Exercise 1a

The LOGISTIC Procedure

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1271.921	894.864
SC	1277.298	959.390
-2 Log L	1269.921	870.864

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	399.0564	11	<.0001
Score	386.3114	11	<.0001
Wald	239.8724	11	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	242.8	108.1	5.0475	0.0247
fixed_acidity	1	0.2750	0.1253	4.8168	0.0282
volatile_acidity	1	-2.5810	0.7843	10.8298	0.0010
citric_acid	1	0.5678	0.8385	0.4585	0.4983
residual_sugar	1	0.2395	0.0737	10.5474	0.0012
chlorides	1	-8.8163	3.3649	6.8648	0.0088
free_sulfur_dioxide	1	0.0108	0.0122	0.7822	0.3765
total_sulfur_dioxide	1	-0.0165	0.00489	11.4093	0.0007
density	1	-257.8	110.4	5.4528	0.0195
pH	1	0.2242	0.9984	0.0504	0.8223
sulphates	1	3.7499	0.5416	47.9397	<.0001
alcohol	1	0.7533	0.1316	32.7644	<.0001

The LOGISTIC Procedure

Odds Ratio Estimates		
Effect	**Point Estimate**	**95% Wald Confidence Limits**
fixed_acidity	1.316	1.030	1.683
volatile_acidity	0.076	0.016	0.352
citric_acid	1.764	0.341	9.127
residual_sugar	1.271	1.100	1.468
chlorides	<0.001	<0.001	0.108
free_sulfur_dioxide	1.011	0.987	1.035
total_sulfur_dioxide	0.984	0.974	0.993
density	<0.001	<0.001	<0.001
pH	1.251	0.177	8.855
sulphates	42.516	14.708	122.900
alcohol	2.124	1.641	2.749

All three global tests are highly significant, p-value < 0.05. At least one of the variables seem important & better than only intercept model because of lower AIC and SC values.Significant predictors (p < 0.05):Alcohol: beta=0.7533, p<0.0001, OR=2.12 (CI=1.64-2.75).Sulphates: beta=3.7499, p<0.0001, OR=42.5 (CI=14.7-122.9).Volatile acidity: beta=−2.5810, p=.0010, OR=0.076 (CI=0.016-0.352).Residual sugar: beta=0.2395, p=.0012, OR=1.27 (1.10-1.47).Total sulfur dioxide: beta=−0.0165, p=.0007, OR=0.984 (0.974-0.993).Fixed acidity: beta=0.2750, p=.0282, OR=1.32 (1.03-1.68).Chlorides: beta=−8.8163, p=.0088, OR ≈ 0 (0.108).Density: beta=−257.8, p=.0195, OR ≈ 0.Not significant after adjustment (p > 0.05):Citric acid (p=0.4983), Free SO2 (p=0.3765), pH (p=0.8223).Based on the individual p-values, the Type 3 table will align closely:Retain: alcohol, sulphates, volatile_acidity, residual_sugar, total_sulfur_dioxide, fixed_acidity, chlorides, density (all have p<0.05 and meaningful effects).remove: citric_acid, free_sulfur_dioxide, pH.
Exercise 1b

The LOGISTIC Procedure

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

Number of Observations Read	1599
Number of Observations Used	1599

Response Profile		
Ordered Value	good_wine	Total Frequency
1	1	217
2	0	1382

Probability modeled is good_wine=1.

Stepwise Selection Procedure

Step 0. Intercept entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L	=	1269.921

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
386.3114	11	<.0001

Step 1. Effect alcohol entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1271.921	1031.888
SC	1277.298	1042.642
-2 Log L	1269.921	1027.888

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	242.0333	1	<.0001
Score	265.2827	1	<.0001
Wald	196.9235	1	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
159.1714	10	<.0001

Note: No effects for the model in Step 1 are removed.

Step 2. Effect sulphates entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1271.921	981.755
SC	1277.298	997.886
-2 Log L	1269.921	975.755

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	294.1658	2	<.0001
Score	307.2853	2	<.0001
Wald	212.6288	2	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
99.3970	9	<.0001

Note: No effects for the model in Step 2 are removed.

Step 3. Effect volatile_acidity entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

The LOGISTIC Procedure

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1271.921	925.263
SC	1277.298	946.771
-2 Log L	1269.921	917.263

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	352.6580	3	<.0001
Score	346.6843	3	<.0001
Wald	231.6963	3	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
43.3044	8	<.0001

Note: No effects for the model in Step 3 are removed.

Step 4. Effect total_sulfur_dioxide entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1271.921	909.553
SC	1277.298	936.439
-2 Log L	1269.921	899.553

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	370.3680	4	<.0001
Score	352.2612	4	<.0001
Wald	231.5612	4	<.0001

The LOGISTIC Procedure

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
27.0559	7	0.0003

Note: No effects for the model in Step 4 are removed.

Step 5. Effect chlorides entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1271.921	902.904
SC	1277.298	935.167
-2 Log L	1269.921	890.904

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	379.0164	5	<.0001
Score	358.2769	5	<.0001
Wald	233.2432	5	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
21.2222	6	0.0017

Note: No effects for the model in Step 5 are removed.

Step 6. Effect fixed_acidity entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1271.921	897.173
SC	1277.298	934.813
-2 Log L	1269.921	883.173

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	386.7474	6	<.0001
Score	367.6364	6	<.0001
Wald	235.2131	6	<.0001

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
13.6969	5	0.0177

Note: No effects for the model in Step 6 are removed.

Step 7. Effect residual_sugar entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	1271.921	894.989
SC	1277.298	938.006
-2 Log L	1269.921	878.989

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	390.9322	7	<.0001
Score	370.1410	7	<.0001
Wald	235.9719	7	<.0001

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
8.1325	4	0.0868

Note: No effects for the model in Step 7 are removed.

Step 8. Effect density entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

The LOGISTIC Procedure

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1271.921	890.076
SC	1277.298	938.470
-2 Log L	1269.921	872.076

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	397.8449	8	<.0001
Score	384.6127	8	<.0001
Wald	241.1927	8	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
1.2166	3	0.7490

Note: No effects for the model in Step 8 are removed.

