Chapitre 2

Estimation par intervalle de confiance

1. Estimation ponctuelle

1.1. Notion d'estimateur

Soit X une variable observée sur une population Ω dont la loi dépend d'un paramètre inconnu θ. Soit $(X_1, X_2, ..., X_n)$ une suite de variables issue de la variable X. Un estimateur T_n de θ sera une variable aléatoire $T_n = T_n(X_1, X_2, ..., X_n)$ fonction de l'échantillon.

Par exemple, $T_n = \overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ est un estimateur de θ.

La valeur de T_n calculée à partir d'un échantillon observé est appelée estimation de θ, elle sera notée $T_n(x_1, x_2, ..., x_n)$.

1.2. Exemples d'estimateurs

1.2.1. Estimateur de la moyenne empirique

On appelle moyenne empirique du caractère X sur un échantillon $(X_1, X_2, ..., X_n)$ la variable aléatoire $\overline{X}$ suivante :

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

On suppose que l'échantillon $(X_1, X_2, ..., X_n)$ est formé de variables aléatoires indépendantes et de même loi que X.

On pose

$$E(X_i) = m \quad \text{et} \quad Var(X_i) = \sigma^2 \quad \forall i.$$

On a :

Espérance de $\overline{X}$: $E(\overline{X}) = m$

Variance de $\overline{X}$: $V(\overline{X}) = \dfrac{\sigma^2}{n}$

En effet, on a:

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} m = \frac{1}{n}(n \times m) = m$$

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2}(n \times \sigma^2) = \frac{\sigma^2}{n}$$

1.2.2. Estimateur de la variance

On appelle variance empirique du caractère X sur un échantillon $(X_1, X_2, ..., X_n)$ la variable aléatoire S^2 suivante :

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

On appelle variance empirique corrigée la variable aléatoire $\overline{S}^2$ suivante:

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

On a donc la relation :

$$nS^2 = (n-1)\overline{S}^2$$

D'où

$$\overline{S}^2 = \frac{n}{n-1} S^2$$

On a :

Espérance de S^2 : $E\left(S^2\right) = \frac{n-1}{n}\sigma^2$

Espérance de $\overline{S}^2$: $E\left(\overline{S}^2\right) = \sigma^2$

On démontre (et nous l'admettons) que, dès lors que la loi *L* admet des moments jusqu'à l'ordre 4, S^2 et $\overline{S}^2$ possèdent une variance. Celle-ci tend vers zéro lorsque n augmente.

1.2.3. Estimation ponctuelle d'une proportion (pourcentage)

Supposons qu'une population Ω est constituée de deux catégories d'individus A et B ; où la catégorie A est en proportion inconnue ***p*** inconnue. Si sur un échantillon représentatif donné on observe une proportion ***f*** d'individus de la catégorie A, alors on estime que: ***p = f***

1.2.4. Propriété d'un estimateur

1.2.4.1. Estimateur sans biais

Un estimateur T_n de θ est dit sans biais si

$$E(T_n) = \theta \ , \ \forall \theta \in \Theta$$

où Θ est appelé espace des paramètres ou espace des états de la nature.

- La moyenne empirique $\overline{X}$ est un estimateur sans biais de la moyenne $E(X_i) = m$

$$E(\overline{X}) = m$$

ici le paramètre θ est m.

- La variance empirique corrigée $\bar{S}^2$ est un estimateur sans biais de la variance $V(X_i) = \sigma^2$

$$E(\bar{S}^2) = \sigma^2$$

Ici le paramètre θ est σ^2.

1.2.4.2. Estimateur asymptotiquement sans biais

Un estimateur T_n de θ est dit asymptotiquement sans biais si
$\forall \theta \in \Theta$, $E(T_n) \to \theta$ *quand* $n \to +\infty$.

La variance empirique S^2 est un estimateur asymptotique sans biais de la variance. On a :

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

Donc :

$$E(S^2) \to \sigma^2 \quad quand \quad n \to +\infty$$

1.2.4.3. Estimateur convergent

Un estimateur sans biais ou asymptotiquement sans biais dont la variance tend vers zéro lorsque n tend vers l'infini est convergent.

- $\overline{X}$ est un estimateur convergent de m car il est sans biais et sa variance $\frac{\sigma^2}{n}$

tend vers zéro lorsque $n \to \infty$.

- S^2 et $\bar{S}^2$ sont des estimateurs convergents de □2 .

En rfésumé
En l'absence de tout autre information sur la variable X, on estime

- que la moyenne inconnue ***m*** de la variable X sur la population Ω est égale à la moyenne de X sur l'échantillon ; c'est-à-dire $m = \bar{x}$.

- que la meilleure estimation ponctuelle de l'écart type inconnu σ (de la population) est donnée en fonction de l'écart type de l'échantillon (σ_e) par la formule : $\sigma = \sigma_e \sqrt{\frac{n-1}{n}} = s\sqrt{\frac{n-1}{n}}$.

1.2.5. Comparaison des estimateurs

Soient T_n et $\tilde{T}_n$ deux estimateurs sans biais d'un paramètre θ . T_n est dit plus efficace que $\tilde{T}_n$ si :

$$\forall\ \theta \in \Theta,\ \ V(T_n) \leq V(\tilde{T}_n)$$

2. Estimation par intervalle de confiance

Entreprendre une estimation ponctuelle est une démarche naturelle. En effet si l'on se trouve placé face à un phénomène aléatoire dépendant d'un paramètre inconnu θ, il est logique de chercher à disposer d'une valeur numérique de ce paramètre. Cependant il existe de nombreuses situations où une telle estimation ponctuelle n'est pas, en elle-même d'un grand intérêt.

