

Probabilidade e Estatística

2009/1

Prof. Fernando Deeke Sasse

CCT-UDESC

Exercícios 2

Distribuição exponencial

1 A distância entre os buracos (significativos) em uma rodovia segue uma distribuição exponencial com uma média de 6 Km.

(a) Qual é a probabilidade de que não haja buracos em um trecho de 10 Km da rodovia ?

(b) Qual é a probabilidade de que haja 3 buracos em 10 Km de rodovia ?

(c) Qual é o desvio padrão da distância entre buracos ?

(d) Qual é a probabilidade de que o primeiro buraco ocorra entre 12 e 15 Km do início da inspeção ?

(e) Qual é a probabilidade de que não haja buracos em dois trechos separados de 5Km, nesta rodovia ?

Solução.

(a)

> **f := (lambda, x) -> lambda*exp(-lambda*x);**

$$f := (\lambda, x) \rightarrow \lambda e^{-\lambda x} \quad (1.1)$$

> **lambda := 1/6;**

$$\lambda := \frac{1}{6} \quad (1.2)$$

X: distância até o próximo buraco.

> **P(10 < X) := evalf(∫₁₀[∞] f(λ, x) dx)**

$$P(10 < X) := 0.1888756028 \quad (1.3)$$

(b) Seja Y o número de buracos em 10Km de rodovia. Como a distância entre os buracos é exponencial, Y é uma variável aleatória de Poisson com $\lambda = 10(1/6) = 5/3$.

> **g := (y, lambda) -> exp(-lambda)*lambda^y/factorial(y) ;**

$$g := (y, \lambda) \rightarrow \frac{e^{-\lambda} \lambda^y}{y!} \quad (1.4)$$

> **g(3., 5.*(1/3));**

$$0.1457373479 \quad (1.5)$$

(c) $\sigma_X = \frac{1}{\lambda} = 6$

(d)

> **evalf(∫₁₂¹⁵ f(λ, x) dx)**

(e)

$$0.05325028458 \quad (1.6)$$

$$> P(5 < X) := \text{evalf}\left(\int_5^{\infty} f(\lambda, x) dx\right)$$

Como os eventos são independentes, a probabilidade desejada é:

$$P(5 < X) := 0.4345982085 \quad (1.7)$$

$$> P(X > 5)^2;$$

$$0.1888756028 \quad (1.8)$$

>

2. Um fabricante de peças recebe ordens pela internet através de dois sistemas diferentes de roteamento. O tempo entre as ordens para cada sistema é tipicamente um distribuição exponencial com média de 3.2 minutos. Ambos sistemas operam independentemente.

(a) Qual é a probabilidade de que nenhuma ordem seja recebida em um período de 6 min ? Em 12 min ?

(b) Qual é a probabilidade de que ambos os sistemas recebam duas ordens entre 10 e 15 min. após o site começar a funcionar ?

(c) Por quê a distribuição de probabilidade conjunta não é necessária para responder os itens anteriores ?

Solução.

(a)

$$> f := (\lambda, x) \rightarrow \lambda e^{-\lambda x}$$

$$f := (\lambda, x) \rightarrow \lambda e^{-\lambda x} \quad (2.1)$$

$$> \lambda := \frac{1}{3.2}$$

$$\lambda := 0.3125000000 \quad (2.2)$$

$$> P(6 < X, 6 < Y) := \int_6^{\infty} \int_6^{\infty} f(\lambda, x) f(\lambda, y) dx dy$$

$$P(6 < X, 6 < Y) := 0.02351774586 \quad (2.3)$$

$$> P(12 < X, 12 < Y) := \int_{12}^{\infty} \int_{12}^{\infty} f(\lambda, x) f(\lambda, y) dx dy$$

> (b) Seja Y o número de ordens em um intervalo de 5 min. Então Y é uma variável aleatória de Poisson com

$$P(12 < X, 12 < Y) := 0.0005530843701 \quad (2.4)$$

$$> \lambda := np$$

$$\lambda := np \quad (2.5)$$

$$> \lambda := \frac{5}{3.2}$$

$$\lambda := 1.5625000000 \quad (2.6)$$

$$> g := (y, \lambda) \rightarrow \frac{e^{-\lambda} \lambda^y}{y!}$$

$$g := (y, \lambda) \rightarrow \frac{e^{-\lambda} \lambda^y}{y!} \quad (2.7)$$

$$> g(2, \lambda)^2$$

>

(c) As variáveis X e Y são independentes.

Distribuição de Erlang

1. O tempo entre falhas em uma máquina laser é exponencialmente distribuído com uma média de 2500h.

(a) Qual é o tempo esperado até a segunda falha?

(b) Qual é a probabilidade de que o tempo até a terceira falha exceda 50,000h ?

