13. Os produtores de um programa de televisão pretendem modificá-lo se for assistido regularmente por menos de um quarto dos possuidores de televisão. Uma pesquisa encomendada a uma empresa especializada mostrou que, de 400 famílias entrevistadas, 80 assistem ao programa regularmente. Com base nos dados, qual deve ser a decisão dos produtores?

Solução: Cada possuidor de televisão assiste ao programa (Sucesso=1) ou não(Fracasso=0). Assim criamos uma variável aleatória de Bernoulli de parâmetro p em

que

$$p = P(X = 1) \quad e \quad q = 1 - p = P(X = 0).$$

Seja $S_n = \sum_{i=1}^{n} X_i$ o número de possuidores de TV que assistem ao programa

Um estimador não viciado para p será:

$$\hat{p} = \frac{S_n}{n} = \bar{X}.$$

O problema pede uma decisão dos produtores do programa se o modificam ou não.

na realidade temos um teste de hipóteses:

$$H_0 : p \geq \frac{1}{4} \quad (\quad \text{não modificam o programa}),$$

versus

$$H_1 : p < \frac{1}{4} \quad (\quad \text{modificam o programa}).$$

Este teste é equivalente a:

$$H_0 : p = \frac{1}{4} \quad (\quad \text{não modificam o programa}),$$

versus

$$H_1 : p < \frac{1}{4} \quad (\quad \text{modificam o programa}).$$

Para fazer um teste de hipótese Bussab&Morettin sugerem na página 353 os seguintes passos:

Passo 1. Fixe qual a hipótese H_0 a ser testada e qual a hipótese alternativa H_1.

Passo 2. Use a teoria estatística e as informações disponíveis para decidir qual estatística (estimador) será usada para testar a hipótese V. Obter as propriedades dessa estatística (distribuição, média, desvio padrão).

Passo 3. Fixe a probabilidade α de cometer o erro de tipo I e use este valor para construir a região crítica (regra de decisão). Lembre que essa região é construída para a estatística definida no passo 2, usando os valores do parâmetro hipotetizados por H_0.

Passo 4. Use as observações da amostra para calcular o valor da estatística do teste.

Passo 5. Se o valor da estatística calculado com os dados da amostra não pertencer à região crítica, não rejeite H_0; caso contrário, rejeite H_0.

Procuraremos, sempre que fizermos teste de hipóteses, distinguir bem esses cinco passos. Vamos seguir fielmente estes passos:

Passo 1. Fixe qual a hipótese H_0 a ser testada e qual a hipótese alternativa H_1.

$$H_0 : p = \frac{1}{4} \quad (\quad \text{não modificam o programa}),$$

versus

$$H_1 : p < \frac{1}{4} \quad (\quad \text{modificam o programa}).$$

Passo 2. Use a teoria estatística e as informações disponíveis para decidir qual estatística (estimador) será usada para testar a hipótese V. Obter as propriedades dessa estatística (distribuição, média, desvio padrão).

Nossa população é descrita por uma variável aleatória discreta com distribuição de Bernoulli de parâmetro p. Seja $(X_1, X_2, \ldots, X_n)$ a nossa amostra aleatória e estatística

$$S_n = \sum_{i=1}^{n} X_i \sim Bin(n, p),$$

com

$$E(S_n) = np \qquad Var(S_n) = npq = E(S_n)q < E(S_n)$$

Se o tamanho da amostra for grande (n=400), no nosso caso, podemos aproximar S_n por uma variável aleatória contínua Y da seguinte maneira:

$$Y \sim N(\mu = np, \sigma^2 = npq),$$

de forma que

$$Z = \frac{Y - np}{\sqrt{npq}} \sim N(0, 1).$$

Poderíamos aproximar $\hat{p}$ por uma variável aleatória contínua W com distribuição normal:

$$W \sim N\left(p, \frac{p(1-p)}{n}\right),$$

de forma que

$$Z = \frac{\sqrt{n}(W - p)}{\sqrt{pq}} \sim N(0, 1).$$

Passo 3. Fixe a probabilidade α de cometer o erro de tipo I e use este valor para construir a região crítica (regra de decisão). Lembre que essa região é construída para a estatística definida no passo 2, usando os valores do parâmetro hipotetizados por H_0.