La méthode d'estimation ponctuelle est une méthode qui n'est donc pas entièrement satisfaisante. En effet, les valeurs estimées varient d'un échantillon à l'autre, et peuvent parfois être très éloignées des vraies caractéristiques ***m, σ*** et ***p*** de la population Ω. Cette estimation est donc peu précise. Pour palier à ce défaut de précision, on définit à partir de l 'échantillon, un intervalle dans lequel, on est sûr de trouver la valeur inconnue du paramètre avec un risque donné. Estimer un paramètre en montrant qu'il appartient avec une probabilité donnée (par exemple 95%) à un intervalle, est ce que l'on appelle réaliser une estimation par intervalle de confiance.

Remarque

Dans la suite nous distinguerons deux types d'échantillons :
- Les petits échantillons dont la taille est inférieure à 30.
- Les grands échantillons dont la taille est supérieure ou égale à 30.

2.1. Intervalle de confiance pour un paramètre réel

Définition

Soit (X_1, …, X_n) un échantillon issu d'une variable aléatoire dont la loi de probabilité dépend d'un paramètre inconnu θ. Soit $\alpha \in]0,1[$ un réel quelconque fixé *a prior*.
On appelle intervalle de confiance pour le paramètre θ au niveau de confiance $1-\alpha$, tout intervalle de la forme $[A_n\ , B_n]$, où A_n et B_n sont deux statistiques sur l'échantillon (X_1, …, X_n) telles que :

$$\Pr[A_n \leq \theta \leq, B_n] = 1-\alpha$$

Remarques

• Le réel α peut être interprété comme le risque que l'intervalle de confiance $[A_n \, , B_n]$ ne contienne pas la vraie valeur du paramètre.

• Si A_n et B_n sont toutes deux finies et aléatoires, on dit que l'intervalle $[A_n \, , B_n]$ est bilatéral. Si l'une des statistiques A_n ou B_n, est certaine ou infinie, l'intervalle est dit unilatéral.

2.2. Intervalle de confiance d'une moyenne

Soit $(X_1, \ldots, X_n)$ un échantillon issu d'une variable aléatoire suivant une loi normale de moyenne inconnue m et d'écart type σ.
Soit α un réel quelconque fixé *a priori*, $\alpha \in]0,1[$. On se propose de déterminer un intervalle de confiance pour le paramètre θ au niveau de confiance $1 - \alpha$:

$$\Pr[A_n \leq m \leq, B_n] = 1 - \alpha$$

Pour construire un intervalle de confiance pour la moyenne m inconnue de la variable X suivant une loi normale N (*m*, σ), on considère la variable suivante :

$$T = \frac{\overline{X}_n - m}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m\right)$$

La quantité $T = \frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m\right)$ est une variable aléatoire qui dépend du paramètre que l'on cherche à estimer, *m*, mais dont la loi ne dépend pas de ce paramètre et est parfaitement connue si l'on connaît le paramètre σ.

- Si σ est connu, $\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m\right)$ est une variable aléatoire pivotable pour le paramètre *m* de N (m, σ). Dans ce cas $\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m\right) \to N(0,1)$
- Si σ est inconnu, on l'estime par $\overline{S}$. Dans ce cas $\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m\right)$suit une loi de Student à $(n-1)$ degrés de liberté.

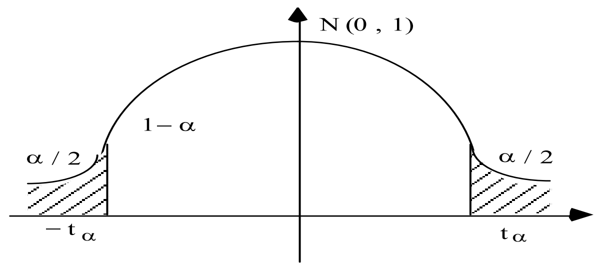
$$\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m\right) \mathrm{d} \to T(n-1)$$

-

2.2.1. Cas où σ est connu

Soit (X_1, …, X_n) l'échantillon indépendant. On utilise comme estimateur de m, la moyenne empirique $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$\overline{X}$ suit alors une loi normale $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$. La variable $\frac{\sqrt{n}\left(\overline{X} - m\right)}{\sigma}$ suit une loi normale N(0 , 1).



$$\Pr\left(-t_\alpha \le \frac{\sqrt{n}\left(\overline{X} - m\right)}{\sigma} \le t_\alpha\right) = 1 - \alpha$$

$$\Pr\left(\overline{X} - \frac{\sigma}{\sqrt{n}} t_\alpha \le m \le \overline{X} + \frac{\sigma}{\sqrt{n}} t_\alpha\right) = 1 - \alpha$$

Ceci conduit à l'intervalle bilatéral symétrique centré en $\overline{X}$:

$$\overline{X} - \frac{\sigma}{\sqrt{n}} t_\alpha \le m \le \overline{X} + \frac{\sigma}{\sqrt{n}} t_\alpha$$

Connaissant la moyenne et l’écart type de l’échantillon, on obtient l’intervalle

$$m \in \left[\overline{x} - t_\alpha \frac{\sigma}{\sqrt{n}} \; ; \; \overline{x} + t_\alpha \frac{\sigma}{\sqrt{n}}\right]$$

t_α est lu sur la table de la loi normale centrée réduite.

