1) **(1,0 ponto) - Considere uma amostra que possui 12 trabalhadores e que revelou os seguintes salários recebidos por um mês de trabalho na empresa xyz: R$ 1.518,00, R$ 1.850,00, R$ 1.752,00, R$ 1.850,00, R$ 2.500,00, R$ 2.000,00, R$ 3450,00, R$ 12.000,00, R$ 3.050,00, R$ 4.400,00, R$ 5.154,25, R$ 6.247,00. Deste modo, elabore um gráfico Box Plot, informando os valores mínimo, primeiro quartil, segundo quartil, terceiro quartil e o valor máximo. Existem *outliers* na distribuição apresentada?**

Passo 1 – Ordenar os dados

1.518,00, 1.752,00, 1.850,00, 1.850,00, 2.000,00, 2.500,00, 3.050,00, 3.450,00, 4.400,00, 5.154,25, 6.247,00, 12.000,00

Passo 2 – Calcular a Mediana (Segundo Quartil – Q2)

A mediana é o valor central que divide o conjunto de dados em duas metades. Como temos 12 observações (um número par), a mediana será a média dos dois valores centrais (o 6º e o 7º).

$$Q2 = \frac{2500,00 + 3050,00}{2}$$
$$Q2 = 2.775,00$$

Passo 3 – Calcular o Primeiro Quartil – Q1

O primeiro quartil é a mediana da metade inferior dos dados. A metade inferior é composta pelos 6 primeiros valores.

Metade inferior = 1.518,00, 1.752,00, 1.850,00, 1.850,00, 2.000,00, 2.500,00

A metade inferior dos dados possui 6 elementos, logo a mediana corresponde a média dos dois valores centrais (3º e 4º)

$$Q1 = \frac{1850,00 + 1850,00}{2}$$
$$Q1 = 1.850,00$$

Passo 4 – Calcular o Terceiro Quartil – Q3

O terceiro quartil é a mediana da metade superior dos dados. A metade superior é composta pelos 6 últimos valores.

Metade superior = 3.050,00, 3.450,00, 4.400,00, 5.154,25, 6.247,00, 12.000,00

A metade superior dos dados possui 6 elementos, logo a mediana corresponde a média dos dois valores centrais (3º e 4º)

$$Q3 = \frac{4400,00 + 5154,25}{2}$$
$$Q3 = 4.777,13$$

Passo 5 – Amplitude Interquartil

A amplitude interquartil corresponde a diferença entre o terceiro e o primeiro quartil. Ela representa a faixa de valores onde se encontram os 50% centrais dos dados.

$$AIQ = Q3 - Q1$$
$$AIQ = 4.777,13 - 1.850,00$$
$$AIQ = 2.927,13$$

Passo 6 - Cálculo dos valores máximo e mínimo (Limites do Box Plot)

Os valores máximo e mínimo no diagrama de caixa são utilizados para identificar outliers.

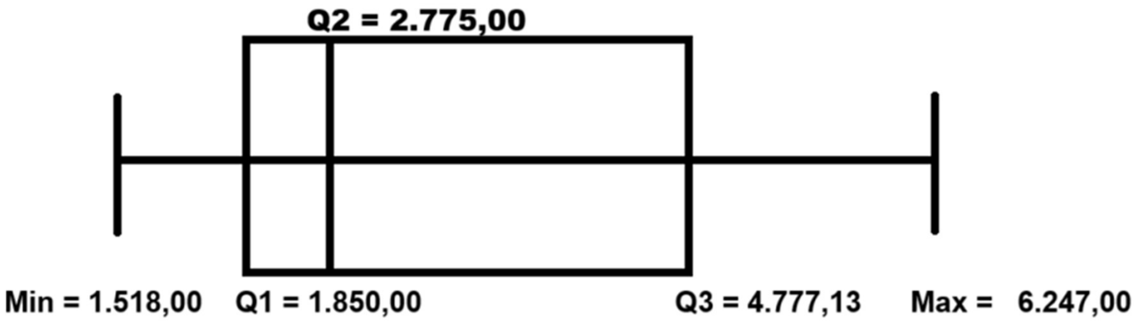
1. Valor mínimo (absoluto): É o menor valor da amostra (R$ 1.518,00).
2. Valor máximo (absoluto): É o maior valor da amostra (R$ 12.000,00)

