



UNIVERSIDADE DE PASSO FUNDO
FACULDADE DE CIÊNCIAS ECONÔMICAS,
ADMINISTRATIVAS E CONTÁBEIS
CENTRO DE PESQUISA E EXTENSÃO DA FEAC
(www.upf.br/cepeac)

Texto para discussão

Texto para discussão Nº 06/2021

**Roteiro RLS (Regressão Linear Simples) e RLM (Regressão Linear
Múltipla) no Software Statística 7.0**

Andre da Silva Pereira
Karolynne Sousa Gomes
Leticia Eduarda Bender
Valeska Morgana Corrêa Batistela

UNIVERSIDADE DE PASSO FUNDO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA
DE ALIMENTOS

**Roteiro RLS (Regressão Linear Simples) e RLM (Regressão Linear
Múltipla) no Software Statística 7.0**

André da Silva Pereira
Karolynne Sousa Gomes
Leticia Eduarda Bender
Valeska Morgana Corrêa Batistela

Passo Fundo-RS

2021

LISTA DE ILUSTRAÇÕES

Figura 1: Gráfico de regressão linear simples.....	2
Quadro 1- Dados contendo variáveis dependente e independente.....	3
Figura 2: Importação da tabela no software Statistica 7.0	4
Figura 3: Passo 1	5
Figura 4: Passo 2	6
Figura 5: Passo 3	6
Figura 6: Passo 4	7
Figura 7: Resultado do exemplo gerado para regressão linear simples	7
Figura 8: Passo 1 (gráfico)	8
Figura 9: Passo 2 (gráfico)	9
Figura 10: Passo 3 (gráfico)	9
Figura 11: Geração do gráfico de dispersão entre as variáveis tempo e distância.	10
Figura 12: Selecionar a opção para regressão linear múltipla.....	11
Figura 13: Resultados gerados para a regressão múltipla.	11
Figura 14: Tabela para a geração do gráfico (Passo 2).	12
Figura 15: Selecionar as variáveis (Passo 3).	12
Figura 6: Gráfico de dispersão RLM.....	13

SUMÁRIO

INTRODUÇÃO	1
1. ROTEIRO DE REGRESSÃO LINEAR SIMPLES (RLS) NO STATISTICA 7.0	2
2. ROTEIRO DE REGRESSÃO LINEAR MÚLTIPLA (RLM) NO STATISTICA 7.0	3
3. DIAGRAMA DE DISPERSÃO	3
4. EXEMPLO DE REGRESSÃO SIMPLES NO STATISTICA 7.0	3
5. EXEMPLO DE ANÁLISE DE REGRESSÃO MÚLTIPLA NO STATISTICA 7.0	10
REFERÊNCIAS	15

APOSTILA

MÉTODOS QUANTITATIVOS

INTRODUÇÃO

A análise de regressão linear simples e múltipla é amplamente utilizada em diversas áreas do conhecimento, para investigar relações entre variáveis e construção de um modelo (equação).

- **Regressão Linear Simples:** definida como uma relação linear entre a variável dependente e uma variável independente.

- **Regressão Linear Múltipla:** relação linear entre várias variáveis independentes e uma variável dependente.

Entre os resultados que são obtidos a partir da análise de correlação e regressão, estão:

- R múltiplo: expressa o nível de correlação entre a variável dependente e independente, quanto maior o valor melhor o modelo;

- R^2 : representa o quanto os dados estão próximos da regressão ajustada, representa o quanto o modelo criado representa os dados;

- R^2 ajustado: é uma versão modificada do R^2 , ajustada para o número de preditores no modelo;

- F: utilizado para determinar se a variabilidade entre as médias do grupo é maior que a das observações dentro dos grupos;

- p: é a probabilidade de se obter teste igual ou muito próximo do observado em uma amostra, sob hipótese nula. Muitas vezes é utilizada significância de 5 % em que pode ser rejeitada a hipótese nula caso o valor-p seja menor que 5 %;

- Erro padrão da estimativa: média de variação de uma média amostral em relação à média da população.

Entre os softwares utilizados para aplicação da regressão linear estão o Excel, a linguagem de programação R, o SPSS e o Statistica 7. Nesta apostila será (1) introduzido conceitos de regressão simples e múltipla, (2) descrito como aplicar os dados e variáveis no software Statistica 7.0, (3) realizada análise dos dados obtidos através dos exemplos.

1. ROTEIRO DE REGRESSÃO LINEAR SIMPLES (RLS) NO STATISTICA 7.0

- Envolve variáveis dependentes (VD) ou independentes (VI);
- VD é explicada ou saída;
- VI é explicativa ou preditora;

A regressão linear apresenta forma linear e pode ser definida através da Equação 1.

