

## Éléments statistiques

*La connaissance de la qualité des données, en sécurisant l'utilisateur, incite davantage à leur réutilisation.*

*Ce décryptage de la norme ISO 19157 a pour vocation de donner un cadre méthodologique pour qualifier les données lors de leur diffusion.*

**L'essor des données ouvertes et géolocalisées et la profusion d'usages existant et à venir nous rend tous progressivement producteur et utilisateur de données géographiques.**

**Les activités régaliennes ou les politiques publiques s'appuient sur de l'information maîtrisée où la qualité des données produites ou utilisées devient un entrant indispensable. Pour autant, tout le monde ne dispose pas des moyens des producteurs institutionnels de données et il paraît utile de fournir des recommandations et des méthodes plus adaptées au contexte de chacun, pour qualifier les données géographiques, communiquer sur les résultats obtenus voire savoir les interpréter. C'est l'objectif que s'est fixé le Cerema en proposant cette collection de fiches, à l'interface des productions et des usages.**

**Les concepts liés à la qualité des données nécessitent des connaissances accrues et abordent des domaines techniques parfois ardu, dont la statistique fait partie. Cette fiche a pour objet de rappeler quelques notions de statistique (variables aléatoires, échantillon, incertitude, niveau de confiance...) nécessaires à la compréhension de la qualité des données géographiques.**

### 1. Préable

Cette fiche aborde quelques notions statistiques utiles à l'évaluation de la qualité d'un jeu de données géographiques. Dans un premier temps, elle rappelle quelques définitions : variables aléatoires, population, échantillon, paramètres de tendance centrale (moyenne, médiane) et paramètres de dispersion (variance, écart-type), etc.

Les paragraphes suivants introduisent les notions d'incertitude et de niveau de confiance. Par exemple, la mesure d'une moyenne basée sur un échantillon permet seulement d'estimer la valeur réelle qui serait mesurée sur

l'intégralité du lot de données. Cette approximation peut être quantifiée par le calcul de l'incertitude et par sa propagation, ce qui permet de déterminer un intervalle dans lequel se trouve la valeur réelle, avec un certain niveau de confiance.

La dernière partie fournit quelques éléments utiles à l'évaluation de la conformité d'un échantillon par rapport à sa représentativité de l'ensemble du lot de données en vue d'évaluations basées sur des dénombrements et sur des écarts-types.



## 2. Les différents types de variables aléatoires

Par certains aspects, l'évaluation de la qualité d'un lot de données passe par des approches statistiques. Une maîtrise de quelques notions de base de cette discipline est nécessaire. C'est l'objet de ce chapitre que de rappeler d'une manière abordable ces quelques notions.

### 2.1 Les variables quantitatives

Ce sont des variables qui prennent des valeurs numériques. Elles sont de deux types :

#### ■ Les variables discrètes

Elles ne prennent leurs valeurs que dans un ensemble de valeurs disjointes ou séparées (par exemple : les nombres entiers).

**Exemple :** le nombre de voies d'une route.

#### ■ Les variables continues

Elles peuvent prendre un nombre infini de valeurs réelles dans un intervalle (par exemple :  $[-2, 5]$  ou  $]-\infty, +\infty[$ ). C'est le cas de la plupart des mesures de longueur, surfaces, volume...

**Exemple :** la largeur d'une route..

### 2.2 Les variables qualitatives

Elles se rapportent à une propriété non quantifiable. Par exemple, une couleur, l'appartenance à une espèce, un type d'objet, un toponyme, etc. Elles peuvent être de deux natures :

#### ■ Les variables nominales

Une variable nominale est une variable qualitative dont les modalités ne sont pas ordonnées.

**Exemple :** la nature d'un document d'urbanisme : POS, PLU, CC, PSMV, SCOT

#### ■ Les variables ordinales

À défaut d'être mesurables, les variables ordinales sont classables. Elles peuvent être classées de façon naturelle ou par ordre croissant ou décroissant (froid, tiède, chaud).

**Exemple :** les échelons territoriaux : commune, canton, arrondissement, département, région, état

## 3. Population

### 3.1 Population

La population est l'ensemble des éléments qui constituent le jeu de données que l'on désire évaluer. Le nombre d'éléments de la population est noté  $N$ .

