

Série d'exercices n°2-Solutions

Exercice 1 : En 1897, l'économiste Vilfredo Pareto (1848-1923), professeur d'économie politique à l'université de Lausanne en Suisse, eut l'idée de modéliser la loi des revenus en postulant que le nombre relatif de personnes dont le revenu dépasse une valeur x est inversement proportionnel à une puissance de x . la définition suivante fut adoptée : une variable aléatoire X , suit une loi de Pareto de paramètres α et θ (on note $X \sim P(\alpha; \theta)$) si sa densité est

$$f(x) = \frac{(\alpha - 1) \theta^{\alpha-1}}{x^\alpha} \text{ pour } x \geq \theta \text{ et } f(x) = 0 \text{ ailleurs, } (\theta > 0, \alpha > 1)$$

On suppose ici que α est connu. On donne les deux caractéristiques de la loi de X ,

$$E(X) = \left(\frac{\alpha - 1}{\alpha - 2}\right) \theta \text{ pour } \alpha > 2 \text{ et } Var(X) = \frac{(\alpha - 1) \theta^2}{(\alpha - 3)(\alpha - 2)^2} \text{ pour } \alpha > 3$$

1. Déterminez l'estimateur T_n de θ par la méthode des moments.
2. Soit $(X_1, \dots, X_n)$ un échantillon de la v.a. X . Posons la statistique V_n :

$$V_n = \min_{1 \leq i \leq n} X_i$$

$$\text{On peut montrer que } E(V_n) = \left(\frac{n(\alpha - 1)}{n(\alpha - 1) - 1}\right) \theta \text{ et } Var(V_n) = \frac{n(\alpha - 1) \theta^2}{(n(\alpha - 1) - 2)(n(\alpha - 1) - 1)^2}$$

Peut-on proposer un e.s.b. de θ en se basant sur la statistique V_n ?

3. Comparez les deux estimateurs T_n et V_n .

Solution de l'exercice 1

1. On doit exprimer θ en fonction des moments de X , on a :

$$E(X) = \left(\frac{\alpha - 1}{\alpha - 2}\right) \theta \Rightarrow \theta = \left(\frac{\alpha - 2}{\alpha - 1}\right) E(X)$$

Cette dernière expression contient le moment d'ordre 1 de X . Un estimateur T_n de θ par la méthode des moments sera défini en remplaçant $E(X)$ par son estimateur la moyenne empirique $\bar{X}$,

$$T_n = \left(\frac{\alpha - 2}{\alpha - 1}\right) \bar{X} \text{ avec } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

2. On va exprimer θ en fonction de l'espérance de V_n , on aura :

$$E(V_n) = \left(\frac{n(\alpha - 1)}{n(\alpha - 1) - 1}\right) \theta \Rightarrow \theta = \left(\frac{n(\alpha - 1) - 1}{n(\alpha - 1)}\right) E(V_n) = E\left[\left(\frac{n(\alpha - 1) - 1}{n(\alpha - 1)}\right) V_n\right]$$

On peut proposer un estimateur sans biais en se basant sur V_n , en le multipliant par le coefficient $\frac{n(\alpha - 1) - 1}{n(\alpha - 1)}$.

On note V'_n cet estimateur,

$$V'_n = \left(\frac{n(\alpha - 1) - 1}{n(\alpha - 1)}\right) V_n$$

On vérifie bien que $E(V'_n) = \theta$

3. On veut comparer les deux estimateurs T_n et V_n . Le premier est un estimateur sans biais, $E(T_n) = \theta$, tandis que le deuxième est biaisé $E(V_n) - \theta \neq 0$, on doit comparer T_n avec l'estimateur corrigé de V_n c'est-à-dire comparer T_n avec V'_n .

