Survey Methods & Design in Psychology

Lecture 10

ANOVA

(2007)

Lecturer: James Neill

Overview of Lecture

- Testing mean differences
- ANOVA models
- Interactions
- Follow-up tests
- Effect sizes

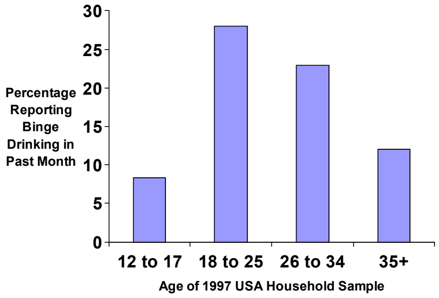
Parametric Tests of Mean Differences

- One-sample *t*-test
- Independent samples *t*-test
- Paired samples *t*-test
- One-way ANOVA
- One-way repeated measures ANOVA
- Factorial ANOVA
- Mixed design ANOVA
- ANCOVA
- MANOVA
- Repeated measures MANOVA

Correlational statistics vs tests of differences between groups

- Correlation/regression techniques reflect the strength of association between continuous variables
- Tests of group differences (t-tests, anova) indicate whether significant differences exist between group means

Are The Differences We See Real?



Major Assumptions

- Normally distributed variables
- Homogeneity of variance
- Robust to violation of assumptions

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.

(in other words these are inferential tools for examining differences in means)

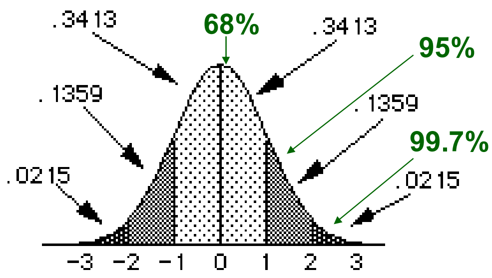
t-tests

- An inferential statistical test used to determine whether two sets of scores come from the same population
- Is the difference between two sample means 'real' or due to chance?

Use of *t* in *t*-tests

- Question: 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



Use of *t* in *t*-tests

- *t* reflects the ratio of differences between groups to within groups variability
- 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*)

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.

Single sample *t*-test

Compare one group (a sample) with a fixed, pre-existing value (e.g., population norms)

E.g., Does a sample of university students who sleep on average 6.5 hours per day (*SD* = 1.3) differ significantly from the recommended 8 hours of sleep?

Independent groups *t*-test

Compares mean scores on the same variable across different populations (groups)

e.g.,
- Do males and females differ in IQ?
- Do Americans vs. Non-Americans differ in their approval of George Bush?

Assumptions (Independent samples *t*-test)

- IV is ordinal / categorical e.g., gender
- DV is interval / ratio e.g., self-esteem
- Homogeneity of Variance
  - If variances unequal (Levene's test), adjustment made
  - Normality – t-tests robust to modest departures from normality: consider use of Mann-Whitney U test if severe skewness
- Independence of observations (one participant's score is not dependent on any other participant's score)

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				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. E Differe
immrec immediate recall-number correct_wave 1	Equal variances assumed	4.784	.029	-10.268	2517	.000	-.896	
	Equal variances not assumed			-10.306	2511.570	.000	-.896	

Paired samples *t*-test

- Same participants, with repeated measures
- Data is sampled <u>within subjects</u>, e.g.,
  - Pre- vs. post- treatment ratings
  - Different factors
    e.g., Voter's approval ratings of candidate X vs. Y

Assumptions- paired samples *t*-test

- DV must be measured at interval or ratio level
- <u>Population of difference scores must be normally distributed (robust to violation with larger samples)</u>
- Independence of observations (one participant's score is not dependent on any other participant's score)

Do females' memory recall scores change over time?

