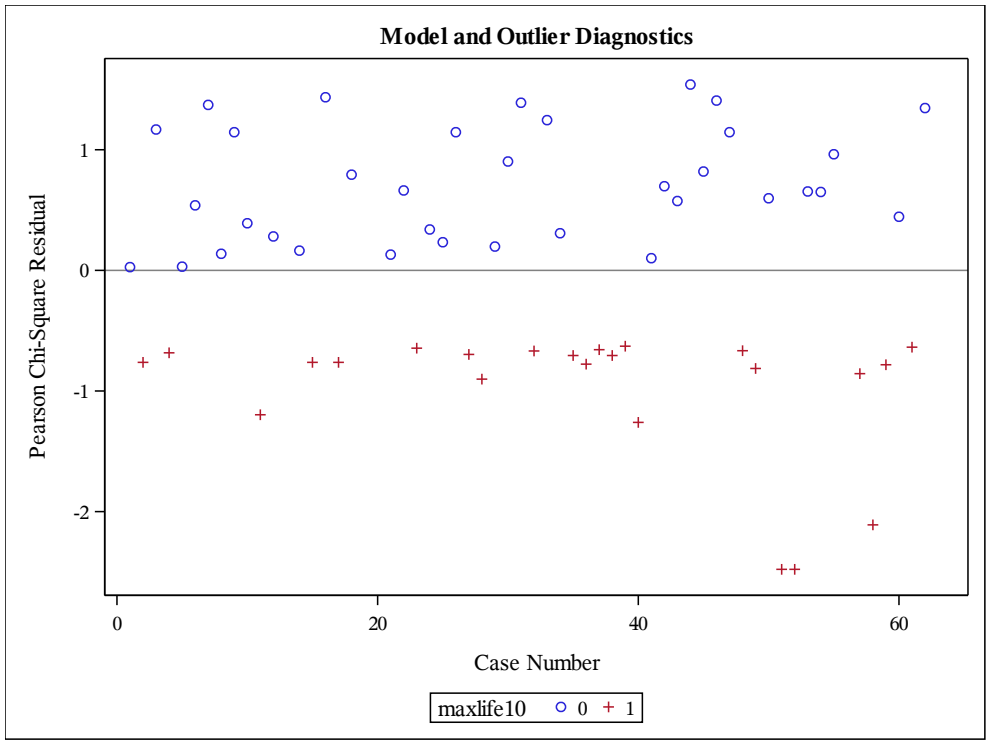
HW 4 Spring 2018 Solution

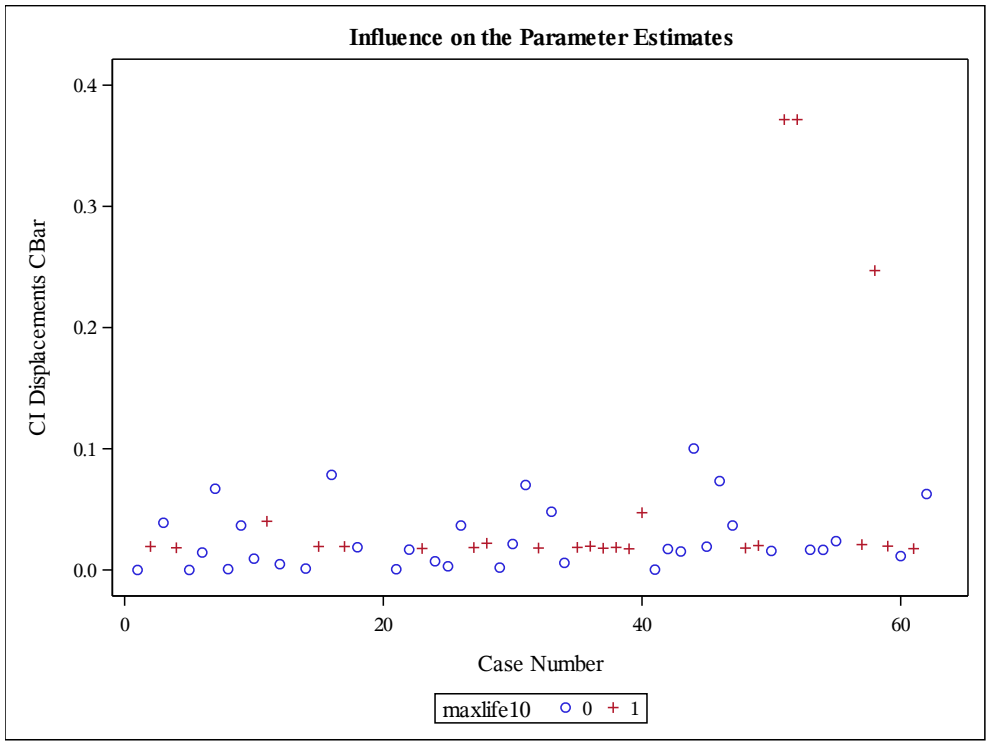
Note: For the Cbar values in this homework solution, the cutoff is 0.5. The Cook's distance cutoff value is 1.

Exercise 1

a. The best model from stepwise selection contains gestationtime as the only predictor. In the diagnostic plots, we find no unduly influential points based on the residuals and Cbar measures (cutoff of 0.5). Thus, we will keep all observations. Note that SAS's message about the validity of the model being questionable (and quasicomplete separation of points) is from Step 2 of the model selection process, which contains overalldangerindex and gestationtime. The final model with only gestationtime is the warning-free choice.

