

*The LOGISTIC Procedure***WARNING:** *The validity of the model fit is questionable.*

HW 4 Spring 2018

JINRAN YANG

## Problem 1

*The LOGISTIC Procedure*

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		2.5451	0.6366

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
gestationtime	1	8.5131	0.0035

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.0833	0.4960	4.7711	0.0289
gestationtime	1	0.0129	0.00442	8.5131	0.0035

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
gestationtime	1.013	1.004	1.022

*The LOGISTIC Procedure*

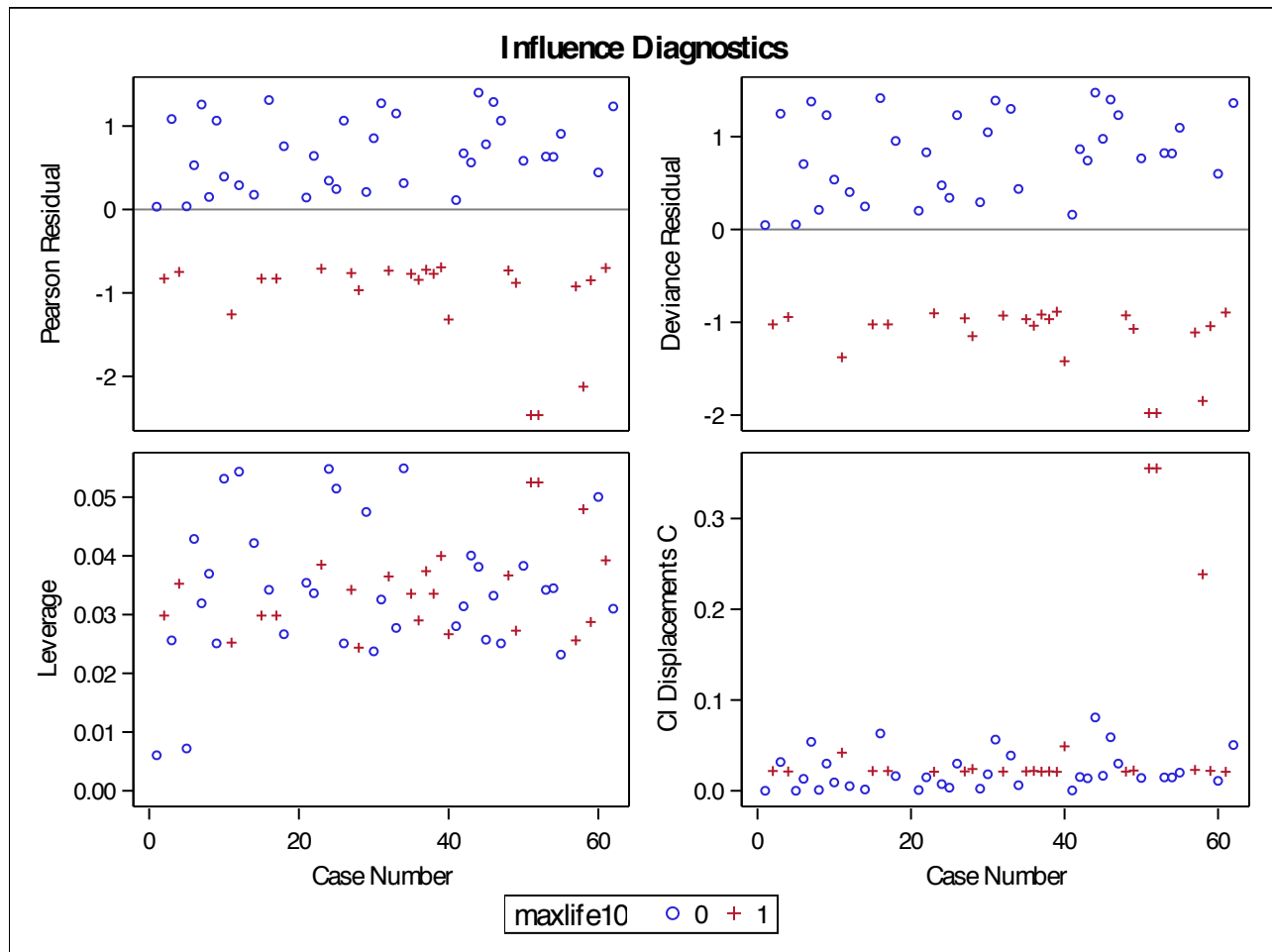
Analysis the final model:

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

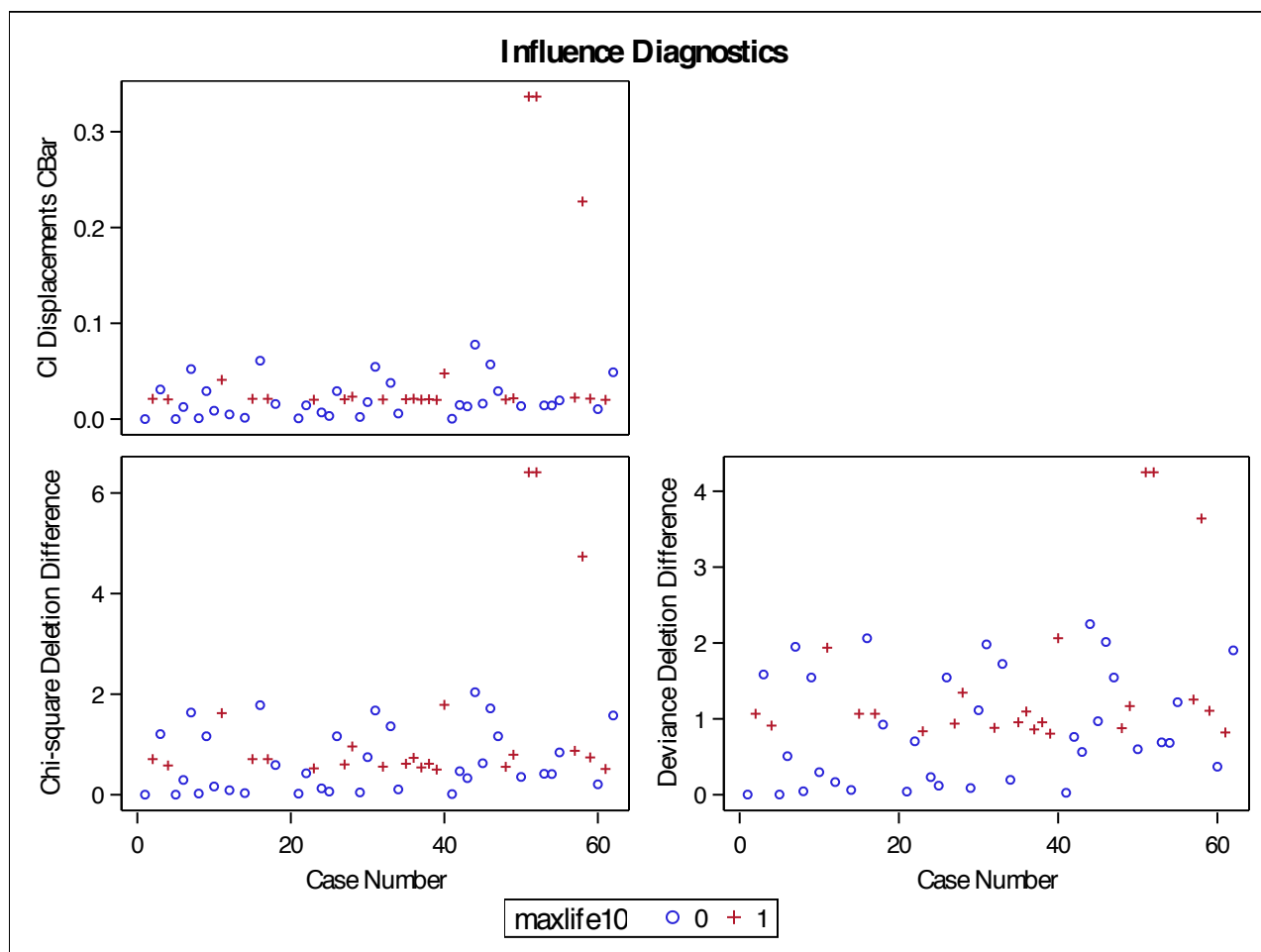
Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.8759	0.4632	3.5760	0.0586
gestationtime	1	0.0119	0.00414	8.2827	0.0040

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

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

*The LOGISTIC Procedure*

### The LOGISTIC Procedure



- (a) According to the result of the stepwise model selection, only one predictor-gestationtime remain in the model. Then, we analysis the final model which is selected by the stepwise model selection, and we can see from the table **Testing Global Null Hypothesis:  $BETA=0$**  that the model is significant. Also, we can see that the gestationtime parameter is significant, since its P value is smaller than 0.05. Then we can look at the result of the influential diagnostics. As we can see from the graph **CI Displacements Cbar**, there is no point larger than 1 so that we do not have influential points in the final model.
- (b) we can see that the **gestationtime** is significant, since its P value is smaller than 0.05 and the odds ratio of the **gestationtime**'s **95% Wald Confidence Limits** doesn't contain 1; P value of the **Hosmer-Lemeshow's test** is larger than 0.05 which means there is a good fitness of the model.
- It can be told from the **Pearson Residuals** and **Deviance Residuals**' graph that the residuals are randomly distributed between -2 and 2, so this model should be fine.

*The LOGISTIC Procedure*

(c) The **95% Wald Confidence Limits** of **gestationtime** odds ratio doesn't contain 1, so it is significant.

The odds of species maximum life span at least 10 years changes by a factor of 1.012 with one unit increase in **gestationtime**, which imply a very slight positive relationship between them.

