Contrôle Continu n°2

Aucun document n'est autorisé. La calculatrice est interdite. Il sera tenu compte du soin apporté à la rédaction et à la présentation. Toute réponse devra être précisément justifiée. Le sujet comporte 2 pages.

Durée de l'épreuve : 1h

Question de cours [1 point]

On dispose d'un modèle statistique dépendant d'un paramètre $\theta \in \mathbb{R}$ inconnu que l'on souhaite estimer.

Vrai ou Faux ? Si $\widehat{\theta_n}$ est un estimateur sans biais de θ alors $\widetilde{\theta_n} = \dfrac{\theta + \widehat{\theta_n}}{2}$ est un estimateur sans biais de θ. Vous justifierez brièvement votre réponse.

Correction :
Faux. La fonction $\widetilde{\theta_n}$ n'est pas un estimateur de θ car elle dépend du paramètre inconnu θ. Parler de biais n'a donc même pas de sens !

Problème [10 points]

Ce sujet est inspiré d'un énoncé proposé dans le cours de M. Fromont : `https://perso.univ-rennes2.fr/system/files/users/fromont_m/PolyTests.pdf`

Contexte :

En janvier 2013, une étude publiée dans la *Revue Américaine de Sociologie* et reprise dans une dépêche AFP affirme que "plus un homme marié accorde de temps aux tâches ménagères, moins il a de relations sexuelles"... Cette dépêche a rapidement été reprise par de nombreux médias français : "Faire le ménage ou l'amour, il faut choisir" (figaro.fr), "Insolite : un homme qui passe l'aspirateur, c'est moins excitant ?" (express.fr) ou "Faire le ménage n'est pas si *sexy*" (europe1.fr).

Cette enquête, menée par Sabino Kornrich, Julie Brines et Katrina Leupp, et intitulée "Égalitarisme, travail ménager et fréquence des rapports sexuels dans le mariage" cherche à démontrer que dans le mariage, la fréquence des relations sexuelles et l'accomplissement des tâches domestiques obéissent à des schémas liés au genre : "il existe une sorte de scénario sexuel bien défini par le genre, dans lequel se conduire selon ce genre est important pour la création du désir sexuel et l'accomplissement de l'acte". Les auteurs étudient ainsi la fréquence des rapports dans le mariage en fonction de la participation des hommes aux travaux domestiques, en différenciant les travaux dits traditionnellement féminins (courses, ménages, cuisine,...), des travaux traditionnellement masculins (entretien de la voiture, jardinage, paiement des factures,...). L'étude stipule que plus un homme marié accorde de temps aux tâches ménagères, moins il a de relations sexuelles et qu'à l'inverse, les couples dans lesquels l'homme participe davantage à des tâches traditionnellement masculines, à une moyenne plus élevée de rapports sexuels. L'enquête conclut cependant que "refuser de participer aux tâches ménagères provoque des conflits dans le couple et l'insatisfaction des épouses", elle-même liée à l'activité sexuelle.

L'étude précédente datant de 2013, nous posons la même question aujourd'hui avec une approche statistique différente. Les chiffres donnés dans la suite de ce sujet sont fictifs.

On admet que le pourcentage de participation aux travaux domestiques traditionnellement féminins d'un homme ayant une fréquence sexuelle mensuelle **inférieure à 4** suit une loi gaussienne d'espérance 30.
On relève les pourcentages de participation aux travaux domestiques traditionnellement féminins de n hommes déclarant avoir **au moins 4** rapports sexuels par mois, et on note $x_1, \ldots, x_n$ ces pourcentages. On suppose que les observations $(x_i)_{i=1,\ldots,n}$ sont issues d'un n-échantillon $(X_1, \ldots, X_n)$ de v.a. indépendantes et de même loi normale d'espérance m et de variance σ^2. Ici, les paramètres $m \in \mathbb{R}^+$ et $\sigma \in \mathbb{R}^+_*$ sont inconnus.

1. *[1.5 pts]* Soit $0 < \alpha < 1$. Déterminer un intervalle de confiance (symétrique) pour m au niveau $(1 - \alpha)$.

