Assignment 2
Statistical Analysis using SPSS

HSA Tutorial
21st November 2017
20.30-22.00

Overview

- Tests and Research Questions
- Comparing means: Independent subjects t-test
- Comparing means: Paired samples t-test
- Categorical data: Crosstabs and chi-square tests
- Comparing multiple means: ANOVA

Inferential statistics

- Use the data from the sample to make inferences (assumptions) about the population
- Is this pattern a real difference?
- Is it just due to chance?
- What conclusions can we draw from our findings?

Inferential statistics

Differences	Relationships
T-test Difference between **2 groups** (mean scores on **continuous** variable) RM or IS	Pearson Correlation Test the strength of a relationship between **2 continuous** variables
1-way Analysis of Variance Difference between **2 or more groups** Impact of 1 IV on 1 DV across groups	Multiple Regression **Multiple independent** variables **1 continuous** dependent variable
2-way Analysis of Variance Impact of 2 IV's on one DV Interaction Effect	Factor Analysis Multiple variables (e.g. Scale items)

Parametric vs. non-parametric

Parametric	Non-parametric

▪ Conventional Statistical Tests (e.g. t-test; ANOVA)
▪ Assumptions
  ▪ Normal Distribution
  ▪ Low Variance
  ▪ Mean is adequate representation
  ▪ Representative of sample
▪ Each test has specific assumptions to be met
	▪ Make no assumptions about the distribution of the data
	  • Median as most accurate measure of central tendency
	▪ More 'robust' than parametric tests
	▪ BUT less power
	▪ e.g. Mann and Whitney; Kruskal-Wallis

Power

- Power (1-β): Likelihood of rejecting H_o when it is false
  - Alpha (α): Lower levels of alpha makes it harder to reject the null hypothesis
  - Effect Size (relationship between variables)
  - Sample Size (power increases with larger samples)

Tests of difference

Independent subjects t-test	
Independent variable:	1 categorical (binary)
Dependent variable:	1 continuous
Non-parametric equivalent:	Mann-Whitney U Test

Tests of difference

Repeated measures t-test	
Independent variable:	1 categorical (binary)
Dependent variable:	1 continuous
Non-parametric equivalent:	Wilcoxon

Tests of difference

1-way between groups ANOVA	
Independent variable:	1 categorical (3+ levels)
Dependent variable:	1 continuous
Non-parametric equivalent:	Kruskal-Wallis

Tests of difference

1-way repeated measures ANOVA	
Independent variable:	1 categorical (3+ levels)
Dependent variable:	1 continuous
Non-parametric equivalent:	Freidman Test

Tests of difference

2-way between groups ANOVA	
Independent variable:	2 categorical (2+ levels)
Dependent variable:	1 continuous
Non-parametric equivalent:	

Testing relationships

Chi-square	
Independent variable:	I categorical
Dependent variable:	I categorical
Non-parametric equivalent:	

Testing relationships

Pearson's correlation	
Independent variable:	2 continuous
Dependent variable:	-
Non-parametric equivalent:	Spearman (RHO)

Exercise 5: Choosing the right statistic

- Is there a difference between the typing speed of an iPad and a laptop computer, using the same participants?
- T-test (repeated measures)

- Do males or females have more positive attitudes to the Internet?
- T-test (independent subjects)

- Is there a relationship between age and frequency of use of the Internet?
- Pearson's correlation

- Is there a difference between those under 30, 30-50, and 50+ in terms of their use of the Internet?
- 1-way ANOVA

- Do those over 65 years and under 65 years differ in terms of owning a smartphone (yes/no)?
- Chi-square test

The t-test

- Compares the average (mean) scores between **2 groups** on some continuous variable
- Independent Variable: Categorical
- Dependent Variable: Continuous

Do users perform better on a typing test using a laptop or a tablet?

Types of t-tests

1. Independent-samples T-test
   - Compare mean scores of 2 *different groups or conditions*
2. Repeated Measures T-Test
   - Compare the *same group* on *different tests (matched pairs)*

Rationale

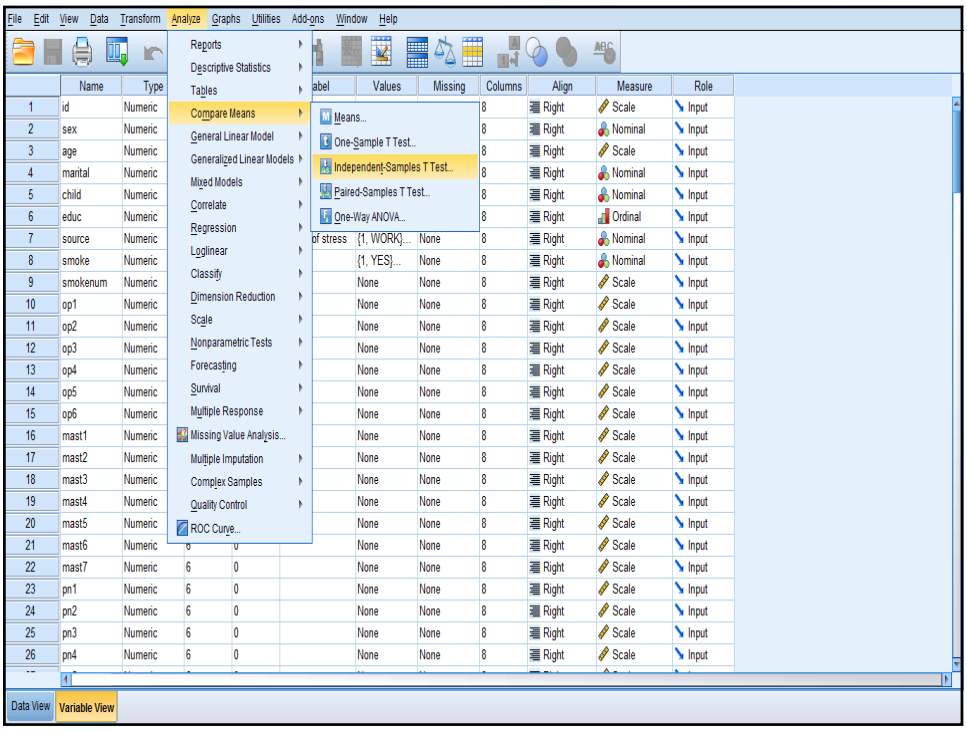
- Compares the sample means of collected data to sample means that would emerge if there was no effect
- The standard error is used to gauge the variability between the group means
  - Small SE = small difference
  - If difference in sample group is bigger than that of estimated we can assume that groups come from different populations
  - As the difference between the groups increase, the more **confidently** we can assume that it is due to the experimental manipulation

