

Tutoriel sur les statistiques

Explorable.com 21.4K reads

Ce tutoriel sur les statistiques est un guide dont l'objectif est de vous aider à comprendre les concepts clés des statistiques et comment ces concepts se rattachent à la méthode scientifique et à la recherche.

Les scientifiques utilisent fréquemment les statistiques pour analyser leurs résultats. Pourquoi les chercheurs utilisent des statistiques? Les statistiques peuvent aider à comprendre un phénomène en confirmant ou en rejetant une hypothèse. Elles sont essentielles pour acquérir des connaissances dans la plupart des théories scientifiques.

Cependant, nul besoin d'être un scientifique; tous ceux qui souhaitent apprendre de quelle façon les statistiques peuvent aider les chercheurs peuvent lire ce tutoriel sur les statistiques pour la méthode scientifique.

Que sont les statistiques?

 **EXPLORABLE**
Quiz Time!



Quiz:
Psychology 101 Part 2



Quiz:
Psychology 101 Part 2



Quiz:
Flags in Europe

[See all quizzes =>](#)

Données de recherche

Cette section du tutoriel explique comment les données sont acquises et utilisées.

Les résultats d'une enquête scientifique contiennent souvent plus d'informations que ce dont le chercheur a besoin. Ces informations ou données sont appelées données brutes.

Pour être en mesure d'analyser les données judicieusement, les données brutes sont traitées ^[1] en "données de sortie" ^[2]. Il existe de nombreuses méthodes pour traiter les données mais essentiellement, le scientifique organise et résume les données brutes dans un bloc pour plus de clarté. Tout type d'information organisée peut être appelé "ensemble de données" ^[3].

Ensuite, les chercheurs peuvent appliquer diverses méthodes statistiques pour analyser et comprendre les données plus facilement et plus précisément. En fonction de la recherche, le chercheur peut aussi vouloir utiliser des statistiques descriptives ^[4] comme dans la recherche exploratoire.

Ce qui est formidable avec les données brutes, c'est que vous pouvez revenir en arrière et vérifier certaines choses si vous soupçonnez que quelque chose se produit différemment de ce que vous aviez prévu. On fait cela après avoir analysé la signification des résultats.

Comme vous voyez plus clairement ce qui se passe, les données brutes peuvent vous inspirer pour de nouvelles hypothèses. Vous pouvez aussi contrôler les variables qui pourraient influencer la conclusion ^[5] (par exemple les variables tierces ^[6]). Dans les statistiques, un paramètre ^[7] est toute valeur numérique caractérisant une population donnée ou un de ses aspects.

Tendance centrale et distribution normale

Cette partie du tutoriel vous aidera à comprendre la distribution, la tendance centrale et comment elles se rapportent aux ensembles de données ^[3].

La plus grande part des données du monde réel sont distribuées normalement ^[8]. La distribution normale est une courbe de totaux ou une distribution des fréquences ^[9] dans laquelle le nombre le plus fréquent se situe près du milieu. De nombreuses expériences se basent sur des conjectures d'une distribution normale ^[10]. C'est une raison pour laquelle les chercheurs mesurent très souvent la tendance centrale ^[11] dans la recherche statistique, comme la moyenne ^[12] (moyenne arithmétique ^[13] ou moyenne géométrique ^[14]), la médiane ^[15] ou le mode ^[16].

La tendance centrale peut donner une assez bonne idée de la nature des données (la moyenne, la médiane et le mode affichent la "valeur moyenne"), en particulier lorsqu'elle est combinée avec une mesure de la distribution des données. Pour mesurer cette distribution, les scientifiques calculent l'écart type ^[17]; mais Il existe d'autres méthodes: la variance ^[18], l'erreur type de la moyenne ^[19], l'erreur type de l'estimation ou la "plage" ^[20] (qui indique les extrêmes dans les données).

Pour créer le graphique de la distribution normale de quelque chose, vous utiliserez normalement la moyenne arithmétique ^[13] d'un "échantillon suffisamment grand" ^[21] et vous devrez calculer l'écart type.

