Australian National University

Digital literacy training

SPSS

Advanced Significance Testing
2020

Table of Contents

Presenter: Nyree Mason nyree.mason@anu.edu.au

The IBM Statistical Package for the Social Sciences (SPSS) was developed as a data management and analysis tool. This course will show you how to perform these common statistical analyses:

- t-Tests
- Chi-Square
- Correlation
- Multiple Linear Regression
- Binary Logistic Regression **(OPTIONAL)**
- ANOVA: Between- and Within-Subjects designs
- ANCOVA **(OPTIONAL)**
- Factor Analysis using Principal Components

We may have time at the end of the session to cover any additional statistics you may require.

The data files

ql.anu.edu.au/training

Save these 3 files:

Employee_Data.sav

Anxiety_2.sav

Car_sales.sav

To start SPSS

1. Log on to the computer using university ID (e.g., 'u1234567') and password.
2. On Windows Desktop slide down to find **IBM SPSS Statistics 25**.
3. SPSS **File → Open → Data** look in university ID folder (e.g., 'u1234567')

Note: You can open a file any time within SPSS using **File > Open > Data**. You can also open any file from another **spreadsheet style** application (e.g., Excel) following the SPSS prompts.

Significance testing (Inferential Statistics)

Important: with all significance testing you **MUST** check the underlying assumptions of each statistical test (e.g., normality, homogeneity of variance). A useful online textbook with extensive explanations of statistics and checking their underlying assumptions using SPSS can be found at: http://www.statisticalassociates.com

Note on Terminology: **Dependent variables** (a.k.a. Response) are those with which you expect to measure some effect. **Independent variables** (sometimes called **Factors** in SPSS) are those that you expect to have an effect on the Dependent Variables. For example, if you think males will be taller on average than females, sex will be the Independent Variable and height will be the Dependent Variable.

Dealing with Missing Data

There are two main options for dealing with **Missing Values** in SPSS. This is an important consideration for some Descriptive and all Inferential Statistical analyses. You can choose the appropriate method by pressing the **Options** button for the analysis.

- **Exclude Cases Listwise:** excludes the data from any case that has a missing data point for one or more of the variables included in the analysis list. **This is the default and can result in the loss of a lot of data.** For example, in the data set below, if you conduct a correlation matrix including all 3 variables, SPSS will only include Case 3, because it has no missing data for any of the variables.
- **Exclude Cases Pairwise:** only excludes missing cases on a paired variable by paired variable basis. This preserves as much of the data set as possible. For example, in the data set below, if you conduct a correlation matrix including all 3 variables, SPSS will include Cases 1-3 in the correlation between Variables 1 & 3, Cases 3-6 in the correlation between Variables 2 & 3.

CASE	Variable 1	Variable 2	Variable 3
1	10	-	15
2	12	-	17
3	15	14	14
4	-	12	10
5	-	10	11

t-TESTS

- Useful for testing the significance of the difference between *two means*. There are 3 types of t-tests: **dependent/paired groups** for use when the cases in group 1 are exactly the same as those in group 2 (or are paired in some way, e.g., husband & wife); **independent groups** for use when the cases in group 1 are different to those in group 2; or **one-sample** when you compare a mean from your data to a hypothetical mean (e.g., a mean found in other research).
- Must have a dependent variable of an interval/ratio level of measurement (both called "Scale" in SPSS) and a nominal/ordinal independent variable with only 2 levels (e.g., yes/no) (except for the one-sample t-test of course).
- To compare the education levels of the 2 sexes, **Analyze → Compare Means → Independent Samples t-Test**. In **Test Variable** move current salary into the box, and then make the **Grouping Variable** gender. Click on **Define Groups** and type in "m" to signify males for **group 1**, and "f" for females for **group 2**. Then click **OK** (or **Paste** & **Run**).

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Current Salary	Male	258	$41,441.78	$19,499.214	$1,213.968
	Female	216	$26,031.92	$7,558.021	$514.258

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
Current Salary	Equal variances assumed	119.669	.000	10.945	472	.000	$15,409.862	$1,407.906	$12,643.322	$18,176.401
	Equal variances not assumed			11.688	344.262	.000	$15,409.862	$1,318.400	$12,816.728	$18,002.996

If Levene's Test is significant ($p<0.05$) then equal variances are NOT assumed, so look at the statistics on the bottom row. This table shows that there is a significant difference between the means ($t(470)=8.458$, $p<0.001$). Please note that there are other underlying assumptions to check for a t-test that are not included in the output.

