Statistics for Data Analysis

Powered by SPSS

Gold Business Partner

Competency Data Science & Business Analytics

Webinar

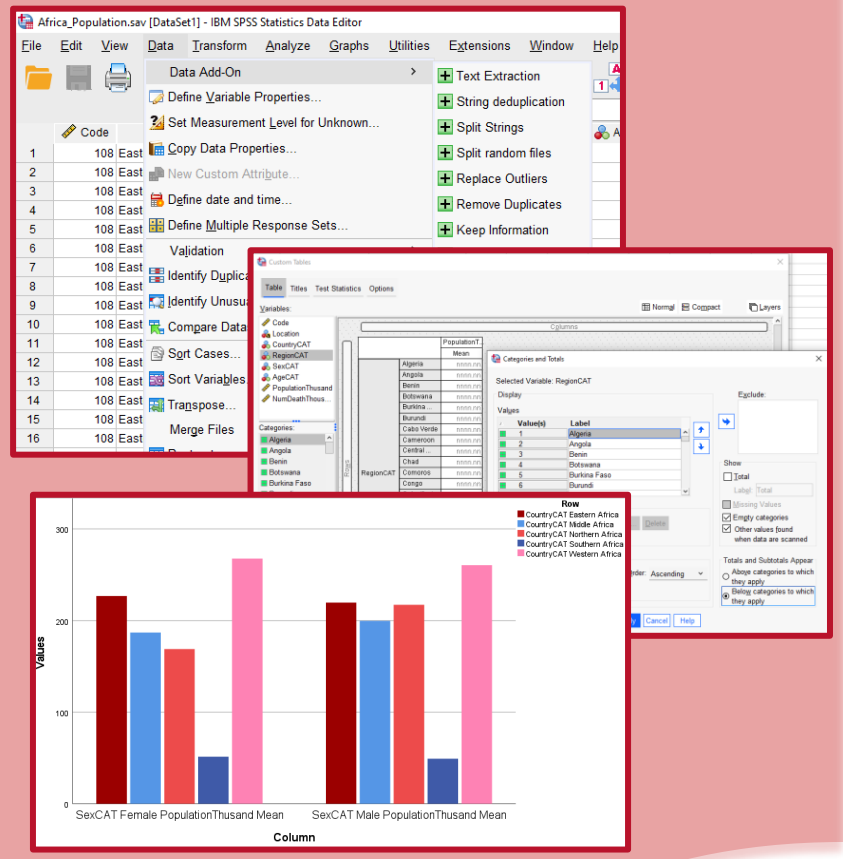
How to choose a statistical test?

Statistics for Data Analysis

Webinar | How to choose a statistical test?

SPSS

- Introduction to Statistics for Data Analysis
- Brief introduction to non-parametric tests
- Test for independent and paired samples
- Test characteristics and examples
- Q&A
- Thanks



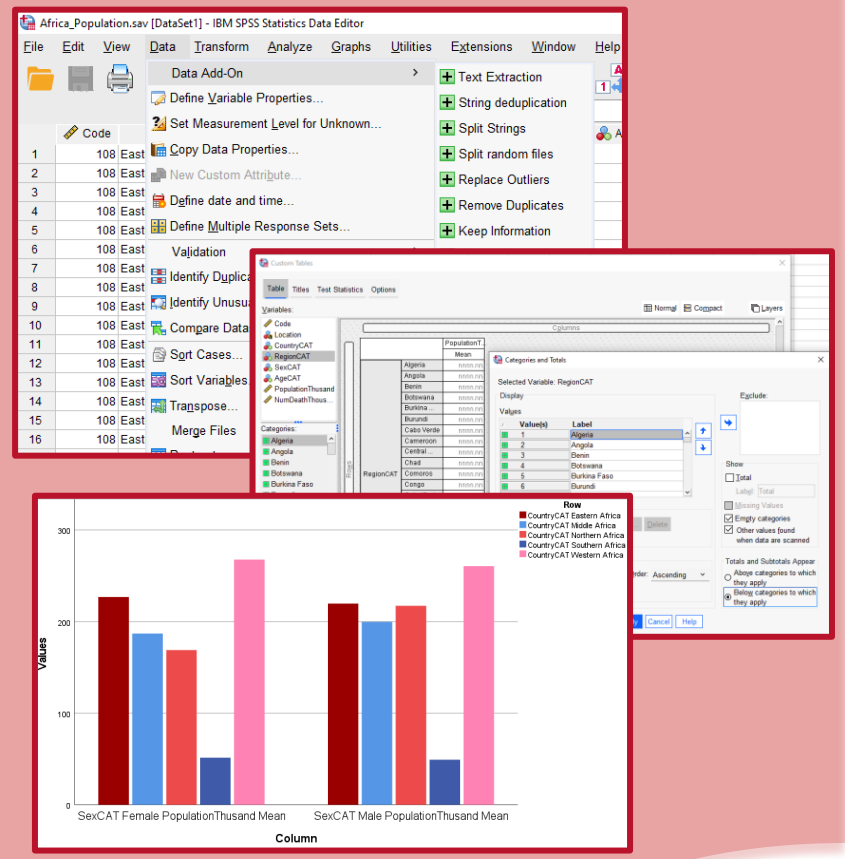
www.statisticsfordataanalysis.com

Statistics for Data Analysis

Webinar | How to choose a statistical test?

