



LABORATÓRIO 4

LABORATORY 4

Gabriel Penha*,

Moisés Augusto†,

Carlos Siebra Boucas‡

RESUMO

Startup é uma expressão em inglês que significa começar do zero. São empresas que possuem ideias inovadoras com potencial de obtenção de lucro em um curto período de tempo. Considerando os diversos exemplos deste tipo de empresa que surgiram recentemente, este trabalho teve como objetivo analisar os dados de 50 *startups*, com intuito de ajustar modelos de regressão linear simples que pudessem prever o lucro. A base de dados utilizada continha informações sobre os gastos com *marketing*, despesas administrativas e em investimento e pesquisa. Para cada uma dessas três variáveis, um modelo linear simples foi ajustado; dentre eles, concluiu-se que o melhor seria aquele que utilizou os investimentos em pesquisa e desenvolvimento de variável explicativa. A performance dos ajustes possivelmente poderia ser melhorada com um modelo que considerasse múltiplas variáveis, com um maior tamanho de amostra e com mais informações sobre as empresas (isto é, mais variáveis preditoras).

Palavras-chave: Startup. Lucro. Gastos. Linear. Modelo.

1 INTRODUÇÃO

Segundo Borba (2017), *Startup* é uma expressão em inglês que significa começar do zero. São empresas que possuem ideias inovadoras com potencial de obtenção de lucro em um curto período de tempo.

Visando conhecer melhor a relação entre lucro e gastos de empresas já estabelecidas no mercado, Dados de 50 diferentes *startups* foram coletados. O conjunto continha informações como **lucros, gastos com pesquisa e desenvolvimento, gastos com marketing e gastos com despesas administrativas**. A base de dados não continha dados faltantes, tampouco repetidos. Todas as observações eram numéricas contínuas, o que facilitou o ajuste dos modelos propostos.

O objetivo deste trabalho foi analisar os dados coletados e ajustar diferentes modelos de regressão linear simples, com intuito de entender a relação entre as variáveis explicativas (os diferentes tipos de gasto) e a variável resposta (lucro), e fornecer conhecimento relevante no que tange ao que investir, quando se monta uma *startup*.

Nas seguintes seções, ter-se-ão a análise exploratória dos dados, os modelos ajustados e suas interpretações (incluindo a variância dos estimadores) e algumas considerações finais.

2 ANÁLISE EXPLORATÓRIA

A base de dados original possuía 50 observações, obtiveram-se as estatísticas descritivas de cada uma das variáveis, a saber:

2.1 Lucro

O lucro médio das empresas foi próximo de 112013,00U.M, em que U.M significa *unidade monetária*. A empresa que mais lucrou faturou por volta de 192262,00U.M, sendo que a menos lucrativa obteve em torno de 14681,00U.M. 25% das empresas faturaram mais que 139766,00U.M e 25% delas faturaram menos que algo em torno de 90139,00U.M.

* Instituto de Matemática e Estatística, Departamento de Estatística, Bacharelado em Estatística; ✉ penha.gabriel@ufba.br.

† Instituto de Matemática e Estatística, Departamento de Estatística, Bacharelado em Estatística; ✉ moises.augusto@ufba.br.

‡ Instituto de Matemática e Estatística, Departamento de Estatística, Bacharelado em Estatística; ✉ carlos.siebra@ufba.br.

2.2 Gastos com pesquisa e desenvolvimento

Algumas *startups* não investiram em pesquisa e desenvolvimento. A que mais investiu desembolsou 165349,20U.M. Em média, as empresas gastaram em torno de 73722,00U.M com pesquisa e desenvolvimento, sendo que cerca de 25% delas gastaram mais que 101603,00U.M e cerca de 25% gastaram menos que 39936,00U.M. Verificou-se a relação entre o lucro e os gastos com pesquisa e desenvolvimento graficamente através de um diagrama de dispersão.

