

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS – ICE_x
DEPARTAMENTO DE ESTATÍSTICA

**CADERNO DE EXERCÍCIOS PARA
ACOMPANHAMENTO DA DISCIPLINA
“INTRODUÇÃO À BIOESTATÍSTICA”**

Autores (em ordem alfabética)

Ana Bárbara Costa Reis
Ilka Afonso Reis (*coordenadora*)
Lourdes Coral Contreras Montenegro (*coordenadora*)
Priscila Brandão Reis

**Esta apostila é parte integrante produzido pelo projeto “Modernização
do Ensino da Disciplina Introdução à Bioestatística – EST179” sob o
Edital PROGRAD 002/2009.**

Segunda Parte:
Solução dos Exercícios

Seção 1: Tipos de Estudos, Variáveis, População e Amostra

Exercício (1.1)

a) **Estudo do tipo:** Observacional

Grupos de Comparação: Gestantes que nunca consumiam peixe;

Gestantes que consumiam pelo menos 15g de peixe uma vez por semana.

b) **Estudo do tipo:** Experimental

Grupos de Comparação: Gestantes que sofrem de periodontite e fizeram tratamento;

Gestantes que fazem tratamento para periodontite.

c) **Estudo do tipo:** Observacional

Grupos de Comparação: Crianças que dormem roncando;

Crianças que não roncam dormindo.

d) **Estudo do tipo:** Experimental

Grupos de Comparação: Ratas grávidas que receberam alimentação normal;

Ratas grávidas alimentadas com certas vitaminas.

e) **Estudo do tipo:** Observacional

Grupos de Comparação: Homens acima de 50 anos que praticam esportes;

Homens acima de 50 anos que não praticam esportes.

f) **Estudo do tipo:** Experimental

Grupos de Comparação: Mulheres com idade média de 53 anos – com pressão arterial normal – e de 58 anos – hipertensas – que se alimentaram de soja;

Mulheres com idade média de 53 anos – com pressão arterial normal – e de 58 anos – hipertensas – que não se alimentaram de soja.

g) **Estudo do tipo:** Experimental

Grupos de Comparação: Ratas que passaram a maternidade;

Ratas que não passaram pela maternidade.

h) **Estudo do tipo:** Observacional

Grupos de Comparação: Pessoas com peso ideal;

Pessoas acima do peso ideal – obesos.

i) **Estudo do tipo:** Experimental

Grupos de Comparação: Ratos que viviam em ambiente com túneis de plástico;

Ratos que ficaram em um gaiola vazia.

j) **Estudo do tipo:** Observacional

Grupos de Comparação: Filhos de mulheres que tomaram mais de um tipo de antibiótico durante a gravidez;

Filhos de mulheres que tomaram um tipo ou nenhum tipo de antibiótico durante a gravidez.

k) **Estudo do tipo:** Observacional

Grupos de Comparação: Filhos de mulheres que fumaram na gravidez;

Filhos de mulheres que não fumaram na gravidez.

l) **Estudo do tipo:** Observacional

Grupos de Comparação: Pessoas que dormem até 7 horas por dia;

Pessoas que dormem mais de 8 horas por dia.

m) **Estudo do tipo:** Experimental

Grupos de Comparação: Mulheres com dificuldades sexuais tratadas com placebo;

Mulheres com dificuldades sexuais tratadas com Viagra.

n) **Estudo do tipo:** Experimental

Grupos de Comparação: Mulheres que receberam placebo;

Mulheres tratadas com compostos à base isoflavona.

Exercício (1.2)

a) O número de filhos de casais residentes em uma cidade.

Variável do tipo quantitativa e discreta.

b) O grau de satisfação na vida sexual de homens com idade acima de 50 anos.

Variável do tipo qualitativa e ordinal.

c) A idade – em anos completos – de homens com câncer de próstata.

Variável do tipo quantitativa e discreta.

d) O estágio de uma determinada doença em pacientes.

Variável do tipo qualitativa ordinal.

e) A quantidade de leite retirado de uma vaca em um dia.

Variável do tipo quantitativa e contínua.

f) A cor da pele de pacientes com câncer de pele.

Variável do tipo qualitativa e nominal.

Exercício (1.3)

a) População: Homens entre 40 e 64 anos.

Amostra: 1456 homens entre 40 e 64 anos.

b) População: Pacientes com doença coronariana.

Amostra: 348 pacientes com doença coronariana.

c) População: crianças que mamaram no peito por seis meses ou mais e outras crianças que o fizeram por menos de noventa dias.

Amostra: 345 crianças.

d) População: Homens e mulheres hipertensos e totalmente sedentários.

Amostra: 207 homens e mulheres hipertensos e totalmente sedentários.

e) População: Crianças cujas mães tomaram remédios contra infecções urinárias ou respiratórias, entre outras, durante o período de gestação.

Amostra: 24.690 crianças cujas mães tomaram remédios contra infecções urinárias ou respiratórias, entre outras, durante o período de gestação.

f) População: Mulheres americanas.

Amostra: 74 mil americanas.

Seção 2: Análise Descritiva de Dados e Exploratória dos Dados

Exercício (2.1)

Estatística é um conjunto de métodos desenvolvidos para coletar, organizar, descrever, analisar, interpretar e concluir sobre dados.

Exercício (2.2)

Tabela 2: Resposta de pacientes após receberem uma determinada vacina

Resposta	Frequencia		Frequencia	Frequencia
	Absoluta	Relativa(%)	Absoluta Acumulada	Relativa(%) Acumulada
Baixa	6	17.14	6	17.14
moderada	18	51.43	24	68.57
Alta	11	31.43	35	100
Total	35	100	---	---

Tabela 3: Nascimentos em Ontário, Canadá, de 1982 a 1992

Cesariana	Frequencia	
	Absoluta	Relativa(%)
Sim	269864	19.61
Não	1106138	80.39
Total	1376002	100

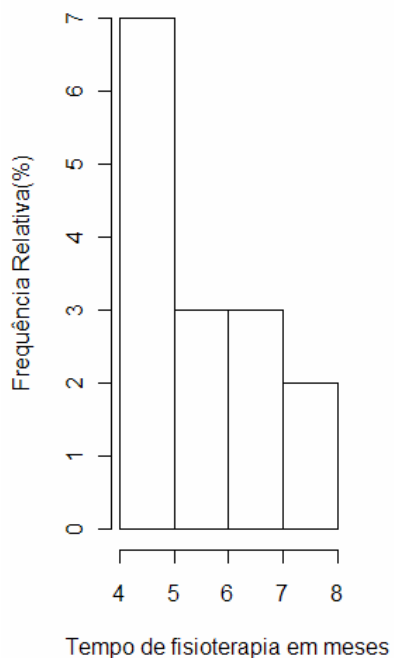
Exercício (2.3)

- a) Fisioterapia: variável do tipo quantitativa e discreta.
- Sequelas: variável do tipo qualitativa e nominal
- Cirurgia: variável do tipo qualitativa e ordinal

b)

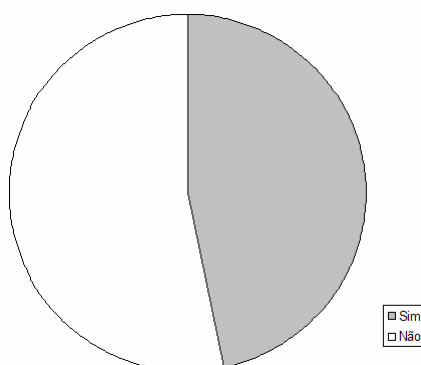
Fisioterapia	Frequência Absoluta	Sequelas	Frequência Absoluta	Cirurgia	Frequência Absoluta
4	2	S	7	A	4
5	5	N	8	M	7
6	3	Total	15	B	4
7	3			Total	15
8	2				
Total	15				

c)

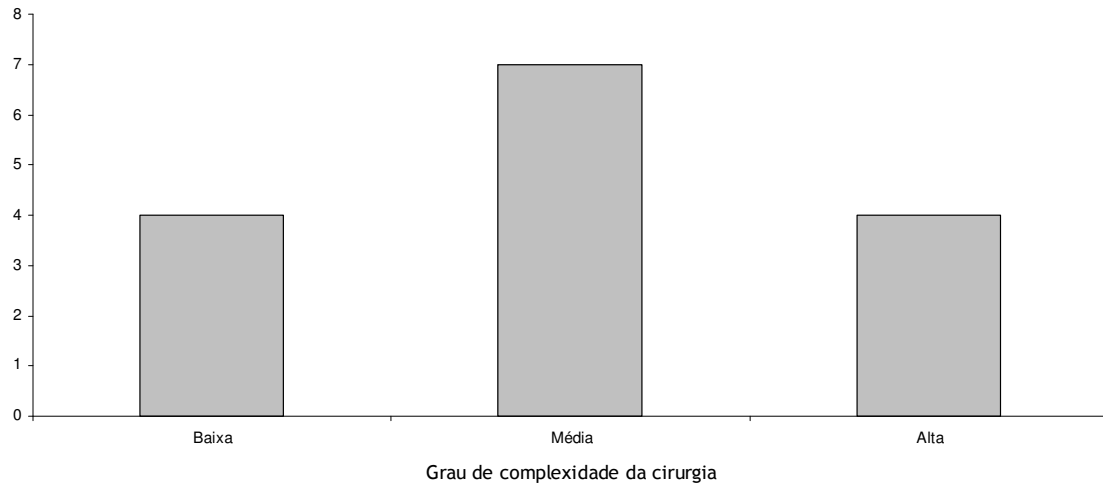


Pelo histograma da variável fisioterapia podemos perceber uma assimetria com concentração à esquerda (cauda à direita) que indica que o tempo de fisioterapia dos pacientes está concentrado em tempos menores. A grande parte dos pacientes se recupera entre 4 e 5 meses.

Presença de sequelas



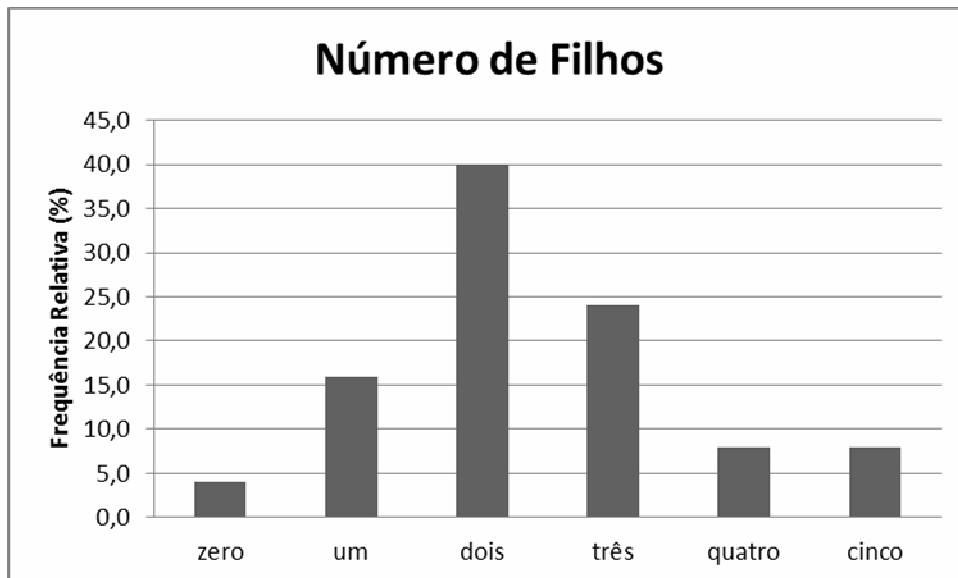
Pelo gráfico de pizza da variável Sequelas percebemos que a quantidade de pacientes sem sequelas é menor.



Pelo gráfico de barras, podemos observar que a maior parte dos pacientes realiza cirurgia de grau médio.

Exercício (2.4)

Analisando a Tabela 2.3, podemos perceber que as famílias mais frequentes são as de dois filhos (40%), seguida pelas famílias de três filhos. Apenas 16% das famílias têm mais de três filhos, mas são ainda mais comuns do que famílias sem filhos.

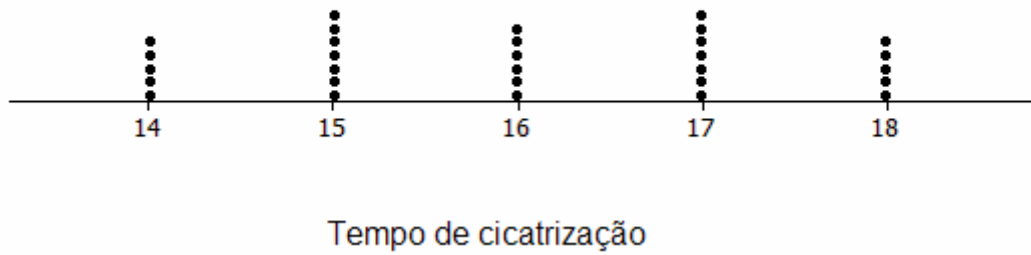


Exercício (2.5)

Os gráficos não são comparáveis. As escalas de peso não são iguais e as frequências são absolutas e não relativas, como deveriam ser para que os gráficos pudessem ser comparados.