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	immrec immediate recall-number correct_ wave 1 & bimrec immediate recall-number correct_w2	1234	.528	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	immrec immediate recall-number correct_ wave 1 - bimrec immediate recall-number correct_w2	-.086	2.204	.063	-.209	.037	-1.369	1233	.171

Assumptions

- IV is ordinal / categorical e.g., gender
- DV is interval / ratio e.g., self-esteem
- Homogeneity of Variance
  - If variances unequal, adjustment made (Levene's Test)
- Normality - often violated, without consequence
  - look at histograms
  - look at skewness
  - look at kurtosis

SPSS Output: Independent Samples *t*-test: Same Sex Relations

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						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the 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

SPSS Output: Independent Samples *t*-test: Opposite Sex Relations

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						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the 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

SPSS Output: Independent Samples *t*-test: Opposite 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)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1 SSR - OSR	.7289	.9645	3.128E-02	.6675	.7903	23.305	950	.000

What is ANOVA? (Analysis of Variance)

- An extension of a *t*-test
- A way to test for differences between *M*s of:
  (i) more than 2 groups, or
  (ii) more than 2 times or variables
- Main assumption:
- DV is metric, IV is categorical

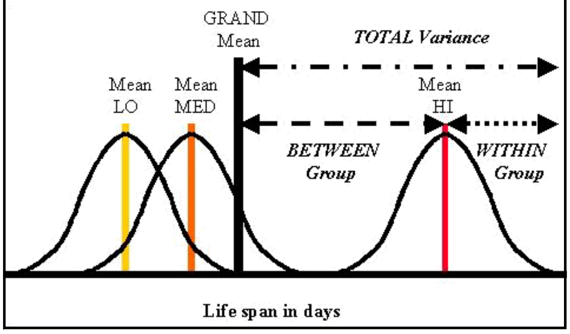
Introduction to ANOVA

- Single DV, with 1 or more IVs
- IVs are discrete
- Are there differences in the central tendency of groups?
- Inferential: Could the observed differences be due to chance?
- Follow-up tests: Which of the *M*s differ?
- Effect Size: How large are the differences?

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* represents the ratio between explained and unexplained variance
- *F* indicates the likelihood that the observed mean differences between groups could be attributable to chance.
- *F* is equivalent to a MLR test of the significance of *R*.

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



Assumptions – One-way ANOVA

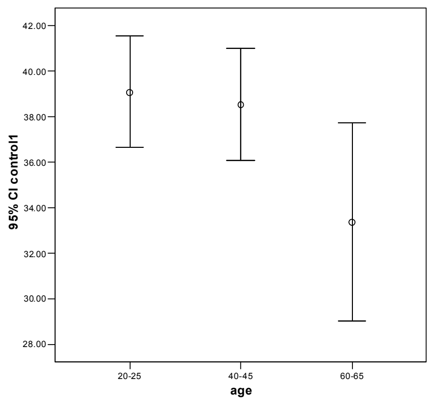
DV must be:

1. Measured at interval or ratio level
2. Normally distributed in all groups of the IV (robust to violations of this assumption if *N*s are large and approximately equal e.g., >15 cases per group)
3. Have approximately equal variance across all groups of the IV (homogeneity of variance)
4. Independence of observations

Example: One-way between groups ANOVA

Does LOC differ across age groups?

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



Descriptives

control1

	N	Mean	Std. Deviation	Std. Error	95% Confi
					Lower Bour
.00 20-25	20	39.1000	5.25056	1.17406	36.642
1.00 40-45	20	38.5500	5.29623	1.18427	36.071
2.00 60-65	20	33.4000	9.29289	2.07795	29.050
Total	60	37.0167	7.24040	.93473	35.146

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			

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

$= 395.433 / 3092.983$

$= 0.128$

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

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

Multiple Comparisons

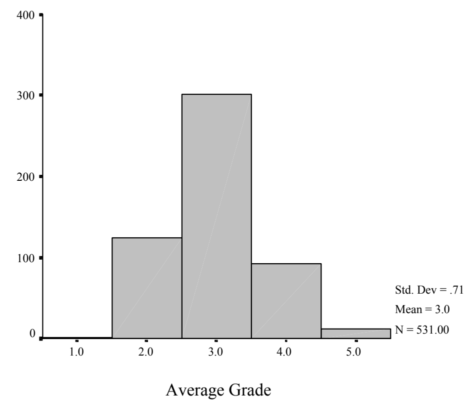
Dependent Variable: control1
Tukey HSD

(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

ONE-WAY ANOVA
Are there differences in Satisfaction levels between students who get different Grades?



