Université de Strasbourg
UFR de Mathématiques et d'Informatique
Han-Ping LI

Année 2018/2019
Statistique
Etude de cas L3

Fiche 3 Intervalles de confiance

Exercice 1 :

On s'intéresse à la taille du lobe frontal des crabes (Leptograpsus variegatus). On prend 200 valeurs de la variable la taille du lobe frontal "FL" du fichier "crabs" de la librairie "MASS" comme population. On note μ et σ^2 la moyenne théorique et la variance théorique de la variable FL. Après une étude statistique préalable (test de la normalité), on suppose raisonnablement que la population admet une loi normale $\mathcal{N}(\mu, \sigma^2)$ avec les deux paramètres inconnus.

a)install.packages("MASS") ; library(MASS)

b) Sauvegarder ces 200 valeurs de "FL" du fichier "crabs" (sans "e") dans une variable nommée **Population**.

c) Générer une réalisation de l'échantillon $(X_1...X_n)$ avec n=28 , puis construire l'intervalle de confiance de μ de niveau de confiance $1-\alpha = 95\%$. Contient-il la valeur de $\mu = mean(Population)$?

d) Générer 100 réalisations de l'échantillon $(X_1...X_n)$ avec n=28 , puis construire 100 réalisations de l'intervalle de confiance de μ de niveau de confiance $1 - \alpha = 95\%$. Quel est le pourcentage des réalisations qui ne contiennent pas le paramètre $\mu = 15.583$?

e) Générer une réalisation de l'échantillon $(X_1...X_n)$ avec n=15 , puis construire l'intervalle de confiance de σ^2 de niveau de confiance $1-\alpha = 90\%$. Contient-il la valeur de $\sigma^2 = var(Population)$?

f) Générer 100 réalisations de l'échantillon $(X_1...X_n)$ avec n=15 , puis construire l'intervalle de confiance de σ^2 de niveau de confiance $1 - \alpha = 90\%$. Quel est le pourcentage des réalisations qui ne contiennent pas le paramètre $\sigma^2 = 12.2173$?

g) Comparer les résultat avec ceux des autres étudiants.

Exercice 1' pour les économistes :

On s'intéresse aux rendements journaliers de bourse de l'indice Standard and Poors 500 de 1990 à 1999. On prend 2780 valeurs du fichier "SP500" de la librairie "MASS" comme population. On note μ et σ^2 la moyenne théorique et la variance théorique du rendement. Après une étude statistique préalable (test de la normalité), on suppose raisonnablement que la population admet une loi normale $\mathcal{N}(\mu, \sigma^2)$ avec les deux paramètres inconnus.

a)install.packages("MASS") ; library(MASS)

b) Sauvegarder ces $N = 2780$ valeurs du rendement du fichier "SP500" dans une variable nommée **Population**.

c) Générer une réalisation de l'échantillon $(X_1...X_n)$ avec n=28 , puis construire l'intervalle de confiance de μ de niveau de confiance $1-\alpha = 95\%$. Contient-il la valeur de $\mu = mean(Population)$?

d) Générer 100 réalisations de l'échantillon $(X_1...X_n)$ avec n=28 , puis construire 100 réalisations de l'intervalle de confiance de μ de niveau de confiance $1 - \alpha = 95\%$. Quel est le pourcentage des réalisations qui ne contiennent pas le paramètre $\mu = 0.04575267$?

e) Générer une réalisation de l'échantillon $(X_1...X_n)$ avec n=15 , puis construire l'intervalle de confiance de σ^2 de niveau de confiance $1-\alpha = 90\%$. Contient-il la valeur de $\sigma^2 = var(Population)$?

f) Générer 100 réalisations de l'échantillon $(X_1...X_n)$ avec n=15 , puis construire l'intervalle de confiance de σ^2 de niveau de confiance $1 - \alpha = 90\%$. Quel est le pourcentage des réalisations qui ne contiennent pas le paramètre $\sigma^2 = 0.8982233$?

g) Comparer les résultat avec ceux des autres étudiants.

