

Durée de l'épreuve : 1h 30mn

Épreuve sans document. La calculatrice est autorisée (sans sa documentation).

Les exercices (encadrés) sont indépendants. Le barème (sur 60) est donné à titre indicatif.

Indiquer les réponses exclusivement sur ce document et aux endroits réservés à cet effet (ne pas écrire dans la marge)

Les logiciels DS3Win, PAC et Statistica ont été utilisés pour analyser les données.

Dossier DEFINITIONS : Description et Inférence (3 points)

1. Expliquer, en deux ou trois phrases maximum, en quoi consiste la distinction entre l'étape descriptive et l'étape inférentielle, lors de l'analyse statistique des données :

2. Citer 2 statistiques inférentielles :

Dossier RORSCHACH : Structures (7 points)

Source : Inspiré de Azoulay, Emmanuelli, Rausch de Traubenberg, Corroyer, & Rozencwajg (1999) - Les données normatives du Rorschach à l'adolescence, *Symposium "Normalcy and pathologies: the specificity of adolescence"*. XVIème congrès International du Rorschach et des méthodes projectives. Amsterdam : 19-23 juillet.

Le test de Rorschach consiste en une dizaine de planches comportant chacune une tache d'encre différente. Chaque planche est présentée au sujet et ce dernier est invité à dire *"tout ce à quoi la planche lui fait penser"*.

Une recherche, portant sur 700 sujets, a visé à constituer une base de données normative des productions en présence des différentes planches. Ces données devraient servir ensuite de référence pour les utilisations individuelles de ce test. On recueille les verbalisations sur sept groupes d'âge (de 13 à 64 ans) composés chacun d'hommes et de femmes de 3 milieux sociaux différents.

On relève, pour chacune des 10 planches présentées, le nombre de kinesthésies (évoqueries d'objets en mouvement) données par le sujet.

On se demande dans quelle mesure la production de kinesthésies au Rorschach varie d'une planche à l'autre et selon l'âge, le milieu social et le sexe.

On ne précisera pas l'indice pour le facteur de groupe.

1. Indiquer quels sont les facteurs (ou VI) étudiés dans cette recherche :

2. Indiquer quelle est la variable dépendante (VD) :

3. Indiquer quels sont, parmi les facteurs :

- le facteur de groupe :

- le(s) G-facteur(s) :

- le(s) T-facteur(s) :

4. Indiquer à l'aide des symboles d'emboîtement (<>) de croisement (*) et de confusion (~) :

- la relation entre les G-Facteur(s) :

- la formule du plan le plus riche :

Dossier LATERAL : De la description à l'inférence (16 points)

Source : Inspiré de Fagard J. et Corroyer, D. (en révision) - *Using a continuous index of laterality to determine how laterality is related to interhemispheric transfer and bimanual coordination in children.*

Cette recherche visait à étudier, en particulier, l'évolution du transfert interhémisphérique chez le jeune enfant. On s'intéresse ici à une partie des données, recueillies chez deux groupes d'enfants droitiers de 5 ans (a1) et 6 ans (a2). Ils ont à comparer, les yeux bandés, les textures de petits coussins. Chaque coussin est placé sur la paume de la main de l'enfant. Chaque enfant réalise plusieurs comparaisons de textures dans deux conditions : les deux textures sont présentées successivement, soit sur une main puis sur l'autre (condition croisée, notée c1), soit sur la même main (condition non croisée, notée c2). On note, pour chaque enfant, le pourcentage d'erreurs dans chacune des deux conditions.

On fait l'hypothèse que, chez les jeunes enfants, la comparaison est plus difficile dans la condition croisée. En effet, dans cette condition, la comparaison nécessite un transfert interhémisphérique des informations tactiles recueillies pour qu'elles puissent être comparées. Or ce transfert interhémisphérique ne s'établit que progressivement au cours du développement.