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		

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		

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 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)

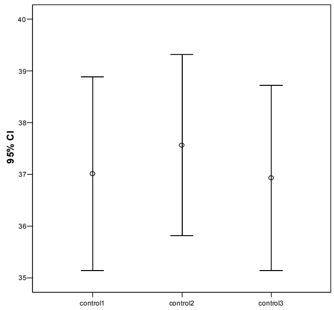
Assumptions - Repeated measures ANOVA

- 1. 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 (Mauchly test of sphericity)
- Note: This assumption is commonly violated, however the multivariate test (provided by default in SPSS output) does not require the assumption of sphericity and may be used as an alternative.
- When results are consistent, not of major concern. When results are discrepant, better to go with MANOVA
- Normality

Example: Repeated measures ANOVA

- Does LOC vary over a period of 12 months?
- LOC measures obtained over 3 intervals: baseline, 6 month follow-up, 12 month follow-up.

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



Descriptive Statistics

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

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

					Epsilon[a]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	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

Tests of Within-Subjects Effects

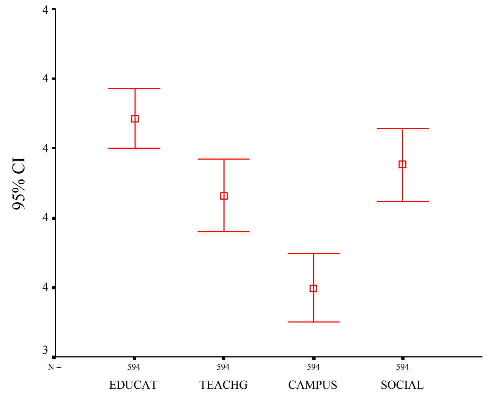
Measure: MEASURE_1

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		

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

Descriptive Statistics

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



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		

Followup Tests

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

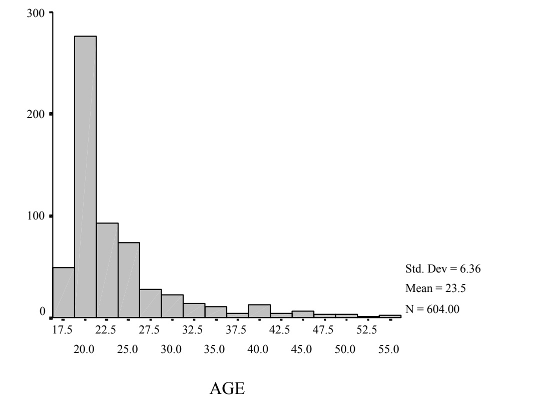
Post hoc

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

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

TWO-WAY ANOVA
Are there differences in Satisfaction levels between Gender and Age?



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

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)

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

TWO-WAY ANOVA
Are there differences in LOC between Gender and Age?

Example: Two-way (factorial) ANOVA

- Main1: Do LOC scores differ by Age?
- Main2: 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)

Example: Two-way (factorial) ANOVA

- Factorial designs test Main Effects and Interactions
- In this example we have two main effects (Age and Gender)
- And one interaction (Age x Gender) potentially explaining variance in the DV (LOC)