Exercice 2 :

On simule $M = 100$ réalisations d'un échantillon $(X_1, \ldots, X_{20})$ (n=20) de la loi $\mathcal{N}(-10, 1.5^2)$. On n'est pas censé de connaitre les deux paramètres $\mu = (-10$ et $\sigma^2 = 1.5^2$. On sauvegarde 100 réalisations de la statistique $\overline{X}$ ainsi que celles de S_c^2.

a) Déterminer 100 réalisation de l'intervalle de confiance de μ au niveau de confiance de **90**%, basée sur ces 100 réalisations. Quel est le pourcentage des réalisations qui ne contiennent pas le paramètre $\mu = -10$? Commenter vos résultats.

b) Comment utiliser R pour représenter ces 100 réalisations de l'intervalle de confiance su un même graphe ?

c) Déterminer 100 réalisation de l'intervalle de confiance de σ^2 au niveau de confiance de 95%, basée sur ces 100 réalisations. Quel est le pourcentage des réalisations qui ne contiennent pas le paramètre $\sigma^2 = 1.5^2$?

d) Ayant un niveau de confiance $95\%(\alpha = 0.05)$, les deux quantiles suivants

$$\mathbb{P}\Big(\frac{(n-1)S_c^2}{\sigma^2} < C_1\Big) = \mathbb{P}\Big(\frac{(n-1)S_c^2}{\sigma^2} > C_2\Big) = \frac{\alpha}{2}.$$

fournissent un choix simple à calculer, et proche du choix optimal. Quelles sont les longueurs des réalisations des IC en c) ?

e) Pour le même niveau de confiance 95%, on définit dans cette question

$$\begin{aligned}\mathbb{P}\Big(\frac{(n-1)S_c^2}{\sigma^2} < C_1\Big) &= 0.0445\\ \mathbb{P}\Big(\frac{(n-1)S_c^2}{\sigma^2} > C_2\Big) &= 0.00539.\end{aligned}$$

On sait qu'ils sont biaisés. Montrer qu'ils ont systématiques une longueurs plus courtes.

Exercice 3 : Intervalle de confiance pour π (la proportion de largeur d'anneaux $\leqslant 0,89$)

On s'intéresse à la largeur des anneaux de croissance des arbres. On prend le fichier "treering" avec 7980 comme étant le cardinal de la population. On note π la proportion de largeurs qui sont inférieures ou égales à 0,89 (unité) et on souhaite construire un intervalle de confiance de π de niveau de confiance de 95%.

a) Convertir ensuite les valeurs des "treering" en 1 ou 0 selon que les valeurs des largeurs $\leqslant 0,89$ (unité) ou non. Sauvegarder ces 7980 valeurs converties dans un vecteur nommé **Population**.

b) Générer une réalisation de l'échantillon $(X_1, ..., X_n)$ de taille $n = 65$.

c) Appliquer Théorème Central Limite pour construire un intervalle de confiance de niveau 95% du paramètre π. Contient-il le paramètre π ?

install.packages("BioStatR") ; library(BioStatR)

d) Comparer avec le résultat par la méthode exacte :
binom.ci(k,n=n1, conf.level=0.95,method="exact")

e) Comparer avec le résultat par la méthode de Wilson :
binom.ci(k,n=n1, conf.level=0.95,method="Wilson")

f) Comparer avec le résultat par la méthode de Wald :
binom.ci(k,n=n1, conf.level=0.95,method="wald")

g) Mêmes questions avec $M = 100$ réalisations.

Exercice 4 :

On simule $M = 100$ réalisations d'un échantillon $(X_1, \ldots, X_{15})$ (n=15) de la loi de Bernoulli $\mathcal{B}(1, 0.3)$. On n'est pas censé de connaitre le paramètre $p = 0.3$. On sauvegarde 100 réalisations de la statistique $\overline{X}$.

a) Déterminer, par la méthode exacte, 100 réalisation de l'intervalle de confiance de p au niveau de confiance de 95%, basée sur ces 100 réalisations. Quel est le pourcentage des réalisations qui ne contiennent pas le paramètre $p = 0.3$? Commenter vos résultats.

b) Comment utiliser R pour représenter ces 100 réalisations de l'intervalle de confiance su un même graphe ?

c) Déterminer, par la méthode de Wilson, 100 réalisation de l'intervalle de confiance de p au niveau de confiance de 95%, basée sur ces 100 réalisations. Quel est le pourcentage des réalisations qui ne contiennent pas le paramètre p ?

d) Comment utiliser R pour représenter ces 100 réalisations de l'intervalle de confiance su un même graphe ?

e) Peut-on utiliser la méthode de Wald ? Pourquoi ?