

Stat 448, homework 4
Shuhui Guo

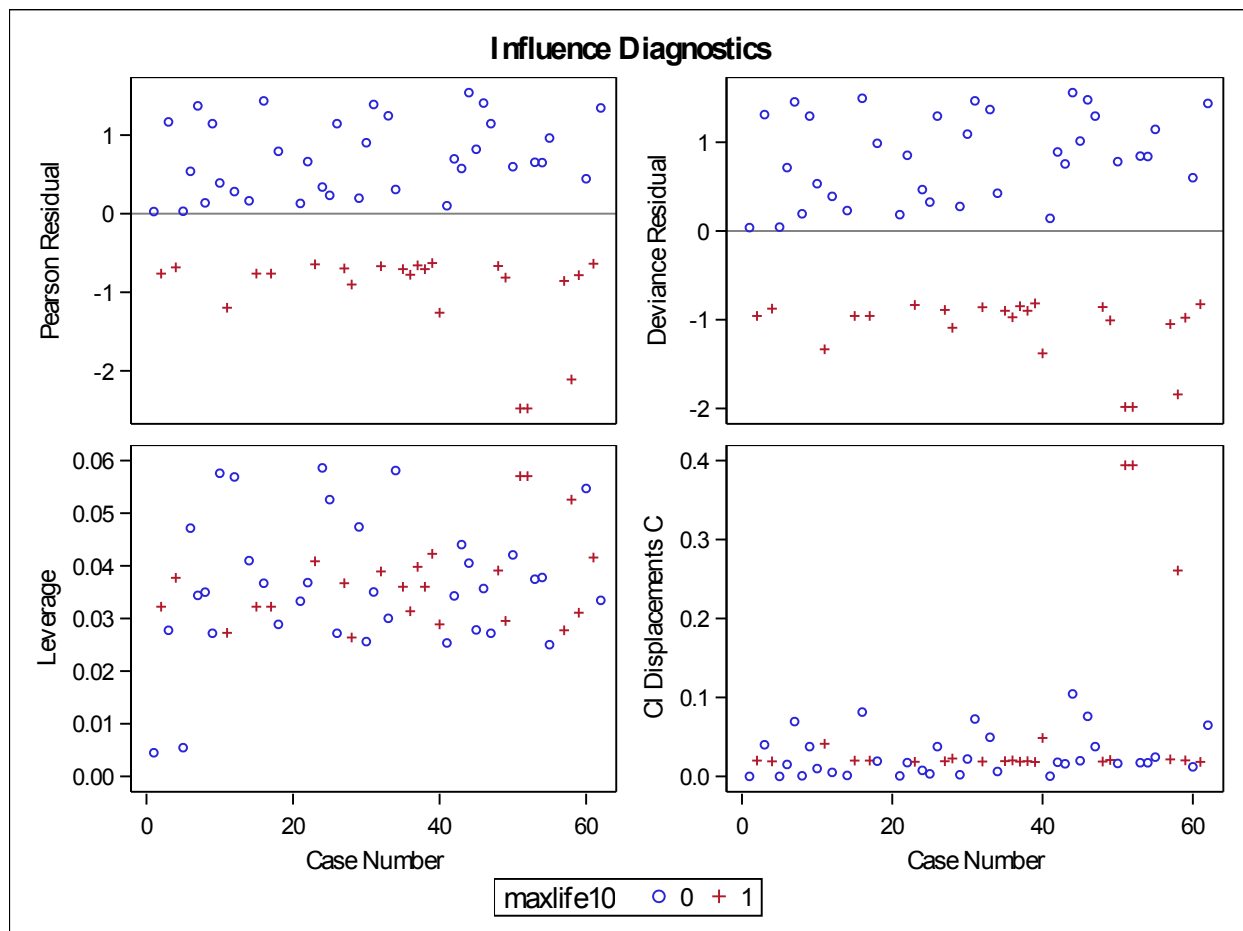
Exercise 1

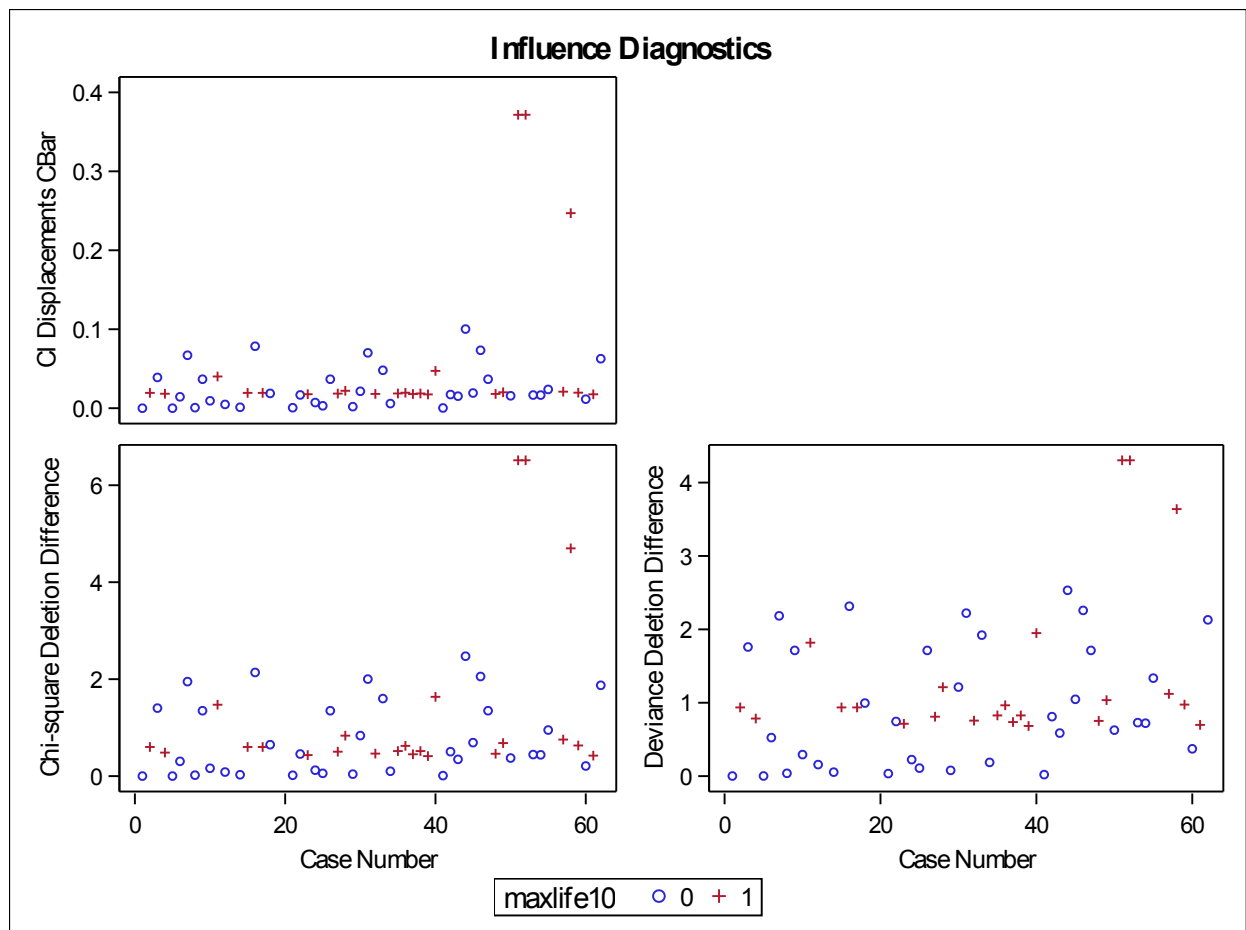
a)

