NORME INTERNATIONALE ISO 3207

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

Interprétation statistique des données — Détermination d'un intervalle statistique de dispersion

Statistical interpretation of data — Determination of a statistical tolerance interval

Première édition — 1975-05-15

CDU 519 (083.4)

Réf. nº : ISO 3207-1975 (F)

Descripteurs : analyse statistique, intervalle statistique de dispersion, table de données, calcul.

Prix basé sur 15 pages

ISO 3207-1975 (F)

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 3207 a été établie par le Comité Technique ISO/TC 69, *Application des méthodes statistiques,* et soumise aux Comités Membres en novembre 1973.

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

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

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

Suède
U.S.A.

Imprimé en Suisse

SOMMAIRE

Page

SECTION UN : PRÉSENTATION FORMELLE DES RÉSULTATS 1

— **Remarques générales** 1

— **Tables**

Table 1 — Intervalle statistique de dispersion unilatéral (variance connue) 2

Table 2 — Intervalle statistique de dispersion bilatéral (variance connue) 3

Table 3 — Intervalle statistique de dispersion unilatéral (variance inconnue) 4

Table 4 — Intervalle statistique de dispersion bilatéral (variance inconnue) 5

SECTION DEUX : EXEMPLES 6

Annexes

A Cas d'une distribution quelconque 9

B Tables statistiques 11

Table 5 — Intervalle statistique de dispersion unilatéral (variance connue) Valeurs du coefficient k_1 $(n, p, 1-\alpha)$ 11

Table 6 — Intervalle statistique de dispersion bilatéral (variance connue) Valeurs du coefficient k'_1 $(n, p, 1-\alpha)$ 12

Table 7 — Intervalle statistique de dispersion unilatéral (variance inconnue) Valeurs du coefficient k_2 $(n, p, 1-\alpha)$ 13

Table 8 — Intervalle statistique de dispersion bilatéral (variance inconnue) Valeurs du coefficient k'_2 $(n, p, 1-\alpha)$ 14

Table 9 — Intervalles statistiques non paramétriques unilatéraux de dispersion — Effectif n de l'échantillon pour une proportion p au niveau de confiance $1-\alpha$ 15

Table 10 — Intervalles statistiques non paramétriques bilatéraux de dispersion — Effectif n de l'échantillon pour une proportion p au niveau de confiance $1-\alpha$ 15

Page blanche

Interprétation statistique des données — Détermination d'un intervalle statistique de dispersion

SECTION UN : PRÉSENTATION FORMELLE DES RÉSULTATS

REMARQUES GÉNÉRALES

1) La présente Norme Internationale spécifie des méthodes permettant, à partir d'un échantillon, de déterminer un intervalle statistique de dispersion, c'est-à-dire un intervalle contenant, avec une probabilité fixée (niveau de confiance) au moins une fraction p de la population dont provient l'échantillon. L'intervalle statistique de dispersion peut être bilatéral ou unilatéral. Les limites de l'intervalle s'appellent «limites statistiques de dispersion»; elles portent aussi le nom de «limites naturelles du processus».

2) Ces méthodes ne sont applicables que si l'on peut admettre que, dans la population considérée, les individus de l'échantillon ont été prélevés au hasard et sont indépendants.

3) Les méthodes décrites ci-après ne sont également applicables qu'à la condition que la distribution du caractère étudié soit normale. La nécessité de cette condition de normalité est plus importante ici, car elle infère davantage que pour des moyennes et des différences entre moyennes exposées dans l'ISO 2854, *Interprétation statistique des données — Techniques d'estimation et tests portant sur des moyennes et des variances.*

4) Pour vérifier l'hypothèse de normalité, on utilise les méthodes exposées dans l'ISO . . ., *Interprétation statistique des données — Tests de normalité*[1].

5) Lorsque l'hypothèse de normalité doit être rejetée ou lorsque, pour une raison quelconque, on a des doutes sur sa validité, on peut envisager de transformer la variable pour la rendre normale ou d'appliquer la méthode décrite dans la remarque introductive de l'annexe A de la présente Norme Internationale.

Il existe également des méthodes qui permettent de déterminer des intervalles statistiques de dispersion pour d'autres types de distributions que des distributions normales. La description de ces méthodes n'a pas été envisagée dans la présente Norme Internationale.

6) Il est souhaitable, lors de la détermination d'un intervalle statistique de dispersion, de donner toutes les 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.

7) On ne doit procéder 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.