*The LOGISTIC Procedure***Problem 2**

Perform the stepwise model selection:

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

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	0.3829	0.9131	0.1759	0.6750
predationindex	1	-3.0348	0.9392	10.4407	0.0012
sleepexposureindex	1	3.6603	1.0719	11.6601	0.0006

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
predationindex	0.048	0.008	0.303
sleepexposureindex	38.872	4.756	317.727

*The GENMOD Procedure*

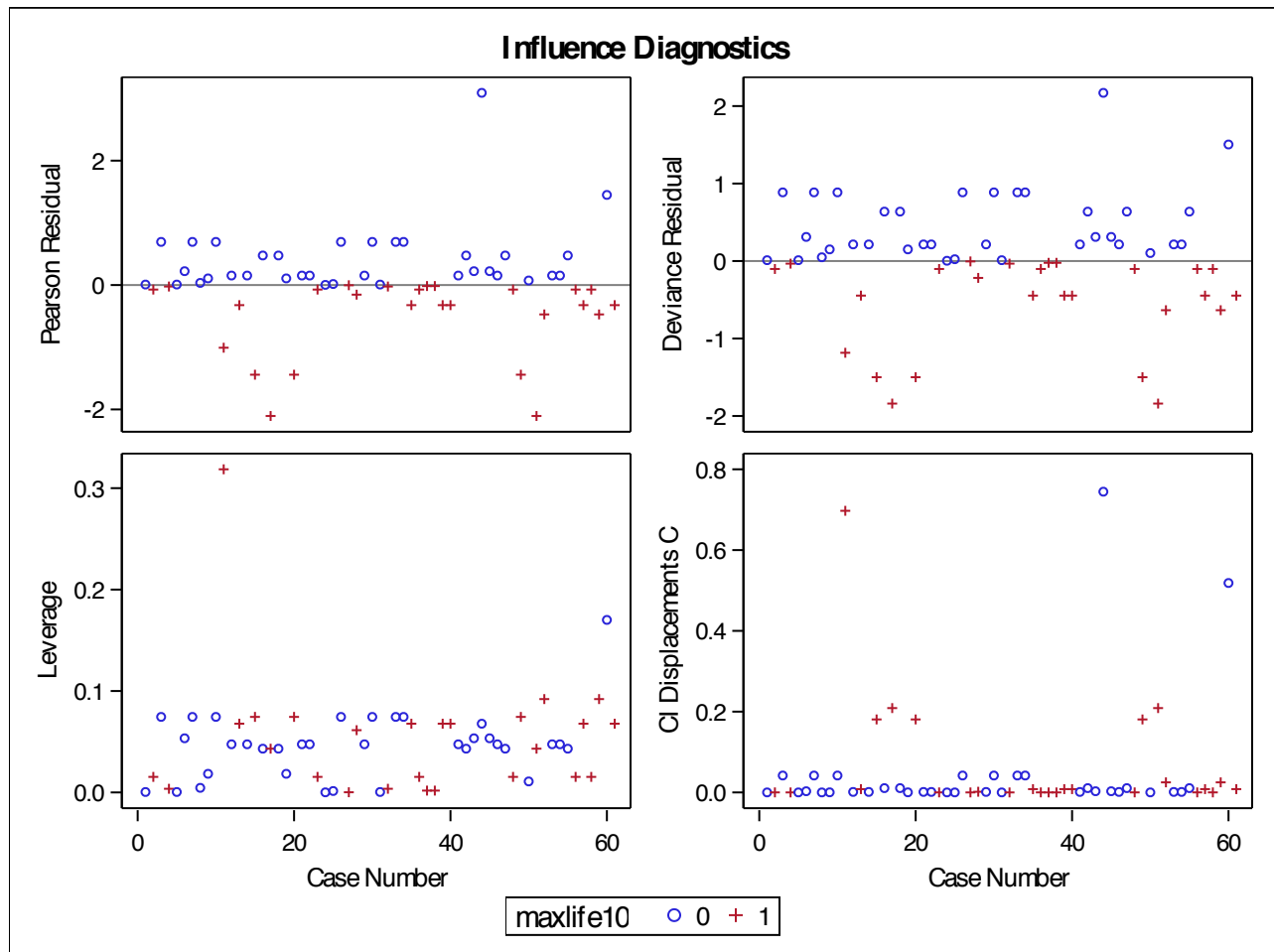
Analysis the final model selected by the stepwise model selection and remove all the influential points:

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

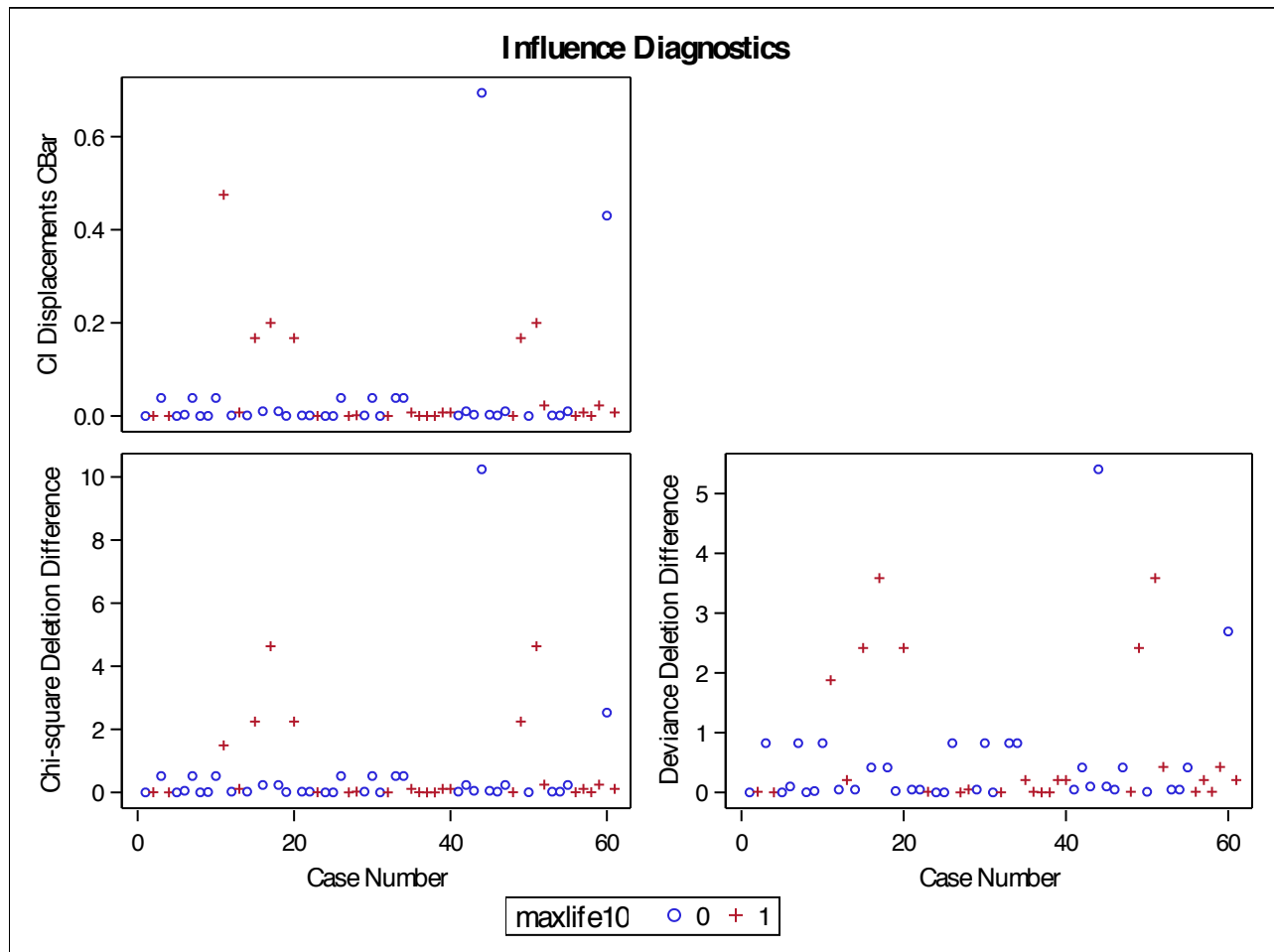
Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
predationindex	0.050	0.008	0.306
sleepexposureindex	42.338	5.246	341.662

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

*The GENMOD Procedure*



### The GENMOD Procedure



(a) As we can see from the graph *CI Displacements Cbar*, there is no point larger than 1 so that we do not have influential points in the final model.

(b) we can see from the table *Testing Global Null Hypothesis: BETA=0* that the model is significant, since the P value are way too smaller than 0.05. Also, we can see that the predictor parameters estimated are significant, since its P value are smaller than 0.05.

P value of the *Hosmer-Lemeshow's test* is larger than 0.05 which means there is a good fitness of the model.

It can be told from the *Pearson Residuals* and *Deviance Residuals'* graph that most residuals are distributed between -2 and 2, but there are some suspicious points really close to 2 or -2. There might some problem about these points.

(c) The 95% Wald Confidence Limits of the odds ratio of both **predationindex** and **sleepexposureindex** doesn't contain 1, so they are both significant.