SPSS

- Introduction to Statistics for Data Analysis
- Brief introduction to non-parametric tests
- Test for independent and paired samples
- Test characteristics and examples
- Q&A
- Thanks



Statistics for Data Analysis

Introduction to NON-Parametric Tests

SPSS

A non-parametric test is a test whose model does not specify conditions about the parameters of the population from which the sample comes

Advantages of non-parametric tests

- Less restrictive assumptions
- Possibility of use even with small samples
- Analysis of different populations
- Analysis of ranks or nominal data

Disadvantages of non-parametric tests

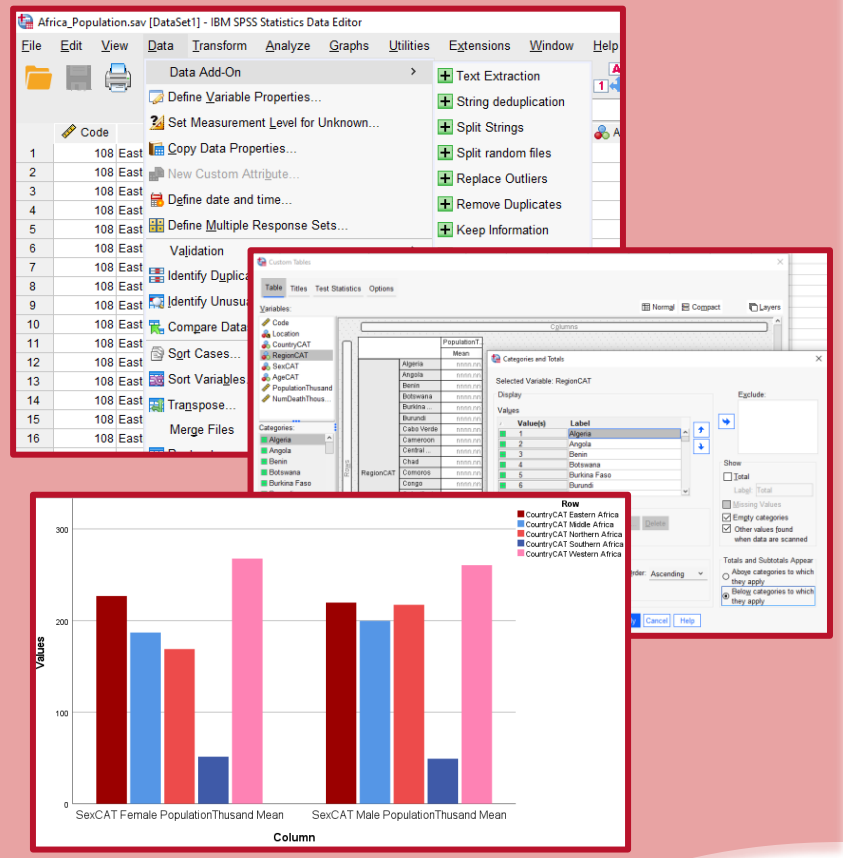
- Less precise with the same information (in case the assumptions were verified)

Statistics for Data Analysis

Webinar | How to choose a statistical test?

SPSS

- Introduction to Statistics for Data Analysis
- Brief introduction to non-parametric tests
- Test for independent and paired samples
- Test characteristics and examples
- Q&A
- Thanks



www.statisticsfordataanalysis.com

Summary of the main statistical tests

From the stratified means, however, it is not possible to quantify whether the difference between the groups is due to a sampling error (to the case that an "unfavorable" sample was extracted) or whether this difference actually exists (statistically significant). The hypothesis is verified that they are actually equal and if the data are consistent with it.

Instrument: T-TEST FOR TWO INDEPENDENT SAMPLES

Instrument: MANN-WHITNEY RANKS TEST

Mann-Whitney U test for 2 independent samples

- $H_0 : Me_1 = Me_2$

- The Mann-Whitney U test procedure is based on precedence, ie it does not use the sum of the ranks, but the sum of the differences between the ranks of the two groups;

A	B
92	156
114	123
82	198
164	83
167	242
110	176
135	185
	217

Mann-Whitney U test for 2 independent samples

1. We combine the data of the two groups into a single whole, arranging the values or scores in ascending order according to the algebraic value.

82	83	92	110	114	123	135	156	164	167	176	185	198	217	242
A	**B**	**A**	**A**	**A**	**B**	**A**	**B**	**A**	**A**	**B**	**B**	**B**	**B**	**B**

2. The number of precedences is counted. The smaller sum of these precedences is the value of U:

$$U = 0 + 1 + 1 + 1 + 2 + 3 + 3 = 11$$

82	83	92	110	114	123	135	156	164	167	176	185	198	217	242
A	**B**	**A**	**A**	**A**	**B**	**A**	**B**	**A**	**A**	**B**	**B**	**B**	**B**	**B**
0		**1**	**1**	**1**		**2**		**3**	**3**					

3. In the event that the hypothesis H1 is true, therefore one sample has a significantly lower median than the other, the value of U will tend to 0, since the data of the group under examination, which must provide the smaller total, will precede all the data of the other group and therefore each of them will have 0 precedence.

Mann-Whitney U test for 2 independent samples

4. In the event that the hypothesis H0 of equality or identity of the two central trends is true, the data of the two groups will be randomly mixed
5. U will tend to an average value (mU), depending on the number of observations present n1 and n2, according to the relationship:

$$\mu_U = \frac{n_1 \cdot n_2}{2} \qquad \sigma_U \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}$$

6. In the case of large samples (n1 or n2> 15), if the null hypothesis H0 is true, the sampling distribution of U is well approximated by the normal distribution, with mean 0 and unit variance:

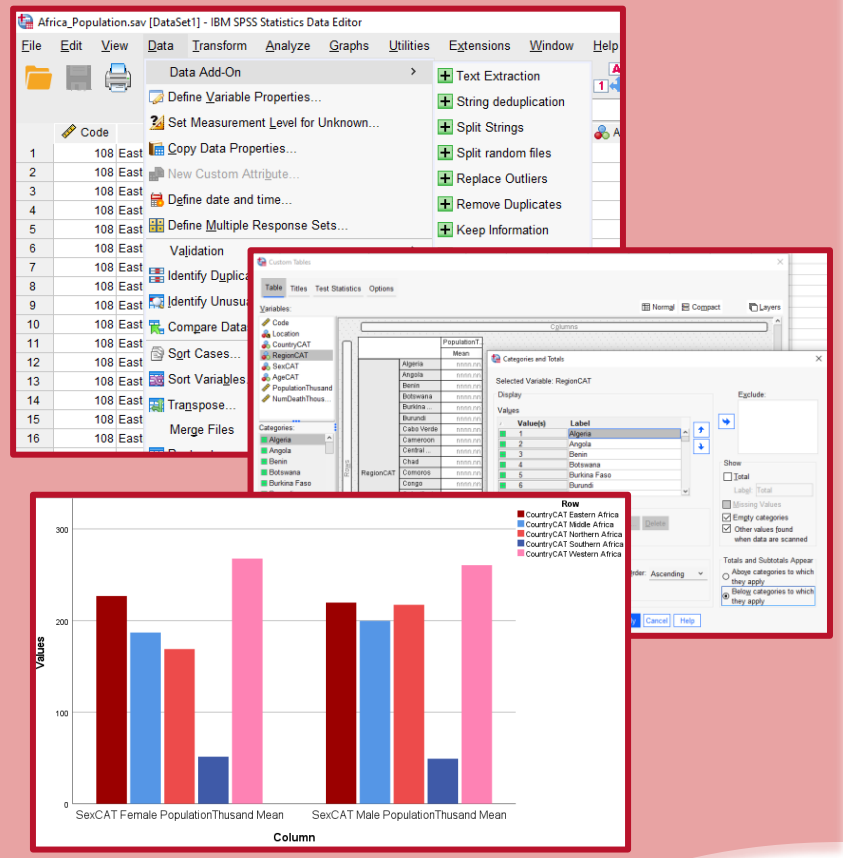
$$z = \frac{U - \mu_U}{\sigma_U}$$

Statistics for Data Analysis

Webinar | How to choose a statistical test?

SPSS

- Introduction to Statistics for Data Analysis
- Brief introduction to non-parametric tests
- Test for independent and paired samples
- Test characteristics and examples
- Q&A
- Thanks

