

Alan Glen Evangelista

Math 562 - Take-Home Lab

For this lab, length-at-age data is examined for a random sample of Yellowstone cutthroat trout (YCT) and rainbow hybrid trout (RHT) from the Snake River. The von Bertalanffy growth function and the Gompertz growth function are commonly applied models for fish growth. Although both are used for fish growth, they both have distinct traits. The von Bertalanffy function is concave down and assumes that growth rate decreases linearly with length. While the Gompertz function has an inflection point and assumes that growth rate decreases exponentially with age. Both use similarly named parameters, i.e.  $t_0$  as a shift parameter,  $L_{\infty}$  (abbreviation: Linf) as the max length asymptote, and  $k$  as the rate that length approaches  $L_{\infty}$ .

In my analysis, the Gompertz growth function with differences in the parameters  $t_0$  and  $L_{\infty}$  (model name: LinfG) fits the data best. Compared to other variations in parameterization of von Bertalanffy and Gompertz functions, the LinfG model has the lowest AIC and a good combination of maximum likelihood value, R-squared value, and number of parameters. A comparison of AIC, log-likelihood, weights, and R-squared can be seen from the tables on the following page. Throughout the trials the likelihoods of the Gompertz models were stronger, suggesting that an inflection point in the growth curve is likely. From each variation of the Gompertz model, all my  $t_0$  values, which represent the location of the inflection point, were between 0.26-1.1. Since they were all above 0, that means the inflection point was included in each Gompertz model I optimized.

As my best model is the LinfG model (Gompertz with  $L_{\infty}$  and  $t_0$  differing between species) and other variations that contain differing  $L_{\infty}$  have higher likelihoods, the theoretical maximum length should differ between species. Between the Gompertz models I used with differing  $L_{\infty}$ 's, there is always a positive adjustment for the RHT (+87 to +200). Even in the von Bertalanffy models I optimized, there is always a positive adjustment for the RHT (+30 to +220). This suggests that the theoretical maximum length for the RHT is higher.

My Gompertz models suggest that the rate of growth is faster for the RHT however my von Bertalanffy models are inconsistent with which species has a higher  $k$  value. Also, my models with varying  $k$ 's don't do significantly better than ones without differing  $k$ 's. This suggests that there is no strong evidence that RHT grow faster than YCT.

```

> aictable(myAIC,339)
  Params  logL      AICc    deltaAICc  weight  cumwt
LinfGmle    6 -1787.236  3586.724         0    0.3243  0.3243
fullGmle    8 -1785.22    3586.877    0.1527    0.3005  0.6248
fullVBmle   8 -1785.804  3588.044    1.3193    0.1677  0.7925
kLGmle      7 -1786.966  3588.271    1.5473    0.1496  0.9421
kGmle       6 -1789.179  3590.612    3.8874    0.0464  0.9885
kLVBmle     7 -1789.934  3594.206    7.4821    0.0077  0.9962
sigGmle     6 -1792.273  3596.799   10.0751    0.0021  0.9983
t0Gmle      5 -1793.754  3597.688   10.9637    0.0013  0.9996
kVBmle      6 -1794.156  3600.564   13.8401    0.0003    1
LinfVBmle   6 -1796.46   3605.173   18.4486    0        1
t0VBmle     5 -1801.011  3612.203   25.4787    0        1
sigVBmle    6 -1808.731  3629.715   42.9911    0        1
nullVBmle   4 -1835.424  3678.967   92.2427    0        1
nullGmle    4 -1837.245  3682.61    95.886     0        1

