# Mathématiques pour les Sciences de la Vie
# Probabilités – Statistique

Printemps 2022

S. Mousset

Université Claude Bernard Lyon I – France

# Estimation

On souhaite connaître la distribution d'une variable dans la population (taille, poids, couleur de cheveux,. . .

# Estimation

On souhaite connaître la distribution d'une variable dans la population (taille, poids, couleur de cheveux,. . .
Faut-il mesurer toute la population ?

# Échantillon

Échantillon :

- Sous-ensemble de la population d'intérêt.

# Échantillon

Échantillon :

- Sous-ensemble de la population d'intérêt.
- Taille de l'échantillon : nombre d'observations $n$.

# Échantillon

Échantillon :

- Sous-ensemble de la population d'intérêt.
- Taille de l'échantillon : nombre d'observations $n$.
- Stratégie d'échantillonnage : comment obtenir les meilleures informations sur la population.

# Échantillon

Échantillon :

- Représentatif de la population (non biaisé).

# Échantillon

Échantillon :

- Représentatif de la population (non biaisé).
- Aléatoire simple (population de taille $N$, échantillon de taille $n$, probabilité identique pour tous d'être échantillonné : $\frac{n}{N}$.

# Échantillon

Échantillon :

- Représentatif de la population (non biaisé).
- Aléatoire simple (population de taille $N$, échantillon de taille $n$, probabilité identique pour tous d'être échantillonné : $\frac{n}{N}$.
- Structuré : on possède une information sur une structuration en $k$ classe de fréquences $p_1$,...,$p_k$, et on constitue l'échantillon en conséquence.

# Échantillon

Échantillon :

- Représentatif de la population (non biaisé).
- Aléatoire simple (population de taille $N$, échantillon de taille $n$, probabilité identique pour tous d'être échantillonné : $\frac{n}{N}$.
- Structuré : on possède une information sur une structuration en $k$ classe de fréquences $p_1$,...,$p_k$, et on constitue l'échantillon en conséquence.

# Échantillon

Échantillon :

- Représentatif de la population (non biaisé).
- Aléatoire simple (population de taille $N$, échantillon de taille $n$, probabilité identique pour tous d'être échantillonné : $\frac{n}{N}$.
- Structuré : on possède une information sur une structuration en $k$ classe de fréquences $p_1$,. . .,$p_k$, et on constitue l'échantillon en conséquence.

Nous nous intéresserons uniquement à l'échantillonnage aléatoire simple.

# Notation

- Population : l'ensemble des individus.

# Notation

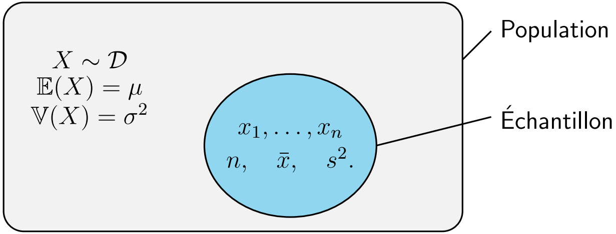
- Population : l'ensemble des individus.
- Échantillon : sous-ensemble de la population choisi au hasard.

$$\begin{aligned} X &\sim \mathcal{D} \\ \mathbb{E}(X) &= \mu \\ \mathbb{V}(X) &= \sigma^2 \end{aligned}$$

Population

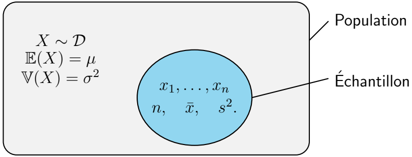
# Notation

- Population : l'ensemble des individus.
- Échantillon : sous-ensemble de la population choisi au hasard.



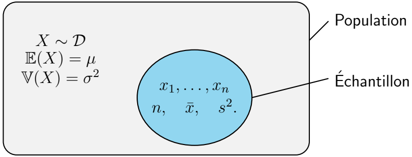
# Notation

- $\mu$ et $\sigma$ sont la moyenne et la variance de la distribution de $X$. On ne peut les obtenir qu'en échantillonnant tout le monde.



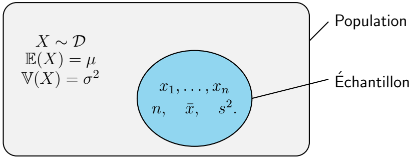
# Notation

- $\mu$ et $\sigma$ sont la moyenne et la variance de la distribution de $X$. On ne peut les obtenir qu'en échantillonnant tout le monde.
- $x_1$, ..., $x_n$ sont les observations de $X$ sur les $n$ individus de l'échantillon.



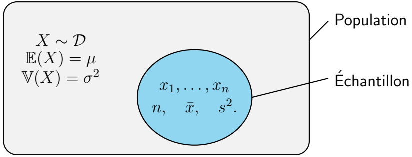
# Notation

- $\mu$ et $\sigma$ sont la moyenne et la variance de la distribution de $X$. On ne peut les obtenir qu'en échantillonnant tout le monde.
- $x_1$, ..., $x_n$ sont les observations de $X$ sur les $n$ individus de l'échantillon.
- $\bar{x}$ et $s^2$ sont les mesures de la moyenne et de la variance du caractère étudié dans l'échantillon.



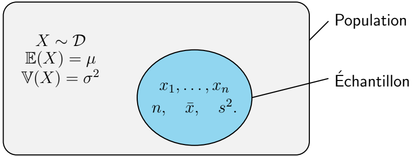
# Notation

- $\mu$ et $\sigma$ sont la moyenne et la variance de la distribution de $X$. On ne peut les obtenir qu'en échantillonnant tout le monde.
- $x_1$, ..., $x_n$ sont les observations de $X$ sur les $n$ individus de l'échantillon.
- $\bar{x}$ et $s^2$ sont les mesures de la moyenne et de la variance du caractère étudié dans l'échantillon.



# Notation

- $\mu$ et $\sigma$ sont la moyenne et la variance de la distribution de $X$. On ne peut les obtenir qu'en échantillonnant tout le monde.
- $x_1$, ..., $x_n$ sont les observations de $X$ sur les $n$ individus de l'échantillon.
- $\bar{x}$ et $s^2$ sont les mesures de la moyenne et de la variance du caractère étudié dans l'échantillon.




On veut estimer $\mu$ et $\sigma^2$ à partir de l'échantillon.
Les estimations sont notées $\hat{\mu}$ et $\hat{\sigma}$.

