

Pour rédiger un rapport de statistique

Résumé

Cette vignette est destinée aux étudiants des formations à finalité professionnelle devant rédiger des comptes-rendus de traitements statistiques réalisés au cours d'études de cas ou en stage. Il se contente de rappeler les règles élémentaires que doit respecter la forme de tels documents.

1 Introduction

Les rapports relatant le traitement informatique et l'analyse statistique d'un jeu de données réelles comportent souvent de nombreuses maladresses tant de forme que de contenu. Ce document, destiné à évoluer, tente d'en dresser une liste synthétique afin, si possible, d'éviter aux promotions à venir de reproduire inlassablement ces mêmes errements.

Compte tenu de la performance des outils informatiques mis à la disposition des étudiants, ceux-ci doivent en utiliser au mieux les fonctionnalités afin de produire des documents percutants dans leur forme et pertinents sur le fond.

2 Forme

Plusieurs siècles de tradition typographique ont abouti à un ensemble de règles de composition que doit respecter tout document technique ; y déroger ne fait que perturber l'attention du lecteur et nuit donc à la bonne compréhension du texte. Ces règles sont très nombreuses, voici les principales.

2.1 Police et style

Pour être facile à lire, un long document doit utiliser une police de caractères à espacements proportionnels avec *serifs* c'est à dire avec empattements des lettres (de type *times* ou *roman*). Ceux-ci guident l'œil et facilitent l'attention du lecteur sur la ligne en cours de déchiffrement. Pour la même raison, une ligne doit comporter à peu près 70 signes, espaces compris. Sa longueur, et ainsi la

taille des marges, sont donc fonction de la taille des caractères. Une des règles essentielles est la *parcimonie* ; il faut éviter les changements intempestifs de police ou de style de caractères. Chaque enrichissement ne doit pas être gratuit mais chargé d'un sens identique pour tout le document :

- une police à espacements fixes (*courrier*) permet d'introduire des extraits de sorties d'un logiciel,
- un mot est mis en *italique* pour attirer l'attention au cours de la lecture,
- un mot en **gras** est repéré dans la page avant que celle-ci ne soit lue,
- le soulignement, indispensable dans un document manuscrit, devient inutile, il est à éviter,
- des caractères plus grands sont utilisés pour titres et sous-titres,
- ...

2.2 Gestion des espaces

2.2.1 Espacements horizontaux

La règle en vigueur est simple à retenir :

- tout signe de ponctuation simple “.” “;” est collé au mot qui le précède et suivi d'un espace,
- tout signe de ponctuation double “;” “:” “?” “!” est précédé d'un espace insécable (pour lui éviter d'apparaître seul en début de ligne) et suivi d'un espace ¹,
- une parenthèse ouvrante (resp. fermante) est précédée (resp. suivie) d'un espace.

Il est préférable de justifier le texte à gauche *et* à droite, la césure automatique permet encore d'en améliorer l'apparence.

2.2.2 Espacements verticaux

Un espacement plus grand que l'interligne standard souligne le découpage du texte en paragraphes. Éviter impérativement les titres de paragraphes seuls en bas de page ou les lignes isolées en début de page ; si le logiciel ne sait pas opérer cette justification verticale, elle est obtenue, en fin de rédaction, en gérant manuellement les sauts de page et les espacements de paragraphes.

1. En anglais ou en français québécois, les signes de ponctuation doubles ne sont pas précédés d'un espace.

2.3 Structure

Le rapport doit être fortement structuré et son plan peut s'inspirer de celui décrit ci-dessous. Les quatre premiers points sont intangibles.

1. Couverture : elle donne l'intitulé des Universités, de la formation, Prénom(s) et Nom(s) du (des) auteur(s), le titre précis, l'année.
2. **Résumé d'auteur** : *Indispensable*, il présente le domaine d'étude et le problème, résume la démarche et les principaux résultats. Il a un objectif totalement différent de l'introduction même si certaines redondances peuvent apparaître. Il doit être suffisamment attractif pour inciter le lecteur pressé à poursuivre son investigation.
3. Table des matières : elle donne la liste des paragraphes et sous-paragraphes avec numérotation des pages du rapport ainsi que de celles des annexes.
4. Introduction et présentation des données : citer les sources, l'origine des données, décrire les variables, préciser les objectifs, indiquer la démarche ou la méthodologie.
5. Description élémentaire : tris à plat, histogrammes, diagrammes, tests, transformations et sélections des variables..., corrélations, tables de contingences, diagrammes, tableau symétrique des nuages de points. . .
6. Traitements multi-variés : techniques de statistique exploratoire multidimensionnelle.
7. Modélisations, estimations des paramètres et tests.
8. Conclusion générale sur les résultats obtenus.
9. Références.
10. Annexes structurées et numérotées.

2.4 Rédaction

Le rapport lui-même est succinct mais rédigé clairement en évitant tout style télégraphique et abréviations ou sigles abscons. En particulier, les libellés des variables et de leurs modalités doivent être explicites dans les commentaires ou faire appel à des codes mnémoniques limpides. Il est vivement conseillé d'utiliser systématiquement le correcteur orthographique.

