Compare Means

Dr. Anahita Babak

Assistant Professor, Community and family medicine department, Isfahan University of Medical Sciences

Email: babakanahita@yahoo.com

Compare Means

Parametric

- One sample t- test
- Paired samples t-test
- Independent samples t-test
- One-way ANOVA

Nonparametric

- Wilcoxon signed-rank test
- Mann Whitney U-test
- Kruscal wallis

One Sample t -Test

- **examines whether the mean of a population is statistically different from a known or hypothesized value**
- **is also known as Single Sample t -Test**
- **the test variable's mean is compared against a "test value", which is a known or hypothesized value of the mean in the population**
- **Test values may come from a literature review, a trusted research organization, legal requirements, or industry standards**

Common Uses

- Statistical difference between a mean and a known or hypothesized value of the mean in the population.
- Statistical difference between a change score and zero:

✓ creating a change score from two variables, and then comparing the mean change score to zero, if the mean change score is not significantly different from zero, no significant change occurred

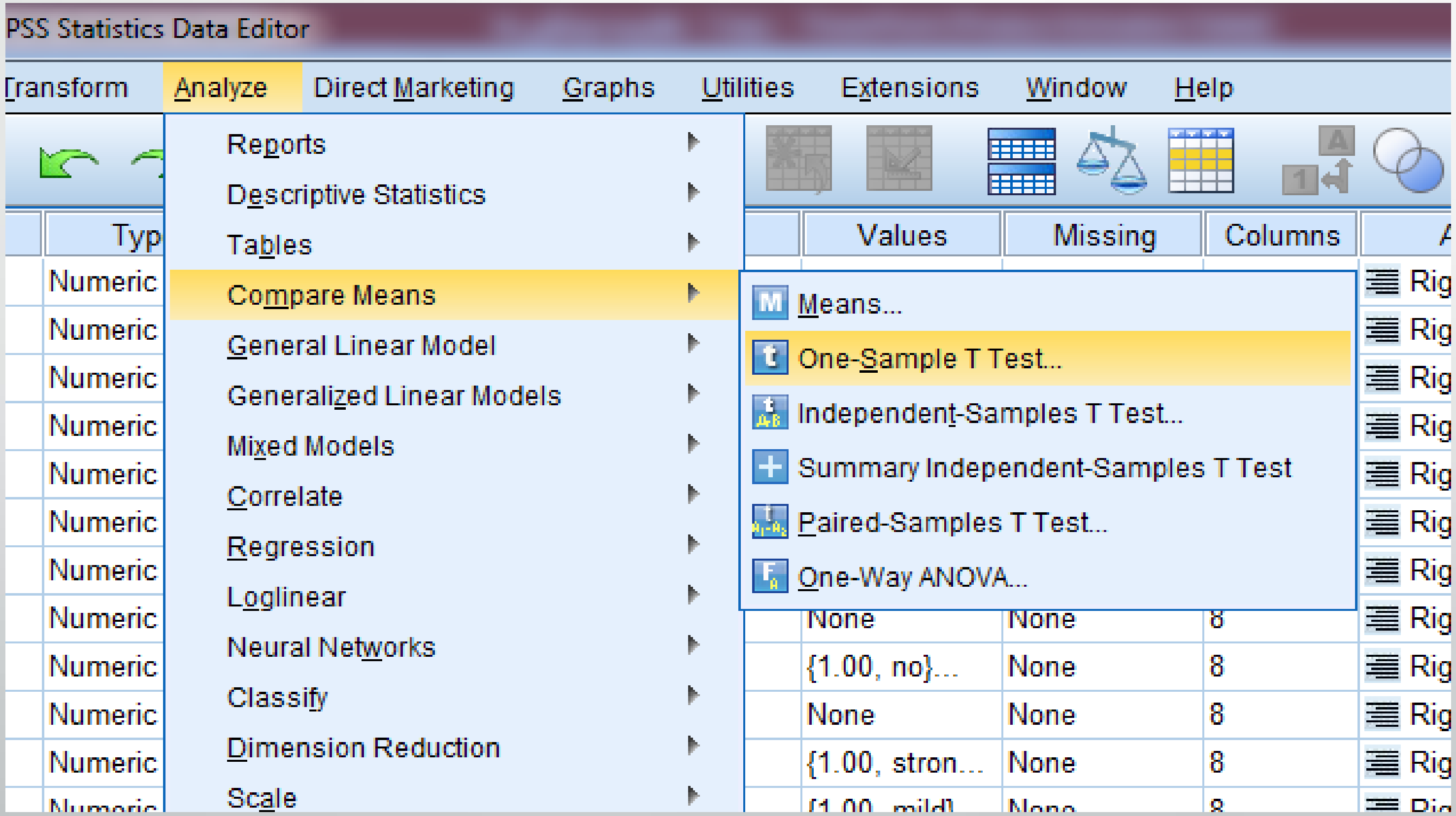
Data Requirements

- **Test variable that is continuous**
- **Scores on the test variable are independent**
- **Random sample of data from the population**
- **Normal distribution (approximately) of the sample and population on the test variable (Among moderate or large samples, a violation of normality may still yield accurate *p* values)**
- **Homogeneity of variances (i.e., variances approximately equal in both the sample and population)**
- **No outliers**

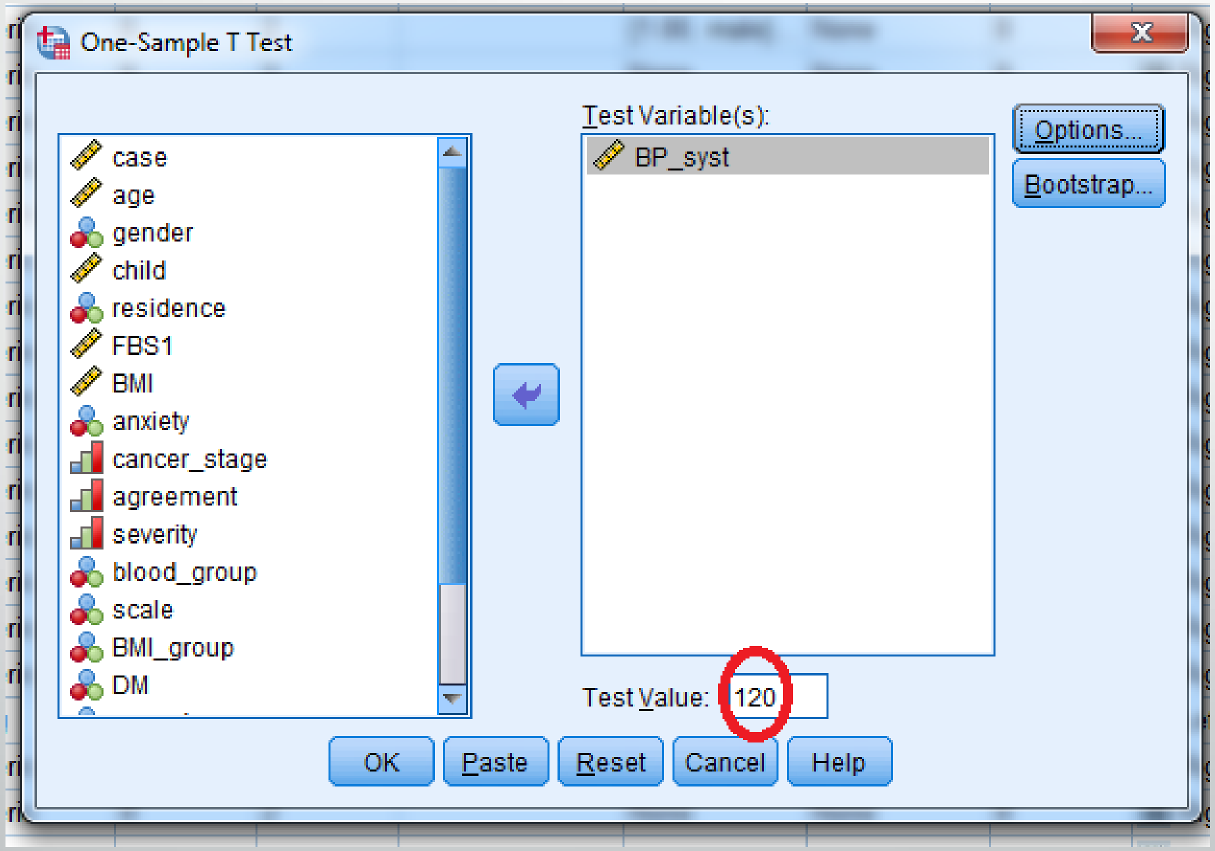
Hypotheses

- **The null hypothesis (*H_0*) and (two-tailed) alternative hypothesis (*H_1*) of the one sample *T* test can be expressed as:**
- **H_0: $\mu = \mu_0$ ("the population mean is equal to the [proposed] population mean")**
  H_1: $\mu \neq \mu_0$ ("the population mean is not equal to the [proposed] population mean")

Running the test



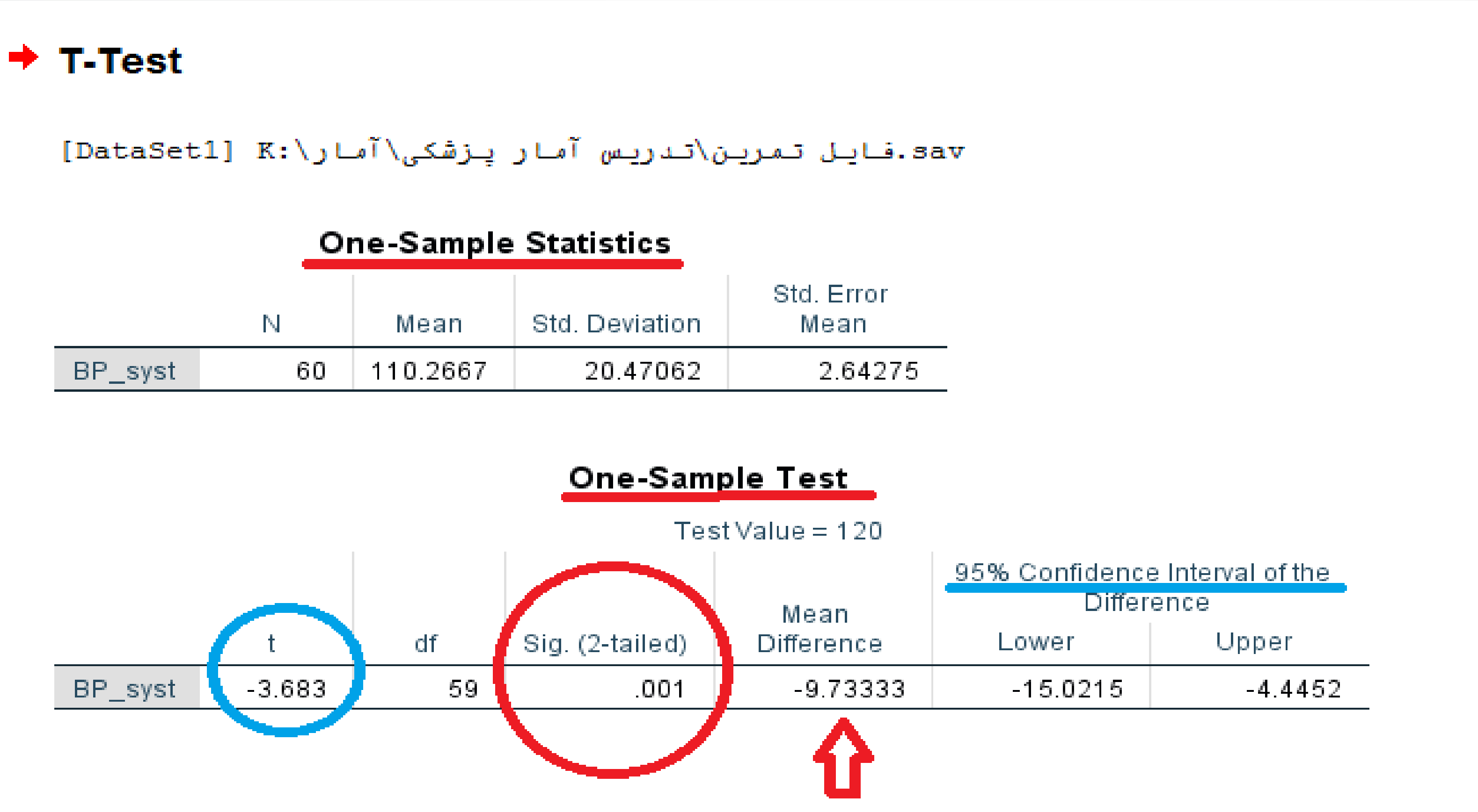
Running the test (cont.)



