



LABORATÓRIO 6 LABORATORY 6

Gabriel Penha*, Moisés Augusto†

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 um modelo de regressão linear múltiplo que explique a relação entre a variável resposta *lucro* e as variáveis explicativas *gastos com marketing* e *gastos com pesquisa e desenvolvimento*. Ajustaram-se dois modelos de regressão linear múltiplo envolvendo essas três variáveis. Um deles utilizou a metodologia tradicional, o outro utilizou a transformação Box-Cox. Vale dizer que os dois ajustes foram bastante parecidos, tendo em vista que o melhor λ encontrado foi aproximadamente 0,98, muito próximo de 1; o caso em que seria adequado não realizar transformações.

Palavras-chave: Startup. Lucro. Gastos. Modelo linear múltiplo. Regressão.

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 um modelo de regressão linear múltiplo, 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 quê 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.

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.

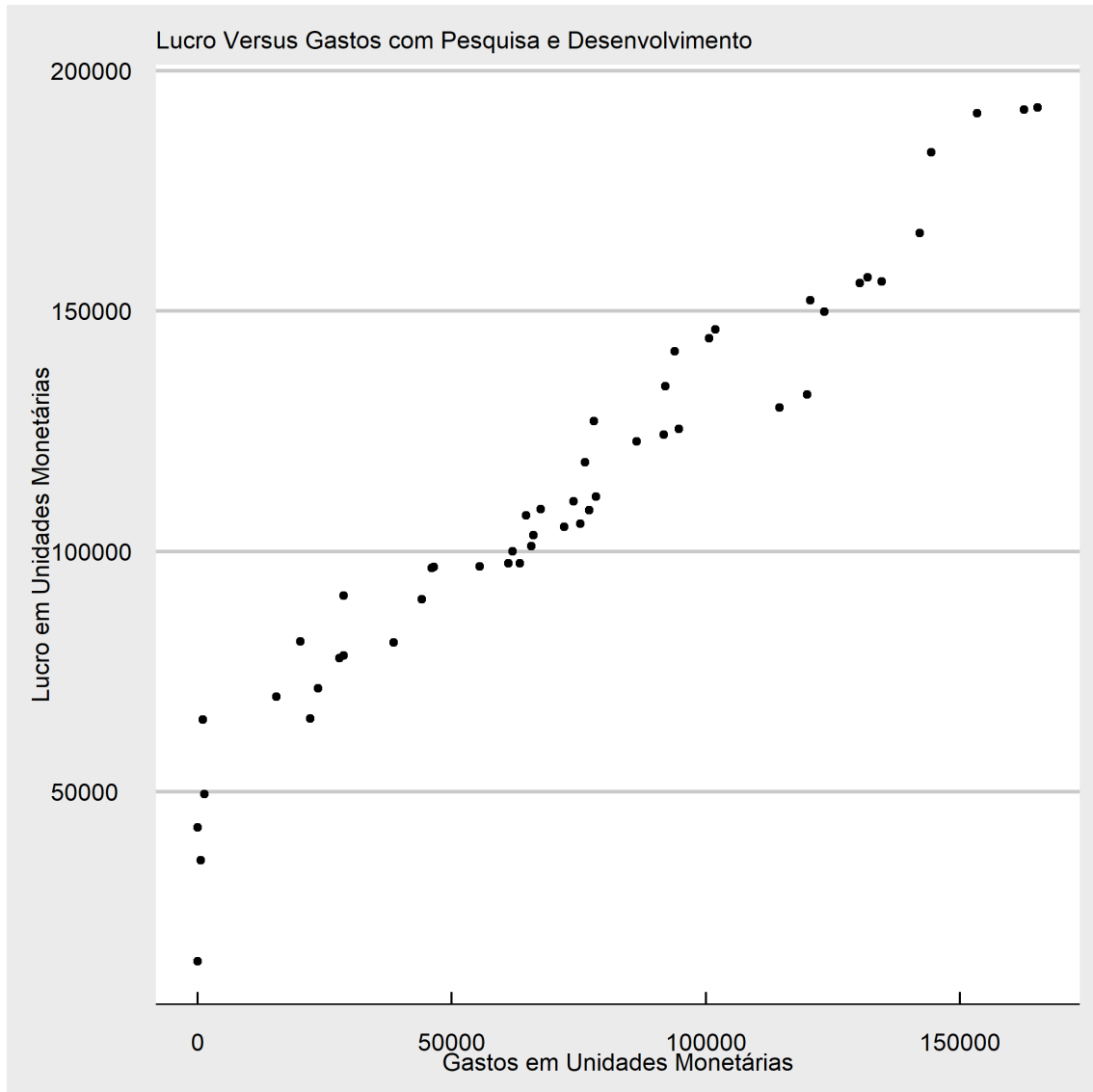


Figura 1 – Diagrama de Dispersão: Lucro (U.M) Vs. Gastos com pesquisa e desenvolvimento (U.M)

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:

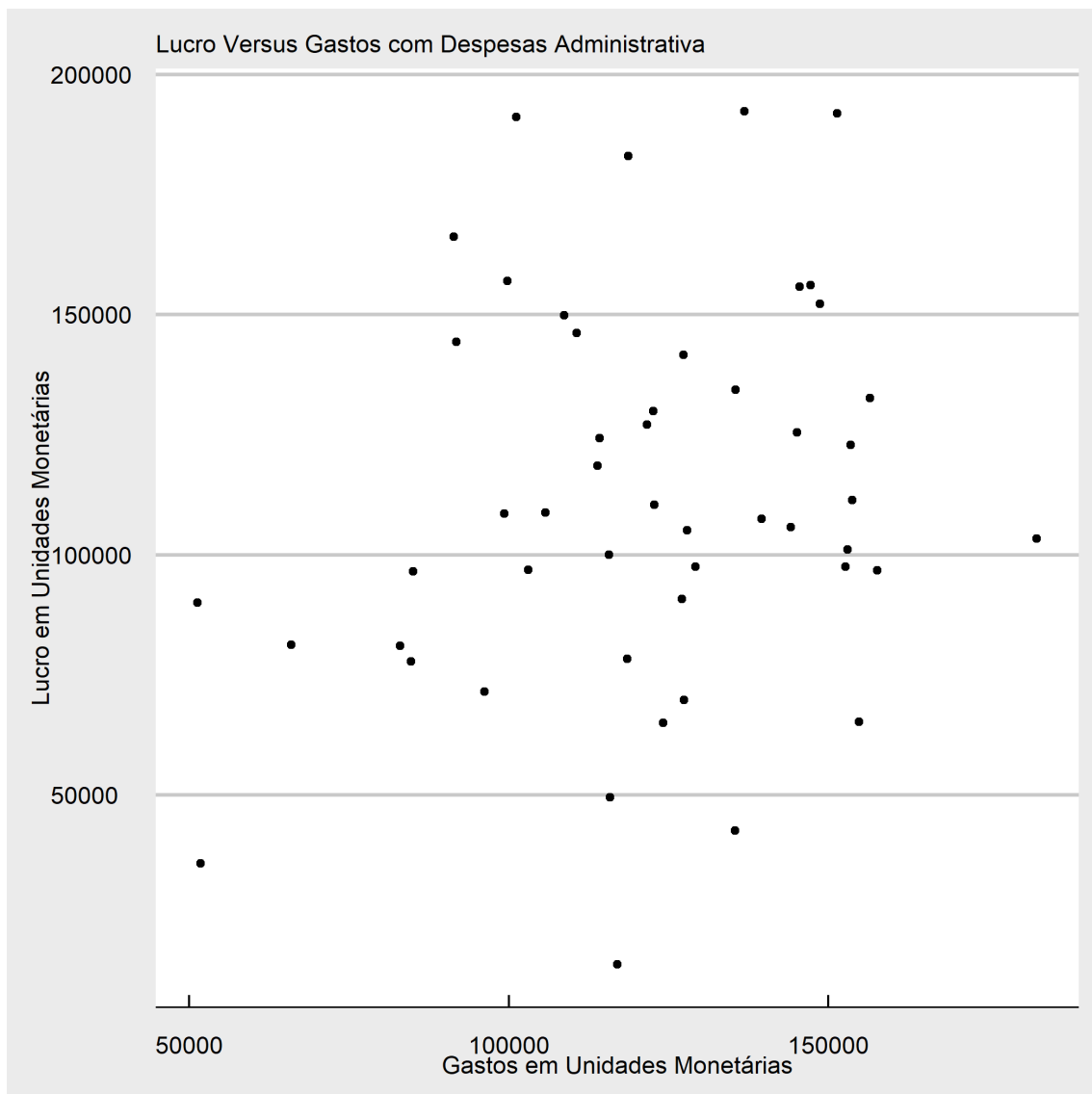


Figura 2 – Diagrama de Dispersão: Lucro (U.M) Vs. Gastos com despesas administrativas (U.M)

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:

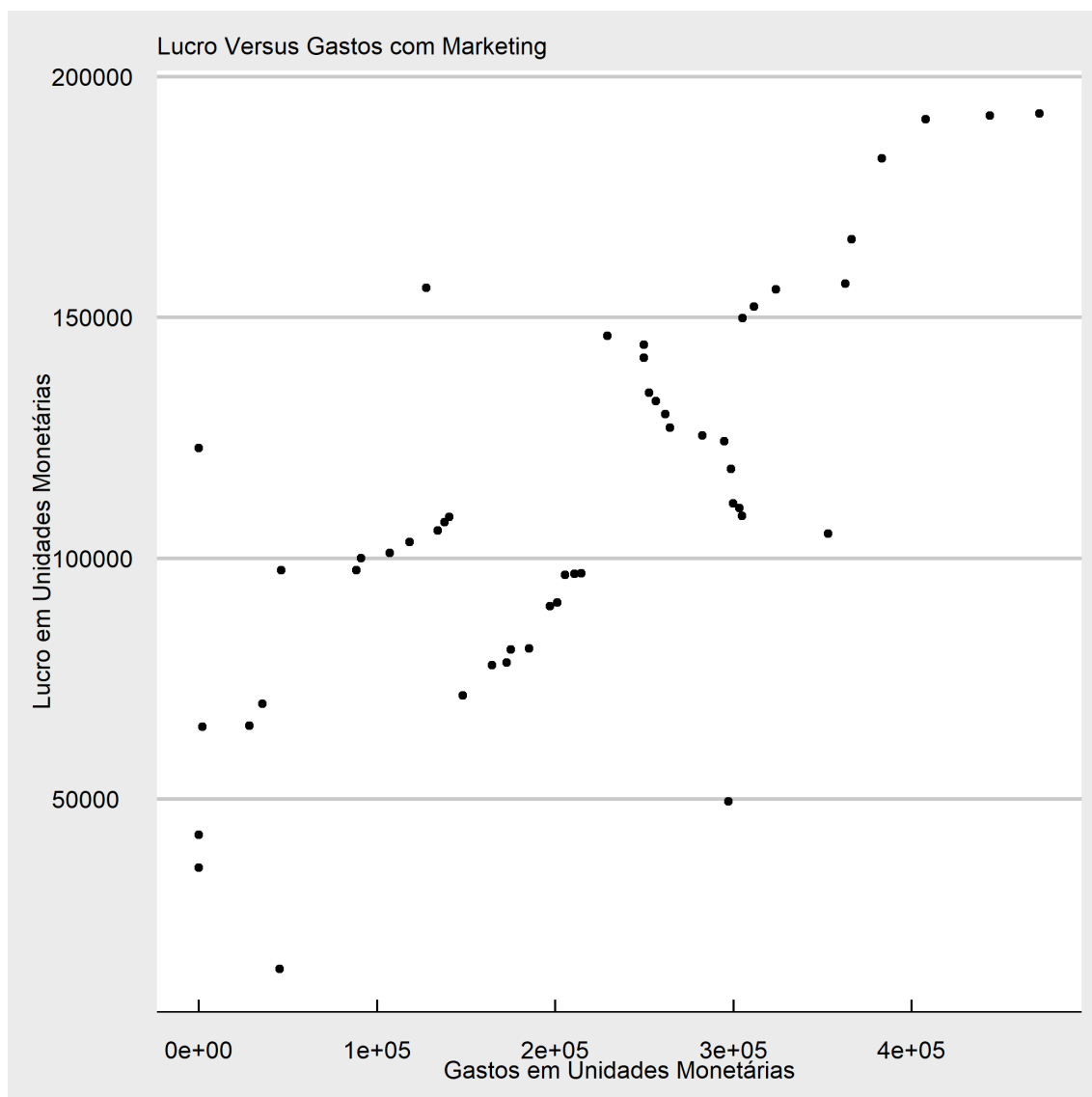


Figura 3 – Diagrama de Dispersão: Lucro (*U.M*) Vs. Gastos com *Marketing* (*U.M*)

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 dois modelos de regressão linear múltipla. Nenhum dos dois considerou a variável gastos com despesas administrativas, por conta de sua baixa correlação com o lucro. O primeiro modelo foi ajustado do modo mais comum, sem considerar nenhum tipo de transformação.

No segundo, aplicou-se a transformação Box-Cox.

3 RESULTADOS

Considere que *gasto_ped* indicará os gastos em pesquisa e desenvolvimento em unidades monetárias e que *gasto_mkt* indicará os gastos com *marketing* também em unidades monetárias. O primeiro modelo de regressão linear (que chamaremos de Modelo 1) ajustado foi da forma:

$$\hat{Y} = 49790,00 + 0,7754\text{gasto_ped} + 0,02745\text{gasto_mkt} \quad (1)$$

Em que \hat{Y} indica o valor estimado para o lucro quando se investe *gasto_ped* em pesquisa e desenvolvimento e *gasto_mkt* em *marketing*.

Este modelo indica que a cada *U.M* investida em pesquisa e desenvolvimento, o lucro médio crescerá em torno de 0,7754*U.M*; além disso, a cada *U.M* investida em *marketing*, o lucro médio crescerá em torno de 0,02745*U.M*. Na ausência de gastos nessas duas áreas, o lucro médio das *startups* é aproximadamente 50000,00*U.M*.