Vamos fixar $\alpha = 0,05$. Nossa regra de decisão será:

Se $X \leq c$ vamos rejeitar H_0. Caso contrário $(X > c)$ não rejeitar.

Como achar c?

$$\alpha = P(rejeitar H_0 \,\big|H_0 \text{ é verdade}) = P(S_n \leq c|p = 0,25) = 0,05.$$

Assim c é o quinto percentil da binomial com $n = 400$ e $p = 0,25$. Vamos usar o **R** para achar

c.

```
>
> ####Questão 13 do Capítulo 12-Bussab-Morettin
>
> n=400;p=0.25;s=80
> mu=n*p;mu
[1] 100
> sigma2=n*p*(1-p);sigma2
[1] 75
> sigma=sqrt(sigma2);sigma
[1] 8.660254
> alfa=0.05
> c=qbinom(alfa,n,p);c
[1] 86
>
> pbinom(86,n,p)###Um pouco maior na realidade  alfa=0,0577.
[1] 0.05775648
> pbinom(85,n,p) ####menor do que alfa.
[1] 0.04519392
>
>
> #####Quinto percentil de Y
>
> c_a=qnorm(0.05,mu,sigma);c_a
[1] 85.75515
>
>
>
> #####Quinto percentil de Z ~N(0,1)
> z_tab=qnorm(0.05);z_tab
[1] -1.644854
>
> ####Vamos padronizar c:  (c-mu)/sigma=z_cal
>
```

```
> c1=mu+z_tab*sigma;c1
[1] 85.75515


> z_cal=(s-mu)/sigma;z_cal;round(z_cal,2)
[1] -2.309401
[1] -2.31
>
>


>
> #####Quinto percentil de  W
>
> p_est=s/n;p_est
[1] 0.2
>
> sigma2_W=p*(1-p)/400;sigma2_W
[1] 0.00046875
> sigma_W=sqrt(sigma2_W);sigma_W
[1] 0.02165064
> c_W=qnorm(0.05,p,sigma_W);c_W;round(c_W,2)
[1] 0.2143879
[1] 0.21
>
>
```

Nossa regra de decisão será:

$$RC = \{s \in \mathbb{N}, s \leq 86\},$$

$$RA = \{s \in \mathbb{N}, 87 \leq s \leq 400\}.$$

Usando a normal padrão:

Nossa regra de decisão será:

$$RC = \{z \in \mathbb{R}, z \leq -1,65\},$$

$$RA = \{z \in \mathbb{R}, z > -1,65\},$$

Nossa regra de decisão será:

Seja $f = \hat{p}$

$$RC = \{f \in \mathbb{R}, 0 \leq f \leq 0,21\},$$

$$RA = \{f \in \mathbb{R}, 0,21 \leq f \leq 1\}.$$

Passo 4. Use as observações da amostra para calcular o valor da estatística do teste.

Na amostra temos que $s = 80$,

$$\hat{p} = \frac{s}{n} = \frac{80}{400} = \frac{1}{5} = 0.20,$$

e

$$z_{cal} = \frac{80 - 100}{\sqrt{75}} = -2,31.$$

Chegamos a etapa final do nosso teste; Vamos decidir:

Passo 5. Se o valor da estatística calculado com os dados da amostra não pertencer à região crítica, não rejeite H_0; caso contrário, rejeite H_0.

Como $s = 80$ pertence à região crítica recomendamos a modificação do programa. Como $z_{cal} = -2,31$ pertence à região crítica recomendamos a modificação do programa. Como $\hat{p} = f = 0,20$ pertence à região crítica recomendamos a modificação do programa.

