

Fátima Vilela Candal - RM563003

Levi Passos Silveira Marques - RM565557

Resposta para a Atividade sobre Agronegócio

Com base nas fontes de dados públicas fornecidas, realizamos uma pesquisa e criamos uma base de dados simulada em Excel para ilustrar os conceitos. Em seguida, apresentamos a análise exploratória em R.

1. Pesquisa em Fontes de Dados Públicas:

Realizamos uma navegação nos sites da CONAB, IBGE, MAPA, Embrapa, INPE e CNA, buscando informações relevantes sobre a produção agrícola, pecuária, dados de mercado, pesquisas e tecnologias no setor do agronegócio brasileiro.

2. Base de Dados em Excel (Simulada):

Criamos uma planilha no Excel com 30 linhas e 4 colunas, exemplificando os tipos de variáveis solicitados. Os dados são fictícios e servem apenas para demonstrar a estrutura da base de dados.

Ano	Área Plantada (hectares)	Tipo de Cultivo	Nível de Tecnologia
2020	1.500,00	Soja	Alto
2021	1.650,50	Milho	Médio
2022	1.400,00	Cana-de-açúcar	Alto
2023	1.700,80	Algodão	Alto
2024	1.550,00	Café	Médio
2020	800,00	Feijão	Baixo
2021	950,20	Arroz	Médio
2022	1.200,00	Soja	Alto
2023	1.350,70	Milho	Médio
2024	1.100,00	Cana-de-açúcar	Alto
2020	600,00	Algodão	Baixo
2021	750,90	Café	Médio
2022	1.000,00	Feijão	Médio
2023	1.150,30	Arroz	Alto
2024	900,00	Soja	Médio
2020	1.800,00	Milho	Alto
2021	1.950,60	Cana-de-açúcar	Alto
2022	1.600,00	Algodão	Médio
2023	1.750,10	Café	Alto
2024	1.450,00	Feijão	Médio
2020	1.050,00	Arroz	Baixo
2021	1.200,40	Soja	Alto
2022	1.400,00	Milho	Médio

2023	1.550,90	Cana-de-açúcar	Alto
2024	1.300,00	Algodão	Médio
2020	700,00	Café	Baixo
2021	850,50	Feijão	Médio
2022	1.100,00	Arroz	Alto
2023	1.250,20	Soja	Médio
2024	1.000,00	Milho	Alto

Descrição das Colunas:

- **Ano:** Variável quantitativa discreta (representa anos).
- **Área Plantada (hectares):** Variável quantitativa contínua (representa a área em hectares).
- **Tipo de Cultivo:** Variável qualitativa nominal (representa diferentes tipos de culturas).
- **Nível de Tecnologia:** Variável qualitativa ordinal (representa níveis de tecnologia em ordem: Baixo, Médio, Alto).

O arquivo Excel com esta base de dados simulada será entregue separadamente.

