Université Djillali Liabes de Sidi
Bel-Abbes
Faculté des Sciences Exactes
Département de Probabilité-Statistique

Master SA-PA.

Cours: Statistique 1.

Cours de Statistique

Présentée par

Pr. ATTOUCH Mohammed Kadi

Master SA-PA

Module. Statistique 1 & *Semestre*. S1

du 07/02 au 06/03 année 2021

Objectif

La statistique Inférentielle est un terme regroupe les méthodes dont l'objectif principal est de préciser un phénomène sur une population globale, à partir de son observation sur un échantillon de cette population. Ce passage ne se fait que moyennant des hypothèses de type probabiliste.

Plan de l'exposé

Echantillonage

- Echantillonnage d'une moyenne
- Echantillonnage d'une variance
- Echantillonnage d'une proportion

Plan de l'exposé

Echantillonage

- Echantillonnage d'une moyenne
- Echantillonnage d'une variance
- Echantillonnage d'une proportion

Estimation paramétrique

- Notion de vraisemblance

Plan de l’exposé

Echantillonage

- Echantillonnage d’une moyenne
- Echantillonnage d’une variance
- Echantillonnage d’une proportion

Estimation paramétrique

- Notion de vraisemblance

Tests statistique

Plan de l'exposé

Echantillonage

- Echantillonnage d'une moyenne
- Echantillonnage d'une variance
- Echantillonnage d'une proportion

Estimation paramétrique

- Notion de vraisemblance

Tests statistique

Intervalle de Confiance

- Intervalle de confiance d'une moyenne
- Intervalle de confiance d'une variance

Plan de l'exposé

Echantillonage

- Echantillonnage d'une moyenne
- Echantillonnage d'une variance
- Echantillonnage d'une proportion

Estimation paramétrique

- Notion de vraisemblance

Tests statistique

Intervalle de Confiance

- Intervalle de confiance d'une moyenne
- Intervalle de confiance d'une variance

Tables Statistique

Echantillonage

Population - échantillon

On appelle population la totalité des unités de n'importe quel genre prises en considération par le statisticien. Elle peut être finie ou infinie.

Un échantillon est un sous-ensemble de la population étudiée.

Qu'il traite un échantillon ou une population, le statisticien décrit habituellement ces ensembles à l'aide de mesures telles que le nombre d'unités, la moyenne, l'écart-type et le pourcentage.

- Les mesures que l'on utilise pour décrire une population sont des paramètres. Un paramètre est une caractéristique de la population.
- Les mesures que l'on utilise pour décrire un échantillon sont appelées des statistiques. Une statistique est une caractéristique de l'échantillon.

Quelques notations

Afin de ne pas confondre les statistiques et les paramètres, on utilise des notations différentes, comme le présente le tableau récapitulatif suivant.

	POPULATION	ÉCHANTILLON
DÉFINITION	C'est l'ensemble des unités considérées par le statisticien.	C'est un sous-ensemble de la population choisie pour étude.
CARACTÉRISTIQUES	Ce sont les paramètres	Ce sont les statistiques
NOTATIONS	N = taille de la population (si elle est finie)	n = taille de l'échantillon
Si on étudie un caractère quantitatif	moyenne de la population $m = \frac{1}{N}\sum_{i=1}^{N} x_i$ écart-type de la population $\sigma_{pop} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2}$	moyenne de l'échantillon $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ écart-type de l'échantillon $\sigma_{ech} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Si on étudie un qualitatif	proportion dans la population p	proportion dans l'échantillon f

Les méthodes d'échantillonnage

Les méthodes empiriques: Utilisées par les instituts de sondage.
Échantillonnage sur la base du jugement: Échantillon prélevé après avis d'experts, qui connaissent la population et sont capable de dire les entités représentatives.
Échantillonnage par des quotas: Échantillon prélevé librement en respectant une composition donnée à l'avance (sexe, âge, CSP).
Les méthodes aléatoires : Reposent sur le tirage au hasard d'échantillons et sur le calcul des probabilités
Échantillonnage aléatoire simple : On prélève dans la population, des individus au hasard, sans remise : tous les individus ont la même probabilité d'être prélevés, et ils le sont indépendamment les uns des autres.
Échantillonnage aléatoire stratifié : Suppose que la population soit stratifiée, i.e. constituée de sous-populations homogènes, les strates. (ex : stratification par tranche d'age). Dans chaque strate, on fait un échantillonnage aléatoire simple, de taille proportionnelle à la taille de strate dans la population (échantillon représentatif). Les individus de la population n'ont pas tous la même probabilité d'être tirés. Nécessite une homogénéité des strates. Augmente la précision des estimations.

Échantillonnage par grappe : on tire au hasard des grappes ou familles d'individus, et on examine tous les individus de la grappe (ex: on tire des immeubles puis on interroge tous les habitants). La méthode est d'autant meilleure que les grappes se ressemblent et que les individus d'une même grappe sont différents, contrairement aux strates.

Les méthodes d'échantillonnage

L'échantillonnage simple consiste à extraire un échantillon de taille n dans une population de taille N par des tirages aléatoires équiprobables et indépendants (tirages avec remise).

- $\Omega = \{\omega_1, \ldots, \omega_N\}$: la population constituée d'éléments appelés unités d'observation.
- X : le caractère que l'on voudrait étudier sur l'ensemble de cette population.
- X_k : le résultat aléatoire du k ème tirage est une v.a qui suit la même loi que X.
- $(X_1, \ldots, X_n)$: les résultats aléatoires des n tirages modélisant l'échantillon étudié.
- x_k: est une réalisation du k-ème tirage de X_k.
- $= (x_1, \ldots, x_n)$: une réalisation de l'échantillon $(X_1, \ldots, X_n)$.

Modélisation d'échantillon aléatoire

Définition

$X_1, \dots, X_n$ sont n v.a. indépendantes et de même loi (celle de X); il est appelé n -échantillon ou échantillon de taille n de X.
Une réalisation $(x_1, \dots, x_n)$ de l'échantillon $(X_1, \dots, X_n)$ est l'ensemble des valeurs observées.

Définition

Une statistique Y sur un échantillon $(X_1, \dots, X_n)$ est une v.a., fonction mesurable des X_k :

$$Y = \varphi\left(X_1, \dots, X_n\right).$$

Après la réalisation, la statistique Y prend la valeur $\varphi(x_i, \dots, x_n)$.

