Homework 2

Sanduni Talagala

January 14, 2020

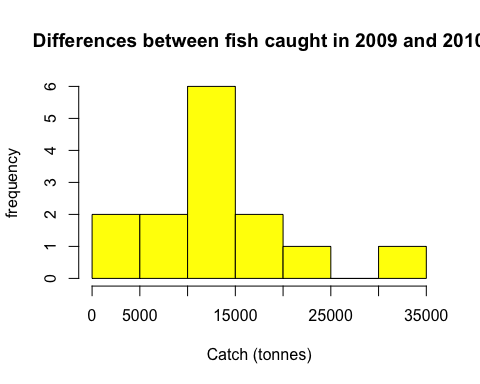
## Question 1: Shrimping and oil spill -> Has the mean shrimp catch in the Gulf changed between the 2009 and 2010 seasons?

**Part a : Hypotheses** H0: There is no difference between the mean catch from 2009 and the mean catch from 2010

HA:The mean catch from 2009 and the mean catch from 2010 are different.

**Part b: histogram** Note: since we are comparing the difference between the 2 years and want to see if the difference is normally distributed, I made a histogram of the difference rather than 2 histograms for the 2 years.

Catch<-read.csv("FishCatch.csv")  
hist(Catch$Difference, main="Differences between fish caught in 2009 and 2010",xlab="Catch (tonnes)", ylab="frequency", col = 'yellow')



**Part c: Type of test**

shapiro.test(Catch$Difference)

##   
## Shapiro-Wilk normality test  
##   
## data: Catch$Difference  
## W = 0.94139, p-value = 0.4365

The differences are normally distributed (p>0.05 in Shapiro-Wilk test). This can also be seen by looking at the histogram of the differences (as it shows a bell shape curve). I assume the sample is random. Furthermore, these are the same lake spots in 2 different years so I will do a **paired t-test** as each lake in 2009 and 2010 has everything (almost - we assume) in common except for the oil spill differences. The question asks if there is a change (does not ask in what direction), so I will run a **two-tailed test**.

**Part d: Run test**

TwoTailedT1<- t.test(Catch$CatchBefore, Catch$CatchAfter, paired=TRUE, alternative = "two.sided")  
TwoTailedT1

##   
## Paired t-test  
##   
## data: Catch$CatchBefore and Catch$CatchAfter  
## t = 6.3808, df = 13, p-value = 2.415e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 8882.005 17975.138  
## sample estimates:  
## mean of the differences   
## 13428.57

**Part e: Critical and observed t** Observed test statistic is calculated above (6.3803). The Critical value is shown below (1.770933)

CriticalValueQ1 <- qt(0.95,13)  
CriticalValueQ1

## [1] 1.770933

**Part f: Conclusion** The observed test statistic is further from zero than the critical value. Furthermore, the samples from the 2 years are significantly different because the p-value is 2.415e-05. Therefore we can **reject the null hypothesis** that the catch from 2009 and 2010 are not different, as catch in the Gulf seemed to have significantly changed between 2009 (before oil spill) and 2010 (after oil spill).

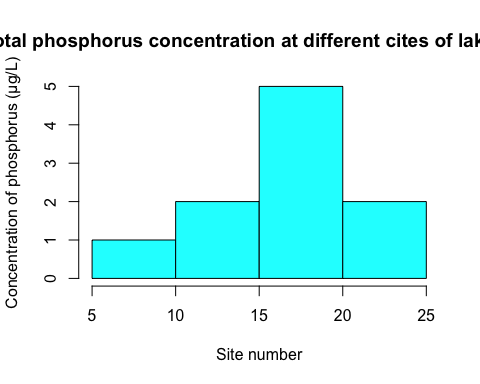
## Question 2: Lake Erie phosphorus ->Are the phosphorous levels in the lake higher than the guidelines?

**Part a: Hypotheses** HO: The mean phosphorus amount in Lake Erier is not greather than the threshold of 15μg/L.

HA:The mean phosphorus amount in Lake Erier is greather than the threshold of 15μg/L.

**Part b: histogram**

ErieP <-read.csv("ErieP.csv")  
hist(ErieP$TotalPhosphorus, main="Total phosphorus concentration at different cites of lake Erie",xlab="Site number", ylab="Concentration of phosphorus (μg/L)", col='cyan')



**Part c: Type of test**

shapiro.test(ErieP$TotalPhosphorus)

##   
## Shapiro-Wilk normality test  
##   
## data: ErieP$TotalPhosphorus  
## W = 0.96489, p-value = 0.8398

The total phosphorus concentration is normally distrubuted (p>0.05 in Shapiro-Wilk test) and the histogram further confirms this (bell-shaped). I assume the data was randomly sampled. Therefore, I will run a **One-Sample t-test** where I will compare the observed values to the known limit. The test will be **one-tailed** as the question asks if the P concentrations are greater (not if they are different).

**Part d: run test**

OneTailedT2 <- t.test(ErieP$TotalPhosphorus, mu=15, alternative = "greater")  
OneTailedT2

##   
## One Sample t-test  
##   
## data: ErieP$TotalPhosphorus  
## t = 0.80178, df = 9, p-value = 0.2217  
## alternative hypothesis: true mean is greater than 15  
## 95 percent confidence interval:  
## 13.27637 Inf  
## sample estimates:  
## mean of x   
## 16.34

**Part e: Critical and observed t** The observed test statistic value can be seen in the above question’s output (0.80178). The critical value is shown below (1.833113).

CriticalValueQ2 <- qt(0.95,9)  
CriticalValueQ2

## [1] 1.833113

**Part f: Conclusion** The observed t statictic (0.80178) is closer to zero than the critical value. Furthermore as seen in the test output, the p-value is greater than 0.05. This means that we **cannot reject the null hypothesis** that the phosphorus levels observed in the lake are not significantly higher than the set maximum (this means the phosphorus levels do not seem to be above tghe set maximum).

