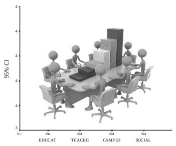
Analysis of Variance



Lecture 9

Survey Research & Design in Psychology

James Neill, 2012

Overview

1. Analysing differences
   1. Correlations vs. differences
   2. Which difference test?
   3. Parametric vs. non-parametrics
2. *t*-tests
   1. One-sample *t*-test
   2. Independent samples *t*-test
   3. Paired samples *t*-test

2

Overview

3. ANOVAs
   1. 1-way ANOVA
   2. 1-way repeated measures ANOVA
   3. Factorial ANOVA
4. Advanced ANOVAs
   1. Mixed design ANOVA (Split-plot ANOVA)
   2. ANCOVA

3

Readings – Assumed knowledge

Howell (2010):

- Ch3 The Normal Distribution
- Ch4 Sampling Distributions and Hypothesis Testing
- Ch7 Hypothesis Tests Applied to Means
- Ch11 Simple Analysis of Variance
- Ch12 Multiple Comparisons Among Treatment Means
- Ch13 Factorial Analysis of Variance

4

Readings

Howell (2010):

- Ch14 Repeated-Measures Designs
- Ch16 Analyses of Variance and Covariance as General Linear Models

See also: Inferential statistics decision-making tree

5

Analysing differences

- Correlations vs. differences
- Which difference test?
- Parametric vs. non-parametric

Correlational vs difference statistics

- Correlation and regression techniques reflect the strength of association
- Tests of differences reflect differences in central tendency of variables between groups and measures.

7

Correlational vs difference statistics

- In MLR we see the world as made of covariation. Everywhere we look, we see relationships.
- In ANOVA we see the world as made of differences. Everywhere we look we see differences.

8

Correlational vs difference statistics

- LR/MLR e.g., What is the **relationship** between gender and height in humans?
- *t*-test/ANOVA e.g., What is the **difference** between the heights of human males and females?

9

Which difference test? (2 groups)

How many groups?
(i.e. categories of IV)

1 group =
one-sample *t*-test

More than 2 groups =
ANOVA models

2 groups:
Are the groups
independent or
dependent?

Independent groups

Dependent groups

Non-para DV =
Mann-Whitney U

Non-para DV =
Wilcoxon

Para DV =
Independent samples t-test

Para DV =
Paired samples *t*-test

Parametric vs. non-parametric statistics

Parametric statistics – *inferential test* that assumes certain characteristics are true of an underlying population, especially the shape of its distribution.

Non-parametric statistics – *inferential test* that makes few or no assumptions about the population from which observations were drawn (distribution-free tests).

Parametric vs. non-parametric statistics

- There is generally at least one non-parametric equivalent test for each type of parametric test.
- Non-parametric tests are generally used when assumptions about the underlying population are questionable (e.g., non-normality).

Parametric vs. non-parametric statistics

- Parametric statistics commonly used for normally distributed interval or ratio dependent variables.
- Non-parametric statistics can be used to analyse DVs that are non-normal or are nominal or ordinal.
- Non-parametric statistics are *less powerful* that parametric tests.

So, *when* do I use a non-parametric test?

Consider non-parametric tests when (any of the following):

- Assumptions, like normality, have been violated.
- Small number of observations (*N*).
- DVs have nominal or ordinal levels of measurement.

Some commonly used parametric & non-parametric tests

Parametric	Non-parametric	Purpose
t test (independent)	Mann-Whitney U; Wilcoxon rank-sum	Compares two independent samples
t test (paired)	Wilcoxon matched pairs signed-rank	Compares two related samples
1-way ANOVA	Kruskal-Wallis	Compares three or more groups
2-way ANOVA	Friedman; χ^2 test of independence	Compares groups classified by two different factors

t-tests

- *t*-tests
- One-sample *t*-tests
- Independent sample *t*-tests
- Paired sample *t*-tests

Why a *t*-test or ANOVA?

- A *t*-test or ANOVA is used to determine whether a sample of scores are from the same population as another sample of scores.
- These are inferential tools for examining differences between group means.
- Is the difference between two sample means 'real' or due to chance?

t-tests

- One-sample
  One group of participants, compared with fixed, pre-existing value (e.g., population norms)
- Independent
  Compares mean scores on the same variable across different populations (groups)
- Paired
  Same participants, with repeated measures

Major assumptions

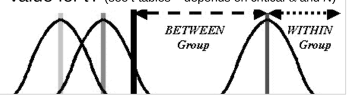
- Normally distributed variables
- Homogeneity of variance

In general, t*-tests and ANOVAs are robust to violation of assumptions, particularly with large cell sizes, but don't be complacent*.

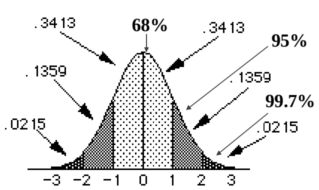
19

Use of *t* in *t*-tests

- *t* reflects the ratio of between group variance to within group variance
- Is the *t* large enough that it is unlikely that the two samples have come from the same population?
- Decision: Is *t* larger than the critical value for *t*? (see *t* tables – depends on critical α and *N*)



Ye good ol' normal distribution



21

One-tail vs. two-tail tests

- Two-tailed test rejects null hypothesis if obtained *t*-value is extreme is either direction
- One-tailed test rejects null hypothesis if obtained *t*-value is extreme is one direction (you choose – too high or too low)
- One-tailed tests are twice as powerful as two-tailed, but they are only focused on identifying differences in one direction.