Exercício (2.6)

a)



O tempo de cicatrização da maioria dos pacientes está entre 15 e 17 dias.

b)

1 | 4 4 4 4 4

1 | 5 5 5 5 5 5 5

1 | 6 6 6 6 6

1 | 7 7 7 7 7 7 7

1 | 8 8 8 8 8

legenda: 1|8 leia-se 18 dias

c) $P_{22} = 15$, $P_{38} = 15$, $P_{76} = 17$

Exercício (2.7)

- a) Esta afirmativa está errada. Na verdade, essa afirmativa refere-se à mediana, pois metade dos valores então abaixo ou são iguais a mediana. A média é um ponto de equilíbrio da distribuição. É onde se concentram mais dados.
- b) Esta afirmativa refere-se ao fato de que a média é o ponto em que mais se concentram os dados da distribuição, logo é o valor mais típico da distribuição. No entanto, é um dos valores típicos do conjunto de dados, que também pode ser resumido usando-se a mediana ou a moda.
- c) Sempre haverá alunos abaixo da média, quando a palavra “média” se referir à estatística que é usada para resumir um conjunto de valores. Esta afirmativa também pode ser interpretada como se a palavra “média” se referisse não à média de um conjunto de dados, mas sim a uma meta que deve ser cumprida em escolas, em geral, essa meta é igual a 60 pontos. Os alunos devem ter 60 pontos ou mais para passar na matéria cursada.

Exercício (2.8)

$$\bar{X} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i}, \text{ em que } k \text{ é o número de classes}$$

$$\bar{X} = \frac{\sum_{i=1}^5 f_i x_i}{\sum_{i=1}^5 f_i} = \frac{2 \times 1675 + 1825 + 8 \times 1975 + 3 \times 2125 + 2 \times 2275}{2 + 1 + 8 + 3 + 2} = 1993,75 \text{ kJ/100g}$$

$$S^2 = \frac{\sum_{i=1}^k f_i (x_i - \bar{X})^2}{\sum_{i=1}^k f_i} = \frac{\sum_{i=1}^k f_i x_i^2 - n \bar{X}^2}{n - 1} = \frac{64045000 - 16 \times 3975039,06}{16 - 1} = 29625$$

Exercício (2.9)

a)

Média	Mediana	Desvio-Padrão
24,385	14	23,571

b)

Grupos	CV
Cura rápida	0,74
Cura normal	0,36
Cura lenta	0,29

O grupo *cura lenta* é o mais homogêneo enquanto que o grupo *cura rápida* é o mais heterogêneo.

c)

Quartil 25%	Quartil 75%	Intervalo inter-quartil
3,25	44,25	44,25 - 3,25 = 41,00

Exercício (2.10)

Os valores da mediana, do primeiro e terceiro quartis são aproximadamente: 13, 12, 14.5, respectivamente. Isso significa que 50 % das taxas de hemoglobina estão abaixo de 13 g/cm³ e 50% estão acima; 25 % das taxas de hemoglobina estão abaixo de 12 g/cm³ e 75% estão acima; 75% das taxas de hemoglobina estão abaixo de 14.5 g/cm³ e 25% estão acima.

Exercício (2.11)

a)

Riqueza da espécie	Mediana	Média	Desvio-Padrão
Floresta	7.500	9.182	4.305
Pântano	4.000	4.864	2.932

A floresta é o habitat com maior riqueza de espécies, pois sua média e mediana possuem valores maiores. Este habitat também apresenta maior variabilidade, já que, apresenta maior desvio-padrão.

b) CV Floresta = 0.469, CV Pântano = 0.60

O grupo mais homogêneo é a Floresta, pois apresenta menor coeficiente de variação.

Exercício (2.12)

a.1) e a.2)

	Min.	1 ^o Qu.	Mediana	Média	3 ^o Qu.	Max.	Desvio-Padrão	CV
Irmãos	65.0	66.5	70.0	69.0	71.0	73.0	2.720	0.039
Irmãs	59.0	62.5	64.0	64.0	65.0	69.0	2.569	0.040

b) É fácil perceber que a altura dos irmãos é maior, pois estes apresentam valores de média, mediana, mínimo, máximo, 1^o e 3^o quartis maiores.

Exercício (2.13)

a) Tabela 1.3: Estatísticas descritivas do ganho de peso (em gramas) de ratos que com dieta a base de bife e cereal

Ganho de Peso	n	Min.	1 ^o Quartil	Mediana	Media	3 ^o Quartil	Max.	D.P.	CV
Bife	20	51.0	77.5	90.0	89.6	102.5	118.0	17.712	0.198
Cereal	20	56.0	74.0	87.0	84.9	95.5	111.0	14.994	0.177

b) Podemos perceber, pela análise dos box-plots, que existe diferença nas medianas e na variabilidade no ganho de peso dos ratos que ingeriram proteína de bife e de cereal. A mediana é maior para o grupo que comeu bife e a variabilidade deste grupo é maior. Além disso, a distribuição do ganho de peso no grupo cereal é um pouco mais simétrica do que a distribuição do ganho de peso no grupo bife.

c) O ganho de peso médio dos ratos que consumiram bife (89,6 g) foi maior do que o dos ratos que consumiram cereal (84,9 g). Pelo coeficiente de variação, percebemos que o grupo que comeu bife possui uma variabilidade um pouco maior do que a variabilidade do outro grupo.

Exercício (2.14)

a.1 e a.2)

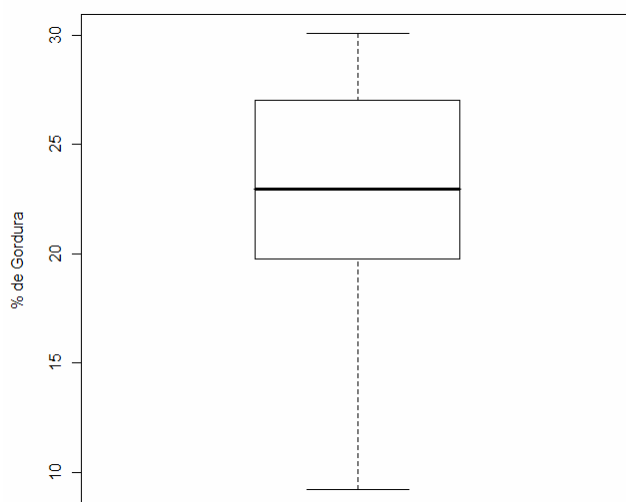
	Média	Moda	1° Quartil	Mediana	3° Quartil
Energia	1989	1970	1928	1980	2088
Proteína	5.875	-	4.375	5.300	7.450
Gordura	22.46	-	19.88	22.95	26.90
Carboidrato	63.69	-	59.30	62.10	67.60
Sódio	139.2	110, 160, 220	105.8	123.0	167.5

Observação: as variáveis Proteína, Gordura e Carboidrato não possuem moda enquanto que a variável Sódio possui três modas.

b)

	Energia	Proteína	Gordura	Carboidrato	Sódio
CV	0.089	0.408	0.266	0.297	0.416

c)



Box-plot do percentual de gordura

O Box-plot apresenta assimetria com concentração à esquerda, o que indica que a maioria dos chocolates apresenta valores maiores para o percentual de gordura.

d) Na figura 2, notamos que as variáveis apresentam medianas bastante diferentes, sendo a da Gordura a menor e a do Sódio a maior. Além disso, percebemos que a variável Sódio apresenta grande variabilidade e seu valor máximo chega a 250 mg. Nota-se também que a variável Gordura apresenta certa assimetria.

Exercício (2.15)

Tabela 1: Estatísticas Descritivas de algumas marcas de chocolate

	Média	Desvio-Padrão	Valor Original		Escore Padronizado	
			Marca A	Marca B	Marca A	Marca B
Energia	1989	177.658	2250	1930	1.47	-0.33
Proteína	5.875	2.400	7.2	3.5	0.55	-0.99
Gordura	22.46	5.967	30.1	24.5	1.28	0.34
Carboidrato	63.69	6.674	59.4	56.4	-0.64	-1.09
Sódio	139.2	57.91	110	40	-0.50	-1.71

O chocolate da marca A se distancia mais do grupo de chocolates quanto ao valor energéticos, ficando a aproximadamente 1.5 desvios-padrão acima da média.

Já o chocolate da marca B destaca-se por apresentar uma quantidade de sódio bem abaixo do valor típico do grupo, com 1.7 desvios-padrão abaixo da média.

Exercício (2.16)

$$Z = \frac{2.09 - 1.70}{0.10} = 3.9$$

Como a altura do atleta está quase 4 desvios acima da média, ele pode ser considerado excepcionalmente alto.

Exercício (2.17)

Por linha:

Resposta	Placebo	Vacina	Total
Baixa	25 (80.6%)	6 (19.4%)	31 (100%)
moderada	8 (30.8%)	18 (69.2%)	26 (100%)
Alta	5 (31.25%)	11 (68.75%)	16 (100%)
Total	38 (47.9%)	35 (52.1%)	73 (100%)

80.6% dos pacientes que tiveram resposta baixa ao tratamento receberam placebo.

30.8% dos pacientes que tiveram resposta moderada ao tratamento receberam placebo.

31.25% dos pacientes que tiveram resposta alta ao tratamento receberam placebo.

A distribuição de pacientes que receberam ou não vacina varia muito dentro das categorias de resposta, principalmente se analisarmos as categorias “moderada” e “alta” em comparação com a categoria “baixa”. Isto é um indício de associação entre as duas variáveis.

Por coluna:

Tabela 2.6: Tratamento de gripe

Resposta	Placebo	Vacina	Total
Baixa	25 (65.78%)	6 (17.2%)	31 (42.5%)
moderada	8 (21.1%)	18 (51.4%)	26 (35.6%)
Alta	5 (13.1%)	11 (31.4%)	16 (21.9%)
Total	38 (100%)	35 (100%)	73 (100%)

A análise desta tabela por coluna é mais interessante, pois vamos analisar como se dá a distribuição dos pacientes nas categorias de resposta ao tratamento dentro de cada tipo de tratamento (vacina ou placebo). Assim, poderemos comparar os dois tratamentos quanto às frequências de cada tipo de resposta ao tratamento. Por exemplo, 13.1% dos pacientes que receberam placebo tiveram resposta “alta” ao tratamento, enquanto 31.4% dos pacientes que receberam vacina tiveram resposta “alta”. Isto também é um indício de que a intensidade de resposta ao tratamento está ligada ao tipo de tratamento (vacina ou não).

Exercício (2.18)

O gráfico de linhas mostra que o número de partos normais é muito maior do que número de cesarianas durante todo o período estudado. O número de partos normais apresenta tendência de crescimento maior do que o número de cesarianas, que, por sua vez, apresenta certa tendência à estabilidade, principalmente no final do período.

Exercício (2.19)

Correlação é a medida da relação entre duas variáveis. Existe correlação quando uma das variáveis está, de alguma forma, relacionada com a outra. Mas atenção: a existência de correlação entre duas variáveis não significa que uma variável é consequência da outra. Somente significa que ambas estão relacionadas de alguma maneira.

Exercício (2.20)

- a) **Falsa.** Duas variáveis podem estar altamente correlacionadas, mas sem nenhuma relação de causa-e-efeito. O exemplo clássico é o gráfico de dispersão que mostra uma alta correlação positiva entre o número de rádios fabricados na Inglaterra e o número de diagnósticos de loucura ao longo do século 20. Ambas as variáveis, na verdade, são influenciadas positivamente pelo desenvolvimento da tecnologia, e não podemos dizer, de maneira nenhuma, algo como “o aumento no número de rádios implica o aumento no número de diagnóstico de loucura”.
- b) **Verdadeira.** Quando agrupamos indivíduos em grupos e resumimos os grupos por meio de médias, a variabilidade dos dados agrupados em médias é menor do que quando usamos os dados originais. Assim, se usarmos estas médias para calcular o coeficiente de correlação, o valor de r tende a aumentar. Isto porque o denominador do coeficiente de correlação r , que mede justamente a variabilidade dos dados, fica menor. Por consequência, o r tende a ficar maior do que o valor calculado com os dados sem agrupamento.
- c) **Falsa.** O valor de r igual a zero somente significa que a correlação LINEAR é inexistente. Isto porque, se duas variáveis tem um relacionamento que não seja linear (por exemplo, quadrático, exponencial, etc.) o coeficiente linear de Pearson (r) será próximo de zero. No entanto, as variáveis continuam correlacionadas, mas NÃO linearmente.

Exercício (2.21)

	Tórax em polegadas (X)	Peso em libras (Y)	XY	X ²	Y ²
	26	80	2080	676	6400
	45	344	15480	2025	118336
	54	416	22464	2916	173056
	49	348	17052	2401	121104
	35	166	5810	1225	27556
	41	220	9020	1681	48400
	41	262	10742	1681	68644
	49	360	17640	2401	129600
	39	204	7956	1521	41616
	31	144	4464	961	20736
Total	410	2544	112708	17488	755448

n = 10

X = Tamanho do tórax, em polegadas.

Y = Peso, em libras.

α = 0,05

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{n \sum_{i=1}^n (x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \left(\sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n y_i \right)^2}}$$

$$r = \frac{84040}{85671,7475} = 0,9809$$

O coeficiente de correlação está mais próximo de um, ou seja, a correlação linear entre Tamanho do Tórax, em polegadas e Peso, em libras, é positiva e forte, indicando que o aumento do tórax do animal é, em geral, acompanhado por um aumento de peso.

Exercício (2.22)

	Sistólica (X)	Diastólica (Y)	XY	X ²	Y ²
	138	82	11316	19044	6724
	130	91	11830	16900	8281
	135	100	13500	18225	10000
	140	100	14000	19600	10000
	120	80	9600	14400	6400
	125	90	11250	15625	8100
	120	80	9600	14400	6400
	130	80	10400	16900	6400
	130	80	10400	16900	6400
	144	98	14112	20736	9604
	143	105	15015	20449	11025
	140	85	11900	19600	7225
	130	70	9100	16900	4900
	150	100	15000	22500	10000
Total	1875	1241	167023	252179	111459

n = 14

X = Sistólica

Y = Diastólica

α = 0,05

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{n \sum_{i=1}^n (x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \left(\sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n y_i \right)^2}}$$

$$r = \frac{11447}{17399,826} = 0,6579$$

Como o coeficiente de correlação é positivo, mas não muito próxima de um, podemos dizer que há uma correlação linear positiva moderada entre uma variação na pressão sistólica e uma variação na pressão diastólica.

