USER MANUAL

MATCP - INTEGRATIVE PROJECT

PREPARED FOR:
LUÍS AFONSO (TPA)

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

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

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

Table of Contents

Table of Images

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:

$$Y_i = \hat{a} + \hat{b}x_i + \varepsilon_i$$

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).
- $\hat{a}$ – intercept, the predicted value of y when the x is 0.
- $\hat{b}$ – 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.

$$S_{xx} = \sum_{i=1}^{n} x_i^{\,2} - n\bar{x}^2 \qquad S_{yy} = \sum_{i=1}^{n} y_i^{\,2} - n\bar{y}^2 \qquad S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

Image 2 - Method_2_3_4_MATCP_Theorical_7_Slide_10

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \qquad \hat{b} = \frac{S_{xy}}{S_{xx}}$$

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).

$$ST = SR + SE$$

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.

$$ST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \qquad SR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \qquad SE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Image 5 - Method_8_9_10_MATCP_Theorical_7_Slide_30_31

The calculations are synthesized in the table:

Source of variation	Sum of squares	Degrees of freedom	Root mean square	F test stats
Regression	SR	1	$MSR = \frac{SR}{1}$	$F = \frac{MSR}{MSE}$
Error	SE	$n-2$	$MSE = \frac{SE}{n-2}$	
Total	ST	$n-1$		

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)

$$R = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \qquad R^2 = \frac{S^2{}_{xy}}{S_{xx}S_{yy}} \qquad R^2_{adj} = \frac{(1-R^2)(n-1)}{n-2}$$

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)

```
n = 499
Sxx= 218955,695
Syy= 7709791182,451
Sxy= -1924158,947
SE= 7692881881,939
SR= 16909300,512
ST= 7709791182,451
avgX= 35,705
avgY= 9063,177
slope= -8,788
intercept= 9376,952
R^2 = 0,002
R =0,047
```

Image 8 - Simple Linear for 95% (Area) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:16909300,512
 MSE:15478635,577
 F0 :1,092
 F de Snedecore : 3,86
-=-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-=-
F0 < F de Snedecor
H0 is accepted -> regression model is not significant
```

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)

```
n = 499
Sxx= 218955,695
Syy= 7709791182,451
Sxy= -1924158,947
SE= 7692881881,939
SR= 16909300,512
ST= 7709791182,451
avgX= 35,705
avgY= 9063,177
slope= -8,788
intercept= 9376,952
R^2 = 0,002
R =0,047
```

Image 10 -Simple Linear for 99% (Area) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:16909300,512
 MSE:15478635,577
 F0 :1,092
 F de Snedecore : 6,686
-=-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-=-
F0 < F de Snedecor
H0 is accepted -> regression model is not significant
```

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)

```
n = 499
Sxx= 123027159,415
Syy= 7709791182,451
Sxy= 737049650,331
SE= 3294163042,797
SR= 4415628139,653
ST= 7709791182,451
avgX= 1517,379
avgY= 9063,177
slope= 5,991
intercept= -27,365
R^2 = 0,573
R =0,757
```

Image 12 - Simple Linear for 95% (Distance) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:4415628139,653
 MSE:6628094,654
 F0 :666,199
 F de Snedecore : 3,86
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
```

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)

```
n = 499
Sxx= 123027159,415
Syy= 7709791182,451
Sxy= 737049650,331
SE= 3294163042,797
SR= 4415628139,653
ST= 7709791182,451
avgX= 1517,379
avgY= 9063,177
slope= 5,991
intercept= -27,365
R^2 = 0,573
R =0,757
```

Image 14 - Simple Linear for 99% (Distance) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:4415628139,653
 MSE:6628094,654
 F0 :666,199
 F de Snedecore : 6,686
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
```

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)

```
n = 499
Sxx= 285,082
Syy= 7709791182,451
Sxy= 1032704,155
SE= 3968842062,444
SR= 3740949120,007
ST= 7709791182,451
avgX= 1,735
avgY= 9063,177
slope= 3622,479
intercept= 2776,47
R^2 = 0,485
R =0,697
```

Image 16 - Simple Linear for 99% (Parking Spaces) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:3740949120,007
 MSE:7985597,711
 F0 :468,462
 F de Snedecore : 6,686
-=-=-=-.-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-.-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
```

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.

<u>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.</u>

Simple Linear for 95% (bedrooms)

```
n = 499
Sxx= 317,852
Syy= 7709791182,451
Sxy= 259482,401
SE= 7497959326,261
SR= 211831856,19
ST= 7709791182,451
avgX= 2,838
avgY= 9063,177
slope= 816,363
intercept= 6746,603
R^2 = 0,027
R =0,166
```

Image 18 - Simple Linear for 95% (Bedrooms) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:211831856,19
 MSE:15086437,276
 F0 :14,041
 F de Snedecore : 3,86
-=-=-=--=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=--=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
```