# Types de données

| Type de données | non regroupées | regroupées en $k$ classes |
|---|---|---|
| Données | $x_1 \dots x_n$ | $x_1 \dots x_k$ |
| Effectifs | $n_1 = \dots = n_n = 1$ | $n_1, \dots, n_k.\ n = \sum_{i=1}^{k} n_i$ |
| Moyenne $\bar{x}$ | $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ | $\bar{x} = \frac{1}{n} \sum_{i=1}^{k} n_i x_i$ |
| Variance $s^2$ | $s^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \left( \frac{1}{n} \sum_{i=1}^{n} x_i \right)^2$ | $s^2 = \frac{1}{n} \sum_{i=1}^{k} n_i x_i^2 - \left( \frac{1}{n} \sum_{i=1}^{k} n_i x_i \right)^2$ |

# Exemple

- Population : Étudiants inscrits en MathSV
- Variable : Taille du prénom
- On a accès aux prénoms des 437 étudiants
- On connaît donc $\mu = 5.9$ et $\sigma^2 = 2.62$
- On calcule $\bar{x}$ et $s^2$ sur un échantillon de 40 étudiants.
- $\bar{x} = 5.8462$ et $s^2 = 3.8225$.

$$\begin{bmatrix} x\_i & n\_i \\ 3 & 1 \\ 4 & 10 \\ 5 & 11 \\ 6 & 3 \\ 7 & 7 \\ 8 & 4 \\ 9 & 2 \\ 13 & 1 \end{bmatrix}$$

# Avec vos calculatrices TI

# Avec vos calculatrices TI


L'écart-type de l'échantillon est nommé $\sigma x$.

# Avec vos calculatrices casio

Entrer les données

MENU 2

Faire les calculs

MENU 2 F2 F6 ▼ F2 2 EXE

EXIT F1

Accéder aux résultats

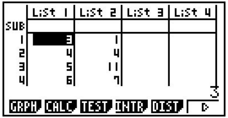
MENU 1

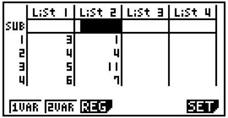
VARS F3 F1

F1 EXE

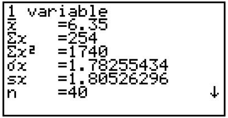
F2 EXE

F5 $x^2$ EXE








# Avec vos calculatrices casio

Entrer les données

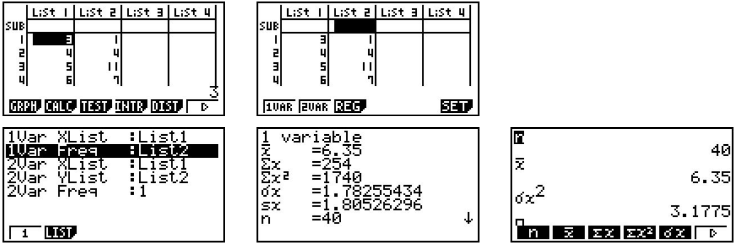
MENU 2

Faire les calculs

MENU 2 F2 F6 ▼ F2 2 EXE
EXIT F1

Accéder aux résultats

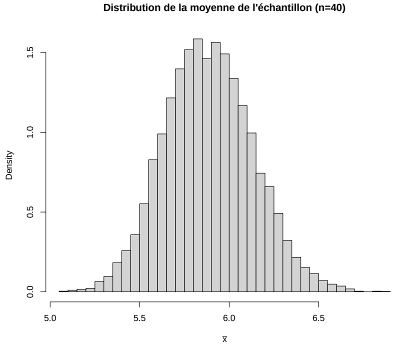
MENU 1
VARS F3 F1
F1 EXE
F2 EXE
F5 $x^2$ EXE




L'écart-type de l'échantillon est nommé $\sigma x$.

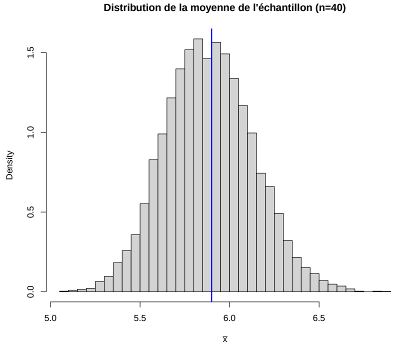
# Distribution de $\bar{x}$

- $\bar{x}$ est une V.A.
- La distribution de $\bar{x}$ tend vers la loi normale.
- La moyenne de $\bar{x}$ est $\mu$
- La variance de $\bar{x}$ est $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$
- $\bar{x}$ est un estimateur non biaisé de $\mu$ : $\hat{\mu} = \bar{x}$.



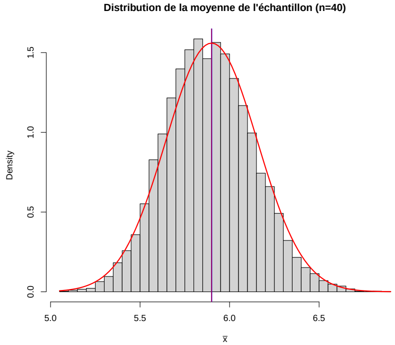
# Distribution de $\bar{x}$

- $\bar{x}$ est une V.A.
- La distribution de $\bar{x}$ tend vers la loi normale.
- La moyenne de $\bar{x}$ est $\mu$
- La variance de $\bar{x}$ est $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$
- $\bar{x}$ est un estimateur non biaisé de $\mu$ : $\hat{\mu} = \bar{x}$.



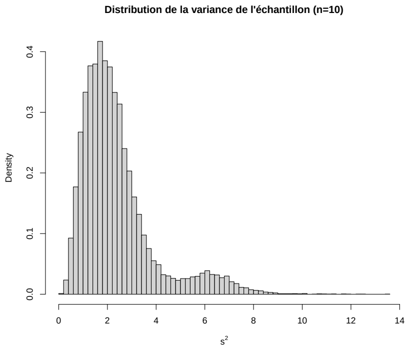
# Distribution de $\bar{x}$

- $\bar{x}$ est une V.A.
- La distribution de $\bar{x}$ tend vers la loi normale.
- La moyenne de $\bar{x}$ est $\mu$
- La variance de $\bar{x}$ est $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}$
- $\bar{x}$ est un estimateur non biaisé de $\mu$ : $\hat{\mu} = \bar{x}$.



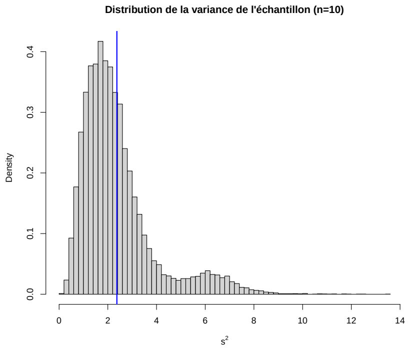
# Distribution de $s^2$

- $s^2$ est une V.A.



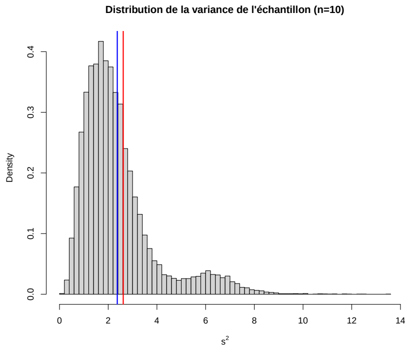
# Distribution de $s^2$

- $s^2$ est une V.A.



