

USER MANUAL

M A T C P - I N T E G R A T I V E P R O J E C T

**PREPARED FOR:**

**LUÍS AFONSO (TPA)**

User Manual

REAL ESTATE USA

Developed by CTRL-ALT-DEFEAT

Contents & Review

Pedro Coelho 1220688@isep.ipp.pt

Luna Silva 1221184@isep.ipp.pt

Diogo Moutinho 1221014@isep.ipp.pt

Vasco Sousa 1221700@isep.ipp.pt

Rafael Araújo 1201804@isep.ipp.pt

Instituto Superior de Engenharia do Porto

MATCP 22/23

18.06.2023



Table of Contents

[Simple Linear Regression 9](#_Toc138015349)

[Overview of Simple Linear Regression 9](#_Toc138015350)

[Simple Linear Regression Model 9](#_Toc138015351)

[Model Significance 10](#_Toc138015352)

[Confidence intervals for prediction values 29](#_Toc138015353)

[Multi Linear Regression 34](#_Toc138015354)

[Overview of Multi Linear Regression 34](#_Toc138015355)

[Multiple Linear Regression Model 34](#_Toc138015356)

[Model Significance 35](#_Toc138015357)

[Coefficient determination 36](#_Toc138015358)

[Hypothesis tests for model coefficients 38](#_Toc138015359)

[Confidence intervals for prediction values 41](#_Toc138015360)

Table of Images

[Image 1 - Method\_1\_MATCP\_Theorical\_7\_Slide\_5 6](#_Toc138015639)

[Image 2 - Method\_2\_3\_4\_MATCP\_Theorical\_7\_Slide\_10 7](#_Toc138015640)

[Image 3 - Method\_5\_6\_MATCP\_Theorical\_7\_Slide\_11 7](#_Toc138015641)

[Image 4 - Method\_7\_MATCP\_Theorical\_10\_Slide\_32 7](#_Toc138015642)

[Image 5 - Method\_8\_9\_10\_MATCP\_Theorical\_7\_Slide\_30\_31 7](#_Toc138015643)

[Image 6 - Table\_1\_MATCP\_Theorical\_7\_Slide30\_31 8](#_Toc138015644)

[Image 7 - Method\_11\_12\_13\_MATCP\_Theorical\_7\_Slide\_12\_13 8](#_Toc138015645)

[Image 8 - Simple Linear for 95% (Area) - Part\_1\_ANOVA 9](#_Toc138015646)

[Image 9 - Simple Linear for 95% (Area) - Part\_2\_ANOVA 9](#_Toc138015647)

[Image 10 -Simple Linear for 99% (Area) - Part\_1\_ANOVA 10](#_Toc138015648)

[Image 11 - Simple Linear for 99% (Area) - Part\_2\_ANOVA 10](#_Toc138015649)

[Image 12 - Simple Linear for 95% (Distance) - Part\_1\_ANOVA 11](#_Toc138015650)

[Image 13 - Simple Linear for 99% (Distance) - Part\_2\_ANOVA 11](#_Toc138015651)

[Image 14 - Simple Linear for 99% (Distance) - Part\_1\_ANOVA 12](#_Toc138015652)

[Image 15 - Simple Linear for 99% (Distance) - Part\_2\_ANOVA 12](#_Toc138015653)

[Image 16 - Simple Linear for 99% (Parking Spaces) - Part\_1\_ANOVA 13](#_Toc138015654)

[Image 17 - Simple Linear for 99% (Parking Spaces) - Part\_2\_ANOVA 13](#_Toc138015655)

[Image 18 - Simple Linear for 95% (Bedrooms) - Part\_1\_ANOVA 14](#_Toc138015656)

[Image 19 - Simple Linear for 95% (Bedrooms) - Part\_2\_ANOVA 14](#_Toc138015657)

[Image 20 - Simple Linear for 99% (Bedrooms) - Part\_1\_ANOVA 15](#_Toc138015658)

[Image 21 - Simple Linear for 99% (Bedrooms) - Part\_2\_ANOVA 15](#_Toc138015659)

[Image 22 - Simple Linear for 95% (Bathrooms) - Part\_1\_ANOVA 16](#_Toc138015660)

[Image 23 - Simple Linear for 95% (Bathrooms) - Part\_2\_ANOVA 16](#_Toc138015661)

