Tester les conditions d'application d'un test paramétrique

Tester la normalité d'une distribution

Exercice 1

1) Lors d'une expérience, les scores observés sur un échantillon de 8 sujets sont les suivants :

Suj	s1	s2	s3	s4	s5	s6	s7	s8
x_i	5	7	8	11	12	13	13	15

On veut étudier la normalité de la distribution des scores dans la population parente. A l'aide d'un tableur, on construit le tableau suivant :

X_i	Z_i	$F(X < x_i)$	$F(X <= x_i)$	Theo	Ecart -	Ecart +
5	-1,5877	0,000	0,125	0,0562	0,0562	0,0688
7	-1,0104	0,125	0,250	0,1562	0,0312	0,0938
8	-0,7217	0,250	0,375	0,2352	0,0148	0,1398
11	0,1443	0,375	0,500	0,5574	0,1824	0,0574
12	0,4330	0,500	0,625	0,6675	0,1675	0,0425
13	0,7217	0,625	0,875	0,7648	0,0148	0,1102
15	1,2990	0,875	1,000	0,9030	0,0280	0,0970
				Maximum	0,1824	0,1398

Justifier la construction de ce tableau et utiliser les tests de Kolmogorov-Smirnov et de Lilliefors pour apporter une réponse au problème posé.

2) Sur un échantillon prélevé au hasard dans une autre population, les scores observés sont :

Suj	s1	s2	s3	s4	s5	s6	s7	s8
x_i	5	5.5	5.5	6	14	16	16	17

Le tableau de calcul, partiellement rempli, est le suivant :

X_i	Z_i	$F(X < x_i)$	$F(X <= x_i)$	Theo	Ecart -	Ecart +
5	-1,0141	0	0,125	0,1553	0,1553	0,0303
5,5	-0,9239	0,125	0,375	0,1778	0,0722	0,1972
6	-0,8338	0,375	0,5	0,2022	0,1728	0,2978
14	0,6085	0,5	0,625	0,7286	0,2286	0,1036
16	0,9690	0,625	0,875			
17		0,875	1			
10,625	Moyenne			Maximum		
5,5469	Ec. type cor.					

Compléter ce tableau et utiliser de même les tests de Kolmogorov-Smirnov et de Lilliefors pour étudier la normalité de la variable dépendante étudiée.

Réponses : 1) Pour le test de K-S au seuil de 5%, la valeur critique est : $D_{crit} = 0.454$. Comme $D_{obs} < D_{crit}$, on conclut sur H_0 : on ne peut pas rejeter l'hypothèse de normalité des scores dans la population parente. Pour le test de Lilliefors au seuil de 5%, on a $L_{crit} = 0.285$, et la conclusion reste identique.

2) La valeur manquante dans la colonne Z_i est $\frac{17 - 10.625}{5.5469}$, c'est-à-dire 1.1493. Les valeurs manquantes de la colonne "Theo" doivent être lues dans une table de la loi normale. Ce sont respectivement : 0.8337 et 0.8748. Les écarts peuvent alors être déterminés en calculant les différences entre la colonne "Theo" et les deux colonnes précédentes. Le maximum des écarts est 0.2978. En gardant un seuil de 5%, les valeurs critiques sont identiques à celles de la question 1. On serait donc conduit à accepter la normalité des scores en utilisant le test de K-S, et à la refuser en utilisant le test de Lilliefors. Comme la moyenne et l'écart type de la distribution théorique ont été estimés à partir de l'échantillon, c'est cette dernière conclusion qu'il convient de retenir ici.

Exercice 2

Dans le cadre d'une étude sur l'acceptation de l'euthanasie en Europe, on a interrogé des échantillons de sujets dans 33 pays européens.
Les sujets devaient notamment répondre à la question :

"Pensez-vous que l'euthanasie (mettant fin à la vie des malades incurables) peut être toujours justifiée, ne peut jamais être justifiée, ou parfois justifiée ?"

en se situant sur une échelle de Likert en dix points, la valeur 1 correspondant à "jamais" et la valeur 10 à "toujours".
Les auteurs indiquent : "en raison des proportions relativement importantes de sujets qui ont répondu 1 ou 10, la variable dépendante ne suit pas une loi normale. Le test de Kolmogorov-Smirnov a confirmé cette absence de normalité".
Pour les Pays-Bas, les réponses des 994 sujets interrogés sont distribuées comme suit :

Réponse	1	2	3	4	5	6	7	8	9	10
Effectif	30	79	64	65	65	117	167	157	125	125

Vérifier l'absence de normalité de cette variable sur la population des Pays-Bas à l'aide des tests de Kolmogorov-Smirnov et de Lilliefors. On pourra utiliser les résultats suivants, dans lesquels X désigne la variable "Réponse" et Z la variable centrée réduite associée :

X_i	n_i	f_i	$F(X < x_i)$	$F(X <= x_i)$	Z_i	Theo	Ecart -	Ecart +
1	30	0.0302	0.0000	0.0302	-2.1415	0.0161	0.0161	0.0141
2	79	0.0795	0.0302	0.1097	-1.7529	0.0398	0.0096	0.0698
3	64	0.0644	0.1097	0.1740	-1.3642	0.0863	0.0234	0.0878
4	65	0.0654	0.1740	0.2394	-0.9755	0.1646	0.0094	0.0748
5	65	0.0654	0.2394	0.3048	-0.5869	0.2786	0.0392	0.0262
6	117	0.1177	0.3048	0.4225	-0.1982	0.4214	0.1166	0.0011
7	167	0.1680	0.4225	0.5905	0.1904	0.5755	0.1530	0.0150
8	157	0.1579	0.5905	0.7485	0.5791	0.7187	0.1282	0.0298
9	125	0.1258	0.7485	0.8742	0.9678	0.8334	0.0849	0.0408
10	125	0.1258	0.8742	1.0000	1.3564	0.9125	0.0383	0.0875

Commentez les résultats obtenus à l'aide d'un logiciel de traitements statistiques (xx désigne la série des 994 observations. :

```
> mean(xx)
[1] 6.506036
```

```
> skewness(xx)
[1] -0.4977117

> kurtosis(xx)
[1] -0.778448

> lillie.test(xx)
Lilliefors (Kolmogorov-Smirnov) normality test
data:  xx
D = 0.1535, p-value < 2.2e-16

> shapiro.test(xx)
Shapiro-Wilk normality test
data:  xx
W = 0.9282, p-value < 2.2e-16

> ad.test(xx)
Anderson-Darling normality test
data:  xx
A = 21.876, p-value < 2.2e-16

> dagoTest(xx)
Title:
 D'Agostino Normality Test
Test Results:
  STATISTIC:
    Chi2 | Omnibus: 108.8585
    Z3  | Skewness: -6.1239
    Z4  | Kurtosis: -8.4472
  P VALUE:
    Omnibus  Test: < 2.2e-16
    Skewness Test: 9.129e-10
    Kurtosis Test: < 2.2e-16
```

Exercice 3

Un chercheur a recueilli des données relatives à deux groupes indépendants de sujets. Avant de réaliser un test de Student, il souhaite tester la normalité de la variable dépendante dans les populations parentes. Plusieurs alternatives s'offrent à lui :

- Un premier collègue lui conseille de tester séparément les données relatives aux deux groupes ;
- Un deuxième collègue lui conseille d'effectuer le test en considérant la réunion des deux ensembles de données ;
- Un troisième collègue lui conseille de calculer dans chacun des deux groupes les écarts à la moyenne du groupe, puis d'effectuer un seul test sur la réunion de tous ces écarts.
- Un quatrième collègue lui conseille de prendre en compte les écarts réduits au lieu des écarts.

Quant à vous, quelle méthode conseilleriez-vous et pourquoi ?

Eléments de réponse. La plus mauvaise méthode est évidemment la 2è. Si la moyenne de la VD est différente dans les deux populations, la distribution sur la réunion des deux populations ne suit pas une loi normale. Dans la première méthode, on effectue deux tests indépendants, et notre conclusion dépend des résultats des deux tests. Le risque de commettre une erreur de type I ou de type II est donc plus élevé que lors de la réalisation d'un seul test. L'une des méthodes 3 et 4 pourrait donc sembler préférable, mais un manque de régularité de la VD dans l'une des populations pourrait se trouver masqué par une bonne régularité dans l'autre. La principale différence entre les méthodes 3 et 4 porte sur l'égalité des variances de la VD dans les populations parentes. Or, le test de Student suppose cette égalité. Le recours à la méthode 4 semble donc inutile. Les méthodes à privilégier seraient donc plutôt les méthodes 1 et 3, avec une préférence pour la méthode 3 si les échantillons sont de très faible effectif.

Exercice 4

Nurcombe et *al.* ont mené en 1984 une étude sur les enfants présentant un poids réduit à la naissance (PRN). Ces enfants posent des problèmes particuliers à leurs parents parce qu'ils sont, en apparence, apathiques et imprévisibles ; en outre, ils risquent de connaître des problèmes physiques et comportementaux. Les données dont on dispose portent sur deux groupes d'enfants ;

- Un groupe expérimental de 25 enfants PRN dont les mères bénéficiaient d'un apprentissage particulier : elles étaient sensibilisées aux signaux émis par ces enfants, afin de leur permettre de mieux répondre à leurs besoins ;
- Un groupe témoin de 31 enfants PRN dont les mères ne bénéficiaient d'aucun programme particulier ;

Il s'agit d'une part, de l'indice de développement mental (IDM) à 6 mois et à 24 mois pour le groupe témoin PRN, et d'autre part de l'IDM à 24 mois pour le groupe expérimental PRN.

On réalise des tests de normalité à l'aide de Statistica sur chacun de ces trois jeux de données. On obtient les résultats suivants :

– pour l'IDM à 6 mois dans le groupe PRN témoin

	Tests de Normalité (Enfants-PRN.sta dans Enfants-PRN.s					
Variable	N	D max	K-S p	Lillief. p	W	p
IDM-6	31	0,12975	p > .20	p > .20	0,96377	0,36577

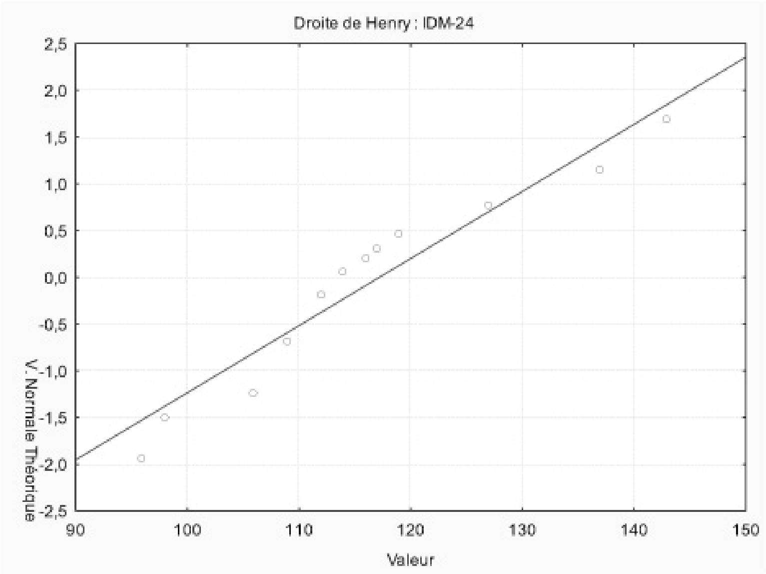
– pour l'IDM à 24 mois dans le groupe PRN témoin

	Tests de Normalité (Enfants-PRN.sta dans Enfants-PRN.s					
Variable	N	D max	K-S p	Lillief. p	W	p
IDM-24	31	0,10030	p > .20	p > .20	0,97346	0,61883

– pour l'IDM à 24 mois dans le groupe PRN expérimental

	Tests de Normalité (Enfants-PRN.sta dans Enfants-PRN.s					
Variable	N	D max	K-S p	Lillief. p	W	p
IDM-24	25	0,16356	p > .20	p < ,15	0,91464	0,03874

Commenter les résultats fournis par Statistica, ainsi que le graphique de la droite de Henry pour la troisième variable :



Réponse : Les trois tests indiquent une distribution normale des variables IDM-6 et IDM-24 dans la population dont est issu le groupe témoin. En revanche, le test de Shapiro-Wilk indique une absence de normalité de la variable IDM-24 dans la population dont est issu le groupe expérimental. Ce résultat n'est pas en accord avec ceux fournis par les deux autres tests, mais on sait que le test de Shapiro-Wilk est plus puissant que les deux autres.

Homogénéité des variances

Exercice 5 *Dossier "pedago"*
Lors d'une expérience pédagogique, on s'intéresse à l'effet comparé de deux pédagogies des mathématiques chez deux groupes de 10 sujets :

- pédagogie traditionnelle (p_1)
- pédagogie moderne (p_2)

On note la performance à une épreuve de combinatoire.

p_1 traditionnelle	
s1	5.0
s2	4.0
s3	1.5
s4	6.0
s5	3.0
s6	3.5
s7	3.0
s8	2.5
s9	1.5
s10	2.5

p_2 moderne	
s11	4.0
s12	5.5
s13	4.5
s14	6.5
s15	4.5
s16	5.5
s17	1.0
s18	2.0
s19	4.5
s20	4.5

1) Vérifier que les paramètres des deux échantillons sont donnés par :

	p_1	p_2
Moyenne	3.250	4.250
Ecart-type	1.365	1.553
Variance	1.863	2.413
Ecart-type corrigé	1.439	1.637
Variance corrigée	2.069	2.681

2) Avant d'appliquer un test de comparaison de moyennes, on veut s'assurer que l'on peut supposer les variances égales dans les populations parentes. Procéder à un test de comparaison de variances permettant de s'en assurer.
Réponses : 2) On obtient $F_{obs} = 1.30$. Or, pour $ddl_1 = 9$, $ddl_2 = 9$ et un seuil de 5%, on lit dans la table : $F_{crit} = 3.18$. L'hypothèse H_0 (égalité des variances) est donc retenue.

Exercice 6
Dans le cadre d'une analyse médicale, deux méthodes de dosage peuvent être utilisées. A partir d'un même prélèvement, on répète 25 fois la méthode A et 30 fois avec la méthode B. Les résultats sont rassemblés dans les tableaux ci-dessous.

Méthode A	
x_i (en g)	n_i
37	1
39	2
40	2
41	4
42	7
43	4
44	2
46	2
47	1
Total	25

Méthode B	
x_i (en g)	n_i
39	2
40	1
41	6
42	9
43	8
44	3
45	1
Total	30

1) Tester l'hypothèse : "les valeurs moyennes obtenues par les deux méthodes sont égales". (Autrement dit, les méthodes sont-elles exactes ?)
2) Comparer les variances des échantillons traités avec les deux méthodes à l'aide du test de Fisher. (Autrement dit, les deux méthodes ont-elles la même précision ?)
Réponses : 1) Les paramètres de statistiques descriptives sont donnés par :

	Méthode A	*Méthode B*
Moyenne	*42.08*	*42.10*
Variance	*4.95*	*1.89*
Variance corrigée	*5.16*	*1.96*

Le test de comparaison des deux moyennes (groupes indépendants) conduit à : $t_{obs} = -.04$, évidemment non significatif aux seuils traditionnels. On ne peut donc pas refuser l'hypothèse H_0 d'égalité des moyennes.
2) La statistique de test suit une loi de Fisher à $ddl_1 = 24$ et $ddl_2 = 29$ degrés de liberté. On obtient : $F_{obs} = 2.63$. Au seuil de 1% unilatéral, on a $F_{crit} = 2.49$. On conclut donc à une différence des variances.

Exercice 7

Au cours de certaines expériences, on est amené à mesurer le *temps de réaction* (TR) des sujets. C'est le temps qui s'écoule entre la présentation d'un stimulus (par exemple, une lampe qui s'allume devant le sujet) et la réaction que ce stimulus doit déclencher (par exemple, presser un bouton).

Première expérience. — Le tableau 1 fournit les TR d'une personne qui a réagi 20 fois à l'allumage d'une lampe rouge. On constate que ces 20 TR ne sont pas égaux. Ces variations d'un moment à l'autre sont imprévisibles à partir des informations dont on dispose dans l'expérience.

Deuxième expérience. — Le sujet voit maintenant s'allumer devant lui une lampe qui peut être rouge, verte ou jaune. il doit réagir si la lampe est rouge, mais ne doit pas réagir dans les deux autres cas. Le tableau 1 fournit 20 TR mesurés dans ces conditions. On observe de nouveau des variations imprévisibles d'un moment à l'autre.

Troisième expérience. — Les conditions sont les mêmes que dans la première expérience (une seule lampe) avec une seule différence : au lieu d'être rouge, la lampe donnant le signal de la réaction est verte. La troisième ligne du tableau donne les résultats. Les temps sont de nouveau différents entre eux.

Numéro d'ordre des 20 présentations	1	2	3	4	5	6	7	8	9	10
1ère expérence	20	15	18	25	17	32	18	17	19	23
2è expérience	32	40	33	37	35	29	42	62	50	39
3è expérience	16	18	19	18	15	18	17	32	23	19

Numéro d'ordre des 20 présentations	11	12	13	14	15	16	17	18	19	20
1ère expérience	19	21	15	22	17	17	21	19	17	23
2è expérience	45	47	52	37	38	39	40	41	42	39
3è expérience	23	20	18	25	15	15	17	23	17	19

1) La dispersion des TR est-elle la même dans chacune des trois conditions expérimentales ? Pour répondre à cette question, comparer deux à deux les variances des trois séries de données à l'aide du test de Fisher.

2) On teste globalement l'homogénéité des variances dans les trois conditions à l'aide des tests de Levene et de Brown-Forsythe. Interpréter les résultats fournis par Statistica :

Variable	Test de Levene d'Homogénéité des Variances (TR.sta) Effets significatifs marqués à p < ,05000							
	SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
TR	76,5613	2	38,2806	781,936	57	13,7181	2,79050	0,06979

Variable	Test d'Homogénéité des Variances de Brown-Forsythe (TR.sta) Effets significatifs marqués à p < ,05000							
	SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
TR	76,8000	2	38,4000	956,050	57	16,7728	2,28942	0,11057

Réponses : 1) Les variances des trois séries sont données par :

	Variance	*Variance corrigée*
1ère expérience	*14.89*	*15.67*
2è expérience	*53.85*	*56.68*
3è expérience	*16.23*	*17,08*

Pour $ddl_1 = 19$ *et* $ddl_2 = 19$ *et un seuil de 5%, on a :* $F_{crit} = 3.00$. *Ici,* $F_{2,1,obs} = 3.61$, $F_{2,3,obs} = 3.31$, $F_{3,1,obs} = 1.09$. *Pour les expériences 1 et 3, l'hypothèse nulle (même variance) peut être retenue. En revanche, l'expérience 2 conduit à une variance différente de celles des deux autres.*
2) Le niveau de significativité de chacun des deux tests est supérieur à 5%. On ne peut donc pas repousser l'hypothèse H_0*, c'est-à-dire l'homogénéité des variances dans les trois groupes.*

Exercice 8 On reprend l'exemple "boulimie" vu en cours. On rappelle ci-dessous les paramètres des données observées :

	Simple	Avec vom.
$\overline{x}_i$	4.61	-0.83
s_{ic}^2	219.04	79.21
n_i	49	32

1) Comparer les résultats observés dans les deux groupes à l'aide d'un test de Student, sans tenir compte d'éventuelles hypothèses sur les variances. La statistique de test à utiliser est alors :

$$t = \frac{\overline{x_1} - \overline{x_2}}{E} \quad \text{avec} \quad E^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

2) Rappeler le résultat du test de comparaison des variances réalisé en cours.
3) Dans le cas de variances hétérogènes, il est conseillé d'utiliser comme statistique de test :

$$t' = \frac{\overline{x_1} - \overline{x_2}}{E} \quad \text{avec} \quad E^2 = \frac{s_{1c}^2}{n_1} + \frac{s_{2c}^2}{n_2}$$

en prenant, comme nombre de degrés de liberté, l'entier le plus proche de la valeur :

$$dl' = \frac{A}{B} \quad \text{avec} \quad A = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \ , \ B = \frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}$$

Calculer la statistique t' et conclure.
Réponses : 1) On obtient $t_{obs} = 1.87$, *non significatif au seuil de 5%.*
3) On obtient $t'_{obs} = 2.064$, *et* $dl' = 78.58$. *On garde donc 79 ddl, et la valeur observée est alors significative d'une différence entre les deux groupes.*

Exercice 9
1) Pour $ddl_1 = 2, ddl_2 = 4$, la densité f de la loi de Fisher-Snedecor est donnée, pour $x \geq 0$ par :

$$f(x) = \frac{8}{(2+x)^3}$$

Construire point par point la courbe de la fonction f.