### *The GENMOD Procedure*

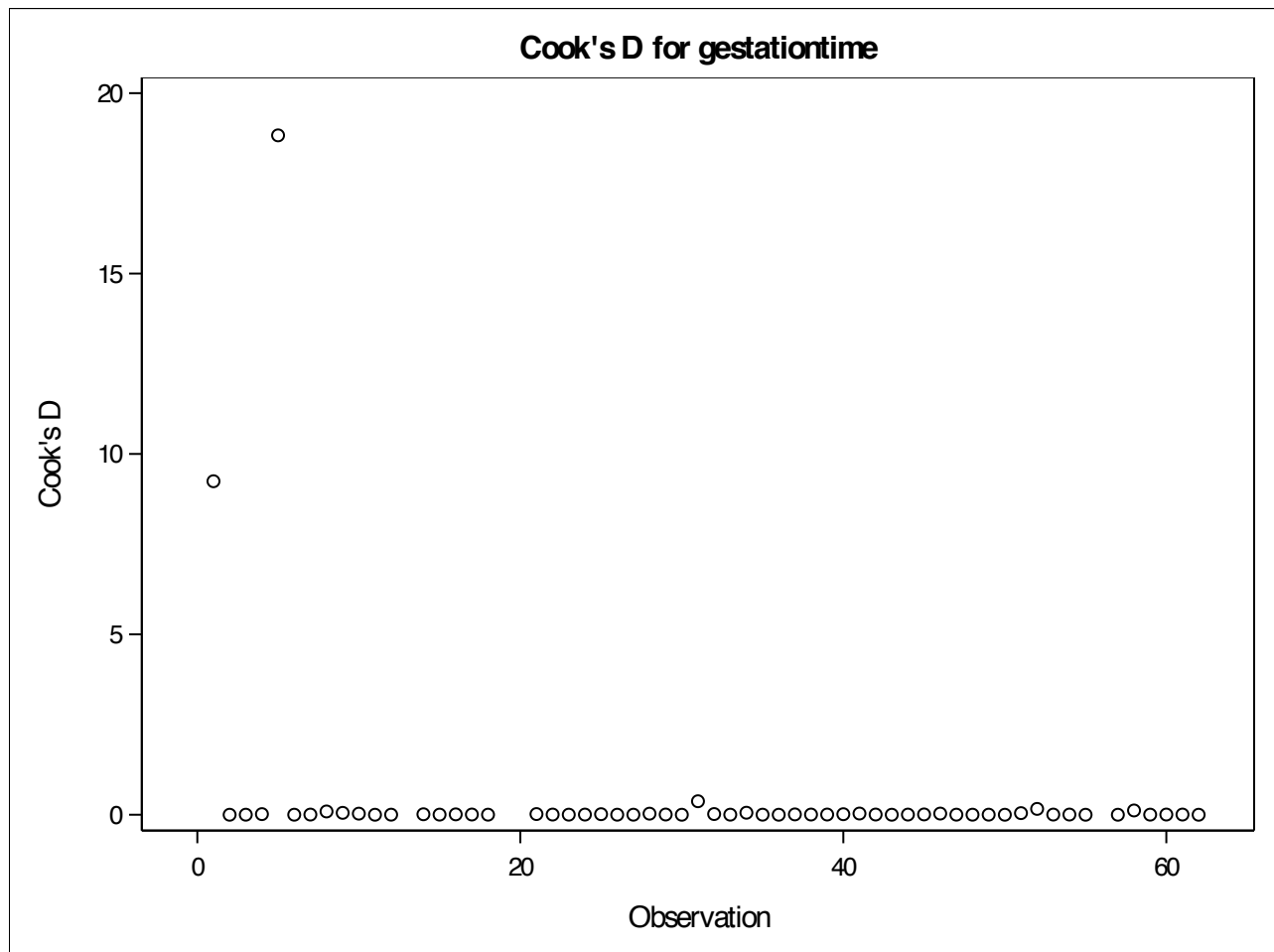
The odds of species maximum life span at least 10 years changed by a factor of 0.05 with one unit increase in **predationindex**, which imply a strong negative relationship between species maximum life span at least 10 years and **predationindex**.

The odds of species maximum life span at least 10 years changed by a factor of 42.338 with one unit increase in **sleepexposureindex**, which imply a strong positive relationship between species maximum life span at least 10 years and **sleepexposureindex**.

#### Problem 3

(a)

Analysis full model to see the influential points:



There are at least two influential points (which are greater than 1) exist. Then we remove them.

*The GENMOD Procedure*

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	44	1496.4925	34.0112
Scaled Deviance	44	44.0000	1.0000
Pearson Chi-Square	44	1678.7128	38.1526
Scaled Pearson X2	44	49.3577	1.1218
Log Likelihood		806.6882	
Full Log Likelihood		-918.5161	
AIC (smaller is better)		1859.0321	
AICC (smaller is better)		1865.1717	
BIC (smaller is better)		1881.1128	

