

Exercices corrigés de statistiques inférentielles.

Exercice 1 Induction

Une entreprise fabrique des sacs en plastique pour les enseignes de distribution. Elle s'intéresse au poids maximal que ces sacs peuvent supporter sans se déchirer.

On suppose ici que le poids maximal que ces sacs peuvent supporter suit une loi normale d'espérance mathématique 58 Kg et d'écart-type 3 Kg.

1. Sur 200 sacs reçus, une grande enseigne de distribution constate un poids moyen de 57,7 Kg.
 - 1.1. Donner un intervalle de confiance bilatéral de la moyenne des poids sur un échantillon de taille 200, au seuil de risque 1 %.
 - 1.2. Quelle est votre conclusion sur le poids moyen constaté ?
2. Donner le poids moyen dépassé dans 97 % des cas, sur un échantillon de taille 200.

Solution

1. Ici on travaille sur un échantillon de taille 200 (la taille de la population étant considérée comme infinie).
 - 1.1. La variable aléatoire \bar{X} égale à la moyenne des poids maximaux sur tout échantillon de taille 200 suit la loi normale $N(58 ; \frac{3}{\sqrt{200}}) = N(58 ; 0,212)$.
On cherche $P(58 - a \leq \bar{X} \leq 58 + a) = 0,99$.
Après avoir posé $T = \frac{\bar{X} - 58}{0,212}$ et lu sur la table de la loi normale centrée réduite, on obtient $a = 0,55$.
Donc l'intervalle de confiance sur tout échantillon de taille 200, de la moyenne des poids, est $[57,45 ; 58,55]$.
 - 1.2. Le poids moyen constaté sur l'échantillon ci-dessus est conforme aux attentes (57,7 Kg appartient à l'intervalle).
2. On cherche $P(\bar{X} > b) = 0,97$ donc après calculs on obtient $b = 57,6$.
Donc le poids moyen dépassé dans 97 % des cas est 57,6 Kg.

Exercice 2 Induction

Les résultats d'une enquête, effectuée sur une population de 1500 salariés d'une entreprise, a montré que dans 65% des cas les individus avaient au moins un crédit en cours.

Trouver la probabilité pour que deux échantillons de 200 personnes chacun, indiquent plus de 10 points d'écart entre les proportions de personnes ayant au moins un crédit en cours.

Solution

Ici il s'agit bien d'une induction puisque le résultat de l'enquête est connu. De plus, la taille de la population ($N = 1500$) n'étant pas plus de 10 fois supérieure à la taille de l'échantillon ($n = 200$), il faut utiliser le coefficient d'exhaustivité.

Soit F la variable aléatoire égale à la proportion de personnes de l'entreprise ayant au moins un crédit en cours sur tout échantillon de taille 200. Comme $n \geq 30$, F suit approximativement la loi normale $N(0,65 ; \sqrt{\frac{0,65 \times 0,35}{200} \times \sqrt{\frac{1500 - 200}{1500 - 1}}}) = N(0,65 ; 0,0314)$.

On considère F_1 et F_2 les variables aléatoires indépendantes, de même loi que F , donnant la proportion de personnes ayant au moins un crédit en cours sur les échantillons 1 et 2 de 200 personnes.

On sait, par des propriétés classiques des variables aléatoires, que $F_1 - F_2$ suit une loi normale telle que : $E(F_1 - F_2) = 0$ et $V(F_1 - F_2) = 2V(F) = 0,00197 = 0,044^2$.

On cherche la probabilité $P(|F_1 - F_2| > 0,1)$.

Or

$$P(-0,1 < F_1 - F_2 < 0,1) = P\left(-\frac{0,1}{0,044} < T < \frac{0,1}{0,044}\right)$$

$$P(-0,1 < F_1 - F_2 < 0,1) = P(-2,27 < T < 2,27)$$

$$P(-0,1 < F_1 - F_2 < 0,1) = 2 \times P(T < 2,27) - 1 = 2 \times 0,9884 - 1 = 0,9768.$$

Donc la probabilité que deux échantillons de 200 personnes chacun, indiquent plus de 10 points d'écart entre les proportions de personnes ayant au moins un crédit en cours, est 2,32%.

Exercice 3 Estimation

Afin de mieux gérer les demandes de crédits de ses clients, un directeur d'agence bancaire réalise une étude relative à la durée de traitement des dossiers, supposée suivre une distribution normale.

Un échantillon non exhaustif de 30 dossiers a donné :

Durée mn	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Effectif	3	6	10	7	3	1

1. Calculer la moyenne et l'écart type des durées de traitement des dossiers de cet échantillon.
2. En déduire les estimations ponctuelles de la moyenne m et de l'écart type σ de la population des dossiers.
3. Donner une estimation de m par intervalle de confiance au seuil de risque 5 %.

