

fishR Vignette - Length-Weight Relationships

Dr. Derek Ogle, Northland College

December 16, 2013

Modeling the relationship between length and weight of a species of fish has been considered a routine analysis for which the results do not warrant publication (Froese 2006) or has been scorned as being of little value (Hilborn and Walters 2001). However, the recent review of methods and the meta-analysis of a large number of length-weight relationships by Froese (2006) demonstrated that a synthetic analysis of length-weight relationships for a species can provide important insights into the ecology of that species.

The relationship between the length and weight of a fish is used by fisheries researchers and managers for two main purposes (Le Cren 1951). First, the relationship is used to predict the weight from the length of a fish. This is particularly useful for computing the biomass of a sample of fish from the length-frequency of that sample. Second, the parameter estimates of the relationship for a population of fish can be compared to average parameters for the region, parameter estimates from previous years, or parameter estimates among groups of fish to identify the relative condition or robustness of the population. By convention, this second purpose is usually generically referred to as describing the *condition* of the species.

In this vignette, length and weight data is discussed in Section 1. The common model used to describe length-weight relationships is introduced, developed, fit, and interpreted in Section 2. Methods for determining if the parameters of the length-weight relationship differ between populations are described in Section 3. Measures of relative condition are discussed in a separate vignette.

This vignette requires functions in the FSA package and a data frame from the fsadata packaged maintained by the author. These packages are loaded into R with

```
> library(FSA)
> library(FSAdata)
```

1 Length-Weight Data

The required data for examining the length-weight relationship for a sample of fish is measurements of the length (L) and weight (W) of individual fish at the time of capture (e.g., Table 1). Any other data about individual fish, such as month or year of capture, are of capture, etc. can also be recorded.

Table 1. Length and weight measurements for a portion of Ruffe from the St. Louis River Harbor, 1992.

month	day	year	individual	length	weight
4	23	1992	1	90	9.3
4	23	1992	2	128	32.5
4	23	1992	3	112	19.0
4	23	1992	4	68	4.4
4	23	1992	5	56	2.1
4	23	1992	6	58	2.8
⋮	⋮	⋮	⋮	⋮	⋮

Three types of length measurements are common in the fisheries literature (Figure 1). *Total length* (TL) is the length from the most anterior to the most posterior point with the tail of the fish compressed to exhibit the longest possible length. *Fork length* (FL) is the length from the most anterior point to the anterior notch in the fork of the tail. For fish without a forked tail, the fork and total lengths are the same. The *standard length* (SL) is the length from the most anterior point to the posterior end of the caudal peduncle. Total length is the most common measurement in fisheries studies, as this is the measurement used in management decisions such as setting minimum lengths.

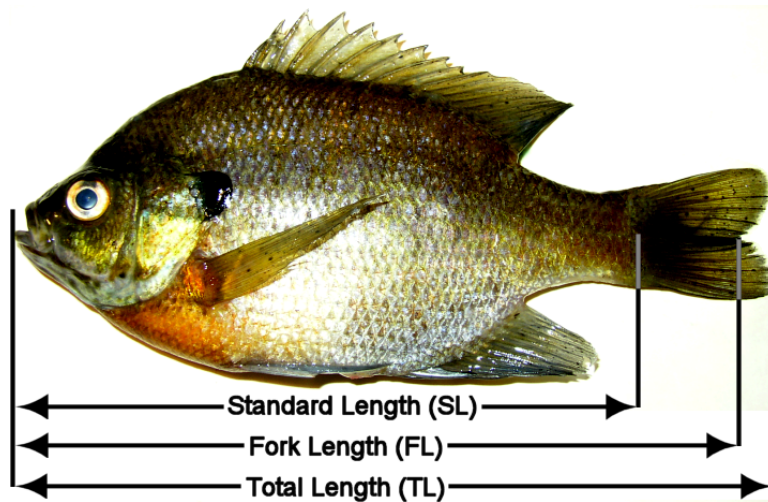


Figure 1. Demonstration of total, fork, and standard length measurements on a bluegill.

