Exercices : Martine Quinio

Estimation et intervalle de confiance

Exercice 1

Un échantillon de 10000 personnes sur une population étant donné, on sait que le taux moyen de personnes à soigner pour un problème de cholestérol élevé est de $7,5\%$. Donner un intervalle dans lequel on soit «sûr» à 95%, de trouver le nombre exact de personnes à soigner sur les 10000.

Correction ▼ [006025]

Exercice 2

Un vol Marseille - Paris est assuré par un Airbus de 150 places ; pour ce vol des estimations ont montré que la probabilité pour qu'une personne confirme son billet est $p = 0.75$. La compagnie vend n billets, $n > 150$. Soit X la variable aléatoire «nombre de personnes parmi les n possibles, ayant confirmé leur réservation pour ce vol».

1. Quelle est la loi exacte suivie par X ?
2. Quel est le nombre maximum de places que la compagnie peut vendre pour que, à au moins 95%, elle soit sûre que tout le monde puisse monter dans l'avion, c'est-à-dire n tel que : $P[X > 150] \leq 0.05$?
3. Reprendre le même exercice avec un avion de capacité de 300 places ; faites varier le paramètre $p = 0.5$; $p = 0.8$.

Correction ▼ [006026]

Exercice 3

Un petit avion (liaison Saint Brieuc-Jersey) peut accueillir chaque jour 30 personnes ; des statistiques montrent que 20% des clients ayant réservé ne viennent pas. Soit X la variable aléatoire : «nombre de clients qui se présentent au comptoir parmi 30 personnes qui ont réservé».

1. Quelle est la loi de X ? (on ne donnera que la forme générale) ; quelle est son espérance, son écart-type ?
2. Donner un intervalle de confiance au seuil 95%, permettant d'estimer le nombre de clients à prévoir.

Correction ▼ [006027]

Exercice 4

Le staff médical d'une grande entreprise fait ses petites statistiques sur le taux de cholestérol de ses employés ; les observations sur 100 employés tirés au sort sont les suivantes.

taux de cholestérol en cg :(centre classe)	effectif d'employés :
120	9
160	22
200	25
240	21
280	16
320	7

1. Calculer la moyenne m_e et l'écart-type σ_e sur l'échantillon.
2. Estimer la moyenne et l'écart-type pour le taux de cholestérol dans toute l'entreprise.
3. Déterminer un intervalle de confiance pour la moyenne.
4. Déterminer la taille minimum d'échantillon pour que l'amplitude de l'intervalle de confiance soit inférieure à 10.

Correction ▼ [006028]

Exercice 5

Sur 12 000 individus d'une espèce, on a dénombré 13 albinos. Estimer la proportion d'albinos dans l'espèce. On comparera les méthodes d'approximation des lois réelles par d'autres lois classiques.

Correction ▼ [006029]

Exercice 6

Une compagnie aérienne a demandé des statistiques afin d'améliorer la sûreté au décollage et définir un poids limite de bagages. Pour l'estimation du poids des voyageurs et du poids des bagages, un échantillon est constitué de 300 passagers qui ont accepté d'être pesés : on a obtenu une moyenne m_e de 68kg, avec un écart-type σ_e de 7 kg.

1. Définir un intervalle de confiance pour la moyenne des passagers. (On admet que le poids des passagers suit une loi normale de moyenne m, d'écart-type σ.)
2. Montrer que l'on peut considérer que le poids des passagers est une variable aléatoire X de moyenne 70 kg, d'écart-type 8 kg.
3. En procédant de même pour le poids des bagages, on admet les résultats :
   – Si le poids maximum autorisé est de 20 kg, le poids des bagages peut être considéré comme une variable aléatoire Y de moyenne 15 kg, d'écart-type 5 kg.
   – La capacité de l'avion est de 300 passagers ; l'avion pèse, à vide, 250 tonnes. Le décollage est interdit si le poids total dépasse 276.2 tonnes. Quelle est la probabilité pour que le décollage soit interdit ?

