

CTeSP

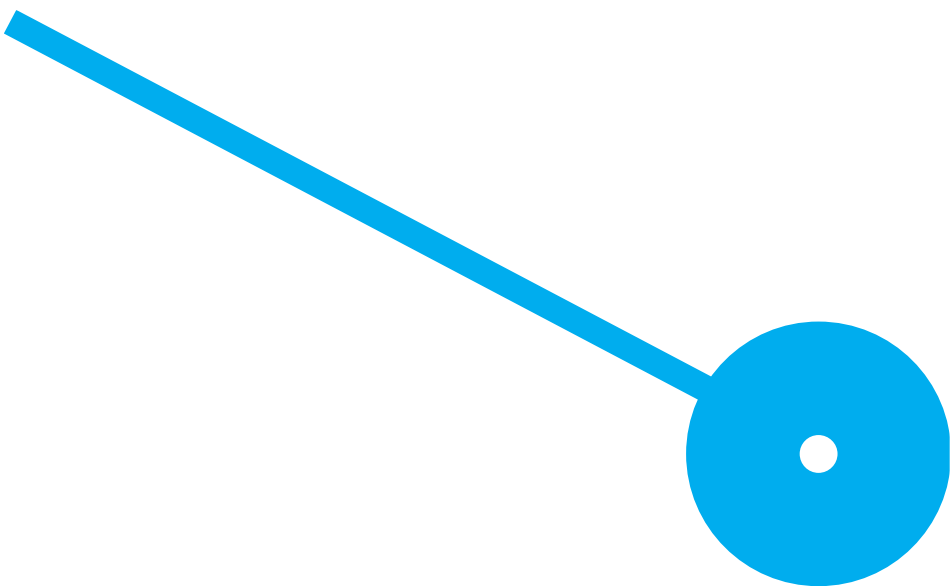
CURSO TÉCNICO SUPERIOR PROFISSIONAL
DESENVOLVIMENTO PARA A WEB E DISPOSITIVOS MÓVEIS

Oficina de Estatística – Tema 6

Alex Oliveira – 8240841

Ricardo Lourenço - 8240764

01/2026



Índice

Índice de Figuras	3
Índice de Tabelas	4
1 Introdução	5
1.1 Contextualização	5
1.2 Objetivos do Trabalho	5
1.3 Caracterização das Variáveis	6
2 Estatística descritiva	7
2.1 Caracterização da população da amostra	7
2.2 Análise univariada das variáveis quantitativas	8
2.2.1 Análise do Lucro	9
2.2.2 Análise das Reparações	9
2.2.3 Análise do Número de Avarias	10
2.3 Identificação de Outliers	11
2.4 Comparações entre Grupos	12
2.4.1 Comparação por Género do Condutor	12
2.4.2 Comparação por Perfil de Cliente	14
2.5 Análise de Correlação e Regressão Linear	15
2.5.1 Modelo 1: Lucro vs Reparações	15
2.5.2 Modelo 2: Lucro vs N Avarias	15
2.5.3 Modelo 3: Reparações vs N Avarias	16
2.6 Síntese da Análise Descritiva	18
3 Estatística Indutiva	19
3.1 Testes de Normalidade	19
3.2 Intervalos de Confiança para Médias Populacionais	20
3.3 Comparação de Médias entre Grupos	21
3.3.1 Comparação do Lucro por Género do Condutor	21
3.3.2 Comparação do Lucro por Perfil de Cliente	22

<u>3.4Significância dos Modelos de Regressão Linear</u>	23
<u>3.4.1Modelo 1: Lucro ~ Reparações</u>	23
<u>3.4.2Modelo 2: Lucro ~ N Avarias</u>	24
<u>3.4.3Modelo 3: Reparações ~ N Avarias</u>	24
<u>3.4.4Síntese dos Modelos</u>	24
<u>3.5Análise de Resíduos do Melhor Modelo</u>	25
<u>3.5.1Normalidade dos Resíduos</u>	25
<u>3.5.2Independência dos Resíduos</u>	25
<u>3.5.3Verificação da Variabilidade entre Grupos</u>	25
<u>3.6Previsões com os Modelos de Regressão</u>	26
<u>3.6.1Previsões "In the Box"</u>	26
<u>3.6.2Previsões "Out of the Box"</u>	27
<u>3.7Síntese da Análise Inferencial</u>	27
<u>4Conclusão</u>	29
<u>4.1Concretização dos Objetivos Propostos</u>	29
<u>4.2Dificuldades Encontradas</u>	30
<u>4.3Passos Futuros</u>	30

1 Introdução

1.1 Contextualização

O presente relatório foi desenvolvido no âmbito da Unidade Curricular de Oficina de Estatística e incide sobre a análise de dados do Tema 6: Stand Automóvel. O estudo foca-se na atividade de um negociante do ramo automóvel que recolheu dados relativos a 17 vendas de veículos usados, todos comercializados com uma garantia contratual de um ano.

A gestão deste setor exige uma compreensão profunda de como as variáveis demográficas, o perfil de fidelização do consumidor e a fiabilidade mecânica do produto interagem para determinar a rentabilidade final. A análise estatística dos dados de vendas permite identificar padrões, testar hipóteses sobre fatores que influenciam o lucro e desenvolver modelos preditivos que apoiem a tomada de decisão empresarial.

Para o efeito, procedeu-se ao tratamento dos dados utilizando o software estatístico R, permitindo uma análise rigorosa que vai desde a descrição exploratória à modelação preditiva, passando pela inferência estatística e validação de pressupostos dos modelos.

1.2 Objetivos do Trabalho

Este estudo estruturou-se para cumprir os seguintes objetivos fundamentais, conforme definido no enunciado:

Estatística Descritiva:

- Caracterizar a população e a amostra em estudo;
- Calcular medidas de localização (média, mediana, quartis), dispersão (desvio-padrão, amplitude interquartil) e forma (assimetria, curtose);
- Criar representações gráficas adequadas (histogramas, boxplots, gráficos de barras e diagramas de dispersão);
- Identificar a presença de outliers através do critério do intervalo interquartil;
- Comparar distribuições entre grupos (género do condutor e perfil de cliente);
- Desenvolver modelos de regressão linear simples e analisar correlações entre variáveis;
- Interpretar os coeficientes de correlação e determinação e avaliar o ajustamento dos modelos.

Estatística Inferencial:

- Identificar distribuições teóricas através do teste de normalidade de Shapiro-Wilk;
- Estimar médias populacionais através de Intervalos de Confiança a 95%;
- Aplicar testes paramétricos (teste t de Student e ANOVA) para comparação de médias entre grupos;

- Aplicar testes não-paramétricos (Wilcoxon-Mann-Whitney e Kruskal-Wallis) como alternativa robusta;
- Avaliar a significância estatística dos coeficientes de regressão;
- Analisar os pressupostos dos modelos de regressão (normalidade e independência dos resíduos);
- Realizar previsões "in the box" (interpolação) e "out of the box" (extrapolação).

