Trabalho de Métodos de Amostragem

Jayme Gomes dos Santos Junior

Curitiba 2018

Sumário

1 Trabalho 1

1.1 Teorema 1

Tendo uma população de tamanho **N**, ao extrair uma amostra de tamanho **n**, temos que o estimador não viezado para $\overline{Y}$ é:

$$\overline{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

1.2 Exercício 1

Tabela 1: População de Tamanho N = 6

População	Valores
y1	2
y2	4
y3	5
y4	7
y5	8
y6	9

Com base na população dada na tabela 1, encontrar a média de todas as combinações possíveis de amostras de tamanhos: n = 2, n = 3 e n = 4.

1.2.1 Resolução

Sabendo que a média populacional é dada por:

$$\overline{Y} = \sum_{i=1}^{N} \frac{y_i}{N} = \frac{2+4+5+7+8+9}{6} = 5.83$$

O objetivo é mostrar que a média de todas as combinações de amostras possíveis terão o mesmo valor da média populacional.

OBS: ys = médias amostrais, P(ys) = probabilidade de cada amostra acontecer e ys.P(ys) = média amostral multiplicada pela sua probabilidade e a média amostral é calculada através do cálculo

$$\sum_{i=1}^{n} ys_i.P(ys_i)$$

1.2.1.1 Para n = 2, temos:

A quantidade de amostras possíveis de tamanho 2 = $\binom{6}{2}$ = 15.

Tabela 2: Amostras Possíveis com n = 2

Amostra	ys	P(ys)	ys.P(ys)
2,4	3	1/15	0.2
2,5	3.5	1/15	0.23
2,7	4.5	1/15	0.3
2,8	5	1/15	0.33
2,9	5.5	1/15	0.37
4,5	4.5	1/15	0.3
4,7	5.5	1/15	0.37
4,8	6	1/15	0.4
4,9	6.5	1/15	0.43
5,7	6	1/15	0.4
5,8	6.5	1/15	0.43
5,9	7	1/15	0.47
7,8	7.5	1/15	0.5
7,9	8	1/15	0.53
8,9	8.5	1/15	0.57

Comparando a média amostral com n = 2: 5.83 e populacional: 5.83, temos que são iguai.

1.2.1.2 Para n = 3, temos:

A quantidade de amostras possíveis de tamanho 2 = $\binom{6}{3}$ = 20.

Tabela 3: Amostras Possíveis com n = 3

Amostra	ys	P(ys)	ys.P(ys)
2,4,5	3.67	1/20	0.18
2,4,7	4.33	1/20	0.22
2,4,8	4.67	1/20	0.23
2,4,9	5	1/20	0.25
2,5,7	4.67	1/20	0.23
2,5,8	5	1/20	0.25
2,5,9	5.33	1/20	0.27
2,7,8	5.67	1/20	0.28
2,7,9	6	1/20	0.3
2,8,9	6.33	1/20	0.32
4,5,7	5.33	1/20	0.27
4,5,8	5.67	1/20	0.28
4,5,9	6	1/20	0.3
4,7,8	6.33	1/20	0.32
4,7,9	6.67	1/20	0.33
4,8,9	7	1/20	0.35
5,7,8	6.67	1/20	0.33
5,7,9	7	1/20	0.35
5,8,9	7.33	1/20	0.37
7,8,9	8	1/20	0.4

Comparando a média amostral com n = 3: 5.83 e populacional: 5.83, temos que são iguai.

1.2.1.3 Para n = 4, temos:

A quantidade de amostras possíveis de tamanho 4 = $\binom{6}{4} = 15$.

Tabela 4: Amostras Possíveis com n = 4

Amostra	ys	P(ys)	ys.P(ys)
2,4,5,7	4.5	1/15	0.3
2,4,5,8	4.75	1/15	0.32
2,4,5,9	5	1/15	0.33
2,4,7,8	5.25	1/15	0.35
2,4,7,9	5.5	1/15	0.37
2,4,8,9	5.75	1/15	0.38
2,5,7,8	5.5	1/15	0.37
2,5,7,9	5.75	1/15	0.38
2,5,8,9	6	1/15	0.4
2,7,8,9	6.5	1/15	0.43
4,5,7,8	6	1/15	0.4
4,5,7,9	6.25	1/15	0.42
4,5,8,9	6.5	1/15	0.43
4,7,8,9	7	1/15	0.47
5,7,8,9	7.25	1/15	0.48

Comparando a média amostral com n = 4: 5.83 e populacional: 5.83, temos que são iguai.

