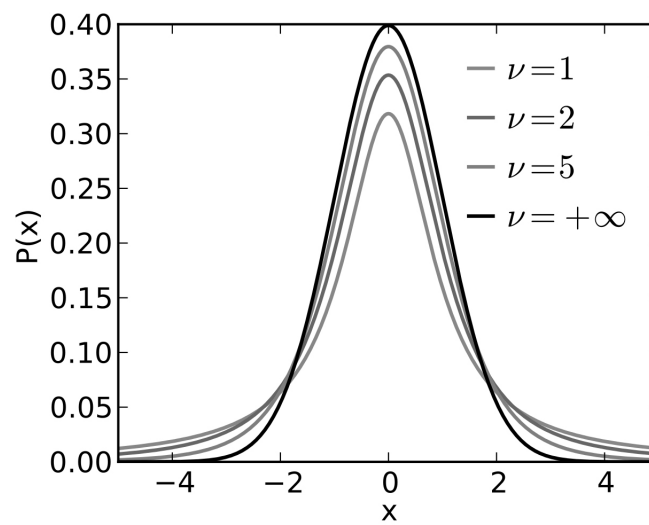




M3.5.1 Statistiques

Corrigé TD2



v0.1

24 mars 2017

Exercice 1. Un commerçant propose à sa clientèle six articles électroménagers. Considérons la population mère constituée par ces six articles codés w_i ($i \in 1, \dots, 6$). Soit X la variable qui représente “le nombre d’unités en stock de chaque article au moment de l’inventaire”.

Ω	w_1	w_2	w_3	w_4	w_5	w_6
X	0	1	2	3	0	1

- Déterminer la loi de probabilité de la variable aléatoire X . Calculer la moyenne $\mu = E(X)$ et la variance $\sigma^2 = Var(X)$, ainsi que σ^4 et le moment centré $\mu_4 = E((X - \mu)^4)$.
- Dans cette population d’effectif $N = 6$, on tire avec remise des échantillons de taille $n = 2$. X_1 est le nombre d’unités en stock pour le premier article tiré et X_2 est le nombre d’unités en stock pour le second. On pose

$$\bar{X} = \frac{1}{2}(X_1 + X_2).$$

Déterminer les valeurs prises par \bar{X} sur tous les échantillons possibles, en déduire sa loi, calculer l’espérance empirique $E(\bar{X})$ et la variance empirique $Var(\bar{X})$. Vérifier les résultats théoriques du cours.

- On considère la statistique S^2 définie par

$$S^2 = \frac{1}{2} \sum_{i=1}^2 (X_i - \bar{X})^2 = \frac{1}{2}(X_1^2 + X_2^2) - \bar{X}^2$$

Déterminer les valeurs prises par S^2 sur tous les échantillons de taille $n = 2$; en déduire sa loi, son espérance et sa variance.

Exercice 2. Dans un pays, les statistiques font ressortir que 64% des ménages possèdent une voiture de tourisme.

Quelle est la probabilité que sur un échantillon au hasard de 225 ménages, la proportion de ceux qui possèdent une voiture soit :

- comprise entre 40% et 70%.
- supérieure à 60%.
- inférieure à 25%.

Exercice 3. Un dispositif de signalisation lumineuse comporte trois lampes ; celle qui est en service est relayée automatiquement en cas de défaillance.

Quelle est la probabilité que l’ensemble fonctionne :

- plus de 5000 heures.
- mois de 4200 heures.

On sait que la durée de vie des lampes utilisées est une variable normale de moyenne 1500 heures et d’écart-type 150 heures.

Exercice 4. Un paquet de tabac produit par la régie des Tabac a un poids moyen de 50 g et un écart-type de 2 g . En supposant que ce tabac soit livré par lots de mille paquets ; quelle est la probabilité que la différence $A - B$ entre les poids de deux lots A et B excède 200 g ?

Exercice 5. Lors d'un concours radiophonique, on note X le nombre de réponses reçues chaque jour. On suppose que X suit une loi normale de paramètres μ et σ . Durant les 10 jours, on a obtenu

$$\begin{aligned}x_1 &= 200, & x_2 &= 240, & x_3 &= 190, & x_4 &= 150, & x_5 &= 220, \\x_6 &= 180, & x_7 &= 170, & x_8 &= 230, & x_9 &= 210, & x_{10} &= 210\end{aligned}$$

Donner une estimation ponctuelle de μ et σ^2 .

Exercice 6. Soit (X_1, X_2, \dots, X_n) un n -échantillon de loi de Bernoulli $\mathcal{B}(p)$ et soit

$$S_n = \sum_{k=1}^n X_k, \quad \bar{X}_n = \frac{S_n}{n}$$

- (a) Montrer que \bar{X}_n est un estimateur sans biais de p .
- (b) On cherche à estimer la variance $\sigma^2 = p(1 - p)$. On propose l'estimateur

$$U_n = \bar{X}_n(1 - \bar{X}_n)$$

- i. Calculer l'espérance et la variance de S_n .
- ii. Montrer que U_n est un estimateur biaisé de σ^2 .
- iii. Donner un estimateur V_n sans biais de σ^2 , fonction de U_n .

Exercice 7. À la veille d'une consultation électorale, on a interrogé cent électeurs constituant un échantillon au hasard. 54 d'entre eux ont déclaré avoir l'intention de voter pour le candidat Y .

Entre quelle limite, au moment du sondage, avec une probabilité de 0.95, la proportion du corps électoral favorable au candidat Y se situe-t-elle ?

Exercice 8. Dans un centre de recrutement, on a mesuré la taille de 400 conscrits. Pour cet échantillon pris au hasard, la taille moyenne \bar{x} est égale à 172 cm et l'écart-type estimé s est égal à 4 cm .

Construire un intervalle qui contienne avec une probabilité 0.99 la taille moyenne de l'ensemble des conscrits de ce centre de recrutement.

Exercice 9. Si l'écart-type de la durée de vie d'un modèle de lampe électronique est estimé à 100 heures , quelle doit être la taille de l'échantillon à prélever pour que l'erreur sur l'estimation de la durée de vie moyenne n'excède pas 20 heures et avec une probabilité

- (a) de 95%.
- (b) de 99%.

Exercice 10. Une machine fabrique des rondelles en série. Le diamètre d est une variable gaussienne dont l'écart-type est égale à 1 millimètre. On prélève au hasard un échantillon de neuf rondelles. Les mesures des diamètres, en millimètre, sont les suivantes :

20.1; 19.9; 20.0; 19.8; 19.7; 20.2; 20.1; 23.1; 22.8

- (a) Construire un intervalle qui contienne avec une probabilité de 0.95 la moyenne m de d .
- (b) Même question en supposant que l'écart-type de d a une valeur inconnue.
- (c) En utilisant les données numériques de (a), l'écart-type de d étant supposé inconnu, construire un intervalle qui contienne avec une probabilité de 0.9 la variance de d .

Exercice 11. Le chiffre d'affaire mensuel de l'entreprise TEX suit une loi normale de moyenne μ inconnue, mais dont l'écart-type est évalué à 50 M Dh. Sur les 12 derniers mois, la moyenne des chiffres d'affaire mensuels a été de 200 M Dh.

Donner une estimation de μ par intervalle de confiance au niveau 0.98.

Exercice 12. Dans une station-service, on suppose que le montant des chèques "essence" suit une loi normale de paramètres μ et σ . On considère un échantillon de taille $n = 50$ et on obtient une moyenne de 130 Dh et un écart-type de 28 Dh.

Donner une estimation de μ par intervalle de confiance au niveau 0.95.

