Pesquisa de mercado e de clientes do Faroeste

Consultores Responsáveis:
Davi Folha Desseaux

Requerente:
João Sábio

Brasília, 9 de novembro de 2025.

ESTAT
consultoria estatística

Sumário

1 Introdução

O seguinte projeto tem como objetivo realizar análises estatísticas que visam compreender melhor o mercado e comércio no Faroeste. Foram feitos estudos sobre receita média por intervalo de tempo, variações de peso e altura e suas possíveis relações e correlações, idades e suas distribuições e produtos mais vendidos nas lojas com maior receita. Essas análises foram realizadas com objetivo de obter soluções embasadas estatisticamente e que agreguem valor à tomada de decisão do cliente. Além disso, gráficos, tabelas e quadros que ajudem na visualização de tudo que foi citado. O nível de significância utilizado será de 5%.

A base de dados representa uma amostra não probabilística por conveniência, composta pelos registros disponíveis no sistema da empresa. Embora não tenha sido realizada uma seleção aleatória, o conjunto de dados é suficientemente amplo e variado, o que permite análises representativas do perfil dos clientes. O banco de dados foi cedido pelo próprio cliente e abrange variáveis quantitativas e qualitativas, sendo as principais utilizadas idade, peso, altura, nome e id de clientes; nome e ID de cidades e lojas; nome, ID e preço dos produtos; data, ID, loja, cliente e quantidade de cada venda.

O software utilizado para análise estatística dos dados foi o R versão 4.4.3. O R é um software de programação gratuito largamente usado na área de estatística e visualização de dados que permite não só o manuseio e análise de bancos de dados, como também a confecção de gráficos, quadros e testes estatísticos.

2 Referencial Teórico

Este relatório é composto por técnicas estatísticas que serão descristas a seguir de acordo com o que foi utilizado em tal estudo.

2.1 Frequência Relativa

A frequência relativa é utilizada para a comparação entre classes de uma variável categórica com c categorias, ou para comparar uma mesma categoria em diferentes estudos.

A frequência relativa da categoria j é dada por:

$$f_j = \frac{n_j}{n}$$

Com:

- $j = 1, ..., c$
- $n_j =$ número de observações da categoria j
- $n =$ número total de observações

Geralmente, a frequência relativa é utilizada em porcentagem, dada por:

$$100 \times f_j$$

2.2 Média

A média é a soma das observações dividida pelo número total delas, dada pela fórmula:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Com:

- $i = 1, 2, ..., n$
- $n =$ número total de observações

2.3 Mediana

Sejam as n observações de um conjunto de dados $X = X_{(1)}, X_{(2)}, \dots, X_{(n)}$ de determinada variável ordenadas de forma crescente. A mediana do conjunto de dados X é o valor que deixa metade das observações abaixo dela e metade dos dados acima.

Com isso, pode-se calcular a mediana da seguinte forma:

$$med(X) = \begin{cases} X_{\frac{n+1}{2}}, \text{para n ímpar} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}, \text{para n par} \end{cases}$$

2.4 Quartis

Os quartis são separatrizes que dividem o conjunto de dados em quatro partes iguais. O primeiro quartil (ou inferior) delimita os 25% menores valores, o segundo representa a mediana, e o terceiro delimita os 25% maiores valores. Inicialmente deve-se calcular a posição do quartil:

- Posição do primeiro quartil P_1:

$$P_1 = \frac{n+1}{4}$$

- Posição da mediana (segundo quartil) P_2:

$$P_2 = \frac{n+1}{2}$$

- Posição do terceiro quartil P_3:

$$P_3 = \frac{3 \times (n+1)}{4}$$

Com n sendo o tamanho da amostra. Dessa forma, $X_{(P_i)}$ é o valor do i-ésimo quartil, onde $X_{(j)}$ representa a j-ésima observação dos dados ordenados.

Se o cálculo da posição resultar em uma fração, deve-se fazer a média entre o valor que está na posição do inteiro anterior e do seguinte ao da posição.

2.5 Variância

A variância é uma medida que avalia o quanto os dados estão dispersos em relação à média, em uma escala ao quadrado da escala dos dados.

2.5.1 Variância Populacional

Para uma população, a variância é dada por:

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

Com:

- $X_i = i$-ésima observação da população
- $\mu =$ média populacional
- $N =$ tamanho da população

2.6 Desvio Padrão

O desvio padrão é a raiz quadrada da variância. Ele avalia o quanto os dados estão dispersos em relação à média.

