Contents

CHAPITRE 7 - ÉCHANTILLONNAGE ET THÉORÈME LIMITE CENTRAL (TLC)

7.1 Généralités

Un échantillon aléatoire est un sous ensemble représentatif de la population.
Un échantillonnage ou un sondage est une sélection ou un prélèvement d'échantillons dans la population.
La taille de la population (N) c'est l'effectif total des éléments ou unités statistiques de la population.
Un échantillon aléatoire de taille n est un échantillon composé d'une suite de n variables aléatoires $(X_1;\ X_2; \dots;\ X_n)$.

On ne dispose pas souvent de données ou d'informations sur la population qui nous intéresse. On devra alors déduire l'information désirée sur la population en choisissant adéquatement un sous-ensemble représentatif de celle-ci, nommé échantillon, et en analysant les données qu'il contient.

La théorie de l'échantillonnage est l'étude des liaisons existant entre une population et les échantillons de cette population. Elle est fondamentale à deux titres :
* estimer les paramètres de la population (e.g. moyenne, variance, proportion) à partir des estimateurs correspondant de l'échantillon ;
* savoir si les différences observées entre deux échantillons sont dues au hasard ou réellement significatives.

Les méthodes d'échantillonnage c'est l'ensemble des méthodes permettant de réaliser un sondage ou de prélever un échantillon de données au sein d'une population ; de manière à reproduire un échantillon aussi représentatif que possible de cette population.
Il existe différentes méthodes d'échantillonnage pour sélectionner les unités statistiques de la population qui seront incluses dans l'échantillon.

7.2 Les méthodes d'échantillonnage

7.2.1 Les méthodes empiriques :

Les plus utilisées par les instituts de sondage. Leur précision ne peut pas être calculée et leur réussite dépend de l'expertise des enquêteurs.

7.2.1.1 Echantillonnage sur la base du jugement :

Ce sont des échantillons prélevés à partir d'avis d'experts, qui connaissent bien la population et sont capables de dire quelles sont les entités représentatives. L'inconvénient est que l'avis des experts est subjectif.

7.2.1.2 Echantillonnage par la méthode des quotas :

Ce sont des échantillons prélevés librement à condition de respecter une composition donnée à l'avance (sexe, âge, CSP, etc.). L'inconvénient est qu'il repose sur la pertinence des catégories retenues.

7.2.2 Les méthodes aléatoires :

Reposent sur le tirage au hasard d'échantillons et sur le calcul des probabilités.

7.2.2.1 Echantillonnage aléatoire simple :

On prélève dans la population, des individus au hasard, sans remise : tous les individus ont la même probabilité d'être prélevés, et ils le sont indépendamment les uns des autres.

7.2.2.2 Echantillonnage aléatoire stratifié :

7.2.2.2.1 Echantillonnage aléatoire stratifié :

On suppose que la population soit stratifiée, i.e. constituée de sous-populations homogènes, les strates. (ex : stratification par tranche d'âge). Dans chaque strate, on fait un échantillonnage aléatoire simple, de taille proportionnelle à la taille de strate dans la population (échantillon représentatif). Les individus de la population n'ont pas tous la même probabilité d'être tirés. Cette méthode nécessite une homogénéité des strates. Mais elle augmente la précision des estimations.

7.2.2.2.2 Echantillonnage par grappe :

On tire au hasard des grappes ou familles d'individus, et on examine tous les individus de la grappe (ex: on tire des immeubles puis on interroge tous les habitants). La méthode est d'autant meilleure que les grappes se ressemblent et que les individus d'une même grappe sont différents, contrairement aux strates.

Il est à noter que la manière de recueillir les données de l'échantillon est fondamental pour que celui-ci soit représentatif de la population étudiée. A titre d'exemple, un sondage réalisé dans la rue sera le plus souvent biaisé si l'enquêteur a tendance à interroger les personnes qui lui semblent le plus aimable. Pour que l'échantillon soit représentatif, il devra être aléatoire. Un échantillon est qualifié d'aléatoire lorsque chaque élément de la population a une probabilité connue et non nulle d'appartenir à la population.

Dans le cas de ce cours, nous nous référerons exclusivement à un échantillonnage aléatoire simple. Cette méthode consiste à prélever au hasard et de façon indépendante n unités d'échantillonnage d'une population de N éléments. Chaque élément de la population possède ainsi la même probabilité d'appartenir à l'échantillon. Ceci pourra être réalisé en numérotant les éléments de la population de 1 à N puis en utilisant une table des nombres aléatoires.

7.3 Echantillon aléatoire d'une suite de n variables indépendamment et identiquement distribuées

7.3.1 Echantillon aléatoire 1 : $(X_1\ ;\ X_2\ ;\ \dots\ ;\ X_n)$ iid avec $X_i \leadsto (m;\ \sigma)$

- Soit X une variable aléatoire dont la distribution est telle que : $\begin{cases} E(X) = m\ (ou\ \mu) \\ V(X) = \sigma^2 \\ \sigma(X) = \sqrt{V(X)} = \sigma \end{cases}$

On dit que X suit une loi de paramètres $m\ et\ \sigma$; on écrit : $X \leadsto (m;\ \sigma)$.

NB : la loi de X n'est pas spécifiée mais on connaît son espérance mathématique (sa moyenne) et son écart-type.

- Soit un échantillon composé d'une suite de n variables aléatoires $(X_1\ ;\ X_2\ ;\ \dots\ ;\ X_n)$ indépendamment et identiquement distribuées selon la loi de X. on a :

i. $$\underbrace{(X_1\ ;\ X_2\ ;\ \dots\ ;\ X_n)}_{\substack{identiquement\ distribuées\ selon\ la\ loi\ de\ X \\ \Leftrightarrow X_i\ suit\ la\ même\ loi\ que\ X}} \Leftrightarrow \begin{cases} E(X_1) = E(X_2) = \cdots = E(X_n) = E(X) = m \\ V(X_1) = V(X_2) = \cdots = V(X_n) = V(X) = \sigma^2 \\ \sigma(X_1) = \sigma(X_2) = \cdots = \sigma(X_n) = \sqrt{V(X)} = \sigma \end{cases}$$

ii. $$\underbrace{(X_1\ ;\ X_2\ ;\ \dots\ ;\ X_n)}_{mutuellement\ indépendants \Leftrightarrow X_i \cap X_j = \emptyset} \quad alors\ \underbrace{cov(X_i; X_j)}_{pour\ i \neq j} = 0$$

iii. $$\underbrace{(X_1\ ;\ X_2\ ;\ \dots\ ;\ X_n)}_{mutuellement\ indépendants} \quad alors\ \underbrace{V(X_i + X_j)}_{pour\ i \neq j} = V(X_i) + 2\underbrace{cov(X_i; X_j)}_{0} + V(X_j) = V(X_i) + V(X_j) = \sigma^2 + \sigma^2 = 2\sigma^2$$

iv. $$\underbrace{(X_1\ ;\ X_2\ ;\ \dots\ ;\ X_n)}_{mutuellement\ indépendants} \quad alors\ \underbrace{V(X_i - X_j)}_{pour\ i \neq j} = V(X_i) - \underbrace{2cov(X_i; X_j)}_{0} + V(X_j) = V(X_i) + V(X_j) = \sigma^2 + \sigma^2 = 2\sigma^2$$

NB : si $\underbrace{(X_1\,;\,X_2\,;\ldots;\,X_n)}_{\acute{e}chantillon\ ind\acute{e}pendant}$ alors $\begin{cases} V(X_i+X_j)=2\sigma^2 \\ V(X_i-X_j)=2\sigma^2 \end{cases}$

Remarque 1 :

$$si\begin{cases} X \leadsto \underbrace{(m\,;\,\sigma)}_{E(X)=m\ et\ \sigma(X)=\sigma} \\ \quad et \\ \underbrace{(X_1\,;\,X_2\,;\ldots;\,X_n)}_{\substack{\acute{e}chantillon\ ind\acute{e}pendant \\ et\ de\ m\hat{e}me\ loi\ que\ X}} \end{cases} alors \begin{cases} E\left(\sum_{i=1}^{i=n} X_i\right)=\sum_{i=1}^{i=n} E(X_i)=n\times m \\ V\left(\sum_{i=1}^{i=n} X_i\right)=\sum_{i=1}^{i=n} V(X_i)=n\times\sigma^2 \end{cases}$$

Remarque 2 :

$$si\begin{cases} X \leadsto \underbrace{(m\,;\,\sigma)}_{E(X)=m\ et\ \sigma(X)=\sigma} \\ \quad et \\ \underbrace{(X_1\,;\,X_2\,;\ldots;\,X_n)}_{\substack{\acute{e}chantillon\ ind\acute{e}pendant \\ et\ m\hat{e}me\ loi\ que\ X}} \end{cases} alors \begin{cases} E\left(\frac{\sum_{i=1}^{i=n} X_i}{n}\right)=\left(\frac{1}{n}\right)\sum_{i=1}^{i=n} E(X_i)=\left(\frac{1}{n}\right)nm=m \\ V\left(\frac{\sum_{i=1}^{i=n} X_i}{n}\right)=\left(\frac{1}{n^2}\right)\sum_{i=1}^{i=n} V(X_i)=\left(\frac{1}{n^2}\right)n\sigma^2=\frac{\sigma^2}{n} \end{cases}$$

Remarque 3 :

$$si\begin{cases} X \leadsto \underbrace{(m\,;\,\sigma)}_{E(X)=m\ et\ \sigma(X)=\sigma} \\ \quad et \\ \underbrace{(X_1\,;\,X_2\,;\ldots;\,X_n)}_{\substack{\acute{e}chantillon\ ind\acute{e}pendant \\ et\ m\hat{e}me\ loi\ que\ X}} \\ \quad et \\ \bar{X}=\frac{\sum_{i=1}^{i=n} X_i}{n} \end{cases} alors \begin{cases} E(\bar{X})=E\left(\frac{\sum_{i=1}^{i=n} X_i}{n}\right)=m \\ V(\bar{X})=V\left(\frac{\sum_{i=1}^{i=n} X_i}{n}\right)=\frac{\sigma^2}{n} \\ \sigma(\bar{X})=\sqrt{\frac{\sigma^2}{n}}=\frac{\sigma}{\sqrt{n}} \end{cases}$$

