

# Introduction to Inferential Statistics

## Table of Contents

---

- [Cross tabulation and Chi-Squared tests](#)
- [Significance testing](#)
- [T-tests](#)
- [Paired sample t-tests](#)
- [One-way ANOVA](#)
- [Repeated Measures ANOVA](#)
- [Course survey](#)

## Getting started

---

We will be using the same dataset we have used in previous sessions, the ESS ([European Social Survey](#)) data that is called **ESS5GB.sav**.

Open this file now, either by double clicking on the file in finder (Mac) or Windows Explorer, or using *file -> open -> data* in SPSS.

Follow the example steps provided, then have a go at the challenges for each section.

## What is inferential statistics?

---

Some notes here on what inferential statistics are, why they are used, and why important.

## Cross tabulation and Chi-Squared tests

---

The chi-squared test is for comparing observed frequencies to expected frequencies - **what is an observed and expected frequency?**

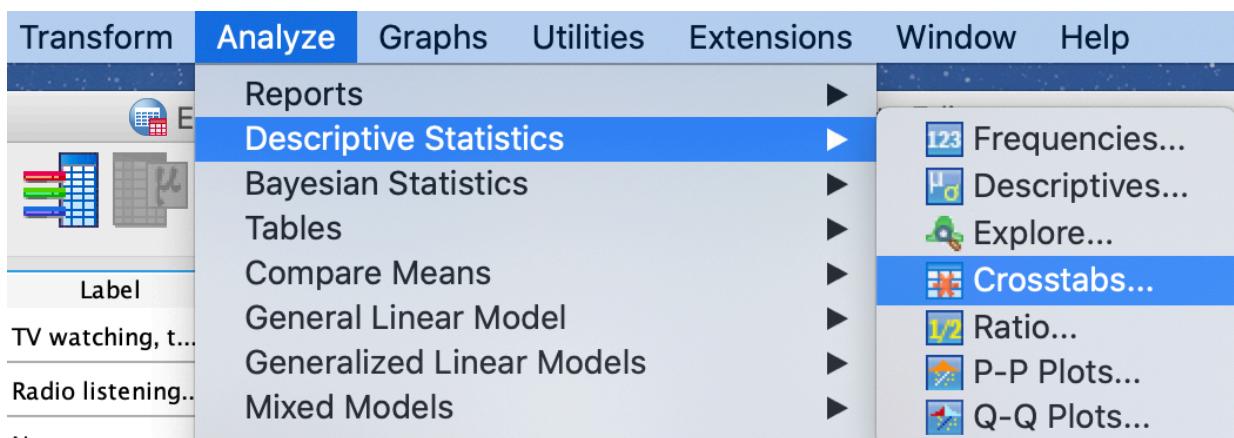
The Chi-squared test looks at associations between two categorical variables. We use it when we are testing frequency data, the number of times something occurs. We review these associations with cross tabulation, that displays these associations in a grid.

For example, in the table below we look at if smokers are ill or not. We can see that out of 56 smokers 39 are ill, compared to 18 out of 78 non-smokers who are not ill. The chi-square

tests lets us assess if this apparent association is significant.

	Smoker	Non-Smoker	Total
Ill	39	18	57
Not ill	17	60	77
Total	56	78	134

To run a chi-square in SPSS we need to go to the crosstabs menu. See the image below how to find crosstabs:



This should bring up the cross tabulation screen. From here we can add our variables to test on.

One might suggest that there is an association between someones gender and if they have been the victim of assault or burglary. As both these variables are categorical, we can use a chi-square to test out this association.

To run a Chi-square test we need to go into the statistics tab, and select chi-square.

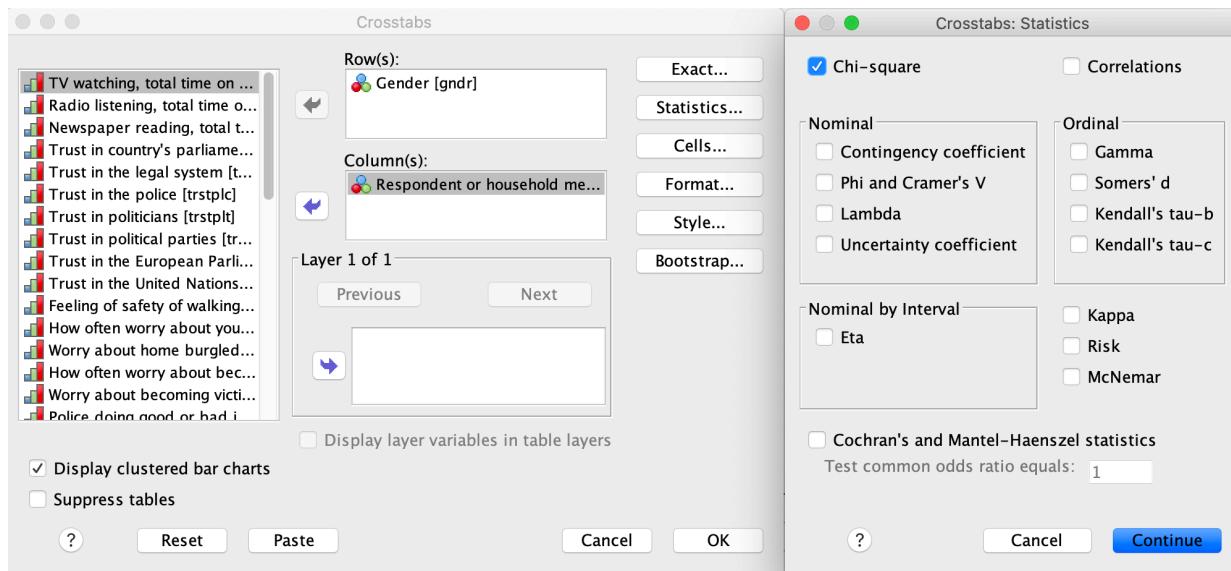
It is also important to get the *expected counts*. To add the expected counts go to cells -> tick expected counts.

