

Durée de l'épreuve : 2 heures

Aucun document n'est autorisé. Seule la calculatrice (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.

**Les pages 9 (données "Erreur Humaine") et 10 (tables et formules) peuvent être détachées et conservées.** Indiquer les réponses exclusivement sur ce document.

### Dossier Cours (4 points)

1) Donner le Cadre de Justification et d'Interprétation (CJI) de l'ANOVA :

2) Que veut dire le sigle ANOVA ?

3) Donner précisément le nom du test d'hypothèse que l'on met en oeuvre pour une ANOVA :

4) Représentation vectorielle des corrélations : on définit le coefficient de corrélation linéaire par le cosinus de l'angle entre deux variables centrées. Si cet angle est égal à  $0^\circ$ , quelle serait la valeur du coefficient de corrélation linéaire ?

5) Lorsque l'on veut analyser descriptivement un tableau de contingence, on utilise, en général, deux approches méthodologiques complémentaires. Indiquer et expliquer brièvement en quoi consistent ces deux analyses :

6) Soit le plan d'expérience suivant :  $S_{10} \langle A_3 * N_2 \rangle * P_2 * J_2 * T_2$ .

Déterminer le nombre de relations binaires ne faisant pas intervenir le facteur A (indiquer la procédure, le détail des calculs et le résultat) :

### Dossier Croyances (12 points)

Source : Stoetzel, J. (1983). *Les valeurs du temps présent*. Paris : PUF. Données citées, pp. 345-346 dans : Werber, B. (1994). *Les Thanatonautes*. Paris : Albin Michel, coll. Livre de Poche.

Dans le tableau 1, ci-après, les résultats d'un sondage effectué sur 300 personnes, choisies au hasard en Europe, de confessions différentes (100 personnes par confession<sup>1</sup>). Chaque sujet devait indiquer s'il croyait : à la Vie après la mort, au Paradis, à l'Enfer, à la Réincarnation, à l'Âme dissociée du corps, en Dieu. Chaque personne pouvait donner plusieurs réponses.

**Question initiale** : *les croyances différent-elles selon les confessions ?*

<sup>1</sup> La catégorie "Non Praticant" regroupe les confessions catholique et protestante.

**Tableau 1**

<i>Croyance</i>	Vie après la mort	Paradis	Enfer	Réincarnation	Âme dissociée du corps	Dieu	<i>Total</i>
<i>Confession</i>							
Catholique	52	45	30	23	66	87	303
Protestante	38	43	16	21	56	75	249
Non Praticant	13	8	3	12	24	23	83
<i>Total</i>	103	96	49	56	146	185	<b>635</b>

1) Que représente la valeur 635, à l'intersection des marges (donner sa notation statistique ainsi que son interprétation en langage naturel) ?

2) Analyse des marges : toutes confessions confondues, on obtient les résultats suivants :

**Tableau 2**

Vie après la mort	Paradis	Enfer	Réincarnation	Âme dissociée du corps	Dieu
16%	15%	8%	9%	23%	29%

Indiquer comment a été calculé le pourcentage de "Enfer" (Tableau 2). Présenter le pourcentage exact arrondi à 3 décimales ci-après :

3) Analyse descriptive de la liaison entre les variables "Confession" et "Croyance" : ci-après le tableau des profils présenté avec des valeurs arrondies à 1 décimale (Tableau 3).

**Tableau 3**

<i>Croyance</i>	Vie après la mort	Paradis	Enfer	Réincarnation	Âme dissociée du corps	Dieu	<i>Total</i>
<i>Confession</i>							
Catholique	<b>17.2%</b>	14.9%	9.9%	7.6%	21.8%	28.7%	100%
Protestante	15.3%	17.3%	6.4%	8.4%	22.5%	30.1%	100%
Non Praticant	15.7%	9.6%	3.6%	14.5%	28.9%	27.7%	100%
<i>Profil Moyen</i>	16.2%	15.1%	7.7%	8.8%	23.0%	29.1%	100%

a) Interpréter à l'aide d'une phrase la valeur 17.2% (en haut à gauche du tableau 3) :

b) Construire le tableau des attractions/répulsions (Tableau 4 à compléter ci-après) :

