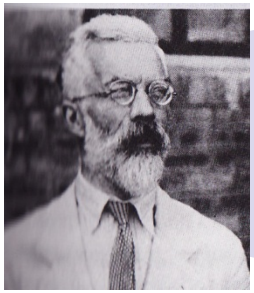
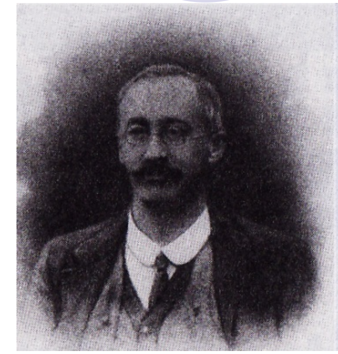


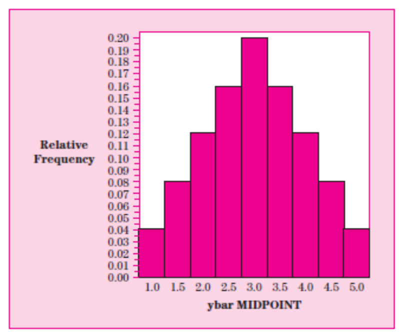
Sir Ronald Fisher

William S. Gosset

Student's t



- **T-tests**
- **One Way Anova**



T-tests and One-way ANOVA using the SPSS

Nikos Comoutos, PhD

- The Student's t-statistic was introduced in 1908 by William Sealy Gosset, using his pen name "Student", when he worked for the Guinness brewery in Dublin, Ireland. He devised the t-test as a way to cheaply monitor the quality of stout.

- The test was published in Biometrika in 1908.

- **Independent t-test**

Perform hypothesis tests concerning the difference in means of two independent populations – samples

- **Dependent t-test**

Perform hypothesis tests concerning the different in means of two dependent (related) Populations - samples

Independent t-test

Independent variable

One independent variable = gender – two levels – males + females

Only one dependent variable

Strength

Assumptions

- The data must be parametric, that is, they should be measured on an interval or ratio scale (if not use a non-paramatetric equivalent test)
- The samples should be randomly selected from the population (the results of the *t* test can be generalised from the sample to the population)
- The two samples should come from populations which have approximately the same variance (i.e., homogeneity of variance assumption). *Use* **Levene test to test this assumption**. *(not significant!!!) (variance:* The average of the **squared** differences from the Mean) – (Standard Deviation is just the square root of Variance)
- The scores of the dependent variable should come from a population which is normally distributed (i.e., normality assumption). *Use* **Histogram with normal curve in the Descriptive Statistics/Frequencies or with *Frequencies* option** *check skewness and kurtosis values (if they are above 1.96 -standard errors, the data are probably not normally distributed)*
- ***t* test is fairly robust to moderate violations of the homogeneity of variance and normality assumptions. Equal variances is not actually required. SPSS will handle it.**

Independent-samples T test

- When two groups are expected to differ in a particular variable

Example one: Elite weight lifting athletes are predicted to have higher scores in strength compared to non-elite ones. (one-tailed test).

Example two: The effect of gender of a sample that suffer from depression in walking using pedometers (two-tailed test).

Hypothesis

- **Null hypothesis** :

