Université de Bretagne Occidentale
U.F.R. de Lettres et Sciences Sociales
BP 814 - 29285 BREST CEDEX

Année Universitaire 2001-2002

Section: Psychologie - DEUG 2è année - UV PSY23B1

Enseignant responsable : F.G. Carpentier

Corrigé de l'épreuve partielle et ponctuelle de Statistiques appliquées à la Psychologie

Exercice 1

Les élèves de sept classes de CE2 de cinq écoles de la région parisienne ont été soumis à un test de compréhension de lecture. La distribution des scores des 165 élèves est donnée par le tableau ci-dessous :

Score	Effectif
7	3
9	2
11	2
13	8
15	7
17	10
19	8

Score	Effectif
21	22
23	32
25	25
27	26
29	14
31	6
Total	165

1) Préciser la population et la variable statistique étudiées. Quel est le type (protocole, effectifs, etc.) du tableau ci-dessus ?

La population est formée par les 165 élèves. La variable étudiée est le score au test de compréhension. C'est une variable numérique. Les observations sont rassemblées dans un tableau d'effectifs.

2) Déterminer le mode et la médiane de la variable "Score".

Le mode est la modalité d'effectif le plus élevé. Il vaut donc 23. La médiane est le score obtenu par le 83è élève. Elle vaut également 23.

3) On procède à un regroupement de la variable "Score" en 6 classes en utilisant comme bornes des classes les valeurs :

$$6 \quad 12 \quad 16 \quad 20 \quad 24 \quad 28 \quad 32$$

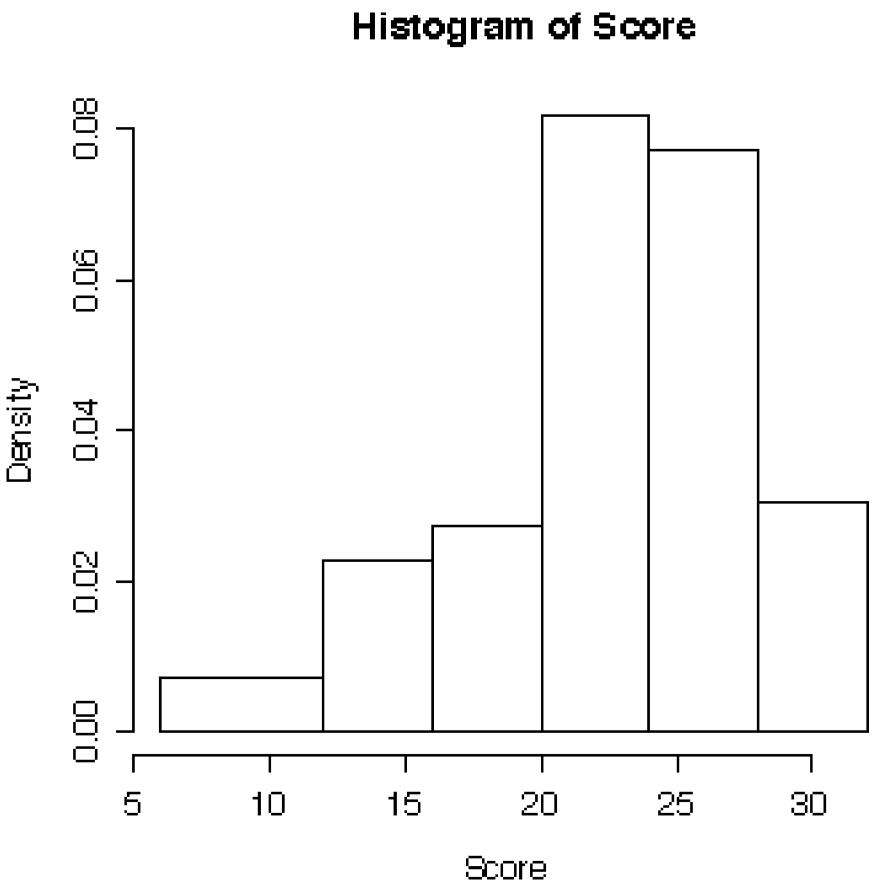
Ce regroupement en classes sera utilisé dans toute la suite de l'exercice.
a) Dresser le tableau des effectifs et celui des effectifs et fréquences cumulés.

