Université Toulouse Le Mirail - 1ère Session - Année 2011/2012

CORRECTION

UE6 PY0106X - PARTIE STATISTIQUE - Mai 2012

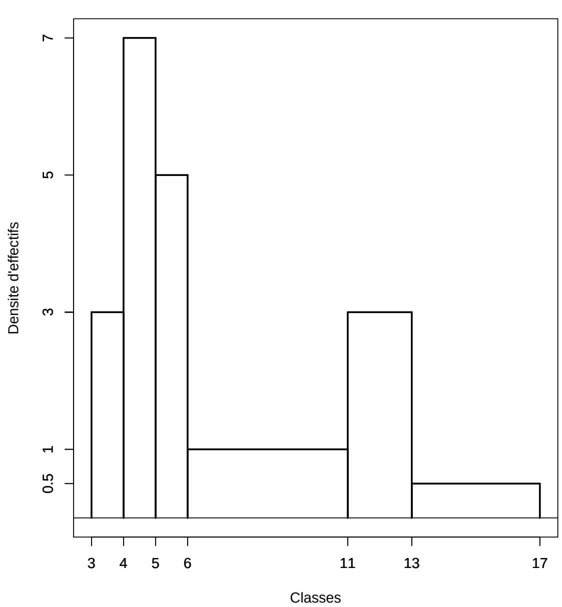
Exercice 1

1) Quelle est la population étudiée? Quelle est la taille de l'échantillon? Préciser la variable étudiée ainsi que son type en détaillant l'ensemble des valeurs possibles :
Population = élèves de CM2 ; Taille de l'échantillon = 28 ; Variable $\boldsymbol{X}$ ="temps de mémorisation" ; valeurs possibles = tous les nombre positifs donc variable quantitative continue.

2)

x_i	n_i	c_i	$n_i \times c_i$	$n_i \times (c_i)^2$	N_i **(eff. cum.)**	$d_i = n_i/a_i$
[3;4[	**3**	**3.5**	**10.5**	**36.75**	**3**	**3**
[4;5[	**7**	**4.5**	**31.5**	**141.75**	**10**	**7**
[5;6[	**5**	**5.5**	**27.5**	**151.25**	**15**	**5**
[6;11[	**5**	**8.5**	**42.5**	**361.25**	**20**	**1**
[11;13[	**6**	**12**	**72**	**864**	**26**	**3**
[13;17[	**2**	**15**	**30**	**450**	**28**	**0.5**
TOTAL	$\boldsymbol{N = 28}$		$\boldsymbol{S_1 = 214}$	$S_2 = 2005$		

3) Représenter la distribution de X ; en déduire le(s) mode(s).



On constate 2 "pics" qui correspondent aux 2 classes modales $[4;5[$ et $[11;13[$. On a donc 2 modes obtenus à partir des centres des 2 classes : 4.5 et 12

4) Calculer la moyenne, la variance et l'écart-type de X.

- $\overline{\boldsymbol{X}} = \frac{\mathbf{214}}{\mathbf{28}} \simeq \mathbf{7.64}$
- $\boldsymbol{Var(X)} = \frac{\mathbf{2005}}{\mathbf{28}} - \mathbf{(7.64)^2} \simeq \mathbf{71.61 - 58.37 = 13.24}$
- $\boldsymbol{\sigma_X} \simeq \sqrt{\mathbf{13.24}} \simeq \mathbf{3.64}$

5) Déterminer la médiane ; que représente cet indice? Expliquez brièvement pourquoi il y a une différence non négligeable entre la moyenne et la médiane.