Note: No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	alcohol		1	1	265.2827		<.0001
2	sulphates		1	2	68.5137		<.0001
3	volatile_acidity		1	3	52.2190		<.0001
4	total_sulfur_dioxide		1	4	15.3862		<.0001
5	chlorides		1	5	6.7950		0.0091
6	fixed_acidity		1	6	7.9181		0.0049
7	residual_sugar		1	7	4.8622		0.0275
8	density		1	8	6.9294		0.0085
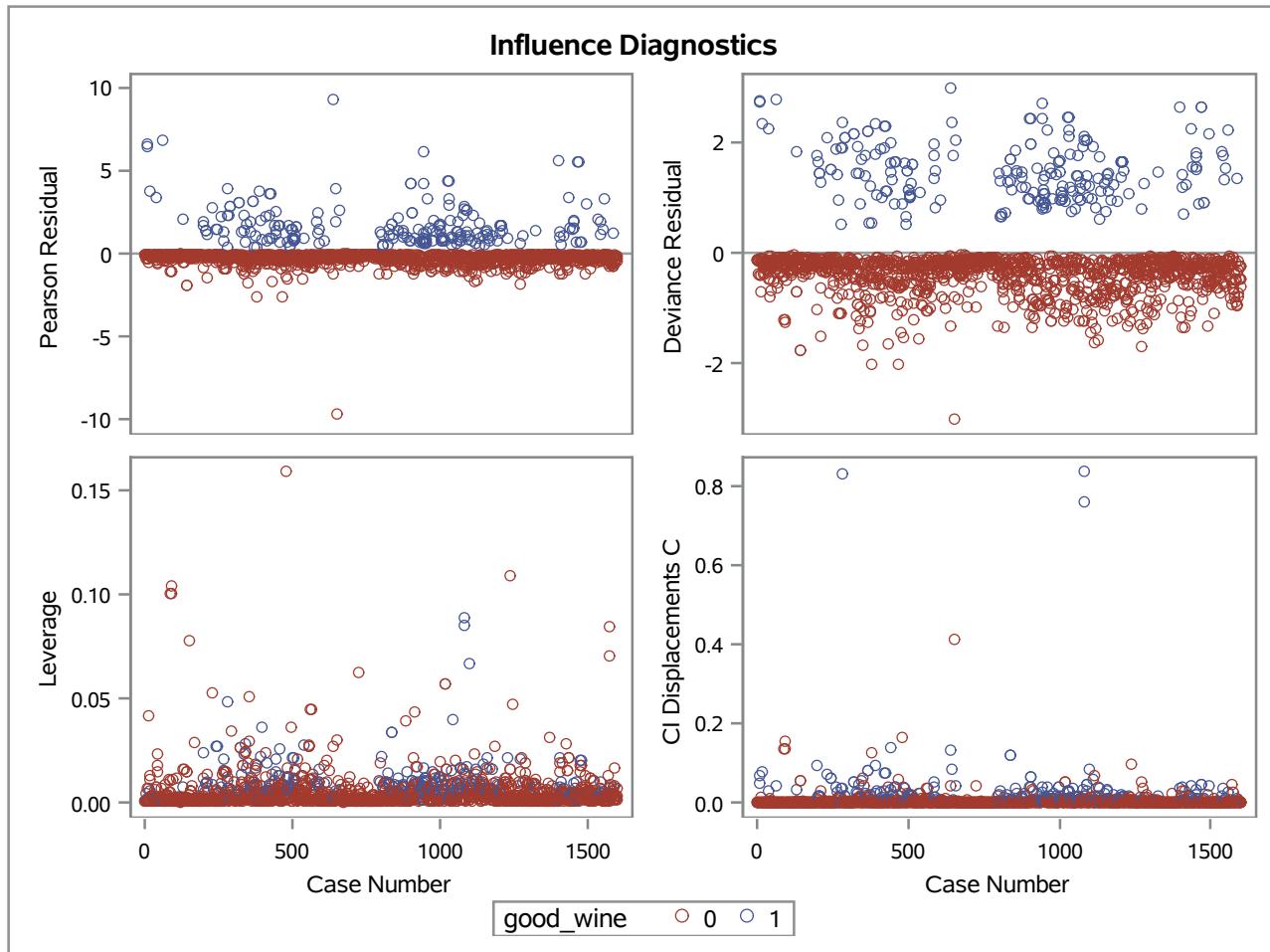
The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	226.8	91.6292	6.1239	0.0133
fixed_acidity	1	0.2812	0.0803	12.2635	0.0005
volatile_acidity	1	-2.9128	0.6467	20.2872	<.0001
residual_sugar	1	0.2328	0.0701	11.0364	0.0009
chlorides	1	-8.4408	3.2589	6.7084	0.0096
total_sulfur_dioxide	1	-0.0136	0.00345	15.5702	<.0001
density	1	-240.9	92.0220	6.8558	0.0088
sulphates	1	3.6987	0.5287	48.9506	<.0001
alcohol	1	0.7823	0.1120	48.7661	<.0001
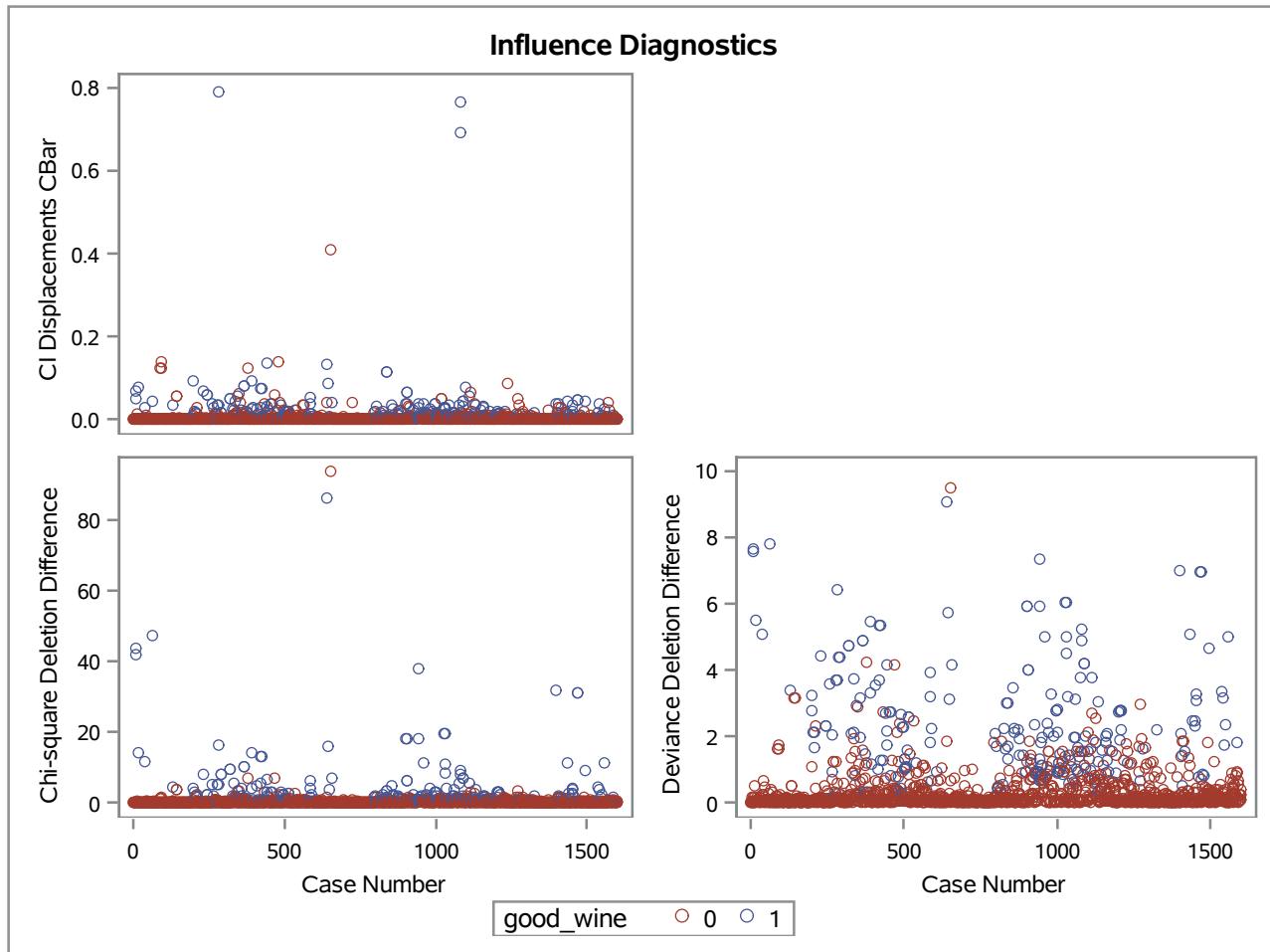
Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
fixed_acidity	1.325	1.132	1.550
volatile_acidity	0.054	0.015	0.193
residual_sugar	1.262	1.100	1.448
chlorides	<0.001	<0.001	0.128
total_sulfur_dioxide	0.986	0.980	0.993
density	<0.001	<0.001	<0.001
sulphates	40.396	14.333	113.850
alcohol	2.186	1.755	2.723

