REVIEW OF SOME PARAMETRIC TESTES

OBJECTIVE

The objective of this lecture is to review some of the parametric tests you study in the previous courses such as:

- T-test (1-sample, 2-sample and paired)
- ANOVA (one and two way)
- Repeated Measures

T-TEST ANALYSIS

First, we review the Measurement Scales. We can measure and convey variables in several ways.

Nominal data:
also called categorical data, are represented by counting the number of times a particular event or condition occurs. For example, you might categorize the political alignment of a group of voters.

Dichotomous variable is a special classification of nominal data; it is simply a measure of two conditions. A dichotomous variable is either discrete or continuous. A *discrete dichotomous* variable has no particular order and might include such examples as gender (male vs. female) or a coin toss (heads vs. tails). A *continuous dichotomous* variable has some type of order to the two conditions and might include measurements such as pass/fail or young/old.

Ordinal Scale

Describe values that occur in some order of rank. However, distance between any two ordinal values holds no particular meaning. For example, imagine lining up a group of people according to height. It would be very unlikely that the individual heights would increase evenly.

Another example of an ordinal scale is a Likert-type scale. This scale asks the respondent to make a judgment using a scale of three, five, or seven items. The range of such a scale might use a 1 to represent strongly disagree while a 5 might represent strongly agree. This type of scale can be considered an ordinal measurement since any two respondents will vary in their interpretation of scale values.

Interval scale

A measure in which the relative distances between any two sequential values are the same. To borrow an example from the physical sciences, we consider the Celsius scale for measuring temperature. An increase from −8 to −7°C degrees is identical to an increase from 55 to 56°C.

Ratio scale

Slightly different from an interval scale. Unlike an interval scale, a ratio scale has an absolute zero value. In such a case, the zero value indicates a measurement limit or a complete absence of a particular condition. To borrow another example from the physical sciences, it would be appropriate to measure light intensity with a ratio scale. Total darkness is a complete absence of light and would receive a value of zero. On a general note, we have presented a classification of measurement scales similar to those used in many introductory statistics texts.

T-TEST

- T test was introduce in 1908 by W.S. Gosset.
- He was a brewer and agricultural statistician for the famous Guinness brewing company in Dublin. It insisted that its employees keep their work secret, so he published under the pseudonym ‘Student’ the distribution. This was one of the first results in modern small-sample statistics.

TYPES OF T-TEST

- **One sample:**

Compare the mean of a sample to a predefined value and used when you have a single sample and are comparing its mean to some value that is assumed to represent a population mean.

- **Dependent (related) samples:**

Compare the means of two conditions in which the same (or closely matched) participants participated

- **Independent (unrelated) samples:**

Used when you have two samples, each representing a different and independent group of subjects. The groups differ along one independent variable with two levels, and you compare the mean of those two groups (between-subjects design).

T-TEST ASSUMPTIONS

- The first assumption made regarding t-tests concerns the scale of measurement. The assumption for a t-test is that the scale of measurement applied to the data collected follows a continuous or ordinal scale, such as the scores for an IQ test.
- The second assumption made is that of a simple random sample, that the data is collected from a representative, randomly selected portion of the total population.
- The third assumption is the data, when plotted, results in a normal distribution, bell-shaped distribution curve.
- The fourth assumption is a reasonably large sample size is used. A larger sample size means the distribution of results should approach a normal bell-shaped curve.
- The final assumption is homogeneity of variance. Homogeneous, or equal, variance exists when the standard deviations of samples are approximately equal.

APPLICATIONS

- To compare the mean of a sample with population mean.
- To compare the mean of one sample with the mean of another independent sample.
- To compare between the values (readings) of one sample in 2 occasions.

ONE SAMPLE T-TEST

- The One-sample t test is used to compare a sample mean to a specific value (e.g., a population parameter; chance performance, etc.).

Assumptions:

- The data must be continuous.
- The data must follow the normal probability distribution.

Check by:

1- graph: (a)p-p or q-q plots (b) Box-plot

2-test: (a) Klomogrov and Smeranov test

(b) Shabero test

- The sample is a simple random sample from its population.

EXAMPLE:

- A teacher wants to see if his students are performing well in the math class. Ten students are chosen randomly and the teacher wants the class to be able to score different than 60 on the math competition. Can the teacher have 95 percent confidence that the mean score for the class would be different than 60?

Students	Scores
1	54
2	65
3	41
4	52
5	47
6	51
7	43
8	66
9	60
10	41

Mean	52

$$s = \sqrt{\frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}}$$

$$= \sqrt{\frac{782}{9}} = 9.321$$

Scores	$x-\bar{x}$	$(x-\bar{x})^2$
54	2	4
65	13	169
41	-11	121
52	0	0
47	-5	25
51	-1	1
43	-9	81
66	14	196
60	8	64
41	-11	121
520/10=52		782

- **The hypothesis:**
  - Null hypothesis: H_0 $\mu = 60$
  - Alternative hypothesis: H_a $\mu \neq 60$
- The α-level: $\alpha = 0.05$

- The assumptions:
- Dependent variables normally distributed.

T test

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$T = \frac{52 - 60}{\frac{9.321}{\sqrt{10}}}$$

The calculated t-value of **-2.714** while the critical t value **2.262** at 9 degrees of freedom

μ_0

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05
df							
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262
10	0.000	0.700	0.879	1.093	1.372	1.812	[illegible]

X bar= sample mean
μ_0 = population mean
s = standard deviation
n = sample size

- **<u>The results:</u>**

  The calculated t-value of 2.714 is larger in magnitude than the critical value of 2.262, therefore we can reject the null hypothesis

- **<u>The conclusion:</u>**

  The teacher now has evidence that the class mean on the test would be different than 60.

STEP BY STEP (SPSS):

one sample data.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	scores	Numeric	8	0		None	None	8	Right	Scale
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

Data View Variable View

one sample data.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	scores	var	var	var	var	var	var	var	var	var	var
1	54										
2	65										
3	41										
4	52										
5	47										
6	51										
7	43										
8	66										
9	60										
10	41										
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											

Data View Variable View

CHECK NORMALITY

one sample data.sav [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

12 :

	scores	var
1	54	
2	65	
3	41	
4	52	
5	47	
6	51	
7	43	
8	66	
9	60	
10	41	
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control
ROC Curve...

Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...

Data View Variable View

Explore...

one sample data.sav [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

12 :

	scores	var	var	var	var	var	var	var	var	var	var	var	var
1	54												
2	65												
3	41												
4	52												
5	47												
6	51												
7	43												
8	66												
9	60												
10	41												
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													

Explore

Statistics...

Plots...

Options...

Bootstrap...

Display

Both

Explore: Plots

Boxplots

Factor levels together

Dependents together

None

Descriptive

Stem-and-leaf

Histogram

Normality plots with tests

Spread vs Level with Levene Test

None

Power estimation

Transformed Power: Natural log

Untransformed

Continue Cancel Help

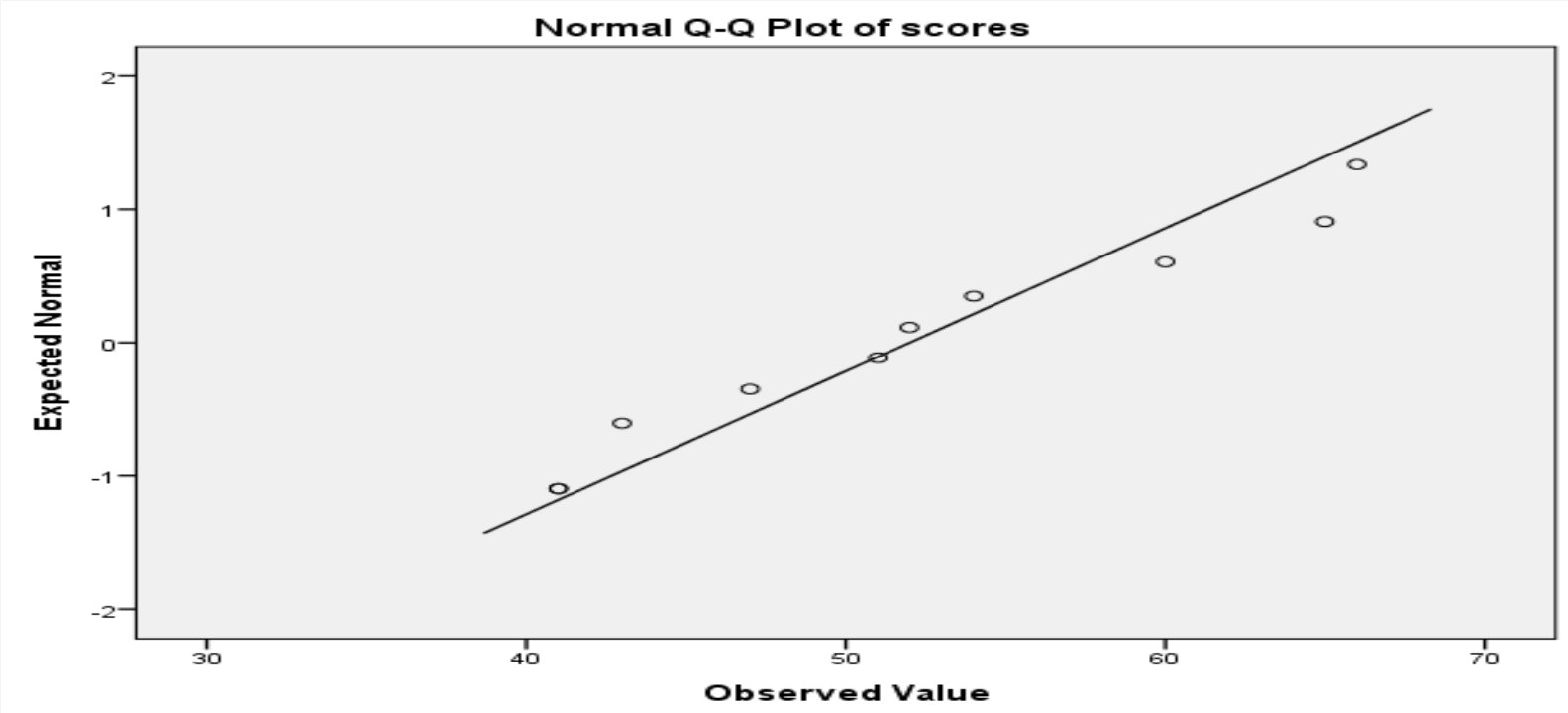
Data View Variable View

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
scores	.133	10	.200*	.917	10	.334

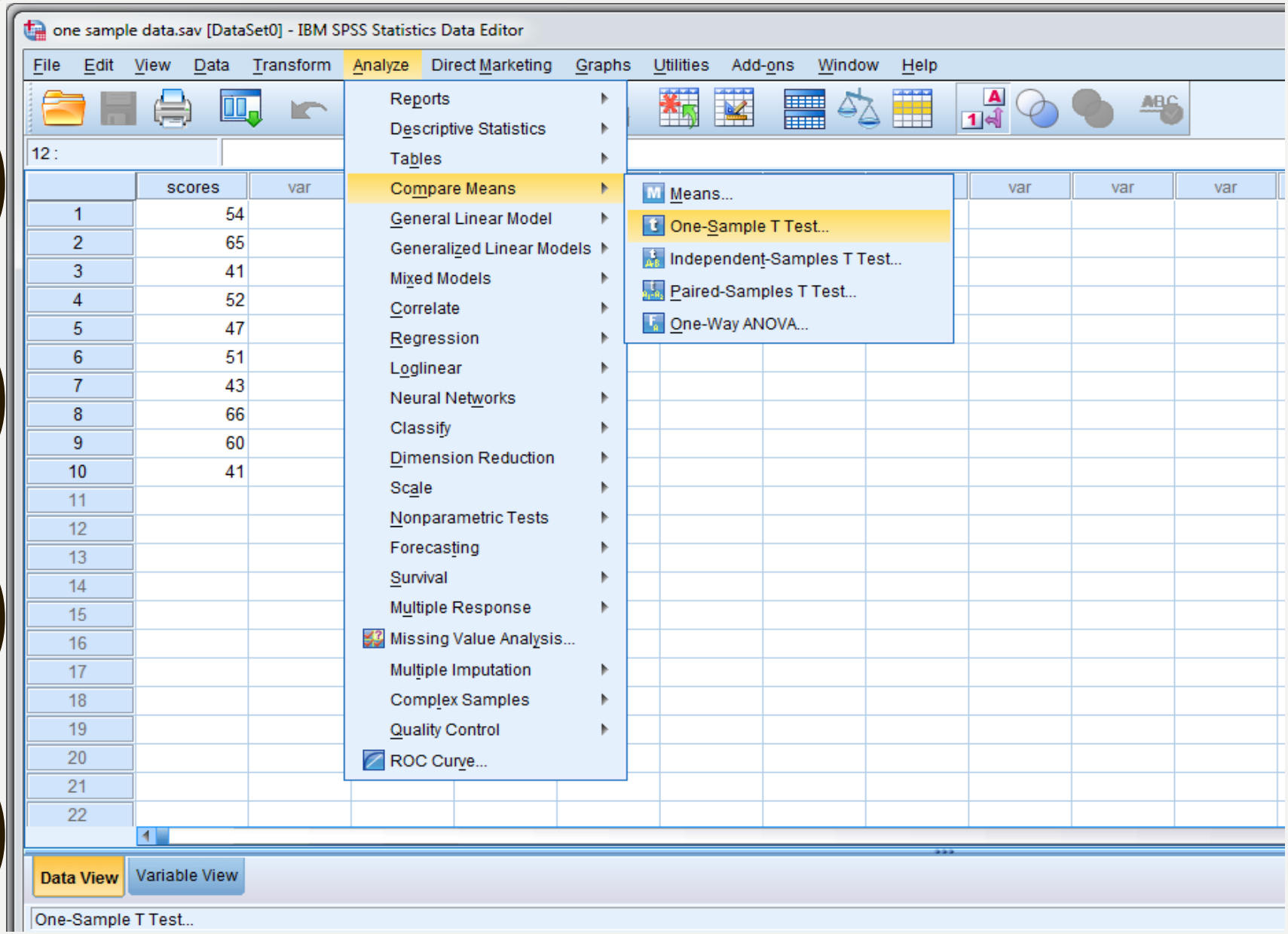
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



- **Value of the Shapiro-Wilk Test is greater than 0.05, the data follows normal.**

ONE SAMPLE T-TEST ANALYSIS



one sample data.sav [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

12 :

	scores	var	var	var	var	var	var	var	var	var	var	var
1	54											
2	65											
3	41											
4	52											
5	47											
6	51											
7	43											
8	66											
9	60											
10	41											
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												

Data View Variable View

One-Sample T Test

Test Variable(s):

scores

Options...

Bootstrap...

Test Value: 60

OK Paste Reset Cancel Help

OUTPUT

NOTE: SPSS does only two-tailed tests. The t-obtained would be the same for a one or two tailed test, but if you are doing a one-tailed test, you can divide the significance by two to calculate the significance of a one-tailed test.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
scores	10	52.00	9.321	2.948

One-Sample Test

	Test Value = 60					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
scores	-2.714	9	.024	-8.000	-14.67	-1.33

RESULTS AND REPORT

Since the p-value (0.024) is less than $\alpha = 0.05$, then we reject the null hypothesis, and so we conclude that the mean score for the class on the test is not equal 60.

Report:

This study was conducted to find out wither the mean score for the student in the math class of the sample is significantly different than the average score of the overall population. At level of significant 0.05, ten students were randomly selected to serve as subjects. By using one sample t-test to analyze the data, the average score was=52 with S.d.=9.32. T-test was 2.714 with df = 9 and p-value 0.024 shows that significantly different than the average score of the overall population at level of significant 0.05. As a results, the teacher now has evidence that the class mean in the course would not equal 60.

PAIRED SAMPLES T-TEST

WHEN TO USE THE PAIRED T-TEST

- A paired t-test is used to compare two population means where you have two samples in which observations in one sample can be paired with observations in the other sample. Examples of where this might occur are:
- Before-and-after observations on the same subjects (e.g. students' diagnostic test results before and after a particular module or course).
- A comparison of two different methods of measurement or two different treatments where the measurements/treatments are applied to the same subjects (e.g. blood pressure measurements using a stethoscope).

Hypothesis testing:

- Null hypothesis H_0: The mean of this sample of differences is zero.
- Alternative hypothesis H_A: The mean is not equal zero.

- Examples
- Before-and-after pairs of measurements after giving a drug

Assumption:

- The dependent variable should be measured at the interval or ratio level (i.e., they are continuous)
- Independent variable should consist of two categorical, "related groups" or "matched pairs".
- There should be no significant outliers in the differences between the two related groups.
- The distribution of the differences in the dependent variable between the two related groups should be approximately normally distributed.

EXAMPLE

- Lets suppose we had a group of 8 students whose accumulative grades were recorded pre and post an improved performance program. Is there a statistically significant difference between the two scores (pre and post) within 5% significance (α=0.05)?

Students	Pre	Post
1	10	12
2	50	52
3	20	25
4	8	10
5	115	120
6	75	80
7	45	50
8	170	175

- Hypothesis:

Null hypothesis: average pre score = average post score

Alternative hypothesis: average pre score = average post score

STEP 1: CALCULATE THE DIFFERENCES:

Students	Pre	Post	Diff.	Diff $(d)^2$
1	10	12	2	4
2	50	52	2	4
3	20	25	5	25
4	8	10	2	4
5	115	120	5	25
6	75	80	5	25
7	45	50	5	25
8	170	175	5	25
			31	137

STEP 2:

- Calculate T statistic:

$\sum d$ = Sum of the differences.

$$T = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D)^2}{n-1}}} = \frac{31}{\sqrt{\frac{8*137-31*31}{8-1}}} = 7.059$$

STEP 3: T TABLE

df	.25	.20	.15	.10	.05	.025
1	1.000	1.376	1.963	3.078	6.314	12.71
2	0.816	1.061	1.386	1.886	2.920	4.303
3	0.765	0.978	1.250	1.638	2.353	3.182
4	0.741	0.941	1.190	1.533	2.132	2.776
5	0.727	0.920	1.156	1.476	2.015	2.571
6	0.718	0.906	1.134	1.440	1.943	2.447
7	0.711	0.896	1.119	1.415	1.895	2.365
8	0.706	0.889	1.108	1.397	1.860	2.306
9	0.703	0.883	1.100	1.383	1.833	2.262
10	0.700	0.879	1.093	1.372	1.812	2.228

STEP 4: COMPARE T CALCULATED WITH T CRITICAL

- Calculated t = 7.06
- Critical t = 2.365

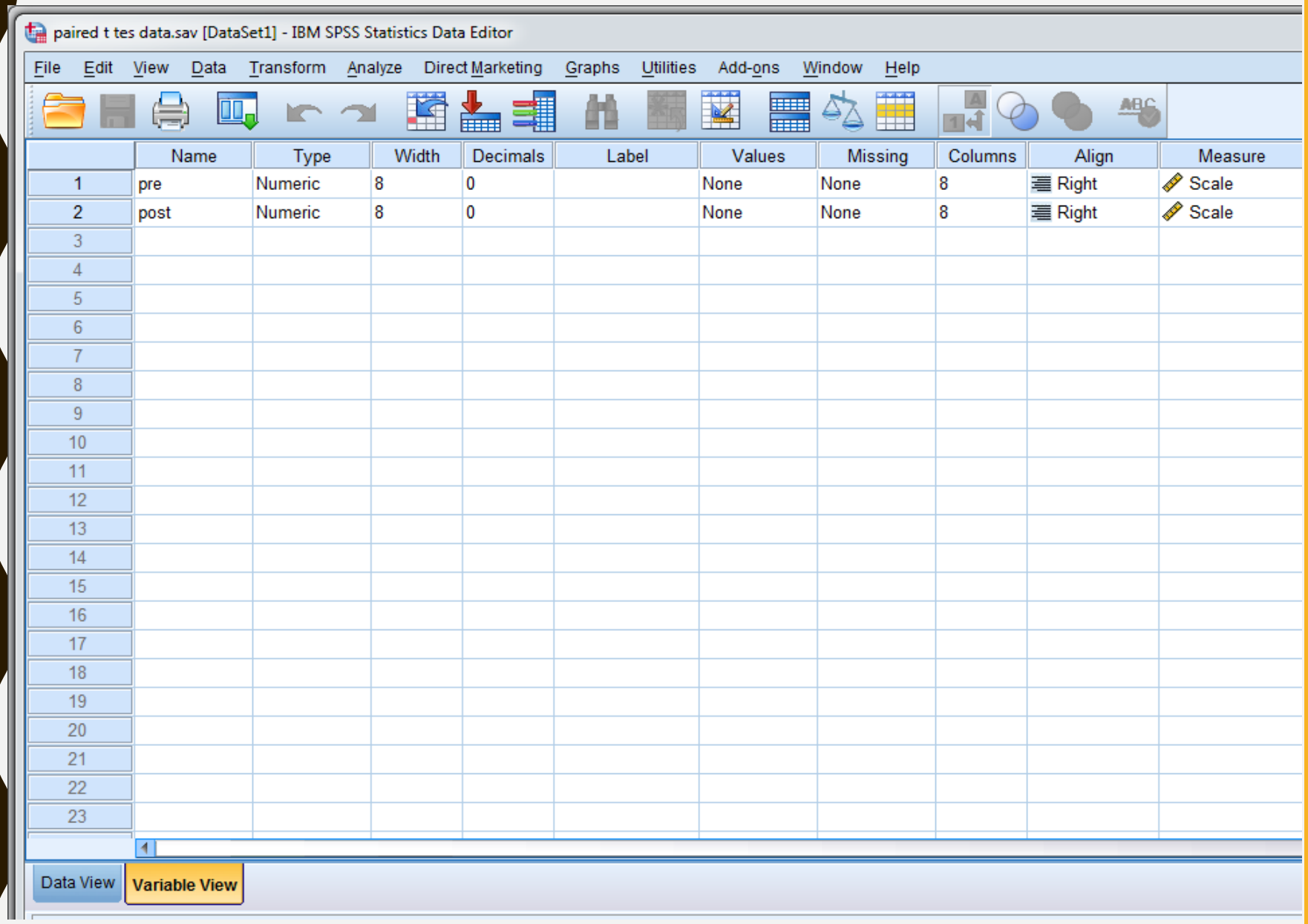
as,

- t calculated > t critical

So,

We reject H_0 and Accept H_a at alpha= 0.05

LETS USE SPSS (DATA ENTERING)



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	pre	Numeric	8	0		None	None	8	Right	Scale
2	post	Numeric	8	0		None	None	8	Right	Scale
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

paired t tes data.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	pre	post	var	var	var	var	var	var	var	var
1	10	12								
2	50	52								
3	20	25								
4	8	10								
5	115	120								
6	75	80								
7	45	50								
8	170	175								
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

Data View Variable View

paired t tes data.sav [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

- Reports
- Descriptive Statistics
- Tables
- Compare Means
  - Means...
  - One-Sample T Test...
  - Independent-Samples T Test...
  - Paired-Samples T Test...
  - One-Way ANOVA...
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Quality Control
- ROC Curve...

8 :

Visible: 2 of 2 Variables

	pre	post	var	var	var	var	var	var	var	var
1	10	1								
2	50	5								
3	20	2								
4	8	1								
5	115	12								
6	75	8								
7	45	5								
8	170	17								
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

Data View Variable View

Paired-Samples T Test...

IBM SPSS Statistics Processor is ready

OUTPUT

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	post	65.50	8	57.540	20.343
	pre	61.63	8	56.642	20.026

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	post & pre	8	1.000	.000

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	post - pre	3.875	1.553	.549	2.577	5.173	7.059	7	.000

REPORT

- A paired-samples t-test was conducted to compare scores pre and post an improved performance program. There was a significant difference in the scores post performance program (M=65.5, SD=57.5) and pre performance program (M=61.6, SD=56.6); t(7)=7.06, p = 0.000. These results suggest that student’s score increased an improved performance program.

INDEPENDENT SAMPLES T-TEST

- The independent samples t-test is probably the single most widely used test in statistics.

- It is used to compare differences between separate groups.

- The independent-samples t-test (independent t-test) compares the means between two unrelated groups on the same continuous, dependent variable. For example, you could use an independent t-test to understand whether there is a difference in test anxiety based on educational level (i.e., your dependent variable would be "test anxiety" and your independent variable would be "educational level", which has two groups: "undergraduates" and "postgraduates").

- Hypothesis testing procedure that uses separate samples for each treatment condition (between subjects design)

NULL AND ALTERNATIVE HYPOTHESES FOR THE INDEPENDENT T-TEST

- The null hypothesis for the independent t-test is that the population means from the two unrelated groups are equal: $H_0: \mu_1 = \mu_2$
- In most cases, we are looking to see if we can show that we can reject the null hypothesis and accept the alternative hypothesis, which is that the population means are not equal: $H_A: \mu_1 \neq \mu_2$

ASSUMPTIONS FOR THE INDEPENDENT T-TEST

- **Independence:** Observations within each sample must be independent (they don't influence each other)
- **Normal Distribution:** The scores in each population must be normally distributed
- **Homogeneity of Variance:** The two populations must have equal variances (the degree to which the distributions are spread out is approximately equal) (Levene's Test for Equality of Variances)

EXAMPLE

- A researcher performed a study which aimed to define the difference in the averages for statistics among scientific department students and literature department students. The sample size was seven students from the scientific department and eight from the literature department; their score shown in the table. Is there a statically significant difference between the two samples within 0.05% significance?

Scientific Department	Literary Department
51	65
50	58
42	76
40	85
55	90
40	60
62	70
60	

65	51
58	50
76	42
85	40
90	55
60	40
70	62
	60
SUM=504	SUM= 400
Mean $\overline{x}_1$ = 504/7=72	Mean $\overline{x}_2$ = 400/8=50

Score for sci. dep.	$x-\overline{x}$	$(x-\overline{x})^2$	Score for lit. dep.	$x-\overline{x}$	$(x-\overline{x})^2$
58	-14	196	50	0	0
76	4	16	42	-8	64
85	13	169	40	-10	100
90	18	324	55	5	25
60	-12	144	40	-10	100
70	-2	4	62	12	144
			60	10	100
SUM=504	SUM=0	SUM=902	SUM= 400	SUM= 0	SUM= 534
$\overline{x}_1$ = 504/7=72	N= 7	$s_1^2 = \frac{902}{6}$ =150.333 $s.d_1$=12.261	$\overline{x}_2$ = 400/8=50	N=8	$s_2^2 = \frac{534}{7}$ =76.285 $s.d_2$=8.734

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

X_1 bar = mean of 1st sample
X_2 bar= mean of 2nd sample
n_1 = sci. sample
n_2 = lit. sample
s_1= Standard deviation of 1st sample
s_2= Standard deviation of 2nd sample

$$t = \frac{72 - 50}{\sqrt{\frac{150.333(7 - 1) + 76.285(8 - 1)}{7 + 8 - 2} \times \left(\frac{1}{7} + \frac{1}{8}\right)}} = \frac{22}{\sqrt{\frac{901.998 + 533.995}{13} \times \left(\frac{15}{56}\right)}}$$

$$= \frac{22}{\sqrt{29.58}} = \frac{22}{5.439} = 4.0445$$

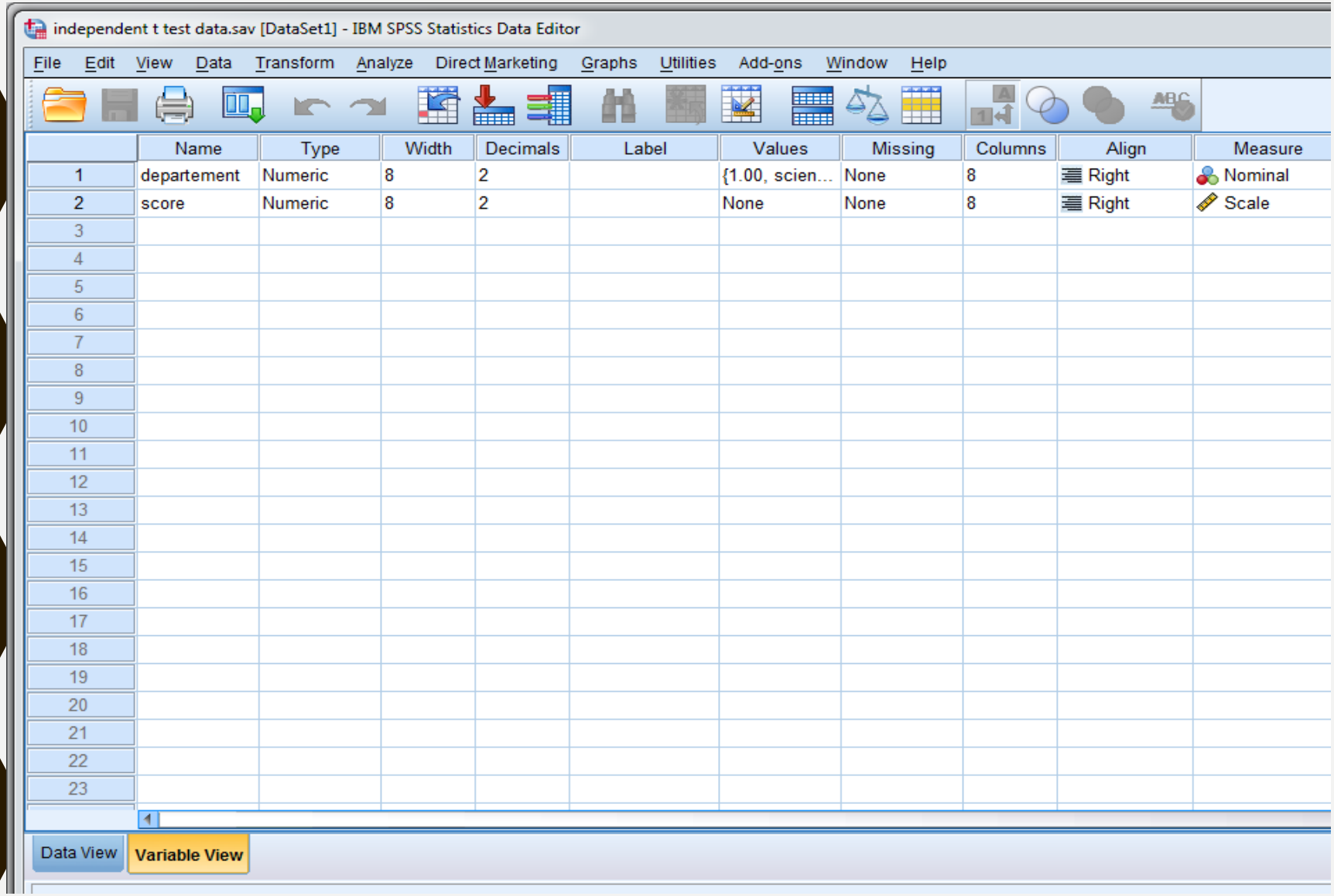
درجات الحرية df	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$ $\frac{\alpha}{2} = 0.025$	$\alpha = 0.01$ $\frac{\alpha}{2} = 0.005$
1	6.31	31.8	12.70	63.65
2	2.92	6.96	4.30	9.92
3	2.35	4.54	3.18	5.84
4	2.13	3.74	2.77	4.60
5	2.01	3.36	2.57	4.03
6	1.94	3.14	2.44	3.70
7	1.89	2.99	2.36	3.49
8	1.86	2.89	2.30	3.35
9	1.83	2.82	2.26	3.25
10	1.81	2.76	2.22	3.16
11	1.7 9	2.71	2.20	3.10
12	1.78	2.68	2.17	3.05
13	1.77	2.65	2.16	3.01
14	1.76	2.62	2.14	2.97
15	1.75	2.60	2.13	2.94

At 13 degree of freedom, t value of 2.16 (critical) < 4.0445 (t calculated)

We reject H_0 and Accept H_A at alpha = 0.05

- **The results:**
  - The calculated t-value of 4.045 is larger in magnitude than the calculated t value of 2.16, therefore we reject the null hypothesis
- **The conclusion:**
  - Scientific department (μ = 72) show significantly more score in than literary department (μ = 50), $t(13) = 4.47$, $p < .05$ (two-tailed).