Cohen's D measure of effect size is also very important to assess. Here the effect size is d=1.042 (Large).

$$d = \frac{M_1 - M_2}{SD_{pooled}} \qquad SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

0.2 = Small, 0.5 = Medium, 0.8 = Large

CHI-SQUARE

- Useful for analysing the significance of the relationship between 2 nominal level variables or the difference in frequencies between the categories of one nominal variable (for the latter use the Chi-Square test under **Analyze → Nonparametric Statistics**).
- Must have at least one nominal level variable and the other may be ordinal/nominal (e.g., minority and employment category). Options for two Ordinal variables are Kendall's Tau ("b" for square tables and "c" for rectangular).
- **Analyze → Descriptive Statistics → Crosstabs** list jobcat as the **row** variable and minority as the **column** variable. Click on **Statistics** and select **Chi Square** and **Phi & Cramer's V** (effect size) click **continue**, then **OK** (or **Paste** and **Run)**.

Employment Category * Minority Classification Crosstabulation

			Minority Classification		
			No	Yes	Total
Employment Category	Clerical	Count	276	87	363
		Expected Count	283.4	79.6	363.0
	Custodial	Count	14	13	27
		Expected Count	21.1	5.9	27.0
	Manager	Count	80	4	84
		Expected Count	65.6	18.4	84.0
Total		Count	370	104	474
		Expected Count	370.0	104.0	474.0

Chi Square compares the observed and expected counts to see if there is an overall significant difference = significant relationship.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	26.172[a]	2	.000
Likelihood Ratio	29.436	2	.000
Linear-by-Linear Association	9.778	1	.002
N of Valid Cases	474		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.92.

<20% of cells have an expected count of less than 5, which means it’s a reliable test. There is a significant relationship between Employment Category and Minority Classification ($\chi^2(2)=26.172$, $p<0.001$).

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.235	.000
	Cramer's V	.235	.000
N of Valid Cases		474	

Cramer’s V ranges from 0 to 1, so this is a smallish and significant effect.

CORRELATION

- Useful for analysing the significance of the relationship between two scale level variables.
- To analyse the bivariate relationships between 3 variables (e.g., current salary, education level and previous experience) choose: **Analyse → Correlate → Bivariate** and move the relevant variables into the **Variables** box. Make sure **Pearson’s** is checked, then click **OK** (or **Paste** and **Run**).

Correlations

		Current Salary	Educational Level (years)	Previous Experience (months)
Current Salary	Pearson Correlation	1	.661**	-.097*
	Sig. (2-tailed)		.000	.034
	N	474	474	474
Educational Level (years)	Pearson Correlation	.661**	1	-.252**
	Sig. (2-tailed)	.000		.000
	N	474	474	474
Previous Experience (months)	Pearson Correlation	-.097*	-.252**	1
	Sig. (2-tailed)	.034	.000	
	N	474	474	474

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations range from -1 to +1. So the relationship between current salary and education level is moderate to strong. It is positive and significant ($r(474)=0.661$, $p<0.001$). There is a very small but significant negative correlation between salary and previous experience ($r(474)=-0.097$, $p>0.05$), and a small negative correlation between previous experience and education level ($r(474)= -0.252$, $p<0.001$).

MULTIPLE LINEAR REGRESSION

Same as correlation but used for determining how well one or more variables can predict/explain another.

Must have one scale dependent variable and one or more scale independent variables (or discrete indicator variables [0,1], or discrete ordinal level variables).

Analyze → Regression → Linear and make current salary the **Dependent** variable, and educational level, previous experience and minority classification the **Independent** variables. Make sure Method = **Enter** (so that all independent variables are entered at the same time).

To get regression diagnostics: in **Statistics** select **Colinearity Diagnostics** and **Casewise Diagnostics** then **Continue**. In **Plots** select **Histogram** and **Normal Probability Plot**. Put **ZPRED** into the X-axis box and **ZRESID** into the Y-axis box. Then **Continue**. In **Save** select **Cook's** and **Leverage**. Then **continue** and **OK** or **Paste** and **Run**.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.667[a]	.445	.442	$12,753.772

a. Predictors: (Constant), Minority Classification, Educational Level (years)

b. Dependent Variable: Current Salary

R Square indicates the percentage of variance explained in current salary by the other predictors (44.5%).