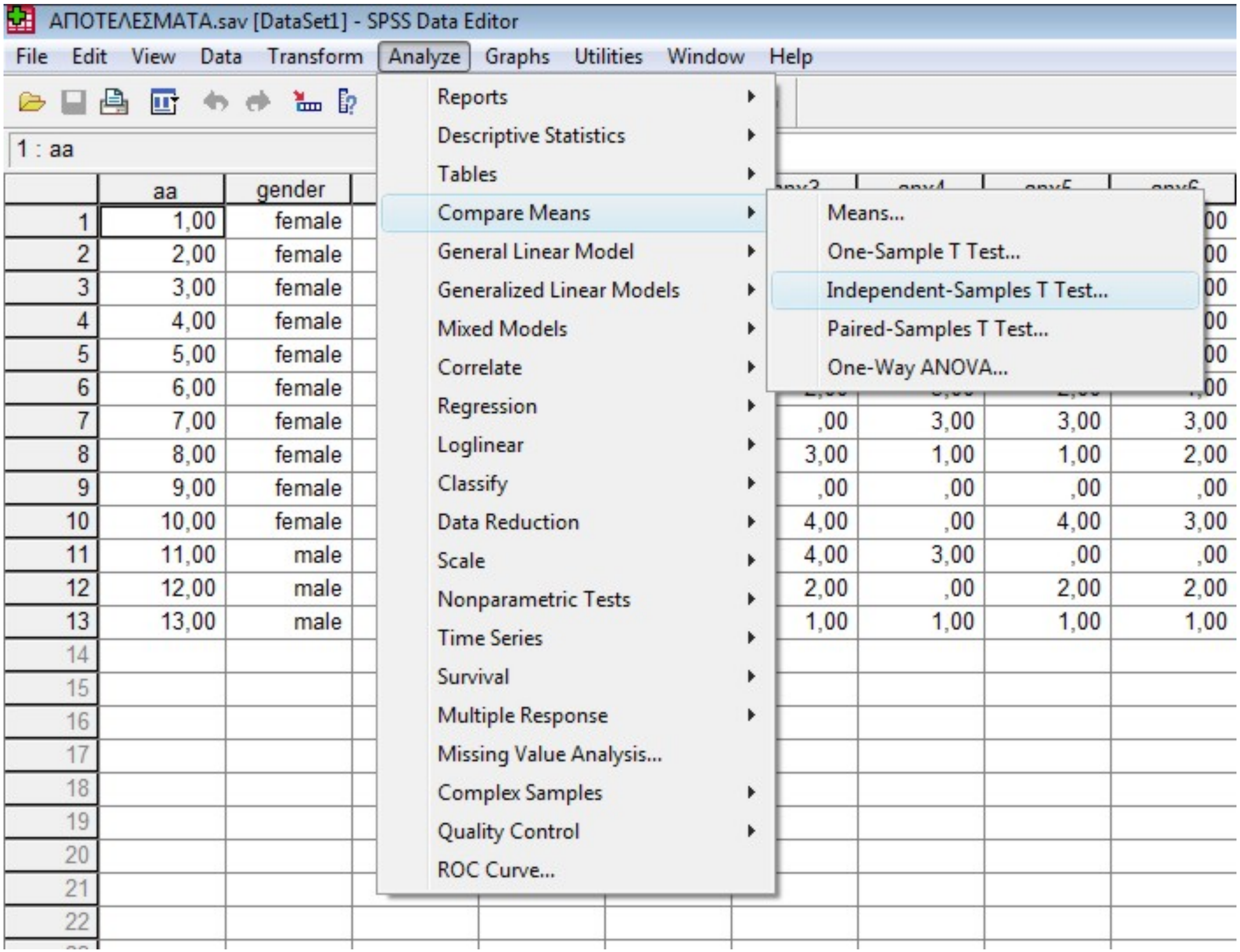
The two means are equal

Ho: $\mu_1 = \mu_2$

- **Alternative hypothesis**

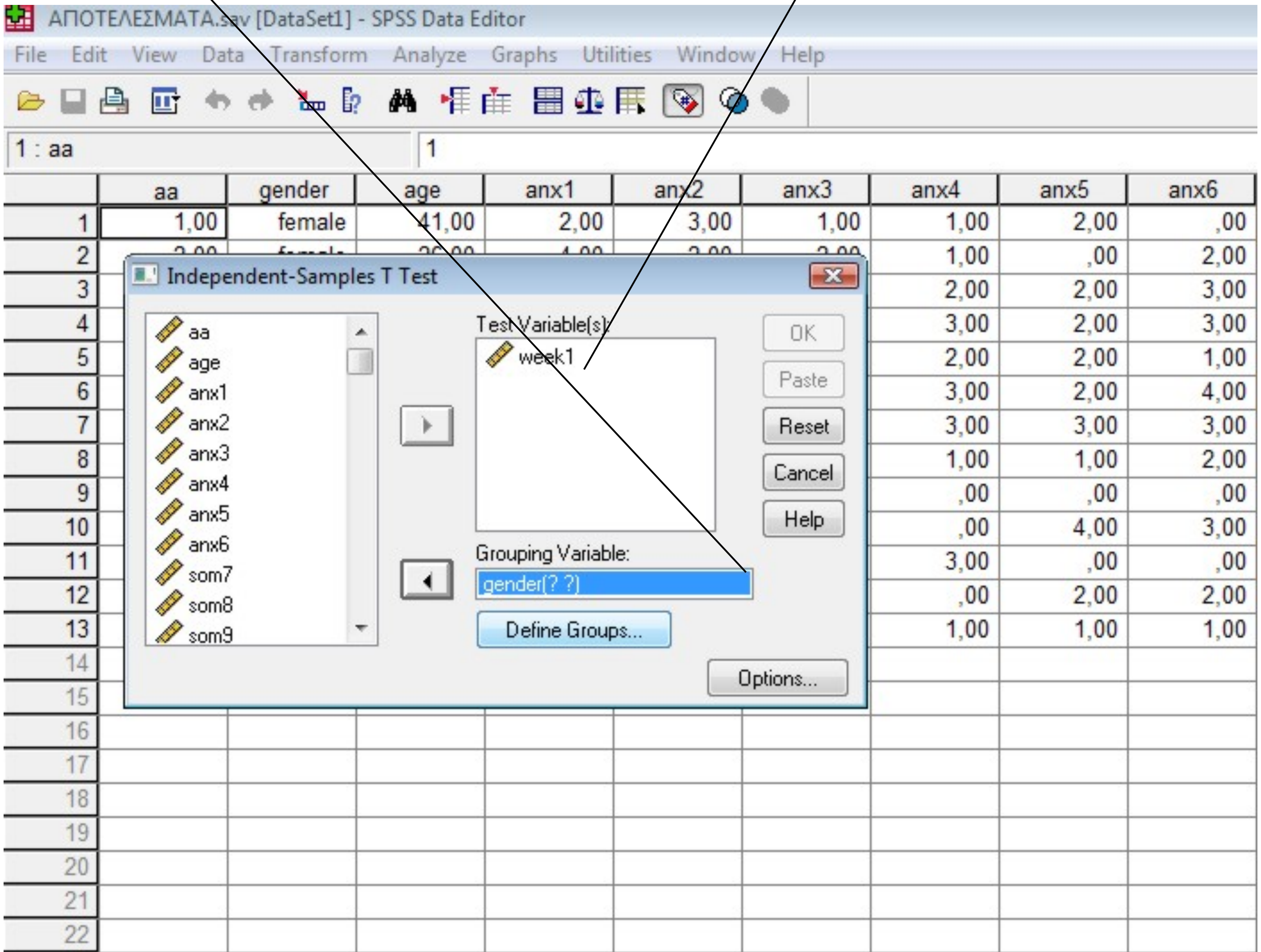
The two means are not equal

H_1: $\mu1 \neq \mu2$

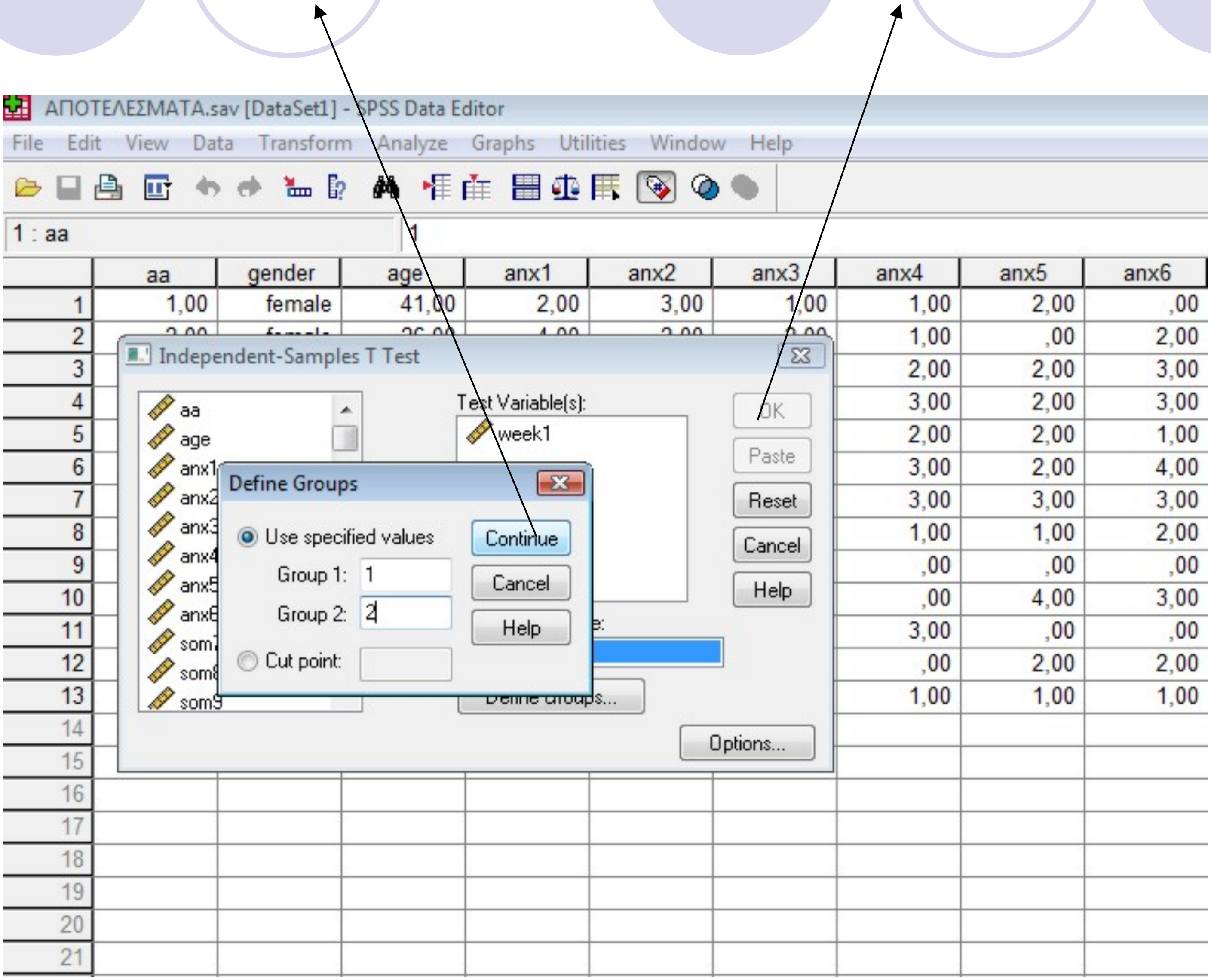

ΑΠΟΤΕΛΕΣΜΑΤΑ.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : aa
aa
gender
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Quality Control
ROC Curve...
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...

Independent, π.χ. gender – two levels, males - females

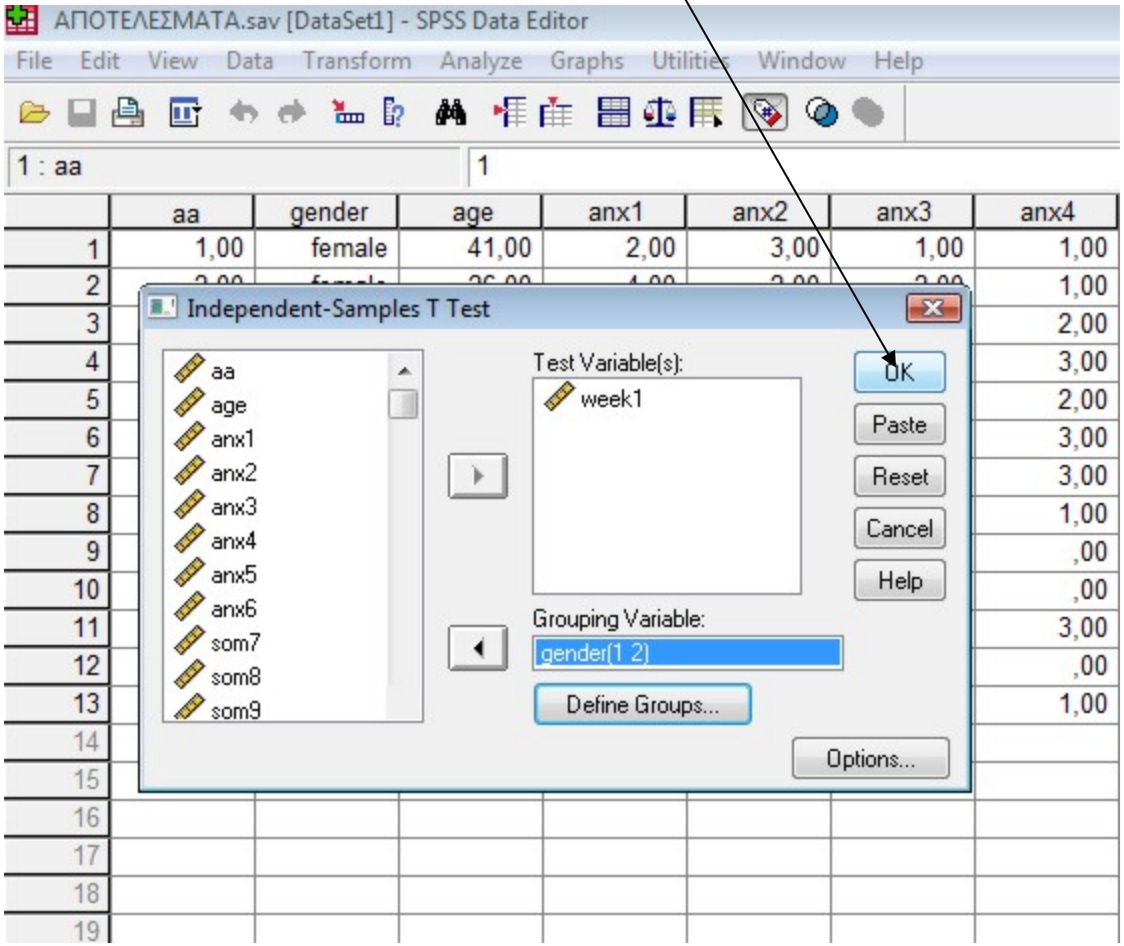
Dependent, π.χ. Number of Steps of week 1



