CE-003: Estatística II - Turma K/O
Avaliações Semanais (1^o semestre 2015)

Semana 3 (av-01)

1. Considere um jogo com um baralho (52 cartas) no qual em uma primeira rodada retira-se duas cartas e em uma segunda rodada retira-se uma carta. O interesse é se as cartas são figuras (valete, dama ou rei) de qualquer naipe. Temos interesse em:
   - obter o espaço amostral;
   - obter a probabilidade de cada ponto amostral;
   - obter a distribuição de probabilidades do número de figuras obtidas nas três cartas.

   Deve-se considerar duas situações, com e sem reposição das cartas entre a primeira e a segunda rodada.

   Solução:
   Notação:

$$F: \text{a carta é uma figura}$$
$$N = \overline{F}: \text{a carta não é uma figura}$$

   - O espaço amostral para as duas situações (com e sem reposição) é o mesmo.

$$\Omega = \{(FF,F);(FF,N);(FN,F);(NF,F);(FN,N);(NF,N);(NN,F);(NN,N)\}$$

   - Já as probabilidades são afetadas por repor ou não as cartas

Ponto amostral	(FF,F)	(FF,N)	(FN,F)	(NF,F)	(FN,N)	(NF,N)	(NN,F)	(NN,N)
Com reposição	$\frac{12}{52}\frac{11}{51}\frac{12}{52}$	$\frac{12}{52}\frac{11}{51}\frac{40}{52}$	$\frac{12}{52}\frac{40}{51}\frac{12}{52}$	$\frac{40}{52}\frac{12}{51}\frac{12}{52}$	$\frac{12}{52}\frac{40}{51}\frac{40}{52}$	$\frac{40}{52}\frac{12}{51}\frac{40}{52}$	$\frac{40}{52}\frac{39}{51}\frac{12}{52}$	$\frac{40}{52}\frac{39}{51}\frac{40}{52}$
Sem reposição	$\frac{12}{52}\frac{11}{51}\frac{10}{50}$	$\frac{12}{52}\frac{11}{51}\frac{40}{50}$	$\frac{12}{52}\frac{40}{51}\frac{11}{50}$	$\frac{40}{52}\frac{12}{51}\frac{11}{50}$	$\frac{12}{52}\frac{40}{51}\frac{39}{50}$	$\frac{40}{52}\frac{12}{51}\frac{39}{50}$	$\frac{40}{52}\frac{39}{51}\frac{12}{50}$	$\frac{40}{52}\frac{39}{51}\frac{38}{50}$

   -

$$X: \text{número de figuras obtidas nas três cartas}$$
$$x \in \{0,1,2,3\}$$

Com reposição

x	0	1	2	3
P[X=x]	P[(NN,N)] $\frac{40}{52}\frac{39}{51}\frac{40}{52}$	P[(FN,N)] + P[(NF,N)] + P[(NN,F)] $\frac{12}{52}\frac{40}{51}\frac{40}{52} + \frac{40}{52}\frac{12}{51}\frac{40}{52} + \frac{40}{52}\frac{39}{51}\frac{12}{52}$	P[(FF,N)] + P[(FN,F)] + P[(NF,F)] $\frac{12}{52}\frac{11}{51}\frac{40}{52} + \frac{12}{52}\frac{40}{51}\frac{12}{52} + \frac{40}{52}\frac{12}{51}\frac{12}{52}$	P[(FFF)] $\frac{12}{52}\frac{11}{51}\frac{12}{52}$

Sem reposição

x	0	1	2	3
P[X=x]	P[(NN,N)] $\frac{40}{52}\frac{39}{51}\frac{38}{50}$	P[(FN,N)] + P[(NF,N)] + P[(NN,F)] $\frac{12}{52}\frac{40}{51}\frac{39}{50} + \frac{40}{52}\frac{12}{51}\frac{39}{50} + \frac{40}{52}\frac{39}{51}\frac{12}{50}$	P[(FF,N)] + P[(FN,F)] + P[(NF,F)] $\frac{12}{52}\frac{11}{51}\frac{40}{50} + \frac{12}{52}\frac{40}{51}\frac{11}{50} + \frac{40}{52}\frac{12}{51}\frac{11}{50}$	P[(FFF)] $\frac{12}{52}\frac{11}{51}\frac{10}{50}$
	0.4471	0.4235	0.1195	0.009955

OBS: no caso sem reposição a v.a. X segue uma distribuição hipergeométrica e as probabilidades podem ser obtidas pela função de probabilidade desta distribuição.

$$X \sim \text{HG}(N = 52, n = 3, k = 12)$$

$$P[X = x] \sim \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$P[X = 0] = \frac{\binom{12}{0}\binom{52-12}{3-0}}{\binom{52}{3}} = 0.4471$$

$$P[X = 1] = \frac{\binom{12}{1}\binom{52-12}{3-1}}{\binom{52}{3}} = 0.4235$$

$$P[X = 2] = \frac{\binom{12}{2}\binom{52-12}{3-2}}{\binom{52}{3}} = 0.1195$$

$$P[X = 3] = \frac{\binom{12}{3}\binom{52-12}{3-3}}{\binom{52}{3}} = 0.009955$$

Semana 4 (av-02)

1. Considere que indivíduos vão fazer um teste online no qual questões serão apresentadas sequencialmente ao candidato. Calcule a probabilidade pedidas nos contextos de cada um dos itens a seguir. Procure identificar: a variável aleatória em questão e sua distribuição de probabilidades.
   (a) Suponha que oito (8) questões são retiradas com reposição (ou seja uma mesma questão pode ser retirada mais de uma vez) de um *banco* de 40 questões dos quais o candidato sabe responder a 25 delas. Qual a probabilidade de acertar três ou mais questões?
   (b) Idem anterior porém supondo agora que as questões não podem se repetir.
   (c) Supondo novamente reposição das questões, o candidato responde até errar pela primeira vez. Qual a probabilidade de acertar pelo menos três questões?
   (d) Idem anterior supondo que responde até errar pela terceira vez.

   Solução:

   (a)

   X : número de questões certas entre oito questões selecionadas ao acaso (com repetição)
   $X \sim \text{B}(n = 8, p = 25/40)$
   $x \in \{0, 1, 2, \ldots 8\}$
   $P[X \geq 3] = 1 - P[X = 0] - P[X = 1] - P[X = 2] = 0.964$

   (b)

   X : número de questões certas entre oito questões selecionadas ao acaso (sem repetição)
   $X \sim \text{HG}(N = 40, K = 25, n = 8)$
   $x \in \{0, 1, 2, \ldots 8\}$
   $P[X \geq 3] = 1 - P[X = 0] - P[X = 1] - P[X = 2] = 0.9783$

   (c)

   X : número de acertos até o primeiro erro
   $X \sim \text{G}(p = 15/50)$
   $x \in \{0, 1, 2, \ldots\}$
   $P[X \geq 3] = 1 - P[X = 0] - P[X = 1] - P[X = 2] = 0.2441$

   (d)