8) Comme il a été dit en 1), le niveau de confiance $1 - \alpha$ est la probabilité pour que l'intervalle statistique de dispersion contienne au minimum une fraction p de la population; le risque que cet intervalle contienne moins d'une fraction p de la population est α. Les valeurs les plus usuelles de $1 - \alpha$ sont 0,95 et 0,99 ($\alpha = 0{,}05$ et $0{,}01$).

Cela signifie que, si des intervalles statistiques de dispersion sont déterminés pour un grand nombre d'échantillons au niveau de confiance 0,95 par exemple, la proportion de ces intervalles qui contiendront au moins la fraction voulue de la population sera voisine de 95 %.

9) Les tables 1 et 2 correspondent au cas où l'écart-type de la population est connu (la moyenne étant inconnue), les tables 3 et 4 au cas où la moyenne et l'écart-type sont inconnus.

Lorsque la moyenne et l'écart-type, ayant respectivement les valeurs m et σ, sont connus, la distribution du caractère étudié (supposée normale) est entièrement déterminée; il y a exactement une fraction p de la population :

$$\left.\begin{array}{l} \text{— à droite de } m - u_p\sigma \\ \text{— à gauche de } m + u_p\sigma \end{array}\right\} \text{ intervalles unilatéraux}$$

— entre $m - u_{(1+p)/2}\sigma$ et $m + u_{(1+p)/2}\sigma$: intervalle bilatéral

où u_p est le fractile d'ordre p de la variable normale réduite.

Les valeurs numériques de u_p peuvent être lues, dans les cas ci-dessus, à la dernière ligne des tables 5 et 6.

10) Les calculs peuvent souvent être fortement simplifiés en effectuant sur les données un changement d'origine et/ou d'unité.

1) En préparation.

TABLE 1 — **Intervalle statistique de dispersion unilatéral** (variance connue)[1]

Caractéristiques techniques de la population étudiée[2]	
Caractéristiques techniques des individus prélevés[2]	
Observations éliminées[3]	
Données statistiques	**Calculs**
Effectif de l'échantillon : $n =$	
Somme des valeurs observées : $\Sigma x =$	$\overline{x} = \frac{\Sigma x}{n} =$
Valeur connue de la variance de la population : $\sigma^2 =$	
d'où l'écart-type : $\sigma =$	$k_1 \ (n, p, 1-\alpha) \ \sigma =$ [6]
Fraction de la population choisie pour l'intervalle statistique de dispersion[4] : $p =$	
Niveau de confiance choisi[5] : $1-\alpha =$	
$k_1 \ (n, p, 1-\alpha) =$	

Résultats

a) Intervalle unilatéral «à gauche»

Il y a une probabilité $1-\alpha$ pour qu'au minimum une fraction p de la population soit inférieure à :

$$L_s = \overline{x} + k_1 \ (n, p, 1-\alpha) \ \sigma =$$

b) Intervalle unilatéral «à droite»

Il y a une probabilité $1-\alpha$ pour qu'au minimum une fraction p de la population soit supérieure à :

$$L_i = \overline{x} - k_1 \ (n, p, 1-\alpha) \ \sigma =$$

1) Un exemple numérique est donné dans la section deux de la présente Norme Internationale : Exemple no 1.
2) Voir paragraphe 6 des Remarques générales.
3) Voir paragraphe 7 des Remarques générales.
4) Voir paragraphe 1 des Remarques générales.
5) Voir paragraphe 8 des Remarques générales.
6) Les valeurs numériques de $k_1 \ (n, p, 1-\alpha)$ peuvent être lues directement dans la table 5 pour différentes valeurs de n, et pour

$$p = 0{,}90;\ 0{,}95;\ 0{,}99$$

$$1-\alpha = 0{,}95 \text{ et } 0{,}99$$

TABLE 2 — **Intervalle statistique de dispersion bilatéral** (variance connue)[1]

Caractéristiques techniques de la population étudiée[2]	
Caractéristiques techniques des individus prélevés[2]	
Observations éliminées[3]	
Données statistiques	**Calculs**
Effectif de l'échantillon : $n =$	
Somme des valeurs observées : $\Sigma x =$	$\bar{x} = \frac{\Sigma x}{n} =$
Valeur connue de la variance de la population $\sigma^2 =$	
d'où l'écart-type : $\sigma =$	$k'_1\ (n, p, 1-\alpha)\ \sigma =$ [6]
Fraction de la population choisie pour l'intervalle statistique de dispersion[4] : $p =$	
Niveau de confiance choisi[5] : $1-\alpha =$	
$k'_1\ (n, p, 1-\alpha) =$	

