Universidade Federal de Pernambuco
Centro de Ciências Sociais Aplicadas
Departamento de Ciências Contábeis e Atuariais

Lista 1

Resumindo os dados

Gráficos

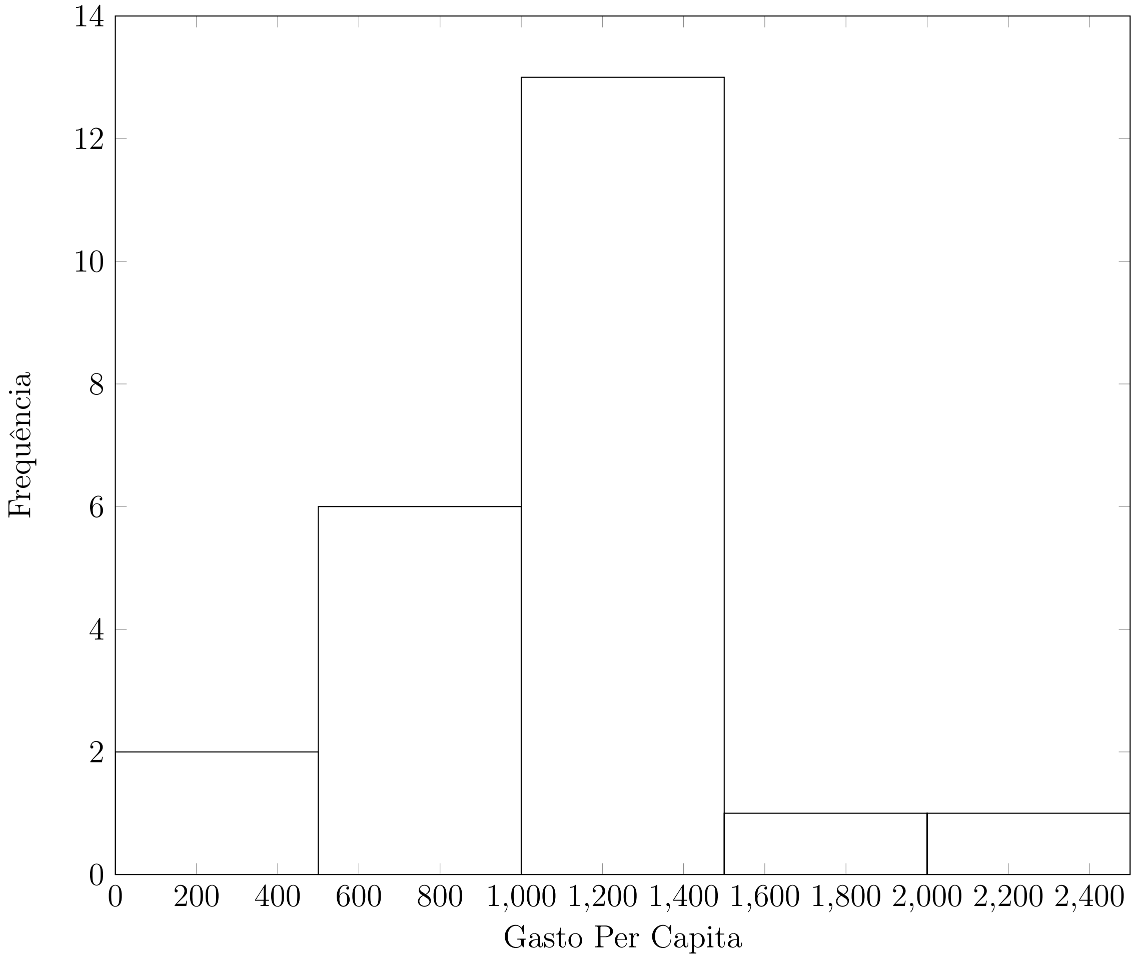
1º) Na tabela abaixo são listados os gastos com os cuidados com a saúde per capita em 1989 para 23 das 24 nações que constituem a Organização para Cooperação Econômica e Desenvolvimento.

Nação	**Gastos Per Capita (US$)**
Austrália	1032,00
Áustria	1093,00
Bélgica	980,00
Inglaterra	836,00
Canadá	1683,00
Dinamarca	912,00
Finlândia	1067,00
França	1274,00
Alemanha	1232,00
Grécia	371,00
Islândia	1353,00
Irlanda	658,00
Itália	1050,00
Japão	1035,00
Luxemburgo	1193,00
Países Baixos	1135,00
Nova Zelândia	820,00
Noruega	1234,00
Portugal	464,00
Espanha	644,00
Suécia	1361,00
Suíça	1376,00
Estados Unidos	2354,00

a) Ordene esses países de acordo com os gastos per capita com a saúde.

Nação	Gastos Per Capita (US$)
Grécia	371,00
Portugal	464,00
Espanha	644,00
Irlanda	658,00
Nova Zelândia	820,00
Inglaterra	836,00
Dinamarca	912,00
Bélgica	980,00
Austrália	1032,00
Japão	1035,00
Itália	1050,00
Finlândia	1067,00
Áustria	1093,00
Países Baixos	1135,00
Luxemburgo	1193,00
Alemanha	1232,00
Noruega	1234,00
França	1274,00
Islândia	1353,00
Suécia	1361,00
Suíça	1376,00
Canadá	1683,00
Estados Unidos	2354,00

b) Construa um histograma para os valores dos gastos per capita.



c) Descreva a forma do histograma.