Remarque : $\bar{X} \leadsto \left(m;\frac{\sigma}{\sqrt{n}}\right),\ avec\ \ E(\bar{X})=m\ et\ \sigma(\bar{X})=\frac{\sigma}{\sqrt{n}}$

7.3.2 Echantillon aléatoire 2 : $(X_1\,;\,X_2\,;\ldots;\,X_n)iid$ avec $X_i \leadsto \underbrace{\mathcal{B}(1\,;\,p)}_{loi\ de\ Bernoulli}$

- Soit X une variable aléatoire qui suit une distribution de Bernoulli de paramètres 1 et p :

$$X \leadsto \underbrace{\mathcal{B}(1\,;\,p)}_{\substack{loi\ de\ Bernoulli \\ p=probabilit\acute{e}\ de\ succ\grave{e}s}} \quad avec \begin{cases} E(X)=p\ (ou\ \pi) \\ V(X)=pq=p(1-p) \\ \sigma(X)=\sqrt{V(X)}=\sqrt{pq}=\sqrt{p(1-p)} \end{cases}$$

- Soit un échantillon composé d'une suite de n variables aléatoires $(X_1\,;\,X_2\,;\ldots;\,X_n)$ indépendamment et identiquement distribuées selon la loi de X. on a :

$$\underbrace{(X_1\ ;\ X_2\ ;\dots\ ;\ X_n)}_{\substack{identiquement\ distribuées\\ selon\ la\ loi\ de\ X\\ \Leftrightarrow X_i\ suit\ la\ même\ loi\ que\ X}} \Leftrightarrow \begin{cases} E(X_1)=E(X_2)=\cdots=E(X_n)=E(X)=p \\ V(X_1)=V(X_2)=\cdots=V(X_n)=V(X)=pq=p(1-p) \\ \sigma(X_1)=\sigma(X_2)=\cdots=\sigma(X_n)=\sqrt{V(X)}=\sqrt{pq}=\sqrt{p(1-p)} \end{cases}$$

i. $$\underbrace{(X_1\ ;\ X_2\ ;\dots\ ;\ X_n)}_{mutuellement\ indépendants \Leftrightarrow X_i \cap X_j = \emptyset} \quad alors\ \underbrace{cov(X_i;X_j)}_{pour\ i\neq j}=0$$

ii. $$\underbrace{(X_1\ ;\ X_2\ ;\dots\ ;\ X_n)}_{mutuellement\ indépendants} \quad alors\ \underbrace{V(X_i+X_j)}_{pour\ i\neq j}=V(X_i)+2\underbrace{cov(X_i;X_j)}_{0}+V(X_j)=V(X_i)+V(X_j)$$

$$d'où\ V(X_i+X_j)=pq+pq=2pq=2p(1-p)$$

iii. $$\underbrace{(X_1\ ;\ X_2\ ;\dots\ ;\ X_n)}_{mutuellement\ indépendants} \quad alors\ \underbrace{V(X_i-X_j)}_{pour\ i\neq j}=V(X_i)-2\underbrace{cov(X_i;X_j)}_{0}+V(X_j)=V(X_i)+V(X_j)$$

$$d'où\ V(X_i-X_j)=pq+pq=2pq=2p(1-p)$$

NB : si $\underbrace{(X_1\ ;\ X_2\ ;\dots\ ;\ X_n)}_{échantillon\ indépendant}$ alors $\begin{cases} V(X_i+X_j)=2pq=2p(1-p) \\ V(X_i-X_j)=2pq=2p(1-p) \end{cases}$

Remarque 1 :

$$si \begin{cases} X \backsim \underbrace{\mathcal{B}(1\ ;\ p)}_{\substack{loi\ de\ Bernoulli\\ p=probabilité\ de\ succès\\ E(X)=p\ et\ V(X)=pq=p(1-p)}} \\ et \\ \underbrace{(X_1\ ;\ X_2\ ;\dots\ ;\ X_n)}_{\substack{échantillon\ indépendant\\ de\ même\ loi\ que\ X}} \end{cases} alors \begin{cases} E\left(\sum_{i=1}^{i=n} X_i\right)=\sum_{i=1}^{i=n} E(X_i)=np \\ V\left(\sum_{i=1}^{i=n} X_i\right)=\sum_{i=1}^{i=n} V(X_i)=n(pq)=np(1-p) \end{cases}$$

Remarque 2 :

$$si \begin{cases} X \backsim \underbrace{\mathcal{B}(1\ ;\ p)}_{loi\ de\ Bernoulli} \\ et \\ \underbrace{(X_1\ ;\ X_2\ ;\dots\ ;\ X_n)}_{\substack{échantillon\ indépendant\\ et\ même\ loi\ que\ X}} \end{cases} alors \begin{cases} E\left(\frac{\sum_{i=1}^{i=n} X_i}{n}\right)=\left(\frac{1}{n}\right)\sum_{i=1}^{i=n} E(X_i)=\left(\frac{1}{n}\right)np=p \\ \underbrace{V\left(\frac{\sum_{i=1}^{i=n} X_i}{n}\right)}_{V\left(\frac{1}{a}X\right)=\frac{1}{a^2}V(X)}=\left(\frac{1}{n^2}\right)\sum_{i=1}^{i=n} V(X_i)=\left(\frac{1}{n^2}\right)n(pq)=\frac{pq}{n}=\frac{p(1-p)}{n} \end{cases}$$

Remarque 3 :

$$\text{si}\begin{cases} X \sim \underbrace{\mathcal{B}(1\,;\,p)}_{\text{loi de Bernoulli}} \\ \text{et} \\ \underbrace{(X_1\,;\,X_2\,;\,\dots\,;\,X_n)}_{\substack{\text{échantillon indépendant} \\ \text{et même loi que } X}} \\ \text{et} \\ K = \sum_{i=1}^{i=n} X_i \\ \text{et} \\ F = \dfrac{K}{n} \end{cases} \quad \text{alors}\begin{cases} E(F) = E\left(\dfrac{K}{n}\right) = E\left(\dfrac{\sum_{i=1}^{i=n} X_i}{n}\right) = p \\ V(F) = V\left(\dfrac{K}{n}\right) = V\left(\dfrac{\sum_{i=1}^{i=n} X_i}{n}\right) = \dfrac{pq}{n} = \dfrac{p(1-p)}{n} \\ \sigma(F) = \sqrt{\dfrac{p(1-p)}{n}} \end{cases}$$

Remarque : $F \sim \left(p;\ \sqrt{\frac{p(1-p)}{n}}\right),\ avec\ \ E(F) = p\ et\ \sigma(F) = \sqrt{\frac{p(1-p)}{n}}$

7.3.3 Exemples de variables d'échantillon (ou estimateurs)

7.3.3.1 Fréquence empirique ou fréquence d'échantillon : F

La fréquence d'échantillon ou fréquence empirique est la variable aléatoire F telle que :

$$F = \frac{K}{n}\ avec \begin{cases} \text{n=taille de l'échantillon} \\ \text{K=nombre de succès dans l'échantillon} \\ \text{F=fréquence de succès dans l'échantillon} \end{cases}$$

$$\underbrace{\hat{p}}_{\substack{\text{proportion de succès} \\ \text{dans la population}}} = \underbrace{F}_{\substack{\text{fréquence de succès} \\ \text{ou} \\ \text{proportion de succès} \\ \text{dans l'échantillon}}} = \frac{\overbrace{K}^{\substack{\text{nombre de succès} \\ \text{dans l'échantiilon}}}}{\underbrace{n}_{\text{taille de l'échantillon}}}$$

But : La fréquence d'échantillon est un estimateur ponctuel de la proportion (p) inconnue de la population statistique. On écrit : $\hat{p} = F = \frac{K}{n}$

Loi de distribution de la fréquence d'échantillon : F suit une distribution de paramètres p et $\sqrt{\frac{p(1-p)}{n}}$, on note :

$$F \sim \left(p\,;\sqrt{\frac{p(1-p)}{n}}\right)\ avec\ E(F) = p\ et\ \sigma(F) = \sqrt{\frac{p(1-p)}{n}}$$

Application 1 – Dans un échantillon aléatoire de 200 composantes électriques fabriquées avec un certain procédé, 15 composantes se sont avérées défectueuses. Donner une estimation ponctuelle de la proportion de pièces défectueuses produites.

Réponse : $\hat{p} = F = \frac{K}{n}$; ici le succès est « composante défectueuse »

D'où : $\underbrace{f}_{\substack{fréquence\ de\ composantes \\ défectueuses \\ dans\ l'échantillon}} = \dfrac{\overbrace{k}^{\substack{nombre\ de\ composantes\ défectueuses \\ dans\ l'échantiilon}}}{\underbrace{n}_{taille\ de\ l'échantillon}} = \frac{15}{200} = 0{,}075\ soit\ 7{,}5\%$

$\hat{p} = f = \frac{k}{n} = \frac{15}{200} = 0{,}075$ soit 7,5% de pièces défectueuses dans l'échantillon.

NB : l'estimateur est une variable en lettres capitales et l'estimation est une valeur en lettres minuscules.

Paramètres de la population	Estimateur (en lettres capitales)	Estimation (en lettres minuscules)
Proportion inconnue (p)	$\hat{p} = F = \frac{K}{n}$	$\hat{p} = f = \frac{k}{n}$
Moyenne inconnue (m)	$\hat{m} = \bar{X} = \frac{\sum_{i=1}^{i=n} X_i}{n}$	$\hat{m} = \bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n}$

Application 2 - le gouvernement d'un pays démocratique souhaite connaître la proportion d'électeurs en accord avec un nouveau texte de loi. Pour cela il organise un sondage sur un échantillon aléatoire. Sur 100 personnes interrogées, 80 pensent voter oui. Donner une estimation ponctuelle de la proportion de personnes favorables au projet de loi.