Résultats

Il y a une probabilité $1-\alpha$ pour qu'au minimum une fraction p de la population soit comprise entre les limites[7] :

$$L_i = \bar{x} - k'_1\ (n, p, 1-\alpha)\ \sigma =$$

$$L_s = \bar{x} + k'_1\ (n, p, 1-\alpha)\ \sigma =$$

1) Un exemple numérique est donné dans la section deux de la présente Norme Internationale : Exemple no 2.
2) Voir paragraphe 6 des Remarques générales.
3) Voir paragraphe 7 des Remarques générales.
4) Voir paragraphe 1 des Remarques générales.
5) Voir paragraphe 8 des Remarques générales
6) Les valeurs numériques de $k'_1\ (n, p, 1-\alpha)$ peuvent être lues dans la table 6 pour différentes valeurs de n, et pour

$$p = 0{,}90;\ 0{,}95;\ 0{,}99$$
$$1-\alpha = 0{,}95 \text{ et } 0{,}99$$

7) Ces limites sont symétriques par rapport à $\bar{x}$, mais elles ne sont pas «symétriques en probabilité». Il n'est pas vrai qu'au niveau de confiance $1-\alpha$, une fraction d'au plus $(1-p)/2$ de la population est inférieure à L_i et une fraction d'au plus $(1-p)/2$ supérieure à L_s.

TABLE 3 — **Intervalle statistique de dispersion unilatéral** (variance inconnue)[1)]

Caractéristiques techniques de la population étudiée[2)] . . . .

Caractéristiques techniques des individus prélevés[2)] . . . .

Observations éliminées[3)] . . . .

Données statistiques	**Calculs**
Effectif de l'échantillon : $n =$	$\overline{x} = \dfrac{\Sigma x}{n} =$
Somme des valeurs observées : $\Sigma x =$	$\dfrac{\Sigma\,(x-\overline{x})^2}{n-1} = \dfrac{\Sigma x^2 - (\Sigma x)^2/n}{n-1} =$
Somme des carrés des valeurs observées : $\Sigma x^2 =$	
Fraction de la population choisie pour l'intervalle statistique de dispersion[4)] : $p =$	$\sigma^* = s = \sqrt{\dfrac{\Sigma\,(x-\overline{x})^2}{n-1}} =$ (estimation de l'écart-type σ)
Niveau de confiance choisi[5)] : $1-\alpha =$	
$k_2\,(n, p, 1-\alpha) =$	$k_2\,(n, p, 1-\alpha)\,s =$ [6)]

Résultats

a) Intervalle unilatéral «à gauche»

Il y a une probabilité $1-\alpha$ pour qu'au minimum une fraction p de la population soit inférieure à :

$$L_s = \overline{x} + k_2\,(n, p, 1-\alpha)\,s =$$

b) Intervalle unilatéral «à droite»

Il y a une probabilité $1-\alpha$ pour qu'au minimum une fraction p de la population soit supérieure à :

$$L_i = \overline{x} - k_2\,(n, p, 1-\alpha)\,s =$$

1) Un exemple numérique est donné dans la section deux de la présente Norme Internationale : Exemple n° 3.
2) Voir paragraphe 6 des Remarques générales.
3) Voir paragraphe 7 des Remarques générales.
4) Voir paragraphe 1 des Remarques générales.
5) Voir paragraphe 8 des Remarques générales.
6) Les valeurs numériques de $k_2\,(n, p, 1-\alpha)$ peuvent être lues dans la table 7 pour différentes valeurs de n, et pour

$$p = 0{,}90;\ 0{,}95;\ 0{,}99$$

$$1-\alpha = 0{,}95 \text{ et } 0{,}99$$

TABLE 4 — **Intervalle statistique de dispersion bilatéral** (variance inconnue)[1]

Caractéristiques techniques de la population étudiée[2] . . . .

Caractéristiques techniques des individus prélevés[2] . . . .

Observations éliminées[3] . . . .