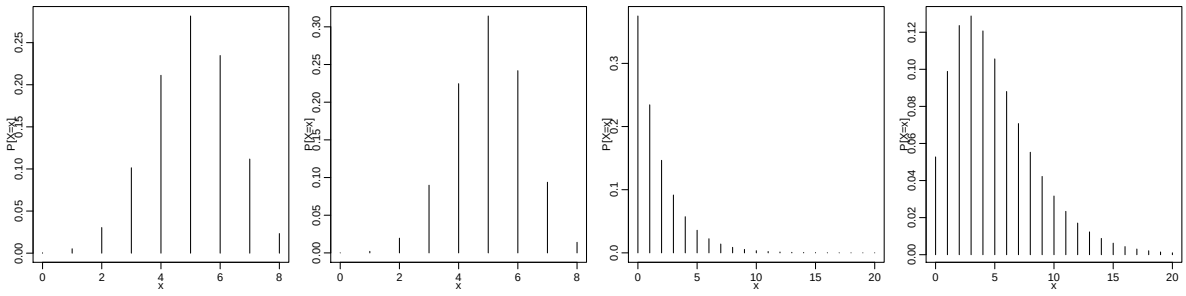
   X : número de acertos até o terceiro erro
   $X \sim \text{BN}(k = 3, p = 15/40)$
   $x \in \{0, 1, 2, \ldots\}$
   $P[X \geq 3] = 1 - P[X = 0] - P[X = 1] - P[X = 2] = 0.7248$

Soluções computacionais (linguagem R):

```
> (q1 <- pbinom(2, size=8, prob=25/40, lower=FALSE))
[1] 0.964
> (q2 <- phyper(2, m=25, n=15, k=8, lower=FALSE))
[1] 0.9783
> (q3 <- pgeom(2, prob=15/40, lower=FALSE))
[1] 0.2441
> (q4 <- pnbinom(2, size=3, prob=15/40, lower=FALSE))
[1] 0.7248
```

Gráficos das distribuições de probabilidades.

```
> par(mfrow=c(1,4))
> par(mar=c(3,3,0.2, 0.2), mgp=c(1.2, 0.6, 0))
> plot(0:8, dbinom(0:8, size=8, prob=25/40), xlab="x", ylab="P[X=x]", type="h")
> plot(0:8, dhyper(0:8, m=25, n=15, k=8), xlab="x", ylab="P[X=x]", type="h")
> plot(0:20, dgeom(0:20, prob=15/40), xlab="x", ylab="P[X=x]", type="h")
> plot(0:20, dnbinom(0:20, size=3, prob=15/40), xlab="x", ylab="P[X=x]", type="h")
```



2. Um vendedor consegue vender, em média, 0,5 unidades de um produto por dia. Calcule as probabilidades de:
   (a) vender alguma unidade em um particular dia;
   (b) não efetuar nenhuma venda em uma semana (considere a semana tendo cinco dias úteis);
   (c) em uma semana (cinco dias úteis) efetuar vendas em ao menos três dias.

Solução:

(a)

$$\begin{aligned} X_1 &: \text{número de vendas em um dia} \\ x_1 &\in \{0, 1, 2, \ldots\} \\ X_1 &\sim \mathrm{P}(\lambda = 0,5) \\ P[X_1 = 0] &= \frac{\mathrm{e}^{0,5} 0,5^0}{0!} = 0.3935 \end{aligned}$$

(b)

$$\begin{aligned} X_2 &: \text{número de vendas em uma semana (cinco dias)} \\ x_2 &\in \{0, 1, 2, \ldots\} \\ X_2 &\sim \mathrm{P}(\lambda = 2,5) \\ P[X_2 = 0] &= \frac{\mathrm{e}^{2,5} 2,5^0}{0!} = 0.08208 \end{aligned}$$

(c)

$$\begin{aligned} X_3 &: \text{número de dias com vendas em uma semana (cinco dias)} \\ x_3 &\in \{0, 1, 2, 3, 4, 5\} \\ X_3 &\sim \mathrm{B}(n = 5, p = P[X_1 = 0] \\ P[X_3 \geq 3] &= P[X_3 = 3] + P[X_3 = 4] + P[X_3 = 5] = 0.6938 \end{aligned}$$

Soluções computacionais (linguagem R):

```
> (q1 <- ppois(0, lambda=0.5, lower=FALSE))
[1] 0.3935
> (q2 <- ppois(0, lambda=0.5*5))
[1] 0.08208
> (q3 <- pbinom(2, size=5, prob=dpois(0, lambda=0.5), lower=FALSE))
[1] 0.6938
```

---

3. Seja a função:

$$f(x) = \begin{cases} 3x^2/8 & 0 < x \leq 2 \\ 0 & \text{caso contrário} \end{cases}$$

(a) Mostre que $f(x)$ é uma função de densidade de probabilidade válida.
(b) Obtenha $P[0,5 < X < 1,5]$.
(c) Obtenha $P[X > 1,2]$.
(d) Obtenha $P[X > 1,2|X > 0,5]$.
(e) Obtenha o valor esperado de X.

Solução:
OBS: Pode-se escrever a f.d.p. utilizando a função indicadora na forma: $f(x) = \frac{3x^2}{8}\mathrm{I}_{(0,2]}(x)$.
(a)

$$\text{Mostrar que: } f(x) \geq 0 \; \forall x \quad \text{e} \quad \int_0^2 f(x)\mathrm{d}x = 1$$

$$\int_0^2 \frac{3x^2}{8}\mathrm{d}x = \frac{3}{8}\frac{x^3}{3}\Big|_0^2 = \frac{3}{8}\frac{2^3 - 0^3}{3} = 1$$

a função acumulada $F(x)$ é dada por:

$$F(x) = \int_0^x f(x)\mathrm{d}x = \frac{3}{8}\frac{x^3 - 0^3}{3} = \frac{x^3}{8}\mathrm{I}_{(0,2]}(x).$$

(b) $P[0,5 < X < 1,5] = \int_{0,5}^{1,5} f(x)\mathrm{d}x = F(1,5) - F(0,5) = 0.406$
(c) $P[X > 1,2] = \int_{1,2}^{2} f(x)\mathrm{d}x = 1 - F(1,2) = 0.784$
(d) $P[X > 1,2|X > 0,5] = \frac{\int_{1,2}^{2} f(x)\mathrm{d}x}{\int_{0,5}^{2} f(x)\mathrm{d}x} = \frac{1-F(1,2)}{1-F(0,5)} = 0.796$
(e)

$$E[X] = \int_0^2 x \cdot f(x)\mathrm{d}x = \frac{3}{8}\left[\frac{2^4 - 0^4}{4}\right] = \frac{3}{2} = 1,5$$

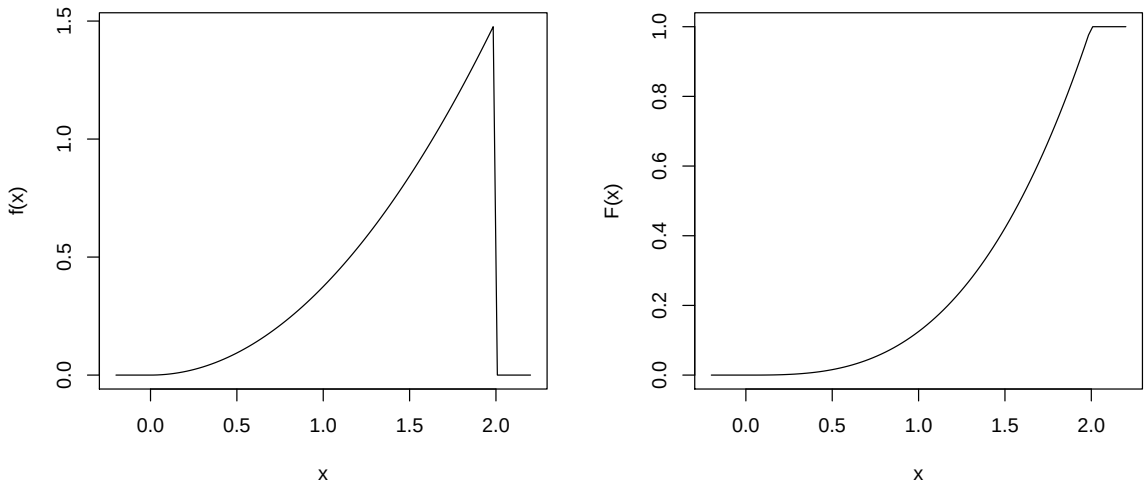


Figura 1: Função de densidade de probabilidade (esquerda) e função de distribuição (direita).