2.6.1 Desvio Padrão Populacional

Para uma população, o desvio padrão é dado por:

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

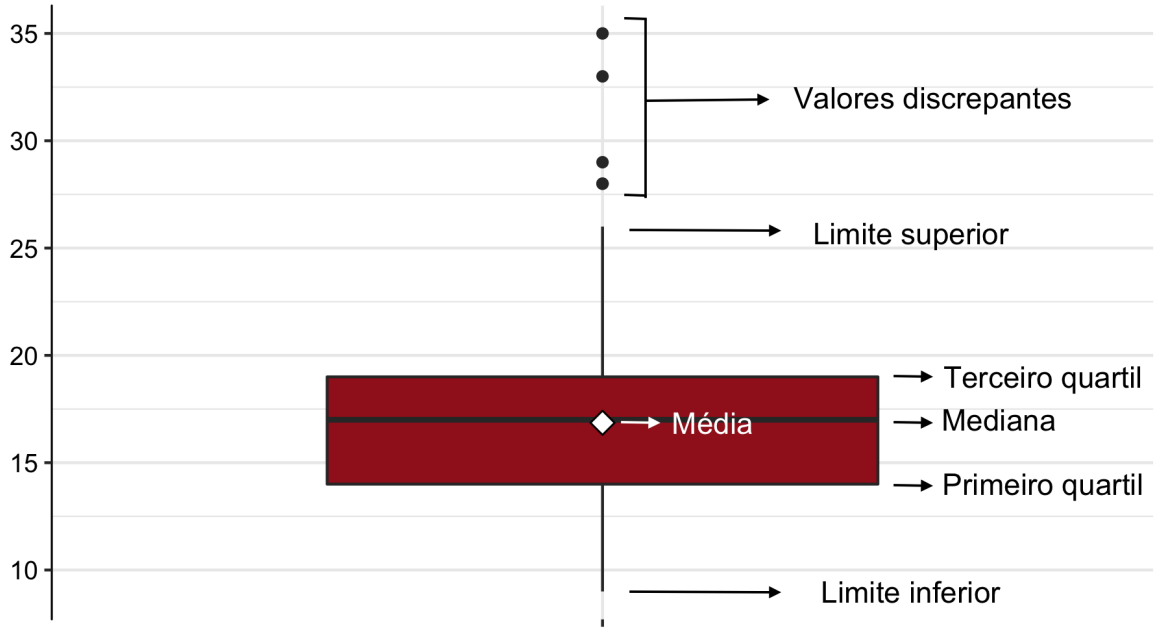
Com:

- $X_i =$ i-ésima observação da população
- $\mu =$ média populacional
- $N =$ tamanho da população

2.7 Boxplot

O boxplot é uma representação gráfica na qual se pode perceber de forma mais clara como os dados estão distribuídos. A figura abaixo ilustra um exemplo de boxplot.

Figura 1: Exemplo de boxplot

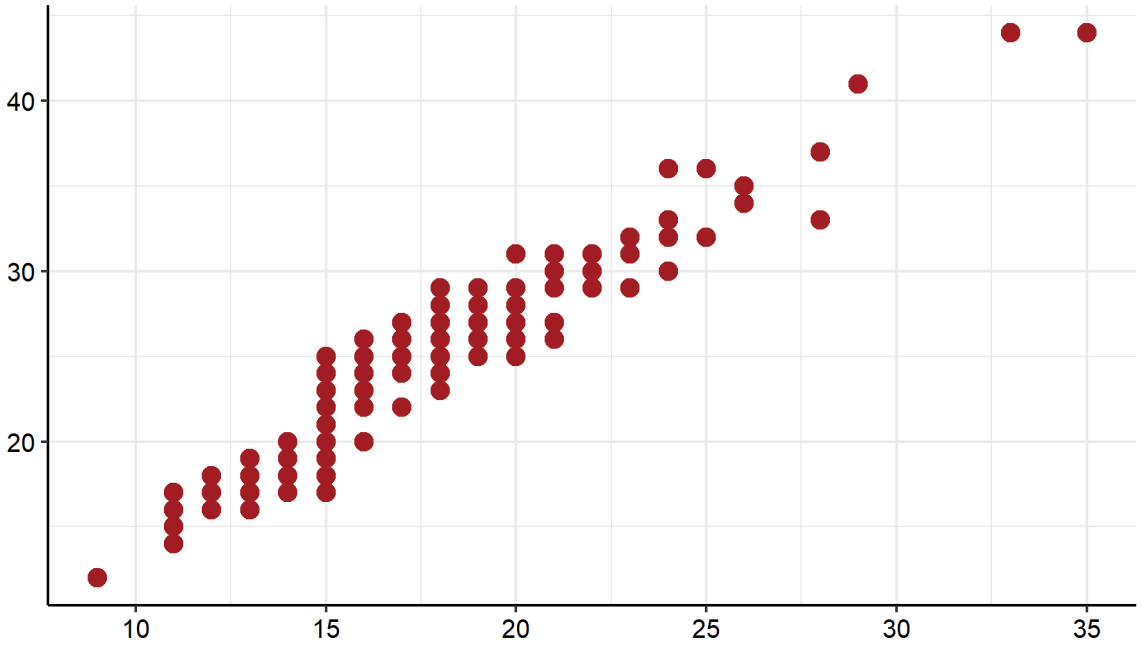


A porção inferior do retângulo diz respeito ao primeiro quartil, enquanto a superior indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada pelo losango branco e os pontos são *outliers*. Os *outliers* são valores discrepantes da série de dados, ou seja, valores que não demonstram a realidade de um conjunto de dados.

