Biostatistiques – BST1

Chapitre 17
Test paramétrique de comparaison de deux moyennes estimées sur deux échantillons indépendants

José LABARERE, PU-PH

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

Plan

- Objectifs
- Introduction
- Test Z de l’écart réduit
- Test t de Student

Objectifs

Comparaison de deux moyennes estimées sur deux échantillons indépendants :

- Identifier une situation requérant ce type de test (test Z écart-réduit ou t de Student)
- Connaitre et vérifier les conditions d'application de ce test
- Interpréter le résultat du test

Plan

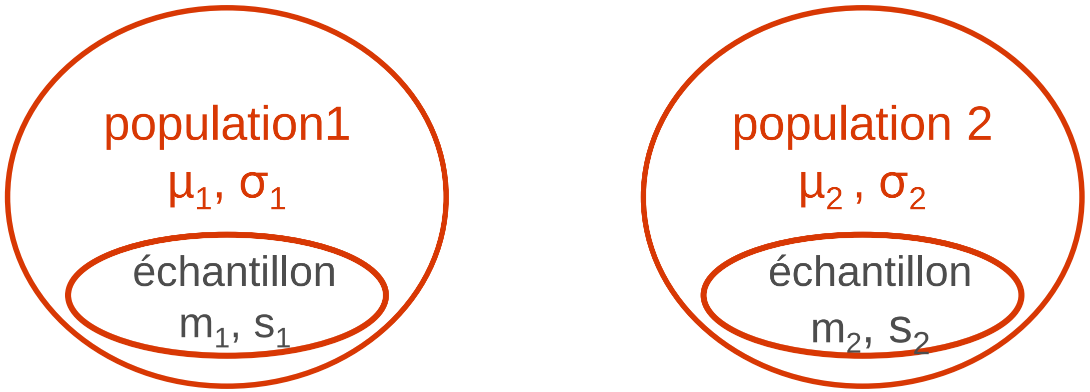
- Objectifs
- Introduction
- Test Z de l’écart réduit
- Test t de Student

Comparaison de 2 moyennes estimées sur 2 échantillons indépendants

Echantillons indépendants (par opposition aux échantillons appariés) :

- les individus de l’échantillon 1 sont différents des individus de l’échantillon 2 (et sans liens avec ceux-ci).
- Les deux échantillons peuvent être d’effectifs différents.

Comparaison de 2 moyennes estimées sur 2 échantillons indépendants



1. Formulation des hypothèses

H0 : $\mu_1 = \mu_2$ (Absence d'association entre la variable quantitative et la variable binaire)

H1 : $\mu_1 \neq \mu_2$

2. Risque α = 0,05 (5%)

3. Choix du test

Test Z de l'écart réduit (n1 $\geq$ 30 et n2 $\geq$ 30)

Test t de Student (normalité de distribution, variances comparables)

Plan

- Objectifs
- Introduction
- Test Z de l’écart réduit
- Test t de Student

Test Z de l'écart réduit

$m_1 \approx \mu_1$

(fluctuations d'échantillonnage)

$m_2 \approx \mu_2$

(fluctuations d'échantillonnage)

Si $n_1 \geq 30$: $m_1 \rightarrow \boldsymbol{N(\mu_1, \sigma_1/\sqrt{n_1})}$

Si $n_2 \geq 30$: $m_2 \rightarrow \boldsymbol{N(\mu_2, \sigma_2/\sqrt{n_2})}$

Rappel : La moyenne m d'une variable quantitative X calculée sur un échantillon d'effectif n est elle-même une variable aléatoire de loi normale $N(\mu, \sigma/\sqrt{n})$, quelle que soit la loi de distribution de la variable X si l'effectif de l'échantillon n ≥30 (Chapitre 11, Théorème central limite)

Test Z de l’écart réduit

$m_1 \approx \mu_1$

(fluctuations d’échantillonnage)

