# An After-Dinner Taste of My Research for Nonstatisticians

Jeff Hart

Department of Statistics, Texas A&M University

# Outline
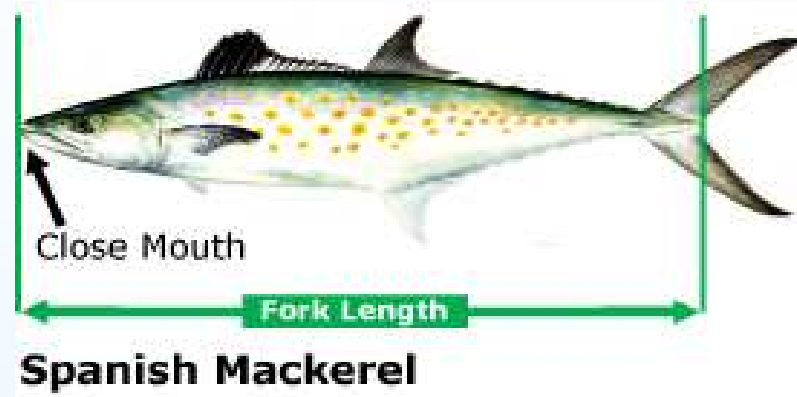
- Introduction to fish data and scatterplots

- Scatterplot smoothing

- The bootstrap

- Hypothesis testing

# Some fish data

- I once helped an animal scientist in analyzing some data collected from fish.

- Study was concerned with a comparison of male vs. female growth rates.

- Two quantities were studied:

  - Fork length

  - Otolith weight

# Fork length



Fork length is measured from the tip of the snout to the fork of the tail.

# Otoliths

- Both fish and humans have otoliths.

- In humans, the **otolithic organs** are in the inner ear. They help us maintain our balance.

- Otoliths in fish are small structures found in their heads and composed of calcium carbonate. They help the fish's balance and hearing.

- Growth rings (as for a tree) are found in the otolith and provide scientists with detailed info on the fish's age and growth.