On présente ci-dessous les données recueillies sur un sous-ensemble de 10 enfants :

	c1	c2
s1a1	30	20
s2a1	35	18
s3a1	40	22
s4a1	28	17
s5a1	27	23
s6a2	18	12
s7a2	15	22
s8a2	20	19
s9a2	21	20
s10a2	11	22

On considère qu'un effet est faible s'il est inférieur à 5 points de pourcentage, et important s'il est supérieur à 10 points de pourcentage.

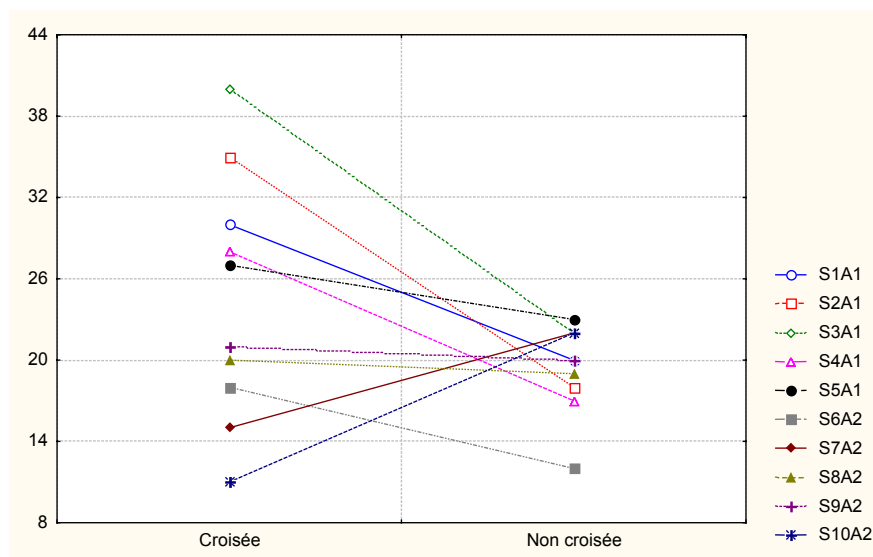
1. Structures (3 points)

- Citer au moins un facteur maintenu constant dans cette expérience :
- Indiquer le plan de recueil des données le plus riche :
- Quelle est la variable dépendante ?

2. Représentation graphique (2 points)

On s'intéresse aux effets individuels de la condition, croisée (c1) ou non croisée (c2), sur les 10 sujets (indépendamment de l'âge).

Commenter ci-dessous, le graphique de la page suivante de ce point de vue (on pourra s'aider également du tableau de données) :



3. Analyse d'un effet (11 points)

On s'intéresse à l'effet du facteur Condition (croisée/non croisée) chez les sujets de 5 ans.

a. Indiquer l'expression formelle de cette question :

b. Indiquer les valeurs du PDP :

c. Calculer l'effet observé, d_{obs} :

$$d_{obs} =$$

d. Citer le nom abrégé et la formule d'un effet calibré adapté à cette question :

e. On trouve $EC = 2.10$. Commenter brièvement ce résultat :

f. On note tout d'abord que la valeur observée du test t de Student est 4.71 (d'où le seuil observé $p < .005$).

Par ailleurs, le logiciel PAC indique également : $p(\text{effet vrai} > 8.09) = 0.90$. Exprimer ce dernier résultat en langage naturel :

g. Rédiger une conclusion détaillée qui résumera toutes les informations recueillies et qui distinguera clairement les deux étapes, descriptive et inférentielle :

Dossier REPRESENTATION : Variances Inter et Intra (3 points)

Soient les deux groupes de valeurs suivantes :

$g_1 (m_1 = 16)$	$g_2 (m_2 = 8)$
12 13 15 17 19 20	4 6 8 9 10 11

Attention ! Il n'est pas demandé de calculer les variances inter et intra dans les exercices ci-dessous.

