Révisions et tests de connaissances

Exercice 1 *Statistiques descriptives* `Données ages`

Dans un groupe de travaux dirigés de licence, on a relevé les dates de naissance de l'ensemble des étudiants :

04/07/74	05/09/75	21/05/74	20/01/76
01/10/76	18/05/75	29/07/76	21/07/77
18/11/76	17/03/76	02/02/77	11/05/65
10/09/77	19/02/76	11/03/76	01/06/77
07/05/77	06/10/74	30/05/77	16/12/73
23/02/77	16/12/75	25/10/74	23/01/73
28/12/76	21/12/75	28/01/72	06/09/75
21/01/77	24/07/61	19/01/76	16/02/74
14/10/75	22/08/64	09/06/75	04/01/77
27/01/74	05/09/63	24/10/76	14/11/74
11/07/76	18/12/77	25/03/75	12/01/73
26/03/72	09/03/74	11/10/69	14/06/60

1) Calculer l'âge en nombre entier d'années de chacun des sujets, au 1/10/1997. Pourquoi s'intéresser à l'âge plutôt qu'à la date de naissance ?

2) Calculer la moyenne, la médiane, le mode, la variance et l'écart-type de l'ensemble des valeurs observées de la variable "âge".

3) On s'intéresse à différents codages de cette variable :
- codage numérique 1 : âge en années *(date de reference : 1/10/1997)*
- codage numérique 2 : âge en mois *(compte au 1/10/97)*
- codage en rangs *(rappel : rang 1 pour le plus jeune, convention habituelle pour les ex- quos : attribution du rang moyen)*
- codage ordinal ; *par exemple, utiliser le codage suivant :*

Age	Code
$19 \leq A < 21$	1
$21 \leq A < 24$	2
$24 \leq A < 27$	3
$27 \leq A < 30$	4
$30 \leq A < 38$	5

- codage binaire : par exemple, coder 0 pour les individus en dessous de la moyenne, 1 pour ceux au dessus.
- codage centré : écarts à la moyenne
- codage centré réduit : écarts réduits.

a) Compléter un tableau du type de celui figurant sur la page suivante.
b) Calculer la moyenne, la variance et l'écart type de la série regroupée en classes (cf. codage ordinal).
4) Construire un histogramme de la distribution en années.
Reponses : 2) $m = 22.90$; $s = 4.28$; $s_c = 4.33$, *Md* $= 21.5$, *Mode* $= 20$.
3-b) $m' = 23.34$; $s'^2 = 16.39$; $s' = 4.05$

Suj.	Date-Nais	Age-1	Age-2	Rang	O d.	Bin.	Cent e	Cent-Red
i42	18/12/77	19	237	1	1	0	-43,9375	-0,86543
i29	21/10/77	19	239	2	1	0	-41,9375	-0,826036
i13	10/09/77	20	240	3	1	0	-40,9375	-0,806339
i08	21/07/77	20	242	4	1	0	-38,9375	-0,766946
i16	01/06/77	20	244	6	1	0	-36,9375	-0,727552
i17	07/05/77	20	244	6	1	0	-36,9375	-0,727552
i19	30/05/77	20	244	6	1	0	-36,9375	-0,727552
i11	02/02/77	20	247	8,5	1	0	-33,9375	-0,668461
i21	23/02/77	20	247	8,5	1	0	-33,9375	-0,668461
i36	04/01/77	20	248	10	1	0	-32,9375	-0,648765
i25	28/12/76	20	249	11	1	0	-31,9375	-0,629068
i09	18/11/76	20	250	12	1	0	-30,9375	-0,609371
i05	01/10/76	20	251	13,5	1	0	-29,9375	-0,589674
i39	24/10/76	20	251	13,5	1	0	-29,9375	-0,589674
i07	29/07/76	21	254	15,5	2	0	-26,9375	-0,530584
i41	11/07/76	21	254	15,5	2	0	-26,9375	-0,530584
i10	17/03/76	21	258	17,5	2	0	-22,9375	-0,451796
i15	11/03/76	21	258	17,5	2	0	-22,9375	-0,451796
i14	19/02/76	21	259	19	2	0	-21,9375	-0,432099
i04	20/01/76	21	260	20,5	2	0	-20,9375	-0,412403
i31	19/01/76	21	260	20,5	2	0	-20,9375	-0,412403
i22	16/12/75	21	261	22,5	2	0	-19,9375	-0,392706
i26	21/12/75	21	261	22,5	2	0	-19,9375	-0,392706
i33	14/10/75	21	263	24	2	0	-17,9375	-0,353312
i02	05/09/75	22	264	25,5	2	0	-16,9375	-0,333615
i28	06/09/75	22	264	25,5	2	0	-16,9375	-0,333615
i35	09/06/75	22	267	27	2	0	-13,9375	-0,274525
i06	18/05/75	22	268	28	2	0	-12,9375	-0,254828
i43	25/03/75	22	270	29	2	0	-10,9375	-0,215434
i40	14/11/74	22	274	30	2	0	-6,9375	-0,136647
i18	06/10/74	22	275	31,5	2	0	-5,9375	-0,11695
i23	25/10/74	22	275	31,5	2	0	-5,9375	-0,11695
i01	04/07/74	23	278	33	2	0	-2,9375	-0,0578595
i03	21/05/74	23	280	34	2	0	-0,9375	-0,0184658
i46	09/03/74	23	282	35	2	1	1,0625	0,0209279
i32	16/02/74	23	283	36	2	1	2,0625	0,0406247
i37	27/01/74	23	284	37	2	1	3,0625	0,0603216
i20	16/12/73	23	285	38	2	1	4,0625	0,0800184
i24	23/01/73	24	296	39,5	3	1	15,0625	0,296684
i44	12/01/73	24	296	39,5	3	1	15,0625	0,296684
i45	26/03/72	25	306	41	3	1	25,0625	0,493652
i27	28/01/72	25	308	42	3	1	27,0625	0,533046
i47	11/10/69	27	335	43	4	1	54,0625	1,06486
i12	11/05/65	32	388	44	5	1	107,063	2,10879
i34	22/08/64	33	397	45	5	1	116,063	2,28606
i38	05/09/63	34	408	46	5	1	127,063	2,50273
i30	27/07/61	36	434	47	5	1	153,063	3,01485
i48	14/06/60	37	447	48	5	1	166,063	3,27091

Exercice 2 *Distributions theoriques - loi binomiale*
Dans le cadre d'une étude sur les familles nombreuses, on veut constituer un échantillon de 100 familles représentatif des familles de 5 enfants.
1) Combien de familles comportant 1 garçon et 4 filles doit-on retenir pour constituer l'échantillon ?
2) Même question pour les familles comportant 0, 2, 3, 4, 5 garçons.
N.B. On fera l'hypothese que la probabilite d'avoir un garcon ou une lle est egale a 0.5
3) Représenter la distribution obtenue à l'aide d'un diagramme en bâtons.
Reponse : Les e ectifs respectifs des familles comportant 0, 1,. . ., 5 garcons sont donnes par : 3, 16, 31, 31, 16, 3

Echantillonnage

Exercice 3
The mean of a large sample is K and σ_K is 2.50([1]). What are the chances that the sample mean misses the true mean by more than : (a) ± 1.00, (b) ± 3.00, (c) ± 10.00.
Reponses : (a) 69%, (b) 23%, (c) moins de 1%.

Exercice 4
Un test comportemental a été étalonné auprès d'une population de 3500 personnes. On sait que la moyenne au test est de 54.24 et que l'écart type est de 6.82. On sait en outre que la variable étudiée est distribuée selon une loi normale.
1) Quelle est la proportion de sujets qui ont une note supérieure à 62 ? Quel est leur nombre ?
2) Quelle est la proportion de sujets qui ont une note comprise entre 45 et 70 ? Quel est leur nombre ?
3) Quelle est la valeur de la variable correspondant au quartile inférieur (Q_1) de la distribution ?
Même question pour le quartile supérieur (Q_3).
4) On tire au hasard, avec remise, un échantillon de 100 sujets de la population étudiée. Soit $\overline{X}$ la variable "moyenne observée sur l'échantillon tiré".
a) Quelle est la loi suivie par cette variable ? Quels sont ses paramètres ?
b) Donner un intervalle rassemblant 95% des valeurs de $\overline{X}$ observées sur de tels échantillons.
Reponses : 1) $P(X > 62) = 0.1291$*, d'ou* $N_1 = 452$.
2) $P(45 \leq X < 70) = 0.9011$ *d'ou* $N_2 = 3153$.
3) $P(X \leq a) = 0.25$ *pour* $a = 49.63$. $P(X \leq a) = 0.75$ *pour* $a = 58.84$.
4) La distribution d'echantillonnage a pour parametres : $\mu = 54.24$ *et* $S = 0.682$*. L'intervalle, centre autour de la moyenne, rassemblant 95% des echantillons est :* $[52.90; 55.58]$.