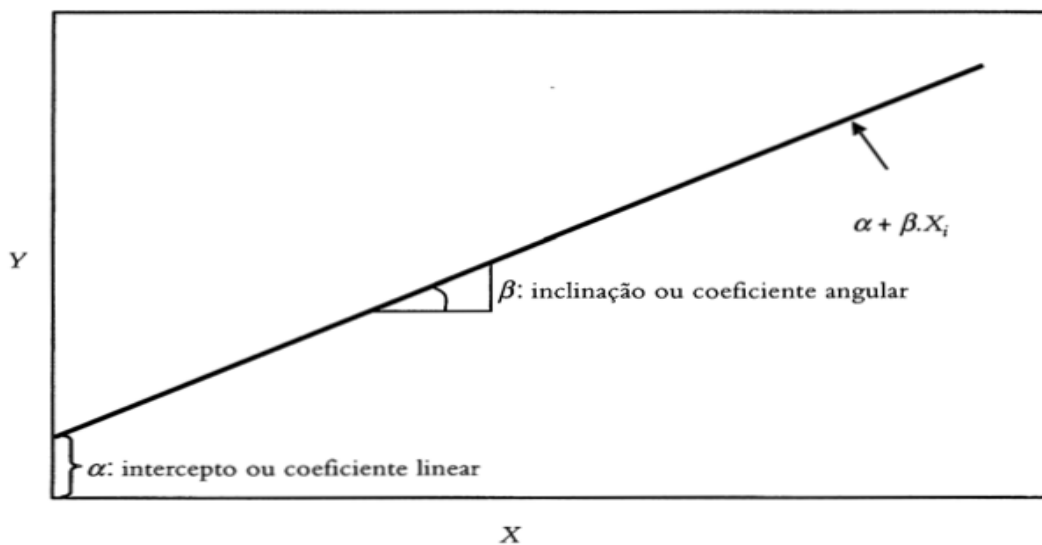
$$Y_i = \alpha + \beta * X_i \quad [\text{Equação 1}]$$

Onde:

Y = valor previsto da variável dependente;

i, α e β = parâmetros estimados do intercepto e inclinação do modelo;

Figura 1: Gráfico de regressão linear simples.



Fonte: FÁVERO; BELFIORE, 2017.

2. ROTEIRO DE REGRESSÃO LINEAR MÚLTIPLA (RLM) NO STATISTICA 7.0

$$Y_i = a + b_1 * X_{1i} + b_2 * X_{2i} + \dots + b_k * X_{ki} + u_i \quad [\text{Equação 2}]$$

Onde:

Y = Fenômeno sendo estudado (variável dependente quantitativa);

a = intercepto (constante ou coeficiente linear);

b_j ($j = 1, 2, \dots, k$) = coeficiente angular;

X_j = variável explicativa;

u = termo do erro (diferença entre Y real e calculado);

i = observações das análises.

Para a interpretação dos resultados das análises de regressão gera-se um diagrama de dispersão, que permite decidir experimentalmente se uma relação linear entre as variáveis deve ser estabelecida.

3. DIAGRAMA DE DISPERSÃO

Por interpretação do diagrama ou gráfico de dispersão, definimos o grau de relação linear entre as variáveis, se é uma relação forte ou fraca. Conforme o modo como se situam os pontos no diagrama ao redor de uma reta que passa entre os pontos. Considera-se a correlação forte quando os pontos se concentram mais, com pequenos desvios, em relação a reta.

4. EXEMPLO DE REGRESSÃO SIMPLES NO STATISTICA 7.0

A seguir, será representado em um exemplo como fazer análise de regressão linear simples e múltipla utilizando o software Statística 7.0. Os dados que serão utilizados para exemplificar estão apresentados no Quadro 1.

Quadro 1- Dados contendo variáveis dependente e independente

Tempo (min) (Y)	Distância (km) (X1)	Quantidade de Cruzamentos (X2)	Quantidade de Cruzamentos (X2)	Quantidade de Cruzamentos (X2)
--------------------	------------------------	--------------------------------------	--------------------------------------	--------------------------------------

15	8	16	32	12
20	6	12	24	20
20	15	30	60	25
40	20	40	39	37
50	25	50	100	32
25	11	22	44	17
10	5	10	20	9
55	32	64	128	60
35	28	56	112	12
30	20	40	80	17

Os dados da tabela do exemplo devem ser importados para a tabela disponível no software Statística 7.0, como mostra a Figura 2.

Figura 2: Importação da tabela no software Statística 7.0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	Tempo y	Distância y	quantidade de	quantidade de	quantidade de	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	15	8	16	32	12														
2	20	6	12	24	20														
3	20	15	30	60	25														
4	40	20	40	39	37														
5	50	25	50	100	32														
6	25	11	22	44	17														
7	10	5	10	20	9														
8	55	32	64	128	60														
9	35	28	56	112	12														
10	30	20	40	80	17														
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			
20																			
21																			
22																			
23																			
24																			
25																			
26																			

Fonte: AUTOR, 2021

Após todos os dados estarem devidamente descritos no programa, devemos abrir a aba "Statistics" e procurar pela função "Multiple Regression".

