



**Instituto Superior de Contabilidade e Administração de Coimbra**  
Instituto Politécnico de Coimbra

# **Trabalho N.º 1 de Análise Estatística de Dados**

Catarina Juliana Martins Auxiliar, 2021134297

Helena Cristina Almeida da Cruz, 2024147830

Nuno Emanuel Lopes Gonçalves, 2015063961

Simão Pedro Tomé Dias, 2020132169

**COIMBRA**

**20 de outubro de 2024**

## RESUMO

O presente relatório foi elaborado no âmbito da disciplina de Análise Estatística de Dados do Mestrado em Análise de Dados e Sistemas de Apoio à Decisão, com a orientação da professora Clara Margarida Pisco Viseu. O principal objetivo do trabalho consiste em aplicar num caso prático, os conteúdos lecionados em aula, que incluem, Estatística Descritiva, Testes Paramétricos e Testes Não Paramétricos, de forma a consolidar os conceitos aprendidos. Nesse sentido, procedeu-se à análise do comportamento das vendas de café numa cadeia de lojas, partindo de uma amostra representativa das vendas dos dois tipos de café mais vendidos e tendo em conta os seguintes fatores: a região onde a loja está localizada (Norte, Centro, Sul), a dimensão da loja (Pequena/Grande), as despesas relacionadas com a promoção do produto (em dólares) e o valor esperado de vendas (em dólares). Ao longo do relatório são apresentadas respostas às 5 questões propostas no caso prático, respostas essas que incluem uma descrição e justificação do método estatístico utilizado, os resultados obtidos no programa de análise estatística SPSS29, e as respetivas conclusões retiradas da análise. De uma forma geral, dados os resultados obtidos, pode concluir-se que as despesas relacionadas com a promoção do produto têm um impacto positivo significativo nas vendas de café, enquanto que os restantes fatores não apresentam influência significativa. Logo, a estratégia da cadeia de lojas deve passar pelo reforço do investimento em estratégias de marketing, pois estas revelam-se essenciais para o seu bom desempenho comercial.

## ÍNDICE GERAL

INTRODUÇÃO.....	1
1 Variáveis em estudo .....	4
2 Resolução das questões propostas .....	4
2.1 Questão 1: Faça uma análise descritiva e comparativa das vendas segundo a dimensão da loja (inclua, na sua análise, gráficos e medidas descritivas adequadas).....	5
2.2 Questão 2: Analise a correlação entre Vendas e Despmark (justifique a escolha do coeficiente de correlação usado e interprete os resultados obtidos).....	9
2.3 Questão 3: Verificação da existência de diferenças significativas entre as médias das vendas dos dois tipos de café, utilizando teste estatístico apropriado.....	11
2.4 Questão 4: Teste se há diferenças significativas entre as vendas (reais) e as vendas esperadas utilizando um teste estatístico apropriado.....	13
2.5 Questão 5: Verifique, usando um teste estatístico adequado, se a região tem impacto significativo nas vendas.....	17
CONCLUSÃO.....	21

## ÍNDICE DE FIGURAS

Figura 1: Variáveis em análise no trabalho .....	4
Figura 2: Distribuição do Valor de Vendas por Dimensão da Loja: Histograma.....	8
Figura 3: Distribuição do Valor de Vendas por Dimensão da Loja: Boxplot.....	8
Figura 4: Distribuição do Valor de Vendas por Dimensão da Loja: Pie Chart .....	9
Figura 5: Gráfico de Dispersão entre Vendas e Despesas de Marketing.....	10

## ÍNDICE DE TABELAS

Tabela 1: Distribuição da amostra por dimensão da loja.....	5
Tabela 2: Estatísticas Descritivas das Vendas em função da Dimensão da Loja .....	5
Tabela 3: Tabela de Correlação de Pearson entre Valor de Vendas e Despesas de Marketing.....	10
Tabela 4: Testes de Normalidade para o Valor das Vendas por Tipo de Café.....	11
Tabela 5: Estatísticas Descritivas das Vendas por Tipo de café.....	12
Tabela 6: Teste de Levene para Igualdade de Variâncias e Teste t de Amostras Independentes para Comparação das Vendas entre os Dois Tipos de Café .....	12
Tabela 7: Medidas de Tamanho de Efeito para Comparação das Vendas entre os Dois Tipos de Café.....	13
Tabela 8: Correlação entre as vendas reais e as vendas esperadas .....	14
Tabela 9: Estatísticas descritivas vendas reais e esperadas .....	14
Tabela 10: Testes de Normalidade para Vendas e Vendas Esperadas.....	15
Tabela 11: Estatísticas descritivas das vendas por tipo de café.....	16
Tabela 12: Teste de Amostras Emparelhadas para Vendas e Vendas Esperadas .....	16
Tabela 13: Medidas de Tamanho de Efeito para Comparação das Vendas com as Vendas esperadas.....	17

---

Tabela 14: Teste de Levene para Homogeneidade das Variâncias das Vendas entre Regiões.....	18
Tabela 15: Testes de Normalidade para o Valor das Vendas por Região.....	19
Tabela 16: Dimensão das amostras por região .....	19
Tabela 17: Teste ANOVA .....	20
Tabela 18: Tamanho do efeito ANOVA.....	20

## INTRODUÇÃO

O objetivo do trabalho realizado e apresentado neste relatório foi o de aplicar num contexto prático, os conceitos teóricos lecionados na disciplina de Análise Estatística de Dados do Mestrado em Análise de Dados e Sistemas de Apoio à Decisão, de forma a consolidar os conhecimentos adquiridos em aula. Os tópicos abordados no trabalho incluem Estatística Descritiva, Testes Paramétricos e Testes Não Paramétricos.