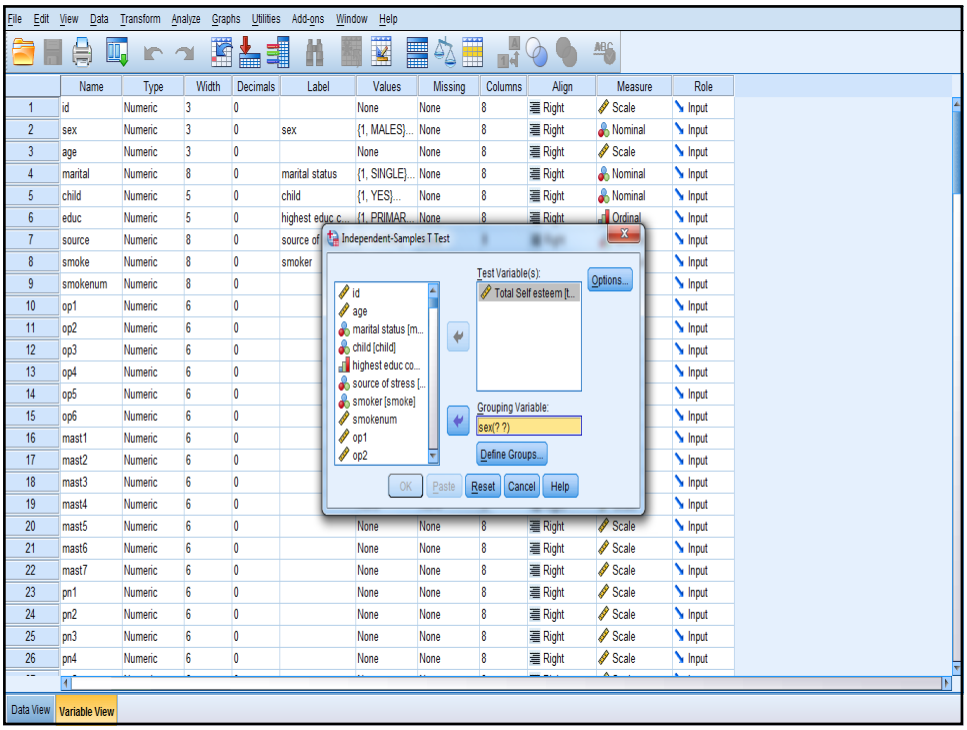
Data for Independent Subjects T-Test

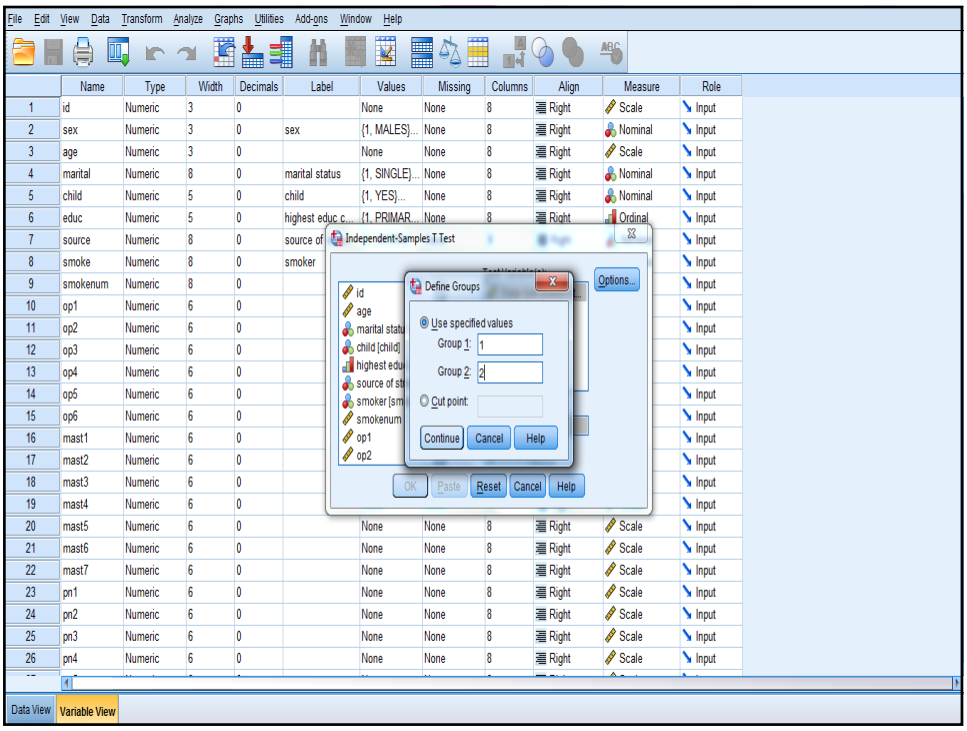
Participant	Group	Self-Esteem Score
1	1	30
2	1	35
3	1	35
4	1	30
5	2	50
6	2	55
7	2	65

Data for Paired Samples T-Test

Participant	Time 1	Time 2
1	30	40
2	35	35
3	45	50
4	50	55
5	65	55
6	40	50
7	35	60




File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Independent-Samples T Test
Test Variable(s):
Total Self esteem [t...
Options...
Grouping Variable:
sex(? ?)
Define Groups...
OK Paste Reset Cancel Help
Data View
Variable View


File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Independent-Samples T Test
Define Groups
Use specified values
Group 1: 1
Group 2: 2
Cut point:
Continue Cancel Help
Options...
OK Paste Reset Cancel Help
Data View
Variable View