Distribution d'échantillonnage d'une moyenne

Soit X le caractère quantitatif que l'on voudrait étudier sur une population infinie et soit $(X_1, X_2, \ldots, X_n)$ un n-échantillon de X.

Définition

La statistique $\overline{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$ est appelée moyenne empirique de X.

Remarque

La moyenne empirique est une variable aléatoire qui prend des valeurs différentes sur chaque échantillon appelées moyennes observées.

Proposition

Soit X une variable aléatoire de moyenne μ et de variance σ^2 et soit $(X_1, X_2, \ldots, X_n)$ un n-échantillon de X. Alors
- $\mathbb{E}(\bar{X}) = \mu$.
- $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

Proposition

La distribution d'échantillonnage de la moyenne est donnée par:

- Si X suit une loi normale $\mathcal{N}(\mu, \sigma)$, alors $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.
- Si la loi de X est quelconque avec $n \ggg 30$, le théorème central limite nous permets d'affirmer que $\overline{X}$ suit une loi normale $\mathcal{N}(\mu, \sigma)$.

Distribution d'échantillonnage d'une variance

Soit $(X_1, X_2, \ldots, X_n)$ un n-échantillon d'une variable aléatoire X de moyenne μ et d'écart-type σ. On définit la statistique

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right) - \bar{X}^2$$

Alors

$$\mathbb{E}\left(S^2\right) = \left(1 - \frac{1}{n}\right)\sigma^2.$$

Remarque

Remarque Comme $1 - \frac{1}{n} < 1$; alors $\mathbb{E}\left(S^2\right) < \sigma^2$.
En moyenne, la variance dans l'échantillon est plus faible que dans la population-mère.

Distribution d'échantillonnage d'une variance

Proposition

Proposition Si le caractère X à étudier suit une loi normale $\mathcal{N}(\mu, \sigma)$ alors $n\frac{S^2}{\sigma^2}$ suit une loi de khi-deux à $(n-1)$ degrés de liberté càd

$$n\frac{S^2}{\sigma^2} \sim \chi^2(n-1).$$

Distribution d'échantillonnage d'une proportion

Soit une population comportant deux modalités A et B. Soit p la proportion d'individus de la population possédant la modalité A. $1 - p$ est donc la proportion des individus de la population possédant la modalité B. On extrait de la population un échantillon de taille n. Soit K_n la v.a qui représente le nombre d'individus dans l'échantillon ayant la modalité A.

Définition

La variable aléatoire $\hat{p} = \dfrac{K_n}{n}$ s'appelle la fréquence empirique. Sa réalisation f est la proportion d'individus dans l'échantillon ayant la modalité A.

Proposition

- Si $n \gg 30$, $np \geq 5$ ou $n(1-p) \geq 5$, par le théorème central limite,

$$\hat{P} \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

- Sinon (le cas ou $n < 30$), la variable K_n suit une loi Binômiale $\mathcal{B}(n, p)$

Estimation Paramétrique

Dilemme

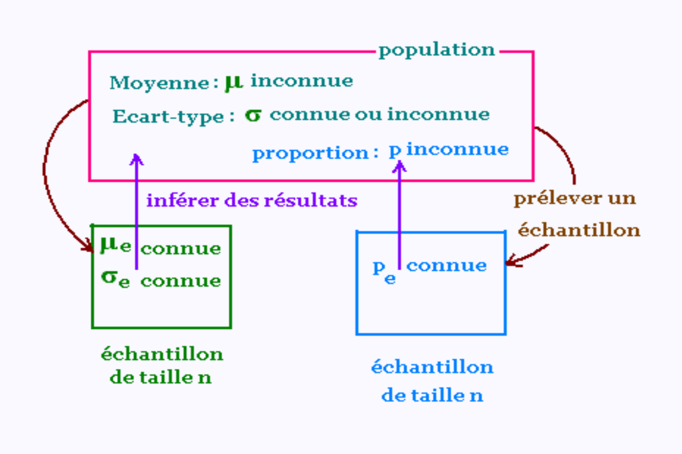


Figure: Inférence statistique-Estimation paramétrique.

Inférence statistique

L'inférence statistique consiste à induire les caractéristiques inconnues d'une population à partir d'un échantillon.

Stat. descriptives (L1)	Stat. inférentielles (L3)
- petite population	- très grande population
- toutes les données	- données d'un échantillon
- on calcule les paramètres	- on extrapole à partir de l'échantillon

Exemple introductif

On a observé 25 séropositifs sur un échantillon de 5000 sujets, soit un taux de 0.5%. Ce taux observé n'a de signification qu'assorti d'une fourchette : le risque que le vrai taux sorte d'une fourchette comprise entre 0.3 % et 0.7 % est acceptable. On peut diminuer ce risque, mais alors la fourchette est plus large, et devient moins intéressante.

Dans ce cours, nous allons apprendre à estimer à l'aide d'un échantillon :

- Dans le cas d'un caractère quantitatif la moyenne m et l'écart-type d'une population.
- Dans le cas d'un caractère qualitatif, la proportion p de la population.

Ces estimations peuvent s'exprimer par une seule valeur (estimation ponctuelle), soit par un intervalle (estimation par intervalle de confiance). Bien sûr, comme l'échantillon ne donne qu'une information partielle, ces estimations seront accompagnées d'une certaine marge d'erreur.

Bias d'un estimateur

Lorsqu'on utilise fréquemment des estimateurs ponctuels on souhaite qu'ils possèdent certaines propriétés. Un paramètre inconnu peut avoir plusieurs estimateurs. Par exemple, pour estimer le paramètre m, moyenne d'une population, on pourrait se servir de la moyenne arithmétique, de la médiane ou du mode.

Estimateur non biaisé.

On note θ le paramètre de valeur inconnue, $\hat{\theta}$ l'estimateur de θ. Un estimateur est *sans biais* si la moyenne de sa distribution d'échantillonnage est égale à la valeur θ du paramètre de la population à estimer, c'est-à-dire si $E(\hat{\theta}) = \theta$.

Si l'estimateur est biaisé, son biais est mesuré par l'écart suivant : BIAIS = $E(\hat{\theta}) - \theta$.