3. Análise Exploratória de Variável Quantitativa em R:

```
# FatimaCandal_RM563003_fase2_cap7
```

```
# LeviMarques_RM565557_fase2_cap7
```

```
#instalacao de pacote
```

```
#install.packages("readxl")
```

```
#comando setwd, determina um diretorio de trabalho.
```

```
setwd("C:/FIAP")
```

```
# Carregar os dados (caminho do arquivo Excel)
```

```
library(readxl)
```

```
dados <- read_excel("agro.xlsx")
```

```
# Escolher a variável quantitativa contínua: Área Plantada (hectares)
```

```
area_plantada <- dados$`Área Plantada (hectares)`
```

```
# Medidas de Tendência Central
```

```
media_area <- mean(area_plantada)
```

```
mediana_area <- median(area_plantada)
```

```
moda_area <- names(sort(table(area_plantada), decreasing = TRUE)[1]) #  
Aproximação da moda para contínuas
```

```
cat("Medidas de Tendência Central para Área Plantada (hectares):\n")
```

```
cat("Média:", media_area, "\n")
```

```
cat("Mediana:", mediana_area, "\n")
```

```
cat("Moda (aproximada):", moda_area, "\n\n")
```

```
# Medidas de Dispersão
```

```
desvio_padrao_area <- sd(area_plantada)
```

```
variancia_area <- var(area_plantada)
```

```
amplitude_area <- max(area_plantada) - min(area_plantada)
```

```
iqr_area <- IQR(area_plantada)
```

```
cat("Medidas de Dispersão para Área Plantada (hectares):\n")
```

```
cat("Desvio Padrão:", desvio_padrao_area, "\n")
```

```
cat("Variância:", variancia_area, "\n")
```

```
cat("Amplitude:", amplitude_area, "\n")
```

```
cat("Intervalo Interquartil (IQR):", iqr_area, "\n\n")
```

```
# Medidas Separatrizes
```

```
quartis_area <- quantile(area_plantada, probs = c(0.25, 0.5, 0.75))
```

```
percentis_area <- quantile(area_plantada, probs = c(0.10, 0.50, 0.90))
```

```
cat("Medidas Separatrizes para Área Plantada (hectares):\n")
```

```
cat("Quartis:\n")
```

```
print(quartis_area)
```

```
cat("\nPercentis (10%, 50%, 90%):\n")
```

```
print(percentis_area)
```

```
cat("\n")
```

```
# Análise Gráfica
```

```
hist(area_plantada, main = "Histograma da Área Plantada (hectares)",
```

```
      xlab = "Área Plantada (hectares)", ylab = "Frequência", col = "lightblue")
```

```
boxplot(area_plantada, main = "Boxplot da Área Plantada (hectares)",
```

```
        ylab = "Área Plantada (hectares)", col = "lightgreen")
```

```
# Adicionar uma linha para a média no histograma
```

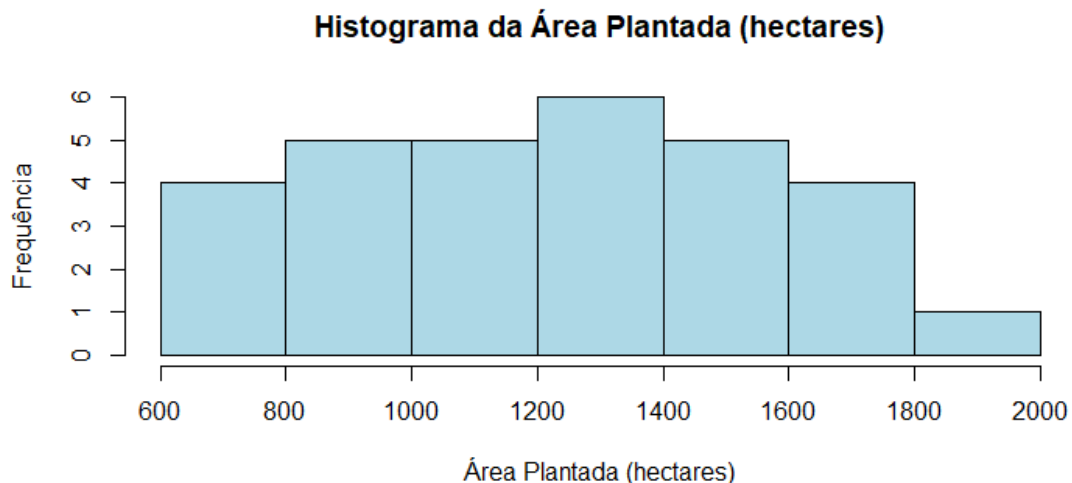
```
abline(v = media_area, col = "red", lty = "dashed", lwd = 2)
```

```
legend("topright", legend = paste("Média =", round(media_area, 2)),
```

