Économie et management.
Licence 2.
Statistiques.

Année
2014 - 2015

Chapitre 3 :

Éléments de théorie d'échantillonnage

Cristel Schiltz.

I _ Introduction.

→ Notion : échantillon aléatoire.
→ Population : N .
→ Moyenne : μ .
→ Écart-type : σ .
→ Dispersion autour de la moyenne.

→ μ inconnu : estimer μ .
→ Inférence statistique : s'appuyer sur les données obtenues par un échantillon n .
→ Pour obtenir des informations sur les données de la population N .

→ **Échantillon aléatoire : représentatif de la population**.
→ Population : N .
→ Échantillon de taille n .
→ Tirage avec remise.

→ **Notion : distribution statistique de moyenne d'échantillon**.
→ Extraction de plusieurs échantillons : $i = 1, \ldots, p$ échantillons.
→ Pour chaque échantillon : $\forall\ i \in [0,\ p]$.
→ Calcul de la moyenne : $E(\bar{x}_i) = \bar{x}_i$.

Échantillon	$\bar{x}_i$
1	$\bar{x}_1$
...	...
i	$\bar{x}_i$
...	...
p	$\bar{x}_p$

→ **Notation**.

	Taille	Moyenne	Écart-type
Population	N	μ	σ
Échantillon	n	$\bar{x}$	s

II _ Distribution d'échantillonnage : théorème Central Limite et loi Normale.

1 _ Conditions d'application.

→ **Théorème Central Limite : conditions d'application**.

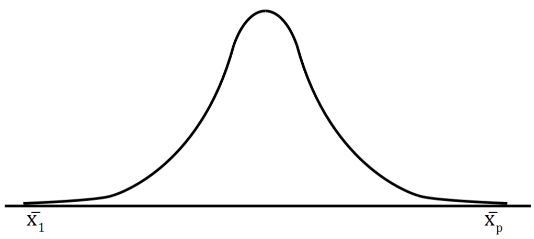
→ **Taille minimum d'échantillon :** $\frac{n}{N} \geq 0{,}05$.

→ **Si** $N \rightarrow +\infty$ **: alors** $n \rightarrow 0$.

→ **Conditions d'application réunies**.

→ Échantillon : tout tirage sera toujours aléatoire et représentatif de la population.

→ Distribution de $\bar{x}_i$: suit une loi normale.



2 _ Moment de la distribution de $\bar{x}_i$.

→ $E(\bar{x}_i) = \frac{1}{n} * \sum_{i=1}^{p} \bar{x}_i$.

→ **Chaque échantillon : représentatif**.

→ $\bar{x}_i \simeq \mu$.

→ $E(\bar{x}_i) = \frac{1}{n} * n * \mu = \mu$.

→ $E(\bar{x}_i) = \mu$.

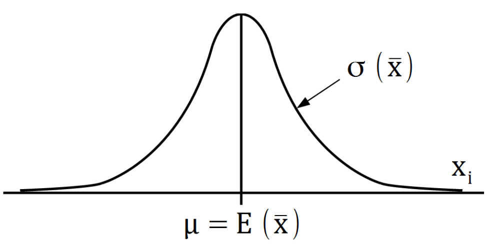
→ **Écart-type de la population :** σ .

→ **Erreur-type d'échantillon :** $\sigma(\bar{x})$.

→ **Si** $N \rightarrow +\infty$ **:** $\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}$.

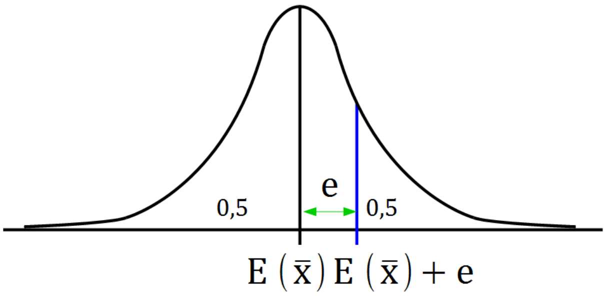
→ **Si** N **fini :** $\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}}$.