[Image 24 - Simple Linear for 99% (Bathrooms) - Part\_1\_ANOVA 17](#_Toc138015662)

[Image 25 - Simple Linear for 99% (Bathrooms) - Part\_2\_ANOVA 17](#_Toc138015663)

[Image 26 - Simple Linear for 95% (Parking Spaces) - Part\_1\_ANOVA 18](#_Toc138015664)

[Image 27 - Simple Linear for 95% (Parking Spaces) - Part\_2\_ANOVA 18](#_Toc138015665)

[Image 28 - Method\_14\_15\_MATCP\_Theorical\_7\_Slide\_23\_24 18](#_Toc138015666)

[Image 29 - Method\_16\_17\_MATCP\_Theorical\_7\_Slide\_23\_24 19](#_Toc138015667)

[Image 30 - Method\_18\_MATCP\_Theorical\_7\_Slide\_22 19](#_Toc138015668)

[Image 31 - Method\_19\_20\_MATCP\_Theorical\_7\_Slide\_23 19](#_Toc138015669)

[Image 32 - Simple Linear for 95% (Area) - Tests 20](#_Toc138015670)

[Image 33 - Simple Linear for 99% (Area) - Tests 21](#_Toc138015671)

[Image 34 - Simple Linear for 95% (Distance) - Tests 22](#_Toc138015672)

[Image 35 - Simple Linear for 99% (Distance) - Tests 23](#_Toc138015673)

[Image 36 - Simple Linear for 95% (Bedrooms) - Tests 23](#_Toc138015674)

[Image 37 - Simple Linear for 99% (Bedrooms) - Tests 24](#_Toc138015675)

[Image 38 - Simple Linear for 95% (Bathrooms) - Tests 24](#_Toc138015676)

[Image 39 - Simple Linear for 99% (Bathrooms) - Tests 25](#_Toc138015677)

[Image 40 - Simple Linear for 95% (Parking Spaces) - Tests 25](#_Toc138015678)

[Image 41 - Simple Linear for 99% (Parking Spaces) - Tests 26](#_Toc138015679)

[Image 42 - Method\_24\_MATCP\_Theorical\_7\_Slide\_21 26](#_Toc138015680)

[Image 43 - Simple Linear for 95% (Area) - Intervals 27](#_Toc138015681)

[Image 44 - Simple Linear for 99% (Area) - Intervals 27](#_Toc138015682)

[Image 45 - Simple Linear for 95% (Distance) - Intervals 27](#_Toc138015683)

[Image 46 - Simple Linear for 99% (Distance) - Intervals 28](#_Toc138015684)

[Image 47 - Simple Linear for 95% (Bedrooms) - Intervals 28](#_Toc138015685)

[Image 48 - Simple Linear for 99% (Bedrooms) - Intervals 28](#_Toc138015686)

[Image 49 - Simple Linear for 95% (Bathrooms) - Intervals 29](#_Toc138015687)

[Image 50 - Simple Linear for 99% (Bathrooms) - Interval 29](#_Toc138015688)

[Image 51 - Simple Linear for 95% (Parking Spaces) - Intervals 29](#_Toc138015689)

[Image 52 - Simple Linear for 99% (Parking Spaces) - Intervals 30](#_Toc138015690)

[Image 53 - Method\_25\_MATCP\_Theorical\_8\_Slide\_4 31](#_Toc138015691)

[Image 54 - Method\_26\_MATCP\_Theorical\_8\_Slide\_6 32](#_Toc138015692)

[Image 55 - Method\_26\_MATCP\_Theorical\_8\_Slide\_14 32](#_Toc138015693)

[Image 56 - Method\_27\_28\_29\_MATCP\_Theorical\_8\_Slide\_14 32](#_Toc138015694)

[Image 57 - Table\_2\_MATCP\_Theorical\_8\_Slide\_16 33](#_Toc138015695)

[Image 166 - Method\_30\_31\_MATCP\_Theorical\_8\_Slide\_17 33](#_Toc138015696)

[Image 59 - Multilinear for 95% - Part\_1\_ANOVA 33](#_Toc138015697)

[Image 60 - Multilinear for 95% - Part\_2\_ANOVA 34](#_Toc138015698)

[Image 61 - Multilinear for 99% - Part\_1\_ANOVA 34](#_Toc138015699)

[Image 62 - Multilinear for 99% - Part\_2\_ANOVA 34](#_Toc138015700)