# Otoliths



2000 μm

# Fish scatterplots



**Males**

**Females**
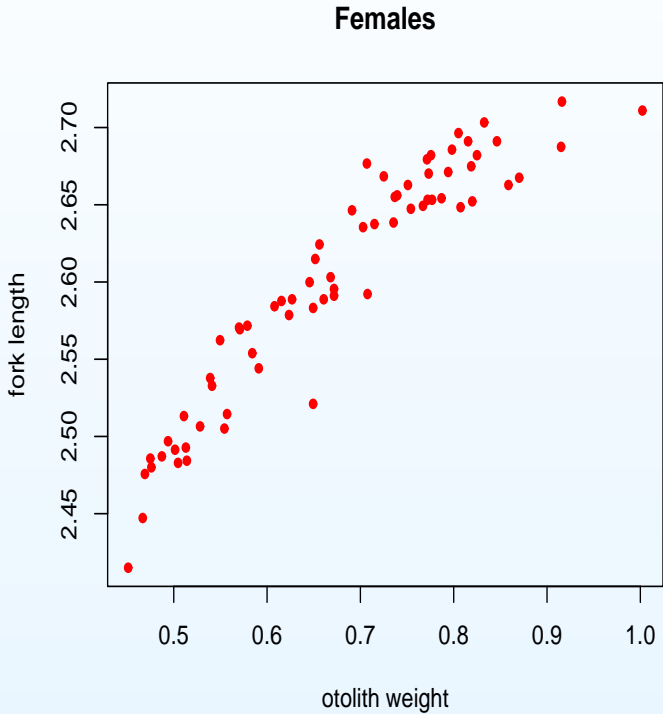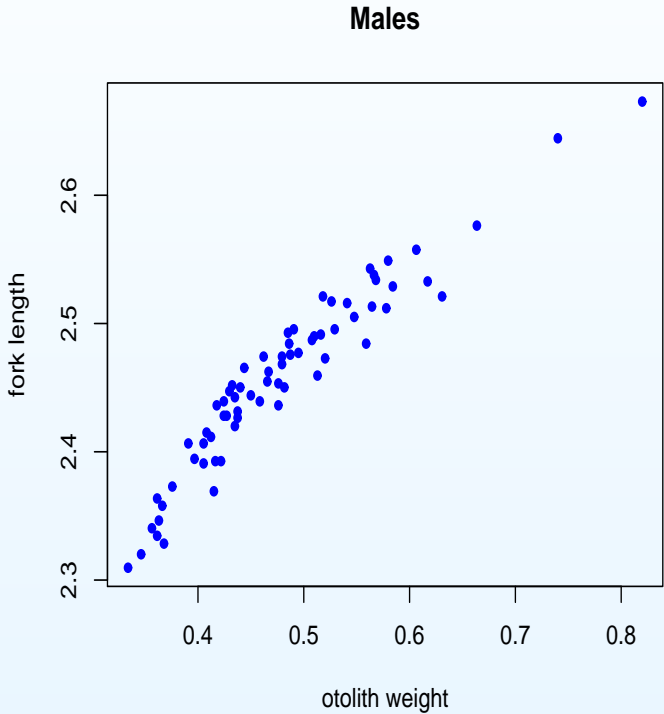
# What is a scatterplot?

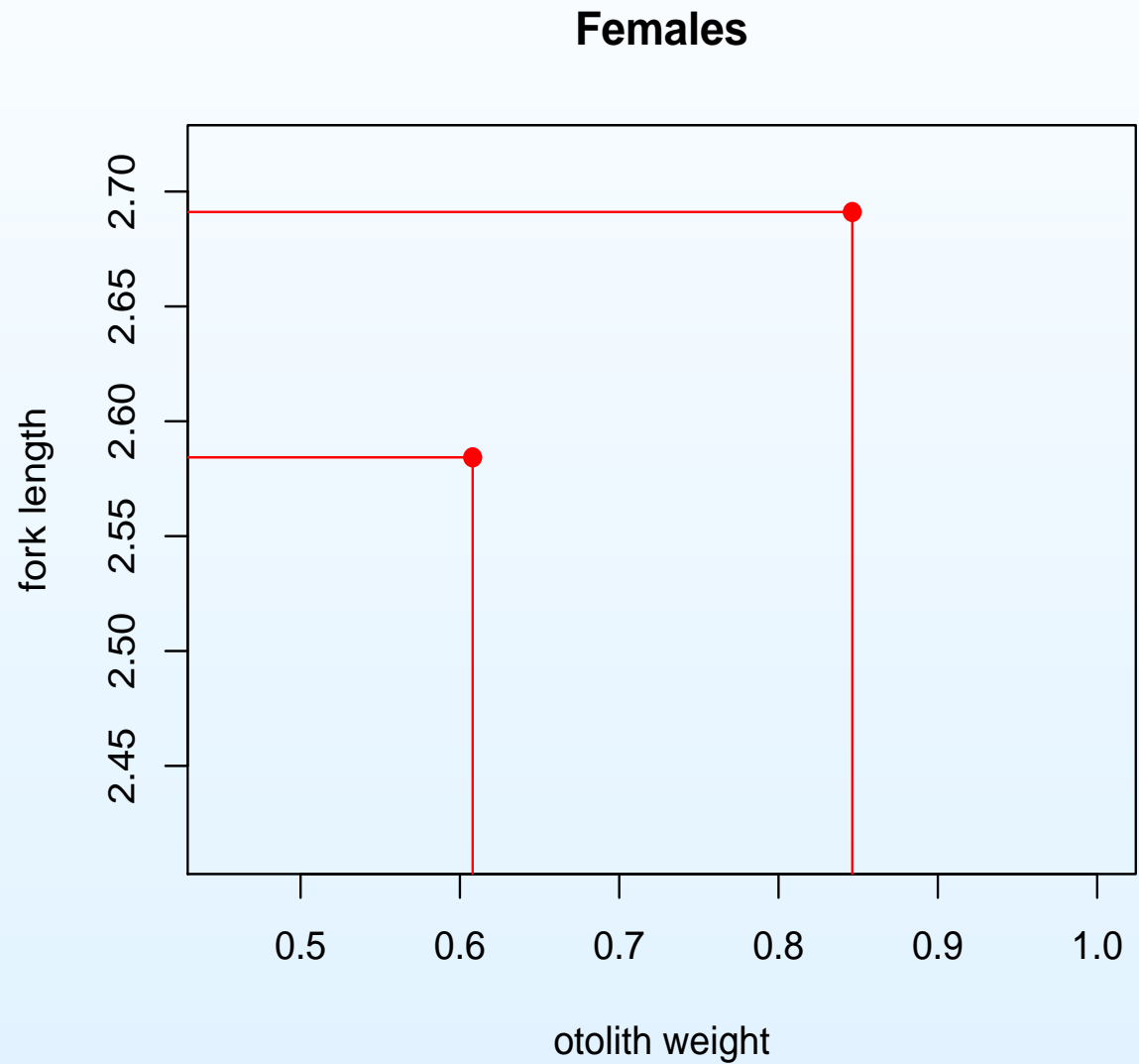A **scatterplot** is a graph like either of the two on the previous page.

