IUT HSE | Probabilités et statistiques
UNIVERSITÉ DE BORDEAUX 1 | 2013/2014

Préparation du contrôle du second semestre - exercices corrigés.

On donne: si Z suit $\mathcal{N}(0,1)$, alors $P[Z > 1.6449] = 0.05$, $P[Z > 2.3263] = 0.01$. $P[|Z| > 1.96] = 0.05$ et $P[|Z| > 2.5758] = 0.01$.

Exercice 1. Soit X une variable aléatoire réelle continue ayant pour densité de probabilité

$$f(x) = \begin{cases} \lambda x & \text{si } 0 \leq x \leq 1 \\ 0 & \text{ailleurs} \end{cases}.$$

(1) Calculez λ.
(2) Calculez la moyenne de X.
(3) Calculez $\mathbb{P}(X \in [0, 1/2])$.

Exercice 2. La durée d'attente à une caisse de supermarché est assimilée à une loi exponentielle. La variable aléatoire égale au délai d'attente suit une loi exponentielle de paramètre $\lambda = 0,04\, min^{-1}$.

(1) Quelle est la probabilité qu'un client attende moins de cinq minutes ?
(2) Quelle est la probabilité qu'il attende plus de 15 minutes ?
(3) Quel est le temps d'attente moyen ?

Exercice 3. Le staff médical d'une grande entreprise fait des 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)	120	160	200	240	280	320
effectif	9	22	25	21	16	7

(1) Estimez la moyenne et l'écart-type empiriques de l'échantillon.
(2) Donnez les intervalles de confiance à 95% et 99% pour la moyenne.

Exercice 4. On souhaite vérifier si la prise de poids d'un jeune mouton en un an (variable Y en kilogrammes) dépend de son poids initial (variable X également en kilogrammes). Sur 100 moutons, on donne les résultats suivants :

$$\sum x_i = 3560; \sum y_i = 4230; \sum x_i^2 = 127200; \sum y_i^2 = 180100; \sum x_i y_i = 151320.$$

(1) Calculez la moyenne et l'écart type empiriques de X et Y.
(2) Calculer le coefficient de corrélation linéaire.
(3) Estimer les paramètres a, b pour la régression linéaire de Y sur X.
(4) Selon ce modèle combien un mouton de poids initial 37 kg devrait-il prendre de poids ?

Exercice 5. Chez un individu adulte, le logarithme du dosage en d-dimères, variable que nous noterons X, est modélisé par une loi normale d'espérance μ et de variance σ^2. La variable X est un indicateur de risque cardio-vasculaire: on considère que chez les individus sains, $\mu = 1$, alors que chez les individus à risque, $\mu = 0$. Dans les deux cas, la valeur de σ^2 est la même : $\sigma^2 = 0.09$.

Le Dr. House ne souhaite pas alarmer inutilement ses patients. Quelles hypothèses H_0 et H_1 choisira-t-il de tester ? Donner la règle de décision pour son test, au seuil de 1%, et au seuil de 5%.

Solution de l'exercice 1. 1) Il faut $f \geq 0$ donc $\lambda \geq 0$ et $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$ soit

$$1 = \int_0^1 \lambda x\,\mathrm{d}x = \left[\frac{\lambda x^2}{2}\right]_0^1 = \frac{\lambda}{2}$$

d'où $\lambda = 2$.

2) $\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)\,\mathrm{d}x = \displaystyle\int_0^1 2x^2\,\mathrm{d}x = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$

3) $\mathbb{P}(X \in [0,1/2]) = \displaystyle\int_0^{1/2} 2x\,\mathrm{d}x = \left[x^2\right]_0^{1/2} = \frac{1}{4}.$

Solution de l'exercice 2. Rappelons que la loi exponentielle de paramètre λ est donnée par la densité

$$f(x) = \begin{cases} \lambda e^{-\lambda t} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}.$$

1) $P(X \leq 5) = \int_0^5 0.04e^{-0.04t}\,\mathrm{d}t = [-e^{-0.04t}]_0^5 = -e^{-0.2} + 1 = 0.18$

2) $P(X \geq 15) = 1 - P(X \leq 15) = 1 - \int_0^{15} 0.04e^{-0.04t}\,\mathrm{d}t = 1 - [-e^{-0.04t}]_0^{15} = 1 + e^{0.6} - 1 = 0.55$

3) $E[X] = \int_0^{+\infty} 0.04te^{-0.04t}\,\mathrm{d}t = [-te^{-0.04t}]_0^{+\infty} + \int_0^{+\infty} e^{-0.04t}\,\mathrm{d}t$
$= -0 + 0 + \left[-\frac{1}{0.04}e^{-0.04t}\right]_0^{+\infty} = \frac{1}{0.04} = 25min$

Solution de l'exercice 3. 1) $\tilde{\mu} = \frac{120\times 9+160\times 22+\cdots+320\times 7}{100} = 214cg$

$\sigma^2 = \frac{120^2\times 9+160^2\times 22+\cdots+320^2\times 7}{100} - 214^2$ donc $\sigma = 55,8$.

