Model One
Fit a logistic regression model with at least two explanatory variables, both of which you think may be related to your response variable, to data from the General Social Science data. Be sure to give a rational for the final model chosen, to interpret model parameters, and to assess model fit. If any of your response variables are continuous be sure to look at regression diagnostics when assessing model fit.

METHODS
Hypothesis: What is the probability of an individual having sexual relations outside of their marriage given their level of education and their gender?

Data:
The data was obtained from the social science survey. The variables from the social science survey used to test the hypothesis were: 34 EDUC Highest level of school completed
42 SEX Respondents sex 1=male, 2=female
487 EVSTRAY Have sex other than spouse while married 1=yes, 2=no, 3=never married

In order to test they hypothesis, the education variable was grouped into three categories high school graduate and below (HIGH), beyond high school to bachelors completion (COLLEGE), and beyond bachelor's (BEYOND). Sample cases that did not respond for any of the variables and cases that replied 3 (never married) for EVSTRAY were eliminated from the data set. This manipulation of the data yielded the following data set:

		Evstray	
Education Level	Gender	Yes	No
High School and below	Male	43	147
	Female	31	227
College Level	Male	41	109
	Female	35	161
Beyond Bachelors	Male	8	34
	Female	13	45

Model Fit: As the hypothesis to be tested contains a dichotomous response/dependent variable and two explanatory/ independent variables the logistic regression model was chosen to fit.
Symbolically the model fit is $\text{logit}(\pi(x_1)) = \log[\pi(x_1)/ 1 - \pi(x_1)] = \alpha + \beta_1(x_{2i}) + \ldots + \beta_j(x_{ji})$

RESULTS
Model Parameters:

```
Analysis Of Parameter Estimates

                                      Standard   Wald 95% Confidence      Chi-
 Parameter              DF   Estimate     Error         Limits           Square
 Intercept               1    -1.5789    0.2614    -2.0913    -1.0665     36.47
 educ        high        1    -0.3011    0.2782    -0.8464     0.2443      1.17
 educ        college     1     0.0503    0.2795    -0.4976     0.5982      0.03
 educ        beyond      0     0.0000    0.0000     0.0000     0.0000       .
 gender      male        1     0.5532    0.1715     0.2170     0.8894     10.40
 gender      female      0     0.0000    0.0000     0.0000     0.0000       .
 Scale                   0     1.0000    0.0000     1.0000     1.0000
```

Model: Logit= $-1.5789 + -0.3011x_{highed} + 0.0503x_{colled} + 0.5532y_{male}$

Fit Statistics/ Diagnostic Information:
Before fitting a logit model, I took a look at the measures of association to get a picture of what was going on with the data. Because all cell counts are greater than five, adjustments were not necessary. In brief the following was found:

Controlling for	Testing	θ =	χ^2	conclusion
Educ= HIGH	Evrstay by gender	2.1420 CL 1.29-3.55	8.9429, df=1, p=0.0028	The variables EVRSTRAY and GENDER are dependent. The odds of a male having sexual relations outside of his marriage is 2.14 times the odds of a female doing so
Educ= COLLEGE	Evrstray by gender	1.7303 CL 1.04-2.89	4.4516, df=1, p=0.0349	The variables EVRSTRAY and GENDER are dependent. The odds of a male having sexual relations outside of his marriage is 1.73 times the odds of a female doing so
Educ= BEYOND	Evrstray by gender	0.8145 CL 0.30- 2.19	0.1664, df=1, p=0.6833	The variables EVRSTRAY and GENDER are independent. The odds of a female having sexual relations outside of her marriage is 1.23 times the odds of a male doing so.

To determine whether the odds ratio's are homogenous I looked at the Breslow-Day statistic, 2.9771 with df=2 and pvalue=0.2257. The null hypothesis is not rejected, thus it can be concluded the conditional odds ratio are comparable across the partial table and homogeneous association is present; additionally, this informs us that our model fitted will not require a three way interaction, or stated another way, the relationship between everstraying and gender is comparable over the three education levels. Because homogeneous association is present the common odds ratio can be determined, the Mantel-Haenzel estimate of 1.7357 indicates the odds of males straying are 1.74 times greater than the odds of a female straying controlling for level of education.
The Cochran-Mantel-Haenzel Statistic (10.5069 with a df=1 and p=0.0012) provides strong evidence against conditional independence, which means that the conditional odds ratios are not all equal to 1, demonstrating evrstray and gender are dependent conditioning for level of education.

To fit the model I compared the saturated model (the model containing all main effects with interaction effects) to the model eliminating the gender*educ interaction. This provided a difference in deviance G^2 of 2.9187, df=2, pvalue=0.2324. To assess model fit, I looked at G^2=2.9772 and X^2=2.9772 with df=2, these values are small which would conclude model fit. Additionally, although our data is not continuous, the Homer and Lemeshow Goodness- of-fit Test pvalue=0.5616 which would support model fit. Furthermore, the largest absolute value of the standardized residuals is 1.04. Thus it can be concluded the interaction term is not needed in the model.