2 Trabalho 2

2.1 Teorema 2

Seja **Y** a variável de interesse e **N** o tamanho da população, considere uma amostra de tamanho **n** extraída pelo processo de amostragem aleatória. A variância da média amostral será dada por:

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

ou

$$V[\overline{y}] = \sum_{i=1}^{n}(\overline{y_i} - \overline{Y})^2 p(ys)_i$$

2.2 Exercício 2

Usando a população do Exerecício 1, calcular a variância da média amostral pelas duas formas para amostras n = 2, n = 3 e n = 4 com o objetivo de mostrar que os resultados são os mesmos usando as duas maneiras para cada tamanho de amostra.

Considerando a tabela 1, temos que:

$$S^2 = \sum_{i=1}^{N}\frac{(y_i - \overline{Y})^2}{N-1} = \frac{(2-5.83)^2}{6-1} + ... + \frac{(9-5.83)^2}{6-1} \cong 6.967$$

2.2.1 Resolução

2.2.1.1 Para n = 2:

1ª Forma:

$$V[\overline{y}] = \left[\frac{N-n}{N}\right]\frac{S^2}{n} = \left[\frac{6-2}{6}\right]\frac{6.967}{2} \cong 2.32$$

2ª Forma:

$$V[\overline{y}] = \sum_{i=1}^{n}(\overline{y_i} - \overline{Y})^2 p(ys)_i = \frac{(3-5.83)^2}{6-1}\frac{1}{15} + \frac{(3.5-5.83)^2}{6-1}\frac{1}{15} + ... + \frac{(8.5-5,83)^2}{6-1}\frac{1}{15} \cong 2.32$$

2.2.1.2 Para n = 3:

1ª Forma:

$$V[\overline{y}] = \left[\frac{N-n}{N}\right]\frac{S^2}{n} = \left[\frac{6-3}{6}\right]\frac{6.967}{3} \cong 1.16$$

2ª Forma:

$$V[\overline{y}] = \sum_{i=1}^{n}(\overline{y_i} - \overline{Y})^2 p(ys)_i = \frac{(3.67-5.83)^2}{6-1}\frac{1}{20} + \frac{(4.33-5.83)^2}{6-1}\frac{1}{20} + ... + \frac{(8-5,83)^2}{6-1}\frac{1}{10} \cong 1.16$$

2.2.1.3 Para n = 4:

1ª Forma:

$$V[\overline{y}] = \left[\frac{N-n}{N}\right]\frac{S^2}{n} = \left[\frac{6-4}{6}\right]\frac{6.967}{4} \cong 0.58$$

2ª Forma:

$$V[\overline{y}] = \sum_{i=1}^{n}(\overline{y_i} - \overline{Y})^2 p(ys)_i = \frac{(4.5-5.83)^2}{6-1}\frac{1}{15} + \frac{(4.75-5.83)^2}{6-1}\frac{1}{15} + ... + \frac{(7.25-5,83)^2}{6-1}\frac{1}{15} \cong 0.58$$

Logo, está demonstrado que as duas formas de cálculo de variância geram o mesmo resultado e que quanto mais aumenta o tamanho da amostra, mais diminui a variãncia, como esperado.

3 Trabalho 3

3.1 Teorema 3

Variânriãncia amostral é um estimador não viezado para a variância populacional.

$$E[s^2] = \sum_{i=1}^{n} s^2 P(ys_i) = S^2$$

Onde:

$$s^2 = \frac{(a_1 - ys)^2 + ... + (a_n - ys)^2}{n-1}$$

Sendo $a_1, ..., a_n$ cada observação da amostra.

3.2 Exercício 3

Provar o teorema 3 utilizando as informações dos exercícios anteriores. Sabemos pelo exercício 2 que: $S^2 = 6,967$.

3.2.1 Resolução

n = 2:

Tabela 5: Variância amostral para n = 2

Amostra	s^2	$s^2P(ys)$
2,4	2	0.13
2,5	4.5	0.3
2,7	12.5	0.83
2,8	18	1.2
2,9	24.5	1.63
4,5	0.5	0.03
4,7	4.5	0.3
4,8	8	0.53
4,9	12.5	0.83
5,7	2	0.13
5,8	4.5	0.3
5,9	8	0.53
7,8	0.5	0.03
7,9	2	0.13
8,9	0.5	0.03

Calculando $E[s^2] = 6.967 = S^2$

n = 3:

Tabela 6: Variância amostral para n = 3

Amostra	s^2	$s^2P(ys)$
2,4,5	2.33	0.12
2,4,7	6.33	0.32
2,4,8	9.33	0.47
2,4,9	13	0.65
2,5,7	6.33	0.32
2,5,8	9	0.45
2,5,9	12.33	0.62
2,7,8	10.33	0.52
2,7,9	13	0.65
2,8,9	14.33	0.72
4,5,7	2.33	0.12
4,5,8	4.33	0.22
4,5,9	7	0.35
4,7,8	4.33	0.22
4,7,9	6.33	0.32
4,8,9	7	0.35
5,7,8	2.33	0.12