Exercice 5
Opinion upon an issue seems about equally divided. How large a sample (N) would you need to be sure (at .01 level) that a deviation of 3% in a sample is not accidental (due to chance) ?
Reponse : 1850

[1] σ_K désigne ici l'éca t type de la dist ibution d'échantillonnage

Introduction aux tests statistiques

Exercice 6
In which of the following experimental problems would it be more important to avoid Type I errors of inference than Type II errors in determining the significance of a difference ?
a) Sex differences in reading rate and comprehension in the fifth grade.
b) Effects of a new drug upon reaction time – especially when the drugs are potent and probably dangerous.
c) Comparison of two methods of learning a new skill.
d) Acceptance of a program which involves much time and money and rejection of a less expensive program.
e) Comparative efficiency of a speed-up and a normal rate of work in a factory.
Reponses : a, c, d, e

Exercice 7
"Le problème lié à l'interprétation d'une hypothèse nulle qui n'a pas été rejetée tourmente depuis plus de 50 ans les étudiants inscrits à un cours de statistiques. Tous les statisticiens s'accordent cependant sur un point : on ne peut jamais prétendre avoir "prouvé" l'hypothèse nulle.
Pour Fisher, un résultat non significatif n'est pas un résultat concluant. D'après lui, on a le choix entre rejeter une hypothèse nulle et suspendre son jugement.
Neymann et Pearson adoptent une approche légèrement différente. Soit on rejette l'hypothèse nulle, soit on l'*accepte*. Toutefois, quand on dit que l'on accepte une hypothèse nulle, cela ne veut pas dire que l'on estime en avoir prouvé l'exactitude. Cela veut dire que l'on fera *comme si* elle était vraie, au moins jusqu'à ce que l'on obtienne des données plus adéquates."

Quels commentaires ces deux opinions suggèrent-elles ? Donner deux exemples de situations dans lesquelles on sera conduit à adopter l'une ou l'autre des deux démarches.

Reponse : Une hypothese nulle peut ne pas être rejetee a cause d'une trop faible taille d'echantillon, ou de l'existence de variables non contrôlees qui ont arti ciellement augmente la dispersion. Il y a des cas ou l'on est conduit a accepter l'hypothese nulle, jusqu'a ce que des informations contraires viennent la contredire. Exemple : absence d'e et indesirable d'un medicament.

Inférence sur une moyenne. Comparaison à une norme

Exercice 8
Dans ce qui suit, on se donne une population de référence dans laquelle la distribution du QI est une loi normale de paramètres $\mu = 100$ et $\sigma = 15$.
a) Dans une classe de rattrapage, on examine un groupe de 5 élèves et on trouve un QI moyen $\overline{QI} = 89.2$. Peut-on dire que le groupe de 5 élèves est, pour le QI, atypique de la population de référence aux seuils traditionnels ?
b) La classe de rattrapage compte 25 élèves ; on examine tous les élèves et on trouve un QI moyen $\overline{QI} = 89.2$. Mêmes questions que précédemment.
Reponses : a) $z_{obs} = -1.61$. *On ne peut pas retenir l'hypothese d'un groupe atypique au seuil de 5% unilateral. b)* $z_{obs} = -3.6$. *Groupe atypique au seuil de 1% unilateral*

Exercice 9

Compas et ses collègues (étude non publiée) ont constaté avec surprise que les jeunes enfants soumis au stress présentent en fait moins de symptômes d'angoisse et de dépression que ce à quoi l'on pourrait s'attendre. Toutefois, ils ont aussi remarqué que les scores obtenus par ces enfants sur une échelle de désirabilité sociale sont étonnament élevés. On sait que la moyenne de population de l'échelle de désirabilité sociale est égale à 3.87. Pour un échantillon de 36 enfants soumis au stress, Compas ***et al.*** ont relevé une moyenne d'échantillon de 4.39, avec un écart type de 2.61.

1) Par quel test pourrait-on savoir si ce groupe présente une tendance accrue à donner des réponses socialement acceptables ?

2) Quelles seraient l'hypothèse nulle et l'hypothèse alternative ?

3) Mettre en œuvre le test choisi et conclure.

Reponse : Il s'agit ici de comparer la moyenne observee sur l'echantillon a la valeur 3.87, consideree comme une norme. Compte tenu de la taille de l'echantillon, on peut utiliser la statistique :

$$Z = \frac{\overline{x} - 3.87}{E} \quad \textit{avec} \quad E^2 = \frac{2.61^2}{36}$$

qui suit une loi normale centree reduite.

On obtient alors $Z_{obs} = 1.20$, ce qui n'est pas signi catif d'une di erence.

Exercice 10

Dans une grande entreprise (américaine), le salaire moyen des hommes possédant entre 3 et 5 ans d'expérience est de 28 000 \$. Les salaires (en milliers de dollars) d'un échantillon aléatoire de 10 femmes possédant entre 3 et 5 ans d'expérience sont les suivants :

24 27 31 21 19 26 30 22 15 36

1) Au niveau descriptif, quelle estimation peut-on faire du salaire moyen féminin ?

2) Y a-t-il des preuves attestant de niveaux de salaires différents pour les hommes et les femmes ?

Reponses : 1) La moyenne des salaires sur l'echantillon est $\overline{x} = 25.1$. On peut donc estimer le salaire feminin moyen a 25 100 \$.

2) On a par ailleurs : $s_c = 6.23$. La statistique de test (comparaison d'une moyenne a une norme sur un petit echantillon) vaut : $t = \frac{25.1-28}{E}$ avec $E^2 = \frac{6.23^2}{10}$, d'ou $t_{obs} = -1.47$. Cette statistique suit une loi de Student a 9 ddl, et donc, pour un test unilateral au seuil de 5%, $t_{crit} = 1.83$. On conclut donc sur H_0 : on n'a pas mis en evidence de di erence entre les hommes et les femmes.

Exercice 11

1) En 2003-2004, la population des étudiants inscrits dans le secteur Lettres à l'UBO était constituée de 3635 femmes et 1338 hommes.

L'échantillon constitué par les étudiants inscrits au centre littéraire de Quimper comprenait 405 femmes pour 174 hommes.

Du point de vue de la variable étudiée, peut-on considérer cet échantillon comme tiré au hasard dans la population des étudiants littéraires de l'UBO (seuil utilisé : 1%) ?

2) De même, la population des étudiants inscrits dans le secteur Sciences était constituée de 1361 femmes pour 2232 hommes.

a) L'échantillon constitué par les étudiants inscrits à l'IUP Génie Mécanique et Productique comprenait 11 femmes pour 138 hommes.

Du point de vue de la variable étudiée, peut-on considérer cet échantillon comme tiré au hasard dans la population des étudiants scientifiques de l'UBO (seuil utilisé : 1%) ?
b) Même question pour les étudiants de l'IUP Génie des Systèmes Indistriels (69 femmes pour 25 hommes).
Reponses : 1) On fait ici un test de comparaison d'une proportion a une norme. La proportion de femmes dans la population est $p_0 = 0.7309$*, tandis que la proportion observee sur l'echantillon est* $f = 0.6995$*, avec une taille d'echantillon* $n = 579$*. On obtient* $Z_{obs} = -1.70$*. On retient donc* H_0*, ce qui revient a repondre positivement a la question posee.*
2) a) On a ici : $p_0 = 0.3788$ *et* $f = 0.0738$ *et donc* $Z_{obs} = -7.67$*. Du point de vue de la repartition par sexe, l'echantillon forme par les etudiants de l'IUP GMP est donc tout a fait atypique de la population des etudiants scienti ques.*
b) On a : $p_0 = 0.3788$ *et* $f = 0.7340$ *d'ou* $Z_{obs} = 7.10$ *et une conclusion similaire a la precedente.*

Tests d'égalité de moyennes, de fréquences

Exercice 12 *Dossier \Plpc", \Plpc-cor"*
Dans une recherche de psychologie, on a présenté à chaque sujet des photographies de chiens, de diverses races, pour moitié à poils courts et pour moitié à poils longs. Chaque sujet évalue le chien présenté en cochant l'une des 6 cases d'une échelle d'attirance. Les évaluations ont été codées numériquement de 0 (faible attirance) à 5 (forte attirance). Dans le tableau ci-dessous, on donne pour chacun des 16 sujets les deux valeurs moyennes des évaluations concernant les deux types de chiens.

	s1	s2	s3	s4	s5	s6	s7	s8
Poils longs	3,50	2,75	3,50	2,75	1,50	3,00	4,00	3,00
Poils courts	2,50	1,25	2,75	0,50	2,00	3,00	2,00	2,00

	s9	s10	s11	s12	s13	s14	s15	s16
Poils longs	2,75	3,25	4,00	3,00	3,75	4,00	4,00	3,25
Poils courts	2,50	0,75	0,50	1,00	1,50	1,75	2,00	2,00

Un *e et individuel* est défini comme la différence entre les deux notes concernant un même sujet. Construire le protocole des effets individuels.
Peut-on affirmer, au seuil de 5%, que les chiens à poils longs sont préférés à ceux à poils courts.
Remarque. La variable statistique de départ (classement sur une échelle d'attirance) est en fait ordinale et non numérique. Nous reprendrons cet exemple avec d'autres outils dans un chapitre ultérieur.
Reponses : Le protocole des e ets individuels a pour moyenne $\overline{d} = 1.5$ *et pour ecart type corrige* $s_c = 1.0448$*. On obtient* $t_{obs} = 5,74$*. La di erence est signi cative au seuil de 5%.*

Exercice 13 *Dossier \Internat"*
Pour une recherche internationale sur l'évaluation des savoir-faire en situation de vie quotidienne, on a construit un test comportant une version en français et une version en anglais. Pour s'assurer de l'équivalence des deux instruments, on administre les deux versions du test à un groupe de 24 paires de sujets, chaque paire comprenant un sujet de langue maternelle anglaise et un sujet de langue maternelle française. Pour constituer les

paires, les sujets sont appariés sur un ensemble de variables permettant de s'assurer de l'équivalence de leur efficience. Le plan du protocole est P24*T2.
On cherche à répondre à la question "les sujets obtiennent-ils le même résultat au test en anglais (t_1) et au test en français (t_2)".