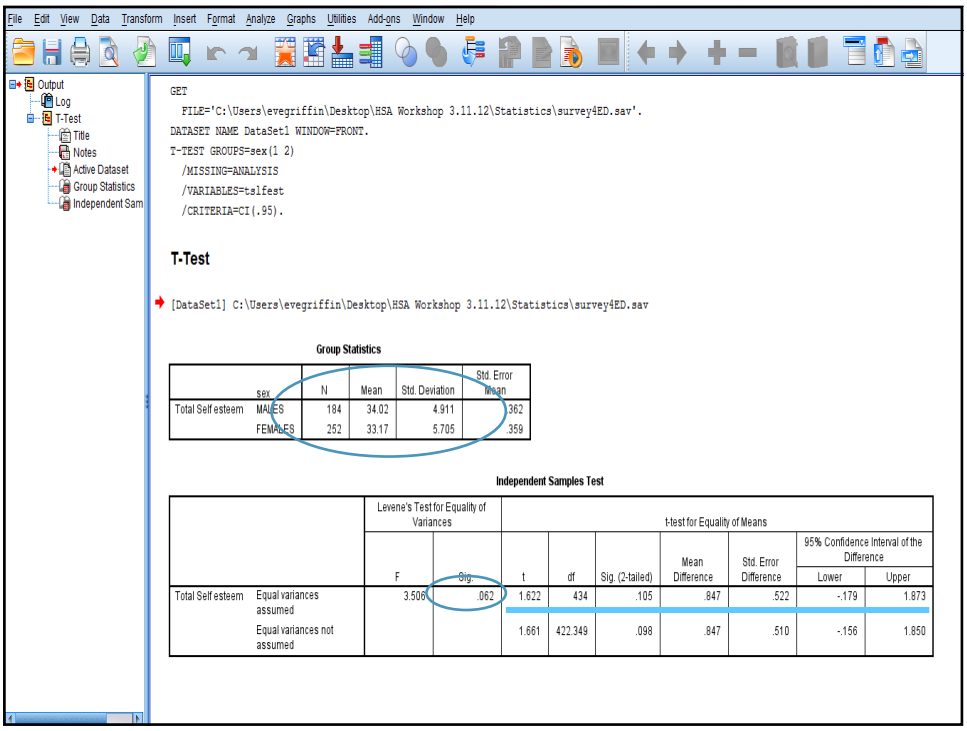
Output: Independent Samples T-Test

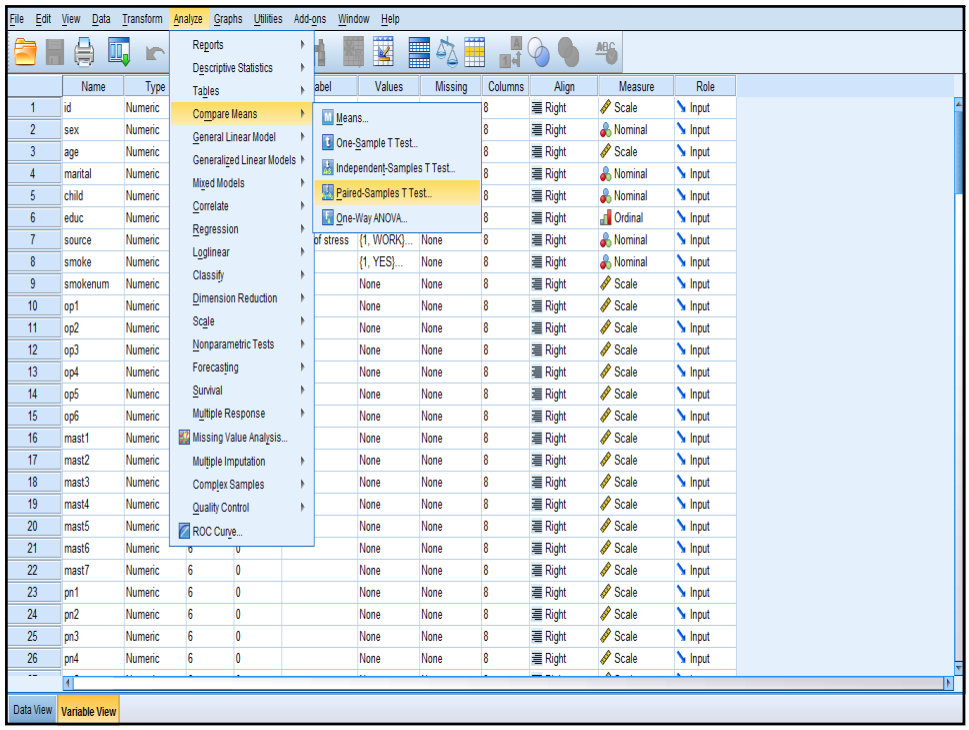
Step 1: Check Descriptives

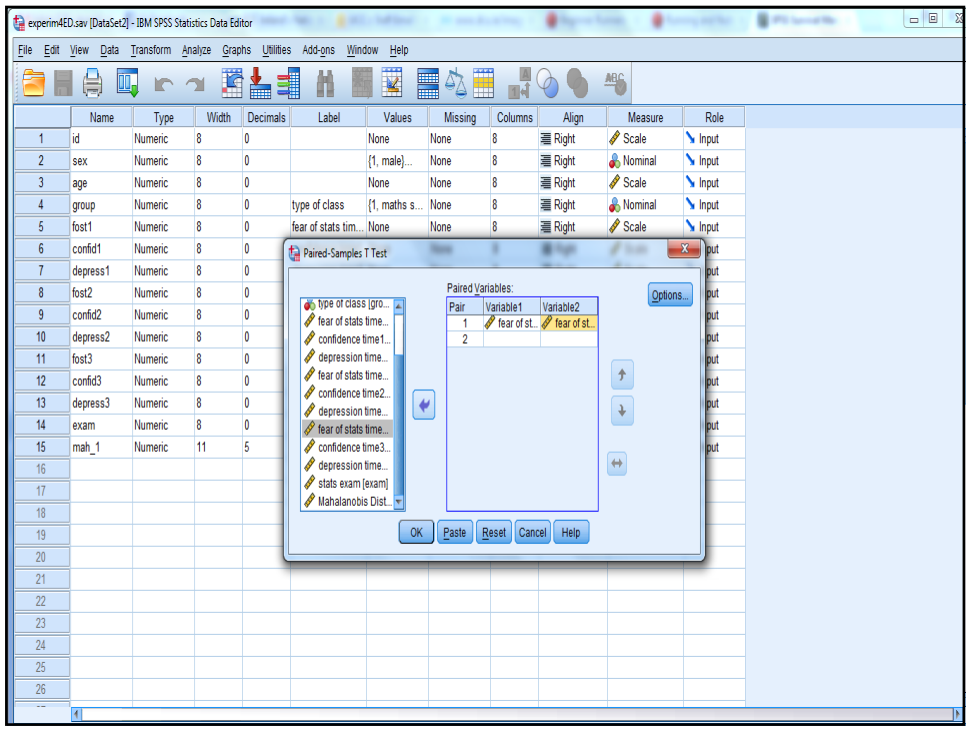
Step 2: Check Levene's Test

Step 3: Read t-test significance level

Step 4: Calculate Effect Size




File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control
ROC Curve...
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...
Name Type Values Missing Columns Align Measure Role
Data View
Variable View