Expected counts are a probability test to determine what count/frequency of an observation we would expect to see from our data. We calculate these using this formula:

`expected counts = (row total * column total)/sample size`. From our smoking example our expected count for smoker and ill would be:

$(57 * 56) / 134 = 23.8209$ ; so we expect 23.82 people who smoke to be ill. As a general rule for chi-square tests, we would expect all expected frequencies to be greater than 5.

We can also select *display clustered bar charts*, which runs a nice visualisation on our cross tabulation. See the image below of what we have done:



Once you run the cross tabulation you will get several tables and a graph that will pop up in your output viewer. Of most interest are the *Crosstabulation* and the *Chi-Square Tests* tables, shown below:

### Gender \* Respondent or household member victim of burglary/assault last 5 years Crosstabulation

		Respondent or household member victim of burglary/assault last 5 years		Total
Gender	Male	Count	206	849
	Female	Expected Count	201.6	853.4
Total	Male	Count	256	1107
	Female	Expected Count	260.4	1102.6
Total		Count	462	1956
		Expected Count	462.0	1956.0
				2418

### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.213 <sup>a</sup>	1	.644		
Continuity Correction <sup>b</sup>	.168	1	.682		
Likelihood Ratio	.213	1	.645		
Fisher's Exact Test				.677	.341
Linear-by-Linear Association	.213	1	.645		
N of Valid Cases	2418				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 201.58.

b. Computed only for a 2x2 table

The cross tabulation gives us a visual summary of our data. We can see that more females

than males responded to both yes and no to the burglary/assault variable, and we had more female responses overall.

The expected counts are all greater than 5, which means we don't have any problems with our data. This is an assumption of the chi-square that has to be met.

If you don't have expected counts in your input, did you select cells -> expected counts from the crosstabs screen?

The chi-square tells us if the association is significant between our variables. The green box shows what we are interested in, which is the *Pearson Chi-Square*. In this case the p value (*Asymptotic Significance (2-sided)*) is greater than 0.05, which means our test is not significant and no association between gender and victim of burglary/assault was found.

## Challenge 1: Run a Chi-Square test

We are interested in testing if subjects who have, and have not, been burgled or assaulted in the last five years have a different attitude in how good a job the police are doing.

Run a Chi-Square test using the `brghmef` (*how often worry about your home being burgled*) and `crmvct` (*respondent or household member victim of burglary/assault*) variables.

1. Open the crosstabs page and add `brghmef` as rows and `crmvct` as columns
2. Tick the *display clustered bar charts* option
3. Select the chi-square option
4. Add observed and expected counts and row and column percentages
5. Run the test - how would you interpret the result?

## Notes on significance testing

---

You will have noticed that with the chi-squared test we just performed we were able to determine if there was, or was not, an association between two variables. This is known as significance testing.

In order to test significance we must first have a hypothesis. This is the process of taking a theory, and turning it into something we can do a specific test or make a prediction on. This is known as hypothesis testing.

Hypothesis testing has two levels, the *research* or *normal* hypothesis and *statistical* hypothesis. Research hypothesis are ideas/suppositions/conjectures that motivate the research; think of these as the larger overall question you want to ask. Statistical hypothesis involve adjusting the research hypothesis so it can be tested using statistical techniques; these are more focused smaller questions. We will focus for now on *statistical* hypothesis

testing.

There are several general steps to hypothesis testing:

- 1) State your hypothesis
- 2) Collect data
- 3) Perform statistical test or tests
- 4) Decide to reject or fail to reject the null hypothesis
- 5) Present/report findings

When stating a statistical hypothesis we always have two, a null ( $H_0$ ) and alternate hypothesis ( $H_a$ ). The alternate hypothesis is what you predict *will* happen, the null hypothesis is what you predict *will not* happen.

For example, if I wanted to test out if it rained more days in the UK than Australia, based on your knowledge of UK and Australian weather you might form the following hypothesis:

- $H_a$ : The UK has, on average, more rainfall than Australia
- $H_0$ : There will be no difference, on average, in rainfall between the UK and Australia

In order to find out if we can reject or fail to reject the null hypothesis, we run statistical tests. We use the p value to determine this. The cut off for rejecting the null hypothesis is 0.05, so if your p-value is less than or equal to 0.05 you can reject your null hypothesis.

The idea with the 0.05 cut off is that there is a less than 5% chance that the results you are seeing are a result of chance.

If we did a statistical test on our rainfall example and found a p-value of 0.01 we could reject the null hypothesis as our test is below the 0.05 cutoff.

## Challenge 2: Make a hypothesis

As we already have the data we will make a hypothesis based on the variables we have. The ESS dataset we are using is survey data with various issues such as politics and policing.

We will use the below to variables to make a hypothesis:

- Respondent or household member victim of burglary/assault last 5 years (crmvt)  
*(categorical/nominal)*
- How successful police are at preventing crimes in country (plcpvcr) *(ordinal)*

Using these variables, what do you predict the difference, on average, will be between victims and non victims of burglary/assault and their attitudes to how successful the police

are at preventing crimes?

Your hypothesis:

- My null hypothesis ( $H_0$ ) is...
- My alternate hypothesis ( $H_a$ ) is...

We will take your null and alternate hypothesis to the next challenge and test it out!

## T-Tests

---

In the last challenge we made a hypothesis and null hypothesis, we will test this out in the challenge. First, we need to learn how to run a t-test in SPSS!

T-tests are for testing if the average (mean) difference between two variables is significant. There are two main options for running t-tests, the independent samples or paired-samples t-test.

The independent samples t-test is used to determine if two means collected from independent samples differ significantly. It is used for a *between-groups* study design, which entails different people being tested on each condition just once.