22

One sample *t*-test

- Compare one group (a sample) with a fixed, pre-existing value (e.g., population norms)
- Do uni students sleep less than the recommended amount?
  e.g., Given a sample of *N* = 190 uni students who sleep M = 7.5 hrs/day (*SD* = 1.5), does this differ significantly from 8 hours hrs/day (α = .05)?

23

One-sample *t*-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Sleep	190	7.53	1.481	.107

One-Sample Test

	Test Value = 8					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Sleep	-4.358	189	.000	-.468	-.68	-.26

Independent groups *t*-test

- Compares mean scores on the same variable across different populations (groups)
- Do Americans vs. Non-Americans differ in their approval of Barack Obama?
- Do males & females differ in the amount of sleep they get?

25

Assumptions (Indep. samples *t*-test)

- **LOM**
  - **IV** is ordinal / categorical
  - **DV** is interval / ratio
- **Homogeneity of Variance**: If variances unequal (Levene's test), adjustment made
- **Normality**: *t*-tests robust to modest departures from normality, otherwise consider use of Mann-Whitney U test
- **Independence of observations** (one participant's score is not dependent on any other participant's score)

26

Do males and females differ in in amount of sleep per night?

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Sleep	male	85	7.31	1.640	.178
	female	105	7.71	1.319	.129

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Sleep	Equal variances assumed	.667	.415	-1.902	188	.059	-.408
	Equal variances not assumed			-1.880	159.618	.065	-.408

Do males and females differ in memory recall?

Group Statistics

	gender_R Gender of respondent	N	Mean	Std. Deviation	Std. Error Mean
immrec immediate recall-number correct_wave 1	1 Male	1189	7.34	2.109	.061
	2 Female	1330	8.24	2.252	.062

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
	4.784	.029	-10.268	2517	.000	-.896	.087	-1.067	-.725
			-10.306	2511.570	.000	-.896	.087	-1.066	-.725

Adolescents' Same Sex Relations in Single Sex vs. Co-Ed Schools

Group Statistics

	Type of School	N	Mean	Std. Deviation	Std. Error Mean
SSR	Single Sex	323	4.9995	.7565	4.209E-02
	Co-Educational	168	4.9455	.7158	5.523E-02

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
SSR	Equal variances assumed	.017	.897	.764	489	.445	5.401E-02	7.067E-02	-8.48E-02	.1929
	Equal variances not assumed			.778	355.220	.437	5.401E-02	6.944E-02	-8.26E-02	.1906

Adolescents' Opposite Sex Relations in Single Sex vs. Co-Ed Schools

Group Statistics

	Type of School	N	Mean	Std. Deviation	Std. Error Mean
OSR	Single Sex	327	4.5327	1.0627	5.877E-02
	Co-Educational	172	3.9827	1.1543	8.801E-02

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
SSR	Equal variances assumed	.017	.897	.764	489	.445	5.401E-02	7.067E-02	-8.48E-02	.1929
	Equal variances not assumed			.778	355.220	.437	5.401E-02	6.944E-02	-8.26E-02	.1906

Independent samples *t*-test

- Comparison b/w means of 2 independent sample variables = *t*-test
  (e.g., what is the difference in Educational Satisfaction between male and female students?)
- Comparison b/w means of 3+ independent sample variables = 1-way ANOVA
  (e.g., what is the difference in Educational Satisfaction between students enrolled in four different faculties?)

Paired samples *t*-test → 1-way repeated measures ANOVA

- Same participants, with repeated measures
- Data is sampled within subjects. Measures are repeated e.g.,:
  - Time e.g., pre- vs. post-intervention
  - Measures e.g., approval ratings of brand X and brand Y

Assumptions (Paired samples *t*-test)

- LOM:
  - IV: Two measures from same participants (w/in subjects)
    - a variable measured on two occasions or
    - two different variables measured on the same occasion
  - DV: Continuous (Interval or ratio)
- Normal distribution of difference scores (robust to violation with larger samples)
- Independence of observations (one participant's score is not dependent on another's score)

Does an intervention have an effect?

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pretest	19.80	20	21.867	4.890
	Posttest	14.40	20	19.198	4.293

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 Pretest - Posttest	5.400	13.527	3.025	-.931	11.731	1.785	19	.090

There was no significant difference between pretest and posttest scores ($t(19) = 1.78$, $p = .09$).

Adolescents' Opposite Sex vs. Same Sex Relations

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SSR	4.9787	951	.7560	2.451E-02
	OSR	4.2498	951	1.1086	3.595E-02

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 SSR - OSR	.7289	.9645	3.128E-02	.6675	.7903	23.305	950	.000

Paired samples *t*-test → 1-way repeated measures ANOVA

- Comparison b/w means of 2 within subject variables = *t*-test
- Comparison b/w means of 3+ within subject variables = 1-way repeated measures ANOVA
  (e.g., what is the difference in Campus, Social, and Education Satisfaction?)

Summary
(Analysing Differences)

- Non-parametric and parametric tests can be used for examining differences between the central tendency of two of more variables
- Learn when to use each of the parametric tests of differences, from one-sample *t*-test through to ANCOVA (e.g. use a decision chart).