1.3 Caracterização das Variáveis

As variáveis em estudo foram classificadas de acordo com a sua natureza técnica:

Variável	Tipo	Descrição
Condutor	Qualitativa nominal	Género do condutor habitual (F=Feminino; M=Masculino)
Avaliação	Qualitativa Ordinal	Perfil do cliente (1=Muito Frequente a 3=Pouco Frequente)
N_Avarias	Quantitativa Discreta	Número de avarias durante o período de garantia
Lucro	Quantitativa Contínua	Margem líquida obtida na venda (em Euros)
Reparações	Quantitativa Contínua	Valor das despesas de reparações (em euros)

Tabela 1 - Caracterização das variáveis do modelo de dados

2 Estatística descritiva

Neste capítulo, procede-se à análise exploratória dos dados, visando descrever as características principais da amostra de 17 veículos vendidos. Esta análise permite uma primeira compreensão das tendências de lucro, custos de reparação e perfis de clientes através de medidas estatísticas, representações gráficas e análise de relações entre variáveis, antes da aplicação de testes inferenciais.

2.1 Caracterização da população da amostra

População: Todas as vendas de veículos usados com garantia de 1 ano realizadas pelo stand automóvel.

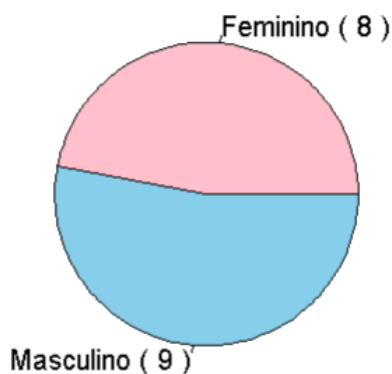
Amostra: 17 vendas de veículos selecionadas para análise.

A amostra é composta por 17 observações, apresentando uma distribuição equilibrada quanto ao género dos condutores (47,1% feminino e 52,9% masculino) e uma diversidade relevante quanto à fidelização dos clientes, com predominância de clientes muito frequentes (41,2%).

Variável	Categoria	Freq. Absoluta (n)	Freq. Relativa (%)
Condutor	Feminino (F)	8	47,1%
	Masculino (M)	9	52,9%
Avaliação	1 (Muito Frequente)	7	41,2%
	2 (Frequente)	5	29,4%
	3 (Pouco Frequente)	5	29,4%

Tabela 2 - Frequências Absolutas e Relativas da Amostra

Distribuição de Veículos por Condutor



Distribuição da Avaliação dos Veículos (%)

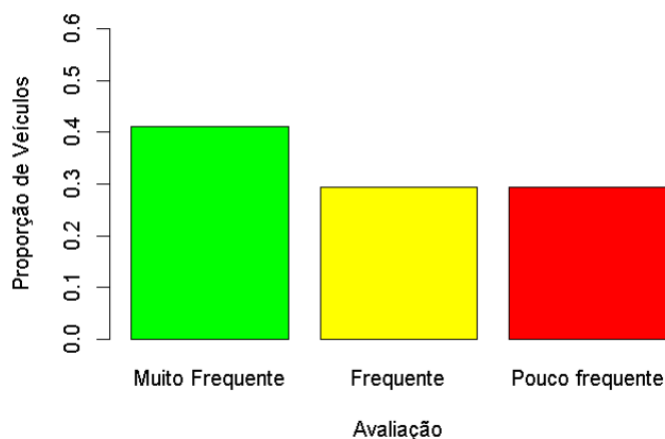


Figura 1 - Distribuição da amostra por Género e Perfil de Cliente

2.2 Análise univariada das variáveis quantitativas

Nesta secção, apresentam-se as medidas de localização, dispersão e forma para as três variáveis quantitativas: Lucro, Reparações e Número de Avarias.

Medida	Lucro (€)	Reparações (€)	N_Avarias
Mínimo	129,70	23,00	0
Q1 (25%)	825,50	67,00	2
Mediana	1118,10	96,00	3
Média	1002,82	97,71	3,29
Q3 (75%)	1373,00	135,00	5
Máximo	1635,60	167,00	6
Desvio-Padrão	465,91	40,05	1.79
CV (%)	46,5%	41,0%	54,5%
Assimetria	-0,576	-0,126	0,088
Curtose	2,09	2,50	1,75

Tabela 3 - Medidas Descritivas das Variáveis Métricas

2.2.1 Análise do Lucro

O lucro médio por venda situa-se nos 1002,82€, com uma mediana de 1118,10€. O desvio-padrão de 465,91€ e o coeficiente de variação de 46,5% indicam variabilidade moderada, sugerindo que diferentes vendas apresentam rentabilidades substancialmente distintas.

A assimetria negativa (-0,576) revela uma distribuição com cauda à esquerda, indicando que existe uma ligeira concentração de valores na parte superior da distribuição (lucros mais elevados), o que é positivo para o negócio. A curtose de 2,09 (próxima de 3) sugere uma distribuição aproximadamente mesocúrtica, semelhante à distribuição normal.

Os valores oscilam entre 129,70€ (mínimo) e 1635,60€ (máximo), com 50% dos lucros situados entre 825,50€ (Q1) e 1373,00€ (Q3), revelando uma amplitude interquartil de 547,50€.

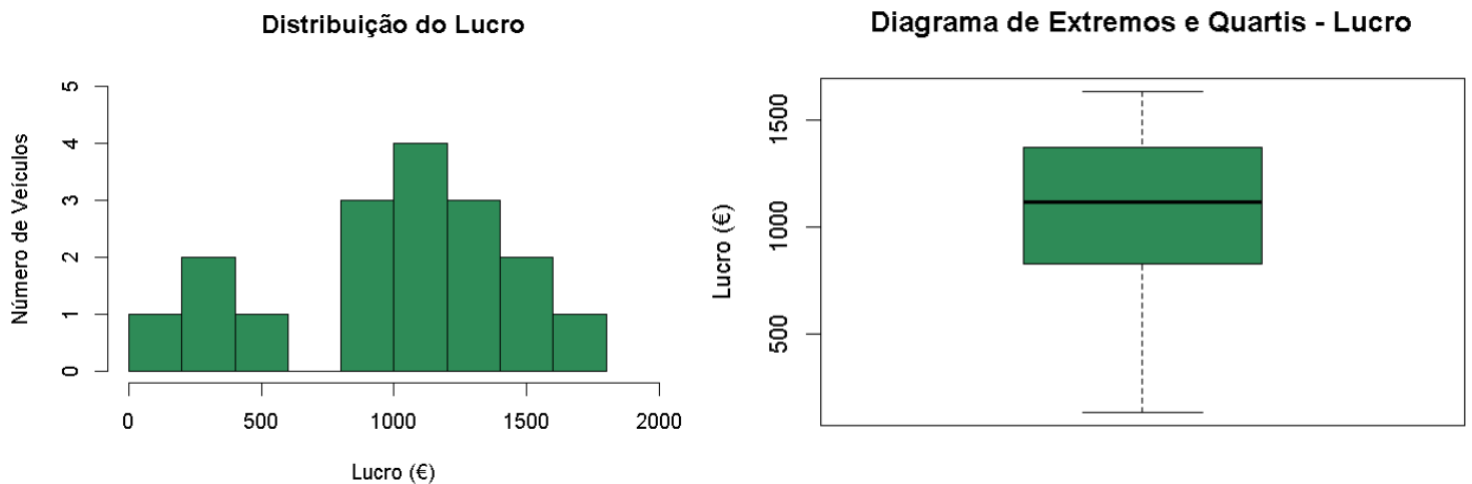


Figura 2 - Histograma e Boxplot da distribuição do Lucro

2.2.2 Análise das Reparações

As despesas de reparação apresentam uma média de 97,71€ e mediana de 96,00€, com valores muito próximos que sugerem distribuição simétrica (confirmado pela assimetria de -0,126, próxima de zero).