**Tableau 4**

<i>Croyance</i>	Vie après la mort	Paradis	Enfer	Réincarnation	Âme dissociée du corps	Dieu
<i>Confession</i>						
Catholique						
Protestante						
Non Praticant						

c) Commenter les attractions de ce tableau 4 :

d) On trouve  $\Phi^2 = 0.0192$ . Interpréter :

f) Formuler une conclusion descriptive qui réponde à la question initiale et qui tienne compte également des différents éléments analysés localement :

4) Analyse inférentielle : on aimerait pouvoir généraliser ces conclusions à une population plus vaste. On utilise pour ce faire la statistique  $\chi^2$ .

a) Calculer  $\chi^2$  et donner le nombre de degrés de liberté (ddl) associé à ce test :

b) Donner le résultat du test :

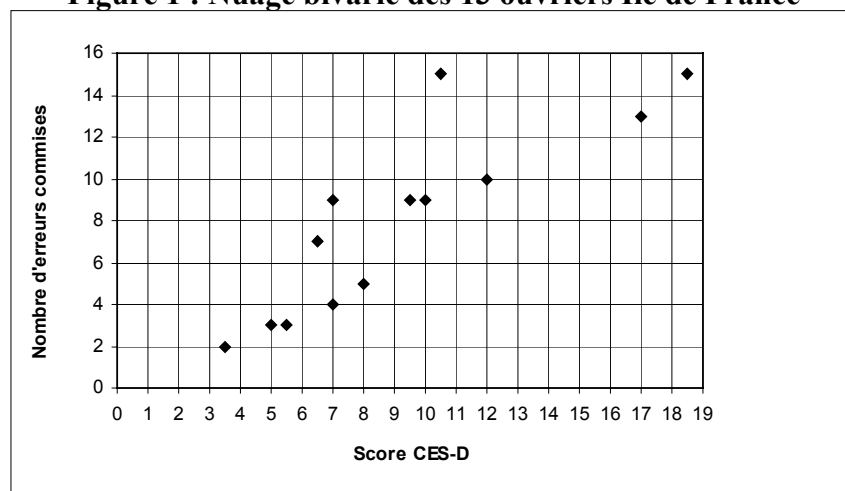
c) Elaborer une conclusion inférentielle :

**Dossier Erreur Humaine : 1<sup>ère</sup> hypothèse (13 points)**  
**Voir présentation des données et le tableau 1 page 9.**

Pour les questions suivantes, on s'intéresse uniquement aux 13 ouvriers de la région Île de France ( $rI$ ).

On fait l'hypothèse d'une relation entre le score CES-D et le nombre d'erreurs commises.

**Figure 1 : Nuage bivarié des 13 ouvriers Île de France**



1) Analyser ce nuage bivarié :

2) Calculer la corrélation entre "Score CES-D" et "Erreurs Commises". Indiquer le résultat arrondi à 4 décimales :

3) On trouve le résultat suivant :  $\frac{Var e}{Var y} = 0.265$  . Indiquer le nom du coefficient correspondant à cette formule et interpréter le résultat :

4) Quelle est la part de variance de "Erreurs Commises" prise en compte par "Score CES-D" ?

5) Élaborer une conclusion descriptive détaillée qui prenne en compte les indices calculés précédemment :

6) Donner l'équation de la droite de régression de "Erreurs Commises" sur "Score CES-D" (présentation à 2 décimales) :

7) Tracer cette droite sur la figure 1 (page précédente).

8) Repérer le sujet *s8* sur la figure 1 (page précédente) et représenter graphiquement son écart résiduel.