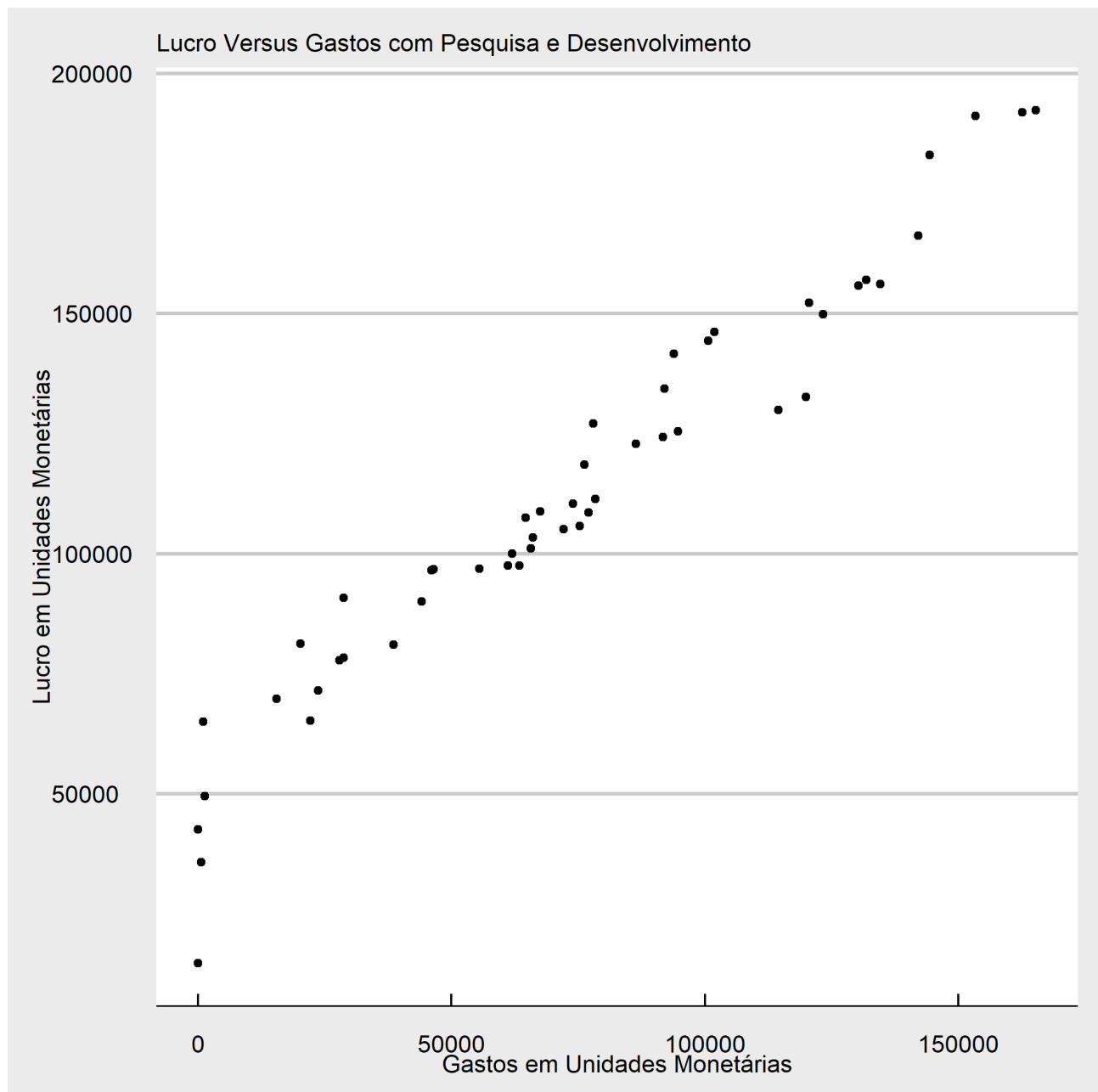


Figura 1 – Diagrama de Dispersão: Lucro (U.M) Vs. Gastos com pesquisa e desenvolvimento (U.M)

Como é possível visualizar pela Figura 1, existem indícios fortes de uma relação linear entre os gastos com pesquisa e desenvolvimento e o lucro, ambos em unidades monetárias. Aparentemente, quanto mais se investe em pesquisa e desenvolvimento, maiores são os lucros.

2.3 Gastos com despesas administrativas

Em média, as *startups* registraram cerca de 121345,00U.M gastos com despesas administrativas. A empresa que mais gastou registrou em torno de 182646,00U.M de despesas, enquanto a que menos gastou teve que pagar cerca de 51283,00U.M.

