Lab 3

Jenna, Dan, Ceili

# Lab 3 Instructions:

Working in a group of 2, 3, or 4 people (singletons allowed if you miss the in-person session or could not attend virtually and find a group), complete the following questions. New groups will be randomly assigned.

Turn in a single for the group *with all group member names on it* after knitting this as a word document with your answers “in-line” (after the questions) and then convert the file to a PDF to submit on gradescope.com.

## Factors related to intra-tendinous morphology of Achilles tendon in runners: Part IV

*-* Ho K-Y, Baquet A, Chang Y-J, Chien L-C, Harty M, Bashford G, et al. (2019) Factors related to intra-tendinous morphology of Achilles tendon in runners. *PLoS ONE* 14(8): e0221183. <https://doi.org/10.1371/journal.pone.0221183>

The article is available at <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0221183>

## Updated code to add “Location” variable and “PainCombs” to the data set:

data(TendonData)  
tendon <- TendonData  
  
 tendonN <- tendon %>%  
 rename(SubjectID = 'Subject ID',  
 Sex = 'Sex (M=1)',  
 CurrPain = 'CurrPain (Y=1)',  
 WaisttoHip = 'Waist to Hip',  
 VISAA = 'VISA-A',  
 Neovascularization = 'Neovascularization (Doppler) (Y=1)',  
 HistPain = 'Hx Pain (Y=1)')  
   
 tendonF <- tendonN %>%   
 mutate(across(where(~ is.character(.x)), as.factor))  
   
   
 tendonF <- tendonF %>%  
 mutate(CurrPain = factor(CurrPain),  
 HistPain = factor(HistPain),  
 Neovascularization = factor(Neovascularization),  
 Sex = factor(Sex),  
 Sex = fct\_recode(Sex,  
 "Female" = "0",  
 "Male" = "1"),  
 Location = factor(substr(SubjectID, 1, 3)),  
 PainCombs = as.numeric(str\_c(HistPain, CurrPain))  
 )  
tendonna1 <- tendonF %>% drop\_na(CSA)  
tendonna2 <- tendonF %>% drop\_na()

1. The data set tendonF has a little work done on some variables. What is different between tendonF and tendonN? In other words, what does mutate(across(where(~ is.character(.x)), as.factor)) do? Hint: This might be a time to run summary on both data sets to get some quick information about all the variables to look for changes.

**The tendonF dataset reclassifies the character variables as factors, which then allows for counts based on groups.**

* We will use the completely cleaned tendonna2 data set throughout the rest of this lab.