Etant donné un échantillon observé $(x_1, \ldots, x_n)$ et une loi de probabilité $\mathbb{P}_\theta$, la vraisemblance quantifie la probabilité que les observations proviennent effectivement d'un échantillon (théorique) de la loi $\mathbb{P}_\theta$.
Prenons l'exemple de 10 lancers de pièce. L'échantillon binaire observé est par exemple : 0 , 1 , 1 , 0 , 1 , 1 , 1 , 0 , 0 , 1 .
Pour un échantillon de taille 10 de la loi de Bernoulli de paramètre p, la probabilité d'une telle réalisation est $p^6(1-p)^4$. Voici quelques valeurs numériques.

p	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$p^6(1-p)^4$	2.6×10^{-5}	1.8×10^{-4}	5.3×10^{-4}	9.8×10^{-4}	1.2×10^{-3}	9.5×10^{-4}	4.2×10^{-4}

Il est naturel de choisir comme estimation de p, celle pour laquelle la probabilité de l'échantillon observé est la plus forte, à savoir ici $p = 0,6$.

On a la définition suivante de la fonction de vraisemblance, pour une variables aléatoire discrète.

Cas Variable aléatoire discrètre

Soit $C = \{c_1, \ldots, c_k\}$ un ensemble fini, $\{\mathbb{P}_\theta\}$ une famille de lois de probabilité sur C, et n un entier. On appelle vraisemblance associée à la famille $\{P_\theta\}$, la fonction qui à un n-uplet $(x_1, \ldots, x_n)$ d'éléments de C et à une valeur θ du paramètre associe la quantité :

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} \mathbb{P}_\theta(x_i).$$

Variable aléatoire discrète continue (VC)

Dans le cas d'un modèle continu, la loi $\mathbb{P}_\theta$ a une densité sur $\mathbb{R}$, et la probabilité pour que l'échantillon prenne une valeur particulière est toujours nulle. Il faut alors remplacer la probabilité $\mathbb{P}_\theta$ par sa densité dans la définition de la vraisemblance.

Soit $\{\mathbb{P}_\theta\}$ une famille de lois de probabilité continues sur $\mathbb{R}$ et n un entier. Notons f_θ la densité de probabilité de la loi $\mathbb{P}_\theta$. On appelle vraisemblance associée à la famille $\{\mathbb{P}_\theta\}$, la fonction qui à un n-uplet $(x_1, \ldots, x_n)$ d'éléments de C et à une valeur θ du paramètre associe la quantité :

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n} f_\theta(x_i) \ .$$

Maximum de vraisemblance

Définition

Supposons que pour toute valeur $(x_1, \ldots, x_n)$, la fonction qui à θ associe $L(x_1, \ldots, x_n, \theta)$ admette un maximum unique. La valeur $\widehat{\theta}$ pour laquelle ce maximum est atteint dépend de $(x_1, \ldots, x_n)$:

$$\widehat{\theta} = \tau(x_1, \ldots, x_n) = \arg\max L(x_1, \ldots, x_n, \theta) \ .$$

On l'appelle estimation par maximum de vraisemblance.
Si $(X_1, \ldots, X_n)$ est un échantillon (théorique) de la loi $\mathbb{P}_\theta$, la variable aléatoire :

$$T = \tau(X_1, \ldots, X_n) \ ,$$

est l'estimateur du maximum de vraisemblance de θ.

Exemple-Maximum de vraisemblance-VD

On prend une variable de Bernoulli, l'ensemble des valeurs possibles est $\{0, 1\}$. Le paramètre inconnu est p. Si $(x_1, \ldots, x_n) \in \{0, 1\}^n$ est un échantillon, la vraisemblance vaut :

$$L(x_1, \ldots, x_n, p) = p^{\sum x_i}(1-p)^{n-\sum x_i} .$$

Son logarithme est :

$$\log(L(x_1, \ldots, x_n, p)) = (\sum x_i) \log p + (n - \sum x_i) \log(1-p) .$$

La dérivée par rapport à p est :

$$\frac{\partial \log(L(x_1, \ldots, x_n, p))}{\partial p} = (\sum x_i)\frac{1}{p} - (n - \sum x_i)\frac{1}{1-p} .$$

Elle s'annule pour :

$$\widehat{p} = \frac{\sum x_i}{n} .$$

Exemple-Maximum de vraisemblance-VD

La dérivée seconde est :

$$\frac{\partial^2 \log(L(x_1,\dots,x_n,p))}{\partial p^2} = -(\sum x_i)\frac{1}{p^2} - (n - \sum x_i)\frac{1}{(1-p)^2}\,.$$

Elle est strictement négative, la valeur $\widehat{p}$ est bien un maximum.
Si $(X_1,\dots,X_n)$ est un échantillon de la loi de Bernoulli de paramètre p, l'estimateur du maximum de vraisemblance de p est :

$$\hat{p} = \frac{\sum X_i}{n} = \overline{X}.$$

Exemple-Maximum de vraisemblance-VC

On considère une expérience de mesure de durée de vie de disques durs d'ordinateurs. À $t = 0$ l'ordinateur (ou, plus vraisemblablement l'appareil de test) est mis en route jusqu'à ce que le disque tombe en panne. On peut raisonnablement supposer que les durées de vie des disques sont indépendantes et on décide de travailler (pour commencer au moins) avec un modèle simple de durée de vie exponentielle de paramètre λ. Si nous notons X_i la variable aléatoire décrivant la durée de vie du ième disque on suppose donc :

$$\mathbb{P}(X_i = x_i|\lambda) = (1/\lambda)exp(-x_i/\lambda).$$

Comme les X_i sont aussi supposées IID on a pour la loi conjointe :

$$\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n|\lambda\right) = \prod_{i=1}^{n}\left((1/\lambda)exp(-x_i/\lambda)\right),$$

Exemple-Maximum de vraisemblance-VC

soit pour la log-vraisemblance :

$$L(x_1, \ldots, x_n, \lambda) = -n \log(\lambda) - \frac{\sum_{i=1}^{n} x_i}{\lambda}.$$

On montre alors immédiatement en annulant la fonction score que :

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{X}.$$

Ainsi l'estimateur du maximum de vraisemblance (EMV) d'une loi exponentielle est la moyenne empirique.