Solução:

Lembramos que uma variável aleatória exponencial descreve o intervalo até que a primeira contagem é obtida em um processo de Poisson. Uma generalização da distribuição exponencial é a distribuição de Erlang, que descreve o intervalo X até que r contagens ocorram em um processo de Poisson. Se $\lambda > 0$ é o valor esperado da variável aleatória de Poisson, então a função densidade de

probabilidade de X , com parâmetros λ e $r=1, 1, \dots$ é dada por

> **restart;**

> **f:=(x,r,lambd)->lambd^r*x^(r-1)*exp(-lambd*x)/(r-1)!;**

$$f := (x, r, \lambda) \rightarrow \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} \quad (3.1)$$

>

O valor esperado e a variância são dados por

$$\mu = \frac{r}{\lambda} \quad \text{e} \quad \sigma^2 = \frac{r}{\lambda^2}.$$

(a)

> **$\lambda := \frac{1}{25000}; r := 2;$**

$$\lambda := \frac{1}{25000}$$

$$r := 2$$

(3.2)

> **$te := \frac{r}{\lambda}$**

$$te := 50000$$

(3.3)

(b)

> **$f(x, 3, \lambda)$**

$$\frac{1}{31250000000000} x^2 e^{-\frac{1}{25000}x} \quad (3.4)$$

> **$int(f(x, 3, \lambda), x=50000.. \infty)$**

$$5 e^{-2}$$

(3.5)

> **$evalf(\%)$**

$$0.6766764160 \quad (3.6)$$

$$> 1 - \text{int}(f(x, 3, \lambda), x=0..50000) \quad 5 e^{-2} \quad (3.7)$$

2. O tempo entre falhas em uma máquina laser é exponencialmente distribuído com uma média de 2500h.

(a) Qual é o tempo esperado até a segunda falha?

(b) Qual é a probabilidade de que o tempo até a terceira falha exceda 50,000h ?

Solução:

Lembramos que uma variável aleatória exponencial descreve o intervalo até que a primeira contagem é obtida em um processo de Poisson. Uma generalização da distribuição exponencial é a distribuição de Erlang, que descreve o intervalo X até que r contagens ocorram em um processo de Poisson. Se $\lambda > 0$ é o valor esperado da variável aleatória de Poisson, então a função densidade de

probabilidade de X , com parâmetros λ e $r=1, 1, \dots$ é dada por

> **restart;**

> **f := (x, r, lambda) -> lambda^r * x^(r-1) * exp(-lambda*x) / (r-1)!;**

$$f := (x, r, \lambda) \rightarrow \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} \quad (4.1)$$

>

O valor esperado e a variância são dados por

$$\mu = \frac{r}{\lambda} \quad \text{e} \quad \sigma^2 = \frac{r}{\lambda^2} .$$

(a)

> $\lambda := \frac{1}{25000}; r := 2;$

$$\lambda := \frac{1}{25000}$$

$$r := 2$$

(4.2)

> $te := \frac{r}{\lambda}$

$$te := 50000$$

(4.3)

(b)

> $f(x, 3, \lambda)$

$$\frac{1}{3125000000000} x^2 e^{-\frac{1}{25000}x} \quad (4.4)$$

> $\text{int}(f(x, 3, \lambda), x=50000.. \infty)$

$$5 e^{-2}$$

(4.5)

> $\text{evalf}(\%)$

$$0.6766764160$$

(4.6)

$$\begin{aligned}
 &> 1 - \text{int}(f(x, 3, \lambda), x=0..50000) \\
 & \qquad \qquad \qquad 5 e^{-2} \qquad \qquad \qquad (4.7)
 \end{aligned}$$

>

Distribuição Normal

- 1 Uma peça deve ser produzida com um comprimento de 100.2 mm e desvio padrão de 0.1 mm.
- (a) Qual é a probabilidade de que um comprimento seja maior que 100.3 mm ou menor que 99.7 mm ?
- (b) Qual deve ser a média de modo que o maior número de comprimentos fique entre 99.7 mm e 100.3 mm ? Determine este número.

Solução:

(a)

$$> f := (\mu, \sigma, x) \rightarrow \exp(-1/2 * (x-\mu)^2 / \sigma^2) / (\text{sqrt}(2 * \text{Pi}) * \sigma);$$

$$f := (\mu, \sigma, x) \rightarrow \frac{e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}}{\sqrt{2 \pi} \sigma} \qquad (5.1)$$

$$> PR := f(100.2, 0.1, x);$$

$$PR := \frac{5.000000000 e^{-50.00000000(x-100.2)^2} \sqrt{2}}{\sqrt{\pi}} \qquad (5.2)$$

$$> P(x > 100.3) := \text{int}(PR, x=100.3..infinity); \qquad P(100.3 < x) := 0.1586552539 \qquad (5.3)$$

$$> P(x < 99.7) := \text{int}(PR, x = -infinity .. 99.7); \qquad P(x < 99.7) := 2.866515719 10^{-7} \qquad (5.4)$$

Ou seja, esta última probabilidade é praticamente zero. Portanto, a probabilidade desejada é aproximadamente 0.16.

(b) A média deve estar exatamente no meio destes comprimentos.

$$> \mu := (99.7 + 100.3) / (2); \qquad \mu := 100.0000000 \qquad (5.5)$$

Como a largura da normal é exatamente $3\sigma = 3(0.1) = 0.3$, devemos ter 99.7 % com comprimentos entre 99.7 mm e 100.3 mm.

> ;

2 A resistência compressora de amostras de cimento pode ser modelada por uma distribuição normal com uma média de 6500 Kg/cm^2 e desvio padrão de 80 Kg/cm^2 .

(a) Qual é a probabilidade de que a resistência de uma amostra seja menor do que 6450 Kg/cm^2 ?

(b) Qual é a probabilidade de que a resistência de uma amostra esteja entre 6400 e 6600 Kg/cm^2 ?

(c) Qual é a resistência excedida por 95 % ?

Solução.