Exercício (2.23)

	Temperatura (X)	Tempo em min (Y)	XY	X ²	Y ²
	55	145,3	7991,5	3025	21112
	61	148,7	9070,7	3721	22112
	49	148,3	7266,7	2401	21993
	62	148,1	9182,2	3844	21934
	70	147,6	10332	4900	21786
	73	146,4	10687	5329	21433
	51	144,7	7379,7	2601	20938
	57	147,5	8407,5	3249	21756
Total	478	1176,6	70318	29070	173063

$$n = 8$$

X = Temperatura

Y = Tempo

$$\alpha = 0,05$$

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{n \sum_{i=1}^n (x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \left(\sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n y_i \right)^2}}$$

$$r = \frac{129,074}{683,6676} = 0,1888$$

A correlação linear entre Temperatura e Tempo é positiva, porém fraca, pois o coeficiente de correlação é mais próximo de zero.

Exercício (2.24)

	Cricrilos em 1min (X)	Temperatura (°F) (Y)	XY	X ²	Y ²
	882	69,7	61475	777924	4858,1
	1188	93,3	110840	1411344	8704,9
	1104	84,3	93067	1218816	7106,5
	864	76,3	65923	746496	5821,7
	1200	88,6	106320	1440000	7850
	1032	82,6	85243	1065024	6822,8
	960	71,6	68736	921600	5126,6
	900	79,6	71640	810000	6336,2
Total	8130	646	663245	8391204	52627

$$n = 8$$

X = Cricrilos em 1 min

Y = Temperatura (°F)

$$\alpha = 0,05$$

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{n \sum_{i=1}^n (x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \left(\sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n y_i \right)^2}}$$

$$r = \frac{53980}{61788,378} = 0,8736$$

Como o coeficiente de correlação é positivo e próximo do valor 1, podemos dizer que há uma correlação linear positiva forte entre o número de cricrilos e a temperatura ambiente.

Exercício (2.25)

	Altura em polegadas (X)	Taxa de Pulsação batimentos por min (Y)	XY	X ²	Y ²
	64,3	76	4886,8	4134,49	5776
	66,4	72	4780,8	4408,96	5184
	62,3	88	5482,4	3881,29	7744
	62,3	60	3738	3881,29	3600
	59,6	72	4291,2	3552,16	5184
	63,6	68	4324,8	4044,96	4624
	59,8	80	4784	3576,04	6400
	63,3	64	4051,2	4006,89	4096
	67,9	68	3633,9	31486,1	42608
	61,4	68	4175,2	3769,96	4624
	66,7	80	5336	4448,89	6400
	64,8	76	4924,8	4199,04	5776
Total	762,4	872	87114	75390,1	102016

n = 12

X = Altura

Y = Taxa de Pulsação

α = 0,05

$$r = \frac{\sum_{i=1}^n [(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{n \sum_{i=1}^n (x_i y_i) - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \left(\sum_{i=1}^n y_i^2 \right) - \left(\sum_{i=1}^n y_i \right)^2}}$$

$$r = \frac{-104}{2711,1621} = -0,0384$$

Como coeficiente de correlação é próximo de zero, a correlação linear entre a altura e a taxa de pulsação por segundo é mais fraca.

Seção 3: Probabilidade

Exercício (3.1)

$$a) E = \{ (M M M), (M M F), (M F M), (F M M), (M F F), (F M F), (F F M), (F F F) \}$$

onde M = Masculino e F = Feminino

$$b) E = \{ (A A), (A B), (A AB), (A O), (B A), (B AB), (B B), (B O), (AB A), (AB B), (AB AB), (AB O), (O A), (O B), (O AB), (O O) \}$$

$$c) E = \{ c: C_{\min} \leq c \leq C_{\max} \}$$

onde c: representa o nível de concentração de oxigênio

C_{\min} : representa o nível mínimo de concentração de oxigênio e

C_{\max} : representa o nível máximo de concentração de oxigênio.

O número de elementos do espaço amostral (E) é infinito.

$$d) E = \{ (F F F), (F F N), (F N F), (N F F), (F N N), (N F N), (N N F), (N N N) \}$$

onde F = Hábito de fumar e N = Hábito de não fumar

$$e) E = \{ (1 1), (1 2), (1 3), (1 4), (1 5), (1 6) \\ (2 1), (2 2), (2 3), (2 4), (2 5), (2 6) \\ (3 1), (3 2), (3 3), (3 4), (3 5), (3 6) \\ (4 1), (4 2), (4 3), (4 4), (4 5), (4 6) \\ (5 1), (5 2), (5 3), (5 4), (5 5), (5 6) \\ (6 1), (6 2), (6 3), (6 4), (6 5), (6 6) \}$$

Exercício (3.2)

$$E = \{1, 2, 3, 4, 5, 6\}$$

a) A: ocorre face par

$$A = \{2, 4, 6\}$$

$$P(A) = \frac{3}{6} = 0.5$$

B: ocorre face ímpar

$$B = \{1, 3, 5\}$$

$$P(B) = \frac{3}{6} = 0.5$$

C: ocorre face menor que 5

$$C = \{1, 2, 3, 4\}$$

$$P(C) = \frac{4}{6} = 0.67$$

D: ocorre face maior ou igual a 4

$$D = \{4, 5, 6\}$$

$$P(D) = \frac{3}{6} = 0.5$$

$$\begin{aligned} \text{b) } P(A \cap B) &= 0, & P(A \cap C) &= \frac{2}{6} = 0.33, & P(A \cap D) &= \frac{2}{6} = 0.33, \\ P(B \cap C) &= \frac{2}{6} = 0.33, & P(B \cap D) &= \frac{1}{6} = 0.17, & P(C \cap D) &= \frac{1}{6} = 0.17 \end{aligned}$$

c) Para o cálculo de probabilidade de união de eventos, serão utilizados os resultados do item, (a) e (b).

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.5 + 0.5 - 0 \\ &= 1, \end{aligned}$$

$$\begin{aligned} P(A \cup C) &= P(A) + P(C) - P(A \cap C) \\ &= 0.5 + 0.67 - 0.33 \\ &= 0.84, \end{aligned}$$

$$\begin{aligned} P(A \cup D) &= P(A) + P(D) - P(A \cap D) \\ &= 0.5 + 0.5 - 0.33 \\ &= 0.67, \end{aligned}$$

$$\begin{aligned} P(B \cup C) &= P(B) + P(C) - P(B \cap C) \\ &= 0.5 + 0.67 - 0.33 \\ &= 0.84, \end{aligned}$$

$$\begin{aligned} P(B \cup D) &= P(B) + P(D) - P(B \cap D) \\ &= 0.5 + 0.5 - 0.17 \\ &= 0.83, \end{aligned}$$

$$\begin{aligned} P(C \cup D) &= P(C) + P(D) - P(C \cap D) \\ &= 0.67 + 0.5 - 0.17 \\ &= 1. \end{aligned}$$

$$\text{d) } P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{\frac{1}{2}} = 0. \text{ Ou seja, como nenhuma face do dado é par e ímpar ao mesmo tempo, então}$$

$$P(A | B) = 0.$$

Como $P(A | B) \neq P(A)$, os eventos A e B não são independentes.

$$\text{e) } P(B | C) = \frac{P(B \cap C)}{P(C)} = \frac{\frac{2}{6}}{\frac{4}{6}} = 0.5. \text{ Ou seja, como, das 4 faces são menores do que 5, duas são ímpares, então}$$

$$P(B | C) = \frac{2}{4}.$$

Como $P(B | C) = P(B)$, os eventos A e B são independentes.

Exercício (3.3)

Tabela 3.1: Tratamento de gripe

Resposta	Placebo	Vacina	Total
Baixa	25	6	31
Moderada	8	18	26
Alta	5	11	16
Total	38	35	73

Sejam os eventos:

$P = \{ \text{o paciente escolhido recebeu tratamento com Placebo} \}$

$V = \{ \text{o paciente escolhido recebeu tratamento com Vacina} \}$

$B = \{ \text{o paciente escolhido obteve resposta Baixa} \}$

M = {o paciente escolhido obteve resposta Moderada}
 A = {o paciente escolhido obteve resposta Alta}

$$a) P(V) = \frac{35}{73} = 0.4795$$

$$b) P(P) = \frac{38}{73} = 0.5205$$

$$c) P(B) = \frac{31}{73} = 0.4246$$

$$d) P(M) = \frac{26}{73} = 0.3562$$

$$e) P(A) = \frac{16}{73} = 0.2192$$

$$f) P(B \cap V) = 6/73 = 0.0822$$

$$g) P(M \cap P) = 8/73 = 0.1096$$

$$h) P(V \cap A) = 11/73 = 0.1507$$

$$i) P(A \cup P) = P(A) + P(P) - P(A \cap P) = \frac{16}{73} + \frac{38}{73} - \frac{5}{73} = 0.6712$$

$$j) P(B \cup V) = P(B) + P(V) - P(B \cap V) = \frac{31}{73} + \frac{35}{73} - \frac{6}{73} = 0.8219$$

Exercício (3.4)

Tabela 3.2: Incidência de DHDA

Sexo	DHDA		
	Presente	Ausente	Totais
Masculino	60	440	500
Feminino	6	494	500
Totais	66	934	1000

Sejam os eventos:

P = {a criança escolhida apresenta Distúrbio}
 A = {a criança escolhida não apresenta Distúrbio}
 M = {a criança escolhida é do Sexo Masculino}
 F = {a criança escolhida é do Sexo Feminino}

$$a) P(P) = \frac{66}{1000} = 0.066$$

$$b) P(A) = \frac{934}{1000} = 0.934$$

$$c) P(M) = \frac{500}{1000} = 0.5$$

$$d) P(F) = \frac{500}{1000} = 0.5$$

$$e) P(P \cap M) = \frac{60}{1000} = 0.06$$

$$f) P(P \cap F) = \frac{6}{1000} = 0.006$$

$$g) P(A \cap M) = \frac{440}{1000} = 0.44$$

$$h) P(A \cap F) = \frac{494}{1000} = 0.494$$

$$i) P(P | M) = \frac{P(P \cap M)}{P(M)} = \frac{60}{500} = 0.12$$

$$j) P(P | F) = \frac{P(P \cap F)}{P(F)} = \frac{6}{500} = 0.012$$

$$k) P(A | M) = \frac{P(A \cap M)}{P(M)} = \frac{440}{500} = 0.88$$

$$l) P(A | F) = \frac{P(A \cap F)}{P(F)} = \frac{494}{500} = 0.988$$

$$m) P(F | P) = \frac{P(F \cap P)}{P(P)} = \frac{6}{66} = 0.0909$$

$$n) P(M | A) = \frac{P(M \cap A)}{P(A)} = \frac{440}{934} = 0.4711$$

Exercício (3.5)

Sejam os eventos: $A = \{\text{o paciente tem Angina}\}$
 $M = \{\text{o paciente é do Sexo Masculino}\}$
 $F = \{\text{o paciente é do Sexo Feminino}\}$

Temos que:

$$P(M) = 0.543; P(F) = 0.457; P(A | M) = 0.673 \text{ e } P(A | F) = 0.566$$

$$a) P(A) = P(A \cap M) + P(A \cap F) = P(A | M) P(M) + P(A | F) P(F) = (0.673)(0.543) + (0.457)(0.566) = 0.6241$$

$$b) P(M | A) = \frac{P(M \cap A)}{P(A)} = \frac{P(A | M) P(M)}{P(A)} = \frac{(0.673)(0.543)}{0.6241} = 0.5855$$

Exercício (3.6)

Sejam os eventos: $N = \{\text{Primeiro nome falso}\}$
 $\bar{N} = \{\text{Primeiro verdadeiro}\}$
 $S = \{\text{Sobrenome falso}\}$
 $\bar{S} = \{\text{Sobrenome verdadeiro}\}$

Temos que:

$$P(N) = 0.54; P(\bar{N}) = 0.46; P(\bar{S} | N) = 0.72; P(S | N) = 0.28 \text{ e } P(\bar{S} | \bar{N}) = 0.25$$

$$P(\bar{N} | \bar{S}) = ?$$

$$\begin{aligned} P(\bar{N} | \bar{S}) &= \frac{P(\bar{N} \cap \bar{S})}{P(\bar{S})} \\ &= \frac{P(\bar{S} | \bar{N}) \times P(\bar{N})}{P(\bar{S})} \\ &= \frac{P(\bar{S} | \bar{N}) \times P(\bar{N})}{P(\bar{S} \cap \bar{N}) + P(\bar{S} \cap N)} \\ &= \frac{P(\bar{S} | \bar{N}) \times P(\bar{N})}{[P(\bar{S} | \bar{N}) \times P(\bar{N}) + P(\bar{S} | N) \times P(N)]} \\ &= \frac{0.25 \times 0.46}{0.25 \times 0.46 + 0.72 \times 0.54} \\ &= 0.228 \end{aligned}$$

Seção 4: Avaliação da Qualidade de Testes Clínicos

Exercício (4.1)

$$\begin{aligned} \text{sensibilidade:} & \quad s = P(+ | D) = 97/100 = 0.97 \\ \text{especificidade:} & \quad e = P(- | \bar{D}) = 98/100 = 0.98 \end{aligned}$$

Como tanto a especificidade quanto a sensibilidade tiveram valores elevados, ou seja, próximos de um, temos que a proteína é um bom biomarcador de apendicite.