Les Tests Statistique

La majorité des tests repose sur le principe suivant : On définit une hypothèse nulle notée H_0 contre l'hypothèse alternative H_1.

$$\textit{Test} : \begin{cases} \text{hypothèse nulle} H_0, \\ \text{hypothèse alternative} H_1. \end{cases}$$

Le test a pour objectif d'accepter ou de rejeter H_0 avec un risque connu à partir des données dont on dispose.
On détermine alors une statistique qui est une variable aléatoire construite à partir des données. Sous H0, cette variable suit une loi de probabilité connue (normale, student, khi-deux,...). On détermine alors l'intervalle de confiance dont lequel doit tomber la statistique avec une probabilité donnée $(1 - \alpha)$, (le plus souvent 95% pour $\alpha = 5\%$).

On définit alors la règle de décision suivante :

- Si la statistique tombe dans l'intervalle, on accepte H_0: Attention, cela ne veut pas dire que H_0 est vraie mais que le test et les données ne permettent pas de voir un écart significatif à H_0.
- Si la statistique ne tombe pas dans l'intervalle, on rejette H_0 avec le risque α de se tromper, (par exemple $\alpha = 5\%$).

Hypothèses statistiques

- Hypothèse nulle H_0 : hypothèse selon laquelle on fixe a priori un paramètre de la population à une valeur particulière.
- Hypothèse alternative H_1 (ou contre-hypothèse) : n'importe quelle autre hypothèse qui differe de l'hypothèse nulle H_0.
- La formulation des hypothèses peut être écrite comme

$$\begin{cases} \theta = \theta_0 & (H_0) \\ \theta \neq \theta_0 & (H_1) \end{cases} \quad \begin{cases} \theta = \theta_0 & (H_0) \\ \theta > \theta_0 & (H_1) \end{cases} \quad \begin{cases} \theta = \theta_0 & (H_0) \\ \theta = \theta_1 & (H_1) \end{cases}$$

$$\begin{cases} \theta = \theta_0 & (H_0) \\ \theta < \theta_0 & (H_1) \end{cases} \quad \begin{cases} \theta \leq \theta_0 & (H_0) \\ \theta > \theta_0 & (H_1) \end{cases} \quad \begin{cases} \theta \geq \theta_0 & (H_0) \\ \theta < \theta_0 & (H_1) \end{cases}$$

Il existe deux types d'erreurs :

- On appelle erreur de première espèce ou erreur de type I, notée α, la probabilité de rejeter H_0 alors qu'elle est vraie. α est aussi appelé niveau ou seuil de signification.

$$\alpha = \mathbb{P}\left(\text{ Rejeter } H_0 / H_0 \text{ vraie }\right) = \mathbb{P}\left(\text{ Choisir } H_1 / H_0 \text{ vraie }\right).$$

- On appelle erreur de deuxième espèce ou erreur de type II, notée β, la probabilité d'accepter H_0 alors qu'elle est fausse.
  $(1-\beta)$ est appelé puissance du test pour H_1, c'est la probabilité de retenir H_1 alors qu'elle est vraie.

$$1-\beta = \mathbb{P}\left(\text{ Rejetet } H_0 / H_1 \text{ vraie }\right)$$

Risques d’un test statistique
Tableau d’un test statistique

On peut résumer les types d’erreurs susceptibles de survenir dans le tableau suivant:

Décision suite au test	Réalité	Réalité
	H_0 vraie	H_0 fausse
Rejet de H_0	Erreur de premier espèce α	Bonne décision
Acceptation de H_0	Bonne décision	Erreur deuxième espèce β

Théorème de Neyman-Pearson

Le théorème de Neyman-Pearson est un théorème qui permet lors de test d'hypothèses simples, de construire le "meilleur" test à risque fixé. On suppose donc qu'on a une variable aléatoire dépendant d'un paramètre et $(x_1, \ldots, x_n)$ un échantillon de valeurs observées de X. La vraisemblance de l'échantillon pour le paramètre est défini par

$$L(x_1, \ldots, x_n, \theta) = \begin{cases} \prod_{i=1}^{n} \mathbb{P}(x_i/\theta) & siXestv.a.dicrte \\ \prod_{i=1}^{n} f(x_i; \theta) & siXestv.a.continue \end{cases}$$

On suppose qu'on doive décider entre $(H_0) : \theta = \theta_0$ et $(H_1) : \theta = \theta_1$.

Théorème de Neyman-Pearson

Alors le test de puissance maximum pour un risque de première espèce a fixé est celui dont la région critique est définie par

$$W = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n / \frac{L(x_1, \ldots, x_n, \theta_0)}{L(x_1, \ldots, x_n, \theta_1)} \leq k \right\}$$

où k est défini par

$$\mathbb{P}\left(\frac{L(x_1, \ldots, x_n, \theta_0)}{L(x_1, \ldots, x_n, \theta_1)} \leq k / \theta = \theta_0 \right) = \alpha$$

Autrement dit, on rejette H_0 si et seulement si $(x_1, ..., x_n)$ appartient à W.

Test de conformité de la moyenne à une référence

Cas d'un petit échantillon $n < 30$ où σ est inconnue

$$\text{Test bilatéral : } \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

On calcule :

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

On détermine t_α pour la loi T de Student à $n - 1$ ddl,
$\left(\mathbb{P}\left(T > t_\alpha\right) = \frac{\alpha}{2}\right)$, et on décide que:
(1) Si $t \in] - t_\alpha; t_\alpha [$, on ne peut rejeter H_0.
(2) Sinon, on rejette H_0 avec une probabilité α de se tromper.

Test de conformité de la moyenne

Exemple: Soit $n = 25$, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, $\bar{x} = 15$ et $s = 9$. Peut on affirmer que la moyenne inconnue de la population est 12.

$$\text{Test bilatéral : } \begin{cases} H_0 : \mu = \mu_0 = 12 \\ H_1 : \mu \neq \mu_0 = 12 \end{cases}$$

Corrigé: On calcule :

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{15 - 12}{\frac{9}{\sqrt{25}}} = 1.67.$$

Le ddl est 24, donc pour $\alpha = 5\% = 0.05$, on trouve $t_\alpha = 2.07$. Comme $t \in \left]-t_\alpha ; t_\alpha\right[= \left]-2.07, 2.07\right[$, on accept donc H_0.

