



**UNIVERSIDADE FEDERAL DE SERGIPE**  
**CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA**  
**DEPARTAMENTO DE ESTATÍSTICA E CIÊNCIAS ATUARIAIS**



**Ana Caroline Santos Florêncio Silva**

**ANÁLISE DE REGRESSÃO MÚLTIPLA COM ESTUDO RELACIONADO À  
EMIÇÃO DOS GASES DE EFEITO ESTUFA NA AMAZÔNIA**

São Cristóvão - SE

2020

Ana Caroline Santos Florêncio Silva

**ANÁLISE DE REGRESSÃO MÚLTIPLA COM ESTUDO RELACIONADO À  
EMISSÃO DOS GASES DE EFEITO ESTUFA NA AMAZÔNIA**

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística e Ciências Atuariais da Universidade Federal de Sergipe, como parte dos requisitos para obtenção do grau Bacharel em 2020.

Orientador: Sadraque Eneas de Figueiredo Lucena

São Cristóvão - SE

2020

Ana Caroline Santos Florêncio Silva

**ANÁLISE DE REGRESSÃO MÚLTIPLA COM ESTUDO RELACIONADO À  
EMISSÃO DOS GASES DE EFEITO ESTUFA NA AMAZÔNIA**

Trabalho de Conclusão de Curso apresentado ao Departamento de Estatística e Ciências Atuariais da Universidade Federal de Sergipe, como parte dos requisitos para obtenção do grau Bacharel em 2020.

Aprovado em 19/03/2020

Sadraque Eneas de Figueiredo Lucena  
Prof. Orientador

Allan Robert da Silva  
Convidado 1

José Rodrigo Santos Silva  
Convidado 2

São Cristóvão - SE

2020

## Resumo

Estudos voltados à preservação do meio ambiente estão se tornando cada vez mais comuns na área da Estatística, e, portanto, torna-se ainda mais necessário compreender a confiabilidade de tais estudos. Este trabalho apresenta e analisa modelos de regressão linear múltipla na predição de valores futuros das emissões de gases de efeito estufa na Amazônia. Além disso, dispõe de um breve tutorial de instalação e um guia de algumas análises estatísticas no software SAS University Edition. Os dados utilizados neste estudo são provenientes do Instituto Nacional de Pesquisas Espaciais que contém estimativas anuais das emissões dos principais gases de efeito estufa na Amazônia: gás carbônico, metano e óxido nitroso. A análise de regressão permitiu constatar que a área desmatada e o gás carbônico absorvido são as variáveis que mais se relacionam com as emissões de gases de efeito estufa e, portanto, se fazem presentes como regressores nos modelos de predição para o gás carbônico e para demais gases. A partir desses resultados podemos concluir que, à medida que a área desmatada aumenta, as emissões de gases poluentes aumentam também, visto que a decomposição e a queima do material orgânico da vegetação liberam esses gases. Por outro lado, à medida que o gás carbônico absorvido aumenta, ou seja, quanto mais áreas preservadas, menores serão as emissões de gases poluentes.

**Palavras-chave:** Regressão Múltipla. Amazônia. Efeito Estufa. SAS.

## Abstract

Studies aimed at preserving the environment are becoming more common in the Statistics area and, therefore, it gets even more necessary to understand the reliability of those studies. This work analyzes models of multiple linear regression to predict future values of greenhouse gas emissions in the Amazon forest. It has a brief installation tutorial and a guide to some of the statistical analysis on SAS University Edition. The database used in this study is from the “Instituto Nacional de Pesquisas Espaciais” that includes estimates of the main greenhouse gases emissions in the Amazon: carbon dioxide, methane and nitrous oxide. The regression analysis showed that the deforested area and the absorbed carbon dioxide are the variables that are most related to the greenhouse gas emissions and, therefore, are present as regressors in the prediction models for carbon dioxide and other gases. From the results, we can conclude that, as the deforested area increases, the emissions of polluting gases increase, since the decomposition and burning of organic material in the vegetation releases those gases. On the other hand, as the absorbed carbon dioxide increases, that is, the more preserved areas, the lower the emissions of greenhouse gases.

**Keywords:** Multiple Regression. Amazon. Rainforest. Greenhouse Effect. SAS.

## Lista de ilustrações

Figura 1 - Obter gratuitamente o software.....	20
Figura 2 - Seleção do sistema operacional do computador .....	20
Figura 3 - Acesso à página inicial do SAS .....	21
Figura 4 - Acesso aos bancos de dados contidos no SAS.....	21
Figura 5 - Importando um banco de dados.....	22
Figura 6 - Importando banco de dados – Janela para importar.....	22
Figura 7 - Importando banco de dados – Visualizando e salvando o arquivo .....	23
Figura 8 - Importando banco de dados – visualizando o código .....	23
Figura 9 - Importando banco de dados – visualizado aba LOG .....	24
Figura 10 - Filtrando uma variável.....	24
Figura 11 - Filtrando uma variável - seleções.....	25
Figura 12 - Summary Statistics - DATA.....	26
Figura 13 - Summary Statistics - OPTIONS .....	27
Figura 14 - Histograma da altura.....	29
Figura 15 - Distribution Analysis - Data.....	30
Figura 16 - Distribution Analysis - Options .....	30
Figura 17 - Histograma - Sale .....	31
Figura 18 - Q-Q Plot - Sales .....	32
Figura 19 - Linear Regression – Data .....	33
Figura 20 - Linear Regression – Model .....	34
Figura 21 - Linear Regression – Model Effects Builder .....	34
Figura 22 - Linear Regression – Options.....	35
Figura 23 - Fit Diagnostic for Weight .....	37
Figura 24 - Weight – Observed x Predicted .....	38
Figura 25 - One-Way ANOVA .....	39
Figura 26 - One-Way ANOVA – Options .....	40
Figura 27 - Box-Plot – Distribuição de MPG (Highway) por Tipo de carro .....	42
Figura 28 - SAS - Graph.....	43
Figura 29 - Bar Chart - Data.....	44
Figura 30 - Total de Vendas por Linha de produto e Região.....	45
Figura 31 - Graph – Boxplot .....	46
Figura 32 - Box plot comparando MGP (cidade) dos carros .....	47
Figura 33 - Graph – Histogram.....	48
Figura 34 - Histograma do volume estocado.....	49
Figura 35 - Graph – Line Chart.....	50
Figura 36 - Média do MGP por origem e drive train .....	51
Figura 37 - Graph – Pie Chart .....	52
Figura 38 - Total de MSRP por Região .....	53

Figura 39 - SAS Studio - Programação .....	54
Figura 40 - Gráficos das séries temporais.....	62
Figura 41 - Valores observados pelos valores preditos ( <i>C02</i> ) .....	65
Figura 42 - Valores observados pelos valores preditos (GEE).....	66
Figura 43 - Gráfico de pontos de alavanca e outliers <i>C02</i> ) .....	68
Figura 44 - Gráfico de pontos de alavanca e outliers (GEE) .....	68
Figura 45 - Gráfico da Distância de Cook ( <i>C02</i> ).....	69
Figura 46 - Gráfico da Distância de Cook (GEE).....	69
Figura 47 - Q-Q Plot dos Resíduos ( <i>C02</i> ) .....	70
Figura 48 - Q-Q Plot dos Resíduos (GEE) .....	71
Figura 49 - Gráfico dos resíduos pelos valores estimados de ( <i>C02</i> ).....	72
Figura 50 - Gráfico dos resíduos pelos valores estimados de (GEE).....	73

## Lista de tabelas

Tabela 1 - Análise Descritiva.....	28
Tabela 2 - Fitted Normal Distribution for sale (Unit Sale) .....	32
Tabela 3 - Análise de Variância (ANOVA).....	36
Tabela 4 - Estimativa de parâmetros.....	36
Tabela 5 - Análise de Variância (ANOVA).....	41
Tabela 6 - Informações e estatísticas para os modelos de regressão com a variável dependente “Emissão de CO <sub>2</sub> ” .....	63
Tabela 7 - Informações e estatísticas para os modelos de regressão com a variável dependente “Emissão de GEE” .....	63
Tabela 8 - Resultado da regressão múltipla para o conjunto de variáveis (CO <sub>2</sub> ).....	64
Tabela 9 - Resultado da regressão múltipla para o conjunto de variáveis (GEE) .....	64
Tabela 10 - Coeficientes .....	65
Tabela 11 - Teste de significância da regressão (CO <sub>2</sub> ) .....	66
Tabela 12 - Teste de significância da regressão (GEE) .....	67
Tabela 13 - Valores de inflação da variância .....	67
Tabela 14 - Testes de normalidade dos resíduos (CO <sub>2</sub> ) .....	70
Tabela 15 - Testes de normalidade dos resíduos (GEE) .....	71
Tabela 16 - Testes de variância constante.....	72
Tabela 17 - Previsão para o modelo 1 .....	73
Tabela 18 - Previsão para o modelo 2 .....	74

## Sumário

<b>1 INTRODUÇÃO</b> .....	<b>10</b>
<b>2 OBJETIVOS</b> .....	<b>12</b>
<b>2.1 Geral</b> .....	<b>12</b>
<b>2.2 Específicos</b> .....	<b>12</b>
<b>3 JUSTIFICATIVA</b> .....	<b>13</b>
<b>4 BREVE REVISÃO LITERÁRIA</b> .....	<b>14</b>
<b>4.1 A Floresta Amazônica e o efeito estufa</b> .....	<b>14</b>
<b>4.2 Regressão Linear Múltipla</b> .....	<b>16</b>
4.2.1 Pressupostos de ajuste de um modelo de regressão linear.....	16
4.2.2 Seleção de variáveis .....	17
<b>4.3 Análise de Variância (One-Way ANOVA)</b> .....	<b>17</b>
<b>5 SAS</b> .....	<b>19</b>
<b>5.1 Instalando o SAS University Edition</b> .....	<b>19</b>
<b>5.2 Abrindo arquivos do SAS</b> .....	<b>21</b>
<b>5.3 Importando banco de dados</b> .....	<b>22</b>
5.3.1 Filtrando uma variável .....	24
<b>5.4 Summary Statistics</b> .....	<b>26</b>
5.4.1 Tasks Roles .....	26
5.4.2 Exemplos e Resultados.....	28
<b>5.5 Distribution Analysis</b> .....	<b>29</b>
5.5.1 Tasks Roles .....	29
5.5.2 Exemplos e Resultados.....	31
<b>5.6 Linear Regression</b> .....	<b>33</b>
5.6.1 Tasks Roles .....	33
5.6.2 Exemplos e Resultados.....	36
<b>5.7 Análise de Variância (One-Way ANOVA)</b> .....	<b>38</b>
5.7.1 Task Role .....	38

5.7.2 Exemplos e Resultados.....	41
<b>5.8 Gráficos.....</b>	<b>42</b>
5.8.1 Bar Chart.....	43
5.8.2 Box Plot.....	45
5.8.3 Histogram.....	47
5.8.4 Line Chart.....	49
5.8.5 Pie Chart.....	51
<b>5.9 Procedures.....</b>	<b>53</b>
5.9.1 Data Step.....	54
5.9.2 Proc Means.....	55
5.9.3 Proc Univariate.....	55
5.9.4 Proc Reg.....	57
5.9.5 Proc ANOVA.....	57
<b>6 METODOLOGIA.....</b>	<b>59</b>
<b>7 RESULTADOS E DISCUSSÕES.....</b>	<b>61</b>
<b>7.1 Análise temporal do desmatamento e emissão de gases de efeito estufa na Amazônia brasileira.....</b>	<b>61</b>
<b>7.2 Análise dos Modelos de Regressão Múltipla.....</b>	<b>62</b>
<b>7.3 Diagnósticos e Análise dos Resíduos.....</b>	<b>67</b>
<b>7.4 Uso da previsão.....</b>	<b>73</b>
<b>8 Conclusões e Considerações Finais.....</b>	<b>75</b>
<b>9 Bibliografia.....</b>	<b>77</b>

## 1 INTRODUÇÃO

A Amazônia, também chamada de Floresta Amazônica, é a maior floresta tropical do mundo e se estende por mais 8 países da América do Sul: Peru, Colômbia, Venezuela, Equador, Bolívia, Guiana, Suriname e Guiana Francesa. Em território brasileiro tem-se a Amazônia Legal, que ocupa uma área de cerca de 5 milhões de quilômetros quadrados e com limites definidos por um viés sociopolítico, e também engloba parte do Cerrado e do Pantanal (SUÇUARANA, 2016).

Com o dispositivo legal da Lei 1.806 de 06.01.1953, foram incorporados à floresta brasileira os Estados do Maranhão, Goiás e Mato Grosso. Com isso, a Amazônia Brasileira passou a ser chamada de Amazônia Legal para fins de planejamento e posteriormente ter seus limites ainda mais estendidos. Segundo Costa (2020), estima-se que a floresta perdeu mais de 20% da sua superfície original e durante os últimos 10 anos, a maior parte das mudanças de cobertura da terra ocorreu por desmatamento causado pela instalação de áreas de pastagens e terras agrícolas (GIBBS et al., 2010).

No Brasil, 26,7% das emissões de gases de efeito estufa são causadas pelo desmatamento, ocorrido majoritariamente na Amazônia, cuja taxa corresponde a 14,1% das emissões nacionais (FAO, 2018). Os gases de efeito estufa (GEE) são os principais responsáveis pelo aquecimento global. Eles são compostos gasosos que absorvem radiação na frequência infravermelha e assim prendem o calor na atmosfera. Com o passar dos anos e com o aumento das atividades humanas, principalmente relacionadas ao setor industrial, o envio de GEE's para a atmosfera desencadeou o aquecimento global (GOLDEMBERG e LUCON, 2008).

Segundo o "Earth Observatory" da NASA, os principais gases de efeito estufa são o dióxido de carbono ( $CO_2$ ), metano ( $CH_4$ ) e óxido nitroso ( $N_2O$ ). O dióxido de carbono, o GEE mais relevante, é originado, principalmente, da queima de combustíveis fósseis como o petróleo, carvão e gás mineral. A segunda atividade responsável pela emissão é o desmatamento. O metano, por outro lado, tem origem majoritária nos processos biológicos como as atividades agropecuárias voltadas à produção de ruminantes, mas a produção de arroz e biomassa também emitem esse

gás. Por fim, o óxido nitroso provém da utilização de fertilizantes e de sistemas de tratamento de esgoto ou solo (JUNGES et al., 2018).

De acordo com o INPE e seu programa de monitoramento da degradação florestal na Amazônia brasileira, o mapeamento da emissão de  $CO_2$  é dividido em emissões por queima e por decomposição, ligadas diretamente ao desmatamento. Também é mapeada a quantidade de  $CH_4$  e  $N_2O$  (INPE, 2019).

Ainda segundo o INPE (2019), através do programa de monitoramento também são disponibilizados dados anuais sobre as emissões líquida de dióxido de carbono, gás carbônico proveniente da queima e decomposição de biomassa, área desmatada e gás carbônico absorvido.

Dessa forma, no presente trabalho foram realizadas análises de regressão, que consistem em obter modelos matemáticos a fim de explicar as variáveis emissão de gás carbônico e emissão de demais gases de efeito estufa por meio dos regressores área desmatada e gás carbônico absorvido. Através desses modelos foram obtidas funções matemáticas que permitiram prever cenários prováveis para as variáveis dependentes.

Este trabalho foi dividido em nove capítulos: introdução, objetivos, justificativa, breve revisão literária, SAS, metodologia, resultados e discussão, e conclusão. Na introdução estão contidas as considerações iniciais acerca do tema e a forma como foi dividido. Os objetivos do trabalho e sua justificativa são apresentadas nos capítulos subsequentes. A revisão bibliográfica aborda a fundamentação teórica utilizada. O capítulo dedicado ao SAS University Studio traz um tutorial de como instalar e manusear o software bem como uma breve introdução às funções nele contidas. A metodologia detalha o processo utilizado para obter os resultados e quais os materiais necessários. O oitavo capítulo contém os resultados do trabalho conforme objetivos apresentados e por fim, no nono capítulo estão as conclusões obtidas.

## **2 OBJETIVOS**

### **2.1 Geral**

Avaliar e quantificar a relação entre desmatamento e efeito estufa na Amazônia a partir do mapeamento realizado pelo INPE, utilizando regressão linear múltipla.

### **2.2 Específicos**

- Analisar, através de dados do INPE, o comportamento da emissão de gases de efeito estufa ao longo dos anos.
- Levantar os dados de área desmatada, emissões de dióxido de carbono, metano e óxido nitroso, e analisar as alterações ocorridas ao longo do tempo, relacionando-as com o contexto político do país.
- Identificar e modelar a relação entre a quantidade de dióxido de carbono emitido e fatores associados.

### 3 JUSTIFICATIVA

Percebendo a necessidade de informar aos governantes e cidadãos quanto ao efeito estufa e como a interferência humana na Floresta Amazônica contribui para o aumento desse efeito, o INPE realiza um mapeamento detalhado sobre fatores importantes relacionados ao problema. Com isso, este trabalho busca observar o comportamento do efeito estufa e fatores relacionados ao longo dos anos de coleta de dados. A evolução da emissão de gases de efeito estufa e o desmatamento são analisados observando o contexto histórico brasileiro em que estão inseridos.

Esta pesquisa se faz relevante pois vivemos em um meio que vem há décadas sendo deteriorado com um acúmulo exagerado de gases poluentes na atmosfera, causados principalmente por atividades humanas, que aceleram o aquecimento do planeta. As consequências disso vem se agravando com o passar dos anos e caso não haja medidas governamentais corretivas e políticas públicas de conscientização da sociedade relacionadas à emissão desses gases, problemas como aumento da temperatura do planeta, derretimento das calotas polares, aumento da incidência solar serão enfrentados.

Do ponto de vista pessoal, este trabalho se torna importante pois através dele serão levantadas respostas quanto a temática ambiental, identificando por métodos estatísticos quais aspectos naturais ou provocados pelo homem interferem no aumento das emissões de gases poluentes para poder realizar previsões a fim de contornar possíveis cenários alarmantes, além de incentivar futuras pesquisas e mudanças positivas no comportamento individual e coletivo relacionadas à preservação do meio ambiente.

## 4 BREVE REVISÃO LITERÁRIA

### 4.1 A Floresta Amazônica e o efeito estufa

É de conhecimento geral que a Floresta Amazônica possui uma enorme biodiversidade, servindo de habitat natural para milhares de espécies de animais e plantas. Além disso, é fonte de inúmeros recursos naturais como o rio Amazonas, considerado um dos rios mais extensos do planeta, responsável pelo armazenamento de 20% da água fresca do mundo e é de onde provém a água usada na agricultura, transporte e produção de alimentos do país. Embora esse posto seja das algas marinhas, a floresta foi, por muito tempo, apelidada de “pulmão do mundo” pela sua grande capacidade de produzir oxigênio através da fotossíntese (LAURANCE, 1999).

De acordo com Laurance (1999), a fotossíntese é um processo físico-químico em que a planta, ao consumir luz solar, água e o gás carbônico presente no ambiente, produz energia na forma de glicose e oxigênio. Com isso, a Amazônia está diretamente ligada à redução de gases de efeito estufa, visto que, o principal deles,  $CO_2$ , é utilizado diariamente pela flora amazônica.

A emissão de grandes quantidades de gases de efeito estufa na Amazônia é consequência de diversos processos distintos, o principal deles é a queimada da floresta em áreas que estão sendo desmatadas com o objetivo de criar áreas voltadas a agricultura e pecuária, ou a queimada provocada por incêndios florestais. Com a presença do fogo, há a liberação não só de gás carbônico, como também do metano, óxido nitroso e monóxido de carbono, principais gases do efeito estufa (FEARNSIDE, 1999)

Ainda segundo Fearnside (1999), nesse processo, há uma parte da biomassa que não é consumida pelo fogo quente e em chamas na queimada inicial, mas que posteriormente será oxidada através do processo de decomposição ou requeimadas, queimadas estas de menor temperatura, juntamente com os resíduos da floresta desmatada *a priori*, onde há maior emissão de metano e óxido nitroso.

Além dos gases de efeito estufa emitidos pela queimada inicial e pela decomposição do material restante, existe a decomposição causada pela exploração madeireira. É comum que áreas de fácil acesso terrestre e fluvial sejam alvos de

desmatamento, em que a vegetação é posteriormente derrubada por conta da agropecuária (GRAÇA, 1997). Esse processo contribui indiretamente na emissão de GEE's pois a retirada de madeiras grandes irá aumentar a potência de futuras queimadas, visto que grandes troncos demoram mais para serem decompostos, aumentando então a emissão de gases, segundo Graça.

O dióxido de carbono e o vapor d'água são grandes absorvedores das radiações solares responsáveis pelo aquecimento da atmosfera, ou seja, quanto maior a presença desses gases na atmosfera, mais quente ficará o ar. Atualmente, a preocupação da comunidade científica é o aquecimento global iminente devido às altas concentrações de GEE's, principalmente de  $CO_2$  (PACHECO e HELENE, 1990).

É sabido que, se as mudanças climáticas provocadas pelo aumento de gases de efeito estufa aumentarem ainda mais nos anos futuros, as consequências no ecossistema do planeta serão fatais (JUNGES et al., 2018). De acordo com Pacheco e Helene (1990), a umidade dos solos será alterada; a composição da fauna e flora será afetada por conta do desequilíbrio biológico; a mudança de temperatura causará alteração no regime de chuvas e conseqüentemente no escoamento das águas e vazão dos rios.

Segundo o Relatório Especial sobre o Aquecimento Global de 1,5°C, do IPCC (2018), o mundo tem que zerar as emissões líquidas o mais rápido possível, tendo como prazo máximo a metade do século, pois aumenta as chances de limitar o aquecimento global a 1,5°C. Zerar as emissões líquidas significa equilibrar as emissões de gases de efeito estufa causadas pela atividade humana com a remoção de gases da atmosfera, também conhecido como o processo natural de remoção de carbono, que consiste na remoção de dióxido de carbono e conseqüente armazenamento dele na atmosfera, com o objetivo de diminuir os impactos nas mudanças de temperatura (MULLIGAN et al., 2019). Segundo Mulligan (2019), isso pode ser feito por exemplo, através da utilização de biomassa florestal como fonte principal de energia, ou com o aumento do plantio de árvores, que absorvem o principal gás causador de efeito estufa em seu processo de respiração.

As florestas secundárias, aquelas que renascem depois do desmatamento primário de uma área, são popularmente conhecidas como "capoeiras". São compostas por uma vegetação pobre em biodiversidade e por isso não costumam ser preservadas. Contudo, a vegetação secundária é eficiente em repor a biomassa que

possuía antes do desmatamento. Durante a sua regeneração, é capaz de absorver 11 vezes mais gás carbônico do que uma vegetação primária. Tal fato é eficaz a longo prazo, evidenciando a importância da preservação dessas áreas (ANGELO, 2016).

## 4.2 Regressão Linear Múltipla

A análise de regressão linear múltipla tem como objetivo explorar as relações entre variáveis através de métodos analíticos e gráficos, que possibilitam prever valores para uma variável resposta (dependente) a partir de um grupo de variáveis independentes (também chamadas de regressores). A precisão e qualidade do modelo também são levadas em conta e dependem do conjunto de dados em análise (DENIS, 2016).

O modelo de regressão linear múltiplo é dado da seguinte forma:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i, \quad (i = 1, \dots, n), \quad (1)$$

em que  $Y_i$  é variável dependente,  $X_i$  representa a  $i$ -ésima variável independente,  $\beta_i$  é o parâmetro do modelo associado a  $X_i$  parâmetros do modelo ou coeficiente da regressão e  $\varepsilon_i$  é o erro aleatório associado à observação  $i$ .

### 4.2.1 Pressupostos de ajuste de um modelo de regressão linear

Existem algumas suposições relacionadas aos resíduos do modelo que devem ser cumpridas para que o modelo de regressão seja adequado, são elas:

- Os resíduos são independentes entre si;
- Normalidade dos resíduos;
- Homocedasticidade;
- Linearidade.

Em um modelo de regressão não é esperado que haja dependência entre os regressores. Em casos onde essa relação é forte, existe a presença de multicolinearidade e isso pode afetar as estimativas dos coeficientes da regressão e consequentemente da capacidade do modelo. A maneira mais eficaz de medir a

presença de multicolinearidade é pelo fator de inflação de variância (VIF), em que valores superiores a 10 indicam problemas na estimativa dos parâmetros (MONTGOMERY e RUNGER, 2012).

#### 4.2.2 Seleção de variáveis

A determinação de quais regressores devem ser utilizados no modelo é um dos pontos mais importantes e cuja dificuldade se encontra na seleção de um conjunto eficiente de variáveis independentes, de maneira que o modelo final seja satisfatório (MONTGOMERY e RUNGER, 2012). Existem métodos computacionais de seleção automática de variáveis comumente utilizados que se baseiam na construção do modelo por meio da adição ou remoção de variáveis. A decisão de manter ou não a variável é feita pela realização do teste F. Os métodos são: “Stepwise”, “Forward” e “Backward” (LIEBSCHER, 2012).

Ainda segundo Montgomery e Runger, a técnica mais comum de comparação entre diferentes modelos encontrados é o  $R^2$  ou  $R^2$  ajustado, porém quanto maior a quantidade de modelos sendo comparado, menor é a garantia de que essa técnica funcione. Outra forma de comparar modelos é por meio dos critérios de Akaike (AIC) e Bayesiano (BIC). Por esta estratégia, o modelo que obtiver os menores valores de AIC e BIC, serão considerados melhores. Não obstante, esses critérios são facilmente influenciados pela quantidade de regressores (EMILIANO et al., 2010).

### 4.3 Análise de Variância (One-Way ANOVA)

Segundo Vieira (2006), a análise de variância é uma técnica estatística cujo objetivo é comparar as médias de pelo menos duas populações distintas. A One-Way ANOVA analisa experimentos que envolvem uma variável resposta e um fator de controle. Para realizar o teste da ANOVA de um fator deve-se estabelecer as hipóteses nula e alternativa, onde:

$H_0$ : As médias populacionais são iguais.

$H_1$ : Ao menos uma das médias é diferente das demais.

Para realizar a análise de variância, deve-se levar em conta os seguintes pressupostos:

- a) As amostras devem ser independentes, ou seja, os dados são coletados aleatoriamente dentro do espaço amostral;
- b) As variâncias entre os grupos devem ser homogêneas e, portanto, cada tratamento contribuirá da mesma forma na soma dos quadrados;
- c) Os resíduos devem seguir uma distribuição normal, entretanto esse pressuposto pode ser violado dependendo do tamanho amostral.

