Crige
la géomatique à la carte

Stéphane ROLLE
Géomaticien

Cerema
CLIMAT & TERRITOIRES DE DEMAIN

Mathieu RAJERISON
Géomaticien

5 modules en 5 dates

02/05 : généralités + éléments de contexte pour le contrôle qualité
04/05 : statistiques basiques, échantillonnage et critère d'exhaustivité
06/05 : critère cohérence logique et précision de position
10/05 : critère cohérence temporelle et précision thématique
13/05 : statistiques avancées et méthode de représentation

- pensez à vous renommer pour plus de clarté

- questions dans le chat de la visio

- supports fournis en fin de formation (fin mai)

Module 5/5

Statistiques avancées

CRIGE (Stéphane Rolle) & Cerema (Mathieu Rajerison)

13 mai 2022

Sommaire

INCERTITUDE

Définition de l'incertitude

Incertitude sur la moyenne d'un échantillon

Incertitude sur un pourcentage d'entités conformes / non conformes

REJET D'UN JEU DE DONNÉES

Évaluation sur la base d'éléments conformes/non conformes

Évaluation de grandeurs mesurables

Incertitude sur une moyenne

Incertitude sur une moyenne

Incertitude

$$\Delta L = 0{,}5m$$

$$L = 12m \pm 0{,}5m$$

Niveau de confiance

95% de chance que $12m - 0{,}5m \leq mesure \leq 12m + 0{,}5m$

« C'est sans aucun doute 12m à 50 cm près »

Incertitude sur une moyenne d'un échantillon

$$x = \overline{x} \pm \Delta\overline{x}$$

$$\Delta\overline{x} = t\frac{s}{\sqrt{n}}$$

s : écart-type

n : taille de l'échantillon

t : coefficient de Student

n	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
t (95%)	12,7	4,3	3,18	2,78	2,57	2,45	2,37	2,31	2,26
t (99%)	63,7	9,93	5,84	4,6	4,03	3,71	3,5	3,36	3,25
n	**12**	**14**	**16**	**18**	**20**	**30**	**50**	**100**	[illegible]
t (95%)	2,2	2,16	2,13	2,11	2,09	2,04	2,01	1,98	1,96
t (99%)	3,11	3,01	2,95	2,9	2,86	2,76	2,68	2,63	2,57

Incertitude sur une moyenne d'un échantillon

$$x = \bar{x} \boxed{\pm \Delta \bar{x}} \qquad \boxed{\Delta \bar{x} = t \frac{s}{\sqrt{n}}}$$

n	2	3	4	5	6	7	8	9	10
t (95%)	12,7	4,3	3,18	2,78	2,57	2,45	2,37	2,31	2,26
t (99%)	63,7	9,93	5,84	4,6	4,03	3,71	3,5	3,36	3,25
n	**12**	**14**	**16**	**18**	**20**	**30**	**50**	**100**	[illegible]
t (95%)	2,2	2,16	$\boxed{2,13}$	2,11	2,09	2,04	2,01	1,98	1,96
t (99%)	3,11	3,01	2,95	2,9	2,86	2,76	2,68	2,63	2,57

Exemple : sur un échantillon de 16 objets, j'observe un écart moyen de 1m et un écart type de 0.2m, si je veux l'incertitude sur un intervalle de confiance de 95 %, alors je dois prendre un coefficient t de 2.13.

Mon incertitude vaut : 2.13 * (0.2/4) = 0.1065, donc je peux dire que sur l'ensemble du jeu de données, l'écart moyen de position est de 1m (plus ou moins 11 cm)

Incertitude sur une moyenne d'un échantillon

n	2	3	4	5	6	7	8	9	10
t (95%)	12,7	4,3	3,18	2,78	2,57	2,45	2,37	2,31	2,26
t (99%)	63,7	9,93	5,84	4,6	4,03	3,71	3,5	3,36	3,25
n	12	14	16	18	20	30	50	100	∞
t (95%)	2,2	2,16	2,13	2,11	2,09	2,04	2,01	1,98	1,96
t (99%)	3,11	3,01	2,95	2,9	2,86	2,76	2,68	2,63	2,57

Plus le nombre d'objets dans l'échantillon augmente, plus t diminue, et plus l'incertitude diminue.

n	2	3	4	5	6	7	8	9	10
t (95%)	12,7	4,3	3,18	2,78	2,57	2,45	2,37	2,31	2,26
t (99%)	63,7	9,93	5,84	4,6	4,03	3,71	3,5	3,36	3,25
n	12	14	16	18	20	30	50	100	∞
t (95%)	2,2	2,16	2,13	2,11	2,09	2,04	2,01	1,98	1,96
t (99%)	3,11	3,01	2,95	2,9	2,86	2,76	2,68	2,63	2,57