Pour un e.s.b. le MSE coïncide avec la variance,

$$\begin{aligned} MSE(T_n) &= Var(T_n) = Var\left(\left(\frac{\alpha-2}{\alpha-1}\right)\bar{X}\right) = \left(\frac{\alpha-2}{\alpha-1}\right)^2 Var(\bar{X}) = \left(\frac{\alpha-2}{\alpha-1}\right)^2 \frac{Var(X)}{n} \\ &= \frac{1}{n} \frac{(\alpha-2)^2}{(\alpha-1)^2} \frac{(\alpha-1)\theta^2}{(\alpha-3)(\alpha-2)^2} = \frac{\theta^2}{n(\alpha-1)(\alpha-3)} \end{aligned}$$

$$MSE(T_n) = \frac{\theta^2}{n(\alpha-1)(\alpha-3)}$$

$$\begin{aligned} MSE(V'_n) &= Var(V'_n) = Var\left(\left(\frac{n(\alpha-1)-1}{n(\alpha-1)}\right)V_n\right) = \left(\frac{n(\alpha-1)-1}{n(\alpha-1)}\right)^2 Var(V_n) \\ &= \frac{(n(\alpha-1)-1)^2}{(n(\alpha-1))^2} \frac{n(\alpha-1)\theta^2}{(n(\alpha-1)-2)(n(\alpha-1)-1)^2} = \frac{\theta^2}{n(\alpha-1)(n(\alpha-1)-2)} \end{aligned}$$

$$MSE(V'_n) = \frac{\theta^2}{n(\alpha-1)(n(\alpha-1)-2)}$$

On peut vérifier que $MSE(V'_n) < MSE(T_n)$ pour $n > 1$:

$$\frac{MSE(T_n)}{MSE(V'_n)} = \frac{n(\alpha-1)(n(\alpha-1)-2)}{n(\alpha-1)(\alpha-3)} = \frac{n(\alpha-1)-2}{\alpha-3}$$

$$\frac{n(\alpha-1)-2}{\alpha-3} > 1 \Leftrightarrow n(\alpha-1)-2 > \alpha-3 \Leftrightarrow n(\alpha-1) > \alpha-1 \Leftrightarrow n > 1, \text{ on rappelle que } \alpha > 3$$

$$\frac{MSE(T_n)}{MSE(V'_n)} > 1 \Leftrightarrow MSE(V'_n) < MSE(T_n)$$

- On conclut que l'estimateur V'_n est plus efficace que l'estimateur T_n .

(On prévoit que l'estimateur V'_n va donner de meilleures estimations que l'estimateur T_n)

Remarque : Quand n est grand le $MSE(V'_n)$ est équivalent à

$$\frac{\theta^2}{n^2(\alpha-1)^2}$$

L'estimateur V'_n a une variance qui converge plus rapidement vers 0 que celle de l'estimateur T_n . En effet la vitesse de convergence de la variance de V'_n est de l'ordre $1/n^2$ tandis que celle de T_n est de l'ordre $1/n$.

Dans le cas de petits échantillons (sondage avec un coût plus réduit), on a plus de chance d'avoir une estimation correcte avec l'estimateur V'_n qu'avec l'estimateur T_n .

Rappelons que pour tout estimateur $\hat{\theta}$ de θ , on a :

$$\text{pour tout } \varepsilon > 0, \quad P(|\hat{\theta} - \theta| \geq \varepsilon) \leq \frac{MSE(\hat{\theta})}{\varepsilon^2}$$

Si $MSE(\hat{\theta}) \rightarrow 0$ quand n devient grand alors $P(|\hat{\theta} - \theta| \geq \varepsilon) \rightarrow 0$ et on aura moins de chance de trouver des valeurs de l'estimateur en dehors de l'intervalle $]\theta - \varepsilon, \theta + \varepsilon[$.

Exercice 2 : On s'intéresse à la fiabilité d'un équipement industriel. Soit X la v.a. égale au nombre de pannes de cet équipement, on va supposer que X suit une loi de Poisson $\mathcal{P}(\lambda)$, où λ est un paramètre inconnu.

$$P(X = x) = e^{-\lambda} \lambda^x / x! \quad \text{avec } x \in \mathbb{N}$$

Soit $(X_1, \dots, X_n)$ un échantillon de la v.a. X . On pose,

$$S_n = \sum_{i=1}^n X_i \quad \text{et} \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

1. Quelle est la loi de S_n . Expliquer pourquoi $\bar{X}$ est un estimateur de λ par la méthode des moments, rappeler ses propriétés.
2. Maintenant, on veut estimer la probabilité qu'il n'y ait aucune panne. On notera cette probabilité θ .
 - a. Exprimer θ en fonction de λ .
 - b. Montrer que $\hat{\theta}_n$, défini ci-dessous, est un estimateur sans biais de θ .

$$\hat{\theta}_n = \left(1 - \frac{1}{n}\right)^{S_n}$$

- c. L'observation d'un échantillon de X a donné les résultats suivants,

2	2	1	1	4	3	4	2	5	6
---	---	---	---	---	---	---	---	---	---

Donner une estimation de θ .