Length measurements are often taken with the aid of a measuring board where the measuring “stick” is embedded into the bottom board and one end of this board is fit with a vertical end piece. The fish to be measured is placed on the bottom board such that the anterior point of the fish is against the vertical end piece and the measurement can be read directly from the embedded measuring stick (Figure 2). Length measurements are subject to very little measurement error (Gutreuter and Krzoska 1994).



Figure 2. Field measurements of length (left) and weight with a spring scale (right).

Two common weight measurements are used. The usual body weight is the weight of the fish as it was captured, whereas the dressed weight is the weight of the fish with the gills and entrails removed. Dressed weight is usually only used when measurements are reported from a commercial fishery.

Weight measurements can be made in the field on fresh specimens or in the lab on fresh-frozen specimens. Weight measurements in the field can be taken with tared spring or electronic balances (Figure 2). However, field measurements can be extremely variable due to differences in fish surface wetness, boat movements, wind, and other adverse environmental conditions (Gutreuter and Krzoska 1994). Substantial variability in weight measurements can occur when fish weigh less than 10% of a scale’s capacity (Gutreuter and Krzoska 1994). Thus, multiple sizes of scales should be taken into the field (Blackwell *et al.* 2000). Wege and Anderson (1978) suggest that the accuracy of the scale should be $\pm 1\%$ of a fish’s body weight for use in relative weight calculations. Weight measurements on frozen fish were roughly 1-9% lighter than the measurements on the same fish when fresh, whereas length measurements were roughly 1-4% shorter on frozen than fresh fish for a variety of species (reviewed in Ogle (2009)).

2 Length-Weight Model

2.1 Model Characteristics

The relationship between the length and weight of a sample of fish tends to have two important characteristics. First, the relationship is not linear (Figure 3). This can be explained intuitively by thinking of length as a linear measure and weight as being related to volume. Thus, as the organism adds a linear amount of length, it is adding a disk of volume with a commensurate weight. Second, the variability in weight increases as the length of the fish increases (i.e., the scatter of the points increases from left-to-right in Figure 3). Thus, variability in weight among shorter fish is less than variability in weight among longer fish. Unfortunately, because of these two characteristics, length-weight data tends to violate the linearity and homoscedasticity (i.e., “constant variance”) assumptions of simple linear regression.

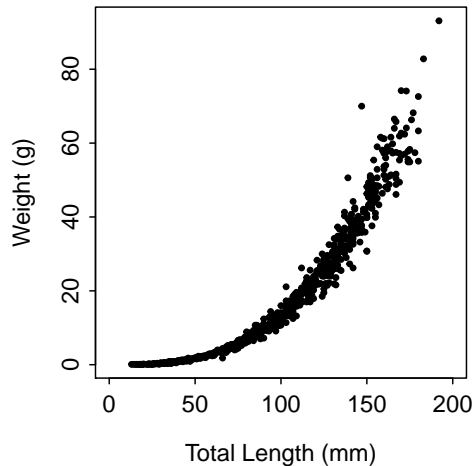


Figure 3. Length and weight of Ruffe from the St. Louis River Harbor, 1992.

These characteristics of length-weight data suggest that a two-parameter power function with a multiplicative error term should be used to model the length-weight relationship. Specifically, the model typically used is

$$W_i = aL_i^b e^{\epsilon_i} \quad (1)$$

where a and b are constants and ϵ_i is the multiplicative error term for the i th fish. The length-weight model (1) can be *transformed* to a linear model by taking the natural logarithms¹ of both sides and simplifying,

$$\log(W_i) = \log(a) + b\log(L_i) + \epsilon_i \quad (2)$$

Thus, with $y = \log(W)$, $x = \log(L)$, slope= b , and intercept= $\log(a)$, (2) is in the form of a linear model. In addition to linearizing the model, this transformation has the added benefit of making the errors additive and stabilizing the variances about the model (i.e., making the scatter around the line nearly constant for all length measurements; Figure 4). With this linearization and stabilization, the usual linear regression methods can be used to fit the relationship between $\log(W)$ and $\log(L)$.

It should be noted that, with the example in Figure 4, the variability on the log scale appears greater for “small” fish. This is because the scale used to measure these fish lacked the required precision to distinguish weights of small fish over a wide length range. It is apparent that fish with a log weight of less than -0.5 should be eliminated from this analysis because of scale imprecision for these fish.

