

Théorie de l'estimation
Théorie de l'estimation

L'objectif de ce cours est de répondre à la problématique suivante:

Comment à partir des informations sur l'échantillon (moyenne, écart-type, La proportion peut-on prévoir celles d'une population?

Problématique

X variable aléatoire d'étude, une population mère caractérisée par deux paramètres exactes (moyenne μ ou écart-type σ) ou encore une fréquence P d'un caractère dans cette population, un échantillon représentatif issu de cette population. La distribution de X dans la population est normale ou quelconque.

L'objectif de la théorie de l'estimation est de répondre à la problématique suivante : comment à partir des informations (moyenne m ou proportion \tilde{P}) calculées sur un échantillon estimer celles d'une population ?

Notons que

Nous distinguerons deux cas :

- *On estime la moyenne μ d'une distribution définie sur la population*
- *On estime la proportion P .*

Vocabulaire

- μ la moyenne théorique dans la population, c'est la valeur attendue
- \bar{X} est la variable aléatoire qui est l'estimateur de μ , c'est la moyenne empirique.
- m est la moyenne de l'échantillon observée.
- La loi de \bar{X} est la loi de probabilité de la moyenne empirique \bar{X} , sa distribution théorique, la distribution d'échantillonnage des moyennes.

1. Estimation et Estimations ponctuelles

l'estimation ponctuelle consiste à donner une valeur unique du paramètre, soit Φ un estimateur de θ .

On définit

$$\text{BIAIS} = E(\Phi) - \theta$$

1. Estimation sans biais

Si Biais = 0 (c'est-à-dire $E(\Phi) = \theta$)

Alors, l'estimateur est dit sans biais ou non biaisé

Exemple: $E(\bar{X}) = \mu$ (Résultat de théorie d'échantillonnage)

*Alors \bar{X} est donc un estimateur sans biais de μ
m la moyenne de l'échantillon observée est une estimation
ponctuelle de μ*

Exemple

- Pour estimer la moyenne **inconnue** d'une population μ , on prélève un échantillon et on calcule la moyenne de cet échantillon.
- Cette moyenne d'échantillon est une estimation ponctuelle de la moyenne μ .

Attention!!

➤ \tilde{S} l'écart-type de l'échantillon n'est pas un estimateur sans biais de σ

Exemple

➤ Pour estimer l'écart-type d'une population σ , on prélève un échantillon et on calcule l'écart-type corrigé $S = \sqrt{\frac{n}{n-1}} \tilde{S}$ de cet échantillon.

Cet écart-type corrigé de l'échantillon est une estimation ponctuelle de l'écart-type σ .

Remarque

➤ L'écart-type de l'échantillon $\tilde{S} = \sqrt{\frac{\sum(x_i - m)^2}{n}}$

➤ l'écart-type corrigé $S = \sqrt{\frac{\sum(x_i - m)^2}{n-1}}$

Exemple

- Pour estimer la fréquence P **inconnue** d'un caractère dans une population, on prélève un échantillon et on calcule la fréquence d'apparition de ce caractère dans l'échantillon.
- Cette fréquence d'apparition est une estimation ponctuelle de la fréquence P .

Estimation par intervalle de confiance de la moyenne :

L'estimation par intervalle de confiance I consiste à construire un intervalle à l'intérieur duquel le paramètre se trouve avec une probabilité donnée.

$$\mu \in I = \left[m - \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}}; m + \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} \right]$$

□ $\mu = m \pm \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} = m \pm \text{erreur}$

□ m le centre de l'intervalle de confiance.

□ l'erreur = demi- longueur de l'intervalle = $\bar{Z}_\alpha \frac{\sigma}{\sqrt{n}}$.

□ $P_r(\mu \in I) = 1 - \alpha$

Remarque:

- \bar{Z}_α est calculé de la table 2 (table de l'écart réduit)
- $P_r(\mu \in I) = 1 - \alpha$ (Cette formule est une conséquence du théorème de la limite centrale)

Population infinie avec ou sans remise

Population finie avec remise

Cas1 : X suit la loi normale $N(\mu; \sigma)$ et σ connu

$$I = \left[m - \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}}; m + \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} \right]$$

cas2

X suit la loi normale $N(\mu; \sigma)$ et σ inconnu, on l'estime par l'estimation

ponctuelle $S = \sqrt{\frac{n}{n-1}} \tilde{S}$

Nous distinguerons deux situations :

Pour $n \geq 30$, on \bar{Z}_α cherche dans la table Normale :

$$I = \left[m - \bar{Z}_\alpha \frac{S}{\sqrt{n}}; m + \bar{Z}_\alpha \frac{S}{\sqrt{n}} \right]$$

Pour $n < 30$, on \bar{t}_α cherche dans la table de Student :

$$I = \left[m - \bar{t}_\alpha \frac{S}{\sqrt{n}}; m + \bar{t}_\alpha \frac{S}{\sqrt{n}} \right]$$

Cas3

X suit la loi quelconque : il faut que $n \geq 30$

Pour σ connu, on cherche \bar{Z}_α dans la table Normale :

$$I = \left[m - \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} ; m + \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} \right]$$

Pour σ inconnu, on cherche \bar{Z}_α dans la table Normale:

$$I = \left[m - \bar{Z}_\alpha \frac{S}{\sqrt{n}} ; m + \bar{Z}_\alpha \frac{S}{\sqrt{n}} \right]$$

Attention!!