Cependant, la distribution de l'échantillonnage ^[22] ne ??sera pas distribuée normalement si la distribution est décalée (naturellement) ou a des valeurs isolées ^[23] (résultats le plus souvent

rare ou erreurs de mesure) qui faussent les données. La distribution-F [24] est un exemple de distribution qui n'est pas distribuée normalement; elle est décalée vers la droite.

Ainsi, les chercheurs vérifient souvent que leurs résultats sont distribués normalement en utilisant la plage, la médiane et le mode. Si la distribution n'est pas distribuée normalement, cela influencera le choix de la méthode ou du test statistique pour l'analyse.

Autres outils

- Quartile [25]
- Trimean [26]

Test des hypothèses - Tutoriel sur les statistiques

Comment savons-nous si une hypothèse est correcte ou non?

Pourquoi utiliser les statistiques pour déterminer cela? [27]

Utiliser des statistiques dans la recherche comporte beaucoup plus que l'utilisation de formules statistiques ou que la familiarisation avec un logiciel de statistiques.

Utiliser des statistiques dans la recherche implique essentiellement

1. Apprendre les statistiques de base [28]
2. Comprendre la relation entre probabilité et statistiques [29]
3. Comprendre les deux branches majeures de la statistique [30]: la statistique descriptive [4] et la statistique inférentielle [31].
4. Comprendre comment les statistiques se rapportent à la méthode scientifique [27].

Les statistiques ne sont pas que des formules et des calculs. (Beaucoup de fausses déductions ont été formulées à cause d'une mauvaise compréhension des concepts statistiques de base)

L'inférence statistique nous aide à tirer des conclusions [5] à partir d'échantillons d'une population.

Quand on mène des expériences [32], un élément essentiel est de tester les hypothèses [33] les unes contre les autres. C'est pourquoi cela constitue une part considérable du tutoriel.

On teste des hypothèses en formulant une hypothèse alternative [34] qui est testée contre l'hypothèse nulle [35], le point de vue commun. Les hypothèses sont testées statistiquement [36] les unes contre les autres.

Le chercheur peut calculer un intervalle de confiance [37] qui définit les limites quand vous considérez qu'un résultat soutient l'hypothèse nulle et quand l'hypothèse de recherche alternative [34] est soutenue.

Cela signifie que les différences entre le groupe expérimental et le groupe témoin [38] ne peuvent pas toutes être acceptées comme soutenant l'hypothèse alternative - le résultat doit varier statistiquement de manière significative [39] pour que le chercheur accepte l'hypothèse alternative. On fait cela en utilisant un test de signification [40] (un autre article [41]).

Attention cependant, le dragage [42], l'espionnage ou la recherche passive des données sans test ultérieur de votre hypothèse dans une expérience contrôlée peut vous amener à trancher à propos de la causalité [43] même s'il n'y a aucun lien avec la vérité [44].

Selon l'hypothèse, vous devrez choisir entre les tests unilatéraux et bilatéraux.

Parfois, le groupe témoin est remplacé par une probabilité expérimentale [45], c'est souvent le cas si la recherche traite un phénomène qui pose un problème d'éthique [46], qui est trop coûteux ou qui prend trop de temps; alors le modèle expérimental véritable [47] est remplacé par une approche quasi-expérimentale [48].

Souvent, il y a un biais de publication lorsque le chercheur confirme que l'hypothèse alternative est correcte au lieu d'obtenir un "résultat nul", en concluant que l'hypothèse nulle fournit la meilleure explication.

Si elles sont appliquées correctement, les statistiques peuvent être utilisées pour comprendre la causalité entre les variables de recherche [49].

Elles peuvent aider à identifier les variables tierces [6], et aussi à les manipuler et à les dissimuler si la personne qui présente les chiffres n'a pas des intentions honnêtes (ou un savoir suffisant) vis-à-vis de ses résultats.

La mauvaise utilisation des statistiques est un phénomène courant qui perdurera probablement tant que les gens ont l'intention d'influencer les autres. Un bon traitement statistique [50] des données expérimentales peut ainsi permettre d'éviter une utilisation des statistiques contraire à l'éthique. La philosophie des statistiques [51] implique de rendre des comptes quand à la bonne utilisation des statistiques, sur la validité statistique [52] et l'éthique dans les statistiques [53].