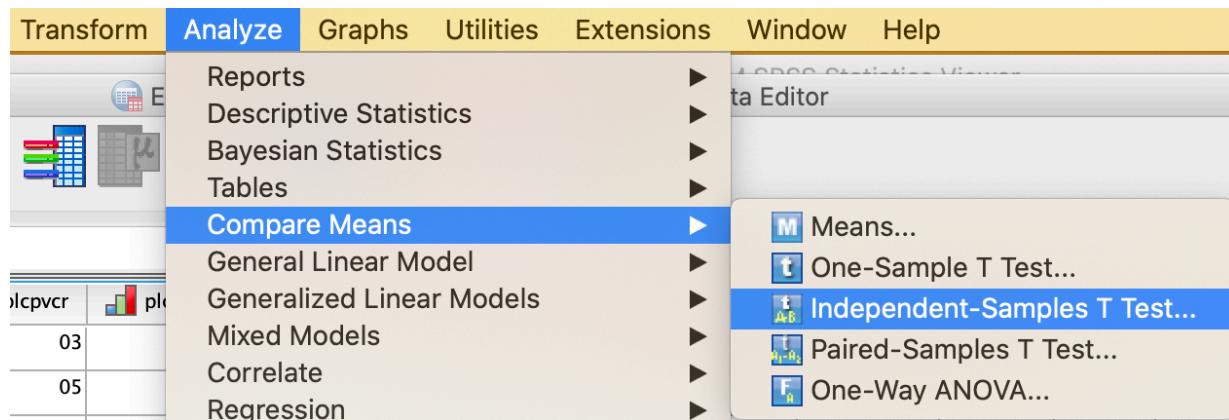
The paired-samples t-test is used to determine if two means from the same sample differ significantly. It is used for a *within-groups* study design, which entails the same person/people doing tests for all conditions.

For example, if we wanted to compare the average rating of two courses, we could design our study two ways:

- *Between-subjects*: each participant rates the just one of the courses
- *Within-subjects*: each participant rates both of the courses

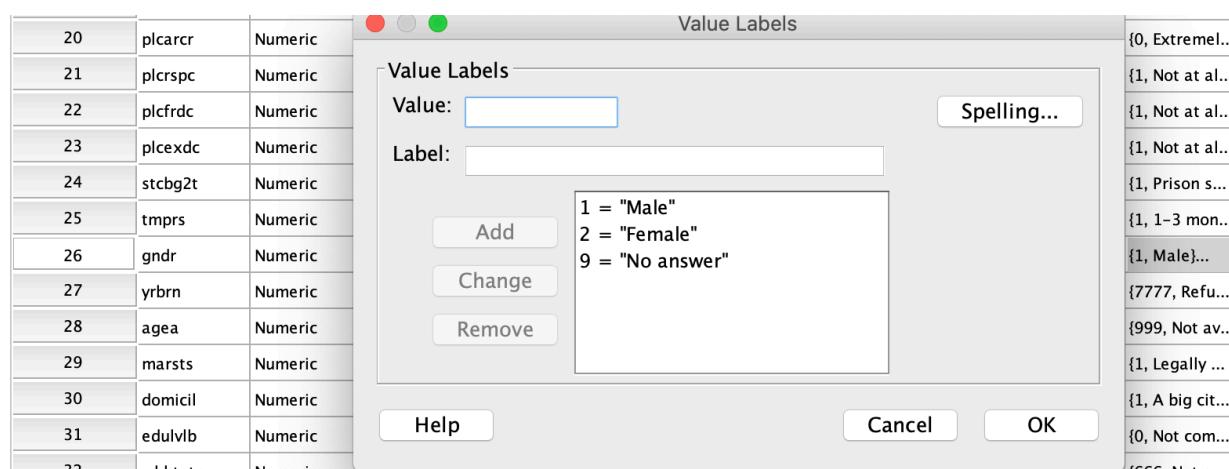
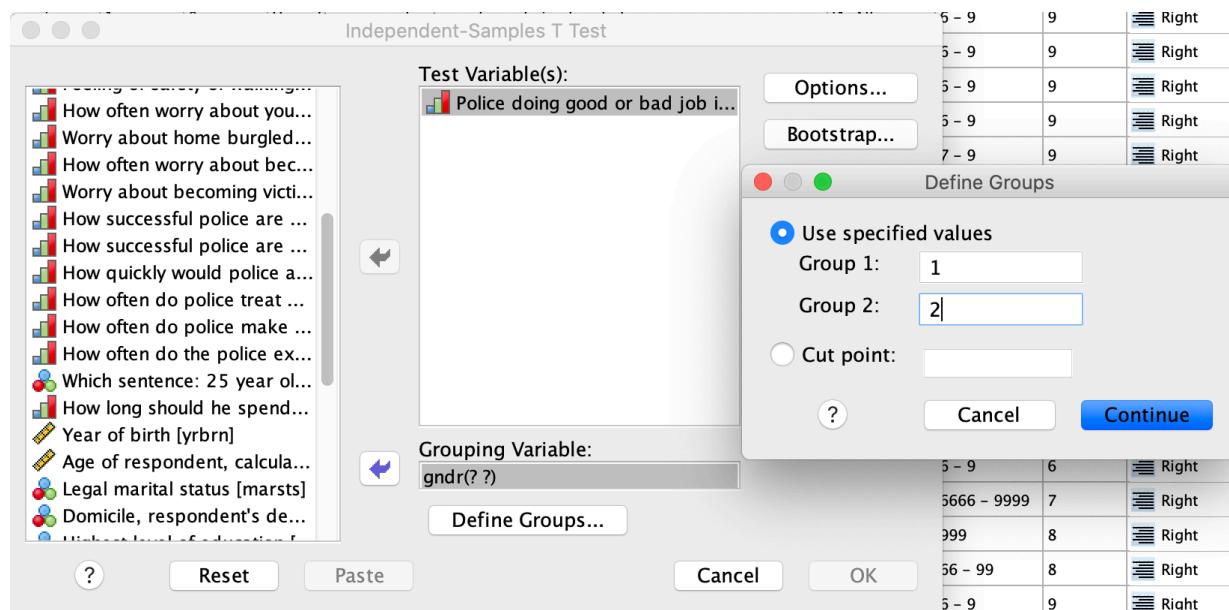
The ESS dataset we are using allows us to use both of these designs. We can use an independent samples t-test to compare two groups response to a question. Or we can use a paired-samples t-test to compare two survey responses.

