

Partie 1 Traitement d'un fichier de notes

Chapitre 1. Description des données étudiées

Chapitre 2. Objectifs de l'étude

2.1	De l'intérêt de préciser des objectifs	19
2.2	Quelques questions préalables (non indépendantes)	19

Chapitre 3. Premières vérifications des données

3.1	Principe presque absolu : les fichiers ne sont jamais propres	21
3.2	Mauvaise lecture du fichier	21
3.3	Quelques erreurs classiques	23
3.4	Détection de valeurs aberrantes	24
3.5	Bilan des valeurs aberrantes	24
3.6	Cerner le problème de chaque valeur aberrante	24
3.7	Corriger, estimer une valeur, supprimer un individu	25
3.8	Que faire s'il y a beaucoup de valeurs aberrantes ?	26
3.9	Grosses erreurs et petites erreurs	27
3.10	Quelques autres types d'erreurs	28
3.11	Bilan-résumé sur la recherche d'erreurs	28

Chapitre 4. Données manquantes

4.1	Remarques préliminaires sur le codage des données manquantes	29
4.2	Bilan des données manquantes	30
4.3	Etude des données manquantes à travers les individus	32
4.4	Répartition des 83 élèves incomplets en 4 groupes homogènes	33
4.5	Groupe des fantômes	34
4.6	Groupe des 15 élèves avec bac incomplet	34
4.7	Groupe des 13 élèves qui n'ont de notes qu'au bac seulement	36
4.8	Groupe des élèves avec quelques valeurs manquantes	37
4.9	Conclusion	38
4.10	Bilan-résumé sur les données manquantes	24

Chapitre 5. Description d'un petit tableau de données : les 15 élèves avec bac incomplet

5.1	Présentation ordonnée d'un tableau	41
5.2	Représentation axiale d'une variable quantitative	42
5.3	Représentation graphique de deux variables quantitatives sur un plan	43

Chapitre 6. Etude d'une variable qualitative : répartition des élèves dans les lycées

6.1	Tri à plat	45
6.2	Diagramme en bâtons triés par effectif décroissant	45
6.3	Le regroupement, moyen efficace de description des données	46
6.4	Diagramme circulaire	46

Chapitre 7. Etude de variables quantitatives : répartition des notes

7.1	Variable discrète ; diagramme en bâtons	49
7.2	Variable continue ou discrète ; histogramme	51
7.3	Moyennes des notes	55
7.4	Quelques notations utiles	57
7.5	Maximum, minimum, étendue	58
7.6	Dispersion autour de la moyenne : écart absolu moyen, écart-type	59
7.7	Boîte de dispersion, médiane, quartile, outlier	61
7.8	Pourcentages par rapport à des valeurs de référence	65
7.9	Que choisir ?	66
7.10	Influence des groupes d'élèves ayant des données manquantes	67
7.11	Exemple de synthèse	68

Chapitre 8. Liaison entre deux variables quantitatives : les notes sont-elles liées entre elles ?

8.1	Le problème	71
8.2	Etude graphique de la liaison entre deux variables quantitatives	71
8.3	Tableau croisé à partir de valeurs de références	74
8.4	Coefficient de corrélation	84
8.5	Distribution de la différence entre deux variables	80
8.6	Que choisir ?	82
8.7	Régression	82

Chapitre 9. Synthèse d'un ensemble de variables quantitatives

9.1	Deux objectifs de l'analyse en composantes principales	91
9.2	Première composante principale	94
9.3	Deuxième composante principale	96
9.4	Représentation des élèves et des lycées	97
9.5	Plan des deux premières composantes : bilan des corrélations entre variables	101
9.6	Troisième et quatrième composantes	102
9.7	Suite du bilan des corrélations (plan 3-4)	103
9.8	Cinquième composante	103
9.9	Conclusion	104

Chapitre 10. Caractérisation d'une sous-population ; élèves avec données manquantes

10.1	Les élèves avec données manquantes proviennent-ils de lycées particuliers ?	105
10.2	Les élèves avec données manquantes ont-ils des notes particulières ?	111
10.3	Autres explorations	115

Chapitre 11. Comparaison entre plusieurs sous-populations : les élèves d'un même lycée

11.1	Sur quelles variables fonder la comparaison ?	117
11.2	Que signifie « comparer plusieurs sous-populations » ?	118
11.3	Comparaison directe des moyennes sur une variable : la note du bac	118
11.4	Probabilité associée à une moyenne calculée pour une sous-population	119
11.5	Comparaison entre les répartitions des notes des lycées	121
11.6	Comparaison entre les extrema	122
11.7	Liaison entre une variable quantitative et une variable qualitative	122
11.8	Comparaison selon deux variables	123
11.9	Comparaison selon l'ensemble des variables ; caractérisation d'un lycée	125
11.10	Conclusion	126

Partie 2. Éléments remarquables et éléments aberrants

Introduction	129
Chapitre 12. Mise en évidence de valeurs remarquables et de valeurs aberrantes	
12.1 Examen systématique des distributions	131
12.2 Intérêt du centrage et de la réduction	131
12.3 Approche systématique pour mettre en évidence des valeurs remarquables	132
12.4 Limites du centrage et de la réduction dans la recherche de valeurs remarquables	134
12.5 Appréciation du caractère aberrant d'une valeur	135
Chapitre 13. Mise en évidence d'individus remarquables	
13.1 Qu'est-ce qu'un individu remarquable ?	139
13.2 Cas où l'on ne prend pas en compte les liaisons entre variables	139
13.3 Cas de deux variables liées linéairement	143
13.4 Procédure de détection systématique d'individus remarquables	131
13.5 Prolongements possibles	148
Chapitre 14. Mise en évidence de variables remarquables	
14.1 Introduction	151
14.2 Asymétrie	151
14.3 Aplatissement	154
14.4 Probabilité associée à un coefficient de forme	156

