STT1000

CHAPITRE 5 INTERVALLES DE CONFIANCE

SOLUTIONS

Intervalle de confiance pour une moyenne

5.1 Dans sa chasse aux insectes volants, la chauve-souris émet des sons de haute fréquence dont elle reçoit ensuite l'écho. Les pulsations, normalement émises à des intervalles de 50 à 100 millièmes de seconde, ralentissent brusquement dès que la chauve-souris perçoit l'insecte. Profitant de ce fait, des expérimentateurs ont étudié la distance à laquelle la chauve-souris perçoit l'insecte. À l'aide d'une caméra et d'un appareil mesurant le rythme des pulsations ils ont pu déterminer la distance qui séparait la chauve-souris de l'insecte au moment précis où l'insecte a été détecté. Cette expérience a été répétée 11 fois. Les distances (en centimètres) sont les suivantes :

62 ; 52 ; 68 ; 23 ; 34 ; 45 ; 27 ; 42 ; 83 ; 56 ; 40

Déterminer un intervalle de confiance à 95 % pour μ, la distance moyenne nécessaire à la détection d'un insecte par la chauve-souris. [36,21 ; 60,51]

5.2 [*Données du tableau* A05] Une vieille règle affirme que la température normale (donc moyenne) du corps est de 98,6 °F (des études récentes, cependant, fixent le chiffre à 98,2). Déterminez, à partir des données du tableau, un intervalle de confiance pour la température moyenne du corps humain. Vos données vous permettent-elles de conclure que l'une ou l'autre des deux études donne des résultats incorrects ? [Données de l'échantillon: $\overline{X}$ = 98,24923 ; S = 0,7331832 ; n = 130]. $98{,}122 \le \mu \le 98{,}38$

Intervalle de confiance pour une proportion

5.3 L'administration d'un service d'urgences souhaite estimer la proportion p de ses patients dont la résidence se situe dans le secteur administratif de l'hôpital. Dans un échantillon de 200 patients, on en trouve 60 dans cette catégorie. Déterminer un intervalle de confiance à 95 % pour p. [0,236 ; 0,364]

5.4 [*Données du tableau* A03] Déterminez un intervalle de confiance pour la proportion des gens dont le score au test verbal (V) est supérieur au score dans le test global (G). Le nombre de cas où le test verbal est supérieur au score global est 9, parmi 28. 14,84 % ; 49,44 %]

5.5 D'une population de comptes de dépenses on tire un échantillon de 120 comptes pour estimer la proportion p des comptes qui comportent un voyage en dehors de la ville. Soit X le nombre de comptes dans l'échantillon qui comportent un voyage en dehors de la ville. Si X = 28, déterminer un intervalle de confiance à 95% pour p. [0,158 ; 0,309]

5.6 Montrez qu'un intervalle construit par la formule $\hat{p} - 1{,}88\,\hat{\sigma}_{\hat{p}} \le p \le \hat{p} + 2{,}05\,\hat{\sigma}_{\hat{p}}$ est un intervalle de confiance à 95% (approximativement) [vous supposerez que $\dfrac{\hat{p} - p}{\hat{\sigma}_{\hat{p}}} \sim N(0 ; 1)$].

Intervalle de confiance pour une variance

5.7 Au numéro 5.1, déterminer un intervalle de confiance à 85 % pour l'écart-type σ des distances.

5.8 Au numéro 5.1, montrer que tous les intervalles de confiance dans le tableau suivant sont de niveau 85 %.

Intervalle de confiance		Largeur de l'intervalle
Limite Inférieure	Limite supérieure	
221,17	1278,45	1057,28
215,19	1007,26	792,07
204,57	830,02	625,45
192,71	735,64	542,93
140,92	601,32	460,41
0,00	587,17	587,17
225,02	∞	∞

5.9 Au numéro 5.2, soit σ, σ_f et σ_h l'écart-type des températures du corps humain, des femmes et des hommes, respective-ment. Déterminer un intervalle de confiance à 90 % pour σ, pour σ_f et pour σ_h. Les écarts-types échantillonnaux sont S = 0,7331832, S_f = 0,7434878 et S_h = 0,6987558. L'échantillon est constitué de 65 femmes et 65 hommes.

