

Corrigé du chapitre 15

Comprendre les données sociales

Mark Paradis, Ph. D.

2 septembre 2026

Voici le corrigé détaillé des exercices de *fin de chapitre* du chapitre 15 de *Comprendre les données sociales*. Le corrigé des exercices situés à l'intérieur de chaque section se trouve dans le livre lui-même, immédiatement après les exercices auxquels il répond.

Faites l'exercice avant de lire le corrigé. Un corrigé lu en premier n'est qu'un texte de plus ; un corrigé lu en second est une vérification de votre raisonnement, et seul le second enseigne quelque chose.

Le corrigé complet, pour tous les chapitres, se trouve à <https://www.markpara-dispolitics.com/books/comprendre-les-donnees-sociales/online/solutions/>.

Chapitre 15 Comparaison de moyennes

Corrigé de l'exercice 15.E1. (a) Un test t pour données appariées. Les mêmes personnes sont mesurées deux fois, avant et après, de sorte que chaque participante ou participant fournit deux scores liés et que les deux colonnes ne sont pas indépendantes.

(b) Une différence est le score d'une personne après moins son score avant. Avec 45 personnes, il y a 45 différences, donc $n = 45$ et $df = 44$ — non 90.

(c) $H_0 : \mu_D = 0$, le changement moyen de confiance est nul. $H_1 : \mu_D \neq 0$, il ne l'est pas. Bilatéral, à moins que la direction n'ait été fixée avant la collecte des données ; un programme peut en principe réduire la confiance autant que l'accroître, en montrant aux participantes et participants tout ce qu'ils ignoraient.

(d) Le test établit si le score moyen de confiance a changé de plus que l'erreur d'échantillonnage ne l'expliquerait au cours des douze semaines. Il n'établit pas que le programme a causé ce changement. En l'absence de groupe témoin, tout ce qui s'est produit d'autre pendant ces douze semaines est une explication rivale — l'amélioration du marché du travail, le changement de saison, la simple familiarisation avec le questionnaire. Le chapitre 3 expose ce qu'un groupe de comparaison procure, et c'est exactement cela.

Corrigé de l'exercice 15.E2. (a) La version de *Welch*. Les deux écarts types valent 12,1 et 16,4, ce qui est un écart de dispersion appréciable, et les deux groupes sont de tailles inégales. C'est la combinaison que l'estimation combinée traite mal. En pratique, Welch est le bon choix par défaut quelle que soit l'allure des variances (§15.3.3).

(b) L'erreur type de Welch vaut

$$SE = \sqrt{\frac{12,1^2}{55} + \frac{16,4^2}{48}} = 2,87$$

et la statistique

$$t = \frac{74,3 - 78,9}{2,87} = -1,60$$

sur 85,5 degrés de liberté, ce qui donne $p = 0,113$. À $\alpha = 0,05$, on rejette H_0 : les deux types de classe diffèrent en rendement moyen.

(c) Le d de Cohen vaut $(74,3 - 78,9)/14,26 = -0,32$, un effet petit à moyen. Le dénominateur est l'écart type *combiné*, et non l'erreur type non combinée qu'utilise le test de Welch : une taille d'effet est une différence mesurée dans les unités où les gens varient. L'intervalle à 95 pour cent pour $\mu_1 - \mu_2$ va de $-10,32$ à $1,12$ points. Remarquez qu'il exclut zéro, ce qui est le même verdict que celui du test.

(d) Cela coûte entièrement la conclusion causale. Les élèves n'ont pas été

assignés aux types de classe, de sorte que les deux groupes peuvent différer par le rendement antérieur, par les écoles qui ont adopté la technologie, par les personnes enseignantes qui se sont portées volontaires pour l'employer, et par tout ce qui a conduit des familles vers une classe plutôt que l'autre. Le test établit une différence entre deux groupes d'élèves ; il ne dit rien de ce que fait la technologie. Comparez la §15.3.4, où la même arithmétique n'autorise une lecture causale que parce que l'assignation était aléatoire.

Corrigé de l'exercice 15.E3. (a) **Étape 1.** $H_0 : \mu = 3,50$; $H_1 : \mu \neq 3,50$, bilatéral.

Étape 2. Un échantillon aléatoire est supposé ; les observations sont indépendantes ; $n = 400$ est très au-dessus de tout seuil à partir duquel le théorème central limite serait douteux. σ est inconnu, donc le test est un test t .

Étape 3. $\alpha = 0,05$, bilatéral, $df = 399$.

Étape 4.

$$SE = \frac{0,91}{\sqrt{400}} = 0,045, \quad t = \frac{3,62 - 3,50}{0,045} = 2,64$$

Étape 5. $p = 0,009$, qui est sous 0,05. On rejette H_0 .

Étape 6. La santé autoévaluée moyenne dans cet échantillon diffère significativement de la moyenne nationale.

