Exercise 1a):

The output for logistic regression model predicting whether a female is a liver patient for females in the data set follows.

The LOGISTIC Procedure

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

Number of Observations Read	135
Number of Observations Used	133

Response Profile		
Ordered Value	LiverPatient	Total Frequency
1	Yes	86
2	No	47

Probability modeled is LiverPatient='Yes'.

Note: 2 observations were deleted due to missing values for the response or explanatory variables.

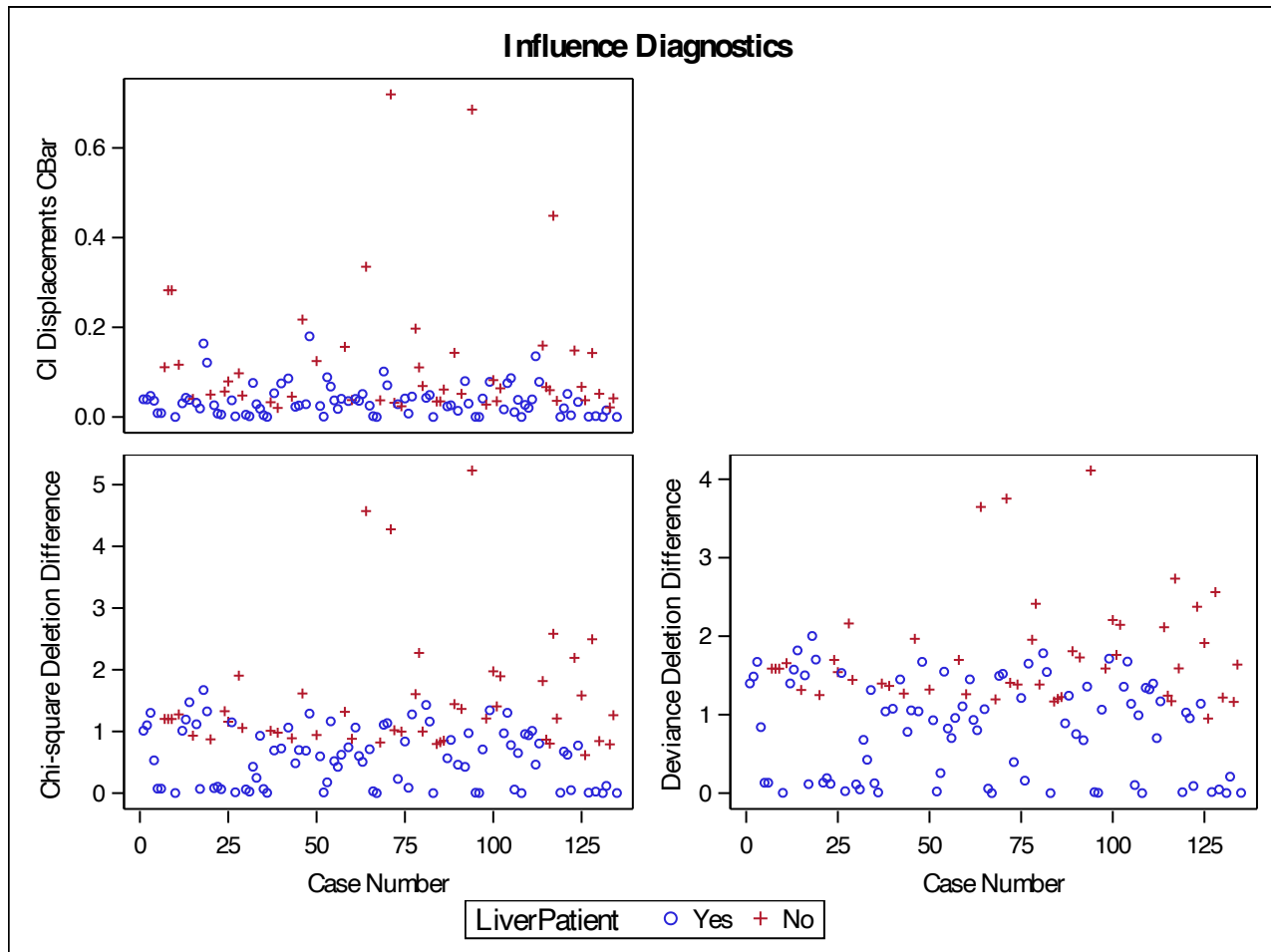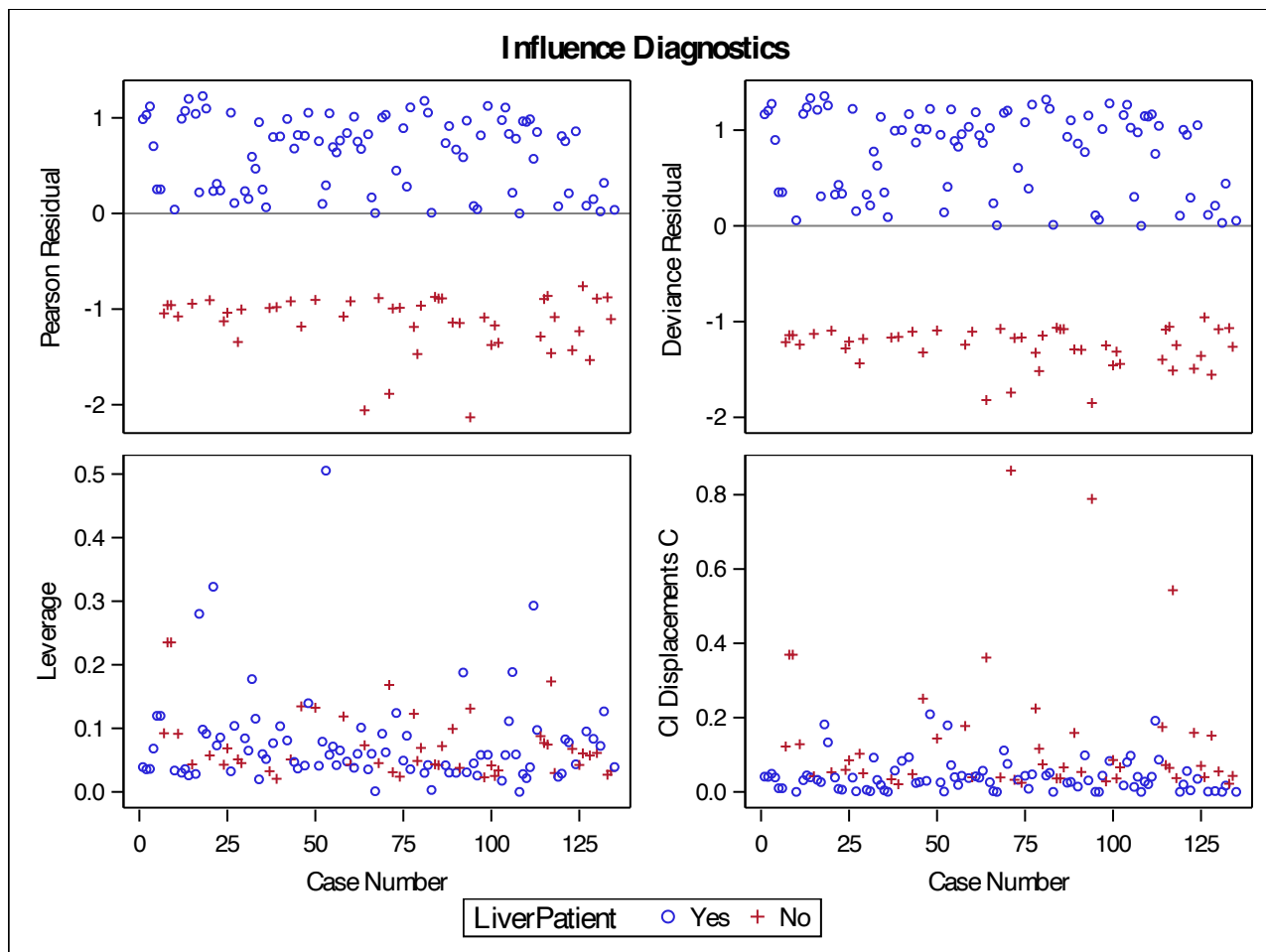| Model Convergence Status |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	165.308
SC	177.662	194.211
-2 Log L	172.771	145.308

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.4635	9	0.0012
Score	16.2424	9	0.0620
Wald	10.8317	9	0.2874

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.9427	3.2385	3.3674	0.0665
Age	1	0.00510	0.0135	0.1423	0.7061
TB	1	-0.6743	1.2841	0.2757	0.5995
DB	1	2.0098	2.3780	0.7143	0.3980
Alkphos	1	0.00159	0.00185	0.7354	0.3911
Alamine	1	0.00269	0.00982	0.0750	0.7841
Aspartate	1	0.00708	0.00880	0.6476	0.4210
TP	1	1.8478	0.9939	3.4559	0.0630
ALB	1	-3.3686	1.9835	2.8842	0.0895
AGRatio	1	4.4397	3.0060	2.1813	0.1397

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Age	1.005	0.979	1.032
TB	0.510	0.041	6.313
DB	7.462	0.071	788.928
Alkphos	1.002	0.998	1.005
Alamine	1.003	0.984	1.022
Aspartate	1.007	0.990	1.025
TP	6.346	0.905	44.516
ALB	0.034	<0.001	1.680
AGRatio	84.745	0.234	>999.999

Association of Predicted Probabilities and Observed Responses

Percent Concordant	72.1	Somers' D	0.441
Percent Discordant	27.9	Gamma	0.441
Percent Tied	0.0	Tau-a	0.203
Pairs	4042	c	0.721

Influence Diagnostics

Influence Diagnostics

Predicted Probabilities for LiverPatient=Yes with 95% Confidence Limits

At TB=2.232 DB=0.957 Alkphos=296.7 Alamine=50.68 Aspartate=63.71 TP=6.626 ALB=3.247 AGRatio=0.944

We first fit the model by using all of the predictors and will see if there are any influential points. By looking at the diagnostics results, we can see that there are no influential points, so now we will perform model selection.