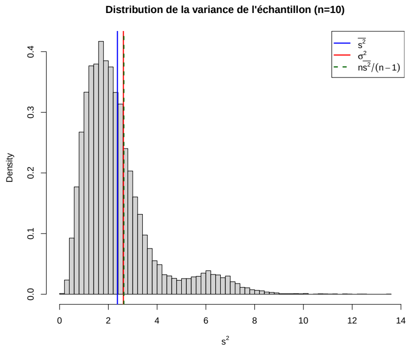
# Distribution de $s^2$

- $s^2$ est une V.A.
- La moyenne de $s^2$ est différente de $\sigma^2$



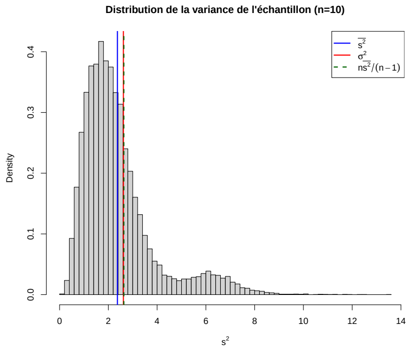
# Distribution de $s^2$

- $s^2$ est une V.A.
- La moyenne de $s^2$ est différente de $\sigma^2$
- On peut montrer que $\mathbb{E}\left(s^2\right) = \frac{n-1}{n}\sigma^2$



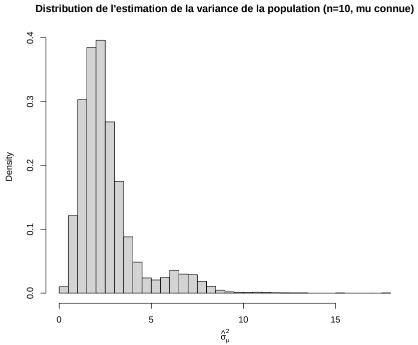
# Distribution de $s^2$

- $s^2$ est une V.A.
- La moyenne de $s^2$ est différente de $\sigma^2$
- On peut montrer que $\mathbb{E}\left(s^2\right) = \frac{n-1}{n}\sigma^2$
- Un estimateur non biaisé de $\sigma^2$ est $\hat{\sigma}^2 = \frac{n}{n-1}s^2$



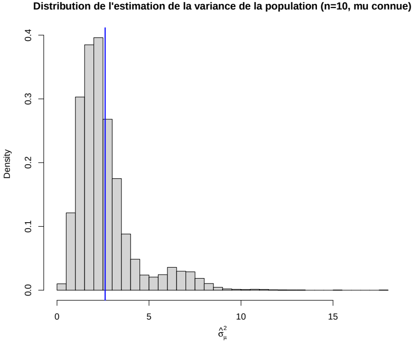
# Comment estimer $\sigma^2$ (cas où $\mu$ est connue)

- Supposons que la moyenne $\mu$ est connue. On calcule $\hat{\sigma}^2_\mu = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$.



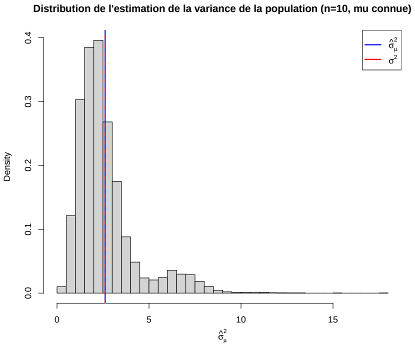
# Comment estimer $\sigma^2$ (cas où $\mu$ est connue)

- Supposons que la moyenne $\mu$ est connue. On calcule $\hat{\sigma}^2_\mu = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$.
- On peut montrer que $\mathbb{E}\left(\hat{\sigma}^2_\mu\right) = \sigma^2$



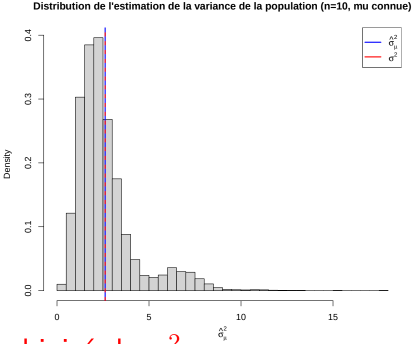
# Comment estimer $\sigma^2$ (cas où $\mu$ est connue)

- Supposons que la moyenne $\mu$ est connue. On calcule $\hat{\sigma}^2_\mu = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$.
- On peut montrer que $\mathbb{E}\left(\hat{\sigma}^2_\mu\right) = \sigma^2$
- Le plus souvent on veut estimer $\mu$ et $\sigma^2$.



# Comment estimer $\sigma^2$ (cas où $\mu$ est connue)

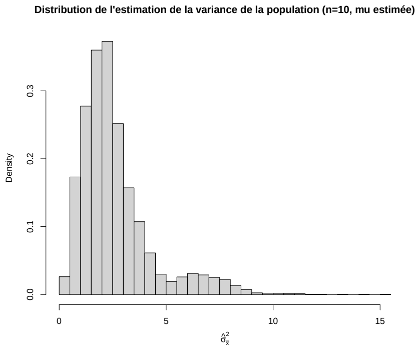
- Supposons que la moyenne $\mu$ est connue. On calcule $\hat{\sigma}^2_\mu = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$.
- On peut montrer que $\mathbb{E}\left(\hat{\sigma}^2_\mu\right) = \sigma^2$
- Le plus souvent on veut estimer $\mu$ et $\sigma^2$.




$\hat{\sigma}^2_\mu = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$ est un estimateur non biaisé de $\sigma^2$.

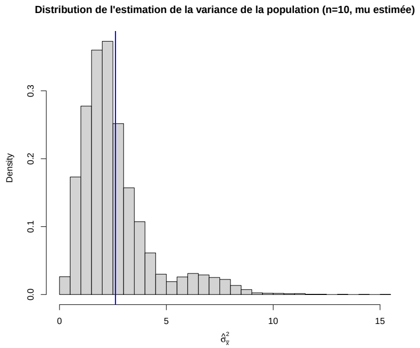
# Comment estimer $\sigma^2$ (cas où $\mu$ est inconnue)

- On calcule $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{j=1}^{n} x_j\right)^2$.



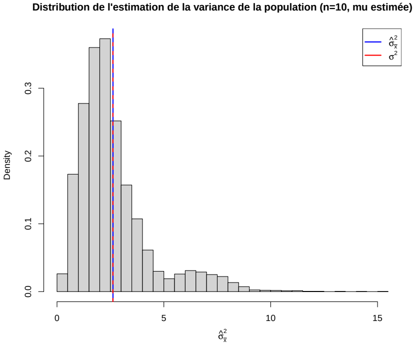
# Comment estimer $\sigma^2$ (cas où $\mu$ est inconnue)

- On calcule $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{j=1}^{n} x_j\right)^2$.
- On peut montrer que $\mathbb{E}\left(s^2\right) = \frac{n-1}{n}\sigma^2$.