Exercice 13. Les salaires mensuels des employés d'une entreprise sont supposés suivre une loi normale de paramètres μ et σ .

- (a) Pour un échantillon de taille $n = 10$, on obtient une moyenne $m = 6500$ Dh et un écart-type $s = 900$ Dh. Donner un intervalle de confiance au niveau 0.95 pour μ .
- (b) Pour un échantillon de taille $n = 100$, on obtient une moyenne $m = 6200$ Dh et un écart-type $s = 850$ Dh. Donner un intervalle de confiance au niveau 0.95 pour μ .

Exercice 14. On se propose d'étudier le corps électoral d'une région.

- (a) 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) Avec un second échantillon de 100 personnes, on a obtenu 45 intentions de vote pour A. En réunissant les deux échantillons, donner un intervalle de confiance pour p au niveau 95%.
- (c) Déterminer la taille n de l'échantillon qui permet d'obtenir un intervalle de confiance de largeur 0.02 sachant qu'une estimation ponctuelle de p a donné la valeur 0.4.
- (d) Pour mieux cerner la population votant pour A, on extrait un échantillon de n personnes ayant l'intention de voter pour A. On suppose que l'âge des individus suit une loi normale de paramètres μ et σ .

- i. Le premier échantillon de taille $n = 15$ a donné une moyenne d'âge de 45 ans et un écart-type de 10 ans. Donner un intervalle de confiance pour μ au niveau 98%.
- ii. Le second échantillon de taille $n = 100$ a donné une moyenne d'âge de 47 ans et un écart-type de 9 ans. Donner un intervalle de confiance pour μ au niveau 98%.
- iii. En gardant l'échantillon de taille 100, donner un intervalle de confiance pour σ^2 au niveau 95%.

Exercice 1. Un commerçant propose à sa clientèle six articles électroménagers. Considérons la population mère constituée par ces six articles codés w_i ($i \in 1, \dots, 6$). Soit X la variable qui représente “le nombre d’unités en stock de chaque article au moment de l’inventaire”.

Ω	w_1	w_2	w_3	w_4	w_5	w_6
X	0	1	2	3	0	1

- (a) Déterminer la loi de probabilité de la variable aléatoire X . Calculer la moyenne $\mu = E(X)$ et la variance $\sigma^2 = Var(X)$, ainsi que σ^4 et le moment centré $\mu_4 = E((X - \mu)^4)$.

Solution.

La loi de probabilité de la variable aléatoire X est

k	0	1	2	3
$P(X = k)$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

La moyenne est

$$\mu = E(X) = \sum_{k=0}^3 k P(X = k)$$

A.N. :

$$\mu = \frac{7}{6}$$

La variance est

$$\sigma^2 = Var(X) = E(X^2) - E(X)^2$$

On a $E(X^2) = \sum_{k=0}^3 k^2 P(X = k)$. A.N. : $E(X^2) = \frac{15}{6}$, donc

$$\sigma^2 = \frac{41}{36} \quad \text{et} \quad \sigma^4 = \frac{1681}{1296}$$

On a

$$\mu_4 = E((X - \mu)^4) = \sum_{k=0}^3 (k - \mu)^4 P(X = k)$$

k	$k - \mu$	$(k - \mu)^4$
0	$-\frac{7}{6}$	$\frac{2401}{1296}$
1	$-\frac{1}{6}$	$\frac{1}{1296}$
2	$\frac{5}{6}$	$\frac{625}{1296}$
3	$\frac{11}{6}$	$\frac{14641}{1296}$

A.N. :

$$\mu_4 = \frac{3345}{1296}$$

- (b) Dans cette population d'effectif $N = 6$, on tire avec remise des échantillons de taille $n = 2$. X_1 est le nombre d'unités en stock pour le premier article tiré et X_2 est le nombre d'unités en stock pour le second. On pose

$$\bar{X} = \frac{1}{2}(X_1 + X_2).$$

Déterminer les valeurs prises par \bar{X} sur tous les échantillons possibles, en déduire sa loi, calculer l'espérance empirique $E(\bar{X})$ et la variance empirique $Var(\bar{X})$. Vérifier les résultats théoriques du cours.

Solution.

Les valeurs prises par \bar{X} sur tous les échantillons possibles sont

$X_1 \backslash X_2$	w_1	w_2	w_3	w_4	w_5	w_6
w_1	0	$\frac{1}{2}$	1	$\frac{3}{2}$	0	$\frac{1}{2}$
w_2	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1
w_3	1	$\frac{3}{2}$	2	$\frac{5}{2}$	1	$\frac{3}{2}$
w_4	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{3}{2}$	2
w_5	0	$\frac{1}{2}$	1	$\frac{3}{2}$	0	$\frac{1}{2}$
w_6	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{1}{2}$	1

La loi de \bar{X} est donc

x_i	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3
$P(\bar{X} = x_i)$	$\frac{4}{36}$	$\frac{8}{36}$	$\frac{8}{36}$	$\frac{8}{36}$	$\frac{5}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

L'espérance empirique est

$$E(\bar{X}) = \sum_i x_i P(\bar{X} = x_i)$$

A.N. :

$$E(\bar{X}) = \frac{42}{36} = \frac{7}{6}$$

et on a bien $E(\bar{X}) = E(X)$.

La variance empirique est

$$Var(\bar{X}) = E(\bar{X}^2) - E(\bar{X})^2$$

On a $E(\bar{X}^2) = \sum_i x_i^2 P(\bar{X} = x_i)$. A.N. : $E(\bar{X}^2) = \frac{69.5}{36}$, donc

$$Var(\bar{X}) = \frac{20.5}{36}$$

et on a bien $Var(\bar{X}) = \frac{\sigma^2}{n}$ avec $n = 2$.

(c) On considère la statistique S^2 définie par

$$S^2 = \frac{1}{2} \sum_{i=1}^2 (X_i - \bar{X})^2 = \frac{1}{2} (X_1^2 + X_2^2) - \bar{X}^2$$

Déterminer les valeurs prises par S^2 sur tous les échantillons de taille $n = 2$; en déduire sa loi, son espérance et sa variance.

Solution.

Les valeurs prises par S^2 sur tous les échantillons de taille $n = 2$ sont

$X_1 \backslash X_2$	w_1	w_2	w_3	w_4	w_5	w_6
w_1	0	$\frac{1}{4}$	1	$\frac{9}{4}$	0	$\frac{1}{4}$
w_2	$\frac{1}{4}$	0	$\frac{1}{4}$	1	$\frac{1}{4}$	0
w_3	1	$\frac{1}{4}$	0	$\frac{1}{4}$	1	$\frac{1}{4}$
w_4	$\frac{9}{4}$	1	$\frac{1}{4}$	0	$\frac{9}{4}$	1
w_5	0	$\frac{1}{4}$	1	$\frac{9}{4}$	0	$\frac{1}{4}$
w_6	$\frac{1}{4}$	0	$\frac{1}{4}$	1	$\frac{1}{4}$	0

La loi de S^2 est donc

x_i	0	$\frac{1}{4}$	1	$\frac{9}{4}$
$P(S^2 = x_i)$	$\frac{10}{36}$	$\frac{14}{36}$	$\frac{8}{36}$	$\frac{4}{36}$

L'espérance empirique est

$$E(S^2) = \sum_i x_i P(S^2 = x_i)$$

A.N. :

$$E(S^2) = \frac{20.5}{36}$$

et on a bien $E(S^2) = \frac{n-1}{n} \sigma^2$ avec $n = 2$.