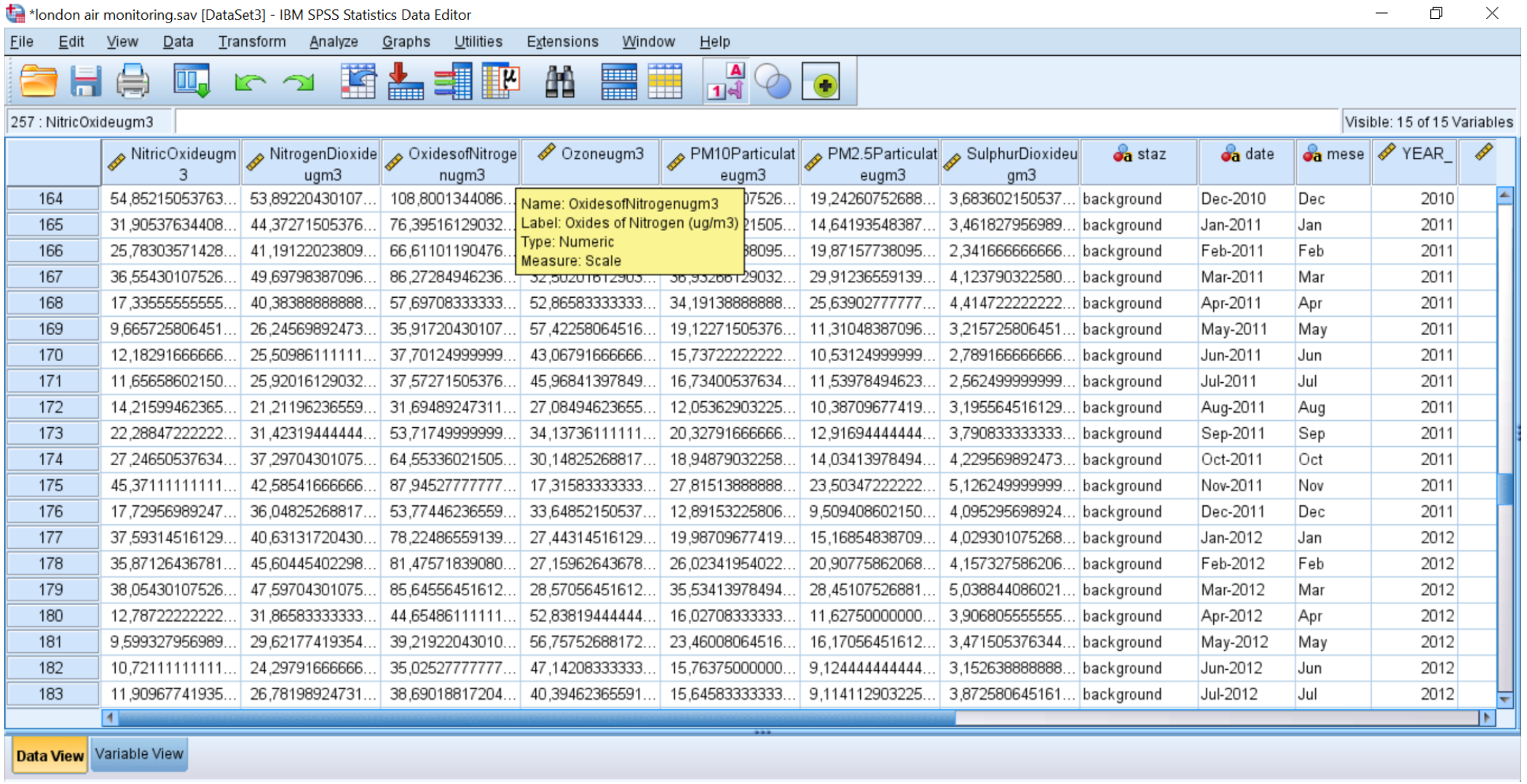


www.statisticsfordataanalysis.com

Statistics for Data Analysis

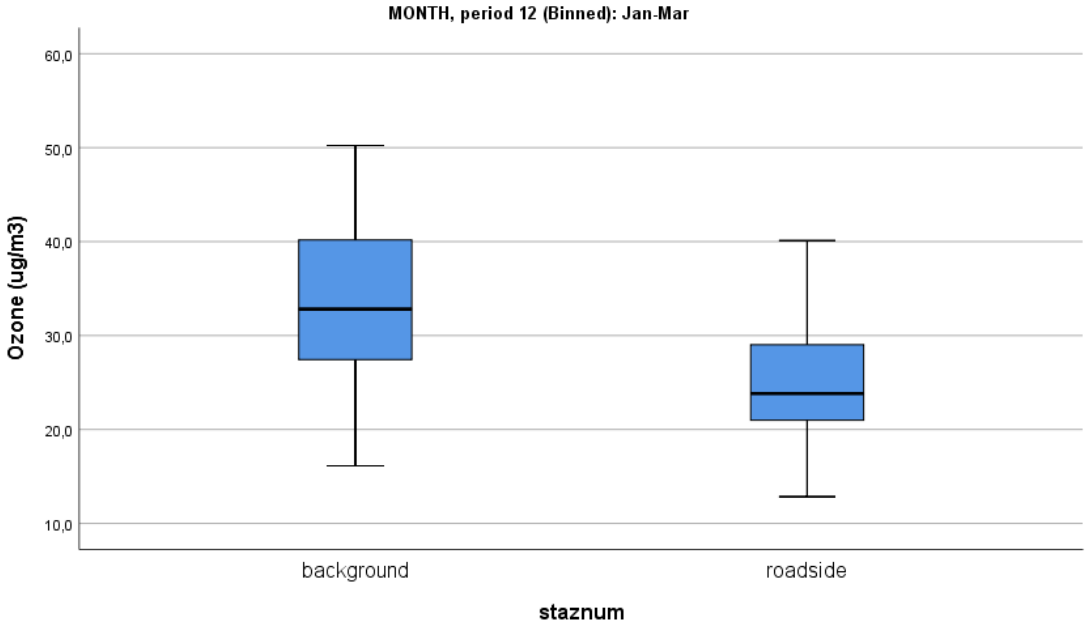
Test characteristics and examples

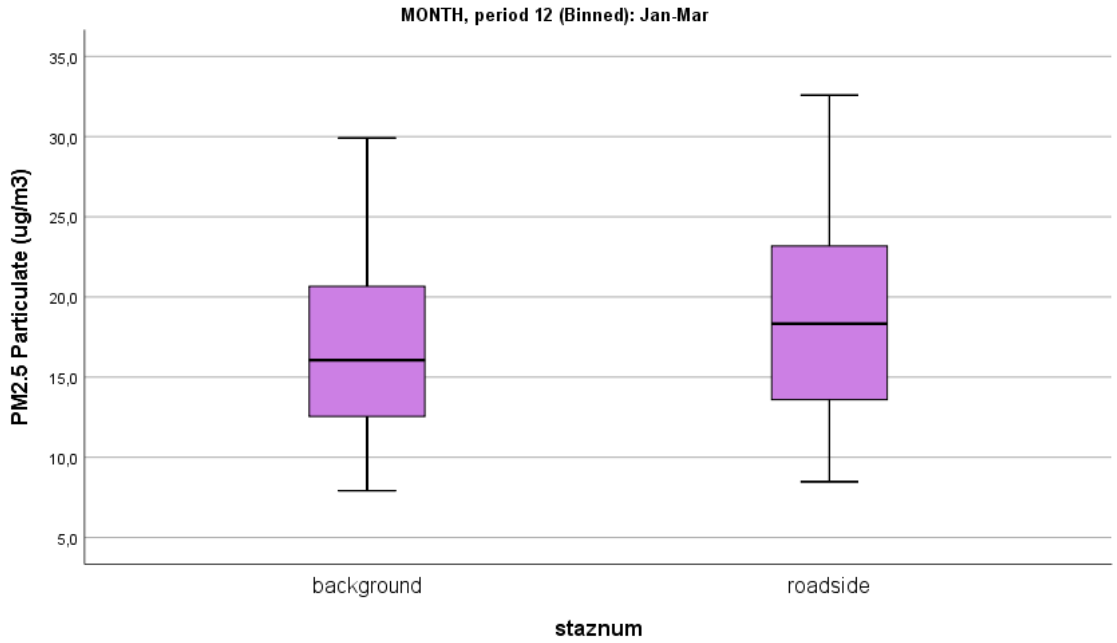
SPSS



	NitricOxideugm3	NitrogenDioxideugm3	OxidesofNitrogenugm3	Ozoneugm3	PM10Particulateugm3	PM2.5Particulateugm3	SulphurDioxideugm3	staz	date	mese	YEAR_
164	54,85215053763...	53,89220430107...	108,8001344086...		...07526...	19,24260752688...	3,683602150537...	background	Dec-2010	Dec	2010
165	31,90537634408...	44,37271505376...	76,39516129032...		...21505...	14,64193548387...	3,461827956989...	background	Jan-2011	Jan	2011
166	25,78303571428...	41,19122023809...	66,61101190476...		...38095...	19,87157738095...	2,341666666666...	background	Feb-2011	Feb	2011
167	36,55430107526...	49,69798387096...	86,27284946236...		...36,95266129032...	29,91236559139...	4,123790322580...	background	Mar-2011	Mar	2011
168	17,33555555555...	40,38388888888...	57,69708333333...	52,86583333333...	34,19138888888...	25,63902777777...	4,414722222222...	background	Apr-2011	Apr	2011
169	9,665725806451...	26,24569892473...	35,91720430107...	57,42258064516...	19,12271505376...	11,31048387096...	3,215725806451...	background	May-2011	May	2011
170	12,18291666666...	25,50986111111...	37,70124999999...	43,06791666666...	15,73722222222...	10,53124999999...	2,789166666666...	background	Jun-2011	Jun	2011
171	11,65658602150...	25,92016129032...	37,57271505376...	45,96841397849...	16,73400537634...	11,53978494623...	2,562499999999...	background	Jul-2011	Jul	2011
172	14,21599462365...	21,21196236559...	31,69489247311...	27,08494623655...	12,05362903225...	10,38709677419...	3,195564516129...	background	Aug-2011	Aug	2011
173	22,28847222222...	31,42319444444...	53,71749999999...	34,13736111111...	20,32791666666...	12,91694444444...	3,790833333333...	background	Sep-2011	Sep	2011
174	27,24650537634...	37,29704301075...	64,55336021505...	30,14825268817...	18,94879032258...	14,03413978494...	4,229569892473...	background	Oct-2011	Oct	2011
175	45,37111111111...	42,58541666666...	87,94527777777...	17,31583333333...	27,81513888888...	23,50347222222...	5,126249999999...	background	Nov-2011	Nov	2011
176	17,72956989247...	36,04825268817...	53,77446236559...	33,64852150537...	12,89153225806...	9,509408602150...	4,095295698924...	background	Dec-2011	Dec	2011
177	37,59314516129...	40,63131720430...	78,22486559139...	27,44314516129...	19,98709677419...	15,16854838709...	4,029301075268...	background	Jan-2012	Jan	2012
178	35,87126436781...	45,60445402298...	81,47571839080...	27,15962643678...	26,02341954022...	20,90775862068...	4,157327586206...	background	Feb-2012	Feb	2012
179	38,05430107526...	47,59704301075...	85,64556451612...	28,57056451612...	35,53413978494...	28,45107526881...	5,038844086021...	background	Mar-2012	Mar	2012
180	12,78722222222...	31,86583333333...	44,65486111111...	52,83819444444...	16,02708333333...	11,62750000000...	3,906805555555...	background	Apr-2012	Apr	2012
181	9,599327956989...	29,62177419354...	39,21922043010...	56,75752688172...	23,46008064516...	16,17056451612...	3,471505376344...	background	May-2012	May	2012
182	10,72111111111...	24,29791666666...	35,02527777777...	47,14208333333...	15,76375000000...	9,124444444444...	3,152638888888...	background	Jun-2012	Jun	2012
183	11,90967741935...	26,78198924731...	38,69018817204...	40,39462365591...	15,64583333333...	9,114112903225...	3,872580645161...	background	Jul-2012	Jul	2012