Limite superior: Q3 + 1,5*AIQ = 4.777,13 + 1,5*2.927,13 = 9.167,83

Limite inferior: Q1 - 1,5*AIQ = 1.850,00 - 1,5*2.927,13 = 1.850,00 – 4.390,69=-2.540,69

O "máximo" da haste do gráfico será o maior valor da amostra que seja **menor ou igual a R$ 9.167,83**. Neste caso, é o **R$ 6.247,00**. O valor de R$ 12.000,00 está acima deste limite e é considerado um **outlier**.

O "mínimo" da haste será o menor valor da amostra que seja **maior ou igual a R$ -2.540,69**. Neste caso, é o próprio **R$ 1.518,00**.



2) **(0,75 pontos) - A passagem de 11 veículos por um radar eletrônico, em uma rodovia, registrou as velocidades apresentadas abaixo (em km/h).**

53	45	46	49	46	77	54	48	41	46	56

a) **Determine sua média, desvio padrão e o coeficiente de variação.**
A média é dada pela fórmula abaixo:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\mu = \frac{53+45+46+49+46+77+54+48+41+46+56}{11} = \frac{561}{11} = 51$$

A média de velocidade é de 51 km/h.

A variância é dada pela fórmula abaixo, considera-se a fórmula da população e não da amostra:

$$\sigma^2 = \frac{\left(\sum_{i=1}^{N} x_i - \mu\right)^2}{N}$$

Cálculo do quadrado das diferenças:

$$\begin{aligned}
&(53-51)^2 = 2^2 = 4 \\
&(45-51)^2 = (-6)^2 = 36 \\
&(46-51)^2 = (-5)^2 = 25 \\
&(49-51)^2 = (-2)^2 = 4 \\
&(46-51)^2 = (-5)^2 = 25 \\
&(77-51)^2 = 26^2 = 676 \\
&(54-51)^2 = 3^2 = 9 \\
&(48-51)^2 = (-3)^2 = 9 \\
&(41-51)^2 = (-10)^2 = 100 \\
&(46-51)^2 = (-5)^2 = 25 \\
&(56-51)^2 = 5^2 = 25
\end{aligned}$$

Deste modo tem-se:

$$\sigma^2 = \frac{938}{11} = 85,27$$

O desvio padrão é a raiz quadrada da variância, conforme apresentado abaixo:

$$\sigma = \sqrt{\frac{\left(\sum_{i=1}^{N} x_i - \mu\right)^2}{N}}$$

Portanto, o valor do desvio padrão é igual a:

$$\begin{aligned}
&\sigma = \sqrt{85,27} \\
&\sigma = 9,23
\end{aligned}$$

O cálculo do coeficiente de variação é dado pela fórmula abaixo:

$$CV = \left(\frac{\sigma}{\mu}\right) \cdot 100\%$$

Dado que temos os valores do desvio padrão e da média, obtém-se:

$$CV = \left(\frac{9,23}{51}\right) \cdot 100\% = 18,1\%$$

b) Se este radar eletrônico foi regulado dando um desconto de 5 km/h nas velocidades dos veículos, qual é a verdadeira média?

A questão afirma que o radar eletrônico foi regulado para ter um "desconto" de 5 km/h. Portanto, isso nos diz que a velocidade registrada é 5 km/h menor que a velocidade real do veículo.

Velocidade registrada = Velocidade verdadeira – 5 km/h

Deste modo, para obter-se a velocidade verdadeira é necessário somar 5 km/h a cada velocidade registrada.

Uma propriedade da média é que, se somarmos uma constante a todos os valores de um conjunto de dados, a nova média será a média original somada a essa constante.