Algorithm converged.
----------------------

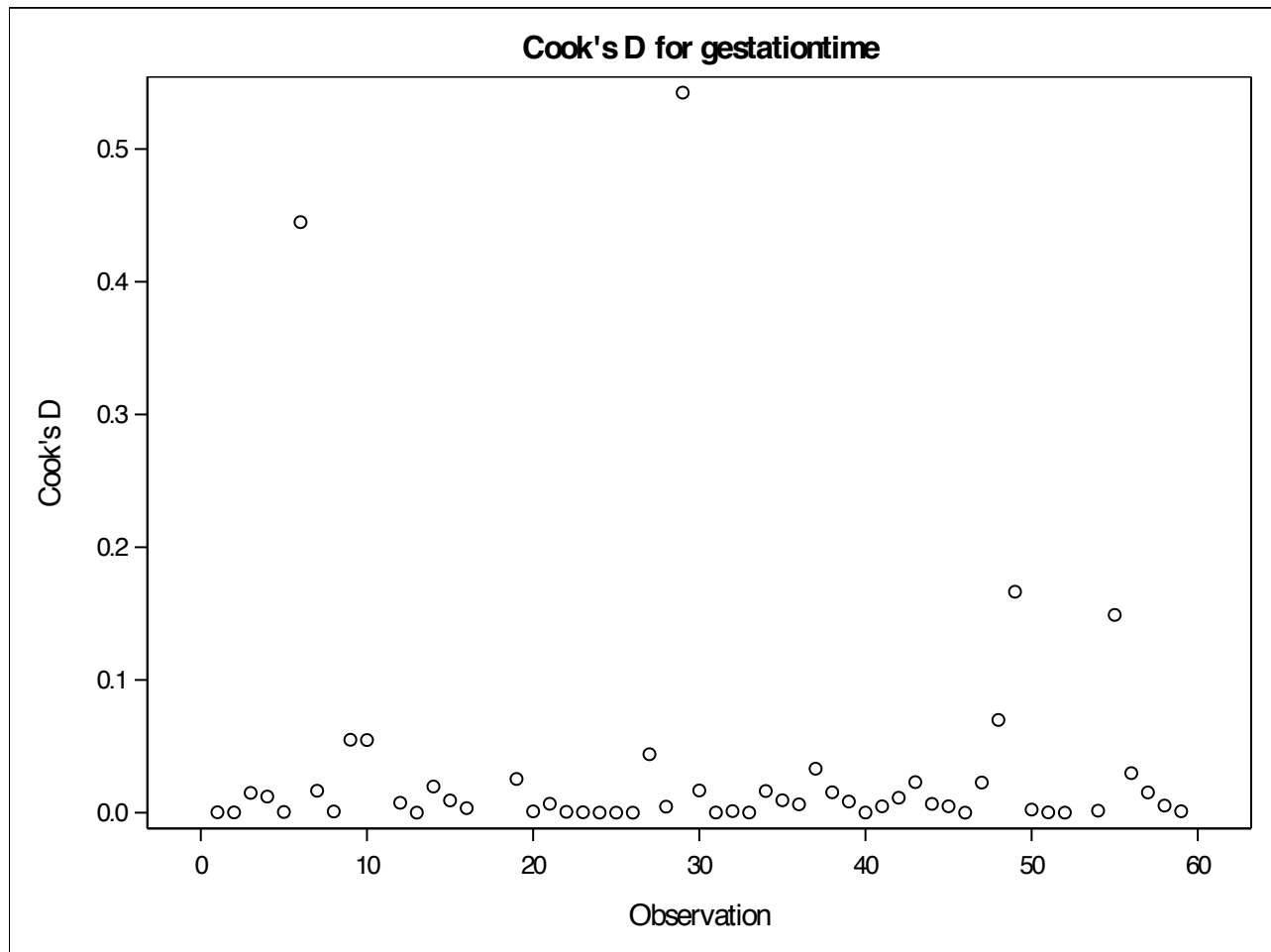
Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	4.4958	0.2380	4.0293	4.9623	356.77	<.0001
bodyweight		1	-0.0019	0.0011	-0.0040	0.0002	3.19	0.0741
brainweight		1	0.0039	0.0009	0.0021	0.0056	18.98	<.0001
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	.	.
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	.	.
Scale		0	5.8319	0.0000	5.8319	5.8319		

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

*The GENMOD Procedure*

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	44	57.08	<.0001	57.08	<.0001
brainweight	2343.3497	1	44	27.46	<.0001	27.46	<.0001
predationindex	2072.5203	4	44	1.99	0.1125	7.96	0.0929
sleepexposureindex	1496.4925	4	44	4.23	0.0055	16.94	0.0020

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

*The GENMOD Procedure*

After 3 times removal, we removed all influential points (considering the overdispersion issues). And we can see from the graph above that there is no point larger than 1.

Also, based on the type III result (since its result will not be affected by the order of the predictors), we find that only **bodyweight** and **sleepexposureindex** are significant (their P values are smaller than 0.05), the rest of predictors are not significant, so we should only keep **bodyweight** and **sleepexposureindex** in the model.

