



**UNIL** | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

---

*Year : 2014*

## DÉVELOPPEMENT D'UN PROCESSUS DE QUANTIFICATION ET D'ÉVALUATION DE CARACTÈRES MANUSCRITS : THÉORIE ET APPLICATIONS

Thiéry Alexandre

Thiéry Alexandre, 2014, DÉVELOPPEMENT D'UN PROCESSUS DE QUANTIFICATION ET D'ÉVALUATION DE CARACTÈRES MANUSCRITS : THÉORIE ET APPLICATIONS

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB\_A09E74B674667

### **Droits d'auteur**

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

### **Copyright**

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

FACULTÉ DE DROIT, DES SCIENCES CRIMINELLES ET D'ADMINISTRATION  
PUBLIQUE  
ÉCOLE DES SCIENCES CRIMINELLES  
INSTITUT DE POLICE SCIENTIFIQUE

---

DÉVELOPPEMENT D'UN PROCESSUS DE QUANTIFICATION ET  
D'ÉVALUATION DE CARACTÈRES MANUSCRITS : THÉORIE ET  
APPLICATIONS

---

THÈSE DE DOCTORAT

par

Alexandre Thiéry

Directeur de thèse  
Prof. Franco Taroni

Jury  
Prof. Silvia Bozza  
Prof. Matthieu Schmittbuhl  
Dr. Raymond Marquis

Président du Jury  
Prof. Pierre Margot

Fond National Suisse N° 132515 et N° 144227

LAUSANNE  
(2014)

ISBN 2-940098-66-2  
Série Criminalistique N° LXII

*Unil*

UNIL | Université de Lausanne  
Ecole des sciences criminelles  
bâtiment Batochime  
CH-1015 Lausanne

## IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de M. Alexandre Thiéry, candidat au doctorat en science forensique, intitulée

« Développement d'un processus de quantification et d'évaluation de caractères manuscrits : théorie et applications »

Le Président du Jury



Professeur Pierre Margot

Lausanne, le 20 novembre 2014





UNIL | Université de Lausanne

FACULTÉ DE DROIT, DES SCIENCES CRIMINELLES ET D'ADMINISTRATION  
PUBLIQUE  
ÉCOLE DES SCIENCES CRIMINELLES  
INSTITUT DE POLICE SCIENTIFIQUE

---

DÉVELOPPEMENT D'UN PROCESSUS DE QUANTIFICATION ET  
D'ÉVALUATION DE CARACTÈRES MANUSCRITS : THÉORIE ET  
APPLICATIONS

---

THÈSE DE DOCTORAT

par

Alexandre Thiéry

Directeur de thèse  
Prof. Franco Taroni

Jury  
Prof. Silvia Bozza  
Prof. Matthieu Schmittbuhl  
Dr. Raymond Marquis

Président du Jury  
Prof. Pierre Margot

Fond National Suisse N° 132515 et N° 144227

LAUSANNE  
(2014)



---

# Avant propos

La finalité d'une expertise en écriture manuscrite est typiquement, dans le cadre des sciences forensiques, de comparer des caractéristiques d'écriture afin d'aider à déterminer si des écrits proviennent de la même main ou non. Le principe premier régissant ces expertises est formulé en deux *lois* dites *fondamentales* : 1) un individu ne peut écrire deux caractères à l'identique et 2) deux individus ne peuvent écrire à l'identique (Morris [1]).

Cependant les expertises en écriture manuscrite ont été soumises à de vives critiques dès le début du XX<sup>ième</sup> siècle par des auteurs comme Poincaré et Locard, insistant sur la nécessité de quantification pour l'expression de la valeur probante de l'indice. Plus récemment, en Amérique du Nord, certaines critiques rappellent que les méthodes employées dans les expertises en écriture manuscrite afin de repérer les caractéristiques utiles et les exploiter en vue d'une conclusion pour l'authenticité d'un document n'ont jamais fait l'objet de recherche empirique. Par ailleurs, ils estiment que le principe sur lequel reposent leurs conclusions relève de ce que certains individus qualifieraient du *bon sens*, ce qui implique que cette matière ne peut être considérée comme scientifique ou suivant une démarche scientifique.

Dès lors, il est nécessaire d'apporter de nouveaux éléments qui permettent de solidifier les bases scientifiques des expertises en écriture manuscrite, et qui soient : 1) fiables, robustes et acceptées parmi les membres de la communauté scientifique ; et 2) intelligibles, accessibles et utiles aux examinateurs et aux différentes professions juridiques. Ainsi, dans l'optique de poursuivre les recherches empiriques en écriture manuscrite selon une démarche scientifique, il semble important de développer des outils permettant de quantifier les caractéristiques de l'écriture manuscrite. A noter que la quantification de caractéristiques de l'écriture est, encore aujourd'hui, un aspect peu utilisé faute de moyens et de méthodes perfectionnés pour l'extraction et l'analyse de l'information.

Ce travail de recherche s'inscrit dans cette démarche. Il présente pour objectif de sélectionner des caractéristiques discriminantes de l'écriture manuscrite, et de les quantifier par de nouvelles méthodes d'extraction de l'information dans le but de discrimination ou d'identification de scripteurs. En d'autres termes, cette étude vise à 1) développer une nouvelle méthode permettant de quantifier l'écriture manuscrite et 2) proposer une méthode statistique pour évaluer ces observations.

Le plan de la recherche est organisé en cinq chapitres :

**Chapitre 1** Ce premier chapitre vise à introduire les sujets relatifs à cette recherche, à savoir : 1) un descriptif de l'examen forensique de l'écriture manuscrite et des caractéristiques discriminantes de l'écriture d'un scripteur ; 2) une description du processus de quantification et d'évaluation de l'écriture manuscrite par des méthodes informatiques,

ainsi qu'une revue des méthodologies et études disponibles dans la littérature scientifique. Ce chapitre se termine par une section portant sur les objectifs de la recherche, à savoir le développement d'un processus de quantification et d'évaluation de l'écriture, la validation de ce processus et son application à des exemples d'intérêt forensique.

**Chapitre 2** L'ensemble des étapes de la méthodologie développée dans le cadre de cette recherche est décrit dans ce chapitre. Cette étude se concentre sur le développement d'une nouvelle méthode d'extraction de l'information, à savoir la quantification du polymorphisme des caractères manuscrits. En revanche, la méthodologie proposée dans cette recherche couvre l'ensemble des étapes permettant de comparer des échantillons d'écriture manuscrite, soit 1) le pré-traitement des caractères, 2) l'extraction de l'information, et 3) l'évaluation des observations.

**Chapitre 3** Dans ce chapitre les trois expériences (expérience 1, 2 et 3) mises en place dans cette recherche sont présentées. Ces expériences ont pour but de répondre aux deux dernières phases de la recherche, à savoir 1) la validation de la méthode de quantification et d'évaluation de l'écriture, qui a été développée dans cette recherche, et 2) l'application de cette méthode à des exemples dans le domaine forensique.

**Chapitre 4** Les résultats découlant des expériences 1, 2 et 3, décrites dans le chapitre 3, sont présentés dans ce chapitre.

**Chapitre 5** Les discussions relatives aux trois phases de la recherche (développement, validation et application) sont données dans ce chapitre. Une conclusion générale de la recherche clôt ce chapitre.

## Notes

Face à la quantité de résultats numériques et graphiques découlant des expériences (voir Chapitre 3), il a été décidé de ne pas inclure les annexes dans ce texte. La totalité des résultats numériques et des figures est disponible sous format électronique.

## Remerciements

J'exprime mes profonds remerciements à mon directeur de thèse, le professeur Franco Taroni pour l'aide compétente qu'il m'a apporté.

J'adresse également mes très sincères remerciements aux membres du jury aussi pour m'avoir fait l'honneur de participer au Jury de soutenance.

Le Fond National Suisse pour la Recherche Scientifique, qui a eu foi en ce projet et a accordé son soutien financier (Fonds N° 132515 et N° 144227).

---

# Table des matières

|                                                                        |            |
|------------------------------------------------------------------------|------------|
| <b>Avant propos</b>                                                    | <b>i</b>   |
| <b>Notes</b>                                                           | <b>ii</b>  |
| <b>Remerciements</b>                                                   | <b>ii</b>  |
| <b>Table des matières</b>                                              | <b>iii</b> |
| <b>1 Introduction</b>                                                  | <b>1</b>   |
| 1.1 L'examen forensique de l'écriture manuscrite . . . . .             | 1          |
| 1.1.1 Caractéristiques de l'écriture . . . . .                         | 2          |
| 1.1.2 Évaluation de l'écriture manuscrite . . . . .                    | 4          |
| 1.1.3 Critiques des expertise en écriture manuscrite . . . . .         | 4          |
| 1.2 État de l'Art . . . . .                                            | 6          |
| 1.2.1 Généralités . . . . .                                            | 7          |
| 1.2.2 Études de Srihari . . . . .                                      | 9          |
| 1.2.3 Études de Marquis . . . . .                                      | 11         |
| 1.2.4 Etat de la recherche . . . . .                                   | 13         |
| 1.3 Objectifs de la recherche . . . . .                                | 14         |
| 1.3.1 Développement d'outils de quantification de l'écriture . . . . . | 14         |
| 1.3.2 Validation du processus . . . . .                                | 16         |
| 1.3.3 Applications de la méthodologie . . . . .                        | 16         |
| <b>2 Méthodologie</b>                                                  | <b>19</b>  |
| 2.1 Pré-traitement . . . . .                                           | 19         |
| 2.2 Extraction de caractéristiques . . . . .                           | 23         |
| 2.2.1 Quadrants . . . . .                                              | 25         |
| 2.2.2 Notation des pixels de l'image . . . . .                         | 28         |
| 2.2.3 Fonction de mesure . . . . .                                     | 28         |
| 2.3 Dissimilarités entre échantillons . . . . .                        | 31         |
| 2.3.1 DTW classiques . . . . .                                         | 31         |
| 2.3.2 Variantes . . . . .                                              | 34         |
| 2.3.3 Dissimilarités entre échantillons . . . . .                      | 36         |
| 2.4 Analyse de dissimilarité . . . . .                                 | 37         |
| 2.4.1 Analyse de la variance . . . . .                                 | 37         |
| 2.4.2 Analyse de matrice de dissimilarité . . . . .                    | 38         |
| 2.5 Rapport de vraisemblance . . . . .                                 | 40         |

|          |                                                                  |            |
|----------|------------------------------------------------------------------|------------|
| 2.5.1    | Généralités . . . . .                                            | 41         |
| 2.5.2    | Cas de la recherche . . . . .                                    | 42         |
| <b>3</b> | <b>Expériences</b>                                               | <b>43</b>  |
| 3.1      | Expérience 1 . . . . .                                           | 44         |
| 3.1.1    | Jeu de données #1 . . . . .                                      | 45         |
| 3.1.2    | Protocole expérimental . . . . .                                 | 45         |
| 3.2      | Expérience 2 . . . . .                                           | 48         |
| 3.2.1    | Jeu de données #2 . . . . .                                      | 49         |
| 3.2.2    | Protocole expérimental . . . . .                                 | 49         |
| 3.3      | Expérience 3 . . . . .                                           | 51         |
| 3.3.1    | Jeu de données #3 . . . . .                                      | 51         |
| 3.3.2    | Protocole expérimental . . . . .                                 | 52         |
| <b>4</b> | <b>Résultats</b>                                                 | <b>55</b>  |
| 4.1      | Expérience 1 . . . . .                                           | 55         |
| 4.1.1    | Représentations visuelles . . . . .                              | 55         |
| 4.1.2    | Analyse Discriminante Linéaire . . . . .                         | 61         |
| 4.1.3    | Récapitulatif des résultats de l'expérience 1 . . . . .          | 62         |
| 4.2      | Expérience 2 . . . . .                                           | 63         |
| 4.2.1    | La largeur, la hauteur et le rapport $W/H$ . . . . .             | 64         |
| 4.2.2    | Etude de la distribution des $F_{ratio}$ . . . . .               | 67         |
| 4.2.3    | Les résultats du calcul des rapports de vraisemblance . . . . .  | 75         |
| 4.2.4    | Représentation par Analyse en Coordonnées Principales . . . . .  | 83         |
| 4.3      | Expérience 3 . . . . .                                           | 90         |
| 4.3.1    | La largeur, la hauteur et le rapport $W/H$ . . . . .             | 90         |
| 4.3.2    | Les résultats du calcul des rapports de vraisemblance . . . . .  | 91         |
| 4.3.3    | Représentation par Analyse en Coordonnées Principales . . . . .  | 99         |
| <b>5</b> | <b>Discussion</b>                                                | <b>103</b> |
| 5.1      | Validation de la méthode . . . . .                               | 103        |
| 5.1.1    | Invariance à la taille et l'épaisseur . . . . .                  | 103        |
| 5.1.2    | Discrimination de scripteur et taux d'erreur . . . . .           | 104        |
| 5.2      | Exemples d'application . . . . .                                 | 106        |
| 5.2.1    | Influence de la taille des caractères . . . . .                  | 106        |
| 5.2.2    | Variation naturelle de l'écriture : influence du temps . . . . . | 108        |
| 5.3      | Limites et avantages de la méthodologie développée . . . . .     | 110        |
| 5.4      | Conclusion . . . . .                                             | 113        |
|          | <b>Bibliographie</b>                                             | <b>115</b> |
|          | <b>Annexes</b>                                                   | <b>126</b> |
| 5.5      | Expérience 1 . . . . .                                           | 126        |
| 5.6      | Expérience 2 . . . . .                                           | 154        |
| 5.7      | Expérience 3 . . . . .                                           | 291        |

# Introduction

L'objectif de ce chapitre est d'introduire les sujets relatifs à la discrimination de scripteurs par méthode informatique et de présenter les objectifs de la recherche. Ce chapitre est organisé de la manière suivante : dans une première partie (section 1.1), une définition des expertises en écriture manuscrite est donnée. Dans une deuxième partie (section 1.2), l'état de l'art introduit les concepts de quantification d'évaluation de l'écriture et présente une sélection de travaux publiés et d'outils informatiques développés à ces fins. Dans une troisième partie (section 1.3), les objectifs et la démarche de la recherche sont présentés ainsi qu'une structure détaillée du texte.

## 1.1 L'Examen forensique de l'écriture manuscrite

L'écriture manuscrite est caractérisée par un mécanisme défini comme perceptivo-moteur ou neuro-musculaire. C'est un phénomène complexe, dont la réalisation met en jeu la coordination de l'ensemble des mouvements du bras, de la main, et des doigts au travers d'un instrument scripturant (Huber et Headrick [2], Hanlen et al. [3]).

L'apprentissage de l'écriture est un long processus. Chaque individu apprend à écrire selon un modèle, comme ceux trouvés dans les manuels scolaires par exemple. Suite à cette période d'apprentissage le processus de l'écriture devient essentiellement automatique, le rendant moins sujet à une pratique consciente. Selon une publication de 1970 étudiant le développement de la vitesse d'écriture (Cormeau et al. [4]), un individu atteint son pic de progression entre l'âge de 7 et 9 ans, phénomène qui s'atténue progressivement jusqu'à l'âge de 13 ans. Ainsi, l'individu forge peu à peu sa propre écriture, variant de plusieurs degrés du modèle initial. Cependant l'écriture d'un individu ne reste pas figée, chaque scripteur possède sa propre variation, et il n'est pas rare qu'un individu adopte plusieurs styles ou systèmes d'écriture (Bohn [5]). L'écriture manuscrite n'est pas définie seulement par le modèle reproduit, mais aussi par sa réalisation.

Par ailleurs, de multiples facteurs peuvent influencer l'écriture. La condition physique de l'individu (*p. ex.* maladie, fatigue, ...), les conditions d'écriture (*p. ex.* position, luminosité, ...) ou les outils (*p. ex.* type d'instrument scripturant, surface d'écriture, ...) sont autant de facteurs, parmi d'autres, qui sont susceptibles d'influencer l'écriture d'un scripteur (Huber et Headrick [2]).

Le plus souvent, les experts en écriture manuscrite sont confrontés au scénario suivant : sur la base de documents de *référence* (dont l'identité du scripteur est connue) et *indiciaire* (aussi appelé document contesté, soit dont l'identité du scripteur est inconnue), il est demandé de savoir si le scripteur connu a écrit le document contesté, ou si un autre scripteur (appelé aussi scripteur alternatif) a écrit le document contesté. Dans

le contexte d'une expertise, ces deux hypothèses, l'une où le scripteur connu a écrit le document contesté (hypothèse  $H_p$ ) l'autre où un scripteur différent a écrit le document contesté (hypothèse  $H_d$ ) reflètent généralement les hypothèses des parties au procès civil ou pénal. En pratique l'examen forensique de l'écriture manuscrite demande en premier lieu d'extraire par observation les caractéristiques d'écriture, puis de les comparer, et enfin, d'évaluer ces observations au regard des hypothèses d'intérêt. Ce processus est appelé le processus *ACE* (l'analyse, la comparaison, et l'évaluation, voir Huber et Headrick [6, 7]). Ashbaugh [8] étend ce processus en ajoutant une étape de vérification (par un second expert par exemple), soit le processus *ACE-V* (voir Figure 1.1). Ce schéma d'analyse et d'interprétation des écritures est répandu dans le domaine forensique [9, 10].

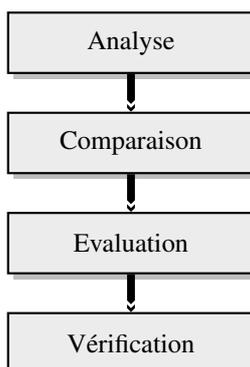


FIGURE 1.1: Illustration du processus *ACE-V* proposé par Ashbaugh [8].

### 1.1.1 Caractéristiques de l'écriture

Les caractéristiques de l'écriture sont formées des habitudes adoptées par le scripteur. Elles sont souvent distinguées en deux catégories : les caractéristiques de classes (ou générales) et les caractéristiques individuelles. Les caractéristiques de classes sont 1) les styles ou 2) les éléments de l'écriture qui sont communs à un groupe d'individus, par exemple une classe sociale. Les caractéristiques individuelles sont les éléments ou qualités de l'écriture qui permettent de différencier chaque individu. De nos jours les caractéristiques de classes perdent progressivement de leur pouvoir discriminatoire au profit des caractéristiques individuelles (Huber et Headrick [2]).

Dans ses travaux, Ellis [11] propose un système à trois étages pour décrire l'écriture manuscrite, sans distinguer les caractéristiques de classes et individuelles. L'unité de base est le graphème. Il s'agit d'une définition abstraite d'une lettre de l'alphabet (*c.-à-d.* sans aucune caractérisation de sa forme), comme la lettre [m] par exemple<sup>1</sup>. Selon cette définition un mot est donc constitué d'un ensemble de graphèmes. Au niveau suivant, chaque graphème est représenté par une série d'allographes, comme les variantes "M", "M", ou "M". Les systèmes d'écriture (d'imprimerie, manuscrit, cursif, *etc...*), induisent un mode de construction particulier pour chaque allographe. Le troisième niveau, un graph, est la représentation graphique de l'allographe sur un support. Ainsi, un allographe peut être produit de manière diverses, que ce soit entre l'écriture de plusieurs scripteurs, ou d'un même individu lors de différentes occasions. En résumé Brault et Plamondon [12] l'expriment ainsi :

---

1. Cette définition peut être étendue aux caractères spéciaux comme le "%" et "&".

« (...) pour un mot donné, chaque lettre apparaît après être passée par trois niveaux de représentation : le graphème (un concept de lettre sans forme précise); l'allographe (représentant un type de lettre précis); et le graph (représentant la séquence ordonnée des mouvements pour former la lettre.) »<sup>2</sup>

A noter que l'écriture est une tâche continue et régulière, et non pas composée d'actions discrètes ou séparées. Bien qu'il y ait des cassures évidentes entre des mots ou des caractères, les mouvements de l'instrument scripturant (*c.-à-d.* la pointe du stylo dans l'espace) sont de manière générale continus et sans cassure, et ne sont pas forcément matérialisés sur le support (Thomassen et Teulings Thomassen et Teulings [13]).

Lors de la comparaison d'écritures, plusieurs caractéristiques similaires peuvent être observées parmi des scripteurs différents. Au contraire, d'autres groupes de caractéristiques auront moins tendance à être dupliqués (Hilton [14], Huber et Headrick [2]). Ce sont ces derniers groupes ou combinaisons de caractéristiques, dans leur ensemble, qui permettent de distinguer un scripteur d'un autre (Osborn [15]). Selon Hilton [16], l'identification de scripteurs peut se décrire comme :

« [les observations des] mêmes caractéristiques distinctives personnelles d'écriture (...) à la fois dans les écritures de référence et indiciaire en nombre suffisant tel que la vraisemblance d'une coïncidence accidentelle soit éliminée (...) »<sup>3</sup>

| Catégorie                  | Éléments                                                                                                                                                                                                                                       |
|----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Éléments de style          | Arrangements, Allographes, Connections, Design et construction des allographes, Dimensions, Inclinaison ou pente, Espacements                                                                                                                  |
| Éléments d'exécution       | Abréviations, Alignements, Commencements et terminaisons, Accents et ponctuations, Embellissements, Lisibilité et qualité de l'écriture, Continuité du tracé, Qualité du tracé, Contrôle de l'instrument scripturant, Mouvements de l'écriture |
| Attributs et habitudes     | Variation naturelle, Persistance                                                                                                                                                                                                               |
| Combinaisons des habitudes | Expansion latérale, Proportion des mots                                                                                                                                                                                                        |

TABLE 1.1: Les 21 éléments, mis en évidence par Huber et Headrick [2], utilisés pour caractériser l'écriture à des fins d'identification.

En revanche cette définition soulève la question de savoir quelles sont les caractéristiques d'écriture qui permettent de discriminer des scripteurs. Huber et Headrick [2]

2. Traduction libre de Brault et Plamondon [12] : « (...) for any given word, each letter appears after having gone through three levels of representation : the grapheme (a concept of a letter without a precise form); the allograph (representing a precise type of letter); and the graph (representing the sequence of appropriate movements for the formation of the letter). »

3. Traduction libre de Hilton [16] : « (...) [the findings of] the same distinctive personal writing characteristics (...) in both known and unknown writing in sufficient number that the likelihood of accidental coincidence is eliminated (...) »

mettent en avant 21 éléments permettant de discriminer les écritures (voir Table 1.1). Cette liste ne se veut ni exhaustive, ni unique. Par exemple les laboratoires du Pays-Bas (Hardy et Fagel [17]) et d’Australie (Found et Rogers [18]) proposent aussi leurs approches, mettant en évidence des différences quant à cette liste. Toutefois, il existe une grande correspondance entre les éléments à considérer dans ces diverses méthodes, et une partie des variations entre ces approches peut s’expliquer simplement par une différence des terminologies employées.

### 1.1.2 Évaluation de l’écriture manuscrite

Depuis l’apparition de l’analyse ADN, le modèle statistique probabiliste utilisé semble avoir été considéré comme le standard pour l’évaluation des observations scientifiques. Dès 1996 Risinger et Saks [19] encouragent l’application de tels modèles probabilistes à d’autres domaines des sciences forensiques. Cependant les modèles probabilistes appliqués à l’indice de nature biologique (ADN) sont basés sur des lois génétiques et peuvent s’appuyer sur la connaissance de fréquences alléliques contenues dans des bases de données de plus en plus nombreuses et riches en profils génétiques. Ces circonstances ne sont pas toujours réunies dans tous les domaines des sciences forensiques (Rudin et Inman [20, 21]). Il faut quand même rappeler que cette méthodologie d’évaluation avait déjà été proposée par Poincaré et al. [22] en 1908 dans le cadre d’une expertise en document dans l’affaire Dreyfus (voir à ce propos Taroni et al. [23]). Il n’est pas sans importance de souligner que déjà dans les années 1886 Bertillon soulignait la nécessité de quantification pour l’expression de la valeur probante de l’indice. Thème repris en 1883 par le même auteur et débattu par Locard en 1920 et 1936 dans le domaine spécifique des expertises en écriture.

L’évaluation de l’écriture manuscrite peut être considérée comme l’étude de la valeur des similarités ou dissimilarités observées, entre l’ensemble des éléments discriminants, mis en évidence lors de la comparaison d’écriture. Les variations naturelles de l’écriture dépendent de facteurs *contrôlés* ou *non-contrôlés*, tout comme chaque élément discriminant peut être influencé par les allographes ou les caractères qui l’entourent. En d’autres termes, la comparaison d’écritures doit prendre en considération les multiples influences qui permettent d’expliquer les changements, subtils ou flagrants, d’une écriture (Huber et Headrick [2]) avant de dégager une conclusion pour la valeur de ces observations.

Par ailleurs, experts et scientifiques s’accordent à dire que les compétences d’experts en document sont principalement acquises au travers de leur expérience du métier (Lee [24]), et que cette discipline souffre cruellement de manque de recherche empirique pour supporter ses conclusions. Plusieurs critiques dénoncent ce manque d’objectivité et de recherche empirique (Risinger et al. [25]).

### 1.1.3 Critiques des expertise en écriture manuscrite

Une expertise en document est vouée à être intégrée dans un processus judiciaire, et l’expert est susceptible d’être convoqué au tribunal pour y exprimer son témoignage. Dès lors, se soulève la question de l’admissibilité des expertises et des témoignages d’experts au tribunal.

Bien qu’il n’existe pas de cadre juridique concernant l’admissibilité des preuves en Europe, les événements dans le système judiciaire américain ont ouvert et alimentent le débat sur les expertises en écriture. De nos jours, les Cours aux États-Unis sont garantes

du bien fondé et de la pertinence des preuves dans leur tribunal. L'admissibilité du témoignage d'expert est ainsi encadrée par les *Federal Rules of Evidence* [26]. Ces règles spécifient les critères à remplir afin qu'une méthode ou technique scientifique soit acceptée au tribunal. Celles-ci sont basées sur les critères dits de Daubert [27], à savoir :

- la théorie ou la technique scientifique a été testée ;
- elle a fait l'objet d'une critique par des pairs ou d'une publication ;
- le taux d'erreur de la technique est connu ;
- le principe est généralement accepté par la communauté scientifique pertinente.

Ainsi, les Cours requièrent que les experts fassent appel à des méthodes scientifiques transparentes, rigoureuses et validées pour examiner et interpréter les preuves. Les expertises en documents manuscrits, l'une des toutes premières disciplines forensiques à voir le jour, n'en sont pas exemptes.

Au final, le manque de recherche empirique et de méthodologie en écriture manuscrite affecte principalement la crédibilité du témoignage des experts. Avant les recommandations liées à *Daubert*, la critique virulente des expertises en écriture manuscrite menée par Risinger, Denbeaux, et Saks [25] suscita un débat animé dans la communauté juridique et d'experts en documents (voir les échanges de Galbraith et al. [28], Moenssens [29] et Risinger et al. [30]). Pour autant, les problématiques du manque de soutien empirique subsistent dans cette période *post-Daubert*, voir à ce propos le rapport de la Cour du District Fédéral de l'Alaska dans l'affaire *United States v. Saelee* en 2001 [31]. Lors du procès la Cour a décidé, au regard des méthodologies et des théories provenant de publications et de conférences, d'exclure l'ensemble du témoignage apporté par l'expert en écriture. Le témoignage de l'expert est considéré, selon les termes employés dans le rapport [31], « ni plus ni moins qu'une série d'observations subjectives » :

*« [la Cour] ne connaît pas les principes de comparaisons que [l'expert] emploierait. Nous ne connaissons pas ce qui est requis pour avoir une similarité valide - qui tendrait à prouver qu'un individu a écrit les deux documents - , ni combien de similarités valides sont nécessaires pour exclure l'éventualité que deux personnes ont essentiellement la même manière d'écrire ces mots en question. En l'absence de principes validés pour effectuer des comparaisons, le témoignage de Mr. Cawley portant sur les similarités ne serait ni plus ni moins qu'une série d'observations subjectives et peu différent d'une opinion infondée sur l'auteur d'un document. Le témoignage de Mr Cawley est tout aussi probable d'induire en erreur ou d'aider le jury à déterminer les faits de cette affaire. Il est par conséquent exclu dans sa totalité. »<sup>4</sup>*

La décision de rejeter le témoignage de l'expert en document ne signifie pas que les tous les témoignages d'experts en écritures soient à l'avenir exclus, ni que les expertises

---

4. Traduction libre de *United States vs. Saelee* [31] « We do not know what principles of comparison Mr. Cawley would employ. We do not know what is required to have a valid similarity — one tending to prove that one person wrote both documents — nor do we know how many valid similarities are necessary to rule out the possibility that two people have essentially the same manner of writing the words in question. The Government would offer a witness as to similarities in handwriting who would speak with apparent authority on a subject in which he would appear to have special knowledge which the court is unable to evaluate on the present record. In the absence of tested principles for making comparisons, Mr. Cawley's testimony as to similarities would itself be nothing more than a set of subjective observations and little different from an unsupported opinion as to the fact of authorship of a document. Mr. Cawley's testimony is as likely to mislead a jury as to assist it in determining the facts of this case. It is therefore excluded entirely. »

en écriture manuscrite soient toutes inadmissibles. L'argumentation de la Cour souligne le manque d'un processus de comparaison précis ainsi que le manque de recherche scientifique dans la méthodologie de l'expert, lui permettant de quantifier des aspects d'importance certaine pour aboutir à une conclusion justifiable. De ce fait, ce rapport met en évidence le besoin de méthodes de quantification de caractéristiques de l'écriture afin de soutenir la recherche empirique et le travail des experts en écriture manuscrite.

## 1.2 État de l'Art

Suite aux critiques visant les expertises en écriture manuscrite et aux divers cas où elles ont échoué à s'établir en tant qu'élément probatoire (voir Risinger [32]), il apparaît nécessaire d'entreprendre des projets de recherche afin d'apporter dans le cadre d'une démarche scientifique, une série d'études empiriques visant la collecte d'informations pertinentes sur lesquelles les experts peuvent s'appuyer pour étayer leurs conclusions. Ainsi de nombreuses études, dans des domaines scientifiques disparates, ont développé des systèmes informatiques visant la caractérisation des aspects de l'écriture afin de pouvoir les quantifier.

L'informatique donne donc l'opportunité aux chercheurs de collecter un grand nombre d'informations (caractères manuscrits) et de minimiser l'erreur humaine (de par la fiabilité des algorithmes), en obtenant ainsi une quantité importante de données utiles. Par ailleurs, l'informatique offre une puissance de calcul supérieure à celle d'un expert et permet d'implémenter des méthodes complexes et d'analyser des quantités de données considérables (Teulings et Thomassen [33]). D'autres domaines tels que la Reconnaissance Optique de Caractères (ROC), et plus généralement la Reconnaissance de Formes ont contribué, bien qu'ayant des objectifs différents, à faire évoluer et progresser les connaissances en analyse de l'écriture manuscrite.

Dans la littérature, la majorité des publications se situe dans le domaine de la Reconnaissance Optique de Caractères, soit principalement à discriminer ou segmenter des caractères [34–43]. Au sein des études qui visent à discriminer des scripteurs ou à étudier les variations de l'écriture, deux catégories se distinguent. D'une part, les études qui focalisent leurs efforts sur l'obtention de faibles taux d'erreur. D'autre part, les projets de recherche qui privilégient l'analyse forensique de l'écriture, comme l'influence de facteurs d'écriture sur l'écriture manuscrite (comme l'instrument scripturant, les conditions d'écriture, ...), l'étude de la variation naturelle de l'écriture d'un individu, ou la sélection des caractéristiques probantes permettant de discriminer les écritures.

L'objectif de cette section de ce premier chapitre est de décrire les différences entre ces deux approches, et d'analyser le soutien que la recherche scientifique a su apporter aux experts en écriture manuscrite. Ainsi l'objectif de cette section n'est pas d'énumérer toutes les méthodes de quantification de caractères ou de discrimination de scripteurs dans la littérature. Cet aspect est régulièrement couvert dans des articles (voir, à titre d'exemple, Vanderkolk [9], Tierney [10]).

Le restant de cette section est organisé de la manière suivante. Dans un premier temps les définitions fondamentales et les nomenclatures utilisées pour décrire les fonctions d'extraction de caractéristiques sont décrites. Dans un deuxième et troisième temps, les deux approches sont présentées au travers des critiques des études de Srihari et Marquis, toutes deux ayant entraîné la publication de plusieurs articles. Dans un quatrième temps une vue plus générale de l'état de la recherche en écriture manuscrite est donnée.

## 1.2.1 Généralités

L'analyse de l'écriture manuscrite par méthode informatique peut être divisée en quatre étapes successives : 1) l'acquisition des échantillons sous format numérique, 2) le pré-traitement des échantillons, 3) l'extraction de caractéristiques et 4) l'analyse statistique. Ces étapes sont illustrées dans la figure suivante (Figure 1.2).

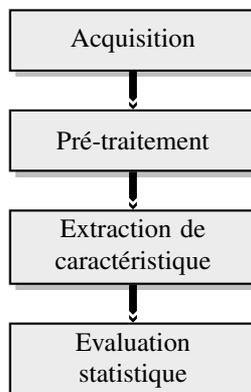


FIGURE 1.2: Schéma d'analyse de l'écriture manuscrite par méthode informatique.

L'acquisition des échantillons d'écriture est faite sous format dynamique (dite *on-line*) ou sous format statique (dite *off-line*). Les données dynamiques sont enregistrées en temps réel au moyen d'une tablette graphique et d'un stylet. L'information des mouvements effectués par l'utilisateur, par exemple la pression et les coordonnées spatiales de la pointe du stylet, est transmise directement à un ordinateur (Verma et al. [44], Buquet [45]). Les données statiques sont les documents numérisés par des scanners ou des appareils photographiques (Plamondon et Lorette [46], Plamondon et Srihari [47]). Ainsi, contrairement aux données dynamiques où les mouvements du stylet sont enregistrés, seul l'*allographe* (la représentation graphique du mot ou caractère sur le support) est disponible.

Un pré-traitement est généralement appliqué aux échantillons. Il s'agit d'une étape purement d'analyse d'image, sur des caractères ou des mots. Les procédés les plus courants sont la réduction du bruit de fond (dû au support lui même, ou la technique d'acquisition), le changement du niveau de couleur (*p. ex.* noir et blanc ou niveau de gris), la segmentation de caractères, ou l'extraction de contours ou de squelettes (Trier et al. [48]) (voir Figure 1.3). L'importance du prétraitement à appliquer aux données est relatif à la technique d'extraction de l'information visée et à la qualité des données (Belaïd et Belaïd [49]).

La méthode d'extraction de caractéristique a comme enjeu principal de quantifier et d'extraire l'information la plus pertinente dans un but de discrimination. Ces méthodes ont donc pour objet d'extraire l'information d'échantillons d'écriture, elles doivent à la fois minimiser l'intra-variabilité et maximiser l'inter-variabilité dans les classes d'écriture (Devijver et Kittler [50], Bunge [51]). Dans le cas de discrimination de scripteurs, analyser les caractéristiques permettant de minimiser la variabilité *intra-scripteur* et maximiser la variabilité *inter-scripteur*, s'avérerait le choix optimal. Les méthodes de caractérisation et de quantification peuvent être classées en différentes catégories :

- *globales* ou *locales*, selon qu'il soit nécessaire (dépendantes du texte) ou non (indépendantes du texte) de connaître la sémantique des mots ou des caractères analysés

(Pecharromán-Bablás [52]) ;

- *conventionnelles* ou *informatiques* : les méthodes conventionnelles sont celles utilisées par les experts en écriture manuscrite. Le plus souvent, elles sont obtenues par un examen visuel, et sont un mélange de mesures quantitatives et qualitatives. Les méthodes informatiques se basent sur un algorithme pour extraire l'information et visent à supprimer la part de choix aléatoires dans le processus lié à la personne (Srihari et al. [53]) ;
- *macro* ou *micro*. On appelle les *macro-informations* les mesures qui peuvent s'appliquer à la totalité de l'échantillon d'écriture (*p. ex.* inclinaison, distribution du niveau de gris). Elles peuvent être extraites au niveau du document, de la ligne, du mot ou du caractère. Les *micro-informations* (*p. ex.* variation du contour, concavité) sont extraites au niveau de l'allographe (Srihari et al. [53]). Vinciarelli [54] propose une autre classification à trois niveaux : *haut-niveau* (extrait au niveau du document, de la ligne, du mot), *moyen-niveau* (extrait au niveau du caractère) et *bas-niveau* (extrait au niveau des segments de caractères) ;

L'évaluation statistique des observations fournies par la méthode d'extraction de caractéristiques se décline en trois catégories : la *classification*, l'*identification*, et la *vérification*. Bien qu'elles aient des méthodologies communes, elles ne partagent pas les mêmes objectifs, comme l'illustre la figure 1.4. La *classification* permet de distinguer des scripteurs en se basant sur des caractéristiques générales de l'écriture (Taylor et Chandler [55]). L'*identification* a pour but de déterminer l'auteur d'un document parmi  $N$  individus présents dans une base de données (Plamondon et Lorette [46], Bensefia et al. [56]). La *vérification* vise à accepter ou rejeter l'appartenance d'un texte à un scripteur, sur la base de comparaisons avec des documents de référence (Huber et Headrick [2], Plamondon et Lorette [46], Gupta et McCabe [57]).

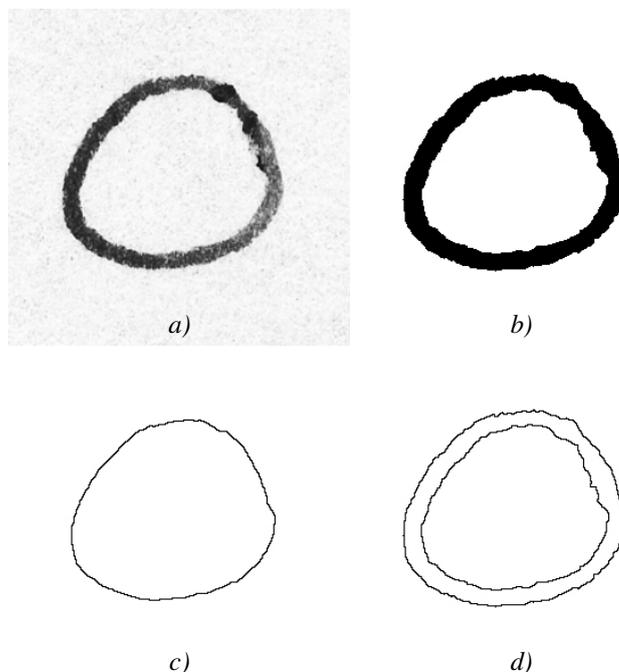


FIGURE 1.3: Exemples de pré-traitement de caractères manuscrits : a) image en niveau de gris, b) image binarisée, c) squelette et d) contour du caractère.

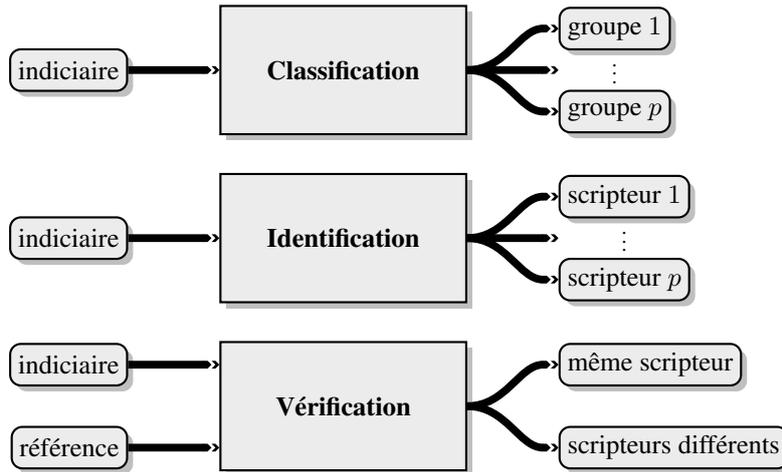


FIGURE 1.4: Schéma présentant les trois objectifs de l'étape d'évaluation de l'écriture, à savoir la *classification*, l'*identification*, et la *vérification* de scripteur.

| Caractéristiques                          | Document | Paragraphe | Mot |
|-------------------------------------------|----------|------------|-----|
| Entropie du niveau de gris                | •        |            |     |
| Seuil de binarisation                     | •        |            |     |
| # de pixels noirs                         | •        | •          | •   |
| # de contours (int./ext.)                 | •        | •          | •   |
| # de composantes de pentes (4-directions) | •        | •          | •   |
| Hauteur moyenne                           | •        | •          | •   |
| Inclinaison moyenne                       | •        | •          | •   |
| Rapport longueur/hauteur                  |          | •          |     |
| Largeur de la marge                       |          | •          |     |
| Longueur                                  |          |            | •   |
| Rapport de la partie supérieure           |          |            | •   |
| Rapport de la partie inférieure           |          |            | •   |

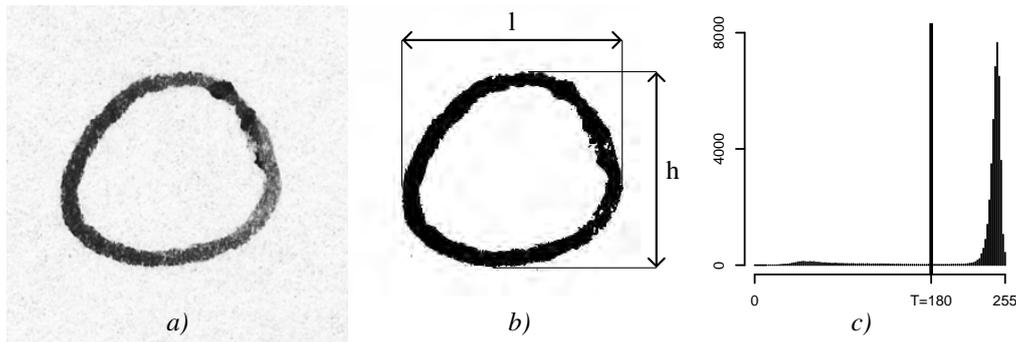
TABLE 1.2: Caractéristiques aux trois niveaux de décomposition (document, paragraphe et mot), utilisé dans l'étude de Srihari et al. [53] (Tableau emprunté de [53]).

## 1.2.2 Études de Srihari

Le but de la recherche de l'équipe de Srihari est d'étudier de manière systématique le principe fondamental des expertises en écriture manuscrite. A savoir, la notion intuitive que l'intra-variabilité de l'écriture des personnes est inférieure à l'inter-variabilité présente dans l'écriture d'un groupe de personnes.

Dans ces études plusieurs caractéristiques de l'écriture manuscrite sont proposées, un jeu de 13 *macro* et 3 *micro* caractéristiques. Le jeu de *macro* caractéristiques (voir Table 1.2.2) est un mélange de caractéristiques *conventionnelles* (par exemple la hauteur des caractères) et de caractéristiques *informatiques* comme le nombre de pixels noirs du caractère. L'étude utilise ces caractéristiques à différents *niveaux*, à savoir au niveau du document, des paragraphes ou des mots. Trois *micro* caractéristiques sont utilisées, ce sont toutes des caractéristiques *informatiques*, en tout elles forment une chaîne binaire

de 512 octets qui exprime les caractéristiques de gradient, de structure et de concavité.



| <i>Conventionnelle</i>                | <i>Informatique</i>                 |
|---------------------------------------|-------------------------------------|
| Longueur ( $l = 29mm$ )               | Seuil de binarisation ( $T = 180$ ) |
| Hauteur ( $h = 26mm$ )                | # de pixels noirs (7630)            |
| Ratio hauteur/longueur ( $r = .897$ ) | Entropie ( $H = 712433.7$ )         |

*d)*

FIGURE 1.5: Exemple d'extraction de caractéristique, selon la méthodologie de Srihari et al. [53]. *a)* image originale, *b)* image binarisée, et ses mesures de longueur et de hauteur, *c)* histogramme des pixels en niveau de gris avec le seuil de binarisation  $T$ , et *d)* tableau des caractéristiques.

L'évaluation des observations dans cette série d'articles est effectué par des tâches de *vérification* et d'*identification* de scripteur. La statistique est donnée sous la forme d'un rapport de vraisemblance, où pour chaque caractéristique extraite est assignée une loi de probabilité. Ces distributions sont considérées comme étant indépendantes l'une de l'autre. Cet aspect d'indépendance peut être critiqué du fait qu'à l'intérieur d'une écriture donnée ces corrélations ont une influence naturelle (par exemple, l'inclinaison influence le rapport longueur/hauteur)

Dans une première étude [53], plus de 1500 scripteurs ont produit chacun trois copies d'un document d'une page. Les résultats d'identification et de vérification de scripteurs sont donnés sous forme de taux de *faux positifs* et *faux négatifs*. La méthode proposée atteint un taux de vérification de 96%, en utilisant les *macros* caractéristiques sur tout le document en plus de 10 mots sélectionnés manuellement pour les *micros* caractéristiques. Cependant, très peu d'informations concernant la prise d'échantillons, le pré-traitement des images, ou le détail des expériences sont fournis. Dans une deuxième étude [58], l'équipe poursuit ses expériences sur une population de 214 jumeaux. L'étude obtient des taux de vérification de 86.7% pour les jumeaux, contre 95.76% pour les non-jumeaux.

Selon l'équipe de Srihari, leurs études sont un soutien à la thèse de individualité de l'écriture. A noter que selon ces auteurs, le fait que les écritures de jumeaux soient moins différenciables que celles de non-jumeaux tend à démontrer que l'écriture est plus similaire (donc moins discriminante) chez les individus partageant le même patrimoine génétique et le même environnement.

Les auteurs, spécialisés dans le domaine de la Reconnaissance Optique de Caractères, se contentent de fournir des taux d'erreur. Cependant, il faut être conscient des potentiels biais associés à la prise d'échantillons, les mesures et l'analyse statistique.

En effet, bien que la taille des populations de ces études soient largement au-dessus de celles habituellement observées dans la majorité des études en écriture manuscrite,

cette caractéristique se heurte à des limitations évidentes. Bien que le texte à recopier par les scripteurs soit fixé (*contrôlé*), extrêmement peu de facteurs influençant l'écriture, et donc pouvant influencer les résultats obtenus et de ce fait les conclusions, sont sous contrôle dans le processus d'échantillonnage utilisé dans ces recherches. Les échantillons ont été obtenus lors de forums académiques, ou par correspondances. Aucune information supplémentaire quant à l'instrument scripturant utilisé n'a été fournie, de même pour le support, les conditions d'écriture, le protocole pour sélectionner les 10 mots dans l'analyse. Tous ces facteurs sont des évidentes sources de biais, car chacune peut influencer la caractéristique de l'écriture d'une personne. Par exemple, outre le biais introduit par la prise d'échantillon non contrôlée, certains facteurs non considérés ont une influence sur plusieurs des caractéristiques utilisées, comme l'instrument scripturant a une influence sur l'épaisseur du trait. Ainsi, il est donc raisonnable de penser que les mesures de binarisation ou d'entropie sont en partie liées à l'instrument scripturant et/ou le support, et de ce fait induit un biais supplémentaire indiscutable. Du point de vue statistique il est discutable de considérer les variables comme étant indépendantes. Finalement, l'interprétation des résultats statistiques et les conclusions tirées dans ces études sont à prendre avec précaution.

Ces études ont été accueillies de manière mitigée dans la communauté scientifique. Ainsi le premier article de 2002 fut rapidement soumis à de vives critiques (voir le commentaire de Saks [59] et la réponse de l'auteur [60]), notamment concernant la validité de ses conclusions sur l'individualité de l'écriture. En 2008, sans prendre en considération les commentaires exprimés par Saks [60], Srihari affirme ses conclusions sur l'individualité de l'écriture.

Les études de Srihari restent des références incontournables dans le domaine des expertises en documents ; le milieu forensique semble être très peu critique vis-à-vis de ces études. Par ailleurs ces articles, financés par un fond du *National Institute of Justice*, sont fréquemment cités en Cour comme preuve de l'individualité de l'écriture [61, 62]. Ces articles sont souvent nommés comme référence pour la préparation de témoignage d'expert ou pour l'apprentissage dans le domaine des expertises en document manuscrit (par exemple, voir Harrison et al. [63], ou King et Logan [64]).

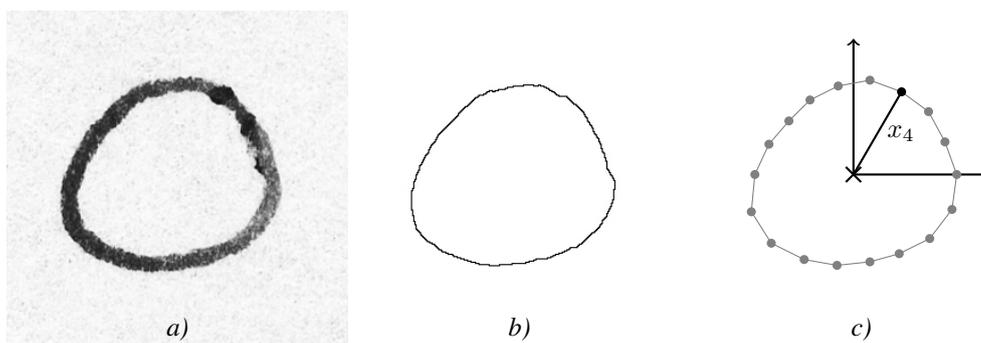
### 1.2.3 Études de Marquis

Le but de la recherche de Marquis, tout comme les études de Srihari, est de donner des éléments en soutien aux *lois fondamentales* de l'écriture. Cependant, au lieu d'utiliser une variété de caractéristiques de l'écriture manuscrite, l'étude propose de quantifier la morphologie des boucles contenues dans les caractères *fermés*<sup>5</sup> à l'aide de la méthodologie connue sous le nom de séries de Fourier. Ainsi les auteurs accordent davantage d'intérêt à la description des variations de l'écriture et l'application de la méthode à des fins forensiques qu'aux taux de *vérification* obtenus.

L'étude décrit de manière précise la méthodologie utilisée (voir Figure 1.6). Chaque boucle contenue dans les caractères échantillonnés est transformée en son squelette. Puis afin d'obtenir une mesure de dimension égale pour chaque échantillon, potentiellement de taille différente, chaque squelette est échantillonné en 128 points, répartis à intervalles réguliers depuis leur barycentre. Les auteurs obtiennent donc pour chaque caractère un signal discret d'indice  $n$ ,  $x[n]$ ,  $n = 0, \dots, 127 = N - 1$ , où  $x[n]$  représente la distance du

---

5. Un caractère, ou une partie d'un caractère, est dit *fermé* lorsque le tracé de l'instrument scripturant forme une figure géométrique fermée, à l'opposé si la figure n'est pas fermée le caractère est dit *ouvert*.



$$\frac{\text{Informatique}}{X_1, X_2, X_3, X_4}$$

d)

FIGURE 1.6: Processus d'extraction de caractéristique, selon Marquis et al. [65]. a) image originale, b) squelette de la boucle, c) échantillonnage du contour (ici avec  $N = 19$ ), et d) tableau des caractéristiques.

point  $n$  au barycentre. Ce signal obtenu traduit la forme de la boucle du caractère. Il est par la suite quantifié par les 4 premières paires de coefficients de sa série de Fourier.

De manière intuitive, les séries de Fourier projettent un signal du domaine spatial dans un domaine des fréquences. Plus généralement, les séries de Fourier permettent de décomposer un signal discret de taille  $N$  (c.-à-d. le signal  $x[n]$ ), en une somme finie de  $N$  fonctions cosinus et sinus appelées harmoniques :  $X[k]$ ,  $k = 0, \dots, N - 1$ . Les harmoniques d'ordre faible représentent les composantes de faible fréquence du signal, et inversement. Ainsi en ne gardant que les premières harmoniques, l'hypothèse faite est que ces composantes sont les plus caractéristiques de la forme de boucles. Pour tout signal discret  $x[n]$ , on appelle transformée de Fourier discrète, l'application :

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-i \times 2nk\pi/N}, \quad (1.1)$$

et son inverse :

$$x[n] = \sum_{k=0}^{N-1} X[k] e^{i \times 2nk\pi/N}. \quad (1.2)$$

La première série d'articles porte sur un jeu de données de 13 scripteurs ayant écrit cinq documents, sur une période de deux semaines. Tous les sujets provenaient du même milieu académique. Les facteurs contrôlés influençant l'écriture sont le texte, l'instrument scripturant et le support. Dans deux premiers articles [65, 66], les auteurs montrent les capacités de discrimination de la méthode sur les caractères  $a$ ,  $d$ ,  $o$  et  $q$ , avec un taux maximum 82.4% pour le caractère  $d$  et minimum de 69.7% pour le caractère  $o$ . Cependant ces études montrent aussi la possibilité de classer les caractères selon leur morphologie ce qui, utilisé à des fins de comparaisons de caractéristiques d'écriture, trouve une application au milieu des experts en écriture. De plus, toujours dans l'optique de fournir une aide scientifique aux experts, une troisième étude [67] a permis d'évaluer l'influence de la taille de l'écriture sur la forme des boucles. Par ailleurs un système de *vérification* de scripteur, utilisant un modèle Bayésien multivarié a été développé (Bozza et al. [68]), faisant appel à la mesure connue sous le nom de *Bayes factor* ou de rapport de vraisemblance. Une

définition plus formelle de cette métrique est donnée dans la section 2.5. Contrairement à la méthodologie employée dans les études de Srihari cette méthode modélise une variation *intra-scripteur* et *inter-scripteur* aboutissant à des résultats montrant des taux de *faux négatifs* de variant de 0.54% à 0.92% et des *faux positifs* variant de 0.14% à 2.75%, suite à des simulations offrant des distributions des valeurs de rapport de vraisemblance permettant ainsi d'évaluer en plus des taux de *faux positifs* et *faux négatifs*, la performance de la méthode par le calcul du taux de *misleading evidence*.

Dans une deuxième série, un groupe de 89 scripteurs a été utilisé à des fins d'*investigation* (Taroni et al. [69]). Par exemple, l'étude propose de tester l'hypothèse qu'il existe, ou non, une différence entre les écritures homme/femme ou droitier/gaucher. Dans cette publication l'aspect d'investigation a été mis en avant par rapport à celui évaluatif. Sur la base d'une écriture la question a été « est-ce que l'expert pourrait offrir une information d'enquête à la police en suggérant l'appartenance du scripteur à une catégorie pré-définie ? ». L'apport investigatif a été mesuré par un rapport de vraisemblance.

L'inconvénient majeur de la méthodologie de Marquis et al. [65] est qu'elle est limitée à des caractères contenant une boucle *fermée* en se focalisant uniquement sur la forme, qui n'est qu'une seule des multiples caractéristiques qu'un expert considère dans son travail. Cependant une unique caractéristique (série de Fourier) décrit la forme d'une boucle, et offre donc l'avantage de pouvoir être assimilé à un processus d'expert.

#### 1.2.4 Etat de la recherche

Les deux recherches présentées dans les sections précédentes (1.2.2 et 1.2.3) illustrent les deux philosophies en analyse forensique de l'écriture. D'un côté, les études qui fonctionnent avec un principe de boîte noire, avec en entrée des documents, des mots ou des caractères, et en sortie une réponse de type binaire (vrai ou faux). D'un autre côté, les études qui étudient les écritures dans l'optique de répondre à des problématiques d'experts. Dans la littérature la proportion de publications étudiant la variation de l'écriture ou l'influence de facteurs [70–75] est très faible comparée à celle qui développe des systèmes de vérification ou identification de scripteurs [76–89]. De même, dans cette faible proportion, les études portent le plus souvent sur l'analyse de signatures ou d'écritures dynamiques. Par exemple, des recherches ont étudié l'influence de conditions de stress (Longstaff et Heath [70]), l'influence de l'alcool (Aşıcıoğlu et Turan [71]), ou l'influence de la position d'écriture sur des signatures (Thiéry et al. [73]).

Les logiciels d'analyse d'écriture manuscrite ont fait leur apparition avec le développement de méthodes de quantification informatique. Ces logiciels sont, en majorité, basés sur des méthodes présentées dans des publications scientifiques. Certains ont trouvé une utilisation quotidienne dans les services de police [90].

Par exemple, *Forensic Information System for Handwriting* (FISH) [91], développé en 1993, est l'un des premiers systèmes informatique dédié à l'analyse de l'écriture. Son rôle est d'assister les examinateurs dans leurs expertises en leur permettant d'annoter leurs échantillons avec des caractéristiques *conventionnelles* et *informatiques*. Au final, au travers d'une base de données d'écriture manuscrite annotée, il permet d'établir un *score* de dissimilarité entre l'échantillon indiciaire et les scripteurs de la base de données. WANDA est le successeur de FISH. Plus récent, il offre des méthodes d'analyses plus modernes et la possibilité à d'autres chercheurs d'implémenter leurs propres méthodes. Dans l'ensemble les deux logiciels partagent la même philosophie. CEDAR-FOX [92] est

un logiciel, développé aux Etats-Unis par l'équipe de Srihari. Comme dans leur recherche, ils y proposent les mêmes méthodes d'extraction et les mêmes statistiques d'*identification* et *vérification*.

Cette brève liste de logiciels met en évidence qu'ils ne représentent pas des travaux de recherche originaux, et devraient être considérés simplement comme des outils informatiques. En effet, ils sont, avant tout autre chose, une implémentation de systèmes ou de procédures publiées au préalable. Ils ont comme avantage de pouvoir calculer des algorithmes impossibles à calculer par un être humain sans support informatique.

Ainsi, force est de constater qu'il subsiste des limites claires en matière de recherche dans le domaine des expertises en écriture manuscrite. Alors qu'un très grand nombre de recherches en écriture manuscrite existent dans la littérature, seule une quantité restreinte de celles-ci visent à quantifier et étudier les variations de l'écriture dans une optique forensique. Ces indicateurs montrent que les recherches en écriture manuscrite ont permis d'apporter un soutien très limité aux experts en analyse de documents contestés.

Il apparaît nécessaire d'entreprendre, au regard des éléments apportés en introduction, des études empiriques visant à soutenir l'examen forensique de l'écriture. C'est-à-dire des études scientifiques qui permettent aux experts en écriture d'étayer leurs conclusions, que ce soit des études qui mettent à disposition des outils pour étudier l'écriture, ou des études qui appliquent ces méthodes pour déterminer l'influence de certains facteurs d'écriture.

## 1.3 Objectifs de la recherche

Dans l'effort de combler une partie des déficiences qui existent dans la recherche en écriture manuscrite, une nouvelle méthode de quantification de l'écriture est développée dans cette étude.

La recherche est organisée en trois phases :

- le **développement** d'une nouvelle méthode d'extraction de caractéristiques pour caractères manuscrits, et la proposition d'un système d'évaluation de ces mesures ;
- la **validation** de cette méthode ;
- et l'**application** de la méthode à des exemples dans le domaine forensique.

La première phase de développement est essentiellement théorique, et fait état d'un chapitre à part entière (Chapitre 2). Pour répondre aux phases de validation et d'application de la méthode proposée trois expériences sont créées, chacune correspondant à un jeu de données spécifique (les expériences 1, 2 et 3 sont décrites dans le chapitre 3). Les résultats de ces expériences sont présentés dans le chapitre 4, enfin les discussions et conclusions de cette recherche sont abordées dans le chapitre 5. Le restant de cette section décrit en détail les trois phases de la recherche.

### 1.3.1 Développement d'outils de quantification de l'écriture

L'outil de quantification développé est divisé en deux phases : 1) le développement d'une nouvelle méthode d'extraction de caractéristiques, et 2) l'intégration de cette mesure dans un système d'évaluation statistique. Le processus développé, soit l'ensemble des deux phases, consiste à reproduire des situations rencontrées lors d'une expertise en écriture manuscrite. Il s'agit de déterminer si le même scripteur a écrit les échantillons

d'écriture du matériel *indiciaire* et de *référence*, ce qu'on appellera un *scénario* fictif. L'approche générale du processus développé dans cette recherche fait appel au processus *ACE* utilisé par les experts en document : les étapes de ce processus sont donc suivies dans le cadre de toute cette recherche (voir Figures 1.1 et 1.4). En reprenant les terminologies employées en Reconnaissance Optique de Caractères (ROC), c'est un processus de *vérification* de scripteur.

Les cinq étapes mises en évidence par les caractères gras, sont celles développées en détail dans le Chapitre 2, dédié aux méthodes utilisées dans cette recherche.

- **Acquisition.** L'étape d'acquisition est fixée à une acquisition *statique* (*off-line*) des documents à l'aide d'un scanner de bureau, afin de reproduire au mieux les conditions lors d'une expertise. Un nombre d'échantillon  $n_s$  est extrait à la fois du matériel *indiciaire* et du matériel de *référence* (le scripteur du matériel de *référence* est connu), soit  $2 \times n_s$  échantillons au total.
- **Pré-traitement.** Le pré-traitement des échantillons est le seul élément du processus de *vérification* de scripteur n'ayant pas de correspondance directe dans le processus *ACE* d'un expert. Cette étape est indispensable pour transformer et assurer la qualité des données (voir section 2.1).
- **Quantification.** La fonction d'extraction de caractéristiques choisie dans le cadre de ce projet calcule un signal permettant de quantifier la morphologie de caractères manuscrits (voir section 2.2). Cette méthode peut être utilisée pour analyser toutes les lettres de l'alphabet. Le polymorphisme des caractères manuscrits est un élément exploité lors de la comparaison d'écritures (Harrison [93] et Hanlen et al. [3]), et grâce aux différences potentielles de formes de caractères qui existent entre scripteurs, il apparaît comme une caractéristique de choix pour étudier et discriminer les écritures de différents scripteurs.
- **Evaluation.** L'étape d'évaluation statistique incorpore les deux étapes de comparaison et d'évaluation du processus *ACE*. Dans le cadre de cette recherche, cette étape est divisée en trois sous-étapes :
  - le **calcul de dissimilarité** entre les caractéristiques extraites. Une métrique de *Dynamic Time Warping* (voir section 2.3) est proposée pour établir une mesure de la dissimilarité entre deux signaux extraits de deux caractères manuscrits ;
  - l'**analyse de dissimilarité**. Une quantité statistique, le  $F_{ratio}$  (section 2.4), est utilisée pour quantifier les différences entre les mesures de dissimilarité des caractères provenant du matériel *indiciaire* et de *référence* ;
  - l'**évaluation**. L'approche classique dans le domaine des sciences forensiques consiste à utiliser une quantité statistique, le rapport de vraisemblance (section 2.5). Cette mesure permet de soutenir l'une des deux hypothèses d'intérêt, par exemple :  $H_p$ , le scripteur connu a écrit le document *indiciaire* ; ou  $H_a$ , un scripteur alternatif a écrit le document *indiciaire*. Les  $F_{ratio}$  observés dans les échantillons d'écriture sont donc évalués par un rapport de vraisemblance.

L'originalité de ce processus est double : 1) plus que de la méthode d'extraction développée, le processus dans son ensemble est composé de méthodes issues de milieux scientifiques différents. A l'image des sciences forensiques, qui sont au croisement de plusieurs sciences. Par exemple le processus utilise des méthodes provenant de la reconnaissance de forme pour la mesure de dissimilarité et d'outils statistiques dédiés à l'étude de systèmes écologiques ; 2) ce processus offre l'opportunité d'utiliser différents

outils statistiques pour évaluer les mesures. Par exemple, une classification hiérarchique ou une analyse par coordonnées principales peuvent venir compléter l'analyse établie par le processus de *vérification* (voir Chapitre 3), Ainsi cette démarche offre l'opportunité d'interpréter un même jeu de données avec différents outils.

### 1.3.2 Validation du processus

La finalité de ce processus de *vérification* de scripteur est de discriminer des caractères manuscrits en fonction de leur morphologie. Cette partie est donc consacrée à l'étude, sous des conditions contrôlées, des étapes du processus de *vérification* de scripteur. La validation du processus est répartie dans les trois expériences de la recherche (expériences 1, 2 et 3). Il s'agit :

- d'analyser et de déterminer quelles sont les mesures, parmi celles mesures développées dans la recherche, qui fournissent des résultats capables de discriminer des formes tout en étant robustes face à la taille et l'épaisseur du trait (expérience 1). En d'autres termes, le but est d'étudier l'étape d'extraction de caractéristiques du processus ;
- de déterminer la capacité du processus de *vérification* à discriminer des scripteurs (expériences 2 et 3). Il s'agit de vérifier que les valeurs extraites aux étapes de calcul du dissimilarité, d'analyse de dissimilarité et d'évaluation statistique 1) soient conformes à la théorie développé dans le Chapitre 2 et 2) permettent de minimiser l'*intra-variabilité* et maximiser l'*inter-variabilité* de l'écriture de scripteurs.

### 1.3.3 Applications de la méthodologie

Dans une dernière phase, deux exemples d'application du processus de *vérification* dans le domaine forensique sont présentés. Ces exemples sont des expériences qui permettent d'analyser des facteurs d'écritures. Il s'agit d'étudier l'influence :

- de la dimension des caractères sur leur morphologie (expériences 1 et 2). Les proportions des caractères ou leurs dimensions relatives, sont des caractéristiques fréquemment utilisées par les experts en écriture manuscrite (Huber et Headrick [2]). Il est alors intéressant de savoir si ces variables peuvent influencer le processus de *vérification* de scripteur. Dans le contexte d'une expertise, un expert peut être amené à déterminer si les mêmes caractéristiques morphologiques peuvent être retrouvées dans des caractères provenant du même scripteur mais avec des dimensions différentes. Sur ce sujet, deux courants sont observés dans la littérature. D'un côté, ceux qui considèrent que les proportions, et donc la morphologie, des caractères est invariante à la dimension des caractères (Harrison [93] et Wann et Nimmo-Smith [94]). De l'autre côté, ceux qui soutiennent que la dimension des caractères est un facteur significatif de la morphologie (Marquis et al. [67], Rogers et Found [95]). Les résultats des expériences permettront de soutenir l'un de ces courants ;
- du temps sur la morphologie des caractères manuscrits (expériences 2 et 3). Dans le contexte d'une expertise, un expert peut-être amené à savoir si les mêmes caractéristiques morphologiques peuvent être retrouvées dans des caractères provenant du même scripteur mais rédigés à des périodes temporelles différentes. Dans le contexte de recherche en écriture manuscrite, le chercheur peut par exemple, s'assurer que les mesures, la méthodologie ou la prise d'échantillon utilisés dans sa recherche ne

sont pas biaisées par l'influence du temps. La variation naturelle de l'écriture, et son évolution dans le temps sont des phénomènes établis et connus des experts en écritures. Par ailleurs, cette variation temporelle a été démontrée pour les signatures *on-line* dans une étude récente de Thiéry et al. [73]. Les résultats de cette expérience permettront de savoir s'il existe une variation naturelle de l'écriture manuscrite en fonction du temps. Ainsi, dans le cas où cette hypothèse apparaîtrait comme vraisemblable, il sera possible d'établir des recommandations pour la prise d'échantillon d'intérêt forensique.



# Méthodologie

Les étapes de la méthodologie développée pour cette recherche sont décrites dans ce chapitre. Cette méthodologie couvre les étapes une fois la prise d'échantillon effectuée (caractères numérisés) jusqu'à l'évaluation statistique (voir section 1.3.1). Le processus développé a pour but, à partir d'échantillons d'écriture provenant de matériel *indiciaire* et de *référence*, de fournir une valeur statistique permettant d'évaluer les observations faites et ainsi soutenir l'hypothèse  $H_p$  : le scripteur connu (soit le scripteur du document de *référence*) a écrit le document *indiciaire* ; où l'hypothèse  $H_d$  : un autre scripteur a écrit le document *indiciaire* (voir Figure 2.1).

Comme indiqué en introduction, dans ce processus des échantillons d'écriture du matériel *indiciaire* et de *référence* analysés forment un *scénario* fictif, en référence aux situations rencontrées lors d'une expertise en écriture manuscrite. On l'appellera alternativement *processus de vérification*, en reprenant les terminologies employées en Reconnaissance Optique de Caractères (ROC). Par ailleurs, il est important de noter que, ce processus étant applicable à tous types de caractères manuscrits (ou plus généralement à une image binarisée), il a été décidé de présenter les échantillons utilisés dans cette recherche dans le chapitre 3.

Le processus de *vérification* est articulé selon les cinq étapes décrites en introduction (voir section 1.3) : le pré-traitement, l'extraction de caractéristiques, la mesure de dissimilarité, l'analyse de dissimilarité et l'évaluation. Les sections qui suivent abordent en détail ces éléments, et permettent étape par étape de découvrir le processus développé.

## 2.1 Pré-traitement

Dans le contexte de cette recherche, le rôle du pré-traitement est d'obtenir pour chaque caractère numérisé une image en noir et blanc de ses contours. Le choix d'étudier les contours est discuté dans la section suivante (section 2.2). L'objectif est que l'image en noir et blanc soit une représentation fidèle du caractère numérisé. En d'autres termes, que le pré-traitement altère au minimum la morphologie des caractères, la morphologie étant l'élément d'écriture qui est mesuré dans la recherche. L'information contenue dans l'image est alors divisée en deux niveaux de description : 1) les pixels du caractère en noir et 2) les pixels de l'arrière plan en blanc. L'objectif de cette section est de décrire le processus de binarisation utilisé dans cette recherche.

Dans le cadre des sciences forensiques, la prise d'échantillons, le format de numérisation (couleurs et résolution<sup>1</sup> de l'image numérisé, par exemple) et le pré-traitement s'inscrivent dans une logique qui contraste avec la recherche en Reconnaissance Optique

---

1. La résolution définit la finesse d'une image numérisée, elle relie le nombre de pixels d'une image à ses dimensions réelles. L'unité utilisée est le *points par pouce (ppp)*.

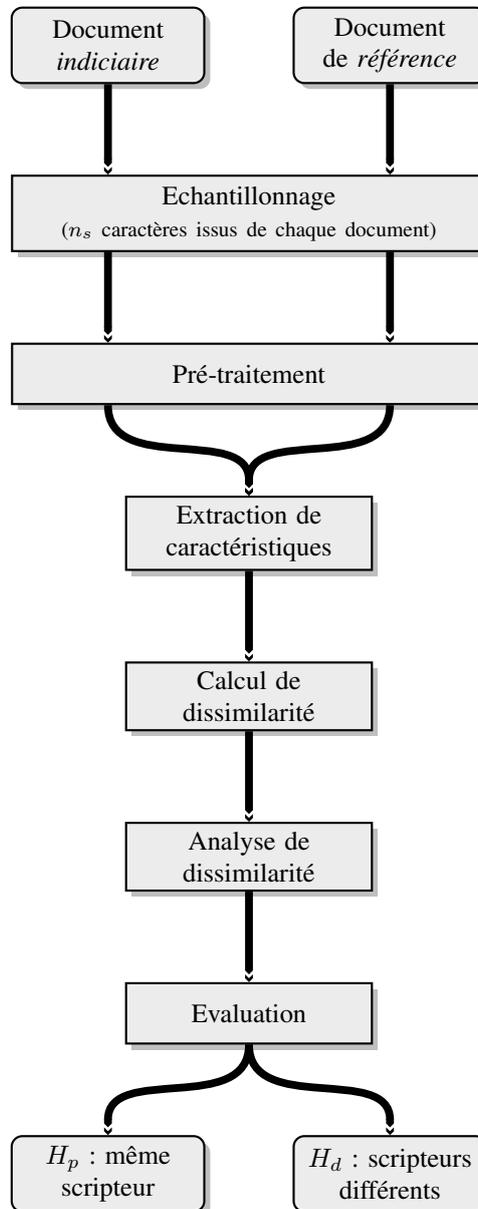


FIGURE 2.1: Illustration des étapes du processus de *vérification* développé dans le cadre de cette recherche. L'objectif de ce processus est, sur la base de caractères provenant du matériel *indiciaire* et de *référence*, d'aboutir à un soutien à l'une des deux hypothèses considérées :  $H_p$  le scripteur du document de *référence* est le scripteur du document *indiciaire*, et  $H_d$  un scripteur différent a écrit le document *indiciaire*.

des Caractères. En effet, dans ce contexte particulier, l'accent est mis sur la capacité des algorithmes à traiter rapidement et en grande quantité des échantillons d'écriture afin de classer et identifier les caractères. De ce fait, dans le domaine de la ROC, les images sont numérisées à une faible résolution et le pré-traitement effectué de manière automatique, le but étant, comme mentionné auparavant, d'identifier des caractères à des lettres. La qualité des données représente par contre un aspect d'extrême importance dans le contexte des sciences forensiques. En effet, la caractérisation précise de l'écriture est fondamentale dans l'expertise en écriture qui vise à mettre en évidence des caractéristiques particulières de l'écriture. Cette contrainte de qualité implique une augmentation des tâches manuelles afin de préserver l'aspect visuel, et donc les caractéristiques morphologiques, des caractères lors du processus de binarisation. Ainsi, ces tâches coûteuses en temps, entraînent une limitation pratique dans la quantité d'échantillons qu'il est possible d'extraire et d'analyser.

Ce processus permet d'obtenir une représentation fidèle de chaque caractère d'intérêt et donc n'altère pas l'information contenue, contrairement à une méthode automatique. Cette combinaison de tâches automatiques et manuelles, donne naissance à une approche dite semi-automatique. Cette pratique représente l'avantage d'avoir une méthode de binarisation uniforme et de pouvoir traiter chaque caractère individuellement.

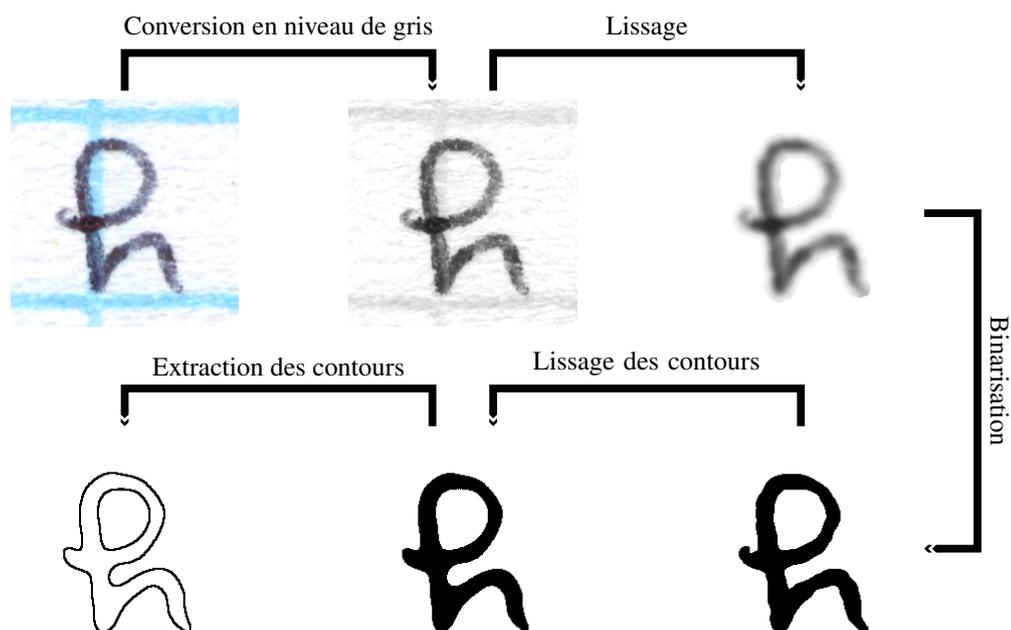


FIGURE 2.2: Illustration des étapes de pré-traitement pour obtenir les contours sont lissés à partir d'un caractère numérisé.

Un caractère numérisé est représenté par l'image  $\mathbf{f}$  de largeur et hauteur  $W_f$  et  $H_f$ , respectivement. On note alors  $f[i, j]$ , le pixel  $(i, j)$ ,  $0 \leq i < W_f$  et  $0 \leq j < H_f$  de l'image  $\mathbf{f}$ . L'algorithme de binarisation proposé transforme une image de couleur  $\mathbf{f}_{\mathbf{RVB}}$ <sup>2</sup> en une image en noir et blanc  $\mathbf{f}_{\mathbf{BW}}$  en effectuant cinq opérations successives (voir Figure 2.2) :

- **Conversion en niveau de gris.**

2. RVB : espace de couleur contenant les trois composantes rouge, verte, et bleue. Chaque pixel  $(i, j)$  de l'image a donc une composante rouge  $f_R[i, j]$ , verte  $f_V[i, j]$  et bleue  $f_B[i, j]$ .

L'image en niveau de couleur  $\mathbf{f}_{\text{RVB}}$  est transformée en niveau de gris. L'image en niveau de gris,  $f_G[i, j]$  est la somme pondérée des composantes de couleur, selon :

$$f_G[i, j] = R \times f_R[i, j] + V \times f_V[i, j] + B \times f_B[i, j] \quad (2.1)$$

où  $(R, V, B)$ ,  $R + V + B = 1$  sont les coefficients de pondération des composantes de couleur. Ces coefficients sont à ajuster en fonction des couleurs de l'instrument scripturant et de l'arrière plan (voir Figure 2.3).

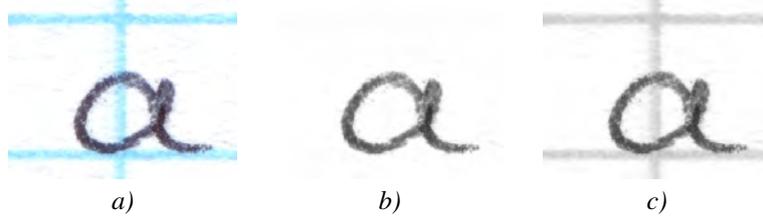


FIGURE 2.3: Illustration de transformation d'image de niveau RVB en niveau de gris. La figure a) représente l'image originale en couleur, et les figures b) et c) les images en niveau de gris, avec  $(R, V, B) = (0, 0, 1)$  et  $(R, V, B) = (1/2, 0, 1/2)$  respectivement.

- **Lissage.** L'image en niveau de gris est lissée par un filtre rectangulaire de taille  $n = 5$  et de noyau  $k(i, j) = 1/n^2$ . Dans ce cas, la valeur de chaque pixel de l'image filtrée  $f'_G[i, j]$  est égale à la valeur moyenne des pixels dans  $f_G$  dans un  $n$ -voisinage de  $(i, j)$ , soit :

$$f'_G[i, j] = \frac{1}{n^2} \sum_{-n}^n \sum_{-n}^n f_G[i' + i, j' + j]. \quad (2.2)$$

- **Gomme et pinceau.** Dans cette étape, l'utilisateur peut corriger les défauts du caractère à l'aide d'outils de gommages et de pinceaux. Le travail effectué par l'utilisateur dans cette étape cruciale permet de garantir la qualité des données.
- **Binarisation.** L'image  $\mathbf{f}'_G$  est ensuite transformée en noir et blanc en fonction du seuil de binarisation  $T_{BW}$  (avec  $0 \leq T_{BW} \leq 255$ ) :

$$f_{BW}[i, j] = \begin{cases} 0, & \text{si } f'_G[i, j] < T_{BW} \text{ (fond)} \\ 1, & \text{sinon (caractère).} \end{cases} \quad (2.3)$$

- **Lissage de contour.** Dans cette étape (voir Figure 2.4 pour une illustration) les contours du caractère binarisé sont extraits, puis lissés en utilisant une série de Fourier (voir Section 1.2.3 pour une définition des séries Fourier), et enfin recomposés sur l'image.

Prenons l'exemple d'un contour appartenant à un caractère, appelons-le  $C_1$  (un caractère manuscrit binarisé peut avoir plusieurs contours  $C_2, C_3, \dots$ , comme illustré dans la Figure 2.4). Ce contour est donc une chaîne composée de  $n$  pixels, décrivant une forme fermée (*c.-à-d.* que le point de départ de la chaîne est adjacent au point d'arrivée). Cette chaîne peut être représentée par les coordonnées  $x$  et  $y$  des pixels qui la composent. Deux cas se présentent, soit la longueur du contour, quantifiée par  $n$ , est :

- inférieure à 80 pixels (où  $\leq 80px$ ), alors la chaîne est considérée comme trop petite pour être lissée et ne subit pas de modifications ;

b)  $> 80px$ , alors la chaîne est lissée par série de Fourier. En pratique le lissage décompose en deux séries de Fourier les deux vecteurs  $x$  et  $y$ , puis les recompose avec 10% du nombre total de coefficient de Fourier<sup>3</sup>, soit  $(n/2) \times \frac{5}{100}$ .

L'étape est répétée pour tous les contours qui existent dans le caractère, dans le cas de la Figure 2.4, elle est répétée pour les contours  $C_2$  et  $C_3$ .

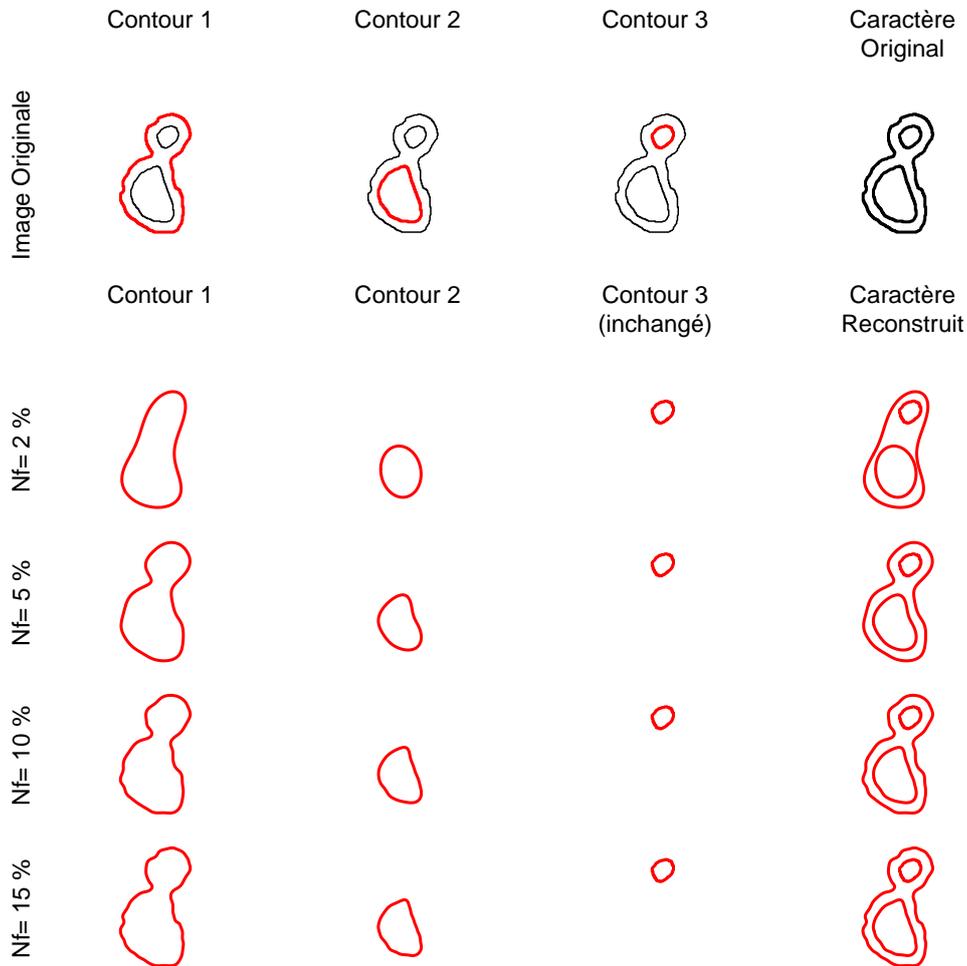


FIGURE 2.4: Illustration de la méthode utilisée pour le lissage de contour des caractères manuscrits.

## 2.2 Extraction de caractéristiques

La méthode de quantification développée pour cette recherche est décrite dans cette section. Elle a pour but d'extraire l'information pertinente contenue dans l'image d'un caractère afin d'en quantifier les caractéristiques morphologiques. Cette méthode est, selon les terminologies introduites dans la section 1.2.1, définie comme une méthode *informatique, locale*, d'extraction de *micro informations* sur des caractères numérisés en noir et blanc.

3. La valeur de 5% fourni des résultats satisfaisant, et a donc été choisi comme standard pour dans cette recherche, voir Figure 2.4.

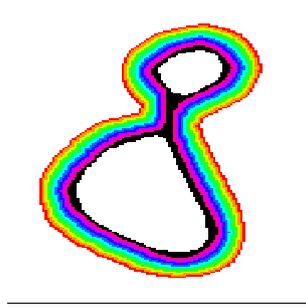


FIGURE 2.5: Exemple des *couches* successives de contours, selon le caractère de la Figure 2.4, représentées en dégradé de couleurs.

Le concept de fonction d'extraction proposée repose sur l'idée que la morphologie d'un caractère est caractérisée par la position relative des pixels qui le composent. Considérons la mesure utilisée par Marquis et al. dans ses travaux : le squelette d'un caractère est en premier lieu quantifié par la distance de ses points, relativement à leur barycentre (*c.-à-d.* le point d'origine). Du fait que la mesure de quantification soit relative à un point d'origine unique, elle apporte une vision incomplète du caractère, en plus d'être restreinte aux seuls caractères contenant des boucles *fermées*. Au contraire, la méthode nouvelle de quantification propose d'utiliser successivement chaque point composant les contours d'un caractère comme centre de mesure. Ainsi, la méthode développée inclut la notion de position relative, et peut être utilisée sur tous les caractères (indépendamment du fait que le caractère ait une boucle *ouverte* ou *fermée*).

La justification d'utiliser les contours du caractère binarisé, au lieu de prendre la totalité de pixels du caractère, est double :

1. les pixels du caractère ne faisant pas partie des contours sont considérés comme de l'information redondante. En effet, en continuant l'exemple du caractère binarisé de la Figure 2.4, considérons les *couches* de contours de la Figure 2.5. Ces contours sont similaires, hormis qu'ils sont produits à des échelles différentes. Or l'objectif étant d'obtenir une mesure caractérisant la morphologie qui n'est pas sensible à la taille et l'épaisseur du caractère, cela implique que cette mesure doit être similaire entre les contours de la Figure 2.5. Par conséquent le fait d'utiliser le caractère *plein* donne lieu à une mesure contenant plus d'information sur l'épaisseur du caractère, *c.-à-d.* l'information portée par le nombre de *couches* de contours.
2. choisir uniquement les pixels composant les contours permet de réduire le temps de calcul. Par exemple, le caractère de la Figure 2.5 la est composé de 5688 pixels, alors que ses contours sont composés de 769 pixels.

Il s'agit d'un processus itératif basée sur la division du plan de l'image en quadrants d'ouverture égale. Dans les prochaines sections de ce texte on donnera une définition géométrique de ces quadrants (section 2.2.1), puis les nomenclatures pour exprimer un caractère binarisé relativement à ces quadrants seront présentés (section 2.2.2). Par la suite, la fonction d'extraction utilisée dans cette recherche pour quantifier l'information de caractères manuscrits, appelée *fonction d'extraction*, sera décrite dans la section 2.2.3.

## 2.2.1 Quadrants

Dans le cadre de la fonction d'extraction de caractéristiques développée, le plan euclidien traditionnel ( $\mathcal{E} = \mathbb{R}^2$ , ayant pour système  $(C, \vec{x}, \vec{y})$ ) est décomposé en  $n_Q$  sous-ensembles distincts appelés quadrants :  $\mathcal{Q}_q^C$ ,  $q = 1 \cdots n_Q$ . Chaque quadrant  $\mathcal{Q}_q^C$ , est caractérisé par son indice  $q$ , son centre  $C$  et son ouverture  $\theta$ <sup>4</sup>(voir Figure 2.6).

De cette manière, les points  $P$  contenus à l'intérieur de ces quadrants sont principalement caractérisés par leur angle relatif au centre  $C$  et l'axe  $\vec{x}$ . De ce fait, il est plus pratique de représenter ces points  $P$  par leurs coordonnées polaires  $(r_C, \phi_C)$ <sup>5</sup>. Plus précisément, un quadrant  $\mathcal{Q}_q^C$  est défini comme l'ensemble des points  $P$  de  $\mathbb{R}^2$  qui sont compris entre les deux demi-droites de centre  $C$  formant un angle avec l'axe  $\vec{x}$  de  $q\theta$  et  $(q+1)\theta$  radians respectivement, soit :

$$\mathcal{Q}_q^C := \forall P \in \mathcal{E}, q\theta_C \leq \phi < (q+1)\theta_C. \quad (2.4)$$

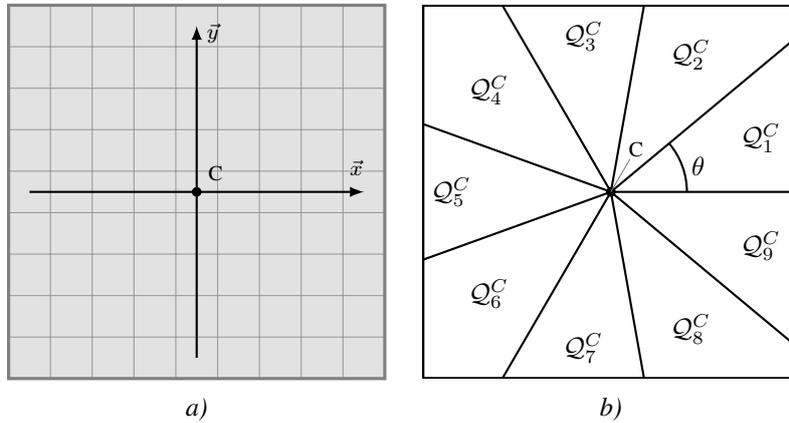


FIGURE 2.6: a) représentation du plan euclidien  $\mathcal{E}$  ayant pour système  $(C, \vec{x}, \vec{y})$ , et b) décomposition du plan  $\mathcal{E}$  en  $n_q = 9$  quadrants (selon la formulation de l'équation 2.4).

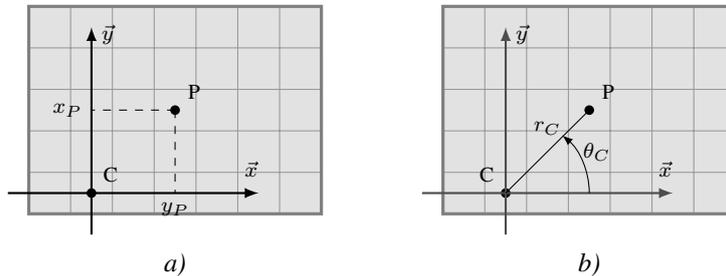


FIGURE 2.7: a) représentation d'un point  $P = (x_C, y_C)$  en coordonnées cartésiennes (soit dans le plan euclidien  $\mathcal{E}$  ayant pour système  $(C, \vec{x}, \vec{y})$ ), et b) en coordonnées polaires  $P = (r_C, \theta_C)$ .

4. A noter que l'angle  $\theta$  est directement lié au nombre de quadrants  $n_Q$ , avec  $\theta = 2\pi/n_Q$ .

5. Dans un système coordonnées polaire les points  $P$  sont définis par leur distance au centre  $r_C$ , et leur angle  $\phi_C$  relatif à l'axe  $\vec{x}$ , voir la Figure 2.7.

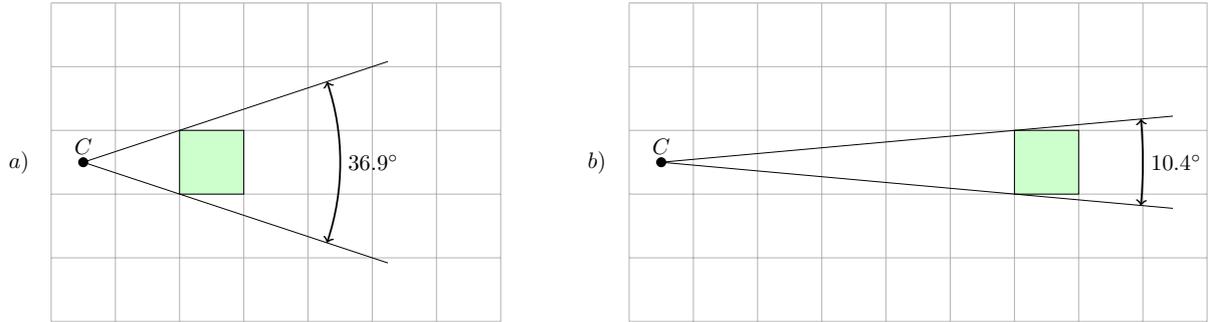


FIGURE 2.8: Illustrations des angles que recouvrent des pixels en fonction de leur distance avec le centre de coordonnées.

Les images de caractères numérisés sont représentées dans des espaces discrets ( $\mathbb{N}^2$ , car échantillonnées en pixels). Ceci entraîne une série de biais dans les mesures utilisant l'angle relatif d'un pixel au point d'origine (l'angle  $\theta_C$ ) :

- Plus la distance entre un pixel et le centre de coordonnées diminue, plus l'étendue que recouvre l'angle entre le pixel et le point d'origine grandit (voir Figure 2.8). Ceci implique qu'un pixel est susceptible d'appartenir à plusieurs quadrants, ou que la précision de mesure de cet angle est trop faible. Etant donné que les images utilisés dans cette recherche proviennent de caractères numérisés, de dimensions de l'ordre à  $300\text{px} \times 300\text{px}$ , ce phénomène ne peut pas être négligé. Les zones où les pixels appartiennent à plusieurs quadrants sont plus nombreuses lorsque ceux-ci sont proches du centre de coordonnées.

Il convient donc d'imposer une distance minimum à l'origine,  $r_{min}$ , pour que les pixels puissent appartenir à un quadrant (*c.-à-d.* les pixels ayant une distance inférieure à  $r_{min}$  n'appartiennent à aucun quadrant), afin de supprimer, en partie, les pixels qui sont problématiques.

- Dans un espace discret certains angles apparaissent plus fréquemment comparés à d'autres (à ce titre, voir la Figure 2.9). En effet, cette figure indique clairement que les angles à  $0$ ,  $45$ ,  $90$  (et par conséquent  $135$ ,  $225$ ,  $270$  et  $335$ ) degrés sont *sur-représentés*. Ceci vient du fait que dans un espace discret, il y a plus de pixels à un angle de  $0^\circ$  (à savoir, tous les pixels où  $x > 0$ ,  $y = 0$ ) qui à  $66.6^\circ$ , par exemple.

Ce phénomène peut donner lieu à des perturbations lorsque l'on applique des mesures sur ces quadrants, comme illustré dans la Figure 2.10 a). Pour limiter l'influence des angles *sur-représentés*, une solution est de placer les pixels correspondant au milieu des quadrants (Figure 2.10 b)). Ainsi en imposant  $n_Q$  multiple de 8 et en posant une phase à l'origine  $\Theta = \theta/2$ , ces angles sont placés au milieu des quadrants.

Au final, un quadrant est défini par quatre paramètres (voir Figure 2.11) :

- (i) le nombre de divisions  $n_Q$ , multiple de 8 (avec  $\theta = 2\pi/n_Q$  et  $\Theta = \theta/2$ ) ;
- (ii) l'indice  $q$ ,  $0 \leq q \leq n_Q - 1$  ;
- (iii) le point d'origine  $C$  ;
- (iv) la distance minimum à l'origine  $r_{min}$ .

|         |         |      |      |      |      |      |      |
|---------|---------|------|------|------|------|------|------|
| 6       | 90      | 80.5 | 71.6 | 63.4 | 56.3 | 50.2 | 45   |
|         | 90      | 78.7 | 68.2 | 59   | 51.3 | 45   | 39.8 |
|         | 90      | 76   | 63.4 | 53.1 | 45   | 38.7 | 33.7 |
| ⋮       | 90      | 71.6 | 56.3 | 45   | 36.9 | 31   | 26.6 |
|         | 90      | 63.4 | 45   | 33.7 | 26.6 | 21.8 | 18.4 |
|         | 90      | 45   | 26.6 | 18.4 | 14   | 11.3 | 9.5  |
| $y = 0$ |         | 0    | 0    | 0    | 0    | 0    | 0    |
|         | $x = 0$ |      | ⋯    |      |      |      | 6    |

FIGURE 2.9: Illustration des angles de pixels relatifs au point d'origine  $(0, 0)$ . Les angles sont donnés en degrés et sont arrondis à une décimale. On aperçoit que les angles 0, 45 et 90 degrés apparaissent plusieurs fois (6 au total), tandis que les autres valeurs (*p.-ex.* 63.4°) n'apparaissent qu'une fois.

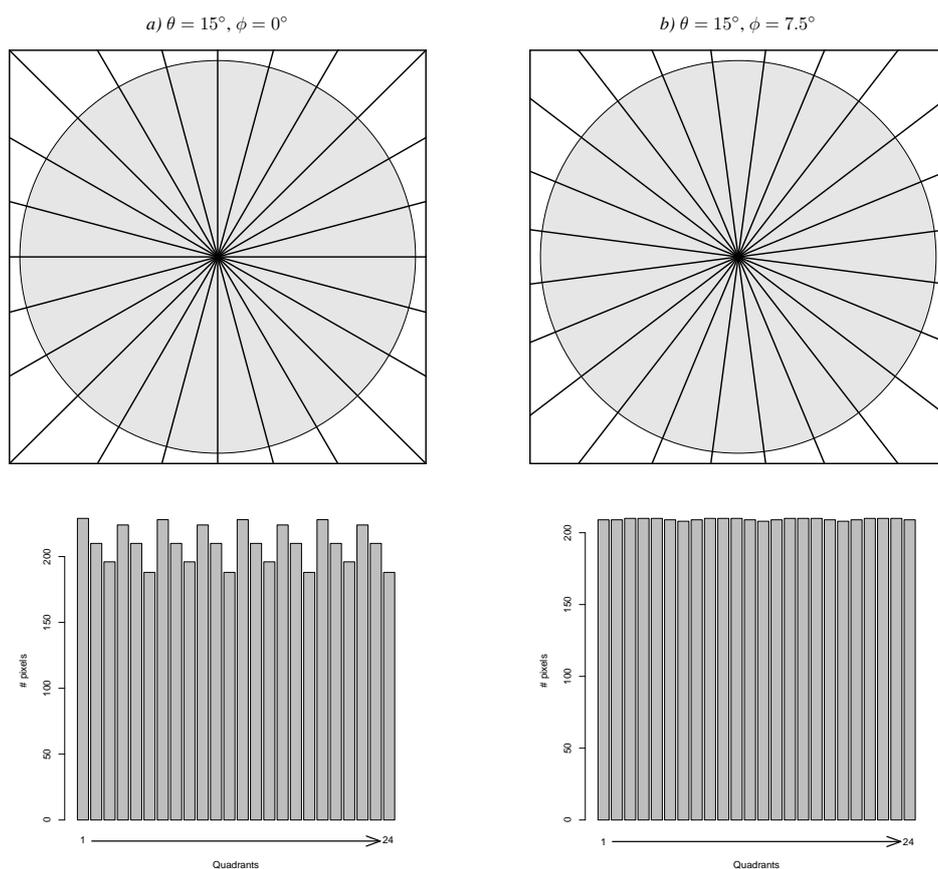


FIGURE 2.10: Illustrations de l'effet des angles *sur-représentés* lors de la mesure du nombre de pixels appartenant à chaque quadrant (pour un rayon de 40 pixels). Deux paramètres jeux sont représentés,  $\theta = 15^\circ, \Theta = 0^\circ$  (colonne a)) et  $\theta = 15^\circ, \Theta = \theta/2$  (colonne b)). Les images du haut illustrent les quadrants, celles du bas le nombre de pixels par quadrant.

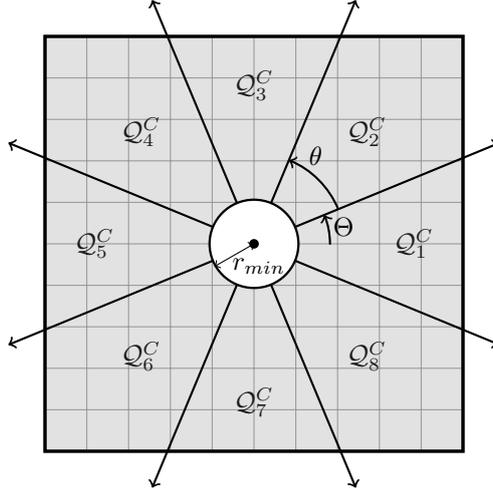


FIGURE 2.11: Illustration des paramètres caractérisant les quadrants  $\mathcal{Q}_q^C$  (à savoir  $n_Q$ ,  $\theta$ ,  $\Theta$  et  $r_{min}$ ).

## 2.2.2 Notation des pixels de l'image

Cette section a pour but d'introduire la notation utilisée dans cette recherche pour définir les pixels d'un caractère numérisé. Précédemment une image noir et blanc en  $\mathbf{f}_{\mathbf{BW}}$  était définie comme une matrice de 0 ou de 1 (section 2.1). Or, la méthode de quantification est basée à la fois sur les contours de caractères en noir et blanc, et sur la division de l'espace en quadrants. Il est donc plus pratique de définir une image comme la liste de points faisant référence aux pixels noirs des contours. Considérons l'exemple de la Figure 2.12 :

- (i) Tous les pixels  $P$  constituant le caractère binarisé sont regroupés dans la liste de pixels  $\mathbf{P}$ . En posant  $n_P$  le nombre des points, cette liste est notée  $\mathbf{P} = (P_1, \dots, P_j, \dots, P_{n_P})$  (voir Figure 2.12-b).
- (ii) Les quadrants étant définis par l'angle relatif d'un point  $P_j$  au centre  $C$ , il est d'intérêt de noter les coordonnées de ces points en coordonnées polaires. On appelle alors  $\mathbf{P}^C$  la liste des points appartenant aux contours avec pour point d'origine  $C$ . On note  $r_j^C$  et  $\theta_j^C$  respectivement la distance et l'angle du point  $P_j$  au centre  $C$  (voir Figure 2.12-c).
- (iii) Enfin, on note  $\mathbf{P}_q^C = (P_{q,1}^C, \dots, P_{q,n_q^C}^C)$  la liste des points qui appartiennent au quadrant  $q$  de centre  $C$ , avec  $n_q^C$  le nombre de pixels dans ce quadrant (voir Figure 2.12 e-f). De manière similaire à l'étape précédente,  $r_{q,j}^C$  et  $\theta_{q,j}^C$  représentent respectivement la distance et l'angle du point  $P_j$  du quadrant  $q$  au centre  $C$ .

## 2.2.3 Fonction de mesure

Dans cette recherche, une *fonction locale*  $f^C(q)$ ,  $q = 1, \dots, n_Q$  est définie comme une mesure qui a pour but d'extraire l'information pertinente des points  $\mathbf{P}_q^C$  contenus dans le quadrant  $\mathcal{Q}_q^C$ , soit le quadrant d'indice  $q$  de point d'origine  $C$ .

Par exemple,  $f^C(q)$  peut mesurer une valeur réelle comme la somme des distances entre les points  $\mathbf{P}$  et le point  $C$  ( $f_1^C(q)$ ) ou leur moyenne ( $f_2^C(q)$ ), ou une valeur discrète

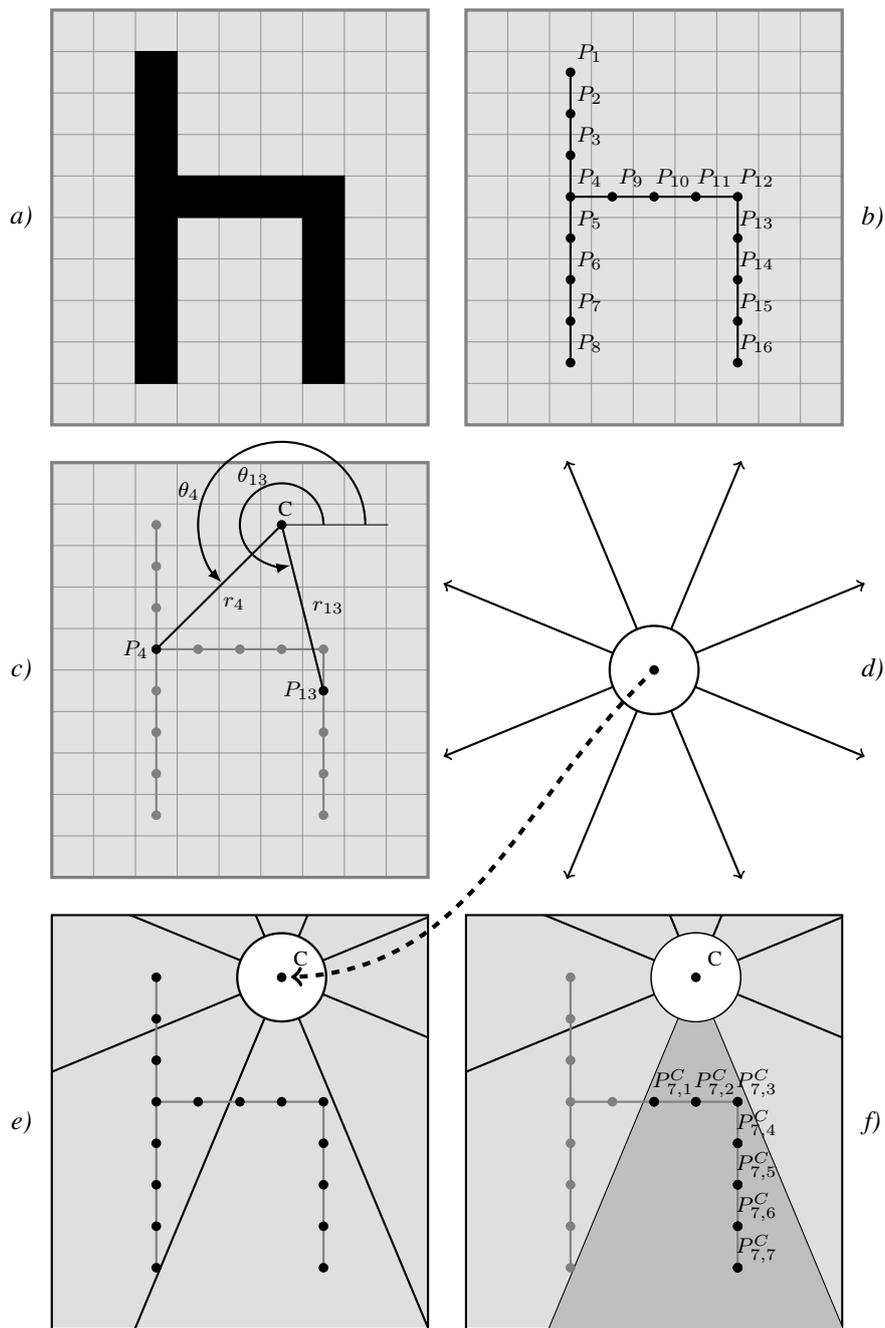


FIGURE 2.12: Illustration de la décomposition d'un caractère binarisé en vecteurs de points. *a)* image noir et blanc, *b)* représentation de la liste de points du caractère contenus dans l'image, *c)* représentation des points en coordonnées polaires et point d'origine  $C$ , *d)* représentation des quadrants ( $n_Q = 8, \Theta = \theta/2$ ), *e)* application du quadrant sur le point  $C$ , *f)* représentation des points appartenant au quadrant  $q = 7$  de centre  $C$ .

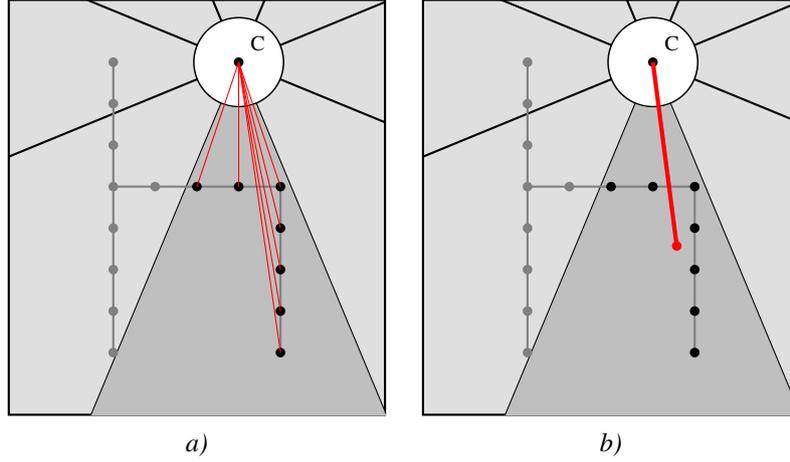


FIGURE 2.13: Illustration des *fonctions locales*  $f_1$  a) et  $f_2$  b) sur le quadrant  $q = 7$  de centre  $C$  ( $n_Q = 8, \Theta = \theta/2$ ).

comme le nombre de pixels dans un quadrant ( $f_3^C(q)$ , voir Figure 2.13) :

$$f_1^C(q) = \sum_{j=1}^{n_q^C} \|\overrightarrow{CP_{q,j}}\| = \sum_{j=1}^{n_q^C} r_{q,j}^C, \quad (2.5)$$

$$f_2^C(q) = \sum_{j=1}^{n_q^C} \frac{\|\overrightarrow{CP_{q,j}}\|}{n_q^C} = \frac{f_1^C(q)}{n_q^C}, \quad (2.6)$$

$$f_3^C(q) = n_q^C. \quad (2.7)$$

La méthode de quantification développée propose d'utiliser successivement chaque point de  $\mathbf{P}$  comme le centre  $C$  de la mesure. La *fonction globale* est donc définie comme la somme des *fonctions locales* pour les  $n_P$  points de  $\mathbf{P}$  :

$$F(q) = \sum_{j=1}^{n_P} f^{P_j}(q), \quad q = 1, \dots, n_Q. \quad (2.8)$$

Comme indiqué précédemment, la *fonction locale* peut prendre plusieurs formes ( $f_1^C(q)$ ,  $f_2^C(q)$  ou  $f_3^C(q)$ , par exemple). Les *fonctions globales* sont classées en deux catégories :

1. Fonctions de quantification : ce sont les *fonctions locales* basées sur une mesure de distance entre les points ( $\|\overrightarrow{P_j C}\|$ ), comme  $f_1^C(q)$  et  $f_2^C(q)$  ;
2. Fonctions de normalisation : ces *fonctions locales* sont basées sur la mesure du nombre de pixels ( $n_q^{P_j}$ ), comme  $f_3^C(q)$ . Elles ont pour but de normaliser les fonctions de quantification.

Au final, la *fonction d'extraction de caractéristiques*  $FE(q)$  (ou *fonction d'extraction*) est définie comme le rapport entre le signal d'une *fonction globale* de quantification et de normalisation (appelées respectivement  $M(q)$  et  $N(q)$ ) :

$$FE(q) = \frac{M(q)}{N(q)}. \quad (2.9)$$

Pour une *fonction d'extraction*, un vecteur de caractéristiques  $(FE(1), FE(2), \dots, FE(n_Q))$  est extrait de chaque caractère. Par construction, les fonctions *globales* (voir page 30) et d'*extraction de caractéristiques* sont des fonctions de quantification invariantes à la translation, mais pas face à la rotation. Dans cette recherche, 6 fonctions de quantification et 3 fonctions de normalisation, soit un total de 18 combinaisons de *fonctions d'extractions* sont étudiées afin d'identifier les fonctions les plus robustes et ayant la meilleure discrimination de formes. Ainsi, à la suite de cette étape, seules les fonctions les plus performantes sont sélectionnées dans le processus de *vérification* de scripteur (voir Section 3.1 – Tableau 3.1 pour la liste des fonctions étudiées).

## 2.3 Dissimilarités entre échantillons

L'objectif de cette section est de comparer les signaux extraits par les *fonctions d'extraction*.

Lors de cette étape, une dissimilarité entre chaque couple d'échantillons est calculée, tout comme un expert qui observe les similarités et les dissimilarités entre les caractéristiques d'écriture. Les signaux  $(FE(1), FE(2), \dots, FE(n_Q))$  extraits de chaque caractère sont directement comparés deux à deux.

La méthode utilisée pour cette étape est le *Dynamic Time Warping* (DTW). Il s'agit d'une méthode qui permet d'aligner deux séquences (ou signaux) de chaque caractère pour former une matrice de dissimilarité  $\mathbf{D}$ . À l'origine créée pour la reconnaissance vocale [96], cette méthode a rapidement rencontré du succès dans d'autres domaines. En science forensique, le DTW est une technique répandue et déjà utilisée dans la vérification ou identification d'écriture et signature *on-line* [97, 98].

Dans un premier temps, la définition et les propriétés du DTW classique sont décrites, puis des exemples de variantes sont donnés. Finalement, une description de l'application du DTW dans le processus de *vérification* développé dans cette recherche est indiquée. Les références [99–101] offrent des détails techniques sur la DTW, ses variantes et ses applications.

### 2.3.1 DTW classiques

Le DTW est une métrique qui permet de comparer et d'aligner deux séquences dans le but de mesurer une distance de dissimilarité entre ces deux séquences. Prenons  $U = (u_1, u_2, \dots, u_n)$  et  $V = (v_1, v_2, \dots, v_m)$  deux signaux échantillonnés à intervalles réguliers, respectivement de longueur  $n$  et  $m$  (voir Figure 2.14). Si chaque élément  $u_i$  et  $v_j$  des signaux  $U$  et  $V$  appartiennent au même espace  $\mathcal{F}$ , alors le DTW est défini comme la fonction :

$$dtw(U, V) : \mathcal{F}^n \times \mathcal{F}^m \rightarrow \mathbb{R}_+, \quad (2.10)$$

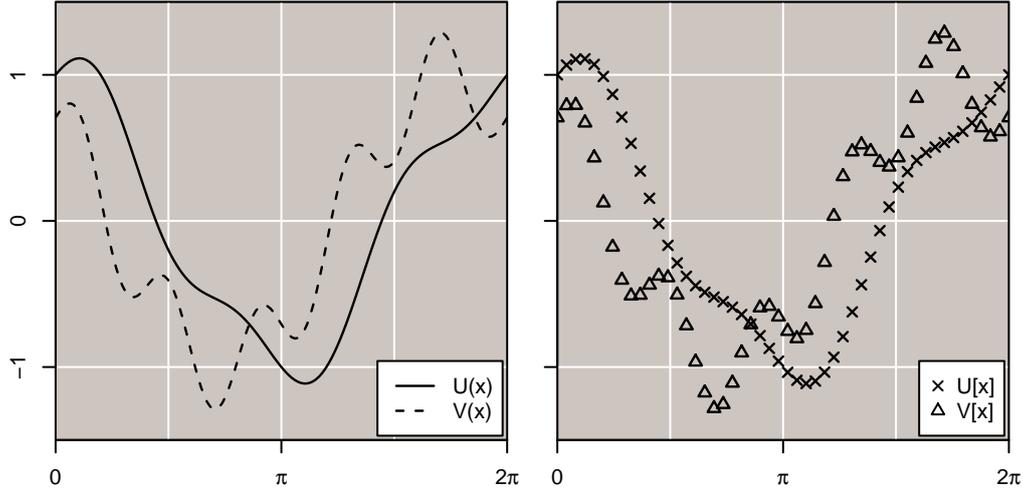


FIGURE 2.14: A droite, exemple de deux fonctions  $U(x) = \cos(x) + 0.2 \times \sin(x/3)$  et  $V(x) = \cos(x + \pi/4) + 0.3 \times \sin(x/5)$  sur  $x \in [0, \pi]$ . A gauche, les deux séquences  $U = (u_1, u_2, \dots, u_n)$  et  $V = (v_1, v_2, \dots, v_m)$  et  $U(x)$  et  $V(x)$  échantillonnés à intervalles réguliers, avec  $n = m = 50$ .

Le principe de l'algorithme de DTW fait intervenir une *fonction de coût*  $c(u, v)$  pour comparer les éléments  $u_i$  et  $v_j$ . A titre d'exemple, les fonctions de coût les plus répandues sont la distance euclidienne  $c(u_i, v_j) = \sqrt{u_i^2 + v_j^2}$  et la distance de Manhattan  $c(u_i, v_j) = |u_i - v_j|$ . Dans le cadre de cette recherche, seule la distance euclidienne est utilisée. La *matrice de coût*  $\mathbf{C}_{n \times m}$  est obtenue en comparant chaque élément de  $U$  et  $Y$  (voir Figure 2.15) :

$$\mathbf{C}_{i,j} = c(u_i, v_j). \quad (2.11)$$

L'objectif est de trouver un chemin qui traverse la matrice  $\mathbf{C}$  de bord en bord, en ayant le moindre coût. Une *séquence d'alignement* (ou chemin) est définie comme le vecteur  $s = (s_1, s_2, \dots, s_L)$  avec  $s_l = (i_l, j_l) \in [1 : n] \times [1 : m], l \in [1 : L]$  satisfaisant les trois conditions suivantes :

- (i) Condition aux frontières :  $s_1 = (1, 1)$  et  $s_L = (n, m)$ .  $\Leftrightarrow$  La séquence commence avec le doublet  $(u_1, v_1)$  et termine à  $(u_n, v_m)$ .
- (ii) Monotonicité :  $i_1 \leq i_2 \leq \dots \leq i_L$  et  $j_1 \leq j_2 \leq \dots \leq j_L$ .  $\Leftrightarrow$  La séquence ne peut pas revenir en arrière,  $i_x \leq i_{x+1}$  et  $j_x \leq j_{x+1}$ .
- (iii) Conditions de pas :  $s_{l+1} - s_l \in \{(1, 0), (0, 1), (1, 1)\}$ .  $\Leftrightarrow$  Contrôle l'avancement de la séquence; pour avancer vers le doublet  $(i, j)$  il faut que l'alignement soit sur l'un des doublets  $(i - 1, j)$ ,  $(i, j - 1)$  et  $(i - 1, j - 1)$ .

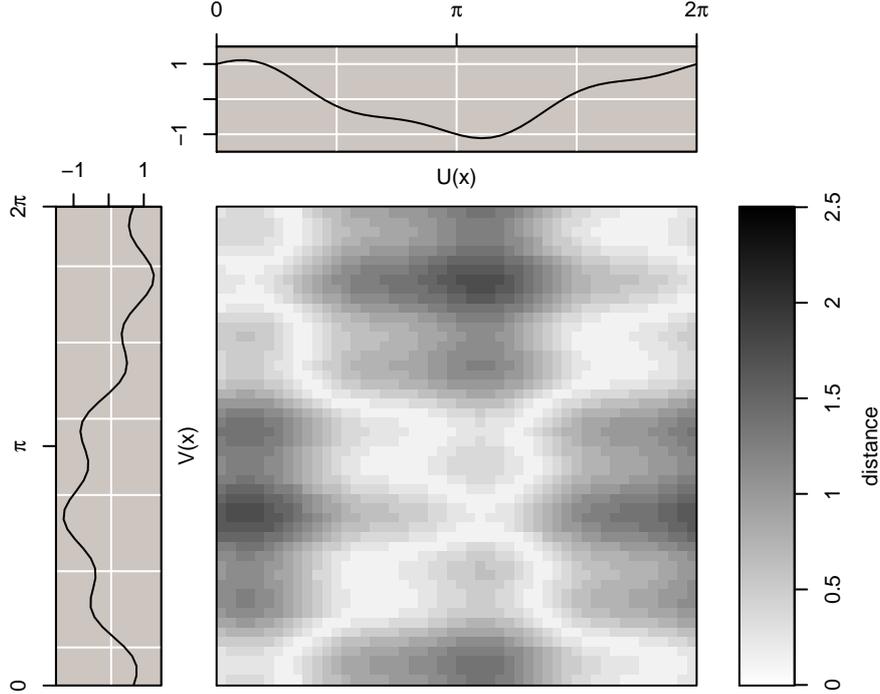


FIGURE 2.15: Illustration de la *matrice de coût*  $\mathbf{C}$  entre les deux séquences  $U = (u_1, u_2, \dots, u_n)$  et  $V = (v_1, v_2, \dots, v_m)$  de la Figure 2.14, obtenue avec la distance euclidienne ( $c(u_i, v_j) = \sqrt{u_i^2 + v_j^2}$ ).

Chaque élément  $s_l = (i_l, j_l)$  représente un alignement entre les éléments  $u_{i_l}$  et  $v_{j_l}$ . Le coût total  $c_s(U, V)$  d'un chemin d'alignement  $s$  entre  $U$  et  $V$  est défini comme la somme des coûts du chemin :

$$c_s(U, V) = \sum_{l=1}^L c(u_{i_l}, v_{j_l}) \quad (2.12)$$

L'alignement optimum  $s^*$  entre  $U$  et  $V$  est le chemin d'alignement ayant le coût  $c_s(U, V)$  le plus bas, parmi tous les chemins d'alignement possibles. La fonction de DTW est alors définie comme le coût total de la séquence d'alignement  $s^*$  :

$$dtw(U, V) = c_{s^*}(U, V) \quad (2.13)$$

$$= \min_s \{c_s(U, V)\} \quad (2.14)$$

Cependant le chemin d'alignement  $s^*$  n'est pas nécessairement une solution unique. De plus, il est évident qu'en pratique, calculer tous les chemins d'alignement possibles entre  $U$  et  $V$  engendre des contraintes de temps trop importantes (temps de calcul). Ainsi, un algorithme est introduit pour permettre de calculer un chemin d'alignement optimal. La *matrice d'accumulation de coût*  $\mathbf{D}_{n \times m}$  est définie comme :

$$\mathbf{D}_{i,j} = dtw(U_i, V_j). \quad (2.15)$$

où  $U_i = (u_1, u_2, \dots, u_i)$  et  $V_j = (v_1, v_2, \dots, v_j)$  sont les sous-séquences de  $U = (u_1, u_2, \dots, u_n)$  et  $V = (v_1, v_2, \dots, v_m)$ . La distance DTW entre  $U$  et  $V$  est obtenue à partir de l'équation 2.15 en posant  $i = n$  et  $j = m$  tel que,  $\mathbf{D}_{n,m} = dtw(U_n, V_m) = dtw(U, V)$ . La procédure de calcul de la matrice  $\mathbf{D}$  est formalisée par les propriétés suivantes :

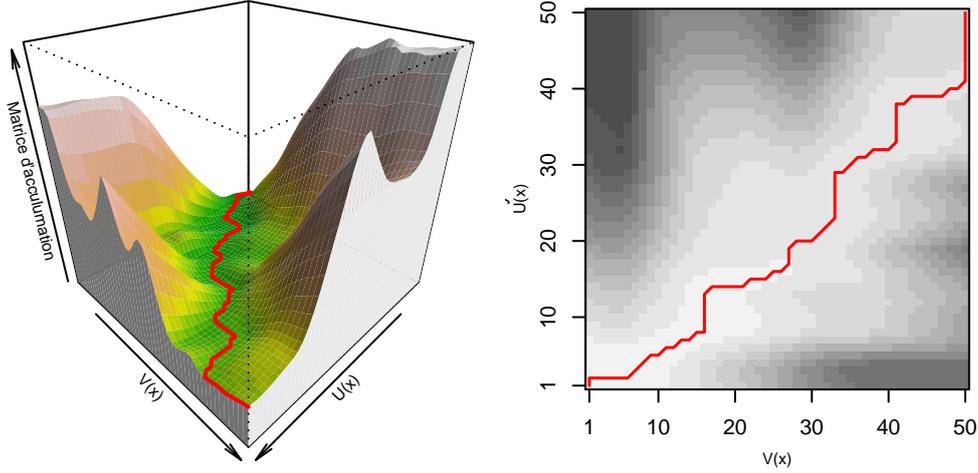


FIGURE 2.16: Illustration en 3 et 2 dimensions, respectivement à gauche et à droite, de la *matrice d'accumulation*  $\mathbf{D}$  et du chemin d'alignement  $s^*$  (en rouge) entre les signaux  $U$  et  $V$  de la Figure 2.14.

$$(i) \mathbf{D}_{i,1} = \sum_{k=1}^i c(u_k, v_1),$$

$$(ii) \mathbf{D}_{1,j} = \sum_{k=1}^j c(u_1, v_k),$$

$$(iii) \mathbf{D}_{i,j} = \min \{ \mathbf{D}_{i-1,j-1}, \mathbf{D}_{i-1,j}, \mathbf{D}_{i,j-1} \} + c(u_i, v_j), \text{ où } i \in [1, n], j \in [1, m].$$

Les conditions (i) et (ii) définissent les conditions aux frontières de la matrice  $\mathbf{D}$ . La condition (iii) régit l'accumulation de la *matrice de coût*  $\mathbf{C}$  dans la matrice  $\mathbf{D}$ . La fonction qui détermine cette accumulation est appelée *fonction de pas*.

### 2.3.2 Variantes

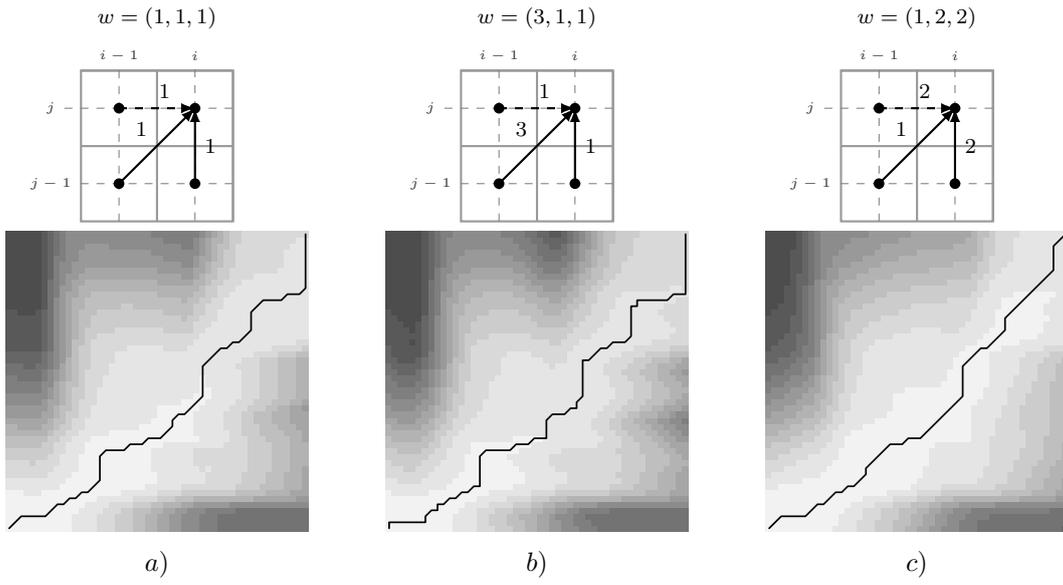
L'algorithme de DTW décrit ci-dessus peut être modifié, que ce soit pour réduire le temps de calcul ou pour mieux contrôler les routes des séquences d'alignement. Cette section présente les trois grandes catégories de variantes DTW les plus répandues : modification des poids, modification de la fonction de pas, modification de la fenêtre de calcul (voir Figure 2.17).