O histograma tem maior frequência entre os intervalos de classe medianos, US$ 500,00 a US$ 1.000,00 e entre US$ 1.000,00 US$1.500,00, sugerindo uma possível distribuição normal dos dados, mas sendo necessário maiores análises para realizar tal afirmativa.

2º) A tabela abaixo categoriza 10.614.000 visitas ao consultório de especialistas de doenças cardiovasculares nos estados unidos por duração de cada visita. Uma duração de 0 (zero) minuto implica que o paciente não teve contato direto com o especialista.

Duração (minutos)	**Número de visitas (milhares)**
0	390
1-5	227
6-10	1023
11-15	3390
16-30	4431
31-60	968
61+	185
Total	10614

Pode-se fazer a afirmação de que as visitas a consultórios de especialistas de doenças cardiovasculares têm duração mais frequente entre 16 e 30 min. Você concorda com essa afirmação? Por quê? Justifique.

Sim, porque o período entre 16 e 30 min apresenta maior frequência que os demais intervalos de tempo de consulta.

Medidas de posição e de dispersão

3º) Na tabela abaixo estão taxas de glicose em miligramas por 100ml de sangue em ratos machos da raça Wistar com 30 dias de idade, que serão usados em um experimento para o teste de determinada droga. Ache a média e a mediana. A moda seria uma informação importante? Por quê?

Tabela 01 - Taxas de glicose em miligramas por 100ml de sangue em ratos machos da raça Wistar com 30 dias de idade

Nº do Rato	**Taxa de glicose**
1	101
2	98
3	97
4	104
5	95
6	105

$\rightarrow$

Nº do Rato	**Taxa de glicose**
5	95
3	97
2	98
1	101
4	104
6	105

$$\text{Média} = \frac{95+97+98+101+104+105}{6} = \frac{600}{6} = 100$$

$$\text{Mediana} = \frac{98+101}{2} = \frac{199}{2} = 99,5$$

Não, porque a frequência de todos os valores é igual a um.

4º) Com os dados apresentados na tabela que segue, calcule o número médio de dentes cariados, para cada sexo.

Tabela 02 – escolares de 12 anos, segundo o número de dentes cariados e o sexo

Nº de dentes cariados	**Sexo**	
	Masculino	**Feminino**
0	16	13
1	2	5
2	3	3
3	2	2
4	2	2

$$\text{Média}_{Masc} = \frac{0*16+1*2+2*3+3*2+4*2}{25} = \frac{0+2+6+6+8}{25} = \frac{22}{25} = 0,88$$

$$\text{Média}_{Fem} = \frac{0*13+1*5+2*3+3*2+4*2}{25} = \frac{0+5+6+6+8}{25} = \frac{25}{25} = 1$$

5º) Para estudar o tempo de latência de um sonífero usado em ratos de laboratório, um pesquisador administrou o sonífero a 10 ratos e determinou o tempo que eles demoravam para dormir. Dos 10 ratos, dois demoraram meio minuto, quatro demoraram 1 minuto, três demoraram 1 minuto e meio e um rato não dormiu. Calcule o tempo médio da latência.

$$\text{Média} = \frac{2*0,5+4*1+3*1,5}{9} = \frac{1+4+4,5}{9} = \frac{9,5}{9} = 1,055$$

6º) Para calcular dois programas de treinamento para executar um serviço especializado, foi feito um experimento. Dez homens foram selecionados ao acaso para serem treinados pelo método A e outros 10 para serem treinados pelo método B. Terminado o treinamento, todos os homens fizeram o serviço e foi registrado o tempo em que cada um desempenhou a tarefa. Os dados estão na tabela a seguir. Encontre para cada método: mínimo, primeiro quartil, mediana, terceiro quartil e máximo. Compare e discuta.

Tabela 03 – Tempo, em minutos, despendido em executar o serviço,segundo o método de treinamento

Método	
A	**B**
15	23
20	31
11	13
23	19
16	23
21	17
18	28
16	26
27	25
24	28

→

Método	
A	**B**
11	13
15	17
16	19
16	23
18	23
20	25
21	26
23	28
24	28
27	31

$$\text{Mediana}_A = \frac{18+20}{2} = \frac{38}{2} = 19$$

$$\text{Mediana}_B = \frac{23+25}{2} = \frac{48}{2} = 24$$

Método A				
Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
11	16	19	23	27
Método B				
Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
13	19	24	28	31

Todas as medidas tendência de A são menores que B, portanto, o método de treinamento A é mais eficiente.

7º) Responda às questões:
a) O valor do desvio padrão pode ser maior do que o valor da média?

Sim, o valor da média pode assumir qualquer valor do conjunto dos números $\mathbb{R}$, enquanto o desvio padrão pertence apenas ao intervalo dos $\mathbb{R}$ maiores ou ou iguais a zero.

b) O valor do desvio padrão pode ser igual ao valor da média?

Sim, a média pode assumir valor igual ao desvio padrão.

c) O valor do desvio padrão pode ser negativo?

