ANALYSIS OF VARIANCE

T-tests; ANOVA; MANOVA

Chapters 17 & 18 Pallant

Topics

1. Experimental designs.
2. t-tests & non-parametric tests for group differences.
3. ANOVA – What is it?
4. Research design issues.
5. Example
6. MANOVA – what is it?

1. Experimental Design

An experiment: Manipulating one or more variables and examining the results.

The treatment (X) is the manipulation of the variable(s).

Test group receives treatment.

Test group O_1 X O_2

Control group O_3 O_4

Control group does not receive treatment.

Observations

Quasi-Experimental Design

- Random assignment (sampling) provides a means of isolation (other causal sources can be ruled out).

- Quasi experiments do not have random assignment, so you must control for other potential causal sources.

Banking Example

- Research Problem: What are the effects of electronic banking (e.g. ATM, telephone, internet) on customer satisfaction and loyalty?

- Design: Quasi-experimental design.

- Analysis Method: Compare group differences with t-tests or ANOVA.

Technique Choice

- **T-tests:** two groups (e.g. male female) or two time points (e.g. pre- and post-intervention.
- **ANOVA:** two or more groups or time points.
- **Paired samples or repeated measures:** Same observations (e.g. people) on more than one occasion, or matched pairs.
- **Between groups or independent samples:** Participants in each group are different (or independent).

Technique Choice 2

- **One-way ANOVA:** One independent variable (e.g. education).
- **Two-way ANOVA:** Two independent variables (e.g. education and gender).
- **MANOVA:** More than one dependent variable (e.g. satisfaction and loyalty).
- **ANCOVA:** Used when controlling for a variable that may influence relationship between dependent and independent variable.

Alternatives

Parametric	**Non-parametric**
Pearson correlation	Spearman correlation
Independent samples t-test	Mann-Whitney U test
Paired samples t-test	Wilcoxian signed rank test
One-way between groups ANOVA	Kruskal-Wallis test
One-way repeated measures ANOVA	Friedman test

Effect Size

- Statistical significance is only important in so far as the differences are substantively meaningful.
- Samples size (we know) affects the power to detect significant differences.
- We can calculate effect size, accounting for sample size, as a measure of the substantive meaning.

parametric (e.g. t-test) cutoffs

Effect size	**Eta squared**	**Cohen's d**
Small Medium Large	.01 (1%) .06 (6%) .138 (13.8%)	.2 .5 .8

2. t-tests

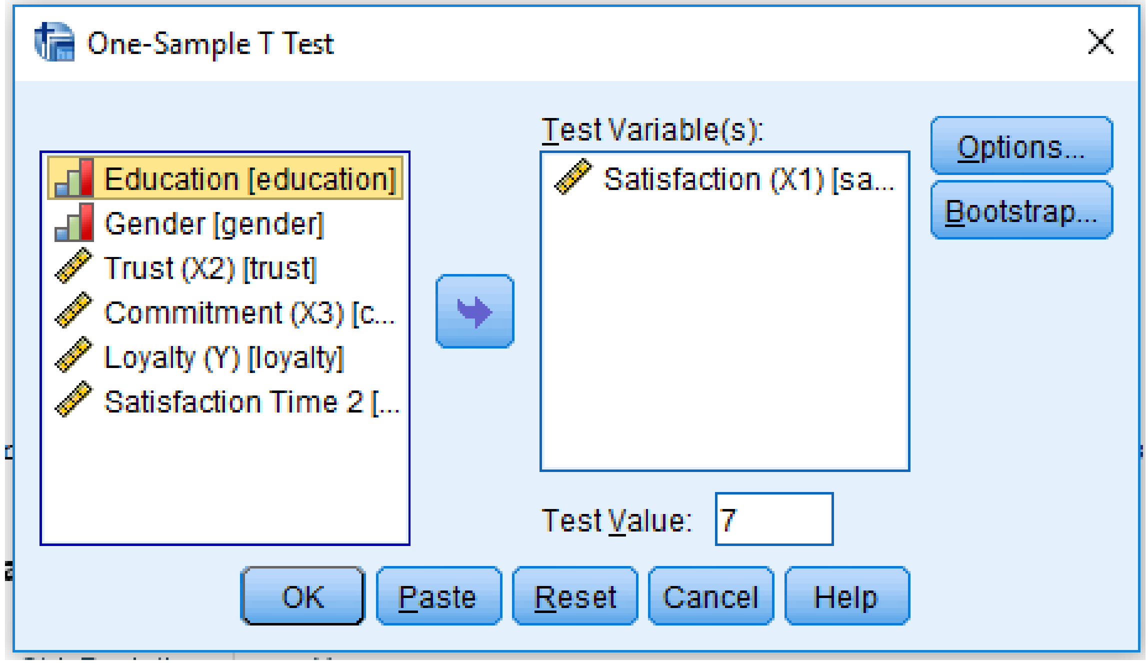
- t-tests assess the statistical significance of the differences between group means.
  - Non-parametric tests use medians.

- Which t-test to use depends on the type of groups you have.

One-Sample t-test

- This procedure tests whether the mean of a single variable differs from a specified constant.

Example: We might want to test whether the average satisfaction for a group of bank customers differs from a specified value (e.g. The previous year’s satisfaction score).



One-Sample T Test
Education [education]
Gender [gender]
Trust (X2) [trust]
Commitment (X3) [c...
Loyalty (Y) [loyalty]
Satisfaction Time 2 [...
Test Variable(s):
Satisfaction (X1) [sa...
Options...
Bootstrap...
Test Value: 7
OK
Paste
Reset
Cancel
Help