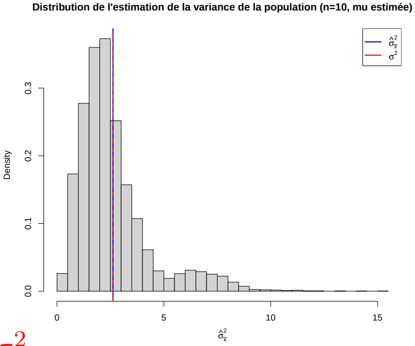
# Comment estimer $\sigma^2$ (cas où $\mu$ est inconnue)

- On calcule $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{j=1}^{n} x_j\right)^2$.
- On peut montrer que $\mathbb{E}\left(s^2\right) = \frac{n-1}{n}\sigma^2$.
- Finalement, on pose $\hat{\sigma}_{\bar{x}}^2 = \frac{n}{n-1}s^2$ et $\mathbb{E}\left(\hat{\sigma}_{\bar{x}}^2\right) = \sigma^2$



# Comment estimer $\sigma^2$ (cas où $\mu$ est inconnue)

- On calcule $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{j=1}^{n} x_j\right)^2$.
- On peut montrer que $\mathbb{E}\left(s^2\right) = \frac{n-1}{n}\sigma^2$.
- Finalement, on pose $\hat{\sigma}_{\bar{x}}^2 = \frac{n}{n-1}s^2$ et $\mathbb{E}\left(\hat{\sigma}_{\bar{x}}^2\right) = \sigma^2$



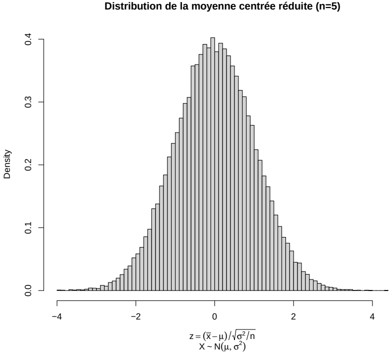

$\hat{\sigma}_{\bar{x}}^2 = \frac{n}{n-1}s^2$ est un estimateur non biaisé de $\sigma^2$.

# Distribution d'échantillonnage

- L'échantillon est composé de $n$ observations $X_1, \dots, X_n$ indépendantes et de même distribution $\mathcal{D}(\mu, \sigma^2)$.
- La moyenne $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est une variable aléatoire. $\mathbb{E}(\bar{X}) = \mu$, $\mathbb{V}(\bar{X}) = \frac{\sigma^2}{n}$
- La distribution de $\bar{X}$ est la distribution d'échantillonnage.
- La distribution de $\bar{X}$ dépend de celle des $X_i$.
- Les fractiles de la distribution de $\bar{X}$ permettent d'établir des intervalles de prédiction (sur la valeur de $\bar{X}$).
- Nous sommes intéressés par l'estimation de $\mu$ : on parle d'intervalles de confiance.

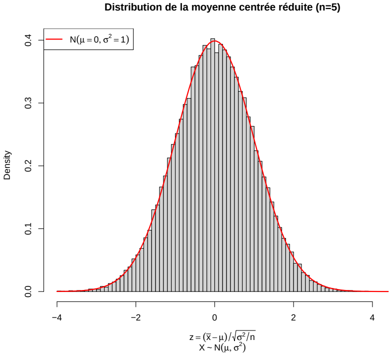
# $X_i \sim \mathcal{N}(\mu, \sigma^2)$ et $\sigma^2$ est connue.

- $\bar{X}$ est la moyenne de $n$ VA. distribuées normalement $X_i \sim \mathcal{N}(\mu, \sigma^2)$.



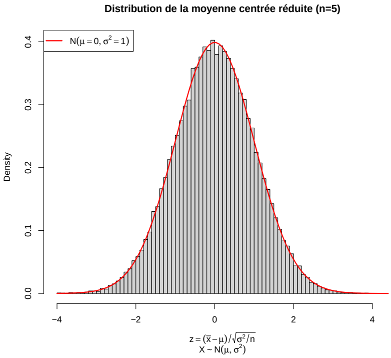
# $X_i \sim \mathcal{N}(\mu, \sigma^2)$ et $\sigma^2$ est connue.

- $\bar{X}$ est la moyenne de $n$ VA. distribuées normalement $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
- Donc $\bar{x}$ suit aussi une loi normale $N\left(\mu, \frac{\sigma^2}{n}\right)$
- $z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$.



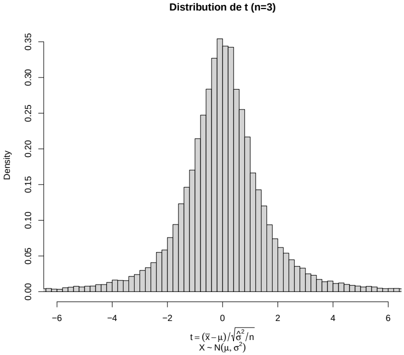
# $X_i \sim \mathcal{N}(\mu, \sigma^2)$ et $\sigma^2$ est connue.

- $\bar{X}$ est la moyenne de $n$ VA. distribuées normalement $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
- Donc $\bar{x}$ suit aussi une loi normale $N\left(\mu, \frac{\sigma^2}{n}\right)$
- $z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$.
- Les fractiles de $\mathcal{N}(\mu = 0, \sigma^2 = 1)$ sont obtenus avec la table de l'écart-réduit.



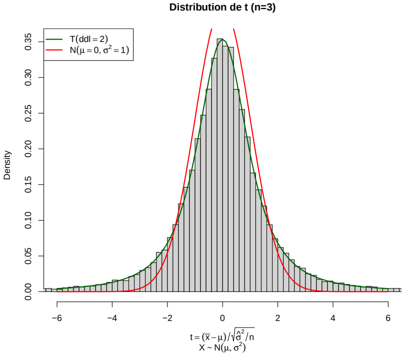
# $X_i \sim \mathcal{N}(\mu, \sigma^2)$ et $\sigma^2$ est inconnue.

- $\bar{X}$ est la moyenne de $n$ VA. iid $X_i \sim \mathcal{N}(\mu, \sigma^2)$.



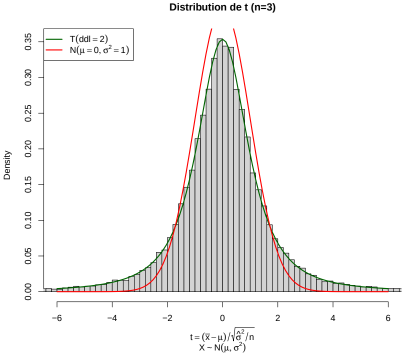
# $X_i \sim \mathcal{N}(\mu, \sigma^2)$ et $\sigma^2$ est inconnue.

- $\bar{X}$ est la moyenne de $n$ VA. iid $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
- $t = \frac{\bar{X}-\mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n-1}}} \sim \mathcal{T}(n-1\text{ddl})$.



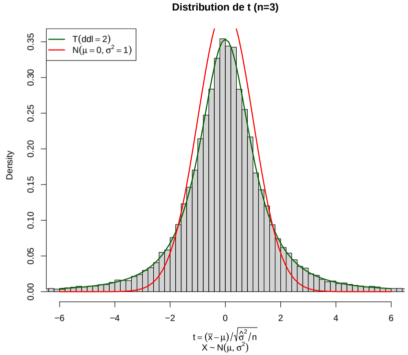
# $X_i \sim \mathcal{N}(\mu, \sigma^2)$ et $\sigma^2$ est inconnue.

- $\bar{X}$ est la moyenne de $n$ VA. iid $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
- $t = \frac{\bar{X}-\mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n-1}}} \sim \mathcal{T}(n-1\mathrm{ddl})$.
- $\mathcal{T}(n-1\mathrm{ddl})$ est la loi de Student à $n-1$ degrés de liberté.