¹Natural logarithms are used throughout this vignette and will be referred to simply as “logarithms” and will be abbreviated with “log.”

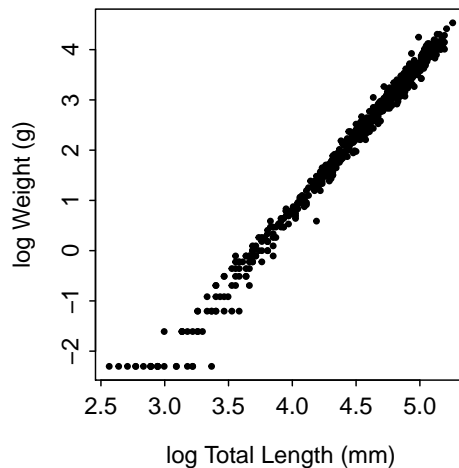


Figure 4. Natural log transformed total length and weight of Ruffe from the St. Louis River Harbor, 1992. Note that the “flaring” of the values in the lower-left corner of the plot is due to minimum weight limitations of the measuring scale.

2.2 Model Fitting in R & Interpretation

2.2.1 Fitting & Basic Summary Results

The lengths and weights of Ruffe² captured in the St. Louis River Harbor, Lake Superior, in 1992 will be examined as an example for fitting the length-weight relationship. These data are stored in the `RuffeSLRH92` data frame available in the `FSAdata` package that is loaded when `FSA` is loaded. These data are loaded into R with

```
> data(RuffeSLRH92)
> str(RuffeSLRH92)

'data.frame': 738 obs. of 11 variables:
 $ fish.id : num 1992 1992 1992 1992 1992 ...
 $ month : int 4 4 4 4 4 4 4 4 4 4 ...
 $ day : int 23 23 23 23 23 23 23 23 23 23 ...
 $ year : int 1992 1992 1992 1992 1992 1992 1992 1992 1992 1992 ...
 $ species : int 1 2 3 4 5 6 7 8 9 10 ...
 $ location: int 160170 160170 160170 160170 160170 160170 160170 160170 160170 160170 ...
 $ length : int 90 128 112 68 56 58 111 111 115 65 ...
 $ weight : num 9.3 32.5 19 4.4 2.1 2.8 16.1 17.9 22.7 3.4 ...
 $ sex : Factor w/ 3 levels "female","male",...: 2 1 2 2 3 2 2 2 2 1 ...
 $ maturity: Factor w/ 11 levels "developing","immature",...: 7 7 7 7 10 7 7 7 7 2 ...
 $ age : int NA NA NA NA NA NA NA NA NA NA 1 ...
```

Before beginning the length-weight model regression, this data frame needed to be “cleaned” by removing all records where measures of either length or weight were missing. This cleaning is accomplished with

```
> ruffe2 <- Subset(RuffeSLRH92, !is.na(weight) & !is.na(length))
```

Fitting a linear model to the log-log transformed length-weight data requires construction of the two new variables $\log(L)$ and $\log(W)$. These values are appended to the `ruffe2` data frame with

²Ruffe are a percid native to Europe that was found in the St. Louis River Harbor, Lake Superior, in the late 1980s. Various U.S. federal agencies monitored the population there until the early 2000s.

```
> ruffe2$logL <- log(ruffe2$length)
> ruffe2$logW <- log(ruffe2$weight)
```

In addition, as mentioned previously when reviewing Figure 4, fish with a measured log weight less than -0.5 should be removed from the analysis because the scale used to weigh fish of these sizes was too imprecise. Thus, these fish are removed with

```
> ruffe3 <- Subset(ruffe2,logW >= -0.5)
> str(ruffe3)

'data.frame': 680 obs. of 13 variables:
 $ fish.id : num 1992 1992 1992 1992 1992 ...
 $ month : int 4 4 4 4 4 4 4 4 4 4 ...
 $ day : int 23 23 23 23 23 23 23 23 23 23 ...
 $ year : int 1992 1992 1992 1992 1992 1992 1992 1992 1992 1992 ...
 $ species : int 1 2 3 4 5 6 7 8 9 10 ...
 $ location: int 160170 160170 160170 160170 160170 160170 160170 160170 160170 160170 ...
 $ length : int 90 128 112 68 56 58 111 111 115 65 ...
 $ weight : num 9.3 32.5 19 4.4 2.1 2.8 16.1 17.9 22.7 3.4 ...
 $ sex : Factor w/ 3 levels "female","male",...: 2 1 2 2 3 2 2 2 2 1 ...
 $ maturity: Factor w/ 11 levels "developing","immature",...: 7 7 7 7 10 7 7 7 7 2 ...
 $ age : int NA NA NA NA NA NA NA NA NA 1 ...
 $ logL : num 4.5 4.85 4.72 4.22 4.03 ...
 $ logW : num 2.23 3.481 2.944 1.482 0.742 ...
```

