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 référence : 1/10/1997)*
- codage numérique 2 : âge en mois *(compté au 1/10/97)*
- codage en rangs *(rappel : rang 1 pour le plus jeune, convention habituelle pour les ex-æquos : 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.
Réponses : 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	Ord.	Bin.	Centre	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 théoriques - 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'hypothèse que la probabilité d'avoir un garçon ou une fille est égale à 0.5
3) Représenter la distribution obtenue à l'aide d'un diagramme en bâtons.
Réponse: Les effectifs respectifs des familles comportant 0, 1,..., 5 garçons sont donnés par: 3, 16, 31, 31, 16, 3

Exercice 3 *Distributions théoriques - 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 ?.
Réponses : 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'hypothèse nulle : l'écart entre la distribution observée et une distribution normale est expliqué par le hasard.*

Exercice 4 *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%.
Réponses numériques : $\chi^2_{obs} = 8,48$, $ddl = 3$, $\chi^2_c = 7,815$. La différence entre les deux sexes est donc significative.

Exercice 5 *Test du* χ^2 *de comparaison à 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 ?

Réponses : $\chi^2_{obs} = 5,2$, $ddl = 3$. *Au seuil de 5%, on ne peut pas refuser l'hypothèse nulle : les écarts constatés sont dus au hasard.*

Exercice 6 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 :

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

Réponses : Sans correction de Yates, on obtient : $\chi^2_{obs} = 15.12$. *La différence est donc significative au seuil de 1%.*

Avec la correction de Yates, on obtient : $\chi^2_{obs} = 14,27$. *En utilisant la formule indiquée ci-dessus, on obtient de même:* $\chi^2_{obs} = 14.27$.

Exercice 7

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'après* 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) 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 aux seuils traditionnels.
Réponses :
1) Les fréquences sont respectivement de $f_1 = \frac{33}{428} = 7.7\%$ *et* $f_2 = \frac{11}{453} = 2.4\%$.
2) Le tableau de contingence, le tableau des effectifs théoriques et celui des contributions au χ^2 *sont donnés par :*

	Placébo	*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 affirmer, au seuil de 1%, l'existence d'une liaison entre le type de traitement et l'apparition de la maladie.*

Exercice 8

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.
Réponses : Les variables mises en jeu dans le tableau de contingence proposé sont d'une part la couleur de la poupée (noire ou blanche), d'autre part le numéro d'ordre de l'expérience (expérience 1 ou expérience 2). L'hypothèse correspondant au test du χ^2 *est donc ici : le choix pour l'une ou l'autre des poupées dépend-il de l'expérience considérée ? ou encore : les deux expériences fournissent-elles des résultats identiques ? On trouve :* $\chi^2_{obs} = 34.15$; $ddl = 1$; *pour* $\alpha = 5\%, \chi^2_{crit} = 3.84$ *et donc les deux expériences fournissent des résultats différents.*

Echantillonnage

Exercice 9

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.
Réponses : (a) 69%, (b) 23%, (c) moins de 1%.

[1] σ_K désigne ici l'écart type de la distribution d'échantillonnage

Exercice 10
Un test comportemental a été étalonné auprès d'un échantillon 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.
Réponses : 1) $P(X > 62) = 0.1291$*, d'où* $N_1 = 452$*.*
2) $P(45 \leq X < 70) = 0.9011$ *d'où* $N_2 = 3153$*.*
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'échantillonnage a pour paramètres : $\mu = 54.24$ *et* $S = 0.682$*. L'intervalle, centré autour de la moyenne, rassemblant 95% des échantillons est :* $[52.90; 55.58]$.

Exercice 11
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) ?
Réponse : 1850

Intervalle de confiance

Exercice 12
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 ?
Réponses: a) $[92.11, 97.89]$*, b)* 97.89

Exercice 13
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 ?
Réponses : Intervalle de confiance : $[0.485; 0.535]$*. On ne peut donc pas assurer que M. X sera élu*

Exercice 14 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

[2]AGCT : Army General Classification Test

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%.
Réponses : 1) L'échantillon de Literacy Digest n'est évidemment pas trop petit : on peut faire des sondages valables sur des échantillons beaucoup plus réduits. Et, contrairement à l'affirmation 2, il n'existe pas d'échantillon trop grand. Le résultat de la question 2 montrera que l'erreur de Literacy Digest n'est pas due à la malchance : si l'échantillon avait été correctement choisi, la probabilité d'obtenir un résultat faux était infime. En revanche, le choix de l'échantillon a été biaisé par la méthode utilisée : on n'a interrogé que des personnes possédant le téléphone (qui n'était pas universellement répandu 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 confiance :* $0.60 - 0.000346 \times 3.3 \leq p \leq 0.60 + 0.000346 \times 3.3$, *c'est-à-dire :* $59.88\% \leq p \leq 60.12\%$.

Exercice 15
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.
Réponse : Intervalle de confiance : $[6.63; 7.19]$ *avec un degré de confiance de 95%.*

Exercice 16
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 caratéristiques de l'agressivité pendant le jeu, on a obtenu une moyenne de 28.53 assortie d'un écart type de 8.10. Déterminer

[3] Expérience rapportée par A. Lieury, Liège, Mardaga, 1992

un intervalle de confiance de la moyenne sur la population, avec un degré de confiance de 99%.
Réponse : On utilise ici le t de Student avec 16 degrés de liberté. On obtient l'intervalle suivant : $[22.79; 34.27]$.

Inférence sur une moyenne. Comparaison à une norme

Exercice 17

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.
Réponses : a, c, d, e

Exercice 18

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 $\overline{QI} = 84.4$. 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 $\overline{QI} = 84.4$. Mêmes questions que précédemment.
Réponses : a) $z_{obs} = -2.326$. *Groupe atypique au seuil de 2.5% unilatéral. b)* $z_{obs} = -5.2$. *Groupe atypique au seuil de 1% unilatéral*

Exercice 19

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.
Réponse : Il s'agit ici de comparer la moyenne observée sur l'échantillon à la valeur 3.87, considérée comme une norme. Compte tenu de la taille de l'échantillon, 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 centrée réduite.

On obtient alors $Z_{obs} = 1.20$*, ce qui n'est pas significatif d'une différence.*

Exercice 20

"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.

Réponse : Une hypothèse nulle peut ne pas être rejetée à cause d'une trop faible taille d'échantillon, ou de l'existence de variables non contrôlées qui ont artificiellement augmenté la dispersion. Il y a des cas où l'on est conduit à accepter l'hypothèse nulle, jusqu'à ce que des informations contraires viennent la contredire. Exemple : absence d'effet indésirable d'un médicament.