Média verdadeira = Média registrada + 5 km/h

$$\mu_{verdadeira} = 51 + 5 = 56$$

3) **(0,75 pontos) - Em uma caixa há 6 bolas vermelhas, 4 bolas azuis e 2 bolas verdes. Se você retirar duas bolas da caixa, uma após a outra e sem reposição, qual a probabilidade de a primeira bola ser azul e a segunda ser vermelha?**

Este é um problema trabalha com probabilidade condicional, dado que a retirada da primeira bola afeta o total de bolas disponíveis para a segunda retirada.

I. Calcular total de bolas:

Total = 6 (vermelhas) + 4 (azuis) + 2 (verdes) = 12 bolas.

II. Probabilidade do 1º evento:

A probabilidade de a primeira bola retirada ser azul é o número de bolas azuis dividido pelo total de bolas.

$$P\left(1^{a}\, SerAzul\right)\frac{NúmeroDeBolasAzuis}{TotalDeBolas} = \frac{4}{12}$$

III. Probabilidade do segundo evento (dado que o primeiro já ocorreu):

Após retirar uma bola azul (**sem reposição**), a caixa agora tem um total de 11 bolas.

O número de bolas vermelhas permanece o mesmo (6).

A probabilidade de a segunda bola ser vermelha, *dado que a primeira foi azul*, é:

$$P\left(2^{a}\,SerVermelha \mid 1^{a}\,FoiAzul\right)\frac{NúmeroDeBolasVermelhas}{TotalDeBolas}=\frac{6}{11}$$

IV. Cálculo da probabilidade total:

Para encontrar a probabilidade de ambos os eventos ocorrerem em sequência, multiplicamos a probabilidade do primeiro evento pela probabilidade do segundo.

$$P=\frac{4}{12}\cdot\frac{6}{11}$$
$$P=\frac{24}{132}$$

Realizando a simplificação da fração acima, obtém-se:

$$P=\frac{2}{11}$$

4) **(0,50 pontos) -** Uma planilha Excel foi importada para o *software* Rstudio, armazenando dados de clientes em um data frame chamado **dados**. Abaixo é apresentado o data frame.

Nome	Categoria	Idade	Vendas_Total	Vendas_Ultimo_Mes	Ativo
Ana	A	28	150.50	50.00	Sim
Bruno	B	35	220.00		Sim
Carla	A	42	95.75	0.00	Não
Daniel	C	22	310.20	120.50	Sim
Eduarda	B	39	180.00	95.00	Sim
Fábio	A	51		75.00	Não

Analise as afirmações a seguir:

I. O comando **dados$Idade > 40** gerará um vetor lógico (TRUE/FALSE) que conterá exatamente dois valores TRUE.

II. A expressão **sum(dados$Vendas_Ultimo_Mes < 10)** retornará o valor 1, referente à linha de Carla, pois comparações com NA (o valor de Bruno) são tratadas como FALSE.

III. O comando **subset(dados, Ativo == "Sim" & Vendas_Total > 300)** retornará um data frame contendo apenas a linha de Daniel.

IV. O comando **dados$Status_Vendas <- ifelse(dados$Vendas_Total > 200, "Alto", "Baixo")** criará uma nova coluna onde o valor para a linha de Fábio será "Baixo".

V. O comando **mean(dados$Vendas_Total)** retornará o valor NA. Para calcular a média ignorando os valores ausentes, o comando correto seria **mean(dados$Vendas_Total, na.rm = TRUE)**.

As alternativas incorretas são:

a) I, II e III
b) III, IV e V
c) II, III e IV
d) II e IV
e) IV e V

I – Verdadeira

Justificativa: A expressão dados$Idade > 40 realiza uma comparação elemento por elemento na coluna Idade. Ao verificar os valores c(28, 35, 42, 22, 39, 51), verifica-se que Carla (42) e Fábio (51) são os únicos com idade maior que 40. Portanto, o vetor lógico resultante terá exatamente dois “valores” TRUE.

II – Falsa

Qualquer operação de comparação com NA, resulta em NA, não em FALSE. O vetor lógico da comparação será c(FALSE, NA, TRUE, FALSE, FALSE, FALSE). Como a função sum() por padrão não remove valores NA (na.rm = FALSE), a soma de um vetor que contém NA também resulta em NA, e não no número 1.

