Chapter 6: Answers

Task 1

Recent research has shown that lecturers are among the most stressed workers. A researcher wanted to know exactly what it was about being a lecturer that created this stress and subsequent burnout. She took 467 lecturers and administered several questionnaires to them that measured: **Burnout** (burnt out or not), **Perceived Control** (high score = low perceived control), **Coping Style** (high score = low ability to cope with stress), **Stress from Teaching** (high score = teaching creates a lot of stress for the person), **Stress from Research** (high score = research creates a lot of stress for the person), and **Stress from Providing Pastoral Care** (high score = providing pastoral care creates a lot of stress for the person). The outcome of interest was burnout, and Cooper's (1988) model of stress indicates that perceived control and coping style are important predictors of this variable. The remaining predictors were measured to see the unique contribution of different aspects of a lecturer's work to their burnout—can you help her out by conducting a logistic regression to see which factor predict burnout? The data are in **Burnout.sav**.

Test

The analysis should be done hierarchically because Cooper's model indicates that perceived control and coping style are important predictors of burnout. So, these variables should be entered in the first block. The second block should contain all other variables and because we don't know anything much about their predictive ability, we should enter them in a stepwise fashion (I chose Forward: LR).

SPSS Output

Step 1:

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	165.928	2	.000
	Block	165.928	2	.000
	Model	165.928	2	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	364.179	.299	.441

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	LOC	.061	.011	31.316	1	.000	1.063	1.040	1.086
	COPE	.083	.009	77.950	1	.000	1.086	1.066	1.106
	Constant	-4.484	.379	139.668	1	.000	.011		

a. Variable(s) entered on step 1: LOC, COPE.

The overall fit of the model is significant both at the first step, $\chi^2(2) = 165.93$, $p < .001$.

Overall, the model accounts for 29.9 – 44.1% of the variance in burnout (depending on which measure R^2 you use).

Step 2:

The overall fit of the model is significant after both at the first new variable (teaching), $\chi^2(3) = 193.34$, $p < .001$, and second new variable (pastoral) have been entered, $\chi^2(4) = 205.40$, $p < .001$

Overall, the final model accounts for 35.6 – 52.4% of the variance in burnout (depending on which measure R^2 you use.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	27.409	1	.000
	Block	27.409	1	.000
	Model	193.337	3	.000
Step 2	Step	12.060	1	.001
	Block	39.470	2	.000
	Model	205.397	4	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	336.770	.339	.500
2	324.710	.356	.524

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	LOC	.092	.014	46.340	1	.000	1.097	1.068	1.126
	COPE	.131	.015	76.877	1	.000	1.139	1.107	1.173
	TEACHING	-.083	.017	23.962	1	.000	.921	.890	.952
	Constant	-1.707	.619	7.599	1	.006	.181		
Step 2[b]	LOC	.107	.015	52.576	1	.000	1.113	1.081	1.145
	COPE	.135	.016	75.054	1	.000	1.145	1.110	1.181
	TEACHING	-.110	.020	31.660	1	.000	.896	.862	.931
	PASTORAL	.044	.013	11.517	1	.001	1.045	1.019	1.071
	Constant	-3.023	.747	16.379	1	.000	.049		

a. Variable(s) entered on step 1: TEACHING.

b. Variable(s) entered on step 2: PASTORAL.

In terms of the individual predictors we could report:

	B (*SE*)	95% CI for *Exp(B)*		
		Lower	Exp(β)	*Upper*
Step 1				
Constant	−4.48** (0.38)			
Perceived Control	0.06** (0.01)	1.04	1.06	1.09
Coping Style	0.08** (0.01)	1.07	1.09	1.11
Final				
Constant	−3.02** (0.75)			
Perceived Control	0.11** (0.02)	1.08	1.11	1.15
Coping Style	0.14** (0.02)	1.11	1.15	1.18
Teaching Stress	−0.11** (0.02)	0.86	0.90	0.93
Pastoral Stress	0.04* (0.01)	1.02	1.05	1.07

Note. R^2 = .36 (Cox & Snell), .52 (Nagelkerke). Model $\chi^2(4)$ = 205.40, p < .001. * p < .01, ** p < .001.