Desta forma, será aberta uma nova janela no programa para selecionarmos as variáveis dependentes e a variável independente. Assim, será gerada tabela em que estarão o valor de R múltiplo, R2 múltiplo, R2 ajustado, F, p, e o erro estimado.

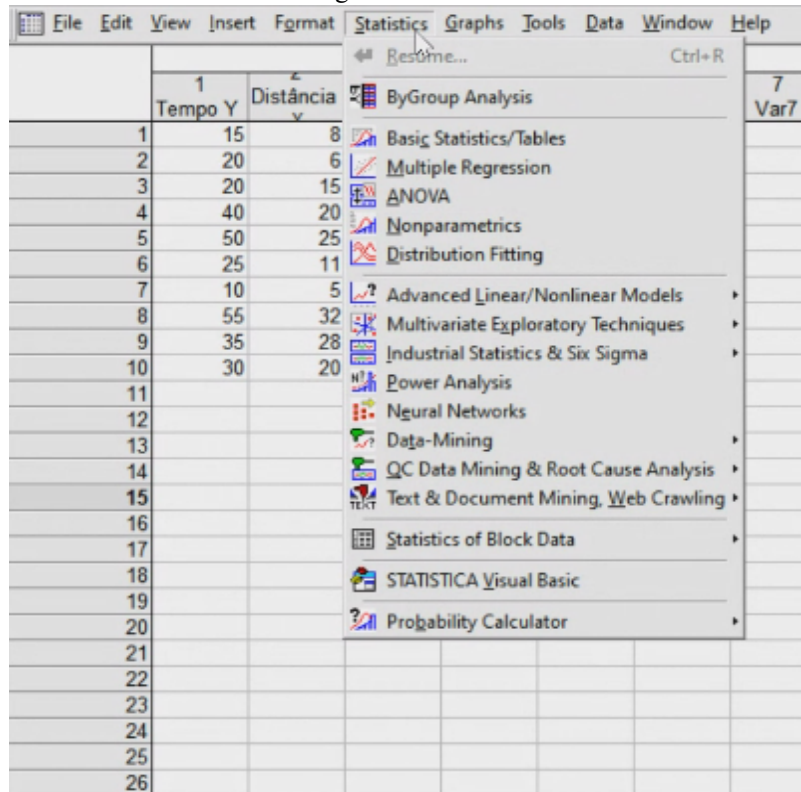
Para realização da análise de regressão linear simples os passos seguintes deverão ser realizados.

1. "Statistics" -Selecionar a opção "Multiple Regression"

2. “Variables”
3. Selecionar as variáveis dependentes e independentes
4. Selecionar a opção “summary Regression Results”.

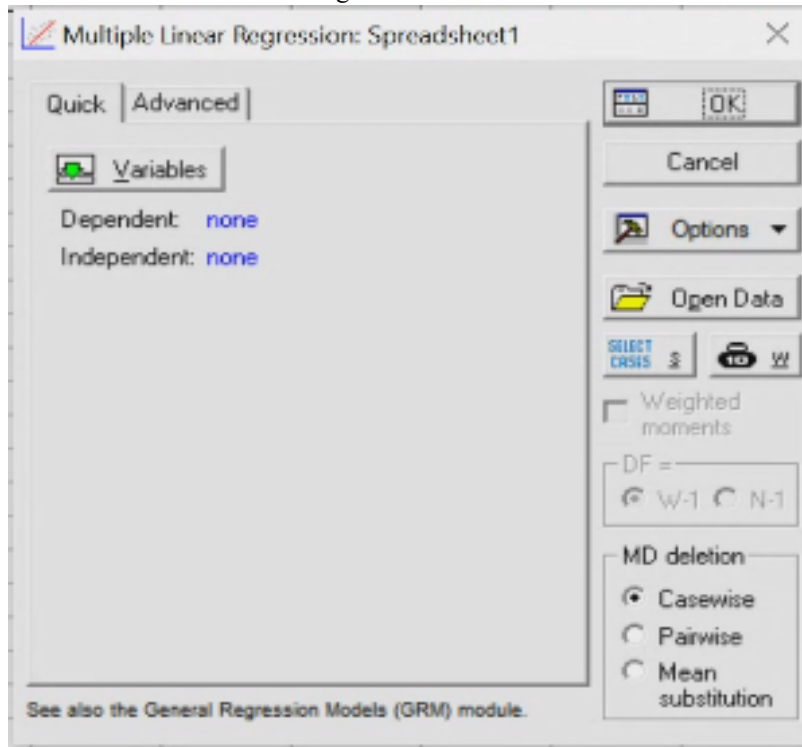
Os passos citados acima estão exemplificados nas figuras abaixo.

