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	$F(X <= Z_i)$	Theo	Ecart
5	-1,5877	0,125	0,0562	0,0688
7	-1,0104	0,25	0,1562	0,0938
8	-0,7217	0,375	0,2352	0,1398
11	0,1443	0,5	0,5574	0,0574
12	0,4330	0,625	0,6675	0,0425
13	0,7217	0,75	0,7648	0,0148
13	0,7217	0,875	0,7648	0,1102
15	1,2990	1	0,9030	0,0970
			Maximum	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	$F(X <= Z_i)$	Theo	Ecart
5	-1,0141	0,125	0,1553	0,0303
5,5	-0,9239	0,25	0,1778	0,0722
5,5	-0,9239	0,375	0,1778	0,1972
6	-0,8338	0,5	0,2022	0,2978
14	0,6085	0,625	0,7286	0,1036
16	0,9690	0,75		
16	0,9690	0,875		
17		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, 0.8337 et 0.8748. Les écarts peuvent alors être déterminés en calculant les différences entre 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.*

Exercice 2

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 3

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

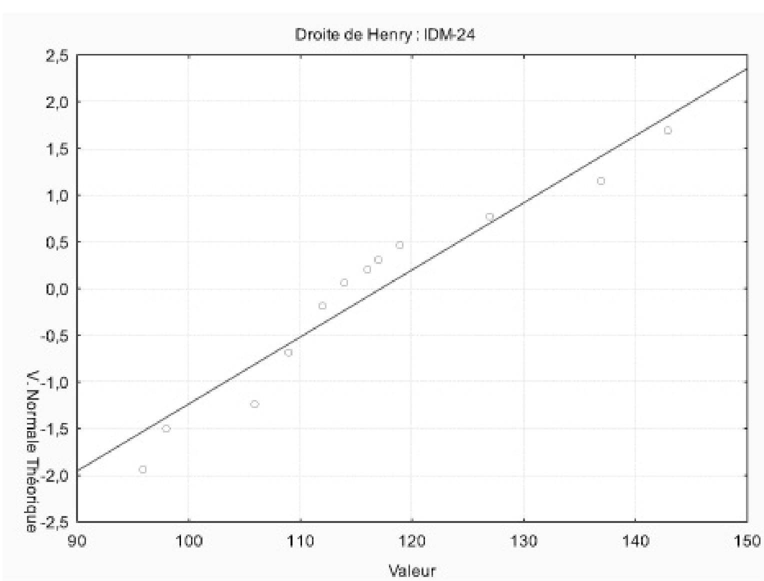
– pour l'IDM à 24 mois dans le groupe PRN témoin

	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 4 *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 5
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 6

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 :

	Test de Levene d'Homogénéité des Variances (TR.sta) Effets significatifs marqués à p < ,05000							
Variable	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

	Test d'Homogénéité des Variances de Brown-Forsythe (TR.sta) Effets significatifs marqués à p < ,05000							
Variable	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 7 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^2_{ic}	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 59 ddl, et la valeur observée est alors significative d'une différence entre les deux groupes.

Exercice 8
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.

Taille d'un effet - Puissance d'un test

Exercice 9
Cette recherche visait à étudier, en particulier, l'évolution du transfert interhémisphérique chez le jeune enfant. On s'intéresse ici à une partie des données, recueillies chez deux groupes d'enfants droitiers de 5 ans (a1) et 6 ans (a2). Ils ont à comparer, les yeux bandés, les textures de petits coussins. Chaque coussin est placé sur la paume de la main de l'enfant. Chaque enfant réalise plusieurs comparaisons de textures dans deux conditions : les deux textures sont présentées successivement, soit sur une main, puis sur l'autre (condition croisée, notée c1), soit sur la même main (condition non croisée, notée c2). On note, pour chaque enfant, le pourcentage d'erreurs dans chacune des deux conditions.

On fait l'hypothèse que, chez les jeunes enfants, la comparaison est plus difficile dans la condition croisée. En effet, dans cette condition, la comparaison nécessite un transfert interhémisphérique des informations tactiles recueillies pour qu'elles puisse être comparées. Or, ce transfert interhémisphérique ne s'établit que progressivement au cours du développement.
On présente ci-dessous les données recueillies sur un sous-ensemble de 10 enfants :

	c1	c2
s1a1	30	20
s2a1	35	18
s3a1	40	22
s4a1	28	17
s5a1	27	23
s6a2	18	12
s7a2	15	22
s8a2	20	19
s9a2	21	20
s10a2	11	22

