
STA401 - Examen terminal - SUJET B (MIN-MAT) - 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 $\mathcal{N}(20; 16)$. Déterminer la valeur a telle que :
 $P(20 - a \leq X \leq 20 + a) = 0,98$.
2. 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.
3. Soit X une variable aléatoire de loi $\mathcal{N}(\mu; \sigma^2)$. Sur un échantillon de taille $n = 37$, on trouve une moyenne empirique de 20 et un écart type empirique 3.
Calculer un encadrement de la moyenne μ vérifiant $P(a \leq \mu \leq b) = 0,99$. Donner une interprétation de cet intervalle ?

Exercice 2 :

Soit X une variable aléatoire qui suit une loi $\mathcal{N}(\mu; \sigma^2)$. La variable X représente le taux d'un indice. On désire savoir si le traitement est efficace en moyenne, donc si la moyenne est descendue à 160, ou si elle est toujours à 200 (valeur de référence). On pose le test suivant : $\mathcal{H}_0 : \mu = 200$ contre $\mathcal{H}_1 : \mu = 160$.

1. Explicitement le risque de première espèce.
2. a) Donner la statistique T de ce test ainsi que sa loi.
b) Montrer que la règle de décision de ce test est :
 $\text{Rejet de } \mathcal{H}_0 \iff T < -t_{1-\alpha}^{n-1} \iff \bar{X} < \mu_0 - t_{1-\alpha}^{n-1} \hat{\sigma} / \sqrt{n}$.
3. Sur un échantillon de taille 30, on trouve une moyenne estimée de X de 181, et un écart type empirique de 51,5. Mettre en oeuvre ce test et conclure pour un seuil de 1% ?
4. a) Expliquer l'intérêt de calculer le risque de seconde espèce pour le test précédent.
b) Calculer ce risque.
c) Déduire la puissance de ce test.
d) Conclure.
5. Calculer la p_{valeur} de ce test. En déduire tous les seuils pour lesquels on rejetterait \mathcal{H}_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 conduisant à l'efficacité.

Exercice 4 :

On étudie le poids des “pub-images” sur deux sites web concurrents A et B. On prend un échantillon de taille 100 mesuré dans sur le site A et un échantillon de taille 150 mesuré sur B. On note X et Y les deux variables aléatoires représentant le poids des “pub-images” mesuré sur A et sur B.

On obtient les résumés numériques suivants :

$$\sum x_i = 38 \quad \sum y_i = 37,5 \quad \sum x_i^2 = 14,6 \quad \sum y_i^2 = 9,67$$

On veut comparer ces 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. Donner toutes les conditions nécessaires à la mise en œuvre de cette méthode.
2. (a) Faire le test de comparaison des variances au seuil de $\alpha = 1\%$. Conclure.
(b) Montrer que ce test donne une p-valeur de 0,27847. Discuter selon les valeurs possibles de α . Que concluez-vous pour $\alpha = 1\%$? [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 $\alpha = 1\%$.

Calculs statistiques :

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

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

Probabilités :

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

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

Formule de Bayes : $P[A | B] = \frac{P[B | A]P[A]}{P[B]} = \frac{P[B | A]P[A]}{P[B | A]P[A] + P[B | \bar{A}]P[\bar{A}]}$

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 $\mathcal{B}(n, p)$: $P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}$; $E[X] = np$ 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 approchée par la loi $\mathcal{N}(np ; np(1 - p))$

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

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

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

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

Estimateurs ponctuels :

$$\hat{p} = F \quad \left\| \quad \hat{\mu} = \bar{X} = \frac{1}{n} \sum X_i \quad \right\| \quad \hat{\sigma}^2 = S'^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

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

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

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

$$IC(\mu, \alpha) = \left[\bar{X} - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{X} + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \quad IC(\mu, \alpha) = \left[\bar{X} - t_{1-\alpha/2}^{n-1} \frac{S'}{\sqrt{n}} ; \bar{X} + t_{1-\alpha/2}^{n-1} \frac{S'}{\sqrt{n}} \right]$$

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

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

Risques des tests :

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

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

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

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

$$\mathcal{H}_0: \sigma^2 = \sigma_0^2 \quad \text{et } \sigma^2 \text{ et } \mu \text{ inconnues : } T = \frac{nS^2}{\sigma_0^2} \rightsquigarrow \chi_{n-1}^2$$

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

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

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

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

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

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

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

Calcul de la p_{valeur} :

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

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

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

Test de comparaison de 2 échantillons appariés :

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

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

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

Test de comparaison de 2 échantillons indépendants :

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

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

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

$$2. \begin{cases} \mathcal{H}_0: \mu_1 = \mu_2 \\ \mathcal{H}_1: \mu_1 \neq \mu_2 \end{cases} \quad \text{si } \sigma_1^2 = \sigma_2^2 \quad T = \sqrt{\frac{n_1 + n_2 - 2}{\frac{1}{n_1} + \frac{1}{n_2}}} \frac{\bar{X} - \bar{Y}}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \rightsquigarrow \mathcal{T}_{n_1+n_2-2}$$

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

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

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).

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

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