5.10 [*Suite du numéro précédent*] Une façon de décider si $\sigma_f = \sigma_h$ consiste à déterminer un intervalle de confiance pour chacun des deux paramètres et de conclure que $\sigma_f \neq \sigma_h$ si les deux intervalles ne se recoupent pas. Est-ce le cas ? [*Ce n'est pas la façon optimale de procéder. Une meilleure façon de la faire est présentée au chapitre* 6].

Exercices divers

5.11 [*Données du tableau* A03] Considérez les sujets comme un échantillon d'une population de jeunes nord-américains de race blanche. Déterminez un intervalle de confiance pour μ, la taille moyenne de la population ; pour μ_f , la taille moyenne des femmes de la population ; et pour μ_h, la taille moyenne des hommes de cette population. La marge d'erreur devrait être à peu près inversement proportionnelle à la racine carrée de la taille de l'échantillon. Est-ce le cas ici ? Discutez.

5.12 [*Données du tableau* A09] Considérez la variable D = [(B1+B2 - (A1+A2)]/2 comme mesure du gain réalisé par l'apprentissage. Déterminez un intervalle de confiance pour le gain moyen dû à l'apprentissage, pour chacune des trois méthodes.

5.13 Une institution scolaire possède 480 salles de cours. Dans un échantillon de 40 salles, on trouve que 5 salles ont un rétroprojecteur endommagé par des vandales. Déterminer un intervalle de confiance à 95 % pour le nombre *M* de rétroprojecteurs endommagés.

5.14 Une compagnie prélève un échantillon de 50 chèques parmi les 2 500 reçus en une journée donnée. On suit le parcours des chèques jusqu'au moment de leur dépôt dans le compte de la compagnie. On constate que 18 des 50 chèques ont mis plus de 5 jours à être déposés.

a) Déterminer un intervalle de confiance à 95% pour la proportion *p* de chèques dont le délai (entre la réception et le dépôt) excède 5 jours ;

b) Déterminez un intervalle de confiance à 95% pour le nombre de chèques dont le délai excède 5 jours. [567 ; 1233]

c) Supposons qu'on veuille faire un échantillonnage sur les chèques de l'année entière (au nombre de 650 000). À un niveau de 95%, quelle est la taille de l'échantillon qu'il faudrait prélever dans les conditions suivantes (vous prendrez pour *p* l'estimation que vous obtenez avec l'échantillon que vous venez de prélever) ?

i) si on accepte une marge de 2% dans l'estimation de la proportion ; 2213

ii) si on accepte une marge d'erreur *relative* (voir le numéro précédent) de 5% de la proportion réelle ; 2732

iii) si on accepte une marge d'erreur de 10 000 chèques dans l'estimation du nombre de chèques qui accusent un délai de plus de 5 jours. 3740

Exercices théoriques

5.15 Quelle est la taille de l'échantillon qu'on devrait prélever pour estimer une proportion *p* avec une probabilité d'au moins 95 % de ne pas se tromper de plus d'un point de pourcentage? On ne connaît pas la valeur de *p* mais on peut être sûr qu'elle n'est pas supérieure à 15%. [Servez-vous du fait que la fonction $p(1-p)$ est croissante sur l'intervalle (0 ; 1/2).] 4898

5.16 Considérons le problème qui consiste à déterminer la taille de l'échantillon qu'on devrait prélever pour qu'avec une probabilité de 95% l'erreur *relative* dans l'estimation de *p* ne dépasse pas 20 % [L'erreur relative est l'erreur exprimée comme proportion de *p* : erreur relative = $|\hat{p} - p|/p$]. Supposer que la meilleure estimation disponible de *p* est $p = 0{,}10$ et estimer la taille de l'échantillon qu'il faut prélever. 864

5.17 Soit X_1 et X_2 un échantillon aléatoire de taille $n = 2$ d'une population $\mathcal{N}(\mu ; \sigma^2)$. Déterminer un intervalle de confiance pour σ^2. [Considérer la distribution de $\dfrac{(X_1 - X_2)^2}{2\sigma^2}$.]

