

STAT 512 HW1

Reading: Ott & Longnecker 12.1-12.6

See Canvas calendar for the due date.

Local Students: Submit your HW (1) through Canvas in pdf format, (2) at class, or (3) to the grader's mailbox in Statistics 102. Do not slide it under my office door!

Distance Students: Submit your HW through Canvas in pdf format.

28 points total, 2 points per question unless otherwise noted.

Notes:

- The data sets from the Ott and Longnecker book can be downloaded from the book's companion website. From the link available from Canvas, download the **ASCII-comma** (CSV, comma delimited) files, then unzip the data. Confusingly, the file extension is .TXT even though the files are actually CSV!
- To import the data, use code something like this:

```
InData <- read.csv("yourfilepath", quote = " ' ")
```

or

```
InData <- read.csv(file.choose(), quote = " ' ")
```

The file.choose() function allows files to be chosen interactively (without typing out the file path location). The quote option used above is because the column names in the original data are (single) quoted (ex: 'STEER').
- For this data, the first column just gives the animal ID (which will not be used for modeling), one way to deal with this is to use the row.names option:

```
InData <- read.csv("yourfilepath", quote = " ' ", row.names = 1)
```
- Always look at the data (using str() and/or View()) after importing!

Read **Problem 12.53** (or 12.56 from the 6th Edition) which deals with cattle feed lot diets.

1. Show the pairwise scatterplots between all 4 variables (Y=Time, X1=Protein, X2=Antibio, X3=Supplem).
2. Calculate pairwise (Pearson) correlations between all 4 variables.
3. Run the 3 simple linear regressions of Time vs each of the above three predictor variables. Show the parameter estimates ("Coefficients" table) and R^2 values. You can just copy/paste the relevant output from R. **(6 pts)**
4. Now run multiple regression of Time on all three predictor variables. Show the parameter estimates ("Coefficients" table) and R^2 value. **We will use this the "full" model for the remaining questions.**
5. Note that (1) the slope estimates from the simple linear regressions are the same as the slope estimates from the "full" model and (2) the R^2 values from the simple linear regressions sum to the R^2 value from the "full" model. In general, this will not be the case (as we saw with the Rice Example). What is different about this data (as compared to the Rice Example)? Hint: Consider the result of question 2.

6. Create plots of (A) Residuals vs Fitted values and (B) QQplot of residuals. Include these plots in your assignment. Thinking about model assumptions, discuss your findings for each plot. **(4 pts)**
7. Interpret the R^2 value from “full” model.
8. Give a one-sentence interpretation of estimated partial regression coefficient for AntiBio in the multiple regression.
9. Working from the “full” model, for each of the four β 's (intercept and three partial regression coefficients) give a p-value for the hypothesis that the true parameter value is zero vs a two-sided alternative. In other words, test $H_0: \beta_i = 0$ versus $H_A: \beta_i \neq 0$.
10. Working from the “full” model, test the null hypothesis that the partial regression coefficient for Protein equals -3.0 versus a two-sided alternative. In other words, test $H_0: \beta_1 = -3$ versus $H_A: \beta_1 \neq -3$. Give a test statistic, p-value and conclusion. **(4 pts) Note:** One approach to this question uses the car package. Remember you need to install a package the first time you use it and load the package every time you use it!