Chapter 8: Regression

Smart Alex's Solutions

Task 1

*In Chapter 3 (Task 6) we looked at data based on findings that the number of cups of tea drunk was related to cognitive functioning (Feng et al., 2010). The data are in the file **Tea Makes You Brainy 716.sav**. Using the model that predicts cognitive functioning from tea drinking, what would cognitive functioning be if someone drank 10 cups of tea? Is there a significant effect?*

Looking at the output below, we can see that we have a useful model, one that significantly improves our ability to predict cognitive functioning. The positive standardized beta value (.078) indicates a positive relationship between number of cups of tea drunk per day and level of cognitive functioning, in that the more tea drunk, the higher your level of cognitive functioning. We can then use the model to predict level of cognitive functioning after drinking 10 cups of tea per day. The first stage is to define the model by replacing the *b*-values in the equation below with the values from the Coefficients output. In addition, we can replace the *X* and *Y* with the variable names so that the model becomes:

$$\begin{aligned}\text{Cognitive functioning}_i &= b_0 + b_1\text{Tea Drinking}_i \\ &= 49.22 + (.460 \times \text{Tea Drinking}_i)\end{aligned}$$

It is now possible to make a prediction about cognitive functioning, by replacing Tea Drinking in the equation with 10:

$$\begin{aligned}\text{Cognitive functioning}_i &= 49.22 + (0.460 \times \text{Tea Drinking}_i) \\ &= 49.22 + (0.460 \times 10) \\ &= 53.82\end{aligned}$$

Therefore, if you drank 10 cups of tea per day, your level of cognitive functioning would be around 53.82.

Descriptive Statistics

	Mean	Std. Deviation	N
Cognitive Function Score (Max = 80)	50.61	9.883	716
Number of Cups of Tea Drunk Per Day	3.03	1.669	716

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	49.218	.764		64.382	.000
	Number of Cups of Tea Drunk Per Day	.460	.221	.078	2.081	.038

a. Dependent Variable: Cognitive Function Score (Max = 80)

Task 2

*Run a regression analysis for the **pubs.sav** data in Jane Superbrain Box 8.1 predicting **mortality** from the number of **pubs**. Try repeating the analysis but bootstrapping the confidence intervals.*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.806[a]	.649	.591	1864.431

a. Predictors: (Constant), Number of Pubs

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	38643389.1	1	38643389.1	11.117	.016[a]
	Residual	20856610.8	6	3476101.80		
	Total	59500000.0	7			

a. Predictors: (Constant), Number of Pubs
b. Dependent Variable: Deaths

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3351.955	781.236		4.291	.005
	Number of Pubs	14.339	4.301	.806	3.334	.016

a. Dependent Variable: Deaths

Bootstrap for Coefficients

Model		B	Bootstrap[a]				
			Bias	Std. Error	Sig. (2-tailed)	95% Confidence Interval	
						Lower	Upper
1	(Constant)	3351.955	-1118.988	1735.337	.024	-9.095E-13	4875.704
	Number of Pubs	14.339	29.225	41.000	.012	10.664	100.000

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Looking at the output tables above, we can see that the number of pubs significantly predicts mortality, $t(6) = 3.33$, $p < .05$. The positive beta value (.806) indicates a positive relationship between number of pubs and death rate in that, the more pubs in an area, the higher the rate of mortality (as we would expect). The value of R^2 tells us that number of pubs accounts for 64.9% of the variance in mortality rate – that's over half!

Looking at the final output table (Bootstrap for Coefficients), we can see that the bootstrapped confidence intervals (I chose percentile for this example) are both positive values – they do not cross zero (10.66, 100.00) – which adds strength to our conclusion that there is a significant positive relationship between number of pubs in an area and its mortality rate.

Task 3

*In Jane Superbrain Box 2.1 we saw some data (**HonestyLab.sav**) relating to people's ratings of dishonest acts and the likeableness of the perpetrator. Run a regression using bootstrapping to predict ratings of dishonesty from the likeableness of the perpetrator.*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.831[a]	.691	.688	1.203

a. Predictors: (Constant), Rating of Likeableness of Perpetrator

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	317.335	1	317.335	219.097	.000[a]
	Residual	141.941	98	1.448		
	Total	459.275	99			

a. Predictors: (Constant), Rating of Likeableness of Perpetrator
b. Dependent Variable: Rating of Deed

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.861	.328		-5.673	.000
	Rating of Likeableness of Perpetrator	.939	.063	.831	14.802	.000

a. Dependent Variable: Rating of Deed

Bootstrap for Coefficients

Model		B	Bootstrap[a]				
			Bias	Std. Error	Sig. (2-tailed)	95% Confidence Interval	
						Lower	Upper
1	(Constant)	-1.861	-.005	.295	.001	-2.502	-1.345
	Rating of Likeableness of Perpetrator	.939	-.001	.066	.001	.817	1.074

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Looking at the output tables above, we can see that the likeableness of the perpetrator significantly predicts ratings of dishonest acts, $t(98) = 14.80$, $p < .001$. The positive standardized beta value (.83) indicates a positive relationship between likeableness of the perpetrator and ratings of dishonesty, in that, the more likeable the perpetrator, the more positively their dishonest acts were viewed (remember that dishonest acts were measured on a scale from 0 = appalling behaviour to 10 = it's OK really). The value of R^2 tells us that likeableness of the perpetrator accounts for 69.1% of the variance in the rating of dishonesty, which is over half!