2.8 Gráfico de Dispersão

O gráfico de dispersão é uma representação gráfica utilizada para ilustrar o comportamento conjunto de duas variáveis quantitativas. A figura abaixo ilustra um exemplo de gráfico de dispersão, onde cada ponto representa uma observação do banco de dados.

Figura 2: Exemplo de Gráfico de Dispersão



2.9 Tipos de Variáveis

2.9.1 Qualitativas

As variáveis qualitativas são as variáveis não numéricas, que representam categorias ou características da população. Estas subdividem-se em:

- **Nominais**: quando não existe uma ordem entre as categorias da variável (exemplos: sexo, cor dos olhos, fumante ou não, etc)
- **Ordinais**: quando existe uma ordem entre as categorias da variável (exemplos: nível de escolaridade, mês, estágio de doença, etc)

2.9.2 Quantitativas

As variáveis quantitativas são as variáveis numéricas, que representam características numéricas da população, ou seja, quantidades. Estas subdividem-se em:

- **Discretas**: quando os possíveis valores são enumeráveis (exemplos: número de filhos, número de cigarros fumados, etc)
- **Contínuas**: quando os possíveis valores são resultado de medições (exemplos: massa, altura, tempo, etc)

2.10 Teste de Hipóteses

O teste de hipóteses tem como objetivo fornecer uma metodologia para verificar se os dados das amostras possuem indicativos que comprovem, ou não, uma hipótese previamente formulada. Ele é composto por duas hipóteses:

$$\begin{cases} H_0 : \text{hipótese a ser testada (chamada de hipótese nula)} \\ H_1 : \text{hipótese alternativa que será aceita caso a hipótese nula seja rejeitada} \end{cases}$$

Essa decisão é tomada por meio da construção de uma região crítica, ou seja, região de rejeição do teste.

2.10.1 Tipos de teste: bilateral e unilateral

Para a formulação de um teste, deve-se definir as hipóteses de interesse. Em geral, a hipótese nula é composta por uma igualdade (por exemplo, $H_0 : \theta = \theta_0$). Já a hipótese alternativa depende do grau de conhecimento que se tem do problema em estudo. Assim, tem-se três formas de elaborar H_1 que classificam os testes em duas categorias:

- **Teste Bilateral**:
  Esse é o teste mais geral, em que a hipótese alternativa consiste em verificar se existe diferença entre os parâmetros de interesse, independentemente de um ser maior ou menor que o outro. Dessa forma, tem-se:

$$H_1 : \theta \neq \theta_0$$

- **Teste Unilateral**:
  Dependendo das informações que o pesquisador possui a respeito do problema e os questionamentos que possui, a hipótese alternativa pode ser feita de forma a verificar se existe diferença entre os parâmetros em um dos sentidos. Ou seja:

$$H_1 : \theta < \theta_0$$

ou

$$H_1 : \theta > \theta_0$$

2.11 Tipos de Erros

Ao realizar um teste de hipóteses, existem dois erros associados: **Erro do Tipo I** e **Erro do Tipo II**.

- **Erro do Tipo I**:
  Esse erro é caracterizado por rejeitar a hipótese nula (H_0) quando essa é verdadeira. A probabilidade associada a esse erro é denotada por α, também conhecido como nível de significância do teste.

- **Erro do Tipo II**:
  Ao não rejeitar H_0 quando, na verdade, é falsa, está sendo cometido o **Erro do Tipo II**. A probabilidade de se cometer este erro é denotada por β.

2.11.1 Nível de significância (α)

O nível de significância do teste é o nome dado à probabilidade de se rejeitar a hipótese nula quando essa é verdadeira; essa rejeição é chamada de **erro do tipo I**. O valor de α é fixado antes da extração da amostra e, usualmente, assume 5%, 1% ou 0,1%.

Por exemplo, um nível de significância de $\alpha = 0,05$ (5%) significa que, se for tomada uma grande quantidade de amostras, em 5% delas a hipótese nula será rejeitada quando não havia evidências para essa rejeição, isto é, a probabilidade de se tomar a decisão correta é de 95%.

2.11.2 Estatística do Teste

A estatística do teste é o estimador que será utilizado para testar se a hipótese nula (H_0) é verdadeira ou não. Ela é escolhida por meio das teorias estatísticas.

2.11.3 P-valor

O **P-valor**, ou nível descritivo, é uma medida utilizada para sintetizar o resultado de um teste de hipóteses. Ele também pode ser chamado de *probabilidade de significância* do teste e indica a probabilidade de se obter um resultado da estatística de teste mais extremo do que o observado na presente amostra, considerando que a hipótese nula é verdadeira. Dessa forma, rejeita-se H_0 quando P-valor $< \alpha$, porque a chance de uma nova amostra possuir valores tão extremos quanto o encontrado é baixa, ou seja, há evidências para a rejeição da hipótese nula.

2.12 Intervalo de Confiança

Quando calcula-se um estimador pontual para o parâmetro, não é possível definir qual a possível magnitude do erro que se está cometendo. Com o objetivo de associar um erro à estimativa, são construídos os intervalos de confiança que se baseiam na distribuição amostral do estimador pontual.

