STAT 652-601
Instructor: Jeff Hart

Assignment #2

The data used in this assignment were obtained in a study of factors that affect diabetes in children (Sockett, et al. (1987), Diabetes, 30, 453–459). The dependent variable is C-peptide concentration at diagnosis, while the two independent variables are age and base deficit, a measure of acidity. The number of points in the data set is $n = 43$.

You may obtain the data via the course website by clicking on “Peptide data” under Data sets. The data set is in three columns of length 43 each. The first column is age (in years), the second column is base deficit and the third is C-peptide concentration.

Please do the following:

1. Find least squares estimates for the model
   \[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \]
   where $Y = \log(\text{C-peptide concentration})$, $x_1 = \text{age}$, and $x_2 = \text{base deficit}$.

2. Perform a test of the null hypothesis that neither age nor base deficit is needed in the model. (Use $\alpha = .05$.)

3. Does it appear that a model with just one of the two independent variables would suffice? Justify your answer.

4. Using the model from 1., find a 95% confidence interval for average $\log(\text{C-peptide concentration})$ for children of age 6 years having base deficit -5.5. Give an interpretation of this interval and state all the assumptions needed for the interval to be valid.

As always, make sure you answer the questions asked in the language of the problem, using your computer output as support for your answer.

This assignment is due Monday, September 22.