Os gastos com despesas administrativas também foram analisados em conjunto aos lucros graficamente:

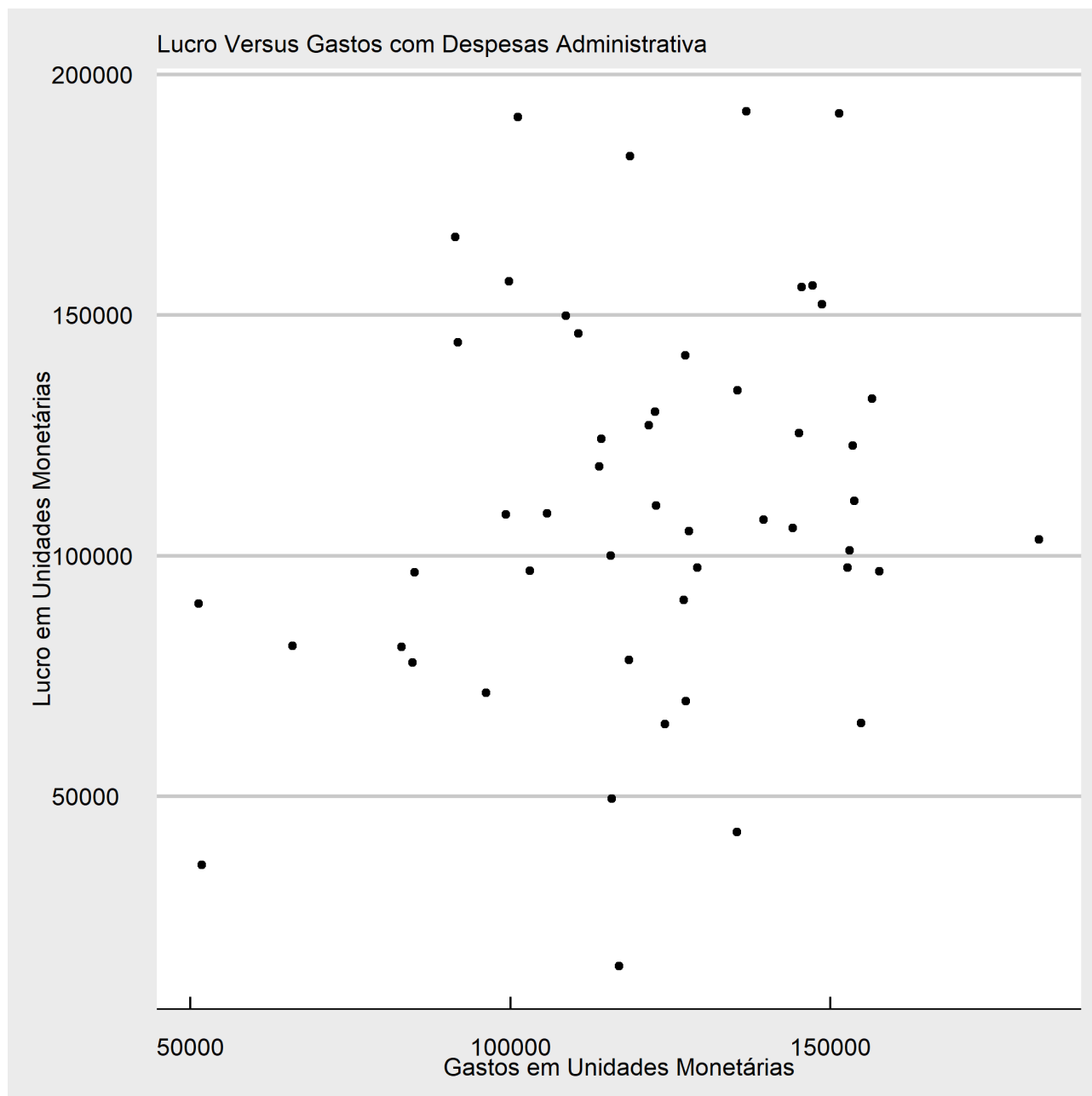


Figura 2 – Diagrama de Dispersão: Lucro ($U.M$) Vs. Gastos com despesas administrativas ($U.M$)

Como é possível observar na Figura 2, não **parece** haver uma relação linear muito evidente entre as variáveis. No entanto, parece que há uma tendência de crescimento do lucro, conforme gasta-se mais despesas administrativas.

