

Licence Economie-Gestion – 2^e année

Matière : Statistiques et probabilités – Éléments de correction
Enseignant : Vincent Jalby

Durée : 2 heures

Exercice I (20 min, 4 points)

Cet article fait référence aux marges d'erreur et intervalles de confiance lors de l'estimation d'une proportion.

1) Le modèle statistique est une variable $X \sim \mathcal{B}(1, p)$ où p représente la proportion à estimer. Si (X_1, \dots, X_n) est un n -échantillon de X , alors on estime p à l'aide de la fréquence empirique, estimateur sans biais et convergent de p :

$$F = \frac{1}{n} \sum_{i=1}^n X_i = \frac{K}{n} \sim \mathcal{N}\left(p, \frac{pq}{n}\right) \quad \text{si } n \text{ est grand}$$

L'intervalle de confiance à 95 % pour p et la marge d'erreur sont

$$ic_{0,95}(p) = \left[f \pm z_{0,975} \sqrt{f(1-f)/n} \right] \quad ME = z_{0,975} \sqrt{f(1-f)/n} \quad \text{avec } z_{0,975} = 1.96$$

On en déduit donc que plus n est faible, plus ME est important. Cela confirme l'affirmation du texte « plus l'échantillon est faible, plus la marge d'erreur est grosse ».

L'exemple pris suppose $f = 0.10$. On a donc, si $n = 500$:

$$ME = 1.96 \sqrt{0.10 \times 0.90 / 500} = 2.63 \% \quad ic_{0,95}(p) = [7.37 \%, 12.63 \%]$$

Mais avec $n = 1000$, on obtient

$$ME = 1.96 \sqrt{0.10 \times 0.90 / 1000} = 1.86 \% \quad ic_{0,95}(p) = [8.14 \%, 11.86 \%]$$

Ce qui confirme les valeurs données dans l'article.

2) Deux candidats A et B mesurés respectivement à 10.5 % et 9.5 % auront des scores compris entre $[10.5 \pm 2.6 \%]$ et $[9.5 \pm 2.6 \%]$. Il est donc possible que le premier ne soit en fait qu'à $10.5 - 2.5 = 8 \%$ et le second à $9.5 + 2.5 = 12 \%$

3) Lors d'une utilisation d'un échantillon aléatoire, la représentativité de l'échantillon, s'il est de taille suffisante, est assurée par le hasard.

Problème

Partie I (25 min, 3.5 points)

1) La population étudiée Ω correspond aux locaux commerciaux de la ville de Limoges. La variable étudiée X représente le loyer HT. L'enquête correspond à un n -échantillon (X_1, \dots, X_n) de X .

Le loyer moyen (sur la population) $\mu = E(X)$ peut être estimé à l'aide de la moyenne (empirique) \bar{X} de l'échantillon :

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

C'est un estimateur sans biais et convergent de μ .

La variance du loyer (sur la population) $\sigma^2 = \sigma_X^2$ peut être estimée à l'aide de la variance (empirique) S^2 de l'échantillon :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

C'est un estimateur sans biais et convergent de σ^2 .

Descriptives

	N	Mean	Median	SD	Variance	Minimum	Maximum	Shapiro-Wilk	
								W	p
Loyer	51	141.73	146.00	25.83	667.08	78.00	185.00	0.97	.279

La sortie informatique porte sur un échantillon de $n = 51$ locaux commerciaux ayant un loyer moyen $\bar{x} = 141.73$ et un écart-type $s = 25.83$. Le test de normalité de Shapiro-Wilk ayant une probabilité critique élevée ($p\text{-value} = 0.279 > 5\%$), on ne peut pas rejeter l'hypothèse de normalité de X . Dans la suite, on supposera donc que $X \sim \mathcal{N}(\mu, \sigma)$.

2) La variable X étant supposée normale, on a

$$K^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \implies \text{IC}_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S^2}{k_{1-\alpha/2}^2}, \frac{(n-1)S^2}{k_{\alpha/2}^2} \right]$$

En prenant $\alpha = 10\%$, on a $k_{0.05}^2 = 34.76$ et $k_{0.95}^2 = 67.50$. D'où

$$\text{ic}_{0.90}(\sigma^2) = \left[\frac{50 \times 667.08}{67.50}, \frac{50 \times 667.08}{34.76} \right] = [494.57, 959.55] \implies \text{ic}_{0.90}(\sigma) = [22.24, 30.98]$$

Partie II (30 min, 5 points)

Dans l'édition de 2025 (basé sur les données de 2024), l'observatoire estimait à 134 € HT par m² et par an le loyer d'un local commercial à Limoges (hors emplacement n°1 et hors zones d'activités commerciales). Les données collectées au premier trimestre 2026 démontrent-elles une augmentation du loyer moyen ?

1) Il s'agit d'un test sur la moyenne (du loyer). L'hypothèse nulle, de stabilité est donc $H_0 : \mu = 134$ et l'hypothèse alternative correspond à l'augmentation du loyer, soit $H_1 : \mu > 134$.

2) Il s'agit d'un test t unilatéral à droite. La variable de décision est

$$T = \frac{\bar{X} - 134}{S/\sqrt{n}} \sim St(n-1) \implies t = \frac{141.73 - 134}{25.83/\sqrt{51}} \approx 2.137$$

La région critique est $W =]t_{0.95}, +\infty[=]1.676, +\infty[$. Comme $t \approx 2.137 \in W$, on rejette l'hypothèse H_0 et on accepte H_1 : le loyer moyen des locaux commerciaux a augmenté en 2026 par rapport à 2024. Le risque pris est le risque de première espèce $\alpha = P(H_1|H_0) = 5\%$.