(a)

> **restart;**

$$f := (\mu, \sigma, x) \rightarrow \frac{e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}}{\sqrt{2\pi}\sigma}$$

$$f := (\mu, \sigma, x) \rightarrow \frac{e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}}{\sqrt{2\pi}\sigma} \quad (6.1)$$

> **mu := 6500; sigma := 80;**

$$\mu := 6500$$

$$\sigma := 80$$

(6.2)

$$P(X < 6450) := \text{evalf}\left(\int_{-\infty}^{6450} f(\mu, \sigma, x) dx\right)$$
$$P(X < 6450) := 0.2659855291$$

(6.3)

Ou seja, praticamente zero.

(b)

$$\text{evalf}(\text{int}(f(\mu, \sigma, x), x = 6400..6600));$$
$$0.7887004525$$

(6.4)

(c)

> **with(stats);**

[*anova, describe, fit, importdata, random, statevalf, statplots, transform*]

(6.5)

$$\text{statevalf}[\text{icdf}, \text{normald}[\mu, \sigma]](0.5e-1);$$
$$6368.411710$$

(6.6)

Ou seja, 6368.41 Kg/cm²

Podemos conferir este resultado:

$$\int_{6368.411}^{10000} f(\mu, \sigma, x) dx$$

$$0.9500009151$$

(6.7)

> ;

Probabilidade Conjunta e Correlação

1 Determine a covariância e a correlação da seguinte distribuição de probabilidade conjunta:

x	-1	-0.5	0.5	1
y	-2	-1	1	2
f_{XY}	1/8	1/2	1/4	1/8

Façamos inicialmente o cálculo manual. A covariância é definida por

$$\sigma_{XY} = E(XY) - E(X)E(Y)$$

> *restart*:

$$\begin{aligned} > \mu_X := & -1 \cdot \left(\frac{1}{8}\right) - 0.5 \cdot \left(\frac{1}{2}\right) + 0.5 \cdot \left(\frac{1}{4}\right) + 1 \cdot \left(\frac{1}{8}\right) \\ & \mu_X := -0.1250000000 \end{aligned} \quad (7.1)$$

$$\begin{aligned} > \mu_Y := & -2 \cdot \left(\frac{1}{8}\right) - 1 \cdot \left(\frac{1}{2}\right) + 1 \cdot \left(\frac{1}{4}\right) + 2 \cdot \left(\frac{1}{8}\right) \\ & \mu_Y := -\frac{1}{4} \end{aligned} \quad (7.2)$$

$$\begin{aligned} > \mu_{XY} := & (-1) \cdot (-2) \cdot \left(\frac{1}{8}\right) + (-0.5) \cdot (-1) \cdot \left(\frac{1}{2}\right) + (0.5 \cdot 1) \cdot \left(\frac{1}{4}\right) + (1 \cdot 2) \cdot \left(\frac{1}{8}\right) \\ & \mu_{XY} := 0.8750000000 \end{aligned} \quad (7.3)$$

$$\begin{aligned} > \sigma_{XY} := & \mu_{XY} - \mu_X \cdot \mu_Y \\ & \sigma_{XY} := 0.8437500000 \end{aligned} \quad (7.4)$$

A correlação é definida por

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \text{ onde } \sigma_X^2 = \sum_x (x - \mu_X)^2 f_{XY}(x, y)$$

Calculemos estas quantidades:

$$\begin{aligned} > & (-1 - \mu_X)^2 \cdot \left(\frac{1}{8}\right) + \left(-\frac{1}{2} - \mu_X\right)^2 \cdot \left(\frac{1}{2}\right) + \left(\frac{1}{2} - \mu_X\right)^2 \cdot \left(\frac{1}{4}\right) + (1 - \mu_X)^2 \\ & \cdot \left(\frac{1}{8}\right) \\ & 0.4218750000 \end{aligned} \quad (7.5)$$

$$\begin{aligned} > \sigma_X := & \text{sqrt}(\%) \\ & \sigma_X := 0.6495190528 \end{aligned} \quad (7.6)$$

$$\begin{aligned} > & (-2 - \mu_Y)^2 \cdot \left(\frac{1}{8}\right) + (-1 - \mu_Y)^2 \cdot \left(\frac{1}{2}\right) + (1 - \mu_Y)^2 \cdot \left(\frac{1}{4}\right) + (2 - \mu_Y)^2 \cdot \left(\frac{1}{8}\right) \\ & 1.6875000000 \end{aligned} \quad (7.7)$$

$$\begin{aligned} > \sigma_Y := & \text{sqrt}(\%) \\ & \sigma_Y := 1.299038106 \end{aligned} \quad (7.8)$$

Note que neste caso podemos também usar a fórmula (as probabilidades podem ser fatoradas em cada termo)

$$\begin{aligned} > \sigma_X := & \sqrt{\mu_{X^2} - \mu_X^2}; \sigma_Y := \sqrt{\mu_{Y^2} - \mu_Y^2} \\ & \sigma_X := \sqrt{\mu_{X^2} - 0.01562500000} \\ & \sigma_Y := \frac{1}{4} \sqrt{16 \mu_{Y^2} - 1} \end{aligned} \quad (7.9)$$

$$\begin{aligned}
 > \mu_{Y^2} := (-2.)^2 \cdot \left(\frac{1}{8}\right) + (-1)^2 \cdot \left(\frac{1}{2}\right) + (1)^2 \cdot \left(\frac{1}{4}\right) + 2^2 \cdot \left(\frac{1}{8}\right) \\
 & \qquad \qquad \qquad \mu_{Y^2} := 1.750000000
 \end{aligned}
 \tag{7.10}$$