O desvio-padrão de 40,05€ e coeficiente de variação de 41,0% indicam variabilidade moderada, embora inferior à observada no lucro. As reparações representam aproximadamente 10% do lucro médio, constituindo um custo significativo, mas controlável da operação.

Os valores variam entre 23,00€ e 167,00€, com amplitude interquartil de 68,00€ (entre Q1=67,00€ e Q3=135,00€).

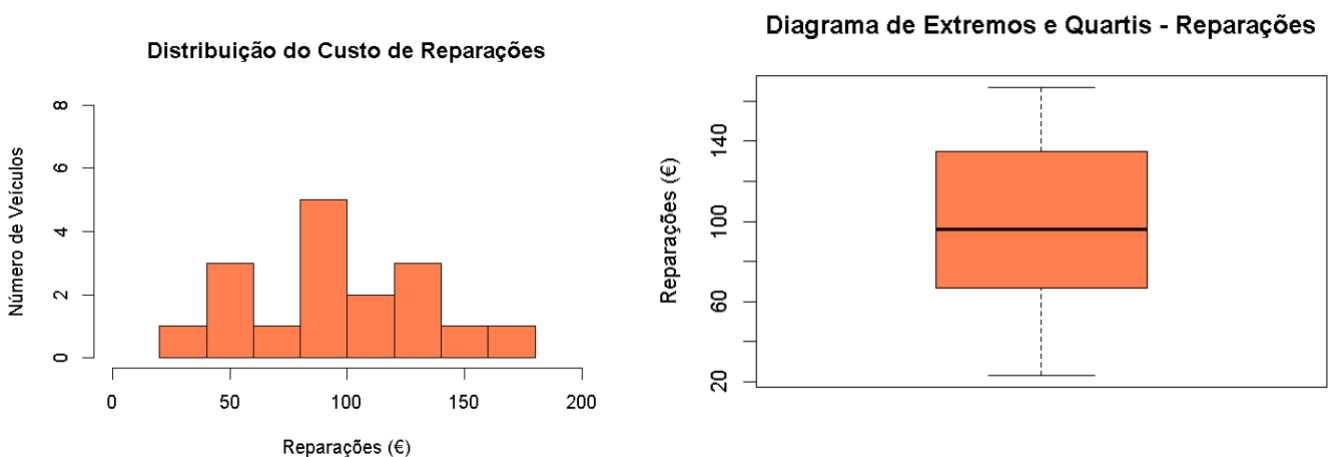


Figura 3 - Histograma e Boxplot da distribuição das Reparações

2.2.3 Análise do Número de Avarias

O número médio de avarias durante o período de garantia é de 3,29, com mediana de 3,00. A distribuição é praticamente simétrica (assimetria = 0,088), variando entre 0 e 6 avarias.

O coeficiente de variação de 54,5% é o mais elevado das três variáveis, refletindo heterogeneidade significativa no estado mecânico dos veículos vendidos. A distribuição de frequências revela que:

- 29,4% dos veículos (5) registaram 3 avarias (valor modal)
- 17,6% (3) registaram 2 ou 5 avarias
- 11,8% (2) registaram 0, 4 ou 6 avarias

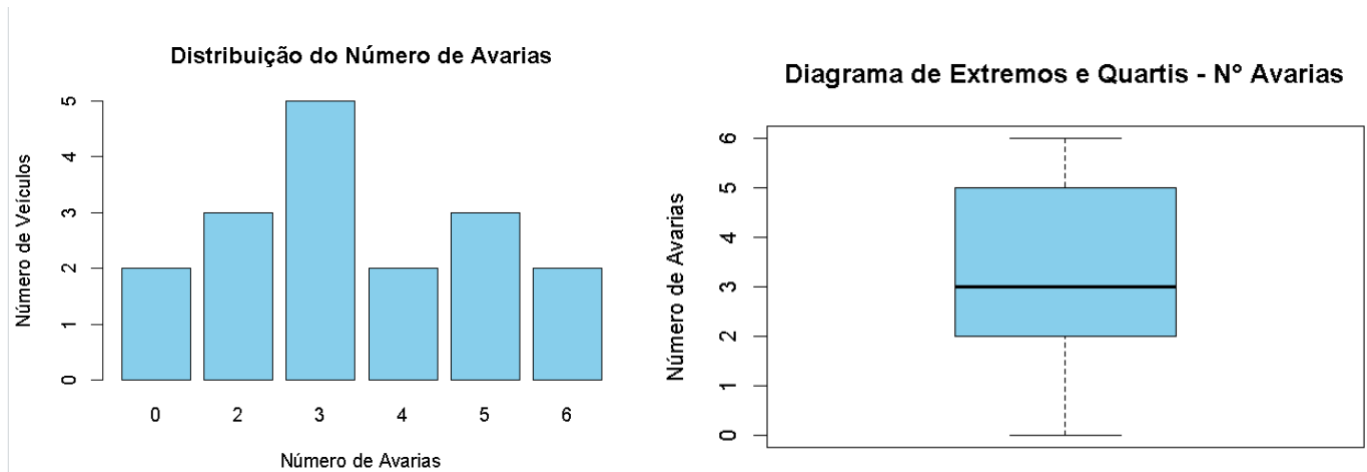


Figura 4 - Gráfico de Barras e Boxplot do Número de Avarias

2.3 Identificação de Outliers

Para verificar a existência de valores atípicos que pudessem distorcer a análise, utilizou-se o critério do intervalo interquartil (IQR). Valores abaixo de $Q1 - 1,5 \times IQR$ ou acima de $Q3 + 1,5 \times IQR$ são considerados outliers.

Análise do Lucro:

- IQR = 547,50€
- Limites: [**-**495,75€ ; 2194,25€]
- Conclusão: Todos os valores estão dentro dos limites. Não há outliers.

Análise das Reparações:

- IQR = 68,00€
- Limites: [**-**35,00€ ; 237,00€]
- Conclusão: Todos os valores estão dentro dos limites. Não há outliers.