Figura 3: Passo 1



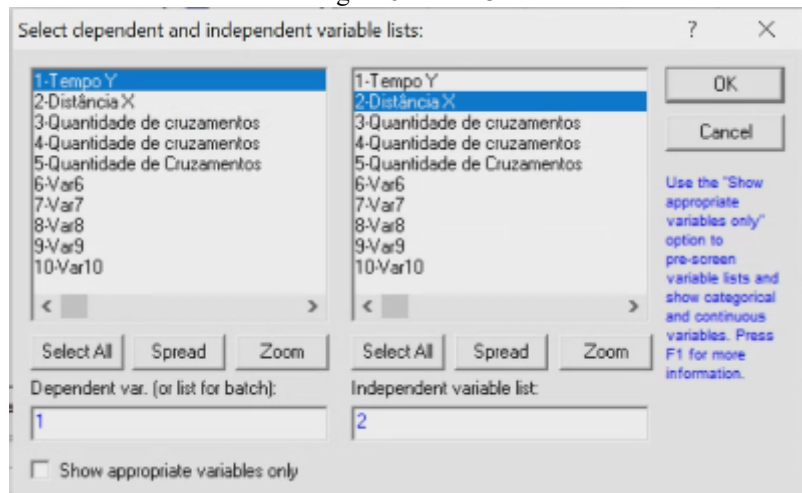
Fonte: AUTOR, 2021

Figura 4: Passo 2



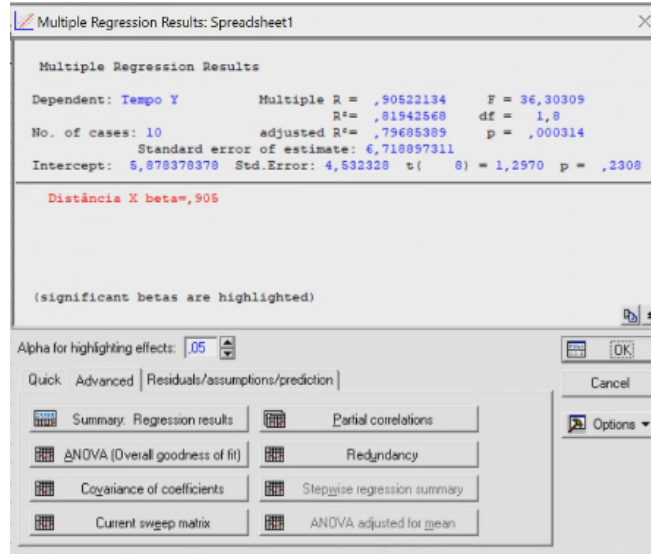
Fonte: AUTOR, 2021

Figura 5: Passo 3



Fonte: AUTOR, 2021

Figura 6: Passo 4



Fonte: AUTOR, 2021

Figura 7: Resultado do exemplo gerado para regressão linear simples

Statistic	Summary Value
Multiple R	0,90522
Multiple R ²	0,81943
Adjusted R ²	0,79685
F(1,8)	36,30309
p	0,00031
Std.Err. of Estimate	6,71890

Fonte: AUTOR, 2021

O “Multiple R” significa a correlação entre as variáveis tempo e distância, como o resultado obtido no exemplo é próximo de 1 (quanto mais próximo de 1 mais forte a correlação) maior significa que a correlação é positiva.

O “Multiple R²” é a explicação do modelo utilizado, portanto em 81,94% dos casos a distância explica chegar atrasado.

O “Adjusted R²” é a confiabilidade do modelo

O “F” é para a comparação dos modelos lineares

O “p” está relacionado com o teste de hipótese, se for igual a 0 significa que o ajuste do modelo não tem efeito, se for menor que 0,05 indica que se pode rejeitar a hipótese nula, ou seja provavelmente será uma adição significativa ao modelo

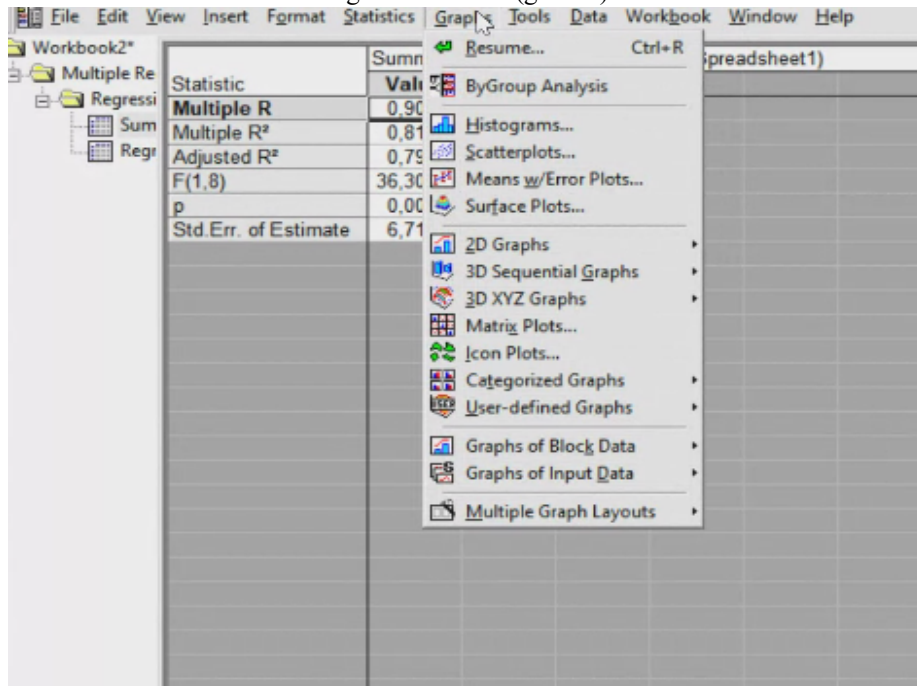
O “Err. Of Estimate” é a distância das médias em relação a linha de regressão

Após a realização da análise um gráfico de dispersão deve ser gerado. Para a interpretação dos resultados.

