

STATISTICS FOR BEHAVIORAL SCIENCES

by GRAVETTER AND WALLNAU · Language: FR

Summary by booksummaryai.com —

<https://booksummaryai.com/country/fr/statistics-for-behavioral-sciences-resume>

Summary

En 60 secondes

Voici un parcours pas à pas de la manière dont les chercheurs en sciences du comportement transforment des chiffres chaotiques en affirmations défendables. On part des populations, des échantillons et de l'écart inévitable entre eux, appelé erreur d'échantillonnage, puis on se dote d'une boîte à outils : distributions de fréquences, tendance centrale, variabilité et scores z . Les probabilités et la distribution des moyennes échantillonnales préparent les tests d'hypothèses, où résident les erreurs de type I et de type II ainsi que la puissance statistique. De là, on avance à travers les tests t , l'ANOVA, la corrélation, la régression, le khi-deux et le test binomial, pour finir avec l'estimation et la machinerie de calcul qui relie le tout.

Pourquoi le lire

Si vous devez un jour lire un article de recherche, mener une étude ou débattre d'un sondage, ce livre vous en donne la grammaire. Il est écrit pour les étudiants en psychologie et en sciences du comportement qui doivent calculer, interpréter et critiquer des statistiques, et pas seulement les reconnaître. Chaque chapitre s'appuie directement sur le précédent, de sorte qu'à la fin vous pouvez suivre une logique unique depuis la moyenne d'un échantillon jusqu'à un intervalle de confiance ou un rapport F .

Les idées clés

Population, échantillon et l'écart qui ne se referme jamais

Un psychologue veut connaître le niveau moyen d'anxiété de tous les étudiants américains. Mesurer chacun d'entre eux est impossible, si bien que l'échantillon se réduit à 200 étudiants d'une seule université. Le psychologue calcule une statistique à partir de ces 200 et espère qu'elle décrive un paramètre de l'ensemble de la population. Le chapitre distingue la statistique descriptive (« procédures statistiques utilisées pour résumer, organiser et simplifier les données ») de la statistique inférentielle (« techniques qui nous permettent d'étudier des échantillons puis de généraliser au sujet des populations dont ils ont été extraits »), les deux grandes catégories de techniques statistiques. Entre la statistique de l'échantillon et le paramètre de la population se glisse l'erreur d'échantillonnage, un écart qui survient naturellement et qu'aucun plan d'expérience habile ne peut éliminer.

Toute conclusion de recherche est une traduction d'un petit groupe vers un grand, et toute traduction perd quelque chose. L'erreur d'échantillonnage n'est pas une faute que vous auriez commise ; c'est le prix à payer pour n'observer qu'une partie d'une population. Une fois que vous acceptez que cet écart est permanent, le reste du livre devient limpide : chaque technique est une manière différente de mesurer, d'encadrer ou de vivre avec cet écart.

À essayer: Prenez n'importe quel chiffre marquant lu cette semaine dans l'actualité et notez quelle population il prétend décrire, qui a été réellement échantillonné, et ce qui pourrait différer entre les deux.

Mettre de l'ordre dans le tas : les distributions de fréquences

Avant tout test ou théorie, vous avez une pile de scores. Le livre montre le premier geste : une tabulation organisée du nombre d'individus situés dans chaque catégorie sur l'échelle de mesure. De ce tableau, la proportion se déduit immédiatement comme f divisé par N , et le pourcentage n'est que cette proportion multipliée par 100.

Une distribution de fréquences est le pont entre les données brutes et toute statistique qui suit. Les proportions et les pourcentages ne sont pas un ornement ; ils constituent le premier résumé honnête de qui s'est retrouvé où. Se familiariser tôt avec f/N fait que les notions ultérieures de probabilités et d'effectifs attendus semblent familières plutôt qu'étrangères.

À essayer: Comptez cette semaine une catégorie de vos propres données, courriels par expéditeur ou pas par jour, et calculez la proportion de votre catégorie la plus fréquente.

La moyenne est un point d'équilibre

Trois prétendants se disputent la représentation d'une distribution. La moyenne est la somme des scores divisée par leur nombre, notée $M = \Sigma X/n$ pour un échantillon et $\mu = \Sigma X/N$ pour une population. Le livre affirme : « La moyenne est le point d'équilibre de la distribution », car la distance totale sous la moyenne égale la distance totale au-dessus d'elle. La médiane occupe le point milieu, et le mode se contente de désigner le score le plus fréquent. Chaque mesure répond à une question légèrement différente sur la position du centre.