O desafio proposto consistia em responder a 5 questões-chave, que permitiam identificar de entre um conjunto de fatores (localização da loja (Norte, Centro, Sul), dimensão da loja (Pequena/Grande), despesas relacionadas com a promoção do produto (em dólares) e o valor esperado de vendas (em dólares), aqueles que tinham maior impacto no volume de vendas de café de uma determinada cadeia de lojas. Para responder a essas questões, partiu-se de uma amostra aleatória representativa das vendas dos dois tipos de café mais vendidos pela cadeia de lojas.

De forma a facilitar a análise e interpretação dos dados, iniciou-se o trabalho com a análise das variáveis e a classificação das mesmas. De seguida, partiu-se para a análise descritiva e comparativa das vendas segundo a dimensão da loja. De forma a avaliar o impacto do montante das despesas relacionadas com a promoção do produto no volume de vendas, foi também realizada uma análise de correlação entre as variáveis Vendas e Despmark. Por último, foram realizados três testes de hipóteses distintos, o primeiro com o intuito de testar a existência de diferenças significativas entre as médias das vendas dos dois tipos de café, o segundo com o objetivo de testar a existência de diferenças significativas entre as vendas reais e as vendas esperadas de café e o terceiro, de forma a testar se a localização geográfica da loja tem impacto significativo no volume de vendas de café.

Todas as questões foram respondidas com base na aplicação dos métodos estatísticos considerados mais apropriados e com base nos resultados gerados através do software de análise estatística SPSS29. Através desta análise, foi possível retirar conclusões relevantes sobre os fatores que mais impactam o volume de vendas na cadeia de lojas.

## 1 Variáveis em estudo

Para facilitar a análise e interpretação dos dados, é fundamental distinguir entre variáveis qualitativas e quantitativas.

As variáveis qualitativas representam categorias ou qualidades que não são medidas numericamente, mas sim classificadas em grupos ou categorias e podem ser do tipo nominal ou ordinal. As variáveis qualitativas nominais não têm uma ordem intrínseca entre as categorias. Por outro lado, as variáveis qualitativas ordinais, têm uma ordem ou hierarquia entre as categorias.

As variáveis quantitativas são medidas numa escala intervalar (onde o ponto zero é arbitrário) ou numa escala de razão (onde o zero representa a ausência da característica) e podem ser do tipo discreto ou contínuo. As variáveis quantitativas discretas assumem valores inteiros ou finitos, já as variáveis quantitativas contínuas podem assumir qualquer valor dentro de um intervalo.

O conjunto de variáveis em análise no presente trabalho é o seguinte:



*Figura 1: Variáveis em análise no trabalho*

No conjunto de variáveis podemos classificar como qualitativas: Região (nominal), Dim (nominal) e Tipo (nominal); e como quantitativas: Vendas (contínua), Despmark (contínua) e Vendasesp (contínua).

## 2 Resolução das questões propostas

Realizada uma análise preliminar e classificação das variáveis em estudo, iniciou-se a resolução das questões propostas.

Começou-se por obter uma amostra aleatória de dimensão  $n=880-5k$ , onde  $k$  correspondia ao número do grupo. Dado que o nosso grupo era o 6, obtivemos uma amostra de dimensão  $n=850$ .

## 2.1 Questão 1: Faça uma análise descritiva e comparativa das vendas segundo a dimensão da loja (inclua, na sua análise, gráficos e medidas descritivas adequadas).

Iniciou-se a resolução das questões com uma análise descritiva e comparativa das vendas segundo a dimensão da loja, onde se obtiveram as tabelas que se apresentam de seguida.

**Resumo de processamento de casos**

	Dimensão da loja	Válido		Casos Omisso		Total	
		N	Porcentagem	N	Porcentagem	N	Porcentagem
Valor de vendas	Grande	341	100,0%	0	0,0%	341	100,0%
	Pequena	509	100,0%	0	0,0%	509	100,0%

**Tabela 1: Distribuição da amostra por dimensão da loja**

Por observação da Tabela 1, podemos confirmar a dimensão da nossa amostra através da soma das vendas por dimensão da loja (Grande /Pequena), certificando-nos de que n=850.

**Descritivas**

Dimensão da loja		Estatística		Estatística do teste Padrão	
Valor de vendas	Grande	Média	288,33	9,096	
		95% de Intervalo de Confiança para Média	Limite inferior	270,44	
			Limite superior	306,22	
		5% da média aparada	279,62		
		Mediana	238,00		
		Variância	28214,227		
		Erro Padrão	167,971		
		Mínimo	63		
		Máximo	678		
		Amplitude	615		
		Amplitude interquartil	278		
		Assimetria	,738	,132	
		Curtose	-,646	,263	
	Pequena	Média	128,02	2,864	
		95% de Intervalo de Confiança para Média	Limite inferior	122,40	
			Limite superior	133,65	
		5% da média aparada	122,48		
		Mediana	116,00		
		Variância	4175,692		
		Erro Padrão	64,620		
		Mínimo	32		
		Máximo	387		
		Amplitude	355		
		Amplitude interquartil	73		
		Assimetria	1,306	,108	
		Curtose	2,046	,216	

**Tabela 2: Estatísticas Descritivas das Vendas em função da Dimensão da Loja**



### **Lojas de Dimensão “Grande”**

#### **Medidas de tendência central dos dados (ou de localização):**

Nas lojas de maior dimensão, o valor médio das vendas é de 288,33 dólares com um intervalo de confiança de 95% (entre 270,44 dólares e 306,22 dólares), o que indica que, com 95% de confiança, a média das vendas se situa nesse intervalo.