2) Pour $ddl_1 = 4, ddl_2 = 4$, la densité g de la loi de Fisher-Snedecor est donnée pour $x \geq 0$ par :

$$g(x) = \frac{6x}{(1+x)^4}$$

Construire point par point la courbe de la fonction g.

Exercice 10

Ref. M. El Ali et al., Analyse des stratégies de coping des joueuses et joueurs de tennis face à une blessure sportive majeure, Annales Médico Psychologiques (2006), doi : 10.1016/j.amp.2005.06.013

Dans la publication citée *supra*, les auteurs rendent compte d'une étude qui répond principalement à un double objectif : tout d'abord, élucider et décrire le choix des stratégies de *coping* (ou stratégies d'ajustement au stress) d'un échantillon de joueuses et joueurs de tennis confrontés à une blessure sportive ; ensuite, analyser de manière qualitative, les différences possibles de sexe expliquant le lien entre stratégies de *coping* et blessure. Douze joueurs de tennis (dont six femmes), âgés de 15 à 62 ans, ont pris part à un entretien semi-directif individuel d'une heure environ. Tous les joueurs sélectionnés avaient une blessure majeure au moment de l'entretien, c'est-à-dire une blessure entraînant un arrêt sportif supérieur à 21 jours. La répartition des choix de *coping* pour l'ensemble des joueurs de l'échantillon est présentée dans le tableau ci-dessous.

RSS	RC	DI	FE	DE	AA	PDIST	RA	S	PD	EE	CSH	RP	EC	AT
2	0	0	0	0	0	0	0	0	0	0	0	2	0	1
4	0	0	0	0	2	0	0	0	0	0	0	4	0	1
5	0	0	0	0	3	0	0	0	0	0	0	5	0	2
6	0	0	0	0	3	1	0	0	0	0	1	7	1	3
6	0	0	0	1	3	2	1	0	1	0	1	7	1	3
6	0	0	0	2	4	2	1	0	1	0	2	7	1	3
8	1	0	0	3	4	2	1	0	1	0	2	7	2	4
8	1	1	1	4	5	2	1	0	1	1	3	8	2	4
9	1	1	1	5	6	3	2	2	2	1	3	9	3	4
10	1	2	1	5	7	3	3	2	2	2	4	9	3	5
10	1	2	1	5	8	4	4	3	2	2	4	10	4	7
12	1	3	2	8	9	4	8	4	3	3	5	12	5	9

Signification des variables. RSS : Recherche de soutien social ; RC : Réévaluation cognitive ; DI : Distraction ; FE : Fuite–évitement, DE : Déni ; AA : Auto-accusation ; PDIST : Prise de distance ; RA : Résignation–acceptation ; S : Substitution ; PD : Pensées désirantes ; EE : Expression des émotions ; CSH : Confiance en soi ; RP : Résolution de problème ; AT : Adhésion thérapeutique.

N.B. Dans ce tableau, les données sont présentées par valeurs croissantes de chaque variable, et ne sont pas appariées par sujet observé.

1) Les auteurs indiquent notamment :

> Une étape préalable à la réalisation d'analyse inférentielle a consisté en la vérification de la multinormalité de nos données. Cette étape est fondamentale dans la mesure où l'existence de données anormales peut biaiser les estimations des termes d'erreurs, les tests de significativité des liens entre les dimensions

> de *coping* et le sexe. (...) Pour confirmer cela, nous avons effectué un test Z de Kolmogorov-Smirnov (K-S) (corrigé par le coefficient de Lilliefors), ainsi que par celui de Shapiro-Wilks'W. Les résultats du test de Kolmogorov-Smirnov, effectués pour chacune des dimensions de l'étude, soutiennent que la distribution des résultats est normale, hormis pour la variable "substitution", qui ne semble pas respecter la courbe de Gauss.

a) Pour la variable RSS, on obtient pour la statistique de Kolmogorov-Smirnov la valeur $D_{obs} = 0.159$. L'hypothèse de normalité de la distribution parente peut-elle être acceptée au seuil de 5% ?

b) Mener une étude analogue (calcul de la statistique de Kolmogorov-Smirnov, réalisation du test de Lilliefors) pour la variable RC. Les paramètres de cette variable sont donnés par : $\overline{x} = 0.5$, $s_c = 0.52$.

c) Pour les autres variables, Statistica fournit les résultats suivants :

	Tests de Normalité (Feuille de données1 dans Coping.stw)					
Variable	N	D max	K-S p	Lillief. p	W	p
DI	12	0,3447	p < ,10	p < ,01	0,7484	0,0026
FE	12	0,3542	p < ,10	p < ,01	0,7321	0,0017
DE	12	0,1821	p > .20	p > .20	0,8840	0,0986
AA	12	0,1593	p > .20	p > .20	0,9683	0,8924
PDIST	12	0,1897	p > .20	p > .20	0,9000	0,1585
RA	12	0,2924	p > .20	p < ,01	0,7520	0,0028
S	12	0,4040	p < ,05	p < ,01	0,6892	0,0007
PD	12	0,2000	p > .20	p < ,20	0,8770	0,0803
EE	12	0,3447	p < ,10	p < ,01	0,7484	0,0026
CSH	12	0,1511	p > .20	p > .20	0,9217	0,3005
RP	12	0,2131	p > .20	p < ,15	0,9690	0,8996
EC	12	0,1941	p > .20	p > .20	0,9155	0,2510
AT	12	0,2215	p > .20	p < ,15	0,9079	0,2006

Quelles sont les variables pour lesquelles l'hypothèse de normalité peut être retenue ? Sur quels résultats statistiques se base le commentaire fait par les auteurs ?

2) Afin d'évaluer l'impact du sexe sur les scores moyens obtenus à chaque dimension de coping, les auteurs ont effectué des tests *t* sur séries indépendantes.
Plus loin, les auteurs indiquent également :

> Dans une deuxième étape, des corrélations partielles bivariées, en contrôlant l'effet du sexe, ont été conduites, afin d'avoir une estimation du degré de liaison (ou force de la relation) entre chaque dimension de coping entre elles.

a) Quelle transformation peut-on faire subir aux données pour *contrôler l'effet du sexe* ?

b) Au vu des traitements envisagés ci-dessus, l'étude de normalité, telle qu'elle a été conduite dans la question 1), était-elle vraiment utile ? Quelles autres études de normalité, plus pertinentes, pourrait-on proposer ?

Réponses : 1) a) Le test convenable est ici celui de Lilliefors, puisque la moyenne et l'écart type sont estimés à partir de l'échantillon. On lit dans la table pour $N = 12$ et $\alpha = 5\%$:

$D_{crit} = 0.242$. *On a donc :* $D_{obs} < D_{crit}$ *et la normalité de la distribution parente est acceptée.*
1) b) Le calcul conduit ici à : $D_{obs} = 0.3315$. *Comme précédemment,* $D_{crit} = 0.242$. *Pour cette variable, c'est l'hypothèse* H_1 *qui est retenue : la normalité de la distribution parente est rejetée.*
1) c) L'examen des p-values fournies par Statistica pour les tests de Lilliefors et de Shapiro-Wilk conduit à accepter la normalité des variables DE, AA, PDIST, PD, CSH, RP, EC et AT au seuil de 5%. On voit que les auteurs ont fondé leur conclusion sur le test de Kolmogorov-Smirnov et non sur ceux de Lilliefors ou de Shapiro-Wilk.
2) a) Pour "contrôler l'effet du sexe", on peut calculer les moyennes par sexe et considérer comme protocole dérivé, les écarts à la moyenne de chaque groupe défini par le sexe.
2) b) Pour justifier l'utilisation du test de Student, il aurait été préférable d'étudier la normalité des variables dans chacun des groupes définis par les modalités de la variable "Sexe". Pour l'étude des coefficients de corrélation, c'est la normalité de l'ensemble des écarts à la moyenne de chaque groupe qu'il faudrait étudier.

Analyse de la variance à un facteur (ANOVA) : comparaison de k moyennes sur des groupes indépendants

Exercice 11

Dans un établissement scolaire, on a réparti les élèves en trois classes de troisième ; les notes ci-dessous sont celles obtenues par les élèves en mathématiques au Brevet des Collèges. Peut-on dire que ces trois classes sont équivalentes ? Si oui, quelles seraient les caractéristiques de la population résultant de la fusion des trois groupes ?

G1	G2	G3
14	8	7
15	18	8
20	3	11
7	12	11
8	15	20
13	8	14
10	7	13
1	11	13
12	8	10
16	14	12
17	14	12
17	9	13
11	9	12
6	9	14
16	10	8

G1	G2	G3
8	14	13
10	15	12
11	14	8
11	13	8
7	10	11
10	12	15
11	10	8
12	12	14
11	12	16
8	11	13
	10	12
	10	15
	10	
	12	

Vérifier l'exactitude des tableaux ci-dessous et conclure.

	G1	G2	G3	Totaux	
n_j	25	29	27	81	
T_j	282	320	323	925	10563,27
Σx_{ij}^2	3600	3782	4091	11473	
T_j^2/n_j	3180,96	3531,03	3864,04	10576,03	
Inter	12,76				
Total	909,73				

Sources de variations	Sommes des carrés	DDL	Carrés moyens	F
Inter	12,76	2	6,38	0,55
Intra	896,97	78	11,50	
Total	909,73	80		

Réponses : Au seuil de 5%, $F_{crit}(2,78) = 3.1$. La différence entre les groupes n'est donc pas significative. De plus, l'obtention d'un F_{obs} inférieur à 1 semblerait indiquer (sans pour autant le montrer) que les classes n'ont pas été constituées au hasard, mais qu'elles ont, au contraire, été rendues artificiellement homogènes : on a composé les trois classes de façon qu'elles soient de niveau équivalent.

Exercice 12

Lors d'une expérience pédagogique, on s'intéresse à l'effet comparé de deux pédagogies des mathématiques chez deux groupes de 10 sujets :

– pédagogie traditionnelle (p_1)
– pédagogie moderne (p_2)

On note la performance à une épreuve de combinatoire.

p_1 traditionnelle	
s1	5.0
s2	4.0
s3	1.5
s4	6.0
s5	3.0
s6	3.5
s7	3.0
s8	2.5
s9	1.5
s10	2.5

p_2 moderne	
s11	4.0
s12	5.5
s13	4.5
s14	6.5
s15	4.5
s16	5.5
s17	1.0
s18	2.0
s19	4.5
s20	4.5

1) Vérifier que les paramètres des deux échantillons sont donnés par :

	p_1	p_2
Moyenne	3.250	4.250
Ecart-type	1.365	1.553
Variance	1.863	2.413
Ecart-type corrigé	1.439	1.637
Variance corrigée	2.069	2.681

2) Ces données expérimentales permettent-elles d'affirmer que la pédagogie a un effet sur les résultats à l'épreuve de combinatoire ?

a) On décompose les données de la manière suivante :

$$\begin{bmatrix} 5.0 & 4.0 \\ 4.0 & 5.5 \\ 1.5 & 4.5 \\ 6.0 & 6.5 \\ 3.0 & 4.5 \\ 3.5 & 5.5 \\ 3.0 & 1.0 \\ 2.5 & 2.0 \\ 1.5 & 4.5 \\ 2.5 & 4.5 \end{bmatrix} = \begin{bmatrix} 3.75 & 3.75 \\ 3.75 & 3.75 \\ 3.75 & 3.75 \\ 3.75 & 3.75 \\ 3.75 & 3.75 \\ 3.75 & 3.75 \\ 3.75 & 3.75 \\ 3.75 & 3.75 \\ 3.75 & 3.75 \\ 3.75 & 3.75 \end{bmatrix} + \begin{bmatrix} -0.5 & 0.5 \\ -0.5 & 0.5 \\ -0.5 & 0.5 \\ -0.5 & 0.5 \\ -0.5 & 0.5 \\ -0.5 & 0.5 \\ -0.5 & 0.5 \\ -0.5 & 0.5 \\ -0.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix} + \begin{bmatrix} 1.75 & -0.25 \\ 0.75 & 1.25 \\ -1.75 & 0.25 \\ 2.75 & 2.25 \\ -0.25 & 0.25 \\ 0.25 & 1.25 \\ -0.25 & -3.25 \\ -0.75 & -2.25 \\ -1.75 & 0.25 \\ -0.75 & 0.25 \end{bmatrix}$$

Vérifier le calcul des sommes de carrés associées à cette décomposition :

$SC_{inter} = 20 \times 0.5^2 = 5$

$SC_{intra} = 1.75^2 + \ldots + 0.25^2 = 42.75$

Dresser le tableau d'analyse de variance et l'utiliser pour comparer les moyennes des deux groupes.

b) Comparer les résultats avec ceux obtenus à l'aide de la statistique T.
Réponses :
Le tableau d'analyse de variance est :

Sources de variation	*Sommes des carrés*	*DDL*	*Carrés moyens*	*F*
Inter	*5,0*	*1*	*5,0*	*2,11*
Intra	*42,75*	*18*	*2,375*	
Total	*47,75*	*19*		

Au seuil de 5%, $F_{crit}(1, 18) = 4.41$*. Hypothèse* H_1 *rejetée.*
Comparaison possible avec les résultats fournis par le T de Student : $t^2_{obs} = (-1.45)^2 = 2.10$*, c'est-à-dire la valeur de* F*.*

Enoncé 13 *Données Bransfor*
On reprend une expérience de Bransford et al. (1972), dans laquelle on demande à des sujets d'écouter le texte suivant :

> "Si les ballons éclatent, le son ne portera pas puisque tout sera bien trop loin du bon étage. Une fenêtre fermée empêchera également le son de porter, surtout depuis que les immeubles récents sont correctement isolés. Comme l'essentiel de l'opération dépend d'une arrivée correcte d'électricité, un fil cassé causerait bien des problèmes. Evidemment, le type peut hurler. Mais la voix humaine n'est pas assez puissante pour porter bien loin. Un problème supplémentaire serait qu'une corde casse sur l'instrument. Alors il serait impossible d'accompagner le message. C'est clair que la meilleure situation impliquerait la plus petite distance. Alors, il y aurait bien moins de problèmes potentiels. Avec un contact en face à face, un bien petit nombre de choses pourrait gêner."

Le but visé par Bransford *et al.* est de montrer l'importance du contexte dans la compréhension et la mémorisation d'un texte. Pour ce faire, ils utilisent quatre groupes expérimentaux :

1. Un groupe "sans contexte" entend simplement le texte.
2. Le groupe "avec contexte avant" regarde une figure suggérant un contexte approprié pendant qu'il entend le texte.
3. Le groupe "avec contexte après" entend le texte puis regarde la figure précédente.
4. Le groupe "avec contexte partiel" regarde une figure suggérant un contexte inapproprié pendant qu'il entend le texte.

A proprement parler cette étude comprend un groupe expérimental (le groupe 2 : contexte pendant) et trois groupes contrôles (les groupes 1, 3 et 4). Les groupes contrôles doivent permettre d'éliminer des explications concurrentes (en particulier, effet facilitateur sur la mémoire de l'imagerie, de l'aspect concret du matériel, etc.). L'expérimentateur s'attend, donc, à observer une performance pour le groupe 2 supérieure aux trois autres groupes.
Il choisit de mesurer le comportement des sujets par deux Variables Dépendantes : une note de compréhension donnée par les sujets (de 0 à 7, avec 0 indiquant l'incompréhension totale), et le nombre d'idées correctement rappelées (Bransford découpe le texte en 14 idées, essayez de les retrouver !). Quoique cette dernière Variable Dépendante soulève de

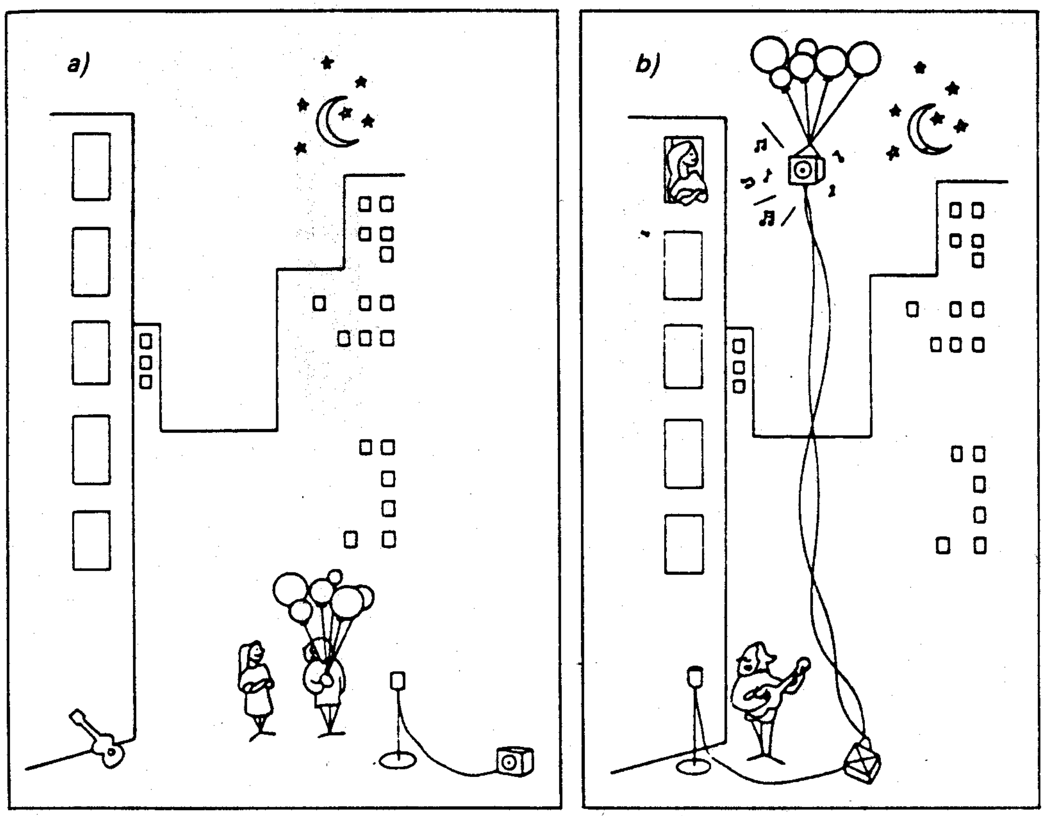


FIG. 1 – Contexte inapproprié (a) et approprié (b) pour l'expérience de Bransford

délicats problèmes de codage (e.g., à partir de quel moment une idée est présente...), elle reflète clairement l'intérêt des auteurs de cette expérimentation.
Dans cette expérience, on utilise vingt sujets répartis en quatre groupes. Les résultats, pour la Variable Dépendante "nombre d'idées rappelées" (maximum 14) se trouvent ci-dessous (mais avant, faites ce que doit faire un bon expérimentateur : prenez une feuille et détaillez les cinq premières étapes du test statistique avant de partir à la pêche aux résultats) :