Click continue and then Oκ



Run the analysis clicking OK



We do not want Levene's Test to be significant. If it is significant we do not have homogeneity of variance between the samples and we report the results from the Equal variances not assumed

Men do more steps compare to women

T-Test

[DataSet1] C:\Users\Nikos\Desktop\DOCS\students\andronikh\ΑΠΟΤΕΛΕΣΜΑΤΑ.sav

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
week1	male	3	76214,00	22871,97943	13205,14
	female	10	39998,00	24162,10552	7640,729

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
week1	Equal variances assumed	,185	,675	2,299	11	,042	36216,000	15754,439	1540,714	70891,29
	Equal variances not assumed			2,374	3,477	,086	36216,000	15256,361	-8778,90	81210,90

Results

$t(11) = 2.30, p < .05$

Two-tailed test because we do not know which of the two genders do more steps than the other.

There is a statistical significance $p < .05$

T-Test

[DataSet1] C:\Users\Nikos\Desktop\DOCS\students\andronikh\ΑΠΟΤΕΛΕΣΜΑΤΑ.sav

Group Statistics

	gender	N	Mean	Std. Deviation	Std. Error Mean
week1	male	3	76214,00	22871,97943	13205,14
	female	10	39998,00	24162,10552	7640,729

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
week1	Equal variances assumed	,185	,675	2,299	11	,042	36216,000	15754,439	1540,714	70891,29
	Equal variances not assumed			2,374	3,477	,086	36216,000	15256,361	-8778,90	81210,90

How we report the test.

In order to examine the hypothesis we used the independent t-test. **The results showed a statistically significant difference in the number of steps t(11) = 2.30, p< .05. Men (*M* = 76.21, *SD* = 22.87) were doing more steps than women (*M*= 40, *S.D* = 24.16).**

Excerpts 11.9–11.10 Comparison of Two Sample Means Using a *t*-Test

The male respondents ($M = 33.88$, $SD = 9.29$) were older than the female respondents ($M = 30.95$, $SD = 8.41$), a statistically significant difference, $t(500) = 3.60$, $p < .01$.

Source: D. Hardina. (1999). Employment and the use of welfare among male and female heads of AFDC households. *Affilia, 14*(2), p. 217.

Paired-samples T Test

Use a dependent samples t-test if you measure

1) the same participants on a dependent variable at two different times (pre – post design)
2) have two separate groups of participants that have been matched based on some characteristic.

Example

- Examine the difference of the number of steps between week1 and week2

Hypothesis

- **Null hypothesis** :

