Math 6010, Fall 2004: Homework

Homework 1 Consider the following hypothetical data; it is regarding a particular bird species:[1]

Wing length (cm)	Age (days)
1.5	4.0
2.2	5.0
3.1	8.0
3.2	9.0
3.2	10.0
3.9	11.0
4.1	12.0
4.7	14.0
5.2	16.0

Use simple linear regression to predict the age of a newly-caught bird whose wing is 4.0 centimeter long. Assuming normal errors, test the hypothesis (95%) that wing length and age are linearly related.

Solution. Let the ages be denoted by $y_1, \ldots, y_9$, and the wing-lengths by $x_1, \ldots, x_9$. You should compute:

$$\begin{array}{lll} \bar{x} \approx 3.46 & s_{\boldsymbol{x}} \approx 1.17 & \\ \bar{y} \approx 9.89 & s_{\boldsymbol{y}} \approx 3.92 & \mathrm{Corr}(x,y) \approx 0.99. \end{array}$$

Now recall that the regression line is described by

$$y = \bar{y} + \widehat{\beta}(x - \bar{x}),$$

where

$$\widehat{\beta} = \mathrm{Corr}(x,y)\,\frac{s_{\boldsymbol{y}}}{s_{\boldsymbol{x}}} \approx 3.3.$$

So the regression line is (approximately) given by

$$y = 9.89 + 3.3(x - 3.46).$$

Plug in $x = 4$ to obtain the regression estimate:

$$y = 9.89 + 3.3(4 - 3.46) \approx 11.21.$$

This answers the first question.

To answer the second question recall that

$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{n s_{\boldsymbol{X}}^2}\right).$$

[1]Borrowed from `http://math.hws.edu/javamath/ryan/Regression.html`

Here, $n = 9$. Pretend that this is large enough. [This is not so good, and will be addressed later on.] Then, one would expect that $\sigma^2 = \mathrm{Var}(\boldsymbol{y}) \approx s_{\boldsymbol{y}}^2 \approx 3.92^2 \approx 15.37$. Also, $s_{\boldsymbol{x}}^2 \approx 1.17^2 \approx 1.37$. So, we would expect $\widehat{\beta}$ to have approximately a normal distribution with mean β and variance

$$\frac{15.37}{9 \times 1.37} \approx 1.25.$$

Under H_0,

$$\widehat{\beta} \approx N\left(0, 1.25\right).$$

Therefore, the sample's $\widehat{\beta} = 9.89$ gives us a P-value of ≈ 0.0. I.e., good reason to believe that H_0 is false; i.e., $\beta \neq 0$; i.e., there is some linear relation between x and y.