Amostra	s^2	$s^2P(ys)$
5,7,9	4	0.2
5,8,9	4.33	0.22
7,8,9	1	0.05

Calculando $E[s^2] = 6.967 = S^2$

n = 4:

Tabela 7: Variância amostral para n = 4

Amostra	s^2	$s^2P(ys)$
2,4,5,7	4.33	0.29
2,4,5,8	6.25	0.42
2,4,5,9	8.67	0.58
2,4,7,8	7.58	0.51
2,4,7,9	9.67	0.64
2,4,8,9	10.92	0.73
2,5,7,8	7	0.47
2,5,7,9	8.92	0.59
2,5,8,9	10	0.67
2,7,8,9	9.67	0.64
4,5,7,8	3.33	0.22
4,5,7,9	4.92	0.33
4,5,8,9	5.67	0.38
4,7,8,9	4.67	0.31
5,7,8,9	2.92	0.19

Calculando $E[s^2] = 6.967 = S^2$

4 Trabalho 4

4.1 Intervalo de Confiança(IC)

O cálculo do IC é feito atravéz da fórmula

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

4.2 Exercício 4

Usando os resultados do Exercício 1, construir o intervalo de confiança considerando n = 2, n = 3 e n = 4 e α - 0.2 e verificar quantas das médias amostrais se encontram dentro do IC.

4.2.1 Resolução

n = 2

$$IC = 5.83 \pm 2.33\sqrt{\frac{6-2}{6}\frac{2.63}{\sqrt{2}}} = 5.83 \pm 3.55 \rightarrow IC = (2.28; 9.38)$$

Com basae nos resultados da Tabela 2 é possível verificar que todas as médias amostrais estão dentro do IC.

n = 3

$$IC = 5.83 \pm 2.33\sqrt{\frac{6-3}{6}\frac{2.63}{\sqrt{3}}} = 5.83 \pm 2.51 \rightarrow IC = (3.32; 8.38)$$

Com basae nos resultados da Tabela 3 é possível verificar que todas as médias amostrais estão dentro do IC.

n = 4

$$IC = 5.83 \pm 2.33\sqrt{\frac{6-4}{6}\frac{2.63}{\sqrt{4}}} = 5.83 \pm 1.78 \rightarrow IC = (4.05; 7.61)$$

Com basae nos resultados da Tabela 4 é possível verificar que todas as médias amostrais estão dentro do IC.

5 Trabalho 5

5.1 Teorema 4

Para calcular o tamanho amostral de uma pesquisa é necessário usar uma amostra piloto, com as informações desta amostra é possível calcular:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \rightarrow n_0 = \left(\frac{Z_{1-\frac{\alpha}{2}} s_p}{r\overline{y}_p}\right)^2$$

OBS: s_p e $\overline{y}_p$ são o desvio padrão e média da amostra piloto

5.2 Exercício

Com base nos dados referentes a QI entregues em sala (Anexo 1), verificar se o tamanho da amostra piloto na folha está adequado e recalcular os parâmetros com o novo tamanho de amostra. Considerar r = 0.03

5.2.1 Resolução

5.2.1.1 Amostra Piloto

As 10 observações aleatórias para a amostra piloto foram colhidas usando o número primo 11 e somando mais 11 para selecionar o próximo, sempre nas linhas da esquerda para a direita, gerando:

Tabela 8: Amostra Piloto

Linha	Coluna	QI
1	11	96
2	2	100
2	13	89
3	4	90

Linha	Coluna	QI
3	15	95
4	6	94
4	17	87
5	8	104
5	19	85
6	10	84

A amostra piloto tem como média amostral $\overline{y}_p = 92.4$ e desvio padrão $s_p = 6.55$.

$$n_0 = \left(\frac{1.96 * 6.95}{0.03 * 92.4}\right)^2 \cong 24$$

$$n = \frac{24}{1 + \frac{24}{1000}} \cong 23$$

Portanto, com base na amostra piloto, o tamanho da amostra deve ser 23.

5.2.1.2 Recalcular Parâmetros

Tabela 9: Dados sorteados após cálculo do n

Linha	Coluna	QI
1	11	96
2	2	100
2	13	89
3	4	90
3	15	95
4	6	94
4	17	87
5	8	104
5	19	85
6	10	84
9	10	89
9	13	89
12	13	77
12	15	97
15	15	91
29	1	99
38	13	87
28	15	83
14	20	100
42	15	80
47	12	90
16	9	88
34	15	88

Ao reaclcular, a média amostral $\overline{y}_p = 90.52$ e desvio padrão $s_p = 6.76$.