Test de conformité de la moyenne à une référence

Cas d'un petit échantillon $n > 30$ où σ est inconnue

$$\text{Test bilatéral :} \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

On calcule :

$$u = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

On détermine z_α pour la loi normale $\mathcal{N}(0, 1)$, $\left(\Phi\left(z_\alpha\right) = 1 - \frac{\alpha}{2}\right)$, et on décide que :

(1) Si $u \in] -z_\alpha ; z_\alpha [$, on ne peut rejeter H_0.

(2) Sinon, on rejette H_0 avec une probabilité α de se tromper.

Test de conformité de la variance

$$\text{Test bilatéral : } \begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$$

Ici on travaille avec la loi Khi-deux à $n-1$ ddl. On cherche les deux nombres a et b tels que : $P\left(Y^2 \geq a\right) = \frac{\alpha}{2}$ et $P\left(Y^2 \geq b\right) = 1 - \frac{\alpha}{2}$. On calcule :

$$y^2 = \frac{n-1}{\sigma_0^2} s^2$$

(1) Si $y^2 \in \left] a; b \right[$, on ne peut rejeter H_0.
(2) Sinon, on rejette H_0 avec une probabilité α de se tromper.

Test de conformité de la variance à une référence

Exemple

$$\text{Test bilatéral : } \begin{cases} H_0 : \sigma^2 = \sigma_0^2 = 90 \\ H_0 : \sigma^2 \neq \sigma_0^2 = 90 \end{cases}$$

Pour $n = 31$, $s^2 = 100$, on a le $\mathrm{ddl} = 31 - 1 = 30.\alpha = 0.05$ donc à partir de la table de la loi χ_2, pour $\frac{\alpha}{2} = 0.025$ donne $a = 46.98$ et $1 - \frac{\alpha}{2} = 0.975$ donne $b = 16.80$.

$$y^2 = \frac{n-1}{\sigma_0^2} s^2 = \frac{30}{90} \times 100 = 33.34$$

Comme $y^2 \in]\, a; b\,[$, on accept l'hypothèse H_0.

Test de conformité de la proportion à une référence

$$\text{Test bilatéral : } \begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

On calcule :

$$u = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

On détermine z_α pour la loi normale $\mathcal{N}(0,1)$, $\left(\Phi(z_\alpha) = 1 - \frac{\alpha}{2}\right)$, et on décide que :

(1) Si $u \in]-z_\alpha ; z_\alpha[$, on ne peut rejeter H_0.

(2) Sinon, on rejette H_0 avec une probabilité α de se tromper.

Test de Proportion

Exemple: Sur un échantillon de taille $n = 400$ de naissances, on a observé 206 mâles, soit une proportion de mâles de $\bar{p} = 206/400 = 0,515$. On se demande si il y a autant de mâles que de femelles. i.e., si $p_0 = 0,5$. On peut effectuer alors le test :

$$\text{Test bilatéral : } \begin{cases} H_0 : p = p_0 = 0.5 \\ H_1 : p \neq p_0 = 0.5 \end{cases}$$

Corrigé: On calcule :

$$u = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.515 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{400}}} = 0.6.$$

Pour $\alpha = 5\% = 0.05$, on a $z_\alpha = 1.96$ Comme $u \in]-z_\alpha ; z_\alpha[$, alors on ne peut rejeter H_0 : donc il est possible que $p = 0.5$

Test de Khi deux

Test d'indépendance du khi-deux entre deux variables qualitatives Le test du khi-deux est largement utilisé pour l'étude de l'indépendance entre deux caractères qualitatifs.

La présentation des résultats se fait sous forme d'un tableau de contingence à deux entrées.

Chaque entrée représente les modalités d'une des variables.

On détermine alors le tableau attendu sous l'hypothèse d'indépendance.

Exemple de Test de Khi deux

Un échantillon de $n = 1000$ personnes ont été interrogées sur une question précise. On a demandé à ces personnes de préciser leur sexe. Nous voudrions savoir si la réponse Y à la question est indépendante du sexe X.

X / Y	Favorable	Défavorable	Indécis	Total
Masculin	$n_{11} = 210$	$n_{12} = 194$	$n_{13} = 91$	$n_{1\bullet} = 495$
Féminin	$n_{21} = 292$	$n_{22} = 151$	$n_{23} = 62$	$n_{2\bullet} = 505$
Tot al	$n_{\bullet 1} = 502$	$n_{\bullet 2} = 345$	$n_{\bullet 3} = 153$	$n = 1000$

Test de Khi deux

Sous hypothèse d'indépendance, la distribution conjointe est simplement le produit des distributions marginales, i.e., $f_{ij} = f_{i\bullet} f_{\bullet j}$. Si l'on estime f_{ij} par $\frac{n_{ij}}{n}$ et $f_{i\bullet}$ par $\frac{n_{i\bullet}}{n}$ et $f_{\bullet j}$ par $\frac{n_{\bullet j}}{n}$, on devrait donc avoir :

$$n_{ij} \approx \frac{n_{i\bullet} n_{\bullet j}}{n}$$

L'idée est de calculer l'écart entre les deux termes: le n_{ij} observé et le n_{ij} prédit ou théorique, si cet écart devient trop important, on devra rejeter l'hypothèse que les variables sont indépendantes.

Test de Khi deux

On calcule :

$$Q = \sum_{i=1}^{k}\sum_{j=1}^{1} \frac{\left(n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet}n_{\bullet j}}{n}}$$

La statistique Q est distribuée suivant une khi-deux à $(k-1)(l-1)$ ddl, où k et l désignent le nombre de valeurs des deux variables.

Test de validité du test : il faut que la fréquence théorique $\frac{n_{i\bullet}n_{\bullet j}}{n} \geq 5$ pour tous i et j.

Nous acceptons H_0 si $Q \leq q_\alpha$ avec $\mathbb{P}\left(Y^2 > q_\alpha\right) = \alpha$.