The LOGISTIC Procedure

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

Number of Observations Read	135
Number of Observations Used	133

Response Profile		
Ordered Value	LiverPatient	Total Frequency
1	Yes	86
2	No	47

Probability modeled is LiverPatient='Yes'.

Note: 2 observations were deleted due to missing values for the response or explanatory variables.

Backward Elimination Procedure

Step   0. The following effects were entered:

Intercept   Age   TB   DB   Alkphos   Alamine   Aspartate   TP   ALB   AGRatio

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	165.308
SC	177.662	194.211
-2 Log L	172.771	145.308

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.4635	9	0.0012
Score	16.2424	9	0.0620
Wald	10.8317	9	0.2874

Step   1. Effect Alamine is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	163.384
SC	177.662	189.397
-2 Log L	172.771	145.384

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.3872	8	0.0006
Score	16.2122	8	0.0394
Wald	10.7249	8	0.2178

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
0.0753	1	0.7837

Step   2. Effect Age is removed:

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

The LOGISTIC Procedure

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	161.502
SC	177.662	184.625
-2 Log L	172.771	145.502

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.2690	7	0.0003
Score	16.1689	7	0.0236
Wald	10.6758	7	0.1534

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
0.1945	2	0.9073

Step   3. Effect TB is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	159.843
SC	177.662	180.076
-2 Log L	172.771	145.843

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	26.9278	6	0.0001
Score	15.4399	6	0.0171
Wald	10.3134	6	0.1121

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
0.5042	3	0.9180

Step 4. Effect Alkphos is removed:

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	158.883
SC	177.662	176.225
-2 Log L	172.771	146.883

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	25.8882	5	<.0001
Score	13.7504	5	0.0173
Wald	9.6890	5	0.0845

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
1.3077	4	0.8601

Step 5. Effect AGRatio is removed:

Model Convergence Status

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	159.511
SC	177.662	173.962
-2 Log L	172.771	149.511

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	23.2607	4	0.0001
Score	12.7100	4	0.0128
Wald	9.7616	4	0.0446

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
3.1875	5	0.6711

Step 6. Effect ALB is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	158.888
SC	177.662	170.450
-2 Log L	172.771	150.888

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.8831	3	<.0001
Score	10.6654	3	0.0137
Wald	8.4241	3	0.0380

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
4.4901	6	0.6107

Step   7. Effect TP is removed:

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	157.757
SC	177.662	166.428
-2 Log L	172.771	151.757

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.0140	2	<.0001
Score	10.4082	2	0.0055
Wald	7.7916	2	0.0203

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
5.2748	7	0.6265

Step   8. Effect DB is removed:

Model Convergence Status

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

The LOGISTIC Procedure

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	174.771	163.433
SC	177.662	169.214
-2 Log L	172.771	159.433

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	13.3381	1	0.0003
Score	7.1865	1	0.0073
Wald	5.8270	1	0.0158

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
9.6079	8	0.2936

Note: No (additional) effects met the 0.05 significance level for removal from the model.

Summary of Backward Elimination					
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	Alamine	1	8	0.0750	0.7841
2	Age	1	7	0.1179	0.7313
3	TB	1	6	0.2963	0.5862
4	Alkphos	1	5	0.7373	0.3905
5	AGRatio	1	4	1.9648	0.1610
6	ALB	1	3	1.3591	0.2437
7	TP	1	2	0.8626	0.3530
8	DB	1	1	2.8598	0.0908

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.1213	0.3061	0.1571	0.6918
Aspartate	1	0.0164	0.00679	5.8270	0.0158

The LOGISTIC Procedure

Odds Ratio Estimates		
Effect	**Point Estimate**	**95% Wald Confidence Limits**
Aspartate	1.017	1.003 1.030

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	60.4	**Somers' D**	0.223
Percent Discordant	38.1	**Gamma**	0.227
Percent Tied	1.6	**Tau-a**	0.103
Pairs	4042	**c**	0.612

Partition for the Hosmer and Lemeshow Test					
		LiverPatient = Yes		**LiverPatient = No**	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	11	5	5.72	6	5.28
2	12	8	6.40	4	5.60
3	14	10	7.70	4	6.30
4	12	7	6.76	5	5.24
5	13	6	7.51	7	5.49
6	13	9	7.73	4	5.27
7	15	7	9.37	8	5.63
8	13	7	8.79	6	4.21
9	13	11	10.08	2	2.92
10	17	16	15.94	1	1.06

