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.
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 :
Le cadre fréquentiste

2) Que veut dire le sigle ANOVA ?
ANalysis Of Variance

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

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°, quelle serait la valeur du coefficient de corrélation linéaire ?
r = +1

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 :
- Analyse locale : comparaison des couples de modalités associées aux deux variables (ou comparaison des liaisons locales)
- Analyse globale : liaison globale entre les 2 variables

6) Soit le plan d'expérience suivant : S10 <A3*N2> * P2 * J2 * T2.
Déterminer le nombre de relations binaires ne faisant pas intervenir le facteur A (indiquer la procédure, le détail des calculs et le résultat) :

$$C_5^2 = \frac{n!}{a!(n-a)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1)(3 \times 2 \times 1)} = \frac{60}{6} = \frac{20}{2} = 10$$

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[1]). 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 diffèrent-elles selon les confessions ?*

Tableau 1

[1] La catégorie "Non Pratiquant" regroupe les confessions catholique et protestante.

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

1) Que représente la valeur *635*, à l'intersection des marges (donner sa notation statistique ainsi que son interprétation précise en langage naturel) ?
n = 635. C'est l'ensemble (le total) des réponses données par les 300 sujets

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

Tableau 2

Vie après mort	Paradis	Enfer	Réincar-nation	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 :
$(49/635) \times 100 = 7.717\%$

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

Croyance / *Confession*	Vie après la mort	Paradis	Enfer	Réincar-nation	Âme dissociée du corps	Dieu	*Total*
Catholique	**17.2%**	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 Pratiquant	15.7%	9.6%	3.6%	14.5%	28.9%	27.7%	*100%*
Profil Moyen	*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) :
Chez les catholiques, 17% des réponses font état d'une croyance de la vie après la mort.

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

Tableau 4

Croyance / *Confession*	Vie après la mort	Paradis	Enfer	Réincar-nation	Âme dissociée du corps	Dieu
Catholique	+	-	+	-	-	-
Protestante	-	+	-	-	-	+
Non Pratiquant	-	-	-	+	+	-

c) Commenter les attractions de ce tableau 4 :
Les catholiques ont tendance à croire à "la vie après la mort" et à l'Enfer.
Les protestants ont plus tendance à croire au Paradis et en Dieu.
Les personnes non pratiquantes ont plus tendance à croire à la réincarnation et "l'âme dissociée du corps".

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

Pour interpréter ϕ^2, il faut calculer le V^2 de Cramer $= \frac{\phi^2}{\phi^2_{Max}}$. Avec ϕ^2_{Max} : plus petite dimension du tableau moins 1. $V^2 = \frac{\mathbf{0.0192}}{\mathbf{2}} = \mathbf{0.0096}$ (ou .01)
La liaison entre "Confession" (ou religion) et "Croyance" est faible ($V^2 = .01 < .04$).

e) Formuler une conclusion descriptive qui réponde à la question initiale et qui tienne compte également des différents éléments analysés localement :
Les 635 réponses données par les 300 personnes interrogées indiquent que les catholiques auraient plus tendance à croire à la vie après la mort et à l'Enfer, les protestants au Paradis et à Dieu et les non pratiquants à la réincarnation et à l'âme dissociée du corps. Cependant les croyances ne semblent pas différer globalement selon la confession des sujets car le V^2 de Cramer indique une liaison faible entre les deux variables ($V^2 < 0.04$).

4) Analyse inférentielle : on aimerait pouvoir généraliser ces conclusions à une population plus vaste. On utilise pour ce faire la statistique χ^2.
a) Calculer χ^2 et donner le nombre de degrés de liberté associé à ce test :
$\chi^2 = n \times \phi^2$; ddl = (J -1) (K - 1). Donc : $\chi^2 = 635 \times 0.0192 = 12.19$ (valeur exacte : 12.17)
ddl = (3 -1) (6 - 1) = 10

