



Nathalie Villa-Vialaneix

Illustration du cours de
Statistique Descriptive I (M1102)

Année scolaire 2013/2014



Université de Perpignan Via Domitia, IUT
STatistique et Informatique Décisionnelle (STID)

Table des matières

1	Généralités sur la statistique descriptive	5
1.1	Population, échantillon	6
1.2	Variables	7
1.3	Mise en forme des données	7
1.3.1	Nettoyage des données	7
1.3.2	Variables quantitatives continues	7
1.3.3	Représentativité	7
1.3.4	Variables quantitatives discrètes	8
1.3.5	Réponses ouvertes	8
2	Représentations graphiques	9
2.1	Variables qualitatives	9
2.1.1	Diagrammes en colonnes	9
2.1.2	Diagrammes en barre	9
2.1.3	Diagrammes circulaires	9
2.2	Variables quantitatives discrètes	9
2.2.1	Diagrammes en bâtons	9
2.2.2	Diagrammes cumulatifs	9
2.2.3	Séries chronologiques	9
2.3	Variables quantitatives continues	10
2.3.1	Histogrammes	10
2.3.2	Polygones cumulatifs	10
2.4	Autres représentations	10
2.4.1	Diagrammes figuratifs	10
2.4.2	Cartes	10
3	Caractéristiques numériques	13
3.1	Caractéristiques de tendance centrale	13
3.1.1	Mode	13
3.1.2	Moyenne	13
3.1.3	Médiane	13
3.2	Caractéristiques de dispersion	13
3.2.1	Étendue	13
3.2.2	Variance et écart type	14
3.2.3	Quantiles, quartiles	14

3.3	Forme d'une distribution	14
3.3.1	Comparaison Moyenne / Médiane	14
3.3.2	Coefficient d'asymétrie	14
3.3.3	Coefficient d'aplatissement	15
3.3.4	Courbe de Lorenz et indice de Gini	15
4	Réponses	17
4.1	Généralités sur la statistique descriptive	17
4.1.1	Variable « Taille » (Question 1.3.2 page 7)	17
4.1.2	Variable « CSP du chef de famille » (Question 1.3.3 page 7)	17
4.1.3	Variable « Note en mathématiques au baccalauréat » (Question 1.3.4 page 8)	18
4.2	Représentations graphiques	19
4.2.1	Variable « Mention au baccalauréat » (Question 2.1.1 page 9)	19
4.2.2	Variable « Âge » (Question 2.2.1 page 9)	20
4.2.3	Variable « Taille » (Question 2.3.2 page 10)	20
4.2.4	Variable « Couleur des yeux » (Question 2.4.1 page 10)	21
4.2.5	Variable « Lieu de naissance » (Question 2.4.2 page 10)	22
4.2.6	Variable « Nombre de frères et sœurs » (Question 3.3.4 page 15)	22

1 Généralités sur la statistique descriptive

Introduction

Les réponses à tous les exercices de ce livret peuvent être utilisées pour des exercices plus avancés. Elles sont donc à remplir directement dans le livret lui-même : un espace dédié est disponible à la fin du livret, page 17.

Les exercices suivants sont relatifs au questionnaire distribué à l'ensemble des étudiants de 1^{ère} année STID et d'APPC que vous avez rempli au début de la première séance. Nous rappelons, ci-dessous, ce questionnaire :

État civil

1. Âge : _____
2. Né(e) en France ? Oui ; Non
3. Si oui, code postal du lieu de naissance : _____
4. Sexe : Homme ; Femme
5. Nombre de frères et soeurs : _____

Informations physiques

1. Taille (en cm) : _____
2. Pointure : _____
3. Couleur des yeux : Bleu ; Marron ; Vert
4. Couleur des yeux de votre mère : Bleu ; Marron ; Vert

Loisirs

1. Couleur de la dernière voiture dans laquelle vous êtes monté : _____
2. Dernier département, autre que l'Aude, dans lequel vous êtes allé : _____
3. Lieu que vous aimeriez visiter : _____
4. Vous intéressez-vous au football ? Pas du tout ; Un peu ; Beaucoup
5. Vous intéressez-vous au rugby ? Pas du tout ; Un peu ; Beaucoup

Informations scolaires

1. Type de baccalauréat : S ; ES ; STI ; STG ; Autre
2. Mention au baccalauréat : Passable ; AB ; B ; TB
3. Catégorie Socio-Professionnelle du chef de famille l'année d'obtention de votre baccalauréat :
 agriculteur exploitant cadre, profession intellectuelle supérieure
 profession intermédiaire employé ouvrier
 artisan, commerçant, profession libérale retraité
 autres
4. Moyenne en mathématiques l'année de Terminale (année d'obtention du baccalauréat) : ____
5. Note en mathématiques au baccalauréat : _____
6. Note en EPS au baccalauréat : _____

1.1 Population, échantillon

Dire, dans ce cadre-ci, quelle est la population. Quelle est sa taille ?

