Homework 3

Name: Zixin Ouyang

Exercise 1

			highwaympg		
			Mean	Std	N
weight_cat	**type**	**ndoors**			
Heavy	**0**	**four**	26.98	4.17	40
		two	26.48	2.75	21
	1	**four**	22.27	3.45	15
		two	23.79	3.96	14
Light	**0**	**four**	34.96	4.42	55
		two	35.12	6.88	49
	1	**four**	29.50	0.71	2
		two	29.00	.	1

a. From the tabulation, it looks like light cars might be more efficient than heavy cars. Also, non-luxury brand cars might be more efficient than luxury brand cars. There is no a big difference across ndoors. We see that the data is not balanced, so we use proc glm.

The GLM Procedure

Dependent Variable: highwaympg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4600.402313	1533.467438	64.91	<.0001
Error	193	4559.546926	23.624595		
Corrected Total	196	9159.949239			

R-Square	Coeff Var	Root MSE	highwaympg Mean
0.502230	15.86876	4.860514	30.62944

Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight_cat	1	2666.244462	2666.244462	112.86	<.0001
type	1	371.828738	371.828738	15.74	0.0001
ndoors	1	1.789603	1.789603	0.08	0.7834

The GLM Procedure

Dependent Variable: highwaympg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4598.612709	2299.306355	97.79	<.0001
Error	194	4561.336529	23.512044		
Corrected Total	196	9159.949239			

R-Square	Coeff Var	Root MSE	highwaympg Mean
0.502035	15.83092	4.848922	30.62944

Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight_cat	1	2707.376160	2707.376160	115.15	<.0001
type	1	370.039587	370.039587	15.74	0.0001

b. Removing ndoors we get our best main effects model. The model is highly significant, both weight_cat and type are highly significant, and the model describes 50.2% of the variation in highway fuel efficiency.

The GLMSELECT Procedure

Data Set	WORK.AUTO
Dependent Variable	highwaympg
Selection Method	Stepwise
Select Criterion	Significance Level
Stop Criterion	Significance Level
Entry Significance Level (SLE)	0.05
Stay Significance Level (SLS)	0.05
Effect Hierarchy Enforced	None

Number of Observations Read	197
Number of Observations Used	197

Class Level Information		
Class	**Levels**	**Values**
weight_cat	2	Heavy Light
type	2	0 1
ndoors	2	four two

Dimensions	
Number of Effects	4
Number of Parameters	7

The GLMSELECT Procedure

Stepwise Selection Summary						
Step	**Effect Entered**	**Effect Removed**	**Number Effects In**	**Number Parms In**	**F Value**	**Pr > F**
0	**Intercept**		1	1	0.00	1.0000
1	**weight_cat**		2	2	167.21	<.0001
2	**type**		3	3	15.74	0.0001

Selection stopped because the candidate for entry has SLE > 0.05 and the candidate for removal has SLS < 0.05.

Stop Details					
Candidate For	**Effect**	**Candidate Significance**		**Compare Significance**	
Entry	ndoors	0.7834	>	0.0500	(SLE)
Removal	type	0.0001	<	0.0500	(SLS)

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 2).

Effects:	Intercept weight_cat type

Analysis of Variance				
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**
Model	2	4598.61271	2299.30635	97.79
Error	194	4561.33653	23.51204	
Corrected Total	196	9159.94924		

Root MSE	4.84892
Dependent Mean	30.62944
R-Square	0.5020
Adj R-Sq	0.4969
AIC	824.00694
AICC	824.21528
SBC	634.85656

Parameter Estimates				
Parameter	**DF**	**Estimate**	**Standard Error**	**t Value**
Intercept	1	30.943058	1.097186	28.20
weight_cat Heavy	1	-8.109582	0.755735	-10.73
weight_cat Light	0	0	.	.
type 0	1	4.048969	1.020623	3.97
type 1	0	0	.	.

The GLMSELECT Procedure

Data Set	WORK.AUTO
Dependent Variable	highwaympg
Selection Method	Stepwise
Select Criterion	Significance Level
Stop Criterion	Significance Level
Entry Significance Level (SLE)	0.05
Stay Significance Level (SLS)	0.05
Effect Hierarchy Enforced	None

Number of Observations Read	197
Number of Observations Used	197

Class Level Information		
Class	**Levels**	**Values**
weight_cat	2	Heavy Light
type	2	0 1

Dimensions	
Number of Effects	4
Number of Parameters	9

The GLMSELECT Procedure

Stepwise Selection Summary						
Step	**Effect Entered**	**Effect Removed**	**Number Effects In**	**Number Parms In**	**F Value**	**Pr > F**
0	**Intercept**		1	1	0.00	1.0000
1	**weight_cat*type**		2	4	65.12	<.0001

Selection stopped because all candidate effects for entry are linearly dependent on effects in the model.

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 1).

Effects:	Intercept weight_cat*type

Analysis of Variance				
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**
Model	3	4607.79707	1535.93236	65.12
Error	193	4552.15216	23.58628	
Corrected Total	196	9159.94924		

Root MSE	4.85657
Dependent Mean	30.62944
R-Square	0.5030
Adj R-Sq	0.4953
AIC	825.60988
AICC	825.92402
SBC	639.74269

Parameter Estimates				
Parameter	**DF**	**Estimate**	**Standard Error**	**t Value**
Intercept	1	29.333333	2.803943	10.46
weight_cat*type Heavy 0	1	-2.530055	2.872064	-0.88
weight_cat*type Heavy 1	1	-6.333333	2.945406	-2.15
weight_cat*type Light 0	1	5.705128	2.844096	2.01
weight_cat*type Light 1	0	0	.	.

c. The statistically significant main predictors are weight_cat and type. The p-value of interaction term weight_cat*type is less than 0.0001, so it is significant and we retain it in the model. The model contains weight_cat, type and their interaction term.

The GLM Procedure

Dependent Variable: highwaympg

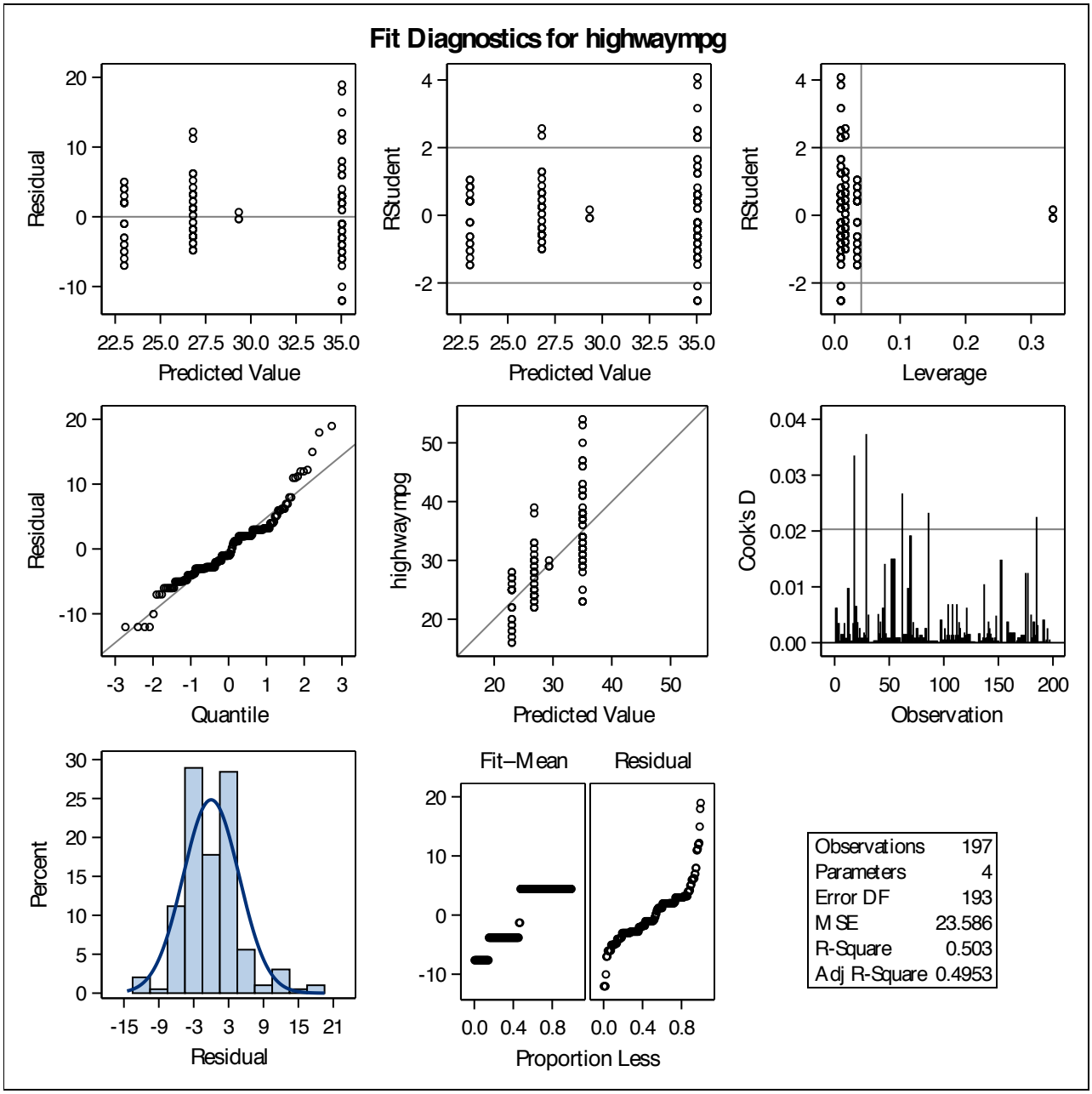
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4607.797074	1535.932358	65.12	<.0001
Error	193	4552.152165	23.586281		
Corrected Total	196	9159.949239			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weight_cat	1	4228.573122	4228.573122	179.28	<.0001
type	1	370.039587	370.039587	15.69	0.0001
weight_cat*type	1	9.184364	9.184364	0.39	0.5334

Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight_cat	1	538.9239061	538.9239061	22.85	<.0001
type	1	229.5685145	229.5685145	9.73	0.0021
weight_cat*type	1	9.1843644	9.1843644	0.39	0.5334

The GLM Procedure

Dependent Variable: highwaympg



The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

weight_cat	highwaympg LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
Heavy	24.9016393	<.0001
Light	32.1858974	

Least Squares Means for Effect weight_cat				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	**2**	-7.284258	-10.289859	-4.278657

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

type	highwaympg LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
0	30.9208701	0.0021
1	26.1666667	

Least Squares Means for Effect type				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	**2**	4.754203	1.748602	7.759804

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

weight_cat	type	highwaympg LSMEAN	LSMEAN Number
Heavy	0	26.8032787	1
Heavy	1	23.0000000	2
Light	0	35.0384615	3
Light	1	29.3333333	4

Least Squares Means for effect weight_cat*type Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: highwaympg				
i/j	1	2	3	4
1		0.0035	<.0001	0.8148
2	0.0035		<.0001	0.1412
3	<.0001	<.0001		0.1893
4	0.8148	0.1412	0.1893	

Least Squares Means for Effect weight_cat*type				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	3.803279	0.964396	6.642161
1	3	-8.235183	-10.264970	-6.205396
1	4	-2.530055	-9.973166	4.913057
2	3	-12.038462	-14.681480	-9.395443
2	4	-6.333333	-13.966512	1.299845
3	4	5.705128	-1.665503	13.075759

The UNIVARIATE Procedure
Variable: resid

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.942247	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.116506	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.45097	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	2.978727	**Pr > A-Sq**	<0.0050

d. The model in part c is significant with a p-value less than 0.0001. The model describes 50.3% of the variation in highway fuel efficiency. Rank the levels from the highest highway mpg to the lowest highway mpg based on LS mean estimates: Light-0, Light-1, Heavy-0, Heavy-1. Based on the result of tests for normality, the normality assumption for the residuals is valid.

