CC0288 - Inferência Estatística I

Sexta Verificação de Aprendizagem - 07/06/2023.

Prof. Maurício

A prova tem 8 questões valendo 12 pontos.

1. (Valor 1 ponto) Seja $X \sim \chi^2(15)$. Responda ao que se pede:

   a. Calcule $p_a = P(X \leq 11,721)$.

   b. Qual a mediana de X?

   Solução:

   Trata do uso da tabela IV do Bussab& Morettin.

   Note que

$$P(X > Med) = 0,50.$$

   Veja a saída do R:

```
####Q1-X ~quiquadrado (15)


> p_a=pchisq(11.721,15);p_a
[1] 0.2999874
> round(p_a,3)
[1] 0.3
> Med=qchisq(0.50,15);Med
[1] 14.33886
> round(Med,3)
[1] 14.339
>
```

2. (Valor 1 ponto) Seja $X \sim t(21)$. Responda ao que se pede:

   a. Calcule $p_a = P(X \leq 0,859)$.

   b. Qual o sexto decil de X?

   Solução:

   Trata do uso da tabela V do Bussab& Morettin.

   Veja a saída do R:

```
>
> ###Q2-X~t(21)
>
>
> p_a=pt(0.859,21);p_a;round(p_a,2)
[1] 0.79998
[1] 0.8
>
> D6=qt(0.60,21);D6;round(D6,3)
[1] 0.2565799
[1] 0.257
>
```

3. (Valor 1 ponto) Seja $X \sim F(15, 15)$. Responda ao que se pede:

   a. Calcule $p_a = P(X \leq 2,40)$.?

   b. Qual o quinto percentil de X?

Solução:

Note que se $X \sim F(15, 15)$ então

$$Y = \frac{1}{X} \sim F(15, 15).$$

Trata do uso da tabela VI do Bussab& Morettin.

Veja a saída do R:

```
> ####Q3-X~F(15,15)
>
> p=pf(2.40,15,15);p;round(p,2)
[1] 0.9497299
[1] 0.95
>
> P5=qf(0.05,15,15);P5
[1] 0.4160691
>
> P95=qf(0.95,15,15);P95
[1] 2.403447
>
> 1/P95
[1] 0.4160691
>
```

4. (Valor 1 ponto) Seja $X \sim N(40, 256)$. Responda ao que se pede:

   a. Calcule $p = P(|X - 40| < 12)$.

   b. Qual o percentil de ordem 5 de X?

Solução:

Sabemos que

$$Z = \frac{X - 40}{5} \sim N(0, 1).$$

$$p = P\left(|X - 40| < 12\right) = P\left(\frac{|X - 40|}{5} < 2,4\right) = P\left(|Z| < 2,4\right)$$

$$p = 2 \times P(0 < Z < 2,4) = 2 \times 0,4918 = 0,9836.$$

Trata do uso da tabela III do Bussab& Morettin.

$$P(X < P_5) = 0,05$$

$$P\left(Z < \frac{P_5 - 40}{5}\right) = 0,05.$$

Sabemos que:

$$P(Z < -1,645) = 0,05$$

$$\frac{P_5 - 40}{5} = -1,645$$

$$P_5 = 40 - 1,645 \times 5 = 31,776.$$

Veja a saída do R:

```
 mu=40;sigma=5
> z=12/sigma;z
[1] 2.4
> p=2*0.4918;p
[1] 0.9836
> pnorm(2.4)
[1] 0.9918025
> pnorm(-2.4)
[1] 0.008197536
> round(pnorm(2.4)-pnorm(-2.4),5)
[1] 0.9836

 1.645*5
```

```
[1] 8.225
> 40-1.645*5
[1] 31.775
> round(qnorm(0.05,40,5),3)####direto no R
[1] 31.776
>
```

5. (Valor 1 ponto) Seja $X \sim N(20, 25)$.
Uma amostra de tamanho 16 é retirada com reposição. Responda ao que se pede:
Calcule:

a. $p_a = P(\bar{X} > 23)$.

b. $p_b = P(12,10167 < S^2 < 43,02667)$

Solução:

Sabemos que

$$\bar{X} \sim N(\mu;,\ \sigma^2/n) = N(20, 25/16)$$

$$Z = \frac{\bar{X} - 20}{5/4} \sim N(0,1).$$

$$p_a = P(\bar{X} > 23) = P(Z > 2,4) = 0,5 - 0,4918 = 0,0082.$$

Sabemos que

$$V = \frac{(n-1)S^2}{\sigma^2} = \frac{15S^2}{25} = \frac{3S^2}{5} \sim \chi^2(15)$$