Test Variable(s): BP_syst

Test Value: 120

Output



T-Test

[DataSet1] K:\آمار\تدریس آمار پزشکی\فایل تمرین.sav

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
BP_syst	60	110.2667	20.47062	2.64275

One-Sample Test

	Test Value = 120					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
BP_syst	-3.683	59	.001	-9.73333	-15.0215	-4.4452

Dr. Anahita Babak

Decision and Conclusions

- Our hypothesized population value was 120 mmHg (the average BP of the overall adult population)
- Our sample value was 110.26 mmHg
- Since *p* = 0.001, we reject the null hypothesis that the mean BP of the sample is equal to the hypothesized population mean
- We conclude that the mean BP is significantly different than 120 mmHg

Based on the results, we can state the following:

- **There is a significant difference in the mean BP of the sample and the overall adult population (*p* = 0 .001)**
- **The average BP of the sample is about 9.74 mmHg lower than the adult population average [95% CI (-15) - (-4.4)]**

Paired samples t-test

- Compares two means that are from the same individual, object, or related units.
- The two means can represent things like:

❖ A measurement taken at two different times (e.g., pre-test and post-test with an intervention administered between the two time points)

❖ A measurement taken under two different conditions (e.g., completing a test under a "control" condition and an "experimental" condition)

❖ Measurements taken from two halves or sides of a subject or experimental unit (e.g., measuring hearing loss in a subject's left and right ears).

Paired samples t-test cont.

- **The purpose of the test is to determine whether there is statistical evidence that the mean difference between paired observations is significantly different from zero**
- **This test is also known as: Dependent *t* Test, Paired *t* Test, Repeated Measures *t* Test**
- **The variable used in this test is known as:**
- **Dependent variable, or test variable (continuous), measured at two different times or for two related conditions or units**

Common Uses

- Statistical difference between two time points
- Statistical difference between two conditions
- Statistical difference between two measurements
- Statistical difference between a matched pair

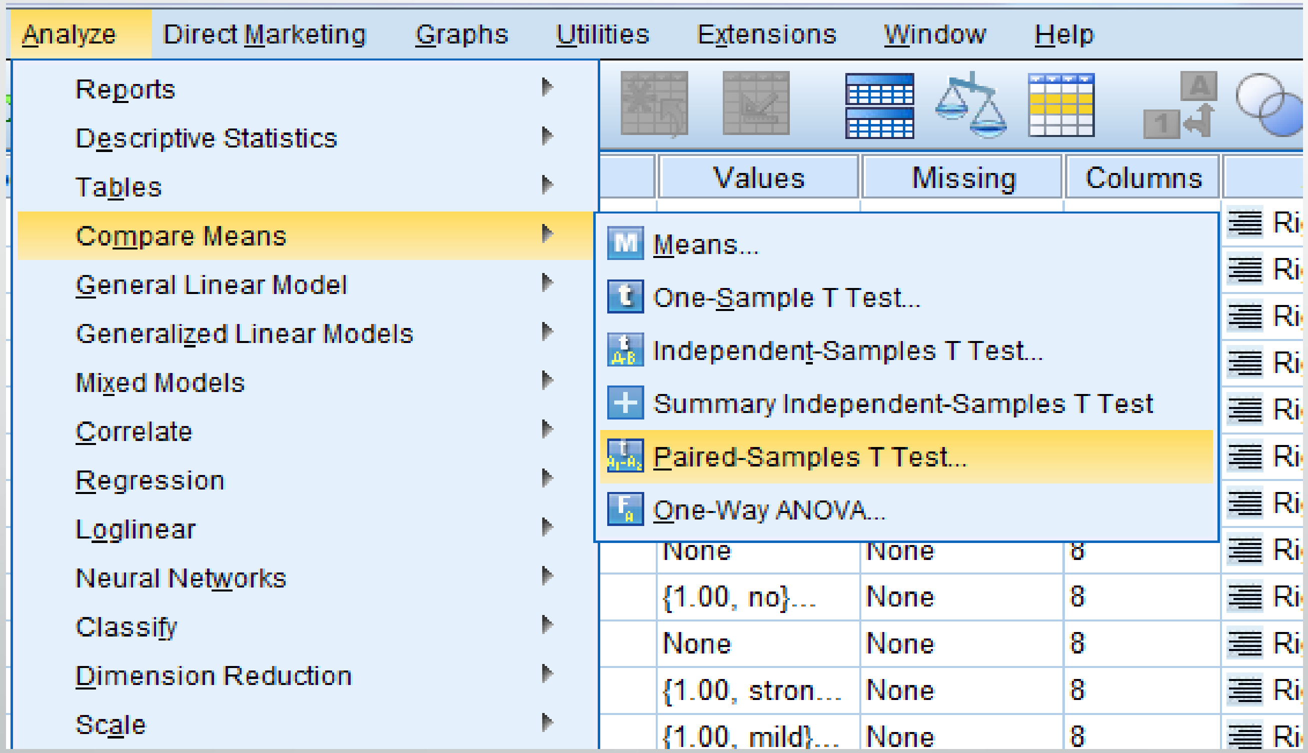
Data Requirements

- Dependent variable that is continuous (i.e., interval or ratio level)
  - **Note:** The paired measurements must be recorded in two separate variables.
- Related samples/groups (i.e., dependent observations)
  - The subjects in each sample, or group, are the same. This means that the subjects in the first group are also in the second group.
- Random sample of data from the population
- Normal distribution (approximately) of the difference between the paired values
- No outliers in the difference between the two related groups

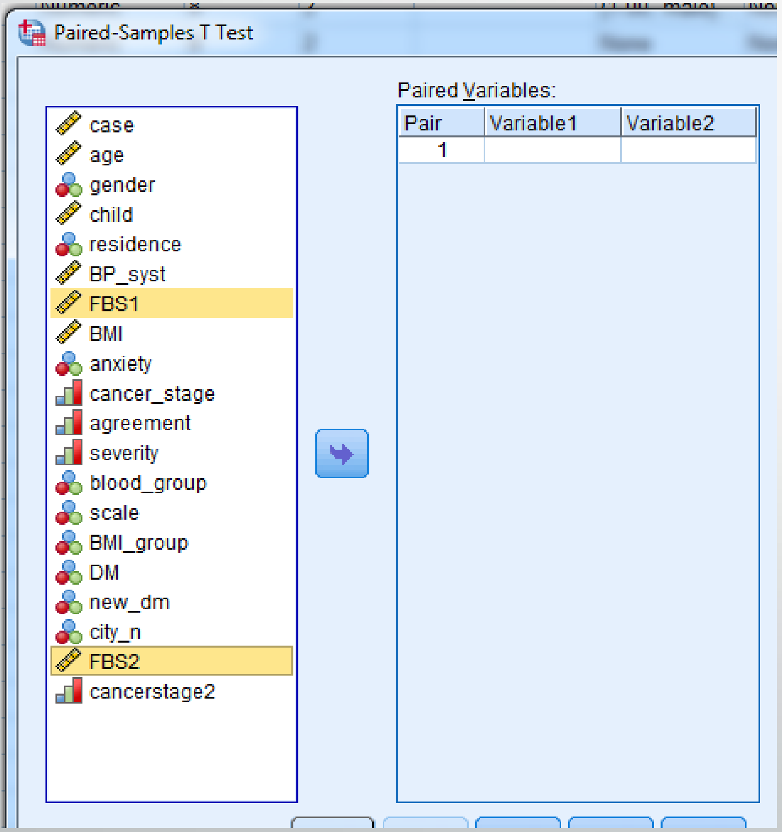
Hypotheses

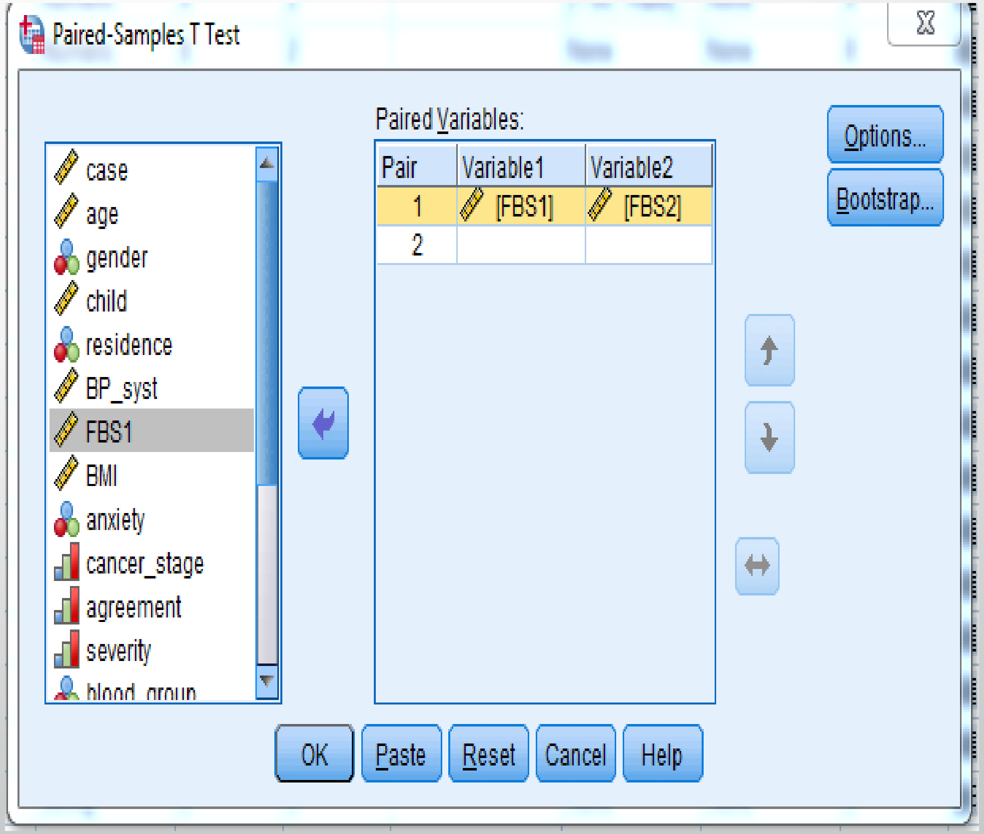
- H_0: $\mu_1 = \mu_2$ ("the paired population means are equal")
  H_1: $\mu_1 \neq \mu_2$ ("the paired population means are not equal")
- OR
- H_0: $\mu_1 - \mu_2 = 0$ ("the difference between the paired population means is equal to 0")
  H_1: $\mu_1 - \mu_2 \neq 0$ ("the difference between the paired population means is not 0")

Running the test

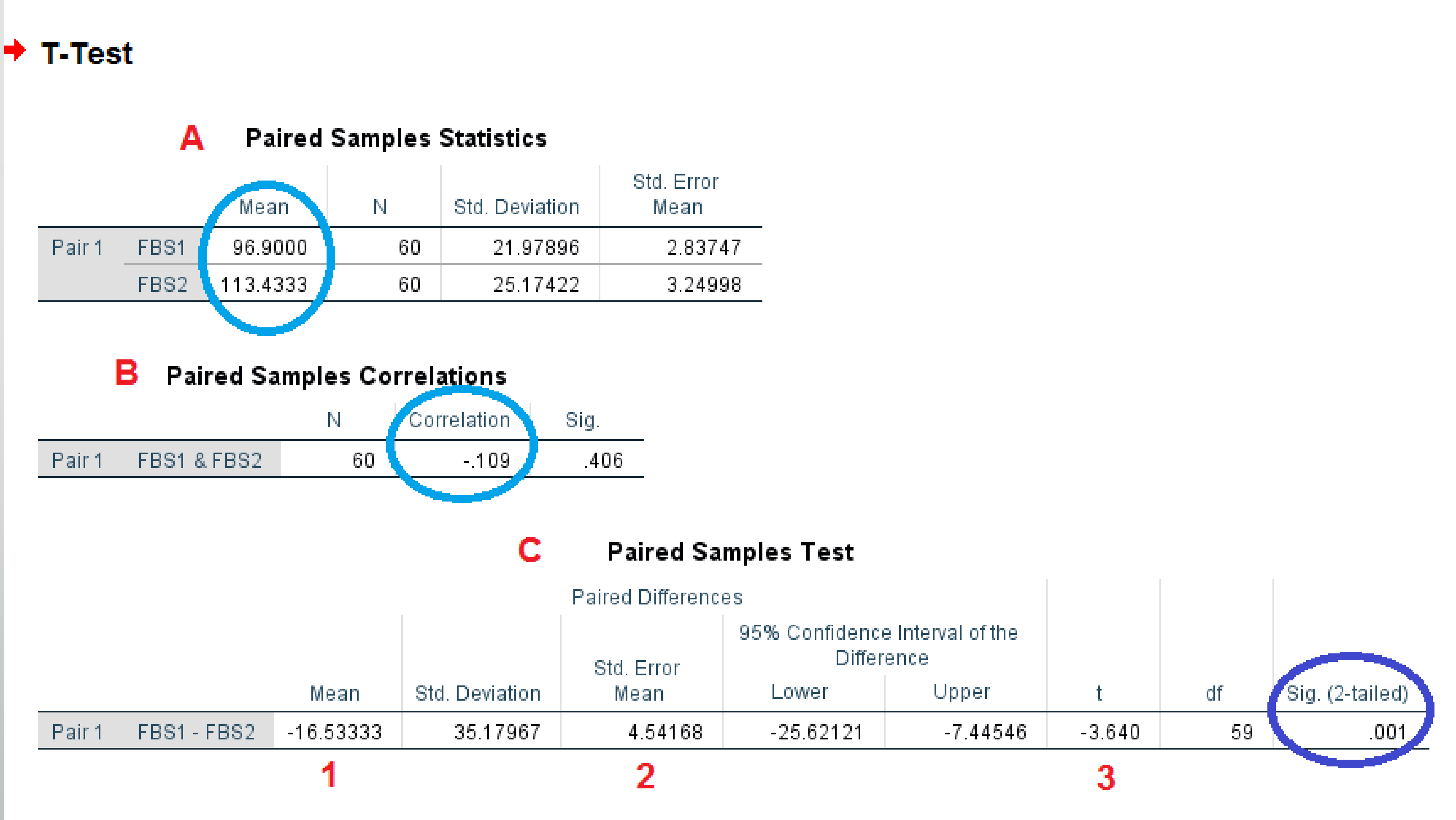


Running the test (cont.)