```

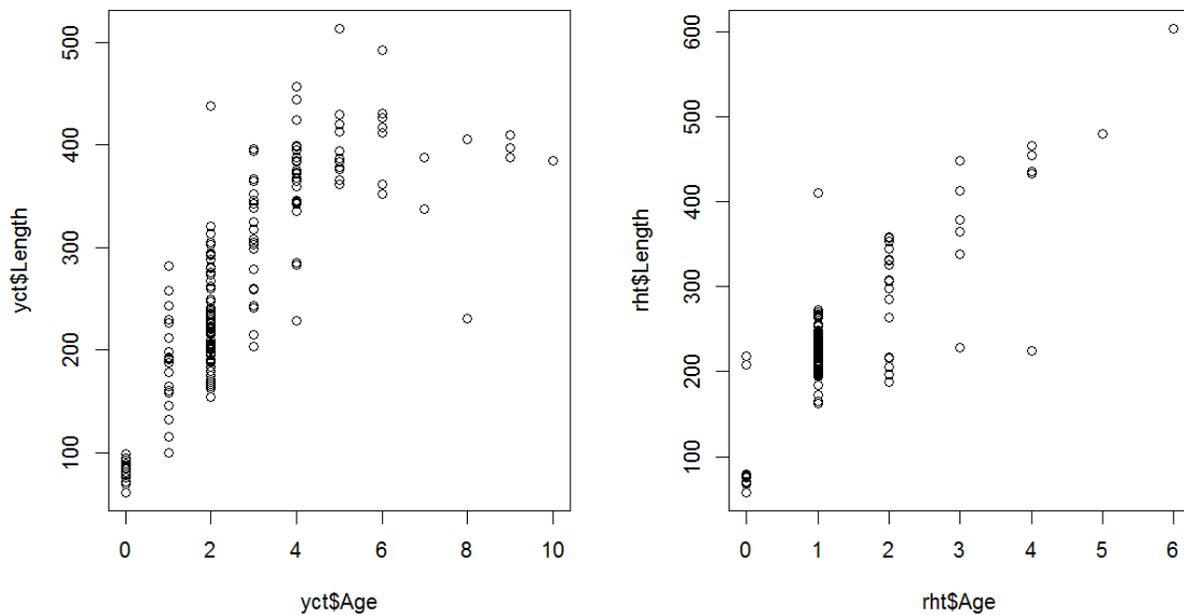
```

> r2table
  R2      Name      Function      Species Specific Parameters
fullGmle  0.2642992  fullGmle     Gompertz      t0, k, Linf, sig
kLGmle    0.2566804  kLGmle       Gompertz      t0, k, and Linf
LinfGmle  0.2554996  LinfGmle     Gompertz      t0 and Linf
fullVBmle 0.2537863  fullVBmle    vonBertalanffy t0, k, Linf, sig
kGmle     0.246913    kGmle        Gompertz      t0 and k
kLVBmle   0.2353792  kLVBmle      vonBertalanffy t0, k, and Linf
sigGmle   0.2330409  sigGmle      Gompertz      t0 and sig
t0Gmle    0.2263115  t0Gmle       Gompertz      just t0
kVBmle    0.2160959  kVBmle       vonBertalanffy t0 and k
LinfVBmle 0.2053665  LinfVBmle    vonBertalanffy t0 and Linf
t0VBmle   0.1837396  t0VBmle      vonBertalanffy just t0
sigVBmle  0.1457038  sigVBmle     vonBertalanffy t0 and sig
nullVBmle 0          nullVBmle    vonBertalanffy none
nullGmle  0          nullGmle     Gompertz      none

```

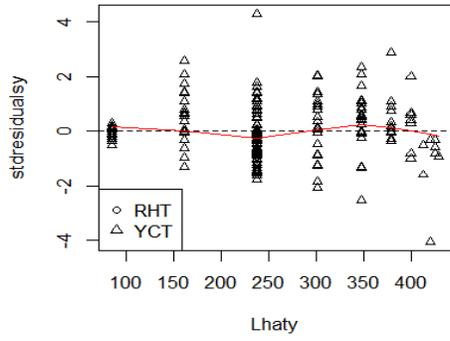
## Statistical Issues:

I used just the normal probability distribution to account for variability in observed lengths. Only the normal distribution was used because the data did not seem to be consistently skewed one way or the other. Without obvious skewness, the symmetry of the normal distribution seems appropriate. However, the model may have benefited by using a right skewed distribution. For the YCT alone, the data looks like it is sometimes slightly right skewed or left skewed. For RHT though, the data has a number of outliers, mostly causing right skewness, as seen below.

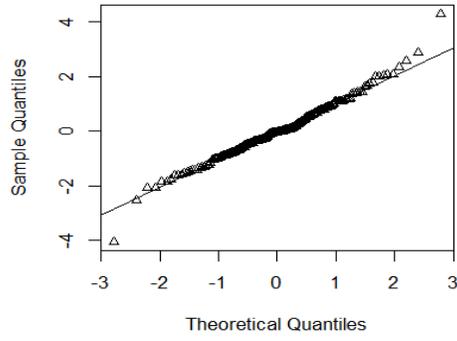


Diagnostic tests on the next page show that the model did fit YCT data well, but the RHT may be too skewed to be compatible with the normal distribution and may benefit from a transformation. The first four plots looking at the YCT data shows a good residual vs. fit and fits vs. observed. However, the next four plots which look at the RHT data shows a few problems with fitting the model to RHT. The first diagnostic shows that the residuals are mostly positive, meaning the model fitted values underestimate the data. The tails of the normal Q-Q plot show that the tails are very off. And the fits vs. observed graphs shows how the fitted values seem to be below the observed. For YCT, the model fit well over most ages, except later years where it overestimates. For RHT, the model fits best at young ages and worsens with increasing age.

