

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

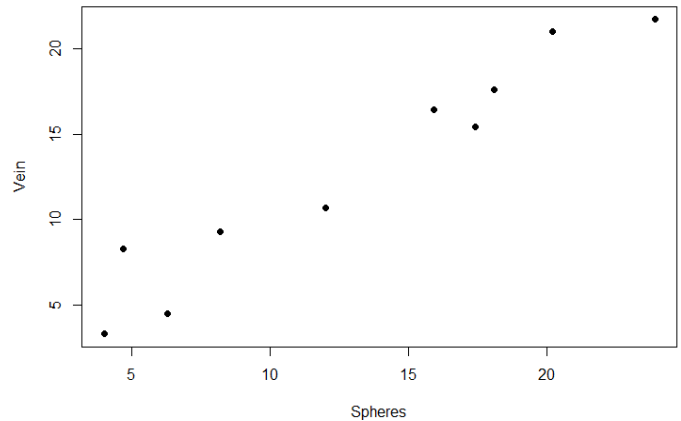
“Spheres” is an experimental method that the researchers hope will predict “Vein,” the standard but difficult method. Some statistics about the data (hint: try labeling these first):

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

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

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

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



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

***Note: the scatterplot (which is step 1) shows a linear pattern.**

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{399.756}{443.201} = 0.902$$

$$b_0 = \bar{Y} - b_1 \bar{X} = 12.82 - 0.902 * 13.07 = 1.03$$

$$\hat{Y}_i = 1.03 + 0.902X_i \text{ (or } Y_i = 1.03 + 0.902X_i + e_i)$$

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

$$R^2 = \frac{SSM}{SST} = \frac{360.5697}{385.256} = 0.94$$

About 94% of the variation in Vein measurement can be explained by the linear relationship with spheres.

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

$$s_e = \sqrt{MSE} = \sqrt{3.09} = 1.758$$

On average, the observed vein measurements are about 1.758 units away from the value predicted by the spheres measurements.

(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?

$$H_0: \beta_1 \leq 0$$

$$H_A: \beta_1 > 0$$

$$\alpha = 0.01$$

$$t^* = \frac{b_1 - \beta_1}{\hat{\sigma}_{b_1}} = \frac{0.902 - 0}{\sqrt{\frac{3.09}{443.201}}} = 10.8$$

$$df = 8$$

$$p\text{-value} = P(t_8 > 10.8) = 0$$

Since the p-value is less than 0.01, we Reject the null hypothesis. There is evidence at the 1% significance level to suggest the two methods are positively associated.

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

$$t_{8, 0.005} = 3.355$$

$$99\% \text{ CI for } \beta_1: b_1 \pm |t| * \hat{\sigma}_{b_1}$$

$$= 0.902 \pm 3.355 \sqrt{\frac{3.09}{443.201}}$$

$$= (0.62, 1.18)$$

Notice this interval is above 0.

**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.*

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

$$\hat{Y}_i = 55.0649 + 0.847126X_i$$

(4) Interpret the “Weight” coefficient.

Weight is used in this equation as the explanatory variable, so the weight coefficient ($\hat{\beta}_1 = b_1$) is the slope of the regression equation:

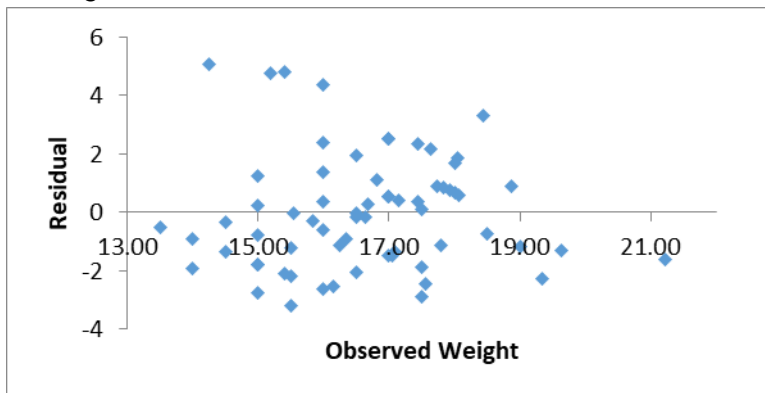
For every additional gram of weight, the expected wing length of a finch increases by 0.847126 mm.

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

$$r^2 = R^2 = (r)^2 = (0.540496)^2 = 0.292$$

About 29.2% of the variation in wing length can be explained by its linear relationship with weight.

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



We are looking for patterns (or lack thereof) in the residual plot. A clear pattern (line/curve) indicates that the data are not linear. A clear pattern (fan-shape) indicates that the errors do not have constant variance (heteroscedasticity).

No obvious patterns emerge in the above plot, so we do not have reason to believe the assumptions of linear data and homoscedasticity (equal variance) are violated.

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

$$s_e = \sqrt{\frac{SSE}{n-2}}$$

$$1.972602 = \sqrt{\frac{SSE}{64}}$$

$$SSE = (1.972602 * 8)^2 \approx 249.03$$

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

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

$$\alpha = 0.05$$

From the Regression Model Output above, the test statistic for weight is $t^* = 5.1$ with $df = n - 2 = 64$.

$$p\text{-value} = 2 * P(t_{64} > 5.1) \approx 0$$

Since the p-value is less than 0.05, we reject the null hypothesis of no linear association. There is enough evidence at the 5% significance level to suggest that weight has a significant linear relationship with wing length.

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

According to the residual plot, the observed weight data ranges from about 13-22 grams. So we can predict for a weight of 16 grams by substituting into the regression equation:

$$\hat{Y}_i = 55.0649 + 0.847126 * 16 = 69.62 \text{ mm}$$

The expected wing length of a finch whose weight is 16 grams is about 69.62 mm.

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

According to the residual plot, the observed weight data ranges from about 13-22 grams. So we cannot predict for a weight of 26 grams. That would be extrapolation. We do not know if the linear pattern continues beyond 21-22 grams of weight.

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

We cannot make a single prediction for the population of all finch whose weight is 16 grams. The regression equation estimates the mean response (mean wing length). A prediction interval estimates the response value for an individual future value, and may be applied to any finch whose weight is 16 grams. It does not estimate the range of the population, nor give a single prediction value.