Sujet	t_1	t_2	Sujet	t_1	t_2	Sujet	t_1	t_2
S1	27.0	25.5	S9	27.0	28.5	S17	21.5	20.5
S2	23.0	23.5	S10	28.5	26.0	S18	19.5	23.5
S3	21.5	18.5	S11	26.5	23.0	S19	23.5	23.0
S4	24.0	22.0	S12	24.0	24.5	S20	21.0	24.0
S5	22.5	28.0	S13	26.0	23.5	S21	23.0	24.0
S6	24.0	19.0	S14	21.5	22.0	S22	20.0	20.5
S7	21.5	19.0	S15	26.5	22.0	S23	20.0	17.5
S8	22.0	25.5	S16	20.0	16.0	S24	21.0	16.0

Reponses : On introduit le protocole des di erences individuelles D *de ni par :* $d_i = t_{i2} - t_{i1}$*. On a :* $\overline{d} = -0.8125$*,* $s_c = 2.9185$*.* $t_{obs} = -1.36386$*. Ici,* $ddl = 23$ *et donc, au seuil de 5% bilateral,* $t_{crit} = 2.07$*. On ne peut donc pas rejeter l'hypothese nulle*

Exercice 14 *Dossier \Manage"*
Dans une étude de psychologie du travail, on cherche à déterminer l'influence de deux conceptions managériales, l'une "hiérarchique" (H), l'autre "démocratique" (D) sur la sécurité dans l'entreprise. A cet effet, on prélève au hasard, dans un vaste ensemble de données, quinze intervalles trimestriels pour lesquels on relève ensuite le nombre d'incidents qui se sont produits sur chacun des sites, par ailleurs comparables à tous égards. Les données résultant des tirages aléatoires sont les suivantes :

Trim.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Conc. H	11	11	14	21	12	10	15	15	17	9	12	12	15	12	11
Conc. D	8	13	12	17	14	9	10	12	13	10	8	13	12	9	9

1) Calculer la moyenne et l'écart type de chacun des deux échantillons proposés. A partir de ces éléments, formuler une conclusion descriptive concernant le problème posé.
2) Les données observées définissent-elles des échantillons indépendants ou appareillés ?
3) Réaliser un test statistique approprié en précisant avec soin l'hypothèse nulle et l'hypothèse alternative, la statistique de test utilisée, la loi suivie par cette statistique et le seuil choisi.
Reponses : 1) $\overline{x}_1 = 13.33$*,* $\overline{x}_2 = 11.27$*,* $s_{1c} = 3.09$*,* $s_{2c} = 2.55$*. 3)* $\overline{d} = 1.87$*.* $s_c = 2.33$*.* $t_{obs} = 3.10$*. La di erence est signi cative au seuil de 1% (test bilateral).*

Exercice 15 *Dossier \Usine"*
Dans une usine, on cherche à voir si un changement de l'environnement (musique dans les ateliers en particulier) peut modifier le rendement. Ce dernier est mesuré ici par le nombre moyen de pièces produites à l'heure par chaque ouvrier.
1) On note pour chacun des 33 ouvriers observés, son rendement avant et après l'introduction de ces changements. Les résultats sont les suivants.

Avant	Après	Avant	Après	Avant	Après
45	48	30	35	40	45
36	40	45	50	40	35
47	53	30	40	38	35
40	40	45	50	35	40
45	46	40	45	40	45
35	30	50	50	35	37
36	40	40	40	38	35
50	60	50	45	50	50
50	60	40	35	45	50
40	40	55	50	30	33
40	40	30	35	38	38

Effectuer les calculs permettant de répondre à la question posée et conclure.
2) On suppose maintenant que les données ci-dessus concernent deux groupes d'ouvriers distincts, les premiers (colonne "Avant") placés dans un environnement ordinaire, les seconds (colonne "Après") placés dans un environnement modifié. Quels sont alors les calculs à effectuer et quelle est alors la conclusion ?
Reponses : 1) $\overline{d} = 2.03$, $s_c = 4.41$. $z_{obs} = 2.64$. *La di erence est signi cative pour un test unilateral au seuil de 1%.*
2) $\overline{x}_1 = 40.85$, $\overline{x}_2 = 42.88$, $s_{1c} = 6.60$, $s_{2c} = 7.52$, $z_{obs} = 1.17$. *Dans le cas d'un test unilateral au seuil de 5%, on ne peut refuser* H_0.

Exercice 16 *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) Ces données expérimentales permettent-elles d'affirmer que la pédagogie a un effet sur les résultats à l'épreuve de combinatoire ?
Reponse : $t_{obs} = -1.45$*. Dans le cas d'un test unilateral au seuil de 5%, on ne peut refuser* H_0*.*

Exercice 17
On veut savoir si une maladie M modifie le taux de certaines protéines dans le sang. On a mesuré leur concentration dans un échantillon de sujets atteints par M et dans un autre échantillon formé de sujets en bonne santé (sujets témoins). Les résultats (dans une unité convenable) sont les suivants :

	Effectifs	Moyenne échantillon	Variance échantillon
Malades	77	141	40
Témoins	33	131	32

Tester l'hypothèse "taux identiques chez les malades et les témoins" contre l'hypothèse :
a) "taux différent chez les malades et les témoins",
b) "taux supérieur chez les malades".
Reponses : $z_{obs} = 8.094$*. Dans les deux cas,* H_0 *est refusee.*

Exercice 18
Dans une étude publiée en 1996, les auteurs se sont intéressés au rôle de la valeur affective d'un texte dans la récupération du souvenir chez les personnes âgées. L'expérience a porté sur 20 sujets ; 10 d'entre eux présentent un déficit mnésique, les 10 autres n'en présentent pas.
Dans un premier temps, on veut s'assurer que les deux groupes ainsi constitués :
- d'une part, sont homogènes du point de vue de l'âge
- d'autre part obtiennent des scores significativement différents au test "Wescher Mémoire".
Les données relatives aux sujets des deux groupes figurent dans les tableaux ci-dessous.

Sujets déficitaires		
	Age	Wescher Mémoire
Sujet 1	80	66
Sujet 2	91	59
Sujet 3	82	84
Sujet 4	87	68
Sujet 5	82	80
Sujet 6	85	75
Sujet 7	84	72
Sujet 8	85	82
Sujet 9	88	78
Sujet 10	87	76

Sujets non déficitaires		
	Age	Wescher Mémoire
Sujet 11	80	113
Sujet 12	81	94
Sujet 13	82	87
Sujet 14	84	98
Sujet 15	85	103
Sujet 16	85	110
Sujet 17	86	97
Sujet 18	89	119
Sujet 19	91	88
Sujet 20	92	91

1) Calculer la moyenne et l'écart type des variables âge et Weschler mémoire pour chacun des deux groupes.
2) Comparer les deux groupes du point de vue de l'âge à l'aide d'un test de comparaison de moyennes.
3) De même, comparer les deux groupes du point de vue de la seconde variable.
Reponses : 1) Pour la variable âge : $\overline{x_1} = 85.1$, $\overline{x_2} = 85.5$, $s_1 = 3.11$, $s_2 = 3.88$. *Pour la seconde variable,* $\overline{x'_1} = 74$, $\overline{x'_2} = 100$, $s'_1 = 7.42$, $s'_2 = 10.40$.
2) On realise un test de comparaison de moyennes sur deux groupes independants. Compte tenu de la taille des echantillons, c'est la loi de Student qui s'applique. Ici, $n_1 = n_2 = 10$ *(groupes equilibres) ;* $ddl = 18$. *Au seuil de 5% unilateral,* $t_{crit} = 1.7341$. *On obtient ici* $t_{obs} = 0.24$. *On retient donc* H_0 *et on conclut a l'homogene•te des âges dans les deux groupes.*
3) Test analogue au precedent. On obtient ici $t_{obs} = 6.10$*, ce qui signi catif d'une di erence aux seuils traditionnels.*

Exercice 19

Le modèle de la mémorisation proposé par Craik et Lockhart (1972) stipule que le degré auquel un sujet se rappelle un matériel verbal est fonction du degré auquel ce matériel a été traité lors de sa présentation initiale. Eysenck (1974) voulait tester ce modèle et examiner s'il pouvait contribuer à expliquer certaines différences relevées entre des sujets jeunes et âgés concernant leur aptitude à se rappeler du matériel verbal. L'étude qu'il a menée incluait 50 sujets dont l'âge se situait entre 18 et 30 ans et 50 sujets compris dans la tranche d'âge 55–65 ans. Dans chacune des tranches d'âge, Eysenck a réparti les 50 sujets dans cinq groupes. Le premier devait lire une liste de mots et se contenter de compter le nombre de lettres de chacun d'eux. Le deuxième groupe devait lire chaque mot et lui trouver une rime. Le troisième groupe devait donner un adjectif qui aurait pu être utilisé pour modifier chaque mot de la liste. Le quatrième devait essayer de se former une image précise de chaque mot. Aucun de ces quatre groupes ne savait qu'il faudrait se rappeler les mots ultérieurement. Enfin, le cinquième groupe, ou groupe d'apprentissage intentionnel, devait lire la liste et mémoriser tous les mots. Après avoir passé trois fois en revue la liste de 27 mots, les sujets devaient retranscrire tous les mots dont ils se souvenaient. Le nombre de mots rappelés par chacun des 100 sujets est indiqué par le tableau 1 page 11.
1) On veut étudier chez les sujets âgés s'il existe une différence de performance entre le groupe 2 (traitement syntaxique) et le groupe 3 (traitement sémantique), en faveur de ce dernier. Le calcul permet d'obtenir les résultats de statistiques descriptives suivants :

