Université Grenoble Alpes
L2 - DLST - UGA **Année 2024**

STA401 - Examen terminal - SUJET A (INM) - 1ère session

Durée : 2 heures
Documents autorisés : Tables statistiques non annotées - Calculatrice
Formulaire examen joint avec ce sujet

Exercice 1 : *Autour du cours*

1. Soit X une variable aléatoire de loi Normale $\mathcal{N}(20;16)$. Calculer la valeur a telle que : $P(20-a \leq X \leq 20+a) = 0,98$.
2. Donner la définition du **niveau de confiance** d'un intervalle de fluctuation de la moyenne d'une loi Normale $\mathcal{N}(\mu;\sigma^2)$.
3. Soit X une variable aléatoire qui suit la loi $\mathcal{N}(20;11^2)$.
   Sur un échantillon de 10 individus, on obtient les résultats suivants : $\sum x_i = 198$ et $\sum x_i^2 = 3930$
   a) Calculer la moyenne et la variance empiriques de cet échantillon.
   b) Calculer un intervalle qui permette de déterminer avec une confiance de 99%, si la moyenne de cet échantillon est cohérente avec le modèle donné.
4. Lors d'une élection en France, un candidat vous demande de faire un test statistique pour savoir s'il sera élu ou pas. On rappelle qu'il lui faut **plus de 50%** des votes pour être élu. Vous voulez minimiser le risque de faire une "fausse joie au candidat" à tort, quelles sont les hypothèses du test paramétrique à poser ? Raisonner sur le risque de 1ère espèce pour cela.
5. a) Soit X une variable de loi $\mathcal{N}(\mu;\sigma^2)$, avec $\mu = 1$ et $\sigma^2 = 4$. Donner la fonction **du logiciel** R à utiliser pour obtenir la valeur : $P(X \leq 1.5)$.
   b) Pour construire un intervalle de confiance de la moyenne d'une loi Normale dont la variance est inconnue, au niveau de confiance de 99%, on utilise un échantillon de taille $n = 20$. Donner la fonction à utiliser avec **le logiciel** R pour obtenir le quantile utile dans cet intervalle.

Exercice 2 :

PARTIE A

On étudie le poids des "pub-images" sur un site web A. On prend un échantillon de taille 100. On note X la variables aléatoire représentant le poids des "pub-images"sur A. On suppose que X suit une loi Gaussienne $\mathcal{N}(\mu;\sigma^2)$.

On obtient les résumés numériques suivants : $\sum x_i = 38$ $\qquad \sum x_i^2 = 14,6$

1. Calculer des estimations sans biais de la moyenne et de la variance de X.
2. Calculer un intervalle de confiance de la moyenne au niveau de confiance de 0,95.
3. Faire le test statistique permettant de déterminer si l'écart type est supérieur à 0,04. Conclure selon la pvaleur de ce test.
4. Poser le test statistique permettant de déterminer si la moyenne est inférieure à 0,4. Le mettre en oeuvre et conclure selon la pvaleur de ce test.

PARTIE B

On veut maintenant comparer le site A précédent avec un autre site concurrent B au travers de ce poids des “pub-images”. On reprend l'échantillon du site A précédent. On prend un échantillon pour B de taille 150. On note Y la variable aléatoire représentant ce poids sur B.

On obtient les résumés numériques suivants : $\sum y_i = 37,5 \qquad \sum y_i^2 = 9,67$

On veut comparer les poids moyens dans les deux sites et déterminer si le site B est celui dont le poids des pub est le moins important.

1. Justifier minutieusement la méthode appropriée qui permet de répondre à ce problème. Préciser les conditions de travail nécessaires à la mise en œuvre de cette méthode.
2. (a) Faire le test de comparaison des variances au seuil de α= 1%. Conclure.
   (b) Montrer que ce test donne une p-valeur de 0,27847. Discuter selon les valeurs possibles de α. Que concluez-vous pour α= 1% ? [Vérifier la cohérence avec (a)].
3. Etablir le test des moyennes qui répond au problème. Calculez la p_{valeur} de ce test.
   Conclure explicitement. Puis, donner l'expression littérale pour un risque α= 1%.