The transformed model, (2), is then fit with `lm()` by submitting a formula of the form $y \sim x$ as the first argument followed by a `data=` argument set equal to the data frame where the variables can be found. The result should be saved to an object to allow extraction of summary information. The model is fit with

```
> lm1 <- lm(logW~logL,data=ruffe3)
```

A quick view of the data and model fit (Figure 5) is produced by submitting the saved `lm()` object to `fitPlot()` as such,

```
> fitPlot(lm1,xlab="log Total Length (mm)",ylab="log Weight (g)",main="")
```

Basic summary information is extracted by submitting the saved `lm()` object to `summary()` as follows

```
> summary(lm1)

Call:
lm(formula = logW ~ logL, data = ruffe3)

Residuals:
    Min       1Q   Median       3Q      Max
-0.7771 -0.0642  0.0043  0.0785  0.5514

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.8368     0.0542   -200    <2e-16
logL         2.9123     0.0117    249    <2e-16

Residual standard error: 0.121 on 678 degrees of freedom
Multiple R-squared: 0.989, Adjusted R-squared: 0.989
F-statistic: 6.18e+04 on 1 and 678 DF, p-value: <2e-16
```

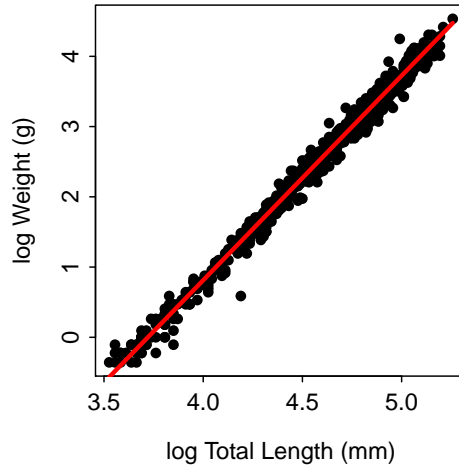


Figure 5. Natural log transformed total length and weight of Ruffe from the St. Louis River Harbor, 1992, with best-fit line superimposed.

From this summary and (Figure 5) it is seen that the model exhibits a tight fit to the transformed data ($R^2 = 0.99$) with the possible exception of a few individuals. The equation of the best-fit line is $\log(W) = -10.84 + 2.91 \cdot \log(L)$ on the transformed scale and $W = 0.000020L^{2.91}$ on the original scale³.

2.2.2 Inferences about Slope

If a fish grows without changing its shape or its density then the fish is said to exhibit *isometric* growth. In this case, the volume of the fish is proportional to any linear measure of its size. If weight is taken as a surrogate of volume (which requires assuming a constant density) and length as the linear measure, then the modeled relationship between length and weight, (1), will have $b = 3$ under isometric growth. Isometric growth in fish is rare (Bolger and Connolly 1989; McGurk 1985). If a fish changes shape or density as it grows, then $b \neq 3$ in (1), and the fish is said to exhibit *allometric* growth. If $b > 3$ then the fish tends to become “plumper” as the fish increases in length (Blackwell *et al.* 2000).