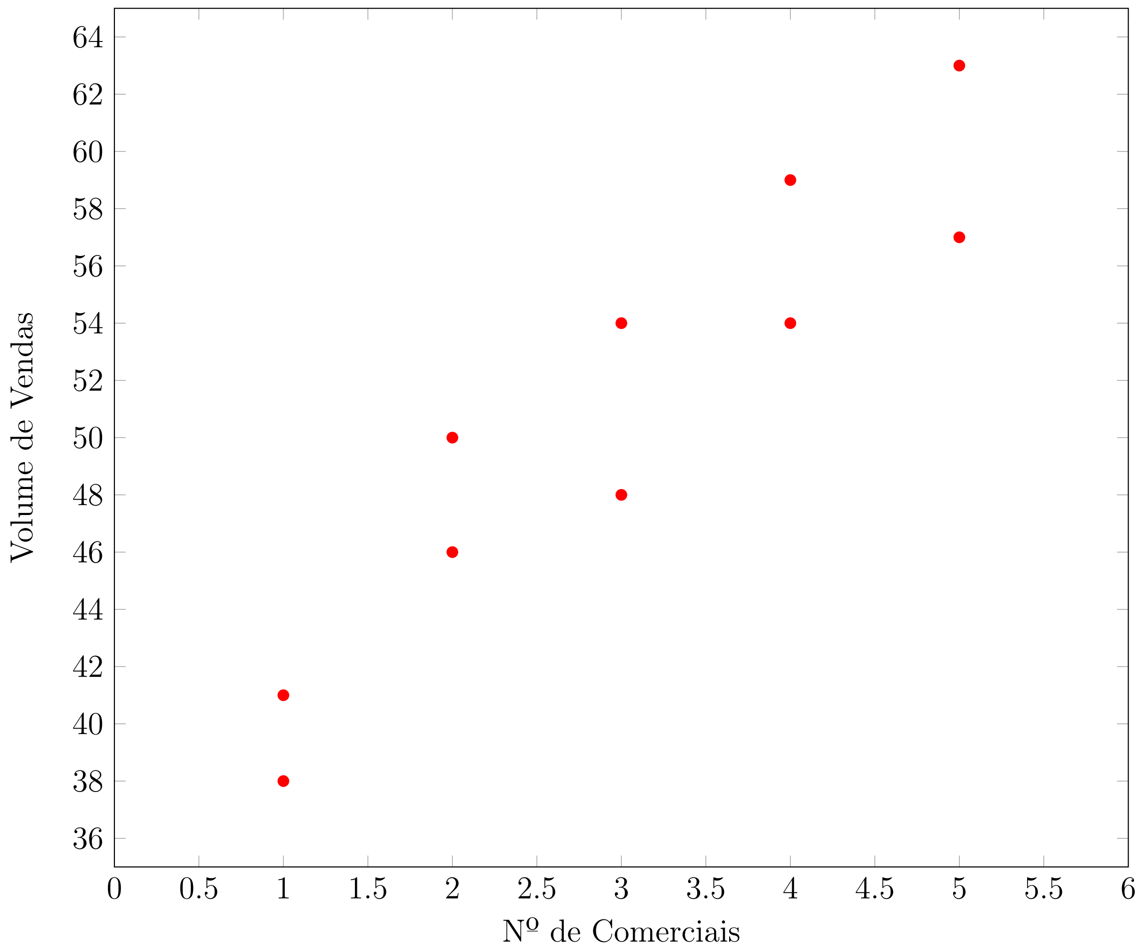
Não, os valores do desvio padrão são resultado da raiz de uma potenciação, portanto os valores são sempre absolutos.

d) Quando o desvio padrão é igual a zero?

O desvio padrão é igual a zero, quando todas as observações numéricas são iguais.

8º) Considere a situação de uma loja de equipamentos de som. Em dez ocasiões durante os três últimos meses , a loja usou comerciais de televisão de fins de semana para promover as vendas. Os gerentes querem verificar se pode ser demonstrada uma relação entre o numero de comerciais mostrados e as vendas na loja durante a semana seguinte. Os dados para as 10 semanas com vendas em centenas de dólares estão mostrados na tabela abaixo. Construa o diagrama de dispersão considerando o número de comerciais como sendo o eixo (x) e as vendas o eixo (y). Faça comentários sobre a relação entre as duas variáveis.

Semana	**Nº de Comerciais(x)**	**Volume de Vendas (por R$ 100)**
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46



O gráfico sugere a existência de uma associação linear direta entre o número de comerciais veiculados e o volume de vendas.

9º) Um estudo foi conduzido para investigar o prognóstico a longo prazo de crianças que sofreram um episódio agudo de meningite bacteriana. Abaixo estão listados os tempos para o ataque apoplético de 13 crianças que tomaram parte no estudo. Em meses, as medidas foram:

0,10 0,20 0,25 4 12 12 24 24 31 36 42 55 96

a) Obtenha as seguintes medidas-resumo numéricas dos dados
i. Média

$$\text{Média} = \frac{0,10 + 0,20 + 0,25 + 4 + 12 + 12 + 24 + 24 + 31 + 36 + 42 + 55 + 96}{13}$$

$$\text{Média} = \frac{336,55}{13} = 25,89$$

ii. Mediana

$$\text{Mediana} = 24$$

iii. Moda

$$\text{Moda} = 12 \text{ e } 24$$

iv. Amplitude

$$\text{Amplitude} = 96 - 0,10 = 95,9$$

v. Intervalo interquartil

$$\text{Intervalo interquartil} = \frac{(42+36)-(4+0,25)}{2} = \frac{73,75}{2} = 36,875$$

vi. Desvio padrão

$(x-\overline{x})$		$(x-\overline{x})^2$
(0,10 - 25,89)	-25,79	665,1241
(0,20 - 25,89)	-25,69	665,9761
(0,25 - 25,89)	-25,64	657,4096
(4 - 25,89)	-21,89	479,1721
(12 - 25,89)	-13,89	192,9321
(12 - 25,89)	-13,89	192,9321
(24 - 25,89)	-1,89	3,5721
(24 - 25,89)	-1,89	3,5721
(31 -25,89)	5,11	26,1121
(36 - 25,89)	10,11	102,2121
(42 - 25,89)	16,11	259,5321
(55 - 25,89)	29,11	847,3921
(96 - 25,89)	70,11	4.915,4121
Total		9.011,3508

$$\text{Desvio Padrão} = \sqrt{\frac{9.011,3508}{13}} = \sqrt{693,18} = 26,33$$

b) Mostre$\Sigma(x-\overline{x})$ é igual a zero.

$$(-25,79)+(-25,69)+(-25,64)+(-21,89)+(-13,89)+(-13,89)+(-1,89)+\\(-1,89)+5,11+10,11+16,11+29,11+70,11=0$$

10º) Em Massachusetts, oito indivíduos sofreram um episodio inexplicável de intoxicação por vitamina D que exigiu hospitalização; pensou-se que essas ocorrências extraordinárias pudessem resultar de uma excessiva suplementação de leite. Os níveis de cálcio e albumina - um tipo de proteína - no sangue para cada indivíduo no momento da internação no hospital são mostrados abaixo:

Cálcio (mmol/l)	**Albumina (g/l)**
2,92	43
3,84	42
2,37	42
2,99	40
2,67	42
3,17	38
3,74	34
3,44	42

→

Cálcio (mmol/l)	**Albumina (g/l)**
2,37	42
2,67	42
2,92	43
2,99	40
3,17	38
3,44	42
3,74	34
3,84	42

↓

Cálcio (mmol/l)	**Albumina (g/l)**
3,74	34
3,17	38
2,99	40
2,37	42
2,67	42
3,44	42
3,84	42
2,92	43

a) Obtenha a média, a mediana, o desvio-padrão e a amplitude dos níveis de cálcio registrados.

$$\text{Média} = \frac{2,37 + 2,67 + 2,92 + 2,99 + 3,17 + 3,44 + 3,74 + 3,84}{8} = \frac{25,14}{8} = 3,1425$$

$$\text{Mediana} = \frac{3,17 + 2,99}{2} = \frac{6,16}{2} = 3,08$$

$$\text{Desvio-padrão} = \sqrt{\frac{2,37^2 + 2,67^2 + 2,92^2 + 2,99^2 + 3,17^2 + 3,44^2 + 3,74^2 + 3,84^2}{8}}$$

$$\text{Desvio-padrão} = \sqrt{\frac{71,89}{8}} = 3$$

$$\text{Amplitude} = 3,84 - 2,37 = 1,47$$

b) Calcule a média, a mediana, o desvio-padrão e a amplitude para os dados de níveis de albumina.

$$\text{Média} = \frac{34+38+40+42+42+42+42+43}{8} = \frac{323}{8} = 40,375$$

$$\text{Mediana} = 42$$

$$\text{Desvio-padrão} = \sqrt{\frac{34^2+38^2+40^2+42^2+42^2+42^2+42^2+43^2}{8}} = \sqrt{\frac{13.105}{8}} = 40,47$$

$$\text{Amplitude} = 43 - 34 = 9$$

c) Para indivíduos saudáveis, o intervalo normal de valores de cálcio é de 2,12 até 2,74 mmol/l, enquanto o intervalo de níveis de albumina é de 32 até 55 g/l. Você acredita que os pacientes que sofreram intoxicação por vitamina D tinham níveis normais de cálcio e de albumina no sangue?

Todos os pacientes apresentavam níveis normais albumina no sangue, enquanto a ampla maioria tinha níveis de cálcio acima da taxa padrão, sugerindo que as internações por intoxicação por vitamina D podem ter uma relação causal com os níveis de cálcio no organismo.

11º) Considere os dados da Questão 10 e faça o que se pede:
a) Construa um intervalo de confiança unilateral de 95% - um limite inferior – para o nível médio verdadeiro de cálcio de indivíduos que sofreram a intoxicação de vitamina D

$$\text{Intervalo inferior} = \overline{x} - z * \frac{\sigma}{\sqrt{n}} = 3,1425 - 1,645 * \frac{3}{\sqrt{8}} = 3,1425 - 1,745 = 1,3975$$

b) Construa um intervalo de confiança unilateral inferior de 95% para o nível médio verdadeiro de albumina desse grupo.

$$\text{Intervalo superior} = \overline{x} + z * \frac{\sigma}{\sqrt{n}} = 40,375 + 1,645 * \frac{40,47}{\sqrt{8}} = 40,375 + 8,322 = 48,697$$

c) Para indivíduos saudáveis, o intervalo normal de valores de cálcio é de 2,12 a 2,74 mmol/l e o intervalo de níveis de albumina é de 32 a 55 g/l. Você acredita os pacientes que sofrem de intoxicação de vitamina D têm níveis normais de cálcio e de albumina no sangue? Os níveis albumina estão dentro de níveis considerados normais, enquanto os

níveis de cálcio ultrapassam a margem superior dos valores normais.