We will first have a go at running an independent samples t-test, the next chapter we will try the paired-samples t-test. To run an independent samples t-test in SPSS we select the compare means menu from analyse, then find independent samples t-test.



The dialog box for independent samples t-test should appear. The test variable is the variable we want to compare the two means on, we have used

`Police doing good or bad job`. The grouping variable is a nominal (categorical) variable, ideally with two groups (or you select two groups), we have used `gender`. SPSS needs you to *define the groups* before you can run the test; for gender they are labeled 1 for male and 2 for female. We can find this information in the variable view from the variables value labels. See the images below on how to do this:

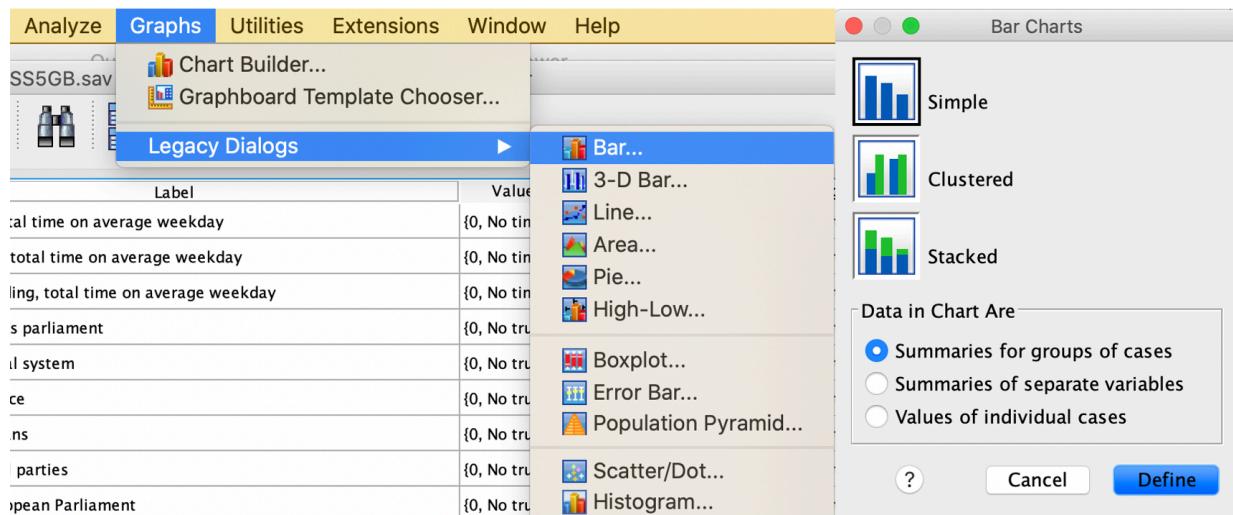


Once you've defined the groups you can run your t-test! You'll get an output with group statistics, which is the descriptive statistics, such as mean and standard deviation for your groups, and the t-test output. The green box in the image below indicates the part of the t-test output we are most interested in. In this case, the *Sig. (2-tailed)* is our p-value, which is showing 0.000. This tells us there is a significant difference between someone's perception of how good a job the police are doing depending on their gender.