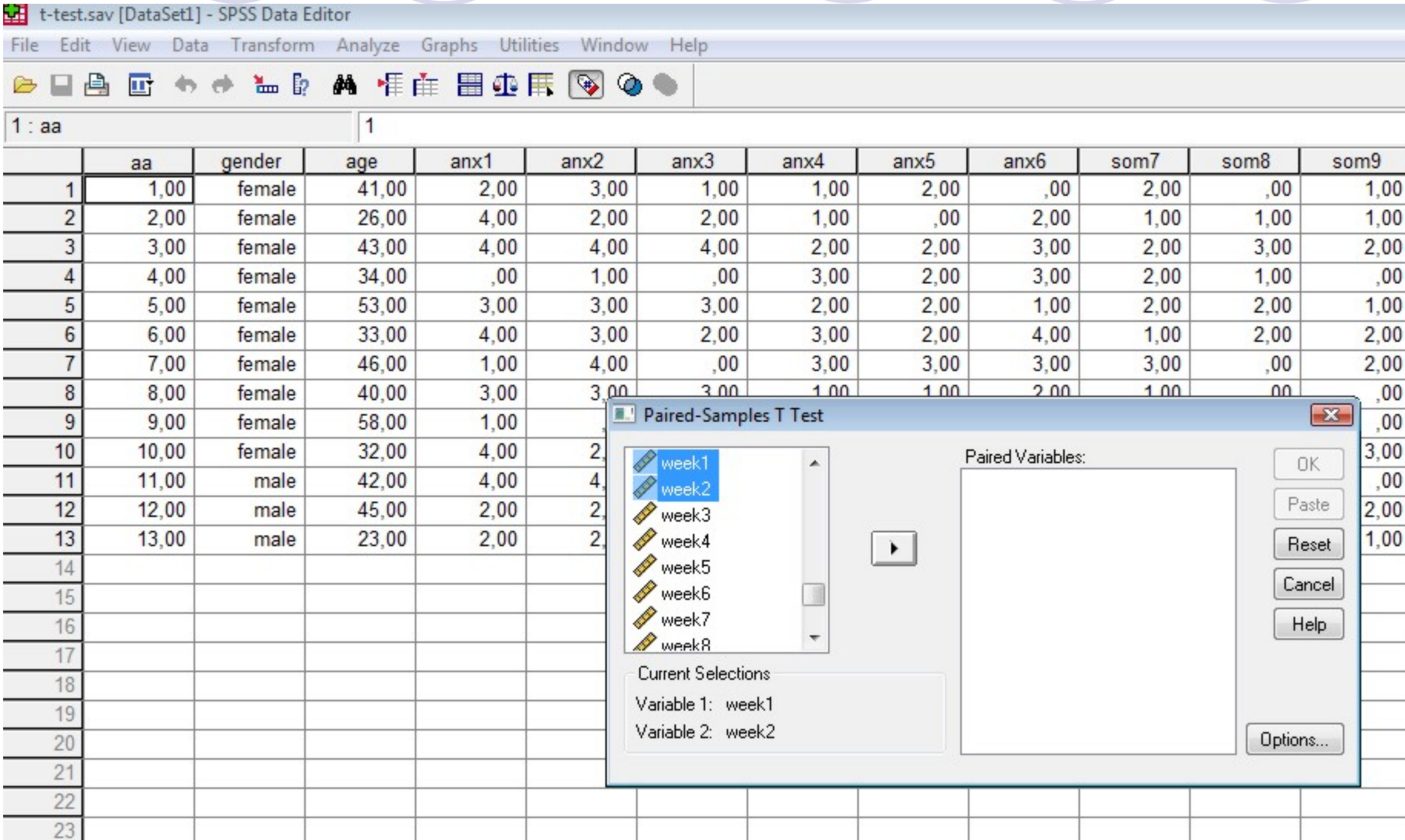
The two means are equal

Ho: $\mu_1 = \mu_2$

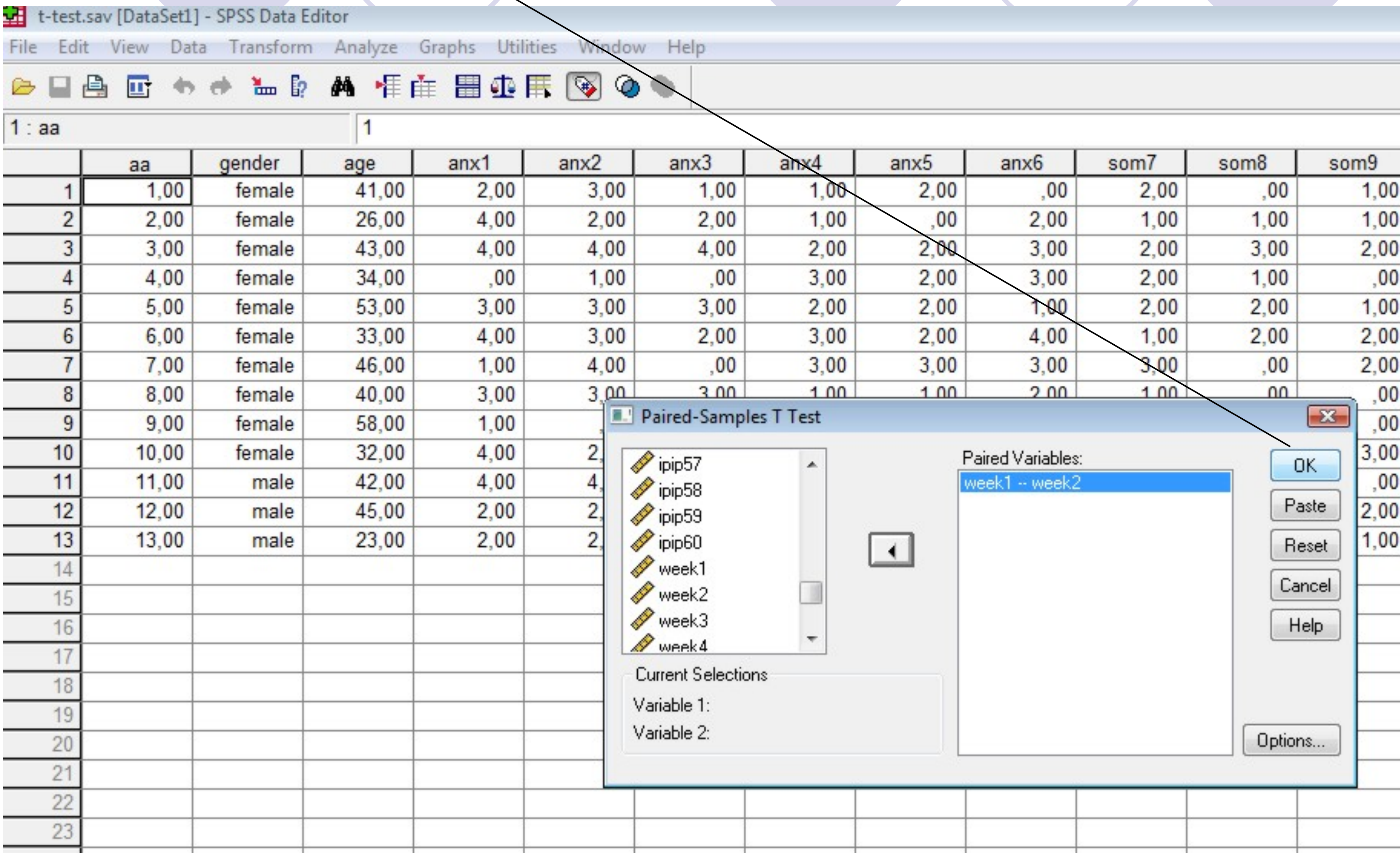
- **Alternative hypothesis**

The two means are not equal

H_1: $\mu1 \neq \mu2$. There is a significant difference between the steps of week 1 and week 2.


t-test.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : aa
aa gender age anx1 anx2 anx3 anx4 anx5 anx6 som7 som8 som9
Paired-Samples T Test
week1
week2
week3
week4
week5
week6
week7
Paired Variables:
OK
Paste
Reset
Cancel
Help
Current Selections
Variable 1: week1
Variable 2: week2
Options...

Click Ok


t-test.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : aa
aa gender age anx1 anx2 anx3 anx4 anx5 anx6 som7 som8 som9
Paired-Samples T Test
ipip57
ipip58
ipip59
ipip60
week1
week2
week3
week4
Paired Variables:
week1 -- week2
OK
Paste
Reset
Cancel
Help
Current Selections
Variable 1:
Variable 2:
Options...

$t(12) = -2.35, p < .05$

the sign of the *t* value is negative. It simply signifies that the mean of the second group is higher than the mean of the first group

T-Test

[DataSet1] C:\Users\Nikos\Desktop\t-test.sav

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	week1	48355,54	13	27879,63280	7732,419
	week2	70234,62	13	30286,85424	8400,062

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	week1 & week2	13	,339	,258

N-1

p < .05

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	week1 - week2	-21879,1	33506,30380	9292,977	-42126,7	-1631,42	-2,354	12	,036

How we report the test.

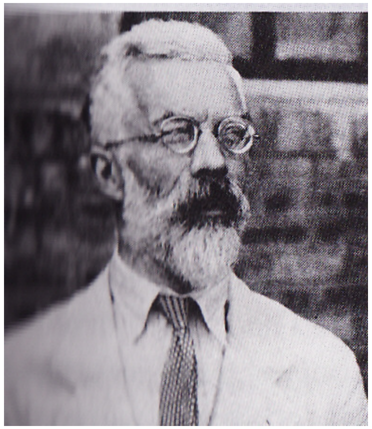
We used paired samples t-test in order to examine our hypothesis... The results showed statistically significant difference between week 1 and week 2, $t(12) = -2.35$, $p < .05$. Participants did more steps in week 2 (*M*= 70.23, *SD* = 30.29*)* than in week1 (*M* = 48.36, *SD* = 27.88*)*

Overall, these practice-teachers' answers were significantly higher at posttest (overall $M = 5.67$) than at pretest (overall $M = 4.91$), $t(74) = 5.76$, $p < .001$ (two-tailed), indicating an increased willingness to view the suicide threat seriously and take specific actions to prevent it.

Source: M. W. Davidson and L. M. Range. (1999). Are teachers of children and young adolescents responsive to suicide prevention training modules? Yes. *Death Studies, 23*, p .65.

ANOVA (Analysis of Variance)

This test is an extension of the Independent Samples *T Test*.



- Invented by R.A. Fisher in the 1920's

One Way Anova - ANOVA (Analysis of Variance)

- Two variables: 1 Categorical, 1 Quantitative
- Do the (means of) the quantitative variables depend on which group (given by categorical variable) the individual is in?
- If categorical variable has only 2 levels:
  - 2-sample t-test
- ANOVA allows for 3 or more groups- levels

Why ANOVA instead of many t-tests?

- Before ANOVA, this was the only option available to compare means between more than two groups.
- The problem with the multiple t-tests approach is that as the number of groups increases, the number of two sample t-tests also increases.
- As the number of tests increases the probability of making a Type I error also increases (Tabachnick & Fidell, 1996)

Advantages

1) ANOVA will identify if *any two* of the group means are significantly different with a *single* test.
2) If the significance level is set at 0.05, the probability of a Type I error for ANOVA = 0.05 regardless of the number of groups being compared.
3) If the ANOVA F-test is significant, further comparisons can be done to determine which groups have significantly different means.

Assumptions (Vincent, 1999)

- The data should be **parametric**, measured on an interval or ratio scale.
- **Independence** (no relationship between the scores of the dependent variable in the different groups) otherwise consider using the *Repeated Measures* ANOVA test.
- **Normality** test with Kolmogorov-Smirnov & Shapiro-Wilk tests.
- **Homogeneity of variance** (Levene's test). We do not want these tests to be significant.

Violation of Assumptions

- According to Vincent (1999), ANOVA is relatively robust to violations of the assumptions of Normality and Homogeneity of variance.
- If you suspect serious violations of the ANOVA assumptions, consider using a non-parametric equivalent test.

Hypothesis

- The Null hypothesis for ANOVA is that the means for all groups are equal:

$$H_o : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$

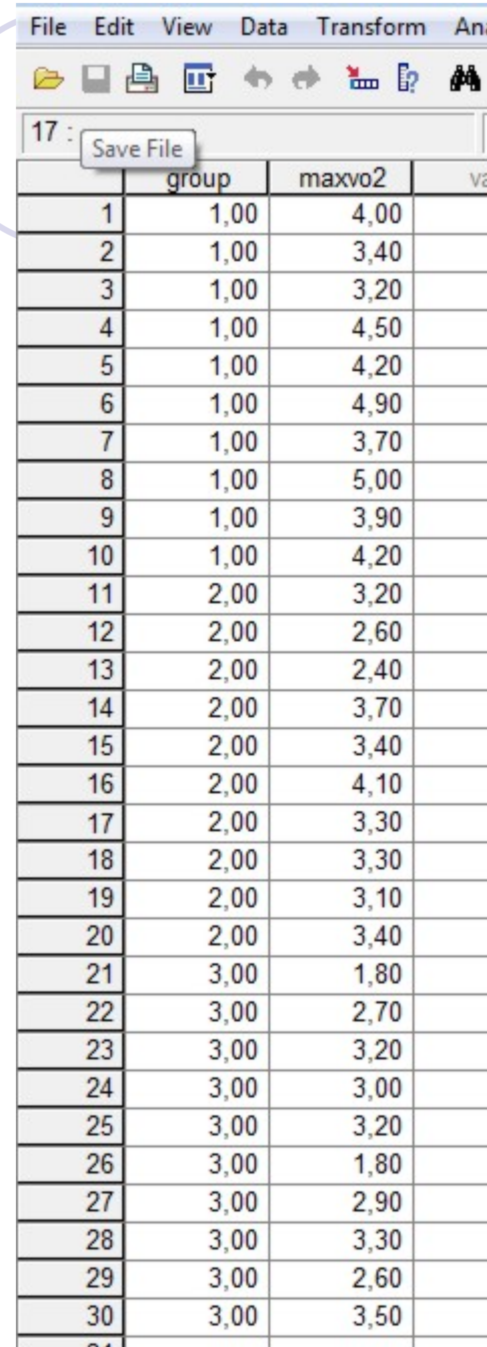
  - Null Hypothesis: $\mu_{control} = \mu_1 = \mu_2$