5.18 On sait que si $X \sim \chi^2_\nu$ et si ν est grand, alors la variable $\sqrt{2X} - \sqrt{2\nu}$ suit approximativement une loi $\mathcal{N}(0 ; 1)$. Servez-vous de ce résultat pour montrer que si *n* est grand, alors *S* suit à peu près une loi $\mathcal{N}(\sigma ; \sigma^2/2(n-1))$. Déterminer un intervalle de confiance à 95% pour σ, sachant qu'un échantillon de taille 250 a donné $S = 39$.

5.19 On peut définir la loi de *Student* en ces termes: soit X et Y deux variables aléatoires *indépendantes*, $X \sim \mathcal{N}(0\,;\,1)$, $Y \sim \chi^2_\nu$. Alors la variable $T = \dfrac{X^2}{\sqrt{Y^2/\nu}} \sim t_\nu$. Montrer que, si $\nu > 2$, $E[T] = 0$, et $\mathrm{Var}[T] = E[T^2]$. [*Suggestion*: Noter que $E(T) = E[X/(Y/\nu)] = \nu\, E[X\,Y^{-1}] = \nu\, E[X]\, E[Y^{-1}]$ où la dernière égalité découle du fait que l'espérance d'un produit de variables indépendantes est égale au produit de leurs espérances; et $\mathrm{Var}[T] = \nu\, E[X^2]\, E[Y^{-1}]$]

5.20 La durée d'une certaine pièce électronique est une variable de loi exponentielle de moyenne β. Un échantillon de 4 pièces donne les durées suivantes (en milliers d'heures) : 2,4 ; 3,0 ; 8,4 ; 11,0. Déterminer un intervalle de confiance à 80 % pour β [utiliser $2\Sigma X_i/\beta$ comme pivot]. $3{,}71 \le \beta \le 14{,}21$

5.21 Soit $X_1, \ldots, X_n$ un échantillon aléatoire tiré d'une population de densité

$$f(x) = \frac{x^{1/\theta-1}}{\theta}\ ;\ 0 < x < 1,\ \theta > 0.$$

a) Considérer les variables $Y_i = -\ln X_i$. Montrer que la fonction de densité de Y_i est $g(y) = \dfrac{1}{\theta} e^{-y/\theta}$.

b) Déterminer un intervalle de confiance pour θ [Considérez la loi des $Z_i = 2Y_i/\theta$ puis celle de leur moyenne $\bar{Z}$].

5.22 Soit une observation d'une population de densité

$$f(x) = \frac{x^{1/\theta-1}}{\theta}\ ;\ 0 < x < 1,\ \theta > 0.$$

a) Déterminer la fonction de répartition de $Y = X^{1/\theta}$.

b) Déterminer un intervalle de confiance à 95% pour θ à partir d'une seule observation tirée de cette population.

5.23 Soit n observations d'une population de densité

$$f(x) = \frac{1}{\theta} e^{-x/\theta}\ ;\ x > 0,\ \theta > 0.$$

Déterminer un intervalle de confiance pour θ en partant du fait que $Y = \dfrac{2}{\theta}\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$.

5.24 Soit $X_1, \ldots, X_n$ un échantillon aléatoire d'une population de loi de densité

$$f(x) = \theta\, x^{-(\theta+1)},\ x > 1,\ \theta > 0.$$

a) Montrer que la variable $Y_i = \theta \ln(X_i)$ suit une loi exponentielle de paramètre 1.

b) Déterminez la distribution de $Z = \sum Y_i$ et déterminez un intervalle de confiance pour θ à partir d'un échantillon de n observations X_1 ; X_2 ; … ; X_n.

c) Montrer que $\dfrac{n-1}{\sum_{i=1}^{n} \ln X_i}$ est un estimateur sans biais de θ.