Soluções computacionais (linguagem R):

```
> require(MASS)
> ## a)
> fx <- function(x) ifelse(x > 0 & x <= 2, (3*x^2)/8, 0)
> integrate(fx, 0, 2)$value
[1] 1
> Fx <- function(x) ifelse(x>0, ifelse(x<=2, (x^3)/8,1), 0)
> Fx(2)
[1] 1
> ## b)
> integrate(fx, 0.5, 1.5)$value
[1] 0.4062
> Fx(1.5)-Fx(0.5)
[1] 0.4062
> ##c)
> integrate(fx, 1.2, 2)$value
[1] 0.784
> 1-Fx(1.2)
[1] 0.784
> ## d)
> integrate(fx, 1.2, 2)$value/integrate(fx, 0.5, 2)$value
[1] 0.7964
> (1-Fx(1.2))/(1-Fx(0.5))
[1] 0.7964
> ## e)
> efx <- function(x) ifelse(x > 0 & x <= 2, x*(3*x^2)/8, 0)
> integrate(efx, 0, 2)$value
[1] 1.5
```

Semana 5 (av-03)

1. Seja uma v.a. X com distribuição normal de média $\mu = 250$ e variância $\sigma^2 = 225$. Obtenha:
   (a) $P[X > 270]$.
   (b) $P[X < 220]$.
   (c) $P[|X - \mu| > 25]$.
   (d) $P[|X - \mu| < 30]$.
   (e) $P[X < 270|X > 250]$.
   (f) o valor x_1 tal que $P[X > x_1] = 0,80$.
   (g) o valor x_2 tal que $P[X < x_2] = 0,95$.
   (h) qual deveria ser um novo valor da média μ para que $P[X < 240] \leq 0,10$?
   (i) com $\mu = 250$ qual deveria ser um novo valor da variância σ^2 para que $P[X < 240] \leq 0,10$?
   (j) qual deveria ser um novo valor da variância σ^2 para que $P[|X - \mu| > 15] \leq 0,10$?

   Solução:

   $$X \sim \mathrm{N}(250, 15^2)$$

   (a) $P[X > 270] = P[Z > \frac{270-250}{15}] = P[Z > 1.3333] = 0.0912$
   (b) $P[X < 220] = P[Z < \frac{220-250}{15}] = P[Z < -2] = 0.0228$
   (c) $P[|X - \mu| > 25] = P[X < 225 \cup X > 275] = P[X < -1.667] + P[X > 1.667] = 0.0956$
   (d) $P[|X - \mu| < 30] = P[220 < X < 280] = P[-2 < X < 2] = 0.9545$
   (e) $P[X < 270|X > 250] = \frac{P[250<X<270]}{P[X>250]} = \frac{0.4088}{0.5} = 0.8176$
   (f) $z = \frac{x_1-250}{15} = -0.842 \longrightarrow x_1 = 237.4$
   (g) $z = \frac{x_2-250}{15} = 1.645 \longrightarrow x_2 = 274.7$

(h) $z = \frac{240-\mu}{15} = -1.282 \longrightarrow \mu = 259.2$

(i) $z = \frac{240-250}{\sigma} = -1.282 \longrightarrow \sigma = 7.8 \longrightarrow \sigma^2 = 60.8$

(j) $P[|X < \mu| > 15] = P[X < \mu - 15 \cup X > \mu + 15] \leq 0,10 \longrightarrow z = \frac{15}{\sigma} = 1.645 \longrightarrow \sigma = 9.1 \longrightarrow \sigma^2 = 83.1$

Comandos em R para soluções:

```
> (qa <- pnorm(270, mean=250, sd=15, lower=FALSE))
[1] 0.09121
> (qb <- pnorm(220, mean=250, sd=15))
[1] 0.02275
> (qc <- 2*pnorm(250-25, mean=250, sd=15))
[1] 0.09558
> (qd <- diff(pnorm(c(250-30,250+30), mean=250, sd=15)))
[1] 0.9545
> (qe <- diff(pnorm(c(250,270), mean=250, sd=15))/pnorm(250, mean=250, sd=15, lower=FALSE))
[1] 0.8176
> (qf <- qnorm(0.80, mean=250, sd=15, lower=FALSE))
[1] 237.4
> (qg <- qnorm(0.95, mean=250, sd=15))
[1] 274.7
> (qh <- 240 - 15 * round(qnorm(0.10), dig=3))
[1] 259.2
> (qi <- (240 - 250)/round(qnorm(0.10), dig=3))
[1] 7.8
> (qj <- 15/round(qnorm(0.95), dig=3))
[1] 9.119
```

Semana 6 (av-04)

1. Suponha que os escores obtidos por estudantes em um teste *online* possam ser bem modelados por uma distribuição normal com média $\mu = 120$ e variância $\sigma^2 = 12^2$.
   (a) Considera-se como estudante de alta performance os que atingem um escore a partir de 135. Qual o percentual esperado de estudantes de alta performance entre todos os que fazem o teste?
   (b) Estudantes com escore abaixo de 100 devem se reinscrever e só podem voltar a fazer o teste após seis meses e os com escore entre 100 e 125 são convidados a refazer o teste após um mês. Quais as proporções de estudantes que deverá se reinscrever e que deverá refazer o teste após um mês?
   (c) Define-se como *quartis* os escores abaixo dos quais espera-se encontrar 25, 50 e 75% dos estudantes. Quais os valores dos escores que definem os quartis?
   (d) Quanto deveria ser o valor μ da média dos escores para que ao menos 30% dos escores fossem de alta performance?
   (e) Há um outro teste que possui média $\mu = 125$ e variância $\sigma^2 = 6^2$. Em qual deles espera-se a maior proporção de estudantes de alta performance?

   Solução:

$$X \sim \mathrm{N}(120, 12^2)$$

(a) $P[X > 135] = P[Z > \frac{135-120}{12}] = P[Z > 1.25] = 0.1056$

(b)

$$P[X < 100] = P[Z < \frac{100-120}{12}] = P[Z < -1.6667] = 0.0478$$

$$P[100 < X < 125] = P[\frac{100-120}{12} < Z < \frac{125-120}{12}] = P[-1.67 < Z < 0.417]$$

(c)

$$P[X < Q_1] = 0,25$$
$$z_1 = -0.674 = \frac{Q_1 - 120}{12}$$
$$Q_1 = 120 - 8.09 = 112$$

Usando o fato de que a distribuição é simétrica temos ainda que:

$$Q_2 = \mu = 120$$
$$Q_3 = 120 + 8.09 = 128$$

(d) $z = \frac{135-\mu}{15} = 0.524 \longrightarrow \mu = 128.7$

(e)

$$X_1 \sim \mathrm{N}(120, 12^2)$$
$$X_2 \sim \mathrm{N}(125, 6^2)$$
$$P[X_1 \geq 135] = P[Z_1 > \frac{135 - 120}{12}] = P[Z_1 > 1.25] = 0.106$$
$$P[X_2 \geq 135] = P[Z_2 > \frac{135 - 120}{12}] = P[Z_2 > 1.67] = 0.0478$$