Dessa forma, considere T um estimador pontual para θ e que a distribuição amostral de T é conhecida. O intervalo de confiança para o parâmetro θ será dado por t_1 e t_2, tal que:

$$P(t_1 < \theta < t_2) = \gamma$$

A probabilidade γ é estabelecida no início do estudo e representa o nível de confiança do intervalo. A interpretação desse resultado é que, se forem tiradas várias amostras de mesmo tamanho e forem calculados intervalos de confiança para cada uma, $100 \times \gamma\%$ dos intervalos irão conter o parâmetro θ. Assim, ao calcular um intervalo, pode-se dizer que há $100 \times \gamma\%$ de confiança de que o intervalo contém o parâmetro de interesse.

2.13 Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson é uma medida que verifica o grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor zero significa que não há relação linear entre as variáveis. Quando o valor do coeficiente r é negativo, diz-se existir uma relação de grandeza inversamente proporcional entre as variáveis. Analogamente, quando r é positivo, diz-se que as duas variáveis são diretamente proporcionais.

O coeficiente de correlação de Pearson é normalmente representado pela letra r e a sua fórmula de cálculo é:

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

Onde:

- x_i = i-ésimo valor da variável X
- y_i = i-ésimo valor da variável Y
- $\bar{x}$ = média dos valores da variável X
- $\bar{y}$ = média dos valores da variável Y

Vale ressaltar que o coeficiente de Pearson é paramétrico e, portanto, sensível quanto à normalidade (simetria) dos dados.

2.14 Teste de Correlação de Pearson

O coeficiente de correlação linear de Pearson indica a força e a direção do relacionamento linear entre duas variáveis quantitativas. É um índice adimensional com valores situados entre -1 e 1, no qual o valor -1 representa total correlação linear negativa entre as variáveis (quando o valor de uma variável cresce, o valor da outra diminui) e o valor 1 representa total correlação linear positiva entre elas (ambas crescem simultaneamente). Esse coeficiente é obtido por meio da fórmula:

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

em que

- $x_i =$ i-ésimo valor da variável X
- $y_i =$ i-ésimo valor da variável Y
- $\bar{x} =$ média dos valores da variável X
- $\bar{y} =$ média dos valores da variável Y
- $r_{Pearson} =$ coeficiente de correlação linear de Pearson amostral

Para o teste de correlação de Pearson, tem-se as seguintes hipóteses:

$$\begin{cases} H_0 : \text{Não há correlação linear entre as variáveis } X \text{ e } Y \\ \qquad (\rho_{Pearson} = 0) \\ H_1 : \text{Há correlação linear entre as variáveis } X \text{ e } Y \\ \qquad (\rho_{Pearson} \neq 0) \end{cases}$$

em que $\rho_{Pearson}$ é o parâmetro a ser testado: coeficiente de correlação linear populacional.

Se X e Y tem distribuição normal, tem-se que a estatística do teste é dada por:

$$t_{Pearson} = \frac{r_{Pearson}\sqrt{n-2}}{\sqrt{1 - r_{Pearson}^2}} \sim t_{n-2}$$

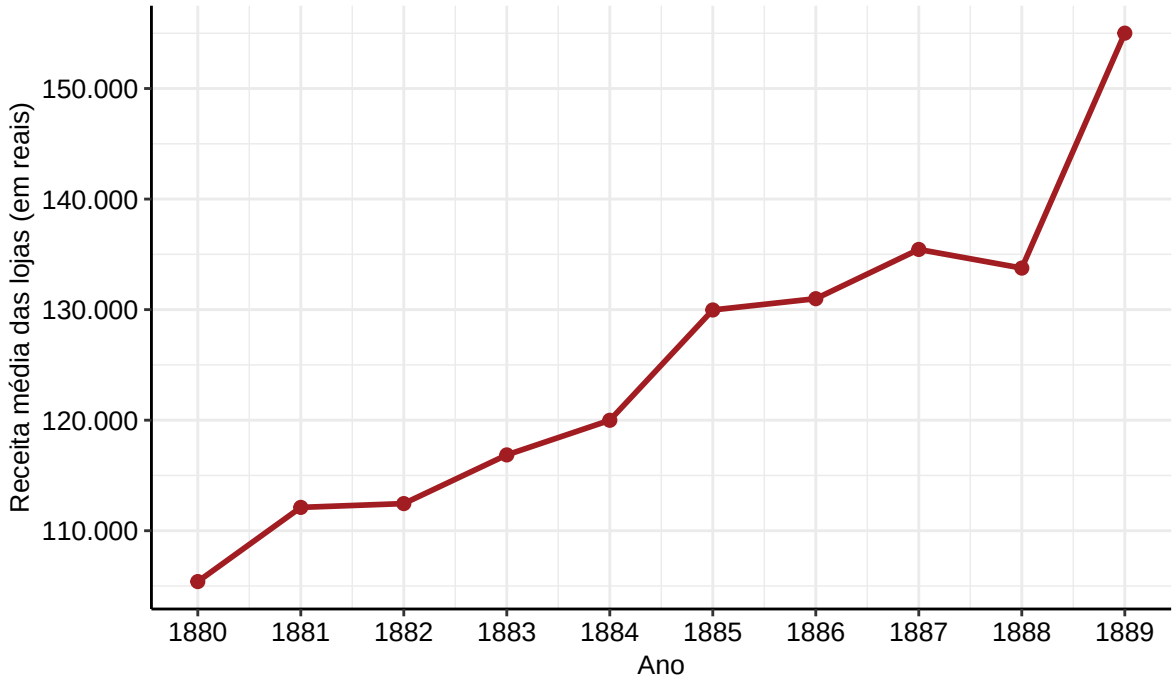
Assim, sob H_0, $t_{Pearson}$ segue uma distribuição *t*-Student com $(n-2)$ graus de liberdade.