(b) $d = (3,62 - 3,50)/0,91 = 0,13$.

(c) Il est détectable, et il est très petit. La différence vaut 0,12 point sur une échelle en cinq points, et $d = 0,13$ est sous le seuil conventionnel d'un petit effet. Avec $n = 400$, l'étude a assez de précision pour distinguer de zéro une différence aussi petite, ce qui est ce que rapporte la valeur p significative — et rien de plus.

Un titre de journal rapporterait la valeur p , parce que « différence significative de santé constatée » est une phrase et que « les deux groupes diffèrent d'environ un dixième de point » n'en est pas une. C'est pourtant la seconde qui décrit le constat. C'est là l'écart entre signification statistique et signification pratique, et il se creuse à mesure que les échantillons grandissent.

Corrigé de l'exercice 15.E4. (a) Parce qu'un test d'hypothèse est une règle de décision appliquée à un échantillon, et qu'un échantillon n'est pas la population. La règle est construite pour se tromper une proportion contrôlée du temps : au plus α du temps quand H_0 est vraie et — c'est la part dont on parle le moins — aussi souvent que la puissance du devis le permet quand H_0 est fausse. Rien dans la démarche ne garantit une réponse juste à une occasion donnée. À la §15.2.6, l'hypothèse nulle est fausse à chaque fois, et le test échoue tout de même à la rejeter dans 32 % des tirages.

(b) Ni l'une ni l'autre, nécessairement. Deux études correctement menées sur la même hypothèse nulle fausse divergeront chaque fois qu'un tirage tombe dans

la région de rejet et que l'autre n'y tombe pas, ce qui, à cette puissance, est un événement courant plutôt qu'étrange. Chercher l'erreur est la mauvaise réaction ; regarder les deux intervalles, et la taille de chaque étude, est la bonne.

(c) Décider la taille de l'échantillon à l'avance, à partir du plus petit effet qui importerait et de la puissance requise pour le détecter — une analyse de puissance. Cela ne peut pas se corriger après coup, pour deux raisons. Une étude qui a déjà produit un résultat non significatif ne peut pas distinguer « aucun effet » de « trop petit pour être vu », parce que c'est précisément ce qu'elle a échoué à mesurer. Et ajouter des cas jusqu'à ce qu'un résultat devienne significatif, puis s'arrêter, change le taux d'erreur du test : la valeur p rapportée ne veut plus dire ce qu'elle dit.

Corrigé de l'exercice 15.E5. (a) Le second. Il rapporte $d = 0,45$ avec un intervalle allant de $-0,20$ à $1,10$, de sorte qu'une lectrice connaît l'ampleur estimée de l'effet et l'étendue des valeurs compatibles avec les données. Le premier rapporte seulement qu'une différence était détectable dans un échantillon de 200, ce qui est un fait sur l'étude.

(b) Non. L'intervalle de la seconde étude contient largement l'ampleur d'effet plausible de la première, et avec $df = 38$ il est assez large pour contenir à peu près n'importe quoi de modeste. Deux résultats sont en contradiction quand leurs intervalles ne se chevauchent pas, non quand l'un franchit un seuil de signification et que l'autre ne le franchit pas.

(c) Quelque chose de cette forme :

Les scores de bien-être étaient plus élevés dans le groupe expérimental ($M = \dots$, $SD = \dots$) que dans le groupe témoin ($M = \dots$, $SD = \dots$), $t(198) = 2,14$, $p = 0,03$, $d = \dots$, IC à 95 % [...].

Les deux moyennes de groupe et les deux écarts types, la taille d'effet et l'intervalle. Tout ce dont une lectrice a besoin pour juger de l'ampleur du constat et pour le combiner à celui de quelqu'un d'autre.

Corrigé de l'exercice 15.E6. Les réponses varieront selon l'article choisi. Une réponse complète travaille les cinq questions de la §15.5 sur le texte même de l'article, et cite l'article là où il avance une affirmation.

Trois choses valent la peine d'être sues avant de commencer. Les articles décrivent fréquemment un devis dans la section méthodologique et rapportent un test qui n'y correspond pas, et le décalage se voit habituellement aux seuls degrés de liberté : un devis apparié sur 30 personnes donne $df = 29$, et $df = 58$ signifie que l'appariement a été ignoré. Les tailles d'effet sont rapportées bien plus souvent en psychologie qu'en économie ou en science politique, de sorte que la réponse à (c) peut en dire davantage sur la discipline que sur les auteurs. Et un article qui

rapporte une seule comparaison significative dit rarement combien en ont été faites — la réponse honnête à (d) est souvent « l'article ne le dit pas », ce qui est en soi un constat sur l'article.

Une réponse défendable à (e) peut être oui comme non. Ce qui la rend défendable, c'est qu'elle nomme laquelle des quatre réponses précédentes l'a emportée.