Looking at the final output table (Bootstrap for Coefficients), we can see that the bootstrapped confidence intervals (I chose percentile for this example) do not cross zero (0.82, 1.07), which gives us confidence in our conclusion that there is a significant relationship between the likeableness of the perpetrator and ratings of dishonest acts.

Additionally, because both of the bootstrapped confidence intervals are positive values, we can conclude that there is a significant *positive* relationship between the likeableness of the perpetrator and ratings of dishonest acts.

Task 4

*A fashion student was interested in factors that predicted the salaries of catwalk models. She collected data from 231 models. For each model she asked them their salary per day on days when they were working (**Salary**), their age (**Age**), how many years they had worked as a model (**Years**), and then got a panel of experts from modelling agencies to rate the attractiveness of each model as a percentage, with 100% being perfectly attractive (**Beauty**). The data are in the file **Supermodel.sav**. Unfortunately, this fashion student bought a substandard statistics textbook and so doesn't know how to analyse her data.☺ Can you help her out by conducting a multiple regression to see which factor predicts a model's salary? How valid is the regression model?*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.429[a]	.184	.173	14.57213	.184	17.066	3	227	.000	2.057

a. Predictors: (Constant), Attractiveness (%), Number of Years as a Model, Age (Years)

b. Dependent Variable: Salary per Day (£)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10871.964	3	3623.988	17.066	.000[a]
	Residual	48202.790	227	212.347		
	Total	59074.754	230			

a. Predictors: (Constant), Attractiveness (%), Number of Years as a Model, Age (Years)

b. Dependent Variable: Salary per Day (£)

To begin with, a sample size of 231 with three predictors seems reasonable because this would easily detect medium to large effects (see the diagram in the chapter).

Overall, the model accounts for 18.4% of the variance in salaries and is a significant fit to the data ($F(3, 227) = 17.07$, $p < .001$). The adjusted R^2 (.17) shows some shrinkage from the unadjusted value (.184), indicating that the model may not generalize well. We can also use Stein's formula:

$$\text{adjusted } R^2 = 1 - \left[\left(\frac{231-1}{231-3-1}\right)\left(\frac{231-2}{231-3-2}\right)\left(\frac{231+1}{231}\right)\right](1-0.184)$$
$$= 1 - 1.031 \times 0.816$$
$$= 1 - 0.841$$
$$= .159$$

This also shows that the model may not cross-generalize well.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	-60.890	16.497		-3.691	.000	-93.396	-28.384		
	Age (Years)	6.234	1.411	.942	4.418	.000	3.454	9.015	.079	12.653
	Number of Years as a Model	-5.561	2.122	-.548	-2.621	.009	-9.743	-1.380	.082	12.157
	Attractiveness (%)	-.196	.152	-.083	-1.289	.199	-.497	.104	.867	1.153

a. Dependent Variable: Salary per Day (£)

In terms of the individual predictors we could report:

	B	*SE B*	β
Constant	–60.89	16.50	
Age	6.23	1.41	.94**
Years as a model	–5.56	2.12	–.55*
Attractiveness	–0.20	0.15	–.08

Note: R^2 = .18 (p < .001). *p < .01, **p < .001.

It seems as though salaries are significantly predicted by the age of the model. This is a positive relationship (look at the sign of the beta), indicating that as age increases, salaries increase too. The number of years spent as a model also seems to significantly predict salaries, but this is a negative relationship indicating that the more years you've spent as a model, the lower your salary. This finding seems very counter-intuitive, but we'll come back to it later. Finally, the attractiveness of the model doesn't seem to predict salaries.

If we wanted to write the regression model, we could write it as:

$$\text{Salary} = \beta_0 + \beta_1\text{Age}_i + \beta_2\text{Experience}_i + \beta_3\text{Attractiveness}_i$$
$$= -60.89 + (6.23\text{Age}_i) - (5.56\text{Experience}_i) - (0.02\text{Attractiveness}_i)$$

The next part of the question asks whether this model is valid.