5.25 Sur une certaine ligne de métro les trains arrivent régulièrement à toutes les θ minutes. Vous arrivez au quai à un moment aléatoire et vous attendez le prochain métro. Soit X votre temps d'attente.

a) Utiliser le fait que $Y = X/\theta$ est de loi uniforme sur $[0\,;\,1]$ pour montrer que l'intervalle $\left[\dfrac{X}{1-\alpha/2} \le \theta \le \dfrac{X}{\alpha/2}\right]$ est un intervalle de confiance à $100(1-\alpha)$ % pour θ.

b) Montrer que pour α fixe et $0 < \beta \leq \alpha$, l'intervalle $\left[\dfrac{X}{1-(\alpha-\beta)} \leq \theta \leq \dfrac{X}{\beta}\right]$ est un intervalle de confiance de niveau 1-α. Quelle est la valeur de β qui donne l'intervalle le plus court?

c) Supposons que le train arrive après13 minutes d'attente:

i) Déterminer un intervalle de confiance à 90% par la formule en a)

ii) Déterminer un intervalle de confiance à 90% par la formule en b)

iii) Déterminer un intervalle à 90% de la forme $[a \leq \theta < \infty)$. Est-ce que l'intervalle vous permettrait de contredire l'affirmation que $\theta = 10$?

d) Vous arrivez n fois au quai et vous notez à chaque fois votre temps d'attente. Montrer que l'intervalle suivant est un intervalle de confiance à 100(1-α) % pour θ en utilisant le fait que la durée moyenne d'attente $\bar{X}$ est approximativement de loi normale: $\left[\dfrac{\bar{X}}{\frac{1}{2}+\frac{z_{\alpha/2}}{\sqrt{12n}}} \; ; \; \dfrac{\bar{X}}{\frac{1}{2}-\frac{z_{\alpha/2}}{\sqrt{12n}}}\right]$

e) Évaluer l'intervalle de confiance par la formule développée en d) à partir des 8 temps d'attente observés suivants (on admettra pour les besoins de ce numéro que n = 8 est assez « grand »): [8,483;17,799]

0,05 ; 3,17 ; 4,61 ; 6,02 ; 6,44 ; 7,67 ; 9,21 ; 10,34

e) Dans un échantillon de taille n, l'estimateur du maximum de vraisemblance de θ est $Y = X_{(n)}$, la plus grande des n valeurs. Montrer que l'intervalle $\left[\dfrac{Y}{(1-\alpha/2)^{1/n}} \; ; \; \dfrac{Y}{(\alpha/2)^{1/n}}\right]$ est un intervalle de confiance de niveau 1-α. Utiliser cette méthode pour déterminer un intervalle de confiance à 90 % pour θ à partir des données présentées en d).
$10,41 \leq \theta \leq 15,04$

5.26 Soit $X_1, \ldots, X_n$ un échantillon d'une population de loi uniforme sur $[0 ; \theta]$. On propose ici trois façons de déterminer un intervalle de confiance pour θ :

Méthode A : Soit $G(y ; \theta)$ la fonction de répartition de $Y = \max\{X_1, \ldots, X_n\}$. L'intervalle de confiance est l'ensemble des valeurs de θ pour lesquelles $\alpha/2 \leq G(y ; \theta) \leq 1 - \alpha/2$.

Méthode B : Soit $H(z ; \theta)$ la fonction de répartition de $Z = \min\{X_1, \ldots, X_n\}$. L'intervalle de confiance est l'ensemble des valeurs de θ pour lesquelles $\alpha/2 \leq H(z ; \theta) \leq 1 - \alpha/2$.

Méthode C : L'intervalle de confiance approximatif est l'ensemble des valeurs de θ pour lesquelles $\left|\dfrac{\bar{X}-\mu(\theta)}{\sigma(\theta)/\sqrt{n}}\right| \leq z_{\alpha/2}$, où $\mu(\theta) = \theta/2$ et $\sigma(\theta) = \theta/\sqrt{12}$.

a) Méthode A

i) Montrer que $G(y) = (y/\theta)^n$, $0 \leq y \leq \theta$;

ii) Montrer que la fonction de densité de Y est $g(y) = ny^{n-1}/\theta^n$, $0 \leq y \leq \theta$;

iii) Montrer que $E(Y) = n\theta/(n+1)$;

iv) Montrer que l'intervalle de confiance est $\dfrac{Y}{(1-\alpha/2)^{1/n}} \leq \theta \leq \dfrac{Y}{(\alpha/2)^{1/n}}$.

