SPSS

statstutor

community project

encouraging academics to share statistics support resources

stcp-rothwell-ANCOVAS

The following resources are associated:
ANOVA in SPSS, Checking normality in SPSS and the SPSS dataset 'Diet.sav'

ANCOVA (Analysis of Covariance) in SPSS

Dependent variable: Continuous (scale)

Independent variables: Categorical factors (two or more independent groups), Scale (continuous) covariates

Common Applications: ANCOVA can be considered as an extension of one-way ANOVA. ANCOVA is used to detect a difference in means of 2 or more independent groups, **whilst controlling for scale covariates**. A covariate is not usually part of the main research question but could influence the dependent variable and therefore needs to be controlled for.

Data: The data set 'Diet.sav' contains information on 78 people who undertook one of three diets. There is background information such as age, gender and height as well as weight lost on the diet (a positive value means they lost weight). The aim of the study was to assess which diet increases weight loss so the independent variable (group) is diet.

	Person	gender	Age	Height	preweight	Diet	weight10weeks	weightLOST
1	1	0	22	159	58	1	54.2	3.8
2	2	0	46	192	60	1	54.0	6.0
3	[illegible]	[illegible]	55	170	[illegible]	[illegible]	63.3	[illegible]
4	[illegible]	[illegible]	33	171	[illegible]	[illegible]	61.1	2.9

Female = 0

Diet 1, 2 or 3

Weight lost after 10 weeks

Before carrying out any analysis, summarise weight lost by diet using some summary statistics. For diet 3 the mean weight lost is greater than the other two diets. The standard deviations are similar so weight lost within each group is equally spread out. One could suggest, however, that a person's height will have an added influence in the amount of weight they lose on a particular diet. This is where ANCOVA comes in useful. ANCOVA stands for 'Analysis of covariance', and it combines the methods used in ANOVA with linear regression on a number of different levels. The resulting output shows the effect of the independent variable after the effects of the covariates have been removed/accounted for.

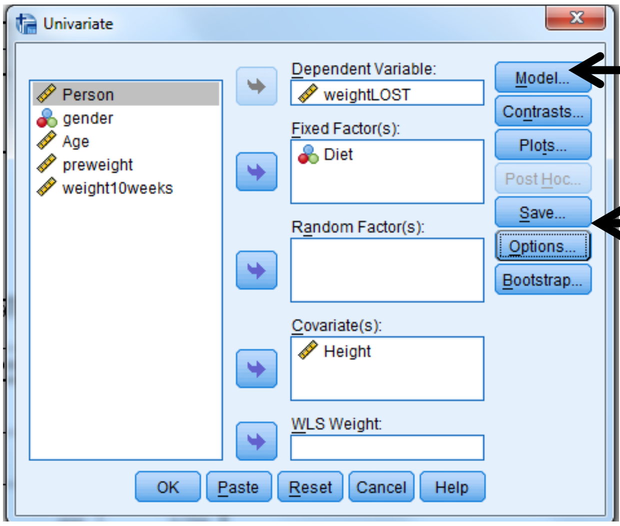
		Diet		
		1	2	3
Weight lost on diet (kg)	Mean	3.30	3.03	5.15
	Standard Deviation	2.24	2.52	2.40

© Joanne Rothwell
University of Sheffield

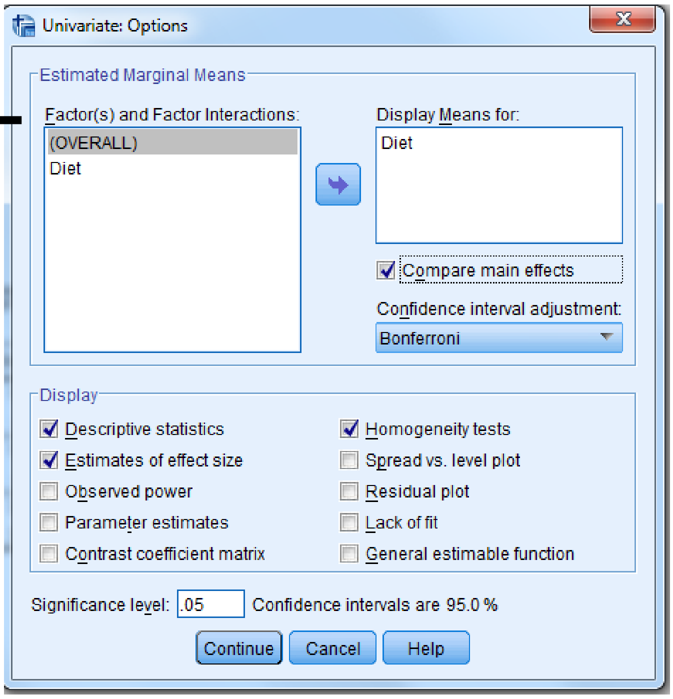
Reviewer: Chris Knox
University of Sheffield

Steps in SPSS

To carry out an ANCOVA, select *Analyze → General Linear Model → Univariate*
Put the dependent variable (weight lost) in the *Dependent Variable* box and the independent variable (diet) in the *Fixed Factors* box. Proceed to put the covariates of interest (height) in the *Covariate(s)* box.



Select the default option *Full Factorial* from the **'Model'** menu



Click on the **Options** button and move the independent variable (diet) over to the *Display Means For* box, click on *Compare main effects* and select *Bonferroni* from the *Confidence interval adjustment* menu to request post hoc tests. Select *Descriptive statistics, Estimates of effect size* and *Homogeneity tests* from the *Display* options.

