UGA – L2/STA401

Partiel STA401 Mars 2025 – 1h30

Documents autorisés : calculatrice programmable, tables statistiques et deux pages recto (ou une RV) manuscrites et personnelles.

Consignes : il sera tenu compte de la rédaction dans la notation. Le barème n'est qu'indicatif.

Exercice 1 (5 pts)

On s'intérresse ici à la variable X définie comme le nombre de carburateurs (`carb`) du jeu de données `mtcars` disponible dans la librairie de base de R et étudié dans le TP2. Pour un échantillon de $n = 32$ véhicules on a observé les effectifs suivants dans l'échantillon $x_1, ..., x_n$:

m_k	1	2	3	4	6	8
n_k	7	10	3	10	1	1
f_k						
F_k						

1. (3.5 pts) Compléter les lignes indiquant les fréquences f_k et les fréquences cumulées F_k (à 10^{-2} près). Tracer le graphe des fréquences cumulées et en déduire par une lecture de ce graphe les valeurs des trois quartiles.

2. (1.5 pts) Calculer la moyenne et l'écart-type empiriques de l'échantillon à 10^{-2} près (Indication : $\sum x_i^2 = 334$).

Exercice 2 (3 pts)

La population mondiale compte aujourd'hui 7,8 milliards d'habitants. 330 millions sont américains (USA) et on sait que 12% des américains appartiennent au groupe des 1% les plus riches de la planète. Quelle est la proportion d'américains dans la sous-population des 1% les plus riches de la planète ?

Exercice 3 (7 pts) :

Selon les normes de santé européennes le poids d'une certaine molécule dans un comprimé C est une quantité aléatoire X de loi normale centrée en $\mu_0 = 150mg$ et d'écart-type connu $\sigma_0 = 5mg$. On considère un échantillon de taille $n = 25$ de la variable X. Dans toutes les applications numériques on donnera les résultats à $10^{-1}mg$ près.

1. (1.5 pts) Soit $X_1, ..., X_n$ un échantillon aléatoire de la variable X de loi $\mathcal{N}(\mu_0, \sigma_0^2)$ Dans quel intervalle centré en μ_0 a t-on une probabilité de $1 - \alpha$ de trouver la variable $\bar{X}_n$? Donner l'expression de l'intervalle en fonction de μ_0, σ_0, n et α et le calculer pour $n = 25$ et au niveau 95%.

2. (1.5 pts) On veut obtenir un intervalle de largeur inférieure à 2% de μ_0. Montrer qu'il faut pour cela prendre un échantillon de taille supérieure ou égale à $n^* = 43$.

3. (1pt) On a prélevé deux échantillons de taille $n = 25$ dans deux lots de comprimés différents. Dans le premier lot on a observé un poids moyen de $151mg$ et dans le second $152mg$. Ces deux échantillons sont-ils conformes à la norme au niveau 95% ?

4. (3 pts) Dans le second lot de comprimés on note Y la variable poids de la molécule et on suppose que Y est de loi $\mathcal{N}(\mu, \sigma_0^2)$ avec μ inconnu et $\sigma_0 = 5mg$.

   (a) (1pt) Quel est l'estimateur de μ ? Quelle estimation obtient t-on avec l'échantillon prélevé dans le second lot ?

   (b) (1pt) Donner l'expression formelle de l'intervalle de confiance symétrique pour le paramètre inconnu μ au niveau de confiance $1 - \alpha$. Calculer l'intervalle obtenu pour l'échantillon précédent au niveau de confiance 95%.

   (c) (1pt) Si μ_0 n'appartient pas à l'intervalle de confiance de niveau 95% on déclarera que le poids moyen de la substance active d'un comprimé du second lot (c'est à dire μ) est différent de μ_0. Quelle est la probabilité de se tromper avec une telle règle de décision ?

Exercice 4 (6 pts) :

En 2025 on se demande s'il est utile de rajouter des parkings à vélo devant le DLST. Le nombre de U mis en place devant le DLST il y a cinq ans a été calculé sous l'hypothèse que 15% des étudiants du DLST qui viennent en cours à 8h sont cyclistes.

Pour répondre, on se propose d'estimer la probabilité p qu'en 2025 un étudiant du DLST vienne en cours à 8h en vélo. En 2025, sur un échantillon de 100 étudiants se présentant au DLST à 8h, 23 sont venus en vélo. On donnera tous les résultats numériques à 10^{-3} près.

1. (1.5 pts) Rappeler l'expression de l'intervalle de confiance de niveau approximatif $1-\alpha$ pour p. Calculer cet intervalle pour l'échantillon observé au niveau de confiance 90%.

2. (2 pts) Pour quelle valeur de risque α obtiendra-t-on un intervalle de confiance de précision $\pm 1\%$ (c'est à dire de largeur 0.02) pour l'échantillon prélevé en 2025 ?

3. (1 pt) Donner l'intervalle (centré en p_0) de fluctuation de la moyenne empirique d'un échantillon de taille n de la loi de Bernoulli de paramètre p_0 au niveau approximatif $1 - \alpha$. Faire l'application numérique pour $n = 100$, $p_0 = 15\%$ et $\alpha = 10\%$.

4. (1 pt) Pour répondre à la question "faut-il rajouter des U devant le DLST ?", on propose la règle de décision suivante : si $\hat{p}$ n'appartient pas à l'intervalle de fluctuation de niveau 90% (calculé dans la question précédente) alors on concluera $p \neq p_0$. Pensez-vous que l'on puisse déclarer que $p \neq p_0$ avec un risque d'erreur de 10% (justifiez) ?

5. (0.5 pt) Si oui, pourrait-on conclure que $p > p_0$ (c'est à dire qu'il faut rajouter des U devant le DLST) ? Autrement dit, préconiseriez-vous à l'UGA d'installer de nouveaux U pour attacher les vélos ?