It seems as though burnout is significantly predicted by perceived control, coping style (as predicted by Cooper), stress from teaching and stress from giving pastoral care. The *Exp(B)* and direction of the beta values tells us that for perceived control, coping ability and pastoral care the relationships are positive. That is (and look back to the question to see the direction of these scales, i.e. what a high score represents), poor perceived control, poor ability to cope with stress and stress from giving pastoral care all predict burnout. However, for teaching, the relationship if the opposite way around: stress from teaching appears to be a positive thing as it predicts not becoming burnt out!

Task 2

A Health Psychologist interested in research into HIV wanted to know the factors that influenced condom use with a new partner (relationship less than 1 month old). The outcome measure was whether a condom was used (Use: condom used = 1, Not used = 0). The predictor variables were mainly scales from the Condom Attitude Scale (CAS) by Sacco, Levine, Reed and Thompson (Psychological Assessment: A journal of Consulting and Clinical Psychology, 1991). **Gender** (gender of the person); **Safety** (relationship safety, measured out of 5, indicates the degree to which the person views this relationship as 'safe' from sexually transmitted disease); **Sexexp** (sexual experience, measured out of 10, indicates the degree to which previous experience influences attitudes towards condom use); **Previous** (a measure not from the CAS, this variable measures whether or not the couple used a condom in their previous encounter, 1 = condom used, 0 = not used, 2 = no previous encounter with this partner); **selfcon** (self-control, measured out of 9, indicates the degree of self-control that a subject has when it comes to condom use, i.e., do they get carried away with the heat of the moment, or do they exert control); **Perceive** (perceived risk, measured out of 6, indicates the degree to which the person feels at risk from unprotected sex). Previous Research (Sacco, Rickman, Thompson, Levine and Reed, in Aids Education and Prevention, 1993) has shown that **gender**, **relationship safety** and **perceived risk** predict condom use. Carry out an appropriate analysis to verify these previous findings, and to test whether Self-control, Previous Usage and Sexual Experience can predict any of the remaining variance in condom use. (1) Interpret all important parts of the SPSS output; (2) How reliable is the final model? (3) What are the probabilities that participants 12, 53 and 75 will used a condom?; and (4) a female, who used a condom in her previous encounter with her new partner, scores 2 on all variables except perceived risk (for which she scores 6). Use the model to estimate the probability that she will use a condom in her next encounter.

The correct analysis was to run a hierarchical logistic regression entering **perceive, safety** and **gender** in the first block and **previous**, **selfcon** and **sexexp** in a second. I used forced entry on both blocks, but you could choose to run a Forward stepwise method on block 2 (either strategy is justified). For the variable **previous** I used an indicator contrast with 'No condom' as the base category.

Block 0

The output of the logistic regression will be arranged in terms of the blocks that were specified. In other words, SPSS will produce a regression model for the variables specified in block 1, and then produce a second model that contains the variables from both blocks 1 and 2. The results from block 1 are shown below. In this analysis we forced SPSS to enter **perceive, safety** and **gender** into the regression model first. First, the output tells us that 100 cases have been accepted, that the dependent variable has been coded 0 and 1 (because this variable was coded as 0 and 1 in the data editor, these codings correspond exactly to the data in SPSS).

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	100	100.0
	Missing Cases	0	.0
	Total	100	100.0
Unselected Cases		0	.0
Total		100	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Unprotected	0
Condom Used	1

Categorical Variables Codings

			Parameter coding	
		Frequency	(1)	(2)
Previous Use with Partner	No Condom	50	.000	.000
	Condom used	47	1.000	.000
	First Time with partner	3	.000	1.000

Classification Table[a,b]

			Predicted		
			Condom Use		
	Observed		Unprotected	Condom Used	Percentage Correct
Step 0	Condom Use	Unprotected	57	0	100.0
		Condom Used	43	0	.0
	Overall Percentage				57.0

a. Constant is included in the model.

b. The cut value is .500

Block 1

The next part of the output tells us about block 1: as such it provides information about the model after the variables **perceive, safety** and **gender** have been added. The first thing to note is that the −2LL has dropped to 105.77, which is a change of 30.89 (which is the value given by the *model chi-square*). This value tells us about the model as a whole whereas the *block* tells us how the model has improved since the last block. The change in the amount of information explained by the model is significant ($\chi 2$ (3) = 30.92, p < 0.0001) and so using perceived risk, relationship safety and gender as predictors significantly improves our ability to predict condom use. Finally, the classification table shows us that 74% of cases can be correctly classified using these three predictors.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	30.892	3	.000
	Block	30.892	3	.000
	Model	30.892	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	105.770	.266	.357