## 5 SAS

O SAS é um software estatístico desenvolvido em meados da década de 70, sendo atualizado com uma versão “point-and-click” em 2004, o que permitiu ao usuário manipular dados e fazer análises sem ter experiência em programação (TOLBERT, 2010). Com essa versão point-and-click, o SAS disponibilizou um guia de referência, SAS Studio Task Reference Guide, que consiste em fazer breves explicações sobre as funções e respectivas técnicas estatísticas, juntamente com exemplos para auxiliar o usuário a fazer uso de tais funções.

A priori, SAS era a sigla correspondente a Statistical Analysis System, mas com as atualizações e maior variedade de serviços ofertados, passou a ser chamado somente de SAS.

O software possui mais de 200 componentes que providenciam diferentes funções, as principais são (CHAPMAN, 2012):

- Base SAS: responsável pela realização de procedimentos estatísticos básicos e gerenciamento de dados;
- SAS/STAT: permite fazer análises estatísticas como ANOVA, análise de covariância, regressão linear e logística, predição de modelos de regressão linear, análise de sobrevivência, análise multivariada e outros;
- SAS/GRAPH: possibilita a elaboração de gráficos detalhados e customizados;
- SAS/ETS: permite realizar análises econométricas e de séries temporais;
- SAS/IML: responsável pela linguagem matricial;
- SAS/QC: permite fazer controle de qualidade;
- Enterprise Guide: possibilita que análises estatísticas sejam feitas apenas apontando e clicando.

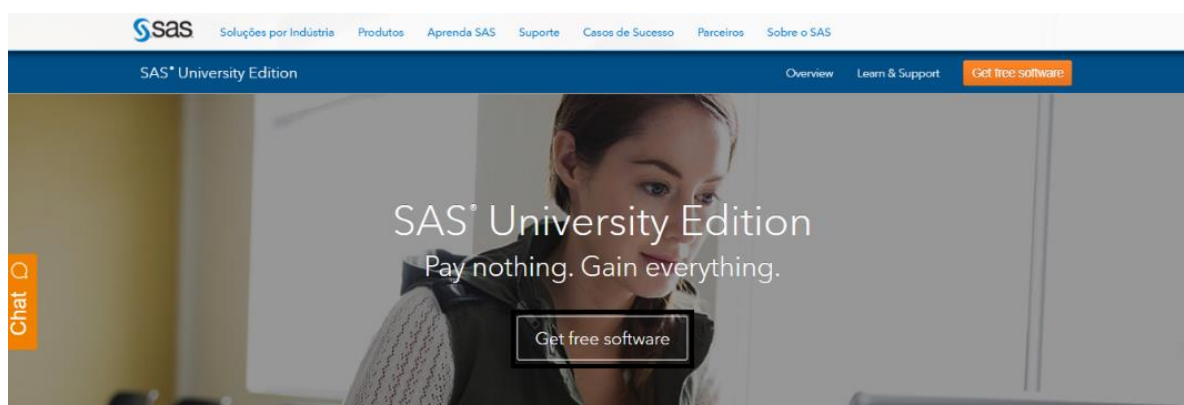
### 5.1 Instalando o SAS University Edition

A fim de possibilitar que estudantes tenham maior acesso ao software, o SAS oferece a opção de fazer o download gratuito da versão SAS University Edition, com

fins universitários e não comerciais. A seguir apresentado o passo-a-passo para download do SAS University Edition versão gratuita.

Ao abrir a página inicial do SAS ([https://www.sas.com/pt\\_br/software/university-edition.html](https://www.sas.com/pt_br/software/university-edition.html)) a opção **Get free software** deverá ser selecionada, conforme Figura 1.

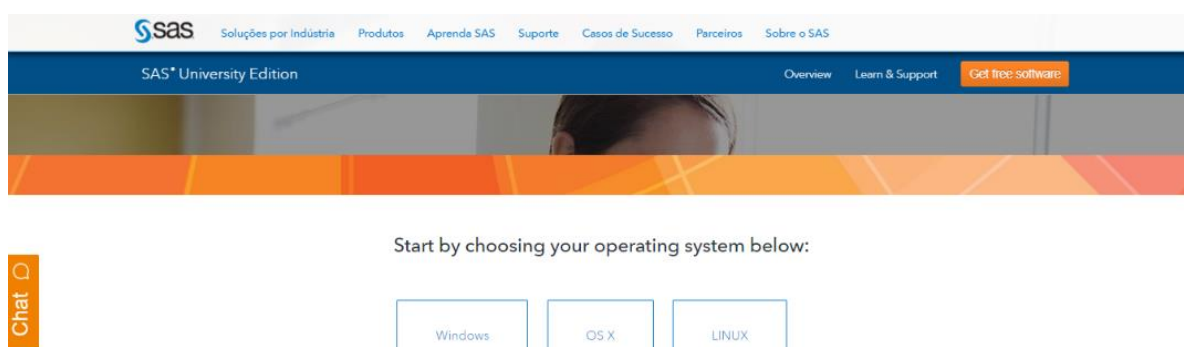
Figura 1 - Obter gratuitamente o software



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Feito isso, o usuário deverá selecionar qual versão do SAS irá fazer download. As opções são **Windows**, **OSX** ou **Linux**, conforme Figura 2. Com isso, terá acesso ao passo-a-passo para instalação do software de virtualização (VirtualBox) e do próprio SAS.

Figura 2 - Seleção do sistema operacional do computador

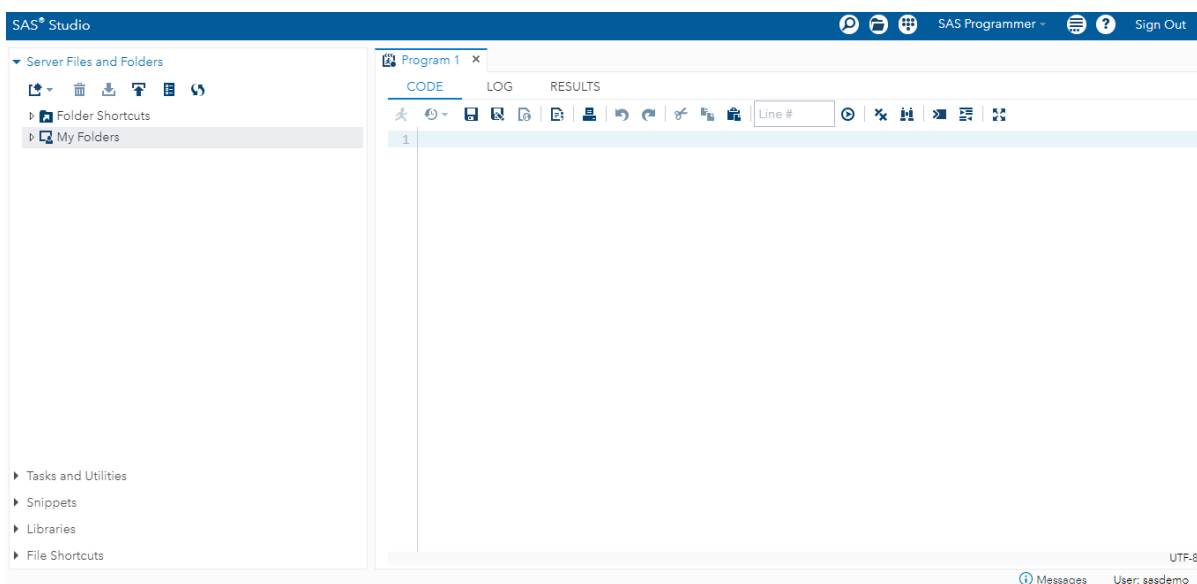


Fonte: Imagem retirada do site <http://sas.com/> (2019)

## 5.2 Abrindo arquivos do SAS

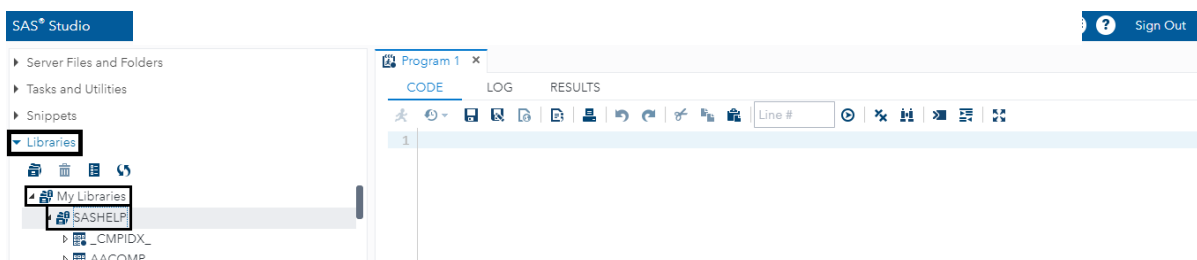
Ao acessar a página inicial do SAS, o usuário irá visualizar a Figura 3. As mais importantes funções estatísticas de análise de dados estão contidas dentro de **Tasks and Utilities**, entretanto, para utilizá-las é preciso ter um banco de dados vinculado ao SAS. O próprio software, por sua vez, possui bancos de dados próprios e para acessá-los, primeiramente deve-se acessar **Libraries**, depois **My Libraries** e por fim **SASHELP**, tendo enfim acesso a todos os bancos já contidos no SAS, como na Figura 4.

Figura 3 - Acesso à página inicial do SAS



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Figura 4 - Acesso aos bancos de dados contidos no SAS



Fonte: Imagem retirada do site <http://sas.com/> (2019)

### 5.3 Importando banco de dados


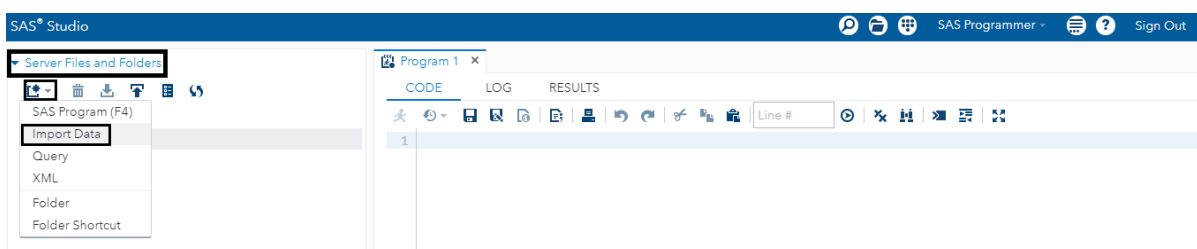
Como informado na página de instalação do SAS, é preciso criar uma pasta intitulada **SASUniversityEdition**, e dentro dela, outra pasta chamada **myfolders**. Para importar algum banco de dados, necessita-se que, primeiramente, o arquivo esteja dentro da pasta **myfolders**. Em seguida, no próprio SAS, para importar o banco de dados segue-se o menu **Server Files and Folders** >  > **Import Data**, conforme a Figura 5.

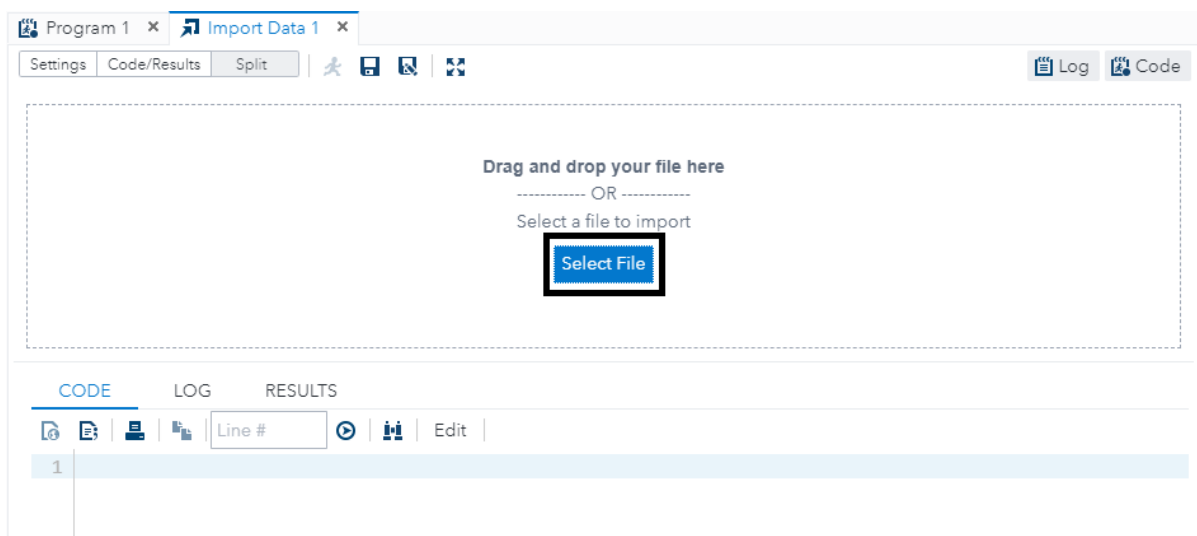
Figura 5 - Importando um banco de dados



Fonte: Imagem retirada do site <http://sas.com/> (2019)

O SAS University Edition suporta arquivos no formato .xls; .mdb; .dbf; .db; .txt; .csv; .asc; .tab; .htm e .html. Ao clicar na tarefa **Import Data**, uma janela irá abrir para que o arquivo externo seja selecionado ou arrastado para a área determinada. Essa janela é mostrada na Figura 6.

Figura 6 - Importando banco de dados – Janela para importar



Fonte: Imagem retirada do site <http://sas.com/> (2019)



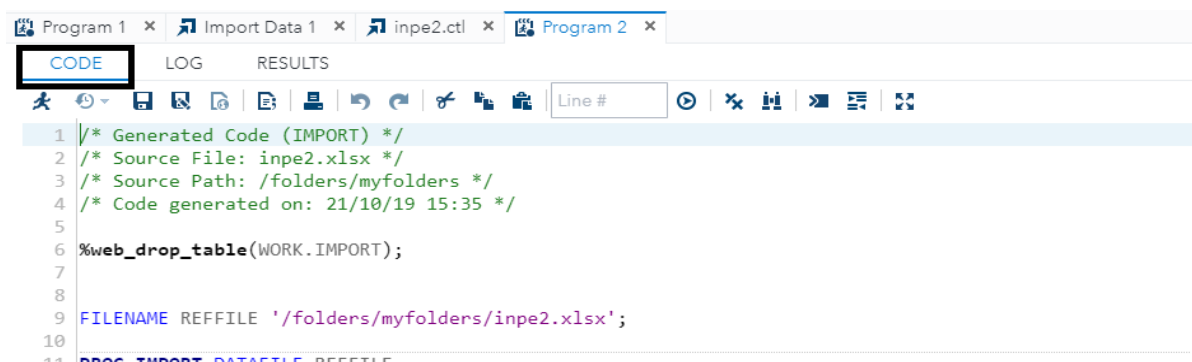
Ao seleccionar o arquivo desejado, ele será aberto na interface principal do SAS e para visualizá-lo deve-se seguir  > **OUTPUT DATA**. Para manter o banco de dados em uma extensão SAS, deve-se clicar no símbolo  e assim, o arquivo poderá ser modificado dentro do software e evita que o mesmo processo de importação seja refeito sempre que o arquivo for importado. Na aba **CODE** o código de importação pode ser visualizado e editado, nela encontra-se informações como o local do arquivo bem como o nome dele. Já em **LOG** são acessados os possíveis erros, avisos e notas importantes referente aos comandos produzidos. Essas informações podem ser vistas na Figura 7, 8 e 9.

Figura 7 - Importando banco de dados – Visualizando e salvando o arquivo



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Figura 8 - Importando banco de dados – visualizando o código



Fonte: Imagem retirada do site <http://sas.com/>

Figura 9 - Importando banco de dados – visualizado aba LOG

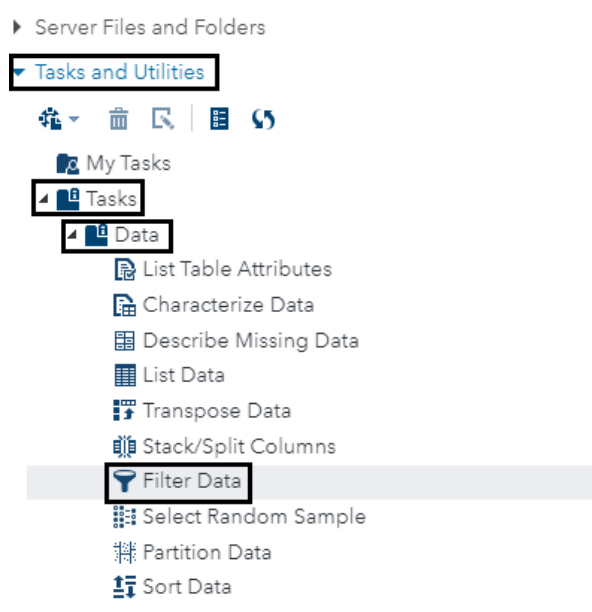


Fonte: Imagem retirada do site <http://sas.com/> (2019)

### 5.3.1 Filtrando uma variável

Se desejar obter alguma informação específica sobre determinada variável, deverá colocar um filtro nela. Para tal, vá em **Tasks and Utilities > Tasks > Data > Filter Data**, conforme mostrado na Figura 10.

Figura 10 - Filtrando uma variável



Fonte: Imagem retirada do site <http://sas.com/> (2019)

O banco de dados deverá ser selecionado em **DATA**, clicando no símbolo indicado na Figura 11. A variável que desejar realizar o filtro deverá ser adicionada em **Variable 1**, clicando no símbolo correspondente mostrado na Figura 11. No campo **Comparison** deve-se escolher qual o critério de comparação que a variável deve seguir. Por exemplo, ao selecionar **less than**, as variáveis menores que o valor selecionado em **Value type** serão filtradas. Para que todas as outras variáveis do banco de dados estejam baseadas nesse filtro, tenha certeza que **All variables** seja selecionada em **OUTPUT DATA SET > Variables to include**. Posteriormente aparecerá uma janela mostrada na Figura 11, onde deverão ser feitas as demais seleções para a filtragem.

Figura 11 - Filtrando uma variável - seleções

The screenshot shows the 'Filter Data' dialog box in SAS. The 'DATA' section is set to 'WORK.IMPORT'. Under 'FILTER 1', 'Variable 1' is set to 'Column'. The 'Comparison' is set to 'Less than', and the 'Value type' is set to 'Enter a value'. The 'Value' field is empty. The 'Logical' option is set to '(none)'. Under 'OUTPUT DATA SET', the 'Data set name' is 'work.filter0002'. The 'Variables to include' is set to 'All variables'. The 'Show Output Data' checkbox is unchecked.

Fonte: Imagem retirada do site <http://sas.com/> (2019)

## 5.4 Summary Statistics

A estatística descritiva é uma área da estatística cujo objetivo é resumir uma série de valores da mesma natureza, organizá-los e descrevê-los através de três maneiras: tabelas, gráficos e medidas descritivas (BUSSAB e MORETTIN, 2003).

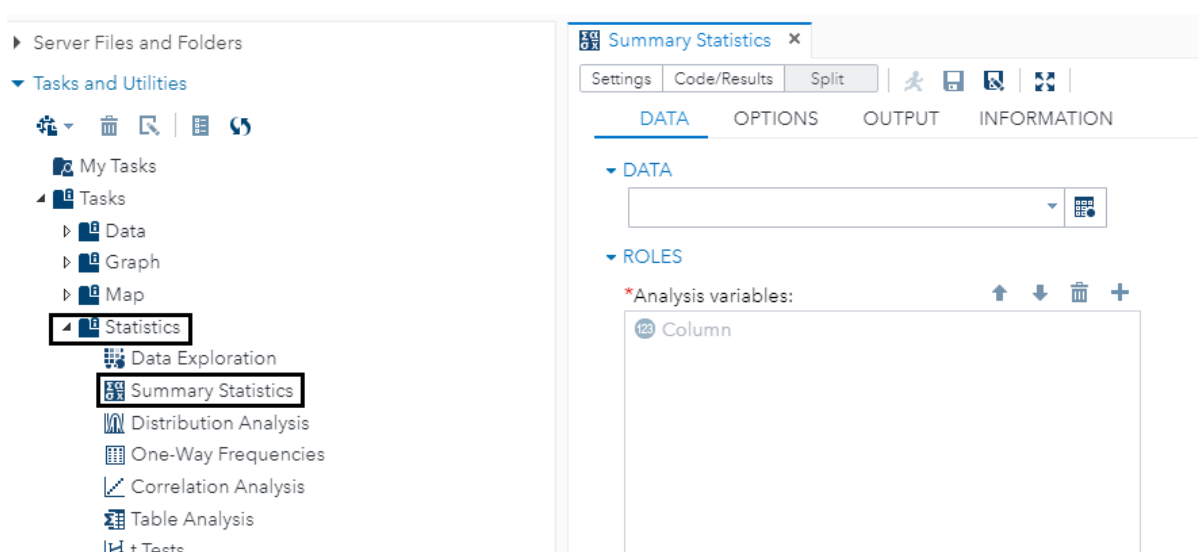
Segundo Bussab e Morettin (2003), a tabela consiste em um quadro que sintetiza um conjunto de dados. O gráfico é uma forma visual de sintetizar os dados que possibilita uma compreensão rápida do fenômeno estudado.

Para evidenciar as características dos dados observados nas tabelas de maneira isolada ou comparativamente, faz-se uso de números ou estatísticas, que estão divididas em medidas de posição e medidas de dispersão.

### 5.4.1 Tasks Roles

Permite fazer estatísticas descritivas para as variáveis, medidas de dispersão e de posição, histogramas e boxplots. Para acessar essa tarefa, siga **Tasks and Utilities > Tasks > Statistics > Summary Statistics**. Após seguir esse procedimento, uma nova aba surgirá contendo espaços para seleções de banco de dados, variáveis para estudo, como mostrado nas Figuras 12 e 13.

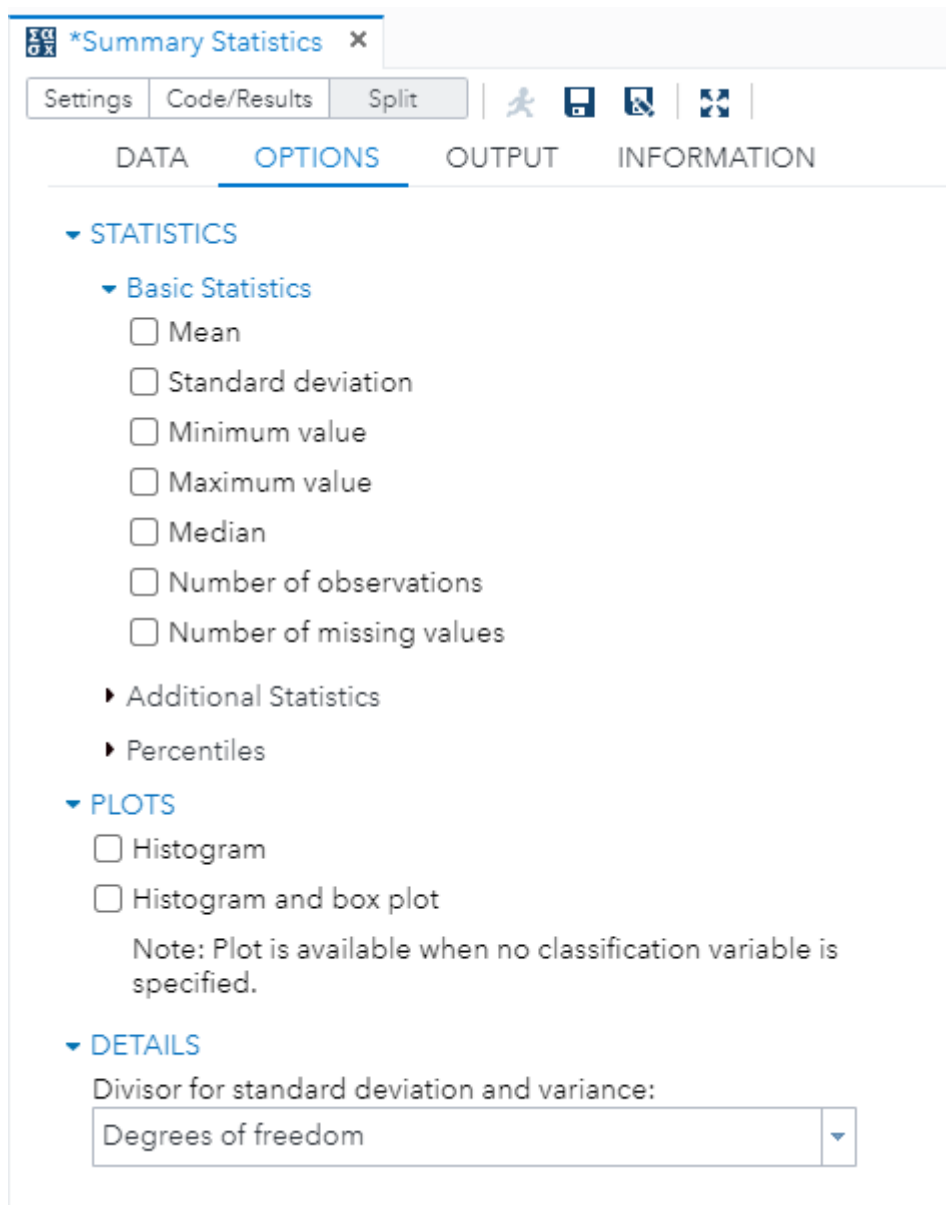
Figura 12 - Summary Statistics - DATA



Fonte: Imagem retirada do site <http://sas.com/> (2019)

- **Analysis variables:** neste espaço as variáveis a serem analisadas serão selecionadas. Note que o software orienta, através de um símbolo, que ali só poderão ser selecionadas variáveis numéricas. Além disso, a presença do asterisco vermelho indica a obrigatoriedade de preenchimento deste campo.
- **Classification variables:** aqui serão selecionadas, caso necessário, variáveis categóricas ou numéricas cuja função será dividir as variáveis anteriores em subgrupos de análise.

Figura 13 - Summary Statistics - OPTIONS



Fonte: Imagem retirada do site <http://sas.com/> (2019)

- **Basic Statistics:** neste campo serão selecionadas as estatísticas básicas como: média, mediana, desvio padrão; estatísticas adicionais como: erro, variância, moda, coeficiente de variação; e percentis que se deseja obter.
- **Plots:** caso opte por não selecionar nenhuma variável categórica, é possível obter um histograma ou box plot.
- **Divisor for standard deviation and variance:** aqui é possível escolher entre graus de liberdade e número de observações para ser o divisor das fórmulas do desvio padrão e variância.