**Modification des poids** Pour favoriser la direction empruntée par le chemin d'alignement (*p. ex.* le chemin rouge dans la Figure 2.16), il est possible d'introduire un vecteur de poids  $w$  dans la fonction de pas (voir Figure 2.17 *a - c*). Dans le cas du DTW classique on définit le poids  $w = (w_d, w_h, w_v)$  pour influencer sur la direction diagonale ( $w_d$ ), horizontale ( $w_h$ ) ou verticale ( $w_v$ ) :

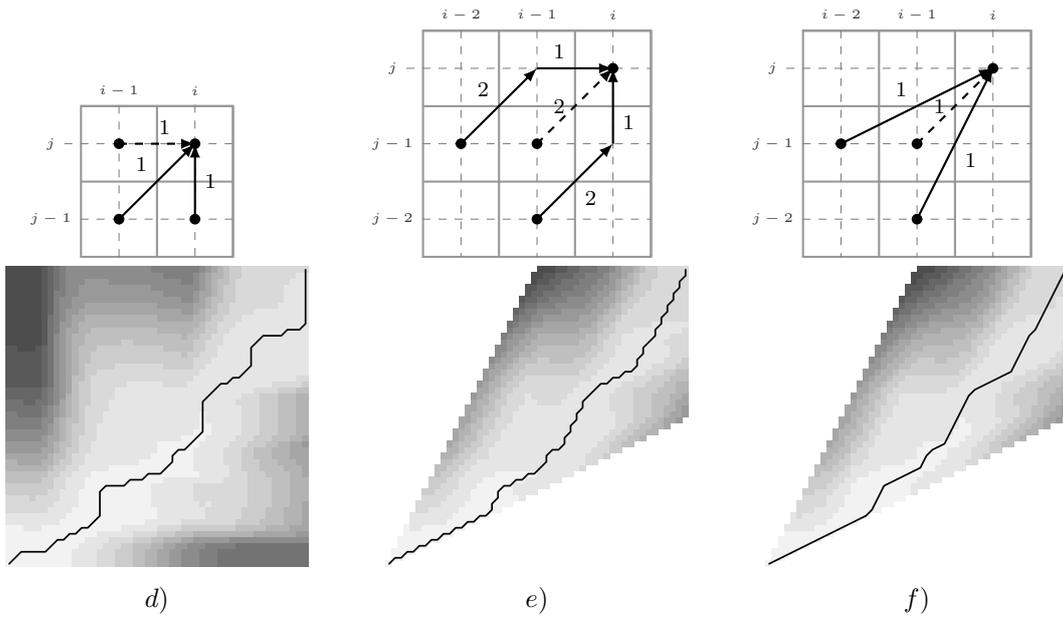
$$D(i, j) = \begin{cases} D(i-1, j-1) + w_d \times c(x_i, y_j) \\ D(i-1, j) + w_h \times c(x_i, y_j) \\ D(i, j-1) + w_v \times c(x_i, y_j) \end{cases} \quad (2.16)$$

Par exemple dans la Figure 2.17 *b*, le poids de l'avancée en diagonale ( $w_d = 3$ ) est supérieur au poids horizontale et verticale ( $w_v = w_h = 1$ ). De ce fait, avancer en diagonale est plus coûteux qu'avancer à l'horizontal et en verticale, comme le montre le chemin

**Modification de pas**



**Modification de pas**



**Taille de la fenetre**

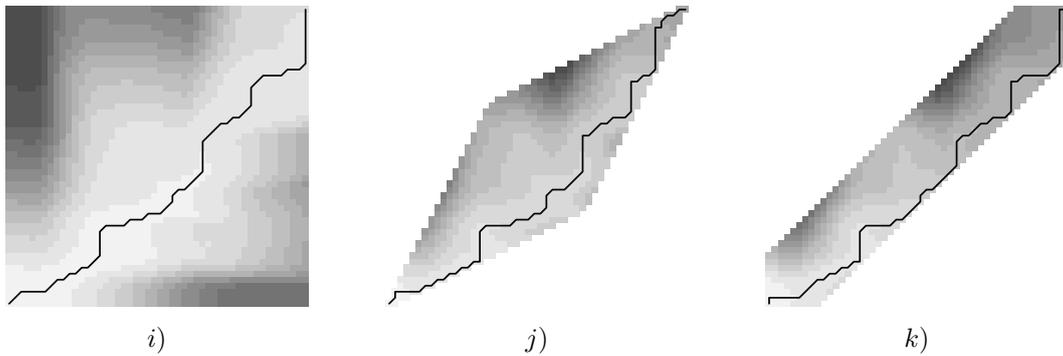


FIGURE 2.17: Illustration des variantes DTW présentées dans la section 2.3.2.

d’alignement. A l’inverse, dans le cas de la Figure 2.17 *c*, le poids de l’avancée en diagonale ( $w_d = 1$ ) est plus faible que les autres poids ( $w_v = w_h = 2$ ), et le chemin d’alignement emprunte plus souvent la diagonale.

**Modification de la fonction de pas** La condition de pas assure que chaque élément de  $U$  est assigné à un élément de  $V$ , et vice versa. Or sous cette condition, un élément de  $U$  peut être assigné à plusieurs éléments consécutifs de  $V$ , et de ce fait crée des alignements verticaux et horizontaux. Une approche intuitive suggère que ces chemins d’alignement sont coincés dans des minimas locaux. Ainsi pour s’affranchir de ces contraintes, il est possible de changer le pas utilisé par la DTW et forcer le chemin à changer de trajet (voir Figure 2.17 *d – f*).

Par exemple, dans l’algorithme DTW classique le coût pour aligner le doublet  $(i, j)$  est calculé à partir des doublets  $(i - 1, j)$ ,  $(i, j - 1)$  et  $(i - 1, j - 1)$  (Figure 2.17 *d*). Il est tout autant possible de calculer le coût à partir des doublets  $(i - 2, j - 1)$ ,  $(i - 1, j - 2)$  et  $(i - 1, j - 1)$ , comme dans la Figure 2.17 *e – f*. A noter que les fonctions de pas des deux images Figure 2.17 *e – f* induisent une restriction de la taille de la fenêtre (*c.-à-d.* qu’une partie de la matrice  $\mathbf{D}$  n’est pas définie).

En pratique, pour toute région restreinte on pose  $C(i, j) = \infty$ , ce qui empêche le chemin d’alignement de passer par cette région. Cette variante a pour principal avantage de permettre de réduire le temps de calcul du processus de DTW.

**Taille de la fenêtre** Une autre approche consiste à imposer une restriction sur les chemins d’alignements possibles, on peut alors définir des régions dans la matrice  $\mathbf{D}$  où un chemin ne peut pas passer. A titre d’exemple, voir l’illustration du parallélogramme d’Itakura et la bande Sakoe et Chiba dans Figure 2.17 *i – l*). En pratique, pour toute région restreinte on pose  $C(i, j) = \infty$ , ce qui empêche le chemin d’alignement de passer par cette région. Cette variante a pour principal avantage de permettre de réduire le temps de calcul du processus de DTW.

Par exemple, les figures 2.17 *k* et *l* illustrent respectivement le parallélogramme d’Itakura et la bande Sakoe et Chiba. Ces fenêtres, comme leurs noms l’indiquent, réduisent la fenêtre de travail à un parallélogramme et à une bande (soit un rectangle avec une rotation à  $45^\circ$ ). Cependant, l’utilisation de fenêtres peut aussi être problématique. En effet, le chemin d’alignement optimal  $s^*$  peut traverser la région restreinte. De ce fait, il est possible que l’alignement  $s$  choisi par le DTW ne coïncide pas avec  $s^*$ .

### 2.3.3 Dissimilarités entre échantillons

L’objectif de cette étape est de calculer les dissimilarités entre le signaux extraits des échantillons par les  $n_{FE}$  *fonctions d’extraction*. Pour connaître les dissimilarités entre caractères il convient de calculer les distances DTW entre chaque signal décrivant un caractère. Ainsi, avec  $n_s$  échantillons provenant à la fois du matériel *indiciaire* et de *référence*, une matrice de dissimilarité  $\mathbf{D}_{2n_s \times 2n_s}$  est obtenue pour chacune des  $n_{FE}$  *fonctions d’extraction*.

Or, il a été montré que l’algorithme de DTW peut être modifié pour changer ses propriétés d’alignement, et ainsi changer les distances entre signaux. Dans cette recherche plusieurs *fonctions de pas* sont étudiées (voir Section 3.1). Les calculs sont effectués à l’aide du logiciel *R* et le package *dtw* (v1.17.1) (Giorgino [102]).

## 2.4 Analyse de dissimilarité

L'objectif de cette section est de fournir une statistique permettant à la fois de réduire le nombre de dimensions d'une matrice à une valeur réelle  $F$ , et de quantifier l'information pertinente contenue dans la matrice de dissimilarité  $\mathbf{D}_{2n_s \times 2n_s}$ .

Or, la matrice  $\mathbf{D}$  est une matrice de dissimilarité, et le DTW n'est pas une mesure euclidienne mais une mesure semi-métrique. Ces contraintes limitent les possibilités de statistiques. Cependant, McArdle et Anderson [103] ont proposé une pseudo-statistique qui reprend les caractéristiques d'une Analyse de la Variance, tout en utilisant les propriétés des matrices de dissimilarité.

Cette section décrit le concept de l'Analyse de la Variance, puis les particularités de matrices de dissimilarité ainsi que leur relation avec l'Analyse de la Variance.

### 2.4.1 Analyse de la variance

Il convient dans un premier temps de définir les notions et concepts de l'Analyse de la Variance (ANOVA pour *ANalysis Of Variance* en anglais). Considérons un phénomène mesurable  $y$  qui est soupçonné d'être influencé par un ou plusieurs facteurs. La question que se pose le chercheur est de savoir si les facteurs ont une influence sur la variable mesurée  $y$ .

En prenant un exemple concret, un chercheur peut mesurer la concentration de Manganèse présent dans le verre (Mn, le phénomène mesurable  $y$ ) de deux bouteilles de la même marque (facteur à 2 niveaux). Le chercheur utilise  $n = 20$  échantillons par bouteille, alors la mesure de la concentration de Mn du  $j^{\text{ième}}$  échantillon de la  $i^{\text{ième}}$  bouteille est notée  $y_{i,j}$ . De manière similaire, le chercheur peut être amené à étudier la concentration de Titane (Ti) dans le verre de ces bouteilles (les concentrations de Mn et Ti sont représentées dans la Figure 2.18). Dans la suite de cette section, nous appellerons  $n$  le nombre d'échantillons,  $m$  le nombre de niveaux du facteur étudié et  $N$  le nombre total d'échantillons.

Le principe de l'ANOVA repose sur le modèle linéaire. A partir de ce modèle, il est possible d'inférer sur la différence entre les moyennes de deux groupes (*c.-à.-d.* les deux bouteilles) en analysant la variance des échantillons. Ainsi le chercheur peut souhaiter de comparer les hypothèses  $H_0$  : les échantillons de verre des deux bouteilles ont la même concentration de Mn ; et  $H_1$  les échantillons de verre des deux bouteilles ont des concentrations différentes.

L'Analyse de la Variance repose sur l'équation :

$$SS_T = SS_R + SS_A. \quad (2.17)$$

Cette équation est appelée *équation fondamentale de l'analyse de la variance*, et sa compréhension est essentielle pour saisir les principes qui régissent une ANOVA. La valeur  $SS_T$  est la somme totale des carrés des écarts (*SCE*, ou *Sum of Squares* en anglais), soit la variance totale des échantillons. Cette somme peut se décomposer en une somme de deux variations :  $SS_A$  la somme des carrés des écarts des moyennes de modalités à la moyenne générale et,  $SS_R$  la somme des carrés des écarts des valeurs observées à la moyenne.

Formellement, dans le cas univarié à un facteur, les mesures  $y$  peuvent être modélisés par un modèle linéaire :

$$y_{i,j} = \mu + \mu_i + \epsilon_{i,j}, \quad (2.18)$$

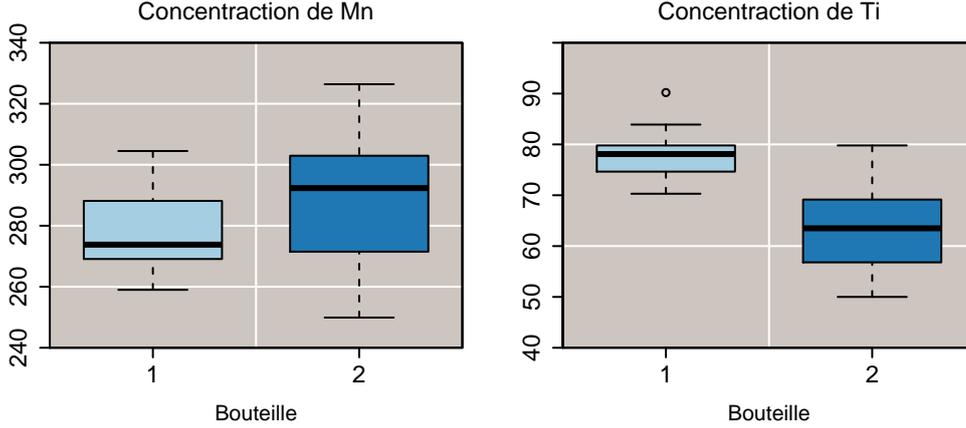


FIGURE 2.18: Exemple ANOVA : représentation des concentrations de Mn (manganèse) et Ti (titane) d'échantillons de verre prélevés sur deux bouteilles. On observe que la concentration de Ti permet une meilleure discrimination entre les deux bouteilles que la concentration de Mn. Les données sont issues du jeu de données `bottle.df` dans le package `R dafs` [104].

où,  $\mu$  est la moyenne de  $y$  commune à chaque espèce,  $\mu_i$  la moyenne de l'espèce  $i$ ,  $\epsilon_{i,j}$  l'erreur de la mesure de l'échantillon.

Intuitivement si les facteurs n'agissent pas alors  $\forall i, \mu_i = 0$ . Sous cette condition ( $H_0$ ),  $E(\frac{SS_A}{m-1}) = \sigma^2$  et  $E(\frac{SS_R}{N-m}) = \sigma^2$  sont deux estimateurs indépendants de  $\sigma$ . Au contraire, si les facteurs agissent alors  $\exists i, \mu_i \neq 0$  ( $H_1$ ), alors  $E(\frac{SS_R}{N-m}) = \sigma^2$  comme pour  $H_0$ , cependant  $E(\frac{SS_A}{m-1}) = \sigma^2 + \frac{1}{r-1} \sum^i n_i \mu_i^2 > \sigma^2$ . On appelle le rapport de ces deux estimateurs la quantité statistique  $F_{ratio}$  (voir [105, 106] pour plus de détails) :

$$F_{ratio} = \frac{SS_A/(m-1)}{SS_R/(N-m)}. \quad (2.19)$$

Ainsi un ratio supérieur à 1 suggère une différence entre les deux estimations de  $\sigma$ , et donc suggère que  $H_0$  soit rejetée. En d'autres termes, l'ANOVA a pour vocation de mesurer si la variabilité des échantillons est due à la variation intra groupe (propre à chaque bouteille) ou la variation inter groupe (propre aux bouteilles de cette marque).

Le modèle de l'ANOVA peut aussi s'écrire sous forme matricielle :

$$\mathbf{y}_{n \times 1} = [y_1, y_2, \dots, y_n]' \quad (2.20)$$

$$\mathbf{X}_{n \times q} = [\mathbf{1}_n, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k] \quad (2.21)$$

$$\boldsymbol{\beta}_{q \times 1} = [\mu_0, \mu_1, \dots, \mu_k]'$$

$$\mathbf{E}_{n \times 1} = [\epsilon_1, \epsilon_2, \dots, \epsilon_n]'$$

où  $\mathbf{X}$  est la matrice de contraste,  $\mathbf{B}$  la matrice des coefficients et  $\mathbf{E}$  la matrice d'erreur. Avec  $q = k + 1$ , le modèle matriciel de (2.21) devient :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}. \quad (2.22)$$

## 2.4.2 Analyse de matrice de dissimilarité

Prenons le cas où la variable  $\mathbf{y}$  n'est pas mesurée directement, mais à travers une matrice de dissimilarité entre les échantillons : la matrice  $\mathbf{D}$  (voir Figure 2.19). C'est par

exemple le cas si le chercheur a seulement une mesure de distance entre chaque échantillon.

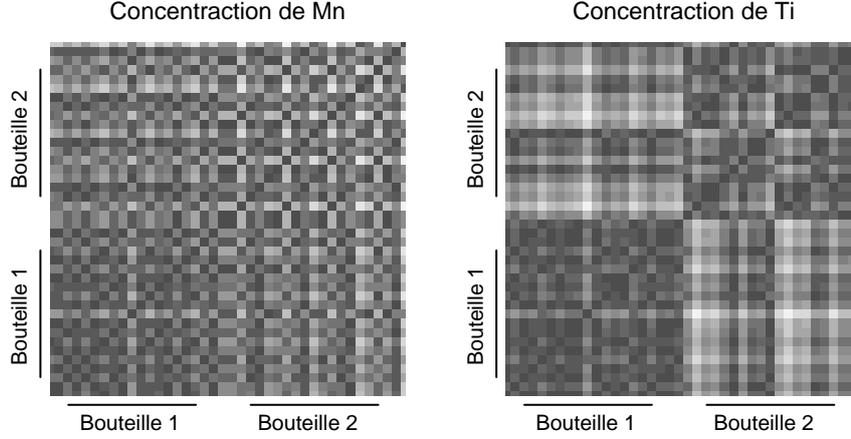


FIGURE 2.19: Exemple ANOVA : représentation matrices de dissimilarité des concentrations de Mn et Ti d'échantillons de verre prélevés sur deux bouteilles (distance euclidienne). Tout comme pour la représentation de la Figure 2.18, on observe que la concentration de Ti permet une meilleure discrimination entre les deux bouteilles que la concentration de Mn.

Pour progresser vers l'analyse de matrices de dissimilarité, il est nécessaire d'introduire la MANOVA (*Multivariate ANalysis Of Variance*), l'extension multivariée de l'ANOVA. Chaque échantillon  $(i, j)$  n'a plus une mesure unique  $y_{i,j}$ , mais est caractérisé par un vecteur  $\mathbf{y}_{i,j} = (y_{i,j}^1, \dots, y_{i,j}^p)$  de  $p > 1$  mesures. Par exemple, en plus de la concentration de Mn le chercheur peut décider d'ajouter comme mesure la concentration de Ti dans le verre ( $p = 2$ ), ou d'autres éléments. Le modèle MANOVA se note :

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (2.23)$$

La MANOVA fait intervenir le produit extérieur de la matrice des valeurs  $\mathbf{Y}'\mathbf{Y}$  (appelé *Total Sum of Squares Cross Product*, SSCP totale). Cette SSCP totale peut-être décomposée en la somme de la SSCP  $a$ ) des valeurs prédites et  $b$ ) des résidus. Posons  $\widehat{\mathbf{Y}}$  la matrice des valeurs prédites de  $\mathbf{Y}$ , la décomposition de la SSCP totale est donc :

$$\mathbf{Y}'\mathbf{Y} = \widehat{\mathbf{Y}}'\widehat{\mathbf{Y}} + \mathbf{E}'\mathbf{E}, \quad (2.24)$$

$$SSCP_T = SSCP_P + SSCP_R \quad (2.25)$$

En relation avec le modèle ANOVA, il est possible à partir de l'équation 2.25 de retrouver l'équation fondamentale de l'analyse de la variance :

$$\begin{aligned} SS_T &= tr(\mathbf{Y}'\mathbf{Y}), \\ SS_A &= tr(\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}}), \\ SS_R &= tr(\mathbf{E}'\mathbf{E}), \\ tr(\mathbf{Y}'\mathbf{Y}) &= tr(\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}}) + tr(\mathbf{E}'\mathbf{E}), \end{aligned} \quad (2.26)$$

Ainsi Anderson Anderson [107] définit la pseudo statistique  $F$  pour vérifier l'hypothèse  $H_0$  :

$$F_{ratio} = \frac{SS_A/(m-1)}{SS_R/(n-m)} = \frac{tr(\widehat{\mathbf{Y}}'\widehat{\mathbf{Y}})/(m-1)}{tr(\mathbf{E}'\mathbf{E})/(n-m)}, \quad (2.27)$$

En utilisant la propriété de la trace ( $tr(\mathbf{A}'\mathbf{B}') = tr(\mathbf{B}'\mathbf{A}')$ ), la même décomposition peut être obtenue à partir du produit extérieur des matrices (McArdle et Anderson [103]).

$$tr(\mathbf{Y}'\mathbf{Y}) = tr(\mathbf{Y}\mathbf{Y}') \quad (2.28)$$

$$\widehat{\mathbf{Y}}\widehat{\mathbf{Y}}' = \mathbf{H}(\mathbf{Y}\mathbf{Y}')\mathbf{H} \quad (2.29)$$

$$\mathbf{E}\mathbf{E}' = (\mathbf{I} - \mathbf{H})(\mathbf{Y}\mathbf{Y}')(\mathbf{I} - \mathbf{H}) \quad (2.30)$$

$$\text{avec } \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}. \quad (2.31)$$

Cette propriété est intéressante car, le produit extérieur d'une matrice peut aussi être obtenu d'une matrice de dissimilarité (Gower [108]). Soit une matrice de dissimilarité  $\mathbf{D}_{n \times n}$ , en posant :

$$\mathbf{G} = \left( \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}' \right) \mathbf{A} \left( \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}' \right), \quad (2.32)$$

alors  $\mathbf{G}$  peut être décomposée directement selon 2.26. Soit avec  $\mathbf{Y}\mathbf{Y}' = \mathbf{G}$ , la pseudo statistique  $F$  devient (Anderson [107]) :

$$F_{ratio} = \frac{tr(\mathbf{H}\mathbf{G}\mathbf{H})/(m-1)}{tr((\mathbf{I} - \mathbf{H})\mathbf{G}(\mathbf{I} - \mathbf{H}))/(n-m)}. \quad (2.33)$$

Dans le cas précis de la recherche, les matrices de dissimilarité  $\mathbf{D}_{2n_s \times 2n_s}$  correspondent à  $n_s$  échantillons qui proviennent à la fois du matériel *indiciaire* et de *référence*. Ce qui équivaut à une analyse à un facteur (deux niveaux, origine de la source : *indiciaire* ou *référence*), soit  $m = 2$  et  $n = 2n_s$ .

Ainsi à l'issue de cette étape un vecteur de  $n_{FE}$  pseudo-statistique  $\mathbf{y} = (F_{ratio,1}, \dots, F_{ratio,n_{FE}})$  est obtenu à partir des  $n_{FE}$  matrices de dissimilarité. Pour simplifier la notation et maintenir une cohésion avec la section suivante, ce vecteur, appelé aussi vecteur de caractéristiques, est noté  $\mathbf{y} = (y_1, \dots, y_{n_{FE}})$ . Les calculs de cette section sont effectués à l'aide du logiciel *R* et le package *adonis* (v2.0.10) ([109]).

## 2.5 Rapport de vraisemblance

En sciences forensiques il est accepté que la valeur d'observations soit exprimée par une métrique appelée le rapport de vraisemblance ( $RV$ ), partie intégrante du théorème de Bayes. Cette méthodologie bayésienne permet d'exprimer le poids des observations au regard des hypothèses d'intérêt dans le cas en examen, et également au juge de se positionner sur les hypothèses d'intérêt.

Cette métrique exprime la signification du vecteur  $\mathbf{y} = (y_1, \dots, y_{n_{FE}})$  des observations, selon les hypothèses posées par les parties en jeu, dans le domaine judiciaire, le procureur et la défense. Ces hypothèses sont généralement nommées  $H_p$  et  $H_d$  respectivement. Cette section décrit dans un premier temps le rapport de vraisemblance, et dans un deuxième temps le cas particulier utilisé dans cette recherche.

Il est utile de préciser que la notation dans cette recherche marque la différence entre les hypothèses :

- $H_0$  et  $H_1$ , faisant référence à l'analyse de matrice de dissimilarité.
- $H_p$  et  $H_d$ , faisant référence au rapport de vraisemblance.

## 2.5.1 Généralités

Posons une variable aléatoire  $y$  et une densité de probabilité  $f(y|\theta)$  spécifiée autour d'un paramètre inconnu  $\theta$ . Un modèle statistique Bayésien est composé par un modèle paramétrique  $\{f(y, \theta), \theta \in \Theta\}$  et une distribution de probabilité a priori sur le paramètre  $\theta$ ,  $\pi(\theta)$ . Les paramètres  $\theta$  appartiennent à l'espace des paramètres  $\Theta$ , divisé en deux sous-ensembles  $\Theta_p$  et  $\Theta_d$  tels que  $\Theta = \Theta_p \cup \Theta_d$ . La question est de savoir si la vraie valeur de  $\theta$  (mais inconnue) appartient à  $\Theta_p$  ou  $\Theta_d$ . Cela revient à tester l'hypothèse :

$$H_p : \theta \in \Theta_p \quad (2.34)$$

appelée hypothèse nulle, contre l'hypothèse :

$$H_d : \theta \in \Theta_d \quad (2.35)$$

appelée hypothèse alternative (Taroni et al. [110]). Par ailleurs les espaces des paramètres ( $\Theta$ ) peuvent être :

- un point fixe de l'espace (*p. ex.*  $\Theta = 0$ ), on appelle alors l'hypothèse associée hypothèse simple.
- un ensemble de l'espace (*p. ex.*  $\Theta \neq 0$  où  $\Theta \leq 1$ ), on appelle alors l'hypothèse associée hypothèse complexe.

L'échelle du  $RV$  est généralement exprimée par son logarithme  $\log(RV) = LRV$ . Dans ce cas, si  $LRV > 0$  les observations sont plus probables sous l'hypothèse  $H_p$ , et vice-versa.

Ces deux hypothèses peuvent être évaluées par la quantité statistique suivante : le *Facteur de Bayes (FB)*. Cette quantité mesure le changement des côtes en faveur des hypothèses en passant des a priori aux a posteriori (Lavine et Schervish [111], Gelman et al. [112]) :

$$BF = \frac{\alpha_p/\alpha_d}{\pi_d/\pi_p}, \quad (2.36)$$

où  $\alpha_p = P(\theta \in \Theta_p|y)$  et  $\alpha_d = P(\theta \in \Theta_d|y)$  sont les probabilités a posteriori des hypothèses  $H_p$  et  $H_d$ ; et  $\pi_p = P(\theta \in \Theta_p)$  et  $\pi_d = P(\theta \in \Theta_d)$  sont les probabilités a priori de  $\Theta_p$  et  $\Theta_d$ , tel que  $\pi_p + \pi_d = 1$ .

Quand les hypothèses d'intérêt sont simples, c'est-à-dire  $\Theta_p = \theta_p$  et  $\Theta_d = \theta_d$ , cette mesure statistique se simplifie et devient un rapport entre vraisemblance ( $RV$ ) [113]. Soit :

$$\alpha_i = \frac{f(y|\theta_i)\pi_i}{f(y|\theta_p)\pi_p + f(y|\theta_d)\pi_d}. \quad (2.37)$$

Alors

$$\begin{aligned} FB &= \frac{f(y|\theta_p)\pi_p}{f(y|\theta_p)\pi_p + f(y|\theta_d)\pi_d} \times \frac{f(y|\theta_p)\pi_p + f(y|\theta_d)\pi_d}{f(y|\theta_d)\pi_d} \times \frac{\pi_d}{\pi_p} \\ &\text{or } \pi_p + \pi_d = 1, \\ \Rightarrow FB &= \frac{f(y|\theta_p)\pi_p}{f(y|\theta_d)\pi_d} \times \frac{\pi_d}{\pi_p} \\ &= \frac{f(y|\theta_p)}{f(y|\theta_d)} = RV. \end{aligned} \quad (2.38)$$

## 2.5.2 Cas de la recherche

Dans le cas de cette recherche,  $\mathbf{y}$  est multivarié ( $p = n_{FE}$ , le nombre de *fonctions d'extraction* sélectionnées). Ces variables sont toutes des  $F_{ratio}$ . En théorie les valeurs des  $F_{ratio}$  sont proches de 1 si le facteur étudié n'a pas d'influence, et supérieures à 1 si le facteur a une influence. L'hypothèse faite est que, la situation où les échantillons d'écriture manuscrite proviennent du même scripteur correspond à la situation où le facteur n'a pas d'influence sur la mesure, et par conséquent un  $F_{ratio}$  proche de 1. Inversement, dans la situation où les échantillons d'écriture manuscrite ne proviennent pas du même scripteur correspond au cas où le facteur a une influence. En d'autres termes, il est attendu que si  $H_p$  est vraie alors  $H_0$  est vraie, et si  $H_d$  est vraie alors  $H_1$  est vraie ( $H_0 \Leftrightarrow H_p$  et  $H_1 \Leftrightarrow H_d$ ).

La loi normale multivariée à  $n_{FE}$  dimensions  $\mathcal{M}_{n_{FE}}$  est utilisée pour modéliser la distribution de  $\mathbf{y}$  :

$$f(\mathbf{y}|\theta) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right), \quad (2.39)$$

où  $\boldsymbol{\mu}$  et  $\boldsymbol{\Sigma}$  sont respectivement le vecteur de moyenne et la matrice de covariance du modèle.

Ainsi, les deux sous-ensembles  $\Theta_p$  et  $\Theta_d$  sont définis comme :

- $\Theta_p = (\boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W)$  : l'espace des paramètres lorsque le matériel *indiciaire* et de *référence* sont issus du même scripteur, soit l'espace de paramètres *intra-scripteur*,
- $\Theta_d = (\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$  : et l'espace des paramètres lorsque le matériel *indiciaire* et de *référence* proviennent de deux scripteurs différents, soit l'espace de paramètres *inter-scripteur*.

En pratique, les paramètres *intra-scripteur*  $\boldsymbol{\mu}_W$  et  $\boldsymbol{\Sigma}_W$ , et les paramètres *inter-scripteur*  $\boldsymbol{\mu}_B$  et  $\boldsymbol{\Sigma}_B$  sont estimés par les vecteurs de  $F_{ratio}$  des 11 scripteurs de l'expérience 2. Ainsi pour ne pas biaiser les résultats des rapports de vraisemblance, lorsqu'un scripteur est utilisé dans le matériel *indiciaire* ou de *référence* il est retiré du jeu de données utilisé pour estimer les paramètres.

Les calculs de cette section sont effectués avec le logiciel *R* et le package *mvtnorm* (v0.9.9997) [114, 115].

## Expériences

Les trois expériences présentées en introduction sont décrites en détail dans ce chapitre. L'objectif de ces expériences est d'appliquer sur un jeu de données, spécifique à chaque expérience, le processus de *vérification* de scripteur développé dans le Chapitre 2. Conjointement, ces expériences permettent de répondre aux deux dernières phases de la recherche, à savoir :

- la **validation** de la méthode de quantification et d'évaluation de l'écriture, qui a été développée dans cette recherche,
- et l'**application** de cette méthode à des exemples dans le domaine forensique.

Les mesures sont effectuées avec  $n_Q = 72$  et  $r_{min} = 15$ <sup>1</sup> (voir section 2.2). Ce chapitre est organisé de la manière suivante : tout d'abord un bref descriptif de chaque expérience est donné ; puis les expériences 1, 2 et 3 sont décrites en détail dans les sections 3.1, 3.2 et 3.3, respectivement.

**Expérience 1** Cette première expérience introduit les différentes combinaisons *fonctions d'extraction* et de *fonction de pas* DTW, mentionnées précédemment dans la section 2.2 et 2.3. L'objectif est de sélectionner les combinaisons qui permettent de discriminer des formes géométriques, tout en étant invariant à leur taille ou leur épaisseur. Contrairement aux deux expériences suivantes (Expériences 2 et 3), le jeu de données de cette expérience est constitué de formes géométriques (voir Figure 3.2 et la section 3.1), et ne contient pas d'échantillons d'écriture manuscrite.

Les expériences 2 et 3 utilisent des échantillons d'écriture manuscrite pour créer des *scénarios* fictifs. Dans le but de limiter l'influence de facteurs d'écriture dans les échantillons, le texte à rédiger, l'instrument scripturant, le support et la surface de rédaction ont été imposés aux scripteurs. A noter que le texte à rédiger dans les deux jeux de données des expériences 2 et 3 sont différents. L'instrument scripturant utilisé, pour les deux bases de données, est un stylo à bille Bic<sup>®</sup>Crystal<sup>™</sup> à encre noire. Les scripteurs ont écrit sur du papier blanc Xerox<sup>®</sup> Business<sup>™</sup> Laser/ Copier/Inkjet 80g/m<sup>2</sup> format A4. Les échantillons ont été numérisés en niveau de gris avec un scanner Imacon<sup>®</sup> Flexitight 2848<sup>™</sup> à une résolution de 1500 ppp.

**Expérience 2** Dans cette deuxième expérience, a été utilisé un jeu de données de 11 scripteurs, tiré des recherches de Marquis et al. [65]. Les objectifs de cette expérience sont

---

1. Des recherches préalables ont permis d'établir que les paramètres  $n_Q = 72$  et  $r_{min} = 15$  permettent un rendement optimal *information/temps de calcul* ; pour cette raison ces paramètres ont été maintenus le long de toute l'étude.

multiples : *a)* valider le processus de *vérification* en étudiant la discrimination de scripteurs des mesures calculées aux étapes de calcul de dissimilarité, d'analyse de dissimilarité, et d'évaluation statistique ; *b)* étudier les variations morphologiques entre caractères de tailles différentes ; et *c)* étudier les variations naturelles de l'écriture des personnes données au cours d'un espace temporel très court.

**Expérience 3** Dans cette troisième expérience, l'un des scripteurs utilisé dans l'expérience 2 (à savoir le scripteur 7) a rédigé trois textes de menace, un an après la première prise d'échantillon. L'objectif principal de cette expérience est d'étudier les variations de l'écriture manuscrite imputable au cours d'un espace temporel long, à travers l'exemple du scripteur 7. Par ailleurs, le schéma de cette expérience est similaire à un cas d'*identification* (voir section 1.2) et permet de contribuer à la validation du processus. Ceci car le taux d'erreur d'*identification* est une valeur permettant de mesurer le potentiel de discrimination de scripteurs de la méthode proposée.

## 3.1 Expérience 1

Dans cette recherche, six *fonctions de quantification*, trois *fonctions de normalisation* (voir Table 3.1) et sept *fonctions de pas* DTW sont étudiées (voir Figure 3.1). L'objectif de cette expérience est d'identifier parmi les 126 combinaisons possibles, les  $n_{FE}$  *fonctions d'extraction* et la *fonction de pas* DTW qui respectent au mieux les conditions fixées en introduction. C'est-à-dire que le processus puisse caractériser la forme des échantillons analysés, tout en étant invariant à leur taille et leur épaisseur.

Conformément aux critères énoncés dans la section 2.2 les *fonctions de quantification* sont basées sur des mesures de distances entre les pixels et le centre de coordonnées et les *fonctions de normalisation* sont basées sur des mesures du nombre de pixels. Il a été décidé de garder les *fonctions de quantification* et de *normalisation* simples, elles sont donc construites avec des fonctions simples, comme la somme, la moyenne ou la médiane. Les sept *fonctions de pas* DTW étudiés proviennent du package *R dtw*, elles sont sélectionnées car leurs *fonctions de pas* sont symétriques. Cet aspect assure d'obtenir une matrice de dissimilarité symétrique, condition essentielle pour poursuivre avec l'analyse de matrices de dissimilarité.

Les facteurs de taille et d'épaisseur sont tous deux partiellement *contrôlés* dans les échantillons d'écriture analysés dans cette étude, grâce au fait que le même support, le même instrument scripturant et le même texte ont été utilisés par les 11 scripteurs. Par ailleurs, il a été démontré que la taille d'écriture (dimension des caractères) est un facteur significatif de la morphologie des caractères manuscrits d'un scripteur (à ce propos, voir Marquis et al. [66]). Ainsi, afin de pouvoir étudier les facteurs de taille et d'épaisseur, un jeu de données de formes géométriques en noir et blanc est créé (voir Figure 3.2). Cette expérience n'utilise pas la totalité du processus de *vérification* de scripteur, l'objectif étant uniquement de trouver, parmi les combinaisons de paramètres possibles, celles qui ne sont pas influencées par les deux facteurs de taille et d'épaisseur et qui discriminent le mieux les formes. Le choix de ces meilleurs paramètres est soutenu part des observations visuelles et des analyses quantitatives.

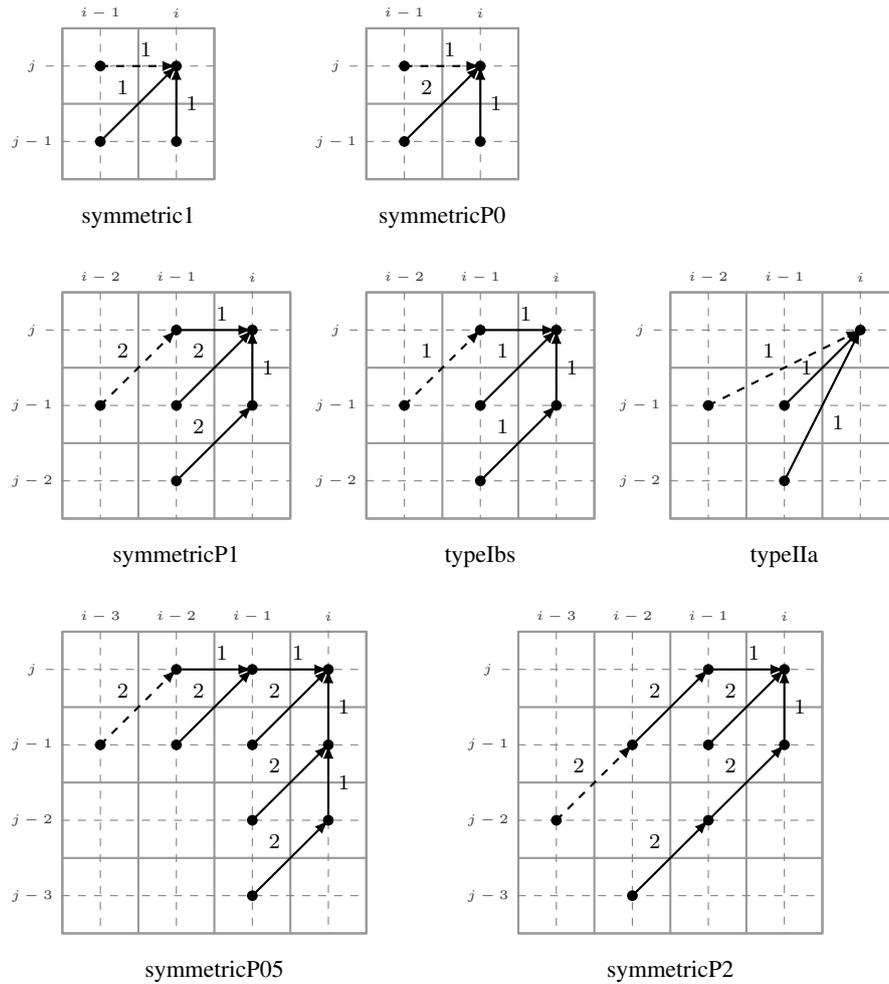


FIGURE 3.1: Illustration des sept *fonctions de pas* DTW étudiées dans l'expérience 1, à savoir les fonctions symmetric1, symmetricP0, symmetricP1, symmetricP2, symmetricP05, typeIbs, typeIIa.

### 3.1.1 Jeu de données #1

Dans ce jeu de données, sept formes géométriques sont reproduites à trois tailles et épaisseurs différentes (soit au total  $7 \times 3 \times 3 = 63$  formes, voir Figure 3.2). L'utilisation de formes géométriques, au lieu d'échantillons d'écriture manuscrite, offre deux avantages :

- elle ne nécessite pas de pré-traitement de l'image, et donc cette absence d'opération élimine toute perte d'informations liées à la binarisation de caractères ;
- les facteurs étudiés (à savoir la taille et l'épaisseur) sont entièrement contrôlés.

Le but recherché dans cette expérience n'est pas d'étudier les performances des méthodes pour toutes les formes possibles, mais d'étudier l'influence des facteurs de taille et d'épaisseur sur la discrimination de formes à partir de ce jeu de données. Ainsi, la liste des formes géométriques ne se veut pas exhaustive.

### 3.1.2 Protocole expérimental

Cette expérience n'utilise pas la totalité des étapes du processus de *vérification* de scripteur. Elle s'arrête à la production de matrices de dissimilarité (voir Figure 3.3),

| <i>fonction de quantification</i> |                                                                                                               | <i>fonction de normalisation</i> |                                    |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------|----------------------------------|------------------------------------|
| Nom                               | Formule                                                                                                       | Nom                              | Formule                            |
| $M_1$                             | $\sum_{i=1}^{n_q^C} (  \overrightarrow{CP_{q,i}}  )$                                                          | $N_1$                            | $n_q^C$                            |
| $M_2$                             | $\frac{\sum_{i=1}^{n_q^C}   \overrightarrow{CP_{q,i}}  }{\text{médiane}_{k,i}   \overrightarrow{CP_{k,i}}  }$ | $N_2$                            | $\frac{n_q^C}{\sum_{k=1}^Q n_k^C}$ |
| $M_3$                             | $\frac{\sum_{i=1}^{n_q^C}   \overrightarrow{CP_{q,i}}  }{\max_{k,i}   \overrightarrow{CP_{k,i}}  }$           | $N_3$                            | 1                                  |
| $M_4$                             | $\frac{\sum_{i=1}^{n_q^C}   \overrightarrow{CP_{q,i}}  }{n_q^C}$                                              |                                  |                                    |
| $M_5$                             | $\frac{\sum_{i=1}^{n_q^C}   \overrightarrow{CP_{q,i}}  }{\text{moyenne}_i   \overrightarrow{CP_{q,i}}  }$     |                                  |                                    |
| $M_6$                             | $\frac{\text{moyenne}_i   \overrightarrow{CP_{q,i}}  }{\text{médiane}_i   \overrightarrow{CP_{q,i}}  }$       |                                  |                                    |

TABLE 3.1: Noms et formules ( $f(q)^C$ ) des six *fonctions de quantification* ( $M_1 - M_6$ ) et des trois *fonctions de normalisation* ( $N_1 - N_3$ ) étudiées dans cette recherche.

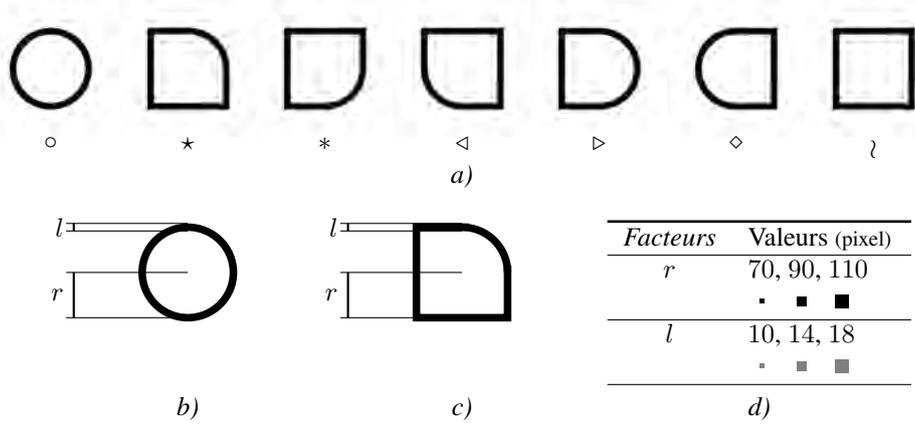


FIGURE 3.2: Illustration du jeu de données 1. a) illustration et légende des sept formes, b) et c) exemples de paramétrisation des formes, et d) valeurs des différents facteurs et leur légende.

l'objectif étant de trouver les meilleures combinaisons de *fonction d'extraction* et de DTW. Ainsi, une matrice de dissimilarité est calculée entre chacun des 63 échantillons, pour chaque :

- des 18 *fonctions d'extraction* ;
- des 7 *fonctions de pas* DTW.

Soit au total, 126 ( $18 \times 7$ ) matrices de dissimilarité  $\mathbf{D}_{63 \times 63}$  sont produites dans cette expérience. Les analyses des matrices de dissimilarité dans cette expérience sont classées en deux groupes (voir Figure 3.3), selon les critères suivants :

- **Observations visuelles** Dans un premier temps, les matrices de dissimilarité sont observées visuellement. Comme souligné dans la section sur l'analyse de matrice de dissimilarité (Section 2.4.2), les corrélations entre la variable mesurée et les facteurs peuvent être observées par la représentation graphique des matrices de dissimilarité  $\mathbf{D}$  (comme dans l'exemple de la Figure 2.19 section 2.4). Ces observations, certes simplistes, permettent toutefois d'exclure rapidement les combinaisons les plus décevantes.

Dans un deuxième temps, ces mêmes matrices de dissimilarité sont étudiées par Analyse par Coordonnées Principales<sup>2</sup> (que l'on appelle PCoA pour *Principal Coordinates Analysis* en anglais, afin de la différencier de l'Analyse en Composantes Principales). Cette méthode permet de représenter les échantillons définis par leurs distances relatives, par des points dans l'espace (voir [117, 118]).

Les matrices de dissimilarité peuvent aussi être analysées par classification hiérarchique<sup>3</sup>. Cette méthode statistique de fouille de données (ou *data mining* en anglais) a pour objectif de regrouper deux à deux les échantillons similaires, sur la base de la matrice de distance calculée. Le principe de l'algorithme consiste à créer, à chaque étape, une partition obtenue en agrégeant deux à deux les éléments les plus proches. On désignera alors par élément à la fois les individus ou objets à classer eux-mêmes et les regroupements d'individus générés par l'algorithme. Il y a différentes manières de considérer le nouveau couple d'éléments agrégés, d'où un nombre important de variations de cette technique [120]. L'algorithme ne fournit pas une partition en  $q$  classes d'un ensemble de  $n$  objets mais une hiérarchie de partitions, se présentant sous la forme d'arbres appelés également dendrogrammes et contenant  $n - 1$  partitions. L'intérêt de ces arbres est qu'ils peuvent donner une idée du nombre de classes existant effectivement dans la population [120].

De ce fait, dans cette expérience une matrice de dissimilarité est illustrée par trois représentations graphiques (représentation brute de la matrice, par PCoA et par classification hiérarchique).

---

2. Les calculs et les représentations d'analyse par coordonnée principale sont effectués avec le logiciel *R* et le package *ade4* (v1.6.2) [116].

3. Les calculs et les représentations de classification hiérarchique sont effectués avec le logiciel *R* et le package *stats* (v3.0.2) [119].

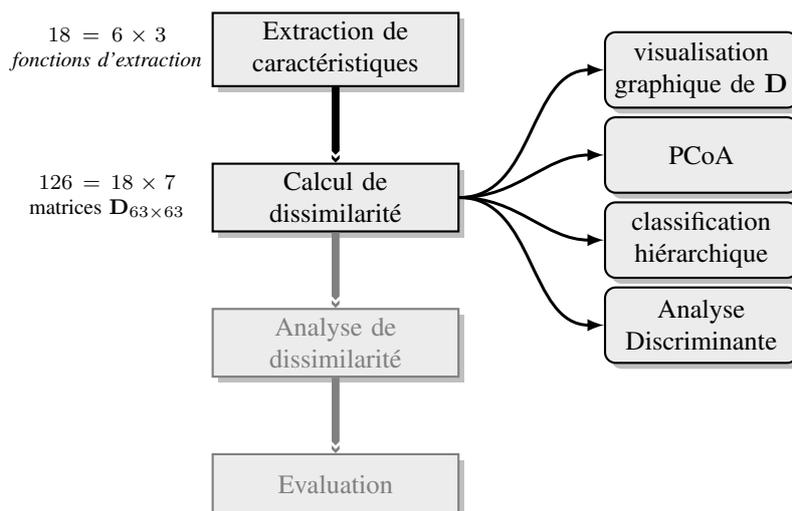


FIGURE 3.3: Illustration du processus d'analyse de l'expérience 1.

- Observation quantitative** Dans un quatrième temps, le potentiel de discrimination des matrices de dissimilarité est étudié par une *Analyse Discriminante Linéaire* (ADL). L'ADL est une méthode statistique qui permet, dans ce cas, de prédire l'appartenance d'une figure géométrique à l'une des sept classes (correspondant aux sept formes géométriques). Il s'agit donc d'analyser les matrices issues des PCoA avec une ADL. Les taux de classifications (soit le pourcentage de figures qui sont prédites dans la bonne classe) obtenus sont étudiés afin de sélectionner les meilleures combinaisons de *fonction d'extraction* et de DTW. On procède à un sous-échantillonnage aléatoire pour obtenir un jeu de données d'apprentissage (utilisé pour estimer les paramètres) et un jeu de données de validation (utilisé pour les prédictions). Les jeux de données d'apprentissage sont composés de 3 échantillons de chaque forme (soit 1/3 des échantillons), et le procédé est itéré 100 fois. Lorsque ce taux de classification tend vers 100%, il indique un meilleur potentiel de discrimination entre les formes géométriques. En revanche, plus le rapport s'approche de la valeur 0%, plus il souligne une difficulté à discriminer les différentes formes.

## 3.2 Expérience 2

L'objectif de cette expérience est, sur la base des  $n_{FE}$  *fonctions d'extraction* sélectionnées dans l'expérience 1, de constituer des *scénarios* fictifs à partir d'un jeu de données de 11 scripteurs, permettant de contribuer à :

- la **validation** de la méthode développée en étudiant la discrimination de scripteurs dans les mesures calculées aux étapes de calcul de dissimilarité, d'analyse de dissimilarité, et d'évaluation statistique ;
- l'**application** de la méthode à des d'intérêts forensiques. A savoir, *a)* étudier les variations morphologiques entre caractères de tailles différentes ; et *b)* étudier les variations naturelles de l'écriture des personnes données au cours d'un espace temporel court (soit  $<$  à 15 jours).

Par ailleurs, la hauteur et la largeur des caractères sont aussi extraites (appelées  $H$  et  $W$ ). Elles sont étudiées afin :

1. d'observer si les scripteurs ont respecté la consigne de l'expérience 2, à savoir d'écrire trois fois plus grand les caractères *grands* que *normaux* ;
2. de supporter la variabilité de l'écriture dans un espace temporel.

### 3.2.1 Jeu de données #2

La population de scripteurs de ce jeu de données est tiré de Marquis et al. [65]. Elle est constituée de 11 collaborateurs de l'*Institut de Police Scientifique*. Ces scripteurs ont été sélectionnés dans l'étude de Marquis et al. [65] parmi 100 scripteurs car ils possédaient la caractéristique de fermer le contour de leurs boucles, notamment sur les caractères a, d, o et q. Chaque scripteur a fourni cinq documents (appelés pages  $I - V$ ) rédigés le long d'une période de 15 jours. Chaque document contient :

- les caractères minuscules a, b, d, e, f, g, h, j, k, l, o, p, q, écrits à la taille habituelle de chaque scripteur (10×, caractères dits *normaux*) ;
- les caractères minuscules a, b, d, e, f, g, h, j, k, l, o, p, q, dits *grands*. Contrairement au point précédent, il a été demandé au scripteur d'écrire ces caractères à une taille trois fois supérieure de leurs caractères *normaux*. Les caractères dits *grands* sont reproduits 3× sur chaque page.

Soit, par scripteur (11) et par lettre (13) : 50 caractères *normaux* et 15 caractères *grands*. Au total 715 échantillons par lettre.

### 3.2.2 Protocole expérimental

Dans cette expérience, six *scénarios* fictifs sont créés (voir Figure 3.4).

- Les *scénarios* 1 à 4 comparent du matériel *indiciaire* d'écriture *normale* à du matériel de *référence* d'écriture *normale*, à quatre niveaux d'échantillonnage  $n_s = 5, 10, 15$  et  $20^4$ . Cette première série permet d'étudier l'influence de la taille d'échantillonnage sur la vérification de scripteurs ;
- Le *scénario* 5 compare du matériel *indiciaire* d'écriture *grande* à du matériel de *référence* d'écriture *grande* avec  $n_s = 5^5$  ;
- Le *scénario* 6 compare du matériel *indiciaire* d'écriture *normale* à du matériel de *référence* d'écriture *grande*. Ces deux derniers *scénarios*, en combinaison avec le *scénario* 1 permettent d'étudier l'influence de la taille de caractères.

Sur la base de la quantification des échantillons de caractères manuscrits, à savoir le vecteur de caractéristique obtenu des  $n_{FE}$  fonctions, un rapport de vraisemblance ( $RV$ ) est calculé pour :

- chaque *scénario* (6) ;
- chacune des 13 lettres du jeu de données ;

---

4. Soit à chaque niveau d'échantillonnage  $n_s$  échantillons du matériel *indiciaire* et  $n_s$  échantillons du matériel de *référence*.

5. La valeur  $n_s = 5$  à été choisie pour les *scénarios* 5 et 6, en raison de la quantité de matériel à disposition. En effet avec 15 échantillons de caractères *grands* par scripteur, il est impossible d'effectuer deux tirages aléatoires (sans remise) de  $n_s = 10$  échantillons d'un même scripteur pour ces caractères. Ainsi dans l'optique de comparer les résultats des différents *scénarios* et de limiter les niveaux d'échantillonnage, les *scénarios* utilisant des caractères *grands* sont limités à  $n_s = 5$ .

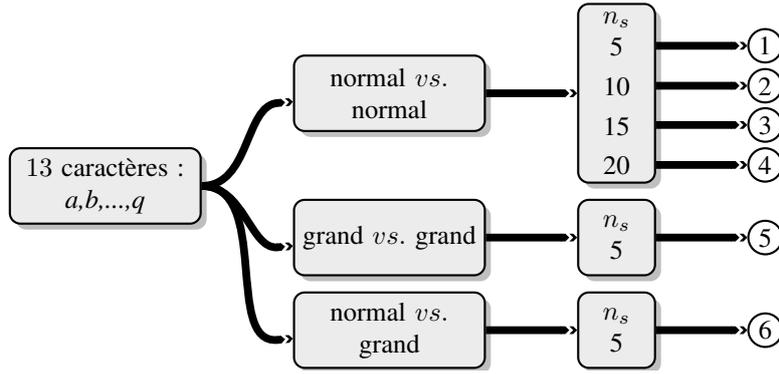


FIGURE 3.4: Illustration des six scénarios 1 – 6 de l’expérience 2.

- chacune des 66 combinaisons de scripteurs dans le matériel de *référence* et *indiciaire*, soit :
  - a) 11 combinaisons lorsque  $H_p$  est vraie, soit lorsque le matériel provient du même scripteur (1 vs. 1, 2 vs. 2, ...);
  - b) 55 combinaisons lorsque  $H_d$  est vraie, soit lorsque le matériel provient de deux scripteurs (1 vs. 2, 1 vs. 3, ...).
- pour chacune des  $n_I = 200$  itérations. Chaque tirage aléatoire des échantillons ( $n_s$  du matériel de *référence* et  $n_s$  du matériel *indiciaire*) des combinaisons précédentes est répété  $n_I$  fois.

Au total  $1029\,600 (= 6 \times 13 \times 66 \times 200)$   $RV$  sont produits dans cette expérience. Cependant les  $n_{FE} \times 1\,029\,600$  matrices de dissimilarité  $\mathbf{D}_{2n_s \times 2n_s}$  nécessaires pour calculer les  $RV$  ne sont pas calculées individuellement. En pratique  $n_{FE}$  matrices de dissimilarité entre tous les échantillons de la même lettre (715) sont calculées, soit  $n_{FE} \times 13$  matrices de dissimilarité  $\mathbf{D}_{715 \times 715}$ . Outre réduire le temps de calcul, cette option permet d’analyser les matrices de dissimilarité  $\mathbf{D}_{715 \times 715}$  par PCoA ou classification hiérarchique. Ainsi, pour chaque calcul individuel des  $F_{ratio}$  et des  $RV$ , la matrice  $\mathbf{D}_{2n_s \times 2n_s}$  est récupéré de la matrice  $\mathbf{D}_{715 \times 715}$ .

Les résultats des rapports de vraisemblance seront principalement donnés en taux d’erreur : c’est-à-dire les taux de *faux positifs* et de *faux négatifs*<sup>6</sup>. S’agissant d’un processus de *vérification*, on appellera aussi ces taux d’erreur « taux de *vérification* ».

Les  $RV$  sont calculés en multivariés, comme indiqué dans la section 2.5. Toutefois on calculera aussi les  $RV$  univariés, à savoir en utilisant qu’une seule des variables  $FE_i$ . Ces résultats univariés permettront de commenter la complémentarité de variables  $FE_i$  et de repérer si les résultats de  $RV$  univariés vont dans le même sens que ceux en multivariés.

Par ailleurs, afin d’appuyer les observations et les conclusions de cette expérience, les dimensions (largeur  $W$  et hauteur  $H$ ) et les ratios ( $W/H$ ) des caractères manuscrits sont aussi étudiés. Les restant des analyses de l’expérience est classé en trois groupes (voir Figure 3.5) :

- Représentations par PCoA.
- Observations des densités  $F_{ratio}$  ;
- Résultats des  $RV$  ;

6. Les taux de *faux positifs* sont les situations où le  $RV$  supporte l’hypothèse  $H_p$  alors que l’hypothèse  $H_d$  est vraie ; et les taux de *faux négatifs* sont les situations où le  $RV$  supporte l’hypothèse  $H_d$  alors que l’hypothèse  $H_p$  est vraie.

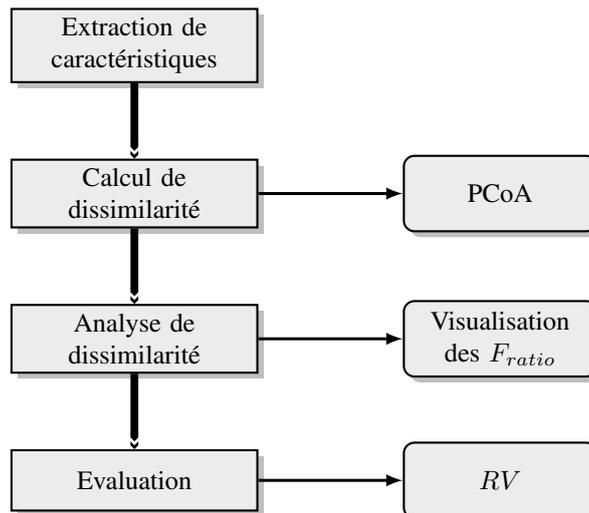


FIGURE 3.5: Illustration du processus d'analyse de l'expérience 2.

### 3.3 Expérience 3

Cette expérience a pour objectif, à travers l'étude de l'écriture du scripteur 7, de contribuer aux phases de validation et d'application de la recherche. Les échantillons d'écriture de cette expérience sont composés de trois textes supplémentaires, rédigés par le scripteur 7 un an après la première prise d'échantillon des cinq pages  $I - V$  de l'expérience 2. On appelle ces textes « textes de menace ». Ainsi sur la base de comparaisons entre les échantillons provenant de ces deux périodes, cette expérience vise à :

1. **valider** le processus de *vérification* de scripteur en étudiant les résultats de l'étape d'évaluation statistique, soit des calculs de rapport de vraisemblance. Il s'agit donc de déterminer si la méthode développée permet d'*identifier* correctement le scripteur 7 comme le scripteur des textes de menace ;
2. **appliquer** cette méthode à des échantillons d'écriture manuscrite afin d'étudier les variations naturelles de l'écriture au cours d'un espace temporel très court (différences entre les trois textes) et long (différence entre les trois textes et les cinq pages de l'expérience 2).

Tout comme dans l'expérience 2, les variables de hauteur  $H$  et de largeur  $W$  sont étudiées afin de supporter la variabilité de l'écriture dans un espace temporel long.

#### 3.3.1 Jeu de données #3

Le scripteur 7 de l'expérience précédente a fourni trois textes de menace (nommés  $i$ ,  $ii$  et  $iii$ , respectivement), une année après la première prise d'échantillons. Les nouveaux échantillons ont été recopiés en écriture script, sur une période de trois jours, chacun sur un jour différent. Les facteurs *contrôlés* (instrument scripturant, papier, numérisation) sont les mêmes que dans l'expérience précédente.

Sur chaque document le scripteur a écrit les lettres de l'alphabet ( $\times 1$ ), et le texte de menace suivant :

*Comme chaque fois, tu as provoqué une discorde inquiétante, espèce de loque ! Petite dinde, tu ne seras en banqueroute que sous peu. Je me moquerai*

| caractère   | a  | b | d  | e  | f | g | h | j | k | l | o  | p | q  |
|-------------|----|---|----|----|---|---|---|---|---|---|----|---|----|
| occurrences | 15 | 1 | 15 | 45 | 1 | 2 | 1 | 0 | 0 | 5 | 15 | 9 | 15 |
| sélectionné | •  |   | •  | •  |   |   |   |   |   |   | •  |   | •  |

TABLE 3.2: Tableau représentant le nombre d’occurrences des caractères contenus dans chaque texte de menace (*i – iii*), et les caractères sélectionnés pour les *scénarios* de l’expérience 3.

*durement de toi parceque tu ne pourras plus manger... dans quelque temps tu deviendras squelettique, disons dans cinq mois, quand tu seras en première page de tous les quotidiens.*

Dans l’optique de comparer les résultats de cette expérience avec plusieurs *scénarios* de l’expérience 2, seuls les lettres ayant plus de 10 occurrences dans le texte de menace sont sélectionnées (voir Table 3.2). De cette manière les *scénarios* de l’expérience 3 peuvent être comparés au *scénarios* 1 et 2 (respectivement  $n_s = 5$  et  $n_s = 10$ ). Soit au total :

- 48 échantillons pour les lettres a, d, o et q (15 + 1 par texte de menace) ;
- 138 échantillons pour la lettre e (45 + 1 par texte de menace).

### 3.3.2 Protocole expérimental

Le protocole expérimental de cette expérience est similaire à celui de l’expérience 2. Les 10 *scénarios* qui la composent sont regroupés en deux expériences : 3 – a et 3 – b (voir Figure 3.6).

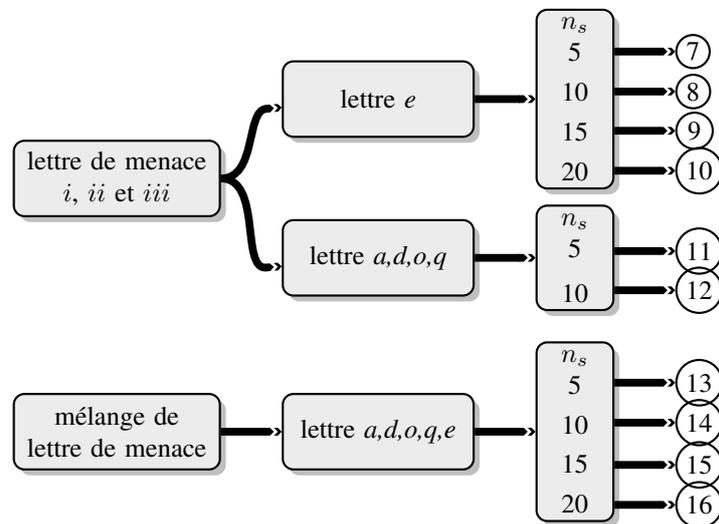


FIGURE 3.6: Illustration des *scénarios* 7 à 12 de l’expérience 3 – a et des *scénarios* 13 à 16 de l’expérience 3 – b.

**Expérience 3 – a** : *scénarios* 7 à 12.

Les *scénarios* de l’expérience 3 – a utilisent comme matériel :

- *indiciaire* des échantillons d’un texte de menace, soit 3 possibilités (textes de menace *i, ii et iii*) ;

- de *référence* des échantillons d'un scripteur du jeu de données #2, 5 pages confondues. Soit 11 possibilités.

Les six *scénarios* créés font varier le nombre d'échantillons ( $n_s$ ) et les lettres analysés (a, d, e, o et q) :

- les 4 premiers *scénarios* (7 à 10) utilisent la lettre e à quatre niveaux d'échantillonnage ( $n_s = 5, 10, 15, 20$ );
- les scénarios 11 et 12 utilisent les lettres a, d, o et q à deux niveaux d'échantillonnage ( $n_s = 5, 10$ ).

### Expérience 3 – b : scénarios 13 à 16.

Les *scénarios* 13 à 16 de l'expérience 3 – b reproduisent les six *scénarios* précédents à deux exceptions : le matériel *indiciaire* et les nombre d'échantillons  $n_s$  (voir Figure 3.6). Ces *scénarios* utilisent comme matériel :

- *indiciaire* des échantillons de l'ensemble des textes de menace ;
- de *référence* des échantillons d'un scripteur du jeu de données #2, 5 pages confondues. Soit 11 possibilités.

Ainsi, avoir comme matériel *indiciaire* un mélange des textes de menace implique que plus d'échantillons sont à disposition pour créer les *scénarios*. Par conséquent les cinq lettres (a, d, o, q et e) sont étudiés aux quatre mêmes niveaux d'échantillonnage ( $n_s = 5, 10, 15, 20$ ).

Dans l'expérience 3, un rapport de vraisemblance ( $RV$ ) est calculé sur la base des mesures  $F_{ratio}$  pour :

- chaque *scénario* ( $6 + 4$ ) ;
- chacune des 5 lettres du jeu de données ;
- (pour l'expérience 3 – a) chacun des 3 textes de menace dans le matériel *indiciaire* ;
- chacun des 11 scripteurs dans le matériel de *référence* ;
- pour chacune des  $n_I = 1000$  itérations<sup>7</sup>. Chaque tirage aléatoire des échantillons ( $n_s$  du matériel de *référence* et  $n_s$  du matériel *indiciaire*) des combinaisons précédente est répété  $n_I$  fois. Soit au total, pour chaque combinaisons 30000  $RV$  ( $3 \times 10 \times 1000 = 30000$ ) lorsque  $H_d$  est vraie et 3000 lorsque  $H_p$  est vraie.

Au total 396 000(=  $(1 \times 4 + 4 \times 2) \times 3 \times 11 \times 1000$ )  $RV$  sont produits dans l'expérience 3 – a et 220 000(=  $4 \times 5 \times 11 \times 1000$ ) dans l'expérience 3 – b. Comme dans l'expérience 2 les  $n_{FE} \times 616\,000$  matrices de dissimilarité  $\mathbf{D}_{2n_s \times 2n_s}$  ne sont pas calculées individuellement. En pratique  $n_{FE}$  matrices de dissimilarité entre tous les échantillons de caractères *normaux* de la même lettre sont calculées. Soit  $685 = 550 + 135$  échantillons pour la lettre e et  $495 = 550 + 45$  échantillons pour les lettres a, d, o, et q. Soit au total respectivement  $n_{FE}$  et  $n_{FE} \times 4$  matrices de dissimilarité  $\mathbf{D}_{685 \times 685}$  et  $\mathbf{D}_{495 \times 495}$ .

Comme dans l'expérience 2, les dimensions et leurs rapports (par exemple *largeur/hauteur*) des caractères sont étudiés. Le restant des analyses de l'expérience 3 est classé en deux groupes :

- Résultats des  $RV$  ;
- Représentations par PCoA.

---

7. L'expérience 3 ayant moins de lettres et combinaisons de scripteurs à analyser que l'expérience 2, le nombre d'itérations dans l'expérience 3 est augmenté à  $n_I = 1000$ .



## Résultats

Les résultats des expériences 1 à 3 sont présentés en suivant les schémas décrits dans le chapitre 3 (voir les Figures 3.3, 3.5). Face à la quantité de résultats numériques et graphiques découlant des expériences, une sélection permettant d’appréhender les tendances générales et les cas particuliers observés dans l’ensemble des résultats est faite. Ces résultats sont présentés dans ce chapitre, le restant des résultats est fourni en annexe.

### 4.1 Expérience 1

Dans cette première expérience, l’accent est mis sur l’étude des différentes mesures de dissimilarités proposées dans cette recherche, soit les *fonctions d’extraction* et les *fonctions de pas* DTW (section 2.2 et 2.3). L’objectif est de sélectionner les combinaisons de *fonctions d’extraction* et de *fonction de pas* DTW qui permettent de discriminer des caractères manuscrits, tout en étant invariant à leur taille ou l’épaisseur. Le choix de ces meilleures combinaisons est effectué par l’étude de la discrimination des formes issues du jeu de données de formes géométriques (voir section 3.1). Il est donc question d’analyser les matrices de dissimilarité entre les échantillons ( $\mathbf{D}_{63 \times 63}$ ) obtenues à la suite du processus de DTW. Les résultats sont présentés en deux sous-sections :

- 1) les résultats des analyses par représentation visuelle (matrice de distance, PCoA et classification hiérarchique) ;
- 2) les résultats des analyses quantitatives. Ce sont les résultats de l’analyse discriminante linéaire.

#### 4.1.1 Représentations visuelles

La première partie de cette expérience porte sur l’analyse visuelle des données, à savoir la représentation visuelle des matrices de dissimilarité, et de leurs représentations par PCoA et classification hiérarchique. En effet, comme indiqué dans la section 2.4, les représentations visuelles des matrices de dissimilarité entre les figures géométriques peuvent faire apparaître les éventuelles corrélations avec les facteurs étudiés (à savoir, la taille, l’épaisseur et la forme des figures géométriques).

Les résultats obtenus sont similaires entre ces différentes représentations, ils indiquent que :

1. chaque *fonction d’extraction* possède sa propre étendue. Par exemple (Figure 4.1), l’étendue des matrices  $\mathbf{D}_{715-715}$  des *fonctions d’extraction*  $M_1 - N_1$  et  $M_3 - N_1$  sont respectivement de  $[169,1; 7987,5]$  et  $[1,0; 7,7]$ <sup>1</sup>. Ceci souligne que les mesures des

---

1. Ces étendues ne prennent pas en compte les éléments de la diagonale, qui sont égaux à 0

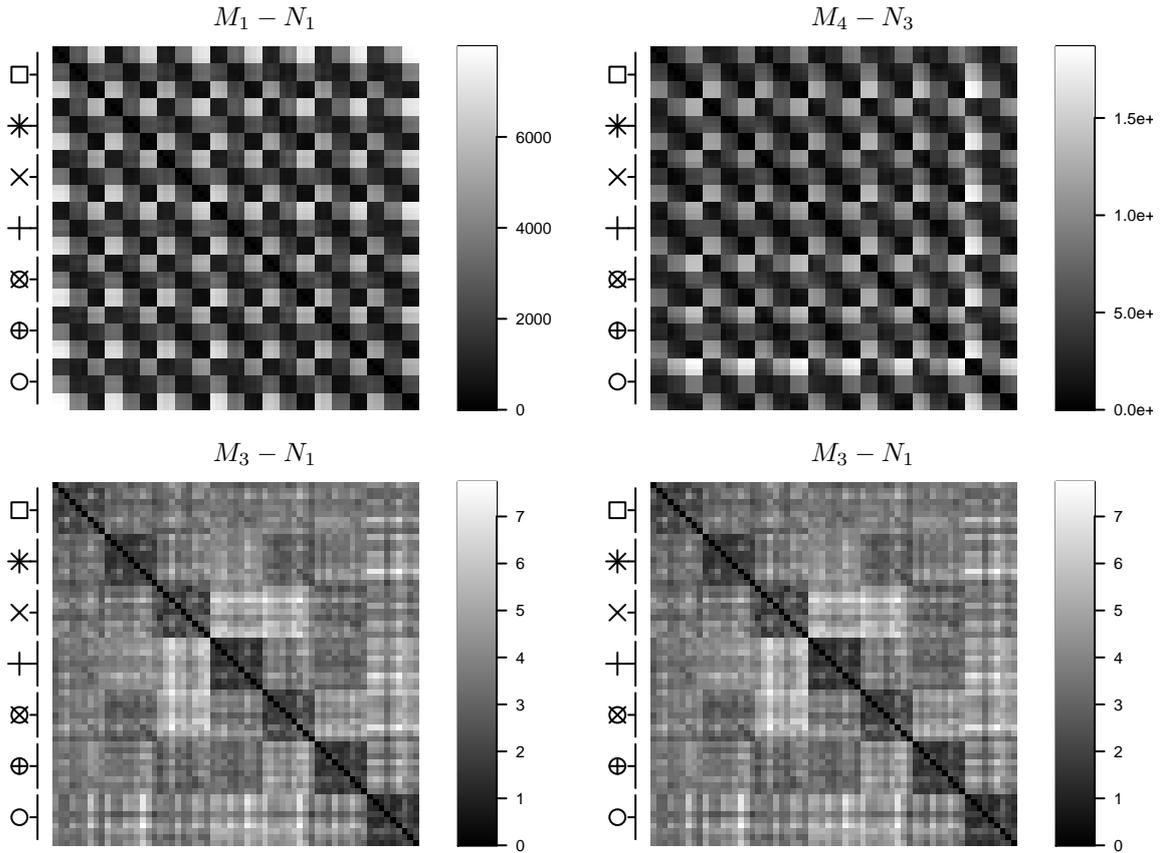


FIGURE 4.1: Illustration des matrices de distance  $\mathbf{D}_{63 \times 63}$  obtenues avec les *fonctions d'extraction*  $M_1 - N_1$ ,  $M_4 - N_3$ ,  $M_3 - N_1$  et  $M_6 - N_2$  pour la *fonction de pas* DTW *symmetricP2*. A noter que le motif observé pour les fonctions  $M_1 - N_1$  et  $M_4 - N_3$  est représentatif d'une fonction qui permet de discriminer les figures géométriques selon leur taille (*c.-à-d.* les distances les plus faibles, en noir, sont entre les figures de même taille). A l'inverse, le motif observé pour les fonctions  $M_3 - N_1$  et  $M_6 - N_2$  est celui d'une fonction permettant de discriminer les figures géométriques selon leur forme (*c.-à-d.* les distances les plus faibles, en noir, sont entre les figures de même forme). Par ailleurs, ce graphique met en évidence que chaque *fonction d'extraction* possède sa propre étendue.

18 *fonctions d'extraction* sont à des échelles différentes. Ainsi, lors de situations où le processus est utilisé en entier (comme dans les expériences 2 et 3), cet aspect montre l'intérêt d'utiliser une métrique comme le  $F_{ratio}$  afin de, en quelque sorte, normaliser les observations dans la matrice de dissimilarité à la même échelle ;

2. deux motifs différents sont observés parmi les combinaisons de *fonction d'extraction* et de *fonction de pas* DTW. D'un côté, les combinaisons qui discriminent les échantillons selon leur taille, comme la *fonction d'extraction*  $M_1 - N_1$ . D'un autre côté, celles qui discriminent selon leur forme géométrique, comme la fonction  $M_3 - N_1$ . Les représentations visuelles montrent que chacune des matrices de dissimilarité possède un schéma similaire à un damier, propre au facteur auquel elle est corrélée (voir Figure 4.1). Ces motifs sont aussi observés avec PCoA (Figure 4.4) et classification hiérarchique (Figure 4.3). En particulier, pour une *fonction de pas* donnée

**Matrices de distances – fonction de pas DTW symmetricP2.**

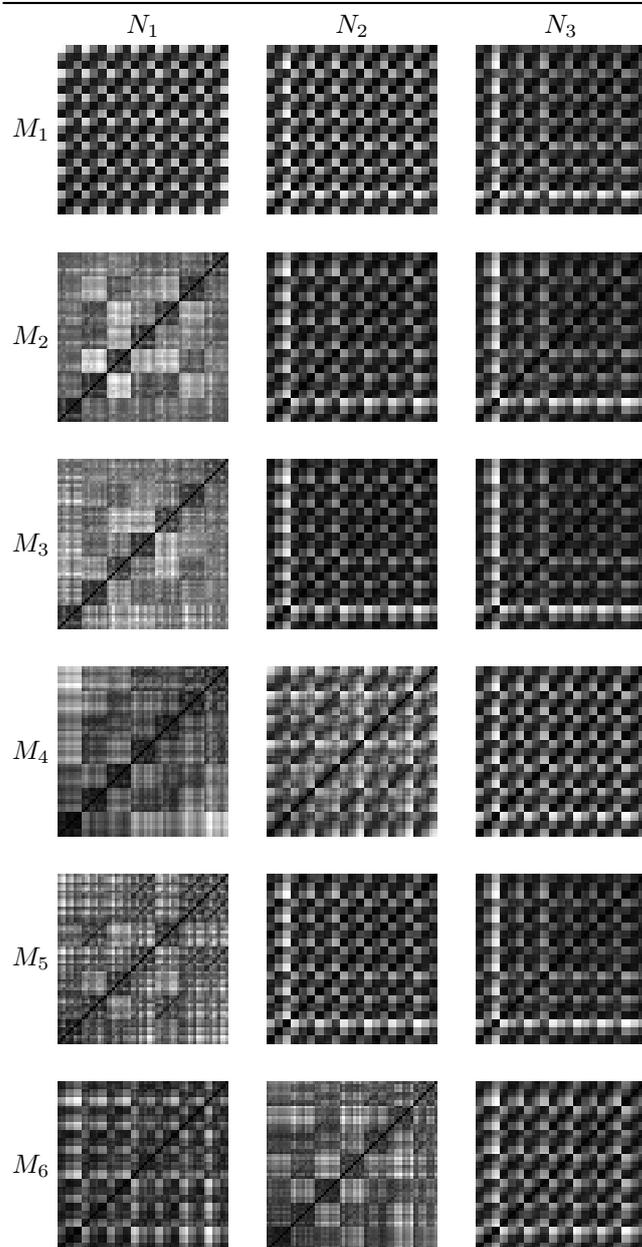


FIGURE 4.2: Illustration des matrices de distance  $\mathbf{D}_{63 \times 63}$  pour les 18 combinaisons de fonctions d'extraction pour la fonction de pas DTW symmetricP2. On observe deux groupes de fonctions, selon les motifs visualisés dans la Figure 4.1 : 1) les fonctions qui permettent de discriminer les échantillons selon leur forme géométrique (à savoir  $M_2 - N_1, M_3 - N_1, M_4 - N_1, M_5 - N_1, M_6 - N_1$ , et  $M_6 - N_2$ ), 2) les fonctions qui discriminent les échantillons selon leur taille (soit le restant des 18 fonctions).



Représentation PCoA – fonction de pas DTW symmetricP2

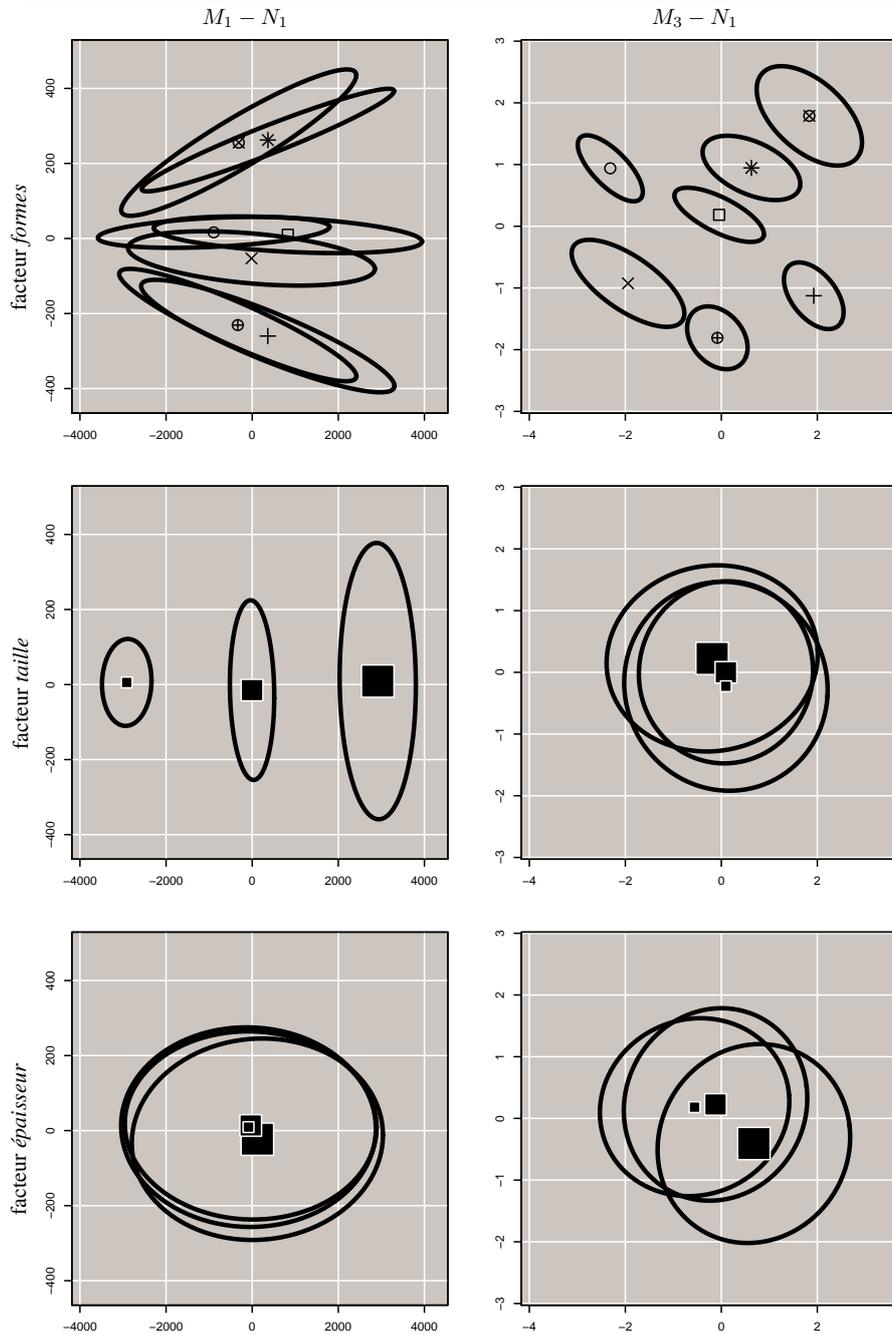


FIGURE 4.4: Représentations PCoA des matrices de dissimilarité  $\mathbf{D}_{63 \times 63}$  pour les *fonctions d'extraction*  $M_1 - N_1$  et  $M_3 - N_1$  et la *fonction de pas* DTW symmetricP2, regroupées par forme, taille et épaisseur. On observe, tout comme indiqué dans les Figures 4.2 et 4.3, deux motifs représentatifs des fonctions qui permettent de classer les échantillons selon leur taille ( $M_1 - N_1$ ), ou selon leur forme géométrique ( $M_3 - N_1$ ). A noter que les deux fonctions ne semblent pas être influencées par l'épaisseur des figures.

**Représentation PCoA – fonction de pas DTW symmetricP2.**

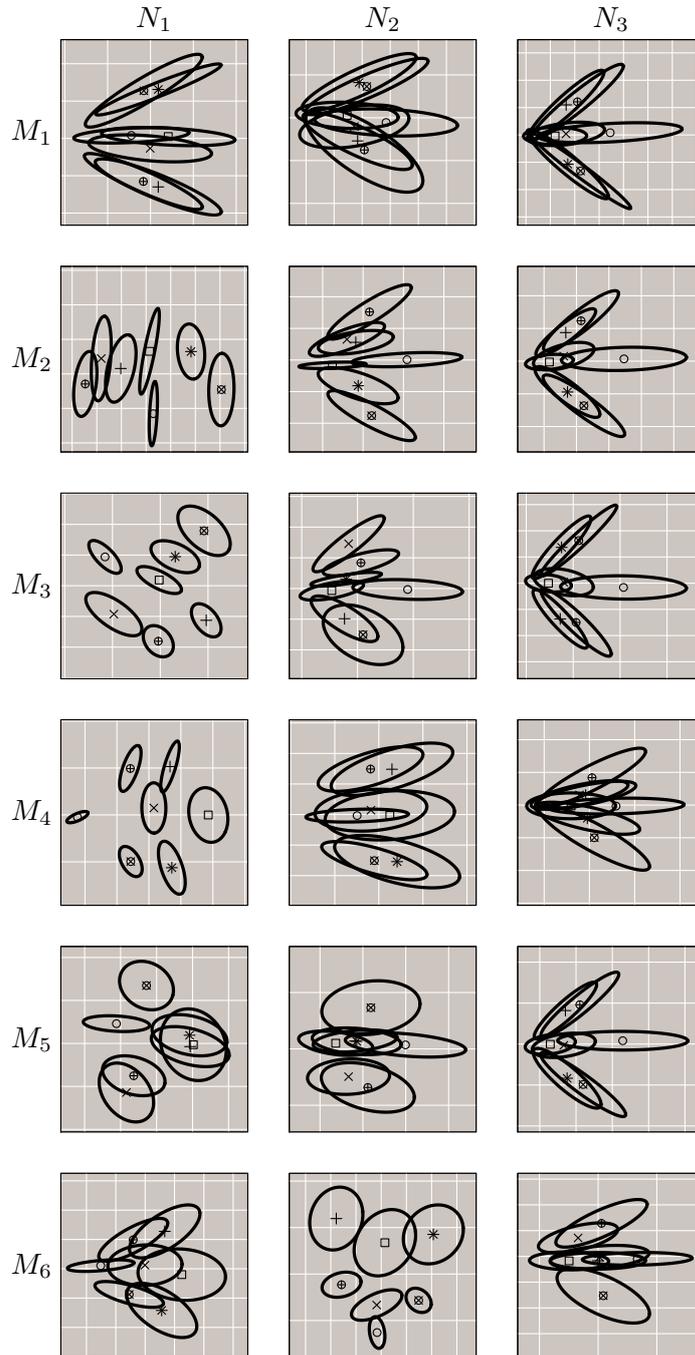


FIGURE 4.5: Représentation par PCoA des 18 combinaisons de *fonctions d'extraction* et la *fonction de pas DTW symmetricP2*, regroupées par forme. Il est mis en évidence, que les fonctions  $M_2 - N_1$ ,  $M_3 - N_1$ ,  $M_4 - N_1$  et  $M_6 - N_2$  sont celles qui permettent de discriminer les formes de manière optimale (c.-à-d. les groupes sont séparés). Par contre, le reste des fonctions ne semblent pas pouvoir discriminer les formes, et plus particulièrement dans les fonctions composées de  $M_1$  ou  $N_3$ .

(voir Figures 4.2 et 4.5) les résultats montrent que :

- toutes les *fonctions d'extraction* composées des fonctions  $M_1$  ou  $N_3$  ne permettent pas de discriminer les formes géométriques. Ceci peut, par exemple, s'observer dans les représentations par PCoA, où les échantillons apparaissent mieux regroupés par leur taille plutôt que par leur forme. Par conséquent ces fonctions ne sont pas de bons candidats pour discriminer et étudier l'écriture de scripteurs ;
- aucune des *fonctions d'extraction* ne semble être corrélée à l'épaisseur des formes géométriques ; en effet, les graphiques de la Figure 4.4 ne montrent pas de lien particulier entre les mesures de dissimilarité et l'épaisseur des formes ;
- seules les *fonctions d'extraction*  $M_2 - N_1$ ,  $M_3 - N_1$ ,  $M_4 - N_1$ ,  $M_5 - N_1$ , et  $M_6 - N_2$  permettent de discriminer les échantillons selon leur forme géométrique. Ainsi, au regard des représentations visuelles, ces *fonctions d'extraction* apparaissent comme le meilleur choix pour quantifier et discriminer l'écriture de scripteurs selon la morphologie de leurs caractères manuscrits.

### 4.1.2 Analyse Discriminante Linéaire

La deuxième partie de cette expérience est axée sur l'étude des taux de classification résultants de l'analyse discriminante linéaire. Ces taux de classification sont utilisés pour quantifier le potentiel de discrimination des matrices de dissimilarité  $\mathbf{D}_{63 \times 63}$ , à savoir plus le taux tend vers 100% plus il indique une meilleure discrimination des formes géométriques. Ainsi ils contribuent à choisir, avec les résultats de la section précédente, les combinaisons de *fonctions d'extraction* et de *fonctions de pas* qui seront exploitées dans les expériences 2 et 3.

Dans un premier temps, les moyennes des taux de classification pour chacune des 18 *fonctions d'extraction* sont étudiées. C'est-à-dire, la moyennes sur les 100 itérations et les 7 *fonctions de pas*. Cette étape permet de déterminer les meilleures  $n_F$  *fonctions d'extraction*. Dans un deuxième temps, les moyennes des taux de classification pour chacune des 7 *fonctions de pas* sont analysées, afin de n'en retenir que la meilleure.

La Figure 4.6 illustre le taux de classification selon les combinaisons de *fonctions d'extraction*. Les résultats indiquent que :

- toutes les *fonctions d'extraction* qui sont basées sur la *fonction de normalisation*  $N_3$  et la *fonction de quantification*  $M_1$  ont un taux moyen de classification inférieur à 50%. Ces résultats sont similaires à ceux des représentations visuelles de la section précédente ; ils soulignent que ces fonctions ne permettent pas d'étudier l'écriture de scripteurs ;
- cinq *fonctions d'extraction* se démarquent. Les fonctions  $M_2 - N_1$ ,  $M_3 - N_1$ ,  $M_4 - N_1$ ,  $M_5 - N_1$  et  $M_6 - N_2$  sont les seules à avoir un taux de classification moyen supérieur à 80%, tandis que ce taux pour le restant des *fonctions d'extraction* ne dépasse pas 65%. Parmi ces cinq fonctions, on remarque que les taux de classification de :
  - a)  $M_2 - N_1$ ,  $M_3 - N_1$ ,  $M_4 - N_1$  et  $M_5 - N_1$  sont en moyenne supérieurs à 90% ;
  - b)  $M_2 - N_1$  et  $M_4 - N_1$  (*écart-type* = 8.83 et 8.78) sont moins variables que ceux de  $M_3 - N_1$ ,  $M_5 - N_1$  et  $M_6 - N_2$  (*écart-type* = 11.42, 11.01 et 10.50).

Par conséquent, ces résultats permettent de désigner les deux *fonctions d'extraction*  $M_2 - N_1$  et  $M_4 - N_1$  comme celles qui ont les taux de classification les plus élevés,

avec un écart-type le plus faible. Cet aspect, combiné avec les résultats de la section précédente, montre que les deux *fonctions d'extraction* sont les meilleurs candidats pour discriminer les formes géométriques, et par conséquent les caractères manuscrits selon leur morphologie.

### Résultats classification linéaire

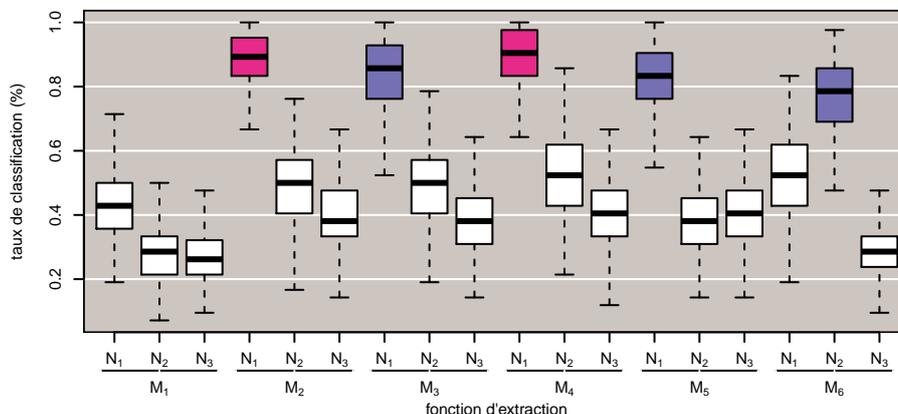


FIGURE 4.6: Boîtes à moustache des taux de classification pour les 18 *fonctions d'extraction* (7 *fonctions de pas* de DTW pour chaque *fonction d'extraction*). Ce graphique met en évidence les fonctions qui possèdent un taux de classification élevé, indiquant une meilleure discrimination des formes. On observe deux groupes caractérisant les *fonctions d'extraction* : les fonctions qui possèdent un taux de classification supérieur à 90% et un écart-type inférieur à 9 (soit  $M_2 - N_1$ ,  $M_4 - N_1$  en rouge), les fonctions avec un taux supérieur à 80% et un écart-type supérieur à 9 ( $M_3 - N_1$ ,  $M_5 - N_1$  et  $M_6 - N_2$  en violet), les fonctions dont le taux est inférieur à 65% (en blanc et noir).

La Figure 4.7 illustre le taux de classification pour les deux *fonctions d'extraction*, retenues précédemment, selon les sept *fonctions de pas* de DTW suggérées dans l'expérience 1 (voir section 3.1). Ces résultats permettent donc de sélectionner la *fonction de pas* avec laquelle on obtient les meilleurs taux de classification. La *fonction de pas symmetricP2* affiche les taux de classification les plus élevés avec 91.5%, puis *symmetricP1* et *typeIbs* (90.3% et 90.9%). Par ailleurs *symmetricP2* a aussi l'écart-type le plus bas (7.92) après *symmetricP1* (7.49).

Pour simplifier le processus de *vérification* de scripteur il est préférable de ne choisir qu'une seule fonction DTW commune à chaque fonction d'extraction. Par ailleurs, les deux fonctions ci-dessus (*typeIbs* et *symmetricP1*) n'améliorent, au mieux, les rapports que de 0.67% par comparaison à la fonction *symmetricP2*. Dès lors la *fonction de pas symmetricP2* apparaît comme le meilleur choix.

### 4.1.3 Récapitulatif des résultats de l'expérience 1

En conclusion, les observations visuelles et quantitatives dans le cadre de cette première expérience, visant à étudier la discrimination des formes géométriques, permettent de choisir les combinaisons des *fonctions d'extraction* et des *fonctions de pas* optimales

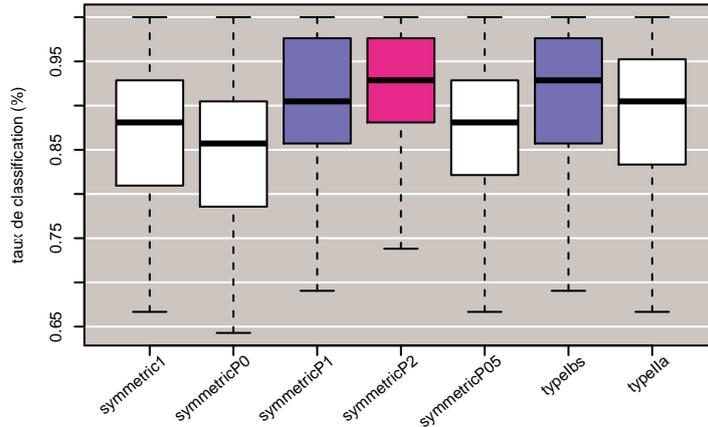


FIGURE 4.7: Boîtes à moustache des taux de classification pour les sept *fonctions de pas* de DTW, avec les deux *fonctions d'extraction* présentant la meilleure discrimination de formes (à savoir  $M_2 - N_1$  et  $M_4 - N_1$ ). Le graphique met en évidence que les taux de classification les plus élevés, indiquant une meilleure capacité de discrimination, et avec un écart-type le plus faible, indiquant une faible variation, est obtenu avec la *fonction de pas symmetricP2*.

pour la quantification et la discrimination de caractères manuscrits. Ainsi deux *fonctions d'extraction* ( $M_2 - N_1$  et  $M_4 - N_1$ , soit  $n_{FE} = 2$ ), et la fonction de pas *symmetricP2* ont été retenues.

Par ailleurs, les résultats soulignent que :

1. chaque *fonction d'extraction* possède sa propre étendue, rendant les valeurs de dissimilarité calculées plus difficiles à comparer entre elles. Cet aspect démontre l'utilité des métriques (ou de mesures) comme le  $F_{ratio}$  pour ajuster les mesures de dissimilarité à la même échelle ;
2. certaines *fonctions d'extraction*, comme la fonction  $M_1 - N_3$ , sont de toute évidence des mesures inadaptées à l'étude des caractères manuscrits, car trop biaisées par les dimensions des caractères. De ce fait ces résultats soulignent, dans le cadre de recherche scientifique, l'importance de s'assurer de la fiabilité des mesures employées afin d'éviter une erreur dans l'interprétation des résultats.

Dans la suite de cette recherche les combinaisons de ces deux *fonctions d'extraction* et de la fonction *symmetricP2* sont notées  $FE_i, i = 1 \dots 2$ . Elles seront utilisées pour quantifier les échantillons d'écriture manuscrite dans les expériences 2 et 3 :

$$FE_1 \Leftrightarrow M_2 - N_1 + symmetricP2,$$

$$FE_2 \Leftrightarrow M_4 - N_1 + symmetricP2$$

## 4.2 Expérience 2

Cette deuxième expérience vise à quantifier et étudier l'écriture manuscrite des 11 scripteurs du jeu de données #2 (voir section 3.2). Chaque scripteur a fourni cinq do-

cuments (appelés pages  $I - V$ ), rédigés le long d'une période de 15 jours. Chaque page contient un ensemble de caractères issus de l'écriture habituelle des scripteurs (caractères dits *normaux*) et de caractères dits *grands* (soit des caractères que les scripteurs ont dû écrire à une taille trois fois supérieure à celle leurs caractères *normaux*). L'expérience 2 présente trois objectifs :

1. valider le processus de *vérification* en étudiant ses paramètres (à savoir le nombre d'échantillons  $n_s$ , le(s) scripteur(s) étudié(s), et la lettre utilisée (a, b, ...)) et savoir si les mesures des étapes de calcul de dissimilarité, d'analyse de dissimilarité, et d'évaluation statistique permettent de discriminer l'écriture de scripteurs ;
2. étudier s'il existe des variations entre les caractères définis comme étant *normaux* et les caractères *grands* ;
3. étudier les variations naturelles de l'écriture des scripteurs du jeu de données #2 au cours d'un espace temporel très court (*c.-à-d.* les échantillons ont été rédigés sur une période de 15 jours).

### 4.2.1 La largeur, la hauteur et le rapport $W/H$

Les variables décrivant les dimensions (la largeur  $W$ , la hauteur  $H$ , le rapport  $W/H$  et la surface  $W \times H$ ) des caractères *normaux* et *grands* sont présentées dans cette section. Elles sont étudiées dans le but de déterminer si :

1. les scripteurs ont respecté la consigne d'écriture pour les caractères *grands*, à savoir qu'ils soient produits à une taille trois fois supérieure à celle de leurs caractères *normaux* ;
2. ces variables varient selon la session d'écriture, soit dans un espace temporel court (*c.-à-d.* les échantillons ont été rédigés sur une période de 15 jours).

La Figure 4.8 illustre les rapports des tailles moyennes  $W$  et  $H$  entre les caractères *normaux* et *grands*. Les résultats sont représentés selon les 11 scripteurs (en haut), et selon les 13 lettres (en bas). Ainsi ces graphiques permettent donc d'évaluer si les scripteurs ont respecté la consigne d'écriture pour les caractères *grands*. Les résultats indiquent que :

- les caractères *grands* ne sont pas de dimensions 3 fois supérieure aux caractères *normaux*, comme demandé dans le protocole de l'expérience 2. En effet, le scripteur 8 est celui qui a respecté au mieux les conditions avec en moyenne des caractères 2 fois plus larges et 2.2 fois plus hauts ;
- en règle générale les scripteurs augmentent plus la hauteur  $H$  de leurs caractères que la largeur  $W$ . Ainsi les scripteurs ont en moyenne des caractères *grands* 1.85 fois plus hauts et 1.59 fois plus larges que leurs caractères *normaux* ;
- les scripteurs ont tendance à mieux respecter la consigne pour certaines lettres que d'autres. Par exemple le rapport moyen des largeurs  $W$  est de 1.7 pour la lettre a, contre 1.4 pour la lettre l. Ainsi on note que les variations des dimensions dépendent essentiellement des scripteurs étudiés, et non pas des lettres.

L'évolution des variables  $W$  et  $H$  des caractères d'un scripteur selon la session d'écriture (soit en fonction des cinq pages  $I - V$  de l'expérience 2) est étudiée pour déterminer s'il existe des variations de l'écriture en fonction du temps. Dans la Figure 4.10 est représenté

## Rapport des dimensions ( $W$ et $H$ ) entre caractères *grands* et *normaux*

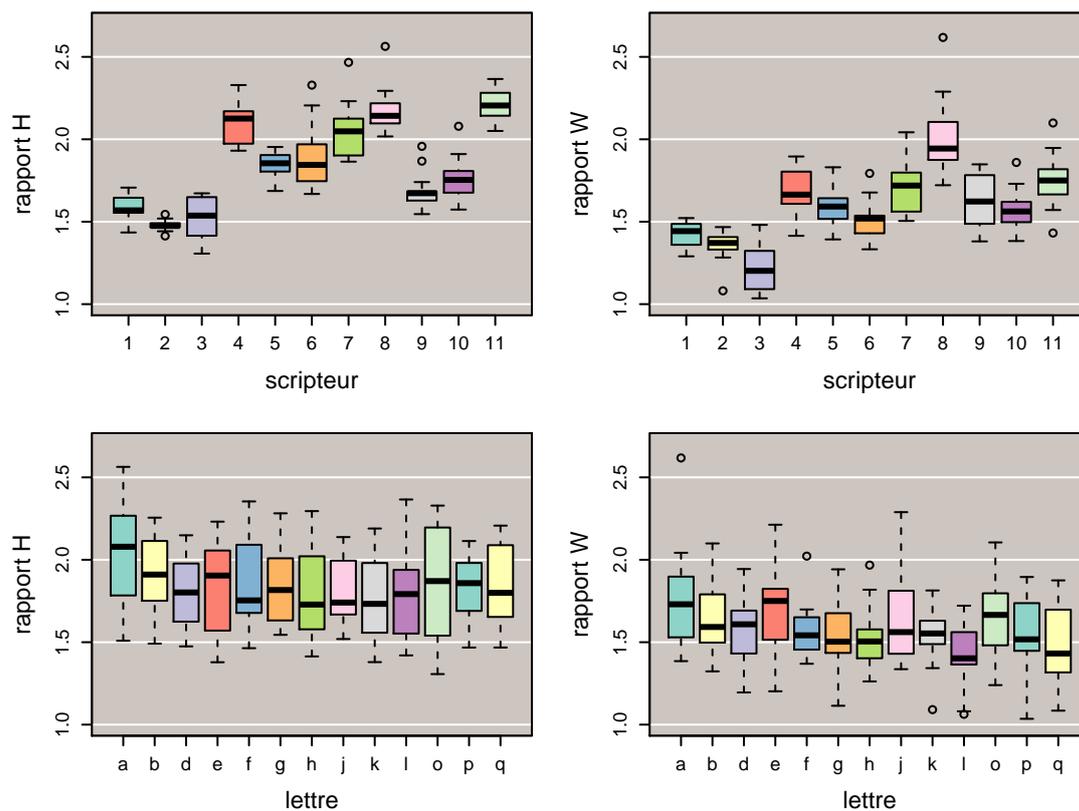


FIGURE 4.8: Illustration des rapports entre la largeur ( $W$ ) et hauteur ( $H$ ) des caractères *grands* et *normaux* pour les 11 scripteurs (en haut) et les 13 lettres (en bas). Les graphiques montrent qu'aucun des caractères *grand* des 11 scripteurs est 3 fois supérieur à la taille (représentés ici par  $W$  et  $H$ ) des caractères issus de leur écriture habituelle, et par conséquent que la consigne n'a pas été respectée par les scripteurs. Par ailleurs on note que le rapport entre largeur ( $W$ ) et hauteur ( $H$ ) est plus caractéristique d'un scripteur que d'une lettre, ce qui souligne la variabilité *inter-scripteur* (entre scripteurs) d'une mesure.

l'évolution de la variable  $W$  pour les scripteurs et lettres 7-g, 7-d et 10-d<sup>2</sup>. Les résultats indiquent que :

- les largeurs  $W$  varient en fonction de la session d'écriture, soit lorsque les échantillons ont été produits à des dates différentes. Par exemple, pour la lettre d et le scripteur 10, les moyennes des  $W$  augmentent progressivement à chaque session. Ces résultats indiquent l'existence d'une variation de l'écriture sur une période temporelle courte ;
- dans une session donnée, ces variations ne sont pas similaires selon le scripteur étudié. Par exemple, pour le scripteur 7, la largeur des caractères *normaux* de la lettre g est en moyenne inférieure dans les sessions I et V en comparaison aux sessions II – IV ; tandis que pour les caractères *grands*, ces valeurs dans sessions III – V sont inférieures aux sessions I, II et IV. On remarque donc, que pour 7-g,

2. Dans le but de simplifier les résultats, seule la variable  $W$  est représentée dans cette section. Le restant des graphiques est disponible dans l'appendice.

## Rapport W/H

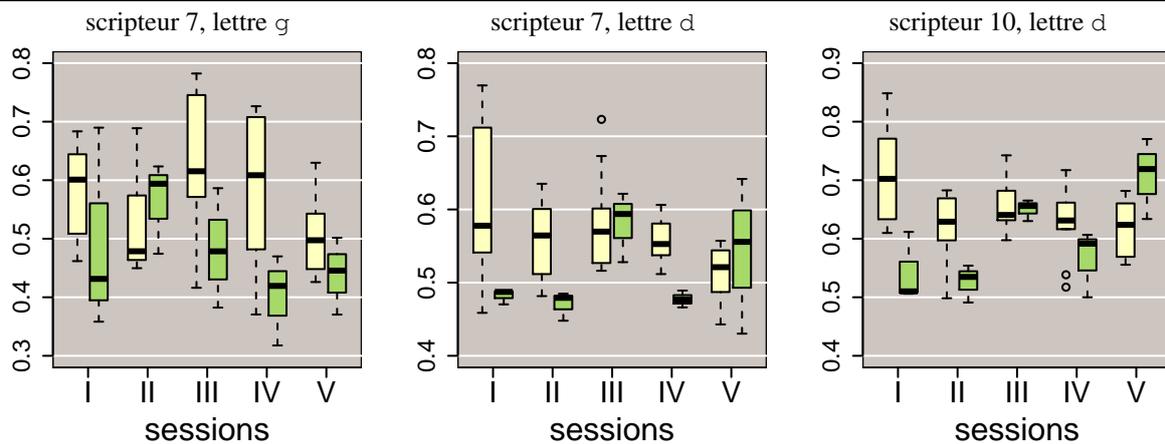


FIGURE 4.9: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ , pour les scripteurs-lettres 7-g, 7-d et 10-d. Par exemple, on observe pour 7-d que les rapports  $W/H$  des caractères *normaux* sont supérieurs aux caractères *grands* pour les sessions  $I$ ,  $II$  et  $IV$ . Concernant 7-d, les rapports  $W/H$  des caractères *normaux* sont aussi supérieurs pour les sessions  $I$ ,  $II$  et  $IV$ . En revanche, il arrive que ces rapports sont inférieurs pour la session  $V$ ) et égaux, pour  $III$ , en comparaison aux caractères *grands*. Ces résultats montrent donc une variation de l'écriture dans un espace temporel court.

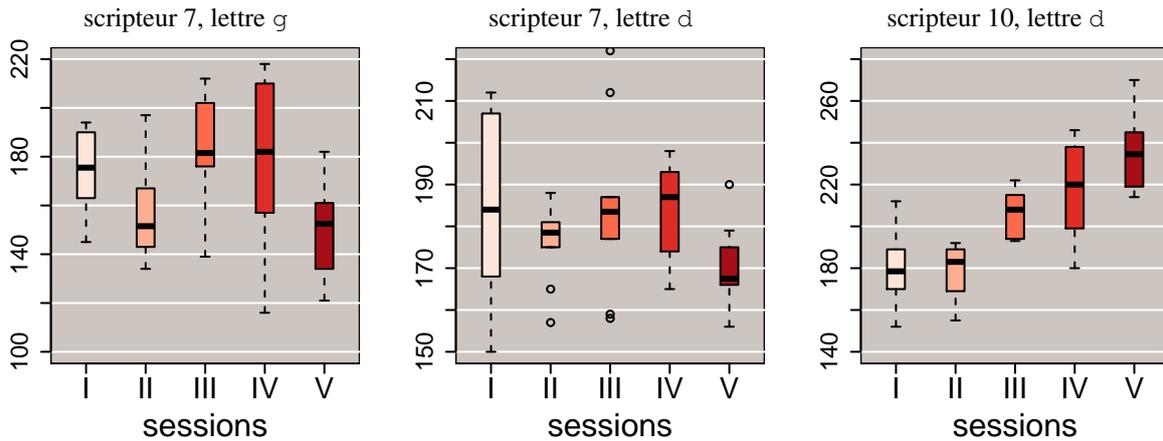
les moyennes des largeurs  $W$  varient entre sessions sans avoir un motif particulier. A l'inverse, les largeurs de 10-d des caractères *normaux* et *grands* suivent la même tendance. Ces résultats mettent en avant la variabilité de l'écriture entre scripteurs.

- ces variations (selon la session) ne sont pas similaires selon la lettre étudiée. Par exemple, pour le scripteur 7, les largeurs des caractères *grands* de la lettre d sont en moyenne inférieures dans les sessions  $I$ ,  $II$  et  $IV$ , alors que pour la lettre g, ceci était valable pour les sessions  $IV - V$ . Ces résultats mettent en avant la variabilité de l'écriture propre à chaque scripteur.

En conclusion ces résultats soulignent que :

1. les 11 scripteurs n'ont pas tous écrit leurs caractères *grands* à une taille trois fois supérieure à celles des caractères *normaux*, comme il l'était demandé lors de la prise d'échantillons. Toutefois, les différences entre les tailles des caractères *normaux* et *grands* sont suffisantes pour considérer que l'expérience 2 permet d'étudier les différences entre caractères de taille différente. Par ailleurs, ces résultats laissent suggérer que les scripteurs n'ont pas autant d'entraînement avec l'écriture *grande* qu'avec l'écriture *normale*. Par conséquent l'écriture des caractères *grands* apparaît comme une écriture inhabituelle des scripteurs, et constitue donc un facteur supplémentaire à considérer dans la discussion des résultats ;
2. les mesures  $W$ ,  $H$ , et  $W/H$  extraites des caractères provenant des 11 scripteurs de l'expérience 2 varient selon la session étudiée, et par conséquent montrent qu'il existe une variation de l'écriture sur un espace temporel court.

### Largeur W, caractères *normaux*



### Largeur W, caractères *grands*

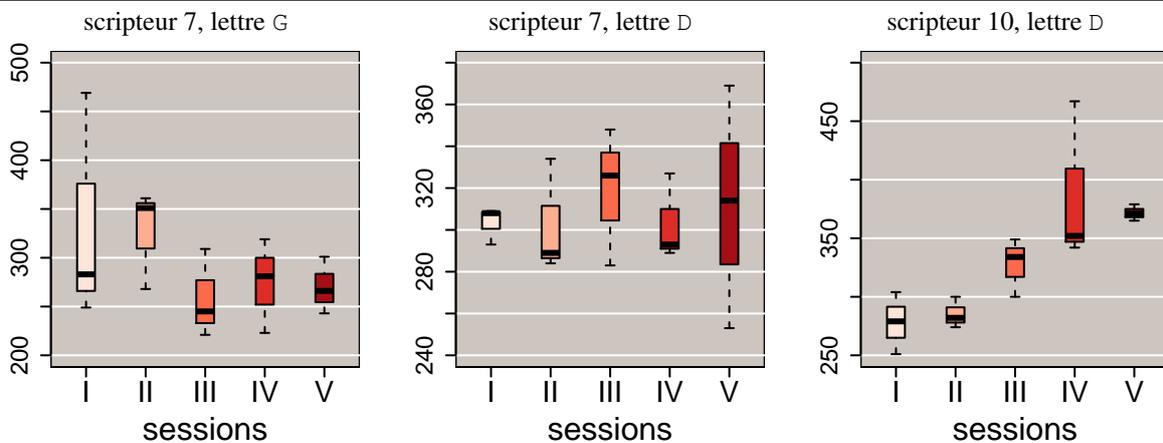


FIGURE 4.10: Boîtes à moustache des largeurs  $W$  selon les pages  $I - V$ , les scripteurs-lettres 7-g, 7-d et 10-d, et les caractères *normaux* et *grands*. On remarque, pour 10-d que les largeurs augmentent graduellement en fonction des cinq sessions. Cette remarque n'est pas valable pour les graphiques du scripteur 7. Par exemple pour les caractères *normaux* de la lettre g, les moyennes des sessions  $I$  et  $V$  sont inférieures aux autres, tandis que pour les caractères *grands*, ce sont les sessions  $III - V$  qui sont inférieures.

## 4.2.2 Etude de la distribution des $F_{ratio}$

Cette section a pour but d'étudier la distribution des  $F_{ratio}$  calculés dans les scénarios de l'expérience 2. Pour rappel, dans le cadre de cette recherche le  $F_{ratio}$  est une métrique statistique qui permet de vérifier si le facteur *scripteur* est significatif ou non (soit un facteur à deux niveaux : le scripteur du matériel *indiciaire* et celui du matériel de *référence*). Il est alors admis que si l'hypothèse  $H_0$  est rejetée (soit non-significatif), alors les scripteurs du matériel *indiciaire* et de *référence* sont différents. A l'inverse, s'il n'y a pas assez d'éléments pour rejeter  $H_0$ , alors il est admis que les scripteurs du matériel *indiciaire* et de *référence* sont les mêmes. A terme, les  $F_{ratio}$  sont utilisés pour le calcul des rapports de vraisemblance de l'expérience (voir les résultats de la section 4.2.3).

Dans la suite de cette étude, les  $F_{ratio}$  obtenus sont divisés en deux groupes, comme décrit précédemment dans la section 2.5 :

- 1) le groupe *intra-scripteur* pour les  $F_{ratio}$  issus des situations où le scripteur du document *indiciaire* et de *référence* est le même (par exemple le scripteur *1 vs 1* ou *2 vs 2*);
- 2) le groupe *inter-scripteur* pour les  $F_{ratio}$  issus des situations où les scripteurs du document *indiciaire* et de *référence* ne sont pas les mêmes (par exemple les scripteurs *1 vs 2* ou *2 vs 7*).

L'analyse conjointe des  $F_{ratio}$  *intra-scripteur* et *inter-scripteur* permet donc de déterminer si les mesures de dissimilarités extraites entre les échantillons sont capables de discriminer les scripteurs. A noter qu'on appelle les  $F_{ratio}$  calculés par le même nom que leur *fonctions d'extraction* respectives, afin de simplifier les notations. De ce fait les  $F_{ratio}$  sont notés  $FE_1$  et  $FE_2$ .

L'objectif est d'étudier à travers l'analyse conjointe des  $F_{ratio}$  *intra-scripteur* et *inter-scripteur* :

1. si les mesures de dissimilarités extraites entre les échantillons sont capables de discriminer les scripteurs (*c.-à-d.* une plus grande séparation entre les distributions *intra-scripteur* et *inter-scripteur* indique une meilleure discrimination);
2. s'il existe des différences entre caractères définis comme étant *normaux* et *grands*.

Tout d'abord, il est nécessaire d'étudier la normalité des  $F_{ratio}$ . Cet aspect est important car dans cette recherche la métrique d'évaluation, à savoir le rapport de vraisemblance, est basée sur une loi normale multivariée. Les représentations visuelles des distributions des  $F_{ratio}$  montrent que les distributions des  $F_{ratio}$  dites *intra-scripteur* et *inter-scripteur* possèdent des représentations asymétriques négatives (penchées à droite de leur moyenne), et ne semblent pas suivre la distribution d'une loi normale (voir Figure 4.11). Pour tenter de corriger une asymétrie négative, il est possible de transformer les données avec des fonctions comme  $1/x$ ,  $1/x^2$  ou  $\log(|x|)$ . La fonction  $\log(|x|)$  donne des résultats s'approchant de la distribution de la loi normale. Ainsi pour la suite des expériences 2 et 3, les résultats des  $F_{ratio}$  sont exprimés par le logarithme de leur valeur absolue  $\log(|FE_i|)$ . Cette transformation permet aux  $F_{ratio}$  de se rapprocher d'une loi normale. Il est également mis en évidence dans la suite de cette recherche que les résultats obtenus dans les étapes de calcul de dissimilarité (section 4.2.4) et d'analyse de dissimilarité (section 4.2.2), à savoir les étapes ne faisant pas intervenir le *RV* et donc où la normalité des  $F_{ratio}$  n'a aucune influence, sont similaires à ceux des rapports de vraisemblance. En conclusion, il convient donc d'accepter la loi normale comme la distribution des  $F_{ratio}$  pour le calcul du rapport de vraisemblance (section 4.2.3). Ainsi lorsque les  $F_{ratio}$  s'approchent de la valeur 1 (et donc les variables  $\log(|FE_i|)$  proches de la valeur 0), ils suggèrent qu'il n'y a pas de différences entre les dissimilarités des échantillons.

La discrimination d'écriture de scripteurs en fonction du nombre d'échantillons est étudiée à travers les *scénarios* 1 à 4. Il s'agit de situations similaires (même procédure d'échantillonnage, même type de caractères, et même jeu de données), où le niveau d'échantillonnage diffère ( $n_s = 5, 10, 15, 20$ , respectivement). Les résultats des distributions globales (tous scripteurs confondus) montrent que (voir Figure 4.12) :

- les moyennes des distributions *intra-scripteur* sont stables en fonction du nombre d'échantillons. Par exemple, dans les cas illustrés dans la Figure 4.12 les moyennes des variables se situent autour de 0, soit un  $F_{ratio}$  proche de 1 ( $\log(1) = 0$ ). Ces

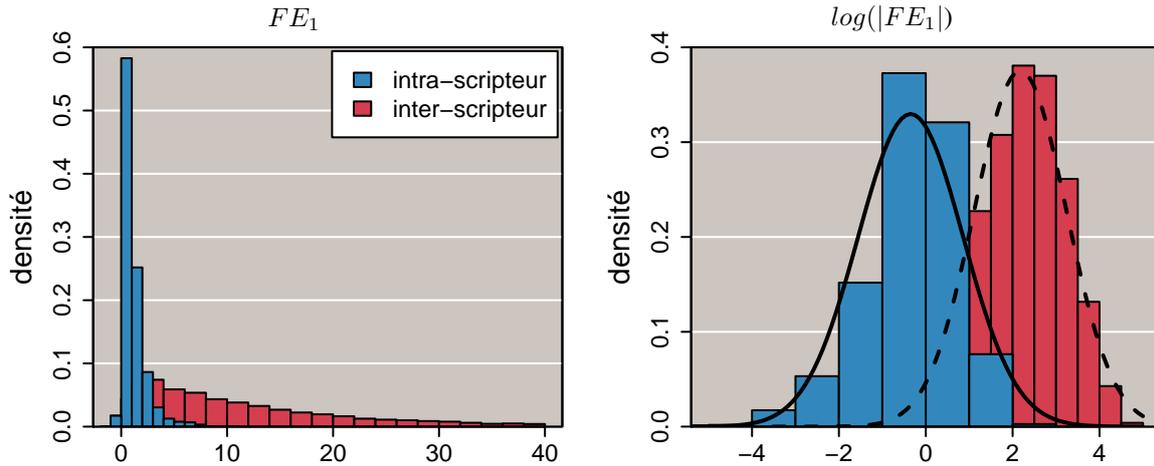


FIGURE 4.11: Illustration des distributions *intra-scripteur* et *inter-scripteur* de a) de la variable  $FE_1$  et b) de leurs transformées respectives avec  $\log(|FE_1|)$ . Les courbes en noir sur la figure b) représentent l'estimation des densités *intra-scripteur* et *inter-scripteur* sous la loi normale. (caractère **q**,  $n_s = 5$ .)

mesures ne suggèrent pas de différences entre les dissimilarités des échantillons, indépendamment du nombre d'échantillons (soit  $H_0$ ) ;

- les moyennes des distributions *inter-scripteur* sont supérieures à  $2 \approx \log(7.4)$  et augmentent en fonction du nombre d'échantillons  $n_s$ . Ceci indique que les dissimilarités calculées entre les échantillons provenant de scripteurs différents ne sont pas similaires (soit  $H_1$ ), et que plus le nombre d'échantillons augmente, plus cette différence s'accroît. Cependant ces augmentations ne sont pas homogènes suivant :
  - a) la lettre étudiée (Figure 4.12). En effet, en prenant la moyenne des variables  $FE_1 - FE_2$ , on remarque qu'elle augmente (entre  $n_s = 5$  et  $n_s = 20$ ) de 2.5 à 3.5 pour la lettre **q**, contre 3 à 4.5 pour la lettre **e** ;
  - b) le(s) scripteur(s) comparé(s). Par exemple, considérons les résultats des comparaisons du scripteur 7 avec les scripteurs 1, 2 et 8 pour la lettre **a** (voir Figure 4.13). Les scripteurs les plus similaires avec 7, dans l'ordre croissant, sont 1, 2, 8 pour  $n_s = 5$  et 2, 1, 8 pour  $n_s = 15$ .

L'influence de la taille des caractères analysés est étudiée à travers les *scénarios* 1, 5 et 6. Les *scénarios* 1 et 5 font appel aux situations où les caractères manuscrits du matériel *indiciaire* et de *référence* sont de même taille. C'est-à-dire les situations où les caractères comparés sont tous soit des caractères dits *normaux*, soit tous des caractères dits *grands*. Inversement, le *scénario* 6 fait appel aux situations où les échantillons du matériel *indiciaire* sont des caractères *grands* et les échantillons du matériel de *référence* des caractères *normaux*. Ces résultats (voir Figure 4.14) font apparaître que :

- si le même type de caractère est comparé (*c.-à-d.* les comparaisons *normal vs. normal* et *normal vs. grand*) les distributions *intra-scripteur* et *inter-scripteur* sont similaires. En effet, les moyennes des distributions divergent légèrement, cependant elles restent proches de la valeur 0, suggérant que les dissimilarités entre les échantillons de même taille d'un scripteur sont similaires. Les distributions *intra-scripteur*

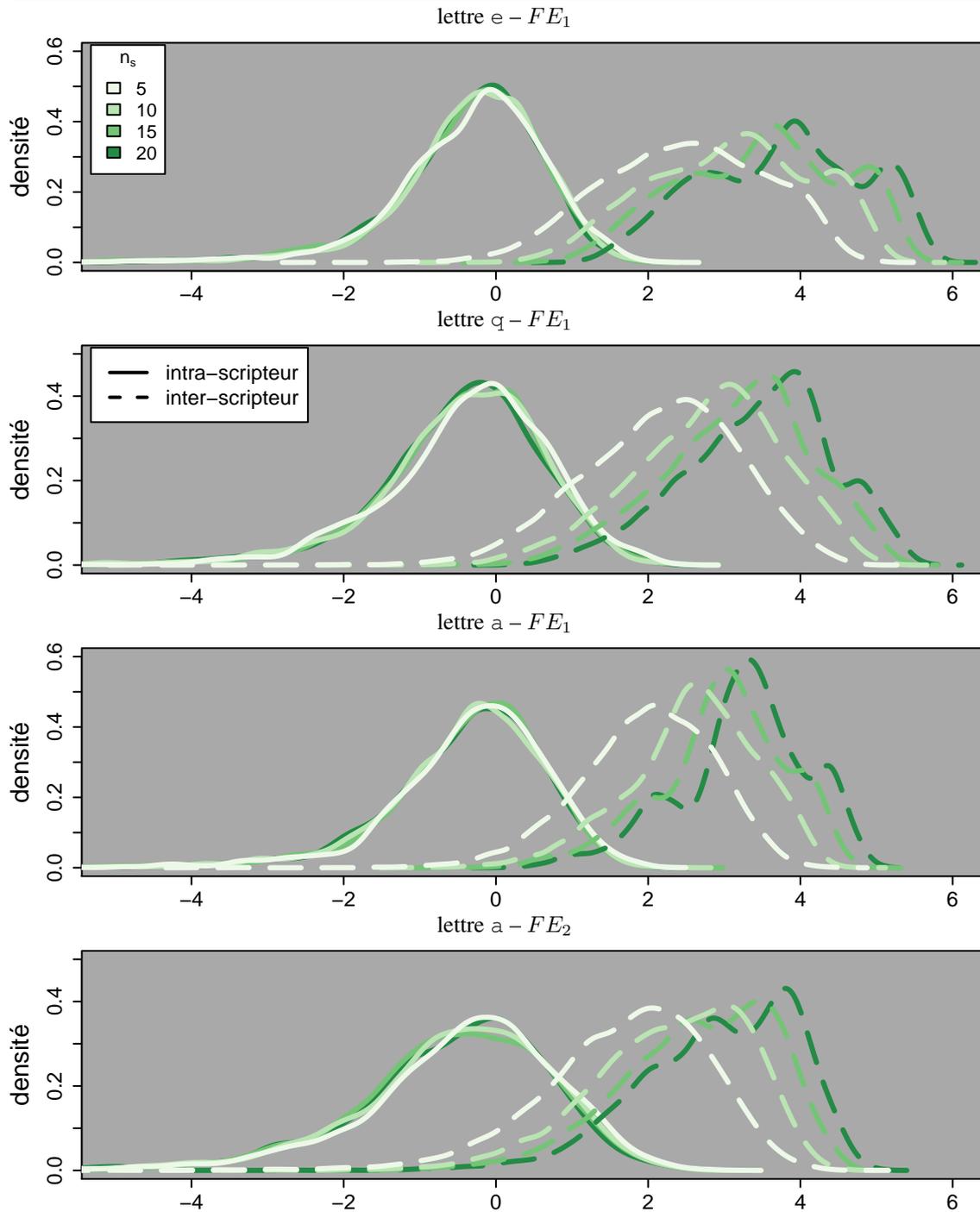


FIGURE 4.12: Exemples d'illustration de distributions des variables  $(FE_i - FE_j)$  obtenues dans les scénarios 1 à 4 de l'expérience 2, soit à différents niveaux d'échantillonnage :  $n_s = 5, 10, 15, 20$ . Ces graphiques mettent en évidence que la distribution *intra-scripteur* des variables  $FE_i$  (c.-à-d. la distribution des  $F_{ratio}$ ) est stable en fonction du nombre d'échantillons  $n_s$ . A l'inverse, les *inter-scripteur* des variables  $FE_i$  augmentent avec  $n_s$ . Ainsi les différences entre les valeurs *intra-scripteur* et *inter-scripteur* des variables s'accroissent lorsque le nombre d'échantillons augmente, indiquant une meilleure discrimination de scripteurs.