Assim

$$p_b = P(12,10167 < S^2 < 43,02667)$$

$$p_b = P\left(\frac{3}{5} \times 12,10167 < \frac{3}{5}\,S^2 < \frac{3}{5} \times 43,02667\right)$$

$$p_b = P\left(7,261 < V < 25,816\right)$$

$$p_b = P(V > 7,261) - P(V > 25,816) = 0,95 - 0,04 = 0,91,$$

usando a tabela IV.

```
> n=16
>
> mu=20;sigma=5
>
>
> sigmaxb=5/sqrt(n); sigmaxb
[1] 1.25
>
>
```

```
> z=(23-20)/sigmaxb;z
[1] 2.4
>
> p_a=1-pnorm(z);p_a
[1] 0.008197536
> round(p_a,4)
[1] 0.0082
> ####V=15S^2/25=3s^2/5 ~qui(15)
>
> q_1=7.261;q_2=25.816
>
> pchisq(q_1,15)
[1] 0.05000173
>
> pchisq(q_2,15)
[1] 0.9599983
>
> s12=5*q_1/3;s12
[1] 12.10167
> s22=5*q_2/3;s22
[1] 43.02667
> #####S^2 é gama de parametros
>
> r=(n-1)/2;r
[1] 7.5
>
> lamb=(n-1)/(2*sigma^2);lamb
[1] 0.3
>
> p1=pgamma(s22,r,lamb);p1
[1] 0.9599983
>
>
>
> p2=pgamma(s12,r,lamb);p2
[1] 0.05000173
>
> p1-p2
[1] 0.9099965
> round(p1-p2,4)
[1] 0.91
>
```

6. (Valor 2 pontos) Em um escritório trabalham 80 secretárias. Para determinar o nível de satisfação com as exigências de suas chefias, uma pré-amostra de 30 secretárias revelou que 12 não estavam satisfeitas com estas exigências. Determine o tamanho da amostra necessário para estimar com 95% de confiança e com erro máximo de de 5%, a proporção de secretárias **satisfeitas** com estas exigências?

Solução:

Veja a saída do R:

```
> ####Questão 6
>
>
> ###X=1  se a secretária está satisfeita.
> ##P(X=1)=p  , X~Ber(p)
>
> N=80 ###Tamanho da população:
> n_0=30;s_0=30-12;s_0 ####Amostra piloto
[1] 18
> gama=0.95;alfa=1-gama;alfa
[1] 0.05
> e=0.05
>
> p_0=s_0/n_0;p_0 ###Estimativa de p na amostra piloto
[1] 0.6
> qnorm(1-alfa/2)
[1] 1.959964
> z_tab=1.96
>
> num=z_tab^2*p_0*(1-p_0)*N;num
[1] 73.75872
> den=e^2*(N-1)+ z_tab^2*p_0*(1-p_0);den
[1] 1.119484
>
> n_1=num/den;n_1
[1] 65.88635
>
> n=ceiling(n_1);n
[1] 66
>
```

7. (Valor 2 pontos) A saída do MINITAB a seguir mostra os resultados de um teste de hipótese para a diferença $\mu_1 - \mu_2$ entre duas médias populacionais.

Two Sample T for X vs Y.

Variável	n	Mean	StDev	SE Mean
X	10	39.31	8,71	a
Y	10	29.12	4.79	b

Difference= $\mu(X) - \mu(Y)$

Estimate for difference: c

95% lower bound for difference: d.

T-test of difference =0 vs (>0); T-value 3.25 P-value=0.003 DF=13

a. (0,25) Qual a hipótese nula? E a hipótese alternativa?
b. (0,25) Este é um teste unilateral ou bilateral? Que tipo de teste é este?
c. (0,5) H_0 pode ser rejeitada ao nível de 1%? Como você sabe?
d. (1) Quais os valores de a, b, c e d?

Solução: Suponha que

Pela saída do Minitab temos:

$$X \sim N(\mu_1, \sigma_1^2) \ \ e \ \ X \sim N(\mu_2, \sigma_2^2)$$

independentes com variâncias desconhecidas e distintas. Se as variâncias fossem iguais os graus de liberdade seriam $18 = n + m - 2$ e não 13.