Exercise 2

a.

Class Level Information		
Class	**Levels**	**Values**
weight_cat	2	Heavy Light
type	2	0 1
ndoors	2	four two

Number of Observations Read	197
Number of Observations Used	197

The GLM Procedure

Dependent Variable: highwaympg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	6271.882240	1567.970560	104.24	<.0001
Error	192	2888.066998	15.042016		
Corrected Total	196	9159.949239			

R-Square	Coeff Var	Root MSE	highwaympg Mean
0.684707	12.66234	3.878404	30.62944

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weight_cat	1	4228.573122	4228.573122	281.12	<.0001
type	1	370.039587	370.039587	24.60	<.0001
ndoors	1	1.789603	1.789603	0.12	0.7305
horsepower	1	1671.479928	1671.479928	111.12	<.0001

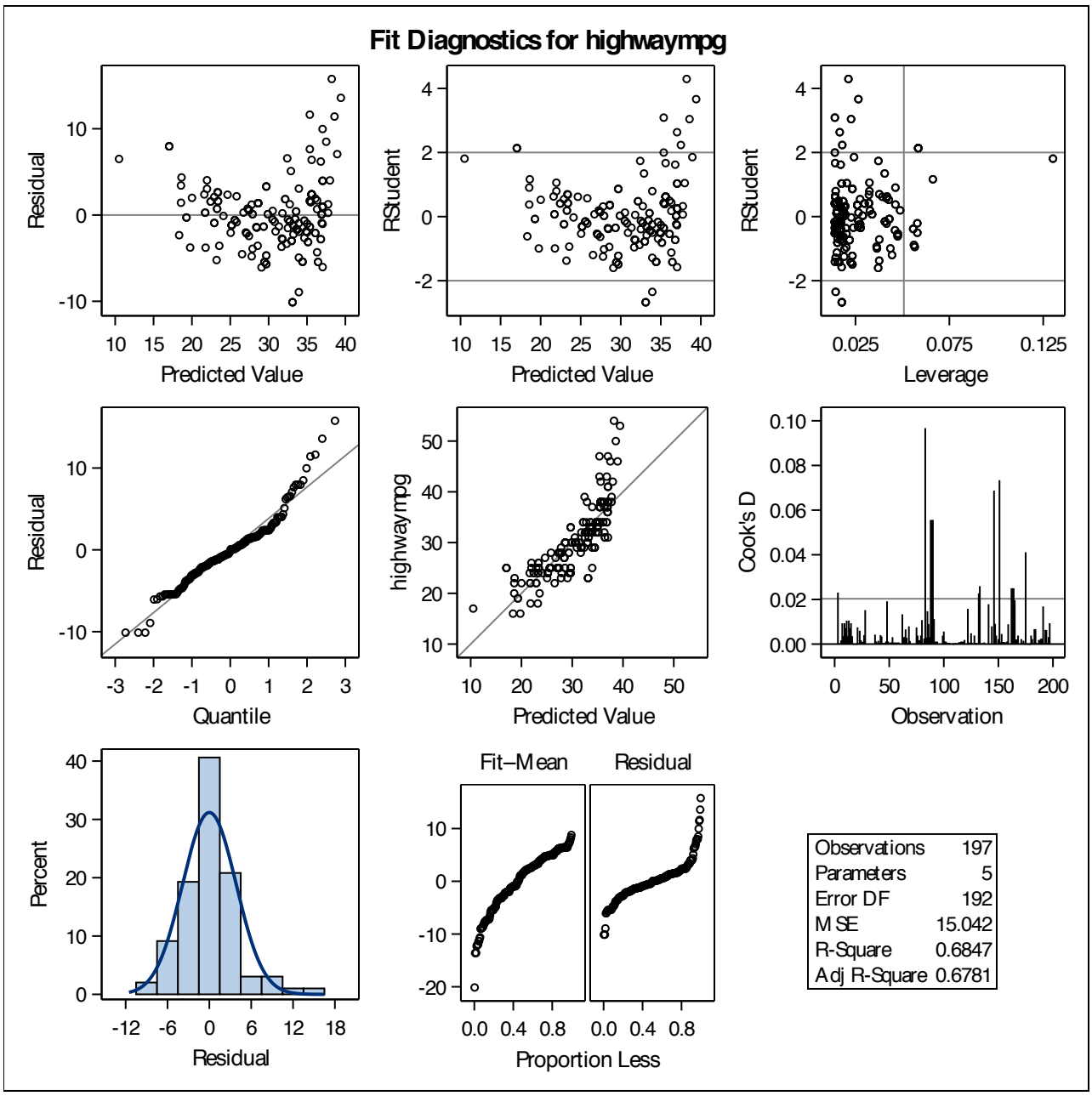
Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight_cat	1	174.183099	174.183099	11.58	0.0008
type	1	11.787770	11.787770	0.78	0.3771
ndoors	1	94.125973	94.125973	6.26	0.0132
horsepower	1	1671.479928	1671.479928	111.12	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	44.34542420	B	1.55738357	28.47	<.0001
weight_cat Heavy	-2.70153281	B	0.79388987	-3.40	0.0008
weight_cat Light	0.00000000	B	.	.	.
type 0	0.77546252	B	0.87598770	0.89	0.3771
type 1	0.00000000	B	.	.	.
ndoors four	-1.43401561	B	0.57326068	-2.50	0.0132
ndoors two	0.00000000	B	.	.	.
horsepower	-0.11887566		0.01127705	-10.54	<.0001

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

The GLM Procedure

Dependent Variable: highwaympg



a. I should not also include weight as a continuous predictor in this model if I include weight_cat as a categorical predictor in my model. Including two weight predictors is redundant.

The GLMSELECT Procedure

Data Set	WORK.AUTO
Dependent Variable	highwaympg
Selection Method	Stepwise
Select Criterion	Significance Level
Stop Criterion	Significance Level
Entry Significance Level (SLE)	0.05
Stay Significance Level (SLS)	0.05
Effect Hierarchy Enforced	None

Number of Observations Read	197
Number of Observations Used	197

Class Level Information		
Class	**Levels**	**Values**
weight_cat	2	Heavy Light
type	2	0 1
ndoors	2	four two

Dimensions	
Number of Effects	5
Number of Parameters	8

The GLMSELECT Procedure

Stepwise Selection Summary						
Step	**Effect Entered**	**Effect Removed**	**Number Effects In**	**Number Parms In**	**F Value**	**Pr > F**
0	**Intercept**		1	1	0.00	1.0000
1	**horsepower**		2	2	355.64	<.0001
2	**weight_cat**		3	3	16.15	<.0001
3	**ndoors**		4	4	6.30	0.0129

Selection stopped because the candidate for entry has SLE > 0.05 and the candidate for removal has SLS < 0.05.

Stop Details					
Candidate For	**Effect**	**Candidate Significance**		**Compare Significance**	
Entry	type	0.3771	>	0.0500	(SLE)
Removal	ndoors	0.0129	<	0.0500	(SLS)

The GLMSELECT Procedure
Selected Model

The selected model is the model at the last step (Step 3).

Effects:	Intercept weight_cat ndoors horsepower

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	3	6260.09447	2086.69816	138.88
Error	193	2899.85477	15.02515	
Corrected Total	196	9159.94924		

Root MSE	3.87623
Dependent Mean	30.62944
R-Square	0.6834
Adj R-Sq	0.6785
AIC	736.77480
AICC	737.08894
SBC	550.90762

Parameter Estimates				
Parameter	DF	Estimate	Standard Error	t Value
Intercept	1	45.384757	1.022665	44.38
weight_cat Heavy	1	-2.741859	0.792137	-3.46
weight_cat Light	0	0	.	.
ndoors four	1	-1.438469	0.572917	-2.51
ndoors two	0	0	.	.
horsepower	1	-0.122436	0.010530	-11.63

b. According to the results of GLMSELECT procedure, the predictors are weight_cat, ndoors and horsepower. However, the predictors chosen from model in exercise 1 are weight_cat and type.

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
weight_cat	2	Heavy Light
ndoors	2	four two

Number of Observations Read	197
Number of Observations Used	197

The GLM Procedure

Dependent Variable: highwaympg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	6260.094471	2086.698157	138.88	<.0001
Error	193	2899.854768	15.025154		
Corrected Total	196	9159.949239			

R-Square	Coeff Var	Root MSE	highwaympg Mean
0.683420	12.65524	3.876229	30.62944

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weight_cat	1	4228.573122	4228.573122	281.43	<.0001
ndoors	1	0.000453	0.000453	0.00	0.9956
horsepower	1	2031.520896	2031.520896	135.21	<.0001