O gráfico será gerado seguindo os passos abaixo.

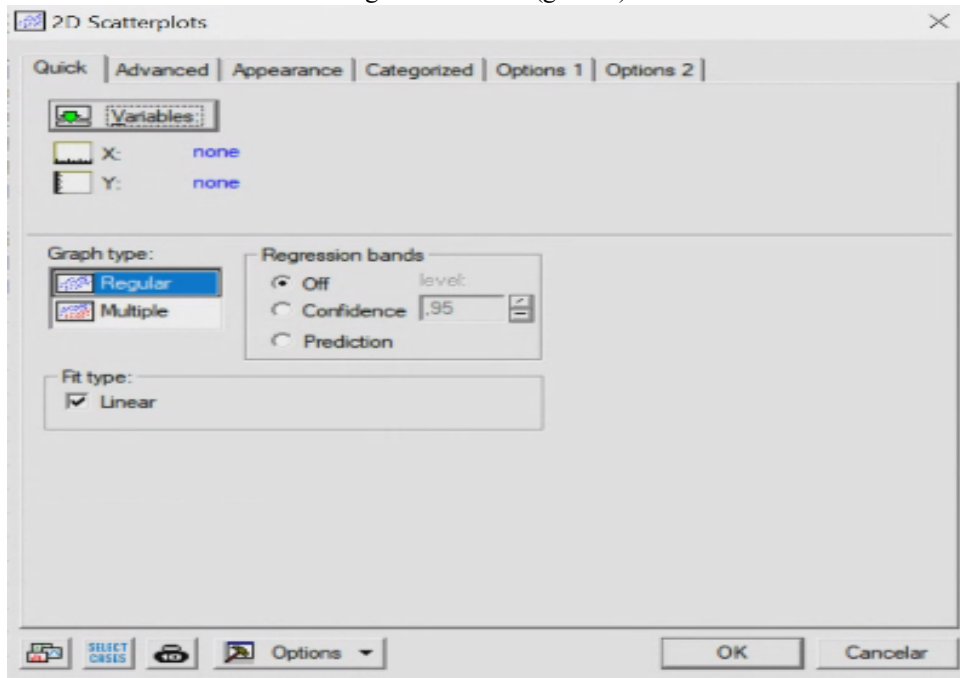
1. “Graphs” - selecione a opção “scatterplots”
2. Selecione as variáveis- “Regular”
3. Selecionar as variáveis dependente e independente (apenas uma por vez).

Figura 8: Passo 1 (gráfico)



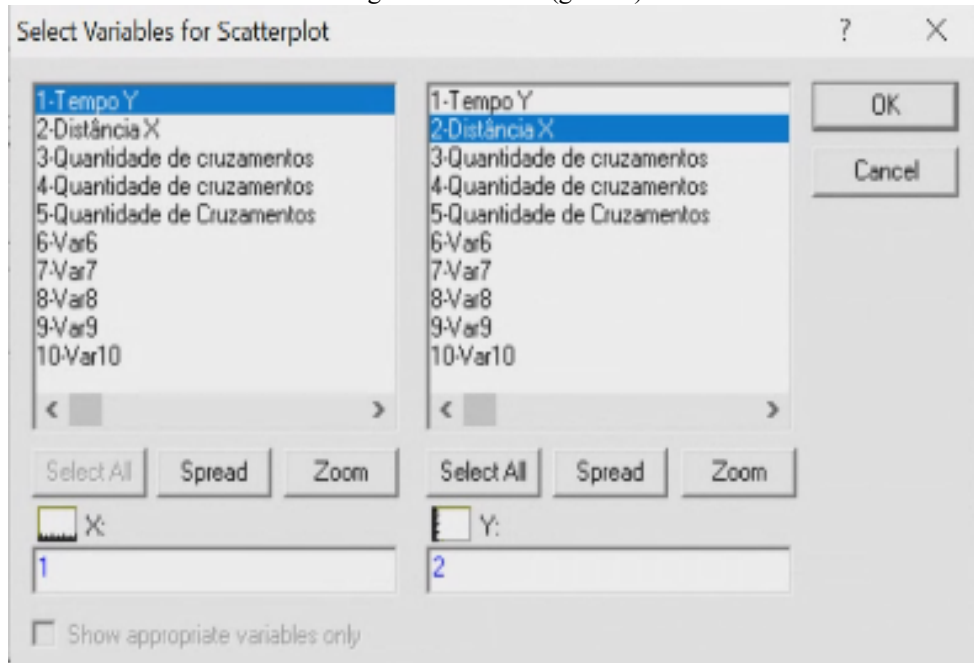
Fonte: AUTOR, 2021

Figura 9: Passo 2 (gráfico)



