NORME INTERNATIONALE 2854

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION •МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ •ORGANISATION INTERNATIONALE DE NORMALISATION

Interprétation statistique des données — Techniques d'estimation et tests portant sur des moyennes et des variances

Statistical interpretation of data — Techniques of estimation and tests relating to means and variances

Première édition — 1976-02-15

CDU 519.28 **Réf. nº : ISO 2854-1976 (F)**

Descripteurs : analyse statistique, essai statistique, estimation, moyenne mathématique, variance.

Prix basé sur 46 pages

AVANT-PROPOS

L'ISO (Organisation Internationale de Normalisation) est une fédération mondiale d'organismes nationaux de normalisation (Comités Membres ISO). L'élaboration de Normes Internationales est confiée aux Comités Techniques ISO. Chaque Comité Membre intéressé par une étude a le droit de faire partie du Comité Technique correspondant. Les organisations internationales, gouvernementales et non gouvernementales, en liaison avec l'ISO, participent également aux travaux.

Les Projets de Normes Internationales adoptés par les Comités Techniques sont soumis aux Comités Membres pour approbation, avant leur acceptation comme Normes Internationales par le Conseil de l'ISO.

La Norme Internationale ISO 2854 a été établie par le Comité Technique ISO/TC 69, *Application des méthodes statistiques,* et soumise aux Comités Membres en octobre 1973.

Elle a été approuvée par les Comités Membres des pays suivants :

Afrique du Sud, Rép. d'	Hongrie	Roumanie
Allemagne	Inde	Royaume-Uni
Australie	Israël	Suisse
Belgique	Italie	Tchécoslovaquie
Brésil	Japon	Thaïlande
Bulgarie	Nouvelle-Zélande	Turquie
Égypte, Rép. arabe d'	Pays-Bas	U.R.S.S.
France	Pologne	Yougoslavie

Les Comités Membres des pays suivants ont désapprouvé le document pour des raisons techniques :

Suède
U.S.A.

Imprimé en Suisse

SOMMAIRE

MOYENNES	Variance	
	connue	inconnue
Comparaison d'une moyenne à une valeur donnée	A	A′
Estimation d'une moyenne	B	B′
Comparaison de deux moyennes	C	C′
Estimation de la différence de deux moyennes	D	D′

VARIANCES	
Comparaison d'une variance à une valeur donnée	E
Estimation d'une variance	F
Comparaison de deux variances	G
Estimation du rapport de deux variances	H

Page blanche

Interprétation statistique des données — Techniques d'estimation et tests portant sur des moyennes et des variances

SECTION UN : PRÉSENTATION DES CALCULS

REMARQUES GÉNÉRALES

1) La présente Norme Internationale spécifie les techniques permettant, à partir d'échantillons :

a) d'estimer la moyenne ou la variance de populations;

b) d'examiner certaines hypothèses concernant la valeur de ces paramètres.

2) Les techniques utilisées ne sont applicables que si l'on peut admettre que, dans chaque population considérée, les individus de l'échantillon ont été prélevés au hasard et sont indépendants. Dans le cas d'une population finie, des individus prélevés au hasard peuvent être considérés comme indépendants si l'effectif de la population est suffisamment élevé, ou si le taux d'échantillonnage est suffisamment petit (par exemple inférieur à 1/10).

3) La distribution du caractère étudié est supposée normale dans chaque population. Cependant, si la distribution ne s'écarte pas trop de la normale, les techniques décrites restent suffisamment valables pour la plupart des applications pratiques, à condition que l'effectif de l'échantillon ne soit pas trop petit. Pour les tableaux A, B, C et D, l'effectif de l'échantillon devrait être de l'ordre de 5 à 10 au minimum; pour tous les autres tableaux, il devrait être de l'ordre de 20 au minimum.[1)]