Com 95% de confiança, tanto os gastos com pesquisa e desenvolvimento quanto os gastos com *marketing* foram variáveis significantes ao modelo. Ou seja, ajudaram a explicar a variabilidade do lucro. O intercepto, isto é, o valor médio de lucro na ausência das duas variáveis explicativas, também foi significativamente diferente de zero.

Uma análise de resíduos foi feita para este modelo, visual e quantitativamente. Os gráficos quantil quantil (Normal) e resíduos *versus* ajustados podem ser visualizados na Figura 4.

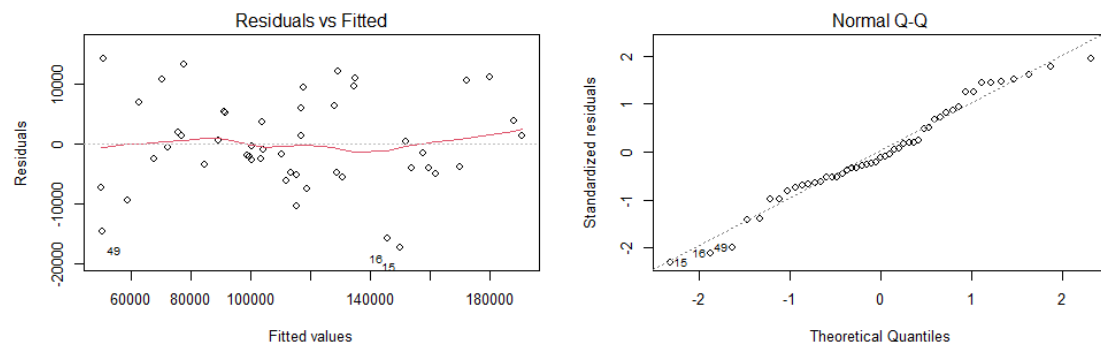


Figura 4 – Análise de resíduos, Modelo 1: Diagrama - Resíduos vs. Ajustados; Quantil - Quantil - Normal

Como é possível observar, os resíduos parecem ser homocedásticos (isto é, sua variabilidade parece ser constante) e também parecem ser normais. Testes foram realizados a fim de verificar essas hipóteses. O teste *shapiro-wilks* apontou para a normalidade dos resíduos, enquanto o teste para homocedasticidade de Goldfeld-Quandt apontou pela constância dos resíduos; de modo que o modelo parece adequado.