Relativamente à mediana, o seu valor indica que 50% das lojas de maior dimensão tiveram vendas abaixo de 238,00 dólares e 50% das lojas tiveram vendas acima desse montante.

A diferença entre a média (288,33 dólares) e a mediana (238,00 dólares) sugere que a distribuição das vendas pode ser assimétrica. Numa distribuição simétrica, a média e a mediana seriam aproximadamente iguais. Neste caso, a média é maior que a mediana, o que pode ser uma indicação de uma distribuição assimétrica à direita, ou que pode significar que algumas lojas de grande dimensão apresentam vendas muito altas.

#### **Medidas de dispersão (variabilidade dos dados):**

No que respeita à variância, podemos constatar que a mesma é alta (28214,227), indicador da grande variabilidade nas vendas entre as lojas de grande dimensão.

O desvio padrão de 167,971 dólares, mais fácil de interpretar, pois encontra-se na mesma unidade de medida dos dados (dólares), confirma a indicação dada pela variância, de que as vendas apresentam uma grande variabilidade em torno da sua média.

A variação dos valores de vendas, entre 63 e 678 dólares, com uma amplitude de 615 dólares, reforça a conclusão retirada sobre a ampla dispersão das vendas nas lojas de grande dimensão.

#### **Medidas de forma (distribuição dos dados):**

O valor de assimetria de 0,738 indica uma assimetria positiva (ou inclinação à direita), o que significa que a cauda direita da distribuição é mais longa do que a cauda esquerda, ou seja, há alguns valores de vendas maiores, que fazem subir o valor da média.

O valor de curtose de -0,646 indica que a distribuição é ligeiramente platicúrtica, ou seja, tem caudas mais achatadas do que a distribuição normal, o que significa que há menos

valores extremos do que numa distribuição normal, com os dados mais concentrados perto da média e menos dispersos nas caudas.

### **Lojas de Dimensão “Pequena”**

#### **Medidas de tendência central dos dados (ou de localização):**

Por sua vez, nas lojas de menor dimensão, o valor médio das vendas é de 128,02 dólares com o intervalo de confiança de 95% (entre 122,40 e 133,65 dólares), o que indica que, com 95% de confiança, a média das vendas se situa nesse intervalo. Comparando com a média obtida para as lojas de dimensão “grande”, pode concluir-se que o volume de vendas em lojas pequenas é significativamente inferior ao das lojas grandes.

Relativamente à mediana, que é de 116,00 dólares, próxima da média, sugere uma distribuição mais simétrica do que no caso das lojas grandes.

#### **Medidas de dispersão (variabilidade dos dados):**

Em termos de variabilidade, a variância (4175,692) e o desvio padrão (64,620) são menores do que os observados para as lojas de grande dimensão, o que indica uma menor dispersão das vendas em torno da média.

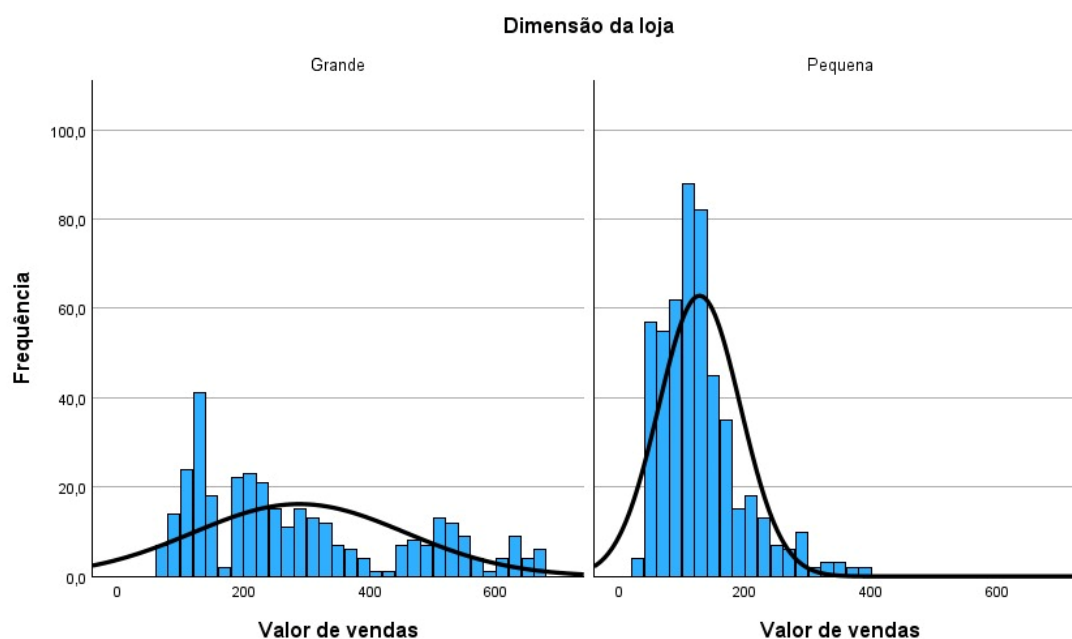
A variação dos valores de vendas entre 32 e 387 dólares, com uma amplitude de 355 dólares, reforça a conclusão retirada sobre a menor dispersão dos dados nas lojas de pequena dimensão.