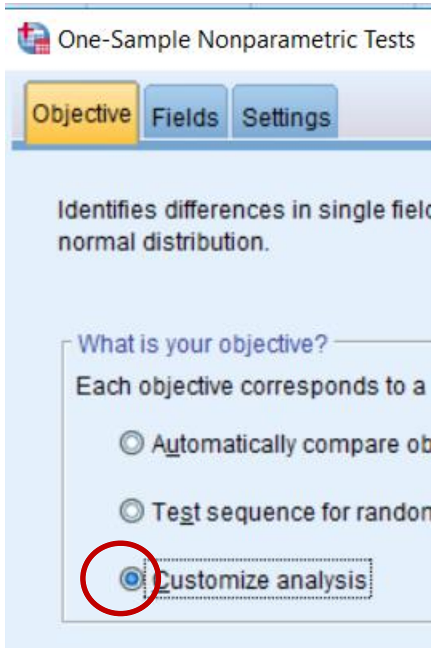
Example

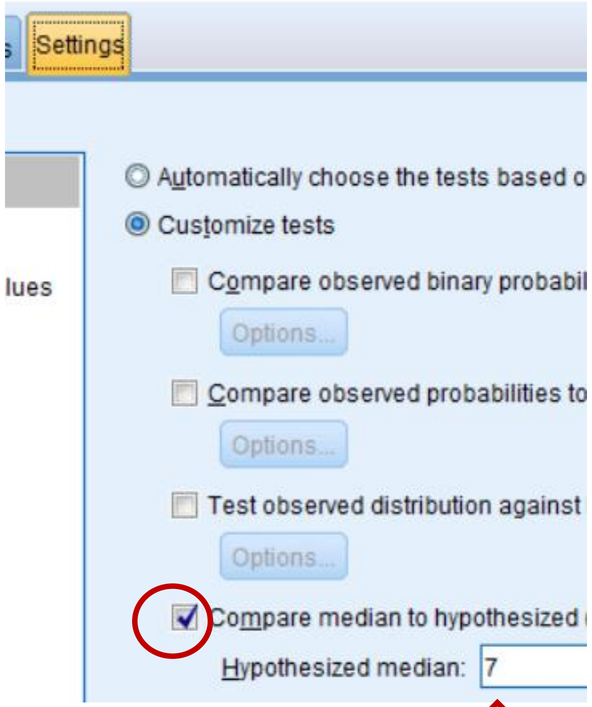
One-Sample Statistics

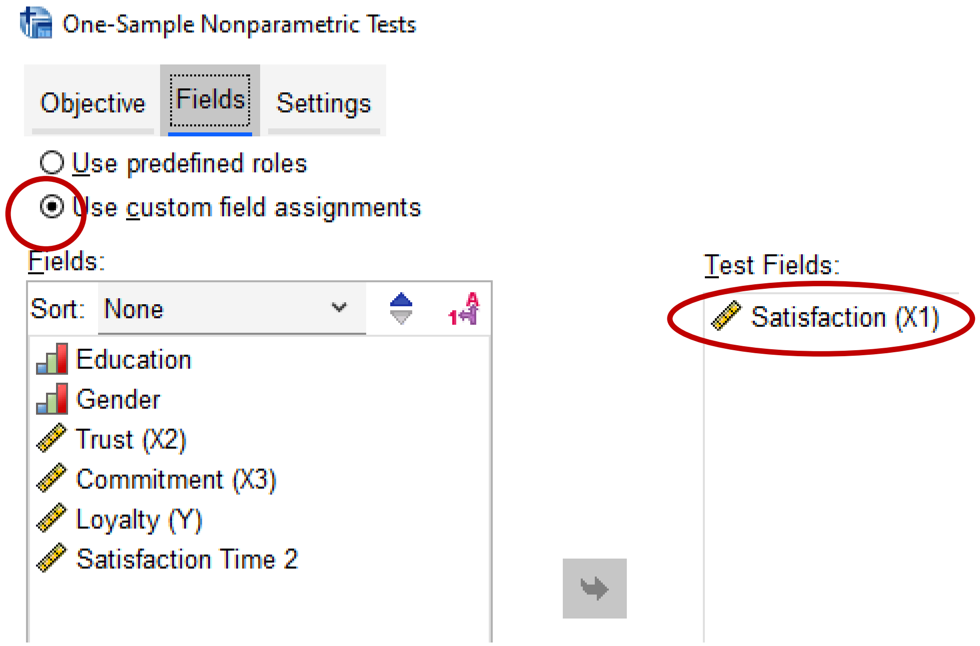
	N	Mean	Std. Deviation	Std. Error Mean
Satisfaction (X1)	353	7.3154	1.49388	.07951

One-Sample Test

	Test Value = 7.0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Satisfaction (X1)	3.967	352	.000	.3154	.1590	.4718







We are assuming a test median of 7.

Non-Parametric (sort of) Equivalent

One-Sample Wilcoxon Signed Rank Test

Satisfaction (X1)

One-Sample Wilcoxon Signed Rank Test Summary

Total N	353
Test Statistic	30866,000
Standard Error	1567,252
Standardized Test Statistic	4,414
Asymptotic Sig.(2-sided test)	<,001

Descriptives

			Statistic	Std. Error
Satisfaction (X1)	Mean		7,3154	,07951
	95% Confidence Interval for Mean	Lower Bound	7,1590	
		Upper Bound	7,4718	
	5% Trimmed Mean		7,3631	
	Median		7,3333	

There is a significant difference from 7.33

Independent Samples t-test = Mann-Whitney

This procedure compares means for two groups of cases. Ideally, for this test, the subjects should be randomly assigned to two groups, so that any difference in response is due to the treatment (or lack of treatment) and not to other factors.

Example: Bank customers are randomly split into two groups. One group receives a treatment (e.g. Promotional material), the other doesn't, and then both groups satisfaction level is measured. We then use the t-test to assess group differences on satisfaction.

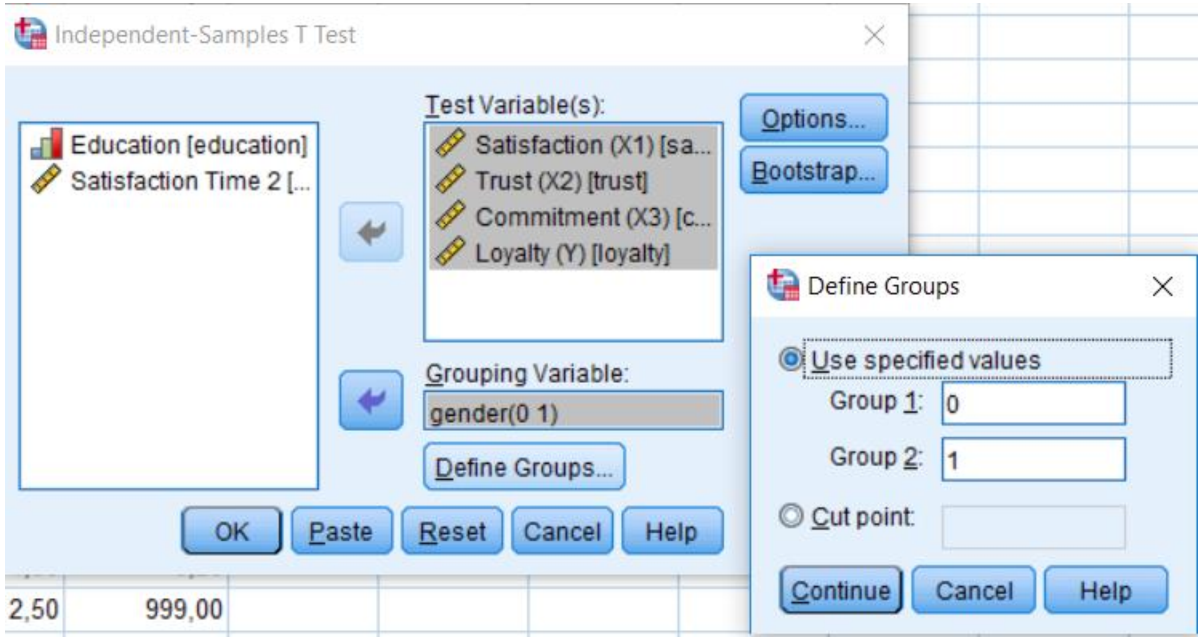
Note: This can be used to compare differences between groups like gender, although it is questionable as to whether you were randomly assigned…

Example

- We want to test for differences between male and female customers at the bank for:
  - Satisfaction
  - Trust
  - Commitment
  - Loyalty

Example Hypotheses:

- Hypothesis 1: Women are more satisfied with the bank than men.
- Hypothesis 2: Women trust the bank more than men.
- Hypothesis 3: Women are more committed to the bank than men.
- Hypothesis 4: Women are more loyal to the bank than men.