Comandos em R para soluções:

```
> (qa <- pnorm(135, mean=120, sd=12, lower=FALSE))
[1] 0.1056
> (qb <- diff(pnorm(c(-Inf, 100, 125), mean=120, sd=12)))
[1] 0.04779 0.61375
> (qc <- qnorm(c(.25, .50, .75), mean=120, sd=12))
[1] 111.9 120.0 128.1
> (qd <- 135 - 12 * round(qnorm(0.70), dig=3))
[1] 128.7
> (qez <- (135 - c(120, 125))/c(12, 6))
[1] 1.250 1.667
> (qep <- pnorm(135, m=c(120, 125), sd=c(12, 6), lower=FALSE))
[1] 0.10565 0.04779
```

---

2. Semana 07 (av-05)

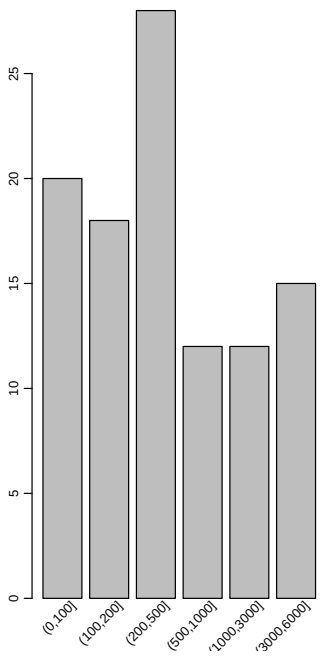
   (a) Obteve-se para análise um conjunto de dados com informações de 105 países sobre (i) renda per capta (em dólares), (ii) taxa de moralidade infantil (por 1000 nascidos vivos), (iii) região sendo os valores: 'Africa'; 'Americas'; 'Asia e Oceania´ e 'Europe', (iv) se o país é ou não exportador de petróleo (sim/não). A seguir são mostrados dos 10 primeiros registros da tabela de dados.

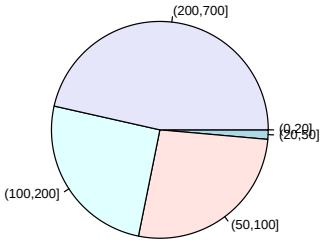
	income	infant	region	oil
Australia	3426	26.7	Asia	no
Austria	3350	23.7	Europe	no
Belgium	3346	17.0	Europe	no
Canada	4751	16.8	Americas	no
Denmark	5029	13.5	Europe	no
Finland	3312	10.1	Europe	no
France	3403	12.9	Europe	no
West.Germany	5040	20.4	Europe	no
Ireland	2009	17.8	Europe	no
Italy	2298	25.7	Europe	no

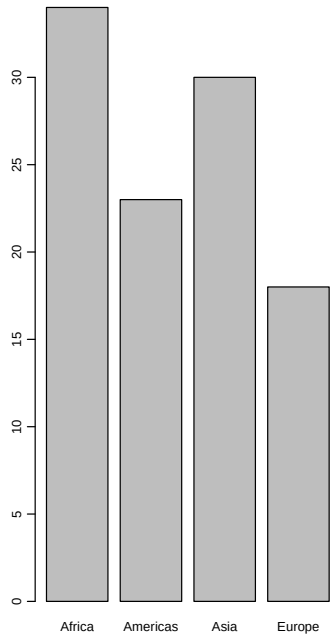
   O objetivo inicial é fazer análises descritivas com estes dados. Com isto em mente responda aos item a seguir.

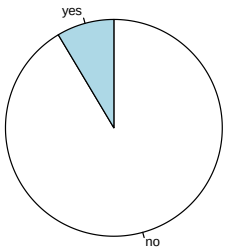
   i. classifique cada uma das variáveis (atributos) da tabela de dados quanto ao seu tipo

   ii. verifique os gráficos a seguir e indique para cada um deles se é ou não o mais adequado para a variável em questão, justificando sua resposta.

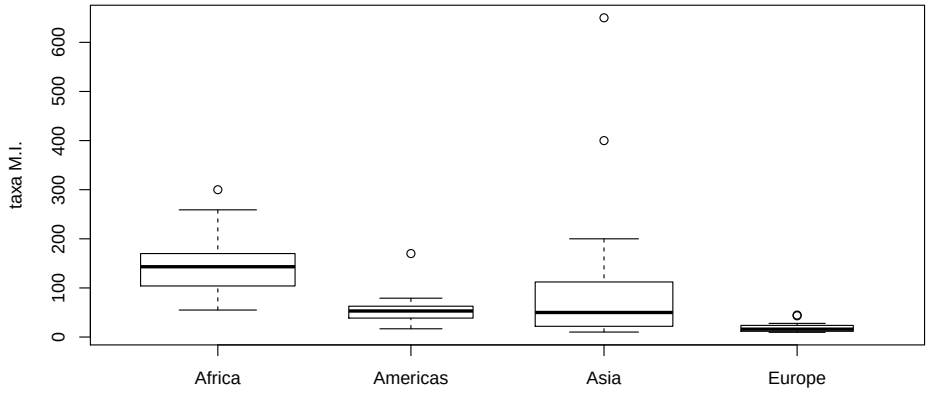








iii. O gráfico a seguir compara as mortalidades infantis entre as regiões. Discuta sua interpretação incluindo comentários sobre: medidas de posições, dispersão, assimetria dos gráficos e pontos discrepantes.

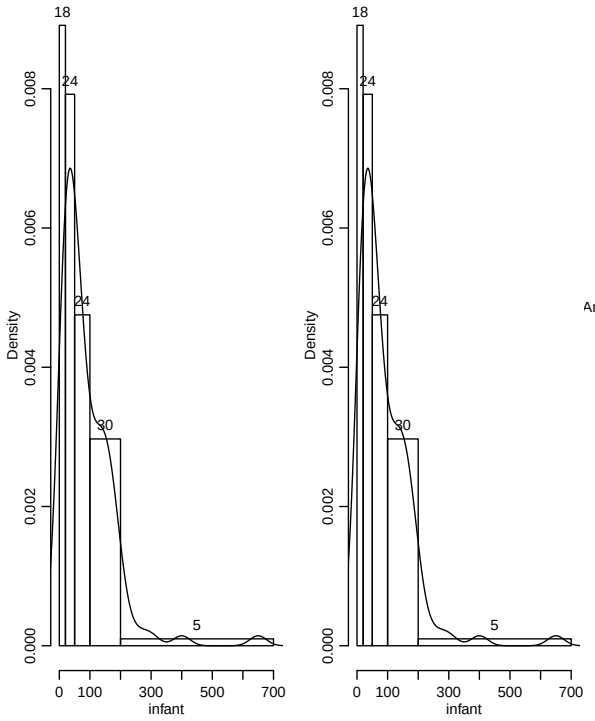


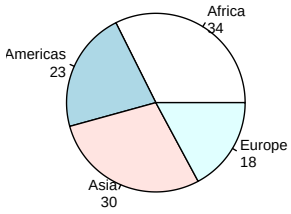
iv. Esboque uma análise utilizando gráficos, tabelas e medidas para investigar se há relação entre: (i) renda e região geográfica, (ii) renda e mortalidade infantil, (iii) região geográfica e produção de petróleo.