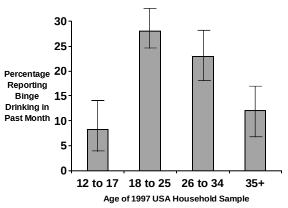
37

t-tests

- Difference between a set value and a variable → one-sample *t*-test
- Difference between two independent groups → independent samples *t*-test = BETWEEN-SUBJECTS
- Difference between two related measures (e.g., repeated over time or two related measures at one time) → paired samples *t*-test = WITHIN-SUBJECTS

38

Are the differences in a sample **generalisable** to a population?



Introduction to ANOVA
(Analysis of Variance)

- Extension of a *t*-test to assess differences in the central tendency (*M*) of several groups or variables.
- DV variance is partitioned into between-group and within-group variance
- Levels of measurement:
  - –Single DV: metric,
  - –1 or more IVs: categorical

40

Example ANOVA research question

Are there differences in the degree of religious commitment between countries (UK, USA, and Australia)?

1. 1-way ANOVA ←
2. 1-way repeated measures ANOVA
3. Factorial ANOVA
4. Mixed ANOVA
5. ANCOVA

41

Example ANOVA research question

Do university students have different levels of satisfaction for educational, social, and campus-related domains ?

1. 1-way ANOVA
2. 1-way repeated measures ANOVA ←
3. Factorial ANOVA
4. Mixed ANOVA
5. ANCOVA

42

Example ANOVA research questions

Are there differences in the degree of religious commitment between countries (UK, USA, and Australia) and gender (male and female)?

1. 1-way ANOVA
2. 1-way repeated measures ANOVA
3. Factorial ANOVA ←
4. Mixed ANOVA
5. ANCOVA

43

Example ANOVA research questions

Does couples' relationship satisfaction differ between males and females and before and after having children?

1. 1-way ANOVA
2. 1-way repeated measures ANOVA
3. Factorial ANOVA
4. Mixed ANOVA ←
5. ANCOVA

44

Example ANOVA research questions

Are there differences in university student satisfaction between males and females (gender) after controlling for level of academic performance?

1. 1-way ANOVA
2. 1-way repeated measures ANOVA
3. Factorial ANOVA
4. Mixed ANOVA
5. ANCOVA ←

45

Introduction to ANOVA

- Inferential: What is the likelihood that the observed differences could have been due to chance?
- Follow-up tests: Which of the *M*s differ?
- Effect size: How large are the observed differences?

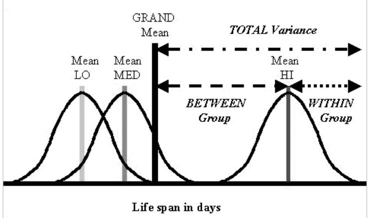
46

F test

- ANOVA partitions the sums of squares (variance from the mean) into:
  - Explained variance (between groups)
  - Unexplained variance (within groups) – or error variance
- *F* = ratio between explained & unexplained variance
- *p* = probability that the observed mean differences between groups could be attributable to chance

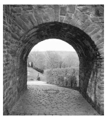
47

F is the ratio of between-group : within-group variance



Follow-up tests

- ANOVA *F*-tests are a "gateway". If *F* is significant, then...



- interpret (main and interaction) effects and
- consider whether to conduct follow-up tests
  - planned comparisons
  - post-hoc contrasts.

49

One-way ANOVA

50

Assumptions – One-way ANOVA

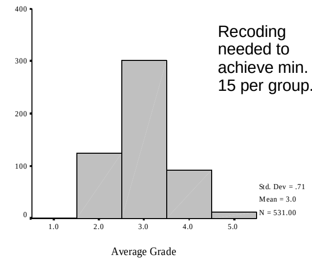
Dependent variable (DV) must be:

- LOM: Interval or ratio
- Normality: Normally distributed for all IV groups (robust to violations of this assumption if *N*s are large and approximately equal e.g., >15 cases per group)
- Variance: Equal variance across for all IV groups (homogeneity of variance)
- Independence: Participants' data should be independent of others' data

51

One-way ANOVA: Are there differences in satisfaction levels between students who get different grades?

52



Recoding needed to achieve min. 15 per group.

53

These groups could be combined.

AVGRADE Average Grade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Fail	1	.2	.2	.2
	2 Pass	125	20.5	23.5	23.7
	3	2	.3	.4	24.1
	3 Credit	299	48.9	56.3	80.4
	4	4	.7	.8	81.2
	4 Distinction	88	14.4	16.6	97.7
	5 High Distinction	12	2.0	2.3	100.0
	Total	531	86.9	100.0	
Missing	System	80	13.1		
Total		611	100.0		

54

The recoded data has more similar group sizes and is appropriate for ANOVA.

AVGRADX Average Grade (R)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.00 Fail/Pass	128	20.9	24.1	24.1
	3.00 Credit	299	48.9	56.3	80.4
	4.00 D/HD	104	17.0	19.6	100.0
	Total	531	86.9	100.0	
Missing	System	80	13.1		
Total		611	100.0		

*SD*s are similar (homogeneity of variance). *M*s suggest that higher grade groups are more satisfied.

Descriptive Statistics

Dependent Variable: EDUCAT

AVGRADX	Mean	Std. Deviation	N
2.00 Fail/Pass	3.57	.53	128
3.00 Credit	3.74	.51	299
4.00 D/HD	3.84	.55	104
Total	3.72	.53	531

Levene's test indicates homogeneity of variance.