LETS USE SPSS (DATA ENTERING)



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	departement	Numeric	8	2		{1.00, scien...	None	8	Right	Nominal
2	score	Numeric	8	2		None	None	8	Right	Scale
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

independent t test data.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	departement	score	var	var	var	var	var	var	var	var
1	scientific	65.00								
2	scientific	58.00								
3	scientific	76.00								
4	scientific	85.00								
5	scientific	90.00								
6	scientific	60.00								
7	scientific	70.00								
8	literary	51.00								
9	literary	50.00								
10	literary	42.00								
11	literary	40.00								
12	literary	55.00								
13	literary	40.00								
14	literary	62.00								
15	literary	60.00								
16										
17										
18										
19										
20										
21										
22										

Data View Variable View

independent t test data.sav [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

15 : score 60.00

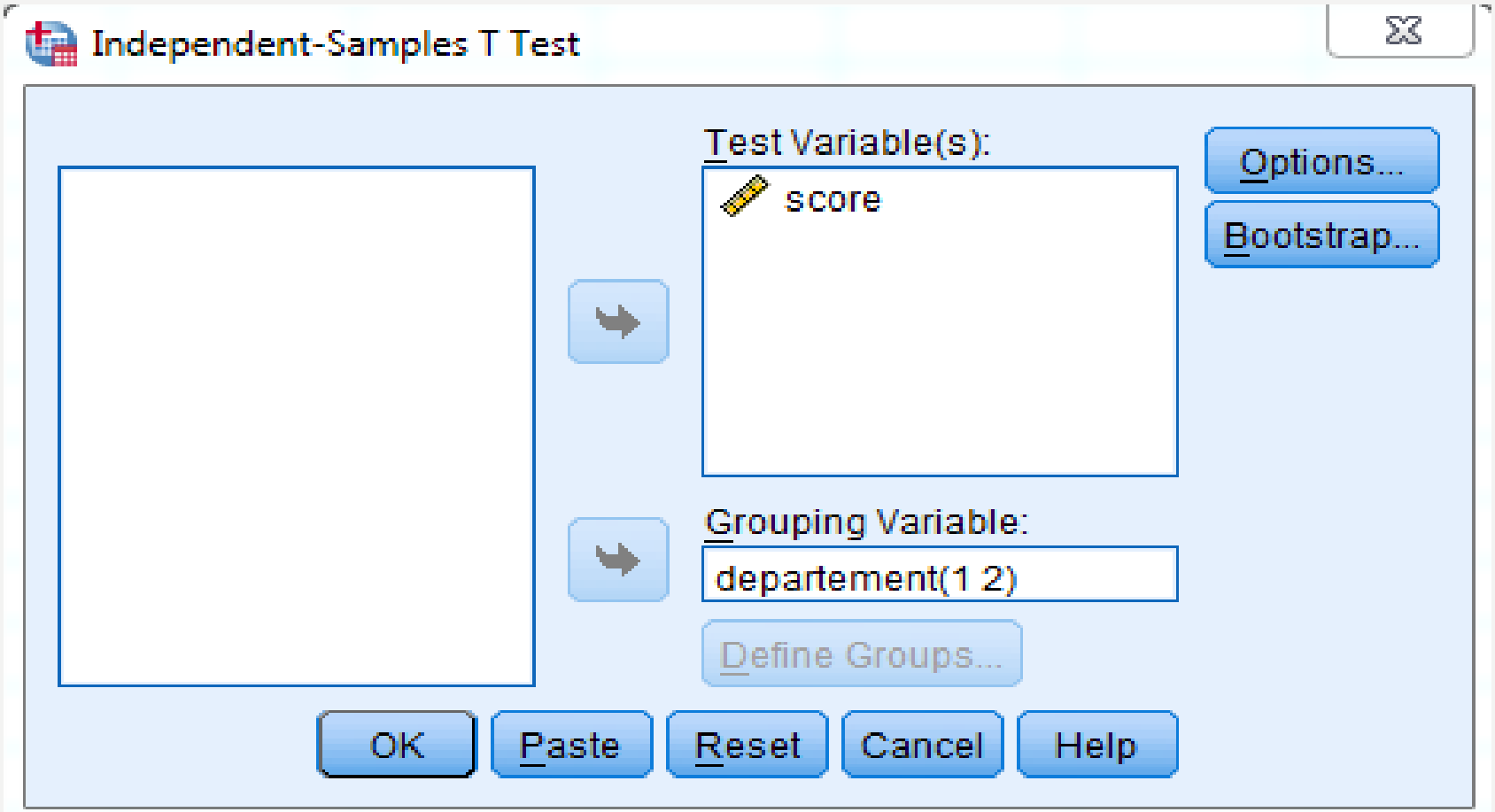
Visible: 2 of 2 Variables

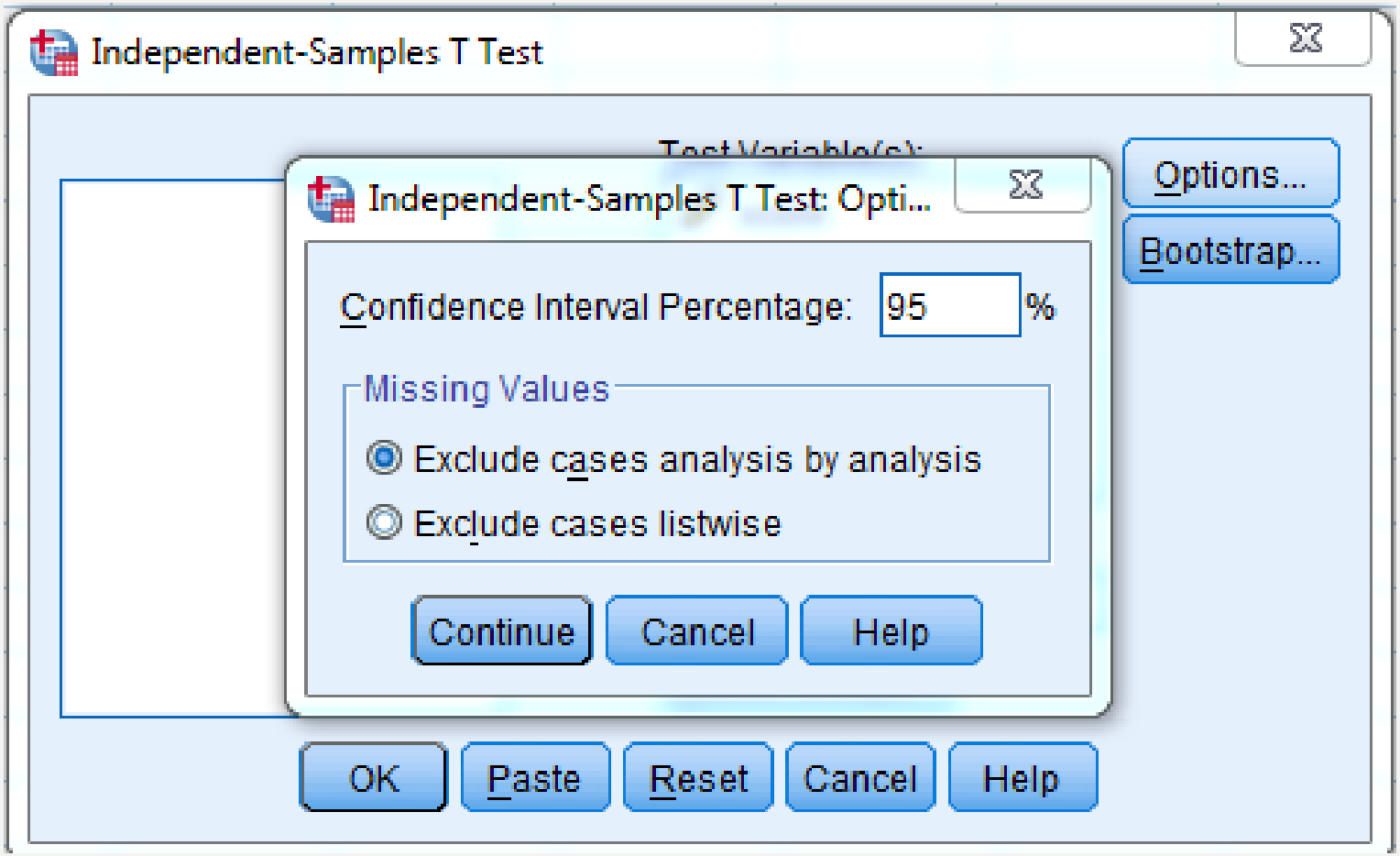
	departement	score
1	scientific	65.0
2	scientific	58.0
3	scientific	76.0
4	scientific	85.0
5	scientific	90.0
6	scientific	60.0
7	scientific	70.0
8	literary	51.0
9	literary	50.0
10	literary	42.0
11	literary	40.0
12	literary	55.0
13	literary	40.0
14	literary	62.0
15	literary	60.0
16		
17		
18		
19		
20		
21		
22		

Analyze:
- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Quality Control
- ROC Curve...

Compare Means:
- Means...
- One-Sample T Test...
- Independent-Samples T Test...
- Paired-Samples T Test...
- One-Way ANOVA...

Data View | Variable View


Independent-Samples T Test
Test Variable(s):
score
Options...
Bootstrap...
Grouping Variable:
departement(1 2)
Define Groups...
OK
Paste
Reset
Cancel
Help


Independent-Samples T Test
Test Variable(s):
Options...
Bootstrap...
Independent-Samples T Test: Opti...
Confidence Interval Percentage: 95 %
Missing Values
Exclude cases analysis by analysis
Exclude cases listwise
Continue
Cancel
Help
OK
Paste
Reset
Cancel
Help

Group Statistics

	departement	N	Mean	Std. Deviation	Std. Error Mean
score	scientific	7	72.0000	12.26105	4.63424
	literary	8	50.0000	8.73417	3.08800

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
score	Equal variances assumed	1.270	.280	4.045	13	.001	22.00000	5.43948	10.24872	33.75128
	Equal variances not assumed			3.951	10.702	.002	22.00000	5.56883	9.70138	34.29862

REPORT

- This study was conducted to find out whether scientific and literary department differ in their scores in statistic course. At level of significant 0.05. By using independent t-test to analyze the data, the average compliance score for scientific was 72 with S.d.=12.26 and for literary the average was 50 with sd = 8.7 . T-test was 4.045 with df= 13 and p-value 0.001 shows that there was a significant difference at level of significant 0.05 between the scientific and literary department.

ONE-WAY ANOVA

ONE-WAY ANOVA

- The one-way ANalysis Of VAriance is used to test the claim that three or more population means are equal.

- This is an extension of the two independent samples t-test.

- The ***response variable*** is the variable you're comparing.

- The ***factor variable*** is the categorical variable being used to define the groups.
  - We will assume *k* samples (groups)

- The *one-way* is because each value is classified in exactly one way.
  - Examples include comparisons by gender, race, political party, color, etc.

Conditions or Assumptions:

- The data are randomly sampled.
- The variances of each sample are assumed equal.
- The residuals are normally distributed.

- The null hypothesis is that the means are all equal:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$

- The alternative hypothesis is that at least one of the means is different

EXAMPLE:

- The statistics classroom is divided into three rows: front, middle, and back.

- The instructor noticed that the further the students were from him, the more likely they were to miss class or use an instant messenger during class.

- He wanted to see at level of sign. 0.05:

Are the students further away did worse on the exams?

The ANOVA doesn't test that one mean is less than another, only whether they're all equal or at least one is different.

$$H_0 : \mu_F = \mu_M = \mu_B$$

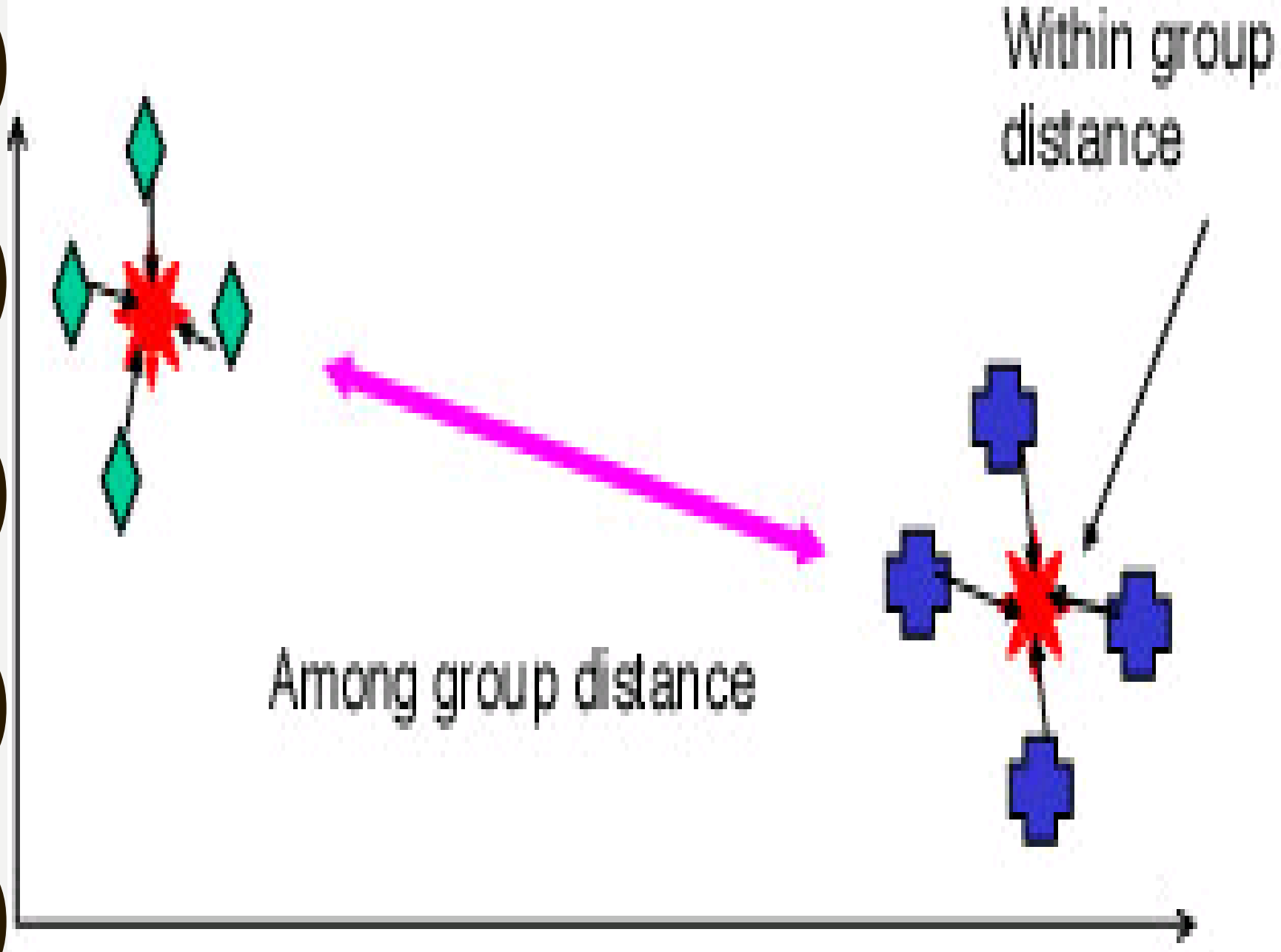
- A random sample of the students in each row was taken.

- The score for those students on the second exam was recorded:
  - Front: 82, 83, 97, 93, 55, 67, 53
  - Middle: 83, 78, 68, 61, 77, 54, 69, 51, 63
  - Back: 38, 59, 55, 66, 45, 52, 52, 61

The summary statistics for the grades of each row are shown in the table below:

Row	FRONT	MIDDLE	BACK
SAMPLE SIZE	7	9	8
MEAN	75.71	67.11	53.50
ST. DEV	17.63	10.95	8.96
VARIANCE	310.90	119.86	80.29

- **VARIATION:**
  - Variation is the Sum of the Squares of the deviations between a value and the mean of the value.
  - Sum of Squares is abbreviated by **SS** and often followed by a variable in parentheses such as **SS(B)** or **SS(W)** so we know which sum of squares we're talking about.


Within group distance
Among group distance

- **ARE ALL OF THE VALUES IDENTICAL?**
  - No, so there is some variation in the data

    This is called the total variation

  - Denoted **SS(Total)** for the total Sum of Squares (variation)
  - Sum of Squares is another name for variation

- **ARE ALL OF THE SAMPLE MEANS IDENTICAL?**
  - No, so there is some variation between the groups.

This is called variation Between Group

  - Sometimes called the variation due to the factor.
  - Denoted **SS(B)** for Sum of Squares (variation) between the groups

- **Are each of the values <u>within each group</u> identical?**
  - No, there is some variation within the groups.

    <u>This is called variation within group</u>

  - Sometimes called the <u>error variation</u>.
  - Denoted **SS(W)** for Sum of Squares (variation) within the groups

- **THERE ARE TWO SOURCES OF VARIATION:**
  - The variation between the groups, **SS(B)**, or the variation due to the factor.
  - The variation within the groups, **SS(W)**, or the variation that can't be explained by the factor so it's called the <u>error variation</u>.

- **Here is the basic one-way ANOVA table**

Source	SS	df	MS	F	p
Between					
Within					
Total					

- **Grand Mean**
  - The grand mean is the average of all the values when the factor is ignored.
  - It is a weighted average of the individual sample means.

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i}$$

$$\bar{\bar{x}} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \mathrm{L} + n_k \bar{x}_k}{n_1 + n_2 + \mathrm{L} + n_k}$$

- Grand Mean for our example is 65.08

$$\bar{\bar{x}} = \frac{7(75.71) + 9(67.11) + 8(53.50)}{7 + 9 + 8}$$

$$\bar{\bar{x}} = \frac{1562}{24}$$

$$\bar{\bar{x}} = 65.08$$

- **Between Group Variation, SS(B)**
  - The between group variation is the variation between each sample mean and the grand mean
  - Each individual variation is weighted by the sample size

$$SS(B) = \sum_{i=1}^{k} n_i \left(\bar{x}_i - \bar{\bar{x}}\right)^2$$

$$SS(B) = n_1\left(\bar{x}_1 - \bar{\bar{x}}\right)^2 + n_2\left(\bar{x}_2 - \bar{\bar{x}}\right)^2 + \mathrm{L} + n_k\left(\bar{x}_k - \bar{\bar{x}}\right)^2$$

The Between Group Variation for our example is SS(B)=1902

$$SS(B) = 7(75.71 - 65.08)^2 + 9(67.11 - 65.08)^2 + 8(53.50 - 65.08)^2$$

$$SS(B) = 1900.8376 \approx 1902$$

- **Within Group Variation, SS(W)**
  - The Within Group Variation is the weighted total of the individual variations.
  - The weighting is done with the degrees of freedom.
  - The df for each sample is one less than the sample size for that sample.

Within Group Variation

$$SS(W) = \sum_{i=1}^{k} df_i s_i^2$$

$$SS(W) = df_1 s_1^2 + df_2 s_2^2 + \mathrm{L} + df_k s_k^2$$

- **The within group variation for our example is 3386**

$$SS(W) = 6(310.90) + 8(119.86) + 7(80.29)$$

$$SS(W) = 3386.31 \approx 3386$$

- After filling in the sum of squares, we have …

Source	SS	df	MS	F	p
Between	1902				
Within	3386				
Total	5288				

- **Degrees of Freedom, df**
  - A degree of freedom occurs for each value that can vary before the rest of the values are predetermined.
  - For example, if you had six numbers that had an average of 40, you would know that the total had to be 240. Five of the six numbers could be anything, but once the first five are known, the last one is fixed so the sum is 240. The df would be 6-1=5
  - The df is often one less than the number of values

- **The <u>Between Group</u> df is one less than the number of groups**
  - We have three groups, so

$$df(B) = 2$$

- **The <u>Within Group</u> df is the sum of the individual df's of each group**
  - The sample sizes are 7, 9, and 8

$$df(W) = 6 + 8 + 7 = 21$$

- The **total df** is one less than the sample size

$$df(Total) = 24 - 1 = 23$$

- Filling in the degrees of freedom gives this …

Source	SS	df	MS	F	p
Between	1902	2			
Within	3386	21			
Total	5288	23			

- **VARIANCES**
  - The variances are also called the **Mean of the Squares** and abbreviated by MS, often with an accompanying variable **MS(B)** or **MS(W)**
  - They are an average squared deviation from the mean and are found by dividing the variation by the degrees of freedom

MS = SS / df

$$Variance = \frac{Variation}{df}$$

- MS(B) = 1902 / 2 = 951.0
- MS(W) = 3386 / 21 = 161.2
- MS(T) = 5288 / 23 = 229.9
  - Notice that the MS(Total) is NOT the sum of MS(Between) and MS(Within).
  - This works for the sum of squares SS(Total), but not the mean square MS(Total)
  - The MS(Total) isn't usually shown

- Completing the MS gives …

Source	SS	df	MS	F	p
Between	1902	2	951.0		
Within	3386	21	161.2		
Total	5288	23			

Special Variances

- The MS(Within) is also known as the pooled estimate of the variance since it is a weighted average of the individual variances

$$S_p^2$$

- The MS(Total) is the variance of the response variable, not technically part of ANOVA table.

F test statistic

- An F test statistic is the ratio of two sample variances.
- The MS(B) and MS(W) are two sample variances and that's what we divide to find F.

$$F = MS(B) / MS(W)$$

- For our data, F = 951.0 / 161.2 = 5.9

- Adding F to the table …

Source	SS	df	MS	F	p
Between	1902	2	951.0	5.9	
Within	3386	21	161.2		
Total	5288	23	229.9		

- The F test is a right tail test
- The F test statistic has an F distribution with df(B) numerator df and df(W) denominator df
- The p-value is the area to the right of the test statistic:

$$P(F_{2,21} > 5.9) = 0.009$$

- Completing the table with the p-value

Source	SS	df	MS	F	p
Between	1902	2	951.0	5.9	0.009
Within	3386	21	161.2		
Total	5288	23			

- The p-value is 0.009, which is less than the significance level of 0.05, so we reject the null hypothesis.

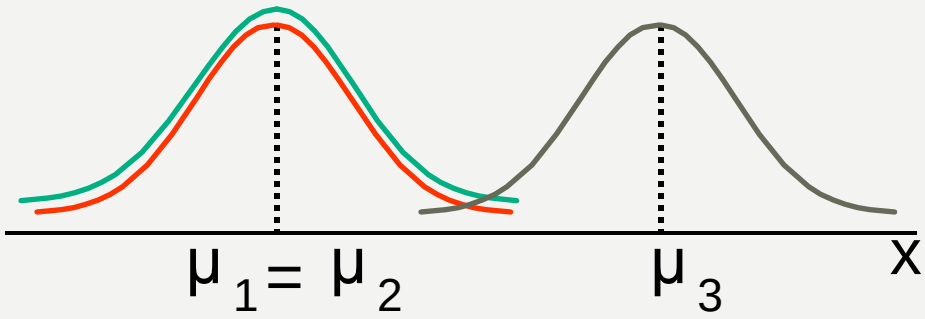
- The null hypothesis is that the means of the three rows in class were the same, but we reject that, so at least one row has a different mean.

- There is enough evidence to support the claim that there is a difference in the mean scores of the front, middle, and back rows in class.

- The ANOVA doesn’t tell which row is different, you would need to look at confidence intervals or run post hoc tests to determine that.

MULTIPLE-COMPARISON PROCEDURE (POST HOC TEST)

- Tells which population means are significantly different
  - e.g.: $\mu_1 = \mu_2 \neq \mu_3$
  - Done after rejection of equal means in ANOVA
- Allows pair-wise comparisons
  - Compare absolute mean differences with critical range



TUKEY-KRAMER CRITICAL RANGE

$$\text{Critical Range} = q_{\alpha}\sqrt{\frac{MSW}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where:

q_{α} = Value from standardized range table with k and N - k degrees of freedom for the desired level of α.

MSW = Mean Square Within

n_i and n_j = Sample sizes from populations (levels) i and j

SPSS-statistical Example

EXAMPLE OF ONE WAY ANOVA

- A manager wants to raise the productivity at his company by increasing the speed at which his employees can use a particular spreadsheet program.

EXAMPLE OF ONE WAY ANOVA

- He employs an external agency which provides training in this spreadsheet program. They offer 3 courses: a beginner, intermediate and advanced course.
- He is unsure which course is needed for the type of work they do at his company.

beginner	intermediate	advanced
5 employees	5 employees	5 employees

- When they all return from the training,
- problem to solve using the spreadsheet program, → time to complete the problem.
- (beginner, intermediate, advanced) → differences in the average time it took to complete the problem.

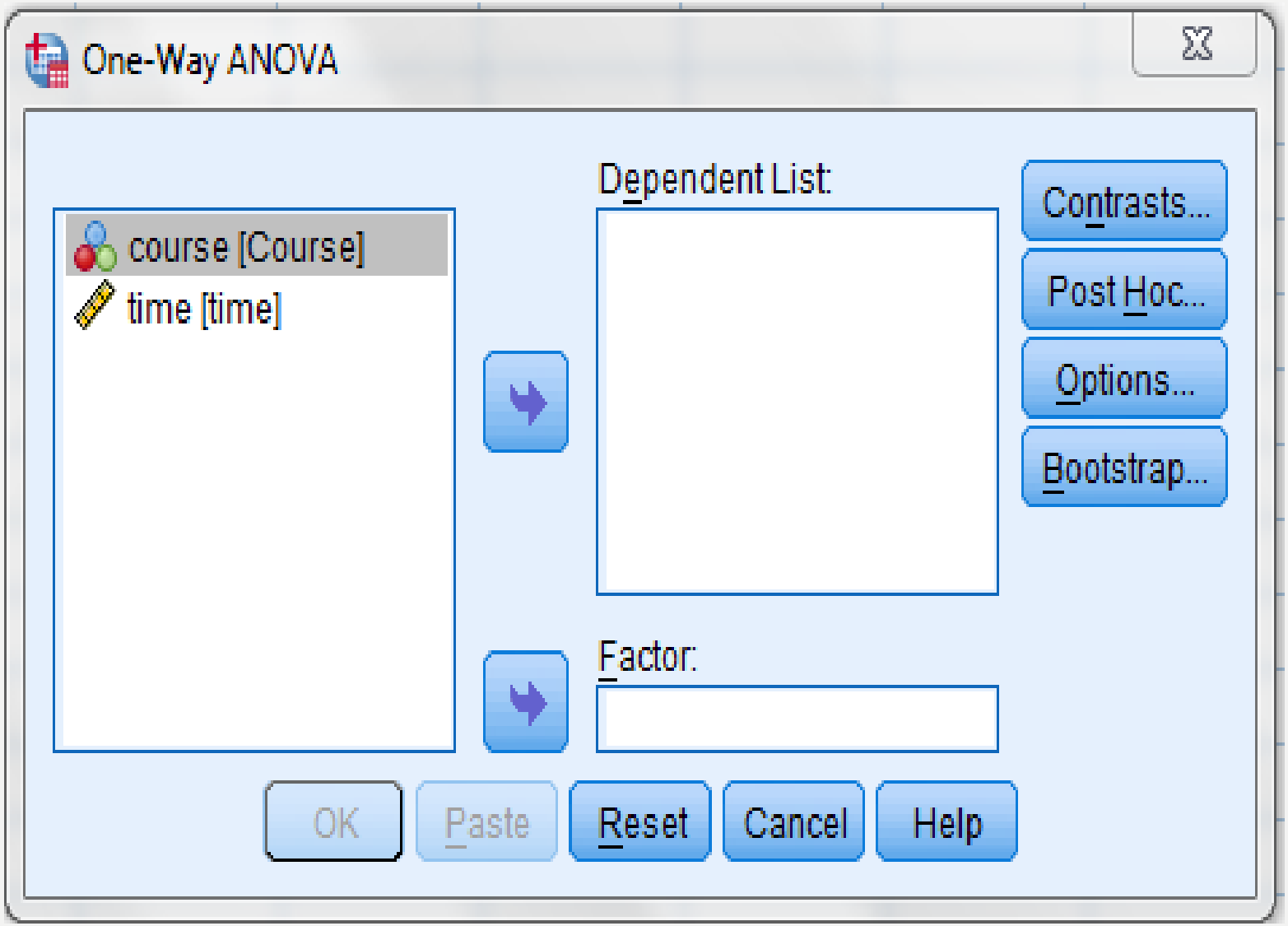
Open data file called "ANOVA_1" and then follow the following steps:

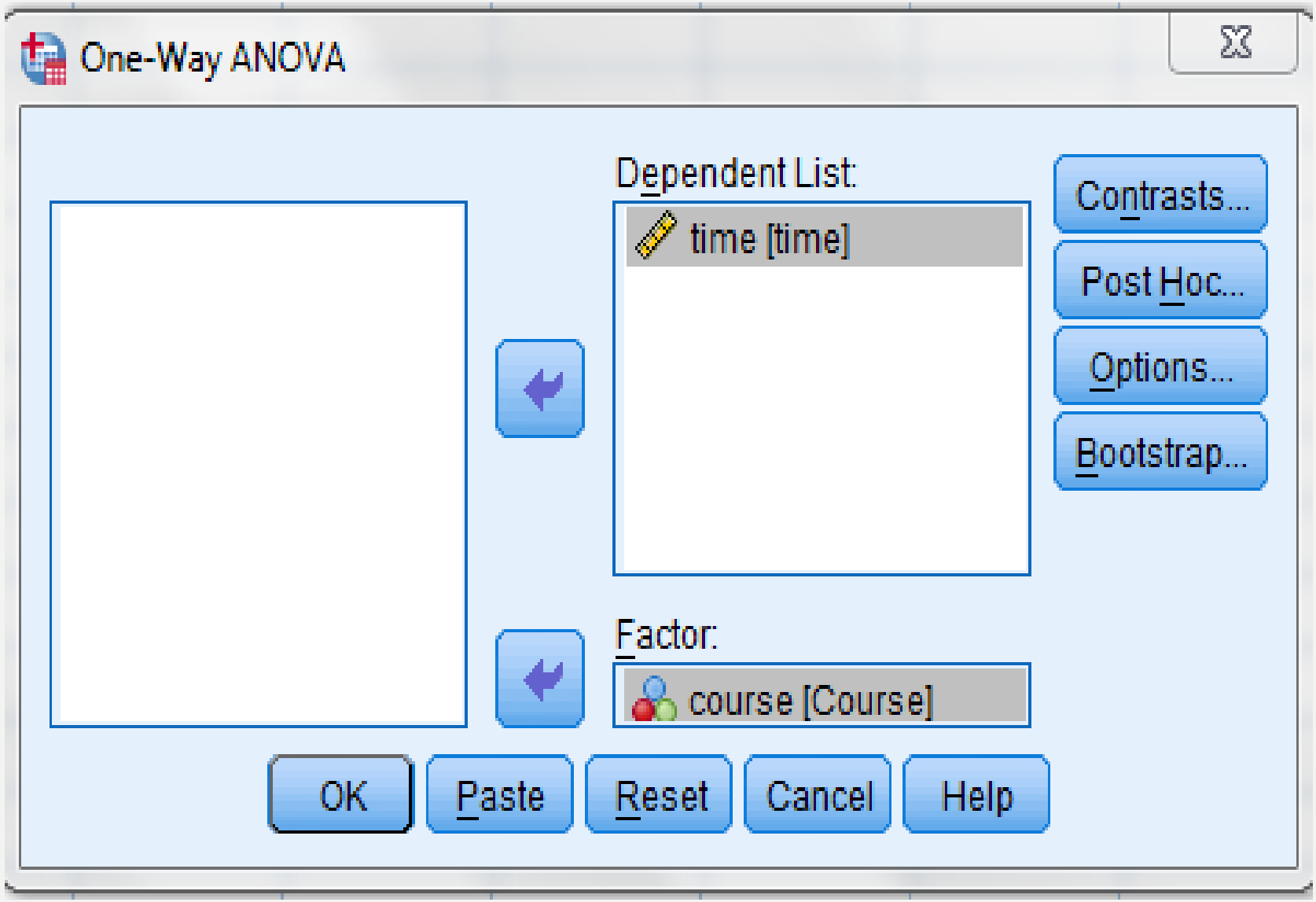
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

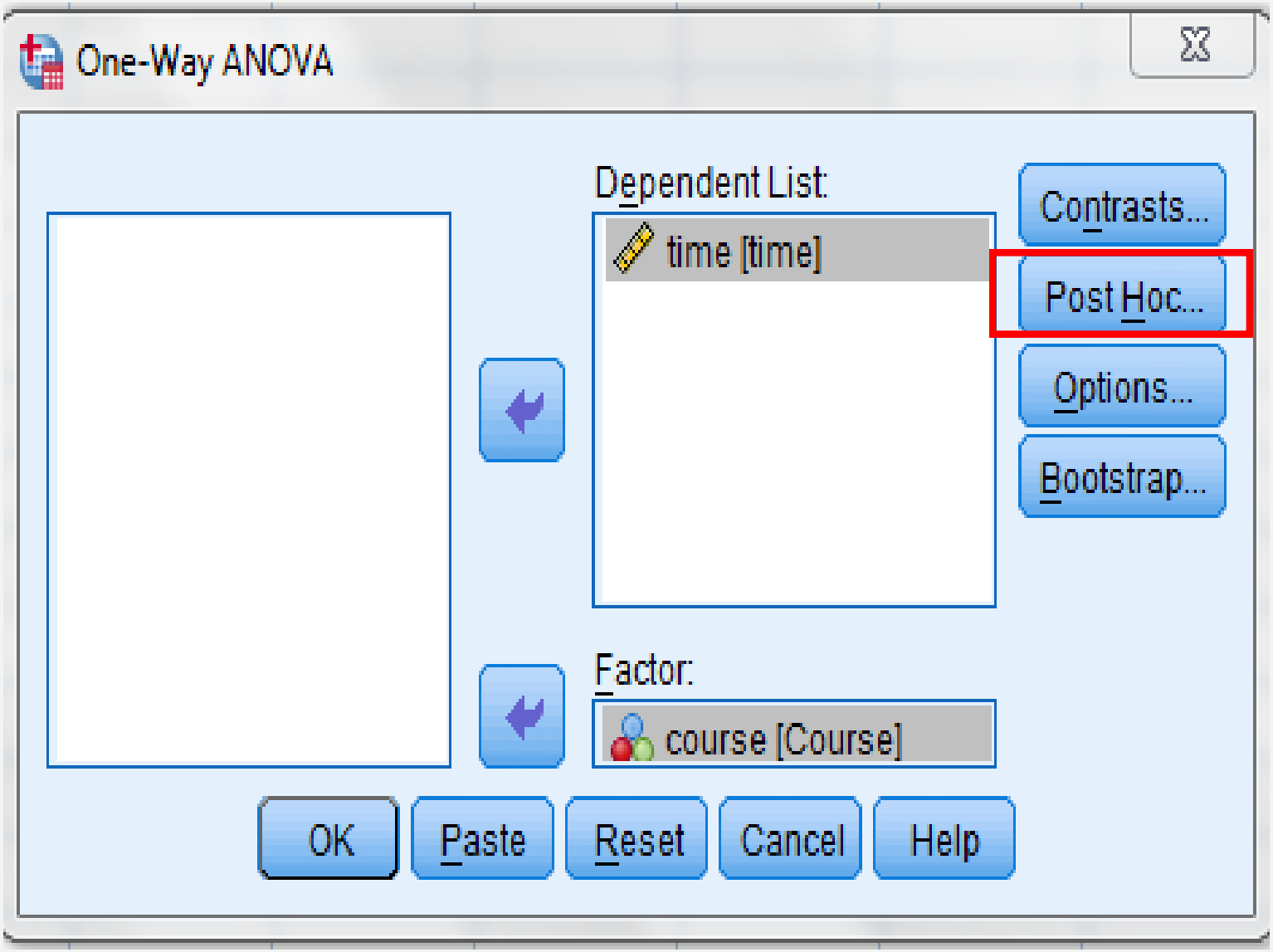
16 : Course

	Course	time
1	beginners	22.
2	beginners	25.
3	beginners	24.
4	beginners	28.
5	beginners	27.
6	intermediate	28.
7	intermediate	27.
8	intermediate	29.
9	intermediate	29.
10	intermediate	33.
11	Advanced	18.
12	Advanced	21.
13	Advanced	20.
14	Advanced	26.
15	Advanced	23.
16		
17		
18		
19		
20		
21		
22		

- Reports
- Descriptive Statistics
- Tables
- Compare Means
  - Means...
  - One-Sample T Test...
  - Independent-Samples T Test...
  - Paired-Samples T Test...
  - One-Way ANOVA...
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Quality Control
- ROC Curve...


