

4

APPLICATIONS SUR LES ANALYSES ASSOCIÉES AU CHAPITRE 10 (ANALYSE UNIVARIÉE, INTERVALLE DE CONFIANCE)

Exercice 1 : intervalle de confiance d'une proportion

Pour estimer la proportion p de foyers qui possèdent, à la date t_0 où est réalisée l'étude, l'appareil *lambda* récemment mis sur le marché, on réalise par prélèvement au hasard un sondage auprès de 2500 foyers et on constate que 500 foyers en possèdent au moins un.

1. Déterminer la nature de la loi que suit le nombre aléatoire N de foyers qui, dans un échantillon de taille 2500 possèdent au moins un appareil *lambda*.

2. Trouver un intervalle de confiance de p de niveau 0,95. [Au préalable, déterminer le nombre b tel que $P(N_{0;1} \leq b) = 0,975$.]

Corrigé de l'exercice 1

1. N suit la loi binomiale $B(2500; p)$ où p est un nombre *certain* de valeur inconnue.

2. Notons $F = N/2500$ la proportion aléatoire d'éléments qui possèdent le caractère considéré « avoir au moins un appareil *lambda* » dans un échantillon de taille 2500.

Page 178, on lit $\frac{F - p}{\sqrt{F(1-F)/2500}} \cong N_{0;1}$. Pour trouver un intervalle de confiance de p avec un niveau de confiance de (0,95), on cherche le nombre $b_{\alpha/2}$ défini par $\Phi(b_{\alpha/2}) = 1 - 0,025$ et donc on lit p. 359 :

$$b_{\alpha/2} = 1,96$$

F prenant la valeur $f = 500/2500 = 0,2$, avec un niveau de confiance de 0,95 on peut affirmer que :

$$f - b_{\alpha/2} \times \sqrt{f(1-f)/2500} \leq p \leq f + b_{\alpha/2} \times \sqrt{f(1-f)/2500}$$

Donc : $0,2 - 0,008 \times 1,96 \leq p \leq 0,2 + 0,008 \times 1,96$, soit $0,18432 \leq p \leq 0,21568$.

Exercice 2 : intervalle de confiance d'une moyenne

Un fabricant de piles électriques affirme que la durée de vie moyenne du matériel qu'il produit est de 170 heures. Un organisme de défense du consommateur prélève au hasard un échantillon de 400 piles et observe une durée de vie moyenne de 165 heures avec un écart-type empirique s de 20 heures. Déterminer un intervalle de confiance de niveau 0,998 pour la durée de vie moyenne de m . Qu'en déduisez-vous *a priori* ?

Corrigé de l'exercice 2

On utilise la propriété « $\frac{\bar{X} - m}{S/\sqrt{n}} \cong N_{0,1}$ » où ici : $n = 400$. \bar{X} prend la valeur 165 et S la valeur 20. Pour trouver un intervalle de confiance de m de niveau $(1 - \alpha)$, on cherche les nombres $b_{\alpha/2}$ tels que $P(N_{0,1} \leq b_{\alpha/2}) = 1 - \alpha/2 = 0,999$ et on lit $b_{\alpha/2} = 3,1$. Donc :

$$0,998 = P[\bar{X} - (3,1) \times S/20 \leq m \leq \bar{X} + (3,1) \times S/20]$$

Avec un niveau de confiance de 99,8 %, on peut affirmer que :

$$165 - (3,1) \times (20/20) \leq m \leq 165 + (3,1) \times (20/20), \text{ soit } 161,9 \leq m \leq 168,1$$

et en déduire que la publicité « $m = 170$ » est *a priori* inexacte. Et l'on pourrait confirmer cette analyse en testant H_0 « $m = 170$ » contre H'_1 « $m < 170$ ».

Exercice 3 : intervalle de confiance d'une proportion

On dispose d'un échantillon de 400 contribuables soumis à l'impôt sur le revenu et auxquels on a demandé s'ils connaissent ou non le montant maximal de réduction d'impôt auxquels ils ont droit pour l'emploi d'une personne à domicile. Le nombre aléatoire X de contribuables qui connaissent ce plafond suit donc une loi $B(400; p)$ où p est la proportion de ceux qui, dans la population P des contribuables soumis à l'impôt sur le revenu, sont informés.