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

According to the above table, the best set of predictors for the model has just one variable: gestationtime.

The influence diagnostics are as below:





According to the above plots, the highest Cbar value is not greater than 0.4, which is in a relatively lower level. Therefore, there is no unduly influential point. So there is no need for refitting the model. The final model will have only one predictor: gestationtime.

b)

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	79.904	66.554
SC	81.965	70.675
-2 Log L	77.904	62.554

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	15.3506	1	<.0001
Score	10.7615	1	0.0010
Wald	8.2827	1	0.0040

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

Based on the above tables, the p-values of Likelihood ratio test, score test and Wald test are all less than 0.05, which indicate that the model is significant. For the estimation of gestationtime, the p-value of the Wald Chi-square test is $0.0040 < 0.05$, which indicates that the estimation of gestationtime is significant. The results of Wald Chi-square test for gestationtime and Wald test for the overall model are the same since there is only one variable in the model. In Hosmer-Lemeshow's test, the p-value is $0.4072 > 0.05$, which does not reject the null hypothesis so we conclude that there is no lack of fit in this model.

There is a remaining issue in the final model. According to the influence diagnostics, there are certain observations with absolute value of Pearson residual greater than 2, which might be high influential points. Nevertheless, there is no obvious trend in residuals and the cook's distance (which refers to Cbar in the plots) does not appear that these points are extremely unduly influential.

c)

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

Based on the above table, the confidence interval of odds ratio estimation does not include 1, which means that the estimation of gestationtime is significant. The estimation of odds ratio for gestationtime is 1.012. Thus for each unit increasing of gestation time, there will be $\exp(1.012) = 2.751$ times increasing of odds of a specie's maximum lifespan being at least 10 years.