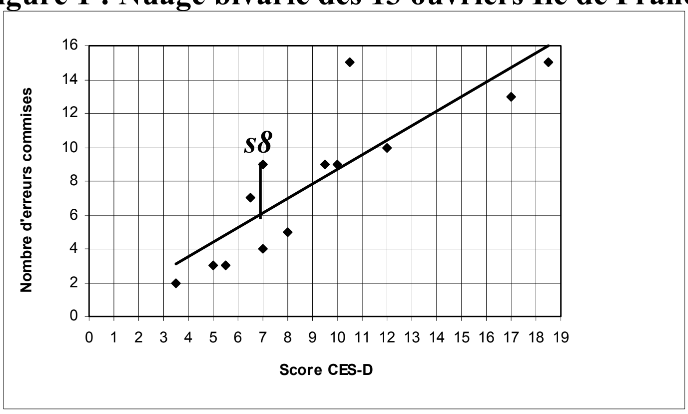
b) Donner le résultat du test :
$\chi^2_{obs} = 12.19 < \chi^2_{[10].05} = 18.31$; résultat non significatif au seuil .05

c) Élaborer une conclusion inférentielle :
D'une manière générale (ou dans la population parente), on ne peut pas conclure à une différence entre les croyances selon la confession des personnes (test du Khi² non significatif au seuil .05). On fait un constat d'ignorance.

Dossier Erreur Humaine : 1ère hypothèse (12points)
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 (*r1*).
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 :
Le nuage est ascendant, de forme allongée, ce qui laisse envisager une liaison assez forte et positive (plus la note au Score CES-D est élevée et plus le nombre d'erreurs augmente). On n'observe pas de valeur atypique, ni de sous-nuages.

2) Calculer la corrélation entre "Score CES-D" et "Erreurs Commises". Indiquer le résultat arrondi à 4 décimales :
$r = +0.8572$

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 :
Il s'agit du coefficient d'indétermination : 1 - R^2 = 0.265. 26.5% de la variance n'est pas prise en compte par la régression de "Erreurs Commises" sur "Score CES-D"

4) Quelle est la part de variance de "Erreurs Commises" prise en compte par "Score CES-D" ?
$R^2 = 1 - 0.265 = 0.735$ *ou* $R^2 = (r)^2 = (.8572)^2 = 0.735$

5) Élaborer une conclusion descriptive détaillée qui prenne en compte les indices calculés précédemment :
Pour ces 13 ouvriers, il existe une corrélation linéaire forte et positive entre le score au CES-D et le nombre d'erreurs commises : plus un sujet est dépressif et plus il aura tendance à commettre des erreurs. 73% (ou 74%) de la variance est prise en compte par la régression linéaire, ce que peut être considéré comme important (> 16%).(ou : r=0.86 important car >.40)

6) Donner l'équation de la droite de régression de "Erreurs Commises" sur "Score CES-D" (présentation à 2 décimales) :
$\tilde{y} = 0.86x + 0.05$ ou $\widetilde{Erreurs} = 0.86 CesD + 0.05$

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 :

	Score CES-D	***Erreurs Commises***
s8	***-0.519***	***+0.232***

Calculer la contribution à la corrélation du sujet *s8* (indiquer précisément la procédure de calcul, l'ampleur et le signe de cette contribution) :
zx×zy =1/13 × [(-0.519) × (+0.232)] = - 0.009. (Sa contribution est négative)

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 :

$$t_{obs} = \sqrt{n-2}\sqrt{\frac{r^2}{1-r^2}} = \sqrt{13-2}\sqrt{\frac{0.735}{1-0.735}} = 5.52(5.53)$$

b) Indiquer le nombre de degrés de liberté (ddl) associé au test choisi :
$n - 2 = 13 - 2 = 11$

c) Donner le résultat du test :

$t_{obs} = 5.52 > t_{[11].001} = 4.437$. *Résultat significatif au seuil .0005 unilatéral* .

d) Élaborer une conclusion inférentielle :

L'existence d'une corrélation positive entre le test CES-D et le nombre d'erreurs commises peut être généralisée à l'ensemble des ouvriers - à la population parente - (test t de Student significatif au seuil .0005 unilatéral).