Independent-Samples T Test
Test Variable(s):
Options...
Bootstrap...
Education [education]
Satisfaction Time 2 [...
Satisfaction (X1) [sa...
Trust (X2) [trust]
Commitment (X3) [c...
Loyalty (Y) [loyalty]
Grouping Variable:
gender(0 1)
Define Groups...
OK
Paste
Reset
Cancel
Help
2,50
999,00
Define Groups
Use specified values
Group 1:
0
Group 2:
1
Cut point:
Continue
Cancel
Help

Example

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Satisfaction (X1)	Male	255	7,1948	1,47773	,09254
	Female	98	7,6293	1,49752	,15127
Trust (X2)	Male	253	5,2767	1,24025	,07797
	Female	103	5,4660	1,25492	,12365
Commitment (X3)	Male	252	4,5265	1,61903	,10199
	Female	104	5,1506	1,37780	,13510
Loyalty (Y)	Male	243	3,6924	,76383	,04900
	Female	102	3,8995	,72205	,07149

Note the larger difference in standard deviation for the commitment variable.

Example continued

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 Diff.	Lower	Upper
Sat	equal	.042	.838	-2.465	351	.014	-.4345	.17628	-.78119	-.08777
	not eq			-2.450	174	.015	-.4345	.17733	-.78448	-.08448
Trst	equal	.006	.940	-1.302	354	.194	-.1893	.14546	-.47541	.09673
	not eq			-1.295	187	.197	-.1893	.14618	-.47772	.09904
Com	equal	7.114	.008	-3.449	354	.001	-.6242	.18097	-.98009	-.26828
	not eq			-3.687	224	.000	-.6242	.16928	-.95777	-.29060
Loy	equal	.776	.379	-2.335	343	.020	-.2071	.08869	-.38158	-.03267
	not eq			-2.390	200	.018	-.2071	.08667	-.37804	-.03621

Not equal variances, use lower t-value.

No significant difference.

Effect Size – Eta^2 for Satisfaction

$$\frac{t^2}{t^2 + (N1 + N2 - 2)}$$

$$\frac{2.465^2}{2.465^2 + (255 + 98 - 2)}$$

$$= .017$$

Small effect
1.7% of variance

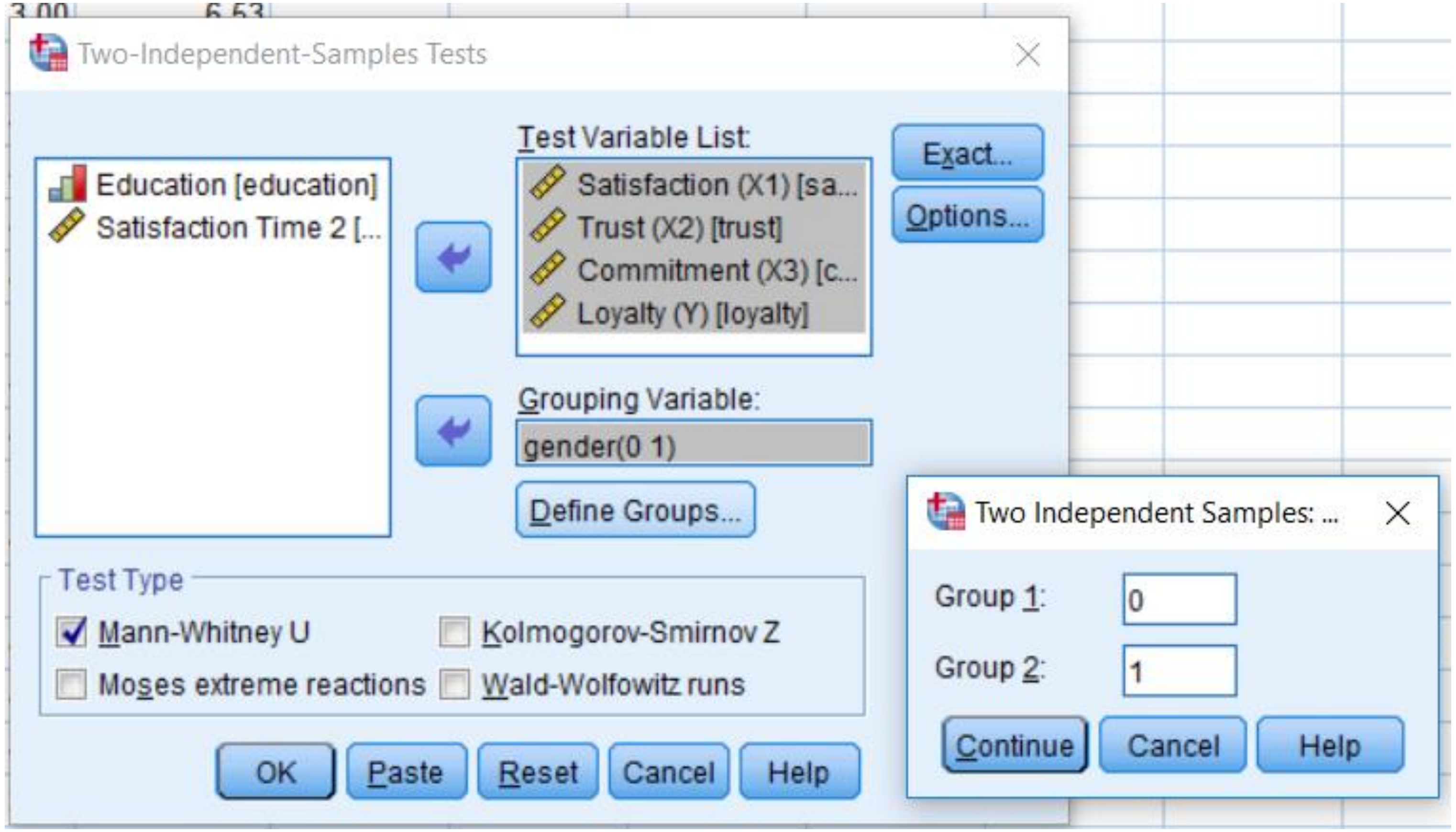
Effect size	Eta squared	Cohen's d
Small Medium Large	.01 (1%) .06 (6%) .138 (13.8%)	.2 .5 .8

Mann-Whitney Example

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Satisfaction (X1) is the same across categories of Groups.	Independent-Samples Mann-Whitney U Test	.004	Reject the null hypothesis.
2	The distribution of Trust (X2) is the same across categories of Groups.	Independent-Samples Mann-Whitney U Test	.150	Retain the null hypothesis.
3	The distribution of Commitment (X3) is the same across categories of Groups.	Independent-Samples Mann-Whitney U Test	.001	Reject the null hypothesis.
4	The distribution of Loyalty (Y) is the same across categories of Groups.	Independent-Samples Mann-Whitney U Test	.014	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.