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.130E+10	2	3.065E+10	188.444	.000[b]
	Residual	7.661E+10	471	162658704.5		
	Total	1.379E+11	473			

a. Dependent Variable: Current Salary

b. Predictors: (Constant), Minority Classification, Educational Level (years)

The ANOVA table indicates this is a significant model (significant amount of variance explained) ($F(2,471)=188.444$, $p<0.001$).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-16539.241	2885.888		-5.731	.000		
	Educational Level (years)	3838.197	205.095	.648	18.714	.000	.982	1.018
	Minority Classification	-3757.640	1428.168	-.091	-2.631	.009	.982	1.018

a. Dependent Variable: Current Salary

The coefficients table shows that both education level (b = 3838.197, $t(471)= 18.714$, $p<0.001$) and minority classification (b = -3757.640, $t(471)=-2.631$, $p<0.01$) are significant predictors. The intercept (b for constant) is also significant. The Tolerance is quite high (close to 1.0=perfect independence) and Variance Inflation Factor (VIF) is low (<5.0), which indicates no collinearity among predictors (predictors aren't correlated to a large extent which an underlying assumption to be met).

Casewise Diagnostics[a]

Case Number	Std. Residual	Current Salary	Predicted Value	Residual
18	4.617	$103,750	$44,871.91	$58,878.089
29	6.164	$135,000	$56,386.50	$78,613.498
32	4.253	$110,625	$56,386.50	$54,238.498
103	3.184	$97,000	$56,386.50	$40,613.498
218	3.055	$80,000	$41,033.71	$38,966.286
274	3.048	$83,750	$44,871.91	$38,878.089
343	4.597	$103,500	$44,871.91	$58,628.089
446	4.617	$100,000	$41,114.27	$58,885.729

a. Dependent Variable: Current Salary

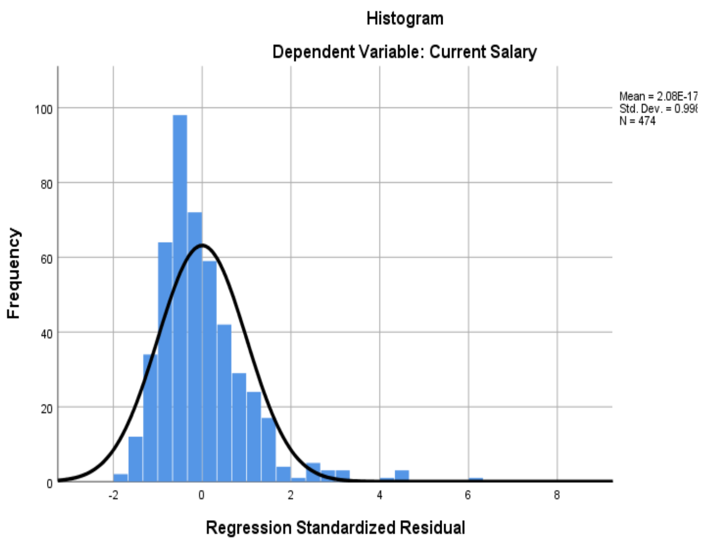
The casewise diagnostics table suggest 8 univariate outliers, the most extreme: case 29.

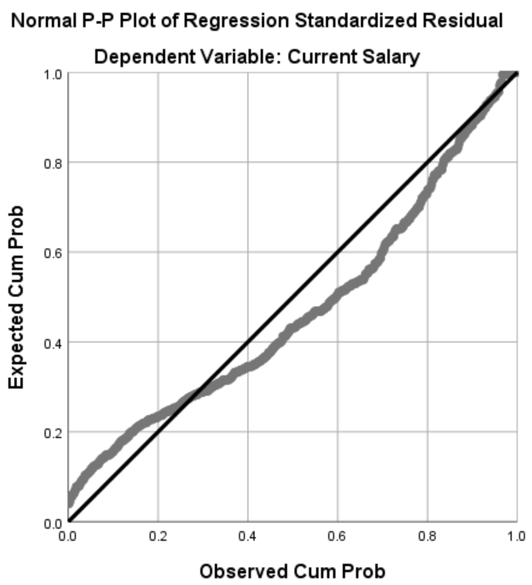
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	$10,408.70	$64,062.89	$34,419.57	$11,384.520	474
Std. Predicted Value	-2.109	2.604	.000	1.000	474
Standard Error of Predicted Value	665.989	1788.029	972.326	290.252	474
Adjusted Predicted Value	$10,029.46	$64,047.17	$34,402.45	$11,393.413	474
Residual	-$22,283.715	$78,613.500	$0.000	$12,726.780	474
Std. Residual	-1.747	6.164	.000	.998	474
Stud. Residual	-1.750	6.195	.001	1.002	474
Deleted Residual	-$22,353.980	$79,406.102	$17.119	$12,824.036	474
Stud. Deleted Residual	-1.754	6.457	.003	1.011	474
Mahal. Distance	.292	8.299	1.996	1.806	474
Cook's Distance	.000	.129	.003	.009	474
Centered Leverage Value	.001	.018	.004	.004	474