→ **Coefficient de correction :** $\sqrt{\frac{N-n}{N-1}}$.



→ **Loi normale : forme.**

→ e : marge d'erreur.



$$\mu \in \left[E\left(\bar{x}\right),\ E\left(\bar{x}\right)+e\right]$$

3 _ Variable aléatoire centrée et réduite.

→ **Variable aléatoire centrée réduite de la loi normale** Z .

→ $Z = \dfrac{E\left(\bar{x}\right)+e-E\left(\bar{x}\right)}{\sigma\left(\bar{x}\right)} = \dfrac{e}{\sigma\left(\bar{x}\right)}$.

→ **Si** $N \to +\infty$ **:** $Z = \dfrac{e}{\frac{\sigma}{\sqrt{n}}}$.

→ **Distribution de** Z **: proportionnelle à la distribution de** $\bar{x}$.

→ **Distribution de** Z .

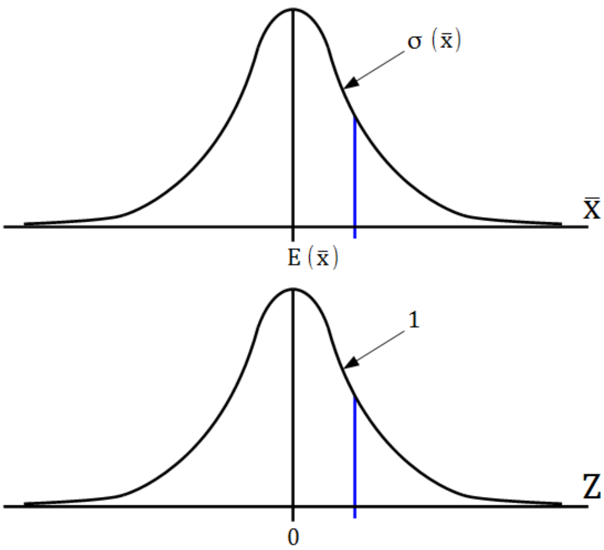
→ Moyenne : 0 .

→ Erreur-type : 1 .

→ **Distribution de** $\bar{x}$.

→ Moyenne : $E\left(\bar{x}\right)$.

→ Erreur-type : $\sigma\left(\bar{x}\right)$.



→ **Table de loi normale centrée et réduite : lecture**.

→ En ligne : unité et première décimale.

→ En colonne : deuxième décimale.

→ **Méthodologie : cinq étapes**.

→ Énoncer la loi : conditions et paramètres.

→ Régionaliser le problème : hachurer la zone cherchée.

→ Calcule de $Z = \dfrac{e}{\frac{\sigma}{\sqrt{n}}}$.

→ Lecture dans la table de loi normale : $P\left(0 < Z_i < Z\right)$.

→ Conclusion.

→ **Exemple : énoncé**.

→ Soit une population N infinie.

→ On extrait un échantillon $n = 30$.

→ De moyenne $E\left(\bar{x}\right) = 51800$.

→ D'écart type d'échantillon $\sigma = 4000$.

→ **Résolution : probabilité que μ soit compris entre $E\left(\bar{x}\right)$ et 52300** .

→ Énoncer la loi : conditions et paramètres.

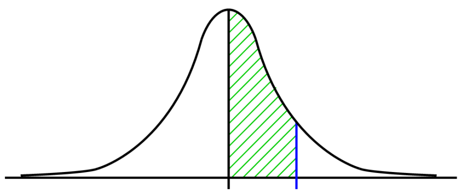
→ Conditions : $\bar{x} \to N$ car $N \to +\infty$.

→ Paramètres.

→ $E\left(\bar{x}\right) = 51800$

→ $\sigma\left(\bar{x}\right) = \dfrac{\sigma}{\sqrt{n}} = \dfrac{4000}{\sqrt{30}} = 730{,}3$.