A test of whether the fish in a population exhibit isometric growth or not can be obtained by noting that b is the estimated slope from fitting the transformed length-weight model. The slope is generically labeled with β such that the test for allometry can be translated into the following statistical hypotheses:

- $H_0 : \beta = 3 \quad \Rightarrow H_0 : \text{“Isometric growth”}$
- $H_A : \beta \neq 3 \quad \Rightarrow H_A : \text{“Allometric growth”}$

Hypothesis tests regarding model parameters can be obtained with a t-test using

$$t = \frac{\hat{\beta} - \beta_0}{SE_{\hat{\beta}}}$$

where $\hat{\beta}$, $SE_{\hat{\beta}}$ and the df are from the linear regression results and β_0 is the specified value in the H_0 . Nearly all statistical packages, R included, print the t and corresponding p - value for $H_0 : \beta = 0$ by default, but not for any hypothesized value other than zero. However, `hoCoef()` can be used to efficiently compute the t and corresponding p - value for non-default hypotheses. This function requires the following arguments:

- a linear model object,

³Note that $a = e^{\text{intercept}} = e^{-10.84} = 0.000020$.

- a number indicating which term to use in the hypothesis test (1 for intercept, 2 for slope),
- a number indicating the specific value in the null hypothesis (i.e., β_0), and
- a character string indicating the direction of the alternative hypothesis ("less", "greater", or "two.sided" (default)).

A test, and confidence interval for b , of whether Ruffe from the St. Louis River Harbor in 1992 exhibited allometric growth or not is constructed with

```
> hoCoef(lm1,2,3)
  term Ho Value Estimate Std. Error    T    p value
    2      3    2.912    0.01172  -7.48 2.305e-13
> confint(lm1)
          2.5 %  97.5 %
(Intercept) -10.943 -10.730
logL         2.889   2.935
```

These results show that Ruffe exhibit allometric growth ($p < 0.00005$) with an exponent parameter (b) between 2.89 and 2.94, with 95% confidence.

2.2.3 Predictions on Original Scale

Predictions of the mean value of the response variable given a value of the explanatory variable can be made with `predict()`. In the length-weight regression, the value predicted is the mean log of weight. Most often, of course, the researcher is interested in predicting the mean weight on the original scale. An intuitive and common notion is that the log result can simply be back-transformed to the original scale by exponentiation. However, back-transforming the mean value on the log scale in this manner underestimates the mean value on the original scale. This observation stems from the fact that the back-transformed mean value from the log scale is equal to the geometric mean⁴ of the values on the original scale. The geometric mean is always less than the arithmetic mean⁵ and, thus, the back-transformed mean always underestimates the arithmetic mean from the original scale.

A wide variety of “corrections” for this back-transformation bias with logarithms have been suggested in the literature. The most common correction for log-transformed data is to multiply the exponentiated back-transformed value by

$$e^{\frac{s_{Y|X}^2}{2}}$$

This correction factor is derived from analysis of normal and log-normal distributional theory.

Predicted values on the logarithmic scale can be back-transformed to the original scale with `exp()`. The correction factor to be applied to this back-transformed value requires isolating the value of $s_{Y|X}$ by appending `$sigma` to `summary()` when a saved `lm` object is given, squaring that value, dividing by two, and then exponentiating that result with `exp()`. The correction factor is computed with⁶

```
> syx <- summary(lm1)$sigma
> ( cf <- exp((syx^2)/2) )           # this is the bias correction factor
[1] 1.007
```

The predicted mean log weight is found with `predict()` when the saved `lm` object is given as the first argument, a data frame with the logged length values is provided as the second argument, and the `interval=`

⁴The geometric mean is defined as the n th root of the product of the n values.

⁵The mean discussed in introductory statistics courses where the values are summed and divided by n is called the arithmetic mean.

⁶The “extra” parentheses on the second command forces R to print the result when the result is being assigned to an object.

argument is set to "c" (for "confidence"). It must be remembered that the data frame in the second argument must contain a variable that is exactly the same as the variable used on the right-hand-side of the formula when constructing the `lm` object (i.e., `logL` in this case). The predicted mean log weight for all 100-mm Ruffe is computed with

```
> ( pred.log <- predict(lm1,data.frame(logL=log(100)),interval="c") )
      fit   lwr   upr
1 2.575 2.566 2.584
```

This result can be used to form a biased predicted mean weight through exponentiation with

```
> ( bias.pred.orig <- exp(pred.log) ) # biased prediction on original scale
      fit   lwr   upr
1 13.13 13.01 13.25
```

Finally, a bias-corrected back-transformed mean weight for all 100-mm fish is found by multiplying the biased back-transformed value by the correction factor from above. This is illustrated with

```
> ( pred.orig <- cf*bias.pred.orig ) # corrected prediction on original scale
      fit   lwr   upr
1 13.23 13.11 13.35
```

Thus, the mean weight of all 100-mm Ruffe in the St. Louis River Harbor in 1992 is between 13.1 and 13.4 g.