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

Dependent Variable: EDUCAT

F	df1	df2	Sig.
.748	2	528	.474

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

a. Design: Intercept+AVGRADX

Tests of Between-Subjects Effects

Dependent Variable: EDUCAT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.306[a]	2	2.153	7.854	.000
Intercept	5981.431	1	5981.431	21820.681	.000
AVGRADX	4.306	2	2.153	7.854	.000
Error	144.734	528	.274		
Total	7485.554	531			
Corrected Total	149.040	530			

a. R Squared = .029 (Adjusted R Squared = .025)

Follow-up tests should then be conducted because the effect of Grade is statistically significant (p < .05).

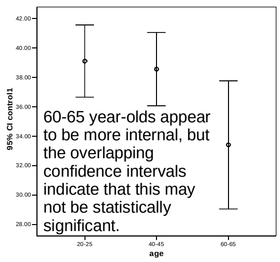
One-way ANOVA: Does locus of control differ between three age groups?

Age

- 20-25 year-olds
- 40-45 year olds
- 60-65 year-olds

Locus of Control

- Lower = internal
- Higher = external



The *SD*s vary between groups (the third group has almost double the *SD* of the younger group). Levene's test is significant (variances are not homogenous).

	N	Mean	Std. Deviation
.00 20-25	20	39.1000	5.25056
1.00 40-45	20	38.5500	5.29623
2.00 60-65	20	33.4000	9.29289
Total	60	37.0167	7.24040

Test of Homogeneity of Variances

control1

Levene Statistic	df1	df2	Sig.
13.186	2	57	.000

ANOVA

control1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	395.433	2	197.717	4.178	.020
Within Groups	2697.550	57	47.325		
Total	3092.983	59			

There is a significant effect for Age (F (2, 59) = 4.18, p = .02). In other words, the three age groups are unlikely to be drawn from a population with the same central tendency for LOC.

Which age groups differ in their mean locus of control scores? (Post hoc tests).

Multiple Comparisons

Dependent Variable: control1

Tukey HSD

(I) age	(J) age	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
.00 20-25	1.00 40-45	.55000	2.17544	.965	-4.6850	5.7850
	2.00 60-65	5.70000*	2.17544	.030	.4650	10.9350
1.00 40-45	.00 20-25	-.55000	2.17544	.965	-5.7850	4.6850
	2.00 60-65	5.15000	2.17544	.055	-.0850	10.3850
2.00 60-65	.00 20-25	-5.70000*	2.17544	.030	-10.9350	-.4650
	1.00 40-45	-5.15000	2.17544	.055	-10.3850	.0850

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

Conclude: Gps 0 differs from 2; 1 differs from 2

Follow-up (pairwise) tests

- Post hoc: Compares every possible combination
- Planned: Compares specific combinations

(Do one or the other; not both)

64

Post hoc

- Control for Type I error rate
- Scheffe, Bonferroni, Tukey's HSD, or Student-Newman-Keuls
- Keeps experiment-wise error rate to a fixed limit

65

Planned

- Need hypothesis before you start
- Specify contrast coefficients to weight the comparisons (e.g., 1st two vs. last one)
- Tests each contrast at critical α

66

Assumptions - Repeated measures ANOVA

Repeated measures designs have the additional assumption of Sphericity:

- Variance of the population difference scores for any two conditions should be the same as the variance of the population difference scores for any other two conditions
- Test using Mauchly's test of sphericity (If Mauchly's W Statistic is $p < .05$ then assumption of sphericity is violated.)

67

Assumptions - Repeated measures ANOVA

- Sphericity is commonly violated, however the multivariate test (provided by default in PASW output) does not require the assumption of sphericity and may be used as an alternative.
- The obtained F ratio must then be evaluated against new degrees of freedom calculated from the Greenhouse-Geisser, or Huynh-Feld, Epsilon values.

68

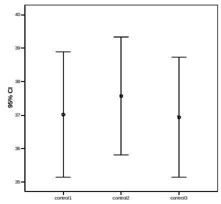
Example: Repeated measures ANOVA

Does LOC vary over time?

- Baseline
- 6 months
- 12 months

69

Mean LOC scores (with 95% C.I.s) across 3 measurement occasions



Not much variation between means.

70

Descriptive statistics

Descriptive Statistics

	Mean	Std. Deviation	N
control1	37.0167	7.24040	60
control2	37.5667	6.80071	60
control3	36.9333	6.92788	60

Not much variation between means.

71

Mauchly's test of sphericity

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	.938	3.727	2	.155	.941	.971	.500

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: Intercept
Within Subjects Design: factor1

Mauchly's test is not significant, therefore sphericity can be assumed.

72

Tests of within-subject effects

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	14.211	2	7.106	2.791	.065
	Greenhouse-Geisser	14.211	1.883	7.548	2.791	.069
	Huynh-Feldt	14.211	1.943	7.315	2.791	.067
	Lower-bound	14.211	1.000	14.211	2.791	.100
Error(factor1)	Sphericity Assumed	300.456	118	2.546		
	Greenhouse-Geisser	300.456	111.087	2.705		
	Huynh-Feldt	300.456	114.628	2.621		
	Lower-bound	300.456	59.000	5.092		

Conclude: Observed differences in means could have occurred by chance (F (2, 118) = 2.79, p = .06) if critical alpha = .05

73

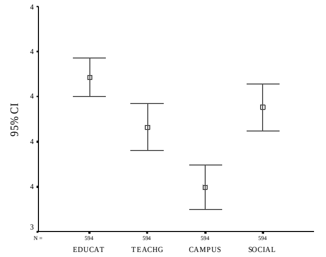
1-way repeated measures ANOVA
Do satisfaction levels vary between Education, Teaching, Social and Campus aspects of university life?

