Estatística - 15 de janeiro de 2015

Proposta de Resolução

Enunciado A

1. Representem X_A e X_B os consumos diários (em kWh) nos edifícios A e B, respetivamente. Do enunciado sabe-se que $X_A \sim \mathrm{N}\left(\mu_A = 620,\, \sigma_A = 15\right)$ e sabe-se que $X_B \sim \mathrm{N}\left(\mu_B = 580,\, \sigma_B = 10\right)$.

   a) Pretende-se calcular $P\left(X_B > X_A\right)$. Então

   $P\left(X_B > X_A\right) = P\left(X_B - X_A > 0\right)$

   Fazendo $X_B - X_A = X_{B-A}$, a nova variável X_{B-A} resulta de uma combinação linear de v.a. Normais, sendo também uma v.a. Normal com

   $\mu = \mu_A - \mu_B = 580 - 620 = -40$ e com

   $\sigma^2 = \sigma_B^2 + \sigma_A^2 = 10^2 + 15^2 = 325$

   Então,

   $$P\left(X_{B-A} > 0\right) = P\left(Z > \frac{0 - -40}{\sqrt{325}}\right) = P\left(Z > 2.2188\right) = 0.0133$$

   b) Pretende-se calcular $P\left(X_B > 600\right)$. Então,

   $$P\left(X_B > 600\right) = P\left(Z > \frac{600 - 580}{10}\right) = P\left(Z > 2\right) = 0.0228$$

   c) Represente-se por Y o número de dias, em 30, em que o consumo de 600 kWh é ultrapassado. Admitindo que os consumos em dias consecutivos são independentes, então $Y \sim \mathrm{B}\left(n = 30,\, p\right)$ com $p = P\left(X_B > 600\right) = 0.0228$, conforme calculado na alínea anterior. Pretende-se calcular $P\left(Y \geq 1\right)$. Utilizando a distribuição Binomial,

   $$P\left(Y \geq 1\right) = 1 - P\left(Y \leq 0\right) = 1 - F(0) = 1 - 0.5014 = 0.4986$$

2. Represente Y o número de acidentes num dia e represente X o intervalo de tempo (em dias) entre acidentes. Do enunciado sabe-se que $Y \sim \mathrm{Poisson}\left(\lambda = 0.01\right)$ e que $X \sim \mathrm{EN}\left(\lambda = 0.01\right)$.

   a) Represente Y_m o número de acidentes mensais. Então,

   $Y_m \sim \mathrm{Poisson}\left(\lambda = 0.01 \times 30 = 0.3\right)$

Pretende-se calcular $P(Y_m \geq 1)$. Utilizando a distribuição de Poisson,

$$P(Y_m \geq 1) = 1 - P(Y_m \leq 0) = 1 - F(0) = 1 - 0.7408 = 0.2592$$

b) Como o processo de Poisson não tem memória, o facto de não terem ocorrido acidentes até maio é irrelevante para o cálculo da probabilidade de acidente em junho. Logo, a resposta é $P(Y_m = 0) = 0.7408$. Alternativamente, utilizando a distribuição Exponencial Negativa, poder-se-ia calcular

$$\begin{aligned} P(X \geq 180 | X \geq 150) &= \frac{P(X \geq 180 \cap X \geq 150)}{P(X \geq 150)} = \frac{P(X \geq 180)}{P(X \geq 150)} = \\ &= \frac{1 - F(180)}{1 - F(150)} = \frac{1 - 0.8347}{1 - 0.7769} = 0.7408 \end{aligned}$$

c) Esta probabilidade pode ser calculada pela distribuição Exponencial Negativa:

$$P(X \geq 180) = 1 - F(180) = 1 - 0.8347 = 0.1653$$

d) O número médio de acidentes por ano (λ_a) é proporcional ao número médio de acidentes diários, ou seja,

$$\lambda_a = \lambda \cdot 30 \cdot 12 = 0.01 \times 360 = 3.6$$

O tempo médio entre acidentes é o inverso do número médio de acidentes, ou seja, $1/\lambda = 1/0.01 = 100$ dias.