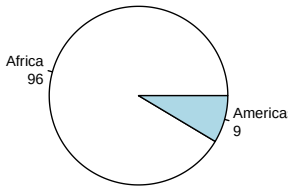
Solução: (parcial)

i. **income (renda):** quantitativa contínua
infant (taxa de mortalidade infantil (x1000)): quantitativa contínua
region (região geográfica): qualitativa nominal
oil (produtor de petróleo - sim/não): qualitativa nominal

ii. Gráficos: comentários e figuras com gráficos sugeridos.
- inadequado pois a variável é quantitativa contínua. O gráfico de barras é recomendado para qualitativas ordinais. Uma opção adequada seria um histograma. Notar classes desiguais e o aspecto diferente dos gráficos.
- inadequado pois a variável é quantitativa contínua. O gráfico de setores é recomendado para qualitativas nominais. Uma opção adequada seria um histograma. Notar classes desiguais e o aspecto diferente dos gráficos.
- inadequado pois a variável é qualitativa nominal. O gráfico de setores seria mais adequado.
- adequado para uma qualitativa nominal.

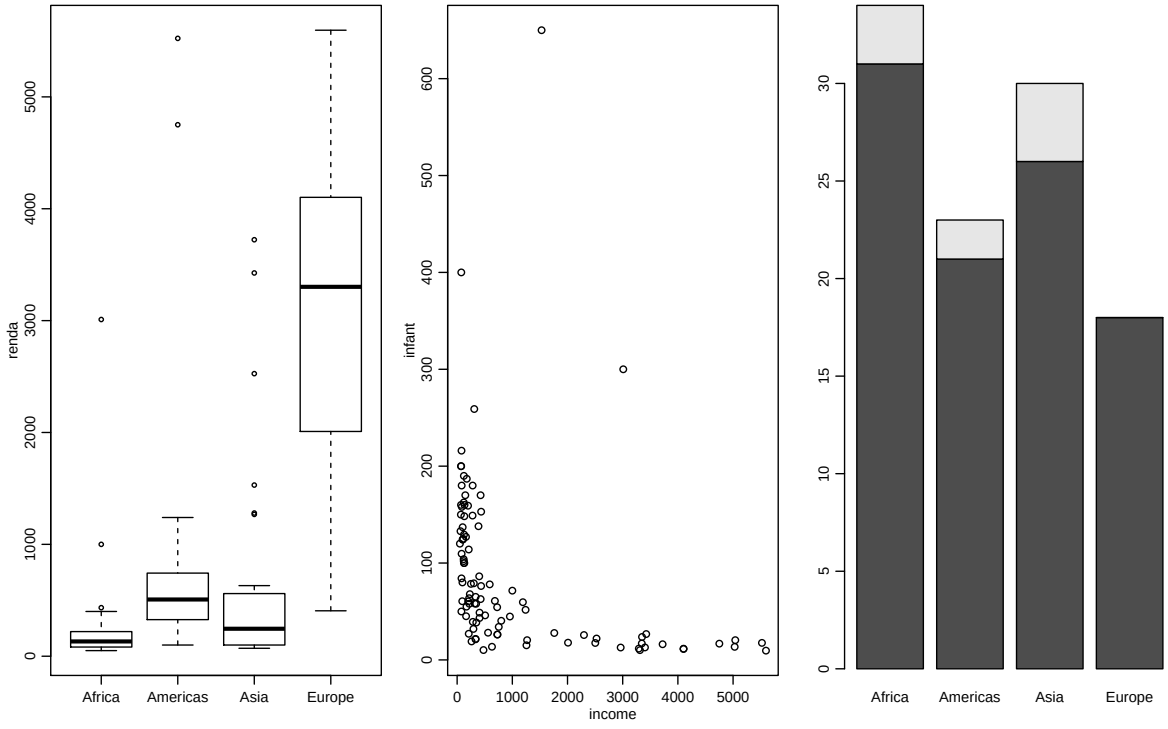




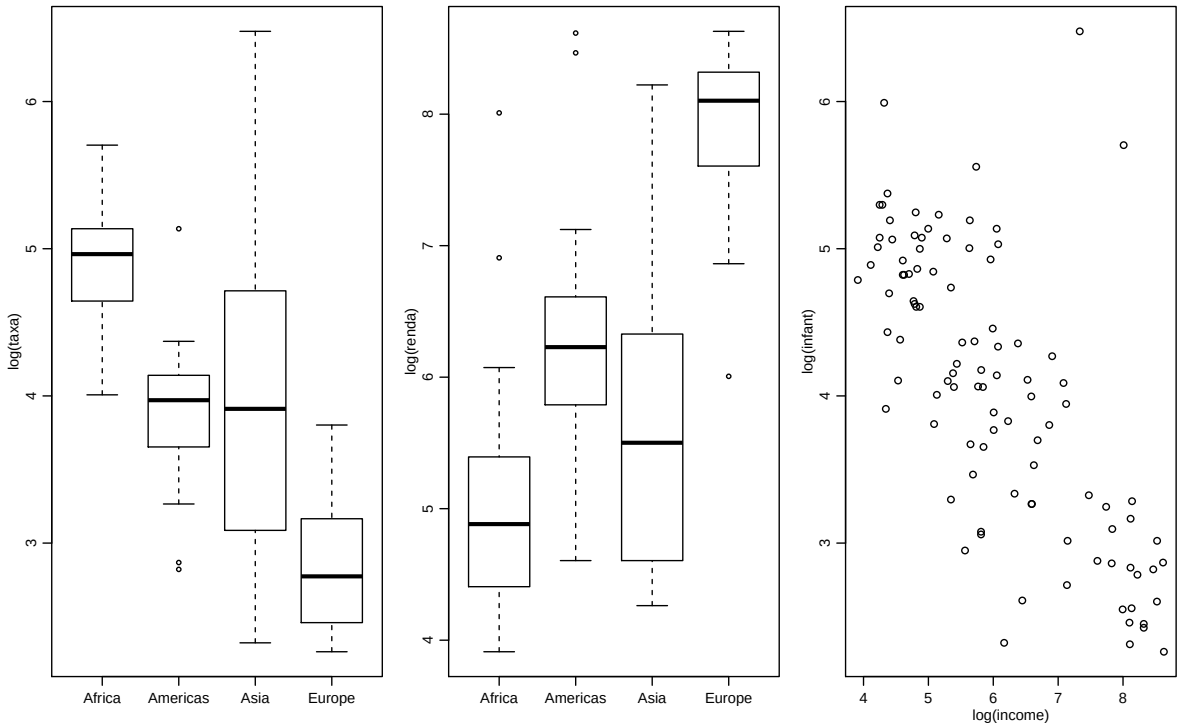


iii.

iv. Apenas alguns possíveis gráficos mostrados aqui. Incluir comentários osbre tabelas e medidas



Transformação: por vezes variáveis são melhro expressas em escalas transformadas. Veja a seguir, por exemplo, gráficos com os logarítmos de renda e taxas.