Two-Independent-Samples Tests
Education [education]
Satisfaction Time 2 [...
Test Variable List:
Satisfaction (X1) [sa...
Trust (X2) [trust]
Commitment (X3) [c...
Loyalty (Y) [loyalty]
Exact...
Options...
Grouping Variable:
gender(0 1)
Define Groups...
Test Type
Mann-Whitney U
Kolmogorov-Smirnov Z
Moses extreme reactions
Wald-Wolfowitz runs
OK
Paste
Reset
Cancel
Help
Two Independent Samples: ...
Group 1:
0
Group 2:
1
Continue
Cancel
Help

Mann-Whitney Example

Which group is lower/higher

Ranks

	Gender	N	Mean Rank	Sum of Ranks
Satisfaction (X1)	Male	255	167,44	42696,50
	Female	98	201,88	19784,50
	Total	353		
Trust (X2)	Male	253	173,55	43908,00
	Female	103	190,66	19638,00
	Total	356		
Commitment (X3)	Male	252	166,79	42031,50
	Female	104	206,87	21514,50
	Total	356		
Loyalty (Y)	Male	243	164,50	39973,50
	Female	102	193,25	19711,50
	Total	345		

MW example continued

Test Statistics[a]

	Satisfaction (X1)	Trust (X2)	Commitment (X3)	Loyalty (Y)
Mann-Whitney U	10056,500	11777,000	10153,500	10327,500
Wilcoxon W	42696,500	43908,000	42031,500	39973,500
Z	-2,851	-1,439	-3,350	-2,458
Asymp. Sig. (2-tailed)	,004	,150	,001	,014

a. Grouping Variable: Gender

sig   not sig   sig   sig

Less than 0.05 means significant difference.

MW example continued

You should report medians because it is non-parametric, but with ordinal data it isn't unusual to get identical medians. Interpretation is tough!

Report

Median

Gender	Satisfaction (X1)	Trust (X2)	Commitment (X3)	Loyalty (Y)
Male	7,3333	5,5000	4,6667	3,7500
Female	8,0000	5,5000	5,3333	3,7500
Total	7,3333	5,5000	5,0000	3,7500

Effect Size - Satisfaction

$$\frac{z}{\sqrt{N}}$$ where N = total number of cases

$$\frac{2.851}{\sqrt{255 + 98}}$$

= .15

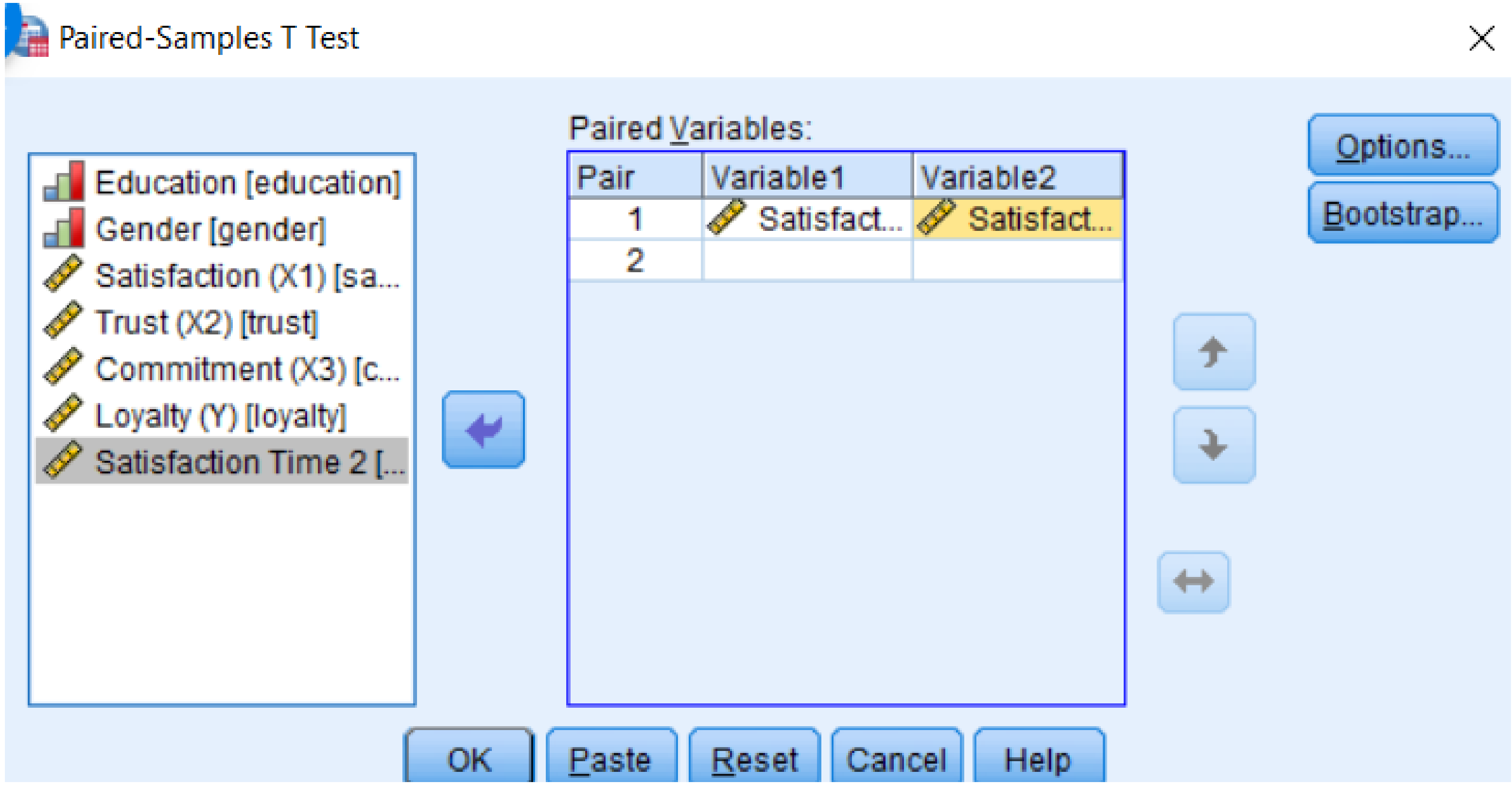
Small effect

Effect size	r
Small Medium Large	.1 .3 .5

Paired-Samples t-test (repeated measures) = Wilcoxian Signed Rank

This tests one sample that has been tested twice (repeated measures) or when there are two samples that have been matched or "paired”.