- **Puisque 28/2=14, on s'intéresse à l'effectif cumulé $\boldsymbol{N_*}$ immédiatement supérieur à 14 soit $\boldsymbol{N_*} = \mathbf{15}$. La classe médiane est donc $[\mathbf{5;6}[$ et la médiane est donnée par son centre : med($\boldsymbol{X}$)=5.5**
- **Cet indice est la valeur de $\boldsymbol{X}$ qui sépare l'échantillon en 2 sous-échantillons de même taille : 50% de l'échantillon prend une valeur plus petite (et donc 50% une valeur plus grande)**
- **En regardant l'histogramme, on constate que les observations se répartissent autour de 2 modes : 1 mode principal autour de 4.5 et 1 mode secondaire à 12. Il est donc normal de trouver une médiane proche du mode principal. En revanche, le mode secondaire impacte la moyenne qui est plus sensible que la médiane aux valeurs "extrêmes". Ceci explique pourquoi on obtient une moyenne plus grande que la médiane.**

6)

6.a) Quel groupe d'élève paraît le plus homogène? Que peut-on en déduire du point de vue de la variance? Pour quel groupe d'élèves la moyenne est-elle la plus représentative?

- **Le groupe 2 est très concentré autour des classes $[\mathbf{6;6.5}[$ et $[\mathbf{6.5;7}[$ alors que que le groupe 1 se disperse sur une étendue beaucoup plus importante ; le groupe 2 est donc plus homogène.**
- **Le groupe 2 étant plus homogène, il possède une variance plus petite que celle du groupe 1.**
- **La moyenne est plus représentative pour le groupe de plus petite variance, donc pour le groupe 2.**

6.b) Pour quel groupe d'élèves l'écart entre la moyenne et la médiane est-il le plus grand?
Les calculs précédents montrent que le groupe 1 présente un écart marqué entre la moyenne et la médiane. De plus, la répartition des observations pour le groupe 2 étant plus homogène (plus concentrée) que celle du groupe 1, l'écart entre moyenne et médiane sera petit ; c'est donc le groupe 2 qui possède le plus petit écart moyenne/médiane.

Exercice 2

1. Préciser les variables étudiées ainsi que leur type. Quelle est la population étudiée? Quelle est la taille de l'échantillon?

   - **La variable X =“assiduité” prend les 2 modalités “assidu” et “non assidu”. Il s'agit donc d'une variable qualitative nominale.**
     La variable Y =“niveau des résultats” prend les 3 modalités “faible”, “moyen” et “bon”. Il s'agit donc d'une variable qualitative ordinale.
   - **Population = étudiants inscrits en 1ère année.**
   - **Taille de l'échantillon $\boldsymbol{N} = 14 + 42 + 97 + 103 + 51 + 26 = 333$.**

2.

X \ Y	faible	moyen	bon	**Marge de $\boldsymbol{X}$**
Assidu	14	42	97	**153**
Non assidu	103	51	26	**180**
Marge de $\boldsymbol{Y}$	**117**	**93**	**123**	**$\boldsymbol{N} = 333$**

3. Déterminer la distribution de Y conditionnellement à X. Représenter graphiquement cette distribution. Que pouvez-vous dire concernant le lien entre X et Y pour cet échantillon?

   -

X \ Y	faible	moyen	bon	TOTAL
Assidu	**9.15**	**27.45**	**63.4**	100
Non assidu	**57.22**	**28.33**	**14.45**	100

Par exemple :

