

MEDI

Durée de l'épreuve : 2 heures.

Aucun document n'est autorisé. Seule la calculatrice est autorisée, sans sa documentation.

Attention : les exercices (encadrés) sont indépendants. Le barème donné à titre indicatif est sur 60 ; la note finale sera donnée sur 20.

La page 9 (extraits des tables des valeurs critiques de la variable χ^2 , de la variable T de Student, de la variable F de Snedecor et formulaire) peut être détachée et conservée.

Indiquer les réponses exclusivement sur ce document.

Dossier Recherches (14 points)

Source : Herman, J. (1986). *Analyse de données qualitatives*. Paris : Masson.

Lors d'une étude sur l'efficacité de 1691 unités de recherche choisies au hasard (au niveau international), on a obtenu les résultats suivants en mettant en relation les Sciences étudiées (*Formelles, Naturelles, Appliquées ou Sociales*) et les méthodes qu'elles utilisent pour l'analyse de leurs données (*Test d'hypothèse, Déduction, Statistique Descriptive ou Modélisation*)

Tableau 1 : tableau de contingence Sciences étudiées / Méthodes utilisées

		<i>Sciences étudiées</i>			
		<i>Formelles</i>	<i>Naturelles</i>	<i>Appliquées</i>	<i>Sociales</i>
<i>Méthodes utilisées</i>	<i>Test d'Hypo.</i>	35	457	178	216
	<i>Déduction</i>	19	83	49	42
	<i>Stat. Descrip.</i>	16	101	50	114
	<i>Modélisation</i>	13	114	72	132
					1691

Question initiale : les Sciences étudiées ont-elles toutes le même profil quant à l'utilisation des méthodes d'analyse de données ?

Analyse descriptive

1) On se demande quelle est la Méthode d'analyse de données la plus fréquemment utilisée, indépendamment des Sciences étudiées (toutes Sciences confondues).

a) Calculer les pourcentages pertinents pour répondre à cette question (les résultats seront présentés arrondis à l'entier le plus proche) :

b) Commenter les résultats obtenus :

2) Compléter le tableau 2 ci-dessous (case vide) :

Tableau 2 : tableau des effectifs théoriques

		<i>Sciences étudiées</i>			
		<i>Formelles</i>	<i>Naturelles</i>	<i>Appliquées</i>	<i>Sociales</i>
Méthodes utilisées	<i>Test d'Hypo.</i>	43.49		182.86	264.07
	<i>Déduction</i>	9.47	86.17	39.83	57.52
	<i>Stat. Descrip.</i>	13.79	125.46	57.99	83.75
	<i>Modélisation</i>	16.25	147.79	68.31	98.65

b) Indiquer précisément comment a été calculée la valeur correspondant à "Test d'hypothèse" pour les "Sciences Naturelles" :

c) Expliquer ce qu'indique la valeur 43.49 (Test d'Hypothèse/Sciences Formelles) :

3) Écart à l'indépendance

a) Construire le tableau des attractions/répulsions ci-après :

Tableau 3 : tableau des attractions/répulsions

		<i>Sciences étudiées</i>			
		<i>Formelles</i>	<i>Naturelles</i>	<i>Appliquées</i>	<i>Sociales</i>
Méthodes utilisées	<i>Test d'Hypo.</i>				
	<i>Déduction</i>				
	<i>Stat. Descrip.</i>				
	<i>Modélisation</i>				

b) Indiquer précisément comment les méthodes d'analyse de données diffèrent selon le type de recherche pratiquée ?

4) Écart global à l'indépendance

a) On trouve $\Phi^2 = 0.0432$. Interpréter cette valeur :

b) Rédiger une conclusion descriptive :

5) Analyse inférentielle : on désire généraliser les résultats obtenus sur cet échantillon à une population plus vaste.

a) Donner le nom du test à utiliser, indiquer la formule et développer le calcul :

b) Calculer le nombre de degrés de liberté associé à ce test (rappeler la formule) :

c) Indiquer le résultat du test :

d) Quelle est la condition préalable pour pouvoir donner une interprétation fréquentiste, en terme de test d'hypothèse, de ce résultat ? Est-elle vérifiée ?