Data Structure

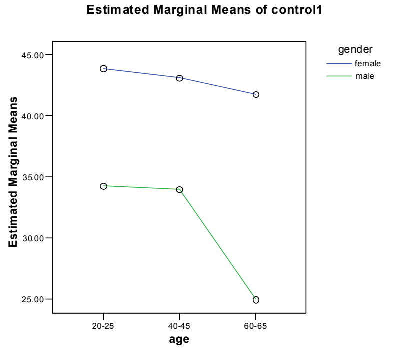
IVs

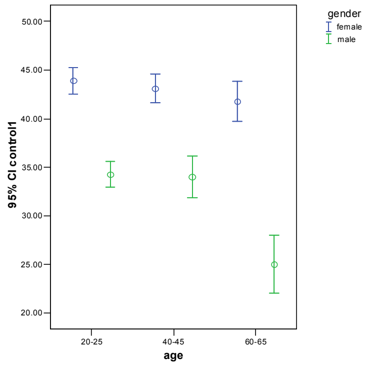
- Age recoded into 3 groups (3)
- Gender dichotomous (2)

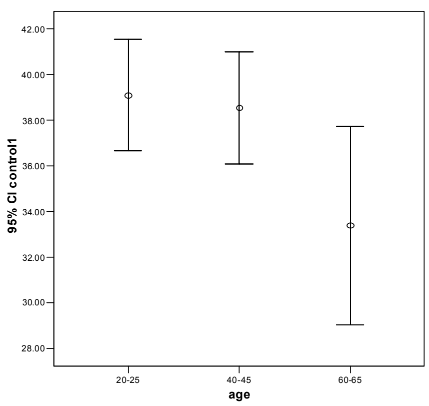
DV

- Locus of Control (LOC)
- Low scores = more internal
- High scores = more external

Plot of LOC by Age and Gender



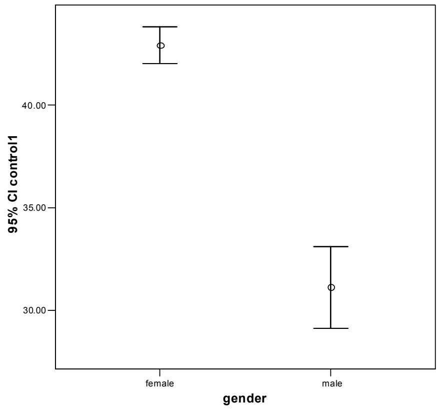




Descriptives

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

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

Descriptive Statistics

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)

Mixed Design ANOVA (SPANOVA)

- It is very common for factorial designs to have within-subject (repeated measures) on *some* (but not all) of their treatment factors.

Mixed Design ANOVA (SPANOVA)

- Since such experiments have mixtures of between subjects and within-subject factors they are said to be of MIXED DESIGN

Mixed Design ANOVA (SPANOVA)

- Common practice to select two samples of subjects
- e.g., Males/Females
    - Winners/Losers

Mixed Design ANOVA (SPANOVA)

- Then perform some repeated measures on each group.
- Males and females are tested for recall of a written passage with three different line spacings

Mixed Design ANOVA (SPANOVA)

- This experiment has two Factors

B/W = Gender (male or Female)

W/I = Spacing (Narrow, Medium, Wide)

- The Levels of Gender vary between subjects, whereas those of Spacing vary within-subjects

CONVENTION

- If A is Gender and B is Spacing the Reading experiment is of the type
- A X (B)
- signifying a mixed design with repeated measures on Factor B

CONVENTION

- With three treatment factors, two mixed designs are possible
- These may be one or two repeated measures
- A X B X (C) or
- A X (B X C)

ASSUMPTIONS

- Random Selection
- Normality
- Homogeneity of Variance
- Sphericity
- Homogeneity of Inter-Correlations

SPHERICITY

- The 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

SPHERICITY

- Is tested by Mauchly’s Test of Sphericity
- If Mauchly’s W Statistic is $p < .05$ then assumption of sphericity is violated

SPHERICITY

- The obtained *F* ratio must then be evaluated against new degrees of freedom calculated from the Greenhouse-Geisser, or Huynh-Feld, Epsilon values.