3 Análises

3.1 Receita média das lojas registrada nos anos de 1880 até 1889

Essa análise tem como objetivo visualizar a evolução da receita média total das lojas, em reais, na região do faroeste nos últimos 10 anos (1880-1889). As variáveis utilizadas foram a renda média das lojas (em reais) e o ano, uma variável quantitativa contínua e uma quantitativa discreta, respectivamente. Além disso, foi utilizada uma taxa de conversão de 5,31 reais para 1 dólar.

Figura 3: Gráfico de linhas da receita média das lojas por ano (em reais)



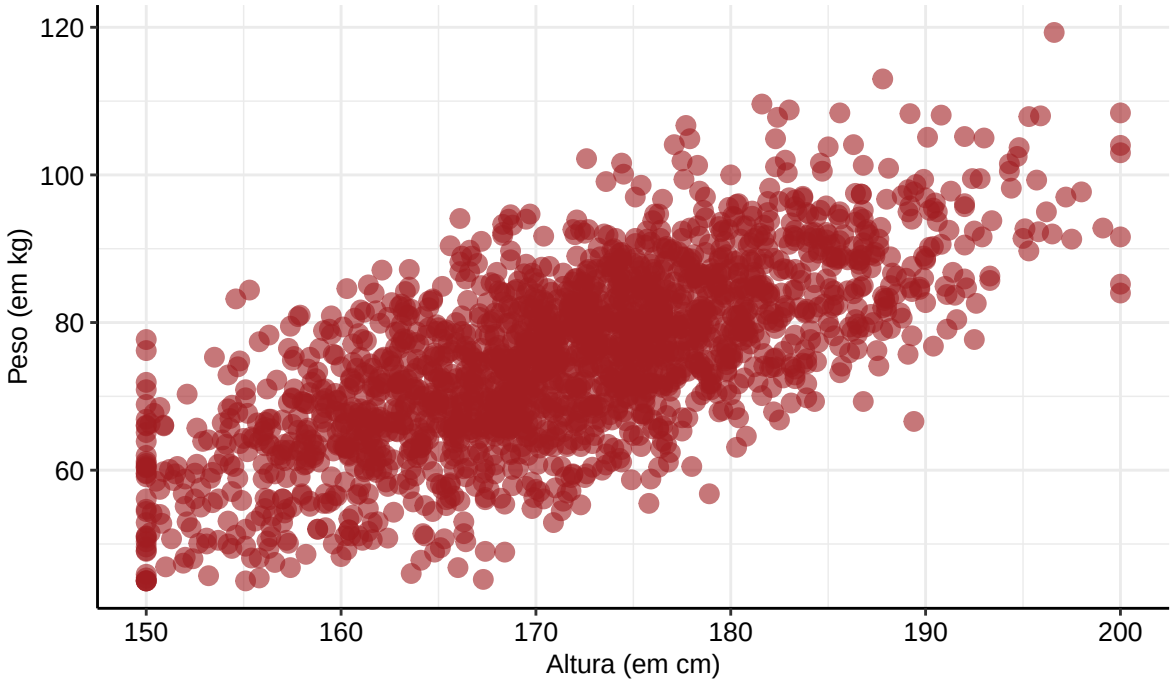
A **Figura 3** evidencia que houve um constante aumento na receita média das lojas, indo de R$105.399,00 no primeiro ano até R$155.009,10 no último ano (com uma pequena queda entre 1887 e 1888, de R$135.444,80 para R$133.757,60). Portanto, ao longo dos 10 anos, a receita média aumentou R$49.610,10. Ou seja, houve um aumento médio de, aproximadamente, R$5.000,00 por ano.

3.2 Variação Peso por Altura

O intuito dessa análise é entender a relação entre o peso (em quilogramas) e altura (em centímetros) dos clientes. As variáveis em questão, citadas anteriormente, são classificadas como quantitativas contínuas e foi elaborado um gráfico de dispersão para observar o comportamento delas. A partir disso, concluiremos se a medida que

o peso aumenta: a altura também aumenta, o contrário ou se não tem diferença. O nível de significância utilizado será de 5%.