4) Un certain nombre de techniques permettent de vérifier l'hypothèse de normalité. Celles-ci feront l'objet d'exemples numériques dans la section deux, et d'un document séparé (encore à préparer). Mais, dans bien des cas, cette hypothèse peut être admise en fonction d'informations autres que celles fournies par les échantillons eux-mêmes. Dans le cas où l'hypothèse de normalité devrait être rejetée, il semble que la marche à suivre appropriée serait de faire appel à des tests non paramétriques ou d'utiliser des transformations permettant de retrouver des populations normales (par exemple $1/x$, $\log(x + a)$, $\sqrt{x + a}$, etc.), mais les conclusions obtenues en appliquant les procédures décrites dans la présente Norme Internationale ne seront directement valables que pour la variable transformée; leur transposition à la variable initiale exige des précautions. Par exemple, **exp(moyenne log** x**)** est égale à la **moyenne géométrique** de x et non pas à la moyenne arithmétique.

Si l'on désire réellement une estimation de la moyenne ou de l'écart-type de la variable X elle-même, alors peu importe que la distribution de la population soit normale ou non, une estimation sans biais de la moyenne m et de la variance σ^2 de la population est fournie par la moyenne $\bar{x}$ et la caractéristique s^2 de l'échantillon.

5) Il est souhaitable d'accompagner chaque opération statistique de toutes indications relatives à l'origine ou à la méthode de prélèvement des données susceptibles d'éclairer leur analyse statistique, notamment l'unité ou la fraction d'unité de mesure la plus petite ayant une signification pratique.

6) Il ne peut être procédé à l'élimination ou à la correction éventuelle de données individuelles apparemment douteuses, que s'il existe des raisons expérimentales, techniques ou évidentes, permettant une justification circonstanciée de cette élimination ou de cette correction. Dans tous les cas, les données éliminées ou corrigées doivent être mentionnées, ainsi que les raisons de leur élimination ou de leur correction.

7) Dans les problèmes d'estimation, le niveau de confiance $1 - \alpha$ est la probabilité pour que l'intervalle de confiance renferme la vraie valeur du paramètre estimé. Ses valeurs les plus usuelles sont 0,95 et 0,99, soit $\alpha = 0{,}05$ et $\alpha = 0{,}01$.

8) Dans les problèmes de tests d'hypothèse, le niveau de signification α est, dans le cas de tests bilatéraux, la probabilité de rejeter l'hypothèse nulle (ou hypothèse testée) lorsque cette hypothèse est vraie (erreur de première espèce); dans le cas des tests unilatéraux, le niveau de signification est la valeur maximale de cette probabilité (valeur maximale de l'erreur de première espèce). Ses valeurs les plus usuelles sont $\alpha = 0{,}05$ (1 chance sur 20) et 0,01 (1 chance sur 100), suivant le risque que l'utilisateur accepte de prendre. Étant donné qu'une hypothèse peut être rejetée en utilisant $\alpha = 0{,}05$, mais acceptée en utilisant 0,01, il est souvent indiqué d'utiliser la phrase «l'hypothèse est rejetée au niveau 5 %» ou, si c'est le cas, «au niveau 1 %». L'attention est attirée sur l'existence d'une erreur de deuxième espèce, erreur qui est commise lorsqu'on accepte

1) Des études concernant la normalité des distributions sont en cours au sein du TC 69/SC 2.

l'hypothèse nulle alors que celle-ci est fausse. Les termes relatifs aux tests statistiques sont définis dans le chapitre 2 de l'ISO 3534, *Statistique — Vocabulaire*[1].

9) Les calculs peuvent souvent être fortement simplifiés en effectuant sur les données un changement d'origine et/ou d'unité. Dans le cas d'observations classées par groupes, on peut se référer aux formules de l'ISO 2602, *Interprétation statistique de résultats d'essais — Estimation de la moyenne — Intervalle de confiance.*

NOTE — Un changement d'origine est essentiel pour obtenir une précision suffisante quand la variance est calculée selon la formule proposée, à l'aide d'un calculateur de faible précision.

10) Les méthodes indiquées dans les tableaux C et C' concernent la comparaison de deux moyennes. Elles supposent que les échantillons correspondants sont indépendants. Pour l'étude de certains problèmes, on peut avoir intérêt à apparier les observations (par exemple dans la comparaison de deux méthodes ou la comparaison de deux instruments). Le traitement statistique des observations appariées fait l'objet de l'ISO 3301, *Interprétation statistique des données — Comparaison de deux moyennes dans le cas d'observations appariées*, mais, dans l'annexe A, un exemple de traitement d'observations appariées est donné. Il utilise formellement les données du tableau A'.