Classification Table[a]

			Predicted		
			Condom Use		
	Observed		Unprotected	Condom Used	Percentage Correct
Step 1	Condom Use	Unprotected	45	12	78.9
		Condom Used	14	29	67.4
	Overall Percentage				74.0

a. The cut value is .500

Hosmer and Lemeshow's goodness-of-fit test statistic tests the hypothesis that the observed data are significantly different from the predicted values from the model. So, in effect, we want a non-significant value for this test (because this would indicate that the model does not differ significantly from the observed data). In this case ($\chi 2$ (8) = 9.70, p = 0.287) it is non-significant which is indicative of a model that is predicting the real-world data fairly well.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9.700	8	.287

The part of the Output labelled *Variables in the Equation* then tells us the parameters of the model for the first block. The significance values of the Wald statistics for each predictor indicate that both perceived risk (Wald = 17.76, $p < 0.0001$) and relationship safety (Wald = 4.54, $p < 0.05$) significantly predict condom use. Gender, however, does not (Wald = 0.41, $p > 0.05$).

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	PERCEIVE	.940	.223	17.780	1	.000	2.560	1.654	3.964
	SAFETY	-.464	.218	4.540	1	.033	.629	.410	.963
	GENDER	.317	.496	.407	1	.523	1.373	.519	3.631
	Constant	-2.476	.752	10.851	1	.001	.084		

a. Variable(s) entered on step 1: PERCEIVE, SAFETY, GENDER.

The values of exp β for perceived risk (exp β = 2.56, $CI_{0.95}$ = 1.65, 3.96) indicate that if the value of perceived risk goes up by one, then the odds of using a condom also increase (because exp β is greater than 1). The confidence interval for this value ranges from 1.65 to 3.96 so we can be very confident that the value of exp β in the population lies somewhere between these two values. What's more, because both values are greater than 1 we can also be confident that the relationship between perceived risk and condom use found in this sample is true of the whole population. In short, as perceived risk increase by 1, people are just over twice as likely to use a condom.

The values of exp β for relationship safety (exp β = 0.63, $CI_{0.95}$ = 0.41, 0.96) indicate that if the relationship safety increases by one point, then the odds of using a condom decrease (because exp β is less than 1). The confidence interval for this value ranges from 0.41 to 0.96 so we can be very confident that the value of exp β in the population lies somewhere between these two values. In addition, because both values are less than 1 we can be confident that the relationship between relationship safety and condom use found in this sample would be found in 95% of samples from the same population. In short, as relationship safety increases by one unit, subjects are about 1.6 times less likely to use a condom.

The values of exp β for gender (exp β = 1.37, $CI_{0.95}$ = 0.52, 3.63) indicate that as gender changes from 0 (male) to 1 (female), then the odds of using a condom increase (because exp β is greater than 1). However, the confidence interval for this value crosses 1 which limits the generalizability of our findings because the value exp β in other samples (and hence the population) could indicate either a positive (*exp(B)* > 1) or negative (*exp(B)* < 1) relationship. Therefore, gender is not a reliable predictor of condom use.

A glance at the classification plot brings not such good news because a lot of cases are clustered around the middle. This indicates that the model could be performing more accurately (i.e. the classifications made by the model are not completely reliable).

```
              Step number: 1

              Observed Groups and Predicted Probabilities

    16 +
       |
       |
F      |
R   12 +                                 C
E      |                            C    C
Q      |                            C    C
U      |                            C    C
E    8 +                            C    C
N      |                          C C    C
C      |      U                   C C    C C
Y      |      UC     U            U U    C C   C C
     4 +      UU     U U          U U    C U   C C
       |    U UU  C  U U   C C    U U    U U   C C
       |UUUUU UU  C  U U   U U    U U    U U   C C    C      C
       |UUUUU UU  U  U U   U U   CU U   CU U   U U   UC  CCC UCCUC
Predicted ---------+--------------+--------------+---------------
  Prob:   0       .25            .5            .75              1
  Group:  UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

          Predicted Probability is of Membership for Condom Used
          The Cut Value is .50
          Symbols: U - Unprotected
                   C - Condom Used
          Each Symbol Represents 1 Case.
```