#### 5.4.2 Exemplos e Resultados

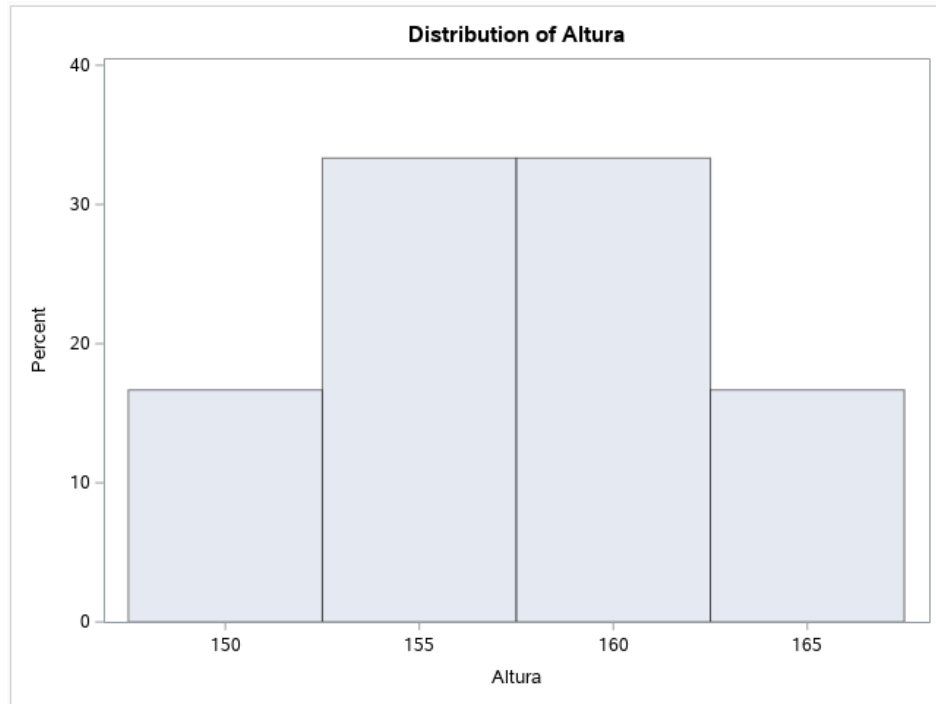
Neste exemplo, solicitou-se um banco de dados contendo informações sobre idade, altura, peso e sexo de estudantes. Encaixando as variáveis numéricas idade, altura e peso no campo **Analysis variables**, selecionando os itens **Mean, Standard deviation, Minimum value, Maximum value, Median** e **Number of observations**, obtém-se a tabela vista na Tabela 1 e no histograma da Figura 14 é mostrada a distribuição da altura dos estudantes.

Tabela 1 - Análise Descritiva

Variable	Label	Mean	Std Dev	Minimum	Maximum	Median	N
Idade	Idade	12.83	1.72	10.00	15.00	13.00	6
Peso	Peso	53.50	4.37	48.00	60.00	53.50	6
Altura	Altura	157.83	5.77	151.00	166.00	157.50	6

Fonte: Elaborado pela autora

Figura 14 - Histograma da altura



Fonte: Elaborado pela autora

## 5.5 Distribution Analysis

Distribuições de probabilidade descrevem o comportamento aleatório de um fenômeno dependente do acaso (JAMES, 1981). Ou seja, é um modelo matemático que descreve o comportamento de uma variável aleatória ou contínua. Ao analisar a distribuição de uma população, é feito um estudo mais aprofundado sobre fenômenos reais, ajudando na compreensão do mesmo.

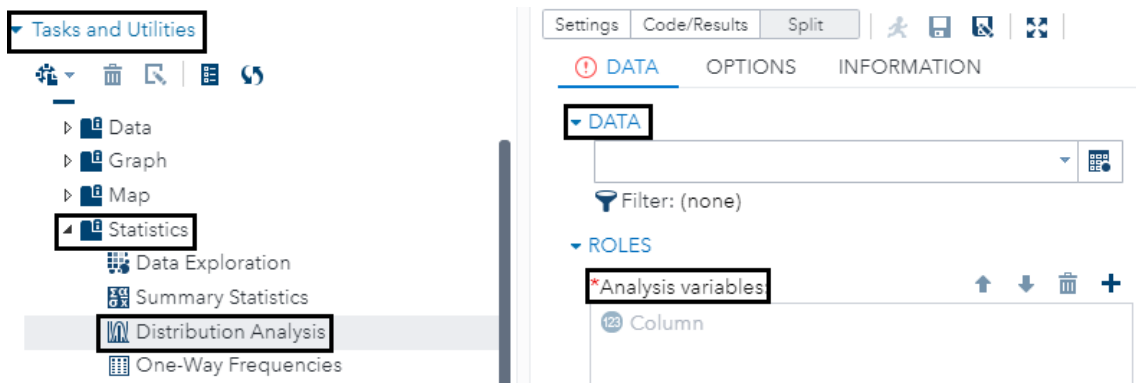
No SAS, essa tarefa provê informações sobre a distribuição dos dados de determinada variável. Possui uma variedade de gráficos como histogramas, gráficos de probabilidade e qq plots que podem ser usados nessa análise (SAS Studio Task Reference Guide).

### 5.5.1 Tasks Roles

Após expandir a pasta **Statistics** na seção **Tasks**, dê um clique duplo em **Distribution Analysis** e a interface irá abrir. Na aba **DATA**, selecione o banco de

dados. Na função **Analysis variables**, deverá escolher no mínimo uma variável numérica, como mostrado na Figura 15.

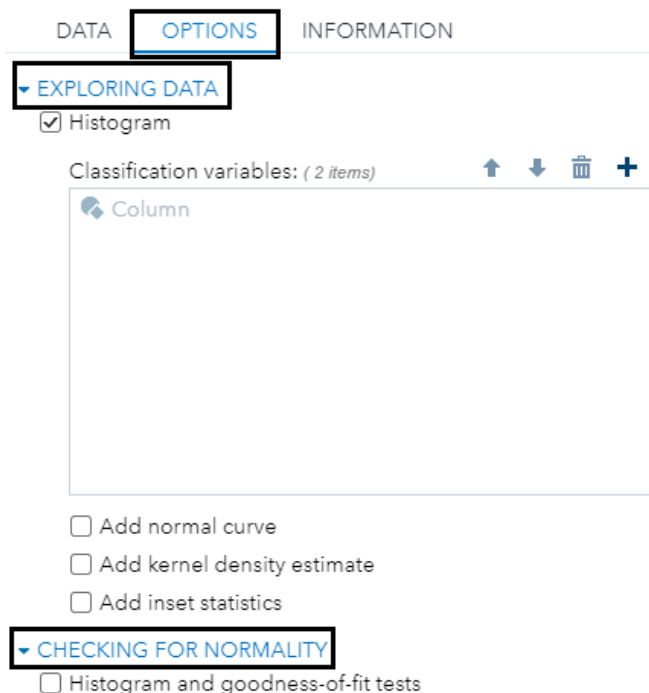
Figura 15 - Distribution Analysis - Data



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Em **OPTIONS**, poderá optar por obter um histograma comparativo, para isso terá de adicionar até duas variáveis categóricas. Consegue ainda adicionar uma curva normal no gráfico, estimativa de densidade kernel ou inserir estatísticas descritivas. O grupo **CHECKING FOR NORMALITY** permite que seja plotado um histograma junto com um teste de bondade de ajustamento, como pode ser visualizado na Figura 16.

Figura 16 - Distribution Analysis - Options

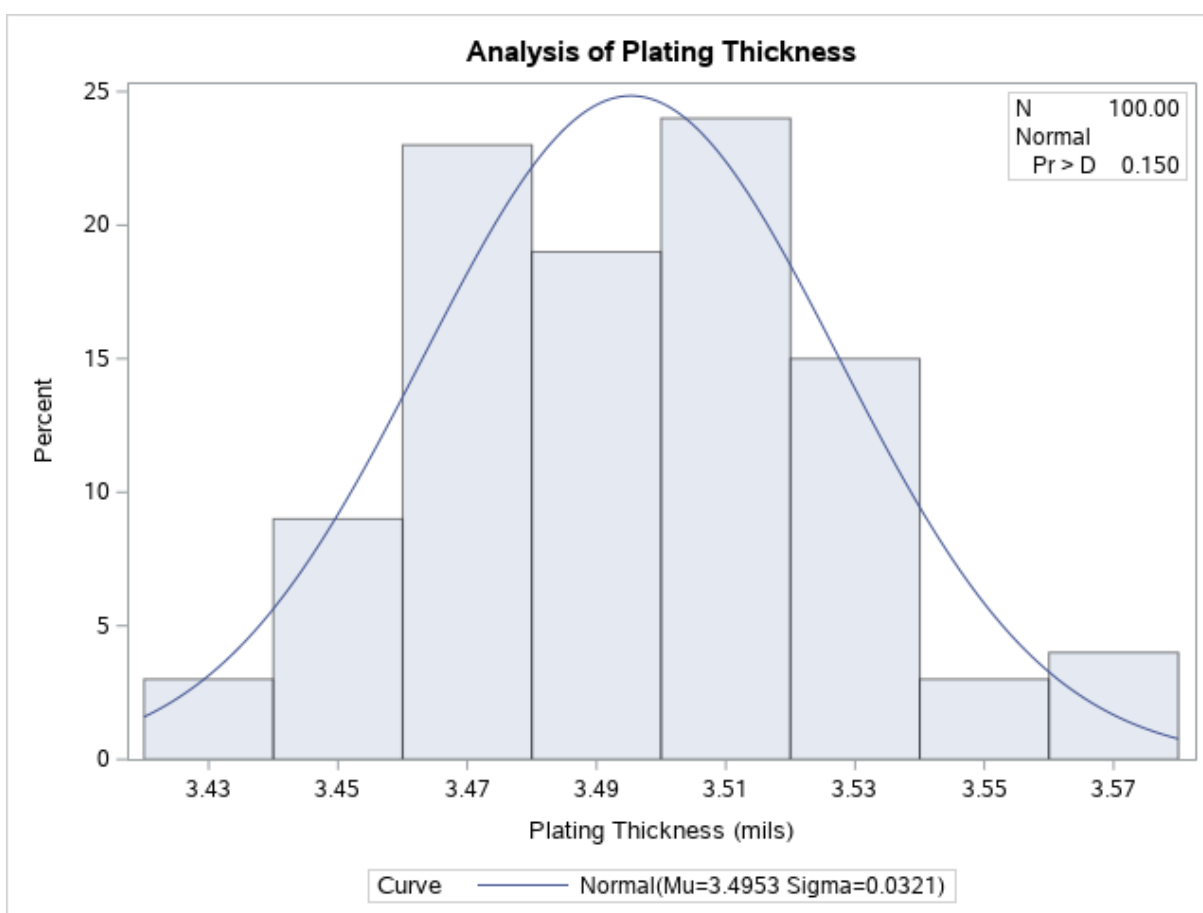


Fonte: Imagem retirada do site <http://sas.com/> (2019)

### 5.5.2 Exemplos e Resultados

Utilizando o banco de dados do SAS, **PRICEDATA**, que contém informações sobre vendas de produtos, será feita uma análise de distribuição das vendas por região. A variável **sale** deverá ser selecionada em **Analysis variables**. Na aba **OPTIONS** e grupo **Exploring Data**, a variável **regionName** deve ser selecionada em **Classification variables**. Para a verificação de normalidade dos dados, em **Checking for Normality**, deve-se selecionar as opções **Histogram and goodness-of-fit tests** e **Normal quantile-quantile plot** (SAS Studio Task Reference Guide). Como é possível observar pelos resultados contidos Figura 17, na Tabela 2 e na Figura 18, os dados são normalmente distribuídos.

Figura 17 - Histograma - Sale



Fonte: Elaborado pela autora

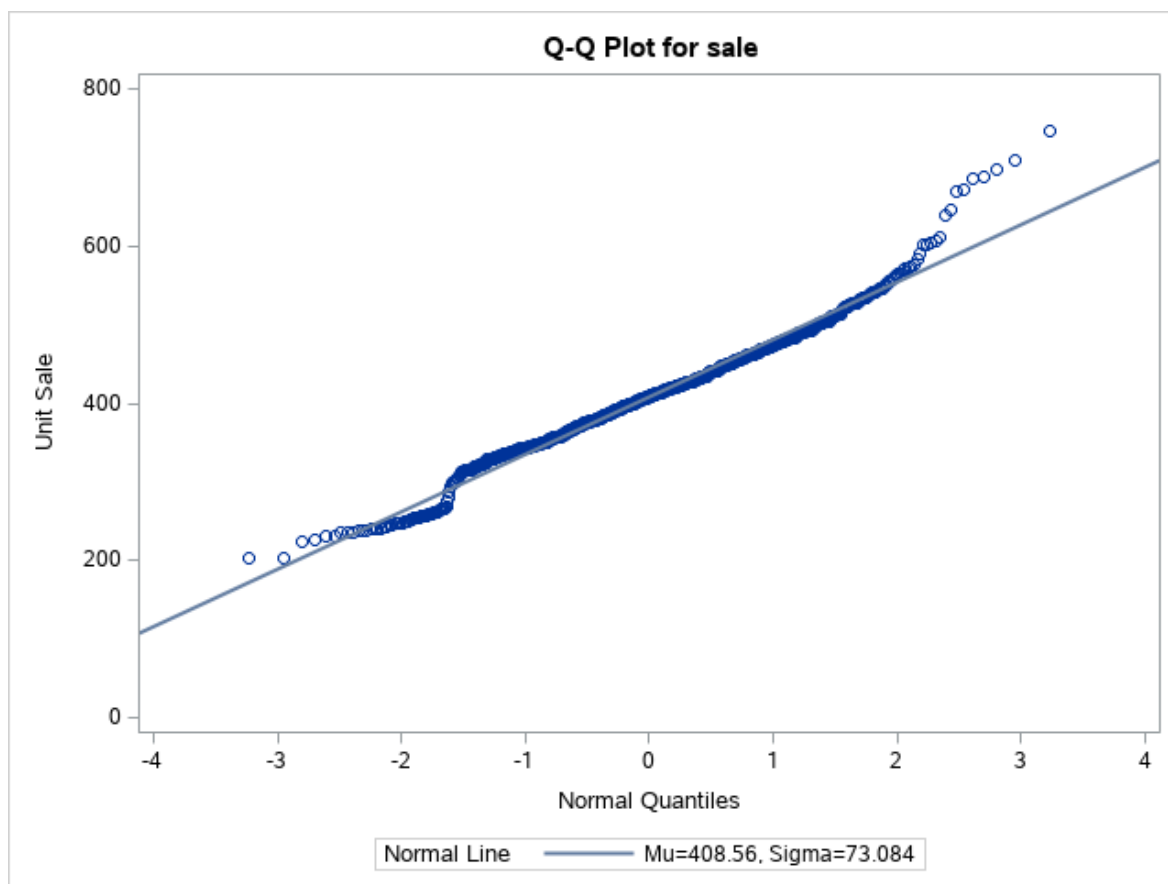
Tabela 2 - Fitted Normal Distribution for sale (Unit Sale)

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.041	Pr > D	< 0.010
	W-			
Cramer-von Mises	Sq	0.440	Pr > W-Sq	< 0.005
Anderson-Darling	A-Sq	3.460	Pr > A-Sq	< 0.005

Fonte: Elaborado pela autora

Os testes de Kolmogorov-Smirnov, Cramer-von Mises e Anderson-Darling são testes não-paramétricos cujo objetivo é verificar se a distribuição de um conjunto de dados adere à Distribuição Normal. Neste caso, a hipótese nula de normalidade é rejeitada ( $p < 0,05$ ).

Figura 18 - Q-Q Plot - Sales



Fonte: Elaborado pela autora

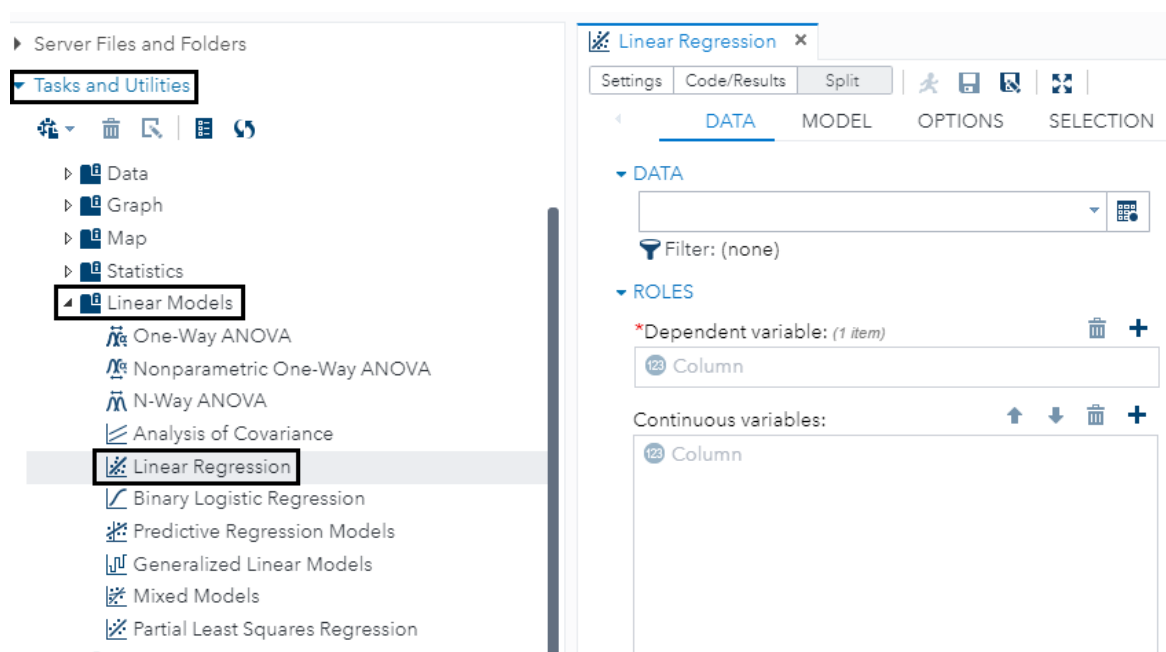
## 5.6 Linear Regression

A regressão linear é uma técnica estatística que consiste em verificar a existência de relação entre uma variável dependente com uma ou mais variáveis independentes.

### 5.6.1 Tasks Roles

A função Linear Regression no SAS é responsável por ajustar um modelo linear para prever resultados da variável dependente a partir de outras variáveis independentes. Essa função provê estatísticas e gráficos para a interpretação dos resultados (SAS Studio Task Reference Guide). Para acessar essa função, siga **Tasks and Utilities > Linear Models > Linear Regression**, conforme Figura 18. Em **DATA**, selecione o banco de dados que deseja analisar, em **Dependent variable**, especifica-se a variável numérica que deseja usar como variável dependente no modelo de regressão. Em **Continuous variables**, deve-se indicar as variáveis numéricas independentes.

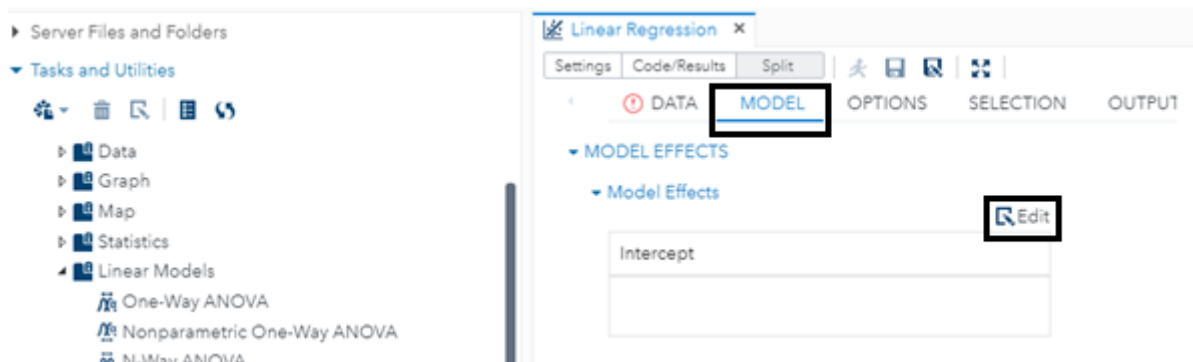
Figura 19 - Linear Regression – Data



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Na aba **MODEL**, é onde o modelo será criado. Em **Edit**, será possível selecionar dentre as variáveis independentes, determinando assim se o modelo será simples ou múltiplo, conforme Figura 20.

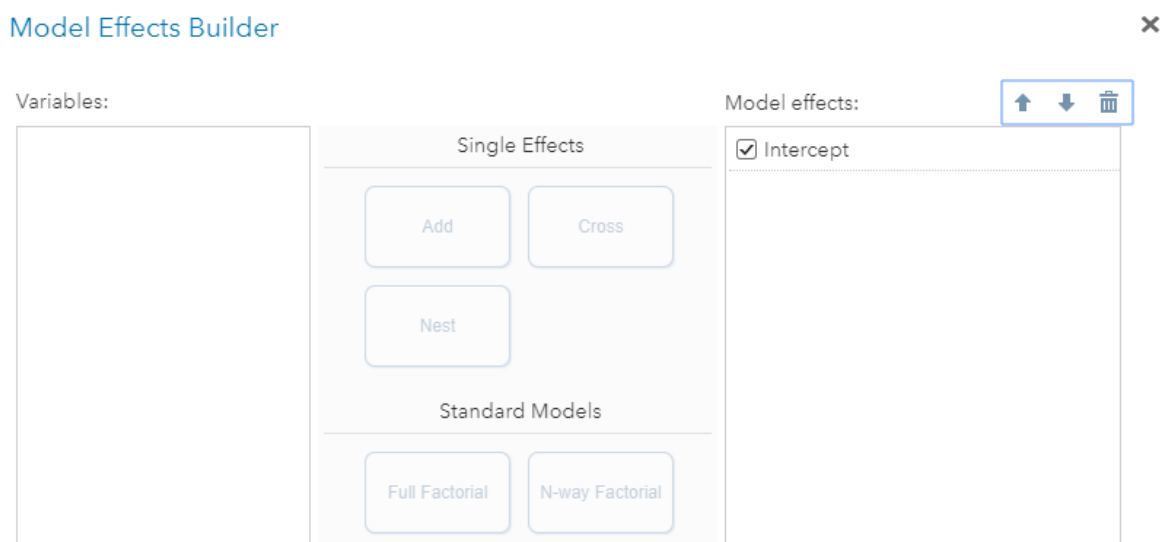
Figura 20 - Linear Regression – Model



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Para criar um modelo tradicional, através da janela **Model Effects Builder**, indique as variáveis que deseja analisar e clique em **Add**. Também é possível criar um modelo com interações, neste caso, deve-se selecionar duas ou mais variáveis e clicar em **Cross**. Além disso, essa aba permite realizar modelos logísticos, onde a variável resposta é binária, como mostrado na Figura 21.

Figura 21 - Linear Regression – Model Effects Builder

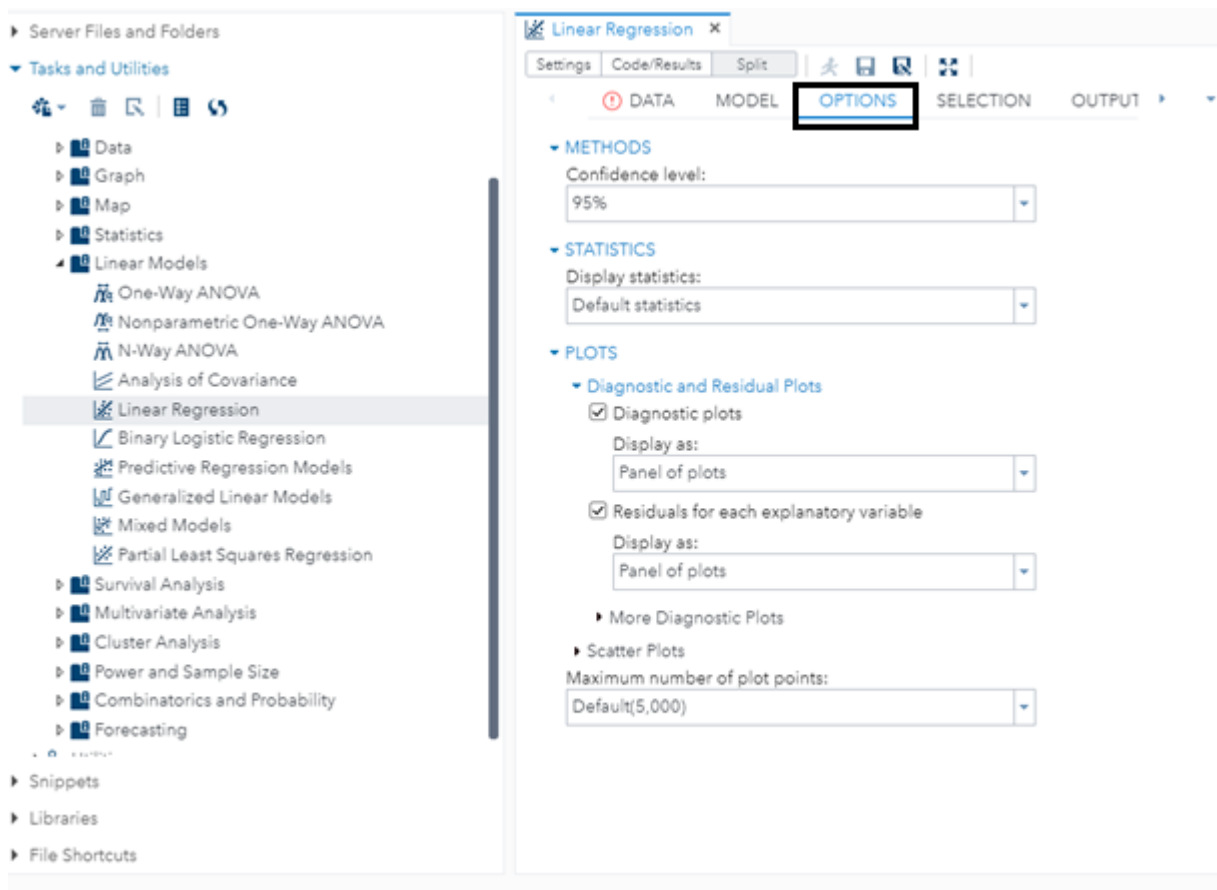


Fonte: Imagem retirada do site <http://sas.com/> (2019)

Em **OPTIONS** é possível configurar as opções do modelo através das seguintes seções, mostradas na Figura 22.

