Exercise 1

(a) Frequency tables comparing city fuel efficiency over 30 mpg and the three possible categorical predictors follow.

Table of cylinders by cityover30mpg			
cylinders	**cityover30mpg**		
Frequency Expected	**0**	**1**	**Total**
eight	5 3.674	0 1.326	5
four	104 111.69	48 40.309	152
six	24 17.635	0 6.3646	24
Total	133	48	181

Table of fuel by cityover30mpg			
fuel	**cityover30mpg**		
Frequency Expected	**0**	**1**	**Total**
diesel	7 11.757	9 4.2431	16
gas	126 121.24	39 43.757	165
Total	133	48	181

Table of drive by cityover30mpg			
drive	**cityover30mpg**		
Frequency Expected	**0**	**1**	**Total**
fwd	67 83.768	47 30.232	114
rwd	66 49.232	1 17.768	67
Total	133	48	181

From the cylinders table, we can see that no six or eight cylinder vehicles have city fuel efficiency over 30 mpg, and roughly 1/3 of the four cylinder vehicles do. From the fuel table, we can see that more than half of the diesel-powered vehicles had city fuel efficiency over 30 mpg, and only about1/4 of gas-powered vehicles did. In terms of drive train, nearly all of the rear wheel drive vehicles had city fuel efficiency of 30 mpg or less, while just over 40% of the front wheel drive vehicles had city fuel efficiency over 30 mpg.

Based on these results, we expect fuel and drive will be useful predictors of city fuel efficiency over 30 mpg. The complete separation for six and eight cylinder vehicles would be problematic for using the cylinders variables, though based on the results we should expect that six and eight cylinder vehicles would be less fuel efficient.

(b) Here we use backward elimination starting with all three categorical predictors in model. With all predictors in the model, the algorithm fails to converge due to the separation in the six and eight cylinder vehicles. This is the reason for the warning. The cylinders variables can be removed without significant loss of information as indicated by the residual chi-square test and the Wald chi-square test used in backward elimination. Neither of the other two terms could be removed without losing significantly more information than expected due to chance, so our final model will contain fuel and drive.

WARNING: The validity of the model fit is questionable.

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
2.1211	2	0.3463

Summary of Backward Elimination					
Step	**Effect Removed**	**DF**	**Number In**	**Wald Chi-Square**	**Pr > ChiSq**
1	**cylinders**	2	2	0.0025	0.9987

(c) Results for our final model follow.

Model Information	
Data Set	WORK.AUTOS
Response Variable	cityover30mpg
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

For the final model containing the fuel and cylinder predictors we see that both the fuel and drive parameter estimates are statistical significant and positive. The global tests for non-zero betas also concur that at least one of the betas should be significantly different from zero. The fuel coefficient compares diesel to gas and the drive coefficient compares front wheel drive to rear wheel drive.

We can also see that the AIC for this model is much lower than for the intercept only model, indicating a better fit than a constant model.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	211.388	157.469
SC	214.586	167.065
-2 Log L	209.388	151.469

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	57.9184	2	<.0001
Score	43.6686	2	<.0001
Wald	14.4814	2	0.0007

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-5.2504	1.2690	17.1172	<.0001
fuel	**diesel**	1	3.0322	1.1005	7.5918	0.0059
drive	**fwd**	1	4.7115	1.2690	13.7849	0.0002

Hosmer and Lemeshow's test is insignificant at a .05 level, so we conclude there are no issues of lack of fit in this model. As with the parameter estimates, the odds ratios are both statistically significant. For fuel, we estimate that the odds of a diesel fuel car to have city fuel efficiency over 30 mpg was about 20.74 times that of a gas fuel car in 1985. Front wheel drive cars are estimated to have had odds of over 30 mpg city fuel efficiency about 111 times that of rear wheel drive cars. Neither confidence interval contains 1, but the intervals are pretty wide. While we can determine the odds of city fuel efficiency over 30 mpg was significantly higher for diesel cars than gas cars and significantly higher for front wheel drive cars than for rear wheel drive cars, we do not have a very precise measure for the actual odds ratio.

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
fuel diesel vs gas	20.744	2.400	179.323
drive fwd vs rwd	111.220	9.247	>999.999

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
0.4729	1	0.4916

Exercise 2

(a) The model with all four continuous predictors definitely fits the data better than a constant model. The AIC for this model is 52.745, which is quite a bit smaller than the AIC of 204.216 for the constant model.

