COVER PAGE

STAT 608 Homework 05 Summer 2017

Please TYPE your name and email address below, then convert to PDF and attach as the first page of your homework upload.

NAME:

EMAIL:

HOMEWORK NUMBER:

STATISTICS 608
Homework 608 S17 05
Due: *11:59 PM, July 8, 2017*

Question 1 [4]

In Exercise 3, page 105 in the textbook you were instructed to use a logarithmic transformation for Ad Revenue and a square root transformation for Circulation. What would be the appropriate transformations according to the Box-Cox technique? (You are not required to implement these transformations.)

Question 2 [6]

This question is related to Exercise 5 on page 224 in the textbook. Assume that $log(Y)$ is an appropriate transformation of the response variable.
Use marginal model plots to *evaluate the fit of the full seven-covariate model.* Describe briefly weaknesses, if any, in the model that these plots reveal.

Question 3 [1+4+4+2]

In the simple linear regression model

$$y_j = \beta_0 + \beta_1 x_j + e_j,$$

$j = 1, \ldots, n$, the predicted values are defined by

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_j$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the least squares estimators of β_0 and β_1.

3.1 Show that the mean of the y - values, $\bar{y}$, equals the mean of the predicted values, $\overline{\hat{y}}$.

3.2 Show that

$$SS_{reg} = \sum (\hat{y}_i - \overline{\hat{y}})(y_i - \bar{y})$$

where

$$\overline{\hat{y}} = n^{-1} \sum \hat{y}_i.$$

3.3 Hence, show that the statistic

$$R^2 = \frac{SSreg}{SST}$$

equals the *square* of the Pearson correlation coefficient between the pairs $(y_j, \hat{y}_j),\ j = 1, \ldots, n$. [Hint: $\frac{a}{b} = \frac{a^2}{ba}$ for $a,\ b \neq 0$.]

3.4 It can be shown (you don't have to) that the result in Question 3.2 is also true in a general linear model setup with $m\ (< n)$ covariates. Suppose $m = 20$ and that you attempt to select the best subset of covariates by maximizing R^2. How many covariates will be in your best subset? Justify your answer.