Exercício (4.2)

$$\begin{aligned} \text{a) sensibilidade:} & \quad s = P(+ | C) = 60/80 = 0.75 \\ \text{especificidade:} & \quad e = P(- | \bar{C}) = 70/75 = 0.93 \end{aligned}$$

$$\begin{aligned} \text{b) Prevalência na população:} & \quad p = 0.518 \\ \text{Prevalência na amostra:} & \quad 80/155 = 0.52 \cong p \end{aligned}$$

Como a prevalência da substância indevida na amostra é próxima da prevalência na população, podemos calcular VPP e VPN diretamente da tabela

$$\begin{aligned} \text{VPP} = P(C | +) &= \frac{60}{65} = 0.92 & \quad \text{PFP} = 1 - \text{VPP} = 0.08 \\ \text{VPN} = P(\bar{C} | -) &= \frac{70}{90} = 0.78 & \quad \text{PFN} = 1 - \text{VPN} = 0.22 \end{aligned}$$

Exercício (4.3)

$$\begin{aligned} \text{a) sensibilidade:} & \quad s = P(+ | D) = 436/450 = 0.97 \\ \text{especificidade:} & \quad e = P(- | \bar{D}) = 495/500 = 0.99 \end{aligned}$$

$$\begin{aligned} \text{b) Prevalência na população:} & \quad p = ? \\ \text{Prevalência na amostra:} & \quad 450/950 = 0.4737 \end{aligned}$$

Considerando que a prevalência da síndrome de Alzheimer na população é igual à prevalência na amostra, podemos calcular os índices pela tabela:

$$\begin{aligned} \text{VPP} = P(D | +) &= 436/441 = 0.9886 & \quad \text{PFP} = 1 - \text{VPP} = 0.0114 \\ \text{VPN} = P(\bar{D} | -) &= 495/509 = 0.9725 & \quad \text{PFN} = 1 - \text{VPN} = 0.0275 \end{aligned}$$

Exercício (4.4)

$$\begin{aligned} \text{a) } s = P(+ | D) &= 54/55 = 0.98 \\ e = P(- | \bar{D}) &= 82/83 = 0.99 \end{aligned}$$

$$\begin{aligned} \text{b) Prevalência na população} &= p = 0.10 \\ \text{Prevalência na amostra} &= 55/138 = 0.40 \end{aligned}$$

$$\text{VPP} = \frac{0.98 \times 0.10}{0.98 \times 0.10 + (1 - 0.99)(1 - 0.10)} = 0.916$$

$$\text{VPN} = \frac{0.99 \times (1 - 0.10)}{0.99(1 - 0.10) + (1 - 0.99)(0.10)} = 0.999$$

$$\text{c) } s = P(+ | D) = 27/31 = 0.87$$

$$e = P(- | \bar{D}) = 81/83 = 0.98$$

d) Prevalência na população: $p = 0.25$

Prevalência na amostra: $31/114 = 0.27$

Como a prevalência da doença na amostra é próxima da prevalência na população, podemos calcular VPP e VPN diretamente da tabela.

$$\text{VPP} = P(D | +) = 27/29 = 0.93$$

$$\text{VPN} = P(\bar{D} | -) = 81/85 = 0.95$$

Seção 5: Variáveis Aleatórias e Distribuição de Probabilidade

Exercício (5.1)

Considere as informações da Questão 3.1, assim, obtemos:

- a) X: número de crianças do sexo feminino entre os três escolhidos
Valores que X pode assumir: $x = 0, 1, 2$ ou 3

Distribuição de probabilidade de X

x	P[X=x]
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

- b) Y: número de indivíduos aidéticos de tipo de sangue AB entre os dois sorteados
Valores que Y pode assumir: $y = 0, 1$ ou 2

Distribuição de probabilidade de Y

y	P[Y=y]
0	9/16
1	6/16
2	1/16
Total	1

- c) Z: número de estudantes que fumam em sala de aula entre os três selecionados
Valores que Z pode assumir: $z = 0, 1$ ou 2

Distribuição de probabilidade de Z

z	P[Z=z]
0	1/8
1	3/8
2	3/8
3	1/8
Total	1

Exercício (5.2)

Seja

Distribuição de probabilidade de X

x	P[X=x]
0	1/4
1	1/2
2	1/4
Total	1

- a) Considere o seguinte exemplo para a distribuição de probabilidade acima.

Experimento: selecionar duas crianças ao acaso e verificar se são do sexo masculino

Espaço amostral: $E = \{M M, M F, F M, F F\}$,

onde $M = \{a \text{ criança selecionada é do sexo masculino}\}$ e

$F = \{a \text{ criança selecionada é do sexo feminino}\}$

X : número de crianças do sexo feminino entre os dois escolhidos.

Valores que X pode assumir: $x = 0, 1$ ou 2 .

Probabilidades: $P[X=x] = 1/4, 1/2$ e $1/4$, correspondente a cada valor da variável aleatória.

b) Valor esperado: $E[X] = 0*(1/4) + 1*(1/2) + 2*(1/4) = 1$

Variância : $Var[X] = (0 - 1)^2*1/4 + (1 - 1)^2 *1/2 + (2 - 1)^2*1/4 = 1/2 = 0.5$

Desvio padrão: $dp[X] = \sqrt{Var(X)} = 0.707$

Exercício (5.3)

Experimento: Uma urna contém três bolas numeradas 1, 2 e 3, seleciona-se duas bolas, uma de cada vez ao acaso e sem reposição da primeira bola.

Espaço amostral:

$E = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$

Cada um dos 6 pontos do espaço amostral E tem a mesma probabilidade de ocorrer, ou seja, $1/6$.

Variável aleatória X : soma dos números das bolas selecionadas.

Valores que X pode assumir: $x = 3, 4$, ou 5 .

Probabilidades de cada valor da variável "soma"		
Valores de soma	Frequência	Probabilidade
3	2	2/6
4	2	2/6
5	2	2/6
Total	6	1

Distribuição de probabilidade de X

x	$P[X=x]$
3	1/3
4	1/3
5	1/3
Total	1

Valor esperado: $E[X] = 3*(1/3) + 4*(1/3) + 5*(1/3) = 4$

Variância : $Var[X] = (3 - 4)^2*1/3 + (4 - 4)^2 *1/3 + (5 - 4)^2*1/3 = 5/3 = 1.67$

Desvio padrão: $dp[X] = \sqrt{Var(X)} = 1.29$

Exercício (5.4)

Experimento: Um determinado tratamento alcança 70% de cura de certa doença quando o mesmo é administrado a pacientes em condições bem definidas.

Espaço amostral:

$$E = \{ (C C C), (C C N), (C N C), (N C C), (C N N), (N C N), (N N C), (N N N) \}$$

onde C = paciente que foi curado e N = paciente que não foi curado

Variável aleatória X : número de pacientes que se submeteram ao tratamento e foram curados

Valores que X pode assumir: $x = 0, 1, 2$ ou 3 .

a) Sabemos que a probabilidade de um indivíduo se submeter ao tratamento alcança uma probabilidade de cura de 70%

Eventos	Probabilidade	x
C C C	$(0.70)^3$	3
C C N	$(0.70)^2 * 0.2$	2
C N C	$(0.70)^2 * 0.2$	2
N C C	$(0.70)^2 * 0.2$	2
C N N	$(0.2)^2 * 0.7$	1
N C N	$(0.2)^2 * 0.7$	1
N N C	$(0.2)^2 * 0.7$	1
N N N	$(0.2)^3$	0
Total	1	

A distribuição de probabilidade da variável aleatória X é dada por

Distribuição de probabilidade de X

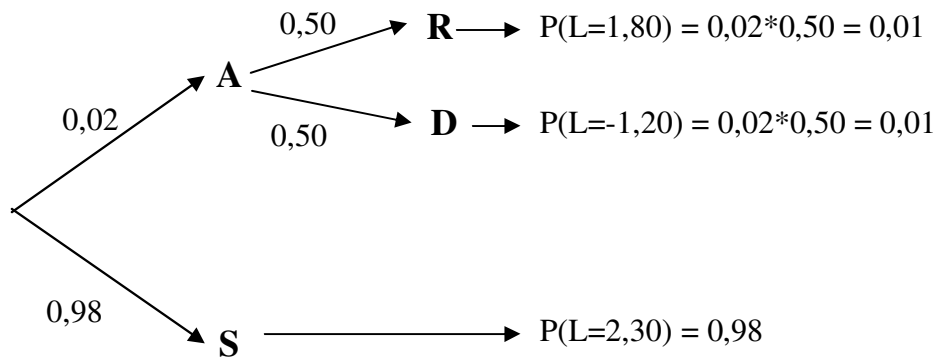
x	P[X=x]
0	$(0.2)^3$
1	$3 * (0.2)^2 * 0.7$
2	$3 * (0.70)^2 * 0.2$
3	$(0.70)^3$
Total	1

Exercício (5.5)

Seja L: lucro por muda produzida

$$L = \begin{cases} 3,50 - 1,20 = 2,30 & , \text{ muda sem ataque} \\ 3,50 - 1,70 = 1,80 & , \text{ muda atacada e recuperada} \\ 0 - 1,20 = -1,20 & , \text{ muda atacada e descartada} \end{cases}$$

O diagrama de árvores (ou árvore de probabilidades), será



Assim, a distribuição da variável aleatória “lucro por muda produzida” é dada por:

1	-1,20	1,80	2,30
P(L=1)	0,01	0,01	0,98

a) O lucro médio por muda produzida é dado por:

$$E(L) = -1,20 * P(L=-1,20) + 1,80 * P(L=1,80) + 2,30 * P(L=2,30) = -1,20 * 0,01 + 1,80 * 0,01 + 2,30 * 0,98 = 2,26$$

Assim, o lucro médio por muda produzida é de R\$2,26.

b) $10000 * E(L) = 10000 * 2,26 = 22600$

Assim, em uma plantação de 10000 mudas, o lucro esperado é R\$22600,00

Seção 6: Distribuição de Probabilidade: Binomial e Poisson

Exercício (6.1)

- ✓ O experimento ter um número fixo de tentativas;
- ✓ As tentativas têm que ser independentes. O resultado de qualquer tentativa individual não afeta as probabilidades nas outras tentativas;
- ✓ Cada tentativa deve ter todos os resultados classificados em duas categorias em geral, chamadas de sucesso e fracasso;
- ✓ A probabilidade de um sucesso permanece constante em todas as tentativas.

Exercício (6.2)

a) X: o número de mudas aproveitáveis em um lote de n=50 mudas.

p: probabilidade de sucesso, ou seja, a probabilidade de uma muda selecionada ao acaso, ser aproveitável é de 0.99.

1-p: probabilidade de fracasso, ou seja, a probabilidade de uma muda selecionada ao acaso, não ser aproveitável é de 0.01

Valores que X pode assumir: $x = 0, 1, 2, 3, \dots$ ou 50

Desse modo, $X \sim \text{Binomial}(n=50, p=0.99)$

Assim,
$$P(X = x) = \binom{50}{x} 0.99^x (1 - 0.99)^{50-x} \text{ para } x = 0, 1, 2, 3, \dots \text{ ou } 50$$

A probabilidade de que pelo menos 45 sejam aproveitáveis é dada por:

$$P(X \geq 45) = P(X = 45) + P(X = 46) + P(X = 47) + P(X = 48) + P(X = 49) + P(X = 50)$$

$$P(X = 45) = \frac{50!}{(50 - 45)!45!} \times (0.99)^{45} \times (0.01)^{50-45} = 0.001$$

$$P(X = 46) = \frac{50!}{(50 - 46)!46!} \times (0.99)^{46} \times (0.01)^{50-46} = 0.0015$$

$$P(X = 47) = \frac{50!}{(50 - 47)!47!} \times (0.99)^{47} \times (0.01)^{50-47} = 0.0122$$

$$P(X = 48) = \frac{50!}{(50 - 48)!48!} \times (0.99)^{48} \times (0.01)^{50-48} = 0.0756$$

$$P(X = 49) = \frac{50!}{(50 - 49)!49!} \times (0.99)^{49} \times (0.01)^{50-49} = 0.3056$$

$$P(X = 50) = \frac{50!}{(50 - 50)!50!} \times (0.99)^{50} \times (0.01)^{50-50} = 0.6050$$

$$P(X \geq 45) = 0.0001 + 0.0015 + 0.0122 + 0.0756 + 0.3056 + 0.6050 \cong 1 \cong 100\%$$

b) O experimento tem um número fixo de tentativas que são independentes, classificadas nas categorias sucesso e fracasso e a probabilidade do sucesso é sempre a mesma.

Exercício (6.3)

X: número de pessoas com alergia quando n=12 pessoas serão selecionadas aleatoriamente.

Valor que X pode assumir: $x = 7$

p: probabilidade de sucesso, ou seja, a probabilidade de cada pessoa ter alergia respiratória é de 0.6.

1-p: probabilidade de fracasso, ou seja, a probabilidade de cada pessoa não ter alergia respiratória é de 0,4.

Desse modo, $X \sim \text{Binomial}(n=12, p=0.6)$

Assim,
$$P(X = 7) = \frac{12!}{(12-7)!7!} \times 0.6^7 \times 0.4^{12-7} = 0.227$$

Exercício (6.4)

Seja X: número de aspirinas defeituosas em uma amostra de $n = 24$ comprimidos.

Valores que X pode assumir: $x = 0$ e 1

p: probabilidade de sucesso, ou seja, a aspirina ser defeituosa é de 0.04.

1-p: probabilidade de fracasso, ou seja, a aspirina não ser defeituosa é de 0.96.

Desse modo, $X \sim \text{Binomial}(n=24, p=0.04)$

Assim,

$$P(X = x) = \binom{24}{x} \cdot (0.04)^x (1 - 0.04)^{24-x} \text{ para } x = 0 \text{ ou } 1.$$

Estamos interessados em saber qual é a probabilidade que o carregamento seja aceito, isto é,

$$P(X \leq 1) = P(x = 0) + P(x = 1)$$

$$P(X = 0) = \frac{24!}{(24-0)!0!} \times (0.04)^0 \times (0.96)^{24-0} = 0.375$$

$$P(X = 1) = \frac{24!}{(24-1)!1!} \times (0.04)^1 \times (0.96)^{24-1} = 0.375$$

$$P(X \leq 1) = 0.375 + 0.375 = 0.75$$

A probabilidade de que o carregamento seja aceito é de 75%.

