CC0288 - Inferência Estatística I

Oitava Verificação de Aprendizagem - 06/07/2023.

Prof. Maurício

aluno:________________________________

A prova tem 5 questões valendo 15 pontos.

1. (Valor 4 pontos) Considere a variável aleatória $X \sim N(\mu, 9)$.

   Seja $X_1, X_2, \ldots, X_9$ uma amostra aleatória da variável aleatória forneceu : $\sum_{i=1}^{9} x_i = 135.$

   a. Queremos testar: $H_0 : \mu = 12 \quad vs \quad H_1 : \mu > 12.$

   Classifique as hipóteses. Calcule o nível descritivo.
   Qual a sua conclusão usando $\alpha = 0,05$?

   Solução:
   $H_0 : \mu = 12(\ Simples) \quad vs \quad H_1 : \mu > 12(\ Composta).$
   A média amostral é dada por:

$$\bar{x} = \frac{135}{9} = 15.$$

Temos que

$$\bar{X} \sim N(\mu, 1).$$

O nível descritivo do teste é dado por:

$$z_{cal} = \frac{15 - 12}{1} = 3.$$

$$\hat{\alpha} = P\left(\bar{X} \geq 15 | \mu = 12\right) = P(Z \geq 3)$$

$$\hat{\alpha} = 0,5 - P(0 < Z < 3) = 0,5 - 0,4987 = 0,0013.$$

```
>
> SX=135;n=9
>
> Xb=SX/9;Xb
[1] 15
> sigma=3
>
> mu_0=12
> z_cal=sqrt(n)*(Xb-mu_0)/sigma;z_cal
[1] 3
>
> nd=1-pnorm(3);nd;round(nd,4)
```

```
[1] 0.001349898
[1] 0.0013
> alfa=0.05
>
> nd> alfa #####Rejeitar H_0.
[1] FALSE
>
> z_tab=qnorm(1-alfa);z_tab
[1] 1.644854
>
> z_tab > z_cal #####Rejeitar H_0.
[1] FALSE
>
>
```

b. Queremos testar:

$$H_0 : \mu = 12 \ \ (Simples) \quad vs \quad H_1 : \mu \neq 12 \ (\ Composta).$$

Classifique as hipóteses. Calcule o nível descritivo.

Forneça o valor tabelado do teste.

Qual a sua conclusão usando $\alpha = 0,05$?
Como

$$p_1 = P(\bar{X} \geq 15 \ |\mu = 12) = 0,0013 < 0,5.$$

Assim

$$nd = 2 \times 0,0013 = 0,0026.$$

2. (Valor 2 pontos) Considere os mesmos dados da questão anterior.

   Além disso sabemos $\sum_{i=1}^{9} x_i^2 = 2212,305.$

   Queremos testar:

   $$H_0 : \mu = 18 \quad vs \quad H_1 : \mu < 18.$$

   Classifique as hipóteses. Calcule o nível descritivo.

   Forneça o valor tabelado do teste.

   Qual a sua conclusão usando $\alpha = 0,04$?

   Como

   $$P(t(8) \leq t_{tab}) = 0,04 = \frac{p}{2}$$

   temos que $p = 0,08$ que não tem na tabela.

   mas

   $$0,05 < 0,08 < 0,10$$

   $$0,025 < 0,04 < 0,05.$$

   Pela tabela temos :

   $$P(t(8) \leq -1,86) = 0,05 \quad e \quad P(t(8) \leq -2,306) = 0,025.$$

   Assim 4

   $$-2,306 < t_{tab} < -1,86.$$

   Note que o valor é $t_{tab} = -2,004$.

```
>
> ######Segunda Questão
>
> n=9
> alfa=0.04
>
> ttab=qt(alfa,n-1);ttab
[1] -2.004152
>
> round(ttab,3)
[1] -2.004
>

> SX=135
> Xb=SX/n;Xb
```

```
[1] 15
>
> S2=(2212.305-135^ 2/n)/(n-1);S2
[1] 23.41312
> s=sqrt(S2);S
[1] 0.1304735
>
> mu_0=18
> tcal=sqrt(n)*(Xb-mu_0)/s;tcal;round(tcal,3)
[1] -1.859999
[1] -1.86
>
> nd=pt(t_cal,n-1);nd;round(nd,2)
[1] 0.04996535
[1] 0.05
>
> nd>alfa
[1] TRUE
> tcal < ttab #####Não rejeitar H_0.
[1] FALSE
>
>
>
```

3. (Valor 3 pontos) Uma fábrica de automóveis anuncia que que seus carros consomem em média, 10 litros de gasolina por 100 quilômetros, com desvio padrão de 0,8 litros. Uma revista desvia que este consumo é maior e resolve testar esta afirmação.