Reprendre l'exemple

Dans notre exemple du tableau, pour $\alpha = 0.05$, on a le ddl est $(3-1)\times(2-1) = 2$, $q_\alpha = 5,99$ et la valeur de Q est :

$$Q = \frac{\left(\frac{495\times502}{1000} - 210\right)^2}{\frac{495\times502}{1000}} + \frac{\left(\frac{495\times345}{1000} - 194\right)^2}{\frac{495\times345}{1000}} + \frac{\left(\frac{495\times153}{1000} - 91\right)^2}{\frac{495\times153}{1000}} + \frac{\left(\frac{505\times502}{1000} - 292\right)^2}{\frac{505\times502}{1000}} + \frac{\left(\frac{505\times345}{1000} - 151\right)^2}{\frac{505\times345}{1000}} + \frac{\left(\frac{505\times153}{1000} - 62\right)^2}{\frac{505\times153}{1000}} = 24,15.$$

On remarque que $Q > q_\alpha$ $(24,15 > 5,99)$, on rejette alors H_0 avec un risque de $\alpha = 0.05$ de se tromper.

Intervalles de Confiance

Dilemme

Figure: Choix de la moyenne de la population à partir d'un échantillon.

La théorie des intervalles de confiance (IC) consiste à construire, autour de l'estimation ponctuelle, un intervalle qui aura une grande probabilité $(1-\alpha)$ de contenir la vraie valeur du paramètre. La probabilité associée à l'intervalle de confiance et exprimée en pourcentage est égale à $1-\alpha$ où α est le seuil de confiance ou niveau de confiance de l'intervalle, exprimé également en pourcentage. Autrement dit,

$$P(LI \leq \theta \leq LS) = 1-\alpha,$$

où

- LI est la limite inférieure de l'intervalle de confiance.
- LS est la limite supérieure de l'intervalle de confiance.
- S est la probabilité associée à l'intervalle d'encadrer la vraie valeur du paramètre.

IC de la moyenne μ de la population

1er cas : Lorsque la taille de l'échantillon est grande $(n \geq 30)$ et la variance de la population de X est connue, on obtient un intervalle de confiance pour μ au seuil de confiance $(1-\alpha)$ de la forme :

$$IC_{(1-\alpha)\%} = \left[\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right]$$

$$\text{i.e., } P\left(\mu \in IC_{(1-\alpha)\%}\right) = (1-\alpha)$$

$\Phi\left(z_\alpha\right) = 1 - \frac{\alpha}{2}$, où Φ est la fdr de la loi normale $\mathcal{N}(0,1)$.

IC de la moyenne μ de la population

1. Ceci est aussi vrai pour de petits échantillons ($n < 30$) lorsque la variable aléatoire X suit une loi normale et que la variance de X est connue.
2. Lorsque la taille de l'échantillon est grande $(n \geq 30)$ et la variance de la population de X est inconnue, on obtient un intervalle de confiance pour la moyenne μ au seuil de confiance $(1 - \alpha)$ de la forme :

$$IC_{(1-\alpha)\%} = \left[\bar{x} - z_\alpha \frac{S}{\sqrt{n}} ; \bar{x} + z_\alpha \frac{S}{\sqrt{n}}\right]$$

$$\text{i.e., } P\left(\mu \in IC_{(1-\alpha)\%}\right) = (1 - \alpha)$$

où S^2 est l'estimateur de la variance, définit par:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

IC de la moyenne μ de la population

Exemple:

On a observé la taille de $n = 200$ hommes Algériens adultes. Après calcul, on a obtenu une moyenne de $\mu = 168cm$. Si on suppose que la variance connue vaut $\sigma^2 = 1$. Donnez un intervalle de confiance à 95% de la vraie moyenne de la population.

Corrigé:

Puisque $\alpha = 0,05(5\%)$, alors $\mathbb{P}(z_\alpha) = 1 - \frac{\alpha}{2} = 0,975$, par suite $z_\alpha = 1,96$.

Finalement

$$IC_{95\%} = [167,86; 168,14], i.e., \mathbb{P}(\mu \in [167,86; 168,14]) = 0,95$$

IC de la moyenne μ de la population

2eme cas : Lorsque la taille de l'échantillon est petite $(n < 30)$ et X suit une loi normale de variance inconnue, on obtient un intervalle de confiance pour μ au seuil de confiance $(1-\alpha)$ de la forme :

$$IC_{(1-\alpha)\%} = \left[\bar{x} - t_\alpha \frac{s}{\sqrt{n}}; \bar{x} + t_\alpha \frac{s}{\sqrt{n}}\right]$$

t_α se déduit da la table student comme suit :

$$\mathbb{P}\left(T > t_\alpha\right) = \frac{\alpha}{2}.$$

IC de la moyenne μ de la population

Exemple:

Un échantillon de 10 appartements (trois et demi) dans un rayon de 1 km de l'université a permis d'estimer le coût moyen du loyer mensuel à 350 par mois et un écart type de 30. Quel est l'intervalle de confiance de 95% pour la moyenne des loyers mensuels? Supposons que les loyers suivent une loi normale.

Corrigé:

pour un coefficient de confiance de 0, 95, on a $\alpha = 0,05$, et donc $\frac{\alpha}{2} = 0,025$. On a $n - 1 = 10 - 1 = 9$ degrés de liberté, alors la table de la distribution Student nous donne $t_\alpha = 2,262$. Finalement, après calcul

$$IC_{95\%} = [328,54\,;371,46].$$

Donc, nous sommes confiants à 95% que la moyenne des loyers mensuels (le vrai paramètre de la population μ), se trouve entre 328.54 et 371.46.

IC de la variance σ^2 de la population

Soit n la taille de l'échantillon et s^2 la variance échantillonnale. Ici, $Y^2 = \frac{(n-1)S^2}{\sigma^2}$ suit la loi de khi-deux à $n-1$ ddl. On cherche les deux nombres a et b tels que : $\mathbb{P}\left(Y^2 \geq a\right) = \frac{\alpha}{2}$ et $\mathbb{P}\left(Y^2 \geq b\right) = 1 - \frac{\alpha}{2}$.