Résultats de l'expérience				
	G.1	G.2	G.3	G.4
	3	5	2	5
	3	9	4	4
	2	8	5	3
	4	4	4	5
	3	9	1	4
$T_{.j}$	15	35	16	21
n_j	5	5	5	5
$\frac{T_{.j}}{n_j}$	3	7	3.2	4.2
Σx_{ij}^2	47	267	62	91

Justifiez les calculs et le tableau d'ANOVA suivants :
Table d'ANOVA :

Source	ddl	SC	CM	F_{cal}	$Pr(F > F_{cal})$
$\mathcal{A}$	3	50.95	16.98	7.22 **	.00288
$\mathcal{S}(\mathcal{A})$	16	37.60	2.35		
Total	19	88.55			

Si on utilise la procédure des valeurs critiques :
** $F_{critique} = 5.29$, au seuil $\alpha = .01$; $F_{cal} > F_{critique}$. On rejette H_0.
Les cinq étapes du test sont évidemment :

1. *Formulation des hypothèses statistiques H_0 et H_1. Ici :*
   H_0 : dans les 4 conditions, les moyennes dans la population parente sont égales
   H_1 : les 4 moyennes ne sont pas toutes égales.
2. *Choix du test : ici, une analyse de variance à un facteur. Statistique : F.*
3. *Distribution de la statistique de test : ici, le F de Fisher Snedecor avec $ddl_1 = 3$ (nombre de groupes – 1) et $ddl_2 = 16$ (nombre d'observations – nombre de groupes).*
4. *Seuil de signification choisi : ici, $\alpha = 1\%$.*
5. *Règle de décision : détermination des zones d'acceptation et de rejet de H_0. Ici, :*
   – *Si $F_{cal} \leq 5.29$, on accepte H_0 (égalité des moyennes)*
   – *Si $F_{cal} > 5.29$, on refuse H_0 et on accepte H_1.*

L'étude pourrait être poursuivie à l'aide de la méthode des contrastes orthogonaux (que nous ne détaillerons pas).
La première étape consiste opposer le groupe 2 aux trois autres groupes en testant l'hypothèse nulle : $3\mu_2 = \mu_1 + \mu_3 + \mu_4$. On calcule : $L_1 = 3\overline{x_2} - \overline{x_1} - \overline{x_3} - \overline{x_4} = 10.6$; $\sum a_j^2 = 3^2 + (-1)^2 + (-1)^2 + (-1)^2 = 12$; $SC_{contraste1} = \dfrac{nL^2}{\sum a_j^2} = 46.81$
Dans la formule précédente, n est le nombre d'observations par groupe. Ici, $n = 5$. Le F de Fisher associé à ce contraste est obtenu en divisant $SC_{contraste1}$ par le carré moyen résiduel 2.35 ; il vaut 19.92. Les degrés de liberté sont 1 et 16. Le résultat est donc significatif d'un comportement du groupe 2 différent de celui des autres groupes.
La méthode peut être poursuivie en opposant le groupe 4 aux groupes 1 et 3 (coefficients appliqués aux quatre moyennes : 1, 0, 1, -2) puis en opposant les groupes 1 et 3 (coefficients appliqués : 1, 0, -1, 0).
Pourquoi s'agit-il de contrastes orthogonaux *?*
Réponse : Les "vecteurs" associés aux coefficients des trois contrastes, à savoir $V_1 = (-1, 3, -1, -1)$, $V_2 = (1, 0, 1, -2)$, $V_3 = (1, 0, -1, 0)$ sont deux à deux orthogonaux (par exemple, $V_1 \cdot V_2 = -1 \times 1 + 3 \times 0 + (-1) \times 1 + (-1) \times (-2) = 0$), ce qui garantit l'indépendance des résultats des trois tests.
Une autre grandeur intéressante est le coefficient (souvent noté η^2) d'estimation de l'intensité de l'effet de la variable indépendante. Dans le cas d'une analyse de variance à un facteur, il est défini par :

$$\eta^2 = \frac{SC_{inter}}{SC_{total}}$$

Il vaut donc ici : $\eta^2 = 0.58 = 58\%$.
Signification : 58% de la variance de la Variable Dépendante est expliquée par la Variable Indépendante (les différentes conditions expérimentales).
η^2 est aussi le carré d'un coefficient de corrélation. η peut en effet être obtenu comme coefficient de la corrélation entre l'ensemble des données observées d'une part, et la série

de données obtenue en remplaçant chaque observation par la moyenne de son groupe d'autre part. Sur notre exemple, soit U la série des données observées et V la série des données du modèle ainsi obtenu.

u_i	3	3	2	4	3	5	9	8	4	9	2	4	5	4	1	5	4	3	5	4
v_i	3	3	3	3	3	7	7	7	7	7	3.2	3.2	3.2	3.2	3.2	4.2	4.2	4.2	4.2	4.2

On obtient : $r(U, V) = 0.7585$ *et* $r^2(U, V) = 0.575$.

Enoncé 14 *Données Loftus*
Elisabeth Loftus (Loftus et Palmer 1974) —– dans une série d'expérimentations sur le thème du témoignage —– désire mettre en évidence l'influence de la tournure d'une question sur la réponse de témoins. Pour ce faire, elle montre à ses sujets, un film décrivant un accident de voiture. Elle pose, ensuite, une série de questions aux sujets. Parmi celles-ci se trouve une des cinq versions d'une question relative à la vitesse des véhicules. Voici ces versions :

1) HIT : About how fast were the cars going when they *hit* each other ? (A environ quelle vitesse allaient les voitures quand elles se sont "percutées").

2) SMASH : About how fast were the cars going when they *smashed* each other ? (To smash : écraser, heurter avec violence).

3) COLLIDE : About how fast were the cars going when they *collided* each other ? (To collide : entrer en collision, s'emboutir).

4) BUMP : About how fast were the cars going when they *bumped* each other ? (To bump : cogner, frapper).

5) CONTACT : About how fast were the cars going when they *contacted* each other ? (To contact : entrer en contact).

Les sujets répondaient en indiquant une vitesse exprimée en miles (nous sommes aux U.S.A). Voici les résultats obtenus (lors d'une réplique de l'expérience) :

HIT	SMASH	COLLIDE	BUMP	CONTACT
22	38	43	47	27
29	40	39	29	24
33	50	32	58	46
50	45	44	34	37
19	48	29	36	31
37	56	44	43	37
33	52	45	25	34
43	47	33	58	18
40	39	48	24	28
34	40	37	31	26

Après avoir identifié les variables dépendante(s) et indépendante(s), vous tirerez les conclusions de cette expérimentation.
Pour vous aider voici quelques statistiques pour chaque groupe :

	$T_{.j}$	$T_{.j}/n_j$	$T^2_{.j}/n_j$	$\Sigma_j x^2_{ij}$
Gr. 1	340	34.0	11560	12338
Gr. 2	455	45.5	20702.5	21043
Gr. 3	394	39.4	15523.6	15894
Gr. 4	385	38.5	14822.5	16241
Gr. 5	308	30.8	9486.4	10060
Total	1882		72095	75576

La variable dépendante est évidemment la vitesse exprimée en miles. La variable indépendante est le type de verbe utilisé pour poser la question sur la vitesse des voitures.
Manifestement, E. Loftus veut montrer que les “sous-entendus” des verbes sont pris en compte par les sujets dans leur décision sur la vitesse (e.g., les sujets utilisent la signification implicite des verbes comme une source d'information). Le point d'importance dans cette expérience est de remarquer que E. Loftus désire généraliser ses résultats à l'ensemble des verbes signifiant quelque chose comme “entrer en contact”. Quoiqu'elle n'ait pas, à proprement parler, sélectionné ses verbes au hasard, elle les juge représentatifs de l'ensemble des verbes de mouvement. Le problème ici est de décider si le facteur expérimental est fixé ou aléatoire. Si l'on admet que les verbes choisis par Loftus représentent un échantillon représentatif, on décidera que le facteur est aléatoire (cf. La polémique initiée par Clark 1973). Si l'on juge que les modalités sont choisies en fait arbitrairement. on décidera que le facteur est fixé, et les conclusions de l'étude se restreignent aux modalités effectivement présentes dans l'expérimentation. Quelle que soit la décision prise, elle sera criticable.
Ici, le distinguo entre facteur fixé et aléatoire peut paraître sans importance car la décision (rejet ou non de l'hypothèse nulle) sera identique dans les deux cas. *Ce ne sera plus le cas dans des plans d'expérience plus complexes.* En fait, l'essentiel de l'argument de Clark (1973) est de montrer qu'une partie des recherches utilisant du matériel linguistique aboutit à des conclusions SCIENTIFIQUES erronées du fait de la confusion entre facteurs fixes et aléatoires (cf. aussi les réponses de Wike et Church 1976). Clark défend l'idée qu'une partie des conclusions de la psychologie du langage est invalide pour avoir cru que des facteurs aléatoires étaient fixes. A cette attaque répond Chastaing (1986) qui démontre méthodologiquement qu'une autre partie de la psychologie du langage est invalide d'avoir cru que des facteurs fixes étaient aléatoires !
Dans le cas présent, le choix entre les deux modèles n'a pas d'influence sur les résultats de l'analyse statistique : on aboutit à des conclusions statistiques identiques (mais pas à des interprétations psychologiques identiques !). L'analyse de variance permet de conclure en tout cas à un effet sur la vitesse estimée, du type de verbe utilisé pour poser la question.
On obtient le tableau d'analyse de variance suivant :

Source	ddl	SC	CM	F_{cal}	$Pr(F > F_{cal})$
Expérimentale	4	1256.52	314.13	4.06 **	.0069
Erreur	45	3481.00	77.36		
Total	49	4737.52			

Ainsi, le type de verbe employé pour interroger les sujets sur la vitesse des véhicules, influence l'estimation qu'ils donnent ($F_{cal}(4, 45) = 4.06,\ p < .05$). On remarque la vitesse élevée induite par *to smash*.

Poursuivons cet exemple en essayant d'apprécier les différences entre ces différents verbes les uns par rapport aux autres) en calculant les t de Student correspondant aux comparaisons par paires.
Comme les groupes sont équilibrés, on a, pour toutes les comparaisons :
$E^2 = CM_{\text{intra}} \left(\frac{1}{10} + \frac{1}{10}\right) = 77.36 \times 0.2 = 15.472$, d'où $E = 3.933$.
Pour chacune des comparaisons, on obtient :

- Hit v/s Smash : $t = \frac{34 - 45.5}{3.933} = -2.92$
- Hit v/s Collide : $t = -1.37$
- Hit v/s Bump : $t = -1.144$
- Hit v/s Contact : $t = 0.81$
- Smash v/s Collide : $t = 1.55$
- Smash v/s Bump : $t = 1.78$
- Smash v/s Contact : $t = 3.74$
- Collide v/s Bump : $t = 0.23$
- Collide v/s Contact : $t = 2.18$
- Bump v/s Contact : $t = 1.95$.

Pour le test LSD de Fisher, la statistique de test suit une loi de Student à 45 ddl. Les valeurs critiques aux seuils de 5% et 1% sont respectivement 2.01 et 2.69 (test bilatéral). Les comparaisons Hit v/s Smash et Smash v/s Contact sont significatives au seuil de 1%, tandis que Collide v/s Contact est significative au seuil de 5%.

Pour le test de Bonferroni-Dunn, le nombre de comparaisons est égal à 10. Pour un seuil global $\alpha_{FW} = 5\%$, il faut prendre un seuil $\alpha_{PC} = 0.5\%$ par comparaison, et on a donc $t_{crit} = 2.95$. La seule comparaison significative est alors Smash v/s Contact.

Pour le test HSD de Tukey au seuil de 5% bilatéral, la valeur critique de la loi des étendues studentisées est $Q_{crit} = 4.018$ lorsque le nombre de groupes est 5 et le nombre de ddl est 45. On a alors $\frac{Q_{crit}}{\sqrt{2}} = 2.84$ et ce sont les paires Hit v/s Smash et Smash v/s Contact qui conduisent à des différences significatives.

N.B. La valeur critique du t de Student à 0.5% et celle de la loi des étendues studentisées ont été obtenues à partir des tables en ligne disponibles à l'adresse : http ://geai.univ-brest.fr/˜carpenti/statistiques/table1.php.

Analyse de la variance à plusieurs facteurs

Enoncé 15 *Dossier "Geometrie"*
Dans une tâche de dénomination de figures géométriques, l'auteur étudie l'évolution du temps de réaction verbale en fonction de la discriminalité des figures.
Dans un premier temps, on présente aux sujets une série de figures. Pour la moitié d'entre eux, la série est constituée de 2 figures, pour l'autre moitié, de 4 figures. Dans chacun des cas, la série est constituée soit de figures facilement discriminables (triangle, carré,...) soit de figures plus complexes (octogone, décagone...).
Dans un deuxième temps, on demande à chaque sujet de nommer une figure tirée au hasard dans la série précédente et on mesure le temps de réaction verbale du sujet.
48 sujets répartis en 4 groupes de 12 ont participé à l'expérience.

Les moyennes des temps de réaction mesurés en millisecondes observés sur chacun des quatre groupes sont indiqués dans le tableau suivant :

	Discriminalité	
Incertitude	Forte	Faible
2 figures	460	510
4 figures	559	864

1) Définir la variable dépendante et les variables indépendantes prises en compte. Quel est le plan d'expérience utilisé ?
2) Au vu du tableau précédent, indiquer s'il semble y avoir une interaction entre les deux facteurs étudiés. Construire un graphe d'interaction. Commenter ce graphe en rédigeant une phrase exprimant comment se traduit l'effet d'interaction.
3) Le tableau d'analyse de variance se présente ainsi :

Sources de var.	ddl	SC	CM	F
Discriminalité	1	3858.3	3858.3	45.06
Incertitude	1	6238.3	6238.3	72,85
Interaction	1	1885.4	1885.4	22,02
Résidu	44	3767.6	85.6	
Total	47	15666.7		

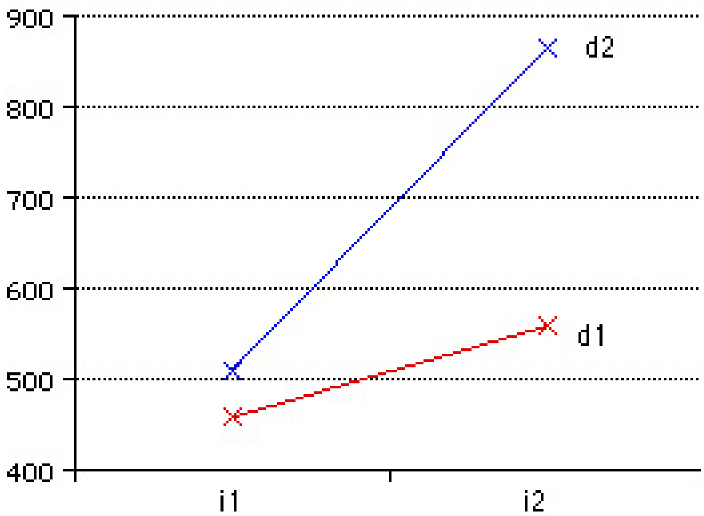
Préciser comment ont été obtenues :

- la valeur 85.6 dans la ligne "résidu" ;
- la valeur 45.06 dans la ligne "discriminalité".

Utiliser la table de la loi de Fisher-Snedecor pour indiquer si les effets principaux et l'effet d'interaction sont significatifs au seuil de 1%.

Réponses : 1) Le plan utilisé est ici $S_{12} < I_2 * D_2 >$.

2) Le temps de réaction augmente lorsque la discriminalité est plus faible. Mais cet effet est d'autant plus important que l'incertitude est élevé.



3) $85.6 = \frac{3767.6}{44}$; $45.06 = \frac{3858.3}{85.6}$. *Au seuil de 1%,* $F_{crit}(1, 44) = 7.2$. *Les effets principaux et l'effet d'interaction sont donc significatifs.*

Enoncé 16

Ref. Adarves-Yorna et al., Crazy innovation or crazy irrelevance ? The contribution of *group norms and social identity to creative behavior, Journal of Experimental Social Psychology, Volume 43, pp 410-416, 2007*

Des chercheurs se sont intéressés aux contributions des normes de groupe et de l'identité sociale sur la créativité. Dans l'expérience qu'ils ont mis en œuvre, quatre conditions expérimentales ont été créées :
– d'une part en manipulant l'identité sociale (saillance de l'identité personnelle v/s saillance de l'identité sociale) ;
– d'autre part en manipulant la norme au sein du groupe : être créatif *avec des mots* ("pour l'espèce humaine il serait bien difficile de communiquer sans les mots") v/s être créatif *avec des images* ("un bon dessin vaut mieux qu'un long discours").

Les participants devaient produire un dépliant sur un sujet donné et le comportement créatif a été mesuré par la proportion d'espace dédié au texte par rapport à l'ensemble de l'espace occupé par les mots et les images dans le dépliant produit.

Dans chaque condition, 8 groupes de sujets ont été observés et les données recueillies ont été agrégées au niveau des groupes. Ces données ont ensuite été traitées par une analyse de variance selon un plan factoriel 2×2 entre les sujets.

Les moyennes observées dans les quatre conditions sont les suivantes :

Saillance	Norme de groupe	Moyenne	Effectif
Personnelle	Mots	25.10	8
Personnelle	Images	48.64	8
Sociale	Mots	54.26	8
Sociale	Images	40.00	8

Le tableau d'analyse de variance produit par Statistica est le suivant :

	SC	Degr. de Liberté	MC	F
ord. origine	56 447.16	1	56 447.16	114.4287
Saillance	842.24	1	842.24	1.7074
Norme de groupe	172.38	1	172.38	0.3494
Saillance*Norme de groupe	2 857.49	1	2 857.49	5.7927
Erreur	13 812.27	28	493.30	

1) Interpréter les résultats fournis par le logiciel.

2) Faire un diagramme d'interaction entre les deux variables indépendantes manipulées par les expérimentateurs et justifier l'affirmation suivante :

> Quand l'identité sociale était saillante, les créations étaient plus en accord avec la norme du groupe. Dans la condition "saillance de l'identité personnelle", les créations étaient inconsistantes avec la norme du groupe.

Enoncé 17

Des chercheurs se sont intéressés à l'effet de l'apprentissage de la musique sur les capacités visuo-spatiales des sujets[1].