Réponse : succès = personne désireuse de voter « oui » ; n = 100 ; k = 80.

Ainsi : $\hat{p} = f = \frac{k}{n} = \frac{80}{100} = 0{,}8$ soit 80% de personnes favorables au projet de loi.

Application 3 - On veut estimer par intervalle de confiance la proportion de personnes qui sont satisfaites du gouvernement au niveau économique. On choisit un échantillon aléatoire de 200 personnes et on obtient une proportion de 64%. Donner une estimation ponctuelle de la proportion de personnes satisfaites dans la population.

Réponse : Succès = personne non satisfaite ; f = 64% = 0,64.

$n = 200$; $\hat{p} = f = 0{,}64$ Soit 64% de personnes satisfaites.

7.3.3.2 Moyenne empirique ou moyenne d'échantillon : $\bar{X}$

La moyenne d'échantillon ou moyenne empirique est la variable $\bar{X}$ telle que :

$$\bar{X} = \frac{\sum_{i=1}^{i=n} X_i}{n}$$

La moyenne d'échantillon $\bar{X}$ suit une distribution non précisée de paramètres m et $\frac{\sigma}{\sqrt{n}}$, on note :

$$\underbrace{(X_1, X_2, \dots, X_n)}_{échantillon}\ iid\ et\ X_i \backsim (m\,;\,\sigma) \Rightarrow \bar{X} \backsim \left(m\,;\frac{\sigma}{\sqrt{n}}\right), avec \begin{cases} E(\bar{X}) = m \\ V(\bar{X}) = \dfrac{\sigma^2}{n} \\ \sigma(\bar{X}) = \dfrac{\sigma}{\sqrt{n}} \end{cases}$$

But : La moyenne d'échantillon est un estimateur ponctuel de la moyenne (m) inconnue de la population statistique. On écrit : $\hat{m} = \bar{X} = \frac{\sum_{i=1}^{i=n} X_i}{n}$

7.3.3.3 Variance empirique ou variance d'échantillon : S^2

La variance d'échantillon ou variance empirique est la variable aléatoire S^2 telle que :

$$S^2 = \frac{\sum_{i=1}^{i=n}(X_i - \bar{X})^2}{n} = \left[\frac{\sum X_i^2}{n}\right] - (\bar{X})^2$$

$$avec \quad E(S^2) = \frac{n-1}{n}\sigma^2$$

7.3.3.4 Variance empirique corrigée ou variance corrigée d'échantillon : S'^2

La variance empirique corrigée ou variance corrigée d'échantillon est la variable aléatoire S'^2 telle que :

$$S'^2 = \frac{\sum_{i=1}^{i=n}(X_i - \bar{X})^2}{n-1} = \left[\frac{\sum X_i^2}{n-1}\right] - \left(\frac{n}{n-1}\right)(\bar{X})^2$$

$$E(S'^2) = \sigma^2$$

7.3.3.5 Relation S^2 et S'^2

$$S^2 = \frac{\sum_{i=1}^{i=n}(X_i - \bar{X})^2}{n} \Rightarrow nS^2 = \sum_{i=1}^{i=n}(X_i - \bar{X})^2 \qquad (1)$$

$$S'^2 = \frac{\sum_{i=1}^{i=n}(X_i - \bar{X})^2}{n-1} \Rightarrow (n-1)S'^2 = \sum_{i=1}^{i=n}(X_i - \bar{X})^2 \qquad (2)$$

$$(1)\ et\ (2)\ donnent\ \ nS^2 = (n-1)S'^2 \qquad (3)$$

$$de\ (3)\ on\ tire \qquad S^2 = \frac{n-1}{n}S'^2 \qquad (4)$$

$$de\ (3)\ on\ tire \qquad S'^2 = \frac{n}{n-1}S^2 \qquad (5)$$

7.3.3.6 écart-type d'échantillon : S

$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}} = \sqrt{\left(\frac{\sum X_i^2}{n}\right) - \bar{X}^2}$$

7.3.3.7 Ecart-type corrigé d'échantillon : S'

$$S' = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} = \sqrt{\left(\frac{\sum X_i^2}{n-1}\right) - \left(\frac{n}{n-1}\right)\bar{X}^2}$$

7.3.3.8 relation S et S'

$$\sqrt{n}S = \sqrt{(n-1)}S' \qquad (1)$$

$$ou\ S = \sqrt{\frac{n}{n-1}}S' \qquad (2)$$

$$ou \;\; S' = \sqrt{\frac{n}{n-1}}S \qquad (3)$$

7.4 Les paramètres de la population et les estimateurs (ou statistiques) d'échantillon

Les paramètres de la population, notés θ, sont des données ou des informations sur la population ; ces paramètres sont souvent inconnus.

Exemples de paramètres usuels de la population (θ) :

i. Moyenne de la population : $\theta = m\ (ou\ \mu)$
ii. Variance de la population : $\theta = \sigma^2$
iii. Proportion de la population : $\theta = p\ (ou\ \pi)$

Un estimateur $\hat{\theta}$ du paramètre θ de la population statistique est une variable aléatoire qui est fonction d'une suite de n variables aléatoires représentant l'échantillon :

$$\hat{\theta}_n = f(X_1;\ X_2; \dots;\ X_n)$$

Une statistique est un estimateur. Une statistique est une variable aléatoire fonction de l'échantillon de variables aléatoires $(X_1;\ X_2; \dots;\ X_n)$

7.5. Tableau des paramètres θ usuels : leurs estimateurs $\hat{\theta}$ usuels et leurs estimations ponctuelles

	$\boldsymbol{\theta}$ Paramètre de la population	$\hat{\boldsymbol{\theta}}$ Estimateur de l'échantillon Les estimateurs sont des variables écrites en lettres capitales : $\boldsymbol{\bar{X}}$, $\boldsymbol{S^2}$, $\boldsymbol{S'^2}$, $\boldsymbol{F}$	Estimation ponctuelle (ou valeur de l'estimateur) Les estimations sont écrites en lettres minuscules : $\boldsymbol{\bar{x}}, \boldsymbol{s^2}, \boldsymbol{s'^2}, \boldsymbol{f}, \boldsymbol{s}; \boldsymbol{s'}$
Taille	N	n	
Proportion	p (ou π) Proportion de succès dans une population	$\boldsymbol{\hat{p} = F = \frac{K}{n}}$ Fréquence d'échantillon	$\boldsymbol{\hat{p}} = \mathrm{f} = \frac{\mathrm{k}}{\mathrm{n}}$ Estimation ponctuelle de la proportion p
Moyenne	m (ou μ) Moyenne de la population	$\boldsymbol{\hat{m} = \bar{X} = \frac{\sum X_i}{n}}$ Moyenne d'échantillon	$\boldsymbol{\hat{m}} = \bar{x} = \frac{\sum x_i}{n}$ Estimation ponctuelle de la moyenne m de la population
Variance	Variance de la population : σ^2	Variance d'échantillon $\widehat{\sigma^2} = S^2 = \frac{\sum (X_i - \bar{X})^2}{n}$ Si $n \geq 30$	Calcul de la variance d'échantillon : s^2 $s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ Calcul Ecart-type d'échantillon : s $\hat{\sigma} = s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$
		Variance corrigée d'échantillon $\widehat{\sigma^2} = S'^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$ Si $n < 30$	Estimation de σ^2: $\widehat{\sigma^2} = s'^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ Estimation de σ: $\hat{\sigma} = s' = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$
Ecart type σ		Relation S^2 et S'^2 $S^2 = \frac{n-1}{n} S'^2$ et $S'^2 = \frac{n}{n-1} S^2$	Relation s et s' $\sqrt{n}s = \sqrt{(n-1)}s'$ et $s = s'\sqrt{\frac{n-1}{n}}$ et $s' = s\sqrt{\frac{n}{n-1}}$

Remarque : s' s'écrit aussi :

$$s' = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\left(\frac{\sum x_i^2}{n-1}\right) - \left(\frac{n}{n-1}\right)\bar{x}^2}$$

7.6 - QUALITÉS D'UN BON ESTIMATEUR : $\hat{\theta}$

Voici les principales qualités que devrait avoir un estimateur :

7.6.1 Absence de biais:

O6 dit qu'un estimateur $\hat{\theta}$ est **non biaisé** si sa valeur espérée $E(\hat{\theta})$ est égale au paramètre à estimer.

$\boldsymbol{\hat{\theta}}$ *est sans biais* si $\boldsymbol{E(\hat{\theta}) = \theta}$

$$\boldsymbol{\hat{\theta}}\ est\ asymptotiquement\ sans\ biais\ si\ \lim_{n \to +\infty} \boldsymbol{E(\hat{\theta}) = \theta}$$

7.6.2 Convergence :

On dit qu'un estimateur non biaisé $\hat{\theta}$ est convergent si sa distribution tend à se concentrer de plus en plus autour du paramètre à estimer, lorsque la taille de l'échantillon augmente.

$$\boldsymbol{\hat{\theta}}\ est\ convergent\ \ si\ \lim_{n \to \infty} V(\hat{\theta}) = \mathbf{0}$$

7.6.3 Efficacité :

Dans la classe de tous les estimateurs non biaisés, le plus efficace est celui qui possède la plus petite variance.

Soient : $\boldsymbol{\widehat{\theta_1}}$ *et* $\boldsymbol{\widehat{\theta_2}}$ si $V(\boldsymbol{\widehat{\theta_2}}) \leq V(\boldsymbol{\widehat{\theta_1}})$ alors $\boldsymbol{\widehat{\theta_2}}$ **est l'estimateur le plus efficace des deux estimateurs** de $\boldsymbol{\theta}$.

Si $\boldsymbol{\hat{\theta}}$ est l'estimateur obtenu par la méthode du maximum de vraisemblance (EMV), $\boldsymbol{\hat{\theta}}$ est efficace si

$$\boldsymbol{I_n(\hat{\theta}) = -E\left[\!\left[\frac{\partial^2 Ln(L)}{\partial \theta^2}\right]\!\right]_{\theta = \hat{\theta}} = \frac{1}{V(\hat{\theta})}}$$

Où $I_n(\hat{\theta}) = -E\left[\!\left[\frac{\partial^2 Ln(L)}{\partial \theta^2}\right]\!\right]_{\theta=\hat{\theta}}$ est appelé la quantité d'information de Fisher.