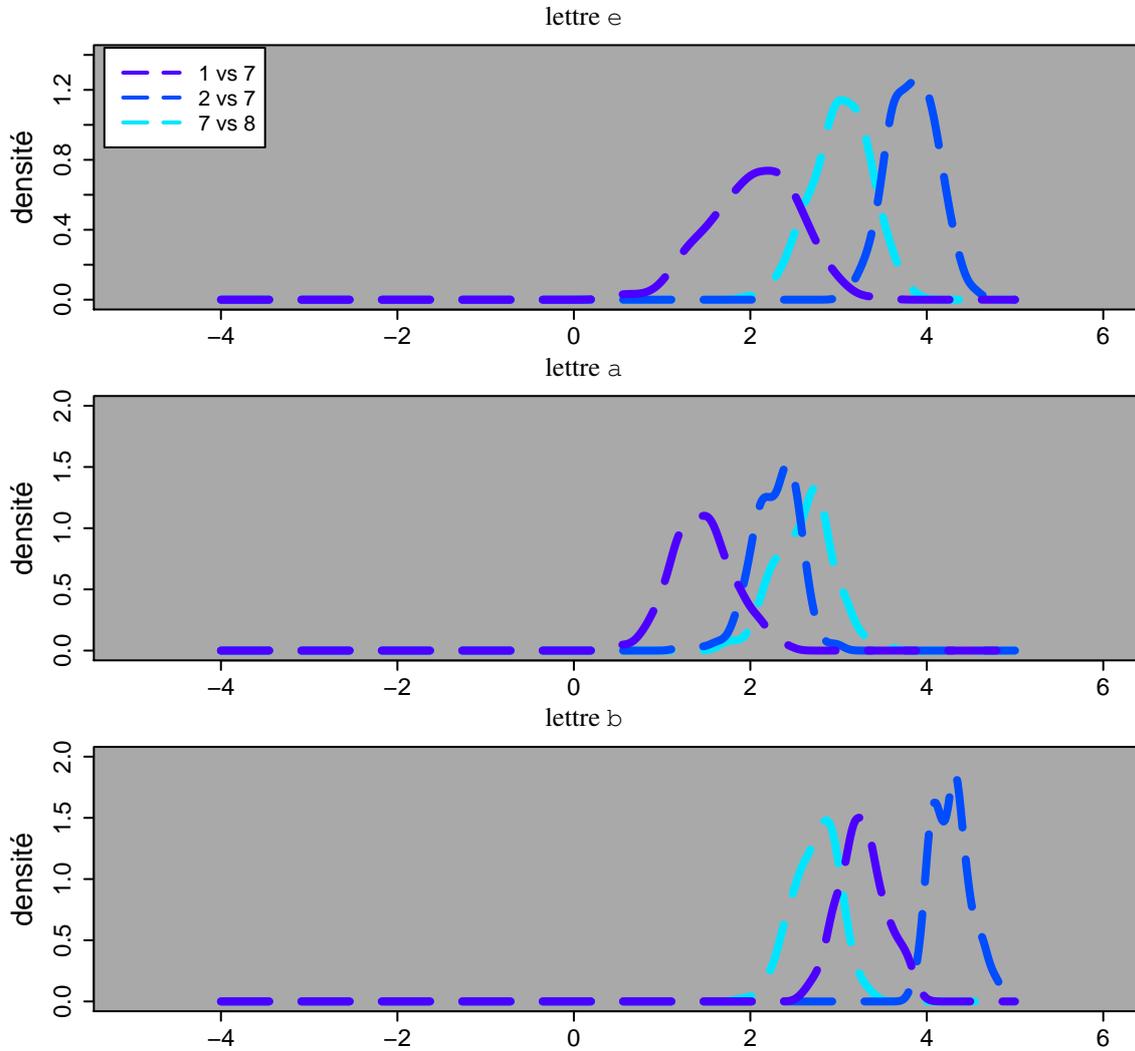


FIGURE 4.13: Exemples d’illustration de distributions *inter-scripteur*, en fonction des comparaisons des scripteurs  $1 vs 7$  et  $2 vs 7$ ,  $7 vs 8$ , obtenues dans le scénario 2 ( $n_s = 5$ ) et la variable  $FE_1$ . Les distributions *inter-scripteur* varient en fonction du scripteur et de la lettre, et ne sont pas similaires à celles observées dans la Figure 4.12. Par exemple, les résultats des comparaisons du scripteur 7 signalent que les scripteurs les plus similaires sont, dans l’ordre croissant, 1, 2, 8 pour la lettre a et 2, 1, 8 pour la lettre e. Par ailleurs, cet aspect souligne que les variations individuelles ne sont pas observables lorsque les scripteurs sont étudiés globalement, comme dans la Figure 4.12.

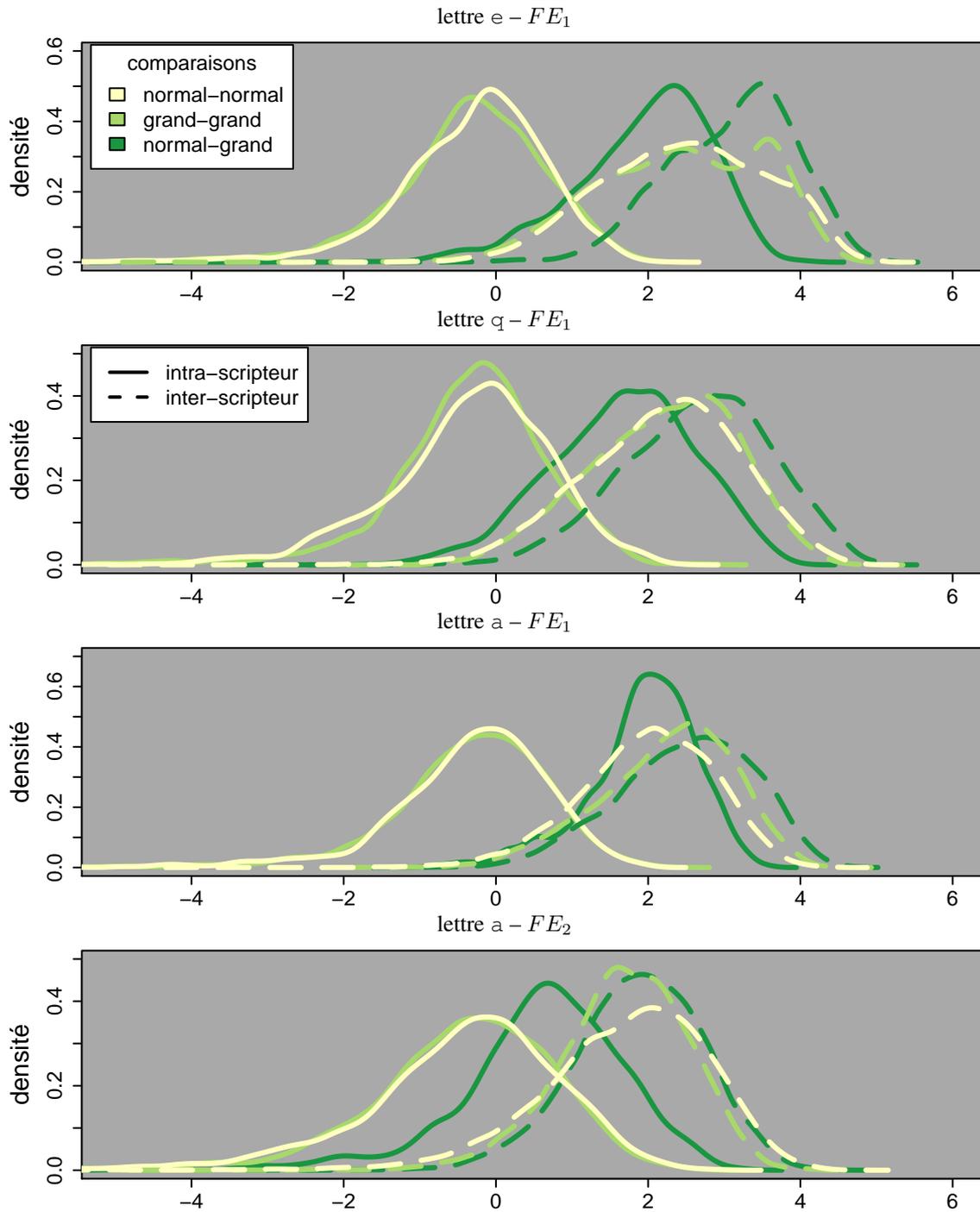


FIGURE 4.14: Exemples d’illustration de distributions *intra-scripteur* et *inter-scripteur* pour les scénarios 1, 5 et 6 de l’expérience 2 (ces scénarios sont les situations où les caractères comparés sont *normaux vs. normaux*, *grands vs. grands*, et *normaux vs. grands*). On observe que pour les scénarios 1 et 5 les distributions *intra-scripteur* et *inter-scripteur* sont similaires. Par contre, les distributions *intra-scripteur* ne sont pas similaires entre le scénario 6 et les scénarios 1 et 5. En effet, les valeurs des variables (*c.-à-d.* les  $F_{ratio}$ ) sont plus hautes pour le scénario 6, suggérant une dissimilarité entre les caractères de dimensions différentes d’un scripteur.

mettent en évidence que lorsque le même type de caractère est comparé, l'hypothèse  $H_0$  est respectée ;

- si les caractères comparés sont de taille différente (*c.-à-d.* les comparaisons *normal vs. grand*) alors la distribution *intra-scripteur* migre en direction des distributions *inter-scripteur*. Dans ce cas, les distributions *intra-scripteur* indiquent que l'hypothèse  $H_0$  n'est pas respectée. De ce fait ces résultats suggèrent que les dissimilarités entre les caractères de taille différente ne sont pas similaires, et donc qu'il existe des différences entre ces caractères ;
- ces deux dernières remarques ne sont pas nécessairement généralisables à tous les scripteurs. En effet, lorsque les scripteurs sont étudiés individuellement (voir Figures 4.13 et 4.15), les résultats indiquent que :
  - les remarques sont similaires pour les distributions *intra-scripteur* ;
  - en revanche les distributions *inter-scripteur* varient en fonction des scripteurs analysés. Cet élément traduit la notion intuitive qu'un individu peut avoir une écriture plus similaire à un scripteur qu'à un autre. Par ailleurs, ceci indique que l'étude des distributions globales des Figures 4.12 et 4.14 (tous scripteurs confondus) masque les variations individuelles qui sont observées dans les Figures 4.13 et 4.15.

Les résultats des Figures 4.13 et 4.15 illustrent les variations individuelles et soulignent, pour un *scénario* donné, la part de variation due au(x) scripteur(s) dans la variation globale. Pour un scripteur donné, les résultats mettent en évidence les remarques générales suivantes :

- les distributions *intra-scripteur* pour les paramètres étudiés ( $n_s$ , lettre, mesure) sont stables et de même ordre de grandeur que pour les distributions *intra-scripteur* globales, indépendamment de la lettre et des scripteurs étudiés. L'étude globale des distributions *intra-scripteur* fournit donc les mêmes résultats que pour les distributions individuelles ;
- la moyenne des distributions *inter-scripteur* varie beaucoup plus comparée à celle des distributions *inter-scripteur* globales. Par exemple, pour la lettre e et le *scénario 2* ( $n_s = 10$ ), la comparaison des scripteurs 1 vs 7 donne un  $F_{ratio}$  en moyenne de 2.04, contre 3.02 pour 7 vs 8 et de 3.84 2 vs 7 (voir Figure 4.13). Ces résultats indiquent que le degré de dissimilarité entre l'écriture de scripteurs différents, mesuré par le  $F_{ratio}$ , dépend du couple de scripteurs étudié. Ainsi il apparaît que certains individus ont une écriture plus similaire à certains scripteurs que d'autres.

En conclusion les résultats de cette section soulignent que :

1. les  $F_{ratio}$  obtenus à partir des matrices de dissimilarité apparaissent comme une métrique permettant de discriminer l'écriture de scripteurs, comme l'indique la séparation entre les distributions *intra-scripteur* et les distributions *inter-scripteur*. En effet, les résultats de ces mesures sont conformes aux hypothèses émises dans la section 2.5, à savoir :
  - a) lorsque les échantillons proviennent du même scripteur, soit les situations faisant référence à l'hypothèse  $H_0$  (*c.-à-d.* qu'il n'y a pas de différence entre les dissimilarités des échantillons, et soutiennent donc l'hypothèse que le scripteur

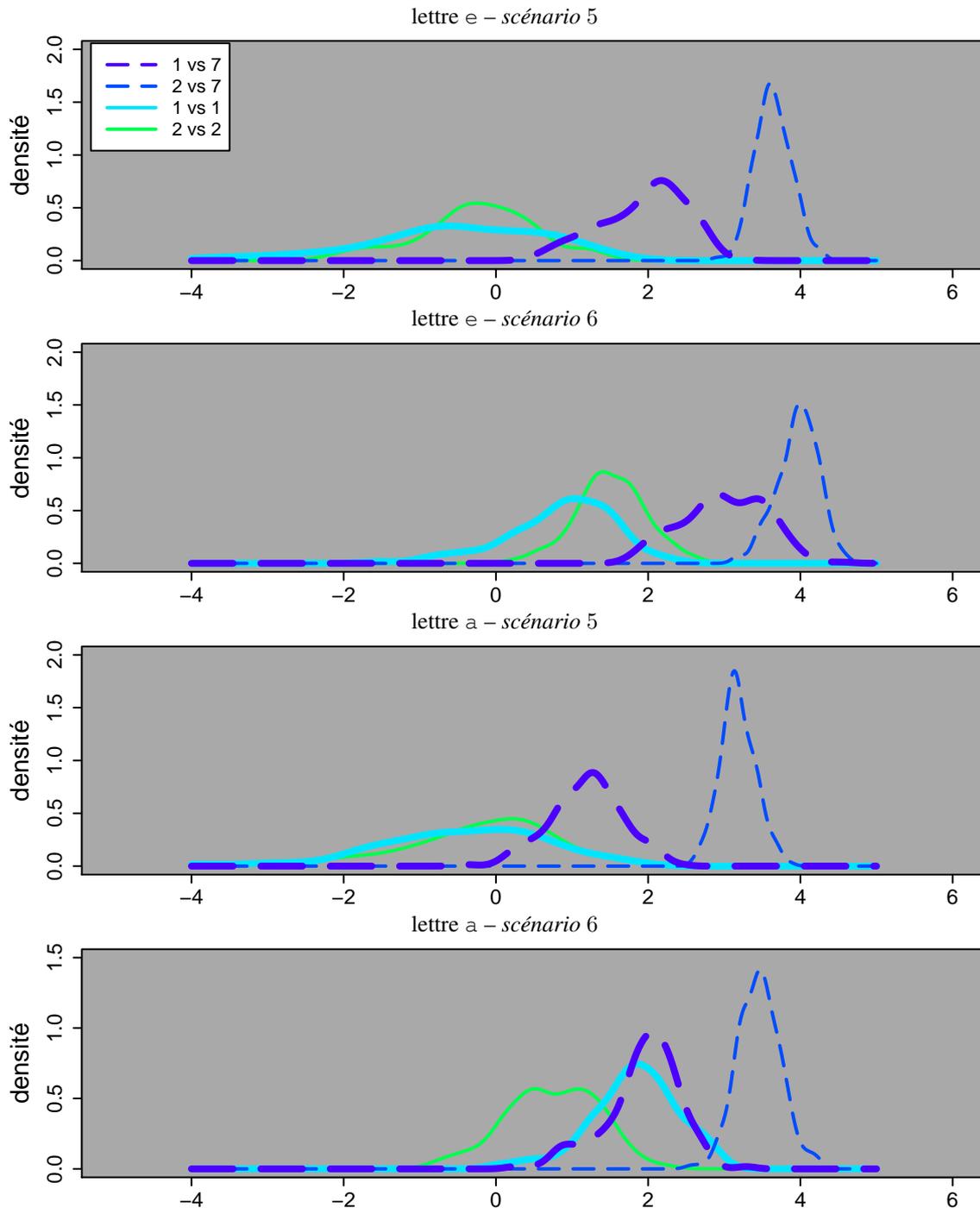


FIGURE 4.15: Exemples d’illustration de distributions *intra-scripteur* (1 vs 1 et 2 vs 2) et *inter-scripteur* (1 vs 7 et 2 vs 7), obtenues dans les scénarios 5 et 6 de l’expérience 2. Les résultats montrent que les distributions *intra-scripteur* du scénario 5 sont stables et centrées autour de la valeur  $0 = \log(1)$ , au contraire celles du scénarios 6 sont éloignées de la valeur 1. Ceci met en évidence que pour un scripteur donné, les dissimilarités sont similaires lorsque le même type de caractère est étudié, et ne sont pas similaires quand les caractères ne sont pas de même type. Par ailleurs, on observe que les distributions, hormis les distributions *intra-scripteur* du scénario 5, varient largement selon le(s) scripteur(s) comparé(s). Ceci indique que l’étude des distributions globales (voir la Figure 4.14) masque les variations individuelles des scripteurs.

du matériel *indiciaire* et de *référence* est le même), les distributions *intra-scripteur* des  $F_{ratio}$  sont centrées autour de la valeur  $0 = \log(1)$ ;

- b) lorsque les échantillons proviennent des scripteurs différents, soit les situations faisant référence à l'hypothèse  $H_1$  (*c.-à-d.* qu'il existe une différence entre les dissimilarités des échantillons, et soutiennent donc l'hypothèse que le scripteur du matériel *indiciaire* et de *référence* sont différents), les distributions *inter-scripteur* ne sont pas centrées sur la valeur 0 et s'en éloignent progressivement lorsque le nombre d'échantillons augmente.

2. il existe une différence entre les caractères *normaux* et *grands* des scripteurs (à savoir, à leur taille habituelle et à une taille équivalente à trois fois celle de l'écriture habituelle, respectivement). En effet les distributions *intra-scripteur* du scénario 6, soit dans la situation où les échantillons proviennent du même scripteur et sont un mélange des caractères *normaux* et *grands*, sont écartés de la valeur 0 et indiquent qu'il existe une différence entre les dissimilarités des échantillons.

### 4.2.3 Les résultats du calcul des rapports de vraisemblance

Cette section présente les résultats du calcul des rapports de vraisemblance de l'expérience 2. Le rapport de vraisemblance est une métrique qui mesure le poids des observations au regard des deux hypothèses d'intérêt,  $H_p$  et  $H_d$ , afin de soutenir l'une d'elles. Donc, dans le cadre des scénarios fictifs de cette recherche, pour soutenir l'hypothèse  $H_p$  : le scripteur connu a écrit le document *indiciaire*; soit pour soutenir l'hypothèse  $H_d$  : un scripteur alternatif a écrit le document *indiciaire*<sup>3</sup>.

L'objectif de cette section est similaire à la section précédente, à savoir déterminer :

1. si les rapports de vraisemblance calculés permettent de discriminer l'écriture de scripteurs;
2. s'il existe des différences entre caractères définis comme étant *normaux* et *grands*.

Les résultats de *vérification multivariés* (soit avec les deux variables  $FE_1$  et  $FE_2$ ) et *univariés* (soit avec l'une des deux variables, ici  $FE_1$ , les résultats *univariés* de  $FE_2$  sont donnés en annexe) des scénarios 1 à 4 sont illustrés dans Figure 4.16 et indiquent que :

- les taux de *faux positifs* et *faux négatifs* baissent lorsque le nombre d'échantillons augmente, de manière similaire à une décroissance exponentielle. Cette baisse des taux d'erreur est donc modélisée par la fonction  $f(n_s) = N_0 e^{-\lambda \times n_s}$ , où  $\lambda$  représente la demi-vie du taux d'erreur en fonction de  $n_s$  (voir Figure 4.17). Cependant cette baisse :
  - est plus importante pour les *faux négatifs* que les *faux positifs*;
  - n'est pas constante lorsque les scripteurs sont étudiés individuellement. Par exemple, les taux de *faux négatifs* du scripteur 7 pour la lettre a sont plus hauts pour  $n_s = 10$  que  $n_s = 15$ .
- les variations des taux d'erreur en fonction de  $n_s$ , lorsque les scripteurs sont étudiés individuellement, sont plus importantes que celles des résultats globaux (moyenne des taux d'erreur tous scripteurs confondus). Par exemple pour la lettre e, les taux

---

3. A noter que chaque scénario est composé de  $n_s$  échantillons du matériel *indiciaire* et de  $n_s$  échantillons provenant du matériel de *référence*.

### Résultats des taux de vérification globaux

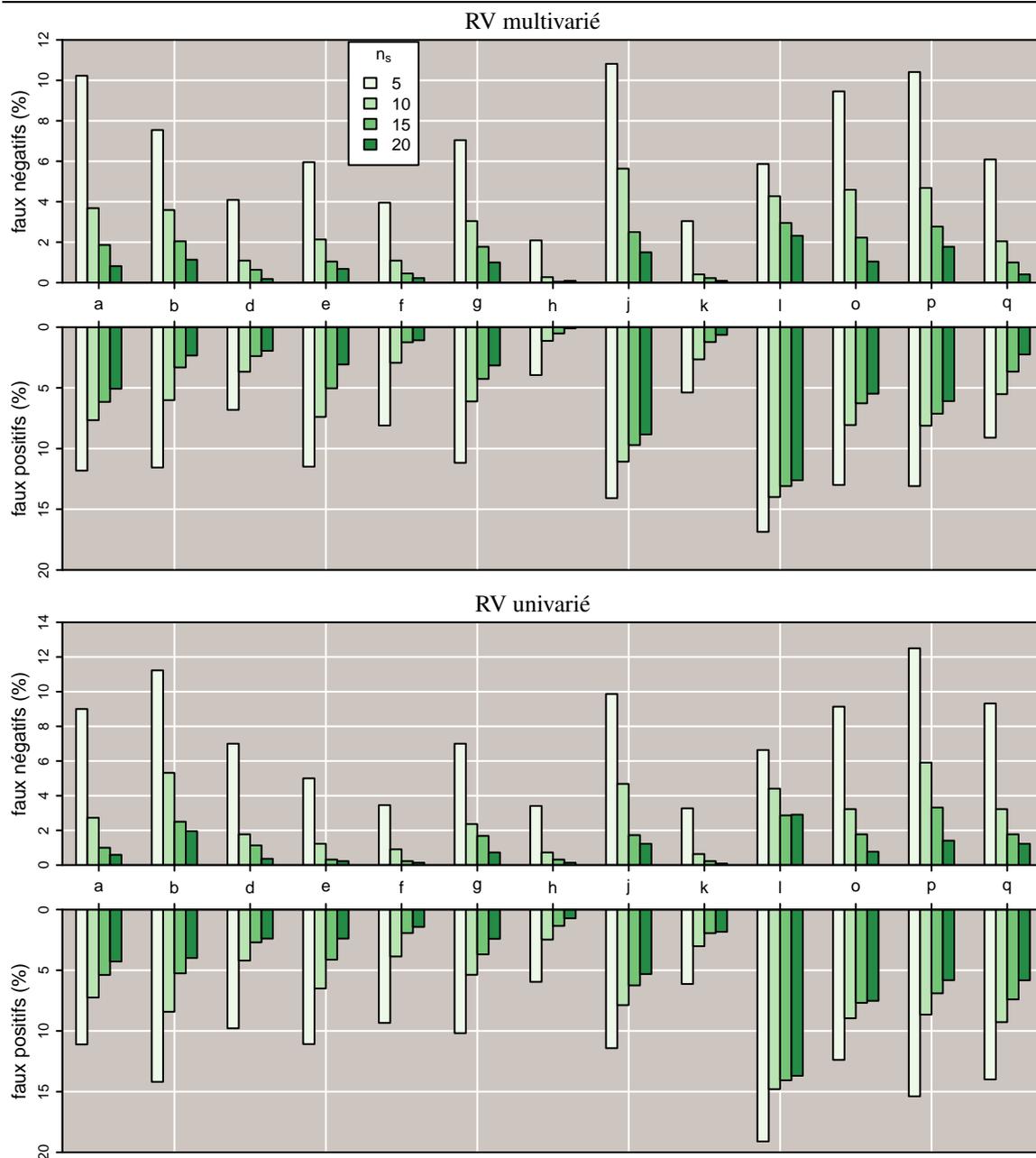


FIGURE 4.16: Illustration de l'influence de  $n_s$  sur les taux *faux négatifs* et *faux positifs* globaux multivariés et univariés ( $FE_1$ ). Les graphiques montrent que les taux d'erreur diminuent lorsque le nombre d'échantillons  $n_s$  augmente. Par ailleurs, on observe que certaines lettres sont plus discriminantes que d'autres, par exemple le taux d'erreur des lettres **h** et **k** sont plus faibles que ceux des lettres **a** et **o**.

de décroissance pour les comparaisons des scripteurs *2 vs 2* et *7 vs 7* (soit des résultats de *faux négatifs*) sont respectivement de  $\lambda = 0.78$  et  $\lambda = 0.09$ , alors que le taux global est de  $\lambda = 0.14$ . On note aussi que les valeurs des *faux négatifs* du scripteur 7 ne diminuent pas avec  $n_s$ , en effet le taux de *faux négatifs* est de 5.5% pour  $n_s = 5$  contre 7% pour  $n_s = 10$ . Ceci relève que les résultats globaux masquent les variations individuelles des scripteurs dans les résultats ;

- les taux globaux de *faux négatifs* sont inférieurs aux taux de *faux positifs* indépendamment de la lettre étudiée, du nombre d'échantillons et du type de *RV* (*multivarié/univarié*). Ceci, à l'exception des lettres **a**, **j** et **p** pour les *RV* univariés  $FE_2$  avec  $n_s = 5$  (par exemple 20% *faux négatifs* > 16.3% *faux positifs* pour la lettre **a**). Ces résultats montrent simplement que l'on obtient de meilleurs résultats lorsque le matériel comparé provient du même scripteur (correspondant aux *faux négatifs*), que lorsque le matériel est issu de deux scripteurs différents (correspondant aux *faux positifs*). En d'autres termes, la méthode développée est plus robuste sous  $H_p$  que sous  $H_d$  ;
- les résultats des rapports de vraisemblance *univariés* et *multivariés* montrent la même tendance générale pour l'influence du nombre d'échantillons, à savoir que, plus la valeur de  $n_s$  augmente, plus le taux d'erreur diminue. Ces résultats montrent que les observations des deux mesures  $FE_1$  et  $FE_2$  mènent aux mêmes conclusions.

L'influence du type de lettre est étudié à travers les *scénarios* 1, 5 et 6. Les *scénarios* 1 et 5 font appel aux situations où les caractères manuscrits du matériel *indiciaire* et de *référence* sont de même taille, soit respectivement des caractères *normaux* et *grands*. Inversement, le *scénario* 6 fait appel aux situations où les échantillons du matériel *indiciaire* sont des caractères *grands* et les échantillons du matériel de *référence* des caractères *normaux*.

Les résultats des taux de *faux positifs* et *faux négatifs* en fonction du type de lettre, pour le cas *univarié* et *multivarié*, sont illustrés dans la Figure 4.18. Ces résultats montrent que pour une lettre et un nombre d'échantillons donnés :

- les taux d'erreur quand les échantillons comparés ne sont pas de même taille sont tous supérieurs aux taux des *scénarios* 1 et 5. La différence minimale entre les taux d'erreur des *scénarios* 1 et 5 comparés au *scénario* 6 est de 3.7 points (lettre **b**, *faux négatifs*). Cependant, ces observations ne sont pas valables lorsque les scripteurs sont étudiés individuellement. Ceci souligne qu'il existe des différences entre les caractères *normaux* et *grands* des 11 scripteurs de la recherche.
- les taux globaux ne sont pas représentatifs des taux d'erreur individuels. Ainsi, le taux global de *faux positifs* pour le *scénario* 6 et la lettre **e** est de 23.6%. Alors que ce même taux est de 23% en comparant les scripteurs 1 et 7 et de 0% avec les scripteurs 2 et 7. Ces résultats mettent en évidence la variabilité de l'écriture au sein d'un groupe de scripteurs. Ils indiquent aussi le fait qu'une mesure qui est discriminante entre deux scripteurs ne l'est pas forcément pour d'autres. Il faut donc considérer cet aspect lors d'une expertise en écriture et ne pas se focaliser sur un seul type de mesure ;
- les résultats des rapports de vraisemblance *univariés* et *multivariés* montrent la même tendance générale pour chaque type de lettre (**a**, **b**, ..., **q**). Cet aspect souligne, tout comme les résultats de la Figure 4.16, que les observations des deux mesures  $FE_1$  et  $FE_2$  mènent aux mêmes conclusions.

### Taux de vérification multivariés: lettre e

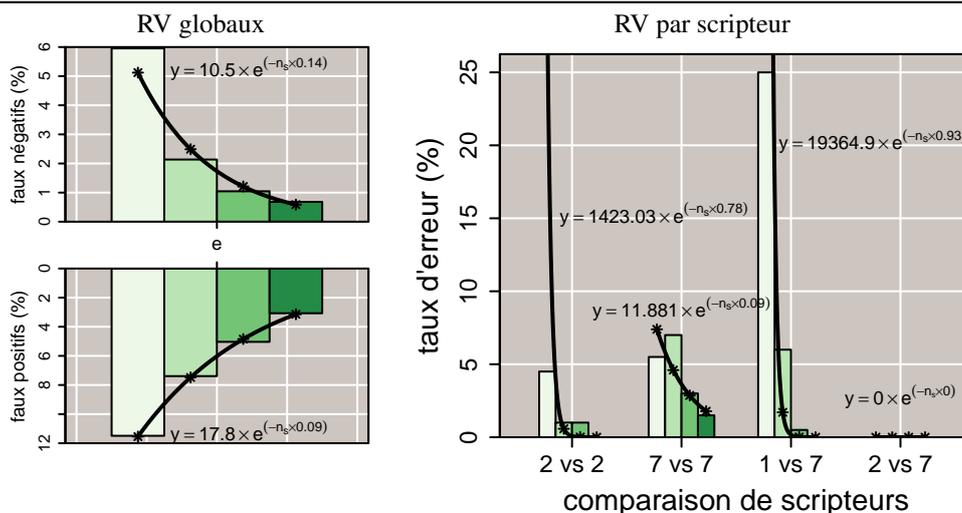


FIGURE 4.17: Illustration de l'influence du nombre d'échantillons ( $n_s$ ) sur les taux de vérification globaux (à gauche) et par scripteur(s) (à droite). Les courbes sur les graphiques représentent le modèle de décroissance exponentielle ajustée au données :  $f(n_s) = N_0 e^{-\lambda \times n_s}$ . On observe que le taux  $\lambda$  est supérieur pour les résultats des *faux négatifs* que pour ceux des *faux positifs*, indiquant une baisse plus rapide du taux *faux négatifs* en fonction du nombre d'échantillons  $n_s$ . En effet sur le graphique de gauche  $\lambda = 0.14$  pour les *faux négatifs* contre  $\lambda = 0.09$  pour les *faux positifs*. Le graphique de droite montre les différences des taux de décroissance lorsque les résultats sont étudiés par scripteurs, par exemple les taux de décroissance pour les comparaisons des scripteurs *2 vs 2* et *7 vs 7* (soit des résultats de *faux négatifs*) sont respectivement de  $\lambda = 0.78$  et  $\lambda = 0.09$ , alors que le taux global est de  $\lambda = 0.14$ . On note que les valeurs des *faux négatifs* du scripteur *7 vs 7* ne diminuent pas forcément avec  $n_s$ , en effet le *faux négatifs* est de 5.5% pour  $n_s = 5$  et 7% pour  $n_s = 10$ . Ces résultats mettent en évidence que les résultats globaux masquent les variations individuelles des scripteurs dans les résultats.

Lorsque les résultats des calculs de rapport de vraisemblance de l'expérience 2 sont regardés plus en détail, ils indiquent que les taux d'erreur ne sont pas similaires en fonction de(s) scripteur(s) étudié(s). Prenons l'exemple de la lettre a (voir la Table 4.1) :

- Dans les scénarios 1 à 5, pour le scripteur 4 les taux de *faux positifs* sont très faibles lorsqu'il est comparé aux scripteurs 5 et 8 (0% chacun pour tous les  $n_s$ ). Inversement les taux de *faux positifs* sont élevés lorsqu'il est comparé au scripteur 6 et augmentent avec  $n_s$  (45% pour  $n_s = 5$  et 85% pour  $n_s = 20$ ), ceci indique que l'écriture de l'écrivain 6 est hautement similaire à celle de l'écrivain 4 (dans le cas de la lettre a). Par contre les taux de *faux négatifs* sont moins variables, et oscillent entre 0% et 15.5%, tous scripteurs et nombre d'échantillons  $n_s$  confondus ;
- Dans le cas du scénario 6, les taux d'erreur sont en moyenne plus élevés que pour les scénarios 1 et 5 (l'étendue est de 0% à 97% pour les *faux négatifs* et les *faux positifs*). Par contre, cette tendance n'est pas généralisable. Par exemple, lorsque le scripteur 4 est comparé aux scripteurs 6 ou 8, le taux d'erreur est 9.5% pour le scénario 6 contre 45% et 52.5% pour les scénarios 1 et 5 ;

## Résultats des taux de vérification globaux

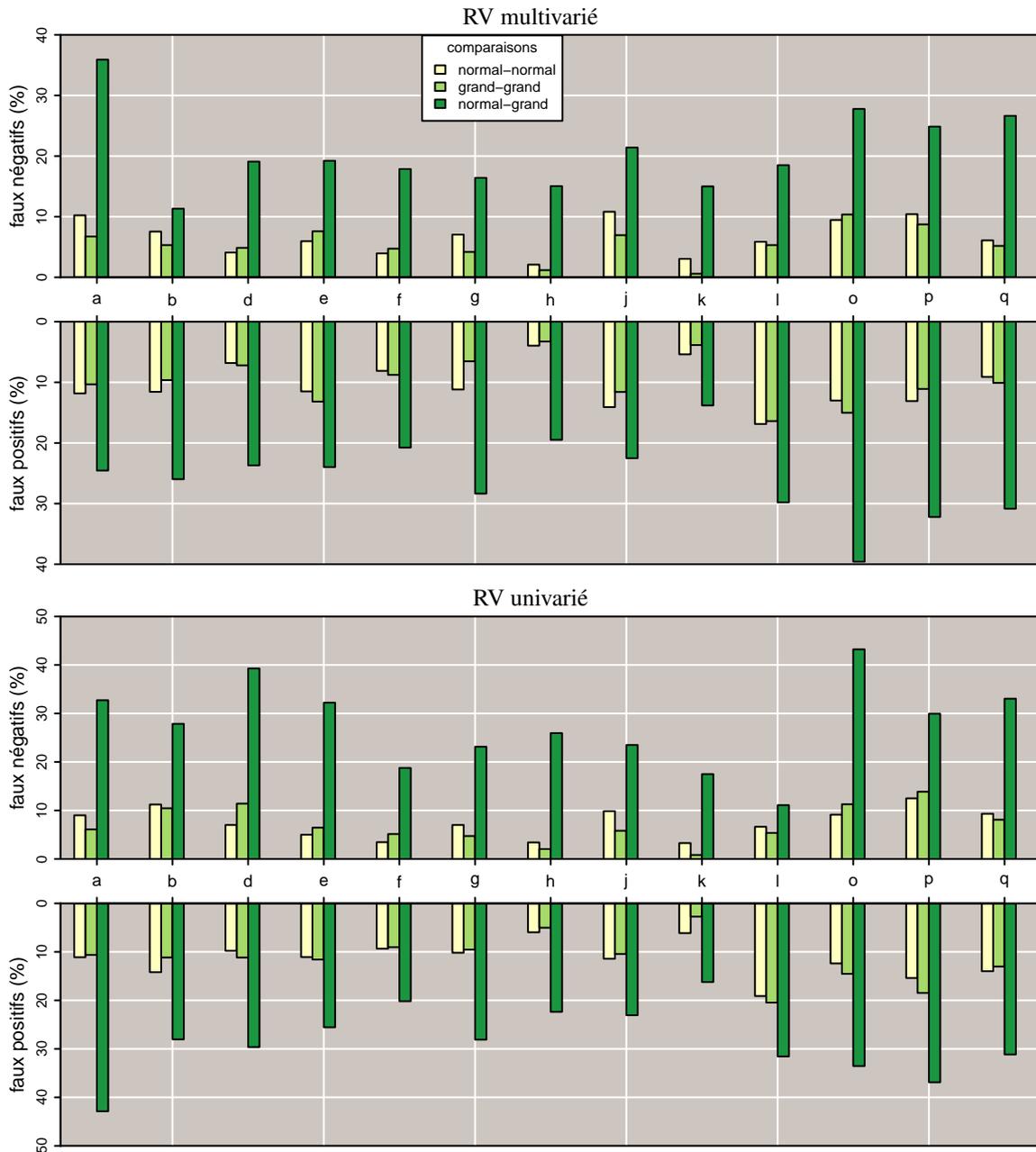


FIGURE 4.18: Illustration de l'influence du type de caractère (les caractères dits *normaux* et *grands* des scripteurs) sur le taux de *faux négatifs* et de *faux positifs* globaux multivariés et univariés (soit les scénarios 1, 5 et 6). Les résultats montrent que les taux de *faux positifs* et *faux négatifs* sont : 1) similaires lorsque les échantillons comparés sont de même taille (les situations où les échantillons sont tous des caractères *normaux* ou *grands*, et 2) inférieurs aux taux d'erreur lorsque les échantillons comparés ne sont pas de même taille. Ces résultats soutiennent l'hypothèse qu'il existe des différences entre des caractères de taille différente.

### Taux de vérification multivariés

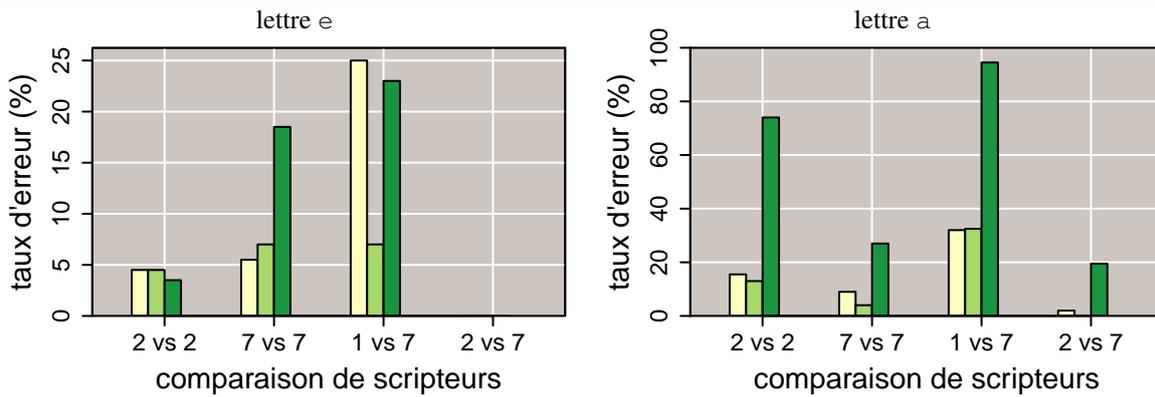


FIGURE 4.19: Illustration de l'influence du type de lettre sur les taux de *vérification* par scripteur (pour les lettres **e** et **a**). On observe, contrairement aux observations de la Figure 4.18, les taux d'erreur lorsque les caractères analysés sont de même dimensions ne sont pas tous inférieurs à ceux obtenus lorsque les caractères sont de dimensions différentes, comme en attestent les résultats de la comparaison des scripteurs *2 vs. 7* pour la lettre **a**. Par ailleurs, dans le cas de *2 vs. 7* et la lettre **e**, le taux d'erreur est supérieur lorsque les caractères comparés sont *normaux* que quand ils sont un mélange *normaux vs grands*.

Ainsi ces résultats mettent en évidence la variabilité de l'écriture au sein de chaque scripteur, et montre que l'étude seule des résultats globaux de *vérification* n'apporte que très peu d'information d'intérêt pour un expert en écriture. Par exemple, le taux d'erreur moyen ne donne aucune information sur la variabilité du taux d'erreur, ainsi que les particularités de chaque scripteurs. De ce fait, les résultats globaux de taux d'erreur ne portent pas autant de valeur dans le cadre forensique que dans celui de la Reconnaissance optique de caractère. Par ailleurs, les résultats soulignent l'intérêt de décrire la prise d'échantillons et les expériences effectuées afin d'interpréter les résultats obtenus. En effet, dans les exemples présentés ci-dessus les résultats diffèrent amplement entre les *scénarios*, alors que les mêmes scripteurs et les mêmes lettres sont étudiés (seuls le type de caractère *normal* ou *grand* change).

La distribution des *LRV*, pour les lettres **h** et **o**, est illustrée dans la Figure 4.21. Ces résultats montrent que :

1. les *LRV* forment des distributions différentes selon si l'hypothèse  $H_p$  est vraie, ou si  $H_d$  est vraie. Par exemple, dans le cas de la lettre **h** avec  $n_s = 20$ , la moyenne est de 17.95 (*écart-type* 7.52) pour  $H_p$  contre  $-13.35$  (*écart-type* 4.69) pour  $H_d$ , soit une différence de 31.3. Ces résultats participent à la validation des hypothèses de construction de la méthode, à savoir que des  $F_{ratio}$  des situations où les échantillons d'écriture proviennent du même scripteur donnent des *RV* supportant  $H_p$  ; et les  $F_{ratio}$  issus des situations où les échantillons d'écriture proviennent de scripteurs différentes donnent des *RV* supportant  $H_d$  ;
2. la différence entre les distributions du *LRV* de  $H_p$  et  $H_d$  augmentent avec le nombre d'échantillons ( $n_s$ ). Par exemple, pour la lettre **h**, cette différence est de 14.01 pour  $n_s = 5$ , contre 31.3 pour  $n_s = 20$ . L'augmentation de cette différence est moins



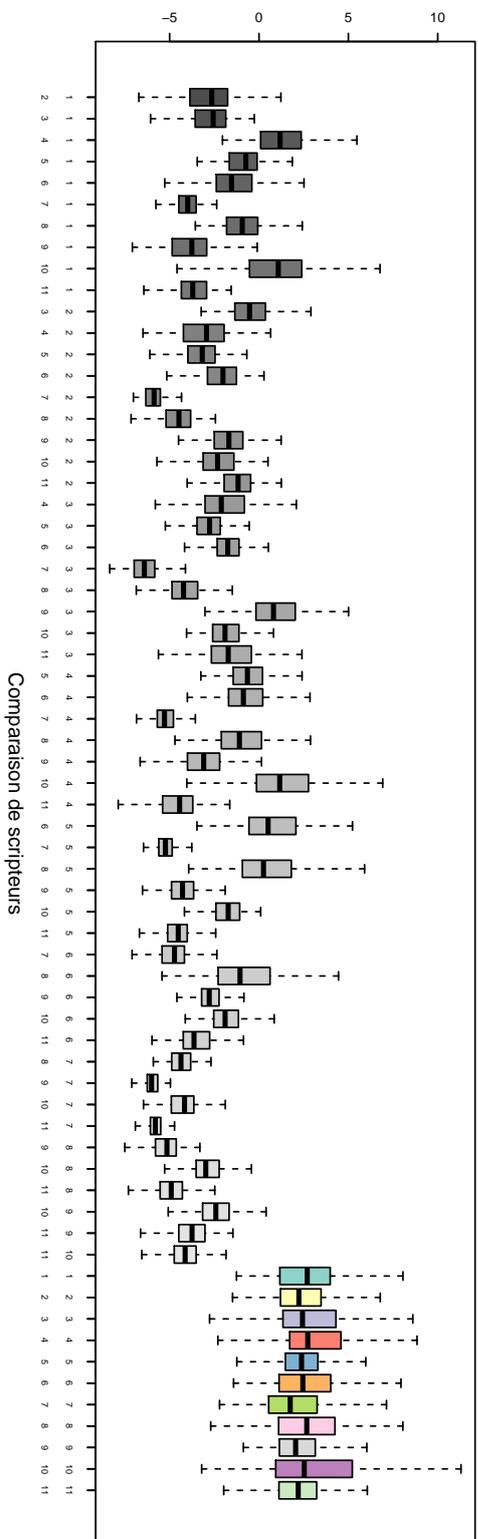
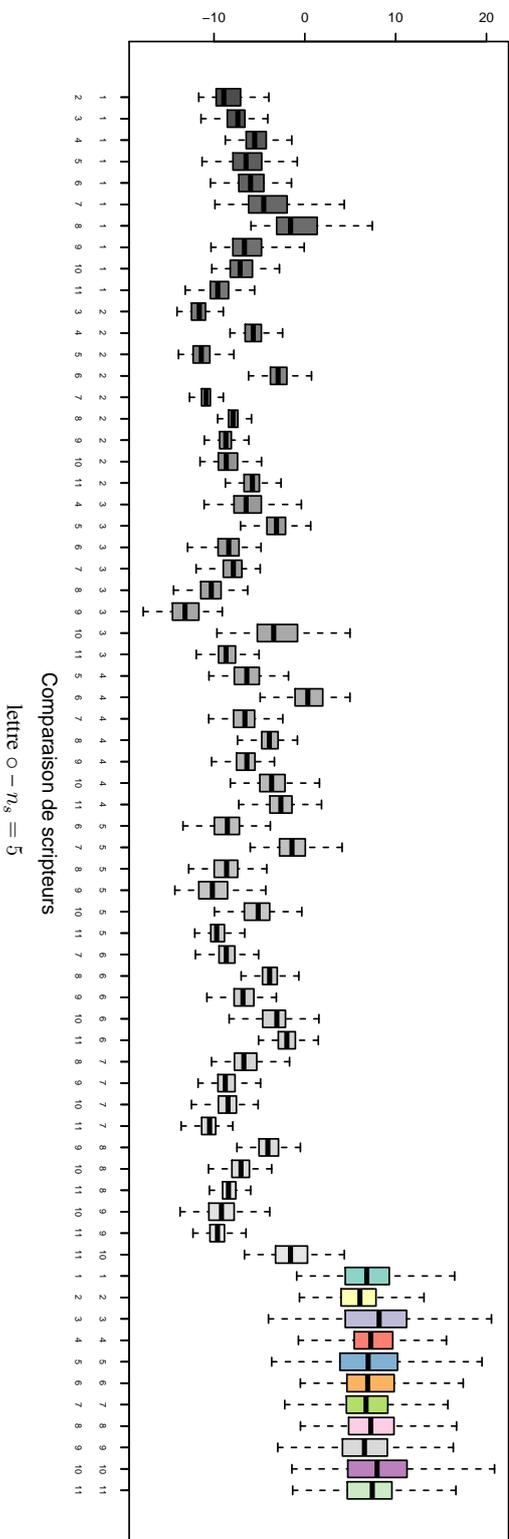


FIGURE 4.20: Illustration de la distribution des  $LRV$  pour les lettres  $h$  et  $o$  avec  $n_s = 5$ , pour chaque comparaison de scripteurs : de gauche à droite les comparaisons de  $1 vs 2$ ,  $1 vs 3$ ,  $\dots$ ,  $10 vs 11$ ,  $1 vs 1$ ,  $\dots$ ,  $11 vs 11$ . On observe d'un côté, que les moyennes de  $LRV$  lorsque  $H_d$  est vraie (*c.-à-d.* les scripteurs du matériel de *référence* et *indiciaire* sont différents), varient selon le scripteur étudié. D'un autre côté, les distributions lorsque  $H_p$  est vraie ont des moyennes similaires. Ces résultats sont donc en accord avec ceux obtenus pour les  $F_{ratio}$ , à savoir que les distributions *inter-scripteur* varient selon le scripteur étudié et les distributions *intra-scripteur* sont stables. On note, par ailleurs, le caractère individuel de chaque comparaison de scripteurs, qui est masqué lorsque les  $LRV$  sont représentés en fonction de  $H_p$  et  $H_d$  (voir Figure 4.21).

notée dans le cas de la lettre **a**, avec 5.43 pour  $n_s = 5$  et 13.83 pour  $n_s = 20$ . Ces informations traduisent le fait que certaines lettres (*p.-ex.* la lettre **a**) sont moins discriminantes que d'autres (*p.-ex.* la lettre **h**).

Ces résultats sont illustrés en détail dans la Figure 4.20, où les 66 possibilités de comparaison de scripteurs sont considérées, au lieu d'étudier les moyennes pour  $H_p$  et  $H_d$  comme dans la Figure 4.21. Ces résultats indiquent que :

1. d'un côté les moyennes de  $LRV$  lorsque  $H_d$  est vraie (*c.-à-d.* les scripteurs du matériel de *référence* et *indiciaire* sont différents), varient selon le scripteur étudié. D'un autre côté, les distributions lorsque  $H_p$  est vraie ont des moyennes similaires. Ces résultats sont donc en accord avec ceux obtenus pour les  $F_{ratio}$ , à savoir que les distributions *inter-scripteur* varient selon le scripteur étudié et les distributions *intra-scripteur* sont stables.
2. le caractère individuel de chaque comparaison de scripteurs. En effet, lorsque l'écriture de deux scripteurs est comparée (soit  $H_d$  vraie), les moyennes de  $LRV$  varient selon les scripteurs. Par exemple, pour la lettre **h**, la moyenne est de  $-0.28$  en comparant le scripteur 1 avec le scripteur 7 (*1 vs 8*), de  $-5.85$  en comparant *1 vs 6*, et de  $-9.24$  en comparant *1 vs 11*.

En conclusion, les résultats présentés dans cette section sont identiques à ceux de la section précédente, à savoir la section 4.2.2, et démontrent que :

1. la méthodologie développée permet de discriminer les scripteurs. En effet, on observe que les hypothèses  $H_p$  et  $H_d$  sont généralement vérifiées. Il s'agit pour  $H_p$ , des cas où le matériel *indiciaire* et de *référence* proviennent du même scripteur (soit  $LRV > 0$ ) et pour  $H_d$ , des cas où le matériel provient de deux scripteurs (soit  $LRV < 0$ ). Par ailleurs, il est nécessaire pour obtenir une meilleure information des variations de l'écriture d'un scripteur, de récolter un grand nombre d'échantillons. En effet, les taux d'erreur des calculs de rapport de vraisemblance diminuent quand le nombre d'échantillons  $n_s$  augmente. Par exemple pour la lettre **h** et  $n_s = 20$ , les taux de *faux négatifs* et de *faux positifs* sont de 0.1% ;
2. il existe des différences entre les caractères *normaux* et *grands* des scripteurs.

Par ailleurs les résultats montrent que :

1. les variations des taux d'erreur en fonction du(des) scripteur(s) étudié(s) sont très importantes. Ainsi il est mis en évidence que ces variations dues aux scripteurs sont masquées dans les résultats globaux ;
2. les résultats des calculs de rapport de vraisemblance *multivariés* et *univariés* sont similaires, et il apparaît que les limites concernant la normalité des  $F_{ratio}$  mentionnés précédemment (voir section 4.2.2) n'impactent pas les conclusions.

#### 4.2.4 Représentation par Analyse en Coordonnées Principales

Cette section présente les résultats des représentations des matrices de dissimilarité, issues de l'expérience 2, par la technique de PCoA. Dans les résultats des calculs des  $F_{ratio}$  et des  $RV$ , pour chaque itération  $2 \times n_s$  échantillons<sup>4</sup> ont été tirés aléatoirement

---

4. Les  $2 \times n_s$  échantillons sont composés 1) des  $n_s$  échantillons du matériel *indiciaire*, et 2) des  $n_s$  échantillons du matériel de *référence*.

## Etendue des rapports de vraisemblance

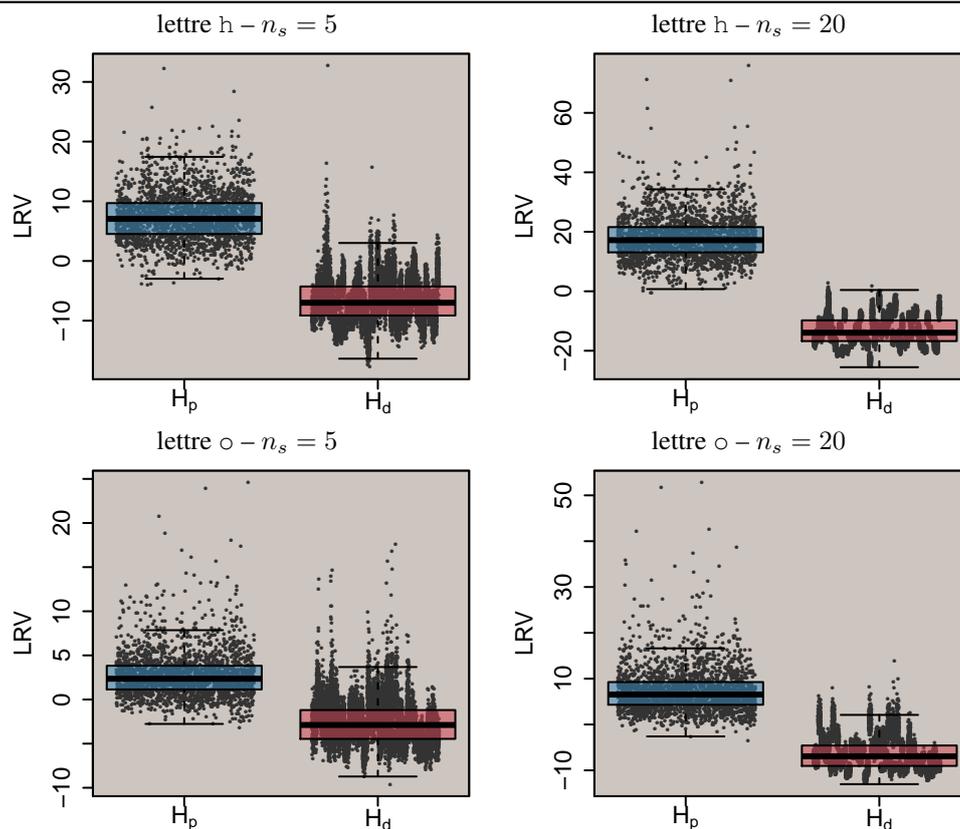


FIGURE 4.21: Illustration de la distribution des  $LRV$  pour les lettres **h** et **o**, avec  $n_s = 5, 20$ . Ces résultats indiquent que les rapports de vraisemblance forment des distributions différentes selon que l'hypothèse  $H_p$  est vraie ou non. Par exemple, pour la lettre **o** avec  $n_s = 5$ , la moyenne est de 2.78 (écart-type 2.67) pour  $H_p$  contre  $-2.65$  (écart-type 2.48) pour  $H_d$ .

des 11 scripteurs, et de ce fait seules les matrices de dissimilarité  $\mathbf{D}_{2n_s \times 2n_s}$  entre ces échantillons ont été analysées. Au contraire, la méthode de PCoA permet d'utiliser la totalité des matrices de dissimilarité (à savoir les matrices  $\mathbf{D}_{665 \times 665}$ ). En effet, la PCoA est une méthode d'ordination qui permet de représenter des échantillons sur un plan en deux dimensions, et peut ainsi représenter la totalité de la matrice de dissimilarité en un seul graphique.

Il s'agit donc de classer les échantillons de ces matrices dans les catégories illustrant les deux types de caractères (*normaux* et *grands*) et les différents scripteurs (1 à 11). Ainsi cette section a pour objectif de déterminer si :

1. les mesures de dissimilarités extraites entre les échantillons sont capables de discriminer les scripteurs (*c.-à-d.* une plus grande séparation des scripteurs dans les représentations par PCoA indique une meilleure discrimination) ;
2. il existe des différences entre caractères définis comme étant *normaux* et *grands* ;
3. les dissimilarités calculées entre les échantillons d'un scripteur varient dans un espace temporel court. En d'autres termes, s'il existe des différences entre les cinq pages  $I - V$  rédigés par les scripteurs dans l'expérience 2.

Les résultats de la Figure 4.22 montrent les représentations de matrices  $\mathbf{D}_{665 \times 665}$  selon la technique de PCoA. Dans ces représentations, les échantillons sont regroupés par leur taille (soit les caractères *normaux* et *grands*) ou par scripteur. Ainsi, ces graphiques permettent d'étudier les différences entre les échantillons issus de caractères de taille différente. Les résultats indiquent :

- qu'il existe des différences, pour un même scripteur, entre les échantillons issus de caractères de taille différente. En effet, les représentations des groupes des caractères *normaux* et *grands* ne sont pas superposés dans les figures, et indiquent que les dissimilarités entre ces échantillons ne sont pas similaires. Ces différences mettent donc en évidence que les caractères manuscrits *grands* ne sont pas similaires (*c.-à-d.* ne sont pas reproduits de manière proportionnelle) aux caractères *normaux* ;
- que les différences entre les caractères *normaux* et *grands* ne sont pas similaires selon la *fonction d'extraction* étudiée : par exemple les différences entre ces caractères pour la lettre **e** sont plus accentuées avec  $FE_1$  qu'avec  $FE_2$  ;
- que les différences entre les caractères *normaux* et *grands* ne sont pas similaires entre lettres : par exemple les différences sont plus accentuées pour la lettre **a** que la lettre **h**.

Ces résultats montrent bien qu'il existe des différences entre les caractères *normaux* et *grands* des scripteurs, comme mis en évidence par les graphiques des PCoA.

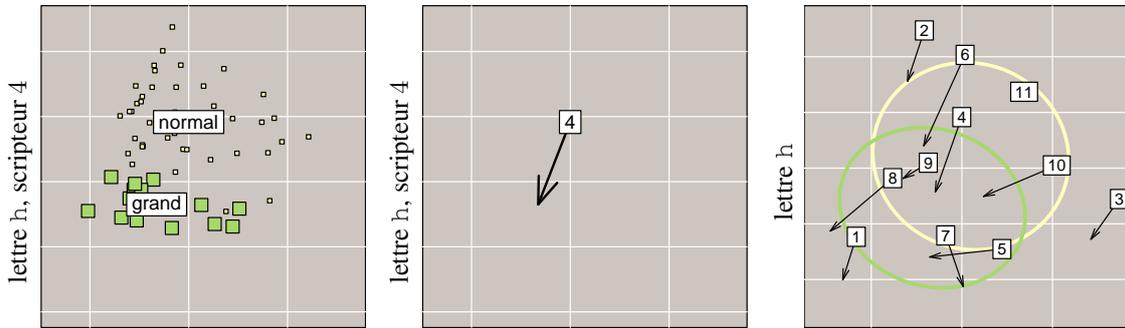
De ce fait ces résultats montrent l'avantage d'un processus de *vérification* qui permet, comme celui développé dans cette recherche, d'utiliser plusieurs outils statistiques pour évaluer les observations issues des échantillons.

Ces résultats, tout comme ceux observés dans les sections précédentes, soulignent l'importance des variations individuelles de l'écriture, et par conséquent démontrent l'intérêt d'étudier des observations issues de l'ensemble des échantillons.

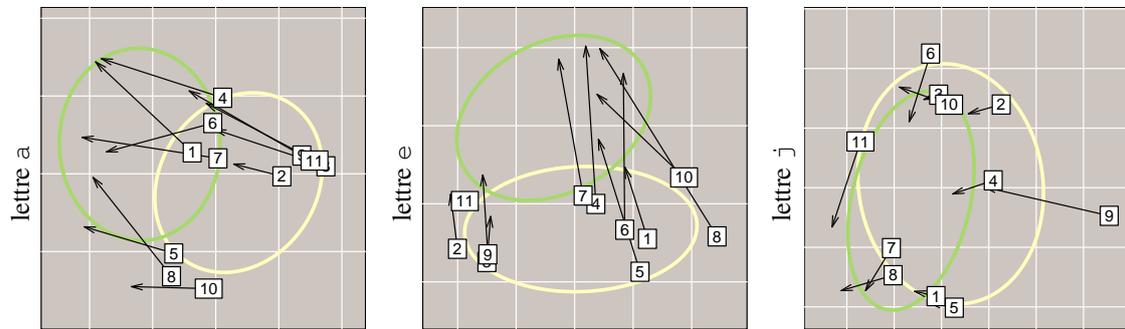
La Figure 4.23 illustre les dissimilarités au sein des cinq pages rédigées  $I - V$  par les scripteurs de l'expérience 2 (voir section 3.2). Dans ces représentations par la technique de PCoA, les échantillons de chaque scripteur sont regroupés selon les pages  $I - V$ , et permettent donc d'étudier les différences entre les échantillons issues de pages différentes, à savoir d'échantillons produits à des jours différents. Ces résultats montrent :

1. indépendamment du scripteur ou de la lettre étudiée, les groupes formant les cinq pages  $I - V$  (représentés par les ellipses dans la Figure 4.23) ne sont pas semblables. Ceci indique qu'il existe des différences entre les échantillons issus de pages différentes (*c.-à-d.* les pages  $I - V$  écrites sous une période de 15 jours). Par exemple pour le scripteur 10 et la lettre **e**, les échantillons de la page  $I$  se différencient des pages  $II$  à  $V$ . Ces observations mettent en évidence la variation d'écriture dans une période temporelle courte ;
2. les différences, notées ci-dessus, ne sont pas similaires selon :
  - le scripteur étudié : par exemple, pour le scripteur 7, les différences entre les sessions  $I - V$  sont plus prononcées avec la lettre **h** qu'avec la lettre **p**.
  - la lettre analysée : dans le cas de la lettre **e**, les différences entre les échantillons des cinq pages sont plus prononcées pour les scripteurs 1 et 10 que pour les scripteurs 4 et 11. Ces résultats soulignent que chaque scripteur possède sa propre variation d'écriture en fonction du temps.

**Représentation PCoA – caractères *normaux/grands* – fonction  $FE_1$**



**Représentation PCoA – fonction  $FE_1$**



**Représentation PCoA – fonction  $FE_2$**

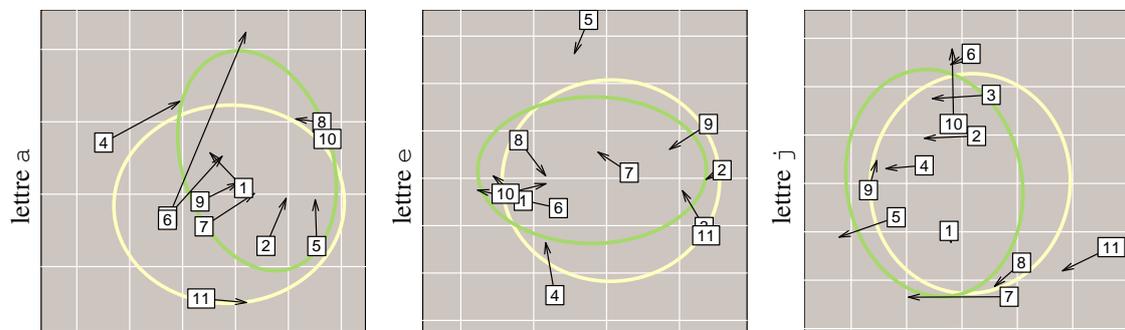


FIGURE 4.22: Représentations par la technique de PCoA de matrices de dissimilarité  $D_{665 \times 665}$  de l'expérience 2 issus de la lettre *e*. Les graphiques des deux dernières lignes montrent les différences entre le centre des caractères *normaux* (représenté par l'extrémité avec un chiffre) et le centre de caractères *grands* (représenté par l'extrémité avec une flèche :  $\rightarrow$ ). Les graphiques du haut illustrent la construction de ces graphiques avec l'exemple de la lettre *h* et le scripteur 4. Ces graphiques illustrent donc les dissimilarités calculées entre les caractères *normaux* et *grands* par les mesures  $FE_i$ . On observe dans les graphiques en bas que les groupes de ces deux types de caractères diffèrent. Ces résultats indiquent donc qu'il existe une différence entre ces deux types de caractères. Par ailleurs, on remarque que la mesure  $FE_1$  est plus sensible à la taille que la mesure  $FE_2$ .

Ces résultats soulignent que chaque scripteur possède sa propre variation d'écriture en fonction du temps.

Les résultats montrent qu'il existe des différences entre les échantillons produits à des jours différents, et de ce fait sont une indication qu'il existe une variation de l'écriture qui est imputable au temps. Par ailleurs, les deux derniers points mettent en évidence que chaque scripteur possède sa propre variation, ainsi les résultats ne permettent pas d'établir de tendance générale concernant l'influence du temps sur l'écriture de scripteurs. Cependant, ils soulignent l'importance de la variabilité au sein de l'écriture d'un scripteur.

La Figure 4.24 illustre les dissimilarités entre les 11 scripteurs de l'expérience 2, avec la technique de PCoA. Les échantillons sont issus des caractères *normaux* de l'expérience 2 et sont donc regroupés par scripteur. Ces représentations permettent d'étudier les différences entre les scripteurs, pour une lettre et une *fonction d'extraction* données, afin de déterminer si les mesures de dissimilarités extraites permettent de discriminer les scripteurs. Les résultats indiquent que :

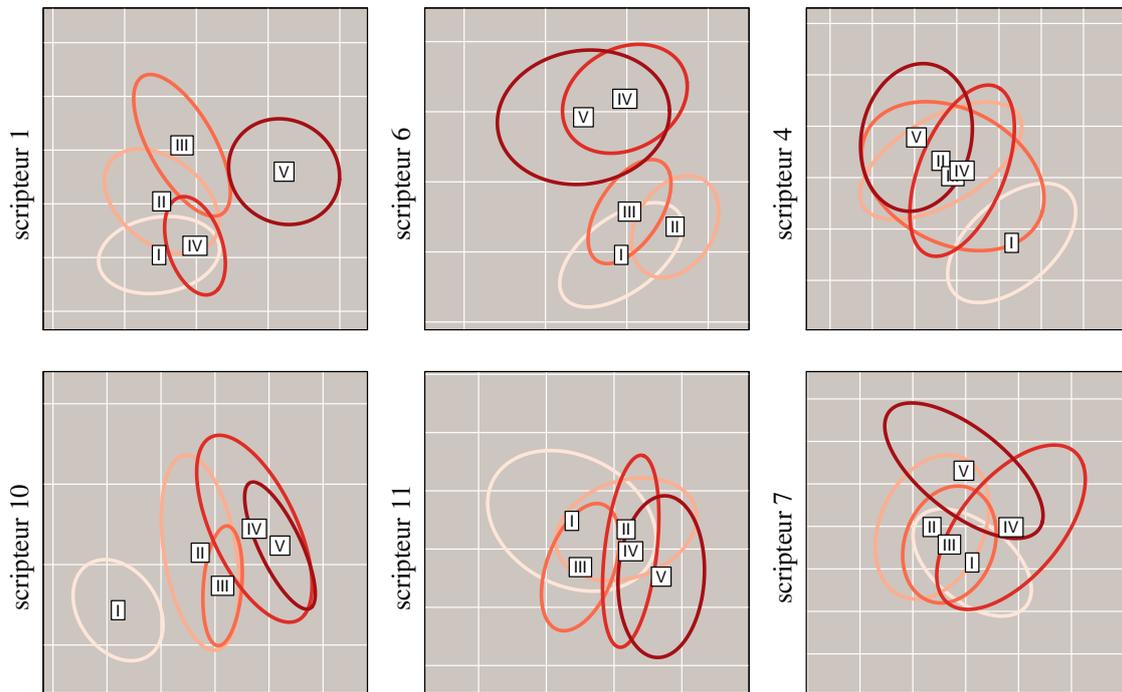
- les mesures de dissimilarités permettent de distinguer les scripteurs selon leur écriture, en effet, une plus grande séparation des scripteurs dans les représentations par PCoA indique une plus grande différence entre les dissimilarités. Par exemple dans le cas de la lettre **h**, on observe que les groupes de scripteurs sont distincts. Ainsi les matrices de dissimilarité calculées depuis les  $n_{FE}$  fonctions sélectionnées apparaissent comme des mesures permettant de discriminer l'écriture de scripteurs, et rejoignent donc les résultats des  $F_{ratio}$  et des calculs de rapport de vraisemblance ;
- le potentiel de discrimination entre scripteurs varie en fonction de la lettre étudiée. Par exemple, les groupes de scripteurs se discriminent de préférence avec la lettre **h** plutôt qu'avec la lettre **e**. Ceci indique que certaines lettres sont plus discriminantes que d'autres. Ainsi, lors d'une expertise en écriture manuscrite, l'expert trouverait peut être plus de caractéristiques discriminantes dans les caractères de la lettre **h** que dans ceux de la lettre **e** ;
- pour une même lettre, les scripteurs ne sont pas regroupés de manière exactement similaire selon la *fonction d'extraction* étudiée. Par exemple, dans le cas de la lettre **e** (voir Figure 4.24) :
  - les scripteurs 11 et 3 sont regroupés pour la *fonction d'extraction*  $FE_2$  ;
  - cependant, les scripteurs 1 et 10 semblent former un groupe à part pour  $FE_2$ , alors qu'ils appartiennent à un groupe plus vaste avec la *fonction d'extraction*  $FE_1$ .

Ces différences indiquent principalement que chaque *fonction d'extraction* mesure une quantité différente. Ces résultats soulignent donc l'intérêt de combiner les  $n_{FE} = 2$  *fonctions d'extraction* afin d'augmenter le potentiel de discrimination lors de l'étape d'évaluation, soit le calcul de rapport de vraisemblance.

En conclusion, les résultats présentés dans cette section soulignent :

1. que les mesures de dissimilarités extraites entre les échantillons sont capables de discriminer les scripteurs (*c.-à-d.* une plus grande séparation des distributions indique une meilleure discrimination) ;
2. qu'il existe des différences entre les caractères *normaux* et *grands* des scripteurs ;

Représentation PCoA – lettre e – fonction  $FE_1$



Représentation – scripteur 7 – fonction  $FE_1$

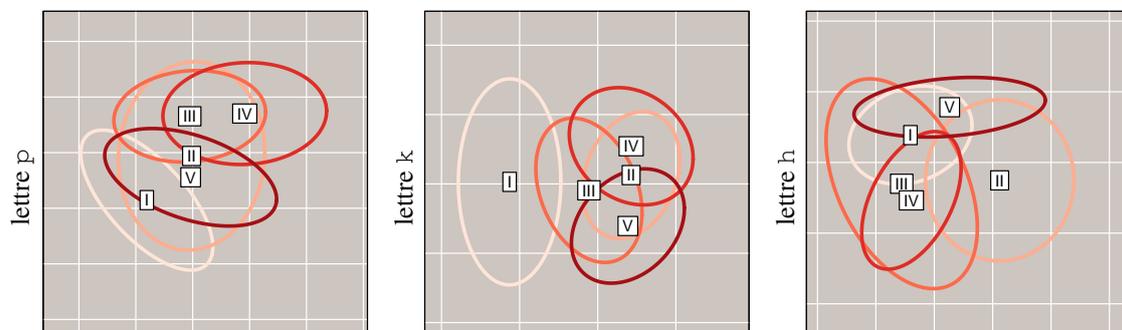
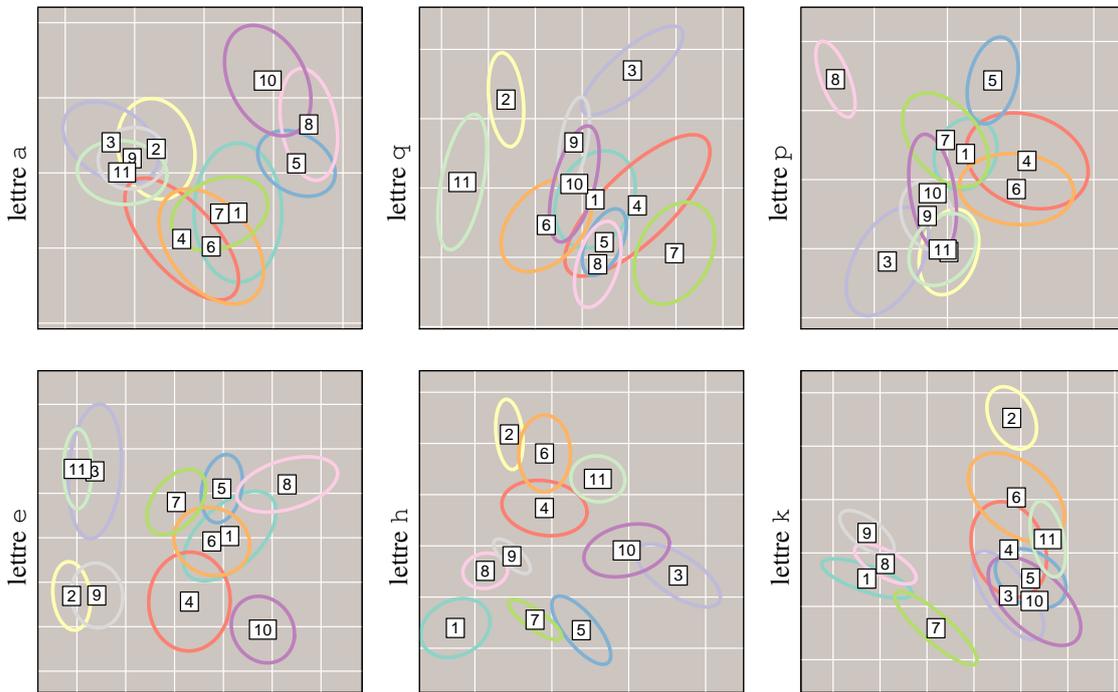


FIGURE 4.23: Représentation par PCoA des matrices de dissimilarité de la lettre e des scribes 1, 4, 6, 7, 10 et 11, et des lettres p, h et k du scribe 7. Les caractères, issus des caractères *normaux* de l'expérience 2, sont groupés selon les cinq pages (I – V) rédigées par les scribes. Ces graphiques montrent qu'il existe une différence entre les dissimilarités des échantillons issus de pages différentes. Par exemple pour le scribe 6 les échantillons peuvent être classés en deux groupes : 1) les pages I – III, 2) et les pages IV – V. Par ailleurs, on observe que ces variations diffèrent entre scribes, par exemple les différences entre les échantillons des cinq pages sont plus prononcées pour les scribes 1 et 10 que pour les scribes 4 et 11. Par ailleurs, les graphiques du scribe 7 mettent en évidence que ces différences varient en fonction de la lettre étudiée. Ces résultats soulignent l'existence d'une variation temporelle courte de l'écriture de scribes.

Représentation PCoA – fonction  $FE_1$



Représentation PCoA – fonction  $FE_2$

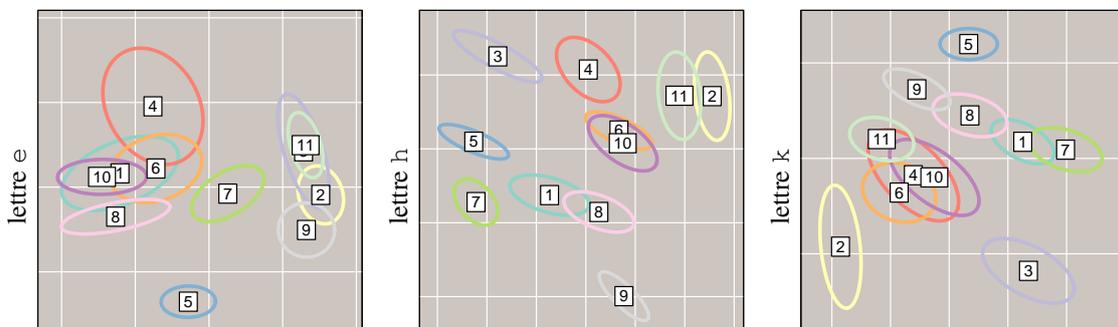


FIGURE 4.24: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* de l'expérience 2, groupés par scripteurs (les lettres a, q, e, h et k, pour la fonction d'extraction  $FE_1$ , et les lettres e, h et k, pour la fonction d'extraction  $FE_2$ ). Il est mis en évidence dans ces graphiques que les mesures permettent de discriminer les scripteurs. En effet, une plus grande séparation des scripteurs dans la représentation par PCoA indique une meilleure discrimination de scripteur. Cependant, certaines lettres sont moins discriminantes que d'autres, par exemple les différences entre scripteurs sont plus distinctes pour les lettres h et k que pour les lettres a et e. On remarque que les mesures de dissimilarités ne sont pas exactement similaires selon la fonction d'extraction étudiée (les scripteurs ne sont pas regroupés de manière similaire selon que  $FE_1$  ou  $FE_2$  soit utilisée).

3. qu'il existe une variation temporelle courte (période  $\leq 15$  jours) de l'écriture des scripteurs.

## 4.3 Expérience 3

L'expérience 3 vise à étudier d'avantage l'écriture du scripteur 7 en ajoutant un jeu de données supplémentaire à l'expérience 2. Ce scripteur a rédigé trois textes de menace *i – iii*, un an après la première prise d'échantillon des cinq pages *I – V* de l'expérience 2. Ainsi sur la base des comparaisons entre les échantillons provenant de ces deux sessions, cette expérience présente deux objectifs (pour une description détaillée de l'expérience 3 voir la section 3.3) :

1. valider le processus de *vérification* en étudiant les résultats du calcul de rapport de vraisemblance ;
2. étudier s'il existe des variations de l'écriture du scripteur 7 au cours d'un espace temporel très court (différences entre textes de menace) et long (différence entre les textes de menace et les cinq pages de l'expérience 2).

Comme indiqué dans la description de l'expérience 3, les résultats sont présentés en trois sous-sections : largeur, hauteur et rapport  $W/H$  (4.3.1) ; les résultats du calcul des rapports de vraisemblance (4.3.2) ; et les résultats des représentations des matrices de dissimilarité par la technique de PCoA (4.3.3).

### 4.3.1 La largeur, la hauteur et le rapport $W/H$

Les dimensions (à savoir, la largeur  $W$  et la hauteur  $H$ ) et le rapport  $W/H$  des caractères du scripteur 7 sont étudiés dans cette section. L'étude de ces variables sur l'ensemble des sessions (*c.-à-d.* les cinq pages de l'expérience 2 et les trois textes de menace de l'expérience 3) permet d'observer s'il existe des variations de l'écriture en fonction d'un espace temporel très court et long. En effet les trois textes de menace ont été rédigés sur trois jours successifs (correspondant à l'espace temporel très court), un an après la prise d'échantillon de l'expérience 2 (correspondant à l'espace temporel long).

Les variables  $W$ ,  $H$  et  $W/H$  des caractères du scripteur 7 pour les caractères **a**, **o** et **e** sont représentées pour les huit sessions (*c.-à-d.* les cinq pages de l'expérience 2 et les trois textes de menace de l'expérience 3 pris individuellement) dans la Figure 4.25. Par ailleurs, la Figure 4.26 illustre les largeurs  $W$  et les hauteurs  $H$ , qui sont regroupées en fonction des deux jeux de données, soit l'ensemble de pages provenant de l'expérience 2 et celles provenant de l'expérience 3. Les observations de ces graphiques montrent que :

- de manière générale, les caractères du scripteur 7 sont moins hauts et larges dans les textes de menace que dans les cinq premières sessions de l'expérience 2. Toutefois, cette remarque n'est pas valable pour tous les caractères : pour la lettre **q** les largeurs  $W$  sont similaires dans l'ensemble des huit sessions. Dans le cas de la lettre **e**, la largeur moyenne  $W$  des caractères de la lettre **e** sont en moyenne plus élevées dans l'expérience 3 que dans l'expérience 2. Ces résultats indiquent donc l'existence d'une variation de l'écriture suite à un espace temporel long ;
- les hauteurs  $H$  et largeurs  $W$  de caractères ne sont pas homogènes entre les trois textes de menace *i – iii*. Par exemple, d'un côté pour la lettre **o**, la largeur  $W$  des

caractères des textes de menace est similaire entre ces trois sessions. D'un autre côté, pour la lettre *e*, la largeur  $W$  augmente successivement dans les trois textes de menace, inversement les hauteurs  $H$  diminuent successivement. Ceci souligne donc qu'il existe une différence entre les caractères sur une période temporelle très courte.

En conclusion les résultats, similaires à ceux précédemment obtenus entre les cinq pages  $I - V$  (voir section 4.2.1), indiquent que :

1. les variables  $W$ ,  $H$  et  $W/H$  du scripteur 7 de l'expérience 3 varient entre les textes de menace et entre les échantillons de l'expérience 2 et 3, et par conséquent qu'il existe une variation temporelle de l'écriture à la fois sur une période très courte et sur une période longue ;
2. les variations de ces variables ( $W$ ,  $H$  et  $W/H$ ) ne sont pas similaires selon la lettre étudiée. De ce fait, sur la base de ces observations, aucune tendance générale ne peut être donnée concernant l'évolution de l'écriture dans un espace temporel.

### 4.3.2 Les résultats du calcul des rapports de vraisemblance

Cette section présente les résultats du calcul des rapports de vraisemblance de l'expérience 3. Dans le cadre des *scénarios* fictifs de cette recherche, le  $RV$  permet d'évaluer les dissimilarités calculées entre des caractères manuscrits provenant du matériel *indiciaire* et de *référence* au regard des deux hypothèses d'intérêt  $H_p$  et  $H_d$  afin de soutenir l'une d'elles. C'est-à-dire de soutenir soit l'hypothèse  $H_p$  : le même scripteur a écrit les documents *indiciaire* et de *référence* ; soit l'hypothèse  $H_d$  : deux scripteurs différents ont écrit les documents *indiciaire* et de *référence* ( $H_d$ ). Cette expérience est divisée en deux sous-expériences :

- 1) l'expérience 3 – *a* (soit les *scénarios* 7 à 12), fait appel aux situations où le matériel *indiciaire* est issu de l'un des trois textes de menace (*c.-à-d.* les sessions *i - iii* du scripteur 7) et le matériel de *référence* est issu des échantillons de l'un des scripteurs de l'expérience 2 (*c.-à-d.* les scripteurs 1 à 11) ;
- 2) l'expérience 3 – *b* (soit les *scénarios* 13 à 16), fait appel aux situations où le matériel *indiciaire* est issu d'un mélange des trois textes de menace et le matériel de *référence* est issu des échantillons de l'un des scripteurs de l'expérience 2.

Les objectifs des résultats présentés dans cette section ont pour but de déterminer :

1. si les rapports de vraisemblance calculés permettent de discriminer l'écriture de scripteurs ;
2. s'il existe des variations de l'écriture en fonction d'un espace temporel très court, en comparant les résultats de l'expérience 3 – *a* et 3 – *b* ;
3. s'il existe des variations de l'écriture en fonction d'un espace temporel long, en comparant les résultats de l'expérience 3 avec ceux de l'expérience 2.

L'expérience 3 – *a* (soit les *scénarios* 7 à 12) fait appel aux situations où le matériel de *référence* est issu des échantillons de l'un des 11 scripteurs de l'expérience 2, et le matériel *indiciaire* est formé par des échantillons d'un des trois textes de menace *i - iii*. Ainsi les résultats des calculs de rapport de vraisemblance de cette expérience permettent de

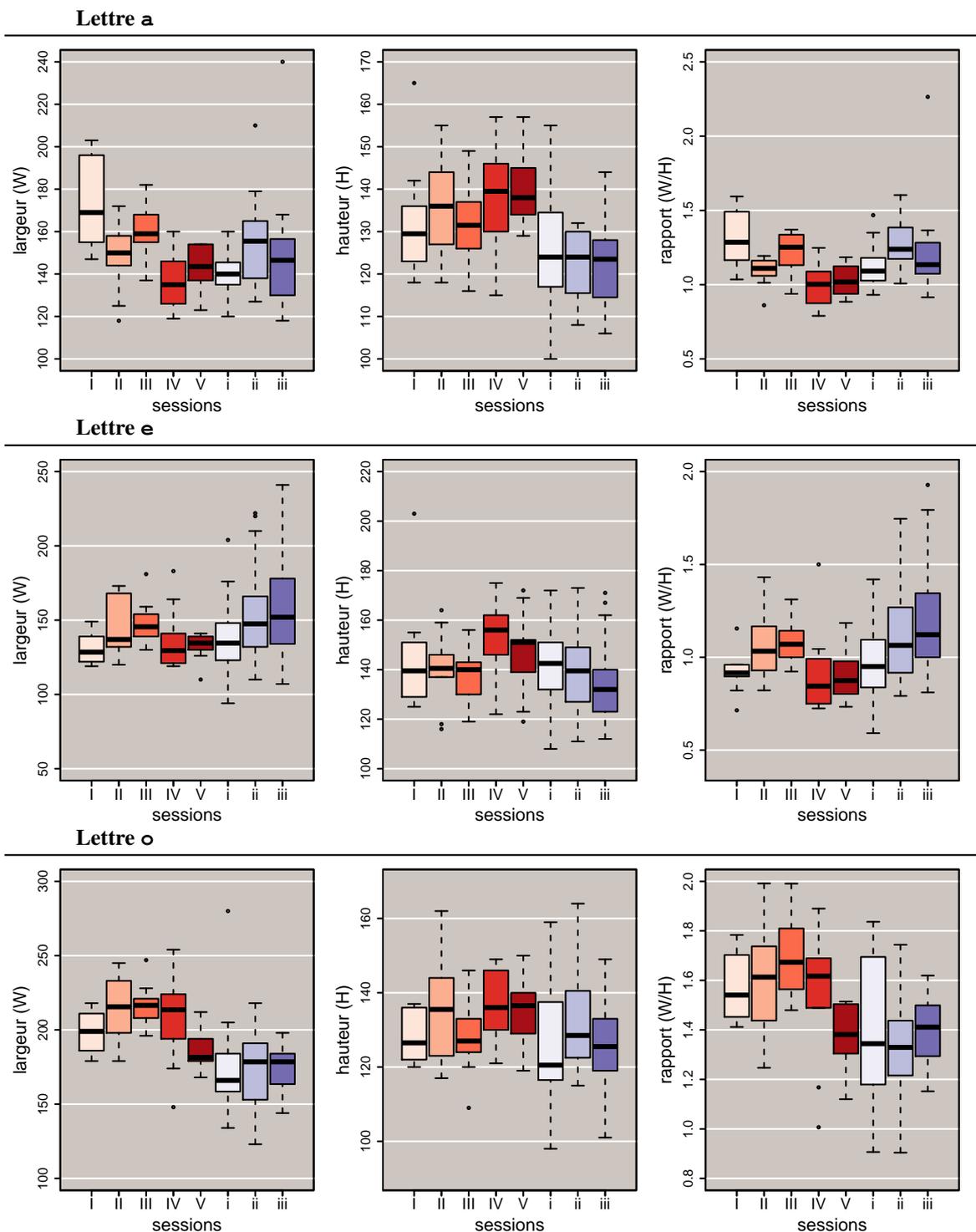


FIGURE 4.25: Boîte à moustache de la hauteur, largeur et du rapport  $W/H$  des caractères a, e et o du scripteur 7, pour les cinq pages ( $I - V$ ) et les trois textes de menace ( $i - iii$ ). Par exemple, on observe pour la lettre a que la hauteur  $H$  des trois textes de menace est inférieure à celle des cinq pages  $I - V$ . Dans le case de la lettre o, c'est avec la largeur  $W$  que l'on remarque des différences entre les pages  $I - V$  et  $i - iii$ . Par ailleurs, on observe que la hauteur  $H$  des caractères des textes de menace est en moyenne inférieure et plus variable que celle des pages  $I - V$ . Dans tous les cas, on note des dissimilarités entre les mesures de chaque session. Ces résultats mettent donc en évidence les variations temporelles de l'écriture manuscrite du scripteur 7.

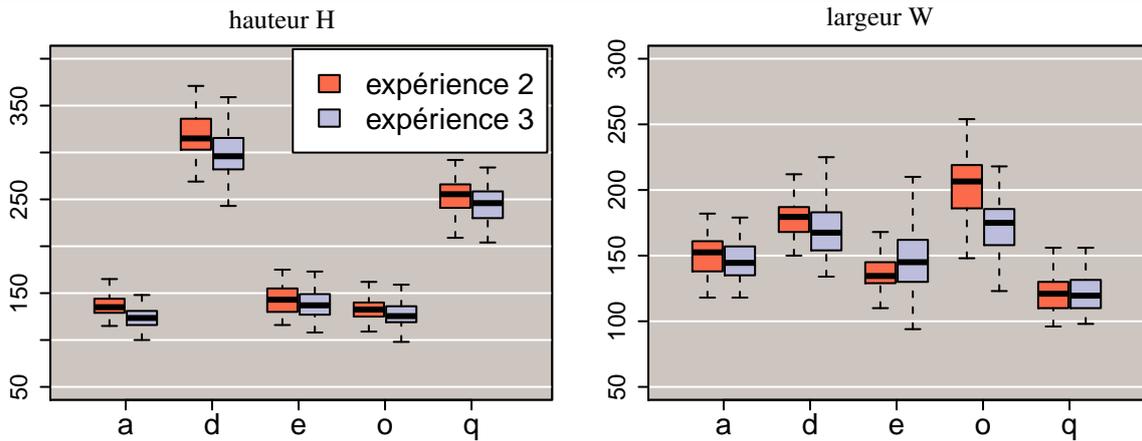


FIGURE 4.26: Boîte à moustache de la hauteur et largeur des caractères a du scripteur 7 issus de l'expérience 2 et 3 (les deux prises d'échantillons différent d'une année). On observe qu'en moyenne la hauteur et la largeur des caractères sont plus faibles dans les textes de menace (expérience 3), hormis pour la largeur  $W$  des caractères de la lettre e qui est en moyenne plus élevée dans l'expérience 2. Ceci met donc en évidence une variation de l'écriture de ce scripteur suite à un espace temporel long. Cependant, comme on peut le constater pour la lettre e, aucune tendance générale ne peut être donnée.

déterminer s'il existe une variation de l'écriture du scripteur 7 dans un espace temporel. Ces résultats sont présentés dans la Figure 4.27, et plus en détails dans la Table 4.2, selon les cinq lettres, du nombre d'échantillons, des scripteurs de référence et des textes de menace. Ils soulignent que :

- les taux de *faux positifs* et *faux négatifs* sont différents de ceux de l'expérience 2, dans laquelle étaient comparés des échantillons rédigés sous une période de 15 jours. Par exemple, d'un côté les taux de *faux positifs* baissent progressivement lorsque le nombre d'échantillons augmente, indépendamment de la lettre (a, d, ...), ou le texte de menace (*i - iii*) étudié, tout comme dans l'expérience 2. De l'autre côté, les taux de *faux négatifs* ( $\sim 0 - 90\%$ ) n'ont pas de tendance générale, contrairement à l'expérience 2. De plus, les taux de *faux négatifs* sont en moyenne supérieurs de 16.9 points en comparaison aux taux de l'expérience 2, contre une baisse moyenne de 2.5 points pour les taux de *faux positifs*. Ces résultats mettent en évidence une variation de l'écriture suite à un espace temporel long ;
- les résultats indiquent qu'il n'y a pas de tendance générale entre les textes de menace. C'est-à-dire que les résultats ne sont pas similaires selon le texte de menace étudié, par exemple :
  - pour la lettre e, les taux de *faux positifs* les plus faibles sont obtenus avec le texte de menace *ii*, alors que les taux de *faux négatifs* les plus faibles sont obtenus avec le texte de menace *i*.
  - pour la lettre d, les taux de *faux négatifs* et *faux positifs* les plus faibles sont obtenus avec le texte de menace *iii*, puis *ii* et finalement *i*.

Ainsi ces résultats mettent en avant l'existence de variation de l'écriture suite à un espace temporel très court ( $\leq 3$  jours) ;

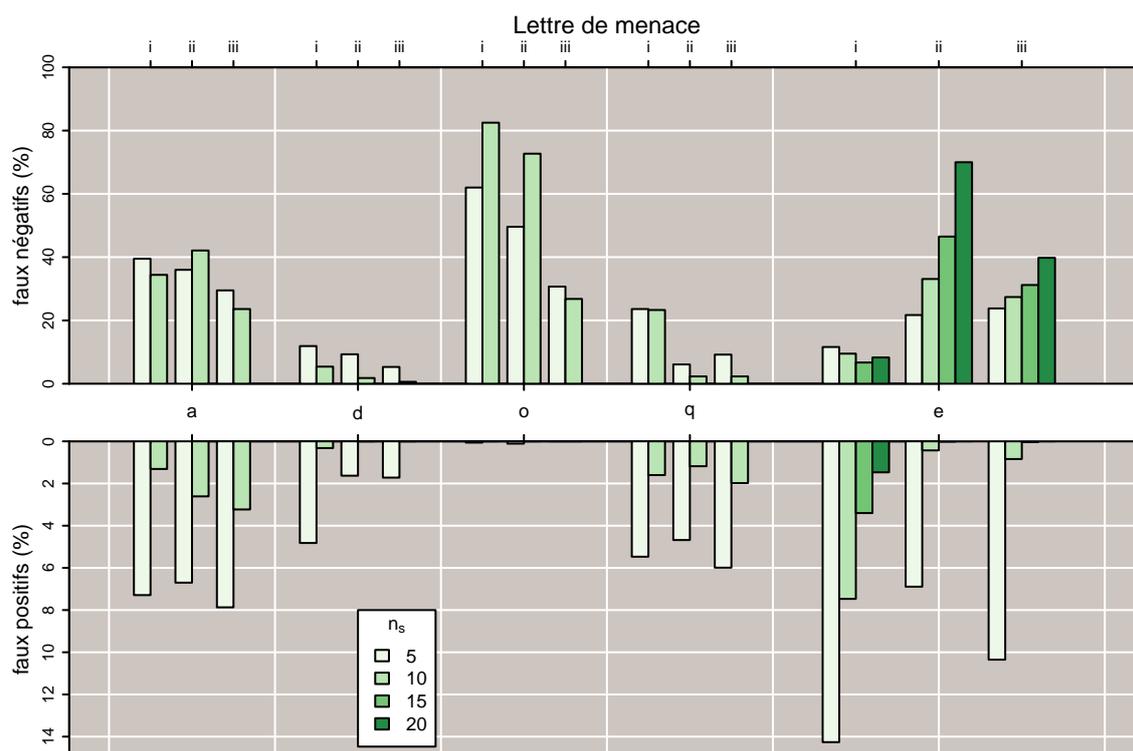


FIGURE 4.27: Illustration des taux moyens de *faux positifs* et *faux négatifs* obtenus des calculs de rapport de vraisemblance de l'expérience 3 – a (scénarios 7 – 12). Soit lorsque le matériel *indiciaire* est issu de l'un des trois textes de menace (scripteur 7) et que le matériel de *référence* est issu des échantillons de l'un des 11 scripteurs de l'expérience 2. On observe que le taux de *faux positifs* diminue lorsque le nombre d'échantillons  $n_s$  augmente, cependant ce n'est pas le cas pour le taux de *faux négatifs*. Par exemple, le taux de *faux négatifs* augmente avec  $n_s$  pour la lettre o alors qu'il diminue pour la lettre d. Ces graphiques mettent en évidence l'écart des résultats obtenus entre les différents textes de menace (i – iii) et les différentes lettres étudiées (a, d, o, q et e).

- les taux de *faux positifs* ( $\sim 0 - 12\%$ ) sont inférieurs aux taux de *faux négatifs* ( $\sim 0 - 90\%$ ), tout comme dans l'expérience 2 ;
- les taux de *faux positifs* varient en fonction du scripteur de *référence* étudié. Par exemple, pour la lettre q, les scripteurs 1, 4, 5, 6 et 8 se distinguent des autres scripteurs par les valeurs de leurs taux de *faux positifs*. En effet ces scripteurs sont les seuls à avoir des *faux positifs* pour cette lettre. Par ailleurs, pour la lettre o, les scripteurs 6 et 8 sont les seuls scripteurs à avoir des *faux positifs* dans le texte de menace i. D'un autre côté, le scripteur 10 n'a pas un seul *faux positifs* pour le texte de menace ii. Ces résultats soulignent l'importance des variations dues aux scripteurs étudiés ;
- Ces scénarios peuvent être visualisés comme de l'*identification* de scripteur, l'un des 11 scripteurs de *référence* étant le scripteur des textes de menace (scripteur 7). Le scripteur 7 est le seul scripteur à avoir des  $LRV = \log(RV)$  positifs pour toutes les lettres, nombre d'échantillons et textes de menace. Cela signifie que les

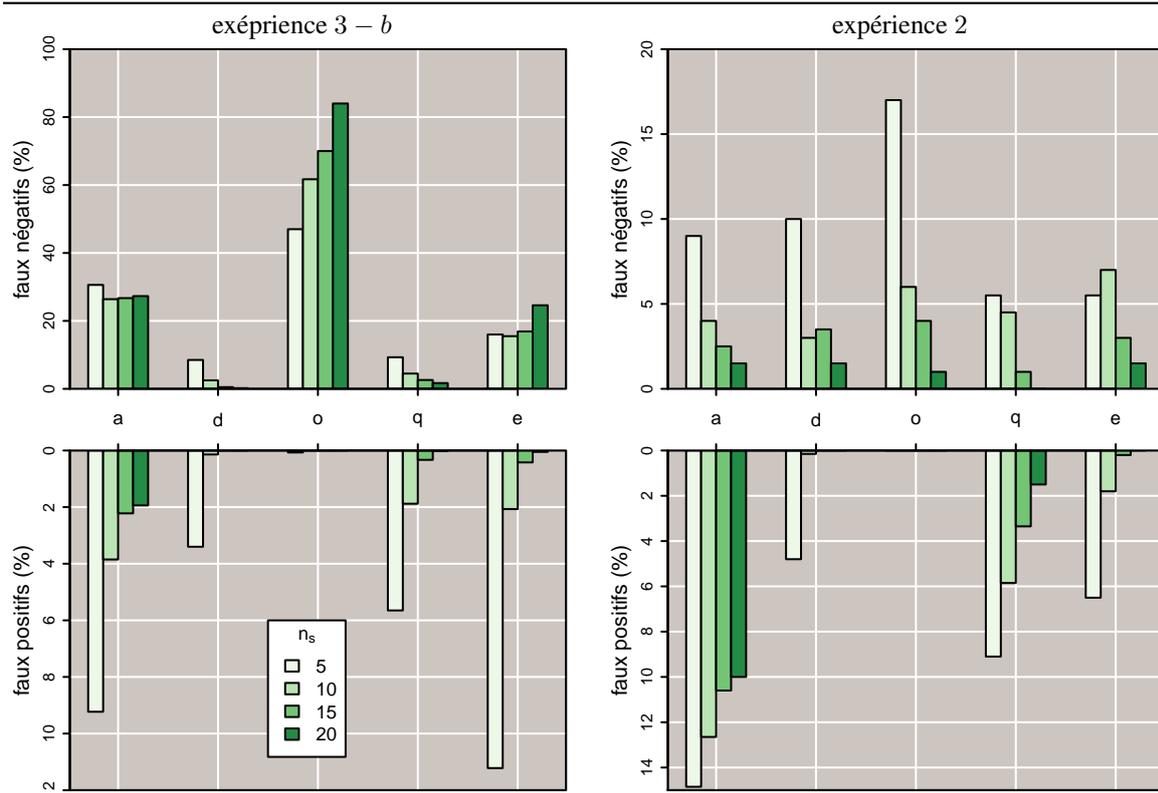


FIGURE 4.28: Illustration des taux moyens de *faux positifs* et *faux négatifs* pour l'expérience 3 – b (à gauche), et les résultats équivalents de l'expérience 2 pour comparaison (à droite). Ces illustrations mettent en évidence que les taux d'erreur ne sont pas similaires entre ces deux expériences. Par exemple les taux *faux négatifs* sont plus faibles lorsque les échantillons comparés ont été écrits dans un espace temporel court (expérience 2), que dans un espace temporel long (expérience 3). Ceci suggère qu'il existe une variation de l'écriture due à un espace temporel long.

valeurs montrent une association positive entre le matériel de *référence* et le matériel *indiciaire*. Dans ce contexte, les résultats indiquent correctement que le scripteur des textes de menace est le scripteur 7. Ces résultats permettent ainsi de contribuer à la validation du processus développé dans cette recherche.

Les résultats des calculs de rapport de vraisemblance de l'expérience 3 – b (soit les scénarios 13 – 16) sont présentés dans la Figure 4.28, et plus en détails dans la Table 4.3. Ce sont les résultats des comparaisons avec des échantillons de l'un des 11 scripteurs de l'expérience 2 en tant que matériel de *référence*, et avec le matériel *indiciaire* composé d'échantillons issus de la combinaisons des trois textes de menace. Ces résultats indiquent que :

- les taux de *faux positifs* et *faux négatifs* sont différents de ceux de l'expérience 2, dans laquelle étaient comparés des échantillons rédigés sous une période de 15 jours. Par exemple les taux de *faux positifs* baissent progressivement lorsque le nombre d'échantillons  $n_s$  augmente, indépendamment de la lettre étudiée (a, d, ...), alors que les taux de *faux négatifs* n'ont pas de tendance générale en fonction de  $n_s$ . Ces

| Lettre     | a                    |     | d   |     | o   |     | q   |     | e   |     |     |     |
|------------|----------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $n_s$      | 5                    | 10  | 5   | 10  | 5   | 10  | 5   | 10  | 5   | 10  | 15  | 20  |
| Scripteurs | lettre de menace $i$ |     |     |     |     |     |     |     |     |     |     |     |
| 1          | 35                   | 0   | 88  | 1   | 0   | 0   | 206 | 26  | 513 | 401 | 276 | 144 |
| 2          | 83                   | 1   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| 3          | 189                  | 8   | 261 | 31  | 0   | 0   | 0   | 0   | 18  | 0   | 0   | 0   |
| 4          | 18                   | 0   | 0   | 0   | 0   | 0   | 294 | 134 | 375 | 179 | 37  | 3   |
| 5          | 0                    | 0   | 101 | 0   | 0   | 0   | 33  | 0   | 149 | 33  | 1   | 0   |
| 6          | 318                  | 122 | 0   | 0   | 1   | 0   | 6   | 0   | 304 | 134 | 26  | 0   |
| 7          | 605                  | 656 | 881 | 946 | 380 | 175 | 764 | 767 | 884 | 905 | 933 | 917 |
| 8          | 0                    | 0   | 32  | 0   | 5   | 0   | 8   | 0   | 55  | 0   | 0   | 0   |
| 9          | 12                   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 3   | 0   | 0   | 0   |
| 10         | 2                    | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 9   | 0   | 0   | 0   |
| 11         | 72                   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |

|            |                       |     |     |     |     |     |     |     |     |     |     |     |
|------------|-----------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Scripteurs | lettre de menace $ii$ |     |     |     |     |     |     |     |     |     |     |     |
| 1          | 3                     | 0   | 18  | 0   | 0   | 0   | 77  | 0   | 190 | 20  | 0   | 0   |
| 2          | 7                     | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 5   | 0   | 0   | 0   |
| 3          | 77                    | 2   | 66  | 0   | 0   | 0   | 0   | 0   | 200 | 18  | 1   | 0   |
| 4          | 24                    | 0   | 0   | 0   | 0   | 0   | 361 | 118 | 120 | 2   | 0   | 0   |
| 5          | 0                     | 0   | 78  | 0   | 0   | 0   | 4   | 0   | 7   | 0   | 0   | 0   |
| 6          | 464                   | 257 | 0   | 0   | 0   | 0   | 2   | 0   | 96  | 1   | 0   | 0   |
| 7          | 640                   | 579 | 907 | 982 | 504 | 273 | 939 | 977 | 783 | 669 | 535 | 300 |
| 8          | 0                     | 0   | 1   | 0   | 11  | 0   | 24  | 0   | 1   | 0   | 0   | 0   |
| 9          | 0                     | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 2   | 0   | 0   | 0   |
| 10         | 0                     | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| 11         | 95                    | 2   | 0   | 0   | 0   | 0   | 0   | 0   | 68  | 2   | 0   | 0   |

|            |                        |     |     |     |     |     |     |     |     |     |     |     |
|------------|------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Scripteurs | lettre de menace $iii$ |     |     |     |     |     |     |     |     |     |     |     |
| 1          | 4                      | 0   | 13  | 0   | 0   | 0   | 139 | 4   | 202 | 24  | 2   | 0   |
| 2          | 0                      | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   |
| 3          | 75                     | 0   | 96  | 0   | 0   | 0   | 0   | 0   | 334 | 48  | 1   | 0   |
| 4          | 189                    | 6   | 0   | 0   | 0   | 0   | 363 | 193 | 159 | 1   | 0   | 0   |
| 5          | 87                     | 0   | 63  | 0   | 0   | 0   | 57  | 1   | 25  | 0   | 0   | 0   |
| 6          | 392                    | 317 | 0   | 0   | 0   | 0   | 4   | 0   | 176 | 8   | 0   | 0   |
| 7          | 705                    | 764 | 947 | 994 | 693 | 732 | 908 | 977 | 762 | 726 | 688 | 602 |
| 8          | 14                     | 0   | 0   | 0   | 0   | 0   | 36  | 0   | 28  | 1   | 0   | 0   |
| 9          | 0                      | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 1   | 0   | 0   | 0   |
| 10         | 0                      | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 23  | 0   | 0   | 0   |
| 11         | 26                     | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 86  | 2   | 0   | 0   |

TABLE 4.2: Résultats des calculs de rapport de vraisemblance de l'expérience 3 –  $a$ , soit des situations où le matériel *indiciaire* est issu de l'un des trois textes de menace ( $c$ -à- $d$ . les sessions  $i$  –  $iii$  du scripteur 7) et le matériel de *référence* est issu des échantillons de l'un des scripteurs de l'expérience 2 ( $c$ -à- $d$ . les scripteurs 1 à 11). Les chiffres indiqués représentent le nombre d'itérations, sur 1000, où les observations soutiennent l'hypothèse  $H_p$  : le matériel *indiciaire* et de *référence* proviennent du même scripteur. Les résultats indiquent que le scripteur des textes de menace est le scripteur 7. Par ailleurs les variations des résultats entre les trois textes de menace mettent en évidence la variation temporelle de l'écriture. Par exemple le taux de *faux négatifs* du scripteur 7 est, avec la lettre  $e$  et  $n_s = 20$ , de 8.3% pour le texte  $i$  (908 rapports de vraisemblance corrects sur 1000) contre 70% pour le texte  $ii$ .

| $n_s$      | 5        | 10  | 15  | 20  | 5        | 10  | 15  | 20  |
|------------|----------|-----|-----|-----|----------|-----|-----|-----|
| Scripteurs | lettre a |     |     |     | lettre d |     |     |     |
| 1          | 25       | 0   | 0   | 0   | 36       | 1   | 0   | 0   |
| 2          | 22       | 0   | 0   | 0   | 0        | 0   | 0   | 0   |
| 3          | 171      | 27  | 0   | 0   | 164      | 7   | 0   | 0   |
| 4          | 114      | 36  | 0   | 0   | 0        | 0   | 0   | 0   |
| 5          | 16       | 0   | 0   | 0   | 125      | 6   | 0   | 0   |
| 6          | 444      | 311 | 222 | 194 | 0        | 0   | 0   | 0   |
| 7          | 694      | 736 | 733 | 727 | 915      | 975 | 995 | 998 |
| 8          | 3        | 0   | 0   | 0   | 15       | 0   | 0   | 0   |
| 9          | 12       | 0   | 0   | 0   | 0        | 0   | 0   | 0   |
| 10         | 0        | 0   | 0   | 0   | 0        | 0   | 0   | 0   |
| 11         | 116      | 11  | 0   | 0   | 0        | 0   | 0   | 0   |

|            |          |     |     |     |          |     |     |     |
|------------|----------|-----|-----|-----|----------|-----|-----|-----|
| Scripteurs | lettre o |     |     |     | lettre q |     |     |     |
| 1          | 1        | 0   | 0   | 0   | 162      | 13  | 0   | 0   |
| 2          | 0        | 0   | 0   | 0   | 0        | 0   | 0   | 0   |
| 3          | 0        | 0   | 0   | 0   | 0        | 0   | 0   | 0   |
| 4          | 0        | 0   | 0   | 0   | 332      | 175 | 33  | 1   |
| 5          | 0        | 0   | 0   | 0   | 42       | 0   | 0   | 0   |
| 6          | 0        | 0   | 0   | 0   | 7        | 0   | 0   | 0   |
| 7          | 530      | 383 | 300 | 160 | 907      | 955 | 974 | 983 |
| 8          | 6        | 0   | 0   | 0   | 22       | 0   | 0   | 0   |
| 9          | 0        | 0   | 0   | 0   | 0        | 0   | 0   | 0   |
| 10         | 0        | 0   | 0   | 0   | 0        | 0   | 0   | 0   |
| 11         | 0        | 0   | 0   | 0   | 0        | 0   | 0   | 0   |

|            |          |     |     |     |
|------------|----------|-----|-----|-----|
| Scripteurs | lettre e |     |     |     |
| 1          | 308      | 115 | 34  | 5   |
| 2          | 0        | 0   | 0   | 0   |
| 3          | 175      | 10  | 1   | 0   |
| 4          | 252      | 34  | 2   | 0   |
| 5          | 59       | 5   | 0   | 0   |
| 6          | 233      | 43  | 5   | 0   |
| 7          | 840      | 845 | 831 | 754 |
| 8          | 30       | 0   | 0   | 0   |
| 9          | 8        | 0   | 0   | 0   |
| 10         | 15       | 0   | 0   | 0   |
| 11         | 42       | 0   | 0   | 0   |

TABLE 4.3: Résultats des calculs de rapport de vraisemblance ( $RV$ ) de l'expérience 3 – b, soit des situations où le matériel *indiciaire* est issu d'un mélange des trois textes de menace (*c.-à-d.* les sessions *i – iii* du scripteur 7) et le matériel de *référence* est issu des échantillons de l'un des scripteurs de l'expérience 2 (à savoir les scripteurs 1 à 11). Les chiffres indiqués représentent le nombre d'itérations, sur 1000, lorsque les observations soutiennent l'hypothèse  $H_p$  : le matériel *indiciaire* et de *référence* proviennent du même scripteur. Les résultats indiquent que le scripteur des textes de menace est le scripteur 7, en effet ce scripteur est celui qui a le plus de  $RV$  indiquant que l'hypothèse  $H_p$  est vraie.

résultats mettent en évidence une variation de l'écriture suite à un espace temporel long ;

- les taux de *faux positifs* sont beaucoup plus faibles et stables dans cette expérience 3 – b que dans l'expérience 3 – a. Par exemple, dans l'expérience 3 – a, la lettre o affiche un taux de *faux positifs* d'environ 80% à  $n_s = 10$ , alors que ce même taux est atteint à  $n_s = 20$  dans l'expérience 3 – b. Ceci indique qu'une prise d'échantillons sur une session (ou 1 journée) n'est pas suffisant pour associer correctement les scripteurs, comme dans l'expérience 3 – a, alors qu'un mélange de trois sessions 3 – b permet de réduire le taux d'erreur. Ces résultats mettent donc en évidence une variation de l'écriture suite à un espace temporel très court ;
- les taux d'erreur ne sont pas similaires entre les lettres étudiées, par exemple :
  - les taux de *faux positifs* de la lettre a varient entre 2% et 9.2%, alors que ces mêmes taux varient de 0% à 0.07% pour la lettre o ;
  - les lettres d et q ont des taux de *faux négatifs* très faibles (0.2 – 9.3%) en comparaison aux autres lettres (> 15.5% pour les lettres a, e et o).
- les taux d'erreur varient en fonction du scripteur de *référence* étudié. Comme dans l'expérience 3 – a, pour la lettre q les scripteurs 1, 4, 5, 6 et 8 se démarquent du restent avec leurs taux de *faux positifs*. D'un autre côté le scripteur 10, ne présente pas de *faux positifs*, alors que c'était uniquement le cas dans le texte de menace *ii*. Ceci met en évidence l'importance des variations dues aux scripteurs étudiés ;
- dans l'expérience 3 – b, comparable à de l'*identification* de scripteur, le scripteur 7 est correctement associé aux textes de menace. En effet, il est le seul scripteur à posséder des valeurs de  $LRV = \log(RV)$  positives pour toutes les lettres. Ces résultats positifs permettent de contribuer à la validation du processus développé dans cette recherche.

La distribution des rapports de vraisemblance est illustrée dans la Figure 4.29, pour les lettres q et o. Ces résultats montrent, au-delà des éléments déjà présentés dans la Table 4.3, que :

1. la distribution des  $LRV$  varie selon les scripteur comparés. Par exemple, la moyenne en comparant les textes de menace du scripteur 7 au scripteur 10 est de  $LVR = -3.95$ , contre  $LRV = -1.38$  en le comparant au scripteur 3. Au même titre que la Figure 4.20, ces résultats soulignent le degré de similarité entre l'écriture d'un caractère x de deux scripteurs, disons A et B, n'est pas nécessairement le même que celui entre les scripteurs B et C, ou A et C. Ils indiquent aussi, que pour un même couple de scripteurs, le degré de similarité de l'écriture d'un caractère x n'est pas nécessairement le même que celui pour un caractère y.
2. les moyennes des  $LRV$  lorsque  $H_d$  est vraie (*c.-à-d.* en comparant des échantillons du scripteur 7 de l'expérience 2 et 3), sont supérieures à 0 pour la lettre h, alors que ceux pour la lettre o sont inférieures à 0, supportant donc l'hypothèse inverse. Ces résultats mettent en avant la variation de la lettre o du scripteur 7 le long d'un espace temporel long (> 1 an).

En conclusion les résultats présentés dans cette section soulignent :

1. qu'il existe des variations dans l'écriture au cours d'un espace temporel court (expérience 3 – b) et long (expérience 3 – b) ;

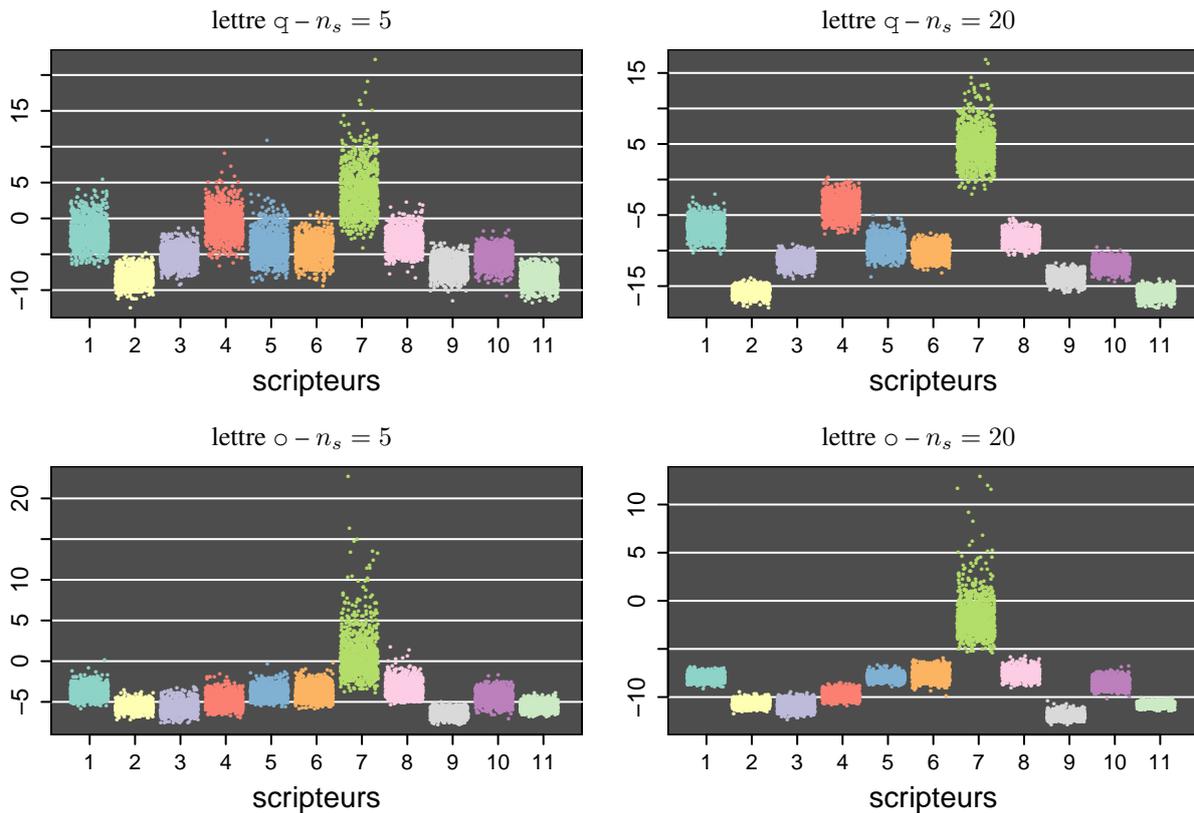


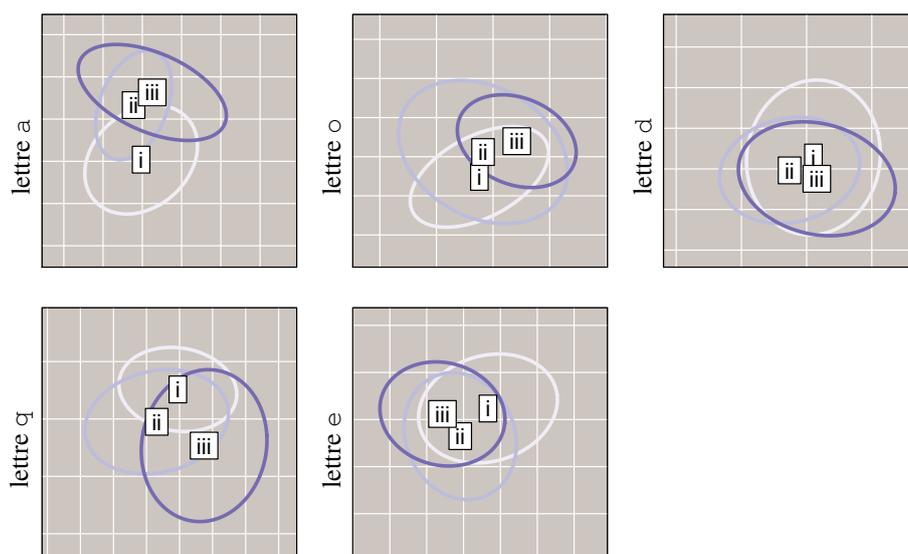
FIGURE 4.29: Illustration de la distribution des  $LRV$  de l'expérience 3–b, pour les lettres q et o, avec  $n_s = 5, 10$ . On observe dans ces résultats le caractère individuel de chaque comparaison de scribeur. En effet, les moyennes de  $LRV$  lorsque  $H_d$  est vraie (*c.-à-d.* les scribeurs du matériel de *référence* et *indiciaire* sont différents), varient selon le scribeur étudié. Lorsque  $H_d$  est vraie (*c.-à-d.* en comparant des échantillons du scribeur 7 de l'expérience 2 et 3), les  $LRV$  obtenus pour la lettre h sont supérieurs à 0, alors que ceux pour la lettre o sont inférieurs à 0, supportant donc l'hypothèse inverse. Ces résultats traduisent la variation de la lettre o du scribeur 7 le long d'un espace temporel long ( $> 1$  an).

2. que les résultats peuvent varier fortement selon la méthode d'échantillonnage utilisée (différences entre les résultats des expériences 2 et 3) ;
3. les variations des taux d'erreur en fonction du(des) scribeur(s) étudié(s) sont très importantes. Ainsi il est mis en évidence que ces variations dues aux scribeurs sont masquées dans les résultats globaux.

### 4.3.3 Représentation par Analyse en Coordonnées Principales

Cette section présente les résultats des représentations des matrices de dissimilarité par PCoA pour le scribeur 7. Soit les dissimilarités qui ont été calculées entre les échantillons issus des expériences 2 et 3. La PCoA est une méthode d'ordination qui permet de représenter les échantillons sur un plan en deux dimensions selon ses deux composantes principales. Par exemple la PCoA peut être utilisée pour visualiser les différences entre

**Représentation PCoA – fonction  $FE_1$**



**Représentation PCoA – fonction  $FE_2$**

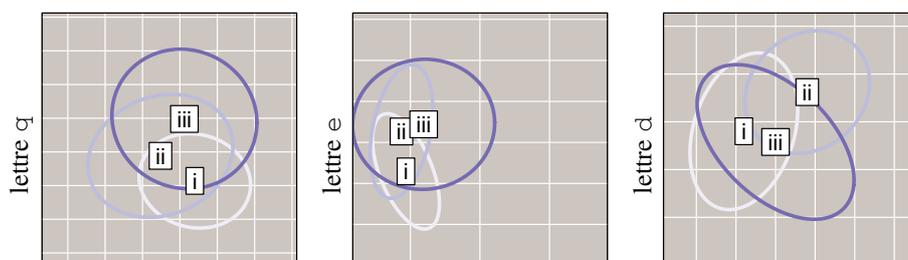


FIGURE 4.30: Représentation des dissimilarités entre les échantillons des trois textes de menace ( $i - iii$ ) par la méthode de PCoA. Ce sont les échantillons rédigés par le scripteur 7, pour les lettres a, e, d, o et q. Les graphiques illustrent donc les variations de l'écriture dans un espace temporel court (les trois textes de menace ont été rédigés en trois jours, chacun à un jour différent). Ces graphiques montrent qu'il existe une différence entre les sessions, indépendamment de la lettre étudiée. Par exemple pour la lettre a, les trois groupes représentant les textes de menace ne sont pas superposés, indiquant que les dissimilarités entre les échantillons qui proviennent des trois textes. Par ailleurs, on remarque que les résultats varient selon la *fonction d'extraction* : dans le cas de la lettre d les différences sont moins notables pour  $FE_1$  que  $FE_2$ . Ces résultats supportent donc qu'il existe une variation quotidienne de l'écriture d'un même scripteur.

les caractères lorsqu'ils sont regroupés par sessions (*c.-à-d.* les trois lettres de menace  $i - iii$  ou les cinq pages  $I - V$ ), et par conséquent déterminer s'il existe une variation temporelle de l'écriture.

Les résultats des représentations par PCoA sont illustrés dans les Figures 4.30 et 4.31. Ces résultats montrent que :

- il existe des différences entre les échantillons des trois textes de menace du scripteur 7. A titre d'exemple, dans la Figure 4.30 les trois groupes représentant les échantillons de la lettre a ne sont pas superposés, indiquant que les dissimilarités entre

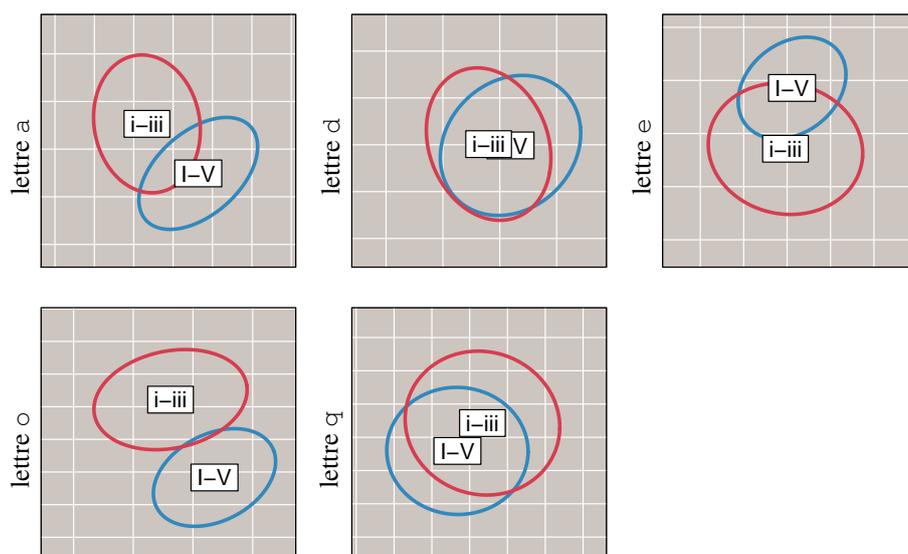


FIGURE 4.31: Représentation, par la méthode de PCoA, des dissimilarités entre les échantillons des trois textes de menace ( $i - iii$ ) rédigés par les scripteurs 7, pour les lettres a, e, d, o et q et la fonction d'extraction  $FE_1$ . Les échantillons sont groupés en deux catégories, 1) l'ensemble de cinq pages  $I - V$  et 2) l'ensemble de trois textes de menace  $i - iii$ . Les graphiques illustrent donc les variations de l'écriture dans un espace temporel long (les cinq pages et les trois textes de menace ont été rédigés à un an d'intervalle). Les graphiques montrent qu'il existe une différence entre ces sessions. Par exemple pour la lettre o, les deux groupes représentant les sessions ne sont pas superposés, indiquant que les dissimilarités entre les échantillons provenant des sessions  $I - V$  et  $i - iii$  diffèrent. On note aussi que ces différences ne sont pas similaires selon la lettre étudiée. Par exemple, les représentations de deux groupes avec la lettre e affichent moins de différences qu'avec la lettre a.

les échantillons issus des trois textes de menace diffèrent. Ceci signifie que l'écriture de ce scripteur varie en fonction d'une période temporelle courte. Ces conclusions sont donc similaires à celles de l'expérience 2 (voir Figure 4.23) ;

- il existe des différences entre les échantillons des textes de menace ( $i - iii$ ) et des pages ( $I - V$ ) du scripteur 7. Par exemple pour la lettre o (Figure 4.31), les deux groupes représentant les sessions ( $i - iii$ ) et ( $I - V$ ) ne sont pas superposés. Ces résultats indiquent une variation des dissimilarités entre les échantillons provenant des sessions de l'expérience 2 et 3, et donc une variation de l'écriture sur une période temporelle longue ;
- les variations sur une période temporelle courte ou longue, énoncées ci-dessus, ne sont pas homogènes selon les lettres étudiées. Ainsi dans le cas de la Figure 4.30, on remarque que les différences entre les trois groupes, représentant les trois textes de menace, sont moins notables avec la lettre d qu'avec la lettre a. Ces résultats soulignent que pour un scripteur donné, chaque lettre possède sa propre variation ;
- les variations sur une période temporelle courte ou longue, énoncées ci-dessus, ne sont pas homogènes selon la fonction d'extraction étudiées. Par exemple pour la lettre d (Figure 4.30), les différences entre les trois textes de menace sont plus

prononcées pour la fonction  $FE_1$  que pour la fonction  $FE_2$ . Ceci indique donc que, pour un scripteur donné, les variations sur une période temporelle sont moins importantes pour certaines caractéristiques de l'écriture que d'autres.

En conclusion, il est donc important de souligner :

1. qu'il existe des variations de l'écriture du scripteur, ces observations sont à la fois présentes sur une période temporelle courte ou longue ;
2. que pour un scripteur donné, chaque lettre possède sa propre variation.

## Discussion

Les résultats des trois phases de la recherche (développement d'un outil informatique de collecte de l'information, validation de ce processus, applications à des situations d'intérêt forensique) sont examinés dans ce chapitre. Dans une première et deuxième section les résultats de validation et d'application de la méthode sont discutés (sections 5.1 et 5.2). Les limites et avantages du processus de *vérification* développé dans cette recherche, sont abordés dans une troisième section (5.3). Enfin, une conclusion générale de la recherche est donnée dans la dernière section (5.4).

### 5.1 Validation de la méthode

La validation de la méthode développée est consacrée à l'étude des mesures calculées aux différentes étapes du processus, et de leur capacité à discriminer des scripteurs selon leur écriture. Deux objectifs ont été formulés en introduction pour cette phase de la recherche (section 1.3.2), à savoir :

- analyser et déterminer quelles sont les mesures, parmi l'ensemble des mesures développées dans la recherche, qui fournissent des résultats capables de discriminer des formes tout en étant robustes face à la taille et l'épaisseur du trait (expérience 1) ;
- déterminer la capacité du processus de *vérification* à discriminer des scripteurs (expériences 2 et 3). En d'autres termes évaluer si les mesures sélectionnées minimisent l'*intra-variabilité* et maximisent l'*inter-variabilité* de l'écriture de scripteurs.

#### 5.1.1 Invariance à la taille et l'épaisseur

L'objectif de l'expérience 1 est de sélectionner  $n_{FE}$  mesures parmi les 126  $F_{ratio}$  potentiels, au travers du jeu de figures géométriques, qui :

- puissent discriminer les figures selon leurs formes géométriques ;
- ne soient pas la taille et l'épaisseur du trait de ces figures.