Characteristics:

- Each dot corresponds to a different fish.

- The horizontal position of a dot indicates a fish's otolith weight.

- The vertical position of a dot indicates a fish's fork length.

# What is a scatterplot?



**Females**

fork length vs. otolith weight

# What's learned from the scatterplots?

- Older fish (i.e., ones with bigger otoliths) are longer. Wow, what a shocker!

- Less obvious questions can also be answered.

  ○ At what <span style="color:red">rate</span> do the fish grow? (Do they grow quickly as they age?)

  ○ Does the rate of growth depend on the fish's age?
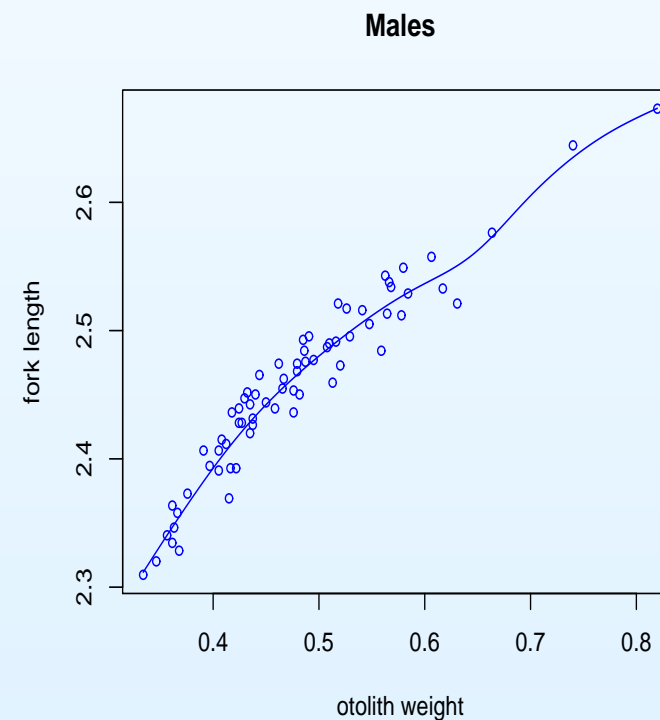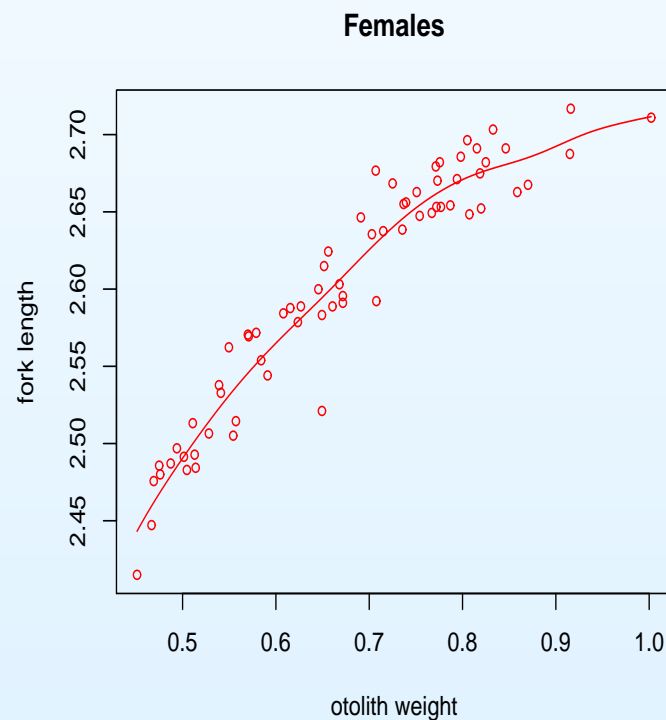
  ○ At what age do the fish reach maturity?

