

2 ANÁLISE DESCRITIVA UNI E BIDIMENSIONAL DE VARIÁVEIS QUANTITATIVAS

2.1 Abrindo arquivos (planilhas de trabalho)

Abra o arquivo de dados correspondente ao exemplo trabalhado no capítulo 1.

FILE → **OPEN WORKSHEET** → em **Examinar**, siga as orientações do seu professor
→ **Desc.mtw** → **Abrir** → **Ok**.

2.2 Síntese numérica de variáveis quantitativas

- a) Sintetize as variáveis quantitativas calculando as medidas usuais: média aritmética, mediana, desvio padrão, primeiro quartil (Q1), terceiro quartil (Q3), o mínimo e o máximo.

1. Escolha a opção **Stat** → **Basic Statistics** → **Display Descriptive Statistics**
2. Selecione para o quadro **Variables** as variáveis **Teste, Experiência e Vendas**.
3. Clique **Ok**.

Como não usamos a opção **Statistics** o Minitab forneceu sua escolha padrão:

N = número de dados;

N* = número de dados omissos (dados em branco);

Mean = média aritmética;

SE Mean = erro padrão estimado da média, que é o desvio padrão dividido pela raiz quadrada do tamanho da amostra. Esta medida é utilizada na inferência sobre a média populacional e será estudada posteriormente;

StDev = desvio padrão;

Minimum = mínimo;

Q1 = primeiro quartil;

Median = mediana;

Q3 = terceiro quartil;

Maximum = máximo.

Poderíamos ter escolhido outras medidas dentre uma lista de estatísticas disponíveis usando a opção **Statistics**, como a variância (**Variance**), o coeficiente de variação (**Coefficient of variation**), a amplitude interquartil (**Interquartile range**), a moda (**Mode**) etc.

- a1) Com as medidas calculadas, complete:

Quadro 2.1: Síntese numérica para as variáveis quantitativas

Medidas	Variáveis		
	Teste (nº questões certas)	Experiência (anos completos)	Vendas (salários mínimos)
Menor valor		1	16
Primeiro quartil		2	
Segundo quartil	6	3	
Terceiro quartil	8		
Maior valor		5	54
Média aritmética	6,067	2,667	
Desvio padrão	1,870	1,234	

- a2) O desvio padrão da variável vendas é de _____ salários mínimos.
- a3) Suponha que um vendedor seja considerado excepcional se seu volume médio de vendas está dois desvios padrões acima da média geral. Quanto ele deve vender para ser considerado excepcional? _____.
- a4) Os 50% que venderam menos, venderam de 16 a _____ salários mínimos.
- a5) O diretor de vendas anunciou que transferirá para outra praça todos os vendedores cujo volume de vendas for inferior ao 1º quartil da distribuição. Qual é o volume mínimo de vendas que um vendedor deve realizar para não ser transferido? _____.
- a6) A menor nota no teste foi _____ e a maior _____. Setenta e cinco por cento acertaram _____ ou mais questões.
- a7) Vinte e cinco por cento dos vendedores têm _____ ou mais anos de experiência.
- a8) Em média, os vendedores têm aproximadamente _____ anos de experiência.

2.3 Gráfico de pontos (Gráfico de dispersão unidimensional)

- b) Construa um gráfico de pontos para a variável Vendas.

1. Escolha na barra de ferramentas **Graph** → **Dotplot** (Gráfico de Pontos);
2. Mantenha a opção **Simple** em **One Y** → **Ok**;
3. Selecione para o quadro **Graph variables** a variável **Vendas**;
4. Clique **Ok**.

Complete:

- b1) O valor máximo de vendas foi de _____ salários mínimos. _____ vendedores tiveram um volume médio mensal de 24 salários mínimos. No intervalo de 36 a 48 salários mínimos, temos _____ vendedor(es).
- b2) Um vendedor foi considerado excepcional se vendeu em média mais de 53,06 salários mínimos no mês, como foi respondido no item (a3) da seção anterior. Qual é o vendedor excepcional? _____ (para responder esta questão, posicione o cursor no ponto correspondente a este vendedor no gráfico e leia o número da linha que o identifica).

2.4 Ramo-e-folhas

- c) Construa um Ramo-e-folhas para a variável Vendas, com um incremento de 5 salários mínimos.

1. Escolha na barra de ferramentas **Graph** → **Stem-and-Leaf**;
2. Selecione para o quadro **Graph variables** a variável **Vendas**;
3. Digite **5** no quadro **Increment** → **Ok**.

O Ramo-e-folhas ficará como a seguir, onde a primeira coluna apresenta uma contagem de elementos, a segunda os ramos e a terceira as folhas.