$$\begin{aligned}
 > \mu_{X^2} := (-1)^2 \cdot \left(\frac{1}{8}\right) + (-0.5)^2 \cdot \left(\frac{1}{2}\right) + (0.5)^2 \cdot \left(\frac{1}{4}\right) + 1^2 \cdot \left(\frac{1}{8}\right) \\
 & \qquad \qquad \qquad \mu_{X^2} := 0.4375000000
 \end{aligned}
 \tag{7.11}$$

$$\begin{aligned}
 > \sigma_X \cdot \sigma_Y \\
 & \qquad \qquad \qquad 0.6495190528 \\
 & \qquad \qquad \qquad 1.299038106
 \end{aligned}
 \tag{7.12}$$

$$\begin{aligned}
 > \rho_{XY} := \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \\
 & \qquad \qquad \qquad \rho_{XY} := 1.000000000
 \end{aligned}
 \tag{7.13}$$

Estes mesmos resultados podem também ser obtidos através dos comandos do pacote Statistics:

> *with(Statistics)* :
 > $X := \langle -1, -0.5, 0.5, 1 \rangle$

$$X := \begin{bmatrix} -1 \\ -0.5 \\ 0.5 \\ 1 \end{bmatrix}
 \tag{7.14}$$

> $Y := \langle -2, -1, 1, 2 \rangle$

$$Y := \begin{bmatrix} -2 \\ -1 \\ 1 \\ 2 \end{bmatrix}
 \tag{7.15}$$

> $W := \left\langle \frac{1}{8}, \frac{1}{2}, \frac{1}{4}, \frac{1}{8} \right\rangle$

$$W := \begin{bmatrix} \frac{1}{8} \\ \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{8} \end{bmatrix}
 \tag{7.16}$$

$$\begin{aligned}
 > \text{Covariance}(X, Y, \text{weights} = W) \\
 & \qquad \qquad \qquad 0.8437500000
 \end{aligned}
 \tag{7.17}$$

$$> \text{Correlation}(X, Y, \text{weights} = W)
 \tag{7.18}$$

> ;

2 Determine a covariância e a correlação da densidade de probabilidade conjunta

$f_{XY}(x, y) = e^{-x-y}$ no domínio $x > 0, y > 0$.

Solução.

Novamente, aqui $\sigma_{XY} := \mu_{XY} - \mu_X \cdot \mu_Y$. Como

$\mu_{XY} = \int_0^{\infty} \int_0^{\infty} xy e^{-x-y} dx dy = \int_0^{\infty} x e^{-x} dx \int_0^{\infty} y e^{-y} dy = \mu_X \cdot \mu_Y$, a covariância e a correlação devem ser zero.

4 Na transmissão de informação digital, a probabilidade de que um sinal que representa um bit tenha uma distorção alta, moderada ou baixa é 0.02, 0.1 e 0.88, respectivamente. Suponha que três sinais são transmitidos e que a quantidade de distorção de cada sinal é independente. Sejam X e Y o número de sinais com distorção alta e moderada, respectivamente.

(a) Determine: $f_{XY}(x, y)$, $f_X(x)$, $E(X)$, $f_Y(y)$, $E(Y|X=1)$.

(b) São X e Y independentes ?

Solução.

(a)

X tem distribuição binomial com $n = 3$ e $p_1 = 0.02$:

> *restart*

> **p1:= 0.2e-1: n := 3:**

> **fX := x-> binomial(n, x)*p1^x*(1-p1)^(n-x);**

$$fX := x \rightarrow \text{binomial}(n, x) p_1^x (1 - p_1)^{n-x} \quad (9.1)$$

> **Lx:= [seq(fX(x), x = 0 .. n)];**

$$Lx := [0.9411920, 0.057624, 0.001176, 0.0000080] \quad (9.2)$$

Y tem distribuição binomial com $n=3$ e $p_2 = 0.1$

> **p2 := 0.1: n := 3: p3:=0.88:**

> **fY := (y) -> binomial(n, y)*p2^y*(1-p2)^(n-y) ;**

$$fY := y \rightarrow \text{binomial}(n, y) p_2^y (1 - p_2)^{n-y} \quad (9.3)$$

> **Ly := [seq(fY(y), y = 0 .. n)];**

$$Ly := [0.7290, 0.243, 0.027, 0.0010] \quad (9.4)$$

> $Pr(X=1, Y=1) := \frac{3!}{1 \cdot 1 \cdot 1} p_1 \cdot p_2 \cdot p_3$

$$Pr(X=1, Y=1) := 0.01056 \quad (9.5)$$

> $Pr(X=2, Y=1) := \frac{3!}{2 \cdot 1} p_1^2 \cdot p_2$

$$Pr(X=2, Y=1) := 0.0001200000000 \quad (9.6)$$

> $Pr(X=3, Y=0) := \frac{3!}{3!} p_1^3$

$$Pr(X=3, Y=0) := 0.000008 \quad (9.7)$$

$$\begin{aligned} > Pr(X=1, Y=2) &:= \frac{3!}{1 \cdot 2} p1 \cdot p2^2 \\ Pr(X=1, Y=2) &:= 0.0006000000000 \end{aligned} \quad (9.8)$$

$$\begin{aligned} > Pr(X=0, Y=1) &:= \frac{3!}{1 \cdot 2} p2 \cdot p3^2 \\ Pr(X=0, Y=1) &:= 0.2323200000 \end{aligned} \quad (9.9)$$

