



Nathalie Villa-Vialaneix

Illustration du cours de
Statistique Descriptive II (M1201)

Année scolaire 2013/2014



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

Table des matières

1 Généralités sur la statistique descriptive bivariée	5
1.1 Mise en forme des données	5
1.2 Distributions marginales	5
1.3 Distributions conditionnelles	5
2 Représentations graphiques	7
2.1 Représentation graphique de deux variables qualitatives	7
2.2 Représentation graphique d'une variable qualitative et d'une variable quantitative	7
2.3 Représentation graphique de deux variables quantitatives	7
3 Étude de la liaison entre une variable qualitative et une variable quantitative	9
3.1 Statistiques conditionnelles	9
3.2 F de Fischer et rapport de corrélation	9
4 Étude de la liaison entre deux variables qualitatives	11
4.1 Indépendance de deux variables	11
4.2 Effectifs théoriques d'indépendance	11
4.3 Indice du χ^2 et C de Cramer	11
5 Régression linéaire	13
5.1 Covariance et coefficient de corrélation linéaire	13
5.2 Droite de régression de Y en X	13
6 Réponses	15
6.1 Généralités sur la statistique descriptive bivariée	15
6.2 Représentations graphiques	18

1 Généralités sur la statistique descriptive bivariée

1.1 Mise en forme des données

1. Effectuer la table de contingence des variables « Couleur des yeux » et « Couleur des yeux de la mère ». Que vaut n_{23} ?
2. Déterminer les fréquences conjointes des variables « Couleur des yeux » et « Couleur des yeux de la mère ». Que vaut f_{23} ?

1.2 Distributions marginales

Déterminer les distributions marginales des variables « Couleur des yeux » et « Couleur des yeux de la mère ». Que vaut $n_{.3}$, $n_{2.}$?

1.3 Distributions conditionnelles

1. Déterminer les distributions de la variable X : « Couleur des yeux » conditionnellement aux différentes modalités de la variable Y : « Couleur des yeux de la mère ».
2. Que vaut $f_{\text{Bleu}}^{[Y = \text{Bleu}]}$? Interprétez cette valeur.

2 Représentations graphiques

2.1 Représentation graphique de deux variables qualitatives

1. Représenter la distribution conjointe des variables « Couleur des yeux » et « Couleur des yeux de la mère ».
2. On considère les variables : « Sexe » et « Avoir eu une mention (supérieure ou égale à AB) ».
 - a) Déterminer la table de contingence de ces deux variables.
 - b) Effectuer le diagramme en barre de la distribution de la « Mention » conditionnellement aux modalités de la variable « Sexe ».
Même question pour la distributions du « Sexe » conditionnellement aux modalités de la variable « Mention ».
Quel graphique permet de savoir si les filles ont mieux réussi que les garçons ?
Conclusion ?
 - c) Effectuer les diagrammes circulaires parallèles de la distribution de « Mention » conditionnellement aux modalités de la variable « Sexe ».
3. Effectuer le diagramme en triangle de la distribution de la variable « Couleur des yeux » conditionnellement aux modalités de la variable « Couleur des yeux de la mère ». Interpréter.

2.2 Représentation graphique d'une variable qualitative et d'une variable quantitative

Représentez, sur deux diagrammes à la même échelle :

- la distribution conjointe des variables « Sexe » et « Note en mathématiques au baccalauréat » ;
- la distribution conjointe des variables « Sexe » et « Note en EPS au baccalauréat ».

Conclusion ?

2.3 Représentation graphique de deux variables quantitatives

Représentez la distribution conjointe des variables « Note en mathématiques au baccalauréat » et « Note en EPS au baccalauréat ». Utilisez deux couleurs différentes pour représenter les modalités correspondant à des filles et les modalités correspondant à des garçons.

3 Étude de la liaison entre une variable qualitative et une variable quantitative

3.1 Statistiques conditionnelles

Soient X et Y , respectivement, les variables « Note en mathématiques au baccalauréat » et « Sexe ».

1. Donnez, pour chacune des deux sous populations

$$\mathcal{P}_1 = \{Y = \text{« Homme »}\} \quad \mathcal{P}_2 = \{Y = \text{« Femme »}\}$$

les valeurs de X .