[Image 63 - Method\_32\_MATCP\_Theorical\_8\_Slide\_21 35](#_Toc138015701)

[Image 64 - Method\_33\_34\_MATCP\_Theorical\_8\_Slide\_18\_21 35](#_Toc138015702)

[Image 65 - Method\_35\_36\_MATCP\_Theorical\_8\_Slide\_21 35](#_Toc138015703)

[Image 66 - Multilinear for 95% - Tests 36](#_Toc138015704)

[Image 67 - Multilinear for 99% - Tests 37](#_Toc138015705)

[Image 68 - Method\_37\_MATCP\_Theorical\_8\_Slide\_18 38](#_Toc138015706)

[Image 69 - Multilinear for 95% - Intervals 39](#_Toc138015707)

[Image 70 - Multilinear for 99% - Intervals 40](#_Toc138015708)

## Simple Linear Regression

### Overview of Simple Linear Regression

In the process of making decisions, it is often necessary to make predictions. When it is possible to establish a relationship between two variables – one, whose values we want to explain (dependent variable), and the other, which is the variable that explains the one mentioned before (independent variable) - the prediction is easier. Simple linear regression is used to estimate the relationship between two variables, more specifically, to establish if there is a statistically significant relationship between the two. Apart from this, it is used when you want to know:

1. How strong the relationship is between the two variables;

2. The value of the dependent variable at a certain value of the independent variables;

### Simple Linear Regression Model

The formula for a simple linear regression is:



Image 1 - Method\_1\_MATCP\_Theorical\_7\_Slide\_5

* 𝑌𝑖 - predicted value of the dependent variable (y) for any given value of the independent variable (x).
* 𝑎̂ – intercept, the predicted value of y when the x is 0.
* 𝑏̂ – regression coefficient – how much we expect y to change as x increases.
* 𝑥𝑖 – independent variable.
* 𝜀𝑖 – error of the estimate or how much variation there is in the estimation of the regression coefficient.

The simple linear regression is applied in tables of value pairs. Each observation is a pair of values, one for each variable. After, is constructed a scatter diagram of the observations.

To estimate the values of 𝑎 and 𝑏 parameters, there is going to be implemented the Minimum Square Method.



Image 2 - Method\_2\_3\_4\_MATCP\_Theorical\_7\_Slide\_10



Image 3 - Method\_5\_6\_MATCP\_Theorical\_7\_Slide\_11

### Model Significance

#### Anova

The Analysis of Variance (ANOVA) consists of calculations that provide information about the levels of variability within a regression model based on the total variation of the Y (dependent variable).



Image 4 - Method\_7\_MATCP\_Theorical\_10\_Slide\_32

* 𝑆𝑇 – total variability of Y observations
* 𝑆𝑅 – part of the variability of Y´s observations that are eliminated when using knowledge of the independent variable to predict Y
* 𝑆𝐸 - part of the variability of observations of Y that remain even knowing the value of x.

Uma imagem com Tipo de letra, texto, file, branco

Descrição gerada automaticamente

Image 5 - Method\_8\_9\_10\_MATCP\_Theorical\_7\_Slide\_30\_31

The calculations are synthesized in the table:

Uma imagem com texto, captura de ecrã, número, file

Descrição gerada automaticamente

Image 6 - Table\_1\_MATCP\_Theorical\_7\_Slide30\_31

#### Coefficient correlation

The correlation coefficients are indicators of the strength of the linear relationship between two different variables, x and y. A linear correlation coefficient that is greater than zero indicates a positive relationship. A value that is less than zero signifies a negative relationship. (Nickolas, 2021)

Uma imagem com Tipo de letra, file, diagrama, branco

Descrição gerada automaticamente

Image 7 - Method\_11\_12\_13\_MATCP\_Theorical\_7\_Slide\_12\_13

Depending on the result of 𝑅 it is possible to reach different conclusions:

* The closer 𝑅 is to zero, the weaker the linear relationship.
* Positive 𝑅 values indicate a positive correlation, where the values of both variables tend to increase together.
* Negative 𝑅 values indicate a negative correlation, where the values of one variable tend to increase when the values of the other variable decrease.