3. Definam-se os seguintes acontecimentos:

GI: ocorrência de gémeos idênticos, com $P(GI) = 0.3$;

$GF = \overline{GI}$: ocorrência de gémeos fraternos, com $P(GF) = 1 - P(GI) = 0.7$;

MM: gémeos do sexo masculino, com $P(MM|GI) = 0.5$ e com $P(MM|GF) = 0.25$;

FF: gémeos do sexo feminino, com $P(FF|GI) = 0.5$ e com $P(FF|GF) = 0.25$;

MF: um gémeo masculino e outro feminino, com $P(MF|GF) = 0.5$.

Pretende-se calcular $P(GI|MM)$. Pela lei da probabilidade condicionada (ou desenvolvendo diretamente com o teorema de Bayes),

$$\begin{aligned} P(GI|MM) &= \frac{P(GI \cap MM)}{P(MM)} = \frac{P(GI \cap MM)}{P(GI \cap MM) + P(GF \cap MM)} \\ &= \frac{P(MM|GI) \cdot P(GI)}{P(MM|GI) \cdot P(GI) + P(MM|GF) \cdot P(GF)} \\ &= \frac{0.5 \times 0.3}{0.5 \times 0.3 + 0.25 \times 0.7} = 0.4615 \end{aligned}$$

4. Represente X o número de funcionários de uma loja. Sabe-se que $X = Y + 1$ e que $Y \sim \mathrm{B}(n = 5,\, p = 0.6)$.

   a) Como Y varia entre 0 e $n = 5$, então, $X = Y+1$ vai variar entre 1 e $n+1 = 6$. Logo, o número mínimo de funcionários por loja é 1 e o número máximo de funcionários por loja é 6.

   b) Pretende-se calcular $P(X = 4)$. Utilizando a distribuição Binomial,

   $$P(X = 4) = P(Y + 1 = 4) = P(Y = 3) = f(3) = 0.3456$$

   c) Represente-se por S o número total de funcionários a trabalhar no centro comercial. S resulta da soma dos funcionários das 240 lojas, representados por X_1, X_2, ..., X_{240}. Admitindo que o número de funcionários das várias lojas é independente entre si, temos uma soma de variáveis independentes e identicamente distribuídas. Logo, $S = X_1 + X_2 + \cdots + X_{240}$. Pelo teorema do limite central, $S \approx$ Normal com $\mu_S = 240 \cdot \mu_X$ e com $\sigma^2_S = 240 \cdot \sigma^2_X$. Os parâmetros de X podem ser obtidos com:

   $$\mu_X = E(Y + 1) = E(Y) + 1 = np + 1 = 5 \times 0.6 + 1 = 4$$

   $$\sigma^2_X = Var(Y + 1) = Var(Y) = np(1 - p) = 5 \times 0.6 \times (1 - 0.6) = 1.2$$

   Logo, $\mu_S = 240 \times 4 = 960$ e $\sigma^2_S = 240 \times 1.2 = 288$
   Pretende-se calcular[1] $P(S > 1000 + 0.5)$:

   $$P\left(S > 1000.5\right) = P\left(Z > \frac{1000.5 - 960}{\sqrt{288}}\right) = P\left(Z > 2.3865\right) = 0.0085$$

5. Utilizando a variável X para representar o diâmetro das peças produzidas, sabe-se que $X \approx$ Normal, com valor esperado, μ, desconhecido, e desvio padrão $\sigma = 0.05$ mm.

   a) Dado o enunciado, pretende-se testar, com $\alpha = 0.5\%$, as seguintes hipóteses:

   $$H_0 : \mu = 10\,\text{mm}$$
   $$H_1 : \mu \neq 10\,\text{mm}$$

   Como se trata de uma população Normal com variância conhecida, a estatística do teste ET é dada por

   $$ET = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

   Se H_0 for verdadeira, $ET \sim \mathrm{N}(0,\, 1)$. Nestas condições, H_0 deve ser rejeitada se $ET < -z(\alpha/2)$ ou se $ET > z(\alpha/2)$, sendo $\alpha/2 = 0.0025$ e $z(\alpha/2) = 2.807$.