L'intervalle de confiance au seuil $(1-\alpha)$ pour la variance inconnue σ^2 de la population est de la forme :

$$IC_{(1-\alpha)\%} = \left[\frac{(n-1)s^2}{a}; \frac{(n-1)s^2}{b}\right].$$

IC de la variance σ^2 de la population

Exemple:
Pour $n = 31$, $s^2 = 100$, on a le $ddl = 31 - 1 = 30$. Pour $\alpha = 0,05$, on a $\frac{\alpha}{2} = 0,025$, ce qui donne $a = 46,98$ et $1 - \frac{\alpha}{2} = 0,975$, ce qui donne $b = 16,80$.
Finalement,

$$IC_{95\%} = [0,3965\,;0,4835].$$

IC de la proportion p de la population

Lorsque n est grand $(n \geq 30)$ et si $\bar{p}$ est la proportion échantillonnale alors un intervalle de confiance pour la proportion p inconnue de la population au seuil de confiance $(1-\alpha)$ de la forme

$$IC_{(1-\alpha)\%} = \left[\bar{p} - z_\alpha \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}; \bar{p} + z_\alpha \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right]$$

avec $\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$, où Φ est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$

IC de la moyenne μ de la population

Exemple:

Le Conseil Electoral (CE) est une instance qui s'occupe des sondages politiques. À l'aide de sondages téléphoniques, les interviewers demandent aux citoyens pour qui ils voteraient si les élections avaient lieu aujourd'hui. Récemment, CE a trouvé que 220 votants sur 500 voterait pour un candidat particulier. Le CE veut estimer l'intervalle de confiance à 95% pour la proportion des votants qui sont en faveur de ce candidat.

Corrigé:

On a $n = 500$, $\overline{p} = 220/500 = 0,44$ et $z_\alpha = 1,96$ obtenue à partir de la table de la loi gaussienne, donc

$$IC_{95\%} = [0,3965\,;0,4835],$$

i.e., SPI est confiant à 95% que la proportion des votants qui favoriseront ce candidat est entre 0.3965 et 0.4835.

Table Loi Normale

Table de Gauss

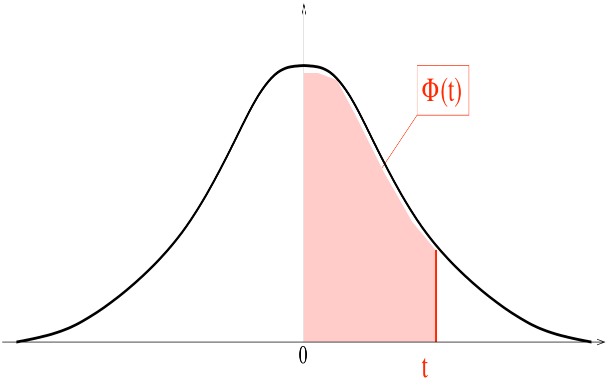
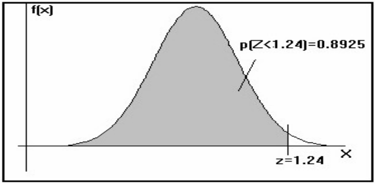


Figure: La table donne $\Phi(t) = P(0 < X < t)$.

Utilisation de la table de Gauss

TABLE DE LA LOI NORMALE CENTREE REDUITE

Lecture de la table: Pour z=1.24 (intersection de la ligne 1.2 et de la colonne 0.04), on a la proportion P(Z < 1,24) = 0.8925



P(Z > 1,96) = 0,025
P(Z > 2,58) = 0,005
P(Z > 3,29) = 0,0005

Rappels:
1/ P(Z > z) = 1 - P(Z < z) et 2/ P(Z < -z) = P(Z > z)
Exemple: Sachant P(Z < 1,24) = 0,8925, on en déduit:
1/ (P(Z > 1,24) = 1 - P(Z < 1,24) = 1- 0,8925 = 0,1075
2/ P(Z < -1,24) = P(Z > 1,24) = 0,1075

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916

t	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
4.0	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Table de Khi deux

Table de χ_2

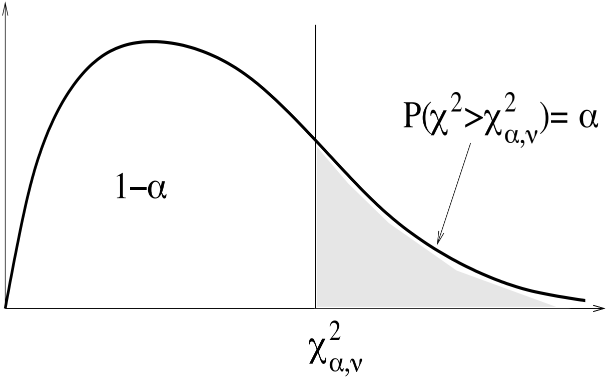


Figure: La table donne $\chi^2_{\alpha,\nu}$ tel que $P(|\chi^2_{\alpha,\nu}| > t_\alpha) = \alpha$.

Utilisation de la table de Khi deux

TABLE TEST Z

α	0,05	0,01	0,001
$\alpha/2$	0,025	0,005	0,0005
Z	1,96	2,58	3,29

DISTRIBUTION DU KHI2

La table donne les valeurs critiques de χ^2 pour un nombre de degrés de liberté (ddl) et pour un seuil repère donnés (α).

Par exemple:

Pour ddl = 3 et α= 0,05 la table indique $\chi^2 = 7{,}81$

Ceci signifie que: $P(\chi^2_{[3]} > 7{,}81) = 0{,}05$

ddl \ α	0,05	0,01	0,001
1	3,84	6,63	10,83
2	5,99	9,21	13,82
3	7,81	11,34	16,27
4	9,49	13,28	18,47
5	11,07	15,09	20,52
6	12,59	16,81	22,46
7	14,07	18,48	24,32
8	15,51	20,09	26,12
9	16,92	21,67	27,88
10	18,31	23,21	29,59
11	19,68	24,72	31,26
12	21,03	26,22	32,91
13	22,36	27,69	34,53
14	23,68	29,14	36,12
15	25,00	30,58	37,70
16	26,30	32,00	39,25
17	27,59	33,41	40,79
18	28,87	34,81	42,31
19	30,14	36,19	43,82
20	31,41	37,57	45,31
21	32,67	38,93	46,80
22	33,92	40,29	48,27
23	35,17	41,64	49,73
24	36,42	42,98	51,18
25	37,65	44,31	52,62
26	38,89	45,64	54,05
27	40,11	46,96	55,48
28	41,34	48,28	56,89
29	42,56	49,59	58,30
30	43,77	50,89	59,70