Output



T-Test

A Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	FBS1	96.9000	60	21.97896	2.83747
	FBS2	113.4333	60	25.17422	3.24998

B Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	FBS1 & FBS2	60	-.109	.406

C Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	FBS1 - FBS2	-16.53333	35.17967	4.54168	-25.62121	-7.44546	-3.640	59	.001

Decision And Conclusions

- **FBS1 and FBS2 were weakly and negatively correlated (r = -0.109, p = 0.4).**
- **There was a significant average difference between FBS1 and FBS2 (p = 0.001).**
- **On average, FBS2 was 16.5 mg/dl higher than FBS1 [95% CI (-25.6, -7.44)]**

Independent Samples t Test

- Compares the means of two independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different
- is also known as: Independent *t* Test, Student *t* Test
- The variables used in this test are known as:

✓Dependent variable, or test variable

✓Independent variable, or grouping variable

Common Uses

- Statistical differences between the means of two groups
- Statistical differences between the means of two interventions
- Statistical differences between the means of two change scores

Data Requirements

- Dependent variable that is continuous (i.e., interval or ratio level)
- Independent variable that is categorical (i.e., two groups)
- Cases that have values on both the dependent and independent variables
- Independent samples/groups (i.e., independence of observations)
- Random sample of data from the population
- Normal distribution (approximately) of the dependent variable for each group
- Homogeneity of variances (i.e., variances approximately equal across groups)
- No outliers

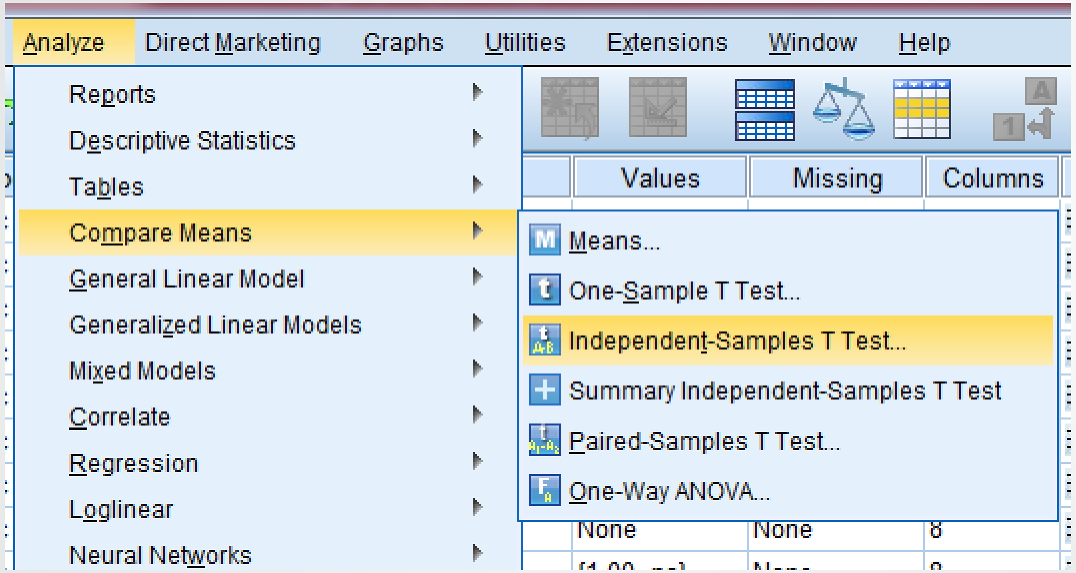
Hypotheses

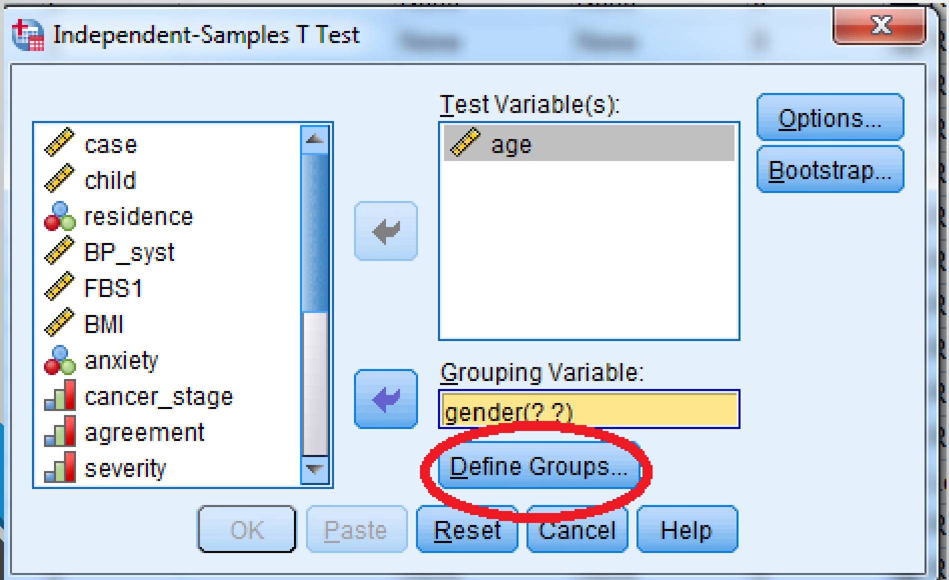
- H_0: $\mu_1 = \mu_2$ ("the two population means are equal")

  H_1: $\mu_1 \neq \mu_2$ ("the two population means are not equal")
- OR
- H_0: $\mu_1 - \mu_2 = 0$ ("the difference between the two population means is equal to 0")

  H_1: $\mu_1 - \mu_2 \neq 0$ ("the difference between the two population means is not 0")

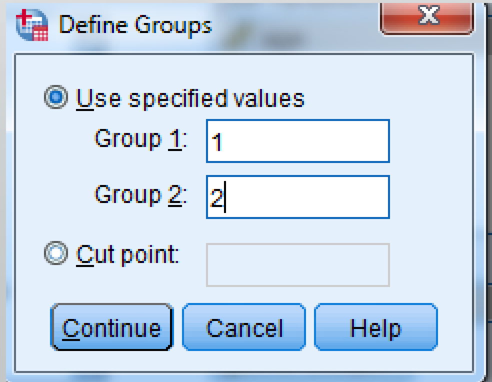
Running the test





Test Variable(s): age

Grouping Variable: gender(? ?)



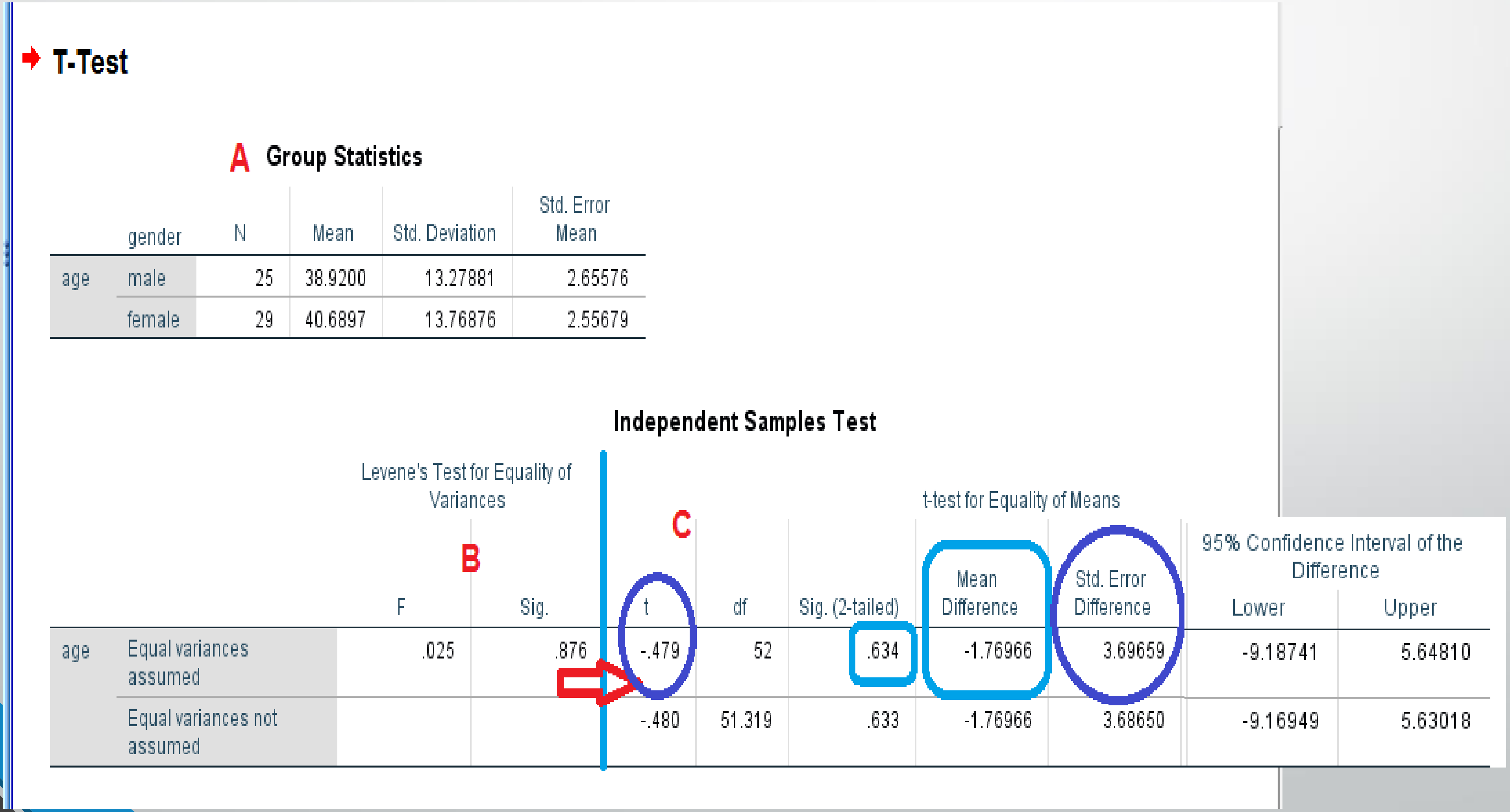
● Use specified values

Group 1: 1

Group 2: 2

○ Cut point:

Output



T-Test

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
age	male	25	38.9200	13.27881	2.65576
	female	29	40.6897	13.76876	2.55679

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
age	Equal variances assumed	.025	.876	-.479	52	.634	-1.76966	3.69659	-9.18741	5.64810
	Equal variances not assumed			-.480	51.319	.633	-1.76966	3.68650	-9.16949	5.63018

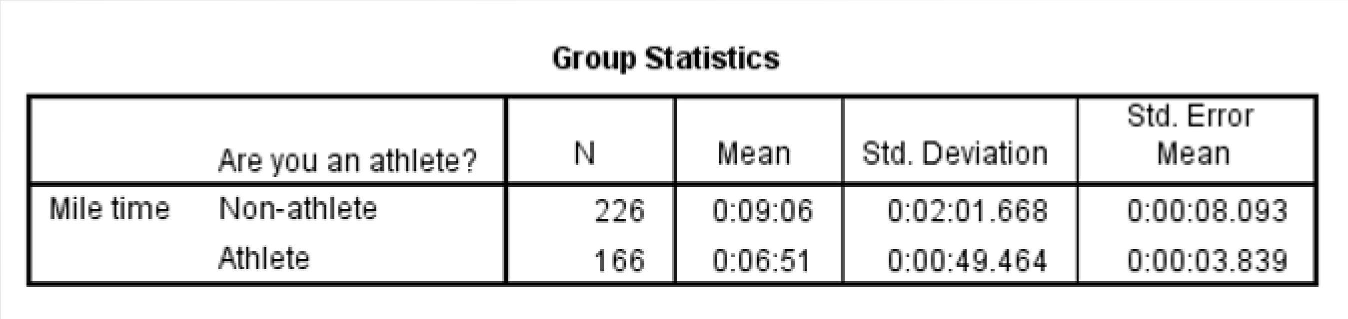
Decision and Conclusions

- **The mean age in female was 1.76 years greater than men**
- **But**
- **This difference was not significant (*p=0.63*).**