[1] Utilizando a correção de continuidade

No caso da amostra fornecida, $\bar{x} = 9.9764$. Logo,

$$ET = \frac{9.9764 - 10}{0.05/\sqrt{5}} = -1.0554$$

De acordo com o critério definido, a hipótese nula não deve ser rejeitada, não havendo evidência de que a média dos diâmetros se tenha desviado significativamente de 10 mm. Procedendo ao cálculo do valor de prova ($v.p.$):

$$v.p. = 2 \cdot P(Z \geq |ET|) = 2 \cdot P(Z \geq |-1.0554|) = 2 \times 0.1456 = 0.2912$$

b) Há evidência de desvios significativos quando H_0 é rejeitada, ou seja, quando $ET < -z(\alpha/2)$ ou quando $ET > z(\alpha/2)$. Desenvolvendo a primeira condição,

$$ET < -z(\alpha/2) \Leftrightarrow \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < -z(\alpha/2) \Leftrightarrow \bar{x} < \mu - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \Leftrightarrow$$
$$\Leftrightarrow \bar{x} < 10 - 2.807\frac{0.05}{\sqrt{5}} \Leftrightarrow \bar{x} < 9.9372$$

ou seja, $\bar{x}_i = 9.9372$. Desenvolvendo a segunda expressão,

$$ET > z(\alpha/2) \Leftrightarrow \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} > z(\alpha/2) \Leftrightarrow \bar{x} > \mu + z(\alpha/2)\frac{\sigma}{\sqrt{n}} \Leftrightarrow$$
$$\Leftrightarrow \bar{x} > 10 + 2.807\frac{0.05}{\sqrt{5}} \Leftrightarrow \bar{x} > 10.0628$$

conclui-se que $\bar{x}_s = 10.0628$.

6. a) Trata-se de um intervalo de confiança para a proporção Binomial. Como se trata de uma amostra de grande dimensão, pode utilizar-se no cálculo a expressão

$$\hat{p} \pm z(\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Na amostra obteve-se $y = 150$ e $n = 400$. Logo, $\hat{p} = y/n = 150/400 = 0.375$. Sabe-se também que $1 - \alpha = 0.95$ e, logo, $z(\alpha/2) = z(0.025) = 1.96$. O intervalo de confiança será

$$0.375 \pm 1.96\sqrt{\frac{0.375(1-0.375)}{400}} \Leftrightarrow [0.3276,\ 0.4224]$$

O resultado exato, caso tivesse sido utilizada a distribuição Binomial diretamente, seria [0.3274, 0.4245].

b) A amplitude do intervalo calculado na alínea anterior é dada por

$$A = 2 \cdot z(\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Resolvendo em ordem a n, fica

$$n = \frac{\left[2 \cdot z(\alpha/2)\right]^2 \hat{p}(1-\hat{p})}{A^2}$$

Para garantir uma amplitude menor que o especificado ($A = 0.05$), em qualquer circunstâncias, deve utilizar-se $\hat{p} = 0.5$. Então,

$$n = \frac{\left[2 \times 1.96\right]^2 \times 0.5 \times (1-0.5)}{0.05^2} = 1536.5835$$

devendo o resultado ser arredondado por excesso, ou seja, $n = 1537$.

Caso se pudesse tomar a amostra da alínea anterior como *amostra piloto* para obtenção de uma estimativa de p, poder-se-ia fazer $\hat{p} = 0.375$. Neste caso o resultado seria $n = 1441$, ligeiramente inferior.