Beckham Carver

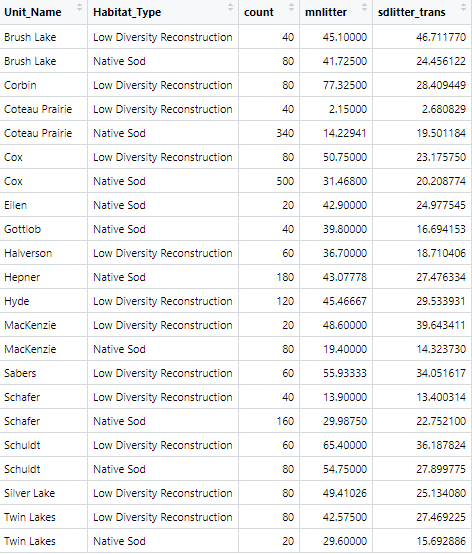
Feb 3, 2022

STAT 4025

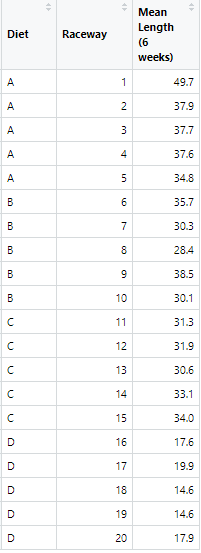
Prof. Robinson

Homework Assignment

1. **At the end of this paper all my code is included**, here is my table for problem 1.



A)



B) Our null hypothesis would be that “we find no difference in mean change between diets A and B” which is mathematically represented as muA – muB = 0. If we are simply looking to prove a difference in diets, this two-tailed hypothesis test would suffice, so our alternative hypothesis is Hₐ = “we find a differences in mean change between diets A and B” or muA – muB ≠ 0. If we were specifically trying to argue for one diet over another- a one-tailed test- we would change our alternative hypothesis accordingly.

C) When performing our hypothesis test, we use mu to represent the population mean. Meaning muA represents the population mean observed from the tanks fed with diet A and muB representing the population mean observed from diet B. The null hypothesis assumes that the difference between these two values will be equal to zero or lack statistical significance to support there being a true difference in means. As for the alternative hypothesis, it assumes that there will be a statistically significant difference in population mean lengths between diet A and B.

D) We will be using the line **t.test(mnLength ~ Diet, var.equal=F, data=trout\_prob2D)** for our t-test. var.equal was set to false because this makes the t-test more scrupulous on the data, in this case using true or false produced the same conclusion with slightly varying specifics.

It is valid to say that RATG would allow us to assume var.equal=TRUE however it is often best practice to use the most meticulous methods and, in this case, assuming equal variance was false was more scrupulous on the data**.** Nonetheless concluding equal variance between 10 experimental units is subjective and either method provides the same conclusion.

Output: data: mnLength by Diet

t = 2.1462, df = 7.355, p-value = 0.06713

alternative hypothesis: true difference in means between group A and group B is not equal to 0

95 percent confidence interval:

-0.632226 14.512226

sample estimates:

mean in group A mean in group B

39.54 32.60

Notably, our p-value was not within out 0.05 threshold, furthermore our confidence interval includes zero, this means we cannot conclude that there is a true difference in population means between diet A and diet B. Therefore, we fail to reject our null hypothesis. Although a difference is observed in these small sample groups, it would be disingenuous to extrapolate these results to the population as our small p-value does not provide confidence that there is a true difference in means.

E) The two groups in our two-group comparison are the tanks fed diet A and the tanks fed diet B. In this case the desired outcome is the highest mean increase in length among the tanks of fish, and group A and B are “competing” to have the largest increase. Proving there is a ‘winner’ of this competition first requires determining if there is a difference between the groups at all. With a p-value of 0.06713 we CANNOT conclude that there is a true difference in means between the groups, therefore pitching the groups against each other in a 1-tailed test was not necessary as the groups essentially ‘tied’ before they raced. The data suggests there is a possibility that diet A could have advantages, however further testing would be needed for any confident conclusions.

F) The p-value 0.06713 represents a 6.137% chance that the **true** difference in population means between groups A and B (39.54 and 32.60 respectively) is zero, and not what was exhibited. That is to say there is a 6.137% chance that there is no real difference between the diets. This value is greater than the requested 5% cutoff for determining true difference in means, which fails to reject our null hypothesis.

**CODE:**

# Beckham Carver

# STAT 4015

# Prof. Robinson

# Homework 3

ppr <- read.csv("PPR\_2022.csv", header = T)

trout <- read.csv("trout\_hatchery\_22.csv", header = T)

library(tidyverse)

# Part 1:

#psuedocode:

# define ppr\_prob1, using ppr for data

# filter for Low Diversity, or Native Sod

# filter remaining for year 2018

# group data structure by WPA (Unit\_Name && Habitat\_type)

# analyze with mean and SD of litter depth

ppr\_prob1 <- ppr %>%

filter((Habitat\_Type=="Low Diversity Reconstruction"

|Habitat\_Type=="Native Sod")

& Monitoring\_Year==2018

& Prioritization=="High")%>%

group\_by(Unit\_Name,Habitat\_Type)%>%

summarise(count=n(),

mnlitter=mean(LitterDepth,na.rm=T),

sdlitter\_trans=sd(LitterDepth,na.rm=T))

# Part 2 (A)

#psuedocode:

# define trout\_prob2, using trout data

# group by both diet and raceway

# within these groups summarize length into a mean

# remove the count column

# rename columns

trout\_prob2 <- trout %>%

group\_by(Diet,raceway)%>%

summarise(count=n(),mn\_length=mean(Length\_6w\_mm,na.rm=T))%>%

mutate(count = NULL)

colnames(trout\_prob2) <- c("Diet","Raceway","Mean Length (6 weeks)")

# Part 2 (D)

# filtering downs to the two needed diets

trout\_prob2D <- trout\_prob2 %>%

filter(Diet=="A" || Diet=="B")

#renaming for convenience

names(trout\_prob2D)[3] <- 'mnLength'

t.test(mnLength ~ Diet, var.equal=F, data=trout\_prob2D)