

# Regressão linear

Passos práticos

Excel & R

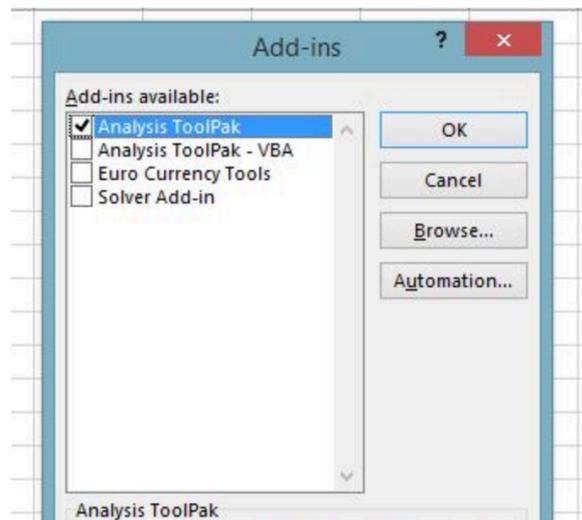
# Excel

- **REGRESSION IN EXCEL**

## Data Analysis Toolpak

The Data Analysis ToolPak is an Excel add-in that provides [data analysis](#) tools for financial, statistical, and engineering data analysis

- Click the File tab, click Options, and then click the Add-Ins category.
- Select Analysis ToolPak and click on the Go button.
- Check Analysis ToolPak and click on OK.



# Excel (1)

- Go to the Data tab > Analysis group > Data analysis.
- Select Regression and click OK.
- Select the Input Y Range as the Dependent Variable of Interest and Input X Range as Independent Variable of Interest. Check the residuals and click OK.

# Excel (2)

- **Interpret Regression Analysis Output**
- **Summary Output**
- The summary output tells you how well the calculated linear regression equation fits your data source.
- The Multiple R is the Correlation Coefficient that measures the strength of a linear relationship between two variables. The larger the absolute value, the stronger is the relationship.
  - 1 means a strong positive relationship
  - -1 means a strong negative relationship
  - 0 means no relationship at all
- R Square signifies the Coefficient of Determination, which shows the goodness of fit. It shows how many points fall on the regression line. In our example, the value of R square is 0.96, which is an excellent fit. In other words, 96% of the dependent variables (y-values) are explained by the independent variables (x-values).
- Adjusted R Square is the modified version of R square that adjusts for predictors that are not significant to the regression model.
- Standard Error is another goodness-of-fit measure that shows the precision of your regression analysis.

# Excel (3)

- **ANOVA**
- ANOVA stands for Analysis of Variance. It provides information about the levels of variability within your regression model.
- Df is the number of degrees of freedom associated with the sources of variance.
- SS is the sum of squares. The smaller the Residual SS viz a viz the Total SS, the better the fit of your model to the data.
- MS is the mean square.
- F is the F statistic or F-test for the null hypothesis. It is very effectively used to test the overall model significance.
- Significance F is the P-value of F.
- **Regression Graph In Excel**
- You can quickly visualize the relationship between the two variables by creating a graph. To create a linear regression graph, follow these steps:
- Select the two variable columns of your data, including the headers.
- Go to Insert tab > Charts group > Scatter Plot.
- You will get a scatter plot in your worksheet.
- Now to add the trend line, right-click on any point and select Add Trend line.

# R

## The R Project for Statistical Computing

- **Linear model**
- `data=read.xlsx("DataRegressionExample1.xlsx")`
- `colnames(data)`  
`"preço" "area" "quartos"`
- `model<-`  
`lm(data$preço~data$area+data$quartos)`
- `summary(model)`

```
model <- lm(df$preço ~ df$area)
```

```
>  
> summary(model)
```

```
Call:  
lm(formula = df$preço ~ df$area)
```

```
Residuals:
```

```
   Min     1Q  Median     3Q    Max  
-117.477 -36.245  -6.399  31.862 235.899
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 11.0383    24.7519   0.446  0.657  
df$area      1.5099     0.1272  11.868 <2e-16 ***
```