La variance empirique est

$$Var(S^2) = E(S^4) - E(S^2)^2$$

On a $E(S^4) = \sum_i x_i^2 P(\bar{X} = x_i)$. A.N. : $E(S^4) = \frac{29.125}{36}$, donc

$$Var(S^2) = \frac{628.25}{1296}$$

et on a bien $Var(S^2) = \frac{n-1}{n^3} ((n-1)\mu_4 - (n-3)\sigma^4)$ avec $n = 2$.

Exercice 2. Dans un pays, les statistiques font ressortir que 64% des ménages possèdent une voiture de tourisme.

Quelle est la probabilité que sur un échantillon au hasard de 225 ménages, la proportion de ceux qui possèdent une voiture soit :

(a) comprise entre 40% et 70%.

Solution.

Soit F_n la proportion des ménages qui possèdent une voiture de tourisme dans l'échantillon. On a

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

avec $X_i \hookrightarrow \mathcal{B}(p)$, $\forall 1 \leq i \leq n = 225$ et $p = 0.64$:

$$X_i = \begin{cases} 1 & \text{si le } i^{\text{e}} \text{ ménage possède une voiture de tourisme} \\ 0 & \text{sinon} \end{cases}$$

On pose

$$U = \frac{F_n - E(F_n)}{\sqrt{V(F_n)}} = \sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}}$$

Les X_i sont supposés indépendantes, et on a $n \geq 30$, donc avec le TCL, U suit approximativement la loi normale centrée réduite : $U \hookrightarrow \mathcal{N}(0, 1)$.

On a

$$\begin{aligned} P(0.4 \leq F_n \leq 0.7) &= P\left(\sqrt{n} \frac{0.4 - p}{\sqrt{p(1-p)}} \leq \sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}} \leq \sqrt{n} \frac{0.7 - p}{\sqrt{p(1-p)}}\right) \\ &= P(a \leq U \leq b) \\ &= \Phi(b) - \Phi(a) \end{aligned}$$

avec Φ la fonction de répartition de la loi normale centrée réduite, et

$$a = \sqrt{225} \frac{0.4 - 0.64}{\sqrt{0.64(1 - 0.64)}} = -7.5 \text{ et } \Phi(-7.5) \simeq 0$$

$$b = \sqrt{n} \frac{0.4 - p}{\sqrt{p(1-p)}} = \sqrt{225} \frac{0.7 - 0.64}{\sqrt{0.64(1 - 0.64)}} = 1.875 \text{ et } \Phi(1.875) \simeq 0.9696$$

donc

$$\boxed{P(0.4 \leq F_n \leq 0.7) = 96.96\%}$$

(b) supérieure à 60%.

Solution.

On a

$$\begin{aligned}P(F_n \geq 0.6) &= P\left(\sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}} \geq \sqrt{n} \frac{0.6 - p}{\sqrt{p(1-p)}}\right) \\&= P(U \geq c)\end{aligned}$$

avec

$$c = \sqrt{225} \frac{0.6 - 0.64}{\sqrt{0.64(1 - 0.64)}} = -1.25$$

donc

$$P(U \geq c) = P(U \leq -c) = \Phi(-c)$$

et

$$\Phi(-c) = \Phi(1.25) = 0.8944$$

donc

$$\boxed{P(F_n \geq 0.6) = 89.44\%}$$

(c) inférieure à 25%.

Solution.

On a

$$\begin{aligned}P(F_n \leq 0.25) &= P\left(\sqrt{n} \frac{F_n - p}{\sqrt{p(1-p)}} \leq \sqrt{n} \frac{0.25 - p}{\sqrt{p(1-p)}}\right) \\&= P(U \leq d) \\&= \Phi(d)\end{aligned}$$

avec

$$d = \sqrt{225} \frac{0.25 - 0.64}{\sqrt{0.64(1 - 0.64)}} = -12.2$$

donc

$$\Phi(d) = \Phi(-12.2) \simeq 0$$

donc

$$\boxed{P(F_n \leq 0.25) \simeq 0\%}$$

Exercice 3. Un dispositif de signalisation lumineuse comporte trois lampes ; celle qui est en service est relayée automatiquement en cas de défaillance.

On sait que la durée de vie des lampes utilisées est une variable normale de moyenne 1500 heures et d'écart-type 150 heures.

Quelle est la probabilité que l'ensemble fonctionne :

(a) plus de 5000 heures.

Solution.

Soit X_i la durée de vie de la lampe i .

On a $X_i \hookrightarrow \mathcal{N}(1500, 150)$, et les X_i sont supposés indépendantes.

Soit $T_n = \sum_{i=1}^3 X_i$ la durée de fonctionnement total.

On a $S_n \hookrightarrow \mathcal{N}(3 \times 1500, \sqrt{3 \times 150^2}) = \mathcal{N}(4500, 150\sqrt{3})$.

Donc $U = \frac{T_n - 4500}{150\sqrt{3}} \hookrightarrow \mathcal{N}(0, 1)$.

On cherche

$$\begin{aligned} P(T_n > 5000) &= P\left(\frac{T_n - 4500}{150\sqrt{3}} > \frac{5000 - 4500}{150\sqrt{3}}\right) \\ &= P\left(U > \frac{10}{3\sqrt{3}}\right) \\ &= 1 - P\left(U \leq \frac{10}{3\sqrt{3}}\right) \\ &= 1 - \Phi\left(\frac{10}{3\sqrt{3}}\right) \end{aligned}$$

On a $\Phi(\frac{10}{3\sqrt{3}}) \simeq 0.9729$, donc

$$P(T_n > 5000) = 2.71\%$$

(b) mois de 4200 heures.

Solution.

On a

$$\begin{aligned} P(T_n < 4200) &= P\left(\frac{T_n - 4500}{150\sqrt{3}} < \frac{4200 - 4500}{150\sqrt{3}}\right) \\ &= P\left(U < -\frac{2}{\sqrt{3}}\right) \\ &= 1 - P\left(U < \frac{2}{\sqrt{3}}\right) \\ &= 1 - \Phi\left(\frac{2}{\sqrt{3}}\right) \end{aligned}$$

On a $\Phi(\frac{2}{\sqrt{3}}) \simeq 0.8759$, donc

$$P(T_n < 4200) = 12.41\%$$

Exercice 4. Un paquet de tabac produit par la régie des Tabac a un poids moyen de 50 g et un écart-type de 2 g. En supposant que ce tabac soit livré par lots de mille paquets ; quelle est la probabilité que la différence $A - B$ entre les poids de deux lots A et B excède 200 g ?

Solution.

Soit X_i et Y_i respectivement les tailles d'un paquet dans deux lots différents A et B .

Le poids des lots est $T_A = \sum_{i=1}^n X_i$ et $T_B = \sum_{i=1}^n Y_i$ avec $n = 1000$.

On cherche $P(|T_A - T_B| > 200)$.

On a $n \geq 30$ et les X_i sont supposés indépendantes, donc avec le TCL, T_A et T_B suivent approximativement la loi normale $\mathcal{N}(nm, \sigma\sqrt{n})$ avec $m = 50$ et $\sigma = 2$.

Alors $T_A - T_B$ suit la loi normale $\mathcal{N}(0, \sigma\sqrt{2n})$, et donc $U = \frac{T_A - T_B}{\sigma\sqrt{2n}}$ suit la loi normale centrée réduite $\mathcal{N}(0, 1)$.

