Homework 1
STAT 334, Spring 2019
Due at the beginning of class on Thursday, April 11

These problems do not require any work on the computer.

1. Problem 1.5, p. 33
2. Problem 1.7, p. 33
3. Problem 1.8, p. 33

4. Problem 1.11, p. 34. Answer the questions below rather than the question given in the book.
(a) Interpret the meaning of β_1 in the context of the problem.
(b) Explain what having $\beta_1 < 1$ means in terms of the success of the training program.
(c) Interpret the meaning of β_0 in the context of the problem.
(d) Is the value of β_0 meaningful? Explain.
(e) Overall, do you think that the training program increased production output. Explain why or why not.

5. Problem 5.3, p. 209.

6. Write out matrix equations to prove the following statements. (You may find Section 5.7 in the text helpful.)
(a) The matrix $\mathbf{X}'\mathbf{X}$ is symmetric.
(b) The matrix $\mathbf{H}$ is symmetric.
(c) The matrix $\mathbf{H}$ is idempotent.

The following problems require *JMP*.

7. Read the introduction to Problem 1.19 on page 35 of the textbook. The **Grade point average** data is on the PolyLearn in the file **GPA.jmp**. Run a regression of GPA (y, in grade points) vs. ACT score (x, in ACT points) and make a scatterplot of the variables with a regression line on it. Answer the following questions about the data instead of the questions in the textbook.
(a) What is the estimated regression function?
(b) Is there a linear association between the freshman GPA and ACT score? State the null and alternative hypotheses for the appropriate test. Explain how you determine the result of the test using a 5% level of significance, then write an interpretation of the result in the context of the data.
(c) What does the R^2 of this regression mean in the context of the data?
(d) Write a sentence explaining the meaning of the slope of this regression with 95% confidence.
(e) Write a sentence explaining the meaning of the intercept of this regression with 95% confidence. Is the intercept meaningful? Explain.

8. Read the introduction to Problem 1.22 on page 36 of the textbook. The **Plastic hardness** data is on the PolyLearn in the file **Plastic.jmp**. Run a regression of hardness (y, in Brinell units) vs. time (x, in hours) and make a scatterplot of the variables with a regression line on it. Answer the following questions about the data instead of the questions in the textbook.
(a) What is the estimated regression function?
(b) What is the value of e_1, the residual of the first observation? Explain how e_1 differs from ε_1, the error for the first observation.
(c) Is there a linear association between the number of times a carton is transferred and the number of ampules broken? State the null and alternative hypotheses for the appropriate test. Explain how you determine the result of the test using a 5% level of significance, then write an interpretation of the result in the context of the data.
(d) Write a sentence explaining the meaning of the slope of this regression with 95% confidence.
(e) Write a sentence explaining the meaning of the intercept of this regression with 95% confidence. Is the intercept meaningful? Explain.

Continued on the next page

The following problems require *R*.

9. Compute the following quantities in *R* using the **Plastic hardness** data and answer questions about them below. A tab-delimited text file of this data is on PolyLearn in the file **Plastic.txt**. The *R* script **STAT334-HW1-Plastic.r** contains code to set up the data into a response vector **y** and a design matrix **X**. You will have to type in the formulas for the different calculations below on your own.

- the coefficient vector **b**
- the hat matrix **H**
- the predicted value vector $\hat{\mathbf{y}}$
- the residual vector **e**
- the sums of squares *SSE*, *SSR*, and *SSTO*
- the mean square error, *MSE*
- the variance covariance matrix $\mathbf{s}^2\{\mathbf{b}\}$

(a) Write out the coefficient vector **b**.
(b) What are the **standard errors** of b_0 and b_1?
(c) What are the **covariance** and **correlation** between b_0 and b_1?
(d) Compute the R^2 of this regression and interpret what it means in the context of the data.
(e) If the hardness of plastic sample #2 were increased by 1, how much would the predicted hardness of plastic sample #2 change?
(f) If the hardness of plastic sample #14 were increased by 1, how much would the predicted hardness of plastic sample #2 change?
(g) Show that the vectors **e** and $\hat{\mathbf{y}}$ are orthogonal to each other. What is the correlation between these two vectors?

10. The data in the file **Cereal1.txt** show the sugar content (as a percentage of weight) of a sample of 18 adults' and 18 children's cereal brands. The *Sugar* column contains the sugar content, while the *Type* column contains the type of cereal—either adults' or children's. We want to estimate the mean sugar content for the two different types of cereals. The *R* script **STAT334-HW1-Cereal1.r** contains code to set up the data into the following matrices:

y the response vector containing the *Sugar* variable
X.A a design matrix for the two means parameterization of the model
X.B a design matrix for the effect parameterization of the model (constraint: $\tau_{Adult} + \tau_{Child} = 0$)
X.C a design matrix for the baseline parameterization of the model (constraint: $\tau_{Adult} = 0$)

Use *R* to compute the coefficient estimates for all three versions of the model.