Voici un autre tutoriel sur les statistiques de qualité qui intègre les statistiques et la méthode scientifique.

Fiabilité et erreur expérimentale

Les tests statistiques utilisent des données provenant d'échantillons. Ces résultats sont ensuite généralisés [54] à la population globale. Comment pouvons-nous savoir si ils reflètent la bonne conclusion [55]?

Contrairement à ce que certains pourraient croire, les erreurs dans la recherche [55] sont une part essentielle du test de signification [41]. Ironiquement, la possibilité d'une erreur est ce qui fait que la recherche est scientifique en premier lieu. Si une hypothèse ne peut pas être falsifiée [56] (par exemple, l'hypothèse a une logique circulaire), elle n'est pas testable [57] et donc non scientifique, par définition.

Si une hypothèse est testable, Soyez ouverts à la possibilité qu'elle se révèle fausse. Statistiquement, cela offre la possibilité d'obtenir des erreurs expérimentales [58] dans vos résultats à cause d'erreurs aléatoires ou d'autres problèmes avec la recherche. Les erreurs expérimentales peuvent également être décomposées en erreurs de type I et erreurs de type II [55]. Les courbes ROC [59] sont utilisées pour calculer la sensibilité entre les vrais positifs et les faux positifs.

Une analyse de la puissance [60] d'un test statistique peut déterminer le nombre d'échantillons dont un test aura besoin pour avoir une valeur-p [61] acceptable afin de rejeter une hypothèse nulle [35] fausse.

La marge d'erreur [62] est liée à l'intervalle de confiance [37] et à la relation entre la signification statistique, la taille des échantillons et les résultats attendus [63]. L'ampleur de l'effet [64] estime l'intensité de la relation entre deux variables dans une population. Il peut aider à déterminer la taille de l'échantillon [21] nécessaire pour généraliser [54] les résultats à l'ensemble de la population.

Reproduire [65] la recherche des autres est également essentiel pour savoir si les résultats peuvent être généralisés ou si ils sont simplement le produit d'une expérience isolée et aléatoire. La réplication peut aider à repérer les erreurs aléatoires [66] et les erreurs systématiques [67] (validité du test [68]).

Le coefficient alpha de Cronbach [69] est utilisé pour mesurer la cohérence interne ou la fiabilité [70] du score d'un test.

Reproduire l'expérience ou la recherche garantit la fiabilité des résultats statistiquement [71].

Si les résultats ont des valeurs isolées, on observe souvent une régression vers la moyenne [72], ce qui fait que le résultat n'est pas statistiquement différent entre le groupe expérimental et le groupe témoin.

Tests statistiques

Ici, nous allons introduire quelques tests et méthodes statistiques utilisés couramment par les chercheurs.

Relation entre les variables

La relation entre les variables [73] est très importante pour les scientifiques. Elle les aide à comprendre la nature de ce qu'ils étudient. Une relation linéaire [74] se produit lorsque deux variables varient proportionnellement, c'est à dire que si une variable augmente, l'autre variable augmentera avec le même ratio. Une relation non linéaire [75] se produit lorsque les variables ne varient pas proportionnellement. La corrélation [76] est un moyen d'exprimer la relation entre deux ensembles de données ou entre deux variables.

Les échelles de mesure [77] sont utilisées pour classer, catégoriser et (le cas échéant) quantifier les variables.

Le coefficient de corrélation de Pearson [78] (ou corrélation produit-moment de Pearson)

exprimera uniquement la relation linéaire entre deux variables. Le coefficient de corrélation des rangs de Spearman [79] est surtout utilisé pour les relations linéaires quand on travaille avec des variables ordinales. Le coefficient de Kendall tau (?) peut être utilisé pour mesurer les relations non linéaires.

La corrélation partielle [80] (et la corrélation multiple) peut être utilisée pour le contrôle d'une variable tierce [6].

Prédictions

L'objectif des prévisions est de comprendre les causes. Corrélation ne signifie pas nécessairement causalité [81]. Avec la régression linéaire, on mesure souvent une variable manipulée [82].

Quelle est la différence entre corrélation et régression linéaire [83]? Essentiellement, une étude corrélationnelle [84] observe les interactions entre les variables [85] tandis qu'une régression linéaire s'intéresse à la ligne la mieux ajustée dans un graphique.