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
6.9186	8	0.5454

The LOGISTIC Procedure

Influence Diagnostics

LiverPatient   ○ Yes   + No

Influence Diagnostics
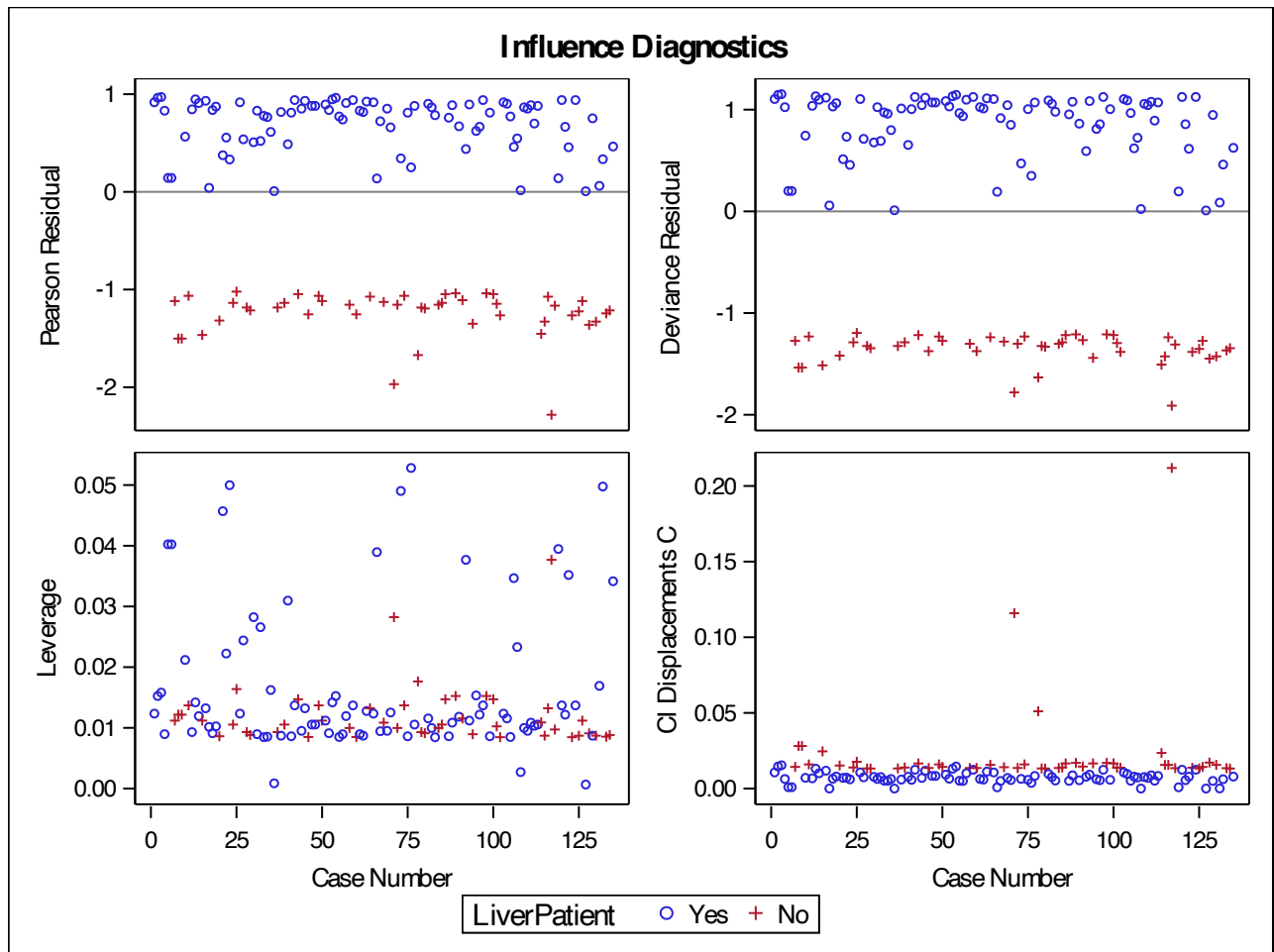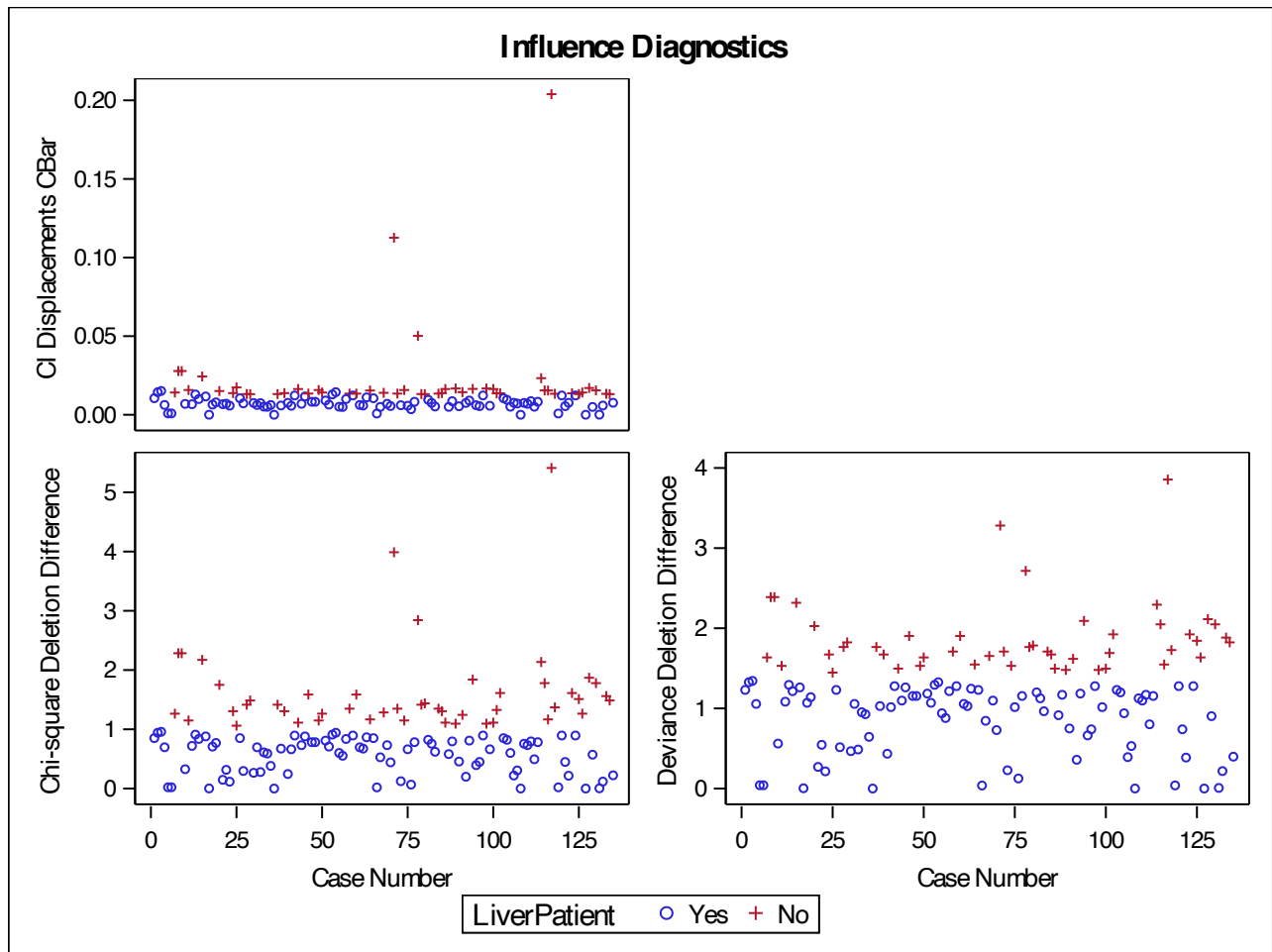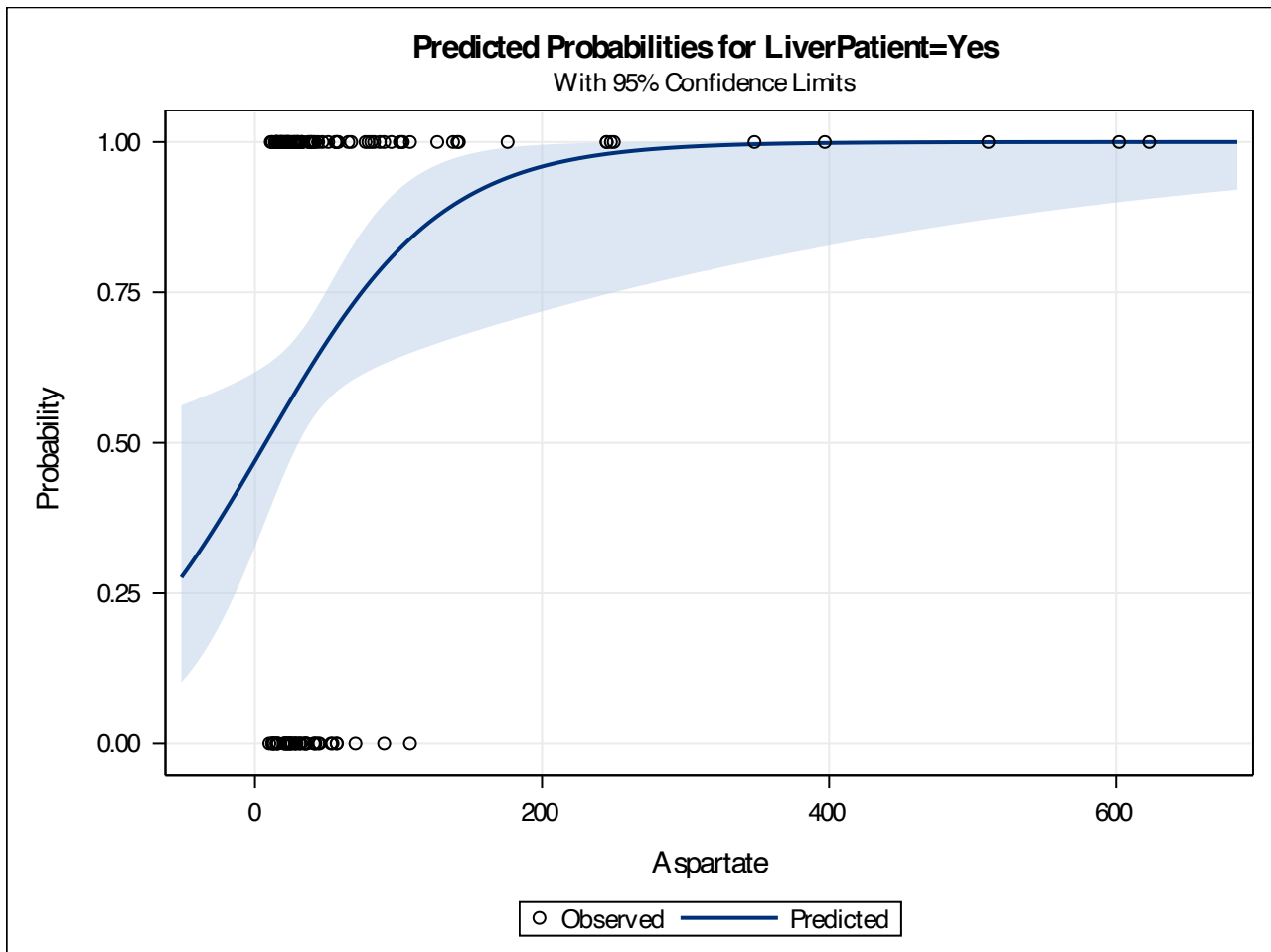
We have now our final model with only Aspartate predictor.

Exercise 1b):

Looking at the table Model Fit statistics, the Log Likelihood column values under Intercept and Covariates can be compared to the Intercept Only column values. We have small values for both AIC and SC(163.433 and 169.214), so we may conclude that model with predictors will be better. The global tests indicate there is at least one significant parameter estimate. Hosmer-Lemeshow'stest gives us p-value of 0.5454, which is larger than 0.05, hence we may conclude that we accept null hypothesis, and our model fits okay. There are no influential points based on cbar and diagnostics look fine.

Exercise 1c):

The odds ratio for Aspartate is 1.017, and it is significant. We performed backward selection and can see now that the confidence interval does not include 1. The odds of an adult female being a liver patient with liver disease increase by a factor of 1.017 with a one unit increase in Aspartate.

Exercise 2a):

The output for logistic regression model predicting whether a male is a liver patient for males in the data set follows.

The LOGISTIC Procedure

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

Number of Observations Read	423
Number of Observations Used	421

Response Profile		
Ordered Value	LiverPatient	Total Frequency
1	Yes	316
2	No	105

Probability modeled is LiverPatient='Yes'.

Note: 2 observations were deleted due to missing values for the response or explanatory variables.

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	474.936	406.917
SC	478.979	447.343
-2 Log L	472.936	386.917

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	86.0193	9	<.0001
Score	47.0936	9	<.0001
Wald	38.0177	9	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.5252	1.5712	2.5832	0.1080
Age	1	0.0182	0.00837	4.7514	0.0293
TB	1	0.0124	0.0835	0.0220	0.8821
DB	1	0.3607	0.2351	2.3544	0.1249
Alkphos	1	0.00124	0.00102	1.4966	0.2212
Alamine	1	0.0148	0.00677	4.7905	0.0286
Aspartate	1	0.00254	0.00370	0.4711	0.4925
TP	1	0.6511	0.4389	2.2010	0.1379
ALB	1	-1.2062	0.8548	1.9915	0.1582
AGRatio	1	0.9500	1.3066	0.5287	0.4671

Odds Ratio Estimates		
Effect	Point Estimate	95% Wald Confidence Limits
Age	1.018	1.002 1.035
TB	1.012	0.860 1.192
DB	1.434	0.905 2.274
Alkphos	1.001	0.999 1.003
Alamine	1.015	1.002 1.028
Aspartate	1.003	0.995 1.010
TP	1.918	0.811 4.533
ALB	0.299	0.056 1.598
AGRatio	2.586	0.200 33.478

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.1	Somers' D	0.562
Percent Discordant	21.9	Gamma	0.562
Percent Tied	0.0	Tau-a	0.211
Pairs	33180	c	0.781

Influence Diagnostics

Influence Diagnostics

The LOGISTIC Procedure

Predicted Probabilities for LiverPatient=Yes with 95% Confidence Limits

At TB=3.722 DB=1.701 Alkphos=284.1 Alamine=90.73 Aspartate=125.9 TP=6.4 ALB=3.069 AGRatio=0.939

We first fit the model by using all of the predictors and will see if there are any influential points. By looking at the diagnostics results, we can see that there are no influential points, so now we will perform model selection.

The LOGISTIC Procedure

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

Number of Observations Read	423
Number of Observations Used	421

Response Profile		
Ordered Value	LiverPatient	Total Frequency
1	Yes	316
2	No	105

Probability modeled is LiverPatient='Yes'.

Note: 2 observations were deleted due to missing values for the response or explanatory variables.

