Eléments de statistique

Répétition 3

6 novembre 2018

Estimation

Le but de l'estimation est d'évaluer un ou plusieurs paramètre(s) de population à partir d'un ou plusieurs échantillon(s).

Exercices

3.1. Le record du monde de marathon est détenu par le Kenyan Dennis Kimetto en 2 h 02 minutes 57 secondes. On tire un échantillon i.i.d. de 9 résultats de ce marathonien et le temps moyen réalisé est d'exactement 125 minutes. Donner l'intervalle de confiance à 95% du temps de course moyen de Dennis Kimetto, sachant que les résultats sur marathon de cet athlète suivent une loi normale dont l'écart-type est de 1 minute.

3.2. On s'intéresse à la durée de vie moyenne des piles de 1,5 volts.

a) On observe 100 piles et on constate que la durée moyenne de vie est de 57,4 heures avec un écart-type de 11,04 heures. Donner l'intervalle de confiance à 99% de la durée de vie moyenne d'une pile de la population.

b) Quelle est la taille d'un échantillon pour avoir un intervalle de confiance à 99% dont la largeur est de 4 heures au plus, si on pense que l'écart-type de la population est égal à 10 heures ?

3.3. Les temps de séchage (en heures) suivants ont été enregistrés lors de différents essais d'une nouvelle peinture au latex : 3.4, 2.5, 4.8, 2.9, 3.6, 2.8, 3.3, 5.6, 3.7, 2.8, 4.4, 4.0, 5.2, 3.0 et 4.8. On suppose que les mesures effectuées représentent un échantillon d'une population normale.

a) Trouvez un intervalle de confiance à 95% du temps de séchage moyen de la peinture.

b) Même question en sachant que l'écart-type de la population est égal à 0,97.

c) Comparez les intervalles de confiance obtenus aux points a) et b). Que peut-on en conclure ?

d) Supposons qu'on fasse une nouvelle mesure s'élevant à 6,9 heures. Cette mesure est-elle surprenante ?

3.4. Supposons qu'on étudie une variable quantitative suivant une loi $\mathcal{N}(\mu; \sigma^2)$ en partant d'un échantillon i.i.d.. Comparez les estimateurs s_n^2 et s_{n-1}^2 du paramètre σ^2 en terme de variance, biais et erreur quadratique moyenne.

Exercices suggérés

3.5. Une élection entre un candidat A et un candidat B va avoir lieu dans un pays dont le corps électoral est constitué de 35 millions de personnes. Un sondage a été réalisé auprès de 1400 électeurs pris au hasard. L'échantillon a donné 680 personnes prêtes à voter pour A. Déterminer l'intervalle de confiance au seuil de 95% de la proportion de personnes favorables au candidat A.

3.6. Comparez l'erreur quadratique moyenne de l'estimateur s^2_{n+1} à celle des autres estimateurs de l'exercice 3.5.

3.7. Une entreprise du secteur métallique effectue un test d'une nouvelle méthode pour lutter contre la corrosion. Un échantillon aléatoire de 100 pièces métaliques n'ayant subi aucun traitement a été imergé dans un bain pendant 24 heures. Les pièces de métal on en moyenne perdu 12,2 millimètre d'épaisseur, avec un écart-type de 1,1 millimètre. Un second échantillon de 200 pièces métalliques fût exposé au traitement anti-corrosif avant d'être immergé à son tour dans un bain pendant 24 heures. Cette fois-ci, la diminution moyenne de l'épaisseur fût de 9,1 millimètres avec un écart-type de 0,9 millimètres.

a) Estimez à l'aide d'un intervalle de confiance à 98%, quelle est la différence de diminution d'épaisseur entre les deux échantillons.

b) Le traitement semble-t-il avoir bien diminué la corrosion moyenne ?

Suggestion : commencez par trouver les formules de distribution d'échantillonage de la différence entre deux moyennes.

3.8. EXAMEN JANVIER 2016

Un magasin vend des cartes mémoires MircoSD dont les capacités de stockage sont de respectivement 8Go, 16Go, 32Go et 64Go. Un client tire un échantillon de 2 cartes MicroSD par tirage i.i.d. dans cet ensemble muni d'une loi uniforme (avec remise, à chaque tirage chaque capacité de stockage a $\frac{1}{4}$ d'être choisie). Soit m_x la moyenne de capacité de stockage des cartes MicroSD de l'échantillon, s^2_x la variance et s^2_{n-1} la variance corrigée.

