

Analyse statistiques de Données en Psychologie 2 (ADP2)

Diplôme : Licence 2 - Psychologie	
Session : janvier 2010 Semestre : 1 Année universitaire 2009 / 2010	Durée de l'épreuve : 1 h 30 Coefficient : 5
Responsable : Denis CORROYER Examen final	Nombre de pages : 8

Aucun document n'est autorisé. La calculatrice n'est pas autorisée.
Les différents exercices (encadrés) sont indépendants.

Le barème, indicatif, est sur 60 ; la note finale sera donnée sur 20.

**INDIQUER LES REPONSES EXCLUSIVEMENT SUR CE DOCUMENT.
NE RIEN ECRIRE DANS LES MARGES.**

Les calculs et les graphiques ont été réalisés avec le logiciel SES-Pegase.

Pour se prononcer sur l'importance des effets on conviendra des valeurs repères suivantes : .20 et .40 pour un indice de type r , $Eta...$, 4% et 16% pour un indice de type R^2 , $Eta^2...$

1. Dossier NOBEL (18 points)

Le tableau suivant rapporte la répartition des 296 prix Nobel attribués (jusqu'en 1980) à quatre pays dans cinq disciplines.

Pays :

- USA (USA),
- Grande-Bretagne (GB),
- Allemagne de l'Ouest (RFA),
- France (FRAN).

Disciplines (type de prix Nobel) :

- Médecine (MEDE),
- Physique (PHYS),
- Chimie (CHIM),
- Littérature (LITT),
- Sciences Économiques (SECO).

Source : Rouanet, Le Roux, Bert (1987) p..33-35. D'après le journal Le Monde

A. Effectifs conjoints et marginaux (4 points)

	MEDE	PHYS.	CHIM	LITT	SECO	TOTAL
USA	55	43	24	8	9	139
GB	19	20	21	6	2	68
RFA	11	14	24	7	0	56
FRAN	7	9	6	11	0	33
TOTAL	92	86	75	32	11	296

Par exemple, la case en haut à gauche indique que 55 de ces 296 prix Nobel ont été attribués aux USA en Médecine.

1. On constate que les USA ont reçu plus de prix Nobel de Chimie (24) que la Grande Bretagne (21).

Pourquoi peut-on dire malgré tout que la GB tend à recevoir relativement plus de prix Nobel de Chimie que les USA ? Quel calcul peut étayer cette affirmation ?

Parce que les USA ont reçu 24 Nobel de Chimie sur un total de 139 alors que la GB en a reçu 21 sur un total de 68, soit relativement plus que les USA.

On pourrait vérifier que le rapport 21/68 (pour la GB) est bien supérieur à 24/139 (pour les USA).

2. On s'intéresse au nombre de prix Nobel de Médecine obtenus par la France (7). Sachant que le rapport 7/92 est égal à 8%, peut-on en conclure que 8% des prix Nobel reçus par la France étaient des prix Nobel de Médecine ? Commenter cette affirmation :

Ce commentaire est erroné. Ce pourcentage de 8% (7/92) est le pourcentage de prix Nobel attribués à la France parmi les prix Nobel de Médecine : 8% des prix Nobel de médecine ont été attribués à la France. Le pourcentage de prix Nobel de Médecine parmi les prix Nobel attribués à la France est égal au rapport 7/33 (pourcentage très supérieur à 8%).

B. Profils et profils moyens (4 points)

On a reporté ci-dessous le tableau des profils des pays et le profil moyen, obtenu à partir du tableau précédent.