7.7- Exemple d'estimateurs non biaisés (sans biais) et convergents : $\boldsymbol{\bar{X}}$, $\boldsymbol{F}$ et $\boldsymbol{S'^2}$

7.7.1 Montrons que F est un estimateur non biaisé et convergent de la proportion p d'une population

- ✓ $F\ est\ un\ estimateur\ sans\ biais\ de\ p\ si\ E(F) = p$

$\hat{p} = F\ \ et\ E(\hat{p}) = \underbrace{E(F) = p}_{déjà\ démontré} \Rightarrow E(F) = p$ donc F est un estimateur sans biais de p.

- ✓ F est un estimateur convergent de p si $\lim_{n\to\infty} V(F) = 0$

$\hat{p} = F\ \ et\ V(\hat{p}) = \underbrace{V(F) = \frac{p(1-p)}{n}}_{déjà\ démontré}\ \ alors\ \lim_{n\to\infty} V(F) = \lim_{n\to\infty} \frac{p(1-p)}{n} = 0;$

F est donc un estimateur convergent de p.

7.7.2 Montrons que $\hat{m} = \bar{X} = \frac{\sum_{i=1}^{i=n} X_i}{n}$ est un estimateur non biaisé et convergent de la moyenne m d'une population

- ✓ $\bar{X}$ *est un estimateur sans biais du paramètre m, si* $E(\bar{X}) = m$

$\hat{m} = \bar{X}$ *et* $E(\hat{m}) = \underbrace{E(\bar{X}) = m}_{\text{déjà démontré}}$ donc $\bar{X}$ est un estimateur sans biais de m.

- ✓ $\bar{X}$ est un estimateur convergent de m si $\lim_{n\to\infty} V(\bar{X}) = 0$

$\hat{m} = \bar{X}$ *et* $V(\hat{m}) = \underbrace{V(\bar{X}) = \frac{\sigma^2}{n}}_{\text{déjà démontré}}$ et $\lim_{n\to\infty} V(\bar{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$; $\bar{X}$ est donc un estimateur convergent de m.

7.7.3 S^2 est un estimateur biaisé de σ^2

- ✓ S^2 est un estimateur biaisé de σ^2

$\widehat{\sigma^2} = S^2$ *et* $E(\widehat{\sigma^2}) = E(S^2)$ *soit* $E\left(\frac{\sum(X_i-\bar{X})^2}{n}\right) = \frac{n-1}{n}\sigma^2 \Rightarrow S'^2$ est un estimateur biaisé de σ^2.

Le biais est : $E(\widehat{\sigma^2}) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \left(\sigma^2 - \frac{\sigma^2}{n}\right) - \sigma^2 = -\frac{\sigma^2}{n}$

7.7.4 S'^2 est un estimateur non biaisé de σ^2

- ✓ S'^2 est un estimateur non biaisé de σ^2

$$\widehat{\sigma^2} = S'^2 \text{ et } E(\widehat{\sigma^2}) = E(S'^2) = E\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}E(S^2) = \frac{n}{n-1}\left(\frac{n-1}{n}\sigma^2\right) = \sigma^2$$

S'^2 est donc un estimateur sans biais de σ^2.

7.7.5 L'estimateur sans biais de σ^2 *est* $S'^2 = \frac{\sum(X_i-\bar{X})^2}{n-1}$.

7.8 Exercices d'application : Estimation ponctuelle de la moyenne (m) et l'écart-type (σ) d'une population

Exercice 1 :

Une entreprise fabrique un certain type de composants électroniques dont la durée de vie X, exprimée en heures, est une variable aléatoire. Des mesures effectuées sur un échantillon aléatoire de taille 50 ont donné les résultats suivants :

$\sum_{i=1}^{50} x_i = 60\,000$; $\sum_{i=1}^{50} x_i^2 = 74.10^6$

a). Donner une estimation ponctuelle de la durée de vie moyenne des composants.

b). Donner une estimation ponctuelle de l'écart-type de cette durée de vie.

Corrigé exercice 1

1. Identifier la variable aléatoire X (unités) et sa loi si possible.
2. Sur le brouillon, repérer les paramètres de la population et les estimations d'échantillon donnés par l'exercice.

Population : paramètres	Echantillon : estimation
$N =$	$n =$
$p =$	$\hat{p} = \mathrm{f} = \frac{\mathrm{k}}{\mathrm{n}} =$
$m =$	$\hat{m} = \bar{x} =$

$\sigma^2 =$	$\widehat{\sigma^2} = s'^2 =$
$\sigma =$	$\hat{\sigma} = s' =$

3. Puis procéder aux calculs exigés pour répondre aux questions.

X = la durée de vie des composants en heures.

a) Estimation ponctuelle de m : n = 50

$\hat{m} = \bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n} = \frac{\sum_{i=1}^{50} x_i}{n} = \frac{60000}{50} = 1200\ heures$ la durée de vie moyennes s'élève à 1200 heures.

b) Estimation ponctuelle de l'écart-type σ : $\hat{\sigma}$= s'=$\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}\ ou\ s' = \sqrt{\frac{n}{n-1}}s\ et\ n = 50$

s'=$\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}} = \sqrt{\left(\frac{\sum x_i^2}{n-1}\right) - \left(\frac{n}{n-1}\right)(\bar{x})^2} = \sqrt{\left(\frac{74000000}{50-1}\right) - \left(\frac{50}{50-1} \times 1200^2\right)} = 202{,}0305\ heures$

l'écart-type de la durée de vie moyenne est de 202,03 heures.

Autre méthode : $s' = \sqrt{\frac{n}{n-1}}s \quad avec\ s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n}} = \sqrt{\left(\frac{\sum x_i^2}{n}\right) - \bar{x}^2}$

- ✓ Variance d'échantillon

$s^2 = \left[\frac{\sum_{i=1}^{50} x_i^2}{n}\right] - \bar{x}^2 = \left[\frac{\sum_{i=1}^{50} x_i^2}{50}\right] - \bar{x}^2 = \left[\frac{74.000.000}{50}\right] - 1200^2 = 40.000$

- ✓ Ecart-type d'échantillon s : s=$\sqrt{40000} = 200$
- ✓ Estimation ponctuelle de l'écart-type σ : $\hat{\sigma}$= s'

$\hat{\sigma}$= s'=$\sqrt{\frac{n}{n-1}}s = \sqrt{\frac{50}{49}} \times 200 = 202{,}0305\ heures$

Exercice 2. On prélève un échantillon de 50 copies d'examen ; et on trouve une note moyenne d'échantillon de 12,3.

Sachant que $\sum_{i=1}^{50}(x_i - \bar{x})^2 = 384$ On suppose que les notes sont normalement distribuées

a). Donner une estimation ponctuelle de la durée de vie moyenne des composants.

b). Donner une estimation ponctuelle de l'écart-type de cette durée de vie.

Indication de corrigé exercice 2

Données : $n = 50; \sigma = inconnu;\ \bar{x} = 12{,}3;$

X : notes; σ inconnu et n = 50 $\Rightarrow\ \hat{\sigma} = s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{384}{50-1}} =$

Exercice 3 - Dans cette partie, on suppose que m et σ sont inconnus.

On a relevé dans le tableau suivant les résultats de 10 pesées d'un même objet :

Masse (en g)	72,20	72,24	72,26	72,30	72,36	72,39	72,42	72,48	72,50	72,54

Les résultats seront arrondis au centième le plus proche.

1°) Calculer la moyenne et l'écart –type de cet échantillon

2°) En déduire les estimations ponctuelles de la moyenne m et de l'écart-type σ de la population.

Réponse - Exercice 3

1°) Calculer la moyenne et l'écart –type de cet échantillon

X : variable aléatoire, X = masse (ou poids) en g ; et x_i: les différentes valeurs de X.

Moyenne de l'échantillon : $\bar{x}$

$$\bar{x} = \frac{\sum x_i}{n} = \underbrace{\frac{723{,}6907}{10}}_{voir\ le\ tableau\ de\ calcul} = 72{,}36907$$

Variance de l'échantillon : s^2

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \left[\frac{\sum x_i^2}{n}\right] - \bar{x}^2$$

$$s^2 = \left[\frac{52372{,}845}{10}\right] - 72{,}36907 = 0{,}01237$$

Écart –type d'échantillon :

$$s = \sqrt{0{,}01237} = 0{,}111$$

Tableau de calcul

	x_i	x_i^2
	72,200	5212,840
	72,240	5218,618
	72,260	5221,508
	72,300	5227,290
	72,360	5235,970
	72,390	5240,312
	72,420	5244,656
	72,480	5253,350
	72,500	5256,250
	72,540	5262,052
Total	$\sum x_i = 723,6907$	$\sum x_i^2 = 52372{,}845$

2°) En déduire les estimations ponctuelles de la moyenne m et de l'écart-type σ de la population.

Estimation ponctuelle de la moyenne d'échantillon m: $\hat{m} = \bar{x}$	Estimation ponctuelle de l'écart-type de la population σ: $\hat{\sigma} = s'$
$\hat{m} = \bar{x} = 72{,}36907$	$s' = \sqrt{\frac{n}{n-1}} \times s = \sqrt{\frac{10}{9}} \times 0{,}111 = 0{,}117$

Exercice 4 - On a contrôlé le dosage d'un produit dans un mélange à la sortie d'une chaîne de conditionnement. Pour un échantillon de 100 lots de 5 kilogrammes de mélange analysés, on a obtenu les résultats suivants, où x_i représente la masse du produit exprimée en grammes et n_i l'effectif correspondant :

x_i	142	144	146	148	150	152	154	156	158	160
n_i	1	5	6	21	32	22	7	4	1	1

1. Calculer la moyenne et l'écart-type de l'échantillon.
2. Donner une estimation ponctuelle de la moyenne et de l'écart-type.

