

Laboratoire 5 – Introduction au test T pour deux échantillons indépendants

Dans le laboratoire précédent, nous avons travaillé sur les données du Programme pancanadien d'évaluation (PPCE) 2013 à l'aide de techniques d'analyse descriptive. Nous nous concentrerons maintenant sur les techniques d'analyse inférentielle. Dans le présent laboratoire, nous ciblerons les différences entre **deux** groupes d'élèves.

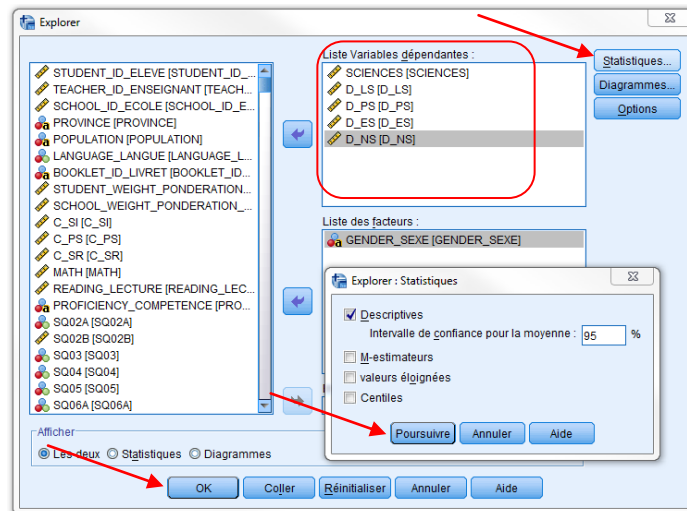
1. Test T pour échantillons indépendants – Incidence du sexe sur le rendement

Vous souhaitez voir s'il existe entre les filles et les garçons une différence de rendement en sciences, en mathématiques et en lecture (sans sous-scores). La première étape consiste à fournir un tableau de statistiques descriptives des variables pertinentes, mais puisque vous avez déjà produit ce tableau (au laboratoire 2), vous n'avez pas à refaire ce travail.

Question :

1. Veuillez formuler l'hypothèse nulle et l'hypothèse alternative pour chacune des variables continues.

Après avoir formulé de façon explicite vos hypothèses, il importe de fournir des statistiques descriptives. Veuillez produire un tableau qui compare le rendement des filles et des garçons en sciences, en mathématiques et en lecture (sans sous-scores). Pour étudier les groupes formés selon le sexe, cliquez sur Analyse → Statistiques descriptives → Explorer → sélectionnez toutes les variables qui vous intéressent et placez-les dans la boîte « Liste Variables dépendantes » → placez GENDER_SEXE dans la boîte « Liste des facteurs » → cliquez sur Statistiques → assurez-vous que « Descriptives » est coché → cliquez sur Poursuivre → cliquez sur OK.



Le tableau suivant pourrait vous aider à produire votre tableau.

Tableau 1

Titre (Veuillez donner un titre à ce tableau.)

	Garçons				Filles			
	Moyenne	Écart-type	Skewness (asymétrie)	Kurtosis (aplatissement)	Moyenne	Écart-type	Skewness (asymétrie)	Kurtosis (aplatissement)
Sciences								
Lecture								
Math.								

Quand vous rédigez un article, vous ne pouvez pas simplement insérer un tableau dans le texte sans donner une explication. Veuillez écrire un court paragraphe pour accompagner votre tableau, comme si vous publiiez ce dernier dans un article.

La prochaine étape consiste à vérifier les trois hypothèses pour vos données : l'indépendance de l'échantillon, la normalité et l'homogénéité de la variance. Si ces hypothèses sont violées, les résultats des analyses risquent d'être trompeurs ou erronés.