Block 2

The output below shows what happens to the model when our new predictors are added (previous use, self-control and sexual experience). This part of the output describes block 2, which is just the model described in block 1 but with a new predictors added. So, we begin with the model that we had in block 1 and we then add **previous**, **selfcon** and **sexexp** to it. The effect of adding these predictors to the model is to reduce the −2 log-likelihood to 87.971 (a reduction of 48.69 from the original model as shown in the *model chi-square* and an additional reduction of 17.799 from the reduction caused by block 1 as shown by the *block* statistics). This additional improvement of block 2 is significant ($\chi 2$ (4) = 17.80, $p < 0.01$) which tells us that including these three new predictors in the model has significantly improved our ability to predict condom use. The classification table tells us that the model is now correctly classifying 78% of cases. Remember that in block 1 there were 74% correctly classified and so an extra 4% of cases are now classified (not a great deal more—in fact, examining the table shows us that only 4 extra cases have now been correctly classified).

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	17.799	4	.001
	Block	17.799	4	.001
	Model	48.692	7	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	87.971	.385	.517

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9.186	8	.327

Classification Table[a]

			Predicted		
			Condom Use		Percentage Correct
Observed			Unprotected	Condom Used	
Step 1	Condom Use	Unprotected	47	10	82.5
		Condom Used	12	31	72.1
	Overall Percentage				78.0

a. The cut value is .500

The section labelled *Variables in the Equation* now contains all predictors. This part of the output represents the details of the final model. The significance values of the Wald statistics for each predictor indicate that both perceived risk (Wald = 16.04, $p < 0.001$) and relationship safety (Wald = 4.17, $p < 0.05$) still significantly predict condom use and, as in block 1, Gender does not (Wald = 0.00, $p > 0.05$). We can now look at the new predictors to see which of these has some predictive power.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	95.0% C.I.for EXP(B) Upper
Step 1[a]	PERCEIVE	.949	.237	16.038	1	.000	2.583	1.623	4.109
	SAFETY	-.482	.236	4.176	1	.041	.617	.389	.980
	GENDER	.003	.573	.000	1	.996	1.003	.326	3.081
	SEXEXP	.180	.112	2.614	1	.106	1.198	.962	1.490
	PREVIOUS			4.032	2	.133			
	PREVIOUS(1)	1.087	.552	3.879	1	.049	2.965	1.005	8.747
	PREVIOUS(2)	-.017	1.400	.000	1	.990	.983	.063	15.287
	SELFCON	.348	.127	7.510	1	.006	1.416	1.104	1.815
	Constant	-4.959	1.146	18.713	1	.000	.007		

a. Variable(s) entered on step 1: SEXEXP, PREVIOUS, SELFCON.

Previous use has been split into two components (according to whatever contrasts were specified for this variable). Looking at the very beginning of the output we are told the parameter codings for **Previous(1)** and **previous(2)**. You can tell by remembering the rule from contrast coding in ANOVA which groups are being compared: that is, we compare groups with zero codes against those with codes of 1. From the output we can see that **Previous(1)** compares the condom used group against the other two, and **Previous(2)** compares the base category of first time with partner against the other two categories. Therefore we can tell that previous use is not a significant predictor of condom use when it is the first time with a partner compared to when it is not the first time (Wald = 0.00, $p < 0.05$). However, when we compare the condom used category to the other categories we find that using a condom on the previous occasion does predict use on the current occasion (Wald = 3.88, $p < 0.05$).

Of the other new predictors we find that self control predicts condom use (Wald = 7.51, $p < 0.01$) but sexual experience does not (Wald = 2.61, $p > 0.05$).

The values of exp β for perceived risk (exp β = 2.58, $CI_{0.95}$ = 1.62, 4.106) indicate that if the value of perceived risk goes up by one, then the odds of using a condom also increase. What's more, because the confidence interval doesn't cross 1 we can also be confident that the relationship between perceived risk and condom use found in this sample is true of the whole population. As perceived risk increase by 1, people are just over twice as likely to use a condom.

The values of exp β for relationship safety (exp β = 0.62, $CI_{0.95}$ = 0.39, 0.98) indicate that if the relationship safety decreases by one point, then the odds of using a condom increase. The confidence interval does not cross 1 so we can be confident that the relationship between relationship safety and condom use found in this sample would be found in 95% of samples from the same population. As relationship safety increases by one unit, subjects are about 1.6 times less likely to use a condom.