- The Alternative hypothesis for ANOVA is that *at least two* of the means are not equal $\mu 1 > \mu 2$, or $\mu 1 < \mu 2$
- The test statistic for ANOVA is the ANOVA

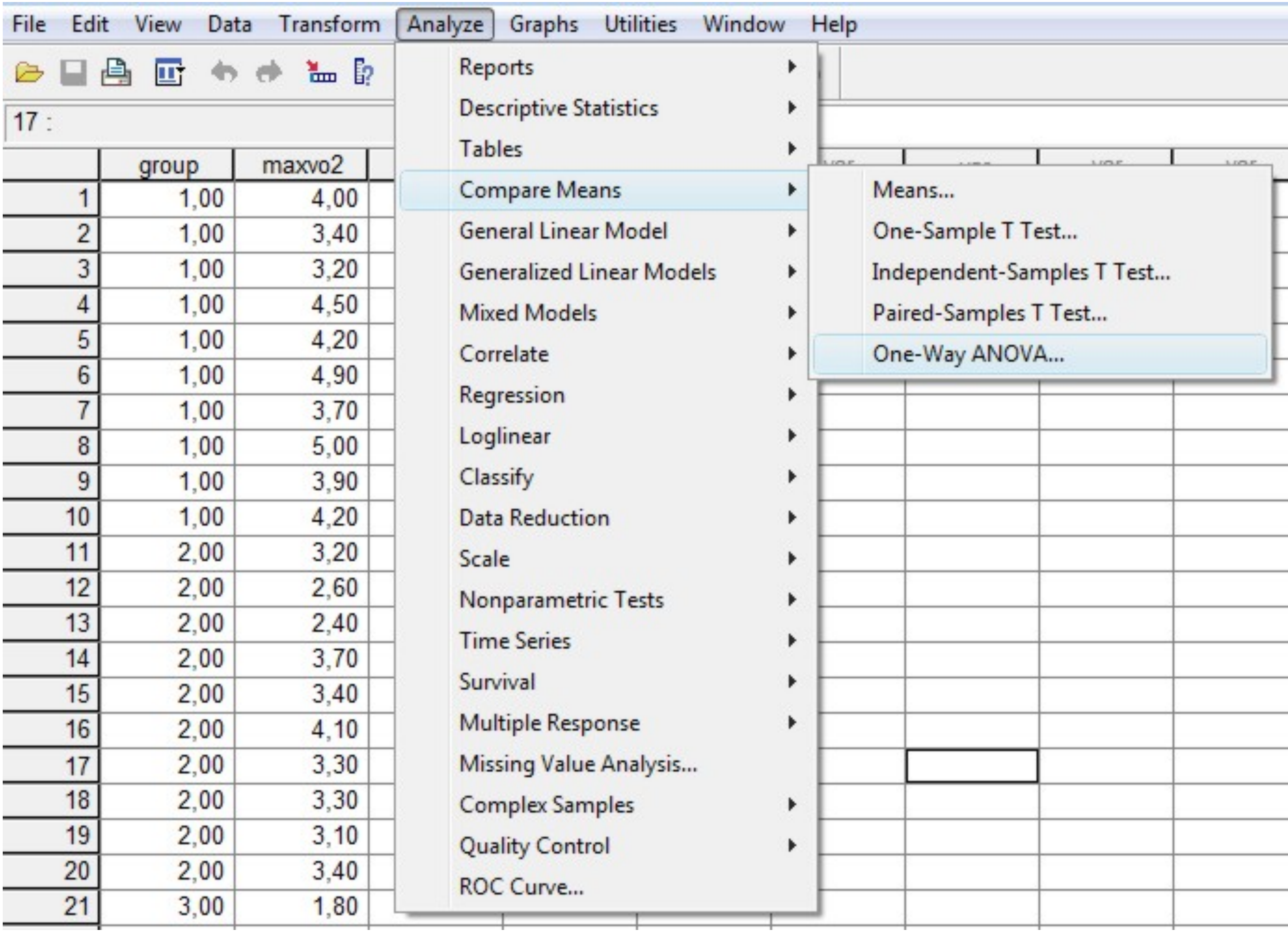
F-statistic.

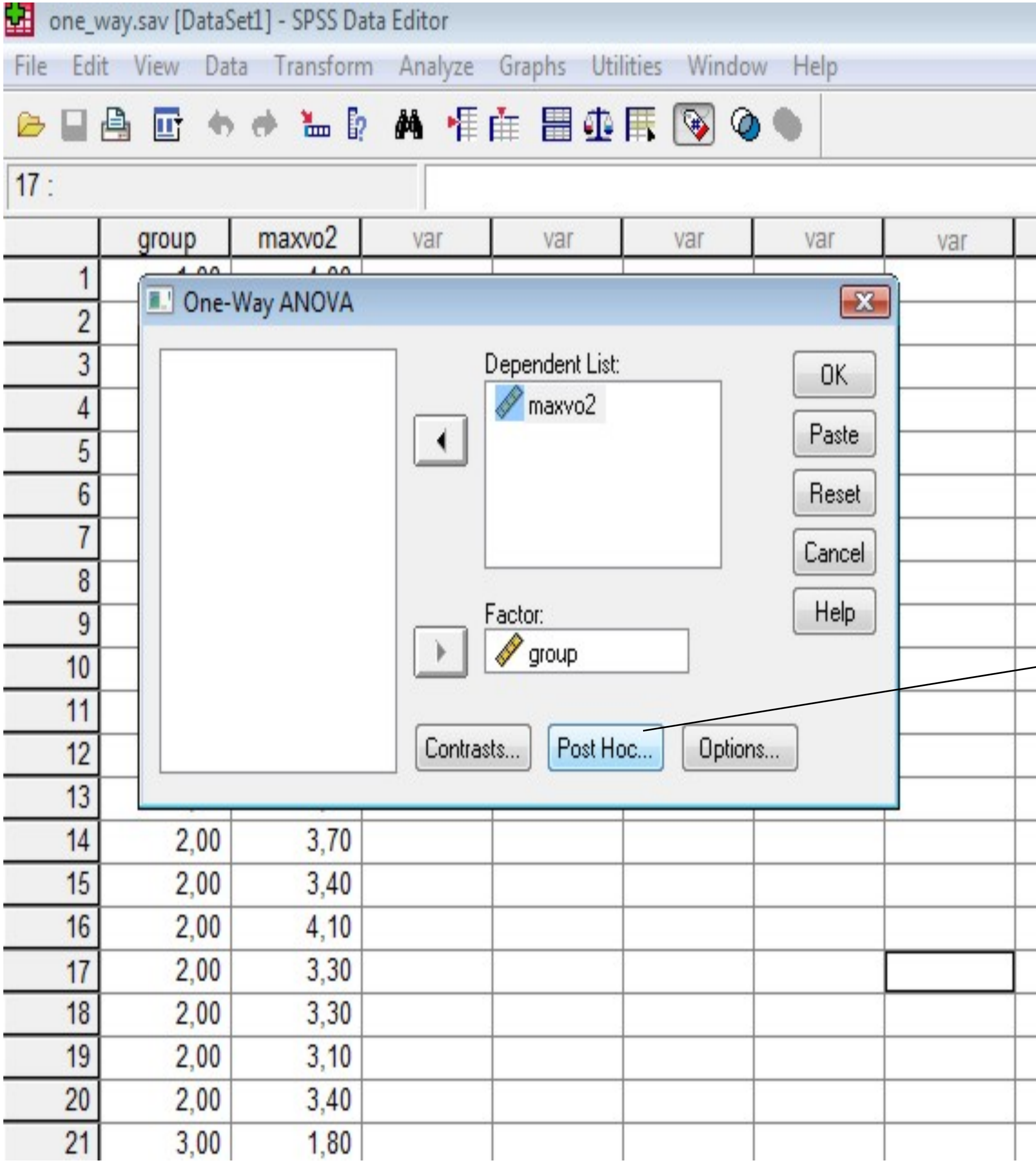
- The ANOVA F-test is a comparison of the average variability between groups to the average variability within groups.
  - The variability within each group is a measure of the spread of the data within each of the groups.
  - The variability between groups is a measure of the spread of the group means around the overall mean for all groups combined.
  - $$F = \frac{\text{average variability between groups}}{\text{average variability within groups}}$$