Classes	Effec.	Amp.	Dens.	c_i	$n_i c_i$	$n_i c_i^2$
$[6, 12[$	7	6	1.17	9	63	567
$[12, 16[$	15	4	3.75	14	210	2940
$[16, 20[$	18	4	4.5	18	324	5832
$[20, 24[$	54	4	13.5	22	1188	26136
$[24, 28[$	51	4	12.75	26	1326	34476
$[28, 32[$	20	4	5	30	600	18000
Total	165				3711	87951

x	Ef. Cumul.	Fr. Cumul.
6	0	0%
12	7	4.2%
16	22	13.3%
20	40	24.2%
24	94	56.9%
28	145	87.9%
32	165	100%

b) Représenter la variable “Score” à l'aide d'un histogramme.

Observez le traitement réservé à la première classe, dont l'amplitude est différente de celle des autres classes. Notez également que la “density” indiquée par le logiciel ne correspond pas à la densité calculée ci-dessus : c'est en fait la densité, divisée par l'effectif total.



c) Calculer la moyenne, la variance et l'écart type de la variable étudiée.

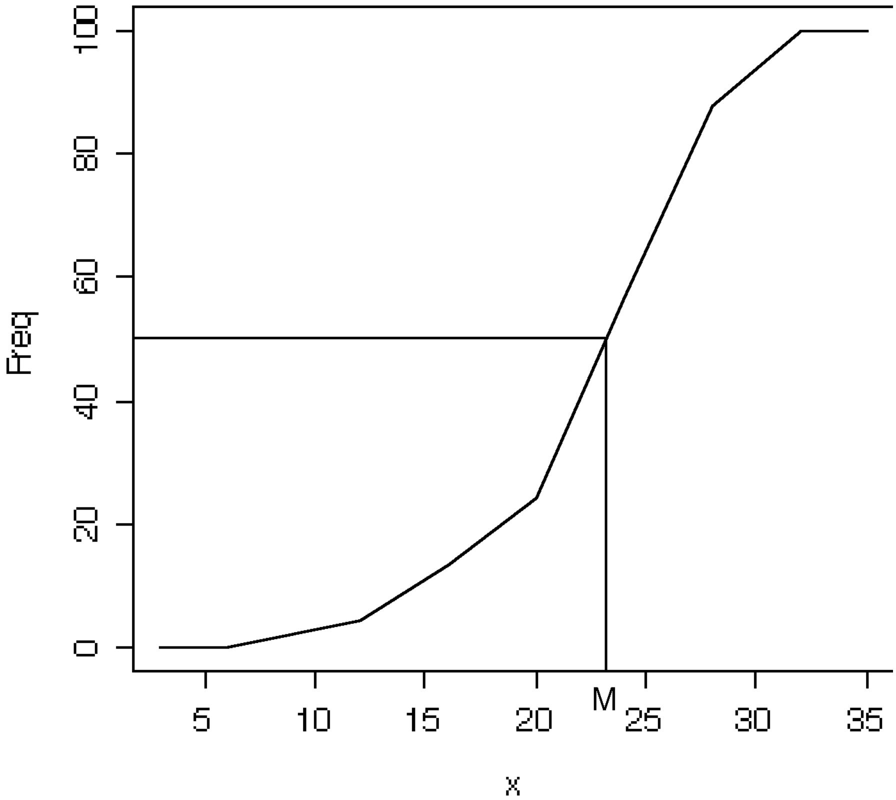
Moyenne : $\mu = \dfrac{3711}{165} = 22.49$

Variance : $V = \dfrac{87951}{165} - 22.49^2 = 27.19$

Ecart type : $\sigma = \sqrt{27.19} = 5.21$

4) Représenter graphiquement la fonction de répartition de cette variable et évaluer à l'aide du graphique la médiane de la série regroupée en classes.