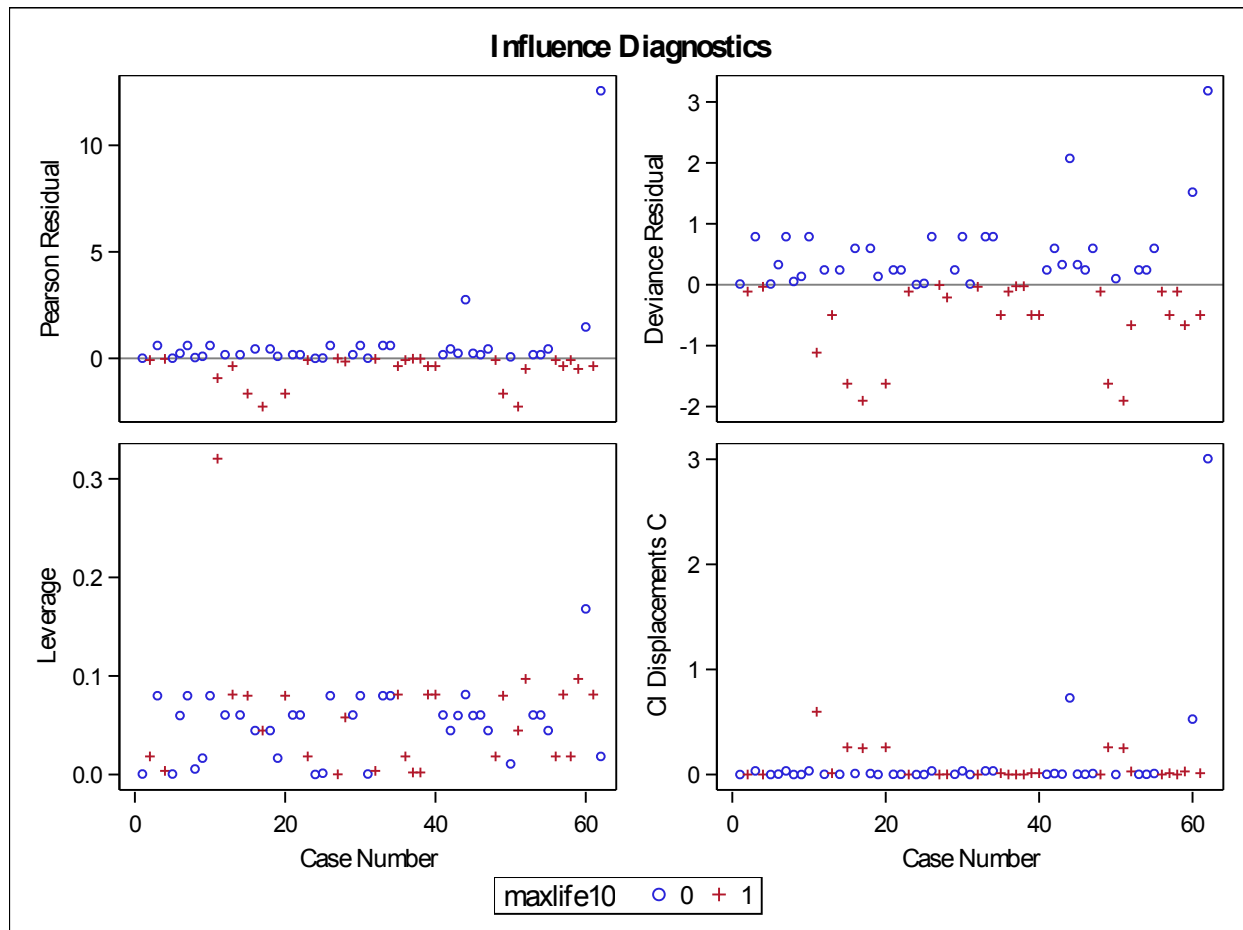
Exercise 2

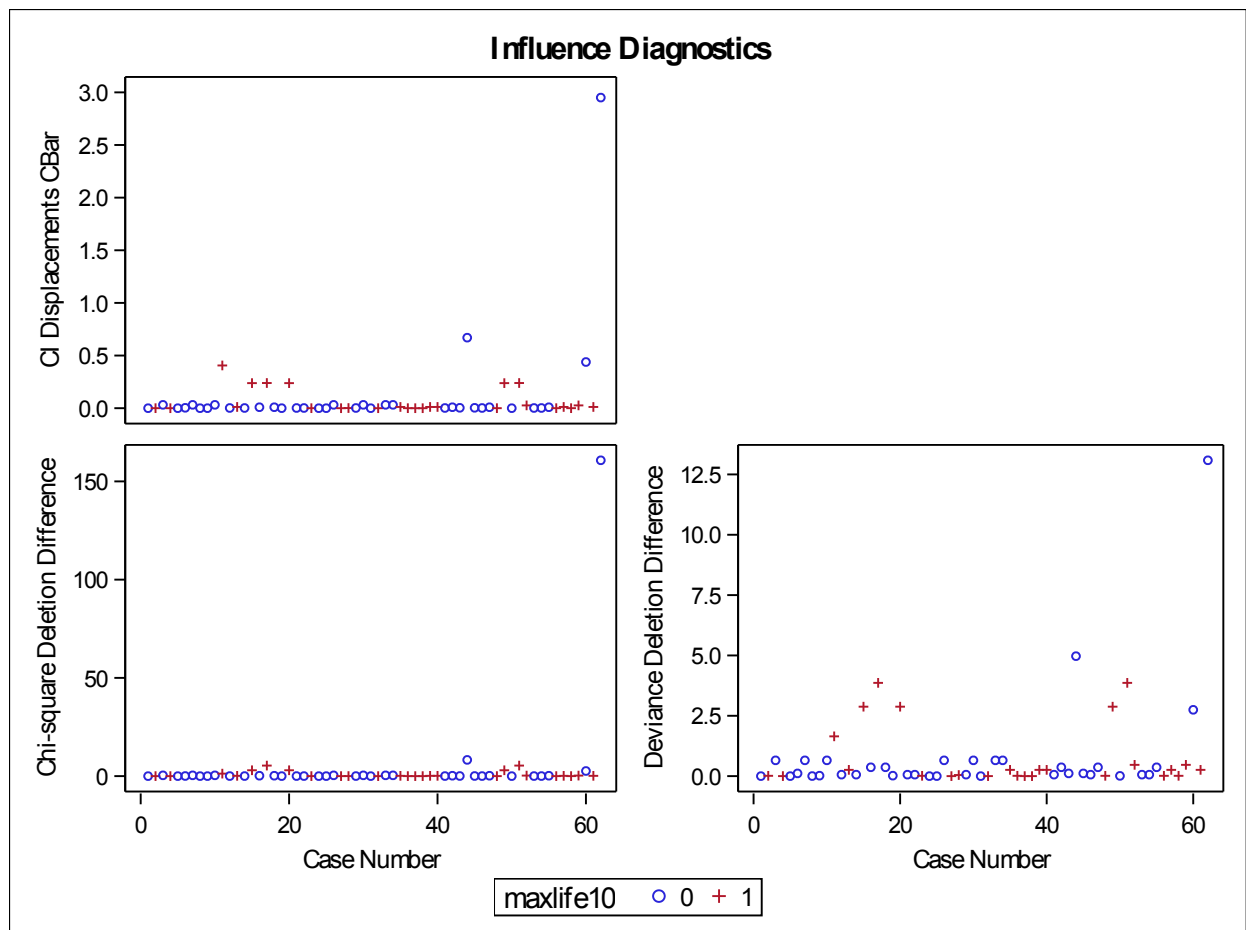
a)

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

According to the above table, the best set of predictors for the model has two variables: sleepexposureindex and predationindex.

The influence diagnostics are as below:





Based on the above plots, the highest Cbar value is around 3.0, which is relatively high and should be treated as an unduly influential point. Therefore, the model should be refitted without this high influential point.

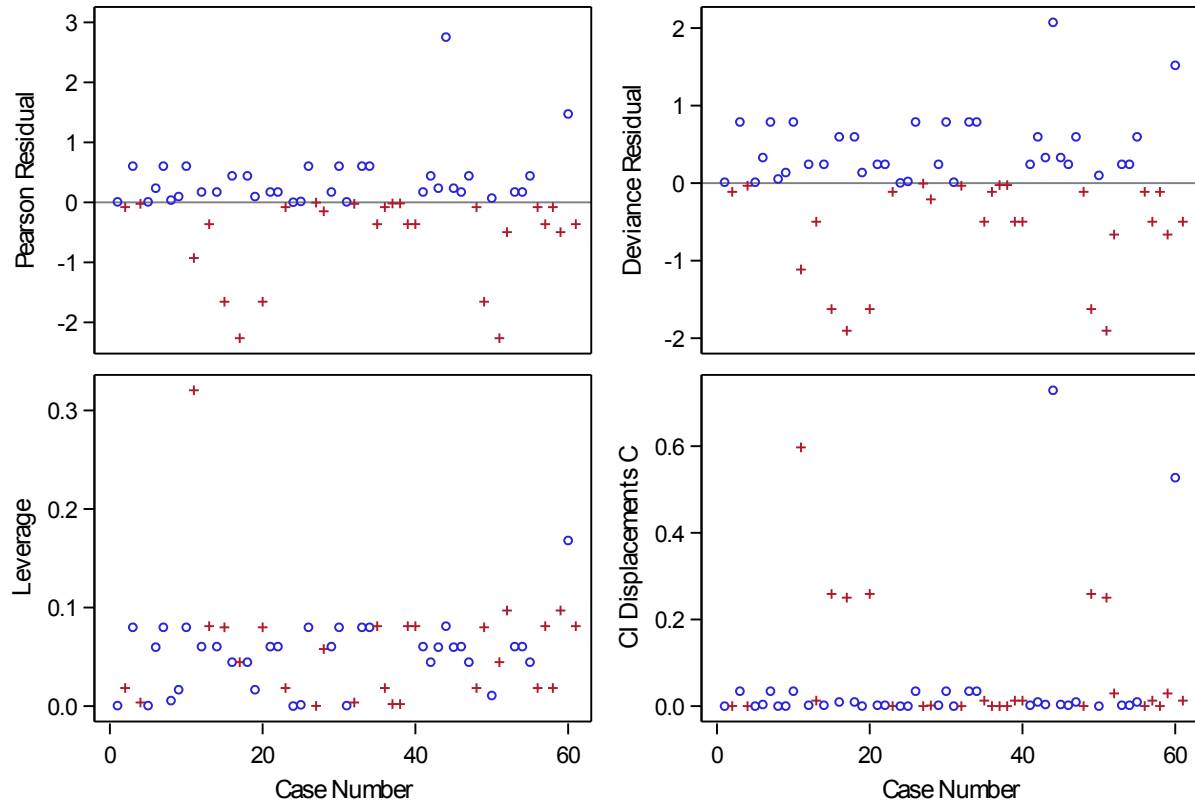
The result after refitting is as below:

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

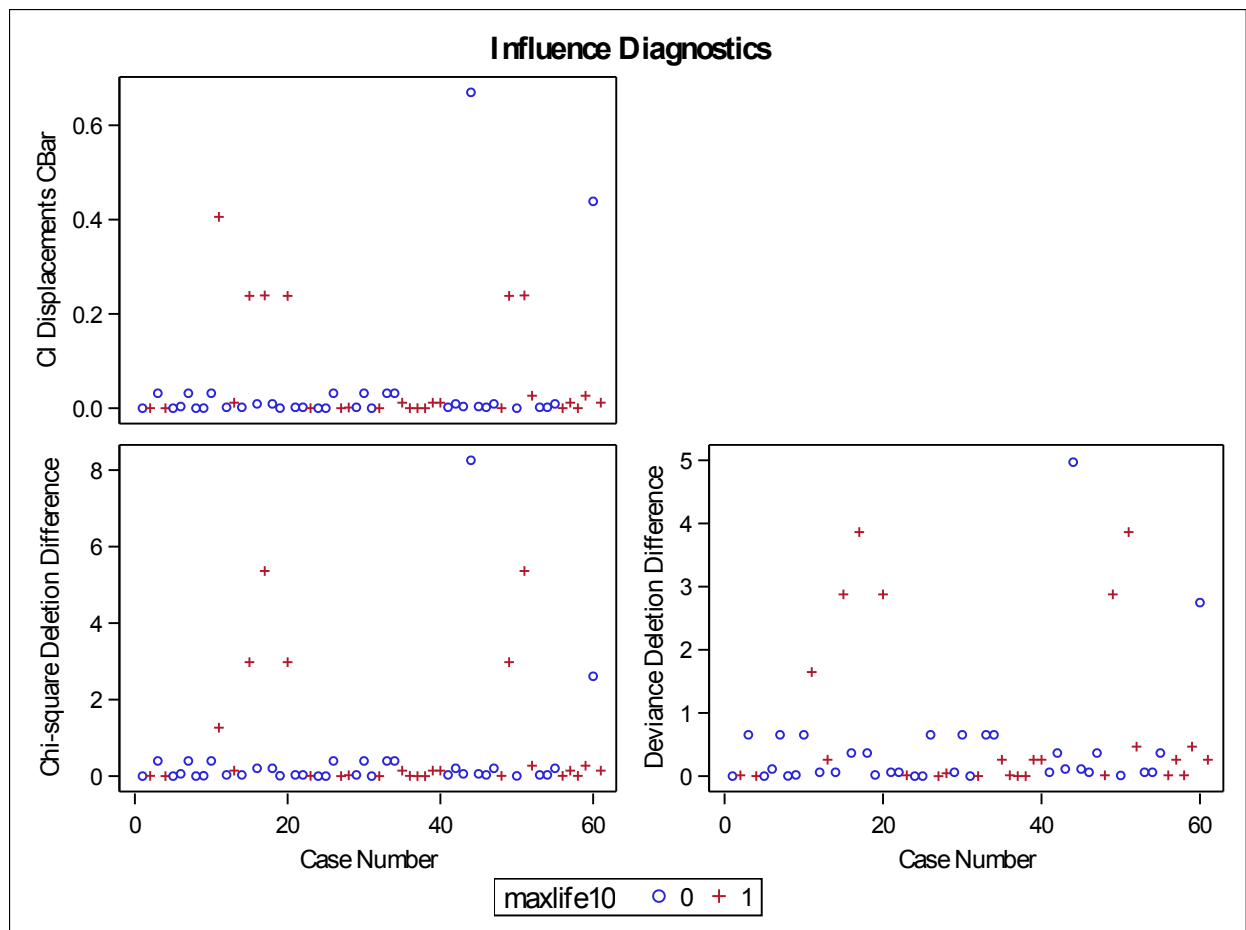
According to the above table, the best set of predictors for the model has two variables: sleepexposureindex and predationindex.

The influence diagnostics are as below:

Influence Diagnostics



maxlife10 0 + 1



For the refitted model, the highest Cbar value is around 0.7, which is relatively low. Therefore, there is no unduly influential point. The final model will have two predictors: sleepexposureindex and predationindex.

b)

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	85.231	38.305
SC	87.342	44.638
-2 Log L	83.231	32.305

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	50.9260	2	<.0001
Score	32.5994	2	<.0001
Wald	12.3609	2	0.0021

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.0256	0.8524	0.0009	0.9760
predationindex	1	-2.9882	0.9204	10.5402	0.0012
sleepexposureindex	1	3.7457	1.0654	12.3605	0.0004

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
1.1307	7	0.9924

Based on the above tables, the p-values of Likelihood ratio test, score test and Wald test are all less than 0.05, which indicate that the model is significant. For the estimations of sleepexposureindex and predationindex, the p-values of the Wald Chi-square test are both less than 0.05, which indicates that the estimations of the two predictors are significant. In Hosmer-Lemeshow's test, the p-value is 0.9924>0.05, which does not reject the null hypothesis so we conclude that there is no lack of fit in this model.

There is a remaining issue in the final model. According to the influence diagnostics, there are certain observations with absolute value of Pearson residual greater than 2, which might be high influential points. Nevertheless, there is no obvious trend in residuals and the cook's distance (which refers to Cbar in the plots) does not appear that these points are extremely unduly influential.

c)

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
predationindex	0.050	0.008	0.306
sleepexposureindex	42.338	5.246	341.662

Based on the above table, the confidence intervals of odds ratio estimation do not include 1 for both of the predictors, which means that the estimations of the two predictors are significant. The estimation of odds ratio for predationindex is 0.050. Thus for each unit increasing of predation index, the odds of a species' maximum lifespan being at least 10 years will change by a factor of $\exp(-0.950)=0.387$. The estimation of odds ratio for sleepexposureindex is 42.338. Thus for each unit increasing of sleep exposure index, there

will be $\exp(42.338)=2.439 \times 10^{18}$ times increasing of odds of a species' maximum lifespan being at least 10 years.