- Possible Example: Bank customers are randomly sampled, then the entire sample’s satisfaction is measured. The entire group receives a treatment (e.g. Promotional material), and then satisfaction is measured again. We then use the t-test to assess before & after differences on satisfaction.


Paired-Samples T Test
Paired Variables:
Pair
Variable1
Variable2
1
Satisfact...
Satisfact...
2
Education [education]
Gender [gender]
Satisfaction (X1) [sa...
Trust (X2) [trust]
Commitment (X3) [c...
Loyalty (Y) [loyalty]
Satisfaction Time 2 [...
Options...
Bootstrap...
OK
Paste
Reset
Cancel
Help

Example

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Satisfaction (X1)	7.3154	353	1.49388	.07951
	Satisfaction Time 2	8.3802	353	1.39484	.07424

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Satisfaction (X1) - Satisfaction Time 2	-1.06478	.63111	.03359	-1.13084	-.99871	-31.699	352	.000

Wilcoxian Example

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Satisfaction (X1) and Satisfaction Time 2 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

3. ANOVA – What is it?

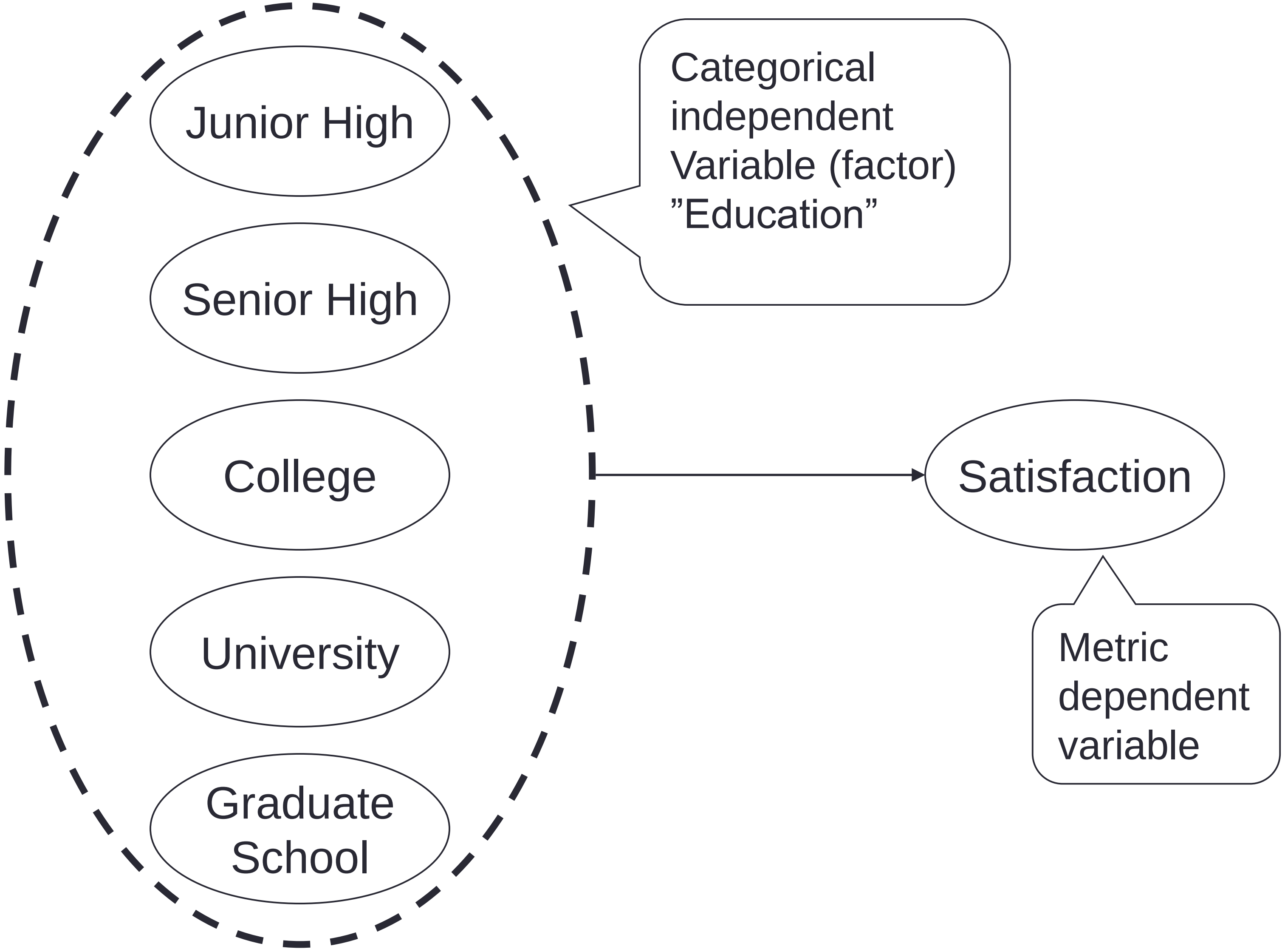
Analysis of Variance – ANOVA – is a dependence technique that measures the differences for a single independent variable defining groups based on one or more metric dependent variables.

- The t-test made an assessment between two groups. With more groups multiple t-tests could be used, but then you inflate the error. ANOVA avoids this problem.

Types of ANOVA

- Between groups (independent samples) – when the groups are different.
- Repeated measures – same group measured on different occasions.

Between Groups ANOVA Example



4. Research Design Issues

Sample Size:

- As a minimum, the observations in each cell must be greater than the number of dependent variables.
- Better: 20 observations per cell (more may be needed for more power).

Blocking factors:

- E.g. Male/Female: If we assume males and females to be different on the dependent variable, we can split the analysis between these two groups to increase the likelihood of finding group differences.

Independence of observations:

- Some spurious variable that causes the observations to be correlated.
- E.g. Dropping stock prices (if not controlled for) may affect all bank respondents in a uniform way.
- No tests – use logic.

Equal variance across groups:

- We are concerned about substantial differences in the variance across groups.
- Gets worse if group sizes are different.
- Levene test.

Multivariate normality:

- No tests – rely on univariate tests.

Multicollinearity:

- Is a bad thing.

Outliers:

- Have a very strong influence, so delete them if possible.