Corrigé exercice 4

1. Calculer la moyenne et l'écart-type de l'échantillon.

X = masse du produit en g

Formules et tableau de calcul

La moyenne d'échantillon : $\bar{x}$

$$\bar{x} = \frac{\sum n_i x_i}{n} = \frac{15010}{100} = 150{,}10\ g$$

La variance

$$s^2 = \left[\frac{\sum n_i x_i^2}{n}\right] - (\bar{X})^2$$

Tableau de calcul

x_i	n_i	$n_i x_i$	$n_i x_i^2$
142	1	142	20164
144	5	720	103680
146	6	876	127896
148	21	3108	459984
150	32	4800	720000
152	22	3344	508288

$$s^2 = \frac{2253932}{100} - (150{,}1)^2 = 9{,}31$$

L'écart-type : s

$$s = \sqrt{9{,}31} = 3{,}051\ g$$

154	7	1078	166012
156	4	624	97344
158	1	158	24964
160	1	160	25600
Total	100	15010	2253932

2. Donner une estimation ponctuelle de la moyenne et de l'écart-type de la population.

Estimation ponctuelle de la moyenne de la population : $\widehat{m} = \bar{x} = 150{,}10\ g$

Estimation ponctuelle de l'écart-type de la population : $\hat{\sigma} = s'$

$$\hat{\sigma} = s' = \sqrt{\frac{n}{n-1}} \times s = \sqrt{\frac{100}{99}} \times 3{,}051 = 3{,}066\ g$$

7.9 Théorème central limite

7.9.1 Théorème central limite $(n > 30)$ et loi de distribution de F

Le théorème central limite (TCL) énonce que pour une taille d'échantillon n très grand, **en pratique pour $\boldsymbol{n > 30}$**, on a :

a. Approximation de la loi binomiale par la loi normale

$$si\ K \backsim \underbrace{\mathcal{B}(n;p)}_{\substack{loi\\ binomiale}}\ et\ \boldsymbol{n > 30} \Rightarrow K \backsim N\left(np\,;\sqrt{np(1-p)}\right) \Rightarrow U = \frac{K - np}{\sqrt{np(1-p)}} \backsim N(0\,;1)$$

b. Loi de distribution de F

$$F = \frac{K}{n}\ et\ \boldsymbol{n > 30} \Rightarrow F \backsim N\left(p\,;\sqrt{\frac{p(1-p)}{n}}\right) \Longrightarrow U = \frac{F - p}{\sqrt{\frac{p(1-p)}{n}}} \backsim N(0\,;1)$$

7.9.2 Théorème central limite $(n > 30)$ et loi de distribution de $\bar{X}$

Le théorème central limite (TCL) énonce que pour une taille d'échantillon n très grand, **en pratique pour $\boldsymbol{n > 30}$**, on a :

a. **Loi de distribution de** $\bar{X}$ si σ connu

$$\text{si } n > 30\ et\ X_i \backsim N(m\,;\,\sigma) \text{ et } \sigma \text{ connu} \Rightarrow \bar{X} \backsim N\left(m\,;\frac{\sigma}{\sqrt{n}}\right) \Longrightarrow U = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \backsim N(0\,;1)$$

b. Loi de distribution de $\bar{X}$ si σ inconnu

$$si\ n > 30\ et\ X_i \backsim N(m\,;\,\sigma) \text{ et } \sigma \text{ inconnu: } \hat{\sigma}=s' \Rightarrow \bar{X} \backsim N\left(m\,;\frac{s'}{\sqrt{n}}\right) \Longrightarrow U = \frac{\bar{X} - m}{\frac{s'}{\sqrt{n}}} \backsim N(0\,;1)$$

7.9.3 loi de distribution de $\bar{X}$ lorsque $n \leq 30$

a. Loi de distribution de $\bar{X}$ si n≤30 et σ connu

$$\text{si } n\leq 30 \text{ et } X_i \sim \mathcal{N}(m;\sigma) \text{ et } \sigma \text{ connu} \Rightarrow \bar{X} \sim N\left(m\,;\frac{\sigma}{\sqrt{n}}\right) \Longrightarrow U = \frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} \sim N(0\,;1)$$

b. Loi de distribution de $\bar{X}$ si **n≤30 et** σ inconnu et σ̂=s'

$$\text{si } n\leq 30 \text{ et } X_i \sim \mathcal{N}(m;\sigma) \text{ et } \sigma \text{ inconnu et } \hat{\sigma}=s' \Rightarrow \bar{X} \sim N\left(m\,;\frac{s'}{\sqrt{n}}\right) \Longrightarrow T = \frac{\bar{X}-m}{\frac{s'}{\sqrt{n}}} \sim T_{n-1}$$

T_{n-1}: *loi de Student* à $(n-1)$ *dégrés de libertés*

Si Z suit une loi de Student à n degrés de libertés : T_n, on écrit :

$Z \sim T_n \;\; avec\; E(T) = 0\; et\; V(T) = \frac{n}{n-2}$ et $n > 2$

Remarque :

a) L'espérance n'existe que pour n > 1. Lorsque l'espérance existe, elle est nulle, puisque la loi est symétrique autour de 0.
b) La variance n'existe que pour n > 2. La variance est toujours supérieure à 1 et tend vers 1 lorsque n tend vers l'infini

7.9.4 loi de distribution de S'^2

Soit la variable W :

$$W = \frac{(n-1)S'^2}{\sigma^2}$$

Il est démontré que la variable aléatoire $W = \frac{(n-1)S'^2}{\sigma^2}$ suit la loi de Khi 2 à (n-1) degrés de liberté (ddl). Et on écrit :

$$W = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$avec \begin{cases} E(W) = n-1 \\ V(W) = 2(n-1) \\ \sigma(W) = \sqrt{2(n-1)} \end{cases}$$

Conclusion

$$Si \begin{cases} Y_i \sim N(m\,;\,\sigma) \\ et\ m\ inconnu \\ et\ \ \hat{m} = \bar{Y} \end{cases} \Rightarrow \frac{Y_i - \bar{Y}}{\sigma} \sim N(0\,;\,1) \Rightarrow W = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2_{n-1}\ \ avec \begin{cases} E(W) = n-1 \\ V(W) = 2(n-1) \\ \sigma(W) = \sqrt{2(n-1)} \end{cases}$$

Remarque : $\sum_{i=1}^{i=n}\left(\frac{Y_i-\bar{Y}}{\sigma}\right)^2$ somme des carrés de n variables normales centrées réduites et liées par la relation $\sum_{i=1}^{i=n}(X_i - \bar{X}) = 0$.

7.9.5 théorème central limite et approximations d'autres lois par la loi normale

Approximation de la loi du Chi-deux (ou Khi-deux) à n degrés de liberté : χ_n^2

$$si\ X \sim \chi_n^2\ et\ n > 30 \Rightarrow X \sim N\left(n\,;\ \sqrt{2n}\right) \Longrightarrow \begin{cases} \dfrac{X-n}{\sqrt{2n}} \sim N(0\,;1) \\ \sqrt{2X} - \sqrt{2n-1} \sim N(0\,;1),\ \ pour\ n\ petit \end{cases}$$

Approximation de la loi de Student à n degrés de liberté : T_n

$$si\ X \sim T_n\ et\ n > 30 \Rightarrow X \sim N(0\,;\ 1)$$

Approximation de la loi de Poisson par la loi normale

$$si\ X \sim \mathcal{P}(\lambda)\ et\ n > 30\ et\ \lambda > 18 \Rightarrow X \sim N\left(\lambda\,;\ \sqrt{\lambda}\right) \Longrightarrow \frac{X-\lambda}{\sqrt{\lambda}} \sim N(0\,;1)$$

7.9.6 Autres approximations de lois

Approximation de la loi hypergéométrique par la loi binomiale

$$Si\ K = \sum_{i=1}^{i=n} X_i \sim \mathcal{H}(N,n;p) \underbrace{\Longrightarrow}_{si\ \frac{n}{N}<0{,}05\ et\ n>30\ et\ p<0{,}1\ (ou\ 0<np<10)} K \sim \mathcal{B}(n;p)$$

Approximation de la loi binomiale par la loi de Poisson

$$Si\ K = \sum_{i=1}^{i=n} X_i \sim \mathcal{B}(n;p) \underbrace{\Longrightarrow}_{n>30\ et\ p<0{,}1\ et\ 0<np<15} K \sim \mathcal{P}(\lambda = E(K) = np)$$

7.10 Tableau des propriétés de la moyenne et de la fréquence d'échantillon en situations « avec remise » ou « sans remise »

Estimateurs d'échantillon	Formule	Espérance mathématique (ou moyenne)	Variance Avec ou sans remise	Ecart-type Avec ou sans remise
Moyenne d'échantillon (ou moyenne empirique)	$\boldsymbol{\bar{X}} = \frac{\sum \boldsymbol{X_i}}{\boldsymbol{n}}$	$E(\boldsymbol{\bar{X}}) = m$	$V(\boldsymbol{\bar{X}}) = \frac{\sigma^2}{n}$, avec remise (Tirage indépendant, non exhaustif); $V(\boldsymbol{\bar{X}}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$, sans remise ; (Tirage non indépendant, exhaustif);	$\sigma(\boldsymbol{\bar{X}}) = \frac{\sigma}{\sqrt{n}}$, avec remise $\sigma(\boldsymbol{\bar{X}}) = \frac{\sigma}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N-1}\right)}$, sans remise;
Fréquence d'échantillon	$\boldsymbol{F} = \frac{\boldsymbol{K}}{\boldsymbol{n}}$	$E(\boldsymbol{F}) = p$	$V(\boldsymbol{F}) = \frac{pq}{n}$, avec remise; $V(\boldsymbol{F}) = \frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)$, sans remise;	$\sigma(F) = \sqrt{\frac{pq}{n}}$, avec remise $\sigma(F) = \sqrt{\frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)}$, sans remise;

NB : Tirage avec remise ou indépendant

Si N est très grand ou N inconnu ou si le taux de sondage $\frac{n}{N} < 0{,}05$ on admet que le tirage se fait avec remise et est dit indépendant.