# $X_i \sim \mathcal{N}(\mu, \sigma^2)$ et $\sigma^2$ est inconnue.

- $\bar{X}$ est la moyenne de $n$ VA. iid $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
- $t = \frac{\bar{X}-\mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n-1}}} \sim \mathcal{T}(n-1\text{ddl})$.
- $\mathcal{T}(n-1\text{ddl})$ est la loi de Student à $n-1$ degrés de liberté.
- $\mathcal{T}(n-1\text{ddl}) \to \mathcal{N}(0,1)$ lorsque $n$ est grand.



# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$,
  $X^2 = \sum_{i=1}^n X_i^2$ alors
  $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$

# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$,
  $X^2 = \sum_{i=1}^{n} X_i^2$ alors
  $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si
  $X_i \sim \mathcal{N}(\mu, \sigma^2)$

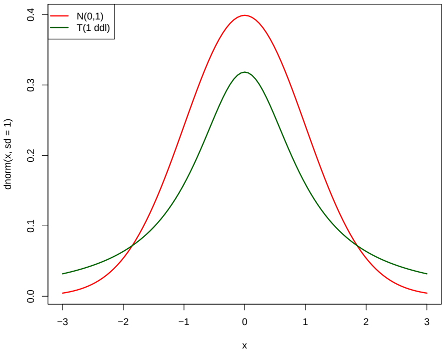
$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n - 1 \text{ ddl})$$

# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$, $X^2 = \sum_{i=1}^n X_i^2$ alors $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n-1 \text{ ddl})$$

- Lorsque $n$ croît $\mathcal{T}(n \text{ ddl}) \to \mathcal{N}(0,1)$.

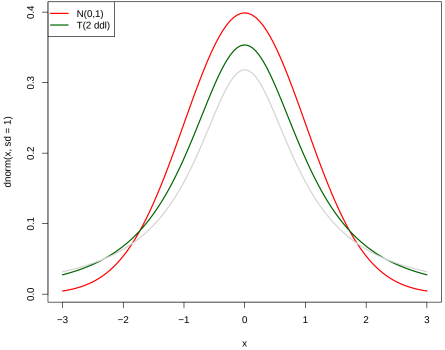


# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$, $X^2 = \sum_{i=1}^n X_i^2$ alors $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n-1 \text{ ddl})$$

- Lorsque $n$ croît $\mathcal{T}(n \text{ ddl}) \to \mathcal{N}(0,1)$.

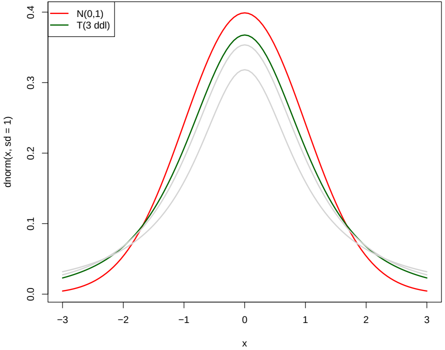


# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$, $X^2 = \sum_{i=1}^{n} X_i^2$ alors $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n - 1 \text{ ddl})$$

- Lorsque $n$ croît $\mathcal{T}(n \text{ ddl}) \to \mathcal{N}(0,1)$.

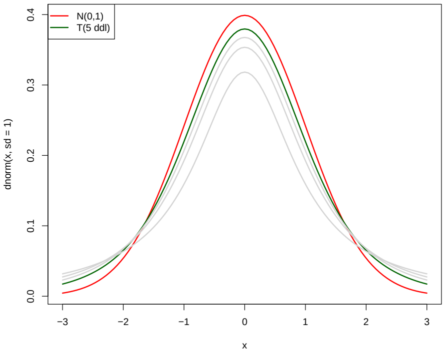


# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$, $X^2 = \sum_{i=1}^n X_i^2$ alors $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n - 1 \text{ ddl})$$

- Lorsque $n$ croît $\mathcal{T}(n \text{ ddl}) \to \mathcal{N}(0,1)$.

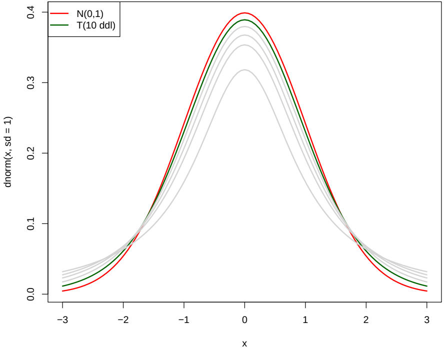


# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$, $X^2 = \sum_{i=1}^n X_i^2$ alors $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n-1 \text{ ddl})$$

- Lorsque $n$ croît $\mathcal{T}(n \text{ ddl}) \to \mathcal{N}(0,1)$.

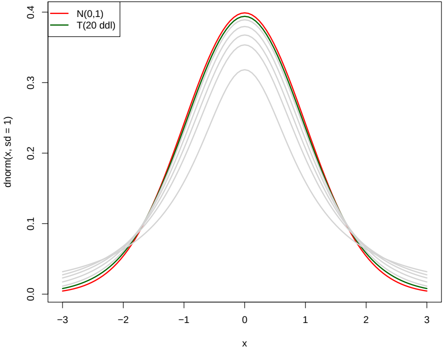


# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$, $X^2 = \sum_{i=1}^n X_i^2$ alors $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n - 1 \text{ ddl})$$

- Lorsque $n$ croît $\mathcal{T}(n \text{ ddl}) \to \mathcal{N}(0,1)$.

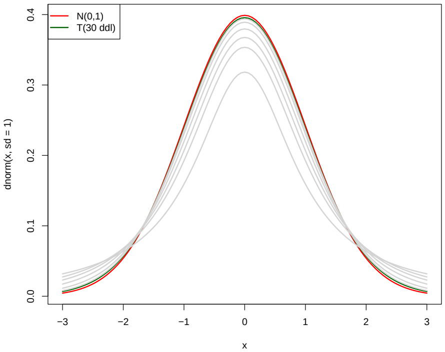


# Loi de $t$ de Student

- Si $X_0, \ldots, X_n$ iid $\mathcal{N}(0,1)$, $X^2 = \sum_{i=1}^{n} X_i^2$ alors $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n - 1 \text{ ddl})$$

- Lorsque $n$ croît $\mathcal{T}(n \text{ ddl}) \to \mathcal{N}(0,1)$.



# Loi de $t$ de Student

- Si $X_0, \dots, X_n$ iid $\mathcal{N}(0,1)$, $X^2 = \sum_{i=1}^{n} X_i^2$ alors $t = \frac{X_0}{\sqrt{X^2/n}} \sim \mathcal{T}(n \text{ ddl})$
- En ce qui nous concerne si $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n - 1 \text{ ddl})$$

- Lorsque $n$ croît $\mathcal{T}(n \text{ ddl}) \to \mathcal{N}(0,1)$.
- Les fractiles de $t$ sont donnés par la table de la loi de Student.

Table de $t$.