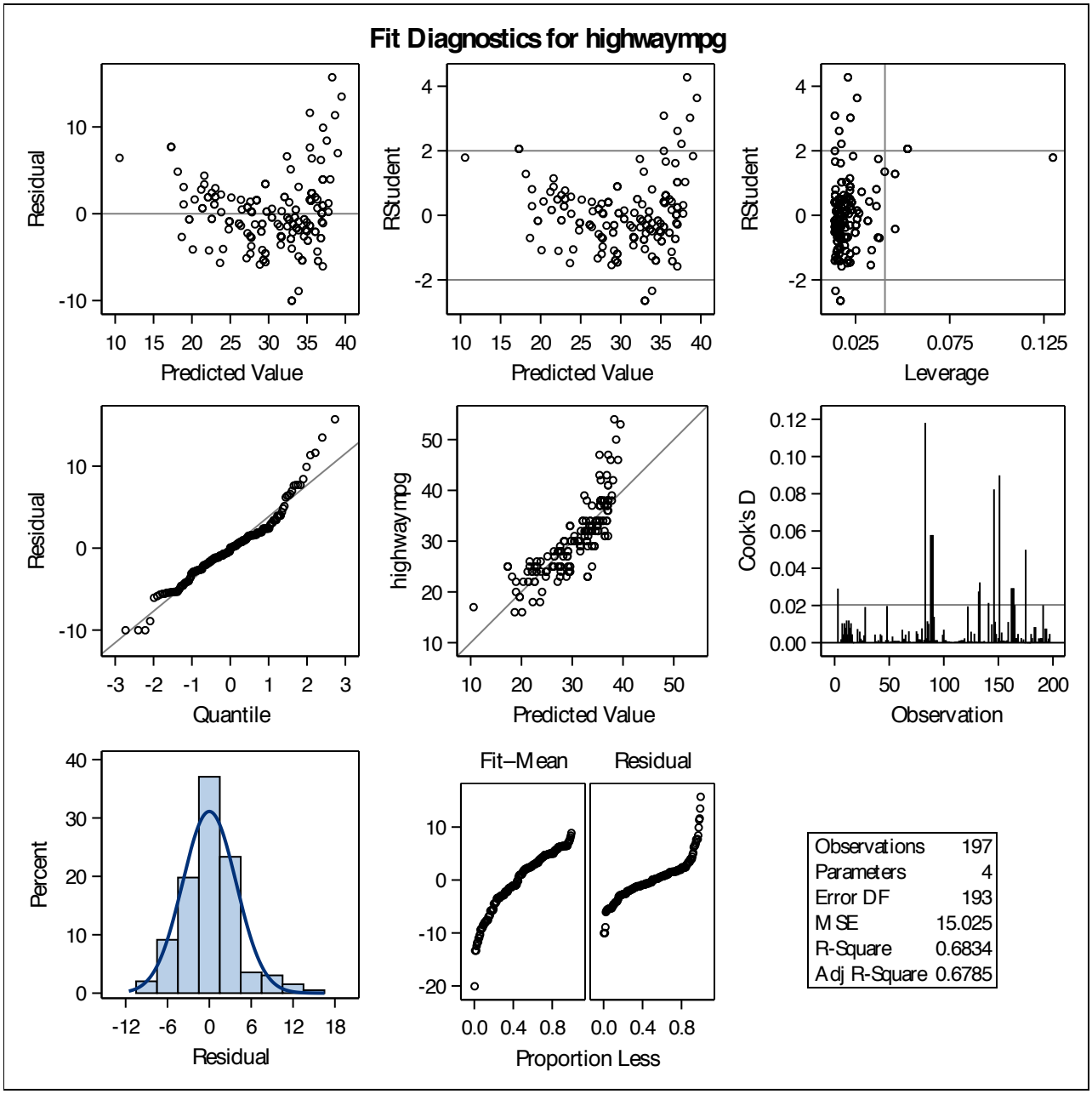
Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight_cat	1	180.014783	180.014783	11.98	0.0007
ndoors	1	94.718735	94.718735	6.30	0.0129
horsepower	1	2031.520896	2031.520896	135.21	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	45.38475713	B	1.02266461	44.38	<.0001
weight_cat Heavy	-2.74185935	B	0.79213747	-3.46	0.0007
weight_cat Light	0.00000000	B	.	.	.
ndoors four	-1.43846852	B	0.57291723	-2.51	0.0129
ndoors two	0.00000000	B	.	.	.
horsepower	-0.12243613		0.01052952	-11.63	<.0001