3 Comparing Length-Weight Regressions

A common problem in fisheries research is to determine if the parameters from a simple linear regression fit are statistically different between two or more populations. For example, a researcher may want to determine if $\log(a)$ or b in (2) differ between sexes, between species, between lakes, or between treatments. These types of questions require fitting a more complicated regression model including indicator and interaction variables. This section shows, through an example, how to use those methods within the context of analyzing length-weight relationships.

Grafton Lake is a lake in the Kejimikujik National Park, Nova Scotia that was made larger by the construction of a dam in 1938. In the early 1990s, the Park Management team decided to remove the dam. Samples of Yellow Perch were collected in 1994 prior to dam removal and in 2000, several years after dam removal, to determine the effects of removing the dam on biological populations. One aspect of this research was to determine if the parameters of the length-weight relationship had changed significantly between these two time periods.

The Grafton Lake perch data set was recorded in the `YPerchGL` data frame distributed in the `FSadata` package. The `year` variable in this data frame is considered to be a numeric rather than group factor variable. Thus, a new variable, `fyear`, must be constructed with the `factor()` function for use in the linear model. The data are loaded, new transformed variables are created, and the new factor variable created with

```
> data(YPerchGL)
> YPerchGL$logFL <- log(YPerchGL$f1)
> YPerchGL$logW <- log(YPerchGL$w)
> YPerchGL$fyear <- factor(YPerchGL$year)
> str(YPerchGL)
```



```
'data.frame': 100 obs. of 6 variables:
 $ fl : int 59 60 63 63 66 69 72 74 78 78 ...
 $ w : num 2.5 2.5 3.3 3.6 3.9 4.1 4.4 4.8 5 5.6 ...
 $ year : int 1994 1994 1994 1994 1994 1994 1994 1994 1994 1994 ...
 $ logFL: num 4.08 4.09 4.14 4.14 4.19 ...
 $ logW : num 0.916 0.916 1.194 1.281 1.361 ...
 $ fyear: Factor w/ 2 levels "1994","2000": 1 1 1 1 1 1 1 1 1 1 ...
```

The full model with indicator and interaction terms is then fit and stored in an object with

```
> lm1 <- lm(logW~logFL*fyear,data=YPerchGL)
```

The analysis of variable table is constructed by submitting the saved `lm` object to `anova()` as such

```
> anova(lm1)
Analysis of Variance Table

Response: logW
          Df Sum Sq Mean Sq F value Pr(>F)
logFL      1  38.8    38.8  4670.1 < 2e-16
fyear      1   0.2     0.2    28.2 6.9e-07
logFL:fyear 1   0.0     0.0     0.4  0.53
Residuals 96   0.8     0.0
```

These results indicate that the interaction terms is not significant ($p = 0.5270$). Thus, there is not enough evidence to conclude that there is difference in slopes in the length-weight relationship between years. The p -value for the indicator variable suggests that there is a difference in intercepts between the two years ($p < 0.00005$). Because the two years have statistically equal slopes but different intercepts, there is a constant difference between the log-transformed weights of fish from the two years regardless of the log-transformed lengths of the fish.

Plots and confidence intervals should be constructed for the model without the interaction term, as it was not significant. The confidence intervals, constructed with

```
> lm2 <- lm(logW~logFL+fyear,data=YPerchGL)
> confint(lm2)

          2.5 %   97.5 %
(Intercept) -10.65020 -9.93167
logFL        2.67174  2.83174
fyear1       0.03154  0.06895
```

show that fish captured in 2000 are between -0.069 and -0.032 smaller, on the log scale, than fish captured in 1994 regardless of the length of the fish. A plot (Figure 6) depicting the final model can be constructed with `fitPlot()` as follows,

```
> fitPlot(lm2,xlab="log Fork Length (mm)",ylab="log Weight (g)",legend="topleft",main="")
```

