Exercice I

Rappelons que si X_i est une variable aléatoire normalement distribuée de moyenne μ et de variance σ^2, alors $E(X_i) = \mu$ et $Var(X_i) = \sigma^2$. Donc : $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1} \mu = \frac{1}{n}(n\mu) = \mu$

Par conséquent, l'estimateur de μ est non biaisé. Maintenant, vérifions l'estimateur de σ^2. Tout d'abord, notons que nous pouvons réecrire la formule pour les estimateur comme suit : $\hat{\sigma}^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \bar{X}^2$

parce que :

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\bar{X} + \bar{X}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2X_i\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}\left(n\bar{X}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2
\end{aligned}$$

Ensuite, en tenant compte des attentes des estimateur, nous obtenons : $E(\hat{\sigma}^2) = \frac{(n-1)\sigma^2}{n}$

comme illustré ici :

$$E(\hat{\sigma}^2) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right] = \left[\frac{1}{n}\sum_{i=1}^{n} E(X_i^2)\right] - E(\bar{X}^2) \tag{1}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\sigma^2+\mu^2) - \left(\frac{\sigma^2}{n}+\mu^2\right) \tag{2}$$

$$= \frac{1}{n}(n\sigma^2+n\mu^2) - \frac{\sigma^2}{n} - \mu^2 \tag{3}$$

$$= \sigma^2 - \frac{\sigma^2}{n} = \frac{n\sigma^2-\sigma^2}{n} = \frac{(n-1)\sigma^2}{n} \tag{4}$$

La première égalité tient de la forme réecrite de l'estimateur. La seconde égalité tient aux propriées de l'attente. La troisième égalité consiste à manipuler les formules alternatives pour la variance, à savoir :

$Var(X) = \sigma^2 = E(X^2) - \mu^2$ et $Var(\bar{X}) = \frac{\sigma^2}{n} = E(\bar{X}^2) - \mu^2$

Les égalités restantes proviennent d'une simple manipulation algébrique. Maintenant, parce que nous avons montré :

$E(\hat{\sigma}^2) \neq \sigma^2$

l'estimateur de σ^2 est un estimateur biaisé.

Donc : $E(S^2) = E\left[\frac{\sigma^2}{n-1}\cdot\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n-1}E\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n-1}\cdot(n-1) = \sigma^2$

La première égalité est valable parce que nous avons effectivement multiplié la variance de l'échantillon par 1. La deuxième égalité est conforme à la loi de l'espérance qui dit que nous pouvons tirer une constante par l'espérance. La troisième égalité est valable en raison des deux faits que nous avons rappelés ci-dessus.

Exercice II

1. Donnez un intervalle de confiance de 95% pour la mpyenne μ.
   Ici, $\alpha = 0.05$. L'intervalle de confiance est :
   $$\left[\bar{x} \pm z_\alpha \frac{\sqrt{\sigma^2}}{\sqrt{n}}\right] = \left[3250 \pm 1.96\frac{\sqrt{1000}}{\sqrt{12}}\right] = [3232; 3268] \text{ à } 95\%$$
2. Donnez un intervalle de confiance de 99% pour la moyenne. Comparez sa largeur avec celle de l'intervalle de la question précédente. Donner votre interprétation.
   Ici, $\alpha = 0,01$. L'intervalle de confiance est :
   $$\left[\bar{x} \pm z_\alpha \frac{\sqrt{\sigma^2}}{\sqrt{n}}\right] = \left[3250 \pm 2.575\frac{\sqrt{1000}}{\sqrt{12}}\right] = [3226.4936; 3273.5064] \text{ à } 99\%$$
   Cet intervalle est plus large que le précédent. Plus la probabilité que la moyenne appartienne à l'intervalle est élevée ($0,99$ au lieu de $0,95$), plus l'intervalle doit être large. Pour obtenir une plus grande confiance, moins de précision doit être acceptée.

3. Si vous utilisez le même échantillon, un intervalle de confiance d'amplitude de 30 psi de large a été défini, quel serait son niveau de confiance ?

   La largeur d'un intervalle de confiance au niveau $1-\alpha$ est la suivante : $2z_\alpha \frac{\sqrt{\sigma^2}}{\sqrt{n}} = 2z_\alpha \frac{\sqrt{1000}}{\sqrt{12}} = 30$ donc $z_\alpha = \frac{30\sqrt{12}}{2\sqrt{1000}} = 1.6432$ alors $\pi_{z_\alpha} = 0.9498 = \frac{1+(1-\alpha)}{2}$. Donc $\alpha = 0.1004$. Et $1-\alpha = 0.8996$.

4. Quel nombre d'essais minimal serait nécessaire pour estimer μ avec une précision de ± 15 psi, au niveau de confiance 0,95 ?

   Pour n essais, la précision de l'intervalle de confiance au niveau 0.95 est la suivante : $\pm 1.96 \frac{\sqrt{1000}}{\sqrt{n}}$, donc $n = \left(\frac{1.96\sqrt{1000}}{15}\right)^2 = 17.07$.

   La taille de l'échantillon doit être d'au moins 18.