1. Proposer une modification d'une valeur du groupe g_1 de manière à diminuer la variance Inter (sans se préoccuper de la variance Intra). Justifier la réponse :

2. Proposer une modification d'une valeur du groupe g_2 de manière à augmenter la variance Intra (sans se préoccuper de la variance Inter). Justifier la réponse :

3. Proposer de modifier deux valeurs du groupe g_2 de manière à augmenter la variance Intra sans modifier la variance Inter. Justifier la réponse :

Dossier FICTIF : Protocoles Dérivés Pertinents (8 points)

Supposons que le tableau suivant rapporte les résultats d'une expérience pour laquelle la formule du plan de recueil des données était : $S < A_2 > * B_2 * C_2$. On s'intéresse à différents effets, notés ci-dessous de manière formelle (A, B/c1, A/c1, A.B, B/c1a1, B.C). Pour chacun de ces effets :

1. Calculer, et reporter dans la colonne correspondante du tableau de droite, le protocole dérivé pertinent (PDP) permettant de calculer l'effet calibré et le T de Student.

2. Indiquer dans la dernière ligne du tableau ("D ou M") si l'effet moyen peut être obtenu, soit en calculant la moyenne générale du PDP (dans ce cas noter "M" dans la case du PDP correspondant), soit en calculant une différence de moyennes entre deux groupes du PDP (dans ce cas noter "D").

	b1c1	b2c1	b1c2	b2c2	A	B/c1	A/c1	A.B	B/c1a1	B.C
s1a1	3	2	8	3						
s2a1	2	4	4	2						
s3a1	0	1	3	4						
s4a2	0	2	1	1						
s5a2	6	7	4	3						
s6a2	2	2	3	1						
	D ou M									

Dossier SOURIRE : Probabilités inverses (8 points)

Source : Toute ressemblance avec des données réelles serait fortuite...

Un sociologue s'intéresse à l'attitude des voyageurs du métro envers leurs voisins pendant le voyage. Il suit les voyageurs pendant 30mn et note leur comportement majoritaire en 3 catégories : Quelconque (Q), plutôt Souriant (S), ou plutôt Agressif (A).

L'hypothèse de ce sociologue étant qu'un certain groupe social (noté F) serait plus souriant et moins agressif qu'un autre groupe (noté I), il note donc également l'appartenance de la personne observée à l'un ou l'autre de ces deux groupes.

Il recueille des données sur 250 personnes, dont 200 personnes du groupe F (80%) et 50 personnes du groupe I (20%). Il constate que, sur 75 personnes plutôt souriantes, 50 appartiennent au groupe F ($50/75 = 0.67$) et 25 au groupe I ($25/75 = 0.33$).

1. Donner une expression formelle (sous la forme $P(A/B)$) des deux proportions calculées ci-dessus :

2. Le sociologue en conclut que les personnes du groupe F sont plus souriantes que celles du groupe I. Pourquoi sa conclusion est-elle erronée ?

3. Quelles proportions faut-il calculer pour comparer les deux groupes du point de vue de leur comportement social (donner des expressions formelles du type $P(A/B)$) ? Calculer ces proportions :

4. Répondre à la question posée :

5. Le sociologue constate 5% de comportements plutôt agressifs parmi les personnes du groupe F contre 4% parmi les personnes du groupe I. Utiliser le théorème de Bayes, rappelé ci-dessous, pour calculer la proportion $P(F/A)$. Indiquer le détail des calculs :

$$P(j/k) = \frac{P(j) \cdot P(k/j)}{\sum_j P(j) \cdot P(k/j)}$$

$$P(F/A) =$$

6. Compte tenu des éléments fournis, présenter les données complètes sous la forme d'une distribution d'effectifs bivariée (tableau de contingence) :

Dossier DEPRESS : Test de typicalité (5 points)

On approxime la distribution des scores à une échelle de dépression de 2046 personnes tout-venant par une distribution normale de moyenne $\mu = 19.90$, et de variance $\sigma^2 = 25$.

1. On se demande quel est, approximativement, le nombre de personnes dont le score est supérieur à $x = 25$? Commencer à développer les calculs nécessaires. Que manque-t-il pour mener ces calculs à terme ?