2.4 Gastos com Marketing

A empresa que mais investiu com Marketing pagou cerca de $471784,00U.M$, enquanto houveram empresas que não desembolsaram quantia alguma. Em média, as empresas registraram em torno de $211025,00U.M$, sendo que ao menos 50% delas gastaram mais que $212716,00U.M$.

Também se verificou graficamente a relação entre investimentos em *Marketing* e lucro, como é possível ver a seguir:

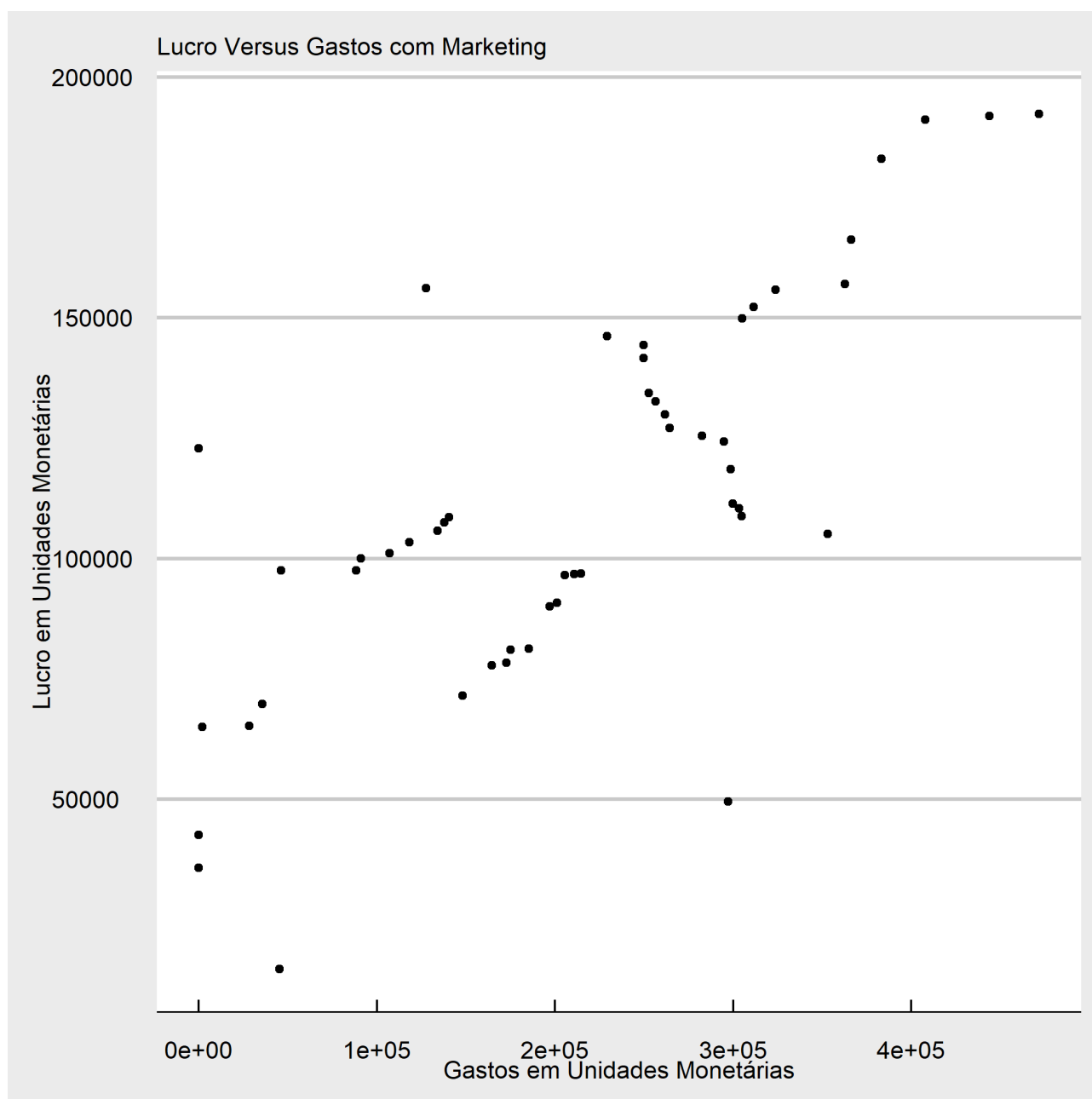


Figura 3 – Diagrama de Dispersão: Lucro (*U.M*) Vs. Gastos com *Marketing* (*U.M*)

Na Figura 3, é possível constatar uma tendência de aumento nos lucros conforme investe-se mais em *Marketing*. Os dados, inclusive, **parecem** ter uma relação linear e para verificar esta relação e as apresentadas nas Figuras 1 e 2, realizaram-se testes de hipótese para a correlação entre cada uma das variáveis explicativas e o lucro.

