Homework set #3 (100 points total)

Numerical Exercises: (20 points)

A. (10 points) A simple linear regression analysis performed on a sample of n=24 points provides the following information: \( b_0 = 1, b_1 = 2, s(b_0) = 0.05, s(b_1) = 0.25, \) \( \text{SST} = 117,287, \) \( \text{SSE} = 30.0 \). Based on this: (i) Test \( H_0: \beta_1 = 0 \) vs \( H_1: \beta_1 \neq 0 \) (select a test statistic, compute its value and use its null distribution to find a p-value; explain your work). (ii) Compute \( R^2 \), and explain the meaning of the figure you obtain.

B. (10 points) Fill in the missing entries in this ANOVA table, relative to a simple linear regression:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>??</td>
<td>1000</td>
<td>??</td>
</tr>
<tr>
<td>Error</td>
<td>??</td>
<td>800</td>
<td>??</td>
<td>??</td>
</tr>
<tr>
<td>Total</td>
<td>81</td>
<td>??</td>
<td>??</td>
<td>??</td>
</tr>
</tbody>
</table>

Using the value of the F statistic and its null distribution, find a p-value for \( H_0: \beta_1 = 0 \) vs \( H_1: \beta_1 \neq 0 \) (explain your work).

“Body fat” data analysis: (80 points)

Please refer to the data description and general guidelines file when performing the following analyses and preparing your write-up.

A. (30 points) For the simple linear regression of body fat percentage on abdomen circumference, produce again the standard output with least square fit information, tests for the regression coefficients, ANOVA table, coefficient of determination, etc. In addition produce the following graphical and test-based diagnostics:
(i) Plot of regular residuals versus the predictor, lack of fit test (with data subsetting), and Breusich-Pagan test for non-constant variance.
(ii) Normal probability plot for regular residuals, with confidence band and AD test.
(iii) Plot of standardized residuals versus the predictor, and “Unusual observations” output. Explain and comment upon any problems you might identify through these diagnostics. If you find outliers that you wish to remove from the data, re-run both the regression and the diagnostics. Conclude with a general evaluation of the simple linear regression model based on all the information you gathered.

B. (30 points) Repeat the analysis in A for the regression of forearm circumference on wrist circumference.

C. (20 points) Back to body fat, plot the regular residuals from the simple linear regression of body fat percentage on abdomen circumference (in A) versus all other quantitative variables in the data set that might be used as additional predictors; Weight, Height, NeckC, ChestC, HipC, ThighC, KneeC, AnkleC, BicepsC, ForearmC, WristC. Is there any visual evidence that other variables might capture something that is left out by abdomen circumference? Do these make sense – i.e. are they proxies for factors useful in predicting body fat, which are not conveyed by abdomen circumference?