11) Les symboles et leur définition utilisés dans la présente Norme Internationale sont conformes à l'ISO 3207, *Interprétation statistique des données — Détermination d'un intervalle statistique de dispersion.*

1) Actuellement au stade de projet.

TABLEAUX

A — Comparaison d'une moyenne à une valeur donnée (variance connue)

A′ — Comparaison d'une moyenne à une valeur donnée (variance inconnue)

B — Estimation d'une moyenne (variance connue)

B′ — Estimation d'une moyenne (variance inconnue)

C — Comparaison de deux moyennes (variances connues)

C′ — Comparaison de deux moyennes (variances inconnues, mais présumées égales)

D — Estimation de la différence de deux moyennes (variances connues)

D′ — Estimation de la différence de deux moyennes (variances inconnues, mais présumées égales)

E — Comparaison d'une variance ou d'un écart-type à une valeur donnée

F — Estimation d'une variance ou d'un écart-type

G — Comparaison de deux variances ou de deux écarts-types

H — Estimation du rapport de deux variances ou de deux écarts-types

TABLEAU A — **Comparaison d'une moyenne à une valeur donnée** (variance connue)

Caractéristiques techniques de la population étudiée (5) . . . .

Caractéristiques techniques des individus prélevés (5) . . . .

Observations éliminées (6) . . . .

Données statistiques	**Calculs**
Effectif de l'échantillon : $n =$	$\bar{x} = \frac{\Sigma x}{n} =$
Somme des valeurs observées : $\Sigma x =$	$[u_{1-\alpha}/\sqrt{n}]\,\sigma =$
Valeur donnée : $m_0 =$	$[u_{1-\alpha/2}/\sqrt{n}]\,\sigma =$
Valeur connue de la variance de la population : $\sigma^2 =$	
D'où l'écart-type : $\sigma =$	
Niveau de signification choisi (8) : $\alpha =$	

Résultats

Comparaison de la moyenne de la population à la valeur donnée m_0

Cas bilatéral :

L'hypothèse de l'égalité de la moyenne de la population à la valeur donnée (hypothèse nulle) est rejetée si l'on a :

$$|\bar{x} - m_0| > [u_{1-\alpha/2}/\sqrt{n}]\,\sigma$$

Cas unilatéraux :

a) L'hypothèse selon laquelle la moyenne de la population n'est pas inférieure à m_0 (hypothèse nulle) est rejetée si l'on a :

$$\bar{x} < m_0 - [u_{1-\alpha}/\sqrt{n}]\,\sigma$$

b) L'hypothèse selon laquelle la moyenne de la population n'est pas supérieure à m_0 (hypothèse nulle) est rejetée si l'on a :

$$\bar{x} > m_0 + [u_{1-\alpha}/\sqrt{n}]\,\sigma$$

NOTE — Les références (5), (6) et (8) renvoient aux paragraphes correspondants des «Remarques générales».

Commentaires

1) Le niveau de signification α (voir § 8 des «Remarques générales») est la probabilité de rejeter l'hypothèse nulle lorsque cette hypothèse est vraie.

2) U désigne la variable normale réduite; la valeur u_α est définie par :