# Mixing males and females

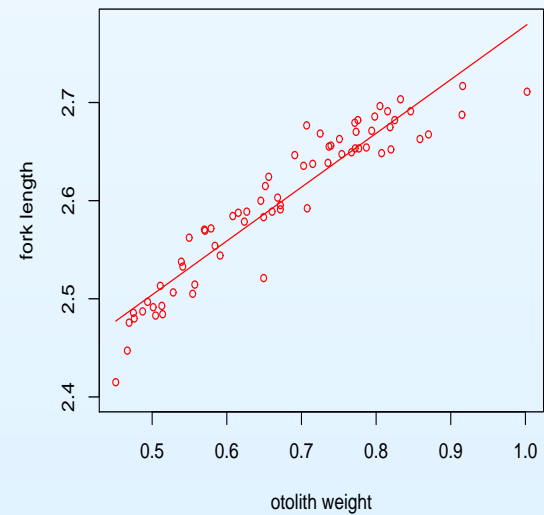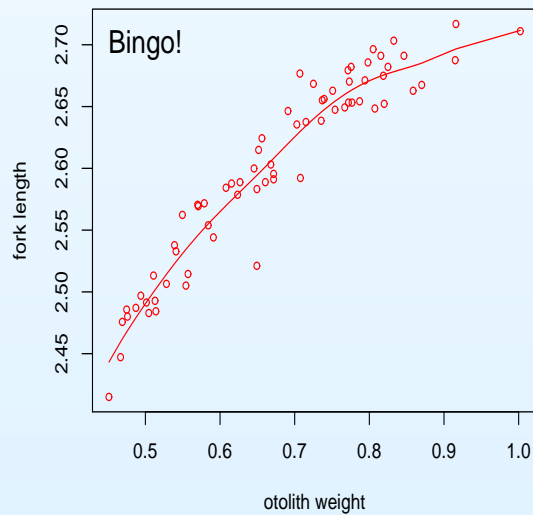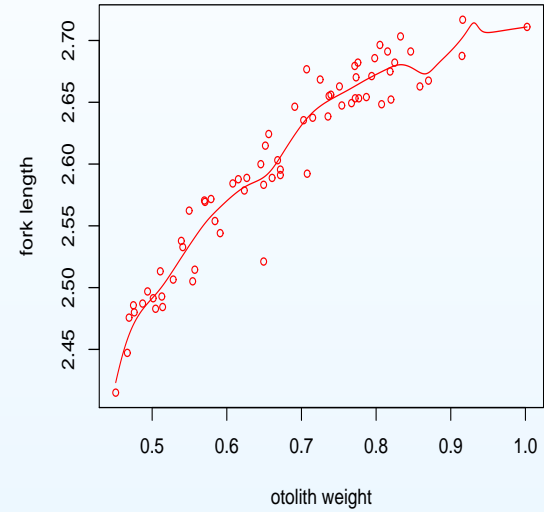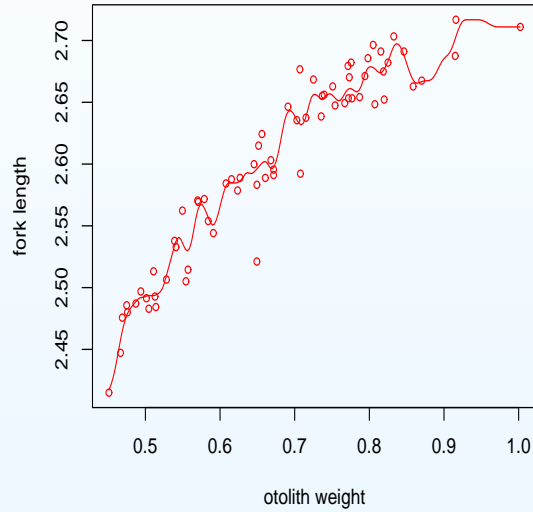**The