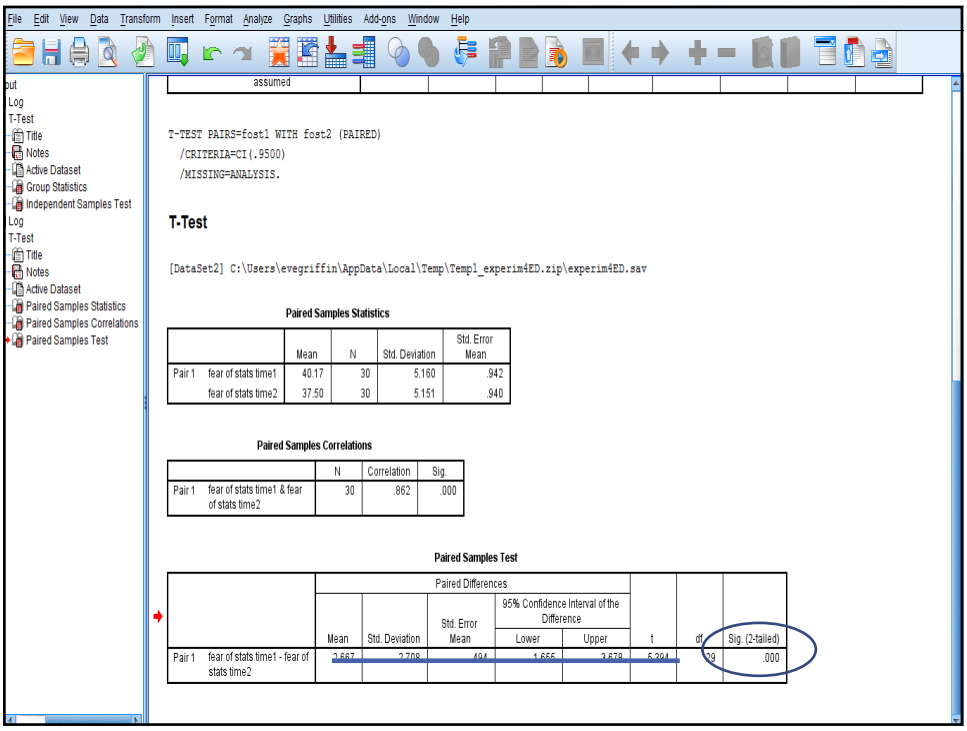

experim4ED.sav [DataSet2] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
Paired-Samples T Test
Paired Variables:
Pair Variable1 Variable2
Options...
OK Paste Reset Cancel Help

Output: Paired Samples T-Test

Step 1: Check Significance

Step 2: Compare Means

Step 3: Calculate Effect Size



Effect Size

- Relates to the **magnitude** of difference between scores
  - Are they really different? Or is it by chance?
- Eta Squared is a common statistic (0 to 1)
  - Represents the proportion of variance the DV that can be explained by the IV
- Use calculation in Pallant (2013) for t-tests

.01 = Small Effect Size
.06 = Moderate Effect Size
.14 = Large Effect Size

Non-Parametric Equivalents

- What if the data is non-normal?

1. Mann-Whitney U-Test
   - Analyses data based on ranks
2. Wilcoxon Signed Ranks Test
   - Similar to Mann-Whitney

ANOVAs: Rationale

- The fewer tests conducted, the better
- Is there a difference in Anxiety Scores across 3 age groups (16-25; 26-39; 40+)?
  - Compare 1 + 2; 2 + 3;1 + 3 using t-tests?
  - Greatly increase probability of committing a Type 1 Error
- Is there a difference in anxiety scores for males and females across 3 age groups?
  - Doing separate ANOVAs does not allow us to test for an interaction between IV's (i.e. gender and age)

ANOVAs

- Comparing several means
- 1-way Analysis of Variance
  - Between-groups

Is there a difference in statistics results across undergraduate classes?

Planned Comparisons/Post-Hoc Analyses

- ANOVA tells you **if there is a difference** but not **where it is**

1. **Planned Comparison (*a priori)***
   - Testing specific hypotheses
   - 1-tailed hypothesis (e.g. Difference between Group 1 and 2)
   - Need to be included beforehand

- **Post-Hoc (*a posteriori)***
  - Conduct a whole set of comparisons, exploring the differences between each of the groups/conditions in a study
  - Designed to guard against the possibility of a increased Type I error due to the large number of different comparisons being made

F-Statistic

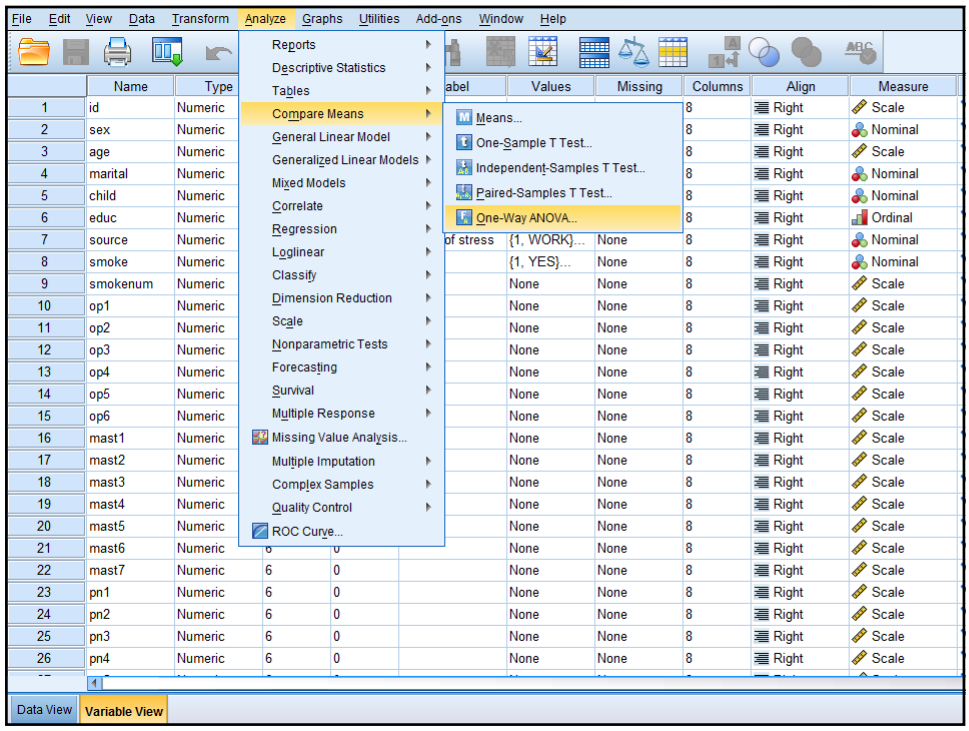
- Testing difference between 3 or more means
  - H_0 = all means are equal