One-Way ANOVA
course [Course]
time [time]
Dependent List:
Factor:
Contrasts...
Post Hoc...
Options...
Bootstrap...
OK
Paste
Reset
Cancel
Help


One-Way ANOVA
Dependent List:
time [time]
Contrasts...
Post Hoc...
Options...
Bootstrap...
Factor:
course [Course]
OK
Paste
Reset
Cancel
Help


One-Way ANOVA
Dependent List:
time [time]
Contrasts...
Post Hoc...
Options...
Bootstrap...
Factor:
course [Course]
OK
Paste
Reset
Cancel
Help

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
- [x] Tukey
- [ ] Tukey's-b
- [ ] Duncan
- [ ] Hochberg's GT2
- [ ] Gabriel
- [ ] Waller-Duncan

Type I/Type II Error Ratio: 100

- [ ] Dunnett

Control Category : Last

Test

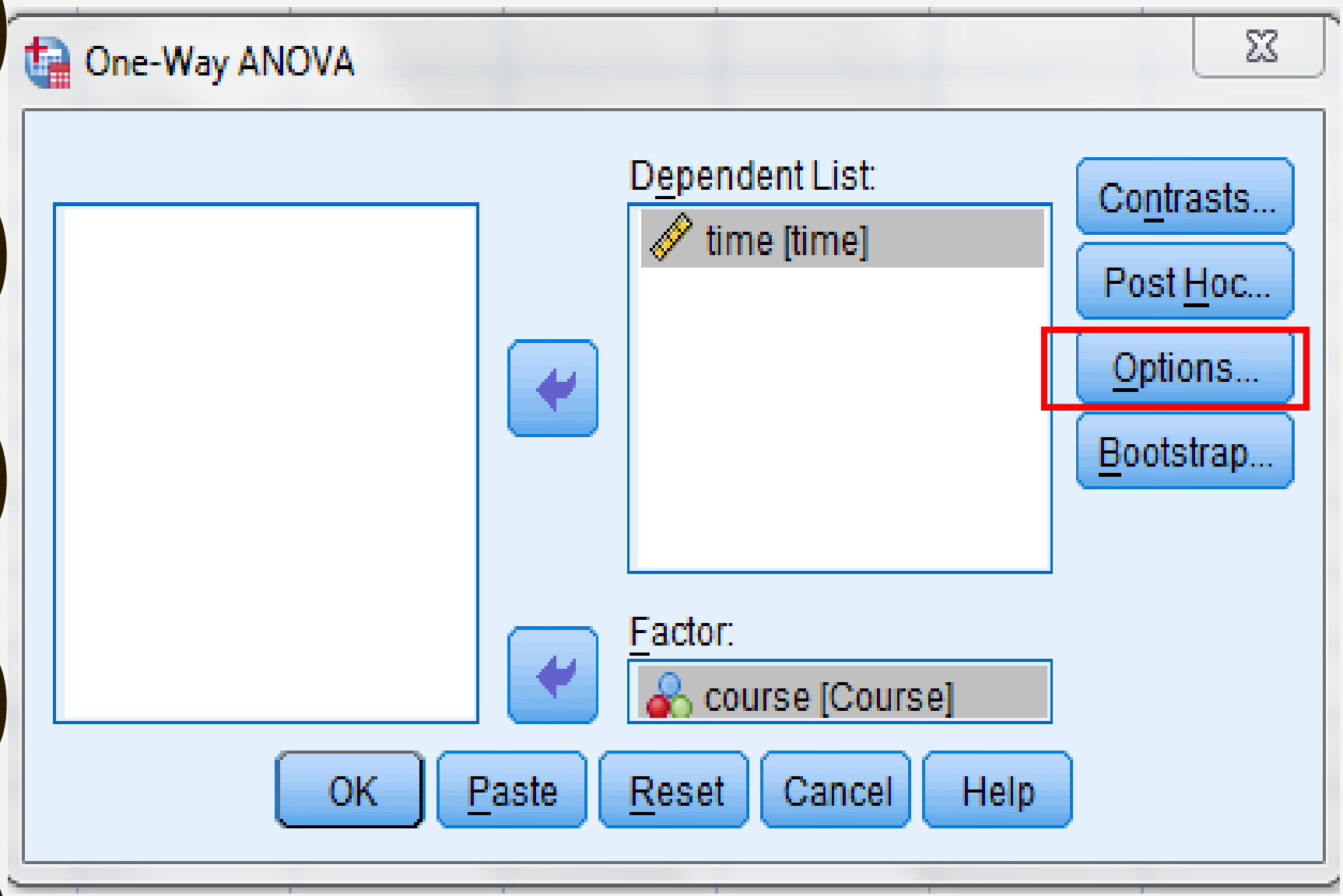
- (•) 2-sided
- ( ) < Control
- ( ) > Control

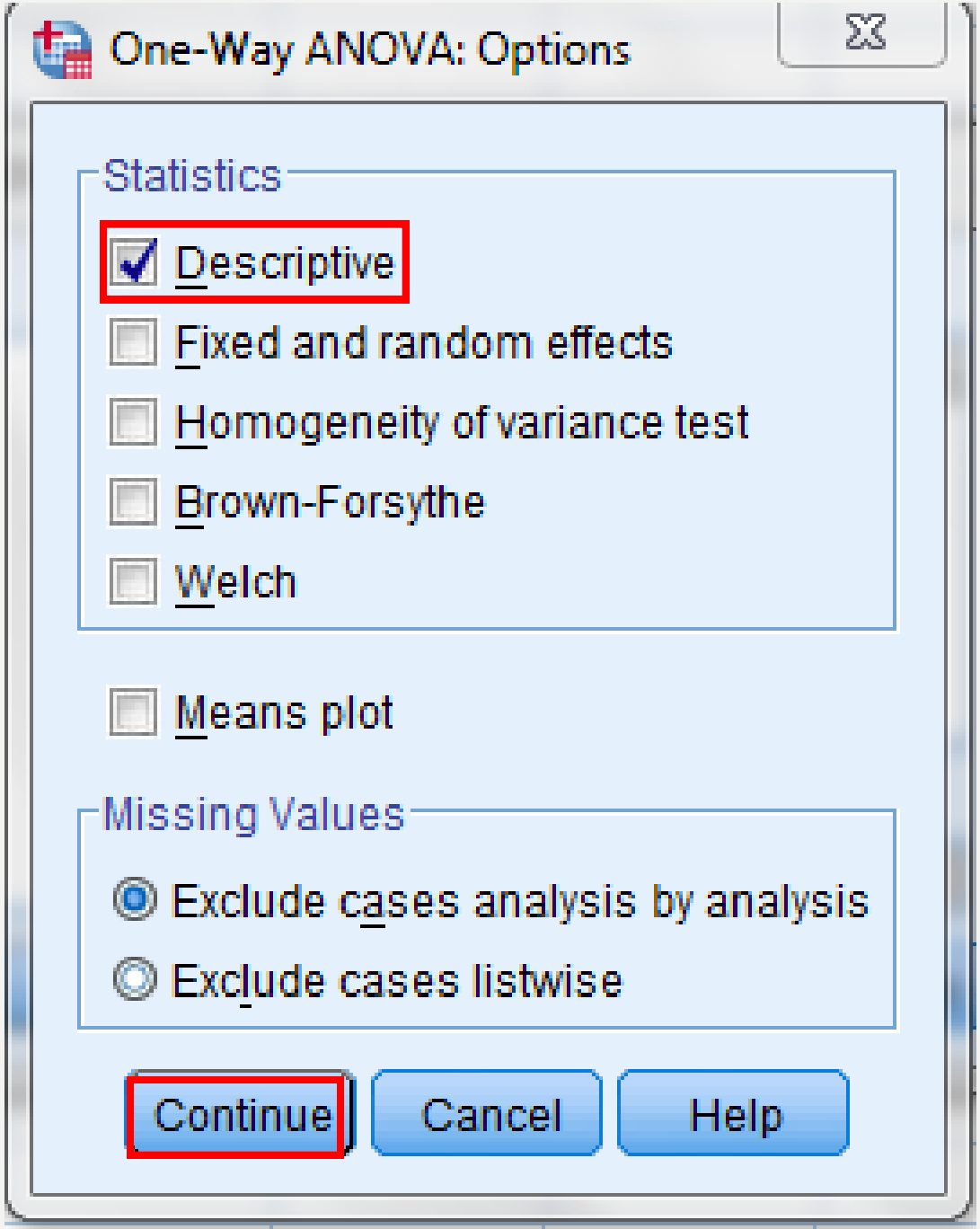
Equal Variances Not Assumed

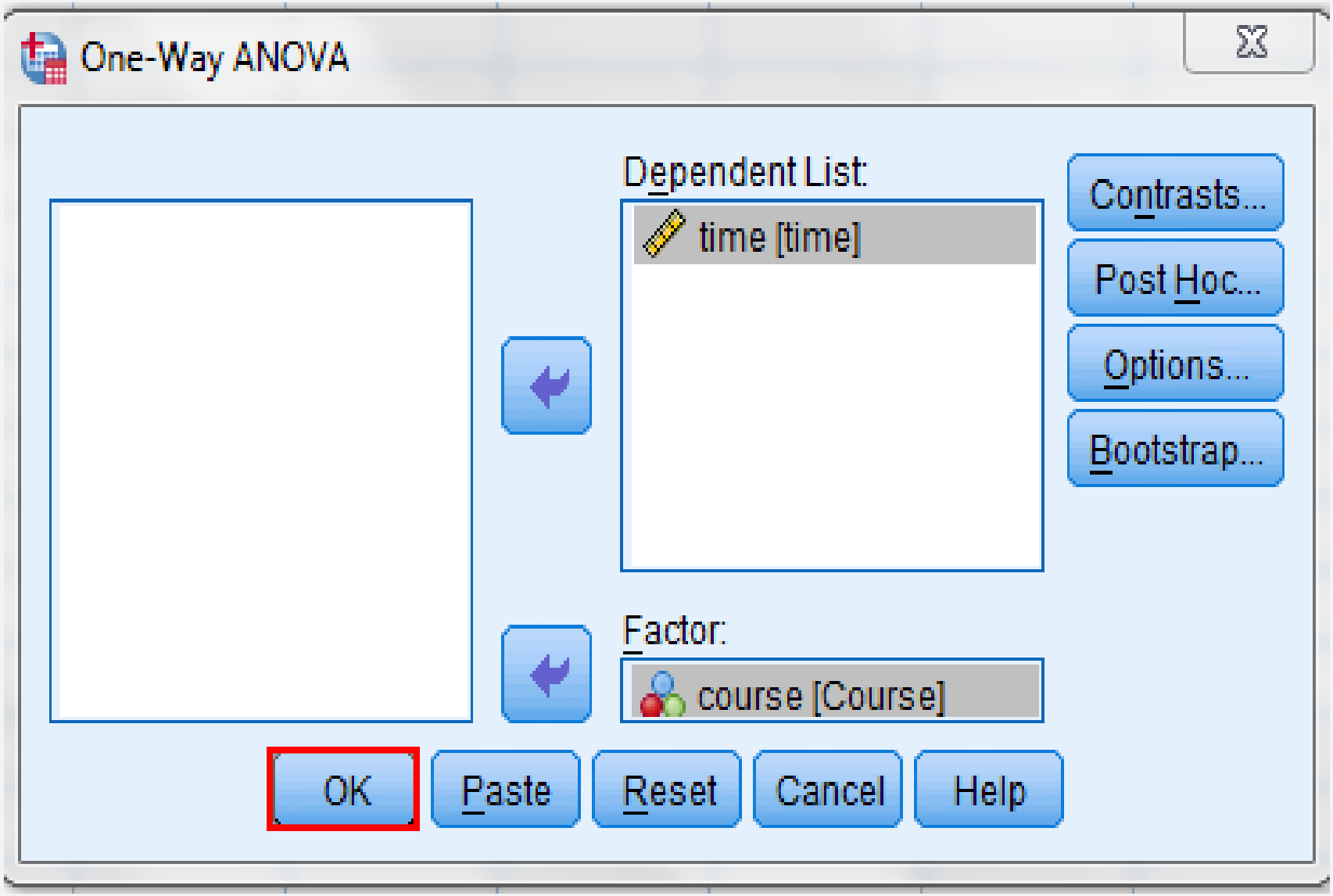
- [ ] Tamhane's T2
- [ ] Dunnett's T3
- [ ] Games-Howell
- [ ] Dunnett's C

Significance level: 0.05

Continue | Cancel | Help


One-Way ANOVA
Dependent List:
time [time]
Contrasts...
Post Hoc...
Options...
Bootstrap...
Factor:
course [Course]
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


One-Way ANOVA
Dependent List:
time [time]
Contrasts...
Post Hoc...
Options...
Bootstrap...
Factor:
course [Course]
OK
Paste
Reset
Cancel
Help

SPSS STATISTICS OUTPUT OF THE ONE-WAY ANOVA

DESCRIPTIVE TABLE

Descriptives

time

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
beginners	5	25.2000	2.38747	1.06771	22.2356	28.1644	22.00	28.00
intermediate	5	29.2000	2.28035	1.01980	26.3686	32.0314	27.00	33.00
Advanced	5	21.6000	3.04959	1.36382	17.8134	25.3866	18.00	26.00
Total	15	25.3333	4.01189	1.03586	23.1116	27.5550	18.00	33.00

ANOVA TABLE

- This is the table that shows the output of the ANOVA analysis and whether we have a statistically significant difference between our group means.

ANOVA

time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	144.533	2	72.267	10.733	.002
Within Groups	80.800	12	6.733		
Total	225.333	14			

there is a statistically significant difference in the mean length of time to complete the spreadsheet problem between the different courses taken

MULTIPLE COMPARAISONS TABLE

Multiple Comparisons

Dependent Variable: time

Tukey HSD

(I) course	(J) course	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
beginners	intermediate	-4.00000-	1.64114	.075	-8.3783-	.3783
	Advanced	3.60000	1.64114	.113	-.7783-	7.9783
intermediate	beginners	4.00000	1.64114	.075	-.3783-	8.3783
	Advanced	7.60000*	1.64114	.002	3.2217	11.9783
Advanced	beginners	-3.60000-	1.64114	.113	-7.9783-	.7783
	intermediate	-7.60000-*	1.64114	.002	-11.9783-	-3.2217-

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

REPORTING THE OUTPUT OF THE ONE-WAY ANOVA

- There was a statistically significant difference between groups as determined by one-way ANOVA ($F(2,12) = 10.733$, $p = .002$).
- A Tukey post-hoc test revealed that the time to complete the problem was statistically significantly lower after taking the Advanced (21.6 ± 3.0 min) compared to the Intermediate (29.2 ± 2.2 min) course and the beginners course (25.2 ± 2.3 min). There were no statistically significant differences between the beginners and intermediate groups ($p = .075$) or between the beginners and Advanced group (p=.113) while it was significance between the intermediate and Advanced (p=.002).

TWO-WAY ANOVA

A TWO-WAY ANOVA is useful when we desire to compare the effect of multiple levels of **two factors** and we have **multiple observations at each level**.

- "Two-Way" means groups are defined by 2 independent variables.
- These IVs are typically called *factors*.
- An experiment in which any combination of values for the 2 factors can occur is called a *completely crossed factorial design.*
- If all cells have the same *n*, the design is said to be *balanced*.
- Still have only 1 dependent variable

- **What kind of variables?**

Continuous (scale / interval / ratio) and 2 independent categorical variables (factors)

- **Common Applications:**

Comparing means of a single variable at different levels of two conditions (factors) in scientific experiments.

- **THE VARIABLES IN THE TWO-WAY ANOVA**

- **There are two kinds of variables Two-Way ANOVA**

ONE DEPENDENT VARIABLE

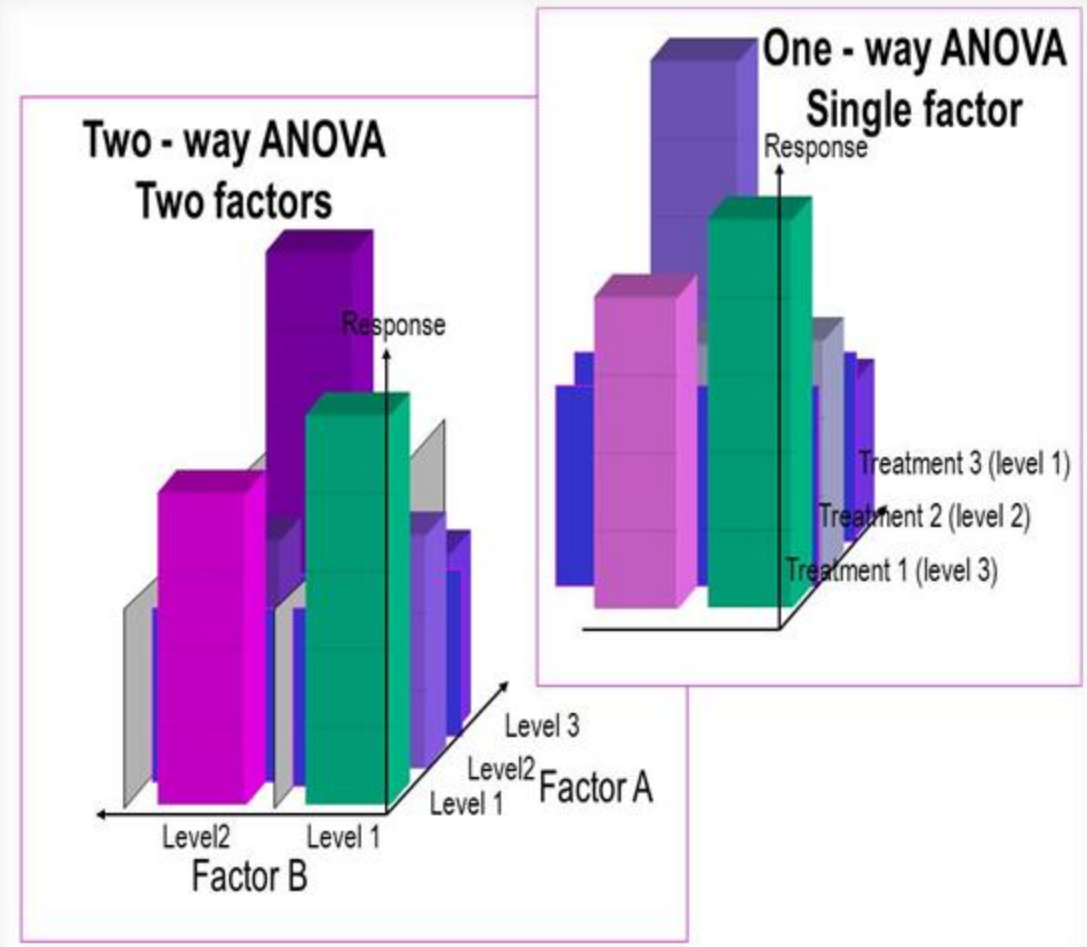
TWO INDEPENDENT VARIABLE

ASSUMPTIONS

- Independent random samples are drawn
- Populations are normally distributed
- Populations have equal variances

- A two-way ANOVA always involves two independent variables.

- Each independent variable, is made up of, or defined by, two or more elements called levels.



- **An Example:**

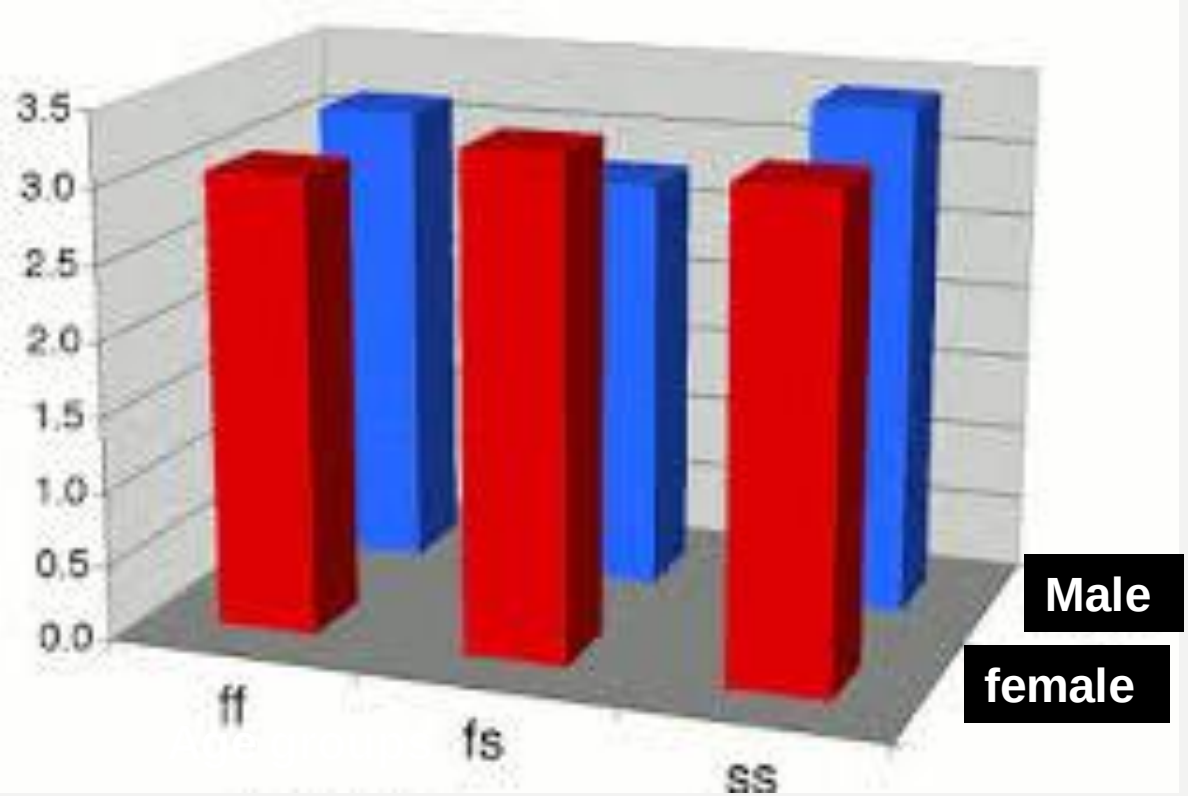
- Let us suppose that we desire to know if patient phobia/ Anxiety in the clinic varies according to **age** and **gender**.
- The variable of interest is therefore patient phobia .
- There are two **factors** being studied - age and gender.

Further suppose that the patients have been classified into three groups or **levels**:

1/ age less than 30, 2/ 30 to 40 3/ above 40

In addition patients have been labeled into gender classification **(levels)**:

- male
- Female


3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
ff
fs
ss
Male
female

- **Testing for Interaction**

There are two versions of the Two-Way ANOVA:

1. The basic version has one observation in each cell - one phobia score from one patient each of the six cells.
2. The second version has more than one observation per cell but the number of observations in each cell must be equal. The advantage of the second version is it also helps us to test if there is any interaction between the two factors.

For instance, in the example above, we may be interested to know if there is any interaction between age and gender.

SOURCES OF VARIATION

- **Suppose that the two factors of interest are: A and B, and**
- a = number of levels of factor A
- b = number of levels of factor B
- N = total number of observations in all cells

$$SST = SS_A + SS_B + SS_{AB} + SSE$$

Degrees of Freedom:

SST
Total Variation

N - 1

SS_A
Variation due to factor A

a – 1

SS_B
Variation due to factor B

b – 1

SS_{AB}
Variation due to interaction between A and B

(a – 1)(b – 1)

SSE
Inherent variation (Error)

N – ab

MEAN SQUARE CALCULATIONS

$$MS_A = \text{Mean square factor A} = \frac{SS_A}{a-1}$$

$$MS_B = \text{Mean square factor B} = \frac{SS_B}{b-1}$$

$$MS_{AB} = \text{Mean square interaction} = \frac{SS_{AB}}{(a-1)(b-1)}$$

$$MSE = \text{Mean square error} = \frac{SSE}{N-ab}$$

TWO-WAY ANOVA: THE F TEST STATISTIC

F Test for Factor A Main Effect

H_0: $\mu_{A1} = \mu_{A2} = \mu_{A3} = \cdots$

H_A: Not all μ_{Ai} are equal

$$F = \frac{MS_A}{MSE}$$

Reject H_0 if $F > F_\alpha$

F Test for Factor B Main Effect

H_0: $\mu_{B1} = \mu_{B2} = \mu_{B3} = \cdots$

H_A: Not all μ_{Bi} are equal

$$F = \frac{MS_B}{MSE}$$

Reject H_0 if $F > F_\alpha$

F Test for Interaction Effect

H_0: factors A and B do not interact to affect the mean response

H_A: factors A and B do interact

$$F = \frac{MS_{AB}}{MSE}$$

Reject H_0 if $F > F_\alpha$

TWO-WAY ANOVA SUMMARY TABLE

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Statistic
Factor A	SS_A	**a – 1**	$MS_A = SS_A/(a-1)$	$\frac{MS_A}{MSE}$
Factor B	SS_B	**b – 1**	$MS_B = SS_B/(b-1)$	$\frac{MS_B}{MSE}$
AB (Interaction)	SS_{AB}	**(a – 1)(b – 1)**	$MS_{AB} = SS_{AB}/[(a-1)(b-1)]$	$\frac{MS_{AB}}{MSE}$
Error	**SSE**	**N – ab**	**MSE =** $SSE/(N-ab)$	
Total	**SST**	**N – 1**		

- Degrees of freedom always add up
  - $N-1 = (N-ab) + (a-1) + (b-1) + (a-1)(b-1)$
  - Total = error + factor A + factor B + interaction
- The denominator of the F Test is always the same but the numerator is different
- The sums of squares always add up
  - $SST = SSE + SS_A + SS_B + SS_{AB}$
  - Total = error + factor A + factor B + interaction

GENERAL EXAMPLE

A researcher is interested in whether emergency rooms patients get more agitated when the machine monitoring their vital signs emits lots of noises. He measures their pulse as a way to index their agitation, and have some of the patients hooked up to machines that are very noisy or machines that have the volume off.

DV = Pulse (beats per minute)

IV1: Monitor Volume (volume on, volume off)

He is also interested in whether men or women tend to have higher pulse rates when in the ER, and in particular, whether each gender responds differently to the presence of noisy monitoring devices.

IV2: Gender (male, female)

A: Monitor Volume (volume on, volume off)

B: Gender (male, female)

A x B: Interaction of monitor volume and gender

	Male	**Female**
Volume on		
Volume off		

KIND OF THE INFORMATION

MAIN EFFECT FOR IV1:

Monitor Volume (volume on, volume off)

Overall, did pulse rate differ on average as a function of monitor volume?

Did the monitor volume matter?

MAIN EFFECT FOR IV2:

Gender (male, female)

Overall, did pulse rates differ on average as a function of gender?

Did gender matter?

INTERACTION IV1xIV2:

Interaction of monitor volume and gender

Did the influence of the monitor's volume on pulse rates depend on whether the person was male or female?

	Male	**Female**
Volume on		
Volume off		

SOME TERMINOLOGY

2x2 2x3 2x4 factorial design

Number of numbers = how many factors (IVs)

How many factors = how many “ways”

one factor = one-way, two factors = two-way, etc.

Number itself = how many levels of that IV

Total number of conditions = product

So a 2x2 design has 4 conditions; a 2x3 has 6 conditions

MAIN EFFECTS

		IV2	
		Male	Female
IV1	Volume on	100	80
	Volume off	70	60

Cell means = means for each condition

So the males whose heart rate was measured with the volume on average had a pulse of 100

The womens whose heart rate was measured with the volume on had a pulse of 80 on average

	IV2			
	Male	**Female**		
IV1 Volume on	100	80	(90)	(100+80)/2
Volume off	70	60	(65)	(70+60)/2
	(85)	(70)		
	(100+70)/2	(80+60)/2		

Marginal means = means for main effects (average of cell means in respective column or row)

		IV2		
		Male	**Female**	
IV1	**Volume on**	100	80	(90)
	Volume off	70	60	(65)
		(85)	(70)	

Main effect for IV1:

Does monitor volume matter? Does it effect pulse rate? 90 vs. 65

Main effect for IV2:

Does gender matter?

Does it effect pulse rate? 85 vs. 70

A main effect lets you know overall whether that IV influenced the DV, ignoring the other IV

Refer to marginal means when interpreting main effects. Look at *F* value and corresponding *p* value for each main effect to see whether it is significant. Is it probably a real effect?

Using a factorial design is like having two separate studies rolled into one:

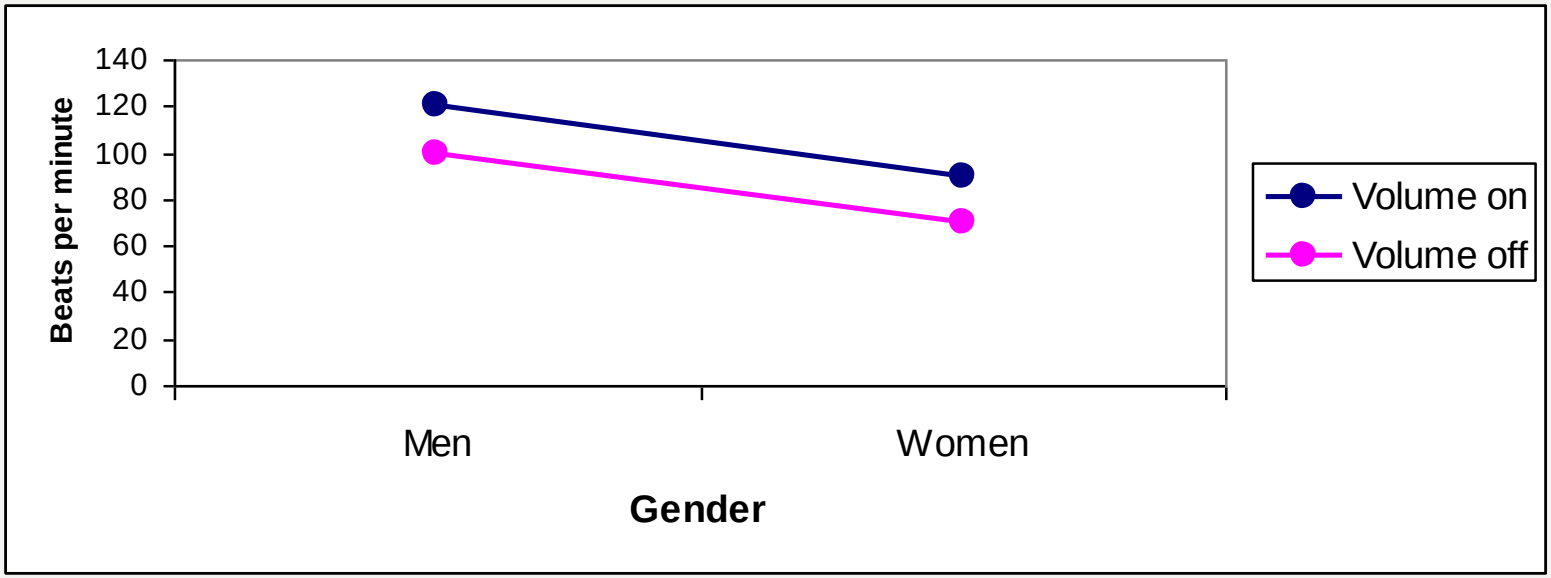
You get to see the overall effects of each variable

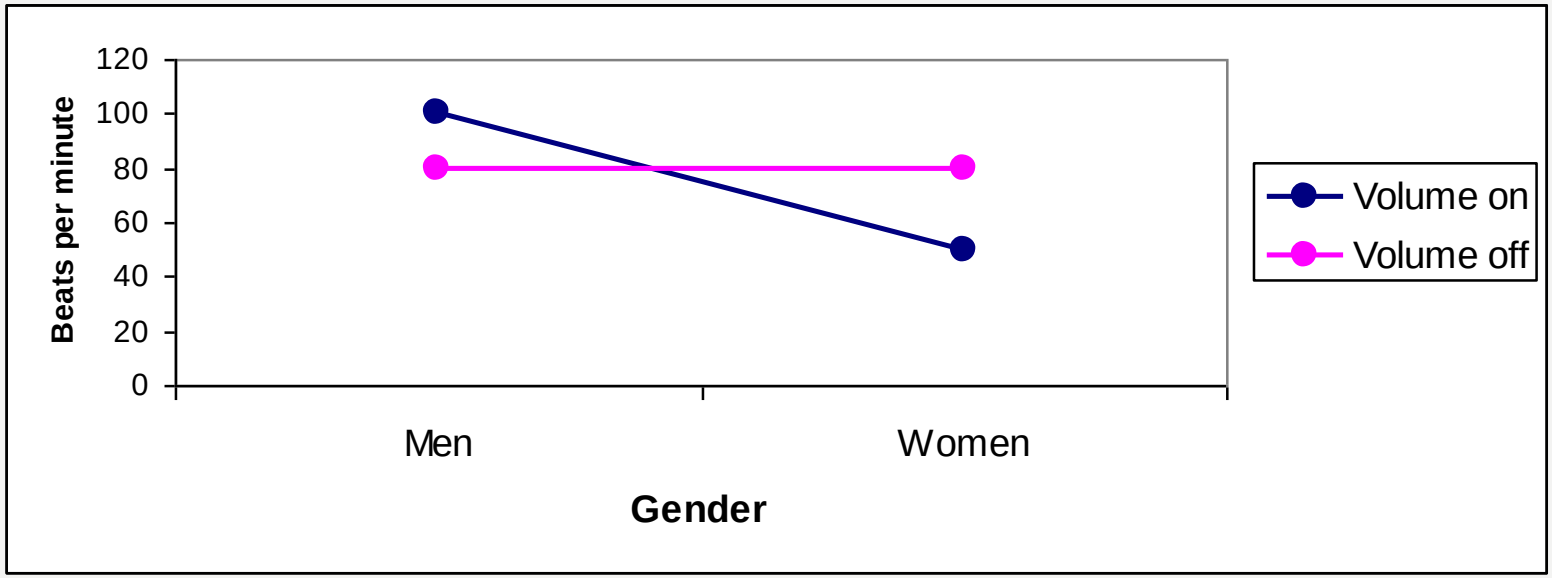
As in a one-way ANOVA, if the main effect is significant and there are only 2 levels of IV, you know where the difference is.

If 3 or more levels of IV, you know that at least one condition is different from the others but do not know which differences are real. You would need to run additional tests to see which differences between 2 conditions are real (e.g., Tukey's HSD if equal n; Fisher's protected t if unequal n)

	Male	**Female**
Very noisy		
Somewhat noisy		
Volume off		

Plot the cell means to see what the data are showing





Factorial designs not only yield info about main effects, but they provide a third – and often critical – piece of information

Interaction:

Main effects do not tell full story; need to consider IV1 in relation to IV2

Do the effects of one IV on the DV depend on the level of the 2nd IV?

Is the pattern of one IV across the levels of the other IV different depending on the level of the other IV (not parallel)?

