# **MA207 - Introduction to Statistics: Lab #7: Hypothesis Tests and Confidence Intervals on Large Data Sets Day 2 - Blackfoot River Fish Data**

In this multi-day lab we will work with several large data sets. For each day you will receive a large data set and will be asked to answer several specific questions related to the data set. For each problem you will need to be able to:

* Filter and sort the data (pivot tables are helpful),
* Find appropriate descriptive statistics for parts of the data set,
* Determine the appropriate statistical tool for the question being asked and complete all of the calculations necessary to answer the questions, and
* Interpret and explain the meaning of your computations in context.

## The Blackfoot River Fish Data:

This data set of 18,532 fish caught over several decades comes from Montana Fish Wildlife and Parks (FWP). For this project, researchers used electrofishing equipment to attract fish to the boat, then dipped them out of the water with nets, measured length in mm and weight in grams. They were often working in cold conditions in late autumn or early spring, so some measurement error is expected. Also, because this is real data, it might be a bit messy.

These data are not from a random sample. The biologists’ goal was to catch all fish within a section of the Blackfoot River every few years to assess the health of the population. Changes over years are important to the biologists. Despite the fact that this is not a random sample, FWP biologists know that they did not catch every fish and they also know that the fish are not stationary within the river. Hence, we can use this data as a sample of the fish *near* that section of river.

Download the Blackfoot River Fish data set from the Moodle page.

**Your Tasks:**

Use appropriate statistical tools to answer the following questions. Provide sufficient detail in your write-up so that anyone with a similar statistics background will understand what you have done. You do not need to turn in your Excel document.

1. Was there a statistical difference between the weight of the fish at Johnsrud in 1989 and in 2006?
   1. Which statistical tool are you using? Why are you using this tool?
   2. Clearly report your numerical results of the statistical test. If you run a hypothesis test and happen to reject your null hypothesis then you need to follow up with a confidence interval estimating the true parameter.
   3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.
2. According to [Headhunters Fly Shop](http://www.headhuntersflyshop.com/blackfoot-river/), most fishermen on most days catch trout that are, on average, 12 inches in length (304.8mm) on the Blackfoot River. Headhunters is in Ovando, MT which is not too far from the Scotty Brown Bridge. Do we have statistical evidence that the 2006 trout near the Scotty Brown Bridge are less than the length advertised by Headhunters?
   1. Which statistical tool are you using? Why are you using this tool?
   2. Clearly report your numerical results of the statistical test. If you run a hypothesis test and happen to reject your null hypothesis then you need to follow up with a confidence interval estimating the true parameter.
   3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.
3. A local fisherman claims that the proportion of *large* fish (over 300 grams (⅔ of a pound)) that he caught at Scotty Brown Bridge in 2002 seemed to be larger than the proportion of *large* fish caught at Johnsrud in the same year. Is there statistical evidence to support or refute his claim? You may find it helpful to add a column of data for “size” and fill it with the words “large” and “small”. To do this, an if statement can be quite efficient. For example, =IF(D2>300,"large","small") will determine whether a weight is over 300 and assign it the word “large” if so.
   1. Which statistical tool are you using? Why are you using this tool?
   2. Clearly report your numerical results of the statistical test. If you run a hypothesis test and happen to reject your null hypothesis then you need to follow up with a confidence interval estimating the true parameter.
   3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.
4. Write a question of your own based on this data, focusing on one particular species of fish (you choose which species and what question to ask). Clearly state your question and then answer it with a hypothesis test or confidence interval.
   1. Which statistical tool are you using? Why are you using this tool?
   2. Clearly report your numerical results of the statistical test.
   3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.