2. On s'intéresse à un groupe de 36 personnes isolées socialement et à qui l'on fait passer la même échelle de dépression. La moyenne des scores de cet échantillon est de 22.30. On se demande si ce groupe de 36 personnes peut être qualifié d'atypique, du point de vue du score de dépression, par rapport à la population des 2046 personnes.

a. Calculer la moyenne et la variance de la distribution d'échantillonnage de la moyenne (DEM) :

$$\text{Moy}(M) =$$

$$\text{Var}(M) = \frac{\sigma^2}{n} \frac{(N-n)}{(N-1)} =$$

b. On veut situer le groupe des 36 personnes par rapport à la population tout-venant, en procédant à un test de typicalité. Développer les calculs nécessaires, aussi loin que possible compte tenu des outils disponibles :

c. Sachant que le calcul final donne $P(M > 22.30) = .0055$, répondre à la question sur l'atypicalité de ce groupe de 36 personnes du point de vue de la dépression :

Dossier BINAIRE : Formules (10 points)

L'ensemble des exercices de ce dossier vise à exprimer les formules des statistiques classiques (moyenne, variance...) dans le cas particulier de deux valeurs numériques (x_1 et x_2 ou m_1 et m_2), de même poids relatifs ($f_1 = f_2 = 1/2$). On suppose également que x_2 (resp. m_2) désigne toujours la plus élevée des deux valeurs.

1/ Calculs (3 points)

On considère tout d'abord un cas particulier où $x_1 = 10$ et $x_2 = 18$.

Calculer la moyenne (m), la différence ($d = x_2 - x_1$), la variance (var), l'écart-type (ety), la variance corrigée (s^2) et l'écart-type corrigé (s) de ces deux valeurs :

statistiques	résultats
m	
d	
var	
ety	
s^2	
s	

Ces résultats pourront permettre de vérifier les formules générales développées ci-après.

2/ Démonstrations (7 points)

On considère maintenant le cas général (x_1 et x_2 quelconques)

a. Sachant que m est à égale distance des deux valeurs x_1 et de x_2 , exprimer les deux écarts à la moyenne ($x_1 - m$) et ($x_2 - m$) en fonction de d (différence entre ces deux valeurs, soit $d = x_2 - x_1$) :

$$(x_1 - m) = \quad \quad \quad (x_2 - m) =$$

b. la formule générale de la variance est $\text{var} = \sum f_i (x_i - m)^2$

Développer cette formule, dans le cas particulier étudié ici (deux valeurs x_1 et x_2 équipondérées) en l'exprimant comme une somme de deux termes :

$$\text{var} =$$

Exprimer alors la variance et l'écart-type comme une fonction de d (cf. a.) :

$$\text{var} = \quad \quad \quad \text{ety} =$$

De manière générale (n valeurs) la variance corrigée $s^2 = (n/n-1) \text{var}$.

Exprimer la variance corrigée et l'écart-type corrigé de deux valeurs équipondérées comme une fonction de d :

$$s^2 = \quad \quad \quad s =$$

c. Sachant que les résultats précédents sont valides pour deux valeurs numériques quelconques (donc pour deux moyennes, par exemple) et de même poids relatif, exprimer la variance inter, la variance inter corrigée et l'écart-type inter corrigé, comme une fonction de la différence entre les deux moyennes ($m_2 - m_1$), notée d :

$$V_{\text{inter}} = \quad \quad \quad S^2_{\text{inter}} = \quad \quad \quad S_{\text{inter}} =$$

d. Il existe plusieurs effets calibrés pour évaluer la différence entre deux groupes indépendants du point de vue d'une variable numérique. L'effet calibré de Cohen (nommé souvent "d de Cohen") est égal à d / S_{intra} (d désigne, dans la formule, la différence entre les deux moyennes).

Rappeler la formule d'un autre effet calibré, noté ECG :

$$\text{ECG} =$$

Compte tenu des résultats précédents, exprimer l'effet calibré de Cohen en fonction de l'effet calibré ECG (dans le cas de deux groupes équilibrés) :

$$\text{"d de Cohen"} = d / S_{\text{intra}} =$$