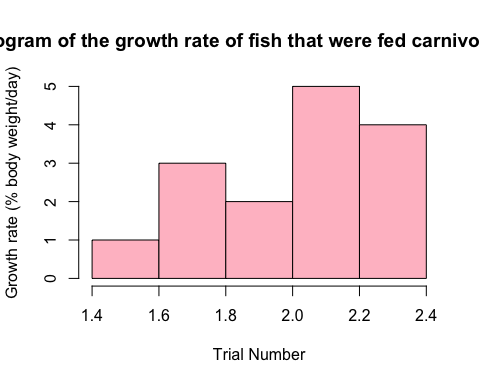
## Question 3: aquaculture -> Does the vegetarian diet differ from the carnivorous diet in terms of growth rate in farmed salmon?

**Part a : Hypotheses** H0: There is no difference in the growth rate between farmed salmon given the vegetarian diet and the carnivorous diet.

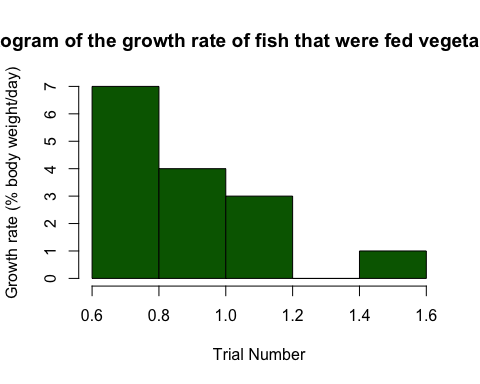
HA: The growth rate between farmed salmon given the vegetarian diet and the carnivorous diet are different.

**Part b: histogram** Note: Here both samples have to be normally distributed, so I will show both distributions.

FishFood <-read.csv("FishDiet.csv")  
hist(FishFood$Carnivorous, main="Histogram of the growth rate of fish that were fed carnivorous diet",xlab="Trial Number", ylab="Growth rate (% body weight/day)", col ='pink')



hist(FishFood$Vegetarian, main="Histogram of the growth rate of fish that were fed vegetarian diet",xlab="Trial Number", ylab="Growth rate (% body weight/day)", col = 'darkgreen')



**Part c: Type of test**

shapiro.test(FishFood$Carnivorous)

##   
## Shapiro-Wilk normality test  
##   
## data: FishFood$Carnivorous  
## W = 0.88016, p-value = 0.0477

shapiro.test(FishFood$Vegetarian)

##   
## Shapiro-Wilk normality test  
##   
## data: FishFood$Vegetarian  
## W = 0.9114, p-value = 0.1423

The data does not look normally distributed. However, for the purposes of this assignment I will run a **Weltch two-sample t-test** to compare the means of these 2 independent groups.Since I am running a weltch t-test, I do not have to worry about the assumption that the variances are equal (even though they are almost the same variances). The test will be **two-tailed** as the question asks if there are any differences (either side)

**Part d: run test**

t.test(FishFood$Carnivorous, FishFood$Vegetarian, alternative = "two.sided")

##   
## Welch Two Sample t-test  
##   
## data: FishFood$Carnivorous and FishFood$Vegetarian  
## t = 12.237, df = 27.862, p-value = 1.008e-12  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.9435752 1.3230915  
## sample estimates:  
## mean of x mean of y   
## 2.053333 0.920000

**Part e: Critical and observed t** The observed value of the test statistic is seen in the above output (12.237) and the critical value is shown below (1.701419).

CriticalValueQ3 <- qt(0.95,27.862)  
CriticalValueQ3

## [1] 1.701419

**Part f: Conclusion**

The observed test statistic (12.237) is further from zero than the critical value (1.701419). Furthermore, the p value is less than 0.05. Therefore we can **reject the null hypothesis** that the growth rate of the fish under the two diets are not different (this means that there is some difference in the growth rate by altering the fish diet).

## Question 4: PCB effects -> is there any interaction between symptoms and the trophic level of the individual in question?

**Part a: Hypotheses** HO: The severity of symptoms from PCB is independent of the trophic level of the species.

HA: The severity of symptoms from PCB is dependent of the trophic level of the species.

**Part b: histogram** For this question, normallity is not an assumption, so I did not make a histogram (as suggested).

**Part c: Type of test**

FirstLevel = c(133,100,17)  
TopPredator = c(4,61,35)  
dat = data.frame('F'= FirstLevel, 'P'=TopPredator)  
chisq.test(dat)$expected

## F P  
## [1,] 97.85714 39.14286  
## [2,] 115.00000 46.00000  
## [3,] 37.14286 14.85714

In this case I’m trying to determine if there is an interaction between tropic levels and the effects of PCB, so I will use a **χ2 contingency analysis**. Here, after I made an expected table I see that I don’t have any expected values less than 1 and no more than 20% of the expected values are less than 5. This means assumptions are not violated.

**Part d: run test**

Testing = chisq.test(dat)  
Testing

##   
## Pearson's Chi-squared test  
##   
## data: dat  
## X-squared = 89.253, df = 2, p-value < 2.2e-16

**Part e: critical and observed value** The observed test statistic value is shown above (89.253). the Critical value is calculated below (2.919986)

CriticalValueQ4 = qt(0.95,2)  
CriticalValueQ4

## [1] 2.919986

**Part f: Conclusion** The observed test statistic is further from zero than the critical value. Furthermore, according to the output the p value is less than 0.05. Taken together, this allows us to **reject the null hypothesis** that there is no interaction between the symptoms and the trophic level, as there seems to be an interaction.