

## Estimation of Means and Prediction of Values in Simple Linear Regression

Can a low barometer reading be used to predict maximum wind speed of an approaching tropical cyclone? For seven tropical cyclones, the following data was recorded.

$X$  = the lowest barometric pressure reading (in millibars) during the approach of the cyclone, and

$Y$  = the maximum wind speed (in miles per hour) of the cyclone when it reached the shore.

The data and some summary statistics are provided below.

|     | Plot |     |     |     |     |     |     | Summary Statistics |        |        |
|-----|------|-----|-----|-----|-----|-----|-----|--------------------|--------|--------|
|     | 1    | 2   | 3   | 4   | 5   | 6   | 7   | Mean               | SD     | r      |
| $X$ | 1004 | 975 | 992 | 935 | 985 | 932 | 953 | 968                | 28.343 | -0.988 |
| $Y$ | 41   | 100 | 65  | 145 | 80  | 151 | 132 | 102                | 42.284 |        |

1. A scatter plot of minimum pressure vs. maximum wind speed suggests that a linear relationship may hold between these variables. Compute the equation of the least-squares regression line for predicting maximum wind speed from minimum pressure.
2. Suppose that a tropical cyclone is headed for shore and that the barometric pressure readings seem to have hit a minimum at 945 millibars. Predict the maximum speed of the cyclone when it reaches the shore.
3. Your answer to question 2 is an estimate of some population parameter. Describe in words what population parameter is being estimated.

4. Find a 95% confidence interval for the mean of the population of  $Y$  values that share the value of  $X = 945$  millibars with the help of the following facts:

$$SE(\hat{\beta}_0 + \hat{\beta}_1 X) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{(n-1)s_x^2}} \quad t = \frac{(\hat{\beta}_0 + \hat{\beta}_1 X) - (\beta_0 + \beta_1 X)}{SE(\hat{\beta}_0 + \hat{\beta}_1 X)} \text{ with } d.f. = n - 2.$$

5. It is nice to have a range of plausible values for the average of the maximum wind speeds of tropical cyclones like the one approaching shore (those with minimum barometric pressures of 945 millibars). This information is provided by the confidence interval computed in question 4. What would be even better is a range of plausible values for the maximum wind speed of the particular tropical cyclone that is currently approaching the shore. Such a range of values is known as a **prediction interval**.
6. A  $100(1 - \alpha)\%$  prediction interval for a single value of  $Y$  corresponding to a particular value of  $X$  is

$$\hat{\beta}_0 + \hat{\beta}_1 X \pm t_{n-2}^{(1-\alpha/2)} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{(n-1)s_x^2}}.$$

Find a 95% prediction interval for the maximum wind speed of the oncoming tropical cyclone.

7. A 95% confidence region for the population regression line  $\mu\{Y|X\} = \beta_0 + \beta_1 X$  is given by

$$\hat{\beta}_0 + \hat{\beta}_1 X \pm \sqrt{2F_{2,n-2}^{(0.95)}} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{(n-1)s_x^2}}.$$

We can be 95% confident that these bounds will contain the population regression line over the observed range. Sketch a picture of a least-squares regression line with its 95% confidence region. Where is the confidence region the narrowest?