Levene's Test for Equality of Variances

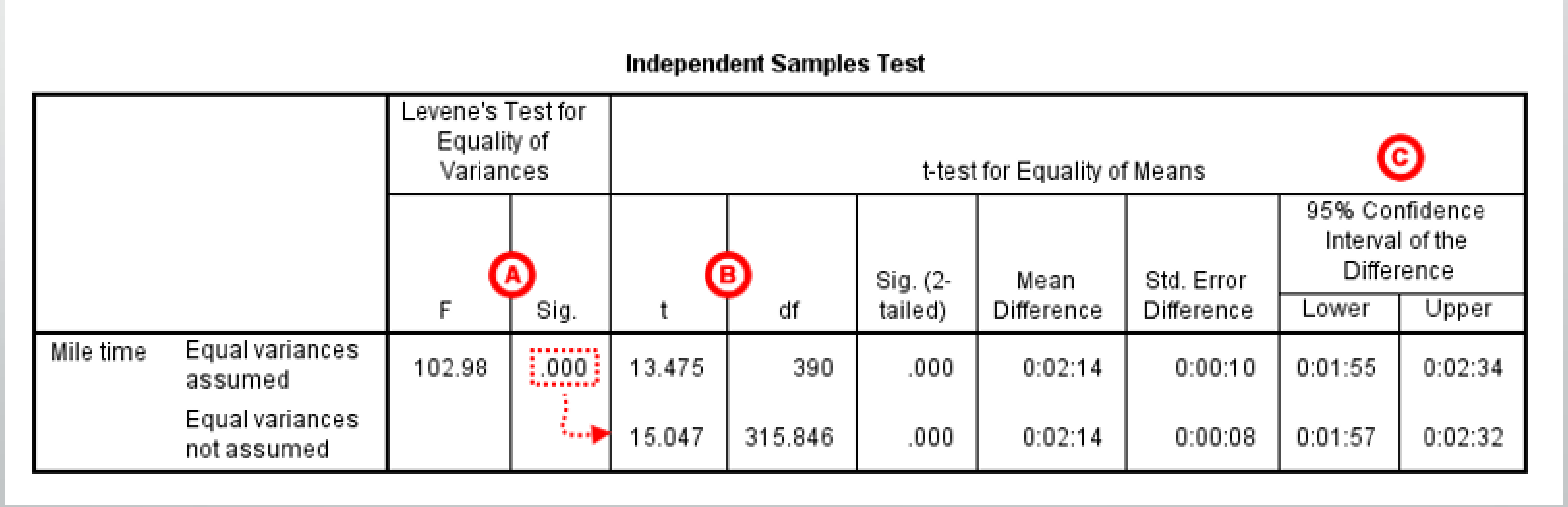
- **The hypotheses for Levene's test are:**
- **H_0: $\sigma_1^2 - \sigma_2^2 = 0$ ("the population variances of group 1 and 2 are equal")**
  H_1: $\sigma_1^2 - \sigma_2^2 \neq 0$ ("the population variances of group 1 and 2 are not equal")
- **This implies that if we reject the null hypothesis of Levene's Test, it suggests that the variances of the two groups are not equal; i.e., that the homogeneity of variances assumption is violated.**

Output



Group Statistics

	Are you an athlete?	N	Mean	Std. Deviation	Std. Error Mean
Mile time	Non-athlete	226	0:09:06	0:02:01.668	0:00:08.093
	Athlete	166	0:06:51	0:00:49.464	0:00:03.839



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Mile time	Equal variances assumed	102.98	.000	13.475	390	.000	0:02:14	0:00:10	0:01:55	0:02:34
	Equal variances not assumed			15.047	315.846	.000	0:02:14	0:00:08	0:01:57	0:02:32

Interpretation

- The *p*-value of Levene's test is printed as ".000" (but should be read as $p < 0.001$ -- i.e., *p* very small), so we we reject the null of Levene's test and conclude that the variance in mile time of athletes is significantly different than that of non-athletes. This tells us that we should look at the "Equal variances not assumed" row for the *t* -test (and corresponding confidence interval) results.

Decision and Conclusions

- **There was a significant difference in mean mile time between non-athletes and athletes ($t_{315.846}$ = 15.047, p < .001).**
- **The average mile time for athletes was 2 minutes and 14 seconds faster than the average mile time for non-athletes**

One-Way ANOVA

- The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

- The variables used in this test are known as:

✓ Dependent variable

✓ Independent variable (also known as the grouping variable, or *factor*)

  - This variable divides cases into two or more mutually exclusive *levels*, or groups

Common Uses

- Statistical differences among the means of two or more groups
- Statistical differences among the means of two or more interventions
- Statistical differences among the means of two or more change scores

Assumptions

- dependent variable: measured at the continuous level
- independent variable: 3 or more categorical, independent groups
- independence of observations
- no significant outliers
- dependent variable: approximately normally distributed for each category of the independent variable
- homogeneity of variances

Hypotheses

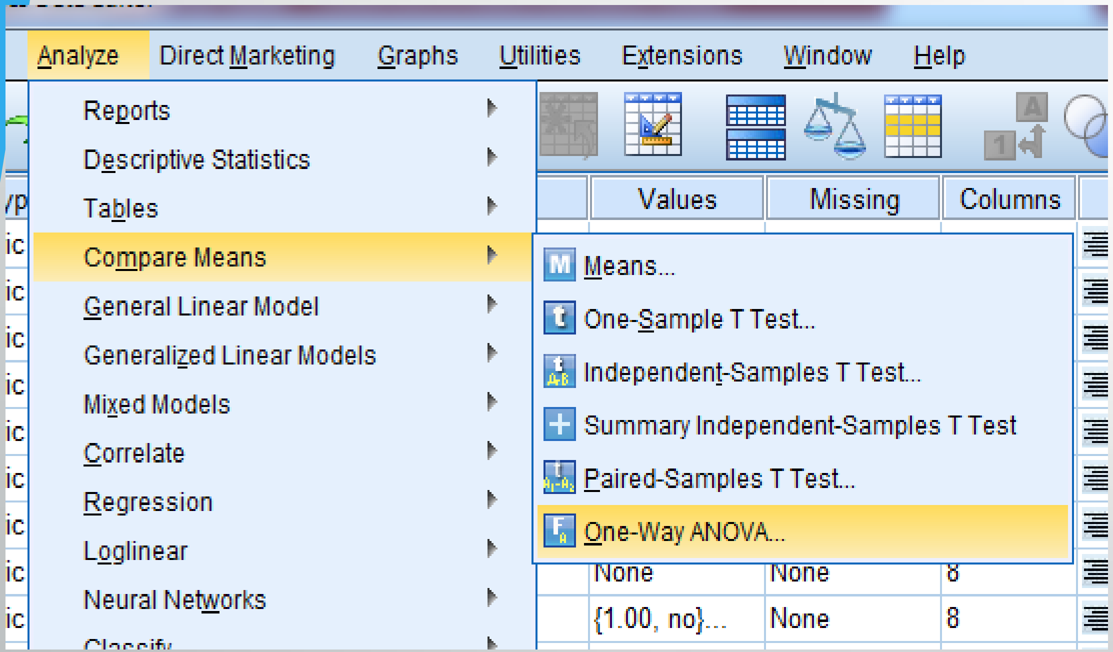
- H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ ("all *k* population means are equal")

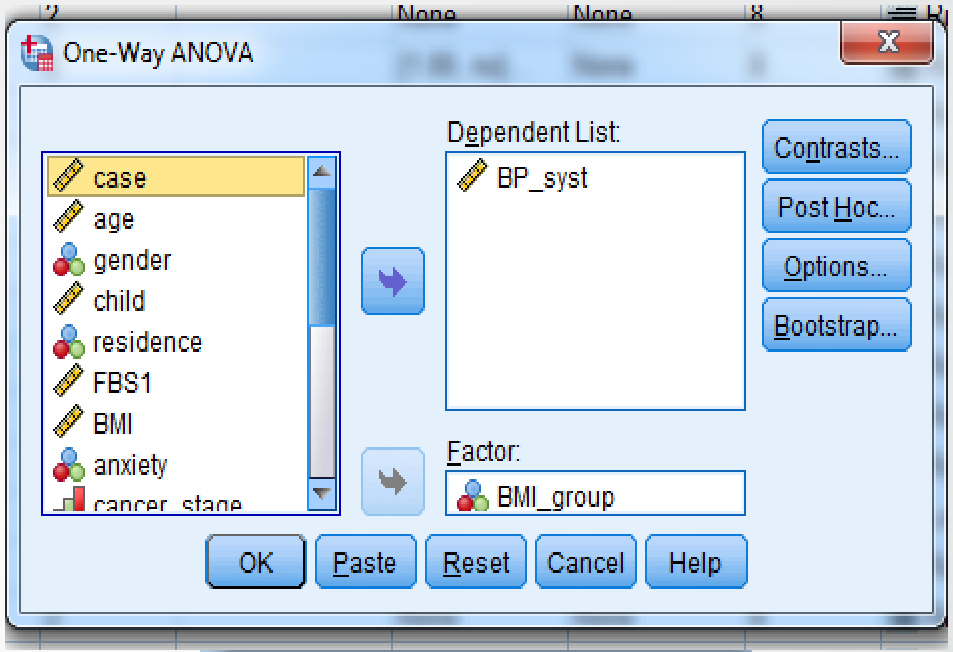
  H_1: At least one μ_i different ("at least one of the *k* population means is not equal to the others")

- where

- μ_i is the population mean of the i^{th} group ($i = 1, 2, \ldots, k$)

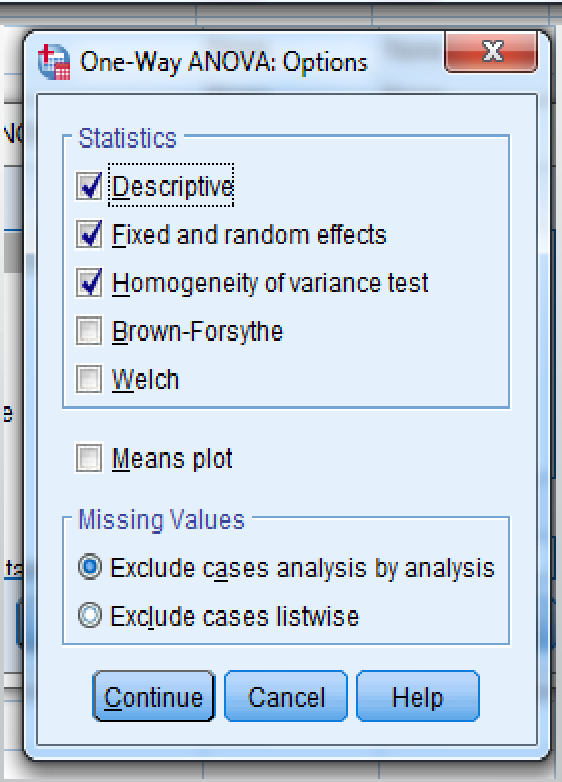
Run the test





Dependent List: BP_syst

Factor: BMI_group



Statistics

☑ Descriptive

☑ Fixed and random effects

☑ Homogeneity of variance test

☐ Brown-Forsythe

☐ Welch

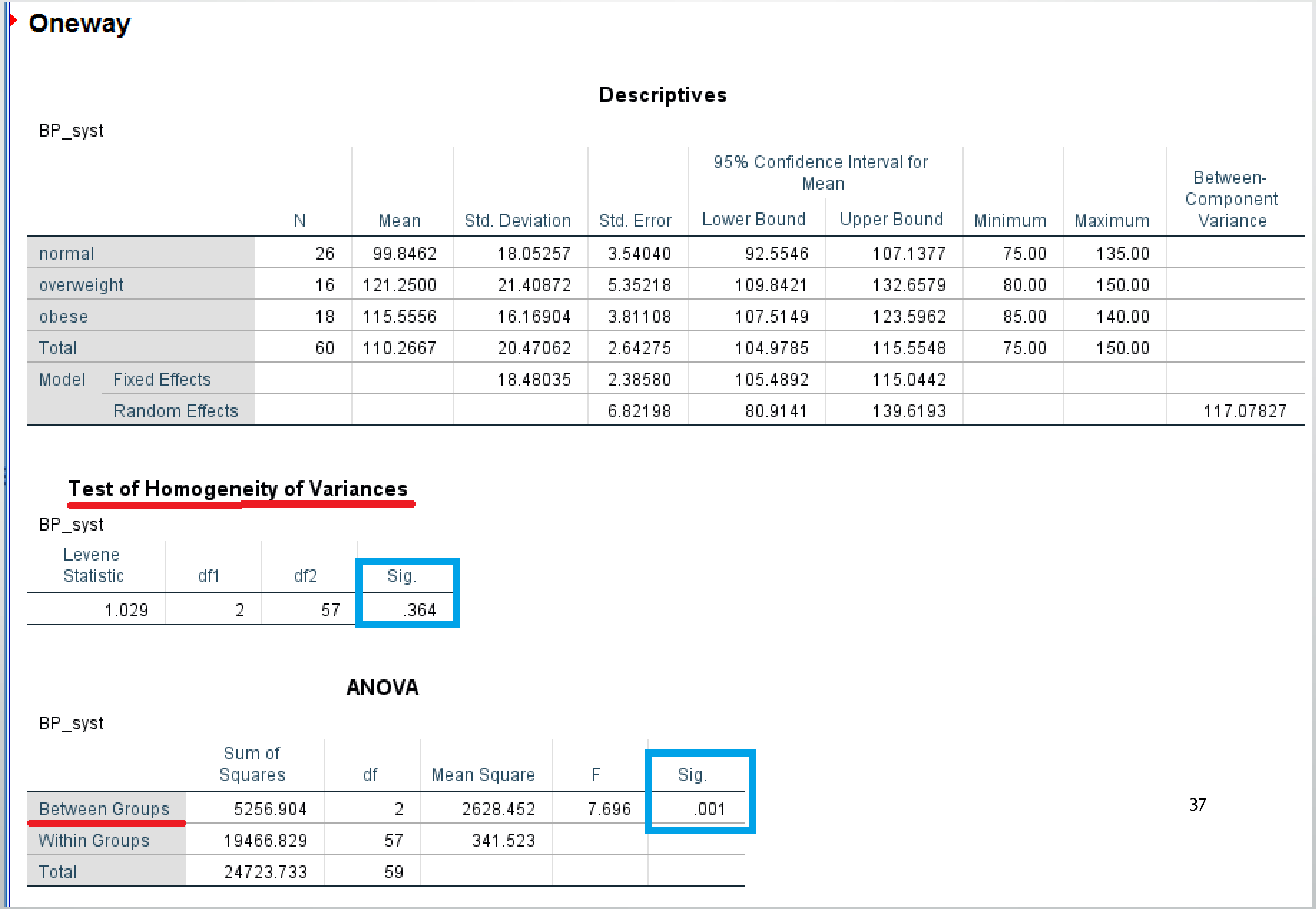
☐ Means plot

Missing Values

● Exclude cases analysis by analysis

○ Exclude cases listwise

Output



Oneway

Descriptives

BP_syst

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
normal		26	99.8462	18.05257	3.54040	92.5546	107.1377	75.00	135.00	
overweight		16	121.2500	21.40872	5.35218	109.8421	132.6579	80.00	150.00	
obese		18	115.5556	16.16904	3.81108	107.5149	123.5962	85.00	140.00	
Total		60	110.2667	20.47062	2.64275	104.9785	115.5548	75.00	150.00	
Model	Fixed Effects			18.48035	2.38580	105.4892	115.0442			
	Random Effects				6.82198	80.9141	139.6193			117.07827