INTERACTIONS

Example 1:

Crossover interaction

	Male		Female	
Volume on	100	>	50	(75)
Volume off	50	<	100	(75)
	(75)	=	(75)	

The main effects would lead you to conclude that neither variable influenced pulse rates. But that is not true. Look at the patterns shown by the cell means. The pattern seen in the differences in average pulse rates between men and women differ depending on whether the volume is on or off

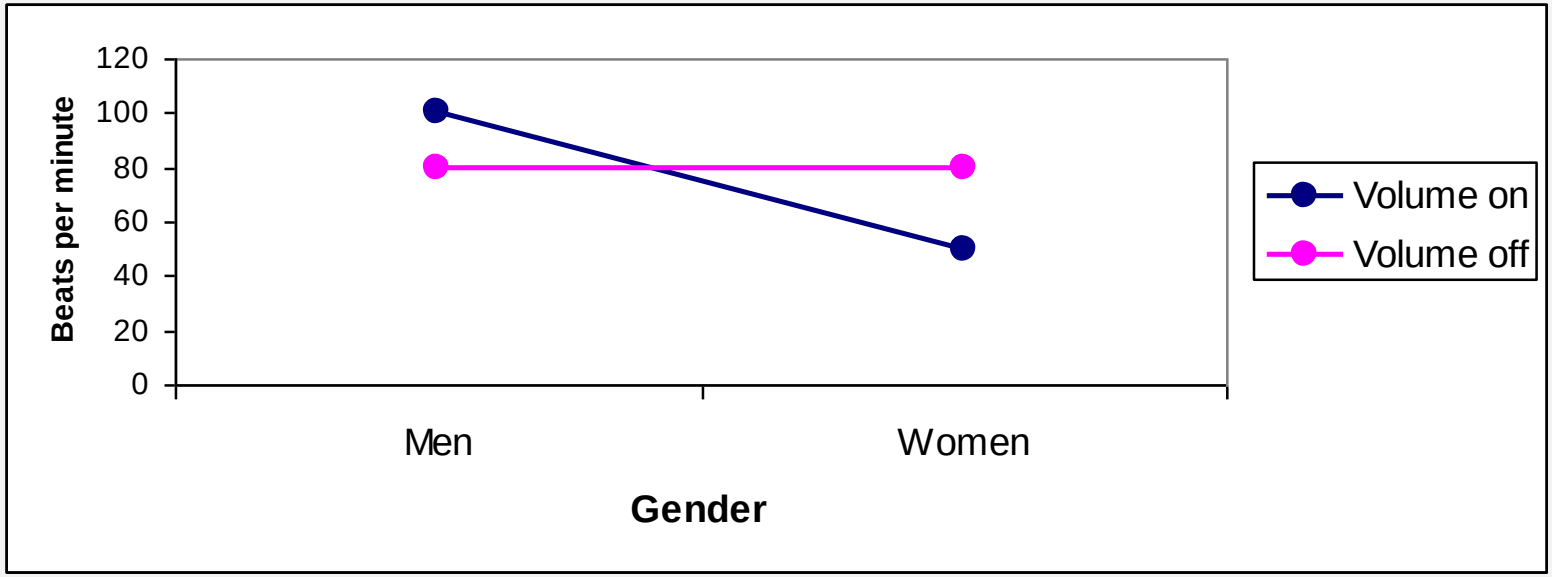
Example 2:

Treatment works for one level but not for other

	Male		Female	
Volume on	100	>	50	(75)
Volume off	80	=	80	(80)
	(90)	>	(65)	

Main effects would lead you to conclude that males always have higher pulse rates on average, but that is only true when the volume is on (not when the volume is off)

Nonparallel lines suggest an interaction is present

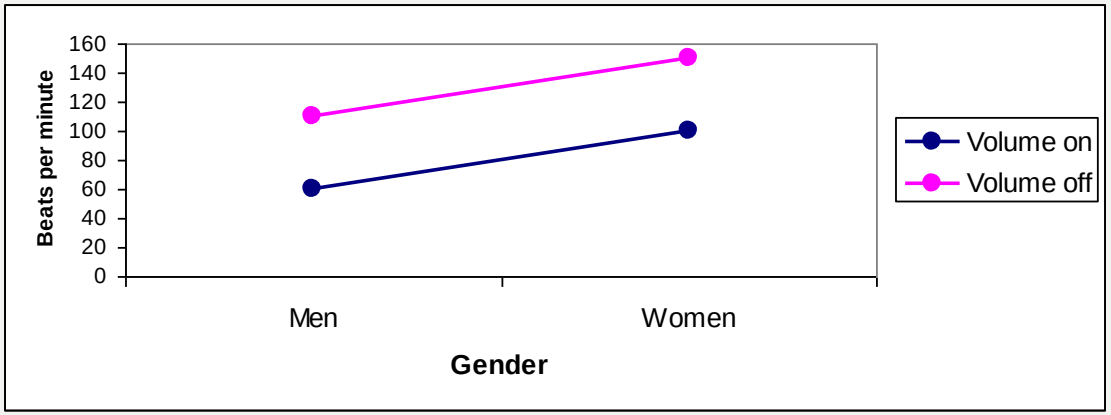


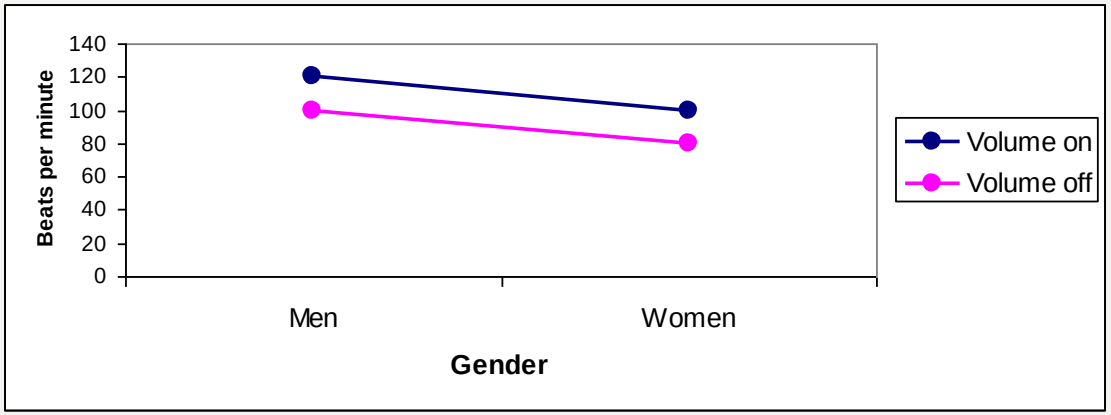
Interactions can be seen easily when line graphs are made

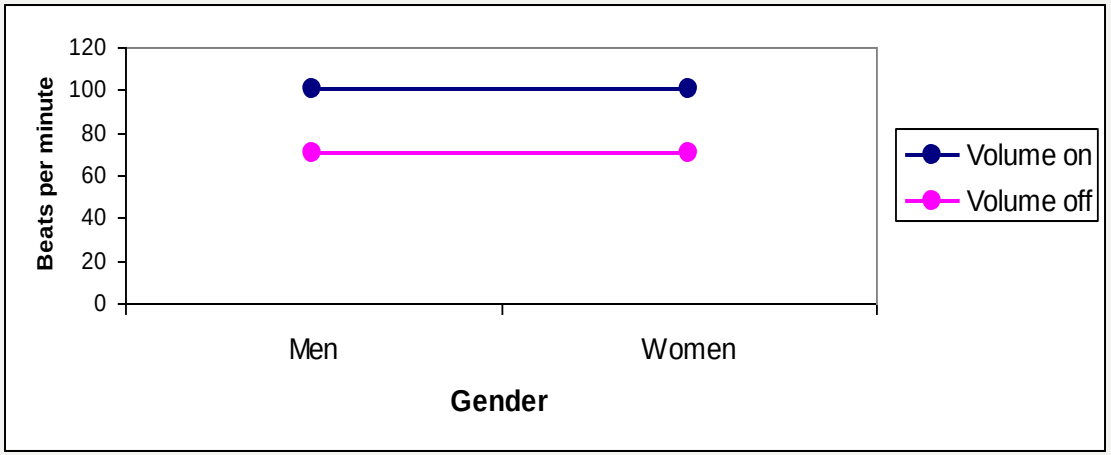
Nonparallel lines = Interaction

Parallel lines = No interaction

Parallel lines suggest there is <u>not</u> an interaction







Example 3: Ordinal Interaction

Sometimes an interaction is significant but it does not change the conclusions you draw from the main effects. The patterns are the same, it is HOW MUCH the difference varies

	Male		Female	
Volume on	60	<	90	(75)
Volume off	50	<	60	(55)
	(55)	<	(75)	

Women always have higher average pulse rates than men but the relative difference is more pronounced when the volume is on than when it is off

Volume on: 60 vs. 90 = 30 point difference

Volume off: 50 vs. 60 = 10 point difference

The volume being on always produced higher average pulse rates than the volume being off, but women are especially affected by this Men: 60 vs. 50 = 10 point difference

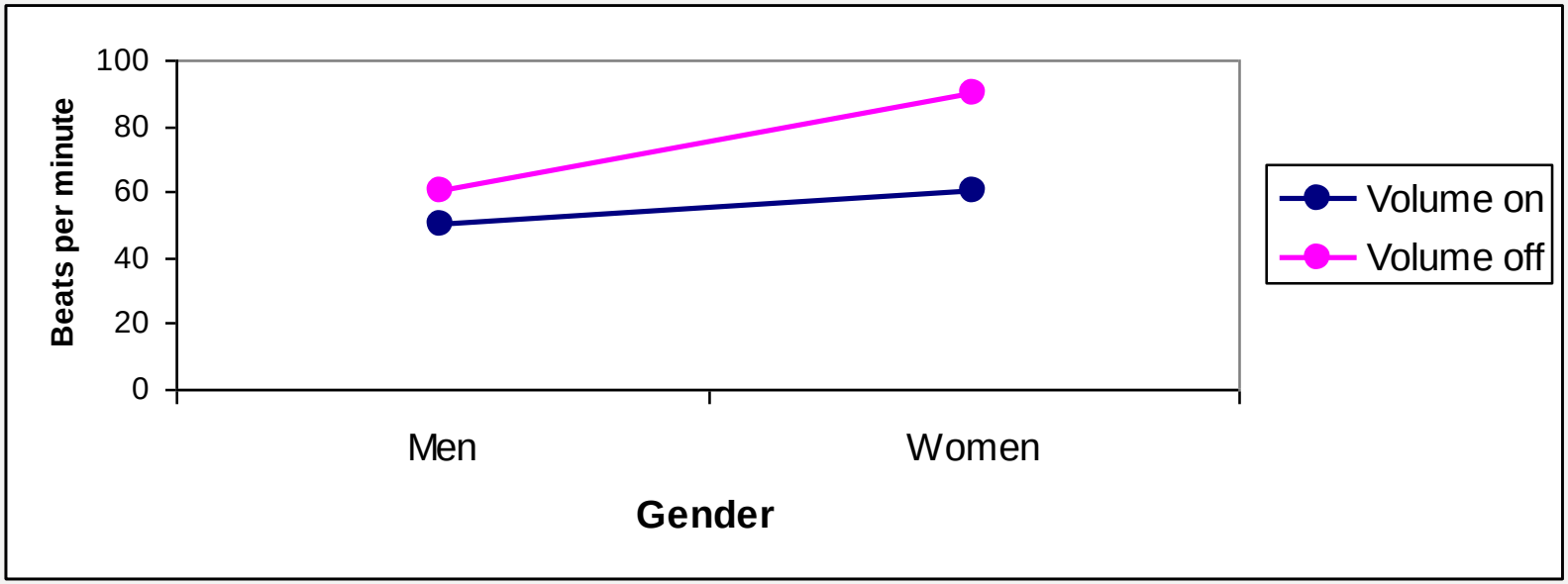
Women: 90 vs. 60 = 30 point difference

SOME REMARKS

Remark 1:

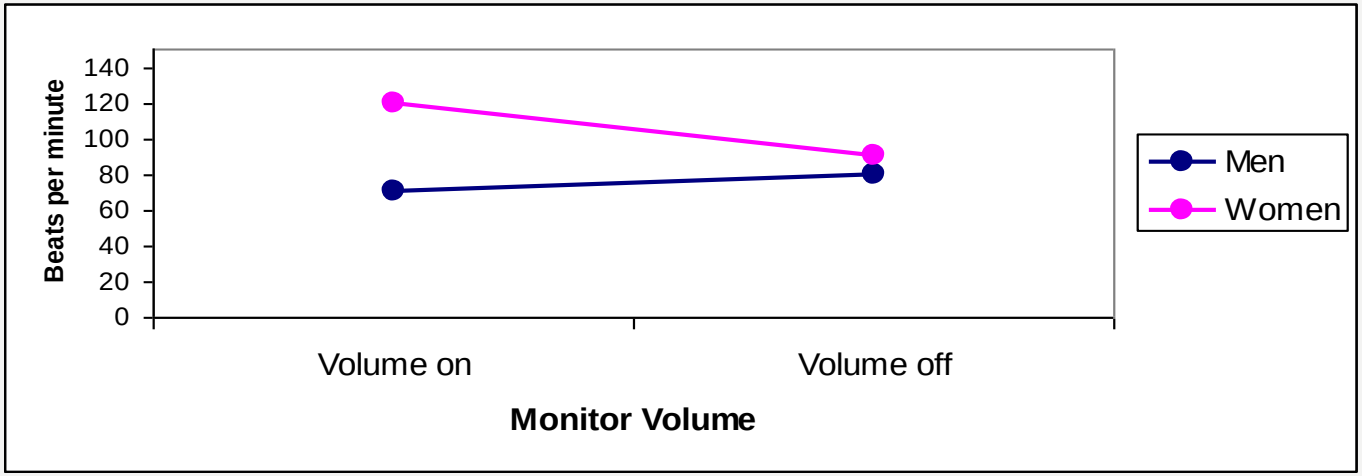
Graphing the data really drive interactions home

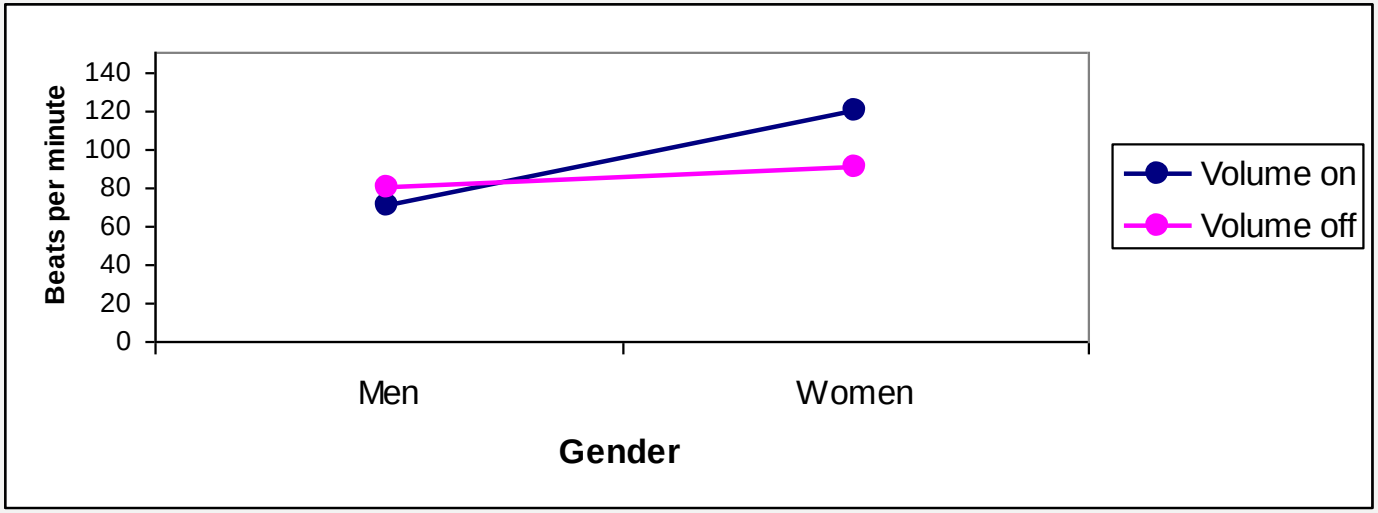
The lines are not parallel but are moving in the same direction (i.e., slopes have same sign but different values); Women always have higher pulses than men but especially so when the volume is off



Remark 2:

Graphs can be made with either variable on the x axis; a good researcher thinks about whether the data tell the story better with IV1 or IV2 on the x axis





Remark 3:

- Remind that when talking about main effects, always use marginal means
- When talking about interactions, always use cell means
- When making graphs, graph the cell means, not the marginal means

HYPOTHESIS TESTING USING ANOVA

Testing for the Main Effect of IV1 and similarly for IV2

NULL HYPOTHESIS H_0:

- The IV1 (IV2) does not influence the DV
- Any differences in average scores between the different conditions of IV1 (IV2) are probably just due to chance (measurement error, random sampling error)
- The (population) means at all levels of the IV1 (IV2) are equal

RESEARCH OR ALTERNATIVE HYPOTHESIS H_A:

- The IV1 does influence the DV
- The differences in average scores between the different conditions are probably not due to chance but show a real effect of the IV1 on the DV
- At least one level of the IV1 has a different (population) mean

Testing for the Interaction of IV1 and IV2

NULL HYPOTHESIS H_0:

- There is not an interaction between IV1 and IV2
- IV1 and IV2 are independent
- The effect of IV1 does not depend on the level of IV2 (and IV2 does not depend on IV1)

RESEARCH OR ALTERNATIVE HYPOTHESIS H_A:

- There is an interaction between IV1 and IV2
- IV1 and IV2 are dependent
- The effect of IV1 does depend on the level of IV2 (and IV2 does depend on IV1)

So, under H_A the (population) cell means cannot be modeled only from the main effects.

HOW TO REPORT TWO-WAY ANOVAS

STEP 1.

Describe the design itself. A two-way ANOVA of [IV1] (level 1, level 2) and [IV2] (level 1, level 2) on [DV] was conducted

[DV] was analyzed in a two-way [between, within, mixed] ANOVA, with [IV1] (level 1, level 2) as a [between subjects; within subjects] variable and [IV2] (level 1, level 2) as a [between subjects; within subjects] variable.

A 2 x 2 factorial [between; within; mixed] ANOVA was conducted on [DV], with [IV1] (level 1, level 2) and [IV2] (level 1, level 2) as the independent variables

For Example:

* A two-way ANOVA of monitor volume (volume on, volume off) and patient's gender (male, female) on pulse rate as measured by beats per minute was conducted.

** The number of beats per minute was analyzed in a two-way mixed factorial ANOVA, with monitor volume (volume on, volume off) manipulated within-subjects and gender (male, female) as a between-subjects variable

** A 2 x 2 between-subjects ANOVA was conducted on pulse rate, with monitor volume and patient's gender as factors

STEP 2:

Report the main effect for IV1:

A significant main effect of [IV1] on [DV] was found, F(dfbet, dferror) = x.xx, p = xxx.

The main effect of [IV1] on [DV] was/was not significant, F(dfbet, dferror) = x.xx, p = xxx.

Step 2a:

If the main effect is significant, then describe it by reporting the marginal means

[DV] was higher / lower for [IV1, Level 1] (M = x.xx) than for [IV1, Level 2] (M = x.xx).

[DV] did not significantly differ between [IV1, Level 1] (M = x.xx) and [IV1, Level 2] (M = x.xx).

FOR EXAMPLE:

A significant main effect of monitor volume on pulse rate was found, F(*, **) = ++, p=* < .05.

Patients' average pulse rate was higher when the volume was on (M = 100.53) than when the volume was off (M = 75.13)

The main effect of monitor volume on pulse rate was not significant, F(*, ***) = +++, p= **>0.05.

This means the average pulse rate did not differ significantly whether the volume was on or off.

So there is no need for an additional sentence, though some people like to say: Thus pulse rates did not differ on average when the volume was on (M = 88.23) or off (M = 84.66)]

STEP 3:

Report the main effect for IV2

STEP 3A:

If the main effect is significant, the describe it by reporting the marginal means

Use same sentence structures as Step 2 and 2a.

FOR EXAMPLE:

A significant main effect of gender on pulse rate was found, F(*, **) = **, p=* < .05. Men's average pulse rate was higher (M = 105.88) than women's (M = 85.31)

The main effect of gender on pulse rate was not significant, F(*, **) = **, p = *.

No need for an additional sentence, though some people like to say: Thus pulse rates did not differ between men (M = 86.25) or women (M = 89.32).

STEP 4:

Report the interaction

The [IV1] x [IV2] interaction was/was not significant, $F(df_{IV1xIV2}, df_{error}) = x.xx$, $p = xxx$

When the interaction is NOT significant (i.e., when $p > .05$), that is all you need to do. This means that the two IVs influence the DV independently from each other.

NOTE:

A significant interaction trumps the main effects!!!

When the interaction is significant, you have to report the cell means and describe their patterns.

You also need to be careful what you say about the main effects. The emphasis should be on the interaction when it is significant.

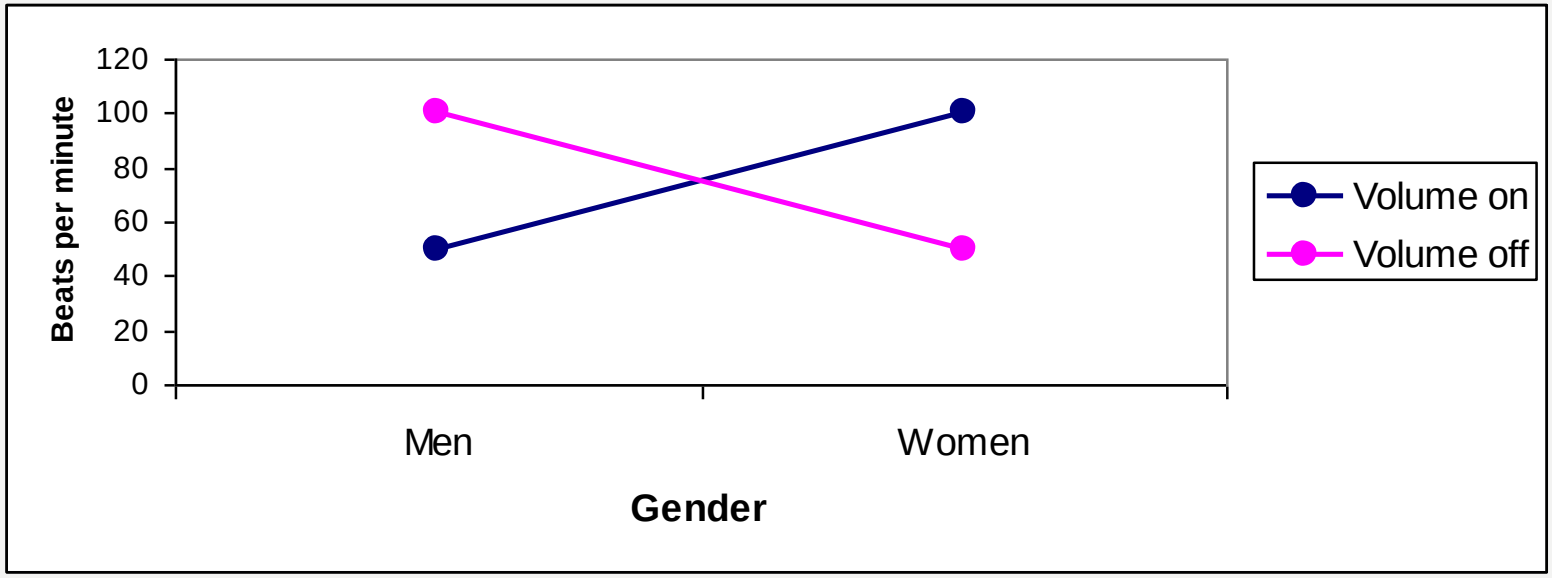
You may examine CELL MEANS and see that the pattern for one IV across the other IV is not the same (i.e., the lines in the graph are not parallel).

When the patterns are different, the statement you make about a main effect is NOT true for all levels of the other IV. So the emphasis in your write up should be about the interaction, not about the main effects.

EXAMPLE 1:

When the volume is on, heart rates are higher for women than for men, but when the volume is off, women have lower heart rates than men.

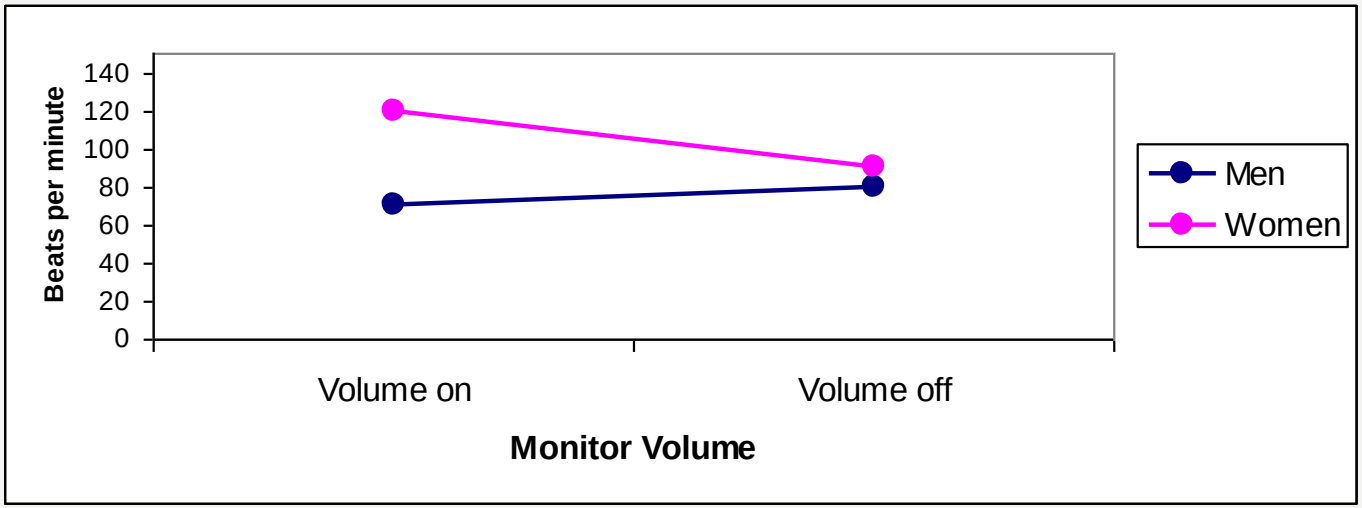
So it is not true that women's heart rates are the same as men's heart rates, as you might have concluded if you only looked at the main effect.



EXAMPLE 2:

Maybe when the volume is on, women's heart rates are higher than men, but when the volume is off, women and men have similar heart rates.

So it is not true that women's heart rates are always higher, as you might have concluded if you only looked at the main effect



In those cases, your write up should emphasize the patterns seen in the interaction, and not the main effects.

SAMPLE WRITE UPS FOR EXAMPLES 1 AND 2

Although the main effect of [IV] on [DV] was significant, F(df, df) = x.xx, p < .xxx, there was a significant interaction between [IV1] and [IV2], F(dfIV1xIV2, dferror) = x.xx, p < .xxx.

The describe the patterns, incorporating CELL MEANS & SDs into the sentence:

EXAMPLE 1:

When the volume of the monitor was on, women’s average heart rate (M=x.xx, SD=x.xx) was higher than men's average heart rate (M=x.xx, SD=x.xx).

However, when the volume was off, women’s heart rates (M=x.xx, SD=x.xx) were lower than men's heart rates (M=x.xx, SD=x.xx), on average.

EXAMPLE 2:

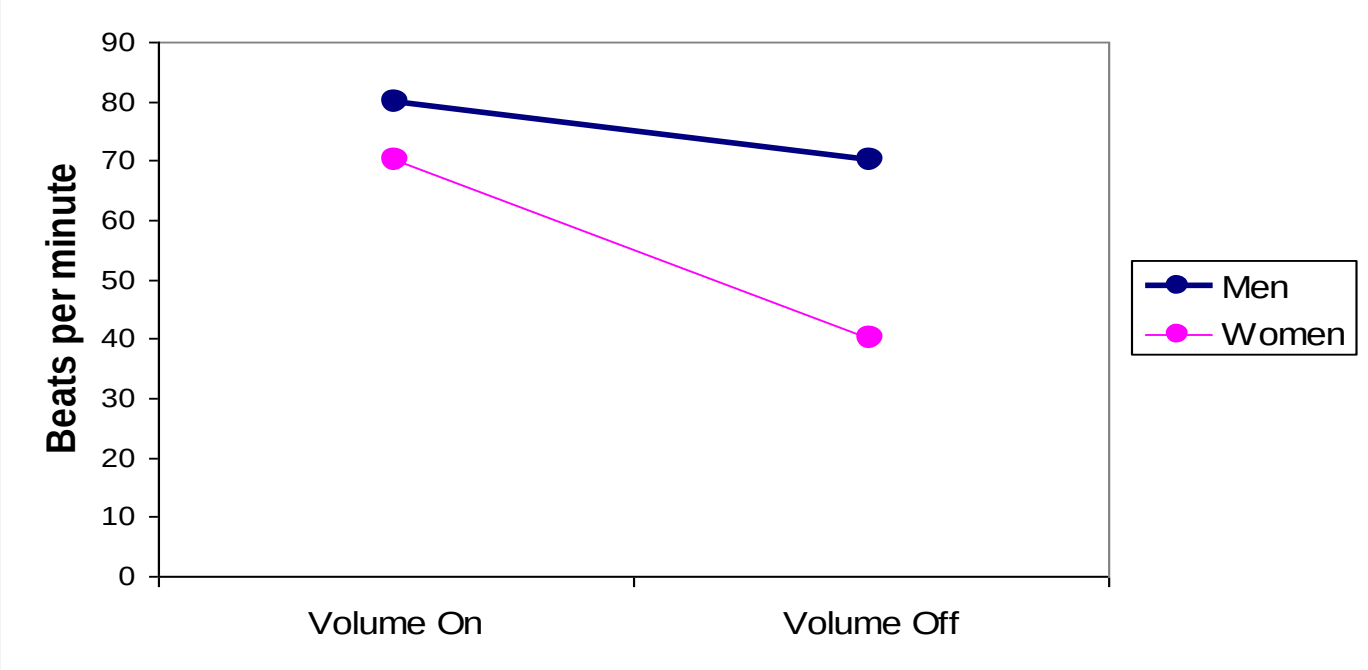
When the volume of the monitor was on, women’s average heart rate (M=x.xx, SD=x.xx) was higher than men's average heart rate (M=x.xx, SD=x.xx).

However, when the volume was off, women’s heart rates (M=x.xx, SD=x.xx) did not differ from men's heart rates (M=x.xx, SD=x.xx), on average.

EXAMPLE 3:

What happens if you examine the cell means and you observe that the pattern for one IV is the same across the other IV –

e.g., when the volume is on, men's heart rate is higher than women's; likewise, when the volume is off, men's heart rate is higher than women's



If the interaction was significant but is saying the same thing as the main effect does, then it indicates that the pattern is more pronounced at one level of an IV than at the other level.

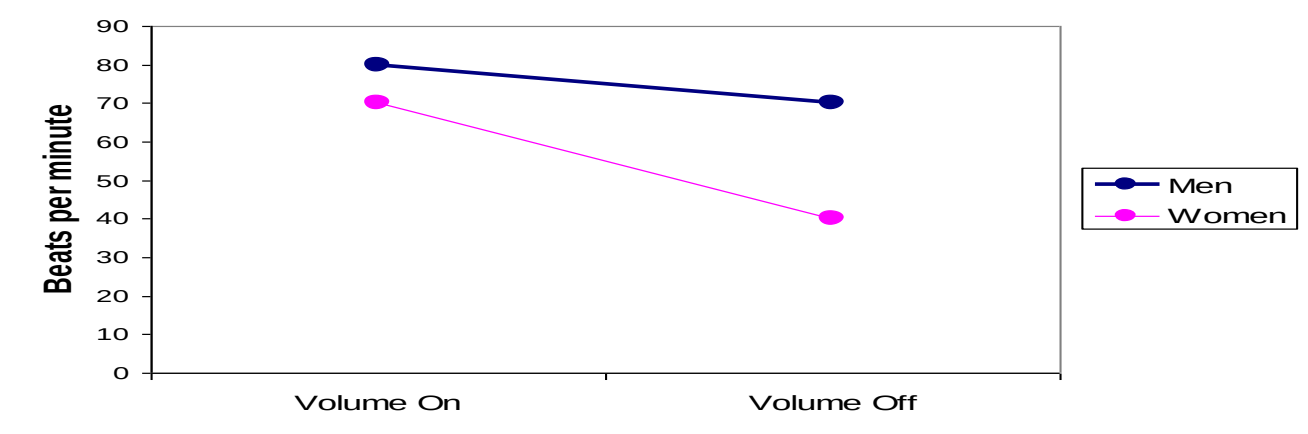
That is, the main effects tell the story (men have higher heart rates than women in the ER, and this is true whether the monitor's volume is on or off).

What is driving the significant interaction can be seen if you drew this as a graph – the lines are moving in the same direction but one slope is steeper than the other. HOW much higher men's heart rates are relative to woman's differs depending on whether the volume is on or off

Sample Write Up for Example 3

The [fill in name of IV1] and [fill in name of IV2] interaction was significant, F(df, df) = x.xx, p < .xxx, though it was consistent with the patterns seen in the main effects.

When the volume was off, average heart rates were higher for men (M=x.xx, SD=x.xx) than for women (M=x.xx, SD=x.xx). Likewise, when the volume was on, heart rates were higher for men (M=x.xx, SD=x.xx) than for women (M=x.xx, SD=x.xx), on average, though the relative difference was more pronounced with the volume off.



EFFECT SIZE IN TWO-WAY ANOVAS

There are 3 different effect sizes that can be calculated:

- Effect size for IV1
- Effect size for IV2
- Effect size for IV1xIV2 interaction

All three are measured by eta squared:

$$\eta^2 = SS_{factor}/SS_{total}$$

Note: SPSS will calculate this for you

Small	0 - .20
Medium	.21 - .40
Large	> .40

SPSS

EXAMPLE

A researcher is interested in whether the amount of time a person spent on watching TV each day was influenced by two factors, their gender and home ownership.

(This will be a 2 X 2 design with two levels of gender (male and female), and two levels of home ownership status (owns own home and doesn't own home).