2.5 Correlação

Utilizou-se um nível de confiança de 95% para a realização dos testes. O primeiro teste verificou a correlação entre o lucro e os gastos com pesquisa e desenvolvimento. O resultado registrado foi uma correlação maior que 97%; o teste de hipóteses obviamente atestou a relação entre as duas variáveis.

O segundo teste foi entre os gastos com despesas administrativas e o lucro. A correlação registrada foi de cerca de 20%. Através do teste de hipóteses, concluiu-se que não haviam indícios fortes o suficiente a favor da relação entre as duas variáveis.

O terceiro e último teste indicou uma correlação próxima a 62%. Através do teste de hipóteses concluiu-se favoravelmente pela relação entre os investimentos em *marketing* e o lucro.

Feito isso, ajustaram-se três modelos de regressão linear simples, um para cada uma dessas duas variáveis. Esperava-se que o modelo ajustando os investimentos em pesquisa e desenvolvimento tivesse uma melhor performance.

3 RESULTADOS

O primeiro modelo de regressão linear simples foi ajustado utilizando os investimentos em pesquisa e desenvolvimento como variável explicativa. Através da análise de resíduos, concluiu-se que uma das observações era um valor discrepante e, então, ela foi retirada do modelo que foi novamente ajustado. Na Equação 1, é possível observar a forma deste modelo:

$$Lucro_i = 5,172 \times 10^4 + 8,278 \times 10^{-1} gasto_ped_i + \epsilon_i \quad (1)$$

Sendo $Lucro_i$ o lucro da i -ésima *startup*, $gasto_ped_i$, o gasto em pesquisa e desenvolvimento da i -ésima *startup* e ϵ_i a diferença entre o valor ajustado para o lucro e o lucro registrado da i -ésima *startup*. Este modelo indica que a cada $U.M$ investida em pesquisa e desenvolvimento, o lucro médio sobe em $8,278 \times 10^{-1}$ e que no caso de não investimento, o lucro médio fica em torno dos $51720,00U.M$.

A um nível de 95% de confiança, a variável ajustada foi significativa. Realizaram-se testes de hipóteses que apontaram que tanto o intercepto da reta (β_0) quanto o coeficiente angular β_1 eram diferentes de 0. Daí, uma análise gráfica de resíduos foi feita. Os gráficos utilizados podem ser visualizados na Figura 4.