males seem to be somewhat younger than the females.**
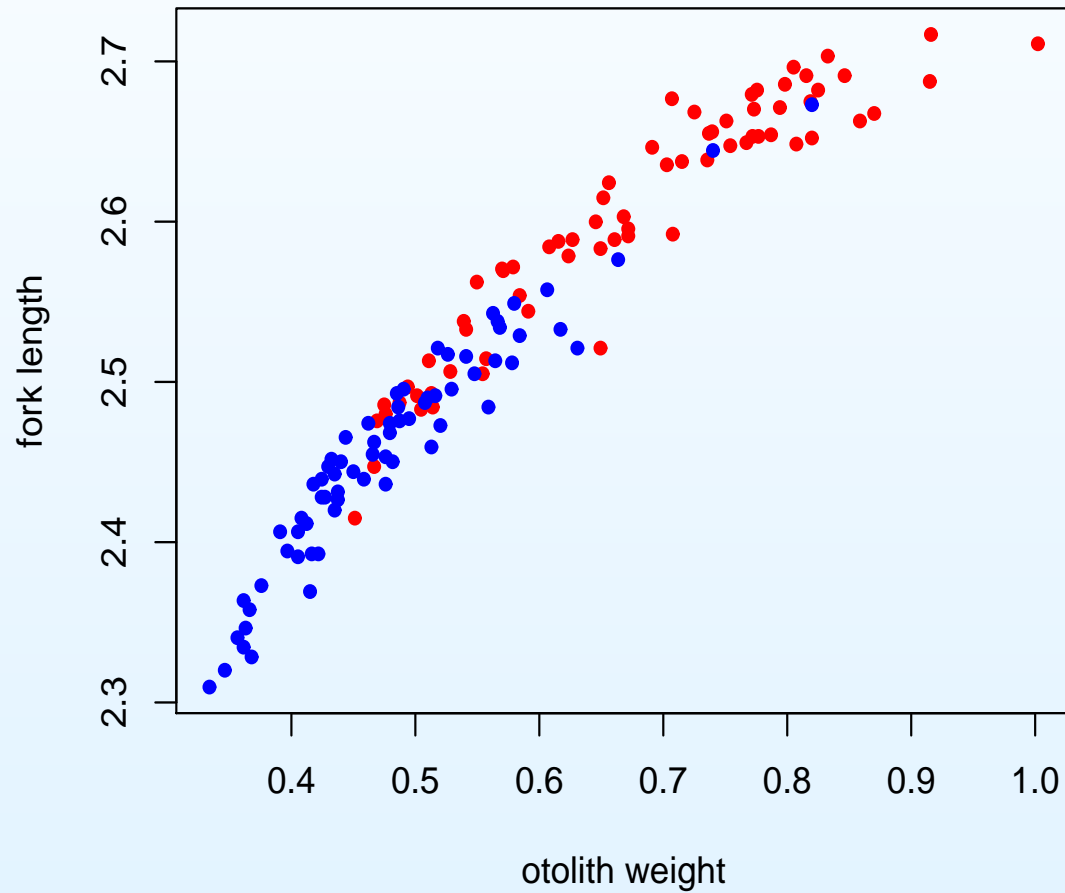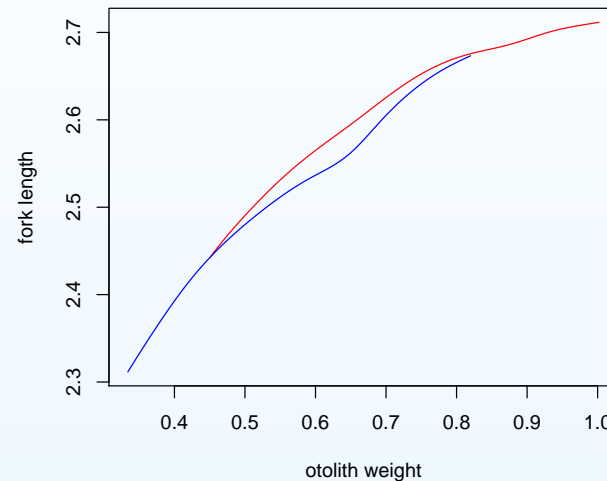
# Scatterplot smoothing