	Groupe 2	Groupe 3
Moyenne	6.9	11.0
Effectif	10	10
Ecart-type	2.02	2.37
Ecart-type corrigé	2.13	2.49

2) Etudier de même s'il existe une différence de performance due à l'âge parmi les sujets du groupe 2.
Reponses : 1) Il s'agit d'une comparaison de moyennes sur deux groupes independants. Avec les donnees fournies, on obtient $t_{obs} = -3.80$ *alors que, pour un test bilateral au seuil de 5%, on obtient* $t_{crit} = 2.10$ ($ddl = 18$)*. Il existe donc une di erence signi cative entre le traitement syntaxique et le traitement semantique.*

	Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5
Sujets âgés	9	7	11	12	10
	8	9	13	11	19
	6	6	8	16	14
	8	6	6	11	5
	10	6	14	9	10
	4	11	11	23	11
	6	6	13	12	14
	5	3	13	10	15
	7	8	10	19	11
	7	7	11	11	11
Sujets jeunes	8	10	14	20	21
	6	7	11	16	19
	4	8	18	16	17
	6	10	14	15	15
	7	4	13	18	22
	6	7	22	16	16
	5	10	17	20	22
	7	6	16	22	22
	9	7	12	14	18
	7	7	11	19	21

TAB. 1 – Données Eysenck

2) Les sujets du groupe 2 forment deux sous-groupes independants du point de vue de l'âge. On obtient $t_{obs} = -0.766$. *La di erence de performance n'est pas signi cative dans ce cas.*

Exercice 20

Deux groupes de séropositifs asymptomatiques ont été constitués sur la base d'une affectation au hasard des patients aux traitements. Le premier groupe, traité à l'aide d'un placebo, comprenait 428 personnes ; le second comprenait 453 personnes traitées par 500 mg/jour d'AZT. Après cinquante-cinq semaines de traitement, 33 personnes sous placebo sont tombées malades, contre seulement 11 dans le groupe AZT. ***D'apres*** Le Monde ***du 18 avril 1990***.

L'analyse statistique visera à répondre à la question : la prise de 500 mg/jour d'AZT retarde-t-elle l'apparition du sida ?

1) Calculer, pour chacun des deux groupes, la fréquence d'apparition de la maladie. Formuler une conclusion descriptive.

2) Comparer les fréquences d'apparition de la maladie dans les deux groupes à l'aide d'un test de comparaison de proportions au seuil de 1% unilatéral.

3) a) On considère les deux variables statistiques *groupe* (modalités : *Placebo* et *AZT*) et *apparition de la maladie* (modalités : *oui* et *non*). Dresser un tableau de contingence résumant les observations précédentes.

b) Comparer les groupes à l'aide d'un test du χ^2, au seuil de 1%.

Reponses :

1) Les frequences sont respectivement de $f_1 = \frac{33}{428} = 7.7\%$ *et* $f_2 = \frac{11}{453} = 2.4\%$.
2) On obtient $z_{obs} = 3.60$. *Les frequences sont donc signi cativement di erentes au seuil de 5% unilateral (*$z_{crit} = 2.33$*).*
2) Le tableau de contingence, le tableau des e ectifs theoriques et celui des contributions au χ^2 *sont donnes par :*

	Placebo	*AZT*	*Total*
oui	*33*	*11*	*44*
non	*395*	*442*	*837*
Total	*428*	*453*	*881*

	Plac.	*AZT*
oui	*21.4*	*22.6*
non	*406.6*	*430.4*

	Plac.	*AZT*
oui	*6.28*	*5.95*
non	*0.33*	*0.31*

Ainsi, $\chi^2_{obs} = 12.89$. *On peut a rmer, au seuil de 1%, l'existence d'une liaison entre le type de traitement et l'apparition de la maladie.*

Exercice 21
Dans une étude devenue classique (1939), des chercheurs ont montré à des enfants noirs une poupée noire et une poupée blanche en leur demandant de choisir celle avec laquelle ils voudraient jouer. Sur 252 enfants, 169 ont choisi la poupée blanche tandis que 83 préféraient la poupée noire.
Un deuxième groupe de recherche a reproduit l'expérience précédente en 1970. Les études n'étaient pas exactement identiques, mais les résultats se sont avérés intéressants : sur 89 enfants noirs, 28 ont choisi la poupée blanche et 61 ont préféré la poupée noire.
Pour les deux études, un test adéquat montre que les enfants ne choisissent pas la poupée au hasard.
Une troisième équipe de chercheurs réunit les résultats précédents dans un tableau de contingence et procède à un test du χ^2 sur ce tableau :

	Exp. 1	Exp. 2
Poupée noire	83	61
Poupée blanche	169	28

Formuler avec précision l'hypothèse nulle et l'hypothèse alternative correspondant à ce test. Réaliser le test et conclure.
Reponses : Les variables mises en jeu dans le tableau de contingence propose sont d'une part la couleur de la poupee (noire ou blanche), d'autre part le numero d'ordre de l'experience (experience 1 ou experience 2). L'hypothese correspondant au test du χ^2 *est donc ici : le choix pour l'une ou l'autre des poupees depend-il de l'experience consideree ? ou encore : les deux experiences fournissent-elles des resultats identiques ? On trouve :* $\chi^2_{obs} = 34.15$; $ddl = 1$ *; pour* $\alpha = 5\%, \chi^2_{crit} = 3.84$ *et donc les deux experiences fournissent des resultats di erents.*

Exercice 22 Un débat télévisé est organisé entre deux candidats à une élection. Un sondage fait auprès d'un échantillon de 200 électeurs a eu lieu avant le débat ; 95 électeurs déclaraient alors vouloir voter pour le candidat A.
Un autre sondage est effectué après le débat. Sur 150 électeurs interrogés, 84 déclarent alors vouloir voter pour le candidat A.
La proportion d'intentions de vote pour A a-t-elle été influencée par le débat ? Répondre à cette question à l'aide d'un test de comparaison de proportions au seuil de 5%.
Reponse. Il s'agit ici d'une comparaison de frequences sur deux groupes independants. On trouve : $f_1 = 0.475$, $f_2 = 0.56$, $p = 0.5114$, *d'ou* $E = 0.054$ *et* $z_{obs} = -1.57$. *Au seuil*

de 5% pour un test unilateral, on a : $z_{crit} = -1.645$*. On n'a donc pas mis en evidence d'in uence du debat sur les intentions de vote pour A.*

Exercice 23 χ^2 de MacNemar
Soit deux questions d'un test, A et B, auxquelles un groupe de 184 sujets a répondu. On voudrait savoir si la fréquence des réussites et des échecs est la même pour A et B. On a :

A : 99 échecs et 85 réussites
B : 64 échecs et 120 réussites.

On sait que 62 sujets ont répondu avec succès aux deux questions.
Justifier le tableau d'effectifs observés suivant :

	B : réussite	B : échec	Total
A : réussite	62	23	85
A : échec	58	41	99
Total	120	64	184

Pour répondre à cette question, le statisticien forme le tableau d'effectifs attendus suivant :

	B : réussite	B : échec
A : réussite	62	40.5
A : échec	40.5	41

Justifier la construction de ce tableau et comparer les deux distributions à l'aide de la distance du χ^2 et d'un test (le nombre de degrés de liberté est ici $ddl = 1$. Pourquoi ?).
Vérifier que le χ^2 de MacNemar peut aussi être obtenu par la formule :

$$\chi^2 = \frac{(|\text{effectif d'une case} - \text{effectif de l'autre case}| - 1)^2}{\text{effectif d'une case} + \text{effectif de l'autre case}}$$

Reponses : Sans correction de Yates, on obtient : $\chi^2_{obs} = 15.12$*. La di erence est donc signi cative au seuil de 1%.*
Avec la correction de Yates, on obtient : $\chi^2_{obs} = 14.27$*. En utilisant la formule indiquee ci-dessus, on obtient de même :* $\chi^2_{obs} = 14.27$*.*

Exercice 24
Lors d'une enquête par questionnaire relative au traitement de la violence dans les établissements scolaires, un groupe de parents est interrogé avant et après la projection d'un film illustrant ce problème. Au total, 45 personnes ont fourni des réponses. On remarque que, avant la projection, 29 personnes pensaient que ces difficultés devaient être traitées en rendant plus sévère le règlement de l'établissement alors que les autres préconisaient des aides individualisées aux élèves concernés. Après la projection, on trouve 18 personnes qui préfèrent une solution de type règlementaire dont 14 avaient déjà cette opinion avant le film.
1) Dresser le tableau de contigence obtenu en croisant la condition (avant ou après la projection) et le type de solution préconisé.
2) L'opinion des sujets est-elle influencée par la projection du film ? Répondre à cette question à l'aide d'un test de comparaison de proportions appariées, puis à l'aide d'un test du χ^2 de Mac Nemar, au seuil de 5% (on suppose l'échantillon de taille suffisante pour pouvoir appliquer les méthodes envisagées).
Reponses. 1)

		Apres Projection		
		Reglement	Aide	Total
Avant projection	Reglement	14	15	29
	Aide	4	12	16
	Total	18	27	45

Methode des proportions appariees : $Z = \frac{15-4}{\sqrt{15+4}} = 2.52$. *Pour un test bilateral au seuil de 5%,* $z_{crit} = 1.96$. *L'opinion des sujets est donc in uencee par la projection du lm.*
Methode du χ^2 *de Mac Nemar (avec correction de Yates) :* $\chi^2_{obs} = \frac{(15-4-1)^2}{15+4} = 5.26$, $ddl = 1$, $\chi^2_{crit} = 3.84$, *et la conclusion reste la même.*

Tests non paramétriques

Exercice 25 *Test du* χ^2 *sur un tableau de contingence.*
Lors d'une enquête, on a interrogé 150 personnes prises au hasard sur leurs connaissances en langues étrangères. Les résultats obtenus sont les suivants :

	Hommes	Femmes
Anglais	37	24
Allemand	9	19
Espagnol	16	15
Aucune	20	10

Les connaissances en langues étrangères dépendent-elles du sexe dans la population dont est issu l'échantillon étudié ? On répondra à cette question en effectuant un test au seuil de 5%.
Reponses numeriques : $\chi^2_{obs} = 8.48$, *ddl* $= 3$, $\chi^2_c = 7.815$. La différence entre les deux sexes est donc significative.