Penser la moyenne comme le pivot d'une balançoire explique pourquoi un score extrême l'entraîne tandis que la médiane reste immobile. L'image du point d'équilibre annonce aussi la logique des écarts, qui reviendront plus tard comme matière première de la variance. Choisir entre moyenne, médiane et mode, c'est choisir quel type de « typique » vous voulez revendiquer.

À essayer: Calculez la moyenne et la médiane de vos dix derniers achats et observez lequel des deux le plus gros achat fait bouger.

La variabilité : à quel point la foule est-elle dispersée ?

Deux classes peuvent afficher la même moyenne et rester des endroits radicalement différents où s'asseoir. L'étendue, la différence entre le score le plus grand et le score le plus petit, est

signalée comme une mesure grossière car elle n'écoute que les extrêmes. Le vrai travail, c'est la variance, la moyenne des écarts élevés au carré : $\sigma^2 = \Sigma(X-\mu)^2/N$ pour une population, et pour un échantillon $s^2 = SS/(n-1)$. La formule de calcul de SS est $\Sigma X^2 - (\Sigma X)^2/n$. L'écart-type est la racine carrée de la variance, ramenant la réponse dans les unités des scores.

Diviser par $n-1$ au lieu de n corrige le biais qu'il y a à utiliser des données d'échantillon pour estimer la variabilité de la population. Les écarts au carré semblent abstraits jusqu'à ce que vous en preniez la racine carrée et retombiez dans des unités que vous pouvez vous représenter.

À essayer: Calculez SS avec la formule de calcul pour cinq petits nombres et vérifiez que le résultat correspond à la somme des écarts au carré obtenue par la méthode longue.

Les scores z placent chaque score sur une même carte

Un étudiant obtient 76 à un examen de psychologie ; un autre obtient 82 à un examen de biologie. Quelle est la meilleure performance ? Les chiffres bruts ne peuvent pas le dire, car les examens ont des moyennes et des dispenses différentes. Le livre convertit chaque score avec $z = (X - \mu)/\sigma$, ou $z = (X - M)/s$ pour un échantillon : le signe indique la direction par rapport à la moyenne, le nombre indique la distance en écarts-types. Transformer les valeurs X en scores z crée une distribution standardisée de moyenne 0 et d'écart-type 1, et la forme de la distribution reste inchangée. Le chapitre utilise l'exemple de la comparaison des scores de deux étudiants à des examens différents — un étudiant obtenant 76 à un examen de psychologie et un autre 82 à un examen de biologie — pour montrer comment les scores z permettent de comparer des scores issus de distributions différentes.

Un score z est une adresse : même forme, nouvelles coordonnées. C'est ce qui rend légitimes les comparaisons entre distributions, car les deux scores parlent désormais la langue des écarts-types.

À essayer: Trouvez la moyenne et l'écart-type de deux échelles qui vous tiennent à cœur, comme vos heures de sommeil et votre temps d'écran, et convertissez une valeur récente de chacune en score z .

La probabilité comme proportion de la courbe

La probabilité reçoit une définition simple : une fraction ou une proportion de tous les résultats possibles, calculée comme le nombre de résultats classés comme A divisé par le nombre total de résultats possibles. Pour une distribution normale, cette proportion est littéralement l'aire sous la courbe. La table de la loi normale centrée réduite devient l'outil de base, listant les proportions correspondant à chaque position de score z , si bien que vous pouvez trouver la probabilité d'un score au-dessus, en dessous ou entre des valeurs z données.

C'est le chapitre où la position devient vraisemblance. Comme les scores z du chapitre 5 standardisent toute distribution normale, une seule table les sert toutes.

À essayer: Utilisez une table z pour trouver la proportion de scores au-dessus d'un écart-type au-dessus de la moyenne, et vérifiez-la avec la règle du 68-95 que vous connaissez peut-être déjà.