Sensibilidade e especificidade
12º) Um estudo registrou que a sensibilidade da mamografia como teste de triagem para

detecção do câncer de mama é 0,85, enquanto sua especificidade é 0,80.
a) Qual a probabilidade de um resultado de teste falso negativo?

$$\mathbb{P}(\text{Falso negativo}) = 1 - especifidade = 1 - 0,8 = 0,2$$

b) Qual a probabilidade de resultado falso positivo?

$$\mathbb{P}(\text{Falso positivo}) = 1 - sensibilidade = 1 - 0,85 = 0,15$$

c) Na população na qual a probabilidade de que uma mulher tenha câncer de mama é 0,0025, qual a probabilidade de que tenha câncer, se sua mamografia for positiva?

$$\mathbb{P}(\text{Câncer de mama}|\text{Mamografia positiva}) = \text{sensibilidade*prevalência}$$

$$\mathbb{P}(\text{Câncer de mama}|\text{Mamografia positiva}) = 0,85 * 0,0025 = 0,002125$$

13º) O Instituto Nacional de Segurança Ocupacional e de Saúde desenvolveu uma definição de casos de síndrome de túnel carpal – uma doença do punho – que incorpora três critérios: sintomas de envolvimento do nervo, história de fatores de risco ocupacional e a presença de materiais de exames físicos. A sensibilidade dessa definição como um teste para a síndrome de túnel carpal é de 0,67; sua especificidade é de 0,58.
a) Em uma população cuja prevalência da síndrome de túnel carpal esta estimada em 15%, qual o valor preditivo de um resultado positivo de teste?

$$\mathbb{P}(\text{Túnel do carpo}|\text{Teste positivo}) = \text{sensibilidade*prevalência}$$

$$\mathbb{P}(\text{Túnel do carpo}|\text{Teste positivo}) = 0,67 * 0,15 = 0,1005$$

b) Como esse valor preditivo se modifica, se a prevalência for somente 10%? E se for 5%?

Se o valor da prevalência passa a assumir o valor de 0,1, o valor preditivo diminui em 33%, se a prevalência assume o valor de 0,05, o valor preditivo diminui em 66%.

14º) Os dados seguintes são tomados de um estudo que investiga o uso de uma técnica chamada ventriculografia radiolinotípica como teste de diagnóstico para se detectar doença da artéria coronária.

Teste	**Doença**		**Total**
	Presente	**Ausente**	
Positivo	302	80	382
Negativo	179	372	551
Total	481	452	933

a) Qual a sensibilidade da ventrículografia radionuclídica nesse estudo? Qual a especificidade?

$$\text{Sensibilidade} = \frac{a}{a+c} = \frac{302}{302+179} = \frac{302}{481} = 0,63$$

$$\text{Especificidade} = \frac{d}{b+d} = \frac{372}{80+372} = \frac{372}{452} = 0,82$$

b) Para uma população cuja prevalência da doença da artéria coronária seja 0,10, calcule a probabilidade de que um indivíduo tenha a doença, sendo que ele apresenta resultado positivo usando a ventriculografia radionuclídica.

$$\mathbb{P}(\text{Doença artéria coronária}|\text{Teste positivo}) = \text{sensibilidade*prevalência}$$

$$\mathbb{P}(\text{Túnel do carpo}|\text{Teste positivo}) = 0,82 * 0,10 = 0,082$$

c) Qual o valor preditivo de um teste negativo.

$$\text{Valor preditivo negativo} = \frac{d}{c+d} = \frac{372}{179+372} = \frac{372}{551} = 0,675$$

Intervalo de confiança, teste de hipótese

15º) Os seguintes dados foram coletados para uma amostra de uma população normal: 10, 8, 12, 15, 13, 11, 6, 5.

a) Qual a estimativa pontual da média populacional?

$$\text{Média} = \frac{5+6+8+10+11+12+13+15}{8} = \frac{80}{8} = 10$$

b) Qual a estimativa pontual do desvio padrão da população?

$(x-\overline{x})$		$(x-\overline{x})^2$
(5 - 10)	-5	25
(6 - 10)	-4	16
(8 - 10)	-2	4
(10 - 10)	0	0
(11 - 10)	1	1
(12 - 10)	2	4
(13 - 10)	3	9
(15 - 10)	5	25
Total		84

$$\text{Desvio Padrão} = \sqrt{\frac{84}{8}} = \sqrt{10,5} = 3,24$$

c) Qual o intervalo de confiança com 95% para a média da população?

$$\text{Intervalo inferior} = \overline{x} - z * \frac{\sigma}{\sqrt{n}} = 10,5 - 1,96 * \frac{3,24}{\sqrt{8}} = 10,5 - 2,245 = 8,255$$

$$\text{Intervalo superior} = \overline{x} + z * \frac{\sigma}{\sqrt{n}} = 10,5 + 1,96 * \frac{3,24}{\sqrt{8}} = 10,5 + 2,245 = 12,745$$

16º) Considere o seguinte teste de hipótese.

$$H_0 : \mu \geq 10$$
$$H_a : \mu < 10$$

Uma amostra com n=50 fornece uma média de amostra de 9,46 e um desvio padrão da amostra de 2.
a) Com α = 0,05, qual o valor crítico para Z? Qual a regra de rejeição?