### 3.2 Échantillon

Un échantillon est un sous-ensemble d'éléments prélevé dans une population et destiné à fournir des renseignements sur cette population. Le nombre d'éléments de l'échantillon est noté  $n$ .

## 4. Les paramètres de tendance centrale et de position

Pour une série de nombres, les paramètres de tendance centrale et de position permettent de voir comment se positionnent les données autour d'une valeur centrale. Trois paramètres permettent de définir cette tendance centrale.

### 4.1 La moyenne arithmétique

La moyenne arithmétique notée  $\bar{x}$  (pour un échantillon  $n$ ) et  $\mu$  (pour une population  $N$ ) est obtenue en divisant la somme des données par le nombre de données.

#### Exemple

Pour la série : 2, 6, 7, 9, 11, la moyenne arithmétique est :  $\bar{x} = (2 + 6 + 7 + 9 + 11)/5 = 35/5 = 7$

La formule générale pour un échantillon comprenant  $n$  éléments est : 
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

### 4.2 La médiane

Pour une série ordonnée de  $n$  nombres d'un échantillon, la médiane  $M$  est le nombre qui se trouve au milieu, tel qu'il y ait autant de nombres avant qu'après.

### Exemple

série : 2, 3, 7, 9, 11       $M = 7$   
série : 2, 3, 6, 7       $M = (3 + 6)/2 = 4,5$

## 4.3 Le mode

Le mode d'une série de nombres est celui apparaissant le plus fréquemment.

**Exemple :** dans la série 4, 3, 1, 5, 2, 1, 6, le mode est 1.

## 5. Les paramètres de dispersion

Les paramètres de dispersion permettent d'évaluer les écarts des données par rapport à la moyenne arithmétique du jeu de données.

### 5.1 L'étendue

Pour un ensemble de données numériques, l'étendue est la différence entre la plus grande et la plus petite valeur.

**Exemple :** deux séries de nombres ayant la même moyenne mais une étendue très différente

série : -10, 0, 10, 20, 30  
étendue :  $30 - (-10) = 40$   
série : 8, 9, 10, 11, 12  
étendue :  $12 - 8 = 4$

### 5.2 La variance

La variance est une mesure servant à caractériser la dispersion d'un échantillon. Pour calculer la variance d'une série statistique ou d'une variable aléatoire, on calcule les écarts entre les différentes valeurs de la série et sa moyenne puis on calcule la moyenne de ces écarts élevés au carré. (Source wikipédia.)

La variance d'un échantillon de  $n$  valeurs, notée  $v$ , est donc la moyenne des carrés des écarts à la moyenne des données de l'échantillon.

Pour un échantillon de  $n$  éléments, la variance s'exprime ainsi :

$$v = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

**Remarque :** pourquoi  $n-1$  et pas  $n$  ?

Contrairement à la moyenne, la variance et l'écart-type présentent naturellement un biais s'ils sont calculés sur un échantillon aléatoire. Intuitivement, on devine que la dispersion globale est sous-estimée sur un échantillon. Pour corriger cet effet, la variance et l'écart-type sont calculés avec les «  $n-1$  » éléments de l'échantillon au dénominateur. On l'appelle écart-type non biaisé. En conservant  $n$  au lieu de  $(n-1)$ , on obtient la variante dite écart-type biaisé.

### 5.3 L'écart-type

L'écart-type d'un échantillon de  $n$  valeurs, notée  $s$ , est la racine carrée de la variance. L'écart-type d'une population est notée  $\sigma$ .

$$s = \sqrt{v} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

## 6. Incertitude

Les mesures réalisées sur des échantillons demeurent des estimations. Pour compléter les valeurs affectées aux estimations, il est indispensable de leur adjoindre l'incertitude de la mesure.

### 6.1 Incertitude sur une moyenne

L'incertitude type sur une moyenne, notée  $\Delta \bar{x}$ , s'exprime en fonction de l'écart-type  $\sigma$  :

$$\Delta \bar{x} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}$$

En résumé, si on réalise  $n$  mesures de  $x$  avec les résultats  $x_1, x_2 \dots x_n$ , on écrira le résultat final sous

la forme :  $x = \bar{x} \pm \frac{\sigma}{\sqrt{n}}$  où  $\bar{x}$  et  $\frac{\sigma}{\sqrt{n}}$

sont les meilleures estimations de la valeur vraie et de l'incertitude-type.