Ayant constaté que 10 personnes de l'échantillon connaissent la réponse, déterminer un intervalle de confiance de p de niveau 0,95.

Corrigé de l'exercice 3

Notons $F = N/400$ la proportion aléatoire d'éléments qui possèdent le caractère considéré.

Page 178, on lit $\frac{F - p}{\sqrt{F(1 - F)/400}} \cong N_{0,1}$. Pour trouver un intervalle de confiance de p avec un niveau de confiance de (0,95), on cherche le nombre $b_{\alpha/2}$ défini par $\Phi(b_{\alpha/2}) = 1 - 0,025$ et donc on lit p. 359 :

$$b_{\alpha/2} = 1,96$$

F prenant la valeur $f = 10/400 = 0,025$, avec un niveau de confiance de 0,95 on peut affirmer que :

$$f - b_{\alpha/2} \times \sqrt{f(1-f)/n} \leq p \leq f + b_{\alpha/2} \times \sqrt{f(1-f)/n}$$

donc : $0,025 - 0,0078 \times 1,96 \leq p \leq 0,025 + 0,0078 \times 1,96$, soit $0,0098 \leq p \leq 0,0402$.

Autre solution plus précise : soit les racines p_1 et p_2 de l'équation du second degré

$$p^2(1 + b_{\alpha/2}^2/n) - p(2f + b_{\alpha/2}^2/n) + f^2 = 0.$$

$$p^2(1 + 1,96^2/400) - p(2 \times 0,025 + 1,96^2/400) + 0,025^2 \text{ soit } 0,014 \leq p \leq 0,045.$$

Exercice 4 : intervalle de confiance d'une moyenne

Une machine produit des pièces en très grande série. Sur un échantillon de 170 pièces, on a relevé une longueur moyenne de 41,3 mm avec un écart-type de 1,3 mm.

Donner un intervalle de confiance de niveau 0,95 pour la longueur moyenne m des pièces produites par la machine. Le logiciel Excel vous donne le résultat suivant :

Arguments de la fonction

INTERVALLE.CONFIANCE

Alpha 0,05 = 0,05

Écart_type 1,3 = 1,3

Taille 170 = 170

= 0,195419088

Renvoie l'intervalle de confiance pour la moyenne d'une population. Consultez l'aide sur l'équation utilisée.

Taille est la taille de l'échantillon.

Résultat = 0,195419088

[Aide sur cette fonction](#) OK Annuler

Corrigé de l'exercice 4

On utilise la propriété « $\frac{\bar{X} - m}{S/\sqrt{n}} \cong N_{0;1}$ » où ici : $n = 170$, \bar{X} prend la valeur 41,3 et S la valeur 1,3. Pour trouver un intervalle de confiance de m_P de niveau $(1 - \alpha)$ on cherche les nombres $b_{\alpha/2}$ tel que $P(N_{0;1} \leq b_{\alpha/2}) = 1 - \alpha/2 = 0,975$ on lit $b_{\alpha/2} = 1,96$.

Donc :

$$0,998 = P[\bar{X} - (1,96) \times S/170^{0,5} \leq m \leq \bar{X} + (1,96) \times S/170^{0,5}]$$

Avec un niveau de confiance de 99,8 %, on peut affirmer que :

$$41,3 - (1,96) \times (1,3/(170^{0,5})) \leq m \leq 41,3 + (1,96) \times (1,3/170^{0,5}),$$

soit $41,05 \leq m \leq 41,495$. Par lecture d'Excel, $41,3 - 0,195 \leq m \leq 41,3 + 0,195$.

Exercice 5 : intervalle de confiance d'une moyenne

Le service commercial d'une société se demande quelle est la satisfaction moyenne de ses clients. Sur un échantillon de 1200 sondés, la satisfaction moyenne est de 7,5 (sur une note de 0 à 10) et l'écart-type standard S est de 2,5.

Construire un intervalle de confiance de la satisfaction moyenne des clients avec un niveau de confiance de 90 %.

Corrigé de l'exercice 5

Il s'agit de déterminer un intervalle de confiance de la valeur moyenne d'un caractère dans une population lorsque l'on dispose d'un échantillon non exhaustif de grande taille extraite de cette population. Pour ce faire, on utilisera la variable $T = \frac{X - m_P}{S/\sqrt{n}}$ dont la distribution est sensiblement égale à la loi normale standard, autrement dit : $T \cong N_{0,1}$ [cf. p 161].