	MEDE	PHYS	CHIM	LITT	SECO	Total
USA	40 %	31 %	17 %	6 %	6 %	100 %
GB	28 %	29 %	31 %	9 %	3 %	100 %
RFA	20 %	25 %	43 %	13 %	0 %	100 %
FRAN	21 %	27 %	18 %	33 %	0 %	100 %
Moyenne	31 %	29 %	25 %	11 %	4 %	100 %

1. Indiquer comment a été obtenu le pourcentage de 40% (cf. en haut à gauche) :

$$40\% = 55 / 139$$

2. Indiquer la formule classique de calcul des effectifs théoriques à partir des effectifs marginaux :

$$\hat{n}_{jk} = \frac{n_j \cdot n_k}{n}$$

3. Indiquer comment appliquer cette formule pour retrouver l'effectif théorique de la case (USA, MEDE), soit 43.2.

$$43.2 = (139 \times 92) / 296$$

4. Indiquer comment retrouver également cet effectif théorique de la case (USA, MEDE) à partir du profil moyen :

$$43.2 = 31\% \times 139$$

C. Taux de liaison (4 points)

Txl	MEDE	PHYS	CHIM	LITT	SECO
USA	+27 %	+6 %	-32 %	-47 %	+74 %
GB	-10 %	+1 %	+22 %	-18 %	-21 %
RFA	-37 %	-14 %	+69 %	+16 %	-100 %
FRAN	-32 %	-6 %	-28 %	+208 %	-100 %

1. Indiquer la formule d'un taux de liaison ou écart relatif :

$$Txl = \frac{(O - T)}{T} \quad \text{ou} \quad Txl = \frac{(n_{jk} - \hat{n}_{jk})}{\hat{n}_{jk}}$$

2. Que signifie le fait que le taux de liaison soit négatif pour la case (USA, CHIM) ?

Cela signifie que le nombre de prix Nobel de Chimie reçus par les USA est inférieur à ce qu'il aurait été s'il y avait eu une stricte indépendance entre les variables Pays (PAYS) et type de prix Nobel reçu (DISC) ou, dit autrement, s'il n'y avait pas de différence entre les pays en ce qui concerne le type de prix Nobel reçus.

3. Commenter la valeur de ce taux de liaison (USA, CHIM) :

Les USA ont reçu presque un tiers (32%) de prix Nobel de Chimie en moins par rapport à ce qu'ils auraient reçu s'il y avait eu indépendance entre ces deux variables (PAYS et DISC).

D. Écart global à l'indépendance (6 points)

Le Φ^2 observé pour ce tableau est égal à 0.15.

1. Pour se prononcer sur l'importance de l'écart à l'indépendance, il est difficile de le faire à partir de ce seul indice (Φ^2). Pourquoi ?

Parce que cet indice a un maximum qui varie selon les dimensions du tableau. Une même valeur du Φ^2 peut, selon la taille du tableau, correspondre dans un cas à un écart faible et dans un autre cas à un écart important à l'indépendance.

2. Indiquer le nom d'un autre indice qui permet de répondre à cette question :

Le V^2 de Cramer

3. Indiquer la formule de cet indice

$$V^2 = \Phi^2 / \Phi^2_{\max}$$

4. Calculer la valeur de cet indice pour cet exemple :

$$\text{De manière générale } \Phi^2_{\max} = \text{Min}(J, K) - 1$$

$$\text{Ici } \Phi^2_{\max} = \text{Min}(4, 5) - 1 = 4 - 1 = 3$$

$$\text{D'où : } V^2 = \Phi^2 / 3 = 0.15 / 3 = 0.05$$

5. A partir de ce résultat, conclure brièvement sur l'importance de l'écart à l'indépendance pour ces données NOBEL :

On peut conclure à un écart modéré à l'indépendance entre le pays (PAYS) et le type de Nobel (DISC) sur ces données car $0.04 < V^2 < 0.16$.

2. Dossier Howellp272 (23 points)

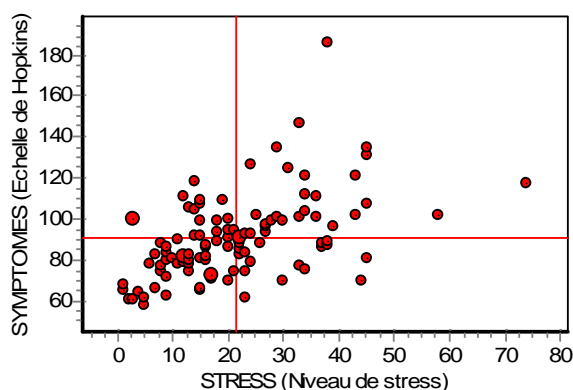
Dans son ouvrage *Méthodes statistiques en Sciences Humaines* (1988, Paris: De Boeck) David C. Howell présente des données qui visaient à étudier la relation entre le niveau de stress (variable STRESS) et la santé mentale chez des universitaires de première année, estimée à partir de l'échelle symptomatique de Hopkins (variable SYMPTOMES).