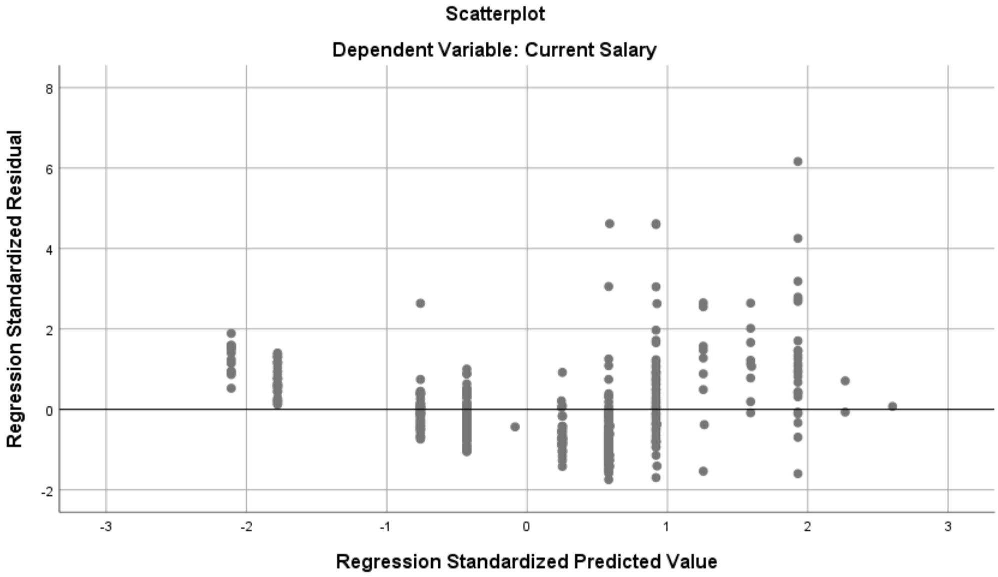
a. Dependent Variable: Current Salary

The table above shows that 2/3 measures indicate no influential multivariate outliers/cases, so the assumption is acceptable. Maximum value for Leverage is less than the critical value of 3[(p+1)/n] = 0.025; for Cook's Distance it is less than the critical value 1; for Mahalanobis' D is greater than $\chi2(3)=7.815$ (the only one to suggests at least one influential outlier).





The Standardised Residuals roughly follow a normal distribution in the Histogram and the line in the Cumulative Probability Plot, therefore the normality assumption is acceptable.



The linearity assumption is not acceptable (there is a curved pattern). The homoscedasticity assumption is also not acceptable (funnel shape = increasing variance, not constant as expected). Ideally the points would fall within +2 or -2 if there were no outliers. The data probably needs transformation and/or generalised linear regression in order to achieve reliable results.

BINARY LOGISTIC REGRESSION (OPTIONAL)

- Same as Regression but used for a binary Dependent Variable (e.g. gender). It tries to find the best linear combination of Independent Variables to explain "group membership".
- Must have one binary dependent variable and one or more scale independent variables (or discrete indicator variables [0,1], or discrete ordinal level variables).
- **Analyze → Regression → Binary Logistic** and make gender the **Dependent** variable, and current salary, educational level and previous experience the **Independent** variables. Make sure Method = **Enter** (so that all independent variables are entered at the same time). If you have any Nominal independent variables, you would select them in **Categorical**.
- In **Save** you can select the same Regression Diagnostics as before. If you select **Group Membership**, this will create a new variable in the data set will the predicted group membership for each case.
- In **Options** select **Hosmer-Lemshow goodness-of-fit** and **Classification plots**. Then **continue** and **OK** (or **Paste** and **Run**).

In **Block 1**:

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	28.922	8	.000

Hosmer and Lemeshow tests the difference between the observed and predicted group membership. A well-fitting model will **not** be significant but this is ($\chi 2(8)=21.522$, $p<0.01$). So it's not a particularly good model.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	470.856[a]	.320	.427

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Nagelkerke R is a pseudo-R^2 value, and ranges from 0 to 1. So, this suggests a moderate effect from all three predictors.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	182.521	3	.000
	Block	182.521	3	.000
	Model	182.521	3	.000

The omnibus test is significant ($\chi2(3)=182.521$, $p<0.001$) which suggests that at least one of the predictors is significantly related to the Dependent variable.

Classification Table[a]

Observed			Predicted sex: Female	Predicted sex: Male	Percentage Correct
Step 1	sex	Female	171	45	79.2
		Male	72	186	72.1
	Overall Percentage				75.3

a. The cut value is .500

This Classification table is an overall indication of how well the model can "predict" gender (75.3% overall and it predicts who is female better than who is male).