- **Confidence level:** especifica o nível de significância usado na construção de intervalos de confiança;
- **Statistics:** é possível incluir estatísticas padrão ou estatísticas adicionais como estimação de parâmetros, soma dos quadrados, correlações parciais, diagnósticos com análise de resíduos e valores preditos, análise de colinearidade e heterocedasticidade;
- **Plots:** inclui gráficos de diagnóstico e de resíduos.

Figura 22 - Linear Regression – Options



Fonte: Imagem retirada do site <http://sas.com/> (2019)

## 5.6.2 Exemplos e Resultados

Utilizando o banco e dados **CLASS** já incluso no SAS que contém informações sobre peso, altura e sexo de estudantes, e com o objetivo de prever o peso de um estudante através de sua altura, aloque a variável **Weight** em **Dependent variable**, **Age** e **Height** em **Continuous variables**. Na aba **MODEL**, selecione as variáveis **Age** e **Height** e clique em **Add** (SAS Studio Task Reference Guide). Com isso, é obtido um modelo onde peso é a variável dependente e idade e altura, as variáveis independentes, conforme resultados apresentados na Tabela 3, Tabela 4, Figura 23, Figura 24.

Tabela 3 - Análise de Variância (ANOVA)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7215.6371	3607.8185	27.23	< .0001
Error	16	2120.0997	132.5062		
Corrected Total	18	9335.7368			

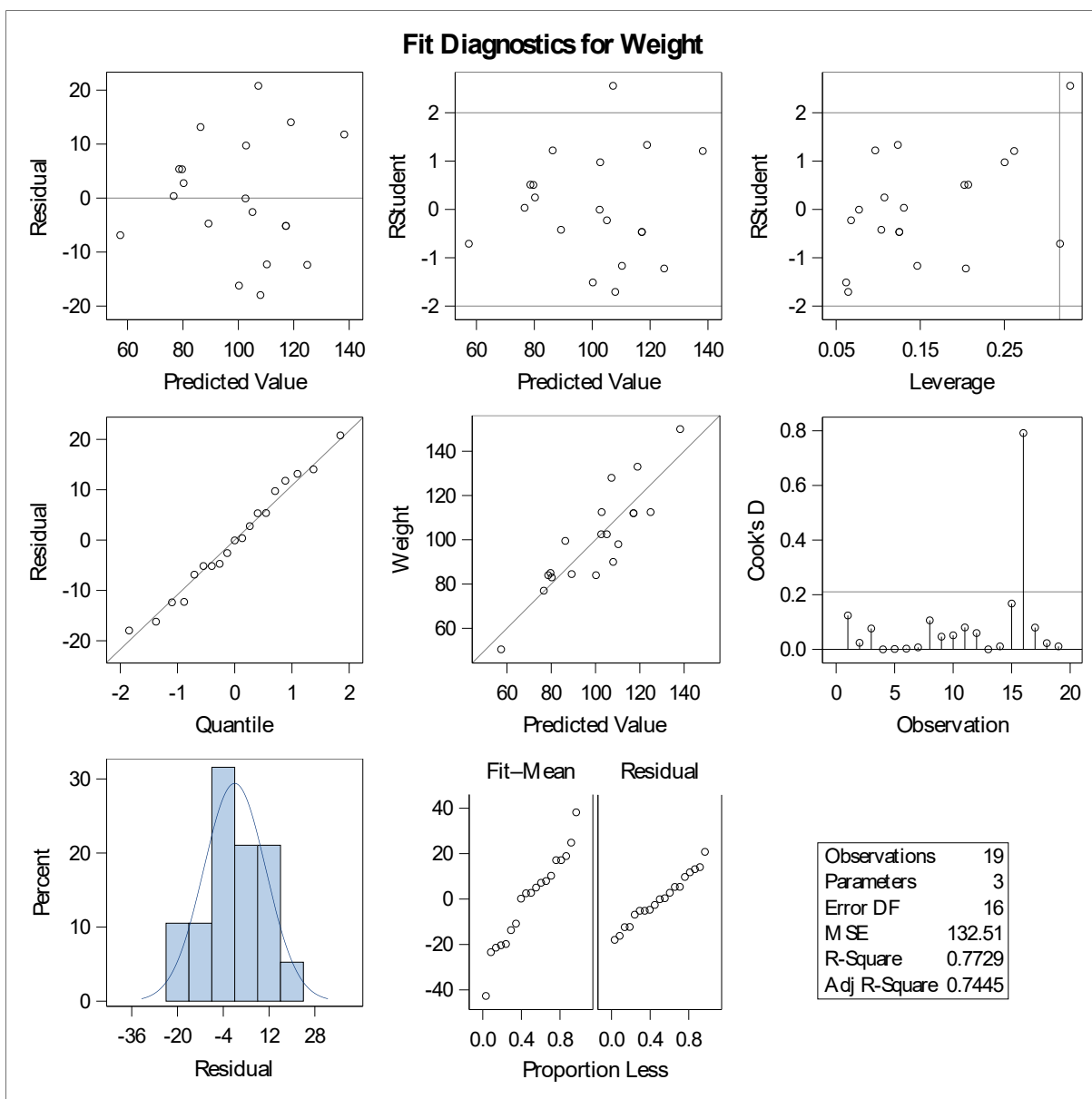
Fonte: Elaborado pela autora

Tabela 4 - Estimativa de parâmetros

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-141.2237	33.3830	-4.23	0.0006
Age	1	1.2783	3.1101	0.41	0.6865
Height	1	3.5970	0.9054	3.97	0.0011

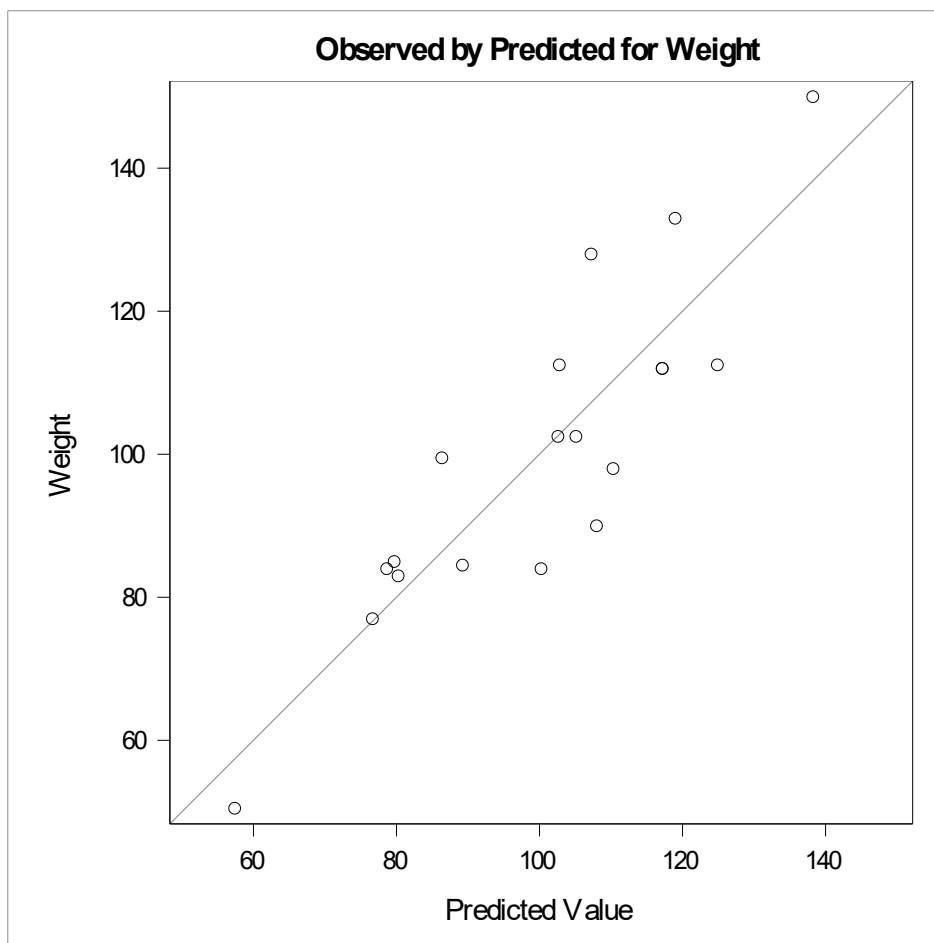
Fonte: Elaborado pela autora

Figura 23 - Fit Diagnostic for Weight



Fonte: Elaborado pela autora

Figura 24 - Weight – Observed x Predicted



Fonte: Elaborado pelo autor

## 5.7 Análise de Variância (One-Way ANOVA)

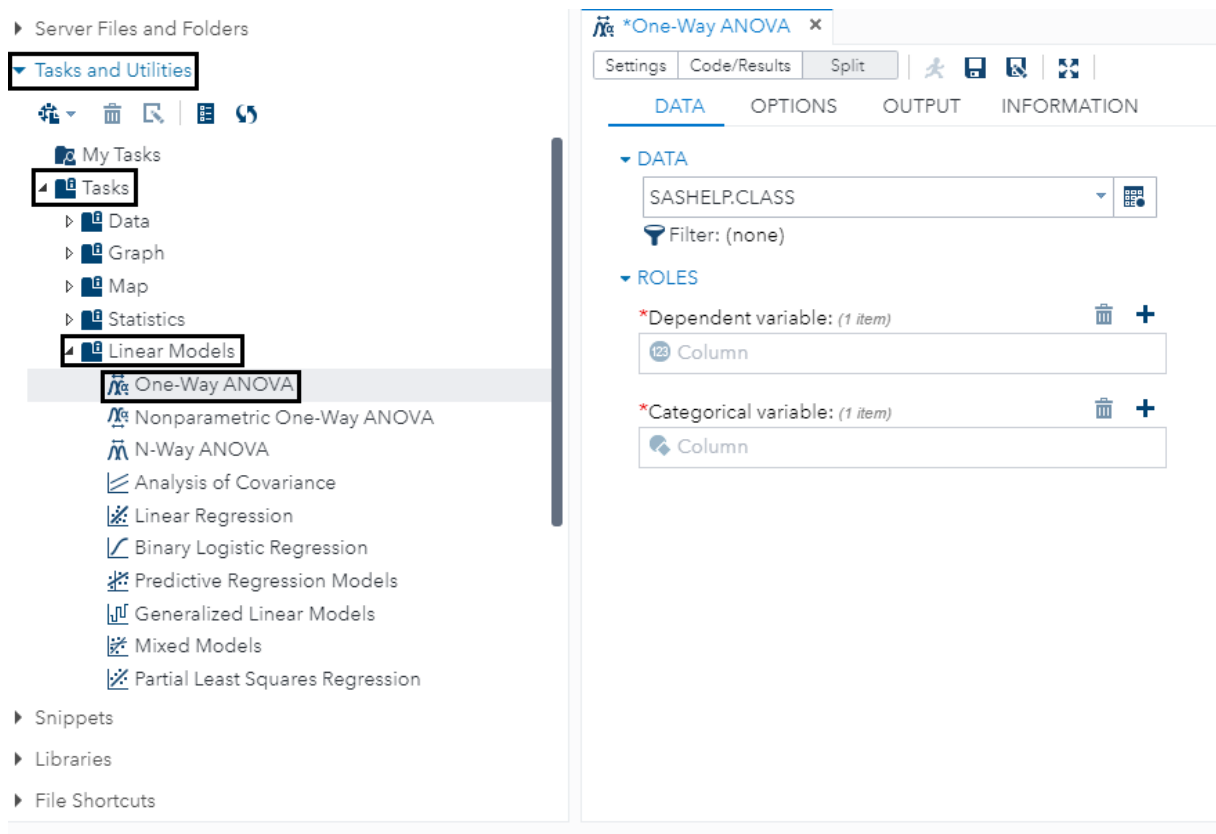
A análise de variância ou ANOVA é uma técnica utilizada para verificar se há diferença significativa entre duas ou mais médias populacionais provenientes de diferentes grupos.

### 5.7.1 Task Role

A tarefa de análise de variância (ANOVA) de um fator no SAS permite que sejam feitos testes e gráficos para verificar a diferença entre as médias de uma única variável categórica e uma variável contínua dependente (SAS Studio Task Reference Guide).

Para acessar essa tarefa, siga **Tasks and Utilities > Tasks > Linear Models > One-Way ANOVA**. Será então aberta a janela conforme Figura 25, onde deve-se indicar a variável contínua em **Dependent variable** e em **Categorical variable**, uma variável categórica que deve ter dois ou mais valores distintos.

Figura 25 - One-Way ANOVA



Fonte: Imagem retirada do site <http://sas.com/> (2019)

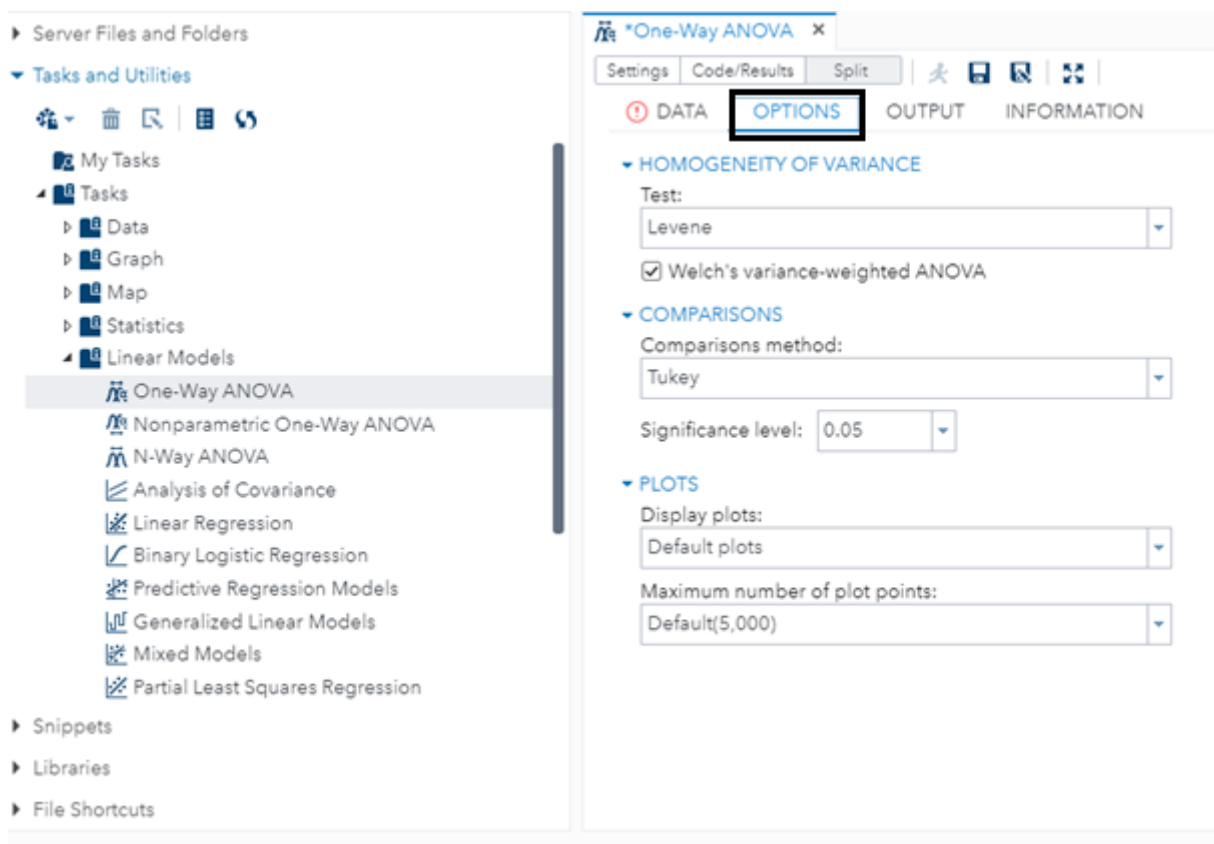
Na aba de opções, **OPTIONS**, na seção **Homogeneity of Variance**, é preciso selecionar o teste que será feito para verificar a homogeneidade das variâncias, conforme Figura 26, sendo eles:

- **None:** especifica que nenhum teste será feito.
- **Barlett:** deve ser usado se houver mais de cinco observações em cada grupo e esses grupos possuírem distribuição normal.
- **Levene:** é o teste padrão para testar a homogeneidade da variância, e é baseado nos quadrados dos resíduos.

- **Brown & Forshythe:** é uma variação do teste de Levene em que variâncias iguais são determinadas pelo desvio absoluto médio.
- **O'Brian:** uma modificação do teste de Levene que é baseado no cálculo do quadrado dos resíduos.

Caso os grupos tenham variâncias diferentes, usa-se o teste **Welch's variance-weighted ANOVA**.

Figura 26 - One-Way ANOVA – Options



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Na seção **COMPARISONS**, deve-se selecionar o método de comparação. Os testes mais usados são:

- **Tukey:** compara os possíveis pares de médias e se baseia na diferença mínima significativa. É poderoso quando aplicado em comparações entre todos os pares (MONTGOMERY, 2013).

- **Fisher**: usado na comparação de pares de vários grupos de tratamento quando o número de observações é inferior a 20. É conhecido por preservar o erro Tipo I se o número de grupos de tratamento for 3 (MEIER, 2006).
- **Bonferroni**: é um dos métodos utilizados para contornar o problema de múltiplas comparações (DUNNET, 1955). Após rejeitar a hipótese nula, o teste permite buscar saber quais médias apresentam diferenças significativas.
- **Scheffe**: usado quando há um grande número de contrastes entre médias (DUARTE, 2008).

### 5.7.2 Exemplos e Resultados

Utilizando o banco de dados **CARS** contido no SAS, contendo informações sobre a mecânica de diversas marcas de carros e com o objetivo de testar as diferenças nas médias de MGP na rodovia (número de milhas que um veículo percorre usando um galão de combustível) por tipo de carro, selecione em **Dependent variable** a variável **MGP\_Highway** e em **Categorical variable**, a variável categórica **Type**.

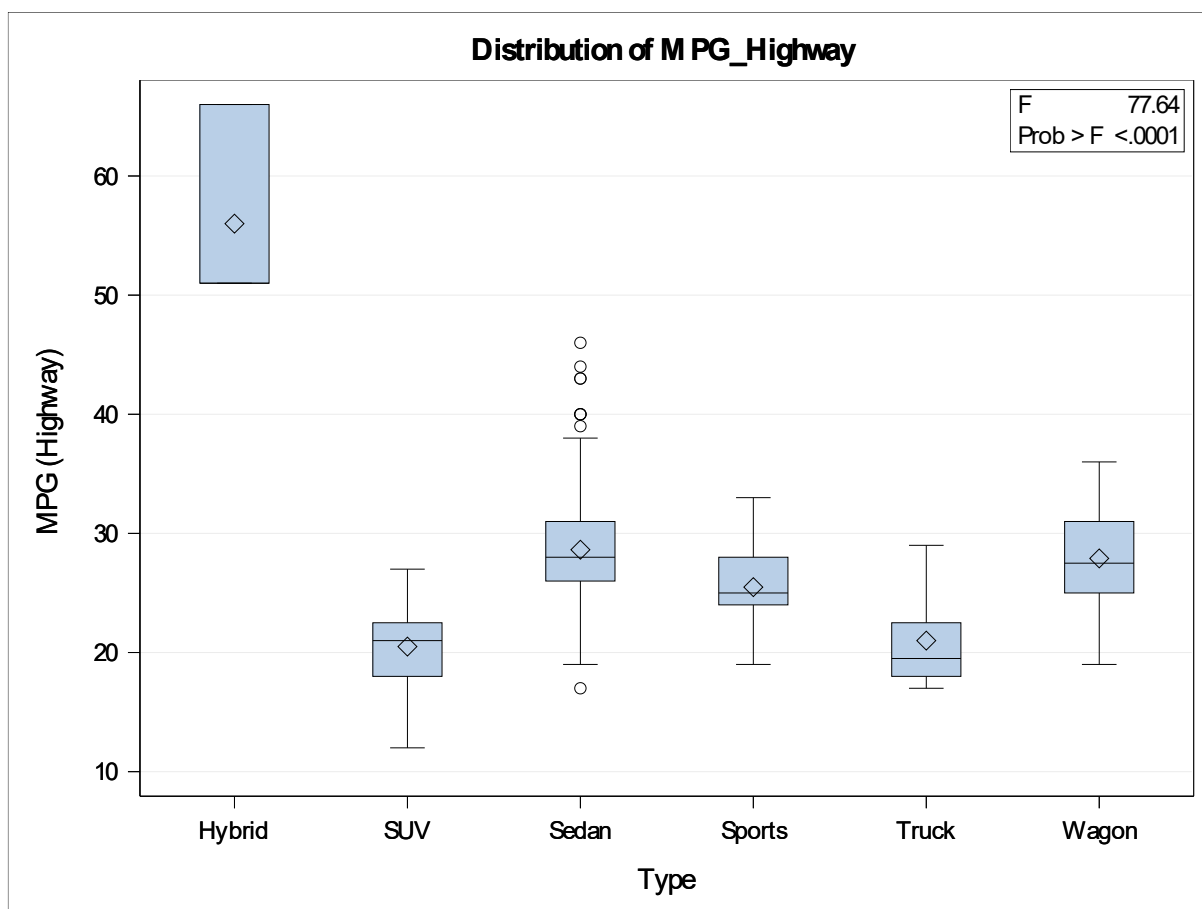
O teste de homogeneidade da variância será o teste de **Levene**, e o método de comparações, o teste **Tukey**. O software irá retornar a tabela da ANOVA e um gráfico da distribuição com boxplot do MGP, conforme Tabela 5 e Figura 27.

Tabela 5 - Análise de Variância (ANOVA)

Source	DF	Sum of Squares	Mean	F Value	Pr > F
			Suqare		
Model	5	6743.47	1348.6958	77.64	< .0001
Error	422	7331.03	17.37		
Corrected Total	427	14074.51			

Fonte: Elaborado pela autora

Figura 27 - Box-Plot – Distribuição de MPG (Highway) por Tipo de carro



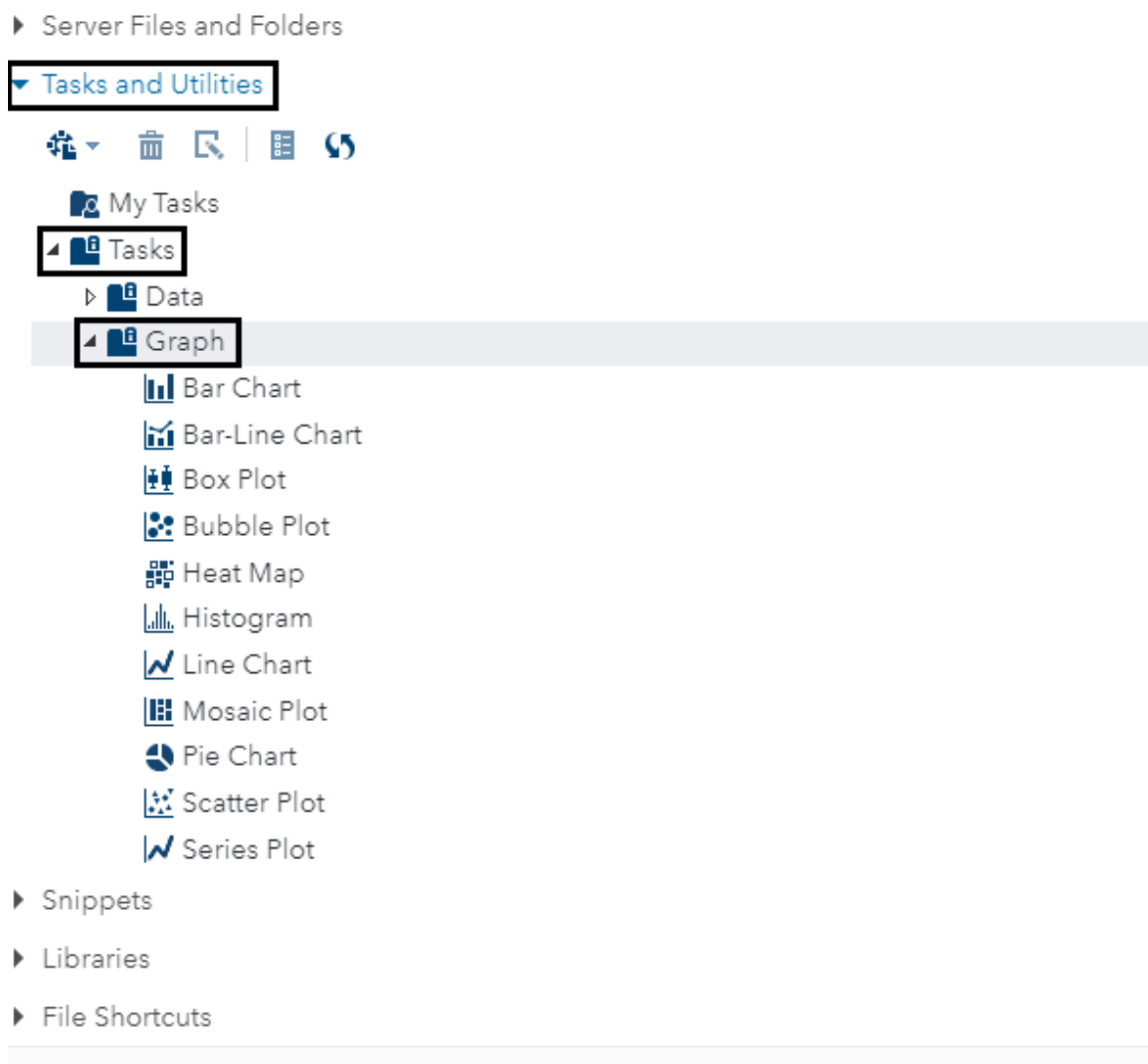
Fonte: Elaborado pela autora

## 5.8 Gráficos

Gráficos, na estatística, são ferramentas de apresentação dos dados, cujo objetivo é produzir em quem está visualizando, uma impressão rápida e sintetizada sobre o fenômeno estudado, facilitando assim sua compreensão. O SAS disponibiliza uma variedade de gráficos a serem trabalhos, e para acessá-los, siga: **Tasks and Utilities > Tasks > Graph**, conforme Figura 28.

São 12 tipos de gráficos à disposição, cada um deles possui uma interface diferente onde devem ser feitas seleções de acordo com seus respectivos objetivos e exigências. Entretanto, todos eles possuem a aba **APPREANCE**, cujo propósito é permitir ao usuário personalizar o gráfico com alteração das cores, referência, título, nota de rodapé e tamanho.