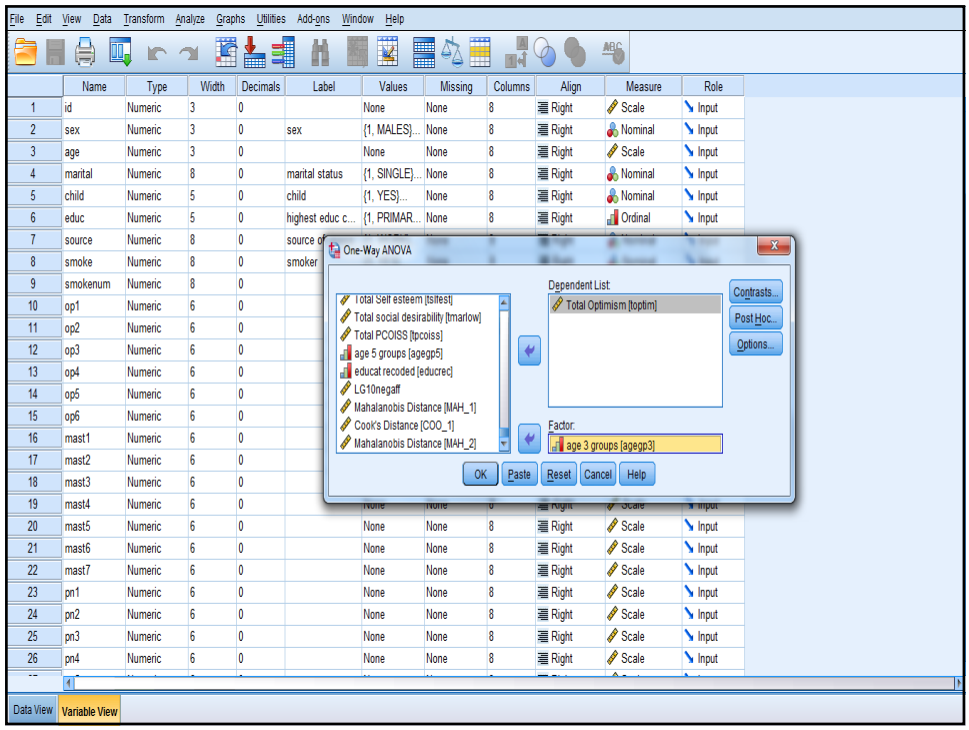
- *F-Statistic*: Compares amount of systematic variance in the data to the amount of unsystematic variance (Model : Error)

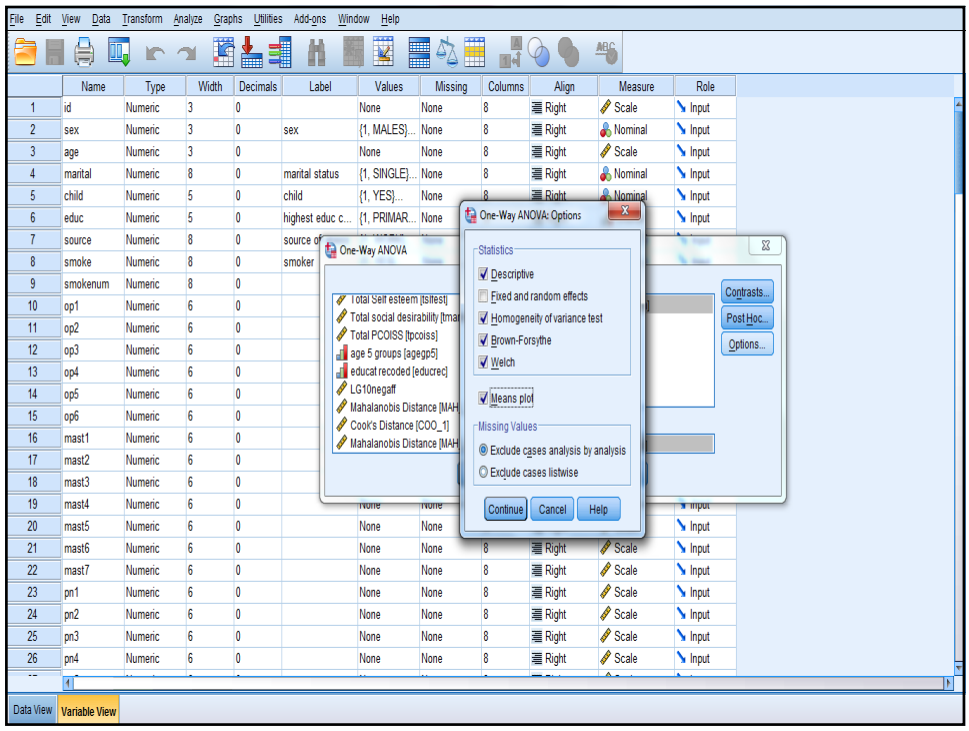
- Omnibus – tells us if there is an overall difference, but not where the difference lies

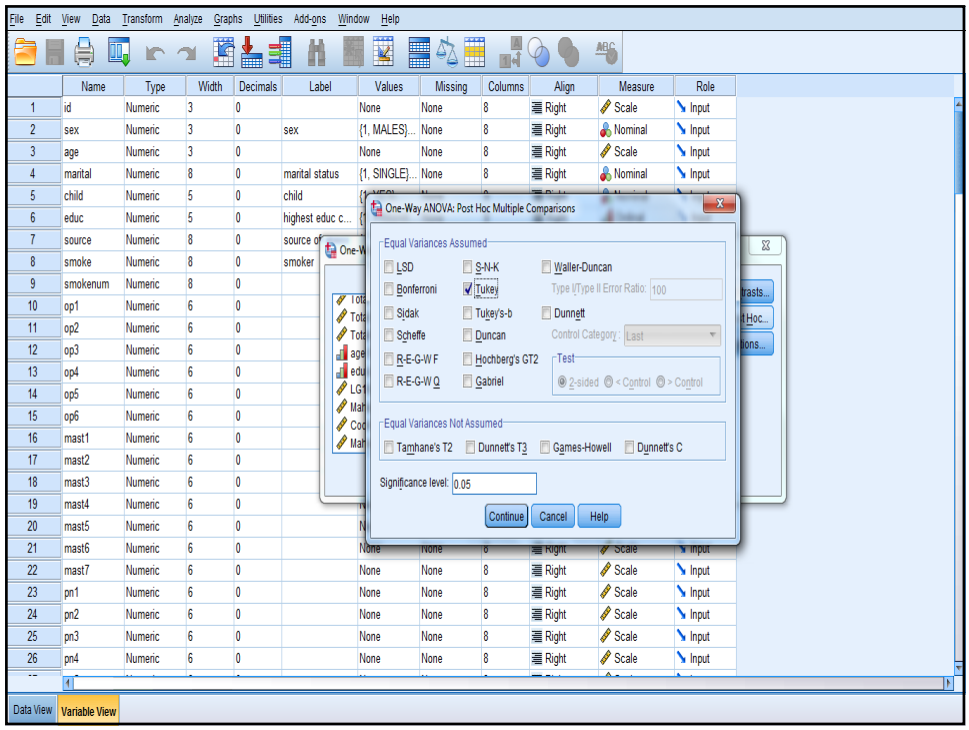
ANOVA Data (1-way between)