1.2 Variables

1. La question « **État civil**, 3 » est-elle une variable sur l'ensemble de la population ? Expliquer.
2. Quelles sont les modalités de la variable relative à la question « **État civil**, 1 » ? À la question « **État civil**, 3 » ? À la question « **État civil**, 5 » ? À la question « **Informations physiques**, 3 » ? À la question « **Loisirs**, 1 » ? À la question « **Loisirs**, 4 » ? À la question « **Informations scolaires**, 4 » ? À la question « **Informations scolaires**, 5 » ?
3. Pour chacune des questions du questionnaire, dire quel est le type de la variable associée ?

Les données brutes (réponses au questionnaire), mises sous la forme d'un tableau, sont fournies sur une feuille à part. Conservez ce tableau pour **l'ensemble des séances de cours** dans lesquelles nous l'utiliserons en guise d'illustration.

1.3 Mise en forme des données

1.3.1 Nettoyage des données

Quelles variables contiennent des réponses :

- inhomogènes ;
- inadaptées (aberrantes) ;
- manquantes ?

Si possible, proposez une explication aux phénomènes observés (et dites quelles peuvent être leurs conséquences dans le cadre d'une étude statistique). Enfin, lorsque c'est possible, corrigez ces valeurs.

1.3.2 Variables quantitatives continues

On s'intéresse à la variable « Taille ».

1. À partir des données brutes, proposer une répartition en classes de mêmes amplitudes.
2. À partir de cette répartition, construire un tableau d'effectifs.
3. Cette répartition en classes vous paraît-elle pertinente ? Justifier.
4. Proposer une autre répartition en classes de tailles inégales.

1.3.3 Représentativité

On s'intéresse à la variable « Catégorie socio-professionnelle du chef de famille ».

1. Faire le tableau de fréquences de cette variable.

2. Comparer-le à celui-ci fourni par l'INSEE sur la population française (données 2005)¹ :

Agriculteurs exploitants	1,6 %
Cadres, professions intellectuelles supérieures	8,9 %
Professions intermédiaires	14,0 %
Employés	17,6 %
Ouvriers	14,6 %
Artisans, commerçants, professions libérales	3,6 %
Retraités	39,2 %
Autres	0,4 %

Peut-on considérer, a priori, que cet échantillon est représentatif de la population française en ce qui concerne la catégorie socio-professionnelle du chef de famille ?

3. Même question si on s'intéresse uniquement à la population active (c'est-à-dire si on ne tient pas compte des retraités). Si l'échantillon n'est pas représentatif, citer des catégories socio-professionnelles sur et sous-représentées. Proposer éventuellement des explications.

Quelle est la conséquence de la représentativité ou de la non représentativité de cet échantillon ?

1.3.4 Variables quantitatives discrètes

On s'intéresse à la variable « Note en mathématiques au baccalauréat ».

1. Construire un tableau d'effectifs, de fréquences et de fréquences cumulées de cette variable (Attention à la définition de la population!).
2. Quelle est la modalité en dessous de laquelle on trouve la moitié des étudiants de la classe ?
3. Le tableau construit à la question 1 permet-il de donner une bonne vision de la répartition des notes en mathématiques dans cette classe ? Proposer un regroupement en classes et faire le tableau d'effectifs associé.

1.3.5 Réponses ouvertes

On s'intéresse à la variable « Lieu de visite souhaité ». Cette question est dite *ouverte*, c'est-à-dire qu'on peut avoir des réponses de niveaux différents (ville, lieu dit, région, pays...).

1. Proposer un regroupement thématique des réponses et, à partir de celui-ci, construire le tableau d'effectifs.
2. Est-il judicieux de calculer les effectifs cumulés pour cette variable ? Justifier votre réponse.

1. Source : <http://www.insee.fr>

2 Représentations graphiques

2.1 Variables qualitatives

2.1.1 Diagrammes en colonnes

1. Effectuer le diagramme en colonnes de la variable « Lieu de visite souhaité » d'après le regroupement thématique précédemment effectué dans la question 1.3.5.
2. Construire le tableau d'effectifs et d'effectifs cumulés de la variable « Mention au baccalauréat ». En déduire, le diagramme en colonne des effectifs cumulés de cette variable.

