Interactive Excel Tool for Computing, Assessing, and Using
EmP Ws Equations

Tool Philosophy

Rather than “input data here”, and “get Ws equation there”, this tool was designed to allow the researcher to participate (by observing or intervening) in the process of setting up the data for equation estimation. In this way, it is hoped that the resulting equation will have less of a “black box” feel, and feel more like a sequence of understandable steps. Excel was chosen as the platform for the tool since it is ubiquitous, and it has graphical and interactive capabilities that most other software packages don’t (in combination, at least) have.

Tool Elements (each element has its own worksheet)

**Raw data:** Simply, a place to organize the length-weight data in a manner that facilitates data quality assurance and analysis.

**Data quality:** The data from a selected sampled population is graphed so that unusual values and other data quality issues can be detected. This sheet is kept hidden unless the user invokes it from the data sheet.

**Collation:** Mean weights for fish in each length class from each study population are recorded. In addition, the number of populations with fish in each length class is noted.

“**Number of fish**”: This sheet, hidden unless invoked by the user from the collation worksheet, tallies the number of fish in each length class from each study population.

**Summarized:** This sheet sets up the data for estimation of quartiles (1\(^{st}\), 2\(^{nd}\), and 3\(^{rd}\)) and means of mean weights for each length class.

**EmP-Ws:** On this sheet, the summarized data is organized for estimation of Ws equations (in particular, length classes with insufficient data) are removed from consideration, and the resulting Ws equations are displayed.

**Ws graphs:** Plots of the resulting equations are done on the log-log and original scales.

**Bootstrap:** A bootstrap study is done to assess precision of the median Ws equations.

**Units:** this sheet allows the user to change measurement units; updated Ws equations and estimates of Ws are then provided.

**Application:** For a chosen study population, scatter plots are drawn of Wr for individual fish, and for means from selected length categories.

**Instructions for Use:**
Raw data:

1. Enter the number of fish populations for which you have length-weight data; the set up datasheet button will organize the requisite number of data entry columns.

2. Enter your data (hopefully, you can cut & paste it from another file).

3. Click the Min & Max button to find the smallest and largest observed lengths in your entire data set. This will be useful in establishing an initial set of length classes to consider.

4. Prior to proceeding with the analysis, we recommend you spending some time with the data quality sheet so that you can be very confident in the validity of the raw data.
Data quality:

1. Enter the number (numbers are those assigned on the raw data sheet) for the population you want to look at.
2. Click to collect the data.
3. A numerical summary of the log(W)-log(L) regression for that population is given in the table on the left of the screen.
4. The data values are plotted in scatter plots with the log-log regression from the entire data set drawn on them to provide a reference point. Example: If the majority of the points for one population fall below (or above) the population regression line, that might be due to length having been measured by a different protocol for that population. At any rate, you can detect such things as well as individual gross data entry errors using this sheet.
Collation:

1. Enter a value for the lower limit of the lowest length class, and a choice for the longest length that you want to consider (here, with 1062 mm selected, the largest length class is from 1060 mm to 1070 mm). Also indicate the bin width (10 mm is usual for most studies; some researchers working with juvenile fish have used 5 mm).
2. **Click** the **collate** button. The mean weight for fish in each length class, from each population, will be recorded.
3. Note for this illustration that the length classes for lengths less than 250 mm are represented by very few populations. We’ll return to this when we discuss the bootstrapping sheet.
4. To take the sample size examination, a little deeper, **click** the **Number of fish** button.
“Number of fish”:

<table>
<thead>
<tr>
<th>Column A contains the total number of fish in each class, summed over all populations.</th>
<th>Length Bins</th>
<th>[Set 1]</th>
<th>[Set 2]</th>
<th>[Set 3]</th>
<th>[Set 4]</th>
<th>[Set 5]</th>
<th>[Set 6]</th>
<th>[Set 7]</th>
<th>[Set 8]</th>
<th>[Set 9]</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>End</td>
<td>Mid</td>
<td>Numfish</td>
<td>Numfish</td>
<td>Numfish</td>
<td>Numfish</td>
<td>Numfish</td>
<td>Numfish</td>
<td>Numfish</td>
<td>Numfish</td>
<td>Numfish</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>16</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>1</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
<td>210</td>
<td>220</td>
</tr>
<tr>
<td>3</td>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
<td>210</td>
<td>220</td>
<td>230</td>
</tr>
<tr>
<td>4</td>
<td>140</td>
<td>150</td>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
<td>210</td>
<td>220</td>
<td>230</td>
<td>240</td>
</tr>
</tbody>
</table>