The Type 3 Analysis shows that the gender variable is needed in the model, while the education variable may not be. However, because the test hypothesis was to examine the relationship between education level and gender on evrstray the variable will be kept in the model.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
educ	2	3.96	0.1381
gender	1	10.44	0.0012

Interpretation:
Regardless of gender, the odds of an individual with an education level beyond bachelor's degree having sex with someone other than their spouse while married is 1.35 times the odds of an individual with an education level of high school graduate or below having sex with someone other than their spouse while married.
Regardless of gender, the odds of an individual with an educational level beyond high school up to college graduate having sex with someone other than their spouse while married is 1.05 times the odds of an individual with an education level beyond college graduate having sex with someone other than their spouse while married.
Regardless of gender, the odds of an individual with an educational level beyond high school up to college graduate having sex with someone other than their spouse while married is 1.29 times the odds of an individual with an educational level of high school graduate or below having sex with someone other than their spouse while married.
Additionally, regardless of education level, the odds of a male having sex with someone other than their spouse while married is 1.74 times the odds of a female having sex with someone other than their spouse while married.

Model Two:
Analyze the General Social Science data using loglinear models and different variables than were used in model one. Be sure to include at least three variables of interest in your model. Once again be sure to give a rationale for the final model chosen, interpret model parameters, and to assess model fit. Illustrate the relationship between model parameters and fitted odds ratios. Be sure that the final model considered in NOT the saturated model. If this is the case for the variables you have chosen then include a fourth variable in your model.

METHODS
Hypothesis: Is there an association between how one views cheating on taxes, extramarital sex, and watching x-rated movies?

Data: The following data from the General Social Science survey was used:

420 XMARSEX1
Is extramarital sex wrong?
1= Always wrong
2= Almost Always wrong
3= Wrong sometimes
4= Not wrong at all
8=Can't choose

224 XMOVIE
Seen an x-rated movie in the past year
0=NAP
1=yes
2=no
8=DK
9=NA

424 TAXCHEAT
Wrong to cheat on taxes?
0=NAP
1=Strongly Agree
2=Agree
3=Not agree/disagree
4=Disagree
5=Strong Disagree
8=Can't choose

The data was condensed eliminating cases that responded 8 (Can't choose or DK), 0 (NAP), or 9 (NA). Additionally the XMARSEX1 variable combined values 1 & 2 re-coding, 1= almost or always wrong, and combined values 3&4 recoding, 2=sometimes or not wrong. The TAXCHEAT variable was similarly modified combining 4&5 and recoded 0= disagree or strongly disagree, and combining 1&2 recoding 1=agree of strongly agree, with the 3 value cases being eliminated.

The data set that was obtained after manipulating the larger data set was:

		Taxcheat	
View on Xmarsex	xmovie	Not wrong	wrong
Almost or always wrong	Yes	41	47
	No	195	63
Sometimes or not wrong	Yes	4	6
	No	10	10

Model Fit: Because the hypothesis is testing the relationship between three categorical variables the loglinear model was chosen. This model does not differentiate between response variables and explanatory variables. Symbollically represented
$\text{Log}(\mu_{ij}) = \lambda + \lambda_i^x + \lambda_j^y + \lambda_k^z + \lambda_{ij}^{xy} + \lambda_{ik}^{xz} + \lambda_{jk}^{yz} + \lambda_{ijk}^{xyz}$

RESULTS

Model Parameters:

```
                        Analysis Of Parameter Estimates

                                                     Standard        Wald 95%
 Parameter                          DF   Estimate       Error    Confidence Limits

 Intercept                           1     2.4406      0.2689     1.9136    2.9676
 taxcheat         notax              1    -0.2138      0.3749    -0.9487    0.5210
 taxcheat         tax                0     0.0000      0.0000     0.0000    0.0000
 xmovie           movie              1    -1.5163      0.1646    -1.8390   -1.1937
 xmovie           nomovie            0     0.0000      0.0000     0.0000    0.0000
 xmar             nosex              1     2.8248      0.2751     2.2856    3.3639
 xmar             sexok              0     0.0000      0.0000     0.0000    0.0000
 taxcheat*xmovie  notax    movie     1     1.1962      0.2443     0.7174    1.6749
 taxcheat*xmovie  notax    nomovie   0     0.0000      0.0000     0.0000    0.0000
 taxcheat*xmovie  tax      movie     0     0.0000      0.0000     0.0000    0.0000
 taxcheat*xmovie  tax      nomovie   0     0.0000      0.0000     0.0000    0.0000
 taxcheat*xmar    notax    nosex     1    -0.8969      0.3837    -1.6490   -0.1448
 taxcheat*xmar    notax    sexok     0     0.0000      0.0000     0.0000    0.0000
 taxcheat*xmar    tax      nosex     0     0.0000      0.0000     0.0000    0.0000
 taxcheat*xmar    tax      sexok     0     0.0000      0.0000     0.0000    0.0000
 Scale                               0     1.0000      0.0000     1.0000    1.0000
```