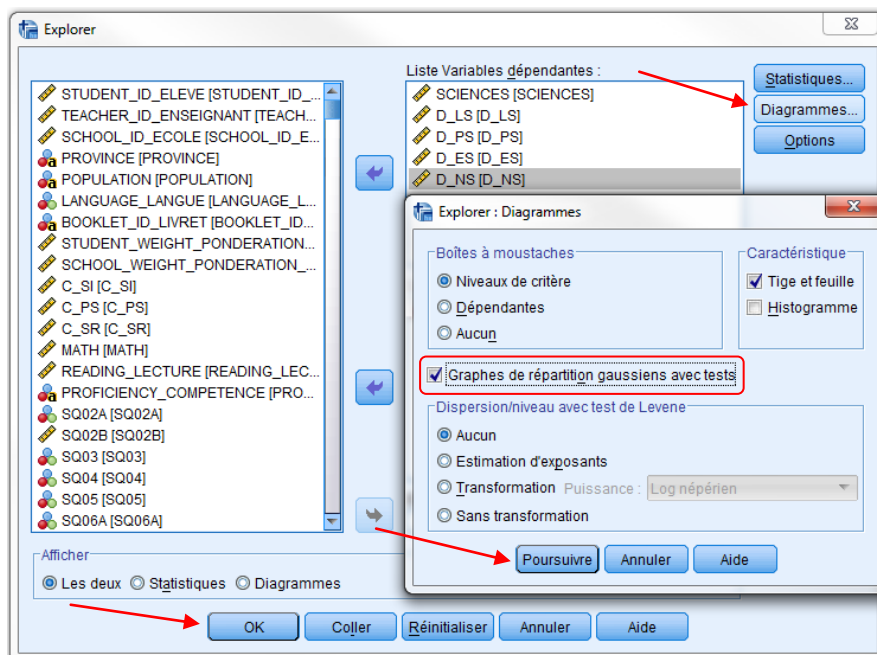
Question :

2. Veuillez dire si vos deux groupes satisfont l'hypothèse d'indépendance de l'échantillon. Si oui, veuillez expliquer pourquoi (en fournissant des éléments de preuve qualitatifs).

Vous devez également fournir des éléments de preuve empiriques de la satisfaction des hypothèses de normalité et d'homogénéité de la variance. Pour vérifier l'hypothèse de

normalité, nous pouvons consulter les histogrammes des données des deux groupes, mais le jugement humain est relativement subjectif. C'est pourquoi nous nous en remettons à des tests statistiques objectifs, tels les tests de normalité de Kolmogorov-Smirnov et de Shapiro-Wilk. Pour lancer ces tests, cliquez sur Analyse → Statistiques descriptives → Explorer → placez les scores dans la boîte « Liste Variables dépendantes » → placez la variable GENDER_SEXE dans la boîte « Liste de facteurs » → cliquez sur Diagrammes → cochez « Graphes de répartition gaussiens avec tests » → cliquez sur Poursuivre → cliquez sur OK.

Veuillez vous assurer que les pondérations pour les élèves sont utilisées tout au long des analyses.



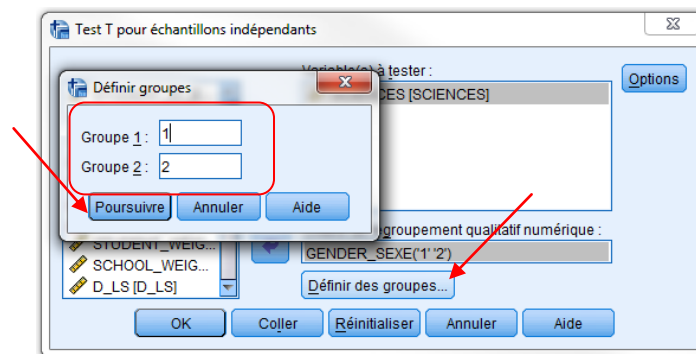
Maintenant, analysez les tests de normalité pour chaque groupe d'élèves. Si les résultats des tests de normalité de Kolmogorov-Smirnov et de Shapiro-Wilk sont significatifs, la normalité des données des deux groupes est significativement différente (l'hypothèse de normalité n'est pas satisfaite).