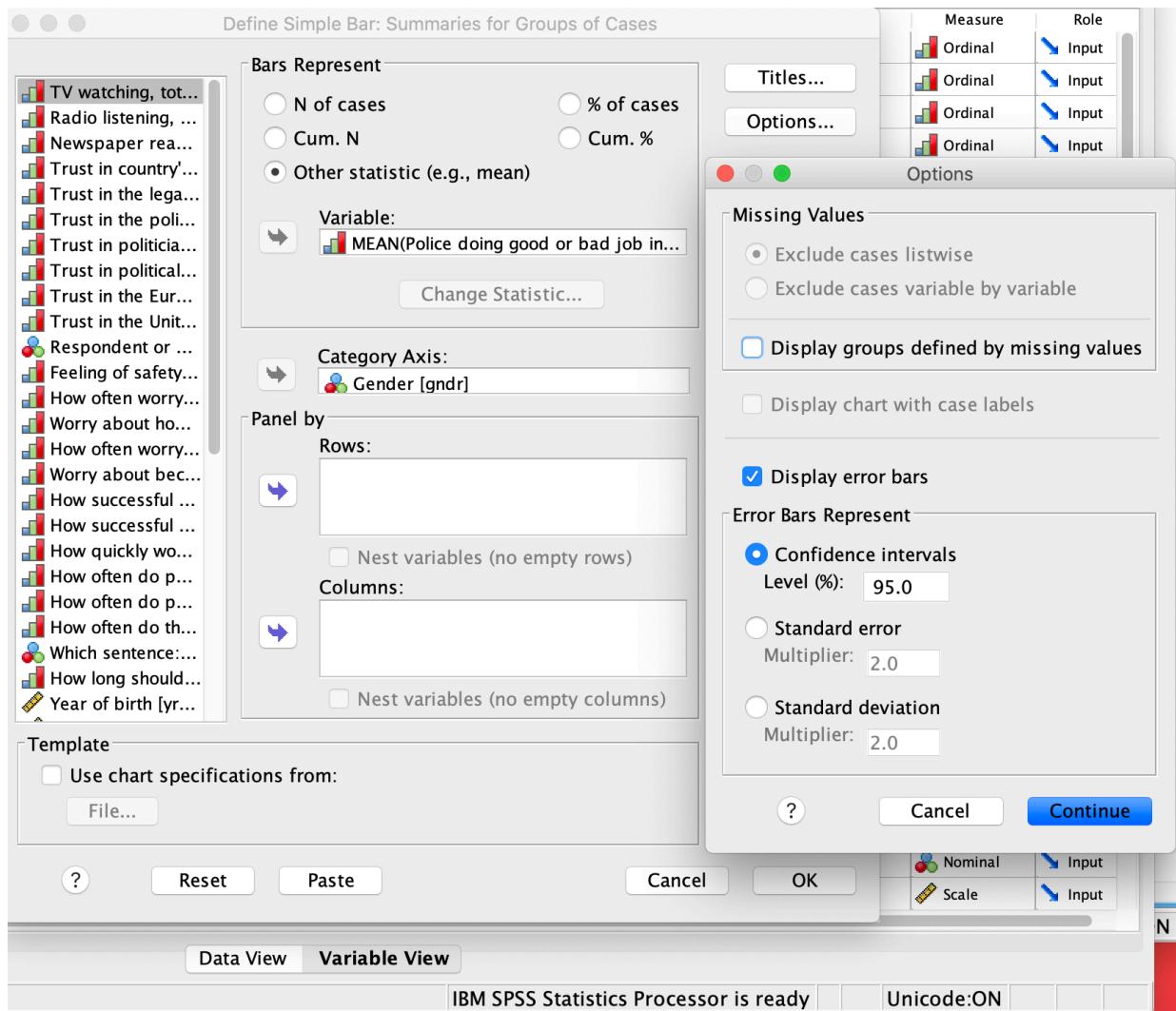
### T-Test

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval for the Difference
Police doing good or bad job in country	Equal variances assumed	18.175	.000	3.783	2408	.000	.118	.031	.057
	Equal variances not assumed			3.753	2190.253	.000	.118	.032	.057

As a bonus, it can be helpful to visualise the mean values with error bars. We first start by making a bar chart.



Once in the bar chart menu, we can add error bars using options. See the image below:



When you select OK, you should get a nice bar chart with error bars.

### Challenge 3: Run an independent samples t-test (based on your hypothesis from challenge 2)

Using the below variables we used to define our hypothesis, run an independent samples t-test in SPSS:

- Respondent or household member victim of burglary/assault last 5 years (crmvtc) (*categorical/nominal*)
- How successful police are at preventing crimes in country (plcpvcr) (*ordinal*)

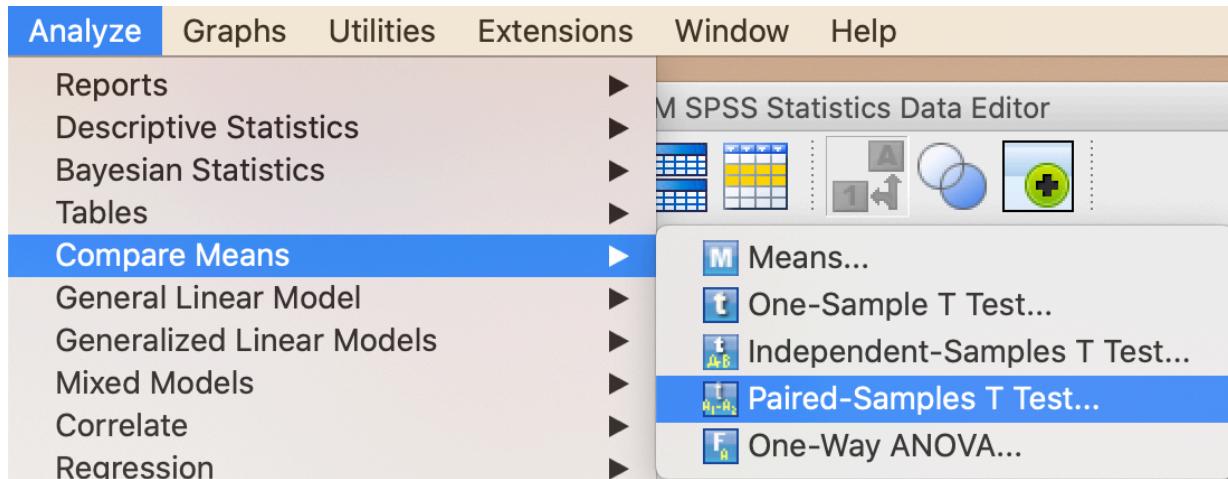
- 1) Make a bar plot of your two variables, with error bars
- 2) Run an independent samples t-test
- 3) Review your output. Was the result significant, and did it match your expectations and hypothesis?

