

Statistique inférentielle

TP 2 : Intervalles de confiance

Le fichier `poids.csv` fournit les poids à la naissance des nouveau-nés d'une maternité observés sur une année. La valeur indiquée est en gramme. Les données sont générées selon une loi normale.

1. Calculer la moyenne (empirique) et la variance (empirique) du poids à la naissance pour l'ensemble des nouveau-nés.
2. Écrire une fonction qui, à partir d'un vecteur de données et d'un entier n , renvoie une liste contenant un échantillon de taille n provenant des données fournies, la moyenne empirique et la variance empirique corrigée de cet échantillon. Estimer à l'aide de cette fonction la moyenne et la variance du poids à la naissance à partir d'échantillons de taille $n = 30, 60$ et 80 .
3. Écrire une fonction qui, à partir d'un vecteur de données normales, renvoie un intervalle de confiance de niveau de risque α pour la moyenne en supposant la variance connue. Déterminer les intervalles de confiance à 95% pour le poids moyen à la naissance à partir des échantillons obtenus à la question 2 (on prendra $\sigma^2 = 438,9^2$).
4. Écrire une fonction qui, à partir d'un vecteur de données normales, renvoie un intervalle de confiance de niveau de risque α pour la moyenne en supposant la variance inconnue. Déterminer les intervalles de confiance à 95% pour le poids moyen à la naissance à partir des échantillons obtenus à la question 2. Comparer les résultats avec ceux de la fonction `t.test`.

Afin d'offrir les premiers soins aux nouveau-nés, la maternité s'approvisionne en lait premier âge auprès d'un fournisseur local. Lors de la livraison, elle constate que 177 cartons de lait sur une commande de 197 respectent les normes d'hygiène de la maternité.

5. Déterminer un intervalle de confiance à 95% pour la proportion de cartons de lait défectueux :
 - en utilisant les formules vues en cours,
 - en utilisant la fonction `binom.test` de R.Comparer les résultats obtenus par les deux méthodes.

Pour aller plus loin... Peignes d'intervalles de confiance dans le modèle gaussien de moyenne inconnue et de variance connue.

1. Enregistrer une réalisation d'une variable aléatoire réelle distribuée selon une loi normale standard dans m_{inc} . Cette réalisation correspond au paramètre de moyenne inconnue pour lequel on va construire des intervalles de confiance (et les représenter graphiquement).
2. Écrire une fonction dont les arguments d'entrée sont N et n , 2 entiers, a , un niveau de risque et v , une variance et dont la sortie est une matrice de taille $N \times 2$ pour laquelle la i -ème ligne correspond aux bornes inférieure et supérieure d'un intervalle de confiance de niveau $1 - a$ pour le paramètre de moyenne dans le modèle gaussien de variance v et de moyenne m_{inc} .

3. Générer 100 intervalles de confiance de niveau 95% pour le paramètre m_{inc} dans un modèle gaussien de variance $v = 3600$ pour des tailles d'échantillon $n = 30, 50$ et 150 .
4. Représenter les peignes de 100 intervalles de confiance pour ces 3 tailles d'échantillon. Vous adopterez le code couleur suivant :
 - En rouge, si la réalisation de l'intervalle de confiance contient m_{inc} ;
 - En bleu, sinon.
5. Combien d'intervalles de confiance sont de couleur bleue ? Est-ce raisonnable ? Que peut-on observer quand la taille de l'échantillon augmente ?