1. [1pt] How many total fish are in this dataset?

18352

1. [1pt] How many different species are represented in this dataset?

4 (Brown, Bull, WTC, RBT)

1. [2pts] What are the names of the two river sections where the fish were collected?

Scotty Brown and Johnsrud

1. [2pts] How many fish had either 0 or “NA” recorded for weight?

1797

1. [2pts] Remove these 0 and “NA” weight fish from the dataset with the subset()function.

Done.

1. [3pts] How does the average weight of fish vary by time and species?

It seems that though the mean weight of fish seem to have fluctuated over the years from 1989 to 2006 all species, on average, have a higher mean weight in 2006 than they did in 1989. Also, all species experienced and increase and decrease in mean weight over the years. Lastly, it seems that the average weight of fish was highest in 1996, though not all species weight peeked at this time.

1. [3pts] Make and interpret a QQplot of the weight of Rainbow Trout. That is, justify your comments about normality of the data by referring to your QQplot.

The weight of Rainbow Trout is not approximately normally distributed. The data plotted against the theoretical normal distribution failed to form a straight line. At X value 1, the data seemed to experience an abnormally large upward trend which indicates a departure from normality.

1. b. [1pt] Which section of river typically supports a higher percentage of the Rainbow Trout population?

Johnsrud.

c. [2pts] Looking at your mosaic plot, explain why the bar for 1996 is so much skinnier than the bars for 1991 or 1993.

It seems the width of each bar was determined by the size of data it includes relative to the total data found in the table. Given that in 1996 the sample size of Rainbow Trout Fish was 854 as opposed to 2058 in 1991 and 1784 in 1993, it makes sense that its’ bar would be much skinnier; its’ sample size is less than half of the sample size in 1991 and 1993.

1. .For this particular habitat, as long as ScottyBrown can support at least 17% of the Rainbow Trout population, the ecosystem is considered healthy and biologists write off any change from year to year to a combination of sample variation and a natural 3 rise and fall of the fish population. Back in 2006, however, some biologists were sounding alarm bells as the percent of Rainbow Trout in ScottyBrown fell to 16.7%. Others argued that, given the sample size, this proportion is not statistically different than the target of 17% and that they should wait to see the results of next year’s sample before jumping to any conclusions. [5pts]

Conduct a hypothesis test to either support or refute the claim made by this second group of biologists. That is, conduct a hypothesis test for H0: P(RBT in ScottyBrown2006) = 0.17 against the alternative that is in fact less. Report your test statistic and p-value and give a conclusion in the context of the problem.

1. Establish the null and alternative Hypothesis

H0: P(RBT in ScottyBrown2006) >= 0.17

HA: P(RBT in ScottyBrown2006) < 0.17

1. Find p̂

p̂ = p/n=556/1771 = 0.31

1. Is n(1- p̂) >= 15? Yes! It has an approximate normal distribution.
2. Find E(p̂) which is p̂. 0.31
3. Find the standard deviation of p̂.

σ= sqrt((p̂(1- p̂))/n) = sqrt((0.31\*(1-0.31))/1771) = 0.011

* Thus, N ~ (0.31,0.011)

1. Find Z which is Z= mean of proportion – mean of population / standard deviation = 0.31-0.17/0.011= 1.273
2. Find p which is P(Z<12.727) = 1
3. Given that the p-value is 1 the null hypothesis holds. Thus, this proportion is not statistically different than the target of 17% and that they should wait to see the results of next year’s sample before jumping to any conclusions.