Test characteristics and examples





Summary of the main statistical tests

There may be a case in which the researcher is interested in evaluating whether the same statistical units report different values after the action of a factor in time or space.

For example, you want to compare some ladies' weight before and after slimming treatment.

For example, You want to evaluate the effect of treatment with a before / after design

Instrument: TEST T FOR TWO PAIRED SAMPLES

Instrument: WILCOXON RANKS TEST

Wilcoxon test for paired data

- H_0 : d median = 0;
- H_1 : d median ≠ 0

1. The difference column is calculated with their sign:

   diff. = X-Y

2. Null differences are eliminated from the analysis;

3. Differences, considered in absolute value, are transformed into their rank.

Ora	Giorno X	Giorno Y	diff.(X-Y)	Rango
6	120	140	*- 20*	- 3,5
7	145	160	*- 15*	- 2
8	305	295	*+ 10*	+1
9	200	230	*- 30*	- 7
10	160	200	*- 40*	-10
11	135	185	*- 50*	-12
12	170	150	+20	+3,5
13	285	515	*-230*	-14
14	290	220	*+ 70*	+13
15	200	225	*- 25*	-5
16	150	180	*- 30*	-7
17	160	190	*- 30*	-7
18	115	115	*0*	-

Wilcoxon test for paired data

4. The sign of the corresponding difference is attributed to each rank

5. The ranks with the same sign add up

6. The smaller total is chosen, which is the value of T :

$$T = 1 + 3{,}5 + 13 = 17{,}5.$$

7. According to the null hypothesis H_0, the difference between the two sets of paired observations should be equal to zero: that is, in the difference column, the sum of the ranks with a positive sign and the sum of the ranks with a negative sign should be equal.

Ora	Giorno X	Giorno Y	diff.(X-Y)	Rango
6	120	140	*- 20*	- 3,5
7	145	160	*- 15*	- 2
8	305	295	*+ 10*	+1
9	200	230	*- 30*	- 7
10	160	200	*- 40*	-10
11	135	185	*- 50*	-12
12	170	150	+20	+3,5
13	285	515	*-230*	-14
14	290	220	*+ 70*	+13
15	200	225	*- 25*	-5
16	150	180	*- 30*	-7
17	160	190	*- 30*	-7
18	115	115	*0*	-

Wilcoxon test for paired data

8. Consequently, if H_0 is true, the smaller total should tend to an expected mean value mT determined by N, the number of differences or ranks, according to the relation :

$$\mu_T \frac{N(N+1)}{4}$$

With the data from the example, $m_{T:}$ 52,5.

8. The significance of the difference between the two paired data series is evaluated by translating it into the significance of the difference between T and m_T; in the example between 17,5 and 52,5.

