ST 312: 10.1-10.2 Practice Problems

Ex 1. The following data are from a study of two methods for measuring the blood flow in the stomachs of dogs:

Subject	1	2	3	4	5	6	7	8	9	10
Spheres	4	4.7	6.3	8.2	12	15.9	17.4	18.1	20.2	23.9
Vein	3.3	8.3	4.5	9.3	10.7	16.4	15.4	17.6	21	21.7

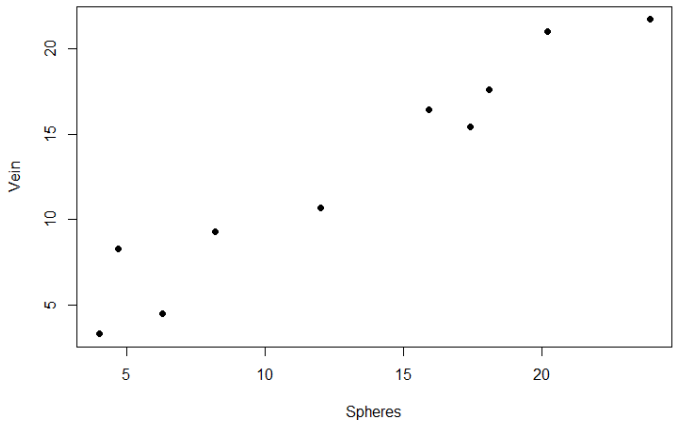
"Spheres" is an experimental method that the researchers hope will predict "Vein," the standard but difficult method. Some statistics about the data (hint: try labeling these first):

$\bar{X} = 13.07,\ \sum_{i=1}^{10}(X_i - \bar{X})^2 = 443.201,$

$\bar{Y} = 12.82,\ \sum_{i=1}^{10}(Y_i - \bar{Y})^2 = 385.256,$

$\sum_{i=1}^{10}\left(\hat{Y}_i - \bar{Y}\right)^2 = 360.5697$, and

$\sum_{i=1}^{10}(X_i - \bar{X})(Y_i - \bar{Y}) = 399.756.$



(1) Find the regression line using least square methods.

(2) What is the proportion of the variation in Vein measurements can be explained by Spheres measurements?

(3) Compute the residual SD, s_e, and interpret its value.

(4) We expect the measurements of the two methods to be positively associated. State hypotheses in terms of the slope of the regression line that express this expectation, and carry out a significance test at $\alpha = 0.01$. What conclusion do you draw?

(5) Find a 99% confidence interval for the slope.

**Note: The results in (4) and (5) may not agree with each other (though here they happen to give the same results) because (5) is equivalent to a 2-tail test while (4) is a right tail test.*

(6) Suppose that we observe a value of Spheres equal to 15.0 for one dog. Give a 99% interval for predicting the Vein value for that dog. (Fill in the missing values in the formula below and find the final interval)

$$99\%\ \_\_\_\_\_\ \text{for}\ Y = \hat{Y} \pm \left|t_{df,\frac{\alpha}{2}}\right| \times \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right)}$$

$$= \_\_\_\_\_ \pm \_\_\_\_\_ \times \sqrt{\_\_\_\_\_\left(1 + \frac{1}{10} + \frac{(15 - 13.07)^2}{443.201}\right)}$$

$$= \_\_\_\_\_ \pm \_\_\_\_\_ \times 1.85 =$$

(7) (True / False) A <u>prediction interval</u> can be used to predict the mean Vein values for those dog with Spheres equal to 15.0.

(8) (True / False) The SE of $\hat{Y}$ for $X = 15$ is the same as the SE of $\hat{Y}$ for $X = 20$.

Ex. 2. Finch Data. To examine the relationship between the wing length (in mm) and the weight (in grams) of finch, a random sample of finch were collected and a linear model was fit to the data. Below is some output.

Regression Statistics	
Multiple R	0.540496
R Square	xxxxx
Standard Error	1.972602
Observations	66

Regression Model Output			
	Coefficients	*Standard Error*	*t Stat*
Intercept	55.0649	2.749791	20.02512
Weight	0.847126	0.164831	5.139344

(1) In the output above, “Multiple R” is the correlation coefficient, r. Describe the relationship (strength & direction) between wing length and weight.

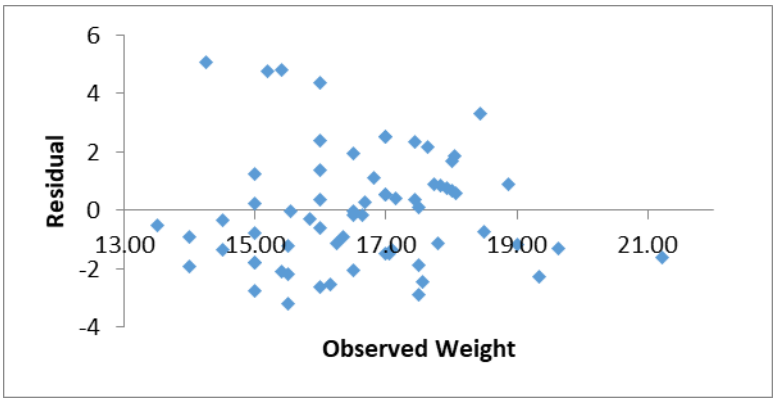
(2) Which variable is the response variable? Use the Regression Model Output to decide.

(3) State the equation of the least squares regression line.

(4) Interpret the “Weight” coefficient.

(5) The r^2 value is missing from the Regression Statistics above. Find it and interpret its value.

(6) Below is a residual plot. Does the plot suggest that any assumptions have been violated? What are we looking for?



(7) In the Regression Statistics above, the “Standard Error” value is the residual standard error. Find the SSE (the sum of squares error).

(8) Is weight a significant predictor of wing length? Use $\alpha = 0.05$ to decide.

(9) Can we predict the expected wing length of a finch whose weight is 16 grams? If so, find it. If not, why?

(10) Can we predict the expected wing length of a finch whose weight is 26 grams? If so, find it. If not, why?

(11) Can we predict the wing length of all finch whose weight is 16 grams? Explain.