   Para tal, analisa 36 automóveis dessa marca , obtendo um consumo médio de 10,4 litros por 100 quilômetros. Considerando que o consumo siga o modelo Normal, o que a revista pode concluir sobre o anúncio da fábrica ao nível de 1%? Qual o tamanho do erro do tipo II se a média for 10,6?

Solução:

Considere a variável aleatória X consumo de gasolina, em litros, em um percurso de 100 km.

$$X \sim N(\mu; 0,64).$$

Queremos testar:

$$H_0 : \mu = 10 \ \ (Simples) \quad vs \quad H_1 : \mu > 10 \ (\ Composta).$$

dados amostrais:

$$n = 36 \ \ ; \ \ \bar{x} = 10,4.$$

Sabemos que:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Como $\alpha = 0,01$ temos $z_{tab} = 2,33$.
A nossa regra de decisão fica:
Se $z_{cal} > z_{tab}$ rejeitar H_0. Caso contrário não rejeitar.
Se H_0 é verdade temos:

$$z_{cal} = \frac{10,4 - 10}{0,8/6} = 3.$$

Como $3 > 2,33$ não podemos aceitar H_0. A fábrica mentiu.
Vamos calcular o tamanho do erro do tipo II. A região de aceitação

$$z_{cal} \leq 2,33$$

é equivalente a:

$$\frac{\bar{X} - \mu_0}{0,8/6} \leq 2,33$$

$$\bar{X} \leq \mu_0 + 2,33 \times \frac{0,8}{6} = 10,31.$$

Logo

$$\beta = P\left(\bar{X} \leq 10,31 \mid \mu = 10,6\right) = P(Z \leq -2,18)$$

$$\beta = P(Z \geq 2,18) = 0,5 - P(0 < Z < 2,18) = 0,5 - 0,48537 = 0,01463.$$

```
> #####Terceira questão
>
> ######H_0: mu=10 vs H_1: mu >10
> n=36;xb=10.4; mu_0=10;sigma=0.8; alfa=0.01
> z_tab=qnorm(1-alfa);z_tab
[1] 2.326348
> z_tab=2.33
>
> z_cal=(xb-mu_0)/(sigma/sqrt(n));z_cal
[1] 3
>
> ###Regra de decisão: Se z_cal >3 rejeitar H_0. caso contrario não.
>
>
> z_cal>ztab ####### Não rejeitar H_0. A fábrica não mentiu.
[1] TRUE
>
>
> #####item b   Rc: Xb >c
>
> c=mu_0+z_tab*(sigma/sqrt(n));c
[1] 10.31067
>
> c=10.31
>
>
> ####Aceitar H_0 quando Xb <=10,31.
>
> ###Beta(mu)=P(Xb<=10,31)=P(Z<=(10,31-mu)/(0.8/6))
>
>
> mu_1=10.6
> z_1=(c-mu_1)/(sigma/sqrt(n));z_1
[1] -2.175
> beta=pnorm(z_1);beta;round(beta,4)
[1] 0.01481506
[1] 0.0148
>
>
```

4. (Valor 4 pontos) Uma empresa não pode produzir mais de 5 % de unidades defeituosas de um artigo em um mesmo lote. Seja p a proporção de unidades defeituosas em um certo lote e suponha que neste lote 100 artigos são sorteados para serem inspecionados. Responda as seguintes questões:

   Seja $X = 1$ se o artigo é defeituoso e $X = 0$ caso contrário.

   Note que:

$$X \sim Ber(p).$$

   a. Qual o parâmetro que se deseja testar?
      Resposta: O parâmetro p é a probabilidade p do artigo ser defeituoso.

   b. Qual é o estimador a ser utilizado e sua distribuição amostral?
      Resposta: Seja

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

      O estimador não viciado para p é
      a

$$\hat{p} = \frac{S}{n}.$$

      Uma distribuição aproximada para $\hat{p}$ é aproximadamente:

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

      de sorte que aproximadamente:

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

   c. Indique as hipóteses a serem testadas, classifique-as e interprete-as.
      Resposta:
      A hipótese nula

$$H_0 : p = 0,05,$$

      que é uma hipótese simples. Ela indica que a empresa fabrica dentro dos padrões especificados.
      A hipótese alternativa

$$H_1 : p > 0,05,$$

      que é uma hipótese composta . Ela indica que a empresa não fabrica dentro dos padrões especificados.

   d. Determine o critério de decisão com nível de significância de 5%.
      Resposta:
      Se H_0 é verdade temos:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1).$$

