Faculté de Droit & des Sciences Économiques — Université de Limoges

ANNÉE UNIVERSITAIRE 2018-2019

1re session — 4e semestre

Licence Economie-Gestion – 2e année

Matière : Statistiques et probabilités – Éléments de correction — **Durée :** 2 heures
Enseignant : Vincent Jalby

Problème

Partie I (25 min, 4,5 points)

1) La population Ω correspond aux étudiants décohabitants de l'Université de Limoges. La variable étudiée X est le loyer du logement. On note $\mu = \mathrm{E}(X)$ le loyer moyen (des étudiants décohabitants de Limoges) et $\sigma = \sigma_X$ son écart-type. L'échantillon (observé) est noté $(x_1, \ldots, x_{90})$. La taille de l'échantillon est $n = 90$.

2) Un estimateur du loyer moyen est la moyenne empirique de l'échantillon :

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

C'est un estimateur sans biais ($\mathrm{E}(\overline{X}) = \mu$) et convergent ($\overline{X} \to \mu$) de μ.

3) La sortie SPSS donne la moyenne observés de l'échantillon ($\bar{x} = 348.03$), son erreur standard se $= 10.001$ correspondant à la précision de l'estimation et l'écart-type de X, $s = 94.876$.

4) Comme l'échantillon est grand ($n = 90 > 50$), on peut supposer que $\overline{X} \hookrightarrow \mathcal{N}(\mu, \sigma/\sqrt{n})$ et $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \hookrightarrow St(n-1)$. D'où l'intervalle de confiance à 95 % du loyer moyen μ :

$$\mathrm{IC}_{0.95}(\mu) = \left[\overline{X} \pm 1.987 \frac{S}{\sqrt{n}}\right] \quad \text{et} \quad \mathrm{ic}_{0.95}(\mu) = \left[348.03 \pm 1.987 \frac{94.876}{\sqrt{90}}\right] = [328.16, 367.90]$$

5) Bien que l'échantillon soit de taille importante, il est nécessaire de supposer que le loyer X suit une loi normale si on souhaite étudier la variance dans la suite.

Partie II (25 min, 4,5 points)

Tout d'abord, on souhaite vérifier que les écarts de loyers ne sont pas trop importants.

1) Un estimateur de la variance des loyers est la variance empirique de l'échantillon

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \quad \text{avec} \quad K^2 = \frac{(n-1)S^2}{\sigma^2} \hookrightarrow \chi^2(n-1)$$

Une estimation de l'écart-type est donnée par $s = \sqrt{s^2} = 94.876$.

2) On obtient alors l'intervalle de confiance de la variance :

$$\mathrm{IC}_{0.95}(\sigma^2) = \left[\frac{(n-1)S^2}{k^2_{0.975}}, \frac{(n-1)S^2}{k^2_{0.025}}\right] \implies \mathrm{ic}_{0.95}(\sigma^2) = \left[\frac{89 \times 9001.471}{118.1}, \frac{89 \times 9001.471}{65.65}\right] = [6783.50, 12203.06]$$

L'intervalle de confiance à 95 % pour l'écart-type des loyers est la racine carrée du précédent, soit $\mathrm{ic}_{0.95}(\sigma) = [82.36, 110.47]$. L'écart-type des loyers est donc compris entre 82.36 et 110.47 € avec un niveau de confiance de 95 %.

3) La sortie STATA reprend les mêmes informations que la sortie SPSS précédente (augmentée de l'IC à 90 % de la moyenne μ). Elle inclut aussi un test sur la variance (ou écart-type) d'hypothèse nulle $H_0 : \sigma = 83$. On peut répondre à la question initiale (les écarts de loyers ne sont pas trop importants), en regardant le test unilatéral H_0 contre $H_1 : \sigma > 83$. On voit que la *p-value* associée à ce test est $0.0277 \approx 3\ \% < 5\ \%$. Cela amène donc à accepter l'hypothèse H_1.

4) Donc, avec un risque de moins de 5 %, on peut conclure que l'écart-type des loyers est supérieur à 83 € (donc présumément important).

Partie III (35 min, 6 points)

On souhaite à présent vérifier si le loyer moyen est significativement inférieur à celui donné par l'OVE.