Solution de l'exercice 2

1. On connaît que si X et Y sont deux v.a. indépendantes de lois respectivement $P(\lambda_1)$ et $P(\lambda_2)$, avec $\lambda_1 > 0$ et $\lambda_2 > 0$ alors $X + Y \sim P(\lambda_1 + \lambda_2)$. Cela se généralise à n v.a. indépendantes de lois de Poisson, la v.a. S_n est la somme de n v.a. indépendantes de même loi $P(\lambda)$. Donc S_n suit une loi de Poisson de paramètre $\lambda + \lambda + \dots + \lambda = n\lambda$,

$$P(S_n = k) = e^{-n\lambda} \frac{(n\lambda)^k}{k!} \quad \text{avec } k \in \mathbb{N}$$

Puisqu'on a : $E(X) = \lambda$, l'estimateur de λ par la méthode des moments est construit tout simplement en remplaçant le moment $E(X)$ par son estimateur $\bar{X}$, d'où $\bar{X}$ est l'estimateur de λ par la méthode des moments. L'estimateur $\bar{X}$ a les propriétés suivantes :

$\bar{X}$ est un e.s.b de λ , $E(\bar{X}) = \lambda$

$\bar{X}$ est un estimateur consistant (convergent) de λ , $\bar{X} \xrightarrow{P} \lambda$ (loi faible des grands nombres)

$$MSE(\bar{X}) = Var(\bar{X}) = \frac{Var(X)}{n} = \frac{\lambda}{n} \rightarrow 0 \quad \text{quand } n \rightarrow +\infty$$

La loi asymptotique de $\bar{X}$ est une loi normale

$$\sqrt{n} \left(\frac{\bar{X} - \lambda}{\sqrt{\lambda}} \right) \sim N(0, 1) \quad \text{quand } n \text{ est grand}$$

On rajoutera que c'est un estimateur efficace : pour tout estimateur sans biais $\hat{\lambda}$ de λ , on vérifie que,

$$Var(\hat{\lambda}) \geq Var(\bar{X}) = \frac{\lambda}{n}$$

2. Soit θ la probabilité qu'il n'y ait aucune panne.

- a. On veut exprimer θ en fonction de λ ,

$$\theta = P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda} \Rightarrow \theta = e^{-\lambda}$$

- b. Montrons que $\hat{\theta}_n$ est un estimateur sans biais de θ . On a $S_n \sim P(n\lambda)$, $S_n \in \mathbb{N}$

$$E(\hat{\theta}_n) = E\left(\left(1 - \frac{1}{n}\right)^{S_n}\right) = \sum_{k=0}^{+\infty} \left(1 - \frac{1}{n}\right)^k P(S_n = k) = \sum_{k=0}^{+\infty} \left(1 - \frac{1}{n}\right)^k e^{-n\lambda} \frac{(n\lambda)^k}{k!}$$

$$E(\hat{\theta}_n) = e^{-n\lambda} \sum_{k=0}^{+\infty} \frac{(n\lambda - \lambda)^k}{k!} = e^{-n\lambda} e^{n\lambda - \lambda} = e^{-\lambda} = \theta$$

$$E(\hat{\theta}_n) = \theta$$

Donc $\hat{\theta}_n$ est un estimateur sans biais de θ .

Remarque : on rappelle que si X est une v.a. discrète à valeurs dans E , alors $E(g(X))$ est par définition, est égal à

$$E(g(X)) = \sum_{x \in E} g(x) P(X = x)$$

- c. Si on calcule s_n , on trouve $s_n = 30$. Alors, une estimation de θ sera

$$\hat{\theta}_n = \left(1 - \frac{1}{n}\right)^{s_n} = \left(1 - \frac{1}{10}\right)^{30} = 0.0424 = 4.24\%$$

La probabilité qu'il n'y ait aucune panne est estimée à 4.24%.

Exercice 3 : Dans une enquête, on a demandé à 1500 personnes de reconnaître leur source la plus importante d'informations ; 65% ont déclaré que l'internet était leur principale source d'informations.