On a

$$\begin{aligned} P(|T_A - T_B| > 200) &= P\left(\frac{|T_A - T_B|}{\sigma\sqrt{2n}} > \frac{200}{\sigma\sqrt{2n}}\right) \\ &= P(|U| > \frac{200}{\sigma\sqrt{2n}}) \end{aligned}$$

En réarrangeant et en évaluant la valeur à partir de la table de la loi normale centrée réduite, on trouve

$$P(|T_A - T_B| > 200) = 2.54\%$$

Exercice 5. Lors d'un concours radiophonique, on note X le nombre de réponses reçues chaque jour. On suppose que X suit une loi normale de paramètres μ et σ . Durant les 10 jours, on a obtenu

$$\begin{aligned}x_1 &= 200, & x_2 &= 240, & x_3 &= 190, & x_4 &= 150, & x_5 &= 220, \\x_6 &= 180, & x_7 &= 170, & x_8 &= 230, & x_9 &= 210, & x_{10} &= 210\end{aligned}$$

Donner une estimation ponctuelle de μ et σ^2 .

Solution.

Un estimateur de μ est $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, donc une estimation ponctuelle de μ est

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{2000}{10}$$

donc

$$\boxed{\bar{x} = 200}$$

Un estimateur de σ^2 est $S^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$, donc une estimation ponctuelle de σ^2 est

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{407000}{10} - 200^2$$

donc

$$\boxed{s^2 = 700}$$

Exercice 6. Soit (X_1, X_2, \dots, X_n) un n -échantillon de loi de Bernoulli $\mathcal{B}(p)$ et soit

$$S_n = \sum_{k=1}^n X_k, \quad \bar{X}_n = \frac{S_n}{n}$$

(a) Montrer que \bar{X}_n est un estimateur sans biais de p .

Solution.

On a

$$E(\bar{X}_n) = E\left(\frac{S_n}{n}\right) = \frac{1}{n} E(S_n) = \frac{1}{n} E\left(\sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{1}{n} \sum_{k=1}^n p = p$$

donc \bar{X}_n est un estimateur sans biais de p .

(b) On cherche à estimer la variance $\sigma^2 = p(1-p)$. On propose l'estimateur

$$U_n = \bar{X}_n(1 - \bar{X}_n)$$

i. Calculer l'espérance et la variance de S_n .

Solution.

On a

$$E(S_n) = E(n\bar{X}_n) = nE(\bar{X}_n) = np$$

et

$$V(S_n) = V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k) = \sum_{k=1}^n \sigma^2 = n\sigma^2$$

en supposant que les X_k sont indépendantes !

ii. Montrer que U_n est un estimateur biaisé de σ^2 .

Solution.

On a

$$\begin{aligned} E(U_n) &= E(\bar{X}_n - \bar{X}_n^2) \\ &= E(\bar{X}_n) - E(\bar{X}_n^2) \\ &= E(\bar{X}_n) - E(\bar{X}_n)^2 + E(\bar{X}_n)^2 - E(\bar{X}_n^2) \\ &= E(\bar{X}_n)(1 - E(\bar{X}_n)) - V(\bar{X}_n) \end{aligned}$$

avec $V(\bar{X}_n) = V\left(\frac{S_n}{n}\right) = \frac{1}{n^2}V(S_n) = \frac{\sigma^2}{n}$. Donc

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

On a donc $E(U_n) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n} \neq 0$, donc U_n est un estimateur biaisé de σ^2 .

iii. Donner un estimateur V_n sans biais de σ^2 , fonction de U_n .

Solution.

Soit

$$V_n = \frac{n}{n-1}U_n$$

V_n est un estimateur sans biais de σ^2 , car

$$E(V_n) = \frac{n-1}{n}E(U_n) = \frac{n-1}{n} \times \frac{n}{n-1}\sigma^2 = \sigma^2$$

Exercice 7. À la veille d'une consultation électorale, on a interrogé cent électeurs constituant un échantillon au hasard. 54 d'entre eux ont déclaré avoir l'intention de voter pour le candidat Y .

Entre quelle limite, au moment du sondage, avec une probabilité de 0.95, la proportion du corps électoral favorable au candidat Y se situe-t-elle ?

Solution.

Soit

$$X_i = \begin{cases} 1 & \text{si la } i^{\text{e}} \text{ personne compte voter pour } Y \\ 0 & \text{sinon} \end{cases}$$

la variable aléatoire qui suit la loi de Bernoulli de paramètre p , avec $1 \leq i \leq n = 100$. On cherche à estimer p par intervalle de confiance.

L'ESBVM de p est

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

donc l'estimation ponctuelle de p dans l'échantillon est $\hat{p} = \frac{54}{100} = 0.54$.

Soit

$$Z = \sqrt{n} \frac{\bar{X} - p}{\sqrt{p(1-p)}}$$

Les X_i sont supposés indépendantes et $n \geq 30$, donc avec le TCL, Z suit approximativement la loi normale centrée réduite $\mathcal{N}(0, 1)$.

En posant

$$T = n\bar{X} = n\hat{p}$$

on a

$$Z = \frac{T - np}{\sqrt{np(1-p)}}$$

Soit u_α le réel tel que

$$P(|Z| \leq u_\alpha) = 1 - \alpha$$

avec $\alpha = 5\%$.

Pour trouver l'intervalle de confiance recherché, il suffit d'écrire $|Z| \leq u_\alpha$ sous la forme $r_1 \leq p \leq r_2$.

$$\begin{aligned} |Z| \leq u_\alpha &\iff \left| \frac{T - np}{\sqrt{np(1-p)}} \right| \leq u_\alpha \\ &\iff \frac{(T - np)^2}{np(1-p)} \leq u_\alpha^2 \\ &\iff \dots \\ &\iff (n + u_\alpha^2)p^2 - (2T + u_\alpha^2)p + \frac{T^2}{n} \leq 0 \\ &\iff r_1 \leq p \leq r_2 \end{aligned}$$

avec r_1 et r_2 les racines de l'équation quadratique en p

$$(n + u_\alpha^2)p^2 - (2T + u_\alpha^2)p + \frac{T^2}{n} = 0$$

On a

$$\Delta = \dots = u_\alpha^2 \left(u_\alpha^2 + 4T \left(1 - \frac{T}{n} \right) \right)$$

donc

$$r_{1,2} = \dots = \frac{\frac{T}{n} + \frac{u_\alpha^2}{2n} \pm u_\alpha \sqrt{\frac{u_\alpha^2}{4n^2} + \frac{T}{n^2} \left(1 - \frac{T}{n} \right)}}{1 + \frac{u_\alpha^2}{n}}$$

En négligeant u_α devant n , on a

$$\begin{aligned} r_{1,2} &= \frac{T}{n} \pm u_\alpha \sqrt{\frac{T}{n^2} \left(1 - \frac{T}{n} \right)} \\ &= \hat{p} \pm u_\alpha \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \end{aligned}$$

Calculons u_α . On a

$$P(|Z| \leq u_\alpha) = 1 - \alpha \iff P(Z \leq u_\alpha) = 1 - \frac{\alpha}{2}$$

On a $1 - \frac{\alpha}{2} = 0.975$, donc à partir de la table de la loi normale centrée réduite, on lit $u_\alpha = 1.96$. A.N. :

$$\begin{aligned} r_1 &= 0.54 - 1.96 \sqrt{\frac{0.54(1 - 0.54)}{100}} = 0.442 \\ r_2 &= 0.54 + 1.96 \sqrt{\frac{0.54(1 - 0.54)}{100}} = 0.638 \end{aligned}$$

donc un intervalle de confiance (asymptotique) de p au niveau de confiance de 95% est

$$\boxed{[r_1, r_2] = [0.442, 0.638]}$$

Exercice 8. Dans un centre de recrutement, on a mesuré la taille de 400 conscrits. Pour cet échantillon pris au hasard, la taille moyenne \bar{x} est égale à 172 cm et l'écart-type estimé s est égal à 4 cm.