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	Age (Years)	Number of Years as a Model	Attractiveness (%)
1	1	3.925	1.000	.00	.00	.00	.00
	2	.070	7.479	.01	.00	.08	.02
	3	.004	30.758	.30	.02	.01	.94
	4	.001	63.344	.69	.98	.91	.04

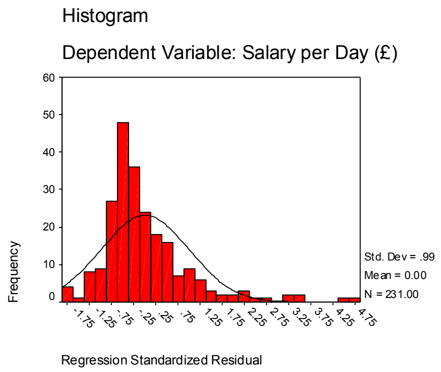
Collinearity Diagnostics[a]

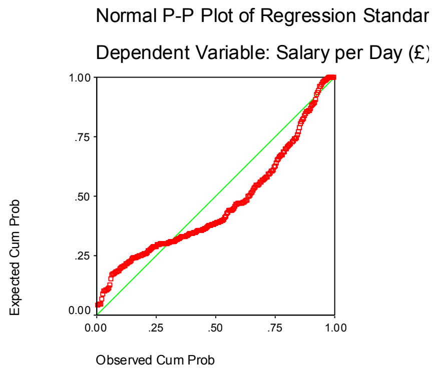
a. Dependent Variable: Salary per Day (£)

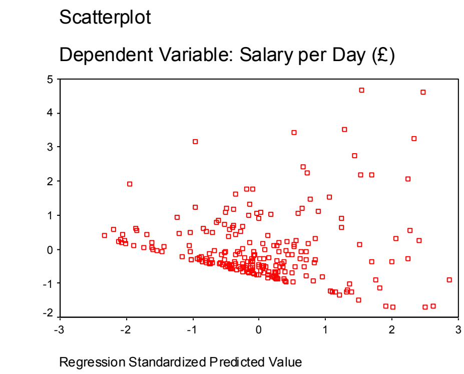
Casewise Diagnostics[a]

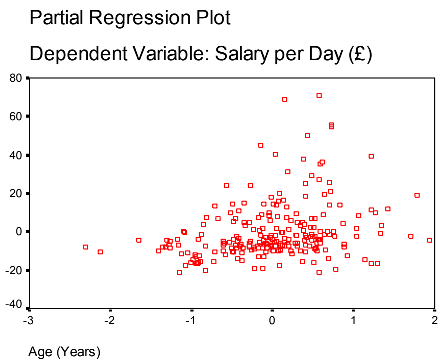
Case Number	Std. Residual	Salary per Day (£)	Predicted Value	Residual
2	2.186	53.72	21.8716	31.8532
5	4.603	95.34	28.2647	67.0734
24	2.232	48.87	16.3444	32.5232
41	2.411	51.03	15.8861	35.1390
91	2.062	56.83	26.7856	30.0459
116	3.422	64.79	14.9259	49.8654
127	2.753	61.32	21.2059	40.1129
135	4.672	89.98	21.8946	68.0854
155	3.257	74.86	27.4025	47.4582
170	2.170	54.57	22.9401	31.6254
191	3.153	50.66	4.7164	45.9394
198	3.510	71.32	20.1729	51.1478

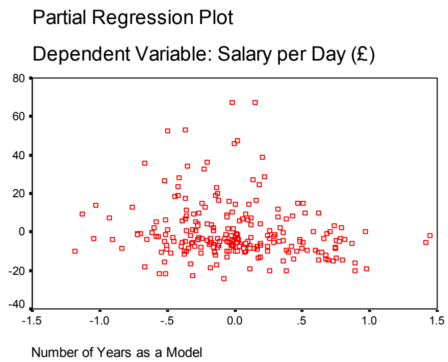
a. Dependent Variable: Salary per Day (£)

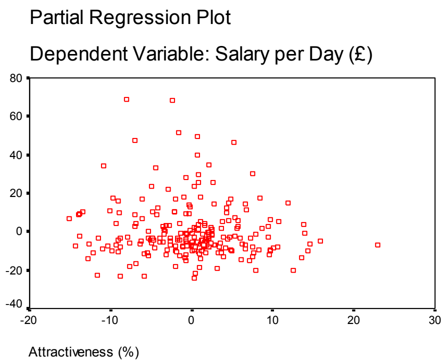












Residuals: There are six cases that have a standardized residual greater than 3, and two of these are fairly substantial (case 5 and 135). We have 5.19% of cases with standardized residuals above 2, so that's as we expect, but 3% of cases with residuals above 2.5 (we'd expect only 1%), which indicates possible outliers.

Normality of errors: The histogram reveals a skewed distribution, indicating that the normality of errors assumption has been broken. The normal P–P plot verifies this because the dashed line deviates considerably from the straight line (which indicates what you'd get from normally distributed errors).