III – Verdadeira

A função subset filtra os dados com base nas condições fornecidas. A condição Ativo == “Sim” seleciona as linhas 1, 2, 4 e 5. Dessas linhas, a condição Vendas_Total > 300 é aplicada, e apenas a linha 4 (Daniel, como vendas de 310,20) satisfaz o critério. O resultado é um data frame com a linha de Daniel.

IV – Falsa

A função ifelse(teste, valor_sim, valor_não) avalia a condição teste. Para a linha de Fábio, o teste é NA > 200, que resulta em NA. A função ifelse propaga o NA do teste para o resultado. Deste modo, o valor na nova coluna Status_Vendas para Fábio será NA, e não “Baixo”.

V - Verdadeira

Por padrão, funções como mean(), sum(), max(), etc., retornam NA se houver qualquer valor NA nos dados. Para informar a função que deve ser ignorado tais valores ausentes antes de fazer o cálculo, o argumento na.rm = TRUE deve ser utilizado.

5) **(0,50 pontos) -** Na estatística as representações gráficas são ferramentas indispensáveis para visualização e interpretação de dados. Existem diferentes tipos de gráficos para sintetizar os dados, são eles: gráficos de barras, histograma, gráficos de linhas, setor e BoxPlot. Um aluno durante a aula de estatística recebeu 4 séries estatísticas de dados para analisar, são elas:

- S1: Evolução do consumo mensal de materiais;
- S2: Participação percentual de cada sócio no capital de uma empresa;
- S3: Quantidade de alunos de uma escola por faixa etária;
- S4: Comparação de distribuição das notas obtidas por três turmas diferentes (A, B e C) em uma mesma prova.

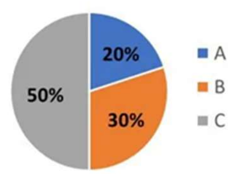
O tipo de gráfico mais adequado para representar essas séries, são:

a) S1: Gráfico de Linhas; S2: Gráfico de Setores; S3: BoxPlot; S4: Histograma.
b) S1: Gráfico de Linhas; S2: BoxPlot; S3: Histograma; S4: Gráfico de Barras.
c) S1: Gráfico de Linhas; S2: Gráfico de Barras; S3: Histograma; S4: Gráfico de Barras.
d) S1: Gráfico de Linhas; S2: Gráfico de Setores; S3: Histograma; S4: BoxPlot.
e) S1: Gráfico de Barras; S2: Gráfico de Setores; S3: Gráfico de Linhas; S4: Histograma.

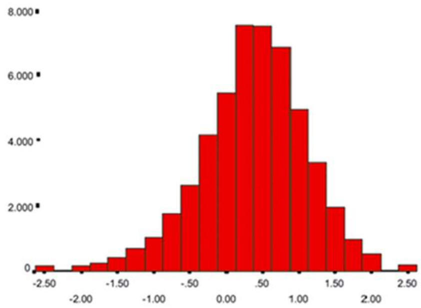
- **S1: Gráfico de Linhas:** Perfeito para "Evolução do consumo mensal". Gráficos de linha são a ferramenta padrão para mostrar a tendência de uma variável contínua ao longo do tempo.
- **S2: Gráfico de Setores:** Ideal para "Participação percentual". Este tipo de gráfico, também conhecido como gráfico de pizza, é projetado especificamente para ilustrar as proporções de diferentes categorias como partes de um todo (100% do capital da empresa).
- **S3: Histograma:** A escolha exata para "Quantidade de alunos por faixa etária". Um histograma representa a distribuição de frequência de uma variável numérica contínua (idade) que foi dividida em intervalos (as faixas etárias).
- **S4: BoxPlot:** A melhor opção para "Comparação de distribuição das notas". O BoxPlot é superior a outras ferramentas para comparar a distribuição de uma variável numérica (notas) entre diferentes grupos (as turmas), mostrando a mediana, a dispersão (intervalo interquartil) e possíveis outliers de cada turma lado a lado.