- Sometimes you can't see the forest for the trees. If there's too much scatter in the scatterplot, it may be difficult to see subtle effects.

- Subtle effects can be seen better by smoothing the scatterplot.



**Females**

**Males**

# How much smoothing is the "right" amount?



Bingo!

# Are the male and female growth curves the same?



- Obviously, these two smooths are at least a little different!

- But are they enough different to convince us that the growth curve of *all* females (of this species) is different than the growth curve of *all* males?

- Important point: *Two different sets of fish of the same sex and species can produce quite different scatterplot smoothers.*

# Variation in male scatterplot smooths
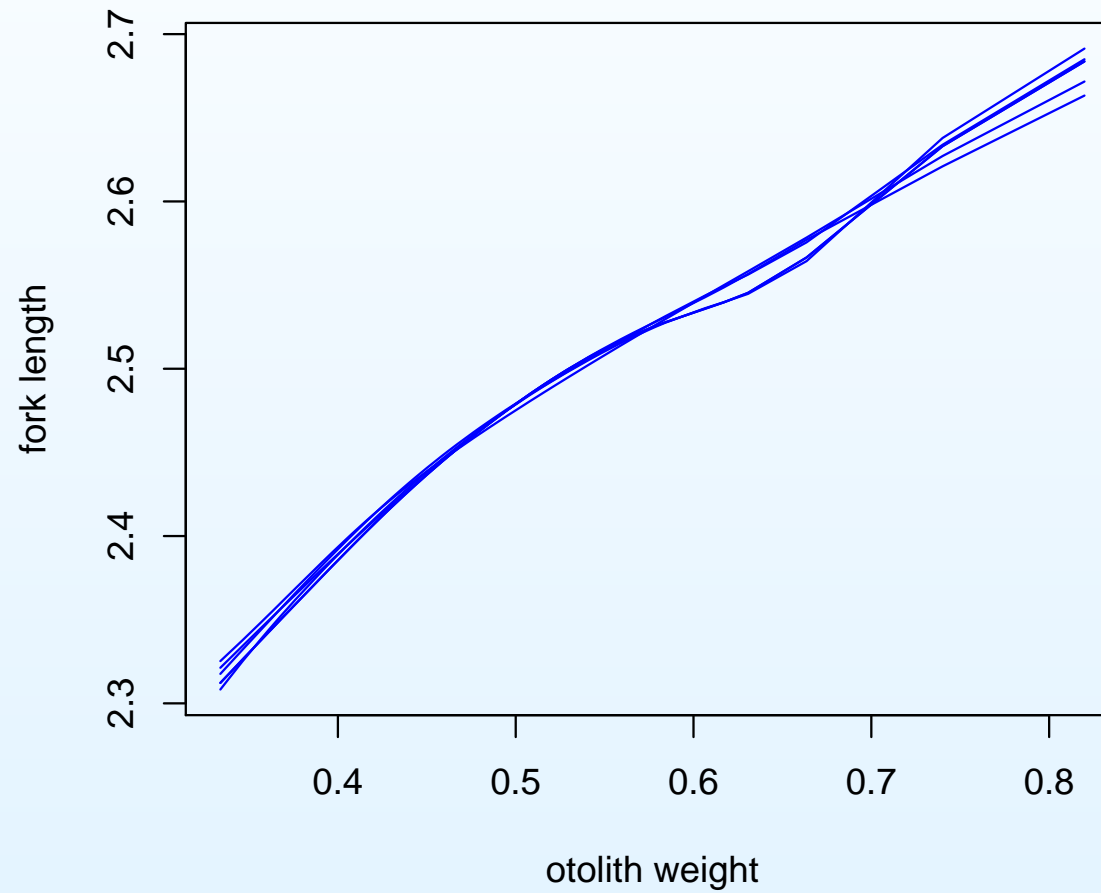
# Difference between two scatterplot smooths