As a result, he is going to test the hypotheses that:

- Sex of respondent has a significant main effect on hours per day spent watching TV
- Home ownership has a significant main effect on hours per day spent watching TV
- Sex of respondent & home ownership have a significant interaction effect on hours per day spent watching TV

- First, open ANOVA_2.sav data file

*ANOVA_2.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	4	0	Respondent ID Nu...	None	None	8	Center	Scale	Input
2	sex	Numeric	1	0	Respondent's Sex	{1, Male}...	None	8	Center	Nominal	Input
3	homeown	Numeric	8	2	OwnsOwnHome	{1.00, Owns Own ...	None	8	Center	Scale	Input
4	tvhours	Numeric	2	0	Hours Per Day Wa...	{-1, NAP}...	-1, 98, 99	8	Right	Ordinal	Input
5	news	Numeric	1	0	How Often Does R ...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
6	wrkstat	Numeric	1	0	Labor Force Status	{0, NAP}...	0, 9	8	Right	Ordinal	Input
7	marital	Numeric	1	0	Marital Status	{1, married}...	9	8	Right	Ordinal	Input
8	agewed	Numeric	2	0	Age When First M...	{0, nap}...	0, 98, 99	8	Right	Ordinal	Input
9	sibs	Numeric	2	0	Number of Brother...	{98, dk}...	98, 99	8	Right	Ordinal	Input
10	childs	Numeric	1	0	Number of Children	{8, Eight or More}...	9	8	Right	Ordinal	Input
11	age	Numeric	2	0	Age of Respondent	{98, DK}...	0, 98, 99	8	Right	Ordinal	Input
12	birthmo	Numeric	2	0	Month in Which R ...	{0, NAP}...	0, 98, 99	8	Right	Ordinal	Input
13	zodiac	Numeric	2	0	Respondents Astro...	{0, NAP}...	0, 98, 99	8	Right	Ordinal	Input
14	educ	Numeric	2	0	Highest Year of Sc...	{97, NAP}...	97, 98, 99	8	Right	Ordinal	Input
15	degree	Numeric	1	0	RS Highest Degree	{0, Less than HS}...	7, 8, 9	8	Right	Ordinal	Input
16	padeg	Numeric	1	0	Father's Highest D...	{0, LT High School...	7, 8, 9	8	Right	Ordinal	Input
17	madeg	Numeric	1	0	Mother's Highest D...	{0, LT High School...	7, 8, 9	8	Right	Ordinal	Input
18	race	Numeric	1	0	Racew of Respond...	{1, white}...	None	8	Right	Ordinal	Input
19	income91	Numeric	2	0	Total Family Income	{0, NAP}...	0, 98, 99	8	Right	Ordinal	Input
20	rincom91	Numeric	2	0	Respondent's Inco...	{0, NAP}...	0, 98, 99	8	Right	Ordinal	Input
21	region	Numeric	1	0	Region of Interview	{0, Not Assigned}...	0	8	Right	Ordinal	Input
22	xnorcsiz	Numeric	2	0	Expanded N.O.R.C...	{0, Not Assigned}...	0	8	Right	Ordinal	Input
23	size	Numeric	4	0	Size of Place in 10...	{-1, Not Assigned}...	-1	8	Right	Ordinal	Input
24	partyid	Numeric	1	0	Political Party Affli...	{0, Strong Democr...	8, 9	8	Right	Ordinal	Input
25	vote92	Numeric	1	0	Voting in 1992 Ele...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
26	polviews	Numeric	1	0	Think of Self as Lib...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
27	cappun	Numeric	1	0	Favor or Oppose D...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
28	gunlaw	Numeric	1	0	Favor or Oppose G...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
29	grass	Numeric	1	0	Should Marijuana ...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
30	relig	Numeric	1	0	Religious Preference	{1, Protestant}...	8, 9	8	Right	Ordinal	Input
31	life	Numeric	1	0	Is Life Exciting or ...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
32	chldidel	Numeric	1	0	Ideal Number of Ch...	{-1, NAP}...	-1, 9	8	Right	Ordinal	Input
33	pillok	Numeric	1	0	Birth Control to Te...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
34	sexeduc	Numeric	1	0	Sex Education in ...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
35	spanking	Numeric	1	0	Favor Spanking to ...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
36	letdie1	Numeric	1	0	Allow Incurable Pat...	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
37	bigband	Numeric	1	0	Bigband Music	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
38	blugrass	Numeric	1	0	Bluegrass Music	{0, NAP}...	0, 8, 9	8	Right	Ordinal	Input
39	country	Numeric	1	0	Country Western	{0, NAP}	0, 8, 9	8	Right	Ordinal	Input

Data View | Variable View

*ANOVA_2.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 68 of 68 Variables

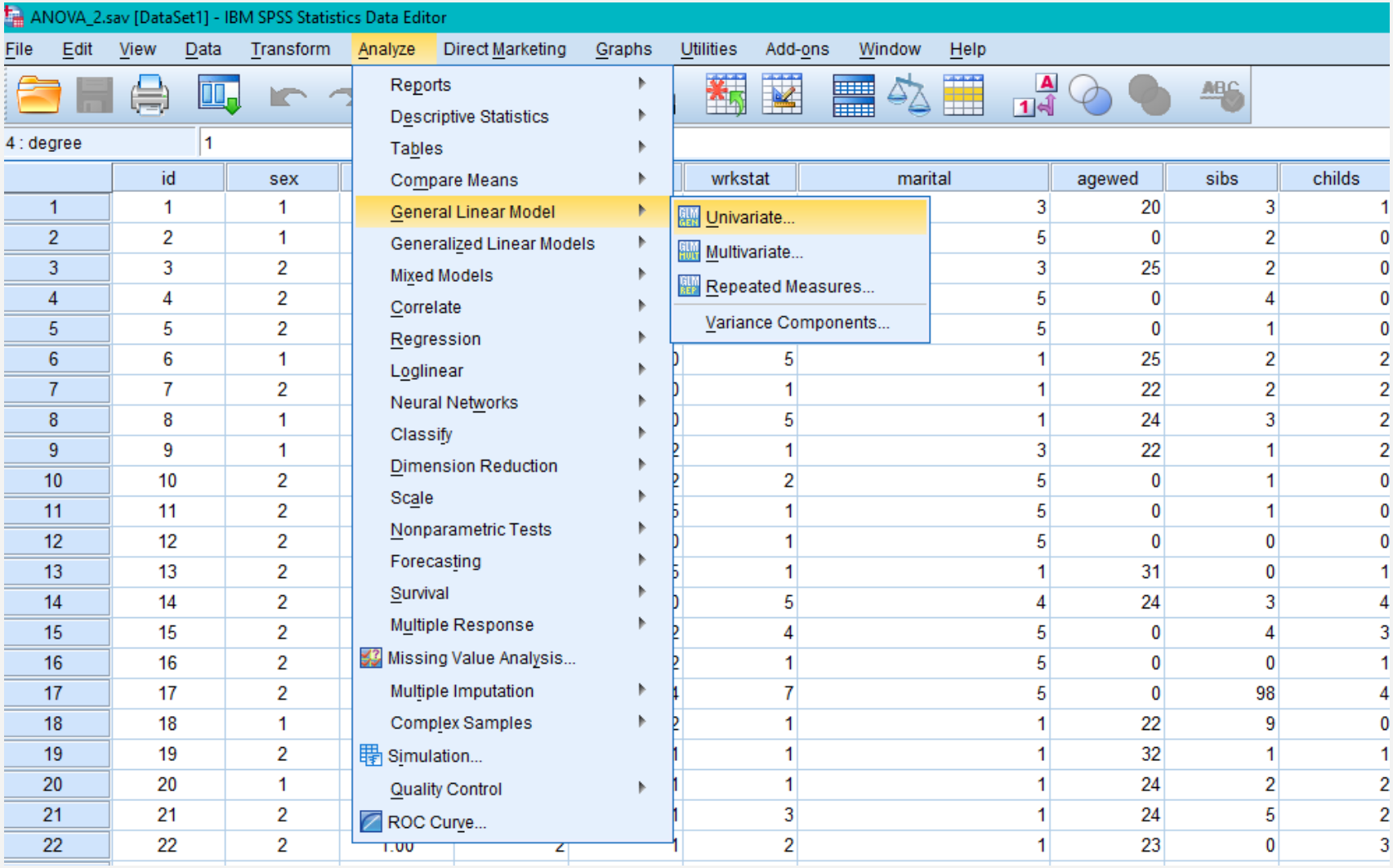
	id	sex	homeown	tvhours	news	wrkstat	marital	agewed	sibs	childs	age	birthmo	zodiac	educ	degree	padeg	madeg	race	income91	rincom91	region	xnorcsiz
1	1	Male	.	4	NAP	Working ful...	divorced	20	3	1	43	May	Taurus	11	High school	NAP	LT High Sc...	white	$40000-49...	$35000-39...	Middle Atla...	City, GT 2...
2	2	Male	Doesn't O...	5	Everyday	Working ful...	never married	nap	2	0	44	August	Virgo	16	Bachelor	LT High Sc...	LT High Sc...	black	$40000-49...	$40000-49...	Middle Atla...	City, GT 2...
3	3	Female	.	2	NAP	Working ful...	divorced	25	2	0	43	February	Aquarius	16	Bachelor	High School	LT High Sc...	white	$40000-49...	$40000-49...	Middle Atla...	City, GT 2...
4	4	Female	Doesn't O...	4	Everyday	Working p...	never married	nap	4	0	45	NA	NA	15	High school	High School	DK	white	Refused	Refused	Middle Atla...	City, GT 2...
5	5	Female	Doesn't O...	1	Everyday	Retired	never married	nap	1	0	78	October	Libra	17	Graduate	High School	Junior Coll...	white	$75000+	NAP	Middle Atla...	City, GT 2...
6	6	Male	.	4	NAP	Retired	married	25	2	2	83	March	Pisces	11	High school	LT High Sc...	LT High Sc...	white	$20000-22...	NAP	Middle Atla...	City, GT 2...
7	7	Female	.	3	NAP	Working ful...	married	22	2	2	55	October	Libra	12	High school	LT High Sc...	LT High Sc...	white	$50000-59...	$25000-29...	Middle Atla...	City, GT 2...
8	8	Male	.	3	NAP	Retired	married	24	3	2	75	November	Sagittarius	12	High school	LT High Sc...	LT High Sc...	white	$12500-14...	NAP	Middle Atla...	City, GT 2...
9	9	Male	Doesn't O...	1	Few Times...	Working ful...	divorced	22	1	2	31	July	Cancer	18	Graduate	High School	High School	white	$75000+	$60000-74...	Middle Atla...	City, GT 2...
10	10	Female	Doesn't O...	1	Few Times...	Working p...	never married	nap	1	0	54	March	Pisces	18	Graduate	LT High Sc...	High School	white	$10000-12...	$1000-2999	Middle Atla...	City, GT 2...
11	11	Female	Doesn't O...	2	Never	Working ful...	never married	nap	1	0	29	April	Taurus	18	Graduate	LT High Sc...	LT High Sc...	white	$30000-34...	$30000-34...	Middle Atla...	City, GT 2...
12	12	Female	.	4	NAP	Working ful...	never married	nap	0	0	23	October	Scorpio	15	High school	NAP	Junior Coll...	white	$40000-49...	$3000-3999	Middle Atla...	City, GT 2...
13	13	Female	Doesn't O...	3	Never	Working ful...	married	31	0	1	61	NA	NA	12	High school	High School	High School	white	$22500-24...	$8000-9999	Middle Atla...	City, GT 2...
14	14	Female	.	6	NAP	Retired	separated	24	3	4	63	March	Aries	4	Less than ...	NAP	LT High Sc...	other	$5000-5999	NAP	Middle Atla...	City, GT 2...
15	15	Female	Doesn't O...	4	Few Times...	Unempl, lai...	never married	nap	4	3	33	March	Pisces	10	Less than ...	DK	DK	other	$10000-12...	$10000-12...	Middle Atla...	City, GT 2...
16	16	Female	Doesn't O...	2	Few Times...	Working ful...	never married	nap	0	1	36	November	Scorpio	14	High school	NAP	High School	black	$12500-14...	$12500-14...	Middle Atla...	City, GT 2...
17	17	Female	Doesn't O...	5	Less Than ...	Keeping ho...	never married	nap	dk	4	39	March	Pisces	8	Less than ...	NA	LT High Sc...	black	$5000-5999	NAP	Middle Atla...	City, GT 2...
18	18	Male	Doesn't O...	2	Few Times...	Working ful...	married	22	9	0	55	January	Capricorn	15	High school	High School	LT High Sc...	other	Refused	Refused	Middle Atla...	City, GT 2...
19	19	Female	Owns Own...	2	Everyday	Working ful...	married	32	1	1	55	September	Libra	16	Bachelor	High School	LT High Sc...	white	$75000+	$22500-24...	Middle Atla...	Unlnc, Lrg ...
20	20	Male	Owns Own...	4	Everyday	Working ful...	married	24	2	2	34	April	Taurus	16	Bachelor	LT High Sc...	High School	white	$75000+	$75000+	Middle Atla...	Unlnc, Lrg ...
21	21	Female	Owns Own...	1	Everyday	Temp not ...	married	24	5	2	36	June	Gemini	14	Junior colle...	Bachelor	Bachelor	white	$60000-74...	$25000-29...	Middle Atla...	Unlnc, Lrg ...
22	22	Female	Owns Own...	2	Everyday	Working p...	married	23	0	3	44	August	Leo	18	Bachelor	High School	High School	white	$75000+	$40000-49...	Middle Atla...	Unlnc, Lrg ...
23	23	Male	Owns Own...	3	Everyday	Retired	widowed	25	2	2	80	May	Taurus	18	Graduate	NAP	LT High Sc...	white	$40000-49...	NAP	Middle Atla...	Unlnc, Lrg ...
24	24	Male	Doesn't O...	3	Less Than ...	Working ful...	never married	nap	5	1	32	July	Leo	16	Bachelor	Bachelor	Bachelor	other	$25000-29...	$15000-17...	Pacific	City, GT 2...
25	25	Male	Doesn't O...	2	Once a W...	Working p...	married	32	7	4	37	September	Virgo	5	Less than ...	LT High Sc...	LT High Sc...	other	$20000-22...	$4000-4999	Pacific	City, GT 2...
26	27	Female	Doesn't O...	3	Once a W...	Working ful...	married	40	5	0	49	September	Virgo	16	Bachelor	Bachelor	Bachelor	other	$30000-34...	$25000-29...	Pacific	City, GT 2...
27	28	Female	.	2	NAP	Working ful...	divorced	16	6	2	45	November	Sagittarius	16	Bachelor	High School	NAP	other	$60000-74...	$60000-74...	Pacific	City, GT 2...
28	29	Female	.	2	NAP	Working ful...	never married	nap	2	0	36	April	Taurus	18	Bachelor	High School	Graduate	white	$30000-34...	$10000-12...	Pacific	City, GT 2...
29	30	Female	Doesn't O...	1	Less Than ...	Working p...	never married	nap	5	0	23	March	Aries	16	High school	NAP	Bachelor	black	$10000-12...	$5000-5999	Pacific	City, GT 2...
30	31	Male	.	3	NAP	Working ful...	never married	nap	1	0	29	May	Taurus	16	Graduate	Bachelor	Bachelor	white	$50000-59...	$50000-59...	Pacific	City, GT 2...
31	32	Female	Doesn't O...	2	Few Times...	Working ful...	never married	nap	3	1	30	February	Aquarius	14	Junior colle...	High School	LT High Sc...	black	$30000-34...	$30000-34...	Pacific	City, GT 2...
32	33	Female	.	2	NAP	Working ful...	never married	nap	8	1	40	July	Cancer	16	Bachelor	LT High Sc...	High School	black	$30000-34...	$30000-34...	Pacific	City, GT 2...
33	34	Female	Doesn't O...	1	Less Than ...	Working ful...	never married	nap	8	2	26	August	Virgo	11	Less than ...	High School	High School	white	$30000-34...	$15000-17...	Pacific	Suburb, Lr...
34	35	Male	Owns Own...	3	Everyday	Retired	married	21	3	4	66	February	Pisces	12	High school	LT High Sc...	LT High Sc...	white	$75000+	NAP	Pacific	Suburb, Lr...
35	36	Female	Owns Own...	2	Few Times...	Unempl, lai...	married	22	11	1	53	October	Scorpio	14	High school	High School	High School	other	$40000-49...	$12500-14...	Pacific	Suburb, Lr...
36	37	Female	Owns Own...	3	Few Times...	Working p...	married	19	2	0	20	December	Sagittarius	12	High school	LT High Sc...	LT High Sc...	other	$8000-9999	$1000-2999	Pacific	Suburb, Lr...
37	38	Female	.	4	NAP	Working ful...	divorced	24	4	2	63	July	Cancer	12	High school	LT High Sc...	LT High Sc...	white	$10000-12...	$10000-12...	Pacific	Suburb, Lr...

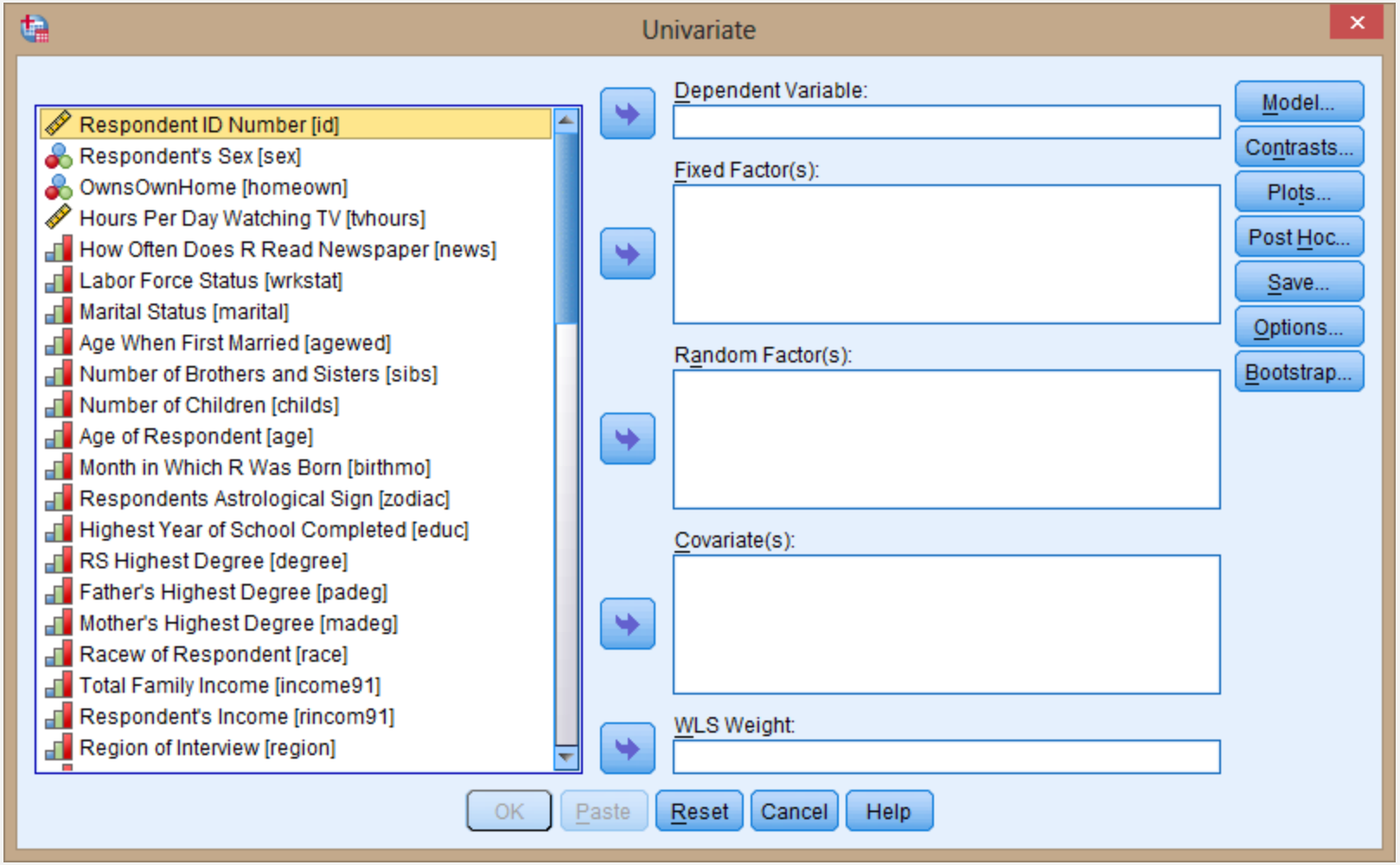
Data View | Variable View

IBM SPSS Statistics Processor is ready | Unicode:ON

GO TO:

Analyze ➔ General Linear Model ➔ Univariate



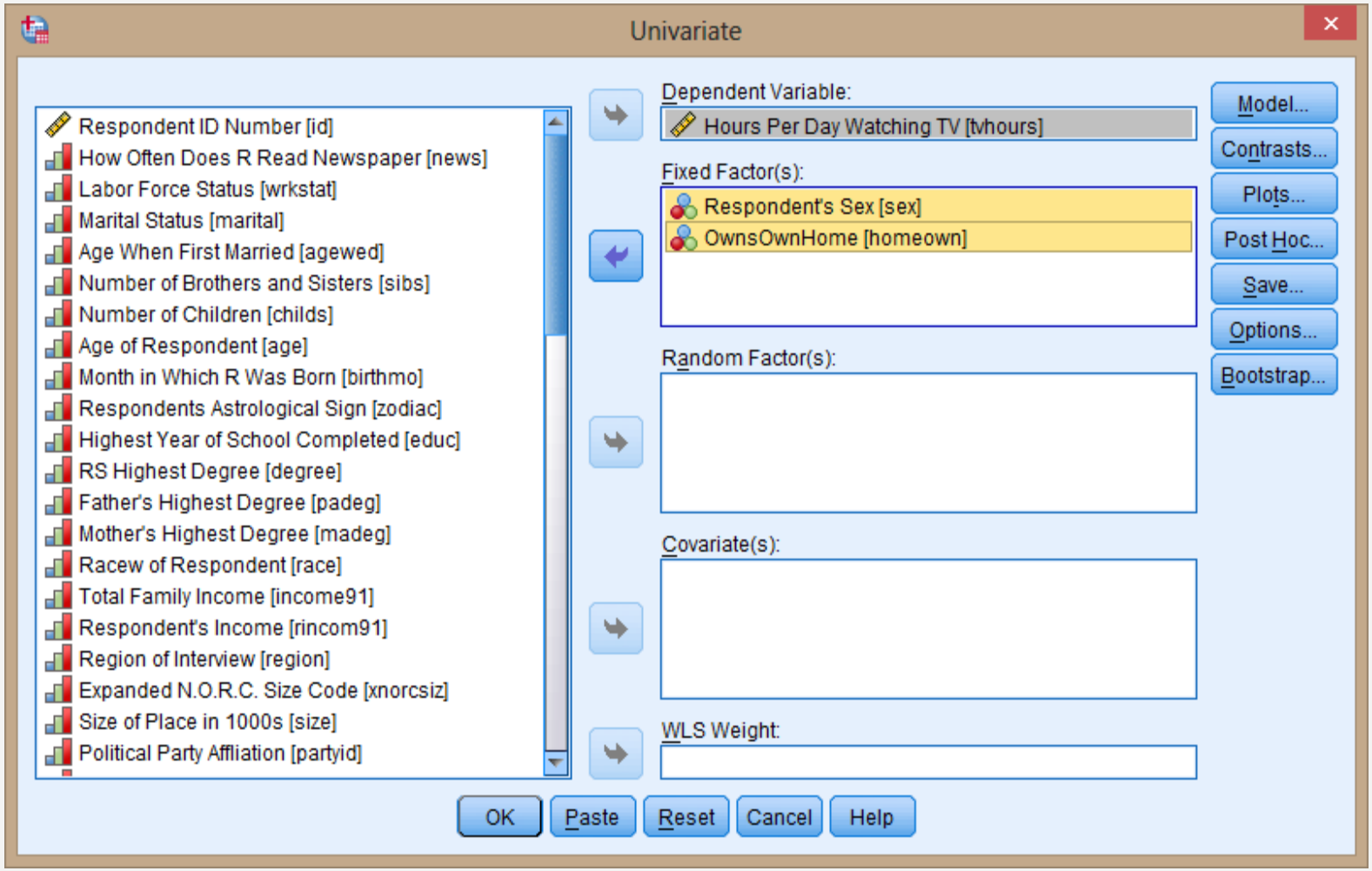

Univariate
Respondent ID Number [id]
Respondent's Sex [sex]
OwnsOwnHome [homeown]
Hours Per Day Watching TV [tvhours]
How Often Does R Read Newspaper [news]
Labor Force Status [wrkstat]
Marital Status [marital]
Age When First Married [agewed]
Number of Brothers and Sisters [sibs]
Number of Children [childs]
Age of Respondent [age]
Month in Which R Was Born [birthmo]
Respondents Astrological Sign [zodiac]
Highest Year of School Completed [educ]
RS Highest Degree [degree]
Father's Highest Degree [padeg]
Mother's Highest Degree [madeg]
Racew of Respondent [race]
Total Family Income [income91]
Respondent's Income [rincom91]
Region of Interview [region]
Dependent Variable:
Fixed Factor(s):
Random Factor(s):
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
Bootstrap...
OK
Paste
Reset
Cancel
Help

Move the variables:

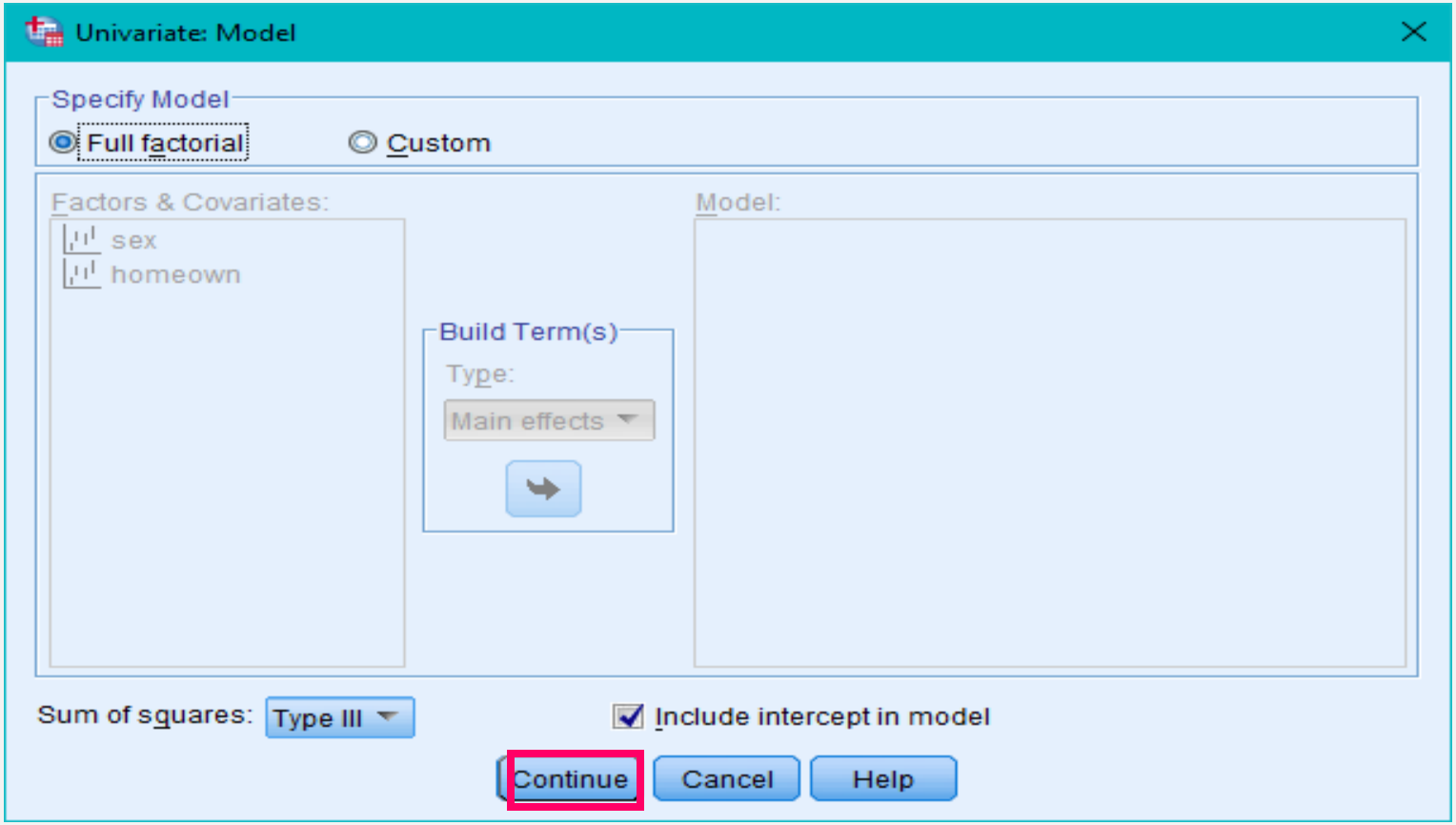
- Respondent's Sex & OwnsOwnHome into the Fixed Factor window

Move the variable:

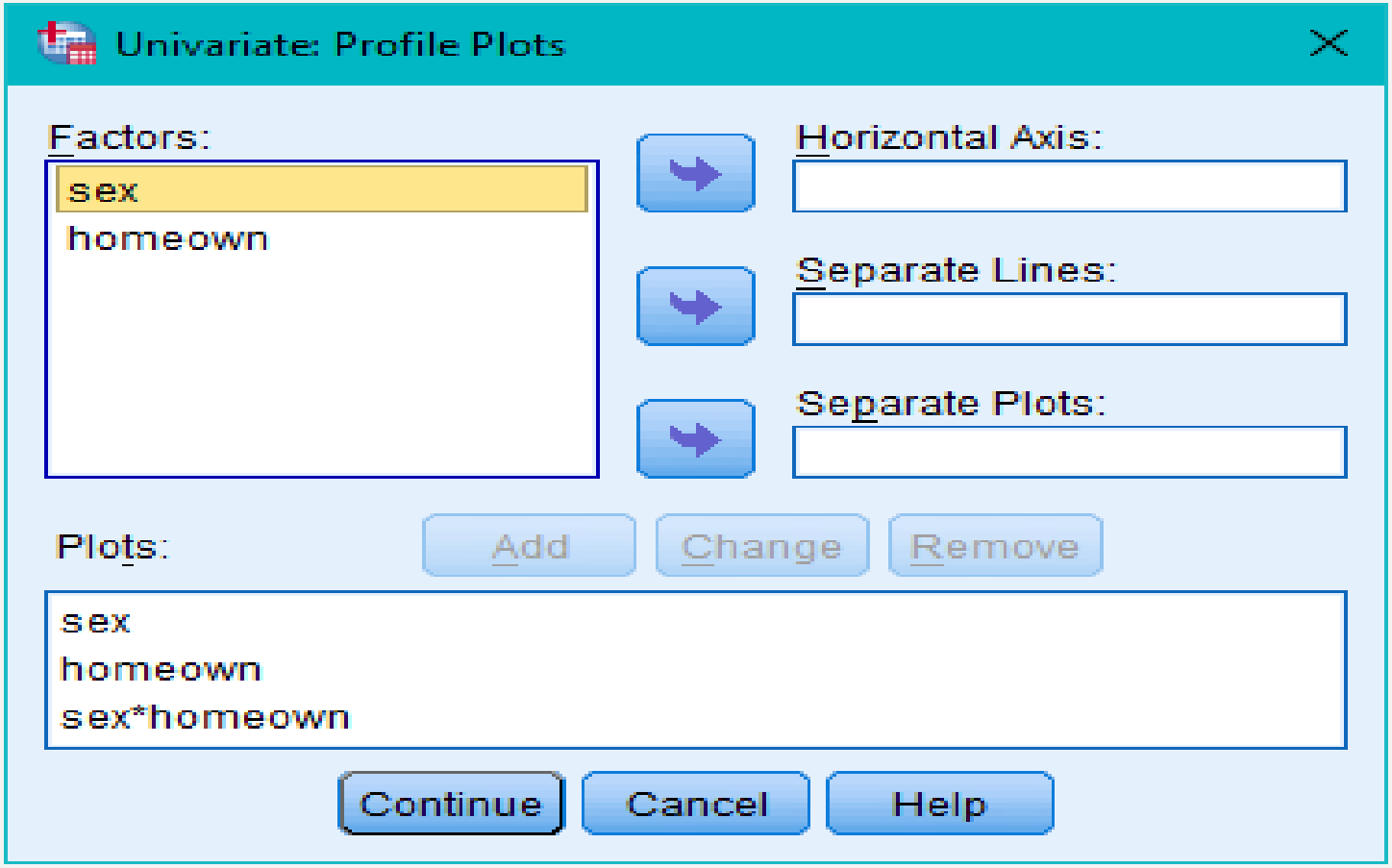
- Hours per Day Watching TV variable into the Dependent Variables window


Univariate
Respondent ID Number [id]
How Often Does R Read Newspaper [news]
Labor Force Status [wrkstat]
Marital Status [marital]
Age When First Married [agewed]
Number of Brothers and Sisters [sibs]
Number of Children [childs]
Age of Respondent [age]
Month in Which R Was Born [birthmo]
Respondents Astrological Sign [zodiac]
Highest Year of School Completed [educ]
RS Highest Degree [degree]
Father's Highest Degree [padeg]
Mother's Highest Degree [madeg]
Racew of Respondent [race]
Total Family Income [income91]
Respondent's Income [rincom91]
Region of Interview [region]
Expanded N.O.R.C. Size Code [xnorcsiz]
Size of Place in 1000s [size]
Political Party Affiliation [partyid]
Dependent Variable:
Hours Per Day Watching TV [tvhours]
Fixed Factor(s):
Respondent's Sex [sex]
OwnsOwnHome [homeown]
Random Factor(s):
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
Bootstrap...
OK
Paste
Reset
Cancel
Help

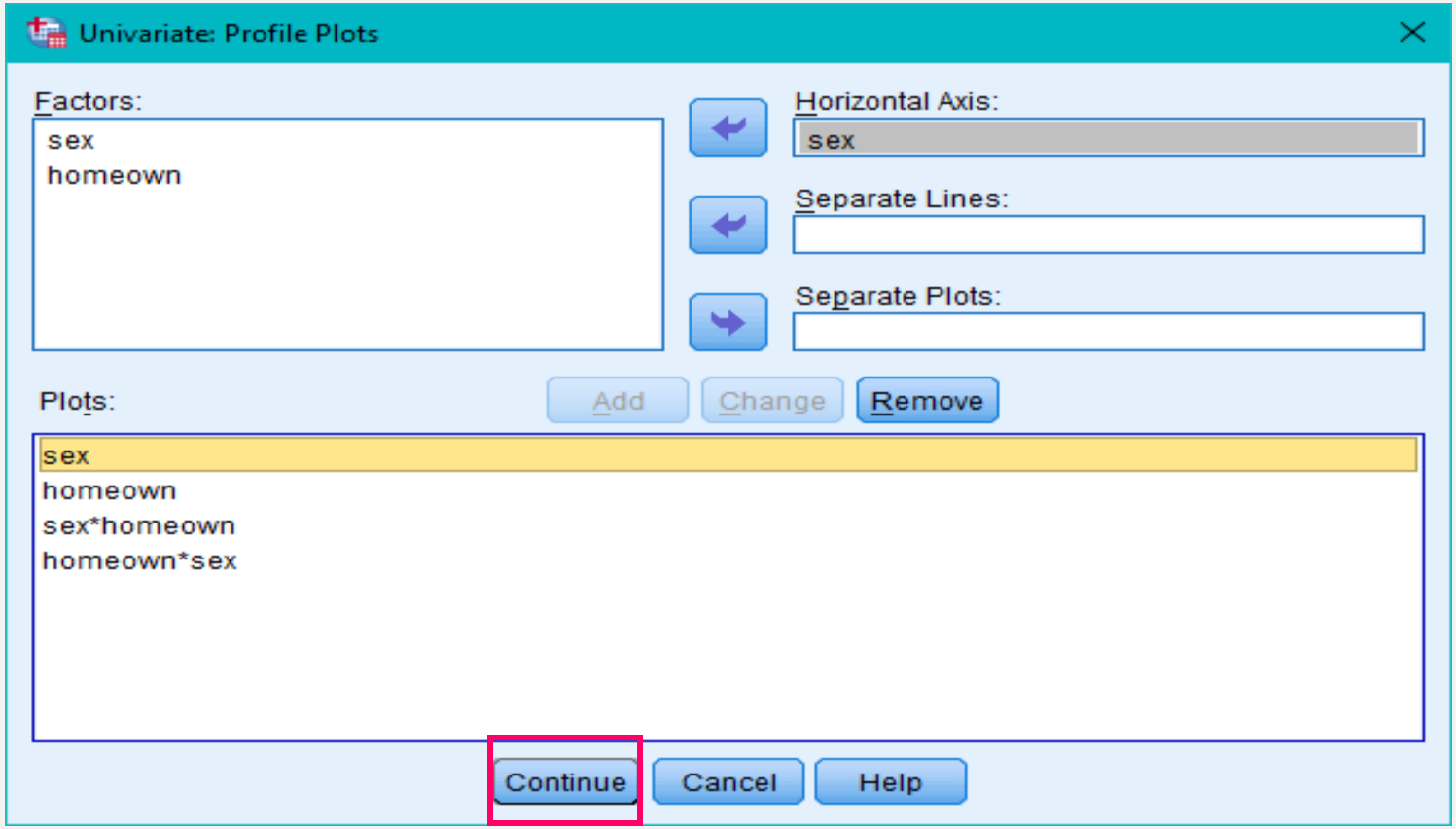
- Click on **Model** ➔ select **Full Factorial** ➔ and **Continue**
- Ignore the Contrasts Button for now



Click on the **Plots** button to select the plots we want.