Exercício (6.5)

X : número de pacientes sobreviventes à cirurgia na amostra $n=10$.

Valores que X pode assumir: $x = 0, 1, 2, \dots$ ou 10

p: probabilidade de sucesso, ou seja, a probabilidade de um paciente sobreviver à cirurgia é igual a 0.9.

1-p: probabilidade de fracasso, ou seja, a probabilidade de um paciente não sobreviver à cirurgia, é igual a 0.1.

Desse modo, $X \sim \text{Binomial}(n=10, p=0.9)$

Assim,

$$P(X = x) = \binom{10}{x} \cdot (0.9)^x (1 - 0.9)^{10-x} \text{ para } x = 0, 1, 2, \dots \text{ ou } 10.$$

a) A probabilidade de que todos sobrevivam

$$P(X = 10) = \binom{10}{10} \times (0.9)^{10} \times (0.1)^{10-10} = \frac{10!}{(10-0)!0!} \times (0.9)^{10} \times (0.1)^{10-10} = 0.3487$$

A probabilidade de que todos os pacientes sobrevivam à cirurgia é de 34.87%.

b) A probabilidade de que ninguém sobreviva

$$P(X = 0) = \binom{10}{0} \times (0.9)^0 \times (0.1)^{10-0} = \frac{10!}{(10-0)!0!} \times (0.9)^0 \times (0.1)^{10-0} = 1.0^{-10} \\ = 0.0000000001$$

A probabilidade de que ninguém sobreviva à cirurgia é muito baixa.

c) A probabilidade de que nove ou mais sobrevivam

$$P(X \geq 9) = P(X = 9) + P(X = 10) = \frac{10!}{(10-9)!9!} \times (0.9)^9 \times (0.1)^{10-9} + 0.3487 = 0.3487 + 0.3874 = 0.7361$$

A probabilidade de que nove ou mais sobrevivam à cirurgia é de 73.61%.

d) A probabilidade de que pelo menos oito sobrevivam

$$P(X \geq 8) = P(X = 8) + P(X = 9) + P(X = 10) = \frac{10!}{(10-8)!8!} \times (0.9)^8 \times (0.1)^{10-8} + 0.3487 + 0.3874 \\ = 0.7361 + 0.1937 = 0.9298$$

A probabilidade de que pelo menos oito sobrevivam é de 92.98%.

e) O número esperado de pacientes submetidos à cirurgia sobreviva é de:

$$E[X] = n.p = 10(0.9) = 9 \text{ pacientes.}$$

A variância de pacientes submetidos à cirurgia sobreviva é de:

$$\text{Var}[X] = n.p.(1-p) = 10(0.9)(0.1) = 0.9$$

E o desvio padrão: $\sqrt{\text{Var}[X]} = 0.9487$ pacientes.

Exercício (6.6)

a) O modelo adequado é o modelo Binomial, pois o número de tentativas (de uma semente germinar) é fixo para cada pacote, e a probabilidade de sucesso (semente germinar) em cada experimento é sempre a mesma, isto é, uma semente germina independente dos outros.

X: número de sementes germinarem em um lote de $n = 10$ sementes

Valores que X pode assumir: $x = 0, 1, 2, \dots$ ou 10

p: probabilidade de sucesso, ou seja, a probabilidade de uma semente germinar é igual a 0.9.

1-p: probabilidade de fracasso, ou seja, a probabilidade de uma semente não germinar é igual a 0.1.

Desse modo, $X \sim \text{Binomial}(n=10, p=0.9)$

Assim,

$$P(X = x) = \binom{10}{x} \cdot (0.9)^x (1-0.9)^{10-x} \text{ para } x = 0, 1, 2, \dots \text{ ou } 10.$$

b) Num lote de 10 sementes, qual a probabilidade de todas germinarem.

$$P(X = 10) = \frac{10!}{(10-0)!0!} \times (0.9)^{10} \times (0.1)^{10-10} = 0.3487$$

A probabilidade de que em um lote de 10 sementes todas as sementes germinarem é de 34.87%.

c) Um cliente comprou um pacote destas sementes Qual a probabilidade de que pelo menos 7 sementes deste pacote germinem?

$$\begin{aligned}
P(X \geq 7) &= P(X = 10) + P(X = 9) + P(X = 8) + P(X = 7) \\
&= 0,3487 + 0,3874 + 0,1937 + \frac{10!}{(10-7)!7!} \times (0,9)^7 \times (0,1)^{10-7} \\
&= 0.9872
\end{aligned}$$

A probabilidade de que pelo menos 7 das 10 sementes do pacote germinem é de 98.72%.

- d) Em 80 destes pacotes, qual o número esperado de pacotes com todas as sementes germinadas.
 $E[X] = n.p = 80 (0.9) = 72$ pacotes.

Exercício (6.7)

- ✓ A variável aleatória X é o número de ocorrências de um evento ao longo de algum intervalo de tempo ou espaço.
- ✓ As ocorrências devem ser:
 - Aleatórias
 - Independentes umas das outras
 - Uniformemente distribuídas sobre o intervalo em uso.

Exercício (6.8)

X: O número de dentes-de-leão encontrados por metro quadrado em uma região.
Valores que X pode assumir: $x = 0, 1, 2, \dots$ (até um limite máximo desconhecido)
 λ : Média de dentes-de-leão encontrados por metro quadrado ($\lambda = 7$).

Supondo $X \sim \text{Poisson}(\lambda = 7)$
Assim,

$$P(X = x) = \frac{e^{-7} \cdot 7^x}{x!} \text{ para } x = 0, 1, 2, \dots$$

- a) Calcule a probabilidade de não se achar qualquer dente-de-leão em uma área de 1m^2 .

$$P(X=0) = \frac{(7,0)^0 e^{-7,0}}{0!} = 9,118 \times 10^{-4} = 0,0009$$

A probabilidade de não se achar qualquer dente-de-leão em uma área de 1m^2 é de 0.09%.

- b) Calcule a probabilidade de que pelo menos um dente-de-leão em uma área de 1m^2 .

$$P(X \geq 1) = 1 - P(X=0) = 1 - 0,0009118 = 0,999.$$

A probabilidade de que pelo menos um dente-de-leão é achado em uma área de 1m^2 é de 99.9%

- c) Calcule a probabilidade de no máximo dois dentes-de-leão em uma área de 1m^2 .

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

$$P(X \leq 2) = \frac{(7,0)^0 \times e^{-7,0}}{0!} + \frac{(7,0)^1 \times e^{-7,0}}{1!} + \frac{(7,0)^2 \times e^{-7,0}}{2!} = 0,0296$$

A probabilidade de que no máximo um dente-de-leão é achado em uma área de 1m^2 é de 2.96%.

Exercício (6.9)

X: O número mensal de adolescentes suicidas no condado.

Valores que X pode assumir: $x = 0, 1, 2, \dots$ (até um limite máximo desconhecido)

λ : média de adolescentes suicidas no condado ($\lambda = 2.75$).

Supondo $X \sim \text{Poisson}(\lambda = 2.75)$

Assim,

$$P(X = x) = \frac{e^{-2.75} \cdot (2.75)^x}{x!} \text{ para } x = 0, 1, 2, \dots$$

Parâmetro da distribuição de Poisson ($\lambda = 2.75$).

a) Qual é a probabilidade de que em um mês escolhido aleatoriamente ocorreu três suicídios de adolescentes

$$P(X=3) = \frac{(2.75)^3 e^{-2.75}}{3!} = \frac{(2.75)^3 e^{-2.75}}{6} = 0.2216$$

A probabilidade de ocorrer três suicídios de adolescentes em um determinado mês é de 22.169%.

b) Qual é a probabilidade de que em um mês escolhido aleatoriamente ocorreu três ou quatro suicídios de adolescentes

$$P(X = 3) + P(X = 4) = 0.2216 + \frac{(2.75)^4 e^{-2.75}}{4!} = 0.2216 + 0.1523 = 0.3739$$

A probabilidade de ocorrerem três ou quatro suicídios de adolescentes em um determinado mês é de 37.39%.

Exercício(6.10)

a) O parâmetro λ é estimado por: $(0 \cdot 190 + 1 \cdot 180 + 2 \cdot 85 + 3 \cdot 35 + 4 \cdot 8 + 5 \cdot 2) / 500 = 0.994$.

b) O cálculo das frequências esperadas segue:

Número de Prímulas por quadrado	Frequências Observadas	Probabilidade $P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}$	Frequências esperadas $P[X=x] \cdot 500$
0	190	0.37009	185.0467
1	180	0.36787	183.9364
2	85	0.18283	91.4164
3	35	0.06058	30.2893
4	8	0.01505	7.5269
5	2	0.00299	1.4963

No quadro acima podemos observar que as frequências observadas e esperadas superestimam quando o número de Prímulas por quadrado, são, 0 e 3, e subestima quando são, 1 e 2, os outros resultados são considerados próximos entre os valores observados e esperados. Assim, podemos concluir que, as plantas de Prímula seguem aproximadamente o modelo de distribuição de Poisson.

c) Cálculo da variância amostral:

$$s^2 = [190(0-0.994)^2 + 180(1-0.994)^2 + 85(2-0.994)^2 + 35(3-0.994)^2 + 8(4-0.994)^2 + 2(5-0.994)^2] / 499 = 1.04$$

Como $E[X] = \lambda = 0,994$ para o modelo de Poisson e a variância amostral encontrada é igual a 1,04, podemos notar que há uma pequena diferença entre estes valores o que indica que é aceitável a afirmação de que $\text{Var}[X] \cong E[X]$ no modelo de Poisson.

Seção 7: Distribuições de Probabilidade: Normal e Faixa de Referência

Exercício (7.1)

- a) $P[Z < 1.87] = 0.9693$
b) $P[Z > 1.28] = P[Z < -1.28] = 0.1003$
c) $P[Z > -1.87] = P[Z < 1.87] = 0.9693$
d) $P[Z < -1.28] = 0.1003$
e) $P[-2.30 < Z < 1.59] = P[Z < 1.59] - P[Z < -2.30] = 0.9334$

Exercício (7.2)

- a) $P[Z < a] = 0.0500$
 a é o percentil 5 da curva Normal Padrão
 $a = 1.645$
- b) $P[Z > a] = 0.1587$
 $P[Z < -a] = 0.1587$
 a é o percentil 15.87 da curva Normal Padrão
 $a = 1.00$.
- c) $P[Z > a] = 0.0500$
 a é o percentil 5 da curva Normal Padrão
 $a = 1.645$
- d) $P[Z < a] = 0.7500$
 a é o percentil 75 da curva Normal Padrão
 $a = 0.67$
- e) $P[-a < Z < a] = 0.8000$
 $P[Z < -a] = 0.1000$
 a é o percentil 10 da curva Normal Padrão
 $-a = -1.285$ e $a = 1.285$

Exercício (7.3)

Considerando a distribuição Normal Padrão, temos que $X \sim N(0,1)$

- a) $P(\mu - \sigma < X < \mu + \sigma) = P\left(\frac{\mu - \sigma - \mu}{\sigma} < Z < \frac{\mu + \sigma - \mu}{\sigma}\right)$
 $= P(-1 < Z < 1)$
 $= P(Z < 1) - P(Z < -1)$
 $= 0.8413 - 0.1587$
 $= 0.6826$
- b) $P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = P\left(\frac{\mu - 1.96\sigma - \mu}{\sigma} < Z < \frac{\mu + 1.96\sigma - \mu}{\sigma}\right)$
 $= P(-1.96 < Z < 1.96)$
 $= P(Z < 1.96) - P(Z < -1.96)$
 $= 0.9750 - 0.025$
 $= 0.95$
- c) $P(\mu - 3\sigma < X < \mu + 3\sigma) = P\left(\frac{\mu - 3\sigma - \mu}{\sigma} < Z < \frac{\mu + 3\sigma - \mu}{\sigma}\right)$
 $= P(-3 < Z < 3)$
 $= P(Z < 3) - P(Z < -3)$
 $\cong 1$
- d) $P(\mu - \sigma < X < \mu + 2\sigma) = P\left(\frac{\mu - \sigma - \mu}{\sigma} < Z < \frac{\mu + 2\sigma - \mu}{\sigma}\right)$
 $= P(-1 < Z < 2)$

$$= P(Z < 2) - P(Z < -1)$$

$$= 0.8185$$

$$\text{e) } P(\mu - 2\sigma < X < \mu + 2\sigma) = P\left(\frac{\mu - 2\sigma - \mu}{\sigma} < Z < \frac{\mu + 2\sigma - \mu}{\sigma}\right)$$

$$= P(-2 < Z < 2)$$

$$= P(Z < 2) - P(Z < -2)$$

$$= 0.9772 + 0.0228$$

$$= 1$$

Exercício (7.4)

X: peso do papel descartado semanalmente em residências

$X \sim \text{Normal}(\mu = 9.4 ; \sigma = 4.2)$

$$\text{a) } P(X > 15) = P\left[\frac{X - 9.4}{4.2} > \frac{15 - 9.4}{4.2}\right]$$

$$= P(Z > 1.33) = 0.0918$$

b) Queremos encontrar “a” tal que

$$P(X < a) = 0.33$$

$$P\left[\frac{X - 9.4}{4.2} < \frac{a - 9.4}{4.2}\right] = 0.33$$

$$P\left[Z < \frac{a - 9.4}{4.2}\right] = 0.33$$

$$\frac{a - 9.4}{4.2} = -0.44$$

$$a = 7.552 \text{ kg}$$

Exercício (7.5)