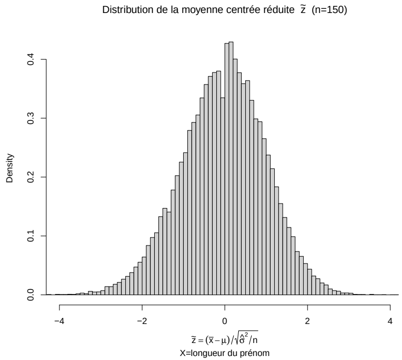
La table donne, pour une probabilité $\alpha$ et un nombre de degrés de liberté (d.d.l.) donnés, la valeur $x$ telle que la probabilité que $t$ égale ou dépasse $x$ en valeur absolue vaut $\alpha$.

| d.d.l. \ $\alpha$ | 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 | 31.599 |
| 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 | 3.768 |
| 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 |
| $\infty$ | 0.126 | 0.674 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.291 |

*Exemple :* Une variable aléatoire suivant une loi de $t$ avec 10 degrés de liberté a une probabilité $\alpha = 0.05$ d'être égale ou supérieure à 2.228 en valeur absolue.

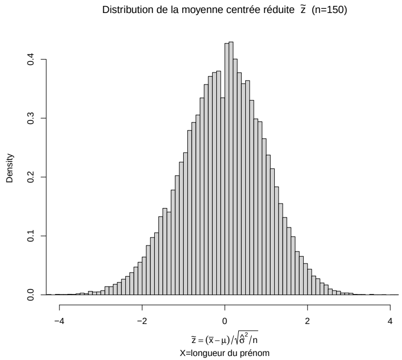
# $X \sim \mathcal{D}(\mu, \sigma^2)$ quelconque mais $n$ est grand : utilisation du TCL

- $\bar{X}$ est la moyenne de $n$ VA. iid d'espérance $\mu$ et de variance $\sigma^2$.



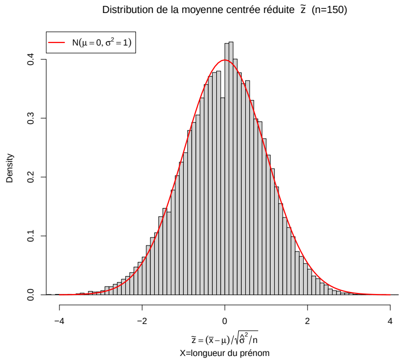
# $X \sim \mathcal{D}(\mu, \sigma^2)$ quelconque mais $n$ est grand : utilisation du TCL

- $\bar{X}$ est la moyenne de $n$ VA. iid d'espérance $\mu$ et de variance $\sigma^2$.
- Le TCL dit que $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.



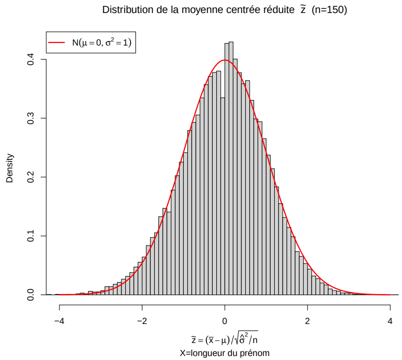
# $X \sim \mathcal{D}(\mu, \sigma^2)$ quelconque mais $n$ est grand : utilisation du TCL

- $\bar{X}$ est la moyenne de $n$ VA. iid d'espérance $\mu$ et de variance $\sigma^2$.
- Le TCL dit que $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
- Donc $Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1)$.



# $X \sim \mathcal{D}(\mu, \sigma^2)$ quelconque mais $n$ est grand : utilisation du TCL

- $\bar{X}$ est la moyenne de $n$ VA. iid d'espérance $\mu$ et de variance $\sigma^2$.
- Le TCL dit que $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
- Donc $Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1)$.
- $T = \frac{\bar{X}-\mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n-1 \text{ ddl})$ Comme $n$ est grand $\mathcal{T}(n-1 \text{ ddl}) \to \mathcal{N}(0,1)$



# Distribution d'échantillonnage de la moyenne $\bar{X}$ : synthèse

| | $\sigma^2$ | $n < 30$ | $n \geqslant 30$ |
|---|---|---|---|
| $X_i \sim \mathcal{N}(\mu, \sigma^2)$ | connue | $z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1)$ | |
| | estimée | $t = \frac{\bar{X}-\mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \mathcal{T}(n-1 \text{ ddl})$ | $\mathcal{T}(n-1 \text{ ddl}) \to \mathcal{N}(0,1)$ |
| $X_i \sim \mathcal{D}(\mu, \sigma^2)$ quelconque | connue | on ne sait pas | $z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1)$ |
| | estimée | on ne sait pas | $t = \frac{\bar{X}-\mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \sim \ \sim \mathcal{T}(n-1 \text{ ddl}) \to \mathcal{N}(0,1)$ |

# Intervalle de prédiction (dit aussi de pari ou de fluctuation)

- Les fractiles des distributions d'échantillonnage permettent d'établir des intervalles de prédiction pour la valeur de $\bar{X}$.
- Ces intervalles de prédiction dépendent de $\mu$, $\sigma^2$ et $n$.
- La lecture des fractiles se fait dans la table de l'écart-réduit (pour $\mathcal{N}(0,1)$) ou celle de la loi de Student ($\mathcal{T}(k \text{ ddl})$)

# Intervalle de prédiction : Exemple

On considère $n = 500$ lancers indépendants d'un dé à 6 faces équilibré. $X_i$ est le résultat du $i$ème lancer. On veut un intervalle de prédiction pour $\bar{X}$ :

$$\mathbb{E}(X_i) = \frac{1}{6}\sum_{i=1}^{6} i = \frac{7}{2}, \quad \mathbb{V}(X_i) = \frac{1}{6}\sum_{i=1}^{6} i^2 - \left(\frac{21}{6}\right)^2 = \frac{35}{12}$$

Un intervalle de prédiction au risque $\alpha$ pour $\bar{X}$ est donné par la table de l'écart-réduit :

$$\mathbb{P}\left(-\varepsilon_\alpha < \frac{\bar{X} - \frac{7}{2}}{\sqrt{\frac{35/12}{500}}} < \varepsilon_\alpha\right) = 1 - \alpha$$

# Intervalle de prédiction : Exemple

Si on souhaite prendre un risque $\alpha = 0.001$, la valeur à utiliser est $\varepsilon_{\alpha=0.001} = 3.29053$. On obtient alors :

$$\mathbb{P}\left(\frac{7}{2} - 3.29053\sqrt{\frac{35/12}{500}} < \bar{X} < \frac{7}{2} + 3.29053\sqrt{\frac{35/12}{500}}\right) = 0.999$$

soit finalement

$$\mathbb{P}\left(3.249 < \bar{X} < 3.751\right) = 0.999$$

Ce qui signifie que la somme des résultats des 500 lancers sera comprise entre 1624 et 1876 avec une probabilité de 0.999.

# Intervalle de confiance

- Contrairement à l'intervalle de prédiction, on cherche à inférer une connaissance à propos de la distribution dans la population à partir d'un résultat sur un échantillon.
- Un intervalle de confiance de la moyenne $\mu$ au risque $\alpha$ est un intervalle dans lequel la moyenne $\mu$ (qui reste inconnue) se trouve avec une probabilité $1 - \alpha$.

$$\mathbb{P}(b_{\text{inf}} < \mu < b_{\text{sup}}) = 1 - \alpha$$

- L'intervalle de confiance résulte d'une expérience sur un échantillon : ses bornes $b_{\text{inf}}$ et $b_{\text{sup}}$ sont donc des variables aléatoires.

