#kruskal wallice test

#csv taken from here. https://data.gov.uk/dataset/1c02d0d2-e3ca-4e89-a567-1f10156d00f8/allocation-of-science-and-research-funding-2011-12-to-2014-15

setwd("/Users/jemy/desktop/advanced statistics") #work computer

rdfunding <- read.csv("rd.csv", sep=",", header=T)

print(rdfunding)

#lets first of all test for normality

typeof(rdfunding$X2011) #wrong data type lets convert

#lets convert to double or numeric to conduct shapiro test

rdfunding$X2011 <- as.numeric(rdfunding$X2011)

rdfunding$x2012 <- as.numeric(rdfunding$x2012)

rdfunding$x2013 <- as.numeric(rdfunding$x2013)

rdfunding$x2014 <- as.numeric(rdfunding$x2014)

#lets now test for normality. Let it begin lets set alpha at 0.05

shapiro.test(rdfunding$X2011)

shapiro.test(rdfunding$x2012)

shapiro.test(rdfunding$x2013)

shapiro.test(rdfunding$x2014)

#unfortunately, we must end the test here. The p is greater than our alpha meaning means are not significantly different. Therefore, we need to find a new dataset.

################################################################

################################################################

seed <- read.csv("RiceFarms.csv", sep=",", header=T)

#price of rice production materials. pesticide, seed, urea fertiliser, phosphorate cost per kg

#lets tets for normality now

shapiro.test(seed$pseed\_cost\_per\_kg) #p is 2.2e-16, less than alpha

shapiro.test(seed$purea\_cost\_per\_kg) #p is 8.018e-12, less than alpha

shapiro.test(seed$pesticide\_cost\_per\_kg)#p is 2.2e-16, less than alpha

shapiro.test(seed$pphosph\_cost\_per\_kg) # p is 1.197e-12, less than alpha

#lets now try visualise this.

boxplot(seed$pseed\_cost\_per\_kg)

boxplot(seed$purea\_cost\_per\_kg)

boxplot(seed$pesticide\_cost\_per\_kg)

boxplot(seed$pesticide\_cost\_per\_kg)

hist(seed$pseed\_cost\_per\_kg)

hist(seed$purea\_cost\_per\_kg)

hist(seed$pesticide\_cost\_per\_kg)

hist(seed$pesticide\_cost\_per\_kg)

#clearly not normal data

Groups.kw = kruskal.test(seed)

Groups.kw

######### data not good and nothing can really be derived from this dataset. On the look out for new data now.

tests <- read.csv("Tests.csv", sep=",", header=T)

boxplot(tests$verbal\_test\_1)

shapiro.test(tests$verbal\_test\_1)

# this data is also not good as P is greater than alpha.

######### ######### ######### ######### ######### ######### ######### ######### #########

weight <- read.csv("weightloss.csv", sep=",", header=T)

#significant differences has been found between the groups.

##lets try the data set, lakes. http://vincentarelbundock.github.io/Rdatasets/doc/DAAG/greatLakes.html <- can be found at this link

#

#the distribution again is normal unfortunately.

#################THIS DATA SET WILL BE THE FINAL DATA SET THAT WILL BE TESTED FOR KRUSKAL############################

rats <- read.csv("fatrats.csv", sep=",", header=T)

attach(rats)

names(rats)

rats$Source <- as.factor(rats$Source)

rats$Source = factor(rats$Source,labels = c("Beef", "Cereal", "Pork"))

class(rats$Source)

Beef <- subset(rats, Source == "Beef")

Cereal <- subset(rats, Source == "Cereal")

Pork <- subset(rats, Source == "Pork")

boxplot(Beef$Gain)

shapiro.test(Beef$Gain) # p = 0.9391

shapiro.test(Cereal$Gain) # p = 0.783

shapiro.test(Pork$Gain) #p = 0.7913

boxplot(Beef$Gain, Cereal$Gain, Pork$Gain) # we can see that all the means are visually normal

qqnorm(Beef$Gain)

qqline(Beef$Gain) #looks very normal

qqnorm(Beef$Gain)

qqline(Cereal$Gain) #looks very normal

qqnorm(Beef$Gain)

qqline(Pork$Gain) #looks very normal

#null hypotheseses are true. We fail to reject the null hypothesis as the means of all groups are not signficantly different.

kruskal.test(Gain ~ Source) #0.5782

#We fail to reject the null hypothesis. Our p-value from kruskals test is greater than our alpha 0.05.

#this tells us that the diet option for the rats do not have significant difference in the level of output.

#lets do a multiple comparisons check

install.packages("FSA")

library("FSA")

dunnTest(Gain ~ Source) #we can see that cereal and pork have an equal means. Other wise, between beef and pork and pork and beef and ceral there is still strong simialirty in means