Através da representação em diagrama de caixa e bigodes (boxplot), confirmou-se visualmente a inexistência de outliers nas variáveis Lucro e Reparações, validando a robustez da amostra para análises subsequentes.

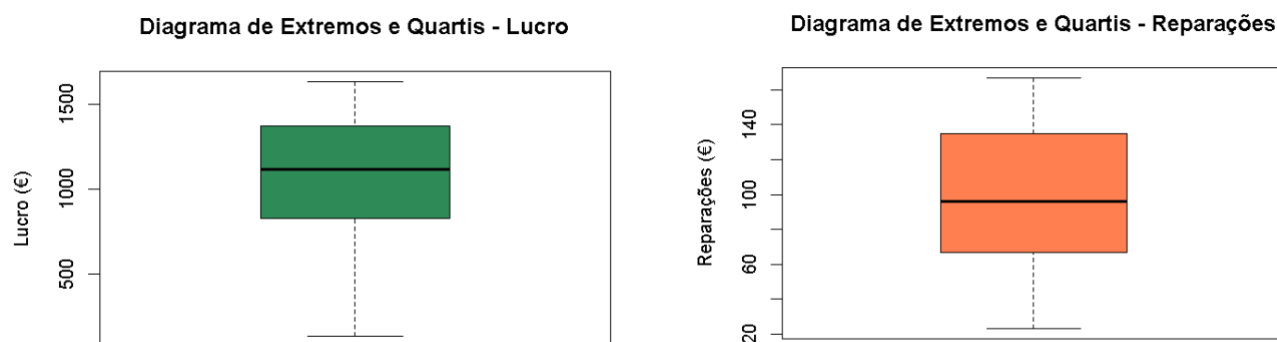


Figura 5 - Boxplots para identificação de outliers

2.4 Comparações entre Grupos

Nesta fase, exploramos como o lucro e as reparações se comportam quando segmentados pelos grupos de interesse: género do condutor e perfil de cliente.

2.4.1 Comparação por Género do Condutor

Lucro: As condutoras femininas geram um lucro médio de 1068,91€ (DP=442,25€), ligeiramente superior aos condutores masculinos, que apresentam 944,08€ (DP=504,74€). Apesar desta diferença de 124,83€ nas médias amostrais, a elevada variabilidade em ambos os grupos (CV>40%) sugere que a diferença pode não ser estatisticamente significativa, hipótese que será testada na secção de inferência.

Reparações: As condutoras femininas registam despesas médias de 113,00€, enquanto os condutores masculinos apresentam 84,11€. Esta diferença de 28,89€ poderá estar relacionada com o tipo de veículos vendidos a cada grupo ou com padrões de utilização distintos.

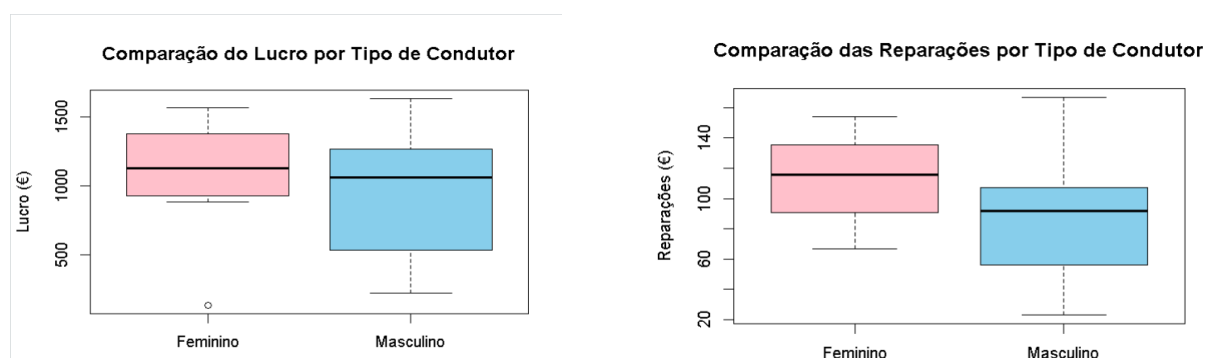


Figura 6 - Comparação do Lucro e Reparações por Tipo/Género do Condutor

2.4.2 Comparação por Perfil de Cliente

A segmentação por perfil de cliente revela diferenças muito acentuadas no lucro:

Perfil	Lucro Médio (€)	Desvio-Padrão (€)
1 - Muito Frequente	1265,27	264,68
2 - Frequente	1116,22	383,33
3 - Pouco Frequente	522,00	395,14

Tabela 4 - Medidas de localização e dispersão do Lucro por Perfil de Cliente (Avaliação)

Os clientes "Muito Frequentes" (Nível 1) geram um lucro médio de 1265,27€, contrastando fortemente com os "Pouco Frequentes" (Nível 3), que geram apenas 522,00€, uma diferença de 143%. Este padrão sugere que a fidelização do cliente está fortemente associada a maior rentabilidade por venda, possivelmente devido a:

- Maior confiança na compra de veículos de maior valor
- Menor necessidade de descontos promocionais
- Perfil de compra mais informado e rápido
-

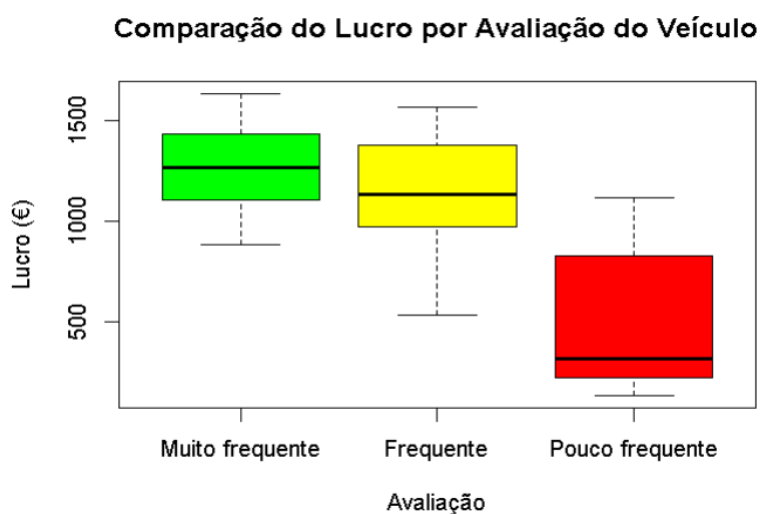


Figura 7 - Comparação do Lucro por Perfil de Cliente

2.5 Análise de Correlação e Regressão Linear

Para finalizar a análise descritiva, estudou-se a relação entre o lucro e as restantes variáveis métricas através do Coeficiente de Correlação de Pearson (r) e do desenvolvimento de modelos de regressão linear simples.