Summary of Stepwise Selection							
	Effect			Number	Score	Wald	
Step	Entered	Removed	DF	In	Chi-Square	Chi-Square	Pr > ChiSq
1	gestationtime		1	1	10.4038		0.0013
2	overalldangerindex		4	2	13.7419		0.0082
3		overalldangerindex	4	1		2.5432	0.6369





b. Now the final model is fitted with one predictor, gestation time. It is significant with a p-value of 0.004. Hence, we can conclude that gestation time is statistically significant in predicting the probability that a species' maximum lifespan will be at least 10 years. The goodness of fit test for the model has a p-

value of 0.407, which indicates the model fit is reasonable. The r-square is 0.233 which means the model predicts poorly.

R-Square	0.2325	Max-rescaled R-Square	0.3147

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.8759	0.4632	3.5760	0.0586
gestationtime	1	0.0119	0.00414	8.2827	0.0040

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
8.2739	8	0.4072

c. The odds ratio is estimated as 1.012 so we can say for a one day increase in gestation time, we expect to see roughly an 1.2% increase in the odds of a species having maximum lifespan at least 10 years. The 95% confidence interval (1.004, 1.020) does not contain 1, thus the odds ratio for gestation time is statistically significant.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
gestationtime	1.012	1.004	1.020

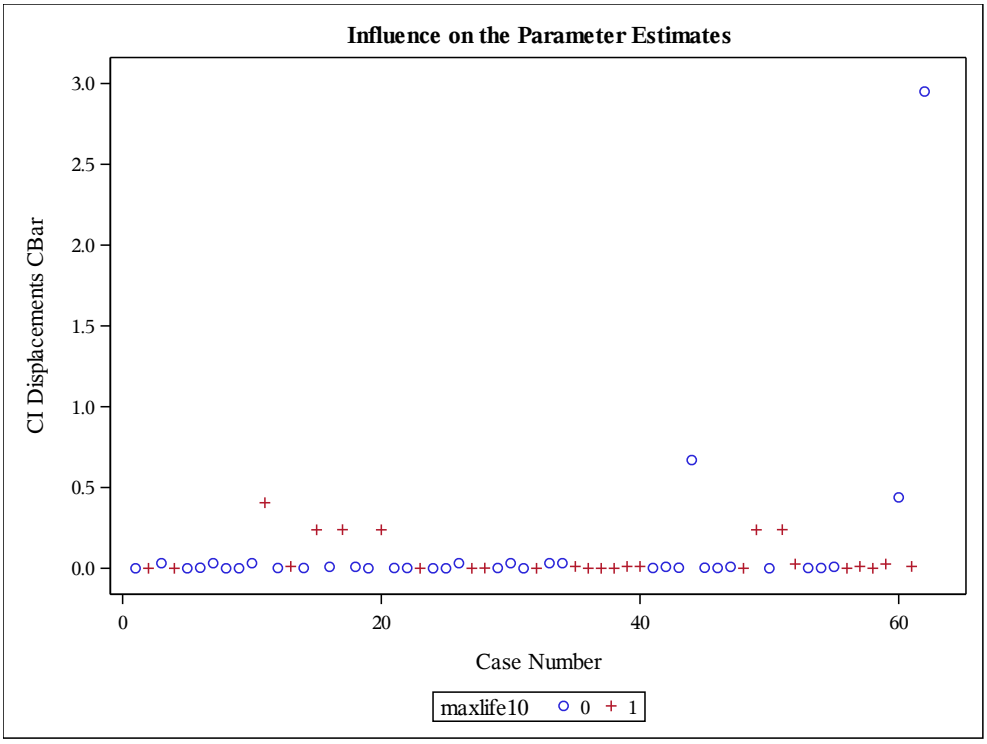
Exercise 2

a. Treating the index variables as continuous, we needed to do several model fittings and re-fittings due to influential points. The idea here is to show the Cbar based on the model that was selected, but refit without finalizing the selection until the influential points were removed. This process could be tedious but does avoid the quasi or complete separation of points issue.

The diagnostic plots for Cbar as well as data print outs of the Cbar values guided us to remove 3 observations due to influence. These observations were 62, 44, and 59 at each iteration of the model fitting and refitting. These 3 points were removed one by one between each model fitting iteration because they were had Cbar values greater than 0.5.

Iteration 1 (Remove largest Cbar value, Obs 62)

Summary of Stepwise Selection							
	Effect			Number	Score	Wald	
Step	Entered	Removed	DF	In	Chi-Square	Chi-Square	Pr > ChiSq
1	sleepexposureindex		1	1	13.9254		0.0002
2	predationindex		1	2	17.6170		<.0001