La distribution des moyennes échantillonnales et le théorème central limite

Prenez tous les échantillons aléatoires possibles de taille n dans une population et recueillez toutes leurs moyennes : cet ensemble est la distribution des moyennes échantillonnales, et son centre se situe exactement à la moyenne de la population, $\mu_M = \mu$. Le théorème central limite livre ensuite le gain : pour toute population de moyenne μ et d'écart-type σ , la distribution des moyennes échantillonnales a pour moyenne μ , pour écart-type σ/n , et tend vers la normale à mesure que n croît ; avec $n \geq 30$ elle est presque parfaitement normale. Cet écart-type reçoit son propre nom, l'erreur type de la moyenne, et il mesure la distance type entre une moyenne d'échantillon et la moyenne de la population. Des échantillons plus grands la réduisent, ce qui est la loi des grands nombres en action.

C'est la charnière de tout le livre : elle explique pourquoi une seule moyenne d'échantillon peut dire quelque chose de fiable sur une population. L'erreur type transforme une inquiétude abstraite — de combien mon échantillon peut-il dévier ? — en un nombre calculable.

À essayer: Simulez 30 échantillons de taille 5 et de taille 30 à partir d'un jeu de données quelconque, tracez les deux ensembles de moyennes, et observez la dispersion s'effondrer à mesure que n croît.

Le test d'hypothèse : quatre étapes, deux façons de se tromper

Le chapitre introduit les quatre étapes du test d'hypothèse : (1) énoncer les hypothèses nulle et alternative, (2) fixer les critères de décision (niveau alpha, généralement .05 ou .01), (3) calculer la statistique de test, et (4) prendre la décision de rejeter ou de ne pas rejeter l'hypothèse nulle (rejeter H_0 si la statistique de test se trouve dans la région critique). L'hypothèse nulle (H_0) est définie comme « un énoncé sur la population qui prédit l'absence d'effet ou l'absence de différence » — « l'hypothèse qui affirme qu'il n'y a aucun changement, aucun effet, aucune différence » — et l'hypothèse alternative (H_1) prédit qu'il existe un effet ou une différence. Pour rendre les enjeux humains, le livre met en scène un procès où l'hypothèse nulle est que l'accusé est innocent : condamner un innocent est une erreur de type I, acquitter un coupable est une erreur de type II. Le niveau alpha (niveau de signification), généralement $\alpha = .05$, est défini comme « une valeur de probabilité utilisée pour définir la notion de « très improbable » dans un test d'hypothèse ». Il fixe le risque d'erreur de type I et délimite les frontières de la région critique pour le rejet de l'hypothèse nulle. Bêta est le risque de la seconde, et la puissance, la probabilité de rejeter correctement une hypothèse nulle fautive, vaut $1 - \beta$ et croît avec des échantillons plus grands et des effets plus importants.

L'analogie du procès rend l'asymétrie inoubliable : le système présume l'innocence, si bien qu'une fausse alerte et un manquement sont des fautes différentes aux coûts différents.

À essayer: Pour votre prochaine décision, écrivez l'hypothèse nulle en mots simples et nommez ce qu'une erreur de type I et une erreur de type II vous coûteraient chacune.

Quand σ est inconnu : la statistique t

La recherche réelle vous livre rarement l'écart-type de la population, si bien que l'ingrédient clé du test z manque. La solution consiste à estimer σ à partir du s de l'échantillon lui-même, ce qui donne une erreur-type estimée $s_M = s/n$ et la statistique $t = (M - \mu)/s_M$. Parce que vous avez estimé, vous payez en incertitude : la distribution t est une approximation de la normale, plus aplatie et plus variable, et sa forme exacte est fixée par les degrés de liberté, $ddl = n - 1$.

La statistique t est de l'honnêteté intégrée dans une formule : substituer une estimation à une valeur connue élargit la distribution pour admettre l'incertitude supplémentaire.

À essayer: Cherchez le t critique pour $ddl = 5$ et $ddl = 30$ à $\alpha = .05$ et comparez chacun à 1,96 pour voir le prix des petits échantillons.

Deux échantillons indépendants et variance groupée

La question change de forme : deux groupes distincts, deux moyennes, une comparaison. La statistique t pour mesures indépendantes divise la différence $M_1 - M_2$ par une erreur-type estimée qui combine l'erreur des deux échantillons, et l'hypothèse nulle s'écrit $H_0 : \mu_1 - \mu_2 = 0$. Quand les deux échantillons diffèrent par leur taille, le livre introduit la variance groupée, une moyenne pondérée des deux variances d'échantillon : $s^2_p = (SS1 + SS2)/(ddl1 + ddl2)$.