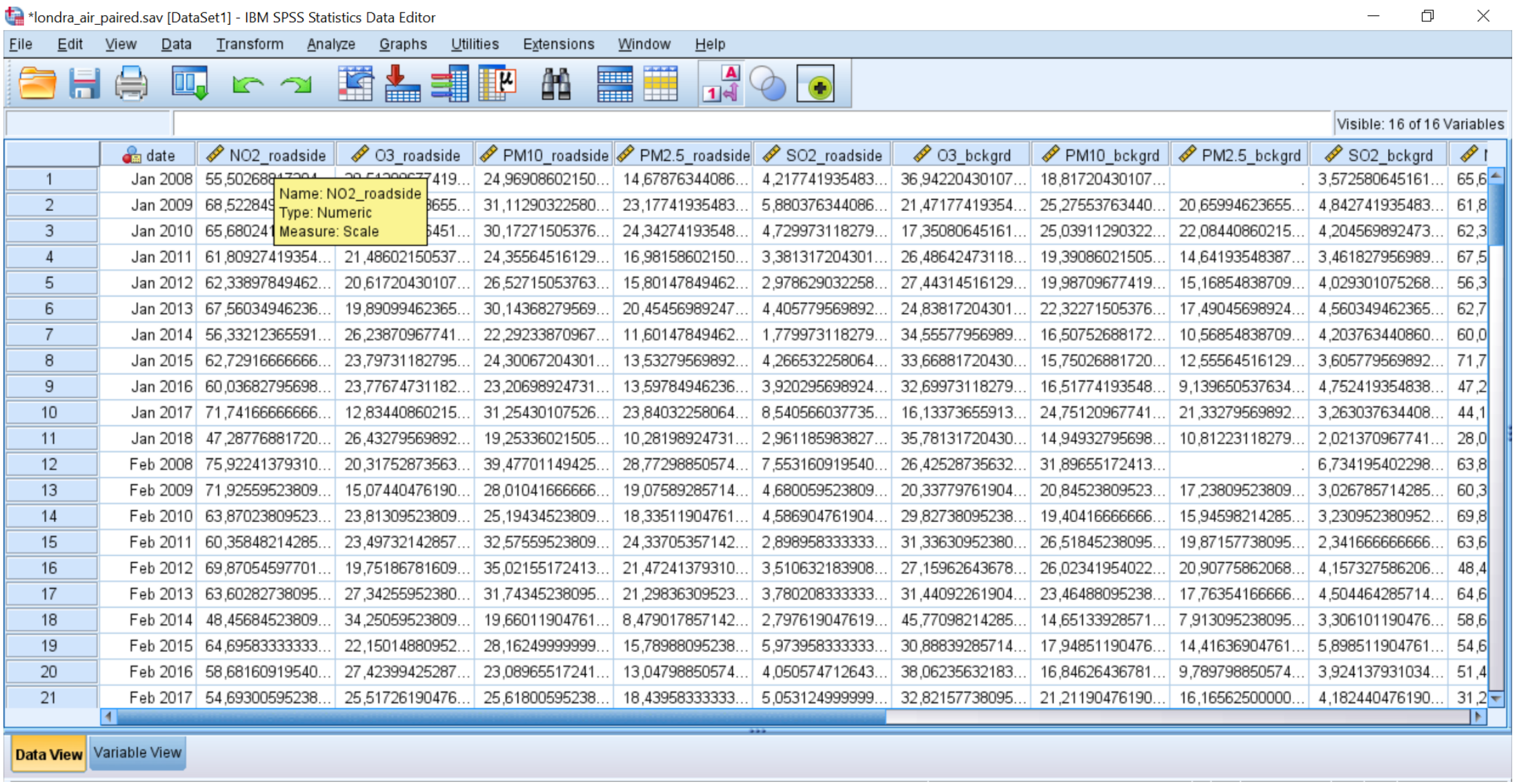
9. For large samples (N > 25), the value of the rank sum (T) is approximately normally distributed:

$$z = \frac{T - \mu_T}{\sigma_T} \qquad \sigma_T \sqrt{\frac{N \cdot (N+1) \cdot (2N+1)}{24}}$$