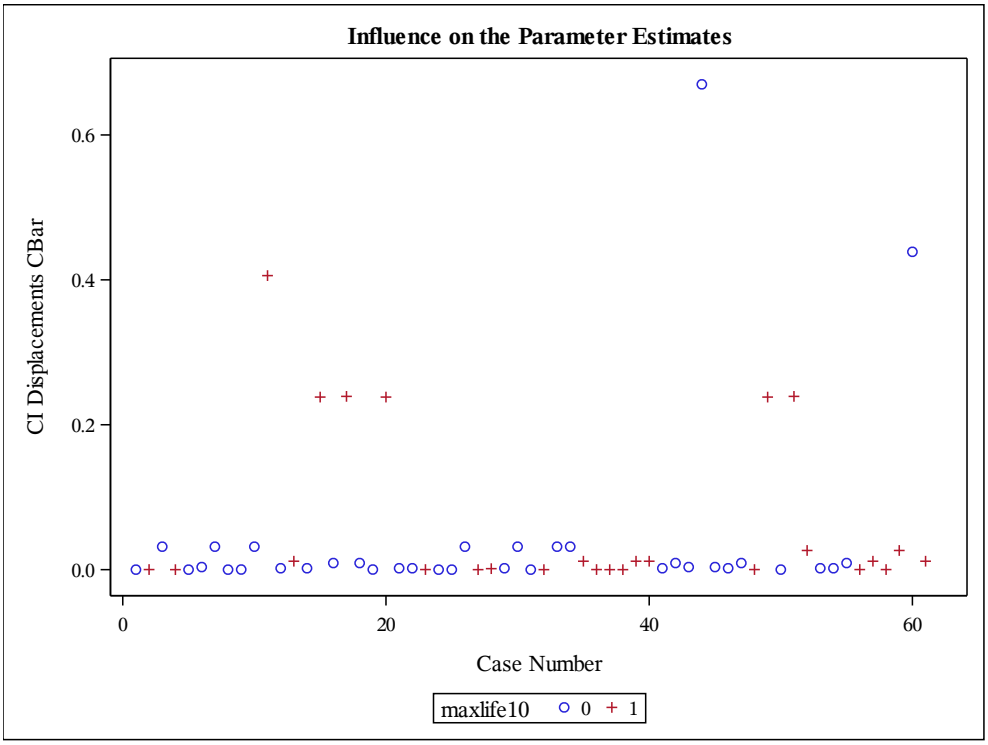


Obs	species	bodyweight	brainweight	totalsleep	maxlifespan	gestationtime	predationindex
44	Phanlang	1.62	11.4	13.7	13	17	2
62	Yellow-b	4.05	17.0	.	13	38	3

Obs	sleepexposureindex	overalldangerindex	maxlife10	cb2
44	1	2	0	0.66986
62	1	1	0	2.95031

Iteration 2 (Remove Obs 44)

Summary of Stepwise Selection							
Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	sleepexposureindex		1	1	13.9254		0.0002
2	predationindex		1	2	17.6170		<.0001

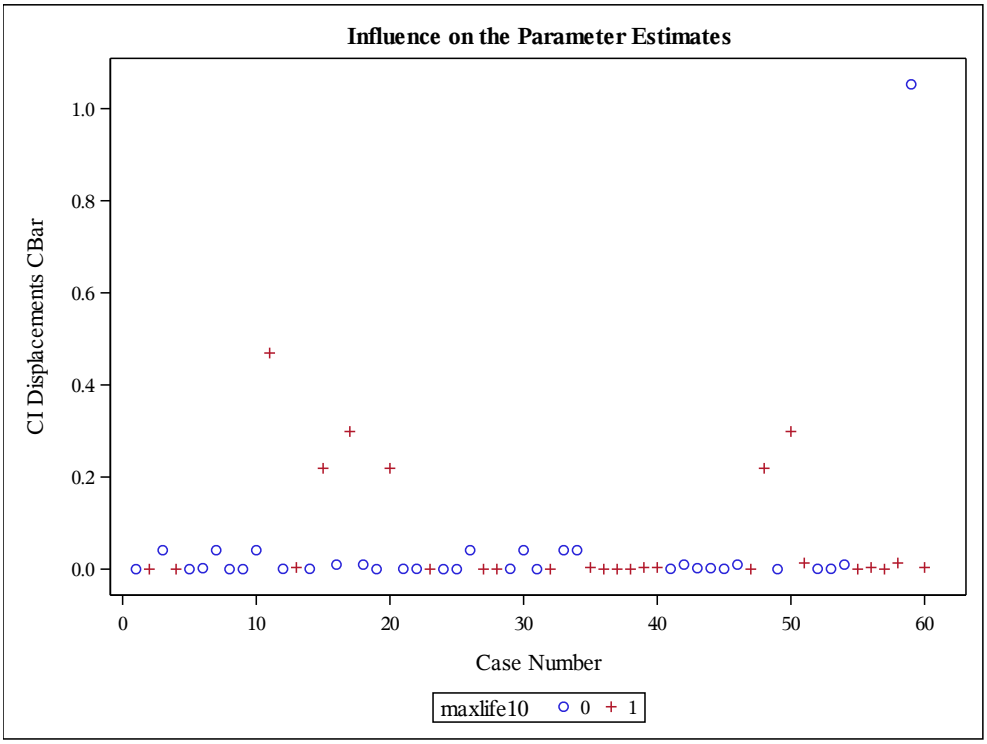


Obs	species	bodyweight	brainweight	totalsleep	maxlifespan	gestationtime	predationindex
44	Phanlang	1.62	11.4	13.7	13	17	2

Obs	sleepexposureindex	overalldangerindex	maxlife10	cb2	cb3
44	1	2	0	0.66986	0.66986

Iteration 3 (Remove Obs 59)

Summary of Stepwise Selection							
Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	sleepexposureindex		1	1	14.8981		0.0001
2	predationindex		1	2	17.8473		<.0001