Construire un intervalle qui contienne avec une probabilité 0.99 la taille moyenne de l'ensemble des conscrits de ce centre de recrutement.

Solution.

Soit

$$T = \sqrt{n-1} \frac{\bar{X} - m}{S}$$

On suppose que les X_i sont indépendantes, et on a $n = 400 \geq 30$, donc avec le TCL et le théorème de Fisher, T suit approximativement la loi de Student à $n - 1$ degrés de liberté : $T \hookrightarrow \text{St}(n - 1)$.

On cherche ε tel que

$$\begin{aligned} 1 - \alpha &= P(|\bar{X}_n - m| \leq \varepsilon) \\ &= P\left(\left|\sqrt{n-1} \frac{\bar{X}_n - m}{S}\right| \leq \sqrt{n-1} \frac{\varepsilon}{S}\right) \\ &= P(|T| \leq \sqrt{n-1} \frac{\varepsilon}{S}) \end{aligned}$$

avec $\alpha = 0.01$.

À partir de la table de la loi de Student, on lit la valeur $t_{n-1, \alpha} = 2.5758$ correspondant à $1 - \alpha = 0.99$ et $n = \infty$.

On a $t_{n-1, \alpha} \simeq \sqrt{n-1} \frac{\varepsilon}{s}$, donc $\varepsilon = \frac{s}{\sqrt{n-1}} t_{n-1, \alpha}$.

$$\text{A.N. : } \varepsilon = \frac{4}{\sqrt{399}} \times 2.5758 = 0.5158.$$

Donc un intervalle de confiance de m au niveau de confiance 0.99% est

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon] = [171.48, 172.52]$$

Exercice 9. Si l'écart-type de la durée de vie d'un modèle de lampe électronique est estimé à 100 heures, quelle doit être la taille de l'échantillon à prélever pour que l'erreur sur l'estimation de la durée de vie moyenne n'excède pas 20 heures et avec une probabilité

(a) de 95%.

Solution.

Soit X_i la variable aléatoire qui représente la durée de vie de la i^e lampe. On suppose que l'écart-type est connu : $\sigma = 100$. On pose

$$U = \sqrt{n} \frac{\bar{X} - m}{\sigma}$$

Les X_i sont supposés indépendantes et identiquement distribuées, et on suppose que $n \geq 30$, donc avec le TCL, U suit approximativement la loi normale centrée réduite : $U \hookrightarrow \mathcal{N}(0, 1)$.

On cherche n tel que

$$\begin{aligned} P(|\bar{X} - m| \leq 20) &= 0.95 \\ \iff P\left(|\sqrt{n}\frac{\bar{X} - m}{\sigma}| \leq \sqrt{n}\frac{20}{\sigma}\right) &= 0.95 \\ \iff P(|U| \leq \sqrt{n}\frac{20}{\sigma}) &= 0.95 \\ \iff 2P(U \leq \sqrt{n}\frac{20}{\sigma}) - 1 &= 0.95 \\ \iff P(U \leq \sqrt{n}\frac{20}{\sigma}) &= 0.975 \\ \iff \Phi\left(\sqrt{n}\frac{20}{\sigma}\right) &= 0.975 \end{aligned}$$

avec Φ la fonction de répartition de la loi normale centrée réduite.

À partir de la table, on trouve $\sqrt{n}\frac{20}{\sigma} = 1.96$, soit

$$n = \left(1.96 \times \frac{\sigma}{20}\right)^2 = \left(1.96 \times \frac{100}{20}\right)^2 = 96.04$$

On prend donc $\boxed{n = 96}$ (et on vérifie que n est bien ≥ 30 pour valider l'application du TCL).

(b) de 99%.

Solution.

On cherche n tel que

$$\begin{aligned} P(|\bar{X} - m| \leq 20) &= 0.99 \\ \iff \dots \\ \iff \Phi\left(\sqrt{n}\frac{20}{\sigma}\right) &= 0.995 \end{aligned}$$

À partir de la table, on trouve $\sqrt{n}\frac{20}{\sigma} = 2.575$, soit

$$n = \left(2.575 \times \frac{\sigma}{20}\right)^2 = \left(2.575 \times \frac{100}{20}\right)^2 = 165.7$$

On prend donc $\boxed{n = 166}$ (et on vérifie que n est bien ≥ 30 pour valider l'application du TCL).

Exercice 10. Une machine fabrique des rondelles en série. Le diamètre d est une variable gaussienne dont l'écart-type est égale à 1 millimètre. On prélève au hasard un échantillon de neuf rondelles. Les mesures des diamètres, en millimètre, sont les suivantes :

20.1; 19.9; 20.0; 19.8; 19.7; 20.2; 20.1; 23.1; 22.8

- (a) Construire un intervalle qui contienne avec une probabilité de 0.95 la moyenne m de d .

Solution.

Soit X_i la variable aléatoire représentant la diamètre de la i^e rondelle.

On a $X_i \hookrightarrow \mathcal{N}(m, \sigma)$. On pose

$$U = \sqrt{n} \frac{\bar{X} - m}{\sigma}$$

On a donc $U \hookrightarrow \mathcal{N}(0, 1)$.

On calcule la moyenne de l'échantillon

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k = \frac{185.7}{9} = 20.63$$

On cherche ε tel que

$$\begin{aligned} 1 - \alpha &= P(|\bar{X} - m| \leq \varepsilon) \\ &= P\left(\left|\sqrt{n} \frac{\bar{X} - m}{\sigma}\right| \leq \sqrt{n} \frac{\varepsilon}{\sigma}\right) \\ &= P(|U| \leq \sqrt{n} \frac{\varepsilon}{\sigma}) \end{aligned}$$

donc

$$\begin{aligned} 1 - \frac{\alpha}{2} &= P(U \leq \sqrt{n} \frac{\varepsilon}{\sigma}) \\ &= \Phi\left(\sqrt{n} \frac{\varepsilon}{\sigma}\right) \end{aligned}$$

À partir de la table de la loi normale centrée réduite, on trouve $u_\alpha = 1.96$ correspondant à $1 - \frac{\alpha}{2} = 0.975$.

On a donc $u_\alpha = \sqrt{n} \frac{\varepsilon}{\sigma}$, soit $\varepsilon = \frac{\sigma}{\sqrt{n}} u_\alpha$.

A.N. : $\varepsilon = \frac{1}{\sqrt{9}} \times 1.96 = 0.653$.

Alors un intervalle de confiance de m au niveau 95% est

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon] = [19.98, 21.29]$$

(b) Même question en supposant que l'écart-type de d a une valeur inconnue.

Solution.

On pose

$$T = \sqrt{n-1} \frac{\bar{X} - m}{S}$$

Dans l'échantillon, on a

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{3845.65}{9} - 20.63^2 = 1.56$$

donc $s = \sqrt{1.56} = 1.249$.

Avec le théorème de Fisher, on a $T \hookrightarrow \text{St}(n-1)$.

On cherche ε tel que

$$\begin{aligned} 1 - \alpha &= P(|\bar{X} - m| \leq \varepsilon) \\ &= P\left(|\sqrt{n-1} \frac{\bar{X} - m}{S}| \leq \sqrt{n-1} \frac{\varepsilon}{S}\right) \\ &= P(|T| \leq \sqrt{n-1} \frac{\varepsilon}{S}) \end{aligned}$$

À partir de la table de la loi de Student, on trouve $t_{n-1, \alpha} = 2.3060$ correspondant à $n-1 = 8$ et $1 - \alpha = 0.95$.