Devemos rejeitar H_0 se

$$z_{cal} > 1,645.$$

que é equivalente a :

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > 1,645$$

Logo

$$\hat{p} > p_0 + 1,645 \times \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$\hat{p} > 0,05 + 1,645 \times \sqrt{\frac{0,05 \times 0,95}{100}}$$

$$\hat{p} > 0,0859$$

ou

$$S > 8,59 \quad \text{ou} \quad S \geq 9.$$

```
> #### Quarta questão  H_0: p=0,05 vs H_1: p >0,05
>
> p_0=0.05
> n=100
> aux=p_0*(1-p_0);aux
[1] 0.0475
>
> aux1=aux/n;aux1
[1] 0.000475
> sqrt(aux1)
[1] 0.02179449
>
> z_tab=1.645
>
> c=p_0+ z_tab*sqrt(aux1);c
[1] 0.08585194
>
> round(c,4)
[1] 0.0859
>
```

e. Com o critério, calcule a probabilidade de aceitar um lote com 7% de defeituosos.
Resposta: Esta questão pede o tamanho do erro do tipo II quando $p = 0,07 = p_1$.
Assim,

$$\beta = P\left(\hat{p} \leq 0,0859 \mid p = 0,07\right)$$

$$\beta = P\left(Z \leq \frac{0,0859 - 0,07}{\sqrt{\frac{0,07\times 0,93}{100}}}\right) = P(Z \leq 0,62) = 0,73224.$$

```
>
> p_1=0.07
>
> aux2=p_1*(1-p_1);aux2
[1] 0.0651
>
> aux3=aux2/n;aux3
[1] 0.000651
> sqrt(aux3)
[1] 0.0255147
>
> z=(0.0859-0.07)/sqrt(aux3);z
[1] 0.6231701
> beta=pnorm(z);beta
[1] 0.7334136
>
>
```

f. Se forem observadas 10 unidades defeituosas, qual o nível descritivo?
Resposta:
Note que:

$$E(S) = 100 \times 0,05 = 5,$$

$$Var(s) = np \times (1-p) = E(S) \times q = 5 \times 0,95 = 4,75.$$

$$nd = P\left(S \geq 10\right) \approx P\left(Z \geq \frac{10-5}{\sqrt{4,75}}\right)$$

$$nd = P(Z \geq 2,29) = 0,5 - 0,48899 = 0,01101.$$

```
> ES=100*0.05;ES
[1] 5
>
>
> VS=100*0.05*0.95;VS
[1] 4.75
>
> sigmaS=sqrt(VS);sigmaS
[1] 2.179449
>
> z=(10-5)/sqrt(4.75);z
```

```
[1] 2.294157
> nda=1-pnorm(z);nda

[1] 0.01089073
```

Note que ele coincide com a solução II. Se elevarmos $2,29$ ao quadrado teremos o valor do qui-quadrado da saída.

O nível descritivo exato é dado por:

$$nde = 0,02819.$$

```
Exact binomial test

data:  10 and 100
number of successes = 10, number of trials = 100, p-value = 0.02819
alternative hypothesis: true probability of success is greater than 0.05
95 percent confidence interval:
0.05526324 1.00000000
sample estimates:
probability of success
0.1
```

```
> 1-pbinom(9,100,0.05)
[1] 0.02818829
>
```

```
> prop.test(10,100,0.05,correct=F)

1-sample proportions test without continuity correction

data:  10 out of 100, null probability 0.05
X-squared = 5.2632, df = 1, p-value = 0.02178
alternative hypothesis: true p is not equal to 0.05
95 percent confidence interval:
0.05522914 0.17436566
sample estimates:
p
0.1

>
> prop.test(10,100,0.05,alternative="greater",correct=F)

1-sample proportions test without continuity correction
```

```
data:  10 out of 100, null probability 0.05
X-squared = 5.2632, df = 1, p-value = 0.01089
alternative hypothesis: true p is greater than 0.05
95 percent confidence interval:
0.06071867 1.00000000
sample estimates:
p
0.1

> prop.test(10,100,0.05,alternative="greater")

1-sample proportions test with continuity correction

data:  10 out of 100, null probability 0.05
X-squared = 4.2632, df = 1, p-value = 0.01947
alternative hypothesis: true p is greater than 0.05
95 percent confidence interval:
0.05689751 1.00000000
sample estimates:
p
0.1

>
```