### Problem 3b

*The GENMOD Procedure*

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	49	1843.3132	37.6186
Scaled Deviance	49	49.0000	1.0000
Pearson Chi-Square	49	2077.0516	42.3888
Scaled Pearson X2	49	55.2134	1.1268
Log Likelihood		724.7210	
Full Log Likelihood		-1091.9264	
AIC (smaller is better)		2195.8528	
AICC (smaller is better)		2197.6028	
BIC (smaller is better)		2207.8968	

Algorithm converged.

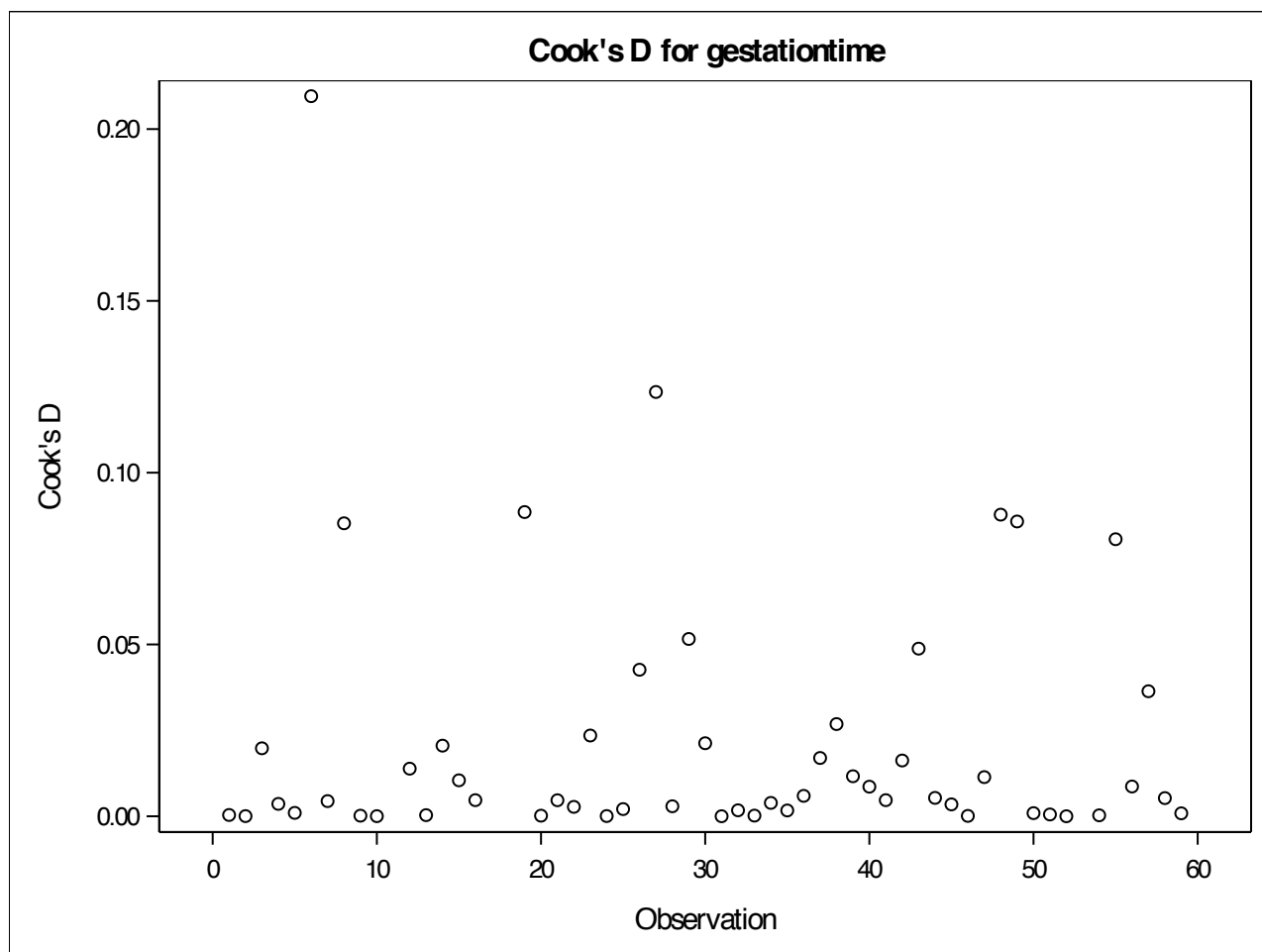
Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	4.6967	0.2193	4.2669	5.1266	458.56	<.0001
brainweight		1	0.0023	0.0004	0.0014	0.0031	26.68	<.0001
sleepexposureindex	1	1	-0.7614	0.2576	-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		

