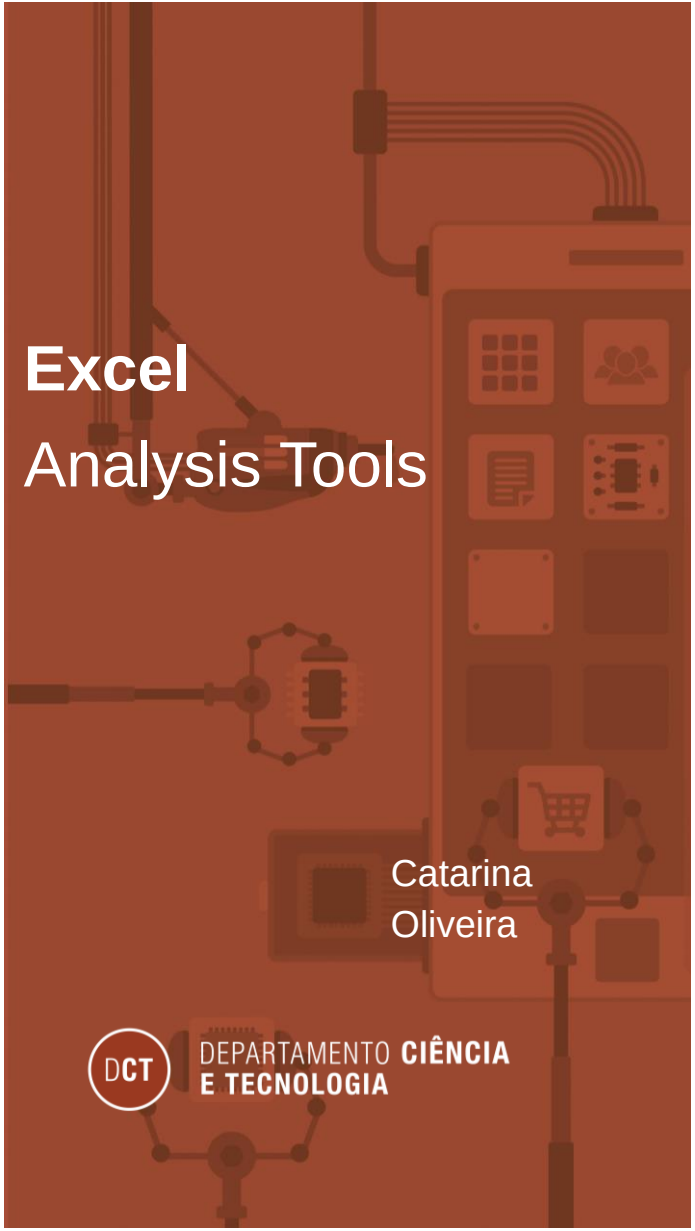


UNIVERSIDADE PORTUCALENSE

CONTENT

Analysis Tools

Analysis Toolpak Add-in

Tools that can be useful for problems of:

- Descriptive statistics
- Statistical inference
- Correlation
- Regression

Analysing:

- Variance
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test
- Fourier Analysis
- Histogram
- Moving Average
- Random number generation
- Rank and Percentile
- Regression
- Sampling
- T-Test
- Z-Test

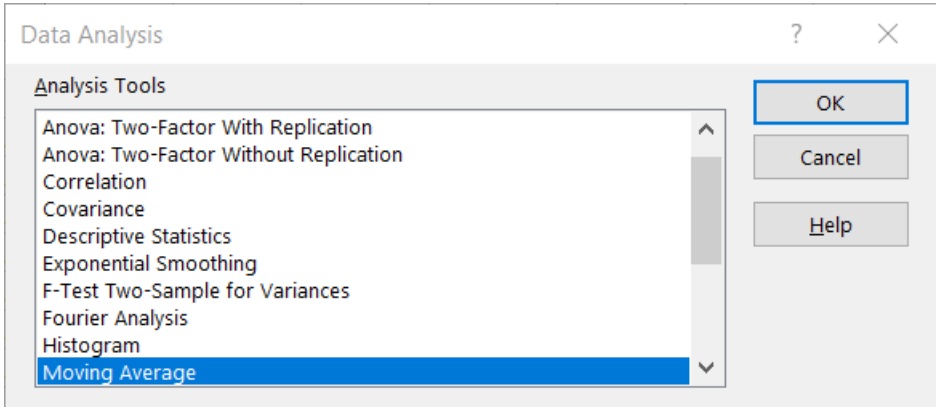
Installing and using the *Analysis Toolpak* Add-in