Exercise 3

a)

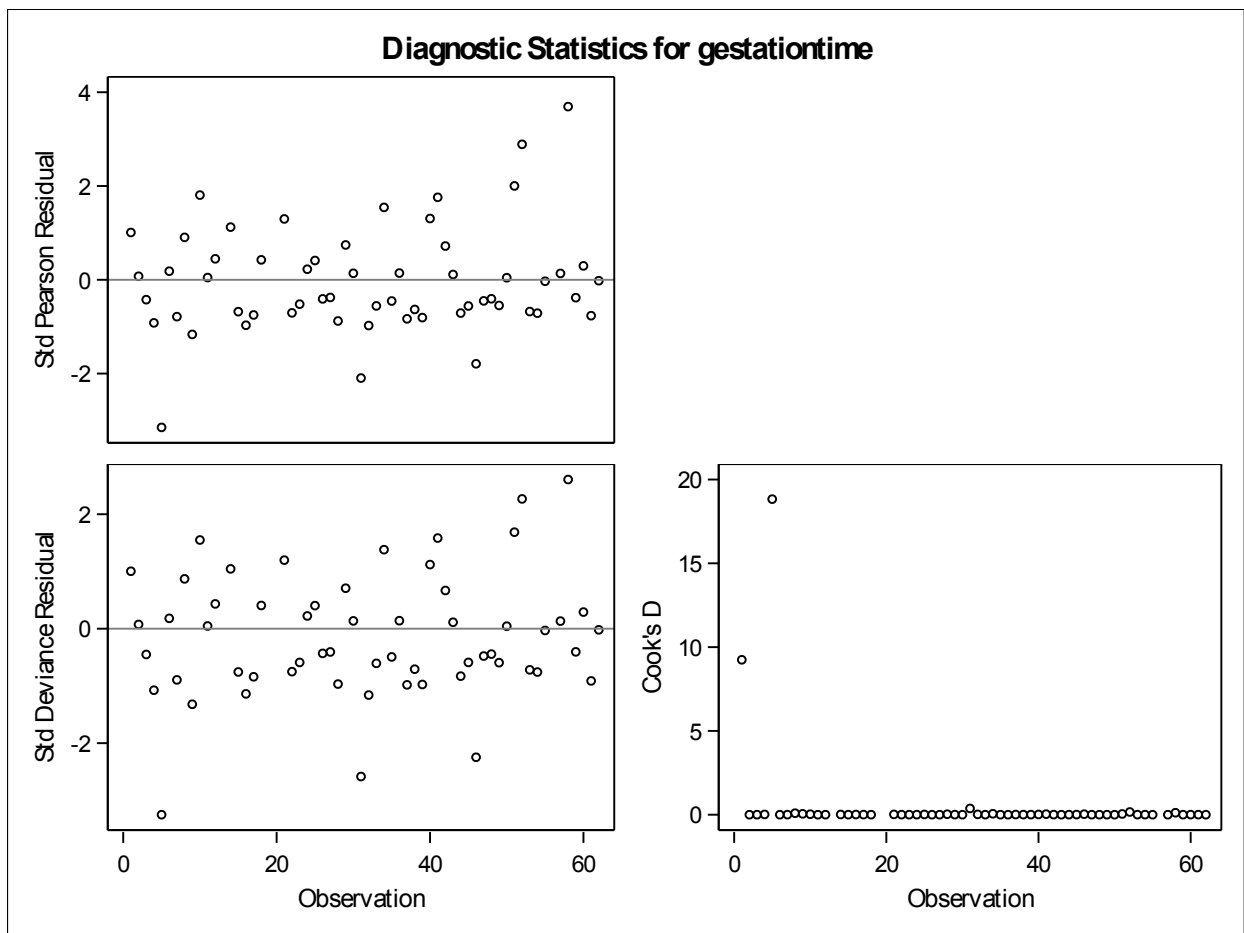
First the full model is fitted as below:

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	5.3562	0.1595	5.0435	5.6688	1127.49	<.0001
sleepexposureindex	1	1	-1.5334	0.3468	-2.2130	-0.8537	19.55	<.0001
sleepexposureindex	2	1	-0.9866	0.3725	-1.7167	-0.2564	7.01	0.0081
sleepexposureindex	3	1	-0.4949	0.3648	-1.2099	0.2202	1.84	0.1749
sleepexposureindex	4	1	-0.6686	0.3578	-1.3698	0.0326	3.49	0.0617
sleepexposureindex	5	0	0.0000	0.0000	0.0000	0.0000	.	.
brainweight		1	0.0004	0.0001	0.0001	0.0007	7.15	0.0075
bodyweight		1	-0.0002	0.0001	-0.0004	0.0001	1.53	0.2162
predationindex	1	1	0.6375	0.3479	-0.0443	1.3193	3.36	0.0669
predationindex	2	1	0.1481	0.3694	-0.5760	0.8721	0.16	0.6886
predationindex	3	1	-0.1691	0.3666	-0.8876	0.5494	0.21	0.6445
predationindex	4	1	0.3479	0.3135	-0.2666	0.9625	1.23	0.2671
predationindex	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	7.3351	0.0000	7.3351	7.3351		

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Intercept	7294.9639						
sleepexposureindex	3884.6633	4	47	15.85	<.0001	63.38	<.0001
brainweight	2955.8727	1	47	17.26	0.0001	17.26	<.0001
bodyweight	2861.7996	1	47	1.75	0.1925	1.75	0.1861
predationindex	2528.7415	4	47	1.55	0.2039	6.19	0.1854

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
sleepexposureindex	4	47	5.52	0.0010	22.08	0.0002
brainweight	1	47	6.75	0.0125	6.75	0.0094
bodyweight	1	47	1.51	0.2259	1.51	0.2198
predationindex	4	47	1.55	0.2039	6.19	0.1854

According to the above tables, sleepexposureindex and brainweight are significant under the significance level of 5%. On the contrast, bodyweight and predationindex are insignificant because the p-values of them are more than 0.05. Nevertheless, before proceeding with model selection, the diagnostics should be checked.



