Note: R code is in courier font

1. Find & download data

# Our data files are located on D2L

# Copy "16 fish.csv" onto your desktop.

1. Open the data file

# reset the working directory to your desktop (use the location info you saved from last week):

# setwd("/Users/Evan/Documents/courses/383.2016")

setwd("/Users/Evan/Desktop")

# alternatively:

# to open data, first change the working directory by using the toolbar:

on a mac: click on *misc/change working directory*

on a pc: click on *file/change dir*

then select the folder where you put the data – the desktop works well; it might be tricky to find the desktop in windows; it’s usually under users/yourname/desktop

# to find your working directory:

getwd()

# In R, use the following to open the data:

fish.data = read.csv("16 fish.csv", header=T)

# now print the data to the screen:

fish.data

# print the first few lines of the dataframe; useful if the data set is large

head(fish.data)

# Attach the names of the variables to the actual vectors (lists) of numbers, and then print the names of variables:

attach(fish.data)

names(fish.data)

1. Make a histogram to inspect invert.biomass;   
   here we just looking to see if the data seem really weird or not; is the distribution odd?

1. The variable called ‘water’ is a factor with two levels: clear and humic (humic means stained brown from humic acids that are formed when organic matter such as leaves decompose in fresh water); first have a look at the variable ‘water’:

water

I would like to know if mean invert.biomass is different in clear versus humic lakes.

Write the code, run an appropriate test, and interpret the output.

t.test(invert.biomass~water)

Welch Two Sample t-test

data: invert.biomass by water

t = 4.0569, df = 41.729, p-value = 0.0002129

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

6.565489 19.567755

sample estimates:

mean in group clear mean in group humic

97.50897 84.44235

Group mean clear > group mean humic. P value < 0.05 reject null where mean is same in clear and humic.

1. Make a boxplot of invert.biomass as a function of water; paste it below and interpret it.

  
median of clear > median of humic

1. Test whether or not invert.biomass have different variance in the two levels of water.
2. Run a non-parametric t-test (i.e., Wilcoxon test) or a Monte Carlo t-test that compares the mean invert.biomass across the two levels of water and interpret the results.
3. A second set of samples were collected at the same locations but the second samples were collected using different sampling methods. Were the two sample methods equivalent or was one biased in comparison to the other?  
   (yes, paired)
4. Interpret the following:

> tapply(invert.biomass, water, mean)

clear humic

97.50897 84.44235

> tapply(invert.biomass, water, ci95)

clear humic

2.854583 5.936681

################## start chunk ####

MC.t.test = function (Y, X, reps) {

mwd = tapply(Y, X, mean)

diff = abs(mwd[1]-mwd[2])

# reps = 999

rrdiff = rep(0, reps)

for (i in 1:reps) {

r.Y = sample(Y, length(Y))

rmwd = tapply(r.Y, X, mean)

rdiff = abs(rmwd[1]-rmwd[2])

rrdiff[i] = rdiff

}

cases = reps - length(which(rrdiff < diff))

p = (cases + 1)/(reps + 1)

dat.out = data.frame(p, diff)

row.names(dat.out) = c("output")

return(dat.out)

}

MC.t.test(Worm.density, Damp, 1000)

################## end chunk ####