Installing:

1. Tab “File” > Command “Options”
2. On the Dialog “Excel Options”
    1. Select “Add-ins”
    2. On the “Manage” select box, choose “Excel Add-ins” and click “Go”
3. On the Add-ins dialog
    1. Select *Analysis Toolpak*

Using:

1. Tab “Data” > Group “Analysis” > Command “Data Analysis”
2. On the dialog “Data Analysis”
    1. Select the tool
    2. Ok
    3. Fill the tool’s dialog



Variance Analysis (ANOVA)

Objective: compare average of different samples

Allows: verify relation between the samples

H0: the samples' averages are equal

Three tools (depending on the number of samples and factors):

- **ANOVA – Single Factor**: performs a simple analysis of variance on data from one or more samples
- **ANOVA – Two Factor with Replication**: is used when data can be classified or have two different dimensions.
- **ANOVA – Two Factor without Replication** : similar to the previous one, but with the difference that there is only one observation for each pair of dimensions

Variance Analysis (ANOVA)

Production	Defects
922	11
1189	10
1109	5
1152	8
893	11
846	12
802	9
1110	14
1044	6
1117	14
1000	5
1065	0

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Production	12	12249	1020,75	16176,56818
Defects	12	105	8,75	17,29545455

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	6144864	1	6144864	758,9126521	1,51182E-18	4,300949502
Within Groups	178132,5	22	8096,931818			
Total	6322996,5	23				

H0: the samples have equal averages

F > F crit → reject H0 → The averages are different

A	B
922	1065
1189	1000
1109	1117
1152	1044
893	1110
846	802
802	846
1110	893
1044	1152
1117	1109
1000	1189
1065	922

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
A	12	12249	1020,75	16176,56818
B	12	12249	1020,75	16176,56818

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0	1	0	0	1	4,300949502
Within Groups	355884,5	22	16176,56818			
Total	355884,5	23				

H0: the samples have equal averages

F < F crit → do nor reject H0 → The averages are equal

Correlation

Allows:

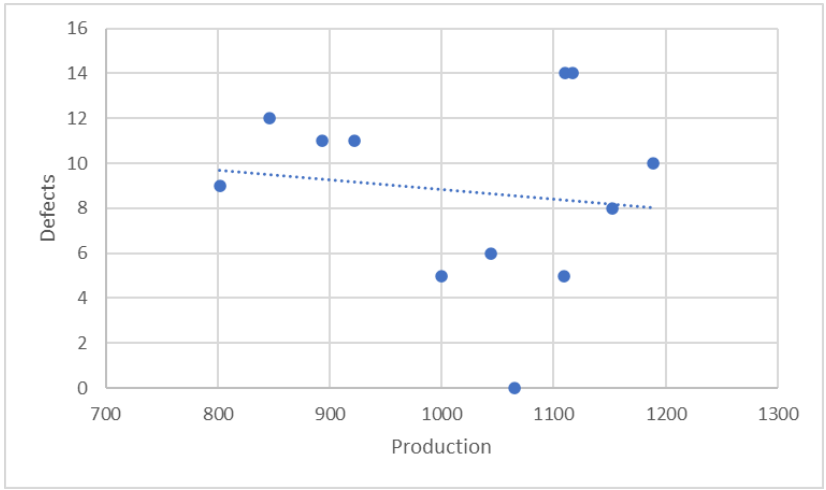
- Analyze and assess the magnitude and direction of the association or correlation between two variables without any type of assumption of functionality (ie: none of the variables is considered dependent on the other variable)

- Determining whether the two measurement variables tend to move together
  - **Positive correlation**: if large values of one variable tend to be associated with large values of the other
  - **Negative correlation**: if small values of one variable tend to be associated with large values of the other
  - **Correlation near zero**: if the values of both variables tend not to be related