Homoscedasticity and independence of errors: The scatterplot of ZPRED vs. ZRESID does not show a random pattern. There is a distinct funnelling, indicating heteroscedasticity. However, the Durbin–Watson statistic does fall within Field's recommended boundaries of 1–3, which suggests that errors are reasonably independent.

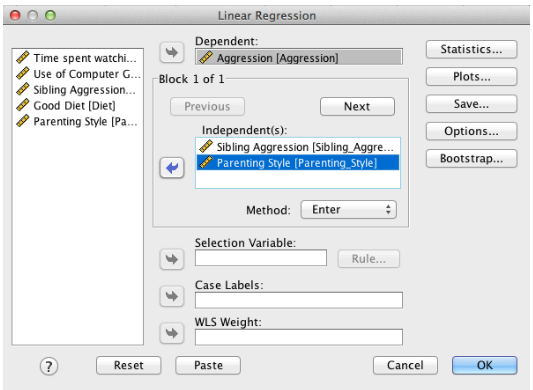
Multicollinearity: For the age and experience variables in the model, VIF values are above 10 (or alternatively, tolerance values are all well below 0.2), indicating multicollinearity in the data. In fact, the correlation between these two variables is around .9! So, these two variables are measuring very similar things. Of course, this makes perfect sense because the older a model is, the more years she would've spent modelling! So, it was fairly stupid to measure both of these things! This also explains the weird result that the number of years spent modelling negatively predicted salary (i.e. more experience = less salary!): in fact if you do a simple regression with experience as the only predictor of salary you'll find it has the expected positive relationship. This hopefully demonstrates why multicollinearity can bias the regression model.

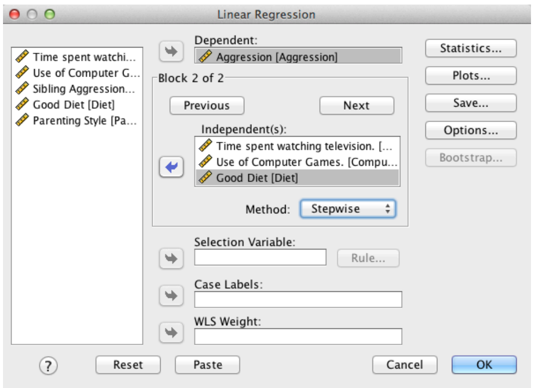
All in all, several assumptions have not been met and so this model is probably fairly unreliable.

Task 5

A study was carried out to explore the relationship between ***Aggression*** *and several potential predicting factors in 666 children who had an older sibling. Variables measured were* ***Parenting_Style*** *(high score = bad parenting practices),* ***Computer_Games*** *(high score = more time spent playing computer games),* ***Television*** *(high score = more time spent watching television),* ***Diet*** *(high score = the child has a good diet low in harmful additives), and* ***Sibling_Aggression*** *(high score = more aggression seen in their older sibling). Past research indicated that parenting style and sibling aggression were good predictors of the level of aggression in the younger child. All other variables were treated in an exploratory fashion. The data are in the file* ***Child Aggression.sav****. Analyse them with multiple regression.*

We need to conduct this analysis hierarchically, entering parenting style and sibling aggression in the first step (forced entry) and the remaining variables in a second step (stepwise):





Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.231[a]	.053	.050	.31125	.053	18.644	2	663	.000	
2	.264[b]	.070	.066	.30875	.017	11.787	1	662	.001	
3	.286[c]	.082	.076	.30697	.012	8.682	1	661	.003	1.911

a. Predictors: (Constant), Parenting Style, Sibling Aggression

b. Predictors: (Constant), Parenting Style, Sibling Aggression, Use of Computer Games.

c. Predictors: (Constant), Parenting Style, Sibling Aggression, Use of Computer Games., Good Diet

d. Dependent Variable: Aggression

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-.006	.012		-.479	.632	-.029	.018					
	Sibling Aggression	.093	.038	.096	2.491	.013	.020	.167	.129	.096	.094	.970	1.031
	Parenting Style	.062	.012	.194	5.057	.000	.038	.086	.211	.193	.191	.970	1.031
2	(Constant)	-.007	.012		-.574	.566	-.030	.017					
	Sibling Aggression	.068	.038	.070	1.793	.073	-.006	.142	.129	.070	.067	.933	1.072
	Parenting Style	.054	.012	.170	4.385	.000	.030	.079	.211	.168	.164	.937	1.067
	Use of Computer Games.	.126	.037	.134	3.433	.001	.054	.197	.186	.132	.129	.918	1.090
3	(Constant)	-.006	.012		-.497	.619	-.029	.017					
	Sibling Aggression	.086	.038	.088	2.258	.024	.011	.161	.129	.087	.084	.908	1.101
	Parenting Style	.062	.013	.194	4.925	.000	.037	.087	.211	.188	.184	.897	1.115
	Use of Computer Games.	.143	.037	.153	3.891	.000	.071	.216	.186	.150	.145	.893	1.120
	Good Diet	-.112	.038	-.118	-2.947	.003	-.186	-.037	-.009	-.114	-.110	.870	1.150