Figura 28 - SAS - Graph

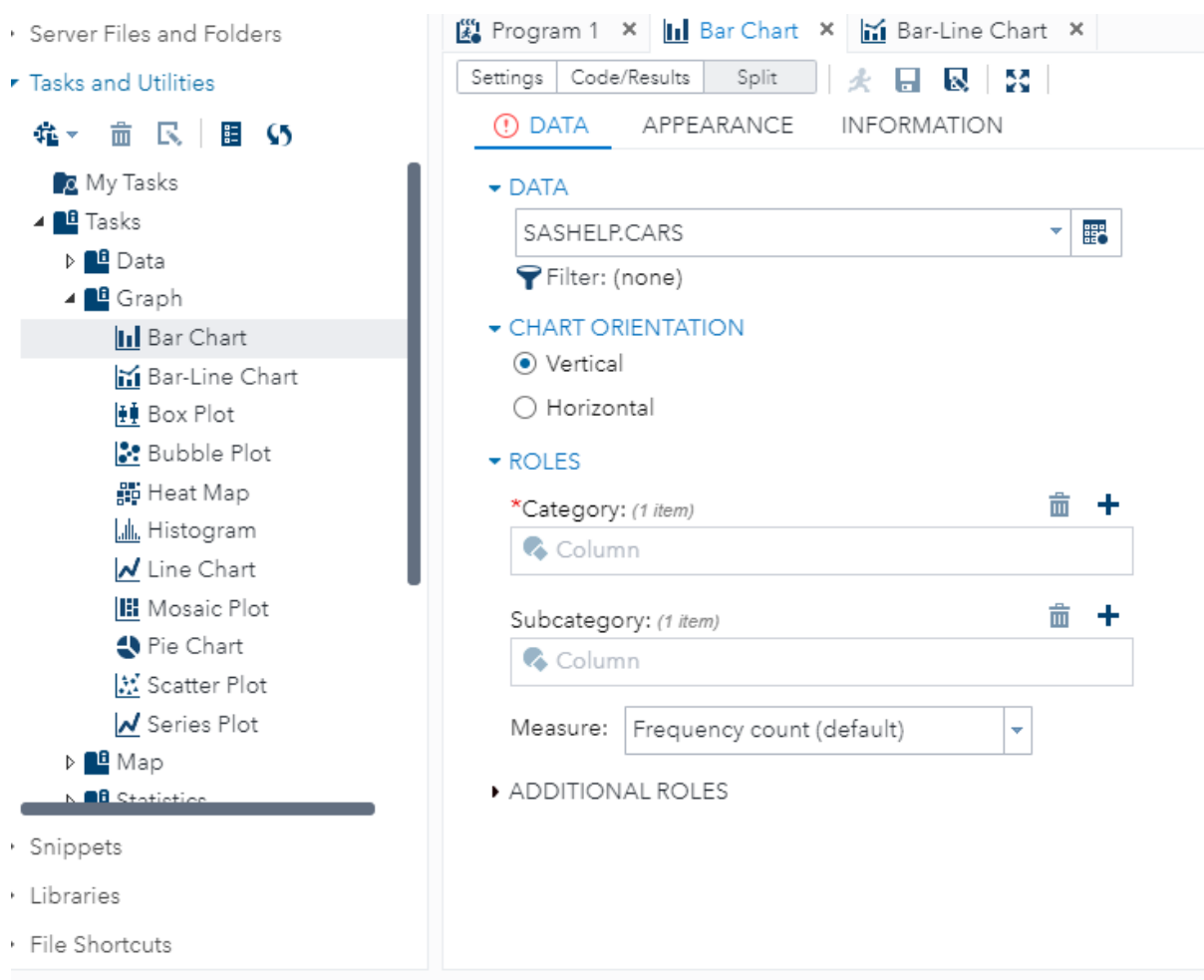


Fonte: Imagem retirada do site <http://sas.com/> (2019)

### 5.8.1 Bar Chart

O gráfico de barras é utilizado para medir a frequência de variáveis categóricas. Ele dispõe as barras retangulares de maneira que esses retângulos correspondam ao número de observações em cada categoria.

Figura 29 - Bar Chart - Data



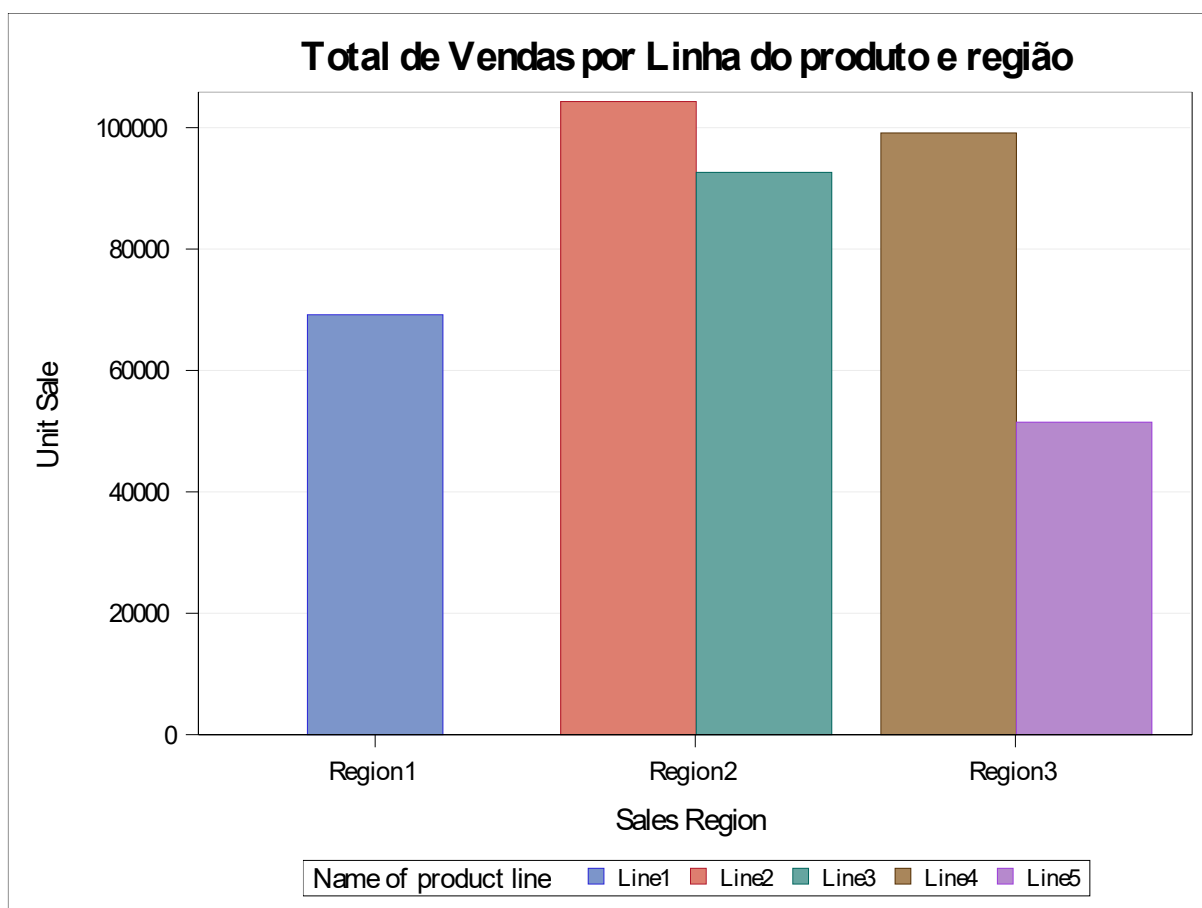
Fonte: Imagem retirada do site <http://sas.com/> (2019)

Ao abrir a interface desse gráfico, deve-se selecionar o banco de dados que deseja estudar e depois, fazer algumas especificações e atribuir variáveis para determinados papéis. Em **Chart Orientation** deve-se escolher se as barras do gráfico serão verticais ou horizontais. A seção **Category** é obrigatória e onde é especificada a variável categórica que classifica as observações. Em **Subcategory**, a variável usada para agrupar os dados deve ser indicada, conforme Figura 29. Também é possível designar a forma de calcular a altura da barra, se pela contagem da frequência, frequência percentual ou através de uma variável numérica.

Um exemplo desse gráfico pode ser visualizado no Figura 30, em que foi utilizado o banco **PRICEDATA** contido no SAS, associando as variáveis **regionName**,

**productLine** e **sale** às seções **Category**, **Subcategory** e **Measure > Variable**, respectivamente.

Figura 30 - Total de Vendas por Linha de produto e Região



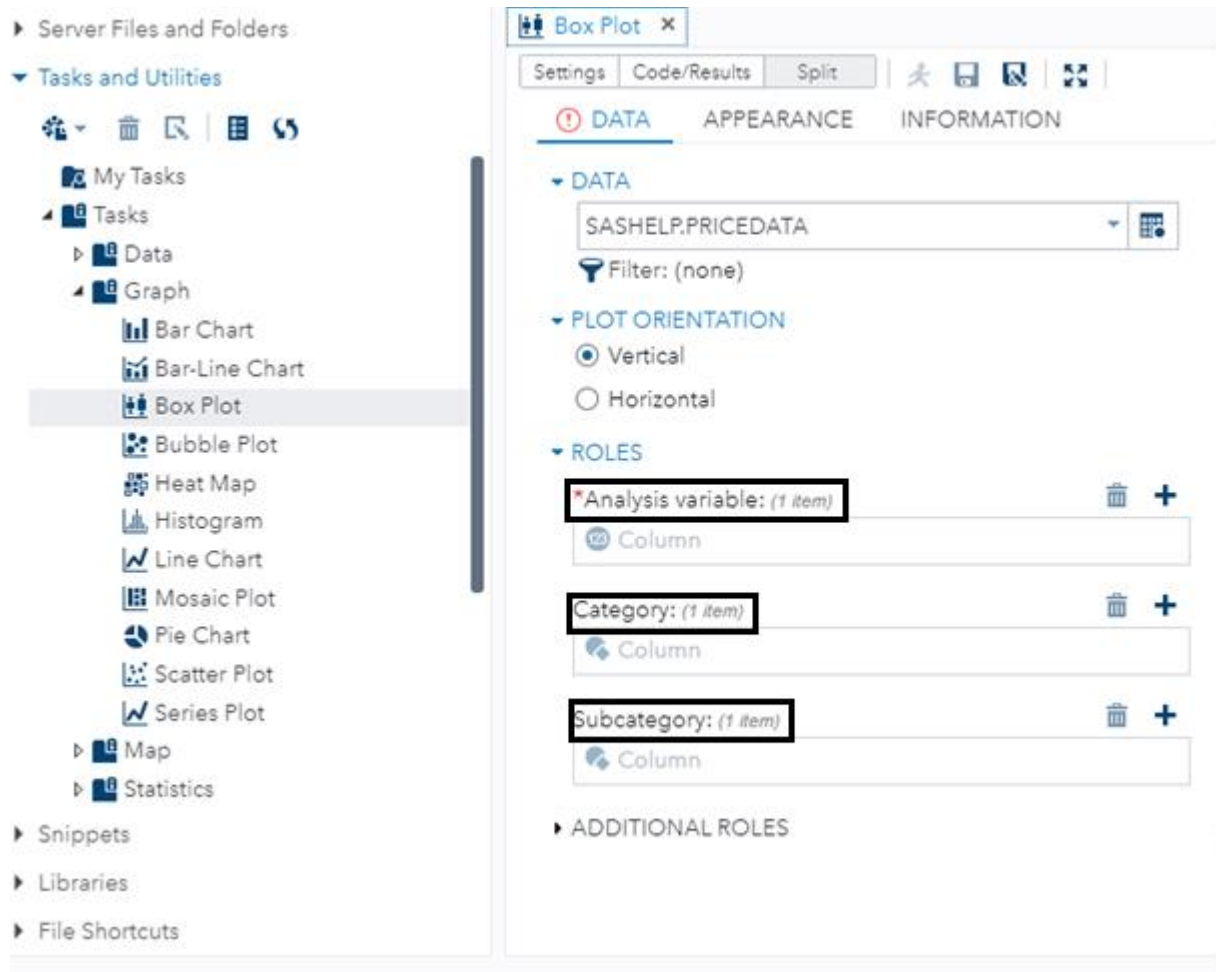
Fonte: Elaborado pela autora

### 5.8.2 Box Plot

O boxplot, ou gráfico de caixa, é formado pelo primeiro e terceiro quartil e pela mediana. É utilizado com o objetivo de analisar a variabilidade dos dados numéricos e comparar visualmente dois ou mais grupos.

Para obter esse gráfico no SAS, é preciso designar a variável de análise no campo **Analysis variable**, para criar um boxplot para cada valor distinto, deve-se selecionar uma variável categórica em **Category**. Se quiser adicionar outra variável categórica para agrupar os boxplots, deve-se adicioná-la em **Subcategory**, conforme Figura 31.

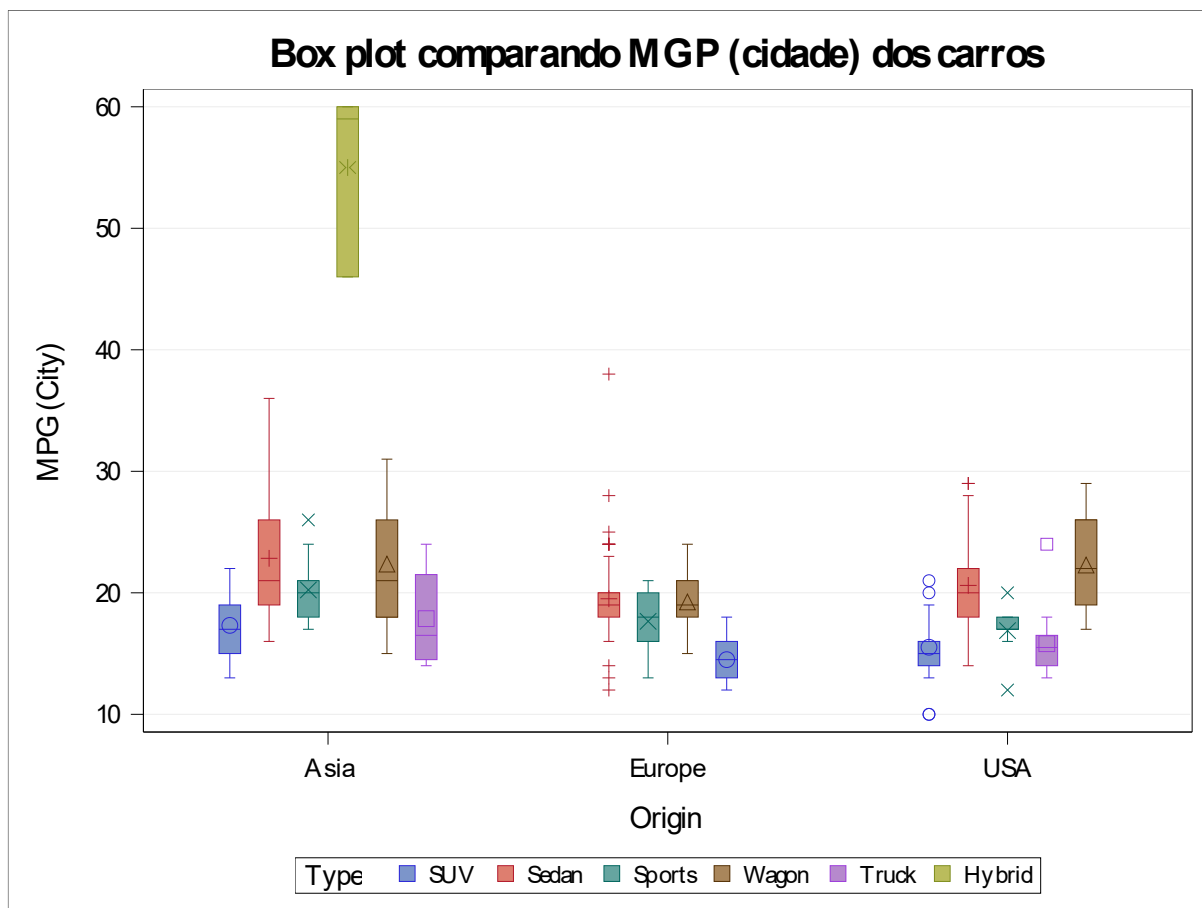
Figura 31 - Graph – Boxplot



Fonte: Imagem retirada do site <http://sas.com/> (2019)

O exemplo mostrado na Figura 32 compara, através do banco de dados **CARS**, quantas milhas por galão (MPG) um carro consome dependendo da área de origem. Dessa forma, aloca-se **MPG\_City** em **Analysis variable**, **Origin** em **Category** e **Type** em **Subcategory**.

Figura 32 - Box plot comparando MGP (cidade) dos carros



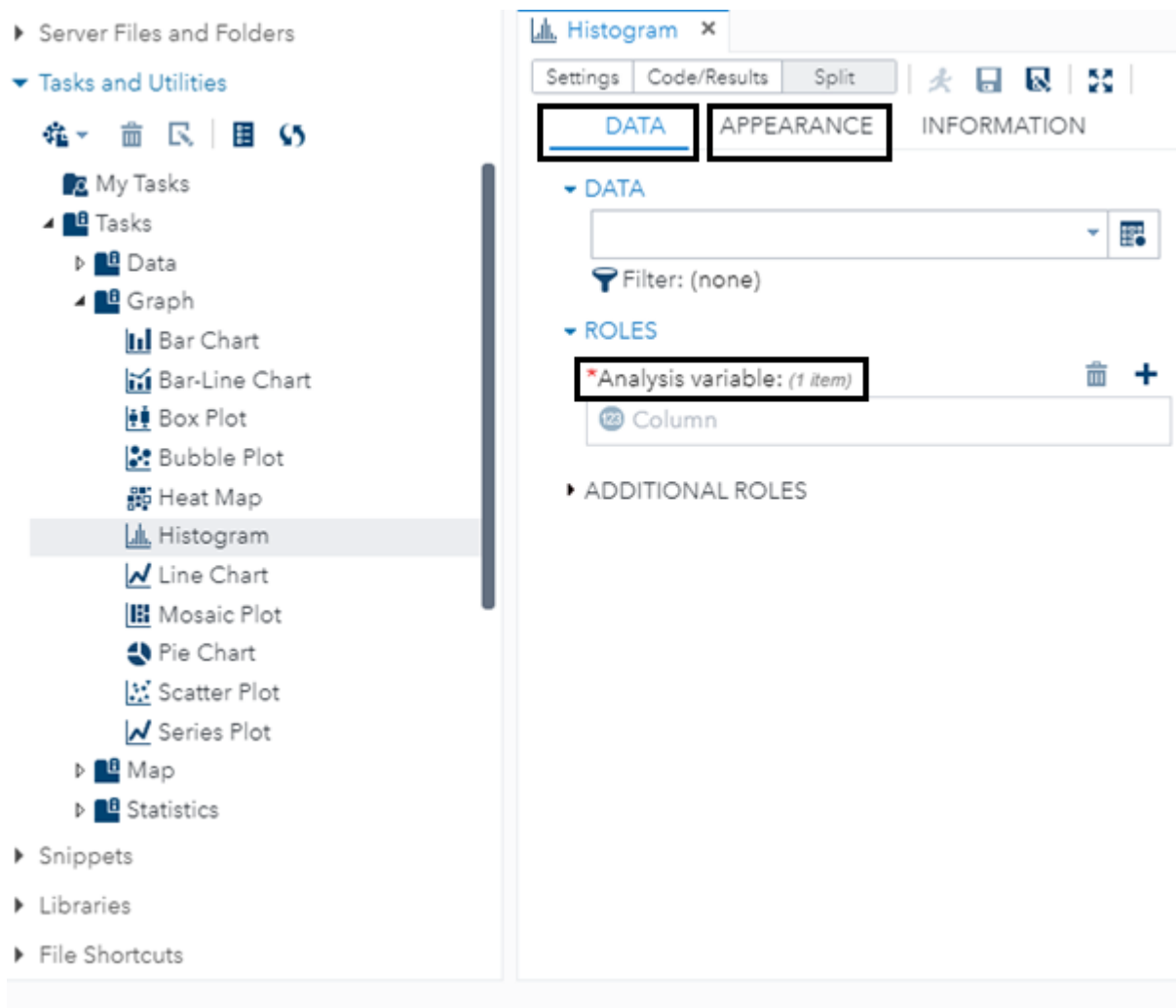
Fonte: Elaborado pela autora

### 5.8.3 Histogram

O histograma é um gráfico conhecido por mostrar a distribuição dos dados e, portanto, indicar a variabilidade do processo. Visualmente é composto por barras retangulares verticais sem espaço entre elas, já que são usadas variáveis quantitativas contínuas em sua composição.

No software, é preciso somente especificar a variável de análise em **Analysis variable** e a escala aplicada no eixo vertical em **Scale**, podendo ser: em porcentagem, contagem ou proporção, conforme Figura 33. Ademais, na aba **APPEARANCE**, é possível adicionar curvas de densidade da distribuição Normal e de Kernel.

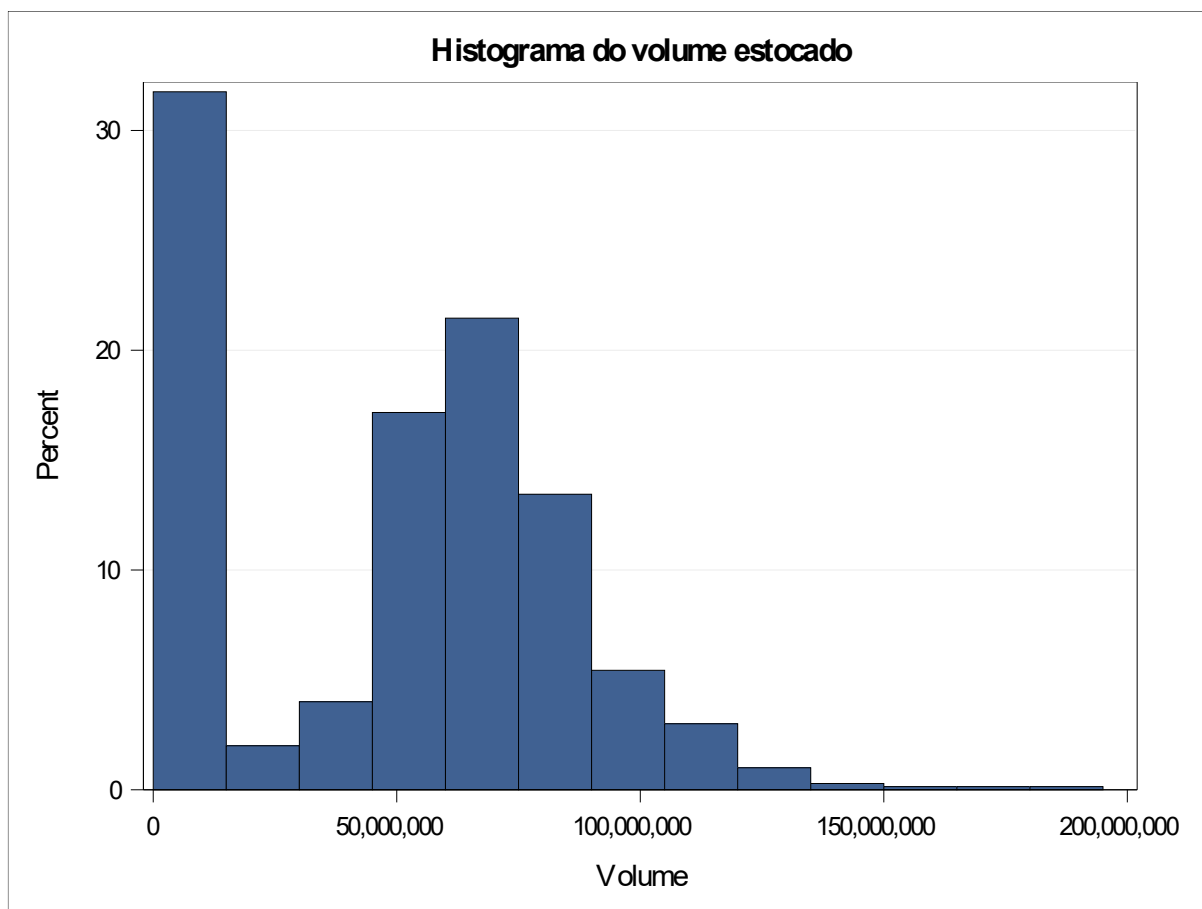
Figura 33 - Graph – Histogram



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Na Figura 34 está, como exemplo, um histograma do volume estocado, onde foi utilizado o banco de dados **STOCKS** e alocado em **Analysis variable** a variável quantitativa contínua **Volume**.