Image 19 - Simple Linear for 95% (Bedrooms) - Part_2_ANOVA

Simple Linear for 99% (bedrooms)

```
n = 499
Sxx= 317,852
Syy= 7709791182,451
Sxy= 259482,401
SE= 7497959326,261
SR= 211831856,19
ST= 7709791182,451
avgX= 2,838
avgY= 9063,177
slope= 816,363
intercept= 6746,603
R^2 = 0,027
R =0,166
```

Image 20 - Simple Linear for 99% (Bedrooms) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:211831856,19
 MSE:15086437,276
 F0 :14,041
 F de Snedecore : 6,686
-=-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
```

Image 21 - Simple Linear for 99% (Bedrooms) - Part_2_ANOVA

Simple Linear for 95% (bathrooms)

```
n = 499
Sxx= 422,228
Syy= 7709791182,451
Sxy= 1170290,286
SE= 4466098504,429
SR= 3243692678,022
ST= 7709791182,451
avgX= 2,405
avgY= 9063,177
slope= 2771,699
intercept= 2397,768
R^2 = 0,421
R =0,649
```

Image 22 - Simple Linear for 95% (Bathrooms) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:3243692678,022
 MSE:8986113,691
 F0 :360,967
 F de Snedecore : 3,86
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
```

Image 23 - Simple Linear for 95% (Bathrooms) - Part_2_ANOVA

Simple Linear for 99% (bathrooms)

```
n = 499
Sxx= 422,228
Syy= 7709791182,451
Sxy= 1170290,286
SE= 4466098504,429
SR= 3243692678,022
ST= 7709791182,451
avgX= 2,405
avgY= 9063,177
slope= 2771,699
intercept= 2397,768
R^2 = 0,421
R =0,649
```

Image 24 - Simple Linear for 99% (Bathrooms) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:3243692678,022
 MSE:8986113,691
 F0 :360,967
 F de Snedecore : 6,686
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
```

Image 25 - Simple Linear for 99% (Bathrooms) - Part_2_ANOVA

Simple Linear for 95% (parking spaces)

```
n = 499
Sxx= 285,082
Syy= 7709791182,451
Sxy= 1032704,155
SE= 3968842062,444
SR= 3740949120,007
ST= 7709791182,451
avgX= 1,735
avgY= 9063,177
slope= 3622,479
intercept= 2776,47
R^2 = 0,485
R =0,697
```

Image 26 - Simple Linear for 95% (Parking Spaces) - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MSR:3740949120,007
 MSE:7985597,711
 F0 :468,462
 F de Snedecore : 3,86
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
```

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 $\hat{a}$ and $\hat{b}$ follow a five-step procedure (Kumar, 2022):

- Formulate null and alternate hypotheses:

$$H_0: a = 0 \ v.s. \ H_1: a \neq 0 \qquad\qquad H_0: b = 0 \ v.s. \ H_1: b \neq 0$$

Image 28 - Method_14_15_MATCP_Theorical_7_Slide_23_24

- Determine the test statistics:

$$T_a = \frac{\hat{a}-a}{s\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}}} \sim t_{n-2} \qquad T_b = \frac{\hat{b}-b}{S/\sqrt{S_{xx}}} \sim t_{n-2}$$

Image 29 - Method_16_17_MATCP_Theorical_7_Slide_23_24

$$s = \sqrt{\frac{1}{n \cdot 2}\sum_{i=1}(y_i - \hat{y}_i)^2}$$

Image 30 - Method_18_MATCP_Theorical_7_Slide_22

- Determine the critical region:

$$t_c = t_{1-\frac{\alpha}{2}}(n-2) \qquad RC =]-\infty, -t_c[U]t_c, +\infty[$$

Image 31 - Method_19_20_MATCP_Theorical_7_Slide_23

- Calculate the statistics.