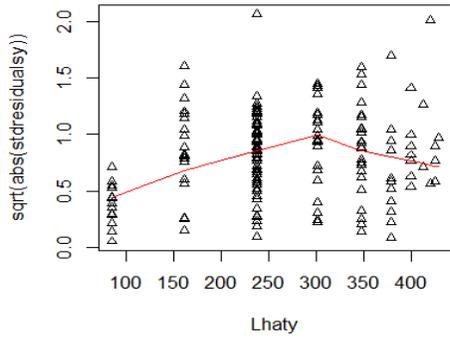
**Residuals vs. Fits**



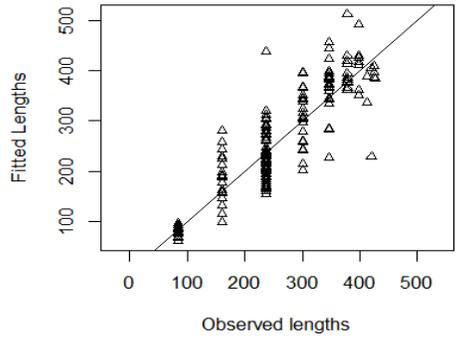
**Normal Q-Q Plot**



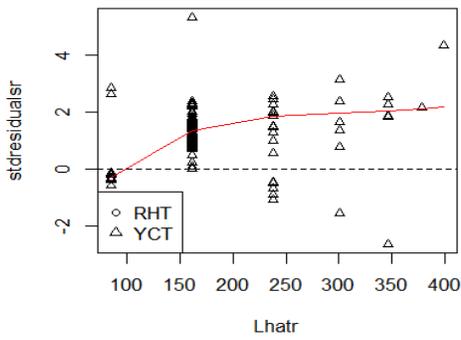
**Scale-Location**



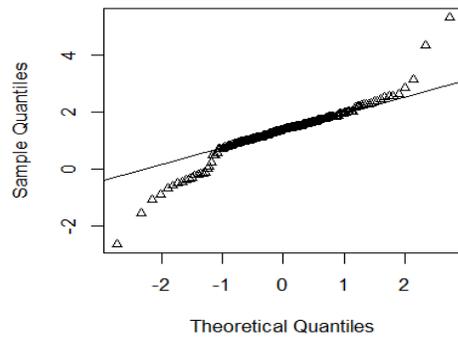
**Fits vs. Observed**



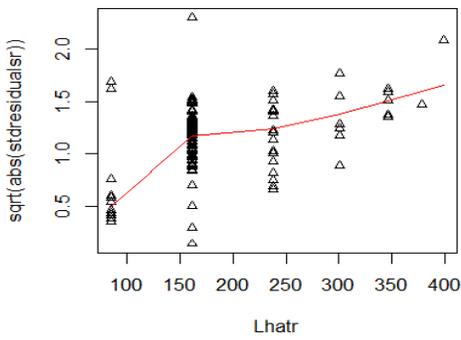
**Residuals vs. Fits**



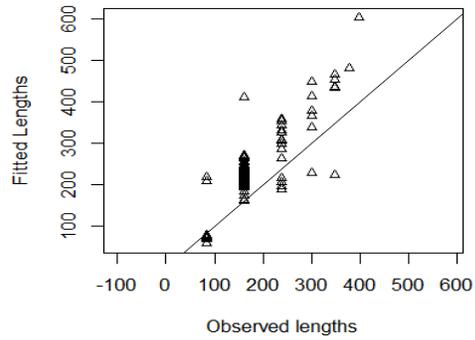
**Normal Q-Q Plot**



**Scale-Location**



**Fits vs. Observed**



Final Model:

The best model LinfG has the following equation:

$$L(t) = L_{\infty} \exp(-\exp(-k(t-t_0)))$$

The YCT has the following parameter value estimates:

$$\begin{aligned} t_0 &= 0.9749940 \pm 2 * 0.09506402 \\ k &= 0.4999927 \pm 2 * 0.04201564 \\ Linf &= 432.4294630 \pm 2 * 13.98548557 \end{aligned}$$

The RHT have the following parameter value estimates: 0.2502821

$$\begin{aligned} t_0 &= 0.6782655 \pm 2 * 0.2502821 \\ k &= 0.4999927 \pm 2 * 0.04201564 \\ Linf &= 519.0069 \pm 2 * 38.35147 \end{aligned}$$

The dispersion parameter is estimated as:

$$\sigma = 46.7855501 \pm 2 * 1.77853283$$

