Statistiques – 4 IF

TD 2 : Test d'hypothèse sur un paramètre

Exercice 1.

Les riverains d'une antenne de téléphonie mobile estiment que le niveau moyen d'ondes émis par l'antenne est nocif pour la santé. Ils souhaitent donc demander le retrait de l'antenne. Le groupe de téléphonie soutient à l'inverse que la limite réglementaire pour le niveau d'ondes n'est pas atteinte.

1. Des experts sont chargés de mettre en place une procédure permettant de trancher entre les deux parties. En raisonnant sur le risque α expliquer pourquoi le choix des hypothèses nulle et alternative peut varier suivant que les experts sont mandatés par le groupe de téléphonie ou par les riverains.

2. Les experts ont relevé 20 fois le niveau des ondes à différents moments de la journée à 150 mètres de l'antenne. On admet que ces niveaux sont distribués selon une loi normale. Le niveau moyen observé est $\overline{x}_n = 0.65$ V/m et l'écart-type observé est $s_n = 0.07$ V/m. Le niveau moyen maximal recommandé par une étude indépendante est de 0.6 V/m. Au vu de ces mesures, les riverains demandent le retrait de l'antenne, mais l'opérateur prétend que le niveau moyen observé n'est pas suffisamment supérieur à 0.6 V/m pour pouvoir conclure avec une confiance de 95 % que le niveau limite toléré est dépassé. Qu'en pensez-vous ?

Exercice 2.

Une machine A fabrique des baguettes métalliques dont la longueur suit une loi gaussienne d'espérance 4mm et d'écart-type 0.10mm.

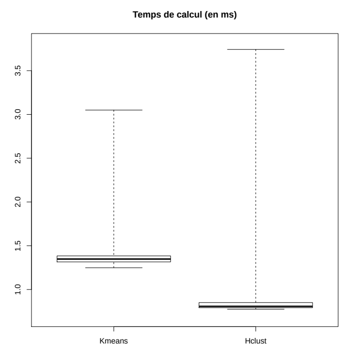
1. Sachant que si les baguettes sont plus grandes que 4.1mm ou plus petites que 3.9mm elles ne sont pas vendables, quelle est la proportion de bonnes baguettes produites par la machine A ?

2. Le prototype d'une nouvelle machine B est testé. Sur un échantillon de 25 baguettes fabriquées, on trouve $\bar{x} = 4mm$ et $s' = 0.08mm$. Tester si au seuil de 5% la machine B est plus précise que la machine A.

Exercice 3. (DS 2018)

On s'intéresse aux algorithmes Kmeans et Hclust, étudiés en cours. Nous cherchons à discriminer les algorithmes selon leur temps de calcul. Notons $T_i^{(K)}$ le temps mis par l'algorithme Kmeans à la i$^{\text{ème}}$ simulation, et $T_i^{(H)}$ le temps mis par l'algorithme Hclust. Les temps seront donnés en millisecondes (ms).

1. Les boxplots des temps $(t_i^{(K)})_{i=1,\ldots,n}$ et $(t_i^{(H)})_{i=1,\ldots,n}$ obtenus sur $n = 100$ simulations sont donnés sur la figure ci-dessous.



   Commenter.

2. On suppose que les $X_i = T_i^{(K)} - T_i^{(H)}$ sont indépendants et de loi normale $\mathcal{N}(m, \sigma^2)$. Pour $n = 30$, on obtient une moyenne empirique $x_n = 0.64$ et une variance empirique $s_n^2 = 3.86$.

   (a) Réaliser le test de (H_0) $m \leqslant 0$ contre (H_1) $m > 0$ au seuil $\alpha = 5\%$. Quel est l'algorithme le plus rapide en moyenne, Kmeans ou Hclust ?

   (b) Quelle commande R permet de réaliser ce test ?

3. On regarde le nombre de fois où l'algorithme Hclust a été plus rapide que l'algorithme Kmeans. Sur $n = 100$ simulations, Hclust a été le plus rapide 86 fois. Peut-on conclure que la probabilité que Hclust soit plus rapide que Kmeans est supérieure à 0.8 ? Ecrire les hypothèses du test réalisé, faire le test et calculer la p-valeur du test. Conclure.