Statistics for Data Analysis

Test characteristics and examples

SPSS

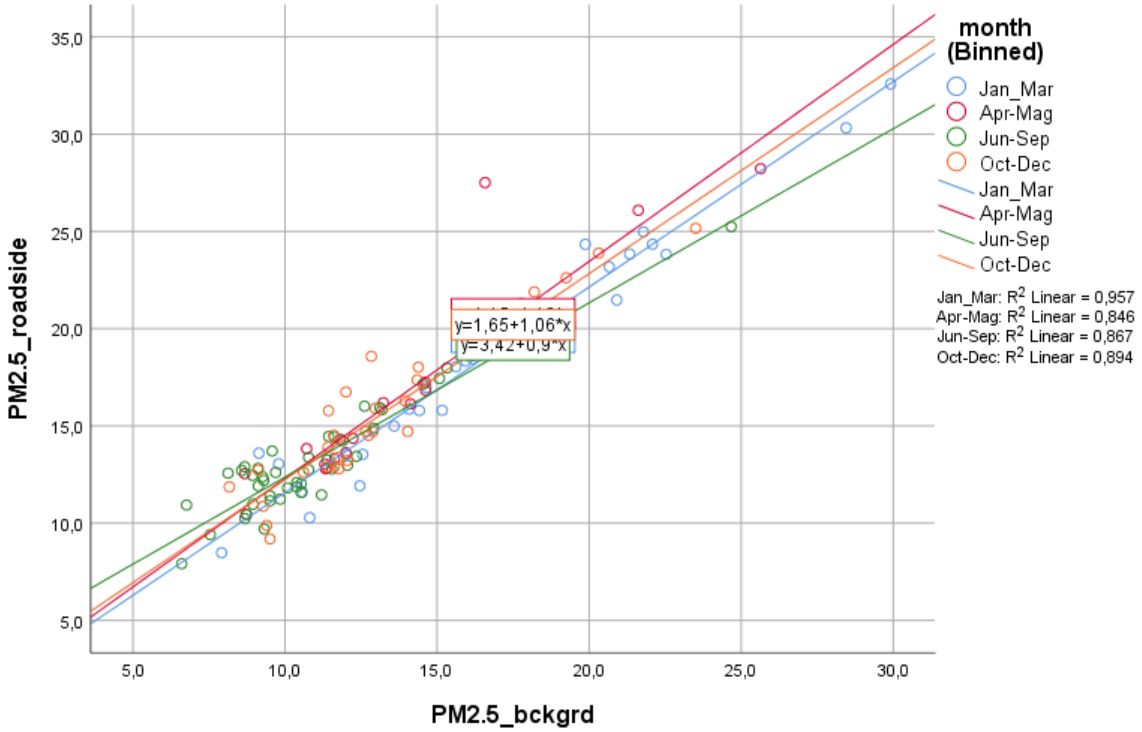


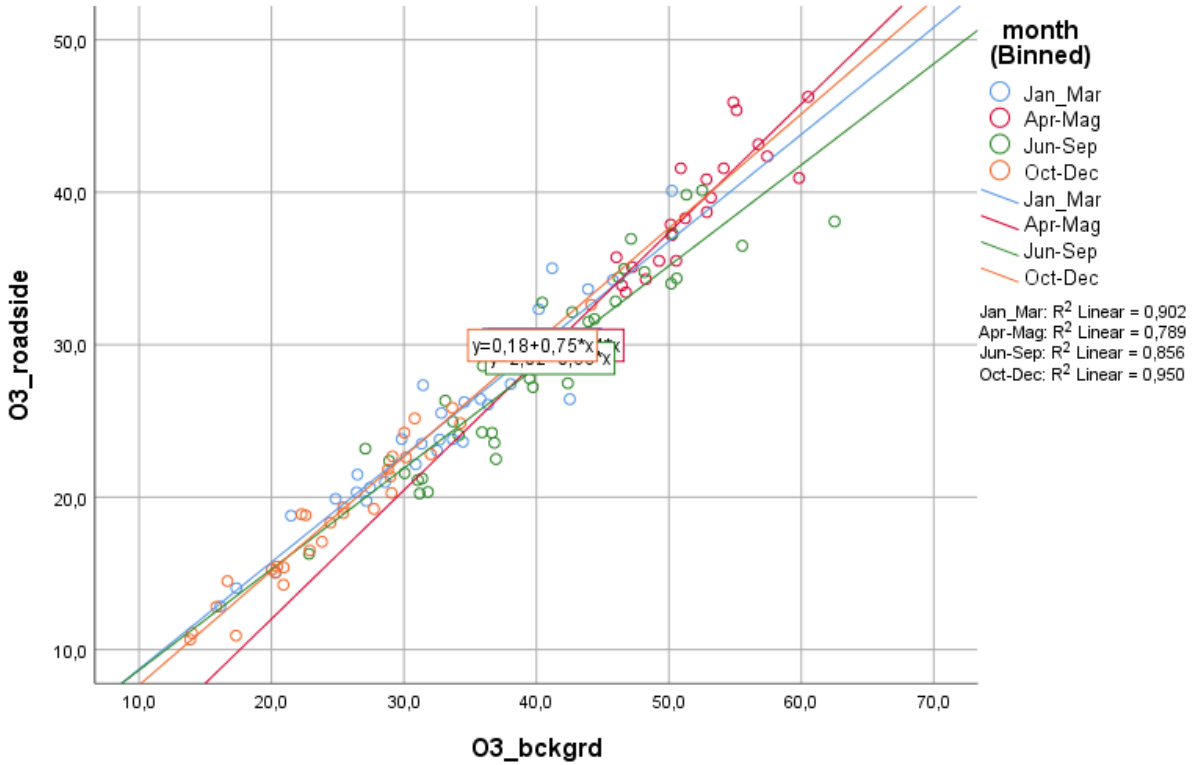
	date	NO2_roadside	O3_roadside	PM10_roadside	PM2.5_roadside	SO2_roadside	O3_bckgrd	PM10_bckgrd	PM2.5_bckgrd	SO2_bckgrd	I
1	Jan 2008	55,5026881...	...419...	24,96908602150...	14,67876344086...	4,217741935483...	36,94220430107...	18,81720430107...	.	3,572580645161...	65,6
2	Jan 2009	68,522849...	...655...	31,11290322580...	23,17741935483...	5,880376344086...	21,47177419354...	25,27553763440...	20,65994623655...	4,842741935483...	61,8
3	Jan 2010	65,680241...	...451...	30,17271505376...	24,34274193548...	4,729973118279...	17,35080645161...	25,03911290322...	22,08440860215...	4,204569892473...	62,3
4	Jan 2011	61,80927419354...	21,48602150537...	24,35564516129...	16,98158602150...	3,381317204301...	26,48642473118...	19,39086021505...	14,64193548387...	3,461827956989...	67,5
5	Jan 2012	62,33897849462...	20,61720430107...	26,52715053763...	15,80147849462...	2,978629032258...	27,44314516129...	19,98709677419...	15,16854838709...	4,029301075268...	56,3
6	Jan 2013	67,56034946236...	19,89099462365...	30,14368279569...	20,45456989247...	4,405779569892...	24,83817204301...	22,32271505376...	17,49045698924...	4,560349462365...	62,7
7	Jan 2014	56,33212365591...	26,23870967741...	22,29233870967...	11,60147849462...	1,779973118279...	34,55577956989...	16,50752688172...	10,56854838709...	4,203763440860...	60,0
8	Jan 2015	62,72916666666...	23,79731182795...	24,30067204301...	13,53279569892...	4,266532258064...	33,66881720430...	15,75026881720...	12,55564516129...	3,605779569892...	71,7
9	Jan 2016	60,03682795698...	23,77674731182...	23,20698924731...	13,59784946236...	3,920295698924...	32,69973118279...	16,51774193548...	9,139650537634...	4,752419354838...	47,2
10	Jan 2017	71,74166666666...	12,83440860215...	31,25430107526...	23,84032258064...	8,540566037735...	16,13373655913...	24,75120967741...	21,33279569892...	3,263037634408...	44,1
11	Jan 2018	47,28776881720...	26,43279569892...	19,25336021505...	10,28198924731...	2,961185983827...	35,78131720430...	14,94932795698...	10,81223118279...	2,021370967741...	28,0
12	Feb 2008	75,92241379310...	20,31752873563...	39,47701149425...	28,77298850574...	7,553160919540...	26,42528735632...	31,89655172413...	.	6,734195402298...	63,8
13	Feb 2009	71,92559523809...	15,07440476190...	28,01041666666...	19,07589285714...	4,680059523809...	20,33779761904...	20,84523809523...	17,23809523809...	3,026785714285...	60,3
14	Feb 2010	63,87023809523...	23,81309523809...	25,19434523809...	18,33511904761...	4,586904761904...	29,82738095238...	19,40416666666...	15,94598214285...	3,230952380952...	69,8
15	Feb 2011	60,35848214285...	23,49732142857...	32,57559523809...	24,33705357142...	2,898958333333...	31,33630952380...	26,51845238095...	19,87157738095...	2,341666666666...	63,6
16	Feb 2012	69,87054597701...	19,75186781609...	35,02155172413...	21,47241379310...	3,510632183908...	27,15962643678...	26,02341954022...	20,90775862068...	4,157327586206...	48,4
17	Feb 2013	63,60282738095...	27,34255952380...	31,74345238095...	21,29836309523...	3,780208333333...	31,44092261904...	23,46488095238...	17,76354166666...	4,504464285714...	64,6
18	Feb 2014	48,45684523809...	34,25059523809...	19,66011904761...	8,479017857142...	2,797619047619...	45,77098214285...	14,65133928571...	7,913095238095...	3,306101190476...	58,6
19	Feb 2015	64,69583333333...	22,15014880952...	28,16249999999...	15,78988095238...	5,973958333333...	30,88839285714...	17,94851190476...	14,41636904761...	5,898511904761...	54,6
20	Feb 2016	58,68160919540...	27,42399425287...	23,08965517241...	13,04798850574...	4,050574712643...	38,06235632183...	16,84626436781...	9,789798850574...	3,924137931034...	51,4
21	Feb 2017	54,69300595238...	25,51726190476...	25,61800595238...	18,43958333333...	5,053124999999...	32,82157738095...	21,21190476190...	16,16562500000...	4,182440476190...	31,2