Model Information	
Data Set	WORK.AUTOS
Response Variable	cityover30mpg
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	204.216	52.745
SC	207.387	68.598
-2 Log L	202.216	42.745

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	159.4713	4	<.0001
Score	62.2535	4	<.0001
Wald	11.9662	4	0.0176

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	16.0666	10.8570	2.1899	0.1389
price1k	1	0.0380	0.2551	0.0222	0.8815
rpm	1	0.00340	0.00165	4.2583	0.0391
enginesize	1	0.1425	0.1161	1.5068	0.2196
hp	1	-0.6630	0.2083	10.1288	0.0015

The global tests are all significant, indicating the present of one or more significant parameter estimates. From the parameter estimates and odds ratios, we can see that rpm and hp are both significant and so we would likely want to retain them. price1k and enginesize are both insignificant both in their parameter estimates and odds ratios, so we may be able to remove one or both of those terms.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
price1k	1.039	0.630	1.712
rpm	1.003	1.000	1.007
enginesize	1.153	0.918	1.448
hp	0.515	0.343	0.775

(b) Using backward selection, price1k, enginesize and rpm are all removed at the .05 level leaving only the hp term in the model. The residual chi-square tests follow showing that an insignificant amount of information is removed as price1k, enginesize and rpm are removed.

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
0.0223	1	0.8814

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
3.1063	2	0.2116

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
5.2055	3	0.1574

Summary of Backward Elimination					
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	price1k	1	3	0.0222	0.8815
2	enginesize	1	2	2.0385	0.1534
3	rpm	1	1	2.2338	0.1350

(c) In the final model, we can see that hp term is significant and negative with an estimate of -.396, so the expected log odds for greater than 30 mpg city fuel efficiency decreased as horsepower increased. The global tests also agree that there are significant non-constant terms in the model.

The AIC for this model is 53.54, which is a lot smaller than the AIC of 210.15 for the constant model, so this model fits the data far better than a constant model.

Model Information	
Data Set	WORK.AUTOS
Response Variable	cityover30mpg
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	210.147	53.538
SC	213.335	59.913
-2 Log L	208.147	49.538

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	158.6087	1	<.0001
Score	56.9695	1	<.0001
Wald	15.4881	1	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	28.8604	7.1140	16.4578	<.0001
hp	1	-0.3961	0.1007	15.4881	<.0001

From Hosmer and Lemeshow's test, we see no evidence of a lack of fit. The p-value of .82 is highly insignificant. From the odds ratio estimate we would expect the odds of having city fuel efficiency over 30 mpg in 1985 to be multiplied by .673 for a one unit increase in horsepower, and the estimate is clearly significant since the interval is entirely less than 1. Put another way, for a one unit increase in horsepower, we would expect a 32.7% decrease in the odds of having city fuel efficiency over 30 mpg.

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
hp	0.673	0.552	0.820

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
5.1159	6	0.5290

Exercise 3

(a) We start by considering the full model—the model with all of the potential predictors included. Based on these results, we conclude that we need to account for underdispersion because the scaled deviance of .2826 is much less than 1.

Model Information	
Data Set	WORK.AUTOS
Distribution	Poisson
Link Function	Log
Dependent Variable	hwaympg

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	167	47.2017	0.2826
Scaled Deviance	167	47.2017	0.2826
Pearson Chi-Square	167	48.0247	0.2876
Scaled Pearson X2	167	48.0247	0.2876
Log Likelihood		13538.0357	
Full Log Likelihood		-486.7354	
AIC (smaller is better)		991.4708	
AICC (smaller is better)		992.5551	
BIC (smaller is better)		1020.0051	

After accounting for the additional dispersion parameter, we see numerous significant terms in the type three analysis and insignificant price1k and rpm terms. We should consider removing each of them separately. The type 1 analysis would lead us to a similar conclusion about terms we might consider removing. Here, the scale has been estimated from the data, so we need to focus on the F tests in the type 1 and type 3 analyses.