Correction ▼ [006030]

Exercice 7

Afin de mieux satisfaire leurs clients, une grande société fournisseur d'accès internet fait ses statistiques sur le nombre d'appels reçus en *hotline*, elle pourra ainsi évaluer le temps d'attente pour le client et le nombre d'employés à mettre au standard ; les résultats de l'enquête portent sur 200 séquences consécutives de une minute, durant lesquelles le nombre d'appels moyen a été de 3 appels par minute. On suppose que les appels sont répartis également dans le temps : on partage un intervalle de temps en unités de une seconde ; alors dans chaque unité de temps, il y a au plus un appel.

1. Quelle est la loi de probabilité du nombre d'appels reçus en 4 minutes ?
2. Montrer que l'on peut approcher cette loi par une loi de Poisson.
3. Donner un intervalle de confiance pour le nombre moyen d'appels en 4 minutes.

Correction ▼ [006031]

Correction de l'exercice 1 ▲

Un intervalle dans lequel on soit «sûr» à 95% de trouver le nombre exact de personnes à soigner sur les 10 000 : $[p - y_\alpha\sqrt{\frac{p(1-p)}{n}}; fe + y_\alpha\sqrt{\frac{p(1-p)}{n}}]$. Fréquence entre 65,7% et 94,3%. Donc entre 698 et 802 personnes sur 10000

Correction de l'exercice 2 ▲

La loi exacte suivie par X est une loi binomiale de paramètres : n, p. $E(X) = 0.75n$ et $\operatorname{Var} X = 0.25 \cdot 0.75n$. Comme $n > 150$, on peut faire l'approximation par la loi normale d'espérance $0,75n$ et d'écart-type $\sigma = \sqrt{0.25 \cdot 0.75n}$. $P[X > 150] \leq 0.05$ si $P[X \leq 150] \geq 0.95$ si : $P[\frac{X-0.75n}{\sqrt{0.25 \cdot 0.75n}} \leq \frac{150-0.75n}{\sqrt{0.25 \cdot 0.75n}}] \geq 0.95$. Dans la table de Gauss, on lit $F(1.645) = 0.95$. On n'a plus qu'à résoudre l'inéquation : $\frac{150.5-0.75n}{\sqrt{0.25 \cdot 0.75n}} \geq 1.645$, dont les solutions sont :

$$0 \leq n \leq 187.$$

Ainsi, en vendant moins de 187 billets, la compagnie ne prend qu'un risque inférieur à 5% de devoir indemniser des voyageurs en surnombre. Faisons varier les paramètres, cela ne pose aucun problème :
$N = 150$, $p = 0.5$. n est solution de l'inéquation : $\frac{150.5-0.5n}{\sqrt{0.5.0.5n}} \geq 1.645$. Solution : $n \leq 272$.
$N = 300$, $p = 0.75$. n est solution de l'inéquation : $\frac{300.5-0.75n}{\sqrt{0.25.0.75n}} \geq 1.645$. Solution : $n \leq 381$.
$N = 300$, $p = 0.5$. n est solution de l'inéquation : $\frac{300.5-0.5n}{\sqrt{0.5.0.5n}} \geq 1.645$. Solution : $n \leq 561$.

Correction de l'exercice 3 ▲

1. La loi de X est la loi binomiale $n = 30$, $p = 0.2$.
2. Un intervalle de confiance au seuil 95%, permettant d'estimer le nombre de clients à prévoir : c'est pour la fréquence : 0.657 ; 0.943. Soit entre 20 et 28 personnes. C'est une large fouchette due à n petit.

Correction de l'exercice 4 ▲

1. On obtient, sur l'échantillon, la moyenne $m_e = 214$, l'écart-type $\sigma_e = 55.77$.
2. La moyenne sur l'entreprise est estimée par m_e. L'écart-type est estimé par : $\widehat{\sigma_e} = \sqrt{\frac{100}{99}}55.77 \simeq 56.05$.
3. On en déduit, au seuil 95%, un intervalle de confiance pour la moyenne : $[m_e - y_\alpha\frac{\widehat{\sigma_e}}{\sqrt{n}}; m_e + y_\alpha\frac{\widehat{\sigma_e}}{\sqrt{n}}] = [203.01; 224.99]$. Ainsi le taux moyen de cholestérol est, à un seuil de confiance 95%, située entre 203 et 225 cg.

