Example: Orange Juice Sweetness

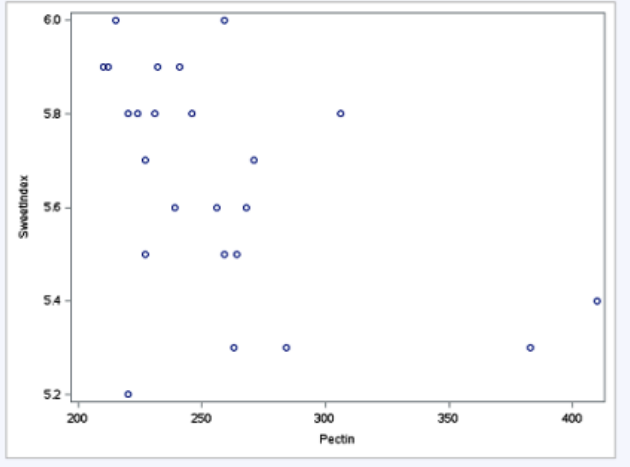
STAT 705, Simple Linear Regression, Part 6

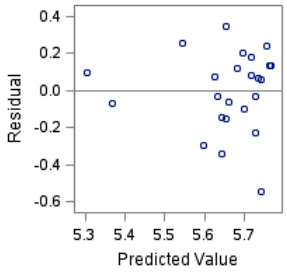
Answers

1. The linear regression model is $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
   (This is directly from the PowerPoint slide in Part 1 of Simple Linear Regression.)

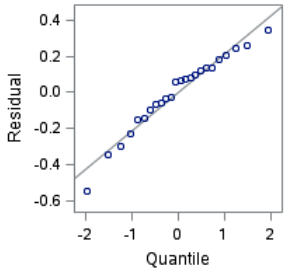
2. We are told that the manufacturer wants to "predict the sweetness from the amount of pectin", so the predictor (X) is the amount of pectin and the response (Y) is sweetness.

3. From the scatterplot, there appears to be a weak linear relationship between these two variables. A linear model does seem reasonable.



4. 

The residual plot has no obvious pattern.



The points on the normal probability plot follow the line fairly closely.

There are no clear violations of the assumptions.

5. The estimated regression equation is Sweetness = 6.252 – 0.00231*Pectin

6. For the ANOVA F test

   a. The hypotheses are H0: $\beta_1 = 0$ vs. Ha: $\beta_1 \neq 0$
   b. The test statistic is F = 6.52
   c. To find the critical value, we use α = 0.05. For the F distribution, the degrees of freedom for the numerator is 1 and the degrees of freedom for the denominator is 22. We do not have 22 degrees of freedom in our F table, but we do have 20 and 24. For 20, the critical value is 4.35 and for 24 the critical value is 4.26. We can deduce that for 22 denominator degrees of freedom, the critical value is somewhere between 4.35 and 4.26.
   d. The p-value is p = 0.0181
   e. Since the p-value is smaller than 0.05, we reject H0 and conclude that the slope β_1 is not 0. In other words, we conclude that there <u>is</u> a linear relationship between these two variables.

7. Consider the t test for H0: $\beta_1 = 0$ vs. Ha: $\beta_1 \neq 0$.

   a. The test statistic is t = –2.55
   b. The critical value is 2.074 (This is from the t table with 22 degrees of freedom and $\alpha/2$ = 0.025)
   c. The p-value is p = 0.0181
   d. Interpret the results of this test.
      Since the p-value is smaller than 0.05, we reject H0 and conclude that the slope β_1 is not 0. In other words, we conclude that there <u>is</u> a linear relationship between these two variables.

8. For a single run of orange juice that has 250 parts per million pectin

   a. a point estimate for the sweetness index is 5.6744
   b. A 95% prediction interval for the sweetness index is (5.2192, 6.1297)

9. For all runs of orange juice that have 250 parts per million pectin

   a. a point estimate for the mean sweetness index is 5.6744
   b. a 95% confidence interval for the mean sweetness index is (5.5825, 5.7664)