Dossier Erreur Humaine : 2ème hypothèse. (9 points)
Voir la présentation des données et les tableaux 1 à 4 p. 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)***

	Score CES-D	*Nombre d'Erreurs*
Moy	9.231	8.000
Var	18.447	18.615

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

	Score CES-D	*Nombre d'Erreurs*
Moy	***13.733***	15.067
Var	***14.696***	65.396

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

	Score CES-D	*Nombre d'Erreurs*
Moy	7.867	6.733
Var	14.316	26.062

1) Pour la région *r2*, calculer la moyenne et la variance à l'aide de la calculette 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} :

V_{intra} = *moyenne (pondérée) des variances*

b) Pour ces 3 régions et pour la VD "Score CES-D", calculer V_{inter} et présenter le résultat arrondi à 2 décimales (donner la procédure de calcul et le développement) :

$$V_{inter} = \frac{13}{43} \times (9.231 - 10.326)^2 + \frac{15}{43} \times (13.733 - 10.326)^2 + \frac{15}{43} \times (7.867 - 10.326)^2 = 6.521$$

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 η^2 (indiquer sa formule) :

$$\eta^2 = \frac{V\ inter}{Vtot} = \frac{13.886}{51.418} = 0.270$$

4) Interpréter η^2 pour les deux VD :

Tant au niveau du score CES-D qu'au niveau des erreurs commises, la part de variance prise en compte par le facteur "Région" peut être considérée comme importante (dans les 2 cas, $\eta^2 > 0.16$, valeur repère)

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

Chez ces 43 sujets, les scores obtenus au CES-D ainsi que le nombre d'erreurs commises dans le travail, diffèrent de façon importante selon la région ($\eta^2 > 0.16$ pour les deux VD).
(les différences individuelles peuvent être expliquées par le facteur Région)

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) :

$$F_{obs} = \frac{n-K}{K-1} \times \frac{V\,inter}{V\,intra} = \frac{43-3}{3-1} \times \frac{13.886}{37.532} = 7.40$$

ddl1 = K -1 = 3 - 1 = 2 ; ddl2 = n - K = 43 - 3 = 40

2) Indiquer le résultat du test :

$F_{obs} = 7.400 > F_{[2,\,40].01} = 5.18$; *résultat significatif au seuil .01*

3) Élaborer une conclusion inférentielle :

D'une manière générale, les performances obtenues par (l'ensemble) les ouvriers différent selon la région d'appartenance. (ou Les résultats trouvés sur l'échantillon peuvent être donc généralisés à l'ensemble des ouvriers travaillant dans les entreprises françaises de même enseigne que celle de l'échantillon) (test F significatif au seuil .01).

Dossier Erreur Humaine : 3ème 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 - Scores CES-D

	r1 : Île de France (n = 13)	*r3 : France Sud (n = 15)*
Moy	9.231	7.867
Var	18.447	14.316

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

$d_{obs} = 9.231 - 7.867 = 1.364$

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

$d_{obs} = 1.36 < 2$. *L'effet est jugé faible selon le critère sémantique.*

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

$\eta^2 = 0.028 < 0.04$ *(valeur repère). Le facteur "Région" n'explique que 3% des différences individuelles. Ce qui jugé faible en regard du critère sémantique.*

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

Pour ces 2 régions, (quelle que soit la variance intra), la variance inter est proche de 0 (0.463), ce qui veut dire que les moyennes des 2 groupes sont proches (de la moyenne générale).

5) Rédiger une conclusion descriptive :

Pour ces 28 sujets, le score moyen obtenu à l'échelle CES-D est plus élevé en Île de France (9.23) qu'en France Sud (7.87). La différence est égale à 1.36 point et peut être qualifiée de faible selon le critère sémantique donné.

B - Analyse inférentielle

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

$$t_{obs} = \sqrt{n-2} \times \sqrt{\frac{V\,inter}{V\,intra}} = \sqrt{28-2} \times \sqrt{\frac{0.463}{16.234}} = 0.861$$

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

ddl = n - 2 = 28 - 2 = 26

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

Pour l'ensemble des ouvriers, on ne peut pas se prononcer sur l'existence d'une différence de scores à l'échelle CES-D (test T non significatif au seuil bilatéral .05). On fait un constat d'ignorance.