a. Dependent Variable: Aggression

Excluded Variables[d]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Time spent watching television.	.049[a]	1.091	.276	.042	.704	1.421	.704
	Use of Computer Games.	.134[a]	3.433	.001	.132	.918	1.090	.918
	Good Diet	-.092[a]	-2.313	.021	-.090	.894	1.119	.894
2	Time spent watching television.	.044[b]	.986	.324	.038	.703	1.423	.703
	Good Diet	-.118[b]	-2.947	.003	-.114	.870	1.150	.870
3	Time spent watching television.	.032[c]	.715	.475	.028	.697	1.436	.669

a. Predictors in the Model: (Constant), Parenting Style, Sibling Aggression

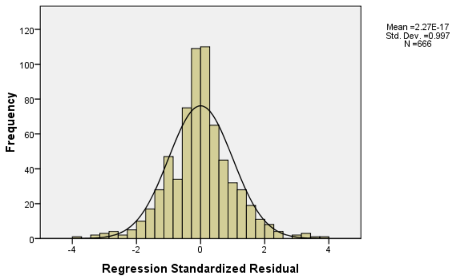
b. Predictors in the Model: (Constant), Parenting Style, Sibling Aggression, Use of Computer Games.

c. Predictors in the Model: (Constant), Parenting Style, Sibling Aggression, Use of Computer Games., Good Diet

d. Dependent Variable: Aggression

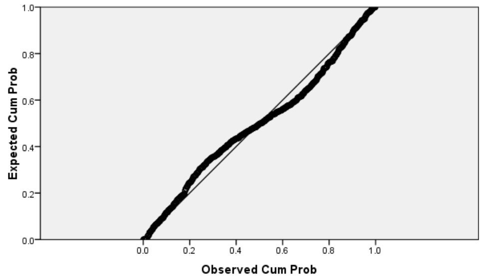
Histogram

Dependent Variable: Aggression



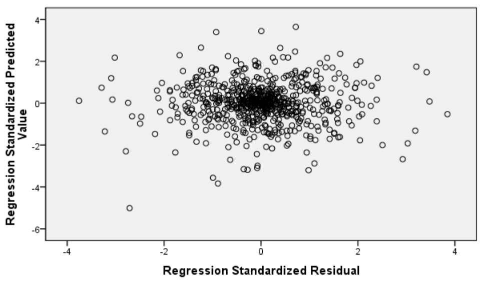
Normal P-P Plot of Regression Standardized Residual

Dependent Variable: Aggression



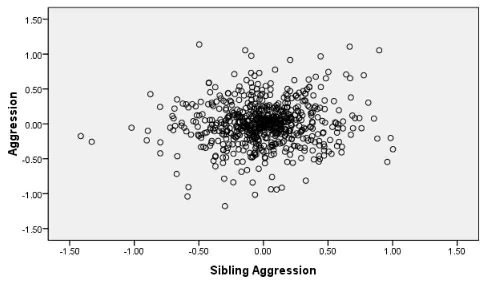
Scatterplot

Dependent Variable: Aggression



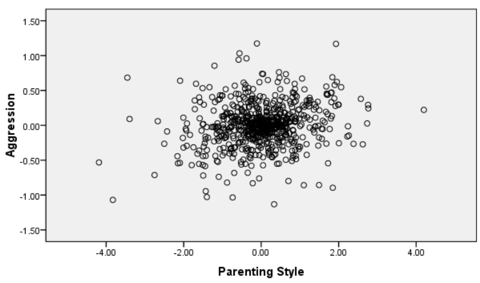
Partial Regression Plot

Dependent Variable: Aggression



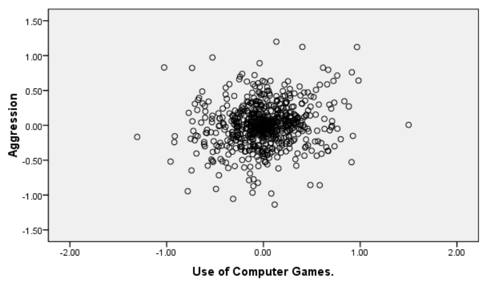
Partial Regression Plot

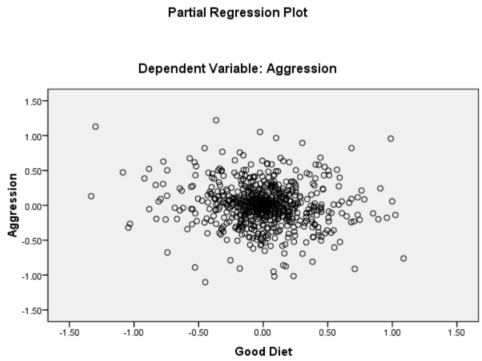
Dependent Variable: Aggression



Partial Regression Plot

Dependent Variable: Aggression





Casewise Diagnostics[a]