6 Trabalho 6

6.1 Exercício 6

Considere que uma petição está sendo lançada para um 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 resultados obtidos na amostra

Tabela 10: Dados do Exercício 6

Nr de Assinaturas	Frequência
42	23
41	4
36	1
32	1
29	1
27	2
23	1
19	1
16	2
15	2
14	1
11	1
10	1
9	1
7	1
6	3
5	2
4	1
3	1

Com média $\overline{y} = 29.64$ e desvio padrão $s = 14.98$

6.1.1 Resolução

$$IC = 29.64 \pm 2.28\sqrt{\frac{676-50}{676}}\frac{14.98}{\sqrt{50}} = 29.64 \pm 2.61 \rightarrow IC = (27.03; 32.25) \cong (27; 32)$$

Portanto, com uma confianç de 80% o intervalo (27;32) contém a verdadeira média de assinaturas por página da petição.

Logo, estima-se que o intervalo (18252;21632) contenha a quantidade total de ssinaturas da petição.

7 Trabalho 7

7.1 Teorema 5

O cálculo do IC para proporção de população finita é dado por:

$$IC = \hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{N-n}{N}} \sqrt{\frac{pq}{n-1}} + \frac{1}{2n}$$

7.2 Exercício 7

Em uma amostragem aleatória simples de tamanho n =100selecionada de uma populaçãode tamanho N = 500, há 37 unidades na classe C. Ache os limites do intervalo com 95% de confiança para proporção e para o número total de unidades da classe C.

7.2.1 Resolução

$$\hat{p} = \frac{37}{100} = 0.37 \;\; e \;\; q = 1 - 0.37 = 0.63 IC = 0.37 \pm 1.96 \sqrt{\frac{500-100}{500}} \sqrt{\frac{0.37 * 0.63}{100-1}} + \frac{1}{2 * 100} = 0.37 \pm 0.09$$

Logo, o IC para a proporção mostra que com 95% de confiança o intervalo (0.28;0.46) contenha a verdadeira proporção e o intervalo (28;46) contenha o número total real de unidades na classe C.

8 Trabalho 8

8.1 Teorema 6

Para calcular a estimação de uma razão, usam-se as seguintes fórmulas:

$$\hat{R} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} \; \hat{V}[\hat{R}] = \frac{(N-n)}{N} \frac{1}{n\overline{x}^2} \left[\frac{\sum_{i=1}^{n} y_i^2 - 2\hat{R} \sum_{i=1}^{n} y_i x_i + \hat{R}^2 \sum_{i=1}^{n} x_i^2}{n-1} \right] \widehat{EP}[\hat{R}] = \sqrt{\hat{V}[\hat{R}]} IC = \hat{R} \pm t_{1-\frac{\alpha}{2}} \widehat{EP}[\hat{R}]$$

8.2 Exercício 8

Considere que uma particular região tem 1350 fazendas produtoras de trigo. Destas fazendas deseja-se conhecer o percentual da área da fazenda com o plantio desta cultura. Para tanto, efetuou-se um levantamento por amostragem aleatória simples de 20 destas fazendas, registrando em cada uma delas a área total da fazenda e a área plantada com trigo.

Tabela 11: Dados das fazendas

Fazenda	Área Total(xi)	Área Plantada(yi)
1	24	12.5
2	19	7.9
3	27	11.1
4	26	10.9
5	27	11.2
6	17	7.0
7	15	7.9
8	7	3.3
9	13	4.1
10	22	8.1
11	8	3.3

Fazenda	Área Total(xi)	Área Plantada(yi)
12	12	5.8
13	13	4.4
14	11	3.7
15	5	2.1
16	5	1.7
17	27	12.3
18	23	3.5
19	31	12.3
20	12	4.4

8.2.1 Resolução

$$\hat{R} = \frac{137.5}{344} = 0.40$$

Portanto, estima-se que 40% da área total das fazendas são plantadas com trigo.

$$\hat{V}[\hat{R}] = \frac{(1350-20)}{1350}\frac{1}{2017.2^2}\left[\frac{\sum_{i=1}^{n} y_i^2 - 2*0.4\sum_{i=1}^{n} y_i x_i + 0.4^2\sum_{i=1}^{n} x_i^2}{20-1}\right] = 0.73\widehat{EP}[\hat{R}] = \sqrt{0.73} = 0.85$$

O erro padrão da estimativa foi de 85%.

$$IC = 0.4 \pm (-2.09)0.85 = 4 \pm (-1.78)$$

Logo, com 95% de confiança o intervalo (38.22%;41.78%) contenha o verdadeiro valor da razão da área plantada com trigo.