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Current Salary	.000	.000	48.558	1	.000	1.000
	Educational Level (years)	.086	.055	2.440	1	.118	1.089
	Previous Experience (months)	.006	.001	30.230	1	.000	1.006
	Constant	-5.535	.733	57.089	1	.000	.004

a. Variable(s) entered on step 1: Current Salary, Educational Level (years), Previous Experience (months).

This table indicates the regression coefficients for each predictor and suggests all Independent variables with the exception of education level are significant predictors of gender.

TWO-WAY ANALYSIS OF VARIANCE (ANOVA)

- Useful for analysing the significance of the differences found between the means for more than two groups. Must have one scale dependent variable (e.g., education level) and at least one nominal/ordinal independent variable with 3 or more levels (e.g., employment category)

- To see if there is a difference between the job categories and minority classifications, and if there's an interaction between the two in terms of Salary: **Analyze → General Linear Model → Univariate** and make salary the **Dependent** variable, and jobcat and minority the **Fixed factors**. In **Options** select **Descriptives, Estimates of effect size** and **Homogeneity tests**, then **Continue**. In **Post Hoc** choose **Tukey** and **Continue**. To create a plot to assess any interaction effect, click on **Plots** and put job category in the **Horizontal Axis** box and minority in the **Separate Lines** box. Tick the **Include Error Bars** box, then click **Add** and **Continue**. Then **OK** (or **Paste** and **Run**).

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

		Levene Statistic	df1	df2	Sig.
Current Salary	Based on Mean	24.720	5	468	.000
	Based on Median	21.271	5	468	.000
	Based on Median and with adjusted df	21.271	5	232.672	.000
	Based on trimmed mean	22.966	5	468	.000

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

a. Dependent variable: Current Salary

b. Design: Intercept + minority + jobcat + minority * jobcat

Levene's test shows that the assumption of homogeneity of variance has been violated based on means (there is a significant difference between the variances $p<0.001$). This means the results from the ANOVA will not be reliable. From the descriptive statistics below, the Management category may be responsible for this problem. There are other statistical methods you can use (e.g., for One-Way ANOVA you can use the Welch test to control for this violation under **Compare Means → One-Way ANOVA**).

Descriptive Statistics

Dependent Variable: Current Salary

Minority Classification	Employment Category	Mean	Std. Deviation	N
No	Clerical	$28,341.09	$7,994.659	276
	Custodial	$31,178.57	$1,658.743	14
	Manager	$63,374.81	$18,164.043	80
	Total	$36,023.31	$18,044.096	370
Yes	Clerical	$26,244.25	$5,772.874	87
	Custodial	$30,680.77	$2,562.920	13
	Manager	$76,037.50	$17,821.961	4
	Total	$28,713.94	$11,421.638	104
Total	Clerical	$27,838.54	$7,567.995	363
	Custodial	$30,938.89	$2,114.616	27
	Manager	$63,977.80	$18,244.776	84
	Total	$34,419.57	$17,075.661	474

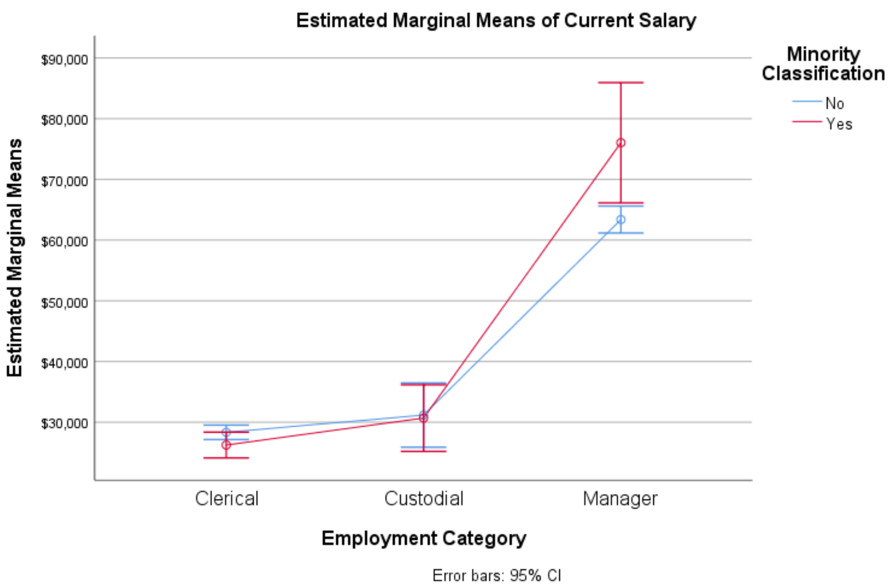
Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	9.034E+10[a]	5	1.807E+10	177.742	.000	.655
Intercept	1.537E+11	1	1.537E+11	1511.773	.000	.764
minority	237964814.4	1	237964814.4	2.341	.127	.005
jobcat	2.596E+10	2	1.298E+10	127.699	.000	.353
minority * jobcat	788578413.1	2	394289206.5	3.879	.021	.016
Error	4.757E+10	468	101655279.9			
Total	6.995E+11	474				
Corrected Total	1.379E+11	473				