Notemos que, como esperado,

$$\begin{aligned} > P(Y=1) &:= Pr(X=0, Y=1) + Pr(X=2, Y=1) + Pr(X=1, Y=1); \\ P(Y=1) &:= 0.2430000000 \end{aligned} \quad (9.10)$$

$$\begin{aligned} > Pr(X=1, Y=0) &:= \frac{3!}{1 \cdot 2} p1 \cdot p3^2 \\ Pr(X=1, Y=0) &:= 0.04646400000 \end{aligned} \quad (9.11)$$

$$\begin{aligned} > Pr(X=0, Y=0) &:= \frac{3!}{3!} p3^3 \\ Pr(X=0, Y=0) &:= 0.681472 \end{aligned} \quad (9.12)$$

$$\begin{aligned} > Pr(X=2, Y=0) &:= \frac{3!}{2 \cdot 1} p1^2 \cdot p3 \\ Pr(X=2, Y=0) &:= 0.001056000000 \end{aligned} \quad (9.13)$$

$$\begin{aligned} > Pr(X=0, Y=2) &:= \frac{3!}{2 \cdot 1} p2^2 p3 \\ Pr(X=0, Y=2) &:= 0.02640000000 \end{aligned} \quad (9.14)$$

$$\begin{aligned} > Pr(X=0, Y=3) &:= \frac{3!}{3!} p2^3 \\ Pr(X=0, Y=3) &:= 0.001 \end{aligned} \quad (9.15)$$

Portanto, formamos a tabela:

	Y	0	1	2	3
X					
3		0.000008			
2		0.001056	0.0001200		
1		0.046464	0.01056	0.000600	
0		0.681472	0.232320000\0	0.02640	0.001

$$\begin{aligned} > E(Y, X=1) &:= 1 \cdot Pr(X=1, Y=1) + 2 \cdot Pr(X=1, Y=2); \\ E(Y, X=1) &:= 0.01176000000 \end{aligned} \quad (9.16)$$

(c) Note que

$$\begin{aligned} > f_X(1) \cdot f_Y(1) \\ &= 0.014002632 \end{aligned} \quad (9.17)$$

que é diferente de $f_{XY}(1, 1) = 0.01056$, dado pela tabela. Portanto, X e Y não são independentes.

Estimadores

1 Dados relativos à espessura, em angstroms, de óxido em semicondutores são listados a seguir: 425, 431, 416, 419, 421, 436, 418, 410, 431, 433, 423, 426, 410, 435, 436, 428, 411, 426, 409, 437, 422, 428, 413, 416.

(a) Calcule um estimador pontual da espessura média de óxido para todos os membros da população.

(b) Calcule um estimador pontual para o desvio padrão da espessura de óxido para todos os membros da população.

(c) Calcule o erro padrão para o estimador pontual da parte (a).

(d) Calcule um estimador pontual da mediana.

(e) Calcule um estimador pontual da proporção de semicondutores na população que tem espessura de óxido maior do que 430.

Solução

(a)

> *restart*

> *with (Statistics) :*

> *L := rtable([425, 431, 416, 419, 421, 436, 418, 410, 431, 433, 423, 426, 410, 435, 436, 428, 411, 426, 409, 437, 422, 428, 413, 416], subtype = Array);*

$$L := \begin{bmatrix} 1..24 \text{ Array} \\ \text{Data Type: anything} \\ \text{Storage: rectangular} \\ \text{Order: Fortran_order} \end{bmatrix} \quad (10.1)$$

> *n := 24.*

$$n := 24. \quad (10.2)$$

> *Mean(L)*

$$423.3333333 \quad (10.3)$$

> ;

(b)

> *s := StandardDeviation(L)*

$$s := 9.082552154 \quad (10.4)$$

(c) O erro padrão é dado por

$$> \sigma_p := \frac{s}{\sqrt{n}}$$

$$\sigma_p := 1.853968195 \quad (10.5)$$

ou

> *StandardError(Mean, L);*

$$1.853968195 \quad (10.6)$$

(d)

> *Median(L)*

$$424. \quad (10.7)$$

```

> k := 0 :
> for i to 24 do
  if L[i] > 430 then k := k + 1 endif
od
> k

```

Ou seja, o estimador pontual da proporção é

7

(10.8)

```

> k
  n

```

0.2916666667

(10.9)

```

> ;

```

2 Duas máquinas produzem um mesmo tipo de peça com um mesmo parâmetro de especificação μ . No entanto, a máquina 1 é mais nova que a máquina 2 e, conseqüentemente, tem um menor variabilidade do parâmetro nas peças produzidas. Sabemos que a variância associada à máquina 1 é σ_1^2 e que a variância associada à máquina 2 é $\sigma_2^2 = a \sigma_1^2$. Suponha que temos observações independentes para o parâmetro na máquina 1 e na máquina 2.

(a) Mostre que

$$\alpha \bar{X}_1 + (1 - \alpha) \bar{X}_2$$

é um estimador não-tendencioso de μ para qualquer $0 < \alpha < 1$.

(b) Determine o erro-padrão do estimador de μ da parte (a).

(c) Qual o valor de α que minimiza o erro padrão do estimador pontual de μ ?

(d) Suponha que $a = 4$ e $n_1 = 2 n_2$. Que valor de α deve ser selecionado para minimizar o erro-padrão do estimador de μ da parte (a)? Quão ruim seria a escolha do parâmetro $\alpha = 0.5$?