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} :

$$f_{k1} = \frac{15}{50} = 0.30 \qquad f_{j1} = \frac{30}{50} = 0.60$$

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

$$\hat{f}_{j1k1} = \frac{9}{50} = 0.18$$

3) Vérifier sur cet exemple que l'effectif théorique $\hat{n}_{j1k1}$ vérifie la propriété $\hat{n}_{j1k1} = \frac{n_{j1} \times n_{k1}}{n}$:

$$\hat{n}_{j1k1} = 9 = \frac{n_{j1} \times n_{k1}}{n} = \frac{30 \times 15}{50}$$

4) Montrer sur cet exemple que la fréquence théorique $\hat{f}_{j1k1}$ vérifie la propriété $\hat{f}_{j1} \times \hat{f}_{k1}$

$$\hat{f}_{j1k1} = \frac{\hat{n}_{j1k1}}{n} = \frac{9}{50} = 0.18 \text{ et } f_{j1} \times f_{k1} = \frac{30}{50} \times \frac{15}{50} = 0.60 \times 0.30 = 0.18$$

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

Pour apprécier la distance existante entre deux profils-lignes j et j'

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	
		La grenouille se trompe	*La grenouille a raison*	*Total*
A	*Il fait beau*	*0.15*	*0.85*	*.70*
B	*Il pleut*	*0.25*	*0.75*	*.30*

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

$$\frac{P(A)\times P(C/A)}{P(A)\times P(C/A)+P(B)\times P(C/B)}=\frac{.70\times .15}{(.70\times .15)+(.30\times .25)}=\frac{.105}{.18}=.583$$

Dans 58% des cas il fera beau même si la grenouille se trompe. Et dans 42% des cas, il pleuvra. Le temps le plus probable est donc le beau temps.

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 même 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^{er} test	2^{nd} 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 :

		2nd Test +	2nd Test -	
1er Test	**+**	**15**	**45**	**60**
	-	**10**	**30**	**40**
		25	**75**	**100**

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

$$f_{test1}=\frac{15+45}{100}=0.60 \qquad f_{test2}=\frac{15+10}{100}=0.25 \qquad dobs=.60-.25=.35$$

$$\text{ou } \frac{45-10}{100}=.35$$

L'effet est considéré important en regard du critère sémantique donné (.35 > .20)

3) Rédiger une conclusion descriptive

Pour ces 100 sujets, le pourcentage d'avis favorables a diminué entre le 1er et le 2nd test (60% au 1er et 25% au 2nd). La différence de 35 points de % peut être considérée comme importante en regard du critère sémantique (35 > 20).

4) Quel test pourrait-on mettre en oeuvre ici pour pouvoir généraliser les résultats obtenus à une population plus vaste non observée ? *Le test du Khi² corrigé de Mac Nemar. Mais on ne pourrait pas le mettre en oeuvre car on ne sait pas si les 100 personnes ont été choisies au hasard (condition d'échantillonnage au hasard nécessaire avant d'effectuer le test).*

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

	Île de France	
Sujets	*Score CES-D*	*Erreurs Commises*
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

	France Nord	
Sujets	*Score CES-D*	*Erreurs Commises*
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

	France Sud	
Sujets	*Score CES-D*	*Erreurs Commises*
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

***<u>3 groupes</u>* (r1, r2 et r3)**	***Score CES-D***	***Erreurs Commises***
Moy	10.326	10.023
Var	22.220	51.418
V_{inter}		13.886
V_{intra}	15.697	37.532
η^2	0.294	

Tableau 5

***<u>2 groupes</u>* (r1 et r3)**	***Score CES-D***
Moy	8.500
Var	16.696
V_{inter}	0.463
V_{intra}	16.234
η^2	0.028

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

ddl \ α	.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 :

α/2 α ddl	.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 ! / Ddl2	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
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... $t_{obs} = \sqrt{n-2} \times \sqrt{\dfrac{V_{inter}}{V_{intra}}}$ $\quad t_{obs} = \sqrt{n-2} \times \sqrt{\dfrac{r^2}{1-r^2}}$

$$F = \frac{(n-G)}{(G-1)} \times \frac{V_{inter}}{V_{intra}} \qquad ddl1 = G - 1 \qquad ddl2 = n - G$$

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

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

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

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

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

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