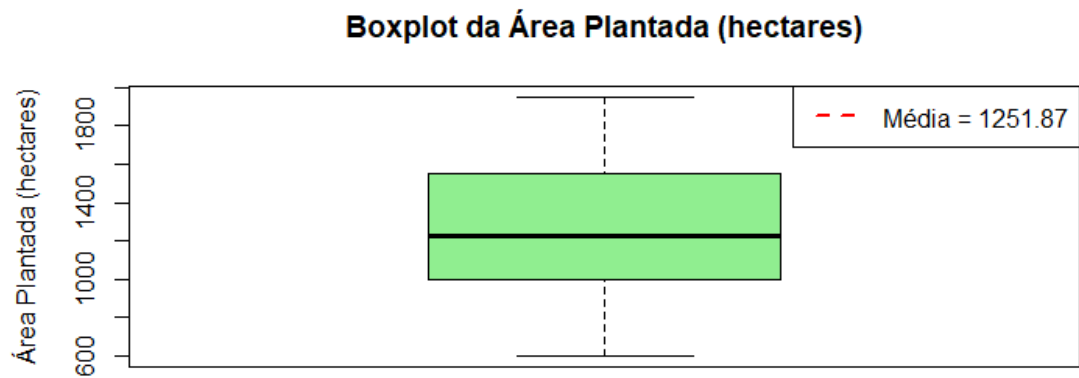
```
      col = "red", lty = "dashed", lwd = 2)
```

Interpretação da Análise da Área Plantada:

- **Medidas de Tendência Central:** A média e a mediana da área plantada fornecem uma ideia do valor central dos dados. A média indica o valor médio da área plantada, enquanto a mediana indica o valor central quando os dados são ordenados. A moda, para dados contínuos, é uma aproximação do valor mais frequente.
- **Medidas de Dispersão:** O desvio padrão e a variância medem a dispersão dos dados em torno da média. Uma maior dispersão indica maior variabilidade na área plantada. A amplitude mostra a diferença entre a maior e a menor área plantada, e o IQR representa a amplitude dos 50% centrais dos dados, sendo menos sensível a outliers.
- **Medidas Separatrizes:** Os quartis dividem os dados em quatro partes iguais, e os percentis em cem partes iguais. Eles ajudam a entender a distribuição dos dados e identificar valores abaixo ou acima de determinados limiares.
- **Análise Gráfica:**
 - O **histograma** mostra a distribuição da frequência da área plantada em diferentes intervalos. A forma do histograma pode indicar se a distribuição é simétrica, assimétrica, etc.



- O **boxplot** resume a distribuição dos dados, mostrando a mediana (linha central na caixa), os quartis (bordas da caixa), os bigodes (que se estendem até um certo limite) e possíveis outliers (pontos fora dos bigodes).



4. Análise Gráfica de Variável Qualitativa em R:

```
# FatimaCandal_RM563003_fase2_cap7
```

```
# LeviMarques_RM565557_fase2_cap7
```

```
#instalacao de pacote
```

```
#install.packages("readxl")
```

```
#comando setwd, determina um diretorio de trabalho.
```

```
setwd("C:/FIAP")
```

```
# Carregar os dados (se ainda não estiverem carregados)
```

```
library(readxl)
```

```
dados <- read_excel("agro.xlsx")
```

```
# Escolher a variável qualitativa nominal: Tipo de Cultivo
```

```
tipo_cultivo <- dados$`Tipo de Cultivo`
```

```
# Criar tabela de frequência
```

```
tabela_frequencia <- table(tipo_cultivo)
```

```
# Análise Gráfica: Gráfico de Barras
```

```
barplot(tabela_frequencia, main = "Distribuição de Tipos de Cultivo",
```

```
       xlab = "Tipo de Cultivo", ylab = "Frequência", col =  
rainbow(length(tabela_frequencia)))
```

```
# Adicionar rótulos com as frequências
```

```
text(x = 1:length(tabela_frequencia), y = tabela_frequencia + 1, labels =  
tabela_frequencia, pos = 3)
```

```
# Escolher a variável qualitativa ordinal: Nível de Tecnologia
```

```
nivel_tecnologia <- factor(dados$`Nível de Tecnologia`, levels = c("Baixo",  
"Médio", "Alto"), ordered = TRUE)
```

```
# Criar tabela de frequência
```

```
tabela_frequencia_tec <- table(nivel_tecnologia)
```

```
# Análise Gráfica: Gráfico de Barras (ordenado)
```

```
barplot(tabela_frequencia_tec, main = "Distribuição de Níveis de Tecnologia",
```