Solução

(a)

$$E(\hat{\mu}) = E(\alpha \bar{X}_1 + (1 - \alpha) \bar{X}_2) = \alpha E(\bar{X}_1) + (1 - \alpha) E(\bar{X}_2) = \alpha \mu + (1 - \alpha) \mu = \mu$$

(b)

$$\begin{aligned}
 s.e.(\hat{\mu}) &= \sqrt{V(\alpha \bar{X}_1 + (1 - \alpha) \bar{X}_2)} = \sqrt{\alpha^2 V(\bar{X}_1) + (1 - \alpha)^2 V(\bar{X}_2)} \\
 &= \sqrt{\alpha^2 \frac{\sigma_1^2}{n_1} + (1 - \alpha)^2 \frac{\sigma_2^2}{n_2}} = \sqrt{\alpha^2 \frac{\sigma_1^2}{n_1} + (1 - \alpha)^2 a \frac{\sigma_1^2}{n_2}} \\
 &= \sigma_1 \sqrt{\frac{\alpha^2 n_2 + (1 - \alpha)^2 a n_1}{n_1 n_2}}
 \end{aligned}$$

(c)

> restart

$$\begin{aligned} > ep2 := \frac{\sigma_1^2 (\alpha^2 n2 + (1 - \alpha)^2 a n1)}{n1 n2} \\ ep2 := \frac{\sigma_1^2 (\alpha^2 n2 + (1 - \alpha)^2 a n1)}{n1 n2} \end{aligned} \quad (11.1)$$

$$\begin{aligned} > eq := \alpha^2 n2 + (1 - \alpha)^2 a n1 \\ eq := \alpha^2 n2 + (1 - \alpha)^2 a n1 \end{aligned} \quad (11.2)$$

$$\begin{aligned} > eq2 := diff(eq, \alpha) = 0 \\ eq2 := 2 \alpha n2 - 2 (1 - \alpha) a n1 = 0 \end{aligned} \quad (11.3)$$

$$\begin{aligned} > \alpha_s := solve(eq2, \alpha) \\ \alpha_s := \frac{a n1}{n2 + a n1} \end{aligned} \quad (11.4)$$

(d)

$$\begin{aligned} > a := 4; n1 := 2 \cdot n2 \\ a := 4 \\ n1 := 2 n2 \end{aligned} \quad (11.5)$$

$$\begin{aligned} > \alpha_s \\ \frac{8}{9} \end{aligned} \quad (11.6)$$

>

O erro padrão neste caso é

$$\begin{aligned} > assume(\sigma_1 > 0) \\ > ep1 := \sqrt{evalf(subs([a=4, n1=2 n2, \alpha=1. \alpha_s], ep2))} \\ ep1 := 0.6666666666 \sigma_1 \sqrt{\frac{1}{n2}} \end{aligned} \quad (11.7)$$

Se $\alpha_s = 0.5$ temos

$$\begin{aligned} > sqrt(evalf(subs([a=4, n1=2 * n2, \alpha=0.5], ep2))); \\ 1.060660172 \sigma_1 \sqrt{\frac{1}{n2}} \end{aligned} \quad (11.8)$$

>

Ou seja, bastante acima do erro mínimo.

3 Uma população normal tem média 100 e variância 25. Qual deve ser o tamanho da amostra aleatória para que o erro padrão da média amostral seja 1.5 ?

Solução

> **restart;**
> **assume(mu, real); assume(sigma, real); assume(sigma>0);**

$$> f := t \rightarrow \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2\sigma^2}(t-\mu)^2}}{\sigma}$$

$$f := t \rightarrow \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}} \quad (12.1)$$

> mu := 100; sigma := 5;

μ := 100

σ := 5

(12.2)

>

>

O erro padrão da média é dado por

$$> el := 1.5 = \frac{\sigma}{\text{sqrt}(n)}$$

$$el := 1.5 = \frac{5}{\sqrt{n}}$$

(12.3)

> solve(el, n)

11.11111111

(12.4)

Portanto, n = 12.

4 A capacidade compressiva do concreto é normalmente distribuída com $\mu = 2500$ psi e $\sigma = 40$ psi. Determine a probabilidade de que uma amostra aleatória de $n = 7$ elementos tenha capacidade compressiva no intervalo de 2490 psi a 2520 psi. Qual o erro padrão da média amostral ?

Solução.

> restart

> σ := 40; μ := 2500; n := 7

σ := 40

μ := 2500

n := 7

(13.1)

O erro padrão é:

$$> \text{sigmaM} := \text{evalf}\left(\frac{\sigma}{\sqrt{n}}\right)$$

sigmaM := 15.11857892

(13.2)

> X1 := 2520; X2 := 2496;