a. R Squared = .655 (Adjusted R Squared = .651)

If homogeneity of variance wasn't a problem, the ANOVA table above suggests a significant difference between the average salaries for at least 2 job categories as you would expect ($F(2,468)=127.699$, $p<0.001$). The effect size of the model is moderate ($R^2=0.655$). There is no main effect for minority classification nor for the interaction effect ($p>0.05$), but there is a significant interaction (F(3.879)=p<.0.05). The effect sizes as shown by Partial Eta Squared, suggests there is only a small interaction, and a moderate effect of job category.



The plot shows that the interaction effect is probably only due to differences in management salaries compared to the rest. This needs further investigation with contrasts.

Multiple Comparisons

Dependent Variable: Current Salary
Tukey HSD

(I) Employment Category	(J) Employment Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Clerical	Custodial	-$3,100.35	$2,011.232	.272	-$7,829.14	$1,628.44
	Manager	-$36,139.26*	$1,220.747	.000	-$39,009.47	-$33,269.05
Custodial	Clerical	$3,100.35	$2,011.232	.272	-$1,628.44	$7,829.14
	Manager	-$33,038.91*	$2,230.514	.000	-$38,283.28	-$27,794.54
Manager	Clerical	$36,139.26*	$1,220.747	.000	$33,269.05	$39,009.47
	Custodial	$33,038.91*	$2,230.514	.000	$27,794.54	$38,283.28

Based on observed means.
The error term is Mean Square(Error) = 101655279.939.
*. The mean difference is significant at the .05 level.

The Post-Hoc Tests for Job Category show that there are significant differences between the average salaries for all three job categories.

ANALYSIS OF COVARIANCE (ANCOVA) (OPTIONAL)

- Traditionally useful for analysing the effects of Nominal/Ordinal variables on a Scale Dependent variable while controlling for another Scale Independent variable. OR for analysing the significance of the difference in relationships between more than one group.
- Must have one scale dependent variable (e.g., education level), one scale independent and at least one nominal/ordinal independent variable (e.g., employment category).
- To see if there is a difference between minority classifications in terms of current salary while controlling for education level in years: **Analyze → General Linear Model → Univariate** and make current salary the **Dependent variable**, education level the **Covariate**, and minority the **Fixed factors**. In **Options** select **Descriptives, Estimates of effect size** and **Homogeneity tests**, then **Continue**. In **Model** choose the **Custom** radio button. Select gender and move into the box, then education level. Next, use the **Ctrl** key to highlight both gender and education, then move into the box specifying **Interaction** in the drop-down menu. Then **Continue** and **OK** (or **Paste** and **Run**).

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

Dependent Variable: Current Salary

F	df1	df2	Sig.
9.591	1	472	.002

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

a. Design: Intercept + minority + educ + minority * educ

This indicates that the homogeneity of variance assumption has been violated (F(1,472)=9.591, p<0.01). Therefore, the results of the ANCOVA will not be reliable.

Tests of Between-Subjects Effects

Dependent Variable: Current Salary

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	6.543E+10[a]	3	2.181E+10	141.413	.000	.474
Intercept	678114663.9	1	678114663.9	4.397	.037	.009
minority	3160526518	1	3160526518	20.493	.000	.042
educ	1.933E+10	1	1.933E+10	125.347	.000	.211
minority * educ	4125150351	1	4125150351	26.747	.000	.054
Error	7.249E+10	470	154227871.2			
Total	6.995E+11	474				
Corrected Total	1.379E+11	473				

a. R Squared = .474 (Adjusted R Squared = .471)

The main effects for minority (F(1, 470)=20.493, p<0.001) and education level (F(1, 470)=125.347, p<0.001) are both significant indicating there is a significant difference between the minority classifications and a significant relationship between education level and current salary. There is also a significant interaction effect (F(1, 470)=26.747, p<0.001), which suggests that the relationship between education and current salary is different within the two groups. So, the groups have significantly different regression slopes.