#### Outputs:

##### Simple Linear for 95% (area)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 8 - Simple Linear for 95% (Area) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 9 - Simple Linear for 95% (Area) - Part\_2\_ANOVA

Comparing the F-statistic to the critical value, we find that F-statistic < F de Snedecore, indicating that the regression model is not significant. The correlation coefficient (R) provides information about the strength and direction of the linear relationship between the variables. In this case, R = 0.047, indicating a weak positive correlation.

##### Simple Linear for 99% (area)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 10 -Simple Linear for 99% (Area) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 11 - Simple Linear for 99% (Area) - Part\_2\_ANOVA

Based on the ANOVA table, the MSR is 16909300.512, and the MSE is 15478635,577. The calculated F-value (F0) is 1.092, and the critical F-value (F de Snedecore) at a significance level of 1% is 6.686. Since F0 is less than F de Snedecore, the null hypothesis (H0) means that the regression model is not significant. This implies that the regression model does not provide a significant improvement over the mean.

##### Simple Linear for 95% (distance)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 12 - Simple Linear for 95% (Distance) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 13 - Simple Linear for 99% (Distance) - Part\_2\_ANOVA

Comparing the F-statistic to the critical value, we find that F-statistic > F de Snedecore, indicating that the regression model is significant. The correlation coefficient (R) provides information about the strength and direction of the linear relationship between the variables. In this case, R = 0.757, indicating a strong positive correlation.

##### Simple Linear for 99% (distance)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 14 - Simple Linear for 99% (Distance) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 15 - Simple Linear for 99% (Distance) - Part\_2\_ANOVA

The R-squared value of 0.573 indicates that approximately 57.3% of the variability in the dependent variable (Y) can be explained by the independent variable (X). The correlation coefficient (R) of 0.757 indicates a moderate positive linear relationship between X and Y.

##### Simple Linear for 99% (parking spaces)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 16 - Simple Linear for 99% (Parking Spaces) - Part\_1\_ANOVA

**Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente**

Image 17 - Simple Linear for 99% (Parking Spaces) - Part\_2\_ANOVA

Based on the ANOVA table, the model's F-statistic is 468.462. The critical F-value for a significance level of 1% is 6,686. Since the calculated F-value is greater than the critical F-value, we can reject the null hypothesis and conclude that the regression model is significant.

The correlation coefficient (R) and the coefficient of determination (R^2) can also be used to assess the model's significance. In this case, the correlation coefficient (R) is 0.697, indicating a moderately strong positive linear relationship between X and Y. The coefficient of determination (R^2) is 0.485, which means that 48.5% of the variability in Y can be explained by the linear relationship with X.

To understand the next outputs, we only need to follow the same line of thought that we used in the previous output, so we are only going to show the outputs.

##### Simple Linear for 95% (bedrooms)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 18 - Simple Linear for 95% (Bedrooms) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 19 - Simple Linear for 95% (Bedrooms) - Part\_2\_ANOVA

##### Simple Linear for 99% (bedrooms)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 20 - Simple Linear for 99% (Bedrooms) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 21 - Simple Linear for 99% (Bedrooms) - Part\_2\_ANOVA

##### Simple Linear for 95% (bathrooms)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 22 - Simple Linear for 95% (Bathrooms) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 23 - Simple Linear for 95% (Bathrooms) - Part\_2\_ANOVA

##### Simple Linear for 99% (bathrooms)

Uma imagem com texto, Tipo de letra, captura de ecrã, número

Descrição gerada automaticamente

Image 24 - Simple Linear for 99% (Bathrooms) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 25 - Simple Linear for 99% (Bathrooms) - Part\_2\_ANOVA

##### Simple Linear for 95% (parking spaces)

Uma imagem com texto, Tipo de letra, captura de ecrã, tipografia

Descrição gerada automaticamente

Image 26 - Simple Linear for 95% (Parking Spaces) - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 27 - Simple Linear for 95% (Parking Spaces) - Part\_2\_ANOVA

#### Hypothesis tests for model coefficients

The hypothesis tests are a statistical procedure that is used to claim about the data of the table of value. The hypothesis tests for model coefficients 𝑎̂ and 𝑏̂ follow a five-step procedure (Kumar, 2022):

* Formulate null and alternate hypotheses:



Image 28 - Method\_14\_15\_MATCP\_Theorical\_7\_Slide\_23\_24

* Determine the test statistics:

A picture containing screenshot, line, diagram, white

Description automatically generated

Image 29 - Method\_16\_17\_MATCP\_Theorical\_7\_Slide\_23\_24

Uma imagem com Tipo de letra, número, file, branco

Descrição gerada automaticamente

Image 30 - Method\_18\_MATCP\_Theorical\_7\_Slide\_22

* Determine the critical region:

A picture containing text, font, line, white

Description automatically generated

Image - Method\_19\_20\_MATCP\_Theorical\_7\_Slide\_23

* Calculate the statistics.
* Make decisions:
  + with a significance level of 𝛼, reject 𝐻0 when

#### Outputs:

##### Simple Linear for 95% (area)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 32 - Simple Linear for 95% (Area) - Tests

#### Intercept Tests:

t-value : 26.941

Critical t-value (tc) : 1.965

|t-value| > tc, leading to the rejection of the null hypothesis. Thus, the intercept is significant.

#### Slope Tests:

t-value : -1.045

Critical t-value (tc) : 1.965

|t-value| <= tc, leading to the acceptance of the null hypothesis. Thus, the slope is not significant.

##### Simple Linear for 99% (area)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 33 - Simple Linear for 99% (Area) - Tests

#### Intercept Tests:

The standard error (SE) of the intercept is 348,058.

The calculated t-value is 26.941, and the critical t-value (tc) at a significance level of 1% is 2,586. Since |t| is greater than tc, the null hypothesis (H0) is rejected. This indicates that the intercept is significantly different from zero.

#### Slope Tests:

The standard error (SE) of the slope is 8,408.

The calculated t-value is -1.045, and the critical t-value (tc) at a significance level of 1% is 2.586. Since |t| is less than or equal to tc, the null hypothesis (H0) is accepted. This suggests that the slope is not significantly different from zero.

##### Simple Linear for 95% (distance)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 34 - Simple Linear for 95% (Distance) - Tests

#### Intercept Tests:

The standard error (SE) of the intercept is 370,576.

The calculated t-value is -0.074, and the critical t-value (tc) at a significance level of 5% is 1,965. Since the absolute value of the calculated t-value (|t|) is less than the critical t-value (|t| < tc), we fail to reject the null hypothesis (H0). This indicates that there is not enough evidence to suggest that the intercept is significantly different from zero.

#### Slope Tests:

The standard error (SE) of the slope is 0,232.

The calculated t-value is 25.811, and the critical t-value (tc) at a significance level of 1% is 1,965. Since the absolute value of the calculated t-value (|t|) is greater than the critical t-value (|t| > tc), we reject the null hypothesis (H0). This suggests that the slope is significantly different from zero.

To understand the next outputs, we only need to follow the same line of thought that we used in the previous output, so we are only going to show the outputs.

##### Simple Linear for 99% (distance)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 35 - Simple Linear for 99% (Distance) - Tests

##### Simple Linear for 95% (bedrooms)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 36 - Simple Linear for 95% (Bedrooms) - Tests

##### Simple Linear for 99% (bedrooms)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 37 - Simple Linear for 99% (Bedrooms) - Tests

##### Simple Linear for 95% (bathrooms)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 38 - Simple Linear for 95% (Bathrooms) - Tests

##### Simple Linear for 99% (bathrooms)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 39 - Simple Linear for 99% (Bathrooms) - Tests

##### Simple Linear for 95% (parking spaces)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 40 - Simple Linear for 95% (Parking Spaces) - Tests

##### Simple Linear for 99% (parking spaces)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 41 - Simple Linear for 99% (Parking Spaces) - Tests

### Confidence intervals for prediction values

The confidence intervals for prediction values can be used to find the average answer of a variable of their true value. The confidence interval at (1 − 𝛼) x 100% for the parameters are given by:

A picture containing font, line, text, diagram

Description automatically generated

Image 42 - Method\_24\_MATCP\_Theorical\_7\_Slide\_21

#### Outputs:

##### Simple Linear for 95% (area)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 43 - Simple Linear for 95% (Area) - Intervals

##### Simple Linear for 99% (area)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 44 - Simple Linear for 99% (Area) - Intervals

##### Simple Linear for 95% (distance)

Uma imagem com texto, captura de ecrã, Tipo de letra, file

Descrição gerada automaticamente

Image 45 - Simple Linear for 95% (Distance) - Intervals

##### Simple Linear for 99% (distance)

Uma imagem com texto, captura de ecrã, Tipo de letra, file

Descrição gerada automaticamente

Image 46 - Simple Linear for 99% (Distance) - Intervals

##### Simple Linear for 95% (bedrooms)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 47 - Simple Linear for 95% (Bedrooms) - Intervals