- Make decisions:
  - with a significance level of α, reject $H0$ when $|t| > t_{1-\frac{\alpha}{2}}(n-2)$

Outputs:

Simple Linear for 95% (area)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :3934,29
-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-
 t = 26,941
 tc =1,965
|t| > tc
 -> H0 rejected

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :3930,338
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
 t = -1,045
 tc =1,965
|t| <= tc
 -> H0 accepted
```

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)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :3934,29
-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-
 t = 26,941
 tc =2,586
|t| > tc
 -> H0 rejected

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :3930,338
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
 t = -1,045
 tc =2,586
|t| <= tc
 -> H0 accepted
```

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)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :2574,509
-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-
 t = -0,074
 tc =1,965
|t| <= tc
 -> H0 accepted

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :2571,922
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
 t = 25,811
 tc =1,965
|t| > tc
 -> H0 rejected
```

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)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :2574,509
-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-
 t = -0,074
 tc =2,586
|t| <= tc
 -> H0 accepted

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :2571,922
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
 t = 25,811
 tc =2,586
|t| > tc
 -> H0 rejected
```

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

Simple Linear for 95% (bedrooms)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :3884,126
-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-
 t = 10,505
 tc =1,965
|t| > tc
 -> H0 rejected

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :3880,225
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
 t = 3,747
 tc =1,965
|t| > tc
 -> H0 rejected
```

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

Simple Linear for 99% (bedrooms)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :3884,126
-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-
 t = 10,505
 tc =2,586
|t| > tc
 -> H0 rejected

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :3880,225
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
 t = 3,747
 tc =2,586
|t| > tc
 -> H0 rejected
```

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

Simple Linear for 95% (bathrooms)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :2997,685
-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-
 t = 6,384
 tc =1,965
|t| > tc
 -> H0 rejected

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :2994,673
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
 t = 18,999
 tc =1,965
|t| > tc
 -> H0 rejected
```

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

Simple Linear for 99% (bathrooms)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :2997,685
-=-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-=-
 t = 6,384
 tc =2,586
|t| > tc
 -> H0 rejected

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :2994,673
-=-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-=-
 t = 18,999
 tc =2,586
|t| > tc
 -> H0 rejected
```

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

Simple Linear for 95% (parking spaces)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :2825,88
-=-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-=-
 t = 8,764
 tc =1,965
|t| > tc
 -> H0 rejected

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :2823,041
-=-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-=-
 t = 21,644
 tc =1,965
|t| > tc
 -> H0 rejected
```

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

Simple Linear for 99% (parking spaces)

```
||-=-=-=- Intercept Hypothesis Test -=-=-=-||
 s :2825,88
-=-=-=-=-=-=-
H0 : a = a0
H1 : a != a0
-=-=-=-=-=-=-
 t = 8,764
 tc =2,586
|t| > tc
 -> H0 rejected

||-=-=-=- Slope Hypothesis Test -=-=-=-||
 s :2823,041
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
 t = 21,644
 tc =2,586
|t| > tc
 -> H0 rejected
```

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 $(1 - \alpha)$ x 100% for the parameters are given by:

$$]\hat{a} - t_c S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}, \hat{a} + t_c S\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}[\qquad]\hat{b} - t_c S\sqrt{\frac{1}{S_{xx}}}, \hat{b} + t_c S\sqrt{\frac{1}{S_{xx}}}[$$

Image 42 - Method_24_MATCP_Theorical_7_Slide_21

Outputs:

Simple Linear for 95% (area)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: 9376,952
 Intercept Standard Error: 348,058
 Intercept Confidence Interval (95.0) -> ] 8693,109; 10060,794[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : -8,788
 Slope Standard Error: 8,408
 Slope Confidence Interval (95.0) -> ] -25,307; 7,732[
```

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

Simple Linear for 99% (area)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: 9376,952
 Intercept Standard Error: 348,058
 Intercept Confidence Interval (99.0) -> ] 8476,966; 10276,938[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : -8,788
 Slope Standard Error: 8,408
 Slope Confidence Interval (99.0) -> ] -30,529; 12,953[
```

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

Simple Linear for 95% (distance)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: -27,365
 Intercept Standard Error: 370,576
 Intercept Confidence Interval (95.0) -> ] -755,45; 700,721[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : 5,991
 Slope Standard Error: 0,232
 Slope Confidence Interval (95.0) -> ] 5,535; 6,447[
```

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

Simple Linear for 99% (distance)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: -27,365
 Intercept Standard Error: 370,576
 Intercept Confidence Interval (99.0) -> ] -985,578; 930,848[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : 5,991
 Slope Standard Error: 0,232
 Slope Confidence Interval (99.0) -> ] 5,391; 6,591[
```

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

Simple Linear for 95% (bedrooms)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: 6746,603
 Intercept Standard Error: 642,208
 Intercept Confidence Interval (95.0) -> ] 5484,833; 8008,373[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : 816,363
 Slope Standard Error: 217,862
 Slope Confidence Interval (95.0) -> ] 388,319; 1244,407[
```