REPEATED MEASURES ANOVA

- Open the anxiety_2.sav file. We'll compare anxiety levels before, during and after the experimental condition. **Analyze → General Linear Model → Repeated Measures** then type in "Time" as the **Within Subject Factor** name. For **number of levels** choose 3 (before, during and after) and click **Add** and then **Define**. As they are all in order, you can highlight all 3 variables at once and then click on the arrow to enter them into the **Within Subjects Variables** box. You can choose a variety of other statistics you may need included in the output by clicking each of the buttons at the right side of the screen. For now, just click **OK**.

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time	.895	2.003	2	.367	.905	.994	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Time

b. 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.

The underlying assumption of sphericity (similar to homogeneity of variance) is acceptable as Mauchly's test is not significant (p>0.05).

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	191.100	2	95.550	51.212	.000	.729
	Greenhouse-Geisser	191.100	1.809	105.612	51.212	.000	.729
	Huynh-Feldt	191.100	1.989	96.086	51.212	.000	.729
	Lower-bound	191.100	1.000	191.100	51.212	.000	.729
Error(Time)	Sphericity Assumed	70.900	38	1.866			
	Greenhouse-Geisser	70.900	34.380	2.062			
	Huynh-Feldt	70.900	37.788	1.876			
	Lower-bound	70.900	19.000	3.732			

There is a significant difference between the 3 means for anxiety ($F(2,38)=51.212$, $p<0.001$) overall. Contrasts will have to be performed to determine where the differences are.

FACTOR ANALYSIS (PRINCIPAL COMPONENTS ANALYSIS)

- Open the **car_sales.sav** file. We'll analyse what aspects of cars seem to "hang together" using Principal Components Analysis because we don't have a causal model in mind: **Analyze → Dimension Reduction → Factor** then add the following into the **Variables** box: vehicle type, price, engine_s, horsepow, wheelbase, width, length, curb_wgt, fuel_cap, mpg and type.
- In **Descriptives** choose **KMO and Bartlett's test of** sphericity the **Continue**. In **Extraction** select **Principal Components** and **Scree plot** then **Continue.** In **Rotation** choose to **Display → Loading plot(s)** and select a **Varimax** rotation method (turns out to be needed for this). In **Scores** select **Save as variables** then **Continue**. Then click **OK** (or **Pate** and **Run**).

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.833
Bartlett's Test of Sphericity	Approx. Chi-Square	1578.819
	df	45
	Sig.	.000

Kaiser-Meyer-Olkin (KMO) is a measure of "factorability". It varies from 0 to 1, and overall should be 0.6 or higher to proceed with Factor Analysis. Bartlett's Test is a similar measure, and if significant the data is factorable also ($\chi 2(45)=1578.819$, $p<0.001$).

Communalities

	Initial	Extraction
Vehicle type	1.000	.930
Price in thousands	1.000	.876
Engine size	1.000	.843
Horsepower	1.000	.933
Wheelbase	1.000	.881
Width	1.000	.776
Length	1.000	.919
Curb weight	1.000	.891
Fuel capacity	1.000	.861
Fuel efficiency	1.000	.860

Extraction Method: Principal Component Analysis.

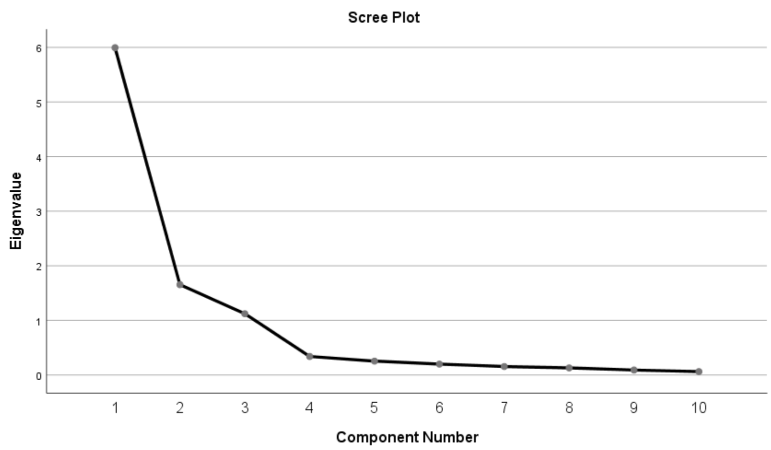
Communalities indicate the amount of variance in each variable that is explained by the others. We want this to be high so these are all good for the model.