Population finie tirage sans remise

On ajoute le facteur d'exhaustivité

$$\sqrt{\frac{N-n}{N-1}}$$

Population finie et tirage exhaustif (sans remise)

Cas1 : X suit la loi normale $N(\mu; \sigma)$ et σ connu

$$I = \left[m - \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}; m + \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

cas2

X suit la loi normale $N(\mu; \sigma)$ et σ inconnu, on l'estime par l'estimation

ponctuelle $S = \sqrt{\frac{n}{n-1}} \tilde{S}$

Nous distinguerons deux situations :

Pour $n \geq 30$, on \bar{Z}_α cherche dans la table Normale :

$$I = \left[m - \bar{Z}_\alpha \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}; m + \bar{Z}_\alpha \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

Pour $n < 30$, on \bar{t}_α cherche dans la table de Student :

$$I = \left[m - \bar{t}_\alpha \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}; m + \bar{t}_\alpha \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

Cas3

X suit la loi quelconque : il faut que $n \geq 30$

Pour σ connu, on cherche \bar{Z}_α dans la table Normale :

$$I = \left[m - \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}; m + \bar{Z}_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

Pour σ inconnu, on cherche \bar{Z}_α dans la table Normale:

$$I = \left[m - \bar{Z}_\alpha \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}; m + \bar{Z}_\alpha \frac{S}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

Attention!!

Parmi les cinq intervalles de confiances, la lecture est faite dans la table de Student (table 3) dans un seul cas , il s'agit du cas $n < 30$ et σ inconnu avec X suit la loi normale $N(\mu; \sigma)$

Exemple

Sur 20 patients a été mesuré le taux de fer sérique (distribution Normale) exprimé en $\mu\text{g}/100\text{ml}$

83.0 ; 98.0 ; 183.3 ; 119.6 ; 78.5 ; 162.6 ; 155.7 ; 147.3 ; 100.1 ; 139.2 ; 172.1 ; 102.0 ; 162.8 ; 113.8 ; 157.4 ; 128.5 ; 136.2 ; 129.3 ; 131.6 ; 157.3.

- 1) Les estimations ponctuelles non biaisées de la moyenne et la variance du taux de fer sérique à partir de cet échantillon*
- 2) L'estimation de la moyenne au risque de $\alpha = 0.05$ par intervalle de confiance*

Solution

1) Les estimations ponctuelles biaisées de la moyenne et la variance du taux de fer sérique à partir de cet échantillon respectivement sont :

$$\text{moyenne} = 132.92 \mu\text{g}/100\text{ml}$$

$$\text{variance corrigée} = S^2 = \frac{n}{n-1} \tilde{S}^2 = 900 (\mu\text{g}/100\text{ml})^2$$

Solution

Données échantillon

$$n = 20 < 30$$

$m = 132,92$ moyenne de l'échantillon

$$S = \sqrt{\frac{\sum_{i=1}^{20} (x_i - m)^2}{n-1}} = 30$$

Données population

Population infinie

σ écart-type de la population

σ inconnu, on l'estime par S

Solution

Pour $n < 30$, la distribution suit la loi normale (condition satisfaite par énoncé), alors :

on \bar{t}_α cherche dans la table de Student :

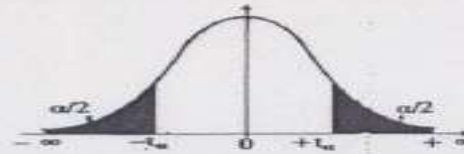
$$I = \left[m - \bar{t}_\alpha \frac{S}{\sqrt{n}} ; m + \bar{t}_\alpha \frac{S}{\sqrt{n}} \right]$$

En utilisant la table 3 de Student, on trouve :

$$\underbrace{\alpha = 0.05 \quad ddl = n - 1 = 19}_{\bar{t}_\alpha = 2.093}$$

On obtient $\mu \in [118.88 ; 146.96]$.

Tabla 6
Distribución t de Student



α $\varepsilon-t$	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.929
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.133	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
35	0.682	0.852	1.052	1.306	1.690	2.030	2.438	2.724	3.592
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.705	3.551
45	0.680	0.850	1.049	1.301	1.679	2.014	2.412	2.690	3.521
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.497
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.461
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.417
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.391
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Estimation par intervalle de confiance de la proportion

Un intervalle de confiance pour la proportion ne se calcule que pour les grands échantillons $n \geq 30$.

L'estimation par intervalle de confiance I consiste à construire un intervalle à l'intérieur du quel le paramètre se trouve avec une probabilité donnée.

$$P \in I = \left[\tilde{P} - \bar{Z}_\alpha \sqrt{\frac{\tilde{P}(1 - \tilde{P})}{n}}; \tilde{P} + \bar{Z}_\alpha \sqrt{\frac{\tilde{P}(1 - \tilde{P})}{n}} \right]$$

Tels que

$$\checkmark n\tilde{P} > 5 \quad n(1 - \tilde{P}) > 5.$$

$$\checkmark P = \tilde{P} \pm \bar{Z}_\alpha \sqrt{\frac{\tilde{P}(1-\tilde{P})}{n}} = \tilde{P} \pm \text{erreur}$$

$\checkmark \tilde{P}$ le centre de l'intervalle de confiance.

\checkmark l'erreur = demi-longueur de l'intervalle

$$= \bar{Z}_\alpha \sqrt{\frac{\tilde{P}(1-\tilde{P})}{n}}.$$

$$\checkmark P_r(P \in I) = 1 - \alpha$$