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

Simple Linear for 99% (bedrooms)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: 6746,603
 Intercept Standard Error: 642,208
 Intercept Confidence Interval (99.0) -> ] 5086,023; 8407,183[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : 816,363
 Slope Standard Error: 217,862
 Slope Confidence Interval (99.0) -> ] 253,025; 1379,701[
```

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

Simple Linear for 95% (bathrooms)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: 2397,768
 Intercept Standard Error: 375,617
 Intercept Confidence Interval (95.0) -> ] 1659,779; 3135,756[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : 2771,699
 Slope Standard Error: 145,886
 Slope Confidence Interval (95.0) -> ] 2485,071; 3058,328[
```

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

Simple Linear for 99% (bathrooms)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: 2397,768
 Intercept Standard Error: 375,617
 Intercept Confidence Interval (99.0) -> ] 1426,522; 3369,013[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : 2771,699
 Slope Standard Error: 145,886
 Slope Confidence Interval (99.0) -> ] 2394,475; 3148,924[
```

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

Simple Linear for 95% (parking spaces)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: 2776,47
 Intercept Standard Error: 316,812
 Intercept Confidence Interval (95.0) -> ] 2154,016; 3398,923[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : 3622,479
 Slope Standard Error: 167,367
 Slope Confidence Interval (95.0) -> ] 3293,646; 3951,312[
```

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

Simple Linear for 99% (parking spaces)

```
||-=-=-=- Intercept Confidence Interval -=-=-=-||
 Intercept: 2776,47
 Intercept Standard Error: 316,812
 Intercept Confidence Interval (99.0) -> ] 1957,276; 3595,663[

||-=-=-=- Slope Confidence Interval -=-=-=-||
 Slope : 3622,479
 Slope Standard Error: 167,367
 Slope Confidence Interval (99.0) -> ] 3189,71; 4055,248[
```

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):

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon$$

Image 53 - Method_25_MATCP_Theorical_8_Slide_4

- Y - predicted value of the dependent variable.
- $\beta 0$ – y-intercept, the predicted value of y when all other parameters are set to 0.
- $\beta 1 X 1$– regression coefficient.
- $X 1$, $X n$– 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:

$$Y = X\beta + \varepsilon$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{1k} \\ \dots & \dots & \dots \\ 1 & x_{n1} & x_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix}$$

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: $X, X^T, X^TX, (X^TX)^{-1}, Y, Y^T, Y^TY, X^TY, X^TX.X^TY = \widehat{\beta}, \beta^T$

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.

$$SQ_T = SQ_R + SQ_E$$

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.

$$SQ_T = Y^TY - n\overline{y}^2 \qquad SQ_R = \beta^TX^TY - n\overline{y}^2 \qquad SQ_E = Y^TY - \beta^TX^TY$$

Image 56 - Method_27_28_29_MATCP_Theorical_8_Slide_14

The calculations are synthesized in the table:

Source of variation	Sum of squares	Degrees of freedom	Root mean square	F test stats
Regression	SQ_R	k	$MQ_R = \frac{SR}{k}$	$F = \frac{MSR}{MSE}$
Error	SQ_E	$n-(k+1)$	$MQ_E = \frac{SE}{n-(k+1)}$	
Total	SQ_T	$n-1$		

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.

$$R^2 = \frac{SQ_R}{SQ_T} = 1 - \frac{SQ_E}{SQ_T} \qquad R^2_{adj} = \frac{(1-R^2)(n-1)}{n-(k+1)}$$

Image 58 - Method_30_31_MATCP_Theorical_8_Slide_17

Outputs:

Multilinear for 95%:

```
SIGNIFICANCE_LEVEL=0.05
n=499,0000
SQT=7709791182,4507
SQR=5746112120,8091
SQE=1963679061,6416
R Squared =0,7453
R Squared Adjusted =0,7427
```

Image 59 - Multilinear for 95% - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MQR:1149222424,1618
 MQE:3983121,8289
 F0 :288,5230
 F de Snedecor : 2,1170
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
```

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%:

```
SIGNIFICANCE_LEVEL=0.01
n=499,0000
SQT=7709791182,4507
SQR=5746112120,8091
SQE=1963679061,6416
R Squared =0,7453
R Squared Adjusted =0,7427
```

Image 61 - Multilinear for 99% - Part_1_ANOVA