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	88.2	Somers' D	0.764
Percent Discordant	11.8	Gamma	0.764
Percent Tied	0.0	Tau-a	0.179
Pairs	299894	c	0.882

The LOGISTIC Procedure

Influence Diagnostics

The LOGISTIC Procedure

Influence Diagnostics

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	226.8	91.6292	6.1239	0.0133
fixed_acidity	1	0.2812	0.0803	12.2635	0.0005
volatile_acidity	1	-2.9128	0.6467	20.2872	<.0001
residual_sugar	1	0.2328	0.0701	11.0364	0.0009
chlorides	1	-8.4408	3.2589	6.7084	0.0096
total_sulfur_dioxide	1	-0.0136	0.00345	15.5702	<.0001
density	1	-240.9	92.0220	6.8558	0.0088
sulphates	1	3.6987	0.5287	48.9506	<.0001
alcohol	1	0.7823	0.1120	48.7661	<.0001

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
7.1585	8	0.5196

Since the p value of the lack of fit test is > 0.05, we reject the null the null hypothesis and there is no need to refine the model further.
Exercise 1c
The global test indicates the model is highly significant (p < 0.0001), meaning at least one wine characteristic helps predict good_wine status. Significant predictors include fixed_acidity, volatile_acidity, residual_sugar, chlorides, total_sulfur_dioxide, density, sulphates, and alcohol (p < 0.05). Citric_acid, free_sulfur_dioxide, and pH are not significant and could be removed from the model.Odds ratios show that higher alcohol and sulphates greatly increase the odds of a wine being rated good, while higher volatile acidity, chlorides, density, and total sulfur dioxide decrease the odds. Fixed acidity and residual sugar also show small positive effects on good_wine status.
Exercise 2

The LOGISTIC Procedure

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	2210.973	1679.625
SC	2216.350	1744.151
-2 Log L	2208.973	1655.625

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	553.3480	11	<.0001
Score	474.0303	11	<.0001
Wald	355.9129	11	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-42.9497	79.4760	0.2920	0.5889
fixed_acidity	1	-0.1360	0.0985	1.9063	0.1674
volatile_acidity	1	3.2817	0.4882	45.1802	<.0001
citric_acid	1	1.2743	0.5627	5.1280	0.0235
residual_sugar	1	-0.0553	0.0538	1.0587	0.3035
chlorides	1	3.9157	1.5693	6.2257	0.0126
free_sulfur_dioxide	1	-0.0222	0.00824	7.2778	0.0070
total_sulfur_dioxide	1	0.0164	0.00288	32.3489	<.0001
density	1	50.9322	81.1508	0.3939	0.5303
pH	1	0.3806	0.7202	0.2793	0.5972
sulphates	1	-2.7951	0.4522	38.2066	<.0001
alcohol	1	-0.8668	0.1042	69.2116	<.0001

The LOGISTIC Procedure

Odds Ratio Estimates		
Effect	**Point Estimate**	**95% Wald Confidence Limits**
fixed_acidity	0.873	0.720	1.059
volatile_acidity	26.621	10.224	69.311
citric_acid	3.576	1.187	10.776
residual_sugar	0.946	0.852	1.051
chlorides	50.185	2.316	>999.999
free_sulfur_dioxide	0.978	0.962	0.994
total_sulfur_dioxide	1.017	1.011	1.022
density	>999.999	<0.001	>999.999
pH	1.463	0.357	6.003
sulphates	0.061	0.025	0.148
alcohol	0.420	0.343	0.516

For the below_avg_wine model, several parameters are significant predictors of low wine quality. Higher volatile acidity, chlorides, citric acid, and total sulfur dioxide all increase the odds of a wine being below average, while higher alcohol, sulphates, and free sulfur dioxide decrease the odds of poor quality. Comparing this to the good_wine model, higher alcohol content and sulphates are consistently associated with better wine quality, whereas higher volatile acidity, chlorides, and total sulfur dioxide are markers of lower-quality wines.

The LOGISTIC Procedure

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

Number of Observations Read	1599
Number of Observations Used	1599

Response Profile		
Ordered Value	below_avg_wine	Total Frequency
1	1	744
2	0	855

Probability modeled is below_avg_wine=1.