Model:

$\text{Log}(\mu_{ij}) = 2.4406 + -0.2138_{notax}^{taxcheat} + -1.5163_{movie}^{xmovie} + 2.8248_{nosex}^{xmar} + 1.1962_{notax*movie}^{taxcheat*movie} + -0.8969_{notax*nosex}^{taxcheat*xmar}$

Fit Statistics/ Diagnostic Information:

Before fitting a model, I took a look at the measures of association to get a picture of what was going on with the data. In brief the following was found:

Controlling for	Variable	θ	χ^2	conclusion
Xmarsex=1	Taxcheat X xmovie	0.2818	25.4323, df=1, p=0.0001	The variables taxcheat and movie are dependent controlling for xmarsex=1. For those who believe cheating on a spouse as almost always and always wrong, the odds of an individual who view cheating on taxes as not wrong, not having viewed an xrated movie in the past year is 3.55 (1/0.2818) times the odds of an individual who has viewed an x-rated movie in the past year.
Xmarsex=2	Taxcheat X xmovie	0.6667	0.2679, df=1, p=0.6048	The variables taxcheat and movie are independent controlling for xmarsex=1. For those who believe cheating on a spouse as sometimes wrong or not wrong as all, the odds of an individual who view cheating on taxes as not wrong, not having viewed an xrated movie in the past year is 1.5 (1/0.6667) times the odds of an individual who has viewed an x-rated movie in the past year.

To determine whether the odds ratio's are homogenous the Breslow-Day statistic, 1.1042 with df=1 and pvalue=0.2933 we do not reject the null hypothesis, thus we can conclude the conditional odds ratio are comparable across the partial table and homogeneous association is present; additionally this informs us that our model fitted will not require a three way interaction term.

Because homogeneous association is present the common odds ratio can be determined, the Mantel-Haenzel estimate of 0.3088 indicates the odds of an individual who view cheating on taxes as not wrong, not having viewed an xrated movie in the past year is 3.24 (1/0.3088) times the odds of an individual who has viewed an x-rated movie in the past year controlling for the views on extramarital sex.

Homogenous association has been determined and to further examine and determine whether conditional independence is present, I looked at the Cochran-Mantel-Haenzel Statistic which was 24.2504 with a df=1 and p<0.0001. Therefore the null hypothesis is rejected indicating 'taxes and xmovie' are dependent, conditioning for xmarsex.

To try to fit a model to the data, the following systematic analysis was performed:

Model	Deviance G^2 (df)	Models compared	Δ G^2 (df)	Conclusion
Saturated (1) gender gpa race	0 (0)			
(2) three way removed	1.0538 (1)	(2)-(1)	(1) P=0.3046	Three way interaction is not needed
(3a) taxcheat*movie removed	24.5638 (2)	(3a)-(2)	23.51 (1) P<.0001	Take out xmovie*xmar Taxcheat*movie and taxcheat*xmar left
(3b) xmovie*xmar removed	1.1575 (2)	(3b)-(2)	0.1037 (1) P= 0.7474	
(3c) taxcheat*xmar removed	5.7263 (2)	(3c)-(2)	4.6725 (1) P=0.0306	
(4a) taxcheat*xmar	25.4158 (3)	(4a)-(3b)	24.2583 (1) p<.001	Both interactions needed. Stop.
(4b) taxcheat*movie	6.5783 (3)	(4b)-(3b)	5.4208 (1) p=0.0199	

The G^2 statistic of 1.1575 and X^2 statistic of 1.2795, both with df=2 demonstrates a good model fit.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	1.1575	0.5787
Scaled Deviance	2	1.1575	0.5787
Pearson Chi-Square	2	1.2795	0.6398
Scaled Pearson X2	2	1.2795	0.6398
Log Likelihood		1308.2345	

Furthermore, with the largest standardized residual being an absolute value of 1.06, and a dissimilarity index of D= 0.012, good model fit is confirmed.

Additionally, the Type 3 Analysis shows the variables and interactions are needed in the model.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
taxcheat	1	0.10	0.7480
xmovie	1	60.66	<.0001
xmar	1	272.97	<.0001
taxcheat*xmovie	1	24.26	<.0001
taxcheat*xmar	1	5.42	0.0199

Interpretation:

Regardless of how one felt about extramarital sex, the odds of those that did not feel cheating on taxes is wrong and had seen an x-rated movie in the past movie is 3.31 times the odds of those that feel cheating on taxes is wrong and have not seen an x-rated movie in the past year.