[1] Effect of musical expertise on visuospatial abilities : Evidence from reaction times and mental imagery, Renaud Brochard, André Dufour and Olivier Després, Brain and Cognition – Volume 54 (2004) pp. 103-109

Dans l'une des expériences réalisées, les chercheurs utilisent 24 sujets (12 musiciens et 12 non-musiciens). Dans une première condition, ils mesurent le temps de réaction des sujets soumis à un stimulus simple : un petit disque lumineux est affiché pendant 70 ms et les sujets doivent appuyer le plus rapidement possible sur un bouton lorsqu'ils aperçoivent le disque. Dans une deuxième condition, les sujets doivent effectuer un choix : le cercle lumineux est soit vert, soit rouge et les sujets doivent appuyer sur la flèche gauche ou la flèche droite d'un clavier selon la couleur du stimulus.
Les données (temps de réaction moyen sur 80 essais dans chaque condition) observées lors d'une reprise de cette expérience sont rassemblées dans le tableau 1.

Sujets	Expertise	Simple	Choix
s1	Mus.	160	303
s2	Mus.	208	230
s3	Mus.	181	272
s4	Mus.	133	236
s5	Mus.	190	283
s6	Mus.	215	282
s7	Mus.	183	261
s8	Mus.	205	291
s9	Mus.	126	322
s10	Mus.	190	268
s11	Mus.	157	284
s12	Mus.	164	233

Sujets	Expertise	Simple	Choix
s13	Non-Mus.	158	323
s14	Non-Mus.	202	345
s15	Non-Mus.	181	341
s16	Non-Mus.	164	340
s17	Non-Mus.	161	244
s18	Non-Mus.	204	333
s19	Non-Mus.	177	333
s20	Non-Mus.	194	287
s21	Non-Mus.	195	324
s22	Non-Mus.	201	311
s23	Non-Mus.	184	304
s24	Non-Mus.	199	354

TAB. 1 – Temps de réaction - Expérience 2

1) Le plan de cette expérience peut être écrit sous la forme $\mathcal{S}_{12} < \mathcal{X}_2 > *\mathcal{C}_2$. Définir les facteurs $\mathcal{S}$, $\mathcal{X}$ et $\mathcal{C}$. Précisez quels sont les niveaux de chacun de ces facteurs. Justifier l'écriture du plan d'expérience.

2) Les données ci-dessus sont traitées par une analyse de variance. Les sommes de carrés relatives aux différentes sources de variation sont indiquées ci-dessous :

Source	Somme des carrés
Expertise musicale	9690.1
Condition expérimentale	160083.0
Interaction Expertise × Condition	4524.1
Sujet(Expertise)	15889.9
Condition × Sujet(Expertise)	15273.9

a) Dresser le tableau d'analyse de variance correspondant.
N.B. L'ordre dans lequel sont indiquées les sources de variation dans le tableau précédent ne correspond pas nécessairement à l'ordre d'apparition dans le tableau d'ANOVA.

b) En utilisant un seuil de 5%, étudier quelles sont les sources de variation dont l'effet est significatif.

Indication de réponse :
Le tableau d'analyse de variance se présente ainsi :

Source	*ddl*	*SC*	*CM*	F_{cal}	*Sig*
X_2	*1*	*9690.1*	*9690.1*	*13.41*	**
$S(X_2)$	*22*	*15889.9*	*722.3*		
C_2	*1*	*160083*	*160083*	*230.6*	**
Interaction X_2C_2	*1*	*4524.1*	*4524.1*	*6.52*	*
Résidu $C_2S_{12}(X_2)$	*22*	*15273.9*	*694.3*		
Total	*47*	*205461*			

Enoncé 18

Ref. Deaux, K. et al., Becoming American : Stereotype Threat Effects in Afro-Caribbean Immigrant Groups, Social Psychology Quaterly, Vol. 70, No 4, 2007, pp. 384-404

Dans une étude publiée en 2007, des chercheurs se sont intéressés aux effets de la menace du stéréotype[2] chez les immigrants originaires des Caraïbes (West Indians) et ont comparé de ce point de vue les immigrants de première génération (nés aux Caraïbes) et ceux de deuxième génération (nés aux USA de parents nés aux Caraïbes).

Dans une partie de l'étude, les chercheurs ont interrogé un échantillon de 270 étudiants originaires des Caraïbes inscrits dans un College de la City University de New York. Cet échantillon comportait 145 sujets immigrés de première génération et 125 sujets immigrés de seconde génération.

Le questionnaire rempli par les sujets comportait deux échelles visant à comparer les stéréotypes culturels perçus relatifs aux Afro-Américains et aux West Indians. Pour chacune des deux cibles et pour 12 adjectifs, les sujets devaient évaluer sur une échelle en 6 points la façon dont le public en général estimait que l'adjectif s'appliquait au groupe cible. On obtenait ainsi pour chacun des deux groupes cibles une évaluation du stéréotype perçu (ou métastéréotype) sur une échelle allant de 12 à 72.

Les statistiques descriptives concernant les résultats observés pour les deux cibles ("AA" : Afro-Américains, "WI" : West Indians), dans chacun des groupes, sont les suivantes :

Statistiques Descriptives par Groupes N=270(pas de VM dans les vars dépendantes)						
Génération	AA Moyennes	AA N	AA Ec-Types	WI Moyennes	WI N	WI Ec-Types
1	34.33	145	10.36	47.53	145	9.91
2	34.88	125	9.00	41.85	125	11.17
TsGrpes	34.59	270	9.74	44.90	270	10.87

La comparaison des deux groupes du point de vue des évaluations du stéréotype perçu pour chacune des deux cibles fournit le résultat suivant :

Tests t ; Classmt : Génération Groupe 1 : 1 Groupe 2 : 2					
	Moyenne 1	Moyenne 2	Valeur t	dl	p
AA	34.33	34.88	-0.4611	268	0.6450
WI	47.53	41.85	4.4294	268	0.000014

[2]La menace du stéréotype se traduit par le comportement d'une personne qui se conforme inconsciemment à un stéréotype négatif portant sur elle en effectuant une tâche.

Enfin, on réalise une analyse de variance avec "Génération" comme facteur d'emboîtement des sujets et "Cible" comme facteur croisé avec les sujets. On obtient le tableau d'ANOVA suivant :

	SC	Degr. de Liberté	MC	F
Génération	884.7	1	884.7	5.827
Sujets	40691.0	268	151.8	
Cible	13652.4	1	13652.4	253.69
Cible*Gén.	1303.6	1	1303.6	24.22
Résidu	14422.5	268	53.8	

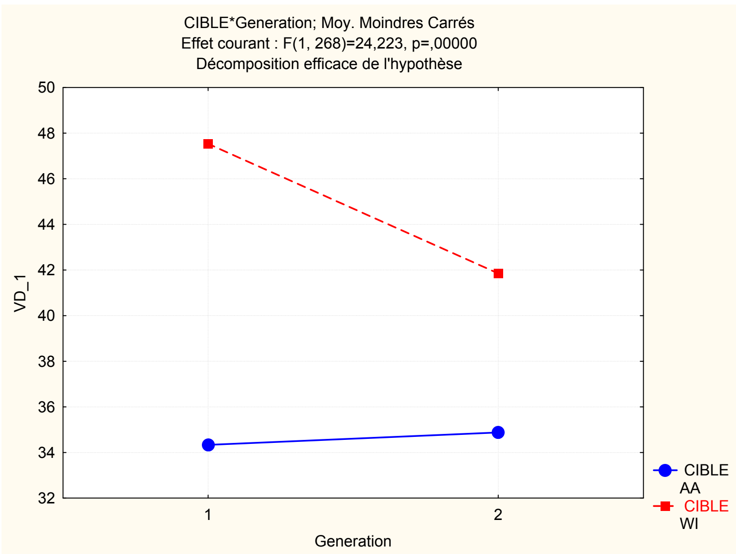
Utiliser ces différents tableaux pour répondre aux questions suivantes.

a) Les étudiants de première et de seconde génération diffèrent-ils dans leurs métastéréotypes concernant les Afro-Américains ? dans leurs métastéréotypes concernant les West Indians ?

b) Interpréter le tableau d'analyse de variance ci-dessus.

c) Réaliser un graphique illustrant l'interaction entre "Génération" et "Cible".

Indications de réponses :
a) On utilise ici le tableau des tests t. Il ne semble pas y avoir de différence significative selon la génération pour le stéréotype perçu relatif aux Afro-Américains (t(268)=-0.46, p=0.64). En revanche, il semble que le stéréotype perçu concernant les West Indians soit significativement différent selon la génération ($t(268) = 4.43, p < 0.001$).
b) Pour le facteur "Génération", on obtient $F_{obs} = 5.83$*, valeur supérieure à la valeur critique au seuil de 5 %* ($F_{crit} = 3,84$)*. L'effet du facteur "Génération" est donc significatif. De même, le facteur "Cible"* ($F(1, 268) = 253.69$) *et l'interaction* ($F(1, 268) = 24.22$) *sont significatifs aux seuils traditionnels.*



c)

Enoncé 19

On vise à l'aide d'un questionnaire à identifier la façon dont les couples homoparentaux sont perçus d'un point de vue évaluatif et normatif à l'aide d'échelles d'accord de format Likert. Deux variables indépendantes sont prises en compte : le sexe du répondant, homme ou femme, et le sexe du couple homoparental, couple d'hommes ou couple de femmes. Les deux questions ci-dessous concernant les couples de femmes puis les couples d'hommes sont posées dans le même questionnaire, ce qui signifie que chaque répondant donne son accord pour les deux modalités de la variable indépendante "sexe du couple homoparental", nommée "Cible" ou "Sexe-couple" dans les tableaux de traitements statistiques.

Question 1.
Voici plusieurs propositions concernant les couples homoparentaux féminins.
Pour chacune, cochez la case qui correspond le mieux à votre opinion, en sachant que 1 signifie que vous êtes tout à fait d'accord avec la proposition et que 5 signifie que vous n'êtes pas du tout d'accord avec la proposition.
Le fait qu'un couple de femmes élève au quotidien un ou plusieurs enfants vous paraît :

	1	*2*	*3*	*4*	*5*
Acceptable					
Concevable					
Légitime					
Normal					
Légal					

Question 2.
Voici plusieurs propositions concernant les couples homoparentaux masculins.
Pour chacune, cochez la case qui correspond le mieux à votre opinion, en sachant que 1 signifie que vous êtes tout à fait d'accord avec la proposition et que 5 signifie que vous n'êtes pas du tout d'accord avec la proposition.
Le fait qu'un couple d'hommes élève au quotidien un ou plusieurs enfants vous paraît :

	1	*2*	*3*	*4*	*5*
Acceptable					
Concevable					
Légitime					
Normal					
Légal					

Pour les trois adjectifs *Acceptable, Normal, Légal*, les valeurs observées sont les suivantes :

		Qu. 1 - couple de femmes			Qu. 2 - couple d'hommes		
No	Sexe	Acceptable	Normal	Légal	Acceptable	Normal	Légal
1	femme	2	3	3	2	3	3
2	femme	1	4	3	4	4	3
3	femme	1	1	1	1	1	1
4	femme	1	1	1	1	1	1
5	femme	1	1	1	1	1	1
6	femme	2	3	2	3	3	2
7	femme	2	2	2	2	2	2
8	femme	2	2	2	2	2	2
9	femme	1	1	1	2	2	2
10	femme	1	1	1	1	1	1
11	femme	2	4	4	4	5	4
12	femme	1	1	1	3	3	3
13	femme	2	3	1	2	3	1
14	femme	5	1	1	5	1	1
15	femme	3	1	2	3	1	2
16	femme	1	1	1	3	4	1
17	femme	3	4	3	3	4	3
18	femme	3	1	1	4	1	1
19	femme	4	1	1	5	1	1
20	femme	1	3	3	1	3	3
21	homme	3	4	1	5	5	5
22	homme	1	1	2	1	1	2
23	homme	1	5	1	1	4	2
24	homme	2	2	2	2	2	2
25	homme	3	2	3	4	3	5
26	homme	5	5	5	5	5	5
27	homme	5	5	5	5	5	5
28	homme	5	5	5	5	5	5
29	homme	1	1	1	3	3	3
30	homme	1	2	2	1	2	2
31	homme	1	1	1	2	3	2
32	homme	2	4	2	3	4	3
33	homme	5	5	5	5	5	5
34	homme	1	1	1	3	3	3
35	homme	1	3	1	1	2	1
36	homme	2	2	2	5	5	5
37	homme	3	3	3	5	5	5
38	homme	2	2	2	5	5	5
39	homme	1	5	1	5	4	1
40	homme	1	5	1	5	5	1

Pour chacun des trois adjectifs, on traite les données à l'aide d'une ANOVA. Les résultats obtenus sont les suivants :

Adjectif : Acceptable

	SC	Degr. de Liberté	MC	F	p
Sexe	8.45	1	8.45	2.48	0.12
Erreur	129.75	38	3.41		
Cible	18.05	1	18.05	25.26	0.000012
Cible*Sexe	1.80	1	1.8	2,52	0.12
Erreur	27.15	38	0.71		

Adjectif : Normal

	SC	Degr. de Liberté	MC	F	p
Sexe	36.45	1	36.45	11.41	0.0017
Erreur	121.35	38	3.19		
Cible	5.00	1	5.00	8.82	0.0051
Cible*Sexe	0.45	1	0.45	0.79	0.38
Erreur	21.55	38	0.57		

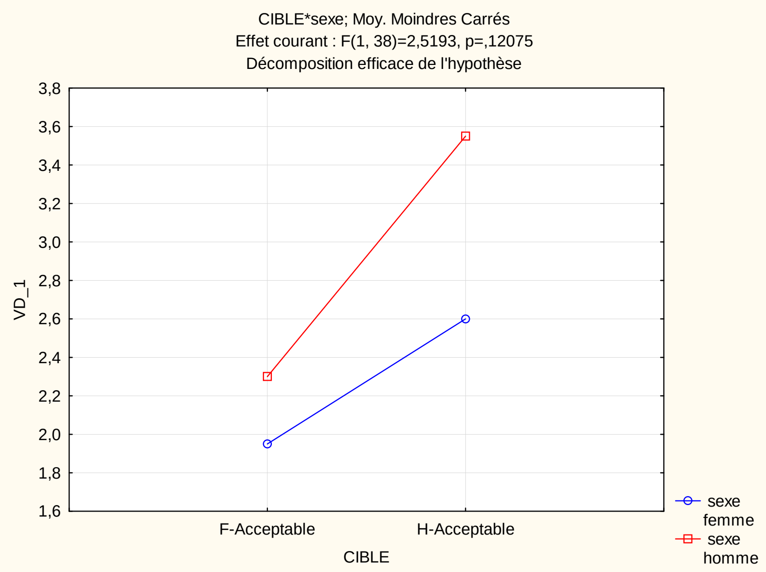
Adjectif : Légal

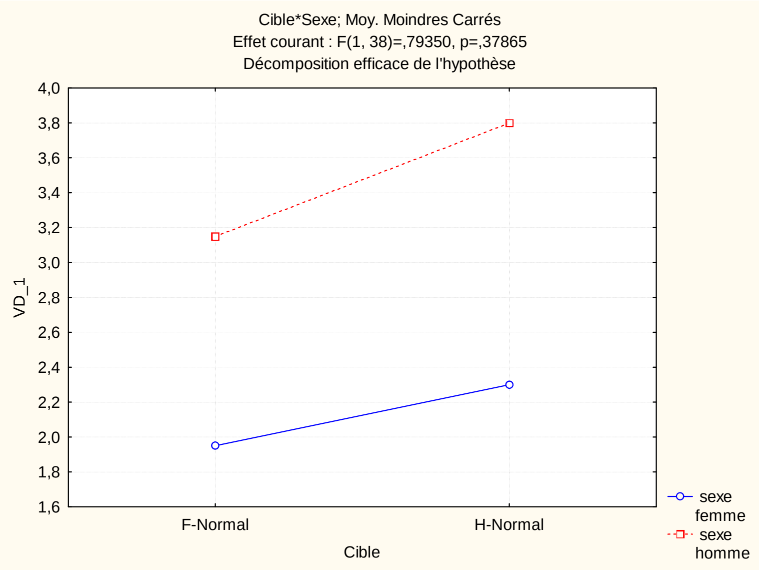
	SC	Degr. de Liberté	MC	F	p
Sexe	20.00	1	20.00	6.75	0.013
Erreur	112.55	38	2.96		
Cible	7.20	1	7.20	15.41	0.00035
Cible*Sexe	4.05	1	4.05	8.67	0.0055
Erreur	17.75	38	0.47		

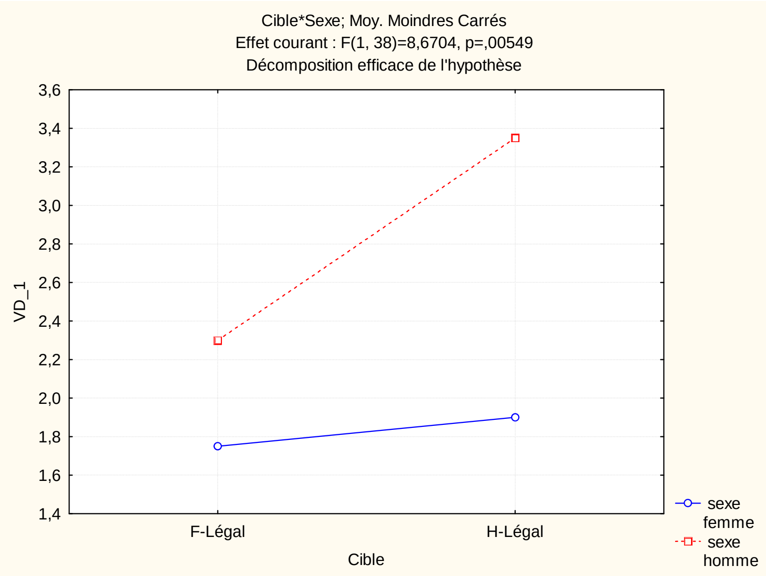
Quels sont les facteurs de variation pris en compte ? Quelles sont les relations d'emboîtement et de croisement entre ces facteurs ?
Interpréter les résultats des trois ANOVA.
Représenter les résultats relatifs à chacun des adjectifs à l'aide d'un graphe d'interaction en utilisant le tableau de moyennes suivant :

sexe	F-Acceptable	F-Normal	F-Légal	H-Acceptable	H-Normal	H-Légal
femme	1.95	1.95	1.75	2.60	2.30	1.90
homme	2.30	3.15	2.30	3.55	3.80	3.35
TsGrpes	2.13	2.55	2.03	3.08	3.05	2.63

*Eléments de réponse : deux facteurs sont ici envisagés : le sexe du sujet interrogé et le sexe du couple homoparental "cible". Les sujets sont emboîtés dans le facteur sexe et croisés avec le facteur cible. Le plan peut s'écrire : Sujet<Sexe>*Cible.*
Lorsque l'on considère un seuil de 5% : pour le premier adjectif, seul la cible a un effet significatif ; pour le deuxième, les deux facteurs sont un effet significatif, mais sans interaction ; pour le troisième, toutes les sources de variation ont un effet significatif.
Les graphes d'interaction sont les suivants :


CIBLE*sexe; Moy. Moindres Carrés
Effet courant : F(1, 38)=2,5193, p=,12075
Décomposition efficace de l'hypothèse
VD_1
3,8
3,6
3,4
3,2
3,0
2,8
2,6
2,4
2,2
2,0
1,8
1,6
F-Acceptable
H-Acceptable
CIBLE
sexe femme
sexe homme


Cible*Sexe; Moy. Moindres Carrés
Effet courant : F(1, 38)=,79350, p=,37865
Décomposition efficace de l'hypothèse
VD_1
4,0
3,8
3,6
3,4
3,2
3,0
2,8
2,6
2,4
2,2
2,0
1,8
1,6
F-Normal
H-Normal
Cible
sexe femme
sexe homme


Cible*Sexe; Moy. Moindres Carrés
Effet courant : F(1, 38)=8,6704, p=,00549
Décomposition efficace de l'hypothèse
VD_1
3,6
3,4
3,2
3,0
2,8
2,6
2,4
2,2
2,0
1,8
1,6
1,4
F-Légal
H-Légal
Cible
sexe femme
sexe homme

Tests non paramétriques sur des groupes indépendants

Enoncé 20

Dans une enquête, on a interrogé 84 hommes et 91 femmes. Les sujets devaient indiquer leur degré d'adhésion à une affirmation, sur une échelle en 5 points. Les résultats sont les suivants :

	Hommes	Femmes
Tout à fait d'accord	10	24
D'accord	15	15
Indifférent	19	21
Opposé	18	17
Tout à fait opposé	22	14

Etudier, à l'aide d'un test de Kolmogorov-Smirnov s'il existe une différence d'opinion entre les hommes et les femmes.
Réponse : Le tableau des fréquences cumulées est donné par :

	Hommes	*Femmes*	*Différence*
Tout à fait d'accord	*0.12*	*0.26*	*0.14*
D'accord	*0.30*	*0.43*	*0.13*
Indifférent	*0.53*	*0.66*	*0.13*
Opposé	*0.74*	*0.85*	*0.10*
Tout à fait opposé	*1.00*	*1.00*	*0.00*

D'où $D = 0.14$. *Pour un test unilatéral, on obtient :* $\chi^2 = 3.673$. *Le niveau de significativité vaut ici* $p = 0.15933$. *On n'a donc pas mis en évidence de différence entre les opinions des deux sexes.*
Pour un test bilatéral au seuil de 5%, *on obtient :* $D_{crit} = 1.36\sqrt{\frac{175}{84\times 91}} = 0.206$ *et la conclusion est analogue.*

Enoncé 21

In a study of correlates of authoritarian personality structure, one hypothesis was that people high in authoritarianism would show a greater tendency to possess stereotypes about members of various national and ethnic groups than would those low in authoritarianism. This hypothesis was tested with a group of 98 randomly selected college women. Each subject was given 20 photographs and asked to "identify" (by matching) as many or as few photographs as they wished. Since, unknown to the subjects, all photographs were of Mexican nationals – either candidates for the Mexican legislature or winners in a Mexican beauty contest and since the matching list of 20 different national and ethnic groups did not include "Mexican", the number of photographs which any subject "identified" constitutes an index of that subject's tendency to stereotype.
Authoritarianism was measured by the F scale of authoritarianism, and the subjects were grouped as "high" and "low" scorers. High scorers were those who scored at or above the median on the F scale ; low scorers were those who scored below the median. The prediction was that these two groups would differ in the number of photographs they identified.