Correlação

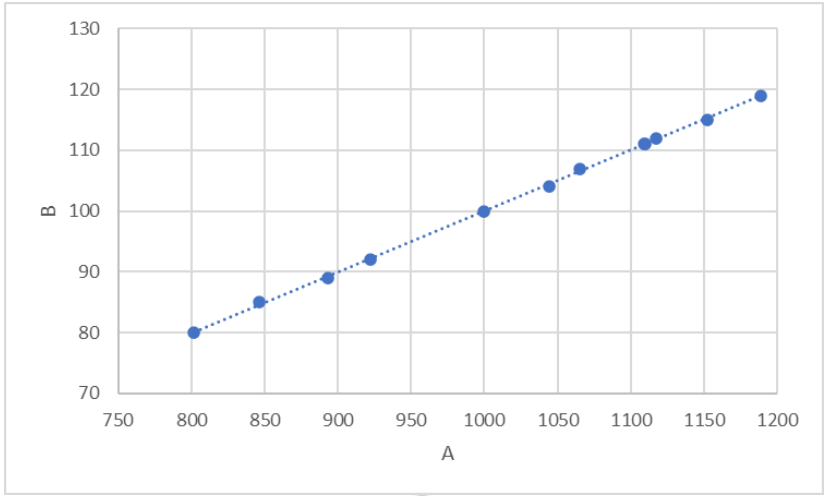
Production	Defects
922	11
1189	10
1109	5
1152	8
893	11
846	12
802	9
1110	14
1044	6
1117	14
1000	5
1065	0

	Production	*Defects*
Production	1	
Defects	-0,12903	1



A	B
922	92
1189	119
1109	111
1152	115
893	89
846	85
802	80
1110	111
1044	104
1117	112
1000	100
1065	107

	A	*B*
A	1	
B	0,999761	1



Covariance

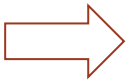
Allows:

- Analyze two variables to determine whether the two variables tend to move together
  - **Positive covariance**: if large values of one variable tend to be associated with large values of the other
  - **Negative covariance**: if small values of one variable tend to be associated with large values of the other
  - **Near-zero covariance**: if the values of both variables tend not to be related

Descriptive statystics

- Allows you to automatically generate an invariant statistics report for data defined as the input range, providing a set of information about the central trend and variability of the data

A	B	C
144	128	132
139	123	134
137	112	131
131	129	111
109	110	126
138	128	128
126	146	107
132	134	105
128	120	126
128	109	117



A		*B*		*C*	
Mean	131,2	Mean	123,9	Mean	121,7
Standard Error	3,065217049	Standard Error	3,686160303	Standard Error	3,419064199
Median	131,5	Median	125,5	Median	126
Mode	128	Mode	128	Mode	126
Standard Deviation	9,693067397	Standard Deviation	11,65666238	Standard Deviation	10,81203034
Sample Variance	93,95555556	Sample Variance	135,8777778	Sample Variance	116,9
Kurtosis	2,525244444	Kurtosis	-0,117946666	Kurtosis	-1,433295139
Skewness	-1,230019145	Skewness	0,381962526	Skewness	-0,548977137
Range	35	Range	37	Range	29
Minimum	109	Minimum	109	Minimum	105
Maximum	144	Maximum	146	Maximum	134
Sum	1312	Sum	1239	Sum	1217
Count	10	Count	10	Count	10
Largest(1)	144	Largest(1)	146	Largest(1)	134
Smallest(1)	109	Smallest(1)	109	Smallest(1)	105
Confidence Level(95,0%)	6,934002702	Confidence Level(95,0%)	8,338673932	Confidence Level(95,0%)	7,734460569

Exponential Smoothing

- Helps to predict a value based on a prior period forecast, adjusted for the error in that previous forecast.
- Uses an adjustment constant, whose magnitude determines the degree of smoothing of forecasts in relation to errors in the previous forecast

F-test

- Performs a two-sample F test to compare the variances of two populations.
- Used to verify the significance of the regression model and analyze the fit caused by the model