Exercice 26 *Dossier \Psymot"*
On dispose des résultats de deux groupes à une épreuve de psycho-motricité, l'un des groupes ayant reçu une éducation particulière.

G1 : 38 34 18 45 5 1 21 22 11 7 52 24 93 17 56

G2 : 79 96 76 89 53 39 78 81 99 77 29 27.

1) a) Etudier, à l'aide d'un test de la médiane au seuil de 5%, si l'éducation a une influence sur les scores des sujets.
b) Que devient le résultat du test si l'on choisit un seuil de 1% ?
2) Reprendre la question précédente à l'aide d'un test de Wilcoxon Mann Whitney au seuil de 1% unilatéral.
Reponses : 1) La mediane de l'ensemble des observations est $M = 39$. *Le tableau de contingence obtenu en croisant \position par rapport a la mediane" et \groupe" est le suivant :*

	G1	G2	Tot.
$\leq M$	11	3	14
$> M$	4	9	13
Tot.	15	12	27

On obtient : $\chi^2_{obs} = 6.24$*. Ce resultat est signi catif d'une di erence entre les groupes au seuil de 5%, mais n'est pas signi catif au seuil de 1%.*
2) $\overline{R_1} = 9.6$*,* $\overline{R_2} = 19.5$ *et* $z_{obs} = -3.22$*. Au seuil de 1%, on refuse l'hypothese* H_0*. C'est ici le groupe G2 qui a l'e ectif le plus faible. Avec les notations de la table du test de Wilcoxon Mann Whitney, on a :* $W = 234$ *et, au seuil de 1 %,* $W'_s = 216$*. On refuse donc* H_0*.*

Exercice 27
Deux enseignants se sont partagé un lot de 20 copies. L'enseignant A a corrigé $n_1 = 8$ copies, l'enseignant B a corrigé $n_2 = 12$ copies, d'où deux groupes de copies. On range les 20 copies par ordre de mérite. Les résultats sont les suivants :

Correcteur A : 2 4 5 7 9 10 12 17
Correcteur B : 1 3 6 8 11 13 14 15 16 18 19 20.

1) Calculer la moyenne des rangs des copies notées par chacun des correcteurs. Quelle conclusion peut-on formuler au niveau descriptif ?
2) A l'aide d'un test de Wilcoxon-Mann-Whitney, étudier si l'on peut conclure à l'hétérogénéïté des deux groupes aux seuils traditionnels.
Reponses : 1) Noter que l'enonce donne le protocole des rangs. On obtient : $\overline{R_1} = 8.25$*,* $\overline{R_2} = 12$ *d'ou, au niveau descriptif, une di erence au bene ce du correcteur B.*
2) Avec les notations de la table du test de Wilcoxon Mann Whitney, on a : $N_1 = 8$*,* $N_2 = 12$*,* $W = 66$ *et, au seuil de 5 %,* $W_s = 62$*,* $W'_s = 106$*. On accepte donc* H_0*.*

Exercice 28
On veut comparer les résultats de deux classes de même niveau scolaire en utilisant une épreuve commune de performances linguistiques. On obtient, selon le barème adopté, les résultats suivants :

Classe A : 41, 49, 14, 81, 18, 54, 95, 13, 38, 15, 5, 17, 25, 96, 8.
Classe B : 80, 96, 86, 99, 89, 32, 7, 49, 98, 12, 67, 46, 26, 18, 80.

1) Effectuer un test de la médiane afin de déterminer si les deux classes obiennent des résultats comparables (seuil 5%).
2) Reprendre le problème à l'aide d'un test de Wilcoxon Mann Whitney.
Reponses : 1) $\chi^2_{obs} = 3.33$*, au seuil de 5%,* $\chi^2_{crit} = 3.84$*. On accepte donc* H_0*.*
2) On a ici : $\sum R_A = 192.5$*,* $\sum R_B = 272.5$*,* $N_1 = N_2 = 15$ *d'ou* $Z_{obs} = 1.6591$*. Pour un seuil de 5% unilateral, on a :* $Z_{crit} = 1.645$*. On accepte donc l'hypothese* H_1 *d'une di erence entre les deux groupes, au bene ce de la classe B. Mais, la di erence entre* Z_{obs} *et* Z_{crit} *est ici tres faible, et il serait sans doute utile de faire une correction tenant compte des ex-aequos.*

Exercice 29
On reprend l'énoncé 12 (chiens à poils longs, chiens à poils courts).
1) On recode les effets en ne retenant que leur signe (+ ou −). A partir du protocole obtenu, procéder au test du signe (N.B. Compte-tenu de la faible taille de l'échantillon, il faut utiliser ici une loi binomiale de paramètres $p = 0.5$ et $n = 15$).
2) Construire le protocole des rangs appliqué au protocole des effets individuels (différence des notes attribuées par un même sujet). L'observation correspondant à un effet nul ne sera pas prise en compte.
3) Procéder au test d'homogénéïté des rangs de Wilcoxon.
Reponses : 1) $D_+ = 14$*,* $D_- = 1$*,* $N = 15$*. Pour une variable* X *distribuee selon une loi*

binomiale de parametres $p = 0.5$ *et* $N = 15$*, on a :* $P(X \leq 1) = 0.00048$*. Sous l'hypothese* H_0*, on a moins de 5 chances sur 10000 d'observer un echantillon aussi extrême. On accepte donc* H_1*.*
Test de Wilcoxon : $N = 15$*,* $T_+ = 118$*,* $T_- = 2$*. D'apres la table, on a, au seuil de 1%,* $T_{m,crit} = 19$*. On accepte donc* H_1*.*

Exercice 30

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. L'étude a porté sur trois 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 ;
- Un groupe d'enfants dont le poids à la naissance était normal.

Le tableau 2 page 17 représente une partie des données observées. Il indique l'indice de développement mental (IDM) à 6 mois et à 24 mois pour le groupe témoin PRN, ainsi que la différence entre les deux indices.

1) Des travaux antérieurs sur de tels enfants laissent penser que les scores IDM des enfants PRN du groupe témoin pouvaient diminuer de façon sensible entre 6 et 24 mois. Réaliser un test de comparaison de moyennes afin de déterminer si les données observées confirment cette hypothèse (choisir un seuil de 5%).

2) Peu satisfait du résultat obtenu à la question 1, un chercheur souhaite comparer les données relatives aux enfants du groupe témoin PRN à l'aide d'un test du signe. Réaliser le test et conclure.

PRN Témoin			
	IDM-6	IDM-24	Diff
s1	124	114	-10
s2	94	88	-6
s3	115	102	-13
s4	110	127	17
s5	116	104	-12
s6	139	104	-35
s7	116	91	-25
s8	110	96	-14
s9	129	104	-25
s10	120	106	-14
s11	105	91	-14
s12	88	102	14
s13	120	104	-16
s14	120	100	-20
s15	116	114	-2
s16	105	109	4
s17	100	109	9
s18	91	119	28
s19	129	91	-38
s20	84	81	-3
s21	91	114	23
s22	116	119	3
s23	100	102	2
s24	113	111	-2
s25	89	80	-9
s26	102	119	17
s27	110	119	9
s28	116	123	7
s29	124	119	-5
s30	126	114	-12
s31	123	132	9
Moyenne	111.00	106.71	-4.29
Var. corrigée	191.8667	167.8129	257.2129

TAB. 2 – Données PRN-Témoin

Reponses : 1) Soit μ_6 et μ_{24} les moyennes de l'IDM a 6 mois et a 24 mois dans la population dont est issu l'echantillon. L'hypothese de recherche se traduit par les hypotheses statistiques suivantes :
H_0 : $\mu_6 = \mu_{24}$
H_1 : $\mu_6 > \mu_{24}$
Il s'agit d'un test de comparaison de moyennes sur des groupes appareilles. La statistique de test est $t = \frac{\overline{d}}{E}$ avec $E^2 = \frac{s_c^2}{n}$. Compte tenu de l'e ectif ($n = 31$), on peut utiliser ici une loi de Student avec 30 ddl ou une loi normale. Pour un seuil de 5% (et un test unilateral avec la zone de rejet \a gauche"), on obtient $t_{crit} = -1.70$ (ou $z_{crit} = -1.65$). On obtient : $E^2 = \frac{257.21}{31} = 8.30$; $E = \sqrt{8.30} = 2.88$; $t_{obs} = \frac{-4.29}{2.88} = -1.49$. L'hypothese H_0 est donc retenue : l'hypothese d'une baisse de l'IDM n'a pas ete demontree.
2) Les hypotheses statistiques s'expriment ici par :
H_0 *: Autant de di erences positives que de di erences negatives dans la population.*
H_1 *: Dans la population parente, la frequence des di erences negatives est supeieure a 50%.*
Comme $N > 30$, on peut utiliser l'approximation de la loi binomiale par une loi normale et prendre comme statistique de test : $Z = \frac{2D - 1 - N}{\sqrt{N}}$ ou $D = \max(D_+, D_-)$. Pour un seuil de 5%, on obtient $z_{crit} = 1.65$. On a ici 12 di erences positives pour 19 di erences negatives, d'ou : $z_{obs} = \frac{38 - 1 - 31}{\sqrt{31}} = 1.07$. On accepte donc l'hypothese H_0, ce qui con rme le resultat de la question 1.