Regardless of whether one has seen an x-rated movie in the past year, the odds of those that feel cheating on taxes is wrong but having extramarital sex is not wrong is 02.45 times the odds of those that did not feel cheating on taxes is wrong but feel extramarital sex is wrong.

Model Three
Using the same variables as in model one, fit a loglinear model. Interpret the model parameters and assess model fit. Compare and contrast the results obtained from fitting model 3 and model one.

METHODS
Hypothesis: What is the relationship between gender, education level and engaging in sexual relations outside of the marriage?

Data: Refer to model one for data description and models of association

Model Fit: Because the hypothesis is testing the relationship between three categorical variables the loglinear model was chosen. This model does not differentiate between response variables and explanatory variables. Symbollically represented
$\text{Log}(\mu_{ij}) = \lambda + \lambda_i^{x} + \lambda_j^{y} + \lambda_k^{z} + \lambda_{ij}^{xy} + \lambda_{ik}^{xz} + \lambda_{jk}^{yz} + \lambda_{ijk}^{xyz}$

RESULTS
Model Parameters:
Analysis Of Parameter Estimates

Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits	
Intercept			1	3.8568	0.1166	3.6283	4.0852
educ	high		1	1.5548	0.1238	1.3121	1.7975
educ	college		1	1.2290	0.1279	0.9783	1.4797
educ	beyond		0	0.0000	0.0000	0.0000	0.0000
gender	male		1	-0.4009	0.0759	-0.5496	-0.2521
gender	female		0	0.0000	0.0000	0.0000	0.0000
evrstray	yes		1	-1.5845	0.2608	-2.0956	-1.0733
evrstray	no		0	0.0000	0.0000	0.0000	0.0000
educ*evrstray	high	yes	1	-0.2953	0.2765	-0.8372	0.2467
educ*evrstray	high	no	0	0.0000	0.0000	0.0000	0.0000
educ*evrstray	college	yes	1	0.0572	0.2777	-0.4871	0.6016
educ*evrstray	college	no	0	0.0000	0.0000	0.0000	0.0000
educ*evrstray	beyond	yes	0	0.0000	0.0000	0.0000	0.0000
educ*evrstray	beyond	no	0	0.0000	0.0000	0.0000	0.0000
gender*evrstray	male	yes	1	0.5532	0.1711	0.2178	0.8886
gender*evrstray	male	no	0	0.0000	0.0000	0.0000	0.0000
gender*evrstray	female	yes	0	0.0000	0.0000	0.0000	0.0000
gender*evrstray	female	no	0	0.0000	0.0000	0.0000	0.0000
Scale			0	1.0000	0.0000	1.0000	1.0000

Model:
$\mathbf{Log(\mu) = 3.8568 + 1.5548^{educ}{}_{high} + 1.2290^{educ}{}_{college} - 0.4009^{gender}{}_{male} - 0.2953^{educ*evrstray}{}_{high*yes} + 0.0572^{educ*evrstray}{}_{college*yes} + 0.5532^{gender*evrstray}{}_{male*yes}}$

Fit Statistics/ Diagnostic Information:

To fit a loglinear model to the data, the following systematic analysis was performed:

Model	Deviance G^2 (df)	Models compared	ΔG^2 (df)	Conclusion
(1) Saturated	( )			na
(2) three way interaction removed	2.9817 (2)	(2)-(1)	2.9817 (1) p= 0.2252	Three way interaction is not needed
(3a) Educ*evrstray Gender*evrstray	3.0301 (4)	(3a)-(2)	0.0484 (2) p= 0.9761	Educ*gender interaction not needed
(3b) Gender*evrstray Educ*gender	6.9407 (4)	(3b)-(2)	3.959 (2) p= 0.1381	
(3c) Gender*educ Educ*evrstray	13.4241 (3)	(3c)-(2)	10.4424 (1) p= 0.0012	
(4a) educ*evrstray	13.5193 (5)	(4a)-(3a)	10.4892 (1) p= 0.0012	educ*evrstray interaction not needed
(4b) gender*evrstray	7.0360 (6)	(4b)-(3a)	4.0059 (2) p= 0.1349	
(5) main effect variables only	17.5251 (7)	(5)-(4b)	10.4891 (1) p= 0.0012	gender*evrstray interaction needed

Although the fit of the model without the educ*evrstray interaction is a good one with a X^2= 13.5608, df=3, and G^2=13.4241,df=3, the dissimilarity index=0.046 which would indicate there could be a significicant non-fit.

```
                 Criteria For Assessing Goodness Of Fit

         Criterion                 DF           Value        Value/DF

         Deviance                   3         13.4241          4.4747
         Scaled Deviance            3         13.4241          4.4747
         Pearson Chi-Square         3         13.5608          4.5203
         Scaled Pearson X2          3         13.5608          4.5203
         Log Likelihood                     3279.8630
```

Additionally, although the adjusted residuals were good with the highest having an absolute value of 2.02, I felt the model that kept the educ*evrstray interaction conceptually made more sense. Furthermore, the dissimilarity index for this model was D=0.014 and the largest absolute value for adjusted residuals was 1.54.