6) Les résultats trouvés en **4-a** et **5-c** paraissent contradictoires. Indiquer pourquoi ils sont compatibles :

Dossier Formules (4 points)

La formule générale du coefficient de corrélation de Bravais-Pearson est :

$$r = \frac{Cov(x, y)}{Ety(x) \times Ety(y)}. \text{ Sachant que } Cov(x, y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{n}, \text{ montrer que dans le cas où}$$

X et Y sont des variables centrées-réduites, r peut s'écrire $\frac{\sum xy}{n}$:

Dossier UE (12 points)

On s'intéresse à la liaison entre les résultats obtenus à deux UE du Deug Sciences Humaines et Sociales de l'Université René Descartes : l'UE de Psychologie Expérimentale (EXPÉ) et l'UE de Psychologie Différentielle (DIFF) pour un échantillon de 9 étudiants de Premier Cycle, choisis au

hasard dans la population des étudiants en Premier Cycle du Centre Universitaire de Boulogne. Les résultats sont les suivants (notes sur 20) :

	<i>e1</i>	<i>e2</i>	<i>e3</i>	<i>e4</i>	<i>e5</i>	<i>e6</i>	<i>e7</i>	<i>e8</i>	<i>e9</i>
EXPÉ	3.5	9.0	10.0	0.5	5.5	6.5	7.5	1.0	10.5
DIFF	1.0	8.0	11.0	1.5	7.5	11.5	9.5	2.0	11.0

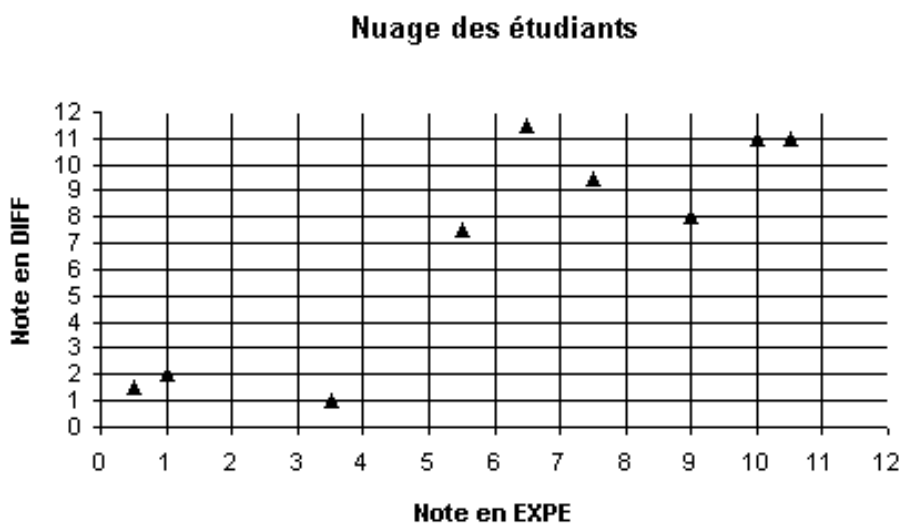
1) Le coefficient de corrélation de Bravais-Pearson entre ces deux variables est égal, arrondi à 2 décimales, à +.88. A l'aide de la calculette, recalculer cette valeur avec une précision de 6 décimales :

$$r_{(EXPÉ, DIFF)} =$$

2) Que deviendrait ce coefficient si la note à l'UE d'EXPÉ était donnée sur 40 et non sur 20, alors que la note en DIFF resterait sur 20 ? Justifier.

3) On cherche à estimer la note en DIFF à partir de la note en EXPÉ. Calculer et indiquer ci-dessous l'équation de la droite de régression (utiliser la calculette ou les formules indiquées page 9 ; présentation des résultats : 2 décimales) :

4) Tracer cette droite de régression sur le graphique ci-après :



5) Représenter le centre de gravité de ce nuage sur le graphique ci-dessus et indiquer ses coordonnées ci-après :

6) Si l'on représentait l'autre droite de régression, qui exprime EXPÉ en fonction de DIFF, sur le graphique ci-dessus, quel serait le point d'intersection des deux droites de régression ?

7) On cherche à apprécier, par un indice global, la qualité de l'ajustement linéaire.

a) Indiquer le nom de cet indice et donner sa valeur :

b) Conclure descriptivement :

8) Représenter sur le graphique précédent les écarts résiduels des étudiants $e1$ et $e6$.