According to the above plot, there are certain high influential points. Also, the points with Cook's distance greater than 1 are printed as below:

Obs	species	bodyweight	brainweight	total leaps	max lifespan	gestation time	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	predbp3	schires3	cd3
1	5	3	0	626.567	1.00865	9.2436
5	5	4	0	760.037	-3.14816	18.8305

Since the dataset is large, the points with Cook's distance greater than 1 are removed at once and the model is refitted as below:

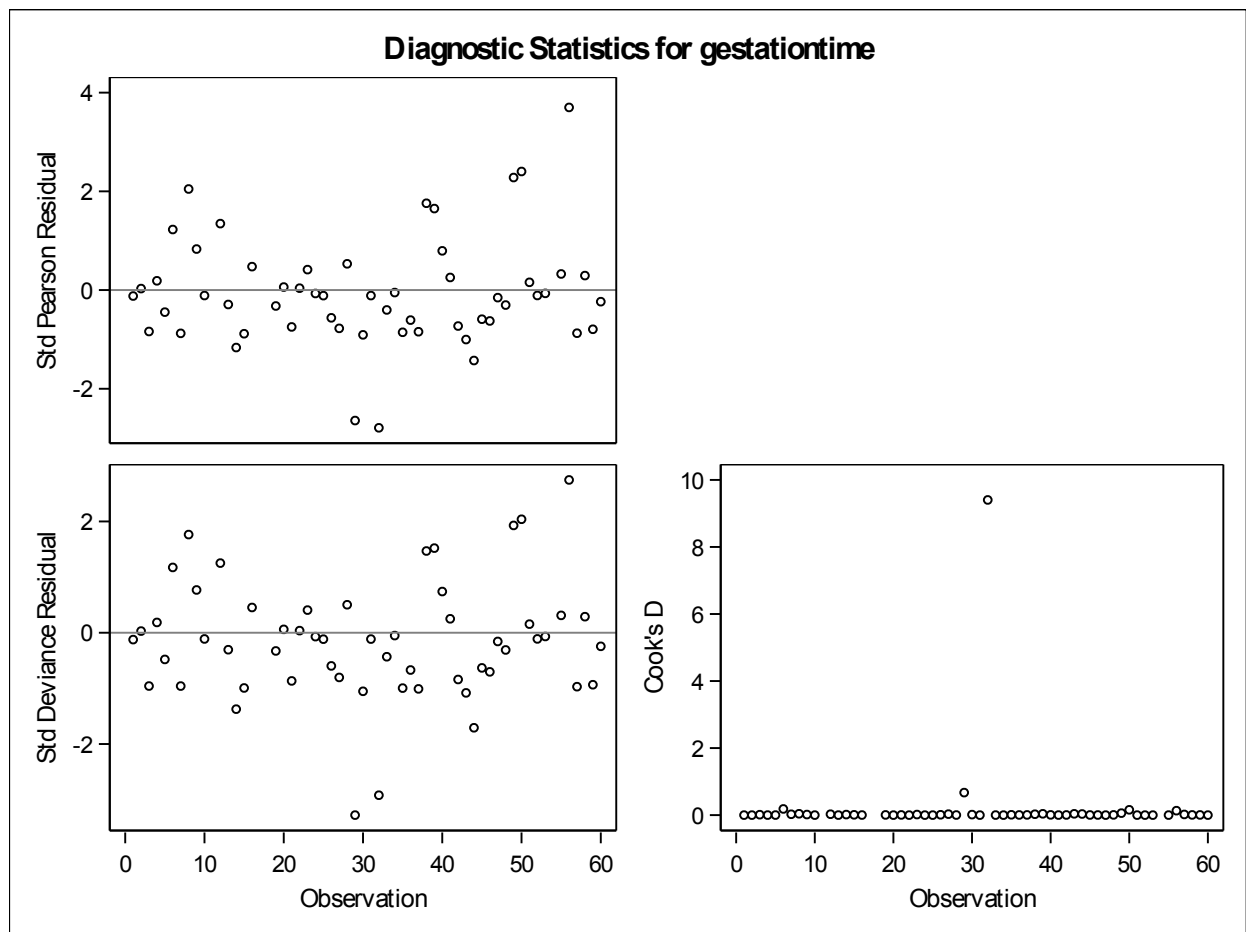
Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	4.7915	0.2245	4.3515	5.2316	455.38	<.0001
sleepexposureindex	1	1	-1.3356	0.3292	-1.9808	-0.6905	16.46	<.0001
sleepexposureindex	2	1	-0.6491	0.3498	-1.3347	0.0364	3.44	0.0635
sleepexposureindex	3	1	-0.2128	0.3448	-0.8886	0.4629	0.38	0.5370
sleepexposureindex	4	1	-0.5180	0.3263	-1.1576	0.1215	2.52	0.1124
sleepexposureindex	5	0	0.0000	0.0000	0.0000	0.0000	.	.
brainweight		1	0.0014	0.0003	0.0007	0.0020	17.29	<.0001
bodyweight		1	0.0006	0.0007	-0.0007	0.0020	0.88	0.3490
predationindex	1	1	0.5515	0.3224	-0.0803	1.1833	2.93	0.0871
predationindex	2	1	0.3812	0.3357	-0.2767	1.0391	1.29	0.2561
predationindex	3	1	0.3839	0.3336	-0.2698	1.0377	1.32	0.2497
predationindex	4	1	0.5889	0.2907	0.0190	1.1587	4.10	0.0428
predationindex	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	6.3035	0.0000	6.3035	6.3035		

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Intercept	5343.2317						
sleepexposureindex	3217.7475	4	45	13.37	<.0001	53.49	<.0001
brainweight	1989.3696	1	45	30.91	<.0001	30.91	<.0001

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
bodyweight	1977.3074	1	45	0.30	0.5844	0.30	0.5817
predationindex	1788.0570	4	45	1.19	0.3279	4.76	0.3125

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
sleepexposureindex	4	45	5.70	0.0008	22.80	0.0001
brainweight	1	45	13.91	0.0005	13.91	0.0002
bodyweight	1	45	0.88	0.3527	0.88	0.3477
predationindex	4	45	1.19	0.3279	4.76	0.3125

According to the above tables, sleepexposureindex and brainweight are significant under the significance level of 5%. On the contrast, bodyweight and predationindex are insignificant because the p-values of them are more than 0.05. Nevertheless, before proceeding with model selection, the diagnostics should be checked.



According to the above plot, there are still certain high influential points. Also, the points with Cook's distance greater than 1 are printed as below:

Obs	species	bodyweight	brainweight	total sleep	maxlife span	gestation time	predation index	sleepexposureindex
32	Man	62	1320	8	100	267	1	1

Obs	overalldangerindex	maxlife10	predbp3	schires3	cd3	predbp	schires	cd
32	1	0	145.282	1.54389	0.055820	354.693	-2.79110	9.40295

After removing the above point, the model is refitted as below:

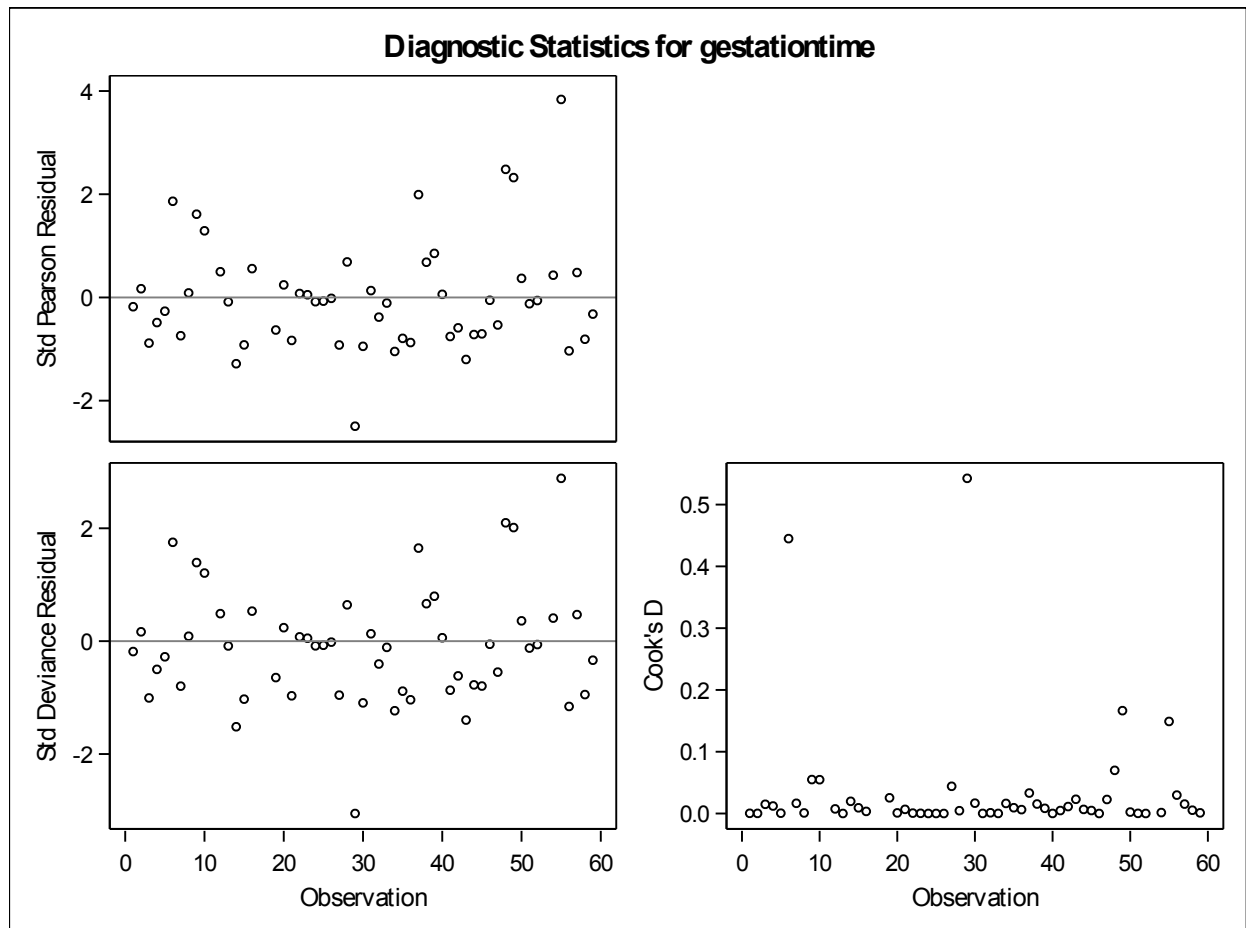
Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	4.4958	0.2380	4.0293	4.9623	356.77	<.0001
sleepexposureindex	1	1	-1.1033	0.3208	-1.7320	-0.4745	11.83	0.0006
sleepexposureindex	2	1	-0.3908	0.3431	-1.0633	0.2816	1.30	0.2547
sleepexposureindex	3	1	-0.2850	0.3329	-0.9376	0.3675	0.73	0.3919
sleepexposureindex	4	1	-0.5823	0.3158	-1.2012	0.0366	3.40	0.0652
sleepexposureindex	5	0	0.0000	0.0000	0.0000	0.0000	.	.
brainweight		1	0.0039	0.0009	0.0021	0.0056	18.98	<.0001
bodyweight		1	-0.0019	0.0011	-0.0040	0.0002	3.19	0.0741
predationindex	1	1	0.4132	0.3121	-0.1986	1.0249	1.75	0.1856
predationindex	2	1	0.3834	0.3210	-0.2457	1.0125	1.43	0.2322
predationindex	3	1	0.4747	0.3219	-0.1562	1.1056	2.17	0.1403
predationindex	4	1	0.7710	0.2910	0.2006	1.3413	7.02	0.0081
predationindex	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	5.8319	0.0000	5.8319	5.8319		

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Intercept	5218.4845						
sleepexposureindex	2847.5407	4	44	17.43	<.0001	69.71	<.0001
brainweight	1843.3132	1	44	29.53	<.0001	29.53	<.0001

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
bodyweight	1738.3751	1	44	3.09	0.0860	3.09	0.0790
predationindex	1496.4925	4	44	1.78	0.1504	7.11	0.1301

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
sleepexposureindex	4	44	4.23	0.0055	16.94	0.0020
brainweight	1	44	17.89	0.0001	17.89	<.0001
bodyweight	1	44	3.15	0.0830	3.15	0.0761
predationindex	4	44	1.78	0.1504	7.11	0.1301

According to the above results, sleepexposureindex and brainweight are significant under the significance level of 5%. On the contrast, bodyweight and predationindex are insignificant because the p-values of them are more than 0.05, which indicates that the two variables should be removed from the model.



According to the residual plot, there is no obvious trend in residuals. Also, there are no points with Cook's distance greater than 1. Therefore, the model selection can be proceeded.