X: nível de colesterol sérico nos homens de 18 a 24 anos

$X \sim \text{Normal}(\mu = 178.1 ; \sigma = 40.7)$

$$\text{a) } P(X < 200) = P\left[\frac{X - 178.1}{40.7} < \frac{200 - 178.1}{40.7}\right]$$

$$= P(Z < 0.54) = 0.7054$$

b) Queremos encontrar “a” tal que

$$P(X < a) = 0.93$$

$$P\left[Z < \frac{a - \mu}{\sigma}\right] = 0.93$$

$$P\left[Z < \frac{a - 178.1}{40.7}\right] = 0.93$$

$$\frac{a - 178.1}{40.7} = 1.47$$

$$a = 237.929 \text{ mg}/100 \text{ mL}$$

c) A faixa de referência de 90% é dada por, $\left[\mu - z_{(\alpha/2)}\sigma ; \mu + z_{(\alpha/2)}\sigma \right]$

Como μ e σ são desconhecidos, vamos estimá-los por $\bar{x} = 34,3$ e $s = 14,2$ e, assim, a faixa de referência de 90% torna-se:

$$\begin{aligned} [\bar{x} - z(\alpha/2).s ; \bar{x} + z(\alpha/2).s] &= [178.1 - z_{(0,05)}(40.7) ; 178.1 + z_{(0,05)}(40.7)] \\ &= [178.1 - 1.645(40.7) ; 178.1 + 1.645(40.7)] \\ &= [111.1485 ; 245.0515] \text{ nível de colesterol sérico} \end{aligned}$$

Exercício (7.6)

X: altura da mulher

$X \sim N(\mu = 161.5; \sigma = 6.4)$

$$\begin{aligned} \text{a) } P(147 \leq X \leq 185) &= P\left(\frac{147 - \mu}{\sigma} < X < \frac{185 - \mu}{\sigma}\right) \\ &= P\left(\frac{147 - 161.5}{6.4} < X < \frac{185 - 161.5}{6.4}\right) \\ &= P(-2.27 < Z < 3.67) \\ &= P(Z < 3.67) - P(Z < -2.27) \\ &= 1 - 0.0116 \\ &= 0.9884 \end{aligned}$$

A porcentagem de mulheres da população que satisfazem a exigência das Forças Armadas é 98.84%.

b) Desenvolvendo através de faixa de referência $[\mu - z_{(\alpha/2)}\sigma ; \mu + z_{(\alpha/2)}\sigma]$

Como μ e σ são desconhecidos, vamos estimá-los por $\bar{x} = 161.5$ e $s = 6.4$ e, assim, a Faixa de Referência de 98% torna-se:

$$\begin{aligned} [\bar{x} - z(\alpha/2).s ; \bar{x} + z(\alpha/2).s] &= [161.5 - z_{(0,01)}(6.4) ; 161.5 + z_{(0,01)}(6.4)] \\ &= [161.5 - 2.33(6.4) ; 161.5 + 2.33(6.4)] \\ &= [146.588 ; 176.412] \text{ limites de altura} \end{aligned}$$

Exercício (7.7)

a) $a = P_{0,05\%}$ e $b = P_{0,95\%}$

Uma pessoa é sadia se estiver na faixa $[P_{0,05\%} ; P_{0,95\%}]$ e não sadia se estiver fora.

b) A especificidade de um teste é estimada como sendo a frequência de negativos entre as pessoas sadias. Para a faixa proposta no exercício, temos que 90% são consideradas negativas.

c) A sensibilidade de um teste é estimada como sendo a frequência de positivos entre as pessoas doentes. Se b é o percentil 5 na população doente, então 95% das pessoas doentes têm valores maiores do que b . E, pelo critério do teste, pessoas com valores acima de b são consideradas positivas. Sendo assim, 95% das pessoas doentes são consideradas positivas pelo teste. Ou seja, a sensibilidade do teste é 95%.

Para pensar: o que acontecerá com a especificidade do teste se aumentarmos a faixa de referência do teste para 95%? E o que acontecerá com a sensibilidade do teste?

Exercício (7.8)

a)
26 | 33
28 | 05378
30 | 68
32 | 9
34 | 27

Figura 7.1: Peso em kg dos meninos

Como o gráfico de ramos e folha mostra que os pesos não estão distribuídos de maneira simétrica, temos que os pesos dos meninos recém nascidos não se distribuem de acordo com o modelo Gaussiano.

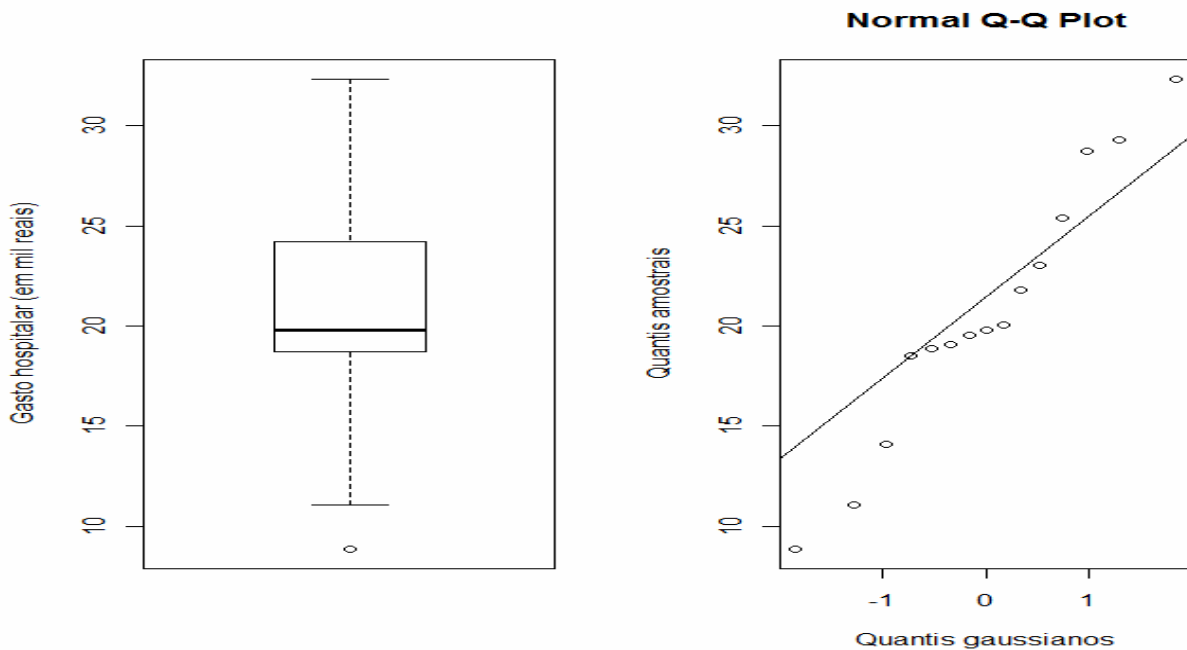
b) A distribuição dos pesos das meninas parece ser simétrica pela análise do box-plot e no qq-plot temos os quantis amostrais próximos dos gaussianos. Assim, podemos supor que o peso de meninas recém-nascidas se distribui de acordo com o modelo Gaussiano.

Exercício (7.9)

a) Os valores referentes a gastos hospitalares parecem se distribuir de forma assimétrica com concentração à esquerda e no qq-plot note-se que os valores dos quantis na parte superior direito encontram-se distante da linha normal o que indica que os quantis amostrais não são próximos dos quantis gaussianos. Assim, podemos supor que os gastos hospitalares não se distribuem de acordo a um modelo Gaussiano.

b)

Quantis amostrais	$f_i = \frac{i-0.5}{n}$	Quantis gaussianos	Quantis amostrais	$f_i = \frac{i-0.5}{n}$	Quantis gaussianos
8,84	0.03	-1.83	20,06	0.57	0.17
11,05	0.10	-1.28	21,80	0.63	0.34
14,11	0.17	-0.97	23,01	0.70	0.52
18,51	0.23	-0.73	25,38	0.77	0.73
18,87	0.30	-0.52	28,75	0.83	0.95
19,07	0.37	-0.34	29,31	0.90	1.28
19,56	0.43	-0.17	32,35	0.97	1.88
19,80	0.50	0			



Os gastos de internação para a amostra dada em questão, parece se distribuir de forma assimétrica com concentração a direita, no qqplot note-se que os valores dos quantis encontram-se fora da linha da normalidade, portanto, podemos supor que os gastos hospitalares para estes dados não segue um modelo Gaussiano.

Exercício (7.10)

Seja X: nota dos alunos

Sabe-se que $X \sim N(80; 5)$

Deseja-se criar uma escala de conceito A, B e C, com $A > B > C$ de modo que

5% dos alunos tenham conceito A

20% dos alunos tenham conceito B

50% dos alunos tenham conceito C

a) Valor de A

$$P(X > A) = 0.05 \Rightarrow P\left[\frac{X - 80}{5} > \frac{A - 80}{5}\right] = 0.05 \Rightarrow P\left[Z > \frac{A - 80}{5}\right] = 0.05 \Rightarrow P\left[Z < -\left(\frac{A - 80}{5}\right)\right] = 0.05$$

$$\text{Logo, } -\left(\frac{A - 80}{5}\right) = -1.645 \Rightarrow A = 88.2$$

Valor de B

$$P(X > B) = 0.25 \Rightarrow P\left[\frac{X - 80}{5} > \frac{B - 80}{5}\right] = 0.25 \Rightarrow P\left[Z > \frac{B - 80}{5}\right] = 0.25 \Rightarrow P\left[Z < -\left(\frac{B - 80}{5}\right)\right] = 0.25$$

$$\text{Logo, } -\left(\frac{B - 80}{5}\right) = -0.68 \Rightarrow B = 83.4$$

Valor de C

$$P(X > C) = 0.5 \Rightarrow P\left[\frac{X - 80}{5} > \frac{C - 80}{5}\right] = 0.5 \Rightarrow P\left[Z > \frac{C - 80}{5}\right] = 0.5 \Rightarrow P\left[Z < \frac{C - 80}{5}\right] = 0.5$$

$$\text{Logo, } \frac{C - 80}{5} = 0 \Rightarrow C = 80$$

b) Sabe-se que um aluno será reprovado se a nota dele for inferior a 70. Então devemos calcular $P(X < 70)$

$$P\left[\frac{X - 80}{5} < \frac{70 - 80}{5}\right] = P[Z < -2] = 0.0228$$

Desse modo, a probabilidade de um aluno ser reprovado é de 2.28%.

c) Considera-se que a distribuição da variável aleatória Y segue o modelo Binomial com parâmetros $n=30$ e probabilidade de sucesso $p = 0.0228$ (probabilidade de ser reprovado), pressupõe-se que cada um dos 30 alunos são independentes e todos com a mesma probabilidade de sucesso p .

d) Considerando o modelo Binomial adequado.

Temos que, $X \sim \text{Binomial}(n=30, p=0.0228)$

Assim,

$$\begin{aligned} P(X = 10) &= \binom{30}{10} \cdot (0.0228)^{10} (0.9772)^{30-10} \\ &= \frac{30!}{10!20!} (0.0228)^{10} (0.9772)^{20} \\ &\cong 0 \end{aligned}$$

A probabilidade de 10 alunos serem reprovados é de aproximadamente 0.

Seção 8: Distribuições Amostrais e Teorema Central do Limite

Exercício (8.1)

Como a amostra é grande, pelo Teorema Central do Limite, temos que: $\bar{X} \sim \text{Normal}\left(13, \frac{3}{\sqrt{60}}\right)$

$$\begin{aligned} \text{a) } P[\bar{X} < 14] &= P\left[\frac{\bar{X} - 13}{\frac{3}{\sqrt{60}}} < \frac{14 - 13}{\frac{3}{\sqrt{60}}}\right] \\ &= P[Z < 2,58] = 0,9951 \end{aligned}$$

A probabilidade do tempo médio de espera ser menor que 14 minutos é 99,51%.

$$\begin{aligned} \text{b) } P[\bar{X} > 11,5] &= P\left[\frac{\bar{X} - 13}{\frac{3}{\sqrt{60}}} > \frac{11,5 - 13}{\frac{3}{\sqrt{60}}}\right] \\ &= P[Z > -3,87] \\ &= P[Z < 3,87] \cong 1 \end{aligned}$$

A probabilidade do tempo médio de espera ser maior que 11.5 minutos é de aproximadamente 100%.

Exercício (8.2)

Como a amostra é grande, pelo Teorema Central do Limite, temos que:

$$\bar{X} \sim \text{Normal}\left(103, \frac{15}{\sqrt{50}}\right)$$

$$\begin{aligned} \text{a) } P[\bar{X} > 110] &= P\left[\frac{\bar{X} - 103}{\frac{15}{\sqrt{50}}} > \frac{110 - 103}{\frac{15}{\sqrt{50}}}\right] \\ &= P[Z > 3,30] \\ &= P[Z < -3,30] = 0,0005 \end{aligned}$$

A probabilidade do QI médio ser maior que 110 é 0.05%.

$$\begin{aligned} \text{b) } P[\bar{X} < 100] &= P\left[\frac{\bar{X} - 103}{\frac{15}{\sqrt{50}}} < \frac{100 - 103}{\frac{15}{\sqrt{50}}}\right] \\ &= P[Z < -1,41] = 0,0793 \end{aligned}$$

A probabilidade do QI médio ser menor que 100 é 7.93%.

Exercício (8.3)

Seja X: número de insetos colonizando uma planta.

Temos que: X segue uma distribuição Normal de média ($\mu = 20$ insetos) e desvio padrão ($\sigma = 5$ insetos)

Assim, $X \sim N(20, 5)$.