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

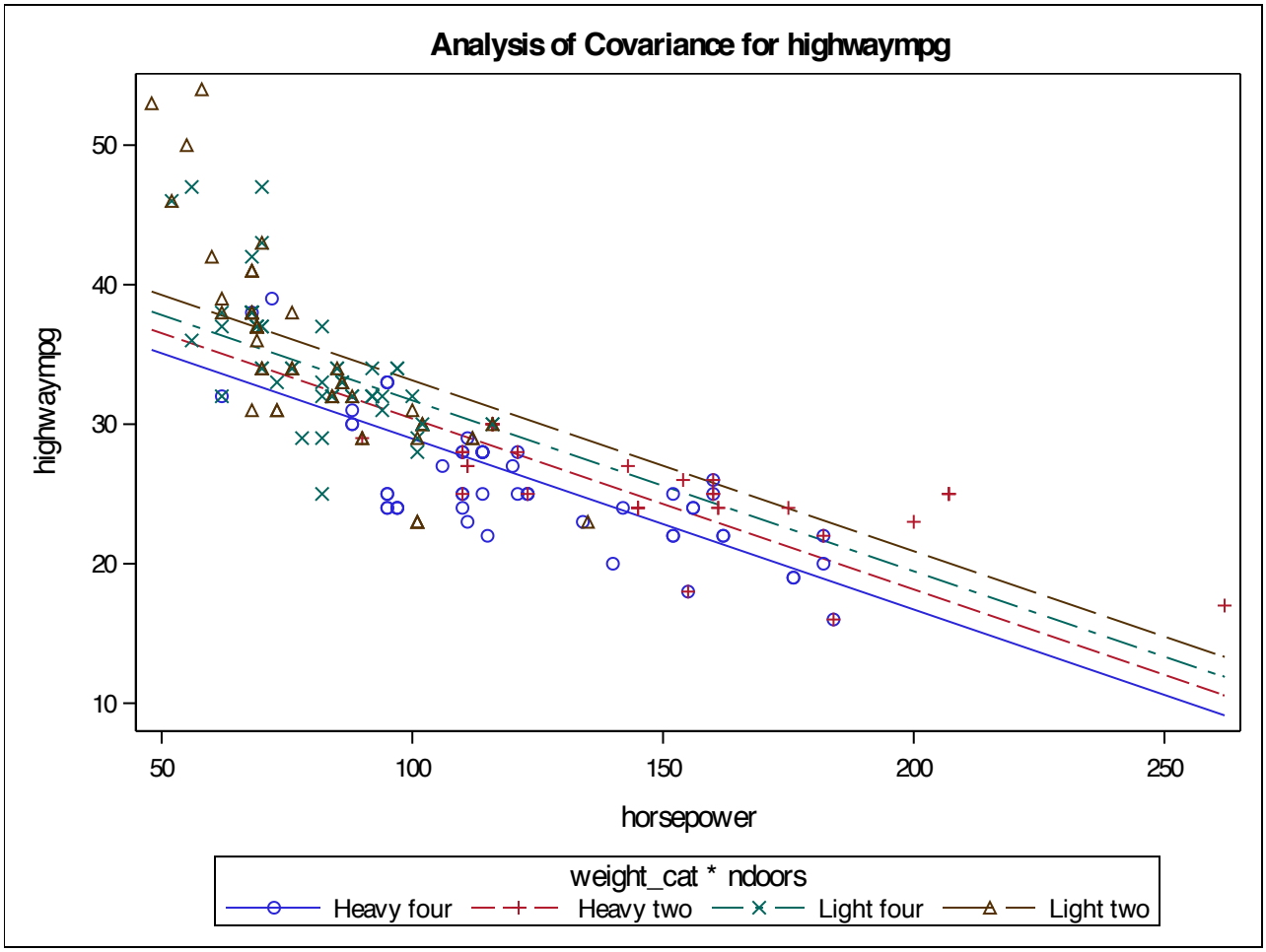
The GLM Procedure

Dependent Variable: highwaympg



The GLM Procedure

Dependent Variable: highwaympg

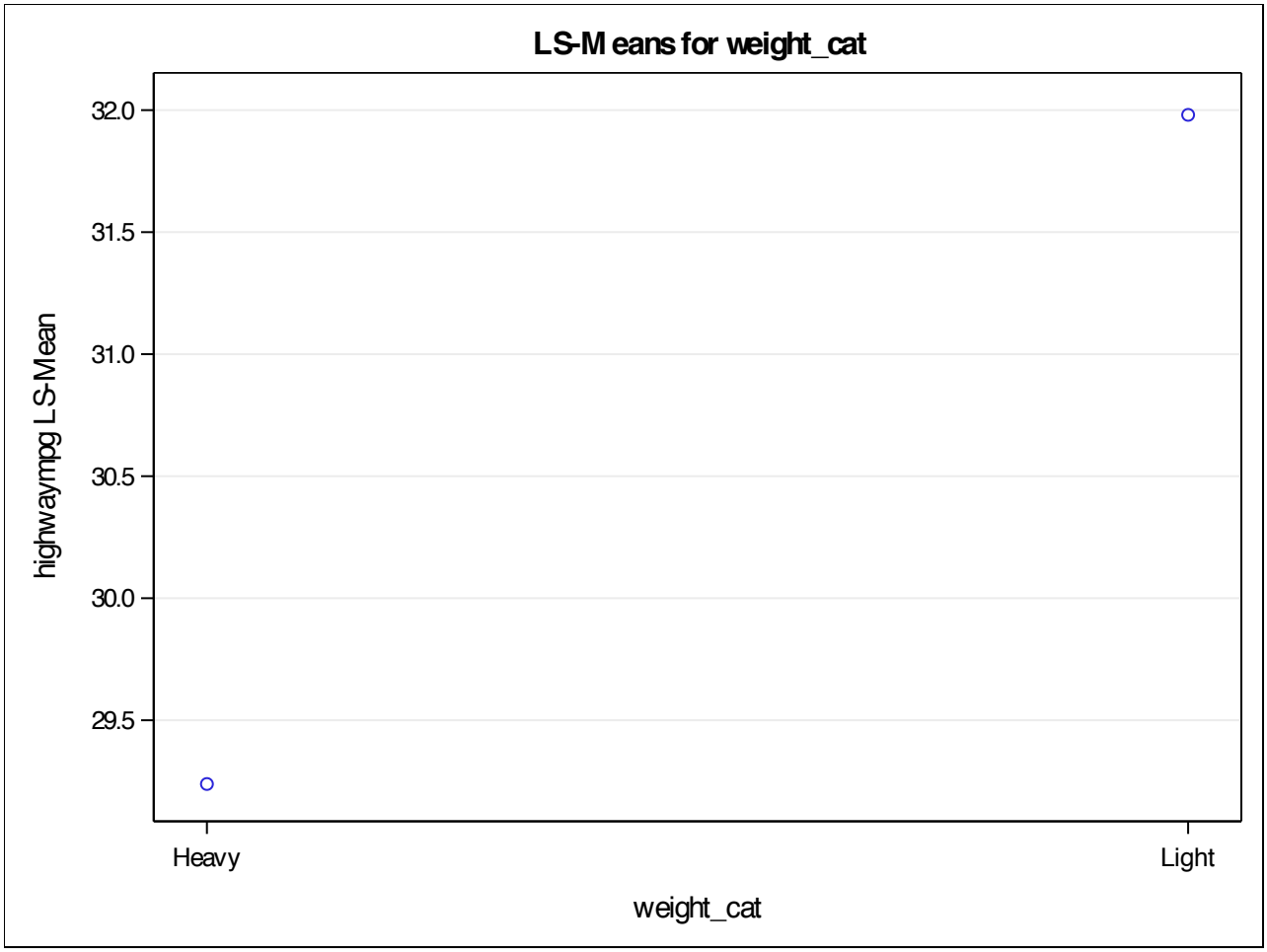


The GLM Procedure
Least Squares Means

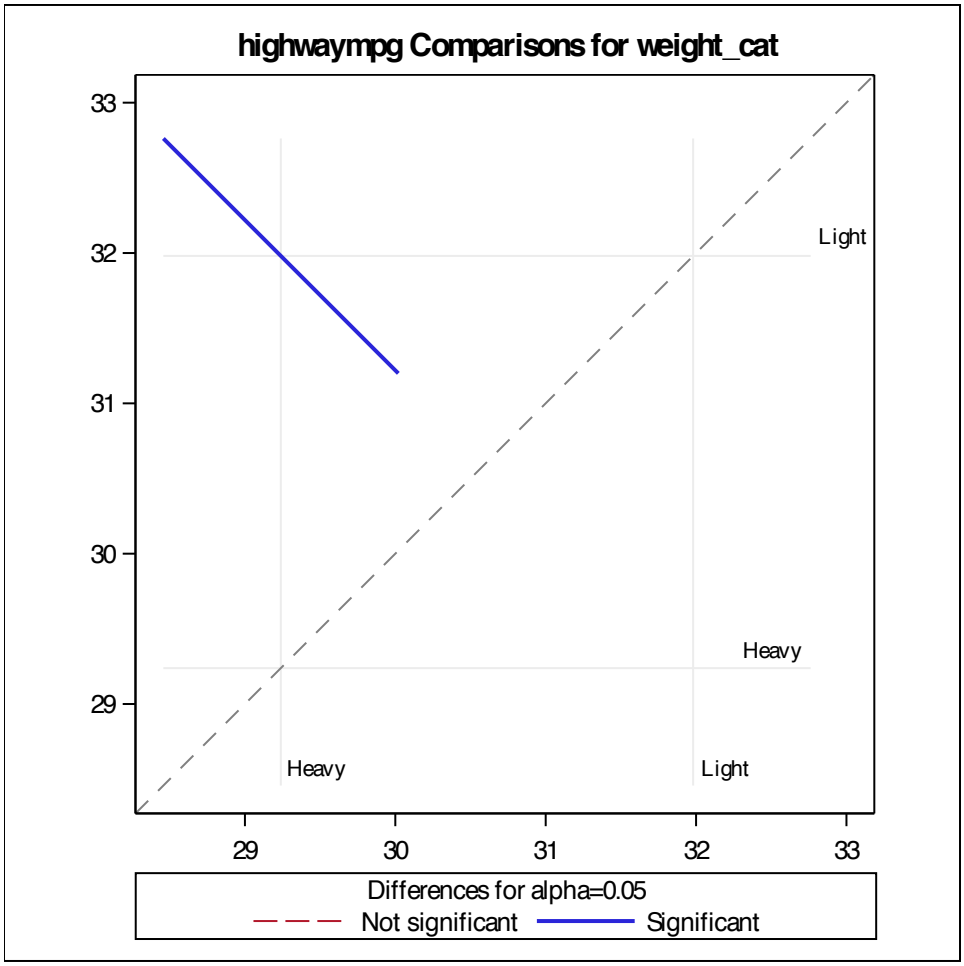
weight_cat	highwaympg LSMEAN	Standard Error	H0:LSMEAN=0	H0:LSMean1=LSMean2
			Pr > |t|	Pr > |t|
Heavy	29.2387836	0.5195726	<.0001	0.0007
Light	31.9806430	0.4504012	<.0001	

weight_cat	highwaympg LSMEAN	95% Confidence Limits	
Heavy	29.238784	28.214014	30.263553
Light	31.980643	31.092302	32.868983

Least Squares Means for Effect weight_cat				
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-2.741859	-4.304217	-1.179502



The GLM Procedure
Least Squares Means

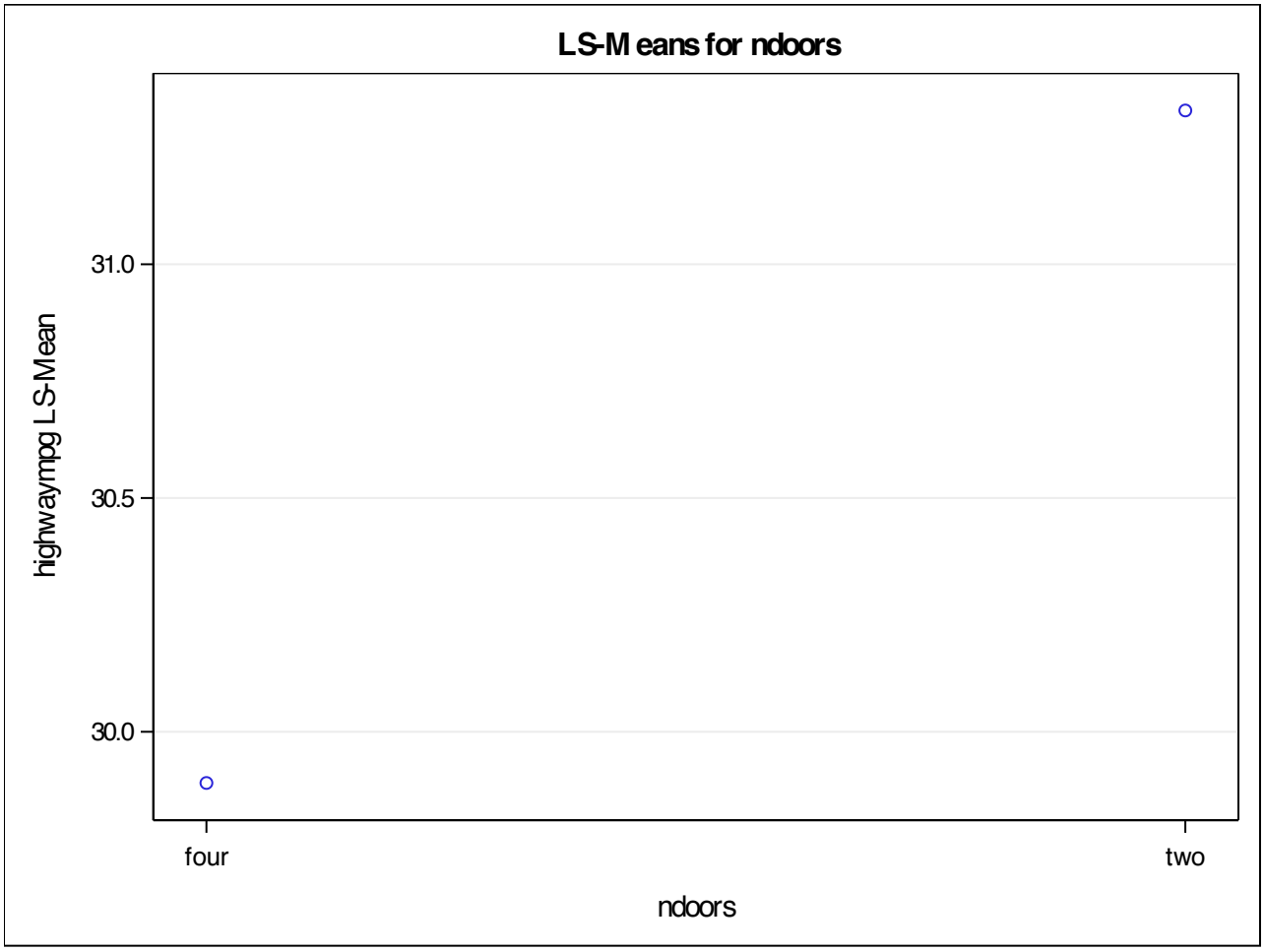


ndoors	highwaympg LSMEAN	Standard Error	H0:LSMEAN=0	H0:LSMean1=LSMean2
			Pr > |t|	Pr > |t|
four	29.8904790	0.3674611	<.0001	0.0129
two	31.3289476	0.4337804	<.0001	

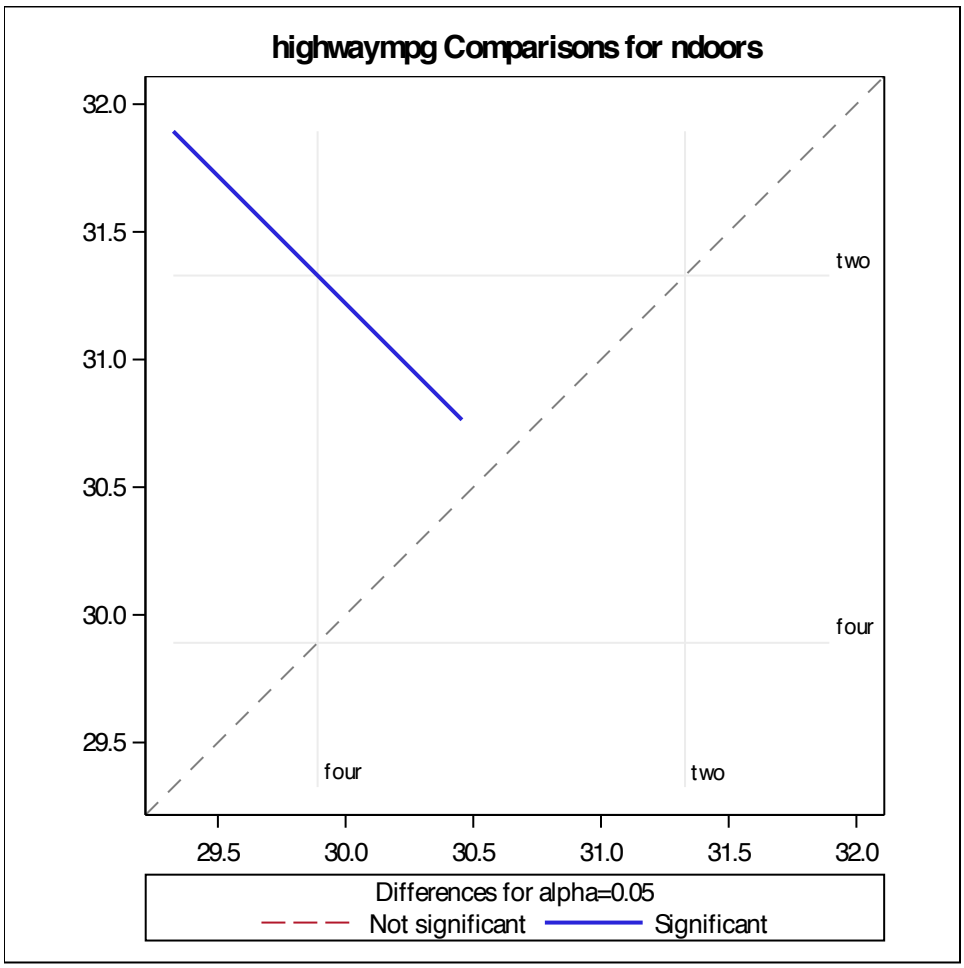
ndoors	highwaympg LSMEAN	95% Confidence Limits	
four	29.890479	29.165724	30.615234
two	31.328948	30.473389	32.184506

The GLM Procedure
Least Squares Means

Least Squares Means for Effect ndoors				
i	j	Difference Between Means	95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-1.438469	-2.568451	-0.308486



The GLM Procedure
Least Squares Means



The UNIVARIATE Procedure
Variable: resid

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.950768	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.113545	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.381356	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	2.443113	**Pr > A-Sq**	<0.0050

c. The model is highly significant with p-value less than 0.0001. The p-values of weight_cat, ndoors and horsepower are less than 0.05 based on the result of type III SS, and the model describes 68.34% of the variation in highway fuel efficiency. The R-square of the ANOVA model in exercise 1 is 50.3%, which is less than 68.34%. Therefore, the ANCOVA model is better in terms of % variation explained. According to the result of tests for normality, the normality assumption for the residuals is valid.