9) Voici les écarts réduits du sujet *s8* pour les deux variables étudiées :

	<i>Score CES-D</i>	<i>Erreurs Commises</i>
<i>s8</i>	<b>-0.519</b>	<b>+0.232</b>

Calculer la contribution à la corrélation du sujet *s8* (indiquer précisément la procédure de calcul, l'ampleur et le signe de cette contribution) :

10) On désire généraliser les conclusions trouvées sur cet échantillon de 13 personnes, à une population plus vaste.

a) Indiquer la procédure adoptée pour cette analyse inférentielle ainsi que le détail des calculs :

b) Indiquer le nombre de degrés de liberté (ddl) associé au test choisi :

c) Donner le résultat du test :

d) Élaborer une conclusion inférentielle :

**Dossier Erreur Humaine : 2<sup>ème</sup> hypothèse (10 points)**  
**Voir la présentation des données et les tableaux 1 à 4 page 9**

On fait l'hypothèse d'une différence de comportement des ouvriers selon leur région de travail, ( $r1$  : Île de France,  $r2$  : France Nord,  $r3$  : France Sud) tant au niveau des scores obtenus à la CES-D, qu'au niveau des erreurs commises.

**A - Analyse descriptive**

**Tableau 6**  
*r1 : Île de France*  
*(n = 13)*

	<i>Score CES-D</i>	<i>Erreurs Commises</i>
<i>Moy</i>	9.231	8.000
<i>Var</i>	18.447	18.615

**Tableau 7**  
*r2 : France Nord*  
*(n = 15)*

	<i>Score CES-D</i>	<i>Erreurs Commises</i>
	13.7 ..	15.067
		65.396

**Tableau 8**  
*r3 : France Sud*  
*(n = 15)*

	<i>Score CES-D</i>	<i>Erreurs Commises</i>
	7.867	6.733
	14.316	26.062

1) Pour la région  $r2$  (France Nord), calculer la moyenne et la variance à l'aide de la calculatrice puis compléter le tableau 6 ci-dessus en présentant les résultats arrondis à 3 décimales.

2) Pour ces 3 régions et pour la VD "Score CES-D", on trouve  $V_{intra} = 15.697$

a) Donner la définition de  $V_{intra}$  :

b) Pour ces 3 régions et pour la VD "Score CES-D", calculer  $V_{inter}$  et présenter le résultat arrondi à 3 décimales (donner la procédure de calcul et le développement) :

3) Pour ces 3 régions et pour la VD "Score CES-D", on trouve  $\eta^2 = 0.294$

Pour ces 3 régions et pour la VD "Erreurs commises", calculer  $\eta^2$  (indiquer sa formule) :

4) Interpréter  $\eta^2$  pour les deux VD :

5) Elaborer une conclusion descriptive, qui prenne en compte les deux VD :

### **B – Analyse inférentielle**

Pour la VD "Erreurs Commises", on aimerait généraliser les résultats obtenus sur cet échantillon de 43 sujets à l'ensemble des ouvriers travaillant dans les entreprises françaises de même enseigne. On a donc mis en oeuvre un test d'hypothèse.

1) Indiquer le nom du test utilisé, le nombre de degrés de liberté (ddl) qui lui est associé ici, développer et donner le résultat du calcul (présentation : 2 décimales) :

2) Indiquer le résultat du test :

3) Elaborer une conclusion inférentielle :

**Dossier Erreur Humaine : 3<sup>ème</sup> hypothèse ( 9 points)**  
**Voir la présentation des données et les tableaux 1, 3 et 5 page 9**

On fait l'hypothèse que les ouvriers de la région Île de France (*r1*) sont plus dépressifs que les ouvriers de la région France Sud (*r3*).

*Critère sémantique* : on jugera un effet important s'il est supérieur ou égal à 5 points, faible s'il est inférieur ou égal à 2 points

### **A - Analyse descriptive**

**Tableau 9**

	<i>r1 : Île de France (n = 13)</i>	<i>r3 : France Sud (n = 15)</i>
<i>Moy</i>	9.231	7.867
<i>Var</i>	18.447	14.316

1) Calculer l'effet moyen du facteur "Région" :

2) Se prononcer sur l'importance de cet effet :

3) On trouve  $\eta^2 = 0.028$ . Interpréter :

4) En s'appuyant sur les résultats donnés Tableau 5 de la page 9, expliquer pourquoi la valeur de  $\eta^2$  est proche de 0 :

5) Rédiger une conclusion descriptive :

### B - Analyse inférentielle

1) On trouve  $t_{obs} = 0.861$ . Indiquer la formule qui a permis de trouver ce résultat :