On a donc $\sqrt{n-1} \frac{\varepsilon}{s} = t_{n-1, \alpha}$, donc $\varepsilon = \frac{s}{\sqrt{n-1}} t_{n-1, \alpha}$.

A.N. : $\varepsilon = \frac{1.249}{\sqrt{8}} \times 2.3060 = 1.018$.

Donc un intervalle de confiance de m au niveau 95% est

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon] = [19.61, 21.65]$$

(On peut remarquer que cet intervalle est moins précis que le premier. C'est normal puisque la connaissance de l'écart-type apporte plus de précision.)

(c) En utilisant les données numériques de (a), l'écart-type de d étant supposé inconnu, construire un intervalle qui contienne avec une probabilité de 0.9 la variance de d .

Solution.

On pose

$$Y = n \frac{S^2}{\sigma^2}$$

Avec le théorème de Fisher, on a $Y \hookrightarrow \chi_{n-1}^2$.

On a $\forall 0 < a < b$, d'une part

$$P(a \leq Y \leq b) = F_{\chi_{n-1}^2}(b) - F_{\chi_{n-1}^2}(a)$$

avec $F_{\chi_{n-1}^2}$ la fonction de répartition de la loi de khi-deux à $n-1$ degrés de liberté.

Et d'autre part,

$$\begin{aligned}P(a \leq Y \leq b) &= P(a \leq n \frac{S^2}{\sigma^2} \leq b) \\&= P(\frac{n}{b} S^2 \leq \sigma^2 \leq \frac{n}{a} S^2)\end{aligned}$$

On équilibre les risques en choisissant a et b tels que

$$F_{\chi_{n-1}^2}(b) = 1 - \frac{\alpha}{2} \text{ et } F_{\chi_{n-1}^2}(a) = \frac{\alpha}{2} \text{ avec } \alpha = 0.10$$

de telle façon à avoir

$$P(\frac{n}{b} S^2 \leq \sigma^2 \leq \frac{n}{a} S^2) = (1 - \frac{\alpha}{2}) - \frac{\alpha}{2} = 1 - \alpha$$

donc $[\frac{n}{b} s^2, \frac{n}{a} s^2]$ est un intervalle de confiance de σ^2 au niveau de confiance de 90%.

On calcule a et b à partir de la table de la loi de khi-deux avec $n - 1 = 8$.

On trouve $b = 15.51$ correspondant à $1 - \frac{\alpha}{2} = 0.95$.

Et $a = 2.73$ correspondant à $\frac{\alpha}{2} = 0.05$.

A.N. :

$$\frac{n}{b} s^2 = \frac{9}{15.51} \times 1.56 = 0.905$$

$$\frac{n}{a} s^2 = \frac{9}{2.73} \times 1.56 = 5.143$$

donc un intervalle de confiance à 90% pour σ est

$$\boxed{[0.905, 5.143]}$$

Exercice 11. Le chiffre d'affaire mensuel de l'entreprise TEX suit une loi normale de moyenne μ inconnue, mais dont l'écart-type est évalué à 50 M Dh. Sur les 12 derniers mois, la moyenne des chiffres d'affaire mensuels a été de 200 M Dh.

Donner une estimation de μ par intervalle de confiance au niveau 0.98.

Solution.

Soit X_i la variable aléatoire représentant le chiffre d'affaire du i^e mois.

On a $X_i \hookrightarrow \mathcal{N}(\mu, \sigma)$. On suppose que l'écart-type est connu : $\sigma = 50$.

On pose

$$U = \sqrt{n} \frac{\bar{X} - m}{\sigma}$$

On a $U \hookrightarrow \mathcal{N}(0, 1)$.

Avec $\alpha = 0.02$, on cherche ε tel que

$$\begin{aligned} 1 - \alpha &= P(|\bar{X} - m| \leq \varepsilon) \\ &= P\left(\left|\sqrt{n} \frac{\bar{X} - m}{\sigma}\right| \leq \sqrt{n} \frac{\varepsilon}{\sigma}\right) \\ &= P(|U| \leq \sqrt{n} \frac{\varepsilon}{\sigma}) \end{aligned}$$

donc

$$\begin{aligned} 1 - \frac{\alpha}{2} &= P(U \leq \sqrt{n} \frac{\varepsilon}{\sigma}) \\ &= \Phi\left(\sqrt{n} \frac{\varepsilon}{\sigma}\right) \end{aligned}$$

À partir de la table de la loi normale centrée réduite, on trouve $u_\alpha = 2.326$ correspondant à $1 - \frac{\alpha}{2} = 0.99$.

On a donc $u_\alpha = \sqrt{n} \frac{\varepsilon}{\sigma}$, soit $\varepsilon = \frac{\sigma}{\sqrt{n}} u_\alpha$.

A.N. : $\varepsilon = \frac{50}{\sqrt{12}} \times 2.326 = 33.57$.

Donc une estimation de μ par intervalle de confiance au niveau 0.98 est

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon] = [166.43, 233.57]$$

Exercice 12. Dans une station-service, on suppose que le montant des chèques “essence” suit une loi normale de paramètres μ et σ . On considère un échantillon de taille $n = 50$ et on obtient une moyenne de 130 Dh et un écart-type de 28 Dh.

Donner une estimation de μ par intervalle de confiance au niveau 0.95.

Solution.

Soit X_i la variable aléatoire représentant le montant du i^e chèque. On a $X_i \hookrightarrow \mathcal{N}(\mu, \sigma)$.

On a dans l'échantillon $\bar{x} = 130$ et $s = 28$.

Soit

$$T = \sqrt{n-1} \frac{\bar{X} - m}{S}$$

On a, avec théorème de Fisher, T suit la loi de Student à $n - 1$ degrés de liberté : $T \hookrightarrow \text{St}(n - 1)$.

Avec $\alpha = 0.05$, on cherche ε tel que

$$\begin{aligned} 1 - \alpha &= P(|\bar{X}_n - m| \leq \varepsilon) \\ &= P\left(\left|\sqrt{n-1} \frac{\bar{X}_n - m}{S}\right| \leq \sqrt{n-1} \frac{\varepsilon}{S}\right) \\ &= P(|T| \leq \sqrt{n-1} \frac{\varepsilon}{S}) \end{aligned}$$

À partir de la table de la loi de Student, pour trouver la valeur $t_{n-1, \alpha}$ correspondant à $1 - \alpha = 0.95$ et $n - 1 = 49$, on fait une interpolation linéaire : $\frac{t_{n-1, \alpha} - 2.0211}{49 - 40} = \frac{2.0086 - 2.0211}{50 - 40}$ qui donne $t_{n-1, \alpha} = 2.0098$.

On a donc $t_{n-1, \alpha} \simeq \sqrt{n-1} \frac{\varepsilon}{s}$, donc $\varepsilon = \frac{s}{\sqrt{n-1}} t_{n-1, \alpha}$.

A.N. : $\varepsilon = \frac{28}{\sqrt{49}} \times 2.0098 = 8.04$.

Donc un intervalle de confiance de μ au niveau de confiance 0.95 est

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon] = [121.96, 138.04]$$

Exercice 13. Les salaires mensuels des employés d'une entreprise sont supposés suivre une loi normale de paramètres μ et σ .

- (a) Pour un échantillon de taille $n = 10$, on obtient une moyenne $m = 6500$ Dh et un écart-type $s = 900$ Dh. Donner un intervalle de confiance au niveau 0.95 pour μ .