S1

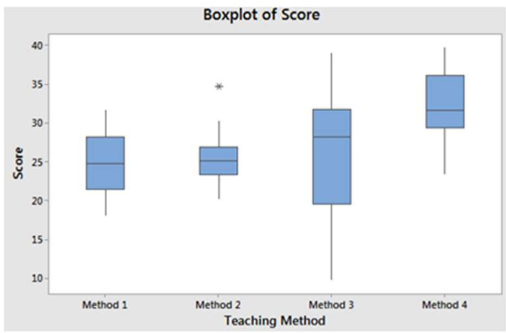
S2



S3



S4



6) **(0,2 pontos) -** Ao nascerem, os bebês são avaliados em diversas frentes para verificar seu estado de saúde. Considere as seguintes variáveis anotadas na ficha

de um recém-nascido: o peso em quilogramas (kg), a altura em centímetros (cm), o sexo e a nota do teste Apgar (um valor inteiro de 0 a 10).

A classificação correta para estas quatro variáveis, respectivamente, é:

I. Contínua, Contínua, Qualitativa Nominal e Quantitativa Discreta.
II. Contínua, Contínua, Qualitativa Ordinal e Quantitativa Contínua.
III. Discreta, Discreta, Qualitativa Nominal e Quantitativa Discreta.
IV. Contínua, Discreta, Qualitativa Nominal e Quantitativa Discreta.
V. Ambas quantitativas contínuas, e ambas qualitativas nominais.

1. Peso (em quilogramas)

- **Tipo:** Quantitativa Contínua.
- **Justificativa:** O peso é uma variável que resulta de uma **medição**. Ele pode assumir qualquer valor dentro de um intervalo, incluindo frações e decimais (ex: 3,450 kg, 2,985 kg). A precisão do valor é limitada apenas pela acurácia da balança. Por ser um número obtido por medição, é classificada como **contínua**.

2. Altura (em centímetros)

- **Tipo:** Quantitativa Contínua.
- **Justificativa:** Assim como o peso, a altura (ou comprimento) de um bebê é obtida através de uma **medição**. Ela também pode assumir qualquer valor em uma escala, como 49,5 cm ou 50,2 cm. Portanto, é uma variável **contínua**.

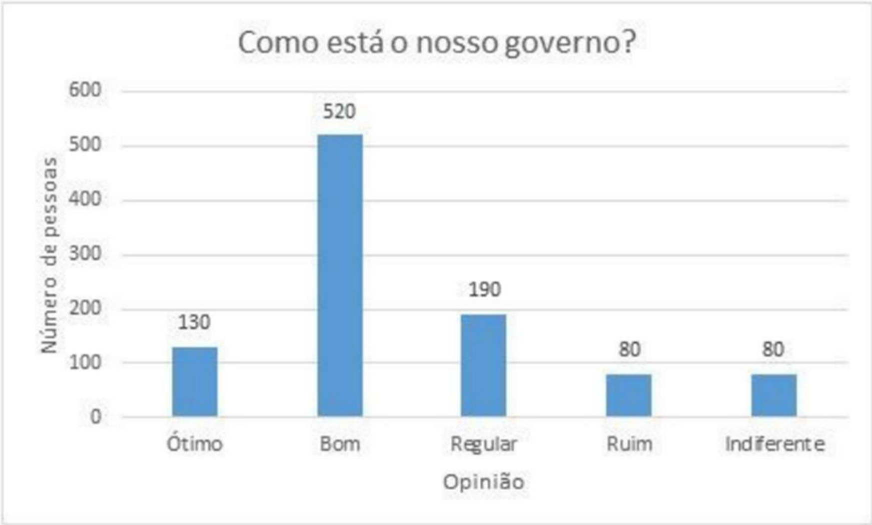
3. Sexo

- **Tipo:** Qualitativa Nominal.
- **Justificativa:** Esta variável descreve uma qualidade ou categoria ("Masculino" ou "Feminino"). Não existe uma ordem ou hierarquia natural entre essas categorias (nenhuma é intrinsecamente "maior" ou "menor" que a outra). Por se tratar de categorias sem uma ordem, é classificada como **nominal**.