Statistics for Data Analysis

Test characteristics and examples

SPSS



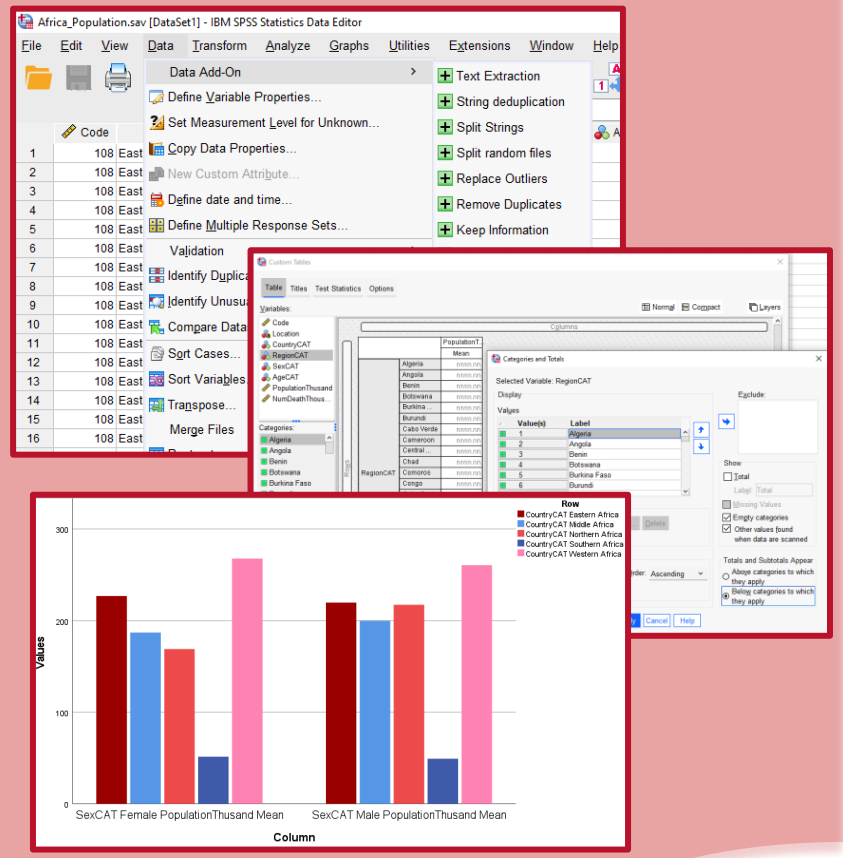


Statistics for Data Analysis

Webinar | How to choose a statistical test?

SPSS

- Introduction to Statistics for Data Analysis
- Brief introduction to non-parametric tests
- Test for independent and paired samples
- Test characteristics and examples
- Q&A
- Thanks

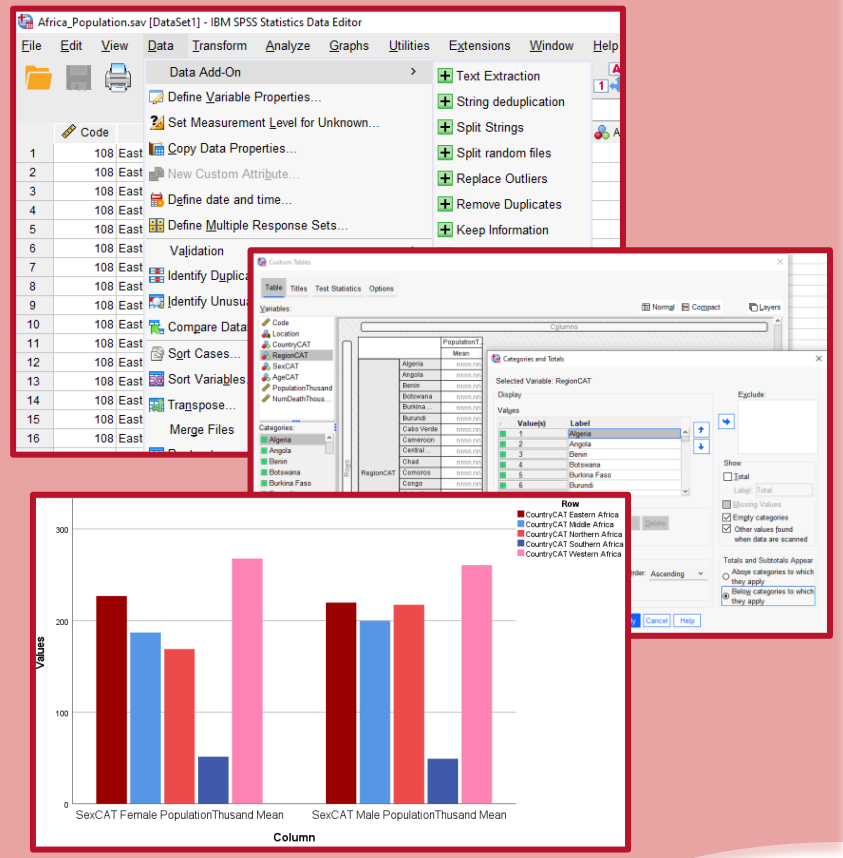


Statistics for Data Analysis

Webinar | How to choose a statistical test?

SPSS

- Introduction to Statistics for Data Analysis
- Brief introduction to non-parametric tests
- Test for independent and paired samples
- Test characteristics and examples
- Q&A
- Thanks



Statistics for Data Analysis

QUESTIONS & ANSWERS

SPSS

Statistics for Data Analysis

Webinar | How to choose a statistical test?

SPSS

@Sales

THANKS

Sales & Training Teams @SPS