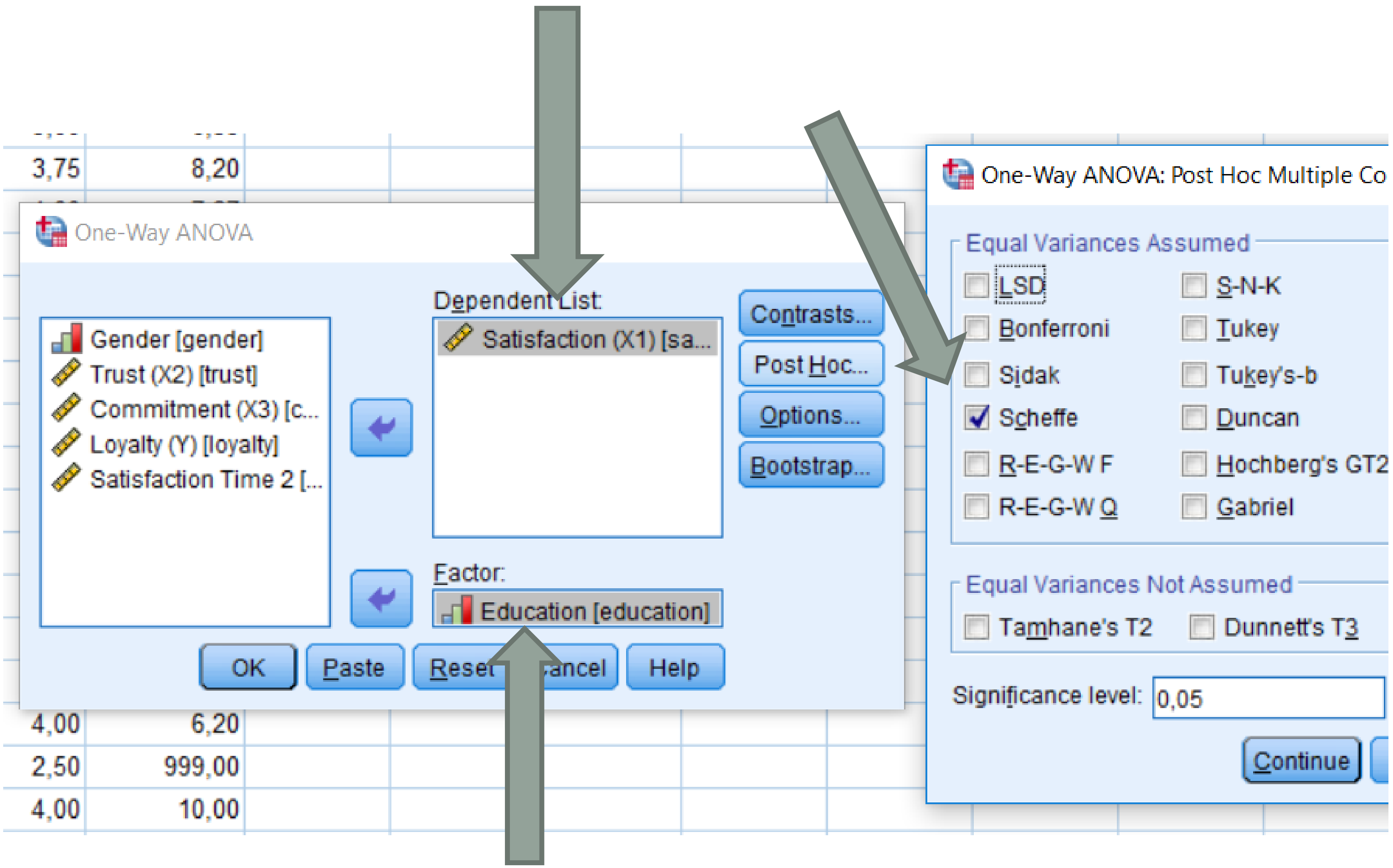
5. Example – Satisfaction and Education

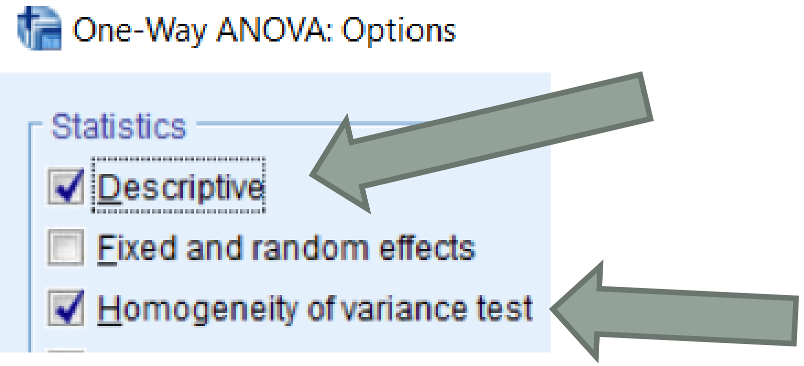
- As a bank, we are concerned about the effect of internet banking on customer satisfaction. One particular aspect we are considering is the level of education and the propensity to use and be satisfied with internet services. We believe they are positively associated. Therefore, we want to test if there are significant differences between groups with different levels of education with regard to satisfaction.

Hypotheses

- H_0: There is no significant difference in the level of satisfaction across levels of education.

- H_A: There is a significant difference in the level of satisfaction across levels of education.


3,75
8,20
One-Way ANOVA
Dependent List:
Satisfaction (X1) [sa...
Gender [gender]
Trust (X2) [trust]
Commitment (X3) [c...
Loyalty (Y) [loyalty]
Satisfaction Time 2 [...
Contrasts...
Post Hoc...
Options...
Bootstrap...
Factor:
Education [education]
OK
Paste
Reset
ancel
Help
4,00
6,20
2,50
999,00
4,00
10,00
One-Way ANOVA: Post Hoc Multiple Co
Equal Variances Assumed
LSD
S-N-K
Bonferroni
Tukey
Sidak
Tukey's-b
Scheffe
Duncan
R-E-G-W F
Hochberg's GT2
R-E-G-W Q
Gabriel
Equal Variances Not Assumed
Tamhane's T2
Dunnett's T3
Significance level:
0,05
Continue


One-Way ANOVA: Options
Statistics
Descriptive
Fixed and random effects
Homogeneity of variance test

Descriptive Statistics

There appears to be a trend, but is it significant?

Descriptives

Satisfaction (X1)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Junior High	33	7,889	1,74934	,30452	7,2686	8,5092	2,67	10,00
Senior High	112	7,574	1,29309	,12219	7,3323	7,8165	3,67	10,00
College	38	7,158	1,34396	,21802	6,7161	7,5996	4,00	9,00
University	106	7,217	1,62848	,15817	6,9034	7,5306	2,67	10,00
Graduate School	64	6,823	1,38519	,17315	6,4769	7,1689	2,67	10,00
Total	353	7,315	1,49388	,07951	7,1590	7,4718	2,67	10,00

Homogeneity

Test of Homogeneity of Variances

Satisfaction (X1)

Levene Statistic	df1	df2	Sig.
1,836	4	348	,121

Greater than 0.05 means that the variances are homogenous, which is what we want for this test (i.e. Big is good).