Considerando uma amostra de $n = 30$ plantas de algodão e pelo Teorema Central do Limite, segue que

$$\bar{X} \sim N\left(20, \frac{5}{\sqrt{30}}\right)$$

$$\text{a) } P[\bar{X} < 18] = P\left[\frac{\bar{X} - 20}{\frac{5}{\sqrt{30}}} > \frac{18 - 20}{\frac{5}{\sqrt{30}}}\right] = P[Z < -2,19] = 0,0143$$

A probabilidade do número médio de insetos observados para estas 30 plantas seja menor do que 18 é 1.43%.

$$\text{b) } \left[\bar{X} - z_{\alpha/2} \frac{5}{\sqrt{30}}; \bar{X} + z_{\alpha/2} \frac{5}{\sqrt{30}}\right] = \left[20 - 1,96 \frac{5}{\sqrt{30}}; 20 + 1,96 \frac{5}{\sqrt{30}}\right] = [18,21; 21,79]$$

O intervalo simétrico que inclua 95% dos valores de \bar{X} é [18.21; 21.79].

Exercício (8.4)

a) Seja X: consumo calórico diário em mulheres adolescentes saudáveis.

Temos que: X segue uma distribuição Normal de média ($\mu = 30$ kcal/kg) e desvio padrão ($\sigma = 5$ kcal/kg)

Assim, $X \sim N(30, 5)$.

Considerando uma amostra de $n = 50$ adolescentes.

Pelo Teorema Central do Limite, temos que a distribuição de $\bar{X} \sim N\left(30, \frac{5}{\sqrt{50}}\right)$

$$\text{b) } P[21,5 < \bar{X} < 24] = P\left[\frac{21,5 - 30}{\frac{5}{\sqrt{50}}} < \frac{\bar{X} - 30}{\frac{5}{\sqrt{50}}} < \frac{24 - 30}{\frac{5}{\sqrt{50}}}\right] = P[-12,02 < Z < -8,48] \cong 0$$

A probabilidade de que \bar{X} esteja entre 21.5 e 24 kcal/kg é aproximadamente 0.

Seção 9: Intervalos de Confiança

Exercício (9.1)

Variável X: peso dos bebês nascidos de mães que usaram cocaína.

Parâmetro μ : peso médio dos bebês nascidos de mães que usaram cocaína.

$$n = 190$$

$$\bar{x} = 2700g$$

$$s = 645g$$

Usaremos $E = z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$, isso é possível, pois o tamanho da amostra é considerado grande. O valor

crítico $z_{\frac{\alpha}{2}} = 1.96$ foi encontrado na Tabela da distribuição Normal, correspondente a $\alpha=0.05 \rightarrow$

$$\frac{\alpha}{2} = 0.025.$$

Calculando a margem de erro:

$$E = z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 1.960 \times \frac{645}{\sqrt{190}} = 91.715$$

Intervalo de 95% de confiança para μ : (σ desconhecido).

$$IC_{\mu}^{95\%} = [\bar{x} - E; \bar{x} + E]$$

$$IC_{\mu}^{95\%} = [2700 - 91.7147; 2700 + 91.715]$$

$$IC_{\mu}^{95\%} = [2608.285; 2791.715]$$

Conclusão: Com 95% de confiança, concluímos que a média de peso dos bebês que nasceram de mães usuárias de cocaína está entre 2608g e 2791g.

Exercício (9.2)

A pressão sanguínea sistólica média de 40 indivíduos está entre 114.4 e 123.4, com 99% de confiança.

Exercício (9.3)

Variável X: temperatura de adultos sadios.

Parâmetro μ : temperatura média de adultos sadios.

$$\bar{x} = 98,2^{\circ}F$$

$$s = 0,62^{\circ}F$$

$$n = 106$$

a) Usaremos $E = z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ isso é possível, pois o tamanho da amostra é considerada grande. O valor

crítico $z_{\frac{\alpha}{2}} = 1,96$ foi encontrado na Tabela da distribuição Normal, correspondente ao $\alpha=0,05 \rightarrow$

$$\frac{\alpha}{2} = 0.025.$$

Calculando a margem de erro:

$$E = z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 1.96 \times \frac{0,62}{\sqrt{106}} = 0.118$$

b) Calculando o intervalo de confiança:

$$IC_{\mu}^{95\%} = [\bar{x} - E; \bar{x} + E]$$

$$IC_{\mu}^{95\%} = [98.2 - 0.118; 98.2 + 0.118]$$

$$IC_{\mu}^{95\%} = [98.082; 98.318]$$

Conclusão: Com 95% de confiança, concluímos que a média da temperatura de 106 adultos sadios está entre 98,08 °F e 98,32 °F.

Exercício (9.4)

Variável: X: consumo de oxigênio do rim de pacientes com a moléstia.

Parâmetro μ : consumo médio de oxigênio do rim de pacientes com a moléstia.

$$\bar{x} = 13,90$$

$$s^2 = 0,67$$

$$n = 5$$

Visto que a variável aleatória X tem distribuição Normal com σ^2 desconhecido e a amostra é pequena, então o Intervalo de Confiança é dado por:

$$\left[\bar{x} - t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}}; \bar{x} + t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}} \right]$$

$$\alpha = 0.10 \rightarrow \frac{\alpha}{2} = \frac{0.10}{2} = 0.05$$

O valor crítico é $t_{\frac{\alpha}{2};(n-1)} = 2.132$, que é encontrado na Tabela t-Student, como o valor crítico correspondente a $n - 1 = 4$ graus de liberdade e a uma área em uma cauda de 0,05.

Calculando Margem de Erro:

$$E = t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}} = 2.132 \times \frac{\sqrt{0.67}}{\sqrt{5}} = 0.780$$

Calculando o Intervalo de Confiança:

$$IC_{\mu}^{90\%} = [\bar{x} - E; \bar{x} + E]$$

$$IC_{\mu}^{90\%} = [13.90 - 0,780; 13.90 + 0,780]$$

$$IC_{\mu}^{90\%} = [13.120; 14.680]$$

Conclusão: Com 90% de confiança, concluímos que o consumo médio de oxigênio do rim de pacientes com a moléstia está entre 13.09 cm³/min e 14.71 cm³/min.

Exercício (9.5)

Variável: X: peso de cada caixa de cereais em gramas.

Parâmetro μ : peso médio das caixas de cereais em gramas.

$$\bar{x} = 450.95$$

$$s^2 = 6.37$$

$$n = 10$$

Visto que a variável aleatória X tem distribuição Normal com amostra, então o Intervalo de Confiança é dado por:

$$\left[\bar{x} - t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}}; \bar{x} + t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}} \right]$$

$$\alpha = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

O valor crítico é $t_{\frac{\alpha}{2};(n-1)} = 2.262$, que é encontrado na Tabela t-Student, como o valor crítico correspondente a $n - 1 = 9$ graus de liberdade e a uma área em uma cauda de 0,025.

Calculando a margem de erro:

$$E = t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}} = 2.262 \times \frac{\sqrt{6.37}}{\sqrt{10}} = 1.805$$

Calculando o Intervalo de Confiança:

$$IC_{\mu}^{95\%} = [\bar{x} - E; \bar{x} + E]$$

$$IC_{\mu}^{95\%} = [13.90 - 1.805; 13.90 + 1.805]$$

$$IC_{\mu}^{95\%} = [12.095; 15.705]$$

Conclusão: Com 95% de confiança, concluímos que o peso médio de cada caixa de cereais está entre 12.095 gramas e 15.705 gramas.

Exercício (9.6)

Variável X: Permanência de pacientes no Hospital das Clínicas de São Paulo.

Parâmetro μ : Permanência média de pacientes no Hospital das Clínicas de São Paulo.

$$\bar{x} = 12 \text{ dias}$$

$$s = 8 \text{ dias}$$

$$n = 25 \text{ pacientes}$$

$$\alpha = 0,02$$

Visto que a variável aleatória X tem distribuição Normal com amostra, então o Intervalo de Confiança é dado por:

$$\left[\bar{x} - t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}}; \bar{x} + t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}} \right]$$

$$\alpha = 0.02 \rightarrow \frac{\alpha}{2} = \frac{0,02}{2} = 0.01$$

O valor crítico é $t_{\frac{\alpha}{2};(n-1)} = 2.492$, que é encontrado na Tabela t-Student, como o valor crítico correspondente a $n - 1 = 24$ graus de liberdade e a uma área em uma cauda de 0,01.

Calculando a margem de erro:

$$E = t_{\frac{\alpha}{2};(n-1)} \frac{s}{\sqrt{n}} = 2.492 \times \frac{\sqrt{8}}{\sqrt{25}} = 1.410$$

Calculando o Intervalo de Confiança:

$$IC_{\mu}^{98\%} = [\bar{x} - E; \bar{x} + E]$$

$$IC_{\mu}^{98\%} = [12 - 1.410; 12 + 1.410]$$

$$IC_{\mu}^{98\%} = [10.590; 13.41]$$

Conclusão: Com 98% de confiança, concluímos que a permanência média dos pacientes no Hospital das Clínicas de São Paulo está entre 10.59 dias e 13.41 dias.

Exercício (9.7)

X: número de meninas que nasceram de pais que usaram o método XSORT.

$$n = 325$$

$$\hat{p} = \frac{x}{n} = \frac{295}{325} = 0.91 \rightarrow \text{Proporção amostral de } x \text{ sucessos em uma amostra de tamanho } n$$

$$\alpha = 0,01$$

O valor crítico é $z_{\frac{\alpha}{2}} = 2.575$, que é encontrado na Tabela da distribuição Normal, como o valor crítico

correspondente a $\frac{\alpha}{2} = 0.0050$.

Calculando a margem de erro:

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2.575 \times \sqrt{\frac{0,91(1-0,91)}{325}} = 0.041$$

Calculando o Intervalo de Confiança:

$$IC_p^{99\%} = [\hat{p} - E; \hat{p} + E]$$

$$IC_p^{99\%} = [0.91 - 0.041; 0.91 + 0.041]$$

$$IC_p^{99\%} = [0.869; 0.951]$$

Conclusão: Com 99% de confiança, concluímos que o método parece ser eficaz, pois a proporção de meninas é substancialmente maior que 0.5.

Exercício (9.8)

X: número de dinamarqueses usuários de telefone celular que desenvolveram câncer no cérebro.

$$n = 420.095$$

$$\hat{p} = \frac{x}{n} = \frac{135}{420.095} = 3.214 \times 10^{-4} \rightarrow \text{Proporção amostral de } x \text{ sucessos em uma amostra de tamanho } n$$

$$\alpha = 0,05$$

O valor crítico é $z_{\frac{\alpha}{2}} = 1.96$, que é encontrado na Tabela da distribuição Normal, como o valor crítico

correspondente a $\frac{\alpha}{2} = 0,0250$.

Calculando a margem de erro:

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96 \times \sqrt{\frac{(3,214 \times 10^{-4}) \times (1 - (3,214 \times 10^{-4}))}{420.095}} = 5,421 \times 10^{-5}$$

Calculando o Intervalo de Confiança:

$$IC_p^{95\%} = [\hat{p} - E; \hat{p} + E]$$

$$IC_p^{95\%} = [(3,214 \times 10^{-4}) - (5,421 \times 10^{-5}); (3,214 \times 10^{-4}) + (5,421 \times 10^{-5})]$$

$$IC_p^{95\%} = [0,0267\%; 0,0376]$$

Conclusão: Com 95% de confiança, concluímos que os usuários de telefones celulares parecem não ter uma taxa de câncer do cérebro ou do sistema nervoso que seja diferente da taxa entre os que não usam telefones celulares, pois 0,0340% está incluído no intervalo de confiança.

Exercício (9.9)

Variável X: Número de dias entre o início dos sintomas da Leishmaniose Visceral e a notificação do caso às autoridades.

Pelo enunciado acima temos:

$$\text{Erro da estimativa: } E = \frac{\sigma}{2}$$

Coefficiente de confiança: $P(E) = \gamma = 0,95$.

O valor crítico $z_{\frac{\alpha}{2}} = 1.96$ foi encontrado na Tabela da distribuição Normal, correspondente a

$$\alpha = 0.05 \quad \rightarrow \quad \frac{\alpha}{2} = 0.025$$

Assim, podemos calcular o tamanho da amostra da seguinte forma:

$$n = \left(z_{\alpha/2} \times \frac{\sigma}{E} \right)^2 = \left(1.96 \times \frac{\sigma}{\sigma/2} \right)^2 = (1.96 \times 2)^2 = 15.366$$

Conclusão: Logo, para que o erro cometido na estimação do número médio de dias entre os sintomas da Leishmaniose Visceral e a notificação do caso às autoridades, seja a metade do desvio-padrão do número de dias, com probabilidade igual a 95%, a pesquisadora deve estudar aproximadamente 16 casos.

Exercício (9.10)

X: número de indivíduos contaminados em uma determinada região.

$$n = 2500$$

$$\hat{p} = \frac{x}{n} = \frac{850}{2500} = 0,34 \rightarrow \text{Proporção amostral de } x \text{ sucessos em uma amostra de tamanho } n$$

Intervalo de confiança com 95% de confiança:

$$\alpha = 0.05$$

O valor crítico é $z_{\frac{\alpha}{2}} = 1.96$, que é encontrado na Tabela da distribuição Normal, como o valor crítico

correspondente a $\frac{\alpha}{2} = 0.0250$.

Calculando a margem de erro:

$$E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1,96 \times \sqrt{\frac{0,34 \times (1-0,34)}{2500}} = 0,018569383$$

Calculando o Intervalo de Confiança:

$$IC_p^{95\%} = [\hat{p} - E; \hat{p} + E]$$

$$IC_p^{95\%} = [0,34 - 0,0186; 0,34 + 0,0186]$$

$$IC_p^{95\%} = [0,3214; 0,3586]$$

Conclusão: Com 95% de confiança, concluímos que a proporção de contaminados na população de uma determinada cidade afetada por um surto epidêmico está entre 0,3214 e 0,3586.

Exercício (9.11)

- a) Variável X: Número de ratos nos quais se desenvolve certo tipo de tumor quando submetidos à radiação.