=

The REG Procedure
Model: MODEL1
Dependent Variable: citympg

Exercise 3

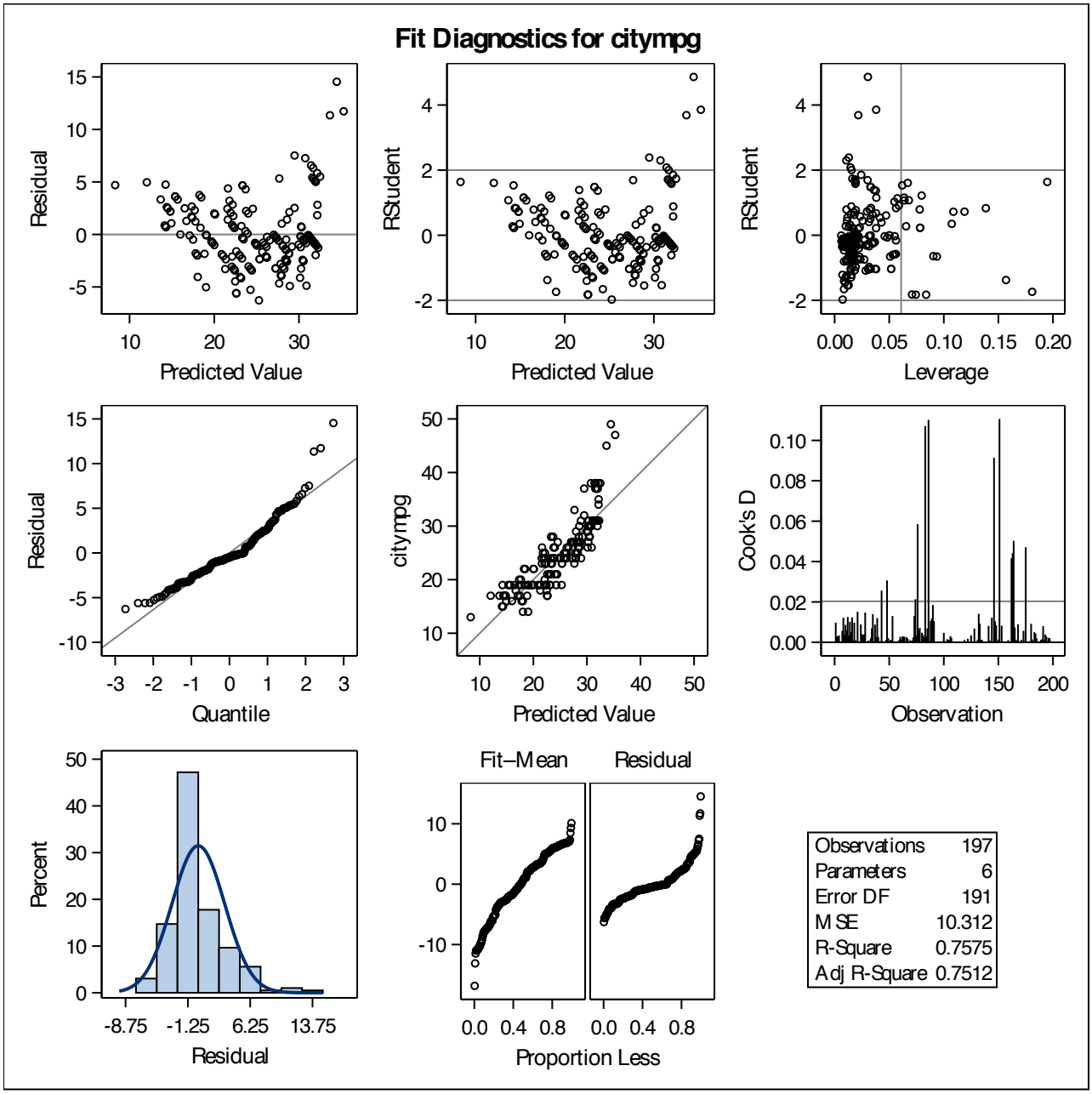
Number of Observations Read	197
Number of Observations Used	197

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	5	6153.86048	1230.77210	119.35	<.0001
Error	191	1969.57099	10.31189		
Corrected Total	196	8123.43147			

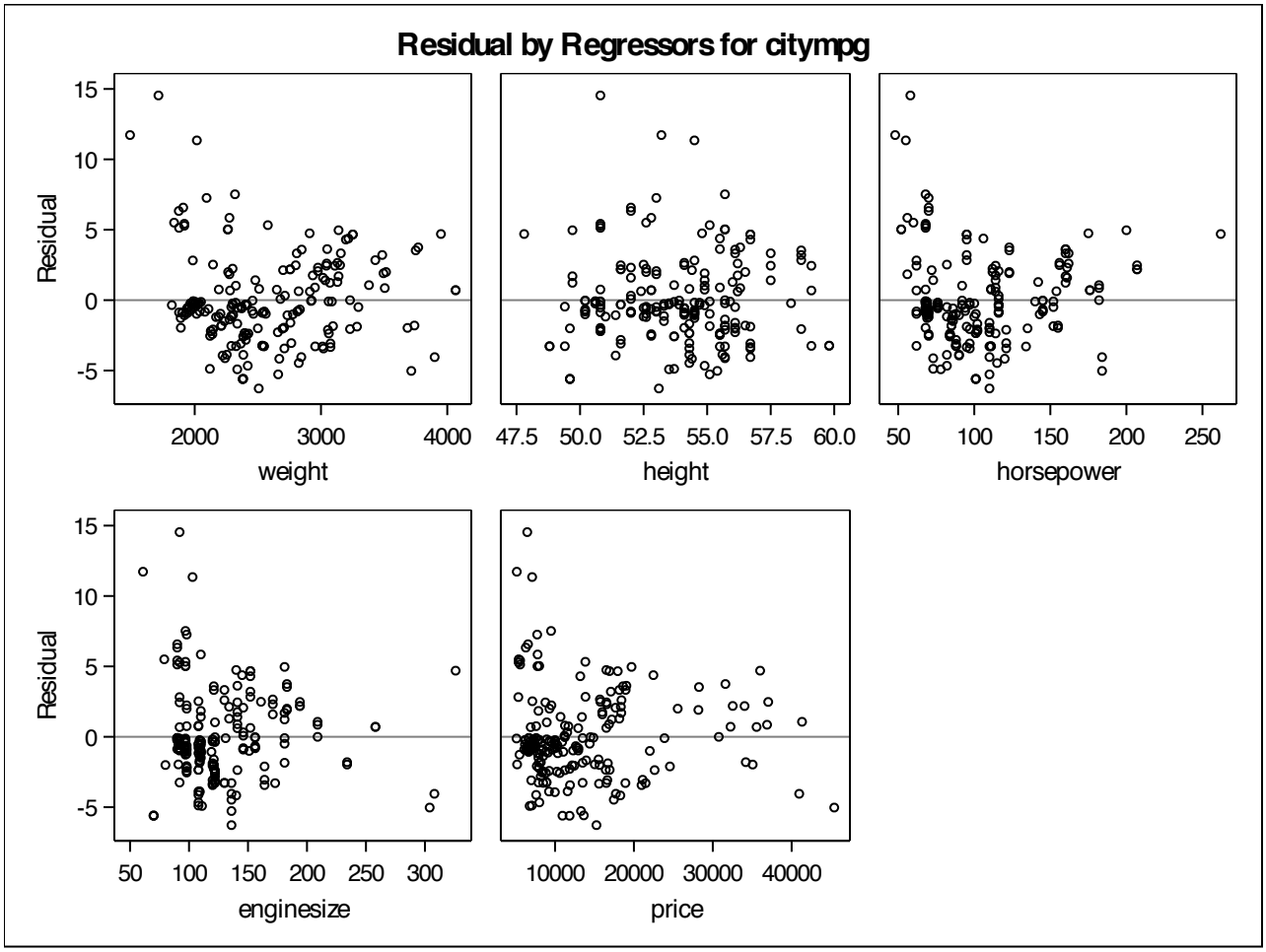
Root MSE	3.21121	**R-Square**	0.7575
Dependent Mean	25.15228	**Adj R-Sq**	0.7512
Coeff Var	12.76708		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	45.64591	5.87030	7.78	<.0001
weight	1	-0.00695	0.00104	-6.69	<.0001
height	1	0.03822	0.11649	0.33	0.7432
horsepower	1	-0.13185	0.01259	-10.47	<.0001
enginesize	1	0.06947	0.01357	5.12	<.0001
price	1	0.00000473	0.00006615	0.07	0.9430

The REG Procedure
Model: MODEL1
Dependent Variable: citympg



The REG Procedure
Model: MODEL1
Dependent Variable: citympg



a. The model is significant with a p-value less than 0.0001. Predictors weight, horsepower, enginesize are significant. 75.75% variance of miles per gallon in the city is explained by the model. Based on significance level of 0.05, height and price can be removed from the model.

The REG Procedure
Model: MODEL1
Dependent Variable: citympg

Summary of Stepwise Selection								
Step	**Variable Entered**	**Variable Removed**	**Number Vars In**	**Partial R-Square**	**Model R-Square**	**C(p)**	**F Value**	**Pr > F**
1	horsepower		1	0.6756	0.6756	62.5393	406.14	<.0001
2	weight		2	0.0406	0.7162	32.5765	27.73	<.0001
3	enginesize		3	0.0412	0.7574	2.1185	32.78	<.0001

b. Predictors height and price are removed by the stepwise selection.