Exercice 31

Dans une recherche médicale, on a examiné 15 patients atteints de la maladie de Parkinson, dans deux conditions : à jeun et sous médicament L-Dopa. Les résultats ont permis d'établir, pour la variable *Vitesse de la marche*, le protocole suivant des effets individuels.

Sujet	Effet	Sujet	Effet	Sujet	Effet
s1	+0,23	s6	+0,12	s11	+0,30
s2	+0,32	s7	+0,33	s12	+0,10
s3	+0,04	s8	+0,28	s13	-0,12
s4	+0,19	s9	0,00	s14	+0,15
s5	+0,22	s10	+0,12	s15	+0,09

Ordonner le protocole des effets individuels et construire le protocole des rangs. L'observation correspondant à un effet nul ne sera pas prise en compte.
Procéder à un test d'homogénéïté des rangs de Wilcoxon et formuler une conclusion en termes d'effet du médicament (seuil : 5% unilatéral).
Reponse : $T_- = 5$, $T_+ = 100$. Avec les notations de la table du test de Wilcoxon, on a : $N = 14$, $T_m = 5$. Or, au seuil de 5% unilateral, $T_{m,crit} = 25$. On accepte l'hypothese alternative au seuil de 5% unilateral.

Exercice 32

On étudie les comportements agressifs d'un groupe d'enfants ayant des difficultés de comportement, pendant une demi-journée, avant et après la projection d'un film d'aventures. Peut-on dire que ce film a eu une influence sur le comportement agressif des enfants ? (on trouvera dans le tableau 3 page 19 trois colonnes indiquant le numéro du sujet, le nombre de comportements agressifs avant la projection et le nombre de comportements agressifs après la projection). Apporter une réponse à la question posée en utilisant un

Sujet	Avant	Après
1	46	71
2	31	79
3	65	27
4	61	39
5	47	75
6	32	28
7	58	32
8	14	36
9	7	61
10	48	83
11	43	80
12	33	28
13	14	72

Sujet	Avant	Après
14	84	35
15	2	28
16	5	39
17	99	53
18	88	66
19	75	40
20	71	84
21	23	14
22	84	15
23	2	99
24	34	36
25	92	13
26	61	87

Tab. 3 – Comportements agressifs

test de Wilcoxon.
Reponse : $T = T_{-} = 194.5$, $z_{obs} = 0.47$. *On ne peut pas refuser l'hypothese nulle.*
En utilisant la table du test de Wilcoxon, on a : $N = 26$, $T_{min} = T_{+} = 156.5$ *et, au seuil de 5% unilateral,* $T_{m,crit} = 110$. *On ne peut donc pas refuser* H_0.

Exercice 33

Dans une crèche, on a observé des enfants en présence de deux puéricultrices différentes ; on a compté, suivant une grille préalablement établie, le nombre de comportements verbaux de ces enfants, pendant une période déterminée. Les résultats sont donnés dans les tableaux suivants. Peut-on dire qu'il y a une différence de comportements verbaux entre les deux observations ?

Séance 1	Séance 2	Diff.	Rang-	Rang+
64	88	24		17
62	42	-20	11	
10	65	55		24
61	87	26		19
17	9	-8	5,5	
20	28	8		5,5
65	87	22		13
10	78	68		27
67	9	-58	25	
25	58	33		22
1	92	91		30
42	36	-6	4	
90	66	-24	17	
87	22	-65	26	
93	4	-89	29	

Séance 1	Séance 2	Diff.	Rang-	Rang+
16	39	23		14,5
16	66	50		23
83	70	-13	8	
45	13	-32	20,5	
33	10	-23	14,5	
21	10	-11	7	
30	54	24		17
67	50	-17	10	
34	2	-32	20,5	
93	10	-83	28	
78	73	-5	2,5	
63	42	-21	12	
49	65	16		9
71	73	2		1
93	98	5		2,5

Répondre à la question posée à l'aide d'un test des signes puis d'un test de Wilcoxon.

Reponses : Test des signes : $D = D_- = 16$, $z_{obs} = 0.18$. *On ne peut pas refuser l'hypothese nulle.*

Test de Wilcoxon : $T_- = 240.5$, $z_{obs} = 0.15$. *Même conclusion.*

Avec les notations de la table du test de Wilcoxon : $N = 30$, $T_{min} = T_+ = 224.5$. *Au seuil de 5% unilateral,* $T_{m,crit} = 151$. *On accepte donc* H_0.

Exercice 34
On dit que dans une famille, les aînés ont tendance à être plus indépendants que leurs cadets. Un chercheur élabore une échelle d'indépendance en 25 points et procède à l'évaluation de 20 aînés et du frère ou de la sœur qui suit directement chacun des aînés. Imaginons qu'il obtienne les résultats suivants :

Paire	Aîné	Cadet
1	8	9
2	13	15
3	8	10
4	5	7
5	12	10
6	15	13
7	5	8
8	15	12
9	16	13
10	5	9

Paire	Aîné	Cadet
11	17	13
12	12	8
13	2	7
14	13	8
15	19	14
16	18	12
17	14	8
18	17	11
19	18	12
20	20	10

Quels sont les individus statistiques et les variables étudiés ?
A partir des données ci-dessus, construire le protocole des rangs signés.
N.B. Malgre leur desordre apparent, les donnees ont ete classees de maniere a faciliter ce travail...
Tester l'hypothèse du chercheur à l'aide d'un test de Wilcoxon, au seuil de 5%.
Reponse : $T_+ = 164$; $T_- = 46$*. En utilisant la table de Wilcoxon, on trouve, au seuil de 5%,* $T_{m,c} = 60$*. L'hypothese \les aines sont plus independants que leurs cadets" est donc con rmee. On peut aussi utiliser l'approximation par une loi normale. On obtient* $Z_{obs} = 2.18$ *et la conclusion est la même.*

Exercice 35
Quelle différence faites-vous entre un test *parametrique* et un test *non parametrique*. Dans quelles circonstances est-il préférable d'utiliser des tests non paramétriques ?

Exercice 36 *Test du* χ^2 *de comparaison a une norme.*
Selon les lois de l'hérédité, la combinaison de deux caractères génétiques A/a et B/b se fait selon les proportions suivantes :

AB	Ab	aB	ab
$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

Sur 568 observations, on a obtenu les résultats suivants :

AB	Ab	aB	ab
327	118	98	25

Quel type de réponse peut-on apporter à la question : les résultats expérimentaux vérifient-ils les lois de l'hérédité ?
Reponses : $\chi^2_{obs} = 5,2$*, ddl* $= 3$*. Au seuil de 5%, on ne peut pas refuser l'hypothese nulle : les ecarts constates sont dus au hasard.*

Exercice 37 *Distributions theoriques - loi normale*
Dans une maternité, on a relevé le poids de 200 enfants à la naissance. Après répartition en 9 classes, les valeurs observées sont les suivantes :

Poids (en kg)	Effectif
[2.0, 2.6[	17
[2.6, 2.8[	13
[2.8, 3.0[	16
[3.0, 3.2[	30
[3.2, 3.4[	41
[3.4, 3.6[	36
[3.6, 3.8[	17
[3.8, 4.0[	14
[4.0, 4.6]	16

1) Calculer la moyenne de la variable "poids", puis la variance d'échantillon et la variance corrigée (c'est-à-dire l'estimation ponctuelle de la variance sur la population).
2) On souhaite tester la normalité de la distribution des poids dans la population parente. A cet effet, on calcule les fréquences théoriques des classes précédentes lorsque la variable suit une loi normale de paramètres $\mu = 3.306$ et $\sigma = 0.504$. On obtient les résultats suivants :

Classe	Fréquence
$]-\infty,\ 2.6[$	0.0808
[2.6, 2.8[	0.0771
[2.8, 3.0[	0.1142
[3.0, 3.2[	0.1447
[3.2, 3.4[	0.1571
[3.4, 3.6[	0.1461
[3.6, 3.8[	0.1163
[3.8, 4.0[	0.0793
$[4.0,\ +\infty[$	0.0844

Refaire les calculs concernant les deux premières lignes et la dernière ligne de ce tableau. Construire le tableau des effectifs théoriques des classes, lorsque l'effectif total est de 200 unités statistiques.
3) On exécute ensuite un test du χ^2 pour comparer les effectifs observés aux effectifs théoriques obtenus à la question précédente. Le tableau obtenu est alors le suivant :

n_i	t_i	Contributions
17	16.16	0.0434
13	15.42	0.3790
16	22.83	2.0431
30	28.95	0.0383
41	31.43	2.9155
36	29.22	1.5745
17	23.26	1.6841
14	15.85	0.2167
16	16.89	0.0466

a) Que représentent les deux premières colonnes de ce tableau ? Refaire le calcul permettant d'obtenir la valeur figurant dans la troisième colonne de la première ligne (c'est-à-dire 0.0434).

b) Calculer la "distance" du χ^2 entre la distribution théorique et la distribution observée.
c) On admet que le nombre de degrés de liberté à considérer ici est $ddl = 6$. Quelle est alors la valeur critique du χ^2 au seuil de 1% ? La valeur du χ^2 observée peut-elle être expliquée par les fluctuations d'échantillonnage, ou est-elle significative d'une absence de normalité des poids dans la population parente ?.
Reponses : 1) $m = 3.306$, $s^2 = 0.2532$, $s_c^2 = 0,2544$.
3c) $\chi^2_{obs} = 8.94$, $\chi^2_c = 16.812$. *On peut accepter l'hypothese nulle : l'ecart entre la distribution observee et une distribution normale est explique par le hasard.*

