

STATISTIQUES PARAMÉTRIQUES ET NON PARAMÉTRIQUES
Sujet donné en Janvier 2007
ÉNONCÉ ET INDICATIONS DE RÉPONSES

N.B. Calculatrices, tables des lois statistiques et résumé de cours autorisés.

Exercice 1

Dans une réplique d'une expérience de Smith (1979), l'auteur désire mettre en évidence l'effet facilitateur du contexte d'acquisition sur le rappel. Dans un premier temps, les sujets apprennent une liste de quatre-vingts mots dans une pièce orange décorée avec des draperies, posters, tableaux, ... Les sujets subissent quelques minutes après leur apprentissage un test de reconnaissance dont le but est simplement de leur laisser croire que l'expérience s'arrête là. Le lendemain, on demande aux sujets de noter les mots dont ils peuvent se souvenir. Le rappel se déroule dans différentes conditions qui définissent cinq groupes expérimentaux (les sujets étant affectés aléatoirement à chaque groupe).

- (1) Même contexte : les sujets se trouvent dans la même pièce orange où ils ont appris la liste la veille.
- (2) Contexte différent : les sujets se trouvent dans une pièce nettement différente de celle où ils ont appris la veille.
- (3) Contexte imaginé : les sujets sont dans la pièce du groupe (2). Mais on leur conseille de se rappeler d'abord la pièce orange (on les aide en leur posant des questions).
- (4) Contexte photographié : les sujets se trouvent dans la condition (3), mais on leur montre plusieurs photographies de la pièce orange.
- (5) Contexte placebo : les sujets sont dans la même pièce que le groupe (2). Mais on leur demande de se rappeler d'abord leur chambre à coucher à titre "d'échauffement intellectuel".

Les scores observés figurent dans le tableau ci-dessous.

| Groupe 1 Même contexte | Groupe 2 Contexte différent | Groupe 3 Contexte imaginé | Groupe 4 Contexte photographié | Groupe 5 Contexte Placebo |
|---------------------------|--------------------------------|------------------------------|-----------------------------------|------------------------------|
| 16 | 12 | 14 | 17 | 12 |
| 20 | 22 | 15 | 21 | 19 |
| 19 | 10 | 20 | 27 | 10 |
| 22 | 7 | 10 | 10 | 7 |
| 25 | 8 | 12 | 15 | 5 |
| 13 | 15 | 21 | 11 | 12 |
| 14 | 12 | 16 | 19 | 15 |

1) L'analyse de variance à un facteur réalisée sous Statistica produit le résultat suivant :

| Tests Univariés de Significativité pour Rappel | | | | | |
|--|---------|------------------|---------|--------|--------|
| Paramétrisation sigma-restreinte | | | | | |
| Décomposition efficace de l'hypothèse | | | | | |
| Effet | SC | Degr. de Liberté | MC | F | p |
| ord. origine | 7815,11 | 1 | 7815,11 | 332,36 | 0,0000 |
| Groupe | 256,46 | 4 | 64,11 | 2,73 | 0,0477 |
| Erreur | 705,43 | 30 | 23,51 | | |

Interpréter ce résultat.

2) Le protocole des rangs appliqué au tableau précédent est le suivant :

| Groupe 1 | Groupe 2 | Groupe 3 | Groupe 4 | Groupe 5 |
|----------|----------|----------|----------|----------|
| 22.5 | 12 | 16.5 | 24 | 12 |
| 28.5 | 32.5 | 19.5 | 30.5 | 26 |
| 26 | 6.5 | 28.5 | 35 | 6.5 |
| 32.5 | 2.5 | 6.5 | 6.5 | 2.5 |
| 34 | 4 | 12 | 19.5 | 12 |
| 15 | 19.5 | 30.5 | 9 | 19.5 |
| 16.5 | 12 | 22.5 | 26 | 1 |
| Moyennes | | | | |
| 25.00 | 12.71 | 19.43 | 21.50 | 11.36 |

a) Réaliser un test non paramétrique permettant de comparer les scores observés dans les différentes conditions.

b) Comparer le résultat de ce test à celui obtenu à l'aide de l'analyse de variance.

1) Les hypothèses de l'ANOVA sont :

H_0 : égalité des moyennes dans les populations parentes des 5 groupes

H_1 : les 5 moyennes ne sont pas toutes égales.

D'après Statistica, $F(4, 30) = 2.73$ et $p = 4,77\%$. On retient donc H_1 .

2) a) S'agissant d'un protocole $S < A >$, le test non paramétrique convenable ici est le test de Kruskal-Wallis. Les hypothèses du test sont :

H_0 : égalité des médianes dans les populations parentes des 5 groupes

H_1 : les 5 médianes ne sont pas toutes égales.