Stem-and-Leaf Display: Vendas

Stem-and-leaf of Vendas N = 15

Leaf Unit = 1,0

```

1   1   6
7   2  113444
(2) 2   79
6   3  002
3   3
3   4
3   4   8
2   5  04

```

A segunda coluna indica os ramos que no exemplo correspondem às dezenas. As folhas indicam as unidades, que foram ordenadas da menor para a maior, em cada ramo. O ramo 1 e a folha 6 correspondem ao valor de venda igual a 16 salários mínimos. A primeira coluna à esquerda dos ramos apresenta uma contagem das observações. O valor (2) entre parênteses indica o número de elementos no ramo que contém a mediana. A contagem nos ramos abaixo do ramo que contém a mediana indica o número de elementos até o valor correspondente à última folha de cada ramo (temos 7 vendedores que venderam 24 salários mínimos ou menos). A contagem nos ramos acima do ramo que contém a mediana indica o número de elementos maiores ou iguais ao valor correspondente à primeira folha de cada ramo (temos 6 vendedores que venderam 30 salários mínimos ou mais).


c1) Complete, analisando o Ramo-e-folhas construído:

- _____ venderam 24 salários mínimos ou menos.
- _____ venderam de 27 a 29 salários mínimos.
- _____ venderam 48 salários mínimos ou mais.
- _____ venderam 30 salários mínimos ou mais.

c2) A forma da distribuição dos valores da variável Vendas é _____.

c3) Compare os valores da média aritmética e da mediana. A mediana 27 salários mínimos está localizada no 3º ramo e a média _____ salários mínimos está localizada no _____ ramo.

d) Construa um Ramo-e-folhas para a variável Vendas, com um incremento de 10 salários mínimos.

1. Clique em  ou aperte **Ctrl+E** para editar a última caixa de diálogo;
2. Digite **10** no quadro **Increment** → **Ok**.

d1) Com base no gráfico, complete a tabela abaixo:

Tabela 2.1: Vendas, em salários mínimos.

Vendas	Frequência
10 — 20	
20 — 30	
30 — 40	
40 — 50	
50 — 60	
Total	

d2) Represente a tabela 2.1 por meio de um **histograma**. Para isso siga os passos:

1. Selecione **Graph** → **Histogram** → Escolha a opção **Simple** → **Ok**;
2. Selecione a variável **Vendas** para o quadro **Graph variables**;
3. Selecione **Labels** → **Data Labels** → Assinale **Use y-values labels** em **Label Type** → **Ok**;
4. Clique **Ok**.

O histograma não ficou da maneira desejada. Serão necessárias algumas modificações para que ele represente a tabela 2.1. Essas modificações serão feitas diretamente no gráfico.

d3) Modifique o histograma fazendo com que sejam exibidos os limites de classe (*cutpoints*) em vez dos pontos médios (*midpoints*). Inicie a primeira classe com o limite de 10 e faça a última classe exibir o limite superior de 60, com a largura de cada classe igual 10 salários mínimos. Para isso, com o gráfico pronto, siga as instruções abaixo:

1. Clique uma vez em cima de qualquer valor do eixo dos x's, com o botão direito do mouse para abrir a caixa de diálogo **Edit X Scale**;
2. Selecione **Binning**;
3. Em **Interval Type**, selecione **Cutpoint**; em **Intervals Definition**, selecione **Midpoint/cutpoint positions** e digite **10:60/10** no quadro correspondente;
4. Clique **Ok**.

Obs.: 10:60/10 é equivalente a digitar os limites de classe: 10 20 30 40 50 60.

Note que este gráfico fornece as mesmas informações que a tabela 2.1. O capítulo 3 tratará especificamente de histograma, sua construção, análise e aplicação no Controle de Qualidade Total (CQT).

2.5 Diagrama em Caixa - *Boxplot*

O *Boxplot* consiste em uma caixa usando os quartis (o primeiro, a mediana e o terceiro quartil) e

e) Faça um *Boxplot* para a variável **Vendas**.

1. Escolha na barra de ferramentas **Graph** → **Boxplot**;
2. Mantenha a opção de gráfico **Simple** em **One Y** → **Ok**;
3. Selecione a variável **Vendas** para o quadro **Graph variables**;
4. Selecione **Labels** e digite um título para o gráfico → **Ok** → **Ok**.

Responda as questões seguintes posicionando (não fique segurando) o cursor na parte que interessa do gráfico (na caixa ou nos pontos discrepantes) ou lendo o valor correspondente no eixo y:

- e1) Qual o valor do primeiro quartil? _____
- e2) Vinte e cinco por cento dos que venderam menos venderam de _____ a _____ salários mínimos.
- e3) Qual o valor da mediana? _____
- e4) Qual o valor do terceiro quartil? _____
- e5) Qual o valor da amplitude interquartil (*IQR*)? _____

- e6) Há algum ponto discrepante? _____ (Sim/Não)
- e7) Quais os valores de vendas considerados discrepantes? _____, _____ e _____ salários mínimos.
- e8) O ponto discrepante igual a 54 salários mínimos corresponde ao vendedor _____ da planilha de dados.
- e9) Apresente o gráfico na forma horizontal. Para isso, repita os passos 1 ao 4 do item (e). Selecione **Scale** → assinale **transpose value and category scales (em Axes and ticks)** → **Ok** → **Ok**.
- e10) Qual é a forma da distribuição? _____

2.6 Análise Bidimensional I (uma variável quantitativa e outra qualitativa)

Síntese numérica e gráfica

Os vendedores argumentam ao diretor que o critério de ser transferido devido às vendas menores que o primeiro quartil não é justo, pois há zona de vendas menos privilegiada. A quem você daria razão?