Backward Elimination Procedure

Step 0. The following effects were entered:

Intercept Age TB DB Alkphos Alamine Aspartate TP ALB AGRatio

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	474.936	406.917
SC	478.979	447.343
-2 Log L	472.936	386.917

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	86.0193	9	<.0001
Score	47.0936	9	<.0001
Wald	38.0177	9	<.0001

Step 1. Effect TB is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	474.936	404.945
SC	478.979	441.329
-2 Log L	472.936	386.945

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	85.9909	8	<.0001
Score	47.0878	8	<.0001
Wald	38.0201	8	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
0.0231	1	0.8791

Step 2. Effect Aspartate is removed:

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

The LOGISTIC Procedure

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	474.936	403.467
SC	478.979	435.808
-2 Log L	472.936	387.467

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	85.4694	7	<.0001
Score	46.7901	7	<.0001
Wald	37.5502	7	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
0.4915	2	0.7821

Step 3. Effect AGRatio is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	474.936	402.205
SC	478.979	430.503
-2 Log L	472.936	388.205

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	84.7314	6	<.0001
Score	46.7888	6	<.0001
Wald	37.5716	6	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
1.1998	3	0.7530

Step 4. Effect Alkphos is removed:

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	474.936	401.843
SC	478.979	426.099
-2 Log L	472.936	389.843

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	83.0930	5	<.0001
Score	43.0511	5	<.0001
Wald	36.9307	5	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
2.4948	4	0.6456

Note: No (additional) effects met the 0.05 significance level for removal from the model.

Summary of Backward Elimination					
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	TB	1	8	0.0220	0.8821
2	Aspartate	1	7	0.4707	0.4927
3	AGRatio	1	6	0.7016	0.4023
4	Alkphos	1	5	1.2995	0.2543

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.4297	0.9551	2.2405	0.1344
Age	1	0.0190	0.00836	5.1573	0.0231
DB	1	0.4484	0.1751	6.5559	0.0105
Alamine	1	0.0188	0.00531	12.5841	0.0004
TP	1	0.4192	0.2052	4.1745	0.0410
ALB	1	-0.7260	0.2925	6.1621	0.0131

Odds Ratio Estimates		
Effect	Point Estimate	95% Wald Confidence Limits
Age	1.019	1.003 1.036
DB	1.566	1.111 2.207
Alamine	1.019	1.008 1.030
TP	1.521	1.017 2.274
ALB	0.484	0.273 0.858

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	77.2	Somers' D	0.544
Percent Discordant	22.8	Gamma	0.544
Percent Tied	0.0	Tau-a	0.204
Pairs	33180	c	0.772

Partition for the Hosmer and Lemeshow Test					
		LiverPatient = Yes		LiverPatient = No	
Group	Total	Observed	Expected	Observed	Expected
1	42	19	19.54	23	22.46
2	42	25	23.26	17	18.74
3	42	27	25.20	15	16.80
4	42	22	27.58	20	14.42
5	42	32	29.99	10	12.01
6	42	32	32.45	10	9.55
7	42	36	35.24	6	6.76
8	42	40	38.70	2	3.30
9	42	40	41.06	2	0.94
10	43	43	42.97	0	0.03

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.3252	8	0.6109

Influence Diagnostics

Influence Diagnostics

Predicted Probabilities for LiverPatient=Yes with 95% Confidence Limits
At DB=1.701 Alamine=90.73 TP=6.4 ALB=3.069
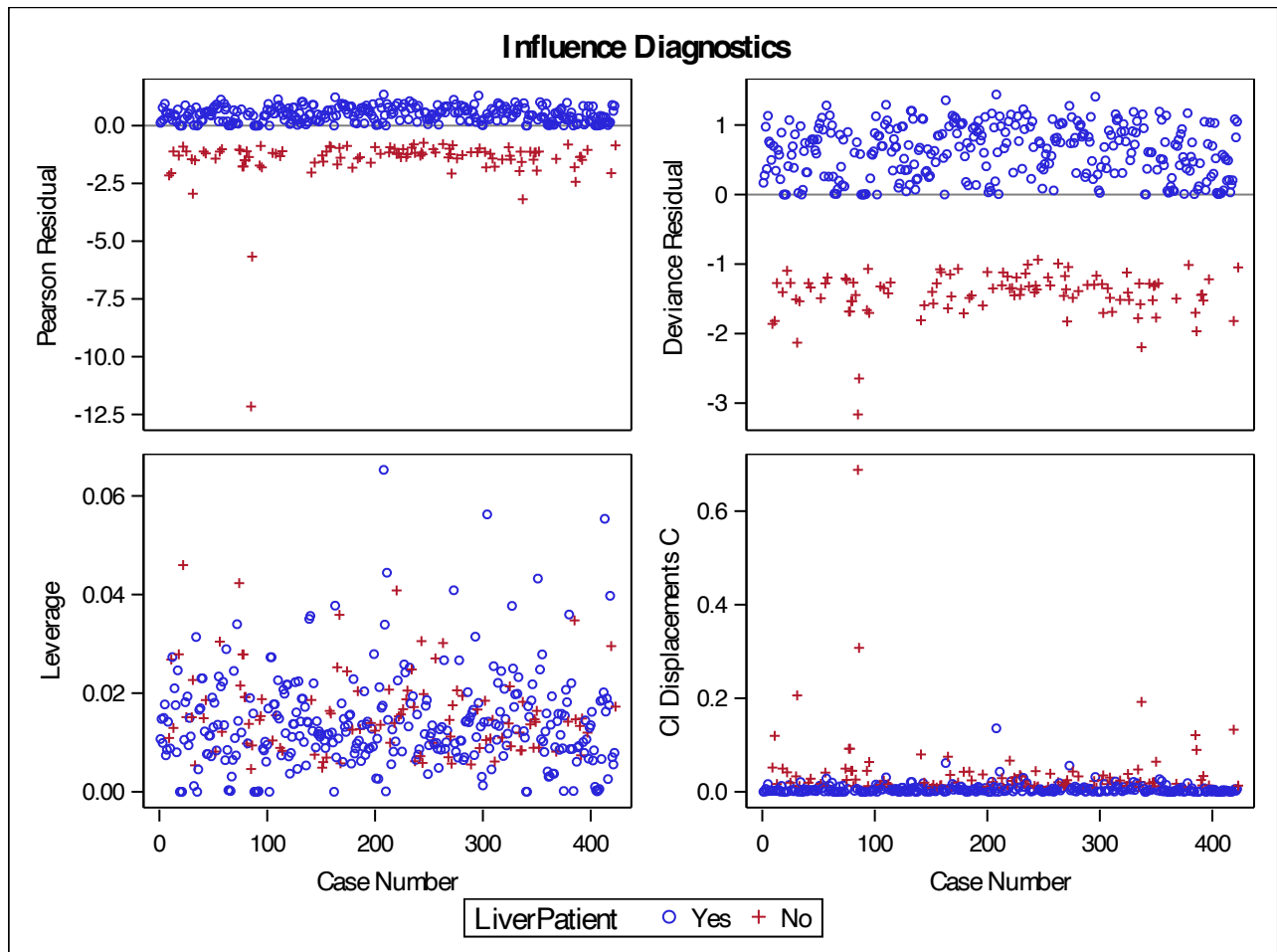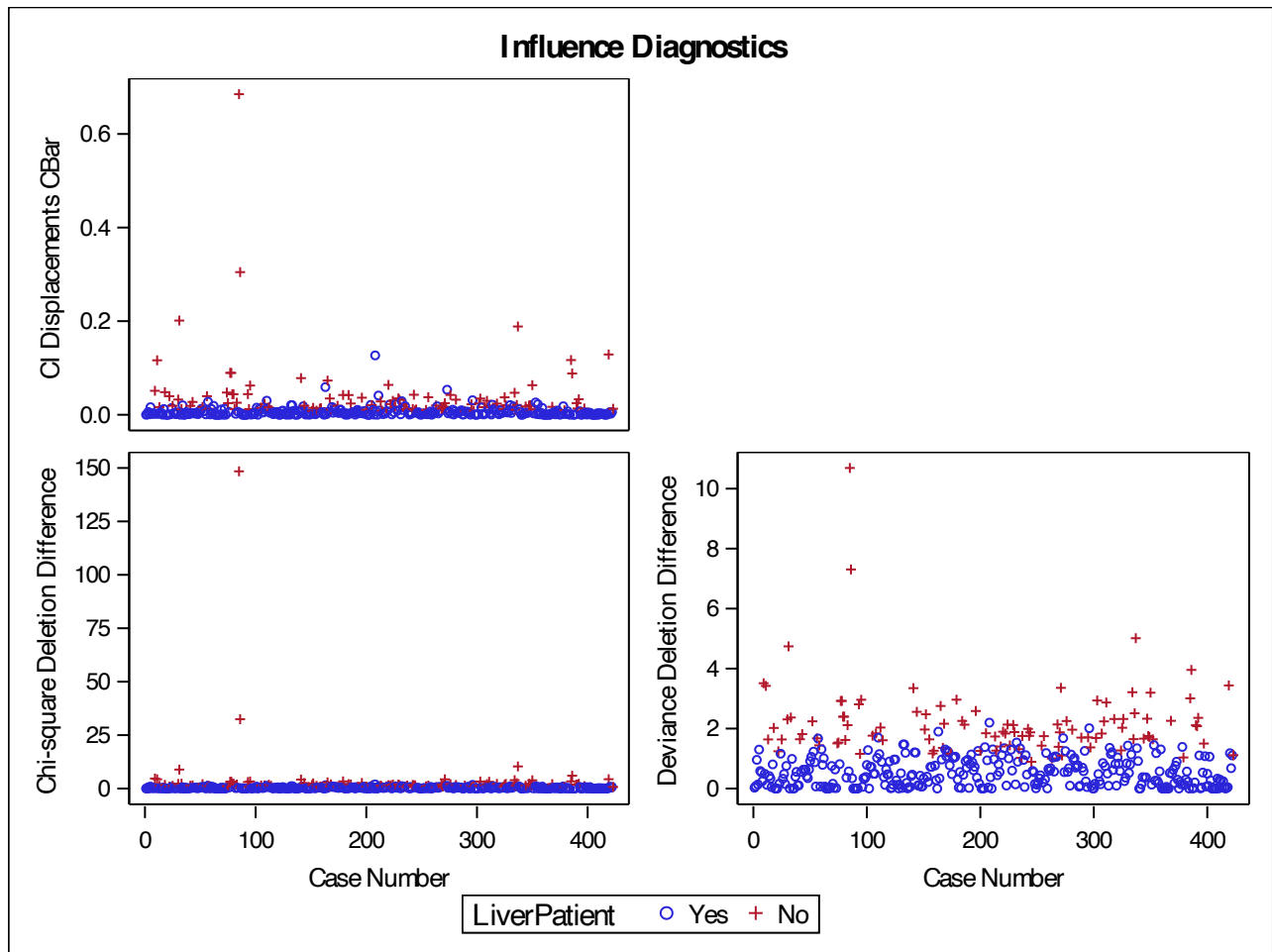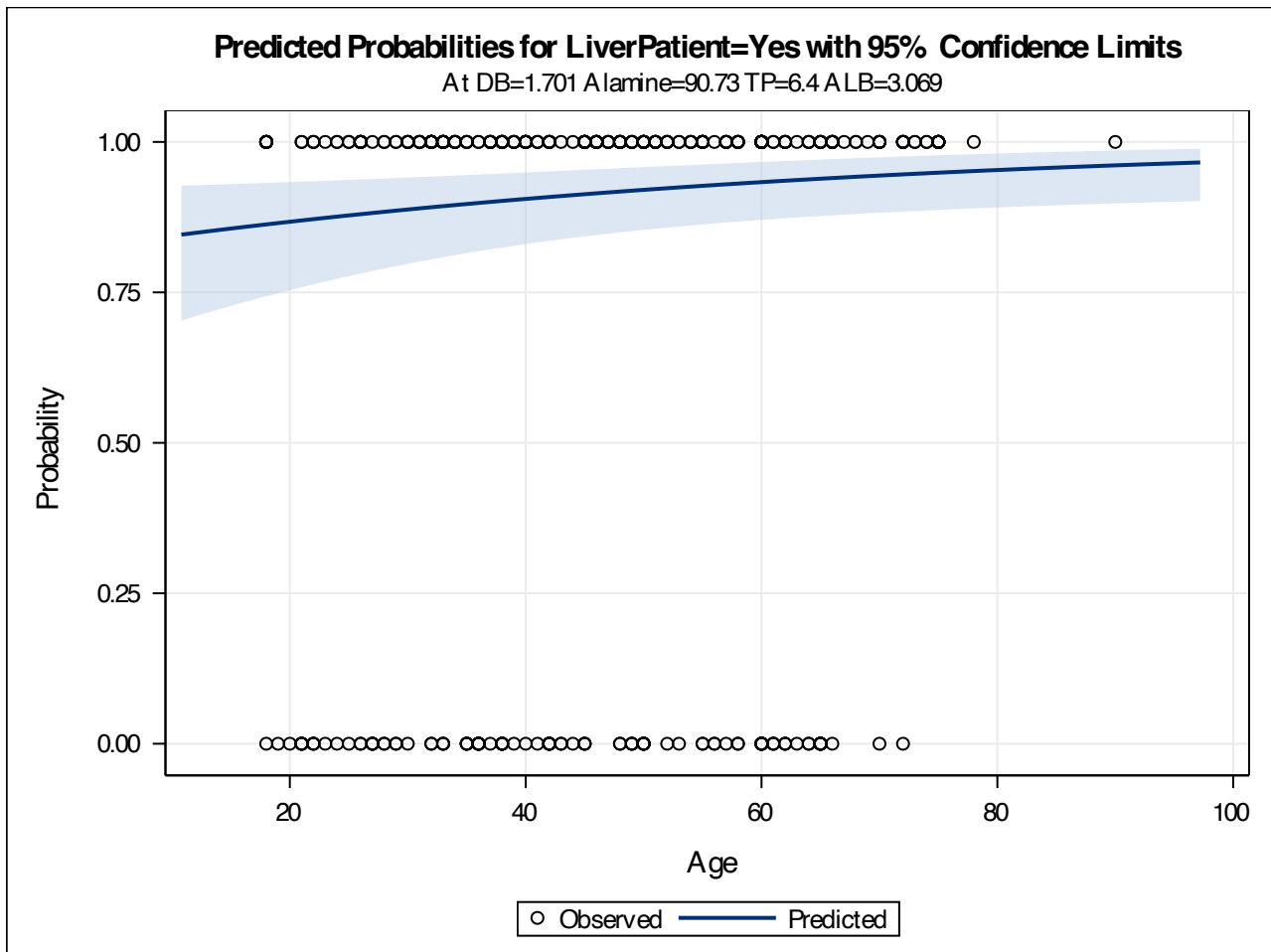
We have now our final model with Age, DB, Alamine, TP and ALB predictors.