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

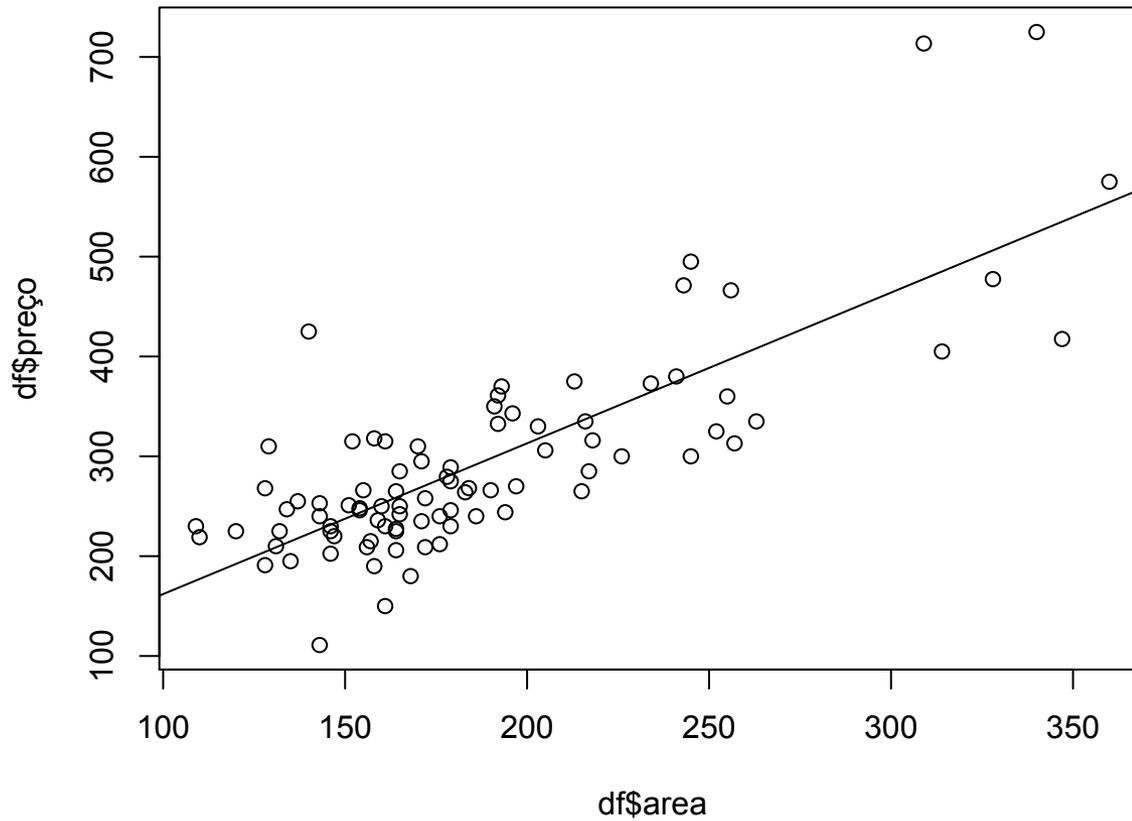
```
Residual standard error: 63.61 on 86 degrees of freedom
```

```
Multiple R-squared:  0.6209, Adjusted R-squared:  0.6165
```

```
F-statistic: 140.8 on 1 and 86 DF, p-value: < 2.2e-16
```

# Regressão com uma variável

# Visualização



# Regressão com 2 variáveis

- Call:
- `lm(formula = data$preço ~ data$area + data$quartos)`
  
- Residuals:
- Min   1Q Median   3Q   Max
- -128.056 -42.814 -7.128 32.466 229.645
  
- Coefficients:
- Estimate Std. Error t value Pr(>|t|)
- (Intercept) -19.2855   31.0475 -0.621   0.536
- data\$area   1.3836   0.1489 9.289 1.4e-14 \*\*\*
- data\$quartos 15.1213   9.4886 1.594   0.115
- \_\_\_\_\_
- Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
  
- Residual standard error: 63.05 on 85 degrees of freedom
- Multiple R-squared: 0.6319,   Adjusted R-squared: 0.6232
- F-statistic: 72.95 on 2 and 85 DF, p-value: < 2.2e-16

# Gráfico