Pelo enunciado acima temos:

- Erro da estimativa: $E=0,02$.

- Coeficiente de confiança: $P(E) = \gamma = 0,90$.

O valor crítico $z_{\frac{\alpha}{2}} = 1.65$ foi encontrado na Tabela da distribuição Normal, correspondente a

$$\alpha=0.10 \rightarrow \frac{\alpha}{2} = 0.05.$$

Como não temos uma informação preliminar sobre p , devemos utilizar $p=0,5$, que maximiza $p(1-p)$. Assim, podemos calcular o tamanho da amostra da seguinte forma:

$$n = \left(\frac{z_{\alpha/2}}{E} \right)^2 \times p(1-p) = \left(\frac{1.65}{0.02} \right)^2 \times 0.25 = 1701.563$$

Logo, para que o erro cometido na estimação da proporção de ratos nos quais se desenvolve certo tipo de tumor quando submetidos a radiação seja no máximo 0,02 com probabilidade igual a 0,90, o pesquisador precisa examinar 1702 animais.

- b) Como seria possível diminuir o tamanho da amostra utilizando a informação adicional de que em geral esse tipo de radiação não afeta mais que 20% dos ratos?

Se p for no máximo 20%, o tamanho da amostra será:

$$n = \left(\frac{z_{\alpha/2}}{E} \right)^2 \times p(1-p) = \left(\frac{1.65}{0.02} \right)^2 \times 0.20 \times 0.80 = 1089$$

Logo, se p for no máximo 20%, para que o erro cometido na estimação da proporção de ratos nos quais se desenvolve certo tipo de tumor quando submetidos a radiação seja no máximo 0,02 com probabilidade igual a 0,90, o pesquisador precisa examinar 1.089 animais.

Exercício (9.12)

$$E = 0.02$$

$$\alpha = 0.05$$

$$z \text{ é tal que } A(z) = 0.975 \rightarrow z = 1.96$$

Como não temos uma informação sobre p , devemos usar $p=0,5$, que maximiza $p(1-p)$.

Assim, podemos calcular o tamanho da amostra da seguinte forma:

$$n = \left(\frac{z}{\epsilon} \right)^2 p(1-p) = \left(\frac{1,96}{0,02} \right)^2 0,25 = 2401$$

O tamanho da amostra deve ser 2401 indivíduos para que as condições acima sejam satisfeitas.

Seja X : concentração da substância A no sangue em mg/cm^3

$$X \sim N(\mu; 0,4^2), \mu > 2.$$

$$P(X < 1.488) = P\left[Z < \frac{1.488 - 2}{0.4} \right] = P[Z < -1.28] = 0.1003$$

Assim, segundo um outro cientista, p é menor ou igual a 0.1003.

A informação acima podem ser utilizada pelo primeiro cientista para reduzir o tamanho da amostra, pois como o valor de p é no máximo 0.1, o valor máximo de $p(1-p)$ é atingido quando $p=0.10$, e assim:

$$n = \left(\frac{z}{\epsilon} \right)^2 p(1-p) = \left(\frac{1,96}{0,02} \right)^2 0,10 * 0,90 = 864.36$$

Neste caso, a informação do segundo cientista ajuda a reduzir o tamanho de amostra para aproximadamente 865 indivíduos.

Seção 10: Conceitos Básicos de Testes de Hipóteses

Exercício (10.1)

$n = 20$; média amostral = 49.35; desvio-padrão amostral = 2.7198 cm.

i. Parâmetro a ser testado:

μ : altura média dos recém nascidos no setor de Pediatria do Hospital das Clínicas da UNICAMP

ii. Hipóteses:

$H_0: \mu = 50$

$H_a: \mu < 50$

iii. Erro tipo I: Concluir que a altura média dos recém nascidos é menor de que 50 cm quando, na verdade, a altura média é 50 cm.

Erro tipo II: Concluir que a altura média dos recém nascidos é igual a 50 cm quando, na verdade, a altura média é menor que 50 cm.

iv. Estatística de teste:

$$T_{obs} = \frac{49.35 - 50}{\frac{2.7198}{\sqrt{20}}} = -1.0688$$

Região de Rejeição: Rejeita-se H_0 se $T_{obs} < -T_{(20-1);0,05} = -1.729$

Verificação: Como -1,0688 está fora da região de rejeição, não rejeitamos H_0 , ao nível de 5% de significância.

v. Valor p:

$$\text{Valor } p = P(t_{19,0,05} < -1.0688) = 0.15$$

Conclusão: Ao nível de 5% de significância, não temos evidências para rejeitar a hipótese de que a altura média de recém nascidos é igual a 50 cm (Valor $p = 0.15$).

Exercício (10.2)

$n = 64$; média amostral = 133 mmHg; desvio-padrão = 16 mmHg.

i. Parâmetro a ser testado:

μ : Pressão sistólica média em uma população de homens

a)

ii. Hipóteses:

$H_0: \mu = 130$

$H_a: \mu > 130$

iii. Erro tipo I: Concluir que a pressão sistólica média em homens é maior do que 130 mmHg, quando, na verdade, a pressão sistólica média é 130 mmHg.

Erro tipo II: Concluir que a pressão sistólica média em homens é igual a 130 quando, na verdade, a pressão sistólica média é maior que 130 mmHg.

iv. Estatística de teste:

$$Z_{obs} = \frac{133 - 130}{\frac{16}{\sqrt{64}}} = 1,5$$

Região de Rejeição: Rejeita-se H_0 se $Z_{obs} > Z_{0,05} = 1.645$

Verificação: Como 1.5 está fora da região de rejeição, não rejeitamos H_0 , ao nível de 5% de significância.

v. Valor p:

$$\text{Valor } p = P(Z > 1.5) = 0.0668$$

Conclusão: Ao nível de 5% de significância, não temos evidências suficientes para rejeitar a hipótese de que a pressão sistólica média é igual a 130 mmHg (Valor $p = 0.0668$).

b)

ii. Hipóteses:

$$H_0: \mu = 130$$

$$H_a: \mu \neq 130$$

iii. Erro tipo I: Concluir que a pressão sistólica média em homens é diferente de 130 quando, na verdade, a pressão sistólica média é 130 mmHg.

Erro tipo II: Concluir que a pressão sistólica média em homens é igual a 130 quando, na verdade, a pressão sistólica média é diferente de 130 mmHg.

iv. Estatística de teste:

$$Z_{obs} = \frac{133 - 130}{\frac{16}{\sqrt{64}}} = 1,5$$

Região de Rejeição: Rejeita-se H_0 se $Z_{obs} < -Z_{0,025}$ ou $Z_{obs} > Z_{0,025}$
 $Z_{obs} < -1.96$ ou $Z_{obs} > 1.96$

Verificação: Como 1.5 está fora da região de rejeição, não rejeitamos H_0 , ao nível de 5% de significância.

v. Valor p:

$$\text{Valor } p = 2 \times P(Z > 1,5) = 2 \times 0,0668 = 0,1336$$

Conclusão: Ao nível de 5% de significância, não temos evidências suficientes para rejeitar a hipótese de que a pressão sistólica média é igual a 130 mmHg (Valor $p = 0.1336$).

Como a hipótese alternativa é bilateral, podemos usar o intervalo de confiança para realizar o teste de hipóteses. O intervalo terá 95% de confiança, já que, o nível de significância é de 5%.

vi. Intervalo de confiança para μ :

$$100(1 - \alpha) = 95$$

$$1 - \alpha = 0.95$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$Z_{0,025} = 1,96$$

$$IC_{\mu}^{95\%} = \left[133 \pm 1,96 \times \frac{16}{\sqrt{64}} \right]$$

$$IC_{\mu}^{95\%} = [133 \pm 3,92]$$

$$IC_{\mu}^{95\%} = [129,08 ; 136,92]$$

Interpretação: A pressão sistólica média está entre 129.08 e 136.92, com 95% de confiança.

Exercício (10.3)

n = 50 (amostra grande), proporção amostral = 7/50 = 0.14

i. Parâmetro a ser testado:

p: proporção de embriões de frangos com anomalia

ii. Hipóteses:

$$H_0: p = 0.25$$

$$H_a: p < 0.25$$

iii. Erro tipo I: Concluir que a proporção de embriões de frangos com anomalia é menor de que 0.25 quando, na verdade, a proporção de embriões de frangos com anomalia é 0.25.

Erro tipo II: Concluir que a proporção de embriões de frangos com anomalia é igual a 0.25 quando, na verdade, a proporção de embriões de frangos com anomalia é menor que 0.25.

iv. Estatística de teste:

$$Z_{\text{obs}} = \frac{0.14 - 0.25}{\sqrt{\frac{0.25(1 - 0.25)}{50}}} = -1.7963$$

Região de Rejeição: Rejeita-se H_0 se $Z_{\text{obs}} < -Z_{0.05} = -1.645$

Verificação: Como -1.7963 está na região de rejeição, rejeitamos H_0 , ao nível de 5% de significância.

v. Valor p:

$$\text{Valor } p = P(Z < -1.7963) = 0.0359$$

Conclusão: Ao nível de 5% de significância, temos evidências para rejeitar a hipótese de que a proporção de embriões com anomalia é igual a 0.25 (Valor p = 0.0359).

Exercício (10.4)

n = 696 (amostra grande), proporção amostral = 0.63

i. Parâmetro a ser testado:

p: Proporção de terapeutas ocupacionais que deixam a profissão

a)

ii. Hipóteses:

$H_0: p = 0,60$

$H_a: p > 0,60$

iii. Erro tipo I: Concluir que a proporção de terapeutas ocupacionais que deixam a profissão é maior do que 0.60 quando, na verdade, a proporção de terapeutas ocupacionais que deixam a profissão é 0.60.

Erro tipo II: Concluir que a proporção de terapeutas ocupacionais que deixam a profissão é igual a 0.60 quando, na verdade, a proporção de terapeutas ocupacionais que deixam a profissão é maior que 0.60.

iv. Estatística de teste:

$$Z_{\text{obs}} = \frac{0.63 - 0.60}{\sqrt{\frac{0.60(1 - 0.60)}{696}}} = 1.6155$$

Região de Rejeição: Rejeita-se H_0 se $Z_{\text{obs}} > Z_{0.05} = 1.645$

Verificação: Como 1.6155 está fora da região de rejeição, não rejeitamos H_0 , ao nível de 5% de significância.

v. Valor p:

$$\text{Valor } p = P(Z > 1.6155) = 0.0526$$

Conclusão: Ao nível de 5% de significância, não temos evidências para rejeitar a hipótese de que a proporção de terapeutas ocupacionais que abandonam a profissão é igual a 0.60 (Valor p = 0.0526).

b)

ii. Hipóteses:

$H_0: p = 0.60$

$H_a: p \neq 0.60$

iii. Erro tipo I: Concluir que a proporção de terapeutas ocupacionais que deixam a profissão é diferente de 0.60 quando, na verdade, a proporção de terapeutas ocupacionais que deixam a profissão é igual a 0.60.

Erro tipo II: Concluir que a proporção de terapeutas ocupacionais que deixam a profissão é igual a 0.60 quando, na verdade, a proporção de terapeutas ocupacionais que deixam a profissão é diferente de 0.60.

iv. Estatística de teste:

$$Z_{\text{obs}} = \frac{0.63 - 0.60}{\sqrt{\frac{0.60(1 - 0.60)}{696}}} = 1.6155$$

Região de Rejeição: : Rejeita-se H_0 se $Z_{\text{obs}} < -Z_{0.025}$ ou $Z_{\text{obs}} > Z_{0.025}$

$$Z_{\text{obs}} < -1.96 \text{ ou } Z_{\text{obs}} > 1.96$$

Verificação: Como 1.6155 está na região de rejeição, rejeitamos H_0 , ao nível de 5% de significância.

v. Valor p:

$$\text{Valor p} = 2 \times P(Z > 1.6155) = 2 \times 0.0526 = 0.1052$$

Conclusão: Ao nível de 5% de significância, não temos evidências suficientes para rejeitar a hipótese de que a proporção de mulheres que abandonam a profissão é igual a 0.60 (Valor p = 0.0526).

Como a hipótese alternativa é bilateral podemos usar o intervalo de confiança para realizar o teste de hipóteses. O intervalo terá 95% de confiança, já que, o nível de significância é de 5%.

vi. Intervalo de confiança para p:

$$100(1 - \alpha) = 95$$

$$1 - \alpha = 0.95$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$Z_{0,025} = 1.96$$

$$IC_p^{95\%} = \left[0.63 \pm 1.96 \times \sqrt{\frac{0.63(0.37)}{696}} \right]$$

$$IC_\mu^{95\%} = [0.63 \pm 0.04]$$

$$IC_\mu^{95\%} = [0.59; 0.67]$$

Interpretação: A proporção de mulheres grávidas que abandonam a profissão está entre 59% e 67%, com 95% de confiança.

Para pensar mais um pouco: se a hipótese de pesquisa fosse que a proporção de mulheres grávidas que abandonam a profissão é diferente de 0.50, qual seria a conclusão da letra **b)**?

Exercício (10.5)

n = 670 (amostra grande), proporção amostral = 0.66

i. Parâmetro a ser testado:

p: Proporção de crianças que tomaram a série completa de vacinas contra a Hepatite B

ii. Hipóteses:

$$H_0: p = 0,60$$

$$H_a: p > 0,60$$

iii. Erro tipo I: Concluir que a proporção de crianças que tomaram a série completa de vacinas é maior do que 0,60 quando, na verdade, a proporção de crianças que tomaram a série completa de vacinas é 0,60.

Erro tipo II: Concluir que a proporção de crianças que tomaram a série completa de vacinas é igual a 0,60 quando, na verdade, a proporção de terapêutas crianças que tomaram a série completa de vacinas é maior que 0,60.