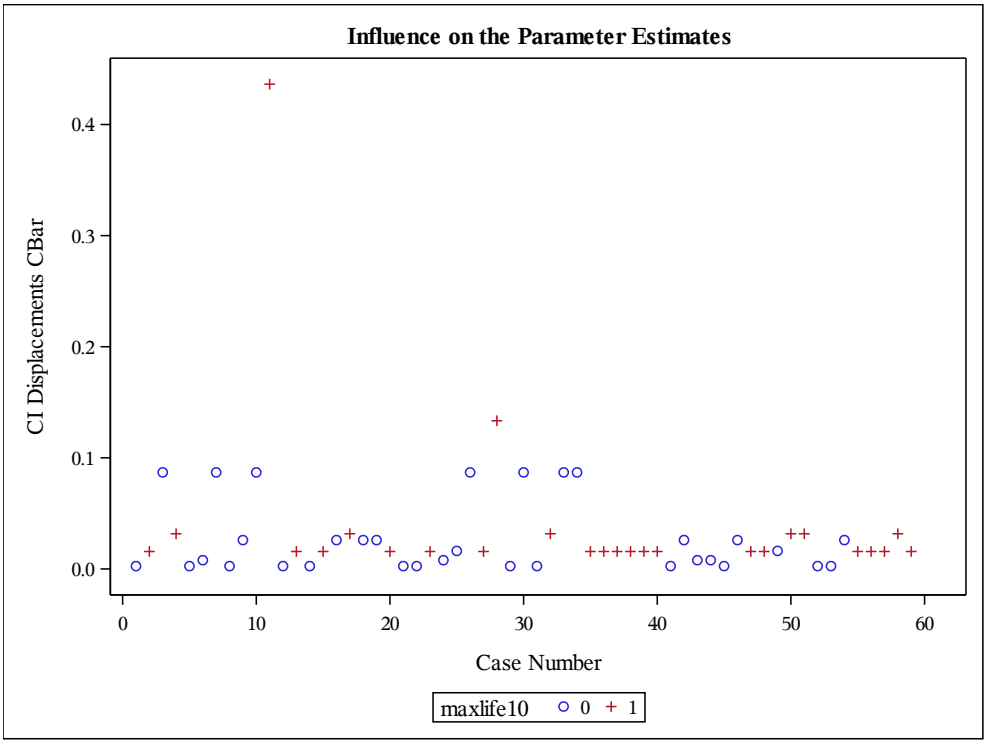


Obs	species	bodyweight	brainweight	totalsleep	maxlifespan	gestationtime	predationindex
59	Vervet	4.19	58	10.3	24	210	4

Obs	sleepexposureindex	overalldangerindex	maxlife10	cb2	cb3	cb4
59	3	4	0	0.43868	0.43868	1.05274

Thus, the final model we chose contained sleepexposureindex as the only predictor with no issues regarding Cbar.

Summary of Stepwise Selection							
	Effect			Number	Score	Wald	
Step	Entered	Removed	DF	In	Chi-Square	Chi-Square	Pr > ChiSq
1	sleepexposureindex		1	1	14.5215		0.0001
2	predationindex		1	2	18.3422		<.0001
3		predationindex	1	1		0.0615	0.8041



b. Now, the final model is fitted with one predictor, sleep exposure index. Sleep exposure index is statistically significant (p-value 0.004) meaning its estimated coefficient is far from 0. Thus, it aids in predicting whether the maximum lifespan of a species will be at least 10 years. The goodness of fit test for the model has a p-value of 0.574, which indicates the model fit is reasonable. The r-square is 0.323 which means the model predicts poorly, but better than the model in Exercise 1 which treated index variables including sleep exposure index as a categorical variable.

R-Square	0.3225	**Max-rescaled R-Square**	0.4320

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-2.0296	0.6259	10.5156	0.0012
sleepexposureindex	1	1.0614	0.3027	12.2927	0.0005

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	**DF**	**Pr > ChiSq**
1.1117	2	0.5736

c. The odds ratio is estimated as 2.890 so we can say for a one-unit increase in sleep exposure index, we expect to see an increase in the odds of a species having maximum lifespan at least 10 years by 2.89 times. The 95% confidence interval (1.597, 5.231) does not contain 1, thus the odds ratio for sleep exposure index (treated as continuous) is statistically significant. This reflects a strong practical increase in lifespan when the sleep exposure index grows.

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
sleepexposureindex	2.890	1.597	5.231