Solution.

Soit X_i la variable aléatoire représentant le salaire mensuel du i^e employé.

On a $X_i \hookrightarrow \mathcal{N}(\mu, \sigma)$. Soit

$$T = \sqrt{n-1} \frac{\bar{X} - m}{S}$$

On a, avec théorème de Fisher, T suit la loi de Student à $n - 1$ degrés de liberté.

Avec $\alpha = 0.05$, on cherche ε tel que

$$\begin{aligned} 1 - \alpha &= P(|\bar{X}_n - m| \leq \varepsilon) \\ &= P\left(\left|\sqrt{n-1} \frac{\bar{X}_n - m}{S}\right| \leq \sqrt{n-1} \frac{\varepsilon}{S}\right) \\ &= P(|T| \leq \sqrt{n-1} \frac{\varepsilon}{S}) \end{aligned}$$

À partir de la table de la loi de Student, on trouve la valeur $t_{n-1, \alpha} = 2.2622$ correspondant à $1 - \alpha = 0.95$ et $n - 1 = 9$.

On a donc $t_{n-1, \alpha} = \sqrt{n-1} \frac{\varepsilon}{s}$, donc $\varepsilon = \frac{s}{\sqrt{n-1}} t_{n-1, \alpha}$.

A.N. : $\varepsilon = \frac{900}{\sqrt{9}} \times 2.2622 = 678.66$.

Donc un intervalle de confiance de μ au niveau de confiance 0.95 est

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon] = [5821.34, 7178.66]$$

- (b) Pour un échantillon de taille $n = 100$, on obtient une moyenne $m = 6200$ Dh et un écart-type $s = 850$ Dh. Donner un intervalle de confiance au niveau 0.95 pour μ .

Solution.

On a

$$1 - \alpha = P(|T| \leq \sqrt{n-1} \frac{\varepsilon}{S})$$

À partir de la table de la loi de Student, pour trouver la valeur $t_{n-1, \alpha}$ correspondant à $1 - \alpha = 0.95$ et $n - 1 = 99$, on fait une interpolation linéaire : $\frac{t_{n-1, \alpha} - 1.9867}{99 - 90} = \frac{1.9840 - 1.9867}{100 - 90}$ qui donne $t_{n-1, \alpha} = 1.9843$.

On a donc $\varepsilon = \frac{s}{\sqrt{n-1}} t_{n-1, \alpha}$. A.N. : $\varepsilon = \frac{850}{\sqrt{99}} \times 1.9843 = 169.52$.

Donc un intervalle de confiance de μ au niveau de confiance 0.95 est

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon] = [6030.48, 6369.52]$$

(On trouve un intervalle plus précis, parce que l'échantillon est plus grand.)

Exercice 14. On se propose d'étudier le corps électoral d'une région.

- (a) 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%.

Solution.

Soit

$$X_i = \begin{cases} 1 & \text{si la } i^{\text{e}} \text{ personne compte voter pour A} \\ 0 & \text{sinon} \end{cases}$$

la variable aléatoire qui suit la loi de Bernoulli de paramètre p , avec $1 \leq i \leq n = 200$. On cherche à estimer p par intervalle de confiance.

L'ESBVM de p est

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

donc l'estimation ponctuelle de p dans l'échantillon est $\hat{p} = \frac{84}{200} = 0.42$.

Soit

$$Z = \sqrt{n} \frac{\bar{X} - p}{\sqrt{p(1-p)}}$$

Les X_i sont supposés indépendants et $n \geq 30$, donc avec le TCL, Z suit approximativement la loi normale centrée réduite $\mathcal{N}(0, 1)$.

En posant

$$T = n\bar{X} = n\hat{p}$$

on a

$$Z = \frac{T - np}{\sqrt{np(1-p)}}$$

Soit u_α le réel tel que

$$P(|Z| \leq u_\alpha) = 1 - \alpha$$

avec $\alpha = 5\%$.

Pour trouver l'intervalle de confiance recherché, il suffit d'écrire $|Z| \leq u_\alpha$ sous la forme $r_1 \leq p \leq r_2$.

$$\begin{aligned} |Z| \leq u_\alpha &\iff \left| \frac{T - np}{\sqrt{np(1-p)}} \right| \leq u_\alpha \\ &\iff \frac{(T - np)^2}{np(1-p)} \leq u_\alpha^2 \\ &\iff \dots \\ &\iff (n + u_\alpha^2)p^2 - (2T + u_\alpha^2)p + \frac{T^2}{n} \leq 0 \\ &\iff r_1 \leq p \leq r_2 \end{aligned}$$

avec r_1 et r_2 les racines de l'équation quadratique en p

$$(n + u_\alpha^2)p^2 - (2T + u_\alpha^2)p + \frac{T^2}{n} = 0$$

On a

$$\Delta = \dots = u_\alpha^2 \left(u_\alpha^2 + 4T \left(1 - \frac{T}{n} \right) \right)$$

donc

$$r_{1,2} = \dots = \frac{\frac{T}{n} + \frac{u_\alpha^2}{2n} \pm u_\alpha \sqrt{\frac{u_\alpha^2}{4n^2} + \frac{T}{n^2} \left(1 - \frac{T}{n} \right)}}{1 + \frac{u_\alpha^2}{n}}$$

En négligeant u_α devant n , on a

$$\begin{aligned} r_{1,2} &= \frac{T}{n} \pm u_\alpha \sqrt{\frac{T}{n^2} \left(1 - \frac{T}{n} \right)} \\ &= \hat{p} \pm u_\alpha \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \end{aligned}$$

Calculons u_α . On a

$$P(|Z| \leq u_\alpha) = 1 - \alpha \iff P(Z \leq u_\alpha) = 1 - \frac{\alpha}{2}$$

On a $1 - \frac{\alpha}{2} = 0.975$, donc à partir de la table de la loi normale centrée réduite, on lit $u_\alpha = 1.96$. A.N. :

$$\begin{aligned} r_1 &= 0.42 - 1.96 \sqrt{\frac{0.42(1 - 0.42)}{200}} = 0.352 \\ r_2 &= 0.42 + 1.96 \sqrt{\frac{0.42(1 - 0.42)}{200}} = 0.488 \end{aligned}$$

donc un intervalle de confiance (asymptotique) de p au niveau de confiance de 95% est

$$\boxed{[r_1, r_2] = [0.352, 0.488]}$$

- (b) Avec un second échantillon de 100 personnes, on a obtenu 45 intentions de vote pour A. En réunissant les deux échantillons, donner un intervalle de confiance pour p au niveau 95%.

Solution.

Les nouveaux paramètres sont $n = 200 + 100 = 300$ et $\hat{p} = \frac{84+45}{300} = 0.43$.

Il suffit de recalculer r_1 et r_2 :

$$r_1 = 0.43 - 1.96\sqrt{\frac{0.43(1-0.43)}{300}} = 0.374$$

$$r_2 = 0.43 + 1.96\sqrt{\frac{0.43(1-0.43)}{300}} = 0.486$$

donc un intervalle de confiance (asymptotique) de p au niveau de confiance de 95% est

$$[r_1, r_2] = [0.374, 0.486]$$

(L'intervalle est plus précis puisque l'échantillon est plus grand.)

- (c) Déterminer la taille n de l'échantillon qui permet d'obtenir un intervalle de confiance de largeur 0.02 sachant qu'une estimation ponctuelle de p a donné la valeur 0.4.

Solution.