Ici $n = 1200$, \bar{X} et S prennent respectivement les valeurs 7,5 et 2,5. Compte tenu de $P(|N_{0,1}| \leq 1,65) = 2P(N_{0,1} \leq 1,65) - 1 \cong 0,90$, avec une probabilité de 90 %, on a :

$$\frac{|7,5 - m_P|}{2,5/\sqrt{1200}} \leq 1,65. \quad \text{Donc} \quad |m_P - 7,5| \leq 0,119 \quad \text{ou de façon équivalente}$$

$$7,381 \leq m_P \leq 7,619.$$

6

APPLICATIONS SUR LES ANALYSES ASSOCIÉES AUX CHAPITRES 15, 16, ET 17 (ANALYSE MULTIVARIÉE)

Exercice 1 : analyse de variance (voir chapitre 15)

Un dirigeant souhaite connaître les impacts de différents types de publicité sur le lieu de vente envisagé sur le chiffre d'affaires. Ainsi, concevant 3 types de publicité P_1 , P_2 et P_3 qui ont des coûts sensiblement égaux, il assigne à 11 lieux de vente assez semblables ces campagnes de promotion selon la répartition suivante : 3 pour P_1 , 4 pour P_2 et 4 pour P_3 . Le relevé du taux de croissance du chiffre d'affaires de chacun des 11 magasins pour la période des promotions est présenté ci-dessous ; ce taux de croissance δ exprimé en % est calculé par référence au chiffre d'affaires réalisé lors de la période précédente de même durée :

promotion P_1	3,1	3,5	4,0	
promotion P_2	2,8	3,6	4,3	4,8
promotion P_3	2,2	2,5	3,1	3,8

En utilisant la statistique de Fisher, avec un niveau de signification de 5 %, tester l'hypothèse H_0 selon laquelle les promotions ont la même influence sur le taux δ d'accroissement du chiffre d'affaires contre l'hypothèse alternative \bar{H}_0 .

Aide résultats sous Excel. Analyse de variance : 1 facteur.

Rapport détaillé

Groupes	Nombre d'échantillons	Somme	Moyenne	Variance
P1	3	10,6	3,53333333	0,20333333
P2	4	15,5	3,875	0,75583333
P3	4	11,6	2,9	0,5

Analyse de variance

Source des variations	Somme des carrés	Degré de liberté	Moyenne des carrés	F	Probabilité	Valeur critique pour F
Entre Groupes	1,94765152	2	0,97382576	1,86638596	0,21615109	4,45897011
À l'intérieur des groupes	4,17416667	8	0,52177083			
Total	6,12181818	10				

Corrigé de l'exercice 1

					\bar{x}_i	n_i	$\sum_j (x_{ij} - \bar{x}_i)^2$	$n_i(\bar{x}_i - \bar{x})^2$
A₁ « publicité P₁ »	$x_{11}=3,1$	$x_{12}=3,5$	$x_{13}=4$		3,5333	3		
A₂ « publicité P₂ »	$x_{21}=2,8$	$x_{22}=3,6$	$x_{23}=4,3$	$x_{24}=4,8$	3,875	4		
A₃ « publicité P₃ »	$x_{31}=2,2$	$x_{32}=2,5$	$x_{33}=3,1$	$x_{34}=3,8$	2,9	4		
Total						11		

$n = 3 + 4 + 4 = 11$. La somme des carrés expliqués par le facteur promotion ou inter-classes $S_E = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$ prend pour valeur 1,94. La somme des carrés résiduels ou intra-classes $S_R = \sum_j (X_{ij} - \bar{X}_i)^2$ prend pour valeur 4,17. La statistique F de l'analyse de variance $F = \frac{S_E/(k-1)}{S_R/(n-k)}$ suit, sous l'hypothèse H_0 , une loi de Fisher-Snedecor $F(2; 8)$ (cf. page 268). Elle prend pour valeur 1,86, donc le risque de rejet à tort de H_0 est ici de $P_0(F \geq 1,86) = P(F_8^2 \geq 1,86) = 0,216$. Avec un tel risque d'erreur, on ne peut pas rejeter l'hypothèse d'absence d'influence du type de promotion.