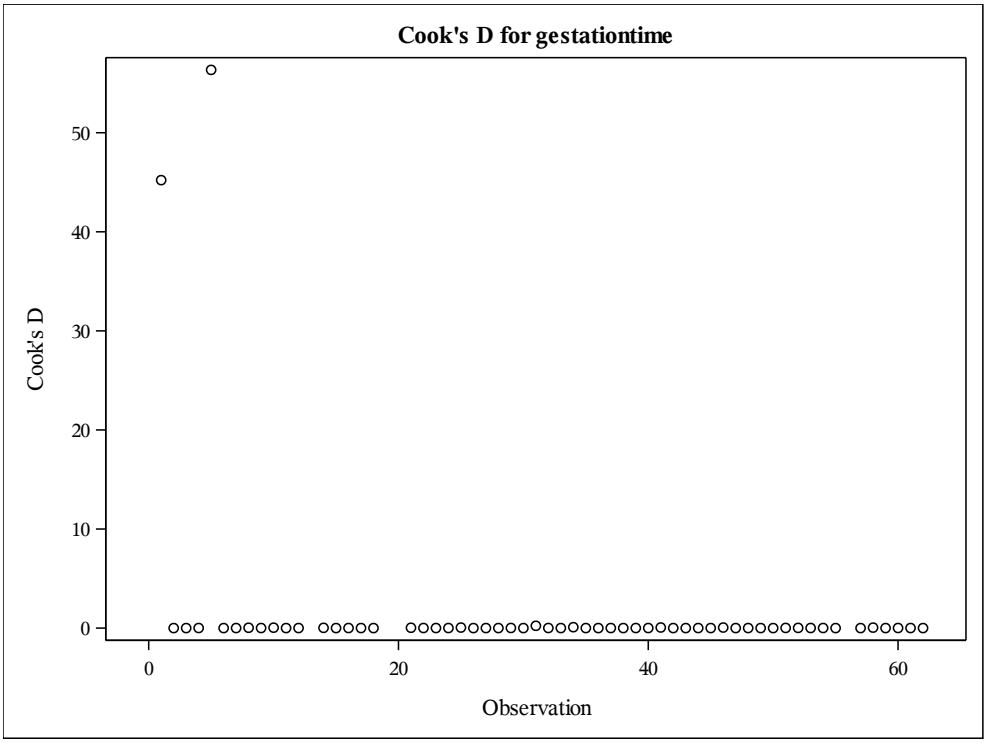
Exercise 3

a. Note that we observe overdispersion since the scale value (scaled deviance 53.101) is much larger than 1 when all 4 predictors are in the model. Hence we should use an overdispersed Poisson model.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	53	2814.3622	53.1012
Scaled Deviance	53	2814.3622	53.1012
Pearson Chi-Square	53	2895.6702	54.6353
Scaled Pearson X2	53	2895.6702	54.6353
Log Likelihood		34922.1124	
Full Log Likelihood		-1589.4546	
AIC (smaller is better)		3188.9092	
AICC (smaller is better)		3190.0631	
BIC (smaller is better)		3199.2115	

When an overdispersed Poisson log-linear model is fitted using the four predictors we needed to do several model fittings and re-fittings due to influential points. The Cook's Distance plot as well as data print outs of the Cook's D values guided us to remove 3 observations (5, 1, and 32) one by one due to influence.

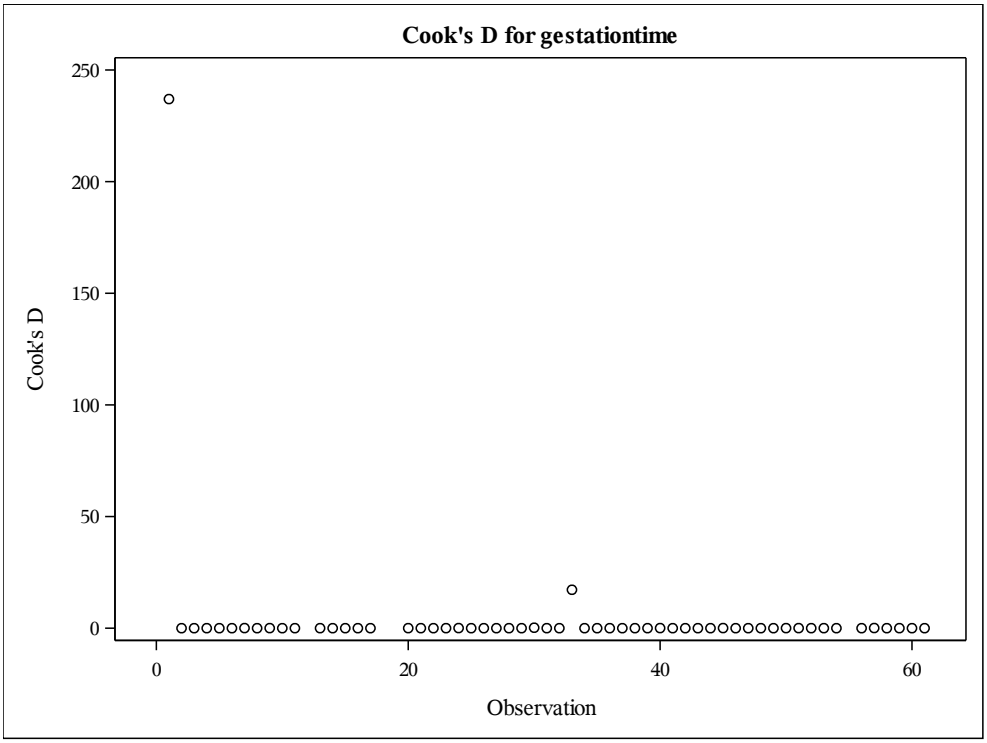
Iteration 1 (Remove Obs 5 with largest Cook’s Distance)



Obs	species	bodyweight	brainweight	totalsleep	maxlifespan	gestationtime	predationindex
1	African	6654	5712	3.3	38.6	645	3
5	Asian el	2547	4603	3.9	69.0	624	3

Obs	sleepexposureindex	overalldangerindex	maxlife10	cdII
1	5	3	0	45.2159
5	5	4	0	56.3444

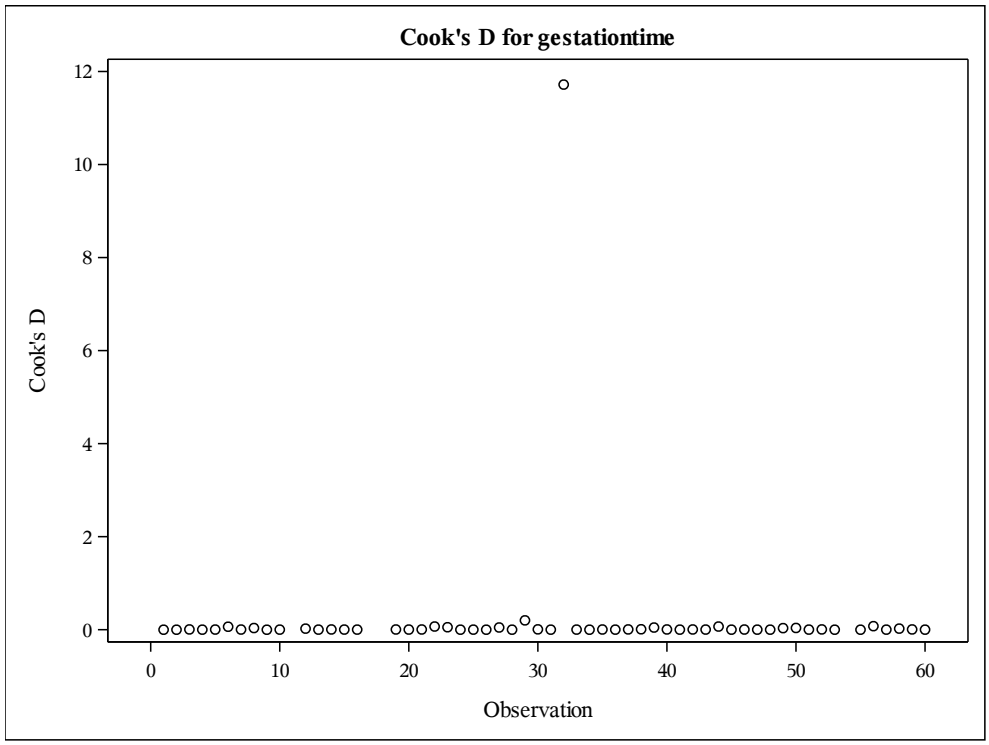
Iteration 2 (Remove Obs 1)