Questions :

3. Quelle est la différence entre le test de normalité de Kolmogorov-Smirnov et celui de Shapiro-Wilk?
4. Lequel des tests (Kolmogorov-Smirnov ou Shapiro-Wilk) devriez-vous utiliser et pourquoi?
5. À l'aide du test de normalité que vous avez choisi à la question 4, veuillez expliquer si vos deux groupes satisfont ou non l'hypothèse de normalité. Veuillez fournir des

éléments de preuve statistiques. Votre réponse devrait suivre la formule « Les données pour cette étude [satisfait/ne satisfait pas] l'hypothèse de normalité, avec une valeur $p < 0,05$. »

Pour vérifier l'homogénéité de la variance, nous pouvons, encore une fois, consulter les histogrammes et les valeurs des variables continues, mais il est difficile de dire si ces valeurs sont bel et bien différentes. C'est pourquoi nous nous en remettons à des tests statistiques, tel le test d'égalité des variances de Levene. Comme avec les tests de normalité, si les résultats du test de Levene sont significatifs, les variances des données des deux groupes sont significativement différentes (l'hypothèse d'homogénéité de la variance n'est pas satisfaite). Le test de Levene est automatiquement exécuté avec le test T. Pour lancer le test T pour des échantillons indépendants (par exemple, selon le sexe), cliquez sur Analyse → Comparer les moyennes → Test T pour échantillons indépendants → placez *toutes les variables pertinentes* dans la boîte « Variables à tester » → placez GENDER_SEXE dans la boîte « Critère de regroupement qualitatif numérique » → Définir des groupes... → utilisez les valeurs précisées → précisez Groupe 1 = 1 et Groupe 2 = 2 → Poursuivre → Options → cliquez sur « Exclure toute observation incomplète » → Poursuivre → OK.



Pour lire les résultats, vous devrez consulter le tableau « Test d'échantillons indépendants ». La colonne « Sig. » vous aidera à déterminer si vos données sont ou non statistiquement différentes entre les deux groupes. Des valeurs inférieures à 0,05 indiquent une différence statistiquement significative. Par exemple, dans la colonne « Sig. », sous « Test de Levene sur l'égalité des variances », dans la rangée « READING_LECTURE », Sig. = 0,821 indique que la variance entre les deux sexes n'est pas statistiquement significative et que, par conséquent, les deux ensembles de données présentent des « variances égales » présumées. Nous lisons donc la rangée supérieure de la variable de la lecture. Puisque la valeur Sig. (bilatérale) pour le test T est de 0,000, donc inférieure à 0,05, il y a une différence statistiquement significative entre le rendement en lecture des filles et celui des garçons.



Test d'échantillons indépendants

		Test de Levene sur l'égalité des variances		Test-t pour égalité des moyennes						
		F	Sig.	t	ddl	Sig. (bilatérale)	Différence moyenne	Différence écart-type	Intervalle de confiance 95% de la différence	
									Inférieure	Supérieure
SCIENCES	Hypothèse de variances égales	.285	.593	-1.438	2991	.150	-5.27193	3.66556	-12.45919	1.91534
	Hypothèse de variances inégales			-1.439	2984.146	.150	-5.27193	3.66478	-12.45769	1.91383
MATH	Hypothèse de variances égales	3.598	.058	-1.572	2991	.116	-4.64839	2.95657	-10.44551	1.14873
	Hypothèse de variances inégales			-1.570	2960.402	.116	-4.64839	2.96037	-10.45298	1.15620
READING_LECTURE	Hypothèse de variances égales	.051	.821	-10.367	2991	.000	-31.37857	3.02687	-37.31353	-25.44361
	Hypothèse de variances inégales			-10.368	2983.491	.000	-31.37857	3.02642	-37.31265	-25.44449

Question :

6. À l'aide du test de Levene, veuillez expliquer si vos deux groupes satisfont l'hypothèse d'homogénéité de la variance. Veuillez fournir des éléments de preuve statistiques. Votre réponse devrait suivre la formule « Les données pour cette étude [satisfont/ne satisfait pas] l'hypothèse d'homogénéité de la variance, $p[</>]0,05$. »

Puisque les tests de vérification des hypothèses ne sont pas l'objet principal de notre recherche, nous n'avons habituellement pas à donner de très nombreux éléments de preuve (par exemple, les statistiques des tests de Kolmogorov-Smirnov et de Shapiro-Wilk et du test F de Levene) pour appuyer les affirmations selon lesquelles les données satisfont ou violent ces hypothèses.