Combiner l'erreur des deux échantillons reconnaît que chaque moyenne porte sa propre incertitude, et l'hypothèse nulle d'une différence nulle donne au test une cible précise. Ce plan prépare aussi le contraste avec les mesures répétées qui vient ensuite.

À essayer: Prenez deux petites listes de nombres de longueurs différentes, calculez chaque SS, groupez les variances et vérifiez que la liste la plus longue domine l'estimation.

Les mêmes personnes, deux fois : le t pour mesures répétées

Au lieu de deux foules, imaginez une seule foule mesurée sous les deux traitements. Le t pour mesures répétées travaille sur des scores de différence, $D = X_2 - X_1$, en testant $H_0 : \mu_D = 0$ avec $t = (MD - \mu_D)/s_{MD}$ et $ddl = n - 1$, où n compte les paires. L'avantage caché du plan est que les différences individuelles s'annulent : comme les mêmes sujets apparaissent dans les deux traitements, l'analyse élimine les problèmes créés par les différences individuelles. Moins de bruit dû à la variation d'une personne à l'autre signifie qu'un véritable effet du traitement est plus facile à détecter.

Chaque participant est son propre témoin, si bien les particularités qui rendent une personne élevée sur les deux mesures se soustraient des scores de différence. Cette réduction de variance est une puissance gratuite, et c'est pourquoi les plans à mesures répétées séduisent les chercheurs aux petits échantillons.

À essayer: Suivez une mesure personnelle avant et après un petit changement cette semaine, calculez vos scores de différence et voyez à quel point ils sont plus stables que les valeurs brutes.

ANOVA et rapport F : comparer plusieurs moyennes à la fois

Trois traitements ou plus rendent les tests t par paires maladroits, si bien l'analyse de variance entre en scène, utilisant les variances d'échantillon comme base de son analyse pour tester simultanément les différences de moyennes entre les groupes. La machinerie repose sur les sommes de carrés : $SS_{total} = \sum X^2 - (G^2/N)$ avec G le total général, SS_{within} comme la somme des SS à l'intérieur de chaque traitement, $SS_{between} = SS_{total} - SS_{within}$, et les degrés de liberté $ddl_{between} = k - 1$, $ddl_{within} = N - k$, avec $MS = SS/ddl$. Le rapport F est le rapport de la variance entre traitements à la variance à l'intérieur des traitements, $F = MS_{between}$. Un rapport F proche de 1,00 n'indique aucun effet du traitement.

L'ANOVA recadre la question : au lieu de demander directement si les moyennes diffèrent, elle demande si la variance entre les groupes dépasse ce que la variance à l'intérieur des groupes laisserait prévoir par hasard. La base de 1,00 du rapport F vous offre une vérification de bon sens instantanée avant toute consultation de table.

À essayer: Divisez n'importe quel jeu de données en trois groupes, calculez la variance des moyennes de groupe par rapport à la variance intra-groupe moyenne, et estimez à l'œil votre propre rapport F.

ANOVA à mesures répétées et deux facteurs à la fois

Le plan avec les mêmes participants change d'échelle : dans l'ANOVA à mesures répétées, les différences individuelles sont mesurées puis retirées du dénominateur du rapport F, ce qui affine le test. Ensuite, le plan gagne une seconde variable indépendante, et l'ANOVA à deux facteurs produit trois rapports F distincts : un effet principal du facteur A, un effet principal du facteur B, et l'interaction entre eux. Une interaction se produit lorsque les effets d'un facteur dépendent des différents niveaux d'un second facteur, et le livre recommande de tracer un graphique des moyennes de cellules pour la voir.

Retirer les différences individuelles du dénominateur, c'est la version ANOVA de l'avantage du t à mesures répétées, appliqué à plusieurs conditions. L'interaction est souvent la véritable découverte, car les effets principaux peuvent masquer des histoires où un traitement aide à un niveau et nuit à un autre.

À essayer: Esquissez deux lignes sur un graphique, une pour chaque niveau du facteur B à travers les niveaux du facteur A, et décidez par vous-même si elles racontent une interaction.