3) La probabilité critique associée au test précédent est

$$p\text{-value} = P(T > 2.137|H_0) = 1 - P(T < 2.137) \approx 1 - 0.98 = 2\%$$

Comme $p\text{-value} = 2\% < 5\%$, cela confirme qu'il faut rejeter H_0 et accepter H_1 . La conclusion aurait été la même avec un risque plus faible de 3% mais pas de 1%.

4) Dans la sortie Stata

```

Stata. ttest Loyer == 134

One-sample t test
-----+-----
Variable | Obs      Mean    Std. err.  Std. dev.  [95% conf. interval]
-----+-----
Loyer   |   51    141.7255   3.616637   25.82795   134.4613   148.9897

      mean = mean(Loyer)
      H0: mean = 134
      Ha: mean < 134
      Pr(T < t) = 0.9812

      t = 2.1361
      Degrees of freedom = 50
      Ha: mean != 134
      Pr(|T| > |t|) = 0.0376

      Ha: mean > 134
      Pr(T > t) = 0.0188
    
```

on retrouve les statistiques descriptives de la sortie jamovi précédente complétées par l'intervalle de confiance pour μ . La statistique T du test sur la moyenne ($H_0 : \mu = 134$) a une valeur $t = 2.1361$. La probabilité critique du test unilatéral à droite ($H_1 : \mu > 134$) est $P(T > t) = 0.0188$ proche du 2% calculé à la question précédente.

Partie III (25 min, 3.5 points)

1) Soit X la variable représentant le loyer des locaux commerciaux de la ville de Limoges et Y celle des locaux commerciaux des zones commerciales. On note μ_X et μ_Y les loyers moyens respectifs, de même pour les écarts-types σ_X et σ_Y .

Descriptives		Zone	N	Mean	Median	SD	Variance	Minimum	Maximum
Loyers	Limoges (Ville)		51	141.73	146.00	25.83	667.08	78.00	185.00
	Zones commerciales		30	132.67	130.00	21.39	457.40	87.00	172.00

On est en présence de deux échantillons (un pour chaque variable). Pour la première (X), on retrouve les données déjà étudiées précédemment ($\bar{x} = 141.73$ et $s_X = 25.83$). Pour la deuxième (Y), sur un échantillon de 30 locaux en zone commerciale, on observe un loyer moyen $\bar{y} = 132.67$ avec un écart-type $s_Y = 21.39$.

On observe donc sur ces échantillons, un loyer moyen plus faible en zone commerciale (132.67 vs 141.73) mais aussi une variabilité un peu plus faible (21.39 vs 25.83).

2) On souhaite donc tester l'égalité des moyennes $H_0 : \mu_X = \mu_Y$ contre $H_1 : \mu_X > \mu_Y$.

Avant cela, on effectue un test de comparaison (homogénéité) des variances $H_0 : \sigma_X = \sigma_Y$ contre $H_1 : \sigma_X \neq \sigma_Y$. La variable de décision est

$$F = \frac{S_X^2}{S_Y^2} \sim F(50, 29) \implies f = \frac{25.83^2}{21.39^2} = 1.46$$

La sortie **jamovi**, sur la ligne *Variance ratio* donne une probabilité critique *p-value* = 0.278 très supérieure au risque habituel (5 ou 10 %) permettant de conclure qu'on ne peut pas rejeter l'hypothèse H_0 d'égalité des variances. Dans la suite, on supposera donc que $\sigma_X = \sigma_Y$.

Homogeneity of Variances Tests					
		F	df	df2	p
Loyers	Levene's	1.73	1	79	.193
	Variance ratio	1.46	50	29	.278

Pour que ce test soit valable, il est nécessaire de supposer la normalité des deux variables X et Y . On a vu que c'était le cas pour X dans la première partie, mais on n'a pas d'information pour Y . Il serait donc préférable d'utiliser le test de Levene, qui ne nécessite pas d'hypothèse de normalité. La probabilité critique de ce second test (*p-value* = 0.193) étant elle aussi très élevée, cela ne change pas la conclusion précédente.

On procède à présent au test de comparaison des moyennes. Comme les variances sont supposées égales, on utilise un test T de Student

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{51} + \frac{1}{30}}} \sim St(51 + 30 - 2) \quad \text{où } S_p \text{ est l'écart-type calculé sur les deux échantillons}$$

D'après la sortie **jamovi**, sur la ligne Student's t, la probabilité critique associée au test bilatéral est *p-value* = 0.109. Celle du test unilatéral à droite ($H_1 : \mu_X > \mu_Y$) est donc *p-value* = 5.45 %. Comme elle est supérieur au risque classique de 5 %, on ne peut pas rejeter l'hypothèse H_0 . Les données collectées ne confirme donc pas que le loyer moyen en zone commerciale est significativement inférieur à celui des locaux commerciaux de la ville.

Independent Samples T-Test				
		Statistic	df	p
Loyers	Student's t	1.62	79.00	.109
	Welch's t	1.70	70.15	.093

Note. $H_a : \mu_{\text{Limoges (Ville)}} \neq \mu_{\text{Zones commerciales}}$

3) La probabilité critique étant très proche du seuil de 5 %, le risque de seconde espèce pris $\beta = P(H_0|H_1)$ est certainement très élevé. Il est donc préférable de ne pas conclure. On pourrait aussi augmenter la taille des échantillons en espérant avoir une conclusion plus nette.