```
||-=-=-=- Significance Model With ANOVA -=-=-=-||
 MQR:1149222424,1618
 MQE:3983121,8289
 F0 :288,5230
 F de Snedecor : 2,8386
-=-=-=-=-=-=-
H0 : b = b0
H1 : b != b0
-=-=-=-=-=-=-
F0 > F de Snedecor
H0 is rejected -> regression model is significant
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
```

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:

$$H_0: \beta_j = 0 \;\; v.s. \;\; \beta_j: a \neq 0$$

Image 63 - Method_32_MATCP_Theorical_8_Slide_21

- Determine the test statistics:

$$T_0 = \frac{\widehat{\beta_j}}{\sqrt{\hat{\sigma}^2 C_{jj}}} \qquad \hat{\sigma}^2 = \frac{SQ_E}{n-(k+1)} = MQ_E$$

Image 64 - Method_33_34_MATCP_Theorical_8_Slide_18_21

$$Cjj - element\ j\ of\ the\ main\ diagonal\ of\ matriz\ C = (X^T X)^{-1}$$

- Determine the critical region:

$$t_c = t_{1-\frac{\alpha}{2}}[n-(k+1)] \qquad RC =]-\infty, -t_c[U]t_c, +\infty[$$

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%:

```
-=-=-=-=-=- Hypothesis Tests (95.0%) -=-=-=-=-=-
Test : H0 : B = 0
       H1 : B =|= 0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 0: 809,2582
observed t -> 1,9800
tc ->1,9647
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 1: -11,0780
observed t -> -2,5921
tc ->1,9647
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 2: 4,8278
observed t -> 16,9583
tc ->1,9647
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 3: -1072,8759
observed t -> -7,9843
tc ->1,9647
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 4: 592,7870
observed t -> 4,3914
tc ->1,9647
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 5: 1695,6582
observed t -> 11,4305
tc ->1,9647
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
```

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%:

```
-=-=-=-=-=- Hypothesis Tests (99.0%) -=-=-=-=-=-
Test : H0 : B = 0
        H1 : B =|= 0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 0: 809,2582
observed t -> 1,9800
tc ->2,5858
|observed t| <= tc, Accepts H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 1: -11,0780
observed t -> -2,5921
tc ->2,5858
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 2: 4,8278
observed t -> 16,9583
tc ->2,5858
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 3: -1072,8759
observed t -> -7,9843
tc ->2,5858
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 4: 592,7870
observed t -> 4,3914
tc ->2,5858
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
Parameter 5: 1695,6582
observed t -> 11,4305
tc ->2,5858
|observed t| > tc, Rejects H0
-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
```

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:

$$]\hat{\beta}_j - t_c S\sqrt{\hat{\sigma}^2 C_{jj}}, \hat{\beta}_j + t_c S\sqrt{\hat{\sigma}^2 C_{jj}}[$$

Image 68 - Method_37_MATCP_Theorical_8_Slide_18

Outputs:

Multilinear for 95%:

```
[-=-=-=-=-=-= Confidence Intervals (95.0%) -=-=-=-=-=-=]



Parameter 0 Confidence Interval (95.0%): ]6,2313, 1612,2851[
Parameter 0 =809,2582
Standard Error: 408,7174
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 1 Confidence Interval (95.0%): ]-19,4748, -2,6812[
Parameter 1 =-11,0780
Standard Error: 4,2737
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 2 Confidence Interval (95.0%): ]4,2685, 5,3872[
Parameter 2 =4,8278
Standard Error: 0,2847
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 3 Confidence Interval (95.0%): ]-1336,8855, -808,8663[
Parameter 3 =-1072,8759
Standard Error: 134,3732
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 4 Confidence Interval (95.0%): ]327,5682, 858,0057[
Parameter 4 =592,7870
Standard Error: 134,9887
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 5 Confidence Interval (95.0%): ]1404,1978, 1987,1185[
Parameter 5 =1695,6582
Standard Error: 148,3448
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
```

Image 69 - Multilinear for 95% - Intervals

Multilinear for 99%:

```
[-=-=-=-=-=-= Confidence Intervals (99.0%) -=-=-=-=-=-=]



Parameter 0 Confidence Interval (99.0%): ]-247,5859, 1866,1023[
Parameter 0 =809,2582
Standard Error: 408,7174
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 1 Confidence Interval (99.0%): ]-22,1288, -0,0271[
Parameter 1 =-11,0780
Standard Error: 4,2737
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 2 Confidence Interval (99.0%): ]4,0917, 5,5640[
Parameter 2 =4,8278
Standard Error: 0,2847
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 3 Confidence Interval (99.0%): ]-1420,3325, -725,4193[
Parameter 3 =-1072,8759
Standard Error: 134,3732
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 4 Confidence Interval (99.0%): ]243,7390, 941,8349[
Parameter 4 =592,7870
Standard Error: 134,9887
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=

Parameter 5 Confidence Interval (99.0%): ]1312,0743, 2079,2420[
Parameter 5 =1695,6582
Standard Error: 148,3448
 -=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=-=
```

Image 70 - Multilinear for 99% - Intervals