Biostatistiques – BST1

Chapitre 12
Estimation par intervalle d'un paramètre

José LABARERE, PU-PH

Arnaud Seigneurin, MCU-PH, Bastien Boussat, MCU-PH, Alexandre Bellier, AHU, Patrice François, PU-PH

Plan

- Objectifs
- Introduction
- Moyenne
- Proportion
- Nombre de sujets nécessaire

Objectifs

Déterminer les bornes de l'intervalle de confiance à $(1 - \alpha)\%$ pour :

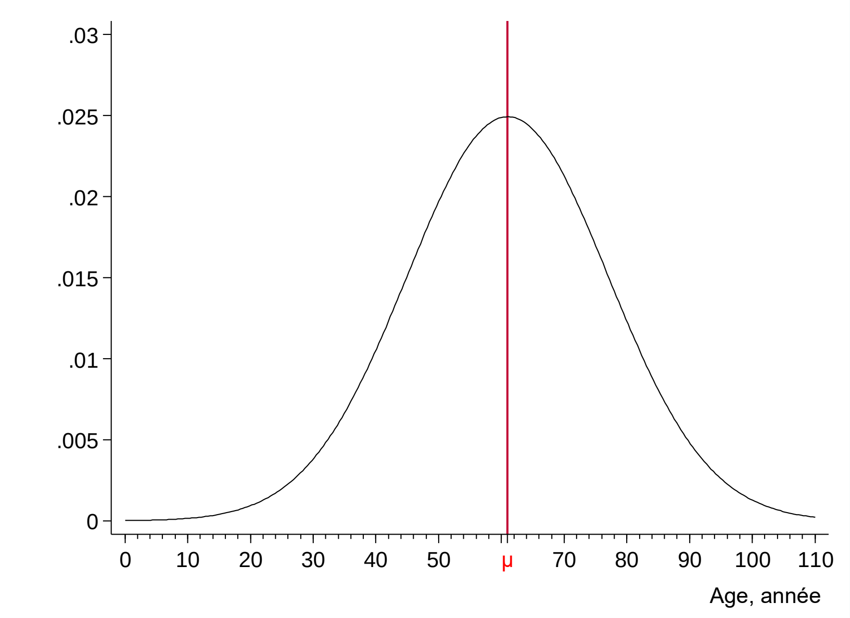
- une moyenne
- une proportion

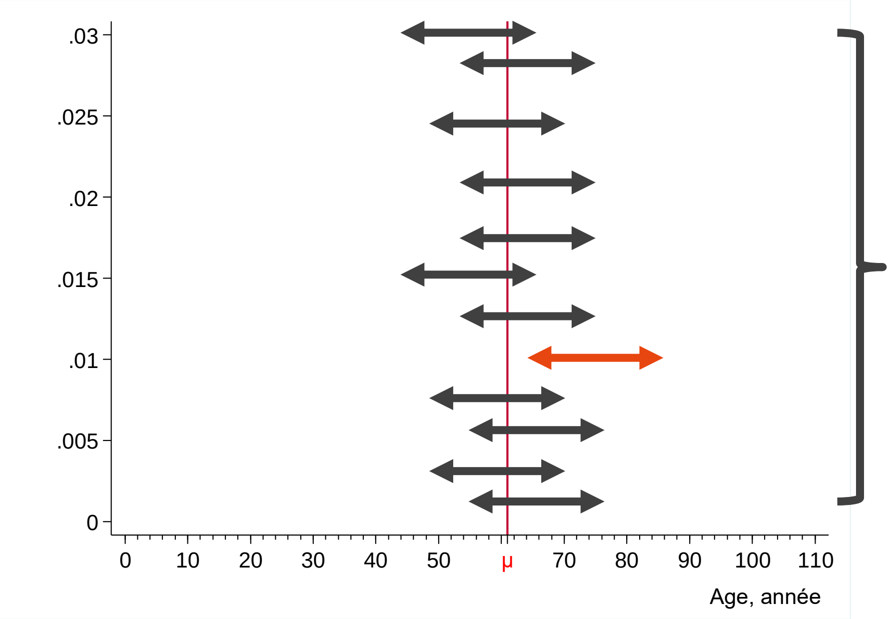
Déterminer le nombre de sujets nécessaire pour estimer un paramètre avec une précision souhaitée