Fourier Analysis

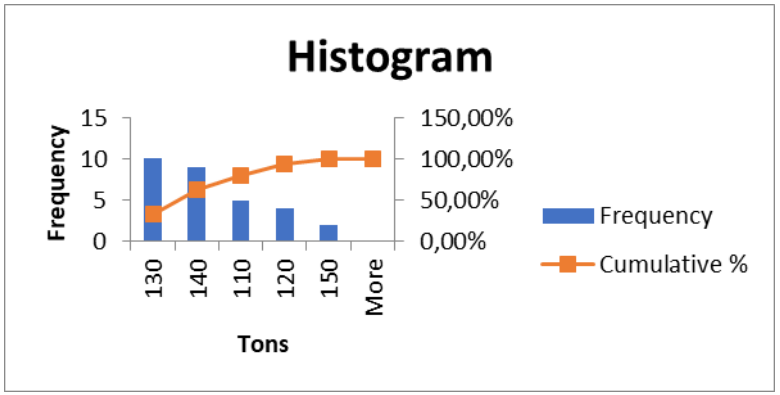
- Allows you to solve problems in linear systems and analyze periodic data, using the Fast Fourier Transform (FFT) method to transform the data.
- Supports inverse transformations, in which the inverse of the transformed data returns the original data

Histogram

- Calculates individual and cumulative frequencies for a range of data cells and data blocks
- Generates data for the number of occurrences of a value in a dataset

A	B	C
144	128	132
139	123	134
137	112	131
131	129	111
109	110	126
138	128	128
126	146	107
132	134	105
128	120	126
128	109	117

Tons	*Frequency*	*Cumulative %*	*Tons*	*Frequency*	*Cumulative %*
110	5	16,67%	130	10	33,33%
120	4	30,00%	140	9	63,33%
130	10	63,33%	110	5	80,00%
140	9	93,33%	120	4	93,33%
150	2	100,00%	150	2	100,00%
More	0	100,00%	More	0	100,00%



Moving Average

- Projects forecast period values based on the average value of the variable over a specified number of preceding periods
- Provides trending information that a simple average of all historical data does not directly show

Random number generation

- Fills a range of cells with a set of independent random numbers taken from one of several distributions

Rank and Percentile

- Produces a table containing the ordinal and percentage position of each value in a dataset.
- It is possible to analyze the relative position of values in a dataset

Regression

- Performs a linear regression analysis using the least squares method to fit a line across a set of observations
- Makes it possible to analyze how a single dependent variable is affected by the values of one or more independent variables

Regression

Production	Defects
922	11
1189	10
1109	5
1152	8
893	11
846	12
802	9
1110	14
1044	6
1117	14
1000	5
1065	0

Only 1% of the variation on Defects is explained by the variation in **Production**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,129030824
R Square	0,016648953
Adjusted R Square	-0,081686151
Standard Error	132,2798918
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2962,552234	2962,552234	0,169308341	0,689407295
Residual	10	174979,6978	17497,96978		
Total	11	177942,25			

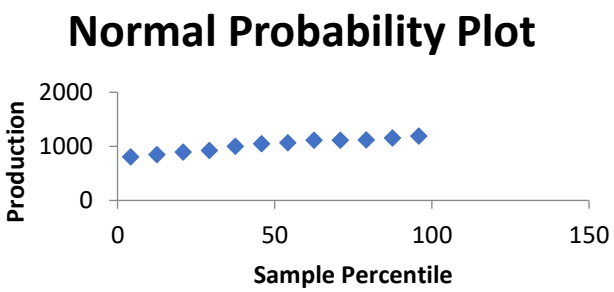
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	1055,278581	92,19484709	11,44617746	4,55009E-07	849,8556601	1260,701502	849,8556601	1260,701502
Defects	-3,946123522	9,590284664	-0,411470948	0,689407295	-25,31460938	17,42236234	-25,31460938	17,42236234

RESIDUAL OUTPUT

Observation	*Predicted Production*	*Residuals*	*Standard Residuals*
1	1011,871222	-89,87122208	-0,712562821
2	1015,817346	173,1826544	1,373114975
3	1035,547963	73,45203679	0,582379869
4	1023,709593	128,2904074	1,017177385
5	1011,871222	-118,8712221	-0,9424954
6	1007,925099	-161,9250986	-1,283857084
7	1019,763469	-217,7634691	-1,726583308
8	1000,032852	109,9671485	0,871897586
9	1031,60184	12,39816032	0,098301413
10	1000,032852	116,9671485	0,927398553
11	1035,547963	-35,54796321	-0,281849477
12	1055,278581	9,721419185	0,07707831