Example

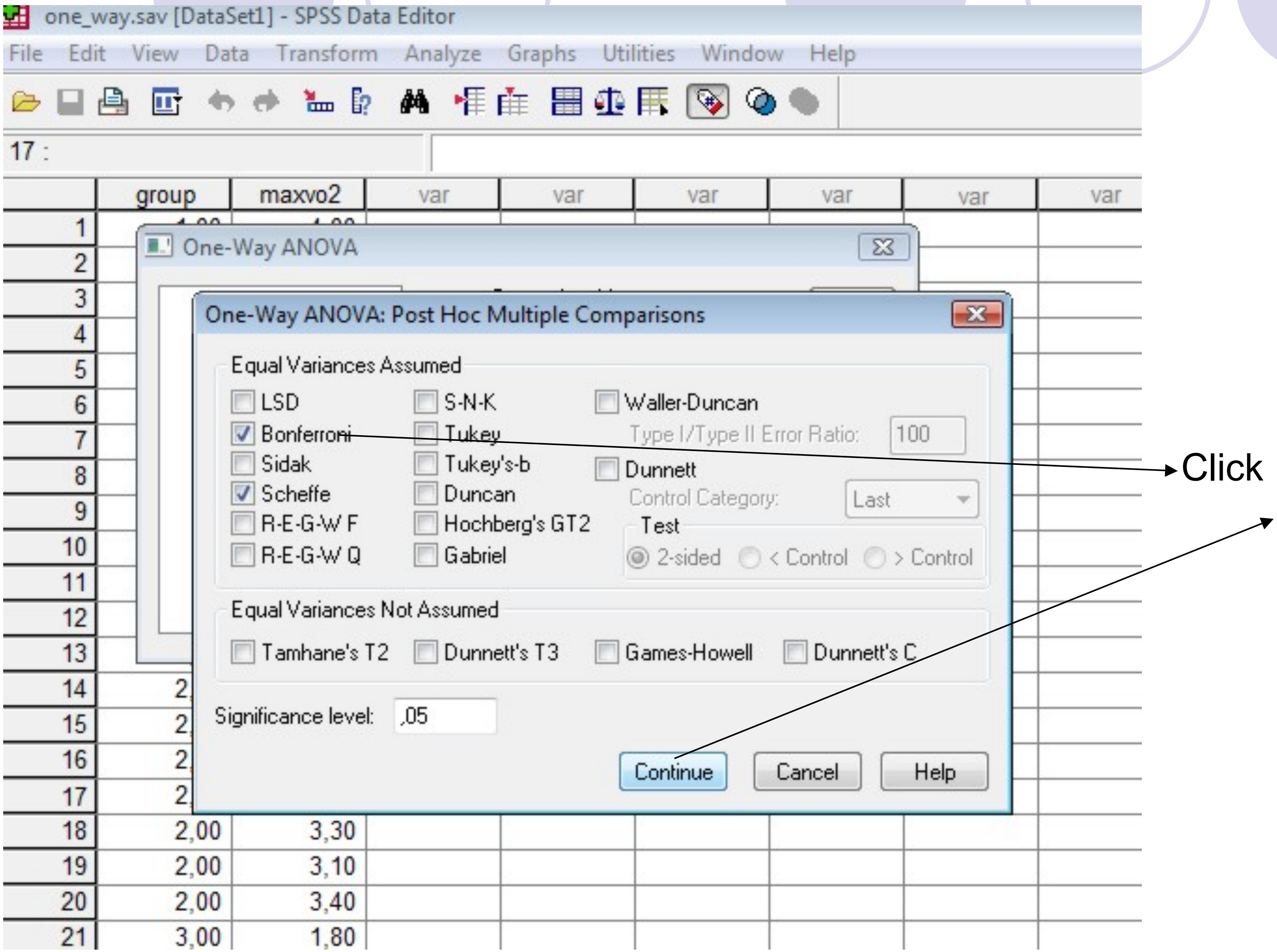


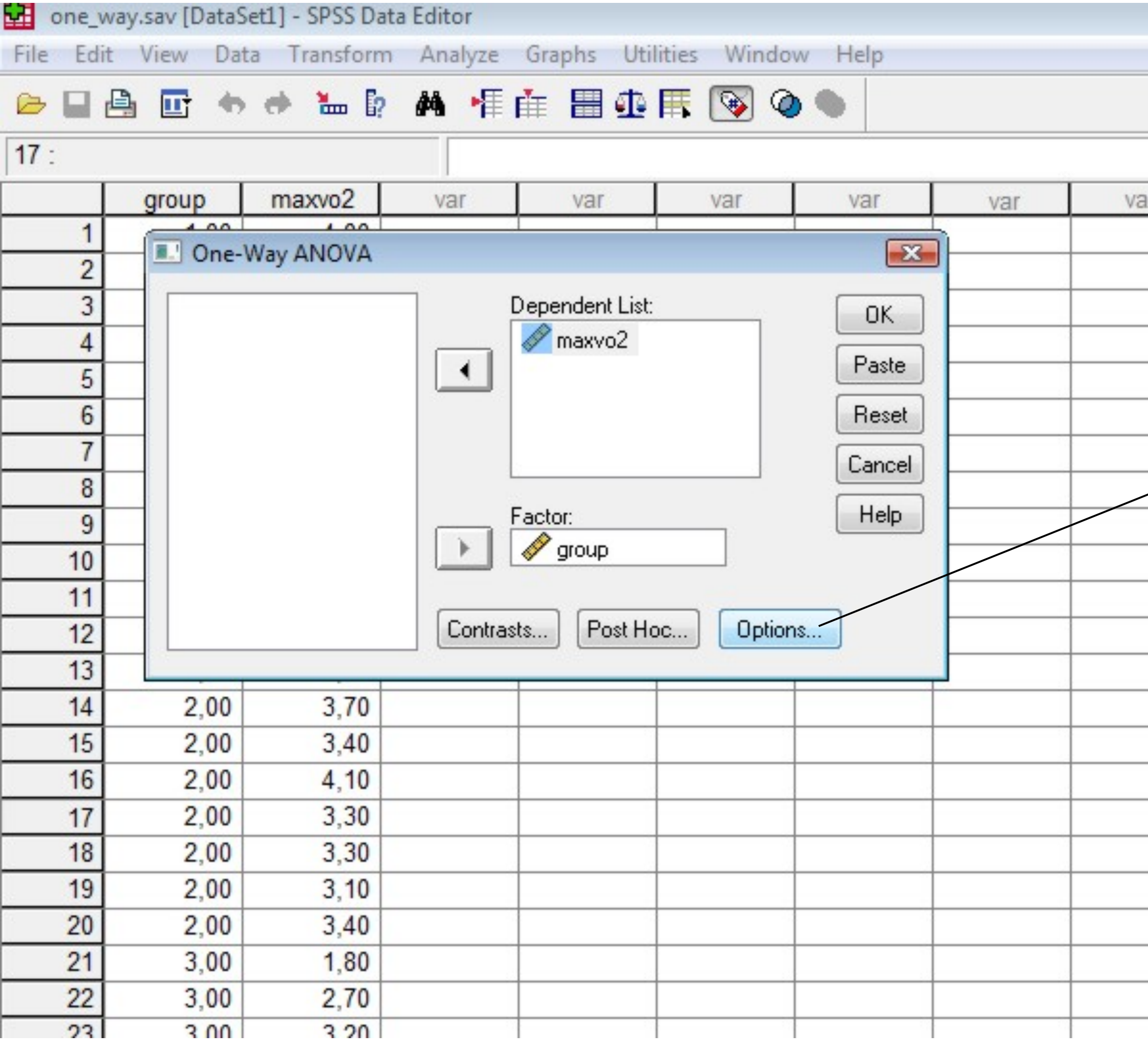
	group	maxvo2
1	1,00	4,00
2	1,00	3,40
3	1,00	3,20
4	1,00	4,50
5	1,00	4,20
6	1,00	4,90
7	1,00	3,70
8	1,00	5,00
9	1,00	3,90
10	1,00	4,20
11	2,00	3,20
12	2,00	2,60
13	2,00	2,40
14	2,00	3,70
15	2,00	3,40
16	2,00	4,10
17	2,00	3,30
18	2,00	3,30
19	2,00	3,10
20	2,00	3,40
21	3,00	1,80
22	3,00	2,70
23	3,00	3,20
24	3,00	3,00
25	3,00	3,20
26	3,00	1,80
27	3,00	2,90
28	3,00	3,30
29	3,00	2,60
30	3,00	3,50


File Edit View Data Transform Analyze Graphs Utilities Window Help
17 :
group
maxvo2
1 1,00 4,00
2 1,00 3,40
3 1,00 3,20
4 1,00 4,50
5 1,00 4,20
6 1,00 4,90
7 1,00 3,70
8 1,00 5,00
9 1,00 3,90
10 1,00 4,20
11 2,00 3,20
12 2,00 2,60
13 2,00 2,40
14 2,00 3,70
15 2,00 3,40
16 2,00 4,10
17 2,00 3,30
18 2,00 3,30
19 2,00 3,10
20 2,00 3,40
21 3,00 1,80
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Quality Control
ROC Curve...
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...