Stepwise Selection Procedure

Step 0. Intercept entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L	=	2208.973

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
474.0303	11	<.0001

Step 1. Effect alcohol entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2210.973	1869.034
SC	2216.350	1879.788
-2 Log L	2208.973	1865.034

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	343.9392	1	<.0001
Score	302.2248	1	<.0001
Wald	250.9727	1	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
198.9236	10	<.0001

Note: No effects for the model in Step 1 are removed.

Step 2. Effect volatile_acidity entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2210.973	1758.503
SC	2216.350	1774.634
-2 Log L	2208.973	1752.503

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	456.4705	2	<.0001
Score	393.1223	2	<.0001
Wald	306.8877	2	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
95.5859	9	<.0001

Note: No effects for the model in Step 2 are removed.

Step 3. Effect total_sulfur_dioxide entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

The LOGISTIC Procedure

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2210.973	1726.371
SC	2216.350	1747.879
-2 Log L	2208.973	1718.371

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	490.6025	3	<.0001
Score	423.1711	3	<.0001
Wald	325.3141	3	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
63.0878	8	<.0001

Note: No effects for the model in Step 3 are removed.

Step 4. Effect sulphates entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2210.973	1694.204
SC	2216.350	1721.090
-2 Log L	2208.973	1684.204

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	524.7691	4	<.0001
Score	451.6755	4	<.0001
Wald	343.2892	4	<.0001

The LOGISTIC Procedure

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
28.4755	7	0.0002

Note: No effects for the model in Step 4 are removed.

Step 5. Effect chlorides entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2210.973	1685.719
SC	2216.350	1717.982
-2 Log L	2208.973	1673.719

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	535.2541	5	<.0001
Score	460.0902	5	<.0001
Wald	347.6940	5	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
18.2629	6	0.0056

Note: No effects for the model in Step 5 are removed.

Step 6. Effect free_sulfur_dioxide entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	2210.973	1678.716
SC	2216.350	1716.356
-2 Log L	2208.973	1664.716

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	544.2574	6	<.0001
Score	468.2190	6	<.0001
Wald	352.9280	6	<.0001

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
9.0385	5	0.1075

Note: No effects for the model in Step 6 are removed.

**Note:** No (additional) effects met the 0.05 significance level for entry into the model.
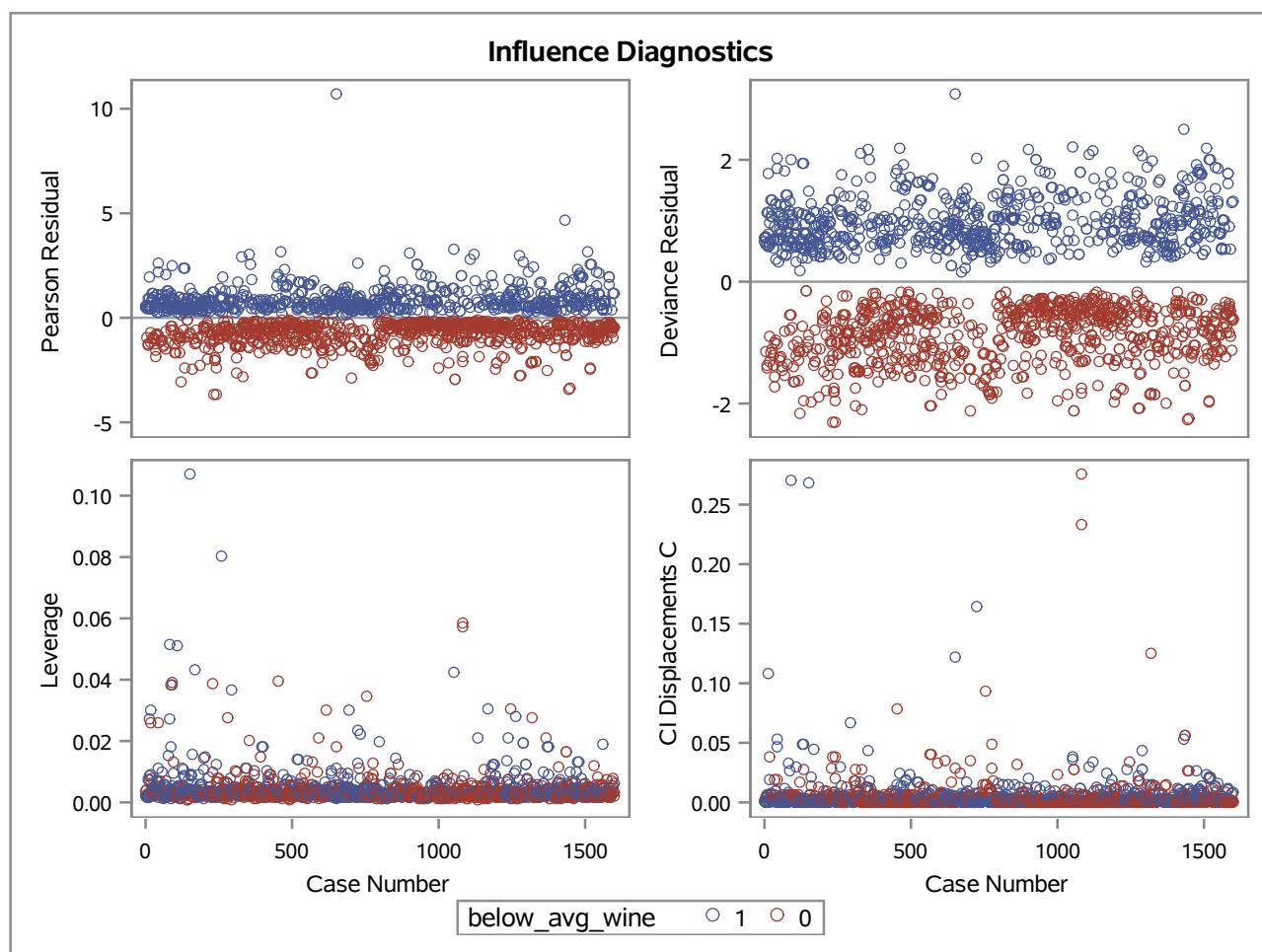
Summary of Stepwise Selection							
	Effect						
Step	**Entered**	**Removed**	**DF**	**Number In**	**Score Chi-Square**	**Wald Chi-Square**	**Pr > ChiSq**
1	alcohol		1	1	302.2248		<.0001
2	volatile_acidity		1	2	110.0827		<.0001
3	total_sulfur_dioxide		1	3	34.3093		<.0001
4	sulphates		1	4	36.1686		<.0001
5	chlorides		1	5	10.5493		0.0012
6	free_sulfur_dioxide		1	6	8.9934		0.0027