Figura 4: Gráfico de dispersão do peso pela altura



Quadro 1: Medidas resumo do peso (kg)

Estatística	Valor
Média	75,19
Desvio Padrão	11,92
Variância	142,00
Mínimo	45,00
1º Quartil	66,90
Mediana	75,30
3º Quartil	83,20
Máximo	119,30

Quadro 2: Medidas resumo da altura (cm)

Estatística	Valor
Média	171,48
Desvio Padrão	9,87
Variância	97,38
Mínimo	150,00
1º Quartil	164,80
Mediana	171,75
3º Quartil	178,00
Máximo	200,00

A **Figura 4** apresentada acima mostra a relação entre altura (em centímetros) e peso (em quilogramas) dos clientes. Cada ponto representa um indivíduo, permitindo visualizar a distribuição conjunta das duas variáveis. Pela **Figura 4**, há indícios de que existe uma relação positiva, ou seja, à medida que a altura aumenta, o peso também tende a ser maior. É exatamente isso que iremos verificar a seguir. Além disso, os Quadros **1** e **2** apresentam as medidas resumo das variáveis, que nos ajudam a entender um pouco de como elas se comportam, sua variabilidade, etc.

Foi escolhido o teste de correlação de Pearson, já que o coeficiente de correlação linear de Pearson indica a força e a direção do relacionamento linear entre duas variáveis quantitativas. As hipóteses do teste são:

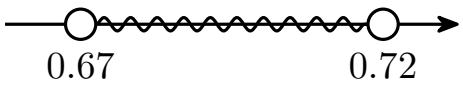
$$\begin{cases} H_0 : \text{Não há correlação linear entre altura e peso} \\ H_1 : \text{Há correlação linear entre altura e peso} \end{cases}$$

Ao realizar o teste, foram encontradas as seguintes conclusões:

Tabela 1

Variáveis	P-valor	Decisão do teste
Peso e Altura	< 0,001	Rejeita H_0

Quadro 3: Intervalo de confiança do coeficiente de correlação de Pearson

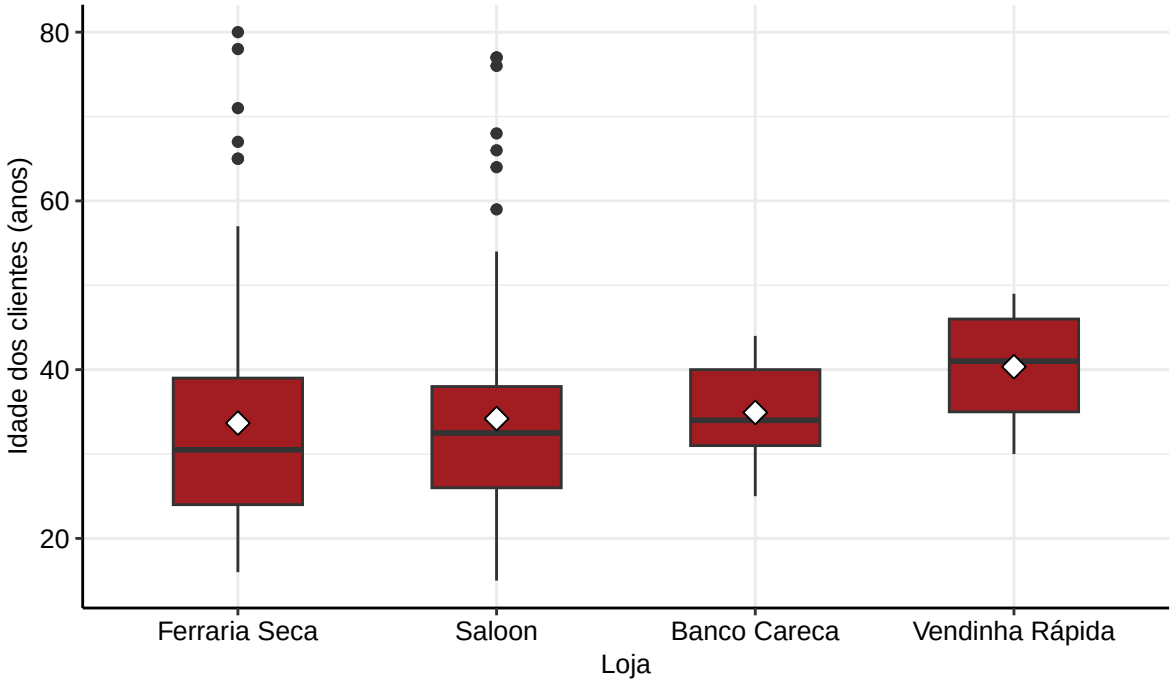
Parâmetro	Intervalo de Confiança (95%)
Coeficiente de Correlação Linear	

A **Tabela 1** mostra os resultados. O coeficiente de correlação linear de Pearson entre peso e altura é 0.697, com um intervalo de confiança de 95% entre 0.63 e 0.74 evidenciado no **Quadro 3**. Isso indica uma correlação positiva forte e estatisticamente significativa, sugerindo que indivíduos mais altos tendem a ter maior peso, e é improvável que essa relação seja fruto do acaso. Em resumo, a medida que a altura cresce, o peso tende a crescer (e vice-versa). O inverso também acontece, a medida que o cliente é mais baixo, o peso tende de maneira moderada a forte ser mais baixo também.