Ainsi cette expérience démontre la volonté d'obtenir un processus qui quantifie le polymorphisme des caractères. Par ailleurs, il est fixé comme contrainte de garder une *fonction de pas* DTW commune pour toutes les *fonctions d'extraction* choisies.

Dans une certaine mesure, la taille du caractère ou l'épaisseur de son tracé peuvent être considérées comme des caractéristiques individuelles lors d'expertises en document. En effet, la pression appliquée au travers de l'instrument scripturant varie entre scripteurs, et peut ainsi influencer l'épaisseur du tracé. De même pour la taille des caractères, qui est susceptible de varier entre les scripteurs. En revanche, l'épaisseur du tracé peut aussi

être influencé par le type d'instrument scripturant (stylo bille, crayon, ...), le support ou la fatigue et le stress du scripteur. De même, la taille des caractères peut être influencée par la présence ou non de lignes directrices sur le document ou de contraintes d'espaces comme dans les formulaires. En d'autres termes, ces caractéristiques sont susceptibles d'être influencées par bien d'autres facteurs, qui sont incontrôlables, et donc à terme peuvent biaiser les mesures effectuées sur les caractères manuscrits. En conclusion, il a été choisi de construire des mesures qui soient indépendantes de la taille et de l'épaisseur des figures géométriques.

En pratique, seules  $n_{FE} = 2$  des 18 *fonctions d'extraction* satisfaisaient les exigences. Par ailleurs, toutes les *fonctions d'extraction* sans normalisation (soit celles composées de  $N_3(q) = 1$ ), où basées sur la *fonction de quantification* la plus simple ( $M_1$ ) sont fortement corrélées à la taille des figures géométriques.

De ce fait, les résultats de l'expérience 1 soulignent l'importance d'étudier les fonctions employées afin de s'assurer que celles-ci ne soient pas une source de biais. Par exemple, en prenant la fonction  $M_1 - N_1$  les échantillons de l'expérience 2 seraient discriminés par la taille de caractères, non pas par leur morphologie. Par ailleurs, ces résultats montrent qu'une simple méthode de normalisation ( $N_1$ ) permet de limiter l'influence des facteurs de taille et d'épaisseur, et donc limiter les biais des *fonctions d'extraction* face à ces facteurs.

### 5.1.2 Discrimination de scripteur et taux d'erreur

Les résultats de la section précédente (soit de l'expérience 1) ont permis de sélectionner les mesures de quantification d'écriture qui seront utilisées pour les expériences 2 et 3. Ainsi, maintenant que l'étape d'extraction de caractéristiques est établie, l'objectif de cette section est de démontrer que les autres étapes du processus de *vérification* permettent de discriminer des scripteurs. Ces étapes concernent respectivement :

- la mesure de dissimilarité, à savoir les matrices de dissimilarité  $\mathbf{D}$  (représentées par la technique de PCoA dans la section 4.2.4) ;
- l'analyse de dissimilarité, à savoir les  $F_{ratio}$  calculés (représentés par les densités dans la section 4.2.2) ;
- l'évaluation statistique, à savoir les rapports de vraisemblance (présentés dans les sections 4.2.3 et 4.3.2).

Les représentations par PCoA permettent d'observer les échantillons d'écriture sur un plan en deux dimensions. De cette manière, il est possible d'étudier le potentiel de discrimination de scripteurs des dissimilarités calculées en comparant la position des groupes de scripteurs sur ces graphiques. Naturellement, plus les échantillons des scripteurs sont séparés, meilleure est la discrimination de scripteurs. Les résultats de ces représentations montrent que les mesures calculées permettent de distinguer les scripteurs selon leur écriture, cependant certaines lettres sont moins discriminantes que d'autres. Par exemple, les groupes de scripteurs se discriminent de préférence avec la lettre **h** qu'avec la lettre **e** (voir Figure 4.24). On note également que les résultats issus des différentes *fonctions d'extraction* ne sont pas homogènes. En effet chacune possède sa propre variation et ne discrimine pas les scripteurs de manière similaire. Ceci montre donc que ces différentes mesures apportent des informations complémentaires.

Ainsi l'ensemble des résultats soulignent que les *fonctions d'extraction* sélectionnées apparaissent comme des mesures utiles pour discriminer des scripteurs.

Tout d'abord, il est important de discuter de la normalité des variables utilisées pour le calcul du rapport de vraisemblance, à savoir les deux  $F_{ratio}$  obtenus :  $FE_1$ ,  $FE_2$ . En effet, dans certains scénarios celles-ci ne respectent pas toujours la loi normale, et ce malgré la transformation préalable des variables pour corriger leur normalité (avec la transformée  $\log(|.|)$ ). Par conséquent, il est possible que la formulation statistique ne soit pas en mesure de modéliser correctement les données, et donc induire des erreurs lorsqu'il s'agit d'évaluer et interpréter les résultats de  $RV$ . Toutefois, plusieurs éléments indiquent la fiabilité des résultats en dépit du fait que la condition de normalité ne soit pas respectée dans tous les cas.

Dans un premier temps, les comparaisons des rapports de vraisemblance *multivariés* et *univariés* (c'est-à-dire les résultats des  $RV$  pour les mêmes scénarios, en considérant toutes les variables  $FE_1$  et  $FE_2$  pour le cas *multivarié*, et en considérant une seule de ces variables, par exemple  $FE_1$ , pour le cas *univarié*) montrent les mêmes tendances et suggèrent, dans toutes les expériences, les mêmes conclusions (voir aussi la section 5.2).

Dans un deuxième temps, tous les résultats ne faisant pas intervenir le calcul du rapport de vraisemblance (soit les résultats des dimensions  $W$  et  $H$ , des représentations par PCoA et des  $F_{ratio}$ ) sont concordants avec les résultats des rapports de vraisemblance.

Par conséquent, il est admis que les distributions des variables ne compromettent pas les interprétations des résultats des rapports de vraisemblance.

Afin de déterminer si l'analyse de dissimilarité est une métrique utile afin de discriminer des scripteurs, il convient de vérifier si les hypothèses  $H_0$  et  $H_1$  sont respectées dans les scénarios. Les représentations des  $F_{ratio}$  issus des résultats des scénarios 1 à 5 (expérience 2)<sup>1</sup> mettent en évidence que ces hypothèses sont tenues dans les expériences. C'est-à-dire, que dans le cas où les échantillons proviennent :

1. du même scripteur, les  $F_{ratio}$  soient autour de 1, correspondant à  $H_0$  vraie ;
2. de scripteurs différents, les  $F_{ratio}$  soient supérieurs à 1, correspondant à  $H_1$  vraie ;

En effet on observe, d'un côté que les valeurs des  $F_{ratio}$  se situent autour de 1 lorsque  $H_0$  est vraie, de l'autre côté que les valeurs s'en éloignent lorsque  $H_1$  est vraie (voir par exemple la Figure 4.12), ce qui est en accord avec les hypothèses.

De ce fait, ces résultats permettent de confirmer le postulat de départ, à savoir que les dissimilarités calculées entre les échantillons d'un même scripteur sont similaires (correspondant à  $H_0$ ), alors que celles calculées entre deux scripteurs sont différentes (correspondant à  $H_1$ ). Ainsi les résultats de l'analyse de dissimilarité indiquent que cette métrique est utile afin de discriminer des scripteurs.

Dans l'étape d'évaluation, l'objectif est celui de savoir si le modèle statistique proposé est adéquat pour interpréter les observations (*c.-à-d.* dans ce cas les  $F_{ratio}$ ), donc déterminer si les hypothèses  $H_p$  et  $H_d$  sont respectées dans le calcul des rapports de vraisemblance des scénarios créés. Entre d'autres termes, il s'agit d'étudier les taux de *faux négatifs* (soit lorsque les documents *indiciaire* et de *référence* proviennent du même scripteur, et que les résultats du  $RV$  indiquent l'inverse) et les taux de *faux positifs* (soit lorsque les documents *indiciaire* et de *référence* proviennent de deux scripteurs différents, et que les résultats du  $RV$  indiquent l'inverse) calculés dans les expériences 2 et 3.

---

1. A noter que dans la phase de validation, seuls les scénarios comparant des caractères manuscrits de même taille sont considérés. Ainsi le scénario 6, où le matériel *indiciaire* est composé de caractères *normaux* et le matériel de *référence* de caractères *grands*, n'est pas pris en compte dans cette étape.

Les résultats de l'expérience 2 montrent que les valeurs des rapports de vraisemblance permettent de discriminer les scripteurs selon leur écriture. Par exemple, pour la lettre **h**, le taux de *faux positifs* est de 0% pour  $n_s = 10, 15, 20$ , alors que le taux de *faux négatifs* ne l'est que pour  $n = 20$ .

Les *scénarios* de l'expérience 3 peuvent s'apparenter à de l'*identification* de scripteur. Il s'agit des situations où l'auteur du document *indiciaire* (le scripteur 7) est présent dans une base de données (les échantillons de l'expérience 2). Sous cette condition, les résultats des calculs de rapport de vraisemblance indiquent sans ambiguïté que le scripteur 7 est l'auteur des textes de menace.

En conclusion l'ensemble des résultats présentés dans cette section permettent de répondre aux deux objectifs de la **validation** du processus développé, à savoir :

1. s'assurer que les fonctions d'extraction de caractéristiques puissent discriminer les caractères en fonction de leur forme, et que cette mesure soit indépendante de la taille et de l'épaisseur des caractères ;
2. démontrer que les aspects théoriques et les hypothèses des différentes étapes du processus sont respectées dans la pratique, et par conséquent, que cette méthodologie est utile pour quantifier et évaluer l'écriture manuscrite.

## 5.2 Exemples d'application

L'objectif de cette phase de la recherche est de fournir des exemples d'application concrets du processus de *vérification* dans le domaine forensique. Deux sujets sont abordés dans cette étude (voir la section 1.3.3) :

- l'analyse de l'influence de la taille des caractères sur leur morphologie (expériences 1 et 2). Les proportions des caractères ou leurs dimensions relatives, sont des caractéristiques fréquemment utilisées par les experts en écriture manuscrite (Huber et Headrick [2]). Sur ce sujet, deux courants sont observés dans la littérature. D'un côté, ceux qui considèrent que les proportions, et donc la morphologie, des caractères est invariante à la taille des caractères (Harrison [93] et Wann et Nimmo-Smith [94]). De l'autre côté, ceux qui soutiennent que la taille des caractères est un facteur significatif de la morphologie (Marquis et al. [67], Rogers et Found [95]). Les résultats obtenus permettent de soutenir l'un de ces courants ;
- le variations des caractères manuscrits au cours du temps (expériences 2 et 3). La variation naturelle de l'écriture, et son évolution dans le temps sont des phénomènes établis et connus des experts en écritures. Cependant cet aspect est souvent négligé dans les études en écriture manuscrite (Thiéry et al. [73]). Les résultats des expériences de cette étude permettent de souligner l'importance de ces variations.

### 5.2.1 Influence de la taille des caractères

L'influence de la taille des caractères sur leur morphologie est étudiée à travers le jeu de données de l'expérience 2, où il a été demandé aux scripteurs d'écrire les caractères 1) à leur taille habituelle et 2) à une taille équivalente à trois fois celle de l'écriture habituelle.

Tout d'abord, les représentations des dimensions des caractères ( $W$ ,  $H$ ,  $W \times H$ ) montrent qu'aucun des scripteurs n'a su respecter la consigne. En effet, les dimensions

des caractères *grands* ne sont pas à une taille trois fois supérieure de l'écriture *normale*. Toutefois, les caractères *grands* sont de dimensions supérieures aux caractères *normaux* et suggèrent que l'expérience 2 permet bien de comparer des caractères des tailles différentes. Par ailleurs la majorité des scripteurs ont tendance à plus augmenter la hauteur que la largeur en fonction de la taille, soit de diminuer le rapport  $W/H$ . Sur ce point, les résultats obtenus rejoignent donc ceux de Rogers et Found [95], et Marquis et al. [65, 67].

Les résultats généraux des rapports de vraisemblance (*scénarios* 1, 5 et 6) permettent d'étudier l'influence de la taille de caractères sur leur morphologie. Les *scénarios* 1 et 5 font appel aux situations où les caractères manuscrits du matériel *indiciaire* et de *référence* sont de même taille, soit respectivement des caractères *normaux* et *grands*. Inversement, le *scénario* 6 fait appel aux situations où les échantillons du matériel *indiciaire* sont des caractères *grands* et les échantillons du matériel de *référence* des caractères *normaux*. Ces résultats indiquent qu'il existe des différences entre les caractères *normaux* et *grands* des scripteurs.

En effet, lorsque un mélange des deux types de caractères sont étudiés (*scénario* 6) les taux d'erreur du processus de *vérifications* sont en moyenne 3.8 fois supérieurs aux taux d'erreur lorsque le même type de caractère est étudié (*scénarios* 1 et 5). Par ailleurs on constate que ce sont les taux de *faux négatifs* (soit les situations où les échantillons d'écriture sont écrits par le même scripteur, mais où la valeur du rapport de vraisemblance indique l'inverse) qui subissent la plus forte augmentation. Par exemple, en considérant les *scénarios* 1 et 6, le taux de *faux négatifs* augmente de 259% contre 167% pour le taux de *faux positifs*.

Cependant, avant de pouvoir tirer une quelconque conclusion, il est nécessaire de s'assurer que les différences observées entre les caractères *normaux* et *grands* ne soient pas dûes à un biais de la mesure. Les efforts et résultats de l'expérience 1 ont permis d'affirmer que les mesures sélectionnées (*fonctions d'extraction* + *fonction de pas DTW*) discriminent les formes malgré une différence dans leur taille. Toutefois, dans l'expérience 1 les proportions sont respectées entre les figures géométriques de tailles différentes, ce qui n'est pas obligatoirement le cas des échantillons d'écritures manuscrite de l'expérience 2. Par ailleurs, les formes (ou la morphologie) des caractères manuscrits sont bien plus complexes et variées que les formes des figures géométriques. Ceci étant les calculs de rapports de vraisemblance des *scénarios* 1, 5 et 6 sont similaires dans le cas *multivarié* et les cas *univariés*, soit selon des fonctions ayant une sensibilité différente à la dimension de caractères analysés. De ce fait, il est admis que les *fonctions d'extraction* étudiées dans cette recherche ne sont pas biaisées par la taille de caractères.

Par ailleurs, plusieurs éléments viennent contredire la généralité de la première observation, à savoir qu'il existe des différences entre les caractères *normaux* et *grands* des scripteurs. En effet, comme indiqué dans la section 5.1.2, les résultats globaux des rapports de vraisemblance masquent les variations individuelles des scripteurs. Par exemple, pour la lettre **a**, le taux de *faux négatifs* global est de 35.9%, alors qu'il est de 8% pour le scripteur 10 et de 95.5% pour le scripteur 6. Ainsi ces résultats indiquent que certains scripteurs ont des caractères *normaux* similaires aux caractères *grands*. Ces résultats ne signifient pas que l'écriture de ces scripteurs ne sont plus les mêmes, mais soulignent les différences observées dans les dissimilarités calculées. Si des proportions de la hauteur et de la largeur des caractères peuvent changer entre les caractères *normaux* et *grands*, les scripteurs ont toujours le même style d'écriture.

En conclusion ces résultats démontrent :

1. qu'il existe des différences entre les caractères *normaux* et *grands* issus 11 scribeurs de l'expérience 2. Par conséquent, ces observations sont une indication que les scribeurs ne reproduisent pas de manière proportionnelle des caractères écrits à des tailles différentes. Ainsi ces conclusions confirment les observations de Marquis et al. [66] et Rogers et Found [95], et sont en désaccord avec les propos de Harrison [93], à savoir l'invariance de l'écriture en fonction de la taille des caractères. Toutefois, il est important de noter que ces résultats dépendent avant tout du scribeur et de la lettre étudiée ;
2. que l'étude de caractères de taille différente peut aussi impliquer une écriture inhabituelle des scribeurs. En supposant que le manque d'expérience dans l'écriture *grande* des scribeurs puisse contribuer aux différences observées, il apparaît nécessaire d'appréhender avec précaution les généralités trouvées dans ces publications à ce sujet. De ce fait, ce facteur (*c.-à-d.* l'écriture inhabituelle) doit faire l'objet d'une attention particulière lors de recherche sur ce sujet afin de mieux le contrôler et ainsi de pouvoir étudier l'influence de la taille des caractères avec plus de précision.

Ainsi, sans pouvoir donner de conclusions communes à tous les scribeurs concernant les différences de caractéristiques entre caractères de tailles différentes, les résultats permettent tout de même de renforcer les recommandations de Marquis et al. [67] sur le protocole d'échantillonnage d'une expertise en écriture manuscrite : il est fortement conseillé de comparer des échantillons du matériel *indiciaire* et de *référence* qui soient de même taille (– dimensions des caractères).

## 5.2.2 Variation naturelle de l'écriture : influence du temps

La variation naturelle de l'écriture en fonction du temps est étudiée dans les expériences 2 et 3. En effet, dans ces expériences, plusieurs scribeurs ont rédigé des documents manuscrits sur différentes sessions (soit différents jours), à savoir :

1. cinq documents sur 15 jours (soit les pages  $I - V$ , représentant un espace temporel court) pour l'expérience 2 ;
2. trois documents sur 3 jours pour l'expérience 3 (soit les textes de menace  $i - iii$  du scribeur 7, représentant espace temporel très court), un an après la prise d'échantillon de l'expérience 2 (le décalage entre la prise d'échantillon de l'expérience 2 avec l'expérience 3 représente un espace temporel long).

Ainsi la comparaison de la quantification des caractères manuscrits issus de toutes ces sessions permet d'étudier les variations de l'écriture sur différents espaces temporels.

Tout au long de cette recherche, la totalité des mesures effectuées (les dimensions  $W$  et  $H$ , les représentations par PCoA et les calculs de rapport de vraisemblance) souligne l'existence d'une variation temporelle de l'écriture. Il est important de noter qu'on observe ces variations indépendamment de la lettre, du scribeur, et de la mesure étudiée, bien que celles-ci ne soient pas homogènes.

Dans un premier temps, les représentations visuelles des caractères (à savoir les dimensions  $H$ , et  $W$ , le rapport  $W/H$  et la technique de PCoA) issus des expériences 2 et 3 indiquent sans ambiguïté qu'il existe une variation des caractéristiques d'écriture en fonction du temps. Par exemple, pour la lettre **e** du scribeur 7, d'un côté les valeurs

des largeurs  $W$  augmentent successivement dans les trois textes de menace, alors que les hauteurs  $H$  diminuent. D'un autre côté, les représentations des échantillons issus des textes de menace par PCoA indiquent des similarités entre les sessions  $ii$  et  $iii$ , et que celles-ci diffèrent avec la session  $i$ .

Dans un deuxième temps, les résultats des rapports de vraisemblance indiquent aussi qu'il existe une variation temporelle de l'écriture manuscrite. En effet, la comparaison des résultats de l'expérience 2 (soit lorsque les documents de *référence* et *indiciaire* ont été rédigés lors du même espace temporel, < 15 jours) et de ceux de l'expérience 3 (soit dans le cas où les documents de *référence* et *indiciaire* ont été rédigés lors de deux espaces temporels distincts, *c.-à-d.* une différence supérieure à 1 an) indiquent une différence entre des écritures d'un même scripteur au cours d'un espace temporel long. Par ailleurs, des variations sont aussi observées lorsque l'espace temporel étudié est court (différence de 1 jour), comme en attestent les différences des résultats des expériences 3 –  $a$  et 3 –  $b$ .

En conclusion ces résultats montrent que :

1. il existe une variation temporelle de l'écriture manuscrite. Les variations de l'écriture dans cette recherche sont observées sur une période de 3 jours, de 15 jours et de 1 an. Par ailleurs, il est important de noter qu'elles sont observées dans tous les échantillons d'écriture, à travers toutes les mesures et les représentations visuelles. Toutefois les résultats ne permettent pas d'affirmer avec certitude si les variations sur une longue période sont plus fortes que des variations sur une courte période. En revanche, il est clair qu'il existe une variation journalière de la morphologie des caractères manuscrits ;
2. la prise d'échantillons est une étape cruciale de la recherche en écriture manuscrite : d'elle dépend, peu importe les méthodes d'extraction ou statistiques employées, tous les résultats et interprétations de la recherche. Ainsi, le matériel utilisé (nombre de scripteurs, texte à copier, nombre de sessions, instrument scripturant, ...) devrait être décrit en détail dans les publications. Cependant, beaucoup d'études en écriture manuscrite ne montrent que très peu d'intérêt à expliquer leur prise d'échantillons (par exemple voir l'article de Srihari et al. [53]). Sans ces informations, il est impossible au lecteur de savoir dans quelles conditions les échantillons ont été obtenus, et par conséquent il est impossible de vérifier, dans une partie tout du moins, la véracité des résultats.

Enfin, il est important de noter que les résultats ne suggèrent pas des différences morphologiques significatives entre des échantillons produits à des jours différents. En effet, les résultats indiquent plutôt que les variations journalières de l'écriture sont telles, que des échantillons pris sur une seule session (*c.-à-d.* sur une journée) ne sont pas suffisants pour observer la totalité de la variabilité de l'écriture d'un scripteur. Par exemple, ces résultats rejoignent les observations de Bohn [5] : il n'est pas rare qu'un individu adopte plusieurs styles ou systèmes d'écriture.

Ainsi, l'ensemble des résultats de cette recherche suggère que, pour obtenir une meilleure information des variations de l'écriture d'un scripteur, il est nécessaire d'obtenir plusieurs échantillons sur plusieurs sessions d'écriture de *référence*. Dans le cas d'un expertise en écriture manuscrite, il est préférable de comparer du matériel *référence* contemporain au matériel *indiciaire*.

## 5.3 Limites et avantages de la méthodologie développée

Cette section aborde avec critique la méthodologie employée dans cette recherche. Les limites et les avantages de la méthodologie sont discutés par ordre des étapes du processus de *vérification* (voir Figure 2.1), à savoir : l'échantillonnage, le pré-traitement, l'extraction de caractéristiques, la mesure de dissimilarité, l'analyse de dissimilarité, et l'évaluation.

Premièrement, le nombre de scripteurs composant le jeu de données est très réduit (11 scripteurs). Par conséquent, il est certain que cet échantillon de scripteurs ne peut être considéré comme représentatif d'une population. De plus, les échantillons d'écriture manuscrite sont limités à une écriture de type *script*, et les scripteurs étaient astreints à recopier le texte qui leur était fourni. Ainsi, comme mentionné précédemment, ces échantillons ne peuvent pas prétendre de capturer la variabilité totale de l'écriture des 11 scripteurs.

Cependant, il est nécessaire de rappeler que l'objectif principal de cette recherche est de proposer une nouvelle méthodologie pour quantifier et évaluer l'écriture manuscrite, soit le processus de *vérification* de scripteur développé. Le but étant celui d'aider un expert ou un chercheur en écriture manuscrite à quantifier et étudier l'écriture manuscrite. Ainsi, au travers de ces expériences, le but est ni d'individualiser l'écriture de chaque scripteur ni de prouver l'individualité de l'écriture, mais de fournir des exemples concrets d'application et d'exploitation de la méthode. Les résultats doivent donc être considérés comme un élément d'analyse d'un style d'écriture (écriture de type *script*), parmi d'autres, pour un groupe de scripteurs donné. En revanche, ces limitations n'empêchent pas d'observer et de quantifier des tendances pour ce groupe limité de scripteurs, et de ce fait de pouvoir comparer les résultats obtenus avec ceux d'autres études sur l'écriture manuscrite (voir section 5.2).

L'impact du processus de binarisation sur la mesure et les résultats n'est pas étudié dans cette recherche. En effet, sachant que la totalité du caractère binarisé est analysé<sup>2</sup>, si chaque utilisateur manipule différemment les images, il est possible que cela ait un impact sur la mesure. Cependant, tous les caractères dans cette recherche ont été binarisés par le même utilisateur, et de ce fait il est impossible de déterminer l'influence de ces manipulations sur le processus de *vérification* de scripteur. Toutefois, les écritures dans cette recherche sont de type *script* et limitent donc fortement les variations entre utilisateurs, contrairement à une écriture cursive où la décision de segmenter les caractères d'un mot peut varier entre utilisateurs. Ainsi, afin de garantir une étude complète du processus de *vérification* développé dans cette recherche, il serait intéressant de connaître l'influence de l'utilisateur.

Les *fonctions d'extraction* et la fonction de dissimilarité DTW sont des mesures *non-conventionnelles* (c'est-à-dire des mesures qui ne font pas référence à des mesures employées par des experts en écriture). Ainsi, un expert ne peut pas relier une différence entre les échantillons et celles suggérées par les matrices de dissimilarité  $\mathbf{D}$  ou par le vecteur

---

2. Plus exactement les contours du caractère sont analysés, contrairement à Marquis et al. [65] ou ce sont les squelettes des boucles qui ont été étudiés.

de caractéristiques  $\mathbf{y} = (FE_1, FE_2)$  à une différence observée dans les caractéristiques qu'utiliserait un expert en écriture.

Cependant la méthode développée dans cette recherche n'a pas pour vocation de remplacer un expert ni, dans son état actuel, d'être utilisée pour fournir une valeur statistique lors d'une expertise. Cette méthode est un outil de quantification permettant d'étudier les variations dans l'écriture manuscrite.

L'utilisation de méthodes informatiques apporte plusieurs avantages. Dans un premier temps, un gain en efficacité et en puissance de calcul qui permet d'analyser les 126 combinaisons de *fonctions d'extraction* et d'étudier 13 lettres sur 11 scripteurs. Sans l'opportunité d'étudier toutes ces combinaisons il est fort probable que la *fonction d'extraction* étudiée ait été la fonction  $M_1 - N_3$ , soit la fonction la plus simple. Mais son importance réside dans le fait qu'elle est une des *fonctions d'extraction* la plus dépendante de la taille du caractère. Dans un deuxième temps, l'utilisation de méthodes informatiques minimise l'erreur humaine. En effet, hormis la procédure de binarisation, le processus de *vérification* développé dans cette recherche ne nécessite pas d'intervention de l'expert ou du chercheur. Ainsi, la quantification des caractères manuscrits et les quantités statistiques (*c.-à-d.* les  $F_{ratio}$  et les rapports de vraisemblance) sont calculés par méthodes informatiques et ne dépendent pas de l'utilisateur.

Cependant il est important de noter que les mesures  $FE_1 - FE_2$  utilisées pour quantifier les échantillons d'écriture dans cette recherche se concentrent sur un type de caractéristique (*c.-à-d.* les distances relatives des pixels), alors qu'un expert est amené à considérer une variété d'éléments (voir section 1.1, Table 1.1).

Le choix de produire des matrices de dissimilarité entre les signaux, au lieu de les quantifier (*p. ex.* avec des séries de Fourier), est discutable. En effet, d'un point de vue simplement pratique les matrices de dissimilarité requièrent plus de temps de calcul (les échantillons doivent être comparés deux à deux) et limitent les possibilités d'analyses statistiques. Toutefois, il a été montré que les signaux des *fonctions d'extraction* n'ont pas de signification réelle pour un expert. Il semble alors plus approprié de transformer les signaux en une mesure procurant plus de sens, comme une mesure de dissimilarité. Ceci étant, ce choix permet de faire un parallèle entre les étapes du processus de *vérification* et le processus *ACE (calcul de dissimilarité – comparaison)*, et donc compenser les limites des *fonctions d'extraction*.

L'analyse de dissimilarité utilisée dans cette recherche (faisant référence aux vecteurs de caractéristiques  $\mathbf{y} = (FE_1, FE_2)$ ) possède l'avantage d'avoir une signification statistique. D'un point de vue théorique les valeurs de ce vecteur sont proches de  $0 = \log(1)$  lorsque  $H_0$  (à savoir l'hypothèse que les dissimilarités mesurées entre les échantillons du matériel *indiciaire* et du matériel de *référence* sont égales) est vraie et s'accroissent lorsque  $H_1$  (à savoir l'hypothèse que les dissimilarités mesurées entre les échantillons du matériel *indiciaire* et du matériel de *référence* ne sont pas égales) est vraie. Ceci en fait un vecteur de caractéristiques idéal pour modéliser les variations *intra-scripteur* et *inter-scripteur*. Les taux d'erreur des rapports de vraisemblance baissent lorsque le nombre d'échantillons augmente. Ces résultats reflètent l'influence de l'échantillonnage sur les propriétés de la méthodologie. Les représentations des PCoA fournissent une indication de la variabilité des échantillons. Ainsi, plus le nombre d'échantillons est faible, plus il est probable que ces échantillons ne traduisent pas la variabilité totale du scripteur. De plus, la méthode d'analyse de matrice de dissimilarité peut confondre une différence de location entre

échantillons (*c.-à-d.* les moyennes) avec leur dispersion (*c.-à-d.* la variation) (voir Warton et al. [121]). Ce phénomène étant plus prononcé quand la taille des échantillons est faible.

Enfin, un rapport de vraisemblance est employé pour évaluer les observations (*c.-à-d.* les  $F_{ratio}$ ). Ceci, afin de supporter l'une des deux hypothèses d'intérêt :  $H_p$ , le scripteur du matériel de *référence* a écrit le document du matériel *indiciare* et  $H_d$ , le scripteur du matériel de *référence* n'est pas celui du matériel *indiciare*.

Le choix du type d'analyse d'évaluation repose sur deux paramètres : celui des hypothèses à tester (*c.-à-d.*  $H_p$  et  $H_d$ ), et celui du modèle (*c.-à-d.* un  $RV$  basé sur un vecteur de  $F_{ratio}$ ). Rappelons que l'objectif de cette recherche est développer, valider et appliquer une nouvelle méthode de quantification et d'évaluation de l'écriture manuscrite. Par conséquent, les rapports de vraisemblance ont été choisis dans le but de donner une valeur statistique qui est connue et acceptée par les chercheurs en sciences forensiques et des experts en documents. Par ailleurs, dans un souci de clarté et de garder une même ligne d'approche tout au long de la recherche, un seul jeu d'hypothèses ( $H_p$  et  $H_d$ ) à été considéré.

Toutefois, il est rappelé tout au long de cette recherche que c'est la combinaison des expériences et des analyses ( $RV$ ,  $F_{ratio}$ , PCoA, hauteur  $H$  et largeur  $W$ ) qui permet d'interpréter les données et de tirer les conclusions. Dans le cas du rapport de vraisemblance ce sont les variations des taux de *vérification* entre *scénarios* qui sont interprétées, plutôt que les taux d'erreur par *scénario*. De plus, il a été montré que tous les résultats issus des expériences 1 à 3 conduisent aux mêmes conclusions. Ainsi, il est admis que les interprétations des rapports de vraisemblance restent fiables (voir section 5.1.2), même si les variables ne respectent pas toujours la distribution d'une loi normale.

Il a été montré que les  $RV$  sont très variables selon la lettre, le *scénario* ou le(s) scripteur(s) étudié(s), sans pour autant avoir de motifs particuliers. Par ailleurs, les taux d'erreur des rapports de vraisemblance n'atteignent pas un taux de 0% dans chaque situation. Ainsi, une large variabilité dans les échantillons d'écriture demeure présente. D'un autre côté, il est d'avis que nous n'avons pas suffisamment de recul et d'expérience sur cette nouvelle méthode de quantification pour que celle-ci soit un outil valide dans le cadre d'une expertise. En effet, tout comme un expert se forme et acquiert des compétences grâce à l'expérience, il est nécessaire de continuer à étudier cette mesure afin d'augmenter les connaissances sur sa validité et sa robustesse, et par ce fait, d'apprendre à connaître et maîtriser les différents biais qu'elle induit. Par conséquent, il serait fortement préjudiciable d'employer cette méthodologie lors d'un procès, En revanche, cette métrique s'avère utile pour étudier des facteurs d'écritures, comme la taille des caractères, ou prouver l'existence d'une variation temporelle de l'écriture. Ce type d'études permet de valider des hypothèses d'intérêts qui peuvent être par la suite utilisées comme éléments dans la méthodologie d'un expert. Cette outil est d'avantage utile pour la recherche en écriture manuscrite, que dans le cadre d'une expertise.

La méthode d'Anderson (Anderson [107], McArdle et Anderson [103]) présentée dans la section 2.4, peut être utilisée en entier pour fournir une *p-valeur* permettant de tester l'hypothèse de l'invariance des caractères manuscrits selon leur taille, ou s'il existe une variation de l'écriture dans un espace temporel. Cette méthode possède l'avantage d'être directement liée à la matrice de dissimilarité. Ceci n'exclut pas l'utilisation de statistique Bayésienne pour analyser les  $F_{ratio}$ . Par exemple, il pourrait être décidé d'inclure d'autres niveaux de variations lors du calcul de rapport de vraisemblance, comme proposé dans Bozza et al. [68] et Taroni et al. [122]. En résumé, comme indiqué tout au long de la

recherche, la méthodologie développée offre plusieurs alternatives d'analyse. Ainsi, dans le cas de futures recherches, le scientifique ou l'expert devrait choisir la méthode d'évaluation en considérant 1) de la validité du modèle, à savoir si les données respectent les hypothèses de construction du modèle, 2) la question à répondre, à savoir ce qui permet de formuler les hypothèses d'intérêt, et 3) l'affinité et le degré de compréhension que le chercheur ou l'expert a de la méthode.

## 5.4 Conclusion

Les différentes étapes de la recherche ont permis de répondre aux objectifs fixés en introduction : de développer, de valider et d'appliquer une nouvelle méthodologie pour quantifier et évaluer l'écriture manuscrite. La méthodologie développée offre une approche originale pour l'étude de l'écriture manuscrite, notamment sur trois points particuliers :

- 1) la fonction d'extraction développée offre une mesure commune à toutes les lettres de l'alphabet. Ainsi les résultats obtenus avec différentes lettres peuvent être comparés plus aisément, ou être intégrés dans le même modèle statistique ;
- 2) le vecteur caractéristique utilisé pour le calcul du rapport de vraisemblance, le vecteur de  $F_{ratio}$ , possède une signification statistique. Cette valeur est aussi une quantité statistique qui permet d'analyser les différences dans les matrices de dissimilarité issus des échantillons ;
- 3) plusieurs outils peuvent être utilisés pour évaluer les observations à différentes étapes du processus de *vérification*. Ainsi, le processus développé offre aux chercheurs ou aux experts tous les outils nécessaires pour évaluer et interpréter les observations effectuées sur des échantillons d'écriture.

De plus, les exemples d'application, à savoir l'étude d'influence de la taille de caractères et de l'influence du temps, montrent l'utilité du processus développé. En effet, bien que les mesures ne soient pas *conventionnelles*, le processus développé peut être utilisé pour quantifier et étudier les facteurs influençant l'écriture manuscrite. Par la suite, le chercheur peut prendre le temps pour un travail plus approfondi et donc recourir à des mesures et des méthodes *conventionnelles*, plus parlantes à un expert en écriture, pour analyser ces facteurs sous un autre angle. Dans le cas de cette recherche, les résultats des exemples d'application ont permis de souligner :

1. qu'il existe des différences entre des caractères écrits à des tailles différentes. En effet, les mesures et observations montrent que les caractères manuscrits ne sont pas reproduits de manière proportionnelle en fonction de leur taille. Ces observations soutiennent celles de Marquis et al. [66] et Rogers et Found [95], et rejettent la théorie de l'invariance de l'écriture en fonction de la taille de Harrison [93]. Ainsi lors de l'étude d'écritures, ou d'une expertise il est conseillé de comparer des échantillons du matériel *indiciaire* et de *référence* qui soient de même dimensions ;
2. qu'il existe des variations naturelles de l'écriture en fonction du temps. De ce fait pour pouvoir capturer avec le plus d'exactitude possible la variabilité de l'écriture d'un scripteur, il est nécessaire d'obtenir plusieurs échantillons sur plusieurs jours d'écriture. En outre, sans que les résultats de cette étude puissent déterminer si la variation sur une longue période est plus importante que celle sur une courte période, il est tout de même conseillé de comparer des échantillons contemporains ;

3. que la prise d'échantillons est une étape très importante dans le cadre de la recherche en écriture manuscrite. En effet, au regard des deux éléments ci-dessus, il est évident que la prise d'échantillons a un impact significatif dans les résultats. Par exemple, le nombre de sessions d'écriture, l'espace temporel qui les sépare, et le type de caractères rédigés par les scripteurs (*c.-à-d.* l'écriture *script* et le type de caractères *normaux/grands*) sont des facteurs qui ont une influence sur les caractéristiques de l'écriture. Or, dans la majorité des publications du domaine de l'écriture manuscrite, l'étape d'échantillonnage est succincte ou incomplète. Il est alors bien souvent impossible au lecteur de savoir quel protocole a été utilisé dans ces recherches, et par conséquent de juger le bien-fondé des résultats et des conclusions. Ainsi il est important que l'étape d'échantillonnage fasse l'objet de plus d'attention dans les publications scientifiques ;
4. que la variabilité au sein d'un groupe de scripteur peut être très haute. Ainsi les recherches qui ne se focalisent que sur les taux globaux calculés sont susceptibles de ne pas observer les tendances individuelles des scripteurs. Par conséquent ces études prennent le risque de tirer des conclusions trop générales qui ne tiennent pas en compte la variabilité individuelle de chaque scripteur.

En conclusion l'étude montre, d'une part l'intérêt d'analyser l'écriture manuscrite avec des méthodes *informatiques*, et d'autre part la complexité de l'écriture manuscrite et de ses facteurs. En effet, les mesures *informatiques* ayant pour but de quantifier l'écriture manuscrite, comme celle développée dans cette recherche, permettent d'analyser avec rapidité et fiabilité de nombreux échantillons d'écriture. Ainsi combiné avec une approche probabiliste pour vérifier des hypothèses, cette démarche permet de sélectionner l'information pertinente contenue dans ces mesures, et donc d'étudier de manière scientifique les caractéristiques de l'écriture manuscrite et ses facteurs. Par exemple, une telle approche a permis de suggérer plusieurs recommandations lors de la prise d'échantillons dans le cadre d'une expertise ou de recherche scientifique, et de vérifier l'hypothèse concernant l'invariance de l'écriture en fonction de la taille des caractères. Ainsi le processus de quantification et d'évaluation de l'écriture manuscrite développé dans cette recherche apporte un nouvel outil pour soutenir la recherche empirique et les experts dans le domaine de l'écriture manuscrite.

---

## Bibliographie

- [1] N. Morris. *Forensic Handwriting Identification : Fundamental Concepts and Principles*. Academic Press, 2000.
- [2] R.A. Huber et A.M. Headrick. *Handwriting Identification : Facts and Fundamentals*. CRC Press, 1999.
- [3] R. Hanlen, P. Manzollilo, R. Muehlberger, et G. Sprey. Survey of handwriting habits areas used by forensic document examiners : Degree of use dans discriminatory power. *Journal of the American Society of Questioned Document Examiners*, 2(1) : 45–50, 1999.
- [4] V-L. M. Cormeau, V. Distrain, J. Toussaint, et E. Bidaine. Normes de vitesse d'écriture. Étude statistique de 1844 écoliers belges de 6 à 13 ans. *Psychologica Belgica*, 10 :247–263, 1970.
- [5] C. E. Bohn. Fundamentals pertaining to signature exemplars. *American Academy of Forensic Sciences*, Dallas, 1974.
- [6] R.A. Huber et A.M. Headrick. The identification process. *Proceedings of the 11th meeting of the International Association of Forensic Sciences*, Vancouver, 1987.
- [7] R.A. Huber. Expert witnesses. *The Criminal Law Quarterly*, 3(2), 1959.
- [8] D. Ashbaugh. *Quantitative-Qualitative Friction Rigde Analysis*. CRC Press, Boca Raton, USA, 1999.
- [9] John R. Vanderkolk. Analysis, Analysis, Comparison, Evaluation, and Verification. In *Forensic Comparative Science*, chapter 6, pages 89–102. Academic Press, San Diego, USA, 2009.
- [10] L. Tierney. Analysis, Comparison, Evaluation, and Verification (ACE-V). In *Encyclopedia of Forensic Sciences*, pages 69–73. Academic Press, Waltham, USA, 2013.
- [11] A. W. Ellis. Splis of the Pen. *Visible Language*, 13(3) :265–282, 1979.
- [12] J. J. Brault et R. Plamondon. A complexity measure of handwritten curves : modeling of dynamic signature forgery. *IEEE Transactions on Systems, Man and Cybernetics*, 23(2) :400–413, 1993.
- [13] A. J. W. M. Thomassen et H-L. H. M. Teulings. *The Development of Handwriting, The Psychology of Written Language*. John Wiley & Sons, New York, USA, 1983.

- [14] O. Hilton. How individual habits are personal writing habits? *Journal of Forensic Sciences*, 28(3) :683–685, 1983.
- [15] A.S. Osborn. *Questioned Documents : A Study of Questioned Documents with an Outline of Methods by which the Facts May be Discovered and Shown*. Lawyers' Co-operative Publishing Company, 1910.
- [16] O. Hilton. *Scientific examination of questioned documents*. Callaghan, 1956.
- [17] H. Hardy et W. Fagel. Methodological aspects of handwriting identification. *Journal of Forensic Document Examination*, 8 :33–69, 1995.
- [18] B. Found et D. K. Rogers. Contemporary issues in forensic handwriting examination. A discussion of key issues in the wake of the starzecpyzel decision. *Journal of Forensic Document Examination*, 8 :1–31, 1995.
- [19] M. D. Risinger et M. J. Saks. Science and nonscience in the court : Daubert meets handwriting identification expertise. *Iowa Law Review*, 82 :21–74, 1996.
- [20] N. Rudin et K. Inman. The shifty paradigm, part I. *The CAC News*, 4th Quarter : 13–16, 2005.
- [21] N. Rudin et K. Inman. The shifty paradigm, part II. *The CAC News*, 1st Quarter : 16–18, 2006.
- [22] H. Poincaré, G. Darboux, et P. Appell. Rapport de MM. les experts Darboux, Appell et Poincaré. In *Affaire Dreyfus ; La Revision du Procès de Rennes ; Enquête de la Chambre criminelle de la Cour de cassation*, pages 3 :500–600. Ligue Française pour la Défense des Droits de l'Homme et du Citoyen, Paris, France, 1908–1909.
- [23] F. Taroni, C. Champod, et P. Margo. Forerunners of Bayesianism in Early Forensic Science. *Jurimetrics Journal*, 38 :183–200, 1998.
- [24] F. Lee. Handwriting identification evidence in the post-Daubert world. *Journal of the American Society of Questioned Document Examiners*, 1(1) :67–68, 1994.
- [25] M. D. Risinger, M. P. Denbeaux, et M. J. Saks. Exorcism of Ignorance as a Proxy for Rational Knowledge : The Lessons of Handwriting Identification "Expertise". *University of Pennsylvania Law Review*, 137 :731–792, 1989.
- [26] *Federal Rules of Evidence*. Legal Information Institute, Cornell Law School, New York, 2010.
- [27] *Daubert vs. Merrel Dow Pharmaceuticals Inc.*, 509 U.S. 579, 113 S.Ct. 2786, 1992.
- [28] O. Galbraith, C. S. Galbraith, et N. G. Galbraith. The "principle of the drunkard's search" as a proxy for scientific analysis. *International Journal of Forensic Document Examiners*, 1 :7–17, 1995.
- [29] A. A. Moenssens. Handwriting Identification in the Post-Daubert World. *UMKC Law Review*, 66 :251–343, Winter 1997.
- [30] M. D. Risinger, M. P. Benbeaux, et M. J. Saks. Brave new "post-Daubert world" - a reply to professor Moenssens. *Steton Hall Review*, 29 :405–485, 1998.

- [31] *United States vs. Saelee*, 162 F.Supp.2d 1097, United States District Court, D. Alaska., 2001.
- [32] M. D. Risinger. Handwriting Identification. In D. L. Faigman, D. H. Kaye, M. J. Saks, J. Sanders, et E. K. Cheng, editors, *Modern Scientific Evidence : The Law and Science of Expert Testimony*, volume 4, pages 377–524. Thompson/West Eagan, Minnesota, USA, 2007.
- [33] H-L. H. M. Teulings et A. J. W. M. Thomassen. Computer-aided analysis of handwriting movements. *Visible Language*, 13(3) :218–231, 1979.
- [34] M. Arivahagan, H. Srinivasan, et S. N. Srihari. A statistical approach to line segmentation in handwritten documents. *Proceedings of the International Society for Optics and Photonics*, 2007.
- [35] K. S. Siddhart, R. Dhir, et R. Rani. Handwritten gurmukhi numeral recognition using different feature sets. *International Journal of Computer Applications*, 28(2), 2011.
- [36] V. N.d More et P. P. Rege. Devanagari handwritten numerals identification based on zenike moments. *IEEE Region 10 Conference*, pages 1–6, 2008.
- [37] H. Singh et R. K. Sharma. Moment in online handwritten character recognition. *Proceedings of the National Conference on Challenges & Opportunities in Information Technology*, 2007.
- [38] G. Abandah, K. Younis, et M. Khedher. Handwritten arabic character recognition using multiple classifiers based on letter form. *Proceedings of the 5th International Conference on Signal Processing, Pattern Recognition & Applications*, 2008.
- [39] G. Abandah et N. Anssari. Novel moment features extraction for recognizing handwritten arabic letters. *Journal of Computer Science*, 5(3) :226–232, 2009.
- [40] I. El-Fegh, Z. S. Zubi, A. A. Elrowayati, et F. A. El-Mouadib. Handwritten arabic words recognition using multi-layer perception and zenike moments. *Proceeding of the 10th International Conference on Evolutionary Computing*, pages 46–51, 2009.
- [41] I. El-Fegh, F. Elmahjoub, B. Alsawady, et A. Baiou. Offline handwritten arabic words recognition using zernike moments and hidden markov models. *Proceedings of the International Conference on Computer Applications and Industrial Electronics*, pages 165–168, 2010.
- [42] M. Thungamani, P. K. Ramakhanth, P. Keshava, et R. K Shravani. Off-line handwritten kannada text recognition using support vector machine using zenike moments. *International Journal of Computer Science and Network Security*, 11(7), 2011.
- [43] K. Bagheri Naoparast et A. Broumandnia. Persian handwritten word recognition using zernike and fourier-mellin moments. *Proceeding of the 5th International Conference on Science of Electronic, Technologies of Information and Telecommunications*, 2009.

- [44] B. Verma, M. Blumenstein, et M. Gosh. A novel approach for structural feature extraction : Contour vs. direction. *Pattern Recognition Letters*, 25 :975–988, 2004.
- [45] A. Buquet. Vérification automatisée des écritures et des signatures. *Revue Internationale de Police Criminelle*, 483 :10–18, 2000.
- [46] R. Plamondon et G. Lorette. Automatic signature verification and writer identification — the state of the art. *Pattern Recognition*, 22(2) :107–131, 1989.
- [47] R. Plamondon et S. N. Srihari. On-Line and Off-Line Handwriting Recognition : A Comprehensive Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1) :63–84, 2000.
- [48] Ø-D. Trier, A. K. Jain, et T. Taxt. Feature extraction methods for character recognition—A survey. *Pattern Recognition*, 29(4) :641–662, 1996.
- [49] A. Belaïd et Y. Belaïd. *Reconnaissance des Formes*. InterÉdition, Paris, France, 1992.
- [50] P. A. Devijver et J. Kittler. *Pattern recognition : A statistical approach*. Prentice Hall, 1982.
- [51] E. Bunge. The role of Pattern Recognition in Forensic Science : An Introduction to Methods. In *Police Research in the Federal Republic of Germany (Kube E., Störzer H.)*. Springer-Verlag, Berlin, Germany, 1991.
- [52] S. Pecharromán-Bablás. *Reconocimiento de escritor independiente de texto basado en características de textura*. PhD thesis, Escuela Politécnica Superior, Universidad Autónoma de Madrid, 2007.
- [53] S. N. Srihari, S-H. Cha, H. Arora, et S. Lee. Individuality of Handwriting. *Journal of Forensic Sciences*, 47(4) :1–17, 2002.
- [54] A. Vinciarelli. A survey on off-line Cursive Word Recognition. *Pattern Recognition*, 35(7) :1433–1446, 2002.
- [55] L. R. Taylor et H. B. Chandler. A system for handwriting classification. *Journal of Forensic Sciences*, 32(6) :1775–1781, 1987.
- [56] A. Bensefia, T. Paquet, et L. Heutte. A writer identification and verification system. *Pattern Recognition Letters*, 26(13) :2080–2092, 2005.
- [57] G. Gupta et A. McCabe. A review of dynamic handwritten signatures verification. Technical report, James Cook University, 1997.
- [58] S. N. Srihari, C. Huang, et H. Srinivasan. On the Discriminability of the Handwriting of Twins. *Journal of Forensic Sciences*, 53(2) :430–446, 2008.
- [59] M. J. Saks. Commentary on : Srihari SN, Cha S-H, Arora H, Lee S. Individuality of handwriting . *Journal of Forensic Sciences*, 48(4) :1–3, 2003.
- [60] S. N. Srihari. Author’s Response to Commentary by Saks. *Journal of Forensic Sciences*, 44(8) :919–920, 2003.

- [61] *United States vs. Carmen Gricco – Memorandum to motion for exclusion of handwriting opinion evidence.*, No. 01-09, Eastern District Court of Pennsylvania, 2002.
- [62] *Robert E. Pettus vs. United States*, No. 08-CF-1361, District of Columbia Court of Appeals, 2012.
- [63] D. Harrison, T. M. Burkes, et D. P. Selger. Handwriting Examination : Meeting the Challenges of Science and the Law. *Forensic Science Communications*, 11(4), 2009.
- [64] C. King et L. Logan. Preparing for a Daubert Hearing : A Study Guide for Forensic Document Examiners. *69th Annual Meeting of the American Society of Questioned Document Examiners*, Philadelphia, Pennsylvania, 2011.
- [65] R. Marquis, M. Schmittbuhl, W. Mazzela, et F. Taroni. Quantification of the shape of handwritten characters : a step to objective discrimination between writers based on the study of the capital character O. *Forensic Science International*, 150 :23–32, 2005.
- [66] R. Marquis, F. Taroni, S. Bozza, et M. Schmittbuhl. Quantitative characterization of morphological polymorphism of handwritten characters loops. *Forensic Science International*, 164 :211–220, 2006.
- [67] R. Marquis, F. Taroni, S. Bozza, et M. Schmittbuhl. Size influence on shape of handwritten characters loops. *Forensic Science International*, 172 :10–16, 2007.
- [68] S. Bozza, F. Taroni, R. Marquis, et M. Schmittbuhl. Probabilistic evaluation of handwriting evidence : likelihood ratio for authorship. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 57(3) :329–341, 2008.
- [69] F. Taroni, R. Marquis, M. Schmittbuhl, A. Biedermann, A. Thiéry, et S. Bozza. The use of the likelihood ratio for evaluative and investigative purposes in comparative forensic handwriting examination. *Forensic Science International*, 214(1–3) :189–194, 2012.
- [70] M. G. Longstaff et R. A. Heath. The influence of tremor on handwriting performance under conditions of low and intermediate physical stress. *Journal of Forensic Document Examination*, 13 :25–44, 2000.
- [71] F. Aşıcıoğlu et N. Turan. Handwriting changes under the effect of alcohol. *Forensic Science International*, 132(3) :201–210, 2003.
- [72] V. Pervouchine et G. Leedham. Extraction and analysis of forensic document examiner features used for writer identification. *Pattern Recognition*, 40(3) :1004–1013, 2007.
- [73] A. Thiéry, R. Marquis, et I. Montani. Statistical evaluation of the influence of writing postures on on-line signatures. Study of the impact of time. *Forensic Science International*, 230(1–3) :107–116, 2013. {EAFS} 2012 6th European Academy of Forensic Science Conference The Hague, 20-24 August 2012.
- [74] C. Li, N. Poon, W. Fung, C. Yang, et M. Phil. Individuality of handwritten arabic numerals in local population. *Forensic Science International*, 50(1) :185–191, 2005.

- [75] M. Rugayah, M. H. Nursakinah, Wing K. Y., Umi K. A., et M. Ismail. Statistical examination of common characteristics for disguised handwriting amongst Malaysian. *Health and the Environment Journal*, 1(2) :30–37, 2010.
- [76] A. A. Brink, Ralph Niels, R. A. van Betenburg, et C. E. van den Heuvel. Towards robust writer verification by correcting unnatural slant. *Pattern Recognition Letters*, 32(3) :449–457, 2011.
- [77] S-H. Cha et S. N. Srihari. Multiple feature integration for writer verification. *Proceeding of the 7th International Workshop on Frontiers in Handwriting Recognition*, Amsterdam, 2000.
- [78] W. Tian et Y. Qiao. Off-line Chinese signature verification based on optimal matching of projection profiles. *Proceeding of the 6th World Congress on Intelligent Control and Automation*, 2006.
- [79] E. Zois et V. Anastassopoulos. Fusion of correlated decisions for writer verification. *Pattern Recognition*, 33(1) :47–61, 2000.
- [80] J. Sung, S-Y. Bang, et S. Choi. A bayesian network classifier and hierarchical gabor features for handwritten numeral recognition. *Pattern Recognition Letters*, 27 :66–75, 2006.
- [81] B. Helli et M. E. Moghaddam. A text-independent persian writer identification based on feature relation graph (FRG). *Pattern Recognition*, 43 :2199–2209, 2009.
- [82] H. E. S Said, T. N. Tan, et K. D. Baker. Writer identification from non-uniformly skewed handwriting images. *Proceeding of the 9th British Conference on Machine Vision*, 1998.
- [83] H. E. S Said, T. N. Tan, et K. D. Baker. Personal identification based on handwriting. *Pattern Recognition*, 33(1) :149–160, 2000.
- [84] Y. Zhu, T. Yang, et Y. Wang. Biometric personal identification based on handwriting. *Proceedings of the 15th International Conference on Pattern Recognition*, Barcelona, Spain, 2000.
- [85] K. Franke et M. Köppen. A computer-based system to support forensic studies on handwritten document. *International Journal on Document Analysis and Recognition*, 3(4) :218–231, 2001.
- [86] M. Tapiador et J. A. Sigüenza. Writer identification method based on forensic knowledge. *Biometric Authentication*, 3072 :555–561, 2004.
- [87] F. Shahabi et M. Rahmati. A new method for writer identification of handwritten Farsi documents. *Proceeding of the 10th International Conference on Document Analysis and Recognition*, Barcelona, Spain, 2009.
- [88] E. Zois et V. Anastassopoulos. Morphological waveform coding for writer identification. *Pattern Recognition*, 33(3) :385–398, 2000.
- [89] A. K. Muda, S. M. Shamsuddin, et M. Darus. Discretization of integrated moment invariants for writer identification. *Proceeding of the 4th International Conference on Advances in Computer Science and Technology*, Langkawi, Malaysia, 2008.

- [90] S. N. Srihari et G. Leedham. A survey of computer methods in forensic document examination . *Proceedings of 11th Conference International on Graphonomics Society*, Melbourne, Australia, 2003.
- [91] K. Franke, L. Schomaker, C. Veenhuis, C. Taubenheim, I. Guyon, L. Vuurpijl, M. van Erp, et G. Zwarts. WANDA : a generic framework applied in forensic handwriting analysis and writer identification. In A. Abraham, M. Köppen, et K. Franke, editors, *Design and application of hybrid intelligent systems*, pages 927–938. IOS Press, Amsterdam, The Netherlands, 2003.
- [92] S. N. Srihari, B. Zhang, C. Tomai, S. Lee, Z. Shi, et C. Shin. A system for handwriting matching and recognition. *Proceedings of the Symposium on Document Image Understanding Technology*, Greenbelt (MD), United-States, 2003.
- [93] W. R. Harrison. *Suspect documents : their scientific examination*. Praeger, 1958.
- [94] J. P. Wann et I. Nimmo-Smith. Evidence against the relative invariance of timing in handwriting. *Quart J, Exp. Psychol.*, 42 :105–119, 1990.
- [95] D. K. Rogers et B. Found. The Objective Measurement of Spatial Invariance in Handwriting. In M. L. Simner, C. G. Leedham, et A. J. W. M. Thomassen, editors, *Handwriting and Drawing Research : Basic and Applied Issues*. IOS Press, 1996.
- [96] J. Liu, Q. Cheng, Z. Zheng, et M. Qian. A DTW-based probability model for speaker feature analysis and data mining. *Pattern Recognition Letters*, 23(11) : 1271–1276, 2002.
- [97] A. Piyush Shanker et A. N. Rajagopalan. Off-line signature verification using DTW. *Pattern Recognition Letters*, 28(12) :1407–1414, September 2007.
- [98] M Faundez-Zanuy. On-line signature recognition based on VQ-DTW. *Pattern Recognition*, 40(3) :981–992, March 2007.
- [99] L. Rabiner et B-H. Juang. *Fundamentals of speech recognition*. Prentice Hall, New Jersey, USA, 1st edition, 1993.
- [100] S. Theodoridis et K. Koutroumbas. *Pattern recognition*. Academic Press, 2nd edition, 2003.
- [101] M. Müller. *Information Retrieval for Music and Motion*. Springer-Verlag, New Jersey, USA, 2007.
- [102] T. Giorgino. Computing and Visualizing Dynamic Time Warping Alignments in R : The dtw Package. *Journal of Statistical Software*, 31(7) :1–24, 8 2009.
- [103] Brian H. McArdle et Marti J. Anderson. Fitting Multivariate Models To Community Data : a Comment on Distance-Based Redundancy Analysis. *Ecology*, 82(1) : 290–297, 2001.
- [104] James Curran et Danny Chang. *dafs : Data analysis for forensic scientists*, 2012.
- [105] J.F. Hair et R.E. Anderson. *Multivariate data analysis*. Prentice Hall Higher Education, 2010.

- [106] N. H. Timm. *Applied Multivariate Analysis*. Springer texts in statistics. Springer-Verlag, New York, USA, 2002.
- [107] M. J. Anderson. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1) :32–46, 2001.
- [108] J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4) :325–338, 1966.
- [109] J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, G. L. Simpson, P. Solymos, M. H. Stevens, et H. Wagner. *vegan : Community Ecology Package*, 2012. URL <http://cran.r-project.org/package=vegan>. R package version 2.0-4.
- [110] F. Taroni, S. Bozza, A. Biedermann, P. Garbolino, et C. Aitken. *Data Analysis in Forensic Science : A Bayesian Decision Perspective*. Statistics in Practice. John Wiley & Sons, 2010.
- [111] M. Lavine et M. J. Schervish. Bayes Factors : What They Are and What They Are Not. *The American Statistician*, 53(2) :119–122, 1999.
- [112] A. Gelman, J. B. Carlin, H. S. Stern, et D. B. Rubin. *Bayesian Data Analysis*. Texts in Statistical Science. Chapman & Hall/CRC, 2nd edition, 2004.
- [113] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer-Verlag, New York, USA, 1985.
- [114] A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, et T. Hothorn. *mvtnorm : Multivariate Normal and t Distributions*, 2014. URL <http://cran.r-project.org/package=mvtnorm>. R package version 0.9-99991.
- [115] A. Genz et F. Bretz. *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg, Germany, 2009.
- [116] S. Dray et A-B. Duffour. The ade4 Package : Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22(4), 2007.
- [117] W. S. Togerson. Multidimensional Scaling : I Theory and Applications. *Psychometrika*, 17(4) :401–419, 1952.
- [118] I. Borg et Groenen P. *Modern Multidimensional Scaling -Theory and applications*. Springer Series in Statistics. Springer-Verlag, New York, USA, 2nd edition, 2005.
- [119] R Core Team. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL <http://www.r-project.org>.
- [120] L. Lebart, M. Piron, et A. Morineau. *Statistique exploratoire multidimensionnelle : visualisations et inférences en fouille de données*. Sciences Sup. Mathématiques. Dunod, Paris, France, 4th edition, 2006.
- [121] D. I. Warton, S. T. Wright, et Y. Wang. Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, 3 : 89–101, 2012.

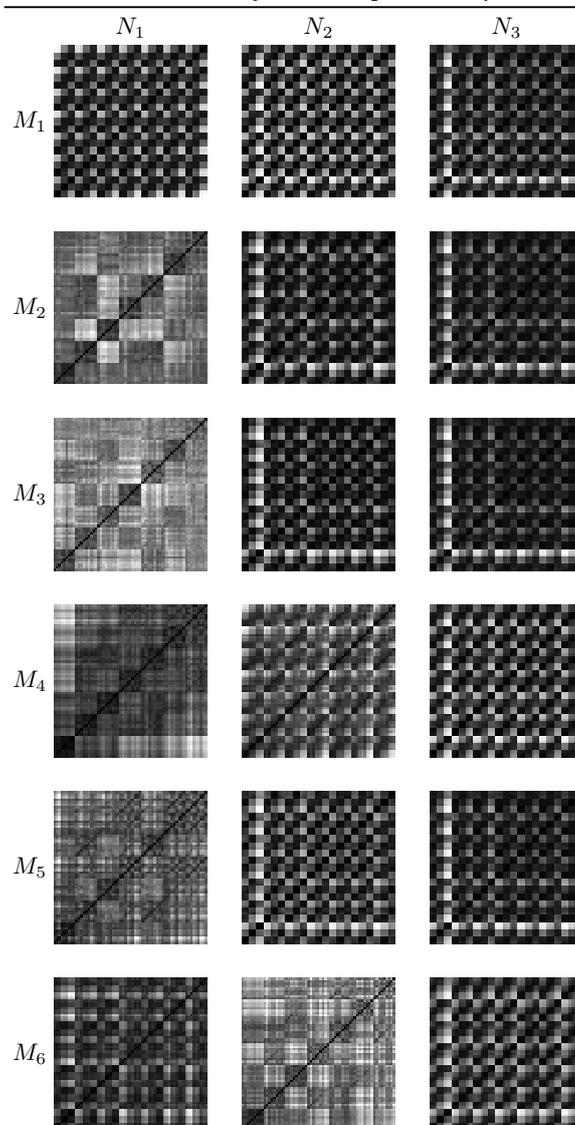
- [122] F. Taroni, R. Marquis, M. Schmittbuhl, A. Biedermann, A. Thiéry, et S. Bozza. Bayes factor for investigative assessment of selected handwriting features. *Forensic Science International*, 242 :265–273, 2014.





## 5.5 Expérience 1

## Matrices de distances

Matrices de distances – fonction de pas DTW *symmetric1*.FIGURE 5.1: Illustration des matrices de distance pour les 18 combinaisons de *fonctions d'extraction* et pour la *fonction de pas DTW symmetric1*.

**Matrices de distances – fonction de pas DTW *symmetricP0*.**

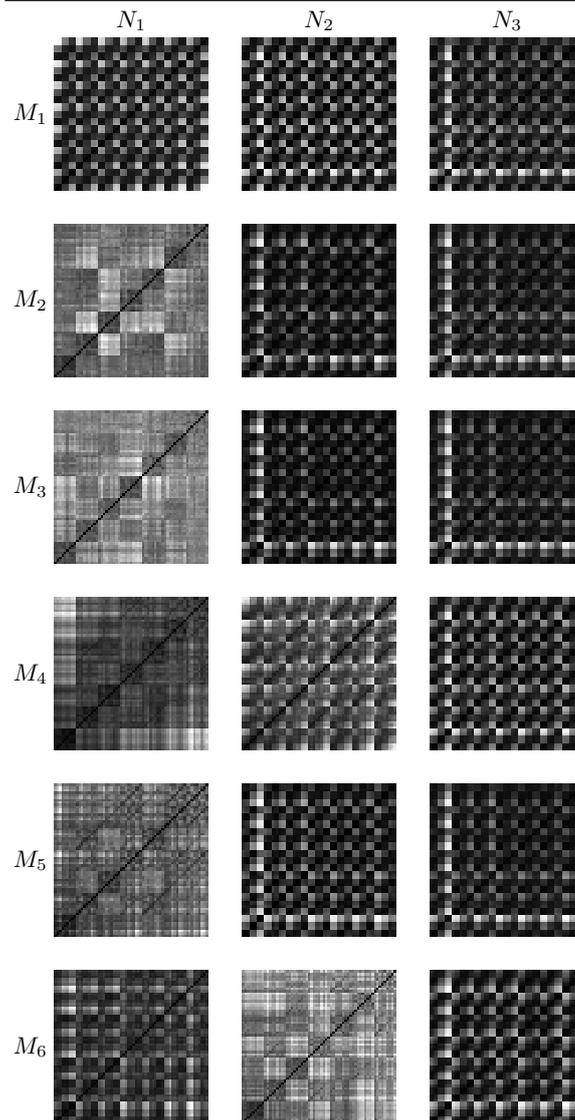


FIGURE 5.2: Illustration des matrices de distance pour les 18 combinaisons de *fonctions d'extraction* et pour la *fonction de pas DTW symmetricP0*.

**Matrices de distances – fonction de pas DTW *symmetricP1*.**

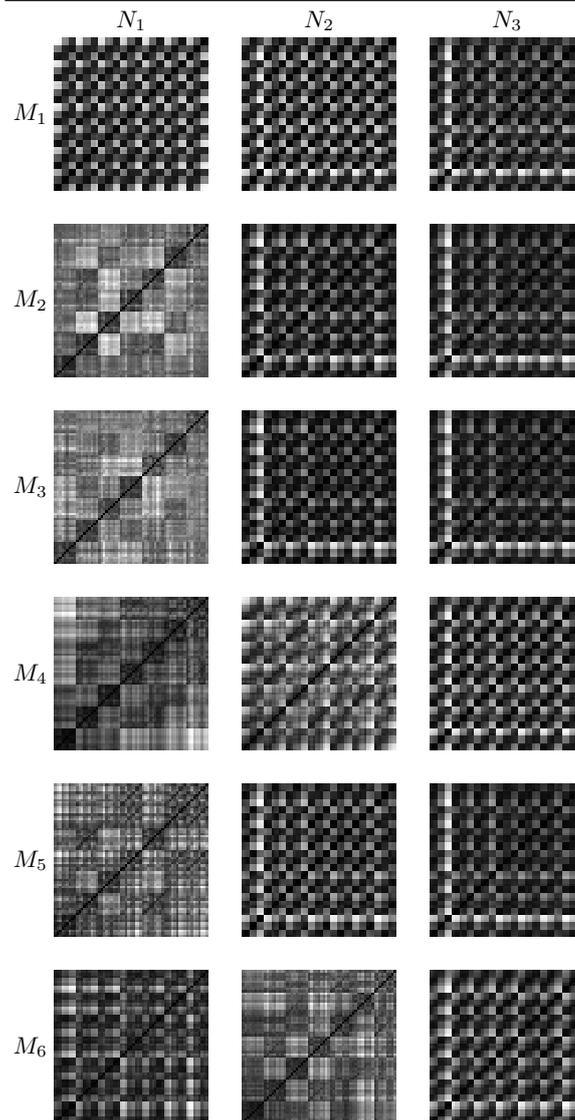


FIGURE 5.3: Illustration des matrices de distance pour les 18 combinaisons de *fonctions d'extraction* et pour la *fonction de pas DTW symmetricP1*.

**Matrices de distances – fonction de pas DTW *symmetricP2*.**

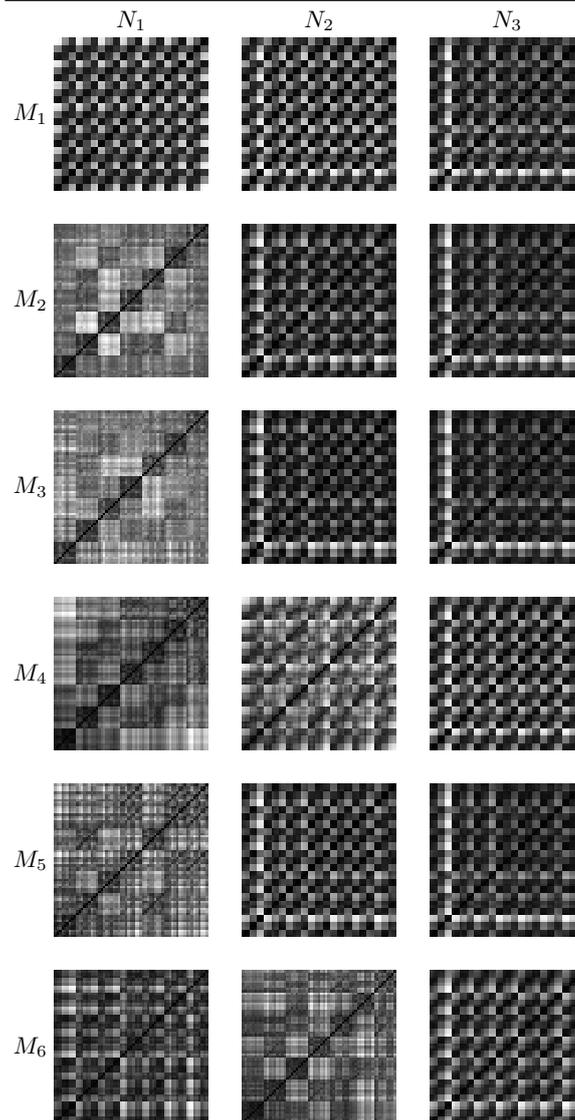


FIGURE 5.4: Illustration des matrices de distance pour les 18 combinaisons de *fonctions d'extraction* et pour la *fonction de pas DTW symmetricP2*.

**Matrices de distances – fonction de pas DTW *symmetricP05*.**

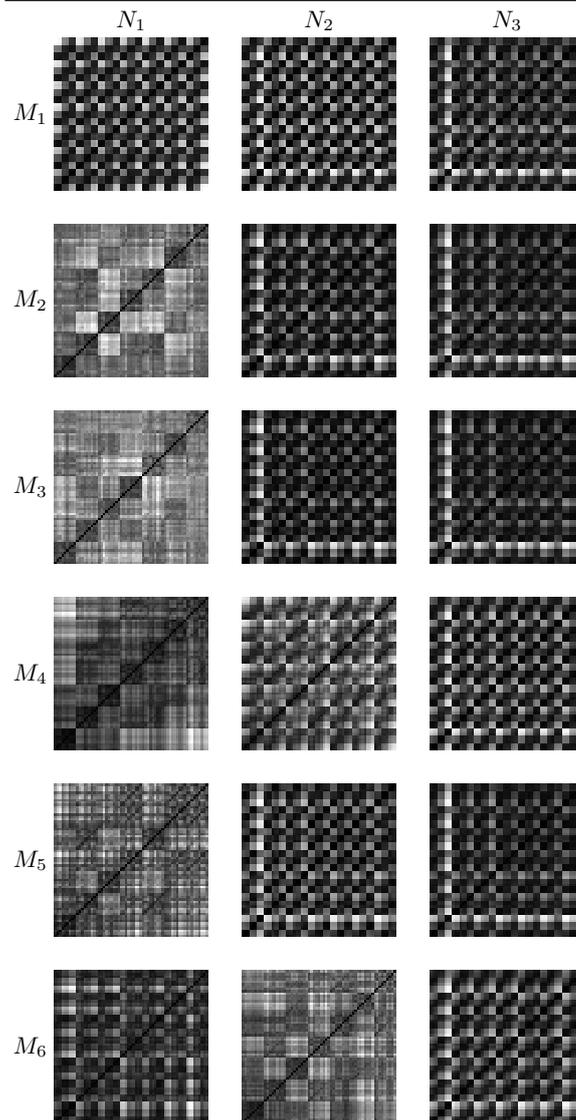


FIGURE 5.5: Illustration des matrices de distance pour les 18 combinaisons de *fonctions d'extraction* et pour la *fonction de pas DTW symmetricP05*.

**Matrices de distances – fonction de pas DTW typeIbs.**

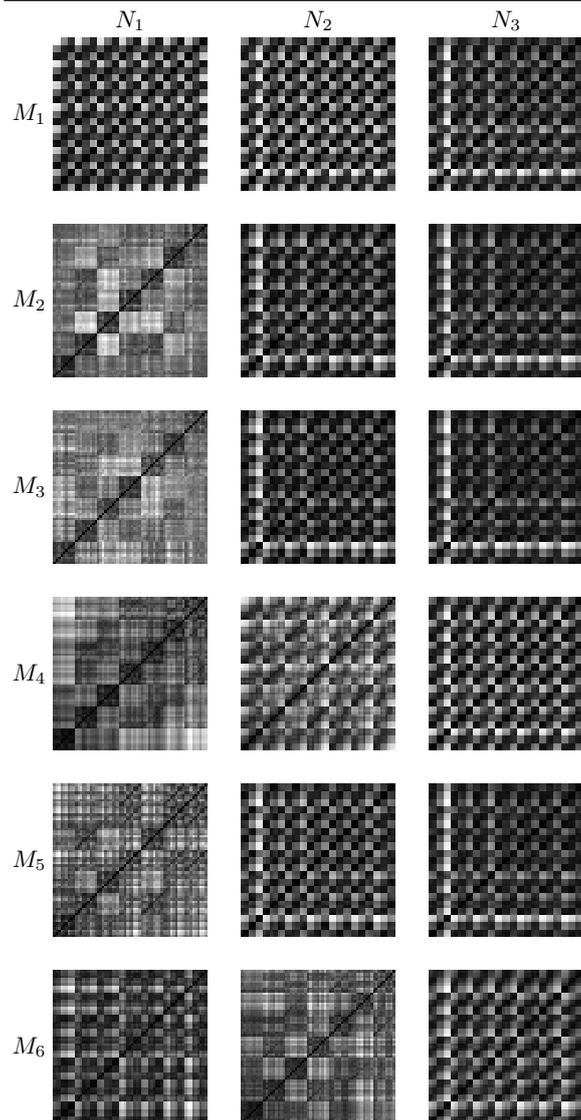


FIGURE 5.6: Illustration des matrices de distance pour les 18 combinaisons de *fonctions d'extraction* et pour la *fonction de pas DTW typeIbs*.

**Matrices de distances – fonction de pas DTW typeIIa.**

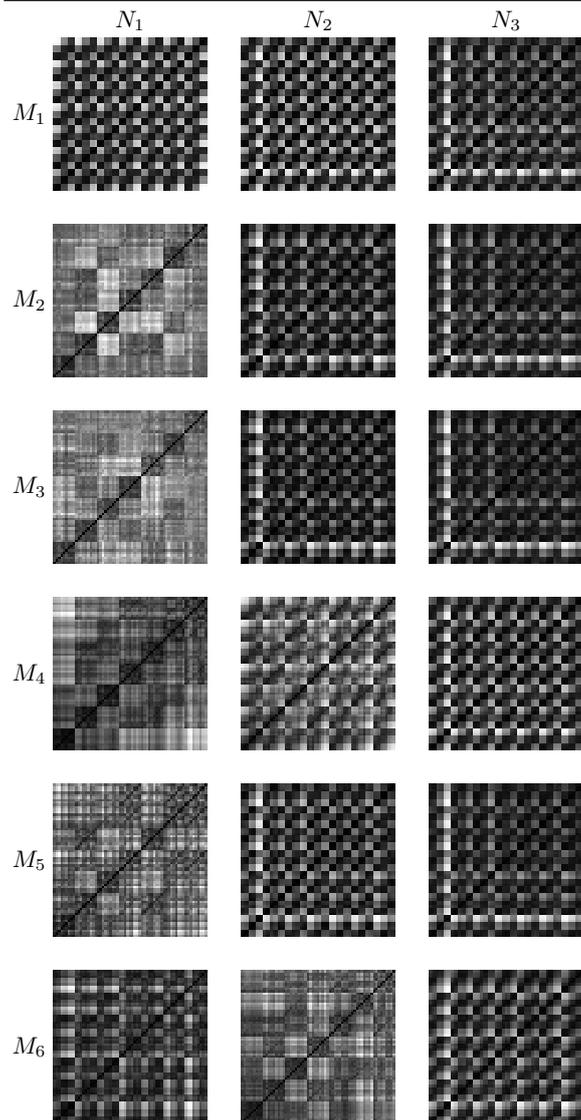


FIGURE 5.7: Illustration des matrices de distance pour les 18 combinaisons de *fonctions d'extraction* et pour la *fonction de pas DTW typeIIa*.

# Analyse par coordonnées principales

**Représentation PCoA – fonction de pas DTW *symmetric1*.**

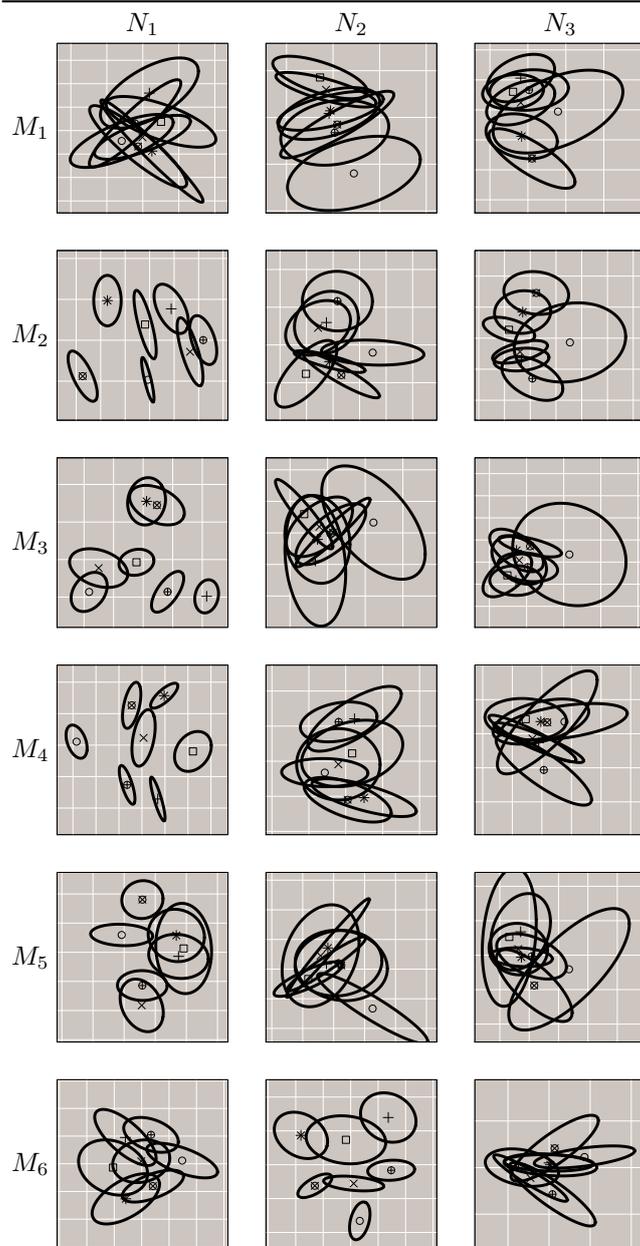


FIGURE 5.8: Représentations PCoA des matrices de dissimilarité pour la *fonction de pas DTW *symmetric1**.

Représentation PCoA – fonction de pas DTW *symmetricP0*.

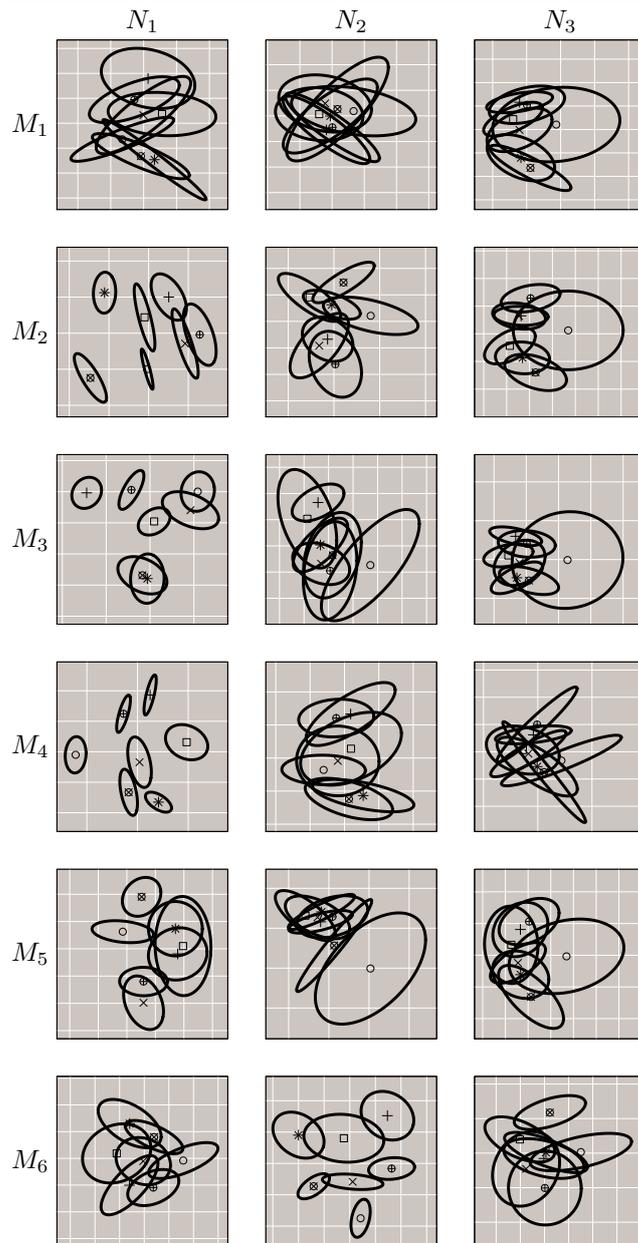


FIGURE 5.9: Représentations PCoA des matrices de dissimilarité pour la *fonction de pas* DTW *symmetricP0*.

Représentation PCoA – fonction de pas DTW *symmetricP1*.

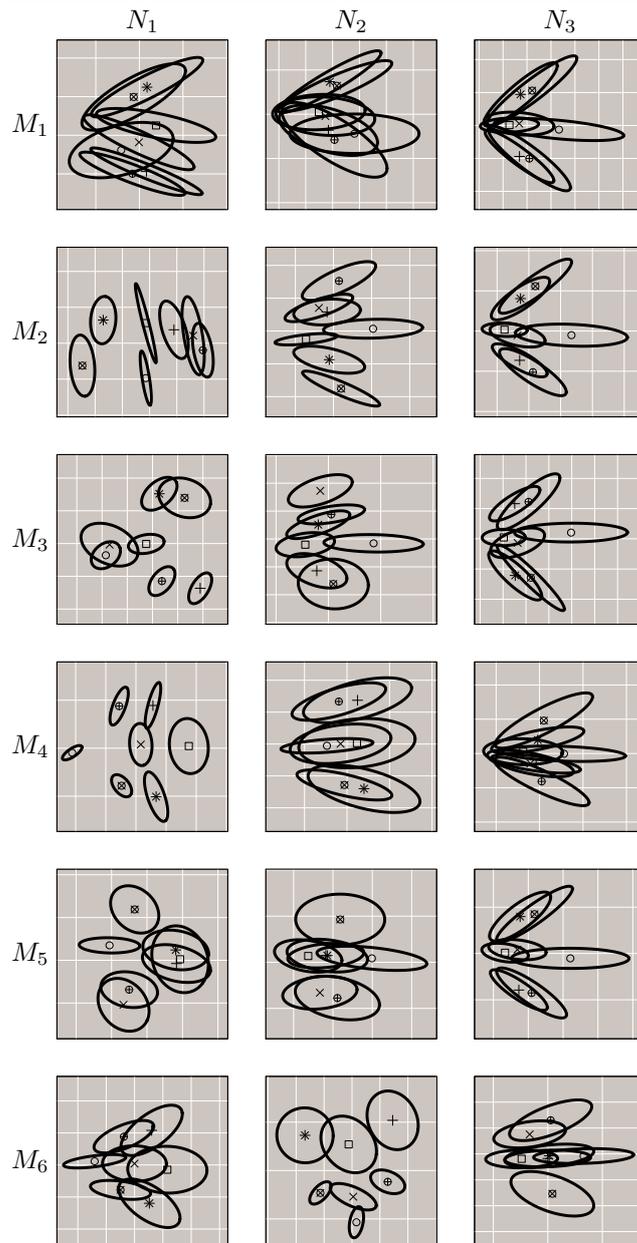


FIGURE 5.10: Représentations PCoA des matrices de dissimilarité pour la *fonction de pas* DTW *symmetricP1*.

Représentation PCoA – fonction de pas DTW *symmetricP2*.

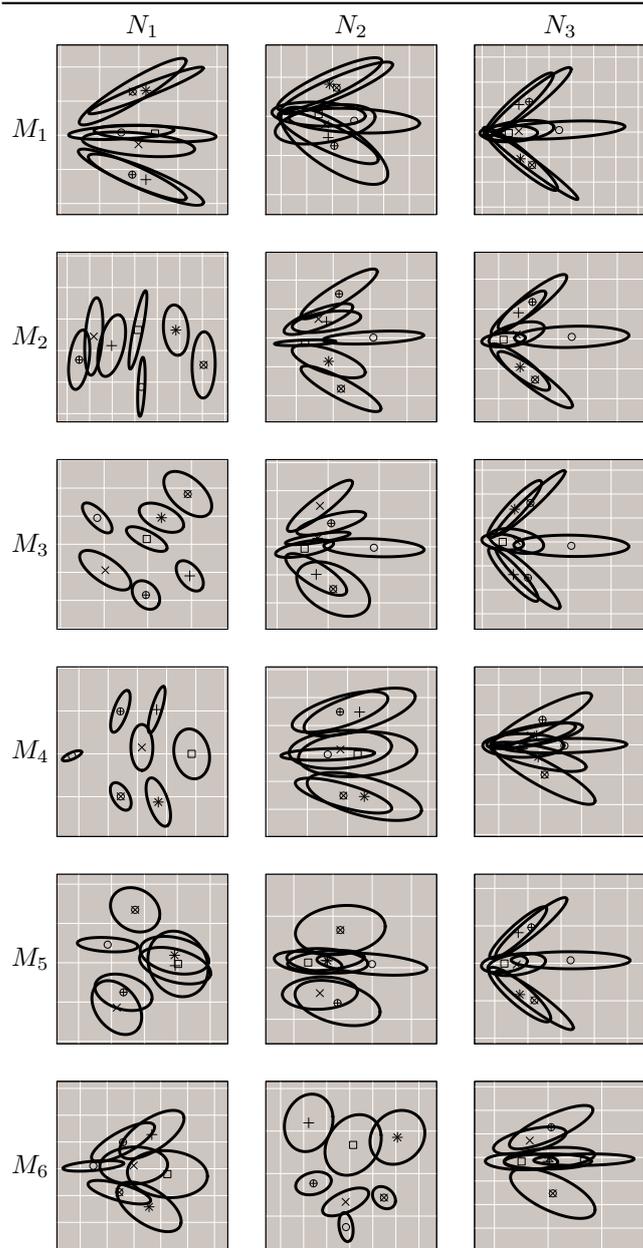


FIGURE 5.11: Représentations PCoA des matrices de dissimilarité pour la *fonction de pas* DTW *symmetricP2*.

Représentation PCoA – fonction de pas DTW *symmetricP05*.

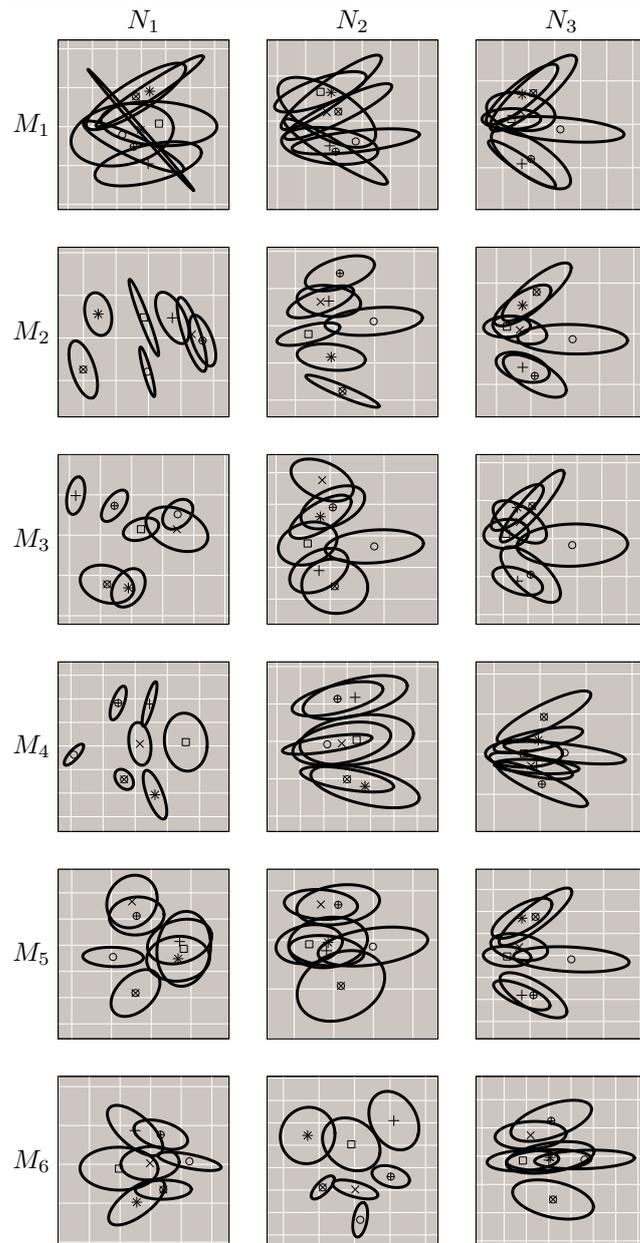


FIGURE 5.12: Représentations PCoA des matrices de dissimilarité pour la *fonction de pas* DTW *symmetricP05*.

Représentation PCoA – fonction de pas DTW *typeIbs*.

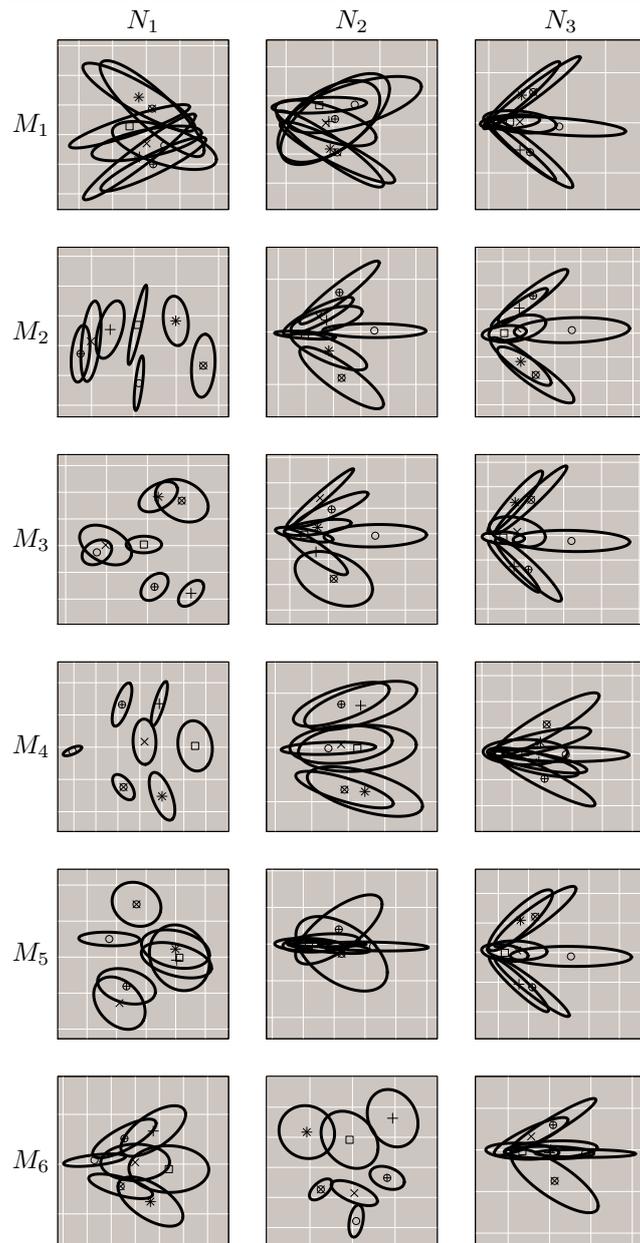


FIGURE 5.13: Représentations PCoA des matrices de dissimilarité pour la *fonction de pas* DTW *typeIbs*.

**Représentation PCoA – fonction de pas DTW typeIIa.**

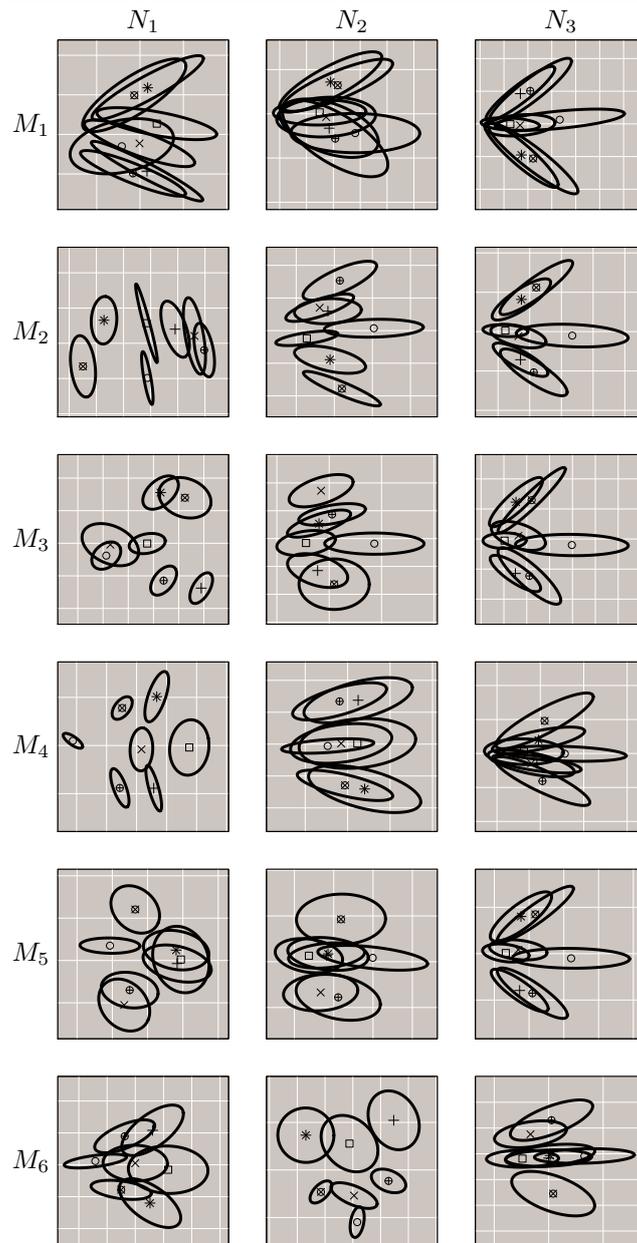


FIGURE 5.14: Représentations PCoA des matrices de dissimilarité pour la *fonction de pas DTW typeIIa*.

# Classification hiérarchique

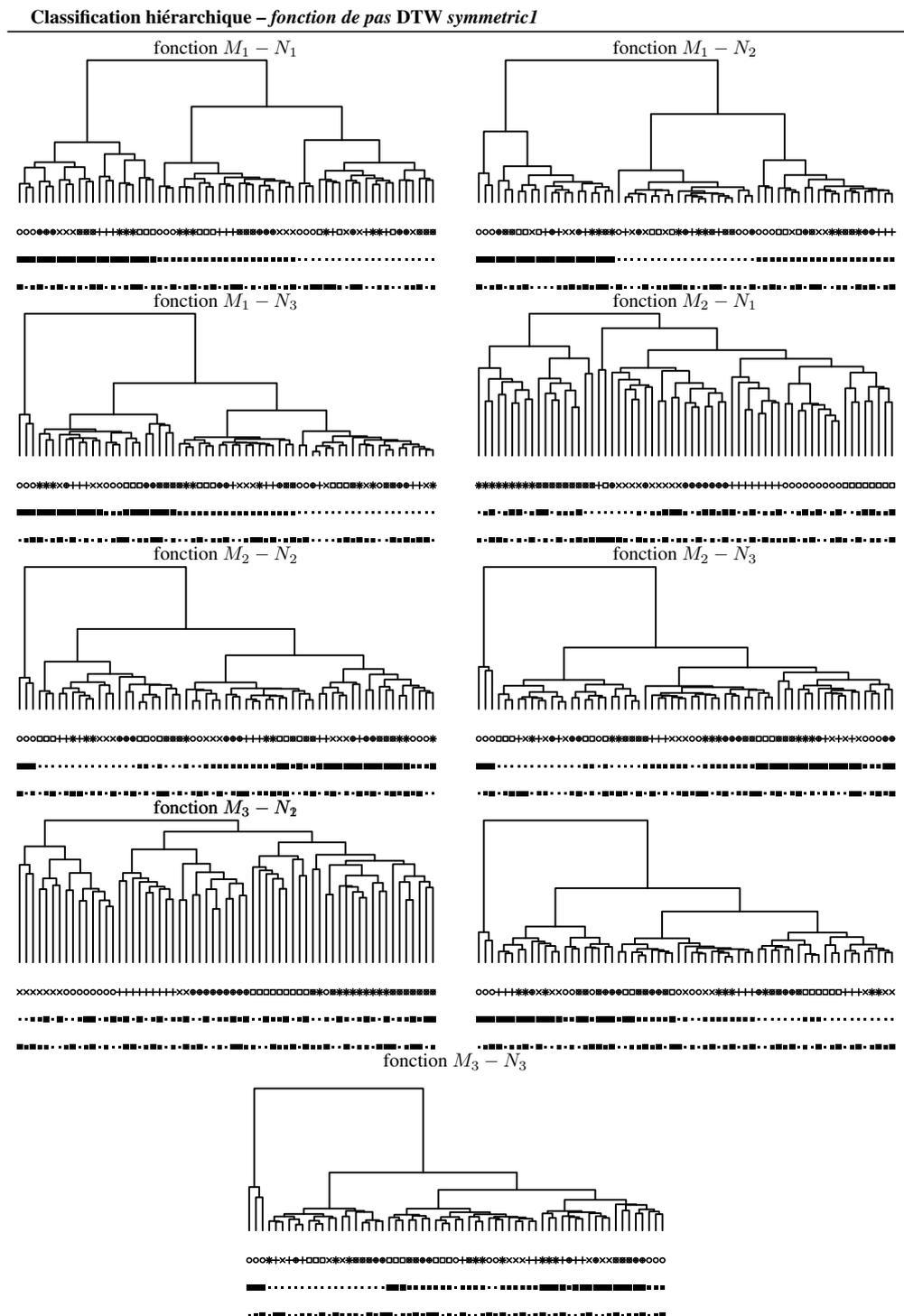


FIGURE 5.15: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *symmetric1*.

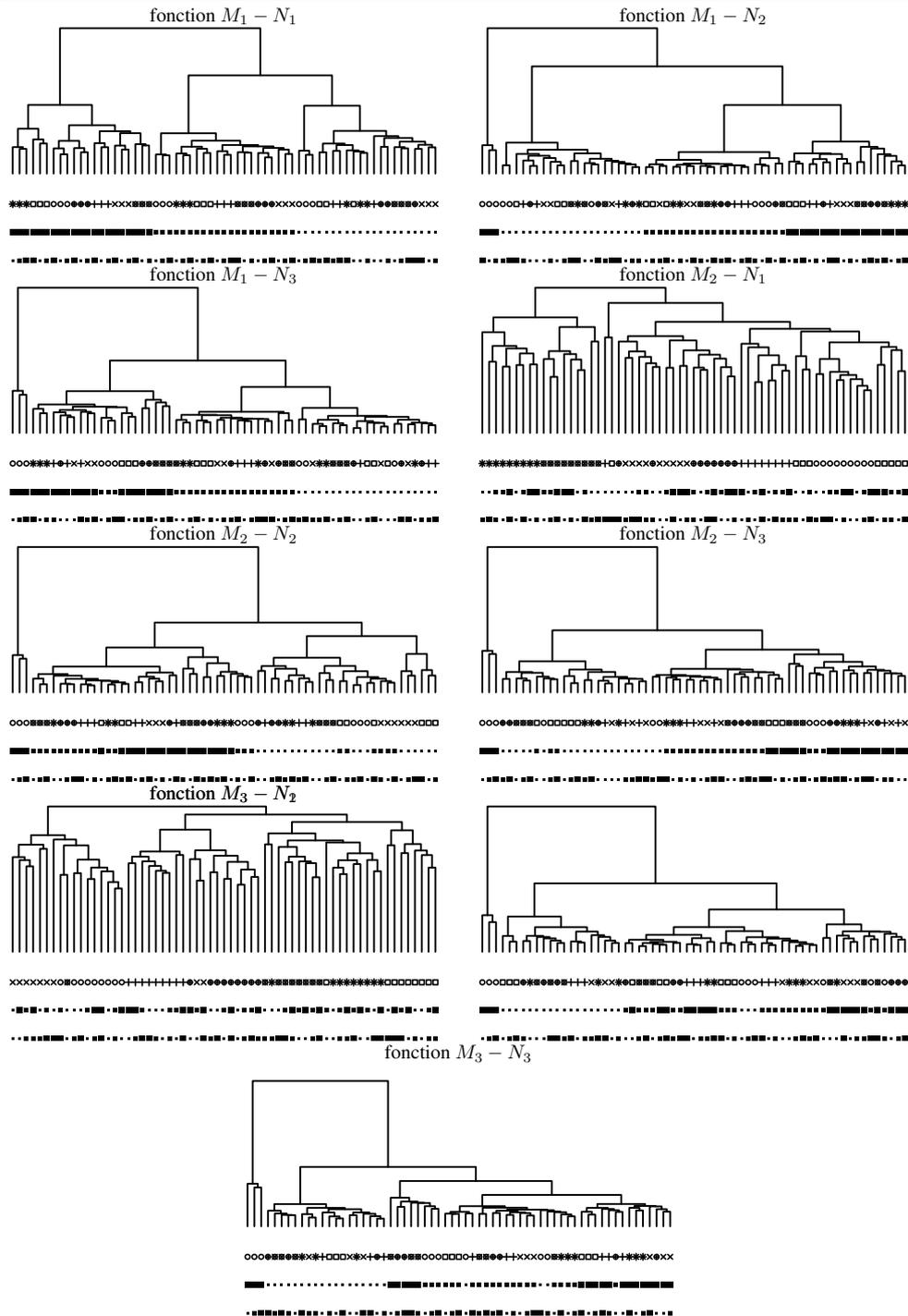


FIGURE 5.16: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *symmetricP0*.

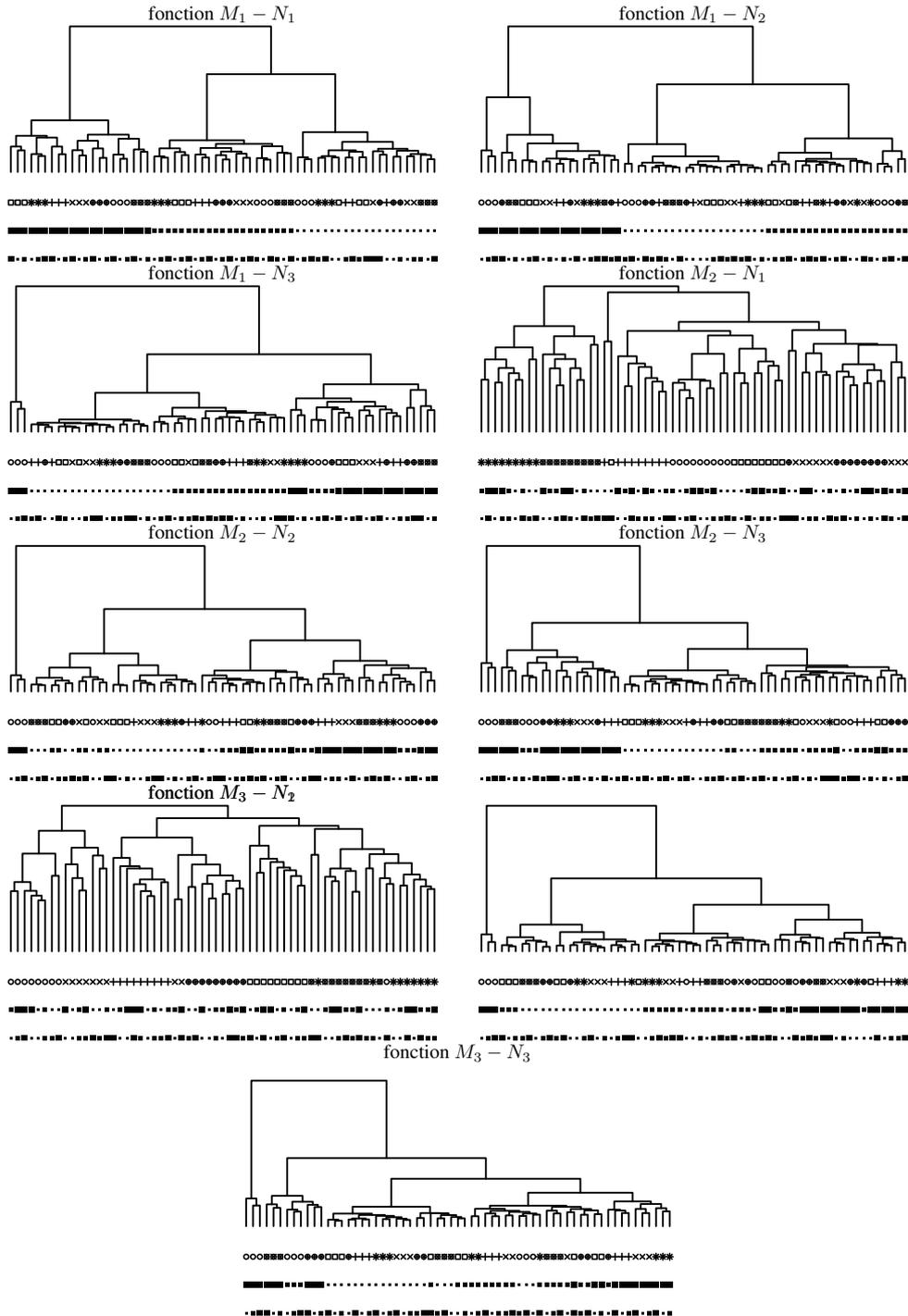


FIGURE 5.17: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *symmetricP1*.

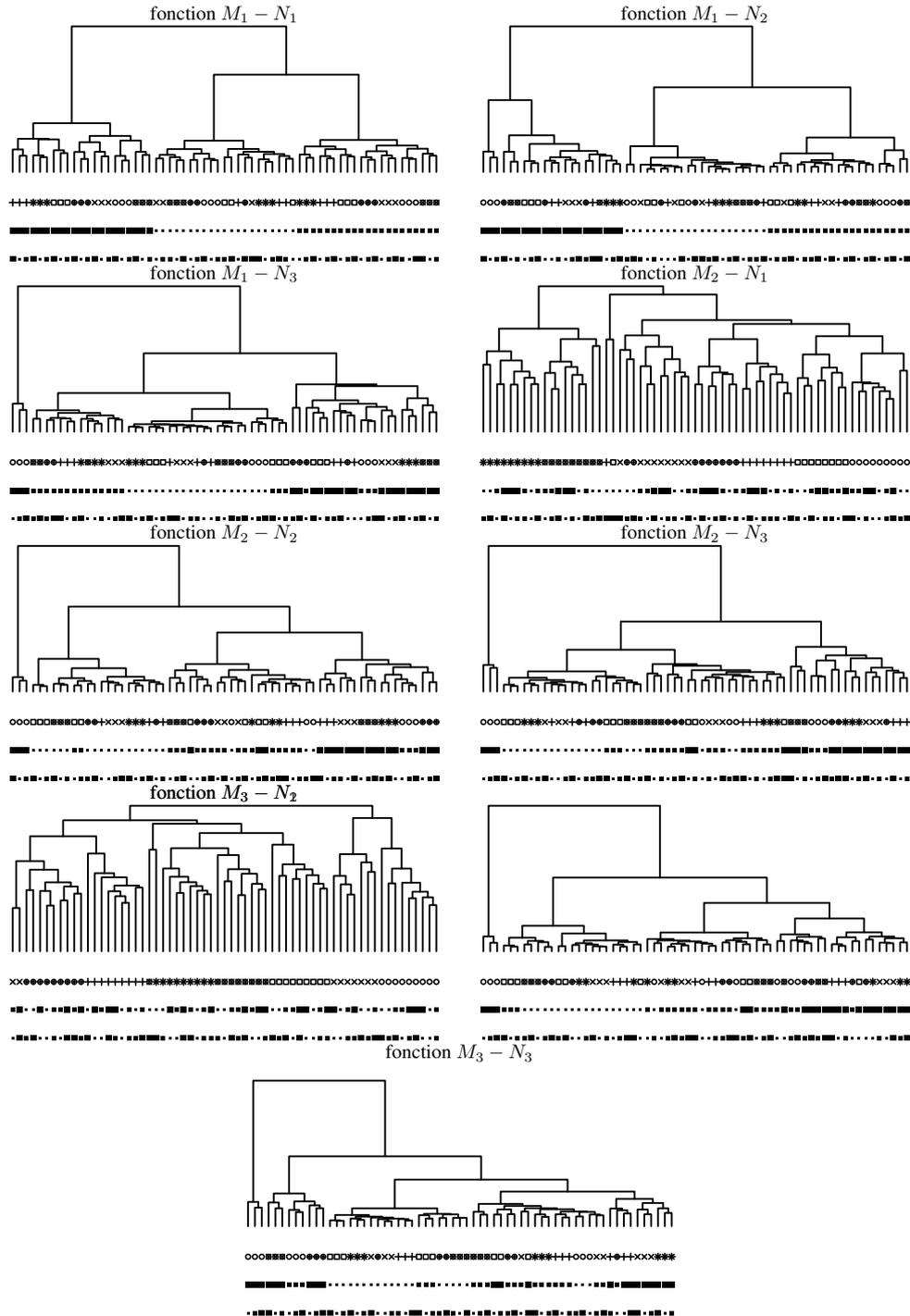


FIGURE 5.18: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW symmetricP2.

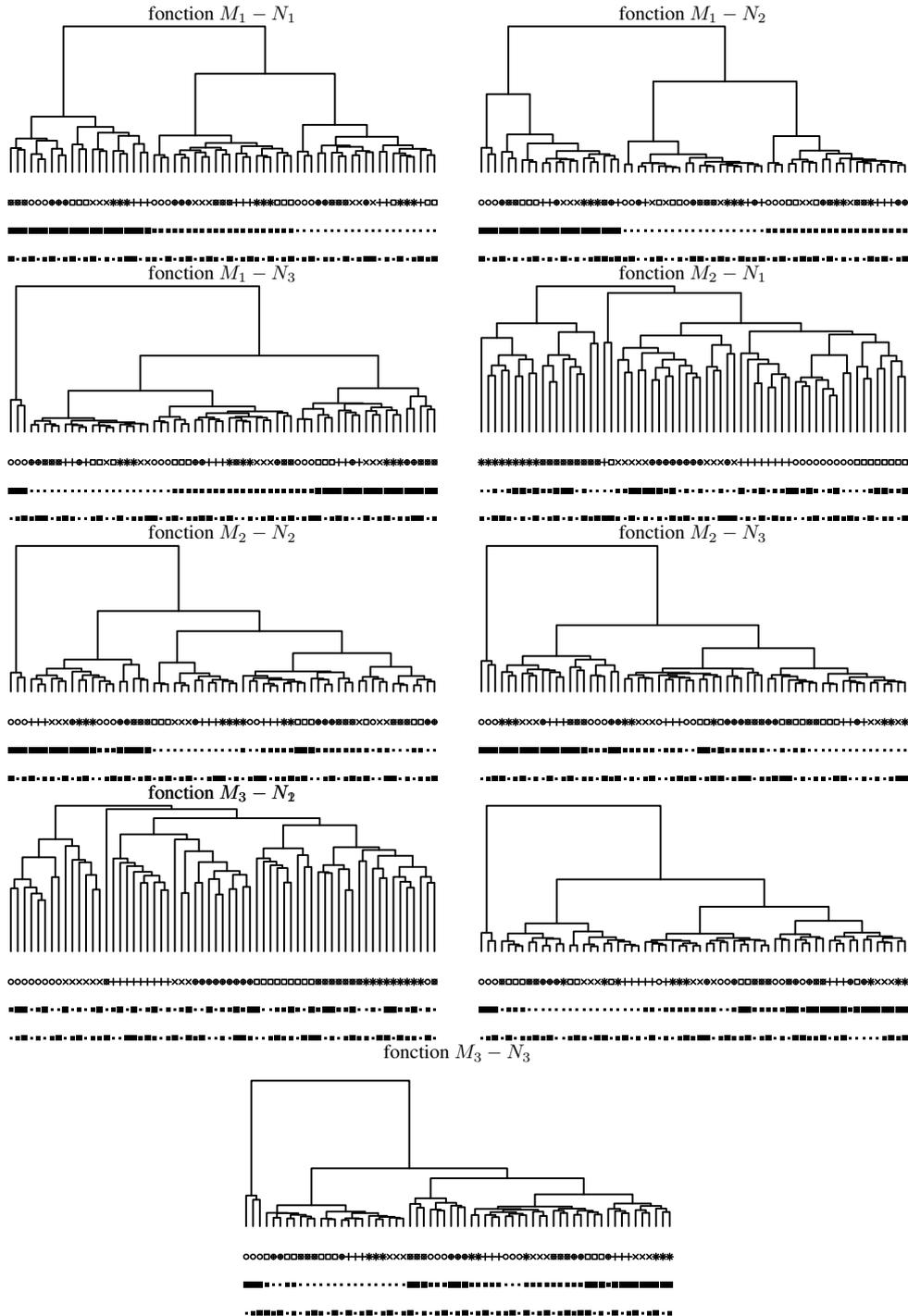


FIGURE 5.19: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *symmetricP05*.

Classification hiérarchique – fonction de pas DTW *typeIbs*

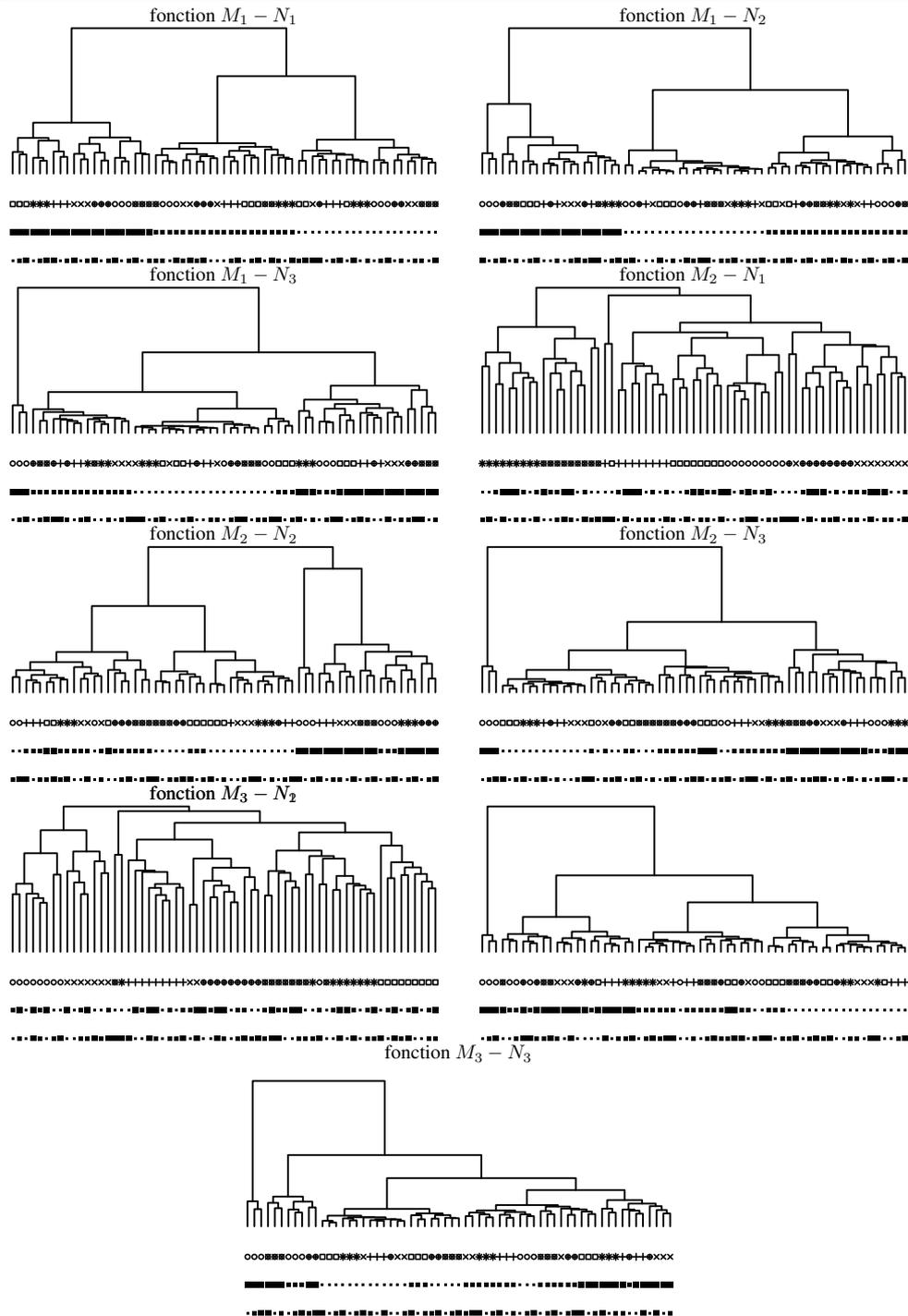


FIGURE 5.20: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *typeIbs*.

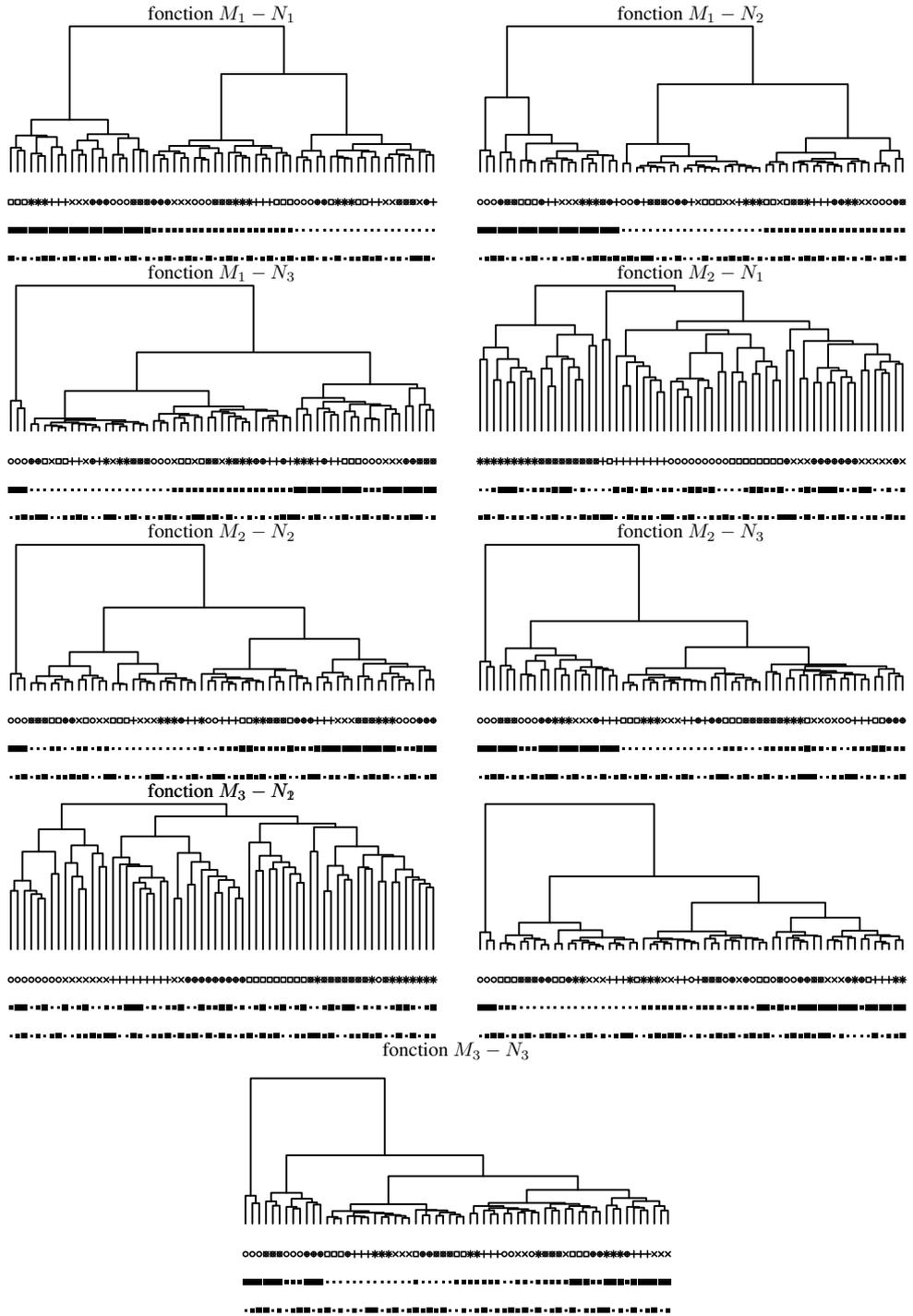


FIGURE 5.21: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *typeIIa*.

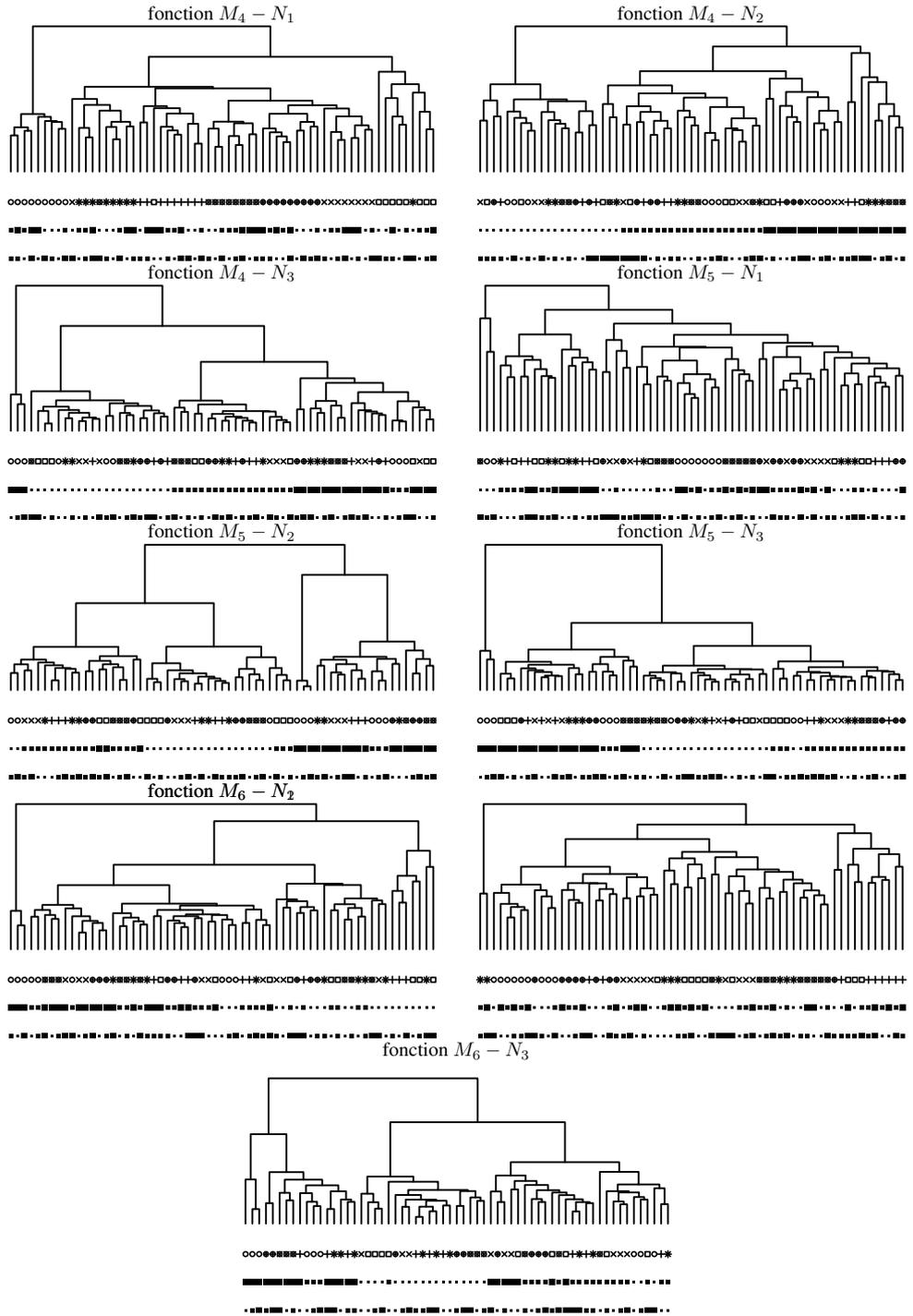


FIGURE 5.22: Illustration des dendrogrammes issus des classifications hiérarchiques pour la *fonction de pas DTW symmetric1*.