HOMOGENEITY OF INTERCORRELATIONS

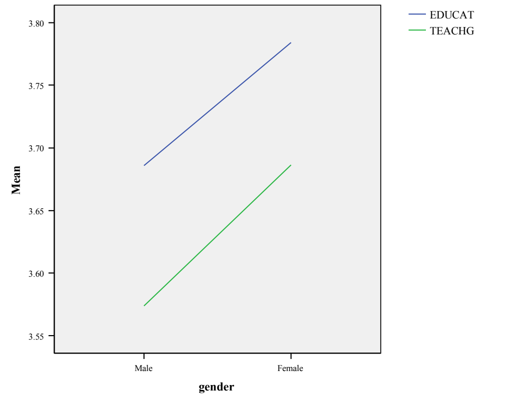
- The pattern of inter-correlations among the various levels of repeated measure factor(s) should be consistent from level to level of the Between-subject Factor(s)

HOMOGENEITY OF INTERCORRELATIONS

- The assumption is tested using Box’s M statistic
- Homogeneity is present when the M statistic is NOT significant at $p > .001$.

Mixed ANOVA or Split-Plot ANOVA

Do Satisfaction levels vary between Gender for Education and Teaching?



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

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

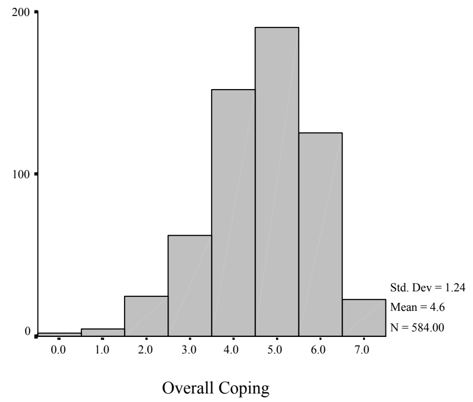
gender	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
0 Male	3.630	.032	3.566	3.693
1 Female	3.735	.029	3.679	3.791

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

ANCOVA
Does Education Satisfaction differ between people who are 'Not coping', 'Just coping' and 'Coping well'?



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		

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

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)

What is ANCOVA?

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

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 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.

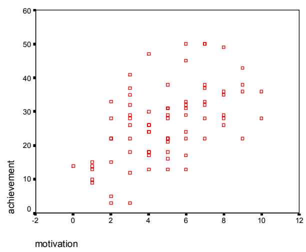
a. Design: Intercept+MOTIV+TEACH

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

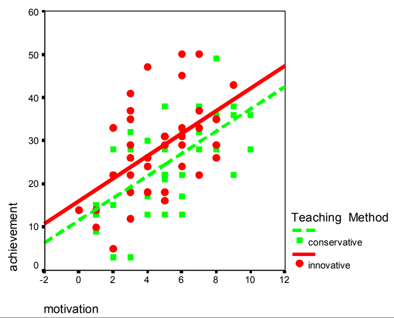
- Check for linearity between CV & DV - check via scatterplot and correlation.



Assumptions of ANCOVA

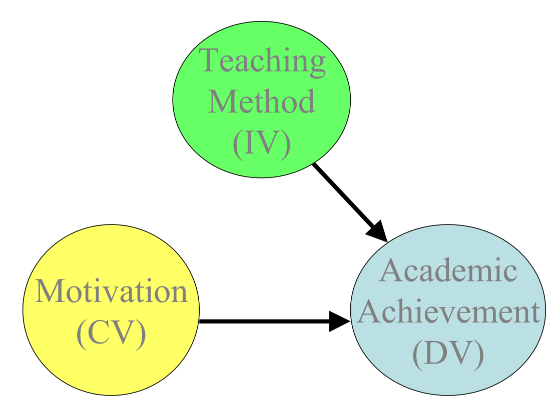
- Homogeneity of regression
  - Estimate regression of CV on DV
  - DV scores & means are adjusted to remove linear effects of CV
  - Assumes slopes of regression lines between CV & DV are equal for each level of IV, if not, don't proceed with ANCOVA
  - Check via scatterplot, lines of best fit.