Model Information	
Data Set	WORK.AUTOS
Distribution	Poisson
Link Function	Log
Dependent Variable	hwaympg

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	167	47.2017	0.2826
Scaled Deviance	167	167.0000	1.0000
Pearson Chi-Square	167	48.0247	0.2876
Scaled Pearson X2	167	169.9118	1.0174
Log Likelihood		47897.6940	
Full Log Likelihood		-486.7354	
AIC (smaller is better)		991.4708	
AICC (smaller is better)		992.5551	
BIC (smaller is better)		1020.0051	

LR Statistics For Type 1 Analysis							
Source	**Deviance**	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
Intercept	247.2911						
fuel	228.5468	1	167	66.32	<.0001	66.32	<.0001
drive	128.2710	1	167	354.78	<.0001	354.78	<.0001
hp	55.9285	1	167	255.95	<.0001	255.95	<.0001
enginesize	51.0173	1	167	17.38	<.0001	17.38	<.0001
cylinders	47.3348	2	167	6.51	0.0019	13.03	0.0015
price1k	47.2817	1	167	0.19	0.6651	0.19	0.6645
rpm	47.2017	1	167	0.28	0.5955	0.28	0.5948

LR Statistics For Type 3 Analysis						
Source	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
fuel	1	167	14.92	0.0002	14.92	0.0001
drive	1	167	8.98	0.0032	8.98	0.0027
hp	1	167	40.17	<.0001	40.17	<.0001
enginesize	1	167	5.54	0.0197	5.54	0.0186
cylinders	2	167	6.34	0.0022	12.69	0.0018
price1k	1	167	0.25	0.6158	0.25	0.6151
rpm	1	167	0.28	0.5955	0.28	0.5948

When we remove rpm from the model, we see that price1k is still highly insignificant

Model Information	
Data Set	WORK.AUTOS
Distribution	Poisson
Link Function	Log
Dependent Variable	hwaympg

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	168	47.2817	0.2814
Scaled Deviance	168	168.0000	1.0000
Pearson Chi-Square	168	48.0811	0.2862
Scaled Pearson X2	168	170.8407	1.0169
Log Likelihood		48102.8610	
Full Log Likelihood		-486.7754	
AIC (smaller is better)		989.5508	
AICC (smaller is better)		990.4130	
BIC (smaller is better)		1014.9146	

LR Statistics For Type 1 Analysis							
Source	**Deviance**	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
Intercept	247.2911						
fuel	228.5468	1	168	66.60	<.0001	66.60	<.0001
drive	128.2710	1	168	356.30	<.0001	356.30	<.0001
hp	55.9285	1	168	257.05	<.0001	257.05	<.0001
enginesize	51.0173	1	168	17.45	<.0001	17.45	<.0001
cylinders	47.3348	2	168	6.54	0.0018	13.08	0.0014
price1k	47.2817	1	168	0.19	0.6644	0.19	0.6638

LR Statistics For Type 3 Analysis						
Source	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
fuel	1	168	15.02	0.0002	15.02	0.0001
drive	1	168	9.21	0.0028	9.21	0.0024
hp	1	168	45.30	<.0001	45.30	<.0001
enginesize	1	168	9.51	0.0024	9.51	0.0020
cylinders	2	168	6.48	0.0019	12.97	0.0015
price1k	1	168	0.19	0.6644	0.19	0.6638

Retaining rpm and removing price1k instead, we see that rpm is highly insignificant, thus we will remove both from the model. Our final model will contain fuel, drive, hp, enginesize, and cylinders.

Model Information	
Data Set	WORK.AUTOS
Distribution	Poisson
Link Function	Log
Dependent Variable	hwaympg

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	171	56.8097	0.3322
Scaled Deviance	171	171.0000	1.0000
Pearson Chi-Square	171	58.0791	0.3396
Scaled Pearson X2	171	174.8209	1.0223
Log Likelihood		41647.0350	
Full Log Likelihood		-499.7304	
AIC (smaller is better)		1015.4607	
AICC (smaller is better)		1016.3078	
BIC (smaller is better)		1040.9598	

LR Statistics For Type 1 Analysis							
Source	**Deviance**	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
Intercept	255.4529						
fuel	237.4928	1	171	54.06	<.0001	54.06	<.0001
drive	133.0720	1	171	314.31	<.0001	314.31	<.0001
hp	70.3837	1	171	188.69	<.0001	188.69	<.0001
enginesize	61.2341	1	171	27.54	<.0001	27.54	<.0001
cylinders	56.8497	2	171	6.60	0.0017	13.20	0.0014
rpm	56.8097	1	171	0.12	0.7291	0.12	0.7286