Calcul de la statistique de test : $K_{obs} = \frac{12}{35 \times 36} (7(25^2 + 12.71^2 + 19.43^2 + 21.5^2 + 11.36^2)) - 3 \times 36 = 9.02$

K suit une loi du χ^2 à 4 ddl. Au seuil de 5%, $\chi_{crit}^2 = 9.488$. On retient donc H_0 .

2) b) On constate que les niveaux de significativité des deux tests sont assez proches. Cependant, celui du premier test est inférieur à 5%, pendant que le second est supérieur à 5%, ce qui nous amène à des conclusions opposées. Cette situation illustre un résultat connu : les tests non paramétriques sont en général moins puissants que les tests paramétriques correspondants.

Exercice 2

Dans le cadre d'une étude sur l'efficacité comparée de l'apprentissage collaboratif en face à face et à distance (on line), les sujets devaient suivre une série de séminaires, les uns en face à face, les autres en enseignement à distance (on line). Dans les deux conditions, les étudiants étaient répartis en petits groupes permettant un travail collaboratif.

Après la phase d'apprentissage, les sujets, répartis en groupes de 4 ou 5 devaient produire un dossier utilisant les techniques apprises lors de l'apprentissage. Sept critères ont été établis, avec un score de 0 à 5 pour chacun d'eux. Le score total était donc compris entre 0 et 35.

Les scores observés pour les groupes de sujets qui ont travaillé en face à face (16 groupes) et pour ceux qui ont travaillé on line (17 groupes) sont les suivants :

| | | | | | | | | | | | | | | | | | |
|-------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| face à face | 12 | 12 | 14 | 15 | 15 | 18 | 19 | 19 | 20 | 20 | 21 | 21 | 21 | 26 | 29 | 31 | |
| on line | 17 | 19 | 22 | 23 | 24 | 24 | 25 | 27 | 28 | 28 | 28 | 29 | 29 | 31 | 31 | 31 | 32 |

Les auteurs ont choisi de comparer les performances des groupes dans les deux conditions à l'aide d'un test de Mann Whitney. Réaliser le test et conclure à un seuil de 5%.

Réponse. Réalisons un test bilatéral. H_0 : égalité des médianes en modes face à face et on line. H_1 : médianes différentes.

On détermine le protocole des rangs, puis on calcule la somme des rangs dans les deux groupes. On obtient : $W_1 = 183.5$ et $W_2 = 377.5$. D'où : $\bar{R}_1 = 11.47$, $\bar{R}_2 = 22.20$, $E^2 = 11.34$ et $Z_{obs} = -3.19$. Au seuil de 5%, $Z_{crit} = 1.96$. On conclut donc sur H_1 .

Exercice 3

Dans un article publié en 2003¹, des chercheurs ont analysé, d'un double point de vue à la fois économique et psychologique, le comportement "pro-environnement" des consommateurs. Une étude précédente avait montré que de nombreux foyers américains étaient prêts à payer leur électricité plus cher, à condition qu'elle soit produite à partir d'énergies renouvelables.

Dans le cadre d'un programme "d'électricité verte", une compagnie du Michigan propose à ses clients de payer une redevance additionnelle, en contrepartie de laquelle la compagnie connecte sur le réseau un bloc photovoltaïque (production d'électricité solaire) d'une puissance déterminée. Afin d'étudier les motivations des souscripteurs de ce programme, une enquête, sous forme de questionnaire envoyé par courrier, a été adressée à 281 participants et à 619 non-participants au programme. Les taux de réponse ont été de 95% pour les participants et de 76% pour les non-participants.

1) Une partie du questionnaire d'enquête comportait neuf questions destinées à évaluer l'altruisme du sujet. Pour chacune d'elles, une échelle de Likert en 5 points était utilisée (modalités : tout à fait d'accord, plutôt d'accord, sans opinion, plutôt opposé, tout à fait opposé).

Exemples :

¹Clark C.F., Kotchen M.J., Moore M.R., Internal and external influences on pro-environmental behavior : Participation in a green electricity program, Journal of Environmental Psychology, Vol. 23, pp. 237-246, 2003

(1) Je me préoccupe d'économiser l'énergie seulement lorsque cela me permet de diminuer mes factures.

(4) Il est de mon devoir d'aider les autres lorsqu'ils sont incapables de s'en sortir seuls.

Les variances observées sur chacun des 9 items sont indiquées dans le tableau ci-dessous :

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-------------------|------|------|------|------|------|------|------|------|------|
| $\text{Var}(X_i)$ | 1.72 | 0.89 | 1.80 | 0.98 | 1.00 | 1.36 | 1.37 | 0.97 | 0.78 |

a) On recode numériquement le degré d'accord des sujets aux différents items, et on additionne les résultats pour obtenir le score de chaque sujet sur une échelle d'altruisme. Quelle précaution doit-on prendre pour effectuer cette cotation ?

b) La variance observée sur l'échelle d'altruisme est $\text{Var}(S) = 29.08$. Calculer un coefficient mesurant la consistance interne entre les différents items. Quel jugement peut-on porter sur la valeur obtenue ?