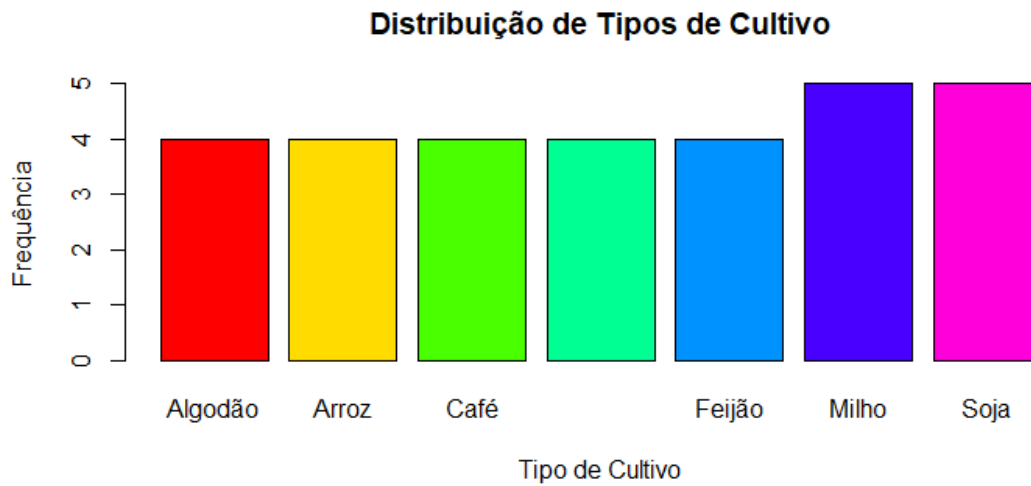
```
       xlab = "Nível de Tecnologia", ylab = "Frequência", col = c("lightcoral",  
"lightskyblue", "lightsalmon"))
```

```
# Adicionar rótulos com as frequências
```

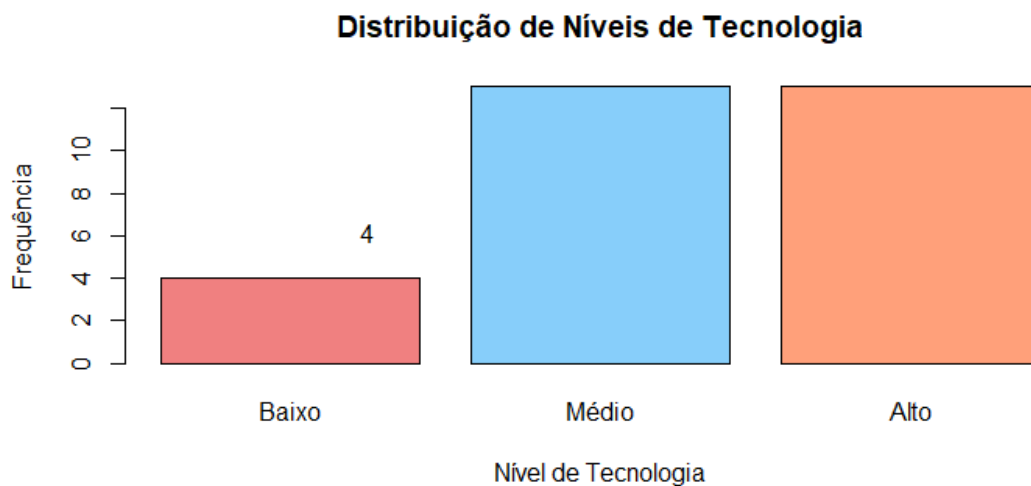
```
text(x = 1:length(tabela_frequencia_tec), y = tabela_frequencia_tec + 1, labels =  
tabela_frequencia_tec, pos = 3)
```