Obs	species	bodyweight	brainweight	totalsleep	maxlifespan	gestationtime	predationindex
1	African	6654	5712	3.3	38.6	645	3
33	Man	62	1320	8.0	100.0	267	1

Obs	sleepexposureindex	overalldangerindex	maxlife10	cdII	cdIII
1	5	3	0	45.2159	236.982
33	1	1	0	0.0918	17.209

Iteration 3 (Remove Obs 32)



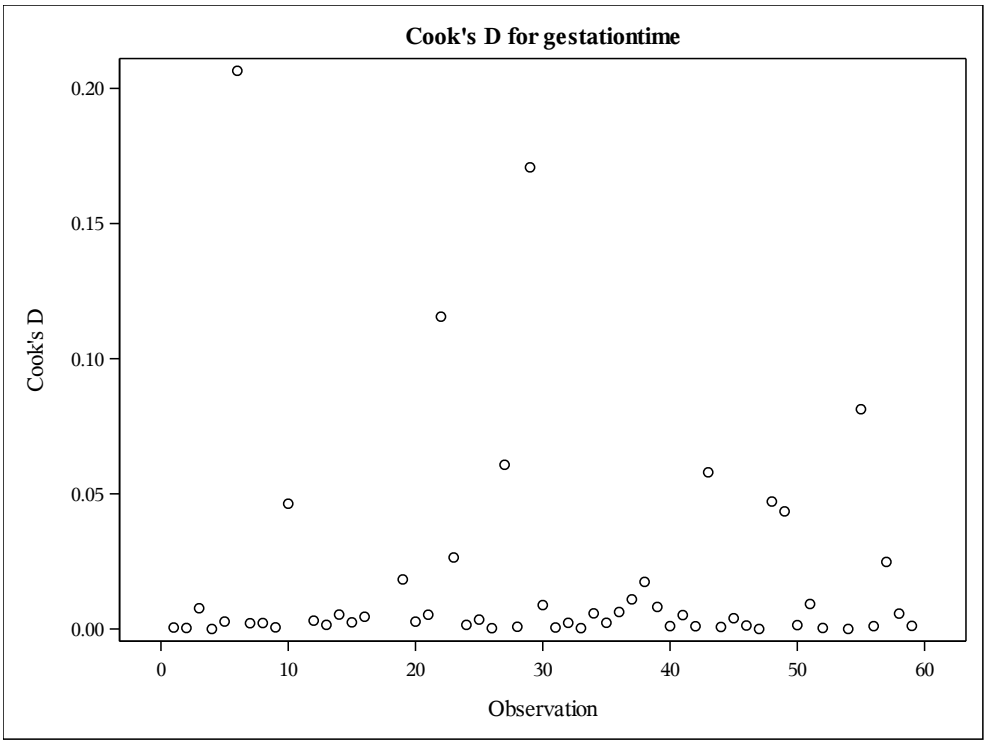
Obs	species	bodyweight	brainweight	totalsleep	maxlifespan	gestationtime	predationindex
32	Man	62	1320	8	100	267	1

Obs	sleepexposureindex	overalldangerindex	maxlife10	cdII	cdIII	cdIV
32	1	1	0	0.091767	17.2087	11.7191

After removing those influential points, we refit the model and choose the best set of predictors. To do this we should fit several competing models and compare their information criterion values (AIC, AICC, BIC). The two best models are: the 4 predictor model with BIC value of 2212.596 and the 3 predictor model (bodyweight, brainweight, sleepexposureindex) with BIC value of 2256.158. We would choose the 3 predictor model since its BIC is not much different from the full model and is simpler (parsimonious). Based on type 1 and type 3 analysis results, we can keep these 3 significant (at 5% level) predictors.