Questions :

7. Vos données satisfont-elles toutes ces hypothèses?
8. On dit que « le test T est très robuste à la violation de la normalité ». Qu'est-ce que ça signifie? (Il vous faudra peut-être recourir à Google pour cette question.)
9. Recommanderiez-vous de poursuivre les analyses de vos données à l'aide d'un test T?

Peu importe votre réponse à la question 9, nous continuerons d'analyser les résultats du test T.

Questions :

10. Pour chacune des trois variables, veuillez indiquer la valeur t, les degrés de liberté et la valeur p observée. (**Remarque** : si l'hypothèse d'homogénéité de la variance n'est pas satisfaite, vous devrez utiliser les données de la seconde ligne.)
11. Selon les valeurs p, lesquelles de vos variables (s'il y en a) se trouvent dans la région critique?
12. Y a-t-il des différences de rendement entre les filles et les garçons en sciences, en mathématiques et en lecture? Répondez à cette question en suivant la formule « Il y a entre les sexes une différence de rendement statistiquement [significative/non significative] pour la variable [nom de la variable], $t([\text{degré de liberté}])=[\text{valeur } t]$, $p[</>]0,05$. » Veuillez vous assurer de suivre les règles de l'American Psychological Association (association américaine de psychologie) dans la rédaction de cette phrase.
13. À l'aide de la différence moyenne et de l'écart type obtenus grâce SPSS, veuillez calculer manuellement le d de Cohen et le pourcentage de la variance expliquée

(Indice : pensez à employer une calculatrice en ligne – Google est utile pour cette question). Qu'indique chacune de ces valeurs (que nous révèle l'ampleur de l'effet)?

****Vous devez calculer l'ampleur de l'effet uniquement pour les variables qui présentent une différence significative entre les deux groupes.****

***Veillez noter** que la différence moyenne est la différence entre les deux moyennes que vous avez indiquées dans votre tableau de statistiques descriptives. Dans les résultats SPSS, elle est simplement appelée « moyenne », mais dans vos documents, vous devrez indiquer qu'il s'agit de la « différence moyenne » pour montrer la différence entre la « moyenne » présentée dans vos analyses descriptives et la « différence moyenne » présentée dans vos analyses inférentielles du test T.

2. Test T pour échantillons indépendants – Incidence du lieu de naissance sur le rendement

Vous avez récemment lu que les élèves au Canada qui sont nés au Canada affichaient de meilleurs résultats aux tests de rendement que ceux nés dans un autres pays. Veuillez noter que ce point porte sur le lieu de naissance (« nés au Canada ») et non nécessairement sur les « élèves canadiens » (nationalité). Vous souhaitez vérifier cette théorie. Veuillez suivre les étapes précédentes pour répondre à la question de recherche suivante : « Les élèves qui ont déclaré être nés au Canada (SQ03) affichent-ils de meilleurs résultats aux tests de rendement en sciences, en mathématiques et en lecture? »

Questions :

- 14.** Veuillez fournir un tableau de statistiques descriptives (moyenne, écart type, erreur standard moyenne, etc.) pour ces nouveaux groupes. Veuillez n'inclure que les statistiques pertinentes.
- 15.** En suivant les étapes précédentes (« Section 1. Test T pour échantillons indépendants – Incidence du sexe sur le rendement »), veuillez expliquer vos résultats dans un **bref** paragraphe (cinq phrases ou moins). Dans ce paragraphe, vous devez a) dire clairement si vos données confirment les trois hypothèses statistiques (il vous faudra peut-être donner séparément une explication pour chacune d'elles), b) répondre à la question de recherche et c) discuter de l'ampleur de l'effet de vos tests inférentiels.
- 16.** SPSS fournit la valeur p pour un test bilatéral. Que serait la valeur p pour un test unilatéral? (Il se peut que vous deviez utiliser Google pour cette question.)

Vous avez maintenant terminé le laboratoire.