Interpretação da Análise dos Tipos de Cultivo e Níveis de Tecnologia:

- **Gráfico de Barras (Tipo de Cultivo):** Este gráfico exibe a frequência de cada tipo de cultivo presente na base de dados. A altura de cada barra representa o número de ocorrências de cada categoria.



- **Gráfico de Barras (Nível de Tecnologia):** Similar ao anterior, este gráfico mostra a frequência de cada nível de tecnologia utilizado. A ordem das barras reflete a ordem natural da variável ordinal (Baixo, Médio, Alto).

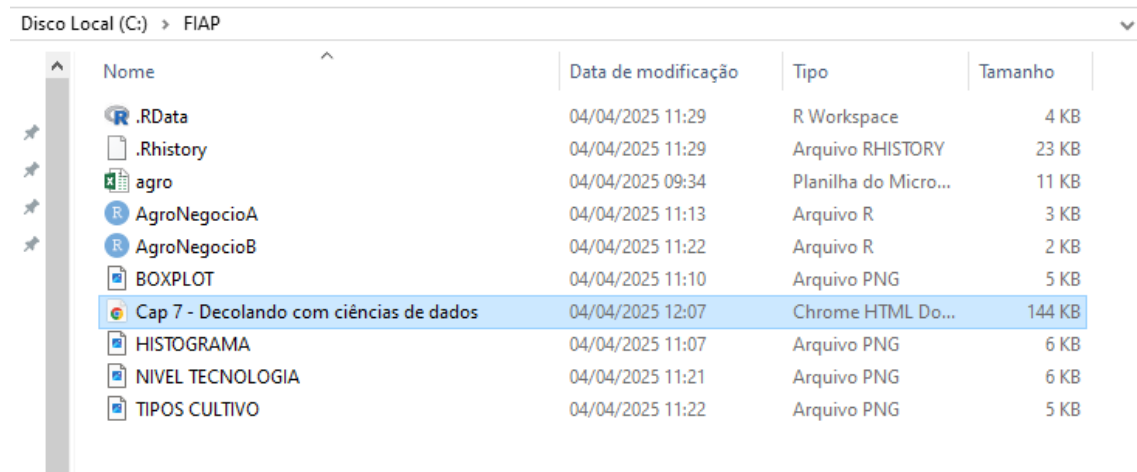


5. Os arquivos (Excel e R) serão entregues separadamente, conforme solicitado:

- Arquivo Excel com a base de dados;
- Arquivo R com os códigos utilizados, contendo na primeira linha do código

FatimaCandal_RM563003_fase2_cap7

LeviMarques_RM565557_fase2_cap7



Nome	Data de modificação	Tipo	Tamanho
.RData	04/04/2025 11:29	R Workspace	4 KB
.Rhistory	04/04/2025 11:29	Arquivo RHISTORY	23 KB
agro	04/04/2025 09:34	Planilha do Micro...	11 KB
AgroNegocioA	04/04/2025 11:13	Arquivo R	3 KB
AgroNegocioB	04/04/2025 11:22	Arquivo R	2 KB
BOXPLOT	04/04/2025 11:10	Arquivo PNG	5 KB
Cap 7 - Decolando com ciências de dados	04/04/2025 12:07	Chrome HTML Do...	144 KB
HISTOGRAMA	04/04/2025 11:07	Arquivo PNG	6 KB
NIVEL TECNOLOGIA	04/04/2025 11:21	Arquivo PNG	6 KB
TIPOS CULTIVO	04/04/2025 11:22	Arquivo PNG	5 KB