UFRJ - CCMN - IM - Departamento de Métodos Estatísticos
Probabilidade e Estatística - Estatística

Prova # 02 19-02-2013

Atenção: Não serão aceitas respostas sem justificativa: as expressões que levaram a alguma resposta numérica devem ser indicadas nos espaços apropriados.

---

1. No Departamento Comercial de certa companhia foi formada uma equipe de vendas com a admissão de 15 vendedores. Cada um deles foi submetido a um teste especialmente concebido para prever seu desempenho futuro. Além do teste, para a admissão, foi considerada a experiência do vendedor. Assim, foram registradas duas variáveis: T, representando o resultado do teste e, E, o número de anos de experiência. Um ano depois, o desempenho de cada vendedor foi medido observando-se a variável V, que representa o volume de vendas médio mensal. Os dados estão na tabela a seguir.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\sum x$	$\sum x^2$
V	57	54	49	46	46	45	45	44	44	42	38	35	29	29	18	621	27159
T	9	8	7	8	8	5	5	7	7	4	6	4	4	4	4	90	586
E	5	3	4	3	4	3	2	2	5	3	2	2	3	4	2	47	163

(a) Construa um ramo e folhas da variável V e calcule a mediana amostral de V.

(b) Calcule a média e a variância amostrais de V.

(c) Comparando as variáveis desempenho no teste (T) e anos de experiência (E), qual delas você diria que parece ser mais importante na previsão do futuro desempenho de vendedores? Por que? Informação útil: $\sum_{i=1}^{15} T_i V_i = 3930$ e $\sum_{i=1}^{15} E_i V_i = 2009$.

(a)

Ramo-e-folhas (ramos: dezenas e folhas: unidades)

1	8
2	9 9
3	5 8
4	2 4 4 5 5 6 6 9
5	4 7

A mediana amostral de V é o valor que ocupa a posição central, quando os valores estão ordenados. Como $n = 15$, a mediana amostral será o valor que ocupa a oitava posição, a saber, 44.

(b)

$\bar{V} = \frac{621}{15} = 41,4$

$s_V^2 = \frac{27159 - 621^2/15}{15-1} = 103,5429.$

(c)

A correlação amostral entre V e T é dada por

$$\frac{3930 - (621 \times 90)/15}{\sqrt{103,5429}\sqrt{586 - (90)^2/15}} = 0,79$$

e, a correlação amostral entre V e E é dada por

$$\frac{2009 - (621 \times 47)/15}{\sqrt{103,5429}\sqrt{163 - (47)^2/15}} = 0,42$$

Logo, a variável que parece ser mais importante na previsão do futuro desempenho de vendedores é T o resultado no teste, por apresentar maior correlação com a variável V em relação à variável E.

2. Suponha que a variável aleatória T: *tempo entre chegadas consecutivas de clientes na fila de um banco* siga distribuição Exponencial com parâmetro λ desconhecido, e tome $(T_1, \ldots, T_n)$ uma amostra aleatória de T. Denote por ψ a mediana de T (lembre que $\int_{-\infty}^{\psi} f(t)dt = \frac{1}{2}$, com $f(t)$ a função de densidade da v.a. T), e seja $T_{md} = (\ln 2)\,\overline{T}$ um estimador de ψ, em que $\overline{T}$ denota a média amostral de $(T_1, \ldots, T_n)$.

   (a) T_{md} é um estimador não-tendencioso para ψ? Por que? **(b)** Obtenha o erro médio quadrático de T_{md} como estimador de ψ. **(c)** O estimador de máxima verossimilhança de ψ é T_{md}? Por que?