L'incertitude-type est correcte dans le cas d'un grand nombre de mesures mais devient imprécise si ce nombre est faible. L'application du coefficient de Student t (qui dépend du nombre de mesures pour un intervalle de confiance donné) permet de tenir compte de cet effet et d'obtenir la nouvelle mesure suivante :

$$x = \bar{x} \pm t \frac{\sigma}{\sqrt{n}}$$

où t est appelé **coefficient de Student**.

### Exemple

Pour un niveau de confiance de 95 % et un nombre de mesures égal à 30,  $t = 2,04$

n	2	3	4	5	6	7	8	9	10
t (95%)	12.7	4.30	3.18	2.78	2.57	2.45	2.37	2.31	2.26
t (99%)	63.7	9.93	5.84	4.60	4.03	3.71	3.50	3.36	3.25

n	12	14	16	18	20	30	50	100	∞
t (95%)	2.20	2.16	2.13	2.11	2.09	2.04	2.01	1.98	1.96
t (99%)	3.11	3.01	2.95	2.90	2.86	2.76	2.68	2.63	2.57

Coefficients de Student fonction du nombre de mesures et de l'intervalle de confiance 95 % et 99 %

## 6.2 Incertitude sur un pourcentage ou proportion

L'incertitude sur un pourcentage ( $\Delta a/a$ ) n'est pas liée à l'écart-type. Elle est fonction du pourcentage évalué, de la taille de l'échantillon et du coefficient de Student.

Soit  $P_0$  un pourcentage observé sur un échantillon de n observations.  $Q_0 = 1 - P_0$  est le pourcentage complémentaire à  $P_0$ . L'incertitude sur la population entière est définie par :

$$t \frac{\sqrt{P_0 \cdot Q_0}}{n} = t \frac{\sqrt{P_0 \cdot (1 - P_0)}}{n}$$

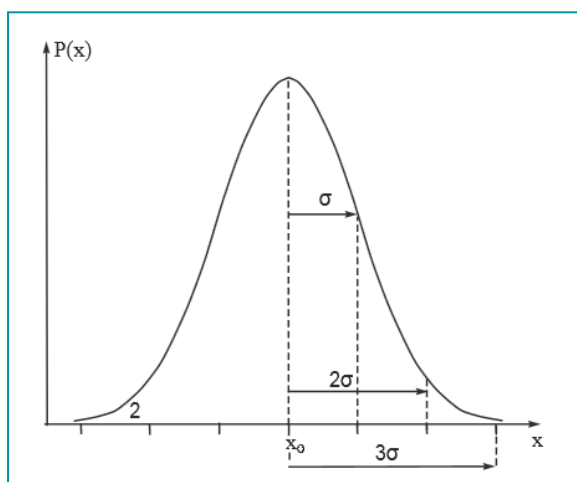
**Exemple** pour un taux d'exhaustivité mesuré de 92 % sur un échantillon de 50 éléments avec un niveau de confiance de 95 % :

$$92 \% \pm 2,01 \sqrt{\frac{(0,92 \times 0,08)}{50}}$$

soit 92 %  $\pm$  8 %

## 7. Incertitude

Lorsque le nombre n des mesures augmente, la distribution des mesures tend vers une courbe gaussienne<sup>1</sup>. Le graphe ci-après représente une courbe de Gauss, aussi appelée « gaussienne » : P(x) est la probabilité pour que la mesure d'une valeur sur un échantillon soit x.  $x_0$  est la vraie valeur sur l'ensemble de la population.



Ainsi, la vraie valeur  $x_0$  recherchée est fonction du nombre de mesures réalisées, de la moyenne obtenue et de l'écart-type calculé avec un certain **niveau de confiance**.

$$68 \% \quad \bar{x} - \sigma < x_0 < \bar{x} + \sigma$$

$$95,4 \% \quad \bar{x} - 2\sigma < x_0 < \bar{x} + 2\sigma$$

$$99,7 \% \quad \bar{x} - 3\sigma < x_0 < \bar{x} + 3\sigma$$

Par exemple, un niveau de confiance de **95 %** exprime la probabilité que la vraie valeur  $x_0$  se trouve à l'intérieur de l'intervalle de confiance  $\pm 2\sigma$

Considérant un niveau de confiance 95 %, une moyenne de l'échantillon de 10 et un écart-type de 2, nous dirons que la valeur estimée est comprise entre  $10 \pm 2 \cdot 2$ , à savoir entre 6 et 14. Si nous choisissons un niveau de confiance plus élevé de 99,7 %, nous dirions qu'elle se situerait entre  $10 \pm 3 \cdot 2$ , donc entre 4 et 16.