Une valeur élevée sur la variable STRESS correspond à un niveau élevé de stress. De même, une valeur élevée sur la variable SYMPTOMES suggère un problème de santé mentale.

Les analyses qui suivent portent sur l'ensemble des données, soit 107 sujets. On a reporté ci-dessous uniquement les valeurs observées pour les 10 premiers sujets :

SUJET	STRESS	SYMPTOMES
s001	30	99
s002	27	94
s003	9	80
s004	20	70
s005	3	100
s006	15	109
s007	5	62
s008	10	81
s009	23	74
s010	34	121
(...)

A. Graphe de corrélation (7 points)



1. Indiquer deux synonymes (autre que graphe de corrélation), français ou anglais, d'un tel graphique :

Nuage bivarié, Nuage des individus, Nuage de points, Diagramme cartésien, Diagramme de corrélation, Diagramme de dispersion, Scatterplot

2. Pourquoi est-il essentiel, lorsqu'on s'intéresse à la liaison entre deux variables numériques, d'étudier ce graphique et ne pas se contenter de calculer le coefficient de corrélation de Bravais-Pearson ? Donner deux arguments :

Vérifier si la liaison est de type linéaire ou d'un autre type (le coefficient de corrélation ne détecte que les liaisons de type linéaire).

Vérifier s'il existe des individus atypiques.

Vérifier s'il existe des sous-groupes d'individus (non détectés par le coefficient de corrélation).

3. Que représentent les deux axes (axe vertical et axe horizontal) au milieu de ce graphique ?

Ce sont les « axes moyens ». Ils permettent de repérer où se situent les moyennes de chacune des deux variables (en abscisse et en ordonnée) et de situer, graphiquement, les individus par rapport à ces moyennes.

4. On peut qualifier la liaison entre ces deux variables de positive. Donner deux arguments pour cela, l'un s'appuyant sur la forme du nuage (a), le second sur les quatre quadrants définis par les deux axes (b) :

a. Le nuage a une pente ascendante (de la gauche vers la droite). Il peut être ajusté par une droite de pente positive.

b. La plupart des individus se trouvent dans les deux cadrans en bas à gauche et en haut à droite . Pour tous les individus situés dans ces deux cadrans, le produit des écarts à la moyenne est positif. D'où une covariance (et par conséquent une corrélation) positive.

B. Corrélation (7 points)

On calcule la valeur du coefficient de corrélation linéaire de Bravais-Pearson ($r = +.51$) et l'intervalle de confiance sur cet indice ($IC_{5\%} = [+.35 ; +.63]$).

1. Au vu de ces informations, on peut savoir quel aurait été le résultat du test t de Student (Significatif/Non significatif) permettant de tester l'hypothèse d'une corrélation parente nulle. Quel aurait été ce résultat ? Justifier votre réponse :

Ce test t de Student aurait été significatif car la valeur 0 est située hors de l'intervalle de confiance (fait partie des valeurs de la corrélation parente pouvant être rejetées).

2. Rédiger une conclusion générale détaillée incluant les aspects descriptifs et inférentiels de cette analyse de la corrélation :

- Sur cet échantillon de 107 sujets, universitaires de première année,
- on constate une liaison de type linéaire
- positive
- ($r = +.51$)
- et forte
- ($r > +.40$)
- entre le niveau de Stress (STRESS) et le score sur l'échelle de santé mentale de Hopkins (SYMPTOM) :
- plus le niveau de stress est élevé, plus les problèmes de santé mentale sont nombreux.
- Il semble qu'une liaison positive existe bien dans la population d'où provient cet échantillon
- (ensemble des universitaires de première année)
- mais on ne peut pas conclure à une liaison forte dans cette population. Il est possible que cette liaison ne soit que modérée.
- ($IC_{5\%} = [+.35 ; +.63]$).