1. Construire un intervalle de confiance pour la proportion de la population qui considère l'internet comme sa principale source d'information. On prendra un niveau de confiance égal à 95%.
2. De quelle taille l'échantillon devrait-il être pour estimer la proportion de la population avec une marge d'erreur de 0.02 ?

Solution de l'exercice 3

1. Soit p la proportion de la population qui considère l'internet comme sa principale source d'information.

On note $\bar{p}$ la proportion des personnes qui ont déclaré que l'internet était leur principale source d'informations parmi $n = 1500$ personnes interrogées.

Pour construire l'intervalle de confiance pour la proportion p , on va utiliser la statistique

$$\sqrt{n} \left(\frac{\bar{p} - p}{\sqrt{\bar{p}(1 - \bar{p})}} \right)$$

de loi asymptotique, la loi normale $N(0, 1)$.

$n \gg 30$, on aura une bonne approximation des probabilités.

On obtient l'intervalle de confiance $I(p)$:

$$I(p) = \left[\bar{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}; \bar{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \right]$$

Pour un niveau de confiance $1 - \alpha$, on lit dans la table de la loi $N(0,1)$ la valeur du quantile d'ordre

$$1 - \frac{\alpha}{2}$$

A.N.

$$1 - \alpha = 95\%, \quad 1 - \frac{\alpha}{2} = 97.5\%, \quad z_{97.5\%} = 1.96, \quad \bar{p} = 0.65$$

$$I(p) = \left[0.65 \pm 1.96 \sqrt{\frac{0.65 * 0.35}{1500}} \right] = [0.65 \pm 0.02414] = [62.59\%, 67.41\%]$$

$$I(p) = [62.59\%, 67.41\%]$$

2. La marge d'erreur ne doit pas dépasser

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} = 0.02$$

On garde le même niveau de confiance qui est de 95%, la valeur de p sera remplacé par la valeur de $\bar{p}$ calculée lors du sondage précédent c'est-à-dire qu'on la remplacera par $\bar{p} = 0.65$

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 1.96 \sqrt{\frac{0.65 * 0.35}{n}} = 0.02 \Rightarrow n = (46.743)^2 = 2184.91$$

On prendra

$$n = 2185$$

Si on veut réaliser un deuxième sondage pour estimer la proportion p de la population avec une marge d'erreur de 0.02, on doit sélectionner aléatoirement 2185 personnes pour former l'échantillon.

Exercice 4 : Les données d'échantillon ESAR suivantes sont issues d'une population de moyenne μ et de variance σ^2 inconnues :

14	13	11	12	9	13	13	11	10	12
----	----	----	----	---	----	----	----	----	----

On donne $\sum_{i=1}^{10} x_i = 118$ et $\sum_{i=1}^{10} (x_i - \bar{x}_{10})^2 = 21.6$

1. Donner une estimation de μ et de σ^2 .
2. Calculez l'intervalle de confiance de μ au niveau de confiance 95%.
3. Calculez l'intervalle de confiance de l'écart-type σ au niveau de confiance 95%.
4. Si on suppose que l'écart-type est connu égal à 2, que devient l'intervalle de confiance de μ au même niveau de confiance 95%.

Solution de l'exercice 4

1. L'estimateur de μ est $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$, une estimation de μ sera

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{118}{10} = 11.8$$

L'estimateur de σ^2 est $S^2 = \frac{1}{9} \sum_{i=1}^{10} (X_i - \bar{X})^2$, une estimation de σ^2 sera

$$s^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2 = \frac{21.6}{9} = 2.4$$

2. La statistique utilisée est

$$\sqrt{n} \left(\frac{\bar{X} - \mu}{S} \right)$$

On supposera pour ce petit échantillon ($n = 10$) que la loi est normale, la loi de la statistique ci-dessus est une Student à $n - 1$ degrés de liberté $t(n - 1)$.

L'intervalle de confiance de μ au niveau de confiance $1 - \alpha = 95\%$,

$$I(\mu) = \left[\bar{X} - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$$

avec $t_{n-1, 1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student $t(n - 1)$.

A.N.

$$t_{n-1, 1-\frac{\alpha}{2}} = t_{9, 97.5\%} = 2.262$$

$$I(\mu) = \left[\bar{x} \pm t_{n-1, 1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right] = \left[11.8 \pm 2.262 \frac{\sqrt{2.4}}{\sqrt{10}} \right] = [11.8 \pm 1.108] = [10.69, 12.91]$$

$$I(\mu) = [10.69, 12.91]$$

3. On utilisera la statistique suivante,

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

de loi de khideux à $n - 1$ degrés de liberté $\chi^2(n - 1)$

On va calculer en premier l'intervalle de confiance de la variance σ^2 puis on va déduire celui de l'écart-type σ .