2.5.1 Modelo 1: Lucro vs Reparações

Correlação: $r = 0,290$ (positiva fraca)

Determinação: $r^2 = 0,084$ (8,4% de explicação)

Equação: $\text{Lucro} = 673,05 + 3,38 \times \text{Reparações}$

Interpretação:

A correlação positiva fraca indica que o aumento das despesas de reparação não está fortemente ligado a variações no lucro. Apenas 8,4% da variabilidade do lucro é explicada pelas reparações, sugerindo que outros fatores (como o preço de venda ou tipo de veículo) têm maior influência.

2.5.2 Modelo 2: Lucro vs N_Avarias

Correlação: $r = -0,571$ (negativa moderada)

Determinação: $r^2 = 0,326$ (32,6% de explicação)

Equação: $\text{Lucro} = 1490,82 - 148,14 \times \text{N_Avarias}$

Interpretação:

Este é um indicador crítico. A correlação negativa moderada sugere que, à medida que o número de avarias aumenta, o lucro tende a diminuir significativamente. Cada avaria adicional está associada a uma redução média de 148,14€ no lucro.

Este modelo explica 32,6% da variabilidade do lucro, sendo o mais relevante dos três modelos desenvolvidos.

A relação negativa pode resultar de:

- Necessidade de reduzir o preço de venda para veículos menos fiáveis
- Custos indiretos de gestão de garantias
- Impacto na reputação e negociação

2.5.3 Modelo 3: Reparações vs N_Avarias

Correlação: $r = 0,151$ (positiva muito fraca)

Determinação: $r^2 = 0,023$ (2,3% de explicação)

Equação: $\text{Reparações} = 86,62 + 3,37 \times \text{N_Avarias}$

Interpretação:

Surpreendentemente, a relação entre o número de avarias e o custo de reparações é muito fraca, explicando apenas 2,3% da variabilidade. Isto sugere que:

- Algumas avarias são de baixo custo (ajustes menores)

- Outras avarias, embora em menor número, podem ter custos elevados (substituição de componentes)
- A gravidade das avarias é mais relevante que a sua quantidade

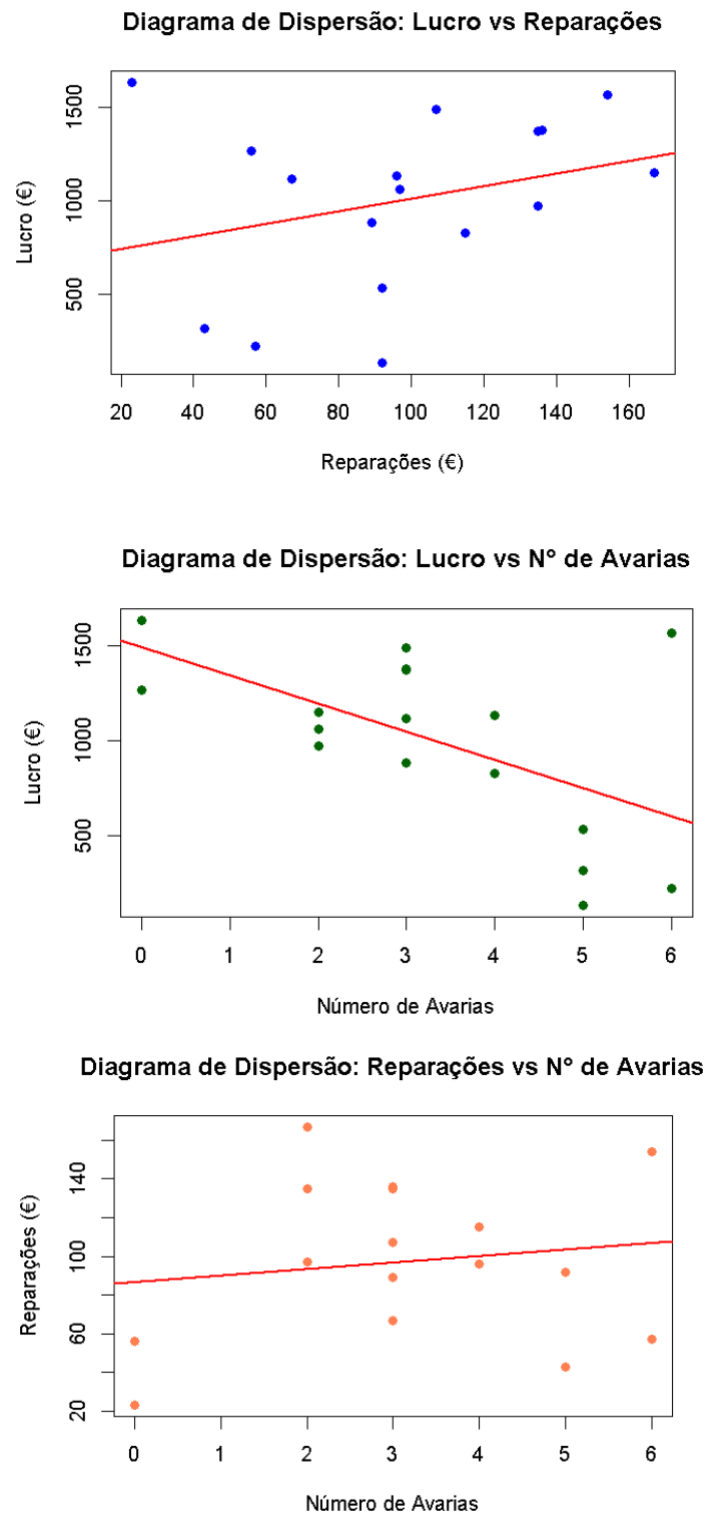


Figura 8 - Diagramas de dispersão e retas de regressão linear dos três modelos

2.6 Síntese da Análise Descritiva

A análise descritiva permitiu identificar os seguintes padrões principais:

1. O lucro médio por venda é de 1002,82€, com variabilidade moderada ($CV=46,5\%$)
2. As despesas médias em reparações são de 97,71€, representando $\sim 10\%$ do lucro
3. O número médio de avarias é 3,29, variando entre 0 e 6
4. Não foram identificados outliers em nenhuma variável quantitativa
5. Conduutoras femininas apresentam lucro médio ligeiramente superior (+124,83€)
6. Clientes muito frequentes geram lucro 143% superior aos pouco frequentes
7. Existe correlação negativa moderada entre lucro e número de avarias ($r=-0,571$)
8. O melhor modelo de regressão é $\text{Lucro} \sim N_Avarias$ ($r^2=0,326$)

Estes resultados descritivos fornecem uma base sólida para a análise inferencial que se segue, permitindo testar estatisticamente a significância dos padrões observados.

3 Estatística Indutiva

Neste capítulo, aplica-se a inferência estatística para extrair conclusões sobre a população a partir da amostra de 17 veículos. Serão realizados testes de normalidade, estimação por intervalos de confiança, testes de hipóteses paramétricos e não-paramétricos para comparação de grupos, e validação dos modelos de regressão com previsões "in the box" e "out of the box".