$m_2 \approx \mu_2$

(fluctuations d’échantillonnage)

$$(\boldsymbol{m_1 - m_2}) \rightarrow \boldsymbol{N}\left(\boldsymbol{\mu_1 - \mu_2}, \sqrt{\frac{\boldsymbol{\sigma_1^2}}{\boldsymbol{n_1}} + \frac{\boldsymbol{\sigma_2^2}}{\boldsymbol{n_2}}}\right)$$

Test Z de l'écart réduit

$m_1 \approx \mu_1$

(fluctuations d'échantillonnage)

$m_2 \approx \mu_2$

(fluctuations d'échantillonnage)

$$(\boldsymbol{m_1} - \boldsymbol{m_2}) \to \boldsymbol{N}\left(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}, \sqrt{\frac{\boldsymbol{\sigma_1^2}}{\boldsymbol{n_1}} + \frac{\boldsymbol{\sigma_2^2}}{\boldsymbol{n_2}}}\right)$$

$$\text{var}\,(m_1 - m_2) = \text{var}\,(m_1) + \text{var}\,(m_2) - \underbrace{2\,\text{cov}\,(m_1, m_2)}_{0}$$

Rappel : var(a-b) = var(a) + var(b) – 2 cov(a,b) cf analogie avec (a-b)² = a² + b² - 2.a.b