Plan

- Objectifs
- Introduction
- Moyenne
- Proportion
- Nombre de sujets nécessaire

Estimation par intervalle





Intervalles de confiance à (1 – α) % de la moyenne théorique μ, calculés à partir d'échantillons

Plan

- Objectifs
- Introduction
- Moyenne
- Proportion
- Nombre de sujets nécessaire

Estimation par intervalle de la moyenne

Théorème central limite

1. La moyenne m d'une variable quantitative X calculée sur un échantillon d'effectif n est elle-même une variable aléatoire
2. Cette variable m suit
   – une loi normale $N(\mu, \sigma/\sqrt{n})$ quelle soit la loi de distribution de la variable X si l'effectif de l'échantillon n ≥30
   – une loi de Student t $(\mu, \sigma/\sqrt{n})$ si la variable X suit elle-même une loi normale (quel que soit l'effectif n de l'échantillon)

Estimation par intervalle de la moyenne

$$m \to N(\mu, \sigma/\sqrt{n})$$

$$\frac{m-\mu}{\sigma/\sqrt{n}} \to N(0,1)$$

$$P\left(-\varepsilon_\alpha \leq \frac{m-\mu}{\sigma/\sqrt{n}} \leq \varepsilon_\alpha\right) = (1-\alpha)$$

$$P\left(m - \varepsilon_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu \leq m + \varepsilon_\alpha \frac{\sigma}{\sqrt{n}}\right) = (1-\alpha)$$

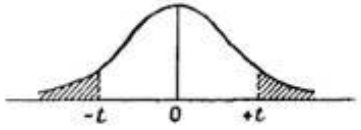
Intervalle de confiance à (1 – α) % de la moyenne théorique μ

$$P\left(m - \varepsilon_\alpha \frac{s}{\sqrt{n}} \leq \mu \leq m + \varepsilon_\alpha \frac{s}{\sqrt{n}}\right) = (1 - \alpha)$$

	n ≥30	**X → N(μ, σ), n <30**
Loi de la moyenne estimée	$m \rightarrow N(\mu, \sigma/\sqrt{n})$	$\frac{m}{s/\sqrt{n}} \rightarrow t_{(n-1)}$
Borne inférieure	$m - \varepsilon_\alpha \frac{s}{\sqrt{n}}$	$m - t_\alpha \frac{s}{\sqrt{n}}$
Borne supérieure	$m + \varepsilon_\alpha \frac{s}{\sqrt{n}}$	$m + t_\alpha \frac{s}{\sqrt{n}}$

Table de t (*).

La table donne la probabilité α pour que t égale ou dépasse, en valeur absolue, une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).



d.d.l. \ α	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	1,000	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,816	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,765	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,741	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,727	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,718	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,711	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,706	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,703	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,700	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,697	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,695	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,694	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,692	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,691	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,690	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,689	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,688	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,688	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,687	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,686	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,686	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,685	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,685	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,684	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,684	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,684	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,683	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,683	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,683	1,055	1,310	1,697	2,042	2,457	2,750	3,646
∞	0,126	0,674	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Valeur de t_α pour :

(n-1) ddl = 5 : 2,57

(n-1) ddl = 10 : 2,22

(n-1) ddl > 30 : 1,96

Plan

- Objectifs
- Introduction
- Moyenne
- Proportion
- Nombre de sujets nécessaire

Estimation par intervalle d'une proportion

$$P\left(p-\varepsilon_{\alpha}\sqrt{\frac{p(1-p)}{n}}\leq\pi\leq p+\varepsilon_{\alpha}\sqrt{\frac{p(1-p)}{n}}\right)=(1-\alpha)$$