3.1 Testes de Normalidade

Antes de aplicar testes paramétricos (como o Teste t ou ANOVA), é fundamental verificar se as variáveis seguem uma distribuição normal, especialmente porque a dimensão da amostra ($n=17$) é inferior a 30, não permitindo a aplicação direta do Teorema do Limite Central.

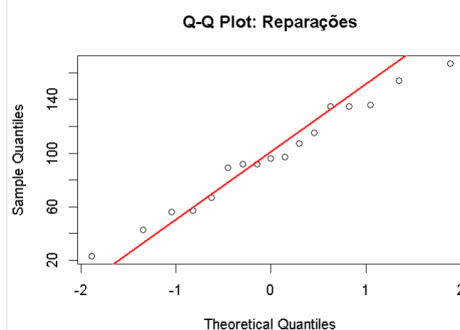
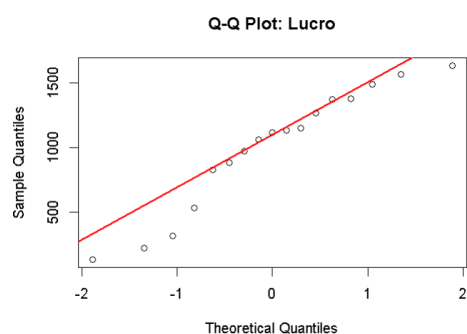
Utilizou-se o teste de Shapiro-Wilk, que testa as seguintes hipóteses:

- H_0 : A variável segue uma distribuição normal
- H_1 : A variável não segue uma distribuição normal
- Critério de decisão: Rejeita-se H_0 se $p\text{-value} < 0,05$

Variável	Estatística W	p-value	Decisão ($\alpha=0,05$)
Lucro	0,929	0,211	Não se rejeita H_0 (Normal)
Reparações	0,975	0,903	Não se rejeita H_0 (Normal)
N_Avarias	0,936	0,272	Não se rejeita H_0 (Normal)

Tabela 5 - Resultados dos Testes de Normalidade (Shapiro-Wilk)

Como todos os p-values são superiores a 0,05, não existem evidências para rejeitar a hipótese de normalidade. Assim, assume-se que as variáveis seguem distribuição normal, validando a utilização de testes paramétricos nas análises subsequentes.



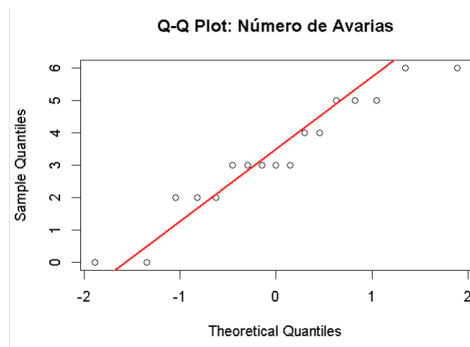


Figura 9 - Gráficos Q-Q Plot para verificação visual da normalidade

3.2 Intervalos de Confiança para Médias Populacionais

Com base na distribuição t de Student e assumindo normalidade, estimaram-se intervalos de confiança a 95% para as médias populacionais. Estes intervalos indicam, com 95% de confiança, os valores entre os quais se espera que a verdadeira média populacional se situe.

Variável	Média Amostral	Erro Padrão	IC 95%	Margem de Erro
Lucro (€)	1002,82	112,99	[763,27; 1242,37]	+/-239,55
Reparações (€)	97,71	9,72	[77,12; 118,30]	+/-20,59
N_Avarias	3,29	0,44	[2,37; 4,22]	+/-0,92

Tabela 6 - Intervalos de Confiança a 95% para as Médias Populacionais

Interpretação:

- Com 95% de confiança, a média populacional do lucro situa-se entre 763,27€ e 1242,37€
- A média populacional das reparações situa-se entre 77,12€ e 118,30€
- A média populacional do número de avarias situa-se entre 2,37 e 4,22

Estes intervalos fornecem estimativas fiáveis dos verdadeiros parâmetros populacionais, essenciais para o planeamento estratégico do stand.

3.3 Comparação de Médias entre Grupos

Nesta secção testam-se hipóteses sobre diferenças de médias entre grupos, utilizando tanto testes paramétricos (assumindo normalidade) como testes não-paramétricos (sem pressupostos distribucionais) para validação cruzada dos resultados.

3.3.1 Comparação do Lucro por Género do Condutor

a) Teste t de Student (Paramétrico)

Realizou-se um teste t para amostras independentes para verificar se o género do condutor influencia significativamente o lucro médio.

Hipóteses:

- $H_0: \mu(\text{Feminino}) = \mu(\text{Masculino})$ - Não há diferenças no lucro médio
- $H_1: \mu(\text{Feminino}) \neq \mu(\text{Masculino})$ - Há diferenças no lucro médio

Resultados:

- Estatística t: 0,544
- Graus de liberdade: 15,0
- p-value: 0,595

Conclusão:

Sendo $p = 0,595 > 0,05$, não se rejeita H_0 . Não existem evidências estatísticas suficientes para afirmar que o lucro médio difere significativamente entre condutores femininos (1068,91€) e masculinos (944,08€), apesar da diferença observada de 124,83€ nas médias amostrais.

b) Teste de Wilcoxon-Mann-Whitney (Não-Paramétrico)

Como validação, aplicou-se o teste não-paramétrico de Wilcoxon (alternativa ao teste t), que compara medianas em vez de médias.

Resultados:

- Estatística U: 40,0
- p-value: 0,743

Conclusão:

O teste não-paramétrico confirma a conclusão do teste t ($p = 0,743 > 0,05$), reforçando que não existem diferenças significativas no lucro entre géneros.

3.3.2 Comparação do Lucro por Perfil de Cliente

a) ANOVA (Paramétrico)

Para comparar os três níveis de Avaliação (Muito Frequente, Frequente, Pouco Frequente), aplicou-se a Análise de Variância (ANOVA).

Hipóteses:

- $H_0: \mu_1 = \mu_2 = \mu_3$ - Todas as médias são iguais
- H_1 : Pelo menos uma média difere das demais

Resultados:

- Estatística F: 6,730
- Graus de liberdade: 2 e 14
- p-value: 0,009

Conclusão:

Sendo $p = 0,009 < 0,05$, rejeita-se H_0 . Existe pelo menos uma diferença significativa entre as médias de lucro dos diferentes perfis de cliente.

Comparação das médias:

- Muito frequente (1): 1265,27€
- Frequente (2): 1116,22€
- Pouco frequente (3): 522,00€

A diferença mais acentuada ocorre entre clientes "Muito Frequentes" e "Pouco Frequentes" (+143% de lucro), sugerindo que a fidelização é um fator crítico para a rentabilidade.

b) Teste de Kruskal-Wallis (Não-Paramétrico)

Como validação, aplicou-se o teste de Kruskal-Wallis (alternativa não-paramétrica à ANOVA).