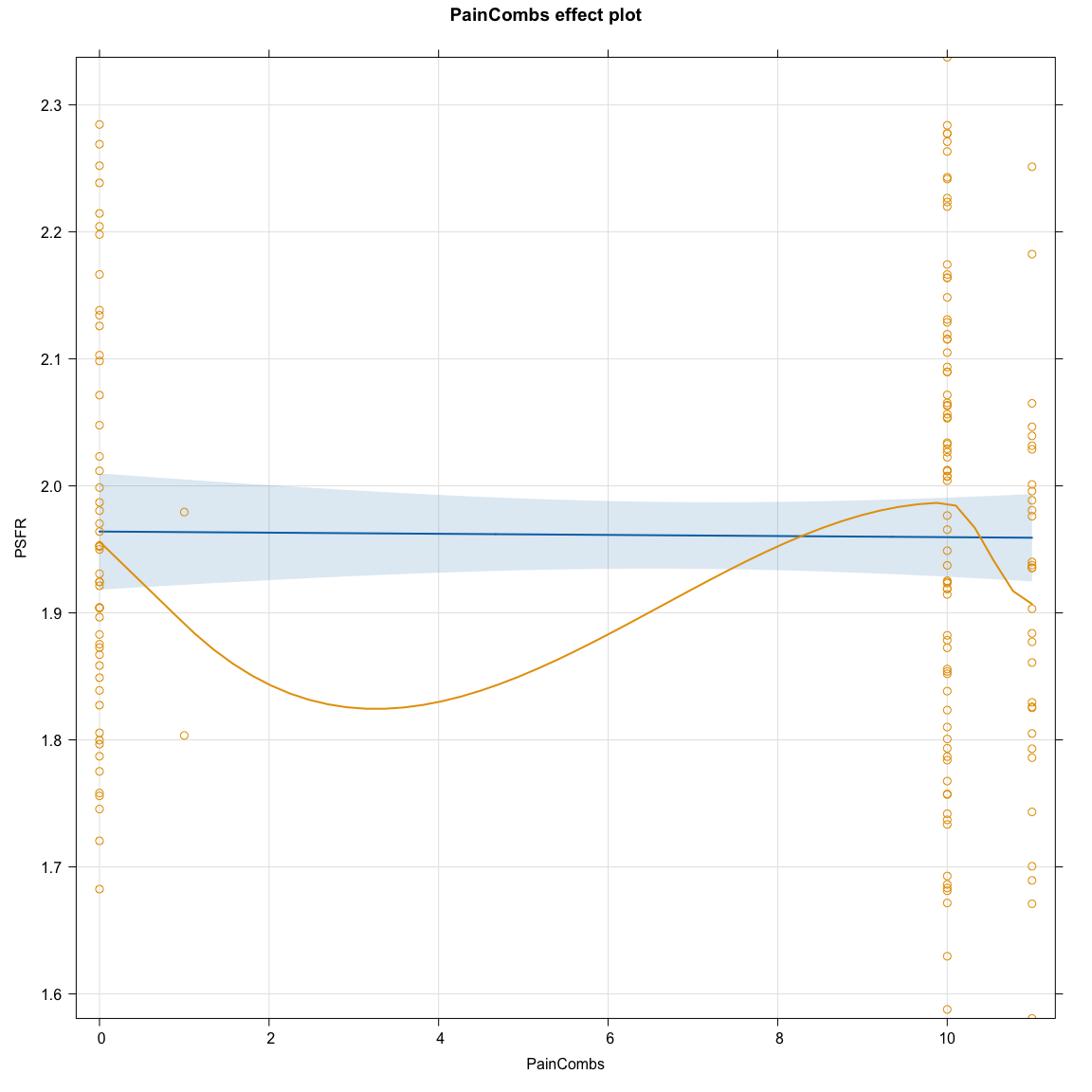
## Multi-category categorical predictor

1. Fit a linear model for PSFR that includes the PainCombs predictor variable only as it currently exists in tendonna2 and make an effects plot of that model. Report and explain the values that are stored in the PainCombs variable. How was the variable treated in your model? Hint: check the PainCombs column in the dataset and understand what the code on line 71 does.

Line 71 strings together the historic pain and current levels and makes it a numeric variable.

**PainCombs has numerical values of 0, 1, 10, and 11. Those correspond to combinations of the History of Pain and Current Pain. We need to dig into the results a bit to understand those levels and give them better names. But this was treated as quantitative in the model and that is not correct. We need to at least convert PainCombs into a factor and it would be even better to make the labels more explicit.**

model\_1 <- lm(PSFR ~ PainCombs, data = tendonna2)  
plot(allEffects(model\_1, residuals = T),1, grid = T)

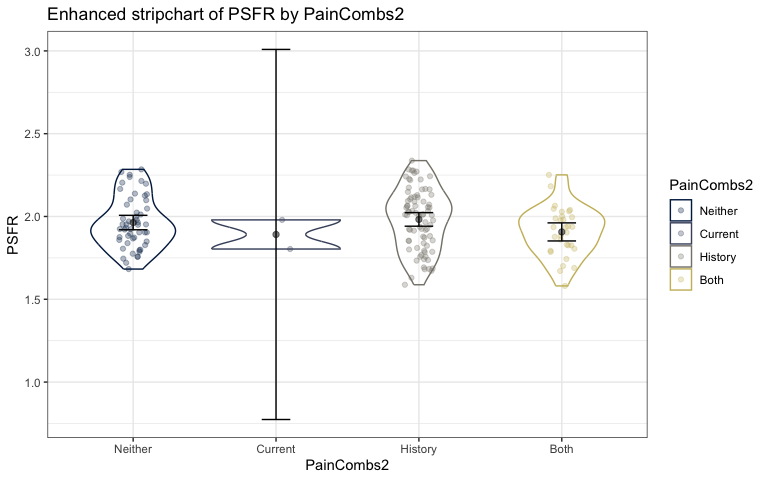


1. Use the following output and the fct\_recode to create a new variable called PainCombs2 that fixes up how PainCombs is labeled as well as generally treating it more correctly. Then generate an enhanced stripchart of PSFR based on the new variable. No discussion and you do not need to modify the axes or title in the plot.

tally(HistPain ~ CurrPain | PainCombs, data = tendonna2)

## , , PainCombs = 0  
##   
## CurrPain  
## HistPain 0 1  
## 0 52 0  
## 1 0 0  
##   
## , , PainCombs = 1  
##   
## CurrPain  
## HistPain 0 1  
## 0 0 2  
## 1 0 0  
##   
## , , PainCombs = 10  
##   
## CurrPain  
## HistPain 0 1  
## 0 0 0  
## 1 79 0  
##   
## , , PainCombs = 11  
##   
## CurrPain  
## HistPain 0 1  
## 0 0 0  
## 1 0 31

