Corrigé MEDI

Durée de l'épreuve : 2 heures
Aucun document n'est autorisé. Seule la calculette (sans sa documentation) est autorisée.
Attention : les exercices (encadrés) sont indépendants. Le barème donné à titre indicatif est sur 60 ; la note finale sera donnée sur 20.
La page 8 (tables et formules) peut être détachée et conservée. Indiquer les réponses exclusivement sur ce document. **Indiquer les réponses exclusivement sur ce document (ne pas écrire dans la marge).**

Dossier Stress&Tumeur (14 points)

Source : données fournies par Alexandre-Nicolas Menchikoff, Laboratoire de Psychologie Expérimentale & équipe enseignante Metus-Medi, Université René Descartes, Paris5.

Lors d'une recherche médicale effectuée sur 93 rats, échantillonnés au hasard, une équipe de chercheurs (Visintainer *et al.*, 1982) a soumis un premier groupe de 30 rats à des chocs électriques qu'ils ne pouvaient éviter. Un deuxième groupe de 30 rats avait la possibilité d'éviter ces chocs en soulevant une trappe qui leur permettait de passer dans une autre cage. Un troisième groupe de 33 animaux ne reçoit aucun choc. Ces 93 animaux ayant tous subi l'implantation d'une tumeur, les auteurs ont dénombré le nombre de rejets spontanés des tumeurs pour chacun des groupes. Le tableau ci-dessous résume les observations recueillies selon les deux variables étudiées chez ces rats : type de réaction de l'organisme vis à vis de la *Tumeur* (rejet/non rejet) et *Niveau de stress* subi (choc inévitable/choc évitable/pas de choc).

Tableau 1 : Effectifs observés

Stress / Tumeur	Choc inévitable	Choc évitable	Pas de choc	Total
Rejet	8	19	18	45
Non-rejet	22	11	15	48
Total	30	30	33	93

Question initiale *: existe-t-il une relation entre le niveau de stress subi et la capacité de l'organisme à se défendre contre certaines tumeurs ?*

A - Analyse descriptive

1) On se demande tout d'abord, avant l'analyse de la liaison, quelle est la réaction à la Tumeur la plus fréquemment observée, indépendamment du stress subi (tous les niveaux de stress confondus).

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

Rejet : $\frac{45}{93} \times 100 = 48\%$ *Non rejet :* $\frac{48}{93} \times 100 = 52\%$

b) Commenter les résultats obtenus :

Le Non rejet de la tumeur est plus fréquemment observé (52%) que le Rejet de celle-ci (48%)

2) Le tableau des effectifs théoriques est le suivant :

Tableau 2 : Effectifs théoriques

Niveau Stress Tumeur	Choc inévitable	Choc évitable	Pas de choc
Rejet	14.52	14.52	15.97
Non-rejet	15.48	15.48	17.03

a) Comment a-t-on obtenu la valeur 15.97 (en haut à droite de ce tableau 2)? Donner la formule utilisée ainsi que le détail des calculs.

$$\hat{n}_{jk} = \frac{n_j \times n_k}{n} = \frac{45 \times 33}{93} = 15.97$$

b) Donner la signification de cet effectif théorique (15.97) en langage naturel :

Si la réaction à la Tumeur et le niveau de Stress étaient indépendants, on aurait observé 16 (15.97) rats dont l'organisme aurait rejeté la tumeur et qui n'auraient subi aucun choc électrique.

3) Le tableau des taux de liaison est le suivant :

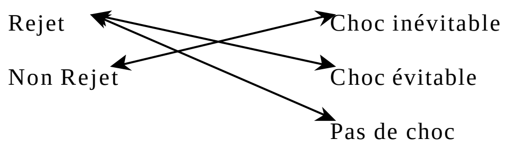
Tableau 3 : Taux de liaison

Niveau Stress Tumeur	Choc inévitable	Choc évitable	Pas de choc
Rejet	-0.45	0.31	0.13
Non-rejet	0.42	-0.29	-0.12

a) Comment a-t-on obtenu la valeur 0.13 (en haut à droite de ce tableau 3) ? Donner la formule utilisée ainsi que le détail des calculs :

$$t^{jk} = \frac{n_{jk} - \hat{n}_{jk}}{\hat{n}_{jk}} = \frac{18 - 15.97}{15.97} = 0.13$$

b) Représenter ci-après le graphe des attractions :



c) Commenter ce graphe des attractions :

On peut constater que l'organisme a tendance a rejeter la tumeur lorsque les rats subissent un niveau de stress moindre ou pas de stress (choc évitable/pas de choc). Par ailleurs, l'organisme a tendance à ne pas rejeter la tumeur lorsque le niveau de stress est élevé (choc inévitable).