O valor crítico para a estatística Z unilateral com significância α = 0,05 é -1,645, pois a hipótese nula afirma que a média é maior ou igual ao valor de referência 10. Deste modo, a regra para rejeição da hipótese nula é obter um valor estatístico para média inferior a -1,645.

b) Calcule o valor da estatística de z. Qual a sua conclusão?

$$\text{z} = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{9,46 - 10}{\frac{2}{\sqrt{50}}} = \frac{-0,54}{\frac{2}{7,07}} = \frac{-0,54}{1} * \frac{7,07}{2} = \frac{-3,8178}{2} = -1,9089$$

Como o valor da estatística Z foi inferior a -1,645, pode-se rejeitar a hipótese nula com de 95% de acurácia.

Teste Qui-quadrado
17º) Os dados seguintes vêm de um estudo concebido para investigar problemas de bebidas entre os estudantes universitários. Em 1983, foi perguntado a um grupo se já dirigira um automóvel depois de beber. Em 1987, depois de atingida a idade legal para consumo de bebidas alcoólicas, foi feita a mesma questão a outro grupo universitário.

Dirigia enquanto bebia	**Ano**		
	1983	**1987**	**Total**
Sim	1.250	991	**2.241**
Não	1.387	1.666	**3.053**
Total	**2.637**	**2.657**	**5.294**

a) Use o teste qui-quadrado para avaliar a hipótese nula de que as proporções de estudantes da população que dirigia enquanto bebia são as mesmas nos dois anos.

Eventos:
X = Ano
Y = Dirigia bêbado

$$(\text{X=1983} \cap \text{Y=Sim}) = \frac{1.250}{5.294} = 0,2361$$

$$(\text{X=1983} \cap \text{Y=Não}) = \frac{1.387}{5.294} = 0,2620$$

$$(\text{X=1987} \cap \text{Y=Sim}) = \frac{991}{5.294} = 0,1872$$

$$(\text{X=1987} \cap \text{Y=Não}) = \frac{1.666}{5.294} = 0,3147$$

$\mathbb{P}(X = x)\mathbb{P}(Y = y)$	Valor esperado (E)
$\frac{2.241}{5.294} * \frac{2.637}{5.294}$	0,2109
$\frac{2.241}{5.294} * \frac{2.657}{5.294}$	0,2124
$\frac{3.053}{5.294} * \frac{2.637}{5.294}$	0,2873
$\frac{3.053}{5.294} * \frac{2.657}{5.294}$	0,2894

GL = (n - 1)*(m - 1) = (2-1)*(2-1) = 1*1 = 1

$$\text{X}^2 = \frac{\Sigma(O_i - E_i)^2}{E_i}$$

$$\frac{(0,2361 - 0,2109)^2}{0,2109} + \frac{(0,2620 - 0,2873)^2}{0,2873} + \frac{(0,1872 - 0,2124)^2}{0,2124} + \frac{(0,3147 - 0,2894)^2}{0,2894}$$

$$\text{X}^2 = \frac{0,0252^2}{0,2109} + \frac{(-0,0253)^2}{0,2873} + \frac{(-0,0252)^2}{0,2124} + \frac{0,0253^2}{0,2894}$$

$$\text{X}^2 = \frac{0,000635^2}{0,2109} + \frac{0,000640^2}{0,2873} + \frac{0,000635^2}{0,2124} + \frac{0,000640^2}{0,2894}$$

$$\text{X}^2 = 0,003011 + 0,002228 + 0,002990 + 0,002211$$

$$\text{X}^2 = 0,001044$$

valor-p ≈ 0,025

valor-p $< \alpha \rightarrow$ Rejeita H_0

b) O que você conclui sobre o comportamento desses estudantes?