# Relation intervalle de prédiction / intervalle de confiance

- Distribution d'échantillonnage, et fractiles de cette distribution

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1) \quad \text{donc} \quad \mathbb{P}\left(-\varepsilon_\alpha < \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} < \varepsilon_\alpha\right) = 1 - \alpha$$

# Relation intervalle de prédiction / intervalle de confiance

- Distribution d'échantillonnage, et fractiles de cette distribution

$$\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1) \quad \text{donc} \quad \mathbb{P}\left(-\varepsilon_\alpha < \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} < \varepsilon_\alpha\right) = 1-\alpha$$

- Intervalle de prédiction de $\bar{X}$ au risque $\alpha$

$$\mathbb{P}\left(\mu - \varepsilon_\alpha\sqrt{\frac{\sigma^2}{n}} < \bar{X} < \mu + \varepsilon_\alpha\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha$$

# Relation intervalle de prédiction / intervalle de confiance

- Distribution d'échantillonnage, et fractiles de cette distribution

$$\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0,1) \quad \text{donc} \quad \mathbb{P}\left(-\varepsilon_\alpha < \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} < \varepsilon_\alpha\right) = 1-\alpha$$

- Intervalle de prédiction de $\bar{X}$ au risque $\alpha$

$$\mathbb{P}\left(\mu - \varepsilon_\alpha\sqrt{\frac{\sigma^2}{n}} < \bar{X} < \mu + \varepsilon_\alpha\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha$$

- Intervalle de confiance de la moyenne $\mu$ au risque $\alpha$

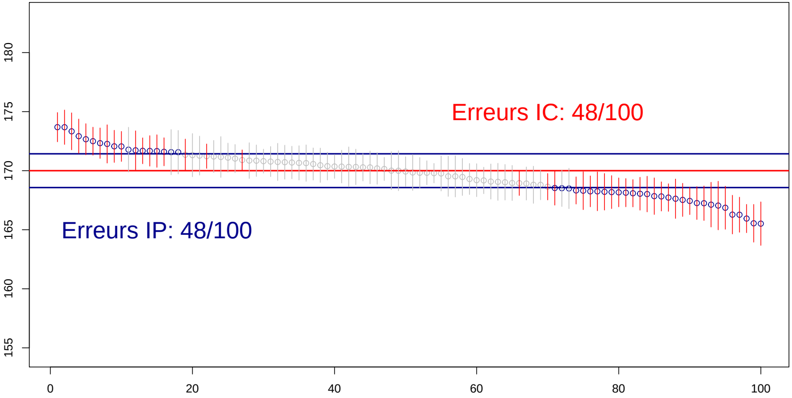
$$\mathbb{P}\left(\bar{X} - \varepsilon_\alpha\sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + \varepsilon_\alpha\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha$$

# Relation intervalle de prédiction / intervalle de confiance

Simulations : $n = 15$, $X_i \sim \mathcal{N}(\mu = 170, \sigma^2 = 64)$

# Relation intervalle de prédiction / intervalle de confiance

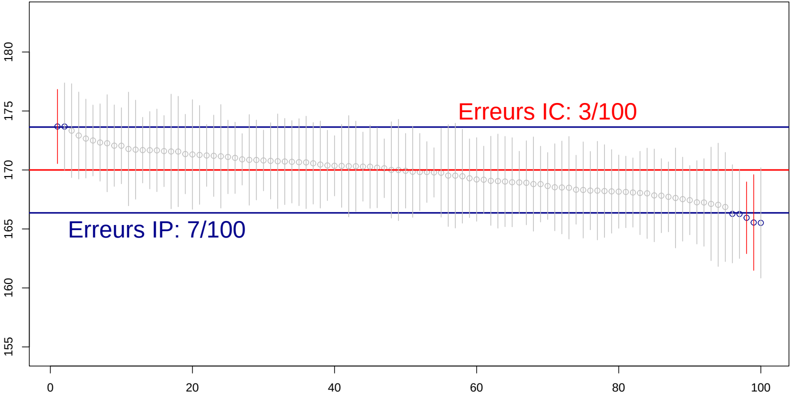
Simulations : $n = 15$, $X_i \sim \mathcal{N}(\mu = 170, \sigma^2 = 64)$




Risque 0.5

# Relation intervalle de prédiction / intervalle de confiance

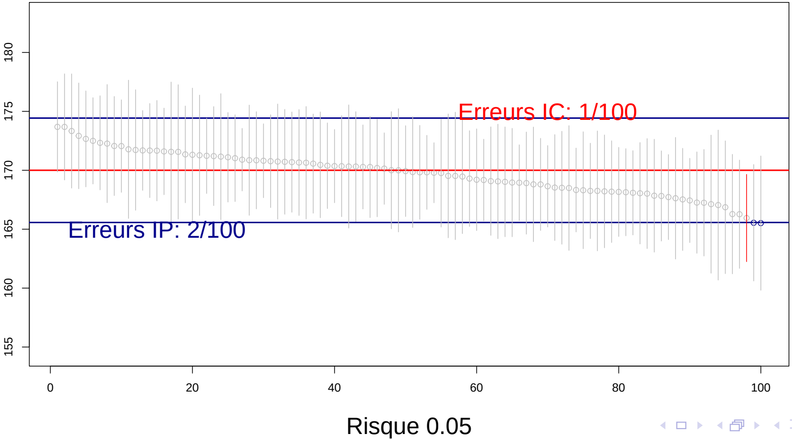
Simulations : $n = 15$, $X_i \sim \mathcal{N}(\mu = 170, \sigma^2 = 64)$




Risque 0.1

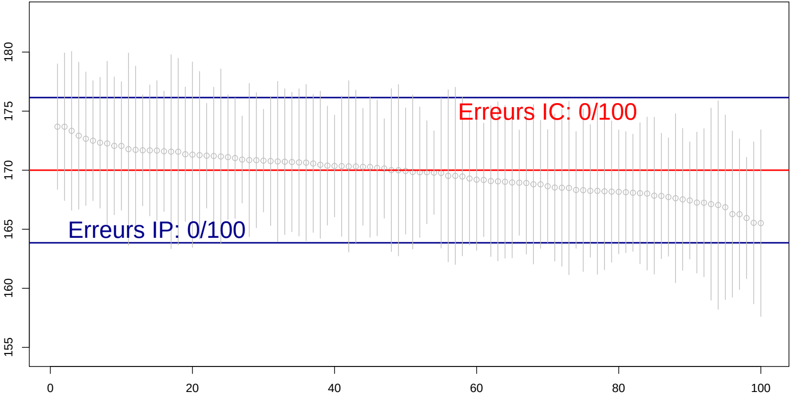
# Relation intervalle de prédiction / intervalle de confiance