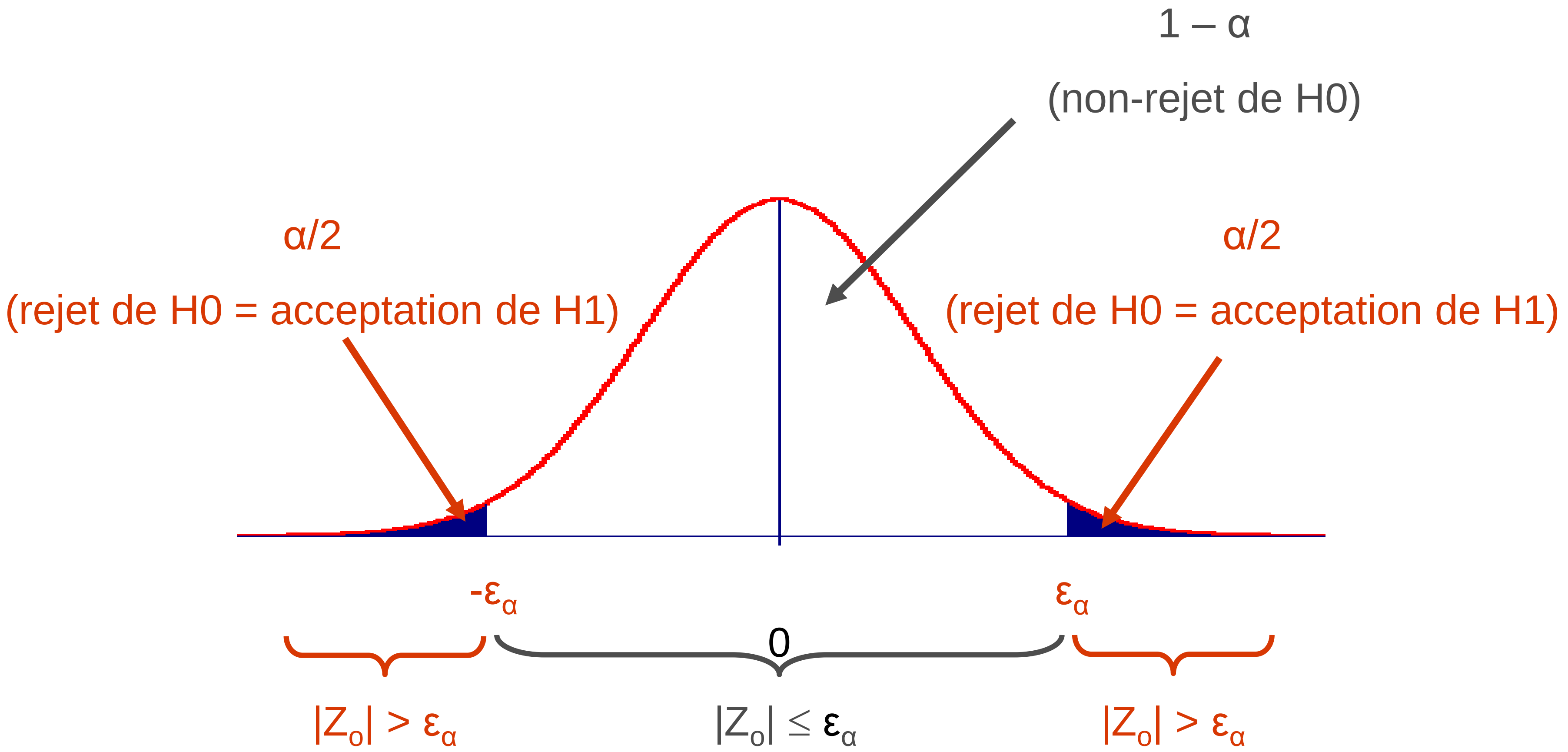
Test Z de l'écart réduit

Sous H0 : $\mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$

$$(m_1 - m_2) \rightarrow N\left(0, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$Z = \frac{(m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0, 1) \qquad Z = \frac{(m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \rightarrow N(0, 1)$$

Densité de probabilité de loi normale centrée réduite (0,1)



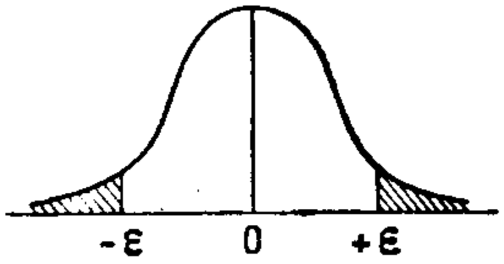
Abscisse : valeurs possibles de Z <u>sous H0</u>

$$\mathbf{Z} = \frac{(\mathbf{m_1} - \mathbf{m_2})}{\sqrt{\frac{\mathbf{s_1^2}}{\mathbf{n_1}} + \frac{\mathbf{s_2^2}}{\mathbf{n_2}}}}$$

Détermination de la valeur de ε_α correspondant à un risque $\alpha = 0,05$

Table de l'écart réduit

La table donne la probabilité α pour que l'écart-réduit dépasse en valeur absolue une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle [-ε,ε]. La probabilité α s'obtient par addition des nombres inscrits en marge



α	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,1	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,2	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,3	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,4	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,5	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,6	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,7	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,8	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,9	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

Détermination du degré de signification associé à Z_o (*P*-value)

Table 1. Demographics and Clinical Characteristics Among Patients In the Development Cohort Who Did or Did Not Develop Critical Illness[a]

Characteristic	Total, mean (SD) [range]	Critical illness: No	Critical illness: Yes
No.	1590	1459	131
Age, mean (SD), y	48.9 (15.7) [1-95]	47.8 (15.2)	61.6 (14.8)
Incubation period, mean (SD), d	5.0 (4.1) [0-24]	4.9 (4.1)	5.7 (4.2)
Admission measures, mean (SD)			
Temperature, °C	37.3 (0.9) [35.5-40.3]	37.4 (0.9)	37.1 (0.9)
Respiratory rate, breaths/min	21.2 (12.0) [12-65]	21.1 (12.4)	23.1 (5.9)
Heart rate, beats/min	88.7 (14.6) [17-205]	88.6 (14.4)	89.7 (16.0)
Blood pressure, mm Hg			
Systolic	126.1 (16.4) [74-187]	125.5 (15.6)	131.4 (22.5)
Diastolic	79.5 (25.6) [40-160]	79 (11.3)	84.7 (76.1)