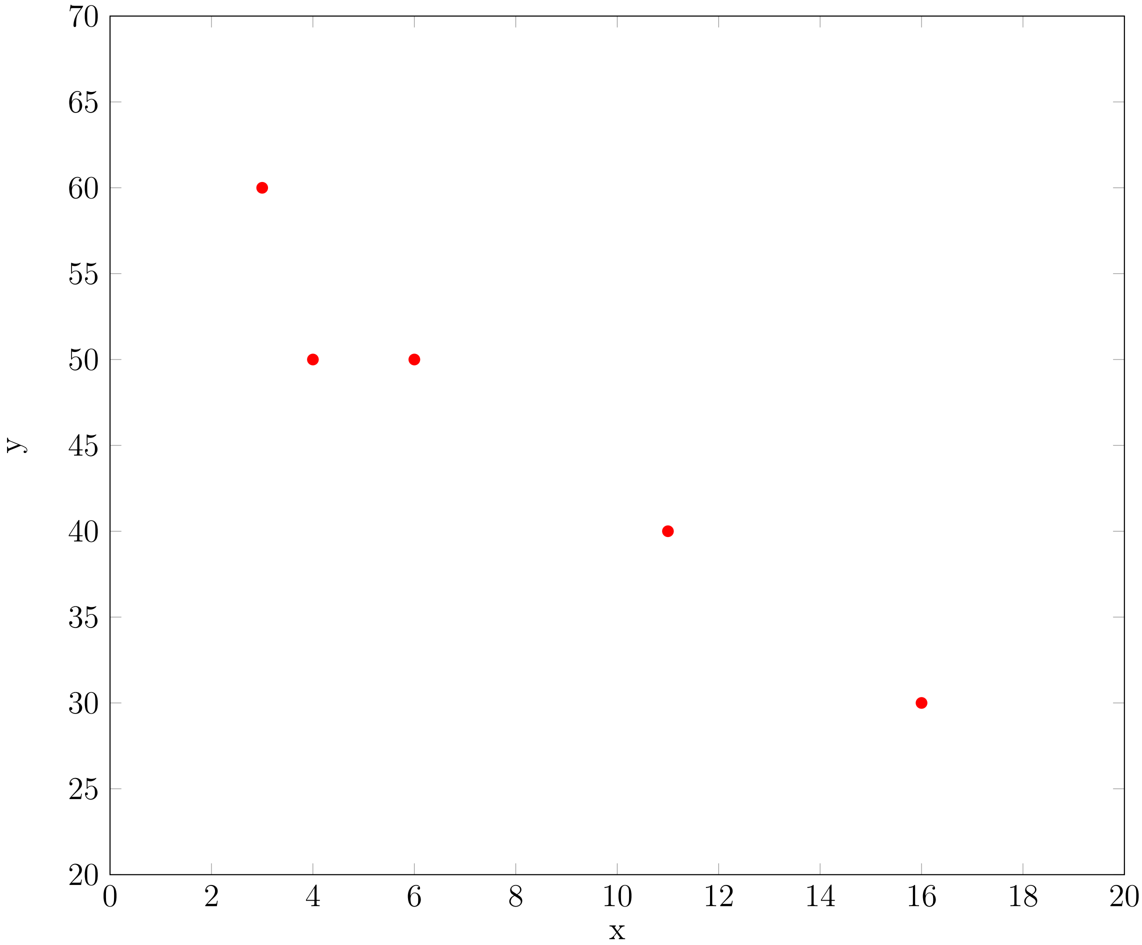
A partir da análise dos dados é possível concluir que existe uma associação entre a variável ano e a variável dirigia bêbado. O que nos faz rejeitar a hipótese H_0.

Correlação e Regressão

18º) Cinco observações tomadas para duas variáveis são apresentadas a seguir:

x_i	4	6	11	3	16
y_i	50	50	40	60	30

a) Desenvolva um diagrama de dispersão com x no eixo horizontal.



b) O que o diagrama de dispersão, desenvolvido no item (a) indica sobre a relação entre as duas variáveis?

Indica uma relação direta inversa, conforme os valores de x crescem os valores de y caem.

c) Calcule e interprete a covariância da amostra.

	x_i	y_i	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
	4	50	-4	4	16	16	-16
	6	50	-2	4	4	16	-8
	11	40	3	-6	9	36	-18
	3	60	-5	14	25	196	-70
	16	30	8	-16	64	256	-128
Σ	40	230			118	520	-240

$$\text{Cov}_{x,y} = \frac{\Sigma(x_i - \overline{x})(y_i - \overline{y})}{n-1} = \frac{-240}{5-1} = \frac{-240}{4} = -60$$

A covariância é negativa, o que significa que a dependência linear entre as variáveis tem sentido oposto.

d) Calcule e interprete o coeficiente de correlação da amostra.

$$\text{S}_x = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{118}{5-1}} = \sqrt{\frac{118}{4}} = \sqrt{29,5} = 5,43$$

$$\text{S}_y = \sqrt{\frac{\Sigma(y_i - \overline{y})^2}{n-1}} = \sqrt{\frac{520}{5-1}} = \sqrt{\frac{520}{4}} = \sqrt{130} = 11,40$$