Exercice 3 :

Deux logiciels (A et B) de compression de fichiers ont été programmés par deux groupes étudiants. Un des logiciels par des étudiants de filière INF, l'autre par des MIN. Douze même fichiers sont codés et compressés par les deux logiciels. Le tableau ci-dessous donne la taille des fichiers **après** compression par A et par B (en ko) :

A	8,40	10,32	9,25	9,78	9,55	10,48	10,20	9,36	10,56	8,76	8,32	8,18
B	8,04	8,04	9,02	7,68	9,12	9,84	8,88	7,16	8,40	6,36	6,92	7,76

Peut-on dire qu'il y a un logiciel de compression significativement plus efficace que l'autre ?

1. a) Analysez ces données. Justifiez en détails les conditions et les hypothèses du test qui permet de répondre à la question.
   b) Mettre en oeuvre ce test pour un seuil de 5%.
2. Conclure en donnant ***tous*** les risques de première espèce concluant à l'efficacité.

Fiche Examen - STA401 - année 2024

Calculs statistiques :

Moyenne empirique : $\overline{x} = \frac{1}{n}\sum x_i$. Fréquence empirique : $f = \frac{k}{n}$ (k : nb de réalisations parmi n individus)

Variance empirique : $s_x^2 = \frac{1}{n}\left(\sum x_i^2\right) - \overline{x}^2 = \frac{1}{n}\sum (x_i - \overline{x})^2$. Ecart-type empirique : s_x

Probabilités :

$P[A \mid B] = \dfrac{P[A \cap B]}{P[B]}$. A et B sont indépendants ssi : $P[A \mid B] = P[A]$ ou $P[A \cap B] = P[A]P[B]$

En notant $\overline{B}$ l'évènement contraire de B, on a : $P[A] = P[A \mid B]P[B] + P[A \mid \overline{B}]P[\overline{B}]$

Formule de Bayes : $P[A \mid B] = \dfrac{P[B \mid A]P[A]}{P[B]} = \dfrac{P[B \mid A]P[A]}{P[B \mid A]P[A] + P[B \mid \overline{A}]P[\overline{A}]}$

Lois :

X suit la loi Bernoulli $\mathcal{B}(p)$: $P[X = 1] = p$; $P[X = 0] = 1 - p$; $E[X] = p$ et $Var[X] = 1 - p$

X suit la loi Binomiale $\mathcal{B}(n, p)$: $P[X = k] = \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k}; E[X] = np$ et $Var[X] = np(1-p)$

$X \rightsquigarrow \mathcal{N}(\mu, \sigma^2) \Leftrightarrow Y = (X - \mu)/\sigma \rightsquigarrow \mathcal{N}(0, 1)$ (centrage-réduction)

T.C.L : (conditions) : la loi $\mathcal{B}(n, p)$ peut être appochée par la loi $\mathcal{N}(np\,;\, np(1-p))$

Inégalité Markov : X v.a. positive, $a > 0$, $P(X \geq a) \leq \dfrac{E(X)}{a}$

Inégalité Bienaymé Tchebychev : X v.a. positive, $a > 0$, $P(\mid X - E(X) \mid \geq a) \leq \dfrac{V(X)}{a^2}$

Intervalles de fluctuation de p (n grand) et de μ (loi $\mathcal{N}(\mu, \sigma^2)$) pour un niveau de confiance $1 - \alpha$:

$IF(p, \alpha) = \left] p - u_{1-\alpha/2} \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \,;\, p + u_{1-\alpha/2} \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \right[$ $IF(\mu, \alpha) = \left] \mu - u_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}} \,;\, \mu + u_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}} \right[$

Estimateurs ponctuels :

$\widehat{p} = F$	$\widehat{\mu} = \overline{X} = \frac{1}{n}\sum X_i$	$\widehat{\sigma^2} = S'^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum (X_i - \overline{X})^2$

Si $(X_1, ..., X_n)$ est un échantillon gaussien de loi $\mathcal{N}(\mu, \sigma^2)$ alors : $\overline{X} \rightsquigarrow \mathcal{N}(\mu, \dfrac{\sigma^2}{n})$

$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0, 1)$ $\dfrac{\overline{X} - \mu}{S'/\sqrt{n}} \rightsquigarrow \mathcal{T}_{n-1}$ $\dfrac{nS^2}{\sigma^2} \rightsquigarrow \mathcal{X}^2_{n-1}$ (n grand) $\dfrac{F - p}{\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1)$

Estimations par intervalles de confiance pour un niveau de confiance $1 - \alpha$:

$IC(\mu, \alpha) = \left] \overline{X} - u_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}} \,;\, \overline{X} + u_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}} \right[$ $IC(\mu, \alpha) = \left] \overline{X} - t^{n-1}_{1-\alpha/2} \dfrac{S'}{\sqrt{n}} \,;\, \overline{X} + t^{n-1}_{1-\alpha/2} \dfrac{S'}{\sqrt{n}} \right[$

$IC(\sigma^2, \alpha) = \left] \dfrac{nS^2}{z^{n-1}_{1-\alpha/2}} ; \dfrac{nS^2}{z^{n-1}_{\alpha/2}} \right[$ $IC(p, \alpha) = \left] F - u_{1-\alpha/2} \dfrac{\sqrt{F(1-F)}}{\sqrt{n}} \,;\, F + u_{1-\alpha/2} \dfrac{\sqrt{F(1-F)}}{\sqrt{n}} \right[$ (n grand)

$IC(\mu, \alpha) = \left] \overline{X} - u_{1-\alpha/2} \dfrac{S}{\sqrt{n}} ; \overline{X} + u_{1-\alpha/2} \dfrac{S}{\sqrt{n}} \right[$ (n grand)

Risques des tests :

$\alpha = P[rejet\ de\ \mathcal{H}_0 \mid \mathcal{H}_0] = P_{\mathcal{H}_0}[\,accepter\ \mathcal{H}_1\,]$ $\beta = P[accepter\ \mathcal{H}_0 \mid \mathcal{H}_1]$.

Puissance d'un test : $1 - \beta$

Tests paramétriques : Si X suit une loi $\mathcal{N}(\mu, \sigma^2)$

$\mathcal{H}_0$: $\mu = \mu_0$ et σ^2 connue : $T = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$; σ^2 inconnue : $T = \dfrac{\overline{X} - \mu_0}{S'/\sqrt{n}} \rightsquigarrow \mathcal{T}_{n-1}$

$\mathcal{H}_0$: $\sigma^2 = \sigma_0^2$ et σ^2 et μ inconnues : $T = \dfrac{nS^2}{\sigma_0^2} \rightsquigarrow \mathcal{X}^2_{n-1}$

$\mathcal{H}_0$: $p = p_0$ et n grand : $T = \dfrac{F - p_0}{\dfrac{\sqrt{p_0(1-p_0)}}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0,1)$

$\mathcal{H}_0$: $\mu = \mu_0$ contre $\mathcal{H}_1 : \mu > \mu_0$ *Rejet de* $\mathcal{H}_0 \iff T > u_{1-\alpha}$ *si* σ^2 *connue*

$\mathcal{H}_0$: $\mu = \mu_0$ contre $\mathcal{H}_1 : \mu \neq \mu_0$ *Rejet de* $\mathcal{H}_0 \iff T < -t^{n-1}_{(1-\alpha/2)}$ *ou* $T > t^{n-1}_{(1-\alpha/2)}$ *si* σ^2 *inconnue*

$\mathcal{H}_0 : \sigma^2 = \sigma_0^2$ contre $\mathcal{H}_1 : \sigma^2 > \sigma_0^2$ *Rejet de* $\mathcal{H}_0 \iff T > z^{n-1}_{1-\alpha}$

$\mathcal{H}_0 : \sigma^2 = \sigma_0^2$ contre $\mathcal{H}_1 : \sigma^2 \neq \sigma_0^2$ *Rejet de* $\mathcal{H}_0 \iff T < z^{n-1}_{\alpha/2}$ *ou* $T > z^{n-1}_{1-\alpha/2}$

$\mathcal{H}_0 : p = p_0$ contre $\mathcal{H}_1 : p > p_0$ *Rejet de* $\mathcal{H}_0 \iff T > u_{1-\alpha}$ *si n grand*

$\mathcal{H}_0 : p = p_0$ contre $\mathcal{H}_1 : p \neq p_0$ *Rejet de* $\mathcal{H}_0 \iff T < -u_{(1-\alpha/2)}$ *ou* $T > u_{(1-\alpha/2)}$ *si n grand*

Calcul de la p_{valeur} :

Test unilatéral inférieur : $p_{valeur} = P_{\mathcal{H}_0}(T < T_{calc})$. Test unilatéral supérieur : $p_{valeur} = P_{\mathcal{H}_0}(T > T_{calc})$

Test bilatéral (lois Normale et Student) : $p_{valeur} = 2 * [1 - P_{\mathcal{H}_0}(T <\mid T_{calc} \mid)]$

Test bilatéral (lois Khi-deux et Fisher) : $p_{valeur} = 2 * Min\,\{P_{\mathcal{H}_0}(T < T_{calc})\,;\,P_{\mathcal{H}_0}(T > T_{calc})\}$

Test de comparaison de 2 échantillons appariés :

Soient $(X_1, ..., X_n)$ et $(X'_1, ..., X'_n)$ deux échantillons appariés. $D_i = X_i - X'_i$. Conditions : $D \rightsquigarrow \mathcal{N}(\mu_d, \sigma_d^2)$.