Number of photographs "identified"	Low scorers	High scorers
0-2	11	1
3-5	7	3
6-8	8	6
9-11	3	12
12-14	5	12
15-17	5	14
18-20	5	6

Source : Siegel, S. (1954). Certain determinants and correlates of authoritarianism. Genetic and Psychological Monographs, 49, 187-229.
Comparer les deux groupes à l'aide d'un test unilatéral de Kolmogorov-Smirnov au seuil de 1%.
Indications de réponses : La comparaison des deux fonctions de répartition est donnée dans le tableau suivant :

	0-2	*3-5*	*6-8*	*9-11*	*12-14*	*15-17*	*18-20*
$S_{44}(X)$	*11/44*	*18/44*	*26/44*	*29/44*	*34/44*	*39/44*	*44/44*
$S_{54}(X)$	*1/54*	*4/54*	*10/54*	*22/54*	*34/54*	*48/54*	*54/54*
$S_{44}(X) - S_{54}(X)$	*.232*	*.355*	*.406*	*.252*	*.143*	*.003*	*.0*

D'où $D = 0.406$, $\chi^2 = 4 \times 0.406^2 \frac{44\times54}{44+54} = 15.99$. H_1 *est retenue*

Enoncé 22

Dans une étude sur l'agressivité chez les jeunes enfants, un expérimentateur observe des binômes d'enfants dans une situation de jeu contrôlée. Il ne peut étudier que deux enfants par jour, et se demande si un biais pourrait apparaître du fait des discussions entre les enfants observés et ceux qui ne l'ont pas encore été. Si tel est le cas, la distribution des scores, ordonnée selon la date d'observation, ne devrait pas être aléatoire. Pour répondre à cette question, on convertit les scores en une variable dichotomique, selon leur position par rapport à la médiane des valeurs observées. On réalise ensuite un test des suites sur la suite des valeurs observées pour cette variable.
Les données sont les suivantes :

Sujet	Score	Pos/Med
1	31	+
2	23	−
3	36	+
4	43	+
5	51	+
6	44	+
7	12	−
8	26	+
9	43	+
10	75	+
11	2	−
12	3	−

Sujet	Score	Pos/Med
13	15	−
14	18	−
15	78	+
16	24	−
17	13	−
18	27	+
19	86	+
20	61	+
21	13	−
22	7	−
23	6	−
24	8	−

Vérifier que le nombre de "runs" est ici : $u = 10$, et que H_0 peut être retenue pour un test bilatéral au seuil de 5%.
Indications de réponse : Ici, $n_1 = n_2 = 12$. *Pour un test bilatéral au seuil de 5%, la table du test des suites donne comme valeurs critiques* $u_1 = 7$ *et* $u_2 = 19$. *La règle de décision est donc :* H_0 *est retenue si* $7 < u_{obs} < 19$, *ce qui est le cas ici.*

Enoncé 23

Sandrine Astor, Marie-Aude Depuiset, La famille explique-t-elle la délinquance des jeunes ?, Dossier No 102 publié par la Caisse d'Allocations Familiales en 2008.

1) Dans ce dossier, les auteurs étudient notamment deux échantillons tirés de la population scolaire, le premier en 1999, le second en 2003. Dans les deux échantillons, l'âge des sujets est compris entre 13 et 19 ans. La répartition par âge des deux échantillons est la suivante :

	Scolarisés 1999		Scolarisés 2003	
Age	Effectif	%	Effectif	%
13 ans	278	12.2%	209	12.9%
14 ans	374	16.3%	270	16.7%
15 ans	341	14.9%	308	19.1%
16 ans	354	15.5%	289	17.9%
17 ans	418	18.3%	251	15.6%
18 ans	357	15.6%	184	11.4%
19 ans	165	7.2%	103	6.4%
Ensemble	2288	100%	1614	100%

On cherche à déterminer s'il existe une différence significative entre les deux échantillons du point de vue de la distribution des âges et on réalise un test de Kolmogorov-Smirnov sur ces données.
a) Pourquoi utilise-t-on le test de Kolmogorov-Smirnov plutôt qu'un test de Student ou de Mann-Whitney ?
b) Exécuter le test et conclure au seuil de 5% bilatéral.

Indications de réponses : On obtient ici $D_{max} = 0.077$. *Pour un seuil de 5%, on a* $D_{crit} = 1.36\sqrt{\frac{n_1+n_2}{n_1 n_2}} = 4.42 \times 10^{-2}$. *Comme* $D_{max} > D_{crit}$, *on retient donc l'hypothèse* H_1 *: il semble y avoir une différence significative dans la distribution des âges des deux groupes.*

Enoncé 24

Les chercheurs en psychologie du sport ont utilisé le terme d'*élan psychologique*[3] pour décrire les variations de performance fondées sur des succès ou échecs récents qui modifient les croyances ou la psychologie des athlètes.
Pour étudier la réalité éventuelle de cet effet au niveau des sports d'équipe, un chercheur[4] a relevé les séries de défaites et de victoires des équipes de la National Basketball Association en 1996/97 et 1997/98.
Chaque équipe joue 82 matches par an. Pour l'une d'entre elles, on observe 45 victoires et 37 défaites, échelonnées chronologiquement comme suit :

[3] *psychological momentum* dans le texte original
[4] Réf. Vergin R.C., Winning streaks in sports and the misperception of momentum, Journal of Sport Behavior, Vol. 23 No 2, 2000, pp. 181-197

V V D V V V D D D D D V V D D V V D V V
D D V V D D V V D D D V V D D D V D D
V V V V V D D V V V V D V V V D D D D V V V V
D V V D D V V D D V V V D V V D D D V V

a) Quel test statistique peut-on utiliser pour étudier si la succession des victoires et des défaites est aléatoire ?

b) Mettre en œuvre le test et conclure au seuil de 5%.
Indications de réponse : Il s'agit ici d'étudier si les successions de "D" et de "V" sont aléatoires ou non, ce qui peut être réalisé à l'aide du test de Wald-Wolfowitz. Le nombre de runs observés est ici $u = 35$. *Compte tenu des tailles d'échantillons (*$n_1 = 45$, $n_2 = 37$*), on utilise l'approximation par une loi normale. On obtient :* $\mu = 41.61$, $\sigma^2 = 19.8586$, $Z_{obs} = -1.37$. *Au seuil de 5% bilatéral,* $Z_{crit} = 1.96$. *Comme* $|Z_{obs}| \leq Z_{crit}$*, on conclut sur* H_0. *L'alternance de défaites et de victoires peut être due simplement à l'effet du hasard.*

Enoncé 25
Un chercheur se demande si l'ordre des hommes et des femmes dans une file d'attente à la caisse d'un théâtre est aléatoire ou non. On a relevé ci-dessous le sexe de 50 clients qui se sont présentés successivement à la caisse :
M F M F M M M F F M F M F M F M M M M F M F M F M M F F F M F M F M F
M M F M M F M M M M F M F M M
Répondre au problème posé à l'aide d'un test des suites.
Indications de réponse : On a ici : $n_1 = 30$ *et* $n_2 = 20$. *Le nombre de runs est* $u = 35$. *En utilisant l'approximation par une loi normale, on obtient* $z_{obs} = 2.83$. *On peut donc rejeter l'hypothèse* H_0 *au seuil de 5% bilatéral.*

Enoncé 26
This example is based on a study of gender differences in aggressiveness of four-year-old boys and girls (Siegel, 1956, page 138).
Twelve boys and 12 girls were observed during two 15-minute play sessions ; each child's aggressiveness was scored (in terms of frequency and degree) during those sessions and a combined single aggressiveness index was derived for each child.
Ces données se trouvent dans le fichier Aggressn.sta (fichier exemple fourni avec Statistica).

Sexe	Score	Rang
GARCON	86	20
GARCON	69	18
GARCON	72	19
GARCON	65	16.5
GARCON	113	22
GARCON	65	16.5
GARCON	118	23
GARCON	45	12
GARCON	141	24
GARCON	104	21
GARCON	41	11
GARCON	50	13

Sexe	Score	Rang
FILLE	55	14
FILLE	40	10
FILLE	22	7
FILLE	58	15
FILLE	16	4.5
FILLE	7	1
FILLE	9	2
FILLE	16	4.5
FILLE	26	8
FILLE	36	9
FILLE	20	6
FILLE	15	3

1) Pourquoi ne semble-t-il pas pertinent d'utiliser un test paramétrique pour comparer ces deux groupes ?
2) Réaliser un test de Wald-Wolfowitz sur ces données. Comparer et vérifier les résultats trouvés avec ceux fournis par Statistica :

Test des Suites de Wald-Wolfowitz (Agressn.sta)
Tests significatifs marqués à p <,05000

	Z	niv. p	Z ajusté	niv. p	Nbe de Suites
AGGRESS	-3,75681	0,000172	3,548100	0,000388	4

3) Réaliser un test de Kolmogorov-Smirnov. Comparer avec les résultats fournis par Statistica :

Test de Kolmogorov-Smirnov (Agressn.sta)
Tests significatifs marqués à p <,05000

	Max Nég Diffférnc	Max Pos Diffférnc	niv. p
AGGRESS	0,00	0,833333	p < .001

4) Réaliser enfin un test de Mann-Whitney.

Test U de Mann-Whitney (Agressn.sta)
Tests significatifs marqués à p <,05000

SommeRgs GARCON	SommeRgs FILLE	U	Z	niv. p	Z ajusté	niv. p
216,0000	84,00000	6,00	3,810512	0,000139	3,812170	0,000138

Enoncé 27

Ref. Ember, M., Size of Color Lexicon : Interaction of Cultural and Biological Factors, American Anthropologist, New Series, Vol. 80, No 2, Jun. 1978, pp. 364-367

Le nombre de mots désignant les couleurs de base varie selon les langues et les cultures. Plusieurs explications ont été avancées pour expliquer ce phénomène. Dans l'article cité ci-dessus, l'auteur étudie si la localisation géographique de la communauté considérée (proximité de l'équateur v/s proximité des pôles, mesurée par la latitude, sans tenir compte de l'hémisphère (Nord ou Sud)) peut expliquer le phénomène. Il indique les données suivantes :

Nombre de mots faible (≤ 5)		
Langue	Latitude	N/S
Baganda	1	N
Batak	2	N
Poto	2	N
Arawak	6	N
Tiv	7	N
Kongo	7	S
Somali	8	N
Toda	12	N
Hanunoo	13	N
Wolof	15	N
Ila	16	S
Songhai	17	N
Shona	19	S
/Xam	20	S
Papago	31	N
Pomo	38	N

Nombre de mots élevé (≥ 6)		
Langue	Latitude	N/S
Bari	10	N
Tagalog	10	N
Tamoul	11	N
Haoussa	12	N
Vietnamien	17	N
Espagnol (Mex.)	19	N
Cantonais	24	N
Zuni	35	N
Japonais	35	N
Coréen	35	N
Mandarin	37	N
Américain	42	N
Bulgare	43	N
Hongrois	47	N
Russe	55	N

L'auteur traite ces données à l'aide d'un test unilatéral de Wilcoxon, Mann et Whitney au seuil de 1 %. Retrouver ses résultats.

Indications de réponses : La somme des rangs dans chacun des groupes F (nombre de couleurs faible) et E (nombre de couleurs élevé) est donnée par : $\sum R_F = 185.5$ *et* $\sum R_E = 310.5$. *Les tables nous donnent les valeurs critiques pour le groupe de plus faible effectif, ici le groupe E. Pour un seuil de 1% unilatéral, les valeurs critiques à gauche et à droite sont 181 et 299. Comme* $\sum R_E > 299$, *on conclut sur* H_1 *: il semble exister un effet de la latitude sur la richesse du lexique des couleurs.*

Enoncé 28

On réalise un test de Mann-Whitney sur deux échantillons de tailles respectives $n_1 = 3$ et $n_2 = 4$. Le protocole des rangs observés sur l'ensemble des 7 observations est le suivant :
Groupe 1 : 1, 2, 4
Groupe 2 : 3, 5, 6, 7
1) Calculer W_1 et W_2. Quelle est la somme des rangs ? Comment peut-on la retrouver ?
2) *Passage des statistiques* W_1 *et* W_2 *aux statistiques* U_1 *et* U_2.
Calculer U_1, puis, pour chacun des sujets du groupe 1, compter le nombre de sujets du groupe 2 qui sont classés après lui. Additionner les 3 décomptes obtenus. Que constate-t-on ?
De même, calculer U_2, puis, pour chacun des sujets du groupe 2, compter le nombre de sujets du groupe 1 qui sont classés après lui. Additionner les 4 décomptes obtenus.
3) On veut étudier l'ensemble des protocoles obtenus en affectant au hasard 3 des 7 sujets dans le groupe 1.
a) Il existe 35 protocoles de ce type. Justifier.
b) Pour chacun de ces 35 protocoles, calculer la somme des rangs dans le premier groupe. Représenter la distribution obtenue à l'aide d'un diagramme en bâtons.
c) Quels sont les protocoles pour lesquels on pourrait conclure sur l'hypothèse alternative H_1 au seuil de 5% unilatéral ?
d) Le protocole observé est-il significatif d'une différence entre les deux groupes ?

4) Les échantillons considérés sont évidemment trop petits pour qu'il soit légitime d'utiliser une approximation par une loi normale. Vérifier cependant que les deux formules données dans le cours conduisent à la même valeur de la statistique Z.

Indications de réponses : La somme des rangs est 28. *Elle peut être retrouvée à l'aide de la formule* $1 + 2 + \ldots + N = \dfrac{N(N+1)}{2}$.

D'après la formule donnée en cours : $U_1 = 12 + 6 - 7 = 11$. *Evaluons, pour chaque sujet du groupe 1, le nombre de sujets du groupe 2 classés après lui :*

Rang du sujet	*Nb de suj. du gr. 2*
1	*4*
2	*4*
4	*3*
Total	*11*

Le nombre de protocoles obtenus en affectant au hasard 3 des 7 sujets dans le groupe 1 est le nombre de manières de sélectionner 3 rangs parmi les 7 possibles. Ce nombre est $C_7^3 = 35$. *Ces 35 protocoles et la somme* W_1 *associée à chacun d'eux sont donnés dans le tableau ci-dessous :*

Rangs des sujets							W_1
1	2	3					6
1	2		4				7
1	2			5			8
1	2				6		9
1	2					7	10
1		3	4				8
1		3		5			9
1		3			6		10
1		3				7	11
1			4	5			10
1			4		6		11
1			4			7	12
1				5	6		12
1				5		7	13
1					6	7	14
	2	3	4				9
	2	3		5			10
	2	3			6		11

Rangs des sujets						W_1
2	3				7	12
2		4	5			11
2		4		6		12
2		4			7	13
2			5	6		13
2			5		7	14
2				6	7	15
	3	4	5			12
	3	4		6		13
	3	4			7	14
	3		5	6		14
	3		5		7	15
	3			6	7	16
		4	5	6		15
		4	5		7	16
		4		6	7	17
			5	6	7	18

En considérant ces protocoles comme équiprobables, on voit que les fréquences des modalités de la variable W_1 *sont données par :*

Mod.	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
Freq.	*1/35*	*1/35*	*2/35*	*3/35*	*4/35*	*4/35*	*5/35*	*4/35*	*4/35*	*3/35*	*2/35*	*1/35*	*1/35*

Le niveau de significativité du protocole observé (autrement dit la fréquence cumulée de la modalité 7) est 2/35, soit une valeur supérieure à 5%. Ce résultat est en accord avec les indications fournies par la table : $W_S = 6$.

Enoncé 29

Un psychologue scolaire veut étudier l'influence du niveau d'étude des mères d'élèves d'une classe de lycéens sur la fréquence de leurs visites auprès de la direction de l'école. Il obtient, pour l'année 1998-1999 :

Niveau d'études	Nombre de visites de la mère
Etudes primaires	4, 3, 0, 7, 1, 2, 0, 3, 5, 1
CEP	2, 4, 1, 6, 3, 0, 2, 5, 1, 2, 1
BEPC	2, 0, 4, 3, 8, 0, 5, 2, 1, 7, 6, 5, 1
Baccalauréat	9, 4, 2, 3
Etudes supérieures	2, 4, 5, 2, 2, 6

1) Montrer que les protocoles de rangs des trois groupes sont donnés par :

```
30.0 25.0 3.0 41.5 9.0 17.5 3.0 25.0 35.0 9.0
17.5 30.0 9.0 39.0 25.0 3.0 17.5 35.0 9.0 17.5 9.0
17.5 3.0 30.0 25.0 43.0 3.0 35.0 17.5 9.0 41.5 39.0 35.0 9.0
44.0 30.0 17.5 25.0
17.5 30.0 35.0 17.5 17.5 39.0
```

2) On demande à un logiciel de traitements statistiques de réaliser un test de Kruskal-Wallis sur ces données. Le résultat produit est le suivant :

```
Kruskal-Wallis rank sum test
data : list(x1, x2, x3, x4, x5) Kruskal-Wallis chi-squared = 2.8532, df = 4,
p-value = 0.5827
```

Refaites les calculs et confirmez les résultats donnés par le logiciel.