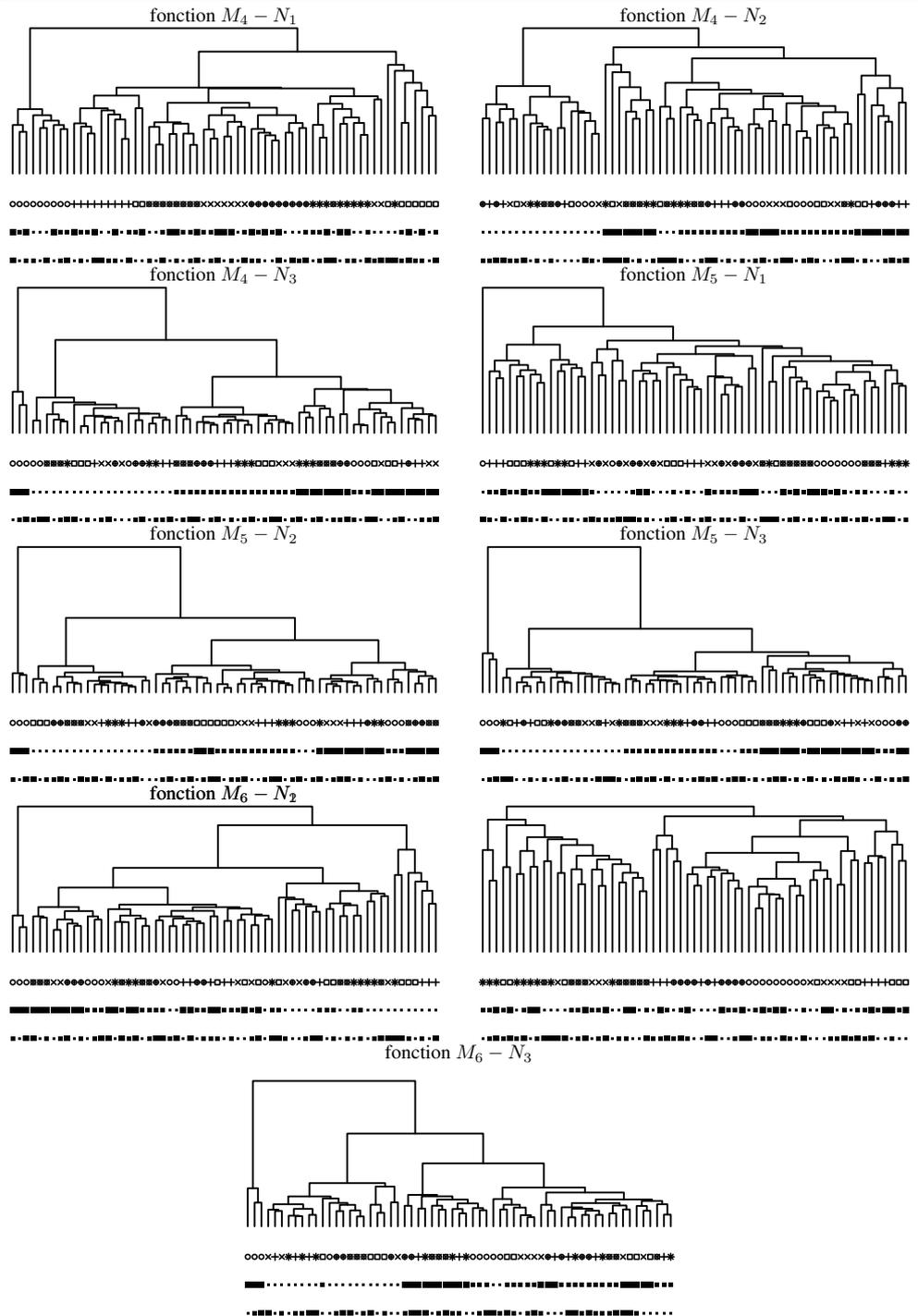


FIGURE 5.23: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *symmetricP0*.

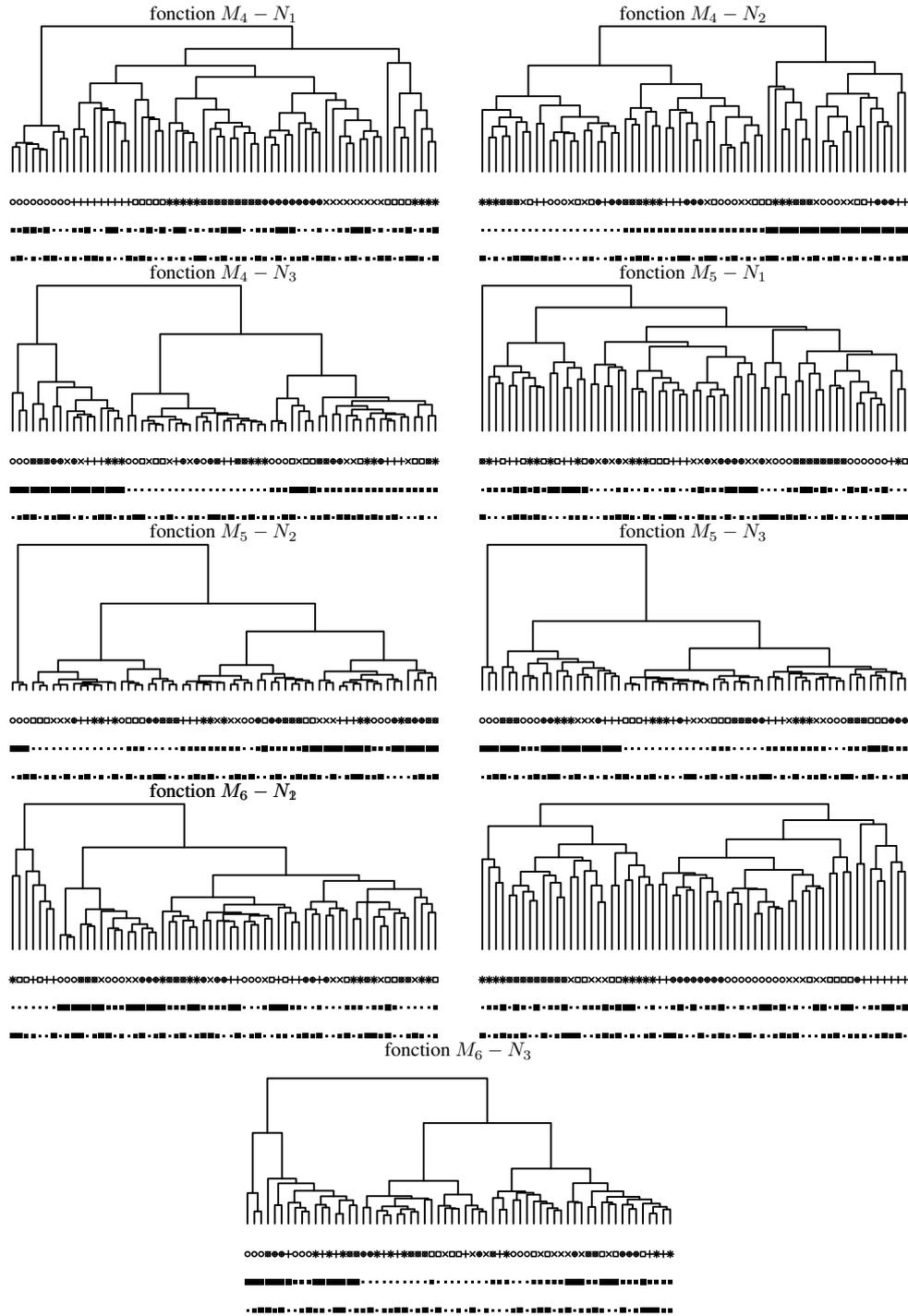


FIGURE 5.24: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *symmetricP1*.

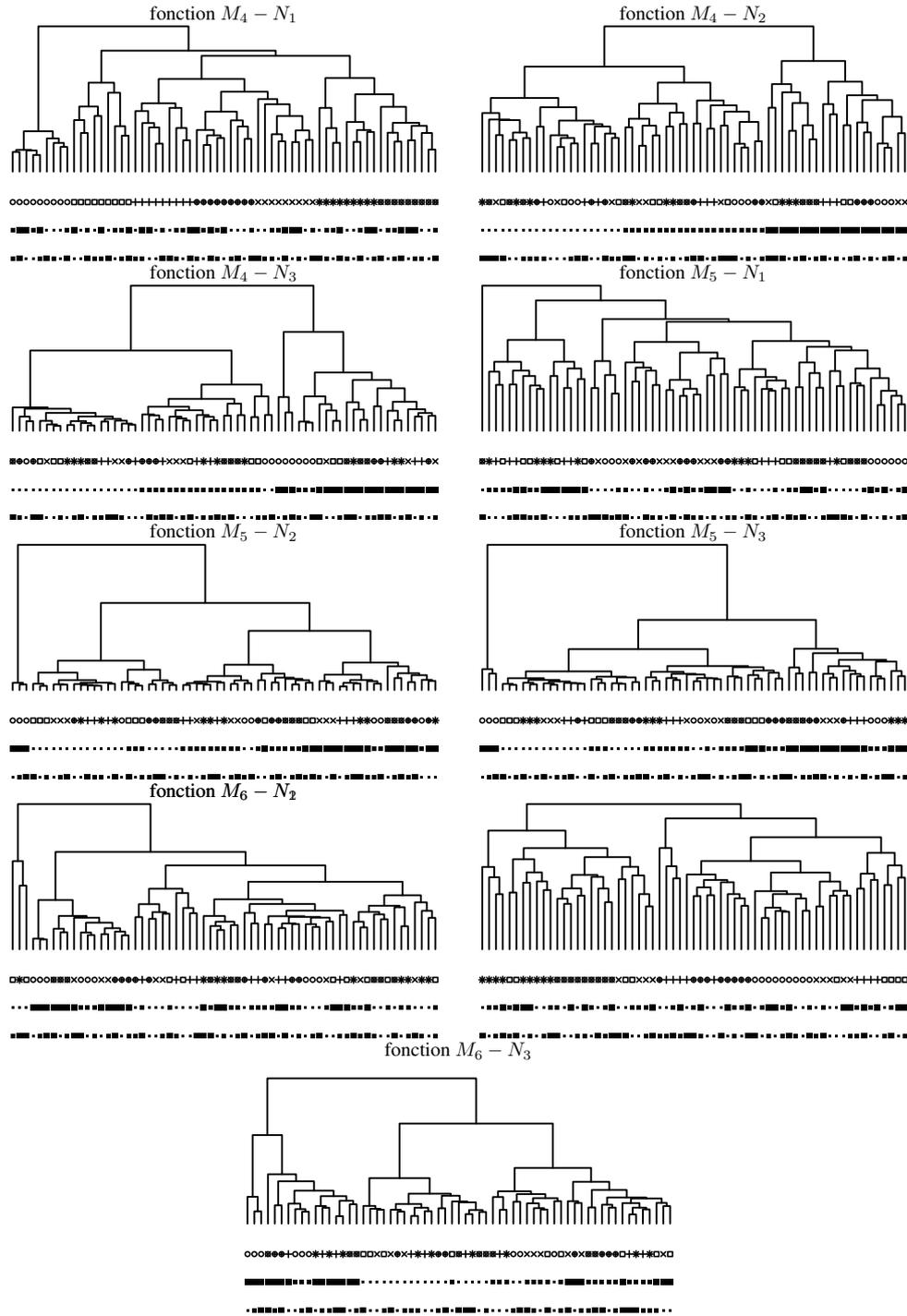


FIGURE 5.25: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *symmetricP2*.

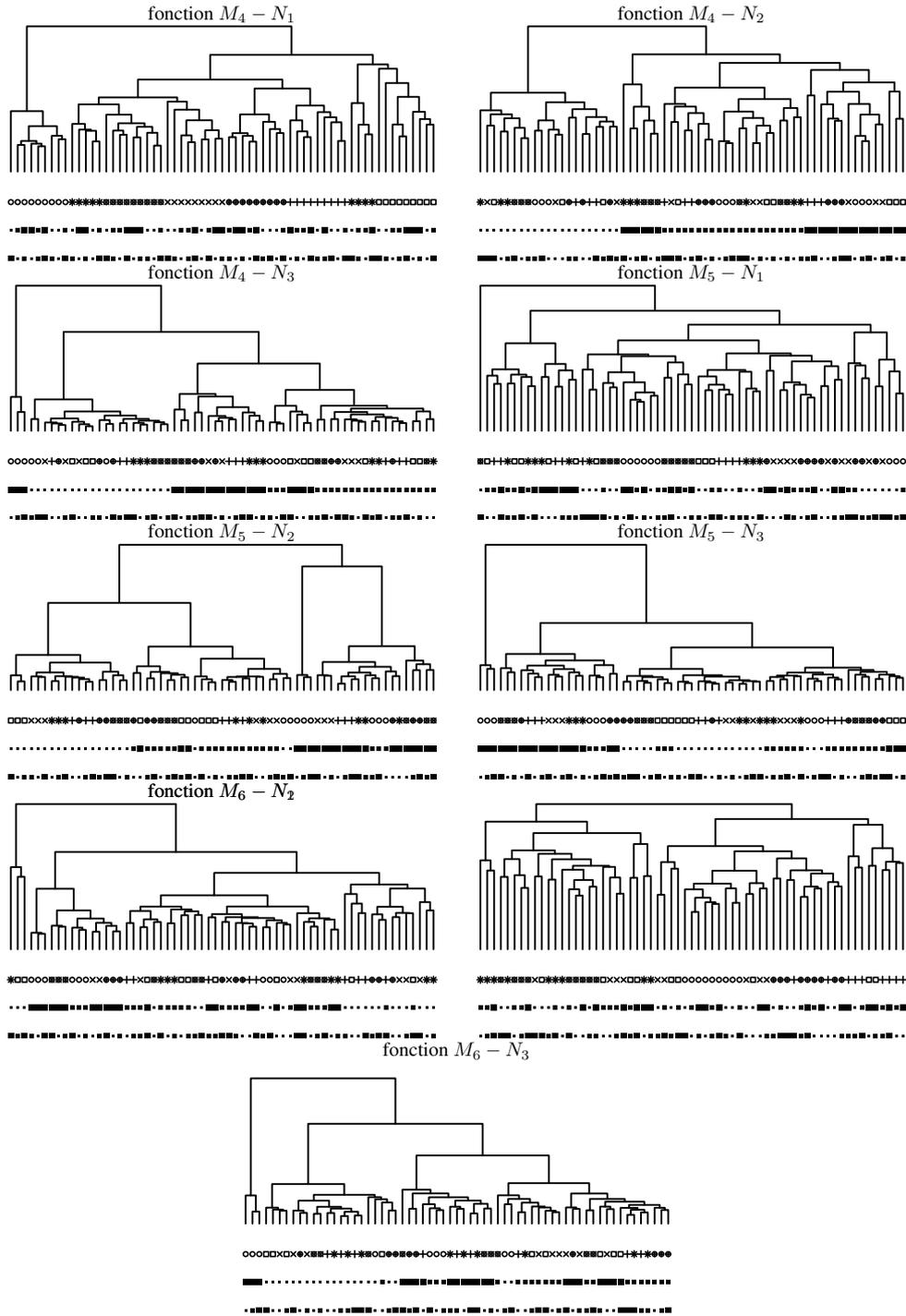


FIGURE 5.26: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW symmetricP05.

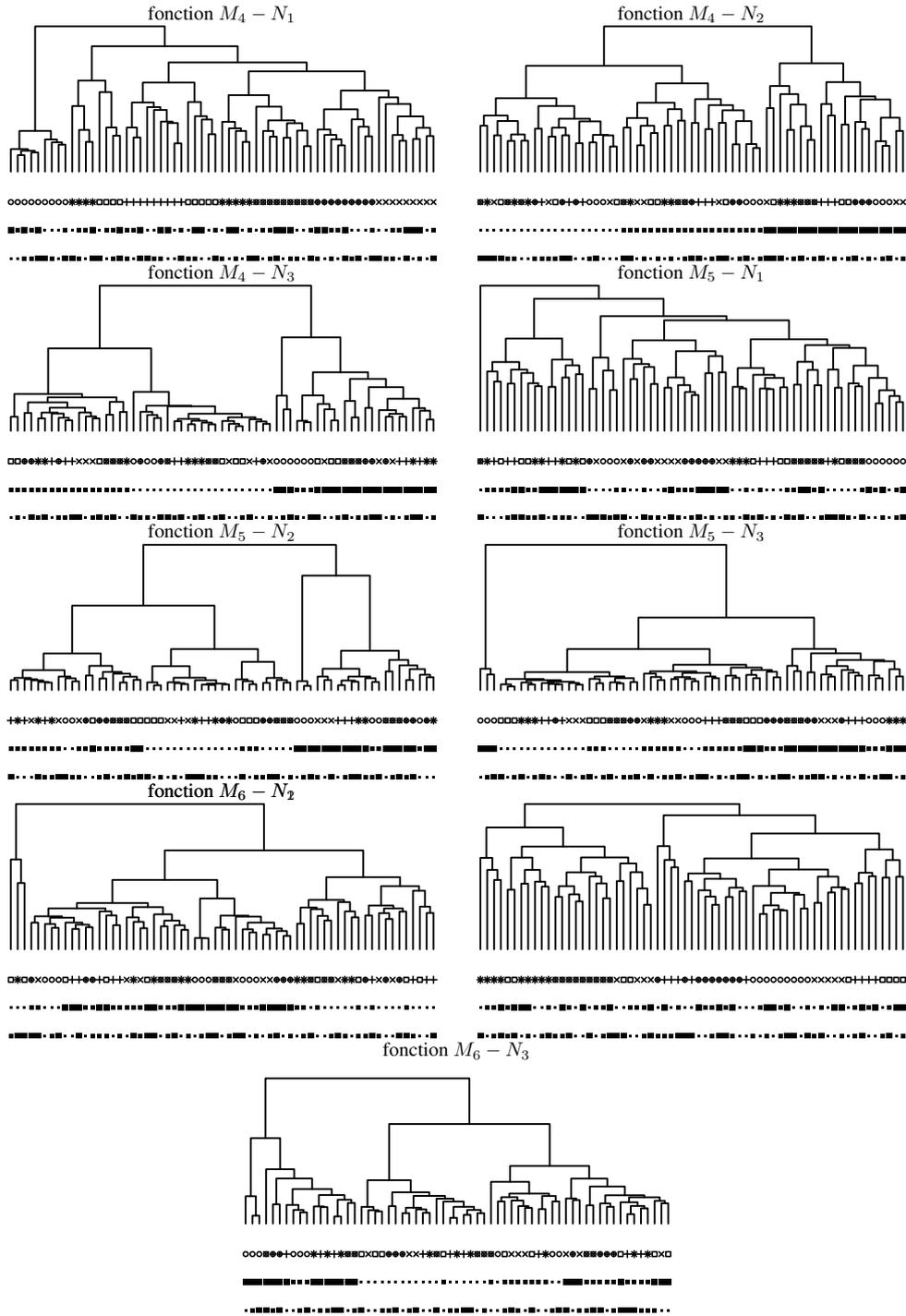


FIGURE 5.27: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *typeIbs*.

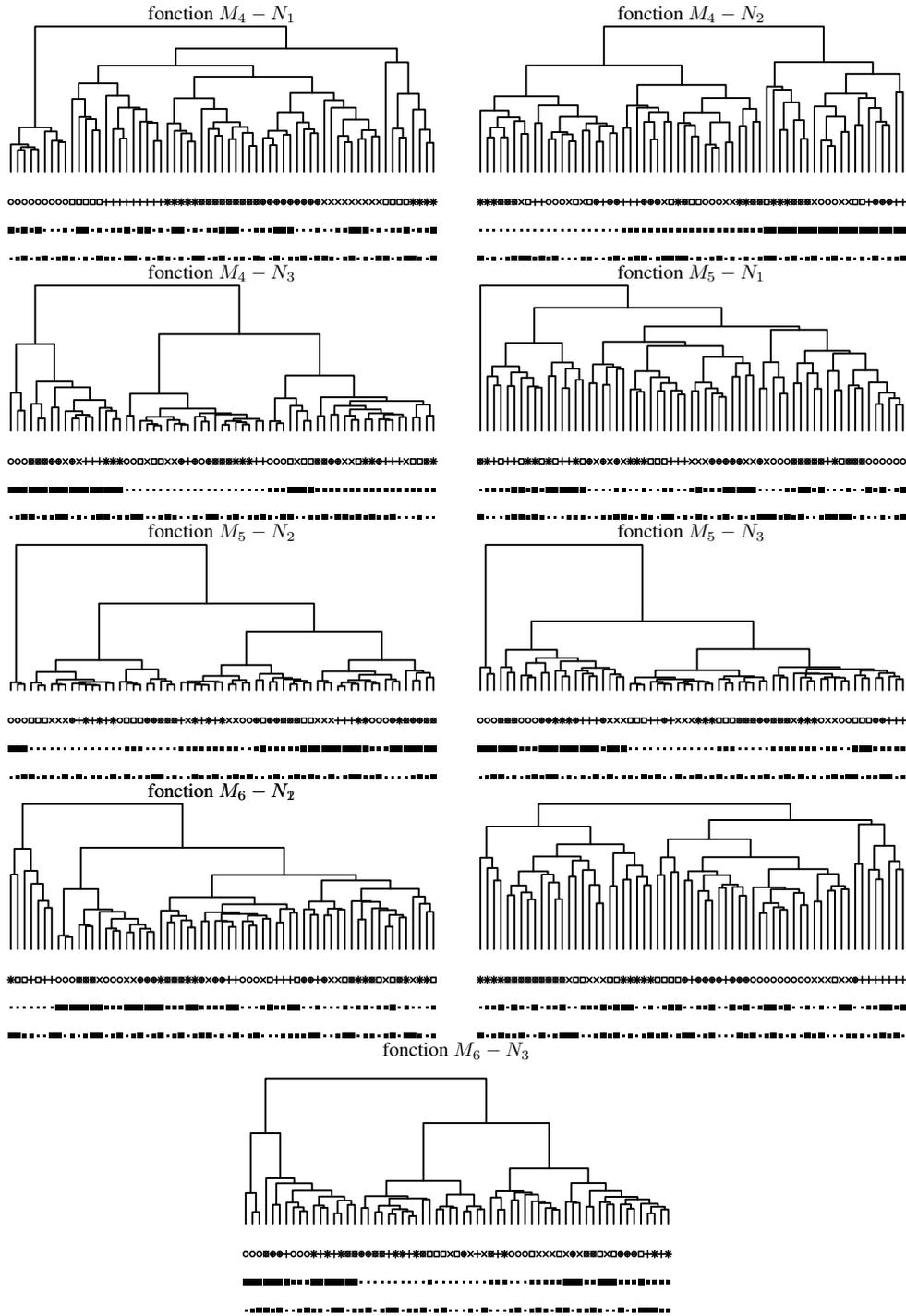


FIGURE 5.28: Illustration des dendrogrammes issus des classifications hiérarchiques pour la fonction de pas DTW *typeIIa*.

## 5.6 Expérience 2

### La largeur, la hauteur et le rapport $W/H$

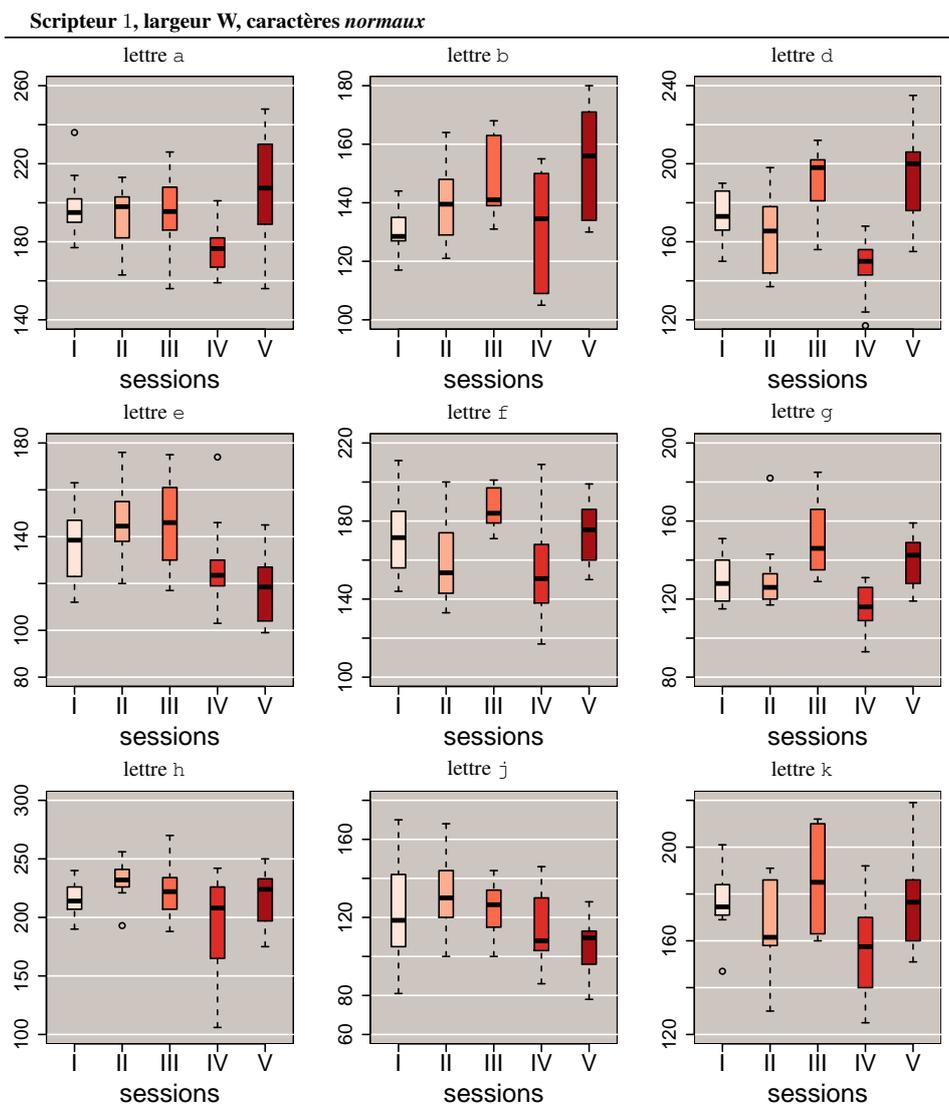


FIGURE 5.29: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

**Scripteur 1, largeur W, caractères *normaux***

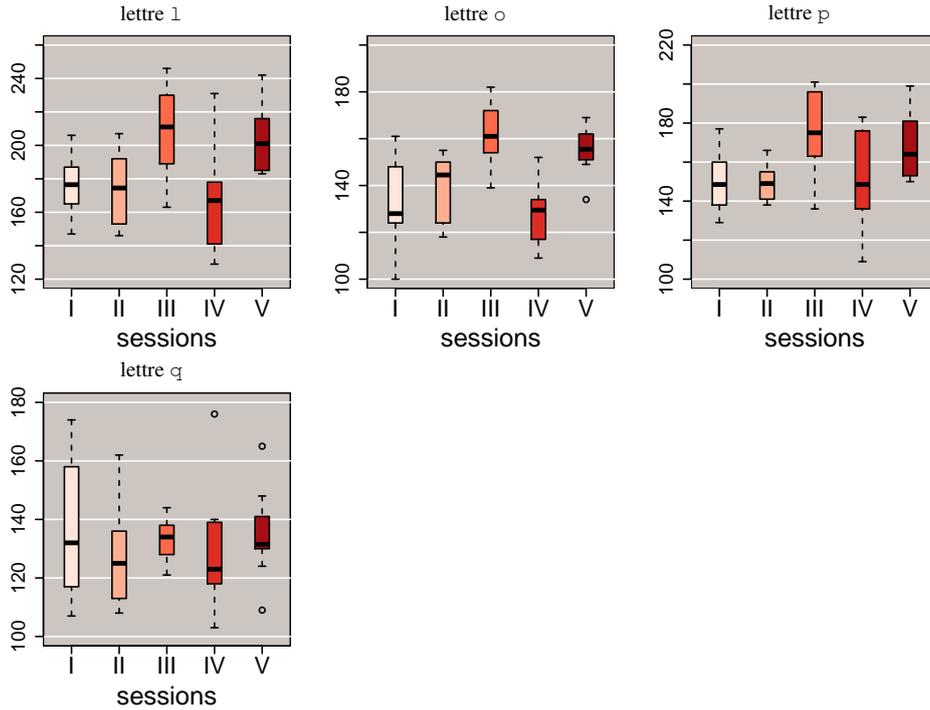


FIGURE 5.30: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

**Scripteur 2, largeur W, caractères *normaux***

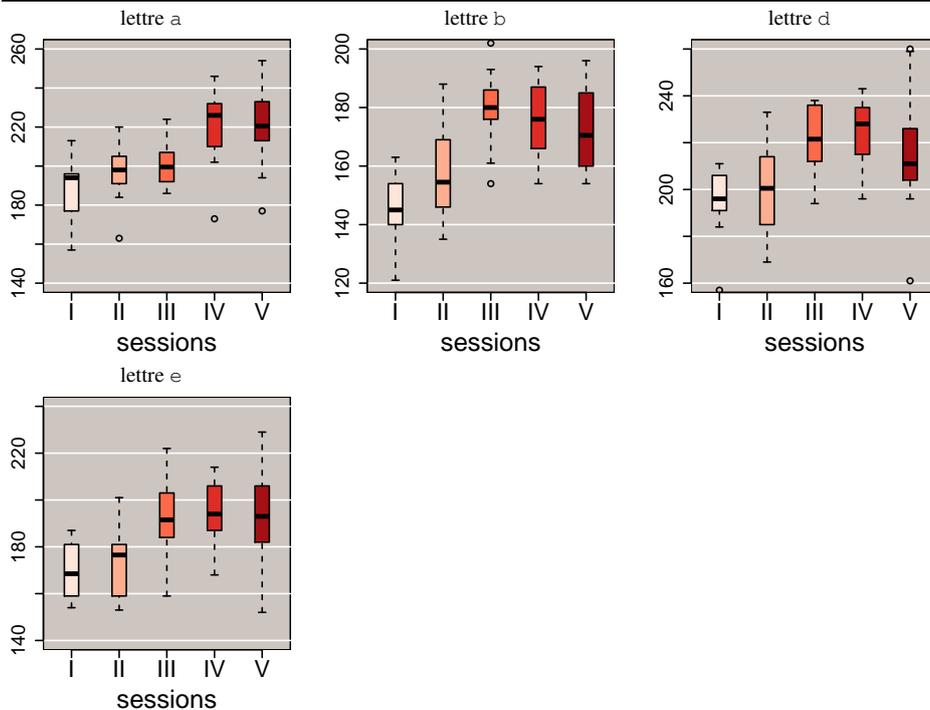


FIGURE 5.31: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

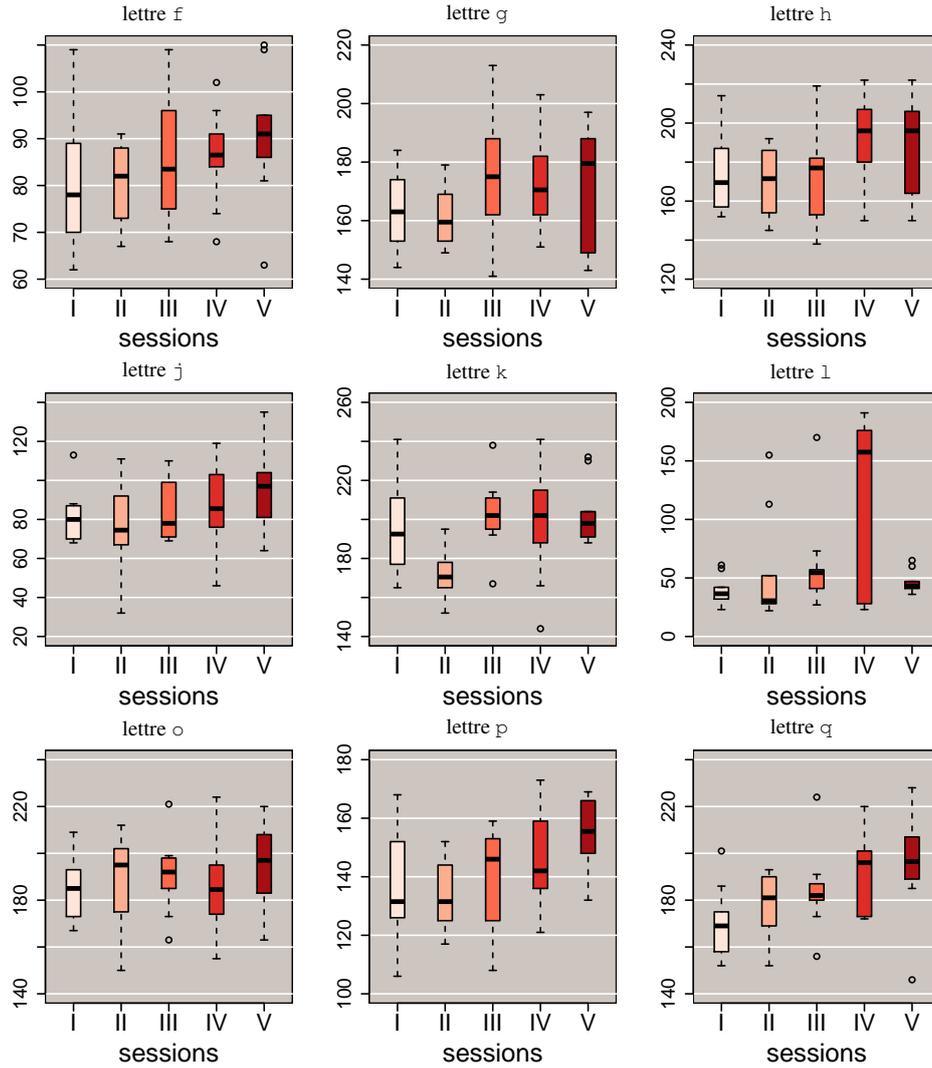


FIGURE 5.32: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

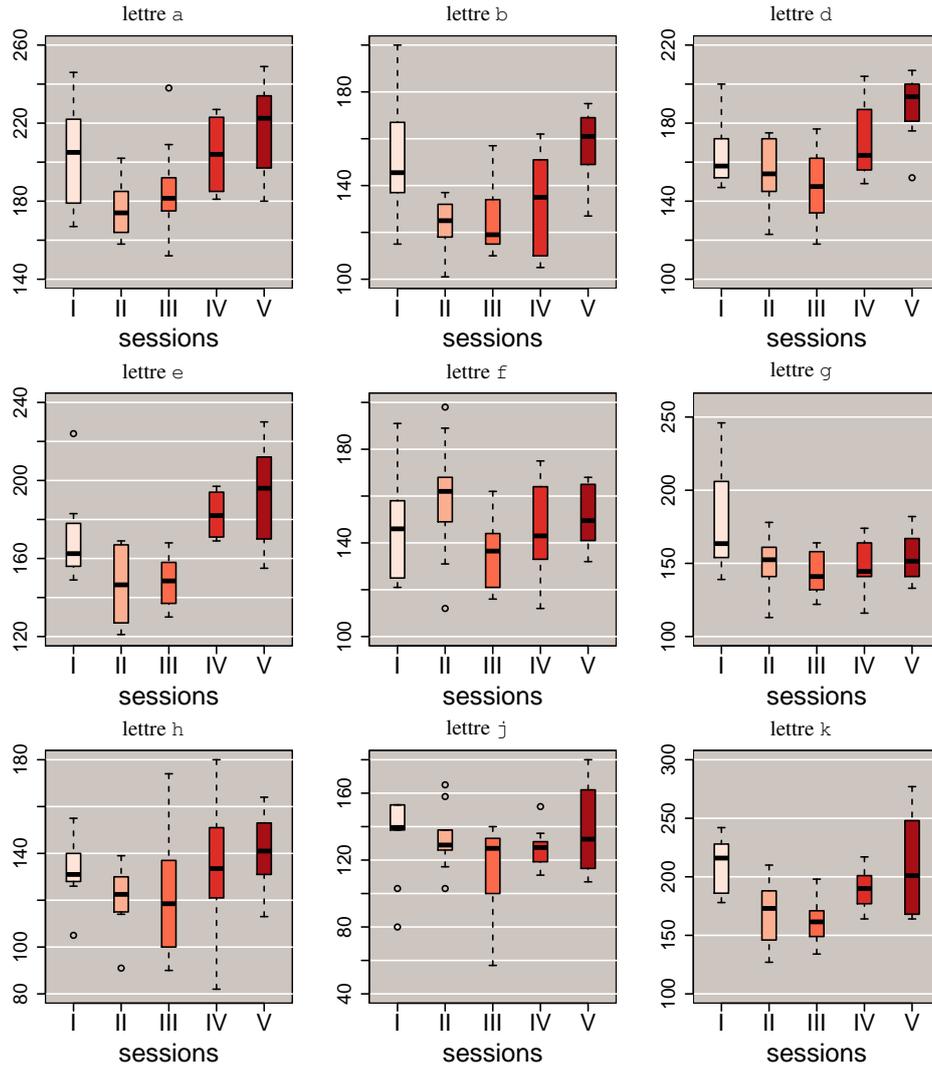


FIGURE 5.33: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 3, largeur W, caractères *normaux*

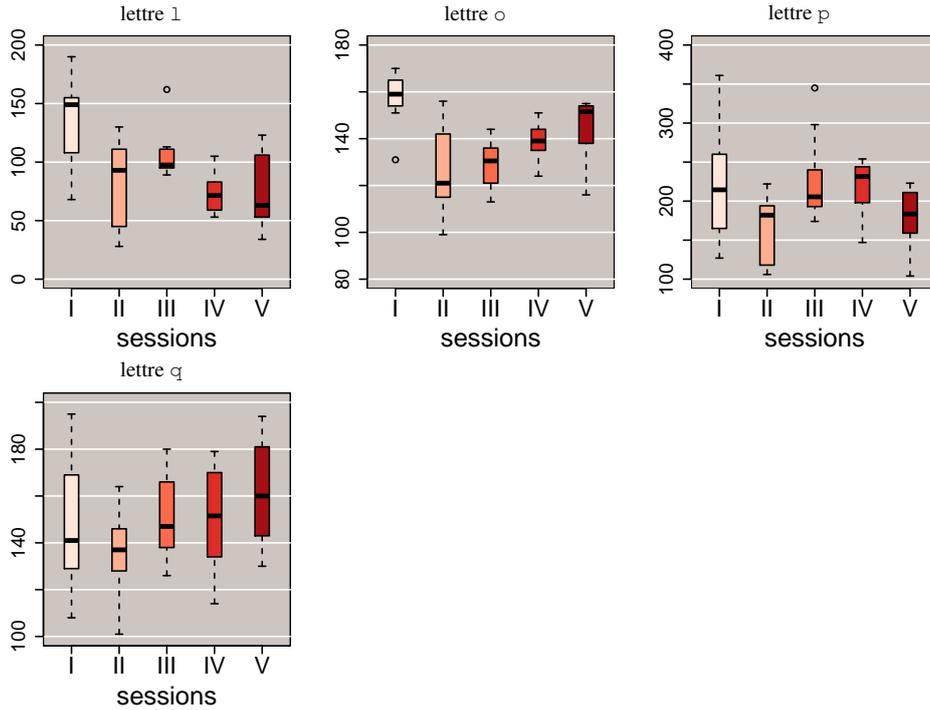


FIGURE 5.34: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 4, largeur W, caractères *normaux*

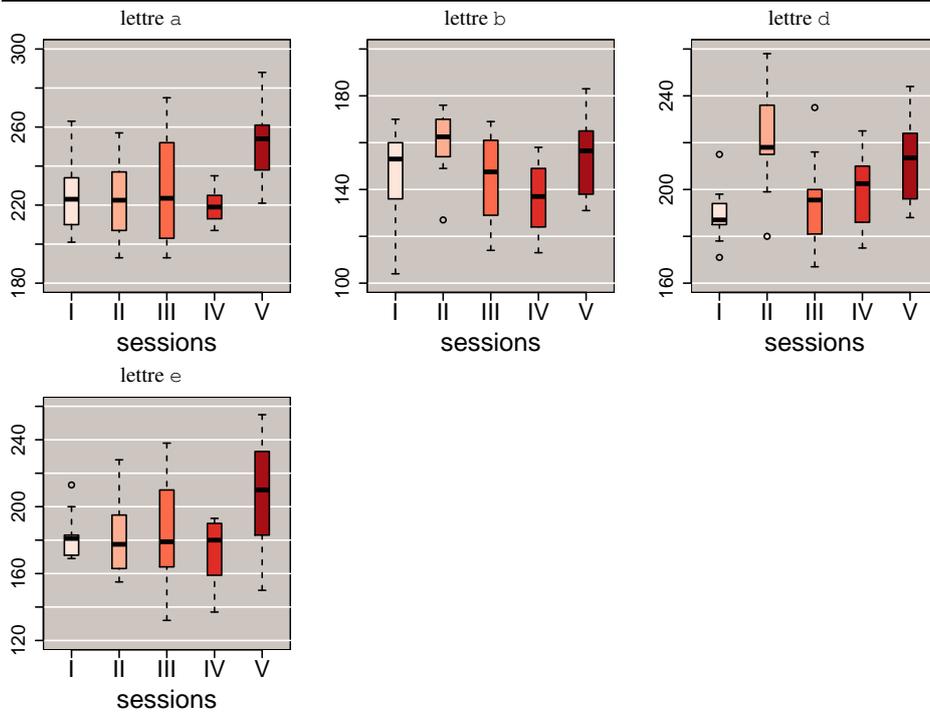


FIGURE 5.35: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

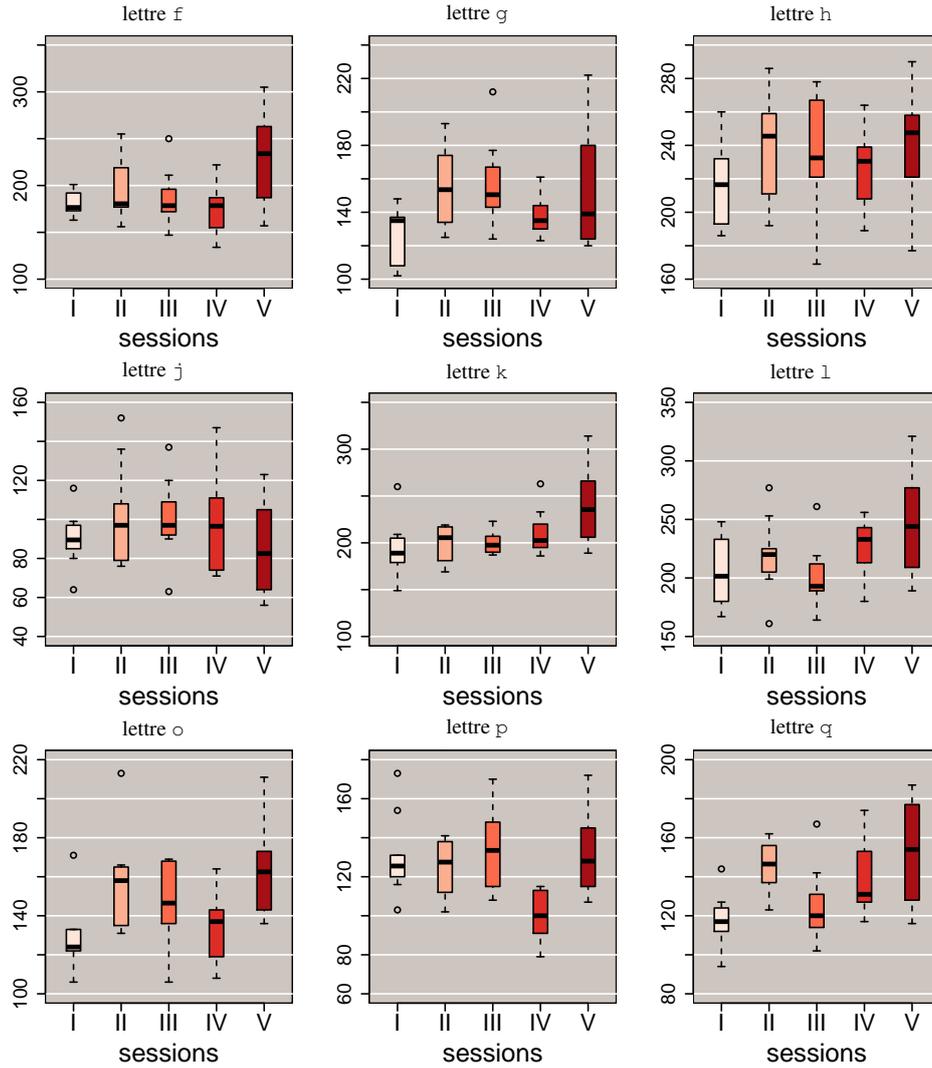


FIGURE 5.36: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

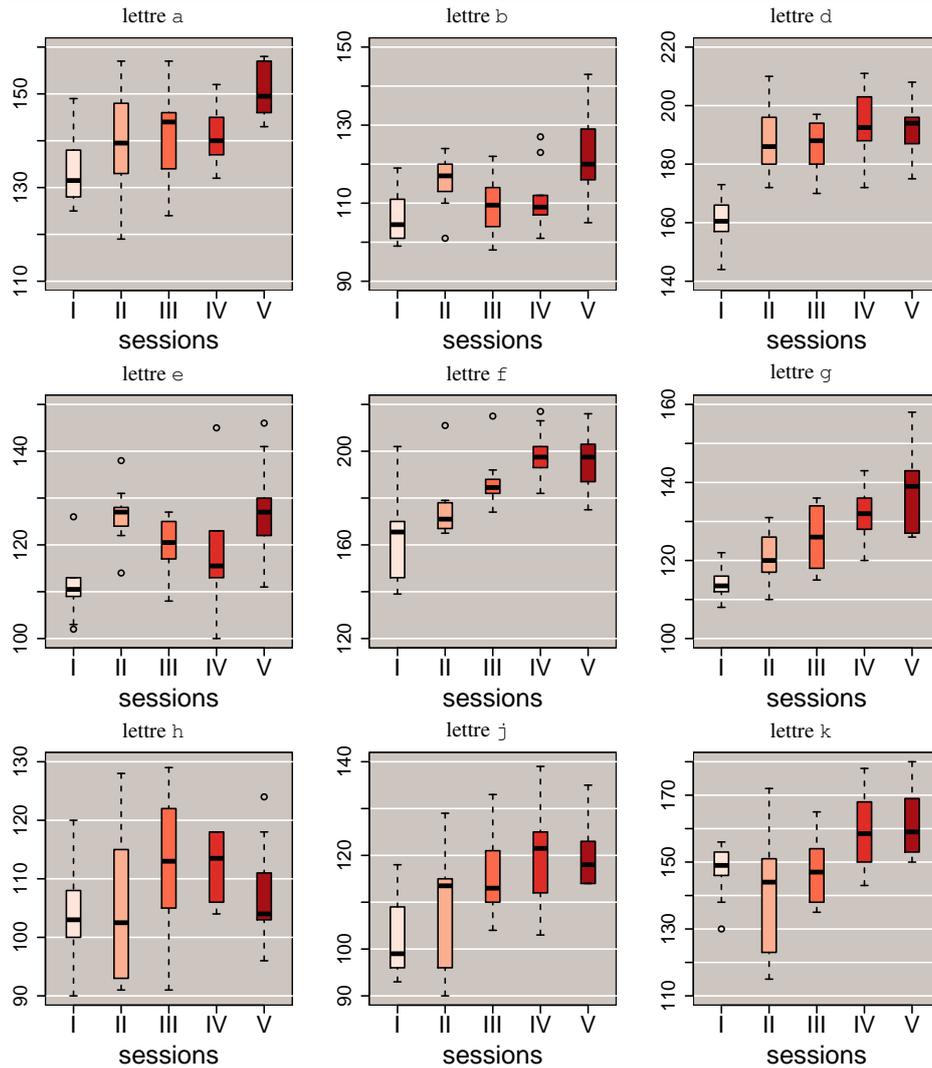


FIGURE 5.37: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 5, largeur W, caractères *normaux*

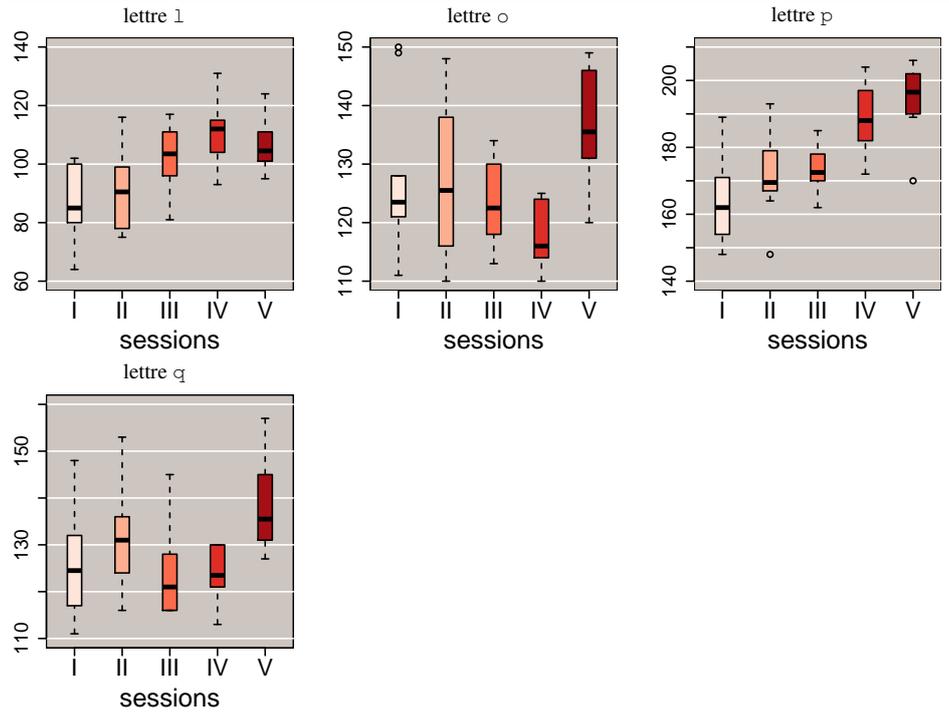


FIGURE 5.38: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 6, largeur W, caractères *normaux*

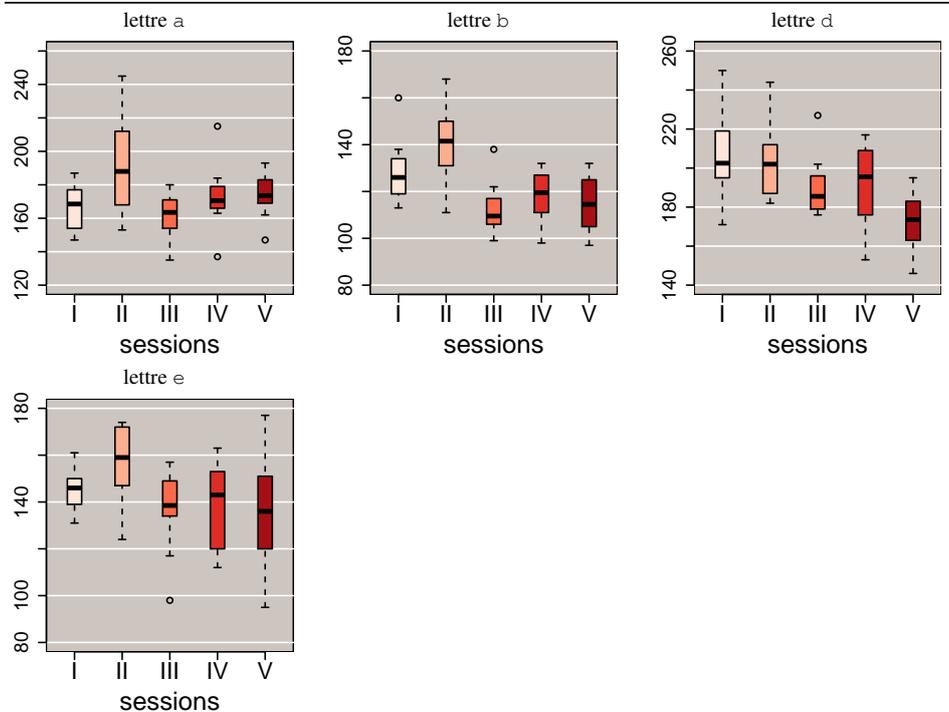


FIGURE 5.39: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

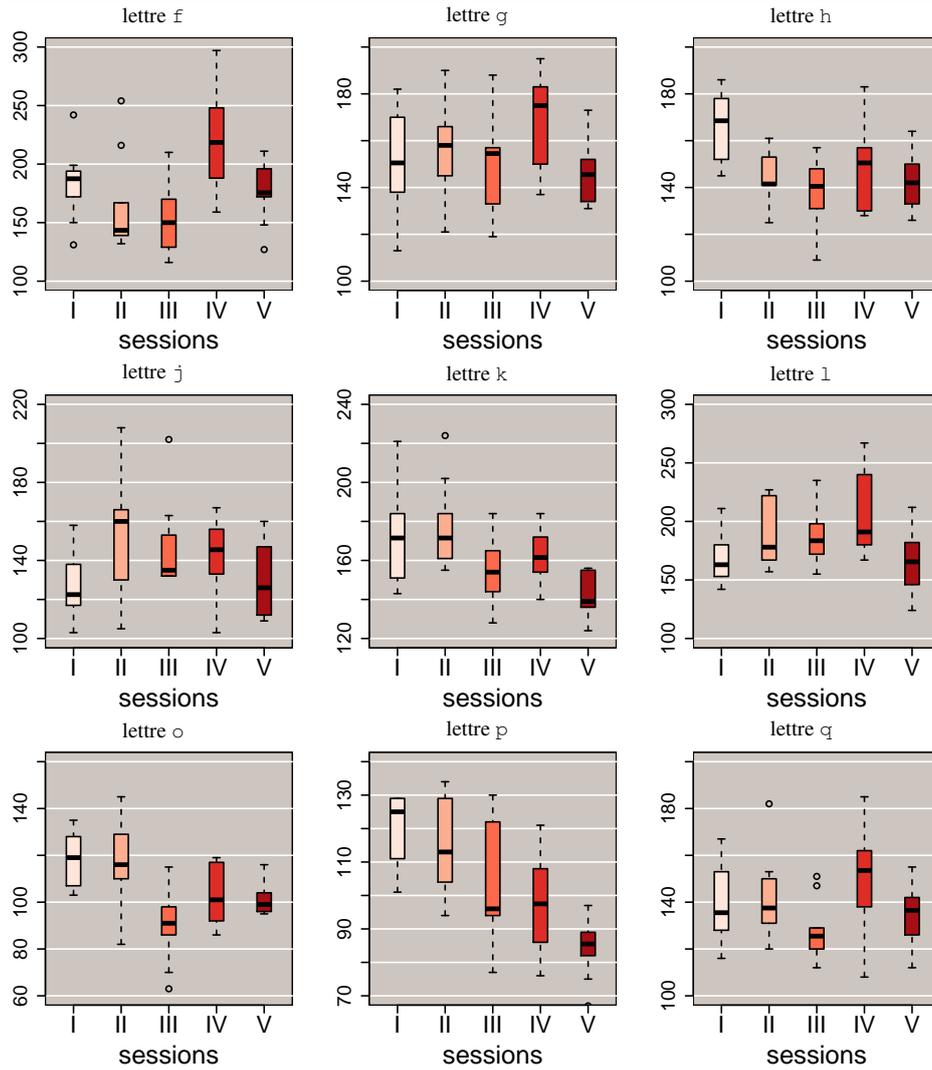


FIGURE 5.40: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

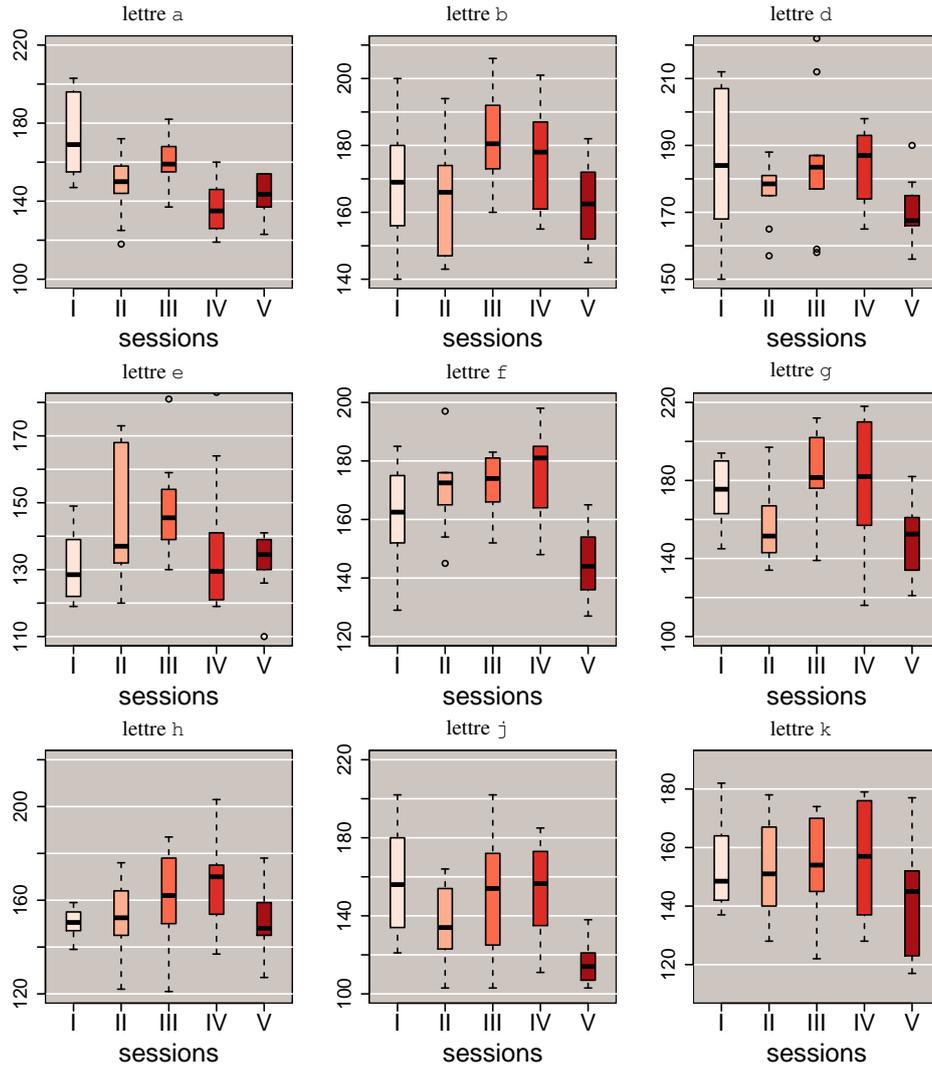


FIGURE 5.41: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 7, largeur W, caractères *normaux*

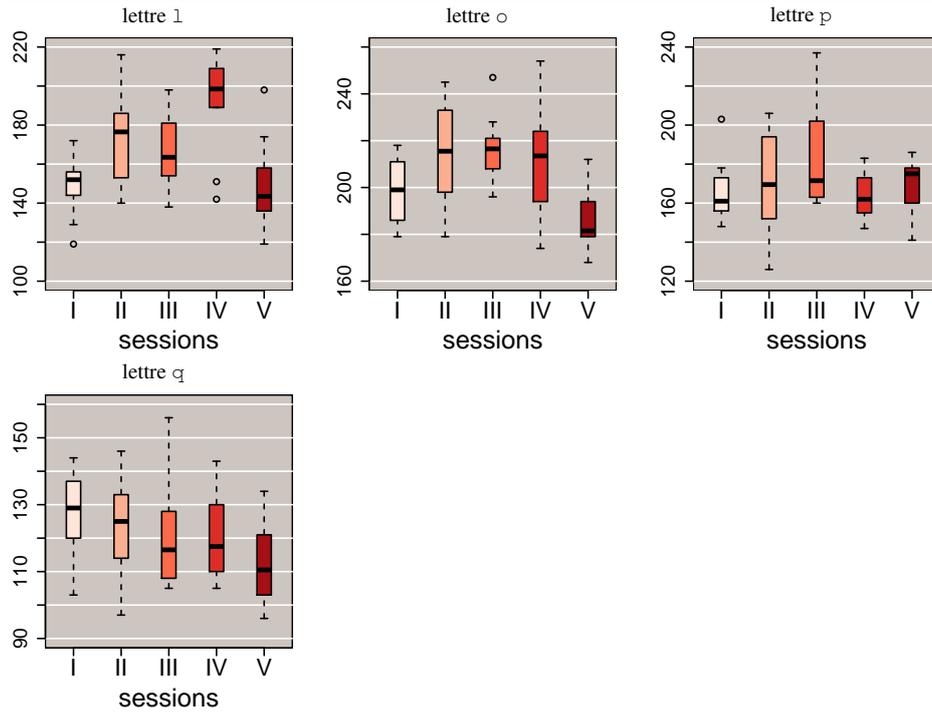


FIGURE 5.42: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 8, largeur W, caractères *normaux*

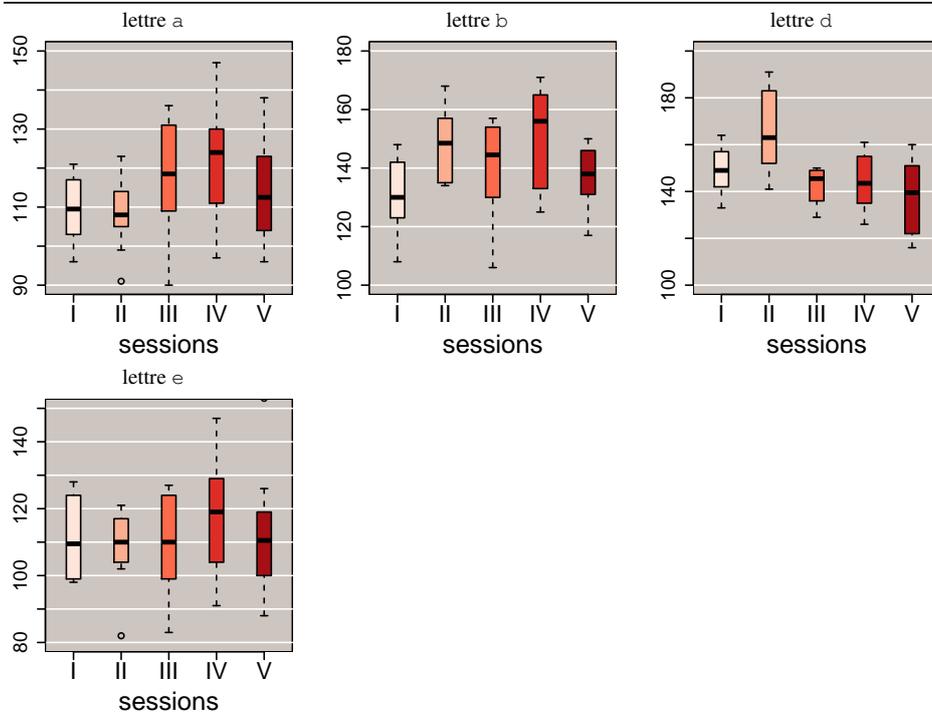


FIGURE 5.43: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

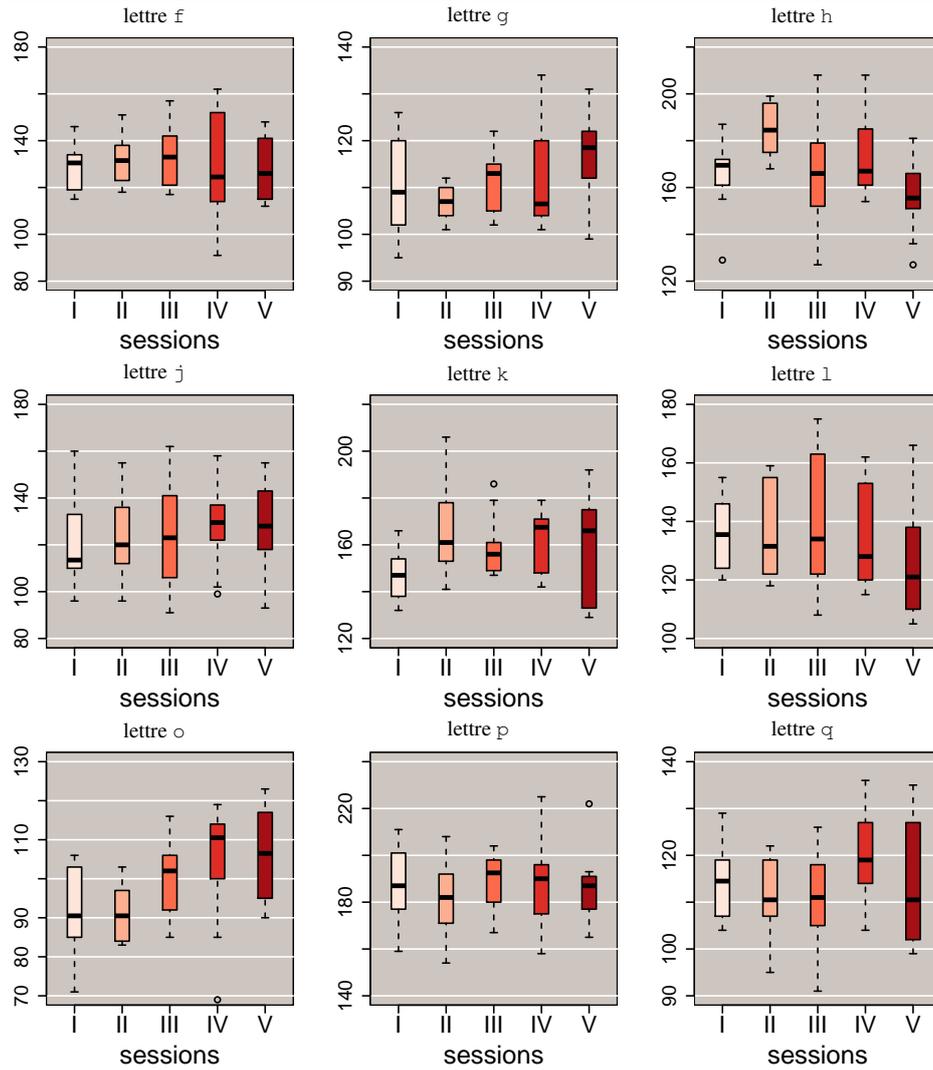


FIGURE 5.44: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

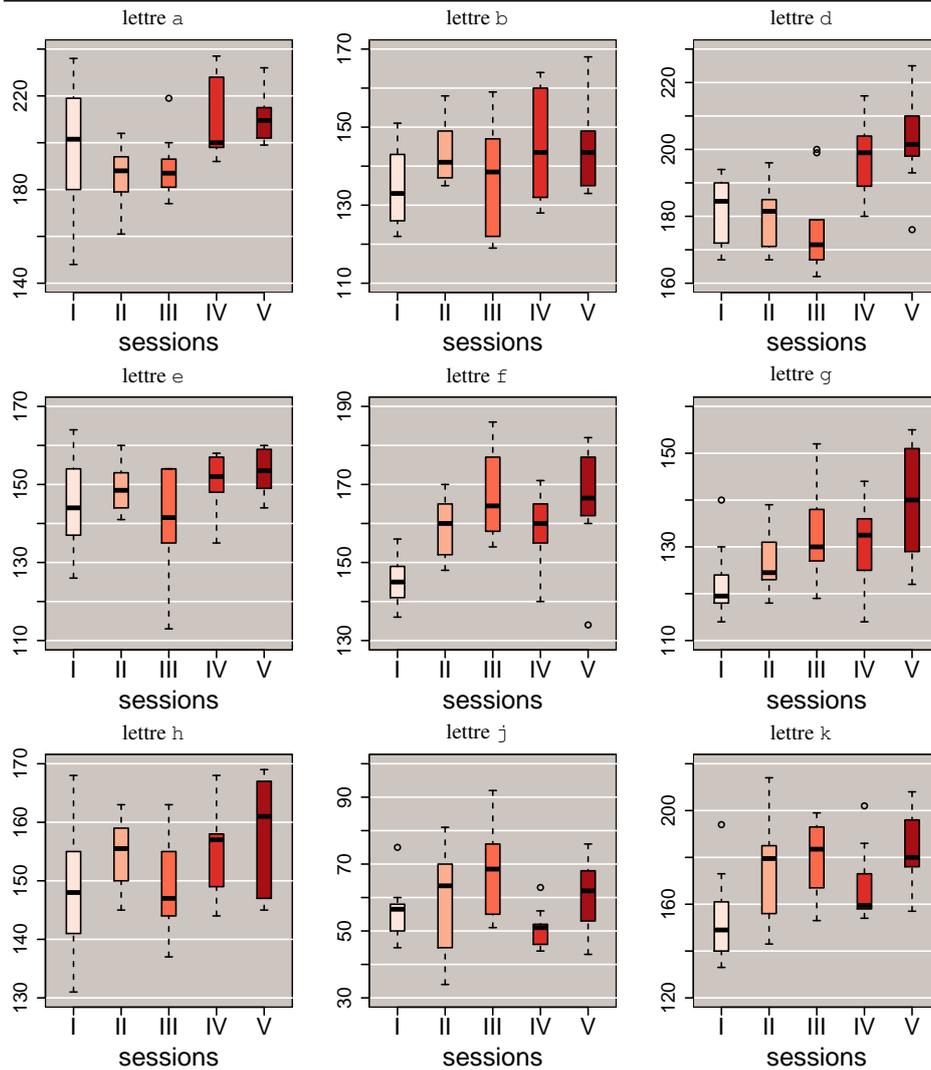


FIGURE 5.45: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 9, largeur W, caractères *normaux*

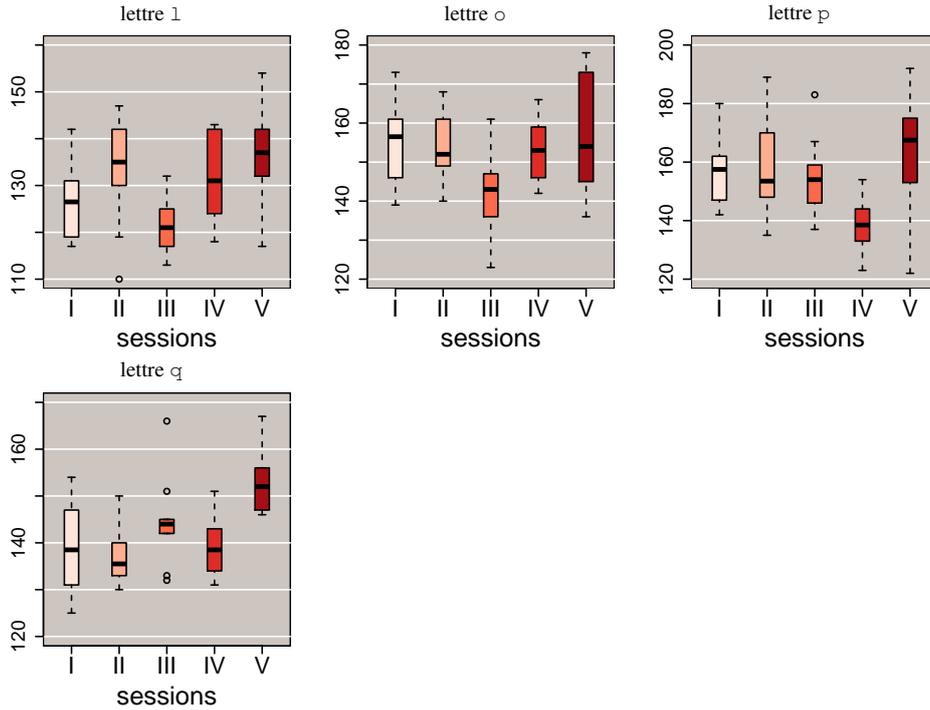


FIGURE 5.46: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 10, largeur W, caractères *normaux*

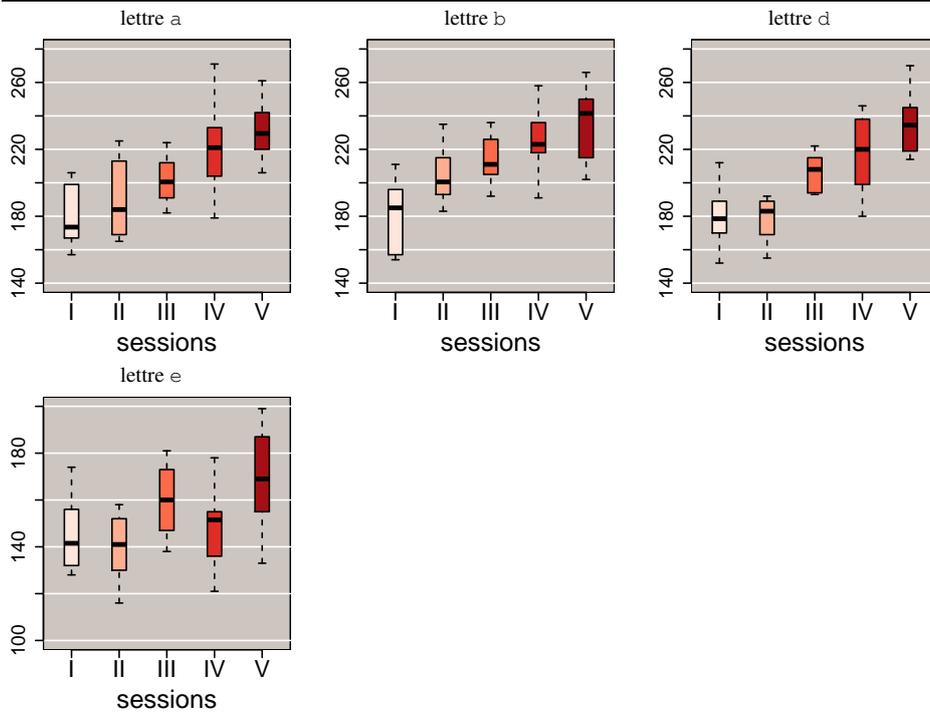


FIGURE 5.47: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

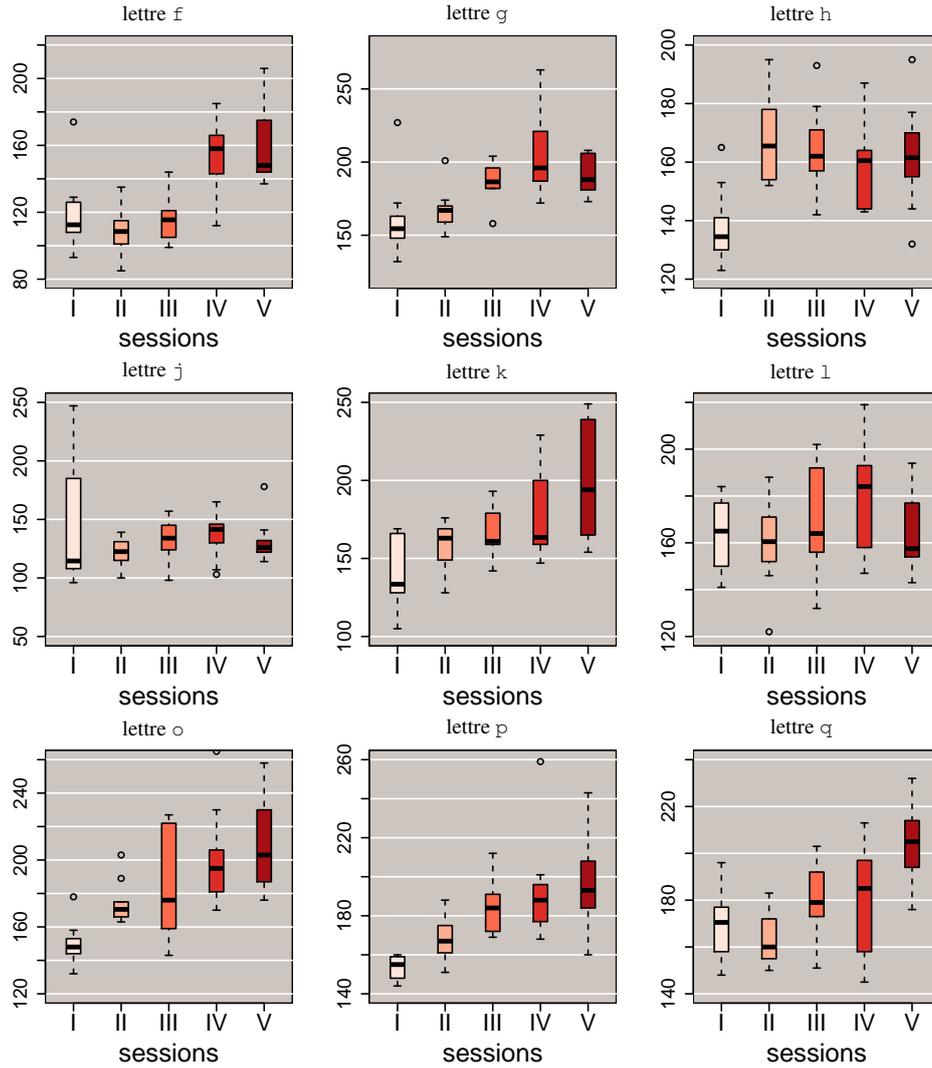


FIGURE 5.48: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

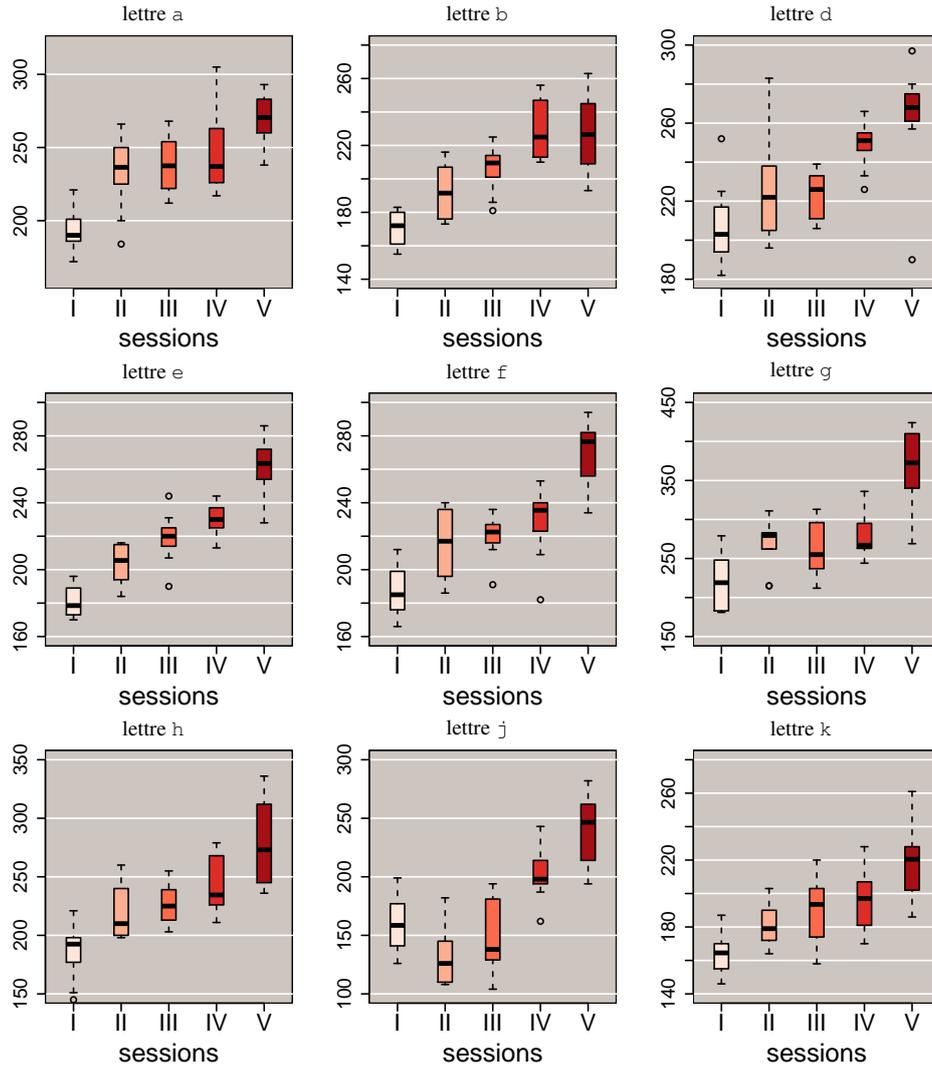


FIGURE 5.49: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

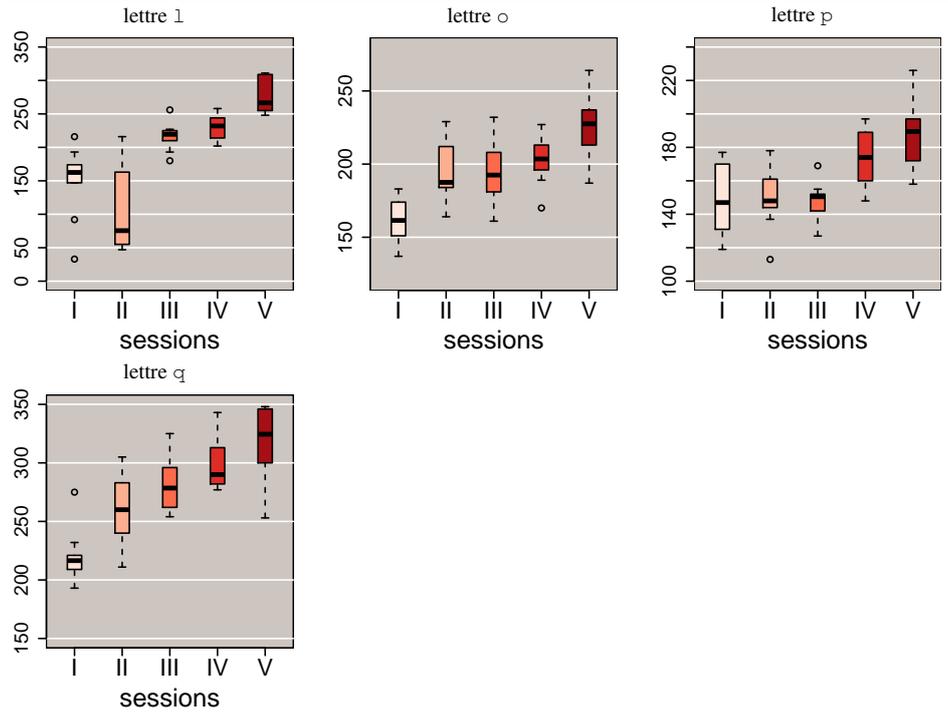


FIGURE 5.50: Boîtes à moustache des largeurs  $W$  pour les caractères *normaux* selon les pages  $I - V$ .

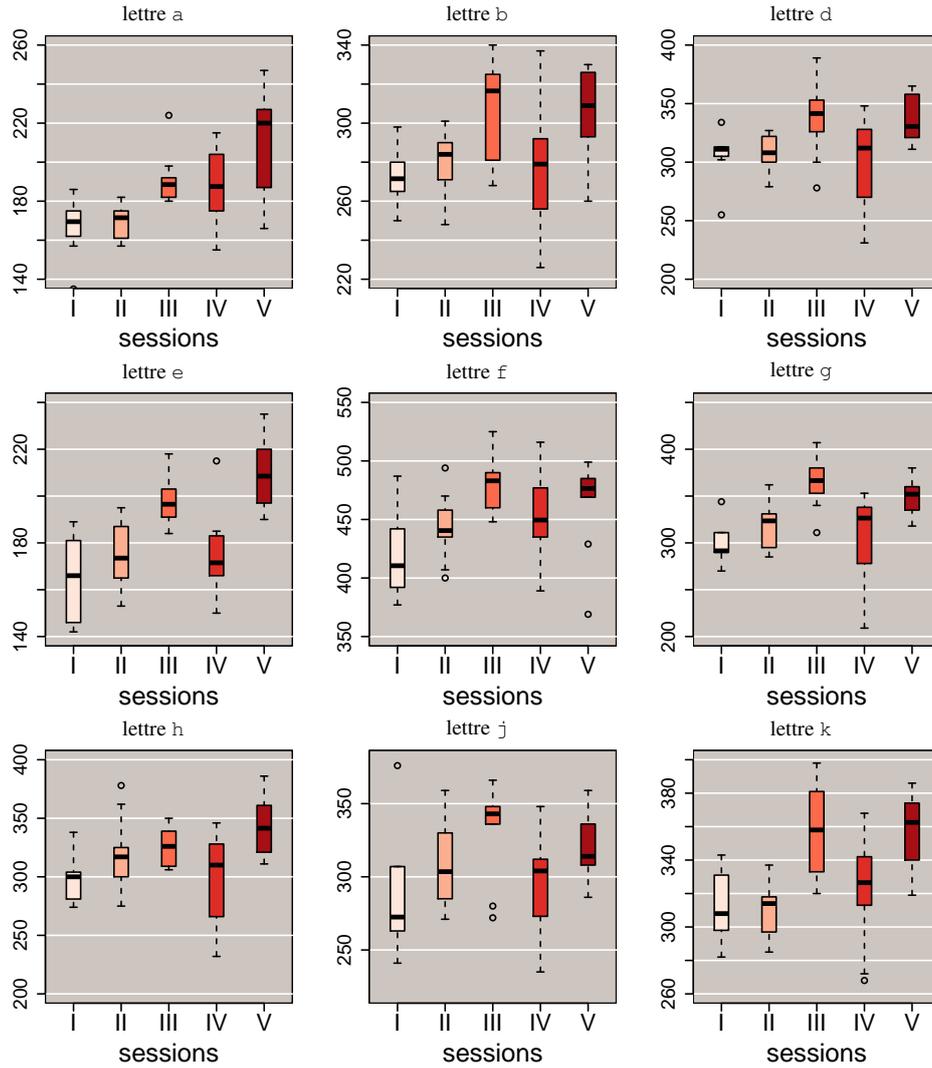


FIGURE 5.51: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

**Scripteur 1, hauteur H, caractères *normaux***

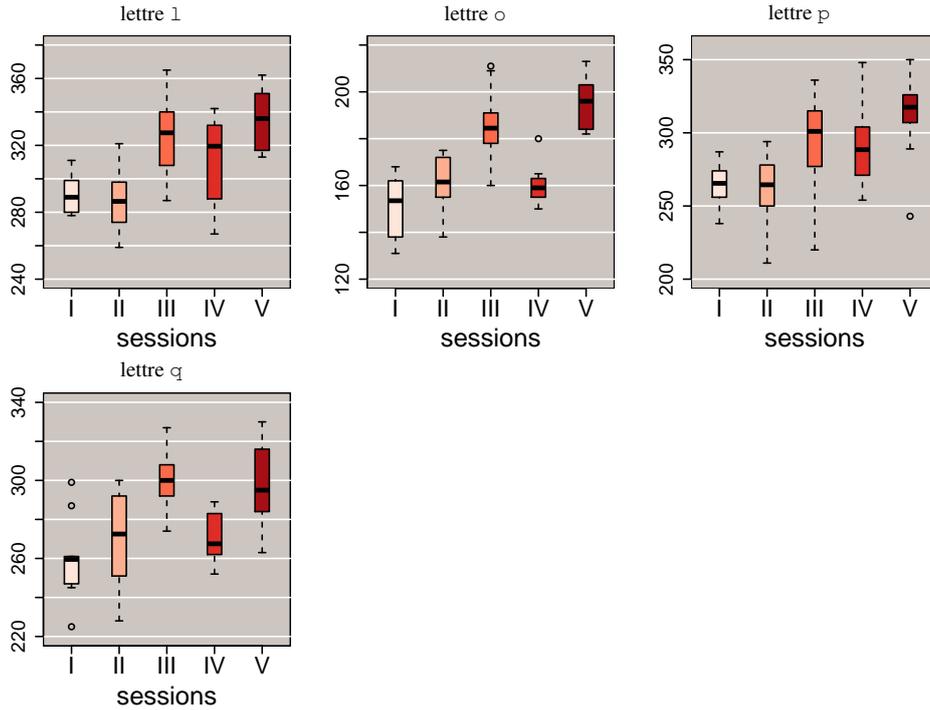


FIGURE 5.52: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

**Scripteur 2, hauteur H, caractères *normaux***

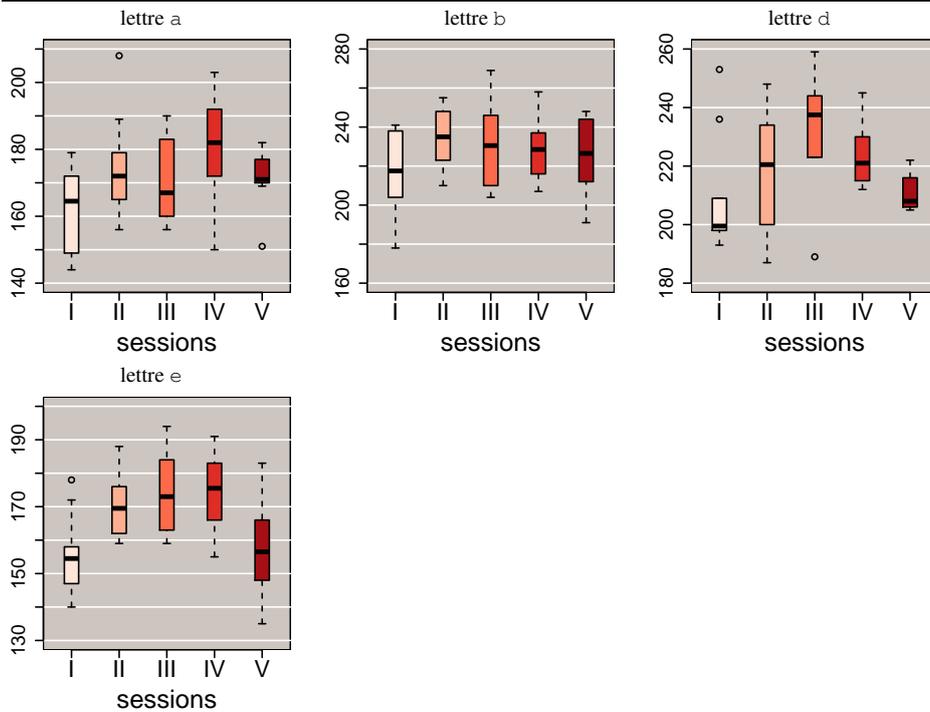


FIGURE 5.53: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

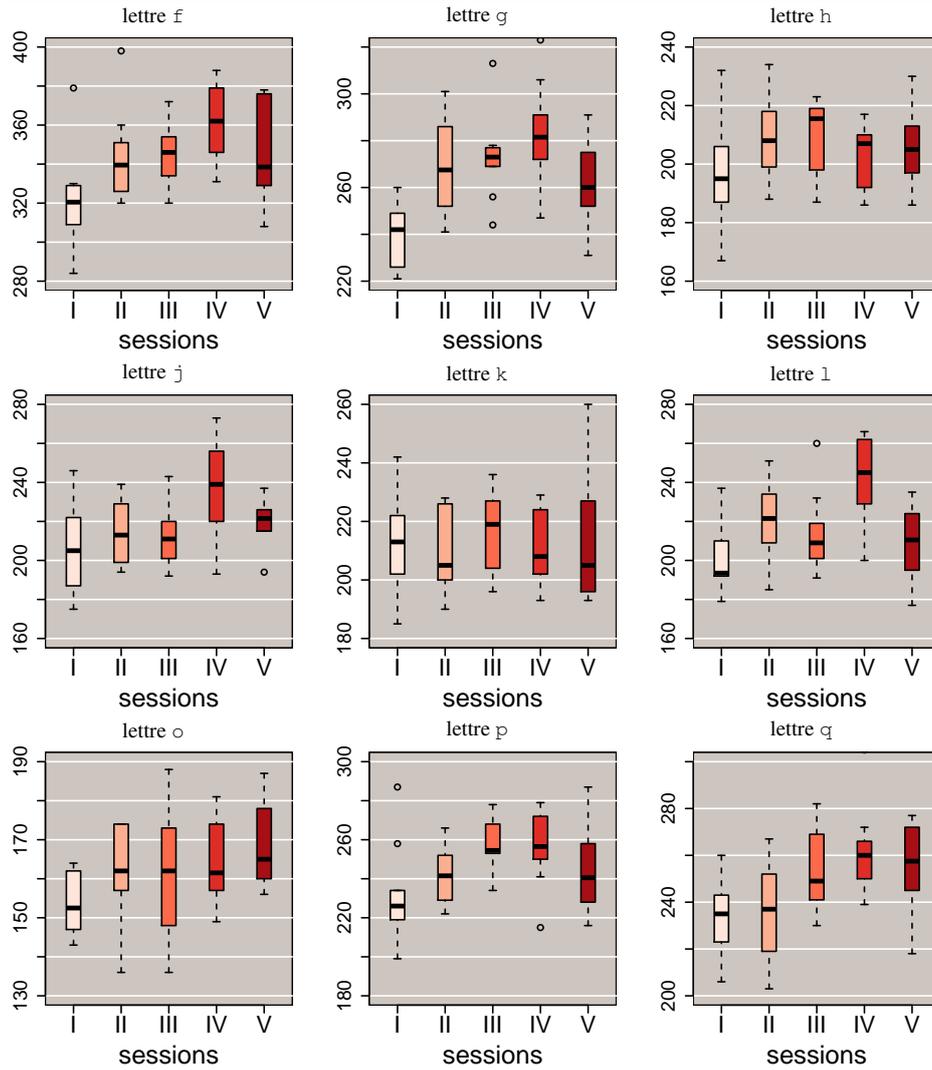


FIGURE 5.54: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

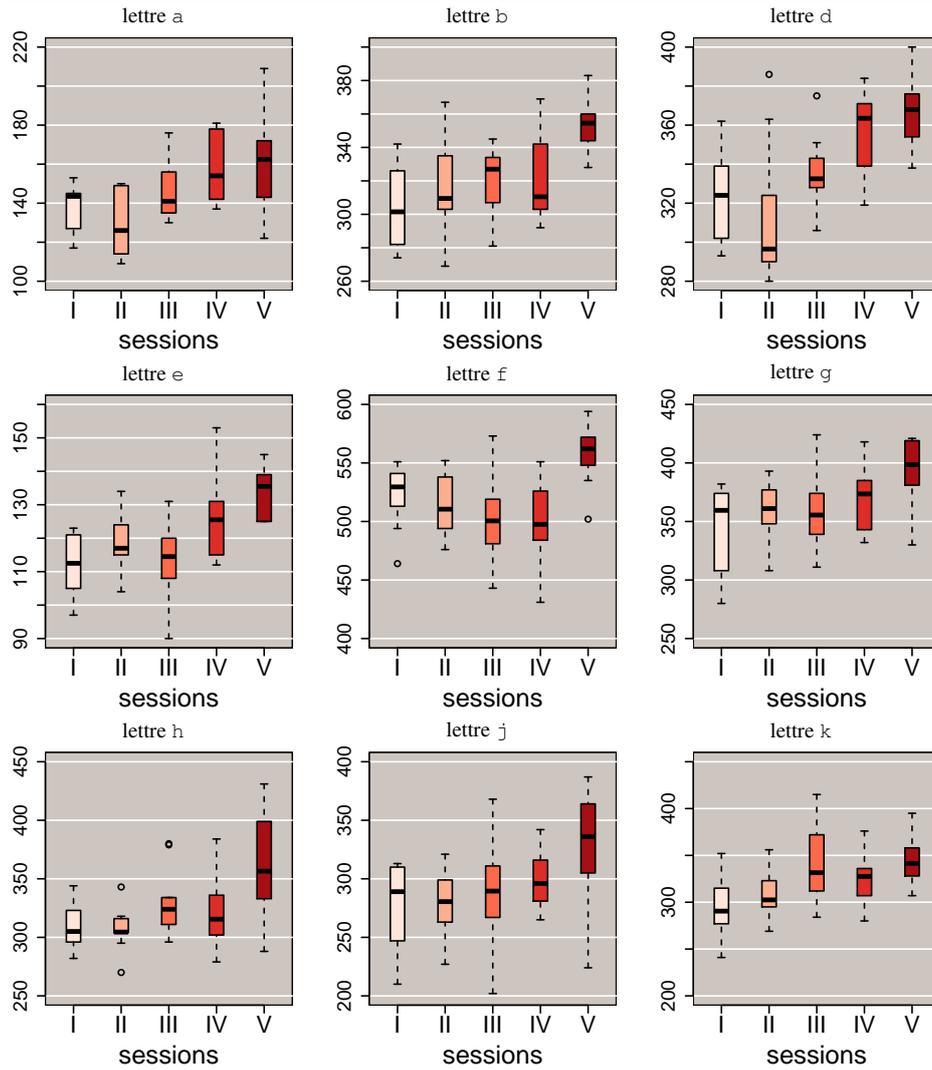


FIGURE 5.55: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 3, hauteur  $H$ , caractères *normaux*

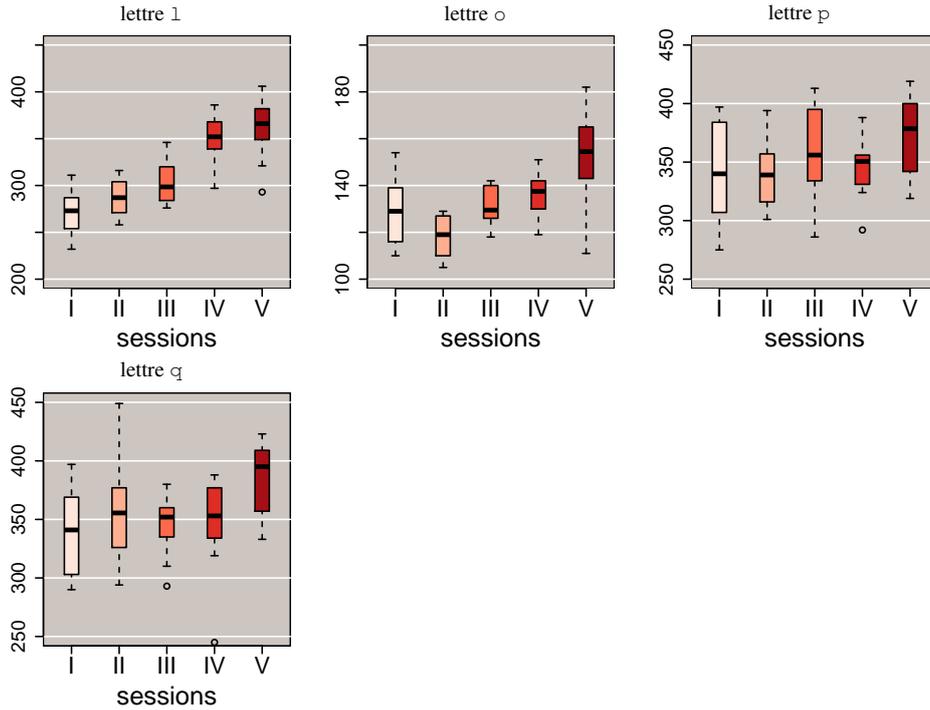


FIGURE 5.56: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 4, hauteur  $H$ , caractères *normaux*

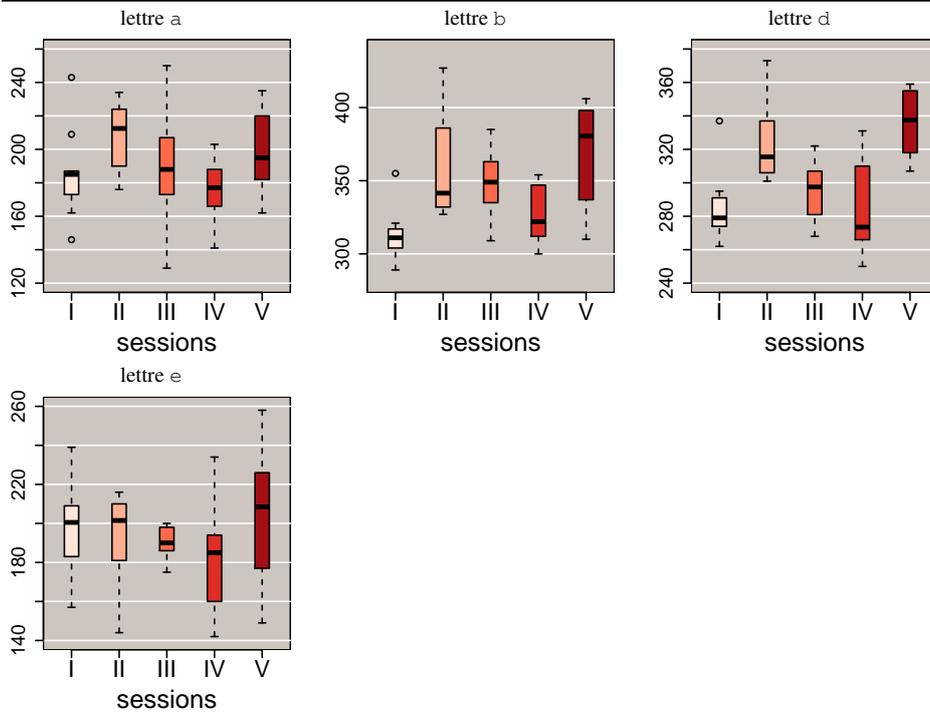


FIGURE 5.57: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

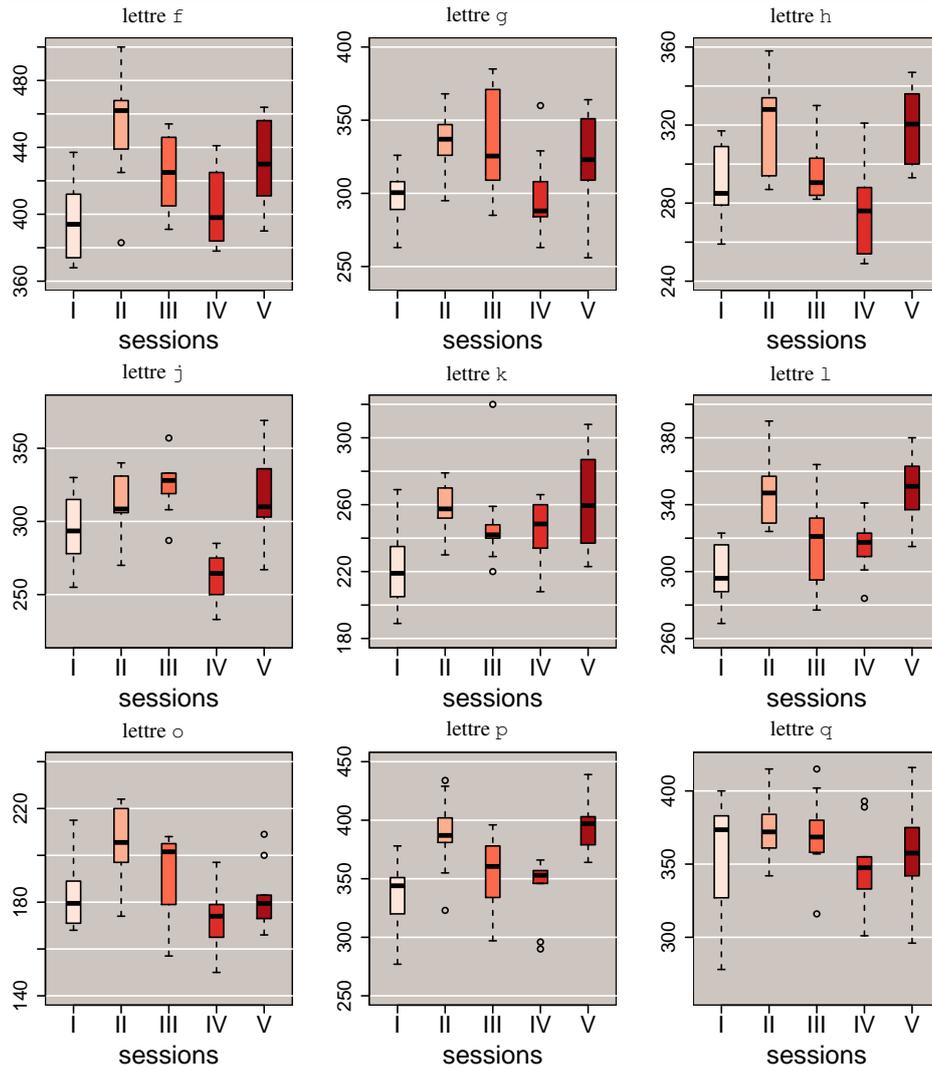


FIGURE 5.58: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

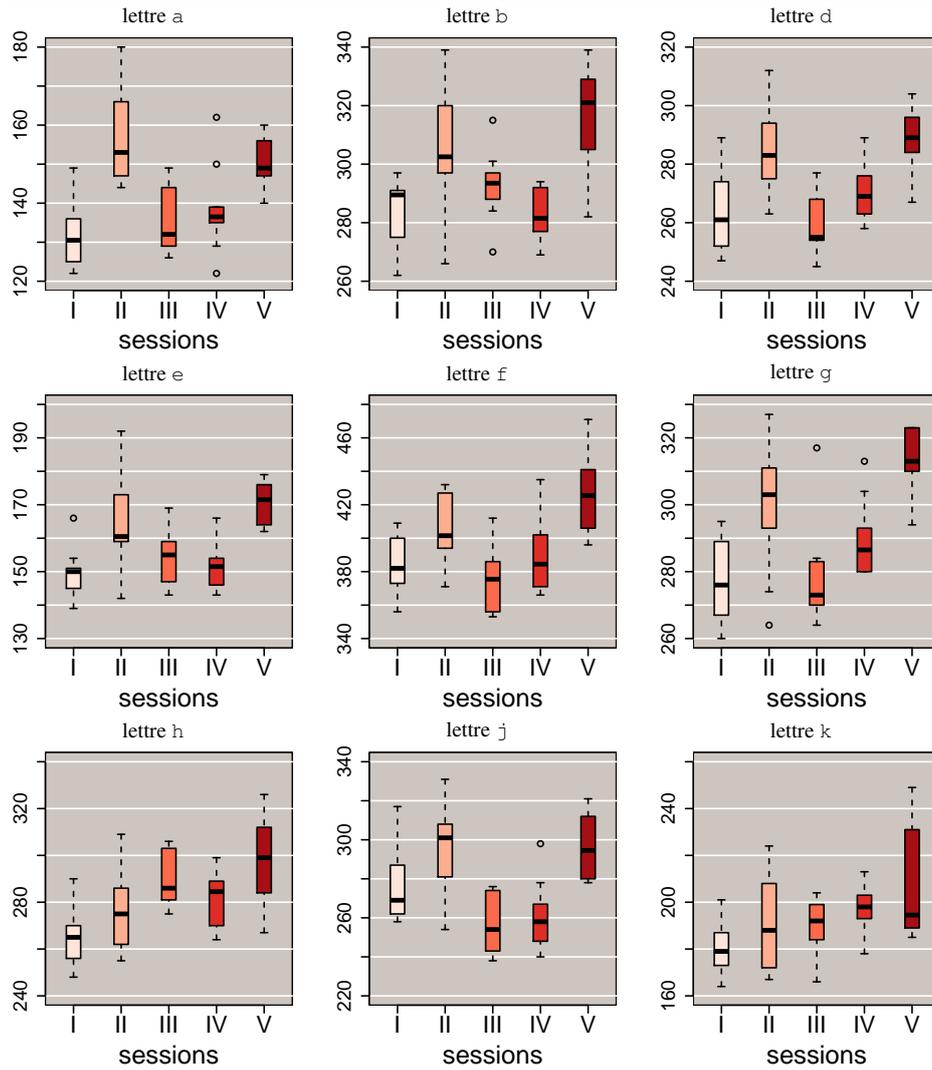


FIGURE 5.59: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 5, hauteur H, caractères *normaux*

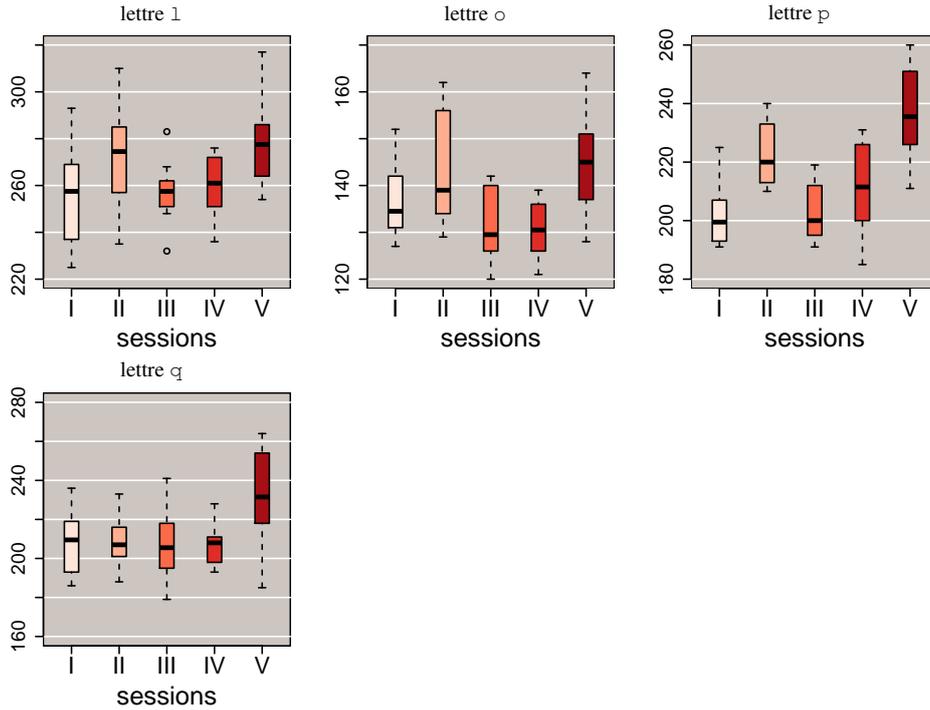


FIGURE 5.60: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 6, hauteur H, caractères *normaux*

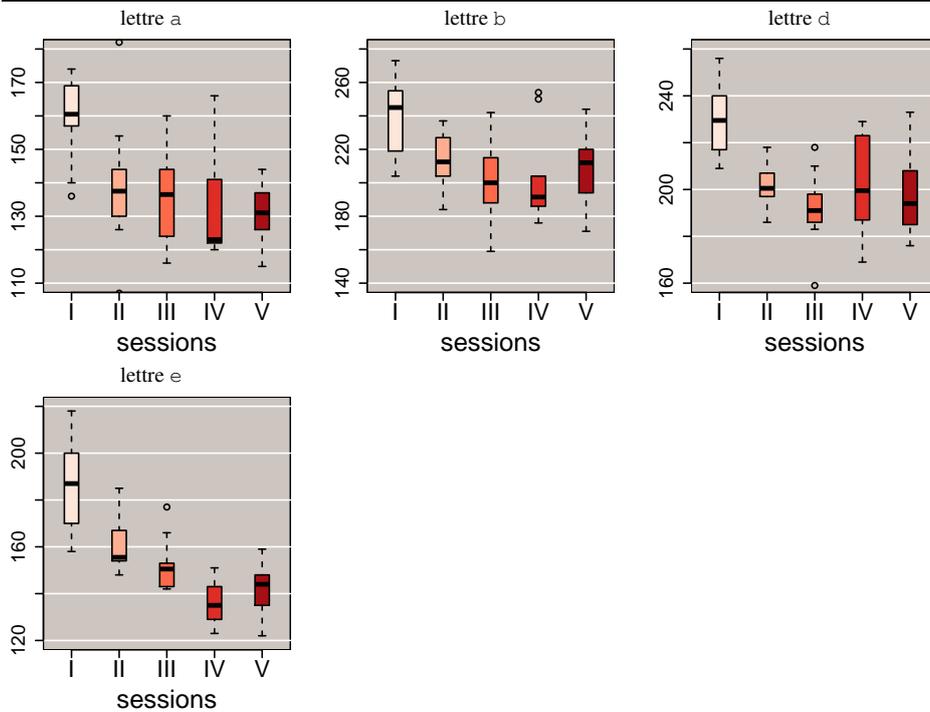


FIGURE 5.61: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

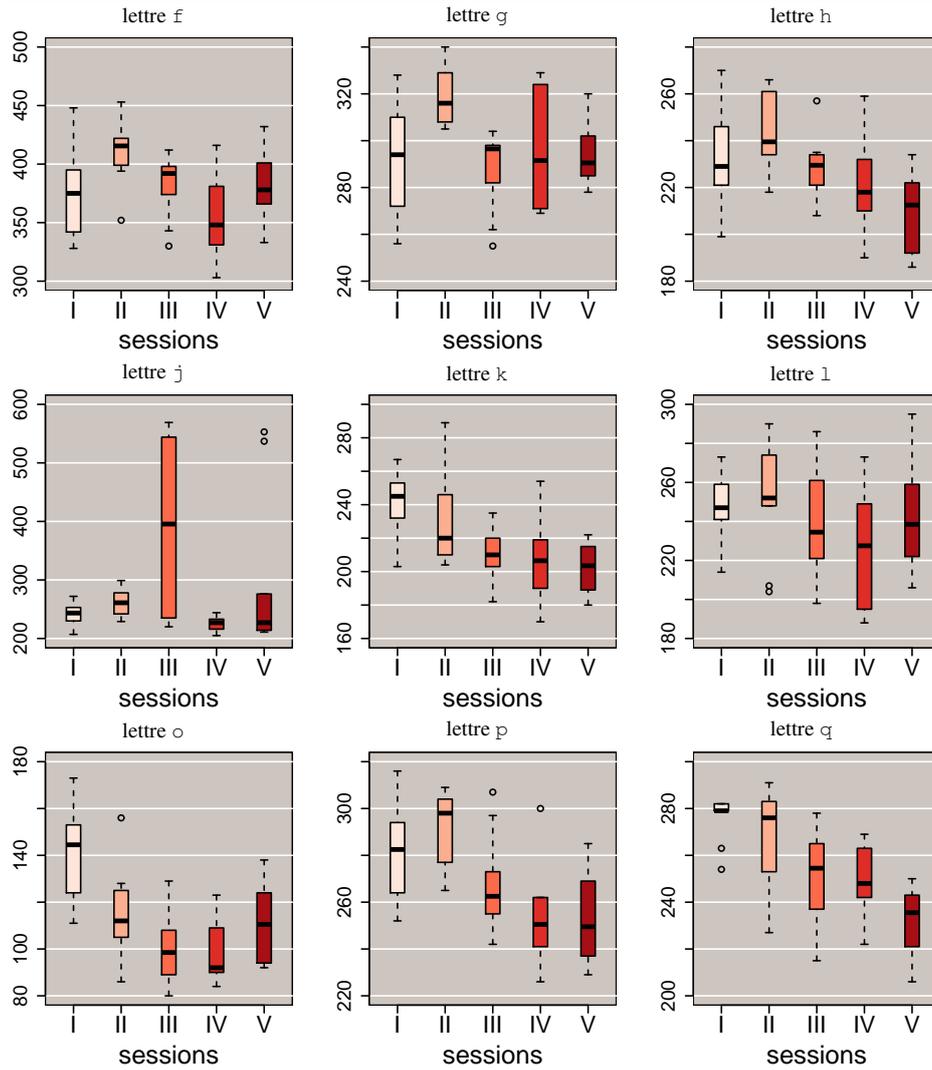


FIGURE 5.62: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

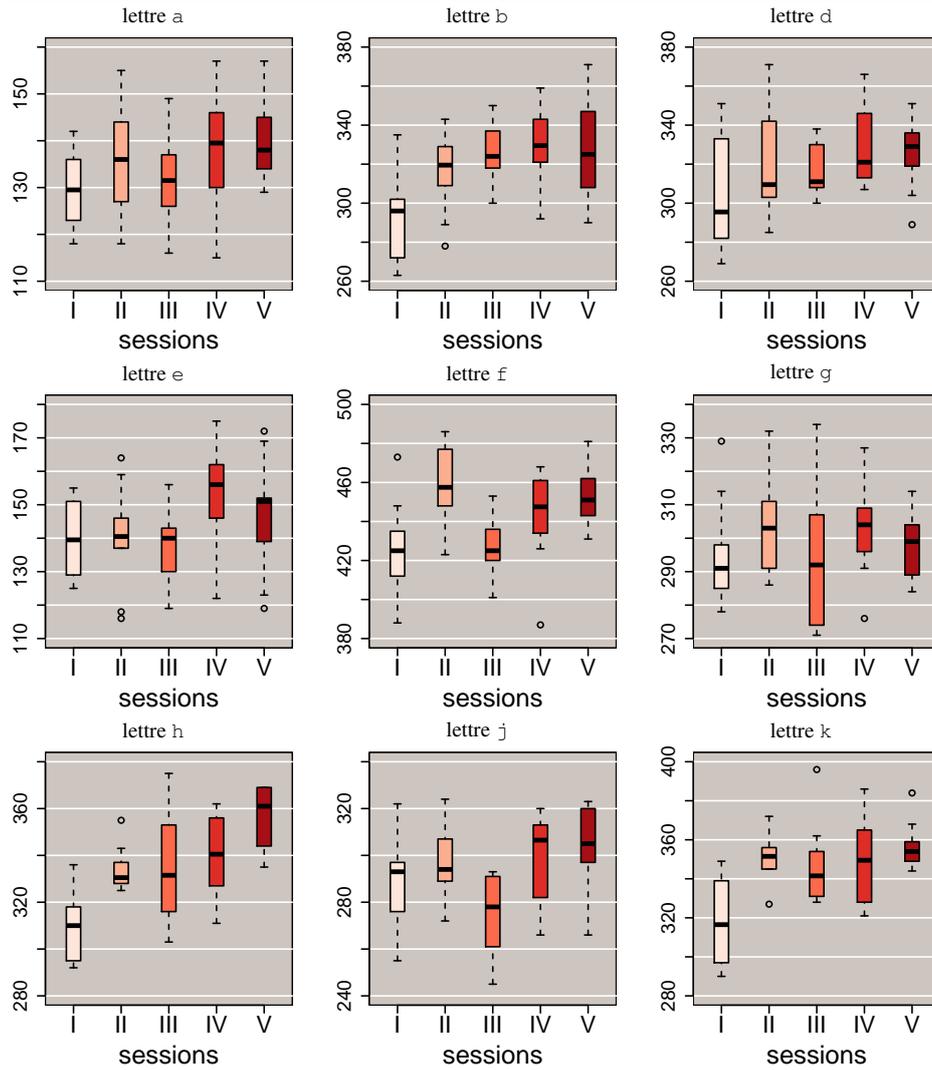


FIGURE 5.63: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 7, hauteur H, caractères *normaux*

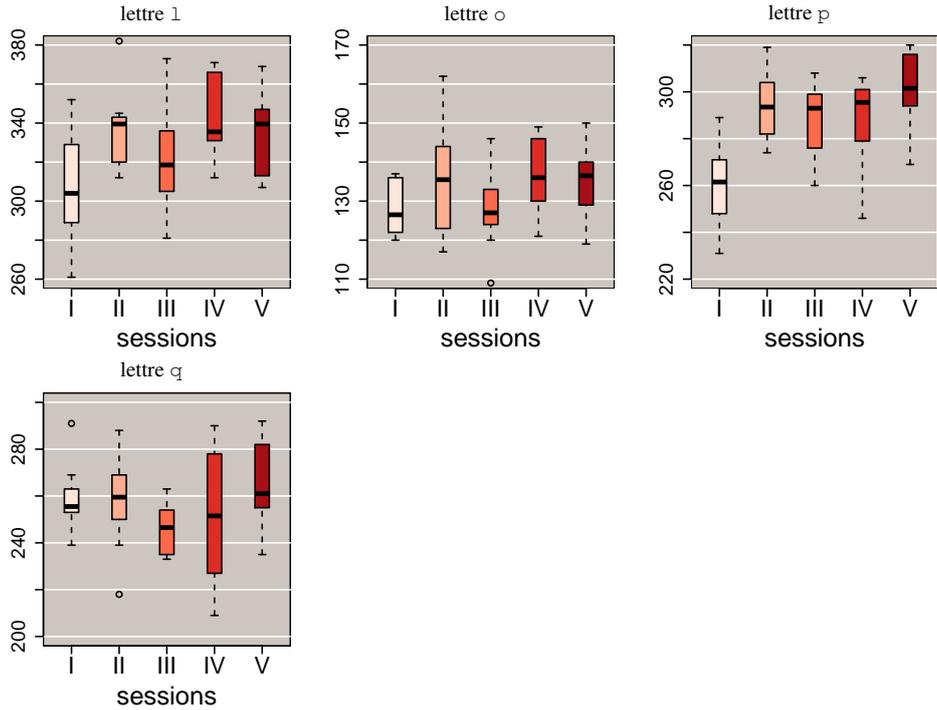


FIGURE 5.64: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 8, hauteur H, caractères *normaux*

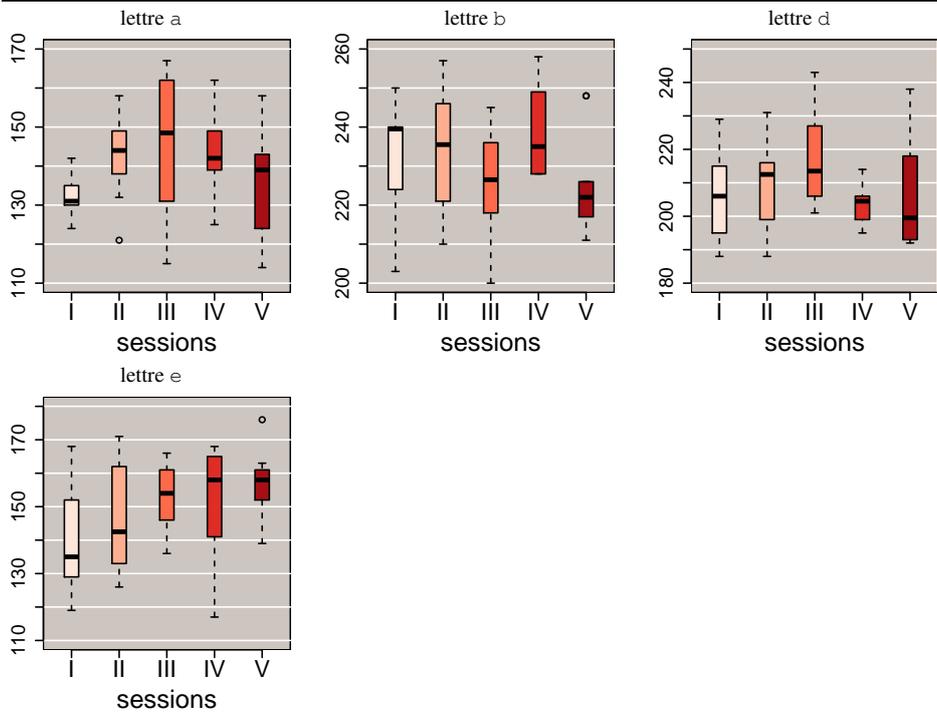


FIGURE 5.65: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

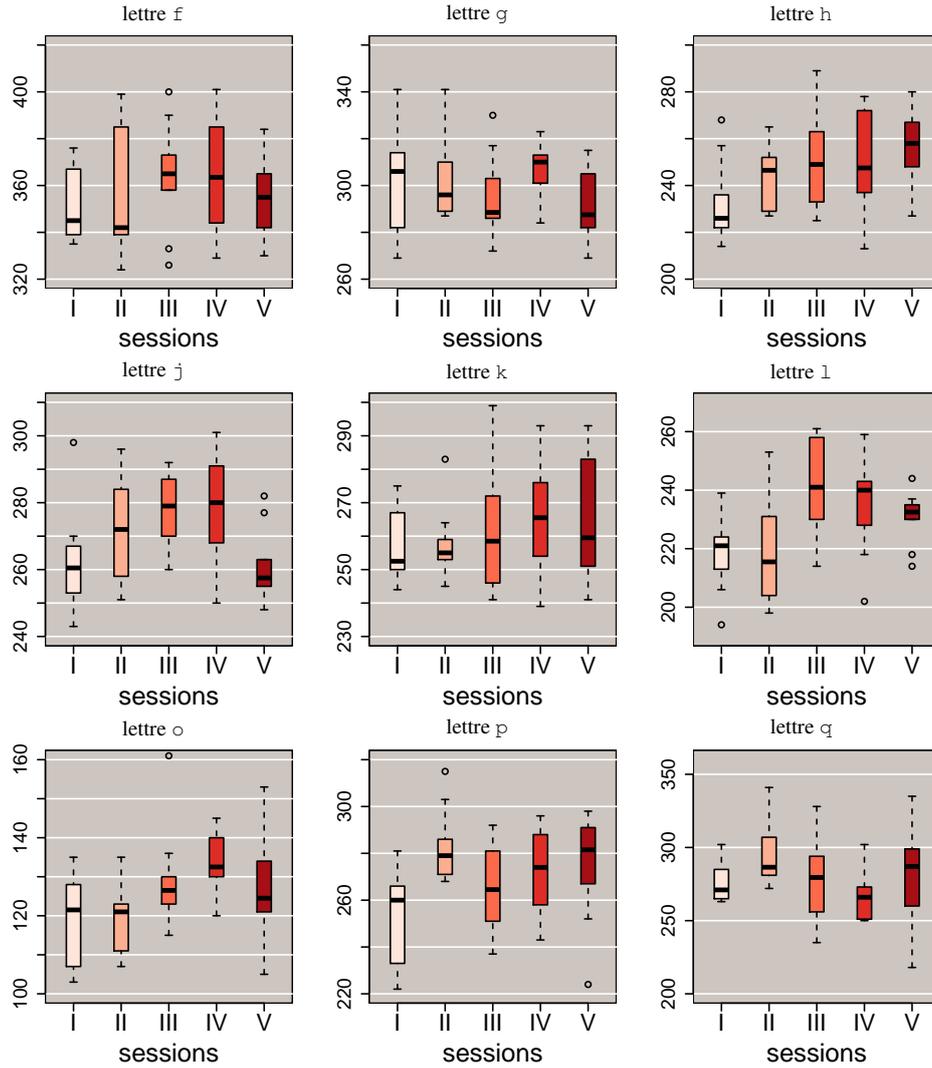


FIGURE 5.66: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

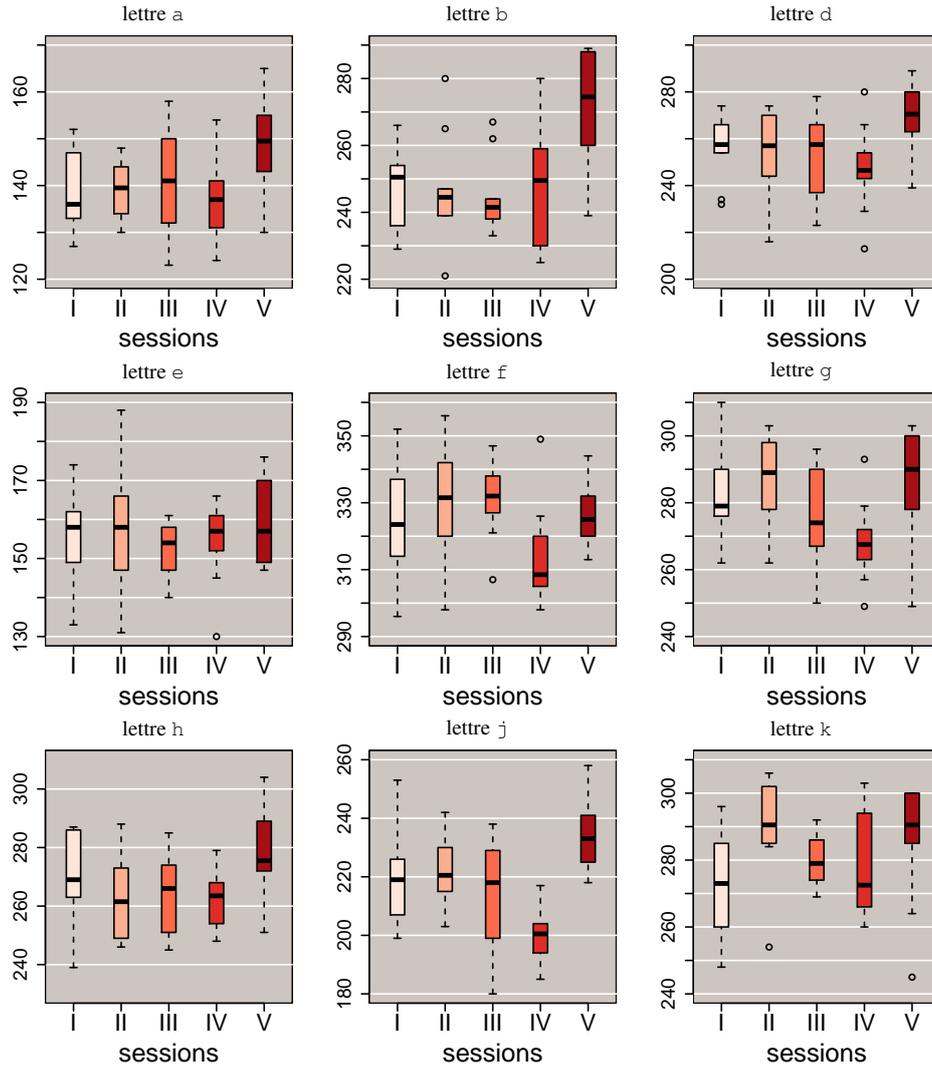


FIGURE 5.67: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 9, hauteur  $H$ , caractères *normaux*

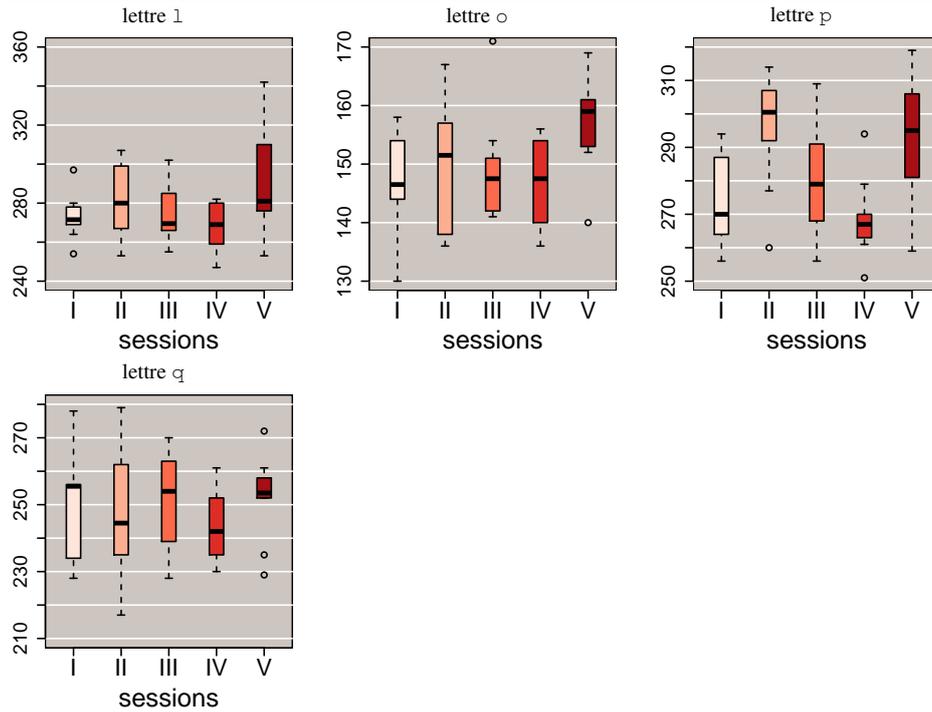


FIGURE 5.68: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

Scripteur 10, hauteur  $H$ , caractères *normaux*

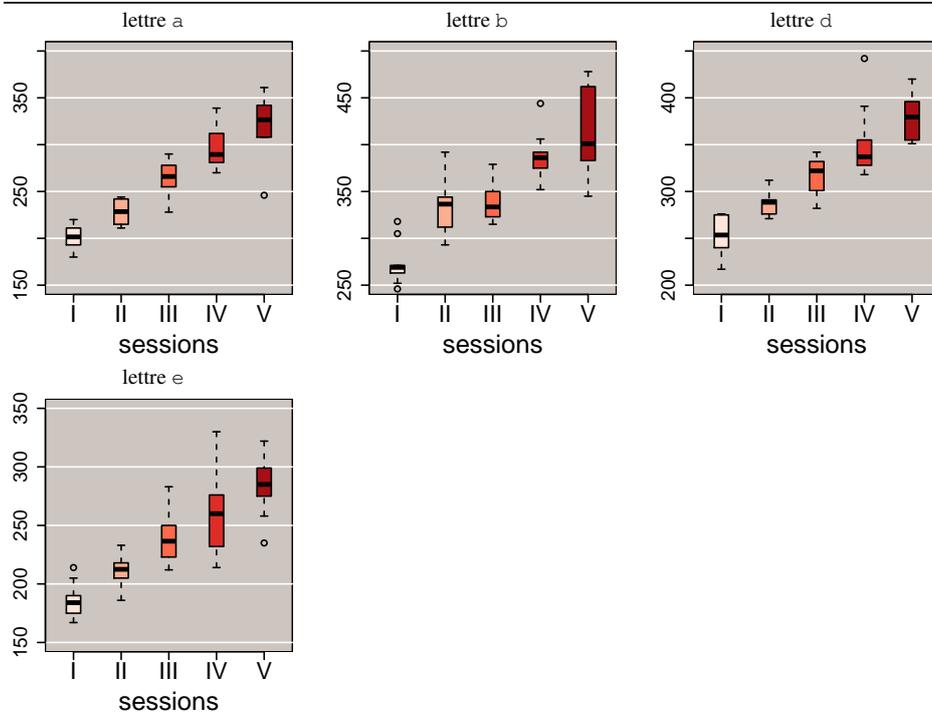


FIGURE 5.69: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

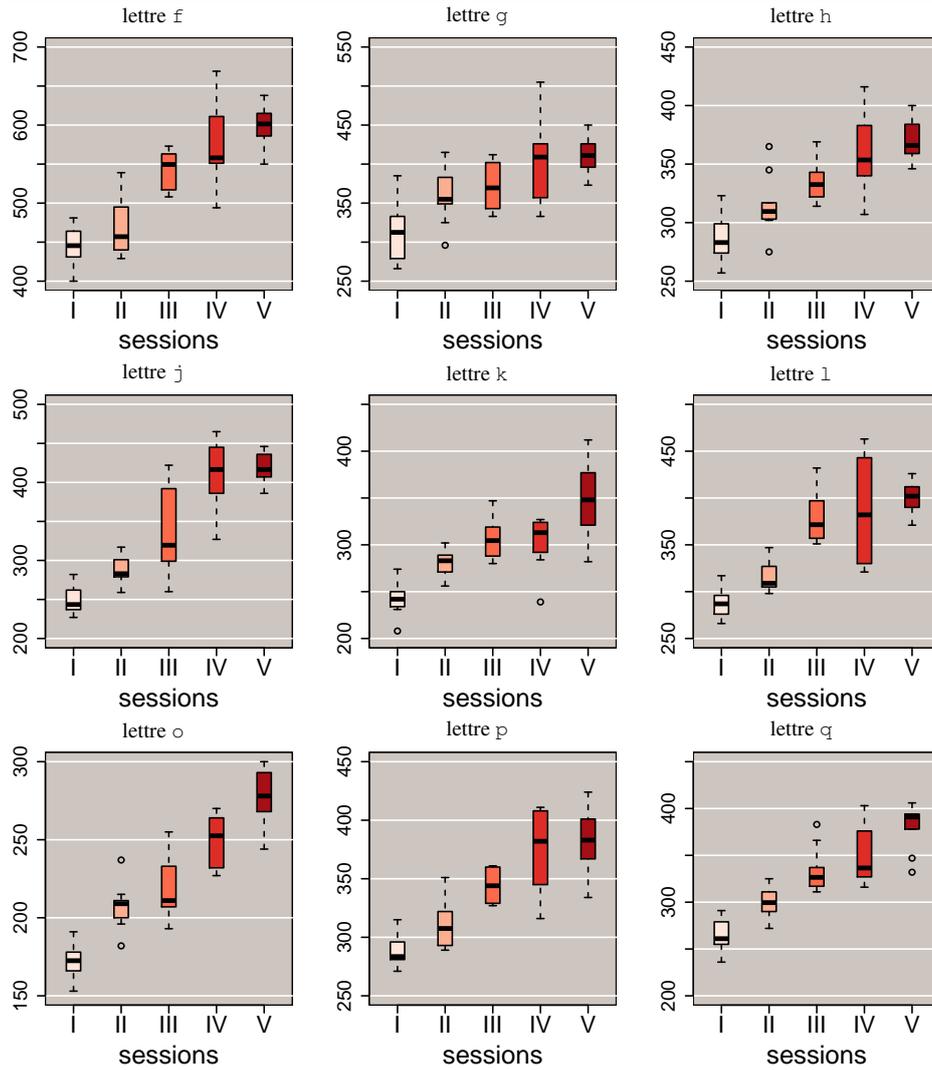


FIGURE 5.70: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

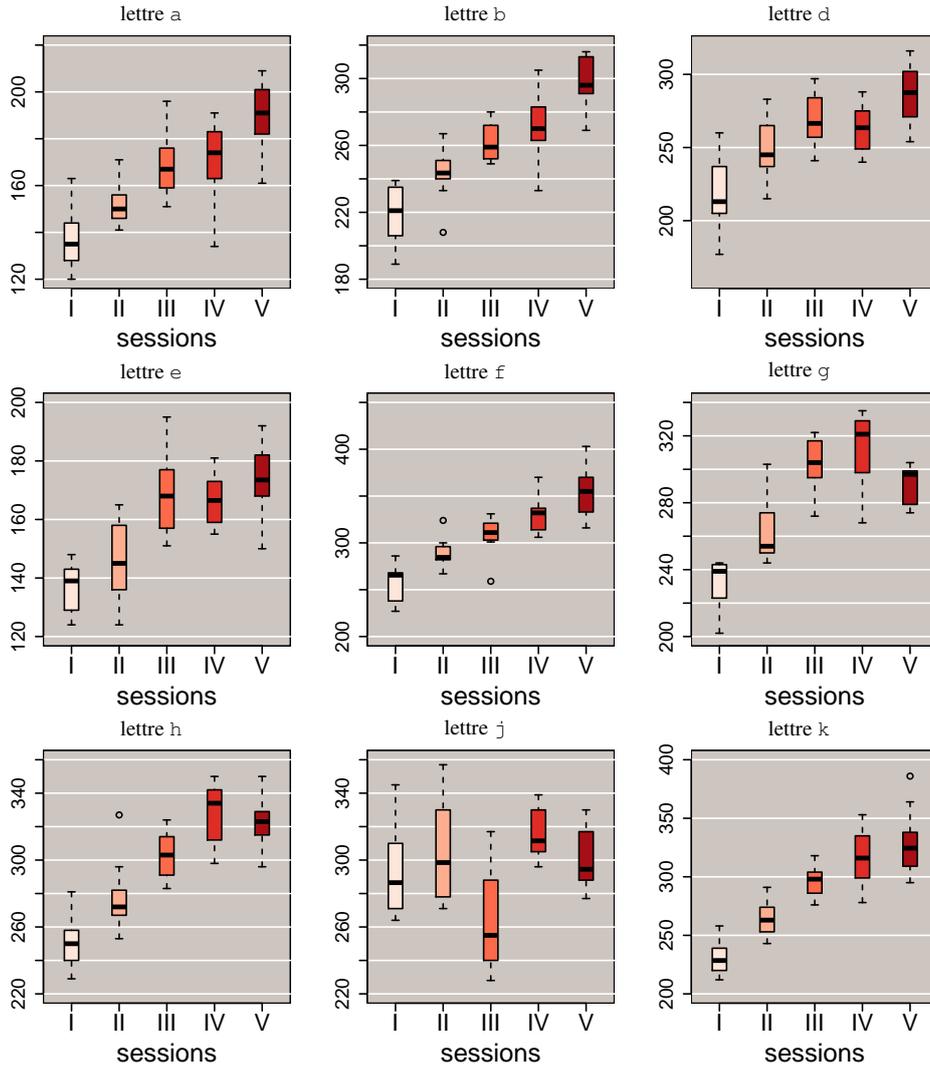


FIGURE 5.71: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

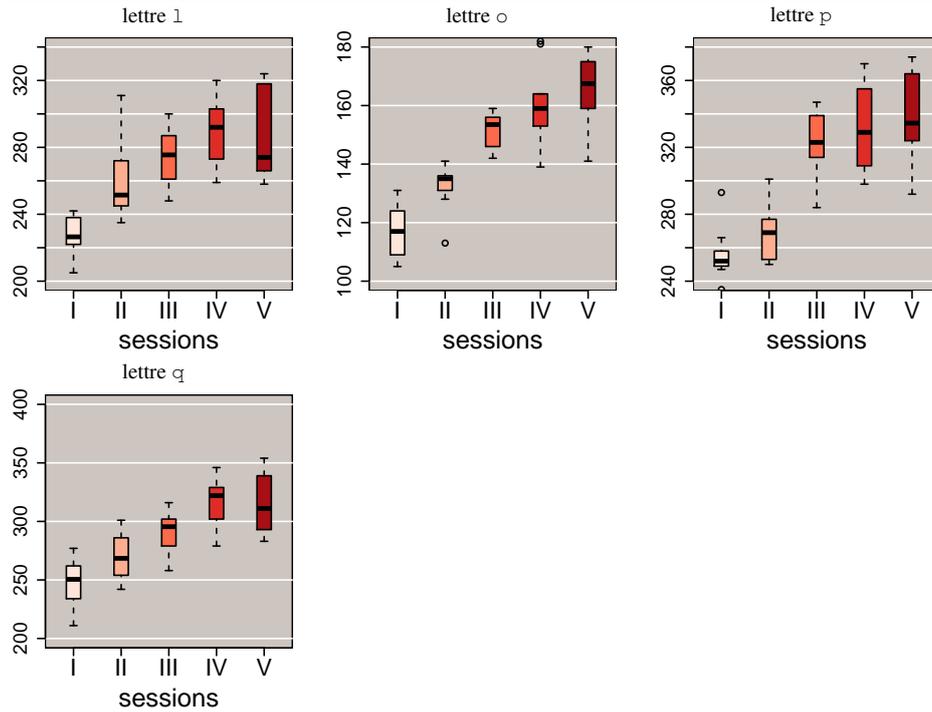


FIGURE 5.72: Boîtes à moustache des hauteurs  $H$  pour les caractères *normaux* selon les pages  $I - V$ .