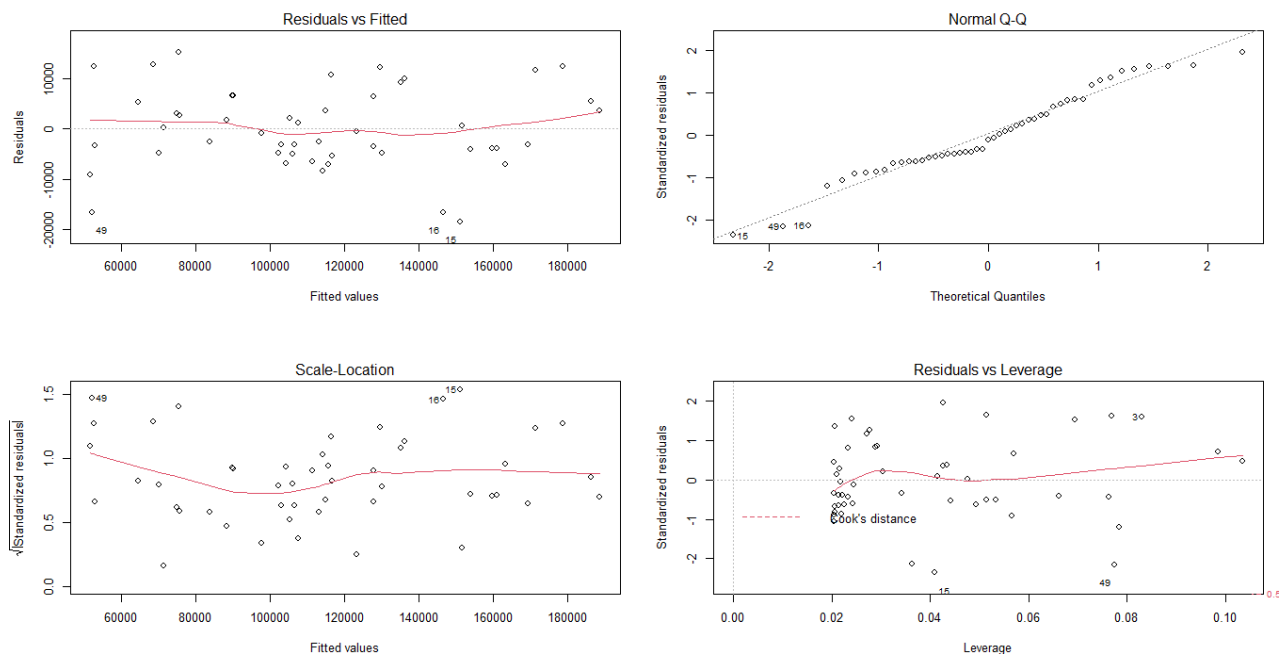


Figura 4 – Análise de Resíduo: Modelo da Equação 1 ($U.M$)

Observando o gráfico quantil-quantil (*Normal QQ*), é possível observar que os resíduos não necessariamente estão se comportando como esperado de uma distribuição normal - apesar de estarem bem próximos disso -. Já analisando o primeiro gráfico, embora padrões não pareçam perceptíveis, é visível que alguns dos resíduos distanciam-se bastante do 0, o que não é um bom indicativo.

Testes de hipóteses para a normalidade e para a homocedasticidade foram realizados. Ambos apontaram para um bom ajuste do modelo - isto é, indicaram tanto a normalidade, quanto a homocedasticidade dos resíduos -. Vale dizer, no entanto, que como a base de dados detém poucas observações, o teste pode estar mais propenso a não rejeitar a hipótese nula (normalidade e homocedasticidade, respectivamente), do que deveria.

O segundo modelo ajustado considerou os gastos com despesas administrativas na tentativa de explicar o lucro. O

modelo pode ser visualizado na Equação 2.

$$Lucro_i = 8,029 \times 10^4 + 2,776 \times 10^{-1} gasto_adm_i + \epsilon_i \quad (2)$$

Sendo $gasto_adm_i$, o gasto em despesas administrativas da i -ésima *startup* e os demais definidos como anteriormente. Este modelo indica que a cada $U.M$ gasta em despesas administrativas aumenta o lucro médio em $2,776 \times 10^{-1}$ e que no caso de não investimento, o lucro médio fica em torno dos 80290,00 $U.M$.

A um nível de 95% de confiança, o gasto com despesas administrativas não foi uma variável significativa para explicar o lucro linearmente; pelo teste de hipóteses realizado, ao nível de 95%, o coeficiente angular β_1 não era diferente de 0. O teste realizado para intercepto da reta (β_0), no entanto, apontou que este valor era diferente de 0. Daí, uma análise gráfica de resíduos foi feita. Os gráficos utilizados podem ser visualizados na Figura 5.