Case Number	Std. Residual	Aggression	Predicted Value	Residual
2	2.281	.77	.0710	.70014
45	-3.067	-.93	.0106	-.94162
47	2.405	.84	.1053	.73842
71	-2.496	-.86	-.0942	-.76622
75	2.126	.74	.0849	.65261
157	3.845	1.13	-.0529	1.18037
163	-2.084	-.68	-.0423	-.63962
169	3.182	.85	-.1251	.97673
182	2.051	.81	.1775	.62946
199	2.505	.58	-.1879	.76897
200	3.026	.75	-.1805	.92899
204	2.080	.63	-.0120	.63837
217	-2.712	-1.30	-.4630	-.83263
221	3.205	1.14	.1543	.98372
266	2.085	.59	-.0533	.64012
270	-3.018	-.73	.1936	-.92649
351	2.386	.74	.0101	.73259
374	2.923	.65	-.2495	.89716
375	2.263	.68	-.0170	.69483
379	-2.789	-1.07	-.2150	-.85618
386	2.388	.65	-.0841	.73290
407	-2.148	-.61	.0502	-.65934
411	-2.188	-.81	-.1394	-.67154
421	-2.045	-.54	.0833	-.62772
431	-2.472	-.82	-.0643	-.75895
439	-3.092	-.85	.1041	-.94922
440	-3.290	-.95	.0624	-1.00982
463	-3.756	-1.15	.0055	-1.15286
482	3.476	1.07	.0025	1.06707
505	-3.223	-1.12	-.1284	-.98938
539	3.416	1.18	.1300	1.04877
589	2.042	.46	-.1671	.62679
630	-2.119	-.63	.0169	-.65047
635	-2.661	-.88	-.0625	-.81672
639	-2.743	-.85	-.0037	-.84210
640	2.024	.56	-.0629	.62135

a. Dependent Variable: Aggression

Based on the final model (which is actually all we're interested in) the following variables predict aggression:

- ✓ Parenting style (b = 0.062, β = 0.194, t = 4.93, p < .001) significantly predicted aggression. The beta value indicates that as parenting increases (i.e. as bad practices increase), aggression increases also.
- ✓ Sibling aggression (b = 0.086, β = 0.088, t = 2.26, p < .05) significantly predicted aggression. The beta value indicates that as sibling aggression increases (became more aggressive), aggression increases also.
- ✓ Computer games (b = 0.143, β = 0.037, t = 3.89, p < .001) significantly predicted aggression. The beta value indicates that as the time spent playing computer games increases, aggression increases also.
- ✓ Good diet (b = –0.112, β = –0.118, t = –2.95, p < .01) significantly predicted aggression. The beta value indicates that as the diet improved, aggression decreased.

The only factor not to predict aggression was:

- ✗ Television (b if entered = 0.032, t = 0.72, p > .05) did not significantly predict aggression.

Based on the standardized beta values, the most substantive predictor of aggression was actually parenting style, followed by computer games, diet and then sibling aggression.

R^2 is the squared correlation between the observed values of aggression and the values of aggression predicted by the model. The values in this output tell us that sibling aggression and parenting style in combination explain 5.3% of the variance in aggression. When computer game use is factored in as well, 7% of variance in aggression is explained (i.e. an additional 1.7%). Finally, when diet is added to the model, 8.2% of the variance in aggression is explained (an additional 1.2%). With all four of these predictors in the model still less than half of the variance in aggression can be explained.

The Durbin–Watson statistic tests the assumption of 'independence of errors', which means that for any two observations (cases) in the regression, their residuals should be uncorrelated (or independent). In this output the Durbin–Watson statistic falls within the recommended boundaries of 1–3, which suggests that errors are reasonably independent.

The scatterplot helps us to assess both *homoscedasticity* and *independence of errors*. The scatterplot of ZPRED vs. ZRESID does show a random pattern and so indicates no violation of the independence of errors assumption. Also, the errors on the scatterplot do not funnel out, indicating homoscedasticity of errors, thus no violations of these assumptions.

Task 6

Repeat the analysis in Labcoat Leni's Real Research 8.1 using bootstrapping for the confidence intervals. What are the confidence intervals for the regression parameters?

Recap of Labcoat Leni 8.1

Ong, et al. (2011). *Personality and Individual Differences*, 50(2), 180-185.