7.11 Distribution d'échantillonnage de la fréquence d'échantillon (F) pour $n > 30$: avec ou sans remise

$n > 30$	Avec remise	$X_i \sim \underbrace{\mathcal{B}(1;p)}_{loi\ de\ Bernoulli} \quad avec \begin{cases} E(X_i) = p \\ V(X_i) = pq \\ \sigma(X_i) = \sqrt{pq} \end{cases}$	$F \rightarrow \mathcal{N}\left(p; \sqrt{\frac{p(1-p)}{n}}\right)$ F : la fréquence d'échantillon converge vers la loi normale pour n>30. En pratique si : np>18 ou si n>30 ou p ≅ 0,5	$U = \frac{\mathrm{F} - \mathrm{p}}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0;1)$
	Sans remise	$K = \sum_{i=1}^{i=n} X_i \sim \underbrace{\mathcal{H}(N;\ n; p)}_{loi\ hypergeométrique}$ $avec \begin{cases} E(K) = np \\ V(K) = npq\frac{N-n}{N-1} \\ \sigma(K) = \sqrt{npq\left(\frac{N-n}{N-1}\right)} \end{cases}$	$F \rightarrow \mathcal{N}\left(p; \sqrt{\frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)}\right)$ si : np>18 ou si n>30 ou p ≅ 0,5	$U = \frac{\mathrm{F} - \mathrm{p}}{\sqrt{\frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)}}$ $U \sim \mathcal{N}(0;1)$
$n \leq 30$	Si la taille de l'échantillon est faible, on a recours aux lois exactes : loi binomiale, loi de poisson, etc.			

CHAPITRE 8 : NOTE DE COURS SUR L'ESTIMATION PAR INTERVALLE DE CONFIANCE D'UNE MOYENNE (m) ET D'UNE PROPORTION (p)

8.1 estimation par intervalle de confiance d'une moyenne de population : *m*

L'estimation de la moyenne m de la population par intervalle de confiance est donnée comme suite :

$$m \in \left]\bar{x} - u_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}} ; \bar{x} + u_{1-\alpha/2} \times \left(\frac{\sigma}{\sqrt{n}}\right)\right[$$

Ou

$$m \in \left]\bar{x} \pm u_{1-\alpha/2} \times \left(\frac{\sigma}{\sqrt{n}}\right)\right[$$

Avec

$$\begin{cases} n = \text{taille de l'échantillon} \\ \bar{x} = \dfrac{\sum x_i}{n} \text{ : moyenne d'échantillon} \\ \dfrac{\sigma}{\sqrt{n}} = \sigma(\bar{X}) \; écart\ type\ de\ \bar{X} \\ u_{1-\alpha/2} = \text{ valeur lue dans la table 2 des fractiles de la loi normale centrée réduite } \mathcal{N}(0;1) \\ 1-\alpha = niveau\ de\ confiance \;\; où \;\; 1-\alpha = P\left(m \in \left]\bar{x} \pm u_{1-\alpha/2} \times \dfrac{\sigma}{\sqrt{n}}\right[\right) \\ \alpha = seuil\ de\ de\ confiance \;\; où \;\; \alpha = P\left(m \notin \left]\bar{x} \pm u_{1-\alpha/2} \times \dfrac{\sigma}{\sqrt{n}}\right[\right) \end{cases}$$

8.2 Conditions d'application des formules d'intervalle de confiance d'une moyenne de population : *m*

$Si\ n > 30\ et\ \sigma\ connu, alors\ \bar{X} \rightsquigarrow \mathcal{N}\left(m; \frac{\sigma}{\sqrt{n}}\right)\ et\ U = \frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0;1)$, on obtient l'intervalle suivant : $$\boldsymbol{m \in \left]\bar{x} - u_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}} ; \bar{x} + u_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right[}$$
$Si\ n > 30\ et\ \begin{cases} \sigma\ inconnu \\ \hat{\sigma} = s' \end{cases} alors\ \bar{X} \rightsquigarrow \mathcal{N}\left(m; \frac{s'}{\sqrt{n}}\right) et\ U = \frac{\bar{X}-m}{\frac{s'}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0;1)$, , on obtient l'intervalle suivant : $$\boldsymbol{m \in \left]\bar{x} - u_{1-\alpha/2} \times \frac{s'}{\sqrt{n}} ; \bar{x} + u_{1-\alpha/2} \times \frac{s'}{\sqrt{n}}\right[}$$
$Si\ n \leq 30\ et\ \sigma\ connu$ et $X \rightsquigarrow \mathcal{N}(m;\sigma)\ alors\ \bar{X} \rightsquigarrow \mathcal{N}\left(m; \frac{\sigma}{\sqrt{n}}\right) et\ U = \frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0;1)$, on obtient l'intervalle suivant : $$\boldsymbol{m \in \left]\bar{x} - u_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}} ; \bar{x} + u_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right[}$$

$Si\ n \leq 30\ et \begin{cases} \sigma\ inconnu \\ \sigma = s' \end{cases}$ et $X \backsim \mathcal{N}(m; s') \longrightarrow \bar{X} \backsim \mathcal{N}\left(m; \frac{s'}{\sqrt{n}}\right) \longrightarrow T = \frac{\bar{X}-m}{\frac{S'}{\sqrt{n}}} \backsim \underbrace{T_{n-1}}_{\substack{loi\ de\ Student \\ à\ (n-1) dégrés\ de\ liberté\ (ddl)}}$ On obtient l'intervalle suivant : $$m \in \left]\bar{x} - t_{(n-1);1-\alpha/2} \times \frac{s'}{\sqrt{n}} ; \bar{x} + t_{(n-1);1-\alpha/2} \times \frac{s'}{\sqrt{n}}\right[$$ $t_{(n-1);1-\alpha/2}$: est lue dans la table de Student par l'intersection de la ligne "(n-1)" et de la colonne "α"

8.3 Formules de la taille d'un échantillon (n) dans le cadre de l'estimation d'une moyenne (m)

$$n \geq \frac{4\left(u_{1-\alpha/2}\right)^2 \sigma^2}{l^2}, avec\ l = amplitude\ ou\ largeur\ de\ l'intervalle$$

$$n \geq \frac{\left(u_{1-\alpha/2}\right)^2 \sigma^2}{\epsilon^2}, avec\ \epsilon = marge\ d'erreur\ \ ou\ \ rayon\ de\ l'intervalle$$

8.4 Exercice pratique de lecture des valeurs $\left(u_{1-\alpha/2}\right)$ dans la table2 de la loi $\mathcal{N}(0;1)$

α Seuil de confiance	$1-\alpha$ Niveau de confiance	$u_{1-\alpha/2}$ Valeur lue dans la table 2 des fractiles de la loi $\mathcal{N}(0;1)$ Par l'intersection de la ligne et de la colonne pour α donné	
$\alpha = 0{,}01$ (1%)	0,99 (99%)	$\alpha = 0{,}01 = \underbrace{0{,}0}_{ligne} + \underbrace{0{,}01}_{colonne}$	$u_{1-\alpha/2} = 2{,}5758$
$\alpha = 0{,}02$ (2%)	0,98 (98%)	$\alpha = 0{,}02 = \underbrace{0{,}0}_{ligne} + \underbrace{0{,}02}_{colonne}$	$u_{1-\alpha/2} = 2{,}3263$
$\alpha = 0{,}05$ (5%)	0,95 (95%)	$\alpha = 0{,}05 = \underbrace{0{,}0}_{ligne} + \underbrace{0{,}05}_{colonne}$	$u_{1-\alpha/2} = 1{,}9600$
$\alpha = 0{,}10$ (10%)	0,90 (90%)	$\alpha = 0{,}10 = \underbrace{0{,}1}_{ligne} + \underbrace{0{,}00}_{colonne}$	$u_{1-\alpha/2} = 1{,}6449$

8.5 Exercice pratique de lecture des valeurs $t_{(n-1);1-\alpha/2}$ dans la table de la loi de Student

$(n-1)$ degrés de liberté	$1-\alpha$ Niveau de confiance	α Seuil de confiance	$t_{(n-1);1-\alpha/2}$ Valeur lue dans la table de la loi de Student par l'intersection de la ligne « (n-1) » et de la colonne « α »	
16	0,99 (99%)	$\alpha = 0{,}01$ (1%)	$\underbrace{16}_{\substack{ligne \\ (n-1)}} \; et \; \underbrace{0{,}01}_{\substack{colonne \\ \alpha}}$	$t_{(n-1);1-\alpha/2} = 2{,}921$
16	0,98 (98%)	$\alpha = 0{,}02$ (2%)	$\underbrace{16}_{\substack{ligne \\ (n-1)}} \; et \; \underbrace{0{,}02}_{\substack{colonne \\ \alpha}}$	$t_{(n-1);1-\alpha/2} = 2{,}583$
16	0,95 (95%)	$\alpha = 0{,}05$ (5%)	$\underbrace{16}_{\substack{ligne \\ (n-1)}} \; et \; \underbrace{0{,}05}_{\substack{colonne \\ \alpha}}$	$t_{(n-1);1-\alpha/2} = 2{,}120$
16	0,90 (90%)	$\alpha = 0{,}10$ (10%)	$\underbrace{16}_{\substack{ligne \\ (n-1)}} \; et \; \underbrace{0{,}10}_{\substack{colonne \\ \alpha}}$	$t_{(n-1);1-\alpha/2} = 1{,}746$

8.6 Exercices d'application de calcul d'un intervalle de confiance d'une moyenne

Application 1

Une entreprise fabrique un certain type de composants électroniques dont la durée de vie X, exprimée en heures, est une variable aléatoire. Des mesures effectuées sur un échantillon aléatoire de taille 50 ont donné les résultats suivants :

$\sum_{i=1}^{50} x_i = 60\,000$; $\sum_{i=1}^{50} x_i^2 = 74.10^6$

a). Donner une estimation ponctuelle de la durée de vie moyenne des composants.
b). Donner une estimation ponctuelle de l'écart-type de cette durée de vie.
c). Donner l'intervalle de confiance à 95%, puis à 99% de cette durée de vie moyenne.
d). Quel aurait dû être la taille de l'échantillon pour que l'intervalle de confiance à 95% de la durée de vie moyenne des composants ait une $\underbrace{amplitude}_{l}$ de 60 heures ?