4. Nota do Teste Apgar

- **Tipo:** Quantitativa Discreta.
- **Justificativa:** A nota Apgar é um valor numérico que resulta de uma **contagem** ou soma de pontos (de 0 a 10). Um bebê pode receber uma nota 7 ou 8, mas nunca uma nota 7,5. Como os valores possíveis são apenas números inteiros e específicos, resultantes de uma contagem, a variável é classificada como **discreta**.

7) Em uma pesquisa de opinião, feita para verificar o nível de aprovação de um governante, foram entrevistadas 1.000 pessoas, que responderam sobre a administração da cidade X, escolhendo uma, e apenas uma, dentre as possíveis respostas: Ótimo, Bom, Regular, Ruim e Indiferente. O gráfico abaixo, apresenta o resultado da pesquisa.



De acordo com o gráfico, pode-se afirmar que o percentual de pessoas que consideram a administração ótima, boa ou regular é igual a:

a) 56%
b) 84%
c) 80%
d) 64%
e) 78%

8) Em uma universidade com 2000 estudantes, foi realizada uma pesquisa sobre a participação em atividades extracurriculares. Os dados coletados são os seguintes:

- 800 estudantes participam de alguma equipe esportiva (Evento E);
- 600 estudantes são membros de clubes acadêmicos (Evento C);
- 240 estudantes participam tanto de equipes esportivas quanto de clubes acadêmicos (Evento $E \cap C$).

Com base nessas informações, analise as seguintes afirmações e classifique-as como **Verdadeira (V)** ou **Falsa (F)**, justificando cada resposta com os cálculos apropriados.

1. A probabilidade de um estudante, selecionado aleatoriamente, ser membro de um clube acadêmico é de 0,4;
2. Os eventos "participar de uma equipe esportiva" e "ser membro de um clube acadêmico" são eventos independentes;
3. A probabilidade de um estudante ser membro de um clube acadêmico, dado que ele já participa de uma equipe esportiva, é de 0,3.

Afirmativa 1: INCORRETA

probabilidade de um estudante, selecionado aleatoriamente, ser membro de um clube acadêmico é de 0,4;

Está afirmação testa o cálculo básico da probabilidade de um evento.

A probabilidade de um estudante ser membro de um clube acadêmico é:

$$P(C) = \frac{NúmeroDeMembrosDeC\,\text{lub}\,es}{TotalDeEstudantes} = \frac{600}{2000} = 0,3$$

Afirmativa 2: CORRETA

Os eventos "participar de uma equipe esportiva" e "ser membro de um clube acadêmico" são eventos independentes;

Para que dois eventos, E e C, sejam independentes, a probabilidade de ambos ocorrerem deve ser igual ao produto de suas probabilidades individuais. A condição de independência é:

$$P(E \cap C) = P(E) \cdot P(C)$$

A probabilidade de participar de uma equipe esportiva (P(E)) é igual a:

$$P(E) = \frac{NúmeroDeAtletas}{TotalDeEstudantes} = \frac{800}{2000} = 0,4$$

Deste modo, tem-se que:

$$P(E \cap C) = P(E) \cdot P(C)$$

$$P(E \cap C) = 0,4 \cdot 0,3$$
$$P(E \cap C) = 0,12$$

Os eventos são, por definição, independentes. A participação em esportes não afeta a probabilidade de um estudante participar de um clube acadêmico, e vice-versa.

Afirmativa 3: CORRETA

A probabilidade de um estudante ser membro de um clube acadêmico, dado que ele já participa de uma equipe esportiva, é de 0,3.

Esta afirmação envolve o conceito de probabilidade condicional. Queremos encontrar a probabilidade do evento C, sabendo que o evento E já ocorreu, ou seja, $P(C \mid E)$.

A fórmula da probabilidade condicional é:

$$P(C \mid E) = \frac{P(E \cap C)}{P(E)}$$

$$P(C \mid E) = \frac{0,12}{0,4} = 0,3$$

O resultado do cálculo corresponde ao valor da afirmação.