$$p=\frac{\sum_{1}^{n}x_i}{n}$$

$$\boldsymbol{S_P}=\sqrt{\frac{p(1-p)}{n}}$$

Conditions d'application

$np_i \geq 5$, $np_s \geq 5$, $n\,(1 - p_i) \geq 5$, $n\,(1 - p_s) \geq 5$

Estimation par intervalle d'une proportion

	No. (%)	[95% CI]
Outcome, No. with data	1581 (,,,)	[,,,]
Died in ICU	405 (26)	$\boxed{[23 \text{ to } 28]}$
Discharged from ICU	256 (16)	[14 to 18]
Still in ICU as of 3/25/2020[a]	920 (58)	[56 to 61]

Abbreviations : CI = confidence interval; ICU = intensive care unit

[a] Patients were admitted between 2/20/2020 and 3/18/2020, with follow-up through 3/25/2020.

$$P\left(p - \varepsilon_\alpha \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + \varepsilon_\alpha \sqrt{\frac{p(1-p)}{n}}\right) = P\left(0{,}26 \; - 1{,}96 \sqrt{\frac{0{,}26(1-0{,}26)}{1581}} \leq \pi \leq 0{,}26 + 1{,}96 \sqrt{\frac{0{,}26(1-0{,}26)}{1581}}\right)$$

Adapté de Grasselli G, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region, Italy. JAMA 2020;323(16):1574-1581.

Plan

- Objectifs
- Introduction
- Moyenne
- Proportion
- Nombre de sujets nécessaire

Précision d'un intervalle de confiance à (1 – α) %

$$P\left(m - \varepsilon_\alpha \frac{s}{\sqrt{n}} \leq \mu \leq m + \varepsilon_\alpha \frac{s}{\sqrt{n}}\right) = (1 - \alpha)$$

Précision i = $\frac{\left(m+\varepsilon_\alpha \frac{s}{\sqrt{n}}\right)-\left(m-\varepsilon_\alpha \frac{s}{\sqrt{n}}\right)}{2} = \frac{\left(m+\varepsilon_\alpha \frac{s}{\sqrt{n}}\right)-\left(m-\varepsilon_\alpha \frac{s}{\sqrt{n}}\right)}{2} = \frac{2\varepsilon_\alpha \frac{s}{\sqrt{n}}}{2} = \varepsilon_\alpha \frac{s}{\sqrt{n}}$

Précision i d'un intervalle de confiance dépend

- du **risque d'erreur α** consenti (qui détermine la valeur de ε_α)
- de l'**effectif de l'échantillon n**

Nombre de sujets nécessaire pour obtenir un intervalle de confiance d’une précision donnée

- Moyenne

$$n = \sigma^2 \frac{\varepsilon_\alpha^2}{i^2}$$

- Proportion

$$n = p(1-p)\frac{\varepsilon_\alpha^2}{i^2}$$

n : nombre de sujets nécessaire

σ^2 : variance théorique de la variable quantitative dans la population

ε_α : valeur lue sur la table de l’écart réduit pour le risque α consenti

i : précision de l’estimation (demi-étendu de l’intervalle de confiance)

Messages clés

Paramètre	Conditions d'application	Intervalle de confiance (1 – α)%
Moyenne	$n \geq 30$	$\left(m - \varepsilon_\alpha \frac{s}{\sqrt{n}}\,;\ m + \varepsilon_\alpha \frac{s}{\sqrt{n}}\right)$
Moyenne	$X \rightarrow N(\mu, \sigma)$, $n < 30$	$\left(m - t_\alpha \frac{s}{\sqrt{n}}\,;\ m + t_\alpha \frac{s}{\sqrt{n}}\right)$
Proportion	$np_i \geq 5$, $np_s \geq 5$, $n(1 - p_i) \geq 5$, $n(1 - p_s) \geq 5$	$\left(p - \varepsilon_\alpha \sqrt{\frac{p(1-p)}{n}}\,;\ p + \varepsilon_\alpha \sqrt{\frac{p(1-p)}{n}}\right)$

Mentions légales