Intervalle de confiance

Exercice 38
For a given group of 500 soldiers the mean AGCT(2) score is 95.00 and the SD is 25.
a) Determine the .99 confidence intervall for the true mean.
b) It is unlikely that the true mean is larger then what value ?
Reponses : a) $[92.11, 97.89]$, *b)* 97.89

Exercice 39
Un institut de sondage a observé sur un échantillon de 1600 personnes prises au hasard, 51% d'intentions de vote en faveur du candidat A.
Estimer la proportion d'intentions de vote en faveur du candidat A dans la population à l'aide d'un intervalle de confiance, avec un degré de confiance de 95%.
Peut-on assurer, avec un risque de 5%, que M. A sera élu ?
Reponses : Intervalle de con ance : $[0.485; 0.535]$. *On ne peut donc pas assurer que M. X sera elu*

Exercice 40 Aux Etats-Unis, en 1936, deux candidats s'affrontent pour l'élection présidentielle : F.D. Roosevelt et G.M. Landon. Plusieurs sondages visant à prévoir le résultat du vote sont réalisés.
Le magazine *Literacy Digest* interroge par téléphone 2 millions d'électeurs sur leurs intentions de vote et prédit ainsi une large victoire de Landon.
G. Gallup, qui vient de fonder son institut de sondages, n'interroge que 3000 personnes et prédit la victoire de Roosevelt.
Et, c'est bien Roosevelt qui sera élu le 3 novembre 1936...
1) Les deux sondages donnent des résultats apparemment contradictoires. Parmi les raisons données ci-dessous, choisir celle (ou celles) qui paraît la plus vraisemblable. Justifier.

1. L'échantillon de Literacy Digest est trop petit.
2. L'échantillon de Literacy Digest est trop grand.
3. L'échantillon de Literacy Digest n'est pas représentatif de l'électorat.
4. Gallup a eu beaucoup de chance alors que Literacy Digest n'en a pas eu.

2) L'histoire n'a pas retenu le pourcentage d'intentions de vote en faveur de Landon observé par Literacy Digest. Supposons, dans cette question, que ce pourcentage était de $f = 60\%$. Soit p le pourcentage (inconnu) d'intensions de vote pour Landon dans la population dont est issu l'échantillon.
Déterminer un intervalle de confiance pour p, avec un degré de confiance de 99.9%.

[2] AGCT : A my Gene al Classification Test

Reponses : 1) L'echantillon de Literacy Digest n'est evidemment pas trop petit : on peut faire des sondages valables sur des echantillons beaucoup plus reduits. Et, contrairement a l'a rmation 2, il n'existe pas d'echantillon trop grand. Le resultat de la question 2 montrera que l'erreur de Literacy Digest n'est pas due a la malchance : si l'echantillon avait ete correctement choisi, la probabilite d'obtenir un resultat faux etait in me. En revanche, le choix de l'echantillon a ete biaise par la methode utilisee : on n'a interroge que des personnes possedant le telephone (qui n'etait pas universellement repandu en 1936).
2) $E^2 = \frac{f(1-f)}{n} = 0.00000012$; $E = 0.000346$. *La table de la loi normale donne :* $P(-z_\alpha \leq Z \leq z_\alpha) = 0.999$ *pour* $z_\alpha = 3.3$. *On obtient donc, comme intervalle de con ance :* $0.60 - 0.000346 \times 3.3 \leq p \leq 0.60 + 0.000346 \times 3.3$, *c'est-a-dire :* $59.88\% \leq p \leq 60.12\%$.

Exercice 41

Quelle est la capacité de la mémoire à court terme ? Pour répondre à cette question, deux chercheurs[3] ont présenté à 210 élèves de lycée une liste de 16 mots communs sur un écran de télévision à raison de 1 mot toutes les deux secondes. La moyenne du rappel est de 6.91 tandis que l'écart type est égal à 2.08. On souhaite estimer, avec un degré de confiance de 95%, la moyenne de la population dont sont issus les sujets composant l'échantillon.
Reponse : Intervalle de con ance : $[6.63; 7.19]$ *avec un degre de con ance de 95%.*

Exercice 42

On a observé un échantillon de 17 enfants présentant des troubles du comportement et âgés de 6 à 8 ans. A l'aide d'une grille comportementale, on a relevé les comportements agressifs produits durant différentes périodes de jeu collectif avec des enfants ne présentant pas ce type de troubles. Sur 60 comportements caractéristiques de l'agressivité pendant le jeu, on a obtenu une moyenne de 28.53 assortie d'un écart type de 8.10. Déterminer un intervalle de confiance de la moyenne sur la population, avec un degré de confiance de 99%.
Reponse : On utilise ici le t de Student avec 16 degres de liberte. On obtient l'intervalle suivant : $[22.79; 34.27]$.

Exercice 43

Un psychologue s'intéresse à la relation entre le sexe (variable X), le statut socio-économique (variable C) et le "locus of control" perçu. Il a pris huit adultes de chaque combinaison sexe-statut socio-économique et leur a administré une échelle portant sur le "locus of control" ; un score élevé indique que le sujet estime contrôler sa vie quotidienne. Les données observées sont rassemblées dans le tableau 4 page 25.
On veut étudier, pour les sujets de statut socio-économique moyen, s'il existe une différence de "locus of control" entre les hommes et les femmes.
Réaliser un test de comparaison de moyennes permettant d'apporter une réponse à la question posée (seuil choisi : 5%).
Reponse : Il s'agit d'une comparaison de deux moyennes sur des groupes independants. On obtient : $t_{obs} = 1.29$, *avec* $14\ ddl$. *Pour un test unilateral au seuil de 5%,* $t_{crit} = 1.76$. *On conclut donc a une di erence selon les sexes.*

Exercice 44 *Dossier \coaction"*

Au cours d'une expérience étudiant l'effet de la coaction sur l'apprentissage, on présente au sujet une liste de syllabes sans signification et on note le nombre d'essais nécessaires à sa mémorisation. Le sujet travaille soit seul, soit en présence d'un autre sujet effectuant

[3] Expé ience appo tée pa A. Lieu y, Liège, Ma daga, 1992

	statut socio-économique		
	Bas	Moyen	Elevé
Hommes	10	16	18
	12	12	14
	8	19	17
	14	17	13
	10	15	19
	16	11	15
	15	14	22
	13	10	20
Femmes	8	14	12
	10	10	18
	7	13	14
	9	9	21
	12	17	19
	5	15	17
	8	12	13
	7	8	16

TAB. 4 – Données "locus of control"

la même tâche que lui. On fait l'hypothèse que les sujets n'apprendront pas aussi vite dans les deux conditions. Les résultats obtenus sur 30 sujets travaillant seuls et 30 sujets travaillant en condition de coaction sont les suivants :

	Seuls	En coaction
Moyenne	5,6	7,5
Ecart type	1,09	1,64

Effectuer les calculs permettant de tester l'hypothèse formulée et conclure.
Reponses : En supposant que l'enonce nous donne des ecarts types corriges, on obtient $t_{obs} = -5.28$*. Au seuil de 1%, les resultats sont signi catifs d'une di erence entre les deux groupes.*

Exercice 45 *N.B. Les trois questions sont independantes*

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. L'étude a porté sur trois 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 ;
- Un groupe d'enfants dont le poids à la naissance était normal.

Le tableau 7 page 28 représente une partie des données, saisies dans la plage A1 :F33 d'une feuille Excel. Il indique d'une part, l'indice de développement mental (IDM) à 6

	A	B	C	D
34		=MOYENNE(B3 :B33)	=MOYENNE(C3 :C33)	=MOYENNE(D3 :D33)
35		=VAR(B3 :B33)	=VAR(C3 :C33)	=VAR(D3 :D33)
36		=VAR.P(B3 :B33)	=VAR.P(C3 :C33)	=VAR.P(D3 :D33)
37		=ECARTYPE(B3 :B33)	=ECARTYPE(C3 :C33)	=ECARTYPE(D3 :D33)
38		=ECARTYPEP(B3 :B33)	=ECARTYPEP(C3 :C33)	=ECARTYPEP(D3 :D33)

Tab. 5 – Formules

	E	F
34		=MOYENNE(F3 :F33)
35		=VAR(F3 :F33)
36		=VAR.P(F3 :F33)
37		=ECARTYPE(F3 :F33)
38		=ECARTYPEP(F3 :F33)

Tab. 6 – Formules (suite)

mois et à 24 mois pour le groupe témoin PRN, ainsi que la différence entre les deux indices et d'autre part l'IDM à 24 mois pour le groupe expérimental PRN.
Les tableaux 5 page 26 et 6 page 26 indiquent les formules utilisées sous Excel pour calculer les paramètres descriptifs indiqués dans la zone B34 :F38 de la feuille.

1) Des travaux antérieurs sur de tels enfants laissent penser que les scores IDM des enfants PRN du groupe témoin pouvaient diminuer de façon sensible entre 6 et 24 mois. Réaliser un test de comparaison de moyennes afin de déterminer si les données observées confirment cette hypothèse (choisir un seuil de 5%).

2) Le programme de formation proposé est-il efficace ? Apporter une réponse à cette question en procédant à un autre test de comparaison de moyennes. Choisir de même un seuil de 5%.