Corrigé

X est la durée de vie des composants. Et X suit une loi normale N(m; σ).
Données : $\bar{x} = \mathbf{1200}$; $\sum_{i=1}^{16} x_i^2 = 74.10^6$; **n= 50;**

c) Estimation ponctuelle de m
$\hat{m} = \bar{x} = 1200$;
d) Estimation ponctuelle de σ

$$n = 50 > 30 \Rightarrow \hat{\sigma} = s = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} = \sqrt{\frac{74000000}{50} - 1200^2} = 200 .$$

c). L'intervalle de confiance à 95% de cette durée de vie moyenne.

$$m \in \left]\bar{x} - u_{1-\alpha/2}\frac{s}{\sqrt{n}} \; ; \bar{x} + u_{1-\alpha/2}\frac{s}{\sqrt{n}}\right[$$

Application numérique : $1 - \alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow u_{1-\frac{\alpha}{2}} = 1{,}96$; $\bar{x}$ =1200 ; $\hat{\sigma}$=s=200.

$$\text{soit } m \in \left]1200 - 1{,}96\frac{200}{\sqrt{50}}\ ;\ 1200 + 1{,}96\frac{200}{\sqrt{50}}\right[$$

$m \in]1144{,}56\ ;\ 1255{,}44[$ Conclusion : Au niveau de confiance de 95 % la durée de vie moyenne des composants électroniques est comprise entre 1144,56 et 1255,44 heures.

Application numérique : $1 - \alpha = 0{,}99 \Rightarrow \alpha = 0{,}01 \Rightarrow u_{1-\frac{\alpha}{2}} = 2{,}5758$; $\overline{x}$ =1200; $\hat{\sigma}$=s=200.

$$m \in \left]\overline{x} - u_{1\text{-}\alpha/2}\frac{s}{\sqrt{n}}\ ;\ \overline{x} + u_{1\text{-}\alpha/2}\frac{s}{\sqrt{n}}\right[;\ \text{soit } m \in \left]1200 - 2{,}5758\frac{200}{\sqrt{50}}\ ;\ 1200 + 2{,}5758\frac{200}{\sqrt{50}}\right[$$

$m \in]1127{,}14;\ 1272{,}85[$

Conclusion : Au niveau de confiance de 99 % la durée de vie moyenne des composants électroniques est comprise entre 1144,56 et 1272,85 heures.

e) Taille de l'échantillon

$$n \geq \left(\frac{2u_{1-\alpha/2}\sigma}{l}\right)^2$$

$$1 - \alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow u_{1-\frac{\alpha}{2}} = 1{,}96;$$

$$n \geq \left(\frac{2u_{1-\alpha/2}\sigma}{l}\right)^2 = \left(\frac{2 \times 1{,}96 \times 200}{60}\right)^2 = 170{,}74$$

Pour un intervalle d'amplitude 60 heures, il faut un échantillon de 171 composants.

Application 2

Exercice 1 : Intervalles de confiance d'une moyenne et d'un écart-type ; taille d'échantillon

Une laiterie produit des camemberts commercialisés sous la marque « Le moine gourmand ». La masse X, exprimée en g, d'un camembert tiré au hasard dans la production, est distribuée selon une loi normale.

On tire un échantillon simple de 17 camemberts que l'on pèse et dont le tableau suivant fournit les masses :

250	254	254	253	256	250	257	251	253	255	250	255	252	261
252	251	255											

1) a) En utilisant une calculatrice donner le poids moyen $\bar{x}$ et la variance $V(X) = s^2$ puis l'écart-type $\sigma(X) = s$ de cet échantillon.
   b) Donner une estimation ponctuelle de la moyenne $\widehat{m}$ et de l'écart-type $\hat{\sigma} = s'$ de la production.
2) Déterminer une estimation par intervalle de confiance à 95% de la masse moyenne m de la production.
3) Le responsable de fabrication des camemberts « Le moine gourmand » souhaite savoir quelle taille minimale donner à un échantillon aléatoire simple pour obtenir un intervalle de confiance pour m, au niveau 95 %, d'amplitude inférieure à 1.

Le responsable précise alors que σ^2 = 6,25. Calculer la taille minimale de cet échantillon.

Corrigé application 2 :

X = masse du camembert en g.

1) a) Calcul du poids moyen $\bar{x}$ et de la variance d'échantillon $V(X) = s^2$ puis l'écart-type $\sigma(X) = s$ de cet échantillon. Formules et tableau de calcul

La moyenne d'échantillon

$$\bar{x} = \frac{\sum n_i x_i}{n} = \frac{4309}{17} = 253{,}47\ g$$

Tableau de calcul

x_i	n_i	$n_i x_i$	$n_i x_i^2$
250	3	750	187500
251	2	502	126002

La variance d'échantillon : s^2

$$s^2 = \frac{\sum n_i (x_i - \bar{X})^2}{n}$$
$$s^2 = \frac{\sum n_i x_i^2}{n} - (\bar{X})^2$$
$$s^2 = \left(\frac{1092341}{17}\right) - (253{,}47)^2$$
$$s^2 = 8{,}014$$

L'écart-type d'échantillon : s

$s = \sqrt{V(X)} = 2{,}831\ g$

252	2	504	127008
253	2	506	128018
254	2	508	129032
255	3	765	195075
256	1	256	65536
257	1	257	66049
261	1	261	68121
Total	$n = \sum n_i = 17$	$\sum n_i x_i = 4309$	$\sum n_i x_i^2 = 1092341$

b) Donner une estimation ponctuelle du poids moyen de la production et de l'écart-type de la population.

Estimation ponctuelle de la moyenne de la population : $\hat{m} = \bar{x} = 253{,}47\ g$

Estimation ponctuelle de l'écart-type de la population : $n = 17 < 30, \Rightarrow\ \hat{\sigma} = s' = \sqrt{\frac{\sum (x_i - \bar{x})}{n-1}}$

$or\ s' = s\sqrt{\frac{n}{n-1}} = 2{,}831 \times \sqrt{\frac{17}{17-1}} \Rightarrow \hat{\sigma} = s' = 2{,}918$g.

NB : l'estimateur sans biais de l'écart-type de la population est : $\hat{\sigma} = s' = \sqrt{\frac{\sum n_i (x_i - \bar{X})^2}{n-1}}$

3. Intervalle de confiance de la moyenne m au niveau de confiance de 95%

N = 17 < 30 ; on a un échantillon de petite taille. De plus l'écart-type de la production σ étant inconnu, on l'estime par $\hat{\sigma} = s'$. On obtient :

$$\bar{X} \backsim \mathcal{N}\left(m; \frac{s'}{\sqrt{n}}\right) \Rightarrow T = \frac{\bar{X} - m}{\frac{s'}{\sqrt{n}}} \backsim T_{n-1}: loi\ de\ Student \text{ à } (n-1)\ dégrés\ de\ liberté.$$

$$\Rightarrow m \in \left]\bar{x} - t_{(n-1);1-\alpha/2}\frac{s'}{\sqrt{n}}\ ;\ \bar{x} + t_{(n-1);1-\alpha/2}\frac{s'}{\sqrt{n}}\right[$$

Application numérique : $\bar{x} = 253{,}47;\ \ \hat{\sigma} = s' = 2{,}918;$ 1-α = 0,95 ⇒ $\alpha = 0{,}05$

$\Rightarrow \alpha = 0{,}05 \Rightarrow (table\ 3\ de\ Student$ à $(17-1)\ 16\ dégrés\ de\ liberté)\ t_{(16);1-\alpha/2} = 2{,}120$

soit $m \in \left]253{,}47 - 2{,}120\frac{2{,}918}{\sqrt{17}}\ ;\ 253{,}47 + 2{,}120\frac{2{,}918}{\sqrt{17}}\right[\ \ et\ \ m \in\ ]\ 251{,}97\ ;\ 254{,}97\ [$

Au niveau de confiance de 95 %, la masse moyenne est comprise entre 251,97 et 254,97 g.

4. La taille minimale de l'échantillon pour $l < 1$

$$n \geq \frac{4\left(\boldsymbol{u}_{1-\alpha/2}\right)^2 \boldsymbol{\sigma}^2}{l^2}, avec\ l = amplitude\ ou\ largeur\ de\ l'intervalle$$

$$l = 2\left(\boldsymbol{u}_{1-\alpha/2} \times \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right)$$

$$\Rightarrow l < 1 \Leftrightarrow 2 \times \boldsymbol{u_{1-\alpha/2}} \times \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}} < \mathbf{1}\ \boldsymbol{soit}\ \frac{4\left(\boldsymbol{u_{1-\alpha/2}}\right)^{\mathbf{2}}\boldsymbol{\sigma}^{\mathbf{2}}}{n} < 1 \Rightarrow n > 4\left(\boldsymbol{u_{1-\alpha/2}}\right)^{\mathbf{2}}\boldsymbol{\sigma}^{\mathbf{2}}$$

Application numérique : l = 1 ; σ^2 = 6,25 ; 1-α = 0,95 ⇒ $\alpha = 0{,}05 \Rightarrow (table\ 2)\ \boldsymbol{u_{1-\alpha/2}} = \mathbf{1{,}96}$

$$n > 4\left(\boldsymbol{u_{1-\alpha/2}}\right)^{\mathbf{2}}\boldsymbol{\sigma}^{\mathbf{2}} = 4 \times 1{,}96^2 \times 6{,}25^2 = 600{,}25\ \ soit\ n = 601\ camemberts$$

