

Trabalho de Métodos de Amostragem

Elen Grochovski GRR20170407

30 de novembro de 2018

Exercícios 1,2,3

Com uma população de tamanho $N = 6$ realizar todas as amostras com tamanho $n = 2, 3, 4$ e estimar média, variância e desvio padrão.

Com base nos teoremas:

TEOREMA 1: Em uma população de tamanho N , selecionamos aleatoriamente, sem reposição, uma amostra de tamanho n , temos que $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ é um estimador não viesado de \bar{Y}

TEOREMA 2: Seja y a variável de interesse e seja N o tamanho da população. Considere que serão selecionados n unidades desta população através do processo de amostragem aleatória simples. A medida de incerteza sobre a média amostral calculada com base na amostra será dada por

$$V[\bar{y}] = E[(\bar{y} - \bar{Y})^2] = \left[\frac{N-n}{N} \right] \frac{S^2}{n}$$

TEOREMA 3: No plano de amostragem aleatória simples, com uma amostra de tamanho n selecionada aleatoriamente de uma população de tamanho N temos que $s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$ é um estimador não viesado de S^2

Intervalo de confiança para a média:

$$\bar{Y} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N-n}{N}} \frac{S}{\sqrt{n}}$$

População

##	Populacao	Valores
## 1	y1	2
## 2	y2	4
## 3	y3	5
## 4	y4	7
## 5	y5	8
## 6	y6	9

Média Populacional

[1] 5.833333

Variância Populacional

[1] 6.966667

Amostra tamanho $n = 2$

##	x1	x2	mean	prob	esp	var	S2
## 1	2	4	3.0	0.067	0.200	0.535	2.0
## 2	2	5	3.5	0.067	0.233	0.363	4.5
## 3	2	7	4.5	0.067	0.300	0.119	12.5
## 4	2	8	5.0	0.067	0.333	0.046	18.0
## 5	2	9	5.5	0.067	0.367	0.007	24.5
## 6	4	5	4.5	0.067	0.300	0.119	0.5
## 7	4	7	5.5	0.067	0.367	0.007	4.5
## 8	4	8	6.0	0.067	0.400	0.002	8.0
## 9	4	9	6.5	0.067	0.433	0.030	12.5
## 10	5	7	6.0	0.067	0.400	0.002	2.0
## 11	5	8	6.5	0.067	0.433	0.030	4.5
## 12	5	9	7.0	0.067	0.467	0.091	8.0
## 13	7	8	7.5	0.067	0.500	0.185	0.5
## 14	7	9	8.0	0.067	0.533	0.313	2.0
## 15	8	9	8.5	0.067	0.567	0.474	0.5

Média das médias das amostra $n = 2$

```
## [1] 5.833333
```

Variância da média da amostra $n = 2$

```
## [1] 2.322222
```

Variância da amostra $n = 2$

```
## [1] 6.966667
```

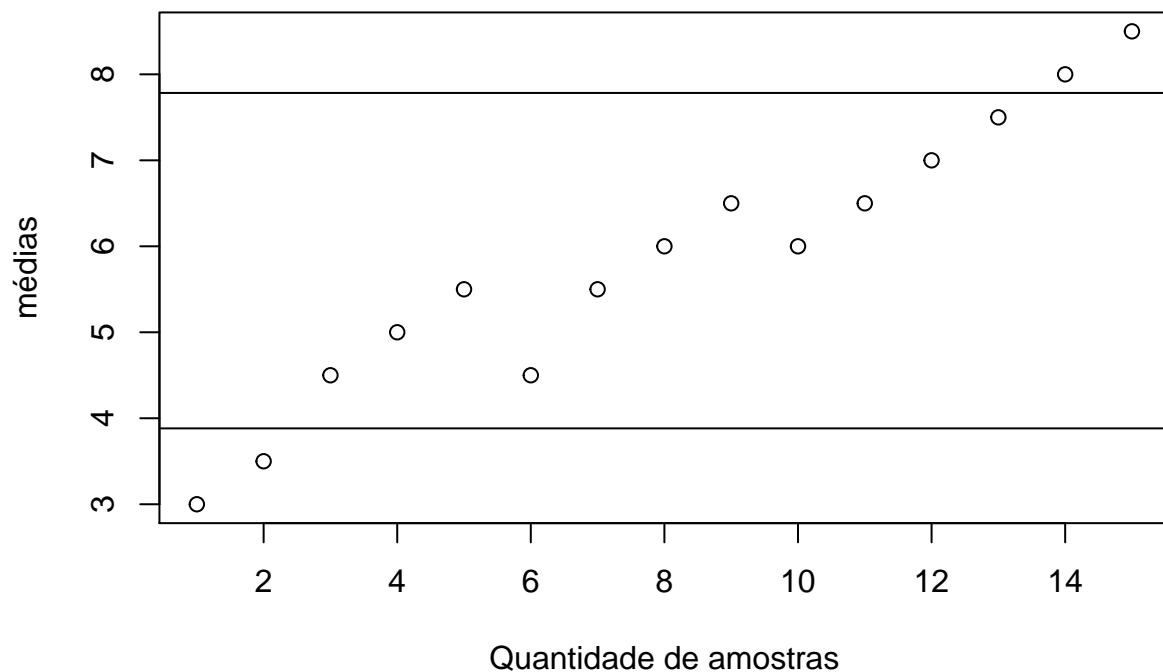
Intervalo de confiança com 80% para as médias da amostra $n = 2$