Par lecture graphique, on retrouve : $M = 23$ environ.



Exercice 2

Le nombre d'heures de sommeil, chaque nuit, des Américains varie considérablement. On estime que cette variable H est distribuée au sein de la population américaine selon une loi normale de paramètres $\mu = 6.86$ et $\sigma = 0.78$.
1) Quelle est la proportion de la population dormant moins de 6 heures par nuit ?

Soit Z la variable normale centrée réduite définie par : $Z = \dfrac{H - 6.86}{0.78}$.
$H = 6$ correspond à $Z = -1.10$. On lit dans la table : $P(Z \leq 1.10) = 0.8643$, d'où $P(H \leq 6) = 1 - 0.8643 = 0.1357$.

2) Quelle est la proportion de la population dormant plus de 8 heures ?

$H = 8$ correspond à $Z = 1.46$. On lit dans la table : $P(Z \leq 1.46) = 0.9279$, d'où $P(H \geq 8) = 1 - 0.9279 = 0.0721$.

3) Donner un intervalle $a \leq H < b$, centré autour de la moyenne et rassemblant 95% de la population (c'est-à-dire tel que $P(a \leq H < b) = 0.95$).

On sait que $P(-t \leq Z < t) = 0.95$ pour $t = 1.96$. Or, $Z = -1.96$ correspond à $H = 6.86 - 1.96 \times 0.78 = 5.33$ et $Z = 1.96$ correspond à $H = 6.86 + 1.96 \times 0.78 = 8.39$.
D'où : $P(5.33 \leq H < 8.39) = 0.95$.

Exercice 3

Dans le cadre d'une recherche sur les représentations sociales de l'allaitement, on a interrogé à l'aide d'un questionnaire 353 femmes réparties en trois groupes : mères d'enfants de moins d'un an (groupe 1), femmes enceintes de leur premier enfant (groupe 2), étudiantes sans enfants (groupe 3). L'un des items du questionnaire concernait le phénomène de la dépression post partum (après l'accouchement). Les sujets devaient indiquer leur degré d'accord (échelle de 1 à 5) avec l'affirmation : "la dépression post partum est due à la solitude et à l'insuffisance de l'accompagnement de la mère". Les résultats observés sont les suivants :

d° d'adhésion	Groupe 1	Groupe 2	Groupe 3
1	12	11	17
2	23	17	47
3	53	33	38
4	31	15	38
5	12	3	3

1) Quelles sont les variables statistiques étudiées. Quelle est la nature de chacune d'elles ? Quelle est la nature (tableau protocole, tableau d'effectifs, tableau de contingence) du tableau ci-dessus ?

On étudie ici la variable "groupe" (nominale, à 3 modalités) et la variable "d° d'adhésion" (ordinale, à 5 modalités). Le tableau proposé est un tableau de contingence.

2) On se propose d'étudier à l'aide d'un test du χ^2 si les trois groupes ont la même opinion vis-à-vis de la question posée.
a) Calculer le tableau des effectifs théoriques correspondant au tableau ci-dessus. Les conditions d'effectif minimum généralement exigées pour utiliser le test du χ^2 sont-elles respectées ?

Les marges calculées à partir du tableau d'effectifs observés et les effectifs théoriques sont donnés dans le tableau ci-dessous.

d° d'adhésion	Groupe 1	Groupe 2	Groupe 3	Total
1	14.8	9.0	16.2	40
2	32.3	19.5	35.2	87
3	46.0	27.8	50.2	124
4	31.2	18.8	34.0	84
5	6.7	4.0	7.3	18
Total	131	79	143	353

Une condition généralement exigée pour réaliser un test du χ^2 est un effectif supérieur ou égal à 5 dans chacune des cases du tableau d'effectifs théoriques. Ici, l'une des cases contient un effectif de 4. Cependant, cela ne concerne qu'une case sur 15, soit moins de 20% des cases du tableau. L'utilisation du test du χ^2 est donc légitime.

b) La valeur χ^2_{obs} observée sur l'échantillon étudié est obtenue comme somme des contributions des différentes cases du tableau.
Détailler le calcul de la contribution de la première case.