a) Ecrivez tous les échantillons possibles et à partir de là, calculez l'espérance et la variance de la variable m_x. (**3 points**)

b) Calculez la moyenne et la variance de la capacité de stockage des cartes MicroSD dans la population. (**2 points**)

c) L'espérance calculée au point a) est-elle égale à la moyenne calculée au point b) ? Si oui, pourquoi ? Sinon, quelle formule les relie ? (**2 points**)

d) Les variances calculées aux points a) et b) sont-elles égales ? Si oui, pourquoi ? Sinon, quelle formule les relie ? (**2 points**)

e) Peut-on utiliser les formules $E\{s^2_x\} = \frac{n-1}{n}\sigma^2_{\mathcal{X}}$ et $E\{s^2_{n-1}\} = \sigma^2_{\mathcal{X}}$ pour calculer les espérances de s^2_x et de s^2_{n-1} ? Justifiez. Si oui, calculez ces deux espérances. (**2 points**)

f) Peut-on utiliser les formules $V\{s^2_x\} = \frac{2(n-1)}{n^2}\sigma^4_{\mathcal{X}}$ et $V\{s^2_{n-1}\} = \frac{2}{n-1}\sigma^4_{\mathcal{X}}$ pour calculer les variances de s^2_x et de s^2_{n-1} ? Justifiez. Si oui, calculez ces deux variances.

(**2 points**)

g) Quel est l'intérêt de corriger la variance ? Quel est le désavantage ? (**1 point**)

Afin de préparer une nouvelle campagne de publicité, le magasin réalise un sondage auprès de ses clients afin d'essayer d'estimer la quantité de mémoire totale réellement utilisée par ceux-ci. Seules 11 personnes ont répondu au sondage. Voici les résultats : 22,774 ; 29,026 ; 24,184 ; 31,389 ; 36,620 ; 30,526 ; 31,890 ; 24,272 ; 26,488 ; 36,057 ; 29.534. On suppose que les mesures effectuées représentent un échantillon d'une population normale.

h) Trouvez un intervalle de confiance à 95% de la quantité de mémoire totale moyenne réellement utilisée par les clients. (**2 points**)

i) Même question en supposant que l'écart-type de la population vaut 5. (**2 points**)

j) L'intervalle de confiance calculé au point i) devait-il forcément être au moins aussi précis que celui calculé au point h) ? (**2 points**)

Solutions des exercices suggérés

3.5. $[0,46; 0,51]$
3.6. L'erreur quadratique vaut $2\frac{\sigma^4}{n+1}$. Plus faible que celles de s_n^2 et s_{n-1}^2.
3.7. a) $[2,8; 3,4]$ b) oui
3.8. a) Il y a 16 échantillons différents :

(8, 8,)	(8, 16)	(8, 32)	(8, 64)
(16, 8,)	(16, 16)	(16, 32)	(16, 64)
(32, 8,)	(32, 16)	(32, 32)	(32, 64)
(64, 8,)	(64, 16)	(64, 32)	(64, 64)

Les moyennes de ces échantillons sont (dans le même ordre) :

8	12	20	36
12	16	24	40
20	24	32	48
36	40	48	64

$E\{m_x\} = 30$ Go $\quad V\{m_x\} = 230$ Go^2

b) $\mu_{\mathcal{X}} = 30$ Go $\quad \sigma^2_{\mathcal{X}} = 460$ Go^2

c) La moyenne des moyennes des échantillons $E\{m_x\}$, donne bien la moyenne de la population $\mu_{\mathcal{X}}$.

d) La variance des moyennes des échantillons $V\{m_x\}$, n'est pas égale à la variance de la population $\sigma^2_{\mathcal{X}}$. Ceci est logique, vu que le calcul de cette dernière se base sur chaque valeur de la population prise une seule fois, tandis que le caclul de chaque m_x se base sur des échantillons de taille 2, dont la moyenne a ainsi plus de chance de se rapporcher de la moyenne de la population. On observera donc moins d'écarts par rapport à la dite moyenne.
La formule du formulaire permet de relier les deux grandeurs correctement :

$\frac{\sigma^2_{\mathcal{X}}}{n} = V\{m_x\} = 230$

e) Oui car ces formules sont valables quelque soit la loi de la variable parente.
$E\{s_x^2\} = 230\text{Go}^2 \quad E\{s_{n-1}^2\} = 460\text{Go}^2$

f) On ne peut pas utiliser les formules du formulaire, on est pas dans le cas gaussien.

g) L'intérêt de corriger la variance est de disposer d'un estimateur non-biaisé. On remarque que cela vient au détriment de la variance de l'estimateur, qui est plus élevée dans le cas de s_{n-1}^2 que de s_x^2.

h) [26,23 Go ; 32,45 Go]

i) [26,387 Go ; 32,297 Go]

j) Le fait de connaître l'écart-type de la population permet en général de construire un IC plus précis. Par contre, si notre estimation de l'écart-type de la population est fortement sous-estimée, il se peut qu'on obtienne malgré tout un intervalle de confiance plus étroit !

Loi du χ^2 : densité de probabilité et fonction de répartition