Exercise 2b):

Looking at the table Model Fit statistics, the Log Likelihood column values under Intercept and Covariates can be compared to the Intercept Only column values. We have small values for both AIC and SC(401.843 and 426.099), so we may conclude that model with predictors will be better. The global tests indicate there is at least one significant parameter estimate. Hosmer-Lemeshow'stest gives us p-value of 0.6109, which is larger than 0.05, hence we may conclude that we accept null hypothesis, and our model fits okay.There are no influential points based on cbar and diagnostics look fine.

Exercise 2c):

The odds ratio for Age is 1.019, for DB is 1.566, for Alamine is 1.019, for TP is 1.521 and for ALB is 0.483 and they are significant.We performed backward selection and can see now that the confidence intervals do not include 1. The odds of an adult male being a liver patient with liver disease increase by a factor of 1.019 with a one unit increase in Aspartate when others are constant.The odds of an adult male being a liver patient with liver disease increase by a factor of 1.566 with a one unit increase in DB when others are constant.The odds of an adult male being a liver patient with liver disease increase by a factor of 1.019 with a one unit increase in Alamine when others are constant.The odds of an adult male being a liver patient with liver disease increase by a factor of 1.521 with a one unit increase in TP when others are constant.The odds of an adult male being a liver patient with liver disease increase by a factor of 0.484 with a one unit increase in ALB when others are constant.By looking at the results we can see that we need more predictors to model predicting whether a male is a liver patient rather than predicting a female.

Exercise 3a):

We first consider a gamma log-linear model for four ounce drinks consumed per day as a function of four different predictors.

The GENMOD Procedure

Model Information	
Data Set	WORK.BUPA
Distribution	Gamma
Link Function	Log
Dependent Variable	four_oz

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	331	307.7732	0.9298
Scaled Deviance	331	379.0762	1.1452
Pearson Chi-Square	331	250.5826	0.7570
Scaled Pearson X2	331	308.6361	0.9324
Log Likelihood		-966.2582	
Full Log Likelihood		-966.2582	
AIC (smaller is better)		1944.5163	
AICC (smaller is better)		1944.7716	
BIC (smaller is better)		1967.4190	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.2618	1.0536	-5.3267	-1.1969	9.59	0.0020
mcv	1	0.0511	0.0117	0.0281	0.0740	19.01	<.0001
alkphos	1	0.0011	0.0025	-0.0039	0.0061	0.19	0.6650
sgot	1	0.0126	0.0065	-0.0002	0.0255	3.74	0.0532
gammagt	1	0.0040	0.0016	0.0008	0.0072	5.87	0.0154
Scale	1	1.2317	0.0852	1.0756	1.4105		

Note: The scale parameter was estimated by maximum likelihood.

LR Statistics For Type 1 Analysis				
Source	2*LogLikelihood	DF	Chi-Square	Pr > ChiSq
Intercept	-1987.4451			
mcv	-1954.8558	1	32.59	<.0001
alkphos	-1953.1587	1	1.70	0.1927
sgot	-1938.9391	1	14.22	0.0002
gammagt	-1932.5163	1	6.42	0.0113

The GENMOD Procedure

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
mcv	1	18.13	<.0001
alkphos	1	0.19	0.6639
sgot	1	3.92	0.0478
gammagt	1	6.42	0.0113

First thing we can notice is that p-values for the predictors alkphos and sgot are larger than 0.05, so they are not statistically significant. However, to get to a final main effects model from here, we would want to remove terms one at a time based on type 3 and type 1 analyses. So both analyses say that we need to remove alkphos variable. Now we fit the model without alkphos.

The GENMOD Procedure

Model Information	
Data Set	WORK.BUPA
Distribution	Gamma
Link Function	Log
Dependent Variable	four_oz

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	332	307.9265	0.9275
Scaled Deviance	332	379.0939	1.1418
Pearson Chi-Square	332	252.3314	0.7600
Scaled Pearson X2	332	310.6498	0.9357
Log Likelihood		-966.3526	
Full Log Likelihood		-966.3526	
AIC (smaller is better)		1942.7051	
AICC (smaller is better)		1942.8870	
BIC (smaller is better)		1961.7907	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.2055	1.0456	-5.2548	-1.1562	9.40	0.0022
mcv	1	0.0512	0.0117	0.0282	0.0741	19.11	<.0001
sgot	1	0.0130	0.0065	0.0002	0.0257	3.97	0.0462
gammagt	1	0.0040	0.0016	0.0008	0.0072	6.05	0.0139
Scale	1	1.2311	0.0851	1.0751	1.4098		

Note: The scale parameter was estimated by maximum likelihood.

