STAT 540, Fall 2015

Due on **12:00 P.M., Friday, September 11th**

Homework 2

Homework format for all STAT 540 homework this term: Please label all problems clearly and turn in an organized homework assignment. You don't need to spend hours producing beautifully typeset homework, but you won't get credit if we can't find or read your answer. Unless noted otherwise, turn in the following (as appropriate for the problem).

- Theoretical derivation (when asked for).
- Numerical results **with an explanation of your solution**, written in complete sentences. If computer code is absolutely necessary to provide context here, then include it–nicely formatted–within the solution (otherwise, see below).
- Appropriate graphics. Use informative labels, including titles and axis labels. Try to put multiple plots on the page by using, for example, the R command `par(mfrow=c(2,2))`.
- **Only as necessary**: Final clean computer code used to answer the problem **attached to the end of your homework**. Only include the rare code excerpts without which we wouldn't be able to figure out what you did. Annotate your code. Number and order the code in order of the problems. When in doubt, leave it out; consider that we will probably never read it.
- Some problems will be relatively open-ended, such as "Here are some data. Analyze them and write a report." I will provide further instructions about reports later. They should be self-contained, with suitable EDA, graphs, numerical results, and **scientific interpretation**. No computer code should be included. The report should be concise: "no longer than necessary".

(1) Consider the problem (7 (b)) in HW 1, i.e. the problem 1.27 in the textbook. Using data from that problem and based on your answer in HW 1, provide solution to the following questions.

(a) Report the value of R^2, the square of the multiple correlation coefficient. Interpret this value.

(b) Perform a t-test of the null hypothesis that the slope of the line is zero. Report the value of the t-statistic, it degrees of freedom, and a p-value. State your conclusion.

(c) Estimate the mean muscle for 55 year old women. Report a standard error for your estimate.

(d) Construct a 95% confidence interval for the mean muscle mass for 55 year old women.
(e) Construct a 99% prediction interval for the observed muscle mass of a randomly selected 75 year old woman.

(2) Experience with a certain type of plastic indicates that a linear relationship exists between the hardness (measured in Brinell units) of items molded from the plastic and the elapsed time since completion of the molding process. To investigate this relationship, sixteen batches of the plastic were made and one test item was molded from each batch. Each item was randomly assigned to one of four predetermined time periods and the hardness of the item was measured after the assigned time had elapsed. The results are shown in the following table. (Textbook 2005, problem 1.22).

Time (hours) (x_i)	Hardness (Brinell units) (y_i)	Estimated Mean ($\hat{y}_i$)	Residual (e_i)
16	199		
16	205		
16	196		
16	200		
24	218		
24	220		
24	215		
24	223		
32	237		
32	234		
32	235		
32	230		
40	250		
40	248		
40	253		
40	246		

(a) Compute the values of the least squares estimates of β_0 and β_1 in the simple linear regression model.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \text{ where } \epsilon_i \sim N(0, \sigma^2),\ i = 1, \cdots, 16$$

(b) Compute the least squares estimate of the condition hardness mean, $\hat{y}_i$ and the corresponding residual, ϵ_i for each of 16 cases in the data set and record the values in the table shown above.
(c) Construct a scatter plot of the observed hardness values versus the elapsed times and display the estimated regression line on your plot. Does it appear that a straight line relationship is appropriate?
(d) Complete the following ANOVA table. Use the values of the residuals computed in part (b) to compute the sum of squared residuals. Use the predicted values $\hat{y}_i$ computed in part (b) to compute the model sum of squares. Provide some details.

Source of variation	df	Sum of squares	Mean squares	F-statistics
Regression model				
Error				
Corrected total				

(e) Test the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 \neq 0$. Report the value of an F-statistic, its degrees of freedom and the resulting p-value. State your conclusion (use $\alpha = 0.05$).
(f) Give a one sentence interpretation of β_1 in the context of this study.
(g) Construct a 95% confidence interval for β_1.
(h) Give a one sentence interpretation of β_0 in the context of this study.
(i) Construct a 95% confidence interval for β_0.
(j) Compute individual 95% confidence intervals for the mean hardness of items molded from this plastic at elapsed times of 16, 30, and 60 hours. These confidence intervals have different widths. At what elapsed time would a 95% confidence interval for the mean hardness have the smallest width?.
(k) Compute the limits for a 95% prediction interval for a future measurement of hardness for an item molded from this plastic after an elapsed time of 60 hours. Explain why these limits are wider than the confidence limits for the mean hardness at 60 hours computed in part (j).

(3) Show the results on pages 21, 26 and 48 in lecture 3, i.e. under the simple linear regression model
(a) $\mathbb{E}(\hat{\beta}_0) = \beta_0$
(b) $\text{cov}(\bar{Y}, \hat{\beta}_1) = 0$
(c) $\sigma^2\{\hat{\beta}_0\} = \text{Var}\{\hat{\beta}_0\} = \sigma^2 \left[1/n + \bar{x}^2/(\sum_{i=1}^n (x_i - \bar{x})^2\right]$
(d) $r = \text{sign}(\hat{\beta}_1)\sqrt{R^2}$ that is the sample Pearson correlation is related to R^2 under simple linear regression (notice, this is only true for simple linear regression model)

(4) Suppose we know in advance that the intercept is zero, so the model is

$$y_i = \beta_1 x_i + \epsilon_i$$

where ϵ_i are independent random variables with zero mean and constant variance σ^2.
(a) Derive the least square estimator of β_1, say $\hat{\beta}_1$.
(b) Show that $\hat{\beta}_1$ is unbiased.
(c) Derive the variance of $\hat{\beta}_1$.
(d) Show that $\hat{\beta}_1$ is the best linear unbiased estimator of β_1 for the model given above.

(5) Textbook problems:
(a) Problems 2.2: interpreting H_0; 2.10 (Estimation versus prediction).
(b) Problems 2.13 parts (a), (b) and (c); 2.65