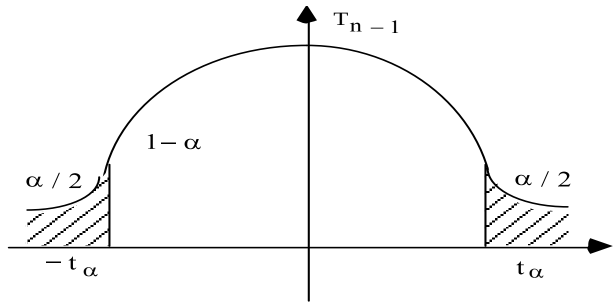
2.2.2. Cas où σ est inconnu (situation la plus fréquente) :et n < 30

La loi de $\overline{X}$ dépendant de σ, on utilise comme estimateur de σ^2 la variance empirique $\overline{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

On sait que la variable aléatoire $\frac{(n-1)\overline{S}_n^2}{\sigma^2}$ suit une loi du Khi-Deux à $(n-1)$ degrés de liberté.

Par suite on a $\frac{\bar{S}_n^2}{\sigma^2}$ suit une loi $\frac{\chi^2(n-1)}{n-1}$.

Le rapport $\frac{\sqrt{n}\left(\bar{X}-m\right)}{\sigma}\frac{\sigma}{\bar{S}_n}=\frac{\sqrt{n}\left(\bar{X}-m\right)}{\bar{S}_n}$ suit une loi de Student à $(n-1)$ degrés de liberté.



$$\Pr\left(-t_\alpha \le \frac{\sqrt{n}\left(\bar{X}-m\right)}{\bar{S}_n} \le t_\alpha\right)=1-\alpha$$

$$\Pr\left(\bar{X}-\frac{\bar{S}_n}{\sqrt{n}}t_\alpha \le m \le \bar{X}+\frac{\bar{S}_n}{\sqrt{n}}t_\alpha\right)=1-\alpha$$

Ce qui conduit à l'intervalle bilatéral symétrique :

$$\bar{X}-\frac{\bar{S}_n}{\sqrt{n}}t_\alpha \le m \le \bar{X}+\frac{\bar{S}_n}{\sqrt{n}}t_\alpha$$

On sait $\bar{S}_n = S_n\sqrt{\frac{n}{n-1}}$, on a :

$$\bar{X}-\frac{S_n}{\sqrt{n-1}}t_\alpha \le m \le \bar{X}+\frac{S_n}{\sqrt{n-1}}t_\alpha$$

Connaissant la moyenne et l’écart type de l’échantillon, on obtient l’intervalle

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

t_α est lu sur la table de la loi de Student à $(n-1)$ degrés de liberté.

2.2.3 Cas où σ est inconnu et n ≥ 30 ("grands échantillons")

Si la taille de l'échantillon est supérieure à 30, la loi de Student est remplacée par la loi normale.

t_α est lu sur la table de la loi normale.

Dans ce cas l'intervalle de confiance s'écrit aussi :

$$\overline{X} - \frac{\overline{S}_n}{\sqrt{n}} t_\alpha \leq m \leq \overline{X} + \frac{\overline{S}_n}{\sqrt{n}} t_\alpha$$

Connaissant la moyenne et l'écart type de l'échantillon, on obtient l'intervalle

$$m \in \left[\overline{x} - t_\alpha \frac{s}{\sqrt{n}} \ ; \ \overline{x} + t_\alpha \frac{s}{\sqrt{n}} \right]$$

t_α est lu sur la table de la loi normale.

Exemple

Un analyste financier étudie les comptes de 200 clients ayant souscrit un emprunt.
A partir d'un échantillon de 20 comptes, il trouve que le solde moyen d'un compte est de 1514,69 Francs avec un écart type égal à 453,34 Francs.
Donner un intervalle de confiance à 95% du solde moyen d'un compte.

Corrigé

Les hypothèses se traduisent par :

$$n = 20$$
$$\overline{x} = 1514{,}69$$
$$s = 453{,}34$$

L'écart type de la population est inconnu, nous devons utiliser la loi de Student à $(n-1)$ degrés de liberté : T_{19}.

$$\alpha = 0{,}05 \qquad\qquad t_\alpha = 2{,}093$$

L'intervalle de confiance ayant 95 chances sur 100 de contenir la valeur vraie de la moyenne m est :

$$\overline{x} - \frac{s}{\sqrt{n-1}} t_\alpha \leq m \leq \overline{x} + \frac{s}{\sqrt{n-1}} t_\alpha$$

$$1514{,}69 - \frac{2{,}093 \times 453{,}34}{\sqrt{19}} \leq m \leq 1514{,}69 + \frac{2{,}093 \times 453{,}34}{\sqrt{19}}$$

$$1297{,}02 \ \leq \ m \ \leq \ 1732{,}36$$

2.3. Estimation par intervalle de confiance d'une proportion

2.3.1. Intervalle de confiance d'une proportion

θ est la proportion p d'individus de P qui possèdent la propriété Q. La variable aléatoire X_i associée au $i^{ème}$ individu a_i d'un échantillon de taille n est définie par :

$$\begin{cases} X_i = 1 \ \ si \ \ a_i \ \text{ possède la propriété } Q_i \\ X_i = 0 \ \text{ sinon} \end{cases}$$

$$T_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{N_n}{n} = F_n$$

F_n est la fréquence de Q dans l'échantillon, X_i suit une loi de Bernoulli de paramètre p.