L'intervalle de confiance de σ^2 au niveau de confiance $1 - \alpha = 95\%$ sera,

$$I(\sigma^2) = \left[\left(\frac{n-1}{b} \right) S^2, \left(\frac{n-1}{a} \right) S^2 \right]$$

Avec a le quantile d'ordre $\alpha/2$ et b le quantile d'ordre $1 - \alpha/2$ de la loi de khideux $\chi^2(n - 1)$. On note

$$a = \chi_{n-1, \alpha/2}^2 \text{ et } b = \chi_{n-1, 1-\alpha/2}^2.$$

A.N.

$$1 - \alpha = 95\%, \quad \frac{\alpha}{2} = 2.5\%, \quad 1 - \frac{\alpha}{2} = 97.5\%, \quad \chi_{n-1, \alpha/2}^2 = \chi_{9, 2.5\%}^2 = 2.7,$$

$$\chi_{n-1, 1-\frac{\alpha}{2}}^2 = \chi_{9, 97.5\%}^2 = 19.02$$

$$I(\sigma^2) = \left[\frac{21.6}{19.02}, \frac{21.6}{2.7} \right] = [1.1356, 8]$$

On déduit l'intervalle de confiance de σ ,

$$I(\sigma) = [\sqrt{1.1356}, \sqrt{8}] = [1.066, 2.83]$$

4. Supposons dans cette question que l'écart-type de la population est connu, $\sigma = 2$. Pour déterminer l'intervalle de confiance de μ , on va utiliser la statistique suivante,

$$\sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right)$$

Cette statistique est de loi normale $N(0, 1)$.

L'intervalle de confiance de μ au niveau de confiance $1 - \alpha = 95\%$,

$$I(\mu) = \left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

avec $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi $N(0,1)$.

A.N.

$$z_{1-\frac{\alpha}{2}} = z_{97.5\%} = 1.96, \sigma = 2,$$

$$I(\mu) = \left[11.8 \pm 1.96 \frac{2}{\sqrt{10}} \right] = [11.8 \pm 1.2396] = [10.56, 13.04]$$

$$I(\mu) = [10.56, 13.04]$$

Exercice 5 : Dans une élection, on s'intéresse à la proportion p des votants qui favorisent le candidat A. On désire estimer p en utilisant un intervalle de niveau de confiance 95%.

1. Quel le nombre minimal d'électeurs qu'on doit observer pour obtenir une erreur d'estimation de $\pm 1\%$?
Un premier sondage sur seulement 100 électeurs a révélé que 23 sont favorables pour le candidat A.
2. Quel est l'intervalle de confiance de p au niveau de confiance 95%.
3. Si on veut réaliser un deuxième sondage, quel est le nombre minimal d'électeurs qu'on doit choisir afin de garantir toujours une erreur d'estimation de $\pm 1\%$.

Solution de l'exercice 5

1. La proportion p sera estimée par $\bar{p} = \frac{1}{n} \sum_{i=1}^n X_i$ avec $X_i \sim B(p)$

$E(\bar{p}) = p$ et $Var(\bar{p}) = p(1-p)/n$, la loi asymptotique de $\bar{p}$ est une loi normale et on a :

$$\sqrt{n} \left(\frac{\bar{p} - p}{\sqrt{p(1-p)}} \right) \sim N(0, 1) \text{ quand } n \text{ est grand}$$

L'intervalle de confiance de la proportion p est

$$I(p) = \left[\bar{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right]$$

L'erreur d'estimation en valeur absolue est égale 1%, on doit vérifier

$$z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} = 0.01$$

On peut exprimer la taille n en fonction des autres quantités,

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{0.01} \right)^2 p(1-p)$$

Dans cette équation la valeur de p est inconnue, on doit la remplacer par une valeur arbitraire $p \in]0, 1[$ et on prendra la valeur centrale $p = 0.5$, cette valeur maximise la fonction $p \mapsto p(1 - p)$ sur $[0, 1]$.