3) On réalise le test de Kruskal-Wallis à l'aide de Statistica, qui fournit les résultats suivants :

```
ANOVA de Kruskal-Wallis par Rangs ; Nb Visites (Données Psy-Sco)
Var. indépendante (classement) : Niveau
Test de Kruskal-Wallis : H ( 4, N= 44) =2,853226 p =,582
```

Ces résultats sont-ils en accord avec les précédents ?

4) Le test de la médiane généralisée, réalisé avec Statistica, donne :

```
Test Médiane, Méd. Globale = 2,50000 ; Nb Visites (Données Psy-Sco)
Var. indépendante (classement) : Niveau
Chi-Deux = 1,895105 dl = 4 p = ,7550
```

Vérifiez les résultats et expliquer pourquoi le niveau de significativité du résultat (.75) est nettement plus élevé que celui du test de Kruskal-Wallis (.58).

Enoncé 30

On réalise une enquête sur la satisfaction professionnelle éprouvée par les personnes actives, selon la profession. La satisfaction professionnelle est mesurée par 18 facteurs, sur une échelle de 1 à 5. La somme des évaluations des 18 facteurs est utilisée comme mesure de la satisfaction professionnelle. Une évaluation élevée correspond à un fort degré de satisfaction professionnelle.

1) Pour un échantillon de 10 juristes et un échantillon de 10 analystes informatiques, les données observées sont les suivantes :

Juristes	41	42	42	44	45	48	50	53	64	76
Analystes	38	44	55	60	62	64	66	71	73	74

a) Un statisticien conseille aux auteurs de l'enquête d'utiliser un test non paramétrique pour étudier ces données. Quelles sont les raisons qui l'amènent à faire ce choix ?
b) Etudier, à l'aide d'un test de Wilcoxon-Mann-Whitney, si la satisfaction professionnelle est plus élevée chez les analystes que chez les juristes (test unilatéral au seuil de 5%). Compte tenu des tailles d'échantillons, on pourra, au choix, utiliser la table du test de Wilcoxon-Mann-Whitney ou l'approximation par une loi normale.

Réponse
a) La variable étudiée est un score numérique calculé à partir de variables ordinales. Bien qu'elle puisse prendre un nombre élevé de valeurs (nombres entiers compris entre 18 et 90), il peut sembler préférable d'utiliser un test ne faisant aucune hypothèse sur la distribution de cette variable dans les populations parentes.
b) On construit tout d'abord le protocole des rangs pour l'ensemble des 20 observations :

Juristes	*2*	*3.5*	*3.5*	*5.5*	*7*	*8*	*9*	*10*	*14.5*	*20*
Analystes	*1*	*5.5*	*11*	*12*	*13*	*14.5*	*16*	*17*	*18*	*19*

La somme des rangs dans le groupe des juristes est 83, celle observée dans le groupe des analystes est 127.
On réalise un test unilatéral en prenant comme hypothèses statistiques :
H_0 *: Les scores des juristes et ceux des analystes s'interclassent de manière homogène.*
H_1 *: La probabilité qu'un score observé chez un juriste soit inférieur à un score observé chez un analyste est supérieure à 50%.*
Les deux échantillons sont ici de taille 10. On prend comme valeur observée de la statistique de test $W = 83$. *La valeur critique (au seuil de 5%) lue dans la table est :* $W_s = 82$.
La règle de décision est donc :
– *Si* $W \leq 82$, *on retient* H_1
– *Si* $W > 82$, *on retient* H_0.
Par conséquent, on retient H_0.

Variantes de cette solution :
On sait que la statistique U de Mann-Whitney est donnée par :

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - W_2$$

$$U = \min(U_1, U_2)$$

On a donc ici : $U_1 = 155 - 83 = 72$, $U_2 = 155 - 127 = 28$, *d'où* $U = 28$. *La table de la statistique U de Mann-Whitney donne, pour un test unilatéral au seuil de 5%,* $U_{crit} = 27$, *et la conclusion reste la même.*

On peut aussi utiliser l'approximation par une loi normale :
Par exemple : $E^2 = \dfrac{21 \times 20 \times 20}{12 \times 10 \times 10} = 7$; $Z = \dfrac{8.3 - 12.7}{\sqrt{7}} = -1.66$.
Or, pour la loi normale centrée réduite la valeur critique correspondant à un test unilatéral "à gauche" est $Z_c = -1.645$. *La conclusion est encore la même.*

2) En fait, le tableau précédent ne concernait qu'une partie des données recueillies. L'étude a porté sur 4 professions : juristes, thérapeutes, ébénistes et analystes informatiques. L'ensemble des scores observés est donné par :

Juristes	41	42	42	44	45	48	50	53	64	76
Analystes	38	44	55	60	62	64	66	71	73	74
Thérapeutes	52	55	59	60	62	78	80	86		
Ebéniste	54	59	64	65	69	79	79			

et le protocole des rangs par :

	Juristes	Analystes	Thérapeutes	Ebénistes
	2	1	10	12
	3.5	5.5	13.5	15.5
	3.5	13.5	15.5	22
	5.5	17.5	17.5	24
	7	19.5	19.5	26
	8	22	31	32.5
	9	25	34	32.5
	11	27	35	
	22	28		
	30	29		
$\overline{R}_i$	10.15	18.8	22	23.5

1) Etudier, à l'aide d'un test de Kruskal-Wallis, si les scores des 4 professions sont significativement différents.
2) Dans le cas où la réponse à la question précédente est affirmative, réaliser les tests post hoc, de type Bonferroni-Dunn, comparant les professions deux à deux.

Réponses
1) Les hypothèses H_0 et H_1 peuvent ici être exprimées par :
H_0 : Les médianes des scores pour les 4 professions sont égales.
H_1 : Ces médianes sont différentes.
La statistique de test K est donnée par :

$$K = \left[\frac{12}{N(N+1)} \sum n_j \overline{R}_j^2\right] - 3(N+1)$$

Elle suit une loi du χ^2 à 3 ddl. La valeur critique, pour un seuil $\alpha = 5\%$ est : $\chi^2_{crit} = 7.815$.
La règle de décision est donc :
– Si $K > 7.815$, on retient l'hypothèse alternative H_1
– Si $K \leq 7.815$, on retient l'hypothèse nulle H_0.
Ici, le calcul donne :

$$K = \frac{12}{35 \times 36}(10 \times 10.15^2 + 10 \times 18.8^2 + 8 \times 22^2 + 7 \times 23.5^2) - 3 \times 36 = 9.16$$

En conclusion, on retient donc l'hypothèse H_1 : les degrés de satisfaction attachés aux 4 professions sont significativement différents.

2) Le nombre de comparaisons post hoc est ici $c = \frac{4 \times 3}{2} = 6$. *Pour obtenir un seuil global* $\alpha_{FW} = 5\%$, *les comparaisons seront donc faites au seuil* $\alpha_{PC} = \frac{5\%}{6} = 0.83\%$, *ce qui conduit, pour des tests bilatéraux à une valeur critique* $Z_c = 2.64$.
Par ailleurs, $\frac{N(N+1)}{12} = 105$.
– Comparaison Juristes v/s Analystes : $E^2 = 105\left(\frac{1}{10} + \frac{1}{10}\right) = 21$, $E = 4.58$ *et* $Z = \frac{18.8 - 10.15}{4.58} = 1.887$. *La différence n'est donc pas significative.*
– Comparaison Juristes v/s Thérapeutes : $E^2 = 105\left(\frac{1}{10} + \frac{1}{8}\right) = 23.65$, $E = 4.86$ *et* $Z = \frac{22 - 10.15}{4.86} = 2.44$. *La différence n'est donc pas significative. Le niveau de significativité (individuel) de cette comparaison est* $p_{PC} = .01477$ *et Statistica, par exemple, indique un niveau de significativité (familywise)* $p_{FW} = .0886$ *(6 fois la p-value individuelle).*
– Comparaison Juristes v/s Ébénistes : $E^2 = 105\left(\frac{1}{10} + \frac{1}{7}\right) = 25.5$, $E = 5.05$ *et* $Z = \frac{23.5 - 10.15}{5.05} = 2.6436$. *La différence est donc (tout juste) significative. Le niveau de significativité (individuel) de cette comparaison est* $p_{PC} = .0082$ *et Statistica, par exemple, indique un niveau de significativité (familywise)* $p_{FW} = .049204$.
– Comparaison Analystes v/s Thérapeutes : $E^2 = 105\left(\frac{1}{10} + \frac{1}{8}\right) = 23.65$, $E = 4.86$ *et* $Z = \frac{22 - 18.8}{4.86} = 0.65$. *La différence n'est donc pas significative.*
– Comparaison Analystes v/s Ébénistes : $E^2 = 105\left(\frac{1}{10} + \frac{1}{7}\right) = 25.5$, $E = 5.05$ *et* $Z = \frac{23.5 - 18.8}{5.05} = 0.93$. *La différence n'est donc pas significative.*
– Comparaison Analystes v/s Ébénistes : $E^2 = 105\left(\frac{1}{8} + \frac{1}{7}\right) = 28.13$, $E = 5.30$ *et* $Z = \frac{23.5 - 22}{5.30} = 0.28$. *La différence n'est donc pas significative.*
La seule différence significative semble être Juristes v/s Ébénistes. On constate sur cet exemple le caractère "conservateur" de ce test post hoc.

Enoncé 31

Ref. Wayne D. Parker, The Socionomic Perspective on Social Mood and Voting : Report on New Mood Measures in the 2006 ANES Pilot Study

Les études concernant le comportement des électeurs américains lors des scrutins nationaux se sont intéressées aux critères déterminant le vote. Pour l'essentiel, ces études se sont intéressées à des critères rationnels et cognitifs. A l'inverse, un chercheur a tenté de définir des critères moins rationnels, plus émotionnels. Ainsi, à partir de questions telles que :

– Lorsque vous pensez à votre futur, êtes-vous généralement optimiste, pessimiste, ou ni l'un ni l'autre ?

– Lorsque vous pensez au futur des Etats-Unis, êtes-vous généralement optimiste, pessimiste, ou ni l'un ni l'autre ?

il a défini deux échelles numériques sur lesquelles les sujets sont évalués : l'échelle "personal mood" ou "humeur personnelle" et l'échelle "social mood" ou "humeur sociale".

Pour chacune de ces échelles, il recherche ensuite s'il existe des différences significatives dans l'électorat, selon diverses variables indépendantes telles que le sexe, le niveau d'étude, la tendance politique, etc. En particulier, pour un échantillon de 160 sujets, on a déterminé :
– les scores sur chacune des deux échelles ;
– la tendance politique, avec trois modalités : républicain, démocrate, indépendant.

Le protocole des rangs appliqué à l'échelle "humeur personnelle" a conduit aux résultats suivants :

Tendance	Effectif	Somme des rangs	Rang moyen
Républicains	62	5 456.5	88.10
Démocrates	64	4 953.5	77.40
Indépendants	33	2 310	70.00
Total	159	12 720	80

N.B. Une non-réponse a été exclue de l'analyse.

Le même traitement, appliqué à l'échelle "humeur sociale" a conduit à :

Tendance	Effectif	Somme des rangs	Rang moyen
Républicains	62	5 838	94.16
Démocrates	65	4 771.5	73.41
Indépendants	33	2 270.5	68.80
Total	160	12 720	80.50

1) Pour chacune des deux échelles, rechercher s'il existe des différences significatives selon la tendance politique (seuil : 5%).

2) Pour l'échelle "humeur sociale", l'auteur fait ensuite un test de Wilcoxon Mann Whitney comparant les Républicains et les Démocrates. Il obtient un niveau de significativité $p = .008$ (test bilatéral).
a) Interpréter cette valeur.
b) Quel résultat obtient-on en réalisant un test post hoc de type Bonferroni-Dunn sur ces données ? Comparer avec le résultat précédent.

Indications de réponses : 1) On réalise ici des tests de Kruskal-Wallis. Pour l'échelle "Humeur personnelle", on obtient $K = 4.12$ ou $K = 3.68$ selon la formule utilisée (différence provoquée par les erreurs dues aux arrondis). Comme $\chi^2_c = 5.99$, on conclut sur H_0. Pour l'échelle "Humeur sociale", on obtient $K = 9.01$, ce qui semble montrer une différence entre les trois groupes.
2) a) Le test de Mann Whitney semble montrer une différence significative entre les Républicains et les Démocrates en ce qui concerne la variable "Humeur sociale", au seuil de 1%.
b) On a ici trois groupes. Le nombre théorique de comparaisons est donc $c = 3$. Pour réaliser un test de Bonferroni avec un seuil $\alpha_{FW} = 5\%$, on réalise donc les tests "par comparaison" au seuil $\alpha_{PC} = 1.67\%$. Pour la comparaison entre les Républicains et les Démocrates, la comparaison donne $Z = -2.52$, ce qui correspond, pour un test bilatéral, à une p-value $p = 1.18\%$, et confirme donc le résultat précédent.

Tests non paramétriques sur des groupes appariés

Enoncé 32
Dans le cadre d'une étude sur le tabagisme chez la femme enceinte, on interroge 100 sujets au 3è et au 8è mois de grossesse. On obtient les résultats suivants :

		Fumeur 8è mois	
		oui	non
Fumeur	oui	35	15
3è mois	non	5	45

Le comportement des sujets est-il le même dans les deux conditions ?
Réponse : Le χ^2 de Mac Nemar vaut ici $\chi^2 = 5$, ce qui est significatif d'une différence de comportement au seuil de 5%.

Enoncé 33
14 sujets sont observés dans deux conditions. On obtient 2 différences positives, 10 différences négatives, 2 différences nulles.
Quel est le niveau de significativité obtenu pour un test unilatéral ? pour un test bilatéral ?
Réponses. La statistique de test D_+ suit une loi binomiale de paramètres $N = 12$ et $p = 0.5$.
Calcul du niveau de significativité de $D_{+,obs}$ pour un test unilatéral :
$P(D_+ = 0) = C_{12}^0\ 0.5^{12} = 0.0002441$
$P(D_+ = 1) = C_{12}^1\ 0.5^{12} = 0.0029297$
$P(D_+ = 2) = C_{12}^2\ 0.5^{12} = 0.0161133$
D'où : $p = P(D_+ \leq 2) = 0.019 = 1.9\%$, pour un test unilatéral.
Pour un test bilatéral :
$p = P(D_+ \leq 2) + P(D_+ \geq 10) = 2P(D_+ \leq 2) = 3.8\%$.

Enoncé 34
40 sujets sont observés dans deux conditions. On obtient 10 différences positives, 30 différences négatives, 0 différence nulle.
Le test des signes met-il en évidence une différence de comportement entre les deux conditions ?
Réponse : On a ici : $D_+ = 10$ et $Z = \dfrac{20 + 1 - 40}{\sqrt{40}} = -3.00$
Au seuil de 1% unilatéral, on retient H_1 : les différences négatives sont significativement plus nombreuses que les différences positives.

Enoncé 35
On a testé huit sujets dans deux conditions A_1 et A_2. On obtient le protocole suivant :

Suj.	A_1	A_2
s1	100	105
s2	70	63
s3	40	50
s4	123	98
s5	92	60
s6	120	78
s7	172	119
s8	173	101

Etudier s'il existe une différence significative entre les deux conditions à l'aide d'un test des rangs signés de Wilcoxon.

Réponse. Contruction du protocole des rangs signés :

Suj.	A_1	A_2	d_i	$\|d_i\|$	r_{i+}	r_{i-}
s1	*100*	*105*	*5*	*5*	*1*	
s2	*70*	*63*	*-7*	*7*		*2*
s3	*40*	*50*	*10*	*10*	*3*	
s4	*123*	*98*	*-25*	*25*		*4*
s5	*92*	*60*	*-32*	*32*		*5*
s6	*120*	*78*	*-42*	*42*		*6*
s7	*172*	*119*	*-53*	*53*		*7*
s8	*173*	*101*	*-72*	*72*		*8*
T					*4*	*32*

On trouve $T_+ = 4$, $T_- = 32$ *et donc* $T_m = 4$.
Au seuil de 5% unilatéral, on lit dans la table : $T_{crit} = 5$.
Comme $T_m < T_{crit}$, *on conclut à une différence significative entre les conditions* A_1 *et* A_2 *au seuil de 5% unilatéral.*

Enoncé 36

Nous nous intéressons à l'influence du style d'interview sur les réponses des sujets à une enquête d'opinion. Nous pourrions entraîner un enquêteur à mener trois types différents d'interviews :

– Interview 1 : intérêt, ton amical, enthousiasme,
– Interview 2 : formalisme, réserve, courtoisie,
– Interview 3 : manque d'intérêt, ton abrupt, formalisme pesant.

L'enquêteur visite ensuite trois groupes de 18 foyers, et utilise le style 1 avec un groupe, le style 2 avec le 2è groupe et le style 3 avec le dernier groupe. Nous obtenons ainsi 18 triplets de foyers, comprenant chacun 3 foyers appariés selon des variables pertinentes. Pour chaque triplet, les trois éléments sont affectés au hasard aux trois conditions (styles d'interview). Nous mesurons ensuite l'effet du style d'interview en notant la réponse faite (oui/non) à un item particulier. Les données obtenues sont les suivantes :

Triplet	Style 1	Style 2	Style 3
1	0	0	0
2	1	1	0
3	0	1	0
4	0	0	0
5	1	0	0
6	1	1	0
7	1	1	0
8	0	1	0
9	1	0	0
10	0	0	0
11	1	1	1
12	1	1	1
13	1	1	0
14	1	1	0
15	1	1	0
16	1	1	1
17	1	1	0
18	1	1	0

1) Etudier à l'aide d'un test Q de Cochran si le style d'interview a une influence sur les réponses obtenues.
2) Le cas échéant, réaliser les tests post hoc comparant les styles d'interview deux à deux.