The difference between the female and male smooths.

# The bootstrap

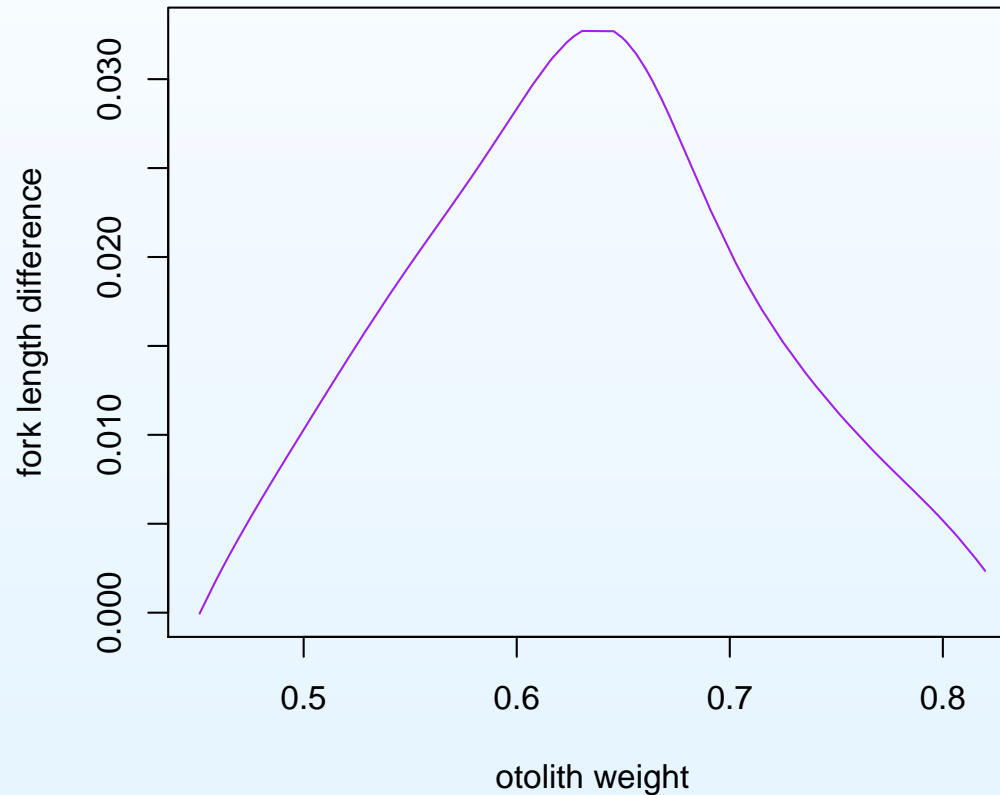- To know whether or not the difference between our two smooths is really unusual, we need to know <span style="color:red">how much the two smooths could change if we got different sets of male and female fish</span>.

- Unfortunately, we have only one set of male and one set of female fish.