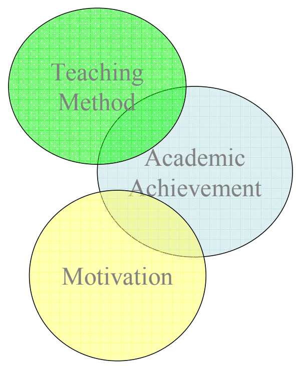
Assumptions of ANCOVA



ANCOVA Example

- 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.





ANCOVA Example

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

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.

ANCOVA & Hierarchical MLR

- Does teaching method affect achievement after controlling for motivation?
  - IV = teaching method
  - DV = achievement
  - CV = motivation
- We could perform hierarchical MLR, with Motivation at step 1, and Teaching Method at step 2.

ANCOVA & Hierarchical MLR

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics R Square Change	F Change	df1
1	.532[a]	.283	.274	9.23685	.283	30.813	
2	.573[b]	.329	.311	8.99722	.045	5.210	

a. Predictors: (Constant), motivation

b. Predictors: (Constant), motivation, dummy coded teaching

ANCOVA & Hierarchical MLR

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2628.975	1	2628.975	30.813	.000[a]
	Residual	6654.913	78	85.319		
	Total	9283.888	79			
2	Regression	3050.744	2	1525.372	18.843	.000[b]
	Residual	6233.143	77	80.950		
	Total	9283.888	79			

a. Predictors: (Constant), motivation

b. Predictors: (Constant), motivation, dummy coded teaching

c. Dependent Variable: achievement

ANCOVA & Hierarchical MLR

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	14.447	2.398		6.024	.000
	motivation	2.465	.444	.532	5.551	.000
2	(Constant)	16.136	2.451		6.585	.000
	motivation	2.593	.436	.560	5.946	.000
	dummy coded teaching	-4.631	2.029	-.215	-2.283	.025

a. Dependent Variable: achievement

1 - Motivation is a sig. predictor of achievement.

2 - Teaching method is a sig, predictor of achievement after controlling for motivation.

ANCOVA Example

- Does employment status affect well-being after controlling for age?
  - IV = Employment status
  - DV = Well-being
  - CV = Age
- Quasi-experimental design - *P*'s not randomly allocated to 'employment status'.

ANCOVA Example

Tests of Between-Subjects Effects

Dependent Variable: Feeling of well being

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	2508.998[a]	3	836.333	8.525	.000	.089
Intercept	248553.317	1	248553.317	2533.544	.000	.906
EMPLOY	2508.998	3	836.333	8.525	.000	.089
Error	25801.609	263	98.105			
Total	420959.000	267				
Corrected Total	28310.607	266				

a. R Squared = .089 (Adjusted R Squared = .078)

- ANOVA - significant effect for employment status

ANCOVA Example

Tests of Between-Subjects Effects

Dependent Variable: Feeling of well being

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	6948.323[a]	4	1737.081	21.305	.000	.245
Intercept	17373.957	1	17373.957	213.085	.000	.449
AGE	4439.325	1	4439.325	54.447	.000	.172
EMPLOY	994.094	3	331.365	4.064	.008	.044
Error	21362.284	262	81.535			
Total	420959.000	267				
Corrected Total	28310.607	266				

a. R Squared = .245 (Adjusted R Squared = .234)

- ANCOVA - employment status remains significant, after controlling for the effect of age.

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

- We can use ANCOVA in survey research when 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 is IV, DV and 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?

Multivariate Analysis of Variance

MANOVA

Generalisation to situation where there are several Dependent Variables.

E.g., Researcher interested in different types of treatment on several types of anxiety.

- Test Anxiety
- Sport Anxiety
- Speaking Anxiety

IV's could be 3 different anxiety interventions:

Systematic Desensitisation

- Autogenic Training
- Waiting List – Control

MANOVA is used to ask whether the three anxiety measures vary overall as a function of the different treatments.

ANOVAs test whether mean differences among groups on a single DV are likely to have occurred by chance.