Tests of Between-Subje[cts Effects]

Dependent Variable: Weight lost on diet (kg)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	71.360[a]	3	23.787	4.0[illegible]4	.010	.142
Intercept	6.945	1	6.945	1.[illegible]5	.278	.016
Height	.266	1	.266	.0[illegible]6	.831	.001
Diet $SS_{Between}$	64.642	2	32.321	5.563	.006	.131
Error SS_{within}	429.913	74	5.810			
Total	1654.350	78				
Corrected Total	501.273	77				

a. R Squared = .142 (Adjusted R Squared = .108)

F = Test statistic
$\frac{MS_{Diet}}{MS_{error}} = \frac{32.321}{5.810} = 5.563$

P = p-value=Sig.
= P(F>5.563) p=0.006

When writing up the results, it is common to report certain figures from the ANCOVA table.

F($df_{between}$, df_{within})= Test Statistic, p = → F(2, 74)=5.563, p=0.006

There is a significant difference in mean weight lost [F(2,74)=5.563, p=0.006] between the diets, whilst adjusting for height. The partial Eta Squared value indicates the effect size and should be compared with Cohen's guidelines (0.2 – small effect, 0.5 – moderate effect, 0.8 – large effect). It can be seen that for Diet the effect size is small (0.13). This value is also used to describe how much of the variance in the dependent variable is explained by the independent variable (13%). Ideally this number would be close to 1.

From this table the influence of the covariate can be determined using the Sig. column.

Post hoc tests

Pairwise Comparisons

Dependent Variable: Weight lost on diet (kg)

(I) Diet	(J) Diet	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	.249	.686	1.000	-1.431	1.930
	3	-1.832*	.681	.026	-3.499	-.165
2	1	-.249	.686	1.000	-1.930	1.431
	3	-2.081*	.684	.010	-3.756	-.406
3	1	1.832*	.681	.026	.165	3.499
	2	2.081*	.684	.010	.406	3.756

Based on estimated marginal means

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

b. Adjustment for multiple comparisons: Bonferroni.

If the main ANOVA is significant, post hoc tests are carried out to see which groups differ. There is a significant difference between diets 1 and 3 ($p = 0.026$) and diets 2 and 3 ($p = 0.01$).

Estimated Marginal Means

The estimated marginal means section of the output gives the adjusted means (controlling for the covariate 'height') for each diet group. This simply means that the effect of 'height' has been statistically removed. From these adjusted means, participants on diet 3 lose the most weight on average after adjusting for height.

Estimates

Dependent Variable: weightLOST

Diet	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Diet 1	3.297[a]	.492	2.316	4.278
Diet 2	3.048[a]	.475	2.101	3.994
Diet 3	5.129[a]	.473	4.187	6.070

a. Covariates appearing in the model are evaluated at the following values: Height = 170.82.

Checking the assumptions for ANCOVA

Assumptions	How to check	What to do if the assumption is not met
Covariates should not be highly correlated (if using more than 1)	Check correlation before performing analysis. Use *Analyse*→ *Correlate*→ *Bivariate* and check that none of the covariates have high correlation values (r>0.8)	If there are some highly correlated covariates, one must select which covariates are of most importance and use those in the model.
Residuals should be normally distributed	Use the **Save** menu within GLM to request the standardised residuals for each subject to be added to the dataset and then use *Analyze* → *Descriptive Statistics* → *Explore* to produce histograms/ QQ plot / Shapiro Wilk tests of residuals.	If the residuals are very skewed, the results of the ANOVA are less reliable. One possible method of solving this issue is transformation of the dependent variable which may help with this assumption violation.
Homogeneity (equality) of variance: The variances (SD squared) should be similar for all the groups.	The Levene's test is carried out if the *Homogeneity of variance test* option is selected in the **Options** menu. If p > 0.05, equal variances can be assumed.	If p < 0.05, the results of the ANOVA are less reliable. One possibility it to transform the data (speak to a statistics tutor for help with this).

Checking the assumptions for this data

Check equality of variances

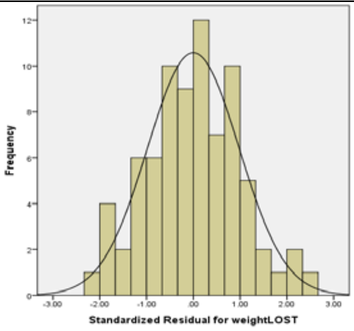
Levene's Test of Equality of Error Variances[a]

Dependent Variable: Weight lost on diet (kg)

F	df1	df2	Sig.
.627	2	75	.537

As p > 0.05, equal variances can be assumed

Check normality



The residuals are normally distributed.

Reporting ANCOVA

A one-way ANCOVA was conducted to compare the effectiveness of three diets whilst controlling for height. Levene's test and normality checks were carried out and the assumptions met.
There was a significant difference in mean weight lost [F(2,74)=5.563, p=0.006] between the diets.
Post hoc tests showed there was a significant difference between diets 1 and 3 (p = 0.026) and diets 2 and 3 (p = 0.01).
Comparing the estimated marginal means showed that the most weight was lost on Diet 3 (mean=5.13kg) compared to Diets 1 and 2 (mean=3.30kg, 3.05kg respectively).