- Statisticians have devised an ingenious method that will allow us to <span style="color:red">simulate</span> variation in scatterplot smooths <span style="color:red">using just our one set of males and one set of females</span>.

- This ingenious method is called the <span style="color:red">bootstrap.</span>

# How the bootstrap works

1. "Tweak" the female data so that it has exactly the same scatterplot smooth as the male data.

2. Scramble the male data and determine the scatterplot smooth for the scrambled data.

3. Scramble the tweaked female data and determine the scatterplot smooth for the scrambled data.

4. Determine the difference between the two new smooths.

5. Repeat steps 2-4 many, many times, say 10,000. (Each time the data are scrambled, they're scrambled differently.)

6. Look at all 10,000 differences to see how they vary one to the next.

# Why is it called "the bootstrap?"

The name was coined by Brad Efron and comes from the saying

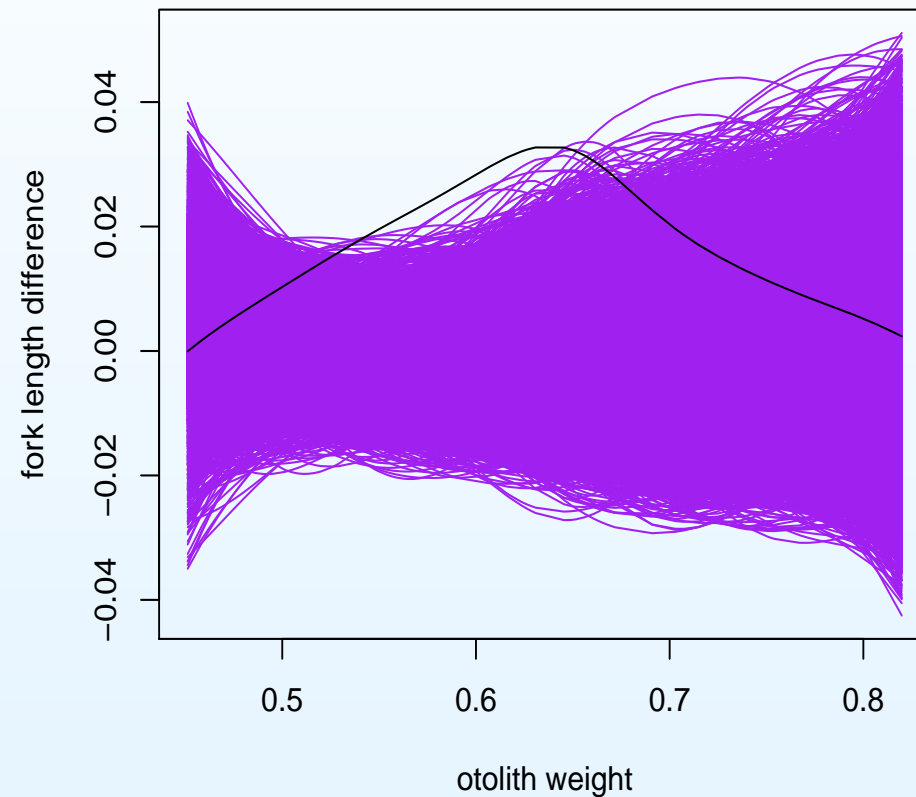<span style="color:red">"lift yourself up by your bootstraps."</span>

Our one set of male and one set of female fish are <span style="color:red">"lifting themselves up by their bootstraps."</span>

Without any help from other sets of fish, the method is still able to determine variation that would occur if we <span style="color:red">did</span> have other sets of fish.

# 10,000 differences obtained by bootstrapping



The black curve is the difference between the female and male smooths.

# Hypothesis testing

Hypothesis:

<span style="color:red">Male and female growth curves are no different.</span>

Since the difference between the female and male smooths strays outside the set of 10,000 simulated differences, it's reasonable to <span style="color:red">reject the hypothesis</span>.

Conclusion:

*Males and females of this species have at least somewhat different growth curves.*