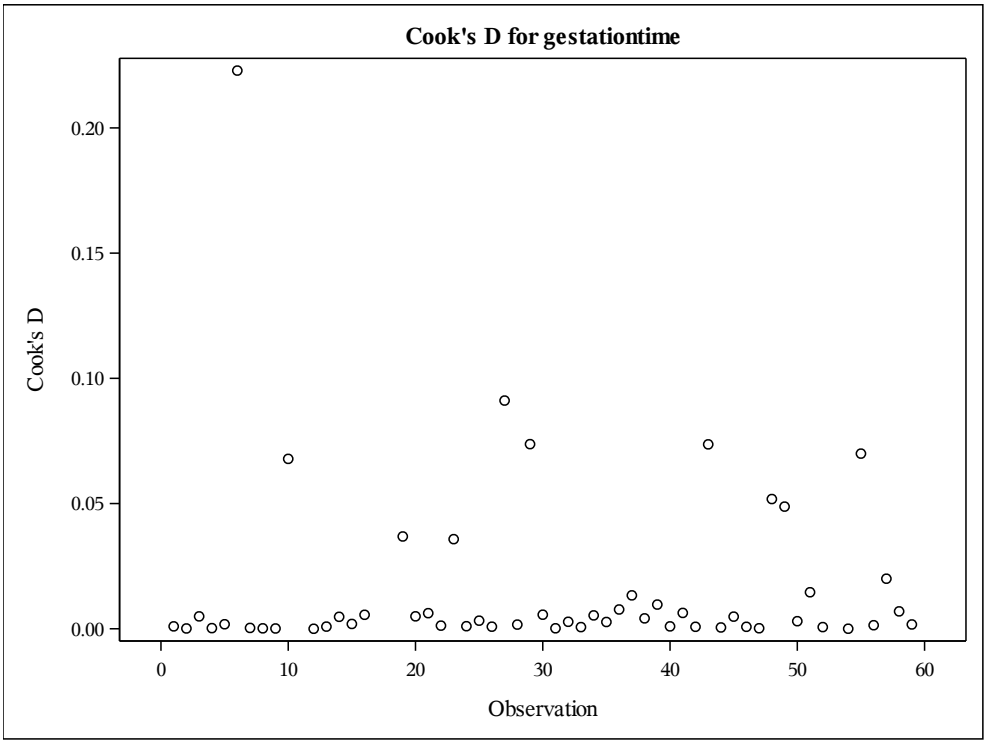
Model with 4 predictors (bodyweight, brainweight, predationindex, sleepexposureindex)

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	50	1852.0195	37.0404
Scaled Deviance	50	50.0000	1.0000
Pearson Chi-Square	50	2159.6427	43.1929
Scaled Pearson X2	50	58.3051	1.1661
Log Likelihood		735.9173	
Full Log Likelihood		-1096.2795	
AIC (smaller is better)		2202.5591	
AICC (smaller is better)		2203.7836	
BIC (smaller is better)		2212.5957	