## Paired Samples t-test

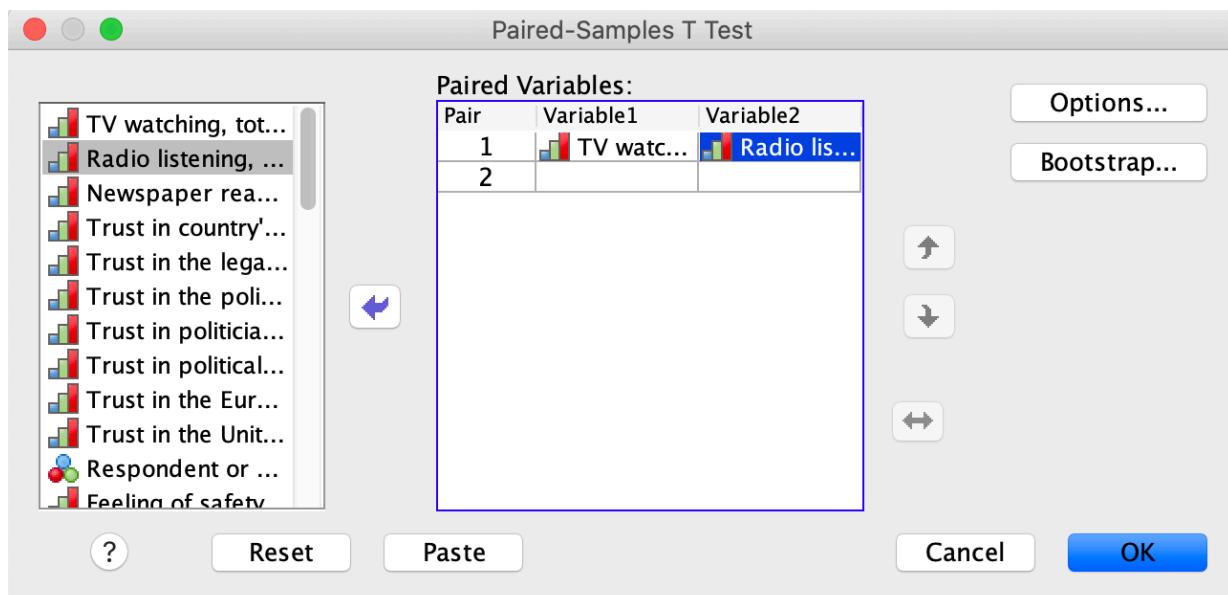
As explained in the previous chapter the paired-samples t-test is used to determine if two means from the same sample differ significantly. We can look too test if there is a difference in survey responses to two related questions.

In this example, we are interested to see if there is a difference in respondents tv watching and radio listening habits. We might suspect that people watch more tv than radio, given all the exciting shows in tv.

To run a paired-samples t-test in SPSS, we go to the analyse menu -> compare means -> paired-samples t test.



This will take you to the paired-samples T test menu. We can test out many pairs of variables, but in this example we will just test one pair: TV watching and Radio listening. You can either drag and drop the variables, or move with the arrow. Once you've moved the variables as per the image below, select *OK* to run the test.



The output you get has three tables. The first provides some descriptive statistics of our variables, in this example we can see average TV watching is 4.97 hours and average radio

listening is 2.85 hours. The second shows correlation between the variables; correlation will be covered in workshop 4! The final table is the t-test, the green box shows us the part we are most interested in, which shows us there is a significant difference between average TV watching and radio listening in respondents. This suggests people who took this survey, on average, spend more time watching TV than listening to the radio.

### → T-Test

**Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 TV watching, total time on average weekday	4.97	2417	2.014	.041
Radio listening, total time on average weekday	2.85	2417	2.617	.053

**Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 TV watching, total time on average weekday & Radio listening, total time on average weekday	2417	-.039	.053

**Paired Samples Test**

	Mean	Std. Deviation	Paired Differences		95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
			Std. Error Mean	Lower	Upper				
Pair 1 TV watching, total time on average weekday - Radio listening, total time on average weekday	2.117	3.365	.068	1.982	2.251	30.928	2416	.000	

## Challenge 4: Run a paired samples t-test

Let's say we are interested in understanding if there is a difference in respondents trust in country's parliament and trust in politicians. We might suspect that people look more favourably on the government than politicians in general.

Our hypothesis might be:

$H_a$ : There will be a difference, on average, in trust between parliament and politicians

$H_0$ : There will be no difference, on average, in trust between parliament and politicians

- 1) Make a bar plot with the mean response of both your variables, making sure to add error bars. *hint: try using the "summaries of separate variables" option*
- 2) Run a paired-samples t-test using the `Trust in country's parliament` and `Trust in politicians` variables
- 3) Review your output. Was the result significant? Did it match up with your expectations and the hypothesis?

# One-way ANOVA

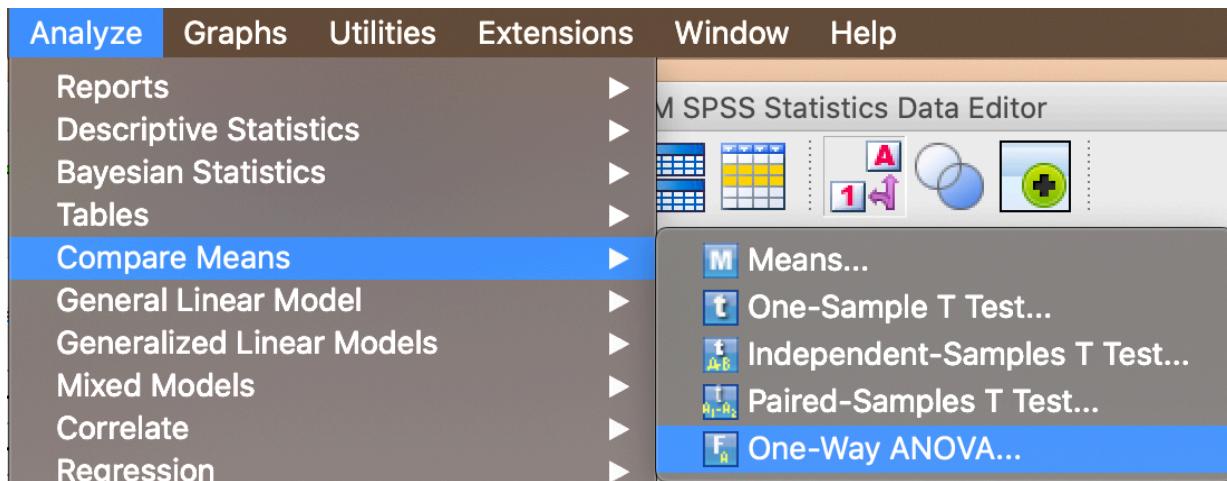
The ANOVA, short for *analysis of variance*, is a statistical technique for testing if the average (means) of three or more variables are different.

Just like the t-test, there are two basic tests we can use: the one-way ANOVA and the repeated measures ANOVA. The one-way ANOVA suits a between-groups design, such as the different responses to a question across three or more groups. The repeated measures ANOVA suits a within-groups design, such as the mean response of three different questions by the same group.