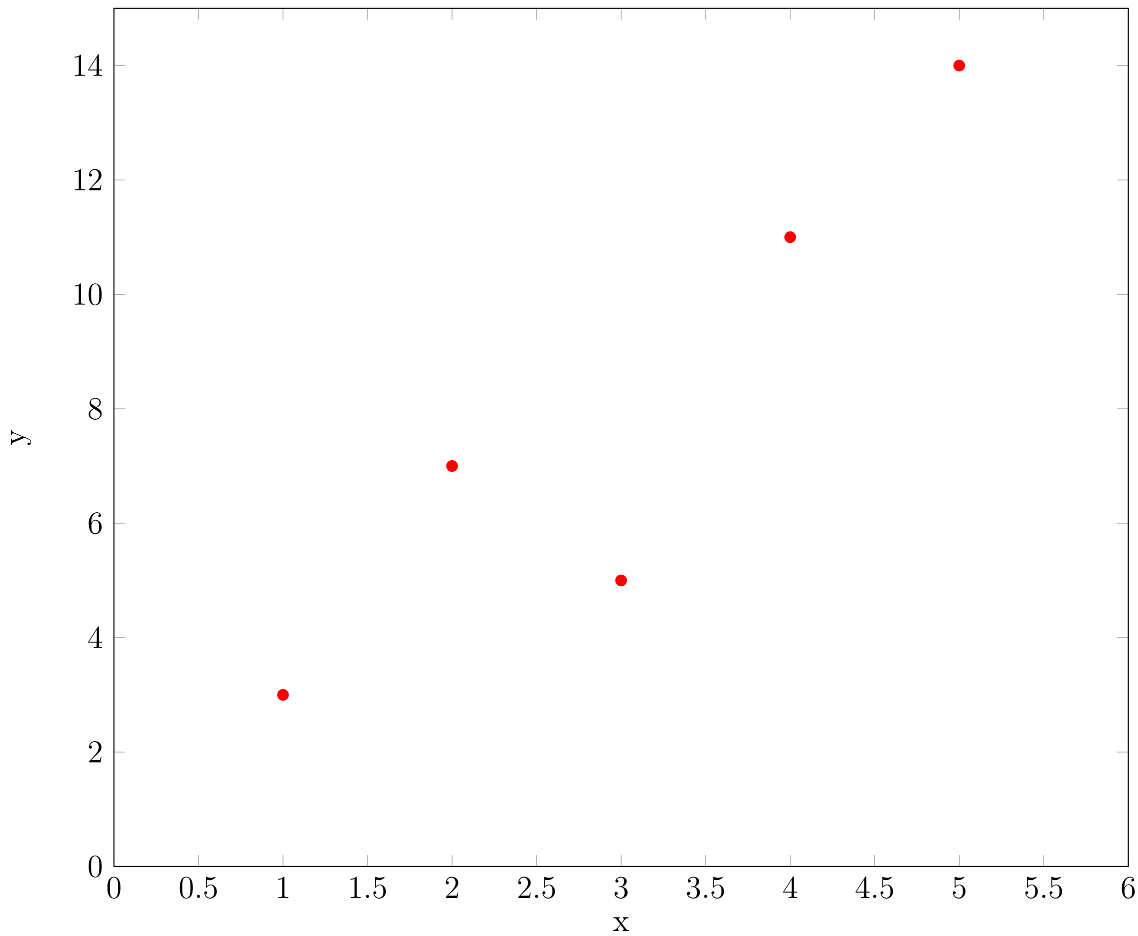
$$\text{r}_{x,y} = \frac{Cov_{x,y}}{S_x S_y} = \frac{-60}{5,43 * 11,40} = -0,9693$$

A variável preditiva x explica, aproximadamente, 96,93% da variação da variável resposta y, e por ter uma correlação negativa as variáveis tem suas variações em sentido inverso.

19º) Dada cinco observações para duas variáveis, x e y.

x_i	1	2	3	4	5
y_i	3	7	5	11	14

a) Desenvolva o diagrama de dispersão par esses dados.



b) O que o diagrama de dispersão, desenvolvido no item (a) indica sobre a relação entre as duas variáveis?

Indica uma relação direta, conforme os valores de x crescem os valores de y seguem o mesmo comportamento.

c) Desenvolva uma equação de regressão estimada calculando os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$ usando qualquer método de estimação e cite o método que usou.

Método dos mínimos quadrados (MMQ)

	x_i	y_i	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
	1	3	-2	-5	4	10
	2	7	-1	-1	1	1
	3	5	0	-3	0	0
	4	11	1	3	1	3
	5	14	2	6	4	12
Σ	15	40			10	26
Média	3	4				

$$\hat{\beta}_1 = \frac{\Sigma((x_i - \overline{x})(y_i - \overline{y}))}{\Sigma(x_i - \overline{x})^2} = \frac{26}{10} = 2,6$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 8 - 2,6 * 3 = 8 - 7,8 = 0,2$$

d) Use a equação de regressão estimada para determinar o valor de y quando x=4.

$$\hat{y} = 0,2 + 2,6 * \hat{x} = 0,2 + 2,6 * 4 = 0,2 + 10,4 = 10,6$$

Regressão logística

20º) Em um estudo que investiga os fatores de risco maternais para sífilis congênita, a doença é tratada como uma variável resposta dicotômica, na qual 1 representa a presença da doença em um recém nascido e 0 sua ausência. Os coeficientes estimados do modelo de regressão logística que contém as variáveis explicativas uso de cocaína ou crack, status marital, número de consultas pré-natais, uso de álcool e nível educacional estão listados abaixo. O intercepto não é dado.

Variável	**Coeficiente**
Uso de cocaína/crack	1,354
Status marital	0,779
Número de consultas pré-natais	-0,098
Uso de álcool	0,723
Nível educacional	0,298

a) Conforme o número de consulta pré-natais aumente, o que acontece com a probabilidade de uma criança nascer com sífilis?

A relação é inversa, quanto maior o número de consultas menores os índices de sífiles congênita em crianças recém-nascidas.

b) O status marital representa uma variável dicotômica, na qual o valor 1 indica que a mulher não é casada e 0 que ela é casada. Qual a chance relativa de que um recém nascido tenha sífilis para mães não casadas versus as mães casadas?

$$e^{\beta x} = e^{0,779} = 2,179$$

c) O uso de cocaína ou de crack também é uma variável aleatória dicotômica; o valor 1 indica que uma mulher usou drogas durante a gravidez e 0 que não usou. Qual a razão de chances estimada de uma criança nascer com sífilis para as mulheres que usaram cocaína ou crack versus as que não usaram?

$$e^{\beta x} = e^{1,354} = 3,873$$

Os filhos de mulheres que usaram cocaína/crack durante a gestação tem 3,873 mais chances de apresentarem sífilis congênita ao nascer do que os filhos de mulheres que não usaram.