Exercice III

La question de savoir si la cuisson à l'huile d'olive réduit le risque de thrombose est étudiée. Pour cela, le logarithme de la concentration en D-dimères, modélisé par une distribution normale, est considéré. Un échantillon de 9 personnes utilisant régulièrement l'huile de tournesol donne une moyenne de $-0,78$, avec un écart-type de $0,27$. Un échantillon de 13 personnes utilisant régulièrement l'huile d'olive donne une moyenne de $-0,97$, avec un écart-type de $0,32$.

1. Testez si la différence entre les variances observées des deux échantillons est significative ou non au seuil 0.05.

   Le test de Fisher doit être appliqué pour déterminer si la différence entre les variances observées des deux échantillons est significative ou non.

$$\begin{aligned} H_0 &: \sigma_x = \sigma_y \\ H_1 &: \sigma_x \neq \sigma_y \end{aligned}$$

   La statistique du test est :

$$F_{obs} == \frac{\frac{9}{8}0.27^2}{\frac{13}{12}0.32^2} = 0.7393$$

   la valeur tabulée du test est

$$F_{tab}(8,12) =$$

   La valeur observée $1,3526$ étant inférieure, l'hypothèse d'égalité des variances doit être acceptée au seuil de 5%.

2. Au seuil de $0,05$, quel test proposeriez-vous pour déterminer si l'huile d'olive réduit considérablement le risque de thrombose ? Quelle est votre conclusion ?

   Après avoir accepté l'égalité des variances, l'application du test de Student pour l'égalité des attentes est justifiée. En notant par X la variable «logarithme de la concentration en D-dimères pour une personne possédant de l'huile de tournesol», et Y la même variable pour les personnes possédant de l'huile d'olive, les hypothèses suivantes doivent être testées :

$$\begin{aligned} H_0 &: \mu_x = \mu_y \\ H_1 &: \mu_x \neq \mu_y \end{aligned}$$

   La statistique observée du test est :

$$t_{obs} = \frac{\sqrt{n_x+n_y-2}}{\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \frac{\bar{x}-\bar{y}}{\sqrt{n_x S_x^2 + n_y S_y^2}} = 1.3055$$

   La statistique tabulée est :

$$t_{tab} = t\,(\alpha = 0.05; \nu = n_x + n_y - 2) = t\,(\alpha = 0.05; \nu = 9+13-2 = 20) = 2.086$$

   Comme $t_{tab} \geq t_{obs}$; par conséquent, l'égalité des attentes ne peut être rejetée : la diminution moyenne observée n'est pas significative au seuil de 5%.

3. Dans une autre étude portant sur 110 utilisateurs d'huile de tournesol, une moyenne de $-0,82$ a été observée, avec un écart-type de $0,29$, tandis que 130 utilisateurs d'huile d'olive ont une moyenne de $-0,93$, avec un écart-type de $0,31$. Trouvez la valeur du test pour déterminer si l'amélioration est significative. Au seuil 0.05, quelle est votre conclusion ?

   Ceci est un test de comparaison des attentes sur de grands échantillons. La statistique de test est :

$$u_{obs} = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{S_X^2}{n_x}+\frac{S_Y^2}{n_y}}} = 2.8366$$

   La statistique tabulée est : $u_{tab} = 1.96$.

   Comme $u_{obs} \geq u_{tab}$; l'hypothèse nulle H_0 est rejetée et on peut en conclure que l'huile d'olive diminue considérablement le risque de thrombose.

Exercice IV

$H_0: \quad X \mapsto Poisson(\lambda)$
$H_1: \quad X$ ne suit pas une distribution de Poisson.

La moyenne de la distribution de Poisson (supposée) étant inconnue, elle doit être estimée à partir des données de la moyenne de l'échantillon : $\bar{x} = \widehat{\lambda} = 0.75$.

En utilisant la distribution de Poisson avec $\widehat{\lambda} = 0.75$, nous pouvons calculer p_i, les probabilités hypothétiques associées à chaque classe. A partir de celles-ci, nous pouvons calculer les fréquences attendues (sous l'hypothèse nulle) :

$$
\begin{aligned}
p_0 &= P(X=0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472 \Rightarrow E_0 = 0.472 * 60 = 28.32 \\
p_1 &= P(X=1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354 \Rightarrow E_1 = 0.354 * 60 = 21.24 \\
p_2 &= P(X=2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.133 \Rightarrow E_2 = 0.133 * 60 = 7.98 \\
p_3 &= P(X=3) = \frac{e^{-0.75}(0.75)^3}{3!} = 0.041 \Rightarrow E_3 = 0.041 * 60 = 2.46
\end{aligned}
$$

La statistique chi-carré peut maintenant être calculée :

$$\chi^2_{cal} = \sum \frac{(O-E)^2}{E} = 2.94$$

Si nous cherchons dans les tableaux de la distribution chi-deux avec $\nu = k - p - 1 = 3 - 1 - 1 = 1$, nous trouvons $\chi^2_{tab}(\alpha = 0.05) = 3.84$.

Donc, $\chi^2_{cal} \leq \chi^2_{tab}$: nous acceptons l'hypothèse nulle H_0.

Nous concluons qu'il n'existe aucune preuve réelle permettant de penser que les données ne suivent pas une distribution de Poisson, bien que le résultat soit limite.

Fin du solution.

"Qui veut chapitrer le prochain, fasse d'abord son examen." Proverbe italien