

# Croisement d'une variable quantitative et d'une variable qualitative

Anas KNEFATI

Université Rennes 2



- 1 Données
- 2 Représentation graphique - Distribution Conditionnelle
- 3 Formules de décompositions
- 4 Rapport de corrélation

- 1 Données
- 2 Représentation graphique - Distribution Conditionnelle
- 3 Formules de décompositions
- 4 Rapport de corrélation

## Présentation des données

- $X$  : Variable qualitative avec  $\ell$  modalités :  $x_1, x_2, \dots, x_\ell$
- $Y$  : Variable quantitative (discrète ou continue) avec  $c$  classes.

## Notes des étudiants pour trois groupes

Note \ Groupe	A	B	C
	13	1	18.7
2	1.5	14.5	
7.8	15	3	
19	12.5	11	
13.5	18.2	15	
8.5	20	7.5	
18.5	9.3	20	
		19	

- 1 Données
- 2 Représentation graphique - Distribution Conditionnelle
- 3 Formules de décompositions
- 4 Rapport de corrélation

# Représentation graphique - Distribution Conditionnelle

## Représentation graphique

On peut tracer pour chaque classe :

- l'histogramme de  $Y$  (ou le diagramme en bâton selon le type de  $Y$ )
- le "box-plot" de  $Y$ .

## Distribution conditionnelle

La moyenne (ou la variance) conditionnelle de  $Y$  sachant que l'on est dans la  $i^{\text{ème}}$  classe de  $X$  : C'est la moyenne (ou la variance) de  $Y$  dans cette classe

## Notations

- $n_i$  : Effectif total de  $Y$  dans la  $i^{\text{ème}}$  classe de  $X$
- $n$  : Effectif total de  $Y$  (toutes classes confondues,  $n = \sum_{i=1}^c n_i$ )
- $\bar{y}$  : Moyenne globale de  $Y$  et  $\sigma^2$  : Variance globale de  $Y$
- $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_\ell$  : Moyennes conditionnelles de  $Y$
- $\sigma_1^2, \sigma_2^2, \dots, \sigma_\ell^2$  : Ses variances conditionnelles

# Exemple :

Note \ Groupe	Groupe		
	A	B	C
	13	1	18.7
	2	1.5	14.5
	7.8	15	3
	19	12.5	11
	13.5	18.2	15
	8.5	20	7.5
	18.5	9.3	20
		19	

	A	B	C
min	2	1	3
Q <sub>1</sub>	7.8	1.5	7.5
Q <sub>2</sub>	13	13.75	14.5
Q <sub>3</sub>	18.5	18.2	18.7
max	19	20	20

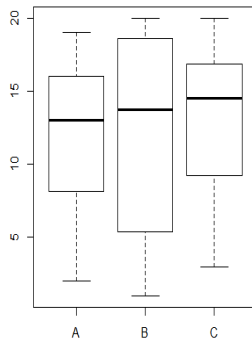


Figure: Box-plots

# Exemple

Notes des étudiants pour trois groupes - Y : Note et X : Groupe

Note \ Groupe	A	B	C	pop. globale
	13	1	18.7	
	2	1.5	14.5	
	7.8	15	3	
	19	12.5	11	
	13.5	18.2	15	
	8.5	20	7.5	
	18.5	9.3	20	
		19		
Effectifs	7	8	7	22
Moyennes $\bar{y}_i$ conditionnelles	11.8	12.1	12.8	$\bar{y} = 12.2$
Variance $\sigma_i^2$ conditionnelles	32	49.9	31.7	$\sigma^2 = 38.6$



# Plan

- 1 Données
- 2 Représentation graphique - Distribution Conditionnelle
- 3 Formules de décompositions
- 4 Rapport de corrélation

# Formules de décompositions

## Formule de la moyenne

$$\begin{array}{lcl} \bar{y} & = & \frac{1}{n} \sum_{i=1}^c n_i \bar{y}_i \\ \text{Moyenne totale} & = & \text{Moyenne des représentants (c-à-d : } \bar{y}_i) \end{array}$$

## Formule de la variance

$$\begin{array}{lcl} \sigma^2 & = & \frac{1}{n} \sum_{i=1}^c n_i \sigma_i^2 + \frac{1}{n} \sum_{i=1}^c n_i (\bar{y}_i - \bar{y})^2 \\ \text{Variance totale} & = & \text{Variance intra-groupe} + \text{Variance inter-groupe} \end{array}$$

## Notations

- $V_{\text{intra}} = \frac{1}{n} \sum_{i=1}^c n_i \sigma_i^2$  : Variance intra-groupe qui regard la dispersion à l'intérieur de chaque groupe.
- $V_{\text{inter}} = \frac{1}{n} \sum_{i=1}^c n_i (\bar{y}_i - \bar{y})^2$  : Variance inter-groupe qui regard la dispersion des représentants.
- On a donc :  $\sigma^2 = V_{\text{intra}} + V_{\text{inter}}$

- 1 Données
- 2 Représentation graphique - Distribution Conditionnelle
- 3 Formules de décompositions
- 4 Rapport de corrélation

## Définition

- Il s'agit d'un indice de liaison entre  $X$  et  $Y$
- Il est noté comme  $\eta^2_{Y/X}$  ou simplement  $\eta^2$
- Le Rapport de corrélation est défini comme la part de  $V_{\text{inter}}$  dans la variance totale :

$$\eta^2 = \frac{V_{\text{inter}}}{V_{\text{totale}}} = \frac{V_{\text{inter}}}{\sigma^2}$$

## Propriétés et interprétation

- $\eta^2$  n'est pas symétrique et il est sans unité.
- $0 \leq \eta^2 \leq 1$
- $\eta^2 = 0$  : Pas de liaison entre  $X$  et  $Y$ .
- $\eta^2 = 1$  : Il y a une liaison totale entre  $X$  et  $Y$
- Plus  $\eta^2$  est grand, plus la liaison entre la variable qualitative  $X$  et la variable quantitative  $y$  est forte.

## Exemple :

- $V_{\text{intra}} = \frac{1}{n} \sum_{i=1}^c n_i \sigma_i^2 = 38.41$
- $V_{\text{inter}} = \frac{1}{n} \sum_{i=1}^c n_i (\bar{y}_i - \bar{y})^2 = 0.19$
- $V = V_{\text{intra}} + V_{\text{inter}} = 38.41 + 0.19 = 38.6$
- $\eta^2 = \frac{V_{\text{inter}}}{V_{\text{totale}}} = \frac{V_{\text{inter}}}{\sigma^2} = \frac{0.19}{38.6} = 0.005$
- Comme  $\eta^2$  est presque nul, alors les notes des étudiants ne changent pas trop selon groupe