74

Descriptive Statistics

	Mean	Std. Deviation
EDUCAT	3.74	.54
TEACHG	3.63	.65
CAMPUS	3.50	.61
SOCIAL	3.67	.65

75



76

Tests of within-subject effects

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
SATISF	Sphericity Assumed	18.920	3	6.307	28.386	.000
	Greenhouse-Geisser	18.920	2.520	7.507	28.386	.000
	Huynh-Feldt	18.920	2.532	7.472	28.386	.000
	Lower-bound	18.920	1.000	18.920	28.386	.000
Error(SATISF)	Sphericity Assumed	395.252	1779	.222		
	Greenhouse-Geisser	395.252	1494.572	.264		
	Huynh-Feldt	395.252	1501.474	.263		
	Lower-bound	395.252	593.000	.667		

77

Factorial ANOVA (2-way): Are there differences in satisfaction levels between gender and age?

78

Factorial ANOVA

- Levels of measurement
  - 2 or more between-subjects categorical/ordinal IVs
  - 1 interval/ratio DV
- e.g., Does Educational Satisfaction vary according to Age (2) and Gender (2)? 2 x 2 Factorial ANOVA

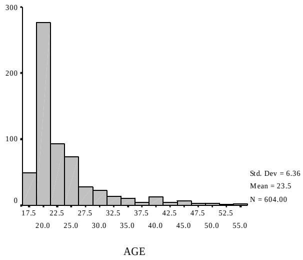
79

Factorial ANOVA

- Factorial designs test Main Effects and Interactions. For a 2-way design:
  - Main effect of IV1
  - Main effect of IV2
  - Interaction between IV1 and IV2
- If
  - significant effects are found and
  - there are more than 2 levels of an IV are involved

then follow-up tests are required.

80

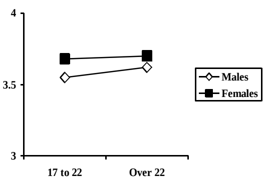


81

AGE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	17	3	.5	.5	.5
	18	46	7.5	7.6	8.1
	19	69	11.3	11.4	19.5
	20	114	18.7	18.9	38.4
	21	94	15.4	15.6	54.0
	22	64	10.5	10.6	64.6
	23	29	4.7	4.8	69.4
	24	29	4.7	4.8	74.2
	25	30	4.9	5.0	79.1
	26	15	2.5	2.5	81.6
	27	16	2.6	2.6	84.3
	28	12	2.0	2.0	86.3
	29	7	1.1	1.2	87.4
	30	7	1.1	1.2	88.6
	31	8	1.3	1.3	89.9
	32	7	1.1	1.2	91.1
	33	7	1.1	1.2	92.2
	34	3	.5	.5	92.7

82



83

Tests of Between-Subjects Effects

Dependent Variable: TEACHG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.124[a]	3	.708	1.686	.169
Intercept	7136.890	1	7136.890	16996.047	.000
AGEX	.287	1	.287	.683	.409
GENDER	1.584	1	1.584	3.771	.053
AGEX * GENDER	6.416E-02	1	6.416E-02	.153	.696
Error	250.269	596	.420		
Total	8196.937	600			
Corrected Total	252.393	599			

a. R Squared = .008 (Adjusted R Squared = .003)

84

Descriptive Statistics

Dependent Variable: TEACHG

AGEX Age	GENDER	Mean	Std. Deviation	N
1.00 17 to 22	0 Male	3.5494	.6722	156
	1 Female	3.6795	.5895	233
	Total	3.6273	.6264	389
2.00 over 22	0 Male	3.6173	.7389	107
	1 Female	3.7038	.6367	104
	Total	3.6600	.6901	211
Total	0 Male	3.5770	.6995	263
	1 Female	3.6870	.6036	337
	Total	3.6388	.6491	600

85

Factorial ANOVA (2-way): Are there differences in LOC between gender and age?

86

Example: Factorial ANOVA

Main effect 1:
- Do LOC scores differ by Age?

Main effect 2:
- Do LOC scores differ by Gender?

Interaction:
- Is the relationship between Age and LOC moderated by Gender? (Does any relationship between Age and LOC vary as a function of Gender?)