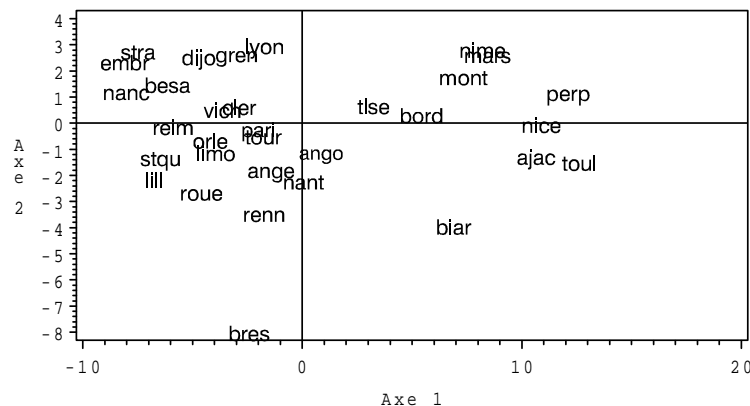


FIGURE 1 – ACP des courbes de températures : premier plan des individus.

Sauf cas très exceptionnel, tout ce qui est calcul, listing... est référencé et reporté en annexe. Chaque annexe a un titre, elles sont numérotées en chiffres romains grandes capitales.

Il n'est pas question, même si c'est très tentant, de lister tous les traitements entrepris sur les données. *Seules* les analyses qui participent à la poursuite des objectifs de l'introduction sont décrites et commentées. Les autres sont signalées pour mémoire comme inintéressantes.

La plus grosse difficulté est l'obtention d'une intégration *harmonieuse* des graphiques pertinents, des tableaux synthétiques et des commentaires associés. Chacun de ces objets statistiques est numéroté et accompagné d'une légende en italique (cf. la figure 1). Ils doivent être de taille réduite, et le texte rédigé doit rester prépondérant en volume.

2.5 Appel de référence

Un appel à une référence bibliographique est incorporé dans le texte plutôt que relégué en note de bas de page. Ainsi, l'algorithme proposé par Dupont (1902, p115-120) est sans doute meilleur que les autres et il est connu que "*ce qui est mieux n'est pas pire*" (Dupond et Dupont 1901, Dupont 1902).

3 Contenu

3.1 Pertinence

Mises à part les erreurs les plus flagrantes concernant l'utilisation des techniques statistiques ou leurs interprétations, les maladresses les plus couramment rencontrées proviennent d'une utilisation impropre de certains termes à la définition précise, de l'introduction d'informations inutiles, inadaptées ou parasites. Les exemples sont nombreux :

- Seule est définie la *corrélation* entre variables quantitatives. Cela n'a pas de sens d'y faire mention pour des variables qualitatives (parler simplement de liaison), encore moi pour des modalités qui sont simplement en correspondance.
 - Tracer une droite de régression à travers un nuage de points qui atteste d'une absence de corrélation ou d'une liaison non-linéaire,
 - surcharge "esthétique" sans signification statistique : diagrammes en perspective, couleurs mal équilibrées.
 - ...
- D'autres erreurs et maladresses à éviter sont listées en annexe I.

3.2 Cohérence

- pas de contradiction entre des commentaires sans explications,
- ne pas mélanger les niveaux d'interprétation (informatique, statistique, économique, sociologique),
- pas de longs discours interprétatifs sur des résultats jugés ensuite douteux,
- ...

4 Conclusion

Les enseignants de statistique qui ont contribué à la rédaction de ce rapport souhaitent vivement que les étudiants le lisent avec une attention soutenue et qu'ils en appliquent les principes dans leurs travaux à venir. La notation tiendra compte du respect ou non de ces quelques règles élémentaires.

Références

Colignon, J.P. (1975), *La ponctuation, Art et Finesse*.

Dupond, J.M. et Dupont, M.J. (1902), "Noix et noisettes de Quercy", *Journal de Statistique rurale*, **3**, 1148-1149.

Dupont, M.J. (1901), *Noix et noisettes* (vol.5 ; 7ème édition), Donud, Cahors.

Imprimerie Nationale (1990), *Lexique des Règles typographiques en usage à l'Imprimerie Nationale*, Imprimerie Nationale.

Ramat, A. (1989), *Grammaire typographique*, Aurel Ramat éditeur, Québec.

Annexes

I Erreurs fréquentes

Liste non-exhaustive (!) des erreurs et maladresses qu'il convient d'éviter lors de l'utilisation des différentes techniques statistiques.

I.1 Outils élémentaires

- traitements et graphiques pas adaptés aux types des variables,
- diagramme en barres d'une variable qualitative issue du codage en classes d'effectifs égaux d'une variable quantitative,
- énorme diagramme en 2 secteurs,
- droite de régression traversant un nuage de points attestant une absence de corrélation ou une liaison non-linéaire ou encore une droite fortement influencée par un individu isolé. . .
- surcharge "esthétique" sans signification statistique : diagrammes en perspective, couleurs,
- jonction de points sans objet,
- tableau ou graphique (axes, échelles) sans légende donc sans signification,
- ...

I.2 Techniques multidimensionnelles

- plans factoriels illisibles car surchargés,
- effet taille conduisant à des graphiques sans intérêt,
- ACP sur des tableaux d'effectifs,
- distributions log-normales ou poissonniennes non transformées et traitées à tort comme des données gaussiennes produisent des kyrielles de points "aberrants",
- influence prépondérante de parasites sur les premiers axes factoriels d'une ACM : données manquantes, non-réponses, modalités à faibles effectifs, liaisons artificielles entre variables redondantes. . .
- ...

I.3 Modélisation

- préciser clairement le modèle et les paramètres à estimer,
- attention aux problèmes numériques dus, par exemple, au mauvais condi-

tionnement de la matrice de covariance des variables explicatives,

– ...