$N_n \to B(n, p)$: loi binomiale

$E(N_n) = n\,p$ $\qquad$ $E(F_n) = p$

$Var(N_n) = n\,p\,q$ $\qquad$ $Var(F_n) = \dfrac{pq}{n}$

On suppose que l'on est dans les conditions d'approximation de la loi binomiale B(n, p) par la loi normale.

Une approximation de la loi de F_n est alors la loi normale $N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)$.

On considère donc que la variable :

$$T_n = \frac{F_n - p}{\sqrt{\dfrac{p(1-p)}{n}}} \to N(0,1).$$

On a alors :

$$P\left(-t_\alpha\sqrt{\frac{p(1-p)}{n}} \le F_n - p \le t_\alpha\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

Connaissant la fréquence f_0 du caractère dans un échantillon de taille n, dans la pratique on remplace dans la racine carrée p par f_0. D'où

$$P\left(-t_\alpha\sqrt{\frac{f_0(1-f_0)}{n}} \le f_0 - p \le t_\alpha\sqrt{\frac{f_0(1-f_0)}{n}}\right) = 1-\alpha$$

Si on pose $\sigma_f = \sqrt{\dfrac{f_0(1-f_0)}{n}}$, alors on a :

$$P\left(f_0 - t_\alpha \sigma_f \le p \le f_0 + t_\alpha \sigma_f\right) = 1-\alpha$$

L'intervalle $\left[f_0 - t_\alpha \sigma_f \ , \ f_0 + t_\alpha \sigma_f \right]$ est l'intervalle de confiance de la proportion p, de niveau de confiance $1 - \alpha$. Ses limites sont :

$$p_1 = f_0 - t_\alpha \sigma_f$$
$$p_2 = f_0 + t_\alpha \sigma_f$$

f_0 est la valeur de F_n calculée à partir de l'échantillon observé.

t_α est lu sur la table de la loi normale centrée réduite.

si $\alpha = 0{,}01$ on a $t_\alpha = 2{,}576$.
si $\alpha = 0{,}05$ on a $t_\alpha = 1{,}96$.
si $\alpha = 0{,}10$ on a $t_\alpha = 1{,}645$.

2.3.2 Marge d'erreur associée à l'estimation d'une proportion

Comme la valeur observée f_0 est sujette à des fluctuations d'échantillonnage, il existera pratiquement toujours un écart entre la valeur observée f_0 et la valeur réelle p. Cet écart, en valeur absolue, constitue la marge d'erreur dans l'estimation de p. Cette quantité s'appelle également la précision du sondage.
La marge d'erreur E sera au plus égale à :

$$t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}} \quad \text{soit} \quad \left| p - f_0 \right| \le t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}} = E$$

L'intervalle de confiance s'exprime donc comme suit :

$$f_0 - E \le p \le f_0 + E$$

Cette marge d'erreur est d'autant plus petite (la précision du sondage est plus grande) que l'intervalle de confiance est plus petit. Pour un même niveau de confiance, on améliore la précision du sondage (diminution de la marge d'erreur) en augmentant la taille de l'échantillon.

Exemple

Quelques jours avant le second tour d'une élection présidentielle, un sondage réalisé sur un échantillon représentatif de 1000 personnes indique un pourcentage d'intentions de vote en faveur d'un candidat X égal à 0,485.
Déterminer un intervalle de confiance pour p, la proportion vraie d'intentions de vote en faveur du candidat X, et ceci avec un niveau de confiance de 90 %, de 95 % et de 99 %. Déterminer la marge d'erreur dans l'estimation de p.

Corrigé :

On a n = 1000 $f_0 = 0{,}485$

L'estimation ponctuelle de p est $f_0 = 0{,}485$

Les intervalles de confiance et les marges d'erreur associées pour les différents niveaux de confiance sont présentés dans le tableau suivant :

$f_0 = 0{,}485$

			Limite inférieure	**Limite supérieure**
α	t_α	$t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}}$	$f_0 - t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}}$	$f_0 + t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}}$
0,10	1,645	0,0260	0,459	0,511
0,05	1,96	0,0310	0,454	0,516
0,01	2,576	0,0407	0,444	0,525

Ainsi pour un niveau de confiance de 95%, l'intervalle $0{,}454 \leq p \leq 0{,}516$ encadre p, la proportion des intentions de vote favorables au candidat X.

La marge d'erreur statistique est 3,1 %. On peut donc considérer, avec un niveau de confiance de 95 %, que l'écart (en valeur absolue) entre la valeur estimée de p (0,485, soit 48,5 %) et sa valeur réelle n'excèdera pas 3,1% avec le nombre d'individus qui a été sondé.