1) On compare les scores observés pour les enfants de 5 ans, dans les deux conditions.
a) Calculer l'effet calibré observé sur les données fournies. Comment peut-on qualifier cet effet ?
b) En prenant l'effet calibré précédent comme estimation de la taille de l'effet, calculer la puissance du test lorsque le seuil est de 5% et la taille de l'échantillon est de 5.
2) On compare les scores observés dans la condition c2 pour les deux groupes d'enfants. Calculer l'effet calibré observé sur les données fournies. Comment peut-on qualifier cet effet ?
Indications de réponses : 1) a) Dans le groupe d'âge a1, la moyenne des différences individuelles est de 12, avec un écart type corrigé de 5.70. L'effet calibré observé est donc $EC = \frac{12}{5.7} = 2.10$. *Cet effet est important.*
1) b) Pour $n = 5$, *la quantité* $\Delta = d\sqrt{n}$ *vaut* $\Delta = 2.10\sqrt{5} = 4.69$. *La consultation de la table pour le seuil de 5% montre que la puissance du test est alors voisine de 100%.*
2) Les moyennes dans les deux groupes sont alors 20 et 19, avec des écarts types corrigés de 2.55 et 4.12, soit un écart type corrigé pondéré de 3.34. L'effet calibré observé est donc : $EC = \frac{20-19}{3.34} = 0.30$. *Cet effet est faible.*

Exercice 10

Nous venons de réaliser une étude qui compare le développement cognitif de bébés d'un an présentant un poids réduit à la naissance. Grâce à une échelle conçue par nos soins, nous avons calculé que les moyennes d'échantillons des deux groupes étudiés étaient respectivement égales à 25 et 30, pour un écart type combiné de 8.
Supposons que nous souhaitions répliquer cette expérience avec 20 sujets par groupe. Si nous posons que les moyennes et écarts types réels ont été correctement estimés, quelle probabilité avons-nous a priori de trouver une différence significative au seuil de 5% lors de la réplication ?
Indications de réponses : L'effet calibré observé est $EC = \frac{25-20}{8} = 0.625$. *En utilisant cette valeur comme estimation de la taille de l'effet d, on obtient* $\Delta = 0.625\sqrt{10} = 1.97$. *Au seuil de 5% bilatéral, la puissance du test envisagé serait de 52%.*

Exercice 11

De nombreux travaux sur l'effet de la pression des pairs ont montré que le score d'influence moyen est de 520 avec un écart type de 80. Un chercheur voudrait montrer qu'une légère modification des conditions génèrerait une moyenne de 500 seulement, et il envisage d'effectuer un test t pour comparer sa moyenne d'échantillon à une moyenne de 520.
1) Quelle est la taille de l'effet en question ?
2) Quelle est la valeur de Δ si la taille de l'échantillon est égale à 100 ?
3) Quelle est la puissance du test ? (seuil choisi : 5%)
4) Représenter par un diagramme la situation décrite.
5) Quelles tailles d'échantillon faudrait-il pour porter la puissance à .80 ? .90 ?
Indications de réponses : La situation proposée est celle de la comparaison d'une moyenne observée à une moyenne de référence, considérée comme une norme. Autrement dit, la moyenne et l'écart type de la population sont ici connus.
1) Le chercheur estime que la taille de l'effet est $d = \frac{520-500}{80} = 0.25$.
2) La statistique de test correspondante ($z = \frac{\overline{x}-\mu}{E}$ *avec* $E^2 = \frac{\sigma^2}{n}$*) s'écrit :* $z = EC\sqrt{n}$*. On a donc* $\Delta = d\sqrt{n} = 0.25\sqrt{100} = 2.5$.
3) Au seuil de 5%, la valeur $\Delta = 2.5$ *correspond à une puissance de 71% (lecture de la table).*
4) Pour une puissance de .80, on devrait avoir $\Delta = 2.80$*, soit* $n = \left(\frac{2.8}{0.25}\right)^2 = 125$*. De même, pour une puissance de .90, il faudrait :* $n = 169$.

Exercice 12

Inventez un exemple simple, comprenant deux groupes, afin de montrer que pour un total de 30 sujets, la puissance augmente à mesure que les tailles d'échantillons se rapprochent de l'égalité.

Exercice 13

Des sondages déjà publiés indiquent que, lors d'un prochain référendum, le oui devrait l'emporter avec 55% des voix.
Vous souhaitez réaliser vous-même un sondage confirmant le vote en faveur du oui. Quelle taille d'échantillon faut-il choisir pour que la probabilité de conclure sur le succès du oui soit d'au moins 70% (seuil utilisé : 5%) ?
Indications de réponses : La situation proposée est la comparaison d'une fréquence observée à une norme. La taille de l'effet envisagé est : $d = \frac{0.55-0.50}{\sqrt{0.55\times0.45}} = 0.10$*. La statistique de test s'écrit ici* $z = EC\sqrt{n}$*, et on a donc* $\Delta = 0.10\sqrt{n}$*. La puissance de 70% sera obtenue pour* $\Delta = 2.5$*, c'est-à-dire* $n = 619$.