→ Régionaliser le problème.


51800 52300

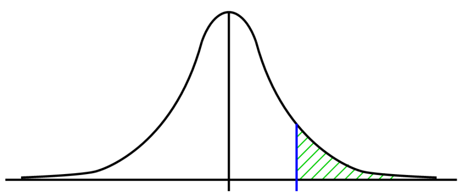
→ $Z = \dfrac{e}{\frac{\sigma}{\sqrt{n}}} = \dfrac{52300 - 51800}{730{,}3} = \dfrac{500}{730{,}3} = 0{,}68$.

→ Lecture dans la table de loi normale : $P\left(0 < Z < 0{,}68\right) = 0{,}2517$.

→ $P\left(51800 < \mu < 52300\right) = P\left(0 < Z < 0{,}68\right) = 0{,}2517$.

→ **Résolution : probabilité que μ soit supérieur à 53000** .

→ Régionaliser le problème.


51800 53000

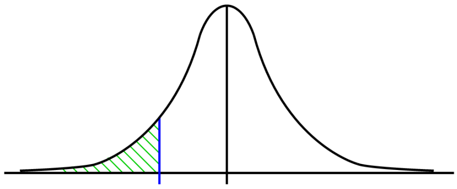
→ $Z = \dfrac{e}{\frac{\sigma}{\sqrt{n}}} = \dfrac{53000 - 51800}{730{,}3} = \dfrac{1200}{730{,}3} = 1{,}64$.

→ Lecture dans la table de loi normale : $P\left(0 < Z < 1{,}64\right) = 0{,}4495$.

→ $P\left(\mu > 53000\right) = 0{,}5 - P\left(E\left(\bar{x}\right) < \mu < 53000\right) = 0{,}5 - P\left(0 < Z < 1{,}64\right) = 0{,}0505$.

→ **Résolution : probabilité que** μ **soit inférieur à** 51000 .

→ Régionaliser le problème.



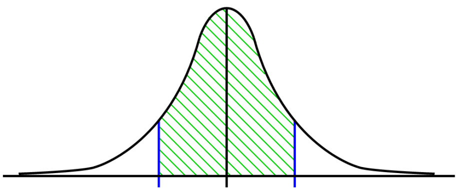
→ $Z = \dfrac{e}{\frac{\sigma}{\sqrt{n}}} = \dfrac{51800 - 51000}{730{,}3} = \dfrac{800}{730{,}3} = 1{,}1$.

→ Lecture dans la table de loi normale : $P\left(0 < Z < 1{,}1\right) = 0{,}3643$.

→ $P\left(\mu < 51000\right) = 0{,}5 - P\left(51000 < \mu < E\left(\bar{x}\right)\right) = 0{,}5 - P\left(0 < Z < 1{,}14\right) = 0{,}1357$.

→ **Résolution : probabilité que** μ **soit compris entre** 51300 **et** 52300 .

→ Régionaliser le problème.



→ $Z = \dfrac{e}{\frac{\sigma}{\sqrt{n}}} = \dfrac{52300 - 51800}{730{,}3} = \dfrac{500}{730{,}3} = 0{,}68$.

→ Lecture dans la table de loi normale : $P\left(0 < Z < 0{,}68\right) = 0{,}2517$.

→ $P\left(51300 < \mu < 52300\right) = 2.P\left(0 < Z < 0{,}68\right) = 0{,}5034$.

III _ Distribution de proportion d'échantillon.

1 _ Conditions d'application.

→ Population de taille N .
→ Échantillon de taille n .

→ Proportion de population P .
→ Proportion d'échantillon $\bar{p}$.

→ **Notion : distribution de proportion moyenne d'échantillon** $\bar{p}_i$.
→ Extraction de plusieurs proportions : $i = 1, \dots, p$ proportions.
→ Pour chaque proportion : $\forall\ i \in [0, p]$.
→ Calcul de la moyenne : $E(\bar{p}_i) = \bar{p}_i$.