2. Pour les deux populations définies à la question précédente, calculer les statistiques conditionnelles :

	\mathcal{P}_1	\mathcal{P}_2	\mathcal{P}
N_j			
\bar{X}_j			
σ_j			

3. Déterminer la variance intra-classes et la variance inter-classes. Vérifier, sur cet exemple, la relation fondamentale de décomposition de la variance.
4. Déterminer, pour les variables X' , « Note en EPS au baccalauréat » et Y , la variance intra-classes et la variance inter-classes. Interpréter les différences entre ces valeurs et celles obtenues à la question précédente.

3.2 F de Fischer et rapport de corrélation

Les variables X , X' et Y sont les mêmes que dans l'exercice précédent.

1. Déterminer, pour les couples (X, Y) et (X', Y) , la valeur de F .
2. Déterminer, pour les couples (X, Y) et (X', Y) , la valeur du rapport de corrélation. Interpréter.

4 Étude de la liaison entre deux variables qualitatives

4.1 Indépendance de deux variables

Soient X et Y les variables, respectivement, « Sexe » et « Avoir eu une mention (ou non) ». Dans ce cas, peut-on dire qu'il y a indépendance entre les deux variables? Que peut-on dire concrètement, à partir de ces deux tableaux, sur la liaison entre les deux variables?

4.2 Effectifs théoriques d'indépendance

1. Que valent les effectifs théoriques d'indépendance des variables « Sexe » et « Avoir eu une mention (ou non) »?
2. Même question pour les variables « Couleur des yeux » et « Couleur des yeux de la mère ». Quelles paires de modalités sont sous-représentées? sur-représentées?

4.3 Indice du χ^2 et C de Cramer

1. Déterminer les contributions au χ^2 pour la table de contingence des variables « Sexe » et « Avoir eu une mention ».
2. Déterminer les contributions au χ^2 pour la table de contingence des variables « Couleur des yeux » et « Couleur des yeux de la mère ». Quelles paires de modalités contribuent le plus au χ^2 ? Interpréter.
3. Déterminer la valeur du χ^2 pour les deux tables de contingence de l'exercice précédent.
4. Déterminer la valeur du C de Cramer pour les deux tables de contingence de l'exercice précédent. Interprétation?

5 Régression linéaire

5.1 Covariance et coefficient de corrélation linéaire

On s'intéresse ici aux variables X , « Note en mathématiques au baccalauréat » et Y , « Note en EPS au baccalauréat ». Pour information, on rappelle le nuage de points de ces deux variables :

1. Calculez la covariance, $\text{Cov}(X, Y)$.
2. Déterminez le coefficient de corrélation linéaire, $r(X, Y)$. Interprétez. Pouvait-on, connaissant le nuage de points, s'attendre à ce résultat ?
3. Soit X' la variable « Moyenne en mathématiques en terminale ».
 - a) Effectuez un graphique similaire au nuage de points des variables X et Y pour les variables X' et X .
 - b) Déterminez le coefficient de corrélation linéaire, $r(X', X)$. Interprétez. Pouvait-on prévoir ce résultat ?

5.2 Droite de régression de Y en X

1. Déterminez l'équation de la droite de régression de Y en X et celle de la droite de régression de X en X' .
2. Une de ces deux droites permet-elle de faire des prévisions raisonnables ? Déterminez la valeur prévue de la note en mathématiques au baccalauréat pour un élève ayant 12 de moyenne annuelle dans cette matière. Commentaires ?

6 Réponses

6.1 Généralités sur la statistique descriptive bivariée

Table de contingence des variables « Couleur des yeux » et « Couleur des yeux de la mère » (Question 1.1 page 5)

Distribution de la variable « Couleur des yeux » conditionnellement aux modalités de la variable « Couleur des yeux de la mère » (Question 1.3 page 5)

6.2 Représentations graphiques

Table de contingence des variables « Mention au baccalauréat » et « Sexe » (Question 2.1 page 7)

Distribution de la variable « Mention au baccalauréat » conditionnellement aux modalités de la variable « Sexe » (Question 2.1 page 7)

Distribution de la variable « Sexe » conditionnellement aux modalités de la variable « Mention au baccalauréat (Question 2.1 page 7)