A.N.

$$z_{1-\frac{\alpha}{2}} = z_{97.5\%} = 1.96$$

$$n = \left(\frac{1.96}{0.01}\right)^2 0.5^2 = 9604$$

On a besoin d'au moins 9604 électeurs qu'on doit observer pour obtenir une erreur d'estimation de $\pm 1\%$.

2. Une estimation de p est $23/100 = 0.23 = 23\%$

L'intervalle de confiance de p au niveau de confiance 95% est

$$I(p) = \left[\bar{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \right]$$

A.N.

$$I(p) = \left[0.23 \pm 1.96 \sqrt{\frac{0.23 * 0.77}{100}} \right] = [0.23 \pm 0.0825]$$

$$I(p) = [14.75\%, 31.25\%]$$

3. L'erreur d'estimation dans le premier sondage est 8.25%. On veut diminuer cette erreur en augmentant la taille n . Dans le choix de la valeur arbitraire de p , on utilisera l'information contenue dans le 1^{er} sondage, à savoir la première estimation de p , c'est-à-dire $\bar{p} = 0.23$.

On cherchera n tel que,

$$1.96 \sqrt{\frac{0.23 * 0.77}{n}} = 0.01$$

C'est-à-dire

$$n = \left(\frac{1.96}{0.01}\right)^2 * 0.23 * 0.77 = 6803.5$$

On doit prendre une taille d'échantillon $n = 6804$ pour garantir une incertitude de l'estimation de l'ordre de 1%.

Exercice 6 : Soit une population normale de moyenne μ inconnue et d'écart-type $\sigma = 1.5$. On y a prélevé un échantillon aléatoire simple d'effectif $n = 9$. Les observations sont les suivantes :

24.3	24.7	23.2	23.8	24.5	26.0	25.4	24.8	23.8
------	------	------	------	------	------	------	------	------

1. Trouvez un intervalle de confiance pour μ . Le niveau de signification est égal à 5%.
2. Déduisez les erreurs absolue et relative de l'estimation.
3. Supposons que σ est inconnue, trouvez un intervalle de confiance pour μ .

Solution de l'exercice 6

1. Une estimation de μ sera

$$\frac{1}{9} \sum_{i=1}^9 x_i = \frac{220.5}{9} = 24.5$$

L'intervalle de confiance de μ au niveau de signification égal à $\alpha = 5\%$.

$$I(\mu) = \left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

A.N.

$$z_{1-\frac{\alpha}{2}} = z_{97.5\%} = 1.96$$

$$I(\mu) = \left[24.5 \pm 1.96 * \frac{1.5}{3} \right] = [24.5 \pm 0.98] = [23.52, 25.48]$$

$$I(\mu) = [23.52, 25.48]$$

2. L'erreur absolue est l'erreur maximale commise en estimant μ par $\bar{X}$, elle est exprimée en unités de X , $|\bar{X} - \mu|$. L'erreur relative est le rapport entre l'erreur absolue et la moyenne,

$$\frac{|\bar{X} - \mu|}{\mu}$$

Dans l'erreur relative, l'effet de l'unité est éliminé, et pour le calcul on remplace dans le dénominateur la moyenne μ par son estimateur $\bar{X}$

$$\frac{|\bar{X} - \mu|}{\bar{X}}$$

L'erreur relative présente l'importance de la grandeur de l'erreur absolue par rapport à la grandeur du paramètre à estimer, elle est exprimée en général en pourcentage %.

A.N.

$$L'erreur absolue = |\bar{x} - \mu| = 0.98$$

$$L'erreur relative = \frac{|\bar{x} - \mu|}{\bar{x}} = \frac{0.98}{24.5} = 0.04 = 4\%$$

La moyenne μ est égale 24.5 ± 0.98 , cette dernière écriture peut être remplacée par une autre écriture utilisant l'erreur relative, $24.5 \pm 4\%$.

L'erreur relative nous permet aussi de comparer deux d'estimations de μ . Si on obtient deux estimations de μ à partir de deux échantillons différents et de même taille n , la meilleure estimation sera celle qui fournit l'erreur relative la plus petite.