3.3 Idade dos clientes de Âmbar Seco a depender da loja

Com o intuito de entender melhor o perfil das idades dos clientes nas diferentes lojas da cidade de Âmbar Seco, serão descritas as características dentro do banco de dados e mostrados quais são os perfis das idades dos clientes para cada loja da cidade. As variáveis ulitizadas foram idade e loja, classificadas como quantitativa discreta e qualitativa nominal, respectivamente. Foram analisados 415 clientes diferentes distribuídos por 4 lojas.

Figura 5: Boxplot das idades dos clientes por loja



Quadro 4: Medidas resumo das idades por loja

Estatística	Banco Careca	Vendinha Rápida	Saloon	Ferraria Seca
Média	34,92	40,35	34,20	33,67
Desvio Padrão	5,57	6,03	12,70	13,31
Variância	31,06	36,39	161,23	177,18
Mínimo	25,00	30,00	15,00	16,00
1º Quartil	31,00	35,00	26,00	24,00
Mediana	34,00	41,00	32,50	30,50
3º Quartil	40,00	46,00	38,00	39,00
Máximo	44,00	49,00	77,00	80,00

Os boxplots apresentados no **Figura 5** ilustram a distribuição das idades dos clientes nas quatro lojas de Âmbar Seco: Ferraria Seca, Banco Careca, Saloon e Vendinha Rápida. Observa-se que as idades dos clientes variam entre as lojas, tanto em termos de centralidade quanto de dispersão. O **Quadro 4** evidencia algumas estatisticas a partir de números precisos que podem passar batidos ao utilizar o boxplot, portanto, também tem enorme valor para a análise.

De forma geral, observa-se que as idades dos clientes variam consideravelmente entre as lojas, com medianas situadas entre 30 e 41 anos e médias entre 33 e 41 anos. A seguir, destacam-se os principais pontos de cada loja:

A Ferraria Seca apresenta a menor média e mediana de idade entre as 4 e a maior variância, o que é bem interessante e explica o fato de que o maior máximo também é desse recinto.

O Banco Careca possui a menor variância e o menor máximo (quase metade do maior, pertencente à Ferraria Seca), o que faz bastante sentido já que é o mais uniforme e concentrado em alguma faixa específica. Nesse caso, a idade vai de 25 a 44 anos, ou seja, do mínimo para o máximo nem dobra. Enquanto isso, outras lojas tem o máximo igual a 5 vezes o mínimo.

Em Saloon, temos a menor idade de todas e uma alta variância, assim como Ferraria Seca. Por isso, ambos tem as duas maiores amplitudes (por muito), sendo de 62 anos em Saloon e 64 na Ferraria Seca. A idade máxima dessa loja é de 77 anos.

Na Vendinha Rápida está a média mais velha, na faixa dos 40 anos. Aqui a variância é pequena se comparada a Saloon e Ferraria Seca e a idade máxima é de 49 anos.

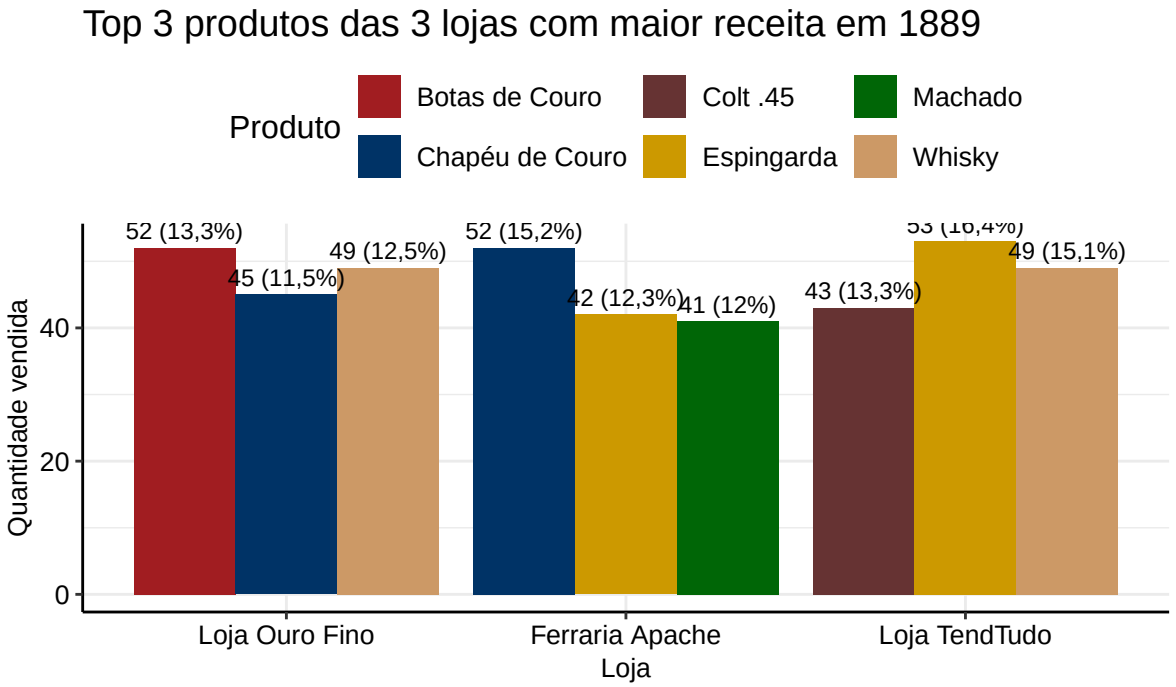
Em termos de tendência central, a média e a mediana de todas está relativamente próxima entre elas para cada loja, o que indica que são razoavelmente bem distribuídas. O que muda é que enquanto o Banco Careca e a Vendinha Rápida variam menos, Saloon e Ferraria Seca têm variâncias muito maiores. Fato que explica a diferença de amplitude entre os dois primeiros e os dois últimos citados.