LR Statistics For Type 1 Analysis				
Source	2*LogLikelihood	DF	Chi-Square	Pr > ChiSq
Intercept	-1987.4451			
mcv	-1954.8558	1	32.59	<.0001
sgot	-1939.3421	1	15.51	<.0001
gammagt	-1932.7051	1	6.64	0.0100

The GENMOD Procedure

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
mcv	1	18.22	<.0001
sgot	1	4.16	0.0413
gammagt	1	6.64	0.0100

Diagnostic Statistics for four_oz
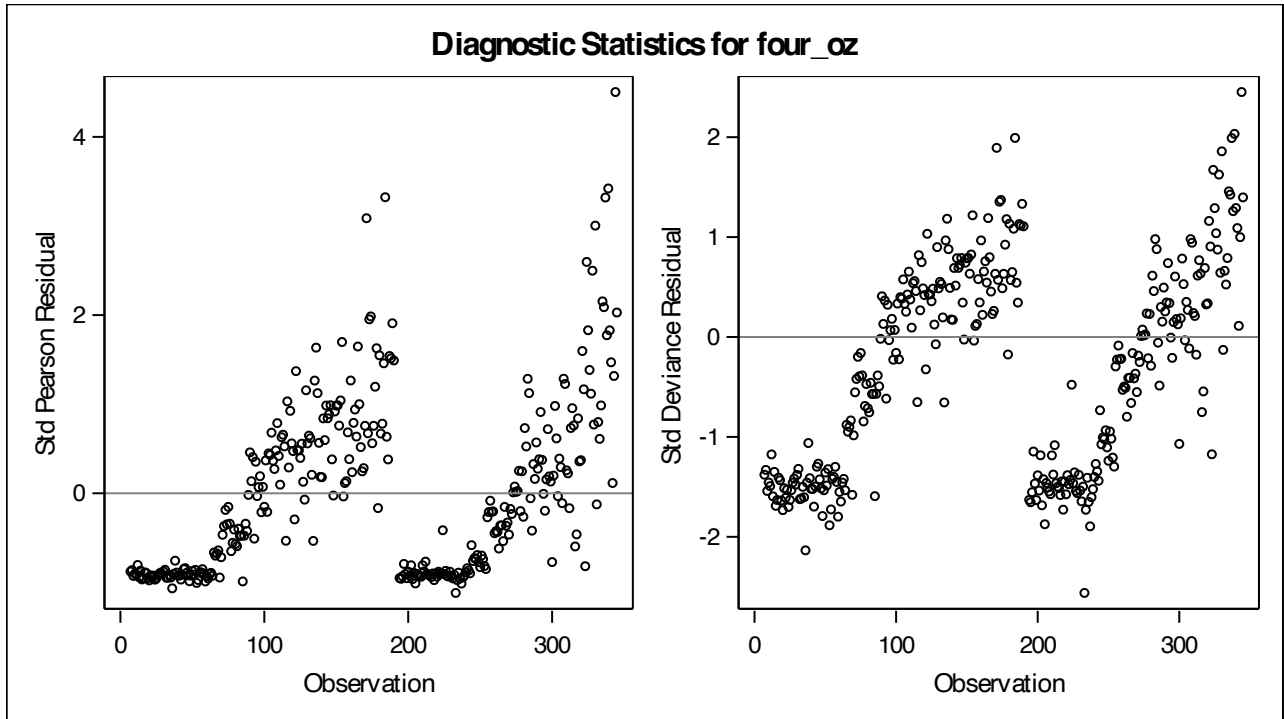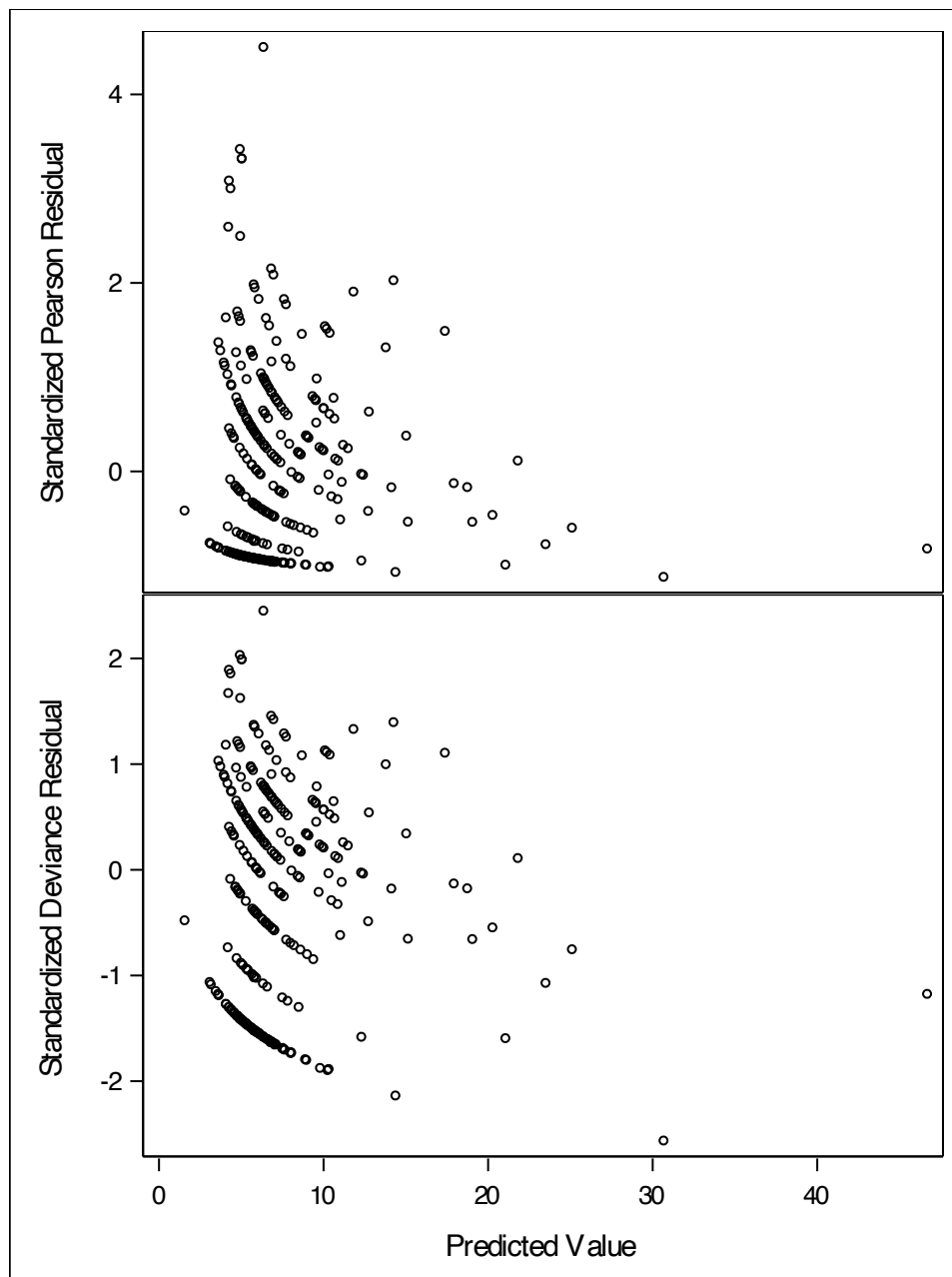
We have removed alkphos variable and now see that sgot has become significant, p-value is 0.0462<0.05.

Exercise 3b):

So we have selected our best model based on Type 1 and Type 3 Analysis. Looking at the p-values for the parameter estimates, we can conclude that all of them are statistically significant. Residuals appear to have an upward trend, and it is also can be seen from the diagnostics plot. For the mcv variable, one increase in four_oz would result in an expected exp(0.0512)=1.053 multiplicative increase in number of four_oz. For sgot variable, one increase in four_oz would result in an expected exp(0.0130)=1.013 multiplicative increase in number of four_oz. For gammagt variable, one increase in four_oz would result in an expected exp(0.0040)=1.004 multiplicative increase in number of four_oz.

Exercise 4a):

We now consider log-linear Poisson model for the same variable(four_oz).

The GENMOD Procedure

Model Information	
Data Set	WORK.BUPA
Distribution	Poisson
Link Function	Log
Dependent Variable	four_oz

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	340	1705.9394	5.0175
Scaled Deviance	340	1705.9394	5.0175
Pearson Chi-Square	340	1722.7243	5.0668
Scaled Pearson X2	340	1722.7243	5.0668
Log Likelihood		2402.9285	
Full Log Likelihood		-1413.7027	
AIC (smaller is better)		2837.4055	
AICC (smaller is better)		2837.5824	
BIC (smaller is better)		2856.6232	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.4612	0.4462	-4.3358	-2.5866	60.16	<.0001
mcv	1	0.0535	0.0049	0.0439	0.0631	118.29	<.0001
alkphos	1	0.0029	0.0011	0.0007	0.0051	6.88	0.0087
sgot	1	0.0066	0.0020	0.0028	0.0105	11.34	0.0008
gammagt	1	0.0036	0.0005	0.0027	0.0045	57.20	<.0001
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	2063.2971			
mcv	1840.9124	1	222.38	<.0001
alkphos	1823.5016	1	17.41	<.0001
sgot	1756.8215	1	66.68	<.0001
gammagt	1705.9394	1	50.88	<.0001

The GENMOD Procedure

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
mcv	1	119.91	<.0001
alkphos	1	6.80	0.0091
sgot	1	11.02	0.0009
gammagt	1	50.88	<.0001