##### Simple Linear for 99% (bedrooms)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 48 - Simple Linear for 99% (Bedrooms) - Intervals

##### Simple Linear for 95% (bathrooms)

Uma imagem com texto, captura de ecrã, Tipo de letra, file

Descrição gerada automaticamente

Image 49 - Simple Linear for 95% (Bathrooms) - Intervals

##### Simple Linear for 99% (bathrooms)

Uma imagem com texto, captura de ecrã, Tipo de letra, file

Descrição gerada automaticamente

Image 50 - Simple Linear for 99% (Bathrooms) - Interval

##### Simple Linear for 95% (parking spaces)

Uma imagem com texto, captura de ecrã, Tipo de letra, file

Descrição gerada automaticamente

Image 51 - Simple Linear for 95% (Parking Spaces) - Intervals

##### Simple Linear for 99% (parking spaces)

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 52 - Simple Linear for 99% (Parking Spaces) - Intervals

## Multi Linear Regression

### Overview of Multi Linear Regression

In the regression analysis, we found situations with more than one independent variable. This regression model takes the multiple regression model (RLM) name. The dependent variable Y may be related to k independent variables.

Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable. So, multiple linear regression is used when you want to know:

1. How strong the relationship is between the two or more independent variables and one dependent variable.
2. The value of the dependent variable at a certain value of the independent variables (Bevans, 2020b)

### Multiple Linear Regression Model

The formula for a multiple linear regression is (Bevans, 2020b):



Image 53 - Method\_25\_MATCP\_Theorical\_8\_Slide\_4

* 𝑌 - predicted value of the dependent variable.
* 𝛽0 – y-intercept, the predicted value of y when all other parameters are set to 0.
* 𝛽1𝑋1– regression coefficient.
* 𝑋1, 𝑋𝑛– independent variable.
* cursive 𝜀𝑖 – model error of the estimate or how much variation there is in our estimate of Y.

The model of the multiple linear regression presented is a system of 𝑛 equations with the matrix representation:

Uma imagem com texto, Tipo de letra, diagrama, file

Descrição gerada automaticamente

Image 54 - Method\_26\_MATCP\_Theorical\_8\_Slide\_6

However, to obtain the matrix representation it is necessary to calculate the following matrix of X and Y variables:

### Model Significance

#### Anova

The Anova calculations for multiple regression are nearly identical to the calculations for simple linear regression, except that the degrees of freedom are adjusted to reflect the number of independent variables in the model.



Image 55 - Method\_26\_MATCP\_Theorical\_8\_Slide\_14

* 𝑆𝑄𝑇 - measures the total variation of observations around the mean.
* 𝑆𝑄𝑅 - measures the variation of the dependent variable.
* 𝑆𝑄𝐸 - measures the variation of the independent variable.



Image 56 - Method\_27\_28\_29\_MATCP\_Theorical\_8\_Slide\_14

The calculations are synthesized in the table:

Uma imagem com texto, captura de ecrã, número, Tipo de letra

Descrição gerada automaticamente

Image 57 - Table\_2\_MATCP\_Theorical\_8\_Slide\_16

### Coefficient determination

The coefficient determination is a measure of the proportion of change in response variable Y that is explained by the regression equation.

Uma imagem com Tipo de letra, file, branco, diagrama

Descrição gerada automaticamente

Image 166 - Method\_30\_31\_MATCP\_Theorical\_8\_Slide\_17

#### Outputs:

##### Multilinear for 95%:

A picture containing text, font, white, algebra

Description automatically generated

Image 59 - Multilinear for 95% - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 60 - Multilinear for 95% - Part\_2\_ANOVA

The mean squares for regression (MQR) is 1149222424,1618, and the mean squares for error (MQE) is 3983121,8289. The F-statistic is calculated as 288.5230, and the critical F-value at a significance level of 0.05 is 2,1170. Since the calculated F-statistic is greater than the critical F-value, we reject the null hypothesis (H0) and conclude that the regression model is significant.

##### Multilinear for 99%:

Uma imagem com texto, Tipo de letra, branco, algebra

Descrição gerada automaticamente

Image 61 - Multilinear for 99% - Part\_1\_ANOVA

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 62 - Multilinear for 99% - Part\_2\_ANOVA

To understand the values on this output, we only need to follow the same line of thought that we used in the previous output.