Simulations : $n = 15,\ X_i \sim \mathcal{N}(\mu = 170, \sigma^2 = 64)$



# Relation intervalle de prédiction / intervalle de confiance

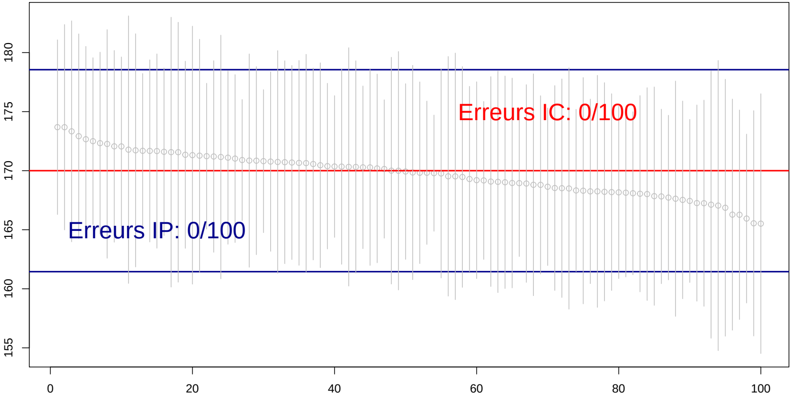
Simulations : $n = 15,\ X_i \sim \mathcal{N}(\mu = 170, \sigma^2 = 64)$




Risque 0.01

# Relation intervalle de prédiction / intervalle de confiance

Simulations : $n = 15$, $X_i \sim \mathcal{N}(\mu = 170, \sigma^2 = 64)$




Risque 0.001

# Intervalle de confiance : Exemple

À l'issue de $n = 500$ lancers de dé, on a obenu une moyenne $\bar{x} = 3.834$. La somme des lancers était $\sum_{i=1}^{500} X_i = 1917$ et la somme des carrés des lancers était $\sum_{i=1}^{500} X_i^2 = 8959$. Peut-on en conclure que le dé est déséquilibré ?

$$\mathbb{P}\left(-\varepsilon_\alpha < \frac{\bar{X} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{n}}} < \varepsilon_\alpha\right) = 1 - \alpha \Leftrightarrow \mathbb{P}\left(-\varepsilon_\alpha < \frac{\mu - \bar{X}}{\sqrt{\frac{\hat{\sigma}^2}{n}}} < \varepsilon_\alpha\right) = 1 - \alpha$$

$$\mathbb{P}\left(\bar{X} - \varepsilon_\alpha \sqrt{\frac{\hat{\sigma}^2}{n}} < \mu < \bar{X} + \varepsilon_\alpha \sqrt{\frac{\hat{\sigma}^2}{n}}\right) = 1 - \alpha$$

# Intervalle de confiance : Exemple

Dans le cas qui nous concerne, on a $\bar{X} = \frac{1917}{500} = 3.834$ et $\hat{\sigma}^2 = \frac{500}{499}\left(\frac{1}{500} \times 8959 - 3.834^2\right) = 3.2248938.$
Si on considère un intervalle de confiance au risque $\alpha = 0.001$, on obtient :

$$\mathbb{P}\left(3.834 - 3.29053\sqrt{\frac{3.2248938}{500}} < \mu < 3.834 + 3.29053\sqrt{\frac{3.2248938}{500}}\right) = 0.999$$

soit finalement

$$\mathbb{P}\left(3.57 < \mu < 4.098\right) = 0.999$$

La valeur théorique de $\mu = 3.5$ pour un dé équilibré ne fait pas partie de l'intervalle de confiance au risque $0.001$.

# Intervalle de confiance de la moyenne $\mu$ : synthèse

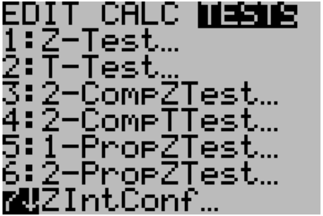
| Dist. de $X_i$ | $\sigma^2$ | $n < 30$ | $n \geqslant 30$ |
|---|---|---|---|
| $\mathcal{D}(\mu, \sigma^2)$ | connue | on ne sait pas | $\mathbb{P}\left(\bar{X} - \varepsilon_\alpha \sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + \varepsilon_\alpha \sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha$ |
| | estimée [1] | on ne sait pas | $\mathbb{P}\left(\bar{X} - \varepsilon_\alpha \sqrt{\frac{\hat{\sigma}^2}{n}} < \mu < \bar{X} + \varepsilon_\alpha \sqrt{\frac{\hat{\sigma}^2}{n}}\right) = 1 - \alpha$ |
| $\mathcal{N}(\mu, \sigma^2)$ | connue | $\mathbb{P}\left(\bar{X} - \varepsilon_\alpha \sqrt{\frac{\sigma^2}{n}} < \mu < \bar{X} + \varepsilon_\alpha \sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha$ | |
| | estimée | $\mathbb{P}\left(\bar{X} - t_{\alpha, n-1\,\mathrm{ddl}} \sqrt{\frac{\hat{\sigma}^2}{n}} < \mu < \bar{X} + t_{\alpha, n-1\,\mathrm{ddl}} \sqrt{\frac{\hat{\sigma}^2}{n}}\right) = 1 - \alpha$ | |

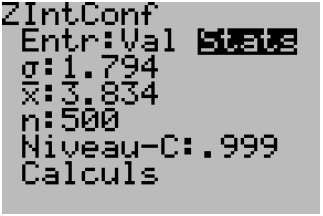
1. Le résultat donné ici tient compte de la convergence $\mathcal{T}(n-1\ \mathrm{ddl}) \to \mathcal{N}(0,1)$

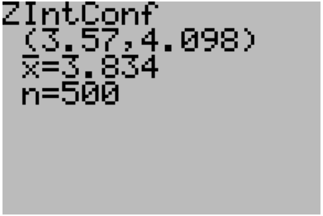
# Avec vos calculatrices TI ou casio

On reprend l'exemple du dé : $n = 500$, $\sum X = 1917$, $\sum X^2 = 8959$ soit $\bar{X} = 3.834$ et $\sqrt{s^2} = 1.794$.

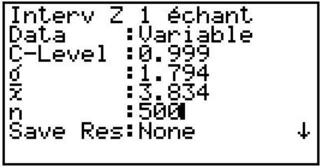
On a un grand échantillon et une distribution quelconque. C'est un IC de type "$z$".

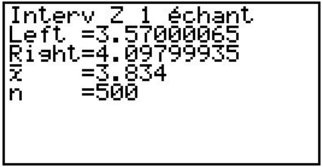



On accède au menu stat.
[MENU][2]
On choisit INTR, puis Z, puis 1-s
[F4][F1][F1]



# Conclusions.

- Nous avons estimé la moyenne $\mu$ de la population grâce à un échantillon.
- D'autres paramètres ($\sigma^2$) peuvent être estimés de la même façon.
- Un estimateur $\hat{\mu}$ est une variable aléatoire.
- Les bornes d'un intervalle de confiance sont des variables aléatoires (contrairement à celles d'un intervalle de prédiction).
- Un intervalle de confiance de $\mu$ au risque $\alpha$ contient la moyenne $\mu$ avec une probabilité $1 - \alpha$.
- Il existe aussi des intervalles de confiance unilatéraux.
- Les intervalles de confiance sont à la base des tests d'hypothèse de conformité.