- **First we get plots for the main effects:**
  - Move the Sex factor into the Horizontal Axis window ➔ click the Add button
  - Move the Homeown factor into the Horizontal Axis window ➔ click the Add button.
- **Next we will get plots for the interaction effect:**
  - Move the Sex factor into the Horizontal Axis window & the Homeown factor into the Separate Lines window ➔ click the Add button
  - Move the Homeown factor into the Horizontal Axis window and the Sex factor into the Separate Lines window ➔ click the Add button

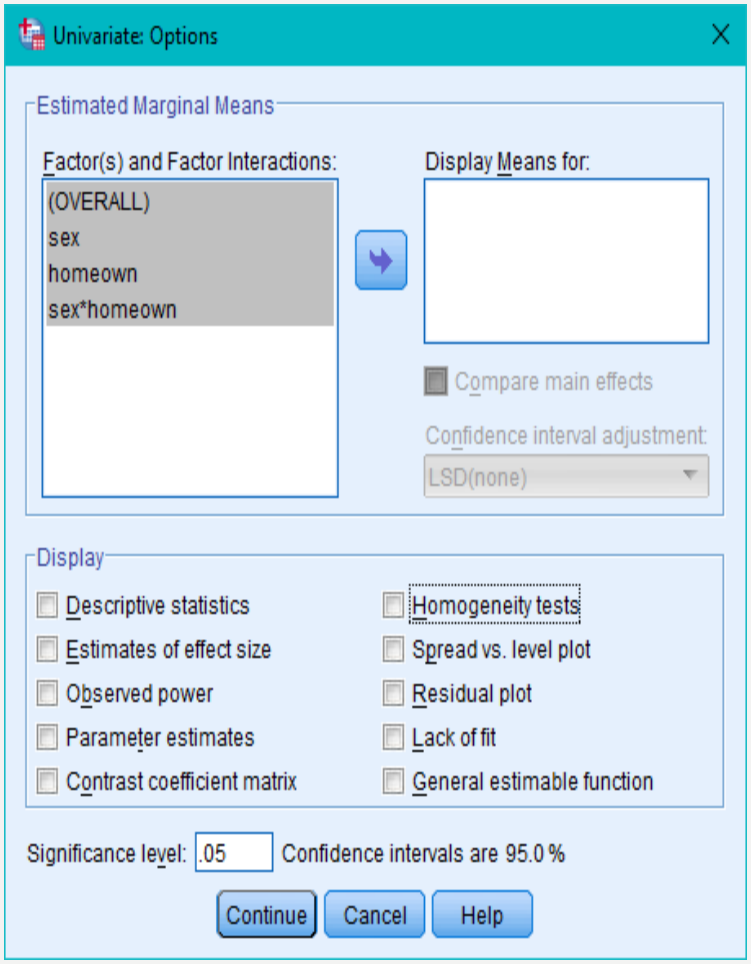


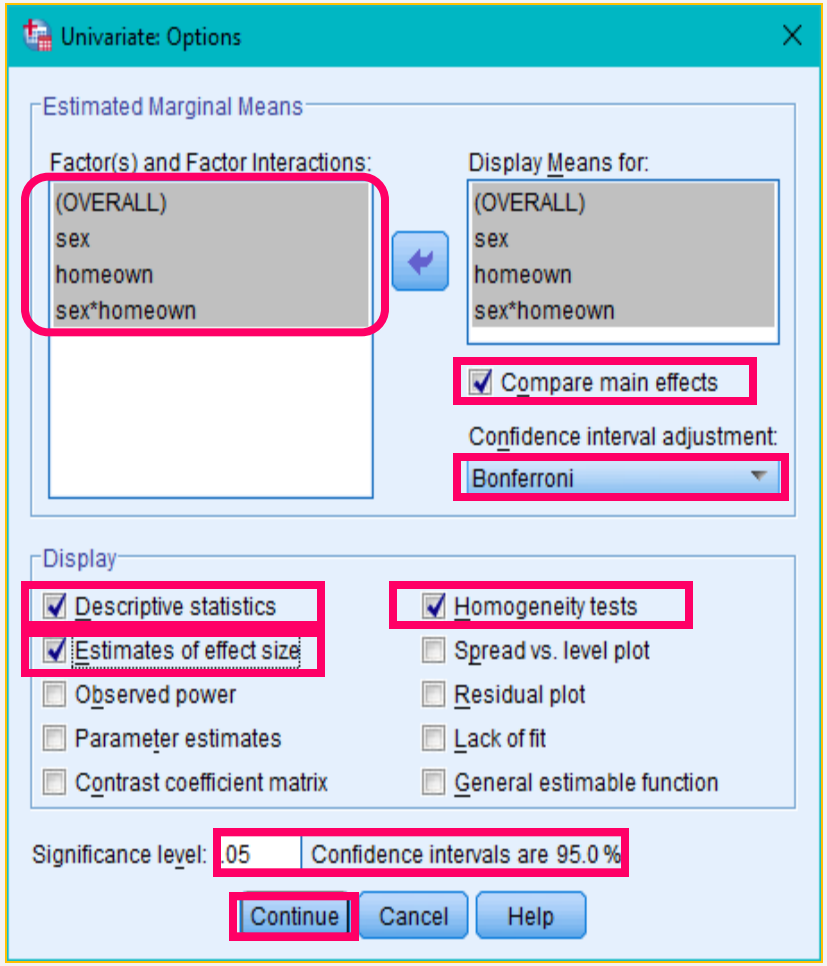
Click Continue

Since the post-hoc tests are only performed when there are more than two levels, we will skip the post-hoc tests button this time because our variables only have two levels.

Otherwise you do the post hoc tests just as you did for one-way ANOVA by moving the factors you want to test into the Post Hoc Tests box and selecting Sheffe and Tamhane tests.

CLICK ON: Options ➔ move all of the Factors (overall, Sex, Homeown, and Sex*Homeown) into the Display Means for box.



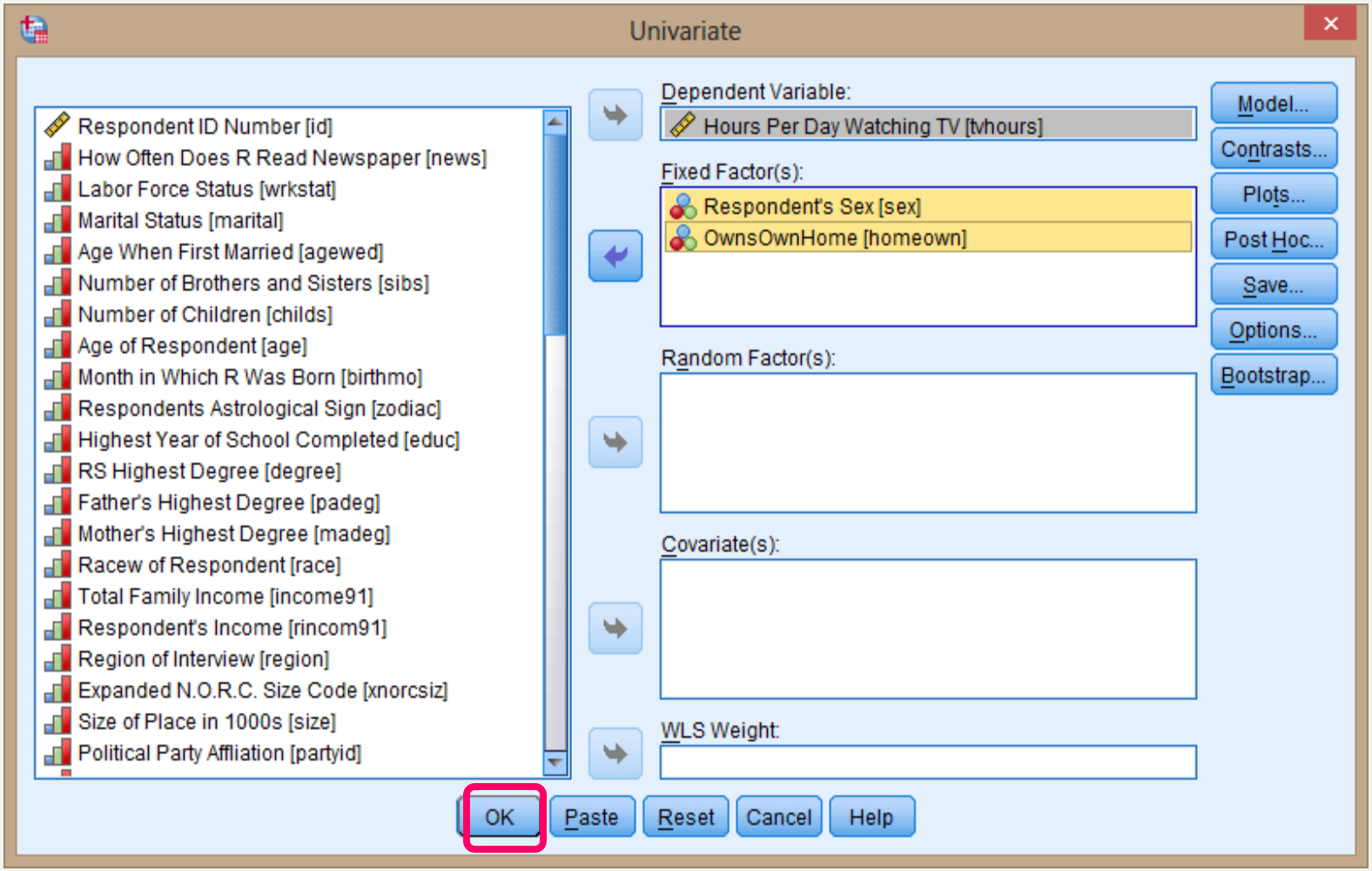


CHECK:

Compare Main Effects ➔ Bonferroni ➔ Descriptive ➔ Estimates of Effect Size ➔ Homogeneity Tests.

Set the level of significance for example 0.05 for confidence interval to 95%

CLICK CONTINUE YOU HAVE:



Then OK, you will have the output results as follows:

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Respondent's Sex	1	Male	451
	2	Female	554
OwnsOwnHome	1.00	Owns Own Home	658
	2.00	Doesn't Own Home	347

Descriptive Statistics

Dependent Variable: Hours Per Day Watching TV

Respondent's Sex	OwnsOwnHome	Mean	Std. Deviation	N
Male	Owns Own Home	2.77	2.197	305
	Doesn't Own Home	2.93	2.249	146
	Total	2.82	2.213	451
Female	Owns Own Home	2.63	1.790	353
	Doesn't Own Home	3.69	3.173	201
	Total	3.02	2.436	554
Total	Owns Own Home	2.70	1.989	658
	Doesn't Own Home	3.37	2.842	347
	Total	2.93	2.339	1005

As you can see in the table of means, there is a trend for females to watch more TV than males and for non- homeowners to watch more TV than homeowners, but there is a particularly pronounced trend for female non-homeowners to watch more TV than everybody else.

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Hours Per Day Watching TV

F	df1	df2	Sig.
9.951	3	1001	.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + sex + homeown + sex * homeown

Tests of Between-Subjects Effects

Dependent Variable: Hours Per Day Watching TV

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	153.307[a]	3	51.102	9.576	.000	.028
Intercept	8063.923	1	8063.923	1511.094	.000	.602
sex	21.146	1	21.146	3.963	.047	.004
homeown	81.594	1	81.594	15.290	.000	.015
sex * homeown	44.587	1	44.587	8.355	.004	.008
Error	5341.817	1001	5.336			
Total	14125.000	1005				
Corrected Total	5495.124	1004				

a. R Squared = .028 (Adjusted R Squared = .025)

1. Sex of respondent has a significant main effect on hours per day spent watching TV.

2. Home ownership has a significant main effect on hours per day spent watching TV.

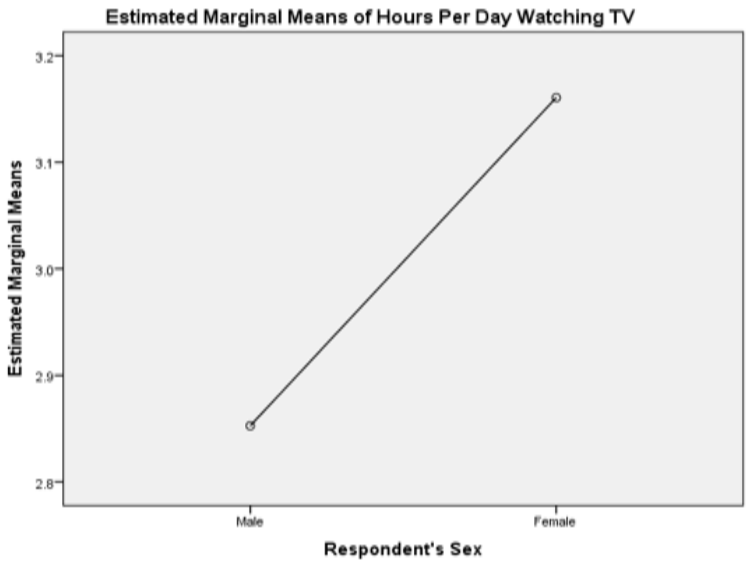
3. Sex of respondent home ownership have a significant interaction effect on hours per day spent watching TV.

Now write a paragraph in which you report the results of the significance tests! Remember that the interpretation of the main effects in a straightforward way is complicated by the significant interaction.
We also need to be a bit skeptical since the partial eta squares are very low and as you will see on the next slide there is a very large SD in one of the conditions

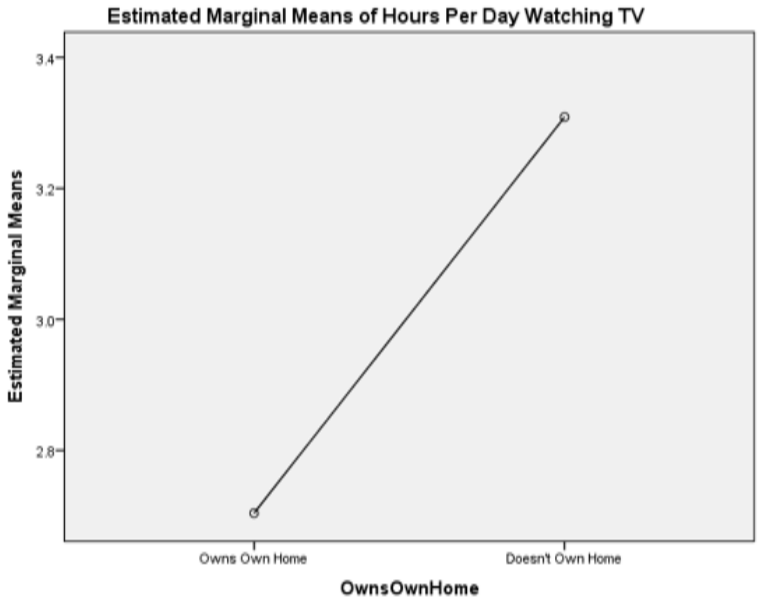
Descriptive Statistics

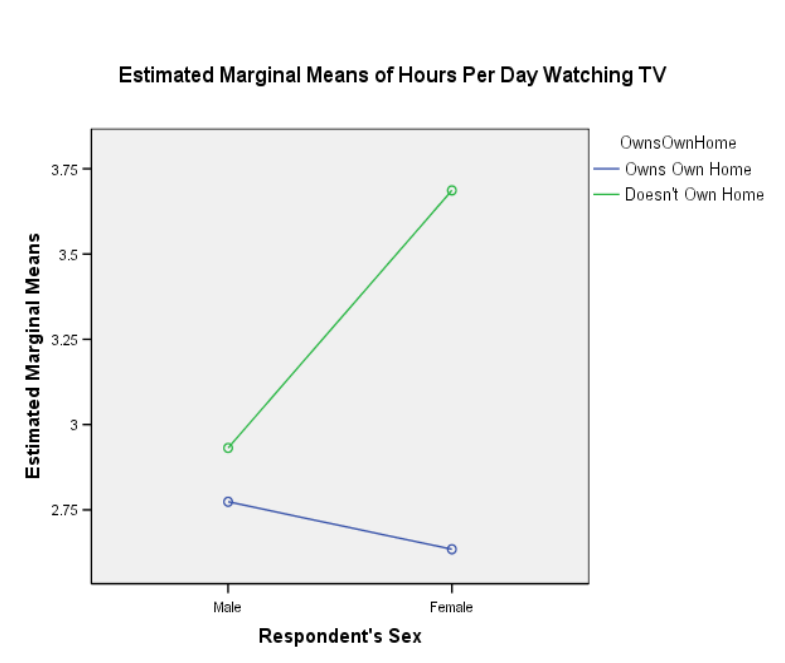
Dependent Variable: Hours Per Day Watching TV

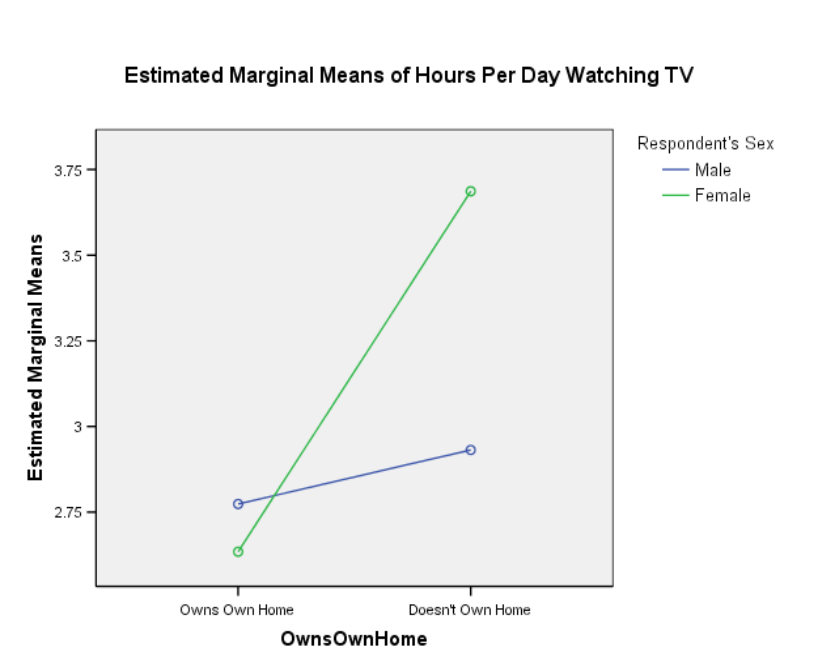
Respondent's Sex	OwnsOwnHome	Mean	Std. Deviation	N
Male	Owns Own Home	2.77	2.197	305
	Doesn't Own Home	2.93	2.249	146
	Total	2.82	2.213	451
Female	Owns Own Home	2.63	1.790	353
	Doesn't Own Home	3.69	3.173	201
	Total	3.02	2.436	554
Total	Owns Own Home	2.70	1.989	658
	Doesn't Own Home	3.37	2.842	347
	Total	2.93	2.339	1005



- As you can see in the table of means, there is a trend for females to watch more TV than males and for non- homeowners to watch more TV than homeowners, but there is a particularly pronounced trend for female non-homeowners to watch more TV than everybody else.







Although the interaction effect is not extremely strong, there is a trend for the relationship between homeownership and hours spent watching TV to be **different** for men than women;

Women who don't own homes are much more likely to spend more time watching TV than owners, compared to men, for whom homeownership makes less of a difference.

To know overall whether that IV influenced the DV, ignoring the other IV (Mane effect ➔ marginal means)

	Male	Female	
Owns Own Home	2.77	2.63	(2.7)
Doesn't Own Home	2.93	3.96	(3.31)
	(2.85)	(3.16)	

Main effect for ownership:

Does ownership matter? Does it effect time spent watching TV? 3.31 vs. 2.7

Main effect for Gender:

Does gender matter?

Does it effect time spent watching TV?

2.85 vs. 3.16

REPORT:

- Time a person spent on watching TV was analyzed in a two-way [between, within, mixed] ANOVA, with gender (male, female) as a [between subjects; within subjects] variable and home ownership (owns own home, doesn't own home) as a [between subjects; within subjects] variable.

- Main effect for Gender:

The main effect of gender on time a person spent on watching TV was significant, F= 3.963, p = 0.047.

Time a person spent on watching TV was higher for [gender, female] (M = 3.16) than for [gender, male] (M = 2.85).

In addition, time a person spent on watching TV was higher for [homeown, doesn't Own Home] (M = 3.31) than for [homeown, owns own Home] (M = 2.7).

The gender x homeown interaction was significant, $F=8.355$, $p = 0.004$

Since the interaction is significant, we need to report the cell means and describe their patterns as follows:

When the person Owns Own Home, women's average time spent on watching TV ($M=2.63$, $SD=1.790$) was lower than men's average time spent on watching TV ($M=2.77$, $SD=2.197$).
However, when the person Doesn't Own Home, women's average time spent on watching TV ($M=3.69$, $SD=3.173$) were higher than men's heart rates ($M=2.93$, $SD=2.249$), on average.

CONCLUSIONS

- T-tests assess if two group means differ significantly.
- ANOVAs compare more than two groups. They use variances instead of means.
- Described two-way analysis of variance
  - Examined effects of multiple factors and interaction

REPEATED MEASURES

ONE WAY REPEATED MEASURES ANOVA

- **A statistical technique for <u>testing for differences between the means of several groups:</u>**
  - **Groups are related in some way.**
  - **The same measurement is made several times on each subject or case.**
  - **Comparing the same subjects under several different treatments.**

Similar to the paired samples T-Test, but no restriction on the number of groups.

- # BETWEEN AND WITHIN FACTORS

Between factors:

A grouping or classification variables such as sex, age, grade levels, treatment conditions etc.

Within factors:

Is the one with multiple measures from a group of people such as time.

Between Subjects ANOVA
Data points in each group are unrelated

Group A	Group B	Group C
5	4	3
6	2	7
9	4	3
2	5	4
9	3	2

Repeated Measures ANOVA
Data points in each group are related

Time A	Time B	Time C
1	2	4
5	5	9
7	8	6
3	5	8
2	2	4

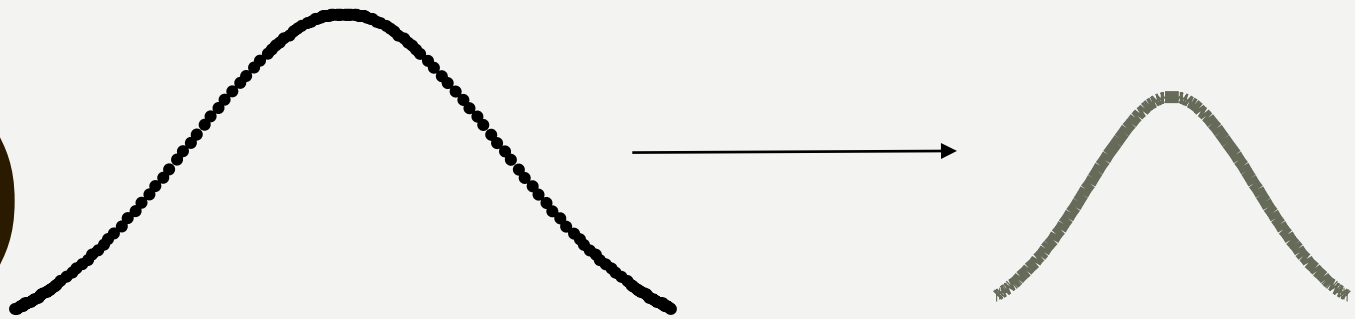
- **THE ANOVA MODEL FOR A SIMPLE REPEATED MEASURES DESIGN:**

REPEATED MEASURES

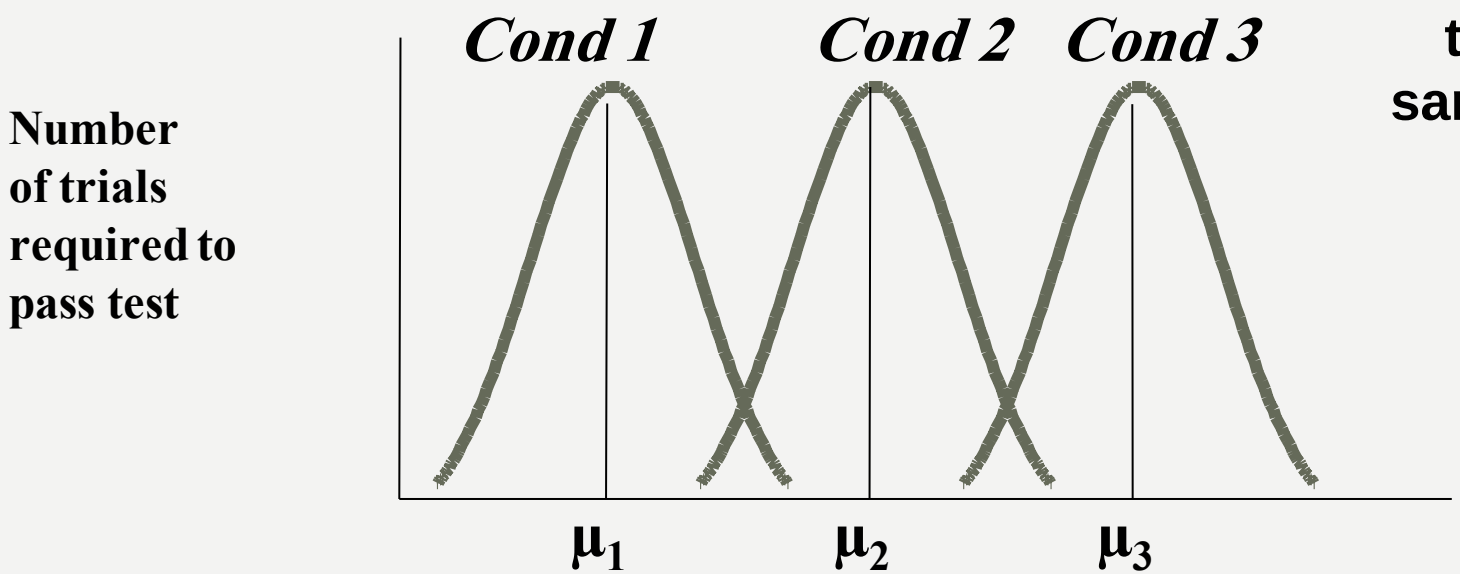
SUBJECTS

y_{11}	y_{12}	y_{13}	$\cdots$	y_{1t}
y_{21}	y_{22}	y_{23}	$\cdots$	y_{2t}
y_{n1}	y_{n2}	y_{13}	$\cdots$	y_{nt}

Population → **draw one sample**



Manipulate the sample



measure effect of IV on a DV

ASSUMPTIONS…

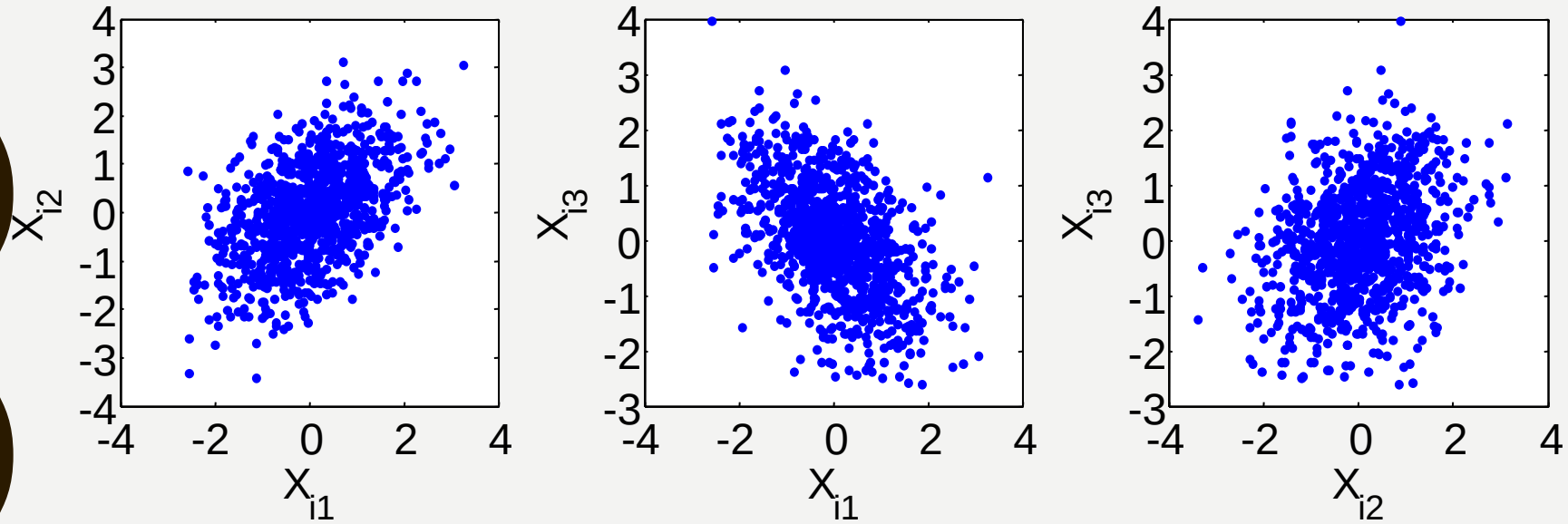
Data in each group should be…

- **Interval scale**
- **Normally distributed**
  - Histograms, box plots
- **Homogeneity of variance**
  - Variance within each condition should be roughly equal
- **RM require complete data for all subjects, i.e. no missing observations for any subjects.**

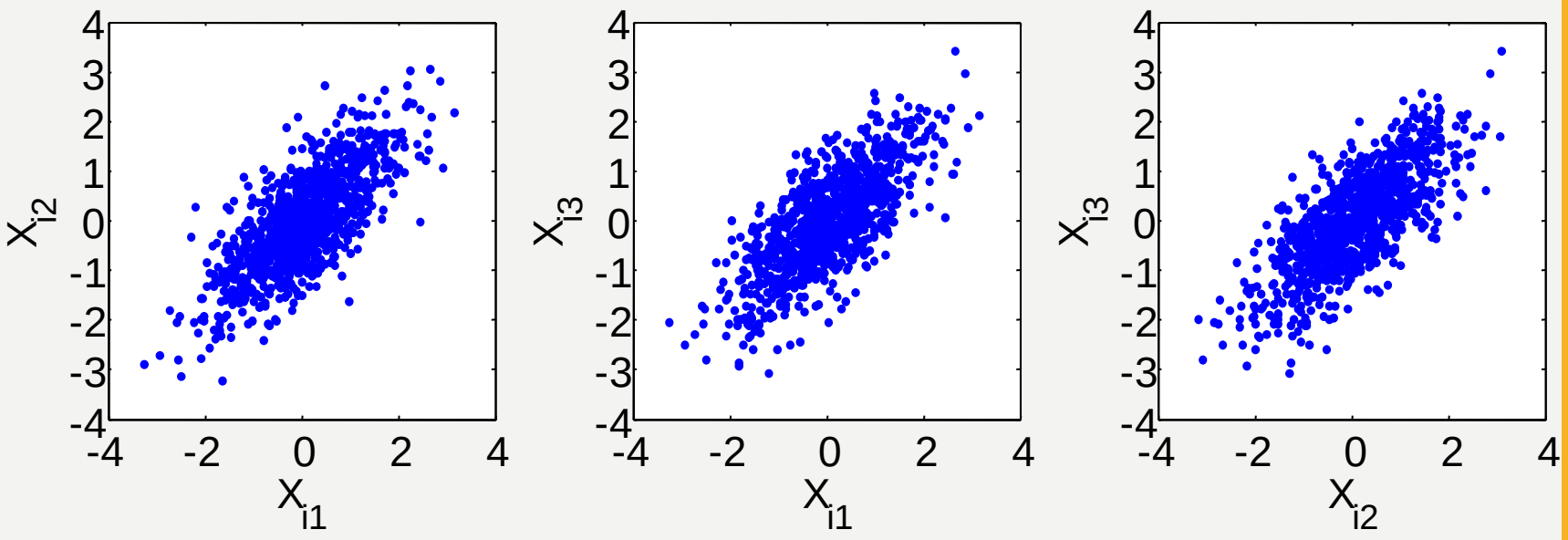
SPHERICITY

- **Sphericity** is the property that the degree of interaction (covariance) between any two different levels of the independent variable is the same.
- **Sphericity** is critical for RM ANOVA because the error term is the average of the pairwise interactions.
- Violations generally lead to inflated F statistics (and hence inflated Type I error).

SPHERICITY DOES NOT HOLD



SPHERICITY DOES HOLD



- ***SO THE SPHERICITY MEAN THAT:***
  - Variance of the differences between conditions is the same.
    - $\text{Variance}_{t1\text{-}t2} \approx \text{Variance}_{t1\text{-}t3} \approx \text{Variance}_{t2\text{-}t3}$
  - Correlation between pairs of groups is the same.
    - $\text{Corr}_{t1\text{-}t2} \approx \text{Corr}_{t1\text{-}t3} \approx \text{Corr}_{t2\text{-}t3}$

This assumption can be tested using Mauchly's Sphericity Test.

TESTING SPHERICITY

- **Mauchly's Test**
  - If $p > .05$, assume equality of variances.

- If p < .05, then the data fails to meet the assumption of sphericity.

Need to use one of the correction factors
E.g. Greenhouse-Geisser
A more conservative test, using different dfs

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
factor1	.077	44.681	9	.000	.414	.448	.250

Which estimate do you use?

Remember!: If the Epsilon estimates <.75, use Greenhouse-Geisser (GG);

If the Epsilon estimates > .75, use Huynh-Feldt (HF).

