Data Exploration
Salamanders on the rocks
(An exciting new libation?)

The goal of this assignment is fourfold:

- Increase your experience with analyzing bivariate data
- Consider different plausible models for bivariate data
- Use transformations of variables
- Solicit your judgments and opinions along the way

We will ask you to calculate statistics and make graphs, and then write coherently about what you see and what you think about what you are seeing. There are not necessarily correct answers except those involving the statistics and graphs. Your judgments will be evaluated on their plausibility given the data, NOT on some external standard of right/wrong, known to scientists but not to us.

Generally, I would suggest working in teams of two. Hand in your final result electronically. The file name should be “OneOfYourNames__SallyMander.”

We have numbered the questions and procedures below not because they represent the “correct sequence” in analysis of the data, but because we want to be able to refer to them in our discussions. (As you know, the exploratory analysis of data is less an a priori well-defined singular path and more of a let-the-data-lead-me-wherever endeavor.)

With my incredibly vast Microsoft Word capability I have inserted text boxes for you to copy-and-paste, or provide answers to graphs and statistics respectively. Consider this a formal writing assignment. Spelling, punctuation, complete sentences – you know, all that English teacher stuff – will form part of the evaluation. Also correct symbols, labeling of graphs and scales -- you know, all that statistics stuff – will be a part of the evaluation of your work. The sizes of the text boxes do not indicate how much you should write; please feel free to write more, or less than is indicated by the size of the text boxes.

The titles of your graphs should make it clear what is being presented, and especially which data (control/experimental, negative/positive targets) are represented. Labels should clearly indicate the variables and units of the data.
The population density is to some extent a function of the availability of resources (food, safe environment, etc.) It certainly makes sense that a population would be denser in an area that had enough food to support it, and available burrows and covering vegetation. The black-bellied salamander, *Desmognathus quadramaculatus*, is usually found under rocks in mountain streams. Could the rocks be important places giving cover and protection against predators, or otherwise important to these creatures?

One way to uncover (so to speak) the relationship, if any, between rock density and salamander preference is to perform an experiment. Investigators created a range of habitats for salamanders by placing different sized rocks and pebbles in a small stream less than 1 meter in depth in the Southern Appalachian Mountains. They cleared small plots in the stream by removing rocks, leaves, and wood from areas 200 cm by 70 cm. Then they re-populated the plots with rocks of different sizes. Three months later, they returned to measure the population density of the salamanders. The scatterplot of their original data are shown in the table. The rock densities are measured in numbers of rocks / per plot, the salamander densities in numbers of salamanders per plot.

<table>
<thead>
<tr>
<th>Rock density (No/plot)</th>
<th>Salamander Density (No/plot)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35.1</td>
<td>5.0</td>
</tr>
<tr>
<td>35.3</td>
<td>4.0</td>
</tr>
<tr>
<td>53.2</td>
<td>6.2</td>
</tr>
<tr>
<td>58.8</td>
<td>6.5</td>
</tr>
<tr>
<td>73.5</td>
<td>7.0</td>
</tr>
<tr>
<td>69.7</td>
<td>8.0</td>
</tr>
<tr>
<td>70.2</td>
<td>8.9</td>
</tr>
<tr>
<td>85.8</td>
<td>9.2</td>
</tr>
<tr>
<td>101.3</td>
<td>10.4</td>
</tr>
<tr>
<td>139.5</td>
<td>12.0</td>
</tr>
<tr>
<td>139.5</td>
<td>9.0</td>
</tr>
<tr>
<td>35.1</td>
<td>5.0</td>
</tr>
</tbody>
</table>

1. Our first step, of course, is to make pictures of the data.

   a) Construct a scatterplot of the data and generally describe any pattern that you see.
b) Does it appear to you that a linear best-fit line would adequately describe the relationship between rock density and salamander density? Explain.

More than one plausible function might be used to fit the data. Possible model candidates for a fit include:

\[ y = \alpha + \beta \ln x + \varepsilon \]
\[ y = \alpha + \beta \sqrt{x} + \varepsilon \]
\[ y = \alpha + \beta \left( \frac{1}{x} \right) + \varepsilon \]
\[ y = \alpha \beta^x + \varepsilon \]
\[ y = \alpha e^{\beta x} + \varepsilon \]

Each of these candidates will require a transformation or transformations of the variable(s). For each candidate model above,

- indicate which variable(s) have been transformed
- fit a linear model to the transformed data
- copy and paste the scatterplots and residual plots.

2. First, consider the models that only require a transformation of the explanatory variable.

a) \[ y = \alpha + \beta \ln x + \varepsilon \]

Scatterplot Residual plot

b) \[ y = \alpha + \beta \sqrt{x} + \varepsilon \]

Scatterplot Residual plot
3. Evaluate the three models in (2) above. Which model gives the best fit to the data? Justify your responses in a few sentences, appealing to specific statistics and/or aspects of your plots.

4. Now consider the models that require a transformation of the response variable.

Each of these candidates will require a transformation or transformations of the variable(s). For each candidate model below,

- indicate which variable(s) have been transformed
- fit a linear model to the transformed data
- copy and paste the scatterplots and residual plots.

a) \( y = \alpha x^\beta + \epsilon \) (Power model)

b) \( y = \alpha e^{\beta x} + \epsilon \) (Exponential model)
5. Evaluate the two models in (4) above. Which model gives the best fit to the data? Justify your responses in a few sentences, appealing to specific statistics and/or aspects of your plots.

6. The streambed is a nonrenewable resource and the number of salamanders that can be sustained is limited. (This limit is the “carrying capacity” of the stream.)
   
a) Which, if any, of these models “make sense” in the light of this finite carrying capacity? What aspect of the model, the statistics and/or the plots are important in determining a model that makes sense?

b) For the model(s) that you feel makes sense, estimate the population carrying capacity of the stream.

c) For the sensible models in part (b) construct a 95% confidence interval or intervals for the population carrying capacity or capacities.

7. If more than one of the models made sense to you, which one would be your top choice? Justify this choice in a few sentences.