2) Indiquer le nombre de degrés de liberté (ddl) associé à ce test :

3) Le chercheur qui a mené cette recherche écrit :  $t_{obs} = 0.861$  ; *résultat non significatif*. Rédiger sa conclusion inférentielle :

### Dossier Formules (4 points)

Soit le tableau de contingence suivant où l'on a fait figurer les effectifs observés (ou effectifs conjoints) et les effectifs théoriques notés entre parenthèses.

	k1	k2	k3	Total
j1	10 (9)	5 (6)	15 (15)	30
j2	5 (6)	5 (4)	10 (10)	20
Total	15	10	25	50

Sur cet exemple :

1) Calculer  $f_{k1}$  et  $f_{j1}$  :

2) Calculer la fréquence théorique  $\hat{f}_{j1k1}$  :

3) Vérifier sur cet exemple que l'effectif théorique  $\hat{n}_{j1k1}$  vérifie la propriété

$$\hat{n}_{j1k1} = \frac{n_{j1} \times n_{k1}}{n} :$$

4) Montrer sur cet exemple que la fréquence théorique  $\hat{f}_{j1k1}$  vérifie la propriété

$$\hat{f}_{j1k1} = f_{j1} \times f_{k1} :$$

5) Dans quel objectif utilise-t-on cette formule :  $d^2(j, j') = \sum_K \frac{f_k^j - f_k^{j'}}{f_k} ?$

### Dossier Météo (4 points)

Source : d'après Frugier, G. (1992). Exercices ordinaires de probabilités. Paris : Marketing, coll. Ellipses.

A Saint-Tropez, il fait beau sept fois sur dix. Le comité des fêtes dispose d'une grenouille pour effectuer ses prévisions météorologiques. Cette grenouille se trompe régulièrement trois fois sur vingt quand il fait beau, et cinq fois sur vingt quand il pleut. Aujourd'hui, elle s'est trompée.

**Question** : *quel est le temps le plus probable aujourd'hui à Saint-Tropez ?*

1) Mettre les informations pertinentes de l'énoncé dans le tableau ci-après (ne pas oublier les intitulés des lignes et des colonnes) :

	C	D	Total
A			
B			

2) Appliquer le théorème de Bayes et répondre à la question :

### Dossier OGM (5 points)

Lors d'un premier test, on demande à 100 personnes de goûter une nouvelle barre chocolatée et de donner une appréciation soit positive (+), soit négative (-) du produit. Une semaine après, lors d'un second test, on fait goûter la même barre chocolatée aux mêmes personnes, sans leur divulguer qu'il s'agit du même produit que celui testé la semaine précédente. On leur indique seulement que le produit de ce second test est fabriqué à partir d'Organismes Génétiquement Modifiés (OGM), puis on leur demande une appréciation. Ci-après les résultats obtenus :

1 <sup>er</sup> test	2 <sup>nd</sup> test	Effectif
+	-	45
+	+	15
-	+	10
-	-	30

**Question** : la révélation de la présence des OGM dans la barre chocolatée a-t-elle un effet sur l'appréciation du produit ?

**Critère sémantique** : on jugera un effet important s'il est supérieur ou égal à 20 points de pourcentage, faible s'il est inférieur ou égal à 10 points de pourcentage.

1) Représenter ces données sous la forme d'un tableau de contingence :

2) Pour ces 100 sujets, évaluer l'effet "révélation de la présence des OGM" dans le produit :

3) Rédiger une conclusion descriptive :

4) Quel test pourrait-on mettre en oeuvre ici pour pouvoir généraliser les résultats obtenus à une population plus vaste non observée ?



## Données "Erreur Humaine"

Lors d'une étude effectuée en Ergonomie, des ouvriers travaillant sur une chaîne de montage dans une grande entreprise ont été évalués à l'aide du test de dépression CES-D (Center of Epidemiologic Studies - Depression Scale). On a ensuite relevé le nombre d'erreurs commises lors de l'exécution de leur tâche.