(a) Using version A of the model, interpret both of the estimated coefficients in the context of the problem.
(b) Using version B of the model, interpret both of the estimated coefficients in the context of the problem.
(c) Using version C of the model, interpret both of the estimated coefficients in the context of the problem.

JMP Instructions

The JMP **Analyze menu** has two commands for regression:

1. **Fit Y by X** works only for simple regression (one explanatory variable) and **cannot do confidence intervals for the mean response or predictions**.
2. **Fit Model** works for simple or multiple regression (more than one explanatory variable). It can also do confidence intervals for the mean response and prediction intervals if you enter the x-values for the predictions into the data table (see below).

Since **Fit Model** does more, you should use the **Fit Model** command for every regression.

1. Run the **Analyze>Fit Model** command.
2. Select the response variable from the variable list at the left and click the **Y** button.
3. Select the explanatory variable from the variable list at the left and click the **Add** button.
4. Optional: If you want to do tests or confidence intervals with an α other than 0.05 (or an interval with other than 95% confidence), click the red triangle next to "Model Specification" and select **Set Alpha Level**. Enter the new α level in the blank, then click **OK**.

After the variables are selected, the Personality blank in the window will say "Standard Least Squares" and the Emphasis blank will say "Effect Leverage". Do not worry about these. Do not change them.

5. Click the **Run** button.

A JMP report window for Fit Model will appear. The different parts of this window are described in more detail on page 47 of the *JMP Documentation* PDF in the "Software Help" section of our Blackboard course. I will describe the highlights below.

The **Parameter Estimates** section of the report window contains the estimated coefficients (sample intercept and sample slope), their standard errors, test statistics, and p-values. The asterisk (*) beside a p-value just means that it is less than α, which is 0.05 unless you changed it in the Fit Model dialog above.

To add confidence intervals for the slope and intercept to the parameter estimates table, click on the red triangle at the top of the window (beside the word "Response") and choose **Regression Reports>Show All Confidence Intervals**. This shows only the confidence intervals for the intercept and slope. It does not show confidence intervals for the mean response or prediction intervals. For those, see below.

Copying Plots and Tables from JMP

1. From any JMP output window, click the selection tool (looks like a "+") in the toolbar or use the keyboard shortcut (S).
2. Click on the content you'd like to copy and highlight it. Click near the edge of the report to select all content. To extend a selection, hold the Shift key.
3. **Click Edit > Copy** (or Control-C)
4. Open the program where you'd like to paste the content. If using Windows, select **Paste > Paste Special**. From the list choose **Picture (Enhanced Metafile)**. If using a Mac, choose **Paste Special > PDF**.

R Instructions

Download R and RStudio to your computer

The latest version of R is available at **http://cran.stat.ucla.edu/**.
The latest version of RStudio Desktop is available at **https://www.rstudio.com/ide/download/**
Follow the download instructions at these sites for help. You should install R before installing RStudio.

Importing Data into R using RStudio

1. Launch RStudio.
2. Use the **File>Import Dataset>From Text (base)…** command.
3. Find the file **Real Estate3.csv** and select it.
4. Make sure the settings are **Heading = Yes, Separator = Tab, Decimal = Period, Quote = Double Quote**.
5. Click **Import**.

The data from the file should now be in R's workspace in an object with the same name as the text file.

Running R Scripts

Use the **File>Open File…** command in *R* to open the script.
Make sure that the data is loaded into *R* before running the script. Follow the instructions above to do this.
Use the **Code>Run Region>Run All** command to run the script.
Text output will appear in the Console window. Plots will appear in the Plots tab. You can view previous plots by clicking on the left arrow under the Plots tab.

Evaluating Formulas in R

You can type matrix formulas into a script file or into the R Console beside the prompt (>).
To evaluate formulas typed into a script file, highlight the text of the formula and select **Code>Run Selected Lines** or press **Ctrl+Enter**.
You may find the following R functions useful:

*	multiplies a scalar by another scalar or a matrix
%*%	multiplies a matrix by a matrix
t(X)	transposes the matrix **X**
solve(A)	finds the inverse of a matrix **A** (**A** must be square)

Copying code and plots from RStudio

You can copy output from the Console window in RStudio as regular text. Paste this into Word for printing.

To copy a graph, click on the **Plots** tab in the bottom right pane of RStudio.
From the **Export** menu inside the pane, select **Copy Plot to Clipboard**.
A new window will appear containing the plot. Click Copy as: **Metafile**, then click the **Copy Plot** button.
You can now paste the plot into Word for printing.