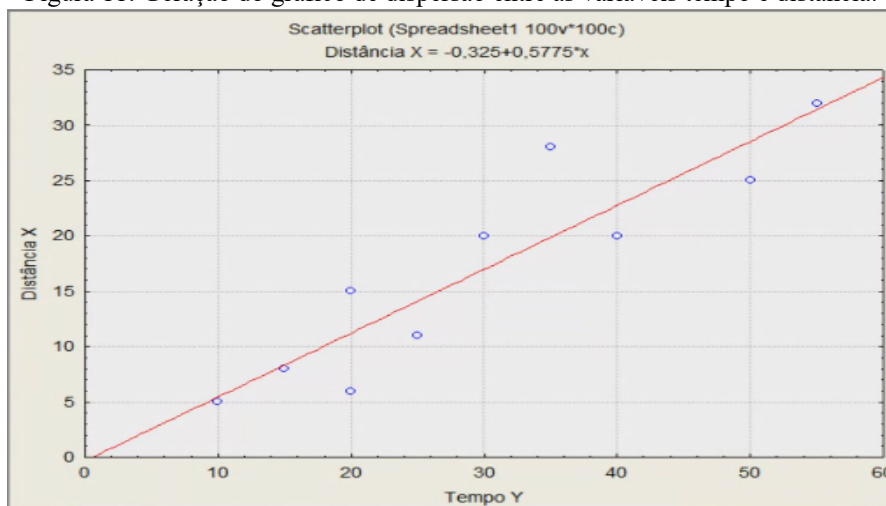
Fonte: AUTOR, 2021

Figura 10: Passo 3 (gráfico)



Fonte: AUTOR, 2021

Figura 11: Geração do gráfico de dispersão entre as variáveis tempo e distância.



Fonte: AUTOR, 2021

O gráfico indica uma correlação linear positiva forte, pois como pode se observar na Figura 10, a linha segue em sentido crescente indicando assim que a correlação apresenta valor próximo de 1.

5. EXEMPLO DE ANÁLISE DE REGRESSÃO MÚLTIPLA NO STATISTICA 7.0

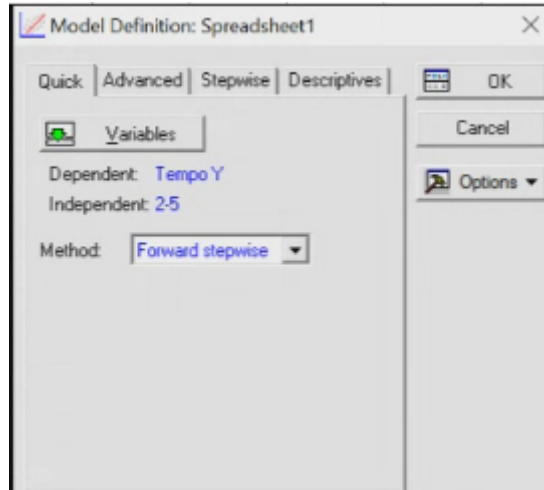
Para realização da análise de regressão múltipla segue-se basicamente os mesmos passos, porém com algumas alterações como descrito a seguir:

1. “Statistics” -Selecionar a opção “Multiple Regression”
2. “Variables” - Selecionar a variável dependente e as variáveis independentes.

Observação: Na regressão múltipla as variáveis são selecionadas todas de uma vez diferente de como ocorre na regressão linear simples onde as variáveis independentes são selecionadas uma por vez.

3. Na tela inicial da definição de variáveis deve-se ir na opção “method” e selecionar a opção “forward stepwise”.
4. E então apertar em “OK” e a tabela de regressão será gerada.

Figura 12: Selecionar a opção para regressão linear múltipla.



Fonte: AUTOR, 2021

Figura 13: Resultados gerados para a regressão múltipla.

Statistic	Summary Value
Multiple R	0,94947
Multiple R ²	0,90150
Adjusted R ²	0,87335
F(2,7)	32,03212
p	0,00030
Std.Err. of Estimate	5,30505

Fonte: AUTOR, 2021

A interpretação dos dados apresentados segue os mesmos padrões de interpretação da análise de regressão linear simples. Porém para o diagrama de dispersão na RLM (regressão linear múltipla) é necessário seguir os passos abaixo.