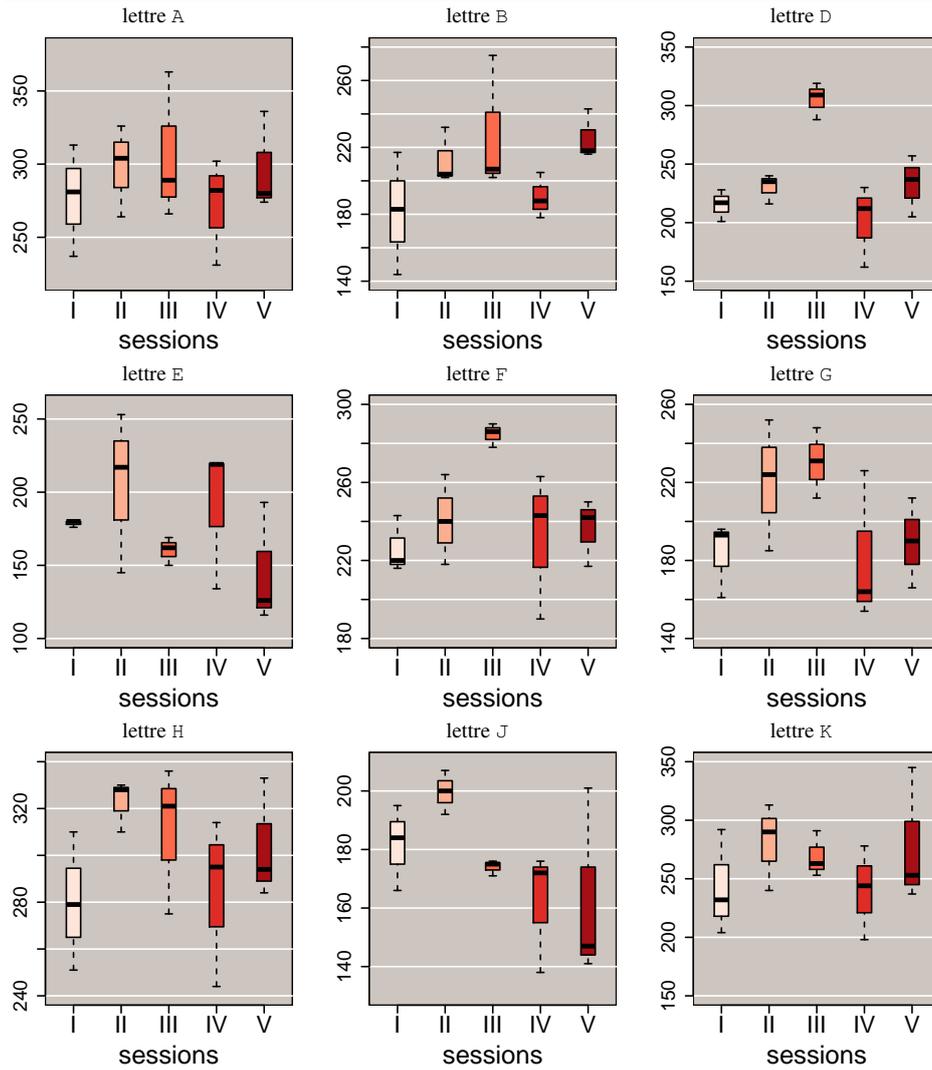


FIGURE 5.73: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 1, largeur W, caractères grands

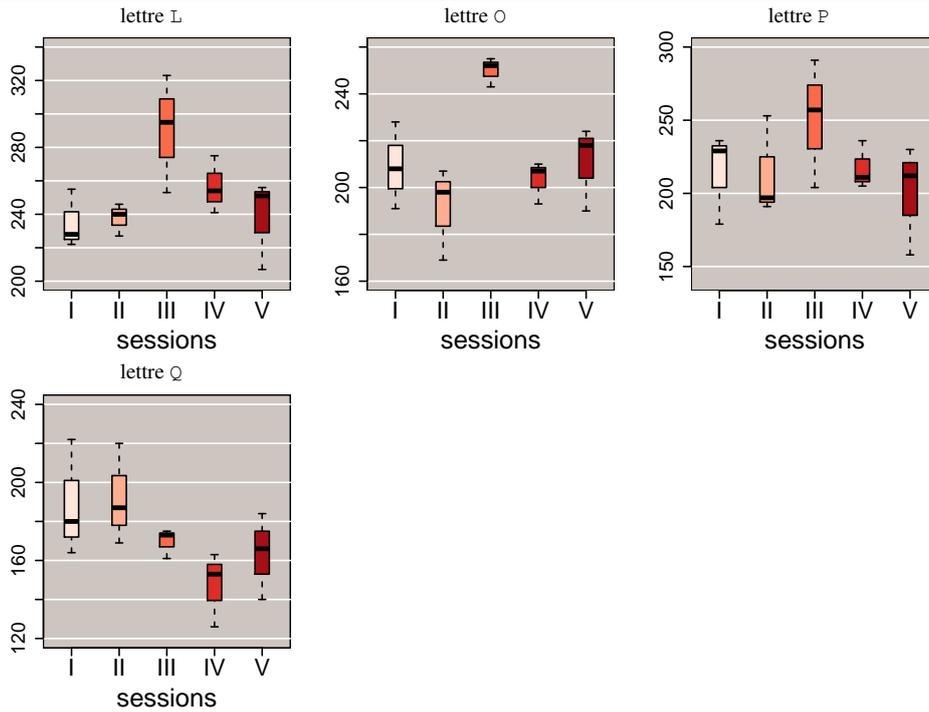


FIGURE 5.74: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 2, largeur W, caractères grands

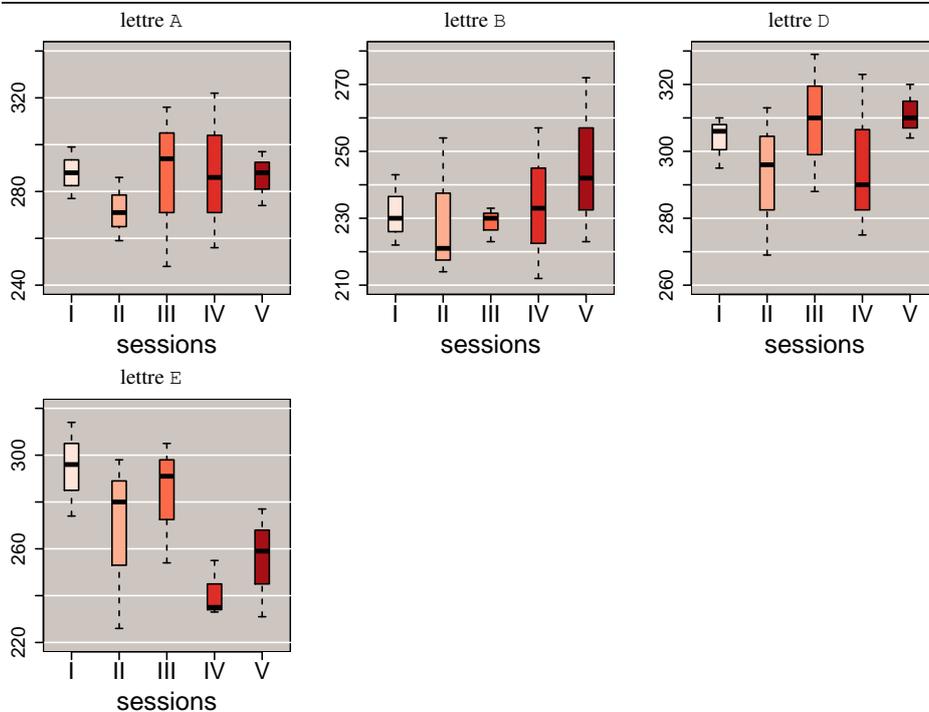


FIGURE 5.75: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

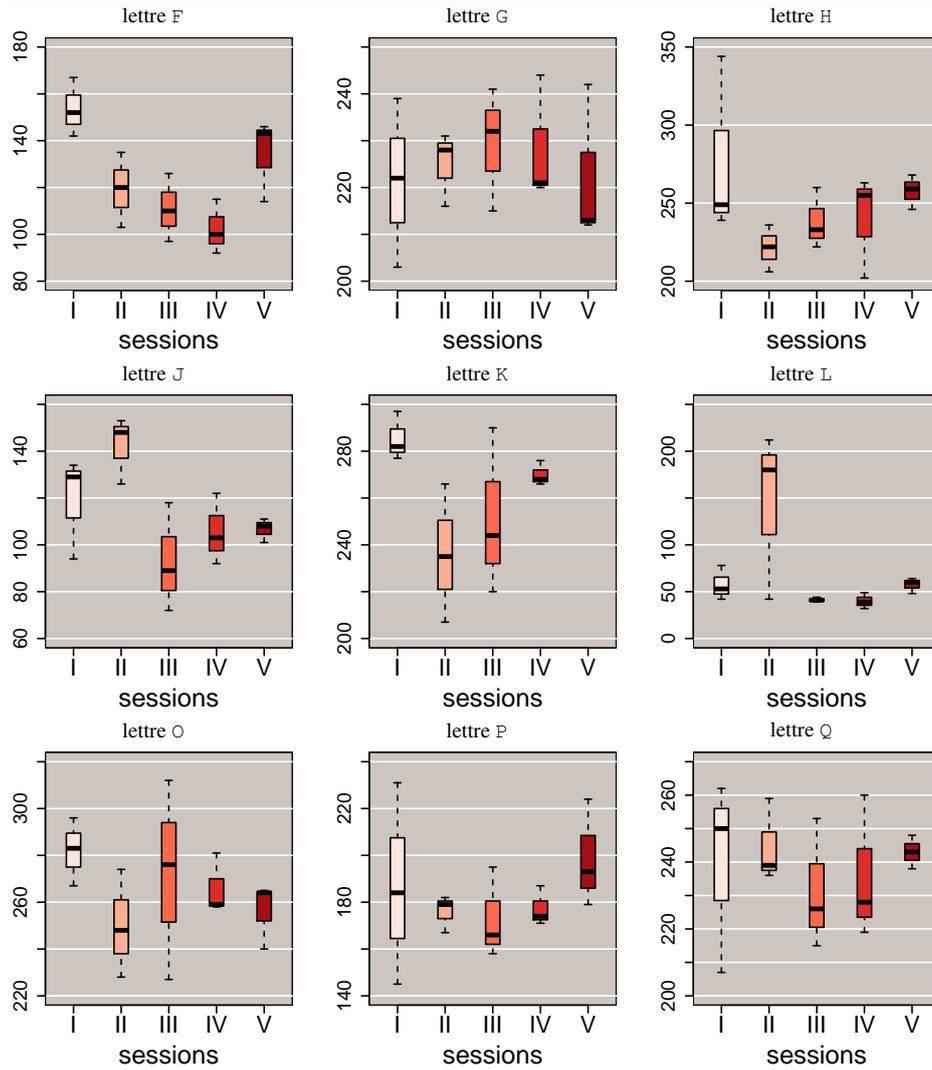


FIGURE 5.76: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

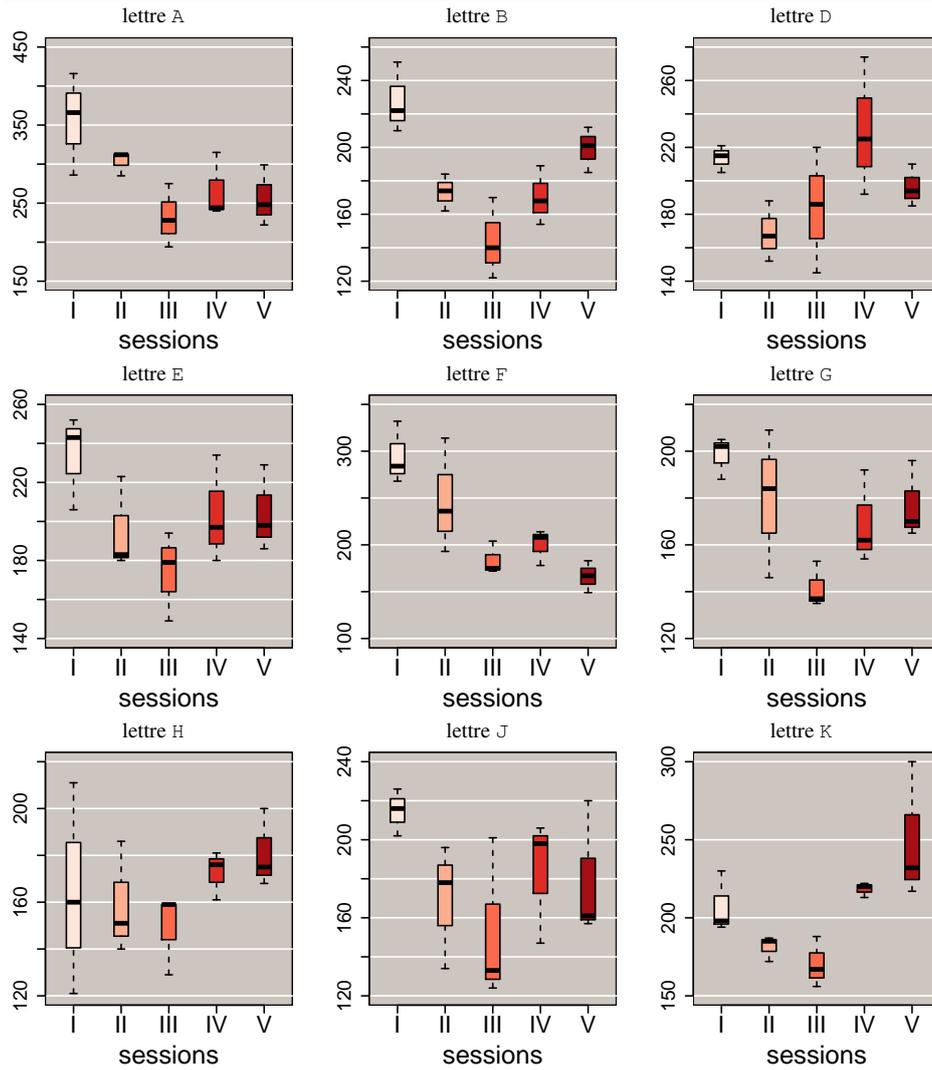


FIGURE 5.77: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 3, largeur W, caractères *grands*

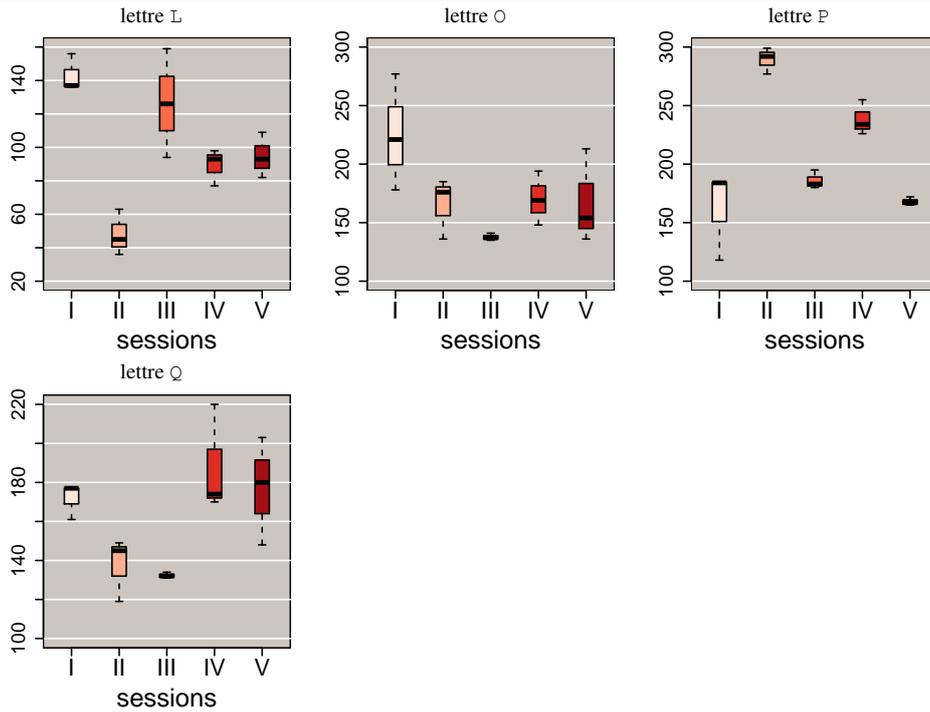


FIGURE 5.78: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages I – V.

Scripteur 4, largeur W, caractères *grands*

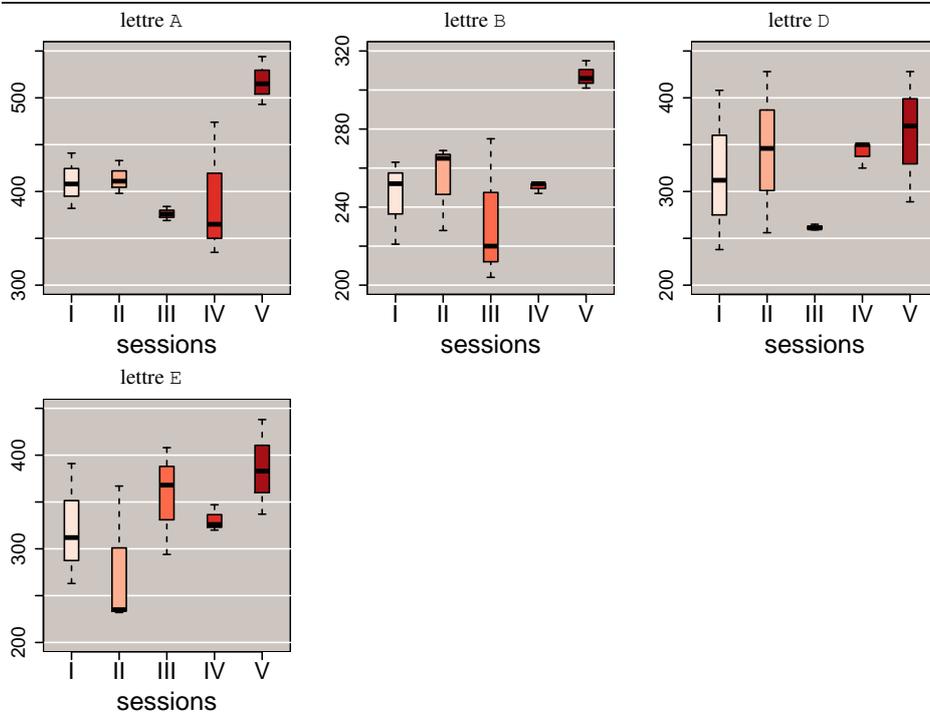


FIGURE 5.79: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages I – V.

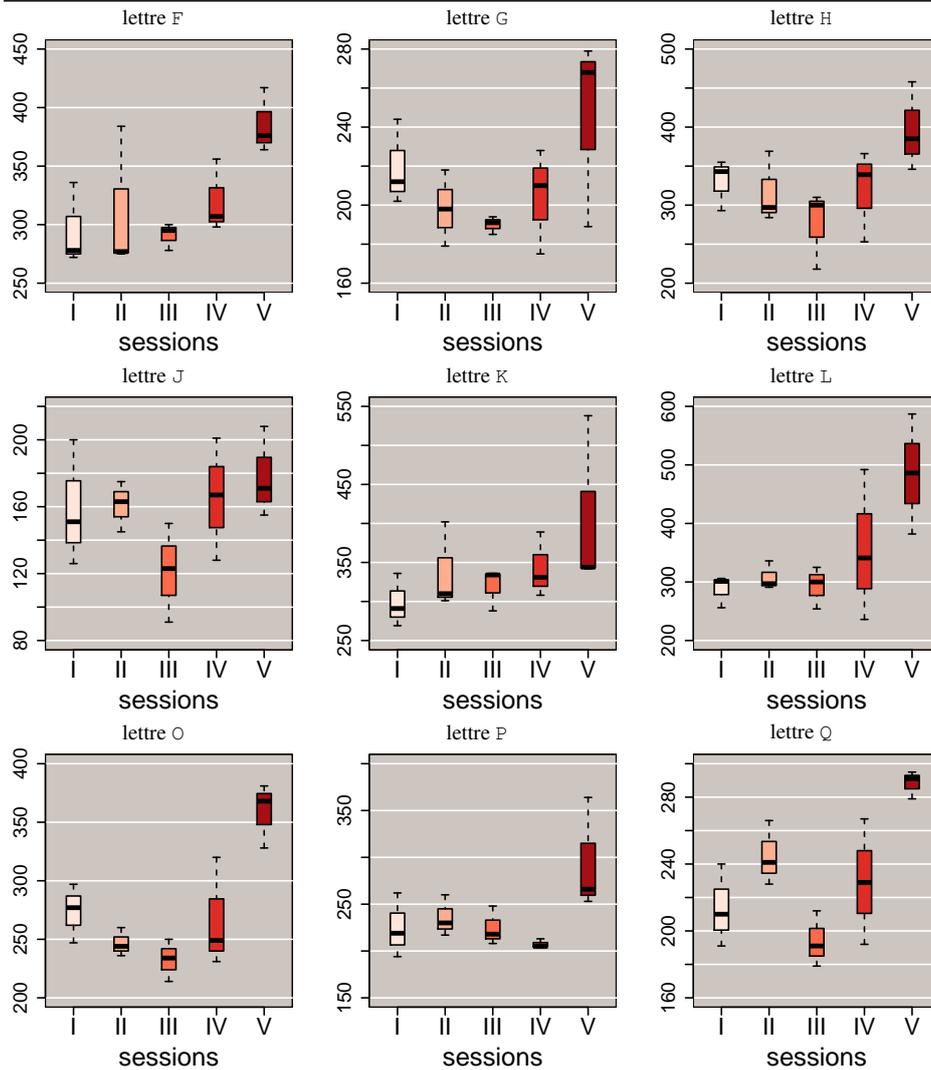


FIGURE 5.80: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

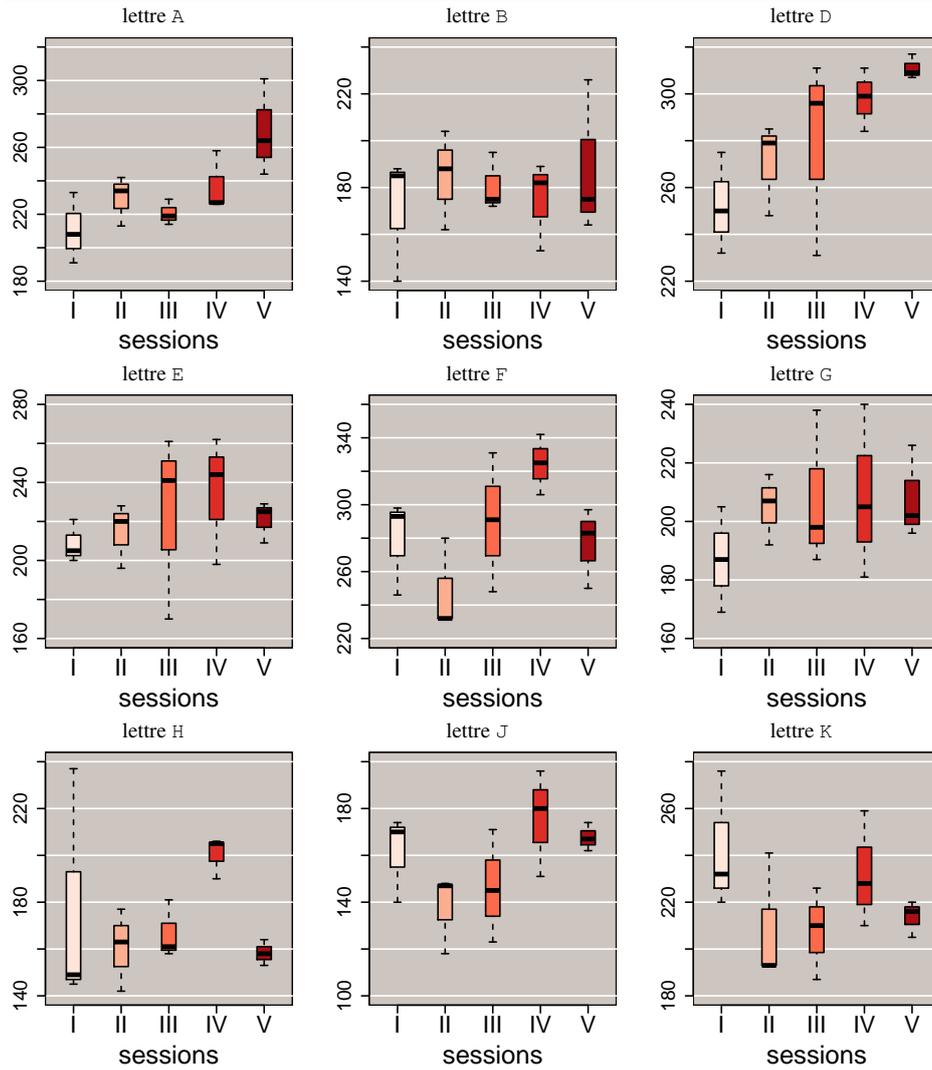


FIGURE 5.81: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 5, largeur W, caractères grands

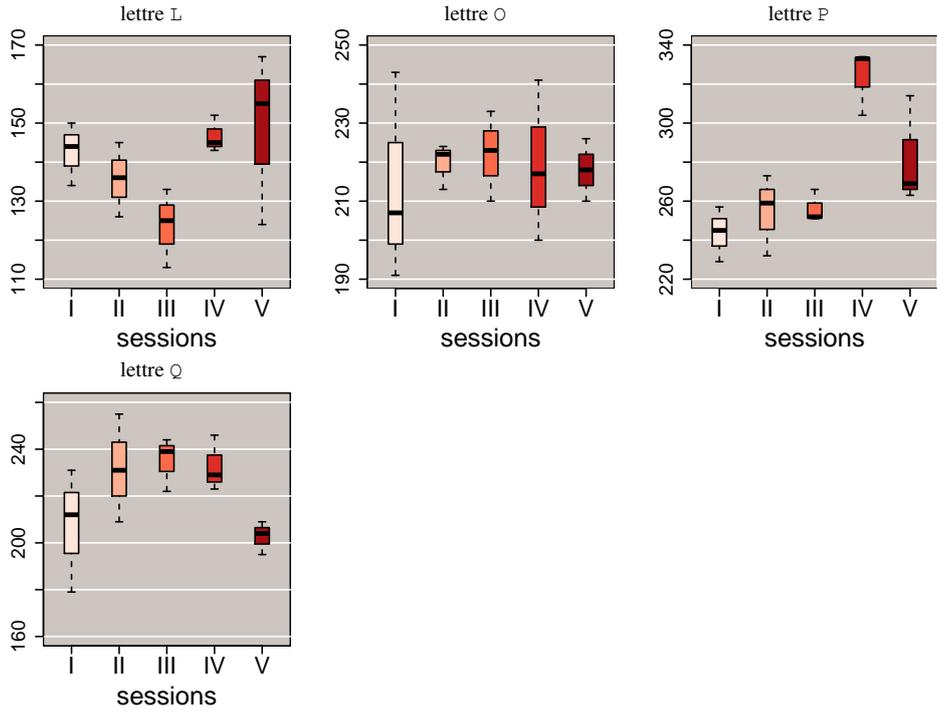


FIGURE 5.82: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages I – V.

Scripteur 6, largeur W, caractères grands

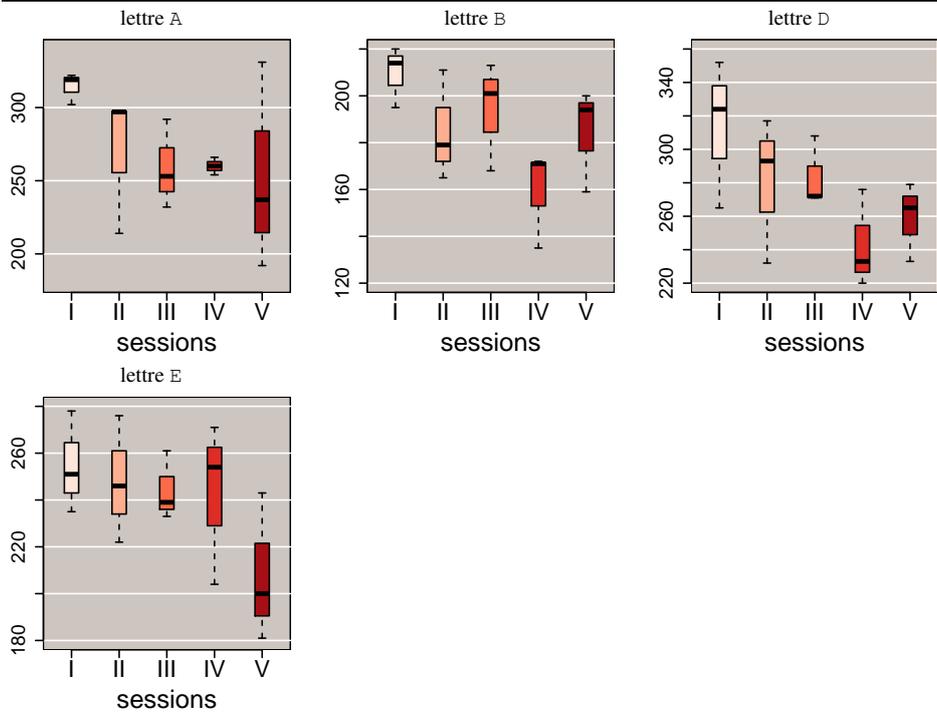


FIGURE 5.83: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages I – V.

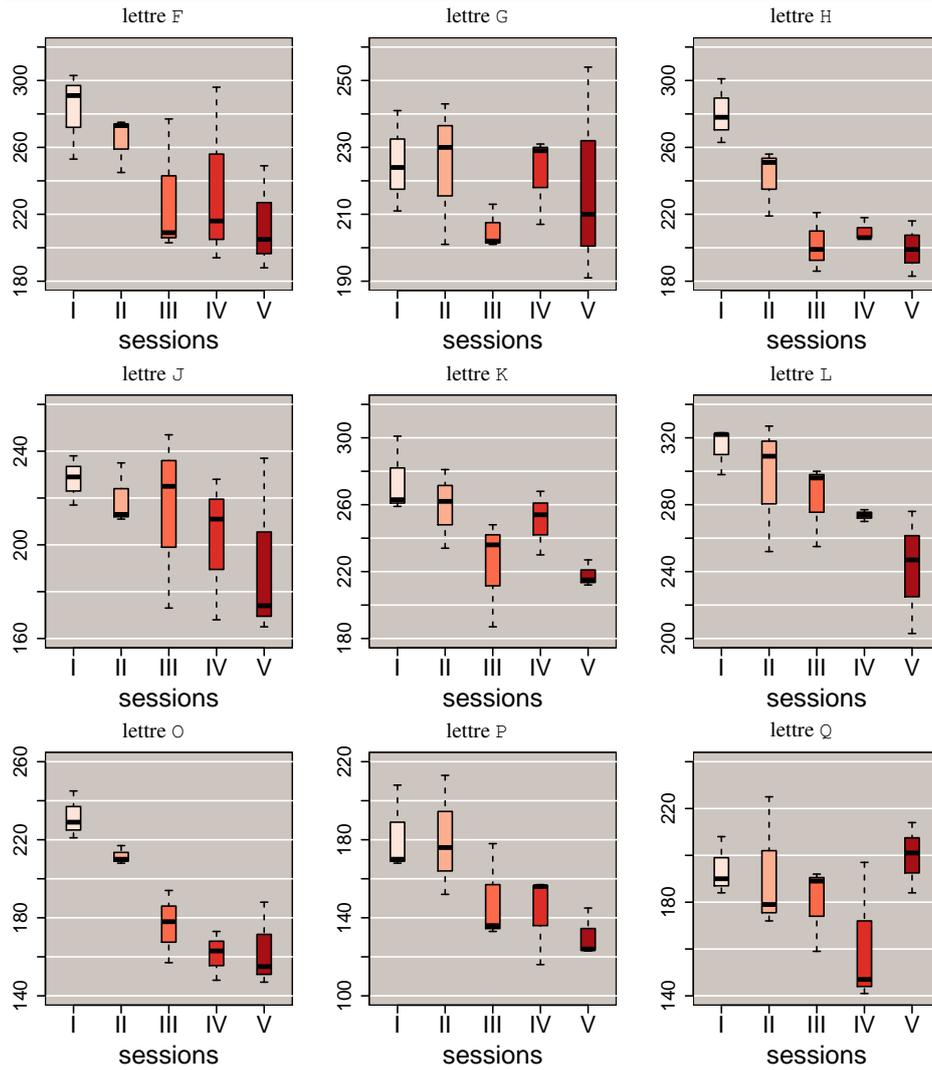


FIGURE 5.84: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

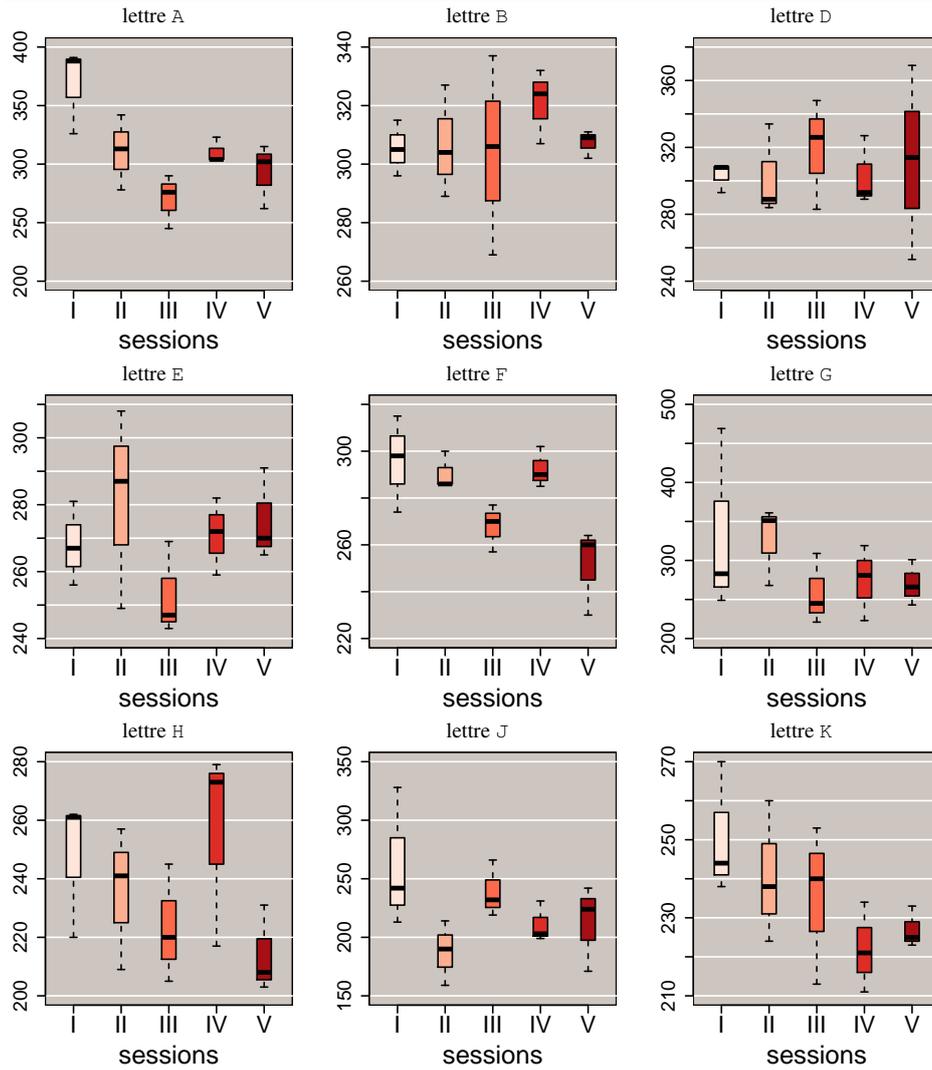


FIGURE 5.85: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

**Scripteur 7, largeur W, caractères grands**

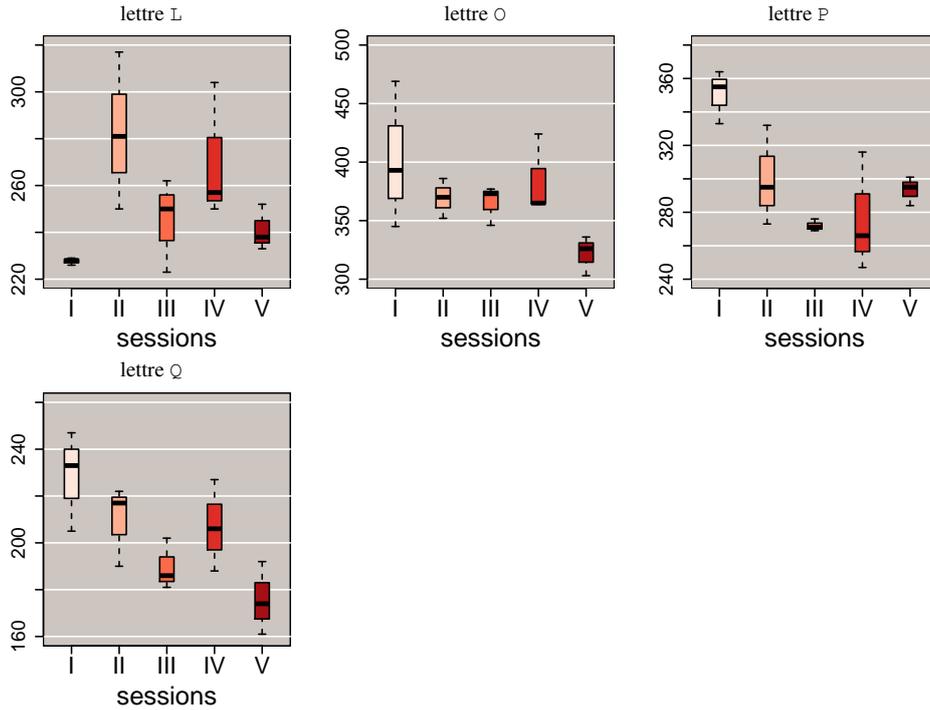


FIGURE 5.86: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

**Scripteur 8, largeur W, caractères grands**

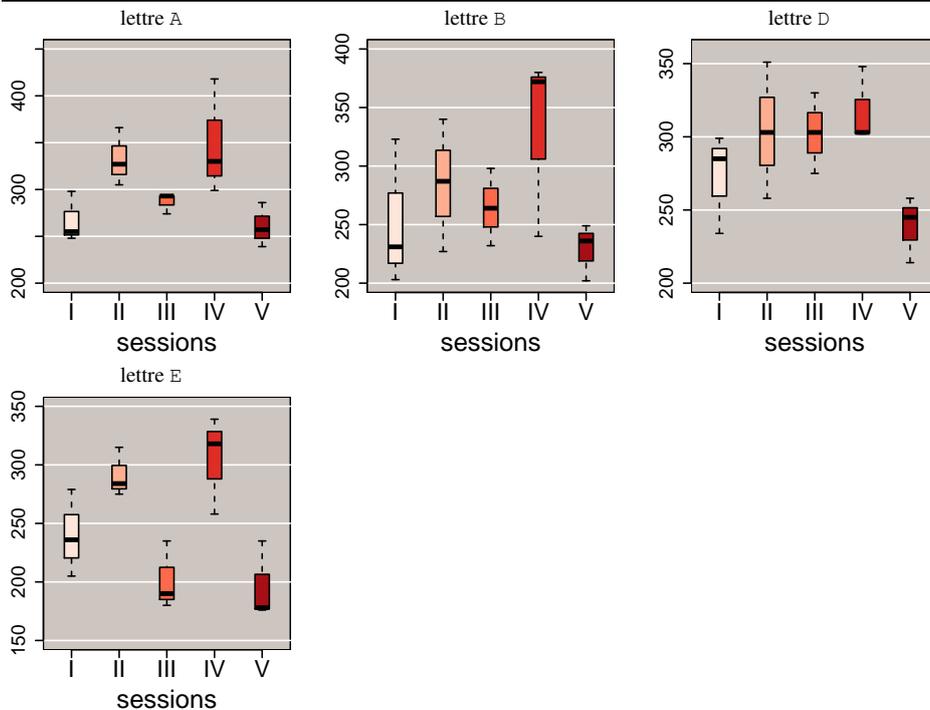


FIGURE 5.87: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

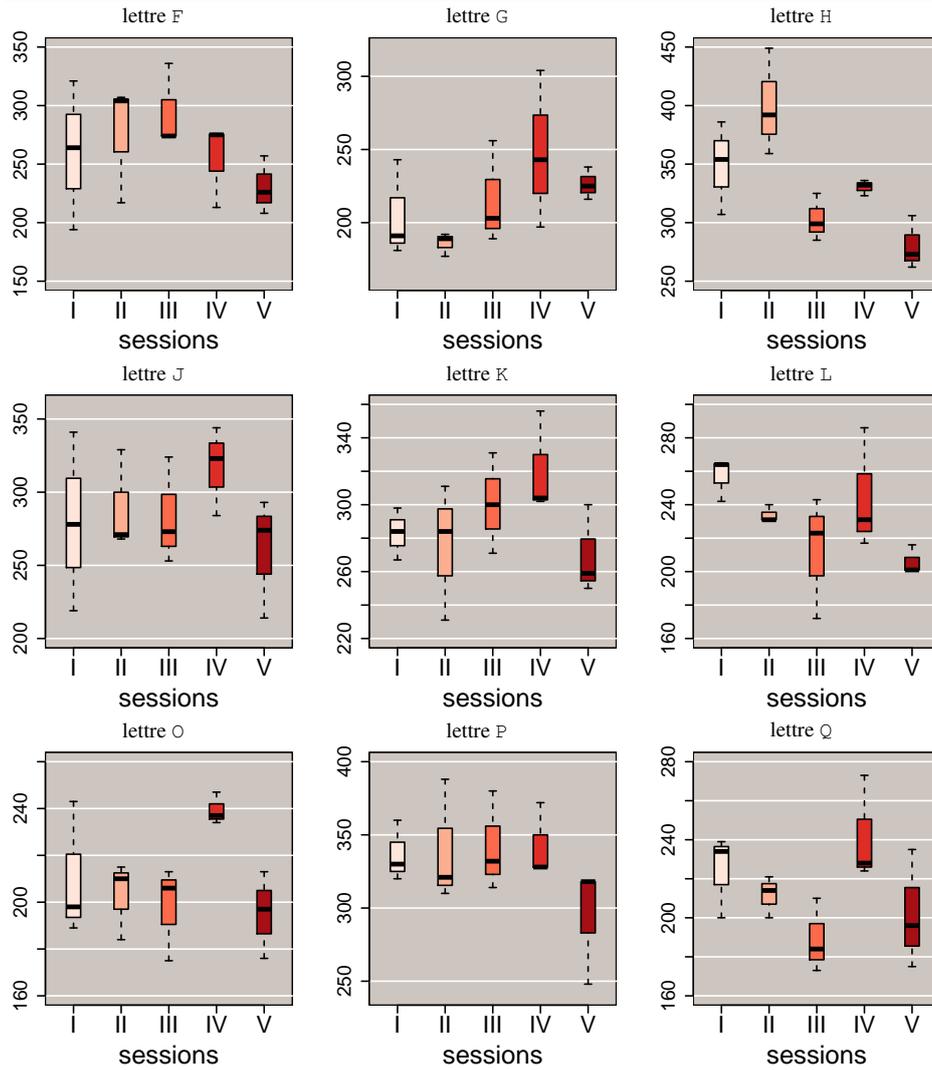


FIGURE 5.88: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

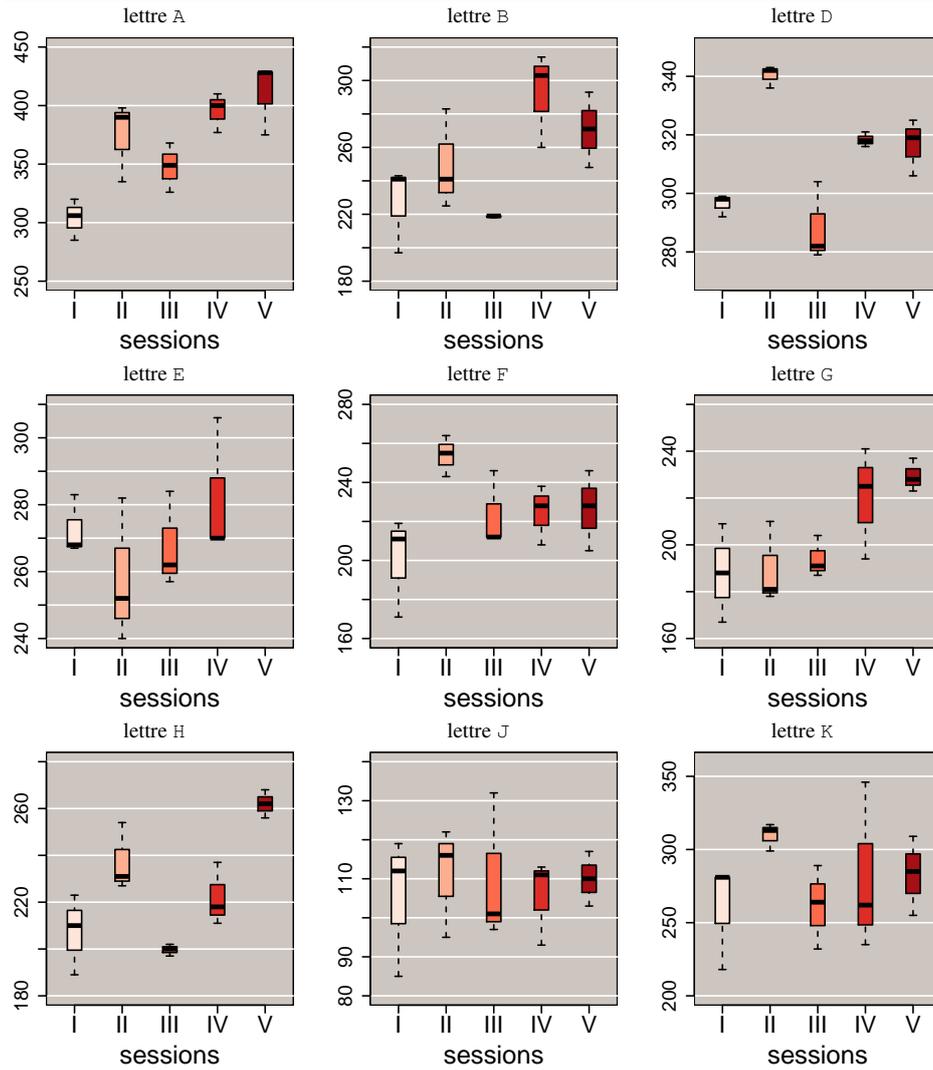


FIGURE 5.89: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 9, largeur W, caractères grands

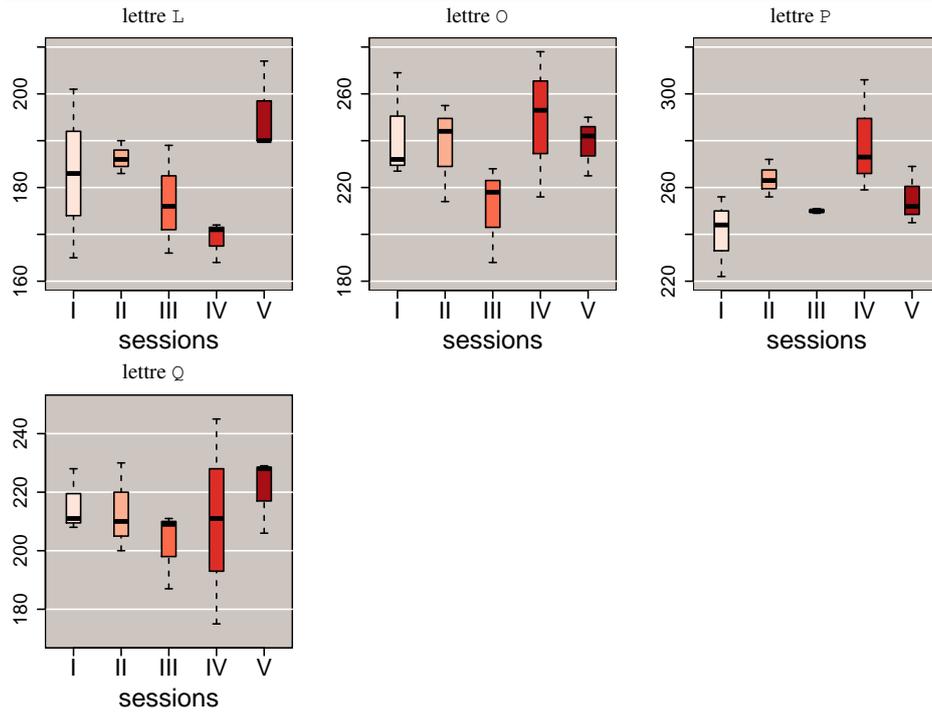


FIGURE 5.90: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages I – V.

Scripteur 10, largeur W, caractères grands

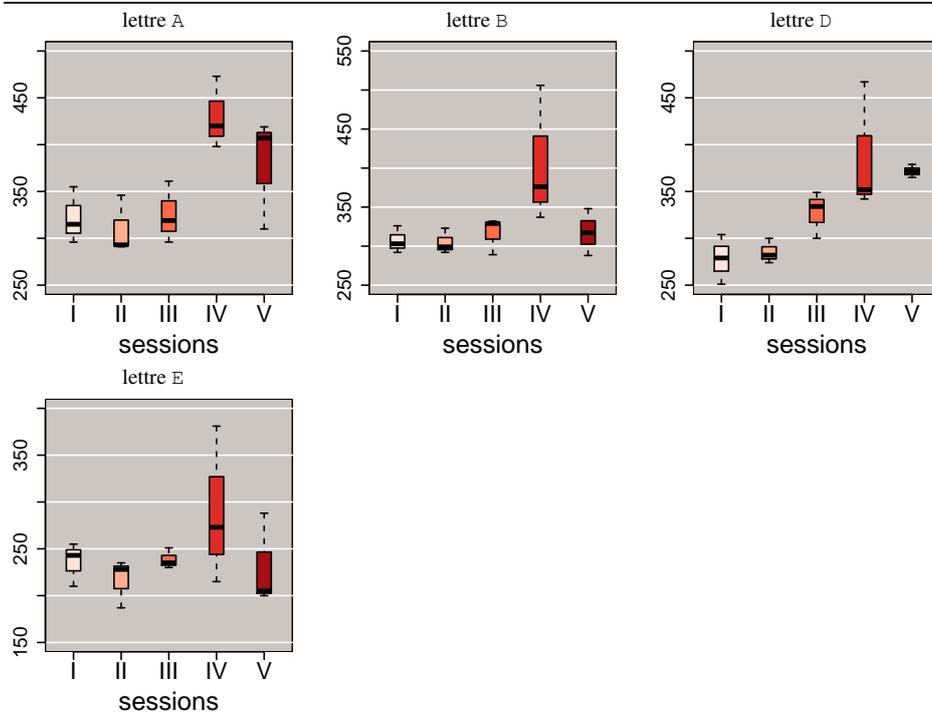


FIGURE 5.91: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages I – V.

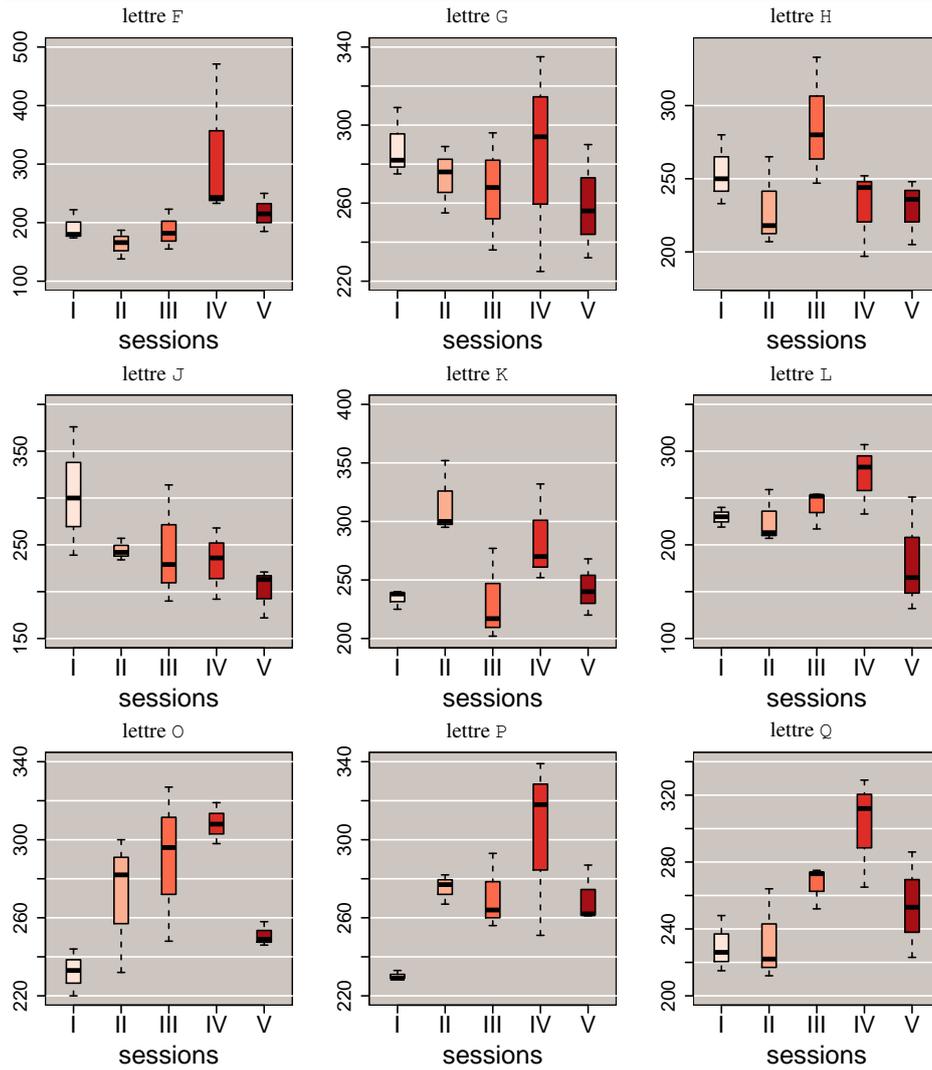


FIGURE 5.92: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

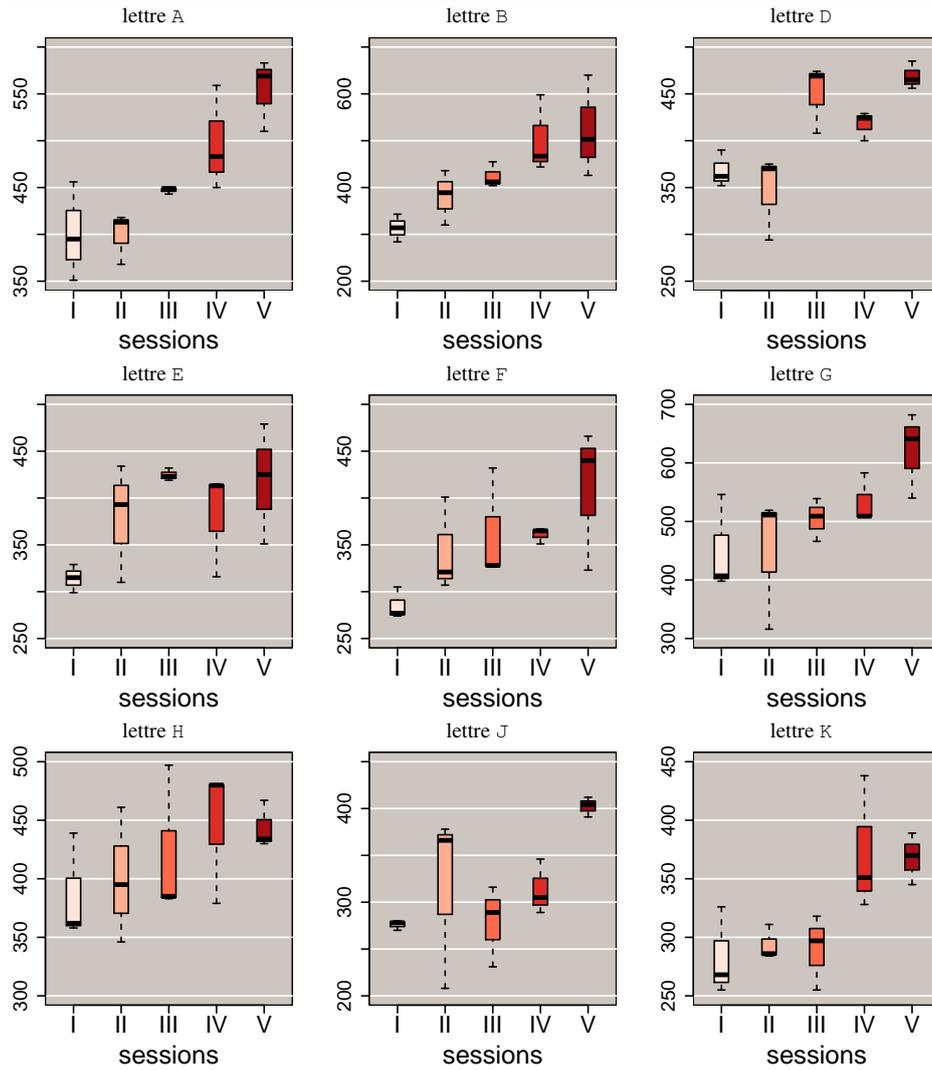


FIGURE 5.93: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

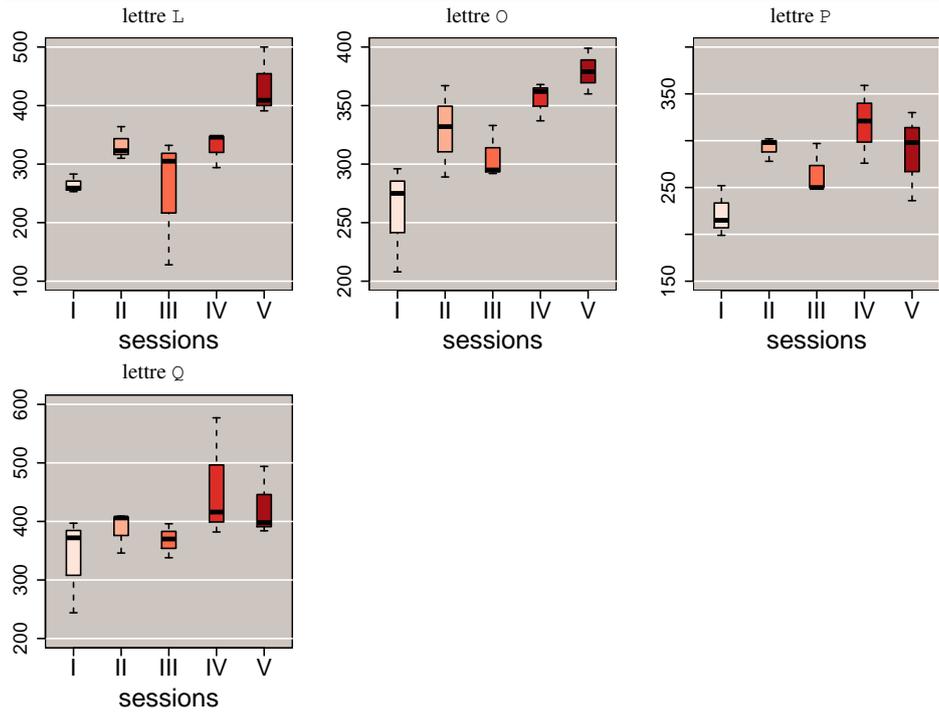


FIGURE 5.94: Boîtes à moustache des largeurs  $W$  pour les caractères *grands* selon les pages  $I - V$ .

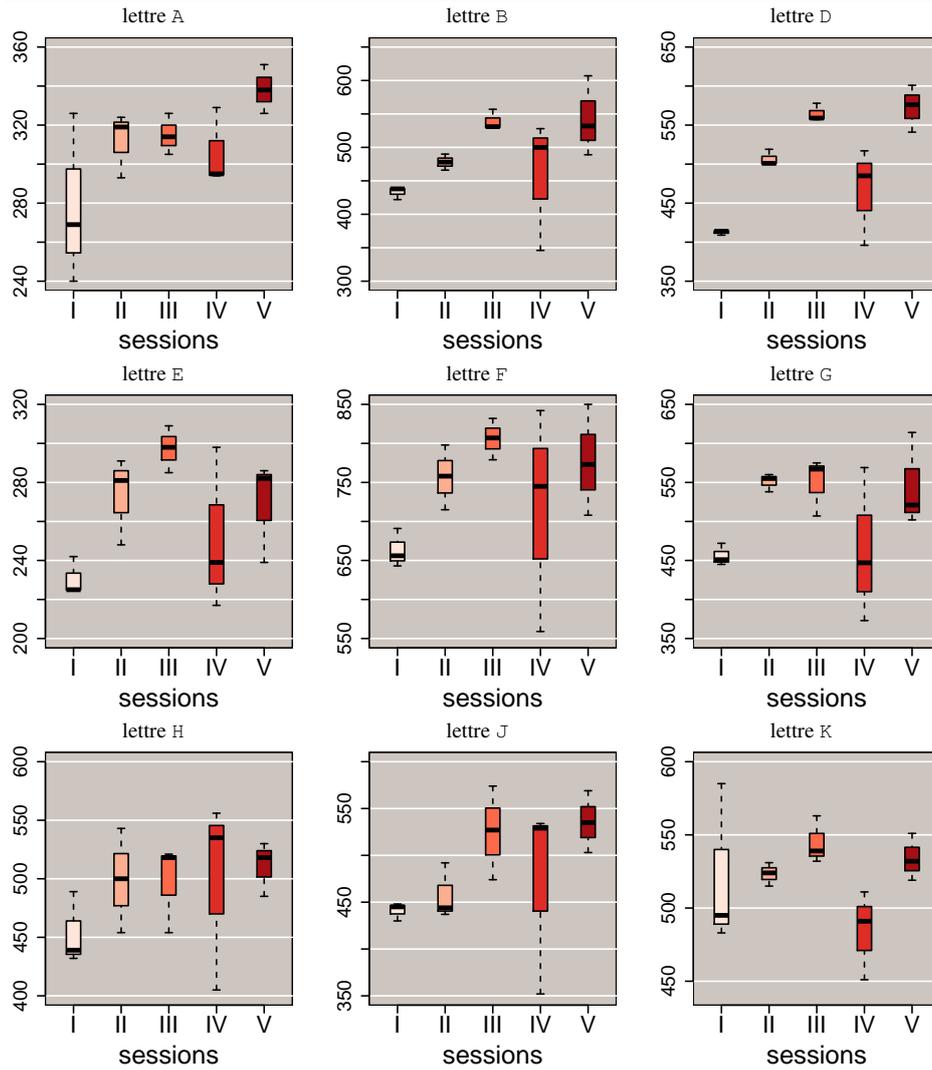


FIGURE 5.95: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

**Scripteur 1, hauteur H, caractères grands**

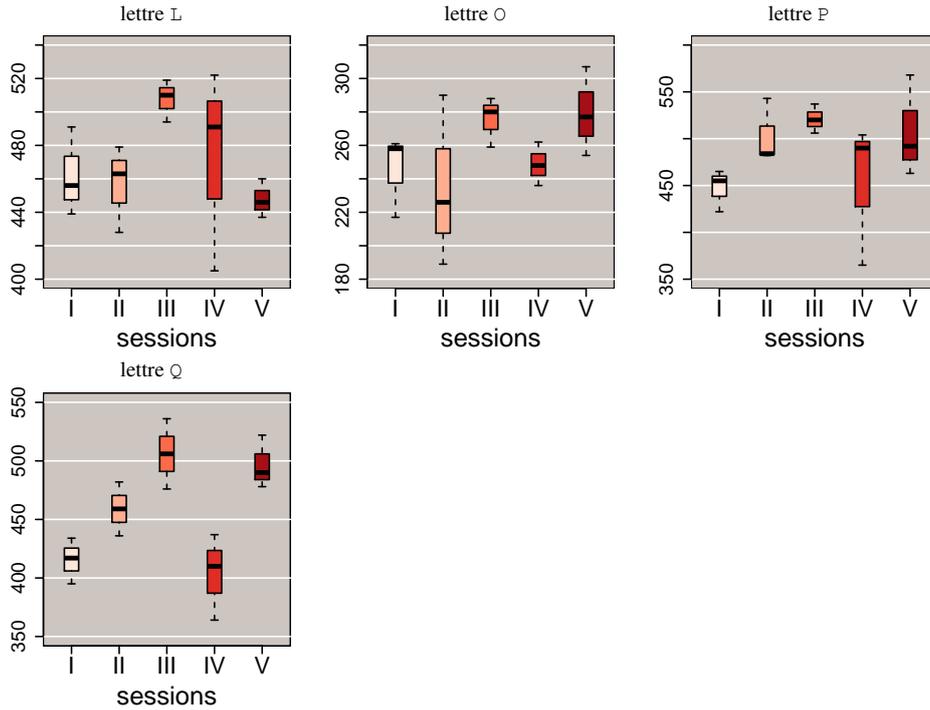


FIGURE 5.96: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

**Scripteur 2, hauteur H, caractères grands**

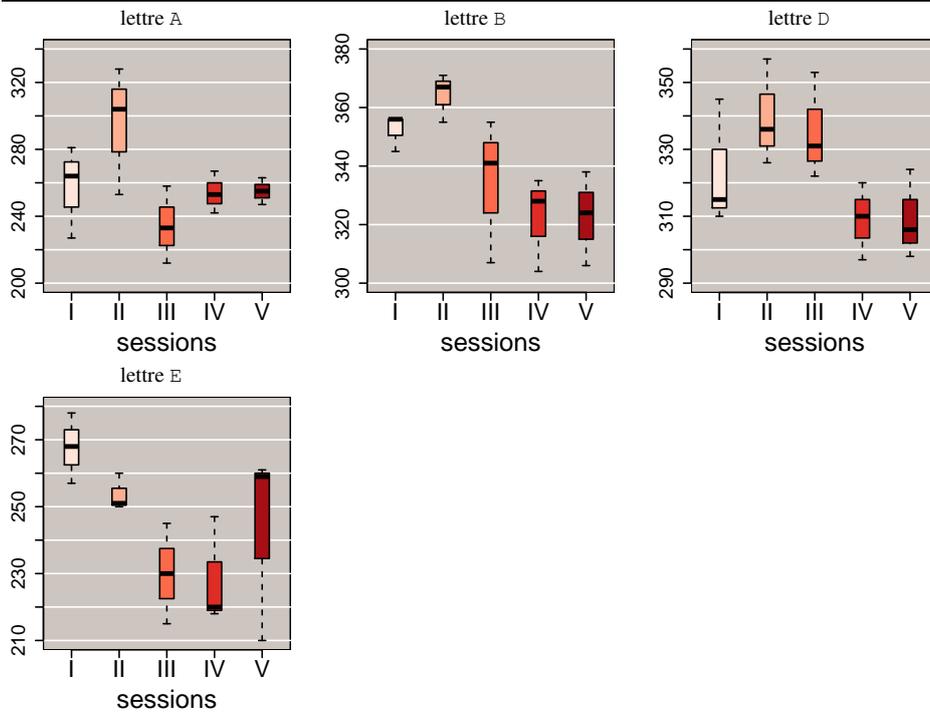


FIGURE 5.97: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

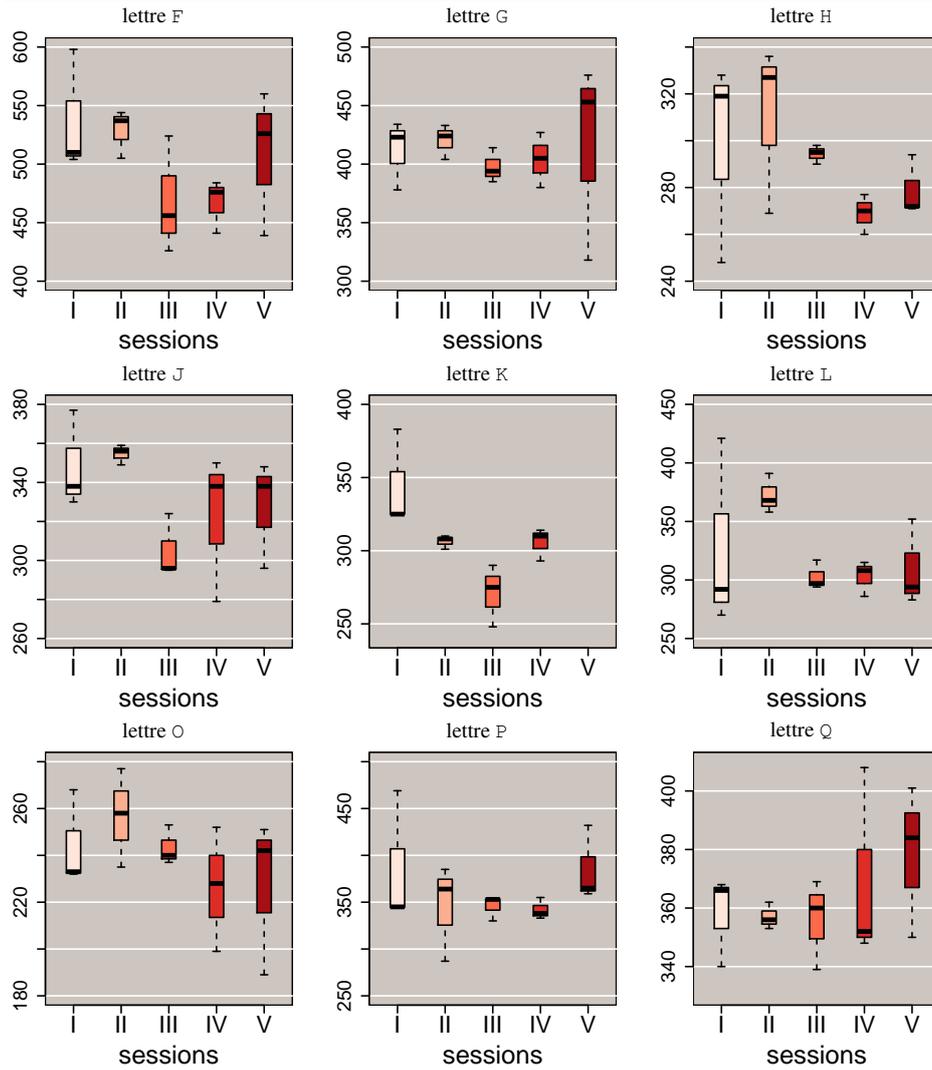


FIGURE 5.98: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

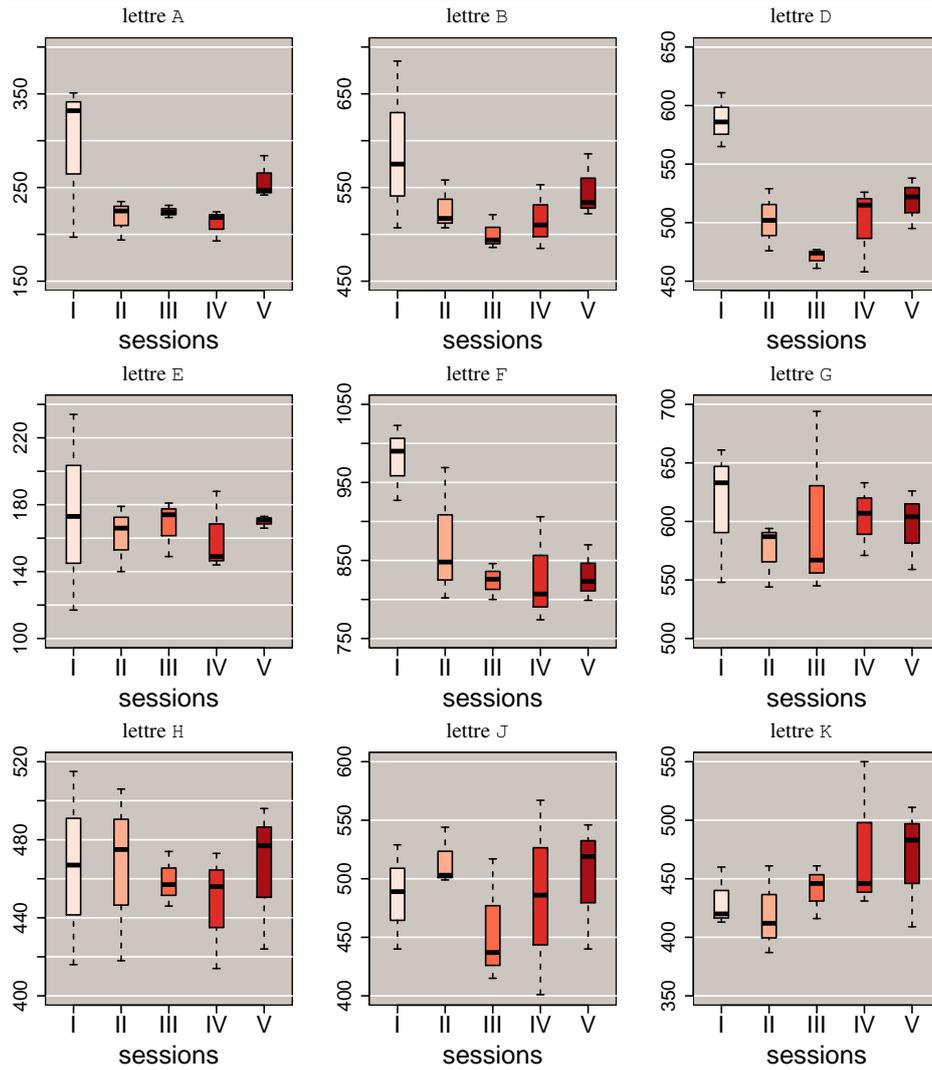


FIGURE 5.99: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

**Scripteur 3, hauteur H, caractères grands**

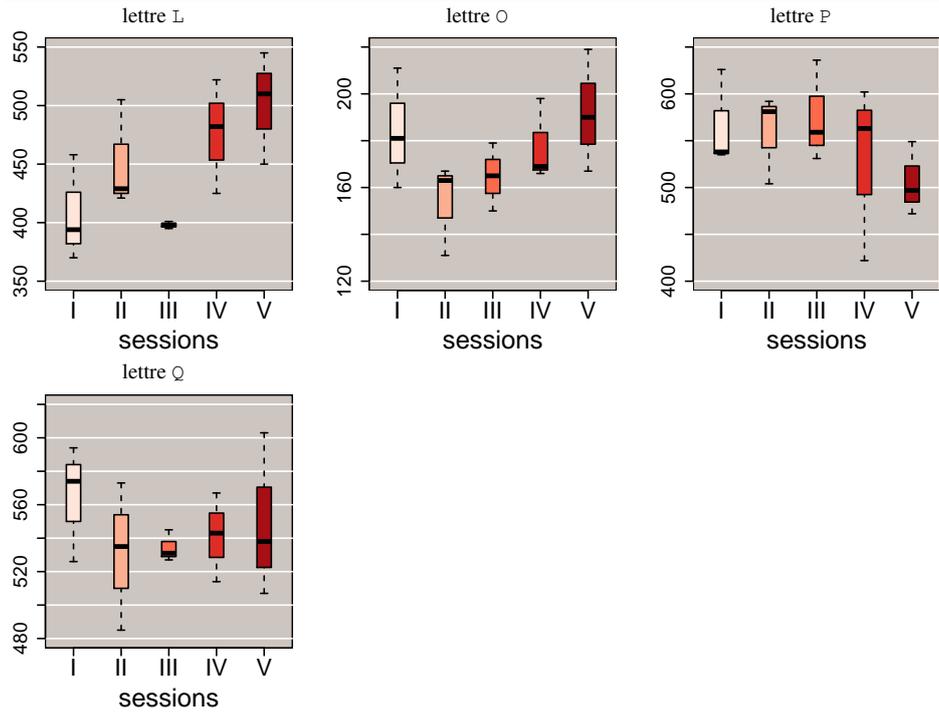


FIGURE 5.100: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

**Scripteur 4, hauteur H, caractères grands**

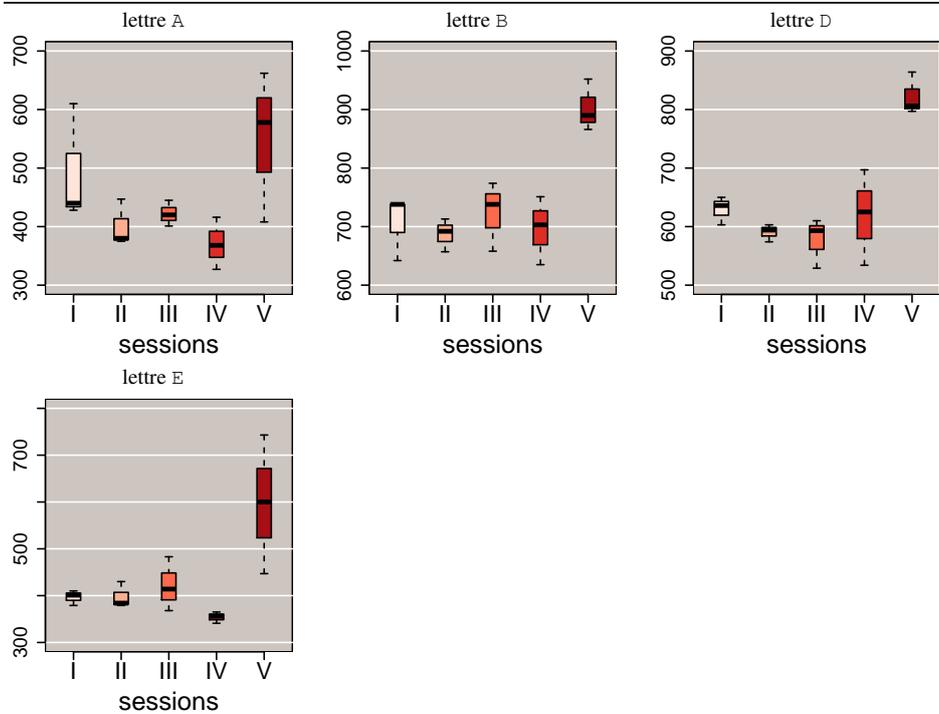


FIGURE 5.101: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

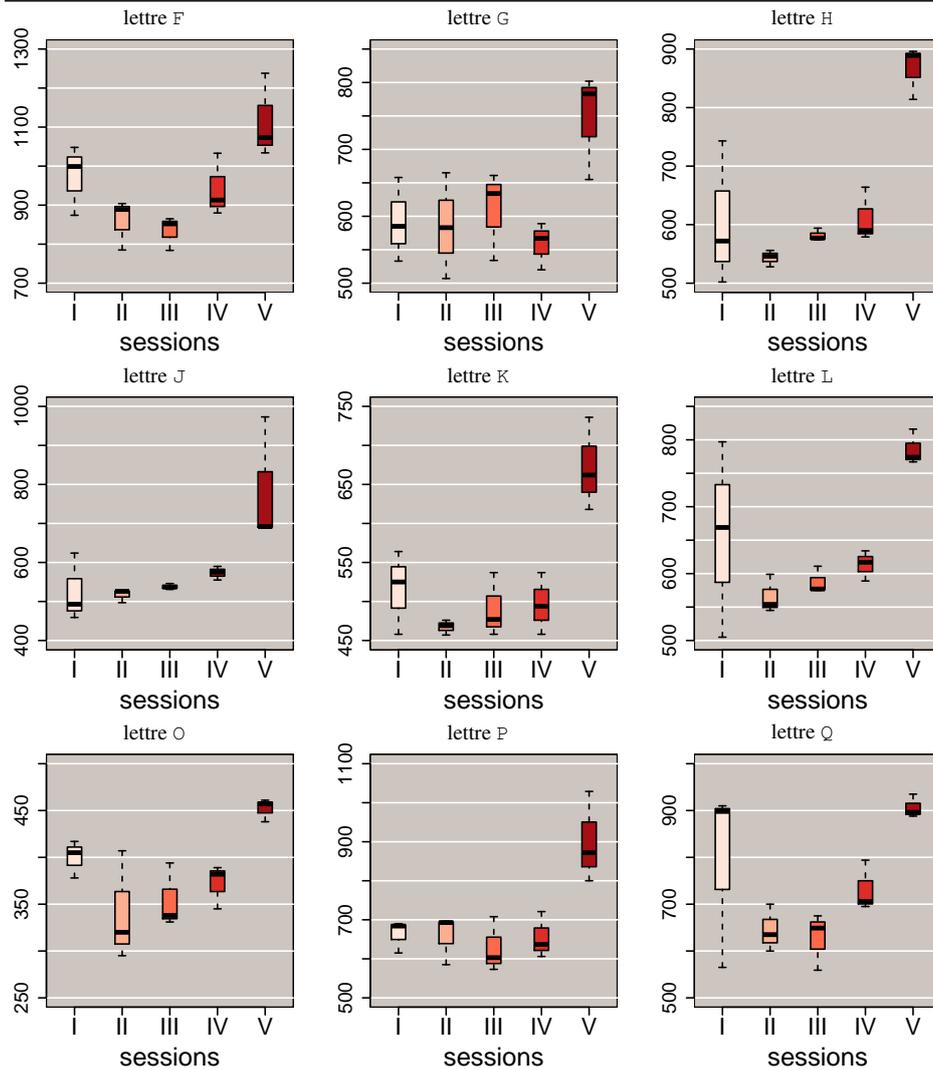


FIGURE 5.102: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

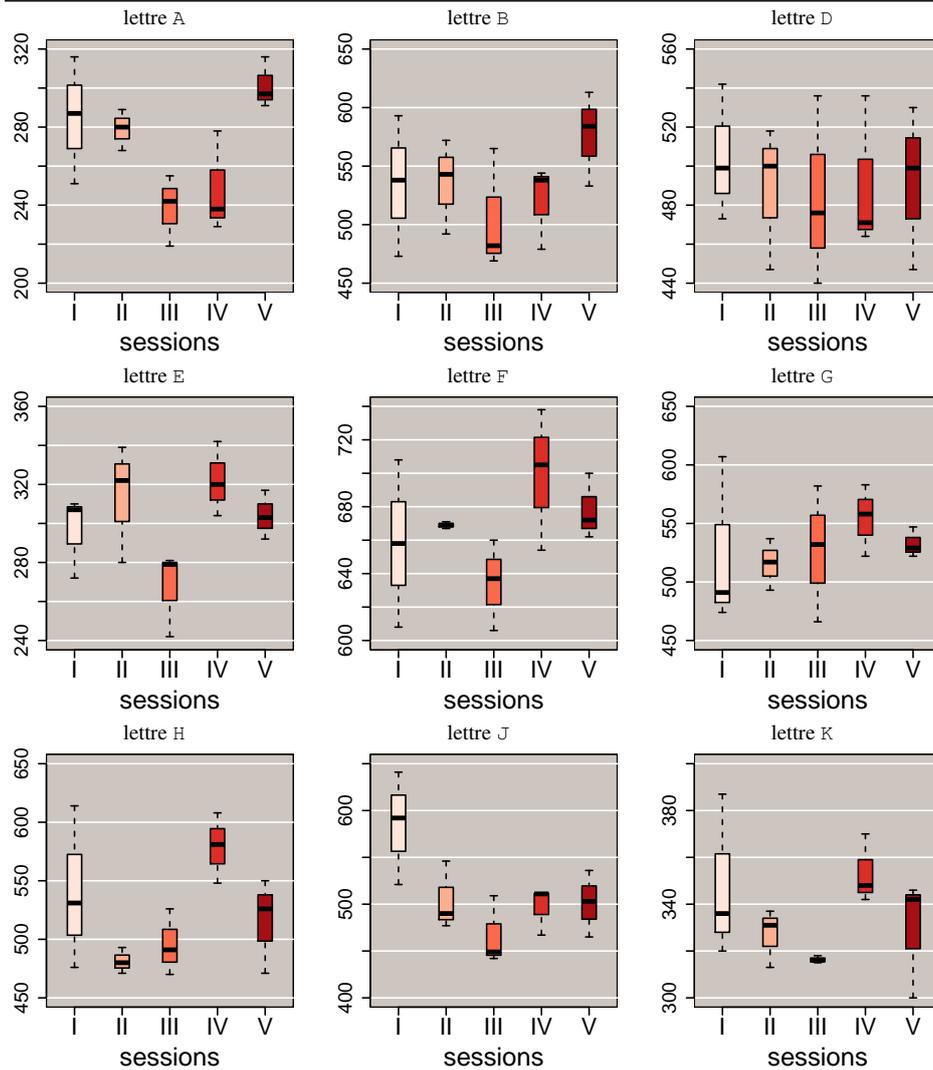


FIGURE 5.103: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 5, hauteur H, caractères *grands*

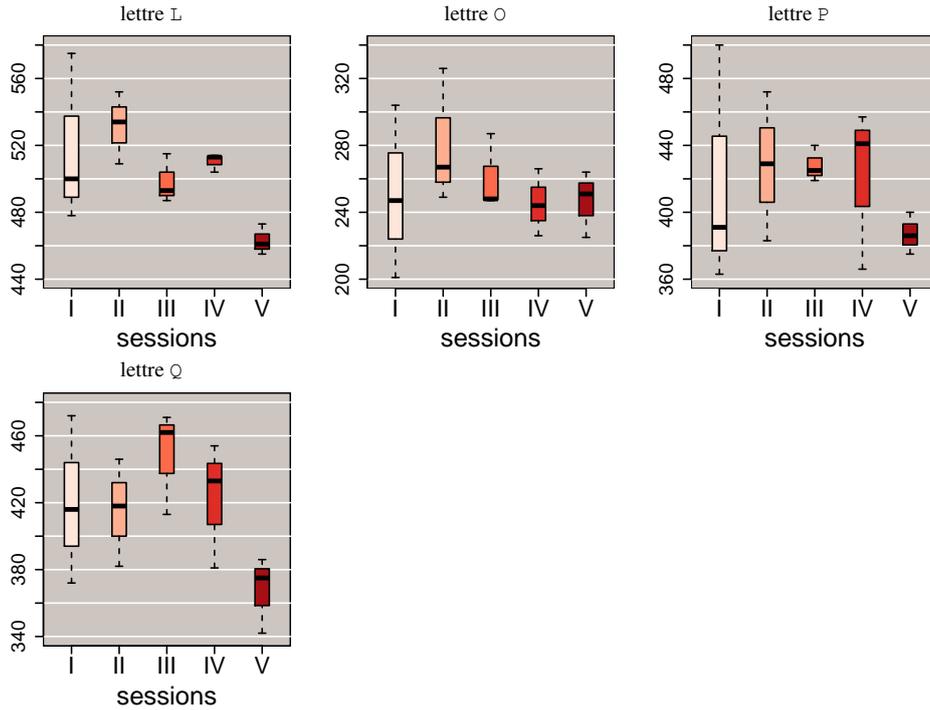


FIGURE 5.104: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 6, hauteur H, caractères *grands*

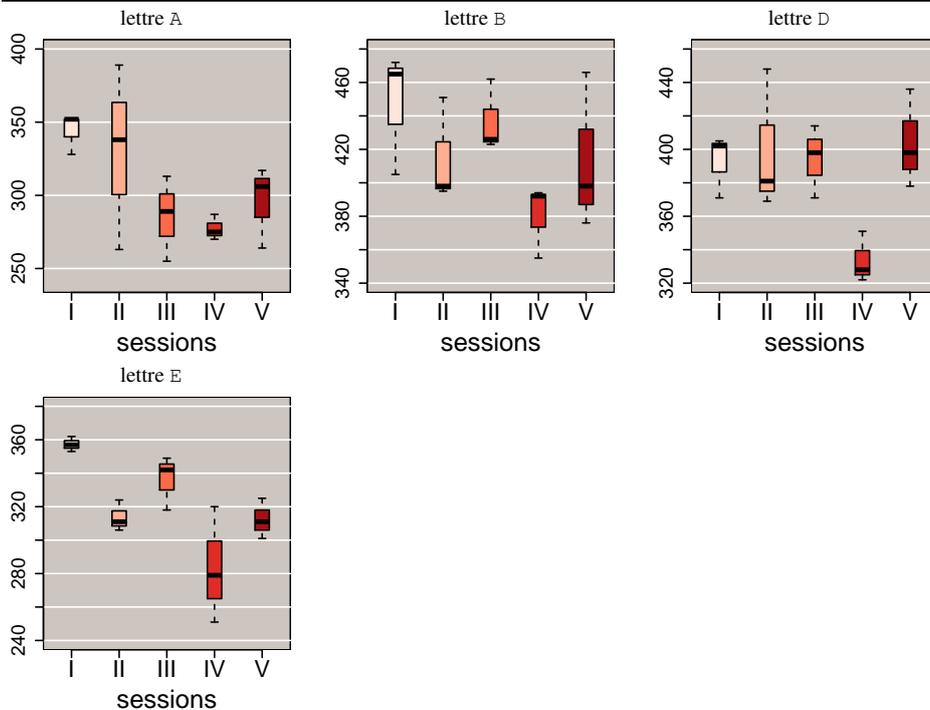


FIGURE 5.105: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

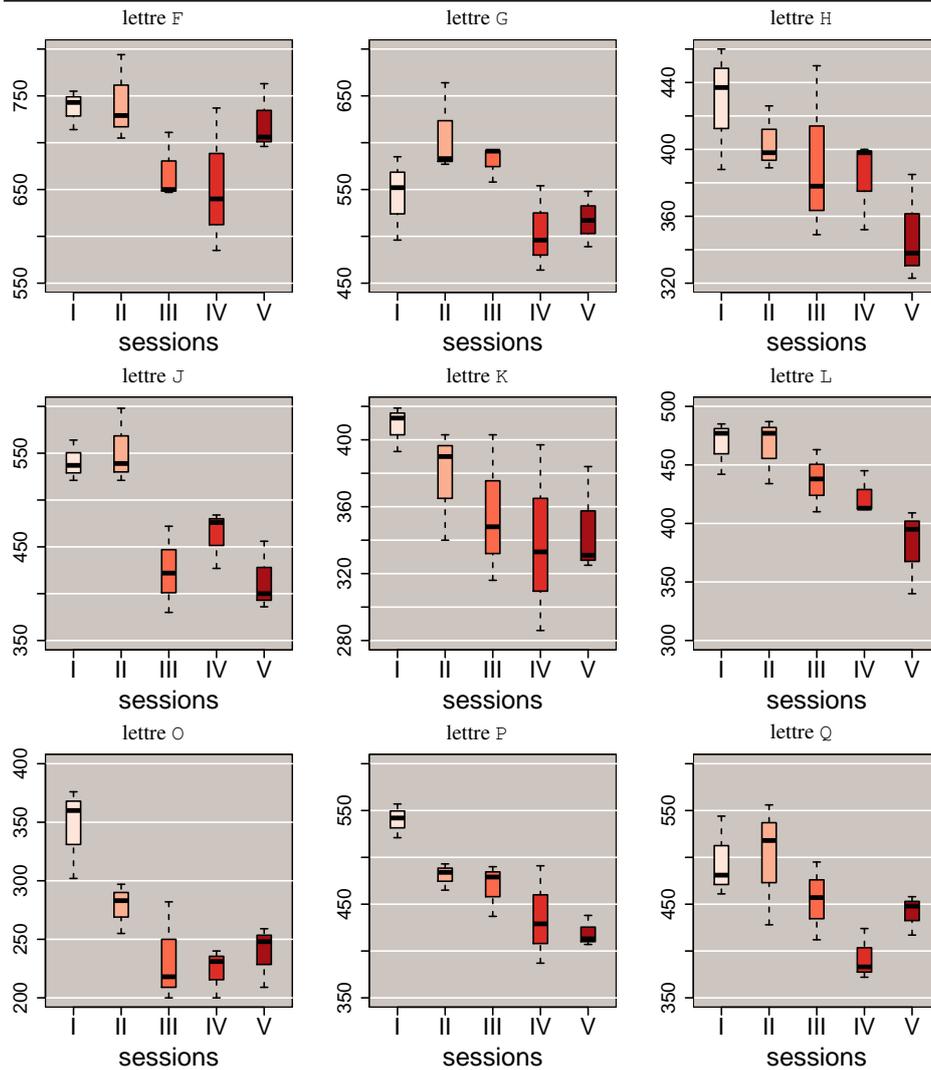


FIGURE 5.106: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

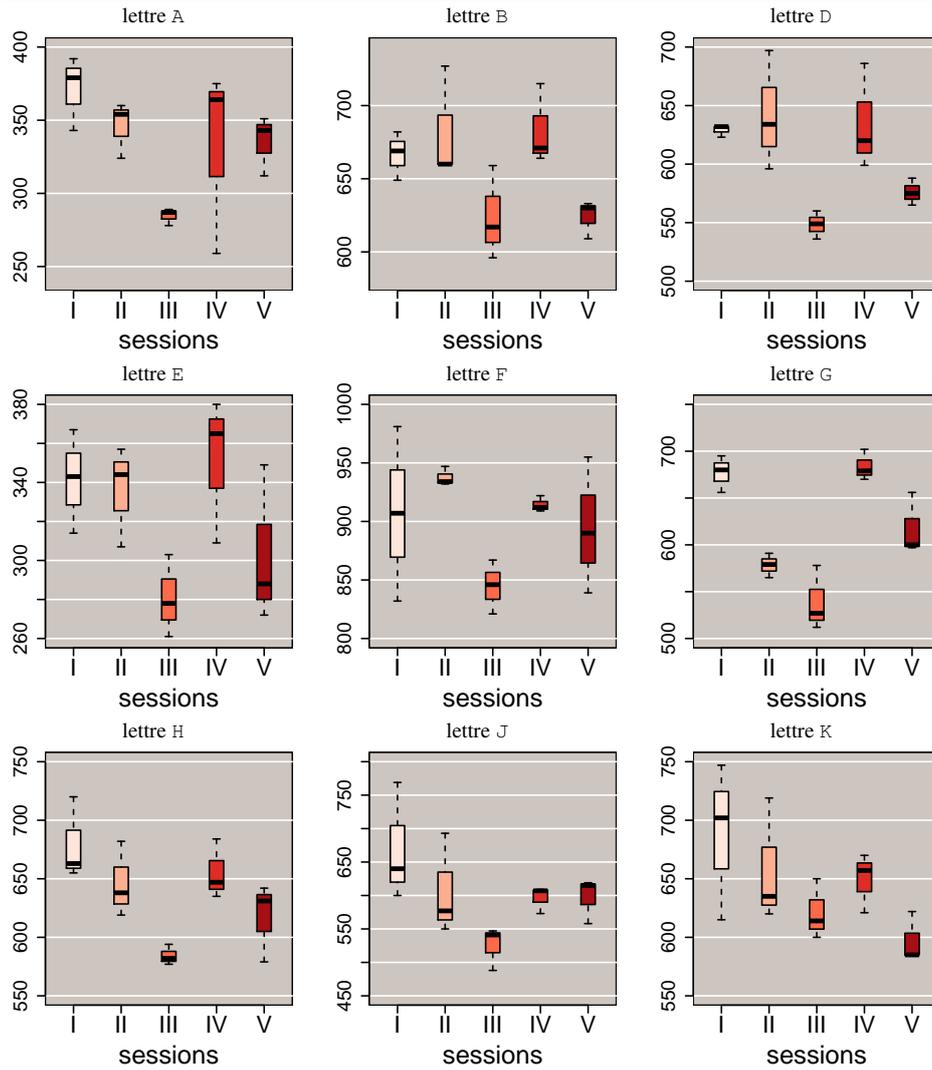


FIGURE 5.107: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 7, hauteur H, caractères *grands*

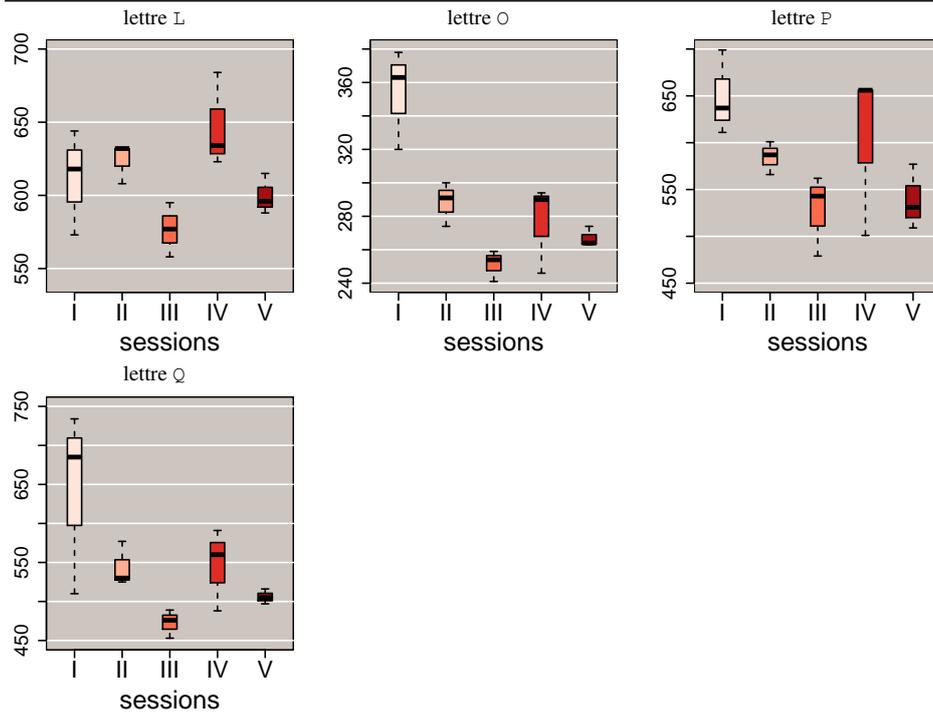


FIGURE 5.108: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 8, hauteur H, caractères *grands*

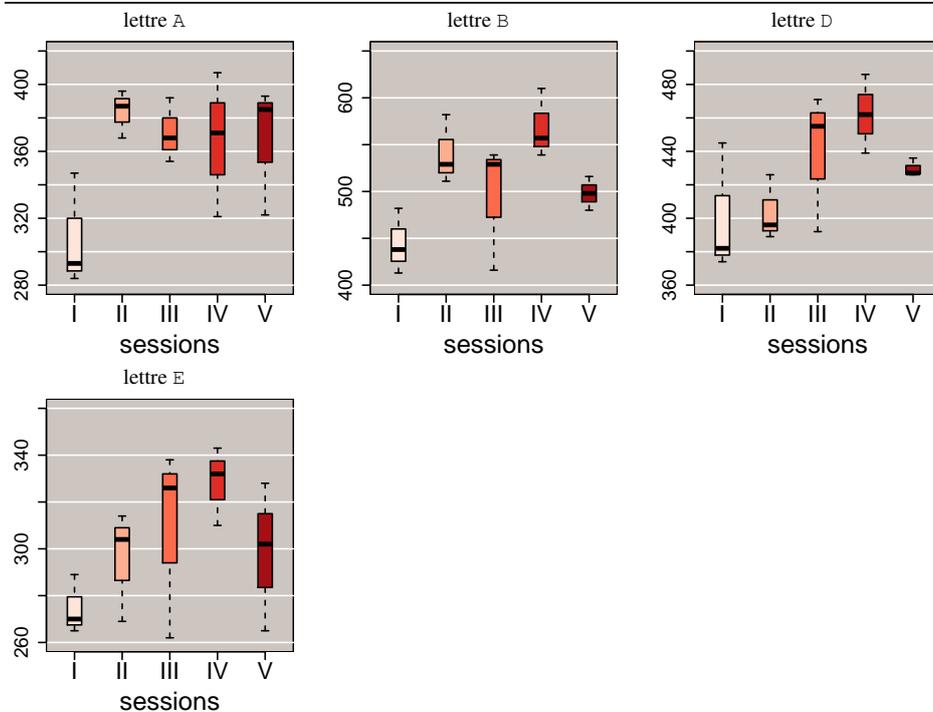


FIGURE 5.109: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

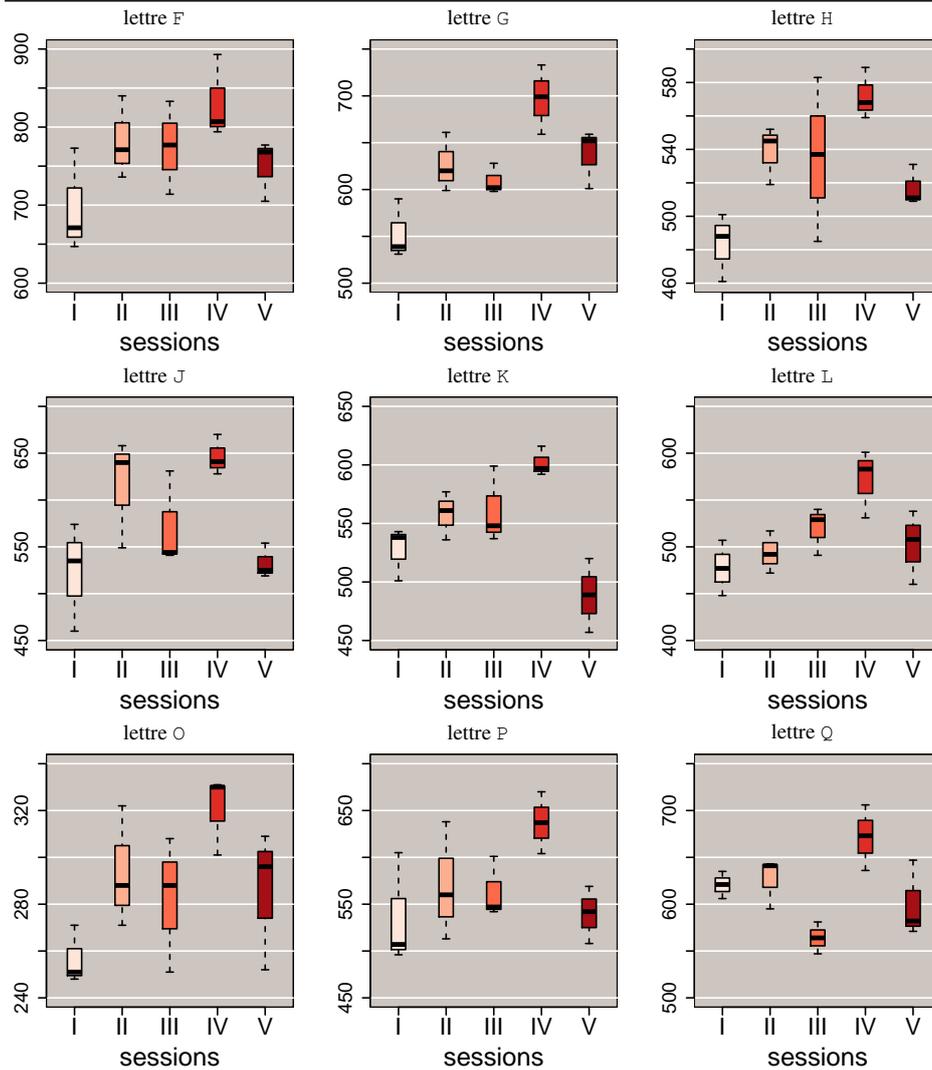


FIGURE 5.110: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

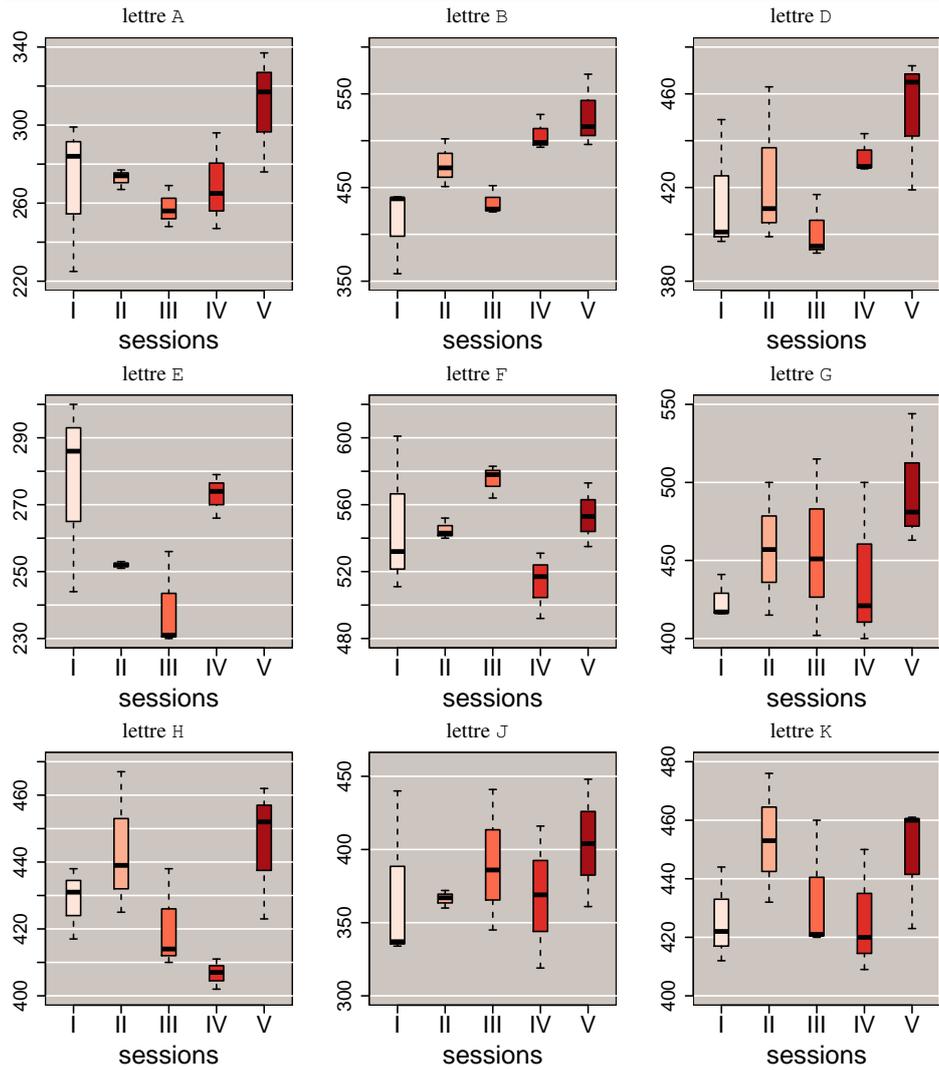


FIGURE 5.111: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 9, hauteur H, caractères *grands*

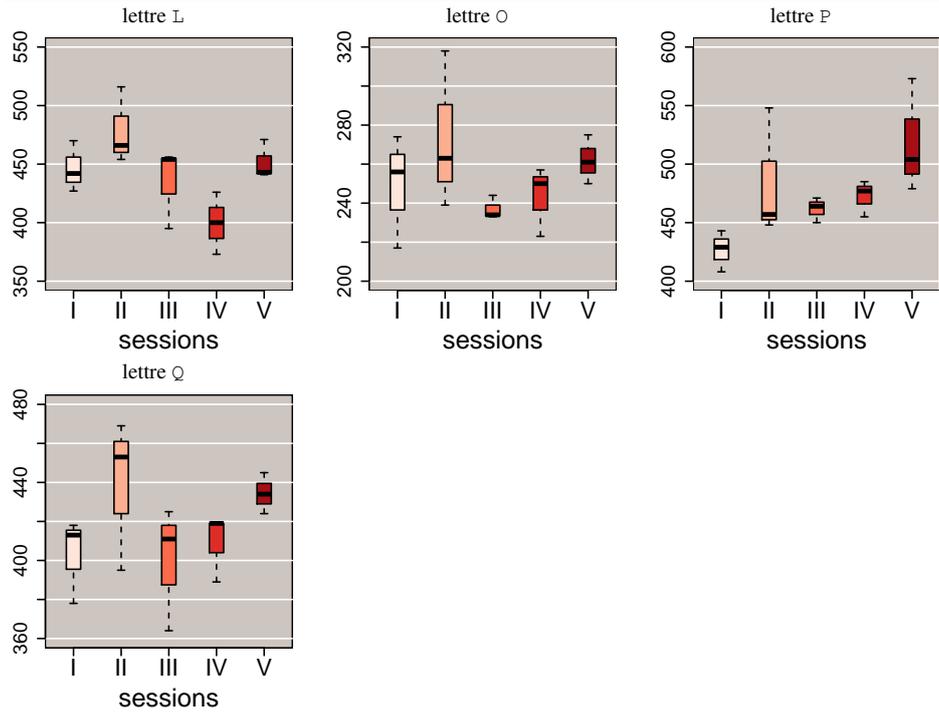


FIGURE 5.112: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

Scripteur 10, hauteur H, caractères *grands*

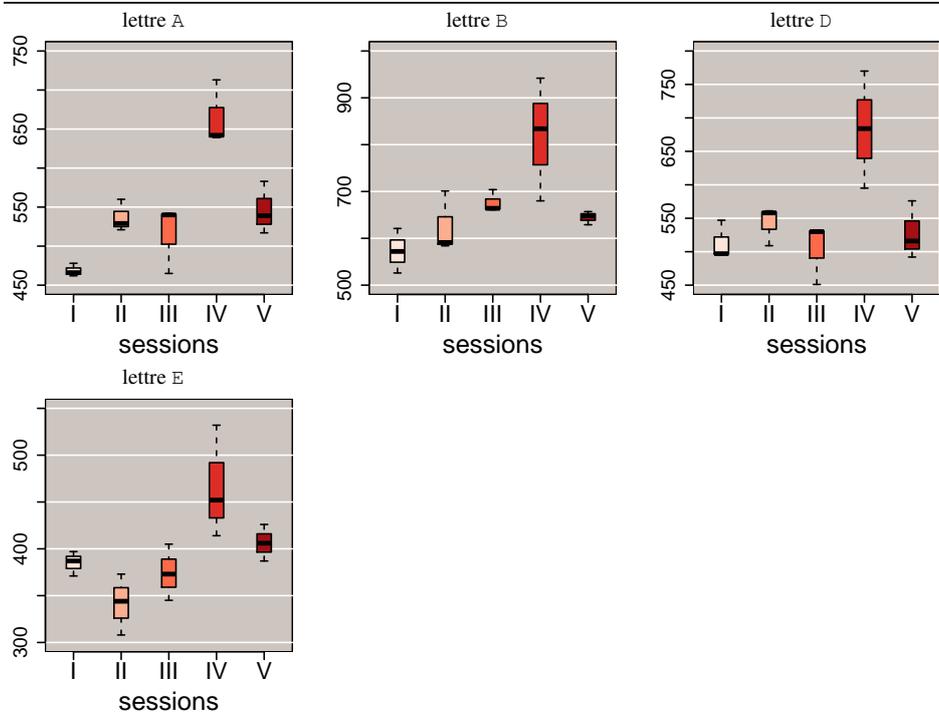


FIGURE 5.113: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

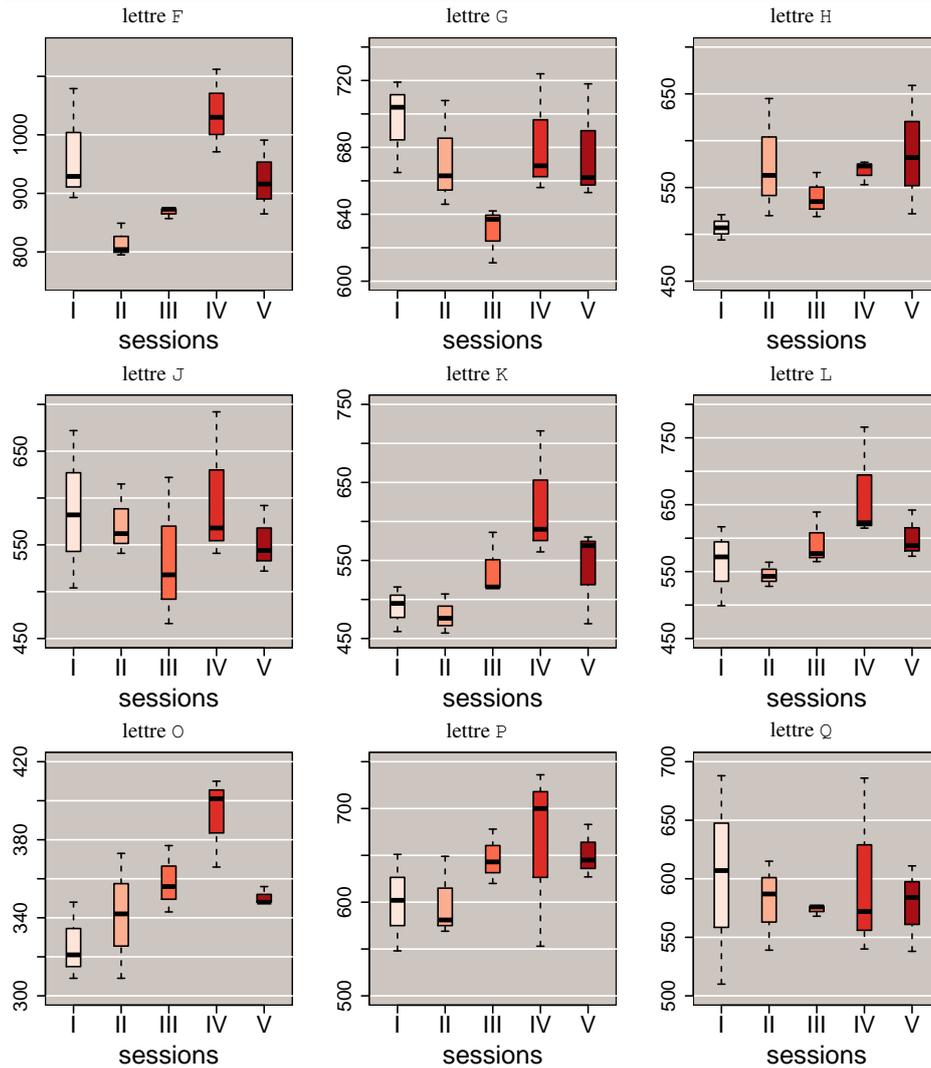


FIGURE 5.114: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

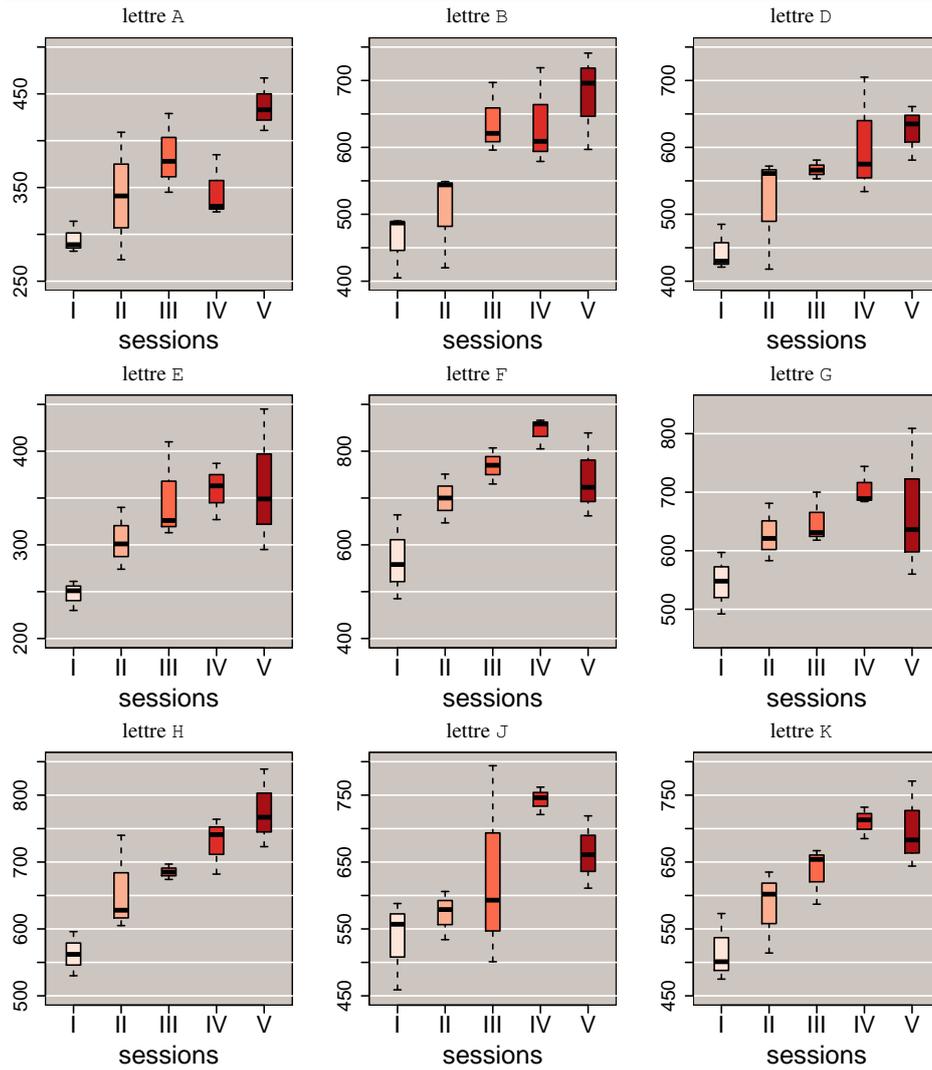


FIGURE 5.115: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

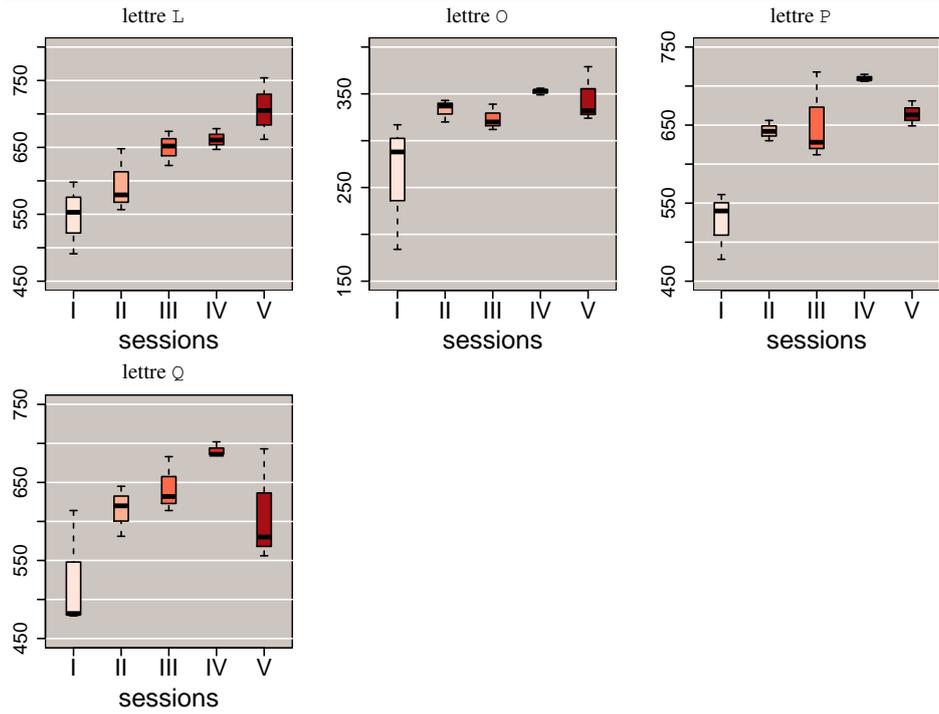


FIGURE 5.116: Boîtes à moustache des hauteurs  $H$  pour les caractères *grands* selon les pages  $I - V$ .

