1. Redraw the plot, enabling plot labels. Highlighting the largest and the smallest brain and body weight brings up the African elephant and the lesser short-tailed shrew.

Change the mouse mode to ‘show coordinates’ and identify humans as the medium-sized creatures with above-average brain weight. Clicking on this point gives the label and coordinates ‘Human  4.1271   7.1854’

2. The first two plots were shown in class – they are

Making the log10 transformation of pressure, replotting and removing linear tend gives

The formal regression fit gives the output

Normal Regression
Kernel mean function = Identity
Response      = log10[Pressure]
Terms         = (Temp)
Coefficient Estimates

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.413347</td>
<td>0.0100331</td>
<td>-41.198</td>
<td>0.0000</td>
</tr>
<tr>
<td>Temp</td>
<td>0.00891110</td>
<td>0.0000494410</td>
<td>180.237</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
R Squared: 0.999569  
Sigma hat: 0.00113596  
Number of cases: 17  
Number of cases used: 16  
Degrees of freedom: 14

Summary Analysis of Variance Table
<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>0.0419193</td>
<td>0.0419193</td>
<td>32485.39</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>14</td>
<td>0.0000180657</td>
<td>1.290404E-6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and from this we get the 95% confidence interval for the slope as 0.0089111 +/- 2.14479*0.000049441, or 0.00891 +/- 0.000011, or (0.00880,0.00902)

The 2.14479 came from the ‘calculate quantile’ capability of Arc, getting the 0.975 quantile of a t with 14 degrees of freedom. We don’t need all these trailing digits; I just list them to make it easier to verify that we are seeing exactly the same things.

Next, fitting the model using the theoretically correct relationship gives

Data set = Forbes, Name of Fit = L2  
Deleted cases are (11)  
Normal Regression  
Kernel mean function = Identity  
Response = log10[Pressure]  
Terms = (u1)  
Coefficient Estimates  
<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>7.31488</td>
<td>0.0316602</td>
<td>231.043</td>
<td>0.0000</td>
</tr>
<tr>
<td>u1</td>
<td>-2179.00</td>
<td>11.6517</td>
<td>-187.012</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R Squared: 0.9996  
Sigma hat: 0.00109482  
Number of cases: 17  
Number of cases used: 16  
Degrees of freedom: 14

Summary Analysis of Variance Table
<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>0.0419206</td>
<td>0.0419206</td>
<td>34973.46</td>
<td>0.0000</td>
</tr>
<tr>
<td>Residual</td>
<td>14</td>
<td>0.000016781</td>
<td>1.198639E-6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(As a matter of interest, Forbes’ theoretically wrong model fits almost as well as the correct one; the sigma-hat being 0.00114 versus 0.00109, a 5% difference).

We test the hypothesis \( \eta_1 = -2111 \) with the t-distributed quantity \((-2179+2111)/11.7617 = -5.78\). This is way out at the edge of a t distribution with 14 degrees of freedom – its P value is 2*2.38352e-05. So the theory and the empiric observations are incompatible.

Why? We can only guess; perhaps there was a problem with Forbes’ barometer and/or thermometer being out of calibration. Miscalibration is a source of subtle errors you can not see using data sets collected from the same instruments.
6.8.1  The fitted regression from Arc gives
Data set = Geyser, Name of Fit = L1
Normal Regression
Kernel mean function = Identity
Response = Interval
Terms = (Duration)

Coefficient Estimates

<table>
<thead>
<tr>
<th>Label</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>33.9878</td>
<td>1.18122</td>
<td>28.774</td>
<td>0.0000</td>
</tr>
<tr>
<td>Duration</td>
<td>0.176863</td>
<td>0.00535212</td>
<td>33.045</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

R Squared:               0.802941
Sigma hat:                6.00355
Number of cases:             270
Degrees of freedom:          268

I suspect the non-technical staff at the
Visitor Center would be most helped by
a little table predicting the interval from
duration.  Such a table, spanning the
durations actually seen in the data, might
be (I have rounded the interval to the
nearest minute, so as not to drown the
message in spurious detail.)

We also would like to give some idea of the uncertainty in the prediction.  A rough and
ready non-technical picture might come from a 95% confidence interval.  With this large
sample we don’t need to get too precise; it would suffice to say that we would expect the
actual interval to be within ± 12 minutes of the tabled prediction.

6.8.2  The Arc prediction capability gives
Data set = Geyser, Name of Fit = L1
Normal Regression
Kernel mean function = Identity
Response = Interval
Terms = (Duration)

Term values = (250)
Prediction = 78.2035, se(pred) = 6.01849, weight = 1
Leverage = 0.0050, Max(h_i) = 0.0140
Estimated population mean value = 78.2035, se = 0.423797

The 95% t quantile is 1.97, so the 95% confidence interval for the mean would be 78.20 ±
1.97*0.424, which is 78.20 ± 0.84

6.8.3  This same calculation gives the 95% prediction interval as 78.2 ± 1.97*6.02,
which is 78.2 ± 11.9

The estimate of the 90% conditional quantile is 78.2 + 1.28*6.02 = 85.9

The checks on the summary statistic are self-verifying.