$\nu \setminus \alpha$	0.99	0.98	0.95	0.9	0.8	0.2	0.1	0.05	0.02	0.01
1	0.000	0.001	0.004	0.016	0.064	1.642	2.706	3.841	5.412	6.635
2	0.020	0.040	0.103	0.211	0.446	3.219	4.605	5.991	7.824	9.210
3	0.115	0.185	0.352	0.584	1.005	4.642	6.251	7.815	9.837	11.345
4	0.297	0.429	0.711	1.064	1.649	5.989	7.779	9.488	11.668	13.277
5	0.554	0.752	1.145	1.610	2.343	7.289	9.236	11.070	13.388	15.086
6	0.872	1.134	1.635	2.204	3.070	8.558	10.645	12.592	15.033	16.812
7	1.239	1.564	2.167	2.833	3.822	9.803	12.017	14.067	16.622	18.475
8	1.646	2.032	2.733	3.490	4.594	11.030	13.362	15.507	18.168	20.090
9	2.088	2.532	3.325	4.168	5.380	12.242	14.684	16.919	19.679	21.666
10	2.558	3.059	3.940	4.865	6.179	13.442	15.987	18.307	21.161	23.209
11	3.053	3.609	4.575	5.578	6.989	14.631	17.275	19.675	22.618	24.725
12	3.571	4.178	5.226	6.304	7.807	15.812	18.549	21.026	24.054	26.217
13	4.107	4.765	5.892	7.042	8.634	16.985	19.812	22.362	25.472	27.688
14	4.660	5.368	6.571	7.790	9.467	18.151	21.064	23.685	26.873	29.141
15	5.229	5.985	7.261	8.547	10.307	19.311	22.307	24.996	28.259	30.578
16	5.812	6.614	7.962	9.312	11.152	20.465	23.542	26.296	29.633	32.000
17	6.408	7.255	8.672	10.085	12.002	21.615	24.769	27.587	30.995	33.409
18	7.015	7.906	9.390	10.865	12.857	22.760	25.989	28.869	32.346	34.805
19	7.633	8.567	10.117	11.651	13.716	23.900	27.204	30.144	33.687	36.191
20	8.260	9.237	10.851	12.443	14.578	25.038	28.412	31.410	35.020	37.566
21	8.897	9.915	11.591	13.240	15.445	26.171	29.615	32.671	36.343	38.932
22	9.542	10.600	12.338	14.041	16.314	27.301	30.813	33.924	37.659	40.289
23	10.196	11.293	13.091	14.848	17.187	28.429	32.007	35.172	38.968	41.638
24	10.856	11.992	13.848	15.659	18.062	29.553	33.196	36.415	40.270	42.980
25	11.524	12.697	14.611	16.473	18.940	30.675	34.382	37.652	41.566	44.314
26	12.198	13.409	15.379	17.292	19.820	31.795	35.563	38.885	42.856	45.642
27	12.879	14.125	16.151	18.114	20.703	32.912	36.741	40.113	44.140	46.963
28	13.565	14.847	16.928	18.939	21.588	34.027	37.916	41.337	45.419	48.278
29	14.256	15.574	17.708	19.768	22.475	35.139	39.087	42.557	46.693	49.588
30	14.953	16.306	18.493	20.599	23.364	36.250	40.256	43.773	47.962	50.892

Table de Student

Table de Student

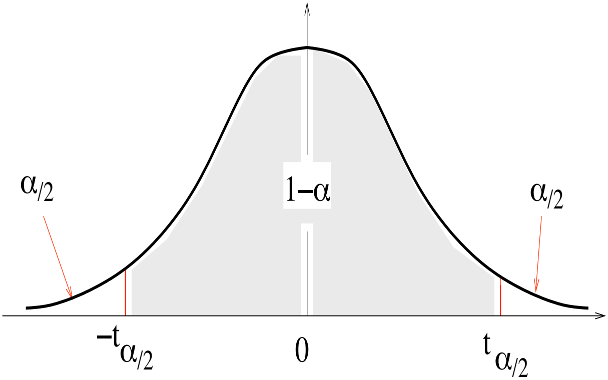


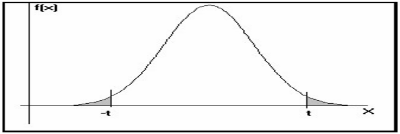
Figure: La table donne $t_{\alpha/2}$ tel que $P(|T_\nu| > t_{\alpha/2}) = \alpha$.

Utilisation de la table de Student

DISTRIBUTIONS DU t DE STUDENT

Table des valeurs critiques bilatérales usuelles

Pour une distribution de Student à ddl degrés de liberté et pour une proportion α (.05, .01 ou .001), la table indique t tel que $P(|T| > t) = \alpha$



Exemple: Pour ddl = 5, on a $P(|T| > 2.571) = .05$ (on note $t_{[5].05}$ cette valeur).

α α/2 ddl	0,05 0,025	0,01 0,005	0,001 0,0005
1	12.706	63.657	636.619
2	4.303	9.925	31.599
3	3.182	5.841	12.924
4	2.776	4.604	8.610
5	2.571	4.032	6.869
6	2.447	3.707	5.959
7	2.365	3.499	5.408
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587
11	2.201	3.106	4.437
12	2.179	3.055	4.318
13	2.160	3.012	4.221
14	2.145	2.977	4.140
15	2.131	2.947	4.073
16	2.120	2.921	4.015
17	2.110	2.898	3.965
18	2.101	2.878	3.922
19	2.093	2.861	3.883
20	2.086	2.845	3.850
21	2.080	2.831	3.819
22	2.074	2.819	3.792
23	2.069	2.807	3.768
24	2.064	2.797	3.745
25	2.060	2.787	3.725
26	2.056	2.779	3.707
27	2.052	2.771	3.690
28	2.048	2.763	3.674
29	2.045	2.756	3.659
30	2.042	2.750	3.646
40	2.021	2.704	3.551
60	2.000	2.660	3.460
120	1.980	2.617	3.373

$\nu \setminus \alpha$	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.82	63.65	636.6
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.59
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.92
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.126	0.254	0.387	0.527	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.460
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291