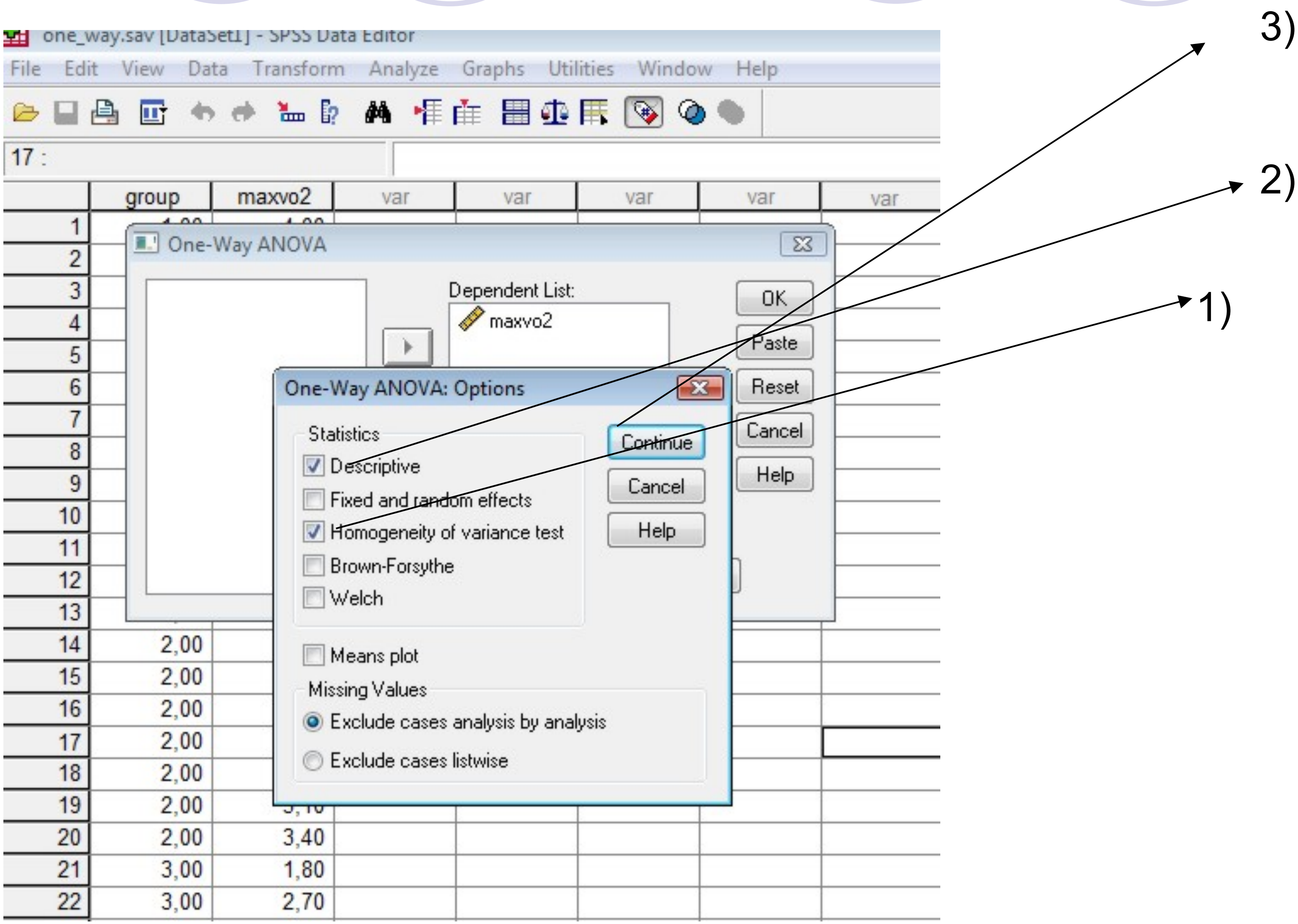
Click Post Hoc


one_way.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
One-Way ANOVA
One-Way ANOVA: Post Hoc Multiple Comparisons
Equal Variances Assumed
LSD
Bonferroni
Sidak
Scheffe
R-E-G-W F
R-E-G-W Q
S-N-K
Tukey
Tukey's-b
Duncan
Hochberg's GT2
Gabriel
Waller-Duncan
Type I/Type II Error Ratio: 100
Dunnett
Control Category: Last
Test
2-sided
< Control
> Control
Equal Variances Not Assumed
Tamhane's T2
Dunnett's T3
Games-Howell
Dunnett's C
Significance level: ,05
Continue
Cancel
Help
Click

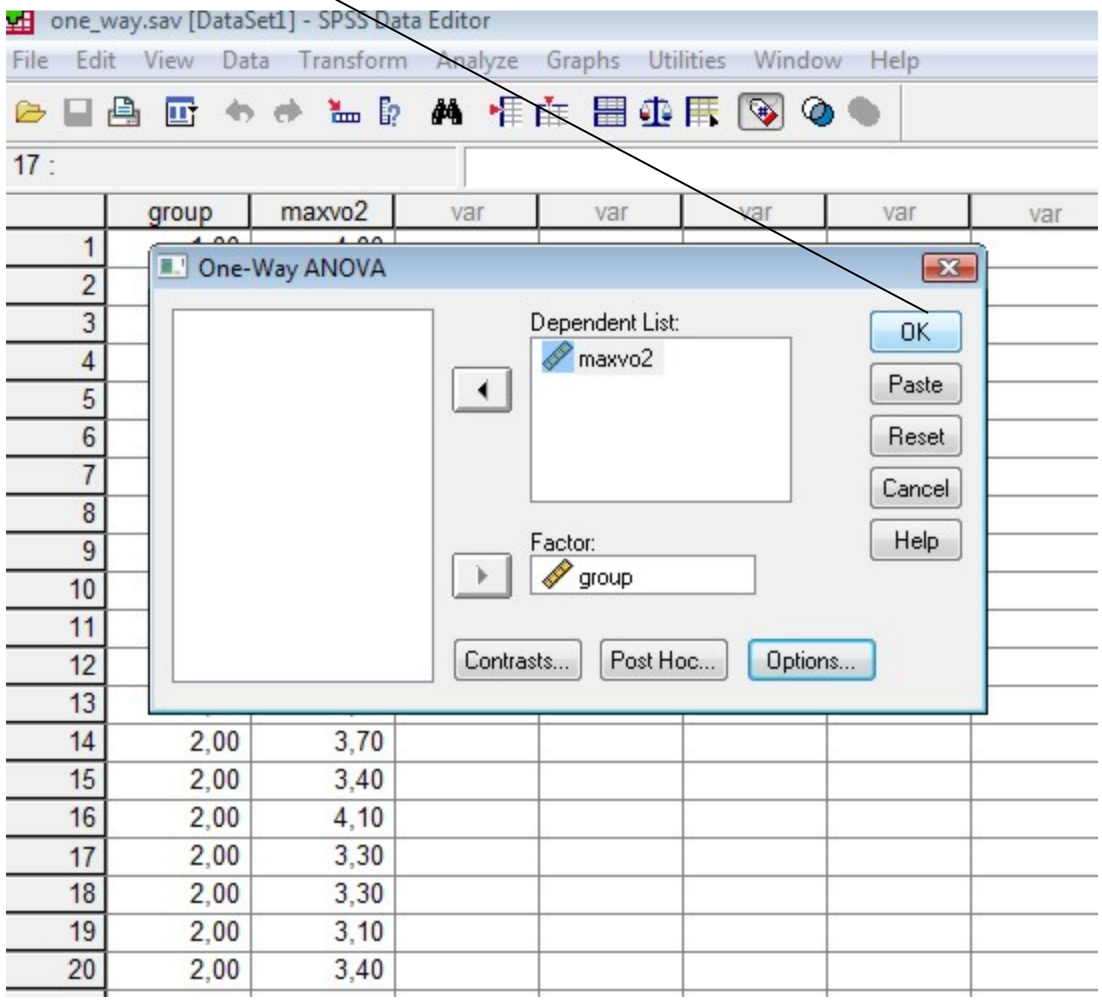


one_way.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
17 :
group
maxvo2
var
One-Way ANOVA
Dependent List:
maxvo2
Factor:
group
OK
Paste
Reset
Cancel
Help
Contrasts...
Post Hoc...
Options...