On a : $Ctr_1 = \dfrac{(12 - 14.8)^2}{14.8} = 0.53$.

c) Le calcul complet donne : $\chi^2_{obs} = 21.25$.
Déterminer le nombre de degrés de liberté. Utiliser la table du χ^2 pour déterminer au seuil de 5% si les variables sont dépendantes entre elles. Formuler une conclusion par rapport à la situation étudiée.

Le nombre de degrés de liberté est ici : $ddl = (l-1)(c-1) = 4 \times 2 = 8$. Pour un seuil de 5%, on lit dans la table : $\chi^2_{crit} = 15.507$. Comme $\chi^2_{obs} > \chi^2_{crit}$, on conclut à un lien entre le groupe d'appartenance du sujet et le degré d'adhésion choisi. Autrement dit, les trois groupes n'ont pas la même opinion vis-à-vis de la question posée.

Exercice 4

Dans une étude sur la relation entre les faits additifs et les faits multiplicatifs, on présente aux sujets une série de problèmes arithmétiques simples sur l'écran d'un ordinateur. Les sujets ont pour consigne de dire si oui ou non, le résultat présenté est correct.
Les problèmes sont de genre additif ou multiplicatif et le résultat présenté peut être "vrai" (ex: 4+3=7), "faux" (ex: 4+3=9) ou "confusion" (ex: 4+3=12) ; on définit ainsi 6 types de problèmes (additif-vrai, multiplicatif-vrai, ...)
Le comportement des sujets observés est mesuré par le temps de réaction (TR) en millisecondes devant chaque problème et par la valeur (vrai/faux) de la réponse fournie. Le tableau ci-dessous donne d'une part, la moyenne des temps de réaction des sujets et d'autre part le taux d'erreurs commises pour chaque type de problème.

Type de problèmes	TR moyen (x_i)	Taux d'erreurs (y_i)
additif-vrai	1359	5.92%
multiplicatif-vrai	1351	5.44%
additif-confusion	1447	8.84%
multiplicatif-confusion	1356	3.50%
additif-faux	1281	2.04%
multiplicatif-faux	1348	2.14%

N.B. On pourra utiliser dans les questions ci-dessous les résultats suivants :
$\sum x_i = 8\,142$; $\sum y_i = 0.278\,8$; $\sum x_i^2 = 11\,062\,692$; $\sum y_i^2 = 0.016\,377\,68$; $\sum x_i y_i = 384.301\,6$.

1) Calculer la moyenne, la variance et l'écart type de chacune des deux variables.

On a : $\mu(TR) = 1357$, $V(TR) = 1843782 - 1357^2 = 2333$, $\sigma(TR) = 48.30$.
$\mu(Tx) = 4.646\%$, $V(Tx) = 27.29 \times 10^{-4} - 0.0464^2 = 5.704 \times 10^{-4}$, $\sigma(Tx) = 0.0239 = 2.39\%$.

2) Calculer la covariance et le coefficient de corrélation.

$Cov(TR, Tx) = 64.0503 - 1357 \times 0.0464 = 1.004$. $\rho = \dfrac{1.004}{48.30 \times 0.0239} = 0.87$.
N.B. Les résultats obtenus dans cette question sont très sensibles aux arrondis faits dans les calculs intermédiaires.

3) Commenter le résultat obtenu. Un temps de réaction élevé favorise-t-il la production d'une réponse correcte ?

La corrélation entre les deux variables est assez forte. En revanche, les temps de réaction élevés sont associés à des taux d'erreurs élevés. Autrement dit, un temps de réaction élevé ne favorise pas la production d'une réponse correcte, au contraire.