→ $\bar{p}$ **suit une loi normale : deux conditions**.
→ **Si** $n * p \geq 5$.
→ **Si** $n(1-p) \geq 5$.

→ **Sinon : nécessaire d'augmenter la taille de** n .

2 _ Moments de la distribution de la moyenne d'échantillon.

a _ Moyenne.

→ $E(\bar{p}_i)$ **: moyenne des proportions moyennes d'échantillons**.
→ Notée : $\bar{p}$.

→ Comme les échantillons sont aléatoires et représentatifs de la population.
→ Alors : $\bar{p}_i \approx p$.

b _ Erreur type.

→ $\sigma(\bar{p})$ **: erreur type d'échantillon**.
→ **Si** N **fini :** $\sigma(\bar{p}) = \sqrt{\frac{p(1-p)}{n}} * \sqrt{\frac{N-n}{N-1}}$.
→ **Si** $N \to +\infty$ **infini :** $\sigma(\bar{p}) = \sqrt{\frac{p(1-p)}{n}} * \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{p(1-p)}{n}}$.
→ Avec $\lim\limits_{N \to +\infty} \sqrt{\frac{N-n}{N-1}} = 1$.

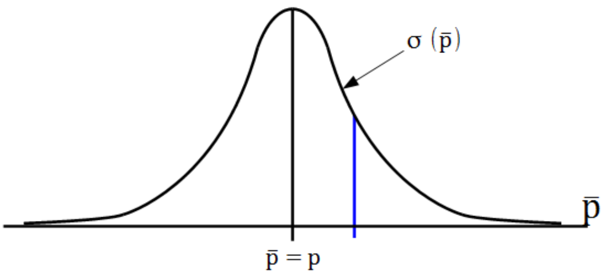
→ **Distribution de** $\bar{p}$ **: pas d'écart type**.

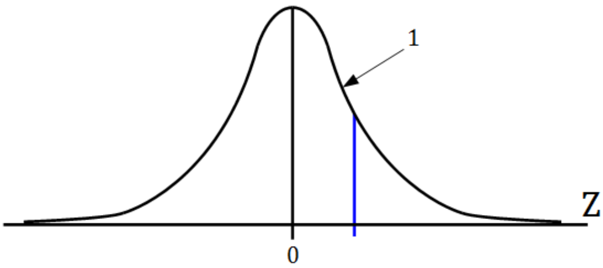
3 _ Variable aléatoire centrée et réduite.

→ **Variable aléatoire centrée réduite de la loi normale** Z .

→ $Z = \dfrac{\bar{p} + e + \bar{p}}{\sigma(\bar{p})}$.

→ **Si** $N \to +\infty$ **:** $Z = \dfrac{e}{\sqrt{\dfrac{p(1-p)}{n}}}$.





4 _ Lecture de la table normale.

→ **Table de loi normale centrée et réduite : lecture**.

→ En ligne : unité et première décimale.

→ En colonne : deuxième décimale.

→ **Méthodologie : cinq étapes**.

→ Énoncer la loi : conditions et paramètres.

→ Conditions : $\bar{p} \to \text{Normale}$ si $\begin{matrix} n * p \geq 5 \\ n(1-p) \geq 5 \end{matrix}$.

→ Paramètres : $\begin{matrix} \bar{p} = p \\ \sigma(p) = \sqrt{\dfrac{p(1-p)}{n}} * \sqrt{\dfrac{N-n}{N-1}} \end{matrix}$

→ Régionaliser le problème : hachurer la zone cherchée.

→ Calcule de $z = \dfrac{\bar{p} + e - \bar{p}}{\sqrt{\dfrac{p(1-p)}{n}}}$

→ Lecture dans la table de loi normale : $P(0 < Z_i < Z)$.

→ Conclusion.