Figura 34 - Histograma do volume estocado



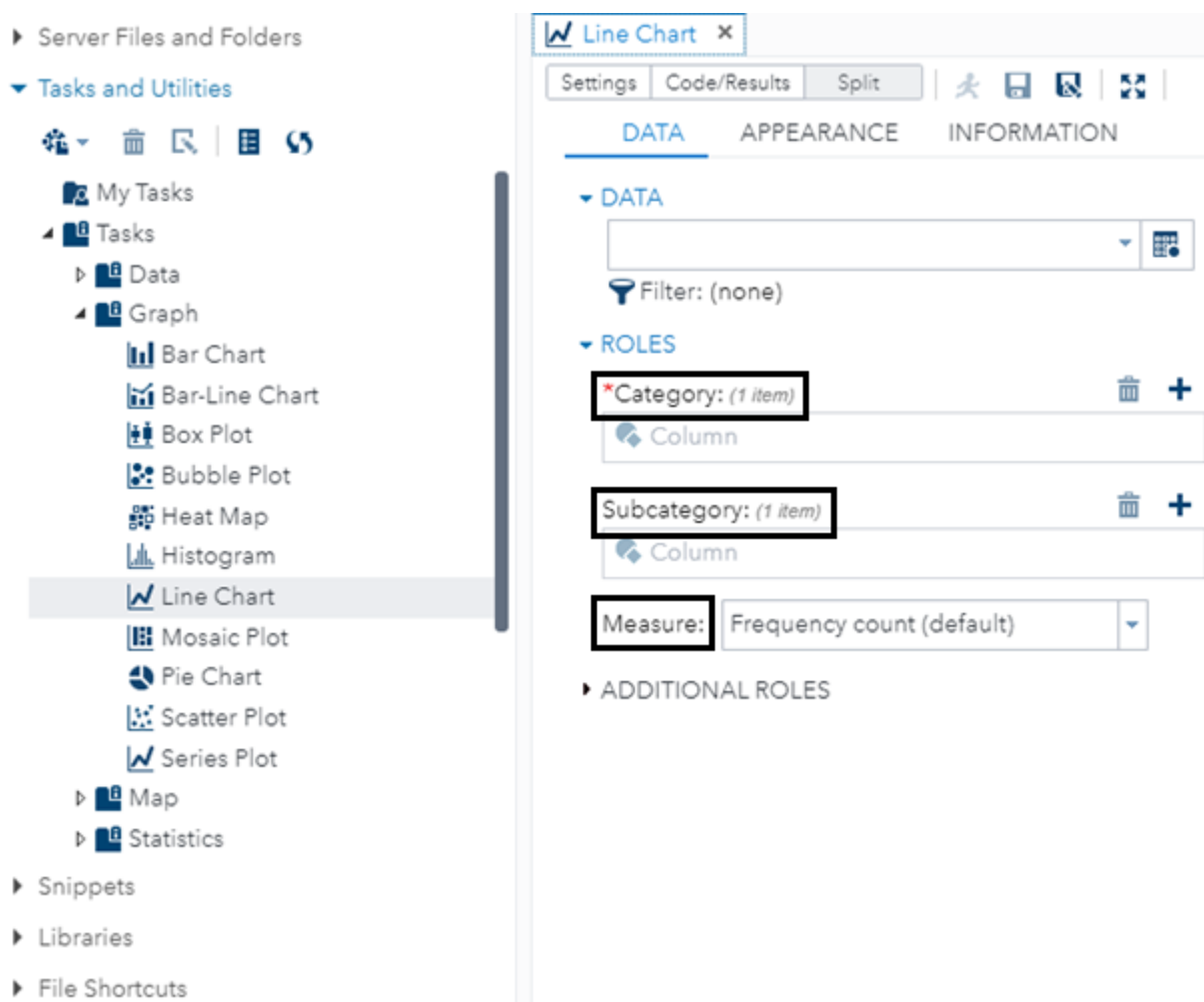
Fonte: Elaborado pela autora

#### 5.8.4 Line Chart

O gráfico de linhas é composto por observações ligadas por segmentos de linha. Tem como objetivo visualizar tendências nos dados em intervalos de tempo e por isso as variáveis devem ser quantitativas discretas.

Essa função no SAS exige a especificação de uma variável para classificar as observações em diferentes categorias na seção **Category**. Em **Subcategory** deve ser selecionar outra variável que classifica os dados em categorias ainda menores, podendo ser identificadas através de legenda. Também é possível escolher como calcular os pontos do gráfico em **Measure**, conforme Figura 35.

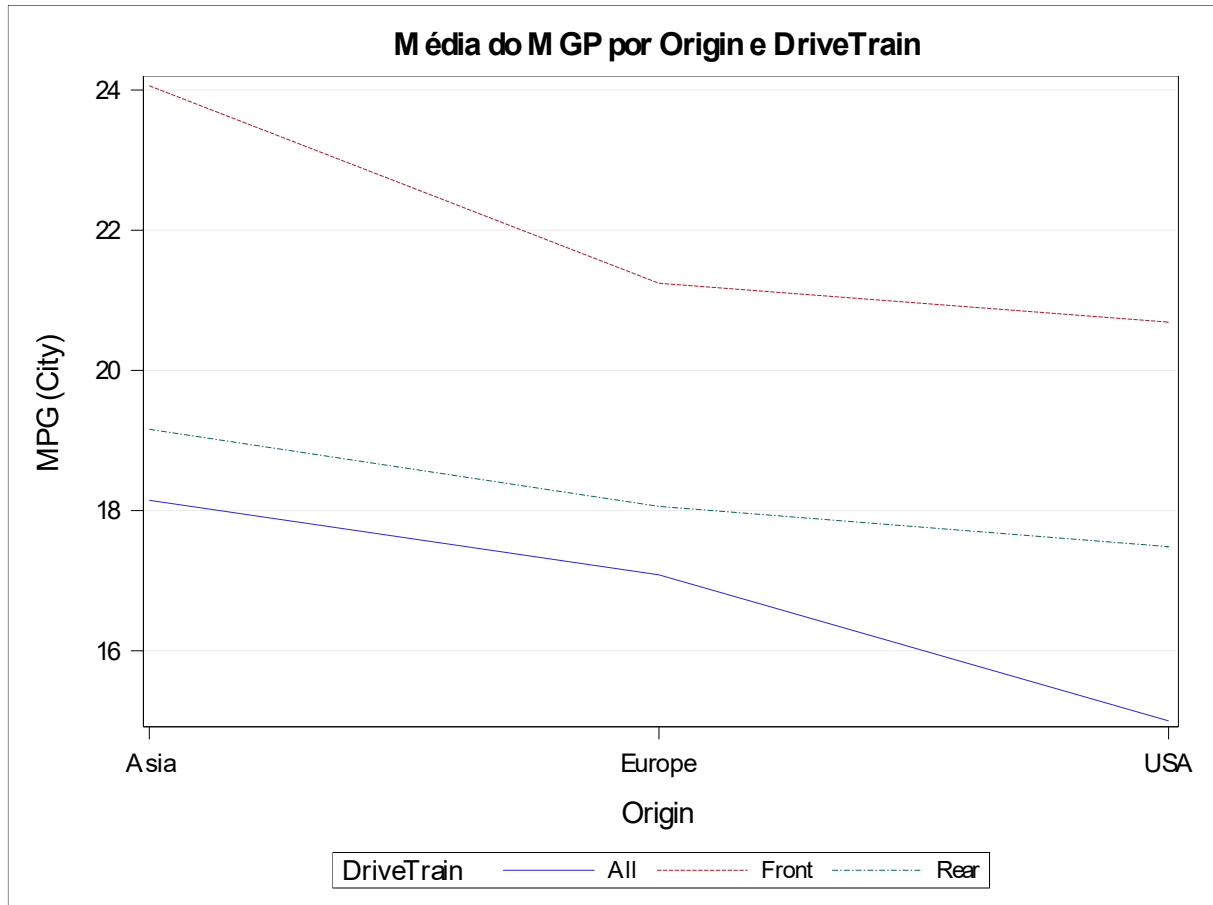
Figura 35 - Graph – Line Chart



Fonte: Imagem retirada do site <http://sas.com/> (2019)

Utilizando o banco de dados **CARS** e alocando as variáveis **Origin**, **DriveTrain**, **MGP\_City** em **Category**, **Subcategory** e **Measure > Variable**, respectivamente, tem-se o gráfico de linhas da média do MGP por origem e drive train, conforme Figura 36.

Figura 36 - Média do MGP por origem e drive train



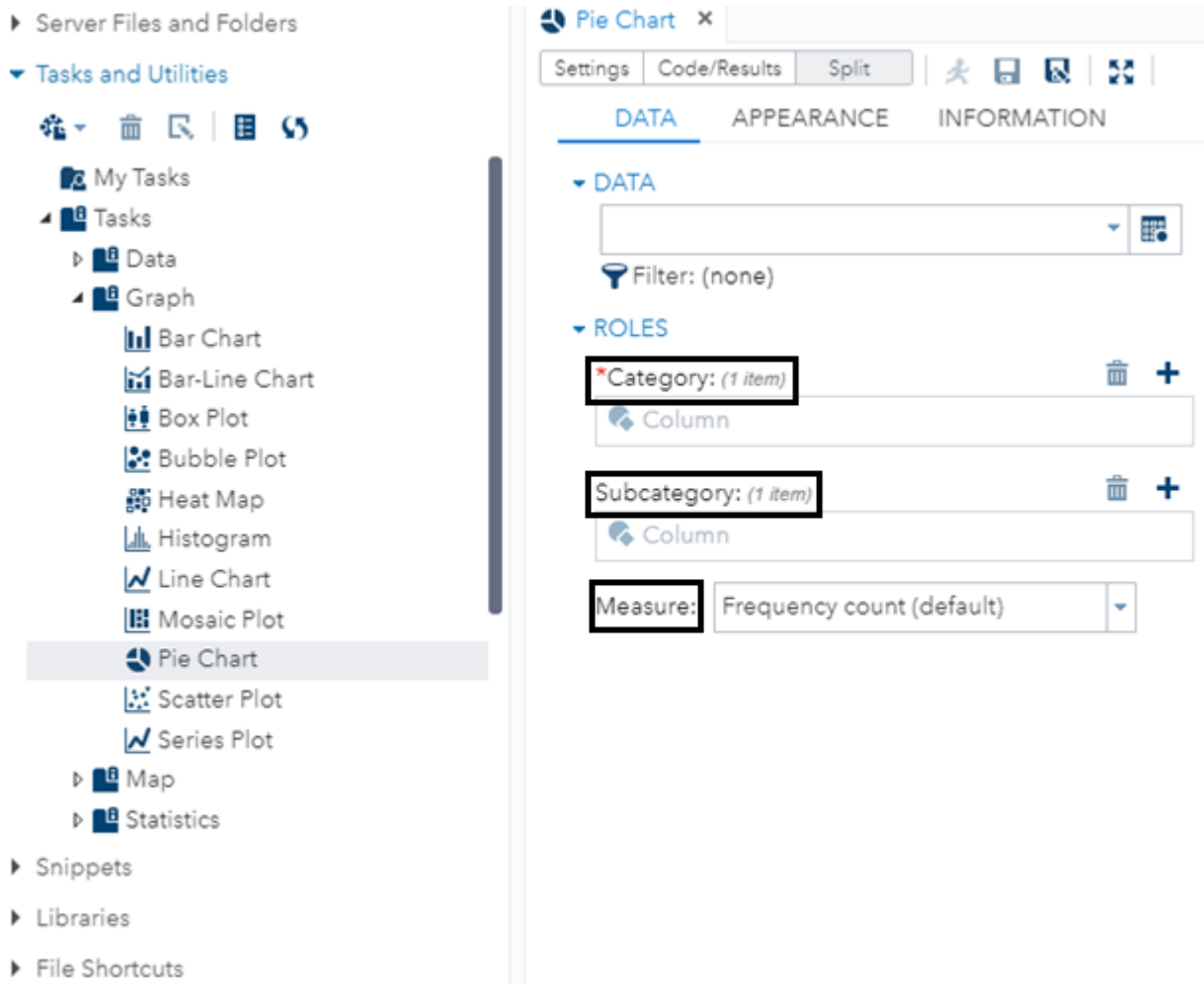
Fonte: Elaborado pela autora

### 5.8.5 Pie Chart

O gráfico de setores, mais conhecido por gráfico de pizza é um diagrama circular em que os valores de cada categoria são proporcionais as suas frequências. O tamanho da fatia representa a contribuição dessa observação para o total do gráfico.

Para obter esse gráfico no SAS, é preciso designar para a seção **Category**, uma variável categórica para classificar as observações em diferentes subconjuntos. Em **Subcategory**, especifica-se uma variável que será usada para agrupar as observações dentro de cada categoria. Para determinar como calcular o tamanho da fatia, é preciso escolher entre: usando a frequência, a porcentagem ou através de outra variável em **Measure**, conforme Figura 37.

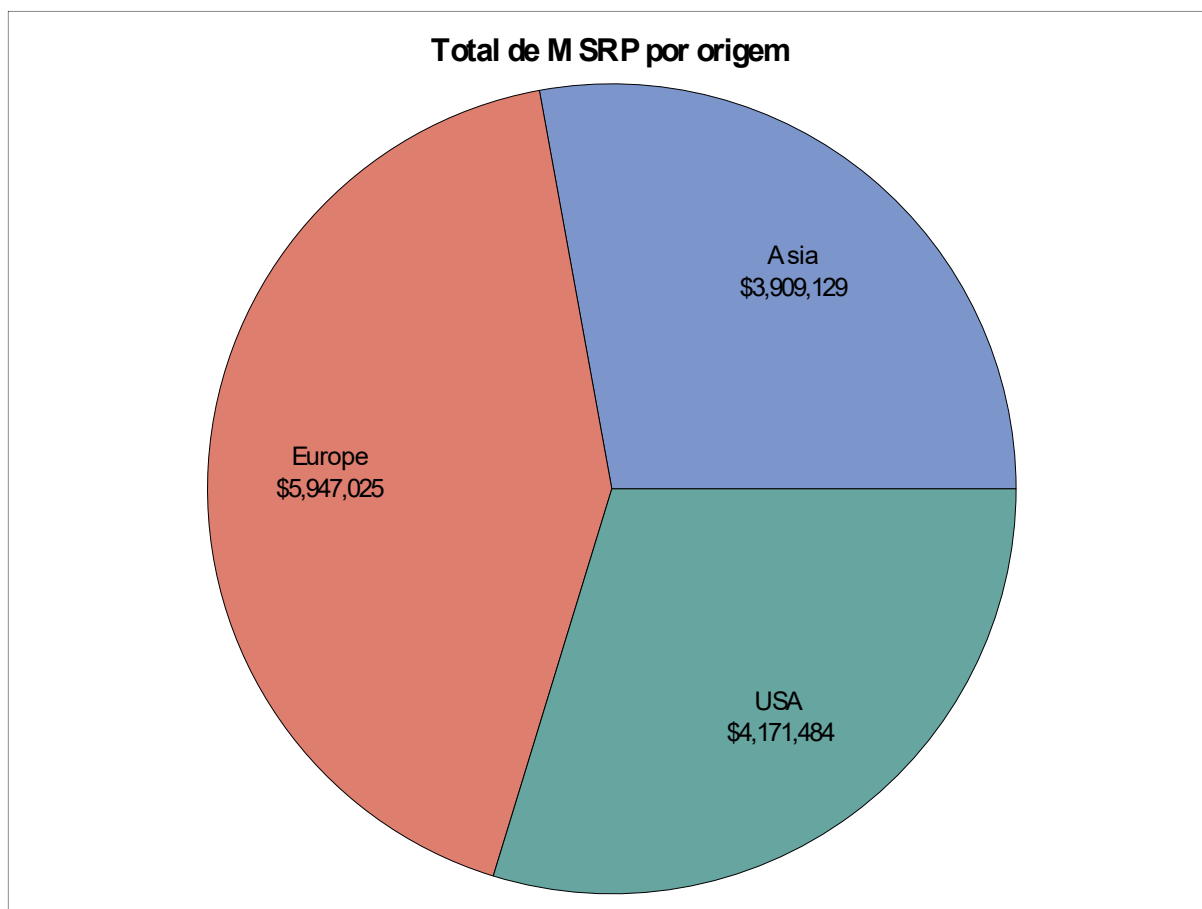
Figura 37 - Graph – Pie Chart



Fonte: Imagem retirada do site <http://sas.com/> (2019)

A Figura 38 exibe um exemplo, utilizando o banco de dados **CARS**, de um gráfico de pizza que mostra o total do preço de varejo do fabricante (MSRP) por região, onde as variáveis **Origin** e **MSRP** foram alocadas nas funções **Category** e **Measure > Variable**.

Figura 38 - Total de MSRP por Região



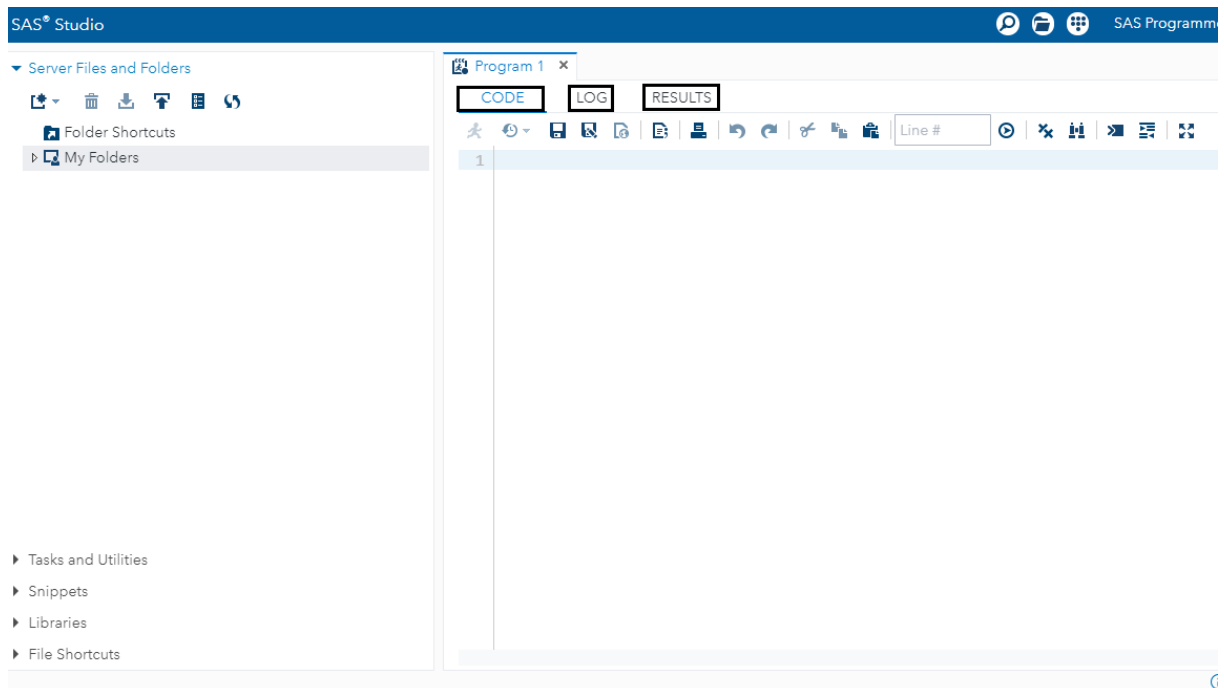
Fonte: Elaborado pela autora

## 5.9 Procedures

Os procedimentos do SAS, chamados de **PROC STEP**, são rotinas de programação que realizam cálculos estatísticos de acordo com a metodologia do software. No **PROC STEP** serão feitas análises dos dados gerados no **DATA STEP**, responsável pela criação de arquivos.

A primeira interface mostrada ao iniciar o SAS Studio conterà as abas **CODE**, **LOG** e **RESULTS**. **CODE** é o espaço reservado para a programação, em **LOG** são exibidos os possíveis erros e sinais de alerta do código criado. Por fim, em **RESULTS** são mostrados os resultados após a sua execução, conforme Figura 39.

Figura 39 - SAS Studio - Programação



Fonte: Imagem retirada do site <http://sas.com/> (2019)

### 5.9.1 Data Step

```

data nomedoarquivo;
    input variável1 variável2 $ variável3;
    datalines;
a b c
d e f
g h i
;
*comentários podem ser inseridos seguidos de um asterisco
PROC PRINT;
  
```

A sintaxe básica para a criação de um arquivo é dada através de instruções e declarações para criá-los fornecendo dados no próprio programa, onde:

- **data**: informa a criação de arquivo nomeado “nomedoarquivo”;

- **input:** estabelece a entrada de dados em três colunas de variáveis. As variáveis categóricas pedem a apresentação do caractere \$ após designá-la;
- **datalines:** dá entrada em quantas linhas forem necessárias de dados, de modo que em cada linha esteja presente a primeira observação de todas as variáveis declaradas, separadas somente por um espaço;
- **PROC PRINT:** indica a impressão do arquivo;
- **run:** indica a execução do programa.

### 5.9.2 Proc Means

Posterior à criação ou importação de um arquivo é possível, através do procedimento MEANS calcular estatísticas simples para as variáveis numéricas inseridas. Algumas das estatísticas fornecidas pelo procedimento são: média (mean), número de observações (n), coeficiente de variação (cv), desvio padrão (std), variância (var), mediana (median).

```
proc means data=nomedoarquivo n mean max min median std var;  
var variavel1 variavel2;  
class variavel3;  
title 'Summary'  
run;
```

- **proc means:** comando para ativar a estatística descritiva. É seguido pela indicação do banco de dados a ser analisado e pela descrição das estatísticas que deseja obter;
- **var:** indica as variáveis numéricas as quais pertencem as observações em análise;
- **class:** especifica as variáveis categóricas que desejar, separando a análise em subgrupos.
- **title:** fornece o título da tabela a ser computada.

### 5.9.3 Proc Univariate

O procedimento UNIVARIATE calcula estatísticas descritivas mais detalhadas, provendo a elaboração, principalmente, de histogramas com curvas de distribuição e Q-Q plots.

```

title 'Histograma';
ods graphic on;
proc univariate data=nomedoarquivo;
    histogram variavel1 / odstitle = title;
run;

title 'Histograma com curva normal';
ods select Histogram ParameterEstimates GoodnessOfFit FitQuantiles Bins;
proc univariate data=nomedoarquivo;
    histogram variavel1 / normal
        odstitle=title;
*normal deve ser adicionado para computar uma curva normal sob o histograma
run;

title 'Normal Q-Q Plot';
ods graphics on;
proc univariate data=nomedoarquivo;
    qqplot variavel2 / odstitle = title;

```

- **ods graphic on:** procedimento para ativar gráficos ODS, próprios do SAS University Studio, que possui uma sintaxe diferente dos procedimentos no SAS tradicional;
- **proc univariate:** declaração que precede a utilização de estatísticas univariadas. É seguida da determinação do arquivo a ser utilizado;
- **histogram/qqplot:** dentro da declaração proc univariate, é especificado o gráfico que deseja computar e a variável utilizada para obtê-lo;

- **ods select:** restringe as estatísticas que devem ser computadas em tabelas diferentes;

#### 5.9.4 Proc Reg

```
proc reg data=nomedoarquivo;

    model vardependente = varindependente1 varindependente2
varindependente3;

    */selection=método;
```

O procedimento REG é responsável por fazer ajustes em modelos de regressão linear através do método dos mínimos quadrados.

A sintaxe básica para obter um modelo de regressão linear irá retornar a análise de variância, valores do coeficiente de determinação, tabela de estimação dos parâmetros, gráficos de diagnóstico da variável dependente, erro médio quadrático, análise dos resíduos, e é composta pelas seguintes declarações:

- **proc reg:** declara que está sendo realizado um procedimento Reg e deve ser sucedido do nome do arquivo que for analisar;
- **model:** especifica as variáveis dependentes e independentes;
- **selection:** indica um método para seleção das variáveis independentes mais significativas. Os métodos possíveis são: FORWARD, BACKWARD, STEPWISE, MAXR, MINR, RSQUARED, ADJRSQ, CP ou NONE.

#### 5.9.5 Proc ANOVA

Esse procedimento realiza a análise da variância de fator único. Normalmente é mais utilizado para experimentos com blocos aleatorizados ou quadrados latinos.

No primeiro, há a aleatorização da replicação de cada combinação de tratamento em cada bloco, no segundo, o número de tratamentos deve ser igual ao número de repetições, ou seja, o número de linhas é igual ao número de colunas.

```
proc anova data = nomedoarquivo;  
    class variavel1 variavel2 variavel3;  
model dependente = variavel1 variavel 2;  
run;
```

- **class**: indica as variáveis categóricas que serão usadas para indicar os grupos de análise;
- **model**: indica as variáveis dependentes e independentes do modelo.

## 6 METODOLOGIA

Foram utilizados os dados fornecidos pelo INPE-EM, serviço do Instituto Nacional de Pesquisas Espaciais (INPE) que se propõe a divulgar dados e estimativas anuais de emissões de gases do efeito estufa no Brasil por mudança de cobertura da terra, em conjunto com o Sistema PRODES, que por sua vez disponibiliza estimativas anuais das emissões para a Amazônia Brasileira. São disponibilizados dados de 1ª Ordem, ou seja, as emissões são captadas no momento da mudança de uso da terra, e de 2ª Ordem, dados que representam o processo gradativo de liberação e absorção dos GEE, gerando estimativas mais precisas e com maior embasamento científico, os quais serão utilizados na análise do presente trabalho.

Essas informações são obtidas através do site oficial do instituto (<http://inpe-em.ccst.inpe.br/>), em uma página voltada somente para downloads de planilhas Excel, contendo estimativas das emissões brutas anuais de 1980 a 2019. Os dados são divididos em “Vegetation Removal” (VR), referente à vegetação primária removida através de queimadas ou decomposição e “Secondary Vegetation” (SV), referente à vegetação secundária, ou seja, aquela que se recuperou por meio de processos naturais após supressão total ou parcial da vegetação primária, de acordo com o Art. 2o da Lei nº 11.428, de 22 de dezembro de 2006 (Brasil, 2006). As variáveis contidas na planilha associadas à vegetação primária removida são: área desmatada, emissão de gás carbônico total, através de queimadas e através de decomposição, emissão dos demais gases (metano e óxido nitroso) por meio de queimadas. Associadas à vegetação secundária, tem-se: área total, área desmatada, emissão de  $CO_2$ , absorção de  $CO_2$  e emissão líquida de  $CO_2$ .

Com o objetivo de criar um modelo de regressão linear, foram feitas análises da correlação entre as variáveis e criação de vários modelos teste para a verificação dos pressupostos. Por meio do software SAS University Edition, esses modelos foram criados através dos métodos de seleção Forward, Backward e Stepwise, e em seguida, a análise dos  $R^2$  ajustados e aplicação dos critérios AIC e BIC para decidir os melhores modelos.

Foram feitas então alterações no banco de dados em relação a algumas variáveis redundantes e com alta correlação entre si:

- Área desmatada total: consiste na soma da área desmatada da vegetação primária e secundária;
- Gás carbônico total: inclui a emissão de  $CO_2$  por queima e decomposição da vegetação primária e secundária.
- Gases de efeito estufa: foi feita a soma da emissão do gás metano e óxido nítrico, pois os valores eram baixos quando comparados à emissão do gás carbônico e altamente correlacionadas.

Grande parte das pesquisas delineadas para examinar o efeito exercido por duas ou mais variáveis independentes sobre uma variável dependente utiliza a análise de regressão múltipla (ABBAD e TORRES, 2002). Foram selecionados dois modelos lineares múltiplos para explicar a emissão de gás carbônico total e a emissão dos demais gases de efeito estufa (GEE), ambas através das variáveis “Área total” desmatada e “Gás carbônico absorvido” referente à vegetação secundária.