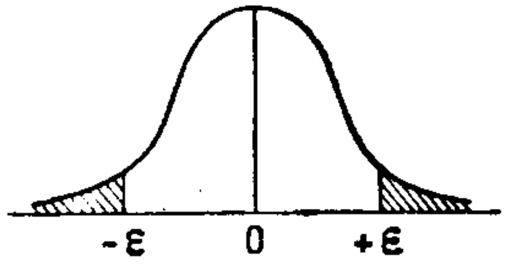
$$Z_o = \frac{(m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(47{,}8 - 61{,}6)}{\sqrt{\frac{15{,}2^2}{1459} + \frac{14{,}8^2}{131}}} = -10{,}2$$

Adapté de Liang W, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Intern Med 2020.

Détermination du degré de signification associé à Z_o (*P*-value)

Table de l'écart réduit

La table donne la probabilité α pour que l'écart-réduit dépasse en valeur absolue une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle [-ε,ε]. La probabilité α s'obtient par addition des nombres inscrits en marge



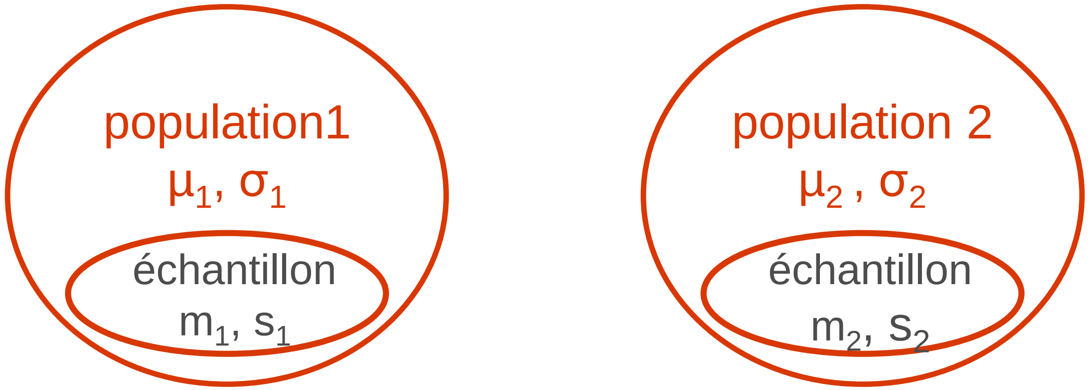
α	1,00E-03	1,00E-04	1,00E-05	1,00E-06	1,00E-07	1,00E-08	1,00E-09	1,00E-10	1,00E-11
ε	3,29	3,89	4,42	4,89	5,33	5,73	6,11	6,47	6,81