2.1.2 Diagrammes en barre

Effectuer le diagramme en barre de la variable « Lieu de visite souhaité » d'après le regroupement thématique effectué dans la question 1.3.5.

2.1.3 Diagrammes circulaires

Effectuer le diagramme circulaire de la variable « Lieu de visite souhaité » d'après le regroupement thématique effectué dans la question 1.3.5.

2.2 Variables quantitatives discrètes

2.2.1 Diagrammes en bâtons

Effectuer le diagramme en bâtons des effectifs et le diagramme en bâtons des effectifs cumulés de la variable « Âge » après avoir effectué son tableau d'effectifs.

2.2.2 Diagrammes cumulatifs

Effectuer le diagramme cumulatif de la variable « Âge ».

2.2.3 Séries chronologiques

Voici la série chronologique du nombre d'étudiants inscrits en IUT de 1990 (année scolaire 1990/1991) à 2005 (année scolaire 2005/2006) :

Année	1990	1995	1998	1999	2000	2001
Nb étudiants IUT	74 328	103 092	114 587	117 407	119 244	118 043
Année	2002	2003	2004	2005		
Nb étudiants IUT	115 465	113 722	112 395	112 597		

Proposer et effectuer deux représentations graphiques de cette série chronologique.

2.3 Variables quantitatives continues

2.3.1 Histogrammes

1. Effectuer l'histogramme et le polygone des effectifs de la variable « Taille » selon les deux regroupements en classes effectués dans la question 1.3.2 page 7.
2. Effectuer l'histogramme des effectifs cumulés de la variable « Taille » selon le regroupement en classes 2.

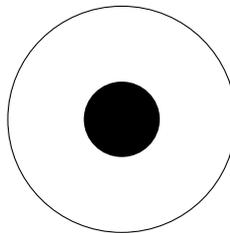
2.3.2 Polygones cumulatifs

Effectuer le polygone cumulatif de la variable « Taille » selon les deux regroupements en classes effectués dans la question 1.3.2 page 7.

2.4 Autres représentations

2.4.1 Diagrammes figuratifs

Utiliser la figure suivante :



pour représenter les effectifs de chacune des modalités de la variable « Couleur des yeux ». On prendra comme échelle pour les rayons intérieur et extérieur, respectivement, 3 cm pour 10 individus et 5 cm pour 10 individus.

2.4.2 Cartes

Utiliser ce fond de carte pour représenter, par région, la variable « Lieu de naissance » :



3 Caractéristiques numériques

Pour les exercices ci-dessous, écrivez les réponses aux questions (valeur de la moyenne, médiane, variance...) à côté de la question pour utilisation postérieure future éventuelle.

3.1 Caractéristiques de tendance centrale

3.1.1 Mode

1. Déterminer le mode de la variable « Couleur des yeux » dont le tableau d'effectifs a été construit dans la question 2.4.1 page 10.
2. Déterminer la classe modale de la variable « Taille » pour les deux regroupements en classes déterminés dans la question 1.3.2 page 7.

3.1.2 Moyenne

1. Déterminer la moyenne de la variable « Âge » dont le tableau d'effectifs a été calculé dans la question 2.2.1 page 2.2.1. Recalculer cette moyenne en supprimant de la population l'individu singulier. Commenter la différence.
2. Déterminer la moyenne de la variable « Taille » directement puis selon les deux regroupements en classes proposés à l'exercice 3.1.1. Commenter la différence.

3.1.3 Médiane

1. Déterminer numériquement et graphiquement la médiane de la variable « Mention au Baccalauréat » dont le tableau d'effectifs et le diagramme en colonnes cumulatif ont été réalisés dans la question 2.1.1 page 9.
2. Déterminer la médiane de la variable « Taille » numériquement, directement (à partir des données initiales) puis selon le **Regroupement 2** en classes donné à l'exercice 3.1.1 dont le tableau d'effectifs cumulés a été déterminé dans l'exercice 2.3.2 page 10.

3.2 Caractéristiques de dispersion

3.2.1 Étendue

Quelle est l'étendue de la variable « Âge » dont le tableau d'effectifs a été calculé dans l'exercice 2.2.1 page 9.

