Exercices de Statistiques

Kamila AKPAKI

January 22, 2025

Exercice 1 : Loi normale et statistiques descriptives

Énoncé complet

1. On mesure le poids d'un bol de céréales supposé suivre une loi normale $X \sim N(\mu, \sigma^2)$, avec $\mu = 1\,\text{kg}$ et $\sigma = 0.2\,\text{kg}$.

   1. $\overline{X}$ donnera 1 kg et s donnera 0.2.
   2. $\overline{X}$ ne donnera pas 1 kg et s ne donnera pas 0.2.
   3. $\overline{X}$ peut donner 1 kg et s peut donner 0.2.
   4. $\overline{X}$ peut donner 1 kg, mais il est peu probable que s donne 0.2.

2. 25 clients achètent chacun $n = 4$ mesures de céréales. On note $\overline{X}_{\text{global}}$ la moyenne des moyennes des clients et s l'écart-type. Quelle affirmation est correcte ?

   1. $\overline{X}_{\text{global}}$ donnera 1 kg et s donnera 0.2.
   2. $\overline{X}_{\text{global}}$ ne donnera pas 1 kg et s ne donnera pas 0.2.
   3. $\overline{X}_{\text{global}}$ peut donner 1 kg mais s ne peut pas donner 0.2.
   4. $\overline{X}_{\text{global}}$ peut donner 1 kg, et il est fort probable que s donne 0.2.

3. Probabilité que le poids moyen des 4 mesures dépasse 1.2 kg.

   1. Il peut utiliser l'écart-type empirique s (non corrigé).
   2. Il doit utiliser l'écart-type corrigé s.
   3. Qu'il utilise s ou s, il aura les mêmes résultats.
   4. Aucune des réponses précédentes.

4. Nombre minimal de mesures pour que la moyenne empirique soit dans un rayon de 0.15 kg avec 95% de confiance :

$$n = \left(\frac{1.96 \cdot 0.2}{0.15}\right)^2$$

Solution détaillée

Question 1 : Moyenne et écart-type empiriques. - On sait que $\overline{X} \sim N(\mu, \sigma^2/n)$ et s dépend des valeurs individuelles de l'échantillon. - $\overline{X}$ peut donner 1 kg par hasard, mais s ne donnera pas nécessairement 0.2 car il est estimé à partir d'un petit échantillon ($n = 4$). - Réponse correcte : **(d)**.

Question 2 : Moyenne des moyennes. - Avec $n = 25$ clients ayant chacun $n = 4$, $\overline{X}_{\text{global}}$ est une moyenne des moyennes individuelles. - $\overline{X}_{\text{global}}$ a une variance réduite, ce qui rend probable qu'il soit proche de 1 kg. - Pour s, l'échantillon devient plus représentatif et tend vers $\sigma = 0.2$. - Réponse correcte : **(d)**.

Question 3 : Probabilité que $\overline{X} > 1.2$ kg. - On calcule avec $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$.

$$Z = \frac{1.2 - 1}{0.2/\sqrt{4}} = 2.$$

$$P(\overline{X} > 1.2) = P(Z > 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228.$$

- Puisque $n = 4$ est petit, on doit utiliser l'écart-type corrigé. - Réponse correcte : **(b)**.

Question 4 : Taille minimale n. - Formule pour un intervalle de confiance donné :

$$n = \left(\frac{z_{1-\alpha/2} \cdot \sigma}{E}\right)^2,$$

où $z_{1-\alpha/2} = 1.96$, $\sigma = 0.2$, $E = 0.15$. - Calcul :

$$n = \left(\frac{1.96 \cdot 0.2}{0.15}\right)^2 = 6.53^2 = 54.$$

- Réponse correcte : **(b)**.

Exercice 2 : Loi géométrique et estimation paramétrique

Énoncé complet

On suppose que le nombre de candidatures avant un recrutement suit une loi géométrique $X \sim \text{Geom}(p)$.

1. Donner l'information de Fisher :

   (a) $I(p) = \frac{1}{p}$.

   (b) $I(p) = \frac{1}{1-p}$.

   (c) $I(p) = \frac{1}{p(1-p)}$.

(d) $I(p) = p(1-p)$.

2. Trouver l'estimateur du maximum de vraisemblance $\hat{p}_n$:

   (a) $\hat{p}_n = \frac{1}{n\sum_{i=1}^n X_i}$.

   (b) $\hat{p}_n = \frac{1}{\sum_{i=1}^n X_i}$.

   (c) $\hat{p}_n = \frac{n}{\sum_{i=1}^n X_i}$.

   (d) $\hat{p}_n = \frac{1}{n^2\sum_{i=1}^n X_i}$.

3. Identifier la statistique pivotale utilisée :

   (a) $Z = \frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1-\hat{p}_n)/n}}$.

   (b) $Z = \frac{\hat{p}_n - p}{\hat{p}_n}$.

   (c) $Z = \hat{p}_n^2(1-\hat{p}_n)$.

   (d) $Z = \hat{p}_n - p$.

4. Donner l'intervalle de confiance asymptotique pour p au niveau $1-\alpha$:

   (a)
$$\left[\hat{p}_n - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right].$$

   (b)
$$\left[\hat{p}_n - z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right].$$

   (c)
$$\left[\hat{p}_n - z_{\alpha}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right].$$

   (d)
$$\left[\hat{p}_n - z_{\alpha/2}\sqrt{\frac{1}{n}}, \hat{p}_n + z_{\alpha/2}\sqrt{\frac{1}{n}}\right].$$

Solution détaillée

Question 1 : Information de Fisher. La fonction de vraisemblance est $L(p) = p^k(1-p)^{n-k}$. L'information de Fisher est calculée à partir de la dérivée seconde de la log-vraisemblance. Réponse correcte : **(c)**.

Question 2 : Estimateur du maximum de vraisemblance. L'estimateur du maximum de vraisemblance pour une loi géométrique est donné par $\hat{p}_n = \frac{1}{\sum_{i=1}^n X_i}$. Réponse correcte : **(b)**.

Question 3 : Statistique pivotale. La statistique pivotale est $Z = \frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1-\hat{p}_n)/n}}$. Réponse correcte : **(a)**.

Question 4 : Intervalle de confiance. L'intervalle de confiance asymptotique est donné par :

$$\left[\hat{p}_n - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right].$$

Réponse correcte : **(a)**.