C. Régression linéaire simple (9 points)

Pour étudier la liaison entre ces deux variables, on procède à une régression linéaire. L'équation de régression obtenue est la suivante :

$$\text{SYMPTOMES}_{\text{préd}} = 0.8 \times \text{STRESS} + 74.$$

1. Le coefficient de régression associé à la variable STRESS (0.8) est positif. Pourquoi pourrait-on s'attendre à ce qu'il soit positif ?

- Deux justifications possibles :
- car ce coefficient est toujours de même signe que la corrélation (cf. $r = +.51$)
 - car le nuage est de pente positive

2. Que nous apprend la valeur 74 ?

C'est le score à l'échelle de Hopkins prédit pour un individu qui aurait un niveau de Stress égal à 0.

3. Interpréter la valeur du coefficient 0.8 :

En moyenne, si le niveau de stress (variable STRESS) augmente de 1 point, le score sur l'échelle de Hopkins (variable SYMPTOMES) augmente de 0.8 point.

4. Calculer la valeur prédite par cette équation pour le sujet s005 (indiquer le détail des calculs effectués) :

$$Pred = 0.8 \times 3 + 74 = 2.4 + 74 = 76.4$$

5. Calculer la valeur résiduelle pour ce sujet :

$$Res = 100 - 76.4 = 23.6$$

6. Énoncer la « Propriété de décomposition de la variance » dans ce contexte de la régression :

$$Var(Y_{obs}) = Var(Y_{pred}) + Var(Y_{res})$$

7. On trouve $R^2 = .26$. Rédiger un commentaire sur la valeur de cet indice :

La connaissance du niveau de Stress permet de prédire 26% de la variance des scores sur l'échelle de Hopkins, soit une part importante ($R^2 = .26 > .16$).

3. Lectures obligatoires (6 points)

A. Permutations (3 points)

Lorsque $n = 3$ personnes (abc) se présentent à l'entrée d'une salle, elles peuvent y entrer dans 6 ordres différents : abc, acb, bac, bca, cab, cba.

a. Quelle formule permet de retrouver ce nombre d'ordres ou « permutations » possibles (P) de n objets ? Indiquer la formule générale, développer le calcul pour $n = 3$ et retrouver le résultat ci-dessus (6 permutations) :

$$P = n !$$

$$P = 3 ! = 3 \times 2 \times 1 = 6$$

b. A partir de cette formule, calculer le nombre de permutations possibles pour 5 personnes :

$$P = 5 ! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

B. Combinaisons (3 points)

Les procédures inférentielles s'appuient sur des procédures d'échantillonnage où on construit tous les échantillons possibles de taille n pouvant être construits à partir de la population de taille N .

Le nombre d'échantillons possibles s'obtient par la formule ci-dessous. Utiliser cette formule pour calculer le nombre d'échantillons de taille $n = 2$ pouvant être constitués à partir d'une population de taille $N = 6$ (indiquer le détail des calculs) :

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{6!}{2!(6-2)!} = \frac{6!}{2! \times 4!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1) \times (4 \times 3 \times 2 \times 1)} = \frac{6 \times 5}{2 \times 1} = \frac{30}{2} = 15$$

4. Dossier Alexithymie (13 points)

Sur 22 adolescents diabétiques on a évalué leur niveau d'alexithymie (déficit de verbalisation des émotions). L'échelle utilisée est le TAS (Toronto Alexithymie Scale). Plus le score est élevé, plus le sujet peut être qualifié d'alexithymique.

Le tableau ci-dessous rapporte le sexe (G, F) et le niveau d'alexithymie (évalué par le TAS) de ces sujets.

Sexe	G	F	G	G	G	G	G	G	F	G	G	G	G	G	F	F	F	G	F	F	F	F
TAS	86	69	73	84	69	76	65	70	80	50	69	63	69	62	71	75	84	69	75	64	71	41

On se demande si le niveau d'alexithymie diffère chez les garçons et les filles.

On dispose des résultats suivants :

	n	Moy	Var
G	13	69.6	80
F	9	70.0	136

On considère un écart inférieur à 6 points comme négligeable, un écart supérieur à 12 points comme notable.