v) Soit $\mathcal{L}(Y)$ la largeur de l'intervalle de confiance. Montrer que $E[\mathcal{L}(Y)]$ tend vers 0 lorsque n tend vers l'infini, que $nE[\mathcal{L}(Y)]$ tend vers $\theta[\ln(1-\alpha/2) - \ln(\alpha/2)]$ et donc que $\sqrt{n}\,E[\mathcal{L}(Y)]$ tend vers 0

b) Méthode B

i) Montrer que $H(z) = 1-\left(1-\frac{z}{\theta}\right)^n$, $0 \le z \le \theta$.

ii) Montrer que la fonction de densité de Z est $h(z) = \frac{n}{\theta}\left(1-\frac{z}{\theta}\right)^{n-1}$, $0 \le z \le \theta$.

iii) Montrer que $E(Z) = \frac{\theta}{n+1}$.

iv) Montrer que l'intervalle $\frac{Z}{1-(\alpha/2)^{1/n}} \le \theta \le \frac{Z}{1-(1-\alpha/2)^{1/n}}$ est un intervalle de confiance à 100(1-α) % pour θ.

v) Soit $M(Z)$ la largeur de l'intervalle de confiance. Montrer que $E[M(Z)]$ tend vers $\theta\frac{\ln(\alpha/2)-\ln(1-\alpha/2)}{\ln(\alpha/2)\ln(1-\alpha/2)}$ lorsque n tend vers l'infini.

c) Méthode C

i) Montrer que l'intervalle de confiance approximatif est $\frac{2\bar{X}}{1+\frac{2z_{\alpha/2}}{\sqrt{12n}}} \le \theta \le \frac{2\bar{X}}{1-\frac{2z_{\alpha/2}}{\sqrt{12n}}}$.

ii) Montrer que l'espérance de la largeur $K(\bar{X})$ de l'intervalle de confiance est $\frac{\theta z_{\alpha/2}}{\sqrt{3n}/2 - z^2_{\alpha/2}/\sqrt{12n}}$.

iii) Montrer que $E[K(\bar{X})]$ tend vers 0 lorsque n tend vers l'infini et que $\sqrt{n}E[K(\bar{X})]$ tend vers $\frac{2\theta z_{\alpha/2}}{\sqrt{3}}$.

d) L'une des trois méthodes est inacceptable. Laquelle? Des deux autres méthodes, laquelle est meilleure?

e) On conçoit aisément pourquoi la méthode B est si peu efficace, comparée à la méthode A: le minimum ne peut contenir beaucoup d'information sur θ, qui est la valeur maximale de la variable.

i) Vérifier que $\hat{\theta}_1 = (n+1)Z$ est sans biais et montrer que $\mathrm{Var}(\hat{\theta}_1) = \frac{n\theta^2}{n+2}$ [Remarquer que la variance ne décroît pas avec n].

ii) Vérifier que $\hat{\theta}_2 = \frac{n+1}{n}Y$ est sans biais et montrer que $\mathrm{Var}(\hat{\theta}_2) = \frac{\theta^2}{n(n+2)}$.

iii) Vérifier que $\hat{\theta}_3 = 2\bar{X}$ est sans biais et montrer que $\mathrm{Var}(\hat{\theta}_3) = \frac{\theta^2}{3n}$.

f) Un échantillon de taille $n = 30$ est tiré d'une population de loi uniforme sur [0 ; θ]. Voici les observations:

0,15	3,82	8,31	12,18	19,61	21,07	27,89	31,36	36,53	42,21
0,22	4,23	8,46	15,33	20,70	21,96	29,46	31,43	38,68	48,64
2,52	4,60	10,22	18,74	20,83	26,00	31,33	36,38	41,98	49,39

Déterminer un intervalle de confiance utilisant chacune des trois méthodes.

Données de base: $y = 49, 390$; $z = 0{,}150$; $\bar{x} = 22, 141$.

Intervalles de confiance à 95 %:

Méthode A: [55,85 ; 49,43]
Méthode B: [1,30 ; 177,82]
Méthode C: [55,81 ; 36,70]