#### **Medidas de forma (distribuição dos dados):**

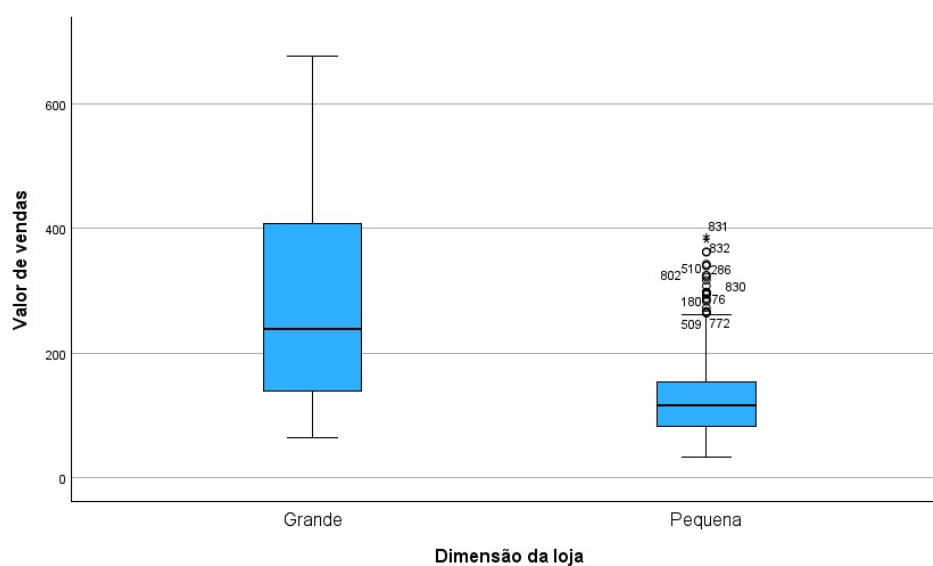
Por fim, avaliando também os valores de assimetria e curtose, constata-se que o valor de assimetria de 1,306 indica uma inclinação mais acentuada à direita, o que sugere que a maioria dos valores é menor do que o valor médio, mas que alguns valores altos estão a distorcer a distribuição.

O valor de curtose de 2,046 indica que a distribuição é leptocúrtica, ou seja, tem caudas mais alongadas do que a distribuição normal, o que significa que há mais valores extremos (outliers) do que numa distribuição normal, com os dados mais dispersos nas caudas.

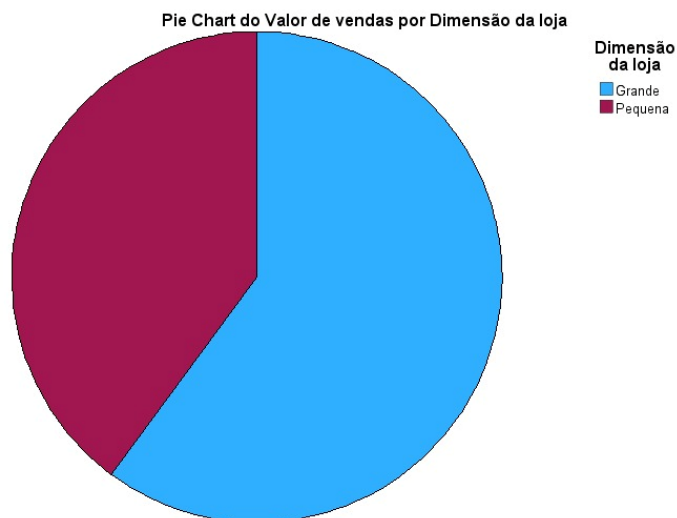
Os gráficos abaixo permitem ilustrar de uma forma mais visual o que foi descrito na análise descritiva.



**Figura 2: Distribuição do Valor de Vendas por Dimensão da Loja: Histograma**



**Figura 3: Distribuição do Valor de Vendas por Dimensão da Loja: Boxplot**



***Figura 4: Distribuição do Valor de Vendas por Dimensão da Loja: Pie Chart***

Em conclusão, a análise descritiva realizada revelou que as lojas de grande dimensão têm vendas significativamente mais elevadas, mas com maior dispersão e variabilidade, refletindo uma distribuição mais achatada e levemente assimétrica, do que as vendas das lojas de pequena dimensão. Por outro lado, as lojas de pequena dimensão apresentam vendas mais concentradas em torno de valores menores, com uma distribuição mais assimétrica e caudas mais alongadas, sugerindo que poucas lojas superam substancialmente a média, a maioria regista vendas relativamente baixas. Estas diferenças indicam padrões de vendas distintos entre os dois tipos de lojas, grandes e pequenas.

## **2.2 Questão 2: Analise a correlação entre Vendas e Despmark (justifique a escolha do coeficiente de correlação usado e interprete os resultados obtidos).**

Como se pretende aferir a correlação entre duas variáveis quantitativas, valor de vendas e despesas de marketing, utilizou-se para o efeito o Coeficiente de Correlação de Pearson e obtiveram-se os resultados apresentados na tabela abaixo.

### Correlações

		Valor de vendas	Despesas de marketing
Valor de vendas	Correlação de Pearson	1	,641**
	Sig. (2 extremidades)		<,001
	N	850	850
Despesas de marketing	Correlação de Pearson	,641**	1
	Sig. (2 extremidades)	<,001	
	N	850	850

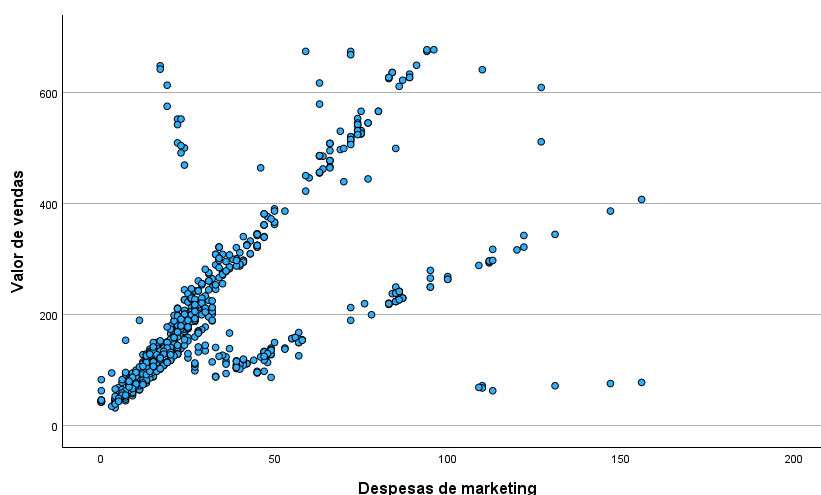
\*\* . A correlação é significativa no nível 0,01 (2 extremidades).