Analyse régressive et autres outils de modélisation

- Régression linéaire [86]
- Régression multiple [87]
- Une analyse de trajectoire est une extension du modèle régressif
- Une analyse factorielle [88] cherche à découvrir les facteurs sous-jacents de quelque chose.
- La méta-analyse [89] utilise fréquemment l'ampleur de l'effet [64]

La probabilité bayésienne [90] prédit la vraisemblance d'événements futurs interactivement, en alternative à la procédure classique.

Tester les hypothèses statistiquement

Le test-t de l'étudiant [91] peut indiquer si l'hypothèse nulle [35] est correcte ou non. Dans la recherche, il est souvent utilisé pour tester les différences entre deux groupes (par exemple, entre un groupe témoin [38] et un groupe expérimental).

Le t-test suppose que les données sont à peu près distribuées normalement et que la variance est égale (ceci peut être testé par le test-F [92]).

Test-t de l'étudiant [93]:

- Test-t indépendant à un échantillon [94]
- Test-t indépendant à deux échantillons [95]
- Test-t dépendant pour les échantillons appariés [96]

Le test du rang signé de Wilcoxon [97] peut être utilisé pour les données non paramétriques.

Un Test-Z [98] est semblable à un test-t mais ne sera généralement pas utilisé sur des échantillons d'une taille inférieure à 30.

Un carré-chi [99] peut être utilisé si les données sont qualitatives plutôt que quantitatives.

Comparer plus de deux groupes

Une analyse de la variance [100] est utilisée quand on souhaite tester s'il existe des variations différentes entre les groupes plutôt que des possibilités différentes. L'analyse de la variance peut également être appliquée sur plus de deux groupes. La distribution-F [24] peut être utilisée pour calculer les valeurs-p [61] dans l'analyse de la variance.

Analyse de la variance

- ANOVA à sens unique [101]
- ANOVA à double sens [102]
- ANOVA factorielle [103]
- Mesures répétées et ANOVA [104]

Statistiques non paramétriques

Quelques méthodes courantes utilisant les statistiques non paramétriques [105]:

- Kappa de Cohen [106]
- Test-U de Mann-Whitney [107]
- Coefficient de corrélation des rangs de Spearman [79]

Autres termes statistiques importants

- Variables discrètes [108]

URL source: <https://explorable.com/fr/tutoriel-sur-les-statistiques>

Liens

- [1] <https://explorable.com/raw-data-processing>
- [2] <https://explorable.com/data-output>
- [3] <https://explorable.com/statistical-data-sets>
- [4] <https://explorable.com/descriptive-statistics>
- [5] <https://explorable.com/drawing-conclusions>
- [6] <https://explorable.com/confounding-variables>
- [7] <https://explorable.com/parameters-and-statistics>
- [8] <https://explorable.com/normal-probability-distribution>
- [9] <https://explorable.com/frequency-distribution>
- [10] <https://explorable.com/normal-distribution-assumptions>
- [11] <https://explorable.com/measures-of-central-tendency>
- [12] <https://explorable.com/statistical-mean>
- [13] <https://explorable.com/arithmetric-mean>
- [14] <https://explorable.com/geometric-mean>
- [15] <https://explorable.com/calculate-median>
- [16] <https://explorable.com/statistical-mode>
- [17] <https://explorable.com/calculate-standard-deviation>
- [18] <https://explorable.com/statistical-variance>
- [19] <https://explorable.com/standard-error-of-the-mean>
- [20] <https://explorable.com/range-in-statistics>
- [21] <https://explorable.com/sample-size>
- [22] <https://explorable.com/sampling-distribution>