Interpretation:
Regardless of education level, the odds of a male having sex with someone other than their spouse while married is 1.74 times the odds of a female having sex with someone other than their spouse while married.

Regardless of gender, the odds of an individual with an education level beyond bachelor's degree having sex with someone other than their spouse while married is 1.34 times the odds of an individual with an education level of high school graduate or below having sex with someone other than their spouse while married.
Regardless of gender, the odds of an individual with an educational level beyond high school up to college graduate having sex with someone other than their spouse while married is 1.05 times the odds of an individual with an education level beyond college graduate having sex with someone other than their spouse while married.
Regardless of gender, the odds of an individual with an educational level beyond high school up to college graduate having sex with someone other than their spouse while married is 1.27 times the odds of an individual with an educational level of high school graduate or below having sex with someone other than their spouse while married.

Compare and contrast models one and three:
With the logit model two variables (gender and education level) are being used to determine the likelihood of the third variable (whether an individual is likely to stray -have sexual relations outside of the marriage). With the loglinear model the association between the three variables is being examined. Thus, conceptually we would expect the loglinear model to include the evrstray*gender interaction as well as the everstray*educ interaction. However, if we fit the loglinear model strictly by looking at the change in deviance we are able to eliminate the everstray*educ interaction; which relates to the TypeIII Test results for the logit model that indicated that 'Education' may not be needed in the model. This is a demonstration of the importance of fitting the models with the hypothesis in mind, so that the model fits not only quantitatively but conceptually as well.

Model Four

Using different variables that were used in models one or two, fit a linear by linear association model using at least three variables. Explain why using this model is better than using a loglinear model. Once again be sure to assess model fit. Furthermore, conduct the Generalized Cochran-Mantel-Haenszel tests and compare and contrast the results.

METHODS

Hypothesis: What is the relationship between an individual's perception of general happiness and their perception of their health?

Data: The following variables were used from the general social science survey:

157 HAPPY general happiness	**159 HEALTH**	**49 MAWRKGRW Mothers employment when R growing up**
0 NAP	0 NAP	0 NAP
1 very happy	1 excellent	1 YES
2 pretty happy	2 good	2 NO
3 not too happy	3 fair	8 DK
8 DK	4 poor	9 NA
9 NA	8 DK	
	9 NA	

The data cases that were missing responses or had a 0 (NAP), 8 (DK), or 9 (NA) response were eliminated. This resulted in the following data set:

		Health			
	Happiness	Excellent	Good	Fair	Poor
Mother employed as a child	Very happy	118	100	16	9
	Pretty happy	127	249	76	13
	Not too happy	23	39	31	13
Mother not employed as a child	Very happy	80	77	13	4
	Pretty happy	66	140	53	14
	Not too happy	4	24	21	12

Model Fit: The data consists of two ordinal variables and a nominal variable. With the ordinal data it is believed that the corners of the table will have greater departures from expected values being 'stronger' relationships between the two variables than within the middle of the table. Thus the linear by linear association model was chosen to be fit, this is a better choice than the loglinear model because it preserves the strength of ordinal (ordered) data, while the loglinear model treats each variable as nominal.

Symbolically represented: $\log(\mu_{ij}) = \lambda + \lambda_i^x + \lambda_j^y + \lambda_k^z + \beta u_i v_j + \lambda_{ik}^{xz} + \lambda_{jk}^{yz}$

RESULTS

Model Parameters:

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept		1	-5.1060	0.6594	-6.3985	-3.8136	59.95
ma	ma	1	0.4715	0.0565	0.3607	0.5823	69.53
ma	female	0	0.0000	0.0000	0.0000	0.0000	.
happy	veryha	1	3.3711	0.2640	2.8538	3.8884	163.12
happy	prettyha	1	2.8341	0.1692	2.5024	3.1658	280.47
happy	notoo	0	0.0000	0.0000	0.0000	0.0000	.
health	exc	1	5.4175	0.3909	4.6513	6.1838	192.05
health	good	1	4.7829	0.3005	4.1939	5.3719	253.32
health	fair	1	2.5012	0.2053	2.0988	2.9037	148.40
health	poor	0	0.0000	0.0000	0.0000	0.0000	.
row*column		1	0.6110	0.0595	0.4944	0.7277	105.40
Scale		0	1.0000	0.0000	1.0000	1.0000	

Model: $\theta = \exp[\beta]$

Fit Statistics/ Diagnostic Information:
With a G^2 = 33.2187, df=16, pvalue= 0.0069 and X^2 = 36.5786, df=16, pvalue=0 .0024 this model is a good fit.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	16	33.2187	2.0762
Scaled Deviance	16	33.2187	2.0762
Pearson Chi-Square	16	36.5786	2.2862
Scaled Pearson X2	16	36.5786	2.2862
Log Likelihood		4570.5321	

The Wald test for the data is $(-5.1060/0.6594)^2$ = 59.96, df=1, pvalue<0.0001, indicating a good fit. Furthermore, the adjusted residuals have an absolute value of three and under, with the cell counts of five or greater this can be considered a good fit.