F-Test

Use these when referring to the tables. 4 is numerator and 348 is denominator.

Less than 0.05, so there is a significant difference between groups.

ANOVA

Satisfaction (X1)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	35,859	4	8,965	4,161	,003
Within Groups	749,694	348	2,154		
Total	785,553	352			

From the tables we see that the critical cutoff value is 2.37, which is smaller than 4.161, indicating a significant difference.

Group Differences

Identifying differences between groups:

- Scheffe
- Tukey’s HSD
- Tukey’s LSD
- Duncan
- Newman-Keuls

All of these have their weaknesses, mostly related to power.

Multiple Comparisons

Dependent Variable: Satisfaction (X1)
Scheffe

(I) Education	(J) Education	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Junior High	Senior High	,3145	,29072	,883	-,5858	1,2148
	College	,7310	,34925	,359	-,3506	1,8126
	University	,6719	,29258	,263	-,2342	1,5780
	Graduate School	1,0660*	,31455	,023	,0919	2,0401
Senior High	Junior High	-,3145	,29072	,883	-1,2148	,5858
	College	,4165	,27555	,684	-,4368	1,2698
	University	,3574	,19889	,521	-,2585	,9734
	Graduate School	,7515*	,22999	,032	,0392	1,4637
College	Junior High	-,7310	,34925	,359	-1,8126	,3506
	Senior High	-,4165	,27555	,684	-1,2698	,4368
	University	-,0591	,27752	1,000	-,9185	,8003
	Graduate School	,3350	,30059	,871	-,5959	1,2658
University	Junior High	-,6719	,29258	,263	-1,5780	,2342
	Senior High	-,3574	,19889	,521	-,9734	,2585
	College	,0591	,27752	1,000	-,8003	,9185
	Graduate School	,3941	,23235	,579	-,3255	1,1136
Graduate School	Junior High	-1,0660*	,31455	,023	-2,0401	-,0919
	Senior High	-,7515*	,22999	,032	-1,4637	-,0392
	College	-,3350	,30059	,871	-1,2658	,5959
	University	-,3941	,23235	,579	-1,1136	,3255

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

Normality?

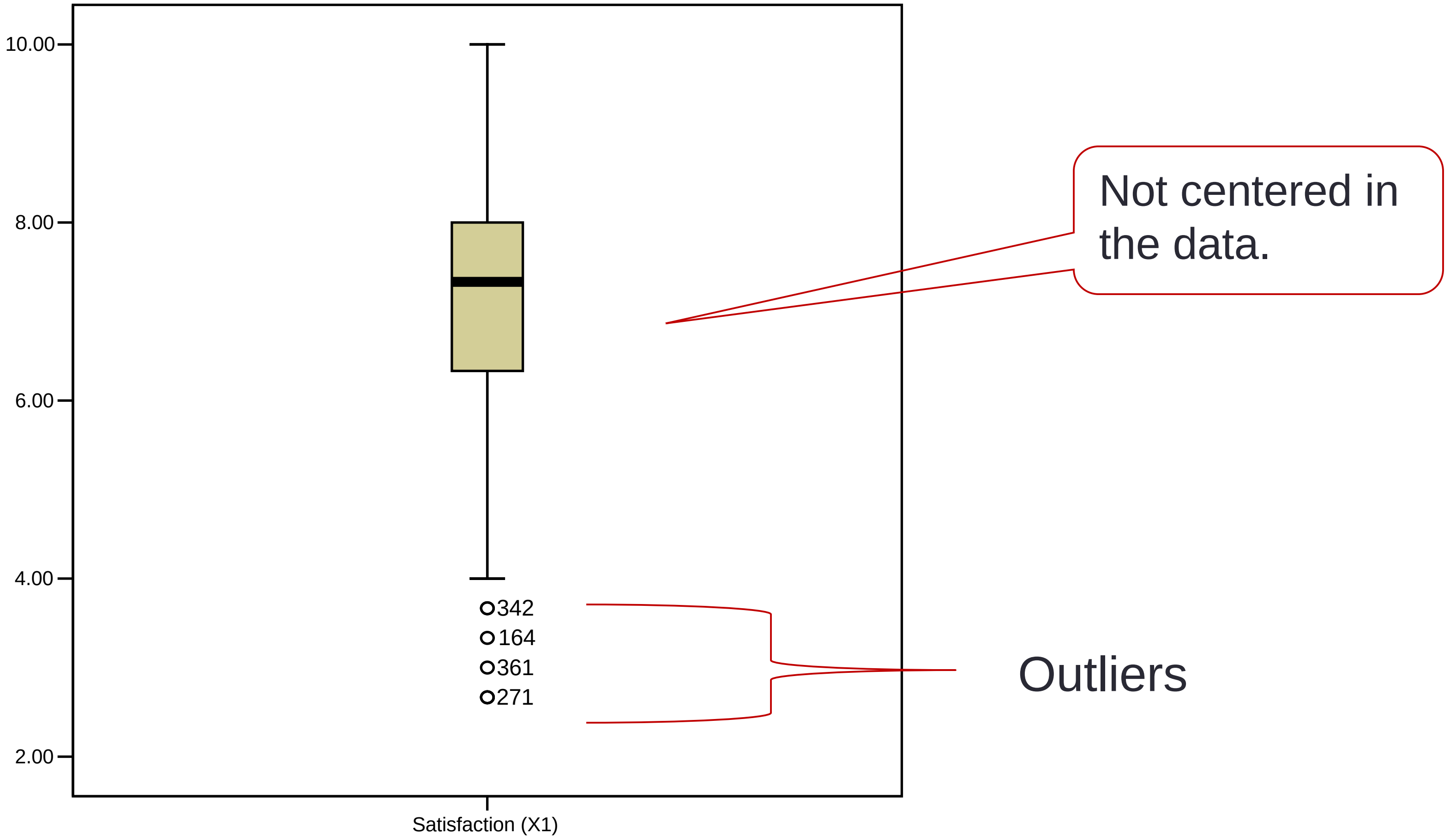
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Satisfaction (X1)	.102	353	.000	.971	353	.000

a. Lilliefors Significance Correction

Indicates that loyalty is not normally distributed (Sig. Is lower than .05).