Données statistiques	**Calculs**
Effectif de l'échantillon : $n =$	$\overline{x} = \frac{\Sigma x}{n} =$
Somme des valeurs observées : $\Sigma x =$	$\frac{\Sigma (x - \overline{x})^2}{n-1} = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n-1} =$
Somme des carrés des valeurs observées : $\Sigma x^2 =$	
Fraction de la population choisie pour l'intervalle statistique de dispersion[4] : $p =$	$\sigma^* = s = \sqrt{\frac{\Sigma (x - \overline{x})^2}{n-1}} =$ (estimation de l'écart-type σ)
Niveau de confiance choisi[5] : $1 - \alpha =$ $k'_2\ (n, p, 1 - \alpha) =$	$k'_2\ (n, p, 1 - \alpha)\ s =$ [6]

Résultats

Il y a une probabilité $1 - \alpha$ pour qu'au minimum une fraction p de la population soit comprise entre les limites[7]

$$L_i = \overline{x} - k'_2\ (n, p, 1 - \alpha)\ s =$$

$$L_s = \overline{x} + k'_2\ (n, p, 1 - \alpha)\ s =$$

1) Un exemple numérique est donné dans la section deux de la présente Norme Internationale : Exemple no 4.
2) Voir paragraphe 6 des Remarques générales.
3) Voir paragraphe 7 des Remarques générales.
4) Voir paragraphe 1 des Remarques générales.
5) Voir paragraphe 8 des Remarques générales.
6) Les valeurs numériques de $k'_2\ (n, p, 1 - \alpha)$ peuvent être lues dans la table 8 pour différentes valeurs de n, et pour

$$p = 0{,}90;\ 0{,}95;\ 0{,}99$$

$$1 - \alpha = 0{,}95 \text{ et } 0{,}99$$

7) Ces limites sont symétriques par rapport à $\overline{x}$, mais elles ne sont pas «symétriques en probabilité». Il n'est pas vrai qu'au niveau de confiance $1 - \alpha$, une fraction d'au plus $(1 - p)/2$ de la population est inférieure à L_i et une fraction d'au plus $(1 - p)/2$ supérieure à L_s.

SECTION DEUX : EXEMPLES

REMARQUES INTRODUCTIVES

On illustrera les tables 1 à 4 par des exemples utilisant les valeurs numériques de l'ISO 2854 (section deux, paragraphe 1 des «Remarques introductives», table X, fil 2) : 12 mesures de la charge de rupture de fil de coton. Il faut noter que le nombre d'observations $n = 12$ retenu pour ces exemples est notablement inférieur à celui préconisé dans l'ISO 2062, *Textiles — Fils sur enroulements — Méthode de détermination de la force de rupture et de l'allongement de rupture du fil individuel — (Appareils à vitesse constante d'accroissement de force, d'allongement ou de déplacement de la pince de traction).*

Le centinewton est l'unité de mesure dans laquelle sont exprimés les données numériques et les résultats des calculs des différents exemples.

x
228,6
232,7
238,8
317,2
315,8
275,1
222,2
236,7
224,7
251,2
210,4
270,7

Ces mesures proviennent d'un lot de 12 000 bobines, d'une même fabrication, emballées dans 120 boîtes de 100 bobines chacune. Douze boîtes ont été prises au hasard dans le lot et une bobine a été retirée au hasard de chacune de ces boîtes. Des éprouvettes de 50 cm de longueur de fil ont été découpées de ces bobines, à environ 5 m de distance de l'extrémité libre. Les essais proprement dits ont été effectués sur les parties centrales de ces éprouvettes. Des informations antérieures permettent d'admettre que la distribution des charges de rupture, mesurées dans ces conditions, est pratiquement normale.

À partir de ces résultats, on trouve :

Effectif de l'échantillon :

$n = 12$

Somme des valeurs observées :

$\Sigma x = 3\,024{,}1$

Moyenne :

$$\overline{x} = \frac{\Sigma x}{n} = 252{,}0$$

Somme des carrés des valeurs observées :

$\Sigma x^2 = 775\,996{,}09$

Somme des carrés des différences avec la valeur moyenne :

$$\Sigma\,(x - \overline{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 13\,897{,}69$$

Estimation de la variance :

$$s^2 = \frac{\Sigma\,(x - \overline{x})^2}{n - 1} = 1\,263{,}4$$

Estimation de l'écart-type :

$s = 35{,}5$

On sait d'autre part, par expérience, qu'à l'intérieur d'un même lot, les charges de rupture sont distribuées suivant une loi très voisine d'une loi normale.

La présentation formelle des calculs ne sera faite que pour la table 3 (intervalle unilatéral, variance inconnue).

EXEMPLES NUMÉRIQUES

Exemple n° 1 — **Intervalle statistique de dispersion unilatéral** (variance connue, table 1)

On suppose que des mesures antérieures ont montré que la dispersion est constante d'un lot à l'autre du même fournisseur, bien que la moyenne ne le soit pas, et est représentée par un écart type $\sigma = 33{,}15$.