The results of the Generalized Cochran-Mantel-Haenszel Test value of 117.0950, df=1, pvalu<0.0001 confirms the dependence of health and happiness conditioning on mothers work status while growing up.

Summary Statistics for happy by health
Controlling for ma
Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	117.0950	<.0001
2	Row Mean Scores Differ	2	118.9363	<.0001

Interpretation:
The parameter estimate for the row*column interaction was 0.611, there was a tendency for people to view health as less good as their happiness declined. The odds ratio for a unit of change in row and column= 1.84 ($e^{.6110}$). If we were to look at the extreme corners of the table (excellent health vs. poor health and very happy vs. not too happy), the odds ratio would be 39.1, so the odds of being in very happy and viewing health as excellent is 39.1 times larger than being not too happy and viewing health as poor.

Model Five

Fit a base-line category logit model to variables of interest that have not yet been considered using at least one explanatory variable. Interpret the results both graphically and in terms of odds ratios. Extra credit will be given if more than one explanatory variable is considered. If you choose to include more than one explanatory variable be sure to consider model fit.

METHODS

Hypothesis: What is the probability that one will find life "exciting" as their SEI increases?

Data: The Data came from the Social Science Survey.
The '644 SEI Respondent socioeconomic index' and '160 LIFE Is life exciting or dull' variable were used. Cases were removed if they were missing a response for either variable and if LIFE was answered with values of 8 ("DK") and 9 ("NA"). The '160 LIFE Is life exciting or dull' variable consists of the following categories: 1=exciting, 2=routine, 3=dull

Model Fit: The data consists of a continuous variable, SEI, and a nominal response category, LIFE, with more than two possible response levels. It could be argued that the LIFE variable is ordinal, however, I felt that these were distinct categories and not necessarily ordered in relation to one another, thus I feel the base-line category logit model was appropriate.

RESULTS

Model Parameters:

Analysis of Maximum Likelihood Estimates

Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.0779	0.4151	0.04	0.8512
	2	1.1788	0.4064	8.41	0.0037
sei	3	0.0444	0.00942	22.26	<.0001
	4	0.0209	0.00939	4.95	0.0261

Model:

log[exciting/dull] = $\log[\pi_1(x_1)/\pi_2(x_2)] = \alpha + \beta_1 x = -0.0779 + 0.0444x$

log[routine/dull] = $1.1788 + 0.0209x$

log[exciting/routine] = $(\alpha_1-\alpha_2) + (\beta_1-\beta_2)x = (-0.0779-1.1788) + (0.0444-0.0209)x = 1.1009 + 0.0235x$

Fit Statistics/ Diagnostic Information:

The MLE intercept X^2=42.29, df=2, pvalue<0.0001 and SEI X^2=51.43, df=2, pvalue<0.001 indicates that both of the variables are needed in the model.

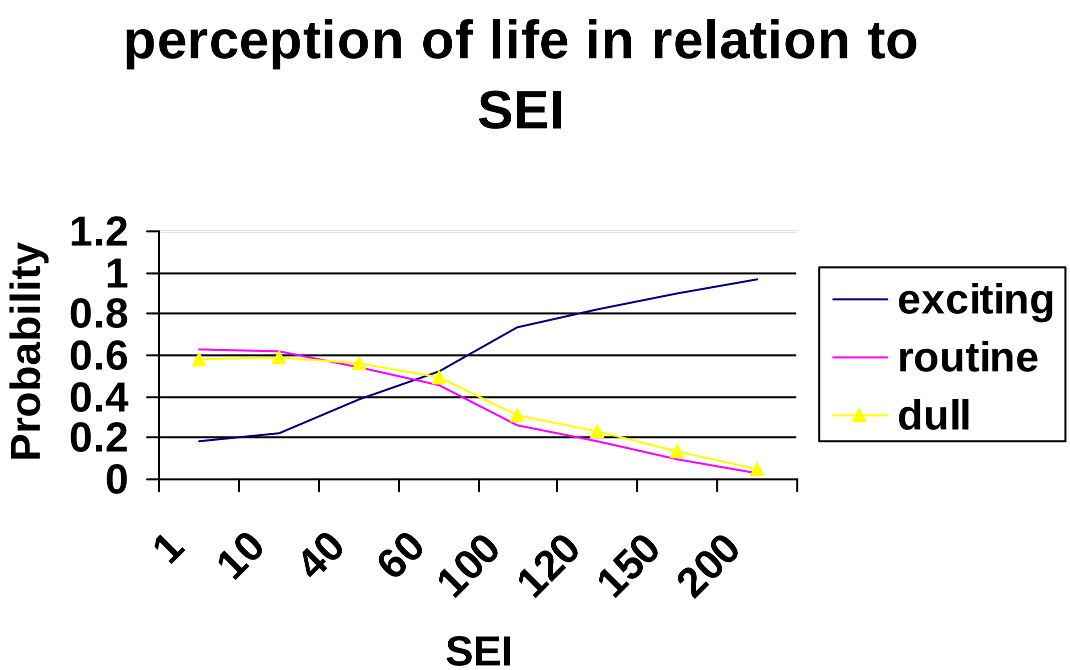
Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	2	42.29	<.0001
sei	2	51.43	<.0001
Likelihood Ratio	376	390.43	0.2933