Correction de l'exercice 5 ▲

Il s'agit ici d'estimer une proportion, suite à une observation qui vaut : $f = \frac{13}{12000} \simeq 1.0833 \times 10^{-3}$.
On peut utiliser une approximation par une loi normale pour la moyenne d'échantillon. On en déduit un intervalle de confiance pour la proportion, au seuil 95% : $I_\alpha = [f - y_\alpha\sqrt{\frac{f(1-f)}{n-1}}; p + y_\alpha\sqrt{\frac{f(1-f)}{n-1}}] \simeq [4.7 \times 10^{-4}, 1.7 \times 10^{-3}]$.
On peut choisir I_α comme intervalle de confiance, au seuil 95%, de la proportion cherchée. Par l'inégalité de Bienaymé-Tchebychev, on a l'intervalle $I = [f - a, f + a]$, avec : $P[|\overline{X} - p| \leq a] \geq 1 - (\frac{\operatorname{Var}\overline{X}}{a^2})$ et $P[|X - p| \leq a] \geq 0.95$ si $1 - \frac{\operatorname{Var}\overline{X}}{a^2} \geq 0.95$, soit $a \geq 1.3979 \times 10^{-3}$. On préfèrera donc la première méthode.

Correction de l'exercice 6 ▲

1. On peut estimer m par la moyenne de l'échantillon : 68 kg, et σ par $\sigma_e\sqrt{\frac{300}{299}} = 7\sqrt{\frac{300}{299}} \simeq 7.0117$ kg. On en déduit un intervalle de confiance pour la moyenne m : $I_\alpha = [67.2; 68.8]$.
2. La borne supérieure de l'intervalle étant de 69 kg, il est raisonnable de prendre 70 kg comme espérance de la variable poids d'un passager.

3. Le décollage est autorisé si le poids total des voyageurs et de leurs bagages ne dépasse pas 26.2 tonnes. Pour chacun des 300 passagers, notons : X_i son poids et Y_i le poids de ses bagages. Faisons l'hypothèse d'indépendance entre les variables X_i et Y_i. Le poids total $Z = \sum_{i=1}^{300}(X_i + Y_i)$ est la somme de 600 variables aléatoires indépendantes ; le théorème central limite s'applique sous cette hypothèse. Comme l'espérance totale est $E(Z) = 300 \cdot (70+15) = 25\,500$ et la variance de Z est : $\operatorname{Var} Z = 300 \cdot (\operatorname{Var} X_i + \operatorname{Var} Y_i)$. Alors Z suit approximativement une loi normale de moyenne $m = 25\,500$, d'écart-type $\sigma = \sqrt{300 \cdot (8^2+5^2)} = 163.4$. Alors $Z' = \frac{Z-m}{\sigma}$ suit approximativement une loi normale centrée réduite. Le décollage est interdit si : $Z > 26\,200$, c'est-à-dire si $Z' > 4.284$. On lit dans la table de Gauss : pour $t = 4$, $F(t) = 0.999\,968 = P[Z' \leq 4]$. Le décollage est interdit pour cause de surcharge pondérale avec une probabilité inférieure à $0.000\,04$.

Correction de l'exercice 7 ▲

1. L'intervalle de temps de 4 minutes est la répétition de 240 secondes, au cours desquelles les appels surviennent de façon indépendante, avec la probabilité d'appel de $\frac{1}{20}$; la loi de probabilité du nombre d'appels reçus en 4 minutes est donc une loi binomiale, de paramètres $n = 240$ et $p = \frac{1}{20}$.
2. Comme $n \geq 30$ et $np \leq 15$, il est possible d'approcher cette loi par une loi de Poisson de paramètre λ estimé par $np = 12$.
3. Un échantillon de taille 200 a été réalisé pour estimer le nombre moyen d'appels par minute ; c'est un échantillon de taille 50 pour la variable précédente (nombre d'appels reçus en 4 minutes) qui suit une loi de Poisson d'espérance et de variance 12. Un intervalle de confiance au niveau 95% pour la moyenne est $I_\alpha = [11; 13]$.
