UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Prova Final de Probabilidade e Estatística **05-12-2018**
Atenção: Não serão aceitas respostas sem justificativa.
Resolver as questões nos espaços apropriados.

Q1) A partir de uma variável aleatória U uniformemente distribuída sobre o intervalo $[0, 1]$, defina duas variáveis aleatórias X e Y da seguinte maneira:

$$X = \begin{cases} 1, \text{se } U \leq 0,3 \\ 0, \text{caso contrário} \end{cases} \quad \text{e} \quad Y = \begin{cases} 1, \text{se } U \leq 0,6 \\ 0, \text{caso contrário} \end{cases} \quad .$$

(a) Obtenha a função de probabilidade conjunta do par (X, Y).

(b) As variáveis aleatórias X e Y são independentes? Justifique.

(c) Calcule $\text{Cov}(X, Y)$.

(d) Calcule $P(X = 1|Y = 1)$ e $P(Y = 0|X = 0)$.

Q2) A distribuição X da temperatura (em graus Celsius) de solidificação de uma substância pode ser modelada por uma Normal. Não se tem conhecimento dos parâmetros de X, mas sabe-se que P(X<50)=0,0227 e P(X>80)=0,1586.

(a) Calcule os parâmetros μ e σ da distribuição de X.

(b) Qual a probabilidade de X pertencer ao intervalo I = [55; 80]?

(c) Se 40 medições da temperatura de solidificação dessa substância forem realizadas, qual a probabilidade da variável X pertencer ao intervalo I em pelo menos 25 das 40 repetições?

Q3) Deseja-se estimar simultaneamente:

i) A média μ, em mg/l, da concentração da substância S em determinado produto químico; e

ii) A proporção p de unidades desse produto nas quais a concentração de S é superior a 5,5 mg/l.

Para isso foi coletada uma amostra piloto composta por 10 unidades amostrais desse produto, para as quais a concentração da substância S, em mg/l, foi medida. Eis os resultados: 12,03 7,75 6,97 7,06 5,86 7,99 4,30 6,12 4,43 5,73. Para simplificar os cálculos, informam-se: $\sum x_i = 68,24$; mg/l e $\sum x_i^2 = 509,79$ (mg/l)^2.
Pergunta-se:

(a) Quais são as estimativas da média μ e do desvio padrão σ da concentração de S, bem como da proporção p, com base na amostra piloto?

(b) Qual o tamanho mínimo de uma nova amostra com base na qual se possa estimar a proporção p de forma que o erro absoluto seja inferior a 0,08 com 95% de probabilidade?

(c) Se for utilizada essa mesma amostra dimensionada no item (b), qual a probabilidade de que na estimação de μ seja cometido um erro máximo de 0,3 mg/l?

Admita que o tamanho da amostra calculado no item (b) é suficientemente grande para que seja válida a aproximação fornecida pelo Teorema Central do Limite.

Q4) O atual tempo de travessia de barcas entre Rio de Janeiro e Niterói é considerado uma variável aleatória com distribuição Normal de média 15 minutos e desvio padrão de 3 minutos. Uma nova barca da empresa concorrente *MSC Barcas* vai entrar em operação e os responsáveis garantem que será mais rápida que as anteriores. Teste a hipótese de que o tempo médio de travessia é igual ou superior a 15 minutos *versus* a alternativa de que ele é inferior a 15 minutos.

(a) Para uma amostra de 20 tempos de travessia com a nova barca, obteve-se um tempo médio de 14,5 minutos. Obtenha a região crítica (ou região de rejeição) do teste considerando um nível de 5%. Tome sua decisão.

(b) Obtenha o p-valor do teste. O resultado está coerente com a conclusão do item (a)? Justifique.

(c) Calcule a probabilidade do erro tipo II, se a nova barca gasta, em média, 2 minutos a menos que as anteriores para completar a travessia.

Boa Prova!

Solução

Q1) (a) Como $U \sim U(0,1)$ segue que:

$$\begin{cases} P(X=1, Y=1) = P(U \leq 0,3) = 0,3, \\ P(X=0, Y=1) = P(0,3 < U \leq 0,6) = 0,3, \\ P(X=0, Y=0) = P(U > 0,6) = 0,4. \end{cases}$$

Além disso, como $Y \geq X$ temos que $P(X=1, Y=0) = 0$.

(b) X e Y não são independentes, pois, por exemplo:

$$P(X=1, Y=0) = 0 \neq (0,3)(0,4) = P(X=1)P(Y=0).$$

(c) Ora, já que $U \sim U(0,1)$, temos que

$$E(X) = P(U \leq 0,3) = 0,3, \quad E(Y) = P(U \leq 0,6) = 0,6$$

Além disso, $E(XY) = P(XY = 1) = P(U \leq 0,3) = 0,3$. Logo,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = (0,3) - (0,3)(0,6) = 0,12.$$

(d) Usando a definição de probabilidade condicional temos,

$$P(X=1|Y=1) = P(X=1, Y=1)/P(Y=1) = (0,3)/(0,6) = 0,5$$

e

$$P(Y=0|X=0) = P(X=0, Y=0)/P(X=0) = (0,4)/(0,7) = 0,5714.$$

Q2) (a) X $\sim N(\mu, \sigma)$. Sabemos que:
$\Phi(\frac{80-\mu}{\sigma}) = 1 - 0,1586 = 0,8414$. Assim, $\frac{80-\mu}{\sigma} \approx 1$
$\Phi(\frac{50-\mu}{\sigma}) = 0,0227$. Assim, $\Phi(\frac{\mu-50}{\sigma}) = 1 - 0,0227 = 0,9773$, então, $\frac{\mu-50}{\sigma} = 2,0009 \approx 2$
Temos 2 equações e 2 incógnitas. Resolvendo o sistema, $\mu = 70$ e $\sigma = 10$.