Interpretation:

Given a 10 point increase in SES the following odds ratios are obtained:
Exciting to dull= exp(0.0444(10)) = 1.5589, Given an increase in socioeconomic status of 10 points, the odds of an individual viewing life as exciting is 1.56 times the odds of an individual viewing life as dull.
Routine to dull= exp(0.0209(10)) = 1.2324, Given an increase in socioeconomic status of 10 points, the odds of an individual viewing life as routine is 1.23 times the odds of an individual viewing life as dull.
Exciting to routine= exp(0.0235(10))= 1.2649, Given an increase in socioeconomic status of 10 points, the odds of an individual viewing life as exciting is 1.26 times the odds of an individual viewing life as routine.

We can also look at this graphically; the following graph illustrates the probability of viewing life as exciting, routine, and dull, given SEI.



It appears that at an SEI of approximately 75 the probability is 0.5 for any of the life categories. As the SEI increases it is much more probable that an individual is likely to find life “exciting” and less likely to find life “routine” or “dull” and conversely, when SEI is low, an individual is much more likely to consider life “routine” or “dull” rather than “exciting”

Model Six

Fit a proportional odds model to variables of interest that have not yet been considered using at least one explanatory variable. Interpret the results both graphically and in terms of odds ratios. Extra credit will be given if more than one explanatory variable is considered. If you choose to include more than one explanatory variable be sure to consider model fit.

METHODS

Hypothesis: What is the likelihood of an individual looking to God for strength and support as their age increases?

Data: The responses for '343 COPE3 Look to God for strength, support' and '32 AGE Age of respondent' were used from the Social Science Survey. Responses with 8 (don't know) or 9 (NA) were removed from the data leaving 1416 data cases. The coding for the COPE3 variable (1=a great deal, 2=quite a bit, 3=somewhat, 4=not at all) was used.

Model Fit: With more than two categories for the response variable of COPE3 and the variable being ordinal the proportional odds model best fits the data.
Symbolically represented: logit $[P(Y \le j)] = \alpha + \beta x$, j=1,...J-1

RESULTS

Model Parameters:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-0.6632	0.1416	21.9423	<.0001
Intercept2	1	0.1805	0.1405	1.6493	0.1991
Intercept3	1	1.5028	0.1495	101.0264	<.0001
age	1	0.00877	0.00287	9.3794	0.0022

Model:
$P(Y \le 1) = -0.6632 + 0.00877x$
$P(Y \le 2) = 0.1805 + 0.00877x$
$P(Y \le 3) = 1.5028 + 0.00877x$
$P(Y \le 4) = 1$

Fit Statistics/ Diagnostic Information:

The test for the proportional odds assumption gives a chi^2 statistic of 0.0187, with df=2, pvalue=0.9907 indicating the model is a good fit for the data.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
0.0187	2	0.9907