Comandos em R para questões e soluções:

```
> #install.packages("car")
> require(car)
> data(Leinhardt)
> #dim(Leinhardt)  # 105 4
> head(Leinhardt, n=10)
> # gráficos item (b)
> par(mar=c(3,3.5, 0.5, 0.3), mgp=c(1.7, 0.7, 0), mfrow=c(1,4))
> BR1 <- c(0, 100, 200, 500, 1000, 3000, 6000)
> with(Leinhardt, barplot(table(cut(income, br=BR1, dig.lab=3)), names.arg=FALSE))
> CL <- names(with(Leinhardt, table(cut(income, br=BR1, dig.lab=4))))
> text(x = 1:6*1.2, par("usr")[3], labels = CL, srt = 45, pos = 2, xpd = TRUE)
> BR2 <- c(0, 20,50, 100, 200, 700)
> par(mar=c(3,3,3,3))
> with(Leinhardt, pie(table(cut(income, br=BR2,dig.lab=4)),radius=0.9, clock=T, init.angle=0))
> par(mar=c(2.8,2.8,0.3,0.3))
> with(Leinhardt, barplot(table(region)))
> par(mar=c(3,3,3,3))
> with(Leinhardt, pie(table(oil),radius=0.9, clock=T))
> # gráfico item (c)
> with(Leinhardt, boxplot(infant ~ region, varwidth=TRUE, ylab="taxa M.I."))
> par(mar=c(2.8,3.5, 0.5, 0.3), mgp=c(1.7, 0.7, 0), mfrow=c(1,4))
> BR1 <- c(0, 20, 50, 100, 200, 700)
> H1 <- with(Leinhardt, hist(infant, breaks=BR1, main=""))
> text(H1$mids, H1$dens, as.character(H1$counts), pos=3)
> with(Leinhardt, lines(density(infant, na.rm=TRUE)))
> BR2 <- c(0, 100, 200, 500, 1000, 3000, 6000)
> H1 <- with(Leinhardt, hist(infant, breaks=BR1, main=""))
> text(H1$mids, H1$dens, as.character(H1$counts), pos=3)
> with(Leinhardt, lines(density(infant, na.rm=TRUE)))
> par(mar=c(3,3,3,3))
> T3 <- with(Leinhardt, table(region))
> P3 <- paste(names(T3), "\n", T3, sep="")
> with(Leinhardt, pie(T3, labels=P3, radius=1))
> T4 <- with(Leinhardt, table(oil))
> P4 <- paste(names(T3), "\n", T4, sep="")
> with(Leinhardt, pie(T4, labels=P4, radius=1))
> par(mar=c(2.8,3.5, 0.5, 0.3), mgp=c(1.7, 0.7, 0), mfrow=c(1,3))
> with(Leinhardt, boxplot(income ~ region, ylab="renda"))
> with(Leinhardt, plot(infant ~ income))
```

```
> with(Leinhardt, barplot(table(oil, region)))
> par(mar=c(2.8,3.5, 0.5, 0.3), mgp=c(1.7, 0.7, 0), mfrow=c(1,3))
> with(Leinhardt, boxplot(log(infant) ~ region, ylab="log(taxa)"))
> with(Leinhardt, boxplot(log(income) ~ region, ylab="log(renda)"))
> with(Leinhardt, plot(log(infant) ~ log(income)))
```

Semana 10 (av-06)

1. O (log-)tempo de processamento de *jobs* submetidos a um servidor de processamento possui um distribuição aproximandamente normal com média de 5,4 e desvio padrão de 1,2 (log-)minutos.
   (a) Qual a proporção esperada de *jobs* que devem ser concluídos em menos do que 5 (log-)minutos?
   (b) Qual a probabilidade de que um lote de 10 *jobs* seja processado em menos que 50 (log-)minutos?
   (c) Qual o intervalo de (log-)tempos ao redor da média de 5,4, que se espera que 95% de lotes de 16 *jobs* sejam processados?
   (d) Quantos *jobs* deveriam ser submetidos em um lote para que o tempo médio de processamento fosse, com probabilidade de 0,90, de no máximo de 6 (log-)minutos.

   Solução:

$$
\begin{aligned}
&X : \text{(log-)tempo de processamento} \\
&X \sim \mathrm{N}(5,4; 1, 2^2) \\
&\overline{X}_n : \text{(log-)tempo médio de processamento de } n \textit{ jobs} \\
&\overline{X}_n \sim \mathrm{N}(5,4; 1, 2^2)
\end{aligned}
$$

   (a) $P[X < 5] = 0.369$
   (b) $P[\sum_{i=1}^{10} X_i < 50] = P[\overline{X}_{10} < 5] = 0.146$
   (c) $P[t_1 < \overline{X}_{16} < t_2] = 0,95 \longrightarrow (t_1 = 4.81, t_2 = 5.99$
   (d) $P[\overline{X}_n < 5,5] = 0,90 \longrightarrow n = 7$

   Comandos em R para soluções:

```
> q1a <- pnorm(5, mean=5.4, sd=1.2)
> q1b <- pnorm(5, mean=5.4, sd=1.2/sqrt(10))
> q1c <- qnorm(c(0.025, 0.975), mean=5.4, sd=1.2/sqrt(16))
> q1d <- ceiling(qnorm(0.9)^2 * 1.2^2/(0.6^2))
```

---

2. Ainda no contexto do problema anterior acredita-se que 18% dos *jobs* encerram-se, por alguma razão) sem produzir o resultado correto.
   (a) Qual a probabilidade de que em um lote de 200 *jobs*, ao menos 80% deles produzam o resultado correto?
   (b) Decide-se fazer uma inspeção diária submetendo um lote de 150 *jobs* de controle. Se 30 ou mais não produzem o resultado correto o serviço é interrompido para uma inspeção na servidora. Qual a probabilidade de uma interrupção desnecessária?
   (c) Qual deveria ser o número limite de *jobs* sem resultado correto para que tal probabilidade de interrupção desnecessária fosse de no máximo 0,15?
   (d) Deseja-se manter o limite de 20% de *jobs* sem resultado correto como critério de interrupção com no máximo 0,15 de probabilidade de uma parada desnecesária. Neste caso, quantos *jobs* de controle deveriam ser testados?

   Solução:

$$
\begin{aligned}
&X : \text{número de } \textit{jobs} \text{ sem resultado correto} \\
&X \sim \mathrm{B}(n; p = 0,18) \\
&p_n : \text{proporção de } \textit{jobs} \text{ sem resultado correto em um lote de } n \textit{ jobs} \text{ submetidos} \\
&p_n \sim \mathrm{N}(0,18; \frac{0,18(1-0,18)}{n})
\end{aligned}
$$

   (a) $P[p_{200} \leq 0,20] = 0.769$
   (b) $P[p_{150} \geq \frac{30}{150}] = 0.262$

(c) $P[p_{150} \geq \frac{C}{150}] \leq 0,15 \longrightarrow C = 32$

(d) $P[p_n \geq 0,20] \leq 0,15 \longrightarrow n = 397$

Comandos em R para soluções:

```
> q2a <- pnorm(0.20, mean=0.18, sd=sqrt(0.18*(1-0.18)/200))
> q2b <- pnorm(30/150, mean=0.18, sd=sqrt(0.18*(1-0.18)/150), lower=FALSE)
> q2c <- ceiling(150 * qnorm(0.15, mean=0.18, sd=sqrt(0.18*(1-0.18)/150), lower=FALSE))
> q2d <- ceiling((qnorm(0.85)^2 * 0.18 * (1-0.18))/(0.20-0.18)^2)
```

Semana 11 (av-07)

1. Foi registrado o número de transações concretizadas em um sistema obtendo-se os seguintes valores:

$$3, 7, 2, 1, 5, 6, 3, 7, 8, 5, 6, 4, 3, 5, 2, 5, 4.$$

Deseja-se ajustar uma distribuição de Poisson ($P[X = x_i] = \mathrm{e}^{-\lambda}\lambda^{x_i}/x_i!$) aos dados. Assumindo-se independência entre os observações,

(a) encontre a expressão do estimador de máxima verossimilhança de λ,

(b) encontre o valor da estimativa de máxima verossimilhança obtida com os dados acima,

(c) usando a estimativa encontrada, encontre a probabilidade estimada de se obter uma ou menos transações em um dia qualquer.