9) On cherche maintenant à généraliser les résultats obtenus par cet échantillon à l'ensemble des étudiants de premier cycle. Indiquer le nom et la formule du test à utiliser ainsi que le nombre de degrés de liberté associé à ce test (sans effectuer de calcul) :

10) La valeur obtenue pour ce test est égale à : 4.87. Indiquer le résultat du test :

Dossier MAC (5 points) HORS PROGRAMME 2003-2004

D'après : Frugier, G. (1992). *Exercices ordinaires de probabilités*. Paris : Marketing, Ellipses.

Olivier est étudiant. Ses revenus l'obligent à se nourrir exclusivement de hamburgers. Il va chez Mac A dans 70% des cas et chez Mac B le reste du temps. Il attrape une gastrite dans 10% des cas chez Mac A et dans 20% des cas chez Mac B.

Aujourd'hui il a une gastrite. Où a-t-il pris, le plus probablement, son dernier hamburger ?

1) Appliquer l'une ou l'autre des deux formules du théorème de Bayes (cf. ci-dessous) pour répondre à la question posée :

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

HORS PROGRAMME 2003-2004

2) En supposant que le nombre total de "repas" mangés est de 100, construire le tableau de contingence qui répartit les hamburgers selon ces deux variables : Mac et Gastrite.

	Gastrite	Pas de Gastrite	Total
Mac A			
Mac B			

Total

3) Vérifier à partir du tableau construit ci-dessus le résultat trouvé en 1) et indiquer le détail des calculs :

Dossier Stratégies (10 points)

On s'intéresse aux stratégies de résolution d'une tâche par 60 enfants, choisis au hasard et testés deux fois à un an d'intervalle. Deux stratégies sont observées : une stratégie A performante et une stratégie B peu performante. On obtient le nombre de réussites suivant pour chaque stratégie utilisée et chaque tranche d'âge testée :

		13 ans	
		Stratégie A	Stratégie B
12 ans	Stratégie A	30	15
	Stratégie B	12	3

On jugera un effet faible s'il est inférieur ou égal à 5 points de pourcentage, important s'il est supérieur ou égal à 10 points de pourcentage.

On fait l'hypothèse que la fréquence d'utilisation de la stratégie A (la plus performante) évolue entre 12 ans et 13 ans.

- 1) La structure de ce tableau de contingence est-elle de type : $S < G >$ ou $S * T$? Justifier.
- 2) Calculer l'effet de l'âge sur la fréquence d'utilisation de la stratégie A dans cet échantillon :
- 3) Élaborer une conclusion descriptive :
- 4) On désire généraliser les résultats à la population des enfants âgés de 12 ans et 13 ans.
 - a) Indiquer le nom du test à utiliser, donner sa formule et développer le calcul :
 - b) Indiquer le résultat du test :
 - c) Pour l'ensemble des enfants âgés de 12 ans et 13 ans, peut-on penser qu'il existe une évolution de la fréquence d'utilisation de la stratégie A ? Justifier.

Dossier Mémoire (15 points)

D'après : Richard, J.F., Lecoutre, M.P. (1994). Inférence statistique. In R. Ghiglione & J.F. Richard (Eds.), *Cours de Psychologie (Tome 4 : Mesures et analyse)*. Paris : Dunod, pp. 387-482.

Deux groupes de sujets : 7 adultes étudiants et 6 enfants de Cours Moyen, choisis au hasard, ont à résoudre une suite de problèmes. Dans chaque problème, ils doivent exécuter mentalement une suite d'additions ; pour le dernier calcul du problème, ils doivent utiliser certains des résultats de calculs précédents et donc rechercher en mémoire les résultats de ces calculs (il peut y avoir selon les cas 2, 3 ou 4 chiffres à récupérer en mémoire). On note l'exactitude de la dernière addition et le temps mis pour faire le dernier calcul. Les données ci-après concernent uniquement la condition "4 chiffres à récupérer en mémoire" pour la variable dépendante "Temps de résolution" (exprimé en demi-secondes) :

Adultes							Enfants					
s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13
28	23	42	33	22	26	15	74	112	61	134	32	70

Moyenne Générale : 51.692

Variance Totale : 1264.213

Question initiale : les enfants mettent-ils plus de temps que les adultes pour faire le calcul dans le cas de 4 chiffres à récupérer en mémoire ?

Critère sémantique : on jugera un effet faible s'il est inférieur ou égal à 15 demi-secondes, important s'il est supérieur ou égal à 30 demi-secondes.