A. Calculs (3 points)

1. Indiquer la procédure de calcul de la moyenne générale, G et F confondus, à partir des moyennes des deux groupes (développer sans calculer le résultat final) :

$$m = \frac{(13 \times 69.6) + (9 \times 70.0)}{22} =$$

2. Indiquer à quoi sert la statistique nommée variance intra-groupes, notée *V_{intra}* :

La variance intra-groupes mesure la dispersion des valeurs à l'intérieur des groupes (par rapport à la moyenne du groupe).

3. Indiquer la procédure de calcul de cette variance intra-groupe (développer avec les valeurs de cet exemple, sans calculer le résultat final).

$$V_{\text{intra}} = \frac{(13 \times 80) + (9 \times 136)}{22} =$$

B. Indices calibrés (5 points)

1. Pour évaluer l'écart entre les groupes on définit des « indices calibrés ». Indiquer quelle est la caractéristique générale d'un indice calibré :

Un indice calibré prend en compte la dispersion intra-groupes (et pas seulement l'écart entre les moyennes des différents groupes).

2. Un exemple d'indice calibré est l'indice noté *Eta*². Quel autre nom désigne cet indice ?

Le rapport de corrélation

3. Indiquer la formule de calcul de cet indice Eta^2 :

$$Eta^2 = V_{inter} / V_{totale}$$

4. Entre quelles limites, minimum et maximum, peut varier cet indice ? Indiquer, pour chacune de ces limites, dans quel cas cette valeur est atteinte :

$$Min = 0 \text{ (ou 0\%)}$$

... quand les moyennes des groupes sont toutes égales, c'est-à-dire lorsque $V_{inter} = 0$ (quelle que soit la dispersion des valeurs à l'intérieur de chaque groupe).

$$Max = 1 \text{ (ou 100\%)}$$

...quand, à l'intérieur de chaque groupe, toutes les valeurs sont égales (mais ces valeurs peuvent différer d'un groupe à l'autre).

C. Inférence (5 points)

On procède à un test t de Student. On trouve $t = 0.08$.

1. Indiquer la formule de calcul du nombre de degrés de liberté (ddl) de cet test et le résultat :

$$ddl = n - 2 = 22 - 2 = 20$$

2. Utiliser l'extrait suivant de la table du test t de Student pour conclure sur le résultat du test (Significatif/Non significatif). Justifier votre réponse :

test non significatif (NS).

($t = 0.08 < 2.09$)

p	.05 (5%)	.01 (1%)	.001 (0.1%)
1	12.71	63.66	636.62
2	4.30	9.92	31.60
3	3.18	5.84	12.92
17	2.11	2.90	3.96
18	2.10	2.88	3.92
19	2.09	2.86	3.88
20	2.09	2.84	3.85
21	2.08	2.83	3.82
22	2.07	2.82	3.79

3. Que peut-on en conclure sur la différence entre G et F dans la population parente ?

Sachant que le test est non significatif, on ne peut pas conclure sur l'existence ou non d'une différence entre G et F dans la population.

On parle aussi, dans ce cas, d'un constat d'ignorance.

4. On calcule également un intervalle de confiance. On a $IC_{5\%} = [-10,8 ; 10,0]$. Que peut-on en conclure (ou ne pas conclure) sur l'importance de la différence parente ? Justifier votre réponse :

Tout d'abord, il est manifeste qu'on ne peut pas conclure à une différence nulle (comme pourrait le suggérer le test t non significatif) entre les deux populations (G et F) bien que cet intervalle inclut la valeur 0 (il est possible que la différence soit nulle). En effet il contient d'autres valeurs que 0 (il est également possible que la différence parente soit de -5, +5, +10, par exemple).

On ne peut pas conclure à une différence faible car l'IC comprend certes des valeurs faibles (entre 0 et 6 points) mais aussi des valeurs modérées (supérieures à 6).

Bref, on ne peut pas se prononcer sur l'ampleur de la différence parente. Tout au plus pourrait-on dire que la différence est faible ou modérée (pas importante).