**Tabela 3: Tabela de Correlação de Pearson entre Valor de Vendas e Despesas de Marketing**

Tendo por base a informação da Tabela 3, o coeficiente de correlação de Pearson (r) entre o valor de venda e as despesas com marketing é de 0,641, o que indica uma correlação positiva moderada a forte, sugerindo que, à medida que as despesas de marketing aumentam, as vendas também tendem a aumentar.

Em termos de significância estatística, o valor -p (Sig. (2 extremidades)) para a correlação é <0,001, o que indica que a correlação observada é estatisticamente significativa ao nível de significância de 0,01, o que confirma que o aumento das despesas de marketing está efetivamente associado ao aumento nas vendas.

Esta correlação pode ser confirmada pelo gráfico de dispersão abaixo, onde os pontos tendem a formar uma linha ascendente. Isso confirma que, à medida que as despesas de marketing aumentam, as vendas também tendem a aumentar.



**Figura 5: Gráfico de Dispersão entre Vendas e Despesas de Marketing**

### 2.3 **Questão 3: Verificação da existência de diferenças significativas entre as médias das vendas dos dois tipos de café, utilizando teste estatístico apropriado.**

De forma a verificar a existência de diferenças estatisticamente significativas entre as médias das vendas dos dois tipos de café, o teste t-Student para amostras independentes trata-se do teste adequado para comparar as médias de dois grupos independentes (neste caso, os dois tipos de café) em relação a uma variável quantitativa dependente (neste caso, as vendas).

O teste assume dois pressupostos: que a variável quantitativa segue uma distribuição normal dentro de cada grupo e a homogeneidade das variâncias entre os dois grupos. Posto isto, é necessário verificar se os dois pressupostos referidos se verificam, antes de avançarmos para a realização do teste.

De forma a confirmar o pressuposto de distribuição normal em cada grupo, testaram-se as seguintes hipóteses:

**H0:** A variável Vendas segue uma distribuição normal dentro de cada grupo (tipo de café).

**H1:** A variável Vendas não segue uma distribuição normal dentro de cada grupo (tipo de café).

Obtiveram-se os seguintes resultados em SPSS:

Testes de Normalidade							
	tipo_cafe	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Estatística	gl	Sig.	Estatística	gl	Sig.
Valor de vendas	1,00	,180	394	<,001	,797	394	<,001
	2,00	,187	456	<,001	,811	456	<,001

a. Correlação de Significância de Lilliefors

**Tabela 4: Testes de Normalidade para o Valor das Vendas por Tipo de Café**

Como o valor- $p < 0,01 < \alpha = 0,05$ , no teste de normalidade Kolmogorov-Smirnov (indicado para amostras de dimensão  $n > 200$ , que é o caso), para os dois grupos (Tipo de Café 1 e 2), rejeita-se H0 e conclui-se que a variável vendas não segue uma distribuição normal dentro de cada grupo.

No entanto, dado que ambos os grupos possuem um número de observações superior a 30 (Tipo café 1:  $n = 394$ ; Tipo café 2:  $n = 456$ ), como se comprova pela Tabela 5, podemos usar o Teorema do Limite Central como justificação para prosseguir com o teste t-Student para amostras independentes, mesmo que os dados não sejam normalmente distribuídos. O referido teorema afirma que, se o tamanho da amostra for suficientemente grande ( $n \geq 30$ ), a distribuição das médias amostrais tende a aproximar-se de uma distribuição normal, independentemente da distribuição original dos dados. Desta forma, pode prosseguir-se com a análise.

Estatísticas de grupo					
	tipo_cafe	N	Média	Desvio Padrão	Erro de média padrão
Valor de vendas	1,00	394	199,66	140,174	7,062
	2,00	456	186,00	142,197	6,659

***Tabela 5: Estatísticas Descritivas das Vendas por Tipo de café***

De forma a averiguar o pressuposto da homogeneidade das variâncias dos dois grupos, realizou-se o teste de Levene, onde se definiram as seguintes hipóteses:

**H0** = As variâncias são iguais entre os dois grupos (Tipo de Café 1 e Tipo de Café 2)

**H1** = As variâncias são diferentes entre os dois grupos (Tipo de Café 1 e Tipo de Café 2)

No SPSS obteve-se a seguinte tabela:

Teste de amostras independentes											
		Teste de Levene para igualdade de variâncias		teste-t para Igualdade de Médias						95% Intervalo de Confiança da Diferença	
		Z	Sig.	t	df	Significância		Diferença média	Erro de diferença padrão	Inferior	Superior
Valor de vendas	Variâncias iguais assumidas	,416	,519	1,406	848	,080	,160	13,660	9,716	-5,411	32,731
	Variâncias iguais não assumidas			1,407	833,446	,080	,160	13,660	9,706	-5,391	32,712

***Tabela 6: Teste de Levene para Igualdade de Variâncias e Teste t de Amostras Independentes para Comparação das Vendas entre os Dois Tipos de Café***

O resultado do teste de Levene mostrou um valor- $p = 0,519 > \alpha = 0,05$ , o que nos leva a aceitar a hipótese nula (H0) de que as variâncias são iguais, e, portanto, deve utilizar-se a linha "Variâncias iguais assumidas" para interpretar o teste t-Student.