8.7 Intervalles de confiances d'une proportion *p*

Si $n > 30$, on applique le théorème central limite $alors\ F \leadsto \mathcal{N}\left(p; \sqrt{\frac{pq}{n}}\right)$ et $U = \frac{F-p}{\sqrt{\frac{pq}{n}}} \leadsto \mathcal{N}(0;1)$

On obtient l'intervalle suivant :

$$p \in \left] f - u_{1-\alpha/2} \times \sqrt{\frac{f(1-f)}{n}} ; f + u_{1-\alpha/2} \times \sqrt{\frac{f(1-f)}{n}} \right[$$

8.8 Formules de la taille (n) d'un échantillon dans le cadre de l'estimation d'une proportion (p) de la population

$$n \geq \frac{4\left(\boldsymbol{u_{1-\alpha/2}}\right)^{\mathbf{2}}\boldsymbol{f}(\mathbf{1}-\boldsymbol{f})}{l^2}, avec\ l = amplitude\ ou\ largeur\ de\ l'intervalle$$

$$n \geq \frac{\left(\boldsymbol{u_{1-\alpha/2}}\right)^{\mathbf{2}}\boldsymbol{f}(\mathbf{1}-\boldsymbol{f})}{\epsilon^2}, avec\ \epsilon = marge\ d'erreur\ \ ou\ \ rayon\ de\ l'intervalle$$

8.9 Exercices d'application de calcul d'un intervalle de confiance d'une proportion

Application 1 – Dans un échantillon aléatoire de 200 composantes électriques fabriquées avec un certain procédé, 15 composantes se sont avérées défectueuses.

a) Soit p la proportion de composantes fabriquées avec ce procédé qui sont défectueuses. Construire un intervalle de confiance pour p ayant un niveau de confiance de 90%.

b) Combien de composantes doit on échantillonner pour estimer la proportion de composantes défectueuses à $\underbrace{3\%\ près}_{\varepsilon=3\%=0{,}03}$ et ceci $\underbrace{18\ fois\ sur\ 20}_{1-\alpha=\frac{18}{20}=0{,}90}$?

Solution . Les données sont : n=200>30 ; k=15 ; 1-α = 0,90 ;

a) $f = \frac{k}{n} = \frac{15}{200} = 0{,}075\ ;\ \ 1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow u_{1-\frac{\alpha}{2}} = 1{,}6449;$

$$p \in \left] f - u_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}}\ ; f + u_{1-\alpha/2}\sqrt{\frac{f(1-f)}{n}} \right[$$

Application numérique :

$$\Rightarrow p \in \left]0{,}075 - 1{,}96\sqrt{\frac{0{,}075(0{,}925)}{200}}\ ;\ 0{,}075 - 1{,}96\sqrt{\frac{0{,}075(0{,}925)}{200}}\right[$$

soit p ∈]0,0385; 0,1115[

Conclusion : au niveau de confiance de 95%, la proportion de composantes défectueuses est comprise entre 3,85% et 11,15%.

b) Taille de l'échantillon

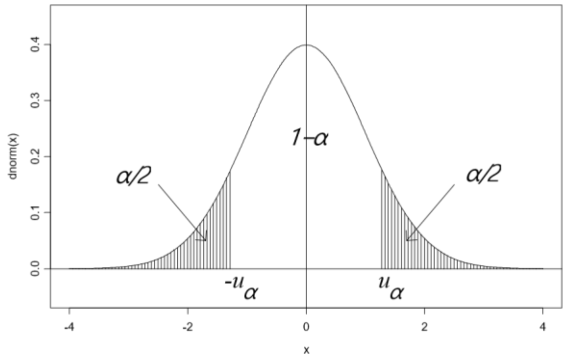
$$n \geq \left(\frac{u_{1-\alpha/2}\sqrt{f(1-f)}}{\varepsilon}\right)^2$$

$f = 0{,}075;\ 1-\alpha = 0{,}90 \Rightarrow \alpha = 0{,}10 \Rightarrow u_{1-\frac{\alpha}{2}} = 1{,}6449$

$n = \left(\frac{u_{1-\alpha/2}\sqrt{f(1-f)}}{0{,}03}\right)^2 = \left(\frac{1{,}6449\sqrt{0{,}075(0{,}925)}}{0{,}03}\right)^2 =$ 208,56 soit 209 composantes.

TABLES STATISTIQUES

TABLE 2 DE LA LOI NORMALE CENTREE REDUITE

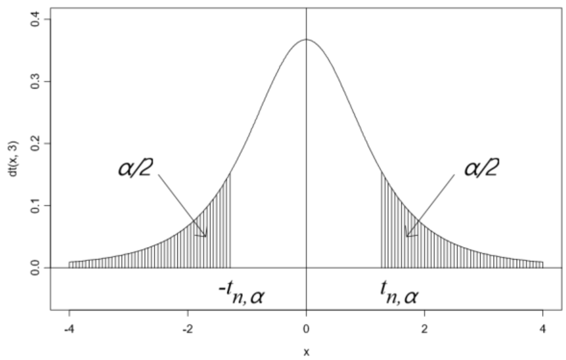


α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	+∞	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	0.6280	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4959	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

Petites valeurs de α

α	0.002	0.001	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
u_α	3.0902	3.2905	3.8906	4.4171	4.8916	5.3267	5.7307	6.1094

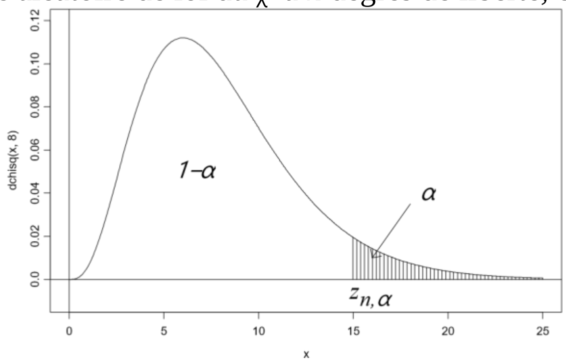
TABLE DE LA LOI DE STUDENT



n \ α	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.62
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
80	0.126	0.254	0.387	0.527	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.416
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
$+\infty$	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

TABLE 4 DE LA LOI DU KHI2 : χ^2

X étant une variable aléatoire de loi du χ^2 à n degrés de liberté, et α un réel de [0,1], la table



n \ α	0.995	0.990	0.975	0.95	0.9	0.8	0.7	0.5	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001
1	0.00004	0.0002	0.001	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	5.02	6.63	7.88	10.80
2	0.01	0.02	0.05	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.61	5.99	7.38	9.21	10.60	13.82
3	0.07	0.11	0.22	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.81	9.35	11.34	12.84	16.27
4	0.21	0.30	0.48	0.71	1.06	1.65	2.19	3.36	4.88	5.99	7.78	9.49	11.14	13.28	14.86	18.47
5	0.41	0.55	0.83	1.15	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	12.83	15.09	16.75	20.52
6	0.68	0.87	1.24	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	14.45	16.81	18.55	22.46
7	0.99	1.24	1.69	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.01	18.48	20.28	24.32
8	1.34	1.65	2.18	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	17.53	20.09	21.95	26.12
9	1.73	2.09	2.70	3.33	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	2.56	3.25	3.94	4.87	6.18	7.27	9.34	11.78	13.44	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.05	3.82	4.57	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	21.92	24.72	26.76	31.26
12	3.07	3.57	4.40	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	4.11	5.01	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	4.66	5.63	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.06	23.68	26.12	29.14	31.32	36.12
15	4.60	5.23	6.26	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	5.81	6.91	7.96	9.31	11.15	12.62	15.34	18.42	20.47	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	6.41	7.56	8.67	10.09	12.00	13.53	16.34	19.51	21.61	24.77	27.59	30.19	33.41	35.72	40.79
18	6.26	7.01	8.23	9.39	10.86	12.86	14.44	17.34	20.60	22.76	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	7.63	8.91	10.12	11.65	13.72	15.35	18.34	21.69	23.90	27.20	30.14	32.85	36.19	38.58	43.82
20	7.43	8.26	9.59	10.85	12.44	14.58	16.27	19.34	22.77	25.04	28.41	31.41	34.17	37.57	40.00	45.31
21	8.03	8.90	10.28	11.59	13.24	15.44	17.18	20.34	23.86	26.17	29.62	32.67	35.48	38.93	41.40	46.80
22	8.64	9.54	10.98	12.34	14.04	16.31	18.10	21.34	24.94	27.30	30.81	33.92	36.78	40.29	42.80	48.27
23	9.26	10.20	11.69	13.09	14.85	17.19	19.02	22.34	26.02	28.43	32.01	35.17	38.08	41.64	44.18	49.73
24	9.89	10.86	12.40	13.85	15.66	18.06	19.94	23.34	27.10	29.55	33.20	36.42	39.36	42.98	45.56	51.18
25	10.52	11.52	13.12	14.61	16.47	18.94	20.87	24.34	28.17	30.68	34.38	37.65	40.65	44.31	46.93	52.62
26	11.16	12.20	13.84	15.38	17.29	19.82	21.79	25.34	29.25	31.79	35.56	38.89	41.92	45.64	48.29	54.05
27	11.81	12.88	14.57	16.15	18.11	20.70	22.72	26.34	30.32	32.91	36.74	40.11	43.19	46.96	49.64	55.48
28	12.46	13.56	15.31	16.93	18.94	21.59	23.65	27.34	31.39	34.03	37.92	41.34	44.46	48.28	50.99	56.89
29	13.12	14.26	16.05	17.71	19.77	22.48	24.58	28.34	32.46	35.14	39.09	42.56	45.72	49.59	52.34	58.30
30	13.79	14.95	16.79	18.49	20.60	23.36	25.51	29.34	33.53	36.25	40.26	43.77	46.98	50.89	53.67	59.70

Pour $n > 30$, on admet que : $\alpha \approx \frac{1}{2}\left(u_{2\alpha} + \sqrt{2n-1}\right)^2$ si $\alpha < \frac{1}{2}$ $z_{n,}$

$\alpha \approx \frac{1}{2}\left(\sqrt{2n-1} - u_{2(1-\alpha)}\right)^2$ si $\alpha \geq \frac{1}{2}$