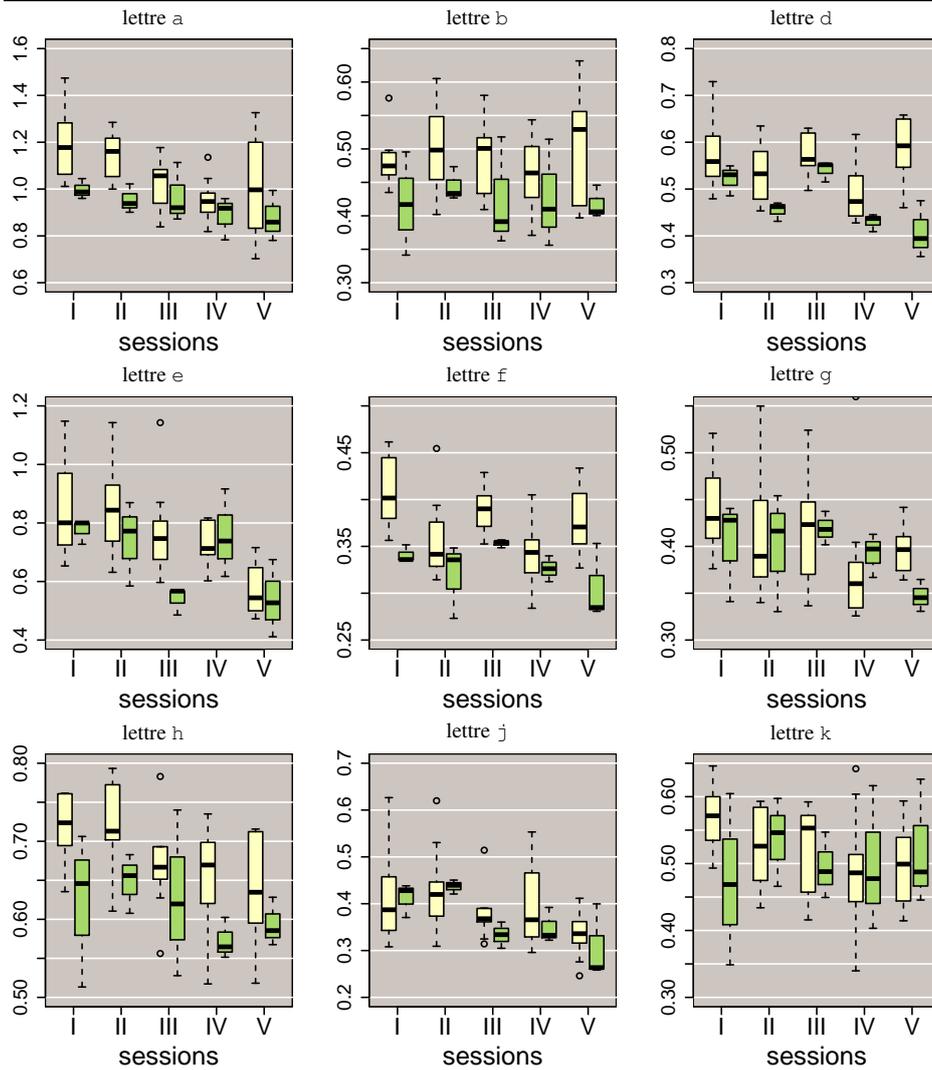


FIGURE 5.117: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

**Scripteur 1, rapport  $W/H$**

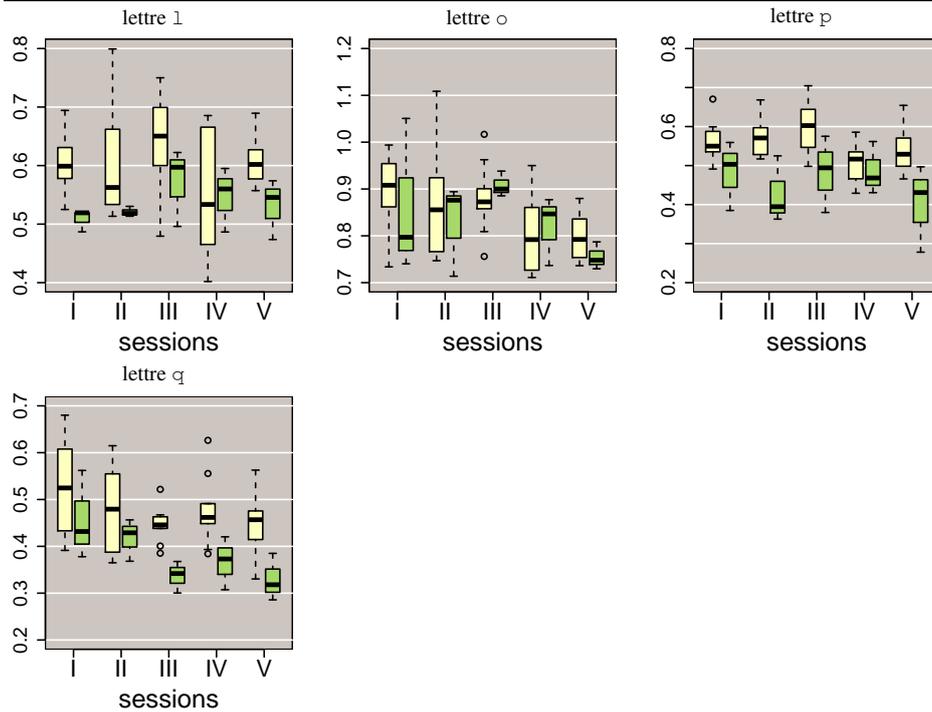


FIGURE 5.118: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

**Scripteur 2, rapport  $W/H$**

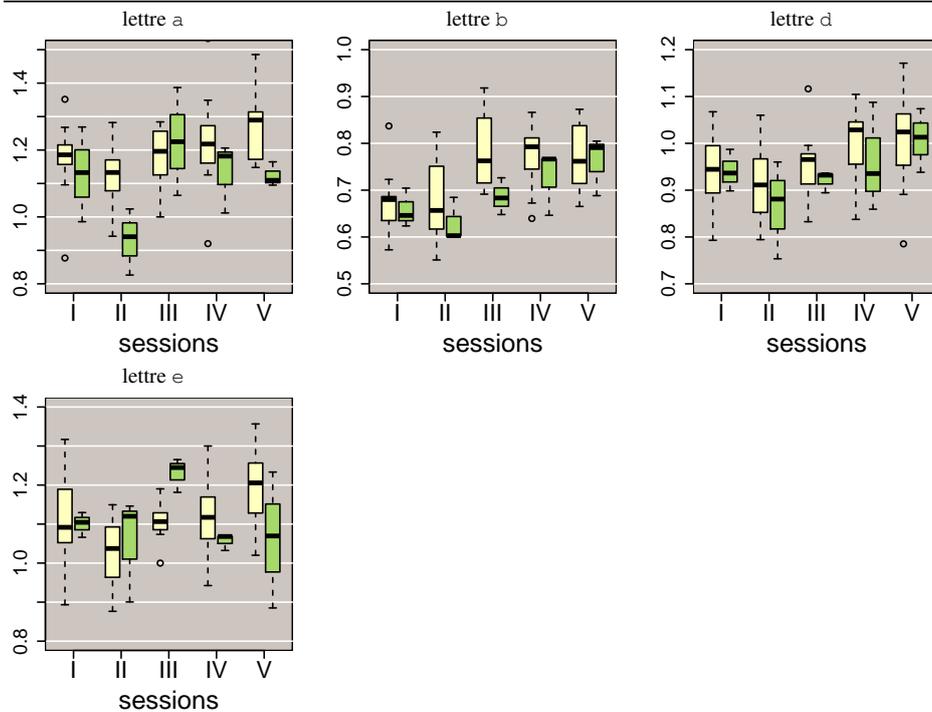


FIGURE 5.119: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

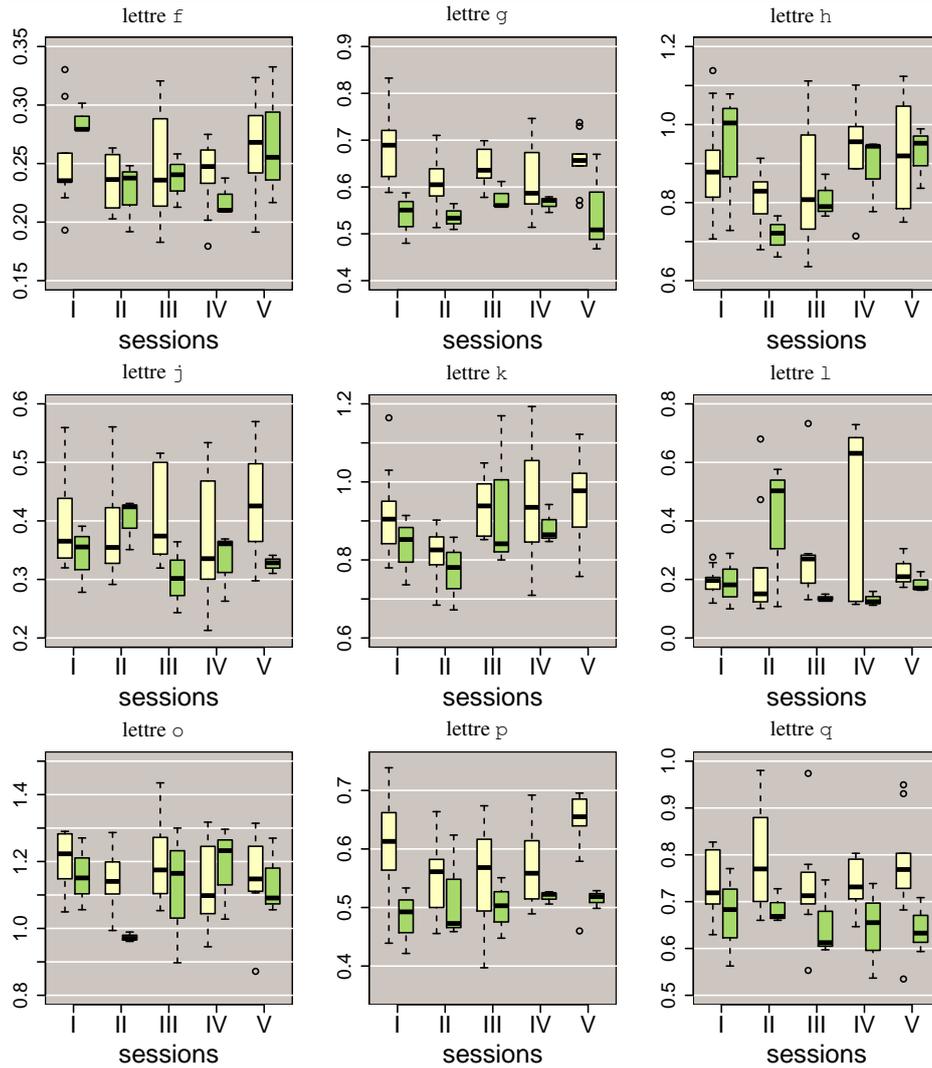


FIGURE 5.120: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

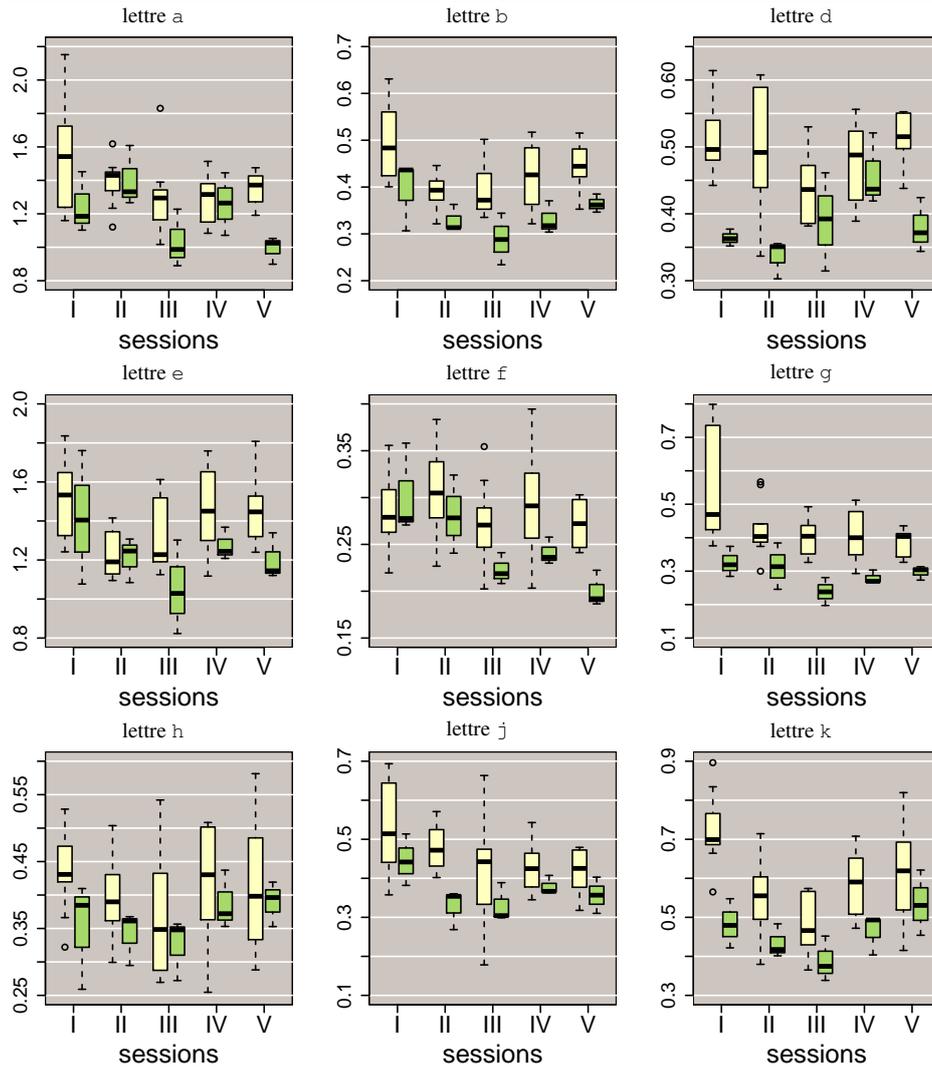


FIGURE 5.121: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

**Scripteur 3, rapport  $W/H$**

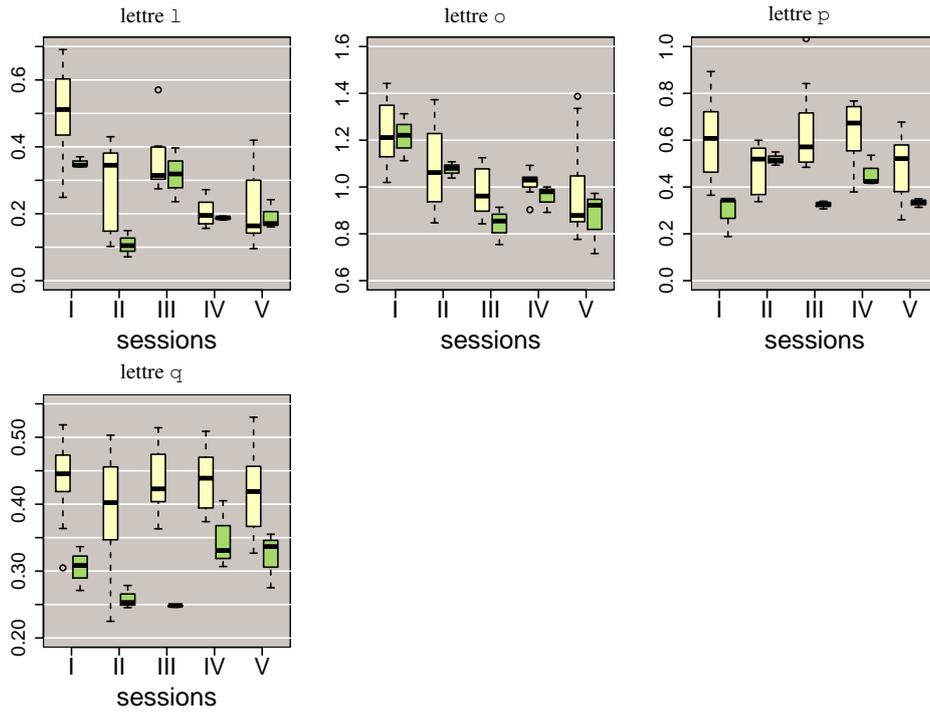


FIGURE 5.122: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

**Scripteur 4, rapport  $W/H$**

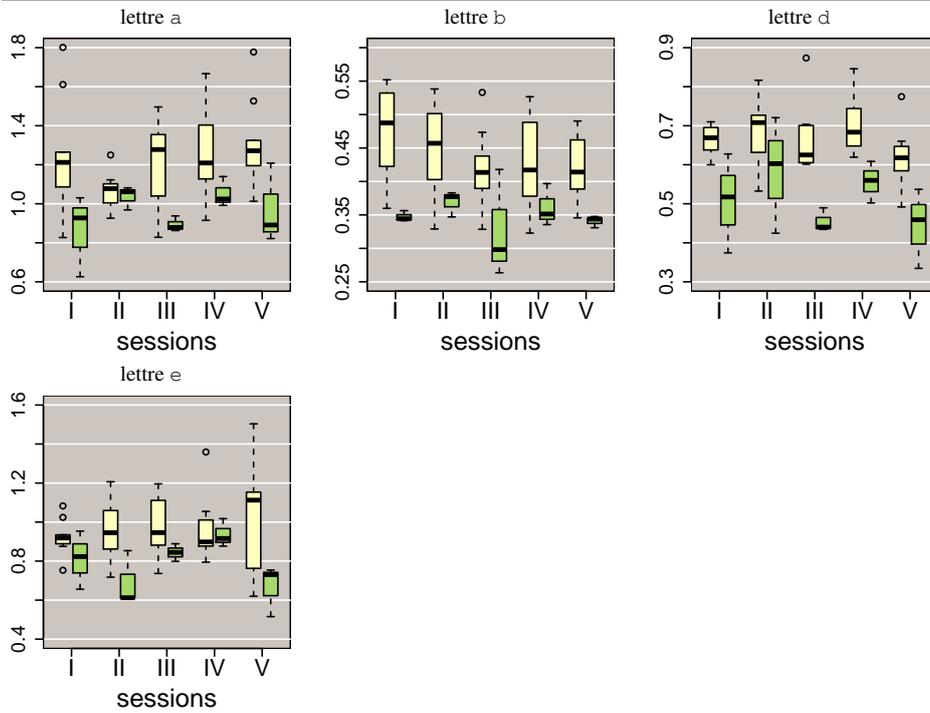


FIGURE 5.123: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

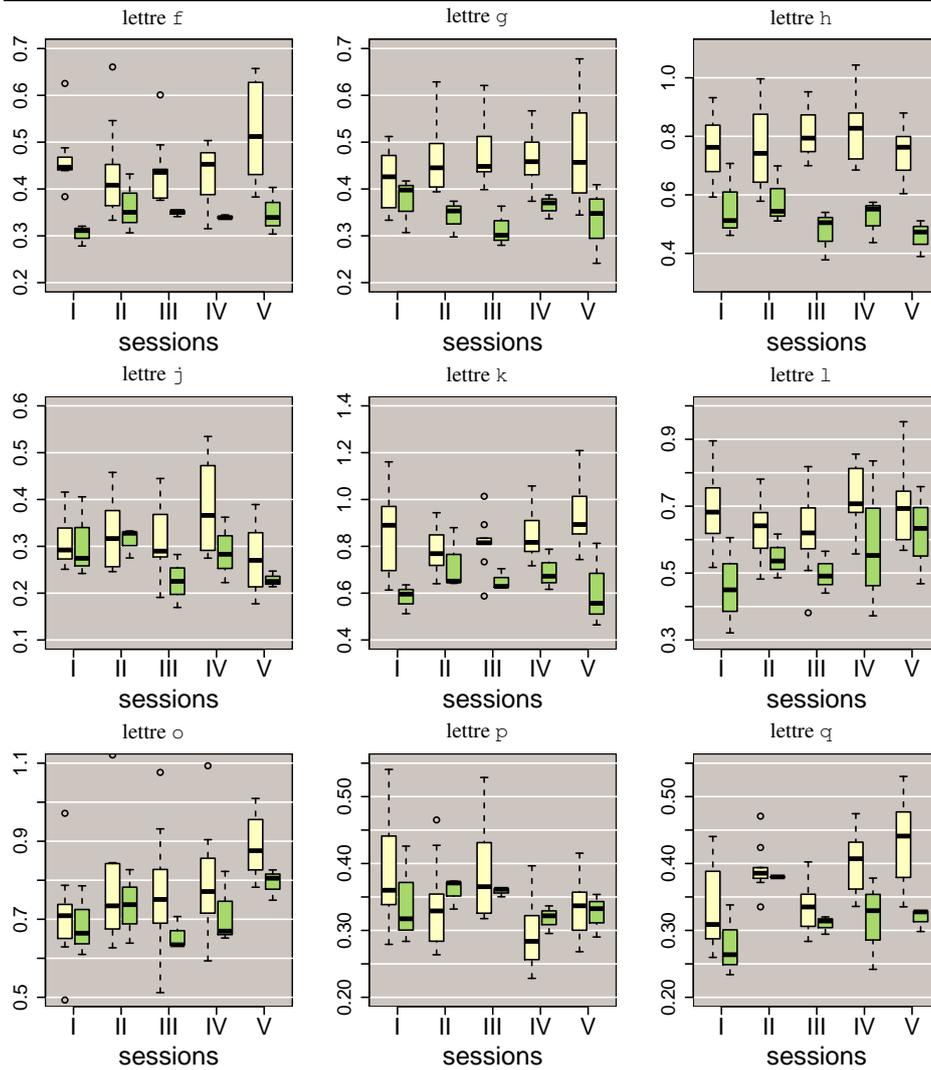


FIGURE 5.124: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

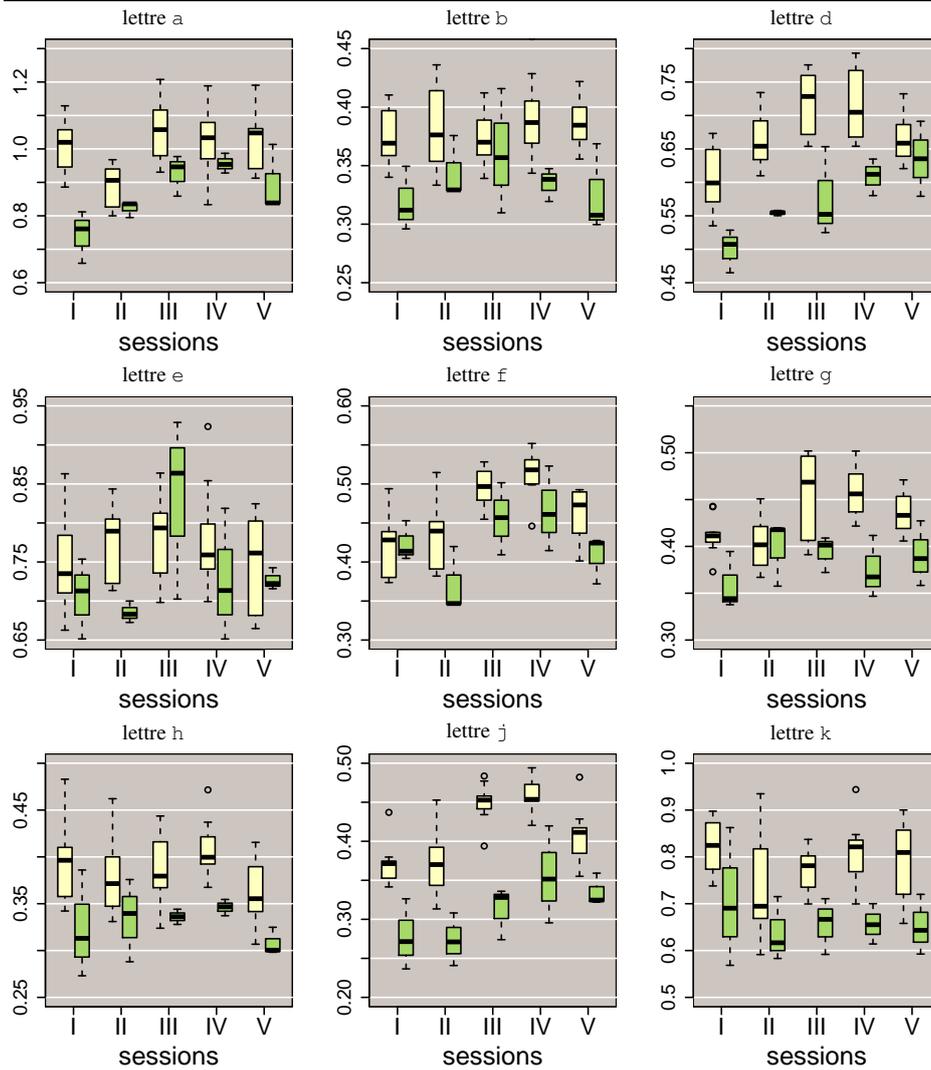


FIGURE 5.125: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

**Scripteur 5, rapport  $W/H$**

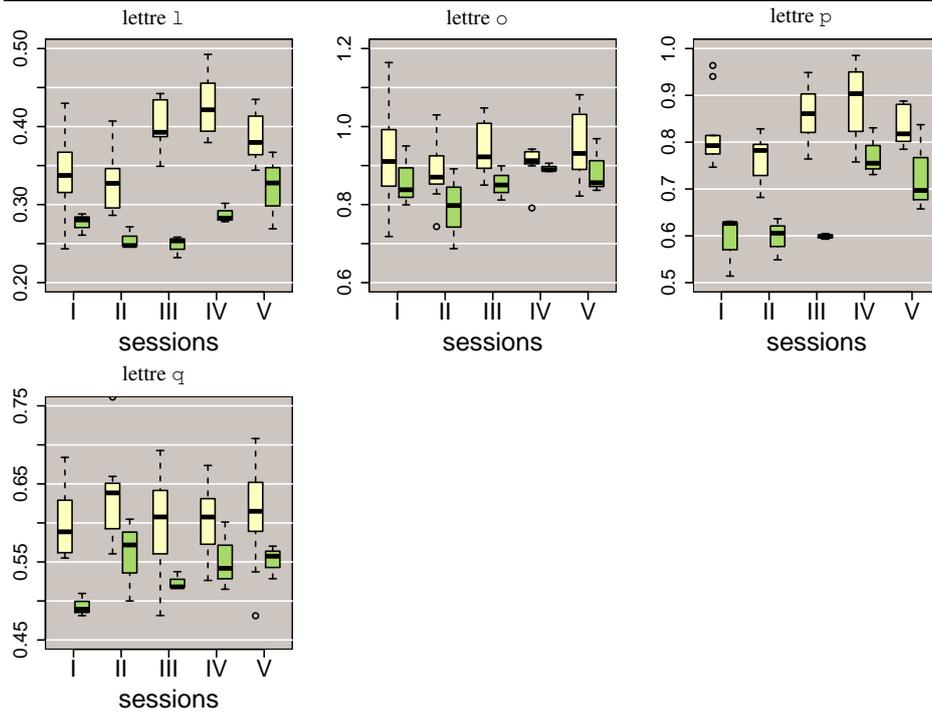


FIGURE 5.126: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

**Scripteur 6, rapport  $W/H$**

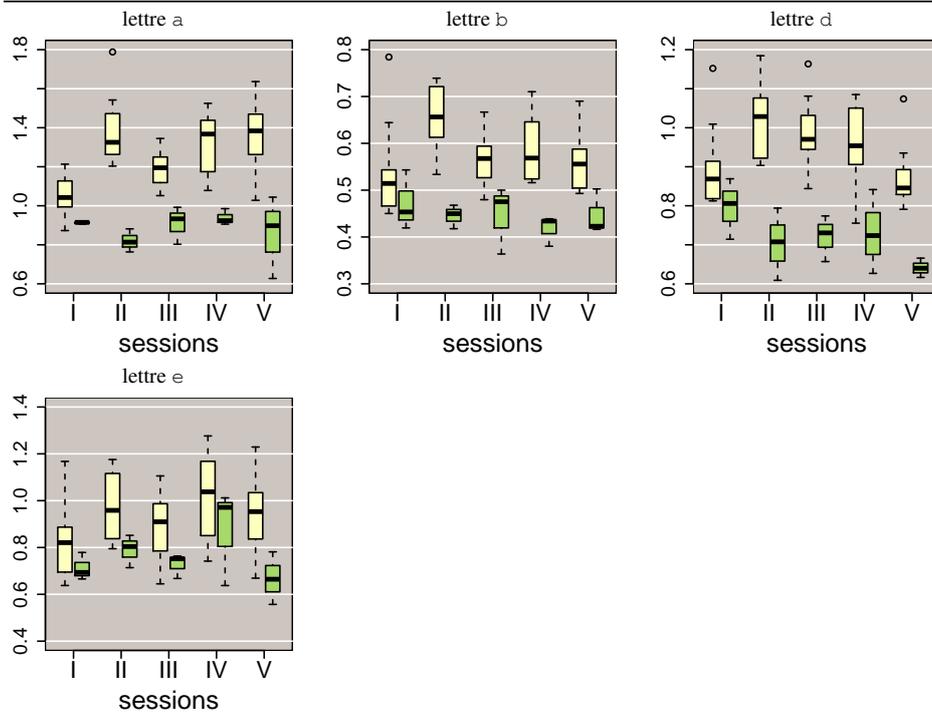


FIGURE 5.127: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

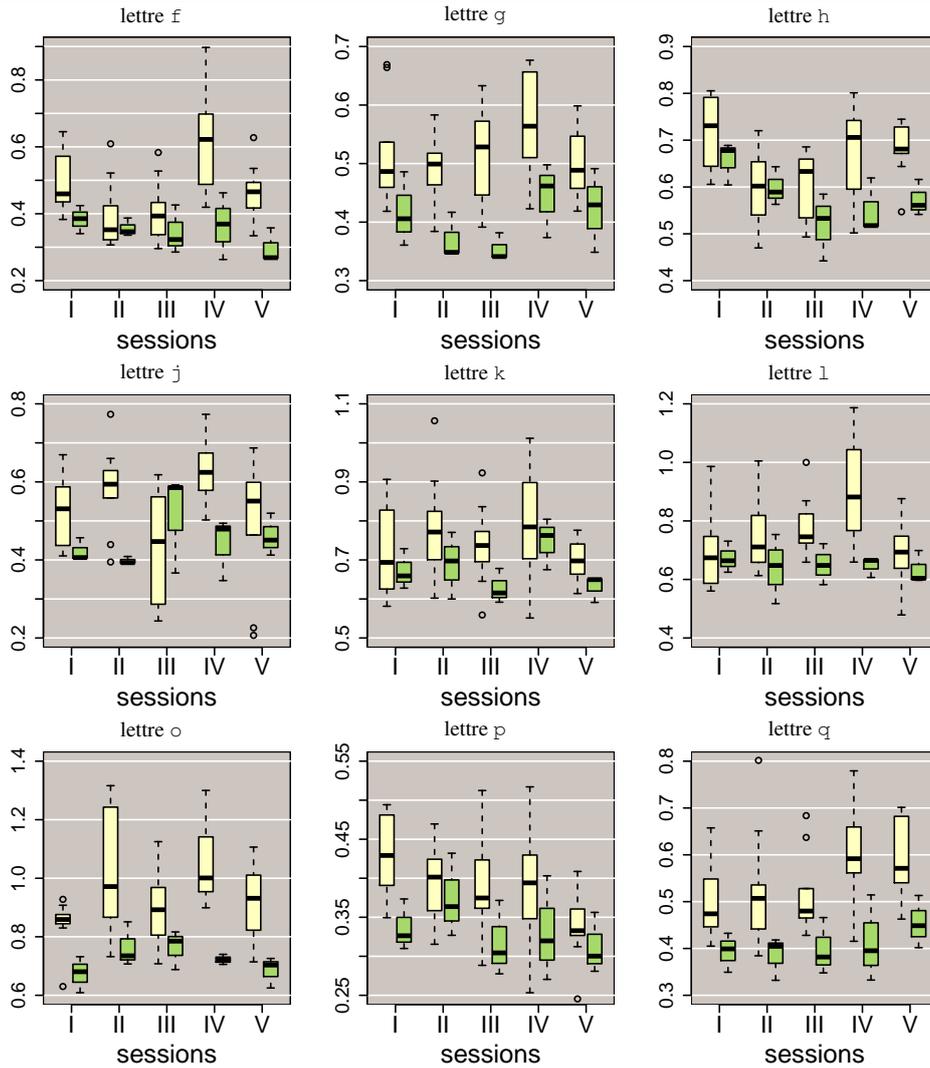


FIGURE 5.128: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

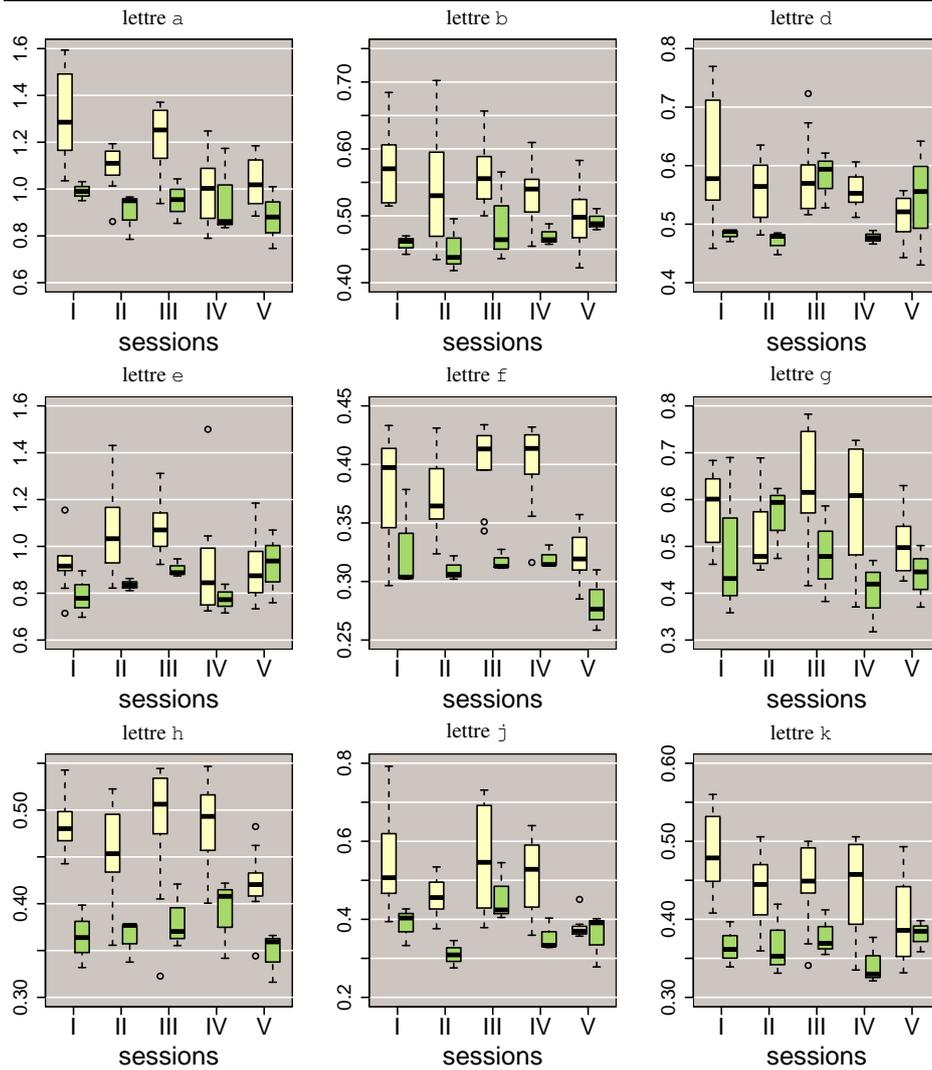


FIGURE 5.129: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

Scripteur 7, rapport  $W/H$

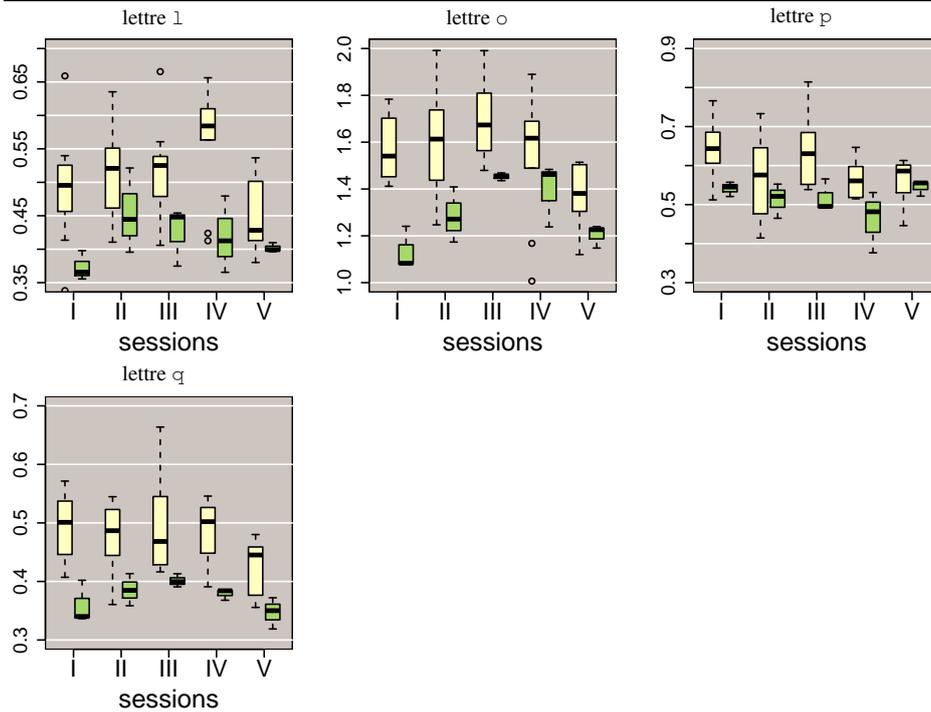


FIGURE 5.130: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

Scripteur 8, rapport  $W/H$

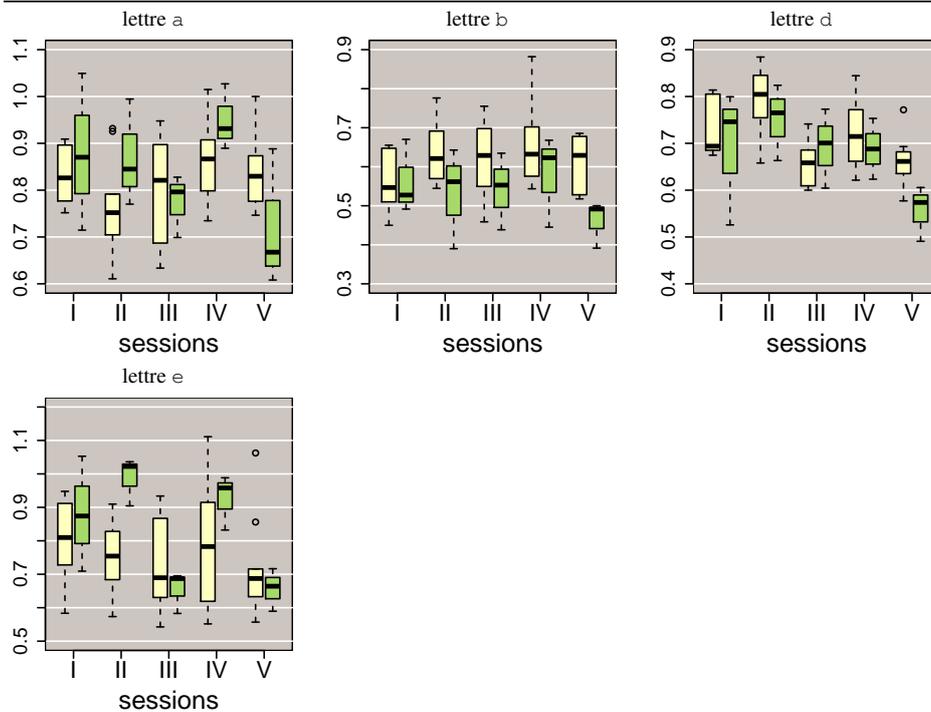


FIGURE 5.131: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

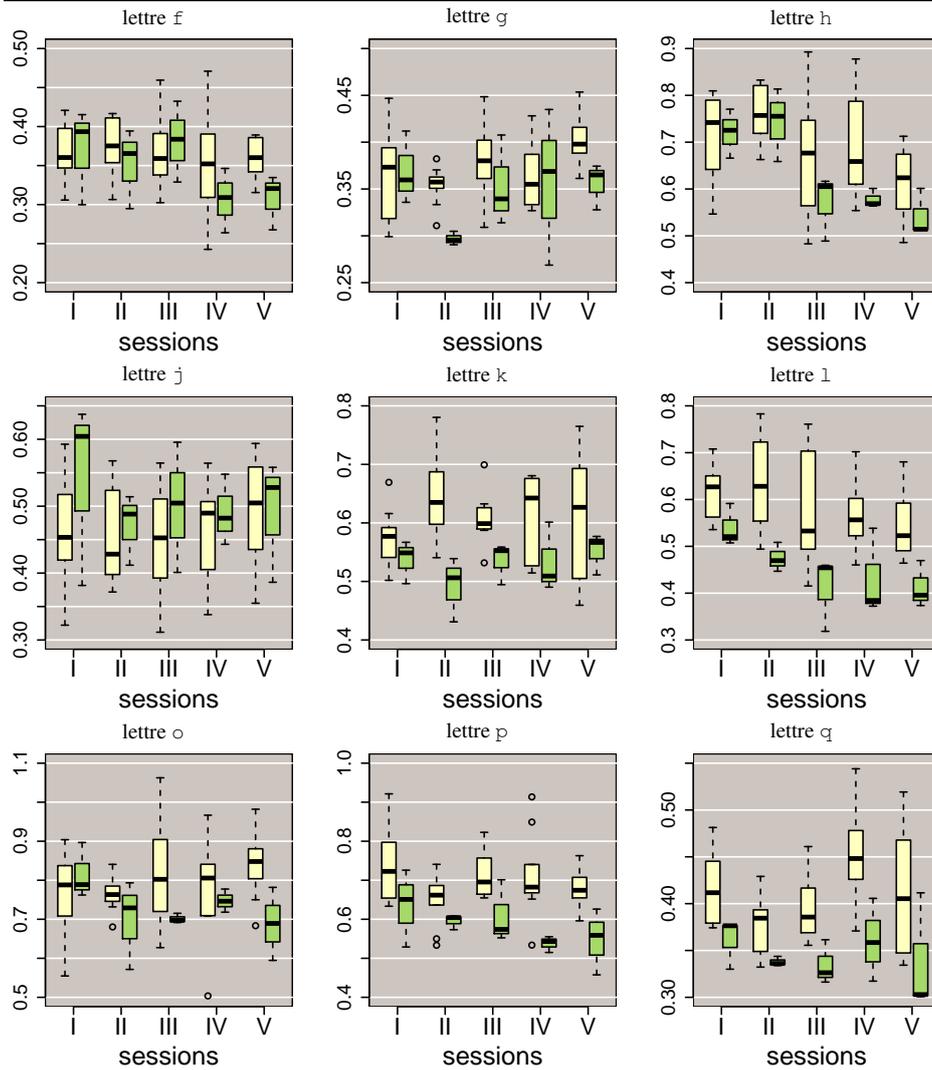


FIGURE 5.132: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

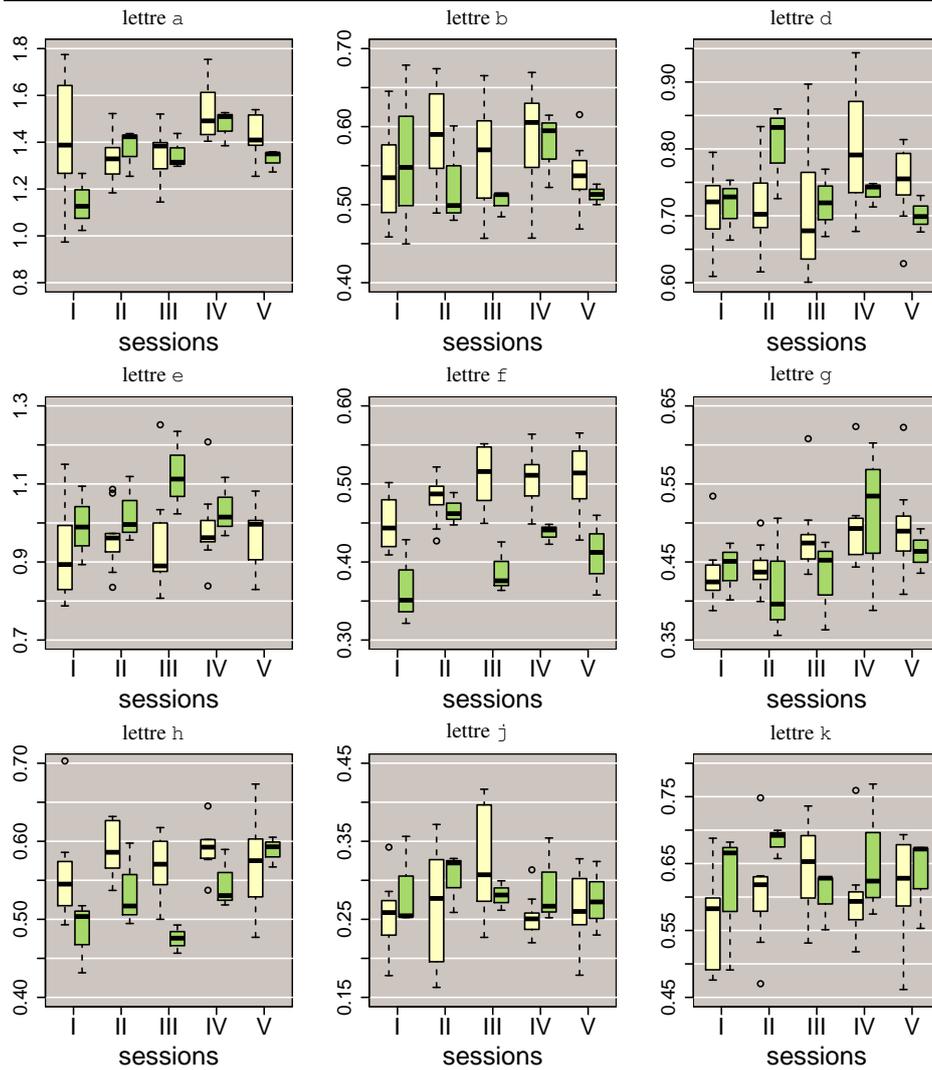


FIGURE 5.133: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

Scripteur 9, rapport  $W/H$

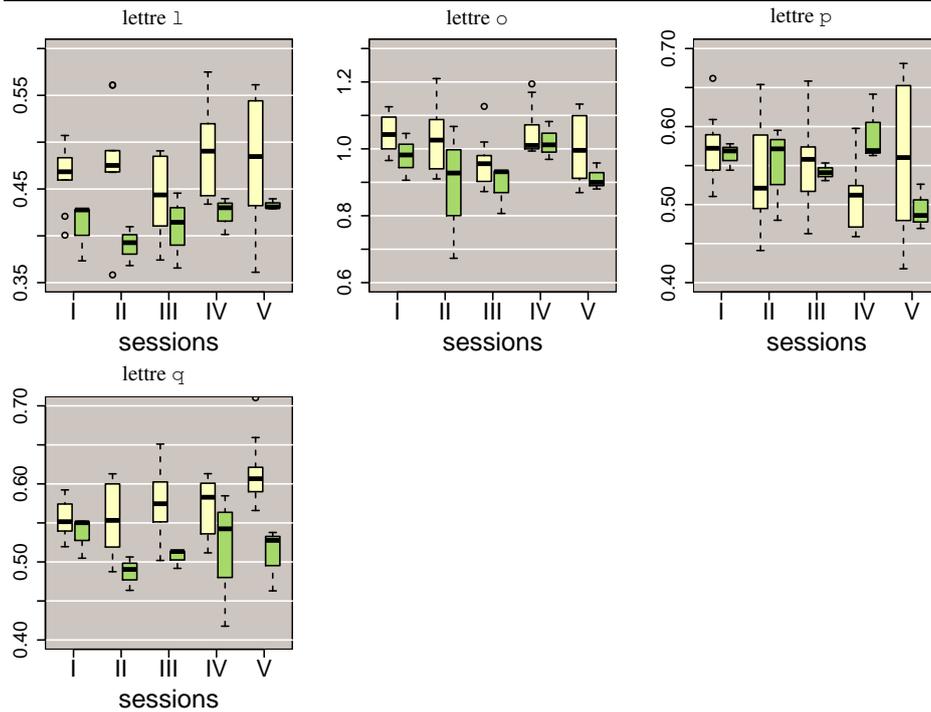


FIGURE 5.134: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

Scripteur 10, rapport  $W/H$

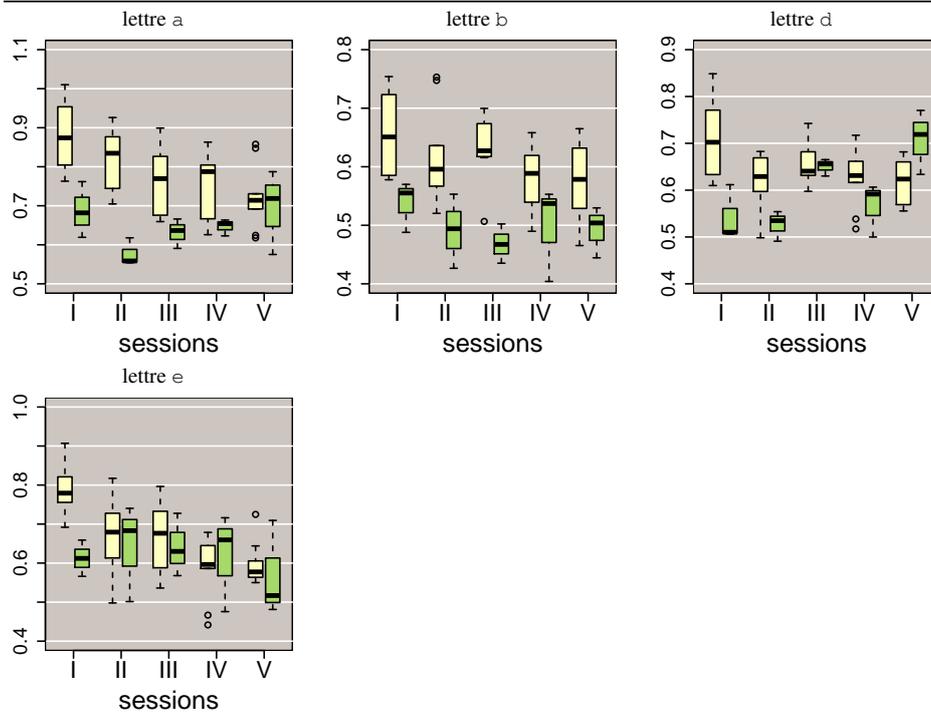


FIGURE 5.135: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

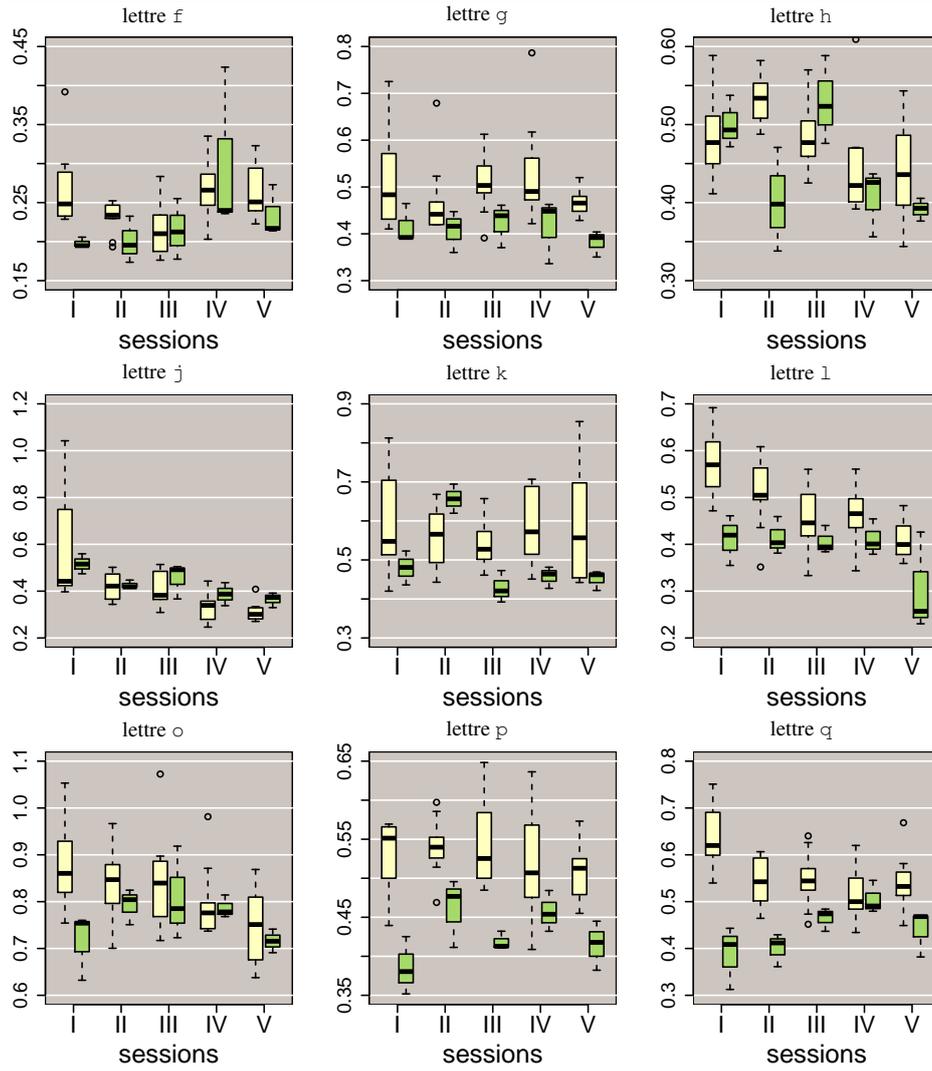


FIGURE 5.136: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

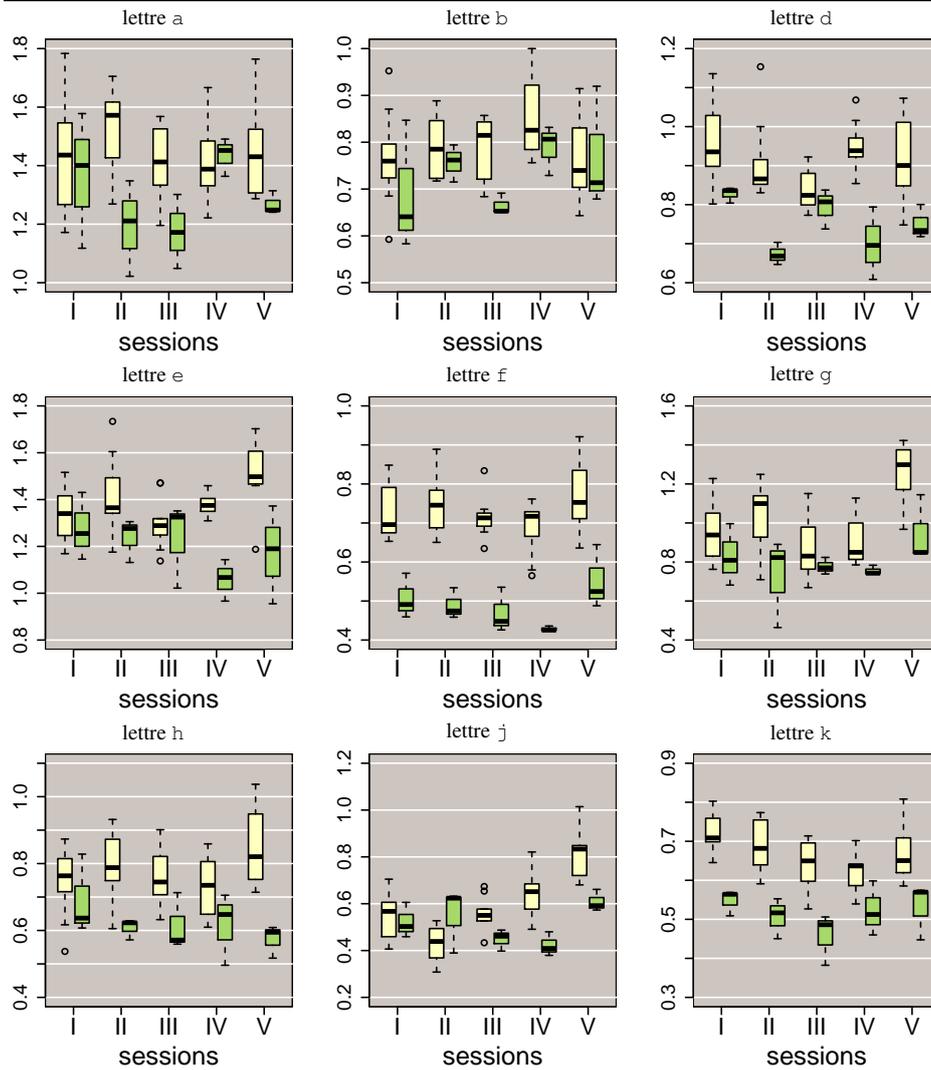


FIGURE 5.137: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

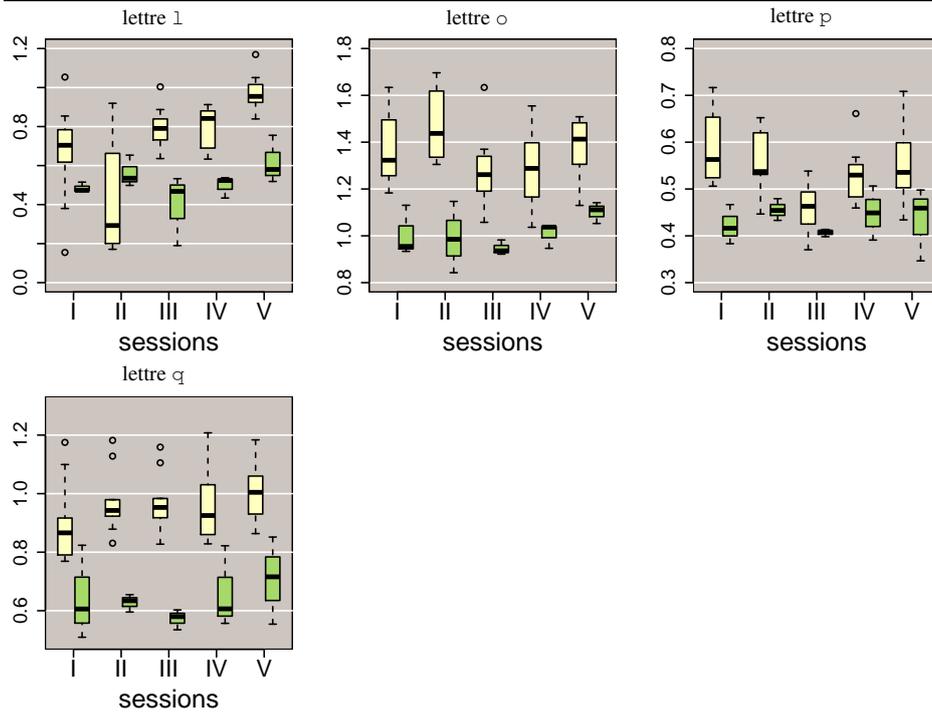


FIGURE 5.138: Boîtes à moustache des rapports  $W/H$  selon les pages  $I - V$ .

# Etude de la distribution des $F_{ratio}$

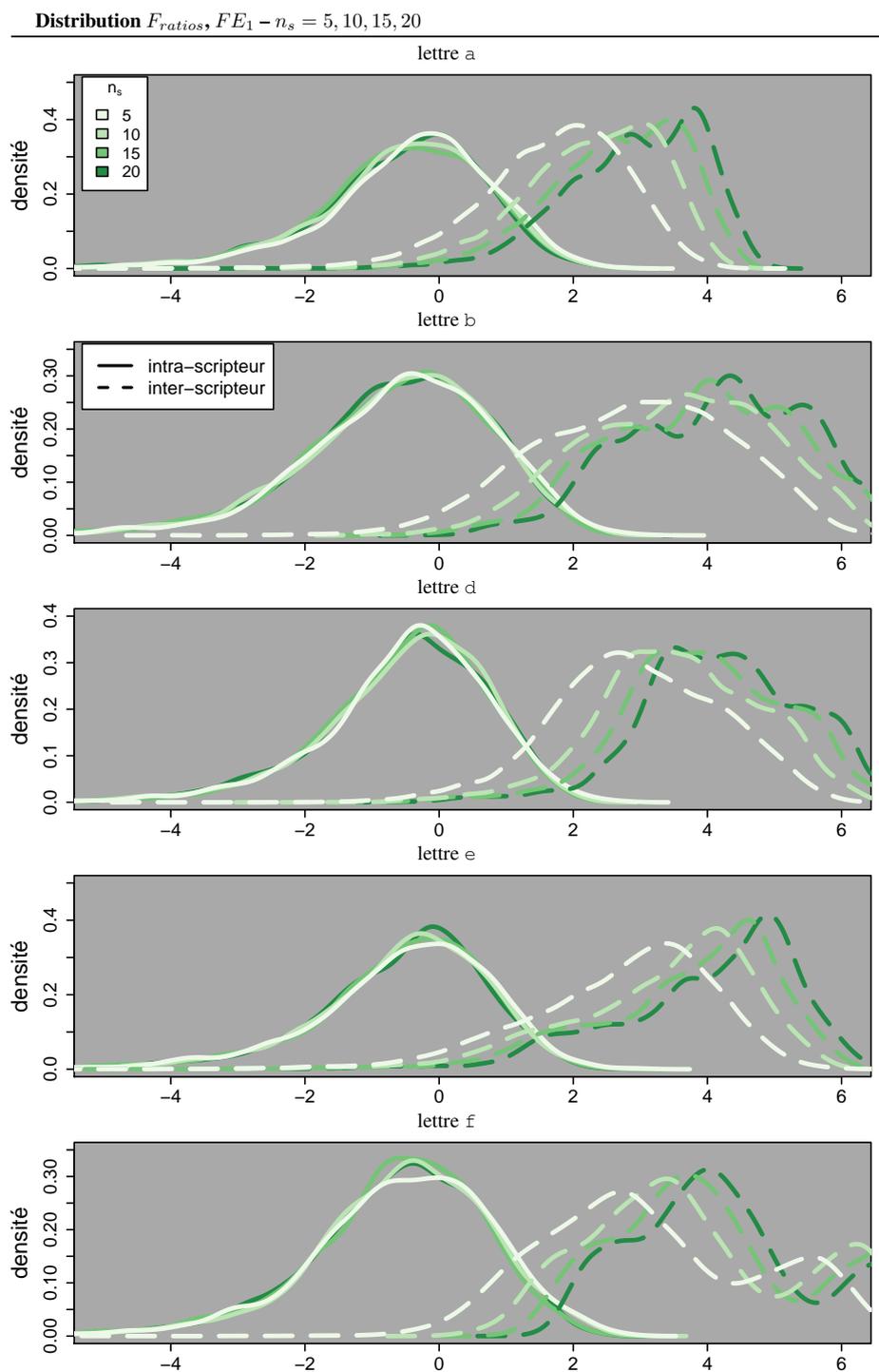


FIGURE 5.139: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1 à 4 de l'expérience 2.

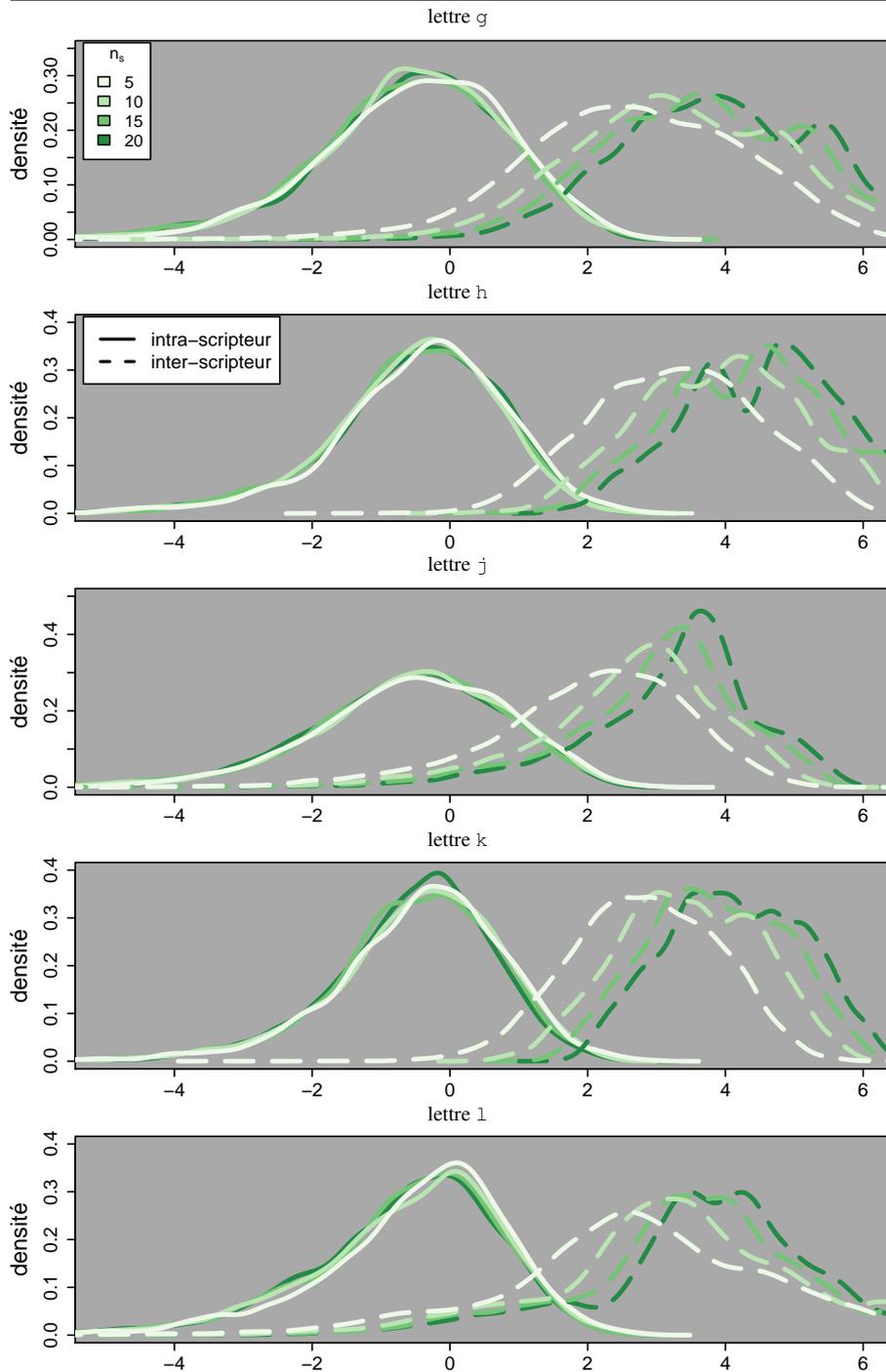


FIGURE 5.140: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1 à 4 de l'expérience 2.

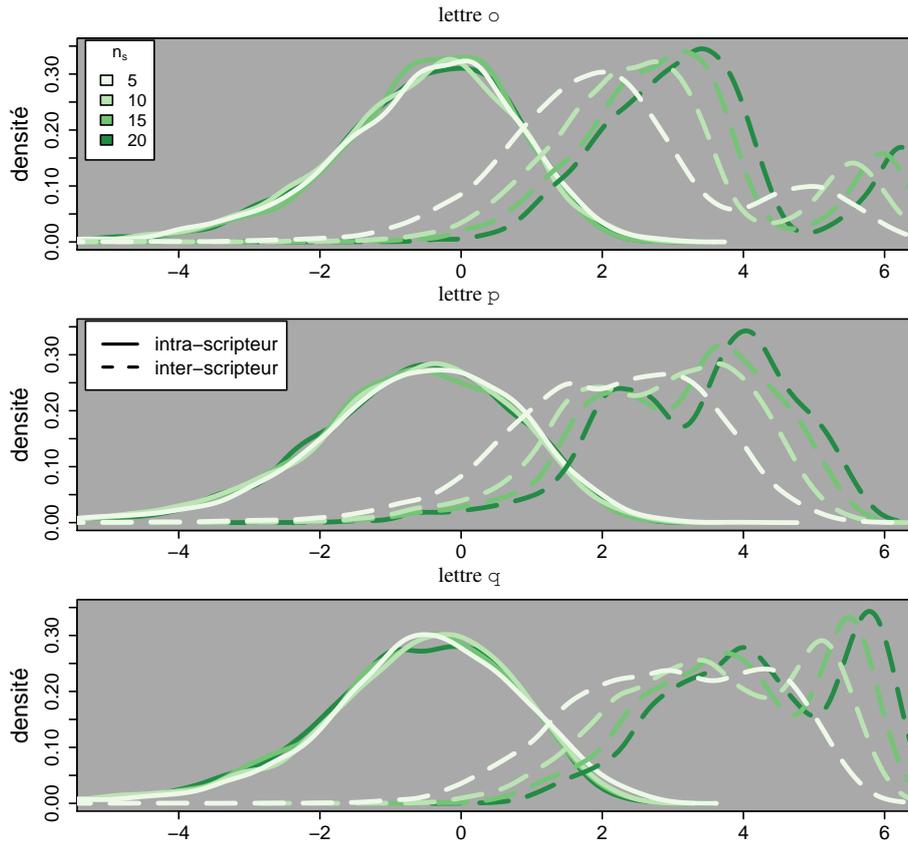


FIGURE 5.141: Distributions des  $F_{ratio}$  obtenus dans les *scénarios* 1 à 4 de l'expérience 2.

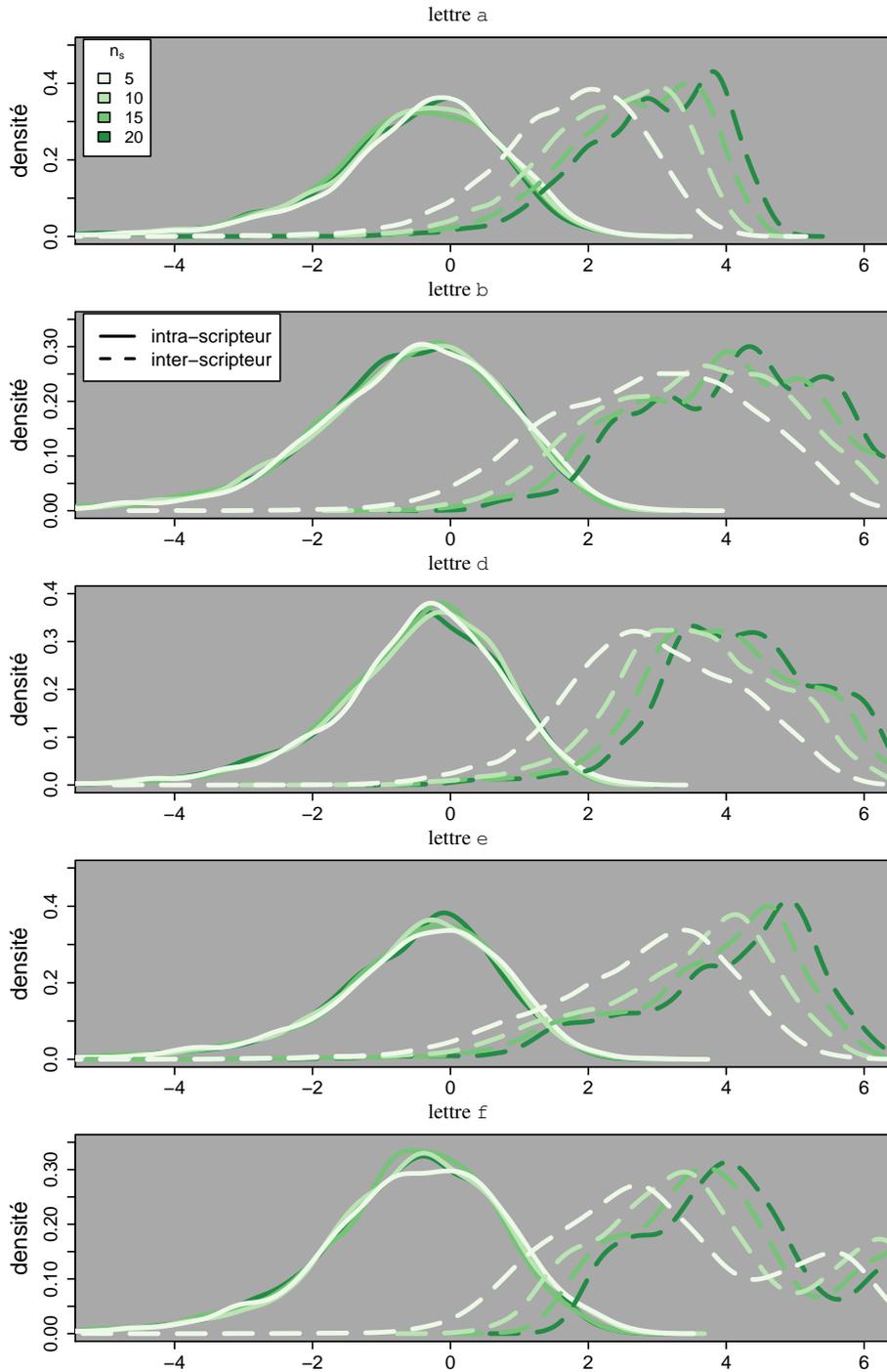


FIGURE 5.142: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1 à 4 de l'expérience 2.

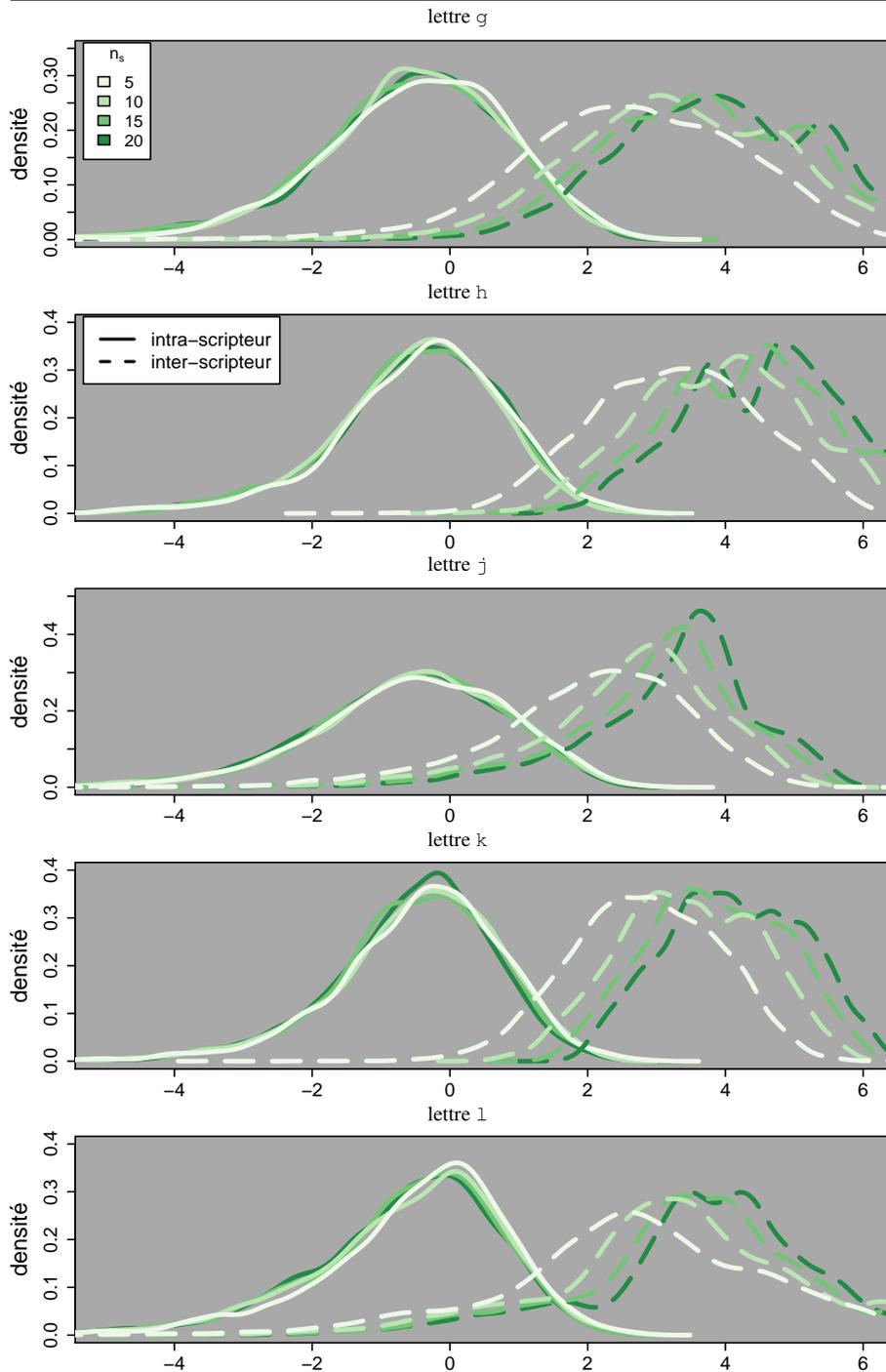


FIGURE 5.143: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1 à 4 de l'expérience 2.

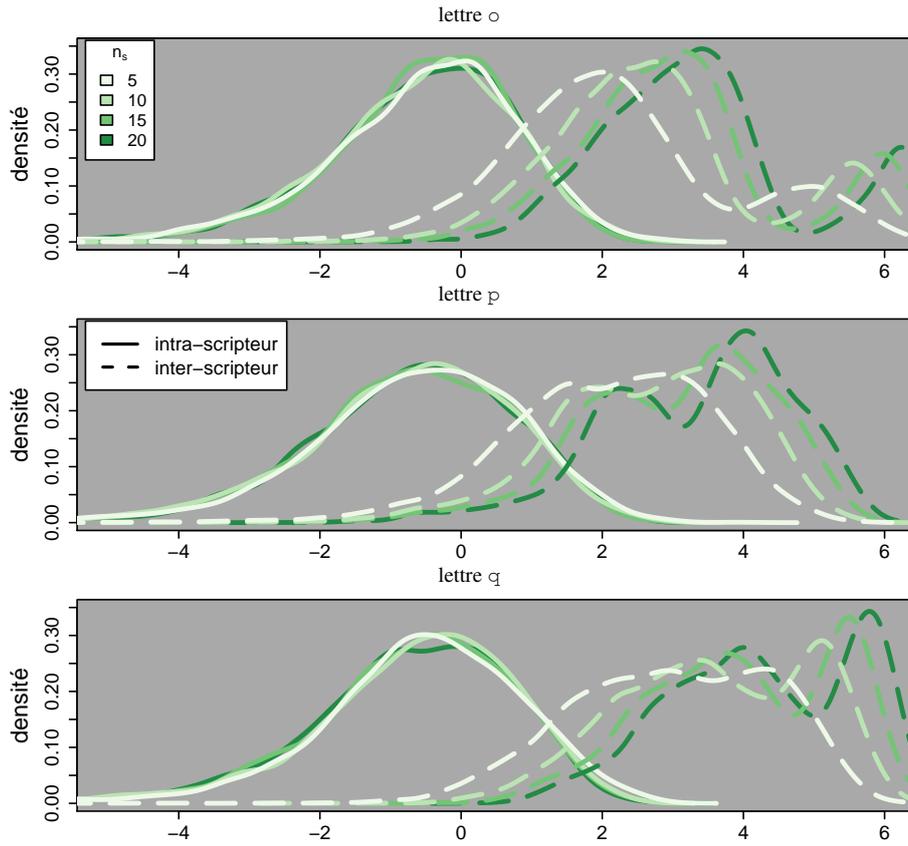


FIGURE 5.144: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1 à 4 de l'expérience 2.

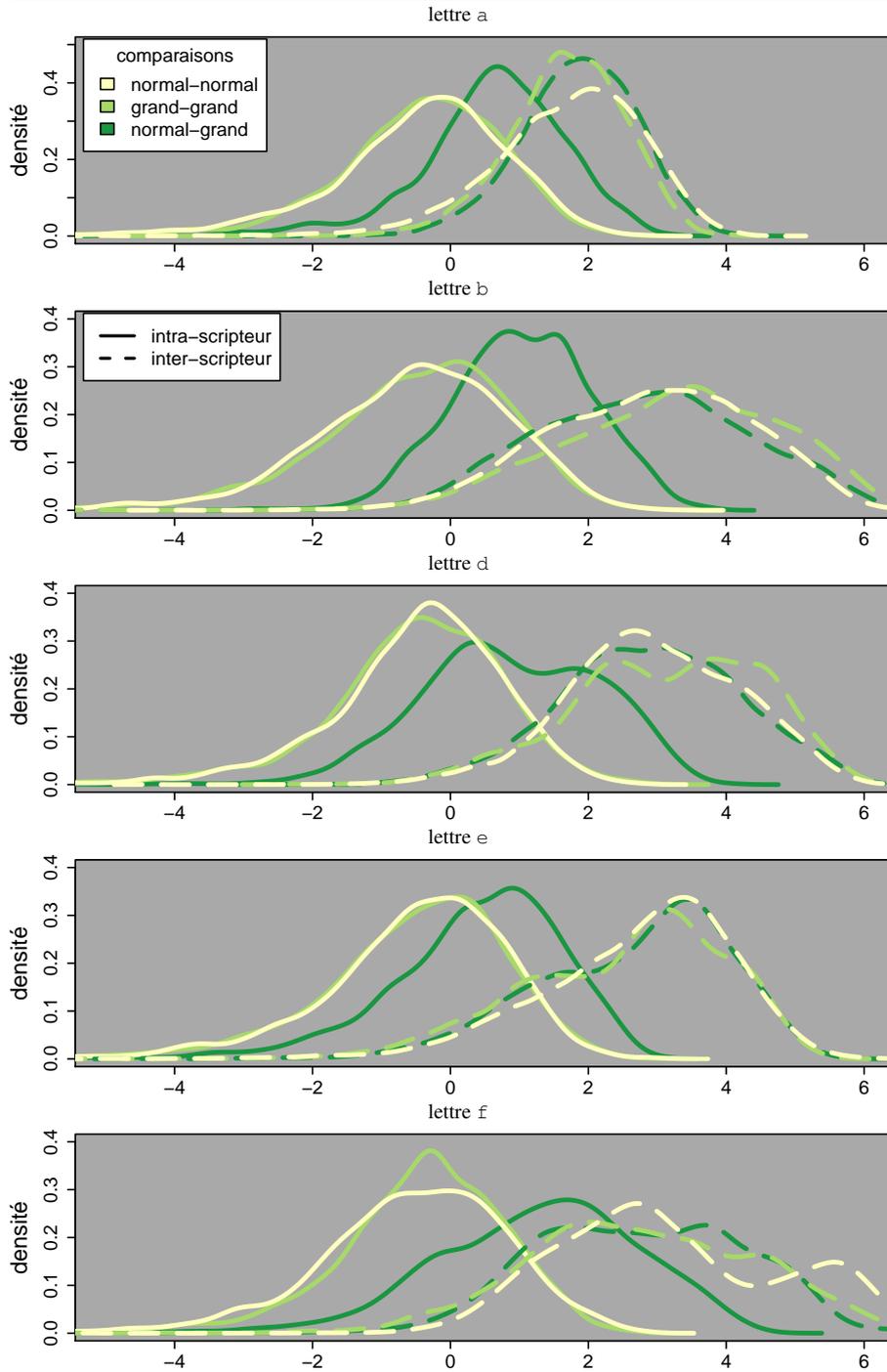


FIGURE 5.145: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1, 5 et 6 de l'expérience 2.

Distribution  $F_{ratios}, FE_1$  – scénarios 1, 5 et 6

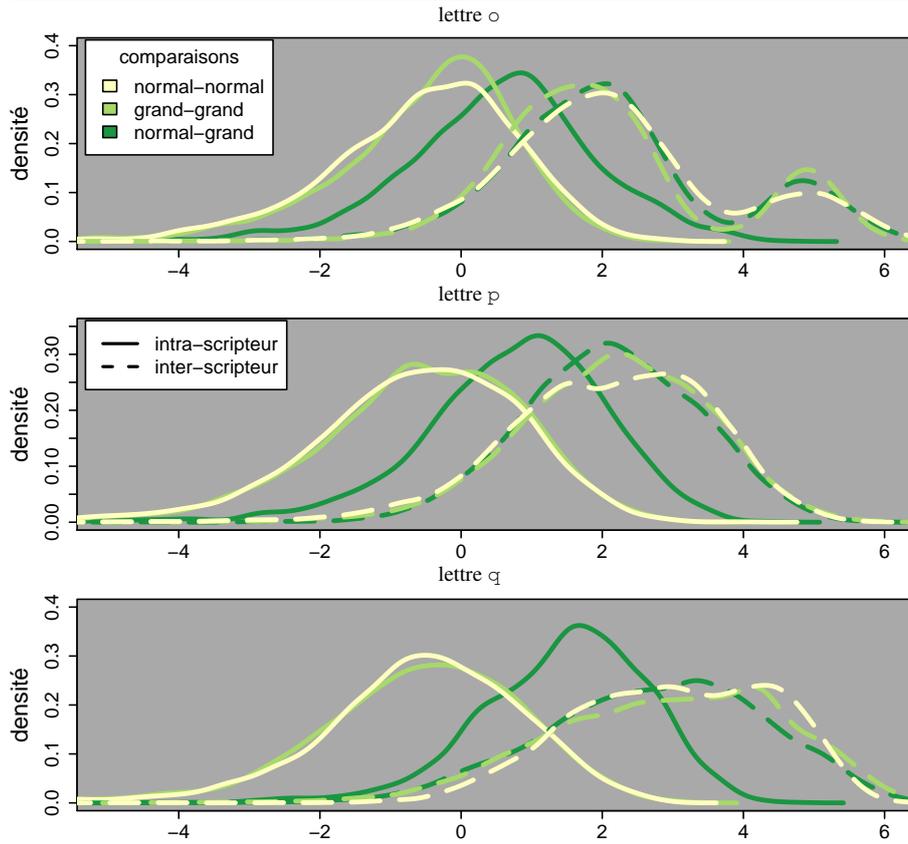


FIGURE 5.146: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1, 5 et 6 de l'expérience 2.

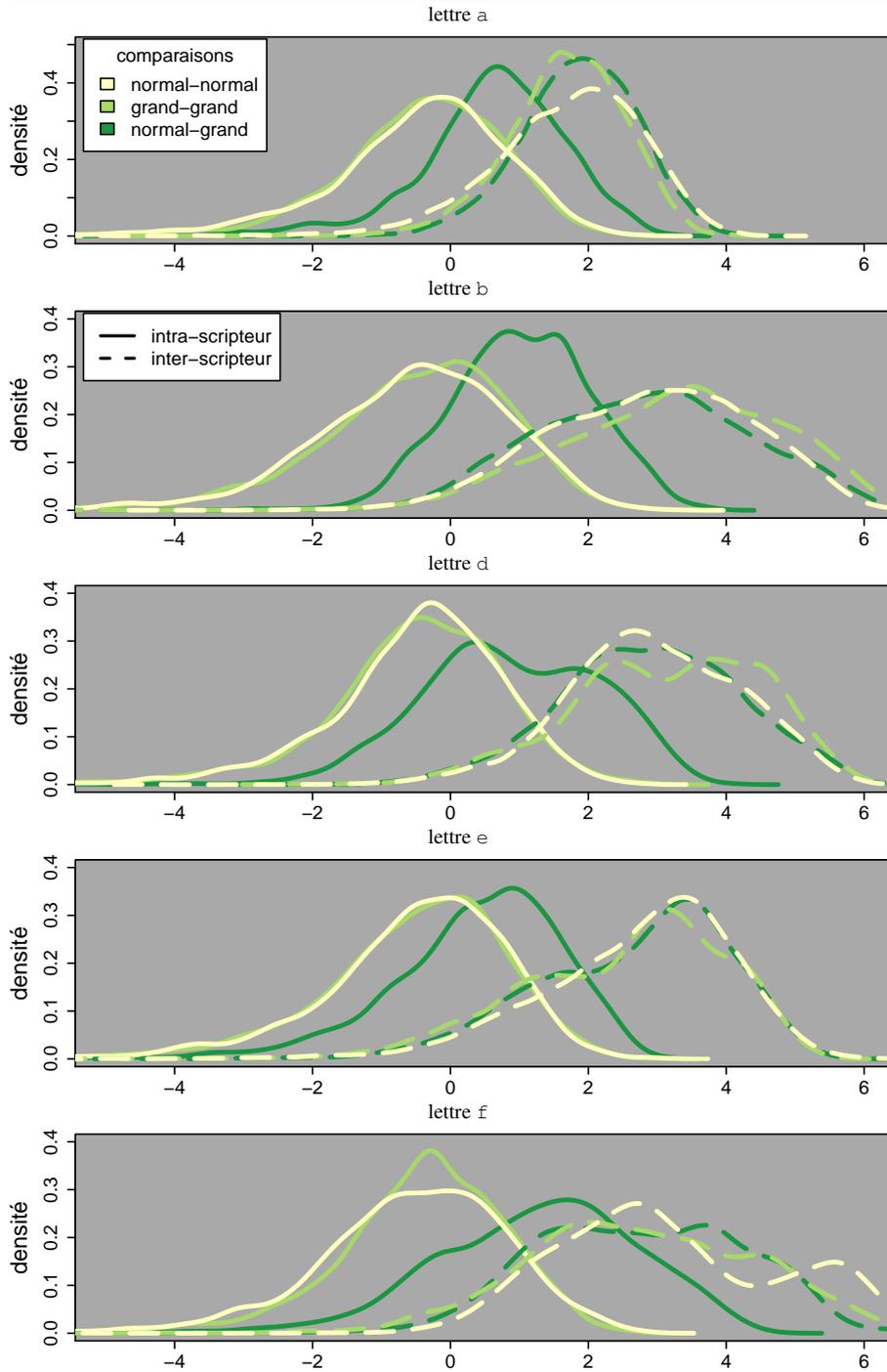


FIGURE 5.147: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1, 5 et 6 de l'expérience 2.

Distribution  $F_{ratios}$ ,  $FE_2$  – scénarios 1, 5 et 6

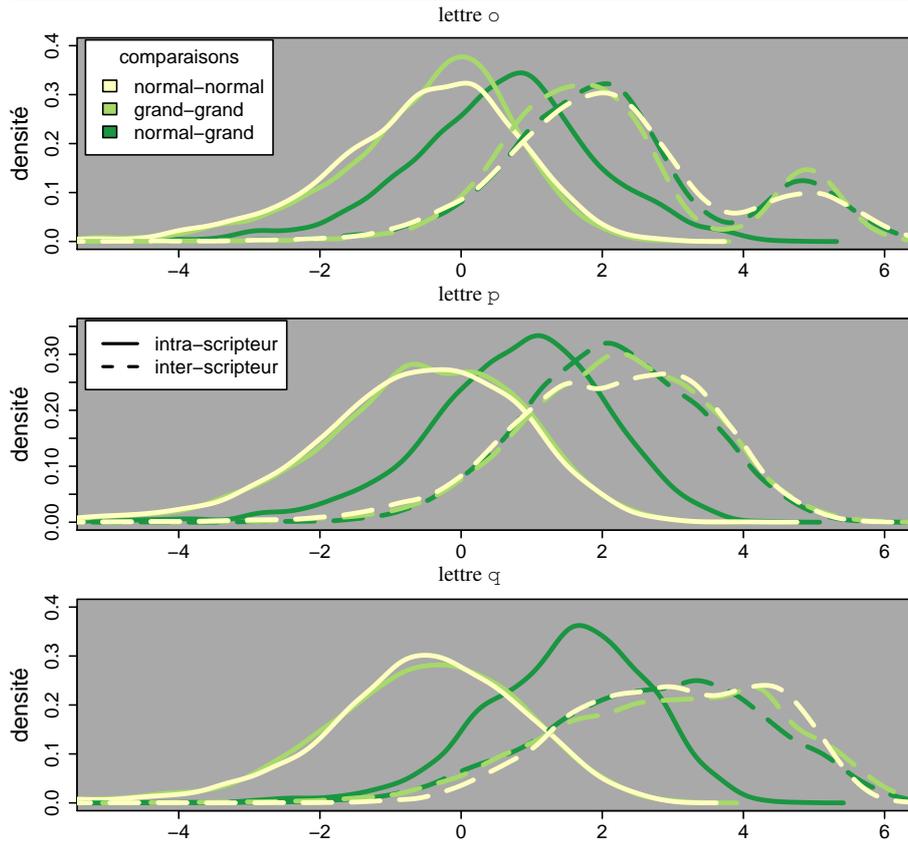


FIGURE 5.148: Distributions des  $F_{ratio}$  obtenus dans les scénarios 1, 5 et 6 de l'expérience 2.

## Les résultats du calcul des rapports de vraisemblance

|                                           |      | Lettre a                 |      |      |      |      |                           |      |      |     |      |                           |      |      |      |      |      |      |      |      |     |    |   |     |
|-------------------------------------------|------|--------------------------|------|------|------|------|---------------------------|------|------|-----|------|---------------------------|------|------|------|------|------|------|------|------|-----|----|---|-----|
|                                           |      | scénario 1 ( $n_s = 5$ ) |      |      |      |      | scénario 2 ( $n_s = 10$ ) |      |      |     |      | scénario 3 ( $n_s = 15$ ) |      |      |      |      |      |      |      |      |     |    |   |     |
| Scripteur                                 | 1    | 2                        | 3    | 4    | 5    | 6    | 7                         | 8    | 9    | 10  | 11   | 1                         | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10  | 11 |   |     |
| 1                                         | 11   | -                        | -    | -    | -    | -    | -                         | -    | -    | -   | -    | 1.5                       | -    | -    | -    | -    | -    | -    | -    | -    | -   | -  | - |     |
| 2                                         | 20   | 15.5                     | -    | -    | -    | -    | -                         | -    | -    | -   | -    | 4                         | 3.5  | -    | -    | -    | -    | -    | -    | -    | -   | -  | - |     |
| 3                                         | 15   | 33.5                     | 5.5  | -    | -    | -    | -                         | -    | -    | -   | -    | 7.5                       | 17.5 | 4    | -    | -    | -    | -    | -    | -    | -   | -  | - |     |
| 4                                         | 34.5 | 1.5                      | 28   | 9    | -    | -    | -                         | -    | -    | -   | -    | 14.5                      | 0    | 19.5 | 3.5  | -    | -    | -    | -    | -    | -   | -  | - |     |
| 5                                         | 0    | 0                        | 0    | 0    | 9    | -    | -                         | -    | -    | -   | -    | 0                         | 0    | 0    | 0.5  | -    | -    | -    | -    | -    | -   | -  | - |     |
| 6                                         | 35.5 | 5.5                      | 52.5 | 45   | 0    | 8.5  | -                         | -    | -    | -   | -    | 11                        | 0    | 67.5 | 23   | 0    | 7.5  | -    | -    | -    | -   | -  | - |     |
| 7                                         | 32   | 2                        | 21.5 | 9    | 0    | 79   | 9                         | -    | -    | -   | -    | 19                        | 0    | 20   | 4.5  | 0    | 83   | 4    | -    | -    | -   | -  | - |     |
| 8                                         | 0.5  | 0                        | 0    | 0    | 15.5 | 0    | 0.5                       | 14.5 | -    | -   | -    | 0                         | 0    | 0    | 1    | 0    | 0    | 0.5  | -    | -    | -   | -  | - |     |
| 9                                         | 6    | 57.5                     | 71   | 4    | 0    | 11   | 1                         | 0    | 9.5  | -   | -    | 0.5                       | 41   | 70   | 0    | 0    | 2    | 0    | 0    | 5    | -   | -  | - |     |
| 10                                        | 3    | 0                        | 0    | 0    | 0    | 0    | 0                         | 10   | 0    | 8.5 | -    | 0                         | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 6    | -   | -  | - |     |
| 11                                        | 0.5  | 4                        | 36.5 | 1    | 0    | 9.5  | 3.5                       | 0    | 0.5  | 0   | 12.5 | 0                         | 0    | 14.5 | 0    | 0    | 1.5  | 0    | 0    | 0    | 4.5 | 0  | 0 | 3.5 |
| <i>scénario 4 (<math>n_s = 20</math>)</i> |      |                          |      |      |      |      |                           |      |      |     |      |                           |      |      |      |      |      |      |      |      |     |    |   |     |
| Scripteur                                 | 1    | 2                        | 3    | 4    | 5    | 6    | 7                         | 8    | 9    | 10  | 11   | 1                         | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10  | 11 |   |     |
| 1                                         | 0.5  | -                        | -    | -    | -    | -    | -                         | -    | -    | -   | -    | 6.5                       | -    | -    | -    | -    | -    | -    | -    | -    | -   | -  |   |     |
| 2                                         | 0    | 0                        | -    | -    | -    | -    | -                         | -    | -    | -   | -    | 0                         | 13   | -    | -    | -    | -    | -    | -    | -    | -   | -  |   |     |
| 3                                         | 0.5  | 2.5                      | 0.5  | -    | -    | -    | -                         | -    | -    | -   | -    | 29.5                      | 55.5 | 2.5  | -    | -    | -    | -    | -    | -    | -   | -  |   |     |
| 4                                         | 0    | 0                        | 3    | 0.5  | -    | -    | -                         | -    | -    | -   | -    | 69                        | 0    | 28   | 3.5  | -    | -    | -    | -    | -    | -   | -  |   |     |
| 5                                         | 0    | 0                        | 0    | 0    | 0    | -    | -                         | -    | -    | -   | -    | 0                         | 0    | 0    | 9    | -    | -    | -    | -    | -    | -   | -  |   |     |
| 6                                         | 0    | 0                        | 85   | 8    | 0    | 1.5  | -                         | -    | -    | -   | -    | 22                        | 0    | 5.5  | 5.5  | 0    | 3.5  | -    | -    | -    | -   | -  |   |     |
| 7                                         | 1.5  | 0                        | 7.5  | 0    | 0    | 91   | 1.5                       | -    | -    | -   | -    | 32.5                      | 0    | 0.5  | 4    | 2    | 48   | 4    | -    | -    | -   | -  |   |     |
| 8                                         | 0    | 0                        | 0    | 0    | 0    | 0    | 0.5                       | -    | -    | -   | -    | 4.5                       | 0    | 1    | 1    | 20   | 78   | 62   | 7.5  | -    | -   | -  |   |     |
| 9                                         | 0    | 20.5                     | 59   | 0    | 0    | 0    | 0                         | 0.5  | -    | -   | -    | 1                         | 1.5  | 80.5 | 0    | 0    | 0    | 0    | 6.5  | -    | -   | -  |   |     |
| 10                                        | 0    | 0                        | 0    | 0    | 0    | 0    | 0                         | 0    | 0.5  | -   | -    | 0                         | 0    | 0    | 0    | 0    | 0    | 0    | 9.5  | -    | -   | -  |   |     |
| 11                                        | 0    | 0                        | 0.5  | 0    | 0    | 0    | 0                         | 0    | 0    | 3   | 0    | 0                         | 0    | 5    | 0    | 0    | 0    | 0    | 11.5 | 0    | 8.5 | 74 |   |     |
| <i>scénario 5 (<math>n_s = 5</math>)</i>  |      |                          |      |      |      |      |                           |      |      |     |      |                           |      |      |      |      |      |      |      |      |     |    |   |     |
| Scripteur                                 | 1    | 2                        | 3    | 4    | 5    | 6    | 7                         | 8    | 9    | 10  | 11   | 1                         | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10  | 11 |   |     |
| 1                                         | 11.5 | -                        | -    | -    | -    | -    | -                         | -    | -    | -   | -    | 11.5                      | -    | -    | -    | -    | -    | -    | -    | -    | -   | -  |   |     |
| 2                                         | 58.5 | 74                       | -    | -    | -    | -    | -                         | -    | -    | -   | -    | 58.5                      | 74   | -    | -    | -    | -    | -    | -    | -    | -   | -  |   |     |
| 3                                         | 97   | 28.5                     | 30   | -    | -    | -    | -                         | -    | -    | -   | -    | 97                        | 28.5 | 30   | -    | -    | -    | -    | -    | -    | -   | -  |   |     |
| 4                                         | 54.5 | 1                        | 57.5 | 28.5 | -    | -    | -                         | -    | -    | -   | -    | 54.5                      | 1    | 57.5 | 28.5 | -    | -    | -    | -    | -    | -   | -  |   |     |
| 5                                         | 41.5 | 0.5                      | 4    | 0    | 25.5 | -    | -                         | -    | -    | -   | -    | 41.5                      | 0.5  | 4    | 0    | 25.5 | -    | -    | -    | -    | -   | -  |   |     |
| 6                                         | 52   | 0.5                      | 5.5  | 9.5  | 0    | 95.5 | -                         | -    | -    | -   | -    | 52                        | 0.5  | 5.5  | 9.5  | 0    | 95.5 | -    | -    | -    | -   | -  |   |     |
| 7                                         | 94.5 | 19.5                     | 50.5 | 12   | 13   | 30.5 | 27                        | -    | -    | -   | -    | 94.5                      | 19.5 | 50.5 | 12   | 13   | 30.5 | 27   | -    | -    | -   | -  |   |     |
| 8                                         | 70.5 | 1                        | 2    | 4    | 6    | 1.5  | 11.5                      | 19.5 | -    | -   | -    | 70.5                      | 1    | 2    | 4    | 6    | 1.5  | 11.5 | 19.5 | -    | -   | -  |   |     |
| 9                                         | 94   | 31                       | 39   | 15   | 0    | 24.5 | 25.5                      | 0    | 35.5 | -   | -    | 94                        | 31   | 39   | 15   | 0    | 24.5 | 25.5 | 0    | 35.5 | -   | -  |   |     |
| 10                                        | 26.5 | 0                        | 0.5  | 0    | 3    | 0    | 0                         | 56.5 | 0    | 8   | -    | 26.5                      | 0    | 0.5  | 0    | 3    | 0    | 0    | 56.5 | 0    | 8   | -  |   |     |
| 11                                        | 74   | 40                       | 16   | 77   | 0    | 50.5 | 46.5                      | 1    | 2.5  | 0   | 40   | 74                        | 40   | 16   | 77   | 0    | 50.5 | 46.5 | 1    | 2.5  | 0   | 40 |   |     |

TABLE 5.1: Résultats des calculs de rapport de vraisemblance de la lettre a obtenus dans l'expérience 2 (*scénarios 1 à 6*).





|           |      | Lettre e                 |      |      |     |      |                           |      |     |    |    |                           |    |     |      |   |     |   |     |     |    |    |      |      |   |      |   |     |   |   |   |    |    |   |   |   |   |   |   |   |   |   |     |     |   |
|-----------|------|--------------------------|------|------|-----|------|---------------------------|------|-----|----|----|---------------------------|----|-----|------|---|-----|---|-----|-----|----|----|------|------|---|------|---|-----|---|---|---|----|----|---|---|---|---|---|---|---|---|---|-----|-----|---|
|           |      | scénario 1 ( $n_s = 5$ ) |      |      |     |      | scénario 2 ( $n_s = 10$ ) |      |     |    |    | scénario 3 ( $n_s = 15$ ) |    |     |      |   |     |   |     |     |    |    |      |      |   |      |   |     |   |   |   |    |    |   |   |   |   |   |   |   |   |   |     |     |   |
| Scripteur | 1    | 2                        | 3    | 4    | 5   | 6    | 7                         | 8    | 9   | 10 | 11 | 1                         | 2  | 3   | 4    | 5 | 6   | 7 | 8   | 9   | 10 | 11 | 1    | 2    | 3 | 4    | 5 | 6   | 7 | 8 | 9 | 10 | 11 |   |   |   |   |   |   |   |   |   |     |     |   |
| 1         | 8    | -                        | -    | -    | -   | -    | -                         | -    | -   | -  | -  | 2.5                       | -  | -   | -    | - | -   | - | -   | -   | -  | -  | 0.5  | -    | - | -    | - | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -   | -   |   |
| 2         | 0    | 4.5                      | -    | -    | -   | -    | -                         | -    | -   | -  | -  | 0                         | 1  | -   | -    | - | -   | - | -   | -   | -  | -  | 0    | 1    | - | -    | - | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -   | -   |   |
| 3         | 0    | 7                        | 8.5  | -    | -   | -    | -                         | -    | -   | -  | -  | 0                         | 0  | 2.5 | -    | - | -   | - | -   | -   | -  | -  | 0    | 0    | 1 | -    | - | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -   | -   |   |
| 4         | 56.5 | 0                        | 0    | 5.5  | -   | -    | -                         | -    | -   | -  | -  | 37                        | 0  | 0   | 2    | - | -   | - | -   | -   | -  | -  | 21.5 | 0    | 0 | 0    | - | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -   | -   |   |
| 5         | 29   | 0                        | 0    | 0.5  | 6.5 | -    | -                         | -    | -   | -  | -  | 23.5                      | 0  | 0   | 1.5  | - | -   | - | -   | -   | -  | -  | 2    | 0    | 0 | 0    | - | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -   | -   |   |
| 6         | 69.5 | 0                        | 0    | 72.5 | 5   | 3    | -                         | -    | -   | -  | -  | 70.5                      | 0  | 0   | 53.5 | 0 | 0   | - | -   | -   | -  | -  | 58.5 | 0    | 0 | 39.5 | 0 | 0.5 | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -   | -   |   |
| 7         | 25   | 0                        | 0.5  | 25   | 0   | 13.5 | 5.5                       | -    | -   | -  | -  | 6                         | 0  | 0   | 11.5 | 0 | 0.5 | 7 | -   | -   | -  | -  | 0.5  | 0    | 0 | 1.5  | 0 | 0   | 3 | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -   | -   |   |
| 8         | 60.5 | 0                        | 0    | 2.5  | 1   | 25.5 | 1                         | 7.5  | -   | -  | -  | 42                        | 0  | 0   | 0    | 0 | 3   | 0 | 1.5 | -   | -  | -  | 19.5 | 0    | 0 | 0    | 0 | 0   | 0 | 0 | 1 | -  | -  | - | - | - | - | - | - | - | - | - | -   | -   |   |
| 9         | 0    | 38                       | 0.5  | 0    | 0   | 0    | 0                         | 0    | 6.5 | -  | -  | 0                         | 22 | 0   | 0    | 0 | 0   | 0 | 0   | 1.5 | -  | -  | 0    | 10.5 | 0 | 0    | 0 | 0   | 0 | 0 | 0 | 0  | 0  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.5 | -   |   |
| 10        | 53.5 | 0                        | 0    | 22   | 0   | 15.5 | 0                         | 28.5 | 0   | 3  | -  | 50.5                      | 0  | 0   | 2    | 0 | 1   | 0 | 8   | 0   | 1  | -  | 36   | 0    | 0 | 0    | 0 | 0   | 0 | 0 | 0 | 0  | 0  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0   | 0.5 | - |
| 11        | 0    | 3                        | 76.5 | 0    | 0   | 0    | 0                         | 0    | 0   | 0  | 7  | 0                         | 0  | 0   | 75.5 | 0 | 0   | 0 | 0   | 0   | 0  | 3  | 0    | 0    | 0 | 87.5 | 0 | 0   | 0 | 0 | 0 | 0  | 0  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0   | 1.5 | - |

|    |      | scénario 4 ( $n_s = 20$ ) |      |    |   |   |     |     |   |   |     |      | scénario 5 ( $n_s = 5$ ) |      |      |     |      |      |     |   |      |    |      | scénario 6 ( $n_s = 5$ ) |      |      |      |    |      |      |    |   |      |    |   |   |   |   |   |   |   |   |   |    |
|----|------|---------------------------|------|----|---|---|-----|-----|---|---|-----|------|--------------------------|------|------|-----|------|------|-----|---|------|----|------|--------------------------|------|------|------|----|------|------|----|---|------|----|---|---|---|---|---|---|---|---|---|----|
|    |      | 1                         | 2    | 3  | 4 | 5 | 6   | 7   | 8 | 9 | 10  | 11   | 1                        | 2    | 3    | 4   | 5    | 6    | 7   | 8 | 9    | 10 | 11   | 1                        | 2    | 3    | 4    | 5  | 6    | 7    | 8  | 9 | 10   | 11 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1  | 1.5  | -                         | -    | -  | - | - | -   | -   | - | - | -   | 11   | -                        | -    | -    | -   | -    | -    | -   | - | -    | -  | 1    | -                        | -    | -    | -    | -  | -    | -    | -  | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 2  | 0    | 0                         | -    | -  | - | - | -   | -   | - | - | -   | 0    | 4.5                      | -    | -    | -   | -    | -    | -   | - | -    | -  | 0.5  | 3.5                      | -    | -    | -    | -  | -    | -    | -  | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 3  | 0    | 0                         | 1    | -  | - | - | -   | -   | - | - | -   | 0    | 34.5                     | 7    | -    | -   | -    | -    | -   | - | -    | -  | 9.5  | 99                       | 0.5  | -    | -    | -  | -    | -    | -  | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 4  | 5.5  | 0                         | 0    | 0  | - | - | -   | -   | - | - | -   | 28.5 | 0                        | 0    | 8.5  | -   | -    | -    | -   | - | -    | -  | 83   | 0                        | 0    | 3.5  | -    | -  | -    | -    | -  | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 5  | 0.5  | 0                         | 0    | 0  | 0 | - | -   | -   | - | - | -   | 47   | 0                        | 0    | 0    | 8.5 | -    | -    | -   | - | -    | -  | 30.5 | 0                        | 0    | 46.5 | 45.5 | -  | -    | -    | -  | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 6  | 41.5 | 0                         | 0    | 15 | 0 | 0 | -   | -   | - | - | -   | 42   | 0                        | 0    | 28   | 0   | 6.5  | -    | -   | - | -    | -  | 65   | 0                        | 0    | 72.5 | 0    | 56 | -    | -    | -  | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 7  | 0    | 0                         | 0    | 0  | 0 | 0 | 1.5 | -   | - | - | -   | 7    | 0                        | 0    | 4.5  | 0   | 1    | 7    | -   | - | -    | -  | 23   | 0                        | 0    | 16   | 0    | 15 | 18.5 | -    | -  | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 8  | 4.5  | 0                         | 0    | 0  | 0 | 0 | 0   | 0.5 | - | - | -   | 52   | 0                        | 0    | 63   | 0   | 79.5 | 61.5 | 5.5 | - | -    | -  | 92.5 | 0                        | 0    | 82   | 0    | 79 | 42   | 17.5 | -  | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 9  | 0    | 2.5                       | 0    | 0  | 0 | 0 | 0   | 0   | 0 | 1 | -   | 0    | 84.5                     | 11.5 | 0    | 0   | 0    | 0    | 1.5 | - | -    | -  | 0    | 85.5                     | 10.5 | 5.5  | 0    | 0  | 1.5  | 0    | 21 | - | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 10 | 13   | 0                         | 0    | 0  | 0 | 0 | 0   | 0   | 0 | 0 | 0.5 | 54.5 | 0                        | 0    | 29.5 | 0   | 23   | 0    | 35  | 0 | 12.5 | -  | 93   | 0                        | 0    | 94   | 0    | 59 | 2.5  | 91   | 0  | 0 | -    | -  | - | - | - | - | - | - | - | - | - | -  |
| 11 | 0    | 0                         | 86.5 | 0  | 0 | 0 | 0   | 0   | 0 | 0 | 1.5 | 0    | 0                        | 0    | 38   | 0   | 0    | 0    | 0   | 0 | 0    | 11 | 0    | 38                       | 73.5 | 0.5  | 0    | 0  | 4    | 0    | 4  | 0 | 44.5 | -  | - | - | - | - | - | - | - | - | - | -  |

TABLE 5.4: Résultats des calculs de rapport de vraisemblance de la lettre e obtenus dans l'expérience 2 (scénarios 1 à 6).

|           |      | Lettre <b>f</b>          |      |     |     |     |                           |     |   |     |    |                           |     |      |     |     |     |   |     |   |   |    |    |     |      |     |   |     |     |     |   |     |     |    |
|-----------|------|--------------------------|------|-----|-----|-----|---------------------------|-----|---|-----|----|---------------------------|-----|------|-----|-----|-----|---|-----|---|---|----|----|-----|------|-----|---|-----|-----|-----|---|-----|-----|----|
|           |      | scénario 1 ( $n_s = 5$ ) |      |     |     |     | scénario 2 ( $n_s = 10$ ) |     |   |     |    | scénario 3 ( $n_s = 15$ ) |     |      |     |     |     |   |     |   |   |    |    |     |      |     |   |     |     |     |   |     |     |    |
| Scripteur |      | 1                        | 2    | 3   | 4   | 5   | 6                         | 7   | 8 | 9   | 10 | 11                        | 1   | 2    | 3   | 4   | 5   | 6 | 7   | 8 | 9 | 10 | 11 | 1   | 2    | 3   | 4 | 5   | 6   | 7   | 8 | 9   | 10  | 11 |
| 1         | 2    | -                        | -    | -   | -   | -   | -                         | -   | - | -   | -  | -                         | 1   | -    | -   | -   | -   | - | -   | - | - | -  | -  | 0   | -    | -   | - | -   | -   | -   | - | -   | -   | -  |
| 2         | 1    | 7                        | -    | -   | -   | -   | -                         | -   | - | -   | -  | -                         | 0   | 1.5  | -   | -   | -   | - | -   | - | - | -  | -  | 0   | 1    | -   | - | -   | -   | -   | - | -   | -   | -  |
| 3         | 46.5 | 9                        | 4.5  | -   | -   | -   | -                         | -   | - | -   | -  | -                         | 14  | 0    | 1.5 | -   | -   | - | -   | - | - | -  | -  | 0.5 | 0    | 0.5 | - | -   | -   | -   | - | -   | -   | -  |
| 4         | 0    | 0                        | 0    | 3   | -   | -   | -                         | -   | - | -   | -  | -                         | 0   | 0    | 0   | 0.5 | -   | - | -   | - | - | -  | -  | 0   | 0    | 0   | 0 | -   | -   | -   | - | -   | -   | -  |
| 5         | 0    | 0                        | 0    | 0   | 2.5 | -   | -                         | -   | - | -   | -  | -                         | 0   | 0    | 0   | 0   | 0.5 | - | -   | - | - | -  | -  | 0   | 0    | 0   | 0 | 0.5 | -   | -   | - | -   | -   | -  |
| 6         | 6    | 0                        | 7    | 0.5 | 0   | 4   | -                         | -   | - | -   | -  | -                         | 0   | 0    | 0   | 0   | 0.5 | - | -   | - | - | -  | -  | 0   | 0    | 0   | 0 | 0   | 0.5 | -   | - | -   | -   | -  |
| 7         | 8.5  | 0                        | 1.5  | 0   | 26  | 0.5 | 5                         | -   | - | -   | -  | -                         | 0   | 0    | 0   | 0   | 0.5 | - | -   | - | - | -  | -  | 0   | 0    | 0   | 0 | 0   | 0   | 0.5 | - | -   | -   | -  |
| 8         | 49   | 0.5                      | 8    | 0   | 51  | 8   | 50.5                      | 1.5 | - | -   | -  | -                         | 6.5 | 0    | 0   | 0   | 3.5 | 0 | 1.5 | - | - | -  | -  | 0   | 0    | 0   | 0 | 0   | 0   | 1   | 0 | 1.5 | 0   | -  |
| 9         | 0    | 0                        | 0    | 0.5 | 0   | 0   | 0                         | 4.5 | - | -   | -  | -                         | 0   | 0    | 0   | 0   | 0   | 0 | 0   | 4 | - | -  | -  | 0   | 0    | 0   | 0 | 0   | 0   | 0   | 0 | 0   | 0.5 | -  |
| 10        | 27   | 78.5                     | 12.5 | 0   | 6.5 | 0   | 29                        | 18  | 0 | 9   | -  | -                         | 3.5 | 84.5 | 0.5 | 0   | 0   | 0 | 5   | 1 | 0 | 1  | -  | 0   | 65.5 | 0   | 0 | 0   | 0   | 0   | 0 | 0   | 0   | 1  |
| 11        | 0    | 0                        | 0    | 0   | 0   | 0   | 0                         | 0   | 0 | 0.5 | 0  | -                         | 0   | 0    | 0   | 0   | 0   | 0 | 0   | 0 | 0 | 0  | -  | 0   | 0    | 0   | 0 | 0   | 0   | 0   | 0 | 0   | 0   | 0  |

|           |   | scénario 4 ( $n_s = 20$ ) |     |   |   |     |   |   |     |   |    |    | scénario 5 ( $n_s = 5$ ) |      |      |    |     |     |     |     |   |    |    | scénario 6 ( $n_s = 5$ ) |      |      |     |      |      |      |      |     |     |      |
|-----------|---|---------------------------|-----|---|---|-----|---|---|-----|---|----|----|--------------------------|------|------|----|-----|-----|-----|-----|---|----|----|--------------------------|------|------|-----|------|------|------|------|-----|-----|------|
| Scripteur |   | 1                         | 2   | 3 | 4 | 5   | 6 | 7 | 8   | 9 | 10 | 11 | 1                        | 2    | 3    | 4  | 5   | 6   | 7   | 8   | 9 | 10 | 11 | 1                        | 2    | 3    | 4   | 5    | 6    | 7    | 8    | 9   | 10  | 11   |
| 1         | 0 | -                         | -   | - | - | -   | - | - | -   | - | -  | -  | 1                        | -    | -    | -  | -   | -   | -   | -   | - | -  | -  | 0.5                      | -    | -    | -   | -    | -    | -    | -    | -   | -   | -    |
| 2         | 0 | 0                         | -   | - | - | -   | - | - | -   | - | -  | -  | 1                        | 4.5  | -    | -  | -   | -   | -   | -   | - | -  | -  | 82                       | 6    | -    | -   | -    | -    | -    | -    | -   | -   | -    |
| 3         | 0 | 0                         | 0.5 | - | - | -   | - | - | -   | - | -  | -  | 30.5                     | 68.5 | 6.5  | -  | -   | -   | -   | -   | - | -  | -  | 99                       | 68.5 | 1    | -   | -    | -    | -    | -    | -   | -   | -    |
| 4         | 0 | 0                         | 0   | 0 | - | -   | - | - | -   | - | -  | -  | 0                        | 0    | 0    | 11 | -   | -   | -   | -   | - | -  | -  | 8.5                      | 0    | 30   | 0.5 | -    | -    | -    | -    | -   | -   | -    |
| 5         | 0 | 0                         | 0   | 0 | 0 | -   | - | - | -   | - | -  | -  | 0                        | 0    | 0    | 0  | 0.5 | -   | -   | -   | - | -  | -  | 49                       | 0    | 8.5  | 0   | -    | -    | -    | -    | -   | -   | -    |
| 6         | 0 | 0                         | 0   | 0 | 0 | 0.5 | - | - | -   | - | -  | -  | 47                       | 0    | 3.5  | 1  | 0   | 8.5 | -   | -   | - | -  | -  | 57                       | 0.5  | 61.5 | 0   | 0.5  | 0.5  | -    | -    | -   | -   | -    |
| 7         | 0 | 0                         | 0   | 0 | 0 | 0   | 0 | - | -   | - | -  | -  | 14                       | 0    | 0.5  | 0  | 0   | 8.5 | 1.5 | -   | - | -  | -  | 68.5                     | 0    | 3    | 0   | 0    | 4.5  | 15   | -    | -   | -   | -    |
| 8         | 0 | 0                         | 0   | 0 | 0 | 0   | 0 | 0 | -   | - | -  | -  | 54                       | 0    | 0    | 0  | 0   | 7.3 | 8.5 | 6   | - | -  | -  | 44.5                     | 0    | 4.5  | 0   | 0    | 96   | 5    | 81   | -   | -   | -    |
| 9         | 0 | 0                         | 0   | 0 | 0 | 0   | 0 | 1 | -   | - | -  | -  | 0                        | 0    | 0    | 0  | 0   | 0   | 0   | 1.5 | - | -  | -  | 0                        | 0    | 0    | 0   | 14.5 | 0    | 59.5 | 0    | 0.5 | 5.5 | -    |
| 10        | 0 | 59.5                      | 0   | 0 | 0 | 0   | 0 | 0 | 0.5 | - | -  | -  | 36.5                     | 30.5 | 52.5 | 0  | 3.5 | 1   | 38  | 10  | 0 | 5  | -  | 84                       | 43.5 | 59.5 | 0   | 22.5 | 15.5 | 42   | 53.5 | 0   | 5   | -    |
| 11        | 0 | 0                         | 0   | 0 | 0 | 0   | 0 | 0 | 0   | 0 | 0  | -  | 0                        | 0    | 0    | 0  | 0   | 0   | 0   | 0   | 0 | 0  | -  | 0                        | 0    | 0    | 0   | 5.5  | 0    | 10   | 0    | 0   | 0   | 81.5 |