Boxplot



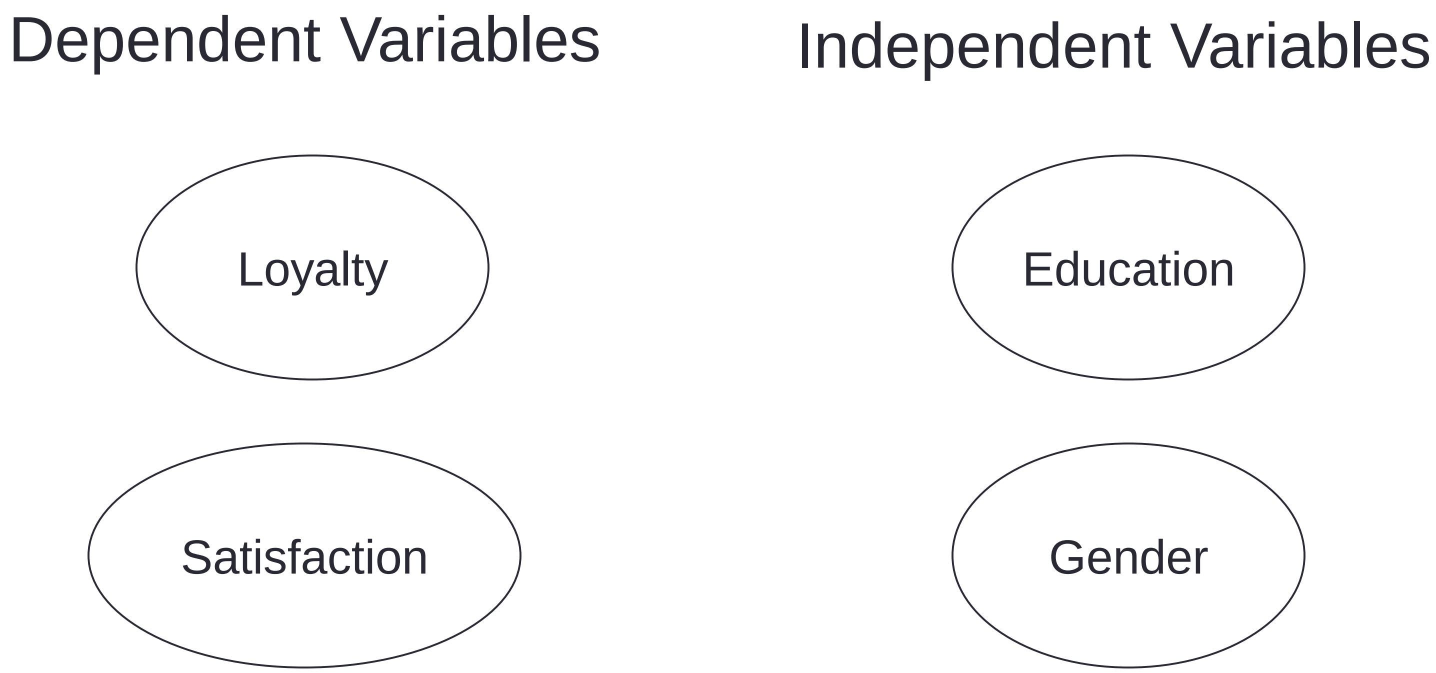
Remove outliers to improve normality, or consider non-parametric tests

6. MANOVA - What is it?

Multiple Analysis of Variance – MANOVA – is a dependence technique that measures the differences for two or more metric dependent variables based on a set of categorical independent variables.

- ANOVA only has one dependent variable.
- A series of univariate ANOVAs ignores the possibility of a composite linear combination of variables that provides evidence of group differences.
  - I.e. ANOVA doesn’t detect all possible group differences.

Example



Research Question: Do significant differences exist for the level of education, accounting for gender, for loyalty and satisfaction?

Between-Subject Factors

Between-Subjects Factors

		Value Label	N
Education	1	Junior High	33
	2	Senior High	105
	3	College	34
	4	University	101
	5	Graduate School	59
Gender	1	Male	236
	2	Female	96

Sample size: Minimum 20 per group – OK.

Note: Group sizes are very different so if there is any problem with unequal covariance matrices, this will make it worse.

Descriptive - Loyalty

Descriptive Statistics

	Education	Gender	Mean	Std. Deviation	N
Loyalty (Y)	Junior High	Male	4.0400	.67191	25
		Fem ale	4.1250	.55097	8
		Total	4.0606	.63747	33
	Senior High	Male	3.7799	.67493	67
		Fem ale	3.9276	.80099	38
		Total	3.8333	.72280	105
	College	Male	3.6635	.79982	26
		Fem ale	3.7500	.79057	8
		Total	3.6838	.78651	34
	University	Male	3.5445	.84916	73
		Fem ale	3.9554	.76089	28
		Total	3.6584	.84241	101
	Graduate School	Male	3.6444	.71607	45
		Fem ale	3.6964	.50171	14
		Total	3.6568	.66776	59
	Total	Male	3.6960	.76248	236
		Fem ale	3.9036	.72853	96
		Total	3.7560	.75762	332

No real big differences between means.

Problem with group sizes.

Multivariate Tests[d]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	Pillai's Trace	.952	3195.8[b]	2.000	321.0	.000	.952	6391.626	1.000
	Wilks' Lambda	.048	3195.8[b]	2.000	321.0	.000	.952	6391.626	1.000
	Hotelling's Trace	19.912	3195.8[b]	2.000	321.0	.000	.952	6391.626	1.000
	Roy's Largest Root	19.912	3195.8[b]	2.000	321.0	.000	.952	6391.626	1.000
EDUCAT	Pillai's Trace	.052	2.164	8.000	644.0	.028	.026	17.314	.858
	Wilks' Lambda	.948	2.184[b]	8.000	642.0	.027	.026	17.471	.862
	Hotelling's Trace	.055	2.203	8.000	640.0	.026	.027	17.626	.866
	Roy's Largest Root	.053	4.306[c]	4.000	322.0	.002	.051	17.225	.929
GENDER	Pillai's Trace	.022	3.633[b]	2.000	321.0	.028	.022	7.267	.669
	Wilks' Lambda	.978	3.633[b]	2.000	321.0	.028	.022	7.267	.669
	Hotelling's Trace	.023	3.633[b]	2.000	321.0	.028	.022	7.267	.669
	Roy's Largest Root	.023	3.633[b]	2.000	321.0	.028	.022	7.267	.669
EDUCAT * GENDER	Pillai's Trace	.042	1.731	8.000	644.0	.088	.021	13.851	.754
	Wilks' Lambda	.958	1.741[b]	8.000	642.0	.086	.021	13.927	.756
	Hotelling's Trace	.044	1.750	8.000	640.0	.084	.021	14.002	.759
	Roy's Largest Root	.041	3.327[c]	4.000	322.0	.011	.040	13.308	.841

a. Computed using alpha = .05

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept+EDUCAT+GENDER+EDUCAT * GENDER