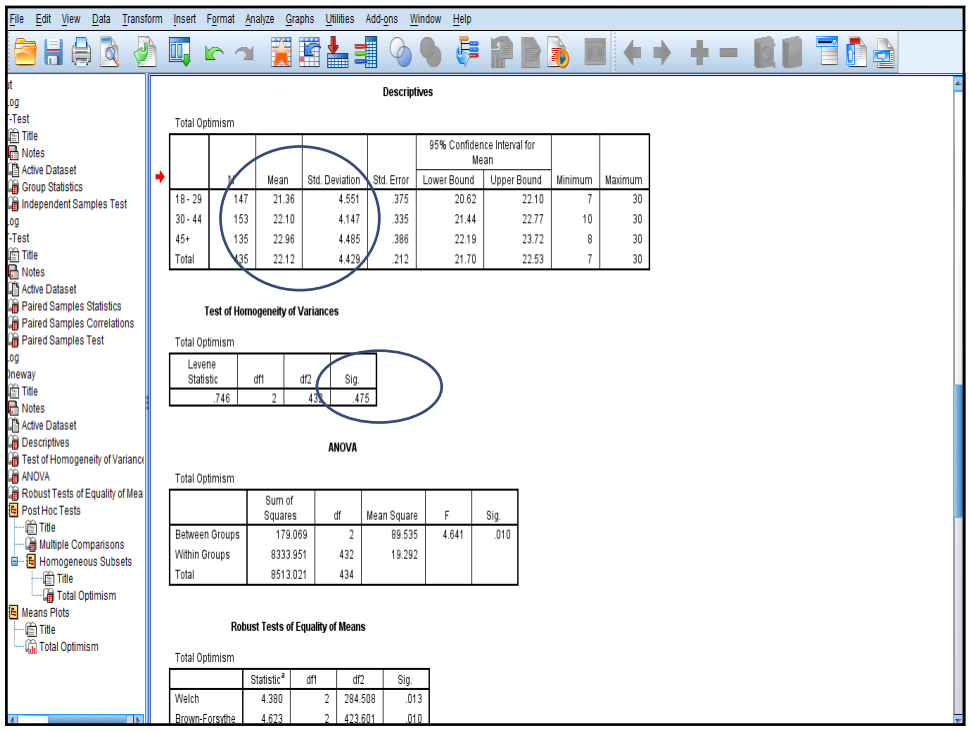
Total Optimism Scores	Age
30	1
45	2
55	3
31	1
78	1
23	2

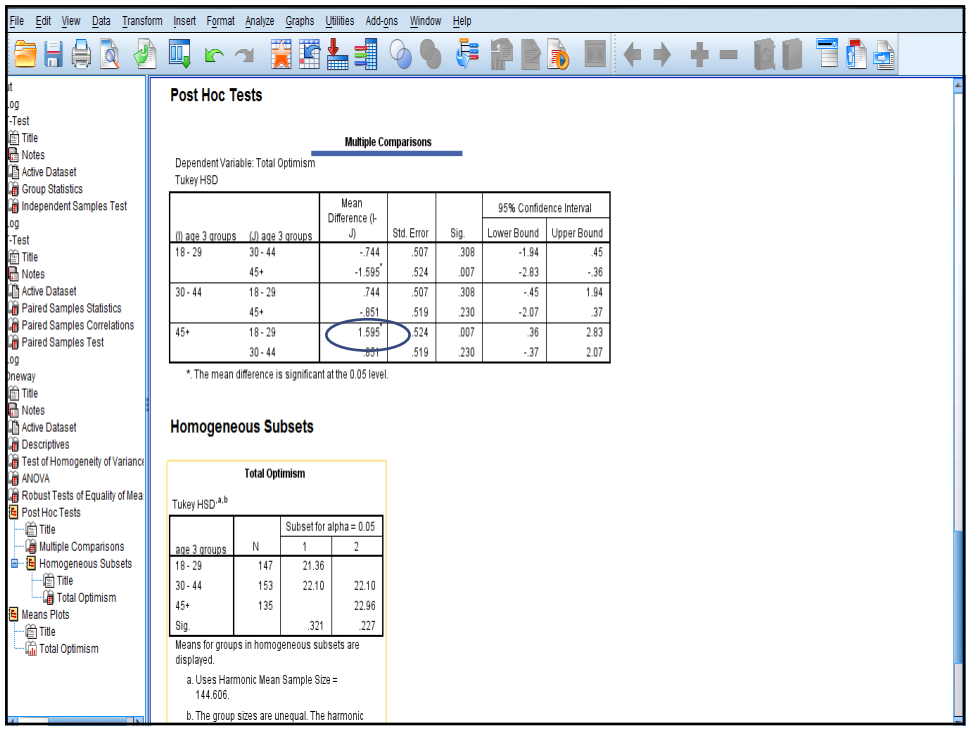



File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
One-Way ANOVA
Dependent List:
Total Optimism [toptim]
Factor:
age 3 groups [agegp3]
Contrasts...
Post Hoc...
Options...
OK Paste Reset Cancel Help
Data View Variable View


File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
One-Way ANOVA
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
Data View Variable View


File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure Role
One-Way ANOVA: Post Hoc Multiple Comparisons
Equal Variances Assumed
LSD
Bonferroni
Sidak
Scheffe
R-E-G-W F
R-E-G-W Q
S-N-K
Tukey
Tukey's-b
Duncan
Hochberg's GT2
Gabriel
Waller-Duncan
Dunnett
Dunnett's C
Games-Howell
Dunnett's T3
Tamhane's T2
Equal Variances Not Assumed
Test
2-sided
< Control
> Control
Control Category
Significance level: 0.05
Continue
Cancel
Help
Data View
Variable View