Solução:

$$\begin{aligned} X &: \text{número de transações concretizadas} \\ X &\sim \mathrm{P}(\lambda) \\ P[X = x_i] &= \frac{\mathrm{e}^{-\lambda}\lambda_i^x}{x_i!} \end{aligned}$$

(a)

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} P[X = x_i] = \prod_{i=1}^{n} \frac{\mathrm{e}^{-\lambda}\lambda_i^x}{x_i!} \\ l(\lambda) &= \log[L(\lambda)] = \sum_{i=1}^{n} \log\left[\frac{\mathrm{e}^{-\lambda}\lambda_i^x}{x_i!}\right] \\ &= \sum_{i=1}^{n} [-\lambda + x_i \log(\lambda) - \log(x_i!)] \\ &= -n\lambda + (\sum_{i=1}^{n} x_i)\log(\lambda) - \sum_{i=1}^{n} \log(x_i!) \\ \frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} &= -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} \\ \frac{\mathrm{d}l(\lambda)}{\mathrm{d}\lambda} &= 0 \longrightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X} \end{aligned}$$

(b) $\overline{x} = 4.47$

(c) $\hat{P}[X \leq 1] = \hat{P}[X = 1] + \hat{P}[X = 1] = \frac{\mathrm{e}^{-4.47}4.47^0}{0!} + \frac{\mathrm{e}^{-4.47}4.47^1}{1!} = 0.0626$

Comandos em R para soluções:

```
> dados <- c(3, 7, 2, 1, 5, 6, 3, 7, 8, 5, 6, 4, 3, 5, 2, 5, 4)
> m <- mean(dados)
> p <- ppois(1, lambda=m)
```

---

2. Uma locadora de veículos que possui uma grande frota decide fazer um estudo sobre vários aspectos relacionados ao desempenho. Para isto vai tomar uma amostra aleatória de 25 de seus veículos para inspeções detalhadas. Várias características serão medidas, mas vamos aqui nos ater apenas ao rendimento de combustível, supondo que a variância do rendimento de toda a frota é de $6,25(km/l)^2$.

(a) Qual a probabilidade do consumo médio aferido nos 25 veículos, diferir do consumo médio de toda a frota em mais que $0,5km/l$? E em mais que $1km/l$?

(b) Qual a margem de erro na estimação do consumo médio da frota para uma confiança de 95%?

(c) Qual deveria ser o tamanho da amostra para que a margem de erro fosse a metade da calculada no item anterior?

(d) Se uma amostra (com $n = 25$) fornecer uma estimativa intervalar de $(11,1\,;\,12,3)km/l$, qual a confiança desta estimativa?

(e) identifique no problema: a população variável aleatória de interesse, o parâmetro de interesse, o estimador, a distribuição amostral, a estimativa pontual e a estimativa intervalar.

Solução:

$$X : \text{consumo de veículo da frota}$$

Distribuição da variável aleatória (população):

$$X \sim \text{Dist.}(\mu_X = \text{E}[X], \sigma_X^2 = \text{Var}[X])$$

distribuição amostral:

$$\overline{X} \approx \text{N}(\mu_{\overline{X}} = \mu_X, \sigma_{\overline{X}} = \sigma_X^2/n)$$

ou, equivalentemente,

$$\overline{X} - \mu \approx \text{N}(0, \sigma^2/n)$$

$$\sigma_X^2 = 6,25\ ;\ \sigma_{\overline{X}}^2 = 0,25\ ;\ \sigma_{\overline{X}} = 0,5$$

(a)

$$P[|\overline{X} - \mu| > 0,5] = P[\frac{|\overline{X} - \mu|}{\sigma} > 0,5/\sigma] = P[|Z| > 1] = 0.317$$

$$P[|\overline{X} - \mu| > 1] = P[\frac{|\overline{X} - \mu|}{\sigma} > 1/\sigma] = P[|Z| > 2] = 0.0455$$

(b)

$$ME = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n}}$$

$$ME = 1.96\frac{2,5}{\sqrt{25}}$$

$$ME = 0.98$$

(c)

$$\frac{ME}{2} = z_{0,95}\sigma_{\overline{X}} = z_{0,95}\frac{\sigma_X}{\sqrt{n^*}}$$

$$n^* = \lceil(\frac{2}{ME})^2 z_{0,95}^2 \sigma_X^2$$

$$n^* = \lceil(\frac{2}{0.98})^2 1.96^2 6,25$$

$$n^* = 100$$

(d)

$$(11,1\,;\,12,3) \equiv 11,7 \pm 0,6$$

$$ME = z_{1-\alpha}\sigma_{\overline{X}} = z_{1-\alpha}\frac{\sigma_X}{\sqrt{n}}$$

$$0,6 = z_{1-\alpha}\frac{2,5}{\sqrt{25}}$$

$$z_{1-\alpha} = \frac{0,6}{0,5}$$

$$1 - \alpha(\text{confiança}) = 0.77(77\%)$$

(e)

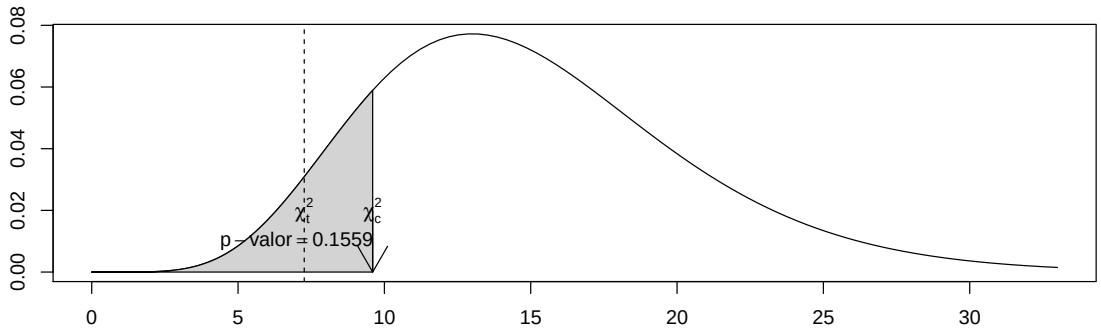
1. O tempo médio, por programador, para executar uma tarefa, tem sido 100 minutos, com um desvio padrão de 15 minutos. Introduziu-se uma modificação nos procedimentos do sistema para diminuir este tempo e também torná-lo mais homogêneo. Após certo período, sorteou-se uma amostra de 16 programadores, medindo-se o tempo para execução de cada um deles. O tempo médio da amostra foi de 85 minutos, e o desvio padrão de 12 minutos. Estes resultados trazem evidências estatísticas de melhoras tanto no tempo médio de execução quanto na maior homogeneidade dos tempos? Em cada critério (tempo médio e homogeneidade), caso afirmativo, forneça estimativas pontuais e intervalares dos parâmetros de interesse.