Diagnostic Statistics for four_oz
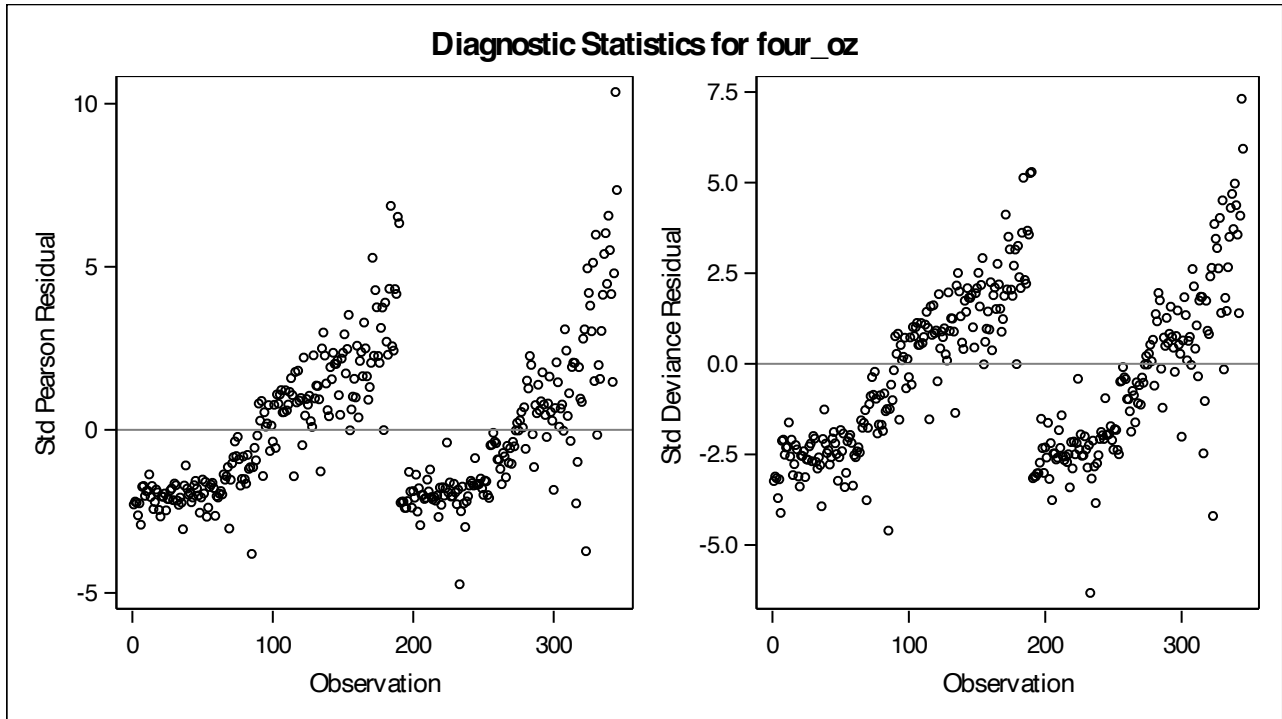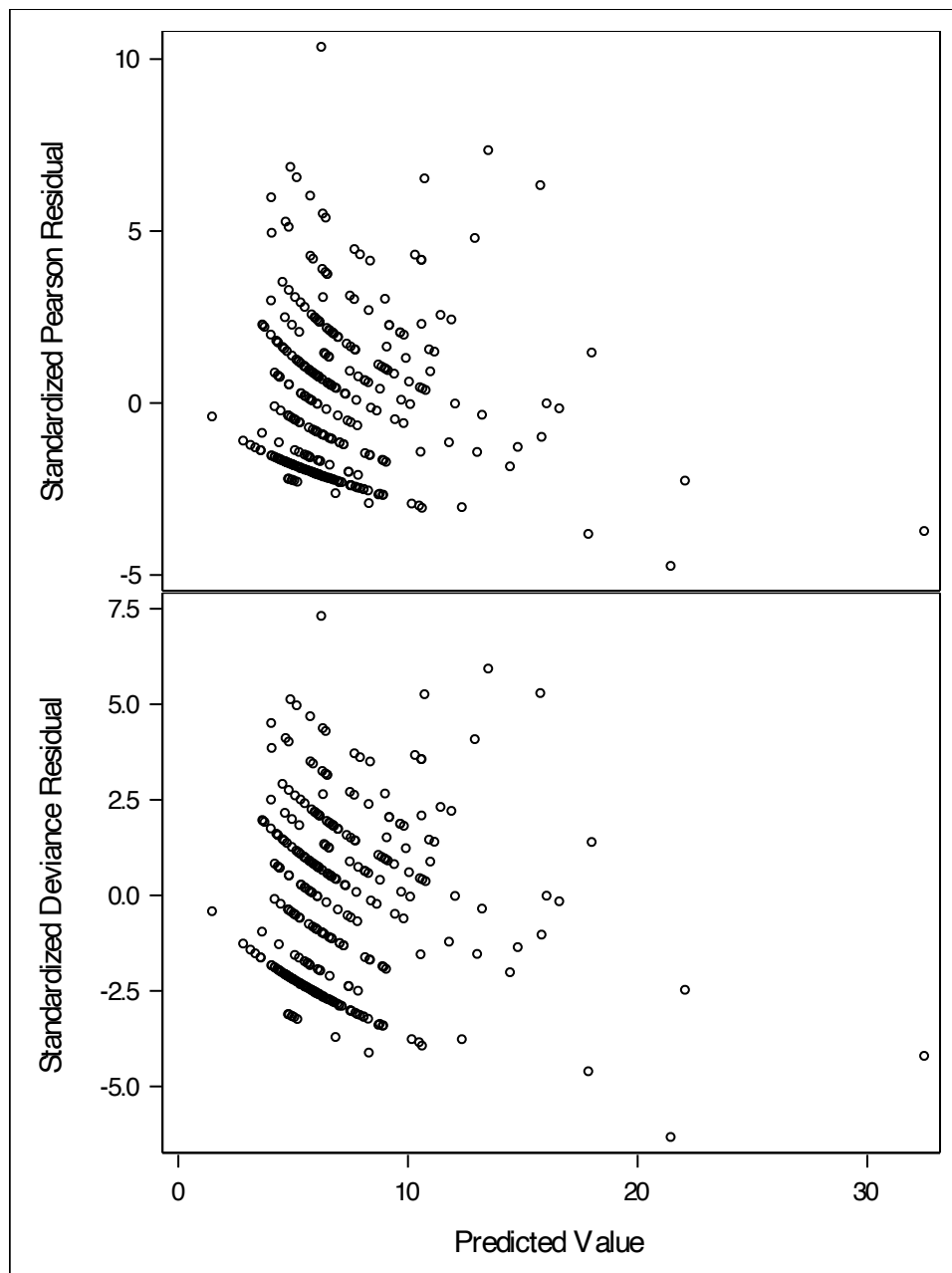
We can see that p-values for all of the variables are less than 0.05, and we can conlude that all of them are significant. This is also can be confirmed by Type 1 and Type 3 analyses.

We can see from the previous model that our scaled deviance over degrees of freedom is far from 1(=5.0175), so there is a need to consider overdispersed Poisson as well.

The GENMOD Procedure

Model Information	
Data Set	WORK.BUPA
Distribution	Poisson
Link Function	Log
Dependent Variable	four_oz

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	340	1705.9394	5.0175
Scaled Deviance	340	340.0000	1.0000
Pearson Chi-Square	340	1722.7243	5.0668
Scaled Pearson X2	340	343.3453	1.0098
Log Likelihood		478.9125	
Full Log Likelihood		-1413.7027	
AIC (smaller is better)		2837.4055	
AICC (smaller is better)		2837.5824	
BIC (smaller is better)		2856.6232	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.4612	0.9995	-5.4202	-1.5021	11.99	0.0005
mcv	1	0.0535	0.0110	0.0319	0.0751	23.58	<.0001
alkphos	1	0.0029	0.0025	-0.0020	0.0078	1.37	0.2415
sgot	1	0.0066	0.0044	-0.0020	0.0152	2.26	0.1327
gammagt	1	0.0036	0.0011	0.0015	0.0057	11.40	0.0007
Scale	0	2.2400	0.0000	2.2400	2.2400		

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	2063.2971						
mcv	1840.9124	1	340	44.32	<.0001	44.32	<.0001
alkphos	1823.5016	1	340	3.47	0.0634	3.47	0.0625
sgot	1756.8215	1	340	13.29	0.0003	13.29	0.0003
gammagt	1705.9394	1	340	10.14	0.0016	10.14	0.0015

The GENMOD Procedure
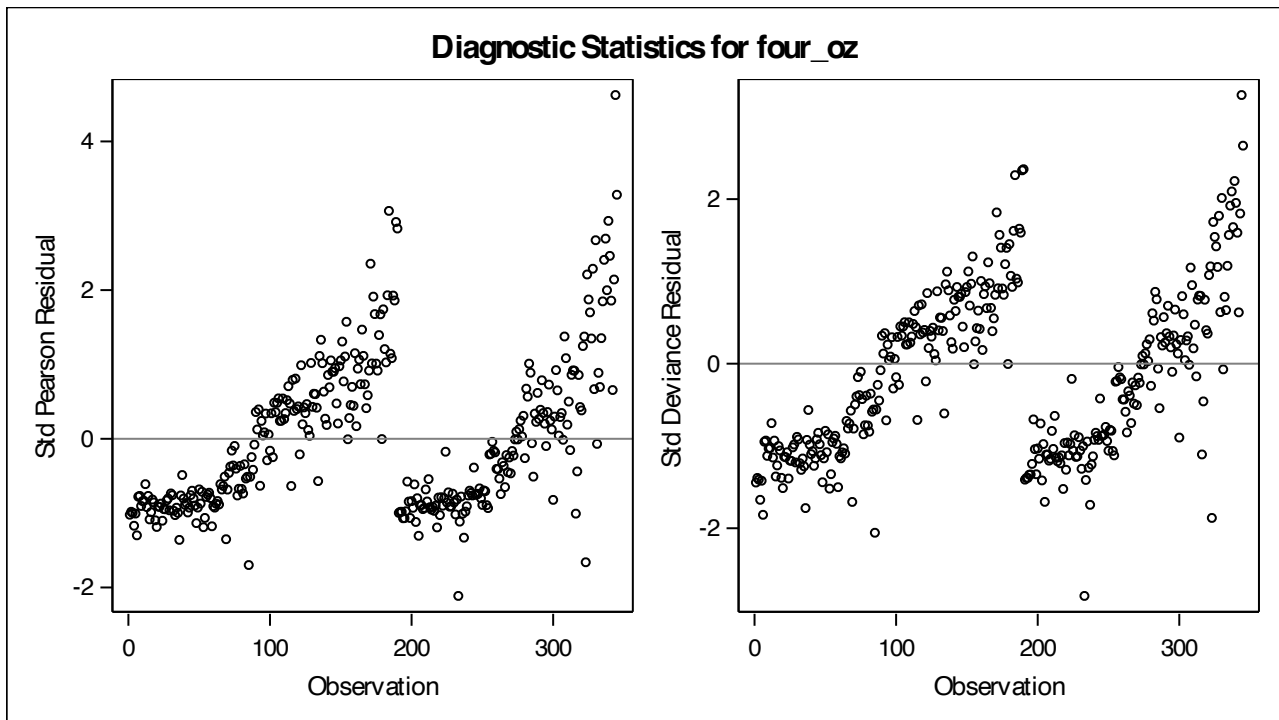
LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
mcv	1	340	23.90	<.0001	23.90	<.0001
alkphos	1	340	1.36	0.2452	1.36	0.2444
sgot	1	340	2.20	0.1393	2.20	0.1384
gammagt	1	340	10.14	0.0016	10.14	0.0015

Diagnostic Statistics for four_oz

First thing we can notice is that p-values for the predictors alkphos and sgot are larger than 0.05, so they are not statistically significant. To get to a final main effects model from here, we would want to remove terms one at a time based on type 3 and type 1 analyses.