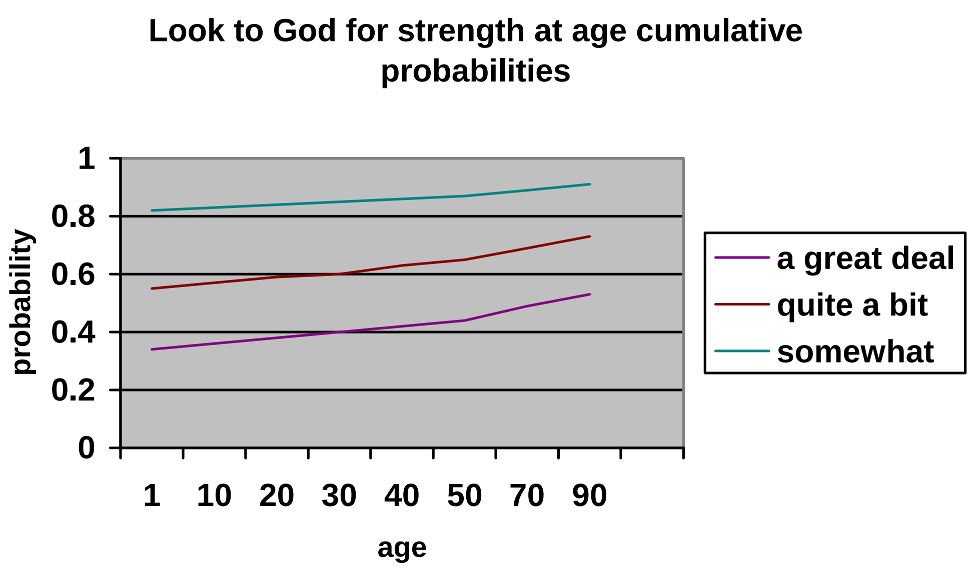
The model with $\beta = 0$ specifies independence between age and godcope3. The likelihood-ratio statistic for an ordinal test of independence (H_0:β=0) is =9.4444 df=1 gives strong evidence of an association (P=0.0021). Similar strong evidence results from the Wald test, using $x^2 = (\beta/ASE)^2 = (0.00877/0.00287)^2 = 9.3794$, pvalue=0.0022.

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	9.4444	1	0.0021
Score	9.3741	1	0.0022
Wald	9.3794	1	0.0022

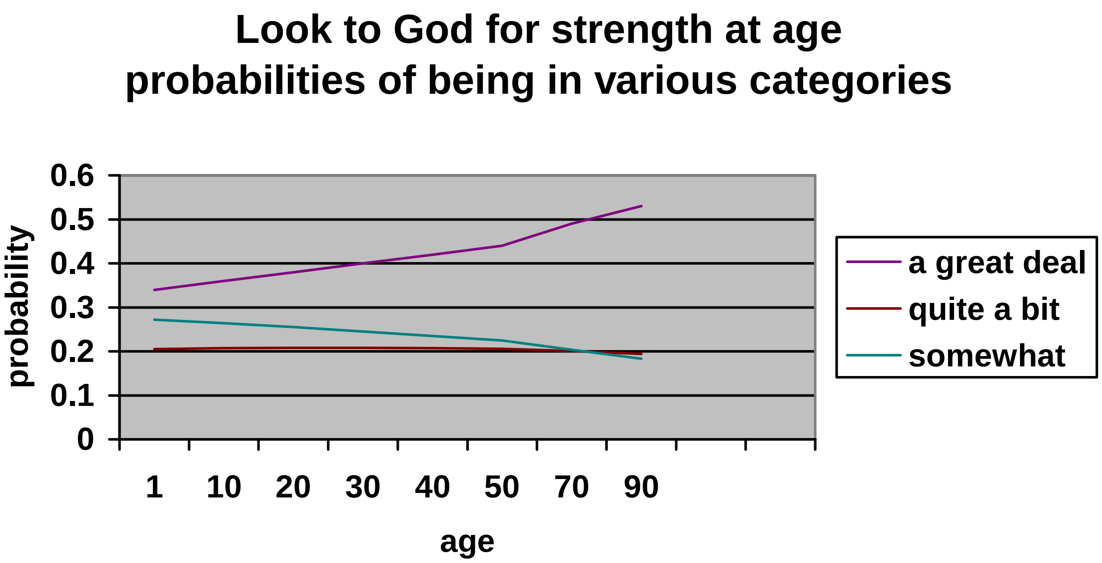
Interpretation:
In terms of odds ratios: $e^{0.00877(\text{age1-age2})}$ = The odds of an individual looking to God for strength is 1.20 times the odds of an individual looking to God for strength who is twenty years younger.

In terms of cumulative probabilities:
$P(Y \le 1) = \exp(-0.6632 + 0.00877x) / 1 + \exp(-0.6632 + 0.00877x)$
$P(Y \le 2) = \exp(0.1805 + 0.00877x) / 1 + \exp(0.1805 + 0.00877x)$
$P(Y \le 3) = \exp(1.5028 + 0.00877x) / 1 + \exp(1.5028 + 0.00877x)$



Cumulatively, the probability of looking to God for strength increases in each category as age increases.

Calculating probabilities:
$P(Y \le 1) = \exp(-0.6632 + 0.00877x) / 1 + \exp(-0.6632 + 0.00877x)$
$P(Y=2) = [\exp(0.1805 + 0.00877x) / 1 + \exp(0.1805 + 0.00877x)] -$
$[\exp(-0.6632 + 0.00877x) / 1 + \exp(-0.6632 + 0.00877x)]$
$P(Y=3) = P(Y \le 3) - P(Y \le 2)$



The probability of looking to God for strength is fairly low for all categories throughout the life span, with the exception of a great deal. The probability of looking to God for strength a great deal rises throughout the life span and after age 80 becomes more probable than not probable.