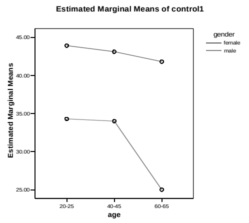
87

Example: Factorial ANOVA

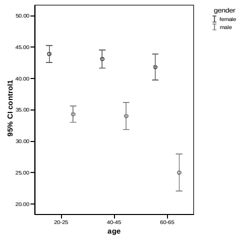
- In this example, there are:
  - –Two main effects (Age and Gender)
  - –One interaction effect (Age x Gender)
- IVs
  - –Age recoded into 2 groups (2)
  - –Gender dichotomous (2)
- DV
  - –Locus of Control (LOC)

88

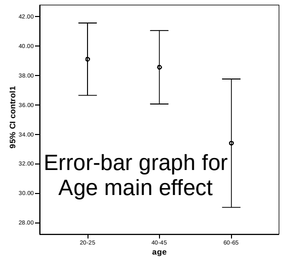
Plot of LOC by Age and Gender



89



90

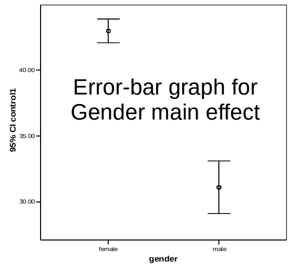


Descriptives for Age main effect

Descriptives

control1

	N	Mean	Std. Deviation
.00 20-25	20	39.1000	5.25056
1.00 40-45	20	38.5500	5.29623
2.00 60-65	20	33.4000	9.29289
Total	60	37.0167	7.24040



Descriptives for Gender main effect

Descriptives

control1

	N	Mean	Std. Deviation
.00 female	30	42.9333	2.40593
1.00 male	30	31.1000	5.33272
Total	60	37.0167	7.24040

Descriptives for LOC by Age and Gender

Dependent Variable: control1

age	gender	Mean	Std. Deviation	N
.00 20-25	.00 female	43.9000	1.91195	10
	1.00 male	34.3000	1.82878	10
	Total	39.1000	5.25056	20
1.00 40-45	.00 female	43.1000	2.02485	10
	1.00 male	34.0000	3.01846	10
	Total	38.5500	5.29623	20
2.00 60-65	.00 female	41.8000	2.89828	10
	1.00 male	25.0000	4.13656	10
	Total	33.4000	9.29289	20
Total	.00 female	42.9333	2.40593	30
	1.00 male	31.1000	5.33272	30
	Total	37.0167	7.24040	60

Tests of between-subjects effects

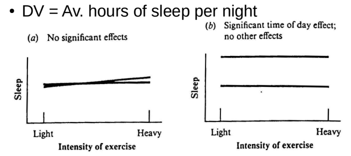
Dependent Variable: control1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2681.483[a]	5	536.297	70.377	.000
Intercept	82214.017	1	82214.017	10788.717	.000
age	395.433	2	197.717	25.946	.000
gender	2100.417	1	2100.417	275.632	.000
age * gender	185.633	2	92.817	12.180	.000
Error	411.500	54	7.620		
Total	85307.000	60			
Corrected Total	3092.983	59			

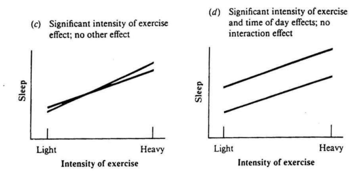
a. R Squared = .867 (Adjusted R Squared = .855)

Interactions

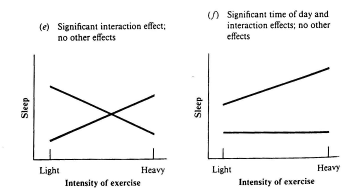
- IV1 = Separate lines for morning and evening exercise.
- IV2 = Light and heavy exercise
- DV = Av. hours of sleep per night



Interactions



Interactions



Mixed design ANOVA (SPANOVA)

- Independent groups (e.g., males and females) with **repeated measures** on each group (e.g., word recall under three different character spacing conditions (Narrow, Medium, Wide)).
- Since such experiments have mixtures of between-subject and within-subject factors they are said to be of **mixed design**
- Since output is split into two tables of effects, this is also said to be **split-plot ANOVA (SPANOVA)**

Mixed design ANOVA (SPANOVA)

- IV1 is between-subjects (e.g., Gender)
- IV2 is within-subjects (e.g., Social Satisfaction and Campus Satisfaction)
- Of interest are:
  - **Main effect** of IV1
  - **Main effect** of IV2
  - **Interaction** b/w IV1 and IV2
- If significant effects are found and more than 2 levels of an IV are involved, then specific contrasts are required, either:
  - A priori (planned) contrasts
  - Post-hoc contrasts

Mixed design ANOVA (SPANOVA)

An experiment has two IVs:

- Between-subjects = Gender (Male or Female) - varies between subjects
- Within-subjects = Spacing (Narrow, Medium, Wide)
- Gender - varies within subjects

Mixed design ANOVA: Design

- If A is Gender and B is Spacing the Reading experiment is of the type A X (B) or 2 x (3)
- Brackets signify a mixed design with repeated measures on Factor B

Mixed design ANOVA: Assumptions

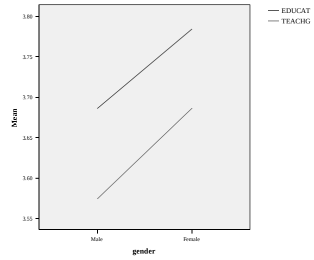
- Normality
- Homogeneity of variance
- Sphericity
- Homogeneity of inter-correlations

Homogeneity of intercorrelations

- The pattern of inter-correlations among the various levels of repeated measure factor(s) should be consistent from level to level of the Between-subject Factor(s)
- The assumption is tested using Box's *M* statistic
- Homogeneity is present when the *M* statistic is NOT significant at $p > .001$.

Mixed design ANOVA: Example

Do satisfaction levels vary between gender for education and teaching?