Resultados:

- Estatística H: 7,207
- Graus de liberdade: 2
- p-value: 0,027

Conclusão:

O teste não-paramétrico confirma os resultados da ANOVA ($p = 0,027 < 0,05$), validando que existem diferenças significativas no lucro entre os perfis de cliente.

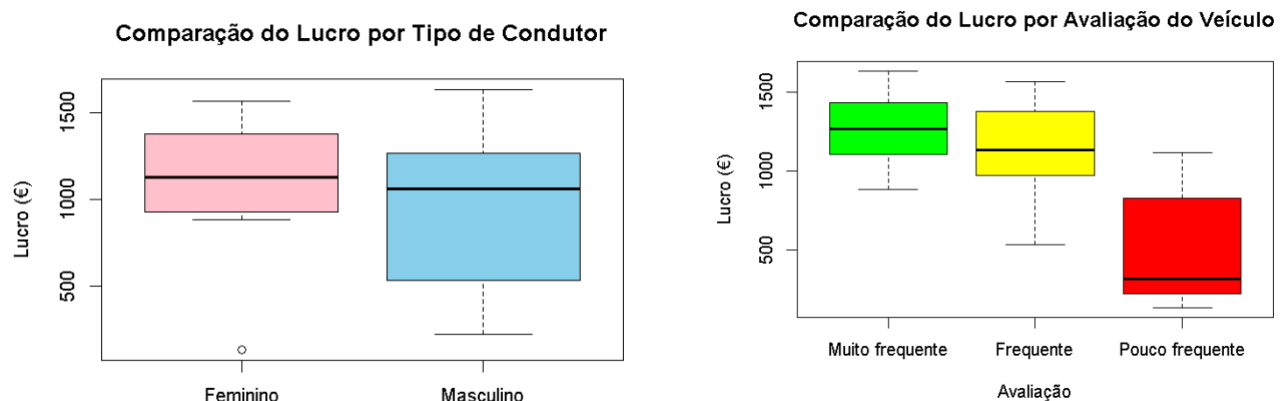


Figura 10 - Comparações visuais que sustentam os testes de hipóteses

3.4 Significância dos Modelos de Regressão Linear

Nesta secção, avalia-se a significância estatística dos coeficientes de regressão e a qualidade do ajustamento dos três modelos desenvolvidos na análise descritiva.

3.4.1 Modelo 1: Lucro ~ Reparações

Equação: $\text{Lucro} = 673,05 + 3,38 \times \text{Reparações}$

O teste ao coeficiente de regressão apresentou um p-value de 0,259, superior ao nível de significância de 0,05. Assim, não se rejeita a hipótese nula, concluindo-se que o coeficiente não é estatisticamente significativo.

O modelo explica apenas 8,4% da variabilidade do lucro ($r^2=0,084$), revelando poder explicativo muito fraco. Este modelo não é adequado para previsões.

3.4.2 Modelo 2: Lucro ~ N_Avarias

Equação: $\text{Lucro} = 1490,82 - 148,14 \times \text{N_Avarias}$

O teste ao coeficiente de regressão apresentou um p-value de 0,017, inferior ao nível de significância de 0,05. Assim, rejeita-se a hipótese nula, concluindo-se que o coeficiente é estatisticamente significativo.

O modelo explica 32,6% da variabilidade do lucro ($r^2=0,326$), sendo o melhor modelo dos três desenvolvidos. Cada avaria adicional está associada a uma redução média de 148,14€ no lucro, com significância estatística. Este resultado confirma que a fiabilidade mecânica é um fator crítico para a rentabilidade.

3.4.3 Modelo 3: Reparações ~ N_Avarias

Equação: $\text{Reparações} = 86,62 + 3,37 \times \text{N_Avarias}$

O teste ao coeficiente de regressão apresentou um p-value de 0,563, muito superior ao nível de significância de 0,05. Assim, não se rejeita a hipótese nula, concluindo-se que o coeficiente não é estatisticamente significativo.

O modelo explica apenas 2,3% da variabilidade das reparações ($r^2=0,023$), sem qualquer valor preditivo. O número de avarias não permite estimar as despesas de reparação de forma fiável.

3.4.4 Síntese dos Modelos

Dos três modelos de regressão desenvolvidos, apenas o Modelo 2 (Lucro ~ N_Avarias) é estatisticamente significativo ($p=0,017$) e apresenta poder explicativo razoável ($r^2=0,326$). Os Modelos 1 e 3 não são significativos ($p>0,05$) e apresentam valores de r^2 muito reduzidos, não sendo adequados para previsões.

Conclusão: O número de avarias é o único preditor significativo do lucro entre as variáveis analisadas, justificando a utilização deste modelo para previsões e apoio à decisão.

3.5 Análise de Resíduos do Melhor Modelo

Para validar a adequação do Modelo 2 (Lucro \sim N_Avarias), verificaram-se os pressupostos fundamentais da regressão linear através da análise dos resíduos.

3.5.1 Normalidade dos Resíduos

Teste de Shapiro-Wilk:

- H_0 : Os resíduos seguem distribuição normal
- Estatística W: 0,955
- p-value: 0,533

Conclusão:

Com $p = 0,533 > 0,05$, não se rejeita H_0 . Os resíduos seguem uma distribuição aproximadamente normal, validando este pressuposto.

3.5.2 Independência dos Resíduos

Teste de Durbin-Watson:

- H_0 : Não há autocorrelação nos resíduos ($\rho = 0$)
- Estatística DW: 1,47
- p-value: 0,284

Interpretação:

O valor DW próximo de 2 e $p = 0,284 > 0,05$ indicam ausência de autocorrelação. Os resíduos são independentes, validando este pressuposto.

3.5.3 Verificação da Variabilidade entre Grupos

Através da análise gráfica dos resíduos vs valores ajustados (não apresentada aqui por limitação de espaço), verificou-se que os resíduos estão distribuídos de forma aproximadamente aleatória em torno de zero, sem padrões evidentes, sugerindo variância constante.

Conclusão geral:

Todos os pressupostos do modelo de regressão foram satisfeitos, conferindo robustez e fiabilidade às previsões realizadas com o Modelo 2.