Essa sequência de constatações permite dizer que o público do Banco Careca e da Vendinha Rápida, no geral, é mais específico e jovem, ficando entre 25 e 50 anos. Enquanto isso, Saloon e Ferraria Seca têm públicos mais distribuídos entre todas as faixasa etárias, levando a crer que para investir em produtos de público-alvo mais jovem especificamente, faz mais sentido olhar para Banco Careca e Vendinha Rápida. Enquanto as outras duas lojas tem potencial para todos os públicos-alvo, não sendo tão nichado.

3.4 Top 3 produtos mais vendidos nas top 3 lojas com maior receita em 1889

Essa análise tem o objetivo de encontrar e visualizar quais são os 3 produtos mais vendidos nas 3 lojas que tiveram a maior receita no ano de 1889. Dessa forma, entender quais foram os produtos, a quantidade vendida e as lojas que mais venderam neste ano. As variáveis utilizadas foram o nome das lojas (qualitativa nominal), nome do produto (qualitativa nominal), quantidade vendida (quantitativa discreta) e receita (quantitativa discreta).

Figura 6: Gráfico de barras bivariado com frequências



A **Figura 6** mostra os 3 produtos mais vendidos das 3 lojas com maior receita, em ordem (da maior loja para menor).Os únicos produtos que se repetem são o whisky (Loja Ouro Fino e Loja TendTudo), o chapéu de couro (Loja Ouro Fino e Ferraria Apache) e a espingarda (Loja TendTudo e Ferraria Apache). Porém, todas as lojas tem pelo menos um produto em comum com as outras, o que sugere uma popularidade desses produtos no Faroeste, independentemente da loja.

A porcentagem em cima de cada coluna é a frequência relativa daquele produto em relação à receita total de cada loja. É interessante observar que em nenhuma dessas lojas os 3 produtos mais vendidos somam mais de 45% da receita. Ou seja, há uma variedade de produtos e vendas, não dependendo de algum específico. Esse fato é positivo para as lojas, pois não dependem de um produto único e conseguem captar o capital de forma variada e dispersa.

4 Conclusões

O objetivo desse projeto era entender melhor o cenário do comércio no Faroeste para estudá-lo e entender dores, possibilidades de expansão, tendências, etc. Compreender características e extrair informaçõe relevantes para o negócio, baseando-se em dados de todo tipo e relacionados à diferentes pontos (receita média, características físicas, idade, produtos).

Em relação à evolução do mercado, as análises de receita média das lojas indicam um crescimento constante ao longo da última década (1880-1889). Houve um aumento médio de aproximadamente R$ 5.000,00 na receita média anual, sugerindo um mercado em expansão e saudável com possibilidades de investimentos.

No que tange ao perfil do consumidor, a análise da correlação entre peso e altura demonstrou haver uma relação positiva estatisticamente significativa. Isso implica que clientes mais altos tendem a ter um peso maior, uma informação útil para estratégias relacionadas a produtos que dependem de dimensões físicas, podendo direcionar melhor os produtos para cada região (ou até cada loja).

Sobre a idade dos clientes de Âmbar Seco, foi possível tirar conclusões relevantes para cada loja referente à faixa etária. Prinipalmente que o público do Banco Careca e da Vendinha Rápida, no geral, é mais específico e jovem, ficando entre 25 e 50 anos. Enquanto Saloon e Ferraria Seca têm públicos mais distribuídos entre todas asfaixasa etárias, levando a crer que para investir em produtos de público-alvo mais jovem especificamente, faz mais sentido olhar para Banco Careca e Vendinha Rápida. Em contrapartida, as outras duas lojas tem potencial para todos os públicos-alvo, não sendo tão nichado.

Por fim, ao analisar os produtos de maior sucesso nas três lojas com maior receita em 1889, verificou-se que produtos como chapéu de couro, whisky e espingarda são populares em todo o Faroeste, pois se repetem entre as líderes de vendas. No entanto, é um ponto positivo que nenhuma das lojas dependa excessivamente de um único item, já que o trio de produtos mais vendidos não somou mais de 45% da receita em nenhum dos casos. Isso demonstra uma variedade saudável de vendas e uma entrada de capital de forma dispersa.