CORRELATION ACROSS TIME POINTS

	Time 1	Time 2	Time 3	Time 4
Time 1	1.00000	0.94035	-0.14150	0.28445
Time 2	0.94035	1.00000	-0.02819	0.26921
Time 3	-0.14150	-0.02819	1.00000	0.27844
Time 4	0.28445	0.26921	0.27844	1.00000

Time 1 and Time 2 are highly positively correlated, but Time 1 and Time 3 are *negatively* correlated!

Thus for this example we might conclude that the conditions for compound symmetry are NOT met.

Total Variance, Between Subjects ANOVA

Between Groups

Variance between the means

Captures the effect of the manipulation / treatment

The bigger the effect of your manipulation, the bigger the Between Groups Variance

Within Groups

Variance within each group

Variance due to random error / chance

NOW WE HAVE FOUR KINDS OF VARIANCE

- **TOTAL VARIANCE**
  - SS_{total}
  - $\sum (x_{ij} - \text{Grand Mean})^2$
- **Variance due to the manipulation**
  - $SS_{treatment}$
  - $n\sum (\text{Group mean}_j - \text{Grand Mean})^2$

- **Variance due to individual differences between participants**
  - $SS_{participants}$
  - No. of conditions $\sum$ (Participant mean - Grand Mean $)^2$

- **Variance due to random error**
  - $SS_{random\ error} = SS_{total} - SS_{treatment} - SS_{participants}$

F Ratio

Compares the variance due to the treatment / manipulation to the variance due to random error / chance

$$\frac{MS_{\text{between groups}}}{MS_{\text{within groups}}}$$

$$\frac{\text{Between Groups Variance}}{\text{Within Groups Variance}}$$

$$\frac{\text{Treatment Effect + Differences due to chance / error}}{\text{Differences due to chance / error}}$$

If the treatment / manipulation has an effect, then…

$$MS_{\text{between groups}} > MS_{\text{within groups}}$$

$$F > 1$$

Repeated measures (no grouping factor, one repeated measures factor (time))

Source	SS	df	MS	F
Between Subject	SS_{Subject}	n - 1	MS_{Subject}	$MS_{\text{Time}} / MS_{\text{Error}}$
Time	SS_{Time}	t - 1	MS_{Time}	
Between Subject Error	SS_{Error}	(n - 1)(t - 1)	MS_{Error}	

EXAMPLE:

Condition / Treatment

Person	1	2	3
1	2	4	7
2	10	12	13
3	22	29	30
4	30	31	34

Condition / Treatment

Person	1	2	3	
1	2	4	7	
2	10	12	13	
3	22	29	30	
4	30	31	34	
Mean	16	19	21	18.67

$SS_{treatment}$: no. of people in each condition $\sum$ (Treatment Mean - Grand Mean)^2

$$4 \sum(16 - 18.67)^2 + (19 - 18.67)^2 + (21 - 18.67)^2$$

Condition / Treatment

Person	1	2	3	Mean per person
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27
4	30	31	34	31.67
Mean	16	19	21	18.67

$SS_{participants}$: no. of conditions $\sum$ (Participant Mean - Grand Mean)2

$$3 \sum(4.33 - 18.67)^2 + (11.67 - 18.67)^2 + \ldots + (31.67 - 18.67)^2$$

Condition / Treatment

Person	1	2	3	Mean per person
1	2	4	7	4.33
2	10	12	13	11.67
3	22	29	30	27
4	30	31	34	31.67
Mean	16	19	21	18.67

$$SS_{\text{random error}} = SS_{\text{total}} - SS_{\text{treatment}} - SS_{\text{participants}}$$

ANOVA TABLE

Source of variation	SS	df	MS	F	p
Between participants (individual differences)	K ∑ (Participant Mean - Grand Mean)^2	n - 1			
Treatment	n ∑ (Treatment mean$_j$ - Grand Mean)^2	K - 1	SS_T / df_T	$\frac{MS_{Treatment}}{MS_{error}}$	Prob. of observing F-value when H_o is true
Random Error	SS_{total} - $SS_{treatment}$ - $SS_{participants}$	(n -1)x (k -1)	SS_e / df_e		
Total	∑ (x_{ij} - Grand Mean)^2	N - 1			

N = Total no. of observations, n = no. of people in each condition, K = no. of conditions

EXAMPLE (1)

- In the following study the experimenter was interested in how the level of a certain enzyme changed in cardiac patients after open heart surgery:

For n = 15 cardiac surgical patients, the enzyme was measured:

- Immediately after surgery (Day 0),
- One day (Day 1),
- Two days (Day 2) and
- One week (Day 7) after surgery.

- **The data are given in the table below:**

Subject	Day 0	Day 1	Day 2	Day 7	Subject	Day 0	Day 1	Day 2	Day 7
1	108	63	45	42	9	106	65	49	49
2	112	75	56	52	10	110	70	46	47
3	114	75	51	46	11	120	85	60	62
4	129	87	69	69	12	118	78	51	56
5	115	71	52	54	13	110	65	46	47
6	122	80	68	68	14	132	92	73	63
7	105	71	52	54	15	127	90	73	68
8	117	77	54	61					

- **The subjects are not grouped (single group).**

- **There is one repeated measures factor Time – with levels:**
  - **Day 0,**
  - **Day 1,**
  - **Day 2,**
  - **Day 7**

THE ANOVA TABLE FOR ENZYME EXPERIMENT

Source	SS	df	MS	F	p-value
Subject	4221.100	14	301.507	32.45	0.0000
Day	36282.267	3	12094.089	1301.66	0.0000
ERROR	390.233	42	9.291		

The Subject Source of variability is modelling the variability between subjects.

The ERROR Source of variability is modelling the variability within subjects.

EXAMPLE (2)

- We might compare a drug that is supposed to reduce cholesterol to placebo where cholesterol is measured every two months over a 12-month period.

- This differs from a block design approach in that we are interested in comparing treatment groups not just looking at whether there has been a change over time a SINGLE group of subjects, i.e.

$$H_o : \mu_{Time\,1} = \mu_{Time\,2} = \Lambda \ \mu_{Time\,T}$$

- **HYPOTHETICAL CHOLESTEROL STUDY**

Group	Initial	2 mo.	4 mo.	6 mo.	8 mo.	10 mo.	12 mo.
DRUG	The individual(s) who measure cholesterol levels at each follow-up are blind to which group the subjects are in.						
Placebo							

QUESTIONS OF INTEREST:

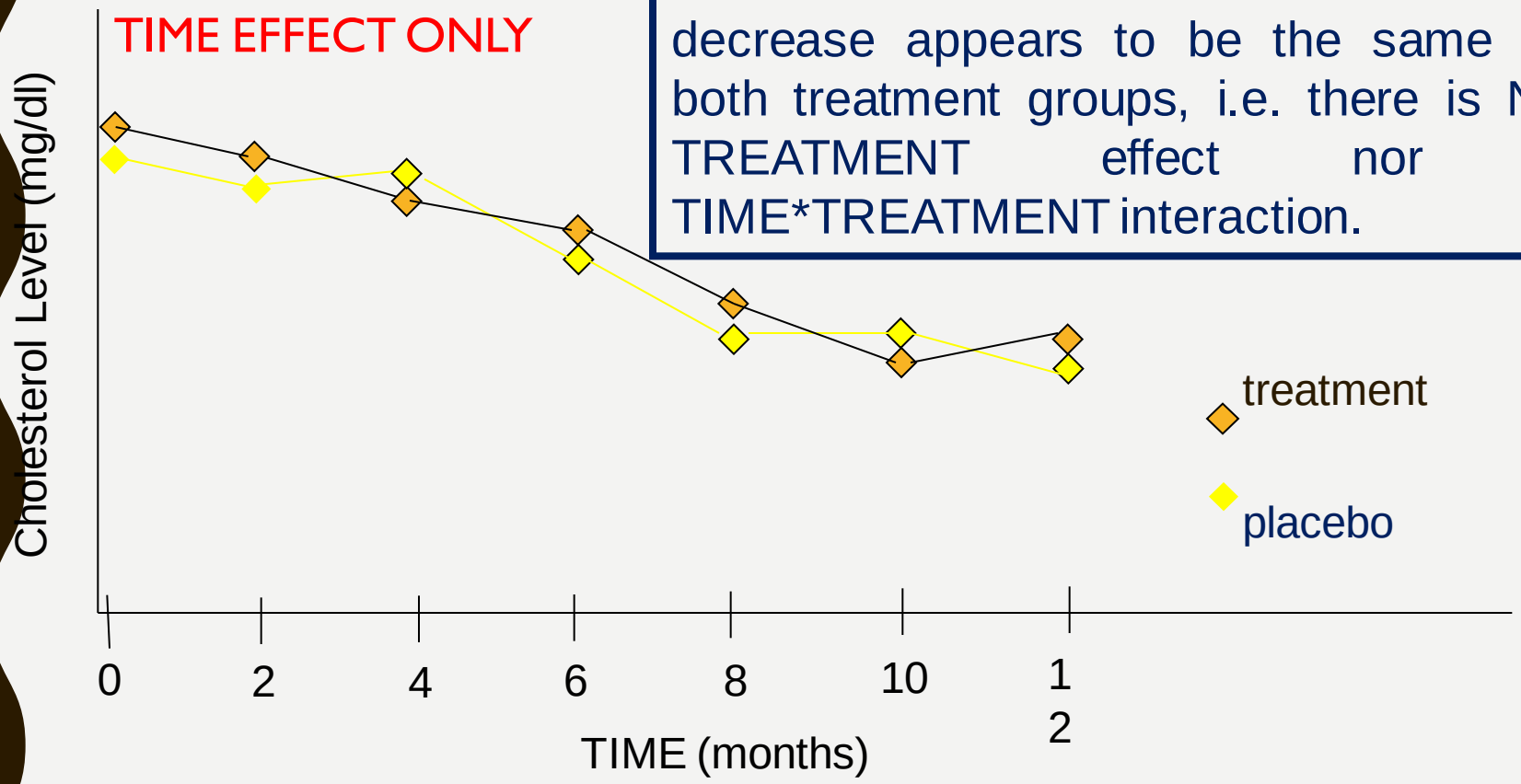
(1) Is there a change in the cholesterol levels of subjects over time, i.e. is there a TIME EFFECT? (within-subjects effect)

(2) Is there a TREATMENT EFFECT? (between-subjects effect)

(3) Is the effect of TIME the same for both TREATMENTS? (within-subjects effect)

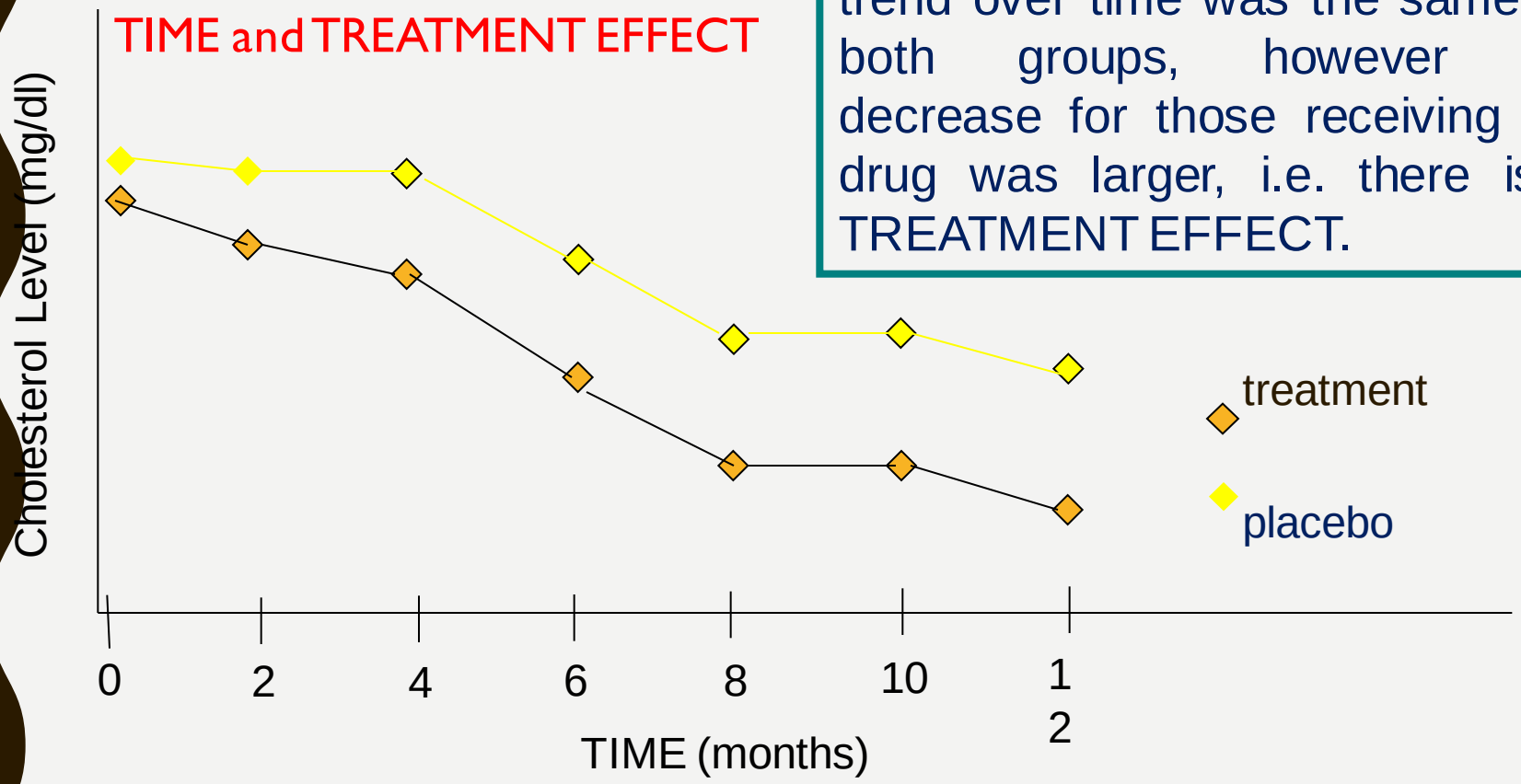
PROFILE PLOTS ILLUSTRATING THE QUESTIONS OF INTEREST

Cholesterol Levels for both groups decreased over TIME however the decrease appears to be the same for both treatment groups, i.e. there is NO TREATMENT effect nor a TIME*TREATMENT interaction.



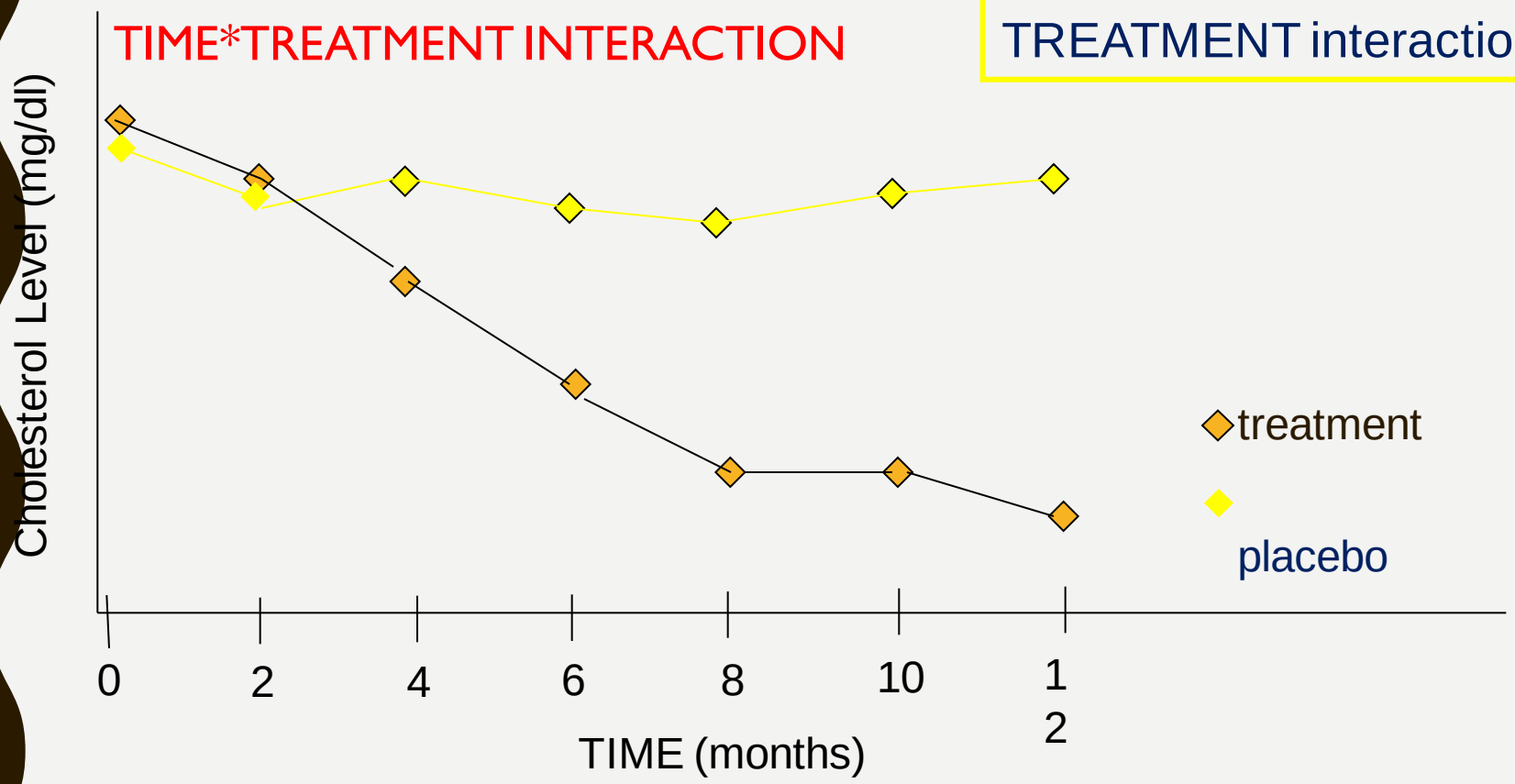
PROFILE PLOTS ILLUSTRATING THE QUESTIONS OF INTEREST

Cholesterol Levels for both groups decreased over TIME and the trend over time was the same for both groups, however the decrease for those receiving the drug was larger, i.e. there is a TREATMENT EFFECT.



PROFILE PLOTS ILLUSTRATING THE QUESTIONS OF INTEREST

Here the effect of time is NOT the same for both groups. Thus we say that there is TIME and TREATMENT interaction.

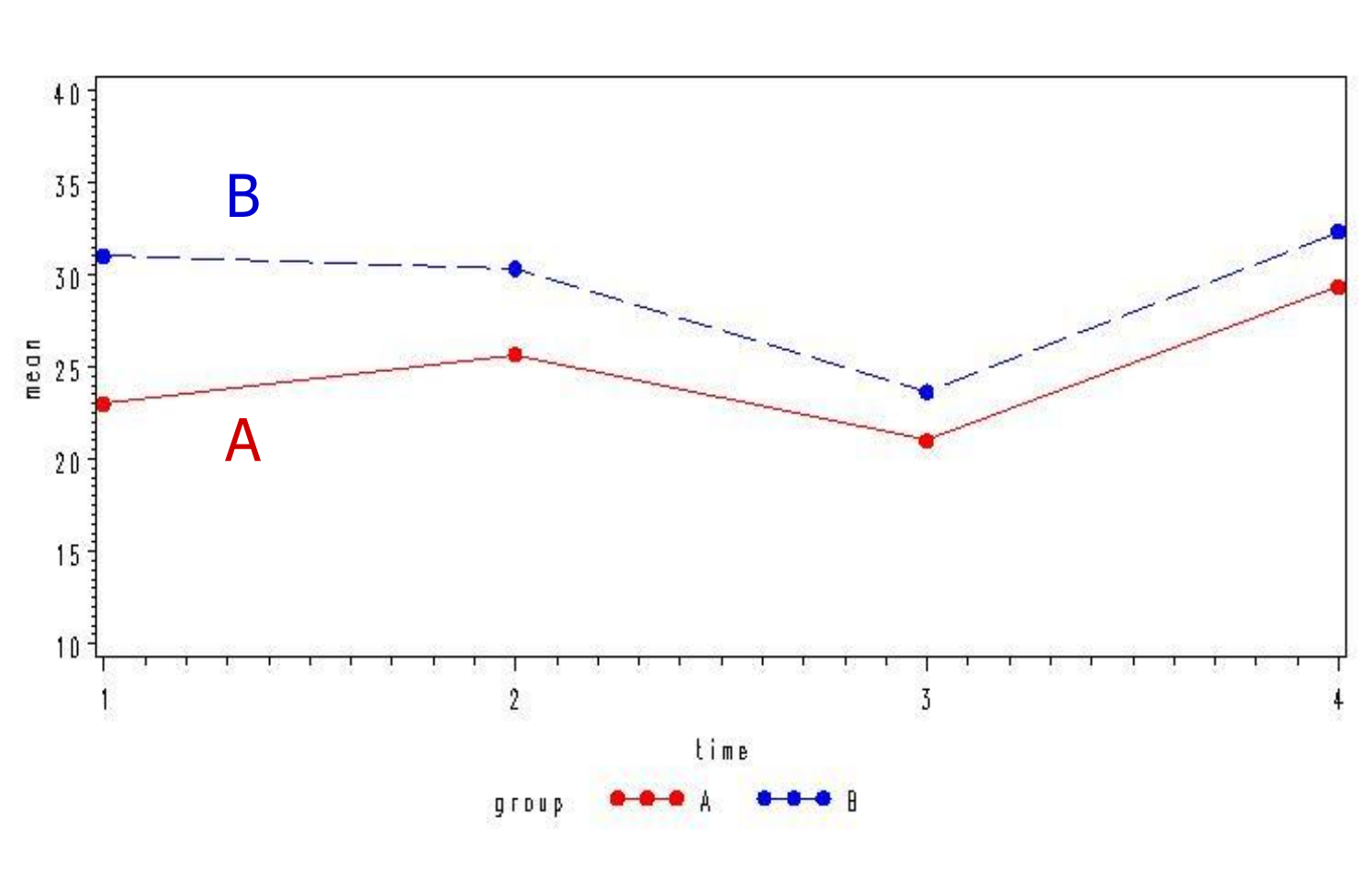


EXAMPLE (3)

Two treatment groups with four measurements taken over equally spaced time intervals (e.g., A = treatment B = placebo)

ID	Group	Time 1	Time 2	Time 3	Time 4
1	A	31	29	15	26
2	A	24	28	20	32
3	A	14	20	28	30
4	B	38	34	30	34
5	B	25	29	25	29
6	B	30	28	16	34

MEAN PROFILE PLOTS BY GROUP



QUESTIONS OF INTEREST

(1) OVERALL, ARE THERE SIGNIFICANT DIFFERENCES BETWEEN TIME POINTS?

From plots it looks like some differences over time, in particular times 3 and 4 look different.

(2) DO THE TWO GROUPS DIFFER AT ANY TIME POINTS, i.e. IS THERE A TREATMENT EFFECT?

From plots it looks like the groups differ at baseline and there are some difference everywhere else.

(3) DO THE TWO GROUPS DIFFER IN THEIR RESPONSES OVER TIME, I.E IS THERE A TIME*TREATMENT INTERACTION?

Their response profiles looks similar over time, though A and B are closer by the end.

NULL HYPOTHESIS SIGNIFICANCE TESTING

- ***Step 1:***

  H_0 – All of the condition means are equal

- ***Step 2:***

  H_{alt} – At least one condition mean is significantly different from the others

- ***Step 3:***

  Collect your data

- ***Step 4:***

  Run the ANOVA

- ***Step 5:***

  Obtain the F statistic and associated p value

- ***Step 6:***

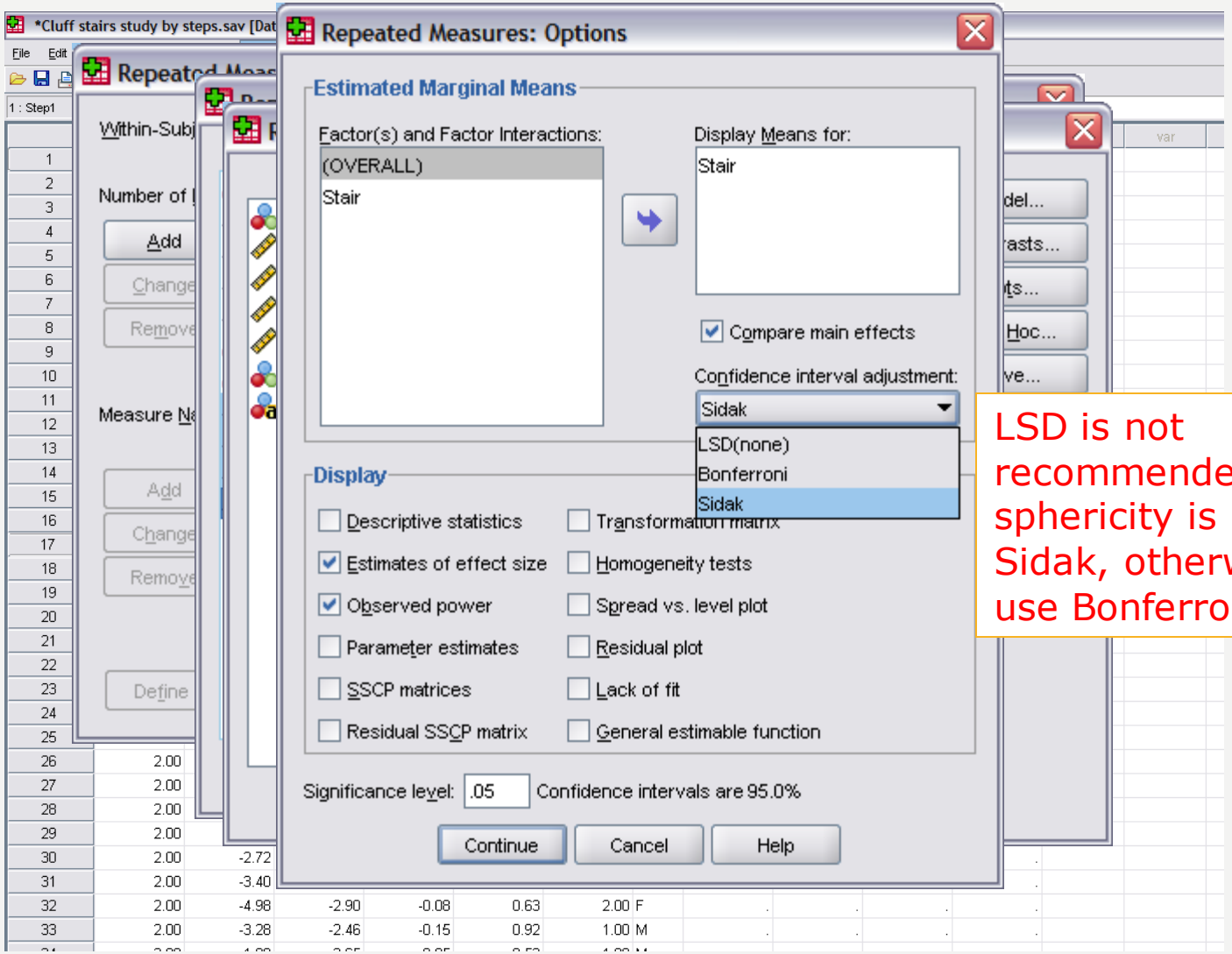
Decide whether to reject or fail to reject H_0 on the basis of the p value

POST HOC TESTING

- Significant F value
  - At least one condition mean is significantly different from the others
- But which one?
- Posthoc tests:
  - Bonferroni
  - Tukey
  - Sidak

GENERAL EXAMPLE

REPEATED-MEASURES ANOVA



LSD is not recommended, if sphericity is OK use Sidak, otherwise use Bonferroni.

REPEATED-MEASURES ANOVA: TEST FOR SPHERICITY

SPSS uses Mauchly's Test of Sphericity.

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subject...	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Stair	.585	7.893	5	.163	.723	.841	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design:
Within Subjects Design: Stair

Since the p-value (Sig. = 0.163) is greater than $\alpha = 0.05$, we accept the null hypothesis that covariance are equal and can "assume sphericity".

REPEATED-MEASURES ANOVA: RESULTS

SPSS shows results for four different assumptions. We can choose the first.

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Stair	Sphericity Assumed	17.422	3	5.807	27.706	.000
	Greenhouse-Geisser	17.422	2.170	8.029	27.706	.000
	Huynh-Feldt	17.422	2.523	6.907	27.706	.000
	Lower-bound	17.422	1.000	17.422	27.706	.000
Error(Stair)	Sphericity Assumed	10.061	48	.210		
	Greenhouse-Geisser	10.061	34.717	.290		
	Huynh-Feldt	10.061	40.360	.249		
	Lower-bound	10.061	16.000	.629		

Since the p-value (Sig. = .000) is less than $\alpha = 0.05$, the null hypothesis is rejected and conclude there is a significant difference across stair steps. Note, a p-value of 0.000 is written $p<0.0005$.

REPEATED-MEASURES ANOVA: TEST FOR BEST FIT

SPSS shows results of fitting polynomials from linear to degree *k*-1.

Tests of Within-Subjects Contrasts

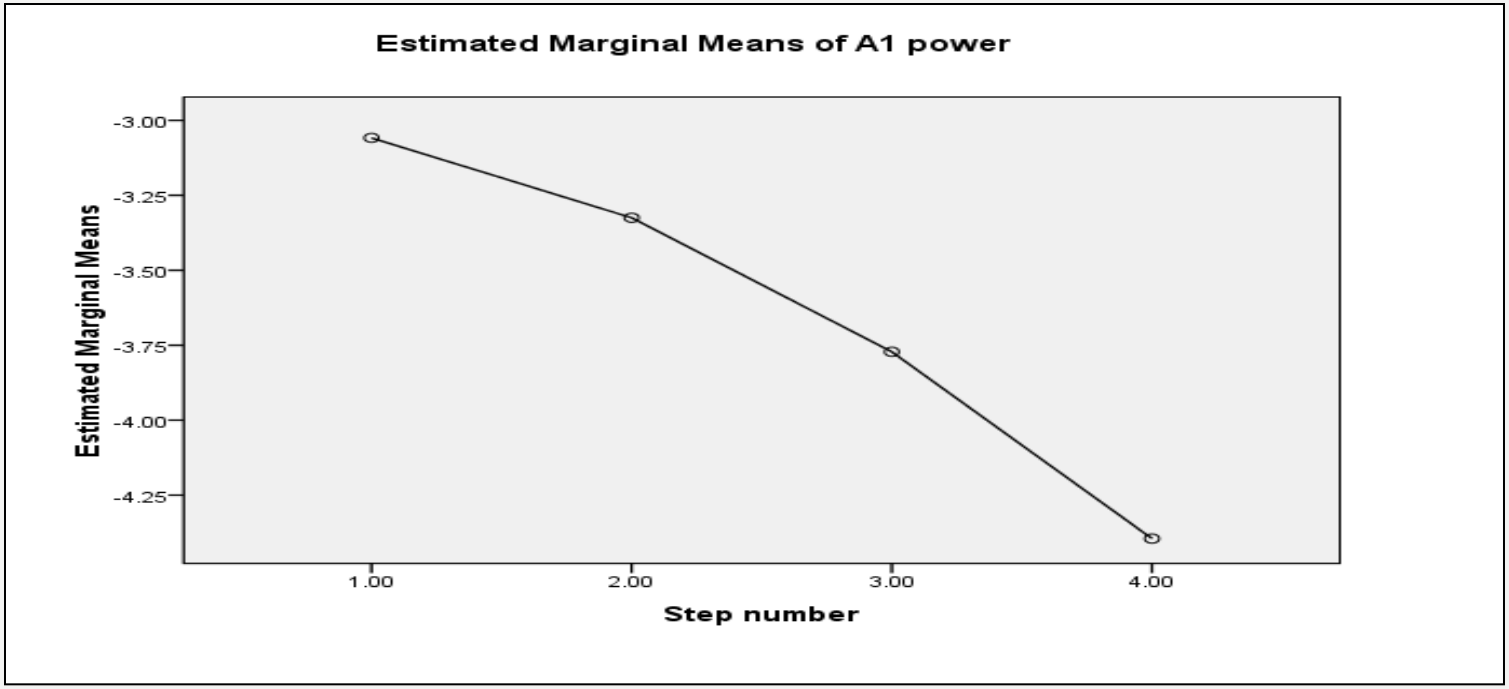
Measure:MEASURE_1

Source	Stair	Type III Sum of Squares	df	Mean Square	F	Sig.
Stair	Linear	16.882	1	16.882	71.498	.000
	Quadratic	.540	1	.540	3.207	.092
	Cubic	2.175E-5	1	2.175E-5	.000	.992
Error(Stair)	Linear	3.778	16	.236		
	Quadratic	2.696	16	.169		
	Cubic	3.587	16	.224		

Since there are only 4 steps, SPSS only tests to a cubic (3rd degree) fit. In this example a linear fit was best. Note, this statistic makes no sense if the DV is not ordered, such as time, age, or date.

REPEATED-MEASURES ANOVA: PLOT OF MARGINAL MEANS

SPSS can plot the group means. This plot shows the means for each step.



Looks like a linear increase in A1 power as people descend the stairs. Note, A1 is a negative power.

REPEATED-MEASURES ANOVA: POST HOC TESTS

- Since there is a significant F we can do *post hoc* testing. If not significant this step IS NOT DONE.
- We will use the **Sidak** *post hoc* test. **Bonferroni** is too conservative. Choose from the **Options…** menu, NOT the **Post Hoc…** menu!

Pairwise Comparisons

Measure:MEASURE_1

(I) Stair	(J) Stair	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	.266	.197	.730	-.326	.858
	3	.714*	.144	.001	.282	1.145
	4	1.336*	.192	.000	.762	1.911
2	1	-.266	.197	.730	-.858	.326
	3	.447*	.132	.023	.051	.843
	4	1.070*	.121	.000	.708	1.432
3	1	-.714*	.144	.001	-1.145	-.282
	2	-.447*	.132	.023	-.843	-.051
	4	.623*	.140	.002	.204	1.042
4	1	-1.336*	.192	.000	-1.911	-.762
	2	-1.070*	.121	.000	-1.432	-.708
	3	-.623*	.140	.002	-1.042	-.204

Based on estimated marginal means

a. Adjustment for multiple comparisons: Sidak.

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

Pairwise Comparisons

Measure:MEASURE_1

(I) Stair	(J) Stair	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	.266	.197	.730	-.326	.858
	3	.714*	.144	.001	.282	1.145
	4	1.336*	.192	.000	.762	1.911
2	1	-.266	.197	.730	-.858	.326
	3	.447*	.132	.023	.051	.843
	4	1.070*	.121	.000	.708	1.432
3	1	-.714*	.144	.001	-1.145	-.282
	2	-.447*	.132	.023	-.843	-.051
	4	.623*	.140	.002	.204	1.042
4	1	-1.336*	.192	.000	-1.911	-.762
	2	-1.070*	.121	.000	-1.432	-.708
	3	-.623*	.140	.002	-1.042	-.204

Based on estimated marginal means

a. Adjustment for multiple comparisons: Sidak.