Após a verificação dos dois pressupostos, pode avançar-se para a realização do teste t-Student. Nesse sentido, formularam-se as seguintes hipóteses:

**H0** = Não há diferenças significativas entre as médias das vendas dos dois tipos de café

**H1** = Há diferenças significativas entre as médias das vendas dos dois tipos de café

O valor- $p = 0,160 > \alpha = 0,05$ , leva-nos a aceitar a hipótese nula (H0), concluindo-se que não há diferenças significativas entre as médias das vendas dos dois tipos de café.

De forma a complementar esta análise, analisou-se adicionalmente o valor d de Cohen, de forma a medir a magnitude da diferença entre as médias das vendas dos dois tipos de café.

Tamanhos de efeitos de amostras independentes					
		Padronizador <sup>a</sup>	Estimativa de ponto	Intervalo de Confiança 95%	
Valor de vendas	d de Cohen	141,263	,097	-,038	,232
	Correção de Hedges	141,388	,097	-,038	,231
	Delta do vidro	142,197	,096	-,039	,231

a. O denominador usado na estimativa dos tamanhos dos efeitos.  
 O d de Cohen usa o desvio padrão agrupado.  
 A correção de Hedges usa o desvio padrão agrupado, além de um fator de correção.  
 O delta de Glass usa o desvio padrão de amostra do grupo de controle (ou seja, o segundo) grupo.

**Tabela 7: Medidas de Tamanho de Efeito para Comparação das Vendas entre os Dois Tipos de Café**

Um valor d de Cohen = 0,097 indica um efeito muito pequeno. O intervalo de confiança de 95% para o d de Cohen varia de -0,38 a 0,232, o que inclui zero, reforçando a conclusão de que não há evidência suficiente para afirmar que existe uma diferença significativa entre as médias de vendas dos dois tipos de café.

Podemos assim concluir que não existem diferenças significativas entre as médias das vendas dos 2 tipos de café.

#### **2.4 Questão 4: Teste se há diferenças significativas entre as vendas (reais) e as vendas esperadas utilizando um teste estatístico apropriado.**

Nesta situação, é esperado, do ponto de vista teórico, que haja uma associação significativa entre as vendas reais e as vendas esperadas. Posto isto, caso essa associação



se confirme, o teste estatístico mais adequado será o teste t-Student para amostras emparelhadas.

Assim, começou-se por analisar o grau de associação entre as duas variáveis. Uma vez que ambas as variáveis são quantitativas, foi aplicado o teste de correlação de Pearson, obtendo-se os resultados abaixo:

**Correlações de amostras emparelhadas**

			Correlação	Significância	
				Unilateral p	Bilateral p
Par 1	Valor de vendas & Valor esperado de vendas	850	,968	<,001	<,001

***Tabela 8: Correlação entre as vendas reais e as vendas esperadas***

O valor do coeficiente de correlação de Pearson é de 0,968, o que indica uma correlação muito forte e positiva entre as vendas reais e as vendas esperadas. Isto sugere que, à medida que o valor das vendas esperadas aumenta, o valor das vendas reais também tende a aumentar. Esta forte correlação indica que as vendas reais estão intimamente relacionadas com as expectativas de vendas, o que pode ser um indicador de que as previsões de vendas são feitas corretamente.

O valor- $p < 0,001 < \alpha = 0,01$ , indica que há uma probabilidade extremamente baixa de que a correlação observada tenha ocorrido por acaso.

Em SPSS obtiveram-se as seguintes estatísticas:

**Estatísticas de amostras emparelhadas**

		Média	N	Desvio Padrão	Erro de média padrão
Par 1	Valor de vendas	192,33	850	141,344	4,848
	Valor esperado de vendas	193,65	850	152,414	5,228

***Tabela 9: Estatísticas descritivas vendas reais e esperadas***

A comparação entre o valor das vendas reais e o valor esperado mostra que, em média, as vendas reais (192,33 dólares) ficaram ligeiramente abaixo das expectativas (193,65 dólares). Embora a diferença entre as médias seja pequena, o desvio padrão mais elevado nas vendas esperadas (152,41 dólares) sugere maior variabilidade nas previsões em comparação com os resultados reais (141,34 dólares).

Dado que se confirma a associação entre as variáveis, o teste estatístico mais adequado será o teste t-Student para amostras emparelhadas.

O teste t-Student para amostras emparelhadas assume que as diferenças entre as variáveis seguem uma distribuição normal. De forma a confirmar o pressuposto, testaram-se as seguintes hipóteses:

**H0:** As diferenças entre as vendas reais e as vendas esperadas seguem uma distribuição normal.

**H1:** As diferenças entre as vendas reais e as vendas esperadas não seguem uma distribuição normal.

Obtiveram-se os seguintes resultados em SPSS:

<b>Testes de Normalidade</b>						
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estatística	gl	Sig.	Estatística	gl	Sig.
Valor de vendas	,180	850	<,001	,809	850	<,001
Valor esperado de vendas	,185	850	<,001	,792	850	<,001

a. Correlação de Significância de Lilliefors

***Tabela 10: Testes de Normalidade para Vendas e Vendas Esperadas***

Como o valor- $p < 0,01 < \alpha = 0,05$  no teste de normalidade Kolmogorov-Smirnov (indicado para amostras de maior dimensão, que é o caso), para as duas variáveis (Vendas e Vendas Esperadas) rejeita-se H0 e conclui-se que as diferenças entre as variáveis não seguem uma distribuição normal.