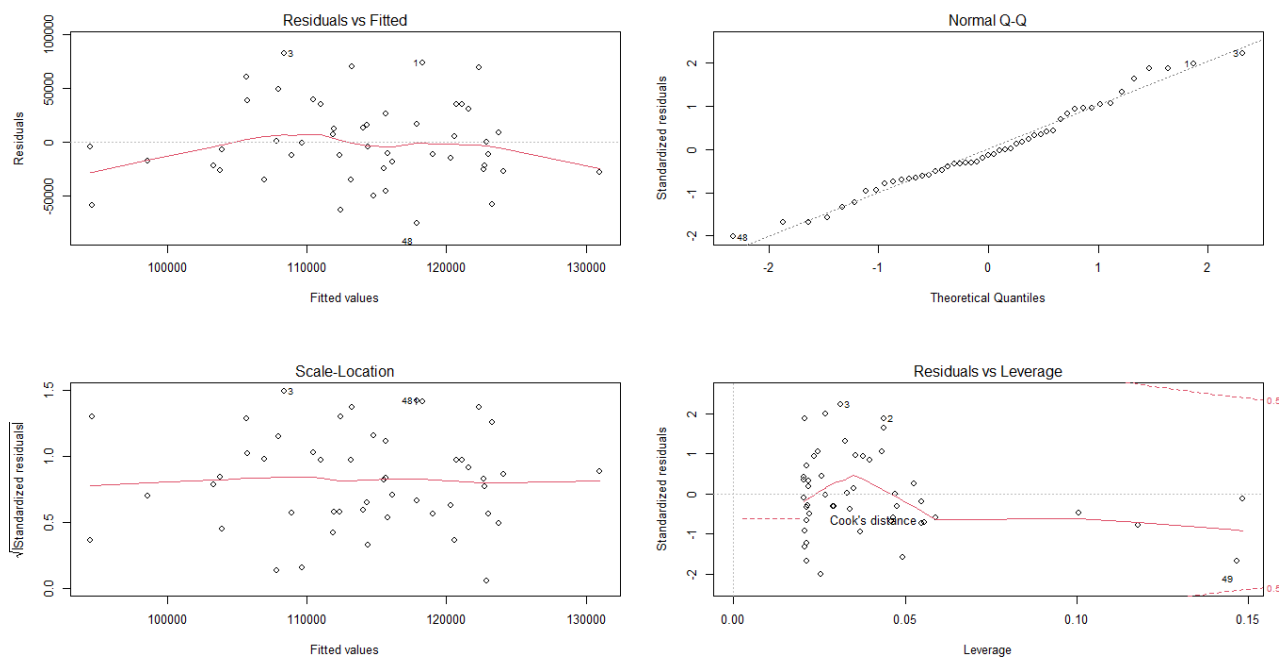


Figura 5 – Análise de Resíduo: Modelo da Equação 2 ($U.M$)

Pela Figura 5, é possível ver que os resíduos não necessariamente seguem uma distribuição normal (algumas das observações se comportam diferentemente do esperado), apesar de estarem bem próximos disso. Quanto à homocedasticidade, é possível notar uma concentração dos resíduos para a faixa entre 110000 $U.M$ e 120000 $U.M$ nos valores ajustados. Além disso, a variabilidade parece crescer um pouco e decrescer levemente depois. Como no caso do modelo 1, testes de hipóteses para testar tanto a normalidade quanto a homocedasticidade foram realizados, ambos ao nível de 95%. Em ambos, os resultados indicaram pela não rejeição da hipótese nula.

Finalmente, o terceiro modelo verificou a relação entre o lucro e os investimentos em *marketing*. Diferentemente dos dois primeiros modelos, a observação discrepante aqui retirada foi a identificada pelo número 47.

$$Lucro_i = 5,959 \times 10^4 + 2,566 \times 10^{-1} gasto_mkt_i + \epsilon_i \quad (3)$$

Em que $gasto_mkt_i$ é o investimento em *marketing* da i -ésima *startup*.

O modelo ajustado indica que a cada uma unidade monetária gasta em *marketing*, o lucro médio obtido é de 0,2566 $U.M$. E no caso de não investimento nesta categoria, espera-se um lucro em torno de 59590,00 $U.M$. Ao nível de 95% a variável explicativa foi considerada significativa. Realizaram-se testes de hipóteses que apontaram que tanto o intercepto da reta (β_0) quanto o coeficiente angular β_1 eram diferentes de 0. Os gráficos para análise dos resíduos podem ser visualizados na Figura 6.