(a)

$$\int_{-\infty}^{\psi} f(t)dt = 1/2 \;\Rightarrow\; \int_{0}^{\psi} \lambda e^{-\lambda t}dt = 1/2 \;\Rightarrow\; 1 - e^{-\lambda\psi} = 1/2 \;\Rightarrow\; e^{-\lambda\psi} = 1/2$$

$$\Rightarrow\; -\lambda\psi = \ln(1/2) \;\Rightarrow\; \psi = -\frac{\ln(1/2)}{\lambda} = -\frac{\ln 1 - \ln 2}{\lambda} \;\Rightarrow\; \psi = \frac{\ln 2}{\lambda}.$$

$$E(T_{md}) \;=\; E((\ln 2)\,\overline{T}) \;=\; (\ln 2) \times E\left(\frac{1}{n}\sum_{i=1}^{n} T_i\right) \;=\; (\ln 2) \times \frac{1}{n} \times n \times E(T) \;=\; (\ln 2) \times \frac{1}{\lambda}$$

$$\Rightarrow\;\; E(T_{md}) = \psi.$$

Portanto, T_{md} é um estimador não-tendencioso para ψ.

(b)

$$EQM(T_{md}) = Var(T_{md}) = Var((\ln 2)\,\overline{T}) = (\ln 2)^2 \times Var\left(\frac{1}{n}\sum_{i=1}^{n} T_i\right)$$

$$= (\ln 2)^2 \times \frac{1}{n^2} \times n \times Var(T) = (\ln 2)^2 \times \frac{1}{n^2} \times n \times \frac{1}{\lambda^2} = \frac{1}{n}\left(\frac{\ln 2}{\lambda}\right)^2 = \frac{\psi^2}{n}.$$

(c) Como $EMV(\lambda) = 1/\overline{T}$ e pela propriedade de invariância dos EMV's, temos que

$$EMV(\psi) = EMV\left(\frac{\ln 2}{\lambda}\right) = \frac{\ln 2}{EMV(\lambda)} = (\ln 2)\overline{T} = T_{md}.$$

Portanto, T_{md} é o estimador de máxima verossimilhança de ψ.

3. Um estudo com 420.095 dinamarqueses usuários de telefones celulares descobriu que 135 deles tinham desenvolvido câncer no cérebro. Uma pesquisa anterior a essa, diz que a taxa de incidência de tipo de câncer era de 0,034% para pessoas que **não** usavam telefones celulares.

   (a) Construa um intervalo de 95% de confiança, não-conservativo, para a proporção de usuários de telefone celular que desenvolvem câncer no cérebro.

   (b) Você diria que os usuários de telefone celular parecem ter uma taxa de incidência de câncer no cérebro diferente da taxa de incidência deste tipo de câncer em pessoas que não usam telefone celular? Realize um teste de hipóteses, adotando um nível de significância de 1% para dar a sua resposta.

   (c) Calcule o p-valor do teste realizado no item anterior.

(a) Seja $\hat{p} = \frac{135}{420.095} = 0,000321 = 0,0321\%$. Temos que,

$$
\begin{aligned}
IC(p, 95\%) &= \hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
&= 0,000321 \pm 1,96 * \sqrt{\frac{0,000321(1-0,000321)}{420.095}} \\
&= (0,000267; 0,000375)
\end{aligned}
$$

Portanto, o intevalo de confiança é (0,0267% ; 0,0375%).

(b) Seja o teste:

$$
\begin{aligned}
H_0 &: p = 0,034\% \\
H_1 &: p \neq 0,034\%
\end{aligned}
$$

Seja a região de rejeição: $\hat{p} > p_0 + z_{1-\frac{\alpha}{2}}\ \sigma_0$ ou $\hat{p} < p_0 - z_{1-\frac{\alpha}{2}}\ \sigma_0$. Logo, temos que:

$$
\begin{aligned}
p_0 + z_{1-\frac{\alpha}{2}}\ \sigma_0 &= 0,034\% + 2,57\sqrt{\frac{0,034\%(1-0,034\%)}{420.095}} = 0,000413 = 0,0413\%, \quad e \\
p_0 - z_{1-\frac{\alpha}{2}}\ \sigma_0 &= 0,034\% - 2,57\sqrt{\frac{0,034\%(1-0,034\%)}{420.095}} = 0,000267 = 0,0267\%
\end{aligned}
$$

Assim, a região é dada por: $\hat{p} < 0,0267\%$ ou $\hat{p} > 0,0413\%$. Portanto, com um nível de 1%, não existem evidências de diferença entre as taxas de incidência de câncer no cérebro de pessoas que usam telefones celulares e as que não usam.