Test of Homogeneity of Variances

BP_syst

Levene Statistic	df1	df2	Sig.
1.029	2	57	.364

ANOVA

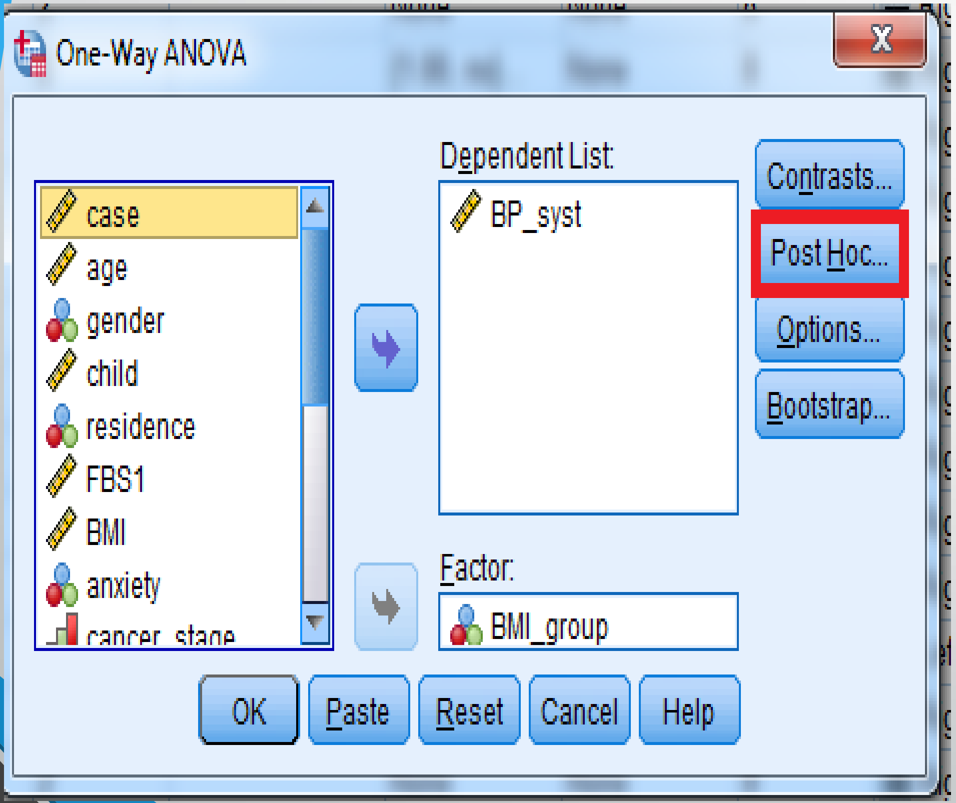
BP_syst

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5256.904	2	2628.452	7.696	.001
Within Groups	19466.829	57	341.523		
Total	24723.733	59			

Discussion and conclusions

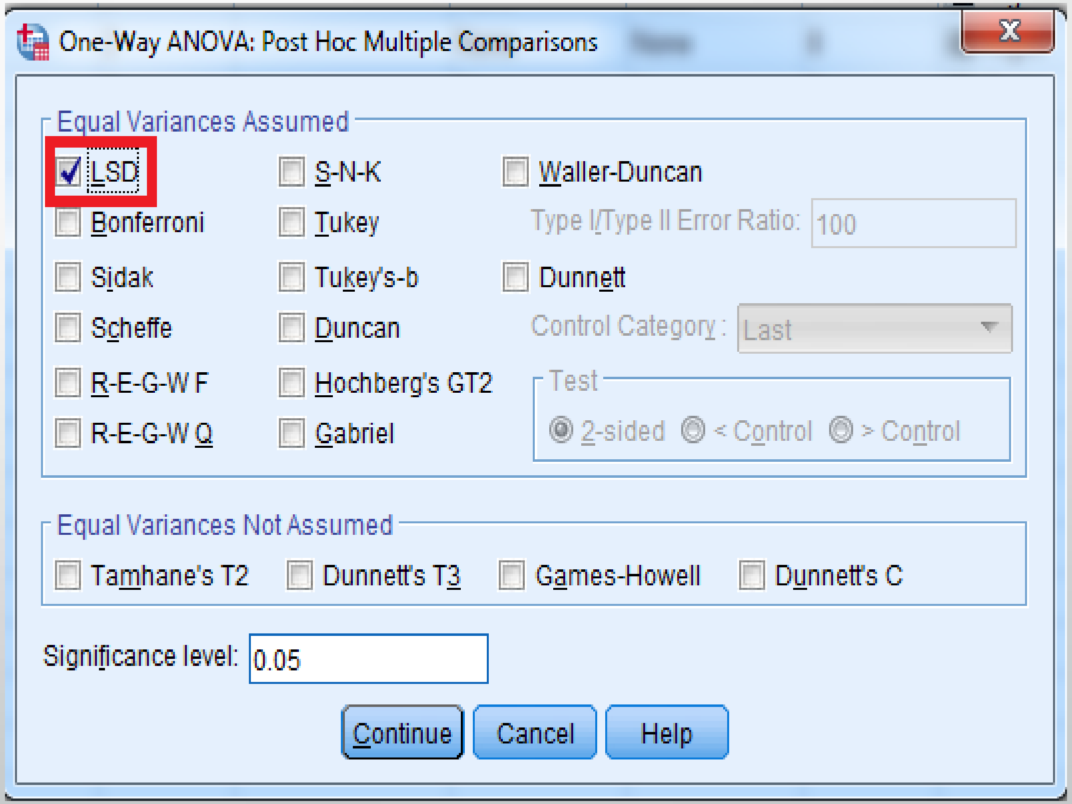
- **We conclude that the mean systolic BP is significantly different for at least one of the BMI groups**
- **Note that the ANOVA alone does not tell us specifically which means were different from one another. To determine that, we would need to follow up with *multiple comparisons* (or *post-hoc*) tests.**

Post-Hoc test



Dependent List: BP_syst

Factor: BMI_group



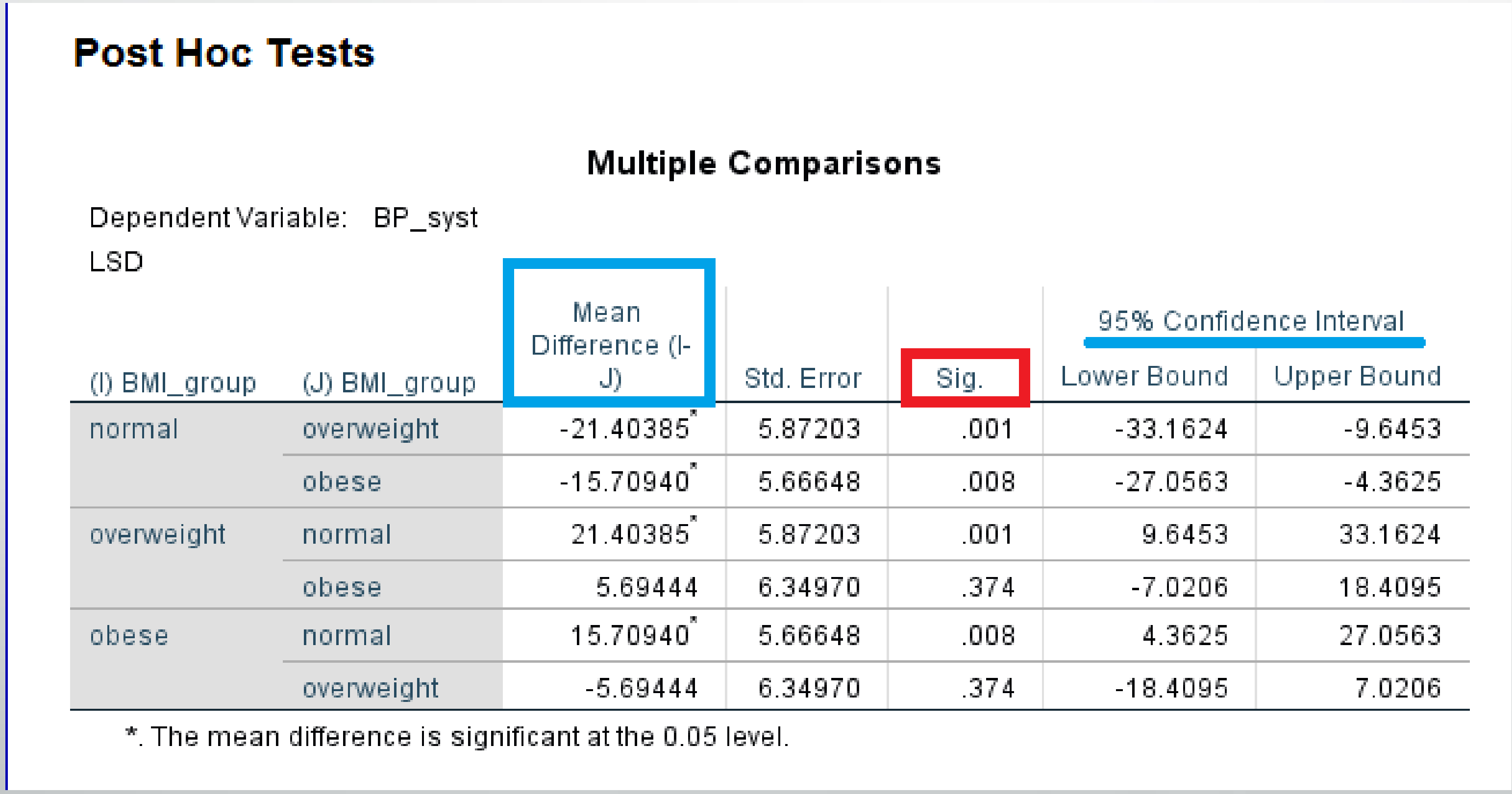
☑ LSD

Type I/Type II Error Ratio: 100

Control Category : Last

Significance level: 0.05

Output



Post Hoc Tests

Multiple Comparisons

Dependent Variable: BP_syst

LSD

(I) BMI_group	(J) BMI_group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
normal	overweight	-21.40385*	5.87203	.001	-33.1624	-9.6453
	obese	-15.70940*	5.66648	.008	-27.0563	-4.3625
overweight	normal	21.40385*	5.87203	.001	9.6453	33.1624
	obese	5.69444	6.34970	.374	-7.0206	18.4095
obese	normal	15.70940*	5.66648	.008	4.3625	27.0563
	overweight	-5.69444	6.34970	.374	-18.4095	7.0206

*. The mean difference is significant at the 0.05 level.