2) Les intervalles de confiance sont de la forme $[\tilde{\mu} - u_\alpha\frac{\sigma}{\sqrt{n}}, \tilde{m}u + u_\alpha\frac{\sigma}{\sqrt{n}}]$ où $u_\alpha = 1.96$ pour 95% et $u_\alpha = 2.5758$ pour 99%.

En effet, les données sont une réalisation de la variable aléatoire $(X_1,\ldots,X_{100})$ où $X_1,\ldots,X_{100}$ sont 100 copies indépendante de la variable aléatoire X: taux de cholestérol chez un employé.

La moyenne empirique est une réalisation de la variable aléatoire $Y = \dfrac{X_1+\cdots+X_{100}}{100}$.

Soit $Z = \dfrac{X_1-\mu+\cdots+X_{100}-\mu}{\sigma\sqrt{100}}$. D'après le théorème centrale limite, Z suit (à peu près) la loi normale $\mathcal{N}(0,1)$ De plus $\dfrac{\sigma Z}{\sqrt{n}} = Y - \mu$ et comme $\tilde{\mu}$ est une réalisation de Y,

$$P(|\mu - \tilde{\mu}| \geq \lambda) = P\left(\left|\frac{\sigma Z}{\sqrt{n}}\right| \geq \lambda\right) = P\left(|Z| \geq \lambda\frac{\sqrt{n}}{\sigma}\right)$$

Comme Z suit $\mathcal{N}(0,1)$ on sait déterminer u_α pour que $P\left(|Z| \geq u_\alpha\right) = \alpha$ et on prend λ pour que $\lambda\dfrac{\sqrt{n}}{\sigma} = u_\alpha$ soit $\lambda = u_\alpha\frac{\sigma}{\sqrt{n}}$.

Solution de l'exercice 4. 1) Les moyennes sont $\mu_X = \frac{1}{100}\sum x_i = \frac{3560}{100} = 35.6$ et $\mu_Y = \frac{1}{100}\sum y_i = \frac{423}{100} = 42.3$.

Les écarts-types sont $\sigma_X^2 = \frac{1}{100}\sum x_i^2 - \mu_X^2 = 4.64$ d'où $\sigma_X = 2.15$ et $\sigma_Y^2 = \frac{1}{100}\sum y_i^2 - \mu_Y^2 = \sum y_i^2 = 11.71$ d'où $\sigma_Y = 3.42$.

2) $\sigma_{XY} = \frac{1}{100}\sum x_i y_i - \mu_X\mu_Y = 7.32$ d'où $\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y} = 0.99$ ce qui indique une très forte corrélation.

3) $a = \frac{\sigma_{XY}}{\sigma_X^2} = 1.58$ et $b = \mu_Y - \mu_X\frac{\sigma_{XY}}{\sigma_X^2} = -13.86$.

4) La droite de régression est $y = ax + b$, selon ce modèle, si $x = 37$, $y = 1.58 \times 37 - 13.86 = 44.6kg$. La prise de poids est donc de $44.6 - 37 = 7.6kg$.

Solution de l'exercice 5. Si Dr. House ne veut pas alarmer inutilement un patient, l'hypothèse qu'il considère comme dangereux de rejeter à tort est: *le patient n'est pas à risque*, donc que sa variable X (la statistique de test) a pour espérance 1. Son hypothèse H_0 est donc $\mu = 1$ (le patient ne prsente pas de risque), qu'il teste contre $H_1 : \mu = 0$ (le patient présente un risque). Il choisira de rejeter des valeurs trop élevées de X. La règle de décision sera donc

$$\text{Rejet de } H_0 \quad \Leftrightarrow \quad X > \ell,$$

où $P_{H_0}[X > l] = \alpha$.

Sous l'hypothèse H_0, la statistique de test X suit la loi $\mathcal{N}(1, 0.09)$, donc $\dfrac{X(1)}{\sqrt{0.09}}$ suit la loi $\mathcal{N}(0, 1)$. Une règle de décision équivalente est :

$$\text{Rejet de } H_0 \quad \Leftrightarrow \quad \frac{X(1)}{\sqrt{0.09}} > \frac{\ell(1)}{\sqrt{0.09}}.$$

Donc $\frac{\ell(1)}{\sqrt{0.09}}$ est la valeur qui a probabilité α d'être dépassée pour une variable aléatoire Z de loi $\mathcal{N}(0, 1)$: $P[Z > 1.6449] = 0.05$, $P[Z > 2.3263] = 0.01$. Au seuil 0.05% la règle de décision du test est:

$$\text{Rejet de } H_0 \quad \Leftrightarrow \quad \frac{X(1)}{\sqrt{0.09}} > 1.6449 \Leftrightarrow X > 1.6449\sqrt{0.09} - 1 = 0.5065.$$

On déclare que le patient présente un risque cardio-vasculaire quand son dosage en d-dimères est supérieur à 0.5065.

Au seuil 0.01 la règle de décision du test est :

$$\text{Rejet de } H_0 \quad \Leftrightarrow \quad \frac{X(1)}{\sqrt{0.09}} > 2.3263 \Leftrightarrow X > 2.3263\sqrt{0.09} - 1 = 0.30.21.$$

Plus le seuil est faible, moins la règle de décision rejette d'individus à risque. Ce qui doit se produire pour rejeter $\mu = 1$ au seuil 0.01 est plus extraordinaire qu'au seuil 0.05.