Réponses.
1) On obtient ici $G_1 = 13$, $G_2 = 13$, $G_3 = 3$, $G = 29$, $\sum L_i^2 = 63$, *d'où* $Q = 16.7$. *Pour ddl* $= 2$ *et un seuil de 1%, la table donne :* $\chi^2_{crit} = 9.21$. *Les fréquences de réponses positives sont donc différentes selon les styles.*
On peut remarquer que 6 des 18 lignes sont entièrement composées de 0 ou entièrement composées de 1. En reprenant le calcul sur les 12 lignes restantes, on constate que la valeur de la statistique de test reste la même. On obtient alors : $G'_1 = 10$, $G'_2 = 10$, $G'_3 = 0$, $G = 20$, $\sum L'^2_i = 36$, *et, comme précédemment,* $Q = 16.7$
2) Les comparaisons Style 1 v/s Style 3 et Style 2 v/s Style 3 utiliseront les mêmes calculs, tandis que la comparaison Style 1 v/s Style 2 conclura évidemment sur H_0. *Réalisons la comparaison Style 1 v/s Style 3.*
Ici : $E^2 = 2\,\dfrac{3 \times 29 - 63}{18^2 \times 3 \times 2} = 0.0247$; $E = 0.16$; $p_1 = 0.72$; $p_2 = 0.17$. *D'où* $Z_{obs} = 3.53$.
Pour un seuil $\alpha_{PC} = 5\%/3 = 1.67\%$ *et un test bilatéral, on a :* $Z_{crit} = 2.39$. *Les deux styles d'interviews conduisent donc à des résultats significativement différents.*

Enoncé 37

Source : Exemple Synchron.sta fourni avec Statistica

Lorsque nous analysons un discours, nous faisons également attention aux signaux visuels ; plus précisément, nous comprenons beaucoup plus facilement un discours lorsque nous pouvons voir le visage de la personne qui parle. D'une certaine manière, "nous lisons tous sur les lèvres", du moins, dans une certaine mesure. Dodd a tenté de déterminer si des enfants âgés de 10 à 16 semaines étaient déjà conscients de la relation entre les mots et le mouvement correspondant des lèvres (de la personne qui les prononce). Dans cette

optique, Dodd a placé les enfants dans une pièce où ils pouvaient voir la personne lisant un texte normal à travers une vitre. Le discours a été diffusé soit simultanément dans la pièce (conditions synchrones), soit avec un décalage de 400 millisecondes (conditions asynchrones). La variable dépendante était la durée pendant laquelle l'enfant regardé le visage à travers la vitre. Nous n'avons formulé aucune hypothèse quant à la condition particulière qui serait le plus susceptible d'attirer l'attention des jeunes enfants (il se peut que le discours asynchrone soit plus intéressant pour eux car il est nouveau, ou au contraire qu'ils détournent leur attention car ils perçoivent que le visage n'est pas à l'origine du discours).
Remarque : les noms d'observations sont constitués des initiales des individus.

	SYNCHRO	DECALAGE
DC	20.3	50.4
MK	17.0	87.0
VH	6.5	25.1
JM	25.0	28.5
SB	5.4	26.9
MM	29.2	36.6
RH	2.9	1.0
DJ	6.6	43.8
JD	15.8	44.2
ZC	8.3	10.4
CW	34.0	29.9
AF	8.0	27.7

Etudier, à l'aide d'un test des signes, puis d'un test des rangs signés de Wilcoxon, si les données recueillies confirment l'hypothèse du chercheur.
Réponses. On observe 2 différences positives sur 12 différences observées. Or, pour la loi binomiale de paramètres $N = 12$ et $p = 0.5$, on a : $P(X \leq 2) = 0.0193$ et, par symétrie, $P(X \leq 2) + P(X \geq 10) = 0.039$. On conclut donc à une différence significative entre les deux conditions, au seuil de 5 %. Après avoir déterminé le protocole des rangs signés, on obtient $T_{min} = 5$ (somme des rangs des différences positives). Or, pour $N = 12$, $\alpha = 2.5\%$ et un test unilatéral, la table donne $T_{m,crit} = 13$. La valeur trouvée est donc significative d'une différence au seuil de 5% bilatéral.

Enoncé 38

Un ergonome désire étudier la forme la plus économique pour un orifice dans lequel des ouvriers doivent faire passer une fiche. Il compare cinq formes d'orifices de moins en moins évasés. Grâce à un appareillage avec cellule photo-électrique, il mesure en millièmes de seconde le temps mis par un ouvrier pour mettre la fiche en position dans l'orifice. Chaque sujet effectue plusieurs essais et l'ergonome note pour chacun le temps médian. Pour 7 sujets, il a obtenu :

Sujet	Taille 1	Taille 2	Taille 3	Taille 4	Taille 5
1	244	417	178	195	452
2	235	307	225	346	613
3	308	290	257	427	438
4	343	305	290	215	534
5	254	263	252	340	469
6	251	291	417	263	445
7	333	414	414	276	441

1) Etudier, à l'aide d'un test de Friedman, si les médianes correspondant aux 5 formes d'orifices sont égales.
2) Le cas échéant, réaliser les tests post hoc comparant les tailles deux à deux.

Réponses. 1) Le protocole des rangs par sujet est donné par :

Sujet	*Taille 1*	*Taille 2*	*Taille 3*	*Taille 4*	*Taille 5*
1	*3*	*4*	*1*	*2*	*5*
2	*2*	*3*	*1*	*4*	*5*
3	*3*	*2*	*1*	*4*	*5*
4	*4*	*3*	*2*	*1*	*5*
5	*2*	*3*	*1*	*4*	*5*
6	*1*	*3*	*4*	*2*	*5*
7	*2*	*3.5*	*3.5*	*1*	*5*
R_j	*17*	*21.5*	*13.5*	*18*	*35*
R_j^2	*289*	*462.25*	*182.25*	*324*	*1225*

$$F_r = \frac{12 \times (289 + 462.25 + 182.25 + 324 + 1225)}{7 \times 5 \times 6} - 3 \times 7 \times 6 = 15.86$$

Pour un seuil de 5% et 4 ddl, on a : $\chi^2_{crit} = 9.49$*. On peut donc rejeter l'hypothèse nulle et conclure qu'il y a une influence de la forme de l'orifice sur le temps d'exécution de la tâche.*
Pour effectuer la correction pour les ex aequo de la 7ème ligne, on calcule :
– la variance de la série (1, 2, 3, 4, 5) : $s_c^2 = 2.5$
– la variance de la série (1, 2, 3.5, 3.5, 5) : $s'^2_c = 2.375$
– Le rapport : $\lambda = \dfrac{7 \times 2.5}{6 \times 2.5 + 2.375} = 1.0072$
– La valeur de F_r *ajustée :* $F'_r = \lambda F_r = 1.0072 \times 15.86 = 15.97$
En traitant cet exemple à l'aide de Statistica, on pourra constater que le logiciel fait cette correction. La conclusion est inchangée.

2) On calcule d'abord $E^2 = \frac{7\times5\times6}{6} = 35$ *d'où* $E = 5.91$*. Le nombre théorique de comparaisons est ici égal à* $\frac{5\times4}{2} = 10$*. Pour un résultat global au seuil de 5%, les tests individuels seront faits au seuil de 0.5%, ce qui correspond, pour des tests bilatéraux, à une valeur critique* $Z_c = 2.80$*.*
Pour la comparaison Taille 1 v/s Taille 2, nous obtenons : $Z_{12} = \frac{17-21.5}{5.91} = -0.76$*. La différence n'est donc pas significative.*
Plus généralement, on peut remarquer que le dénominateur E est indépendant de la comparaison faite. Deux conditions seront donc significativement différentes si la différence des sommes de rangs est, en valeur absolue, supérieure à $Z_cE = 2.80 \times 5.91 = 16.61$*. Ces différences sont données par :*

	Taille 1	*Taille 2*	*Taille 3*	*Taille 4*	*Taille 5*
Taille 1	—	*4.5*	*3.5*	*1*	*18*
Taille 2	*4.5*	—	*8*	*3.5*	*13.5*
Taille 3	*3.5*	*8*	—	*4.5*	*21.5*
Taille 4	*1*	*3.5*	*4.5*	—	*17*
Taille 5	*18*	*13.5*	*21.5*	*17*	—

Ainsi, selon ce test, les différences significatives sont observées entre la taille 5 et les tailles 1, 3 et 4.

Enoncé 39

L'hypnose : dans une expérimentation pratiquée en 1975, Lehman a enregistré le "potentiel cutané" en millivolts chez 8 sujets qui, par ailleurs, étaient interrogés sur la coloration psychique "crainte, joie, tristesse, calme" sous hypnose. Voici le tableau des observations :

	fear	joy	sadness	calmness
1	23.1	22.7	22.5	22.6
2	57.6	53.2	53.7	53.1
3	10.5	9.7	10.8	8.3
4	23.6	19.6	21.1	21.6
5	11.9	13.8	13.7	13.3
6	54.6	47.1	39.2	37
7	21.0	13.6	13.7	14.8
8	20.3	23.6	16.3	14.8

Etudier si l'effet de la coloration psychique sur le potentiel cutané est significatif à l'aide d'un test non paramétrique.

Réponse : le tableau des rangs par sujet s'écrit :

	1	*2*	*4*	*3*
	1	*3*	*2*	*4*
	2	*3*	*1*	*4*
	1	*4*	*3*	*2*
	4	*1*	*2*	*3*
	1	*2*	*3*	*4*
	1	*4*	*3*	*2*
	2	*1*	*3*	*4*
R_j	*13*	*20*	*21*	*26*
R_j^2	*169*	*400*	*441*	*676*

D'où $F_r = \dfrac{12}{8 \times 4 \times 5}(169 + 400 + 441 + 676) - 3 \times 8 \times 5 = 6.45$

La différence entre les conditions n'est pas significative.

Enoncé 40

Supposons que l'on demande à trois mélomanes d'une revue d'écouter 6 versions différentes d'une symphonie de Beethoven et de les ranger séparément suivant l'organisation des plans sonores (qui ressortissent de l'organisation spatiale des instruments, laquelle varie en général grandement selon le chef d'orchestre). Les trois séries indépendantes de rangs données par les trois mélomanes A, B, C sont exposées dans le tableau suivant :

	a	b	c	d	e	f
A	1	6	3	2	5	4
B	1	5	6	4	2	3
C	6	3	2	5	4	1

Les six versions sont-elles appréciées de la même façon ? Répondre à cette question à l'aide d'un test de Friedman.
Réponse : On a $F = 2.429$*. On n'a pas mis en évidence de différence entre les différentes versions.*

Corrélation et régression linéaires

Enoncé 41

On étudie la relation entre l'autoritarisme des étudiants et leur conformisme social. L'autoritarisme des sujets et leur conformisme social sont appréciés par le passage de tests.

étudiant	conformisme	autoritarisme
A	82	42
B	98	46
C	87	39
D	40	37
E	116	65
F	113	88
G	111	86
H	83	56
I	85	62
J	126	92
K	106	54
L	117	81

Vérifier que les protocoles des rangs sont donnés par :

étudiant	conformisme	autoritarisme	d_i^2
A	2	3	1
B	6	4	4
C	5	2	9
D	1	1	0
E	10	8	4
F	9	11	4
G	8	10	4
H	3	6	9
I	4	7	9
J	12	12	0
K	7	5	4
L	11	9	4
			52

Calculer la valeur du coefficient de corrélation de rangs de Spearman et tester la significativité de ce coefficient.

Réponse : on trouve $R_s = 0.82$, significatif au seuil de 5% bilatéral.

Calculer de même le coefficient τ de Kendall et tester sa significativité.

Réponse : $\tau = 0.67$. Significatif au seuil de 5% bilatéral

Exercice 42

Des chercheurs se sont intéressés à la relation entre la familiarité des noms de personnes et l'apparition de blocages dans une tâche de dénomination de visages. Plusieurs études antérieures montrent que les blocages portant sur des noms de personnes familières sont plus fréquents que les blocages portant sur des noms de personnes moins bien connues (effet de familiarité inversé). Cependant, l'effet inverse a été obtenu dans une étude de laboratoire au cours de laquelle le nombre d'essais de récupération était contrôlé (effet de familiarité direct).

Dans leur étude, les chercheurs étudient notamment la corrélation entre le taux de blocage et le score de familiarité de la personne. Les individus statistiques sont ici les 32 stimulis (photographies de personnalités connues du show business).

a) Le coefficient de corrélation de Pearson obtenu est $r = -0.455$. Ce coefficient de corrélation est-il significatif d'une corrélation entre les deux variables étudiées ?

b) A titre de contrôle, les auteurs ont également calculé le coefficient de corrélation des rangs de Spearman. Ils ont obtenu : $R_s = -0.515$. Etudier, de même, la significativité d'un tel coefficient.

c) Comment peut-on interpréter ces coefficients de corrélation : semblent-ils indiquer un effet de familiarité direct ou un effet de familiarité inversé ?

Réponses
a) Soit ρ le coefficient de corrélation entre les deux variables dans la population parente. L'hypothèse nulle est $\rho = 0$, pendant que l'hypothèse alternative est $\rho \neq 0$. Pour $n - 2 = 30$ ddl, la table du coefficient de corrélation donne $r_{crit} = 0.4487$ pour un test bilatéral au seuil de 1%. Comme $|r| > r_{crit}$, on retient l'hypothèse alternative : il existe une corrélation non nulle entre les deux variables.
Variante :
On peut aussi utiliser ici la statistique $T = \sqrt{n-2}\ \frac{r}{\sqrt{1-r^2}}$. On obtient : $T_{obs} = -2.80$. Or, pour un test bilatéral au seuil de 1%, avec ddl $= 30$ la table indique : $T_{crit} = 2.75$. La conclusion est identique.

b) Au seuil de 1%, pour $n = 30$ et $n = 35$, la table du test de corrélation des rangs de Spearman indique respectivement $R_c = 0.467$ et $R_c = 0.433$. La valeur observée $R_s = -0.515$ indique donc une corrélation des rangs significative entre les deux variables.
Variante :
On peut aussi utiliser la statistique : $Z = \sqrt{N-1}\ R_s$. On a ici $Z_{obs} = -2.87$. Or, pour un test bilatéral au seuil de 1%, la table de la loi normale indique : $Z_{crit} = 2.575$. Comme $|Z_{obs}| > Z_{crit}$, on conclut encore sur l'hypothèse H_1 : il existe une corrélation des rangs significative entre les deux variables.

c) Les coefficients de corrélation trouvés sont négatifs. Autrement dit, plus le score de familiarité est élevé, plus le taux de blocages est faible. Cette étude semble donc montrer un effet de familiarité directe et non l'inverse.

Exercice 43

Dans un article publié en 2004, des chercheurs ont analysé le rôle du contrat psychologique, selon l'influence de ses contenus relationnels et transactionnels, dans la prévention du harcèlement moral au travail.
Pour tester leurs hypothèses, ils ont soumis un questionnaire semi-structuré à un échantillon de 265 travailleurs dans trois PME de l'Italie du Centre.

Le contrat psychologique a été appréhendé à l'aide d'une série de 12 items concernant le rapport à l'organisation, que les participants évaluaient sur une échelle de type Likert en 5 points (1= "beaucoup moins que ce qui était promis" ; 5 = "beaucoup plus que ce qui était promis"). Le questionnaire permet de saisir les deux dimensions du contrat psychologique, relationnelle (6 items) et transactionnelle (6 items).
La dimension "violation du contrat psychologique" a été mesurée à l'aide d'une série de quatre items, avec une échelle en 7 points. Un exemple d'item de la violation est : "je me sens trahi par mon organisation".

La variable RCP_R (respect du contrat psychologique relationnel) est la moyenne des scores des 6 items du questionnaire se rapportant à cette dimension. On définit de même les variables RCP_T (respect du contrat psychologique transactionnel) et VCP (violation du contrat psychologique).

1) Le tableau ci-dessous est un extrait des données observées (variables RCP_R et VCP pour 10 sujets).

Sujet	RCP_R	VCP
S1	1.67	3.75
S2	1.83	3.50
S3	2.00	4.00
S4	2.33	4.50
S5	2.50	1.50
S6	2.67	1.75
S7	3.00	4.25
S8	3.17	0.50
S9	3.33	1.00
S10	3.50	0.75

a) Pour étudier la corrélation entre ces deux variables, on choisit de calculer un coefficient de corrélation des rangs de Spearman plutôt qu'un coefficient de corrélation de Bravais-Pearson. Pourquoi ?

b) Calculer le coefficient de corrélation des rangs de Spearman sur cet ensemble de données.

Réponse : $R_s = -0.62$

2) Pour l'ensemble des observations, les auteurs ont obtenu les coefficients de corrélation des rangs indiqués ci-dessous.

	RCP_R	RCP_T	VCP
RCP_R	1		
RCP_T	0.34	1	
VCP	-0.59	-0.14	1

Parmi ces coefficients de corrélation, quels sont ceux qui sont significatifs à 5% , à 1% ?

Réponse : On peut utiliser la statistique de test $Z = \sqrt{N-1}R_s$ *avec* $N = 265$.
Pour un seuil de 5% bilatéral, on a $Z_{crit} = 1.96$ *et donc* $|R_{s,crit}| = \frac{1.96}{\sqrt{264}} = 0.12$. *Tous les coefficients de corrélation trouvés sont significatifs au seuil de 5%.*
Pour un seuil de 1% bilatéral, on a $Z_{crit} = 2.57$ *et donc* $|R_{s,crit}| = \frac{2.57}{\sqrt{264}} = 0.158$. *Le coefficient* $R_s(VCP, RCP_T)$ *n'est pas significatif pour ce seuil.*

Exercice 44

Un questionnaire a été soumis à 15 sujets. Les réponses aux questions $Q1$ à $Q7$, données sur une échelle de Likert en 7 points sont indiquées ci-dessous :

Suj	$Q1$	$Q2$	$Q3$	$Q4$	$Q5$	$Q6$	$Q7$	S
s1	4	5	4	6	4	7	4	34
s2	4	3	4	5	5	4	4	29
s3	4	4	3	5	7	5	3	31
s4	4	3	6	5	4	2	5	29
s5	4	6	7	5	5	3	4	34
s6	5	6	6	8	4	5	6	40
s7	5	5	5	5	6	6	5	37
s8	4	3	2	5	5	5	3	27
s9	4	4	5	3	7	4	4	31
s10	4	3	5	3	1	3	4	23
s11	6	6	7	6	5	5	7	42
s12	2	4	3	1	4	4	5	23
s13	4	3	3	3	6	6	5	30
s14	5	6	5	4	3	3	5	31
s15	7	5	7	7	5	4	6	41

1) La matrice des corrélations entre les questions est donnée par :

	$Q1$	$Q2$	$Q3$	$Q4$	$Q5$	$Q6$	$Q7$
$Q1$	1.00	0.49	0.64	0.71	0.07	0.08	0.57
$Q2$	0.49	1.00	0.61	0.48	-0.01	0.11	0.52
$Q3$	0.64	0.61	1.00	0.49	-0.17	-0.39	0.64
$Q4$	0.71	0.48	0.49	1.00	0.08	0.23	0.32
$Q5$	0.07	-0.01	-0.17	0.08	1.00	0.40	-0.14
$Q6$	0.08	0.11	-0.39	0.23	0.40	1.00	0.00
$Q7$	0.57	0.52	0.64	0.32	-0.14	0.00	1.00

Quelles sont les questions entre lesquelles la corrélation est significative au seuil de 5% bilatéral ?

2) On souhaite évaluer la cohérence des réponses aux questions 1 à 7 ci-dessous.
Les moyennes et variances des variables $Q1$ à $Q7$ et de leur somme S sont données par :

	Moyenne	Variance
$Q1$	4.40000	1.25714
$Q2$	4.40000	1.54286
$Q3$	4.80000	2.60000
$Q4$	4.73333	3.06667
$Q5$	4.73333	2.35238
$Q6$	4.40000	1.82857
$Q7$	4.66667	1.23810
S	32.13333	34.98095

a) Calculer le coefficient α de Cronbach correspondant et commenter.

b) Pour deux des sept questions, la suppression de la question entraîne l'augmentation du coefficient. Sans faire de calculs, indiquer quelles sont vraisemblablement ces questions.

Indication : On obtient $\alpha = .7036$. Les questions dont la suppression entraînerait une hausse de α sont $Q5$ et $Q6$.

Exercice 45 *Données Budget*
Il s'agit d'un extrait d'une enquête (ONU 1967) sur les budgets-temps (temps passé dans différentes activités au cours de la journée).
Les colonnes comprennent 3 variables numériques, le temps passé en : Profession (PROF), Transport (TRAN) et loisirs (LOIS). Les temps sont notés en centièmes d'heures. Le code suivant est utilisé pour identifier les lignes :
H : hommes, F : femmes, A : actifs, N : non actifs, M : mariés, C : célibataires,
U : USA, W : pays de l'ouest, E : Est sauf Yougoslavie, Y : Yougoslavie.

