STA302/1001 Autumn 2017 Homework #1 v2

11 September. With thanks to Alison Gibbs and Becky Lin

Instructions: These aren't for credit, so don't hand them in. They relate to early parts of Chapter 2 of the textbook, and you should be comfortable with them all after the lectures in Week 3.

1. What's wrong with the following simple linear regression model?

$$E(Y_i|X = x_i) = \beta_0 + \beta_1 x_i + e_i$$

2. (a) For the simple linear regression model, what is the implication if $\beta_0 = 0$ so that the model is $Y_i = \beta_1 x_i + e_i$?

(b) Derive the least squares estimator of β_1 for the model $Y_i = \beta_1 x_i + e_i$.

(c) For a simple linear regression model, what is the implication if $\beta_1 = 0$ so that the model is $Y_i = \beta_0 + e_i$?

(d) Derive the least squares estimator of β_0 for the model $Y_i = \beta_0 + e_i$ and show that it is unbiased.

3. Show:

$$\text{(a)} \quad \sum_{i=1}^{n} \hat{e}_i x_i = 0 \qquad \text{(b)} \quad \sum_{i=1}^{n} \hat{e}_i \hat{y}_i = 0$$

4. Consider a simple linear regression model. Assume all of the standard assumptions hold, and suppose that $\beta_0 = 10$, $\beta_1 = 5$, and $\sigma^2 = 4$.

(a) What is the conditional distribution of $Y|X = x$ when $x = 0$? when $x = 5$?

(b) When $x = 2$, what is the conditional probability that Y is between 16 and 20?

just compute E and Var for Y, then use cdf of normal distribution

5. (Source: Exercise 1.11 in Kutner et al.) The regression function relating production output by an employee after taking a training program (Y) to the production output before the training program (X) is $E(Y|X = x) = 20 + 0.95x$, where x ranges from 40 to 100. An observer concludes that the training program does not raise production output on the average because β_1 is not greater than 1.0. Comment.

6. (Source: Exercise 2.3 in Kutner et al.) A member of a student team playing an interactive marketing game received the following computer output when studying the relation between advertising expenditures (x) and sales (y) for one of the team's products:

- Estimated regression equation: $\hat{y} = 350.7 - 0.18x$
- Two-sided p-value for estimated slope: 0.91

The student stated: "The message I get here is that the more we spend on advertising this product, the fewer units we sell!" Comment.

not statistically significant, so no linear relation (null beta_1 = 0) between expenditures and sales

7. The `oldfaithful.txt` data set (on Portal) contains data on 21 consecutive eruptions of Old Faithful geyser in Yellowstone National Park. It is believed that one can predict the duration of the next eruption (`eruption`) from the time elapsed since the last eruption (`waiting`). That is, Y is the "eruption" and X is the "waiting" in the data set.

```
q2data = read.table("oldfaithful.txt",header=TRUE)
str(q2data)     #check the type of each column (variable) in the data set
```

```
## 'data.frame':    272 obs. of  2 variables:
##  $ eruption: num  3.6 1.8 3.33 2.28 4.53 ...
##  $ waiting : int  79 54 74 62 85 55 88 85 51 85 ...
```

```
head(q2data,10) # have a look of the first 10 data lines
```

```
##    eruption waiting
## 1     3.600      79
## 2     1.800      54
## 3     3.333      74
## 4     2.283      62
## 5     4.533      85
## 6     2.883      55
## 7     4.700      88
## 8     3.600      85
## 9     1.950      51
## 10    4.350      85
```

fit <- lm(eruption~waiting, data=q2data)
summary(fit)

(a) Fit a simple linear regression (show R code).

(b) Show the summary output of the simple linear regression.

(c) The estimated linear regression model is:

$$\widehat{eruption} = b_0 + b_1 waiting$$

What are your estimates for b_0 and b_1?

8. Bonus (because you don't need to know R Markdown in this course). Consult the provided template, `HW1-q8-template.Rmd`.

In part (a), a detailed proof is given to show you how to type a proof with left alignment in R Markdown. Learn from (a), then type your solution of (b) and (c) in the same way. Here is a reference for the Latex code to produce mathematical symbols: http://web.ift.uib.no/Teori/KURS/WRK/TeX/symALL.html