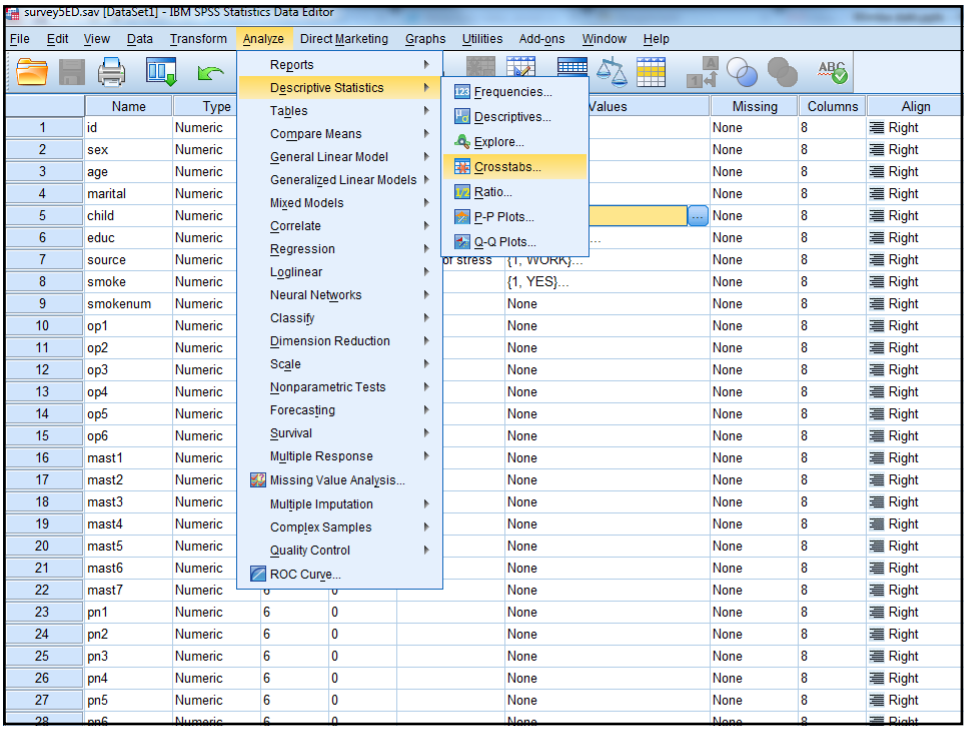

File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Descriptives
Total Optimism
95% Confidence Interval for Mean
Test of Homogeneity of Variances
Levene Statistic
ANOVA
Sum of Squares
Robust Tests of Equality of Means
Welch
Brown-Forsythe