We will cover repeated measures ANOVA in the next section, for now we will look at the one-way ANOVA.

In this example we are interested in finding out if different legal marital statuses of respondents (never married, widowed, married etc.) have different mean weekly TV watching time. One might suggest that we would see differences in those groups. We will use the `TV watching, total time on average weekday` as our dependent variable (what we will measure on), and the `Legal marital status` variable as our categorical variable.

To run a one-way ANOVA we start by going to analyze -> compare means -> one-way ANOVA.



When the one-way ANOVA dialog box appears, we first move the variables we want to test on. TV watching goes to dependent list, and legal marital status goes to factor.

We will want to add some extra elements to our one-way ANOVA; see images below. First we select *options*, and tick the descriptive, homogeneity, and means plot. Then we want to select *Post hoc* and select *Tukey*; post hoc tests help us to find out if we have a significant result, where this difference occurred.

Now select ok to run the one-way ANOVA, and you should get several tables in your output viewer. In the image below we have highlighted two key things to look at. The green box shows us the ANOVA result, which has a *sig* (p value) less than 0.000, so very significant. The post hoc tests show us the interactions between each of the groups (like a t-test). The blue box shows us a significant post hoc test, suggesting our significant ANOVA result was, in part, driven by the difference mean TV watching time between those who are legally married and those that are widowed.

## ANOVA

TV watching, total time on average weekday

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	181.901	5	36.380	8.697	.000
Within Groups	5772.406	1380	4.183		
Total	5954.307	1385			

## Post Hoc Tests

### Multiple Comparisons

Dependent Variable: TV watching, total time on average weekday

Tukey HSD

(I) Legal marital status	(J) Legal marital status	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Legally married	In a legally registered civil union	-.674	.758	.949	-2.84	1.49
	Legally separated	-.140	.382	.999	-1.23	.95
	Legally divorced/civil union dissolved	-.621	.253	.140	-1.34	.10
	Widowed/civil partner died	-1.193*	.259	.000	-1.93	-.45
	None of these (NEVER married or in legally registered civil union)	-.306	.239	.796	-.99	.38
In a legally registered civil union	Legally married	.674	.758	.949	-1.49	2.84
	Legally separated	.534	.786	.984	-1.71	2.78
	Legally divorced/civil union dissolved	.053	.732	1.000	-2.04	2.14
	Widowed/civil partner died	-.519	.734	.981	-2.61	1.58
	None of these (NEVER married or in legally registered civil union)	.368	.727	.996	-1.71	2.44

Have a look at the rest of the post hoc tests to see if you can see any other significant interactions.

---

In our example, our assumptions for the ANOVA were not met as the Levene's test was significant (see the *Test of Homogeneity of Variances* table). This means there is unequal variance between our groups or samples, so our results of the ANOVA are not trustworthy. In this instance there are other options we should use, which we don't have time to cover in this workshop. There is a [good guide available](#) which should help, if when running an ANOVA, your Levene's test is significant.

## Challenge 5: Run an One-way ANOVA

We want to expand our previous hypothesis from challenge 2 about how successful the police are at preventing crimes. We've seen there is a significant difference between respondents who have been victims of assault/burglary. We might be interested to see if those that have been victims of assault/burglary have differences in responses to what sentence to give a burglar, and their perception of how good the police are at preventing

crimes.

Our hypothesis might be:

$H_a$ : There will be a difference, on average, on perception of how successful the police are, and determining a prison sentence for a burglar by respondents who have been victims of assault/burglary

$H_0$ : There will be no difference, on average, on perception of how successful the police are, and determining a prison sentence for a burglar by respondents who have been victims of assault/burglary

- 1) Select only cases that responded yes to the *been a victim of assault/burglary* variable
- 2) Run a one-way ANOVA using the

how successful are the police are at preventing crimes as your dependent, and

which sentence: 25 year old male, house burglary, second time (stcbg2t) as your factor

- 3) Be sure to add a post hoc test, descriptives, means plot and a test of homogeneity
- 4) Review your output, do you see a significant difference?
- 5) Tell SPSS to select all cases again if you have not removed the filter already

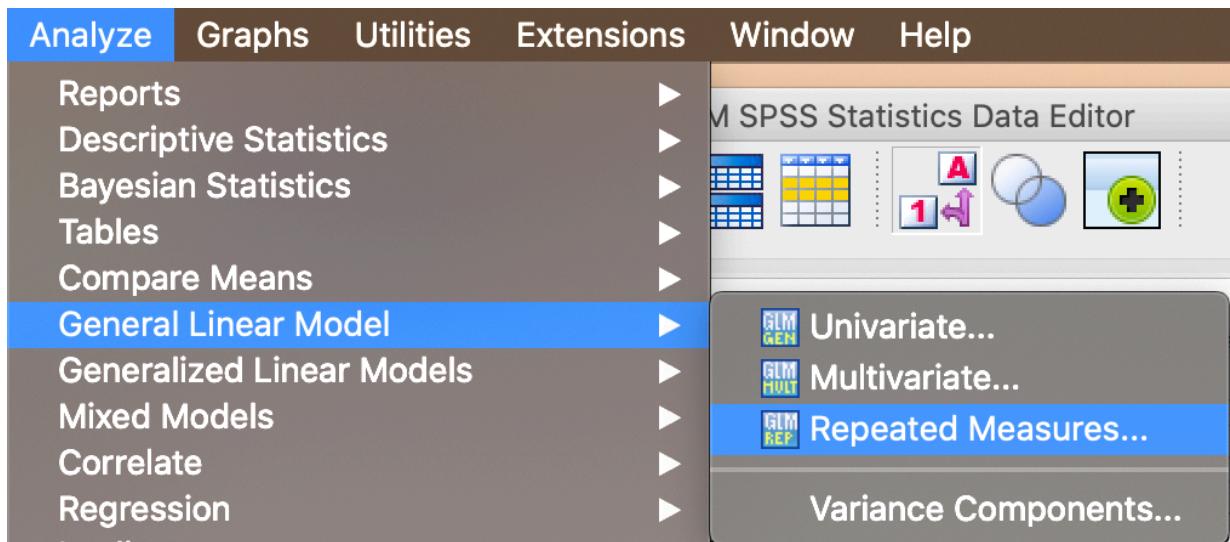
## Repeated measures ANOVA

---

Our final inferential test we will cover in this workshop will be the repeated measures ANOVA. Similar to the paired-samples t-test, we can compare average responses across three or more variables.