O teste proposto na saída testou:

$$H_0 : \mu_1 - \mu_2 = 0,$$

que é a nossa hipótese nula versus

$$H_0 : \mu_1 - \mu_2 > 0$$

que é a nossa hipótese alternativa. Portanto temos um teste unilateralà direita.

Olhando a saída temos que o nível descritivo vale

$$nd = 0,003 < 0,01 = \alpha,$$

portanto devemos rejeitar H_0.

Temos que a é o erro padrão estimado de $\bar{X}$ e vale:

$$a = \frac{s_1}{\sqrt{n}} = \frac{8,71}{\sqrt{10}} = 2,75$$

Temos que b é o erro padrão estimado de $\bar{Y}$ e vale:

$$b = \frac{s_2}{\sqrt{m}} = \frac{4,79}{\sqrt{10}} = 1,51.$$

A estimativa de $\mu_1 - \mu_2$ é representada por **c** que é dada por:

$$c = \bar{x} - \bar{y} = 39,31 - 29,12 = 10,19.$$

O valor de **d**, $4,62$, aparece na saída do R:

$$P(T(13) \leq 1,771) = 0,95$$

A variância estimada de $\bar{Y} - \bar{X}$ é dada por:

$$S^2 = \frac{s_1^2}{n} + \frac{s_2^2}{m} = \frac{s_1^2 + s_2^2}{10} = 9,8808.$$

O erro padrão estimado é dado por

$$S = \sqrt{S^2} = 3,14$$

O erro de estimação

$$e = t_{tab} * S = 5,67.$$

O limite inferior é dado por

$$l_i = \bar{x} - \bar{y} - e = 10,19 - 5,67 = 4,62.$$

```
 S2=(s1^2+s2^2)/10;S2
[1] 9.88082
>
> S=sqrt(S2);S
[1] 3.143377
>
> t1*S
[1] 5.566921
>  delta_est-t1*S
[1] 4.623079
>
```

```
 xb = 39.31; s1 = 8.71; n=10
> yb = 29.12; s2 = 4.79; m=10
>
>
>
> a = s1 / sqrt(n);a;round(a,2)
[1] 2.754344
[1] 2.75
```

```
>
> b = s2 / sqrt(m);b;round(b,2)
[1] 1.514731
[1] 1.51
> c=xb-yb;c
[1] 10.19
> A =  s1^2/n;B= s2^2/m
> num = (A+B)^2;num
[1] 97.6306
> den = A^2/(n-1) + B^2/(m-1);den
[1] 6.97977
> r1 = num/den;r1
[1] 13.98765
> r=floor(r1);r
[1] 13
> tcal = (xb-yb)/sqrt(A+B);tcal;round(tcal,2)
[1] 3.241736
[1] 3.24
> ####Vamos construir o intervalo de confiança unilateral para mu_1-mu_2
> gama=0.95
> t1=qt(0.95,13);t1
[1] 1.770933
> t1=1.771
> d=delta_est - t1*sqrt(A+B);d
[1] 4.623079
> d=delta_est - t1*sqrt(A+B);d;round(d,2)
[1] 4.623079
[1] 4.62
>
```

8. (Valor 3 pontos) O teor de água (em percentagem) para sete tijolos foi medido logo após a fabricação e, em seguida, os tijolos foram colocados em um forno para secagem. Os resultados foram os seguintes:

Amostra Local	X=Antes da Secagem	Y= Depois da Secagem
1	6,56	3,60
2	5,33	2,00
3	7,12	3,95
4	7,28	4,47
5	5,85	2,12
6	8,19	5,77
7	8,18	3,00

Determine um intervalo de confiança de 90% para a redução na percentagem média do teor de água após a secagem.

Para responder a esta pergunta escolha a solução correta entre as 3 apresentadas. Justifique bem sua escolha . Fale das suposições envolvidas? As populações são dependentes ou independentes? As variâncias são iguais ou não? As populações são normais ou (X, Y) tem distribuição normal bidimensional?

Pode usar os resultados diretos da saída não precisa fazer as contas.