3.2.2 Variance et écart type

1. Déterminer l'écart type de la variable « Âge » dont le tableau d'effectifs est rappelé dans l'exercice 2.2.1 page 9.
Recalculer cet écart type en supprimant de la population l'individu singulier.
2. Déterminer l'écart type de la variable « Taille » directement puis selon les deux regroupements en classes proposés à l'exercice 3.1.1. Commenter la différence.
3. Déterminer (directement) la moyenne et l'écart type de la variable « Pointure ». En déduire son coefficient de variation et le comparer au coefficient de variation de la variable « Taille ». Commenter.
4. Déterminer directement la moyenne et l'écart type des variables « Note en mathématiques au baccalauréat » et « Note en EPS au baccalauréat ». Au vu de ces résultats, peut-on dire qu'il était plus facile d'avoir la moyenne en mathématiques ou bien la moyenne en EPS ? Était-il plus facile d'avoir 10 en mathématiques ou bien en EPS ?

3.2.3 Quantiles, quartiles

1. Déterminer numériquement et graphiquement les quartiles de la variable « Mention au Baccalauréat » dont le tableau d'effectifs et le diagramme en colonnes ont été déterminés dans l'exercice 2.1.1 page 9.
2. Déterminer les quartiles de la variable « Pointure ». En déduire le coefficient interquartile de cette variable.
3. Déterminer numériquement et graphiquement les quartiles de la variable « Taille » selon le regroupement en classes de l'exercice 3.1.3. En déduire la boîte à moustaches de cette variable ainsi que son coefficient interquartile. Comparer ce dernier au coefficient interquartiles de la variable « Pointure ».

3.3 Forme d'une distribution

3.3.1 Comparaison Moyenne / Médiane

En comparant la moyenne et la médiane de la variable « Pointure », en déduire la forme de la distribution de cette variable. Confirmer ces conclusions en construisant le diagramme en bâtons de cette variable.

3.3.2 Coefficient d'asymétrie

Calculer le coefficient d'asymétrie de la variable « Taille » selon le regroupement en classes proposé dans l'exercice 3.1.3. Commenter et comparer avec l'histogramme déterminé dans l'exercice 2.3.2 de la page 10.

3.3.3 Coefficient d'aplatissement

Calculer le coefficient d'aplatissement de la variable « Taille » selon le regroupement en classes proposé dans l'exercice 3.1.3. Commenter et comparer avec l'histogramme de cette répartition en classes déterminé dans l'exercice 2.3.2 de la page 10.

3.3.4 Courbe de Lorenz et indice de Gini

- Effectuer le tableau d'effectifs de la variable « Nombre de frères et sœurs ».
- En déduire la courbe de Lorenz de cette variable ainsi que son indice de Gini. Commenter.
- Que deviendraient ces valeurs si chacun de vous se retrouvait avec un frère ou une sœur de plus ?
- Que deviendraient ces valeurs si chacun de vous se retrouvait avec le double de frères et sœurs ?
- Que peut-on déduire des trois courbes de Lorenz et des trois indices de Gini ainsi calculés ?

4 Réponses

4.1 Généralités sur la statistique descriptive

4.1.1 Variable « Taille » (Question 1.3.2 page 7)

Tableau d'effectifs : classes de même amplitude

Tableau d'effectifs : autres classes

4.1.2 Variable « CSP du chef de famille » (Question 1.3.3 page 7)

Tableau d'effectifs sur l'ensemble de la population

Tableau d'effectifs restreint aux actifs**4.1.3 Variable « Note en mathématiques au baccalauréat » (Question 1.3.4 page 8)****Tableau d'effectifs, de fréquences, de fréquences cumulées****Tableau d'effectifs du regroupement en classes**

4.2 Représentations graphiques

4.2.1 Variable « Mention au baccalauréat » (Question 2.1.1 page 9)

Tableau d'effectifs et d'effectifs cumulés

Diagramme cumulatif en colonnes

4.2.2 Variable « Âge » (Question 2.2.1 page 9)**Tableau d'effectifs et d'effectifs cumulés****4.2.3 Variable « Taille » (Question 2.3.2 page 10)****Tableau d'effectifs cumulés, histogramme et polygone cumulatif : classes de même amplitude**

Tableau d'effectifs cumulés, histogramme et polygone cumulatif : autre regroupement en classes**4.2.4 Variable « Couleur des yeux » (Question 2.4.1 page 10)****Tableau d'effectifs**

4.2.5 Variable « Lieu de naissance » (Question 2.4.2 page 10)

Tableau d'effectifs

4.2.6 Variable « Nombre de frères et sœurs » (Question 3.3.4 page 15)

Tableau d'effectifs et indice de Gini