$$P\,[U < u_\alpha] = \alpha$$

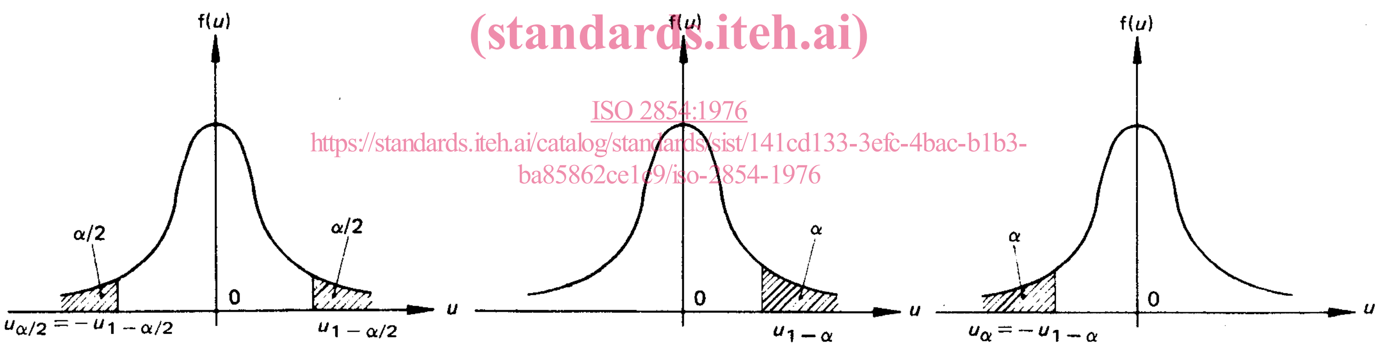
La loi de U étant symétrique par rapport à l'origine, $u_\alpha = -u_{1-\alpha}$.

On a donc :

$$P\,[U > u_\alpha] = 1 - \alpha$$

$$P\,[-u_{1-\alpha/2} < U < u_{1-\alpha/2}] = 1 - \alpha$$

Loi de U (loi normale réduite)



Cas bilatéral

Cas unilatéraux

3) $\sigma/\sqrt{n}$ est l'écart-type de la moyenne $\bar{x}$, dans un échantillon de n observations.

4) Pour la commodité des calculs, les valeurs de $u_{1-\alpha}/\sqrt{n}$ et $u_{1-\alpha/2}/\sqrt{n}$ sont données dans la table 1 de l'annexe B, pour $\alpha = 0{,}05$ et $\alpha = 0{,}01$.

EXEMPLE : voir section deux, «Notes explicatives et exemples».

TABLEAU A′ – **Comparaison d'une moyenne à une valeur donnée** (variance inconnue)

Caractéristiques techniques de la population étudiée (5) . . . .

Caractéristiques techniques des individus prélevés (5) . . . .

Observations éliminées (6) . . . .

Données statistiques	**Calculs**
Effectif de l'échantillon : $n =$	$\bar{x} = \frac{\Sigma x}{n} =$
Somme des valeurs observées : $\Sigma x =$	$\frac{\Sigma\,(x - \bar{x})^2}{n-1} = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n-1}$
Somme des carrés des valeurs observées : $\Sigma x^2 =$	$\sigma^* = s = \sqrt{\frac{\Sigma\,(x-\bar{x})^2}{n-1}} =$
Valeur donnée : $m_0 =$	$[t_{1-\alpha}(\nu)/\sqrt{n}\,]\,s =$
Degrés de liberté : $\nu = n - 1 =$	
Niveau de signification choisi (8) : $\alpha =$	$[t_{1-\alpha/2}(\nu)/\sqrt{n}\,]\,s =$

Résultats

Comparaison de la moyenne de la population à la valeur donnée m_0 :

Cas bilatéral :

L'hypothèse de l'égalité de la moyenne de la population à la valeur donnée (hypothèse nulle) est rejetée si l'on a :

$$|\bar{x} - m_0| > [t_{1-\alpha/2}(\nu)/\sqrt{n}\,]\,s$$

Cas unilatéraux :

a) L'hypothèse selon laquelle la moyenne de la population n'est pas inférieure à m_0 (hypothèse nulle) est rejetée si l'on a :

$$\bar{x} < m_0 - [t_{1-\alpha}(\nu)/\sqrt{n}\,]\,s$$

b) L'hypothèse selon laquelle la moyenne de la population n'est pas supérieure à m_0 (hypothèse nulle) est rejetée si l'on a :

$$\bar{x} > m_0 + [t_{1-\alpha}(\nu)/\sqrt{n}\,]\,s$$

NOTE — Les références (5), (6) et (8) renvoient aux paragraphes correspondants des «Remarques générales».