The REG Procedure
Model: MODEL1
Dependent Variable: citympg

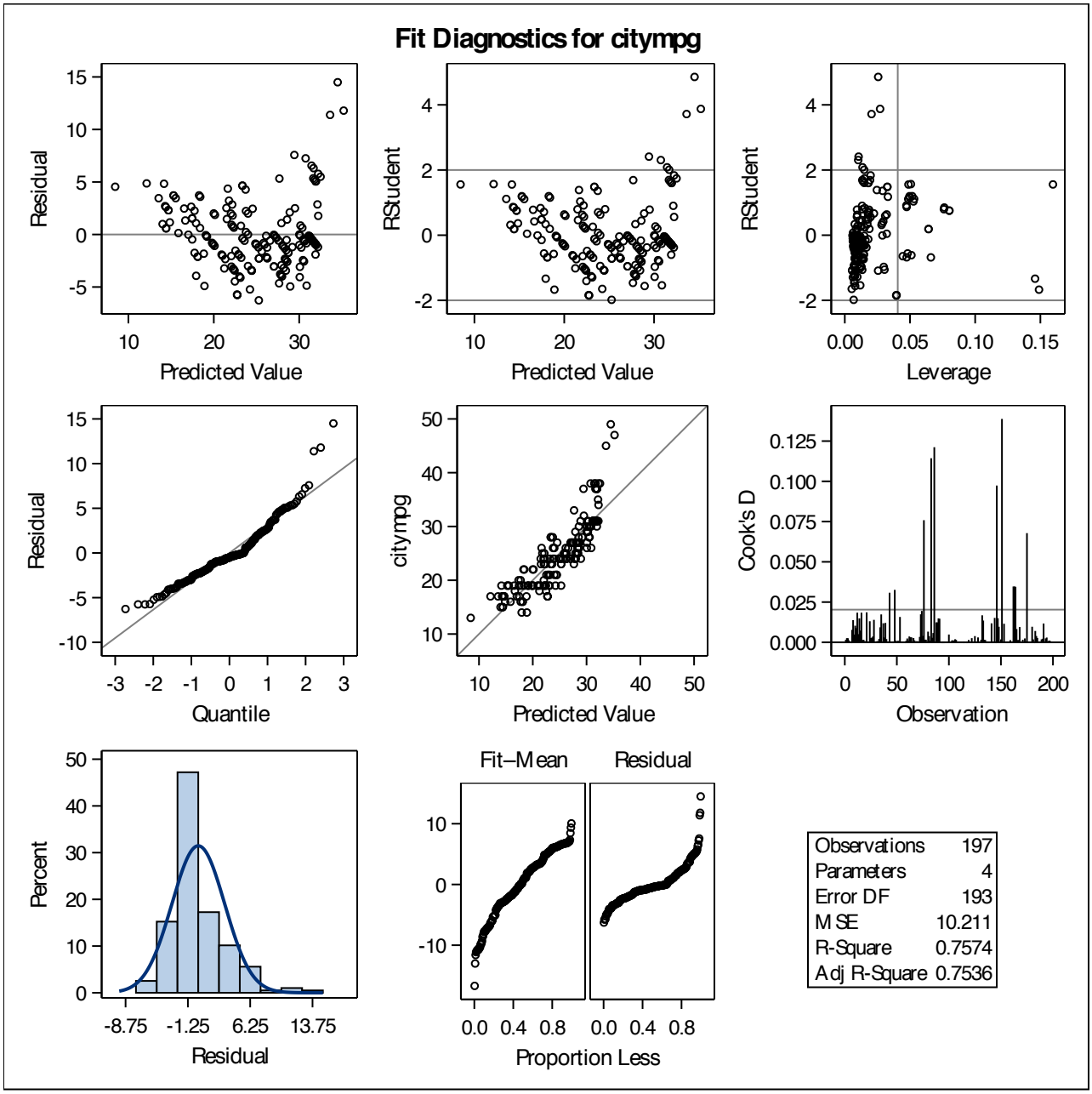
Number of Observations Read	197
Number of Observations Used	197

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	6152.63902	2050.87967	200.84	<.0001
Error	193	1970.79246	10.21136		
Corrected Total	196	8123.43147			

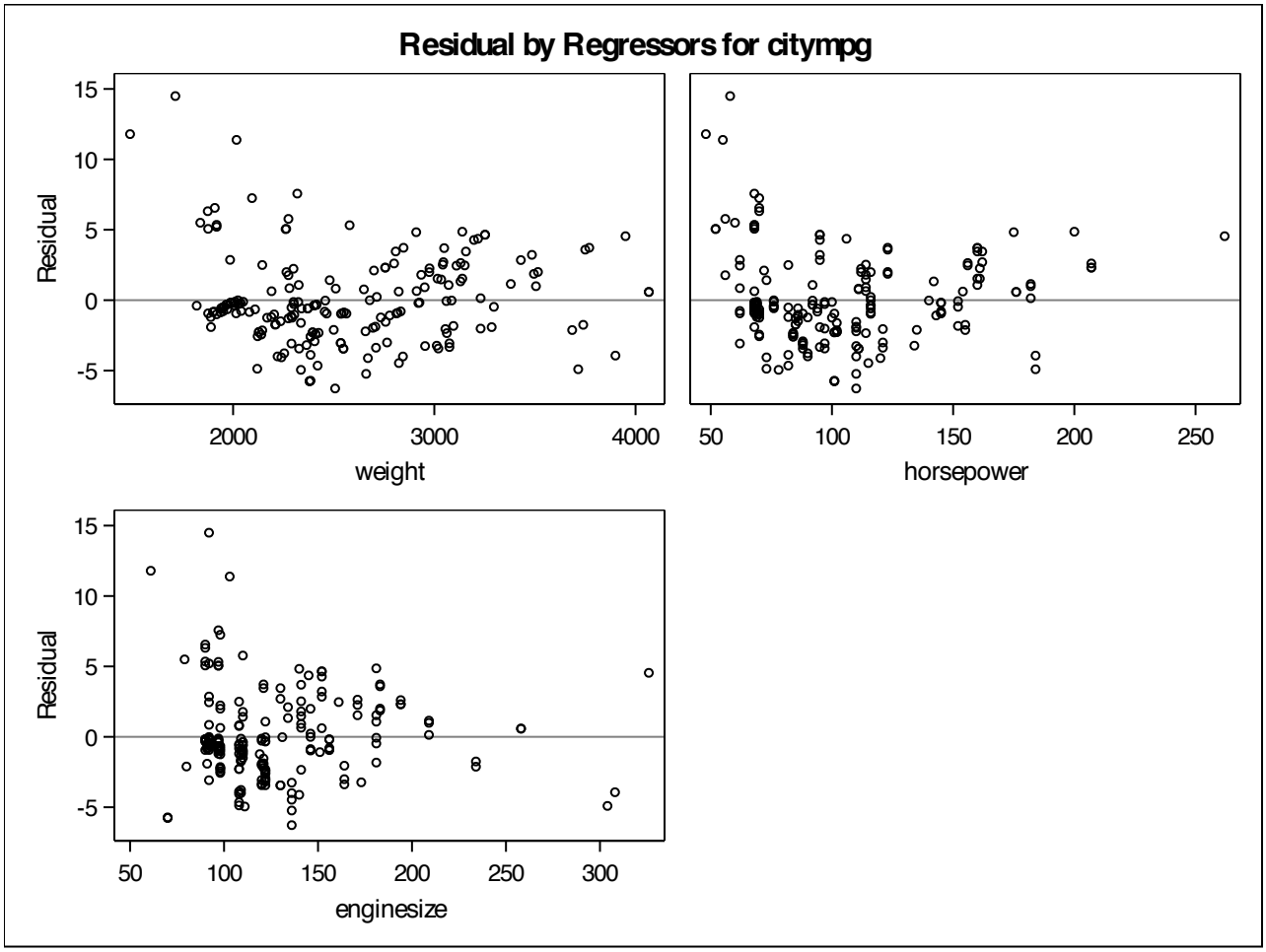
Root MSE	3.19552	**R-Square**	0.7574
Dependent Mean	25.15228	**Adj R-Sq**	0.7536
Coeff Var	12.70470		

Parameter Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	1	47.41053	1.24649	38.04	<.0001	0
weight	1	-0.00674	0.00084456	-7.99	<.0001	3.72741
horsepower	1	-0.13321	0.01096	-12.16	<.0001	3.26470
enginesize	1	0.06929	0.01210	5.73	<.0001	4.93844

The REG Procedure
Model: MODEL1
Dependent Variable: citympg



The REG Procedure
Model: MODEL1
Dependent Variable: citympg



c. All VIFs are less than 5, so there is no predictor with high VIF. I keep weight, horsepower and enginesize in my model.

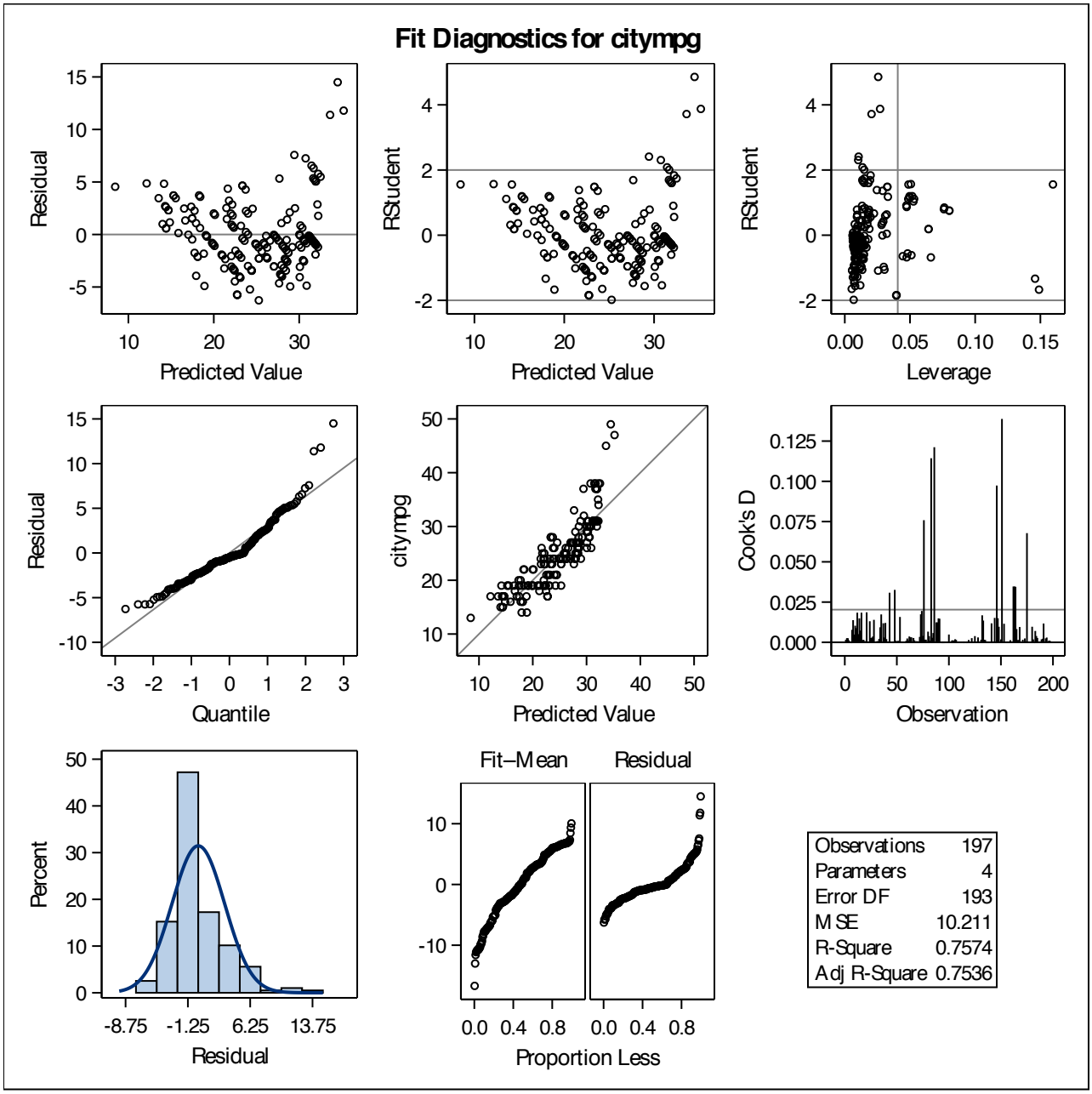
The REG Procedure
Model: MODEL1
Dependent Variable: citympg