2.3.3 Taille d'échantillon requise pour estimer la proportion p avec une marge d'erreur E

On peut facilement déterminer la taille minimale de l'échantillon requise pour une marge d'erreur fixée à l'avance.
On résout pour n, l'équation :

$$E = t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}}$$

On doit distinguer deux cas :

Cas n° 1 :
On connaît une valeur approximative f_0 de p obtenue d'un sondage préalable sur un petit échantillon (par exemple $n = 30$). Pour une marge d'erreur désirée E et un niveau de confiance $1-\alpha$, la taille d'échantillon minimale requise est :

$$n = \frac{t_\alpha{}^2 \; f_o(1-f_0)}{E^2}$$

Cas n° 2 : Si on ne peut obtenir par un sondage préalable une valeur approximative de p, on la fixe à 0,50. Cette valeur représente le cas le plus défavorable. Dans ce cas, la taille d'échantillon requise pour un niveau de confiance $1-\alpha$ est :

$$n = \frac{t_\alpha{}^2\, f_o(1-f_0)}{E^2} = \frac{t_\alpha{}^2\,(0{,}5)(0{,}5)}{E^2} = \frac{t_\alpha{}^2}{4E^2}$$

On donne dans le tableau ci-dessous les tailles d'échantillon requises pour divers niveaux de confiance et diverses valeurs de la marge d'erreur. On suppose ici que nous n'avons aucune connaissance de p, nous la fixons à 0,5.

$n = \frac{t_\alpha{}^2}{4E^2}$		**Niveau de confiance**		
	$1-\alpha$	0,90	0,95	0,99
	t_α	1,645	1,960	2,576
Marge d'erreur E	$4E^2$	**Taille n**	**Taille n**	**Taille n**
0,01	0,0004	6764	9604	16587
0,015	0,0009	3006	4268	7372
0,02	0,0016	1691	2401	4147
0,025	0,0025	1082	1537	2654
0,03	0,0036	752	1067	1843
0,035	0,0049	552	784	1354
0,04	0,0064	423	600	1037
0,045	0,0081	334	474	819
0,05	0,01	271	384	663
0,06	0,0144	188	267	461
0,07	0,0196	138	196	339
0,08	0,0256	106	150	259
0,09	0,0324	84	119	205
0,10	0,04	68	96	166

Exemple 2

Un groupe d'étudiantes inscrites en Sciences de la Santé vont effectuer un sondage auprès de la population étudiante pour estimer le pourcentage d'adeptes du tabagisme. La population étudiante est d'environ 8 000.

1) Déterminer la taille d'échantillon requise pour assurer une marge d'erreur (en valeur absolue) n'excédant pas 5 %, avec un niveau de confiance de 95 %. Une enquête similaire effectuée , il y a 3 ans, indiqua que 32 % d'individus fumaient régulièrement.

2) Déterminer la taille d'échantillon requise en supposant qu'on a aucune information sur p.

Corrigé :

1) $t_\alpha = 1{,}96$; $f_0 = 0{,}32$; $E = 0{,}05$

La taille d'échantillon minimale requise est :

$$n = \frac{t_\alpha{}^2 \; f_o(1-f_0)}{E^2} = \frac{(1{,}96^2)(0{,}32)(1-0{,}32)}{(0{,}05)^2} = 335$$

2) $t_\alpha = 1{,}96$; $f_0 = 0{,}5$; $E = 0{,}05$

Dans ce cas, la taille d'échantillon requise est :

$$n = \frac{t_\alpha{}^2}{4E^2} = \frac{(1{,}96)^2}{4(0{,}05)^2} = 385$$

Énoncés des exercices du chapitre 1

A) Intervalle pour une proportion

Exercice 1

Un sondage sur la popularité du Premier Ministre indique que 51% des personnes interrogées sont favorables à sa politique. Construire un intervalle de confiance de niveau 0,95 pour la proportion p de sénégalais favorables à cette politique, sachant que ce sondage a été réalisé auprès de n = 100 personnes. Même question si n = 1000. Quelle aurait dû être la taille d'échantillon pour que l'intervalle soit de longueur inférieure à 4% ?

Exercice 2

A la sortie d'une chaîne de montage, 20 véhicules automobiles tirés au sort sont testés de façon approfondie. Sachant que deux d'entre eux présentent des défauts graves et doivent repasser dans la chaîne, construire un intervalle de confiance de niveau 0,95 pour la proportion p de véhicules défectueux.

Exercice 3

Une firme fabriquant une poudre de lessive envisage de modifier la présentation de son produit tout en majorant son prix. L'entreprise fabrique une petite série avec la nouvelle présentation et la propose à la clientèle dans un petit nombre de magasins témoins choisis en fonction du réseau de commercialisation de son produit. On constate que sur 300 décisions d'achat, 156 se sont portées sur la nouvelle présentation et 144 sur l'ancienne.

a) Quel est l'intervalle de confiance de la proportion des ménagères qui préfèrent la nouvelle présentation $(\alpha = 0{,}05)$?

b) Déterminer la taille minimale de l'échantillon permettant de connaître à $\pm 1\%$ près (au plus) la proportion des ménagères préférant la nouvelle présentation au seuil de 5%.

Exercice 4

Sur la base d'une ligne de pauvreté correspondant à une consommation de 2400 calories par personne et par jour, les résultats du QUID ont permis d'évaluer la proportion des ménages sénégalais en dessous du seuil de pauvreté à 53,9 % en 2001.