$$\bar{Y} = 5,833$$

$$IC = 5.833 \pm 1,95$$

$$\bar{Y}_{min} = 3.88$$

$$\bar{Y}_{max} = 7.78$$



Amostra tamanho $n = 3$

##	x1	x2	x3	mean	prob	esp	var	S2
## 1	2	4	5	3.667	0.05	0.183	0.235	4.667
## 2	2	4	7	4.333	0.05	0.217	0.112	12.667
## 3	2	4	8	4.667	0.05	0.233	0.068	18.667
## 4	2	4	9	5.000	0.05	0.250	0.035	26.000
## 5	2	5	7	4.667	0.05	0.233	0.068	12.667
## 6	2	5	8	5.000	0.05	0.250	0.035	18.000
## 7	2	5	9	5.333	0.05	0.267	0.013	24.667
## 8	2	7	8	5.667	0.05	0.283	0.001	20.667
## 9	2	7	9	6.000	0.05	0.300	0.001	26.000
## 10	2	8	9	6.333	0.05	0.317	0.013	28.667
## 11	4	5	7	5.333	0.05	0.267	0.013	4.667
## 12	4	5	8	5.667	0.05	0.283	0.001	8.667
## 13	4	5	9	6.000	0.05	0.300	0.001	14.000
## 14	4	7	8	6.333	0.05	0.317	0.013	8.667
## 15	4	7	9	6.667	0.05	0.333	0.035	12.667
## 16	4	8	9	7.000	0.05	0.350	0.068	14.000
## 17	5	7	8	6.667	0.05	0.333	0.035	4.667
## 18	5	7	9	7.000	0.05	0.350	0.068	8.000
## 19	5	8	9	7.333	0.05	0.367	0.112	8.667
## 20	7	8	9	8.000	0.05	0.400	0.235	2.000