Social media websites such as Facebook seem to have taken over the world. These websites offer an unusual opportunity to carefully manage your self-presentation to others (i.e., you can try to appear to be cool when in fact you write statistics books, appear attractive when you have huge pustules all over your face, fashionable when you wear 1980s heavy metal band T-shirts, and so on). Ong et al. (2011) condcted an interesting study that examined the relationship between narcissism and behaviour on Facebook in 275 adolescents. They measured the **Age**, **Gender** and **Grade** (at school), as well as extroversion and narcissism. They also measured how often (per week) these people updated their Facebook status (**FB_Status**), and also how they rated their own profile picture on each of four dimensions: coolness, glamour, fashionableness and attractiveness. These ratings were summed as an indicator of how positively they perceived the profile picture they had selected for their page (**FB_Profile_TOT**). They hypothesized that narcissism would predict, above and beyond the other variables, the frequency of status updates, and how positive a profile picture the person chose. To test this, they conducted two hierarchical regressions: one with **FB_Status** as the outcome and one with **FB_Profile_TOT** as the outcome. In both models they entered **Age**, **Gender** and **Grade** in the first block, then added extroversion (**NEO_FFI**) in a second block, and finally narcissism (**NPQC_R**) in a third block. The data from this study are in the file **Ong et al. (2011).sav**. Labcoat Leni wants you to replicate their two hierarchical regressions and create a table of the results for each.

OK, so I have already shown you how to run the two regressions (see dialog boxes in the solution to Labcoat Leni 8.1). Run these regressions again, but this time clicking on the Bootstrap... button and select ☑ Perform bootstrapping to activate bootstrapping, and to get a 95% confidence interval click ◉ Percentile or ◉ Bias corrected accelerated (BCa). For this analysis, let's ask for a bias corrected (BCa) confidence interval. The other thing is that bootstrapping doesn't appear to work if you ask SPSS to save diagnostics, therefore, click on Save... to open the dialog box and *make sure that everything is deselected*. Back in the main dialog box, click on OK to run the analysis.

If you look at the bootstrapped confidence intervals for the first regression (see output below) you will see that they don't change the results as reported in Ong et al. (2011). The main benefit of the bootstrap confidence intervals and significance values is that they do not rely on assumptions of normality or homoscedasticity, so they give us an accurate estimate of the true population value of *b* for each predictor.

Bootstrap for Coefficients

Model		B	Bootstrap[a]				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
						Lower	Upper
1	(Constant)	3.383	-.176	1.993	.084	-.330	6.652
	Gender	-.775	-.010	.320	.023	-1.418	-.183
	Age	-.033	.016	.172	.826	-.398	.372
	Grade	-.444	-.022	.282	.107	-.978	.031
2	(Constant)	.830	-.226	2.480	.710	-4.463	5.008
	Gender	-.691	-.009	.307	.027	-1.290	-.115
	Age	-.006	.018	.177	.968	-.360	.428
	Grade	-.486	-.022	.281	.079	-1.031	.011
	Extraversion - Total	.052	.000	.029	.076	-.007	.113
3	(Constant)	.650	-.127	2.418	.775	-4.422	5.198
	Gender	-.943	-.009	.312	.004	-1.571	-.321
	Age	-.010	.010	.173	.944	-.362	.357
	Grade	-.522	-.012	.274	.054	-1.057	-.034
	Extraversion - Total	.011	.000	.029	.716	-.049	.072
	NPQC-R Total	.066	3.575E-005	.020	.002	.025	.107

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

So basically, Ong et al.'s prediction was still supported in that, after controlling for age, grade and gender, narcissism significantly predicted the frequency of Facebook status updates over and above extroversion. b = .21 [.025, .107], p < .01.

If you look at the bootstrapped confidence intervals for the second regression (table below), you will see that they also do not change the results as reported in Ong et al. (2011).

Bootstrap for Coefficients

Model		B	Bootstrap[a]				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
						Lower	Upper
1	(Constant)	8.782	-.932	6.564	.155	-5.025	18.392
	Gender	1.290	-.021	.592	.037	.218	2.336
	Age	.150	.081	.542	.767	-.766	1.559
	Grade	.099	-.095	.618	.865	-1.053	1.044
2	(Constant)	-3.461	-1.179	7.883	.652	-19.048	7.992
	Gender	1.475	-.018	.551	.007	.473	2.447
	Age	.365	.091	.594	.527	-.690	1.810
	Grade	-.245	-.105	.658	.711	-1.429	.706
	Extraversion - Total	.224	.003	.042	.001	.141	.325
3	(Constant)	-3.169	-.923	6.674	.622	-16.335	6.456
	Gender	.582	-.012	.609	.335	-.554	1.706
	Age	.337	.071	.504	.493	-.521	1.542
	Grade	-.258	-.085	.578	.662	-1.262	.610
	Extraversion - Total	.104	.005	.047	.031	.014	.211
	NPQC-R Total	.173	-.003	.036	.001	.105	.231

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

These results show that after controlling for age, grade and gender, narcissism significantly predicted the Facebook profile picture ratings over and above extroversion, *b* = 0.37 [0.105, 0.23], *p* = .01.