Estimer le nombre de pauvres sur l'échantillon des 1000 prochaines naissances. (Prendre un niveau de confiance de 95 %)

Donner la marge d'erreur dans l'estimation de la proportion de pauvres.

(**Source :** Questionnaire Unifié des Indicateurs de Développement (QUID) de l'Enquête Sénégalaise Auprès des Ménages II, 2001)

Exercice 5

Lors d'un sondage, sur un échantillon de 200 personnes, on a recueilli 84 intentions de vote en faveur du parti A. Soit p la proportion de votes pour A ; donner un intervalle de confiance pour p au niveau 95 %.

B) Intervalle pour une moyenne

Exercice 6

Une enquête réalisée par un constructeur d'imprimante sur un échantillon de 200 machines montre que la durée moyenne de vie de la tête d'impression matricielle à aiguilles est de 155 millions d'impacts avec un écart-type égal à 70 millions d'impacts.
Estimer la durée de vie moyenne des têtes d'impression de ce type de machines au seuil $\alpha = 0{,}05$.

Exercice 7

Sur un échantillon de 100 câbles métalliques, on a observé une résistance moyenne à la rupture de 2,63 tonnes.

a) Sachant que de façon générale, la résistance de ce type de câble présente un écart-type $\sigma = 0{,}025$ tonne, estimer sa résistance moyenne au seuil $\alpha = 0{,}05$.

b) Déterminez quelle devrait être la taille minimale de l'échantillon pour connaître la résistance moyenne de ce type de câble à $\pm$ 1 kg près (au plus) au seuil de 5%.

c) On suppose maintenant que l'écart-type de la population est inconnu. Un échantillon de 17 câbles indique une résistance moyenne à la rupture égale à 2,66 tonnes avec un écart-type égal à 0,02 tonne. Estimez la résistance moyenne au seuil de 5%.

Exercice 8

Un fabricant de piles électriques affirme que la durée de vie moyenne du matériel qu'il produit est de 170h. Un organisme de défense des consommateurs prélève au hasard un échantillon de 100 piles et observe une durée de vie moyenne de 158h avec un écart-type empirique s de 30h.

a) Déterminer un intervalle de confiance de niveau 0,99 pour la durée de vie moyenne m.

b) Peut-on accuser ce fabricant de publicité mensongère ?

Exercice 9

Dans une station-service, on suppose que le montant des chèques « essences » suit une loi normale de paramètres m et σ. On considère un échantillon de taille 50 et on obtient une moyenne de 13 000 FCFA et un écart-type de 2 800 FCFA. Donner une estimation de m par intervalle de confiance au niveau 0,95.

Corrigés des exercices du chapitre

Exercice 1

L'intervalle $\left\lfloor f_0 - t_\alpha \sigma_f \; ; \; f_0 + t_\alpha \sigma_f \right\rfloor$ est l'intervalle de confiance de la proportion p de sénégalais favorables à cette politique.

On a $\sigma_f = \sqrt{\dfrac{f_0(1-f_0)}{n}}$, $f_0 = 0{,}51$, $t_\alpha = 1{,}96$

$P\{-t_\alpha \leq N(0,1) \leq t_\alpha\} = 0{,}95$, $\alpha = 5\%$

- Par n = 100, les limites de l'intervalle de confiance sont :

$$p_1 = f_0 - t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}} = 0{,}51 - 1{,}96\sqrt{\frac{0{,}51 \times 0{,}49}{100}}$$
$$p_2 = f_0 + t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}} = 0{,}51 + 1{,}96\sqrt{\frac{0{,}51 \times 0{,}49}{100}}$$

$$p_1 = 0{,}51 - 0{,}098 = 0{,}41 \quad \text{et} \quad p_2 = 0{,}51 + 0{,}098 = 0{,}61$$

On obtient alors : $0{,}41 \leq p \leq 0{,}61$
Longueur = 0,61 – 0,41 = 0,20

- Pour n = 1000, les limites de l'intervalle de confiance deviennent :

$$p_1 = 0{,}51 - 1{,}96\sqrt{\frac{0{,}51 \times 0{,}49}{1000}} = 0{,}48$$
$$p_2 = 0{,}51 + 1{,}96\sqrt{\frac{0{,}51 \times 0{,}49}{1000}} = 0{,}54$$

D'où l'intervalle :

$$0{,}48 \leq p \leq 0{,}54$$

longueur = 0,54 – 0,48 = 0,06

- Taille de l'échantillon

La longueur de l'intervalle de confiance est : $(f_0 + t_\alpha \sigma_f) - (f_0 - t_\alpha \sigma_f) = 2t_\alpha \sigma_f$

La longueur est donc : $l_n = 2t_\alpha \sqrt{\dfrac{f_0(1-f_0)}{n}}$

Pour que $l_n \leq 0{,}04$ il faut que :

$$n \geq 2500 \; t_\alpha^2 \; f_0 \,(1 - f_0) \text{ , soit}$$
$$n \geq 2500\,(1{,}96)^2\,(0{,}51)\,(0{,}49) \text{ ,}$$
$$n \geq 2400$$