Média das médias das amostra $n = 3$

```
## [1] 5.833333
```

Variância da média da amostra $n = 3$

```
## [1] 1.161111
```

Variância da amostra $n = 3$

```
## [1] 6.966667
```

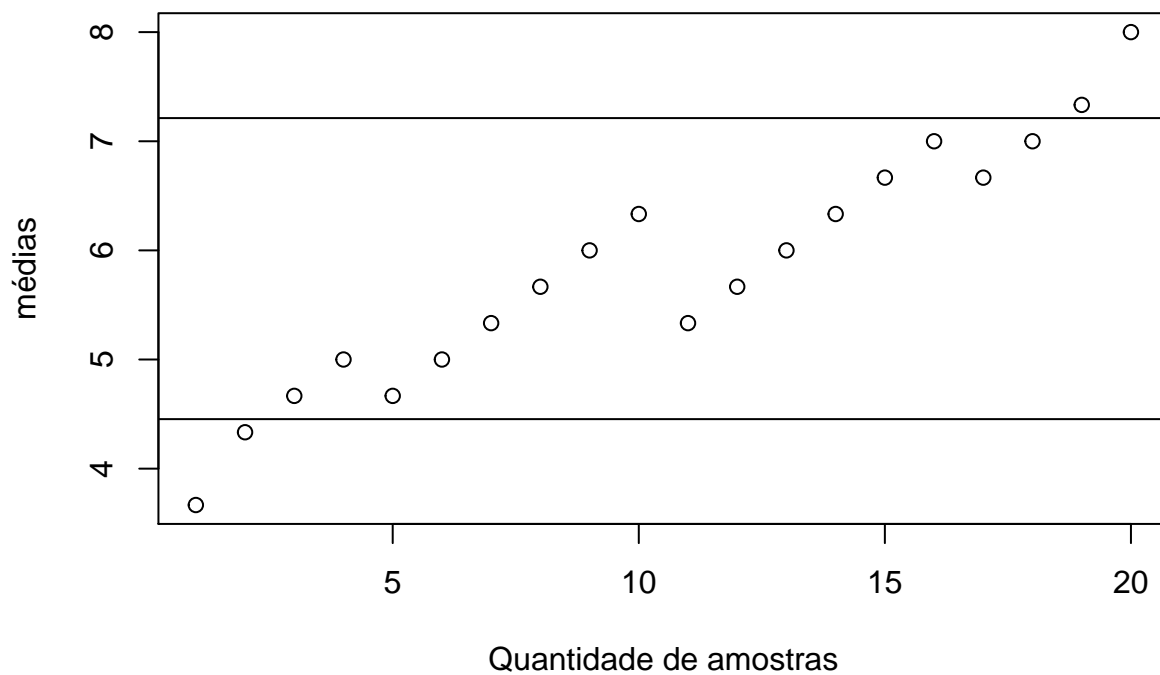
Intervalo de confiança com 80% para as médias da amostra $n = 3$

$\bar{Y} = 5,833$

$IC = 5.833 \pm 1,37$

$\bar{Y}_{min} = 4.45$

$\bar{Y}_{max} = 7.21$



Amostra tamanho $n = 4$

```
##      x1 x2 x3 x4 mean  prob   var   S2
```

```
## 1  2  4  5  7  4.50 0.067 0.119 13.00
## 2  2  4  5  8  4.75 0.067 0.078 18.75
## 3  2  4  5  9  5.00 0.067 0.046 26.00
## 4  2  4  7  8  5.25 0.067 0.023 22.75
## 5  2  4  7  9  5.50 0.067 0.007 29.00
## 6  2  4  8  9  5.75 0.067 0.000 32.75
## 7  2  5  7  8  5.50 0.067 0.007 21.00
## 8  2  5  7  9  5.75 0.067 0.000 26.75
## 9  2  5  8  9  6.00 0.067 0.002 30.00
## 10 2  7  8  9  6.50 0.067 0.030 29.00
## 11 4  5  7  8  6.00 0.067 0.002 10.00
## 12 4  5  7  9  6.25 0.067 0.012 14.75
## 13 4  5  8  9  6.50 0.067 0.030 17.00
## 14 4  7  8  9  7.00 0.067 0.091 14.00
## 15 5  7  8  9  7.25 0.067 0.134  8.75
```