La moyenne arithmétique calculée sur un échantillon est alors une estimation de cette valeur avec une incertitude  $\bar{x} \pm t \frac{\sigma}{\sqrt{n}}$

**Remarque :** la valeur de 95 % est fréquente dans de nombreuses études statistiques et nous retiendrons plus particulièrement cette valeur pour les mesures liées aux critères qualité faisant référence à ces méthodes.

1 Il s'agit d'une loi de probabilité appelée « loi normale ».

## 8. Propagation des incertitudes

### 8.1 Variation d'une fonction à une variable

Soit une mesure  $x \pm \Delta x$  et  $y = f(x)$  une fonction de  $x$ . Quelle est alors l'incertitude sur  $y$  ?

Lorsque  $\Delta x$  est petit,  $f(x)$  est remplacé au voisinage de  $x$  par sa tangente (on calcule alors la dérivée de la fonction  $f$ ) :

$$\Delta y = \left| \frac{df}{dx} \right| \Delta x$$

On fait l'approximation  $\Delta y = dy$  valable si  $\Delta x$  est petit et  $f(x)$  varie « lentement ».

### 8.2 Incertitude d'une fonction à plusieurs variables

Supposons que  $y$  dépende de plusieurs grandeurs  $x$ ,  $z$  et  $t$  mesurées avec les incertitudes  $\Delta x$ ,  $\Delta z$ ,  $\Delta t$  :

$$Y = f(x, z, t)$$

L'erreur maximum possible sur  $y$  est alors :

$$\Delta y = \left| \frac{df}{dx} \right| \Delta x + \left| \frac{df}{dz} \right| \Delta z + \left| \frac{df}{dt} \right| \Delta t$$

**En pratique**, avec par exemple deux variables :

$$\text{L'addition : } y = x + z \quad \Delta y = \Delta x + \Delta z$$

$$\text{La soustraction : } y = x - z \quad \Delta y = \Delta x + \Delta z$$

**Dans les cas de l'addition et de la soustraction, les incertitudes absolues s'ajoutent.**

$$\text{La multiplication : } y = xz \quad \Delta y = z\Delta x + x\Delta z$$

$$\text{La division : } y = x/z \quad \Delta y = (z\Delta x + x\Delta z) / z^2$$

on a alors dans le cas de la multiplication et de la division :

$$\frac{\Delta y}{y} = \frac{\Delta x}{x} + \frac{\Delta z}{z}$$

**Dans les cas de la multiplication et de la division, les incertitudes relatives (proportions) s'ajoutent.**

## 9. Échantillonnage

L'échantillonnage est basé sur des probabilités et implique une sélection aléatoire des éléments de l'échantillon (voir la fiche méthode d'échantillonnage). La norme ISO 19157 propose :

- une méthode d'évaluation des éléments conformes/non-conformes dans un échantillon, elle est par exemple applicable aux critères d'exhaustivité et de précision thématique ;
- une méthode d'évaluation où intervient une grandeur mesurable avec écart-type dans un échantillon, elle est par exemple applicable au critère de précision de position ou dans tout autre critère où intervient une valeur numérique plus ou moins précise.

### 9.1 Évaluation des éléments conformes/non-conformes dans un échantillon

L'objectif est de déterminer en fonction de la taille d'une population et de celle de l'échantillon si le nombre d'objets non conformes ou manquants permet de valider ou non le jeu de données contrôlé.

La méthode proposée par la norme s'appuie sur 4 paramètres :

- $N$  : la taille de la population, correspondant au nombre d'éléments dans la base ;
- $n$  : la taille de l'échantillon que l'on prélève et sur lequel sont effectués les contrôles ;

- $p_0$  : le taux de rejet acceptable ou **limite d'acceptation de la qualité** (LAQ) éventuellement précisée dans les spécifications (0,5 % d'erreur, 1 % d'erreur, etc.) ;
- NC : le niveau de confiance. Il est généralement fixé à 95 %.