B – Analyse descriptive : analyse globale de la liaison entre les deux variables

1) On trouve $\Phi^2 = 0.0952$. Interpréter cette valeur :

On interprète Phi² à l'aide du V² de Cramer : $\frac{f^2}{f^2_{Max}}$ *avec* F^2_{max} *: plus petite dimension du tableau moins une. Ici, la plus petite dimension est égale à 2, donc* $F^2_{max} = 1$

Donc $V^2 = F^2 = 0.0952$.

La liaison entre les deux variables Stress et Tumeur est moyenne (0.04 < V² < 0.16).

2) Rédiger une conclusion descriptive :

Pour les 93 rats observés, on peut constater qu'il existe une relation entre le niveau de stress subi et la tendance de l'organisme à rejeter ou non la tumeur. Cette liaison est considérée comme moyenne (V² de Cramer compris entre 0.04 et 0.16).

C- Analyse inférentielle.
On se demande si on peut généraliser les résultats obtenus sur ces 93 rats à l'ensemble des rats ayant les mêmes caractéristiques que celles observées sur cet échantillon. Pour ce faire, on va utiliser un test statistique.
1) Indiquer ci-après le nom du test à utiliser, sa formule, ainsi que le nombre de degré de liberté (ddl) qui lui est associé :
Test du Khi² (c^2).
$c^2 = n \times f^2$.
Nombre de ddl : (J – 1) ´ (K – 1) = (2 – 1) (3 – 1) = 2

2) Mise en œuvre du test.
a) Développer le calcul de ce test :
$c^2 = 93 \times 0.0952 = 8.85$

b) Donner le résultat du test :
$c^2_{obs} = 8.85 > c^2_{[2].05} = 5.99$. *Résultat significatif au seuil .05*

c) Rédiger une conclusion inférentielle :
Pour l'ensemble des rats ayant les mêmes caractéristiques que celles observées sur l'échantillon, on peut conclure qu'il existe une relation entre le niveau de stress subi et la tendance de l'organisme à rejeter ou non la tumeur (test du Khi² significatif au seuil .05)

Dossier Mémoire&Sémantique (16 points)

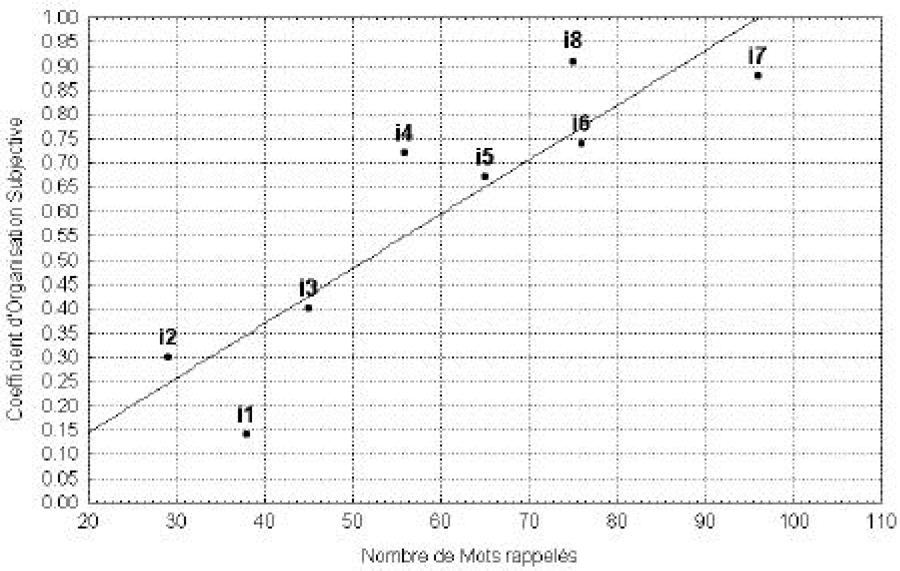
Source : données fournies par Alexandre-Nicolas Menchikoff, Laboratoire de Psychologie Expérimentale & équipe enseignante Metus-Medi, Université René Descartes, Paris 5.