[23] <https://explorable.com/statistical-outliers>
[24] <https://explorable.com/f-distribution>
[25] <https://explorable.com/quartile>
[26] <https://explorable.com/trimean>
[27] <https://explorable.com/why-statistics-matter>
[28] <https://explorable.com/learn-statistics>
[29] <https://explorable.com/probability-and-statistics>
[30] <https://explorable.com/branches-of-statistics>
[31] <https://explorable.com/inferential-statistics>
[32] <https://explorable.com/fr/realisation-dune-experience>
[33] <https://explorable.com/hypothesis-testing>
[34] <https://explorable.com/research-hypothesis>
[35] <https://explorable.com/null-hypothesis>
[36] <https://explorable.com/statistical-hypothesis-testing>
[37] <https://explorable.com/statistics-confidence-interval>
[38] <https://explorable.com/scientific-control-group>
[39] <https://explorable.com/statistically-significant-results>
[40] <https://explorable.com/significance-test-2>
[41] <https://explorable.com/significance-test>
[42] <https://explorable.com/data-dredging>
[43] <https://explorable.com/es/causa-y-efecto>
[44] <https://explorable.com/truth-and-theory>
[45] <https://explorable.com/experimental-probability>
[46] <https://explorable.com/fr/ethique-dans-la-recherche>
[47] <https://explorable.com/fr/conception-experimentale-veritable>
[48] <https://explorable.com/fr/le-modele-quasi-experimental>
[49] <https://explorable.com/research-variables>
[50] <https://explorable.com/statistical-treatment-of-data>
[51] <https://explorable.com/philosophy-of-statistics>
[52] <https://explorable.com/statistical-validity>
[53] <https://explorable.com/ethics-in-statistics>
[54] <https://explorable.com/what-is-generalization>
[55] <https://explorable.com/type-i-error>
[56] <https://explorable.com/falsifiability>
[57] <https://explorable.com/testability>
[58] <https://explorable.com/experimental-error>
[59] <https://explorable.com/roc-curve-analysis>
[60] <https://explorable.com/statistical-power-analysis>
[61] <https://explorable.com/p-value>
[62] <https://explorable.com/statistics-margin-of-error>
[63] <https://explorable.com/statistical-significance-sample-size>
[64] <https://explorable.com/effect-size>
[65] <https://explorable.com/replication-study>
[66] <https://explorable.com/random-error>
[67] <https://explorable.com/systematic-error>
[68] <https://explorable.com/test-validity>
[69] <https://explorable.com/cronbachs-alpha>
[70] <https://explorable.com/statistics-reliability>
[71] <https://explorable.com/statistical-reliability>
[72] <https://explorable.com/regression-to-the-mean>
[73] <https://explorable.com/relationship-between-variables>
[74] <https://explorable.com/linear-relationship>
[75] <https://explorable.com/non-linear-relationship>
[76] <https://explorable.com/fr/la-correlation-statistique>
[77] <https://explorable.com/measurement-scales>
[78] <https://explorable.com/pearson-product-moment-correlation>

[79] <https://explorable.com/spearman-rank-correlation-coefficient>
[80] <https://explorable.com/partial-correlation-analysis>
[81] <https://explorable.com/fr/correlation-et-causalite>
[82] <https://explorable.com/independent-variable>
[83] <https://explorable.com/correlation-and-regression>
[84] <https://explorable.com/correlational-study>
[85] <https://explorable.com/variables-and-statistics>
[86] <https://explorable.com/linear-regression-analysis>
[87] <https://explorable.com/multiple-regression-analysis>
[88] <https://explorable.com/factor-analysis>
[89] <https://explorable.com/meta-analysis>
[90] <https://explorable.com/bayesian-probability>
[91] <https://explorable.com/students-t-test>
[92] <https://explorable.com/f-test>
[93] <https://explorable.com/students-t-test-2>
[94] <https://explorable.com/independent-one-sample-t-test>
[95] <https://explorable.com/independent-two-sample-t-test>
[96] <https://explorable.com/dependent-t-test-for-paired-samples>
[97] <https://explorable.com/wilcoxon-signed-rank-test>
[98] <https://explorable.com/z-test>
[99] <https://explorable.com/chi-square-test>
[100] <https://explorable.com/fr/anova-0-fr>
[101] <https://explorable.com/one-way-anova>
[102] <https://explorable.com/two-way-anova>
[103] <https://explorable.com/factorial-anova>
[104] <https://explorable.com/repeated-measures-anova>
[105] <https://explorable.com/nonparametric-statistics>
[106] <https://explorable.com/cohens-kappa>
[107] <https://explorable.com/mann-whitney-u-test>
[108] <https://explorable.com/discrete-variables>