PROBABILITY OUTPUT

Percentile	*Production*
4,166666667	802
12,5	846
20,83333333	893
29,16666667	922
37,5	1000
45,83333333	1044
54,16666667	1065
62,5	1109
70,83333333	1110
79,16666667	1117
87,5	1152
95,83333333	1189



Regression

A	B
922	92
1189	119
1109	111
1152	115
893	89
846	85
802	80
1110	111
1044	104
1117	112
1000	100
1065	107

99,9% of the variation in **B** is explained by the variation in **A**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,999760629
R Square	0,999521316
Adjusted R Square	0,999473447
Standard Error	2,918529458
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	177857,0719	177857,0719	20880,60009	6,18632E-18
Residual	10	85,17814198	8,517814198		
Total	11	177942,25			

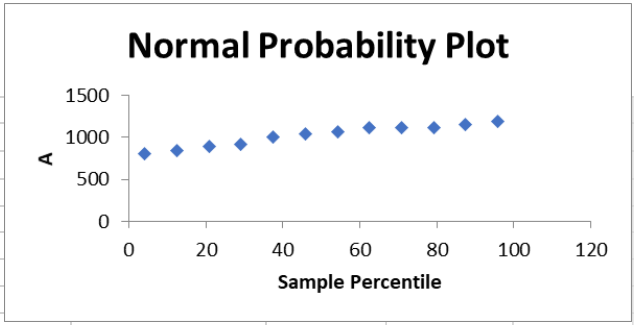
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	4,575560611	7,082578769	0,646030317	0,532803987	-11,20540832	20,35652954	-11,20540832	20,35652954
B	9,954361855	0,06888774	144,5012114	6,18632E-18	9,800870406	10,1078533	9,800870406	10,1078533

RESIDUAL OUTPUT

Observation	*Predicted A*	*Residuals*	*Standard Residuals*
1	920,3768513	1,623148707	0,583298113
2	1189,144621	-0,144621384	-0,051971443
3	1109,509727	-0,509726543	-0,183176396
4	1149,327174	2,672826036	0,960512353
5	890,5137657	2,486234273	0,893458347
6	850,6963183	-4,696318306	-1,687678766
7	800,924509	1,07549097	0,386490684
8	1109,509727	0,490273457	0,176185695
9	1039,829194	4,170806444	1,498829724
10	1119,464088	-2,464088398	-0,885499959
11	1000,011746	-0,011746135	-0,004221116
12	1069,692279	-4,692279122	-1,686227236

PROBABILITY OUTPUT

Percentile	*A*
4,166666667	802
12,5	846
20,83333333	893
29,16666667	922
37,5	1000
45,83333333	1044
54,16666667	1065
62,5	1109
70,83333333	1110
79,16666667	1117
87,5	1152
95,83333333	1189



Sampling

- Allows you to create a sample of a population, treating the input range as the study population
- Sampling is very useful when a study population is too large to be processed and analyzed
- Allows you to create random and periodic samples

T-Test

Two Paired Samples for Averages

- Used when there is a natural pairing of observations in the samples, such as when a group of samples is tested twice, before and after an experiment.
- Performs a Student's t-test to determine whether observations made before and after a treatment are likely to result from distributions with equal population means.
- Does not assume that the variances of both populations are equal

Two Samples with Equal Variances

- Performs a two-sample Student's t-test
- It assumes that the two data sets come from distributions with the same variances.
- It can be used to determine whether the two samples are likely to have resulted from distributions with equal population means

Two Samples with Unequal Variances

- Performs a two-sample Student's t-test
- It assumes that the two data sets come from distributions with unequal variances
- It can be used to determine whether the two samples are likely to have resulted from distributions with equal population means

Z-Test

- Performs a two-sample Z test for means with known variances
- Used to test the null hypothesis that there is no difference between the two population means, against the unilateral or bilateral alternative hypotheses
- If the variances are not known, the ZTEST function should be used instead.

UNIVERSIDADE
PORTUCALENSE

Do conhecimento à prática.