## 7 RESULTADOS E DISCUSSÕES

### 7.1 Análise temporal do desmatamento e emissão de gases de efeito estufa na Amazônia brasileira

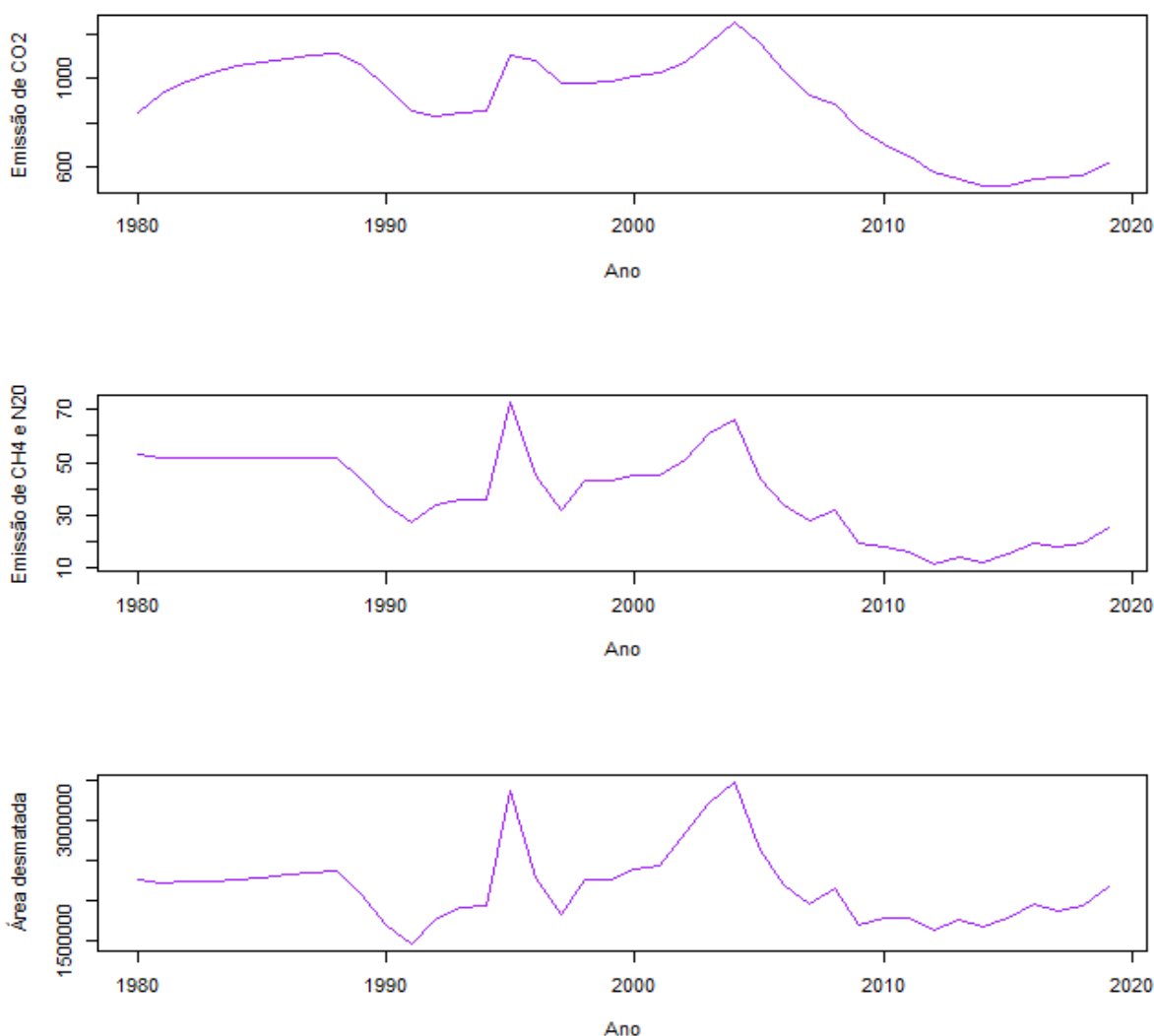
Conforme Figura 40, desde os anos 1980 é possível perceber uma tendência crescente nas emissões de gás carbônico na Amazônia até o final da mesma década, quando então passou a haver “picos e vales” irregulares até 2004, o ano do seu pico, e a partir dali apresentou um comportamento decrescente até 2019. Os gráficos das emissões de  $CH_4$  e  $N_2O$  e o da área desmatada têm comportamentos semelhantes, contudo apresentaram nos primeiros 8 anos de observações, uma tendência constante até o final da década de 80, ocorrendo desde então uma sucessão irregular de flutuações. As emissões de  $CH_4$  e  $N_2O$  tiveram seu pico no ano de 1996 e a área desmatada no ano de 2004, entretanto, assim como as emissões de gás carbônico, elas apresentaram um comportamento decrescente de 2004 até meados da década de 2010.

Cerca de 60% da quantidade de  $CO_2$  emitido na atmosfera, desde 1980, provém do desmatamento (FEARNSIDE, 2002). Em 2004 foram postas em prática pelo governo federal ações para conter o alto desmatamento e conseqüente emissão de gases de efeito estufa, entre elas o sistema de monitoramento por satélite, a criação de áreas protegidas e a identificação de municípios críticos para o desmatamento (BRANDÃO JR. et al., 2015). De acordo com os dados obtidos, isso reduziu as emissões de gás carbônico, metano e óxido nitroso e a área desmatada em 68%, 82%, 80% e 41%, respectivamente, considerando o período entre 2004 e 2014, o que provavelmente justifica os picos e quedas analisadas anteriormente.

A partir de 2013, entretanto, o desmatamento e as emissões dos gases poluentes voltaram a crescer, quando os efeitos do Novo Código Florestal, aprovado em 2012 e que dava anistia aos desmatadores, começaram a surgir, revertendo a tendência de queda dos anos anteriores (MARIA, 2013). Tal efeito vem se agravando por conseqüência da grave crise política que o Brasil vem enfrentando, paralelamente ao enfraquecimento do Ministério do Meio Ambiente, principalmente no ano de 2019, em que diversas medidas prejudiciais foram tomadas, como o fim das reservas legais

e a contenção de multas por exemplo (TRIGUEIRO, 2019). Logo, houve um aumento de 12,3%, 8,9% e 31,7% no desmatamento, emissão de gás carbônico e emissões de metano e óxido nitroso, incremento não visto desde 2004.

Figura 40 - Gráficos das séries temporais



Fonte: Elaborado pela autora

## 7.2 Análise dos Modelos de Regressão Múltipla

A Tabela 6 resume as informações e estatísticas encontradas para os modelos com a variável dependente “Emissão de CO2”. Enquanto a Tabela 7 apresenta as mesmas análises, porém para os modelos com a variável dependente “Emissão de GEE”.

Tabela 6 - Informações e estatísticas para os modelos de regressão com a variável dependente “Emissão de CO<sub>2</sub>”

Método	R <sup>2</sup>	R <sup>2</sup> <sub>ajustado</sub>	Variáveis	VIF máx	AIC	F	P-valor
Forward	100%	100%	área, gee, co2 absorvido, co2 líquido	132,74	72,98	1415	<.0001
Stepwise, Backward	100%	100%	co2 absorvido, co2 líquido	2,19	70,68	2868	<.0001
Manual	73,64%	72,21%	área, co2 absorvido	1,04	495,41	51,67	<.0001

Fonte: Elaborado pela autora

Tabela 7 - Informações e estatísticas para os modelos de regressão com a variável dependente “Emissão de GEE”

Método	R <sup>2</sup>	R <sup>2</sup> <sub>ajustado</sub>	Variáveis	VIF máx	AIC	F	P-valor
Forward	99,25%	99,17%	área, co2, co2 absorvido, co2 líquido	9411,1	151,1	1164	<.0001
Stepwise, Backward	82,12%	81,15%	co2 absorvido, co2 líquido	2,19	274,15	84,96	<.0001
Manual	98,84%	98,77%	área, co2 absorvido	1,04	164,91	1570	<.0001

Fonte: Elaborado pela autora

Os modelos foram escolhidos baseados nos algoritmos de seleção descritos no item 4.2.2 deste trabalho. Avaliando-se as funções obtidas no que diz respeito aos coeficientes de determinação (R<sup>2</sup>) e às estatísticas F, pode-se afirmar que todos são capazes de explicar de maneira satisfatória as variáveis respostas.

Pelos métodos adotados percebe-se que a seleção manual, stepwise e backward não apresentam evidências de multicolinearidade entre as variáveis (VIF<10). Levando em conta que o principal objetivo deste trabalho é explicar as emissões de gases de efeito estufa através da área desmatada, a função determinada pelos métodos stepwise e backward não inclui esta variável, apesar de possuir o menor AIC para a variável dependente “Emissão de CO<sub>2</sub>”, o que torna o modelo selecionado manualmente a melhor opção em ambos os casos.

Portanto, a melhor função encontrada para explicar a emissão de gás carbônico apresentando um erro padrão residual de 111,4, está indicada na Equação 2 e na Tabela 8.

$$\text{Emissão de } CO_2 = 559,4578 + 0,00028*(\text{Área desmatada}) + 2,0794*(CO_2 \text{ absorvido}) \quad (2)$$

Tabela 8 - Resultado da regressão múltipla para o conjunto de variáveis ( $CO_2$ )

Variável	Graus de Liberdade	Coefficiente	Erro padrão	Valor t	Pr >  t
Intercepto	1	559,4578	108,3598	5,16	<.0001
Área desmatada	1	0,0002	0,00004	7,06	<.0001
$CO_2$ absorvido	1	2,0794	0,35582	5,84	<.0001

Fonte: Elaborado pela autora

Mantendo os mesmos regressores, a melhor função encontrada para explicar a emissão dos demais gases de efeito estufa com um erro padrão residual de 1,788, está apresentada na Equação 3 e na Tabela 9.

$$GEE = 5,38945 + 0,00002569*(\text{Área desmatada}) + 0,17709*(CO_2 \text{ absorvido}) \quad (3)$$

Tabela 9 - Resultado da regressão múltipla para o conjunto de variáveis (GEE)

Variável	Graus de Liberdade	Coefficiente	Erro padrão	Valor t	Pr >  t
Intercepto	1	5,38945	1,74038	3,1	<.0001
Área desmatada	1	0,000025	0,0000064	39,95	<.0001
$CO_2$ absorvido	1	0,17709	0,00571	30,99	<.0001

Fonte: Elaborado pela autora

Os regressores contidos na Tabela 8 e na Tabela 9 são significativos e, portanto, participam do modelo. Conforme a Tabela 10, no que tange à variável resposta  $CO_2$ , o coeficiente de correlação múltipla indica uma alta correlação entre os valores preditos e os observados. Segundo o coeficiente de determinação, este modelo explica 73,64% dos valores observados. Para a variável resposta  $GEE$ , o coeficiente de correlação múltipla indica uma correlação ainda maior e o coeficiente

de determinação aponta que, neste segundo modelo, 98,84% das variáveis são explicadas, conforme valores também dispostos na Tabela 10.

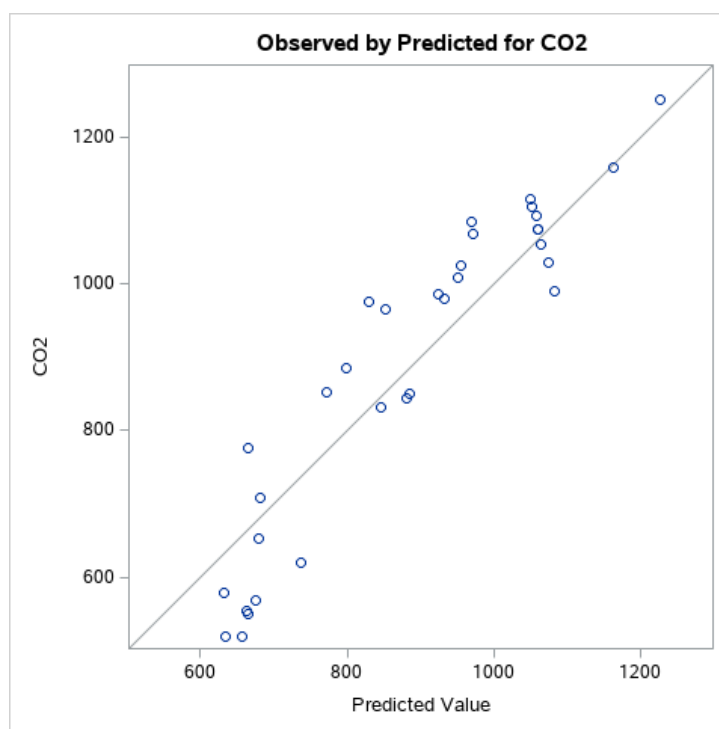
Tabela 10 - Coeficientes

<b>Coeficientes</b>		
	<i>CO<sub>2</sub></i>	GEE
Correlação	0,8581	0,9941
Determinação	0,7364	0,9884
Ajustado	0,7221	0,9877

Fonte: Elaborado pela autora

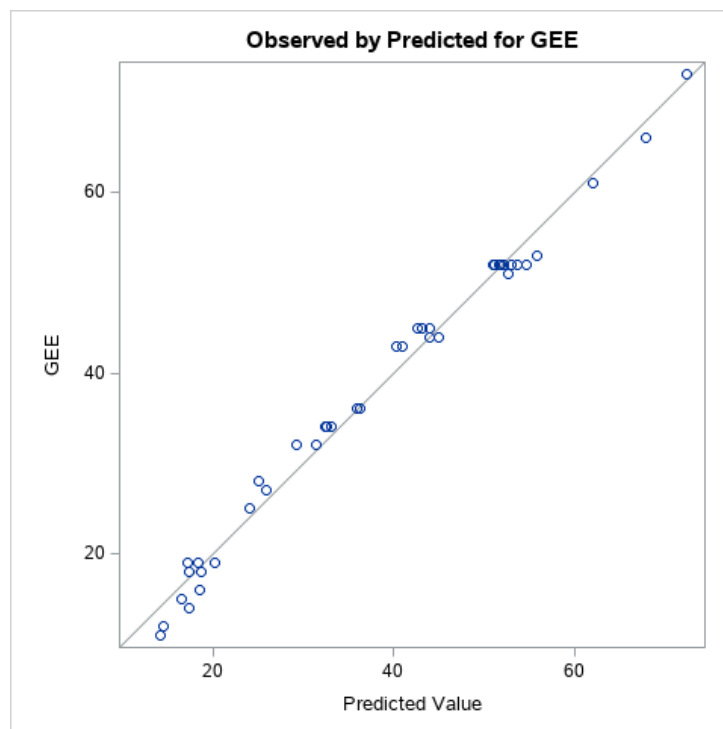
O poder de predição do modelo para a variável  $CO_2$ , mostrado na Figura 41, indica que os pontos observados estão próximos da reta de regressão. Já na Figura 42, o poder de predição para a variável GEE é ainda maior visto que a reta de regressão passa sobre a maior parte dos pontos observados, o que pode ser confirmado pelo valor dos coeficientes de determinação. Os gráficos apontam a qualidade dos modelos de predição e são coerentes com os coeficientes de determinação.

Figura 41 - Valores observados pelos valores preditos ( $CO_2$ )



Fonte: Elaborado pela autora

Figura 42 - Valores observados pelos valores preditos (GEE)



Fonte: Elaborado pela autora

O valor F de Snedecor encontrado através da análise de variância, apresentou, para o modelo 1, o valor de 51,67 e uma significância menor que 1% e para o modelo 2, o valor de 1569,81 e uma significância menor que 1%. Os resultados indicam que não há evidências de que os modelos sejam inadequados, conforme Tabela 11 e Tabela 12.

Tabela 11 - Teste de significância da regressão ( $CO_2$ )

Fonte de Variação	Graus de Liberdade	Soma dos Quadrados	Média Quadrática	F	Valor P
Regressão	2	1281351	640675	51,67	<.0001
Erro ou resíduo	37	458770	12399		
Total	39	1740121			

Fonte: Elaborado pela autora

Tabela 12 - Teste de significância da regressão (GEE)

<b>Fonte de Variação</b>	<b>Graus de Liberdade</b>	<b>Soma dos Quadrados</b>	<b>Média Quadrática</b>	<b>F</b>	<b>Valor P</b>
Regressão	2	10042	5021,0279	1569,81	<.0001
Erro ou resíduo	37	118	3,1985		
Total	39	10160			

Fonte: Elaborado pela autora

### 7.3 Diagnósticos e Análise dos Resíduos

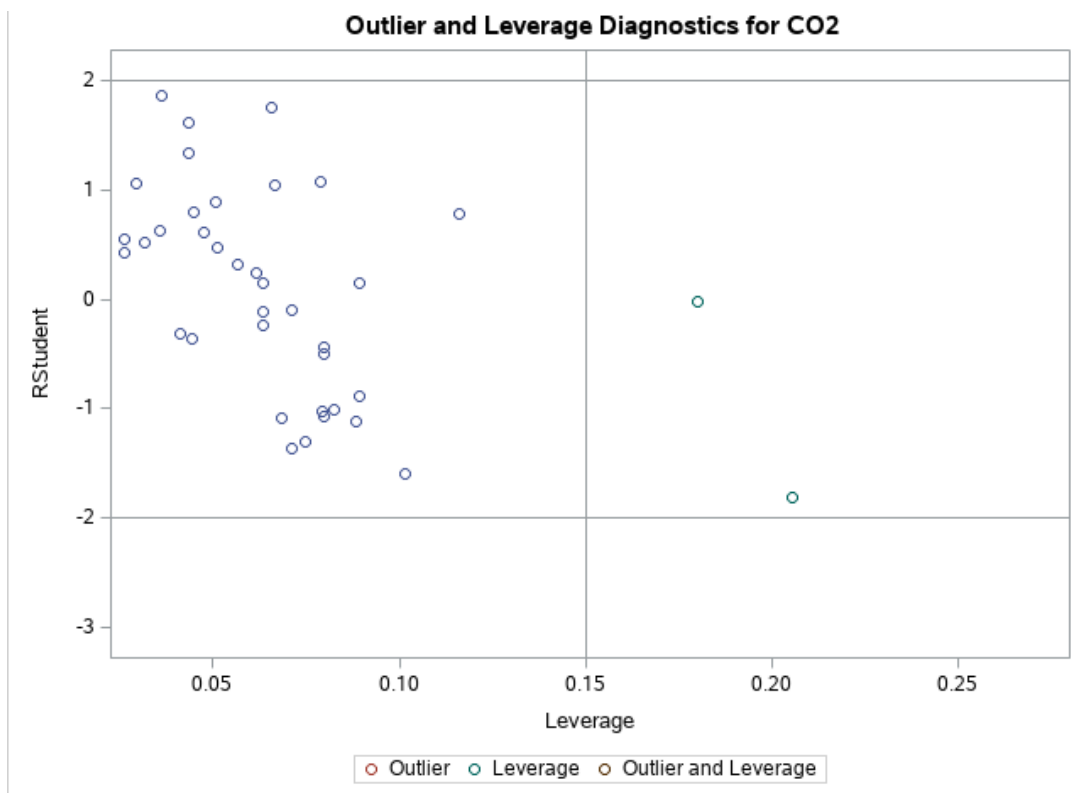
Os valores dos fatores de inflação da variância (VIF) indicados na Tabela 13, não apontaram alta correlação entre as variáveis independentes, visto que eles foram menores do que 5, não apresentando multicolinearidade.

Tabela 13 - Valores de inflação da variância

<b>Variável</b>	<b>VIF</b>
Área desmatada	1,037416
CO <sub>2</sub> absorvido	1,037416

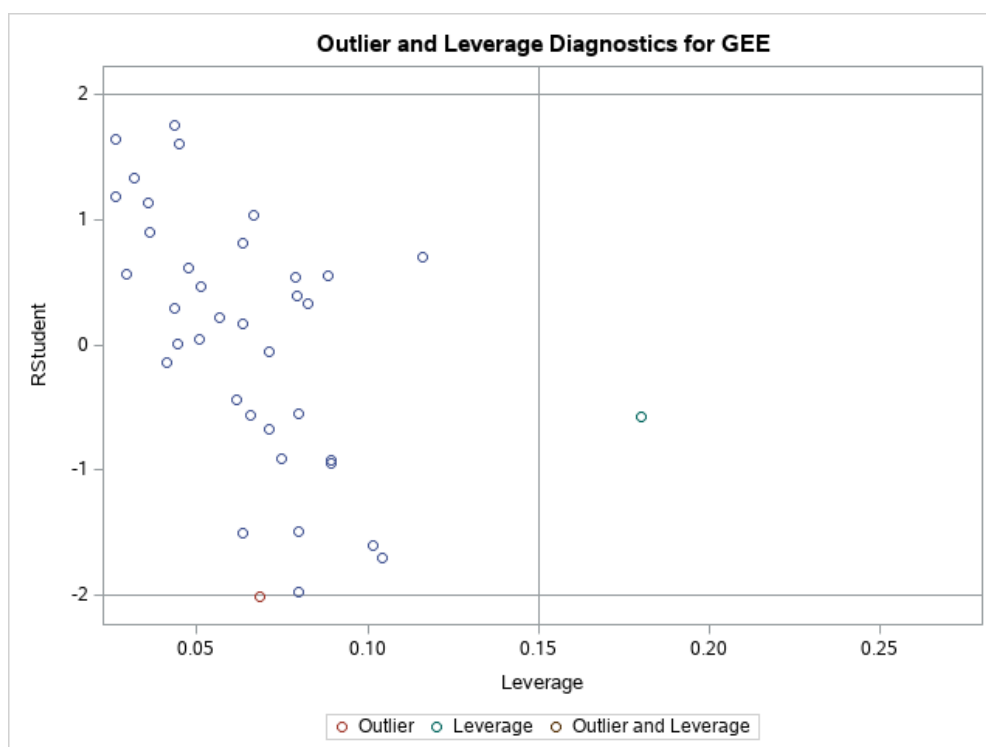
Fonte: Elaborado pela autora

Com a análise dos gráficos de dispersão dos resíduos, os outliers e pontos de alavanca considerados influentes foram retirados da amostra. O gráfico de Lverage disposto na Figura 43 e Figura 44 após o saneamento da amostra, mostra a existência de alguns pontos atípicos, contudo não são considerados influentes, uma vez que a maioria deles estão localizados entre o intervalo -2 e +2. Além disso, não há indicação de nenhuma observação influente, em ambos os casos, dado que no gráfico da distância de Cook todos os valores são inferiores a 0,5, conforme Figura 45 e Figura 46.

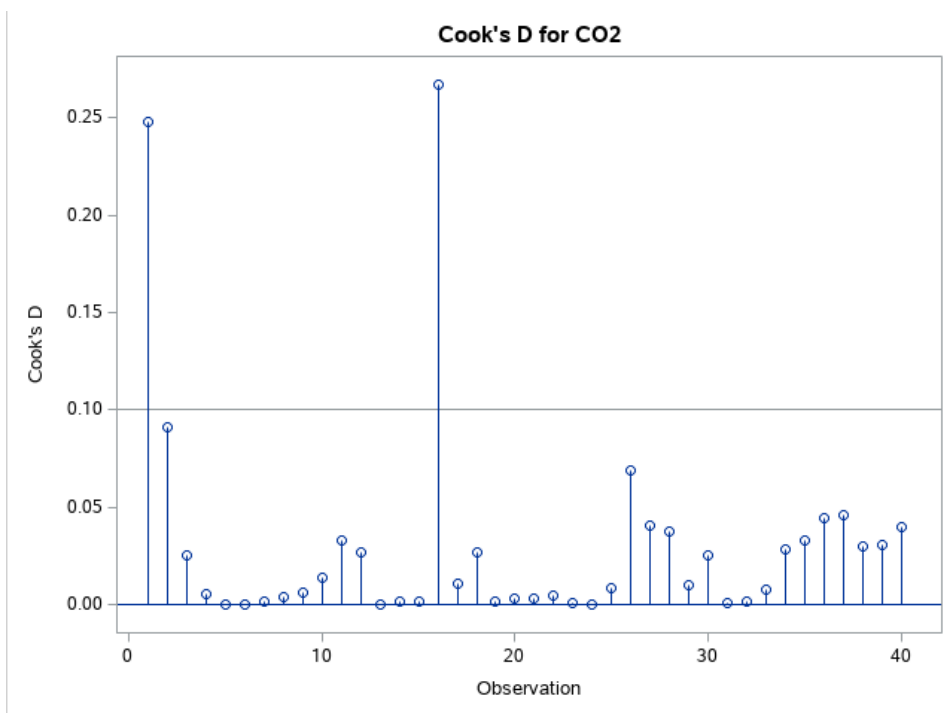
Figura 43 - Gráfico de pontos de alavanca e outliers ( $CO_2$ )

Fonte: Elaborado pela autora

Figura 44 - Gráfico de pontos de alavanca e outliers (GEE)

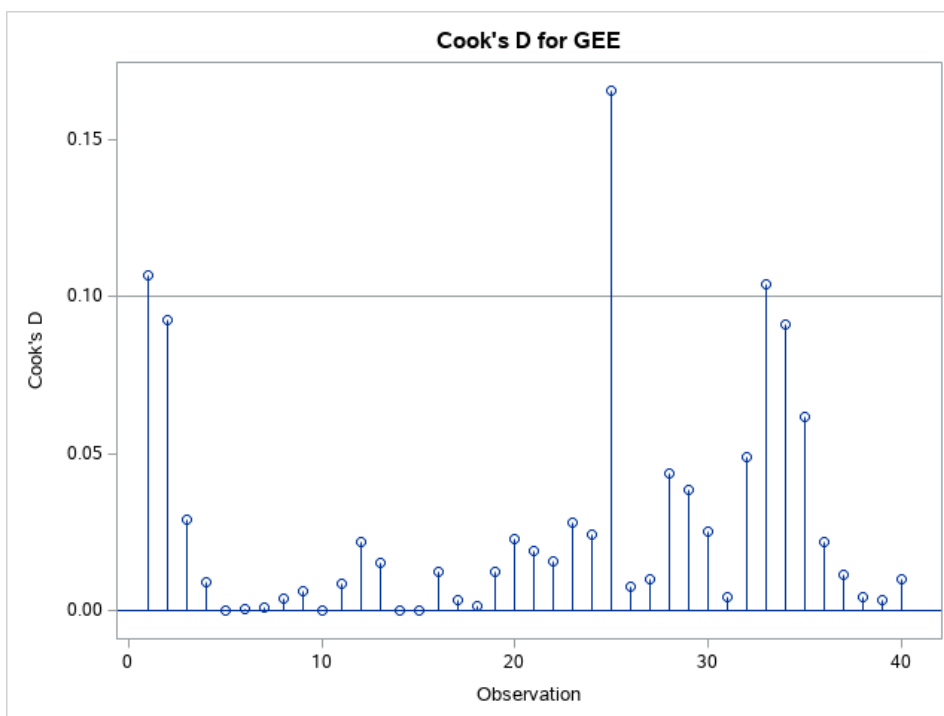


Fonte: Elaborado pela autora

Figura 45 - Gráfico da Distância de Cook ( $CO_2$ )

Fonte: Elaborado pela autora

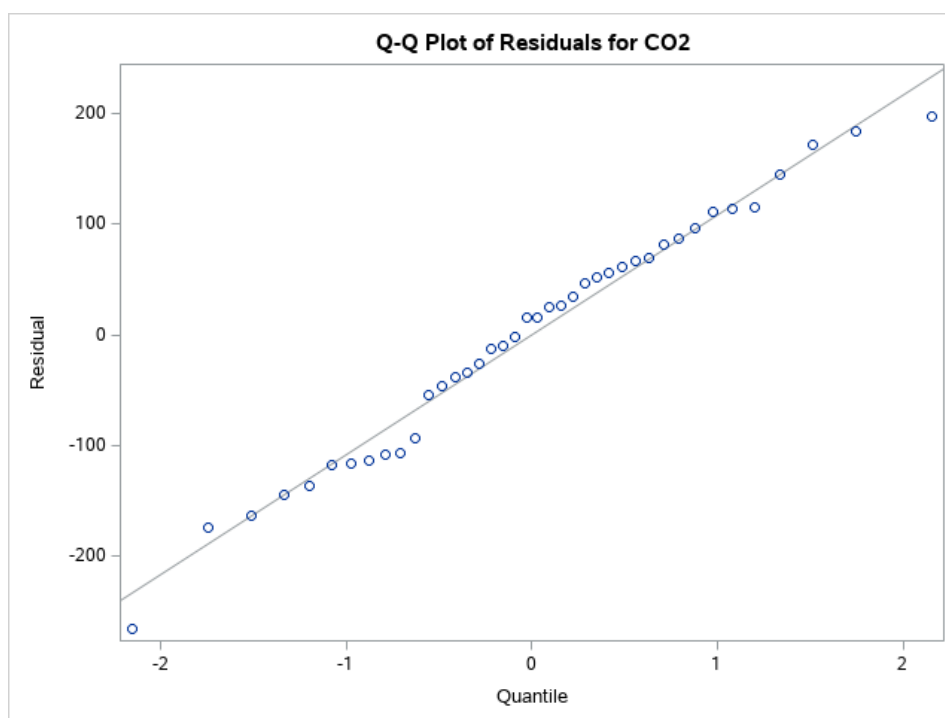
Figura 46 - Gráfico da Distância de Cook (GEE)



Fonte: Elaborado pela autora

Quanto à normalidade, observou-se que esse pressuposto também foi atendido devido à disposição dos pontos sobre a reta no Q-Q Plot dos resíduos na Figura 47 e Figura 48, e o diagnóstico foi ratificado através dos testes Shapiro-Wilk, Lilliefors e Anderson-Darling, conforme Tabela 14 e Tabela 15.