Budget	PROF	TRAN	LOIS
HAU	610	140	315
FAU	475	90	305
FNU	10	0	430
HMU	615	140	305
FMU	179	29	373
HCU	585	115	385
FCU	482	94	336
HAW	653	100	330
FAW	511	70	262
FNW	20	7	368
HMW	656	97	321
FMW	168	22	311
HCW	643	105	388
FCW	429	34	392
HAY	650	140	365

Budget	PROF	TRAN	LOIS
FAY	560	105	235
FNY	10	10	380
HMY	650	145	358
FMY	260	52	295
HCY	615	125	475
FCY	433	89	408
HAE	650	142	334
FAE	578	106	228
FNE	24	8	398
HME	652	133	310
FME	436	79	231
HCE	627	148	463
FCE	434	86	380
Moy	451	86	346
Ety	223	47	63

1) Représenter le nuage de points correspondant aux variables PROF et TRAN, puis celui correspondant aux variables PROF et LOIS.

2) Calculer la covariance et le coefficient de corrélation pour le couple de variables (PROF, TRAN), puis pour le couple (PROF, LOIS). Dans chacun des deux cas, la corrélation est-elle significative ?

3) Déterminer l'équation de la droite de régression de TRAN selon les valeurs de PROF. Quelle est la part de la variance de TRAN qui est "expliquée" par PROF ?

Réponses : 2) Cov(PROF,TRAN)=9805.12 , r(PROF,TRAN)=0.93 ; Cov(PROF,LOIS)=-2651.87 , r(PROF,LOIS)=-0.19. Seule la corrélation entre PROF et TRAN est significative. L'équation de la droite de régression est : TRAN=0.1977 PROF - 3.15. La part de la variance de TRAN "expliquée par" PROF est de $\frac{Var(\widehat{TRAN})}{Var(TRAN)} = r^2 = 0.87$.

Exercice 46 *Données Tailles*
Le tableau ci-dessous donne la taille de 10 garçons (variable Z) ainsi que la taille de leurs parents (le père X et la mère Y).

	X	Y	Z
i1	160.0	161	165.0
i2	165.0	155	162.5
i3	170.0	155	165.0
i4	172.5	165	175.0
i5	175.0	170	180.0
i6	180.0	166	177.5
i7	185.0	167	180.0
i8	187.5	172	190.0
i9	190.0	175	195.0
i10	195.0	168	187.5

On cherche s'il existe une relation entre la taille du fils et celle de ses parents et, si oui, quelle est la part respective de la mère et du père. Pour cela, on procède à la régression de Z sur X et Y. On donne les résultats intermédiaires suivants :

$\sum X_i = 1780$; $\sum Y_i = 1654$; $\sum Z_i = 1777,5$
$\sum X_i^2 = 318012.5$, $\sum Y_i^2 = 273974$; $\sum Z_i^2 = 317068.75$
$\sum X_iY_i = 294932.5$; $\sum X_iZ_i = 317437.5$; $\sum Y_iZ_i = 294632.5$
$Var(X) = 117.25$; $Var(Y) = 40.24$; $Var(Z) = 111.81$
$Cov(X,Y) = 52.05$; $Cov(X,Z) = 104.25$; $Cov(Y,Z) = 63.40$.

1) Quels sont les coefficients de corrélation des variables prises deux à deux ?

2) On utilise un logiciel de traitement statistique pour déterminer l'équation du plan de régression de Z par rapport à X et Y. On obtient :

$Z = 0.4455X + 0.9993Y - 66.83.$

Calculer les valeurs estimées de Z pour chacun des 10 individus statistiques (variable $\hat{Z}$).
On donne par ailleurs : $\sum \hat{Z}_i^2 = 317060.06$ et $\sum Z_i\hat{Z}_i = 317054.34$.

3) Déterminer le coefficient de corrélation multiple.

4) Quelle est la proportion de variance prise en compte par la régression ?

5) Les coefficients de corrélation partielle sont donnés par : $R_{xz,y} = 0.91; R_{yz,x} = 0.95$
Quel est, de la taille du père et de celle de la mère, le meilleur prédicteur de la taille du fils ?

6) Prédire la taille d'un garçon, sachant que son père mesure 188cm et sa mère 171cm.

Réponses : N.B. Calculs exécutés à l'aide d'un logiciel de traitement statistique.

1) Les coefficients de corrélation des variables prises deux à deux sont donnés par : r(X,Y)=0.76 ; r(X,Z)=0.91 ; r(Y,Z)=0.95.

2) Les valeurs estimées de Z sont données par :

i1	*i2*	*i3*	*i4*	*i5*	*i6*	*i7*	*i8*	*i9*	*i10*
165.34	*161.57*	*163.8*	*174.90*	*181.01*	*179.24*	*182.47*	*188.58*	*192.69*	*187.92*

3) Coefficient de corrélation multiple : R=0.991.

4) D'où $R^2 = 0.98$*. Le modèle explique 98% de la variance observée de la variable Z. Cette proportion est très élevée, mais il s'agit de données fictives....*
5) Le coefficient de corrélation partielle le plus élevé est celui liant taille de la mère et taille du fils.
6) Taille du fils si X=188 et Y=171 : Z = 188.

Exercice 47

Aux élections européennes de juin 1984, les votes pour la liste du Front National ont été très variables dans l'espace et leur comparaison avec d'autres variables fait apparaître un certain nombre de relations. Les variables choisies dans le tableau 2 sont les suivantes :
– LPEN : pourcentage de voix de la liste FN ;
– ETRA : pourcentage d'étrangers dans la population en 1982 ;
– DELI : Nombre pondéré de délinquance pour 100 habitants en 1980 ;
– CRCH : Taux mensuel moyen d'évolution du chômage entre le 31.08.81 et le 30.04.83 ;
– TXCH : Pourcentage de chômeurs dans la population active au 30.09.83 ;
– URBA : Pourcentage de population urbaine en 1982.
Les individus statistiques sont ici les régions de France Métropolitaine (ILEF=Ile de France, CHAM=Champagne-Ardennes, etc.). L'échelle régionale n'est certainement pas la meilleure pour une telle étude et les conclusions valent pour les agrégats spatiaux et non des personnes.[5]
Le tableau 3 donne les valeurs des coefficients de corrélation des variables prises deux à deux. On voit que les votes pour l'extrême droite sont fortement corrélés à trois variables : taux d'urbanisation (URBA), taux de délinquance (DELI) et taux d'étrangers (ETRA). D'autre part, ces trois variables sont fortement corrélées entre elles, elles sont donc partiellement redondantes.
Une première régression multiple est réalisée en utilisant les 5 variables explicatives. Le coefficient de corrélation multiple vaut $R = 0.934$ et le coefficient de détermination, $R^2 = 0.872$
Les coefficients de corrélation partielle entre la variable LEPEN et chacune des autres variables sont alors ceux indiqués dans le tableau 4.
On peut tester la significativité de ces coefficients de corrélation. Le nombre de degrés de liberté à prendre en compte est $21 - 6 = 15$. Au seuil de 5%, $r_{crit} = 0.4821$.
On retire alors la variable qui a le plus faible coefficient de corrélation partielle, c'est-à-dire URBA et on réalise une régression multiple de la variable LPEN par rapport aux quatre variables explicatives restantes.
On trouve alors : $R = 0.930$, $R^2 = 0.865$ et les nouveaux coefficients de corrélation partielle indiqués dans le tableau 5. Notons que R ne change pratiquement pas : la variable URBA n'apporte pas d'information supplémentaire par rapport aux quatre variables restantes.

[5]D'après *Initiation aux méthodes statistiques en Géographie*, Groupe Chadule, Masson Ed., 1994

REG	LEPEN	ETRA	DELI	CRCH	TXCH	URBA
ILEF	14.5	13.3	6	0.23	7.1	93.6
CHAM	10.7	5.4	4	0.07	9.5	62.4
PICA	10.8	4.6	4	0.22	9.7	60.7
HNOR	8.9	3.3	4	0.01	11	69.1
CENT	9.3	5.1	3	0.51	7.8	62.9
BNOR	7.6	1.7	4	0.38	9.8	53.4
BOUR	10.1	5.4	3	0.72	8.6	57.9
NORD	9.1	4.8	4	0.21	11.8	86.4
LORR	12.4	8	4	0.51	9.2	72.4
ALSA	12.5	8.1	3	1.25	7.4	73.2
FCOM	12	7.4	4	0.19	8.2	58.8
PAYS	6.8	1.4	3	0.58	9.6	60.1
BRET	6.8	0.7	3	0.84	9.4	55.6
POIT	6.7	1.7	3	0.48	10	50.5
AQUI	8.3	4.6	4	0.85	9.5	64.6
MIDI	8.1	4.8	3	0.54	8.5	59.3
LIMO	4.8	2.7	3	0.57	6.9	50.9
RHON	12.9	9.1	4	0.57	7.5	76.9
AUVE	7.4	4.6	2	0.85	8.3	58.2
LANG	13.2	6.5	4	1.44	11.4	70.7
PROV	19	8.2	6	1.13	10.5	89.6

Tab. 2 – Voix du FN aux élections européennes de 1984

	LEPEN	ETRA	DELI	CRCH	TXCH	URBA
LEPEN	1	0.81	0.76	0.25	0.05	0.77
ETRA		1	0.62	0.08	-0.35	0.76
DELI			1	-0.14	0.19	0.75
CRCH				1	0.00	0.08
TXCH					1	0.13
URBA						1

Tab. 3 – Corrélations entre les variables

	ETRA	DELI	CRCH	TXCH	URBA
LPEN	0.6910802	0.52652694	0.5494687	0.45550366	-0.2161422

Tab. 4 – Corrélation partielle entre LEPEN et les 5 variables

	ETRA	DELI	CRCH	TXCH
LEPEN	0.73354021	0.49258622	0.5282723	0.4116398

Tab. 5 – Corrélation partielle entre LEPEN et 4 variables

Pour tester la significativité de ces coefficients, on prend ici $ddl = 16$ et donc $r_{crit} = 0.4683$. Retirons de même la variable qui a le plus faible coefficient de corrélation partielle, c'est-à-dire TXCH.
On trouve alors : $R = 0.915$, $R^2 = 0.838$ et les coefficients de corrélation partielle du tableau 6.

	ETRA	DELI	CRCH
LEPEN	0.6829056	0.6825633	0.5516189

Tab. 6 – Corrélation partielle entre LEPEN et 3 variables

A ce stade, $r_{crit} = 0.4683$, et tous les coefficients sont significatifs. On peut donc dire que les votes pour l'extrême-droite aux élections européennes de 1984, à l'échelle régionale, ont varié en fonction de trois circonstances : le taux d'étrangers, le taux de délinquance, et à un degré moindre, l'évolution du chômage.
La régression n'a pas été faite dans un but de prévision, et l'équation de régression n'a qu'un intérêt limité :

$$\text{LPEN} = 0.52\ \text{ETRA} + 1.69\ \text{DELI} + 2.37\ \text{CRCH} - 0.4$$

Analyse de médiation

Ref. Costarelli, S., Callà, R.-M.. Self-directed negative affect : The distinct roles of ingroup identification and outgroup derogation, Current research in Social Psychology, Volume 10 No 2, 2004.

Le Sud-Tyrol est une région de l'Italie du Nord dans laquelle coexistent une population de langue italienne et une population de langue allemande. La population de langue allemande a fait l'objet d'une discrimination négative durant le régime fasciste, puis a bénéficié de dispositions favorables ensuite. De ces événements résulte un fort sentiment d'appartenance à un groupe pour les membres de chacun de ces deux groupes ethniques. Une enquête par questionnaire a été menée en 2002 auprès d'un échantillon de 71 lycéens italophones. En particulier, les sujets devaient se positionner sur des échelles unipolaires à 6 points (cotées de 0 à 5, 0=pas du tout, 5=extrêmement), selon leur opinion relativement aux deux communautés. Pour moitié, les adjectifs utilisés étaient à connotation positive (par exemple : les germanophones : ne sont pas du tout/sont extrêmement sympathiques), et pour moitié, les adjectifs utilisés étaient à connotation négative (par exemple : antipathique, repoussant, méprisable). En calculant un score moyen par sujet pour les échelles de même connotation, appliquées à la même cible ethnique, on obtient ainsi pour chaque sujet quatre mesures comprises dans l'intervalle de 0 à 5 :

- l'évaluation positive de l'endogroupe, notée ici ENDOP
- l'évaluation positive de l'exogroupe, notée ici EXOP
- l'évaluation négative de l'endogroupe, notée ici ENDON
- l'évaluation négative de l'exogroupe, notée ici EXON.

Par ailleurs, le questionnaire comportait également des questions permettant d'évaluer deux autres variables, également dans l'intervalle de mesure de 0 à 5 :

– l'intensité de l'identification à l'endogroupe, notée ici IDENT ;
– l'estime négative de soi (*self-directed negative affect*), notée ici SDNA.

A titre d'information (non utilisée dans la suite de l'exercice), on donne les paramètres descriptifs des variables observées :

Description	Notation	Moyenne	Ecart type
Identification à l'endogroupe	IDENT	3.61	0.57
Estime négative de soi	SDNA	3.07	0.63
Evaluation positive de l'endogroupe	ENDOP	4.39	0.71
Evaluation positive de l'exogroupe	EXOP	3.66	0.66
Evaluation négative de l'endogroupe	ENDON	0.56	0.50
Evaluation négative de l'exogroupe	EXON	1.58	0.59

1) On étudie si la prise en compte pour chaque groupe cible d'une évaluation positive et d'une évaluation négative est réellement justifiée.
a) Calculer le coefficient de corrélation des variables ENDOP et ENDON d'une part, celui des variables EXOP et EXON d'autre part. On donne les covariances et les écarts-types non corrigés de variables concernées :
- Cov(ENDOP,ENDON)=−0.1960, s(ENDOP)=0.7050, s(ENDON)=0.4965
- Cov(EXOP, EXON)=−0.2341, s(EXOP)=0.6553, s(EXON)=0.5858.

Réponse. On obtient : $r(ENDOP, ENDON) = \dfrac{-0.1960}{0.7050 \times 0.4965} = -0.56$ *et de même,* $r(EXOP, EXON) = -0.61$.

b) Quels sont les signes de ces coefficients de corrélation ? Comment peut-on interpréter ces signes.

Réponse. Ces coefficients de corrélation sont tous deux négatifs, ce qui n'a rien de surprenant : à un score élevé d'opinions positives correspond un score faible d'opinions négatives et vice-versa ; les variables ENDOP et ENDON semblent corrélées mais varient en sens contraires. Il en est de même pour EXOP et EXON.

c) Ces coefficients de corrélation sont-ils significatifs d'une dépendance entre les variables ? Justifier au seuil de 5%.

Réponse. Ici, l'effectif de l'échantillon est $N = 71$ *et le nombre de degrés de liberté est donc ddl=69. On peut utiliser* $r_{crit} = 0.2319$ *comme valeur critique au seuil de* 5% *bilatéral. Comme dans chacun des deux cas, les coefficients de corrélation observés sont, en valeur absolue, très supérieurs à* r_{crit}*, on conclut à une dépendance entre ENDOP et ENDON et également à une dépendance entre EXOP et EXON.*

d) Peut-on cependant dire que les variables ENDOP et ENDON sont réciproques l'une de l'autre ? Complètement réciproques l'une de l'autre ? Commenter de même le résultat obtenu pour EXOP et EXON.

Réponse. ENDOP et ENDON sont ainsi partiellement réciproques l'une de l'autre. En revanche, elles ne sont pas complètement réciproques, le coefficient trouvé s'écarte notablement de celui ($r = -1$*) qu'on aurait trouvé si les sujets avaient répondu de façon exactement symétrique à chaque question "positive" et à la question "négative" associée (on aurait alors eu : ENDON* $= 5 -$ *ENDOP). Un commentaire en tous points semblable peut être fait pour les variables EXOP et EXON.*

2) On cherche à expliquer les variations de la variable "estime négative de soi" (SDNA) par celles des autres variables.
On définit une variable notée DEROG (*outgroup derogation,* ou partialité envers l'exogroupe) en formant la différence EXON−ENDON.
a) La régression linéaire de la variable SDNA sur la variable IDENT fournit les résultats suivants :

	Synthèse de la Régression ; Variable Dép. : SDNA F(1.69)=4.0587 p<.04784 Err-Type de l'Estim. : .62106					
	Bêta	Err-Type de Bêta	B	Err-Type de B	t(69)	niveau p
OrdOrig.			2.129557	0.472590	4.506145	0.000026
IDENT	0.235700	0.116994	0.260511	0.129309	2.014632	0.047842

Interprétez ces résultats. L'effet de IDENT sur SDNA est-il significatif au seuil de 5% ?

Réponse. L'équation de régression obtenue est : SDNA $= 2.13 + 0.26$ *IDENT et le coefficient de IDENT est significativement différent de 0 (*$t = 2.01, p < 0.05$*). L'effet de IDENT sur SDNA est donc significatif au seuil de* 5%.

b) La régression linéaire de IDENT sur DEROG fournit les résultats suivants :

	Synthèse de la Régression ; Variable Dép. : DEROG F(1.69)=8.4320 p<.00495 Err-Type de l'Estim. : .52667					
	Bêta	Err-Type de Bêta	B	Err-Type de B	t(69)	niveau p
OrdOrig.			-0.129500	0.400765	-0.323132	0.747572
IDENT	0.329994	0.113642	0.318421	0.109657	2.903799	0.004948

Interprétez ces résultats. L'effet de IDENT sur DEROG est-il significatif au seuil de 5% ?

Réponse. L'équation de régression obtenue est : DEROG $= -0.13 + 0.32$ *IDENT et le coefficient de IDENT est significativement différent de 0 (*$t = 2.90, p < 0.01$*). L'effet de IDENT sur DEROG est donc significatif au seuil de* 1%.

c) Enfin, on réalise une régression linéaire multiple de SDNA sur les variables DEROG et IDENT. Les résultats sont alors les suivants :

	Synthèse de la Régression ; Variable Dép. : SDNA F(2.68)=5.1029 p<.00861 Err-Type de l'Estim. : .60028					
	Bêta	Err-Type de Bêta	B	Err-Type de B	t(68)	niveau p
OrdOrig.			2.172575	0.457119	4.752756	0.000011
IDENT	0.140000	0.119789	0.154737	0.132398	1.168723	0.246596
DEROG	0.290005	0.119789	0.332182	0.137210	2.420970	0.018154

Etudier l'effet de chacun des deux prédicteurs sur la variable dépendante SDNA. L'effet de la variable IDENT sur SDNA peut-il être en partie expliqué par l'intermédiaire (la médiation) d'une autre variable ?

Réponse L'équation de régression multiple obtenue est : $SDNA = 2.17 + 0.15\ IDENT + 0.33\ DEROG$.

Le tableau des résultats indique que l'effet de IDENT n'est pas significatif ($t = 1.16, p \geq 0.24$) *alors que celui de DEROG est significatif à* 2%.

Il semble donc que ce soit la variable DEROG, plutôt que la variable IDENT, qui ait un effet sur SDNA. L'effet constaté au a) serait donc essentiellement un effet dû au rôle de médiation joué par la variable DEROG.