Lors d'une expérience portant sur la mémorisation d'une liste de 112 mots appartenant à diverses catégories sémantiques : instruments, minéraux, animaux etc., on demande à 8 sujets adultes de langue maternelle française, échantillonnés au hasard, de rappeler dans l'ordre qu'ils veulent (rappel libre) le plus de mots possible. On retient 2 variables dépendantes (VD) : le nombre de mots rappelés en 10 minutes (*Variable X*) et un coefficient d'organisation subjective (*Variable Y*) susceptible de varier entre 0 et 1 selon la qualité des regroupements catégoriels effectués. Ce coefficient vaut zéro si les catégories dont relèvent les mots sont restituées dans un ordre aléatoire ; il vaut 1 si les mots sont tous regroupés par catégorie sémantique, quel que soit le nombre d'items rappelés pour chacune d'entre elles. Le tableau ci-dessous donne les résultats obtenus par les 8 sujets.

	s1	s2	s3	s4	s5	s6	s7	s8
Nombre de mots rappelés (*Variable X*)	38	29	45	56	65	76	96	75
Coefficient d'organisation (*Variable Y*)	0.14	0.30	0.40	0.72	0.67	0.74	0.88	0.91

Question initiale *: existe-t-il une relation entre le nombre de mots rappelés et la qualité des regroupements catégoriels effectués ?*

1) Construire le nuage bivarié ci-après (ne pas oublier les légendes) :



2) Commenter ce nuage bivarié :
Le nuage est allongé (linéaire) et ascendant, ce qui laisse supposer que la relation entre les 2 variables est forte et positive. Plus le nombre de mots rappelés est élevé, et plus le sujet a tendance à les regrouper en catégories sémantiques.

3) 3) Le coefficient de corrélation linéaire (r_{bp}) est égal à 0.89. Donner la valeur de ce coefficient arrondie à 4 décimales :
$r_{bp} = 0.8925$

4) Déduire de la valeur calculée en 3) la part de variance de la variable *"Coefficient d'organisation"* prise en compte par la régression linéaire de *"Coefficient d'organisation"* sur *"Nombre de mots rappelés"*. Indiquer le nom du coefficient utilisé ainsi que sa formule. Donner la valeur arrondie à 2 décimales.
$R^2 = (r)^2 = (0.8925)^2 = 0.80.$

5) A partir de ces indices (cf. calculs effectués en 3) et 4)), que peut-on dire de la liaison entre *"Nombre de mots rappelés"* et *"Coefficient d'organisation"* pour ces 8 sujets ?
Pour ces 8 sujets, on peut constater l'existence d'une liaison linéaire forte et positive entre les variables "Nombre de Mots rappelés" et "Coefficient d'Organisation Subjective". Le coefficient de corrélation linéaire est égal à +0.89 et cette valeur peut être considérée comme importante (> .40) et la part de variance prise en compte par la régression est considérée également comme importante (R^2 = .80 > .16). Les sujets qui fournissent un grand nombre de Mots rappelés ont tendance à les organiser en classes sémantiques.

6) Calculer et reporter ci-après l'équation de la droite de régression de "*Coefficient d'organisation*" sur "*Nombre de mots rappelés*" (2 décimales) :
$\tilde{y} = 0.01x - 0.08$

7) Représenter cette droite sur le nuage de points ci-dessus.

8) Interpréter la valeur du coefficient de régression a :
Quand le nombre de mots rappelés augmente d'une unité (de 1), le coefficient d'organisation subjective augmente de 0.01.

9) Analyse inférentielle. Que peut-on inférer pour l'ensemble des adultes de langue maternelle française ?

a) Indiquer le nom du test que l'on va utiliser, ainsi que sa formule, pour répondre à cette question et développer le calcul ci-après :

Test T de Student : $t_{obs} = \sqrt{n-2} \times \frac{r}{\sqrt{1-r^2}} = \sqrt{8-2} \times \frac{0.8925}{\sqrt{1-0.7965}} = 4.85$

b) Indiquer le résultat du test :

$t_{obs} = 4.85 > t_{[6].01} = 3.707$. *Résultat significatif au seuil unilatéral .005*

c) Rédiger une conclusion inférentielle :

Pour l'ensemble des adultes de langue maternelle française, on peut inférer qu'il existe une liaison positive entre le nombre de mots rappelés et les regroupements sémantiques effectués (test T de Student significatif au seuil .005 unilatéral)

Dossier Maternelle (13 points)

Source : D'après Hays, W.L. (1994). ***Statistics****. Fort Worth : Harcourt Brace College Publishers (Fifth edition).*

Quinze enfants âgés de 7 ans, échantillonnés au hasard dans une ville anglaise, ont été répartis de manière équilibrée en trois groupes de 5 enfants, selon le nombre d'années de préscolarisation effectuées (ou non) en maternelle :

g1 : pas de préscolarisation en maternelle, *g2* : 1 an de préscolarisation, *g3* : plus de 1 an de préscolarisation. Chaque enfant a passé une épreuve de "maturité sociale" (notée sur 20).