3. L'intervalle de confiance de μ est

$$I(\mu) = \left[\bar{X} \pm t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$$

A.N.

$$s^2 = \frac{1}{8} \sum_{i=1}^9 (x_i - \bar{x})^2 = \frac{5.9}{8} = 0.7375 \Rightarrow s = 0.8588$$

$$t_{n-1, 1-\frac{\alpha}{2}} = t_{8, 97.5\%} = 2.306 \text{ (lecture dans la table de Student)}$$

$$I(\mu) = \left[24.5 \pm 2.306 * \frac{0.8588}{3} \right] = [24.5 \pm 0.66] = [23.84, 25.16]$$

Exercice 7 : Un comptable pense que les problèmes de liquidité d'une entreprise sont une conséquence directe de l'encaissement lent des comptes fournisseurs. Le comptable prétend qu'au moins 70% des comptes fournisseurs datent de plus de deux mois. Un échantillon de 120 comptes fournisseurs a révélé que 88 dataient de plus de deux mois.

1. Quelle est l'estimation ponctuelle de la proportion p de comptes fournisseurs datant de plus de deux mois. Ensuite, donnez un intervalle de confiance à 99% pour p .
2. Déterminez la taille minimum n de l'échantillon à prélever pour que l'erreur absolue de l'estimation (au niveau de confiance 99%) soit inférieure à 3%.

Solution de l'exercice 7

1. Une estimation ponctuelle de p sera,

$$\bar{p} = \frac{88}{120} = 0.7333 = 73.33\%$$

L'intervalle de confiance de p est

$$I(p) = \left[\bar{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \right]$$

A.N.

$$n = 120, z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.01}{2}} = z_{99.5\%} = 2.57$$

$$I(p) = \left[73.33\% \pm 2.57 \sqrt{\frac{73.33\% * 26.67\%}{120}} \right] = [73.33\% \pm 10.375\%]$$

$$I(p) = [62.96\%, 83.70\%]$$

2. L'erreur absolue de l'estimation de l'intervalle de confiance précédent est égale $10.37\% > 3\%$.

On doit chercher à augmenter la taille n pour atteindre l'incertitude de 3%. Quelle est la valeur de n telle qu'on a :

$$\left| z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \right| \leq 3\%$$

Ce qui donne

$$n \geq \left(\frac{z_{1-\frac{\alpha}{2}}}{0.03} \right)^2 * \bar{p}(1-\bar{p})$$

A.N.

$$n \geq \left(\frac{2.57}{0.03} \right)^2 * 73.33\% * 26.67\% = 1433.75$$

$$\Leftrightarrow n \geq 1434$$

La taille minimale à garantir pour que l'erreur absolue soit inférieure ou égale à 3% est

$$n = 1434$$

Exercice 8 : Une entreprise mène une politique de formation du personnel. Au cours de l'année l'ensemble de ces 200 employés ont suivi une même formation intitulée «FC». Le directeur s'intéresse à l'ancienneté de ces employés ainsi qu'à leur satisfaction de la FC.

- I. Le directeur a demandé à un nombre n de ces employés choisis au hasard avec remise (ESAR) leurs anciennetés dans l'entreprise : la moyenne de l'échantillon est $\bar{x} = 68$ mois et l'écart-type échantillon $s = 24$ mois. Déterminer les intervalles de confiance à 90% de l'ancienneté moyenne des employés dans les cas suivants $n = 9$ et $n = 50$.
- II. Le directeur a demandé à 50 de ses employés s'ils étaient satisfaits de la FC ; 40 se sont déclarés satisfaits.
 1. Supposons que les employés sont choisis au hasard et avec remise. Déterminer l'intervalle de confiance à 90% de la proportion des employés satisfaits de la formation.
 2. Supposons maintenant que les employés sont choisis au hasard et sans remise.
 - a. Dans ces conditions peut-on affirmer que L'ESSR donnera approximativement les mêmes résultats que l'ESAR? pourquoi?
 - b. Proposer un intervalle de confiance à 90% de la proportion des employés satisfaits de la formation.
 - c. Déterminer la taille minimum n de l'échantillon nécessaire afin que l'erreur absolue de l'estimation (toujours à 90%) ne dépasse pas 5%.

Solution de l'exercice 8

I. Posons μ = ancienneté moyenne des 200 employés.

On sélectionne n parmi 200, la moyenne observée est $\bar{x} = 68$ mois et l'écart-type observé est

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 24 \text{ mois}$$

L'intervalle de confiance de μ est,

$$I(\mu) = \left[\bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$$

Le $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ d'une loi non encore déterminée. L'échantillon prélevé est un ESAR, donc on ne va pas apporter de correction pour la variance.