Tests of within-subjects contrasts

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SATISF	Type III Sum of Squares	df	Mean Square	F	Sig.
SATISF	Linear	3.262	1	3.262	22.019	.000
SATISF * GENDER	Linear	1.490E-02	1	1.490E-02	.101	.751
Error(SATISF)	Linear	88.901	600	.148		

Tests of 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.
Intercept	16093.714	1	16093.714	29046.875	.000
GENDER	3.288	1	3.288	5.934	.015
Error	332.436	600	.554		

1. gender

Measure: MEASURE_1

gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
0 Male	3.630	.032	3.566	3.693
1 Female	3.735	.029	3.679	3.791

2. satisf

Measure: MEASURE_1

satisf	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	3.735	.022	3.692	3.778
2	3.630	.027	3.578	3.682

What is ANCOVA?

- Analysis of *Cov*ariance
- Extension of ANOVA, using 'regression' principles
- Assesses effect of
  - one variable (IV) on
  - another variable (DV)
  - after controlling for a third variable (CV)

ANCOVA
(Analysis of Covariance)

- A covariate IV is added to an ANOVA (can be dichotomous or metric)
- Effect of the covariate on the DV is removed (or partialled out) (akin to Hierarchical MLR)
- Of interest are:
  - Main effects of IVs and interaction terms
  - Contribution of CV (akin to Step 1 in HMLR)
- e.g., GPA is used as a CV, when analysing whether there is a difference in Educational Satisfaction between Males and Females.

Why use ANCOVA?

- **Reduces** variance associated with covariate (CV) from the **DV error** (unexplained variance) term
- Increases power of *F*-test
- May not be able to achieve experimental control over a variable (e.g., randomisation), but can measure it and statistically control for its effect.

Why use ANCOVA?

- Adjusts group means to what they would have been if all *P*s had scored identically on the CV.
- The differences between *P*s on the CV are removed, allowing focus on remaining variation in the DV due to the IV.
- Make sure hypothesis (hypotheses) is/are clear.

Assumptions of ANCOVA

- As per ANOVA
- Normality
- Homogeneity of Variance (use Levene's test)

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

Dependent Variable: achievement

F	df1	df2	Sig.
.070	1	78	.792

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

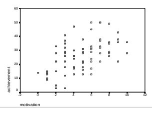
a. Design: Intercept+MOTIV+TEACH

Assumptions of ANCOVA

- Independence of observations
- Independence of IV and CV
- Multicollinearity - if more than one CV, they should not be highly correlated - eliminate highly correlated CVs
- Reliability of CVs - not measured with error - only use reliable CVs

Assumptions of ANCOVA

- Check for linearity between CV & DV - check via scatterplot and correlation.
- If the CV is not correlated with the DV there is no point in using it.



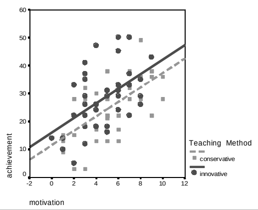
Assumptions of ANCOVA

Homogeneity of regression

- Assumes slopes of regression lines between CV & DV are equal for each level of IV, if not, don't proceed with ANCOVA
- Check via scatterplot with lines of best fit

118

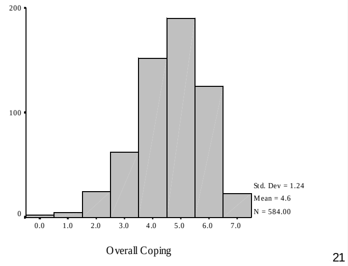
Assumptions of ANCOVA



119

ANCOVA example 1: Does education satisfaction differ between people with different levels of coping ('Not coping', 'Just coping' and 'Coping well') with average grade as a covariate?

120



21

COPEX Coping

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00 Not Coping	94	15.4	16.1	16.1
	2.00 Coping	151	24.7	25.9	42.0
	3.00 Coping Well	338	55.3	58.0	100.0
	Total	583	95.4	100.0	
Missing	System	28	4.6		
Total		611	100.0		

122

Descriptive Statistics

Dependent Variable: EDUCAT

COPEX Coping	Mean	Std. Deviation	N
1.00 Not Coping	3.4586	.6602	83
2.00 Just Coping	3.6453	.5031	129
3.00 Coping Well	3.8142	.4710	300
Total	3.7140	.5299	512

123

Tests of Between-Subjects Effects

Dependent Variable: EDUCAT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11.894[a]	3	3.965	15.305	.000
Intercept	302.970	1	302.970	1169.568	.000
AVGRADE	2.860	1	2.860	11.042	.001
COPEX	7.400	2	3.700	14.283	.000
Error	131.595	508	.259		
Total	7206.026	512			
Corrected Total	143.489	511			

a. R Squared = .083 (Adjusted R Squared = .077)

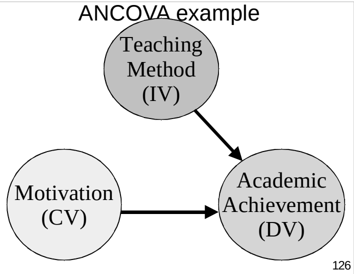
124

ANCOVA Example 2: Does teaching method affect academic achievement after controlling for motivation?

- IV = teaching method
- DV = academic achievement
- CV = motivation
- Experimental design - assume students randomly allocated to different teaching methods.