Task 7

*Coldwell, Pike and Dunn (2006) investigated whether household chaos predicted children's problem behaviour over and above parenting. From 118 families they recorded the age and gender of the younger sibling (**Child_age** and **Child_gender**). They then interviewed the child about their relationship with their mum using the Berkeley Puppet Interview (BPI), which measures (1) warmth/enjoyment (**Child_warmth**), and (2) anger/hostility (**Child_anger**). Higher scores indicate more anger/hostility and warmth/enjoyment, respectively. Each mum was interviewed about their relationship with the child resulting in scores for relationship positivity (**Mum_pos)** and relationship negativity (**Mum_neg**). Household chaos (**Chaos**) was assessed using the Confusion, Hubbub, and Order Scale. The outcome variable was the child's adjustment (**sdq**): the higher the score, the more problem behaviour the child is reported to be displaying. The data are in the file **Coldwell et al. (2006).sav**. Conduct a hierarchical regression in three steps: (1) enter child age and gender; (2) add the variables measuring parent–child positivity, parent–child negativity, parent–child warmth and parent–child anger; (3) add chaos. Is household chaos predictive of children's problem behaviour over and above parenting?*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.076[a]	.006	−.016	.17627	.006	.273	2	93	.762
2	.256[b]	.065	.002	.17471	.060	1.418	4	89	.235
3	.331[c]	.110	.039	.17149	.044	4.373	1	88	.039

a. Predictors: (Constant), Gender of child 2, age of younger child
b. Predictors: (Constant), Gender of child 2, age of younger child, mum PC Relationship POS Child 2, BPI mum anger & hostility Child 2, mum PC relationship Negative Child 2, BPI mum warmth + enjoyment Child 2
c. Predictors: (Constant), Gender of child 2, age of younger child, mum PC Relationship POS Child 2, BPI mum anger & hostility Child 2, mum PC relationship Negative Child 2, BPI mum warmth + enjoyment Child 2, CHAOS -- Mum & Dad combined report

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.017	2	.008	.273	.762[a]
	Residual	2.890	93	.031		
	Total	2.907	95			
2	Regression	.190	6	.032	1.038	.406[b]
	Residual	2.717	89	.031		
	Total	2.907	95			
3	Regression	.319	7	.046	1.548	.162[c]
	Residual	2.588	88	.029		
	Total	2.907	95			

a. Predictors: (Constant), Gender of child 2, age of younger child
b. Predictors: (Constant), Gender of child 2, age of younger child, mum PC Relationship POS Child 2, BPI mum anger & hostility Child 2, mum PC relationship Negative Child 2, BPI mum warmth + enjoyment Child 2
c. Predictors: (Constant), Gender of child 2, age of younger child, mum PC Relationship POS Child 2, BPI mum anger & hostility Child 2, mum PC relationship Negative Child 2, BPI mum warmth + enjoyment Child 2, CHAOS -- Mum & Dad combined report
d. Dependent Variable: Mum & Dad sdq all items child 2

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.237	.155		7.978	.000
	age of younger child	.001	.002	.027	.266	.791
	Gender of child 2	.025	.036	.073	.700	.485
2	(Constant)	.923	.334		2.763	.007
	age of younger child	.001	.002	.064	.608	.545
	Gender of child 2	.025	.036	.071	.682	.497
	BPI mum anger & hostility Child 2	.007	.017	.041	.388	.699
	BPI mum warmth + enjoyment Child 2	-.006	.025	-.026	-.232	.817
	mum PC Relationship POS Child 2	.003	.005	.050	.471	.639
	mum PC relationship Negative Child 2	.013	.006	.236	2.184	.032
3	(Constant)	.741	.339		2.184	.032
	age of younger child	.002	.002	.070	.675	.501
	Gender of child 2	.026	.036	.075	.736	.464
	BPI mum anger & hostility Child 2	.003	.017	.021	.197	.844
	BPI mum warmth + enjoyment Child 2	.001	.025	.002	.021	.984
	mum PC Relationship POS Child 2	.002	.005	.046	.444	.658
	mum PC relationship Negative Child 2	.011	.006	.203	1.887	.062
	CHAOS -- Mum & Dad combined report	.075	.036	.218	2.091	.039

a. Dependent Variable: Mum & Dad sdq all items child 2

Looking at the output tables above, we can conclude that household chaos significantly predicted younger sibling's problem behaviour over and above maternal parenting, child age and gender, $t(88) = 2.09$, $p < .05$. The positive standardized beta value (.218) indicates that there is a positive relationship between household chaos and child's problem behaviour. In other words, the higher the level of household chaos, the more problem behaviours the child displayed. The value of R^2 (.11) tells us that household chaos accounts for 11% of the variance in child problem behaviour.