Média das médias das amostra $n = 4$

```
## [1] 5.833333
```

Variância da média da amostra $n = 4$

```
## [1] 0.5805556
```

Variância da amostra $n = 4$

```
## [1] 6.966667
```

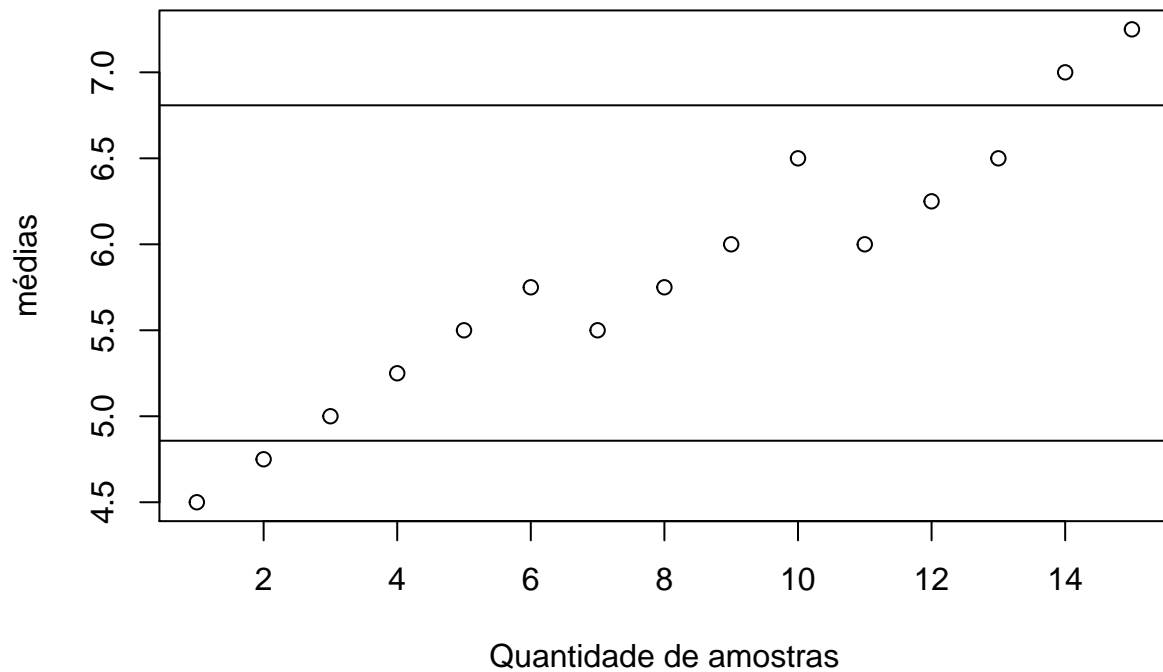
Intervalo de confiança com 80% para as médias da amostra $n = 4$

$\bar{Y} = 5,833$

$IC = 5.833 \pm 0,98$

$\bar{Y}_{min} = 4.86$

$\bar{Y}_{max} = 6.81$



Exercício 4

TABELA ANEXADA AO FINAL

Utilizando a fórmula do RSTUDIO `sample(1:1000, size = 10, replace = F)` gerei 10 números aleatórios dentre os 1000. Os números gerados foram: 410, 686, 28, 491, 927, 658, 721, 883, 931, 399. Retirei da tabela os QI's: 101, 93, 103, 95, 94, 95, 92, 88, 89, 83 e com eles obtive os seguintes resultados:

Média:

```
## [1] 93.3
```

Desvio padrão:

```
## [1] 5.604463
```

Intervalo de confiança com 95% para média

$\bar{y} = 93,30$

$IC = 93,30 \pm 3,45$

$\bar{y}_{min} = 89,85$

$\bar{y}_{max} = 96,75$

Exercício 5

Amostra piloto do Exercício anterior.

fórmula:

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

Onde

$$n_0 = \left(\frac{Z_{1-\frac{\alpha}{2}} S_p}{r \bar{y}_p} \right)^2$$

S_p e \bar{y}_p são estimativas calculadas com base em uma amostra piloto.

$r = 0,03$ é o erro aceitável pelo pesquisador.

O nível de confiança é de $(1 - \alpha) = 95\%$ dado por $z = 1.96$

Com essas informações nós temos os resultados abaixo:

n_0 :

[1] 15.37738

n :

[1] 15

Exercício 6

Considere que uma petição está sendo lançada para determinado pedido de uma certa causa. Na petição, ao todo, existem 676 páginas onde cada uma delas tem um determinado número de assinaturas. Foi selecionada uma amostra de 50 páginas registrando-se o número de assinaturas em cada uma delas. Com base nesta amostra estime o número total de assinaturas e determine o intervalo com 80% de confiança. Na tabela abaixo são apresentados os seguintes resultados obtidos na amostra.

##	Yi	Fr	Total
## 1	42	23	966
## 2	41	4	164
## 3	36	1	36
## 4	32	1	32
## 5	29	1	29
## 6	27	2	54
## 7	23	1	23
## 8	19	1	19
## 9	16	2	32
## 10	15	2	30
## 11	14	1	14
## 12	11	1	11
## 13	10	1	10
## 14	9	1	9
## 15	7	1	7
## 16	6	3	18
## 17	5	2	10
## 18	4	1	4
## 19	3	1	3