Commentaires

1) Le niveau de signification α (voir § 8 des «Remarques générales») est la probabilité de rejeter l'hypothèse nulle lorsque cette hypothèse est vraie.

2) $t(\nu)$ désigne la variable de Student à $\nu = n - 1$ degré de liberté; la valeur $t_\alpha(\nu)$ est définie par :

$$P\,[t(\nu) < t_\alpha(\nu)] = \alpha$$

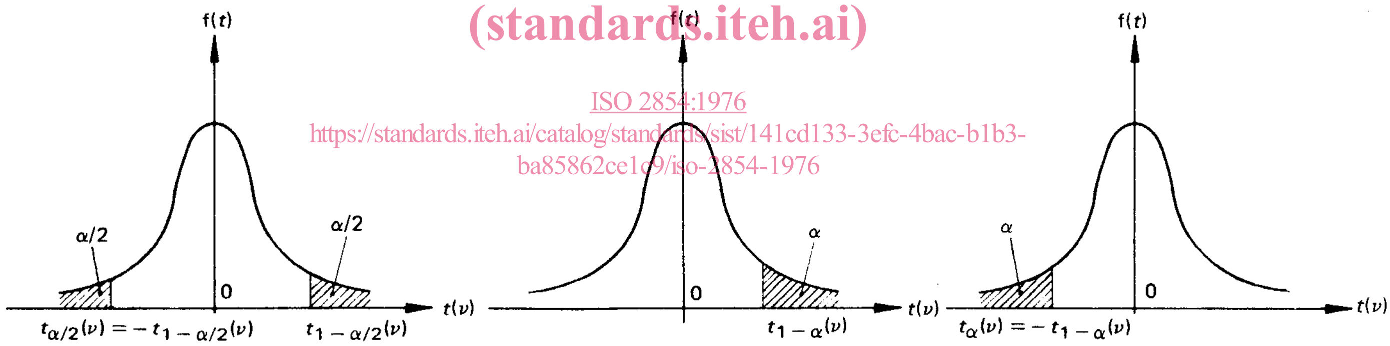
La loi de $t(\nu)$ étant symétrique par rapport à l'origine, $t_\alpha(\nu) = - t_{1-\alpha}(\nu)$.

On a donc :

$$P\,[t(\nu) > t_\alpha(\nu)] = 1 - \alpha$$

$$P\,[- t_{1-\alpha/2}(\nu) < t(\nu) < t_{1-\alpha/2}(\nu)] = 1 - \alpha$$

Loi $t(\nu)$ de Student à $\nu = n - 1$ degrés de liberté



Cas bilatéral

Cas unilatéraux

3) $\sigma^*/\sqrt{n}$ est l'écart-type estimé de la moyenne $\overline{x}$, dans un échantillon de n observations.

4) Pour la commodité des calculs, les valeurs de $t_{1-\alpha/2}(\nu)/\sqrt{n}$ et $t_{1-\alpha}(\nu)/\sqrt{n}$ sont données dans la table IIb de l'annexe B, pour $\alpha = 0{,}05$ et $\alpha = 0{,}01$.

EXEMPLE : voir section deux, «Notes explicatives et exemples».

TABLEAU B — **Estimation d'une moyenne** (variance connue)

Caractéristiques techniques de la population étudiée (5) . . . .

Caractéristiques techniques des individus prélevés (5) . . . .

Observations éliminées (6) . . . .

Données statistiques	Calculs
Effectif de l'échantillon : $n =$	$\bar{x} = \frac{\Sigma x}{n} =$
Somme des valeurs observées : $\Sigma x =$	$[u_{1-\alpha}/\sqrt{n}]\,\sigma =$
Valeur connue de la variance de la population : $\sigma^2 =$	
D'où l'écart-type : $\sigma =$	$[u_{1-\alpha/2}/\sqrt{n}]\,\sigma =$
Niveau de confiance choisi (7) : $1 - \alpha =$	

Résultats

Estimation de la moyenne m de la population :

$$m^* = \bar{x} =$$

Intervalle de confiance bilatéral :

$$\bar{x} - [u_{1-\alpha/2}/\sqrt{n}]\,\sigma < m < \bar{x} + [u_{1-\alpha/2}/\sqrt{n}]\,\sigma$$

Intervalles de confiance unilatéraux :

$$m < \bar{x} + [u_{1-\alpha}/\sqrt{n}]\,\sigma$$

ou

$$m > \bar{x} - [u_{1-\alpha}/\sqrt{n}]\,\sigma$$

NOTE — Les références (5), (6) et (7) renvoient aux paragraphes correspondants des «Remarques générales».