No entanto, dado que ambas as variáveis possuem números de observações maiores que 30 (Vendas:  $n = 850$ ; Vendas esperadas:  $n = 850$ ), como se comprova pela Tabela 11, podemos usar o Teorema do Limite Central referido acima como justificação para prosseguir com o teste t-Student para amostras emparelhadas, mesmo que os dados não sejam normalmente distribuídos. Desta forma, pode prosseguir-se com a análise.

**Resumo de processamento de casos**

	Válido		Casos Omisso		Total	
	N	Porcentagem	N	Porcentagem	N	Porcentagem
Valor de vendas	850	100,0%	0	0,0%	850	100,0%
Valor esperado de vendas	850	100,0%	0	0,0%	850	100,0%

**Tabela 11: Estatísticas descritivas das vendas por tipo de café**

Para a realização do teste t-Student para amostras emparelhadas, formularam-se as seguintes hipóteses:

**H0:** Inexistência de diferenças significativas entre os valores de vendas reais e os valores de vendas esperados.

**H1:** Existência de diferenças significativas entre os valores de vendas reais e os valores de vendas esperados.

Em SPSS obtiveram-se os seguintes resultados:

Teste de amostras emparelhadas										
		Diferenças emparelhadas						Significância		
		Média	Desvio Padrão	Erro de média padrão	95% Intervalo de Confiança da Diferença		t			
					Inferior	Superior				
Par 1	Valor de vendas - Valor esperado de vendas	-1,312	39,008	1,338	-3,938	1,314	-,980	849	,164	,327

**Tabela 12: Teste de Amostras Emparelhadas para Vendas e Vendas Esperadas**

A análise do teste de amostras emparelhadas indica que a diferença média entre o valor das vendas reais e o valor de vendas esperado é de -1,312, com um desvio padrão de 39,008.

O valor de  $t = -0,980$  e o valor- $p = 0,327 > \alpha = 0,05$  indicam que não há evidências estatísticas para rejeitar a hipótese nula, de que não existem diferenças significativas entre as vendas reais e as vendas esperadas. O intervalo de confiança de 95% (de -3,938 a 1,314) abrange o valor zero, o que confirma a conclusão retirada da não rejeição de H0. A discrepância observada entre as médias das vendas reais e esperadas pode ser atribuída à variação aleatória e não a uma diferença significativa entre os dois grupos.

De forma a complementar a análise, analisamos adicionalmente o valor d de Cohen, no entanto, uma vez que o teste t-Student para amostras emparelhadas não mostrou uma diferença significativa entre as vendas reais e as vendas esperadas, não seria necessário analisar o valor d de Cohen, pois este irá apenas reforçar a conclusão já retirada.

**Tamanhos de efeitos de amostras em pares**

Par 1	Valor de vendas - Valor esperado de vendas	d de Cohen	Padronizador <sup>a</sup>	Estimativa de ponto	Intervalo de Confiança 95%	
					Inferior	Superior
			39,008	-,034	-,101	,034
		Correção de Hedges	39,042	-,034	-,101	,034

a. O denominador usado na estimativa dos tamanhos dos efeitos.

O d de Cohen usa o desvio padrão de amostra da diferença média.

A correção de Hedges usa o desvio padrão de amostra da diferença média, além de um fator de correção.

***Tabela 13: Medidas de Tamanho de Efeito para Comparação das Vendas com as Vendas esperadas***

Um valor d de Cohen = 0,034 indica um efeito muito pequeno. O intervalo de confiança de 95% para o d de Cohen varia de -0,101 a 0,034, o que inclui zero, reforçando a conclusão de que não há evidência suficiente para afirmar que existe uma diferença significativa entre as vendas reais e as vendas esperadas de café.

## **2.5 Questão 5: Verifique, usando um teste estatístico adequado, se a região tem impacto significativo nas vendas.**

Para verificar se a região tem um impacto significativo nas vendas, o teste estatístico adequado é o One-Way ANOVA (Análise de Variância). Este teste é adequado quando se pretende comparar as médias de uma variável dependente entre três ou mais grupos, neste caso, permite comparar as médias das vendas entre várias regiões e verificar se pelo menos uma delas difere significativamente das outras.

Posto isto, podemos desde logo definir 2 hipóteses:

**H<sub>0</sub>** = As médias das vendas são iguais em todas as regiões (a região não tem impacto significativo nas vendas).

**H<sub>1</sub>** = Pelo menos uma das médias das vendas difere significativamente (a região tem um impacto significativo nas vendas).

No entanto, antes de avançarmos para o teste ANOVA é necessário, em primeiro lugar, testar três pressupostos: independência das amostras, distribuição normal da variável dependente (Vendas) e homogeneidade das variâncias.

No que diz respeito à independência das amostras, o pressuposto pode dar-se como garantido decorrente da aleatoriedade da seleção da amostra.

Relativamente ao pressuposto de homogeneidade das variâncias das vendas entre regiões, este foi testado usando o teste de Levene. Nesse sentido, definiram-se as seguintes hipóteses:

**H<sub>0</sub>** = As variâncias das vendas são iguais entre as regiões.

**H<sub>1</sub>** = Pelo menos uma das variâncias das vendas é diferente entre as regiões.

No SPSS obteve-se a seguinte tabela:

Testes de homogeneidade de variâncias					
		Estatística de Levene	df1	df2	Sig.
Valor de vendas	Com base em média	,717	2	847	,489
	Com base em mediana	,157	2	847	,855
	Com base em mediana e com gl ajustado	,157	2	812,443	,855
	Com base em média aparada	,264	2	847	,768

**Tabela 14: Teste de Levene para Homogeneidade das Variâncias das Vendas entre Regiões**

O resultado do teste de Levene apresentou um valor- $p=0,489 > \alpha = 0,05$ , logo, não se rejeita H<sub>0</sub>, ou seja, assume-se a homogeneidade das variâncias.

