Stat 6021: Homework Set 2

1. (R required) For this question, you will use the dataset "Copier.txt" for this question. This is the same data set that you used in the last homework. The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, *Serviced* is the number of copiers serviced and *Minutes* is the total number of minutes spent by the service person.

   It is hypothesized that the total time spent by the service person can be predicted using the number of copiers serviced. Fit an appropriate linear regression and answer the following questions:

   (a) Produce an appropriate scatterplot and comment on the relationship between the total time spent by the service person and the number of copiers serviced.

   (b) What is the correlation between the total time spent by the service person and the number of copiers serviced? Interpret this correlation contextually.

   (c) Can the correlation found in part 1b be interpreted reliably? Briefly explain.

   (d) Obtain the 95% confidence interval for the slope, β_1.

   (e) Suppose a service person is sent to service 5 copiers. Obtain an appropriate 95% interval that predicts the total service time spent by the service person.

   (f) What is the value of the residual for the first observation? Interpret this value contextually.

   (g) What is the average value of the all the residuals? Is this value surprising (or not)? Briefly explain.

2. (No R required) A substance used in biological and medical research is shipped by airfreight to users in cartons of 1000 ampules. The data consist of 10 shipments. The variables are number of times the carton was transferred from one aircraft to another during the shipment route (*transfer*), and the number of ampules found to be broken upon arrival (*broken*). We want to fit a simple linear regression. A simple linear regression model is fitted using R. The corresponding output from R is shown next, with some values missing.

```
Call:
lm(formula = broken ~ transfer)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2000     0.6633  ______ ________ ***
transfer      4.0000     0.4690  ______ ________ ***

Residual standard error: 1.483 on 8 degrees of freedom

...

Analysis of Variance Table

Response: broken
          Df Sum Sq Mean Sq F value    Pr(>F)
transfer   1  160.0   160.0  ______ ________ ***
Residuals  8   17.6     2.2
```

The following values are also provided for you, and may be used for the rest of this question: $\bar{x} = 1$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 10$.

(a) Carry out a hypothesis test to assess if there is a linear relationship between the variables of interest.

(b) Calculate a 95% confidence interval that estimates the unknown value of the population slope.

(c) A consultant believes the mean number of broken ampules when no transfers are made is different from 9. Conduct an appropriate hypothesis test (state the hypotheses statements, calculate the test statistic, and write the corresponding conclusion in context, in response to his belief).

(d) Calculate a 95% confidence interval for the mean number of broken ampules and a 95% prediction interval for the number of broken ampules when the number of transfers is 2.

(e) What happens to the intervals from the previous part when the number of transfers is 1? (Describe what happens without calculating)

(f) What is the value of the F statistic for the ANOVA table?

(g) Calculate the value of R^2, and interpret this value in context.

3. (No R required) Suppose that the population slope for a straight-line relationship between y and x is 0.

   (a) Describe how the straight line would look in a plot of y versus x.

   (b) Explain why a slope that is equal to 0 would indicate that y and x are not linearly related, and why a slope that is not equal to 0 would indicate that y and x are linearly related.