Commentaires

1) Le niveau de confiance $1 - \alpha$ (voir § 7 des «Remarques générales») est la probabilité pour que l'intervalle de confiance calculé renferme la valeur vraie de la moyenne.

2) U désigne la variable normale réduite; la valeur u_α est définie par :

$$P\,[U < u_\alpha] = \alpha$$

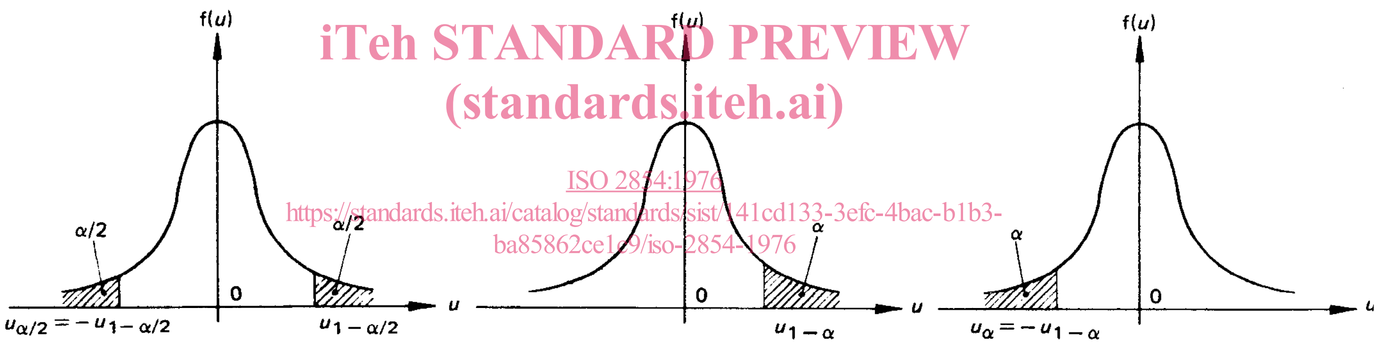
La loi de U étant symétrique par rapport à l'origine, $u_\alpha = - u_{1-\alpha}$.

On a donc :

$$P\,[U > u_\alpha] = 1 - \alpha$$

$$P\,[- u_{1-\alpha/2} < U < u_{1-\alpha/2}] = 1 - \alpha$$

Loi de U (loi normale réduite)



Cas bilatéral

Cas unilatéraux

3) $\sigma/\sqrt{n}$ est l'écart-type de la moyenne $\bar{x}$, dans un échantillon de n observations.

4) Pour la commodité des calculs, les valeurs de $u_{1-\alpha/2}/\sqrt{n}$ et $u_{1-\alpha}/\sqrt{n}$ sont données dans la table I de l'annexe B, pour $\alpha = 0{,}05$ et $= 0{,}01$.

EXEMPLE : voir section deux, «Notes explicatives et exemples».

TABLEAU B′ – **Estimation d'une moyenne** (variance inconnue)

Caractéristiques techniques de la population étudiée (5) . . . .

Caractéristiques techniques des individus prélevés (5) . . . .

Observations éliminées (6) . . . .