$|Z_o| = 10,2 \rightarrow$ *P*-value <0,001

Plan

- Objectifs
- Introduction
- Test Z de l’écart réduit
- Test t de Student

Test t de Student

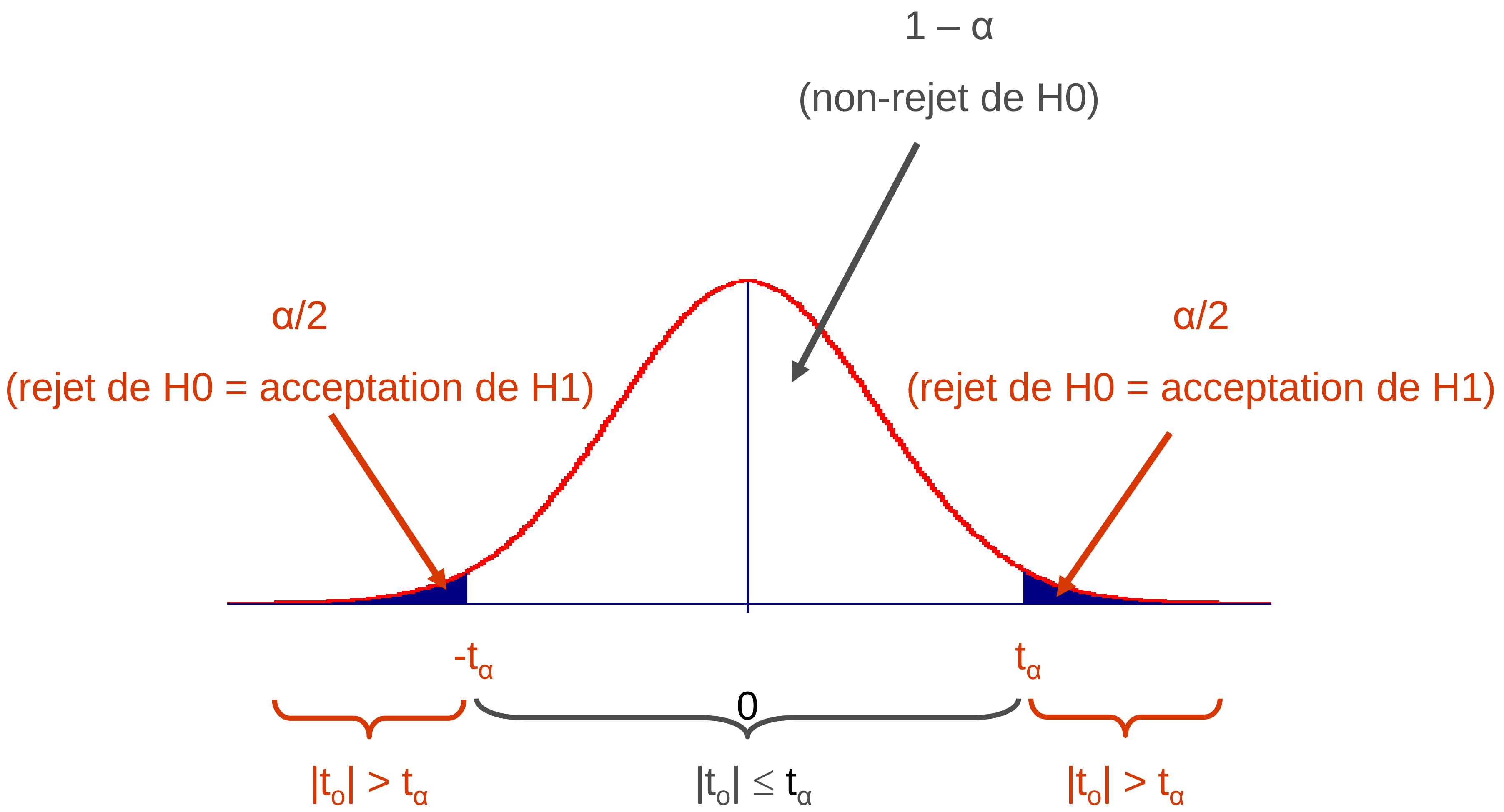


Sous H0 : $\mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$

$$t = \frac{(m_1 - m_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \rightarrow t_{(n_1+n_2-2)ddl}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