The GENMOD Procedure

Model Information	
Data Set	WORK.BUPA
Distribution	Poisson
Link Function	Log
Dependent Variable	four_oz

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	342	1725.4831	5.0453
Scaled Deviance	342	342.0000	1.0000
Pearson Chi-Square	342	1770.7497	5.1776
Scaled Pearson X2	342	350.9721	1.0262
Log Likelihood		474.3365	
Full Log Likelihood		-1423.4746	
AIC (smaller is better)		2852.9491	
AICC (smaller is better)		2853.0195	
BIC (smaller is better)		2864.4798	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.3887	0.9841	-5.3175	-1.4598	11.86	0.0006
mcv	1	0.0564	0.0109	0.0351	0.0777	26.89	<.0001
gammagt	1	0.0045	0.0009	0.0028	0.0063	26.69	<.0001
Scale	0	2.2462	0.0000	2.2462	2.2462		

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	2063.2971						
mcv	1840.9124	1	342	44.08	<.0001	44.08	<.0001
gammagt	1725.4831	1	342	22.88	<.0001	22.88	<.0001

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
mcv	1	342	27.22	<.0001	27.22	<.0001
gammagt	1	342	22.88	<.0001	22.88	<.0001

Diagnostic Statistics for four_oz
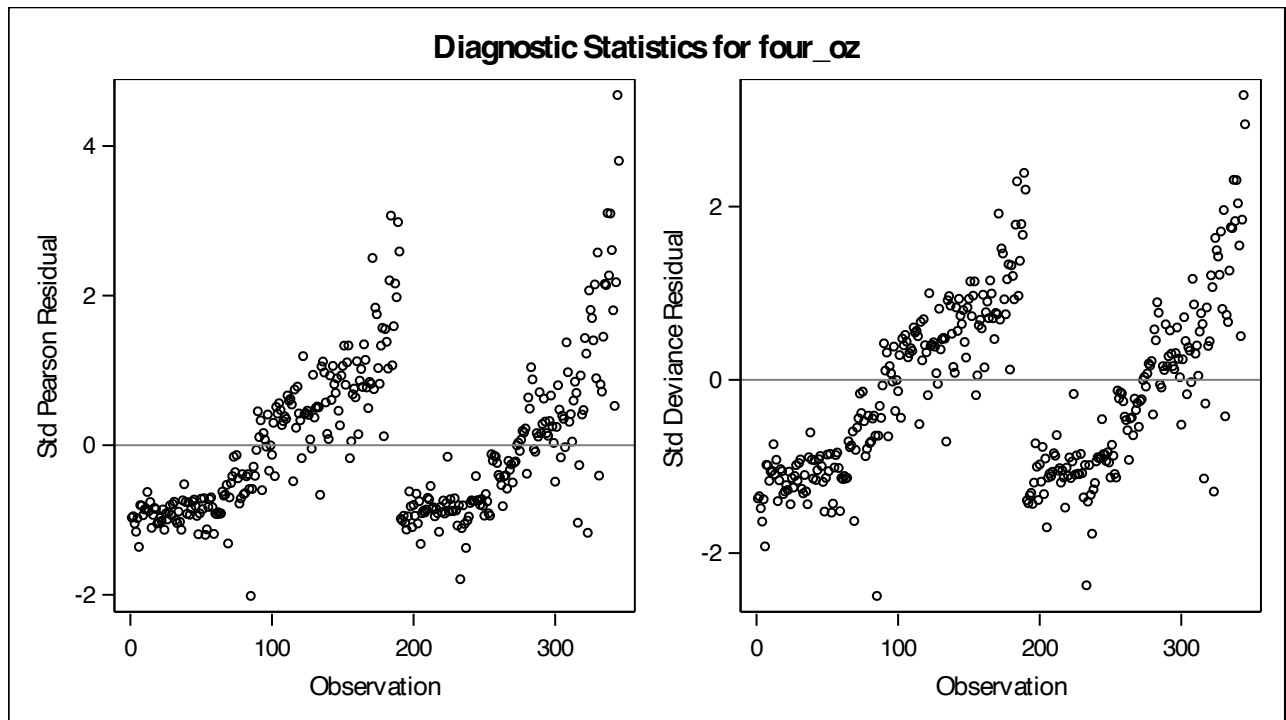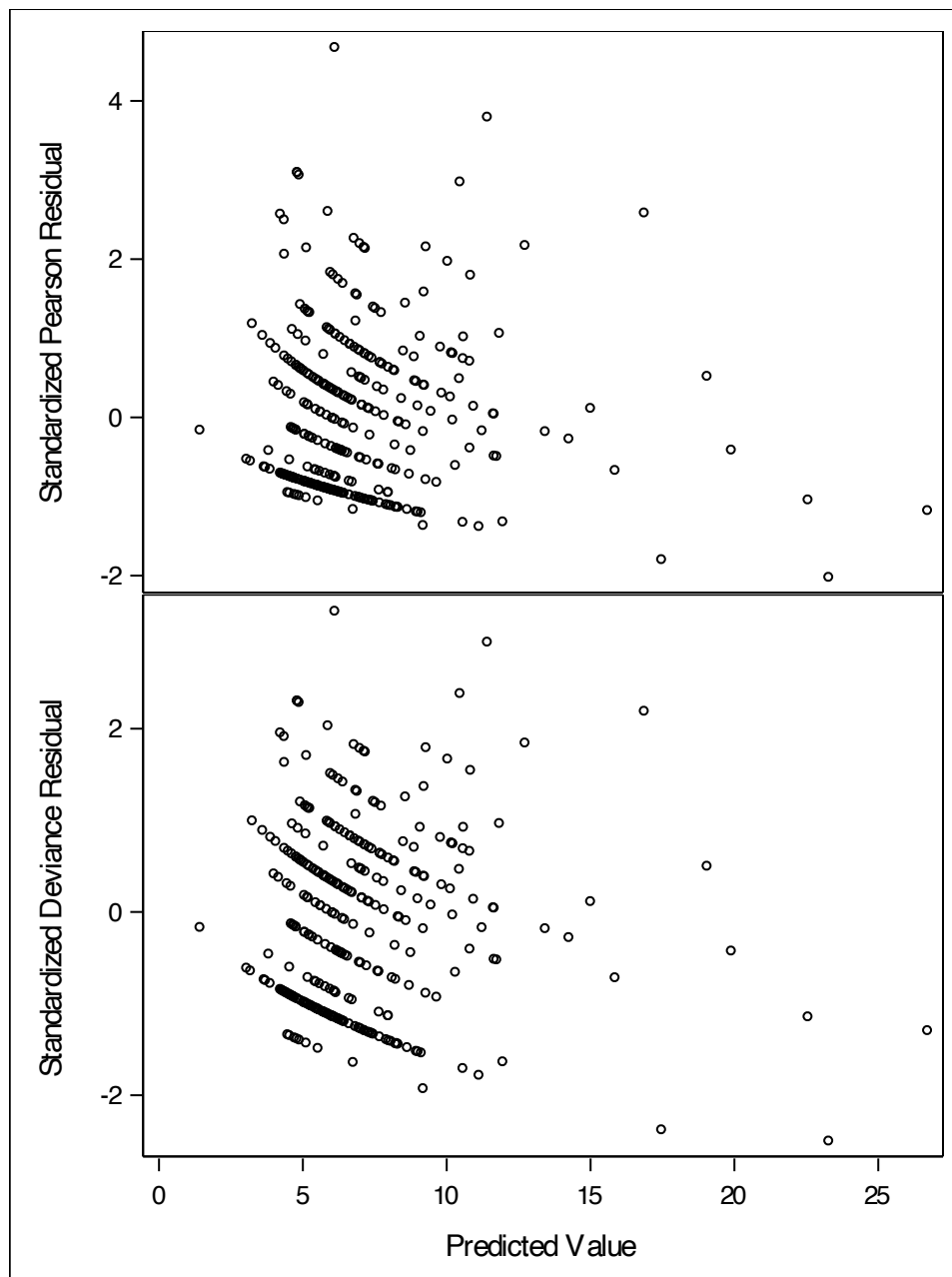
We removed both alkphos and sgot predictors since they were not significant, and Type 1 analysis also confirmed that we can remove alkphos. However removing just alkphos was not enough, so Type 3 analysis showed that sgot could be removed first actually.

Exercise 4b):

Now we got our final models for regular log-linear Poisson and overdispersed Poisson.

1. Log-linear Poisson: Looking at the p-values for the parameter estimates, we can conclude that all of them are statistically significant.Residuals appear to have an upward trend, and it is also can be seen from the diagnostics plot. For the mcv variable, one increase in four_oz would result in an expected exp(0.0535)=1.055 multiplicative increase in number of four_oz. For alkphos variable, one increase in four_oz would result in an expected exp(0.0029)=1.003 multiplicative increase in number of four_oz.For sgot variable, one increase in four_oz would result in an expected exp(0.0066)=1.007 multiplicative increase in number of four_oz. For gammagt variable, one increase in four_oz would result in an expected exp(0.0036)=1.004 multiplicative increase in number of four_oz. 2. Overdispersed Poisson: Looking at the p-values for the parameter estimates, we can conclude that all of them are statistically significant. Residuals appear to have an upward trend, and it is also can be seen from the diagnostics plot. For the mcv variable, one increase in four_oz would result in an expected exp(0.0564)=1.058 multiplicative increase in number of four_oz. For gammagt variable, one increase in four_oz would result in an expected exp(0.0045)=1.0045 multiplicative increase in number of four_oz. As we can see the diagnostic plots were almost identical in all cases. However, looking at the estimates we can conclude that it will be reasonable to consider overdispersed Poisson model, since it is simple(involves only two predictors) and the estimates are little bit higher compared to the estimates in the previous (Gamma and regular Poisson models).