Exercice 2 : tests sur les régressions linéaires (voir chapitre 16)

On connaît la consommation annuelle d'un produit au cours des 4 dernières années :

Année t	2008	2009	2010	2011
Quantité q	99	104	108	113

Une représentation graphique permet de constater que les points $M_1 = (1, 99)$, $M_2 = (2, 104)$, $M_3 = (3, 108)$ et $M_4 = (4, 113)$ sont sensiblement alignés.

1. En utilisant le rapport détaillé indiqué ci-dessous, donner l'équation de la droite de régression de q sur t qui résulte de la méthode des moindres carrés ordinaire et en déduire une estimation de la consommation pour 2012.

2. On a le modèle théorique suivant « $Q_t = at + b + \varepsilon_t$ où a et b sont des constantes de valeurs inconnues et où les variables aléa ε_t sont indépendantes et suivent une même loi normale centrée ». Sans utiliser le rapport détaillé rappelé ci-dessous, trouver un intervalle de confiance de a de niveau 0,95 puis comparer le résultat obtenu à celui indiqué ci-dessous.

Rappel. $(\hat{a} - a) \times \sqrt{(n-2)V(t)}/\hat{\sigma}_\varepsilon = t_{n-2}$ (variable de Student) où $\hat{\sigma}_\varepsilon^2$ prend pour valeur $V_R(q)$, la variance résiduelle déduite de la régression linéaire.

Rapport détaillé

Statistiques de la régression	
Coefficient de détermination multiple	1,00
Coefficient de détermination R ²	1,00
Coefficient de détermination R ²	1,00
Erreur-type	0,32
Observations	4,00

Analyse de variance

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	1,00	105,80	105,80	1058,00	0,00
Résidus	2,00	0,20	0,10		
Total	3,00	106,00			

	Coefficients	Erreur-type	Statistique t	Probabilité	Limite inférieure pour seuil de confiance = 95 %	Limite supérieure pour seuil de confiance = 95 %
Constante	94,50	0,39	244,00	0,00	92,83	96,17
Année t	4,60	0,14	32,53	0,00	3,99	5,21

Corrigé de l'exercice 2

1. L'équation de la droite de régression de q sur t qui résulte de la méthode des moindres carrés ordinaire a pour équation « $q_t = at + b$ avec $a = 4,6$ et $b = 94,5$ », où $t = 1$ pour l'année 2008, $t = 2$ pour l'année 2009, etc. Pour 2012, c'est-à-dire pour $t = 5$, on obtient : $q_t = 4,6 \times 5 + 94,5 = 117,5$ qui est une estimation de la consommation pour 2012.

2. On a le modèle théorique suivant : « $Q_t = at + b + \varepsilon_t$ où a et b sont des constantes de valeurs inconnues et où les variables aléa ε_t sont indépendantes et suivent une même loi normale centrée ».

Pour déterminer un intervalle de confiance de a de niveau 0,95, on utilise la propriété :

« $(\hat{a} - a) \times \sqrt{(n-2)V(t)}/\hat{\sigma}_\varepsilon = t_{n-2}$ (variable de Student) » avec $n = 4$.

On cherche $c_{\alpha/2}$ tel que $P(-c_{\alpha/2} \leq t_2 \leq c_{\alpha/2}) = 0,95$ ou $P(t_2 \leq c_{\alpha/2}) = 1 - \alpha/2 = 0,975$.

Donc $c_{\alpha/2} = 4,303$ et $P(-4,303 \leq t_2 \leq 4,303) = 0,95$.

Par suite : $P(-4,303 \leq (\hat{a} - a) \times \sqrt{(n-2)V(t)}/\hat{\sigma}_\varepsilon \leq 4,303) = 0,95$ ou de façon équivalente :

$$P(-4,303 \times \hat{\sigma}_\varepsilon / \sqrt{(n-2)V(t)} + \hat{a} \leq a \leq 4,303 \times \hat{\sigma}_\varepsilon / \sqrt{(n-2)V(t)} + \hat{a}) = 0,95.$$

$V(t)$ étant égal à 1,25 ; \hat{a} et $\hat{\sigma}_\varepsilon$ prenant respectivement les valeurs 4,6 et $(0,2/4)^{0,5}$ on en déduit que :

$$-4,303 \times 0,22/1,58 + 4,6 \leq a \leq 4,303 \times 0,22/1,58 + 4,6.$$

Avec un niveau de confiance de 95 %, on peut considérer que a est compris entre 3,99 et 5,21.