X1 := 2520

$$X2 := 2496 \quad (13.3)$$

$$\begin{aligned} > ZZ := \frac{XM - \mu}{\text{sigma}M} \\ ZZ := 0.06614378278 XM - 165.3594570 \end{aligned} \quad (13.4)$$

$$\begin{aligned} > z1 := \text{subs}(XM=X1, ZZ); z2 := \text{subs}(XM=X2, ZZ) \\ z1 := 1.3228756 \\ z2 := -0.2645752 \end{aligned} \quad (13.5)$$

$$\begin{aligned} > f := (t, m1, s1) \rightarrow \frac{1}{2} \frac{e^{-\frac{1}{2} \frac{(t-m1)^2}{s1^2}}}{\sqrt{\pi} s1} \\ f := (t, m1, s1) \rightarrow \frac{1}{2} \frac{\sqrt{2} e^{-\frac{1}{2} \frac{(t-m1)^2}{s1^2}}}{\sqrt{\pi} s1} \end{aligned} \quad (13.6)$$

$$\begin{aligned} > P1 := \text{evalf}(\text{int}(f(t, 0, 1), t=-\infty ..z1)) \\ P1 := 0.9070616246 \end{aligned} \quad (13.7)$$

$$\begin{aligned} > P2 := \text{evalf}(\text{int}(f(t, 0, 1), t=-\infty ..z2)) \\ P2 := 0.3956683635 \end{aligned} \quad (13.8)$$

$$\begin{aligned} > P1 - P2 \\ 0.5113932611 \end{aligned} \quad (13.9)$$

>

5 A vida efetiva de um componente utilizado em um motor é uma variável aleatória com uma média de 5000 h e desvio padrão de 40 h. A distribuição de vida efetiva é próxima à distribuição normal. O fabricante do motor introduz uma melhoria no processo de manufatura deste componente que aumenta a vida média para 5050 h e decresce o desvio padrão para 30 h. Suponha que uma amostra aleatória de $n1 = 16$ componentes é selecionada a partir do processo antigo e $n2 = 25$ do processo melhorado. Qual é a probabilidade de que a diferença nas duas médias amostrais seja de ao menos 25 horas? Suponha que os dois processos são independentes.

Solução

A distribuição da média amostral X1 é normal, com média

$$\begin{aligned} > X1M := 5000 \\ X1M := 5000 \end{aligned} \quad (14.1)$$

e desvio padrão

$$\begin{aligned} > \sigma_{1M} := \frac{40}{\text{sqrt}(16)} \\ \sigma_M := 10 \end{aligned} \quad (14.2)$$

A distribuição da média amostral X2 é normal, com média e desvio padrão amostral dados por

$$\begin{aligned} > X2M := 5050 \\ X2M := 5050 \end{aligned} \quad (14.3)$$

$$\begin{aligned} > \sigma_{2M} := \frac{30}{\text{sqrt}(25)} \\ & \qquad \qquad \qquad \sigma_{2M} := 6 \end{aligned} \tag{14.4}$$

A distribuição de $X_{2M} - X_{1M}$ é normal, com média $\mu_2 - \mu_1 = 5050 - 5000 = 50$ e variância

$$\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1} = 6^2 + 10^2 = 136$$

Queremos agora determinar

$$P(X_{2M} - X_{1M} \geq 25)$$

Variável normalizada

$$\begin{aligned} > Z := \text{evalf}\left(\frac{25 - 50}{\sqrt{136}}\right) \\ & \qquad \qquad \qquad Z := -2.143732314 \end{aligned} \tag{14.5}$$

>

$$\begin{aligned} > \text{with (Statistics) : } X := \text{RandomVariable(Normal(0, 1));} \\ & \qquad \qquad \qquad X := _R \end{aligned} \tag{14.6}$$

$$\begin{aligned} > 1 - \text{CDF}(X, Z) \\ & \qquad \qquad \qquad 0.9839728293 \end{aligned} \tag{14.7}$$

>

Usamos aqui o seguinte fato:

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if \bar{X}_1 and \bar{X}_2 are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \tag{7-9}$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

6 Seja a função densidade de probabilidade

$$f(x) = \frac{1}{\theta^2} x e^{-x/\theta}, \quad 0 \leq x < \infty, \quad 0 < \theta < \infty$$

Determine o estimador de máxima verossimilhança para θ .

Solução.

A função de máxima verossimilhança verossimilhança da amostra é

> restart

$$> f_i := \frac{x e^{-\frac{x}{\theta}}}{\theta^2}$$

$$f_i := \frac{x e^{-\frac{x}{\theta}}}{\theta^2} \quad (15.1)$$

> L1 := simplify($\prod_{i=1}^n f_i$)

$$L1 := \left(\frac{x}{\theta^2}\right)^n \prod_{i=1}^n e^{-\frac{x}{\theta}} \quad (15.2)$$

> L2 := subs($\prod_{i=1}^n e^{-\frac{x}{\theta}} = \exp\left(-\frac{n \cdot Xb}{\theta}\right)$, L1)

$$L2 := \left(\frac{x}{\theta^2}\right)^n e^{-\frac{nXb}{\theta}} \quad (15.3)$$

> L3 := log(L2)

$$L3 := \ln\left(\left(\frac{x}{\theta^2}\right)^n e^{-\frac{nXb}{\theta}}\right) \quad (15.4)$$

> eq2 := diff(L3, θ) = 0

$$eq2 := \frac{2 \left(\frac{x}{\theta^2}\right)^n n e^{-\frac{nXb}{\theta}}}{\theta} + \frac{\left(\frac{x}{\theta^2}\right)^n n Xb e^{-\frac{nXb}{\theta}}}{\theta^2} = 0 \quad (15.5)$$

> eq3 := simplify(eq2)

$$eq3 := \frac{n(-2\theta + Xb)}{\theta^2} = 0 \quad (15.6)$$

> solve(eq3, θ)

$$\frac{1}{2} Xb \quad (15.7)$$

>

Ou seja, o estimador de θ é

$$EST(\theta) = \frac{1}{2n} \sum_{i=1}^n x_i$$

7 Ilustre através de um exemplo o fato de que a média amostral é um estimador não tendencioso. Verifique a validade da fórmula de estimação da variância da média amostral.

