Université Grenoble Alpes - (UJF)
L2 Licence Sciences et Technologies - STA401 **MAI 2024**

STA401 - Examen terminal -A- 1ème session - correction

Exercice 1 - "autour du cours"

1. $P(20-a<X<20+a) = P(-\frac{a}{4}<\frac{X-20}{4}<\frac{a}{4})=0,98 \iff P(\frac{X-20}{4}<\frac{a}{4})=0,99 \iff \frac{a}{4}=2,3263 \iff a=9,3052$
2. $1-\alpha = P(a<\overline{X}<b)$; l'intervalle de fluctuation étant [a;b] centré sur μ
3. a) $\overline{x}=19,8$, $s^2=0.96$
   Intervalle de fluctuation de la moyenne, $\alpha=0,01$:
   $\left[\mu \pm u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = \left[20 \pm 2,5758\frac{11}{\sqrt{10}}\right] = [11,04;28,96]$; puisque $\overline{x}$ est dans cet intervalle, la moyenne de cet échantillon est cohérente avec le modèle donné avec une probabilité de 0.99.
4. Par définition, le risque de première espèce est le risque minimisé dans un test : $\alpha = P_{\mathcal{H}_0}(\mathcal{H}_1)$ = P(accepter $\mathcal{H}_1$ alors qu'en réalité $\mathcal{H}_0$ est vraie).
   Donc α = P (faire une fausse joie à tort) = P (déclarer vainqueur / il n'est pas vainqueur) = $P(\ p > 0,5\ /\ p \leq 0,5\ /) = P_{\mathcal{H}_0}(\mathcal{H}_1)$. Donc : $\begin{cases} H_0\ :\ p=\ 0.5 \\ H_1\ :\ p>\ 0.5 \end{cases}$
5. a) pnorm(1.5,1,2)
   b) qt(0.995,19)

Exercice 2 :

Partie A :

1. Calculs : $\hat{\mu}=\overline{x}=0,38$, $\widehat{\sigma^2}=0,001616$.
2. Intervalle de confiance de la moyenne, variance inconnue, $\alpha=0,05$:
   $\left[\overline{x} \pm t^{n-1}_{1-\alpha/2}\frac{s'}{\sqrt{n}}\right] = \left[0,38 \pm 1,985\frac{\sqrt{0,001616}}{\sqrt{100}}\right] = [0,372020;0,3879796]$
3. Test : $\begin{cases} \mathcal{H}_0 :\ \sigma^2 = 0,0016 \\ \mathcal{H}_1 :\ \sigma^2 > 0,0016 \end{cases}$ Test paramétrique de la variance unilatéral.
   La statistique est : $T = \frac{nS^2}{0,0016}$ suit la loi $\mathcal{X}^2_{\mathrm{n}-1}$; Rejet de $\mathcal{H}_0 \ \Leftrightarrow\ T > z^{n-1}_{1-\alpha}$
   La valeur calculée de T est 100 ; $p_{val} = P(T>100) = 0.453$. Donc on accepterait $\mathcal{H}_0$ avec un risque de se tromper β inconnu, mais il faudrait courir des risques supérieur à 45% pour accepter $\mathcal{H}_1$. On peut dire que l'écart type n'est pas supérieure à 0,04.
4. Test : $\begin{cases} \mathcal{H}_0 :\ \mu = 0,4 \\ \mathcal{H}_1 :\ \mu < 0,4 \end{cases}$ Test paramétrique de la moyenne unilatéral (avec variance inconnue).
   La statistique est : $T = \frac{\overline{X}-0.4}{S'/\sqrt{n}}$ suit la loi $\mathcal{T}_{\mathrm{n}-1}$; Rejet de $\mathcal{H}_0 \ \Leftrightarrow\ T < -t^{n-1}_{1-\alpha}$
   La valeur calculée de T est $-4,97494$; $p_{val} = P(T<-4,97494) = 1.376 * 10^{-6}$
   La pvaleur étant très faible, on peut conclure avec un risque très faible de se tromper que la moyenne est inférieure à 0,4.