On veut calculer la limite L_i telle qu'on puisse affirmer, avec un niveau de confiance $1 - \alpha = 0{,}95$, que dans une proportion au moins égale à 0,95 (95 %), la charge de rupture des éléments susceptibles d'être prélevés dans le lot et mesurés dans les mêmes conditions est supérieure à L_i.

La table 5 donne :

$$k_1\ (12;\,0{,}95;\,0{,}95) = 2{,}12$$

D'où :

$$L_i = \overline{x} - k_1 \sigma = 252{,}0 - 2{,}12 \times 33{,}15 = 181{,}7$$

On trouverait naturellement une limite L_i plus faible si l'on prenait une fraction plus élevée de la population (par exemple 99 %) et/ou un niveau de confiance plus élevé (par exemple $1 - \alpha = 0{,}99$).

Exemple n° 2 — **Intervalle statistique de dispersion bilatéral** (variance connue, table 2)

Dans les mêmes conditions que dans l'exemple n° 1, on veut calculer les limites L_i et L_s telles qu'on puisse affirmer, avec un niveau de confiance $1 - \alpha = 0{,}95$, que, pour une fraction du lot au moins égale à $p = 0{,}90$ (90 %), la charge de rupture est comprise entre L_i et L_s. La table 6 donne :

$$k'_1\ (12;\,0{,}90;\,0{,}95) = 1{,}89$$

D'où :

$$L_i = \overline{x} - k'_1\,\sigma = 252{,}0 - 1{,}89 \times 33{,}15 = 189{,}3$$

$$L_s = \overline{x} + k'_1\,\sigma = 252{,}0 + 1{,}89 \times 33{,}15 = 314{,}7$$

On devra éviter de croire que, à gauche de L_i, ne se trouvent que 5 % (au maximum) de la population et de même à droite de L_s.

On a d'ailleurs vu, à propos de l'exemple n° 1, que la limite à gauche de laquelle il n'y a que 5 % au maximum de la population (95 % au minimum à droite de L_i) est $L_i = 181{,}7$.

Exemple n° 3 — **Intervalle statistique de dispersion unilatéral** (variance inconnue, table 3)

On suppose maintenant que l'écart-type de la population est inconnu et doit être estimé à partir de l'échantillon. On se donnera les mêmes conditions que lorsque l'écart-type est connu (exemple n° 1), soit $p = 0{,}95$ et $1 - \alpha = 0{,}95$.

La présentation formelle des données numériques est donnée ci-dessous.

Caractéristiques techniques de la population étudiée : le lot consiste en une fourniture de fil de coton reçue, le 1969-08-03, du fournisseur H, et composée de 12 000 bobines emballées dans 120 boîtes de 100 bobines chacune.

Caractéristiques techniques des individus prélevés : 12 boîtes ont été prises au hasard dans le lot et une bobine a été retirée au hasard de chacune de ces boîtes. Des éprouvettes de 50 cm de longueur de fil ont été découpées de ces bobines, à environ 5 m de distance de l'extrémité libre. Les essais proprement dits ont été effectués sur les parties centrales de ces éprouvettes.

Observations éliminées : aucune.

Données statistiques

Effectif de l'échantillon :

$n = 12$

Somme des valeurs observées :

$\Sigma x = 3\,024{,}1$

Somme des carrés des valeurs observées :

$\Sigma x^2 = 775\,996{,}09$

Fraction de la population choisie pour l'intervalle statistique de dispersion :

$p = 0{,}95$ (95 %)

Niveau de confiance choisi :
$1 - \alpha = 0{,}95$
$k_2\ (12; 0{,}95; 0{,}95) = 2{,}74$
(valeur lue dans la table 7)

Calculs

$$\overline{x} = \frac{3\,024{,}1}{12} = 252{,}0$$

$$\frac{\Sigma\,(x - \overline{x})^2}{n-1} = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n-1}$$

$$= \frac{775\,996{,}09 - (3\,024{,}1)^2/12}{11}$$

$$= 1\,263{,}4$$

$$\sigma^* = s = \sqrt{1\,263{,}4} = 35{,}5$$

$$k_2\ (12; 0{,}95; 0{,}95)\ s = 97{,}3$$

Résultats

Au niveau de confiance 0,95, on peut affirmer qu'une fraction au moins égale à 0,95 (95 %) des charges de rupture du lot est supérieure à :

$$L_i = 252{,}0 - 2{,}74 \times 35{,}5 = 154{,}7$$

On notera que la valeur de L_i est plus faible que dans l'exemple n° 1 (variance connue) parce que l'emploi de s, estimation aléatoire de σ, entraîne une valeur plus élevée du coefficient k (2,74 au lieu de 2,12).