- Cas où $n = 9$, on supposera ici la loi de X celle d'une loi normale $N(\mu, \sigma)$, on a

$$\sqrt{n} \left(\frac{\bar{X} - \mu}{S} \right) \sim t(n-1)$$

A.N.

$$n = 9, \quad 1 - \alpha = 90\%, \quad z_{1-\frac{\alpha}{2}} = t_{n-1, 1-\frac{\alpha}{2}} = t_{8, 95\%} = 1.86$$

$$I(\mu) = \left[68 \pm 1.86 \frac{24}{3} \right] = [68 \pm 14.88]$$

$$I(\mu) = [53.12, 82.88]$$

- Cas où $n = 50$, on a $n > 30$, on utilise le TCL

$$\sqrt{n} \left(\frac{\bar{X} - \mu}{S} \right) \sim N(0, 1) \text{ pour } n \text{ assez grand}$$

A.N.

$$n = 50, \quad 1 - \alpha = 90\%, \quad z_{1-\frac{\alpha}{2}} = z_{95\%} = 1.65$$

$$I(\mu) = \left[68 \pm 1.65 \frac{24}{\sqrt{50}} \right] = [68 \pm 5.60]$$

$$I(\mu) = [62.4, 73.6]$$

Plus n augmente, plus l'amplitude de $I(\mu)$ diminue, on aura ainsi plus de précision dans l'estimation de μ .

II. Posons p la proportion des employés satisfaits de la formation.

1. Soit la v.a. $X = \begin{cases} 1 & \text{si l'employé est satisfait de la FC} \\ 0 & \text{sinon} \end{cases}$

$$P(X = 1) = p; X \sim B(p) \text{ et } Y = \sum_{i=1}^{50} X_i \sim B(50, p)$$

$$\bar{p} = \frac{Y}{50} \text{ est un estimateur de } p.$$

$n = 50 > 30$, on utilise le TCL,

$$I(p) = \left[\bar{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \right]$$

A.N.

$$n = 50, \quad \bar{p} = \frac{40}{50} = 0.8 = 80\%, \quad z_{1-\frac{\alpha}{2}} = z_{95\%} = 1.65$$

$$I(p) = \left[80\% \pm 1.65 \sqrt{\frac{80\% * 20\%}{50}} \right] = [80\% \pm 9.33\%]$$

$$I(p) = [70.67\%, 89.33\%]$$

2. L'échantillon est un ESSR,

a. On calcule le taux de sondage,

$$\frac{n}{N} = \frac{50}{200} = 25\% > 5\%$$

L'ESSR ne donnera pas les mêmes résultats que l'ESAR puis que le taux de sondage dépasse le seuil 5%. On doit corriger la variance de $\bar{p}$,

$$Var(\bar{p}) = \left(1 - \frac{n}{N} \right) \frac{p(1-p)}{n}$$

b. L'intervalle de confiance de p sera,

$$I(p) = \left[\bar{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\left(1 - \frac{n}{N}\right) \frac{\bar{p}(1-\bar{p})}{n}} \right]$$

A.N.

$$I(p) = \left[80\% \pm 1.65 \sqrt{0.75 * \frac{0.8 * 0.2}{50}} \right] = [80\% \pm 8.08\%]$$

$$I(p) = [71.92\%, 88.08\%]$$

c. On doit augmenter la taille n pour que l'erreur absolue

$$z_{1-\frac{\alpha}{2}} \sqrt{\left(1 - \frac{n}{N}\right) \frac{\bar{p}(1-\bar{p})}{n}}$$

soit inférieure ou égale à 5%.

A.N.

$$\begin{aligned} 1.65 \sqrt{\left(1 - \frac{n}{200}\right) \frac{0.8 * 0.2}{n}} &\leq 0.05 \Leftrightarrow 1.65^2 * \left(1 - \frac{n}{200}\right) * 0.16 \leq 0.0025 * n \\ &\Leftrightarrow 0.4356 - 0.002178 * n \leq 0.0025 * n \\ &\Leftrightarrow 0.4356 \leq 0.004678 * n \\ &\Leftrightarrow n \geq 93.12 \end{aligned}$$

La taille minimale sera $n = 94$.

On a besoin d'une taille de l'échantillon minimale de 94 employés, afin de garantir que l'erreur absolue de l'estimation ne dépasse pas 5%.