La largeur de l'intervalle de confiance est

$$r_2 - r_1 = 2u_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

On résout l'équation

$$2u_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{x}} = 0.02$$

avec $\hat{p} = 0.4$ et x réel. Alors le n cherché est : $n = \lceil x \rceil$.

On a donc

$$x = \hat{p}(1-\hat{p})\left(\frac{2}{0.02}u_\alpha\right)^2 = 0.4 \times (1-0.4) \times \left(\frac{2}{0.02} \times 1.96\right)^2 = 9219.84$$

donc

$$n = 9220$$

- (d) Pour mieux cerner la population votant pour A, on extrait un échantillon de n personnes ayant l'intention de voter pour A. On suppose que l'âge des individus suit une loi normale de paramètres μ et σ .
- i. Le premier échantillon de taille $n = 15$ a donné une moyenne d'âge de 45 ans et un écart-type de 10 ans. Donner un intervalle de confiance pour μ au niveau 98%.

Solution.

Soit U_i la variable aléatoire représentant l'âge du i^e individu de l'échantillon.

On a $U_i \hookrightarrow \mathcal{N}(\mu, \sigma)$, $\bar{u} = 45$, et $s = 10$. Soit

$$T = \sqrt{n-1} \frac{\bar{X} - m}{S}$$

Avec le théorème de Fisher, T suit la loi de Student à $n - 1$ degrés de liberté.

Avec $\alpha = 0.02$, on cherche ε tel que

$$\begin{aligned} 1 - \alpha &= P(|\bar{U} - m| \leq \varepsilon) \\ &= P\left(\left|\sqrt{n-1} \frac{\bar{U} - m}{S}\right| \leq \sqrt{n-1} \frac{\varepsilon}{S}\right) \\ &= P(|T| \leq \sqrt{n-1} \frac{\varepsilon}{S}) \end{aligned}$$

À partir de la table de la loi de Student, on lit la valeur $t_{n-1, \alpha} = 2.6245$ correspondant à $1 - \alpha = 0.98$ et $n - 1 = 14$.

On a donc $t_{n-1, \alpha} = \sqrt{n-1} \frac{\varepsilon}{s}$, donc $\varepsilon = \frac{s}{\sqrt{n-1}} t_{n-1, \alpha}$.

A.N. : $\varepsilon = \frac{10}{\sqrt{14}} \times 2.6245 \simeq 7.0$.

Donc un intervalle de confiance de μ au niveau de confiance 98% est

$$[\bar{u} - \varepsilon, \bar{u} + \varepsilon] = [38, 52]$$

- ii. Le second échantillon de taille $n = 100$ a donné une moyenne d'âge de 47 ans et un écart-type de 9 ans. Donner un intervalle de confiance pour μ au niveau 98%.

Solution.

On a

$$1 - \alpha = P(|T| \leq \sqrt{n-1} \frac{\varepsilon}{S})$$

À partir de la table de la loi de Student, pour trouver la valeur $t_{n-1, \alpha}$ correspondant à $1 - \alpha = 0.98$ et $n - 1 = 99$, on fait une interpolation linéaire : $\frac{t_{n-1, \alpha} - 2.3685}{99 - 90} = \frac{2.3642 - 2.3685}{100 - 90}$, qui donne $t_{n-1, \alpha} = 2.3646$.

On a donc $\varepsilon = \frac{s}{\sqrt{n-1}} t_{n-1, \alpha}$. A.N. : $\varepsilon = \frac{9}{\sqrt{99}} \times 2.3646 \simeq 2.1$.

Donc un intervalle de confiance de μ au niveau de confiance 98% est

$$[\bar{u} - \varepsilon, \bar{u} + \varepsilon] = [44.9, 49.1]$$

- iii. En gardant l'échantillon de taille 100, donner un intervalle de confiance pour σ^2 au niveau 95%.

Solution.

On pose

$$Z = n \frac{S^2}{\sigma^2}$$

Avec le théorème de Fisher, on a $Z \hookrightarrow \chi_{n-1}^2$.

On a $\forall 0 < a < b$, d'une part

$$P(a \leq Z \leq b) = F_{\chi_{n-1}^2}(b) - F_{\chi_{n-1}^2}(a)$$

avec $F_{\chi_{n-1}^2}$ la fonction de répartition de la loi de khi-deux à $n - 1$ degrés de liberté. Et d'autre part,

$$\begin{aligned} P(a \leq Z \leq b) &= P(a \leq n \frac{S^2}{\sigma^2} \leq b) \\ &= P(\frac{n}{b} S^2 \leq \sigma^2 \leq \frac{n}{a} S^2) \end{aligned}$$

On équilibre les risques en choisissant a et b tels que

$$F_{\chi_{n-1}^2}(b) = 1 - \frac{\alpha}{2} \text{ et } F_{\chi_{n-1}^2}(a) = \frac{\alpha}{2} \text{ avec } \alpha = 0.05$$

de telle façon à avoir

$$P(\frac{n}{b} S^2 \leq \sigma^2 \leq \frac{n}{a} S^2) = (1 - \frac{\alpha}{2}) - \frac{\alpha}{2} = 1 - \alpha$$

donc $[\frac{n}{b} S^2, \frac{n}{a} S^2]$ est un intervalle de confiance de σ^2 au niveau de confiance de 95%.

On ne peut pas utiliser directement la table de la loi de khi-deux pour calculer a et b , puisqu'elle ne donne pas les valeurs correspondantes à $n - 1 = 99$.

Mais puisque le nombre de degrés de liberté est grand, on peut approximer χ_{n-1}^2 par $\mathcal{N}(n-1, 2(n-1))$. Dans ce cas, $\frac{Z-(n-1)}{\sqrt{2(n-1)}}$ suit approximativement la loi normale centrée réduite $\mathcal{N}(0, 1)$.

On a

$$\begin{aligned} 1 - \frac{\alpha}{2} &= F_{\chi_{n-1}^2}(b) \\ &= P(Z \leq b) \\ &= P(\frac{Z - (n-1)}{\sqrt{2(n-1)}} \leq \frac{b - (n-1)}{\sqrt{2(n-1)}}) \\ &= \Phi(\frac{b - (n-1)}{\sqrt{2(n-1)}}) \end{aligned}$$

On trouve $\frac{b-(n-1)}{\sqrt{2(n-1)}} = 1.96$ correspondant à $1 - \frac{\alpha}{2} = 0.975$, donc $b = 1.96 \times \sqrt{2 \times 99} + 99 = 126.58$.

On a de façon similaire

$$\begin{aligned}\frac{\alpha}{2} &= F_{\chi_{n-1}^2}(a) \\ &= \Phi\left(\frac{a - (n-1)}{\sqrt{2(n-1)}}\right) \\ &= 1 - \Phi\left(-\frac{a + (n-1)}{\sqrt{2(n-1)}}\right)\end{aligned}$$

ce qui donne

$$1 - \frac{\alpha}{2} = \Phi\left(-\frac{a - (n-1)}{\sqrt{2(n-1)}}\right)$$

On trouve $-\frac{a-(n-1)}{\sqrt{2(n-1)}} = 1.96$ correspondant à $1 - \frac{\alpha}{2} = 0.975$, donc $a = -1.96 \times \sqrt{2 \times 99} + 99 = 71.42$.

A.N. :

$$\begin{aligned}\frac{n}{b}s^2 &= \frac{99}{126.58} \times 9^2 = 63.35 \\ \frac{n}{a}s^2 &= \frac{99}{71.42} \times 9^2 = 112.28\end{aligned}$$

donc un intervalle de confiance à 95% pour σ est

$$[63.55, 112.28]$$