Calcule o nível descritivo do teste:

$$nd = \hat{\alpha} = P\left(S \geq 80 \middle| p = 0,25\right) = \sum_{s=0}^{80} \binom{400}{s} \, 0,25^{s} \, 0,75^{400-s} = 0,01085.$$

Calculamos para as demais opções.

```
>
> nd=pbinom(80,400,0.25);nd;round(nd,5)
[1] 0.0108469
[1] 0.01085
>
>
> nd1=pnorm(z_cal);nd1;round(nd1,5)
[1] 0.01046067
[1] 0.01046
> z_2=(p_est-p)/( sqrt(p*(1-p)/n));z_2
[1] -2.309401
>
> nd2=pnorm(z_2);nd2;round(nd2,5)
[1] 0.01046067
[1] 0.01046
>
```

Faça o teste usando **binom.test**:

```
> binom.test(80,400,p=0.25,"less")

Exact binomial test

data:  80 and 400
number of successes = 80, number of trials = 400, p-value = 0.01085
alternative hypothesis: true probability of success is less than 0.25
95 percent confidence interval:
0.0000000 0.2357625
sample estimates:
probability of success
0.2

>
```

Analise cuidadosamente a saída.

Faça o teste usando **prop.test** com e sem fator de correção.

```
>
> ###Com Correção de continuidade:
>
> prop.test(80,400,p=0.25,"less")

1-sample proportions test with continuity correction

data:  80 out of 400, null probability 0.25
X-squared = 5.07, df = 1, p-value = 0.01217
alternative hypothesis: true p is less than 0.25
95 percent confidence interval:
0.0000000 0.2361813
sample estimates:
p
0.2

>
>
> ### Sem Correção de continuidade:
>
> prop.test(80,400,p=0.25,"less",correct=F)

1-sample proportions test without continuity correction

data:  80 out of 400, null probability 0.25
X-squared = 5.3333, df = 1, p-value = 0.01046
alternative hypothesis: true p is less than 0.25
```

```
95 percent confidence interval:
0.0000000 0.2348638
sample estimates:
p
0.2

>
```

Analise cuidadosamente a saída.

Construa um intercalo de confiança para p com $\gamma = 1 - \alpha = 95\%$ de confiança.

Da tabela da normal padrão temos que $z_{tab} = 1,96$ pois

$$P(-1,96 < Z < 1,96) = 0.95.$$

O $IC(p, 95\%)$ é dado por:

$$\hat{p} \pm z_{tab}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0,2 \pm 1,96 \times \sqrt{\frac{0,16}{400}}.$$

$$IC(p, 95\%) \;=\; 0,2 \pm 1,96 \times 0,02 = 0,2 \pm 0,0392$$

$$IC(p, 95\%) \;= (0,1608; 0,2392)\,.$$

Fazendo no **R** temos:

```
>
> gama=0.95;gama
[1] 0.95
> alfa=1-gama
>
> z_tab=qnorm(1-alfa/2);z_tab;round(z_tab,2)
[1] 1.959964
[1] 1.96
>
>
> s=80;n=400
>
> p_est=s/n
>
> epm_est=p_est*(1-p_est)/(n-1);epm_est
[1] 0.0004010025
>
> sqrt(epm_est)
[1] 0.02002505
```

```
>
> ICp95=p_est +c(-1,1)*z_tab*sqrt(epm_est)
>
> ICp95;round(ICp95,4)
[1] 0.1607516 0.2392484
[1] 0.1608 0.2392
>
>
> prop.test(80,400,p=0.25,correct=F)

1-sample proportions test without continuity correction

data:  80 out of 400, null probability 0.25
X-squared = 5.3333, df = 1, p-value = 0.02092
alternative hypothesis: true p is not equal to 0.25
95 percent confidence interval:
0.1637371 0.2419703
sample estimates:
p
0.2

>
> ####Note que o intervalo dado pelo $R$ é diferente do calculado.
>
> ####Vamos tentar explicar esta diferença?
>
>
>
```