LR Statistics For Type 3 Analysis						
Source	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
fuel	1	171	20.98	<.0001	20.98	<.0001
drive	1	171	9.38	0.0026	9.38	0.0022
hp	1	171	19.22	<.0001	19.22	<.0001
enginesize	1	171	31.24	<.0001	31.24	<.0001
cylinders	2	171	6.61	0.0017	13.22	0.0013
rpm	1	171	0.12	0.7291	0.12	0.7286

(b) The type one and type 3 analysis F statistics clearly show that the terms in the model are statistically significant. There are some noticeable changes in the scale estimates as we remove

terms in this case, so we cannot compare the AIC for this model with the previous ones. If we had obtained a scale estimate and held that constant across models, we could compare directly.

Model Information	
Data Set	WORK.AUTOS
Distribution	Poisson
Link Function	Log
Dependent Variable	hwaympg

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	172	56.8497	0.3305
Scaled Deviance	172	172.0000	1.0000
Pearson Chi-Square	172	58.1627	0.3382
Scaled Pearson X2	172	175.9726	1.0231
Log Likelihood		41861.0578	
Full Log Likelihood		-499.7504	
AIC (smaller is better)		1013.5007	
AICC (smaller is better)		1014.1557	
BIC (smaller is better)		1035.8124	

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	4.1363	0.0858	3.9681	4.3045	2322.81	<.0001
fuel	diesel	1	0.1389	0.0279	0.0842	0.1936	24.80	<.0001
fuel	gas	0	0.0000	0.0000	0.0000	0.0000	.	.
drive	fwd	1	0.0678	0.0222	0.0244	0.1113	9.37	0.0022
drive	rwd	0	0.0000	0.0000	0.0000	0.0000	.	.
hp		1	-0.0022	0.0004	-0.0031	-0.0013	25.41	<.0001
enginesize		1	-0.0034	0.0005	-0.0045	-0.0023	39.02	<.0001
cylinders	eight	1	0.1135	0.0709	-0.0256	0.2525	2.56	0.1097
cylinders	four	1	-0.1260	0.0375	-0.1995	-0.0526	11.32	0.0008
cylinders	six	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	0.5749	0.0000	0.5749	0.5749		

Note: The scale parameter was estimated by the square root of DEVIANCE/DOF.

In our final model, we can see that all parameter estimates are significantly different from 0 with the exception of the estimate comparing eight and six cylinder cars. The positive estimate for

diesel fuel indicates higher highway fuel efficiency for diesels than gas powered cars. The positive estimate for front wheel drive indicates higher highway fuel efficiency for front wheel drive cars than rear wheel drive cars. The negative estimates for hp and engine size indicate reduced fuel efficiency as horsepower and engine size increase. The negative estimate for four cylinders indicates a reduction in expected highway fuel efficiency compared to six cylinder vehicles.

Quantitatively, we need to exponentiate the parameter estimates to get the expected multiplicative change in highway fuel efficiency as these predictors change. After taking the exponential of the significant parameter estimates, we see that we expect multiplicative factors of 1.15 for diesel as compared to gas, 1.07 for front wheel drive as compared to rear wheel drive, .9978 for a one unit increase in horsepower, .997 for a one unit increase in engine size, and .882 for four cylinders as compared to six cylinders.

LR Statistics For Type 1 Analysis							
Source	**Deviance**	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
Intercept	255.4529						
fuel	237.4928	1	172	54.34	<.0001	54.34	<.0001
drive	133.0720	1	172	315.93	<.0001	315.93	<.0001
hp	70.3837	1	172	189.66	<.0001	189.66	<.0001
enginesize	61.2341	1	172	27.68	<.0001	27.68	<.0001
cylinders	56.8497	2	172	6.63	0.0017	13.27	0.0013

LR Statistics For Type 3 Analysis						
Source	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
fuel	1	172	24.18	<.0001	24.18	<.0001
drive	1	172	9.41	0.0025	9.41	0.0022
hp	1	172	26.19	<.0001	26.19	<.0001
enginesize	1	172	38.93	<.0001	38.93	<.0001
cylinders	2	172	6.63	0.0017	13.27	0.0013