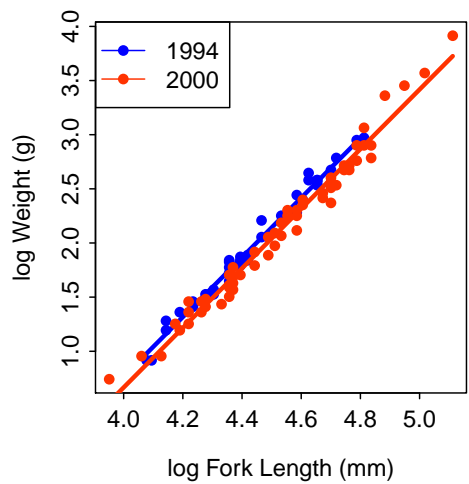


Figure 6. Fitted line from the “parallel lines” model fit to the Grafton Lake Yellow Perch data.

References

- Blackwell, B. G., M. L. Brown, and D. W. Willis. 2000. Relative weight (Wr) status and current use in fisheries assessment and management. *Reviews in Fisheries Science* 8:1–44. [2](#), [6](#)
- Bolger, T. and P. L. Connolly. 1989. The selection of suitable indices for the measurement and analysis of fish condition. *Journal of Fish Biology* 34:171–182. [6](#)
- Froese, R. 2006. Cube law, condition factor and weight-length relationships: history, meta-analysis and recommendations. *Journal of Applied Ichthyology* 22:241–253. [1](#)
- Gutreuter, S. and D. J. Krzoska. 1994. Quantifying precision of *in situ* length and weight measurements of fish. *North American Journal of Fisheries Management* 14:318–322. [2](#)
- Hilborn, R. and C. J. Walters. 2001. *Quantitative Fisheries Stock Assessment: Choice, Dynamics, & Uncertainty*. 2nd edition, Chapman and Hall, New York, 570 pp. [1](#)
- Le Cren, E. D. 1951. The length-weight relationship and seasonal cycle in gonad weight and condition in the perch (*Perca flavescens*). *Journal of Animal Ecology*. 20:201–219. [1](#)
- McGurk, M. D. 1985. Effects of net capture on the postpreservation morphometry, dry weight, and condition factor of Pacific herring larvae. *Transactions of the American Fisheries Society* 114:348–355. [6](#)
- Ogle. 2009. The effect of freezing on the length and weight measurements of ruffe (*Gymnocephalus cernuus*). *Fisheries Research* 99:244–247. [2](#)
- Wege, G. W. and R. O. Anderson. 1978. Relative weight (Wr): A new index of condition for largemouth bass. In G. D. Novinger and J. G. Dillard (editors), *New Approaches to the Management of Small Impoundments, Special Publication*, volume 5, pp. 79–91, American Fisheries Society. [2](#)

Reproducibility Information

Version Information

- **Compiled Date:** Mon Dec 16 2013
- **Compiled Time:** 9:52:51 PM
- **Code Execution Time:** 2.26 s

R Information

- **R Version:** R version 3.0.2 (2013-09-25)
- **System:** Windows, i386-w64-mingw32/i386 (32-bit)
- **Base Packages:** base, datasets, graphics, grDevices, methods, stats, utils
- **Other Packages:** FSA_0.4.3, FSAdata_0.1.4, gdata_2.13.2, knitr_1.5.15
- **Loaded-Only Packages:** bitops_1.0-6, car_2.0-19, caTools_1.16, cluster_1.14.4, evaluate_0.5.1, formatR_0.10, Formula_1.1-1, gplots_2.12.1, grid_3.0.2, gtools_3.1.1, highr_0.3, Hmisc_3.13-0, KernSmooth_2.23-10, lattice_0.20-24, MASS_7.3-29, multcomp_1.3-1, mvtnorm_0.9-9996, nlme_3.1-113, nnet_7.3-7, plotrix_3.5-2, quantreg_5.05, sandwich_2.3-0, sciplot_1.1-0, SparseM_1.03, splines_3.0.2, stringr_0.6.2, survival_2.37-4, tools_3.0.2, zoo_1.7-10
- **Required Packages:** FSA, FSAdata and their dependencies (car, gdata, gplots, Hmisc, knitr, multcomp, nlme, plotrix, quantreg, sciplot)