Partie B :

1. Test de comparaison de deux échantillons indépendants : il y a deux groupes différents de 100 images et 150 images, X_1 et X_2 doivent suivre des lois normales, on veut comparer les moyennes ($\mu_1 > \mu_2$) [images du site B sont moins lourdes que dans le A]. Pour faire ce test, il faut que les variances sont égales.

2. (a) Le test : $\mathcal{H}_0 : \sigma_1^2 = \sigma_2^2$ contre $\mathcal{H}_1 : \sigma_1^2 \neq \sigma_2^2$. La statistique : $T = \dfrac{S_2'^2}{S_1'^2} \rightsquigarrow \mathcal{F}_{(n_2-1;n_1-1)}$

   Calculs des variances estimées : $\widehat{\sigma_1^2} = 0,00161616$ et $\widehat{\sigma_2^2} = 0,00197986$. Valeur de la statistique : $1,225$. Décision : $Rejet de \mathcal{H}_0 \Longleftrightarrow T > f_{(1-\alpha/2)}^{(n_2-1;n_1-1)}$. Ici, $f_{(0,995)}^{(149;99)} \simeq 1,6$ (encadrement avec les tables [1,54 ;1,67]). On ne rejette donc pas $\mathcal{H}_0$, on conclut que les variances sont égales avec une probabilité de 99%.

   (b) On trouve $p_{val} = 2 * P(T > 1.225) = 0,27847$, alors pour $\alpha < 0,27847$ on conclurait que les variances ne sont pas différentes, pour $\alpha > 0,27847$ on conclut que les variances sont différentes. Donc, pour α= 1%, on accepte l'égalité des variances, ce qui est cohérent avec (a).

3. On pose le test : $\mathcal{H}_0 : \mu_1 = \mu_2$ contre $\mathcal{H}_1 : \mu_1 > \mu_2$. Puisque $\sigma_1^2 = \sigma_2^2$ et grands échantillons, la statistique du test est : $T = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0;1)$.

   Rejet de $H_0 \Leftrightarrow T > u_{(1-\alpha)}$. Valeur de la statistique : $24,0943$.
   Avec la table $\mathcal{N}(0;1)$ ou la calculatrice, on trouve : $p_{val} = P(T > 24,0943) \simeq 0$.
   En conséquence, pour α= 1%, $\alpha > p_{val}$, donc on accepte $\mathcal{H}_1$. On conclut donc, avec un risque quasiment nul de se tromper, que le poids moyen des images est plus faible dans le site B.

Exercice 3

1. a) On dispose de deux échantillons appariés car ce sont les 12 mêmes fichiers qui ont été compressés. On veut tester la moyenne des différences, et savoir si elle est significativement positive (algo A - algo B) afin de conclure à l'efficacité de B ou pas. On suppose que $D = X_A - X_B \rightsquigarrow \mathcal{N}(\mu_d, \sigma_d^2)$.

   Le test est : $\begin{cases} \mathcal{H}_0 : \mu_A = \mu_B \\ \mathcal{H}_1 : \mu_A > \mu_B \end{cases} \Longleftrightarrow \begin{cases} \mathcal{H}_0 : \mu_d = 0 \\ \mathcal{H}_1 : \mu_d > 0 \end{cases}$ La statistique est $T = \dfrac{\overline{D}\sqrt{n-1}}{S_d} \rightsquigarrow \mathcal{T}_{n-1}$

   b) Calculs : $\overline{d} = 1,3283$, $s_d = 0,834474$, $T_{calc} = 5,2795$. Lecture du quantile 1.7959. On accepte donc $\mathcal{H}_1$ (B est plus efficace) avec un risque de 5% de se tromper.

2. Calcul de la $p_{valeur} = P_{\mathcal{H}_0}(T > T_{calc}) = 1 - P(T < 5,2795) < 1 - 0,9995$, donc $p_{valeur} < 0,0005$ (valeur exacte avec la calculatrice $1,3.10^{-4}$)
   Pour tous risques $\alpha > p_{valeur}$ (par exemple 5% ou 1%), on accepte $\mathcal{H}_1$. On conclut que l'agorithme B est plus efficace, donc que les fichiers ont des tailles en moyenne plus réduites avec des risques très faibles de se tromper.