MANOVA tests whether mean differences among groups on a combination of DV's are likely to have occurred by chance.

MANOVA advantages over ANOVA

1. By measuring several DV's instead of only one the researcher improves the chance of discovering what it is that changes as a result of different treatments and their interactions.

e.g., Desensitisation may have an advantage over relaxation training or control, but only on test anxiety. The effect is missing if anxiety is not one of your DV's.

2. When there are several DV's there is protection against inflated Type 1 error due to multiple tests of likely correlated DV's.

3. When responses to two or more DV's are considered in combination, group differences become apparent.

LIMITATIONS TO MANOVA

As with ANOVA attribution of causality to IV's is in no way assured by the test.

- The best choice is a set of DV's uncorrelated with one another because they each measure a separate aspect of the influence of IV's.
- When there is little correlation among DV's univariate F is acceptable.
- Unequal cell sizes and missing data are problematical for MANOVA.

- Reduced Power can mean a non-significant Multivariate effect but one or more significant Univariate F's!
- When cell sizes of greater than 30 assumptions of normality and equal variances are of little concern.

- Equal cell sizes preferred but not essential but ratios of smallest to largest of 1:1.5 may cause problems.

- MANOVA is sensitive to violations of univariate and multivariate normality. Test each group or level of the IV using the split file option.
- Multivariate outliers which affect normality can normally be identified using Mahalanobis distance in the Regression sub-menu.

- Linearity: Linear relationships among all pairs of DV's must be assumed. Within cell scatterplots must be conducted to test this assumption.

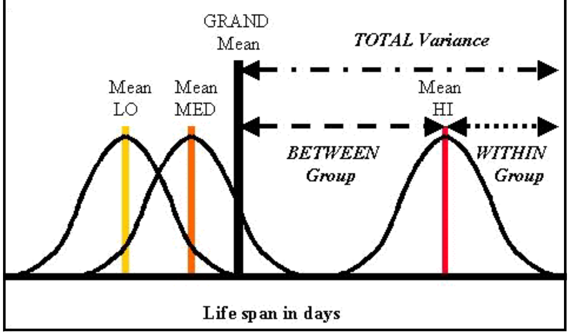
Homogeneity of Regression: It is assumed that the relationships between covariates and DV's in one group is the same as other groups. Necessary if stepdown analyses required.

- Homogeneity of Variance: Covariance Matrices similar to assumption of homogeneity of variance for individual DV's.
- Box's M test is used for this assumption and should be non-significant p>.001.
- Multicollinearity and Singularity: When correlations among DV's are high, problems of multicollinearity exist.

WILKS' LAMBDA

- Several Multi-variate Statistics are available to test significance of Main Effects and Interactions.
- Wilks' Lambda is one such statistic

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



Effect Size: Eta-squared (η^2)

- Analagous to R^2 from regression
- $= SS_{between} / SS_{total} = SS_B / SS_T$
- = proportion of variance in Y explained by X
- = Non-linear correlation coefficient
- = proportion of variance in Y explained by X
- Ranges between 0 and 1
- Interpret as for r^2 or R^2; a rule of thumb: .01 is small, .06 medium, .14 large

Effect Size: Eta-squared (η^2)

- The eta-squared column in SPSS F-table output is actually partial eta-squared (η_p^2).

□ η^2 is not provided by SPSS – calculate separately.

- R^2-squared at the bottom of SPSS F-tables is the linear effect as per MLR – however, if an IV has 3 or more non-interval levels, this won't equate with η^2.

Results - Writing up ANOVA

- Establish clear hypotheses
- Test the assumptions, esp. LOM, normality, univariate and multivariate outliers, homogeneity of variance, *N*
- Present the descriptive statistics (text/table)
- Consider presenting a figure to illustrate the data

Results - Writing up ANOVA

- F results (table/text) and direction of any effects
- Consider power, effect sizes and confidence intervals
- Conduct planned or post-hoc testing as appropriate
- State whether or not results support hypothesis (hypotheses)