3) Peu satisfait du résultat obtenu à la question 1, un chercheur souhaite comparer les données relatives aux enfants du groupe témoin PRN à l'aide d'un test du signe. Réaliser le test et conclure.

1) Soit μ_6 et μ_{24} les moyennes de l'IDM à 6 mois et à 24 mois dans la population dont est issu l'échantillon. L'hypothèse de recherche se traduit par les hypothèses statistiques suivantes :
H_0 : $\mu_6 = \mu_{24}$
H_1 : $\mu_6 > \mu_{24}$
Il s'agit d'un test de comparaison de moyennes sur des groupes appareillés. La statistique de test est $t = \frac{\overline{d}}{E}$ avec $E^2 = \frac{s_c^2}{n}$. Compte tenu de l'effectif ($n = 31$), on peut utiliser ici une loi de Student avec 30 ddl ou une loi normale. Pour un seuil de 5% (et un test unilatéral avec la zone de rejet "à gauche"), on obtient $t_{crit} = -1.70$ (ou $z_{crit} = -1.65$).
En utilisant les valeurs des cellules D34 et D35 du tableau Excel fourni, on obtient : $E^2 = \frac{257.21}{31} = 8.30$; $E = \sqrt{8.30} = 2.88$; $t_{obs} = \frac{-4.29}{2.88} = -1.49$. L'hypothèse H_0 est donc retenue : l'hypothèse d'une baisse de l'IDM n'a pas été démontrée.

2) Il s'agit maintenant d'une comparaison de moyennes sur deux groupes indépendants. On utilise la méthode adaptée aux petits échantillons (car $n' = 25$). On a :
H_0 *:* $\mu = \mu'$
H_1 *:* $\mu < \mu'$.
La statistique de test suit une loi de Student avec 54 ddl. Pour un seuil de 5% (et un test unilatéral), on obtient $t_{crit} = -1.67$.
En utilisant les valeurs des cellules C34, F34, C36 et F36, on obtient :
$E^2 = \dfrac{31 \times 162.40 + 25 \times 154.4}{54}\left(\dfrac{1}{31} + \dfrac{1}{25}\right) = 11.90$ *d'où* $E = 3.45$ *et* $t_{obs} = \dfrac{-10.49}{3.45} = -3.04$.
Comme $t_{obs} < t_{crit}$*, on en déduit que la différence entre les deux groupes est significative d'un effet du programme de formation, au bénéfice du groupe expérimental.*
3) Les hypothèses statistiques s'expriment ici par :
H_0 *: Autant de différences positives que de différences négatives dans la population.*
H_1 *: Dans la population parente, la fréquence des différences négatives est supéieure à 50%.*
Comme $N > 30$*, on peut utiliser l'approximation de la loi binomiale par une loi normale et prendre comme statistique de test :* $Z = \dfrac{2D - 1 - N}{\sqrt{N}}$ *où* $D = \max(D_+, D_-)$*. Pour un seuil de 5%, on obtient* $z_{crit} = 1.65$*. On a ici 12 différences positives pour 19 différences négatives, d'où :*
$z_{obs} = \dfrac{38 - 1 - 31}{\sqrt{31}} = 1.07$*. On accepte donc l'hypothèse* H_0*, ce qui confirme le résultat de la question 1.*

	A	B	C	D	E	F
1	PRN Témoin				PRN expérimental	
2		IDM-6	IDM-24	Diff		IDM-24
3	s1	124	114	-10	s'1	96
4	s2	94	88	-6	s'2	127
5	s3	115	102	-13	s'3	127
6	s4	110	127	17	s'4	137
7	s5	116	104	-12	s'5	114
8	s6	139	104	-35	s'6	119
9	s7	116	91	-25	s'7	109
10	s8	110	96	-14	s'8	109
11	s9	129	104	-25	s'9	143
12	s10	120	106	-14	s'10	109
13	s11	105	91	-14	s'11	116
14	s12	88	102	14	s'12	114
15	s13	120	104	-16	s'13	143
16	s14	120	100	-20	s'14	109
17	s15	116	114	-2	s'15	117
18	s16	105	109	4	s'16	127
19	s17	100	109	9	s'17	112
20	s18	91	119	28	s'18	112
21	s19	129	91	-38	s'19	98
22	s20	84	81	-3	s'20	137
23	s21	91	114	23	s'21	112
24	s22	116	119	3	s'22	109
25	s23	100	102	2	s'23	119
26	s24	113	111	-2	s'24	106
27	s25	89	80	-9	s'25	109
28	s26	102	119	17		
29	s27	110	119	9		
30	s28	116	123	7		
31	s29	124	119	-5		
32	s30	126	114	-12		
33	s31	123	132	9		
34		111.00	106.71	-4.29		117.20
35		191.8667	167.8129	257.2129		160.8333
36		185.6774	162.3996	248.9157		154.4000
37		13.8516	12.9543	16.0379		12.6820
38		13.6264	12.7436	15.7771		12.4258

TAB. 7 – Données PRN

Exercice 46
Un chercheur a étudié les souvenirs de professeurs d'université à l'égard de visages d'étudiants. Ces enseignants avaient côtoyé ces étudiants au cours de séances de travail en petits groupes pendant dix semaines. Deux semaines après, les enseignants devaient reconnaître les visages des étudiants à partir de photos mélangées avec des photos de visages qu'ils ne pouvaient avoir vu. Un an après, la même tâche était demandée. Les résultats détaillés sont les suivants :

		1 an		
		Reconnu	Non reconnu	Total
Deux semaines	Reconnu	81	46	127
	Non reconnu	8	49	57
	Total	89	95	184

Le chercheur fait l'hypothèse d'une baisse significative, au risque de 5%, des taux de reconnaissance avec le temps. Les résultats observés confirment-ils cette hypothèse ?
Reponse. En utilisant un test de comparaison de proportions appariees, on obtient : $z_{obs} = \frac{46-8}{\sqrt{46+8}}$*. Pour un test unilateral (cf. l'hypothese de recherche), on a* $z_{crit} = 1.645$*. L'hypothese du chercheur est donc con rmee.*

Exercice 47 *Dossier \grsang"*
En France, la fréquence du groupe sanguin A est $p_F = 0,45$.
1) Sur un échantillon de $n_S = 400$ Suédois on a trouvé 192 individus de groupe A. La fréquence du groupe sanguin A chez les Suédois diffère-t-elle de p_F ?
2) Sur un échantillon de 100 Ecossais, on a observé 32 individus de groupe A. Cette fréquence diffère-t-elle de celle observée chez les Suédois ?
3) Le groupage sanguin trouvé pour les 400 Suédois est le suivant :
A : 192 ; B : 40 ; AB : 23 ; O : 145.
Diffère-t-il significativement du groupage théorique français :
A : 0,45 ; B : 0,08 ; AB : 0,03 ; O : 0,44.
4) Pour l'échantillon des 100 Ecossais la répartition des groupes est la suivante :
A : 32 ; B : 15 ; AB : 3 ; O : 50.
Les fréquences de ces groupes sont-elles différentes chez les Suédois et les Ecossais ?
Reponses : 1) Comparaison d'une frequence observee a une frequence theorique. $z_{obs} = 1.20$*. On ne peut pas refuser* H_0*.*
2) Comparaison de deux frequences observees. $z_{obs} = 2.877$*. On refuse* H_0 *au seuil de 5% dans le cas d'un test bilateral.*
3) Test de χ^2 *de comparaison d'une distribution observee a une distribution theorique.* $\chi^2_{obs} = 18.34$*,* $ddl = 3$*. On rejette* H_0 *au seuil de 5%.*
4) Test du χ^2 *sur un tableau de contingence.* $\chi^2_{obs} = 11.43$*,* $ddl = 3$*. On rejette* H_0 *au seuil de 5%.*

Exercice 48 *Basket-Ball*
On a relevé la taille de 31 garçons de 14 ans, entraînés aux épreuves de basket-ball. Les garçons sont répartis en deux groupes d'effectifs $n_1 = 12$ et $n_2 = 19$, selon le jugement porté par l'entraîneur (groupe g_1 : jugement positif ; groupe g_2 : jugement négatif). Les données sont les suivantes :
g_1 : 188 184 178 178 177 177 176 174 173 164 163 152
g_2 : 195 193 189 189 188 186 181 181 180 179 179 177 176 176 174 174 172 171 167

1) Effectuer un test de la médiane visant à déterminer si le jugement porté par l'entraîneur est lié à la taille des sujets.
2) Reprendre les mêmes données et conclure à l'aide d'un test de Wilcoxon-Mann-Whitney.
Reponses : 1) La mediane generale vaut $M = 177$. On peut former le tableau de contingence suivant :

	$\leq M$	$> M$	*Total*
$g1$	*8*	*4*	*12*
$g2$	*8*	*11*	*19*
Total	*16*	*15*	*31*

Apres avoir forme le tableau des e ectifs theoriques sous hypothese d'independance des deux variables, on obtient : $\chi^2_{obs} = 1.78$. Ici, $ddl = 1$, et on conclut a l'homogene•te des deux groupes au seuil de 5%.
2) Apres construction du protocole des rangs, on obtient $\overline{R_1} = 19.54$, $\overline{R_2} = 13.76$ et $z_{obs} = 1.72$. Au seuil de 5%, l'hypothese H_0 est acceptee dans le cas d'un test bilateral. Il est di cile de conclure nettement dans le cas d'un test unilateral.
Avec les notations de la table du test de Wilcoxon Mann Whitney, on a : $W = 234.5$ et, au seuil de 5 %, $W'_s = 234$. Comme precedemment, il est di cile de conclure nettement.