Para você responder este item, é necessário seguir os seguintes passos:

- f) Calcule as medidas descritivas, para a variável Vendas, de acordo com as categorias da variável Zona. Analise os resultados, levando em conta as medidas de posição e de variabilidade. Não esqueça de calcular os coeficientes de variação.

- Escolha a opção **Stat** na barra de ferramentas → **Basic Statistics** → **Display Descriptive Statistics**;
- Selecione para o quadro **Variables** a variável **Vendas**; em seguida, posicione o cursor no quadro **By variables (optional)** e selecione a variável **Zona**;
- Clique em **Statistics** e selecione as estatísticas: **Mean, Standard Deviation, Coefficient of variation** (coeficiente de variação), **Minimum, Maximum, First quartile, Median, Third quartile** e **Mode** → **Ok** → **Ok**.

f1) Complete:

Quadro 2.2: Sumário da variável Vendas, de acordo com a Zona para a qual o vendedor foi designado.

Medidas	Zona		
	Norte	Oeste	Sul
Menor valor		21	21
1º Quartil		22,5	22,5
Segundo quartil			
3º Quartil	41,5	31	49
Maior valor	54	32	50
Média aritmética	29,8		34,6
Moda	-		-
Desvio padrão	14,41	4,6	13,56
Coeficiente de variação	48,36%		

f2) Analisando o quadro acima, responda:

Em média, a zona pior para vendas é a _____.

Os valores observados foram mais homogêneos em torno da respectiva média na zona _____. Justifique _____.

Gráfico de pontos por categoria

g) Faça o gráfico de pontos para a variável **Vendas**, de acordo com a variável **Zona**.

1. Escolha na barra de ferramentas **Graph** → **Dotplot** → Escolha **With Groups** na opção **One Y** e clique **Ok**;
2. Selecione para o quadro **Graph Variables** a variável **Vendas**;
3. Clique no quadro correspondente a **Categorical variables for grouping** e selecione a variável **Zona** → **Ok**.

Analisando o gráfico, nota-se que na zona oeste os pontos estão mais concentrados nos valores menores de vendas e é a única zona que possui moda.

Diagrama em caixa por categoria

h) Faça um Desenho Esquemático (Boxplot) para a variável **Vendas** separado por **zona**.

1. Escolha na barra de ferramentas **Graph** → **Boxplot** → Escolha **With Groups** na opção **One Y** e clique **Ok**;
2. Selecione para o quadro **Graph Variables** a variável **Vendas**;
3. Clique no quadro correspondente a **Categorical variables for grouping** e selecione a variável **Zona** → **Ok**.

h1) Complete:

os valores das amplitudes interquartílicas são: _____, _____ e _____ para as zonas norte, oeste e sul, respectivamente. Vinte e cinco por cento das maiores vendas estão entre _____ e _____; _____ e _____; _____ e _____; nas zonas norte, oeste e sul, respectivamente.

h2) Analise o gráfico observando o que ocorreu com a zona oeste, em comparação com as outras zonas.

2.7 Análise Bidimensional II (variáveis quantitativas)

i) Qual das variáveis, Teste ou Experiência observada na admissão, é mais importante para julgar um futuro candidato ao emprego? Para responder esta questão, resolva os itens (i1), (i2), (i3) e (i4).

Diagrama de dispersão

i1) Faça um gráfico de dispersão para as variáveis: Teste e Vendas.

1. Escolha na barra de ferramentas **Graph** → **Scatterplot** → Selecione o tipo **Simple** e clique **Ok**;
2. Selecione para **Y** a variável **Vendas** e para **X** a variável **Teste** (na linha 1) → **Ok**.

Analise o gráfico:

Coeficiente de correlação linear

i2) Calcule o coeficiente de correlação linear entre as variáveis: Teste e Vendas.

1. Selecione a opção **Stat** na barra de ferramentas → **Basic Statistics** → **Correlation**;
2. Selecione para o quadro **Variables** as variáveis **Vendas e Teste**;
3. Desmarque a opção **Display p-values**;
4. Clique **Ok**.

O coeficiente de correlação entre Teste e Vendas é igual a: _____.

i3) Refaça os itens (i1) e (i2) para as variáveis Experiência e Vendas.

Analise o diagrama de dispersão:

O coeficiente de correlação entre Experiência e Vendas é igual a: _____.

i4) Comparando os itens anteriores, a variável _____ é mais adequada para explicar a variação nas vendas.

O coeficiente de correlação e o diagrama de dispersão serão estudados com mais detalhe no capítulo 4.