Le tableau 1 regroupe les valeurs usuelles de ces quatre paramètres. Il s'utilise ainsi pour :

1. évaluer la taille de la population à contrôler ( $N$ ) ;
2. évaluer la taille ( $n$ ) d'un échantillon à constituer ;
3. réaliser le contrôle et compter le nombre d'éléments non conformes ;
4. l'ensemble de la population est rejeté si le nombre d'éléments non conformes est égal ou supérieur à la limite de rejet du tableau selon la limite d'acceptation ( $p_0$ ) désirée.

L'intérêt de ce tableau réside, dans un premier temps, dans ses deux premières colonnes qui nous permettent d'évaluer la taille d'un échantillon en fonction de la taille de la population à évaluer pour un niveau de confiance de 95 %.

Le nombre de rejets acceptable selon les limites d'acceptation de la qualité (LAQ) logiquement définies dans les spécifications ou dans un cahier des charges présente surtout un intérêt quand il s'agit de spécifier la qualité du livrable suite à une commande.

Niveau de confiance = 95 %			$p_0$ = LAQ (limite d'acceptation de la qualité)					
Taille de la population (N)		Taille de l'échantillon (n)	0,5 %	1,0 %	2,0 %	3,0 %	4,0 %	5,0 %
De	à		Limite de rejet					
1	8	Toutes	1	1	1	1	1	1
9	50	8	1	1	1	2	2	2
51	90	13	1	1	2	2	2	3
91	150	20	1	2	2	3	3	4
151	280	32	1	2	3	3	4	4
281	400	50	2	3	3	4	5	6
401	500	60	2	3	4	5	6	7
501	1200	80	3	3	5	6	7	8
1 201	3200	125	3	4	6	8	10	11
3 201	10 000	200	4	6	8	11	14	16
10 001	13 000	315	5	7	12	16	20	23
13 001	15 000	500	6	10	16	23	28	34
15 001	50 000	800	9	14	24	33	42	51
	> 50 000	1250	12	20	34	49	63	76

Tableau 1

**Remarque :** le tableau 1, proposé dans la norme est construit sur la loi hypergéométrique. Cette loi de probabilité permet de décrire l'expérience suivante : on dispose d'une urne contenant un nombre N de boules noires ou blanches. On connaît la proportion  $p_0$  de boules noires dans l'urne. On tire « n » boules dans l'urne, sans remise. La loi hypergéométrique de paramètres (N,  $p_0$ , n), nous donne alors la probabilité d'avoir tiré X boules noires dans l'urne.

#### ■ Fourniture d'un tableau de calcul

Nous fournissons un tableau de calcul « *Calcul\_loi\_hypergéométrique.xls* » capable de personnaliser cette méthode. Basé sur la loi hypergéométrique, il utilise comme paramètres :

- la taille de la population (N) ;
- la taille de l'échantillon (n) ;
- la limite d'acceptation de la qualité ( $p_0$ ).

À partir de ces trois valeurs, le tableau indique la limite de rejet en fonction du niveau de confiance.

Taille de la population	50000
Taille de l'échantillon	800
Limite d'acceptation $p_0$	0,01
Limite de rejet	Niveau de confiance
0	0,0%
1	0,3%
2	1,3%
3	4,1%
4	9,7%
5	18,8%
6	31,0%
7	45,1%
8	59,3%
9	71,8%
10	81,9%
11	89,1%
12	93,9%
13	96,8%
14	98,4%
15	99,3%

Si l'on souhaite par exemple travailler avec un niveau de confiance de 95 %, il suffit de trouver la valeur de niveau de confiance la plus proche de 95 % ou légèrement supérieure, et d'identifier la limite de rejet correspondante.

**Exemple :** on contrôle l'exhaustivité des bâtiments dans un jeu de données en comprenant 2 440.

L'échantillon devra en contenir : 125 d'après le tableau.

Le contrôle qualité montre qu'il manque 2 bâtiments soit un taux d'omission de 2/125 égal à 1,6 %.