Figura 47 - Q-Q Plot dos Resíduos ( $CO_2$ )



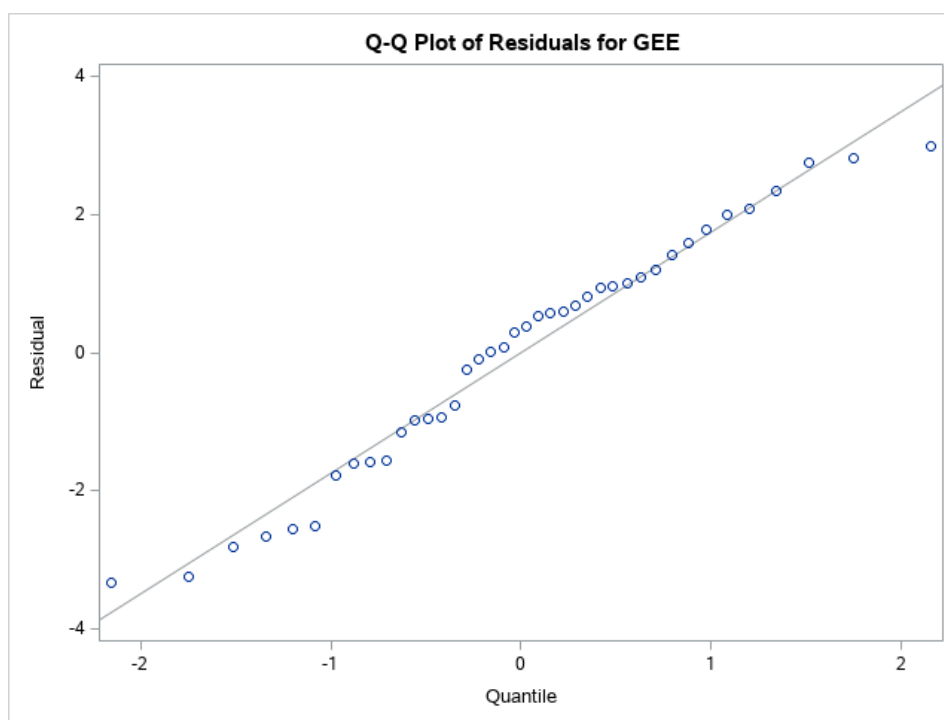
Fonte: Elaborado pela autora

Tabela 14 - Testes de normalidade dos resíduos ( $CO_2$ )

<b>Teste</b>	<b>Estatística</b>	<b>P-Valor</b>
Shapiro-Wilk	0,98051	0,7086
Anderson-Darling	0,27387	0,6465
Lilliefors	0,08868	0,5948

Fonte: Elaborado pela autora

Figura 48 - Q-Q Plot dos Resíduos (GEE)



Fonte: Elaborado pela autora

Tabela 15 - Testes de normalidade dos resíduos (GEE)

<b>Teste</b>	<b>Estatística</b>	<b>P-Valor</b>
Shapiro-Wilk	0,9664	0,2753
Anderson-Darling	0,27387	0,6465
Lilliefors	0,08868	0,5948

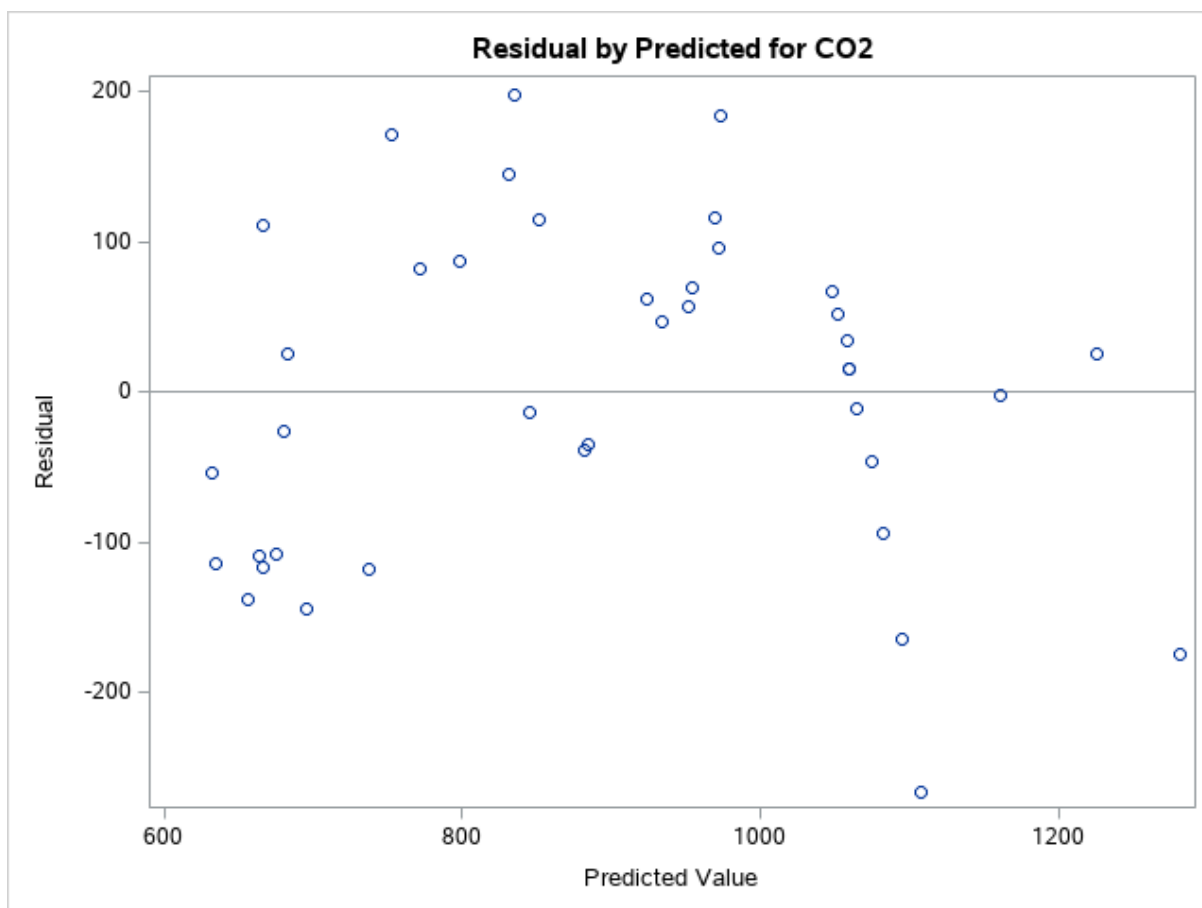
Fonte: Elaborado pela autora

A homocedasticidade dos resíduos foi testada conforme Tabela 16 e ainda foi observado através dos gráficos da Figura 49 e Figura 50 que os resíduos se distribuem de maneira aleatória e com amplitudes semelhantes em torno do zero, o que confirma a suposição de variância constante e independência dos resíduos.

Tabela 16 - Testes de variância constante

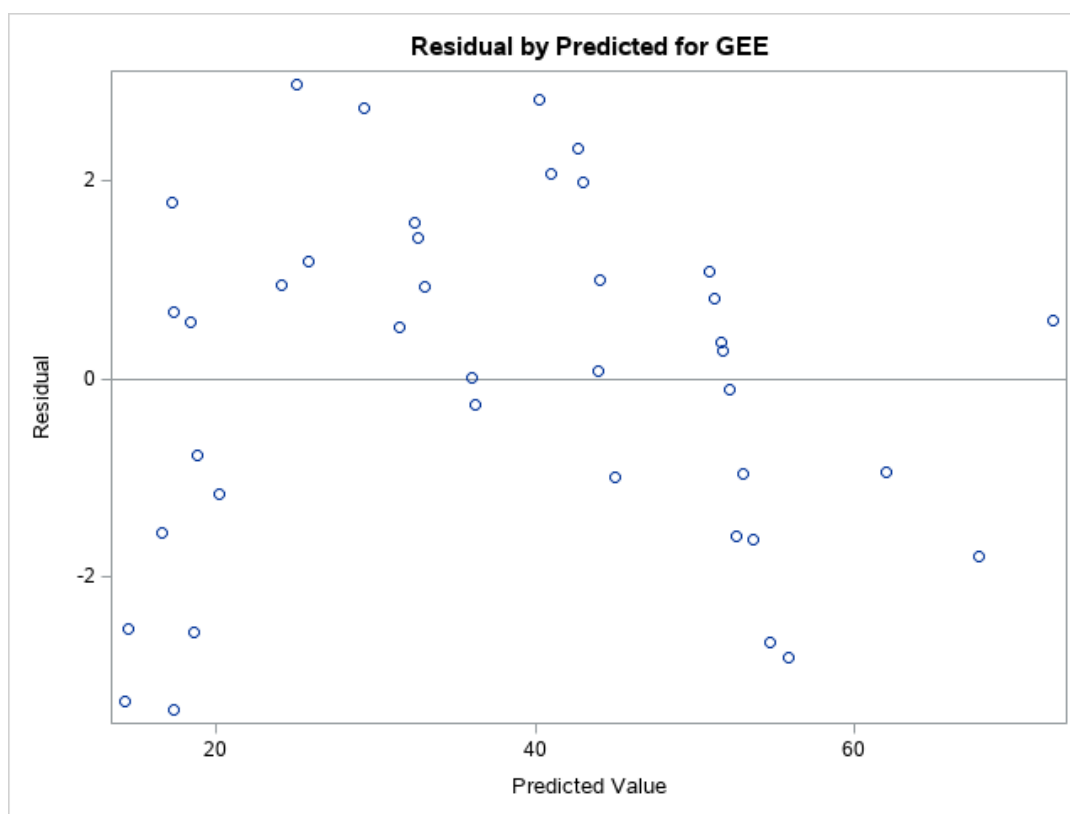
Teste	Estatística		P-Valor	
	$CO_2$	GEE	$CO_2$	GEE
NCV	0,6039	1,6554	0,43709	0,1982
Beusch-Pagan	1,0021	3,0712	0,6059	0,2153

Fonte: Elaborado pela autora

Figura 49 - Gráfico dos resíduos pelos valores estimados de ( $CO_2$ )

Fonte: Elaborado pela autora

Figura 50 - Gráfico dos resíduos pelos valores estimados de (GEE)



Fonte: Elaborado pela autora

#### 7.4 Uso da previsão

Uma vez que os modelos são significativos, a emissão de gases de efeito estufa pode ser estimada. Desta forma, os modelos podem ser utilizados para prever as emissões caso os valores das variáveis preditoras sofressem uma variação de 10% a mais e 10% a menos na última observação, por exemplo. Os valores previstos e os intervalos de confiança a 95% estão na Tabela 17 e Tabela 18.

Tabela 17 - Previsão para o modelo 1

	<b>Previsão</b>	<b>Intervalo de Confiança</b>
10% a mais	754,4534	[669,0571; 839,8496]
10% a menos	718,9996	[664,2104; 773,7889]

Fonte: Elaborado pela autora

Tabela 18 - Previsão para o modelo 2

	<b>Previsão</b>	<b>Intervalo de Confiança</b>
10% a mais	25,9133	[24,5417; 27,2848]
10% a menos	22,1817	[21,3017; 23,0616]

Fonte: Elaborado pela autora

De acordo com os resultados obtidos, um aumento de 10% nas observações da área desmatada e gás carbônico absorvido resultaria em um crescimento de 21% nas emissões de gás carbônico e de 4% nos outros gases poluentes, taxa considerada dentro do esperado. Em relação às emissões de  $CO_2$ , esse cenário seria considerado preocupante, visto que esse crescimento não é observado desde 1995, quando a taxa das emissões foi de 30% a mais.

Por outro lado, em um cenário onde houvesse uma queda de 10% no desmatamento e na absorção de gás carbônico, observaríamos um aumento de 15% nas emissões de gás carbônico, que ainda seria a maior taxa observada dos últimos 25 anos. Já as emissões de outros gases poluentes decairiam em 18%, a maior queda observada desde 2012.

Através das observações realizadas desde 1980, caso o cenário em que as melhores observações até então se repetissem, ou seja, a menor área desmatada e a maior quantidade de gás carbônico absorvido durante o processo de regeneração, a previsão é de que a emissão de  $CO_2$  e dos demais gases poluentes decaíssem em 30% e 79% respectivamente.

## 8 Conclusões e Considerações Finais

Após as análises realizadas no presente trabalho, foi possível compreender melhor a situação ambiental do país atualmente, as consequências para a população e para o futuro do ecossistema brasileiro. Nota-se que a tendência de cada aspecto analisado possui ligação direta com o contexto histórico do respectivo momento.

Da regressão linear múltipla foi constatado que a área desmatada e a quantidade de gás carbônico absorvido durante o processo de regeneração da floresta secundária influenciam na emissão de gases poluentes. Em ambos os modelos essas variáveis tiveram um bom valor em significância.

As análises aqui realizadas foram segmentadas de modo que duas funções de regressão foram criadas com mesmas variáveis dependentes e diferentes variáveis independentes, sendo elas: emissão de gás carbônico e emissão dos demais gases poluentes (metano e óxido nitroso). Em termos de análises de resíduos e diagnósticos, ambos foram considerados adequados, mas o modelo para previsão dos demais gases foi considerado mais consistente, tal fato sendo corroborado por apresentar valores menores dos Critérios de Informação de Akaike e Bayesiano.

Foi observado e ratificado que as variáveis preditoras são diretamente proporcionais às variáveis respostas. Quanto maior a área desmatada por meio de queimadas ou decomposição, maior a emissão dos gases. Por outro lado, apesar de os resultados dos coeficientes terem mostrado que quanto maior absorção de gás carbônico, maior a emissão de gases poluentes, os valores de absorção são medidos em escala negativa, assim sendo, as variáveis são inversamente proporcionais, ou seja, quanto maior for a absorção da vegetação secundária de  $CO_2$  em valor absoluto, menores serão as emissões de gases poluentes e, portanto confirma a necessidade de preservar a vegetação secundária a longo prazo.

Através do uso da predição, foi possível constatar que aumentar em 10% e diminuir em 10% as variáveis regressoras concomitantemente não foi eficaz em nenhum dos modelos. Diminuir em 10% o desmatamento e a absorção é eficiente para a redução da emissão do gás metano e óxido nitroso, mas não foi suficiente para a redução de gás carbônico. Para que haja uma redução eficiente da emissão dos gases é necessário reduzir ainda mais a área desmatada e aumentar a quantidade de

gás carbônico absorvido. Uma boa meta seria diminuir em 10% a área desmatada e aumentar em 10% o gás carbônico absorvido. O melhor cenário seria buscar alcançar os valores de área desmatada e absorção de gás carbônico dos anos 1991 e 2019, respectivamente, onde teve-se a menor área desmatada e a maior quantidade de gás carbônico absorvido pela floresta secundária, respectivamente.

Também foi criado um manual básico do software aqui utilizado, SAS University Studio, visto que ainda é uma ferramenta pouco conhecida e utilizada pelos estudantes e profissionais que fazem uso de técnicas estatísticas. Nele foi priorizado o esclarecimento quanto à instalação e interface do software, além de introduzir alguns conceitos estatísticos por meio de programação.

Por fim, o trabalho aqui apresentado alcançou seu objetivo inicial de explicar as emissões de gases poluentes causadores do efeito estufa e de criar uma função de previsão para os mesmos. Este estudo investigou e estimou o impacto humano no ecossistema brasileiro servindo, com isso, como base de futuras discussões deste gênero.

## 9 Bibliografia

ABBAD, G; TORRES, C V. **Regressão Múltipla Stepwise e Hierárquica em Psicologia Organizacional: Aplicações, Problemas e Soluções**. Estudos de Psicologia, Brasília, Df, v. 7, n. 7, p.19-29, 2002. Disponível em: <<http://www.scielo.br/pdf/epsic/v7nspe/a04v7esp.pdf>>. Acesso em: 25 jan. 2020.

ANGELO, Claudio. **Floresta regenerada é esponja de carbono**. 2016. Disponível em: <http://www.observatoriodoclima.eco.br/floresta-regenerada-e-esponja-de-carbono/>. Acesso em: 05 jan. 2020.

BRANDÃO JR. A, et al. **Evolução das emissões de gases de efeito estufa no Brasil (1990-2013) Setor de mudança de uso da terra**. 2015. Disponível em: <<https://imazon.org.br/PDFimazon/Portugues/livros/Evolucao%20Emissao%20Gases%20Efeito%20Estufa%20Brasil%201990-2013.pdf>>. Acesso em: 15 jan. 2020.

BRASIL. INPE. **Emissões brutas: Amazônia Brasileira: Emissões brutas por desmatamento corte raso**. 2019. Disponível em: <<http://inpe-em.ccst.inpe.br/emissoes-brutas-amz/>>. Acesso em: 13 out. 2019.

BUSSAB, W; A MORETTIN, P. **Estatística Básica**. 5. ed. São Paulo: Saraiva, 2003. 526 p.

CHAPMAN, David D. et al. **Determine what SAS Version and Components are available**. SAS Blog. 2013. Disponível em: <<https://www.lexjansen.com/nesug/nesug12/ma/ma10.pdf>>. Acesso em: 16 out. 2019.

COSTA, Camilla. **'A grande mentira verde': como a destruição da Amazônia vai além do desmatamento**. 2020. Disponível em: <https://www.bbc.com/portuguese/brasil-51317040>. Acesso em: 21 mar. 2020.

DENIS, Daniel J. **Applied Univariate, Bivariate, and Multivariate Statistics**. Journal of Statistical Software. Massachusetts, p. 2-4. set. 2016.

DUNNETT, C W. **A multiple comparisons procedure for comparing several treatments with a control.** Journal Of The American Statistical Association. Oxfordshire, p. 1096-1121. dez. 1955. Disponível em: <<https://www.jstor.org/stable/2281208?seq=1>>. Acesso em: 16 jan. 2020.

EMILIANO, P. A. et al. **Crítérios de Informação de Akaike Versus Bayesiano: Análise Comparativa.** Simpósio Nacional de Probabilidade e Estatística - 19º SINAPE. São Pedro-SP, 2010.

FAO. 2018. **The State of the World's Forests 2018 - Forest pathways to sustainable development.** Rome. Licence: CC BY-NC-SA 3.0 IGO.

FEARNSIDE, P M et al. **TROPICAL FOREST BURNING IN BRAZILIAN AMAZONIA: MEASUREMENT OF BIOMASS LOADING, BURNING EFFICIENCY AND CHARCOAL FORMATION AT ALTAMIRA, PARÁ.** Forest Ecology And Management, Netherlands, v. 123, n. 1, p.65-79, 11 out. 1999.

FEARNSIDE, Philip M. **Fogo e emissão de gases de efeito estufa dos ecossistemas florestais da Amazônia brasileira.** Estudos Avançados, [s.l.], v. 16, n. 44, p.99-123, abr. 2002. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0103-40142002000100007>.

GIBBS, H K et al. **Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s.** PNAS, Arizona, v. 107, n. 38, p.16732-16737, 21 set. 2010. ISSN 0027-8424.

GOLDEMBERG, José; LUCON, Oswaldo. **Energia, Meio Ambiente & Desenvolvimento.** 3. ed. São Paulo: Universidade de São Paulo, 2008.

GRAÇA, P M L. **Conteúdo de carbono na biomassa florestal da Amazônia e alterações após à queima.** 1997. 105 f. Dissertação (Mestrado) - Curso de Ciências

Florestais, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, São Paulo, 1997.

INPE-EM. **Estimativa de emissões de gases de efeito estufa (GEE) por mudança de cobertura da terra.** 2019. Disponível em: <<http://inpe-em.ccst.inpe.br>>. Acesso em: 15 jan. 2020.

JAMES, B.R. **Probabilidade: um curso em nível intermediário.** Livros Técnicos e Científicos. Editora S.A., 1981.

JUNGES, Alexandre Luis et al. **EFEITO ESTUFA E AQUECIMENTO GLOBAL: UMA ABORDAGEM CONCEITUAL A PARTIR DA FÍSICA PARA EDUCAÇÃO BÁSICA.** Experiências em Ensino de Ciências, Cuiabá, v. 13, n. 5, p.126-151, 23 nov. 2018. Disponível em: <[http://if.ufmt.br/eenci/artigos/Artigo\\_ID531/v13\\_n5\\_a2018.pdf](http://if.ufmt.br/eenci/artigos/Artigo_ID531/v13_n5_a2018.pdf)>. Acesso em: 12 out. 2019.

JUNGES, Alexandre Luis; SANTOS, Vinícius Yuri; MASSON, Neusa Teresinha; SANTOS, Francineide Amorim Costa. **EFEITO ESTUFA E AQUECIMENTO GLOBAL: UMA ABORDAGEM CONCEITUAL A PARTIR DA FÍSICA PARA EDUCAÇÃO BÁSICA.** Experiências em Ensino de Ciências, Mato Grosso, p.126-151, 23 nov. 2018. Disponível em: [http://if.ufmt.br/eenci/artigos/Artigo\\_ID531/v13\\_n5\\_a2018.pdf](http://if.ufmt.br/eenci/artigos/Artigo_ID531/v13_n5_a2018.pdf). Acesso em: 17 dez. 2019.

LAURANCE, W. F. **Gaia's Lungs Are rainforests inhaling Earth's excess carbon dioxide?** Natural History, North Carolina, v. 108, n. 2, p.96-96, mar. 1999.

LIEBSCHER, Eckhard. **A Universal Selection Method in Linear Regression Models.** Open Journal Of Statistics, New York, v. 2, n. 2, p.153-162, 23 abr. 2012.

MARIA, A. **Desmatamento alça novo voo na Amazônia.** 2013. Disponível em <<https://envolverde.cartacapital.com.br/desmatamento-alca-novo-voo-na-amazonia/>>. Acesso: em 17 jan. 2020.

MEIER, U. **A note on the power of Fisher's least significant difference procedure.** Pharm Stat 5: 253–263, 2006.

MONTGOMERY, D C. **Design and Analysis of Experiments.** 3. ed. Nova Jersey: John Wiley and Sons, 2013.

MONTGOMERY, D.C.; RUNGER, G.C. **Estatística aplicada e probabilidade para engenheiros.** 5. ed. Rio de Janeiro: Ltc, 2012.

MULLIGAN, J., ELLISON, G. and LEVIN, K. **Foundational Questions on Carbon Removal in the United States.** Working Paper. Washington, World Resources Institute. 2018 Disponível em: <<https://www.wri.org/publication/foundational-questions-on-carbon-removal-usa>>. Acesso em: 16 out. 2019.

PACHECO, M., & HELENE, M. **Atmosfera, fluxos de carbono e fertilização por CO<sub>2</sub>.** Estudos Avançados, v. 4, n. 9, p. 204-220, 1990. Disponível em: <<https://www.revistas.usp.br/eav/article/download/8567/10118>>. Acesso em: 16 out. 2019.

SAS Institute Inc. 2018. SAS Studio 5.1: Users Guide. Cary, NC: SAS Instiutute Inc.

SUÇUARANA, Monik da Silveira. **Floresta Amazônica: Biomass.** Disponível em: <<https://www.infoescola.com/biomass/floresta-amazonica/>>. Acesso em: 05 out. 2019.

TOLBERT, William. **How to Win Friends and Influence People with the SAS Output Delivery System.** Clinical Medicine & Research. Massachusetts, v. 8, n. 3-4, p. 189-190, 01 dez. 2010. Disponível em: < <http://www.clinmedres.org/content/8/3-4/189.3.abstract>>. Acesso em: 16 out. 2019.

TRIGUEIRO, A. **15 pontos para entender os rumos da desastrosa política ambiental do governo Bolsonaro.** 2019. Disponível em: <<https://g1.globo.com/natureza/blog/andre-trigueiro/post/2019/06/03/15-pontos-para->

[entender-os-rumos-da-desastrosa-politica-ambiental-no-governo-bolsonaro.ghtml](#)>.  
Acesso em: 17 jan. 2020.

VIEIRA, S. **Análise de Variância (Anova)**. São Paulo: Atlas, 2006. 216 p.