The values of exp β for gender (exp β = 1.00, $CI_{0.95}$ = 0.33, 3.08) indicate that as gender changes from 0 (male) to 1 (female), then the odds of using a condom do not change (because exp β is equal to 1). The confidence interval crosses 1, therefore, gender is not a reliable predictor of condom use.

The values of exp β for previous use (1) (exp β = 2.97, $CI_{0.95}$ = 1.01, 8.75) indicate that if the value of previous usage goes up by one (i.e. changes from not having used one or being the first time to having used one), then the odds of using a condom also increase. What's more, because the confidence interval doesn't cross 1 we can also be confident that this relationship is true in the whole population. If someone used a condom on their previous encounter with this partner (compared to if they didn't use one, or if it is their first time) then they are three times more likely to use a condom. For previous use (2) the value of exp β (exp β = 0.98, $CI_{0.95}$ = 0.06, 15.29) indicates that if the value of previous usage goes up by one (i.e. changes from not having used one or having used one to it being their first time with this partner), then the odds of using a condom do not change (because the value is very nearly equal to 1). What's more, because the confidence interval crosses 1 we can tell that this is not a reliable predictor of condom use.

The value of exp β for self-control (exp β = 1.42, $CI_{0.95}$ = 1.10, 1.82) indicates that if self-control increases by one point, then the odds of using a condom increase also. The confidence interval does not cross 1 so we can be confident that the relationship between relationship safety and condom use found in this sample would be found in 95% of samples from the same population. As self-control increases by one unit, subjects are about 1.4 times more likely to use a condom.

The values of exp β for sexual experience (exp β = 1.20, $CI_{0.95}$ = 0.95, 1.49) indicate that as sexual experience increases by one unit, then the odds of using a condom increase slightly. However, the confidence interval crosses 1, therefore, sexual experience is not a reliable predictor of condom use.

A glance at the classification plot brings good news because a lot of cases that were clustered in the middle are now spread towards the edges. Therefore, overall this new model is more accurately classifying cases compared to block 1.

```
              Step number: 1

              Observed Groups and Predicted Probabilities

        16 +                                                            +
           |                                                            |
           |                                                            |
F          |                                                            |
R       12 +                                                            +
E          |                                                            |
Q          |                                                            |
U          | U                                                          |
E        8 + U                                                          +
N          | U                                                          |
C          | U                                                          |
Y          | U          C                                               |
         4 + U    U     CC                                              +
           | U U UU   U UU    U    C            C     C       C  C  C   |
           |UUUUUUU  CU UU    U    U  CC  C     U  C  C  C   CC  CCCC   |
           |UUUUUUU CCUUUU  U UUUUCUU CUUCUU C CUUUUCCCU C  CCCCCUUCCCCC|
Predicted ---------+--------------+--------------+--------------+----------
  Prob:   0        .25            .5             .75                  1
  Group:  UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

          Predicted Probability is of Membership for Condom Used
          The Cut Value is .50
          Symbols: U - Unprotected
                   C - Condom Used
          Each Symbol Represents 1 Case.
```

Testing for Multicollinearity

Multicollinearity can affect the parameters of a regression model. Logistic regression is equally as prone to the biasing effect of collinearity and it is essential to test for collinearity following a logistic regression analysis (see the main book for details of how to do this). The results of the analysis are shown below. From the first table we can see that the tolerance values for all variables are all close to 1 and are much larger than the cut-off point of 0.1 below which Menard (1995) suggests indicates a serious collinearity problem. Myers (1990) also suggests that a VIF value greater than 10 is cause for concern and in these data the values are all less than this criterion.

The output below also shows a table labelled *Collinearity Diagnostics*. In this table, we are given the eigenvalues of the scaled, uncentred cross-products matrix, the condition index and the variance proportions for each predictor. If any of the eigenvalues in this table are much larger than others then the uncentred cross-products matrix is said to be ill-conditioned, which means that the solutions of the regression parameters can be greatly affected by small changes in the predictors or outcome. In plain English, these values give us some idea as to how accurate our regression model is: if the eigenvalues are fairly similar then the derived model is likely to be unchanged by small changes in the measured variables. The *condition indexes* are another way of expressing these eigenvalues and represent the square root of the ratio of the largest eigenvalue to the eigenvalue of interest (so, for the dimension with the largest eigenvalue, the condition index will always be 1). For these data the condition indexes are all relatively similar showing that a problem is unlikely to exist.