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	6152.63902	2050.87967	200.84	<.0001
Error	193	1970.79246	10.21136		
Corrected Total	196	8123.43147			

Root MSE	3.19552	**R-Square**	0.7574
Dependent Mean	25.15228	**Adj R-Sq**	0.7536
Coeff Var	12.70470		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	47.41053	1.24649	38.04	<.0001
weight	1	-0.00674	0.00084456	-7.99	<.0001
horsepower	1	-0.13321	0.01096	-12.16	<.0001
enginesize	1	0.06929	0.01210	5.73	<.0001

The REG Procedure
Model: MODEL1
Dependent Variable: citympg



The UNIVARIATE Procedure
Variable: resid (Residual)

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.932588	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.150348	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.614709	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	3.147646	**Pr > A-Sq**	<0.0050

c. All Cook's distances are less than 1, so there is no observation with high Cook's distance. The model is highly significant with p-value less than 0.0001. 75.74% of variance of miles per gallon in the city is explained by the model. The p-value of parameters are all less than 0.0001. Increase of one unit in weight can cause miles per gallon in the city decrease by 0.00674. Increase of one unit in horsepower causes miles per gallon in the city decrease by 0.13321. Increase of one unit in enginesize causes miles per gallon in the city increase by 0.06929. According to the result of tests for normality, the the normality assumption for the residuals is valid.

Exercise 4

Class Level Information		
Class	**Levels**	**Values**
weight_cat	2	Heavy Light
ndoors	2	four two

Number of Observations Read	197
Number of Observations Used	197

The GLM Procedure

Dependent Variable: highwaympg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	7101.366830	1183.561138	109.24	<.0001
Error	190	2058.582409	10.834644		
Corrected Total	196	9159.949239			

R-Square	Coeff Var	Root MSE	highwaympg Mean
0.775263	10.74653	3.291602	30.62944

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weight_cat	1	4228.573122	4228.573122	390.28	<.0001
ndoors	1	0.000453	0.000453	0.00	0.9948
horsepower	1	2031.520896	2031.520896	187.50	<.0001
horsepowe*weight_cat	1	780.115352	780.115352	72.00	<.0001
horsepower*ndoors	1	36.028931	36.028931	3.33	0.0698
horsep*weight*ndoors	1	25.128075	25.128075	2.32	0.1294

Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight_cat	1	784.308437	784.308437	72.39	<.0001
ndoors	1	2.958595	2.958595	0.27	0.6019
horsepower	1	2826.658240	2826.658240	260.89	<.0001
horsepowe*weight_cat	1	687.362793	687.362793	63.44	<.0001
horsepower*ndoors	1	0.118218	0.118218	0.01	0.9169
horsep*weight*ndoors	1	25.128075	25.128075	2.32	0.1294

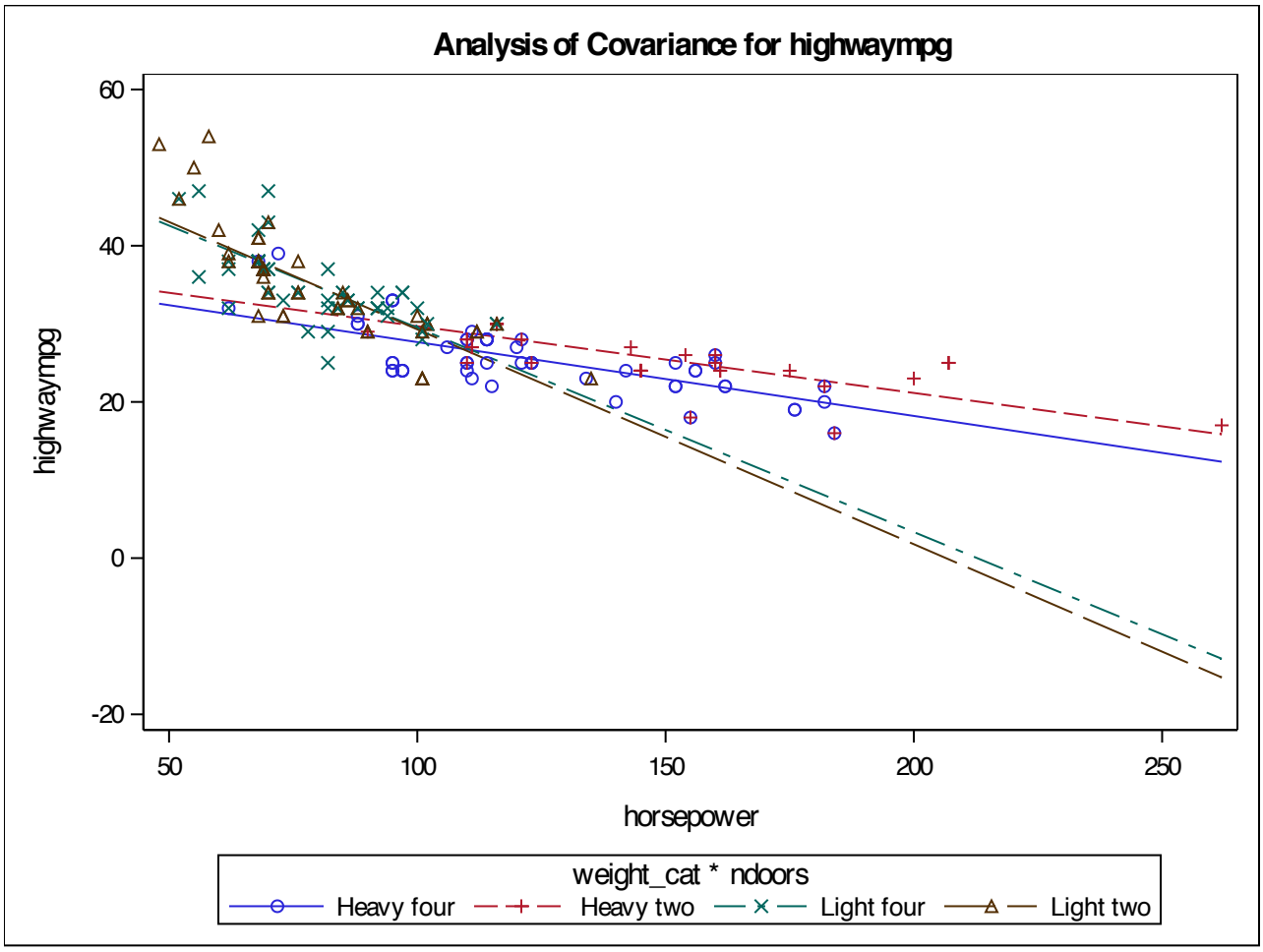
Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		56.79644514	B	1.82583912	31.11	<.0001
weight_cat	Heavy	-18.54713717	B	2.17992107	-8.51	<.0001
weight_cat	Light	0.00000000	B	.	.	.
ndoors	four	-1.13402262	B	2.17013257	-0.52	0.6019
ndoors	two	0.00000000	B	.	.	.
horsepower		-0.27508615	B	0.02248979	-12.23	<.0001
horsepowe*weight_cat	Heavy	0.18958379	B	0.02218536	8.55	<.0001

The GLM Procedure

Dependent Variable: highwaympg

Parameter	Estimate		Standard Error	t Value	Pr > |t|
horsepowe*weight_cat Light	0.00000000	B	.	.	.
horsepower*ndoors four	0.01346431	B	0.02737169	0.49	0.6234
horsepower*ndoors two	0.00000000	B	.	.	.
horsep*weight*ndoors Heavy four	-0.02251532	B	0.01478447	-1.52	0.1294
horsep*weight*ndoors Heavy two	0.00000000	B	.	.	.
horsep*weight*ndoors Light four	0.00000000	B	.	.	.
horsep*weight*ndoors Light two	0.00000000	B	.	.	.

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.



According to the plot of analysis of covariance for highwaympg, the equal slope assumption is not valid. Based on the result of type III SS, the p-value of interaction horsepower*weight*ndoors is 0.1294, which is insignificant, so we remove it first.