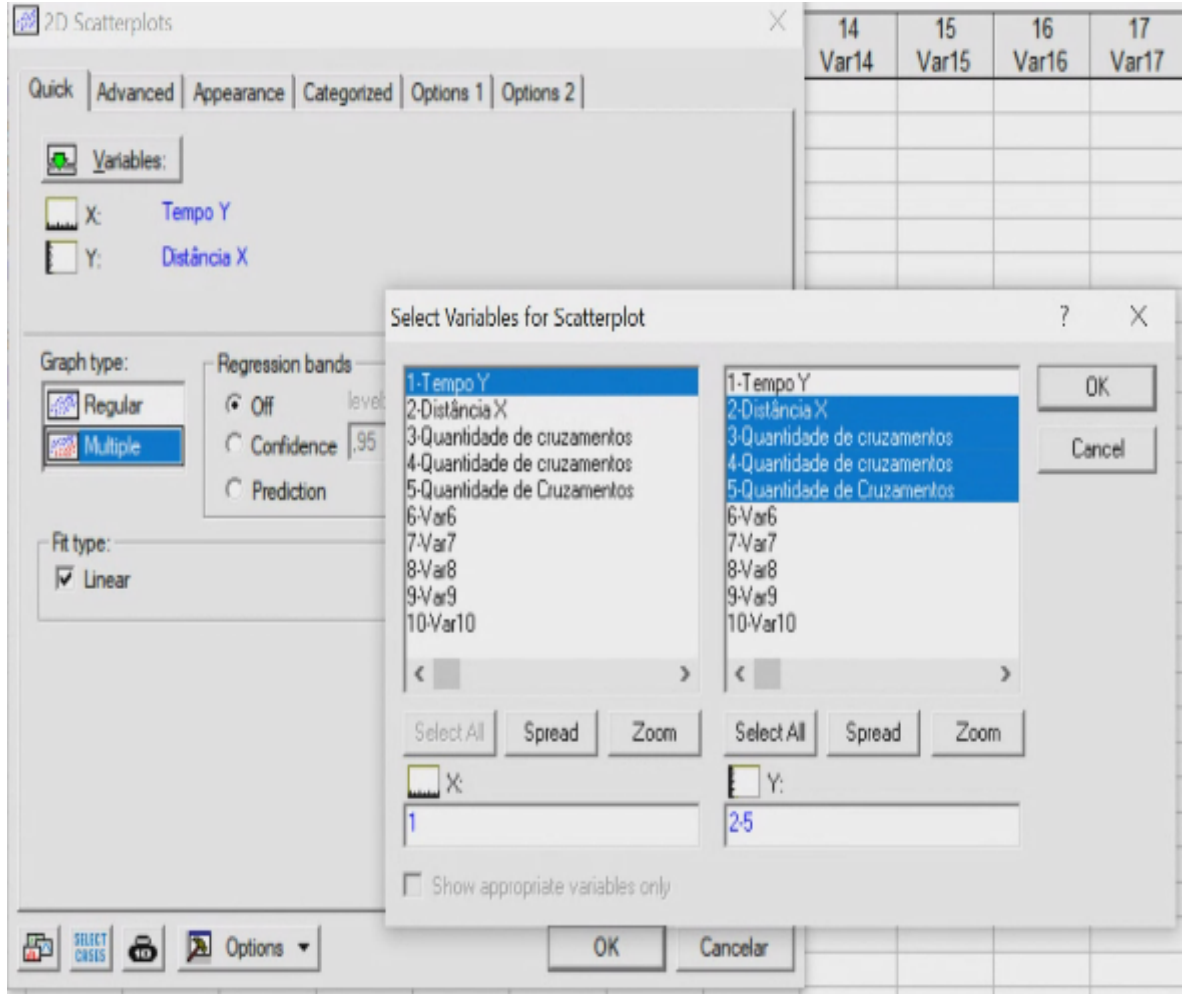
Geração do gráfico de dispersão RLM

1. “Graphs” - selecione a opção “scatterplots”
2. Selecione no painel inicial das variáveis a opção “multiple”
3. Apertar na opção “variáveis” e selecionar as variáveis para geração do gráfico.

Observação: Aqui todas as variáveis independentes serão selecionadas de uma só vez.

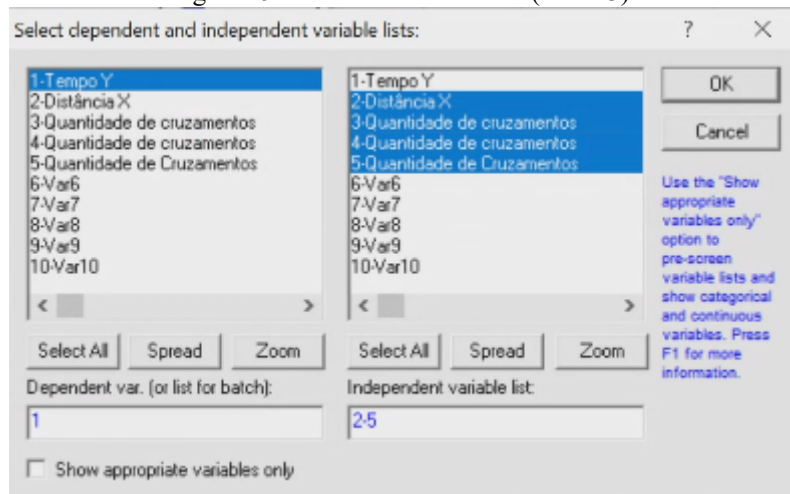
4. Aperta em “OK” e o gráfico será gerado.