Présentation des résultats : 2 décimales

A - On cherche à répondre à la question initiale en comparant ces 2 groupes.

1) Analyse descriptive

a) Calculer l'effet moyen du facteur "Âge" :

b) Élaborer une conclusion :

2) Calculer la variance inter de ce protocole (indiquer la procédure de calcul) :

$V_{inter} =$

3) Calculer la variance intra de ce protocole (indiquer la procédure de calcul) :

$V_{intra} =$

5) Analyse inférentielle : on désire pouvoir généraliser les résultats obtenus sur l'échantillon à une population plus vaste.

a) Quel test va-t-on utiliser (indiquer seulement la formule du test et le nombre de degrés de libertés associé à ce test, sans calculer) ?

b) La valeur obtenue pour ce test est égale à : 3.76. Indiquer le résultat du test :

B - Un troisième groupe constitué de 10 sujets (adultes âgés sélectionnés dans une maison de retraite parce qu'ils ont obtenu les meilleures performances à un pré-test) vient participer à l'expérience "Mémoire".

Voici ses résultats : $m^{g^3} = 100.100$ $V^{g^3} = 409.490$

Par ailleurs, on a calculé pour les 3 groupes de sujets (Adultes/Enfants/Adultes âgés) :

$V_{inter} = 977.915$ $V_{intra} = 490.539$

1) Comment pourrait-on trouver la moyenne générale des 3 groupes (indiquer seulement la procédure, sans effectuer de calcul) ?

2) Analyse descriptive

a) On aimerait savoir si le facteur "Âge" (sur les 3 groupes) peut expliquer descriptivement les différences individuelles. Calculer la statistique appropriée pour répondre à cette question (donner le nom de la statistique, sa formule, le détail des calculs et le résultat) :

b) Interpréter ce résultat :

3) Analyse inférentielle

a) Quelle procédure pourrait-on envisager pour généraliser les résultats obtenus à une population plus vaste (indiquer seulement le nom du test) ?

b) Quelles sont les conditions de validité de ce test ?

c) Répondre à la question inférentielle : peut-on généraliser les résultats à une population plus vaste ?

- Extrait de la table des valeurs critiques de la variable χ^2 :

a	.05	.01	.001
1	3.84	6.63	10.83
2	5.99	9.21	13.82
3	7.81	11.34	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.52
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.12
9	16.92	21.67	27.88
10	18.31	23.21	29.59

- Extrait de la table des valeurs critiques de la variable T de Student :

a/2	.025	.005	.0005
a	.05	.01	.001
ddl			
5	2.571	4.032	6.869
6	2.447	3.707	5.959
7	2.365	3.499	5.408
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587
11	2.201	3.106	4.437
12	2.179	3.055	4.318
13	2.160	3.012	4.221
14	2.145	2.977	4.140
15	2.131	2.947	4.073

Extrait de la table des valeurs critiques de la variable F de Snedecor :

ddl1	!	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ddl2	a															
19	.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23
	.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15
	.001	15.1	10.2	8.28	7.27	6.62	6.18	5.85	5.59	5.39	5.22	5.08	4.97	4.87	4.78	4.70
20	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09
	.001	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.94	4.82	4.72	4.64	4.56
21	.05	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20	2.18
	.01	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.12	3.07	3.03
	.001	14.6	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.81	4.70	4.60	4.51	4.44
22	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07	3.02	2.98
	.001	14.4	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.70	4.58	4.49	4.40	4.33

- Quelques formules...

$$t_{obs} = \sqrt{n-2} \times \sqrt{\frac{V_{inter}}{V_{intra}}}$$

$$t_{obs} = \sqrt{n-2} \times \sqrt{\frac{r^2}{1-r^2}}$$

$$F = \frac{(n-G)}{(G-1)} \times \frac{V_{inter}}{V_{intra}}$$

$$ddl = G-1 \quad ddl = n-G$$

$$Cov(x,y) = \frac{\sum_i (x^i y^i)}{n} - \bar{x} \bar{y}$$

$$a = \frac{Cov(x,y)}{Var x} = r \times \frac{Ety y}{Ety x}$$

$$b = \bar{y} - a \bar{x}$$

$$t^{jk} = \frac{n_{jk} - \hat{n}_{jk}}{\hat{n}_{jk}}$$

$$c^2_{obs} = \frac{(n_{+-} - n_{-+})^2}{n_{+-} + n_{-+}}$$