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

The results now show that steps 1 and 2 are not significantly different for each other but are different from 3 and 4 and steps 3 and 4 are different from all the other steps. This is a better result than the factorial ANOVA.

EXAMPLE:

REPEATED MEASURES

- Subjects undergoing orthodontic treatment, were examined to see the amount of pain perception due to orthodontic treatment. Pain was measured after placing the appliance in different five times (3, 6, 9, 12, 15 minute) periods for the purpose of collecting data. The goal of the experiment was to determine if the pain will progress during treatment or would it stop at any point.

Data file: **RM Orthodontic Treatment**

Subject	Minute 3	Minute 6	Minute 9	Minute 12	Minute 15
1	7	7	23	36	70
2	12	22	26	26	20
3	11	6	9	31	30
4	10	18	16	40	25
5	6	12	9	28	37
6	13	21	30	55	65
7	5	0	2	10	11
8	15	18	22	37	42
9	0	2	0	16	11
10	6	8	27	32	54
ΣC	85	114	164	311	365
Mean =	8.5	11.4	16.4	31.1	36.5

Repeated Measures ANOVA: Data Entry

RM Orthodontic Treatment.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 6 of 6 Variables

	subject	ThreeMin	SixMin	NineMin	TwelveMin	FiftyMin	var
1	1.00	7.00	7.00	23.00	36.00	70.00	
2	2.00	12.00	22.00	26.00	26.00	20.00	
3	3.00	11.00	6.00	9.00	31.00	30.00	
4	4.00	10.00	18.00	16.00	40.00	25.00	
5	5.00	6.00	12.00	9.00	28.00	37.00	
6	6.00	13.00	21.00	30.00	55.00	65.00	
7	7.00	5.00	.00	2.00	10.00	11.00	
8	8.00	15.00	18.00	22.00	37.00	42.00	
9	9.00	.00	2.00	.00	16.00	11.00	
10	10.00	6.00	8.00	27.00	32.00	54.00	
11							
12							
13							

Data View | Variable View

IBM SPSS Statistics Processor is ready

Each level of a within subjects factor is entered as a separate variable.
Fatigue (3, 6, 9, 12, 15 min)

RM Orthodontic Treatment.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	subject	Numeric	8	2		None	None	8	Right	Scale	Input
2	ThreeMin	Numeric	8	2		None	None	8	Right	Scale	Input
3	SixMin	Numeric	8	2		None	None	8	Right	Scale	Input
4	NineMin	Numeric	8	2		None	None	8	Right	Scale	Input
5	TwelveMin	Numeric	8	2		None	None	8	Right	Scale	Input
6	FiftyMin	Numeric	8	2		None	None	8	Right	Scale	Input
7											
8											
9											
10											

Data View Variable View

IBM SPSS Statistics Processor is ready

STEPS IN SPSS IS:

General Linear Model ➔ Repeated Measures

RM Orthodontic Treatment.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 6 of 6 Variables

	subject	ThreeMin	elveMin	FiftyMin	var
1	1.00	7.0	36.00	70.00	
2	2.00	12.0	26.00	20.00	
3	3.00	11.0			
4	4.00	10.0			
5	5.00	6.0			
6	6.00	13.0			
7	7.00	5.0			
8	8.00	15.0	37.00	42.00	
9	9.00	.0	16.00	11.00	
10	10.00	6.0	32.00	54.00	
11					
12					
13					

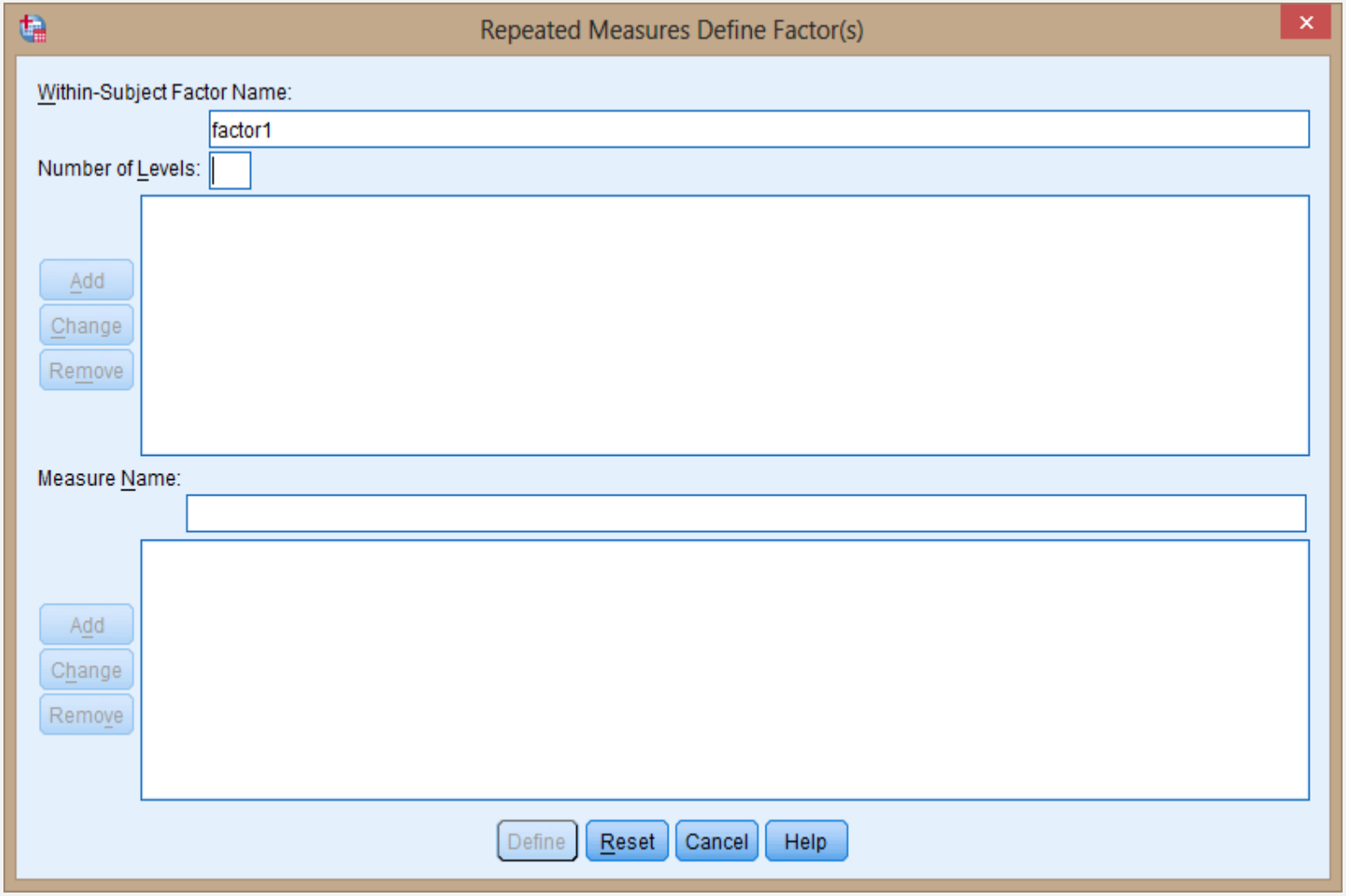
Analyze:
- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis

General Linear Model:
- Univariate...
- Multivariate...
- Repeated Measures...
- Variance Components...

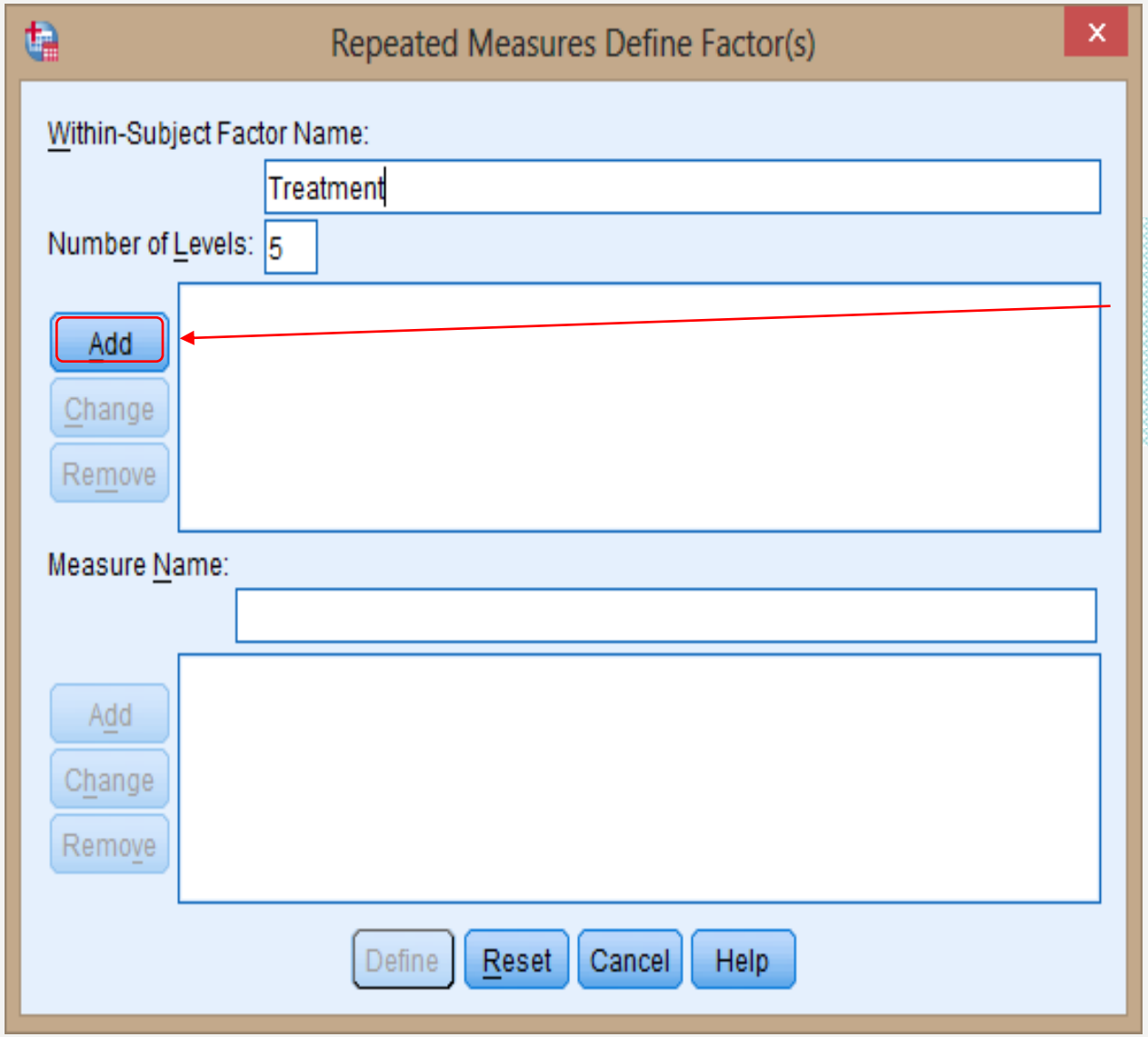
Data View | Variable View

Repeated Measures...

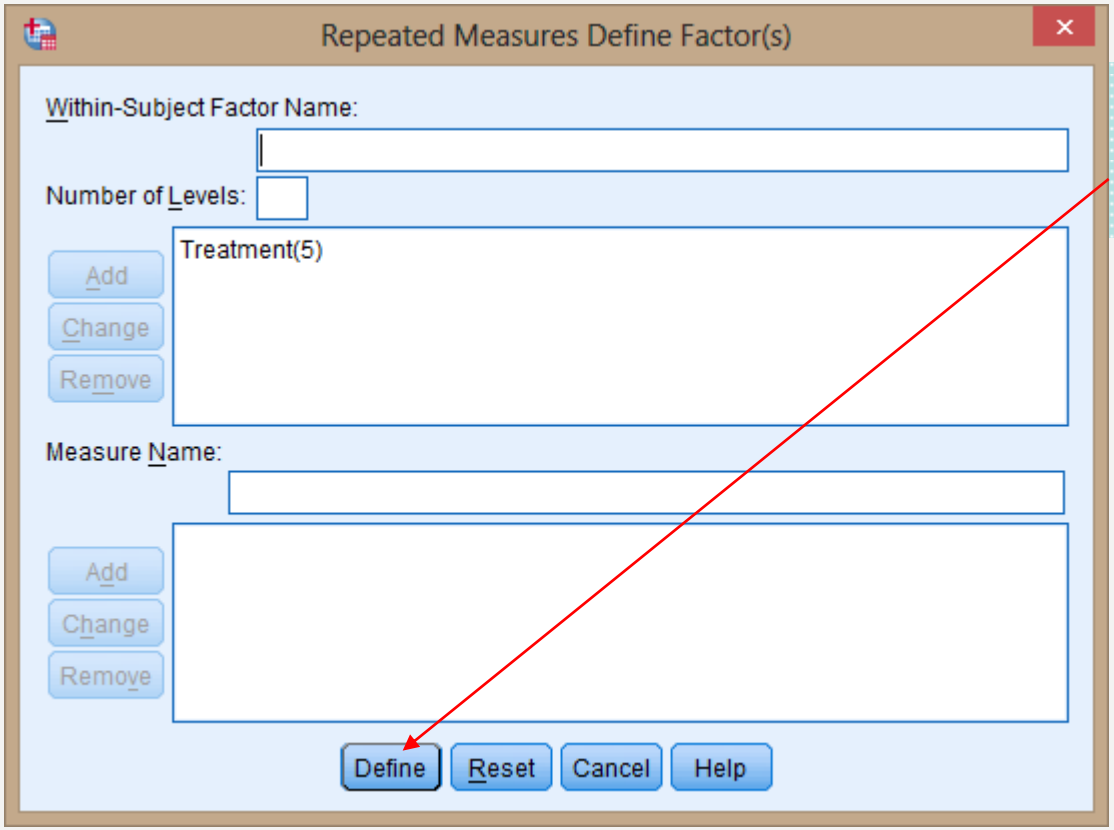
tatistics Processor is ready


Repeated Measures Define Factor(s)
Within-Subject Factor Name:
factor1
Number of Levels:
Add
Change
Remove
Measure Name:
Add
Change
Remove
Define
Reset
Cancel
Help

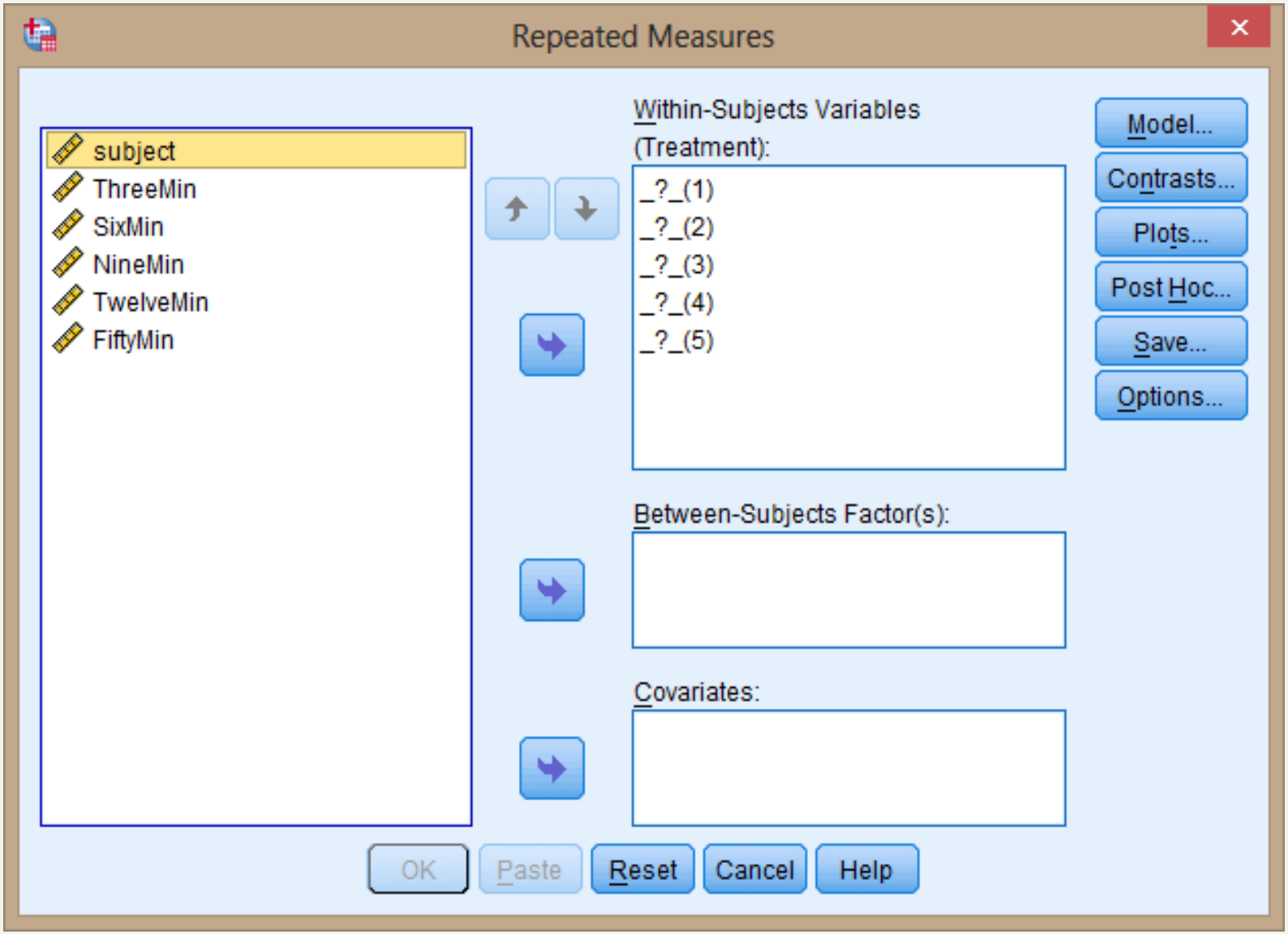
Name and Define the Within Subjects Factors

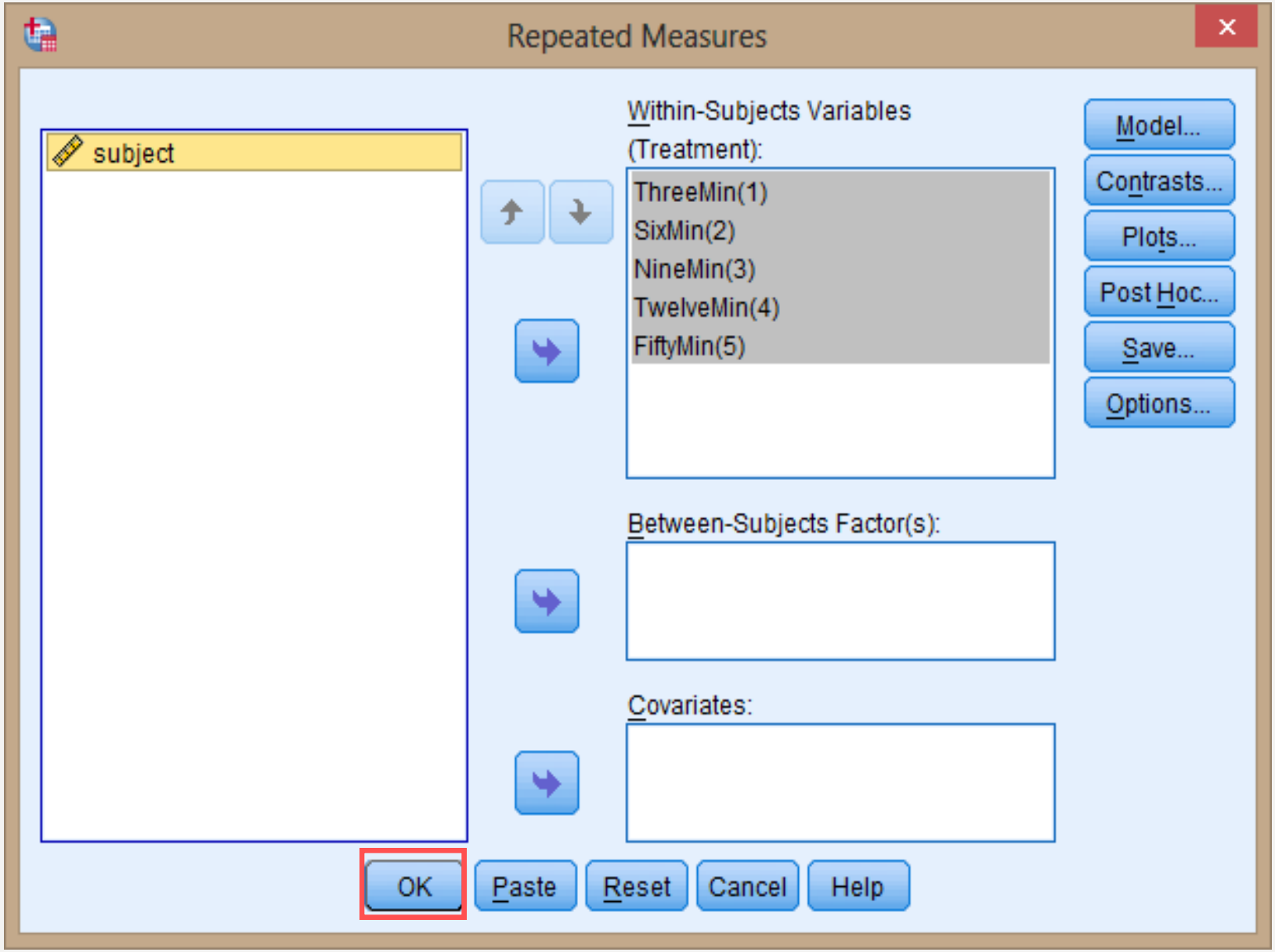


Click Add to enter each within subjects factor.



Click Define to define both Within and Between Subjects Factors.


Repeated Measures
subject
ThreeMin
SixMin
NineMin
TwelveMin
FiftyMin
Within-Subjects Variables (Treatment):
_?_(1)
_?_(2)
_?_(3)
_?_(4)
_?_(5)
Between-Subjects Factor(s):
Covariates:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
OK
Paste
Reset
Cancel
Help



To check ANOVA Assumptions: Sphericity click OK

Repeated Measure ANOVA Assumptions: Sphericity?

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Treatment	.024	27.594	9	.001	.371	.428	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept

Within Subjects Design: Treatment

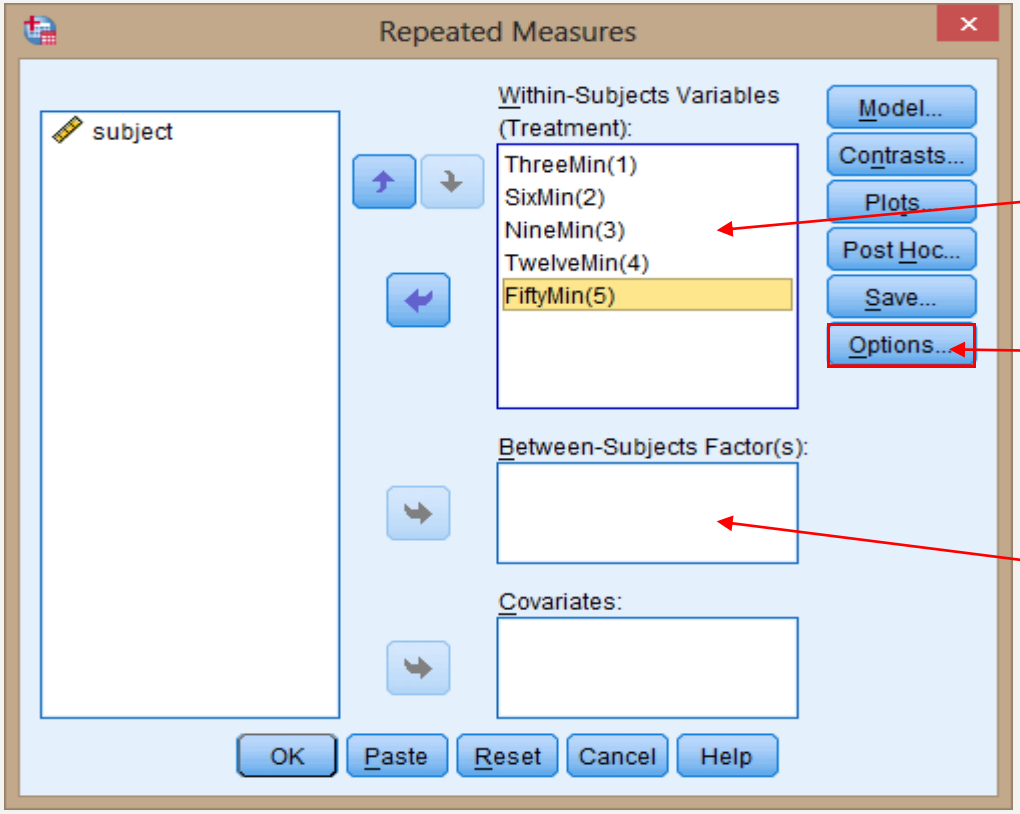
b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

You don't want this to be significant.

Mauchly's Test of Sphericity indicated that sphericity was violated

Since Sphericity is violated and epsilon (0.371) <0.75, we use the Greenhous-Geisser

Defining Within & Between Subjects Factors

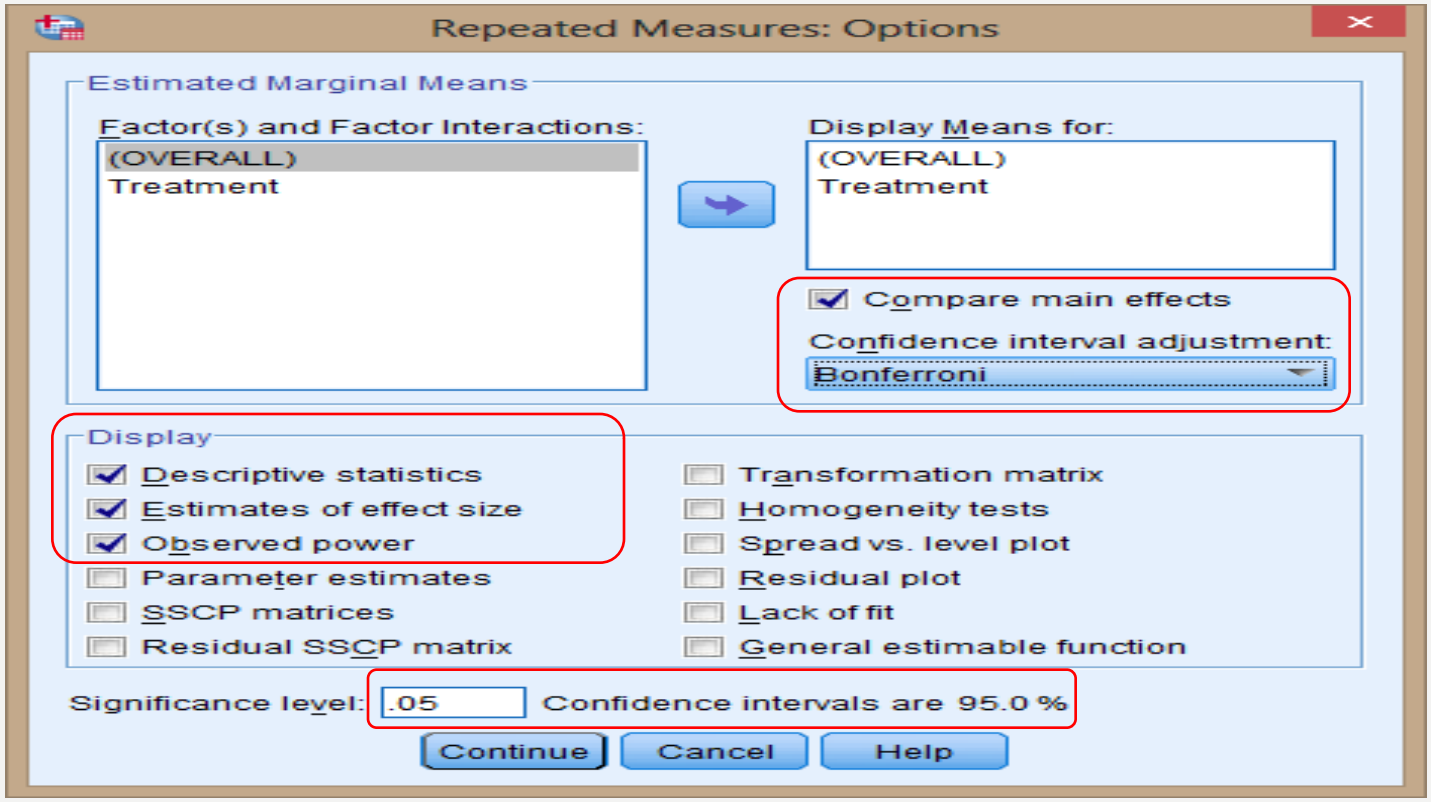


Within Subjects Factors

Click Options

Between Subjects Factors

Repeated Measures Options



SPSS Output

General Linear Model

Within-Subjects Factors

Measure: MEASURE_1

treatmnt	Dependent Variable
1	Minute_3
2	Minute_6
3	Minute_9
4	Minute_12
5	Minute_15

Descriptive Statistics

	Mean	Std. Deviation	N
Minute_3	8.5000	4.50309	10
Minute_6	11.4000	7.96102	10
Minute_9	16.4000	10.80329	10
Minute_12	31.1000	12.55610	10
Minute_15	36.5000	21.13055	10

Multivariate Tests[c]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
treatmnt	Pillai's Trace	.866	9.694[b]	4.000	6.000	.009	.866	38.777	.934
	Wilks' Lambda	.134	9.694[b]	4.000	6.000	.009	.866	38.777	.934
	Hotelling's Trace	6.463	9.694[b]	4.000	6.000	.009	.866	38.777	.934
	Roy's Largest Root	6.463	9.694[b]	4.000	6.000	.009	.866	38.777	.934

a. Computed using alpha = .05

b. Exact statistic

c.
Design: Intercept
Within Subjects Design: treatmnt

SPSS Output: Within Subjects Factors

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Treatment	Sphericity Assumed	6115.880	4	1528.970	18.359	.000	.671	73.437	1.000
	Greenhouse-Geisser	6115.880	1.485	4117.754	18.359	.000	.671	27.268	.995
	Huynh-Feldt	6115.880	1.710	3575.916	18.359	.000	.671	31.400	.998
	Lower-bound	6115.880	1.000	6115.880	18.359	.002	.671	18.359	.967
Error(Treatment)	Sphericity Assumed	2998.120	36	83.281					
	Greenhouse-Geisser	2998.120	13.367	224.289					
	Huynh-Feldt	2998.120	15.393	194.776					
	Lower-bound	2998.120	9.000	333.124					

a. Computed using alpha = .05

Since Sphericity was violated we use the adjusted values:

$F(1.485, 13.367) = 18.36$,

SPSS Output: Between Subjects Effects

Tests of Between-Subjects Effects

Measure: MEASURE 1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	21590.420	1	21590.420	45.801	.000	.836	45.801	1.000
Error	4242.580	9	471.398					

a. Computed using alpha = .05

Since p value<0.05, we reject H_0 and accept H_A that at least one time period different in pain sensation.

SPSS Output: Effect Size & Confidence Intervals

1. Grand Mean

Measure: MEASURE_1

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
20.780	3.070	13.834	27.726

Estimates

Measure: MEASURE_1

treatmnt	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	8.500	1.424	5.279	11.721
2	11.400	2.517	5.705	17.095
3	16.400	3.416	8.672	24.128
4	31.100	3.971	22.118	40.082
5	36.500	6.682	21.384	51.616

Post hoc Tests for Main Effects (Treatment means)

Pairwise Comparisons

Measure: MEASURE_1

(I) treatmnt	(J) treatmnt	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-2.900	1.656	.114	-6.647	.847
	3	-7.900*	2.718	.017	-14.049	-1.751
	4	-22.600*	3.194	.000	-29.826	-15.374
	5	-28.000*	6.354	.002	-42.375	-13.625
2	1	2.900	1.656	.114	-.847	6.647
	3	-5.000	2.380	.065	-10.385	.385
	4	-19.700*	2.848	.000	-26.143	-13.257
	5	-25.100*	6.457	.004	-39.708	-10.492
3	1	7.900*	2.718	.017	1.751	14.049
	2	5.000	2.380	.065	-.385	10.385
	4	-14.700*	2.633	.000	-20.657	8.743
	5	-20.100*	4.792	.002	-30.941	-9.259
4	1	22.600*	3.194	.000	15.374	29.826
	2	19.700*	2.848	.000	13.257	26.143
	3	14.700*	2.633	.000	8.743	20.657
	5	-5.400	4.525	.263	-15.635	4.835
5	1	28.000*	6.354	.002	13.625	42.375
	2	25.100*	6.457	.004	10.492	39.708
	3	20.100*	4.792	.002	9.259	30.941
	4	5.400	4.525	.263	-4.835	15.635

Based on estimated marginal means

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

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

For Example:

- 4 (12 min) is diff from: 1, 2, 3 (3, 6, 9 min)

==========

- 5 (15 min) is diff from: 1, 2, 3 (3,6,9 min)

REPORT

1. Data.

2. Assumptions. We assume that the assumptions for the one way ANOVA repeated measures design are met.

3. Aim : to determine if the pain will progress during treatment or would it stop at any point.

4. Hypotheses.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_A: not all μ's are equal

5. Using one-way ANOVA repeated measures, from Manchly's test of sphericity the Sig less than 0.05 sphericity violated and by Greenhouse the sphericity assumed.
6. Using F test and P-value, we reject H_0 and accept H_A that at least one time period different in pain sensation. So we must do paired test for every pairwise and the results reported as follows from pairwise comparisons table (you may do a similar report for μ_2 with the other means and μ_3 , μ_4 , μ_5) : For example:

- H_0: $\mu_1 = \mu_2$ Versus H_1: $\mu_1 \neq \mu_2$ (since P=0.114>0.05, we accept H_0)
- H_0: $\mu_1 = \mu_3$ Versus H_1: $\mu_1 \neq \mu_3$ (since P=0.017<0.05, we reject H_0)
- H_0: $\mu_1 = \mu_4$ Versus H_1: $\mu_1 \neq \mu_4$ (since P=0.000<0.05, we reject H_0)
- H_0: $\mu_1 = \mu_5$ Versus H_1: $\mu_1 \neq \mu_5$ (since P=0.002<0.05, we reject H_0)

….. etc.

SUMMARY

In this lecture, we review some Parametric Testes such as T-test (1-sample, 2-samples and paired samples), one and two way ANOVA and repeated measures.

In the next lectures, we will focus on some Non- Parametric Tests which are the content of STAT 333 course.

End of the Lecture