Total Variance Explained

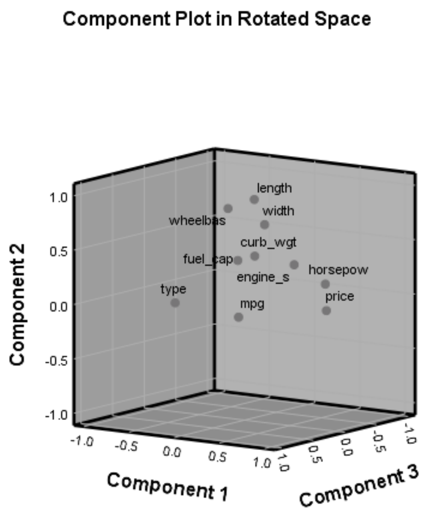
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.994	59.938	59.938	5.994	59.938	59.938	3.220	32.199	32.199
2	1.654	16.545	76.482	1.654	16.545	76.482	3.134	31.344	63.543
3	1.123	11.227	87.709	1.123	11.227	87.709	2.417	24.166	87.709
4	.339	3.389	91.098						
5	.254	2.541	93.640						
6	.199	1.994	95.633						
7	.155	1.547	97.181						
8	.130	1.299	98.480						
9	.091	.905	99.385						
10	.061	.615	100.000						

Extraction Method: Principal Component Analysis.

This shows the amount of variance explained by each Factor. The Factors are not correlated with each other in PCA with Varimax Rotation (orthogonal), therefore this is unique variance. Overall 87.709% of the variance is explained by all 3 Factors.



This is a Scree Plot of the eigenvalues for each component and we want to extract the components on the steep slope (1, 2 and 3).



The Component plot initially does not clearly suggest 3 factors. You can rotate this in 3-Dimensions by double-clicking on the graph and going to **Edit → 3D Rotation** then click on an axis and drag it around to see if 3 factors really do fit the data best.

Rotated Component Matrix[a]

	Component		
	1	2	3
Vehicle type	-.101	.095	**.954**
Price in thousands	**.935**	-.003	.041
Engine size	**.753**	.436	.292
Horsepower	**.933**	.242	.056
Wheelbase	.036	**.884**	.314
Width	.384	**.759**	.231
Length	.155	**.943**	.069
Curb weight	.519	.533	**.581**
Fuel capacity	.398	.495	**.676**
Fuel efficiency	-.543	-.318	**-.681**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

The Rotated Component Matrix helps you interpret what each component represents. Component 1 seems to cluster Price, Engine Size and Horsepower ("prestige"). Component 2 clusters Wheelbase, Width and Length ("size"). Component 3 clusters Type, Curb Weight, Fuel Capacity, Fuel Efficiency ("economy/utility").

Other resources

Training notes

To access training notes, visit the Research & learn webpage anulib.anu.edu.au/research-learn and select the skill area followed by the relevant course. You can register for a workshop and find other information.

Research & learn how-to guides

Explore and learn with the ANU Library's how to guides (anulib.anu.edu.au/howto). Topics covered are:

- Citations & abstracts
- E-books
- EndNote
- Evaluating Sources
- Finding books and more
- Finding journal articles and more
- Finding theses
- Increasing your research impact
- ORCID iD (Open Researcher and Contributor ID)
- Research Data Management
- Text and Data Mining
- Topic analysis

Subject guides

Find subject-specific guides (anulib.anu.edu.au/subjectguides) and resources on broad range of disciplines. Such as:

- Asia Pacific, Southeast Asia and East Asian studies
- Business, economics, art, music and military studies
- Criminal, human rights and taxation law
- History, indigenous studies, linguistics and philosophy
- Biological, environment, physical & mathematical sciences, engineering & computer science, health & medicine

Navigating the sea of scholarly communication

An open access course designed to build the capabilities researchers need to navigate the scholarly communications and publishing world. Topics covered include finding a best-fit publisher, predatory publishing, data citations, bibliometrics, open access, and online research identity. Five self-paced modules, delivered by international and local experts/librarians (anulib.anu.edu.au/publishing).

Online learning

Online learning is available through ANU Pulse, which can be accessed from both on and off campus by all ANU staff and students (ql.anu.edu.au/pulse).

Modules available in ANU Pulse

- Microsoft Office (Access, Excel, OneNote, Outlook, PowerPoint, Project, Visio, Word)
- Microsoft Office (Mac)
- Adobe suite (Illustrator, Photoshop)
- Type IT

Training

A range of workshops are offered to help with your academic research and studies (anulib.anu.edu.au/training-register).

Feedback

Please provide feedback about webinars on the online feedback form (ql.anu.edu.au/libwebinar).