Pour une limite d'acceptation de la qualité de 0,5 %, le tableau propose un nombre de 3 bâtiments manquants comme limite de rejet. (10 bâtiments si la LAQ était fixée à 4 %).

Même si le taux d'omission mesuré (1,6 %) est supérieur à la limite d'acceptation de la qualité (0,5 %), c'est le nombre d'omissions effectives (égal à 2) qui primera. Ce nombre étant inférieur à la limite de rejet égale à 3, le lot de données ne sera pas rejeté.

## 9.2 Évaluation pour un contrôle qualité avec un écart-type

La seconde méthode proposée par la norme ISO 19157 utilisant un tableau basé sur une loi statistique, concerne un contrôle qualité à partir d'un écart-type.

Elle permet de contrôler une grandeur mesurable en posant une limite d'acceptation de la qualité (LAQ). Cette méthode s'appuie sur la loi de Fisher.

## Le principe de cette méthode est le suivant :

Considérons une donnée numérique mesurable comme une variable aléatoire qui suit une loi normale avec un certain écart-type  $\sigma$ . Plus  $\sigma$  est petit, meilleure est la qualité. Dans la démarche qualité, on met en place une borne limite d'acceptation de la qualité (LAQ) sur cet écart-type  $\sigma$  en posant que si  $\sigma > LAQ$ , la qualité n'est pas reconnue.

Comme on ne peut pas contrôler la totalité des  $N$  données de la population, on sélectionne un échantillon aléatoire de  $n$  données et on calcule l'écart-type  $s$  sur les  $n$  données de l'échantillon.

### La loi de Fisher :

Le rapport des variances  $s^2/\sigma^2$  suit alors une loi statistique appelée loi de Fisher de paramètres  $(n-1, N-1)$ . On se réfère alors à une table de Fisher qui indique pour  $n$  et  $N$ , la valeur maximale acceptable du rapport  $s^2/LAQ^2$  pour un seuil de rejet de 0,05 % (soit un niveau de signification de 95 %). ([https://fr.wikipedia.org/wiki/Loi\\_de\\_Fisher#Table\\_de\\_Fisher-Snedecor](https://fr.wikipedia.org/wiki/Loi_de_Fisher#Table_de_Fisher-Snedecor))

Si le rapport  $s^2/LAQ^2$  dépasse la valeur donnée dans la table avec un niveau de signification de 95 %, alors notre hypothèse est rejetée (avec un seuil de rejet de 0,05 %).

Au contraire, si le rapport est inférieur à la valeur de la table de Fisher notre hypothèse n'est pas rejetée et la série de données est validée. Dans la plupart des cas,  $N$  est beaucoup plus grand que  $n$ , on peut alors assimiler  $N$  à l'infini :  $N = \infty$ .

Le rapport  $s^2/LAQ^2$  doit alors être comparé avec les valeurs de la table de Fisher avec les paramètres  $(n-1, \infty)$ , mais il est également possible de préférer comparer les écart-types  $s$  et LAQ aux variances  $s^2$  et  $LAQ^2$ .

Dans ce cas, il suffit de comparer le rapport  $s/LAQ$  avec la racine carrée des valeurs maximales données dans une table de Fisher. C'est le choix qui est fait dans le tableau proposé par la norme ISO 19157, avec un niveau de signification de 95 %.

L'écart type sera considéré trop élevé si  $s/LAQ > \sqrt{F}$

Taille de la population (N)		Taille de l'échantillon (n)	$\sqrt{F}$
De	à		
26	50	5	1,54
51	90	7	1,45
91	150	10	1,37
151	280	15	1,30
281	400	20	1,26
401	500	25	1,23
501	1 200	35	1,20
1 201	3 200	50	1,16
3 201	10 000	75	1,13
10 001	35 000	100	1,12
35 001	150 000	150	1,09
150 001	500 000	200	1,08
> 500 000		200	1,08

Tableau 2 - table de Fisher pour un niveau de signification de 95 %

### ■ Fourniture d'un tableau de calcul

Nous fournissons un tableau de calcul « *Calcul loi de Fisher.xls* » capable de personnaliser cette méthode. Basé sur la loi de Fisher, il utilise comme paramètres :

- la taille de la population équivalente à  $N = \infty$  ;
- la taille de l'échantillon ;
- un niveau de signification de 95 %.