Corrélation, régression et la droite qui s'ajuste

Deux variables, un nuage de points : la corrélation de Pearson mesure la direction et le degré de la relation linéaire, allant de -1,00 pour une corrélation négative parfaite à +1,00 pour une corrélation positive parfaite, avec $r = SP/(SSX \cdot SSY)$. Son carré donne le coefficient de

détermination, r^2 , la proportion de variance d'une variable expliquée par sa relation avec l'autre. Pour des données ordinales, ou des tendances monotones mais non linéaires, la corrélation de rang de Spearman prend le relais. Le livre est sans détour sur les limites : la corrélation ne démontre pas la causalité, et un problème de troisième variable ou une relation de cause à effet inversée peuvent expliquer ce que vous observez. La régression construit alors la droite de prédiction $\hat{Y} = bX + a$ par la méthode des moindres carrés, avec une pente $b = SP/SSX$ et une ordonnée à l'origine $a = MY - bMX$, et l'erreur standard de l'estimation mesure la distance standard entre la droite et les points de données réels, tombant à zéro seulement lorsque $r = \pm 1,00$.

Corrélation et régression sont deux visions d'une même relation : r décrit sa force et sa direction, tandis que l'équation de régression la transforme en prédictions.

À essayer: Choisissez deux variables que vous pouvez consigner chaque jour pendant une semaine, calculez une corrélation approximative, puis listez une variable tierce qui pourrait expliquer le lien avant d'y croire.

Khi-deux et binomiale : tester des catégories

Toute donnée n'est pas un nombre sur une échelle ; certaines ne sont que des catégories. La statistique du khi-deux, $\chi^2 = \sum(fO - fE)^2/fE$, compare les fréquences observées aux fréquences attendues pour des données nominales. Le test d'ajustement demande si les proportions de l'échantillon correspondent à une distribution de population spécifiée, avec ddl = $k - 1$ sur k catégories. Le test d'indépendance demande si deux variables catégorielles sont liées, en travaillant à partir d'une matrice de fréquences observées avec ddl = $(R - 1)(C - 1)$ et des fréquences attendues par cellule $f_e = (\text{total de ligne} \times \text{total de colonne})/N$. Pour le cas particulier d'exactement deux catégories, le test binomial s'applique, et l'exemple du chapitre est celui d'un chercheur lançant une pièce de monnaie de nombreuses fois pour voir si les faces observées correspondent à l'attente sous $P = Q = 0,50$.

Ces tests vous libèrent entièrement des moyennes et des variances : les données sont des décomptes, et la question est de savoir si les décomptes correspondent à une affirmation.

À essayer: Lancez une pièce de monnaie 20 fois, calculez le nombre attendu de faces sous $P = 0,50$, et calculez de combien votre nombre observé s'écarte.

Les meilleures citations

- "La moyenne est le point d'équilibre de la distribution" - Gravetter et Wallnau
- "l'écart, ou l'erreur, qui existe naturellement entre une statistique d'échantillon et le paramètre de population correspondant" - Gravetter et Wallnau
- "l'hypothèse qui énonce qu'il n'y a aucun changement, aucun effet ou aucune différence" - Gravetter et Wallnau

- "lorsque les effets d'un facteur dépendent des différents niveaux d'un second facteur" - Gravetter et Wallnau

Limites

La douceur du livre est aussi son plafond : il s'appuie sur des procédures détaillées et des formules plutôt que sur le raisonnement mathématique plus profond qui les sous-tend, si bien que les lecteurs en quête de démonstrations ou de fondements théoriques de la mesure devront chercher ailleurs. Les tailles d'effet et les préoccupations modernes comme la réplication, les intervalles de confiance au-delà du chapitre sur l'estimation, ou les flux de travail logiciels occupent une place limitée face au calcul manuel. Les éditions varient aussi dans leur traitement des critiques du NHST. Considérez-le comme une première passe qui vous rend capable de calculer et d'interpréter, non comme le dernier mot sur la pratique statistique.

En une phrase

Les statistiques forment une chaîne qui va de l'échantillon à la population, et chaque maillon, de la moyenne au rapport F, existe pour mesurer à quel point votre échantillon risque de vous induire en erreur.

Pour qui

Les étudiants en psychologie et en sciences du comportement qui doivent valider un cours de statistiques sans perdre leur raison, en particulier ceux qui craignent les formules plus que l'échec. Idéal aussi pour les chercheurs et les cliniciens qui mènent des études mais veulent enfin comprendre ce que leur logiciel fait réellement sous le capot.