A média ponderada é dada por $\frac{\sum_{i=1}^n (y_i f r)}{\sum_{i=1}^n f r}$

[1] 29.42

A variância é dada por $\frac{\sum_{i=1}^n (y_i - \mu)^2}{\sum_{i=1}^n f r}$

[1] 14.98011

Total estimado de assinaturas

[1] 19888

Intervalo de confiança com 80% para média

$$\bar{y} = 29,42$$

$$IC = 29,42 \pm 2,48$$

$$\bar{y}_{min} = 26,94$$

$$\bar{y}_{max} = 31,90$$

Intervalo de confiança com 80% para o total de assinaturas

$$IC = 19.888 \pm 1676$$

$$\overline{assmin} = 18.212$$

$$\overline{assmax} = 21.564$$

Exercício 7

Um antropólogo tem interesse em estudar os habitantes de uma ilha. Seu particular interesse é conhecer o percentual de habitantes que pertencem ao grupo sanguíneo O. O pesquisador aceita como razoável um erro de mais ou menos %. O grau de confiança desejado na pesquisa é de 95%. Sabe-se que ao todo, existem na ilha, 3200 nativos. Calcular o tamanho da amostra.

$$d = 0,05$$

$$\alpha = 0,05$$

$$Z = 1 - \alpha = 95\% = 1,96$$

$$n = \frac{\left(\frac{Z}{d}\right)^2 PQ}{1 + \left(\frac{Z}{d}\right)^2 \frac{PQ}{N}}$$

O resultado dessa conta é

[1] 343

Portanto, precisamos de 343 nativos para ter 95% de confiança que o erro será de $\pm 5\%$

Considere que o pesquisador selecionou uma amostra com o tamanho calculado 343. Destes 343 nativos 145 tinham o grupo sanguíneo O. Estime o número de nativos pertencentes a este grupo sanguíneo e calcule o intervalo com 95% de confiança para esse número.

Estimando a proporção de nativos pertencentes ao tipo sanguíneo O utilizamos $\bar{p} = \frac{a}{n}$

[1] 0.42

Agora estimamos o total de nativos pertencentes ao tipo sanguíneo **O** utilizando $\hat{A} = Np$

[1] 1344

Calculando o intervalo de confiança com a fórmula

$$p \pm \left[Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N-n}{N}} \sqrt{\frac{pq}{n-1}} + \frac{1}{2n} \right]$$

Intervalo de confiança com 95% para a proporção de nativos pertencentes ao grupo sanguíneo **O**

$$\bar{p} = 0,42$$

$$IC = 0,42 \pm 0,05$$

$$\bar{p}_{min} = 0,37$$

$$\bar{p}_{max} = 0,37$$

Intervalo de confiança com 95% para o total de nativos pertencentes ao grupo sanguíneo **O**

$$\hat{A} = 1344$$

$$IC = 1344 \pm 160$$

$$\hat{A}_{min} = 1184$$

$$\hat{A}_{max} = 1504$$

Exercício 8

Considere uma particular região onde existem 1350 fazendas que produzem trigo. Deseja-se estimar o percentual da área total plantada com esta cultura nesta região. Para tanto selecionou-se uma amostra com 20 fazendas registrando-se as áreas total da fazenda e plantada com trigo (ambas em hectares).

##	Fazenda	Area_Total	Area_Plantada
## 1	1	24	12.5
## 2	2	19	7.9
## 3	3	27	11.1
## 4	4	26	10.9
## 5	5	27	11.2
## 6	6	17	7.0
## 7	7	15	7.9
## 8	8	7	3.3
## 9	9	13	4.1
## 10	10	22	8.1
## 11	11	8	3.3
## 12	12	12	5.8
## 13	13	13	4.4
## 14	14	11	3.7
## 15	15	5	2.1
## 16	16	5	1.7
## 17	17	27	12.3
## 18	18	23	9.5
## 19	19	31	12.3
## 20	20	12	4.4

Como estamos estimando uma razão, podemos utilizar a fórmula $\hat{R} = \frac{\bar{y}}{\bar{x}}$ nos dando o resultado

```
## [1] 0.4171512
```

Para o calculo do erro padrão vamos precisar da fórmula

$$\widehat{EP[\hat{R}]} = \sqrt{\frac{(N-n)}{N}} \frac{1}{\sqrt{n\bar{x}}} \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{R}x_i)^2}{N-1}}$$

que nos dá o seguinte resultado

```
## [1] 0.1243528
```