   Correction :
   On pose
   $$\hat{\sigma}_n^2 \ := \ \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \,,$$
   la variance empirique corrigée ; et $\hat{\sigma}_n := \sqrt{\hat{\sigma}_n^2}$. Alors, d'après le Théorème de Fisher, $\frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - m)$ suit une loi de Student $\mathrm{St}(n-1)$ à $n-1$ degrés de liberté : en particulier, c'est une fonction pivotale. En notant $q_{\mathrm{St}(n-1)}$ la fonction quantile de la loi $\mathrm{St}(n-1)$, on en déduit
   $$\mathbb{P}\left(q_{\mathrm{St}(n-1)}\left(\frac{\alpha}{2}\right) \leq \frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - m) \leq q_{\mathrm{St}(n-1)}\left(1 - \frac{\alpha}{2}\right) \right) \ = \ 1 - \alpha \,.$$
   Et donc
   $$\mathbb{P}\left(\overline{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} q_{\mathrm{St}(n-1)}\left(1 - \frac{\alpha}{2}\right) \leq m \leq \overline{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} q_{\mathrm{St}(n-1)}\left(\frac{\alpha}{2}\right) \right) \ = \ 1 - \alpha \,,$$
   et comme la loi $\mathrm{St}(n-1)$ est symétrique, on peut écrire $-q_{\mathrm{St}(n-1)}(\alpha/2) = q_{\mathrm{St}(n-1)}(1 - \alpha/2)$. Par conséquent, un intervalle de confiance pour m de niveau $1 - \alpha$ est donné par
   $$\mathrm{IC}_\alpha^{(1)} \ := \ \left[\overline{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}} q_{\mathrm{St}(n-1)}\left(1 - \frac{\alpha}{2}\right) \,,\, \overline{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} q_{\mathrm{St}(n-1)}\left(1 - \frac{\alpha}{2}\right) \right] .$$

2. *[1.5 pts]* Déterminer un intervalle de confiance au niveau $(1 - \alpha)$ pour l'écart-type σ.

   Correction :
   D'après le théorème de Fisher, $(n-1)\frac{\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-1)$, donc en notant $q_{\chi^2(n-1)}$ la fonction quantile de la loi $\chi^2(n-1)$, on a
   $$\mathbb{P}\left(q_{\chi^2(n-1)}\left(\frac{\alpha}{2}\right) \leq (n-1)\frac{\hat{\sigma}_n^2}{\sigma^2} \leq q_{\chi^2(n-1)}\left(1 - \frac{\alpha}{2}\right) \right) \ = \ 1 - \alpha \,.$$
   Et donc
   $$\mathbb{P}\left(\frac{(n-1)\hat{\sigma}_n^2}{q_{\chi^2(n-1)}(1 - \frac{\alpha}{2})} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}_n^2}{q_{\chi^2(n-1)}(\frac{\alpha}{2})} \right) \ = \ 1 - \alpha \,.$$

Par conséquent, un intervalle de confiance pour σ de niveau $1-\alpha$ est donné par

$$\mathrm{IC}_\alpha^{(2)} := \left[\sqrt{\frac{(n-1)\hat{\sigma}_n^2}{q_{\chi^2(n-1)}(1-\frac{\alpha}{2})}}\, , \, \sqrt{\frac{(n-1)\hat{\sigma}_n^2}{q_{\chi^2(n-1)}(\frac{\alpha}{2})}}\right]. \tag{1}$$

3. On souhaite tester l'hypothèse nulle (H_0) : "$m = 30$" contre l'alternative (H_0) : "$m \neq 30$".
   (a) *[0.5 pt]* Que cherche à montrer ce problème de test d'hypothèses ? Vous donnerez une interprétation du choix des hypothèses (H_0) et (H_1) en relation avec le contexte explicité précédemment.

   Correction :
   Avec ce choix d'hypothèse nulle et alternative, on souhaite tester si le pourcentage de participation aux travaux domestiques traditionnellement féminins d'un homme influe sur le nombre moyen de rapports sexuels.

   (b) *[1.5 pts]* À partir d'un intervalle de confiance, construire une région critique $\mathrm{RC}_\alpha^{(1)}$ de niveau α pour ce problème de test. On vérifiera que la région critique est bien de niveau α.

   Correction :
   À partir de l'intervalle de confiance $\mathrm{IC}_\alpha^{(1)}$ au niveau $(1-\alpha)$ pour m , on définit la région critique de niveau α par

$$\mathrm{RC}_\alpha^{(1)} := \left\{(X_1,\dots,X_n)\in\mathbb{R}^n;\ 30\notin \mathrm{IC}_\alpha^{(1)}\right\} ,$$

   ou encore,

$$\mathrm{RC}_\alpha^{(1)} = \left\{(X_1,\dots,X_n)\in\mathbb{R}^n\ ;\ |30-\overline{X}_n| > \frac{\hat{\sigma}_n\, q_{St(n-1)}(1-\frac{\alpha}{2})}{\sqrt{n}}\right\} .$$

   On rejette donc (H_0) au niveau α si et seulement si les observations sont dans la région critique $\mathrm{RC}_\alpha^{(1)}$.
   On a alors

$$\mathbb{P}_{30}\left((X_1,\dots,X_n)\in \mathrm{RC}_\alpha^{(1)}\right) = 1-\mathbb{P}_{30}\left(\mathrm{IC}_\alpha^{(1)}\ni 30\right) = 1-(1-\alpha)=\alpha,$$

   et la région critique $\mathrm{RC}_\alpha^{(1)}$ est bien de niveau α.

