Initials: _____

PHP 2511
Spring Midterm Examination
Wednesday, 16 March 2016

Name: _____

- Put your name on the first page and your initials at the top of each page.
- Write your solutions directly on the exam and show all your work.
- You have 80 minutes for the examination, from 10:30 to 11:50. Papers will be collected at 11:50 and you must stop working then.
- This is a closed book exam. You may not use any notes, computer or other resources.
- The exam has 10 problems and 11 pages. Some problems span several pages, so be sure to read the exam thoroughly. Please check your exam to insure it is complete. The last page has been left blank to give you extra space to show your work.
- Some parts of a question may require the answer to an earlier part. If you cannot solve the earlier part, you can still receive partial credit for the later parts: make up a reasonable answer for the earlier part and use that in solving the later parts.
- Show your work and explain your reasoning; the final answer is not as important as the process by which you arrived at that answer. We can more easily give you partial credit if you have explained your steps.
- Be careful to specify the units of measurement where applicable.
- All R output is in a separate packet.

Problem Scoring		
Problem	Point Value	Problem Grade
1	5	
2	11	
3	5	
4	13	
5	9	
6	8	
7	8	
8	8	
9	10	
10	8	
Total	85	

The Data

These data are from a randomized study to evaluate an antiepileptic drug in reducing the frequency of simple or partial seizures. The treatment variable is coded 0 for placebo and 1 for the drug under study, seizures is the outcome variable and indicates the number of seizures experienced over a two-week period during treatment. The baseline variable is the pre-treatment seizure rate over 8 weeks prior to randomization and is a potential adjustment variable in addition to the age of the patient (variable: age).

Variable	Description
Seizures	Number of seizures over 2 week period.
Treatment	Type of Treatment
0	Placebo
1	Drug
Baseline	Seizure Rate pre-treatment.
age	age of patient.

1. (5 points) We are interested in predicting the number of seizures. We run simple linear regressions on the variables in the data set and the results are as reported below:

	Estimate	95% CI of Estimate	p-Value	R^2
treatment	-0.776	(-8.421 , 6.868)	0.84289899	0.00069
baseline	0.439	(0.352 , 0.526)	0	0.63218
age	0.018	(-0.593 , 0.629)	0.95465464	6e-05

Which variable explains number of seizures best on its own? How do you know this?

2. (11 points) We decide to build a multiple regression model. *(See the R output for the summary of this)*

   a. (6 points) Are there any changes in the estimates from the simple linear regressions? Explain.

b. (5 points) Interpret the effect of `baseline` in the context of this problem.

3. (5 points) One of your colleagues suggest that the reason `treatment` is proving to have been insignificant is that there may be an interaction between the `baseline` and `treatment`. What would interaction mean in this case?

4. (13 points) You decide to run the model with the interaction term added in. *(See the R output for the summary of this)*

   a. (5 points) Does adding the interaction term in improve the model? Explain.

   b. (8 points) Write out the 2 models that we now have to explain these variables effect on the number of seizures. *(No need to actually add any coefficients just write out the equations and leave the additions).*

5. (9 points) You decide to keep moving with the model you have above since it uses all of the variables in the model and has a decent R^2. You move onto testing the model with residual plots. What do these residual plots tell you about the fit? *(See the R output for the plots.)*

6. (8 points) You then move onto marginal model plots. What does this marginal model plot tell you about how this model fits the data? *(See the R output for the plots.)*

7. (8 points) You decide that before moving on you wish to test for outliers. You use a Cook's Distance plot. Do you find any outliers? If so how many are clearly outliers? *(See the R output for the plots.)*

8. (8 points) You remove the largest outlier and proceed from there. Does the summary change from before the outlier was removed? If so, in what way? *(See the R output for the summary.)*

9. (10 points) Do the plots displayed show that this model excluding the outlier has a better fit? Explain. *(See the R output for the plots.)*

10. (8 points) The PI on the study is concerned that there was a problem with the randomization. They feel the treatment does do something but someone made an error in the design. If there is no relationship between treatment and baseline or treatment and age, then we can rule out randomization as being a flaw. Luckily you know logistic regression and you run 2 simple logistic regressions. *(See the R output for the summaries.)*

    a. (4 points) Is the PI right? Do you find associations between treatment and age/baseline?

    b. (4 points) Interpret the effect of baseline on treatment. (*Do this regardless of significance.*)