On aimerait savoir s'il existe une différence de performances à l'échelle de maturité entre ces trois groupes. Les données sont les suivantes :

g1	4	8	11	5	3
g2	9	14	15	10	12
g3	15	18	20	10	14

$m^{g1} = 6.20$ $\quad V^{g1} = 8.56$

$m^{g2} = 12.00$ $\quad V^{g2} = 5.20$

$m^{g3} = \mathbf{15.40}$ $\quad V^{g3} = \mathbf{11.84}$

Les résultats seront donnés arrondis à 2 décimales

1) Calculer la moyenne et la variance pour le groupe *g3* et reporter les valeurs ci-dessus (il n'est pas nécessaire de donner le détail des calculs).

2) Calculer la moyenne générale des trois groupes (donner le détail des calculs) :

$$Moy = \frac{(6.20+12+15.40)}{3} = \frac{(5\times6.20)+(5\times12)+(5\times15.40)}{15} = 11.20$$

3) Calculer la variance intra (donner la formule utilisée et le détail des calculs) :

$$Vintra = \frac{(8.56+5.20+11.84)}{3} = \frac{(5\times8.56)+(5\times5.20)+(5\times11.84)}{15} = 8.53$$

4) On trouve *Vinter* = 14.43. Calculer η² (donner la formule utilisée et le détail des calculs) :

$$h^2 = \frac{V\ inter}{V\ tot} = \frac{14.43}{22.96} = 0.63$$

5) Rédiger une conclusion descriptive :

Pour ces 15 sujets, la préscolarisation explique 63% des différences individuelles. Ce qui peut être considéré comme important (h² = 63% > 16% - valeur-repère). Si l'on considère les moyennes, descriptivement parlant, on constate que c'est le groupe qui a suivi plus d'un an de préscolarisation en maternelle (mg3 = 15.40) qui est le plus performant (mature) et le groupe qui n'a pas suivi de préscolarisation (mg1 = 6.20) est le moins performant (ou toute idée de comparaison de moyenne).

6) Analyse inférentielle. Que peut-on inférer quant à la population parente ?

a) Quelle est la population parente ?

Les enfants anglais (de la ville de passation) âgés de 7 ans

b) Pour répondre à la question inférentielle, on utilise le test F de Fisher-Snedecor. On trouve : $F_{obs} = 10.14$. Donner le nombre de degrés de liberté (ddl) associé à ce test :

ddl1 = K - 1 = 3 - 1 = 2
ddl2 = n - K = 15 - 3 = 12

c) On trouve dans la table des valeurs critiques du F de Fisher-Snedecor, pour le nombre de degrés de liberté (ddl) associé à cette statistique :

.05	**3.89**
.01	**6.93**
.001	**13.00**

Indiquer le résultat du test :

$F_{obs} > F_{[2,12].01} = 6.93$; Résultat significatif au seuil .01.

d) Rédiger une conclusion inférentielle :

Il existe une différence de performances à l'échelle de maturité sociale dans la population parente (F significatif au seuil .01)
(ou chez les enfant de 7 ans)

Dossier Régime (7 points)

Source : D'après Hays, W.L. (1994). ***Statistics****. Fort Worth : Harcourt Brace College Publishers (Fifth edition).*

Lors d'une étude sur les régimes effectués dans le but de perdre du poids, 15 couples (maris et femmes) ont été testés. On a relevé leur poids (en kg) avant régime, puis 2 semaines après. On aimerait savoir si les femmes perdent plus de poids que les hommes. Ci-dessous, les résultats obtenus après 2 semaines de régime (perte de poids en kg) :

$m^h = 3.7$ $V^h = 2.1$
$m^f = 4.9$ $V^f = 0.8$

Critère sémantique *: on jugera un effet important s'il est supérieur à 1kg, faible s'il est inférieur à 0.5 kg.*

1) Calculer l'effet moyen (d_{obs}) du facteur "Sexe" :

$d_{obs} = 4.9 - 3.7 = 1.2$ kg

2) Rédiger une conclusion descriptive :

Chez ces 15 couples, on peut constater que la perte en kg est plus importante chez les femmes (4.9 kg) que chez les hommes (3.7 kg). L'effet du régime peut être considéré comme important en regard du critère sémantique (dobs = 1.2kg > 0.5kg).