### Hypothesis tests for model coefficients

The hypothesis tests are a statistical procedure that is used to claim the data of the table of values. The hypothesis tests for all the regression coefficients follow a five-step procedure (Kumar, 2022):

* Formulate null and alternate hypotheses:



Image 63 - Method\_32\_MATCP\_Theorical\_8\_Slide\_21

* Determine the test statistics:



Image 64 - Method\_33\_34\_MATCP\_Theorical\_8\_Slide\_18\_21

𝐶𝑗𝑗 − 𝑒𝑙𝑒𝑚𝑒𝑛𝑡 𝑗 𝑜𝑓 𝑡ℎ𝑒 𝑚𝑎𝑖𝑛 𝑑𝑖𝑎𝑔𝑜𝑛𝑎𝑙 𝑜𝑓 𝑚𝑎𝑡𝑟𝑖𝑧 𝐶 =

* Determine the critical region:



Image 65 - Method\_35\_36\_MATCP\_Theorical\_8\_Slide\_21

* Calculate the statistics.
* Make decisions:
  + The rejection of 𝐻0 allows us to conclude that the regressor 𝑥𝑗 has explanatory power. The non-rejection of 𝐻0 allows us to conclude that the regressor 𝑥𝑗 can be “deleted”.

#### Outputs:

##### Multilinear for 95%:

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 66 - Multilinear for 95% - Tests

For each parameter, the null hypothesis (H0) states that the coefficient is equal to zero, and the alternative hypothesis (H1) states that the coefficient is not equal to zero. The t-statistic is calculated for each parameter, and the critical t-value at a significance level of 0.05 is 1,9647. If the absolute value of the observed t-statistic is greater than the critical t-value, we reject the null hypothesis. Based on the hypothesis tests, all the parameters in the model are significant at a 5% significance level.

##### Multilinear for 99%:

Uma imagem com texto, captura de ecrã, Tipo de letra, documento

Descrição gerada automaticamente

Image 67 - Multilinear for 99% - Tests

To test the significance of each individual independent variable, hypothesis tests are performed. The significance level is set at 0.01 (1%). The results of the tests are as follows:

* Parameter 0: The coefficient is 809,2582. The observed t-value is 1.9800, which is less than the critical t-value of 2.5858. Therefore, the null hypothesis (H0: B = 0) is accepted, indicating that the parameter is not statistically significant.
* Parameter 1: The coefficient is -11,0780. The observed t-value is -2.5921, which is greater than the critical t-value of -2.5858. Thus, the null hypothesis is rejected, indicating that the parameter is statistically significant.
* Parameter 2: The coefficient is 4,8278. The observed t-value is 16.9583, which is greater than the critical t-value of 2.5858. Hence, the null hypothesis is rejected, indicating that the parameter is statistically significant.
* Parameter 3: The coefficient is -1072,8759. The observed t-value is -7.9843, which is greater than the critical t-value of -2.5858. Therefore, the null hypothesis is rejected, suggesting that the parameter is statistically significant.
* Parameter 4: The coefficient is 592,7870. The observed t-value is 4.3914, which is greater than the critical t-value of 2.5858. Thus, the null hypothesis is rejected, indicating that the parameter is statistically significant.
* Parameter 5: The coefficient is 1695,6582. The observed t-value is 11.4305, which is greater than the critical t-value of 2.5858. Hence, the null hypothesis is rejected, suggesting that the parameter is statistically significant.

### Confidence intervals for prediction values

The confidence intervals for prediction values can be used to find the average answer of a variable of their true value. The confidence interval at (1 − 𝛼) x 100% for all regression coefficients are given by:

Uma imagem com Tipo de letra, tipografia, escrita à mão, caligrafia

Descrição gerada automaticamente

Image 68 - Method\_37\_MATCP\_Theorical\_8\_Slide\_18

### 

#### Outputs:

##### Multilinear for 95%:

Uma imagem com texto, captura de ecrã, Tipo de letra, número

Descrição gerada automaticamente

Image 69 - Multilinear for 95% - Intervals

##### Multilinear for 99%:

Uma imagem com texto, captura de ecrã, Tipo de letra, documento

Descrição gerada automaticamente

Image 70 - Multilinear for 99% - Intervals