Pr\_1\_scratch

Robert McNeil

2024-09-03

#look at all ten, correct with bonferonnis method

#Lets read in CSV file and load packages we’ll need to run code.

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(dplyr)  
  
#lme4 is for a mixed effects model- if it errors when you load it may need dependencies first  
library(lme4)

## Loading required package: Matrix  
##   
## Attaching package: 'Matrix'  
##   
## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

#load dataset   
gum\_treatment <- read.csv("/home/robmcneil/Documents/advanced\_data/Project1\_data.csv")

Details:

Dental researchers were interested in a new gel treatment for gum disease. In their study, subjects were randomly assigned to one of 5 treatments: no treatment and four levels of an active substance in a gel. The lowest level was a placebo (=1) and then another control group (=2), like the “no treatment” group. The remaining three levels were low (=3), medium (=4), and high (=5) concentrations of the active ingredient in the gel. The patients were recruited via a single midwestern dental research clinic. There were 130 participants. Participants were asked to rub the gel on their gums twice daily. The measurements being followed over time were whole-mouth average pocket depth and whole mouth average attachment loss. Visits were at baseline and 1 year. Pocket depth and attachment loss were measured at many sites within each participant’s mouth and then averaged (at each visit). The variable called “sites” gives the number of sites used in the averages. Pocket depth and attachment loss are both measures of how far the gums have pulled away from the teeth, hence smaller values are better. Whole-mouth average is used because the measurements within mouth are highly correlated in a complicated way, which is not yet fully understood by dental researchers. Additional demographic information was collected on race (5=white, 4=Asian, 2=African American, 1=Native American), gender (1=male, 2=female), age, and smoking status (yes=1/no=0). Missing values are denoted by an “NA”. The primary question of Interest of this study is whether treatment results in lower average pocket depth and attachment loss at one year. The data file is project1\_data.csv, which can be accessed in the Project 1 on Canvas.

###### 

#data inspection with str function- number of observations, :

str(gum\_treatment)

## 'data.frame': 130 obs. of 11 variables:  
## $ id : int 101 102 103 104 105 106 107 108 109 110 ...  
## $ trtgroup : int 4 5 2 3 1 4 3 2 5 1 ...  
## $ gender : int 2 2 2 2 2 2 2 2 2 2 ...  
## $ race : int 5 5 5 5 2 5 5 5 2 5 ...  
## $ age : num 44.6 35.6 47.9 55.2 43.8 ...  
## $ smoker : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ sites : int 162 162 144 138 168 168 156 162 156 162 ...  
## $ attachbase : num 2.43 2.54 2.88 4.96 1.77 ...  
## $ attach1year: num 2.58 NA 3.08 5.3 1.45 ...  
## $ pdbase : num 3.25 3.01 3.12 5.22 3.36 ...  
## $ pd1year : num 3.41 NA 3.12 4.89 2.9 ...

#need to do a data quality check- make sure all subject ID's occur once

# Summary statistics for baseline and 1-year measurements:  
# minumum, maximum, median, mean   
#frequenct proceure for demographic distrobution  
summary(gum\_treatment$attachbase)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.8951 1.5792 2.0275 2.1461 2.5849 5.0893

summary(gum\_treatment$attachoneyear)

## Length Class Mode   
## 0 NULL NULL

#address missing data - through a mixed effects model

#look for high correlation between X values #f stastic for the full model?

model <- lmer(Reaction ~ Days + (Days | Subject), data = sleepstudy)  
model

## Linear mixed model fit by REML ['lmerMod']  
## Formula: Reaction ~ Days + (Days | Subject)  
## Data: sleepstudy  
## REML criterion at convergence: 1743.628  
## Random effects:  
## Groups Name Std.Dev. Corr  
## Subject (Intercept) 24.741   
## Days 5.922 0.07  
## Residual 25.592   
## Number of obs: 180, groups: Subject, 18  
## Fixed Effects:  
## (Intercept) Days   
## 251.41 10.47

#create dummy variables for groups   
gum\_treatment <- gum\_treatment %>%  
 mutate(  
 group1 = if\_else(trtgroup == '1', 1, 0),  
 group2 = if\_else(trtgroup == '2', 1, 0),  
 group3 = if\_else(trtgroup == '3', 1, 0),  
 group4 = if\_else(trtgroup == '4', 1, 0),  
 group5 = if\_else(trtgroup == '5', 1, 0)  
 )  
  
#create change variable for time   
  
gum\_treatment<- gum\_treatment %>%  
 mutate(  
 change\_attach = attach1year - attachbase,  
 change\_pd = pd1year - pdbase  
 )

#linear regression  
  
model\_attach\_change <- lm(change\_attach ~ trtgroup + gender + race + age + smoker + sites, data = gum\_treatment)  
summary(model\_attach\_change)

##   
## Call:  
## lm(formula = change\_attach ~ trtgroup + gender + race + age +   
## smoker + sites, data = gum\_treatment)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.96167 -0.17009 0.02172 0.16792 0.52683   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.398371 0.455509 -0.875 0.3840   
## trtgroup 0.017159 0.019790 0.867 0.3881   
## gender -0.095071 0.057537 -1.652 0.1018   
## race 0.029473 0.028406 1.038 0.3021   
## age -0.005194 0.002832 -1.834 0.0698 .  
## smoker 0.044708 0.057534 0.777 0.4391   
## sites 0.003279 0.002467 1.329 0.1870   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2688 on 94 degrees of freedom  
## (29 observations deleted due to missingness)  
## Multiple R-squared: 0.1027, Adjusted R-squared: 0.04548   
## F-statistic: 1.794 on 6 and 94 DF, p-value: 0.1086

#model for pocket depth   
  
model\_pdbase\_change <- lm(change\_pd ~ trtgroup + gender + race + age + smoker + sites, data = gum\_treatment)  
summary(model\_pdbase\_change)

##   
## Call:  
## lm(formula = change\_pd ~ trtgroup + gender + race + age + smoker +   
## sites, data = gum\_treatment)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.54239 -0.16839 -0.00042 0.15519 0.65628   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.176076 0.453765 0.388 0.6989   
## trtgroup 0.020915 0.019714 1.061 0.2914   
## gender -0.131350 0.057316 -2.292 0.0242 \*  
## race 0.023877 0.028297 0.844 0.4009   
## age -0.003224 0.002821 -1.143 0.2560   
## smoker -0.017070 0.057313 -0.298 0.7665   
## sites -0.001630 0.002458 -0.663 0.5088   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2678 on 94 degrees of freedom  
## (29 observations deleted due to missingness)  
## Multiple R-squared: 0.07407, Adjusted R-squared: 0.01497   
## F-statistic: 1.253 on 6 and 94 DF, p-value: 0.2866