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	8.1498	0.8090	101.4768	<.0001
volatile_acidity	1	2.8959	0.3707	61.0342	<.0001
chlorides	1	4.4210	1.4321	9.5305	0.0020
free_sulfur_dioxide	1	-0.0239	0.00799	8.9262	0.0028
total_sulfur_dioxide	1	0.0175	0.00271	41.8292	<.0001
sulphates	1	-2.7059	0.4276	40.0533	<.0001
alcohol	1	-0.8594	0.0707	147.6339	<.0001

The LOGISTIC Procedure

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
volatile_acidity	18.100	8.753	37.430
chlorides	83.181	5.024	>999.999
free_sulfur_dioxide	0.976	0.961	0.992
total_sulfur_dioxide	1.018	1.012	1.023
sulphates	0.067	0.029	0.154
alcohol	0.423	0.369	0.486

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.9	Somers' D	0.638
Percent Discordant	18.1	Gamma	0.638
Percent Tied	0.0	Tau-a	0.317
Pairs	636120	c	0.819

Influence Diagnostics

below_avg_wine ○ 1 ○ 0

The LOGISTIC Procedure

Influence Diagnostics

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	8.1498	0.8090	101.4768	<.0001
volatile_acidity	1	2.8959	0.3707	61.0342	<.0001
chlorides	1	4.4210	1.4321	9.5305	0.0020
free_sulfur_dioxide	1	-0.0239	0.00799	8.9262	0.0028
total_sulfur_dioxide	1	0.0175	0.00271	41.8292	<.0001
sulphates	1	-2.7059	0.4276	40.0533	<.0001
alcohol	1	-0.8594	0.0707	147.6339	<.0001

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
11.6270	8	0.1686

After removing influential points and letting stepwise pick the predictors, the final below-average quality model kept six variables: volatile acidity, chlorides, free sulfur dioxide, total sulfur dioxide, sulphates, and alcohol. All are statistically significant($p < 0.003$). Wines are more likely to be below average when volatile acidity, chlorides, or total sulfur dioxide are higher; they are less likely to be below average when alcohol, sulphates, or free sulfur dioxide are higher. The Hosmer-Lemeshow test ($p = 0.169$) shows no evidence of lack of fit.

Good wines (rating > 7) are associated with more alcohol and more sulphates, and with less volatile acidity, chlorides, and total sulfur dioxide. In contrast, below-average wines (rating < 5) are much more likely when volatile acidity, chlorides, and total SO2 are higher, and less likely when alcohol, sulphates, and free SO2 are higher.

Exercise 3a

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Poisson
Link Function	Log
Dependent Variable	rings

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	4170	2470.3512	0.5924
Scaled Deviance	4170	2470.3512	0.5924
Pearson Chi-Square	4170	2688.7344	0.6448
Scaled Pearson X2	4170	2688.7344	0.6448
Log Likelihood		54606.5583	
Full Log Likelihood		-9802.5113	
AIC (smaller is better)		19619.0227	
AICC (smaller is better)		19619.0495	
BIC (smaller is better)		19663.3841	

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	1.5503	0.0452	1.4618	1.6389	1176.70	<.0001
sex	F	1	0.0112	0.0114	-0.0113	0.0336	0.95	0.3293
sex	I	1	-0.1130	0.0144	-0.1412	-0.0848	61.71	<.0001
sex	M	0	0.0000	0.0000	0.0000	0.0000	.	.
length		1	-0.7087	0.2529	-1.2044	-0.2130	7.85	0.0051
diameter		1	2.4908	0.3077	1.8877	3.0938	65.53	<.0001
height		1	1.3270	0.1445	1.0438	1.6103	84.33	<.0001
whole_weight		1	-0.0873	0.0274	-0.1410	-0.0335	10.13	0.0015
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

LR Statistics For Type 1 Analysis				
Source	Deviance	DF	Chi-Square	Pr > ChiSq
Intercept	4139.3003			
sex	3261.7861	2	877.51	<.0001
length	2613.1519	1	648.63	<.0001
diameter	2536.7855	1	76.37	<.0001

The GENMOD Procedure

LR Statistics For Type 1 Analysis				
Source	**Deviance**	**DF**	**Chi-Square**	**Pr > ChiSq**
height	2480.5560	1	56.23	<.0001
whole_weight	2470.3512	1	10.20	0.0014

LR Statistics For Type 3 Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
sex	2	77.28	<.0001
length	1	7.82	0.0052
diameter	1	64.93	<.0001
height	1	63.31	<.0001
whole_weight	1	10.20	0.0014

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Poisson
Link Function	Log
Dependent Variable	rings

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	4170	2470.3512	0.5924
Scaled Deviance	4170	4170.0000	1.0000
Pearson Chi-Square	4170	2688.7344	0.6448
Scaled Pearson X2	4170	4538.6350	1.0884
Log Likelihood		92176.9137	
Full Log Likelihood		-9802.5113	
AIC (smaller is better)		19619.0227	
AICC (smaller is better)		19619.0495	
BIC (smaller is better)		19663.3841	

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	1.5503	0.0348	1.4822	1.6185	1986.29	<.0001
sex	F	1	0.0112	0.0088	-0.0061	0.0284	1.61	0.2050
sex	I	1	-0.1130	0.0111	-0.1347	-0.0913	104.16	<.0001
sex	M	0	0.0000	0.0000	0.0000	0.0000	.	.
length		1	-0.7087	0.1947	-1.0902	-0.3272	13.26	0.0003
diameter		1	2.4908	0.2368	2.0266	2.9549	110.62	<.0001
height		1	1.3270	0.1112	1.1090	1.5450	142.35	<.0001
whole_weight		1	-0.0873	0.0211	-0.1287	-0.0459	17.10	<.0001
Scale		0	0.7697	0.0000	0.7697	0.7697		

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

LR Statistics For Type 1 Analysis							
Source	**Deviance**	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
Intercept	4139.3003						
sex	3261.7861	2	4170	740.63	<.0001	1481.26	<.0001
length	2613.1519	1	4170	1094.91	<.0001	1094.91	<.0001
diameter	2536.7855	1	4170	128.91	<.0001	128.91	<.0001

The GENMOD Procedure

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
height	2480.5560	1	4170	94.92	<.0001	94.92	<.0001
whole_weight	2470.3512	1	4170	17.23	<.0001	17.23	<.0001

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
sex	2	4170	65.22	<.0001	130.45	<.0001
length	1	4170	13.20	0.0003	13.20	0.0003
diameter	1	4170	109.61	<.0001	109.61	<.0001
height	1	4170	106.86	<.0001	106.86	<.0001
whole_weight	1	4170	17.23	<.0001	17.23	<.0001

After fitting rings ~ sex + length + diameter + height + whole_weight with scale=deviance, the significant predictors ($p < 0.05$) are sex=I (infant) (fewer rings than males), length (negative), diameter (positive), height (positive), and whole_weight (slightly negative). Sex=F (female) is not significant vs males ($p \approx 0.21$), meaning females and males have similar ring counts once size is controlled.