4. On souhaite maintenant tester si le pourcentage de participation aux travaux domestiques traditionnellement féminins d'un homme ayant au moins 4 rapports sexuels par mois est égal ou bien inférieur à 30%.
   (a) *[0.5 pt]* Proposer un choix d'hypothèses nulle (H_0) et alternative (H_1).

   Correction :
   On souhaite tester l'hypothèse nulle (H_0) : "$m = 30$" contre l'alternative (H_1) : "$m < 30$".

(b) *[2 pts]* Construire une région critique $\mathrm{RC}_\alpha^{(2)}$ de niveau α. On utilisera un intervalle de confiance.

Correction :
On va rejeter l'hypothèse nulle pour les "petites" valeurs de m. On sait qu'un intervalle de confiance bilatéral au niveau $(1-\alpha)$ pour m est

$$\mathrm{IC}_\alpha^{(3)} = \left]-\infty\, , \overline{X}_n + \frac{\hat{\sigma}_n q_{\mathrm{St}(n-1)}(1-\alpha)}{\sqrt{n}}\right] .$$

En effet, on a

$$\mathbb{P}\big(\mathrm{IC}_\alpha^{(3)} \ni m\big) = \mathbb{P}\left(m \leq \overline{X}_n + \frac{\hat{\sigma}_n}{\sqrt{n}} q_{\mathrm{St}(n-1)}(1-\alpha)\right) = \mathbb{P}\left(\frac{\sqrt{n}}{\hat{\sigma}_n}(\overline{X}_n - m) \geq q_{\mathrm{St}(n-1)}(\alpha)\right) = 1-\alpha\,.$$

À partir de l'intervalle de confiance $\mathrm{IC}_\alpha^{(3)}$ au niveau $(1-\alpha)$ pour m , on définit la région critique de niveau α par

$$\mathrm{RC}_\alpha^{(2)} := \left\{(X_1, \ldots, X_n) \in \mathbb{R}^n;\ 30 \notin \mathrm{IC}_\alpha^{(3)}\right\} ,$$

ou encore,

$$\mathrm{RC}_\alpha^{(2)} = \left\{(X_1, \ldots, X_n) \in \mathbb{R}^n\ ;\ \overline{X}_n - 30 < -\frac{\hat{\sigma}_n q_{St(n-1)}(1-\alpha)}{\sqrt{n}}\right\} .$$

(c) *[1.5 pts]* Calculer la p-valeur $\hat{\alpha}$ du test construit à partir de la région critique $\mathrm{RC}_\alpha^{(2)}$. On exprimera cette p-valeur en fonction de la fonction de répartition d'une loi usuelle.

Correction :
Calculons la p-valeur $\hat{\alpha}$ du test. Comme la fonction $q_{\mathrm{St}(n-1)}(1-\alpha)$ est décroissante et continue en α, $\hat{\alpha}$ satisfait

$$\begin{aligned}
& \tfrac{\hat{\sigma}_n q_{\mathrm{St}(n-1)}(1-\hat{\alpha})}{\sqrt{n}} &= 30 - \overline{X_n} \\
\Leftrightarrow\ & q_{\mathrm{St}(n-1)}(1-\hat{\alpha}) &= \tfrac{\sqrt{n}(30-\overline{X}_n)}{\hat{\sigma}_n} \\
\Leftrightarrow\ & 1-\hat{\alpha} &= \mathbb{P}\left(T_{n-1} \leq \tfrac{\sqrt{n}(30-\overline{X}_n)}{\hat{\sigma}_n}\right) \\
\Leftrightarrow\ & \hat{\alpha} &= 1 - \mathbb{P}\left(T_{n-1} \leq \tfrac{\sqrt{n}(30-\overline{X}_n)}{\hat{\sigma}_n}\right) ,
\end{aligned}$$

où T_{n-1} est une variable de loi de Student à $(n-1)$ degrés de liberté.

5. *Applications numériques*
   (a) *[0.5 pt]* On suppose que l'intervalle de confiance construit à la Question 1 vaut $[21,82\ ;\ 29,87]$ au niveau 95%. Quelle est la conclusion du test construit à partir de la région critique $\mathrm{RC}_\alpha^{(1)}$ pour $\alpha = 5\%$?

   Correction :
   Comme $30 \notin [21,82; 29,87]$, on rejette l'hypothèse nulle selon laquelle $m = 30$ au profit de l'alternative $m \neq 30$ au niveau 5%.

(b) *[0.5 pt]* On suppose que la p-valeur $\hat{\alpha}$ du test associé à la région critique $\mathrm{RC}_{\alpha}^{(2)}$ déterminée en Question 4c vaut 0,024. Conclure.

Correction :
Comme $\hat{\alpha} < 0.05$, on rejette l'hypothèse nulle selon laquelle $m = 30$ au profit de l'alternative $m < 30$ ce qui semble cohérent au vu du résultat de l'étude précédemment citée.