Exercice 2

Le paramètre p étant une proportion de défectueux, nous retenons ici un intervalle unilatéral de la forme $0 \leq p \leq p_2$,

avec $p_2 = f_0 + t_\alpha \sqrt{\dfrac{f_0(1-f_0)}{n}}$, n = 20 et $f_0 = \dfrac{2}{20} = 0{,}10$

t_α est lu sur la table de la loi normale : $P\ \{N(0,1) \le t_\alpha\} = 0{,}95$

$\pi(t_\alpha) = 0{,}95$, donc $t_\alpha = 1{,}65$.

$$p_2 = 0{,}10 + 1{,}65\sqrt{\frac{0{,}10 \times 0{,}90}{20}} = 0{,}21$$

L'intervalle de confiance est donc : $0 \le p \le 0{,}21$

Exercice 3

a) $n = 300$ $\quad f_0 = \frac{156}{300} = 0{,}52$

Au seuil de 5%, l'intervalle de confiance de la proportion p des ménagères qui préfèrent la nouvelle présentation est :

$$p_1 = 0{,}52 - 1{,}96\sqrt{\frac{0{,}52 \times 0{,}48}{300}} = 0{,}463$$

$$p_2 = 0{,}52 + 1{,}96\sqrt{\frac{0{,}52 \times 0{,}48}{300}} = 0{,}577$$

b) Il s'agit de déterminer n tel que :

$p_2 - p_1 \le 2\ \Delta$ (à $\pm \Delta$ % près). Soit :

$$(f_0 + t_\alpha\ \sigma_f) - (f_0 - t_\alpha\ \sigma_f) \le 2\,\Delta$$

avec $\sigma_f = \sqrt{\frac{f_0(1-f_0)}{n}}$, $t_\alpha = 1{,}96$, $f_0 = 0{,}52$ $\quad \Delta = 0{,}01$

D'où

$t_\alpha \bullet \sigma_f \le \Delta$

$$t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}} \le \Delta \text{ , soit } \quad n \ge \frac{f_0(1-f_0)t_\alpha^2}{\Delta^2}$$

Nous en déduisons :

$$n \ge \frac{1{,}96^2 \times 0{,}52 \times 0{,}48}{(0{,}01)^2}$$

n ≥ 9589 personnes.

Exercice 4

Il s'agit de construire l'intervalle de confiance de la proportion p des ménages sénégalais pauvres.

$n = 1000$; $f_o = 0{,}539$; $n_i = n \times p_i$; $t_\alpha = 1{,}96$

La marge d'erreur est : $me = t_\alpha\sqrt{\frac{f_0 \times (1-f_0)}{n}} = 0{,}0309 = 3{,}09$ %

Les bornes de l'intervalle de confiance sont :

$$p_1 = f_0 - me = 0{,}5081 \quad ; \quad n_1 = 1000 \times p_1 = 508$$

$$p_2 = f_0 + me = 0{,}5699 \quad ; \quad n_2 = 1000 \times p_2 = 570$$

Le nombre de pauvres est estimé entre 508 et 570 sur les 1000 prochaines naissances.

Exercice 5

$$n = 200 \; ; \; f_0 = \frac{84}{200} = 0{,}42 \; ; \; t_\alpha = 1{,}96$$

L'intervalle de confiance de la proportion de votes pour A est :

$$p_1 = f_0 - t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}} = 0{,}3516 \quad \text{et} \quad p_2 = f_0 + t_\alpha \sqrt{\frac{f_0(1-f_0)}{n}} = 0{,}4884$$

L'intervalle de confiance est : $0{,}3516 \leq p \leq 0{,}4884$.

Exercice 6

n = 200 $\overline{X}$ = 155 millions d'impacts s = 70 millions d'impacts

L'écart-type de la population est inconnu, nous devons utiliser la loi de Student à (n – 1) degrés de liberté. Comme $n = 200$, on approxime la loi de Student par la loi normale N(0,1), $t_\alpha = 1{,}96$. L'intervalle de confiance ayant 95 chances sur 100 de contenir la valeur vraie de la moyenne m est :

$$m_1 = \overline{X} - t_\alpha \frac{S}{\sqrt{n-1}} = 155 - 1{,}96 \, \frac{70}{\sqrt{199}}$$

$= 155 - 9{,}726 = 145{,}3$ millions d'impacts

$$m_2 = \overline{X} + t_\alpha \frac{S}{\sqrt{n-1}} = 155 + 1{,}96 \, \frac{70}{\sqrt{199}}$$

$= 155 + 9{,}726 = 164{,}73$ millions d'impacts

D'où l'intervalle :

$$145{,}3 \leq \mathrm{m} \leq 164{,}7$$

$\mathrm{n} > 30$ t_α est donc lu sur la table de la loi normale.