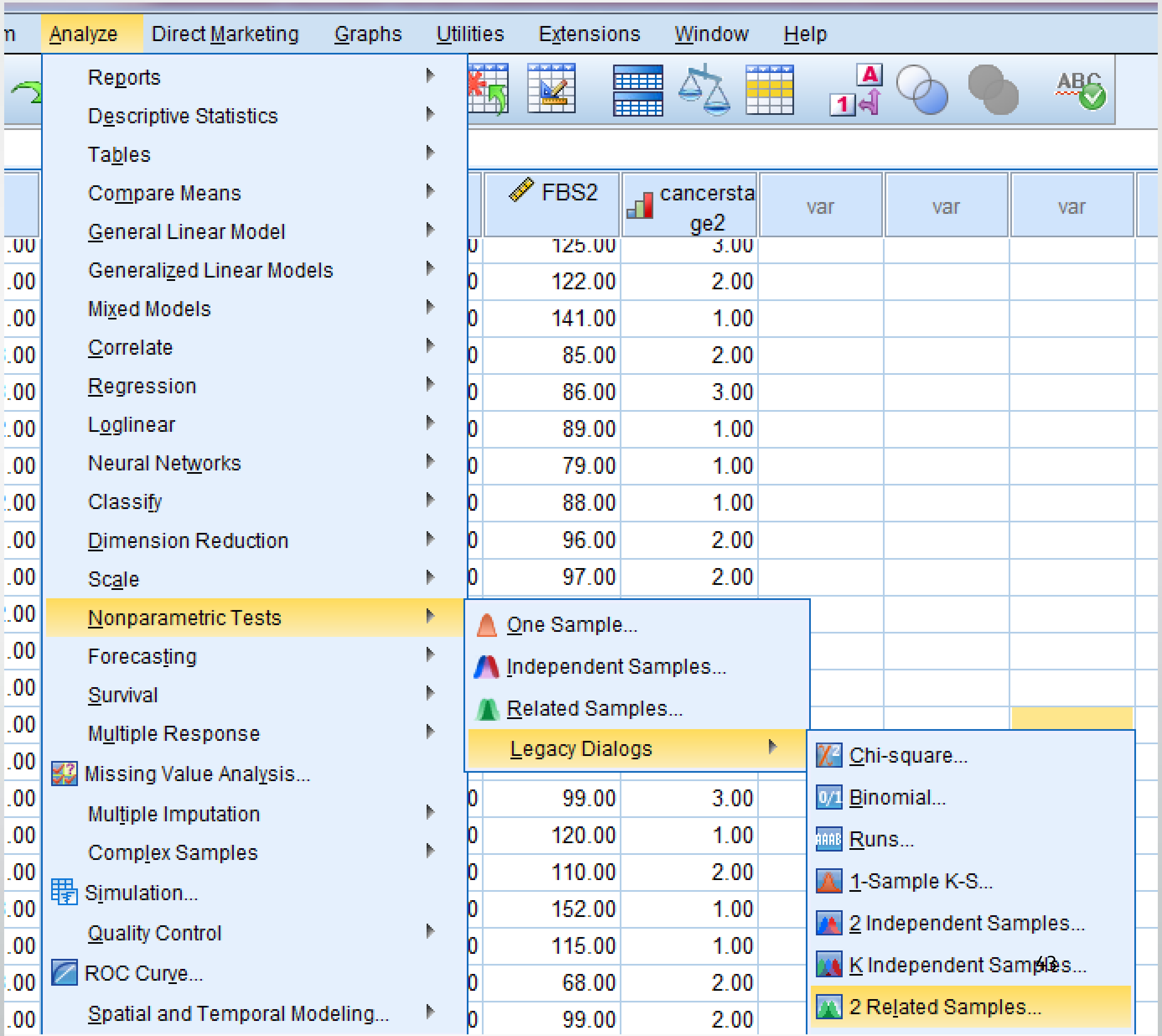
Wilcoxon signed-rank test

- **The Wilcoxon signed rank test should be used if the differences between pairs of data are non-normally distributed.**
- **The null hypothesis for this test is that the medians of two samples are equal.**
- **This test is parametric equivalent of paired dependent samples t-test**

Assumptions

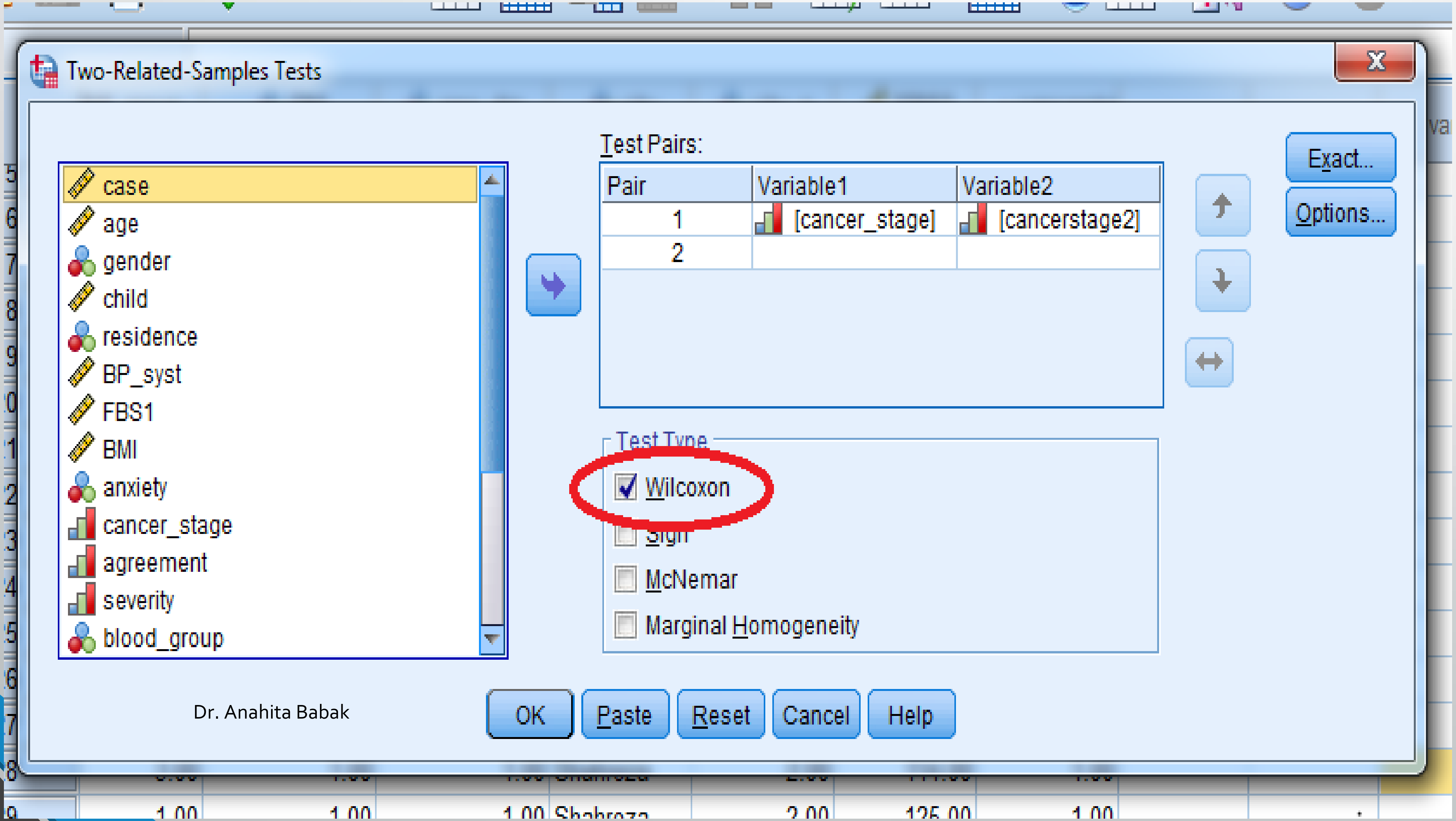
- The dependent variable should be measured at the ordinal or continuous level (Likert items, IQ score, VAS 0-10)
- The independent variable should consist of two categorical, "related groups" or "matched pairs". "Related groups" indicates that the same subjects are present in both groups

Running the test



Dr. Anahita Babak

Running the test (cont.)



Two-Related-Samples Tests

Test Pairs:

Pair	Variable1	Variable2
1	[cancer_stage]	[cancerstage2]
2		

Test Type

☑ Wilcoxon

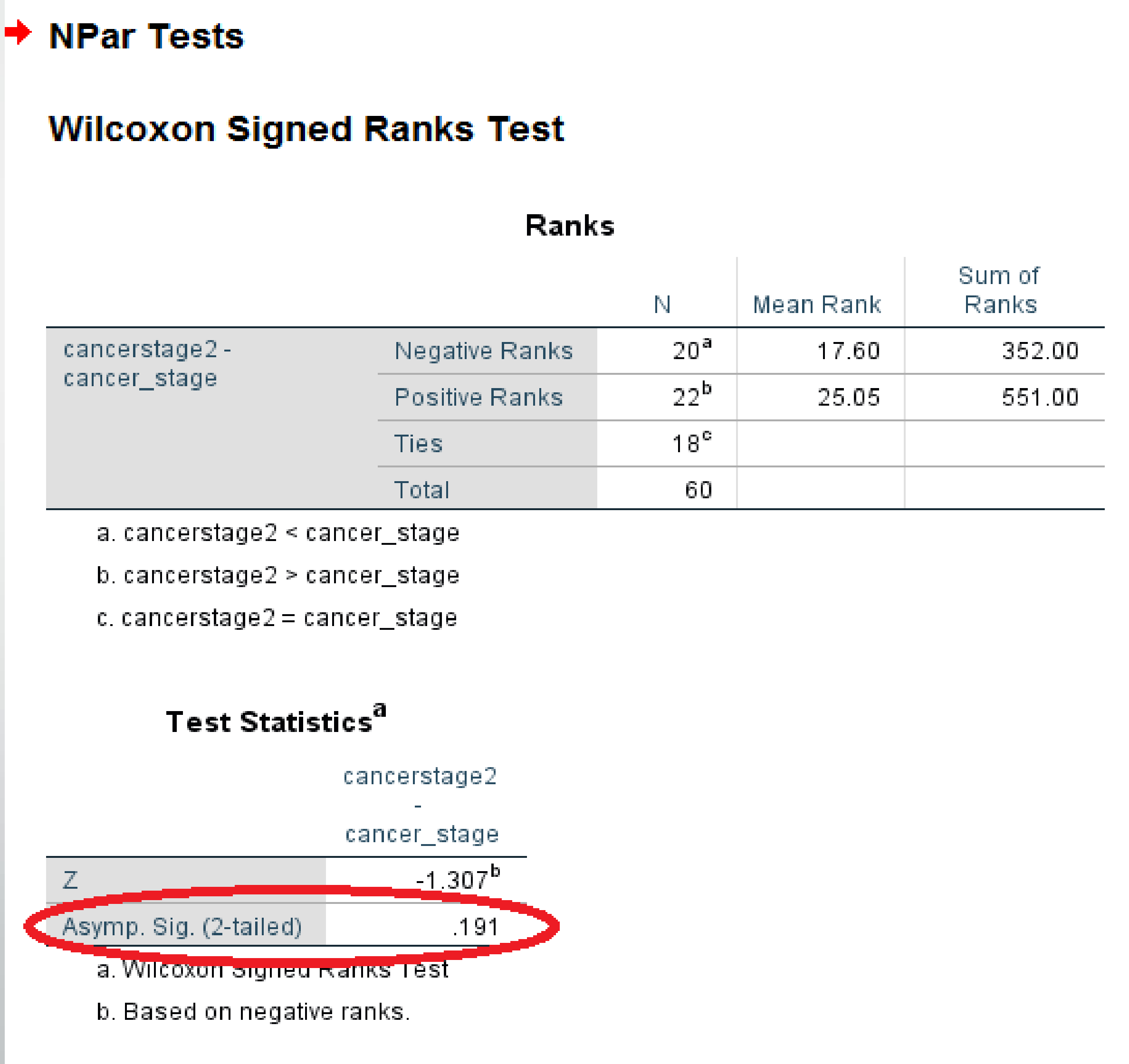
☐ Sign

☐ McNemar

☐ Marginal Homogeneity

Dr. Anahita Babak

Output



NPar Tests

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
cancerstage2 - cancer_stage	Negative Ranks	20[a]	17.60	352.00
	Positive Ranks	22[b]	25.05	551.00
	Ties	18[c]		
	Total	60		

a. cancerstage2 < cancer_stage

b. cancerstage2 > cancer_stage

c. cancerstage2 = cancer_stage

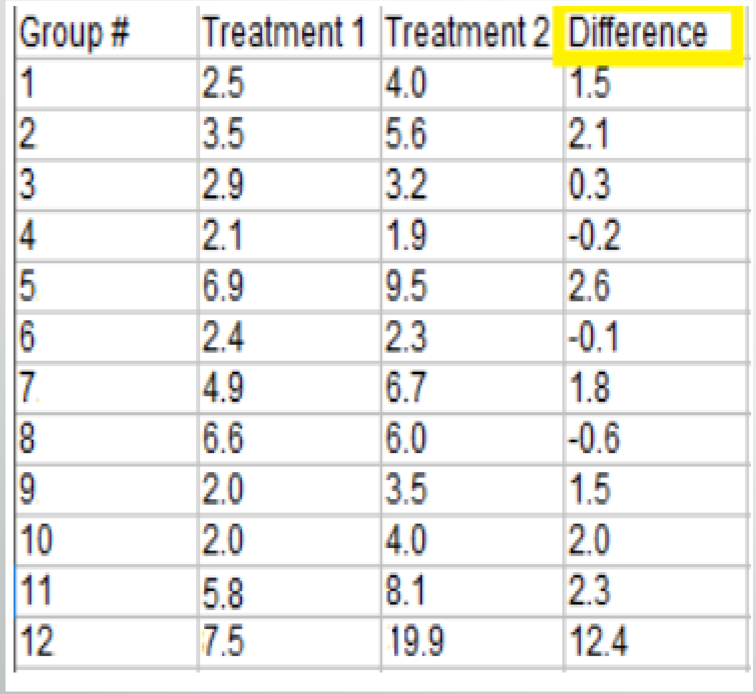
Test Statistics[a]

	cancerstage2 - cancer_stage
Z	-1.307[b]
Asymp. Sig. (2-tailed)	.191

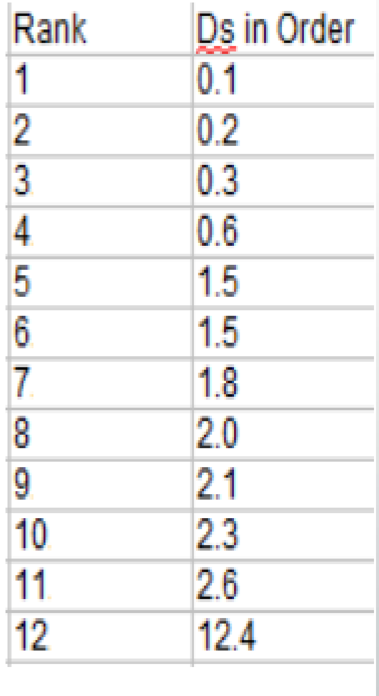
a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Rank