$$\mathbf{9.15} \simeq \frac{\mathbf{14}}{\mathbf{153}}$$

$$\mathbf{27.45} \simeq \frac{\mathbf{42}}{\mathbf{153}}$$

$$\vdots$$

$$\mathbf{14.45} \simeq \frac{\mathbf{26}}{\mathbf{180}}$$

•

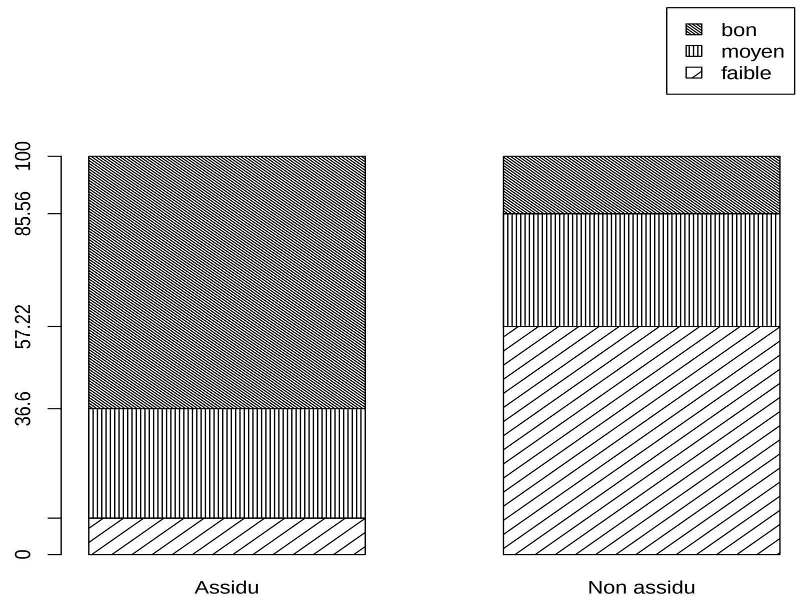


Figure 1: **Distribution de Y conditionnellement à X.**

- **On constate un changement important dans la distribution de Y lorsqu'on passe de la modalité "Assidu" à la modalité "Non assidu" de X. Cela indique qu'il y a un lien entre X et Y pour cet échantillon.**

4. **Tableau des effectifs théoriques**

X \ Y	faible	moyen	bon
Assidu	53.76	42.73	56.51
Non assidu	63.24	50.27	66.49

Exemple :

$$\mathbf{53.76 \simeq \frac{153 \times 117}{333}}$$

$$\mathbf{42.73 \simeq \frac{153 \times 93}{333}}$$

$$\vdots$$

Tableau des contributions

X \ Y	faible	moyen	bon
Assidu	29.41	0.01	29.01
Non assidu	25	0.01	24.66

Exemple :

$$\mathbf{29.41 \simeq \frac{(14 - 53.76)^2}{53.76}}$$

$$\mathbf{0.01 \simeq \frac{(42 - 42.73)^2}{42.73}}$$

$$\vdots$$

5. $\boldsymbol{\chi^2 = 29.41 + 0.01 + 29.01 + 25 + 0.01 + 24.66 = 108.1}$

6. En déduire le V de Cramér ϕ_c. Que peut-on dire de l'intensité du lien entre X (assiduité) et Y (niveau des résultats) pour cet échantillon?

   - $\boldsymbol{\phi_c = \sqrt{\frac{108.1}{333 \times \{\min(2,3) - 1\}}} = \sqrt{\frac{108.1}{333}} \simeq 0.57}$
   - **Puisque $\boldsymbol{0.5 < \phi_c \leq 1}$, l'intensité du lien entre $\boldsymbol{X}$ et $\boldsymbol{Y}$ est plutôt forte pour cet échantillon**

7. <u>Question bonus (hors barème)</u>.

X \ Y	faible	moyen	bon
Assidu	14	42	84
Non assidu	103	**309**	**618**

Pour que $\boldsymbol{\phi_c = 0}$, il suffit que la distribution de $\boldsymbol{Y}$ ne change pas lorsque qu'on passe de $\boldsymbol{x_1 =}$"assidu" à $\boldsymbol{x_2 =}$"non assidu". Il faut donc que les proportions observées pour $\boldsymbol{Y}$ lorque $\boldsymbol{X = x_1}$ soit les mêmes que celles pour $\boldsymbol{Y}$ lorque $\boldsymbol{X = x_2}$. Or $\boldsymbol{14 \times 3 = 42}$ et $\boldsymbol{42 \times 2 = 84}$. C'est pourquoi, afin de respecter les mêmes proportions, on prendra 309 et 618 car $\boldsymbol{309 = 103 \times 3}$ et $\boldsymbol{618 = 309 \times 2}$.