Exercice 7

a) L'écart-type de la population est connu, on utilise alors la loi normale centrée réduite $\frac{\sqrt{n}\left(\overline{X} - m\right)}{\sigma}$ suit une loi N (0 , 1)

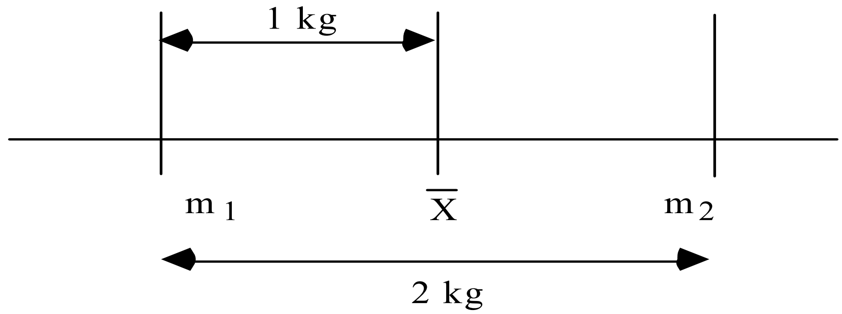
$$\mathrm{n} = 100 \;,\; \overline{X} = 2{,}63 \;,\; \sigma = 0{,}025$$

Au seuil $\alpha = 5\%$, la résistance moyenne est comprise entre les deux limites de confiance : $m_1 = \overline{X} - t_\alpha \frac{\sigma}{\sqrt{n}} = 2{,}63 - 1{,}96 \times \frac{0{,}025}{\sqrt{100}}$ et $m_2 = \overline{X} + t_\alpha \frac{\sigma}{\sqrt{n}} = 2{,}63 + 1{,}96 \times \frac{0{,}025}{\sqrt{100}}$

Soit : m_1 = 2,6251 tonnes et m_2 = 2,6349 tonnes

b) La précision de l'estimation est fournie par l'amplitude de la fourchette que constitue l'intervalle de confiance : $m_2 - m_1$.
Connaître la résistance moyenne à ± 1 kg près, implique donc :

$$m_2 - m_1 \leq 2\,kg = 0{,}002 \ \ tonne..$$



Soit :

$$\left(\overline{X} + t_\alpha \frac{\sigma}{\sqrt{n}}\right) - \left(\overline{X} - t_\alpha \frac{\sigma}{\sqrt{n}}\right) \leq 0{,}002$$

D'où

$$2\,t_\alpha \frac{\sigma}{\sqrt{n}} \leq 0{,}002$$

$$n \geq \frac{t_\alpha^2 \sigma^2}{(0{,}001)^2}$$

$n \geq 2401$

c) L'écart-type de la population σ est inconnu et n < 30 , nous devons alors utiliser la loi de Student à (n − 1) degrés de liberté.

$$n = 17 \quad , \qquad \overline{X} = 2{,}66 \ , \qquad s = 0{,}02$$

Au seuil $\alpha = 0{,}05$ et pour 16 degrés de liberté, la valeur lue est $t_\alpha = 2{,}12$.
La résistance moyenne est comprise alors entre les deux limites de confiance.

$$m_1 = \overline{X} - t_\alpha \frac{S}{\sqrt{n-1}} = 2{,}66 - 2{,}12 \times \frac{0{,}02}{\sqrt{16}}$$

$$m_2 = \overline{X} + t_\alpha \frac{S}{\sqrt{n-1}} = 2{,}66 + 2{,}12 \times \frac{0{,}02}{\sqrt{16}}$$

D'où l'intervalle :

$$2{,}6494 \ \leq \ m \ \leq \ 2{,}6706$$

Exercice 8

a) L'écart-type de la population σ est inconnu, nous devons utiliser la loi de Student à (n – 1) degrés de liberté.

$$n = 100 \quad , \quad \overline{X} = 158 \quad , \quad s = 30$$

Au seuil $\alpha = 0{,}01$, il suffit de lire t_α dans la table de la loi normale centrée réduite (car n > 30) , $t_\alpha = 2{,}576$.

D'où l'intervalle bilatéral.

$$\overline{X} - t_\alpha \frac{S}{\sqrt{n-1}} \leq m \leq \overline{X} + t_\alpha \frac{S}{\sqrt{n-1}}$$

$$158 - 2{,}576 \frac{30}{\sqrt{99}} \leq m \leq 158 + 2{,}576 \frac{30}{\sqrt{99}}$$

Soit : 150 ≤ m ≤ 166

b) Nous allons construire pour m un intervalle unilatéral à gauche (0 ≤ m ≤ a).

$$P\left\{\sqrt{n-1} \times \frac{\left(\overline{X} - m\right)}{S} \geq t_\alpha\right\} = 0{,}99$$

On lit t_α sur la table de la loi normale , P { N (0 , 1) > t_α} $= 0{,}99$ implique que P { N (0 , 1} ≤ t_α} $= 0{,}01$ donc $t_\alpha = -\,2{,}37$ car $\pi\,(2{,}37) \simeq 0{,}99$.

Déterminons la borne supérieure de l'intervalle.

$$\sqrt{n-1} \times \frac{\left(\overline{X} - m\right)}{S} \geq -2{,}37$$

D'où : $m \leq \overline{X} + 2{,}37 \dfrac{S}{\sqrt{n-1}}$, soit m ≤ 165

On a alors :

P { 0 ≤ m ≤ 165 } = 0,99

Il y a donc 99% de chance que cette publicité soit mensongère.

Exercice 9

L'écart type de la population σ est inconnu, nous devons utiliser la loi de Student à $n-1$ degrés de liberté.

$$\overline{X} = 13000\,;\ S = 2800\,;\ n = 50\,;\ t_\alpha = 1{,}96 \text{ pour } \alpha = 0{,}05$$

m est compris entre les deux limites de confiance :

$$m_1 = \overline{X} - t_\alpha \frac{S}{\sqrt{n-1}} = 12216$$

$$m_2 = \overline{X} + t_\alpha \frac{S}{\sqrt{n-1}} = 13784$$