This sheet, hidden unless invoked by the user from the collation worksheet, tallies the number of fish in each length class from each study population. On the left of the sheet is the total number of fish in each length class, summed over all the study populations. Note, in this illustration, the paucity of data for the smaller length classes.
1. Click to have the data re-organized for computation of summary statistics.
2. This sheet sets up the data for estimation of quartiles (1st, 2nd, and 3rd) and means of mean weights for each length class. The available means for each length class are sorted, and laid out in the relevant row (at this point, connection to specific study populations is not retained).
3. The mean and median are computed for all length classes with at least one mean, and the 1st and 3rd quartiles are computed for length classes with three or more means.
4. You are likely tired, by now, of our noting the paucity of data for shorter length classes, so we won’t bother doing that again.
5. Turn to the EmP-Ws sheet for results.
1. For sake of keeping neatly organized, one more data summary is done, namely to organize the relevant data (excluding length classes with too little or no data) for each of the target summary statistics.

2. Once done, the regression equations are computed for each summary statistic (all three quartiles and the mean). The choice of using a linear regression or a quadratic is given (one of us, at least, is inclined to always use a quadratic, but that choice properly belongs to the researcher). As an aid to assessing the utility of a quadratic, the standard error of the quadratic term is displayed.

3. The yellow buttons that take you to other worksheets are self-explanatory.
Plots of the resulting equations are done on the log-log and original scales. Michael and Kate:
I’m not sure if this sheet is all that useful, so I’m not going to elaborate on this one just now…
Bootstrap:

Notes to M & K:

1. The bootstrapping currently does only 100 replicates. Formerly, that took many many hours; now it takes only a few minutes. It would be easy to give the user the option of how many to do. Your thoughts?
2. The graph showing the approximate 95% bounds for the median Emp-Ws is not very useful I think (by necessity of scale, the three lines (median Ws, and upper and lower bounds) are going to be hard to distinguish. Kill it?
3. I’m not sure whether the ability to get a numerical summary of the bootstrap results for individual lengths adds much utility. The relevant information is visible on the graph. If you think it is useful (or at least, interesting), we can keep it. Otherwise, its removal would make the page cleaner.

1. Each bootstrap replicate consists of the following steps:
   a. Randomly select, with replacement, \( n \) integers from 1 to \( n \), where \( n \) is the number of populations being used to form the Ws equation. This represents the act of
randomly selecting \( n \) populations from the conceptual collection of all such populations.

b. For each selected population, randomly select, with replacement \( n_i \) pairs of lengths and weights, where \( n_i \) is the number of fish in the \( i^{th} \) sampled population. This represents the act of randomly sampling from that particular population.

c. Collate, organize, and compute the EmP-median \( W_s \) equation for that replicate.

2. Repeat Step 1 a large number of times.

3. For each length class, compute a relative margin of error (RMoE) as follows. Let

\[
SE_i = \frac{SD_i}{\sqrt{n_i}}
\]

be the standard error of the mean of the estimated \( W_s \) for that length class. Then \( 2 \times SE_i \) is the approximate margin of error (MoE) for a classical 95% confidence interval (the MoE is the plus/minus part of the C.I. calculation. Then MoE divided by the mean for that length class reports the precision of the mean on a relative scale.

4. In the illustration, notice that the RMoE is quite large for the smaller length classes, due to the paucity of data therein. This observation might impel the researcher to re-do the \( W_s \) computations using a larger lower limit.
On occasion, one might wish to present standard weight in units other than those used to generate the Ws in the first place. We assume here that researchers are likely to use for length and weight millimeters and grams, or inches and pounds. This sheet lets you choose which of those units you want to consider.
Application:

1. Enter length-weight data for a population whose relative weight you are interested in studying.
2. Click the summarize data button.
3. Scatter plots are drawn of Wr for individual fish, and for means from selected length categories. For the illustrated “test” population, the Individual values graph shows that most of the fish are above the median; on average, the means in each length category are near or above the third quartile.