Solution

1. On a $m_e = 26,3$ et $\sigma_e = 12,3$.

2. L'estimation ponctuelle de la moyenne est donc $\hat{m} = 26,3$ et l'estimation ponctuelle de l'écart type vaut $\hat{\sigma} = \sqrt{\frac{n}{n-1}} \cdot \sigma_e = 12,5$.

3. La variable aléatoire \bar{X} suit la loi normale $N(m; \frac{\sigma}{\sqrt{n}})$. Or ici σ est inconnu donc il faut utiliser la table de la loi de Student.

On cherche la valeur du réel a tel que : $P(\bar{X} - a \leq m \leq \bar{X} + a) = 0,95$.

On pose la variable aléatoire $S = \frac{\bar{X} - m}{\frac{\hat{\sigma}}{\sqrt{n}}}$ qui suit la loi de Student à 29 ddl.

On a $P(-\frac{a}{2,282} \leq S \leq \frac{a}{2,282}) = 0,95$ donc par lecture sur la table de la loi de Student à 29 ddl on obtient $a = 2,0452 \times 2,282 = 4,67$. On obtient finalement l'intervalle de confiance $[21,63 ; 30,97]$.

Remarque : L'utilisation de la table de la loi normale, possible car $n \geq 30$, conduit à l'intervalle $[21,83 ; 30,77]$.

Exercice 4 Estimation

La société G@E a mis au point un logiciel de gestion destiné essentiellement aux PME.

Après une enquête, dans la région Aquitaine, auprès de 100 entreprises déjà équipées d'un matériel informatique (micro-ordinateur) apte à recevoir ce logiciel, la société G@E décide de fixer le prix de vente à 200 €.

Elle espère diffuser son produit auprès de 68% des PME de la région (cette valeur constituera la proportion de ventes sur l'échantillon).

On peut admettre que les 100 PME interrogées constituent un échantillon représentatif des 4 500 PME formant le marché potentiel.

1. Déterminer l'intervalle de confiance de la proportion p des entreprises intéressées par le logiciel, au seuil de risque 1%.
2. Quelle aurait dû être la taille de l'échantillon pour que l'amplitude de l'intervalle de confiance soit de 20 points (erreur de 0,1).

Solution

1. La variable aléatoire F égale à la proportion d'entreprises intéressées par le logiciel sur tout échantillon de taille 100 suit

approximativement la loi normale $N(p ; \sqrt{\frac{0,68 \times 0,32}{100 - 1}}) = N(p ; 0,0469)$.

On cherche la valeur du réel a tel que : $P(F - a \leq p \leq F + a) = 0,99$.

On pose $T = \frac{F - p}{0,0469}$ et T suit la loi normale centrée réduite.

On a alors : $P(-\frac{a}{0,0469} \leq T \leq \frac{a}{0,0469}) = 0,99$ donc par lecture sur la table 2 on a :

$$a = 0,0469 \times 2,576 = 0,12.$$

Donc l'intervalle de confiance de la proportion p des entreprises intéressées par le logiciel, au seuil de risque 1%, est [56% ; 80%].

Remarques :

1. Ici l'échantillonnage est exhaustif mais le coefficient d'exhaustivité a été négligé car la taille de la population est grande par rapport à la taille de l'échantillon.
2. On peut aussi utiliser les abaques disponibles sur le formulaire. On obtient alors le résultat par lecture directe sur la table.

2. La variable aléatoire F égale à la proportion d'entreprises intéressées par le logiciel sur tout échantillon de taille n suit approximativement la loi $N(p ; \sigma^*)$ où $\sigma^* = \sqrt{\frac{p(1-p)}{n}}$.

Comme précédemment, on cherche a tel que $P(-\frac{a}{\sigma^*} \leq T \leq \frac{a}{\sigma^*}) = 0,99$ et donc par lecture sur la table de la loi normale centrée réduite $\frac{a}{\sigma^*} = 2,575$ d'où $a = 2,575 \sigma^*$.

Or $\sigma^* = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$ où p est inconnue.

On démontre¹ que la grandeur $p(1-p)$ est maximale pour $p = \frac{1}{2}$ donc quelle que soit la valeur de p $\sqrt{p(1-p)} \leq \frac{1}{2}$, donc $\sigma^* \leq \frac{1}{2\sqrt{n}}$ et donc $2,575 \sigma^* \leq \frac{2,575}{2\sqrt{n}}$.

Il suffit alors de choisir n tel que $\frac{2,575}{2\sqrt{n}} \leq 0,1$ donc $\sqrt{n} \geq \frac{2,575}{0,2}$ donc $n \geq 165,77$.

En conclusion, si on prend $n = 166$ on est sûr que l'amplitude de l'intervalle de confiance sera inférieure ou égale à 0,2. Cette valeur ne dépend pas de l'échantillon choisi.

¹ On peut par exemple étudier la fonction $x \rightarrow x(1-x)$.