5. (Valor 2 pontos) Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da variável aleatória

$$X \sim N(\mu, 4).$$

Queremos testar:

$$H_0 : \mu = 10 \quad vs \quad H_1 : \mu > 10.$$

Mostre que a região crítica do teste mais poderoso de tamanho $\alpha = 0,05$ é da forma:

$$\bar{x} \geq c.$$

Mostre que

$$c = 10 + \frac{3,29}{\sqrt{n}}.$$

Para $n = 25$

temos

$$c = 10,658.$$

Resposta: Vamos começar pelo final:

Para $n = 25$ temos

$$c = 10 + \frac{3,29}{\sqrt{n}} = 10 + \frac{3,29}{\sqrt{25}} = 10 + \frac{3,29}{5} = 10 + \frac{6,58}{10} = 10,658.$$

Será que existe teste mais poderoso de tamanho $\alpha = 0,05$ para testar;

$$H_0 : \mu = 10 \quad vs \quad H_1 : \mu = a > 10.$$

$$X \sim N(\mu, 4)$$

Sua f.d.p. é dada por:

$$f(x|\mu) = (8\pi)^{-1/2}\, e^{-\frac{(x-\mu)^2}{8}}\, I_A(x), A = (-\infty, \infty).$$

Note que:

$$L(\mu; \mathbf{x}) = \prod_{i=1}^{n} f(x_i|\mu) = \prod_{i=1}^{n} (8\pi)^{-1/2}\, e^{-\frac{(x_i-\mu)^2}{8}}.$$

$$L(\mu; \mathbf{x}) = (8\pi)^{-n/2} \prod_{i=1}^{n} e^{-\frac{1}{8}\,(x_i-\mu)^2}$$

$$L(\mu;\mathbf{x}) = (8\pi)^{-n/2}\, e^{-\frac{1}{8}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

Se a hipótese nula é verdadeira temos $\mu = 10$:

$$L_0(\mathbf{x}) = (8\pi)^{-n/2}\, e^{-\frac{1}{8}\sum_{i=1}^{n}(x_i-10)^2}.$$

Se a hipótese alternativa é verdadeira temos $\mu = a$:

$$L_1(\mathbf{x}) = (8\pi)^{-n/2}\, e^{-\frac{1}{8}\sum_{i=1}^{n}(x_i-a)^2}.$$

Pelo Lema de Neyman-Pearson , utilizando a razão de verossimilhança simples, temos que o teste mais poderoso será aquele com região crítica dada por

$$A_1^* = \{\mathbf{x}; \frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} \geq k\}.$$

Vamos com calma:

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} = \frac{(8\pi)^{-n/2}\, e^{-\frac{1}{8}\sum_{i=1}^{n}(x_i-a)^2}}{(8\pi)^{-n/2}\, e^{-\frac{1}{8}\sum_{i=1}^{n}(x_i-10)^2}}.$$

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} = \exp\left(-\frac{1}{8}\left(\sum_{i=1}^{n}(x_i-a)^2 - (x_i-10)^2\right)\right)$$

Assim

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} = \exp\left(-\frac{1}{8}\sum_{i=1}^{n}(x_i^2 - 2a\, x_i + a^2 - x_i^2 + 20\, x_i - 100)\right)$$

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} = \exp\left(-\frac{1}{8}\sum_{i=1}^{n}(-2(a-10)x_i + a^2 - 100)\right)$$

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} = \exp\left(\frac{2(a-10)}{8}\sum_{i=1}^{n}\, x_i - \frac{n(a^2-100)}{8}\right)$$

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} = \exp\left(\frac{a-10}{4}\;\sum_{i=1}^{n}\, x_i - \frac{n(a^2-100)}{8}\right)$$

Seja

$$s = x_1 + x_2 + \ldots + x_n$$

De

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} \geq k$$

temos

$$\log\left(exp\left(\frac{s}{50} - 3,4n\right)\right) \geq \log(k)$$

$$\frac{a-10}{4} \sum_{i=1}^{n} x_i - \frac{n(a^2-100)}{8} \geq \log(k)$$

$$\frac{(a-10)}{4} \sum_{i=1}^{n} x_i \geq \frac{n((a-10)(a+10))}{8} + \log(k)$$

Como $a > 10$ temos:

$$\sum_{i=1}^{n} x_i \geq \frac{4}{a-10} \frac{n((a-10)(a+10))}{8} + \frac{4}{a-10} \log(k)$$

Dividindo por n temos:

$$\frac{s}{n} \geq \frac{a+10}{2} + \frac{4}{n(a-10)} \log(k)$$

Logo

$$\bar{x} \geq c.$$

que é a nossa região crítica é da forma do teste M.P. procurado pois só depenseu fo fato que $a > 10$.

Assim é a nossa região UMP para o teste

$$H_0 : \mu = 10 \quad versus \quad H_0 : \mu => 10; \quad .$$