Données statistiques	**Calculs**
Effectif de l'échantillon : $n =$	$\bar{x} = \frac{\Sigma x}{n} =$
Somme des valeurs observées : $\Sigma x =$	$\frac{\Sigma (x - \bar{x})^2}{n-1} = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n-1} =$
Somme des carrés des valeurs observées : $\Sigma x^2 =$	$\sigma^* = s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}} =$
Degrés de liberté : $\nu = n - 1 =$	$[t_{1-\alpha}(\nu)/\sqrt{n}\,]\, s =$
Niveau de confiance choisi (7) : $1 - \alpha =$	$[t_{1-\alpha/2}(\nu)/\sqrt{n}\,]\, s =$

Résultats

Estimation de la moyenne m de la population :

$$m^* = \bar{x} =$$

Intervalle de confiance bilatéral :

$$\bar{x} - [t_{1-\alpha/2}(\nu)/\sqrt{n}\,]\, s < m < \bar{x} + [t_{1-\alpha/2}(\nu)/\sqrt{n}\,]\, s$$

Intervalles de confiance unilatéraux :

$$m < \bar{x} + [t_{1-\alpha}(\nu)/\sqrt{n}\,]\, s$$

ou

$$m > \bar{x} - [t_{1-\alpha}(\nu)/\sqrt{n}\,]\, s$$

NOTE – Les références (5), (6) et (7) renvoient aux paragraphes correspondants des «Remarques générales».

Commentaires

1) Le niveau de confiance $1 - \alpha$ (voir § 7 des «Remarques générales») est la probabilité pour que l'intervalle de confiance calculé renferme la valeur vraie de la moyenne.

2) $t(\nu)$ désigne la variable de Student à $\nu = n - 1$ degrés de liberté; la valeur $t_\alpha(\nu)$ est définie par :

$$P\,[t(\nu) < t_\alpha(\nu)] = \alpha$$

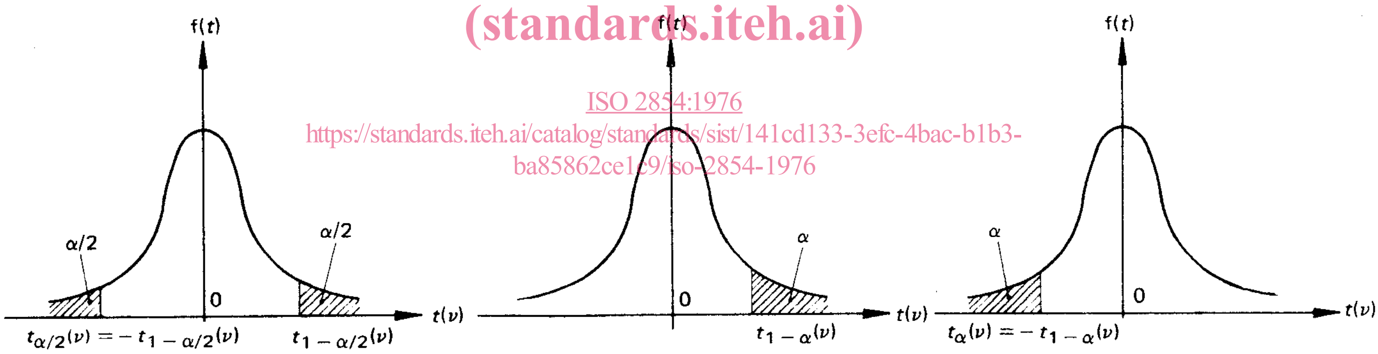
La loi de $t(\nu)$ étant symétrique par rapport à l'origine, $t_\alpha(\nu) = - t_{1-\alpha}(\nu)$.

On a donc :

$$P\,[t(\nu) > t_\alpha(\nu)] = 1 - \alpha$$

$$P\,[- t_{1-\alpha/2}(\nu) < t(\nu) < t_{1-\alpha/2}(\nu)] = 1 - \alpha$$

Loi $t(\nu)$ de Student à $\nu = n - 1$ degrés de liberté



Cas bilatéral

Cas unilatéraux

3) $\sigma^*/\sqrt{n}$ est l'écart-type estimé de la moyenne $\bar{x}$, dans un échantillon de n observations.

4) Pour la commodité des calculs, les valeurs de $t_{1-\alpha/2}(\nu)/\sqrt{n}$ et $t_{1-\alpha}(\nu)/\sqrt{n}$ sont données dans la table IIb de l'annexe B, pour $\alpha = 0{,}05$ et $\alpha = 0{,}01$.

EXEMPLE : voir section deux, «Notes explicatives et exemples».