Quarante-trois ouvriers, choisis au hasard, et répartis en 3 groupes selon leur région de travail ont participé à cette expérience (pour  $r1$  : Île de France,  $n_{g1}=13$  ; pour  $r2$  : France Nord,  $n_{g2}=15$  , pour  $r3$  : France Sud,  $n_{g3}=15$ ). Ci-après sont présentés les résultats obtenus pour chaque groupe et pour chaque Variable Dépendante (VD) étudiée : "Score CES-D" (sur 20) et "Erreurs Commises".

**Tableau 1**

<i>Île de France</i>		
<i>Sujets</i>	<i>Score CES-D</i>	<i>Erreurs Commises</i>
S1	12.00	10
S2	10.00	9
S3	18.50	15
S4	17.00	13
S5	3.50	2
S6	5.00	3
S7	6.50	7
S8	7.00	9
S9	5.50	3
S10	7.00	4
S11	8.00	5
S12	9.50	9
S13	10.50	15

**Tableau 2**

<i>France Nord</i>		
<i>Sujets</i>	<i>Score CES-D</i>	<i>Erreurs Commises</i>
s14	17.00	25
s15	17.50	24
s16	13.50	17
s17	14.00	16
s18	15.00	12
s19	12.50	9
s20	10.00	8
s21	18.00	26
s22	17.00	23
s23	19.00	28
s24	13.50	7
s25	9.50	6
s26	7.50	6
s27	6.00	4
s28	16.00	15

**Tableau 3**

<i>France Sud</i>		
<i>Sujets</i>	<i>Score CES-D</i>	<i>Erreurs Commises</i>
s29	11.50	7
s30	4.00	3
s31	8.50	8
s32	11.00	8
s33	12.00	9
s34	10.50	15
s35	9.50	14
s36	7.00	3
s37	6.00	4
s38	5.50	4
s39	16.50	18
s40	4.50	2
s41	2.00	1
s42	4.50	2
s43	5.00	3

**Tableau 4**

<b>3 groupes</b> <i>(r1, r2 et r3)</i>	<i>Score CES-D</i>	<i>Erreurs Commises</i>
<i>Moy</i>	10.326	10.023
<i>Var</i>	22.220	51.418
<i>V<sub>inter</sub></i>		13.886
<i>V<sub>intra</sub></i>	15.697	37.532
$\eta^2$	0.294	

**Tableau 5**

<b>2 groupes</b> <i>(r1 et r3)</i>	<i>Score CES-D</i>
<i>Moy</i>	8.500
<i>Var</i>	16.696
<i>V<sub>inter</sub></i>	0.463
<i>V<sub>intra</sub></i>	16.234
$\eta^2$	0.028

**- Extrait de la table des valeurs critiques de la variable  $\chi^2$  :**

ddl	$\alpha$	.05	.01	.001
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 :**

ddl	$\alpha/2$ $\alpha$	.025 .05	.005 .01	.0005 .001
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	$\alpha$															
35	.05	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.07	2.04	2.01	1.99	1.96
	.01	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.80	2.74	2.69	2.64	2.60
	.001	12.9	8.47	6.79	5.88	5.30	4.89	4.59	4.36	4.18	4.03	3.90	3.79	3.70	3.62	3.55
40	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52
	.001	12.6	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.75	3.64	3.55	3.47	3.40
45	.05	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	2.01	1.97	1.94	1.92	1.89
	.01	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.67	2.61	2.55	2.51	2.46
	.001	12.4	8.09	6.45	5.56	5.00	4.61	4.32	4.09	3.91	3.76	3.64	3.53	3.44	3.36	3.29

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

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

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

$t^{jk} = \frac{n_{jk} - \hat{n}_{jk}}{\hat{n}_{jk}}$   $\Phi^2 = \sum_{J,K} Cta_{jk}$   $\chi^2_{obs} = \frac{(n_{+-} - n_{-+})^2}{n_{+-} + n_{-+}}$

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

$n! = 1 \times 2 \times 3 \times \dots \times n = n \times (n-1) \times (n-2) \times \dots \times 1$

$\binom{n}{a} = C_n^a = \frac{n!}{a!(n-a)!}$   $\binom{n}{a} = \binom{n}{n-a}$

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