2) On a demandé à un échantillon de 20 sujets choisis parmi les participants au programme de classer par ordre de préférence les cinq motivations environnementales suivantes :

1. La réduction de la pollution de l'air améliorera la santé des écosystèmes naturels
2. La réduction de la pollution de l'air profitera aux personnes résidant dans ma région
3. Ma santé, et celle de ma famille pourront s'améliorer car le programme améliorera la qualité de l'air
4. La diminution des émissions de dioxyde de carbone liées à la production d'électricité ralentira la vitesse du réchauffement du climat
5. J'éprouve de la satisfaction à participer à ce programme, indépendamment de ses effets environnementaux.

Les résultats observés sont les suivants :

| | Motivation 1 | Motivation 2 | Motivation 3 | Motivation 4 | Motivation 5 |
|------------|--------------|--------------|--------------|--------------|--------------|
| s1 | 2 | 4 | 3 | 1 | 5 |
| s2 | 1 | 2 | 4 | 3 | 5 |
| s3 | 1 | 4 | 2 | 3 | 5 |
| s4 | 1 | 3 | 2 | 4 | 5 |
| s5 | 1 | 3 | 2 | 4 | 5 |
| s6 | 2 | 4 | 3 | 1 | 5 |
| s7 | 1 | 2 | 3 | 4 | 5 |
| s8 | 1 | 4 | 2 | 3 | 5 |
| s9 | 2 | 3 | 1 | 4 | 5 |
| s10 | 4 | 3 | 5 | 1 | 2 |
| s11 | 2 | 1 | 5 | 3 | 4 |
| s12 | 3 | 2 | 1 | 5 | 4 |
| s13 | 5 | 3 | 4 | 1 | 2 |
| s14 | 3 | 2 | 1 | 5 | 4 |
| s15 | 3 | 1 | 5 | 2 | 4 |
| s16 | 2 | 3 | 1 | 4 | 5 |
| s17 | 2 | 1 | 4 | 3 | 5 |
| s18 | 1 | 2 | 4 | 3 | 5 |
| s19 | 1 | 2 | 3 | 4 | 5 |
| s20 | 3 | 1 | 4 | 2 | 5 |
| $\sum R_i$ | 41 | 50 | 59 | 60 | 90 |

Etudier à l'aide d'un test approprié si les rangs moyens des cinq motivations sont significativement différents (seuil : 5%)

3) Lors d'études précédentes, des auteurs ont identifié l'égoïsme, le social-altruisme et l'intérêt pour la biosphère comme prédicteurs des comportements pro-environnement. Cependant, lorsque le comportement pro-environnement nécessite la volonté de payer un surcoût pour la protection de l'environnement, ils ont trouvé que seul l'égoïsme est un prédicteur fiable de tels comportements. Les résultats précédents sont-ils en accord avec ces conclusions ?

Réponses : 1) a) Sur l'item 4, un score élevé indique un altruisme fort, alors que c'est l'inverse pour l'item 1. Il faut donc inverser l'échelle pour certains items, de manière qu'un score élevé corresponde systématiquement à un fort niveau d'altruisme (par exemple).

1) b) On calcule ici le coefficient α de Cronbach.

On obtient : $\alpha = \frac{9}{8} \left(1 - \frac{1.72+0.89+1.80+0.98+1.00+1.36+1.37+0.97+0.78}{29.08} \right) = 0.7044$. La valeur du coefficient est légèrement supérieure à 0.7. La cohérence du questionnaire peut être jugée assez bonne.

2) C'est un test de Friedman qu'il convient de faire ici. H_0 affirme à l'égalité des médianes correspondant aux 5 motivations, H_1 correspond à des médianes différentes. Le calcul donne : $F_r = \frac{12}{20 \times 5 \times 6} (41^2 + 50^2 + 59^2 + 60^2 + 90^2) - 3 \times 20 \times 6 = 27.24$. En utilisant la table du test de Friedman, ou celle du χ^2 à 4 ddl, on obtient : $F_{crit} = 9.4$. On conclut donc sur H_1 .

3) Ce sont les motivations 3 et 5 qui sont les moins altruistes. Ces motivations ne se retrouvent pas classées dans les rangs les moins élevés. Les résultats de cette étude contredisent donc l'affirmation selon laquelle "lorsque le comportement pro-environnement nécessite la volonté de payer un surcoût pour la protection de l'environnement, seul l'égoïsme est un prédicteur fiable de tels comportements".