(c) Seja o p-valor,

$$
\begin{aligned}
\tilde{\alpha} &= P(|Z| \geq |z_0|) = P\left(|Z| \geq |\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}|\right) \\
&= P(|Z| \geq 0,67) = 1 - P(-0,67 < Z < 0,67) \\
&= 1 - [\Phi(0,67) - \Phi(-0,67)] = 2 - 2\Phi(0,67) = 0,5028.
\end{aligned}
$$

4. Seja μ a média populacional da variável desempenho (em quilômetros por litro de combustível) dos carros de um determinado modelo e suponha que essa variável seja Normalmente distribuída. Foi realizado um teste $H_0 : \mu \leq 10$ contra $H_1 : \mu > 10$ com base numa amostra aleatória de tamanho $n = 9$ automóveis desse modelo. O p-valor calculado com base nos dados resultou ser igual a 0,10, o que sugere que H_0 seja aceita. Sabe-se também que o coeficiente de variação amostral da variável desempenho foi calculado em 0,3.

   (a) Determine os valores da média e do desvio-padrão amostrais da variável desempenho. **(b)** Deseja-se agora, usando esses mesmos dados, testar $H_0 : \mu \leq 8$ contra $H_1 : \mu > 8$. Com base nas tabelas de probabilidade fornecidas, indique o p-valor, identificando os extremos do intervalo de menor amplitude que o contém em seu interior. **(c)** Você diria que os dados apontam para a rejeição de H_0 ou para a não rejeição de H_0 do teste realizado no item (b)? Por que?

(a) No primeiro teste temos $\begin{cases} H_0 : \mu \leq 10 \\ H_1 : \mu > 10 \end{cases}$.

Já que $\tilde{\alpha} = 0,10$, a estatística de teste é $T = \dfrac{\bar{X} - 10}{S/\sqrt{9}}$.

Sob H_0, T segue uma distribuição t-de-Student com 9-1=8 graus de liberdade, tal que

$$\frac{\bar{x} - 10}{s/\sqrt{9}} = t_{1-\tilde{\alpha};8} = t_{0,9;8} = 1,397 \qquad (*)$$

Por outro lado, $CV = \dfrac{s}{\bar{x}} = 0,3$ implica que $s = 0,3\bar{x} \qquad (**)$.

Substituindo (**) em (*), obtemos

$$\frac{\bar{x} - 10}{0,3\bar{x}/3} = 1,397.$$

Resolvendo essa equação, chegamos a $\bar{x} = 11,62$ e, consequentemente, $s = 0,3 \times 11,62 = 3,49$.

(b) No segundo teste, temos $\begin{cases} H_0' : \mu \leq 8 \\ H_1' : \mu > 8 \end{cases}$.

Assim sendo, a estatística de teste passa a ser $T' = \dfrac{\bar{x} - 8}{s/\sqrt{9}}$.

Sob H_0', T' segue uma distribuição t-de-Student com 9-1=8 graus de liberdade.

Como os dados são os mesmos, $\bar{x} = 11,62$ e $s = 3,49$ tal que $T' = \frac{11,62-8}{3,49/3} = 3,12$.

Consultando a tabela da t-de-Student, vemos que na linha correspondente a 8 g.l., este valor está compreendido entre 2,896 (correspondente a $1 - \alpha = 0,99$) e 3,355 (correspondente a $1 - \alpha = 0,995$). Daí, decorre que $0,005 < \tilde{\alpha} < 0,01$.

(c) Como o p-valor é bem pequeno (menor que 0,01), os dados apontam para a rejeição de H_0.