From the Poisson model, it looks like the number of rings (which indicates an abalone's age) depends mostly on its diameter, height, and length. In contrast, heavier or longer abalones have slightly fewer rings when size is already accounted for, suggesting weight or length alone doesn't always mean older age. Overall, shell thickness and height are the best indicators of an abalone's age. Females vs males show no meaningful difference in rings once body size is included.

Exercise 3b

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Poisson
Link Function	Log
Dependent Variable	rings

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	4170	2470.3512	0.5924
Scaled Deviance	4170	2470.3512	0.5924
Pearson Chi-Square	4170	2688.7344	0.6448
Scaled Pearson X2	4170	2688.7344	0.6448
Log Likelihood		54606.5583	
Full Log Likelihood		-9802.5113	
AIC (smaller is better)		19619.0227	
AICC (smaller is better)		19619.0495	
BIC (smaller is better)		19663.3841	

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	1.5503	0.0452	1.4618	1.6389	1176.70	<.0001
sex	F	1	0.0112	0.0114	-0.0113	0.0336	0.95	0.3293
sex	I	1	-0.1130	0.0144	-0.1412	-0.0848	61.71	<.0001
sex	M	0	0.0000	0.0000	0.0000	0.0000	.	.
length		1	-0.7087	0.2529	-1.2044	-0.2130	7.85	0.0051
diameter		1	2.4908	0.3077	1.8877	3.0938	65.53	<.0001
height		1	1.3270	0.1445	1.0438	1.6103	84.33	<.0001
whole_weight		1	-0.0873	0.0274	-0.1410	-0.0335	10.13	0.0015
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
sex	2	77.28	<.0001
length	1	7.82	0.0052
diameter	1	64.93	<.0001

The GENMOD Procedure

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
height	1	63.31	<.0001
whole_weight	1	10.20	0.0014

Variable sex has a p value > 0.05 so it statistically insignificant.

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Poisson
Link Function	Log
Dependent Variable	rings

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	4172	2547.6308	0.6106
Scaled Deviance	4172	2547.6308	0.6106
Pearson Chi-Square	4172	2783.3025	0.6671
Scaled Pearson X2	4172	2783.3025	0.6671
Log Likelihood		54567.9185	
Full Log Likelihood		-9841.1512	
AIC (smaller is better)		19692.3023	
AICC (smaller is better)		19692.3167	
BIC (smaller is better)		19723.9891	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4162	0.0423	1.3333	1.4992	1119.84	<.0001
length	1	-0.8190	0.2518	-1.3125	-0.3255	10.58	0.0011
diameter	1	2.8135	0.3047	2.2164	3.4106	85.28	<.0001
height	1	1.4284	0.1415	1.1511	1.7058	101.92	<.0001
whole_weight	1	-0.0699	0.0272	-0.1232	-0.0166	6.62	0.0101
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
length	1	10.52	0.0012
diameter	1	84.34	<.0001
height	1	74.75	<.0001
whole_weight	1	6.65	0.0099

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Poisson
Link Function	Log
Dependent Variable	rings

Number of Observations Read	4177
Number of Observations Used	4177

Class Level Information		
Class	Levels	Values
sex	3	F I M

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	4172	2547.6308	0.6106
Scaled Deviance	4172	4172.0000	1.0000
Pearson Chi-Square	4172	2783.3025	0.6671
Scaled Pearson X2	4172	4557.9359	1.0925
Log Likelihood		89360.4174	
Full Log Likelihood		-9841.1512	
AIC (smaller is better)		19692.3023	
AICC (smaller is better)		19692.3167	
BIC (smaller is better)		19723.9891	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4162	0.0331	1.3514	1.4811	1833.84	<.0001
length	1	-0.8190	0.1967	-1.2046	-0.4334	17.33	<.0001
diameter	1	2.8135	0.2381	2.3469	3.2801	139.66	<.0001
height	1	1.4284	0.1106	1.2117	1.6451	166.90	<.0001
whole_weight	1	-0.0699	0.0212	-0.1115	-0.0283	10.84	0.0010
Scale	0	0.7814	0.0000	0.7814	0.7814		

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

The GENMOD Procedure

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
length	1	4172	17.23	<.0001	17.23	<.0001
diameter	1	4172	138.12	<.0001	138.12	<.0001
height	1	4172	122.40	<.0001	122.40	<.0001
whole_weight	1	4172	10.90	0.0010	10.90	0.0010

Removing any influential points and refitting the model based on chosen predictores:

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE_CLEAN
Distribution	Poisson
Link Function	Log
Dependent Variable	rings