Group #	Treatment 1	Treatment 2	Difference
1	2.5	4.0	1.5
2	3.5	5.6	2.1
3	2.9	3.2	0.3
4	2.1	1.9	-0.2
5	6.9	9.5	2.6
6	2.4	2.3	-0.1
7	4.9	6.7	1.8
8	6.6	6.0	-0.6
9	2.0	3.5	1.5
10	2.0	4.0	2.0
11	5.8	8.1	2.3
12	7.5	19.9	12.4



Rank	Ds in Order
1	0.1
2	0.2
3	0.3
4	0.6
5	1.5
6	1.5
7	1.8
8	2.0
9	2.1
10	2.3
11	2.6
12	12.4

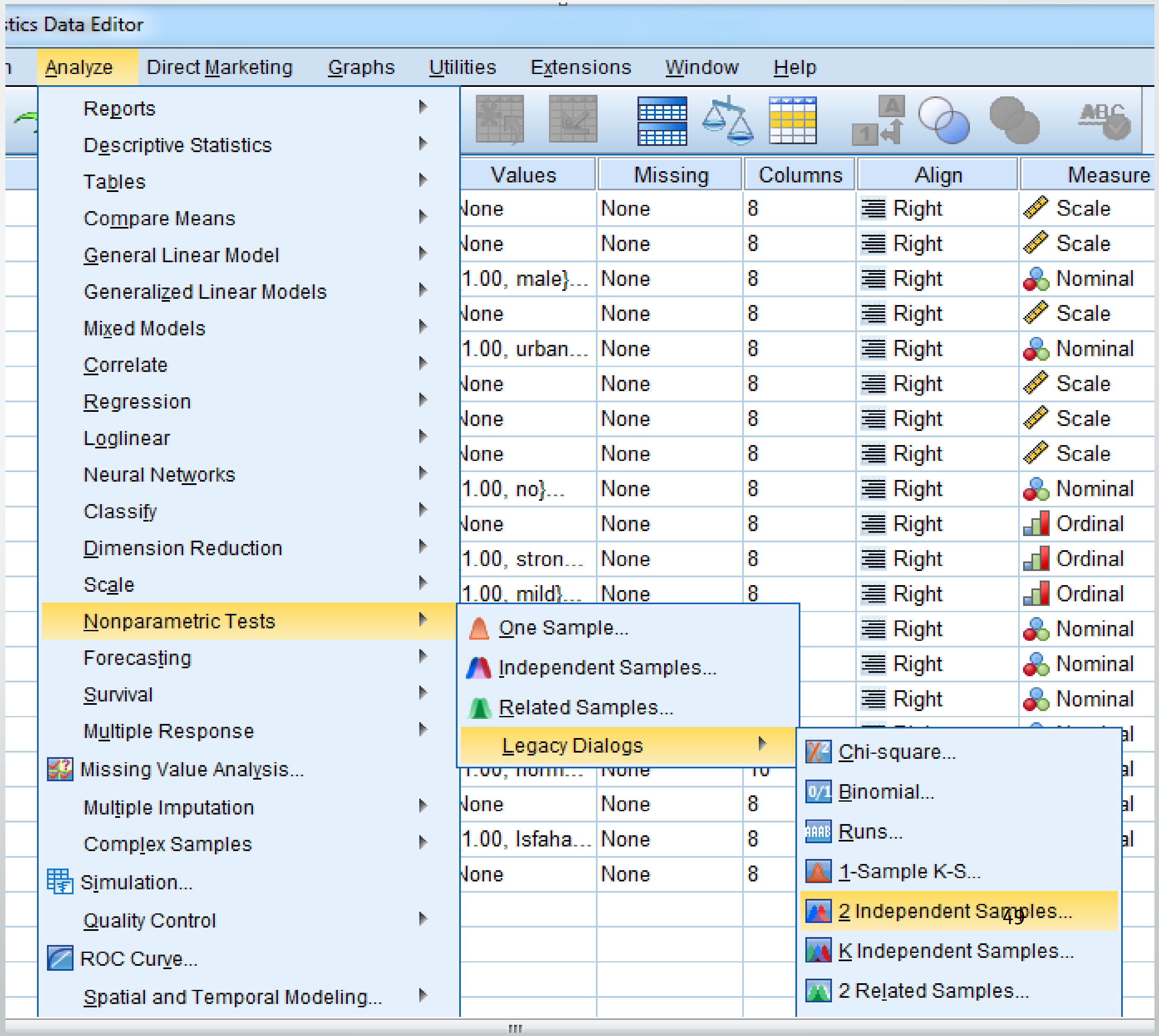
Mann Whitney U-test

- **The Mann-Whitney U test is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed**
- **Equivalent Parametric test: independent samples t-test**
- **The null hypothesis asserts that the *medians* of the two samples are identical.**

Assumptions

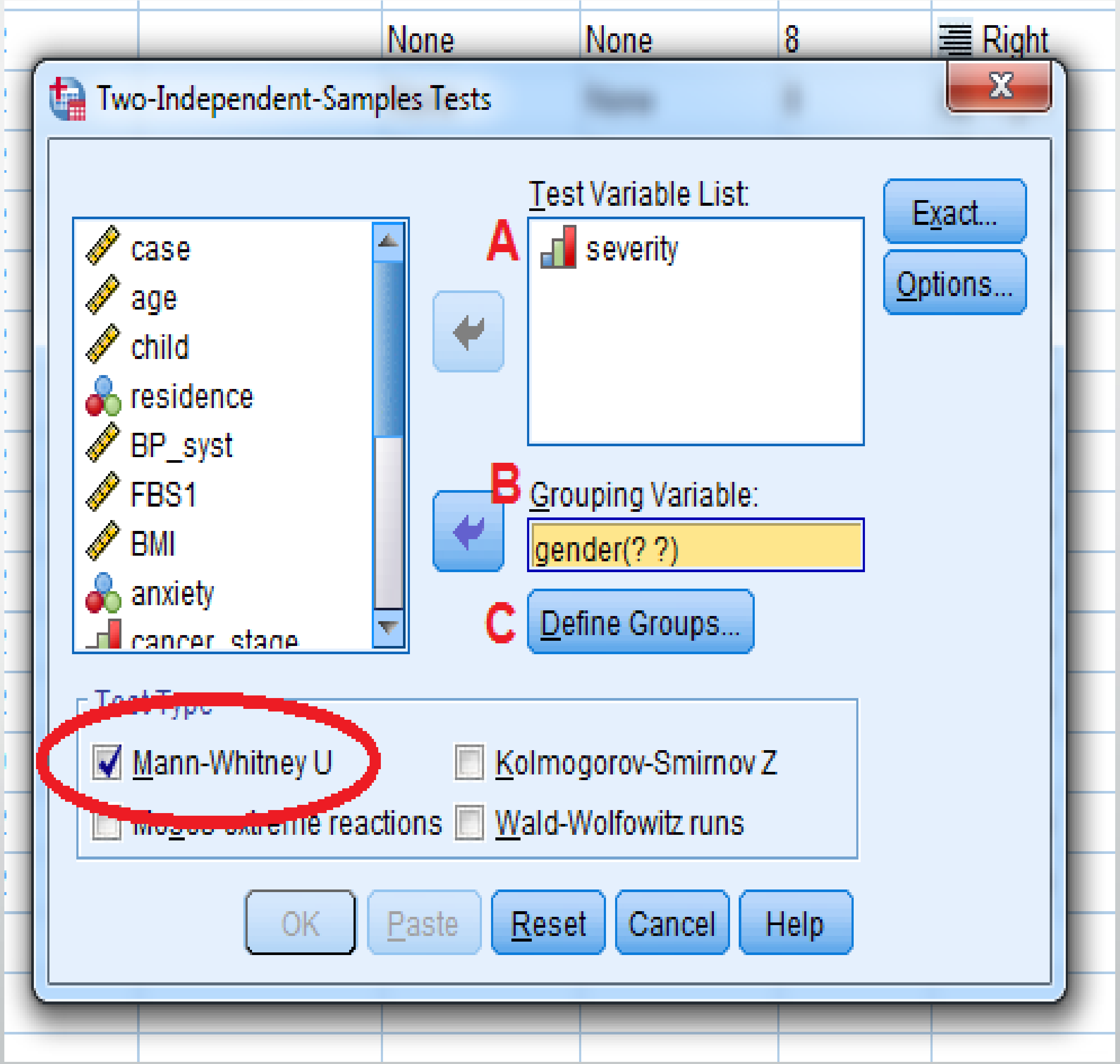
- dependent variable should be measured at the ordinal or continuous level (Likert items, IQ score…)
- independent variable should consist of two categorical, independent groups (gender, smoker…)
- independence of observations (there is no relationship between the observations in each group or between the groups themselves. For example, there must be different participants in each group with no participant being in more than one group.)
- A Mann-Whitney U test can be used when your two variables are not normally distributed

Running the test



Dr. Anahita Babak

Running the test (cont.)



Two-Independent-Samples Tests

Test Variable List: severity

Grouping Variable: gender(? ?)

Test Type

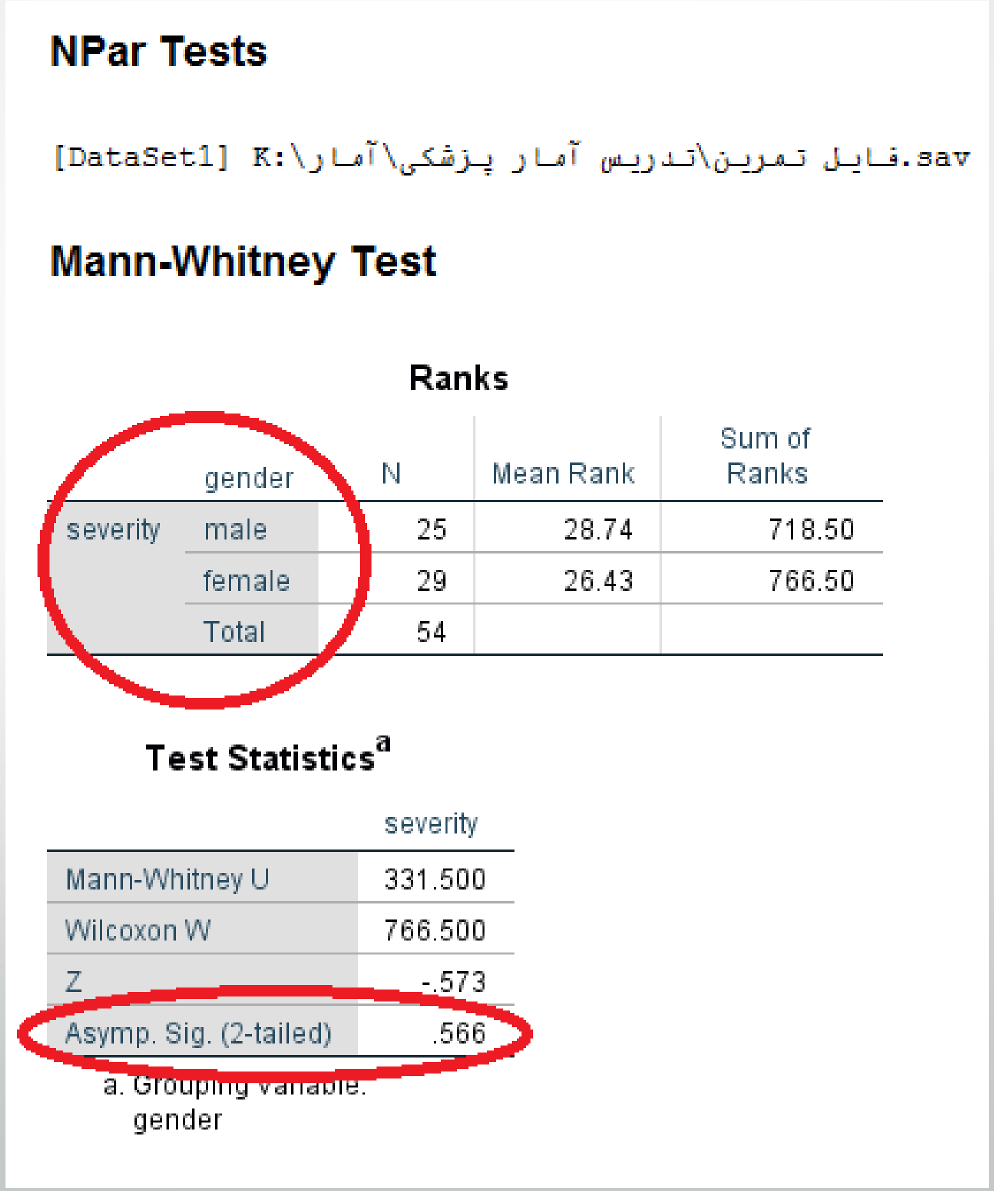
☑ Mann-Whitney U

☐ Kolmogorov-Smirnov Z

☐ Moses extreme reactions

☐ Wald-Wolfowitz runs

Output



NPar Tests

```
[DataSet1] K:\آمار\پزشکی آمار تدریس\تمرین فایل.sav
```

Mann-Whitney Test

Ranks

	gender	N	Mean Rank	Sum of Ranks
severity	male	25	28.74	718.50
	female	29	26.43	766.50
	Total	54		