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Perceived Risk	.849	1.178
	Relationship Safety	.802	1.247
	GENDER	.910	1.098
2	Perceived Risk	.740	1.350
	Relationship Safety	.796	1.256
	GENDER	.885	1.130
	Previous Use with Partner	.964	1.037
	Self-Control	.872	1.147
	Sexual experience	.929	1.076

a. Dependent Variable: Condom Use

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions						
				(Constant)	Perceived Risk	Relationship Safety	GENDER	Previous Use with Partner	Self-Control	Sexual experience
1	1	3.137	1.000	.01	.02	.02	.03			
	2	.593	2.300	.00	.02	.10	.55			
	3	.173	4.260	.01	.55	.76	.08			
	4	9.728E-02	5.679	.98	.40	.13	.35			
2	1	5.170	1.000	.00	.01	.01	.01	.01	.01	.01
	2	.632	2.860	.00	.02	.06	.43	.10	.00	.02
	3	.460	3.352	.00	.03	.10	.01	.80	.00	.00
	4	.303	4.129	.00	.07	.01	.24	.00	.00	.60
	5	.235	4.686	.00	.04	.34	.17	.05	.50	.00
	6	.135	6.198	.01	.61	.40	.00	.00	.47	.06
	7	6.510E-02	8.911	.98	.23	.08	.14	.03	.03	.31

a. Dependent Variable: Condom Use

The final step in analysing this table is to look at the variance proportions. The variance of each regression coefficient can be broken down across the eigenvalues and the variance proportions tell us the proportion of the variance of each predictor's regression coefficient that is attributed to each eigenvalue. These proportions can be converted to percentages by multiplying them by 100 (to make them more easily understood). In terms of collinearity, we are looking for predictors that have high proportions on the same *small* eigenvalue, because this would indicate that the variances of their regression coefficients are dependent (see Field, 2004). Again, no variables appear to have similarly high variance proportions for the same dimensions. The result of this analysis is pretty clear cut: there is no problem of collinearity in these data.

Residuals

Residuals should be checked for influential cases and outliers. As a brief guide, the output lists cases with standardized residuals greater than 2. In a sample of 100, we would expect around 5-10% of cases to have standardized residuals with absolute values greater than this. For these data we have only 4 cases and only 1 of these has an absolute value greater than 3. Therefore, we can be fairly sure that there are no outliers.

Casewise List[b]

Case	Selected Status[a]	Observed Condom Use	Predicted	Predicted Group	Temporary Variable Resid	Temporary Variable ZResid
41	S	U**	.891	C	-.891	-2.855
53	S	U**	.916	C	-.916	-3.294
58	S	C**	.142	U	.858	2.455
83	S	C**	.150	U	.850	2.380

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2.000 are listed.

Question 3

The values predicted for these cases will depend on exactly how you ran the analysis (and the paremeter coding used on the variable 'previous'). Therefore, your answers might differ slightly from mine.

Case Summaries[a]

	Case Number	Predicted Value	Predicted Group
12	12	**.49437**	Unprotected
53	53	**.88529**	Condom Used
75	75	**.37137**	Unprotected

a. Limited to first 100 cases.

Question 4

Step 1: Logistic Regression Equation:

$$P(Y) = \frac{1}{1+e^{-Z}}$$

$$Where \quad Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

Step 2: Use the values of β from the SPSS output (final model), and the values of *X* for each variable (from question) to construct the following table:

Variable	β_i	X_i	$\beta_i X_i$
Gender	0.0027	1	0.0027
Safety	-0.4823	2	-0.9646
Sexexp	0.1804	2	0.3608
Previous (1)	1.0870	1	1.0870
Previous (2)	-.0167	0	0
Selfcon	0.3476	2	0.6952
Perceive	0.9489	6	5.6934

Step 3: Place the values of $\beta_i X_i$ into the equation for *z* (remembering to include the constant).

$$
\begin{aligned}
z &= -4.6009 + 0.0027 - 0.9646 + 0.3608 + 1.0870 + 0 + 0.6952 + 5.6934 \\
&= 2.2736
\end{aligned}
$$

Step 4: Replace this value of *z* into the logistic regression equation:

$$
\begin{aligned}
P(Y) &= \frac{1}{1+e^{-Z}} \\
&= \frac{1}{1+e^{-2.2736}} \\
&= \frac{1}{1+0.10} \\
&= 0.9090
\end{aligned}
$$

Therefore, there is a 91% chance that she will use a condom on her next encounter.