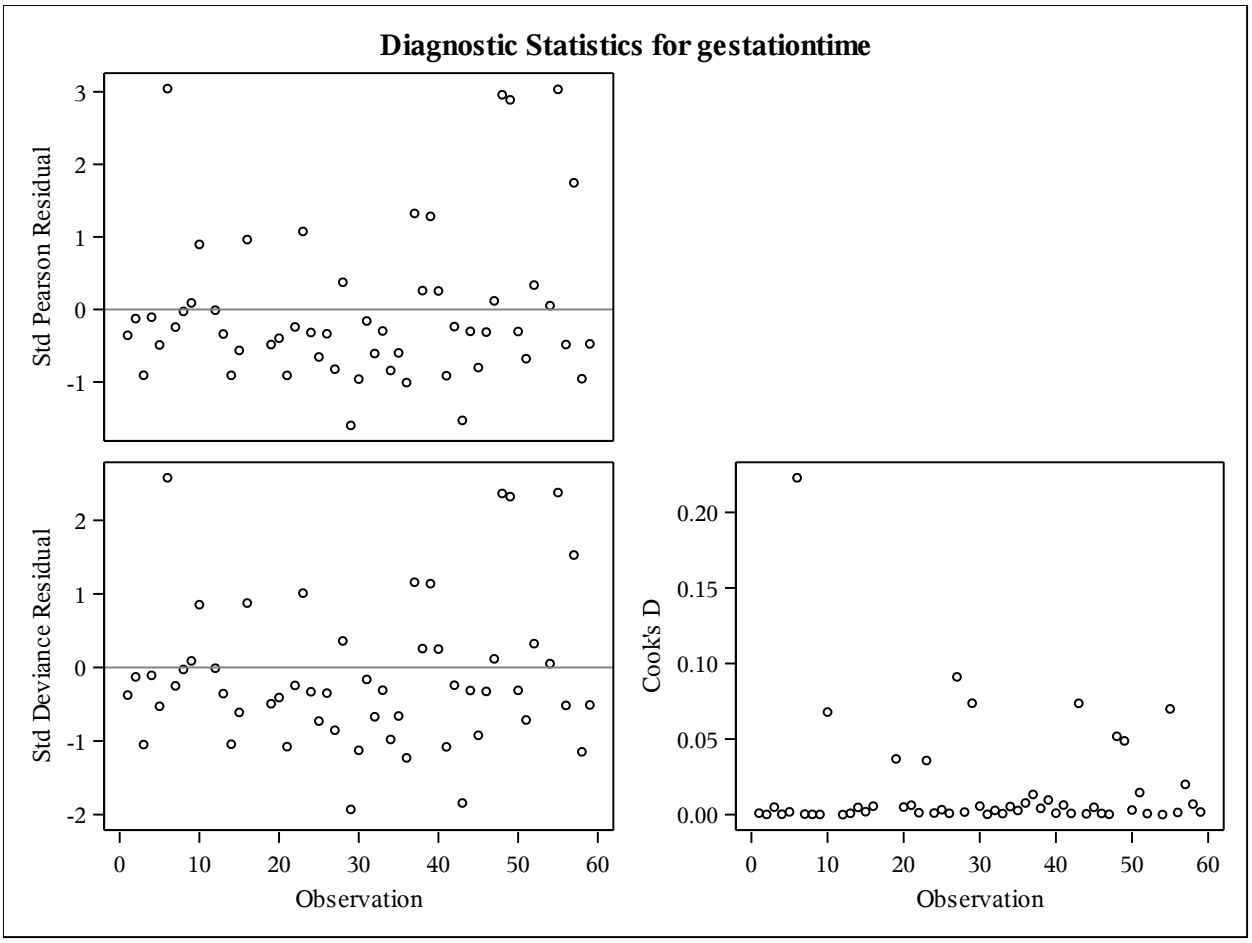


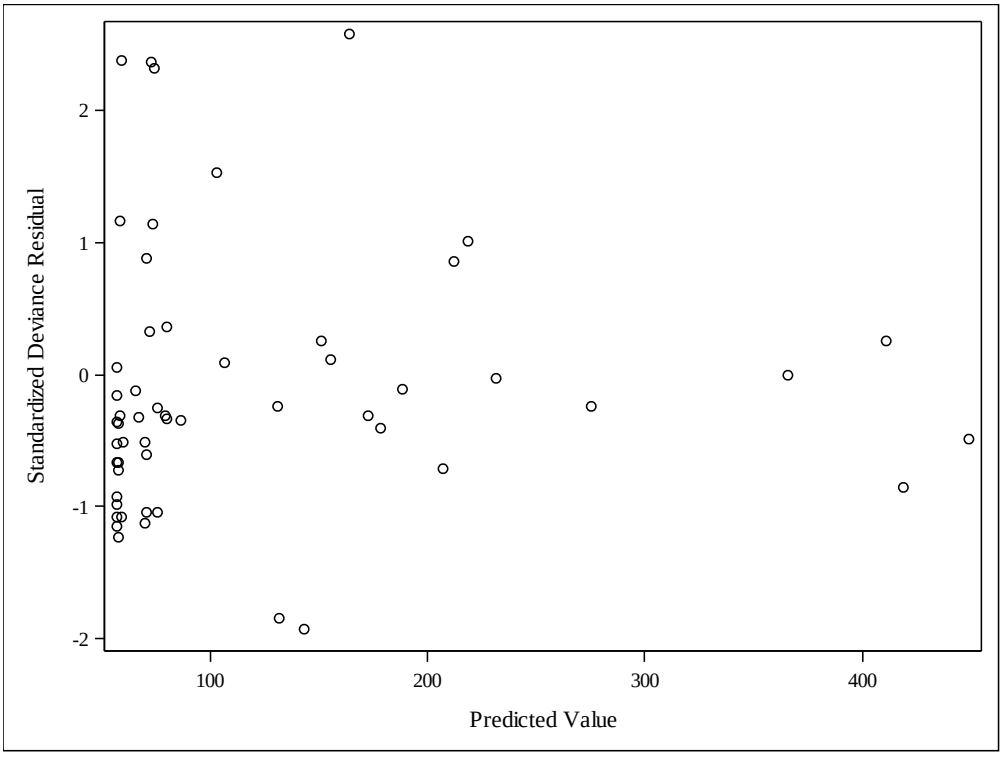
Model with 3 predictors (bodyweight, brainweight, sleepexposureindex)

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	51	1899.5890	37.2468
Scaled Deviance	51	51.0000	1.0000
Pearson Chi-Square	51	2205.1697	43.2386
Scaled Pearson X2	51	59.2042	1.1609
Log Likelihood		731.1997	
Full Log Likelihood		-1120.0643	
AIC (smaller is better)		2248.1286	
AICC (smaller is better)		2248.9286	
BIC (smaller is better)		2256.1579	



b. We fit the log-linear model for gestation time using bodyweight, brainweight and sleepexposureindex as predictors. From the Cook's Distance and residuals plots, we do not have any influential observations of concern. There is no need to remove any further points. We can also see that there is no trend in the residuals vs. predicted values, so the distribution and link choices are reasonable.


Diagnostic Statistics for gestationtime
Std Pearson Residual
Std Deviance Residual
Cook's D
Observation
Observation


Standardized Deviance Residual
Predicted Value

c. From type 1 and type 3 analysis tables, we can see that all predictors are significant (at 5% level) in the model. The parameter estimates are -0.002, 0.0034 and 0.200 for bodyweight, brainweight and sleepexposureindex respectively. The bodyweight has negative value so it implies that there's an expected decrease in log gestation time for a one-unit increase in bodyweight is -0.002 while holding the other predictors constant. The expected gestation time would decrease by exp(-0.002)=0.998 times. The remaining parameter estimates are positive indicating an expected increase in log gestation time. For a one-unit increase in brainweight, the expected log count of gestation time increases by .0034 (multiplicative increase of exp(0.0034)=1.003 in gestation time) holding all other predictors constant. For a one-unit increase in sleepexplosure index, the expected log count of gestation time increases in by 0.200 (multiplicative increase of exp(0.200)=1.22) holding all other predictors constant. There are no other concerns in the diagnostics.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	3.8464	0.1790	3.4955	4.1972	461.70	<.0001
bodyweight	1	-0.0020	0.0010	-0.0039	-0.0001	4.31	0.0379
brainweight	1	0.0034	0.0008	0.0019	0.0049	20.46	<.0001
sleepexposureindex	1	0.2001	0.0573	0.0878	0.3124	12.19	0.0005
Scale	0	6.1030	0.0000	6.1030	6.1030		

LR Statistics For Type 1 Analysis							
Source	**Deviance**	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
Intercept	5218.4845						
bodyweight	3277.2063	1	51	52.12	<.0001	52.12	<.0001
brainweight	2343.3497	1	51	25.07	<.0001	25.07	<.0001
sleepexposureindex	1899.5890	1	51	11.91	0.0011	11.91	0.0006

LR Statistics For Type 3 Analysis						
Source	**Num DF**	**Den DF**	**F Value**	**Pr > F**	**Chi-Square**	**Pr > ChiSq**
bodyweight	1	51	4.26	0.0442	4.26	0.0391
brainweight	1	51	18.17	<.0001	18.17	<.0001
sleepexposureindex	1	51	11.91	0.0011	11.91	0.0006