Il suffit de rentrer la taille de l'échantillon pour obtenir la valeur du rapport  $s/LAQ$ .

### Exemple d'utilisation n°1

Dans un lot de données de 800 objets ponctuels, on mesure le positionnement de 35 objets sur le terrain pour en contrôler la précision de position (taille de la population  $N = 800$  ; taille de l'échantillon  $n = 35$ , et  $\sqrt{F}$  est égal à 1,20 d'après le tableau de valeurs).

La limite d'acceptation de la qualité de l'intervalle de confiance de l'écart-type estimé a été fixée à 10 cm.

Or l'écart-type mesuré sur les 35 objets est de 17 cm.

Nous pouvons calculer :  $s/LAQ = 17/10 = 1,7$ .

Le rapport  $s/LAQ$  est donc supérieur à la valeur acceptable indiquée dans la troisième colonne : 1,20.

**Conclusion : le contrôle qualité n'est pas validé.**

### Exemple d'utilisation n°2

Dans un lot de données comprenant 450 plaques d'égout, on doit mesurer le positionnement de 25 plaques pour en contrôler la précision de position (taille de la population  $N = 450$ , taille de l'échantillon  $n = 25$ , et est égal à 1,23 d'après le tableau de valeurs).

La limite d'acceptation de la qualité de l'intervalle de confiance de l'écart-type estimé a été fixée à 19 cm.

Or l'écart-type mesuré sur les 25 objets est de 21 cm.

Nous pouvons calculer :  $s/LAQ = 21/19 = 1,10$

Le rapport  $s/LAQ$  est donc inférieur à la valeur acceptable indiquée dans la troisième colonne : 1,20.

**Conclusion : le contrôle qualité est validé.**



## Série de fiches « Qualifier les données géographiques »

Fiche n° 01	Connaitre la qualité d'une donnée géographique fiabilise son utilisation
Fiche n° 02	Généralités sur la qualité des données géographiques
Fiche n° 03	Éléments de contexte pour le contrôle qualité
<b>Fiche n° 04</b>	<b>Éléments statistiques</b>
Fiche n° 05	Méthodes d'échantillonnage
Fiche n° 06	Modes de représentation
Fiche n° 07	Critère de cohérence logique
Fiche n° 08	Critère d'exhaustivité
Fiche n° 09	Critère de précision thématique
Fiche n° 10	Critère de précision de position
Fiche n° 11	Critère de qualité temporelle



### Contributeurs

Fiche réalisée sous la coordination de Gilles Troispoux et Bernard Allouche (Cerema Territoires et ville)

#### Rédacteurs

Yves Bonin (Cerema Méditerranée), Arnaud Gallais (Cerema Ouest)

#### Contributeurs

Mathieu Rajerison, Silvio Rousic (Cerema Méditerranée)

#### Relecteurs

Benoît David (Mission information géographique MTES/CGDD), Stéphane Rolle (CRIGE PACA), Magali Carnino (DGAC), Stéphane Lévêque (Cerema Territoires et ville).

#### Maquettage

Cerema Territoires et ville  
Service édition

#### Impression

Jouve  
Mayenne



### Contact

accueil.dtectv@cerema.fr

Date de publication 2017  
ISSN : 2417-9701  
2017/58

**Boutique en ligne : [catalogue.territoires-ville.cerema.fr](http://catalogue.territoires-ville.cerema.fr)**

#### La collection « Connaissances » du Cerema

Cette collection présente l'état des connaissances à un moment donné et délivre de l'information sur un sujet, sans pour autant prétendre à l'exhaustivité. Elle offre une mise à jour des savoirs et pratiques professionnelles incluant de nouvelles approches techniques ou méthodologiques. Elle s'adresse à des professionnels souhaitant maintenir et approfondir leurs connaissances sur des domaines techniques en évolution constante. Les éléments présentés peuvent être considérés comme des préconisations, sans avoir le statut de références validées.

© 2017 - Cerema  
La reproduction totale ou  
partielle du document doit  
être soumise à l'accord  
préalable du Cerema.

Aménagement et développement des territoires - Ville et stratégies urbaines - Transition énergétique et climat - Environnement et ressources naturelles - Prévention des risques - Bien-être et réduction des nuisances - Mobilité et transport - Infrastructures de transport - Habitat et bâtiment