$\mathcal{H}_0$: $\mu_d = 0$ contre $\mathcal{H}_1 : \mu_d \neq 0$; $T = \dfrac{\overline{D}\sqrt{n-1}}{S_d} \rightsquigarrow \mathcal{T}_{n-1}$; $\{$*Rejet de* $\mathcal{H}_0 \iff T < -t^{n-1}_{(1-\alpha/2)}$ *ou* $T > t^{n-1}_{(1-\alpha/2)}\}$

Pour un test unilatéral supérieur (si inférieur, échanger X et X') : *Rejet de* $\mathcal{H}_0 \iff T > t^{n-1}_{1-\alpha}$

Test de comparaison de 2 échantillons indépendants :

Soients $(X_1, ..., X_{n_1})$ et $(Y_1, ..., Y_{n_2})$ deux échantillons de loi $\mathcal{N}(\mu_1, \sigma_1^2)$, et $\mathcal{N}(\mu_2, \sigma_2^2)$, indépendants.

1. $\begin{cases} \mathcal{H}_0 : \sigma_1^2 = \sigma_2^2 \\ \mathcal{H}_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$ $T = \dfrac{S_1'^2}{S_2'^2} \rightsquigarrow \mathcal{F}_{(n_1-1;n_2-1)}$; *Rejet de* $\mathcal{H}_0 \iff T < f^{(n_1-1;n_2-1)}_{(\alpha/2)}$ *ou* $T > f^{(n_1-1;n_2-1)}_{(1-\alpha/2)}$

   Pour un test unilatéral supérieur : *Rejet de* $\mathcal{H}_0 \iff T > f^{(n_1-1;n_2-1)}_{(1-\alpha)}$

2. $\begin{cases} \mathcal{H}_0 : \mu_1 = \mu_2 \\ \mathcal{H}_1 : \mu_1 \neq \mu_2 \end{cases}$ *si* $\sigma_1^2 = \sigma_2^2$ $T = \sqrt{\dfrac{n_1+n_2-2}{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\dfrac{\overline{X}-\overline{Y}}{\sqrt{n_1S_1^2+n_2S_2^2}} \rightsquigarrow \mathcal{T}_{n_1+n_2-2}$

   Rejet de $\mathcal{H}_0 \iff T < -t^{(n_1+n_2-2)}_{(1-\alpha/2)}$ *ou* $T > t^{(n_1+n_2-2)}_{(1-\alpha/2)}$(Test unil.sup. : *Rejet de* $\mathcal{H}_0 \iff T > t^{(n_1+n_2-2)}_{(1-\alpha)}$)

   Cas de grands échantillons : $T = \dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1)$; Test unil. sup : *Rejet de* $\mathcal{H}_0 \iff T > u_{(1-\alpha)}$

Test de comparaison de 2 proportions (2 échantillons indépendants) :

Soient p_1 et p_2 les proportions d'un même caractère mesuré sur deux populations différentes de taille $n_1$et n_2 (grands).

$\begin{cases} \mathcal{H}_0 : \; p_1 = p_2 \\ \mathcal{H}_1 : \; p_1 \neq p_2 \end{cases}$ La statistique est $T = \dfrac{F_1 - F_2}{\sqrt{F(1-F)(\dfrac{1}{n_1}+\dfrac{1}{n_2})}} \rightsquigarrow \mathcal{N}(0,1)$ avec $F = \dfrac{n_1F_1 + n_2F_2}{n_1+n_2}$

Test bilatéral : *Rejet de* $\mathcal{H}_0 \iff T < -u_{(1-\alpha/2)}$ *ou* $T > u_{(1-\alpha/2)}$; (Test unil. sup. : *Rejet de* $\mathcal{H}_0 \iff T > u_{(1-\alpha)}$)