**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	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

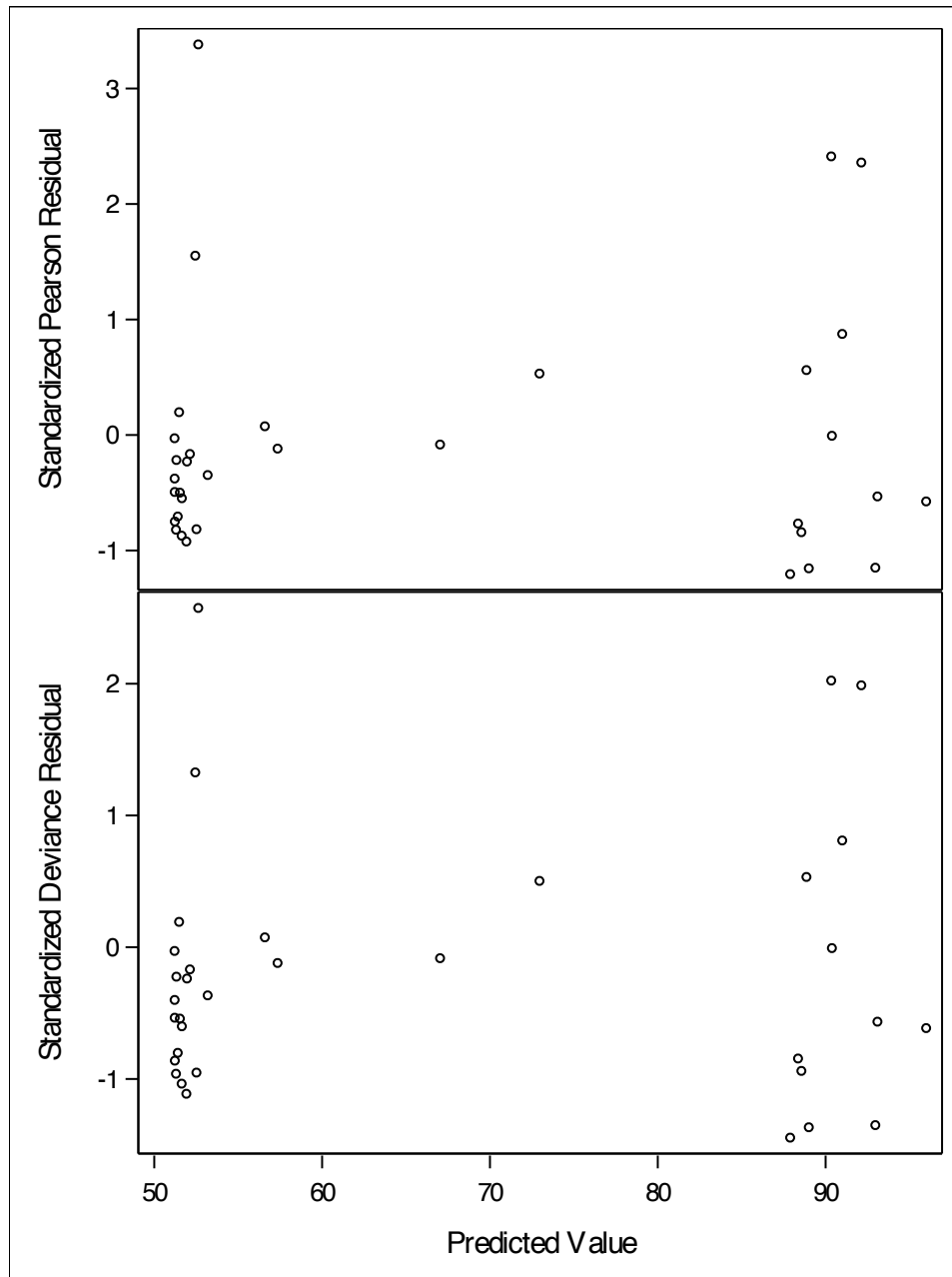
*The GENMOD Procedure*

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



As we can see from the above graph, there is no point larger than 1 so that we do not have influential points in the final model.

**Problem 3C**



By looking at the residuals plots from the above graph, we find that most residuals are distributed between -2 and 2, but they are some suspicious residual larger than 2 which might be problem. And the points seem not really random which might also be a issue.

By looking at *the Analysis Of Maximum Likelihood Parameter Estimates*, we find that the **brainweight** and **sleepexposureindex is 1** are significant, since the P value of these two are smaller than 0.05. The rest are not significant. So that we should consider the **brainweight** and **sleepexposureindex** are significant and these two should be included in the model.

Thus, gestation time chanced by a factor of  $e^{0.0023}$  (1.002303) with one unit increase in **brainweight**, which imply a very slightly positive relationship between gestation time and **brainweight**. So, species with larger brain weight tend to have larger expected number of days in a



full-term pregnancy

As for the **sleepexposureindex**, gestation time changed by a factor of  $e^{-0.7614}$  (approximately 0.467) when animals are least exposed (**sleepexposureindex is 1**) during sleep compared to the situation that animals are largest exposed (**sleepexposureindex is 5**), which imply that the gestation time tend to be shorter when animals are least exposed during sleep. So, species which are least exposed when sleep tend to have shorter expected number of days in a full-term pregnancy.