Figura 14: Tabela para a geração do gráfico (Passo 2).



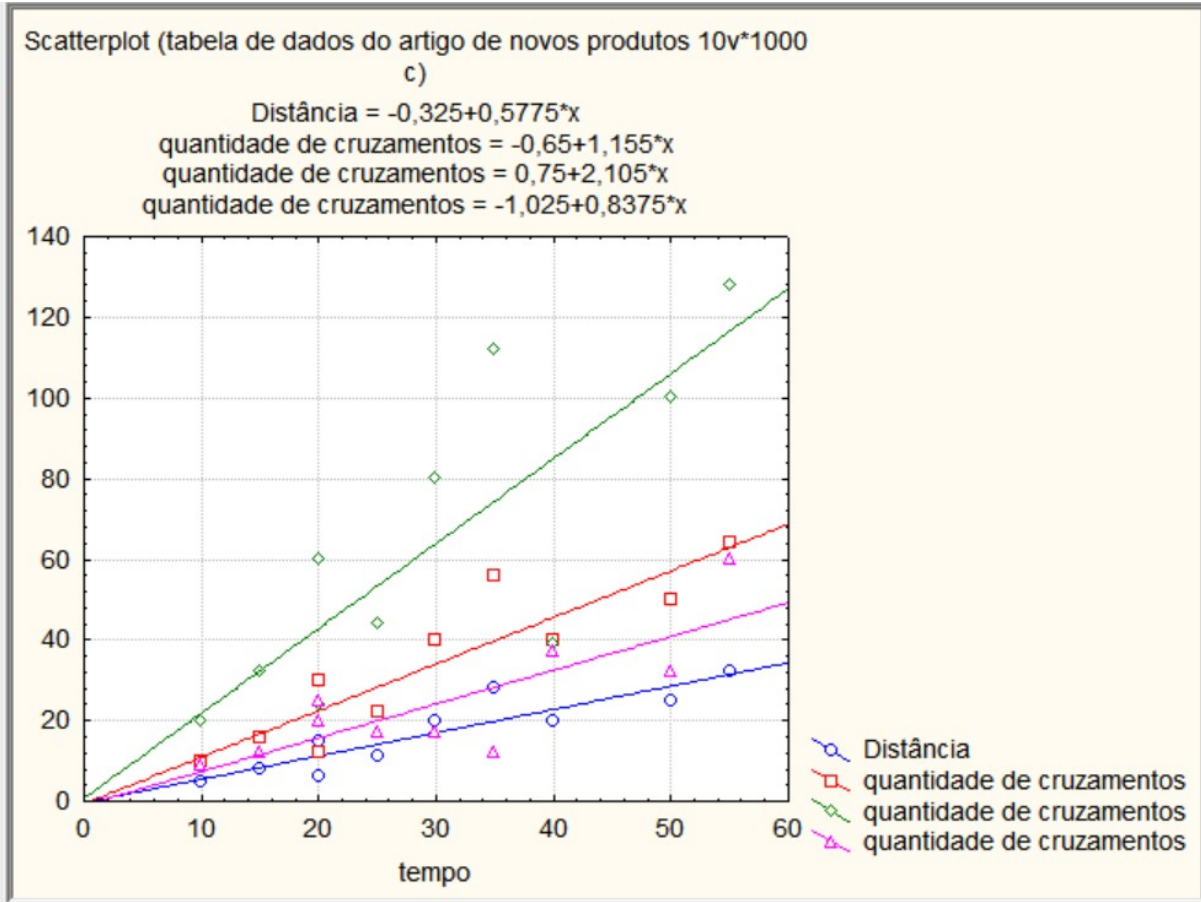
Fonte: AUTOR, 2021

Figura 15: Selecionar as variáveis (Passo 3).



Fonte: AUTOR, 2021

Figura 6: Gráfico de dispersão RLM.



Fonte: AUTOR, 2021

O tempo possui forte correlação com todas as variáveis independentes e que a hipótese do modelo utilizado não é nula.

CONCLUSÃO

O software Statistica 7.0 pode ser utilizado para as análises de regressão linear, tanto simples quanto múltiplas. Apresentando as ferramentas necessárias para que os dados de regressão linear sejam interpretados corretamente, como por exemplo o diagrama de dispersão.

REFERÊNCIAS

HENRIQUES, Carla. Análise de regressão linear simples e múltipla. **Departamento de Matemática. Escola Superior de Tecnologia de Viseu.** Portugal, 2011.

FÁVERO, L. P.; BELFIORE, P. **Manual de análise de dados: Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®.** 1. ed. Rio de Janeiro: Elsevier, 2017, p. 1219.