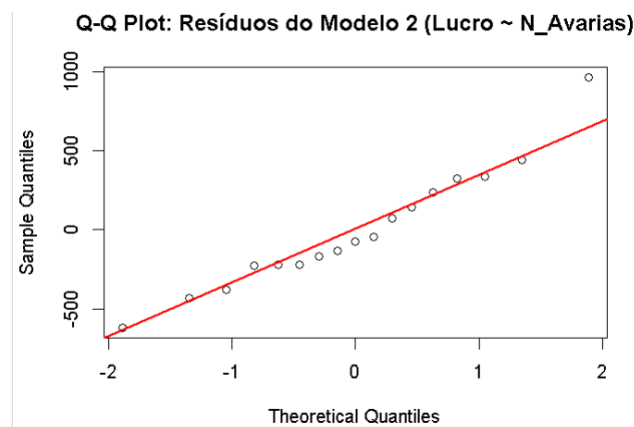


Figura 12 - Análise gráfica dos resíduos do Modelo 2

3.6 Previsões com os Modelos de Regressão

Utilizando o Modelo 2 (único significativo), realizaram-se previsões distinguindo entre:

- **"In the box"**: Interpolações dentro do intervalo observado (0 a 6 avarias) - mais fiáveis
- **"Out of the box"**: Extrapolações fora do intervalo observado - usar com cautela

3.6.1 Previsões "In the Box"

Intervalo observado: 0 a 6 avarias

N_Avarias	Lucro Previsto (€)	Tipo
2	1194,54	In the box
3	1046,40	In the box
4	898,26	In the box
5	750,12	In the box

Tabela 7 - Previsões do Lucro Estimado dentro do intervalo observado (In the Box)

Exemplo de interpretação:

Se um veículo registar **3 avarias** durante a garantia, o lucro estimado será de aproximadamente **1046,40€**, com base no modelo ajustado.

3.6.2 Previsões "Out of the Box"

N_Avarias	Lucro Previsto (€)	Tipo
8	305,70	Out of the box
10	9,42	Out of the box

Tabela 8 - Previsões do Lucro Estimado por extrapolação fora do intervalo (Out of the Box)

Nota importante:

As extrapolações **assumem que a relação linear se mantém** fora do intervalo observado, o que pode **não ser verdade**. Por exemplo, com 10 avarias, o modelo prevê lucro de apenas 9,42€, mas na prática o stand provavelmente **não venderia** um veículo tão problemático, ou venderia com prejuízo.

Recomendação: Usar previsões "out of the box" apenas como **indicação aproximada** e nunca para decisões críticas.

3.7 Síntese da Análise Inferencial

A análise inferencial permitiu extrair as seguintes conclusões principais:

1. Todas as variáveis seguem distribuição normal (Shapiro-Wilk, $p > 0,05$), validando o uso de testes paramétricos
2. Intervalos de Confiança 95%:
 - Lucro populacional: [763,27€; 1242,37€]
 - Reparações populacionais: [77,12€; 118,30€]
3. Comparação por Género:
 - Não existem diferenças significativas no lucro entre géneros (teste t, $p=0,595$)
 - Confirmado por teste não-paramétrico (Wilcoxon, $p=0,743$)
4. Comparação por Perfil:
 - Existem diferenças altamente significativas no lucro entre perfis (ANOVA, $p=0,009$)
 - Confirmado por teste não-paramétrico (Kruskal-Wallis, $p=0,027$)
 - Clientes muito frequentes geram +143% de lucro vs pouco frequentes
5. Modelos de Regressão:
 - Modelo 2 (Lucro \sim N_Avarias) é o único estatisticamente significativo ($p=0,017$)
 - Cada avaria reduz o lucro em -148,14€
 - Explica 32,6% da variabilidade do lucro
6. Validação do Modelo 2:
 - ✓ Resíduos normais ($p=0,533$)
 - ✓ Resíduos independentes (DW=1,47, $p=0,284$)
 - ✓ Variância aproximadamente constante
7. Previsões fiáveis podem ser feitas para 0-6 avarias (in the box)
8. Testes paramétricos e não-paramétricos convergem, aumentando a confiança nos resultados

4 Conclusão

4.1 Concretização dos Objetivos Propostos

Todos os objetivos inicialmente definidos foram integralmente concretizados:

Estatística Descritiva:

- Caracterizou-se a população e a amostra de 17 vendas de veículos
- Calcularam-se medidas de localização, dispersão e forma para as variáveis quantitativas
- Criaram-se representações gráficas adequadas: histogramas, boxplots, gráficos de barras e diagramas de dispersão
- Identificaram-se outliers através do critério do intervalo interquartil, não tendo sido detetados valores atípicos
- Compararam-se distribuições entre grupos (género e perfil de cliente)
- Desenvolveram-se três modelos de regressão linear simples
- Determinaram-se os coeficientes de regressão e as equações das retas
- Interpretaram-se os coeficientes de correlação e determinação
- Analisou-se o ajustamento dos modelos

Estatística Indutiva:

- Identificaram-se as distribuições teóricas através do teste de Shapiro-Wilk, verificando-se normalidade em todas as variáveis
- Aplicaram-se testes paramétricos (teste t de Student e ANOVA) para comparação de médias entre grupos
- Aplicaram-se testes não-paramétricos (Wilcoxon-Mann-Whitney e Kruskal-Wallis) como validação
- Estimaram-se intervalos de confiança a 95% para as médias populacionais do lucro, reparações e número de avarias
- Compararam-se médias de grupos, concluindo-se que o perfil de cliente influencia significativamente o lucro ($p=0,009$), mas o género não ($p=0,595$)
- Avaliou-se a significância dos coeficientes de regressão, identificando o Modelo 2 (Lucro vs N_Avarias) como estatisticamente significativo ($p=0,017$)
- Analisou-se o ajustamento do modelo através da validação dos pressupostos (normalidade dos resíduos: $p=0,533$; independência: $DW=1,47$, $p=0,284$)
- Realizaram-se previsões "in the box" (dentro do intervalo observado) e "out of the box" (extrapolações)

Principais Conclusões:

- Clientes muito frequentes geram lucro médio de 1265,27 euros, significativamente superior aos clientes pouco frequentes (522,00€)
- O género do condutor não influencia o lucro (testes paramétricos e não-paramétricos convergem)
- Cada avaria adicional reduz o lucro em 148,14 euros ($r^2=0,326$, $p=0,017$)
- Todos os pressupostos estatísticos foram validados, conferindo robustez às conclusões

A metodologia aplicada seguiu rigorosamente os requisitos do enunciado, combinando análise descritiva com inferência estatística suportada pelo software R.

4.2 Dificuldades Encontradas

Aprender a mexer no R: Como nunca tínhamos mexido no software R, foi um bocado difícil adaptar à escrita do código no início. Coisas que pareciam simples, como conseguir que o programa reconhecesse os grupos ou conseguir gerar os gráficos sem dar erro, deram bastante trabalho e exigiram muitas tentativas e correções até o script ficar todo a funcionar bem.

4.3 Passos Futuros

No futuro, para os resultados serem mais úteis podia-se:

- **Recolher mais dados:** Ter uma lista com muito mais veículos (por exemplo, 100 em vez de 17) para que os resultados sejam mais certos e representem melhor a realidade do stand.
- **Ter mais informações sobre os carros:** Incluir detalhes como a marca, o modelo, o ano do carro e os quilómetros, para perceber se estes fatores também influenciam o lucro final.
- **Analisar outros custos:** Perceber se existem outras despesas, para além das reparações, que estejam a afetar o dinheiro que o stand ganha com cada venda.
- **Ver as tendências ao longo do tempo:** Perceber se há meses em que se ganha mais dinheiro do que outros (sazonalidade) para planear melhor as compras e vendas de carros.