Densité de probabilité de la loi t de Student à $(n_1 + n_2 - 2)$ ddl



Abscisse : valeurs possibles de t sous H0

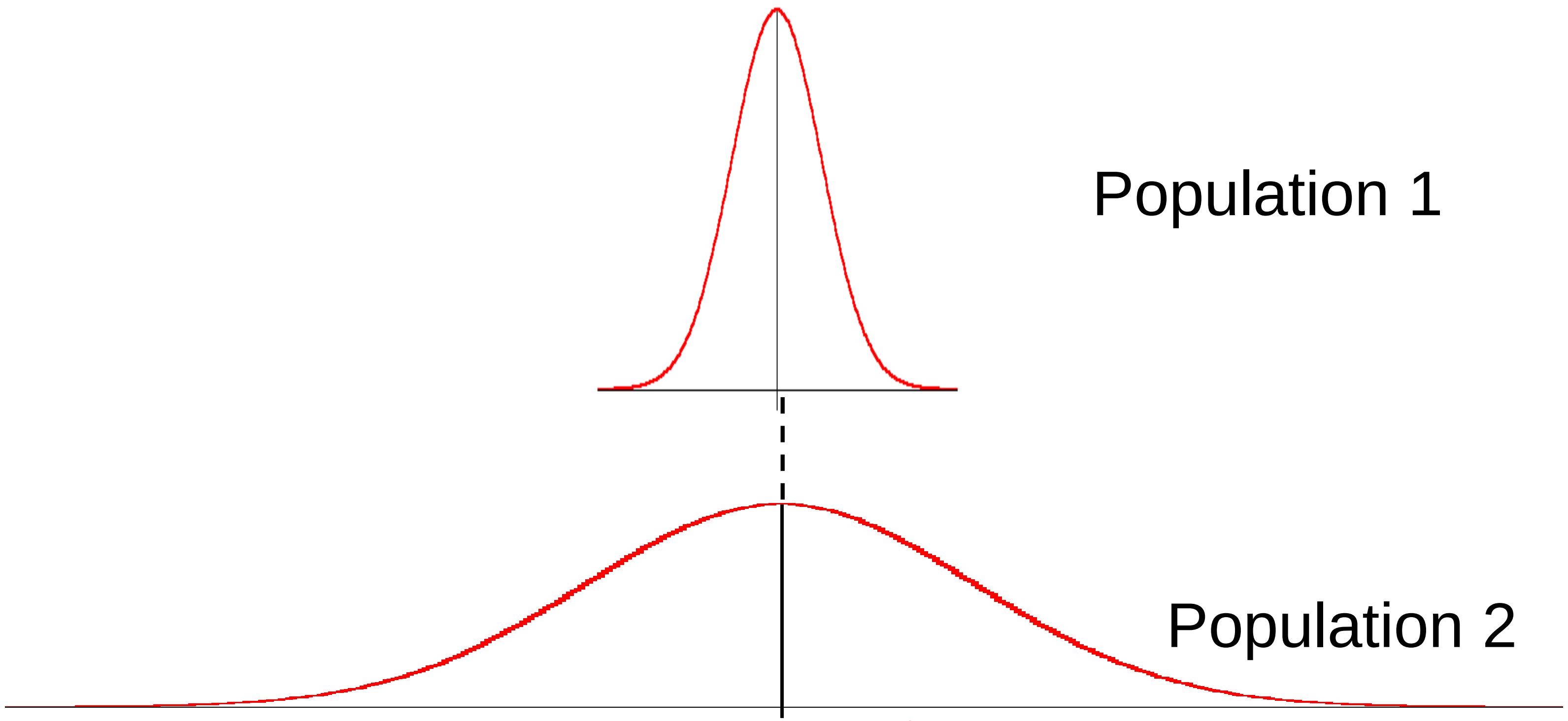
$$t = \frac{(m_1 - m_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Test t de Student : conditions d'application

- La distribution de la variable continue est normale dans les 2 populations

- Les variances σ_1^2 et σ_2^2 sont comparables (rapport 1 à 3)

Comparabilité des variances



$$\left.\begin{array}{c}\mu_1 = \mu_2 \\ \sigma_1^2 < \sigma_2^2\end{array}\right\}$$

Distributions très différentes bien que les moyennes soient égales

Messages clés

m_1	m_2	effectif	test	conditions
estimée	estimée	$n_1, n_2 \geq 30$	Z	-
(échantillons indépendants)		n_1, n_2	$t_{(n_1 + n_2 - 2)\ \text{ddl}}$	Normalité σ^2 comparables

Mentions légales