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	4171	2446.1905	0.5865
Scaled Deviance	4171	4171.0000	1.0000
Pearson Chi-Square	4171	2662.2903	0.6383
Scaled Pearson X2	4171	4539.4717	1.0883
Log Likelihood		93115.5281	
Full Log Likelihood		-9788.4619	
AIC (smaller is better)		19586.9239	
AICC (smaller is better)		19586.9383	
BIC (smaller is better)		19618.6094	
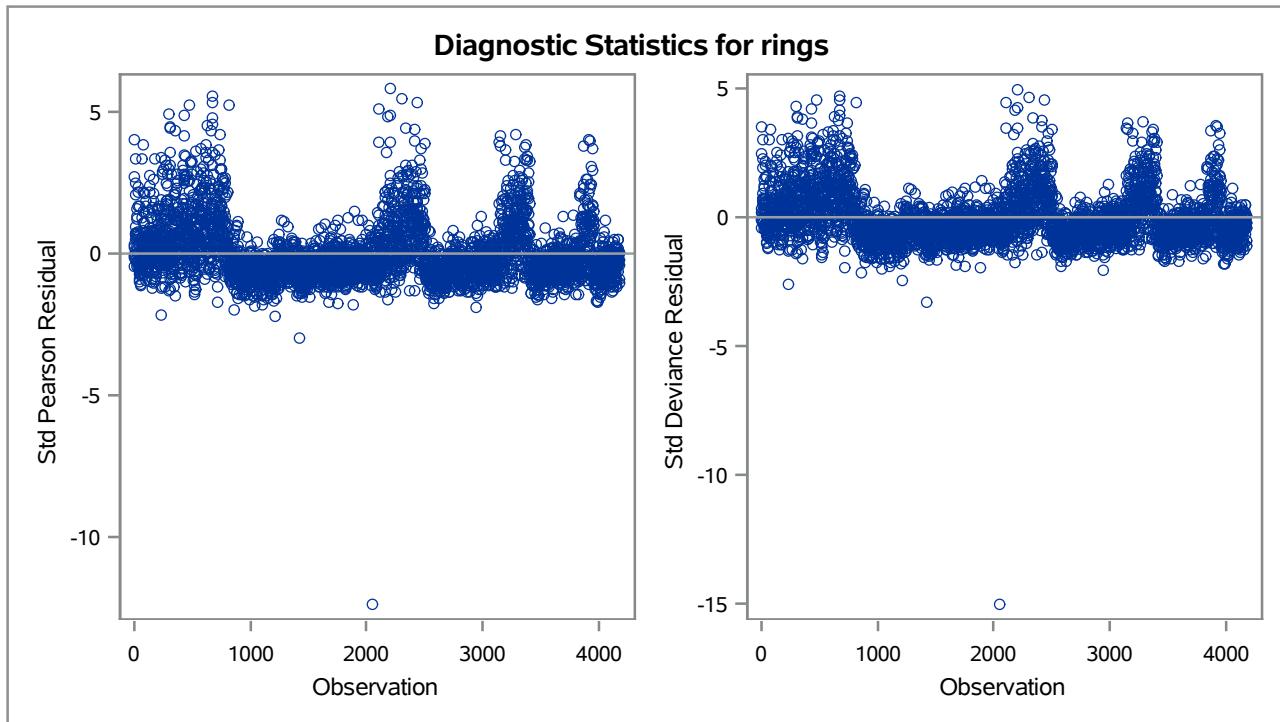
Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.3493	0.0330	1.2845	1.4140	1667.58	<.0001
length	1	-0.8157	0.1928	-1.1935	-0.4378	17.90	<.0001
diameter	1	2.3476	0.2358	1.8854	2.8097	99.12	<.0001
height	1	3.6783	0.2041	3.2783	4.0783	324.81	<.0001
whole_weight	1	-0.1411	0.0217	-0.1836	-0.0986	42.33	<.0001
Scale	0	0.7658	0.0000	0.7658	0.7658		

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

LR Statistics For Type 3 Analysis						
Source	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
length	1	4171	17.80	<.0001	17.80	<.0001
diameter	1	4171	98.27	<.0001	98.27	<.0001
height	1	4171	300.29	<.0001	300.29	<.0001
whole_weight	1	4171	42.84	<.0001	42.84	<.0001

Exercise 3c

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Poisson
Link Function	Log
Dependent Variable	rings

The GENMOD Procedure

Diagnostic Statistics for rings

The standardized Pearson and deviance residuals are centered around zero with no strong pattern or curvature, which suggests that the log-linear Poisson model fits the data well. A few extreme residuals are visible, but they represent isolated observations and do not indicate a major violation of assumptions.
Exercise 4

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Gamma
Link Function	Log
Dependent Variable	rings

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	4170	226.7494	0.0544
Scaled Deviance	4170	4214.4446	1.0107
Pearson Chi-Square	4170	254.7937	0.0611
Scaled Pearson X2	4170	4735.6845	1.1357
Log Likelihood		-9241.0652	
Full Log Likelihood		-9241.0652	
AIC (smaller is better)		18498.1303	
AICC (smaller is better)		18498.1649	
BIC (smaller is better)		18548.8291	

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	1.3862	0.0308	1.3258	1.4466	2023.75	<.0001
sex	F	1	0.0091	0.0088	-0.0081	0.0263	1.07	0.3012
sex	I	1	-0.1108	0.0100	-0.1303	-0.0912	123.17	<.0001
sex	M	0	0.0000	0.0000	0.0000	0.0000	.	.
length		1	-0.5009	0.1920	-0.8772	-0.1245	6.80	0.0091
diameter		1	2.2122	0.2371	1.7475	2.6770	87.05	<.0001
height		1	3.1230	0.2208	2.6901	3.5558	200.00	<.0001
whole_weight		1	-0.1885	0.0209	-0.2295	-0.1476	81.52	<.0001
Scale		1	18.5864	0.4031	17.8128	19.3935		

Note: The scale parameter was estimated by maximum likelihood.

LR Statistics For Type 3 Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
sex	2	149.87	<.0001
length	1	6.81	0.0091
diameter	1	86.53	<.0001

The GENMOD Procedure

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
height	1	216.15	<.0001
whole_weight	1	79.50	<.0001

Variable sex has a p value > 0.05 so it statistically insignificant.

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Gamma
Link Function	Log
Dependent Variable	rings

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	4172	234.9565	0.0563
Scaled Deviance	4172	4215.7866	1.0105
Pearson Chi-Square	4172	265.7952	0.0637
Scaled Pearson X2	4172	4769.1198	1.1431
Log Likelihood		-9315.9989	
Full Log Likelihood		-9315.9989	
AIC (smaller is better)		18643.9978	
AICC (smaller is better)		18644.0179	
BIC (smaller is better)		18682.0218	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.2602	0.0295	1.2024	1.3181	1822.79	<.0001
length	1	-0.6757	0.1953	-1.0584	-0.2929	11.97	0.0005
diameter	1	2.5330	0.2401	2.0624	3.0035	111.32	<.0001
height	1	3.4039	0.2242	2.9645	3.8432	230.58	<.0001
whole_weight	1	-0.1694	0.0213	-0.2111	-0.1277	63.44	<.0001
Scale	1	17.9428	0.3890	17.1963	18.7217		

Note: The scale parameter was estimated by maximum likelihood.