Test Statistics[a]

	severity
Mann-Whitney U	331.500
Wilcoxon W	766.500
Z	-.573
Asymp. Sig. (2-tailed)	.566

a. Grouping Variable: gender

Kruskal-Wallis Test

- can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable.
- It is considered the nonparametric alternative to ANOVA
- Example:

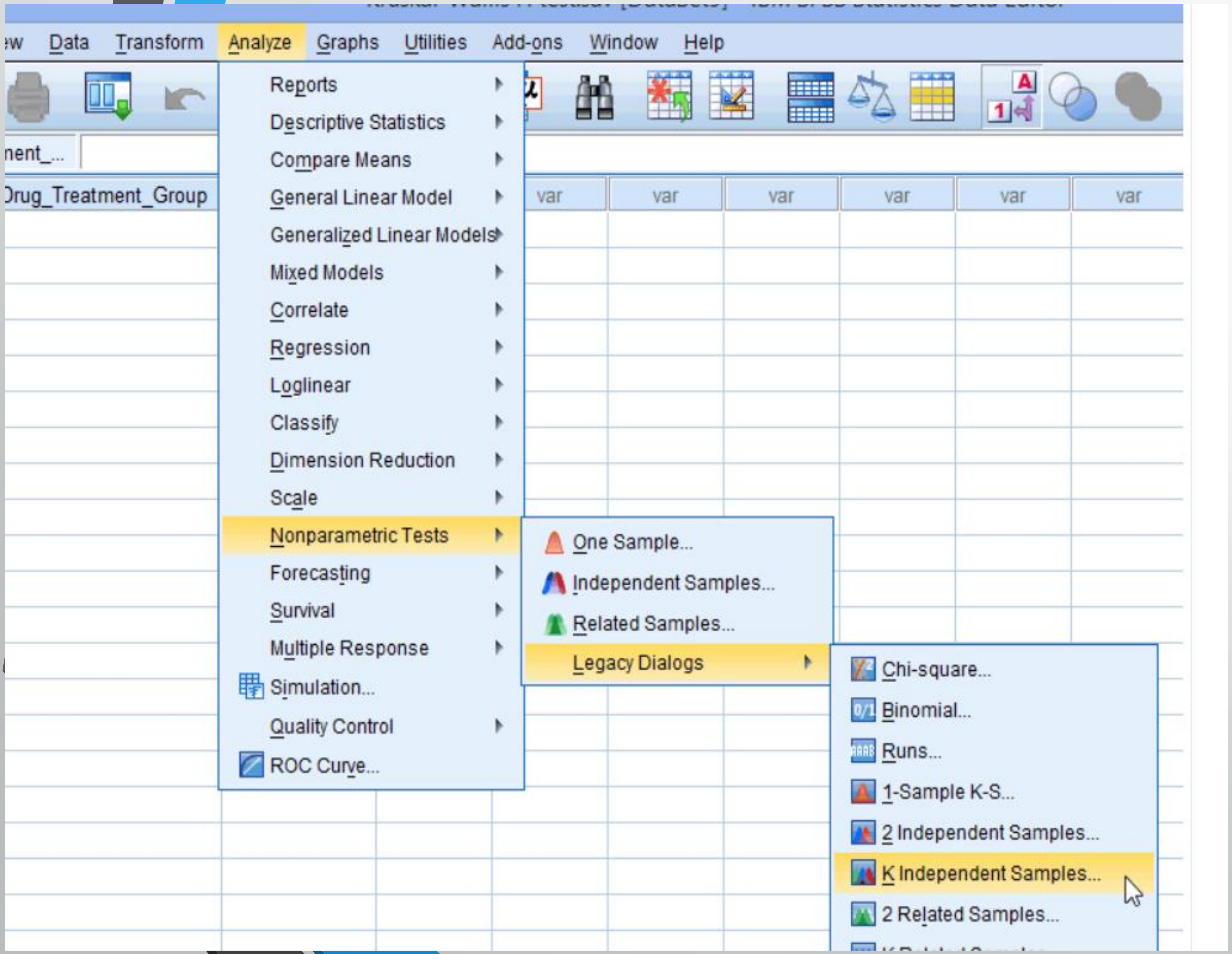
✓ dependent variable: "exam performance‟ on a continuous scale from 0-100

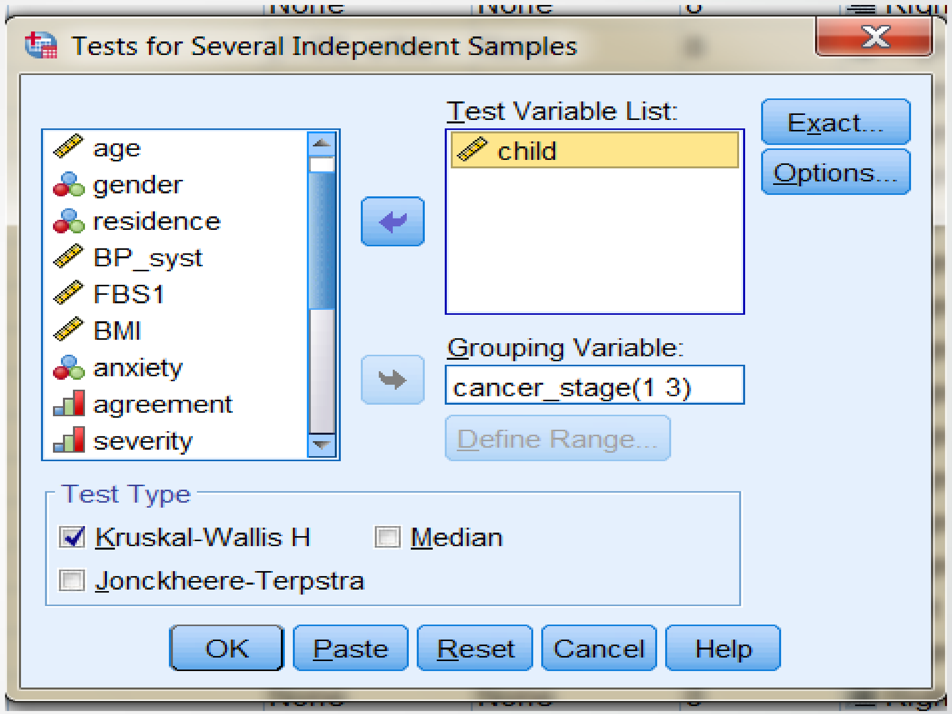
✓ independent variable: "low", "medium" and "high" test anxiety levels

Assumptions

- dependent variable measured at the ordinal or continuous level
- independent variable consist of 3 or more categorical, independent groups
- independence of observations

Running the test




Tests for Several Independent Samples

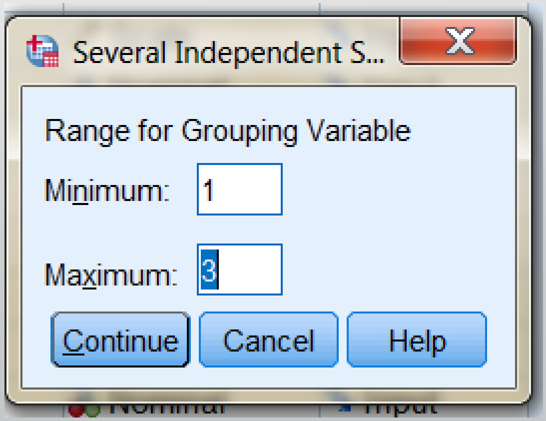
Test Variable List: child

Grouping Variable: cancer_stage(1 3)

Test Type

☑ Kruskal-Wallis H ☐ Median

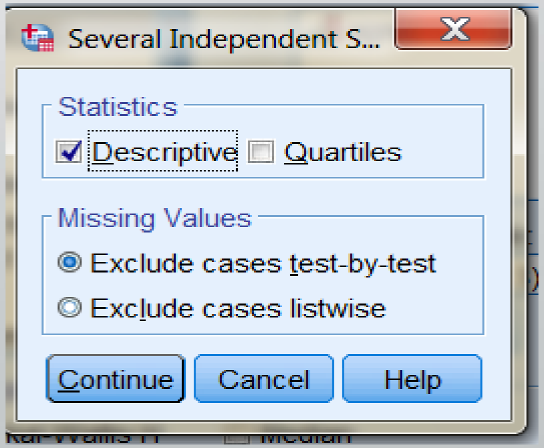
☐ Jonckheere-Terpstra


Several Independent S...

Range for Grouping Variable

Minimum: 1

Maximum: 3


Several Independent S...

Statistics

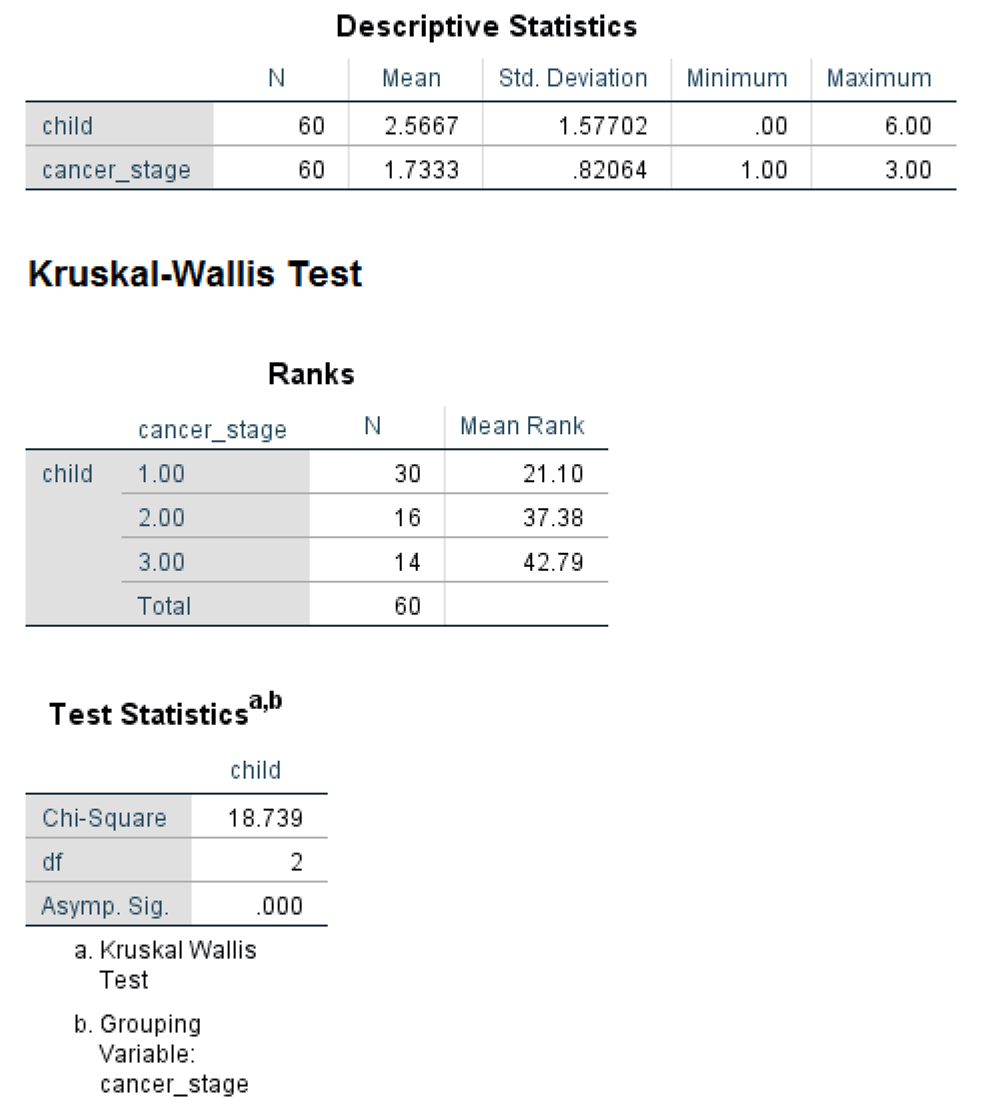
☑ Descriptive ☐ Quartiles

Missing Values

● Exclude cases test-by-test

○ Exclude cases listwise

Output



Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
child	60	2.5667	1.57702	.00	6.00
cancer_stage	60	1.7333	.82064	1.00	3.00

Kruskal-Wallis Test

Ranks

	cancer_stage	N	Mean Rank
child	1.00	30	21.10
	2.00	16	37.38
	3.00	14	42.79
	Total	60	

Test Statistics[a,b]

	child
Chi-Square	18.739
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: cancer_stage

References

- Introduction to Biostatistics and Research Methods, fifth edition, P.S.S Sunder Rao, Dr. PH (2012)
- URL: https://libguides.library.kent.edu/SPSS
- https://statistics.laerd.com/