TABLE 5.5: Résultats des calculs de rapport de vraisemblance de la lettre **f** obtenus dans l'expérience 2 (scénarios 1 à 6).



|           |      | Lettre h                 |    |    |     |     |                           |   |     |     |    |                           |    |   |     |    |    |   |   |     |     |     |    |    |   |   |   |     |   |   |   |   |    |     |
|-----------|------|--------------------------|----|----|-----|-----|---------------------------|---|-----|-----|----|---------------------------|----|---|-----|----|----|---|---|-----|-----|-----|----|----|---|---|---|-----|---|---|---|---|----|-----|
|           |      | scénario 1 ( $n_s = 5$ ) |    |    |     |     | scénario 2 ( $n_s = 10$ ) |   |     |     |    | scénario 3 ( $n_s = 15$ ) |    |   |     |    |    |   |   |     |     |     |    |    |   |   |   |     |   |   |   |   |    |     |
| Scripteur |      | 1                        | 2  | 3  | 4   | 5   | 6                         | 7 | 8   | 9   | 10 | 11                        | 1  | 2 | 3   | 4  | 5  | 6 | 7 | 8   | 9   | 10  | 11 | 1  | 2 | 3 | 4 | 5   | 6 | 7 | 8 | 9 | 10 | 11  |
| 1         | 0.5  | -                        | -  | -  | -   | -   | -                         | - | -   | -   | -  | -                         | 0  | - | -   | -  | -  | - | - | -   | -   | -   | -  | 0  | - | - | - | -   | - | - | - | - | -  | -   |
| 2         | 0    | 0.5                      | -  | -  | -   | -   | -                         | - | -   | -   | -  | -                         | 0  | 0 | -   | -  | -  | - | - | -   | -   | -   | -  | 0  | 0 | - | - | -   | - | - | - | - | -  | -   |
| 3         | 0    | 0                        | 6  | -  | -   | -   | -                         | - | -   | -   | -  | -                         | 0  | 0 | 2   | -  | -  | - | - | -   | -   | -   | -  | 0  | 0 | 0 | - | -   | - | - | - | - | -  | -   |
| 4         | 0.5  | 0                        | 0  | 1  | -   | -   | -                         | - | -   | -   | -  | -                         | 0  | 0 | 0   | 0  | -  | - | - | -   | -   | -   | -  | 0  | 0 | 0 | 0 | -   | - | - | - | - | -  | -   |
| 5         | 0    | 0                        | 1  | 0  | 3.5 | -   | -                         | - | -   | -   | -  | -                         | 0  | 0 | 0   | 0  | 0  | - | - | -   | -   | -   | -  | 0  | 0 | 0 | 0 | 0.5 | - | - | - | - | -  | -   |
| 6         | 0.5  | 3.5                      | 0  | 58 | 0   | 1   | -                         | - | -   | -   | -  | -                         | 0  | 0 | 0   | 17 | 0  | 0 | - | -   | -   | -   | -  | 0  | 0 | 0 | 0 | 7.5 | 0 | 0 | - | - | -  | -   |
| 7         | 14.5 | 0                        | 0  | 0  | 26  | 0   | 2.5                       | - | -   | -   | -  | -                         | 2  | 0 | 0   | 0  | 13 | 0 | 0 | -   | -   | -   | -  | 0  | 0 | 0 | 0 | 4.5 | 0 | 0 | 0 | - | -  | -   |
| 8         | 31   | 0                        | 0  | 0  | 0   | 0   | 0                         | 1 | -   | -   | -  | -                         | 19 | 0 | 0   | 0  | 0  | 0 | 0 | -   | -   | -   | -  | 14 | 0 | 0 | 0 | 0   | 0 | 0 | 0 | 0 | -  | -   |
| 9         | 3.5  | 0                        | 0  | 0  | 0   | 0   | 0                         | 0 | 2.5 | -   | -  | -                         | 0  | 0 | 0   | 0  | 0  | 0 | 0 | 0.5 | -   | -   | -  | 0  | 0 | 0 | 0 | 0   | 0 | 0 | 0 | 0 | 0  | -   |
| 10        | 0    | 0                        | 20 | 6  | 0   | 5   | 0                         | 0 | 0   | 3.5 | -  | -                         | 0  | 0 | 6.5 | 0  | 0  | 0 | 0 | 0   | 0   | -   | -  | 0  | 0 | 2 | 0 | 0   | 0 | 0 | 0 | 0 | 0  | 0   |
| 11        | 0    | 0                        | 0  | 8  | 0   | 9.5 | 0                         | 0 | 0   | 30  | 1  | -                         | 0  | 0 | 0   | 0  | 0  | 0 | 0 | 0   | 4.5 | 0.5 | -  | 0  | 0 | 0 | 0 | 0   | 0 | 0 | 0 | 0 | 0  | 0.5 |

|           |   | scénario 4 ( $n_s = 20$ ) |     |   |   |   |   |   |   |   |    |    | scénario 5 ( $n_s = 5$ ) |   |   |     |     |    |     |   |     |    |    | scénario 6 ( $n_s = 5$ ) |   |      |      |      |      |     |      |     |    |      |
|-----------|---|---------------------------|-----|---|---|---|---|---|---|---|----|----|--------------------------|---|---|-----|-----|----|-----|---|-----|----|----|--------------------------|---|------|------|------|------|-----|------|-----|----|------|
| Scripteur |   | 1                         | 2   | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 1                        | 2 | 3 | 4   | 5   | 6  | 7   | 8 | 9   | 10 | 11 | 1                        | 2 | 3    | 4    | 5    | 6    | 7   | 8    | 9   | 10 | 11   |
| 1         | 0 | -                         | -   | - | - | - | - | - | - | - | -  | -  | 1                        | - | - | -   | -   | -  | -   | - | -   | -  | -  | 1.5                      | - | -    | -    | -    | -    | -   | -    | -   | -  | -    |
| 2         | 0 | 0                         | -   | - | - | - | - | - | - | - | -  | -  | 0                        | 0 | - | -   | -   | -  | -   | - | -   | -  | -  | 0                        | 0 | -    | -    | -    | -    | -   | -    | -   | -  | -    |
| 3         | 0 | 0                         | 1   | - | - | - | - | - | - | - | -  | -  | 0                        | 0 | 2 | -   | -   | -  | -   | - | -   | -  | -  | 0                        | 0 | 3.5  | -    | -    | -    | -   | -    | -   | -  | -    |
| 4         | 0 | 0                         | 0   | 0 | - | - | - | - | - | - | -  | -  | 0                        | 0 | 0 | 0.5 | -   | -  | -   | - | -   | -  | -  | 10                       | 0 | 67.5 | 58.5 | -    | -    | -   | -    | -   | -  | -    |
| 5         | 0 | 0                         | 0   | 0 | 0 | - | - | - | - | - | -  | -  | 1.5                      | 0 | 0 | 0.5 | 1.5 | -  | -   | - | -   | -  | -  | 28.5                     | 0 | 18.5 | 0    | 10.5 | -    | -   | -    | -   | -  | -    |
| 6         | 0 | 0                         | 0   | 1 | 0 | 0 | - | - | - | - | -  | -  | 0                        | 0 | 0 | 6.5 | 0   | 0  | -   | - | -   | -  | -  | 39.5                     | 0 | 7    | 99.5 | 2    | 8.5  | -   | -    | -   | -  | -    |
| 7         | 0 | 0                         | 0   | 0 | 0 | 0 | - | - | - | - | -  | -  | 0                        | 0 | 0 | 0   | 51  | 0  | 2   | - | -   | -  | -  | 11.5                     | 0 | 0    | 0    | 90   | 0    | 14  | -    | -   | -  | -    |
| 8         | 4 | 0                         | 0   | 0 | 0 | 0 | 0 | - | - | - | -  | -  | 0                        | 0 | 0 | 0   | 0   | 0  | 0.5 | - | -   | -  | -  | 85.5                     | 0 | 0    | 5.5  | 0    | 34.5 | 0   | 1.5  | -   | -  | -    |
| 9         | 0 | 0                         | 0   | 0 | 0 | 0 | 0 | 0 | - | - | -  | -  | 0                        | 0 | 0 | 0   | 0   | 0  | 0   | 4 | -   | -  | -  | 37.5                     | 0 | 0    | 0.5  | 0    | 2    | 0   | 85.5 | 3   | -  | -    |
| 10        | 0 | 0                         | 0.5 | 0 | 0 | 0 | 0 | 0 | 0 | - | -  | -  | 0                        | 0 | 0 | 0   | 0   | 0  | 0   | 0 | 0.5 | -  | -  | 19.5                     | 0 | 0    | 76.5 | 35.5 | 78   | 2.5 | 8    | 2.5 | 1  | -    |
| 11        | 0 | 0                         | 0   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0                        | 0 | 0 | 1   | 0   | 33 | 0   | 0 | 0   | 80 | 1  | 19.5                     | 0 | 0    | 78.5 | 0    | 48.5 | 0   | 2    | 0.5 | 75 | 63.5 |

TABLE 5.7: Résultats des calculs de rapport de vraisemblance de la lettre h obtenus dans l'expérience 2 (scénarios 1 à 6).

| Scripteur |     | Lettre j                 |      |    |     |      |                           |     |      |   |      |                           |      |    |    |     |      |      |      |      |      |      |
|-----------|-----|--------------------------|------|----|-----|------|---------------------------|-----|------|---|------|---------------------------|------|----|----|-----|------|------|------|------|------|------|
|           |     | scénario 1 ( $n_s = 5$ ) |      |    |     |      | scénario 2 ( $n_s = 10$ ) |     |      |   |      | scénario 3 ( $n_s = 15$ ) |      |    |    |     |      |      |      |      |      |      |
|           |     | 1                        | 2    | 3  | 4   | 5    | 1                         | 2   | 3    | 4 | 5    | 1                         | 2    | 3  | 4  | 5   |      |      |      |      |      |      |
| 1         | 13  | -                        | -    | -  | -   | -    | -                         | -   | -    | - | 6    | -                         | -    | -  | -  | 1   | -    | -    | -    | -    | -    | -    |
| 2         | 1.5 | 10.5                     | -    | -  | -   | -    | -                         | -   | -    | - | 0    | 8.5                       | -    | -  | -  | 0   | 2    | -    | -    | -    | -    | -    |
| 3         | 6   | 42.5                     | 7.5  | -  | -   | -    | -                         | -   | -    | - | 0    | 37                        | 2.5  | -  | -  | 0   | 35.5 | 2    | -    | -    | -    | -    |
| 4         | 4.5 | 83                       | 32   | 11 | -   | -    | -                         | -   | -    | - | 0    | 86.5                      | 20   | 7  | -  | 0   | 89.5 | 13.5 | 2.5  | -    | -    | -    |
| 5         | 24  | 1                        | 6.5  | 5  | 3   | -    | -                         | -   | -    | - | 7    | 0                         | 0    | 1  | 4  | -   | 0    | 0    | 0    | 1    | -    | -    |
| 6         | 16  | 42                       | 64.5 | 35 | 6.5 | 9.5  | -                         | -   | -    | - | 6    | 54                        | 69.5 | 19 | 1  | 3   | 0.5  | 61   | 80   | 13.5 | 0    | 6    |
| 7         | 9.5 | 0                        | 0.5  | 0  | 0   | 8.5  | -                         | -   | -    | - | 1    | 0                         | 0    | 0  | 0  | 2.5 | 0    | 0    | 0    | 0    | 0    | 0.5  |
| 8         | 4.5 | 0                        | 0.5  | 0  | 0   | 66.5 | 16                        | -   | -    | - | 0    | 0                         | 0    | 0  | 61 | 8   | 0    | 0    | 0    | 0    | 0    | 47.5 |
| 9         | 0   | 11                       | 0.5  | 16 | 0   | 9.5  | 0                         | 0   | 19.5 | - | 14.5 | 0                         | 2.5  | 0  | 1  | 0   | 0.5  | 0    | 0    | 0    | 0    | 6    |
| 10        | 0   | 61                       | 78.5 | 54 | 0   | 65.5 | 0                         | 0   | 3.5  | 0 | 43.5 | 82                        | 29.5 | 0  | 72 | 0   | 29   | 78.5 | 12.5 | 0    | 72.5 | 0    |
| 11        | 0   | 0.5                      | 11   | 0  | 0   | 1    | 5                         | 7.5 | 0    | 2 | 17   | 0                         | 0    | 0  | 0  | 1.5 | 0    | 0    | 0    | 0    | 0    | 4.5  |

| Scripteur |     | scénario 4 ( $n_s = 20$ ) |      |      |   |    |    |   |   |   |     |      | scénario 5 ( $n_s = 5$ ) |     |     |      |      |     |   |      |      |      |      | scénario 6 ( $n_s = 5$ ) |      |    |   |      |     |   |   |   |    |    |   |   |   |   |   |   |   |   |   |    |    |
|-----------|-----|---------------------------|------|------|---|----|----|---|---|---|-----|------|--------------------------|-----|-----|------|------|-----|---|------|------|------|------|--------------------------|------|----|---|------|-----|---|---|---|----|----|---|---|---|---|---|---|---|---|---|----|----|
|           |     | 1                         | 2    | 3    | 4 | 5  | 6  | 7 | 8 | 9 | 10  | 11   | 1                        | 2   | 3   | 4    | 5    | 6   | 7 | 8    | 9    | 10   | 11   | 1                        | 2    | 3  | 4 | 5    | 6   | 7 | 8 | 9 | 10 | 11 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|           |     | 1                         | 1.5  | -    | - | -  | -  | - | - | - | -   | -    | -                        | 6.5 | -   | -    | -    | -   | - | -    | -    | -    | -    | 4                        | -    | -  | - | -    | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -  | -  |
| 2         | 0   | 3.5                       | -    | -    | - | -  | -  | - | - | - | -   | 0    | 7.5                      | -   | -   | -    | -    | -   | - | -    | -    | 2.5  | 3.5  | -                        | -    | -  | - | -    | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - |   |    |    |
| 3         | 0   | 23                        | 0.5  | -    | - | -  | -  | - | - | - | 0   | 45.5 | 3.5                      | -   | -   | -    | -    | -   | - | -    | 2    | 53.5 | 0.5  | -                        | -    | -  | - | -    | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - |   |    |    |
| 4         | 0   | 93                        | 4    | 0    | - | -  | -  | - | - | - | 0   | 77.5 | 50.5                     | 3.5 | -   | -    | -    | -   | - | -    | 32   | 68   | 77   | 3                        | -    | -  | - | -    | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - |    |    |
| 5         | 0   | 0                         | 0    | 0    | 0 | -  | -  | - | - | - | 7   | 0    | 0                        | 2   | 3   | -    | -    | -   | - | 71   | 3.5  | 9    | 12.5 | 42                       | -    | -  | - | -    | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - |    |    |
| 6         | 0.5 | 65                        | 87.5 | 12.5 | 0 | 0  | -  | - | - | - | 0   | 73.5 | 2                        | 0   | 4.5 | -    | -    | -   | 0 | 20   | 96.5 | 32.5 | 0    | 5                        | -    | -  | - | -    | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - |    |    |
| 7         | 0   | 0                         | 0    | 0    | 0 | 2  | -  | - | - | - | 0.5 | 0    | 0                        | 0   | 12  | -    | -    | -   | 6 | 0    | 4    | 4    | 2.5  | 7                        | 21.5 | -  | - | -    | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - |    |    |
| 8         | 0   | 0                         | 0    | 0    | 0 | 0  | 29 | 4 | - | - | 0   | 0    | 0                        | 0   | 0   | 40.5 | 9.5  | -   | 0 | 1    | 0    | 1.5  | 0    | 2                        | 67.5 | 56 | - | -    | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - |    |    |
| 9         | 0   | 0                         | 0    | 0    | 0 | 0  | 0  | 0 | 2 | - | 0.5 | 53   | 22                       | 73  | 0   | 4    | 0    | 7.5 | - | 24.5 | 94   | 47.5 | 97.5 | 38.5                     | 31   | 0  | 0 | 51.5 | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - |    |    |
| 10        | 0   | 22                        | 74   | 3.5  | 0 | 72 | 0  | 0 | 0 | - | 0   | 80.5 | 1                        | 0   | 69  | 0    | 0    | 8   | - | 13   | 86   | 21   | 0    | 76.5                     | 0.5  | 0  | 0 | 45   | -   | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - |    |    |
| 11        | 0   | 0                         | 0    | 0    | 0 | 0  | 0  | 0 | 0 | 3 | 0   | 0    | 0                        | 0   | 1.5 | 12   | 21.5 | 0   | 6 | 2    | 9.5  | 0.5  | 0    | 4.5                      | 50   | 58 | 0 | 2    | 3.5 | - | - | - | -  | -  | - | - | - | - | - | - | - | - | - | -  | -  |

TABLE 5.8: Résultats des calculs de rapport de vraisemblance de la lettre j obtenus dans l'expérience 2 (scénarios 1 à 6).









| Scripteur | Lettre q                 |     |     |     |     |                           |     |     |     |     |                           |     |      |     |     |   |     |     |    |   |
|-----------|--------------------------|-----|-----|-----|-----|---------------------------|-----|-----|-----|-----|---------------------------|-----|------|-----|-----|---|-----|-----|----|---|
|           | scénario 1 ( $n_s = 5$ ) |     |     |     |     | scénario 2 ( $n_s = 10$ ) |     |     |     |     | scénario 3 ( $n_s = 15$ ) |     |      |     |     |   |     |     |    |   |
|           | 1                        | 2   | 3   | 4   | 5   | 1                         | 2   | 3   | 4   | 5   | 1                         | 2   | 3    | 4   | 5   | 1 | 2   | 3   | 4  | 5 |
| 1         | 11                       | -   | -   | -   | -   | -                         | -   | -   | -   | -   | 4.5                       | -   | -    | -   | -   | - | -   | -   | -  | - |
| 2         | 1.5                      | 3.5 | -   | -   | -   | -                         | -   | -   | -   | -   | 0                         | 0.5 | -    | -   | -   | - | -   | -   | -  | - |
| 3         | 0                        | 0   | 7   | -   | -   | -                         | -   | -   | -   | -   | 0                         | 0   | 2.5  | -   | -   | - | -   | -   | -  | - |
| 4         | 43.5                     | 0   | 30  | 11  | -   | -                         | -   | -   | -   | -   | 35                        | 0   | 10.5 | 2   | -   | - | -   | -   | -  | - |
| 5         | 44                       | 0   | 0   | 2.5 | 7   | -                         | -   | -   | -   | -   | 29                        | 0   | 0    | 0.5 | -   | - | -   | -   | -  | - |
| 6         | 26.5                     | 0   | 0   | 8   | 0   | 8                         | -   | -   | -   | -   | 2.5                       | 0   | 0    | 0   | 1.5 | - | -   | -   | -  | - |
| 7         | 5.5                      | 0   | 0   | 62  | 0.5 | 3                         | 5.5 | -   | -   | -   | 0                         | 0   | 0    | 0   | 4.5 | - | -   | -   | -  | - |
| 8         | 9.5                      | 0   | 0.5 | 70  | 0.5 | 18.5                      | 20  | 1.5 | -   | -   | 1.5                       | 0   | 0    | 62  | 0   | 1 | 1.5 | 2.5 | -  | - |
| 9         | 12.5                     | 0   | 0   | 0   | 2.5 | 0                         | 0   | 0   | 4.5 | -   | 0                         | 0   | 0    | 0   | 0   | 0 | 0   | 0   | 1  | - |
| 10        | 53.5                     | 0.5 | 0   | 1   | 12  | 0                         | 0   | 0   | 71  | 2.5 | -                         | 47  | 0    | 0   | 0.5 | 0 | 0   | 0   | 56 | 2 |
| 11        | 0                        | 1.5 | 0   | 0   | 0   | 0                         | 0   | 0   | 0   | 0   | 5.5                       | 0   | 0    | 0   | 0   | 0 | 0   | 0   | 1  | 0 |

| Scripteur | scénario 4 ( $n_s = 20$ ) |     |   |    |     |     |   |   |    |     |     | scénario 5 ( $n_s = 5$ ) |     |      |     |    |     |   |   |      |     |      | scénario 6 ( $n_s = 5$ ) |      |      |      |      |     |     |     |    |     |      |   |   |   |   |   |
|-----------|---------------------------|-----|---|----|-----|-----|---|---|----|-----|-----|--------------------------|-----|------|-----|----|-----|---|---|------|-----|------|--------------------------|------|------|------|------|-----|-----|-----|----|-----|------|---|---|---|---|---|
|           | 1                         | 2   | 3 | 4  | 5   | 6   | 7 | 8 | 9  | 10  | 11  | 1                        | 2   | 3    | 4   | 5  | 6   | 7 | 8 | 9    | 10  | 11   | 1                        | 2    | 3    | 4    | 5    | 6   | 7   | 8   | 9  | 10  | 11   |   |   |   |   |   |
|           | 1                         | 0.5 | - | -  | -   | -   | - | - | -  | -   | -   | 6                        | -   | -    | -   | -  | -   | - | - | -    | -   | -    | 0.5                      | -    | -    | -    | -    | -   | -   | -   | -  | -   | -    | - | - | - | - | - |
| 2         | 0                         | 0.5 | - | -  | -   | -   | - | - | -  | -   | 0   | 4.5                      | -   | -    | -   | -  | -   | - | - | -    | -   | 75.5 | 26.5                     | -    | -    | -    | -    | -   | -   | -   | -  | -   | -    | - | - | - | - | - |
| 3         | 0                         | 0   | 0 | -  | -   | -   | - | - | -  | -   | 0   | 0                        | 5   | -    | -   | -  | -   | - | - | -    | -   | 0.5  | 0                        | 27   | -    | -    | -    | -   | -   | -   | -  | -   | -    | - | - | - | - | - |
| 4         | 22                        | 0   | 0 | 1  | -   | -   | - | - | -  | -   | 30  | 0                        | 1   | 5    | -   | -  | -   | - | - | -    | -   | 55   | 0                        | 47.5 | 1    | -    | -    | -   | -   | -   | -  | -   | -    | - | - | - | - | - |
| 5         | 1.5                       | 0   | 0 | 0  | 0.5 | -   | - | - | -  | -   | 1   | 0                        | 0   | 6    | 1.5 | -  | -   | - | - | -    | -   | 93   | 0                        | 0    | 34.5 | 56.5 | -    | -   | -   | -   | -  | -   | -    | - | - | - | - | - |
| 6         | 0                         | 0   | 0 | 0  | 0   | 0.5 | - | - | -  | -   | 81  | 0                        | 0   | 36   | 7.5 | 7  | -   | - | - | -    | -   | 98.5 | 0                        | 36.5 | 97.5 | 42.5 | 6    | -   | -   | -   | -  | -   | -    | - | - | - | - | - |
| 7         | 0                         | 0   | 0 | 15 | 0   | 0   | 0 | 0 | -  | -   | 0   | 0                        | 0   | 64.5 | 0   | 0  | 2   | - | - | -    | -   | 1    | 0                        | 3.5  | 92.5 | 0    | 7.5  | 88  | -   | -   | -  | -   | -    | - | - | - | - |   |
| 8         | 0                         | 0   | 0 | 41 | 0   | 0   | 0 | 0 | -  | -   | 0   | 0                        | 0   | 54   | 0   | 9  | 7.5 | 3 | - | -    | -   | 8    | 0                        | 28   | 9.5  | 0    | 24   | 8.5 | 3   | -   | -  | -   | -    | - | - | - | - |   |
| 9         | 0                         | 0   | 0 | 0  | 0   | 0   | 0 | 0 | 0  | 0   | 7.5 | 16                       | 0   | 0    | 2   | 0  | 0   | 0 | 3 | -    | -   | 95.5 | 26                       | 0    | 42   | 92.5 | 46.5 | 1.5 | 2.5 | 11  | -  | -   | -    | - | - | - | - | - |
| 10        | 27                        | 0   | 0 | 0  | 0   | 0   | 0 | 0 | 17 | 0.5 | -   | 43                       | 4.5 | 0    | 6   | 33 | 6.5 | 0 | 0 | 71.5 | 6.5 | -    | 99                       | 24.5 | 1    | 69   | 84   | 66  | 15  | 46  | 44 | 8   | -    | - | - | - | - |   |
| 11        | 0                         | 0   | 0 | 0  | 0   | 0   | 0 | 0 | 0  | 0   | 1   | 0                        | 0   | 0    | 0   | 0  | 0   | 0 | 0 | 0    | 0   | 13.5 | 50.5                     | 3    | 0    | 0.5  | 0.5  | 36  | 0   | 0.5 | 0  | 1.5 | 65.5 | - | - | - | - | - |

TABLE 5.13: Résultats des calculs de rapport de vraisemblance de la lettre q obtenus dans l'expérience 2 (scénarios 1 à 6).

# Représentation par Analyse en Coordonnées Principales

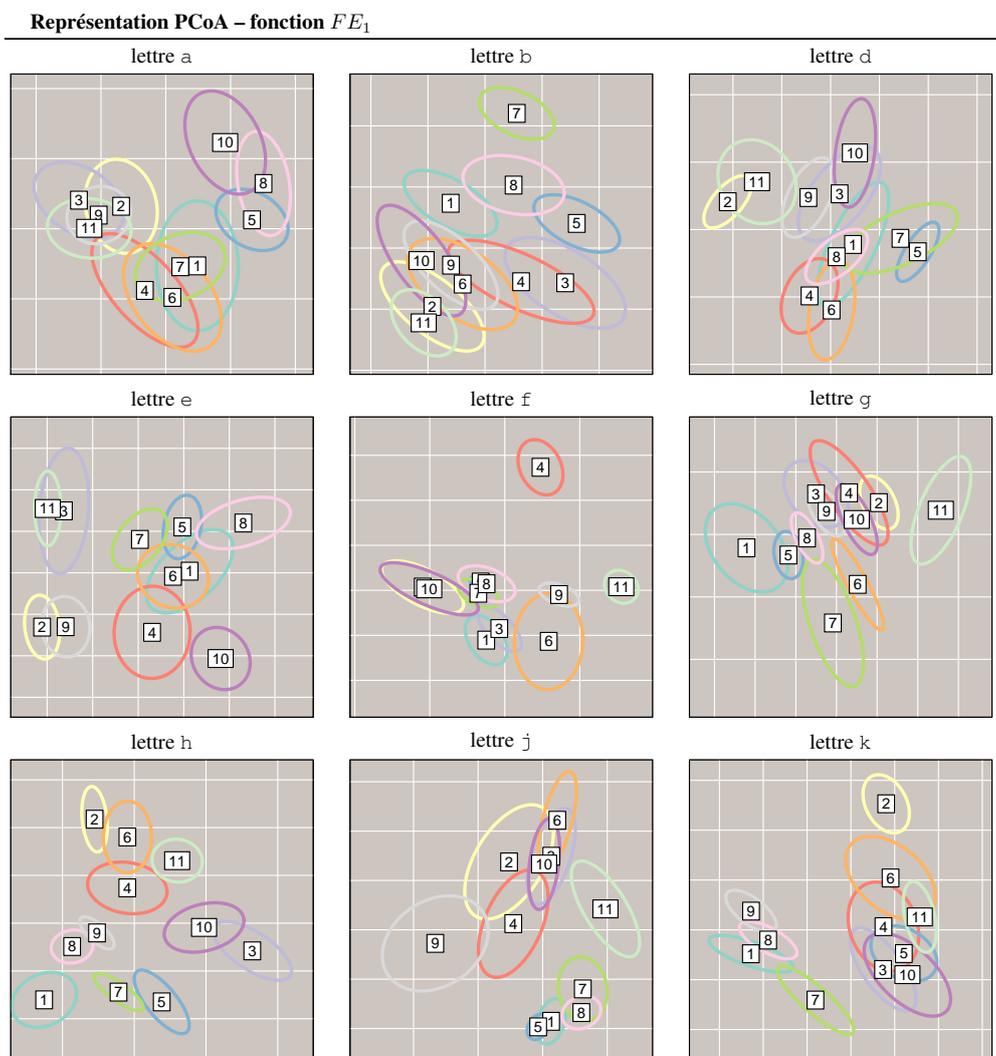


FIGURE 5.149: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* de l'expérience 2, groupés par scripteurs.

Représentation PCoA – fonction  $FE_1$

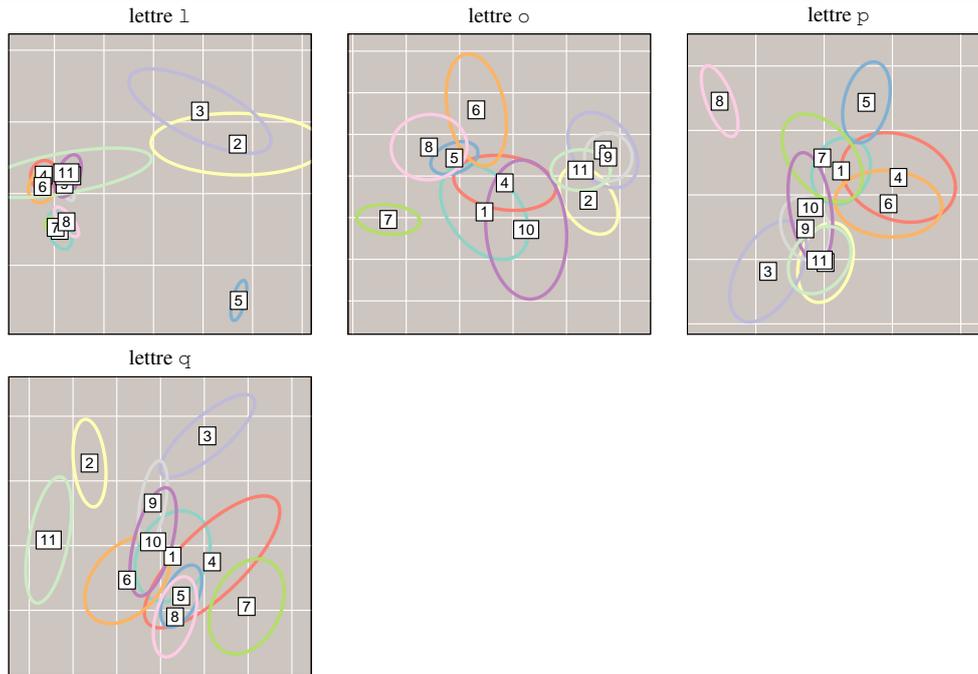


FIGURE 5.150: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* de l'expérience 2, groupés par scripteurs.

Représentation PCoA – fonction  $FE_2$

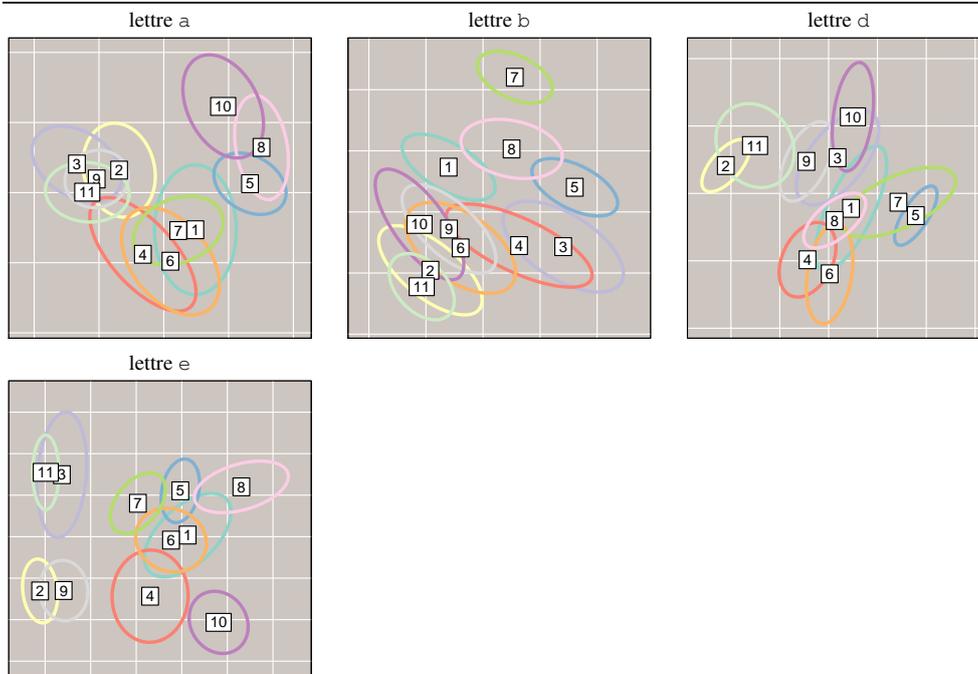


FIGURE 5.151: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* de l'expérience 2, groupés par scripteurs.



FIGURE 5.152: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* de l'expérience 2, groupés par scripteurs.

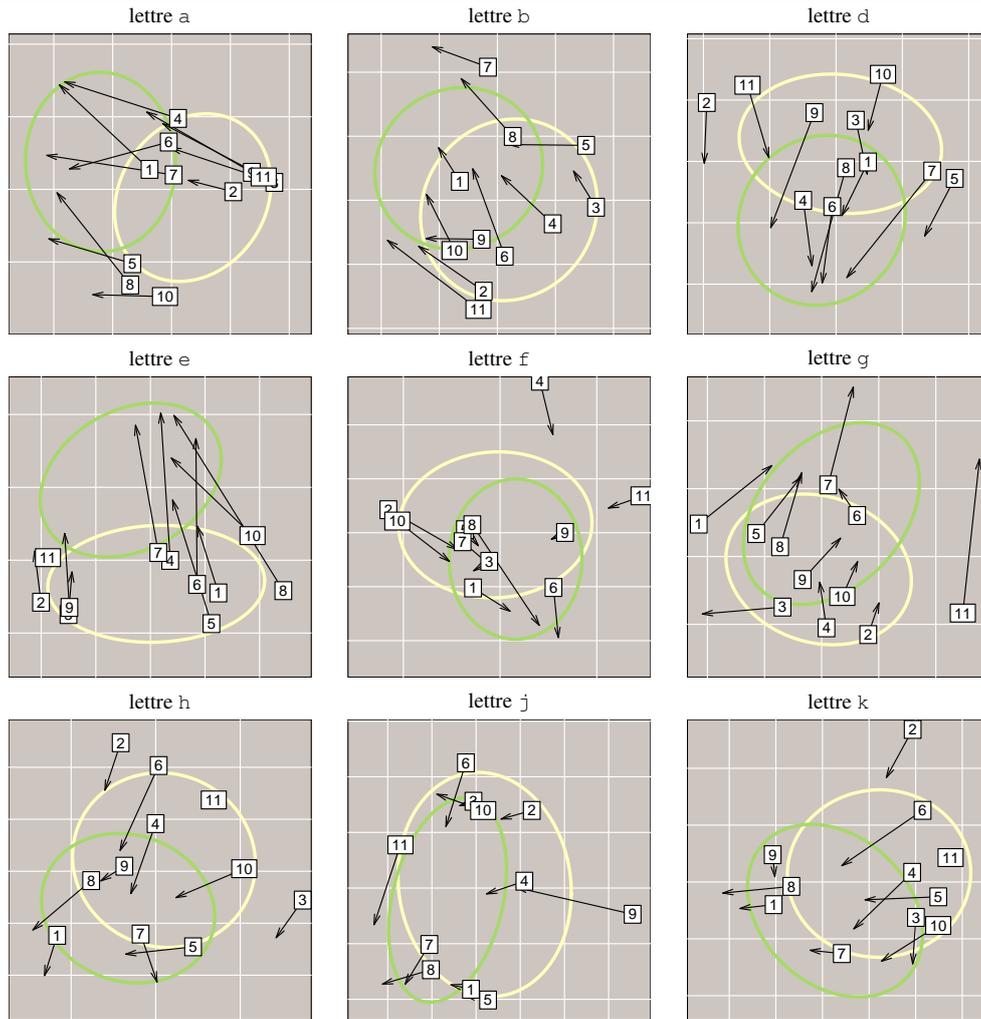


FIGURE 5.153: Représentations par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés par type de caractère (*normaux* et *grands*).

Représentation PCoA – caractères normaux/grands – fonction  $FE_1$

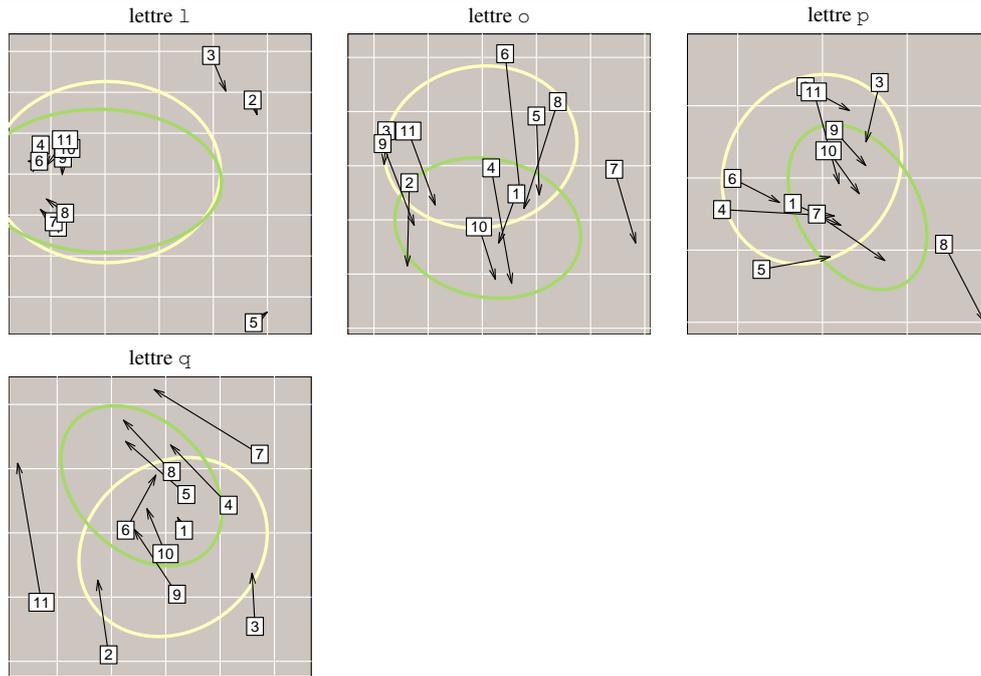


FIGURE 5.154: Représentations par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés par type de caractère (*normaux* et *grands*).

Représentation PCoA – caractères normaux/grands – fonction  $FE_2$

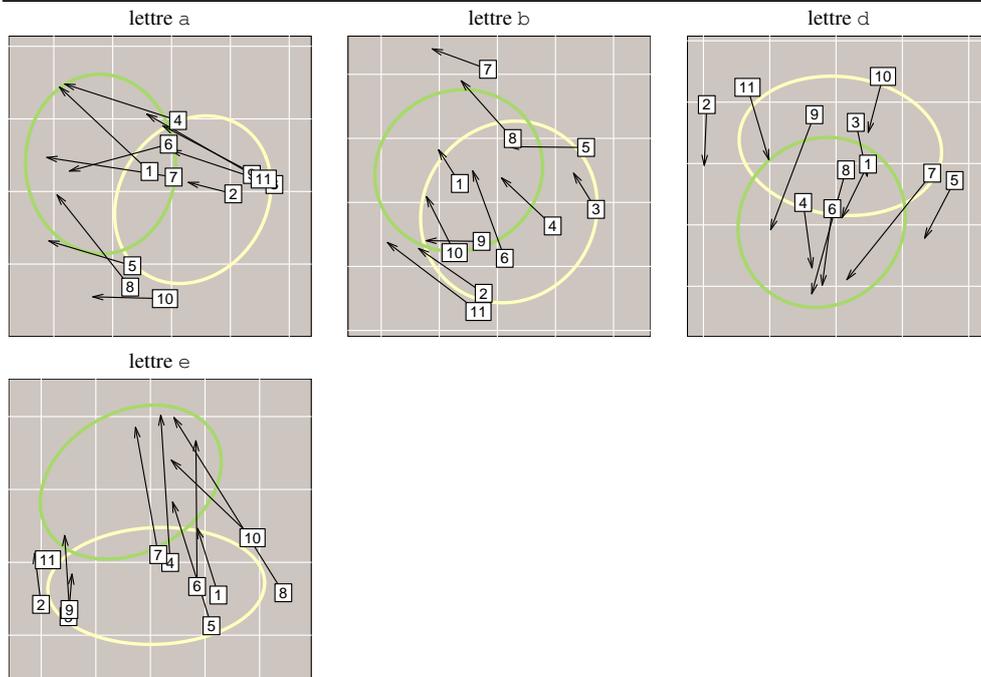


FIGURE 5.155: Représentations par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés par type de caractère (*normaux* et *grands*).



FIGURE 5.156: Représentations par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés par type de caractère (*normaux* et *grands*).

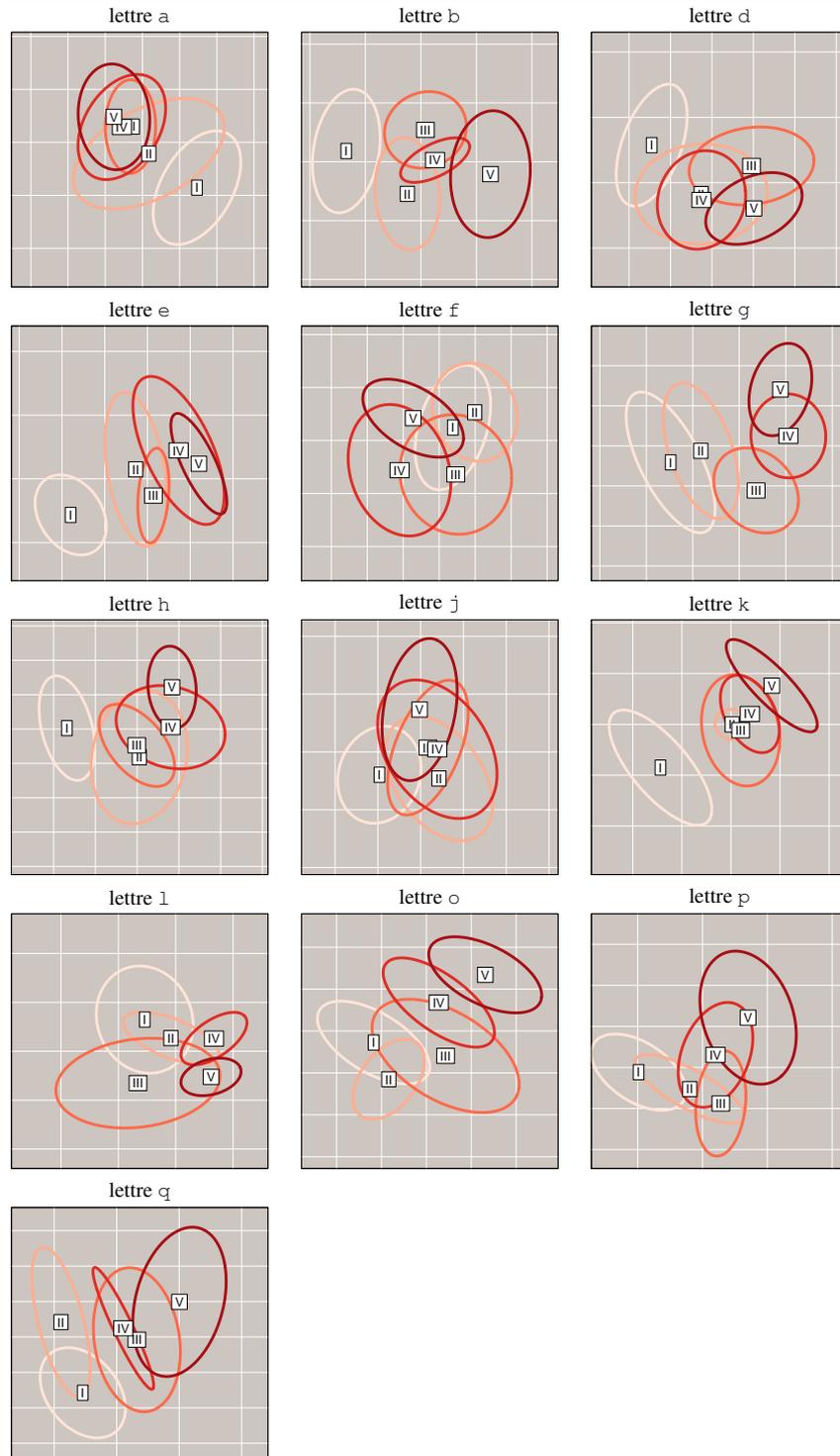


FIGURE 5.157: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

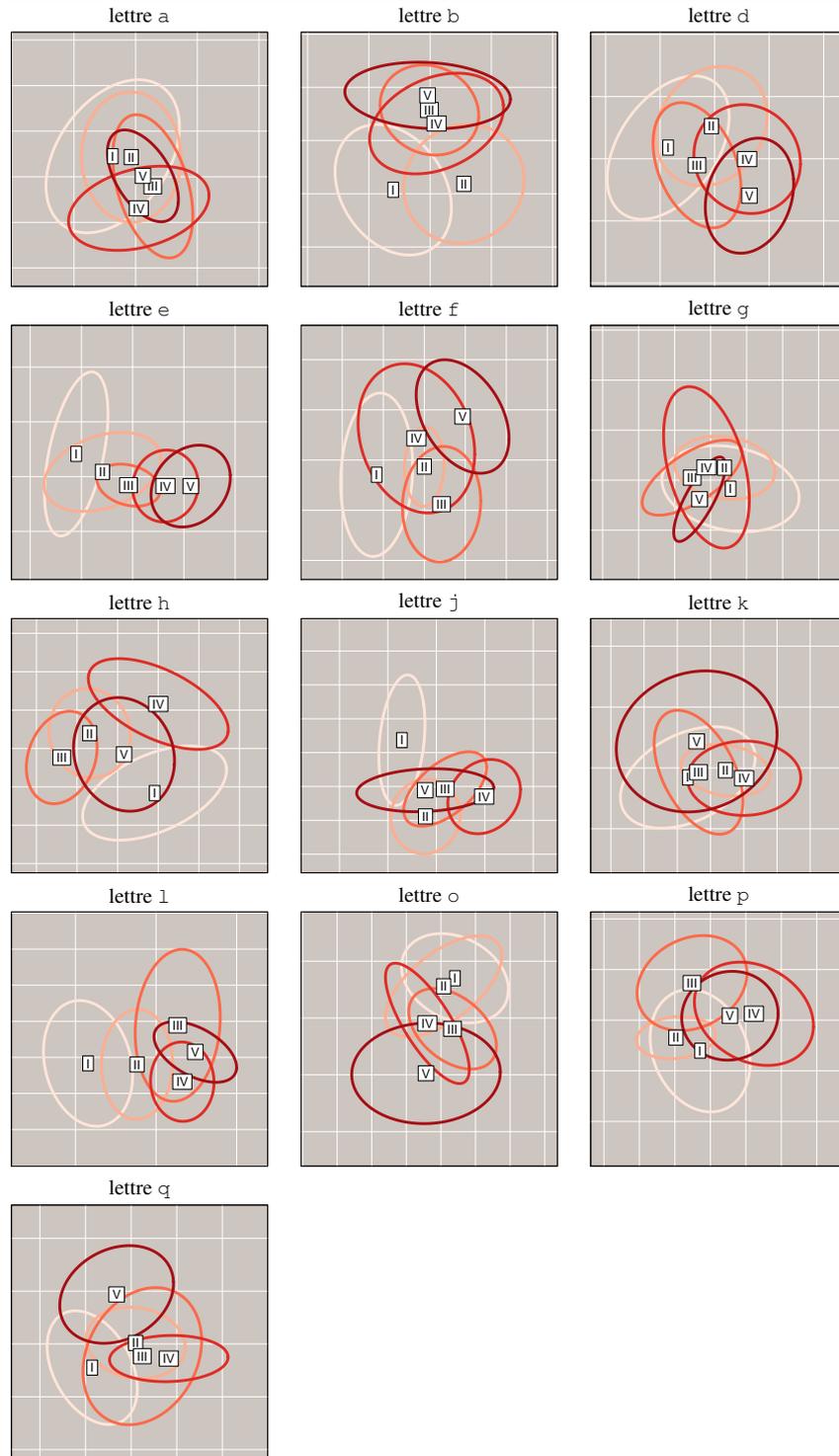


FIGURE 5.158: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).



FIGURE 5.159: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

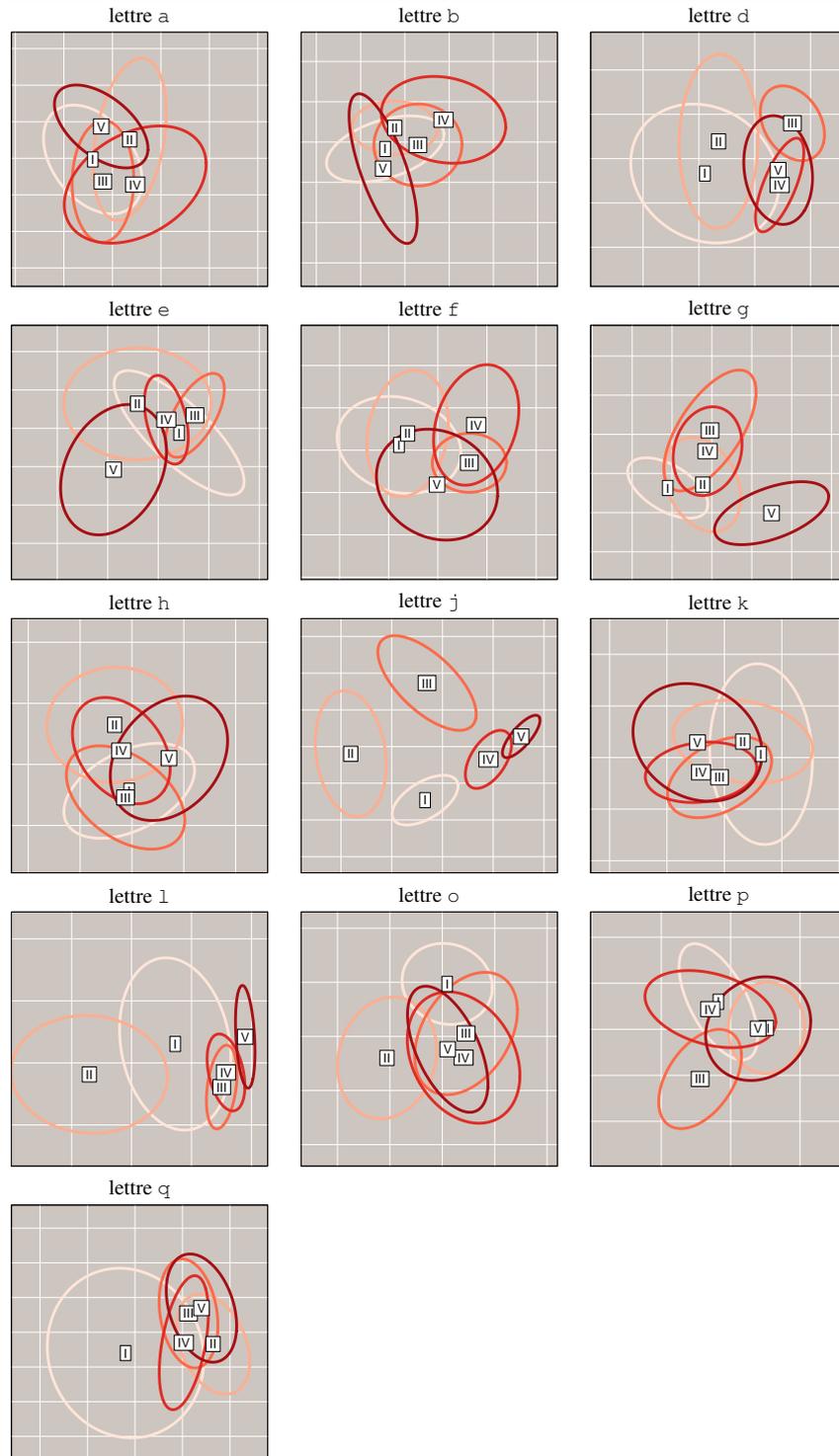


FIGURE 5.160: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

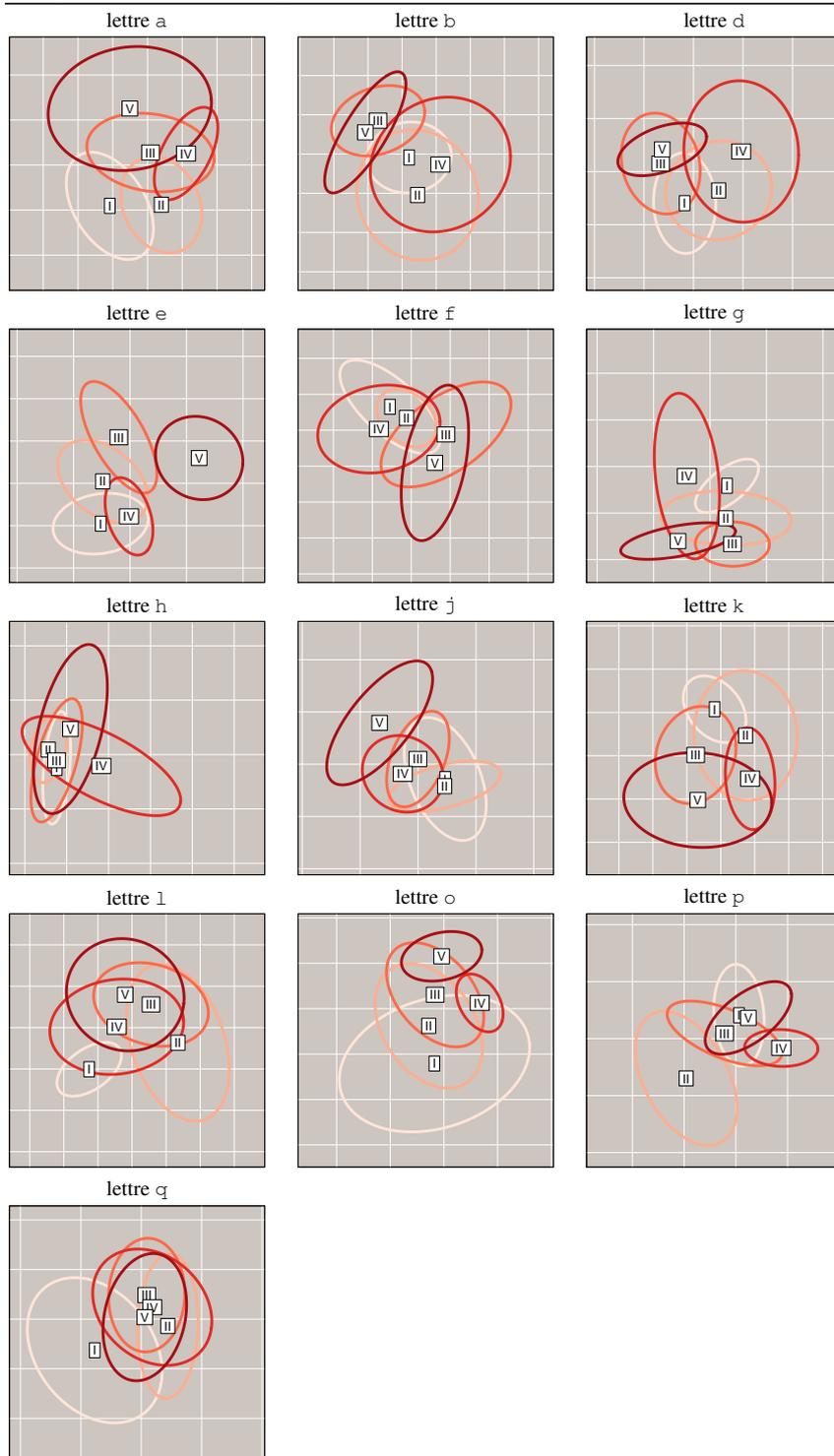


FIGURE 5.161: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

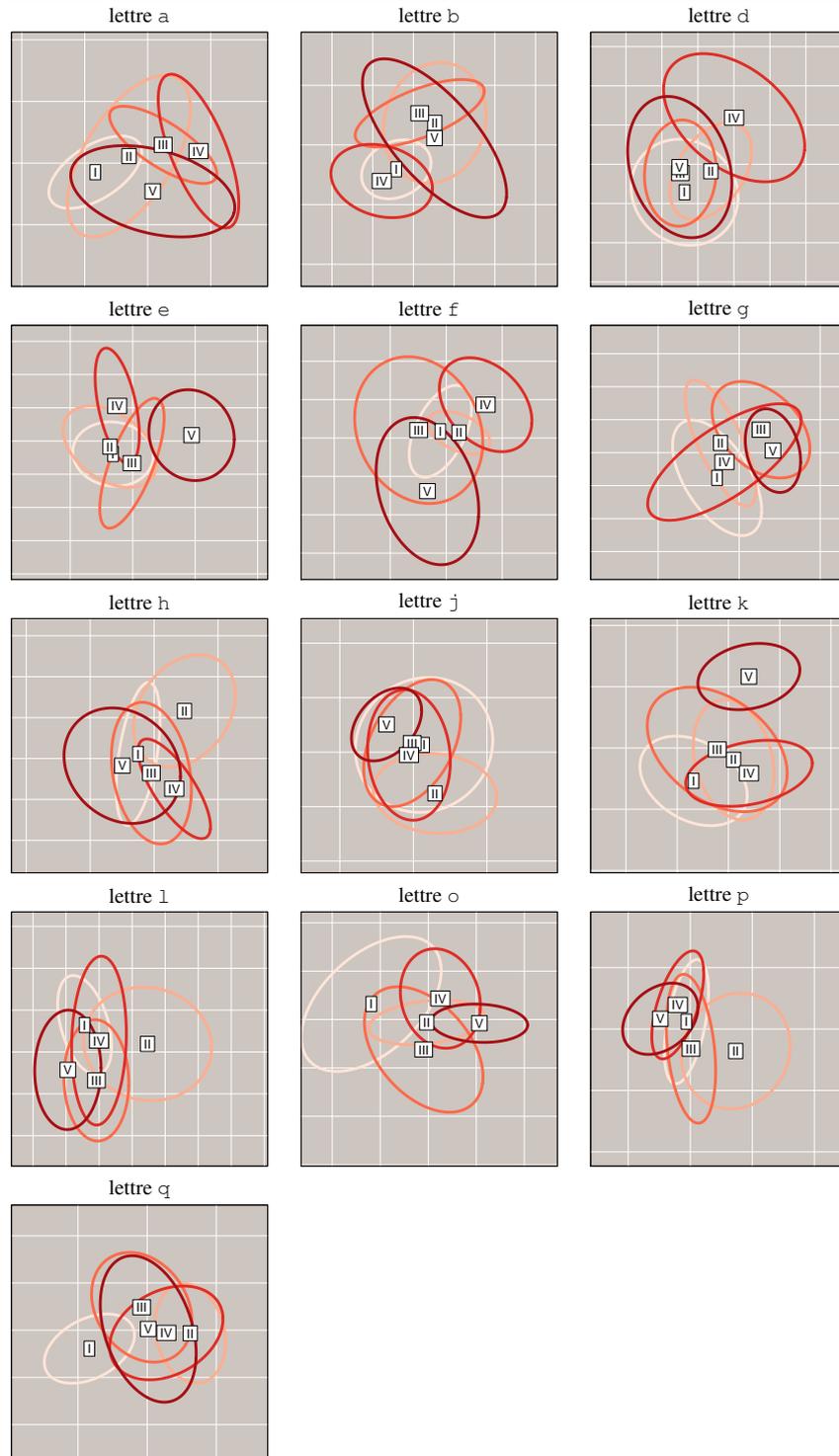


FIGURE 5.162: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

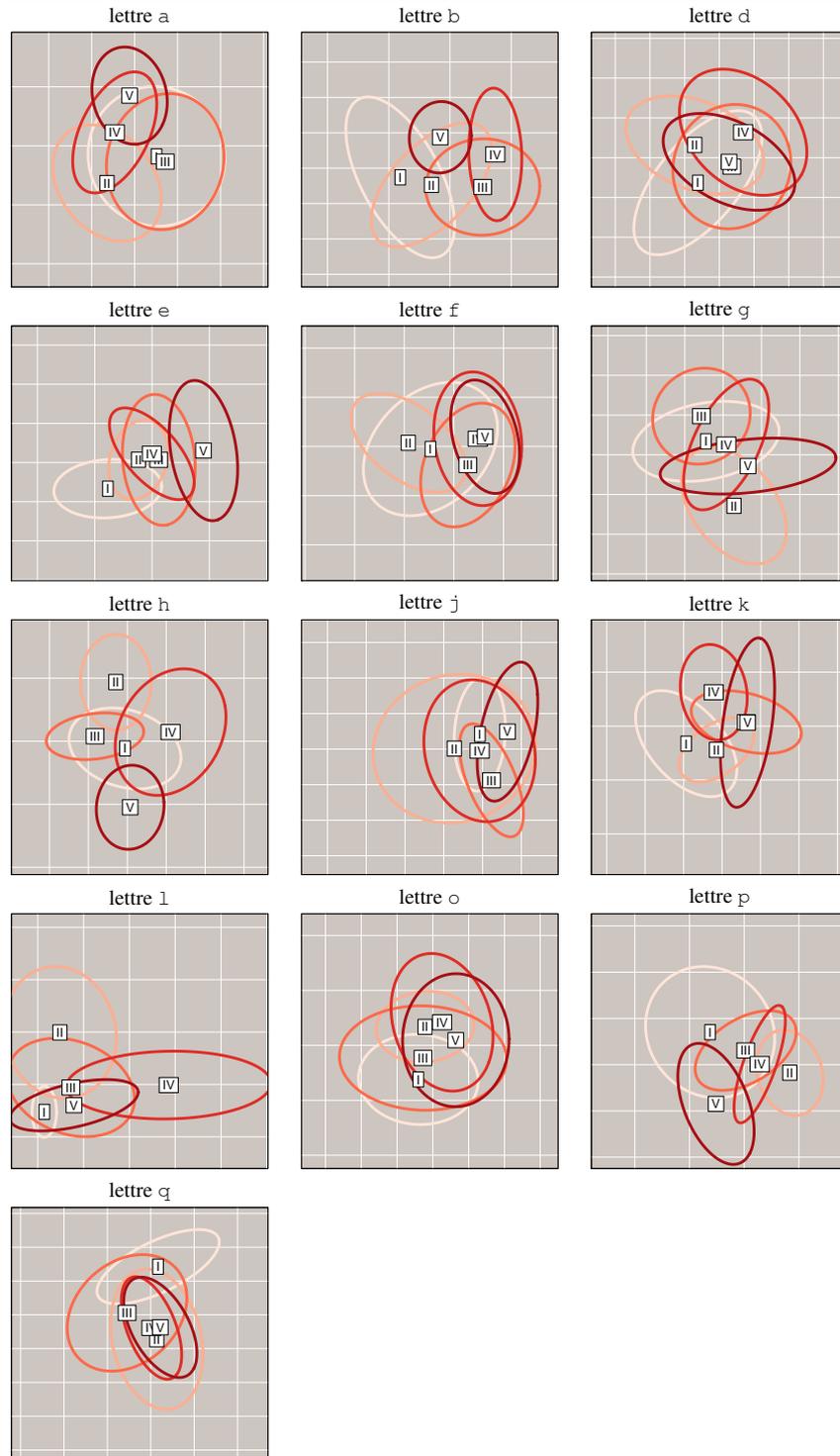


FIGURE 5.163: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

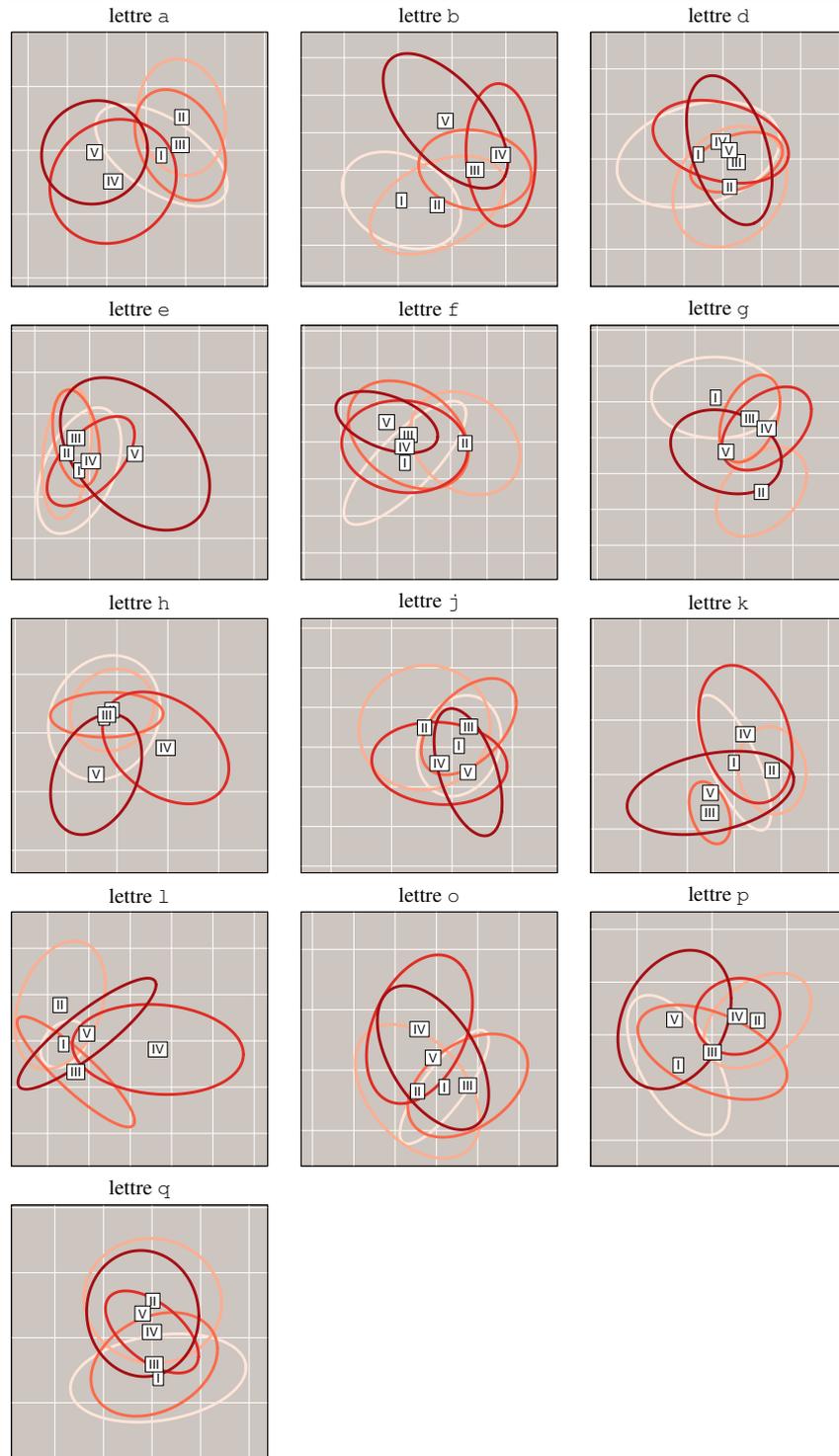


FIGURE 5.164: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

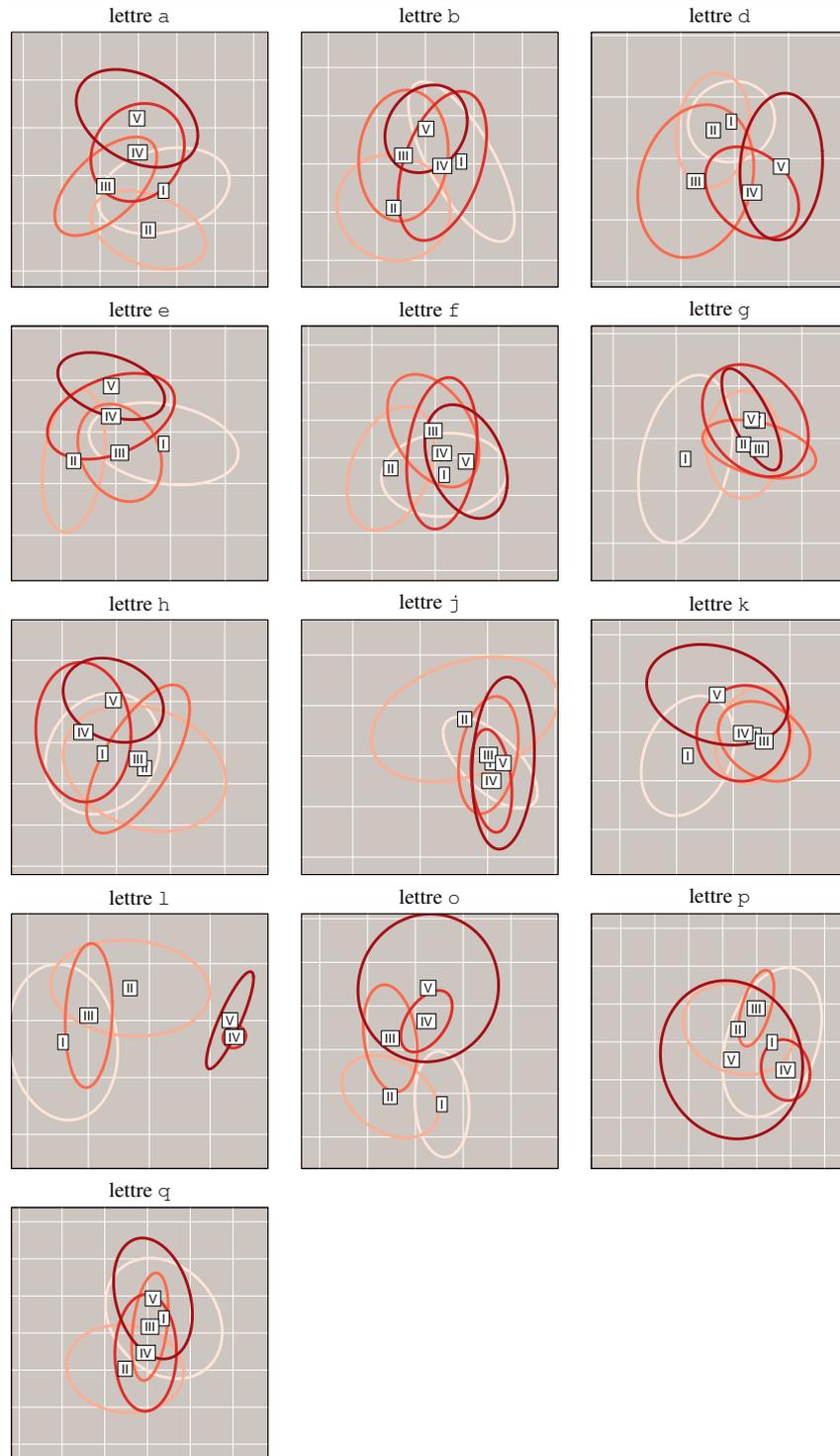


FIGURE 5.165: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

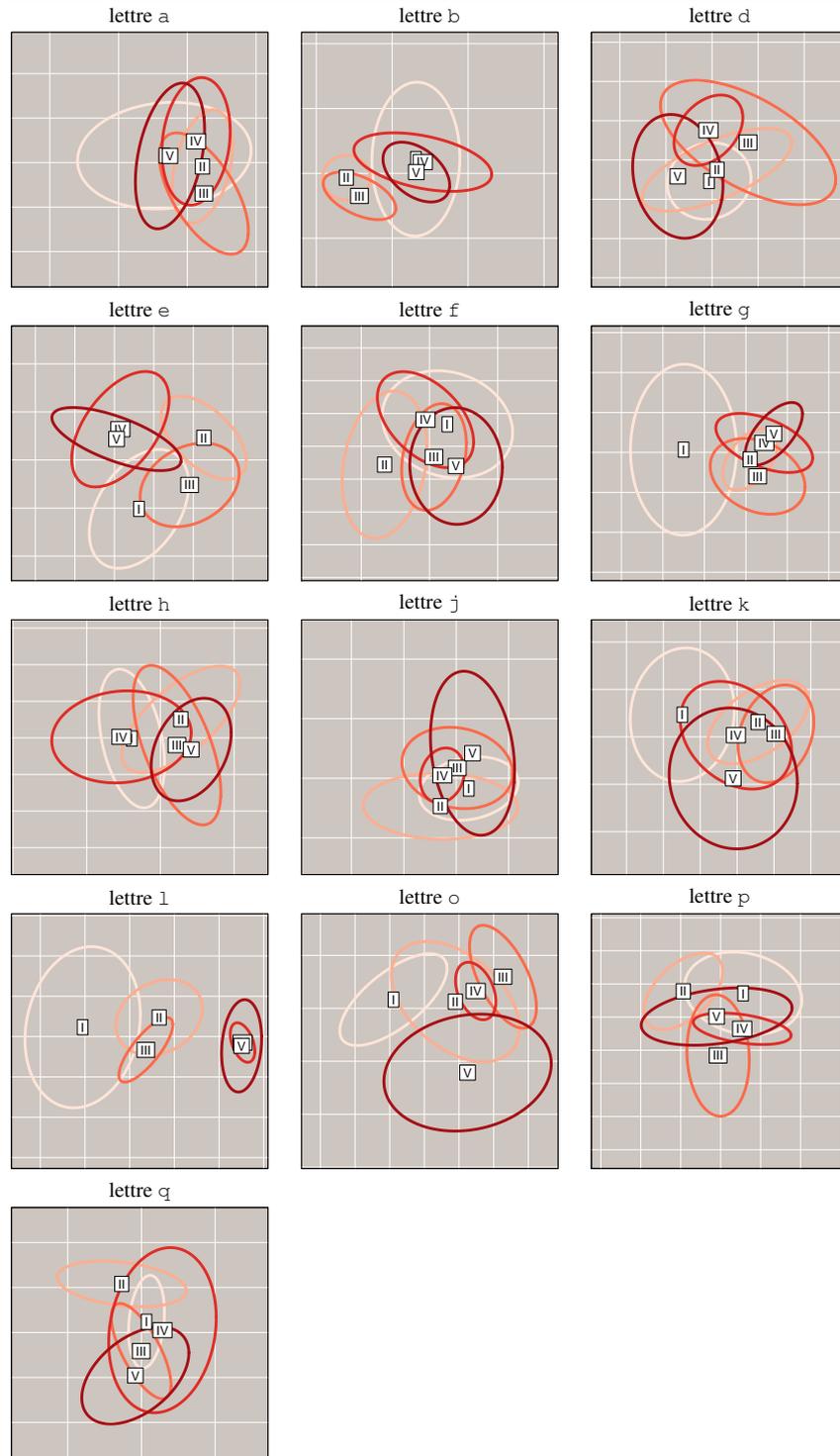


FIGURE 5.166: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

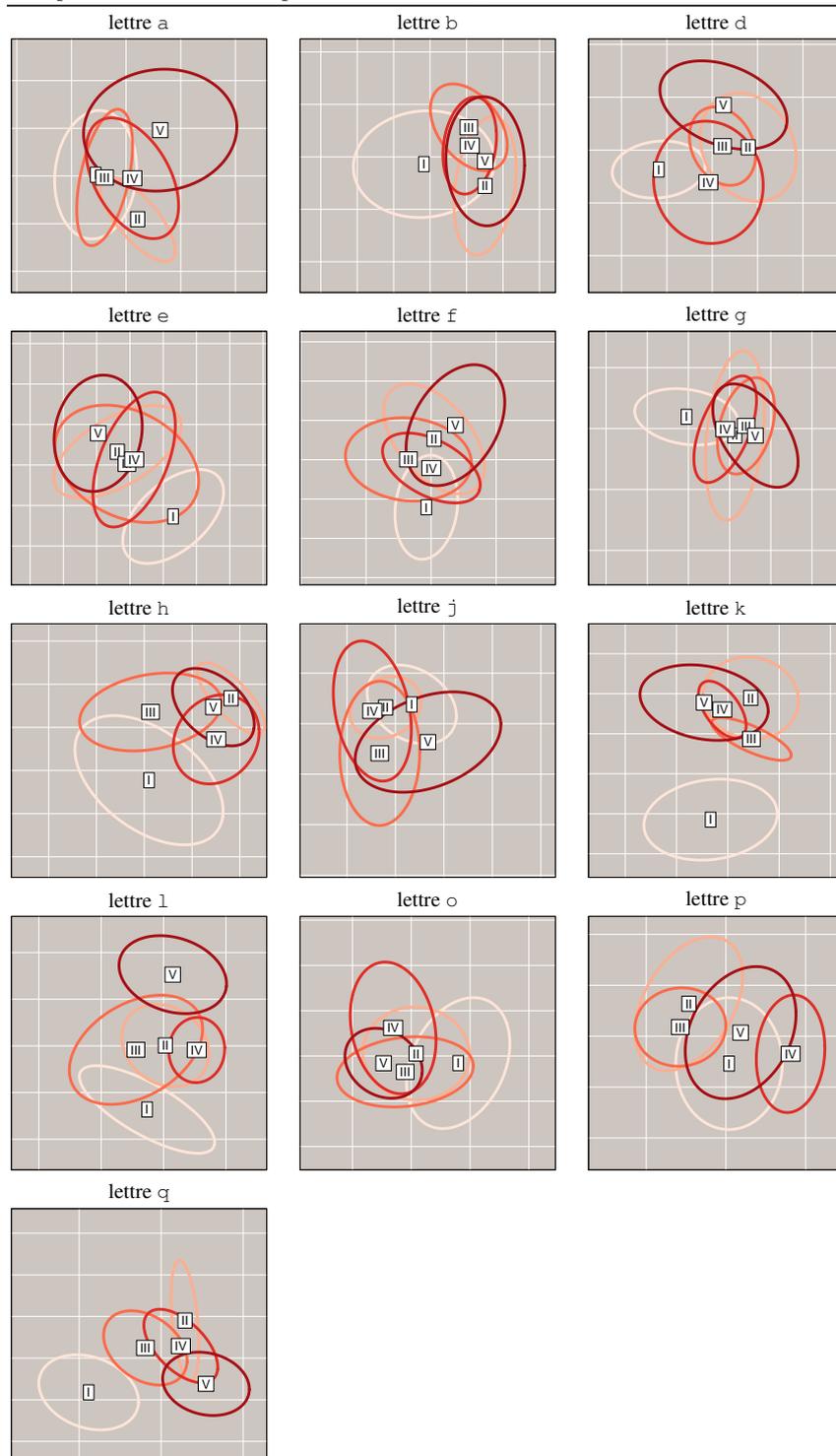


FIGURE 5.167: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

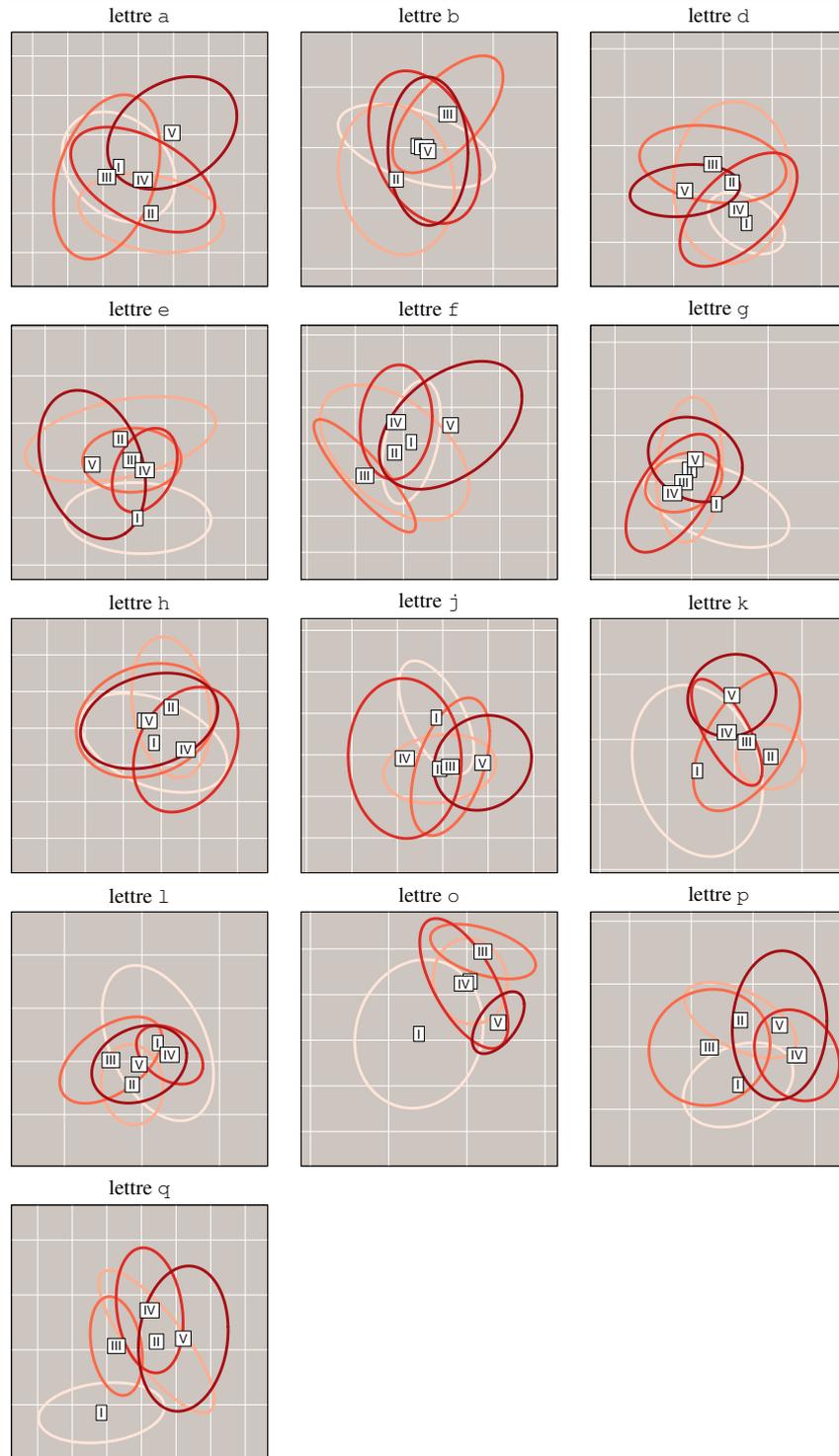


FIGURE 5.168: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

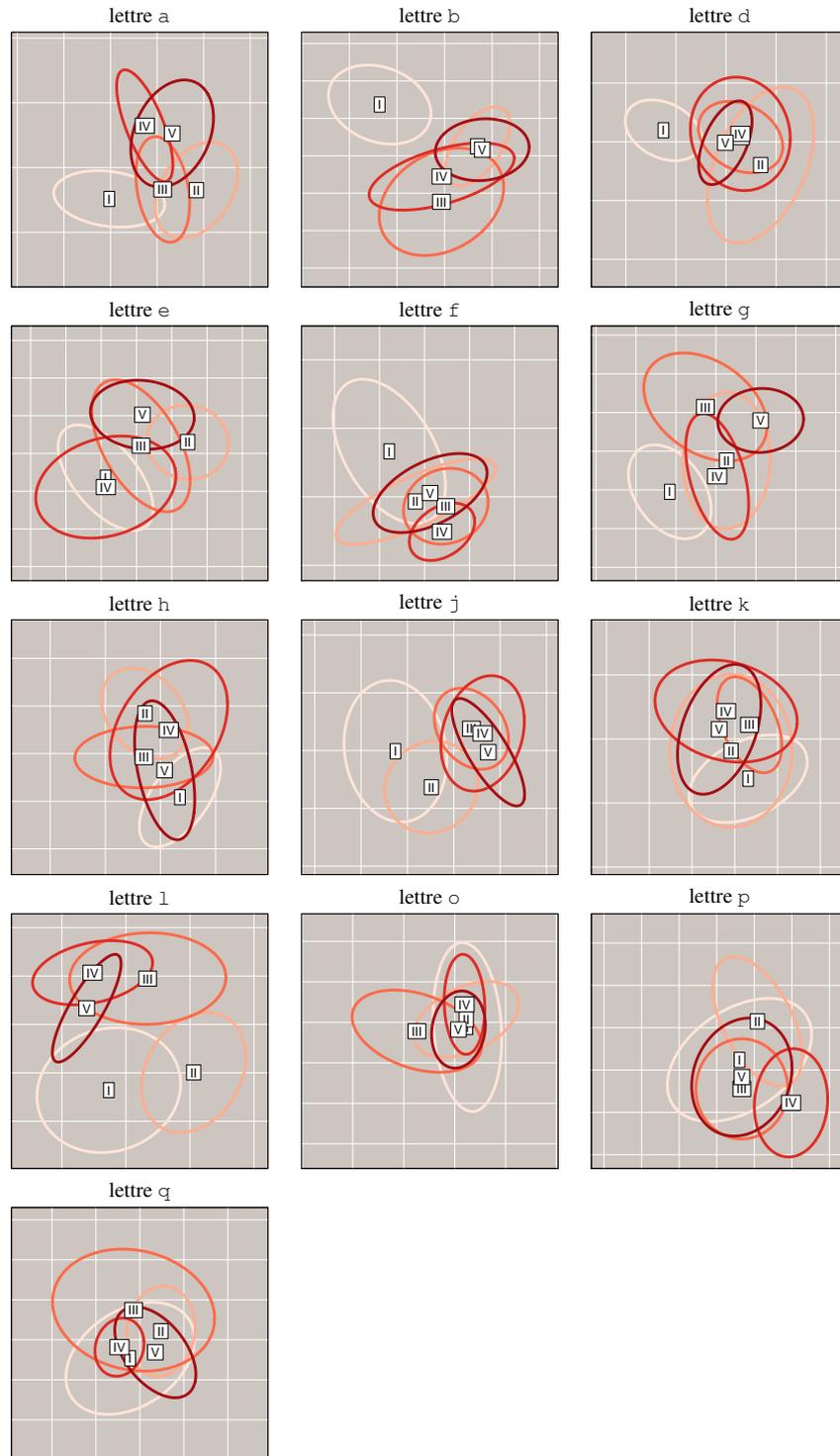


FIGURE 5.169: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

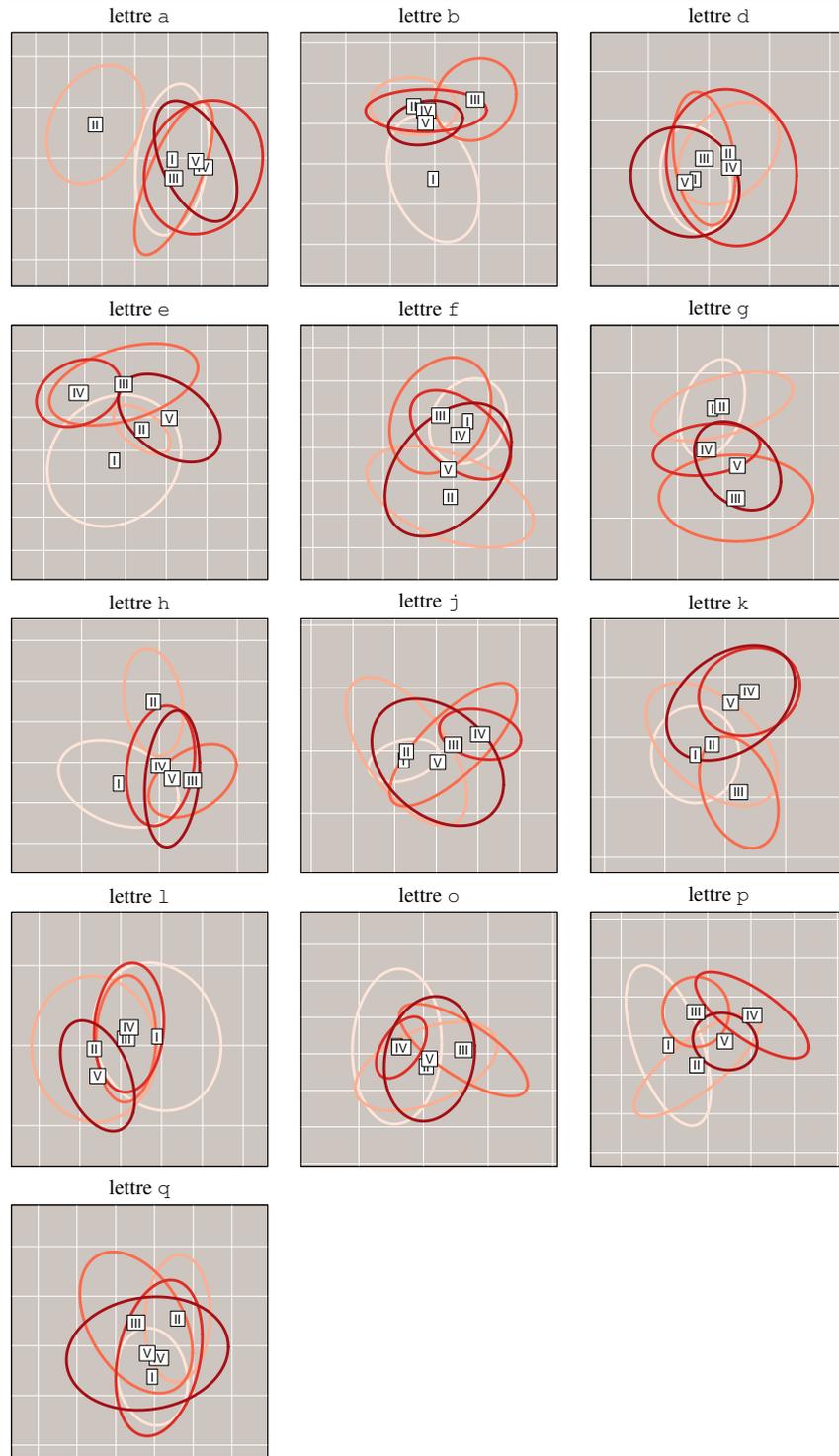


FIGURE 5.170: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

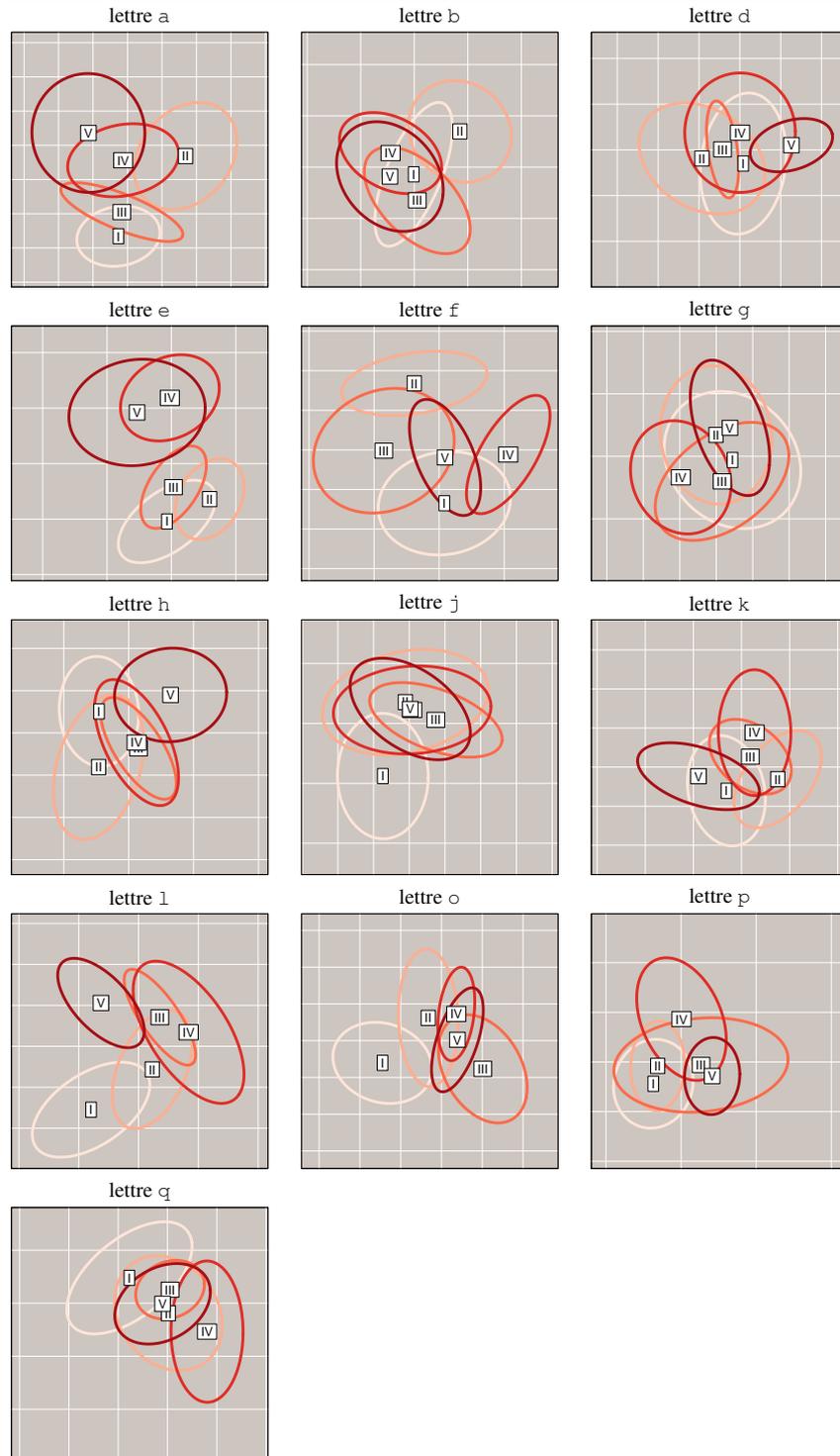


FIGURE 5.171: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

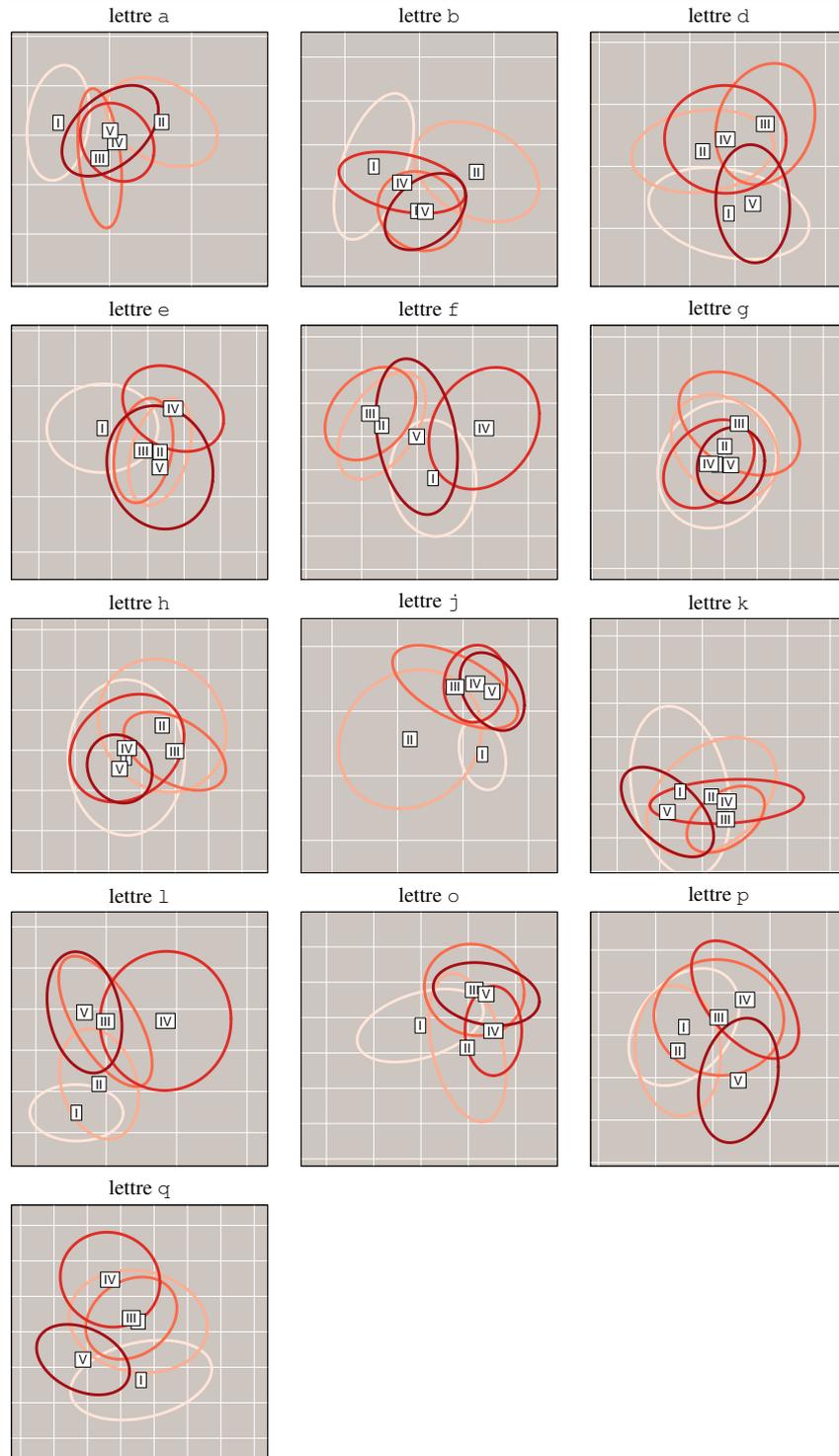


FIGURE 5.172: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

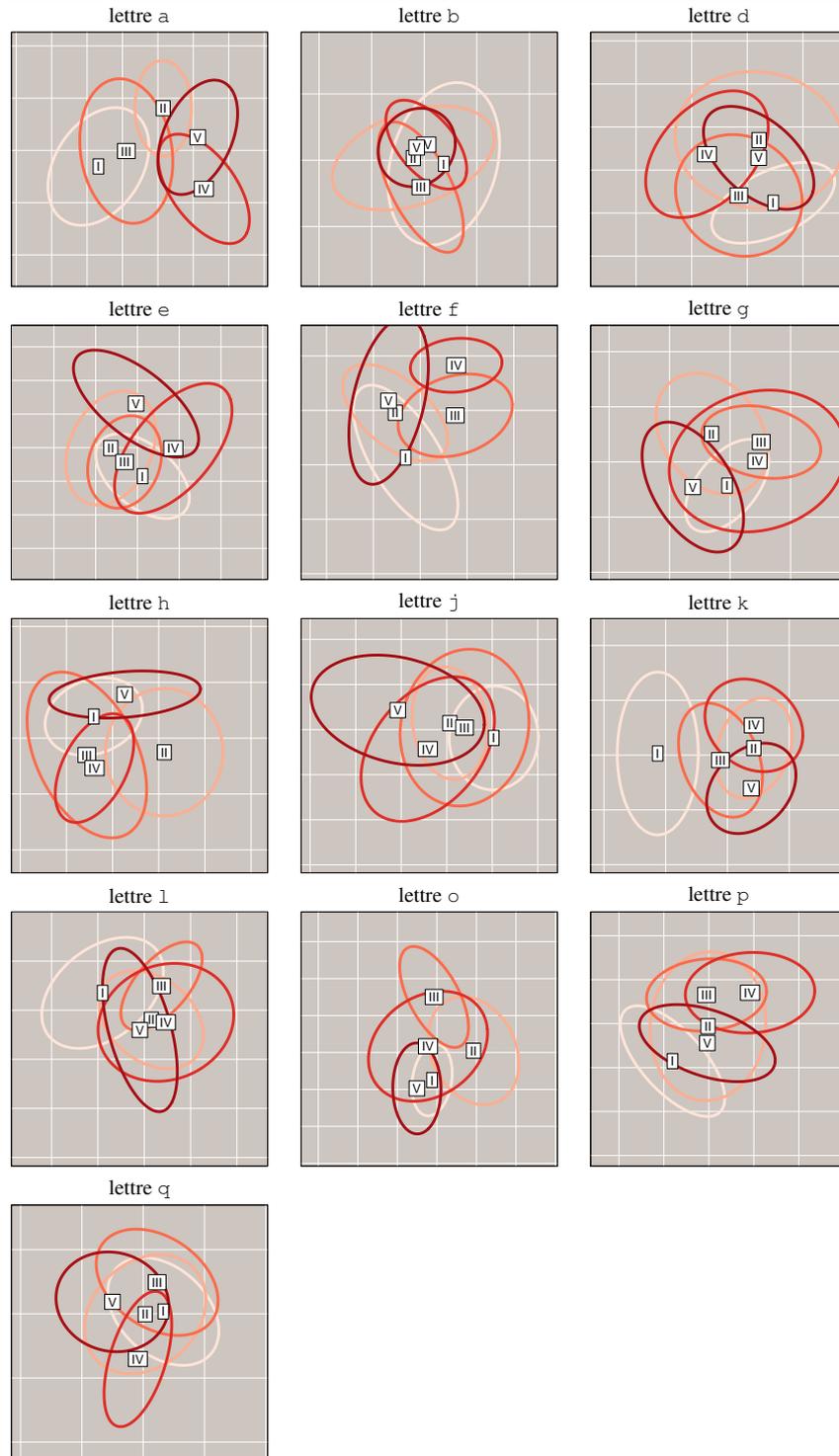


FIGURE 5.173: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

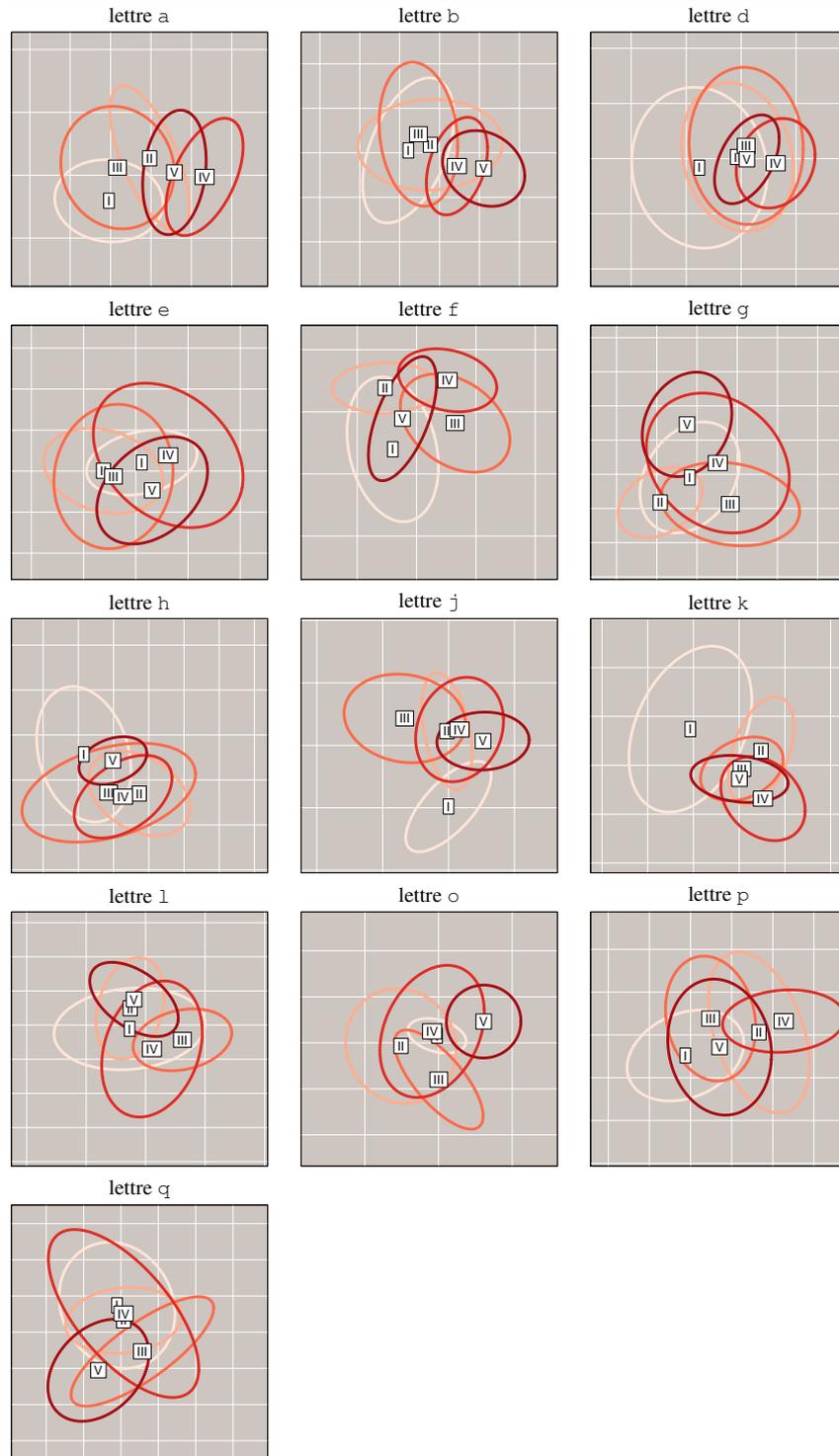


FIGURE 5.174: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

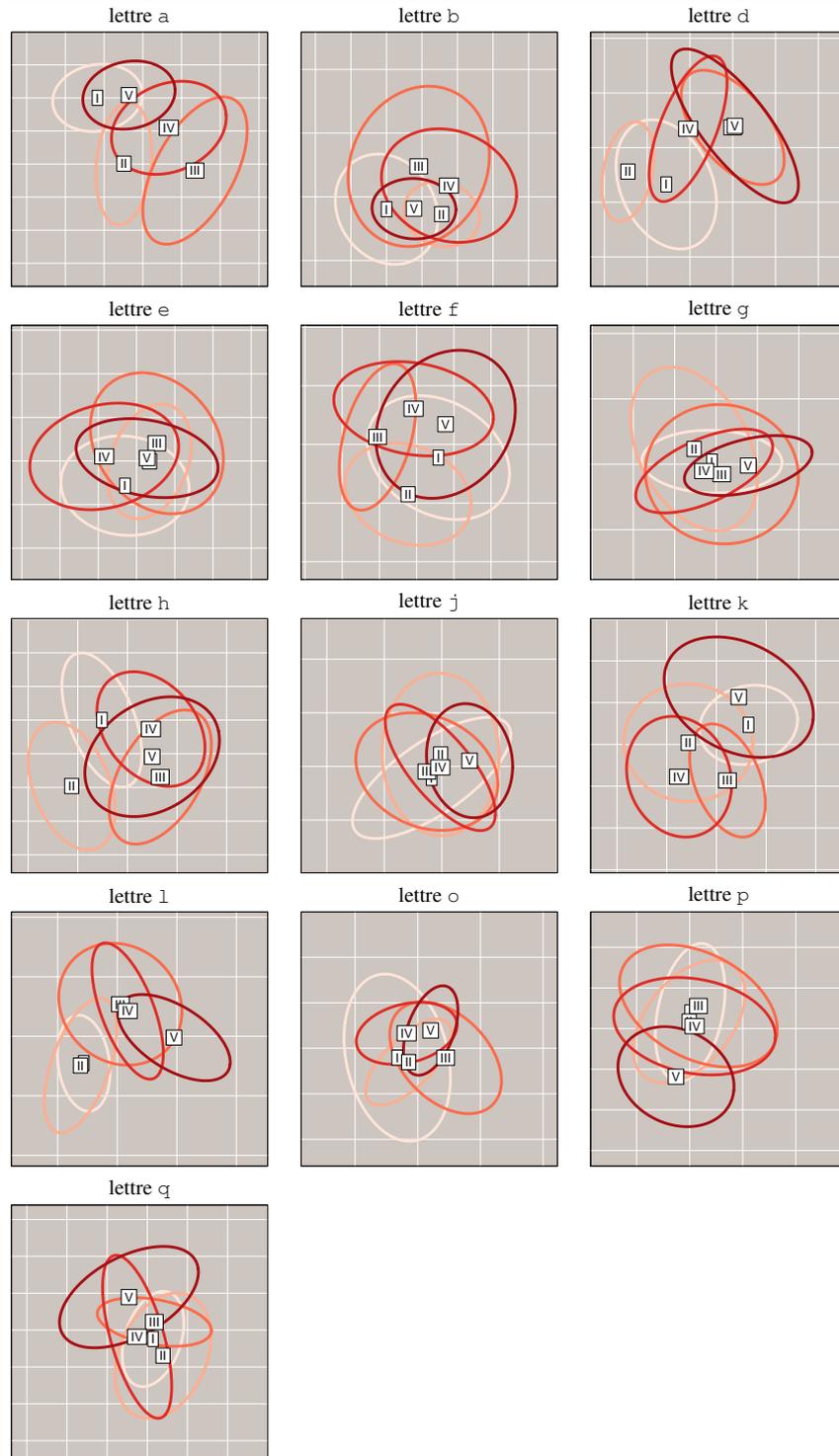


FIGURE 5.175: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

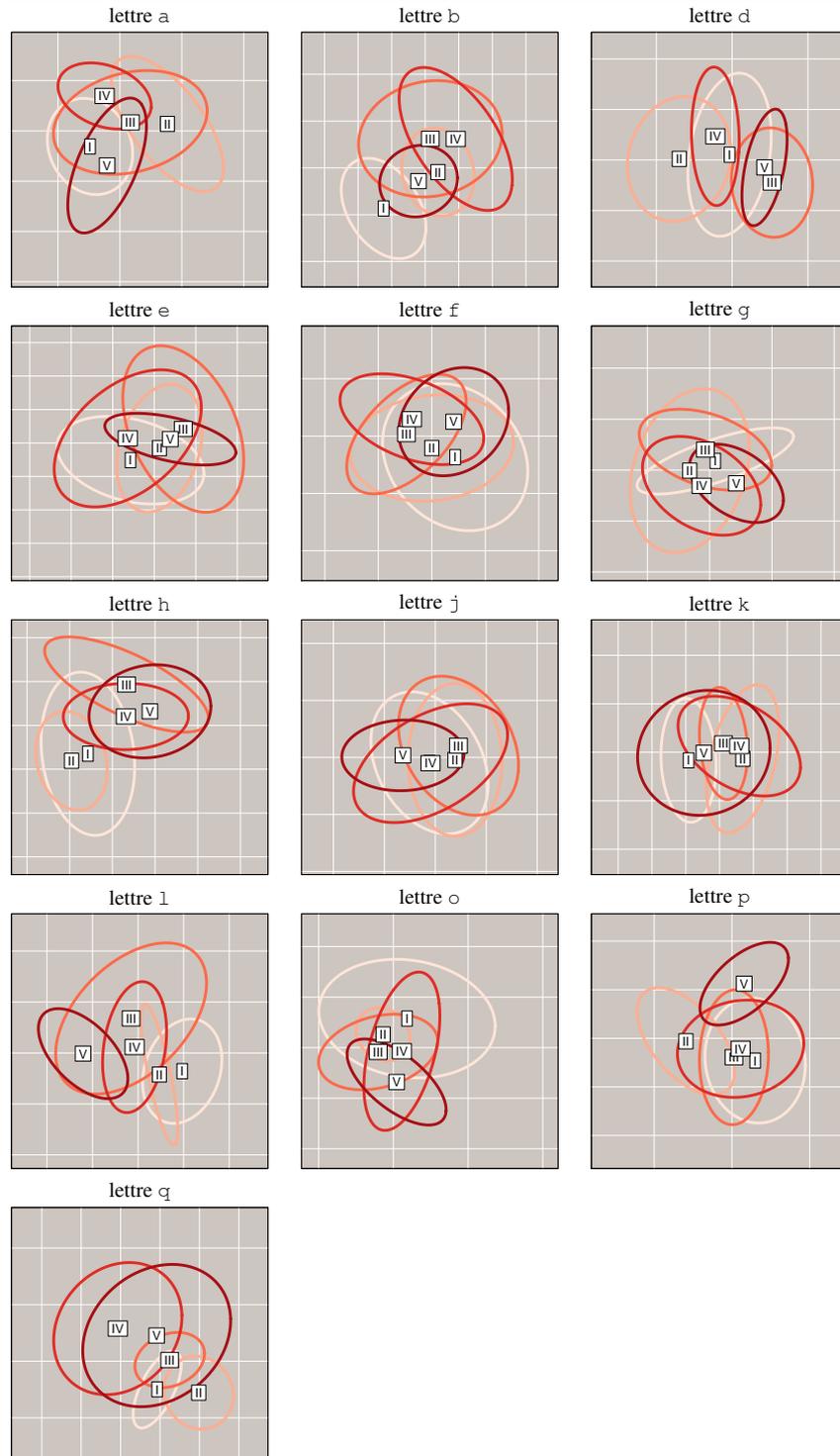


FIGURE 5.176: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages (I – V).

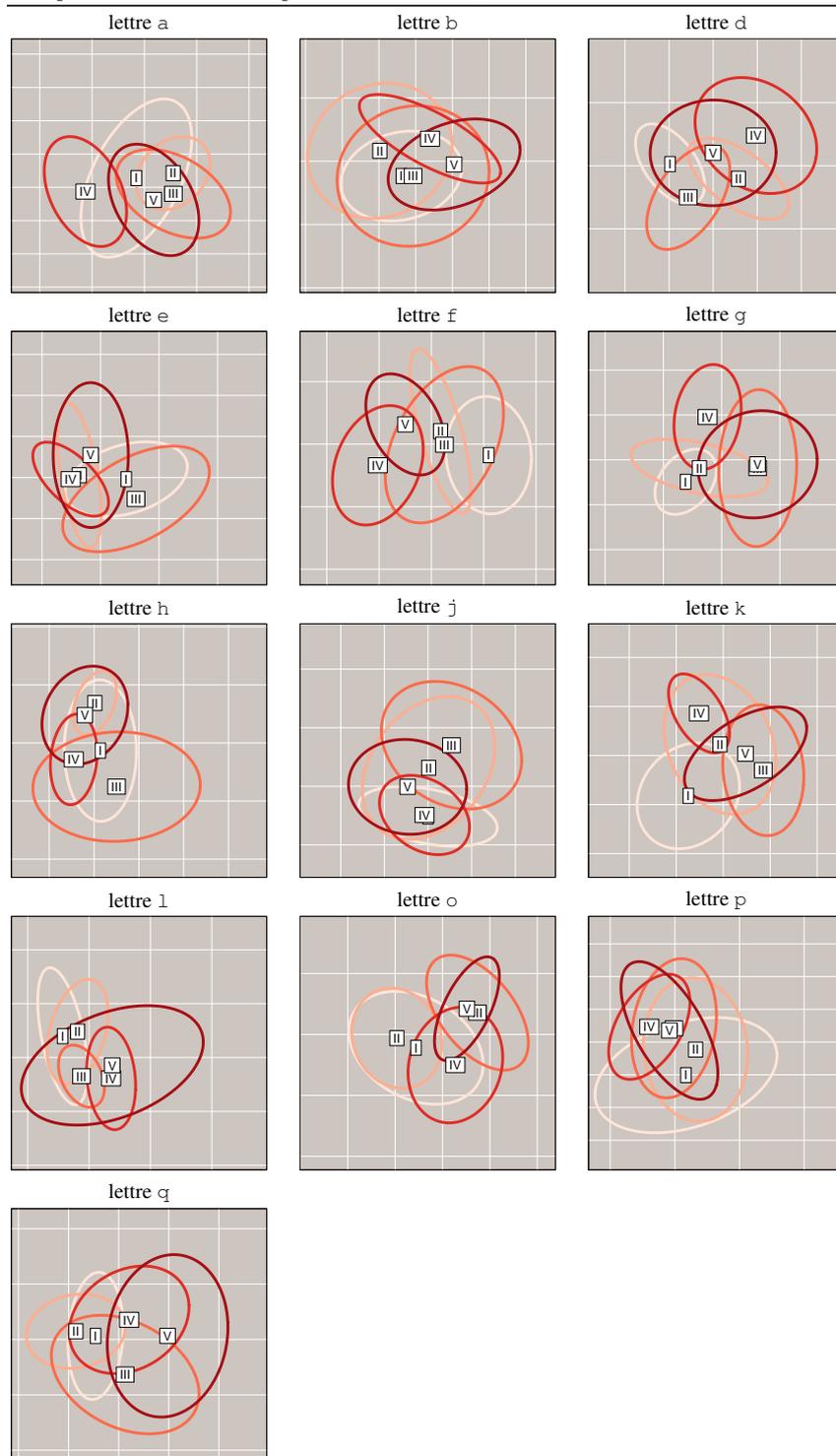


FIGURE 5.177: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

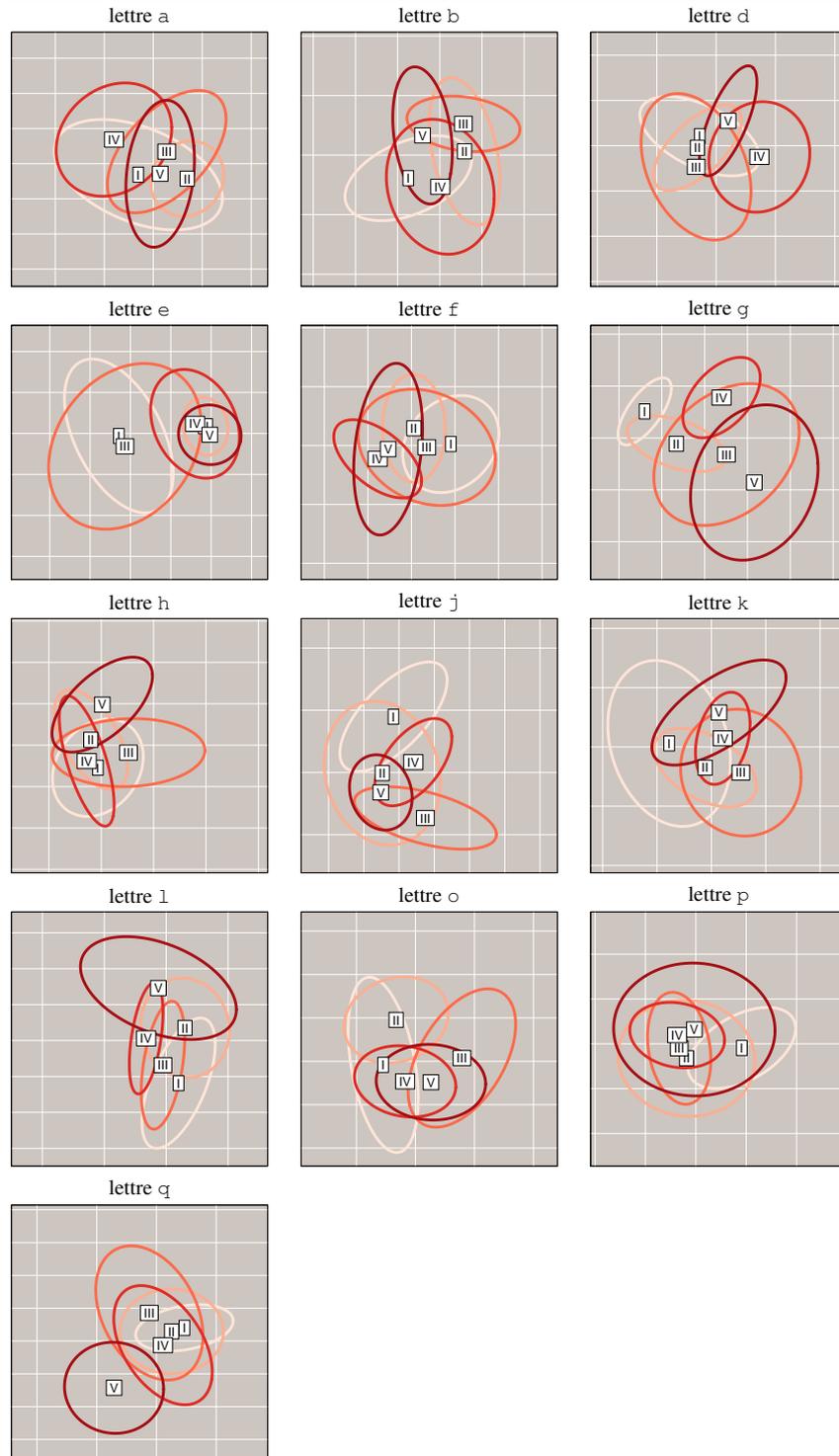


FIGURE 5.178: Représentation par la technique de PCoA de matrices de dissimilarité de l'expérience 2, groupés selon les cinq pages ( $I - V$ ).

## 5.7 Expérience 3

### Représentation par Analyse en Coordonnées Principales

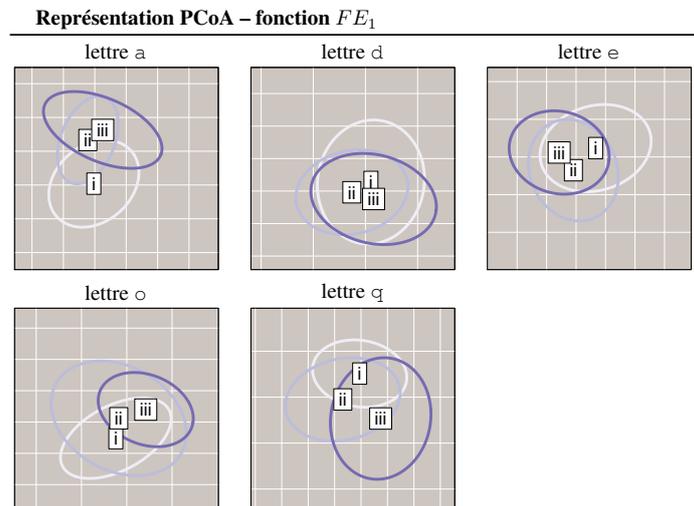


FIGURE 5.179: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* du scripteur 7 de l'expérience 3, groupés par sessions ( $i - iii$ ).

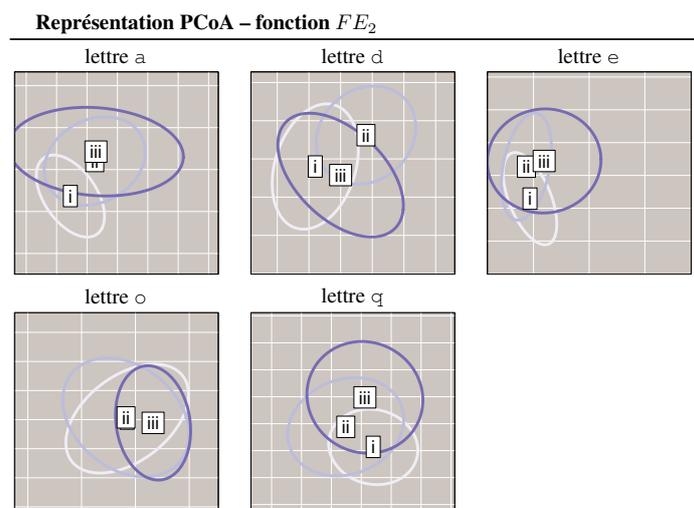


FIGURE 5.180: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* du scripteur 7 de l'expérience 3, groupés par sessions ( $i - iii$ ).

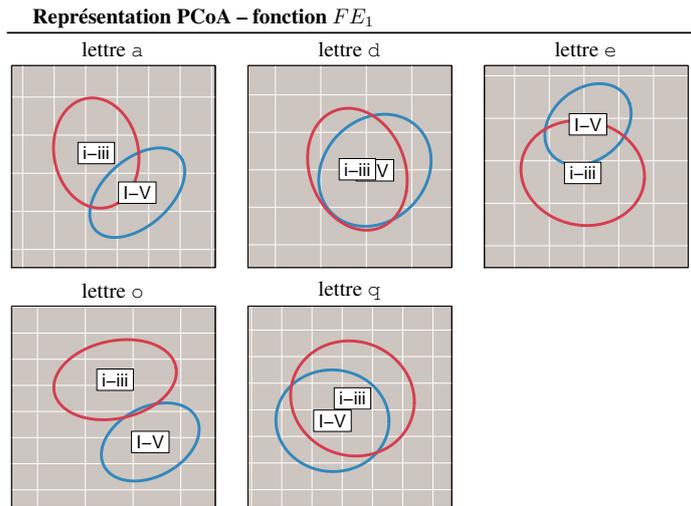


FIGURE 5.181: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* du scripteur 7 des expériences 2 et 3.

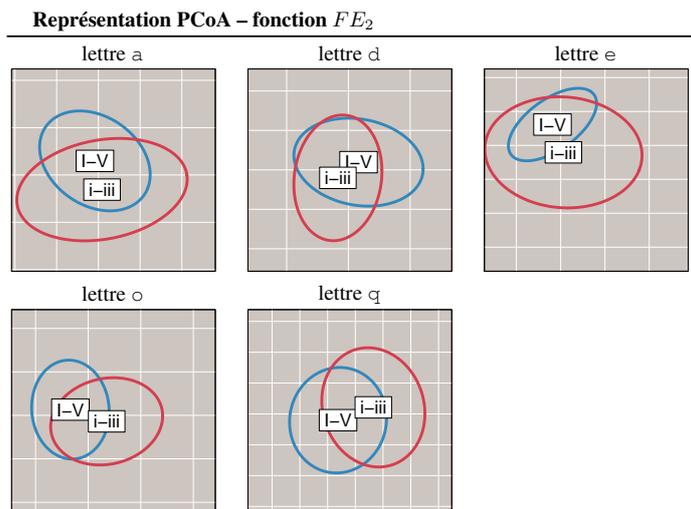


FIGURE 5.182: Illustration par PCoA des matrices de dissimilarité pour les caractères *normaux* du scripteur 7 des expériences 2 et 3.



ISBN 2-940098-66-2  
Série Criminalistique N° LXII

ISBN 2-940098-66-2  
Série Criminalistique N° LXII