(b) $P(55 < X < 80) = \Phi(\frac{80-70}{10}) - \Phi(\frac{55-70}{10}) = \Phi(1) - \Phi(-1,5) = \Phi(1) + \Phi(1,5) - 1$
$P(55 < X < 80) = 0,8413 + 0,9332 - 1 = 0,7715 \approx 0,77$.

(c) Seja Y = Número de repetições com tempo no intervalo I, Y $\sim Bin(n = 40; p = 0,77)$.
Como $np(1-p) = 40 \times 0,77 \times 0,23 = 7,084 \geq 3$, podemos aproximar Y por
$Y_c \sim N(\mu = np = 30,8; \sigma^2 = np(1-p) = 7,084 = 2,66^2)$. Então,
$P(Y \geq 25) \approx P(Y_c > 24,5) = 1 - \Phi(\frac{24,5-30,8}{2,66}) = 1 - \Phi(-2,37) = \Phi(2,37) = 0,9911$.

Q3) (a) As estimativas de μ, σ e p a partir da amostra piloto são, respectivamente:
$\bar{x} = \frac{\sum x_i}{10}$ = 6,82 mg/l ; $s = \sqrt{\frac{\sum {x_i}^2 - \frac{(\sum x_i)^2}{10}}{9}}$ = 2,23 mg/l; $\hat{p} = \frac{8}{10} = 0,8$.

(b) Aproximando Binomial por Normal, e usando a estimativa de p fornecida pela piloto, o tamanho mínimo da nova amostra para garantir que
P[$|\hat{p} - p| <$ 0,08] = 0,95, é $n = (\frac{z_{0,975}}{0,08})^2 p(1-p) \approx (\frac{1,96}{0,08})^2 0,8 \times (1-0,8) \approx 97$.
Uma regra prática afirma que, para garantir a validade da aproximação da Binomial pela Normal, devemos ter np(1-p) $\geq$ 3. Ora, como $97 \times 0,8 \times (1-0,8)$ = 15,52 ≥ 3, essa condição está sendo atendida neste caso.

(c) Aplicando o TCL, e usando a estimativa de σ fornecida pela piloto, temos:
P[$|\bar{X} - \mu| < 0,3] \approx 2\Phi(\frac{0,3\sqrt{97}}{2,23}) - 1 = 2\Phi(1,32) - 1 = 0,82$.

Q4) (a) As hipóteses de interesse são:

$$\begin{cases} H_0: & \text{o tempo médio de travessia é igual ou superior a 15 minutos} \\ H_1: & \text{o tempo médio de travessia é inferior a 15 minutos} \end{cases}$$

As hipóteses envolvem o parâmetro μ e podem ser escritas como:

$$\begin{cases} H_0: & \mu \geq 15 \\ H_1: & \mu < 15 \end{cases}$$

Uma vez que o teste unilateral envolve a média populacional, consideramos a média amostral $\bar{X}$ para construir a estatística de teste e usamos o fato de que $\bar{X} \sim N(\mu, \sigma^2/n = 3^2/20)$. A estatística de teste sob H_0 é:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 15}{3/\sqrt{20}} \sim N(0,1)$$

Dada a amostra de tamanho $n = 20$, sob a hipótese nula, o valor observado de Z é:

$$z_{obs} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{14,5 - 15}{3/\sqrt{20}} = -0,745.$$

e podemos decidir pela a aceitação ou rejeição de H_0 sob o nível de significância de 5% já previamente estipulado. Note que a regra de decisão pode ser formulada em termos da estatística Z ou em termos da própria média amostral $\bar{X}$. Rejeitamos H_0 se

$$z_{obs} < -1,64,$$

ou equivalentemente, se

$$\bar{x} < 15 - 1,64\frac{3}{\sqrt{20}} = 13,89$$

Como $z_{obs} > -1,64$, estamos na região de aceitação da hipótese e a decisão é não rejeitar a hipótese de que o tempo médio de travessia das barcas seja superior ou igual a 15 km/l. Da mesma forma, como $\bar{x} = 14,5$ não pertence à região crítica, não rejeitamos a hipótese H_0 ao nível de significância de 5%.

(b) Como o teste é unilateral a esquerda, o p-valor será dado por:

$$\begin{aligned} p - valor = P\left(\bar{X} \leq \bar{x}_{obs} \mid H_0 \text{ verdadeira}\right) &= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{14,5 - 15}{3/\sqrt{20}}\right) \\ &= P(Z \leq -0,745) = 1 - P(Z \leq 0,745) \\ &= 0,226. \end{aligned}$$

Como $p - valor > \alpha$, concluímos pela não rejeiçao de H_0.

(c) Assumindo $\mu = 13$ a probabilidade do erro do tipo II é dada por

$$\begin{aligned} \beta(13) &= P(\text{não rejeitar } H_0 \mid H_0 \text{ falsa}) \\ &= P(\bar{X} \geq 13,89 \mid \mu = 13) \\ &= P\left(\frac{\bar{X} - 13}{3/\sqrt{20}} \geq \frac{13,89 - 13}{3/\sqrt{20}}\right) \\ &= P(Z \geq 1,32) = 1 - \Phi(1,32) \\ &= 1 - 0,906 = 0,094 \end{aligned}$$

Assim, se $\mu = 13$ minutos, a probabilidade de concluirmos de forma equivocada que H_0 não deveria ser rejeitada é 0,094.