Solução:

i. Teste da variância

$$H_0 : \sigma^2 \geq (15')^2 \text{ vs } H_a : \sigma^2 < (15')^2 (\text{unilateral})$$
$$\alpha = 0,05 \longrightarrow \chi^2_t = 7.26$$
$$\chi^2_c = \frac{(n-1)S^2}{\sigma^2_0} = \frac{(16-1)12^2}{15^2} = 9.6$$
$$\text{p-valor} = 0.156$$

Decisão: não rejeita-se H0



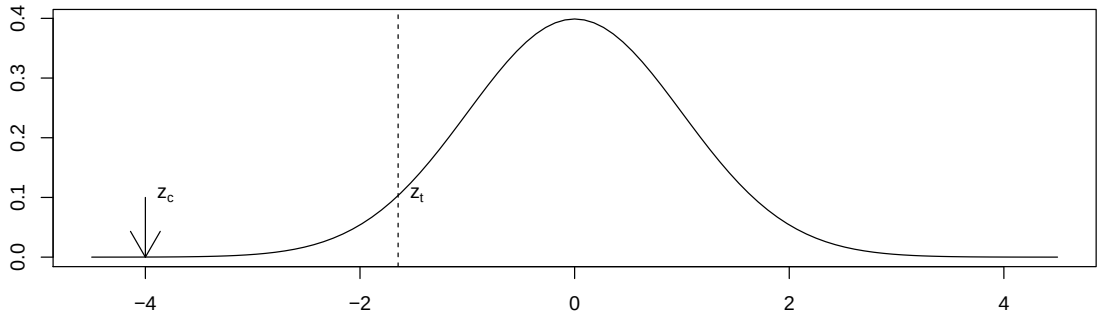
Comandos em R para soluções:

```
> (chi2t <- qchisq(0.05, df=16-1))
[1] 7.261
> (chi2c <- (16-1)*(12^2)/(15^2))
[1] 9.6
> (pvalor <- pchisq(chi2c, df=16-1))
[1] 0.1559
> curve(dchisq(x, df=16-1), from=0, to=33, xlab="", ylab="")
> polygon(cbind(c(chi2c, 0, seq(0, chi2c,l=100)),
+               c(0, 0, dchisq(seq(0, chi2c, l=100), df=16-1))),
+               col="lightgray")
> text(7, 0.01, substitute(p-valor==PV, list(PV=pvalor)))
> abline(v=chi2t, lty=2)
> arrows(chi2c, 0.02, chi2c, 0)
> text(chi2t, 0.02, expression(chi[t]^2))
> text(chi2c, 0.02, expression(chi[c]^2))
> D0 <- "não rejeita-se H0"
> Da <- "rejeita-se H0"
> ## resultado
> ifelse(chi2c > chi2t, D0, Da)
[1] "não rejeita-se H0"
> ## equivalentemente
> ifelse(pvalor > 0.05, D0, Da)
[1] "não rejeita-se H0"
```

ii. Teste da média

$$H_0 : \mu \geq 100 \text{ vs } H_a : \mu < 100 (\text{unilateral})$$
$$\alpha = 0,05 \longrightarrow z_t = -1.64$$
$$z_c = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{85 - 100}{15/\sqrt{16}} = -4$$
$$\text{p-valor} = 0.00003$$

Decisão: rejeita-se H0



Comandos em R para soluções:

```
> (zt <- qnorm(0.05))
[1] -1.645
> (zc <- (85-100)/(15/sqrt(16)))
[1] -4
> (pvalor <- pnorm(zc))
[1] 3.167e-05
> curve(dnorm(x), from=-4.5, to=4.5, xlab="", ylab="")
> abline(v=zt, lty=2)
> arrows(zc, 0.1, zc, 0)
> text(zt, 0.1, expression(z[t]), pos=4)
> text(zc, 0.1, expression(z[c]), pos=4)
> D0 <- "não rejeita-se H0"
> Da <- "rejeita-se H0"
> ## resultado
> ifelse(zc > zt, D0, Da)
[1] "rejeita-se H0"
> ## equivalentemente
> ifelse(pvalor > 0.05, D0, Da)
[1] "rejeita-se H0"
```

---

2. Um fabricante garante que 90% dos equipamentos que fornece a uma fábrica estão de acordo com as especificações exigidas. O exame de uma amostra de 200 peças deste equipamento revelou 25 defeituosas. Teste (estatisticamente) a afirmativa do fabricante para os níveis de significância de 5 e 1%.

Solução:

X : número de defeituosas

$X \sim \mathrm{B}(n = 200, p = 0,10)$

Distribuição amostral utilizada:

$$\hat{p} \sim \mathrm{N}\left(\mu_p = p, \sigma_p^2 = \frac{p(1-p)}{n}\right)$$

Hipóteses:

$H_0 : p \leq 0,10$ vs $H_a : p > 0,10$(unilateral)

Nível de significância:

$\alpha = 0,05 \longrightarrow z_t =$

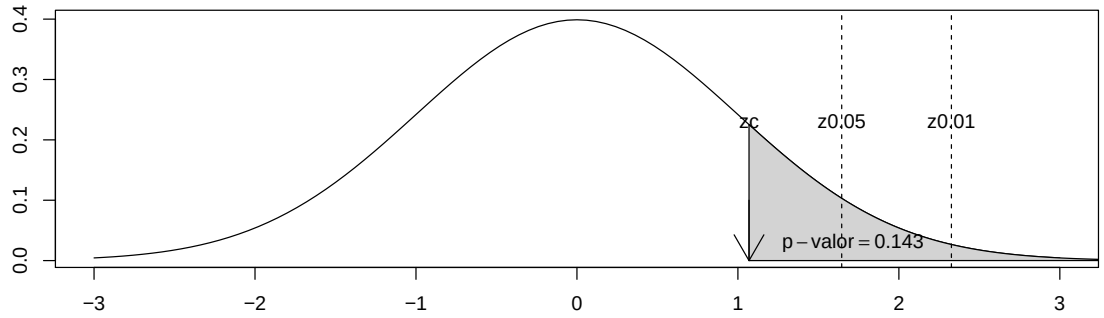
$\alpha = 0,01 \longrightarrow z_t =$

Estimativa pontual:

$$\hat{p} = \frac{25}{200} = 0,125$$

Estatística de teste:

$$z_c = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0,125 - 0,10}{\sqrt{\frac{0,10(1-0,10)}{200}}} = 1.07$$



Comandos em R para soluções:

```
> ## calculos passo a passo
> p.est <- 25/200
> sd.p <- sqrt(p.est*(1-p.est)/200)
> zc <- ((p.est)-0.10)/sd.p
> D0 <- "não rejeita-se H_0"
> Da <- "rejeita-se H_0"
> ## para 5%
> ifelse(zc < qnorm(0.95), D0, DA)
[1] "não rejeita-se H_0"
> ## para 1%
> ifelse(zc < qnorm(0.99), D0, DA)
[1] "não rejeita-se H_0"
> ## alternativamente, usando p p-valor
> (pv <- pnorm(zc, lower=FALSE))
[1] 0.1425
> ifelse(pv > 0.05, D0, DA)
[1] "não rejeita-se H_0"
> ifelse(pv > 0.01, D0, DA)
[1] "não rejeita-se H_0"
> ##
> ## Comandos para gráfico ilustranto resultado do teste
```

```
> curve(dnorm(x), from=-3, to=3, ylab="", xlab="")
> polygon(cbind(c(zc,seq(zc, 4, l=100),4), c(0, dnorm(seq(zc, 4, l=100)), dnorm(4))), col="lightgray")
> abline(v=qnorm(c(0.95, 0.99)), lty=2)
> arrows(zc, 0.1, zc, 0)
> text(c(zc, qnorm(c(0.95, 0.99))), 0.2, c("zc", "z0.05", "z0.01"), pos=3)
> text(1.2, 0.025, substitute(p-valor==PV, list(PV=round(pv, dig=3))), pos=4)
> ##
> ## teste já implementado no R
> prop.test(25, 200, 0.10, alternative="greater", conf.level=0.95)

1-sample proportions test with continuity correction

data:  25 out of 200, null probability 0.1
X-squared = 1.1, df = 1, p-value = 0.1
alternative hypothesis: true p is greater than 0.1
95 percent confidence interval:
 0.08933 1.00000
sample estimates:
    p
0.125

> prop.test(25, 200, 0.10, alternative="greater", conf.level=0.99)

1-sample proportions test with continuity correction

data:  25 out of 200, null probability 0.1
X-squared = 1.1, df = 1, p-value = 0.1
alternative hypothesis: true p is greater than 0.1
99 percent confidence interval:
 0.07831 1.00000
sample estimates:
    p
0.125
```