Para avaliar a normalidade das distribuições das vendas em cada grupo, utilizamos o teste de Kolmogorov-Smirnov. Foram definidas as seguintes hipóteses:

**H<sub>0</sub>**: A distribuição das vendas segue uma distribuição normal dentro de cada região.

**H<sub>1</sub>**: A distribuição das vendas não segue uma distribuição normal dentro de cada região.

Obtiveram-se os seguintes resultados em SPSS:

Testes de Normalidade							
	Região	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Estatística	gl	Sig.	Estatística	gl	Sig.
Valor de vendas	Centro	,126	287	<,001	,869	287	<,001
	Norte	,225	305	<,001	,783	305	<,001
	Sul	,218	258	<,001	,757	258	<,001

a. Correlação de Significância de Lilliefors

**Tabela 15: Testes de Normalidade para o Valor das Vendas por Região**

Como o valor- $p < 0,01 < \alpha = 0,05$  no teste de normalidade Kolmogorov-Smirnov (indicado para amostras de maior dimensão, que é o caso), para as três variáveis (Região Centro, Norte e Sul) rejeita-se  $H_0$  e, portanto, podemos concluir que a distribuição das vendas não segue uma distribuição normal nas três regiões (Centro, Norte e Sul).

No entanto, dado que ambas as variáveis possuem números de observações superiores a 30 (Norte:  $n = 305$ ; Centro:  $n = 287$ ; Sul:  $n = 258$ ), como se comprova pela Tabela 16, podemos usar o Teorema do Limite Central referido acima como justificação para prosseguir com o teste ANOVA, mesmo que os dados não sejam normalmente distribuídos. Desta forma a prosseguir com a análise.

Postos			
	Região	N	Posto médio
Valor de vendas	Norte	305	424,06
	Centro	287	430,71
	Sul	258	421,40
	Total	850	

**Tabela 16: Dimensão das amostras por região**

Posto isto, estamos em condições de avançar para a análise do teste ANOVA, com base nas hipóteses definidas acima.

Resultados obtidos no SPSS:

# ANOVA

Valor de vendas

	Soma dos Quadrados	df	Quadrado Médio	F	Sig.
Entre Grupos	1870,149	2	935,074	,047	,954
Nos grupos	16959648,962	847	20023,198		
Total	16961519,111	849			

**Tabela 17: Teste ANOVA**

## Tamanhos do efeito do ANOVA<sup>a,b</sup>

		Estimativa de ponto	Intervalo de Confiança 95%	
			Inferior	Superior
Valor de vendas	Eta quadrado	,000	,000	,001
	Epsilon quadrado	-,002	-,002	-,001
	Efeito fixo do Omega quadrado	-,002	-,002	-,001
	Efeito aleatório do Omega quadrado	-,001	-,001	,000

a. Eta quadrado e Epsilon quadrado são estimados com base no modelo de efeito fixo.

b. As estimativas negativas, mas menos tendenciosas, são mantidas, não arredondadas para zero.

**Tabela 18: Tamanho do efeito ANOVA**

Com base nos resultados obtidos, podemos concluir que não existem diferenças estatisticamente significativas nas médias de vendas entre as regiões analisadas. O valor- $p = 0,954 > \alpha = 0,05$ , sugere que as variações nas vendas observadas entre as diferentes regiões podem ser atribuídas ao acaso, e não a fatores significativos relacionados com a localização geográfica. Assim, a hipótese nula de que não há diferenças nas médias das vendas entre as regiões permanece válida.

Caso o resultado do teste ANOVA indicasse a existência de diferenças significativas entre as médias de vendas das várias regiões, adicionalmente, teriam de ser realizados testes post-hoc, de forma a identificar quais grupos são significativamente diferentes entre si. No entanto, neste caso, esses testes não são necessários.



## CONCLUSÃO

O presente trabalho teve como principal objetivo aplicar os conceitos teóricos aprendidos na disciplina de Análise Estatística de Dados a um caso prático de análise de vendas de café numa cadeia de lojas. Através da utilização do software estatístico SPSS29, foram realizadas análises descritivas e testes estatísticos para avaliar os fatores que mais influenciam o volume de vendas.

A análise descritiva mostrou diferenças claras entre as vendas de lojas de diferentes dimensões, com as lojas de grande dimensão a apresentar volumes de vendas mais elevados, mas também uma maior variabilidade, enquanto as lojas de pequena dimensão tiveram vendas mais concentradas em torno de valores menores.

A correlação entre as despesas de marketing e as vendas revelou que um aumento nas despesas de marketing está associado a um aumento nas vendas, indicando que este é um fator significativo no desempenho das lojas.

No entanto, ao comparar as vendas dos dois tipos de café mais vendidos, a análise não encontrou diferenças significativas, sugerindo que ambos os tipos de café têm desempenhos de vendas semelhantes.

Da mesma forma, não foram encontradas diferenças estatisticamente significativas entre as vendas reais e as vendas esperadas, indicando que as previsões de vendas feitas pela empresa estão alinhadas com o desempenho real.

Por último, a análise da influência da localização geográfica das lojas nas vendas revelou que a região em que uma loja se encontra (Norte, Centro ou Sul) não tem um impacto significativo nas vendas.

Com base nos resultados obtidos, conclui-se que as despesas associadas à promoção do produto têm um impacto positivo e significativo nas vendas de café. Em contrapartida, os restantes fatores analisados não demonstraram influência significativa. Assim, recomenda-se que a estratégia da cadeia de lojas se concentre em aumentar o investimento em ações de marketing, uma vez que estas se mostram fundamentais para o sucesso comercial da empresa.