1) On souhaite donc faire un test sur la moyenne μ. L'hypothèse de *statu quo* est donnée pour l'OVE, $H_0 : \mu = 362$, l'hypothèse de recherche est donnée par l'énoncé $H_1 : \mu < 362$.

2) On utilise la statistique

$$T = \frac{\overline{X} - 362}{S/\sqrt{n}} \hookrightarrow St(n-1) \quad \text{sous l'hypothèse } H_0$$

On acceptera l'hypothèse H_1 si $\overline{X}$ prend des valeurs significativement inférieure à 362 et donc si T prend des valeurs significativement inférieure à 0. D'où

$$\alpha = 5\,\% = P(H_1|H_0) = P(T < \lambda|H_0) \implies \lambda = -t_{0.95} = -1.662 \quad \text{et} \quad t = \frac{\bar{x} - 362}{s/\sqrt{n}} = \frac{348.03 - 362}{94.876/\sqrt{90}} \approx -1.3968$$

La région critique est donc $W =]-\infty, -1.662]$. Comme $t = -1.3968 \notin W$, on ne peut par rejeter H_0. Le risque pris est le risque de seconde espèce β. Le loyer moyen (des étudiants…) n'est pas significativement inférieur à Limoges par rapport à la moyenne nationale.

3) La probabilité critique associée au test est

$$p\text{-}value = P(T < -1.3968) = 1 - P(T < 1.3968) \approx 1 - 0.925 \approx 7.5\,\%$$

4) En prenant un risque de première espèce $\alpha = 10\,\%$, on aurait $p\text{-}value = 7.5\,\% < 10\,\% = \alpha$. Cela nous amènerait à conclure l'hypothèse H_1 : le loyer moyen à Limoges est inférieur à la moyenne nationale.

5) La sortie R correspond à un test bilatéral sur la moyenne. On retrouve la valeur observée de la statistique T, $t = -1.3966$. La probabilité critique bilatérale est $p\text{-}value_b = 0.166$, d'où la valeur unilatérale $p\text{-}value_u = 0.166/2 = 0.083$ qui est proche du résultat trouvé précédemment (0.075).

Partie IV (35 min, 5 points)

1) La population Ω correspond aux étudiants décohabitants de l'Université de Limoges. La variable étudiée X est la variable dichotomique $X(\omega) = 1$ si ω habite seul, $X(\omega) = 0$ si ω habite à plusieurs. Donc $X \hookrightarrow \mathcal{B}(1, p)$ où p représente la proportion d'étudiants habitant seuls.

2) Soit $K = X_1 + \dots X_n$ le nombre d'étudiants habitant seuls parmi l'échantillon $(X_1, \dots, X_n)$. Un estimateur de p est la fréquence empirique $F = K/n$. C'est un estimateur sans biais et convergent de p. Comme $n > 50$, on a $F \underset{\sim}{\hookrightarrow} \mathcal{N}(p, \sqrt{pq/n})$. Une estimation de p est donnée par $f = 37/85 \approx 0.4353$ (en ne tenant pas compte des 5 réponses manquantes).

3) On souhaite tester $H_0 : p = 0.526$ contre $H_1 : p \neq 0.526$ avec un risque $\alpha = 10\,\%$. Pour cela, on utilise la statistique Z :

$$Z = \frac{F - 0.526}{\sqrt{\frac{0.526(1-0.526)}{85}}} \underset{\sim}{\hookrightarrow} \mathcal{N}(0,1) \quad \text{sous } H_0$$

La région critique (bilatérale) du test est donc $]-\infty, -1.645] \cup [1.645, +\infty[$. Comme

$$z = \frac{0.4353 - 0.526}{\sqrt{\frac{0.526(1-0.526)}{85}}} \approx -1.6746 \in W$$

on rejette l'hypothèse H_0 et on accepte l'hypothèse H_1 avec un risque de 10 % : la part d'étudiants de Limoges vivant seuls est différente du pourcentage national.

4) La sortie SAS donne dans le premier tableau les effectifs et pourcentages des étudiants vivant seuls dans l'échantillon (on note bien les 5 données manquantes). Le second tableau correspond au test précédent. La probabilité critique bilatérale est $p\text{-}value = 0.094 < 0.10$. Cela confirme la conclusion précédente de rejeter H_0 avec un risque de 10 %.