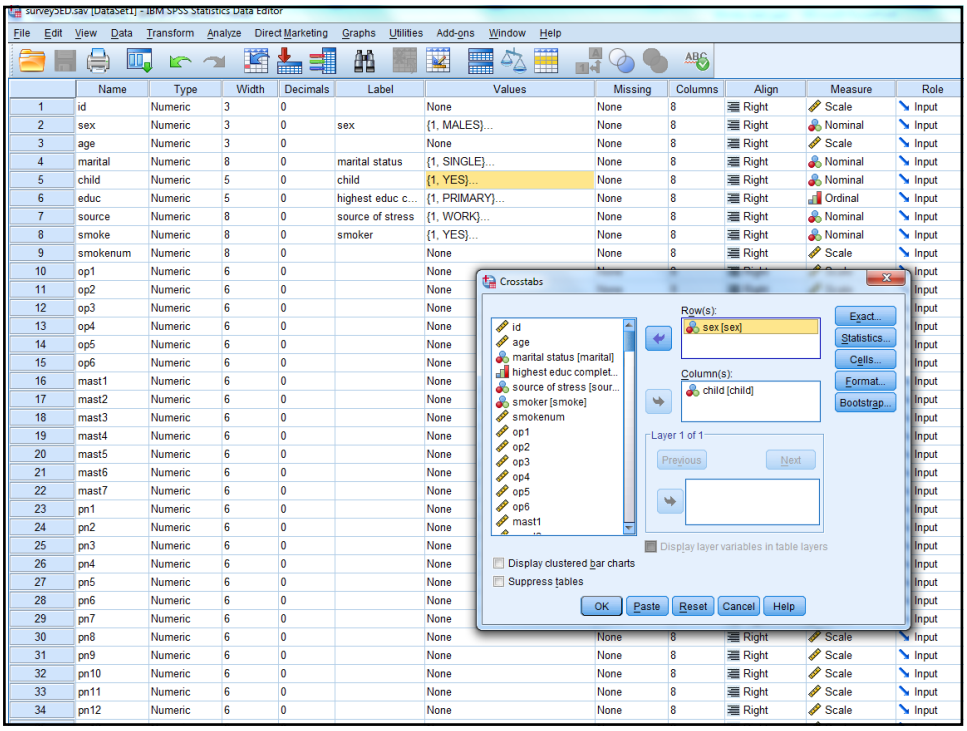


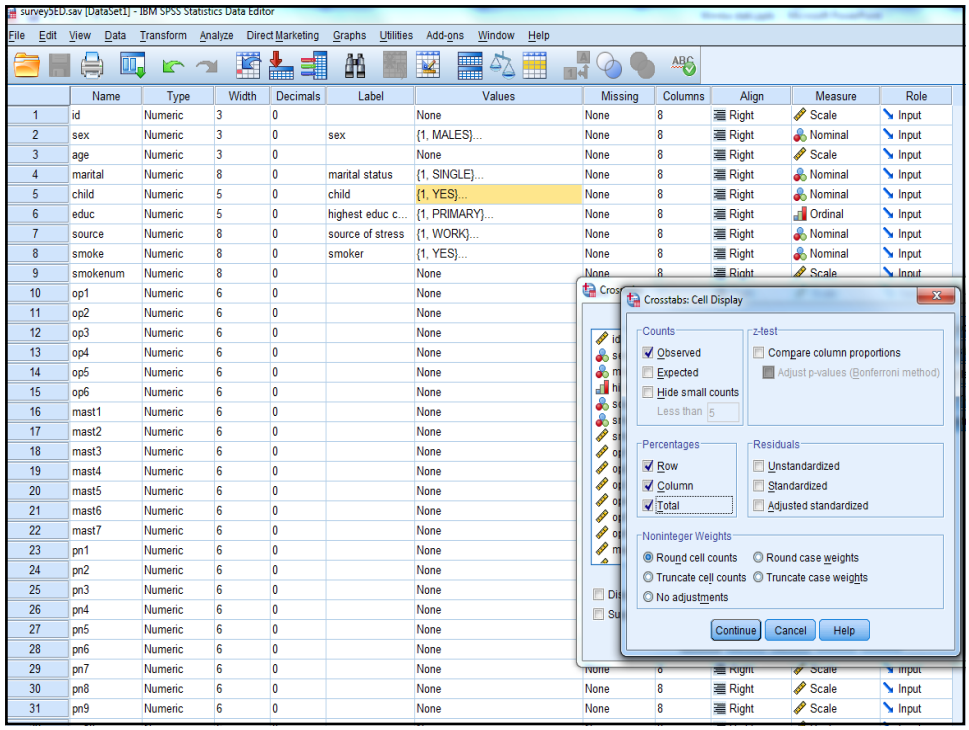
Chi-Square Test

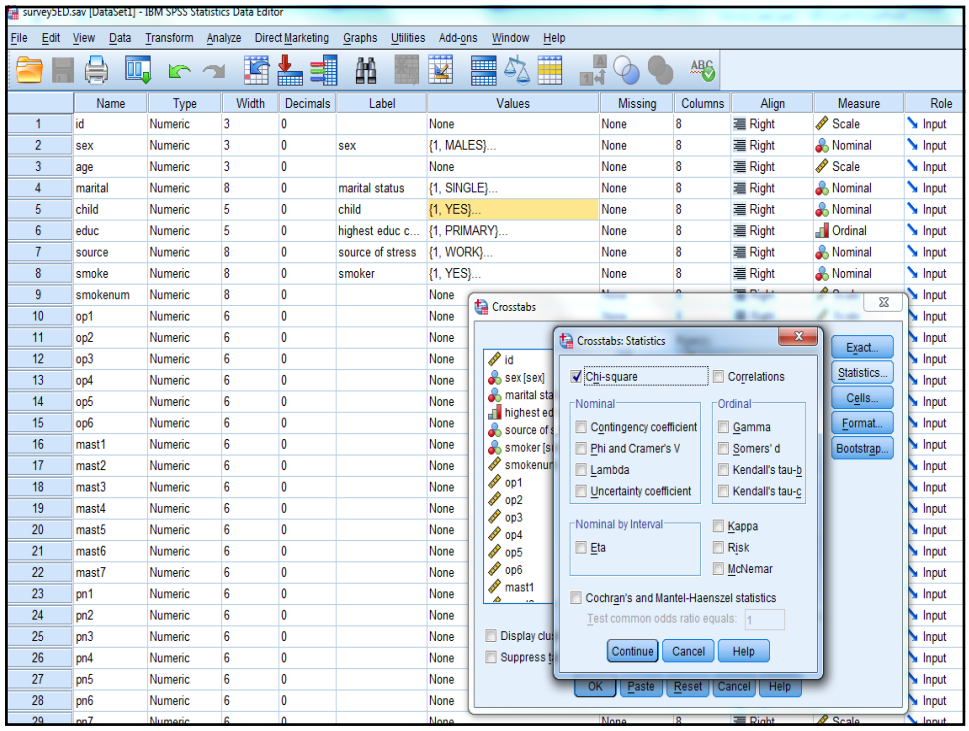
- Tests the strength of association between two variables
  - Independent Variable: Categorical
  - Dependent Variable: Categorical

What is the gender distribution of undergraduate students in relation to their choice of subjects?


survey5ED.sav [DataSet1] - 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
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...
Name Type Values Missing Columns Align
1 id Numeric None 8 Right
2 sex Numeric None 8 Right
3 age Numeric None 8 Right
4 marital Numeric None 8 Right
5 child Numeric None 8 Right
6 educ Numeric None 8 Right
7 source Numeric {1, WORK}... None 8 Right
8 smoke Numeric {1, YES}... None 8 Right
9 smokenum Numeric None None 8 Right
10 op1 Numeric None None 8 Right
11 op2 Numeric None None 8 Right
12 op3 Numeric None None 8 Right
13 op4 Numeric None None 8 Right
14 op5 Numeric None None 8 Right
15 op6 Numeric None None 8 Right
16 mast1 Numeric None None 8 Right
17 mast2 Numeric None None 8 Right
18 mast3 Numeric None None 8 Right
19 mast4 Numeric None None 8 Right
20 mast5 Numeric None None 8 Right
21 mast6 Numeric None None 8 Right
22 mast7 Numeric None None 8 Right
23 pn1 Numeric 6 0 None None 8 Right
24 pn2 Numeric 6 0 None None 8 Right
25 pn3 Numeric 6 0 None None 8 Right
26 pn4 Numeric 6 0 None None 8 Right
27 pn5 Numeric 6 0 None None 8 Right


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 3 0 None None 8 Right Scale Input
2 sex Numeric 3 0 sex {1, MALES}... None 8 Right Nominal Input
3 age Numeric 3 0 None None 8 Right Scale Input
4 marital Numeric 8 0 marital status {1, SINGLE}... None 8 Right Nominal Input
5 child Numeric 5 0 child {1, YES}... None 8 Right Nominal Input
6 educ Numeric 5 0 highest educ c... {1, PRIMARY}... None 8 Right Ordinal Input
7 source Numeric 8 0 source of stress {1, WORK}... None 8 Right Nominal Input
8 smoke Numeric 8 0 smoker {1, YES}... None 8 Right Nominal Input
9 smokenum Numeric 8 0 None None 8 Right Scale Input
Crosstabs
id
age
marital status [marital]
highest educ complet...
source of stress [sour...
smoker [smoke]
smokenum
op1
op2
op3
op4
op5
op6
mast1
Row(s):
sex [sex]
Column(s):
child [child]
Layer 1 of 1
Previous
Next
Exact...
Statistics...
Cells...
Format...
Bootstrap...
Display layer variables in table layers
Display clustered bar charts
Suppress tables
OK Paste Reset Cancel Help


survey5ED.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Crosstabs: Cell Display
Counts
Observed
Expected
Hide small counts
Less than 5
z-test
Compare column proportions
Adjust p-values (Bonferroni method)
Percentages
Row
Column
Total
Residuals
Unstandardized
Standardized
Adjusted standardized
Noninteger Weights
Round cell counts
Round case weights
Truncate cell counts
Truncate case weights
No adjustments
Continue
Cancel
Help


survey5ED.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Crosstabs
Crosstabs: Statistics
Chi-square
Correlations
Nominal
Contingency coefficient
Phi and Cramer's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals: 1
Continue
Cancel
Help
Exact...
Statistics...
Cells...
Format...
Bootstrap...
OK
Paste
Reset
Cancel
Help

Crosstabs

[DataSet1] H:\Personal\Oscail\HSA 2013-2014\Tutorials\HSA Workshop 9.11.13\Statistics\survey5ED.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
sex * child	438	99.8%	1	0.2%	439	100.0%

sex * child Crosstabulation

			child		
			YES	NO	Total
sex	MALES	Count	80	104	184
		% within sex	43.5%	56.5%	100.0%
		% within child	43.2%	41.1%	42.0%
		% of Total	18.3%	23.7%	42.0%
	FEMALES	Count	105	149	254
		% within sex	41.3%	58.7%	100.0%
		% within child	56.8%	58.9%	58.0%
		% of Total	24.0%	34.0%	58.0%
Total		Count	185	253	438
		% within sex	42.2%	57.8%	100.0%
		% within child	100.0%	100.0%	100.0%
		% of Total	42.2%	57.8%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.200[a]	1	.655		
Continuity Correction[b]	.122	1	.727		
Likelihood Ratio	.200	1	.655		
Fisher's Exact Test				.695	.363
Linear-by-Linear Association	.200	1	.655		
N of Valid Cases	438				

Assignment

- 3 exercises

1: Describing and Summarising Data (20%)

2: Tests of Difference (30%)

3: Analysing categorical data and Multivariate Analysis (50%)

Deadline: Monday 4th December 2017

Background to dataset

Questionnaire Survey

The data used for these exercises were collected as part of a recent survey of student **attitudes to research** in three institutions. The students were informed that some of the data they provided would be used by other students to carry out statistical exercises.

The questionnaire comprised five sections:

Section 1: Collects information on experience of research; respondents were asked to indicate, yes, no, or don't know.

Section 2: Contains 25 statements about research. Respondents were asked to indicate their level of agreement or disagreement with the statements.

Section 3: Contains 10 statements about mathematics. Respondents were asked to indicate their level of agreement or disagreement with the statements.

Section 4: Asks for personal details including gender, age, Leaving Certificate results, institution and intentions with regard to postgraduate study.

Section 5: Open ended questions (these have been removed from the dataset)

Questions?