Partie 3. Fiches techniques

Fiche 1. Construction du tableau de données, type de variables, codage

1 Variable qualitative (ou nominale)	161
2 Variable qualitative ordonnée	162
3 Variable indicatrice	164
4 Variable quantitative	165
5 Variable ordinale	167
6 Fréquence et tableau de contingence	167

Fiche 2. Données manquantes

1 Quelques types de données manquantes	171
2 Prise en compte dans les traitements statistiques	173
3 Conclusion	174

Fiche 3. Mesure de la dispersion d'une variable quantitative

1 Pourquoi mesurer la dispersion ?	175
2 Ecart absolu moyen (E_m)	175
3 Ecart-type (s) et variance (s^2)	176
4 Etendue et écart interquantile	179
5 Niveau d'un échantillon ou d'une population	180

Fiche 4. Représentation simultanée de deux variables quantitatives

1 Exemple dans lequel le choix des unités de mesure s'impose	181
2 Exemple dans lequel le choix des unités de mesure pose problème	181

Fiche 5. Liaison entre deux variables quantitatives

1	Du graphique au coefficient de corrélation	185
2	Peut-on apprécier le caractère plus ou moins fortuit d'un coefficient de corrélation ?	188
3	Coefficient de corrélation et forme du nuage de points associé	189
4	Matrice des corrélations	190

Fiche 6. Liaison entre deux variables qualitatives

1	Tableau des données et tableau de référence	191
2	Cas de deux variables ayant chacune deux modalités	193
	2.1 Approche fondée sur l'un des effectifs	193
	2.2 Approche fondée sur le critère du χ^2	195
	2.3 Lien entre le critère du χ^2 et l'approche fondée sur un seul effectif	196
3	Généralisation au cas de deux variables ayant un nombre quelconque de modalités	197
4	Quelques prolongements	200
	4.1 Mesurer une liaison et étudier une liaison	200
	4.2 Cas d'un grand nombre de variables qualitatives	200
	4.3 Test classique du χ^2	201
	4.4 Indépendance entre trois variables	201

Fiche 7. Comparaison entre deux moyennes

1	Cas de moyennes d'une même variable définie sur deux ensembles d'individus	203
	1.1 Présentation d'un exemple	203
	1.2 Calcul et utilisation d'une probabilité associée	204
	1.3 Valeur-test	205
	1.4 Test t classique	206
2	Cas de moyennes de deux variables définies sur le même ensemble d'individus	209
	2.1 Spécificité des données appariées	209
	2.2 Calcul et utilisation d'une probabilité associée	209
	2.3 Test t classique dans le cas de données appariées	211
	2.4 Données appariées et graphique	212

Fiche 8. Liaison entre une variable quantitative et une variable qualitative

1	Données, problématique	213
2	Approche graphique	213
3	Trois variabilités en présence : totale, inter-classes, intra-classes	214
4	Comparaison entre les variabilités : rapport de corrélation	215
5	Test F de Fisher	217
6	Vers l'analyse de variance à plusieurs facteurs	221

Fiche 9. Distribution de variables quantitatives, observées ou aléatoires

1	Distribution observée, distribution théorique, variable aléatoire	227
2	Distribution de variables observées	227
	2.1 Représentations graphiques	227
	2.2 Moyenne et variance	228
	2.3 Distribution conjointe de deux variables	229
	2.4 Indépendance entre deux variables	229
3	Distribution de variables aléatoires discrètes	229
	3.1 Loi uniforme	229
	3.2 Loi binomiale et loi hypergéométrique	230
	3.3 Moyenne et variance	232
	3.4 Distribution conjointe de deux variables	232
	3.5 Indépendance entre deux variables	233

4	Distribution de variables aléatoires continues	233
4.1	Loi uniforme sur un intervalle	235
4.2	Loi normale	235
4.3	Quelques autres lois	237
4.4	Distribution conjointe et indépendance entre deux variables	238
5	Modèle et variable observée	240

Fiche 10. Indicateur statistique et probabilité associée

1	Pourquoi utiliser des probabilités dans l'examen d'un ensemble de données ?	241
2	Modèle de tirage au hasard intérieur aux données	241
3	Utilisation de la probabilité associée	247
3.1	Portées respectives de l'indicateur statistique et de la probabilité associée	247
3.2	Peut-on porter un jugement absolu sur une probabilité associée ?	248
4	Cas de données obtenues à partir d'un échantillon tiré au hasard dans une population	249
4.1	Domaine de l'inférence statistique classique	249
4.2	Principe de l'inférence statistique classique	249
4.3	Seuils, erreurs et risques	251
4.4	A quelle population généraliser les résultats observés ?	253
4.5	En conclusion	253

Fiche 11. Distribution d'une moyenne

1	Moyenne d'un échantillon	255
2	Moyenne de deux ou de plusieurs variables	257

Index	259
--------------	-----

Bibliographie	230
----------------------	-----