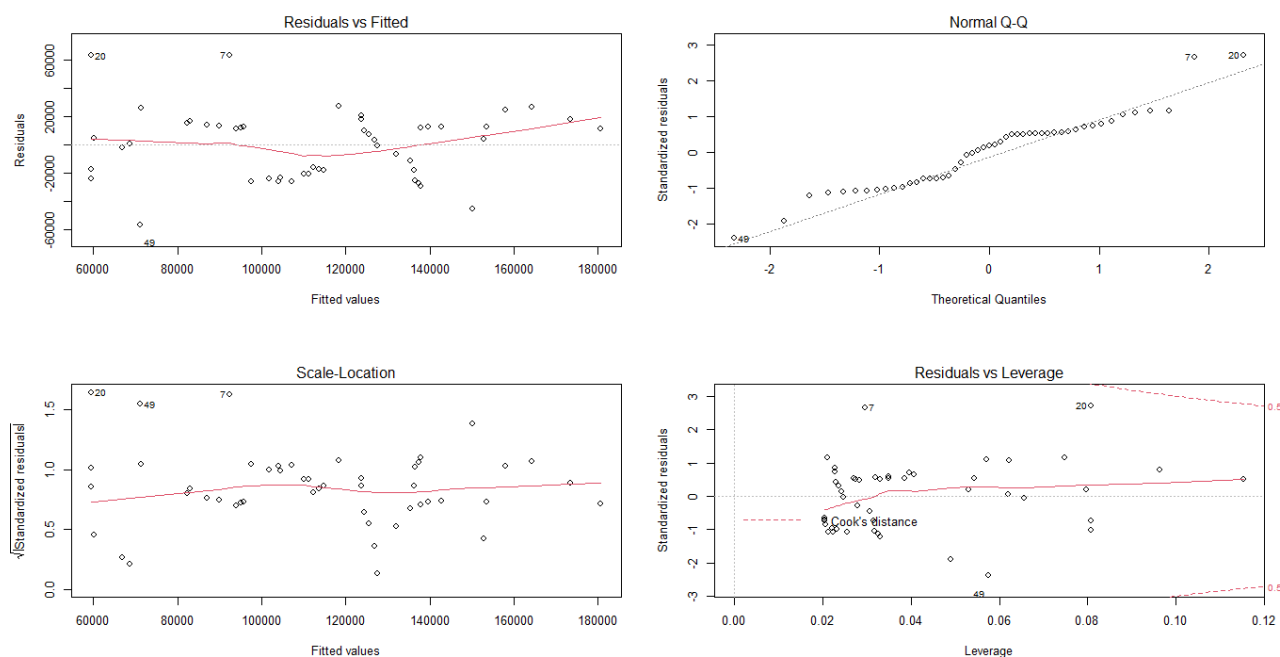


Figura 6 – Análise de Resíduo: Modelo da Equação 2 (*U.M*)

Como é possível observar no gráfico quantil-quantil, os resíduos não parecem se comportar como o esperado de uma distribuição normal. Já o gráfico de resíduos e ajustados indica mais alguns valores discrepantes, o que levanta a hipótese de má adequação do modelo, que talvez melhorasse se se acrescentassem mais algumas variáveis preditoras.

O teste de hipóteses utilizado indicou pela não normalidade nos resíduos do modelo 3, enquanto o teste para verificar a homocedasticidade indicou que a variabilidade era constante.

4 CONSIDERAÇÕES FINAIS

Considerando o mencionado, conclui-se que, possivelmente, os modelos ajustados podem ser melhorados. Uma alternativa seria o modelo de regressão linear múltipla, incluindo ou todas as três variáveis aqui estudadas, ou ao menos as variáveis que indicam o investimento em *marketing* e o investimento em pesquisa e desenvolvimento.

Dito isso, entre os modelos ajustados, o modelo 1 foi considerado o melhor, tanto pela sua performance - e bons indicativos de significância e relação linear entre a variável preditora e a predita -, quanto pela análise de resíduos; apesar de ainda não ser um modelo ideal.

Uma análise mais rebuscada seria interessante. Para tal, recomenda-se fortemente um tamanho maior de amostra e, possivelmente, observações de mais informações relevantes sobre as *startups*.

REFERÊNCIAS

BORBA, G. M. Empreendedorismo por meio de startup: um estudo de caso em uma startup na cidade de Criciúma/SC, 2017.

ANEXO A – CÓDIGOS UTILIZADOS PARA ANÁLISE (NO R)

É possível obter o projeto utilizado para análise de dados no R com o *link* a seguir:

<https://drive.google.com/drive/folders/1JbvK0pC2hVDMlq25oK0x9oc6Xm6lVcCd?usp=sharing>

DECLARAÇÃO DE RESPONSABILIDADE

O(s) autor(es) é(são) o(s) único(s) responsável(eis) pelas informações contidas neste documento.