Explique detalhadamente os elementos da saída do R que você escolheu.

```
> X=c(656,533,712,728,585,819,818)/100;X
[1] 6.56 5.33 7.12 7.28 5.85 8.19 8.18
>
> Y=c(360,200,395,447,212,577,300)/100;Y
[1] 3.60 2.00 3.95 4.47 2.12 5.77 3.00
>
> mean(X);var(X);sd(X)
[1] 6.93
[1] 1.195333
[1] 1.093313
> mean(Y);var(Y);sd(Y)
[1] 3.558571
[1] 1.781114
[1] 1.334584
> cov(X,Y);cor(X,Y)
[1] 1.0856
[1] 0.7440111
>
> D=X-Y;D
[1] 2.96 3.33 3.17 2.81 3.73 2.42 5.18
>
> mean(D);var(D);sd(D)
[1] 3.371429
[1] 0.8052476
[1] 0.8973559
>
> ?t.test
> t.test(X,Y)

Welch Two Sample t-test

data:  X and Y
t = 5.1703, df = 11.553, p-value = 0.0002634
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
1.944544 4.798314
sample estimates:
mean of x mean of y
6.930000  3.558571

>
> t.test(X,Y,paired=TRUE)

Paired t-test

data:  X and Y
t = 9.9403, df = 6, p-value = 5.993e-05
alternative hypothesis: true mean difference is not equal to 0
```

```
95 percent confidence interval:
2.541513 4.201344
sample estimates:
mean difference
3.371429

>
> t.test(X,Y,var.equal=TRUE)

Two Sample t-test

data:  X and Y
t = 5.1703, df = 12, p-value = 0.000233
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
1.950671 4.792186
sample estimates:
mean of x mean of y
6.930000  3.558571

>


>
> t.test(X,Y,conf.level=0.90)$conf.int
[1] 2.205475 4.537383
attr(,"conf.level")
[1] 0.9
>
>
> t.test(X,Y,,var.equal=TRUE,conf.level=0.90)$conf.int
[1] 2.209237 4.533621
attr(,"conf.level")
[1] 0.9
>
>
> t.test(X,Y,paired=TRUE,conf.level=0.90)$conf.int
[1] 2.712363 4.030494
attr(,"conf.level")
[1] 0.9
>
```

Solução:

Vemos claramente que as populações são dependentes pois os dados são do tipo antes e depois, isto é, os dados são pareados.

Na saída do R temos que o coeficiente de correlação amostral é 0,74 evidenciando a dependência.

A solução correta é a 2.

A suposição básica é que

$$(X, Y) \sim N_2(\mu_1, \mu_2, \rho, \sigma_1^2, \sigma_2^2).$$

$$D = X - Y \sim N(\mu_D, \sigma_D^2)$$

Seja

$$\mu_D = \mu_1 - \mu_2$$

O intervalo pedido é da forma:

$$\bar{D} \pm t_{tab}\,\frac{s_d}{\sqrt{n}},$$

que é equivalente a testar:

$$H_0 : \mu_D = \mu_1 - \mu_2 = 0,$$

versus

$$H_1 : \mu_D = \mu_1 - \mu_2 \neq 0.$$

Pela saída temos:

$$n = 7 \;\; \bar{d} = 3,37143 \;\; s_d = 0,897.$$

Pela tabela da t temos

$$P(T(6) \leq)1,943 = 0,95$$

O erro padrão estimado :

$$e = \frac{s_d}{\sqrt{n}} = \frac{0,897}{\sqrt{7}} = 0,66.$$

Assim

$$3,371429 \pm 0,6590658$$

$$l_i = 2,712363 \quad .l_s = 4,030495,$$

que é o último intervalo apresentado.

Por isso vamos rejeitar H_0.

```
> t1=qt(0.95,6);t1
[1] 1.94318
>
> n=7
>
> s_d=0.8973559;db=3.371429
>
> e=t1*s_d/sqrt(7);e
[1] 0.6590658
> IC90=db+c(-1,1)*e;IC90
[1] 2.712363 4.030495
>
```

O valor t que aparece é calculado como:

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{\sqrt{n}\,\bar{d}}{s_d}$$

```
> t_cal=sqrt(n)*db/s_d;t_cal
[1] 9.940273
> round(t_cal,4)
[1] 9.9403
>
```

O nível descritivo é calculado como:

$$nd = P(|T(6)| > |t_{cal}|) = 2P(T(6) > 9,9403) =$$

```
 p1=pt(9.9403,6,lower.tail=F);p1
[1] 2.996553e-05
>  nd=2*p1;nd
[1] 5.993106e-05
```