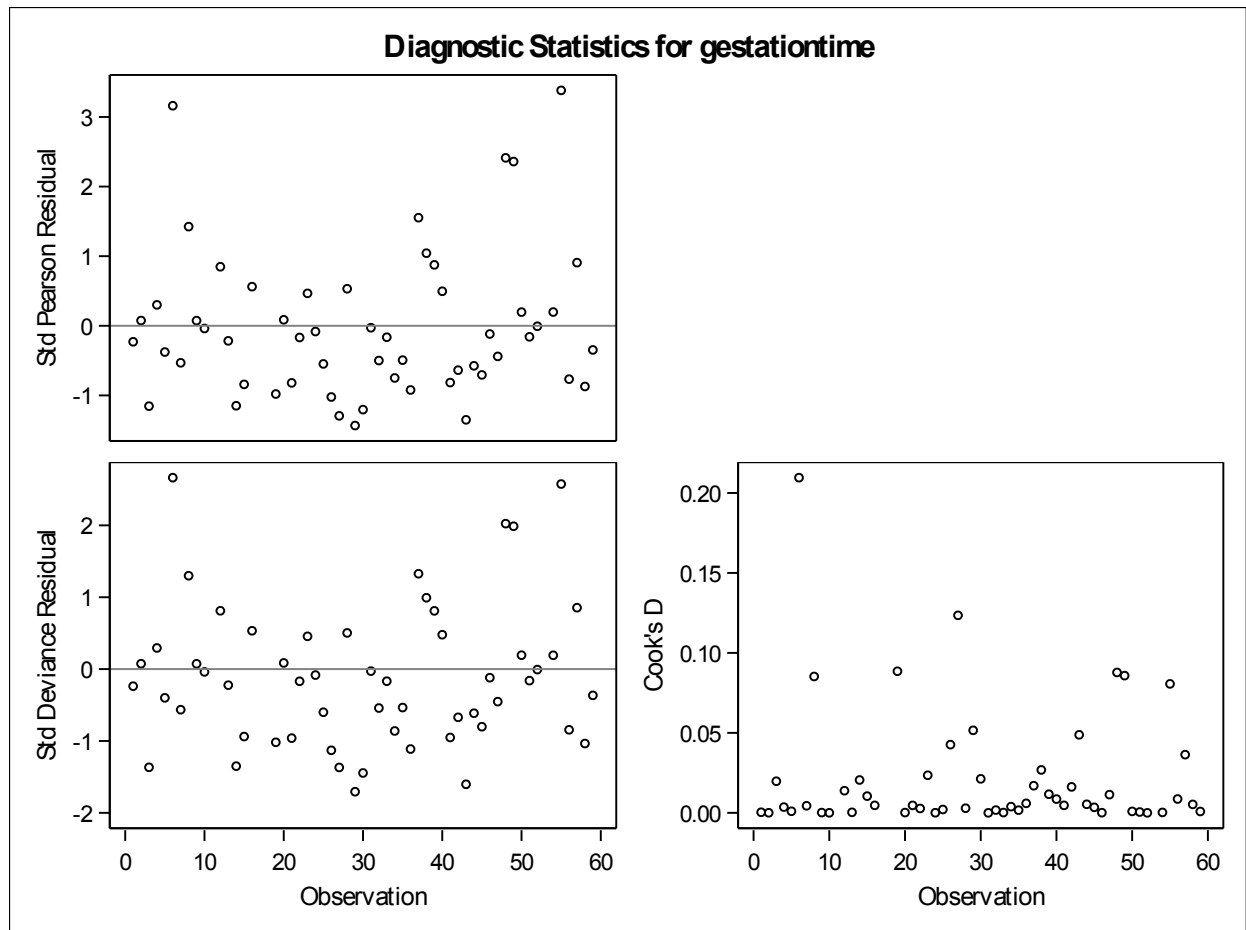
After removing bodyweight and predationindex, the model is refitted as below:

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	4.6967	0.2193	4.2669	5.1266	458.56	<.0001
brainweight		1	0.0023	0.0004	0.0014	0.0031	26.68	<.0001
sleepexposureindex	1	1	-0.7614	0.2576	-1.2663	-0.2566	8.74	0.0031
sleepexposureindex	2	1	-0.2210	0.2835	-0.7767	0.3347	0.61	0.4357
sleepexposureindex	3	1	0.1797	0.2746	-0.3586	0.7180	0.43	0.5129
sleepexposureindex	4	1	-0.0323	0.2592	-0.5404	0.4757	0.02	0.9008
sleepexposureindex	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	6.1334	0.0000	6.1334	6.1334		

LR Statistics For Type 1 Analysis							
Source	Deviance	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
Intercept	5218.4845						
brainweight	2412.5561	1	49	74.59	<.0001	74.59	<.0001
sleepexposureindex	1843.3132	4	49	3.78	0.0093	15.13	0.0044

LR Statistics For Type 3 Analysis						
Source	Num DF	Den DF	F Value	Pr > F	Chi-Square	Pr > ChiSq
brainweight	1	49	26.69	<.0001	26.69	<.0001
sleepexposureindex	4	49	3.78	0.0093	15.13	0.0044

According to the above tables, brainweight and sleepexposureindex are both significant under the significance level of 5%. The best set of predictors for the model includes two variables: brainweight and sleepexposureindex.



According to the residual plot, there is no obvious trend in residuals. Also, there are no points with Cook's distance greater than 1.

b)

Based on the results from part a, brainweight and sleepexposureindex should be retained for the final model.

According to the residual plot in part a, there is no obvious trend in residuals. Also, the highest Cook's distance is around 0.2, which is relatively low. Therefore, there is no unduly influential point and there is no need for refitting the model. The terms which should be retained for the final model are brainweight and sleepexposureindex.

c)

The estimates of the variables are as below:

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	4.6967	0.2193	4.2669	5.1266	458.56	<.0001
brainweight		1	0.0023	0.0004	0.0014	0.0031	26.68	<.0001
sleepexposureindex	1	1	-0.7614	0.2576	-1.2663	-0.2566	8.74	0.0031
sleepexposureindex	2	1	-0.2210	0.2835	-0.7767	0.3347	0.61	0.4357
sleepexposureindex	3	1	0.1797	0.2746	-0.3586	0.7180	0.43	0.5129
sleepexposureindex	4	1	-0.0323	0.2592	-0.5404	0.4757	0.02	0.9008
sleepexposureindex	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	6.1334	0.0000	6.1334	6.1334		

According to this table, the estimates of brainweight and sleepexposureindex=1 are significant. The coefficient for brainweight is estimated as 0.0023. This means that for each one-unit increase in brain weight, the expected number of gestation time increases by a factor of $\exp(0.0023)=1.0023$. For the five-level categorical predictor sleepexposureindex, the estimated coefficients of levels 1, 2, 3, 4 are related to level 5. The coefficient for sleepexposureindex=1 is estimated as -0.7614, and $\exp(-0.7614)=0.4670$. This means that the expected number of gestation time for level 1 of sleep exposure index is 0.4670 of that for level 5 of sleep exposure index. The coefficient for sleepexposureindex=2 is estimated as -0.2210, and $\exp(-0.2210)=0.8017$. This means that the expected number of gestation time for level 2 of sleep exposure index is 0.8017 of that for level 5 of sleep exposure index. The coefficient for sleepexposureindex=3 is estimated as 0.1797, and $\exp(0.1797)=1.1969$. This means that the expected number of gestation time for level 3 of sleep exposure index is 1.1969 times than that for level 5 of sleep exposure index. The coefficient for sleepexposureindex=4 is estimated as -0.0323, and $\exp(-0.0323)=0.9682$. This means that the expected number of gestation time for level 4 of sleep exposure index is 0.9682 of that for level 5 of sleep exposure index.

There is a remaining issue in the final model. According to the diagnostics, there are certain observations with absolute value of Pearson residual greater than 2, which might be high influential points. Nevertheless, there is no obvious trend in residuals and the cook's distance does not appear that these points are unduly influential. So this issue is not a big problem.