Si je veux avoir des chances de me situer dans l'intervalle de confiance, par exemple 99 % de chances, alors t augmente, et l'incertitude augmente aussi (l'intervalle est plus grand) : plus moins 15 cm au lieu de 11 cm

Incertitude sur un pourcentage

Incertitude sur un pourcentage

n	2	3	4	5	6	7	8	9	10
t (95%)	12,7	4,3	3,18	2,78	2,57	2,45	2,37	2,31	2,26
t (99%)	63,7	9,93	5,84	4,6	4,03	3,71	3,5	3,36	3,25
n	12	14	16	18	20	30	50	100	[illegible]
t (95%)	2,2	2,16	2,13	2,11	2,09	2,04	2,01	1,98	1,96
t (99%)	3,11	3,01	2,95	2,9	2,86	2,76	2,68	2,63	2,57

Exemple :

Taux d'exhaustivité (e) : 92%
Échantillon : 30
Niveau de confiance : 95%
Coefficient de Student : 2,04

$$\Delta \overline{x} = t \frac{s}{\sqrt{n}}$$

$$\Delta P = 2.04 \times \frac{\sqrt{0{,}92 \times (1 - 0{,}92)}}{30} = 0{,}018 = 1{,}8\ \%$$

« Le taux d'exhaustivité est de 92 %, plus moins 1.8 % »

Évaluation des objets conformes/non conformes

Évaluation des éléments conformes/non conformes

Niveau de confiance = 95 %			p_0 = LAQ (limite d'acceptation de la qualité)					
Taille de la population (N)		Taille de l'échantillon	0,5 %	1,0 %	2,0 %	3,0 %	4,0 %	5,0 %
De	à	(n)	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

Evaluation des éléments conformes/non conformes

taille de l'échantillon: **125** Limite de rejet: **3 bâtiments manquants**

Niveau de confiance = 95 %			p_0 = LAQ (limite d'acceptation de la qualité)					
Taille de la population (N)		Taille de l'échantillon	0,5 %	1,0 %	2,0 %	3,0 %	4,0 %	5,0 %
De	à	(n)	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

1- mon nombre de zones d'occupation du sol à contrôler est de **2440**, je dois prendre un **échantillon de 125 objets**.

2- je souhaite la plus petite tolérance d'une mauvaise classification (limite d'acceptation de la qualité faible **LAQ = 0,5 %**)

3- 4 objets sont mal classés (par ex. urbain au lieu d'agricole, industriel au lieu d'urbain), ma donnée est considérée non conforme. Si seulement 2 objets sont mal classés, alors ma donnée est conforme.

→ Si je choisis une LAQ plus élevée (plus grande tolérance), je pourrai tolérer un nombre d'objets mal classés plus grand, par ex. à 2 % de la LAQ, le nombre d'objets pouvant être non conformes monte à 6 au lieu de 3 initialement.

Évaluation d’une grandeur mesurable

Évaluation d’une grandeur mesurable

Taille de la population

↓

Taille de l’échantillon

↓

Contrôle de l’échantillon

↓

Calcul de s

↓

Calcul de s/LAQ

↓

Comparaison avec $\sqrt{F}$

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	1,3
281	400	20	1,26
401	500	25	1,23
501	1 200	35	1,2
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 de Fisher pour un niveau de signification de 95%

Exemple

Taille de la population (N)		Taille de l'échantillon (n)	√F
De	à		
26	50	5	1,54
51	90	7	1,45
91	150	10	1,37
151	280	15	1,3
281	400	20	1,26
401	500	25	1,23
501	1 200	35	1,2
1 201	3 200	50	1,16
3 201	10 000	75	1,13
10 001	35 000	100	[illegible]
35 001	150 000	150	[illegible]
150 001	500 000	200	1,08
> 500 000		200	1,08

REJECT

N = 800 objets

n = 35 objets

LAQ = 10 cm

s = 17 cm

s/LAQ = 1,7

s/LAQ > 1,2

Soit un jeu de données de **800 bouches à incendie**, j'extrais un échantillon de **35 objets** dont je contrôle la position dans l'espace.

Ma t**olérance est de 10cm**. Je calcule l'écart type (s = 17).

Je divise l'écart type par ma limite d'acceptation de la qualité (s/LAQ –> 17/10=1,7)

Je compare le résultat au coefficient de Fisher (cf tableau=1,2)

Le résultat est supérieur à ce chiffre, jeu de données est considéré comme non **non conforme**.

Autres exemples : hauteur de bâtiments, distance aux arrêts de transport