For this example, we can expand our paired-samples t-test example. We want to test if there is a difference between respondents average TV watching, radio listening, and newspaper reading. One might guess that we would see a difference as TV seems to be a more popular media than newspaper or radio.

To run a repeated measures ANOVA we go to analyze -> general linear model -> repeated measures.



A dialog box will appear and ask you to define your factors. This is asking you to tell SPSS how many related variables you will be comparing. In this case, we have made a factor called habits, with three levels (we have three variables). Select *add* once you've add the information, and then define.



## Repeated Measures Define Factor(s)

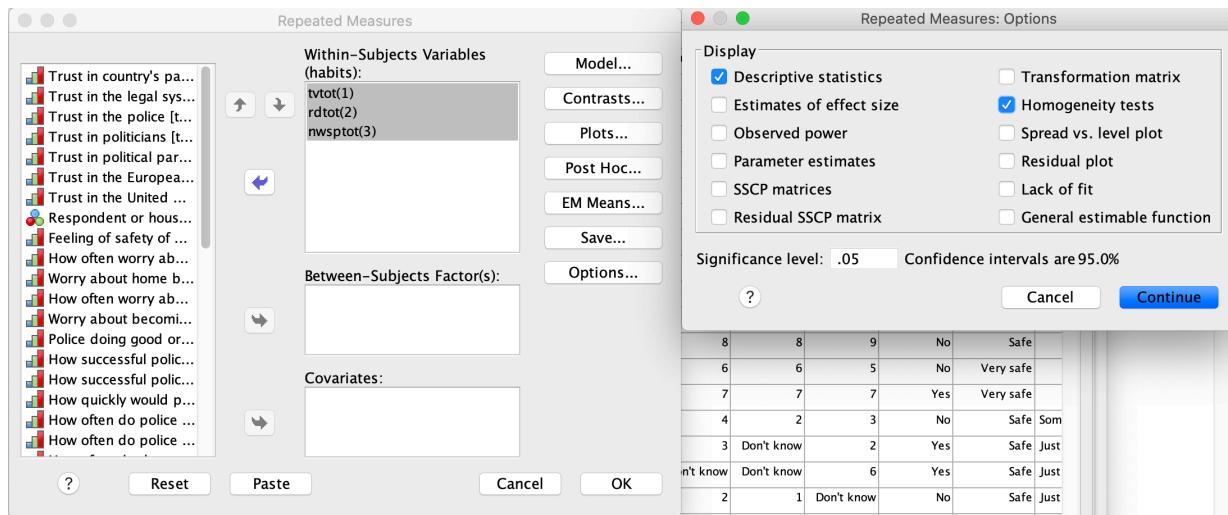
Within-Subject Factor Name:

Number of Levels:

habits(3)

Measure Name:

We now get a familiar dialog box, this time for repeated measures ANOVA. Add the TV, radio, and newspaper variables as your within-subject factors. In options, select descriptives and homogeneity tests. Select ok and run your repeated measures ANOVA.



You will see several tables appear, the two you are most interested in are shown in the image below. You will notice there are several tests, shown in the green box in the image below; to know which test to run, check out [this useful flow chart](#).

Using this flow chart, we would use the huynh-feldt test. In this particular case, all tests are significant, telling us there is a significant difference between respondents TV, reading, and newspaper habits.

### Mauchly's Test of Sphericity<sup>a</sup>

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Epsilon <sup>b</sup>	Lower-bound
habits	.837	430.283	2	.000	.860	.860	.860	.500

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: habits

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.

### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	
habits	Sphericity Assumed	16647.712	2	8323.856	1979.721	.000	
	Greenhouse-Geisser	16647.712	1.719	9683.343	1979.721	.000	
	Huynh-Feldt	16647.712	1.720	9677.101	1979.721	.000	
		Lower-bound	16647.712	1.000	16647.712	1979.721	.000
Error(habits)	Sphericity Assumed	20299.621	4828	4.205			
	Greenhouse-Geisser	20299.621	4150.176	4.891			
	Huynh-Feldt	20299.621	4152.853	4.888			
	Lower-bound	20299.621	2414.000	8.409			

## Challenge 6: Run a repeated measures ANOVA

For the sixth challenge we will expand our question from challenge 4, and look at the differences in respondents trust of the political system using the following variables:

- Trust in country's parliament
- Trust in politicians
- Trust in political parties
- Trust in the European parliament

We might suspect that, on average, respondents have differing levels of trust for each of these political elements.

Our hypothesis might be:

$H_a$ : There will be a difference, on average, in peoples trust in different elements of the political system  $H_0$ : There will be no difference, on average, in peoples trust in different elements of the political system

- 1) Make a bar plot with the mean response of all your variables, making sure to add error bars. *hint: try using the "summaries of separate variables" option*
  - 3) Run a repeated measures ANOVA with your four variables. When defining your *within\_subject factor* be sure to add 4 levels and a sensible name, like politics. Finally, add the descriptives and homogeneity tests options.
  - 3) Review your output. Was the result significant? Did it match up with your expectations and the hypothesis?
- 

*Note we did not have time to fully explain ANOVAs, how they work, and how they are related to linear models (which we will cover in SPSS session 4).* You can read more about ANOVAs [here](#) if you are interested.

## Final challenge - complete the course survey

---

[Click this link to complete the survey](#)