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
length	1	11.98	0.0005
diameter	1	110.34	<.0001
height	1	247.86	<.0001
whole_weight	1	62.11	<.0001

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Gamma
Link Function	Log
Dependent Variable	rings

Number of Observations Read	4177
Number of Observations Used	4177

Class Level Information		
Class	**Levels**	**Values**
sex	3	F I M

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	4172	234.9565	0.0563
Scaled Deviance	4172	4172.0000	1.0000
Pearson Chi-Square	4172	265.7952	0.0637
Scaled Pearson X2	4172	4719.5860	1.1313
Log Likelihood		-9316.1144	
Full Log Likelihood		-9316.1144	
AIC (smaller is better)		18642.2289	
AICC (smaller is better)		18642.2432	
BIC (smaller is better)		18673.9156	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.2602	0.0297	1.2021	1.3184	1803.86	<.0001
length	1	-0.6757	0.1963	-1.0604	-0.2909	11.85	0.0006
diameter	1	2.5330	0.2413	2.0600	3.0060	110.16	<.0001
height	1	3.4039	0.2253	2.9622	3.8455	228.19	<.0001
whole_weight	1	-0.1694	0.0214	-0.2114	-0.1275	62.78	<.0001
Scale	0	17.7565	0.0000	17.7565	17.7565		

Note: The Gamma scale parameter was estimated by DOF/DEVIANCE.

The GENMOD Procedure

Lagrange Multiplier Statistics		
Parameter	**Chi-Square**	**Pr > ChiSq**
Scale	0.2295	0.6319

LR Statistics For Type 3 Analysis						
Source	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
length	1	4172	11.87	0.0006	11.87	0.0006
diameter	1	4172	110.62	<.0001	110.62	<.0001
height	1	4172	252.57	<.0001	252.57	<.0001
whole_weight	1	4172	61.91	<.0001	61.91	<.0001

Removing any influential points and refitting the model based on chosen predictors:

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE_CLEAN
Distribution	Gamma
Link Function	Log
Dependent Variable	rings

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	4171	228.9700	0.0549
Scaled Deviance	4171	4171.0000	1.0000
Pearson Chi-Square	4171	262.4657	0.0629
Scaled Pearson X2	4171	4781.1694	1.1463
Log Likelihood		-9260.1660	
Full Log Likelihood		-9260.1660	
AIC (smaller is better)		18530.3319	
AICC (smaller is better)		18530.3463	
BIC (smaller is better)		18562.0175	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.2390	0.0292	1.1817	1.2963	1794.77	<.0001
length	1	-0.6918	0.1939	-1.0718	-0.3117	12.73	0.0004
diameter	1	2.3451	0.2386	1.8775	2.8128	96.61	<.0001
height	1	4.3141	0.2262	3.8707	4.7574	363.74	<.0001
whole_weight	1	-0.1940	0.0211	-0.2354	-0.1527	84.55	<.0001
Scale	0	18.2164	0.0000	18.2164	18.2164		
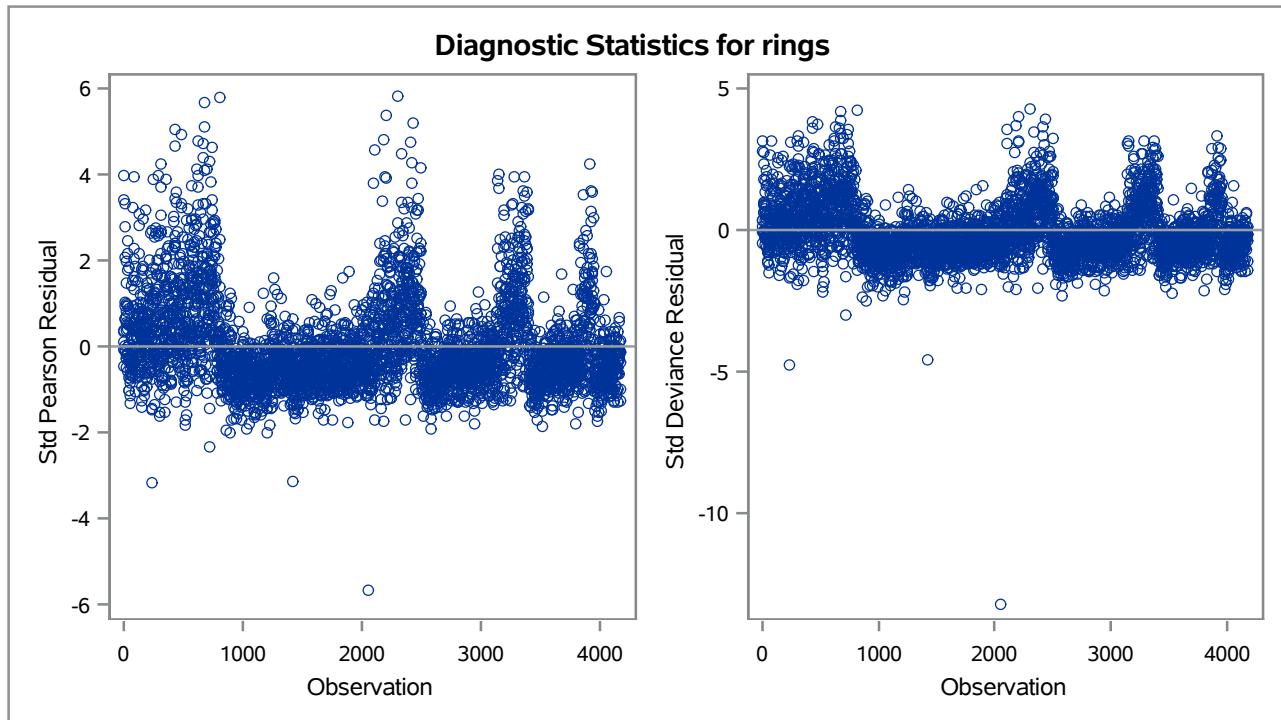
Note: The Gamma scale parameter was estimated by DOF/DEVIANCE.

LR Statistics For Type 3 Analysis						
Source	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
length	1	4171	12.75	0.0004	12.75	0.0004
diameter	1	4171	96.98	<.0001	96.98	<.0001
height	1	4171	368.03	<.0001	368.03	<.0001
whole_weight	1	4171	83.17	<.0001	83.17	<.0001

The GENMOD Procedure

Model Information	
Data Set	WORK.ABALONE
Distribution	Gamma
Link Function	Log
Dependent Variable	rings

The GENMOD Procedure

The Gamma model shows better fit than the Poisson: lower AIC (18,530 vs 19,587), and scaled deviance ≈ 1; residual plots show fewer extreme outliers and more homogeneous spread than the Poisson model. Overall, the Gamma log-linear model is the more reasonable choice here, with meaningful, significant predictors and better diagnostics.