Click


one_way.sav [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
17 :
group maxvo2 var var var var var
One-Way ANOVA
Dependent List:
maxvo2
OK
Paste
Reset
Cancel
Help
One-Way ANOVA: Options
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
Continue
Cancel
Help
14 2,00
15 2,00
16 2,00
17 2,00
18 2,00
19 2,00
20 2,00 3,40
21 3,00 1,80
22 3,00 2,70
3)
2)
1)

Click Ok to run the analysis



One-way ANOVA

It is parametric test. We don't use repeated T tests for these popularities cause we have a big probability of **Type Error I=*Reject the null hypothesis when it actually true.***

Standard deviation of the sample means

Number of subjects, which were randomly assigned to the groups

M or $\bar{X} = \frac{\sum X}{N}$

Descriptive information concerning the VO_2MAX

Standard deviation and variance are based upon all of the scores in group and not just the high and low or the upper and lower quartile points

Dependent Variable

Minimum value in group 1

Maximum value in group 1

Groups

Descri[ptives]

MAXVO2

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	10	4.1000	.5907	.1868	3.6775	4.5225	3.20	5.00
2.00	10	3.[illegible]	.4882	.1544	2.9008	3.5992	2.40	4.10
3.00	10	2.8000	.5925	.1874	2.3761	3.2239	1.80	3.50
Total		3.3833	.7693	.1404	3.0961	3.6706	1.80	5.00

Measures the average amount by which all the vales deviate from the mean

S or $\sigma = \sqrt{\left[\frac{\sum (X-M)^2}{(N-1)}\right]}$

Nearly equal variances in the scores of the dependent variable. The variance is simply the SD squared

Number of the values in the data that are free to change

Looking for significance bigger than .05 if this is true we have homogeneity. we can test homogeneity even with Hartley's Test by hand

Test of Homo[gen]eity of Variances

MAXVO2

Levene Statistic	df1	df2	Sig.
.424	2	27	.659

numbers that can freely change and the mean remain constant.

When we reject only the lowest 2.5% or the highest 2.5% making a total of 5% is referred as a one-tailed or directional test.

Reject the Ho

When we reject extremes in both tails, we have a two-tailed or nondirectional test.

The inferential test comparing three means

Willing of taking 5% chance of making mistake. We look for significance <.05 or <.001

medical situation (drugs) tested simply means that the null hypothesis being

Add 1 three sample means are being compared

How different is the variance between two variables

Ratio = Between Variance / Within Variance

(4.358 ÷ .313) = 13.934

It is the sum of the squares of the deviation of each score of a mean

Reject Ho when an outcome falls in the lowest 5% of the distribution; we will reject Ho whenever an individual score falls in the shaded area. Whenever a score as low as has a probability of 0.05.

$SS_B = \Sigma(\bar{x}_{group} - \bar{x}_{grand})^2$

$SS_W = \Sigma(x - \bar{x}_{group})^2$

$SS_B + SS_W$

$x - \bar{x}$ lost one df

Test statistic

ANOVA

MAXVO2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.717	2	4.358	13.934	.000
Within Groups	8.445	27	.313		
Total	17.162	29			

Subtract 1 from each sample size
[(10-1)+(10-1)+(10-1)] = 27

SS 8.717/2

SS 8.445/27

The sum of df for the between and within is equal to 29. One less than the total number used in the analysis

Setting the a level at .05 means that we accept the 5% probability of making a type I error = reject the null hypothesis when is true. a (alpha)

p = α

The ability to specify the probability of error follows directly from the logic of hypothesis testing.

type II error = accept the null hypothesis when is false

or fail to reject Ho when it's false p = β (beta)

statistical power = reject the null hypothesis when is false

5% chance of making an error is too great risk to take & suggest that we make our criterion much more stringent by rejecting only the lowest 1% the distribution. The more stringent you make your criterion the more like to make type II error.

Decision	Ho True	Ho False
Reject Ho	Type I error p = α	Correct decision p = 1 - β = power
Don't Reject Ho	Correct decision p = 1 - α	Type II error p = β

Multiple Comparisons

Dependent Variable: MAXVO2

	(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.
Tukey HSD	1.00	2.00	.8500*	.2501	.006
		3.00	1.3000*	.2501	.000
	2.00	1.00	-.8500*	.2501	.006
		3.00	.4500	.2501	.189
	3.00	1.00	-1.3000*	.2501	.000
		2.00	-.4500	.2501	.189
Bonferroni	1.00	2.00	.8500*	.2501	.006
		3.00	1.3000*	.2501	.000
	2.00	1.00	-.8500*	.2501	.006
		3.00	.4500	.2501	.250
	3.00	1.00	-1.3000*	.2501	.000
		2.00	-.4500	.2501	.250

$\rightarrow (\bar{x}_1 - \bar{x}_2)$

where is exactly the statistically difference

Regarding VO_2MAX, post hoc tests revealed a significant difference between the control group and the experimental groups according to the conservative Bonferroni and Tukey's tests.

Notes

If we use **Hartlys test** by hand

We have to calculate the ratio between the largest variance and the smallest variance of our groups, the result is compared with the critical value related with the degrees of freedom.The Ho is that the variance are equal.

$Fmax=.5925^2/.4882^2=1.57$

F(9,9) has a critical value of 3.15 so there is no evidence to contradict the variance

Null Hypothese:Ho=$m_1=m_2=m_3$

If this hypothesis is correct we expect to find that the between groups variance and the within groups variance are very small and that their ratio (F) is close to 1.

When populations are not normal or have unequal variances

I)identify and eliminate outliers

II)transform simple data

III)use other test that hasn't got rigorous assumptions.

Post Hoc Tests –follow up tests are performed to examine where exactly differences are among the groups. What caused the null hypotheses to be rejected. The above tests are written with the order of conservatism, and e.g. differences have to be larger to be recognized by the first compared with the last one:

- Bonferroni test
- Sceffe's test
- Tukey' test
- Tukey's (b) test
- Newman-Kenls test
- Duncan's New Multiple Range

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Honestly significant different

MAXVO2

	GROUP	N	Subset for alpha = .05	
			1	2
Tukey HSD[a]	3.00	10	2.8000	
	2.00	10	3.2500	
	1.00	10		4.1000
	Sig.		.189	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 10.000.

There is not significant difference between 2 + 3.

Mean Value

Because it is in the second it's different HSD.

The 1 group is significant different from ② ③

Conclusion?

We can conclude that the training method 1 was the most effective for enhancing the students VO2MAX. Also the training method 2 did not differ in effectiveness even from the control group that the did not have any treatment

How to report the test

- We used One-Way Anova in order to examine the differences in VO2max between different training methods. The results showed significant differences in VO2max between the training methods $F(2, 27) = 13.93$, $p < .001$. The post hoc (Bonferroni) test showed that the training methods 2 and 3 did not have significant differences, whereas training method 1 had significant differences between 2 and 3.

Excerpts 12.4–12.6 The Null Hypothesis in a One-Way ANOVA

$H_0: \mu_{Seniors} = \mu_{Juniors} = \mu_{Sophmores}$

Source: F. E. Ede, B. Panigrahi, and S. E. Calcich. (1998). African American students' attitudes toward entrepreneurship education. *Journal of Education for Business, 73 (5)*, p. 294.

The null hypothesis for this study, $H_0: \mu_I = \mu_{II} = \mu_{III}$, stated that there were no differences in mean scores on the curricular need scale for each of the 30 curricular need statements among the three groups.

Source: K. W. Lambrecht. (1991). Curricular preparation needs. *Journal of Sports Management, 5*(1), p. 51.

A one-way analysis of variance was calculated for each socialization variable. The null hypothesis in each case was that the population means for each socialization variable were equal across the [five] levels of termination.

Source: D. E. Martin and R. A. Dodder. (1991). Socialization experiences and level of terminating participation in sports. *Journal of Sports Management, 14*(2), p. 121.

Excerpts 12.11–12.13 Presentation of a One-Way ANOVA's Results Without a Table

A series of ANOVAs was performed with group (SLI, AC, and YC) as the independent variable and the different measures of arithmetic performance as the dependent variable. . . . Performance [in one of these ANOVAs] was assessed by determining the mean percentage of correct responses for the 20 sets of numbers. These scores were subjected to a one-way ANOVA, which yielded a significant group effect, $F(2,29) = 44.32$, $p < .001$.

Source: B. B. Fazio. (1999). Arithmetic calculation, short-term memory, and language performance in children with specific language impairment: A 5-year follow-up. *Journal of Speech, Language, and Hearing Research*, 42(2), p. 420.

(*Continued*)

Excerpt 12.10 A Table of *M*s, *SD*s, and *N*s With the *F*-value and *p*

We found significant attitudinal differences for the four groups based on the measure of aversion to women who work ($p < .001$). In Table 2 a summary of the analysis of variance shows that Euro-American women were the least averse to women who work while Mexican-American men were the most averse.

Table 2

Analysis of Variance: Sex, Ethnicity, and Aversion to Women Who Work

	M	*SD*	*N*	*F*	*p*
Euro-American Male	2.24	0.51	80	25.34	< .001
Mexican-American Male	2.54	0.59	80		
Euro-American Female	1.82	0.53	80		
Mexican-American Female	2.07	0.51	80		

Source: S. Valentine and G. Mosley. (1998). Aversion to women who work and perceived discrimination among Euro-Americans and Mexican-Americans. *Perceptual and Motor Skills, 86,* p. 1031.

Time for SPSS!!!