3) On trouve *Vintra* = 1.468 et *Vtotale* = 1.824. En déduire *Vinter* (donner le détail des calculs) :

Vinter = = Vtot - Vintra = 1.824 - 1.468 = 0.356

4) Analyse inférentielle. On se demande si on peut généraliser ces résultats à l'ensemble des couples ayant les mêmes particularités que ceux testés dans cette étude ?

a) Indiquer le nom et la formule du test que l'on pourrait mettre en œuvre ainsi que le nombre de degrés de liberté qui lui est associé :

Test T de Student : $t_{obs} = \sqrt{n-2} \times \sqrt{\frac{V\,inter}{V\,intra}}$

ddl = n -2 = 30 - 2 = 28

b) Est-il pertinent ici de généraliser ces résultats ? Pourquoi ? (justifier).

Non, car l'échantillon n'a pas été tiré au hasard.

Dossier Décomposition de la variance (10 points)

1) Donner une définition de la variance intra :

C'est la moyenne des variances

2) Donner une définition de la variance inter :

C'est la variance des moyennes

3) Proposer des modifications des données (fictives) ci-dessous de manière à modifier les valeurs de certaines statistiques.

Il est inutile de recalculer les nouvelles valeurs des statistiques après les modifications. Par contre, <u>expliquer pour chaque cas pourquoi vous avez procédé ainsi</u>.

a) Modifier une, ou s'il le faut plusieurs valeur(s) de *g1* de manière à diminuer la variance intra (sans se préoccuper de la variance inter):

g1	12	8	11	9	*g2*	19	15	12	14	12	18
	11										

Pour diminuer la variance intra, il suffit de diminuer la variance de l'un des groupes. Pour cela, il suffit de rapprocher de la moyenne une des valeurs de ce groupe (diminuer une valeur supérieure à la moyenne, ce que l'on a fait ici : 12 est devenu 11, ou augmenter une valeur inférieure à la moyenne).

b) Modifier une, ou s'il le faut plusieurs valeur(s) de *g1* de manière à augmenter la variance inter (sans se préoccuper de la variance intra) :

g1	12	8	11	9	*g2*	19	15	12	14	12	18
	11										

La moyenne de g1 étant inférieure à la moyenne de g2, on diminue une valeur de g1 (12 devient 11, par exemple) de manière à diminuer la moyenne de ce groupe. En augmentant l'écart entre les moyennes, on augmente la variance inter.

c) Modifier une, ou s'il le faut plusieurs valeur(s) de *g1* de manière à augmenter la variance inter <u>sans modifier la variance intra</u>.

g1	12	8	11	9	*g2*	19	15	12	14	12	18
	11	*7*	*10*	*8*							

On sait que la variance est invariante si l'on ajoute (ou retranche) une même valeur à toutes les valeurs. En diminuant toutes les valeurs de g1 de 1 point, on ne change pas sa variance, et on ne change donc pas la variance intra. Par contre, on augmente la variance inter en augmentant de 1 point l'écart entre les moyennes.

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

a ddl	.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** a ddl	.025 .05	.005 .01	.0005 .001
1	12.706	63.657	636.619
2	4.303	9.925	31.599
3	3.182	5.841	12.924
4	2.776	4.604	8.610
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

- Quelques formules... $t_{obs} = \sqrt{n-2} \times \sqrt{\frac{V_{inter}}{V_{intra}}}$ $\qquad$ $\mathbf{t_{obs}} = \sqrt{n-2} \times \frac{r}{\sqrt{1-r^2}}$

$$F = \frac{(n-K)}{(K-1)} \times \frac{V_{inter}}{V_{intra}} \qquad dd1 = K - 1 \qquad ddl2 = n - K$$

$$r = \frac{Cov(x, y)}{Ety\ x \quad Ety\ y} \qquad Cov(x, y) = \frac{\sum_i (x^i\ y^i)}{n} - \bar{x}\ \bar{y} \qquad a = \frac{Cov(x, y)}{Var\ x} = r \times \frac{Ety\ y}{Ety\ x} \qquad b = \bar{y} - a\,\bar{x}$$

$$t^{jk} = \frac{n_{jk} - \hat{n}_{jk}}{\hat{n}_{jk}} \qquad F^2 = \sum_{J,K} Cta_{jk} \qquad c^2_{obs} = \frac{(|n_{+-} - n_{-+}| - 1)^2}{n_{+-} + n_{-+}}$$

Les différents calculs ont été réalisés à l'aide des logiciels Statistica et DS3-Win