Solução.

Tomemos a lista

```
> restart
> with (Statistics) :
> LI := [2, 3, 6, 8, 11]
                                LI := [2, 3, 6, 8, 11]                (16.1)
```

```
> n1 := nops(LI)
                                n1 := 5                            (16.2)
```

```
>
> k1 := rand(1..n1)
k1 := proc ( )
                                (16.3)
```

```
    proc ( ) option builtin = RandNumberInterface; end proc (6, 5, 3) + 1
end proc
```

```
> k1 ( )
                                5                                (16.4)
```

Geramos todas as 200 amostras de 2 elementos da população dada pela lista L1:

```
> NA := 200 :
> for j from 1 to NA do
    LLI[j] := [LI[k1( )], LI[k1( )]]
od:
> LL1b := {seq(LLI[h], h=1..NA)} :
> LL1c := [op(LL1b)]
LL1c := [[2, 2], [2, 3], [2, 6], [2, 8], [2, 11], [3, 2], [3, 3], [3, 6], [3, 8], [3, 11], [6, 2], [6, 3], [6, 6], [6, 8], [6, 11], [8, 2], [8, 3], [8, 6], [8, 8], [8, 11], [11, 2], [11, 3], [11, 6], [11, 8], [11, 11]]
                                (16.5)
```

```
> NI := nops(LL1c)
                                NI := 25                            (16.6)
```

Cada uma destas amostras tem média:

```
> for k from 1 to NI do
    XI[k] := Mean(LL1c[k])
od:
> LLL1 := [seq(XI[m], m=1..NI)];
LLL1 := [2., 2.500000000, 4., 5., 6.500000000, 2.500000000, 3., 4.500000000, 5.500000000, 7., 4., 4.500000000, 6., 7., 8.500000000, 5., 5.500000000, 7., 8., 9.500000000, 6.500000000, 7., 8.500000000, 9.500000000, 11.]
                                (16.7)
```

A média deste estimador é:

```
> MI := Mean(LLL1)
                                MI := 6.                            (16.8)
```


$LL[j] := [L[k()], L[k()], L[k()], L[k()], L[k()], L[k()], L[k()]]$
od:

Cada uma destas amostras tem média:

> **for** k **from** 1 **to** NA **do**

$X[k] := Mean(LL[k])$

od:

> $LLL := [seq(X[m], m=1..NA)];$

$LLL := [6.700000000, 9.314285714, 8.228571429, 6.771428571, 6.428571429,$ **(16.17)**

$6.457142857, 5.914285714, 8.485714286, 6.657142857, 15.65714286, 4.642857143,$

$7.671428571, 10.37142857, 15.61428571, 8.485714286, 7.785714286, 6.714285714,$

$7.142857143, 7.685714286, 9.028571429, 7.471428571, 7.128571429, 6.757142857,$

$7.400000000, 6.214285714, 5.357142857, 7.300000000, 8.257142857, 8.242857143,$

$7.857142857, 7.571428571, 6.985714286, 5.028571429, 14.98571429, 6.128571429,$

$26.34285714, 7.585714286, 16.90000000, 7.257142857, 6.271428571]$

A média deste estimador é:

> $M := Mean(LLL)$

$M := 8.570000000$ **(16.18)**

Notemos que

> $Mean(L)$

8.509615385 **(16.19)**

de modo que M é bem próximo do valor verdadeiro. Verificaremos agora a validade da fórmula para a variância da média.

A variância da população (variância verdadeira) é dada por

> $V := \frac{sum((L[m] - Mean(L))^2, m=1..n)}{n}$

$V := 86.02471525$ **(16.20)**

A fórmula para a variância da média, nos diz que esta pode ser estimada por

> $\sqrt{\frac{V}{7}}$

3.505601951 **(16.21)**

pois 7 é o número de elementos nas amostras. O desvio padrão verdadeiro da distribuição amostral das médias (erro padrão das médias) é

> $\sqrt{\frac{sum((LLL[m] - M)^2, m=1..NA)}{NA}}$

3.987913883 **(16.22)**

O que concorda aproximadamente com a estimação.

8 Sejam três amostras aleatórias com tamanhos $n_1 = 20$, $n_2 = 10$, $n_3 = 8$, tomadas de uma população com média μ e variância σ^2 . Sejam S_1^2, S_2^2, S_3^2 , as variâncias das amostras. Mostre que

$S^2 = \frac{(20 S_1^2 + 10 S_2^2 + 8 S_3^2)}{38}$ é um estimador não tendencioso de σ^2 .

Solução

Utilizaremos o fato de que a variância de qualquer amostra da população é um estimador não tendencioso da variância da população σ^2 , ou seja, $E(S_i^2) = \sigma^2$, $i = 1, 2, 3$. De fato, S^2 definido acima é um estimador não tendencioso de σ^2 , pois

$$\begin{aligned} E(S^2) &= \frac{E(20 S_1^2 + 10 S_2^2 + 8 S_3^2)}{38} = \frac{1}{38} (20 E(S_1^2) + 10 \cdot E(S_2^2) + 8 E(S_3^2)) \\ &= \frac{1}{38} (20 \sigma^2 + 10 \sigma^2 + 8 \sigma^2) = \sigma^2 \end{aligned}$$