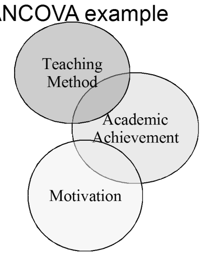
125

ANCOVA example



126

ANCOVA example



ANCOVA example 2

Tests of Between-Subjects Effects

Dependent Variable: achievement

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	189.113[a]	1	189.113	1.622	.207	.020
Intercept	56021.113	1	56021.113	480.457	.000	.860
TEACH	189.113	1	189.113	1.622	.207	.020
Error	9094.775	78	116.600			
Total	65305.000	80				
Corrected Total	9283.888	79				

a. R Squared = .020 (Adjusted R Squared = .008)

- A one-way ANOVA shows a non-significant effect for teaching method (IV) on academic achievement (DV)

ANCOVA example 2

Tests of Between-Subjects Effects

Dependent Variable: achievement

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	3050.744[a]	2	1525.372	18.843	.000	.329
Intercept	2794.773	1	2794.773	34.525	.000	.310
MOTIV	2861.632	1	2861.632	35.351	.000	.315
TEACH	421.769	1	421.769	5.210	.025	.063
Error	6233.143	77	80.950			
Total	65305.000	80				
Corrected Total	9283.888	79				

a. R Squared = .329 (Adjusted R Squared = .311)

- An ANCOVA is used to adjust for differences in motivation
- *F* has gone from 1 to 5 and is significant because the error term (unexplained variance) was reduced by including motivation as a CV.

ANCOVA & hierarchical MLR

- ANCOVA is similar to hierarchical regression – assesses impact of IV on DV while controlling for 3rd variable.
- ANCOVA more commonly used if IV is categorical.

Summary of ANCOVA

- Use ANCOVA in survey research when you can't randomly allocate participants to conditions e.g., quasi-experiment, or control for extraneous variables.
- ANCOVA allows us to statistically control *for one or more* covariates.

Summary of ANCOVA

- Decide which variable(s) are IV, DV & CV.
- Check assumptions:
  - normality
  - homogeneity of variance (Levene's test)
  - Linearity between CV & DV (scatterplot)
  - homogeneity of regression (scatterplot – compares slopes of regression lines)
- Results – does IV effect DV after controlling for the effect of the CV?

Effect sizes

Three effect sizes are relevant to ANOVA:

- **Eta-square (η^2)** provides an overall test of size of effect
- **Partial eta-square (η_p^2)** provides an estimate of the effects for each IV.
- **Cohen's *d***: Standardised differences between two means.

Effect Size: Eta-squared (η^2)

- Analagous to R^2 from regression
- = $SS_{between}$ / SS_{total} = SS_B / SS_T
- = prop. of variance in Y explained by X
- = Non-linear correlation coefficient
- = prop. of variance in Y explained by X
- Ranges between 0 and 1

Effect Size: Eta-squared (η^2)

- Interpret as for r^2 or R^2
- Cohen's rule of thumb for interpreting η^2:
  - .01 is small
  - .06 medium
  - .14 large

ANOVA

ontrol1

	Sum of Squares	df	Mean Square	F	Sig.
etween Groups	395.433	2	197.717	4.178	.020
Vithin Groups	2697.550	57	47.325		
otal	3092.983	59			

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

$= 395.433 / 3092.983$

$= 0.128$

Eta-squared is expressed as a percentage: 12.8% of the total variance in control is explained by differences in Age

136

Effect Size: Eta-squared (η^2)

- The eta-squared column in SPSS *F*-table output is actually partial eta-squared (η_p^2). Partial eta-squared indicates the size of effect for each IV (also useful).
- η^2 is not provided by SPSS – calculate separately:
  - $= SS_{between} / SS_{total}$
  - = prop. of variance in Y explained by X
- R^2 at the bottom of SPSS *F*-tables is the linear effect as per MLR – if an IV has 3 or more non-interval levels, this won't equate with η^2.

137

Results - Writing up ANOVA

- Establish clear hypotheses – one for each main or interaction or covariate effect
- Test the assumptions, esp. LOM, normality and *n* for each cell, homogeneity of variance, Box's *M*, Sphericity
- Present the descriptive statistics (*M*, *SD*, skewness, and kurtosis in a table, with marginal totals)
- Present a figure to illustrate the data (bar, error-bar, or line graph)

138

Results - Writing up ANOVA

- Report on test results – Size, direction and significance (*F*, *p*, partial eta-squared)
- Conduct planned or post-hoc testing as appropriate, with pairwise effect sizes (Cohen's *d*)
- Indicate whether or not results support hypothesis (hypotheses)

139

Summary

- Hypothesise each main effect and interaction effect.
- *F* is an omnibus "gateway" test; may require follow-up tests.
- Conduct follow-up tests where sig. main effects have three or more levels.

140

Summary

- Choose from mixed-design ANOVA or ANCOVA for lab report
- Repeated measure designs include the assumption of sphericity

141

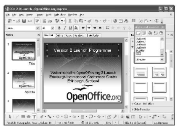
Summary

- Report on the size of effects potentially using:
  - Eta-square (η^2) as the omnibus ES
  - Partial eta-square (η_p^2) for each IV
  - Standardised mean differences for the differences between each pair of means (e.g., Cohen's *d*)

142

Open Office Impress

- This presentation was made using Open Office Impress.
- Free and open source software.
- http://www.openoffice.org/product/impress.html



143