Como bônus, aplicou-se a transformação Box-Cox no conjunto de dados. O melhor λ encontrado foi tal que: $\lambda = 0,98989898....$ O que indica que a transformação ajustaria um modelo extremamente similar ao modelo 1.

Dito isso, o modelo 2 foi da forma:

$$\hat{Y} = 45230,00 + 0,6896gasto_ped + 0,02432gasto_mkt \quad (2)$$

Em que \hat{Y} , $gasto_ped$ e $gasto_mkt$ são definidos como anteriormente. Este modelo indica que na ausência de gastos com pesquisa e desenvolvimento e *marketing*, o lucro médio seria de 45230,00U.M, um pouco inferior ao Modelo 1. Além disso, a cada acréscimo em U.M nos gastos em pesquisa e desenvolvimento, o lucro cresceria cerca de 0,6896U.M; e a cada acréscimo nos gastos com *marketing* o lucro cresceria cerca de 0,02432U.M. Ambas as variáveis foram significantes a 95%; bem como o intercepto foi significativamente diferente de zero.

Na Figura 5, é possível ver os gráficos de resíduos vs. Ajustados e quantil-quantil (Normal) para o Modelo 2.

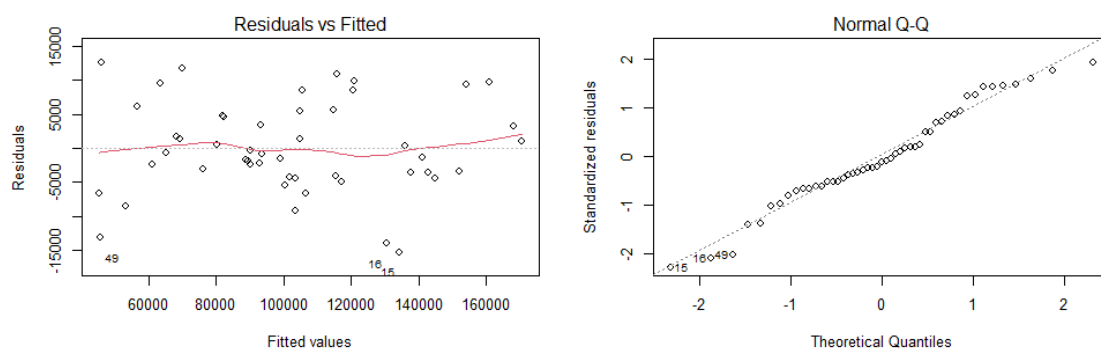


Figura 5 – Análise de resíduos, Modelo 2: Diagrama - Resíduos vs. Ajustados; Quantil - Quantil - Normal

Assim como no Modelo 1 (e como esperado, dado o valor de λ), os resíduos do Modelo 2 não parecem violar nenhum dos pressupostos analisados. Assim como anteriormente, os testes de hipóteses apontaram tanto para a normalidade, quanto para a homocedasticidade dos resíduos.

4 CONSIDERAÇÕES FINAIS

Considerando o mencionado, conclui-se que, ambos os modelos parecem adequados para explicar o lucro. Vale dizer, no entanto, que para um melhor refinamento dos ajustes, um tamanho de amostra maior seria adequado. Além disso, outras variáveis explicativas não incluídas nessa base de dados poderiam ser consideradas.

É interessante ressaltar que como o valor de λ na transformação Box-Cox foi muito próximo de 1, os modelos acabaram por ser bastante similares; isso acontece pois a transformação Box-Cox é da forma: $\frac{Y^\lambda - 1}{\lambda}$ quando λ é diferente de 0. Portanto, na situação que $\lambda = 1$, os valores de Y , a variável resposta, apenas sofrem um decréscimo de 1, o que de fato não costuma alterar muito o ajuste do modelo.

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 todos os laboratórios feitos até agora através do *link* a seguir:

<https://github.com/Mkyou/labs-regressao>

DECLARAÇÃO DE RESPONSABILIDADE

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