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
weight_cat	2	Heavy Light
ndoors	2	four two

Number of Observations Read	197
Number of Observations Used	197

The GLM Procedure

Dependent Variable: highwaympg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	7076.238754	1415.247751	129.73	<.0001
Error	191	2083.710484	10.909479		
Corrected Total	196	9159.949239			

R-Square	Coeff Var	Root MSE	highwaympg Mean
0.772519	10.78358	3.302950	30.62944

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weight_cat	1	4228.573122	4228.573122	387.61	<.0001
ndoors	1	0.000453	0.000453	0.00	0.9949
horsepower	1	2031.520896	2031.520896	186.22	<.0001
horsepowe*weight_cat	1	780.115352	780.115352	71.51	<.0001
horsepower*ndoors	1	36.028931	36.028931	3.30	0.0707

Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight_cat	1	918.696565	918.696565	84.21	<.0001
ndoors	1	10.608217	10.608217	0.97	0.3253
horsepower	1	2834.064893	2834.064893	259.78	<.0001
horsepowe*weight_cat	1	767.310411	767.310411	70.33	<.0001
horsepower*ndoors	1	36.028931	36.028931	3.30	0.0707

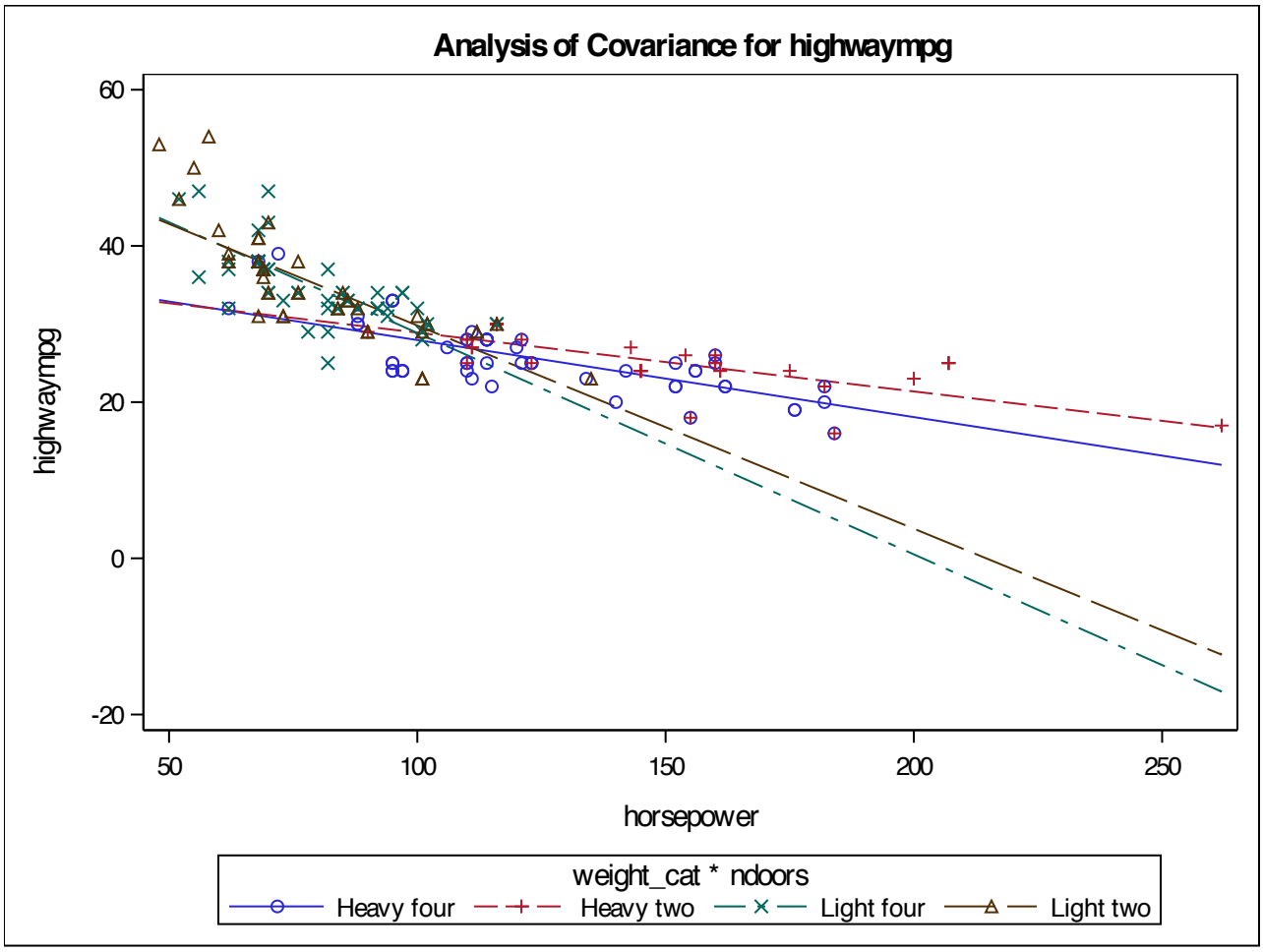
Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		55.82479321	B	1.71663195	32.52	<.0001
weight_cat	Heavy	-19.39985620	B	2.11404618	-9.18	<.0001
weight_cat	Light	0.00000000	B	.	.	.
ndoors	four	1.38785383	B	1.40742262	0.99	0.3253
ndoors	two	0.00000000	B	.	.	.
horsepower		-0.26017014	B	0.02031478	-12.81	<.0001
horsepowe*weight_cat	Heavy	0.18488867	B	0.02204585	8.39	<.0001
horsepowe*weight_cat	Light	0.00000000	B	.	.	.

The GLM Procedure

Dependent Variable: highwaympg

Parameter		Estimate		Standard Error	t Value	Pr > |t|
horsepower*ndoors	four	-0.02336918	B	0.01285937	-1.82	0.0707
horsepower*ndoors	two	0.00000000	B	.	.	.

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.



After first removal, the p-value of interaction horsepower*ndoors is 0.0707, which is insignificant, so we remove it. The plot of analysis of covariance for highwaympg still shows that the equal slope assumption is not valid.

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
weight_cat	2	Heavy Light
ndoors	2	four two

Number of Observations Read	197
Number of Observations Used	197

The GLM Procedure

Dependent Variable: highwaympg

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	7040.209823	1760.052456	159.42	<.0001
Error	192	2119.739415	11.040309		
Corrected Total	196	9159.949239			

R-Square	Coeff Var	Root MSE	highwaympg Mean
0.768586	10.84805	3.322696	30.62944

Source	DF	Type I SS	Mean Square	F Value	Pr > F
weight_cat	1	4228.573122	4228.573122	383.01	<.0001
ndoors	1	0.000453	0.000453	0.00	0.9949
horsepower	1	2031.520896	2031.520896	184.01	<.0001
horsepowe*weight_cat	1	780.115352	780.115352	70.66	<.0001

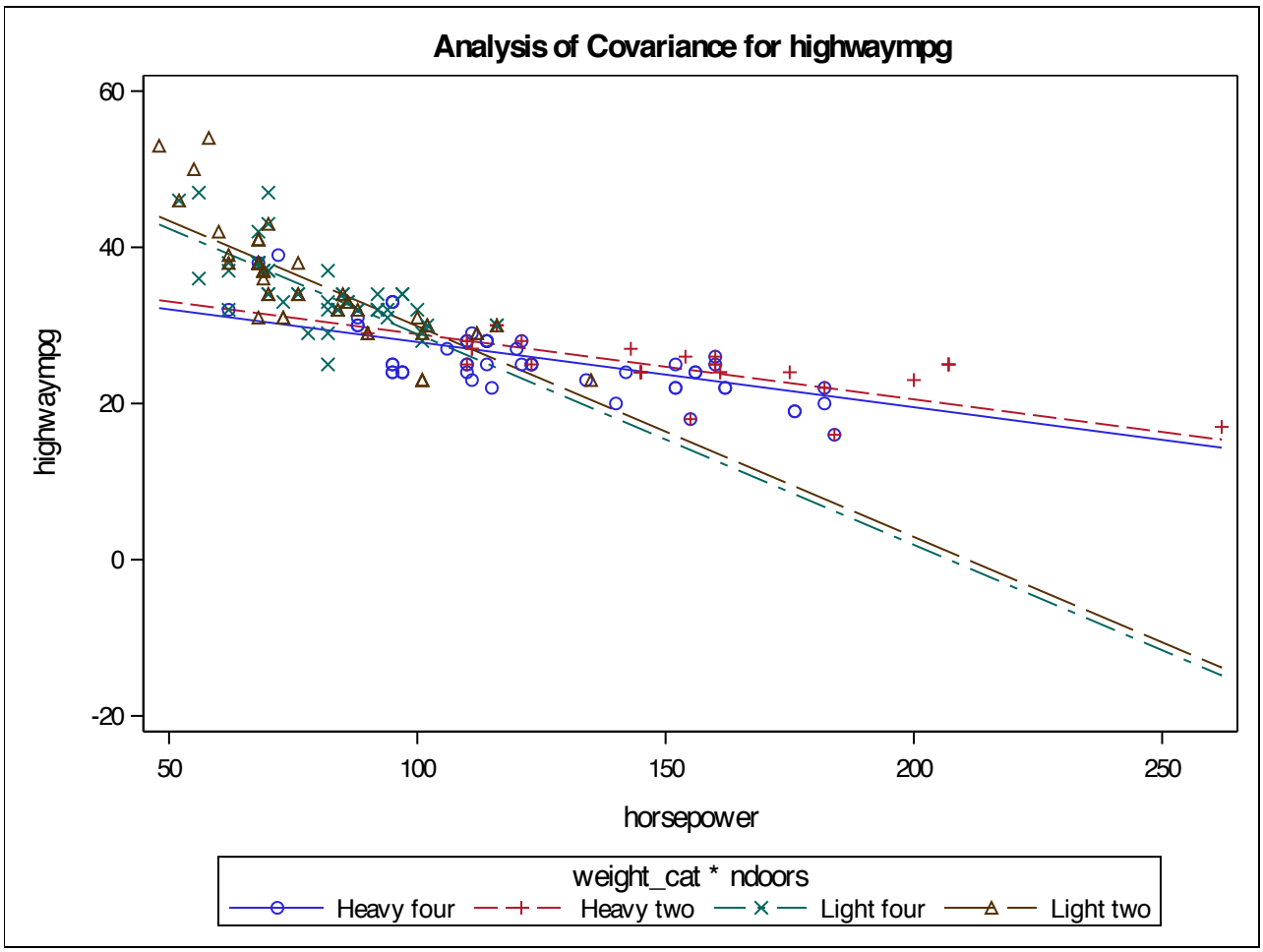
Source	DF	Type III SS	Mean Square	F Value	Pr > F
weight_cat	1	945.873596	945.873596	85.67	<.0001
ndoors	1	46.129747	46.129747	4.18	0.0423
horsepower	1	2799.409356	2799.409356	253.56	<.0001
horsepowe*weight_cat	1	780.115352	780.115352	70.66	<.0001

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		56.88248838	B	1.62460971	35.01	<.0001
weight_cat	Heavy	-19.64469467	B	2.12236148	-9.26	<.0001
weight_cat	Light	0.00000000	B	.	.	.
ndoors	four	-1.00927040	B	0.49375059	-2.04	0.0423
ndoors	two	0.00000000	B	.	.	.
horsepower		-0.26984232	B	0.01972239	-13.68	<.0001
horsepowe*weight_cat	Heavy	0.18630798	B	0.02216372	8.41	<.0001
horsepowe*weight_cat	Light	0.00000000	B	.	.	.

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

The GLM Procedure

Dependent Variable: highwaympg



After second removal, only the interaction horsepower*weight_cat left. According to the plot of analysis of covariance for highwaympg, the equal slope assumption is not valid. Therefore, we cannot run the model without the interaction effect. The p-value of interaction horsepower*weight_cat is 0.0423, so we keep it.

The UNIVARIATE Procedure
Variable: resid

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.974294	**Pr < W**	0.0011
Kolmogorov-Smirnov	**D**	0.084256	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.27019	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	1.527326	**Pr > A-Sq**	<0.0050

According to the result of tests for normality, the the normality assumption for the residuals is valid.