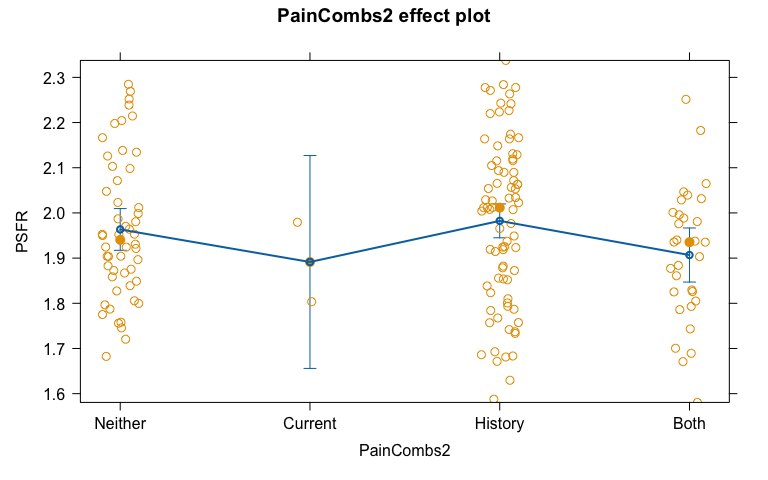
tendonna2 <- tendonna2 %>% mutate(PainCombs2 = factor(PainCombs),  
 PainCombs2 = fct\_recode(PainCombs2,  
 "Neither" = "0",  
 "Current" = "1",  
 "History" = "10",  
 "Both" = "11"))  
  
enhanced\_stripchart(PSFR ~ PainCombs2, data = tendonna2)



1. Fit a linear model for PSFR with the new version of PainCombs as a predictor and make an effects plot from the model. Then write out the estimated model, defining all of your indicator variables.

where is 1 when PainCombs is CP and 0 if else, is 1 when PainCombs is HP and 0 if else, is 1 when PainCombs is PP and 0 if else

model\_2 <- lm(PSFR ~ PainCombs2, data = tendonna2)  
  
plot(allEffects(model\_2, residuals = T))



summary(model\_2)

##   
## Call:  
## lm(formula = PSFR ~ PainCombs2, data = tendonna2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.39465 -0.11606 -0.00146 0.12211 0.35510   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 1.96342 0.02339 83.938 <2e-16  
## PainCombs2Current -0.07201 0.12154 -0.592 0.554  
## PainCombs2History 0.01898 0.03012 0.630 0.530  
## PainCombs2Both -0.05662 0.03827 -1.479 0.141  
##   
## Residual standard error: 0.1687 on 160 degrees of freedom  
## Multiple R-squared: 0.02926, Adjusted R-squared: 0.01106   
## F-statistic: 1.608 on 3 and 160 DF, p-value: 0.1897

1. Uncomment and replace mymodelname with your model name with the corrected version of PainCombs. The code will then extract and display four rows of the model matrix - note that there should be 4 columns and some might be hidden in your initial view in RStudio. Explain the pattern of the contents of the model matrix for each of the rows. Code to print the same rows in the data set for the predictor variable are provided to help you explain the pattern in the design matrix - assuming that the new variables are the last two columns in the data set of 17 and 18.

model.matrix(model\_2) %>% as.data.frame() %>% slice(1,2,10,42)

## (Intercept) PainCombs2Current PainCombs2History PainCombs2Both  
## 1 1 0 1 0  
## 2 1 0 0 0  
## 10 1 0 0 1  
## 42 1 1 0 0

tendonna2 %>% dplyr::select(17:18) %>% slice(1,2,10,42)

## # A tibble: 4 × 2  
## PainCombs PainCombs2  
## <dbl> <fct>   
## 1 10 History   
## 2 0 Neither   
## 3 11 Both   
## 4 1 Current

**The first column in the model matrix is all 1s since it is for the intercept in the model (all observations have in them. Subject #1 (row 1) in the dataset has a history of pain but not current pain, so gets a 0, 1, 0 pattern with the only 1 in the column for history of pain only. Subject #2 (row 2) has no pain, so gets all indicators set to 0. Subject #10 (row 10) has both a history of pain and current pain, so gets 0, 0, and 1 with the indicator set to 1 for the both level. And Subject #42 (row 42) has current pain but not a history of pain, so gets an indicator of 1 for the current pain column but 0s for the other two.**

1. Use favstats to obtain the estimated means for PSFR for each of the levels of the new PainCombs variable. No discussion.

favstats(PSFR ~ PainCombs2, data = tendonna2)

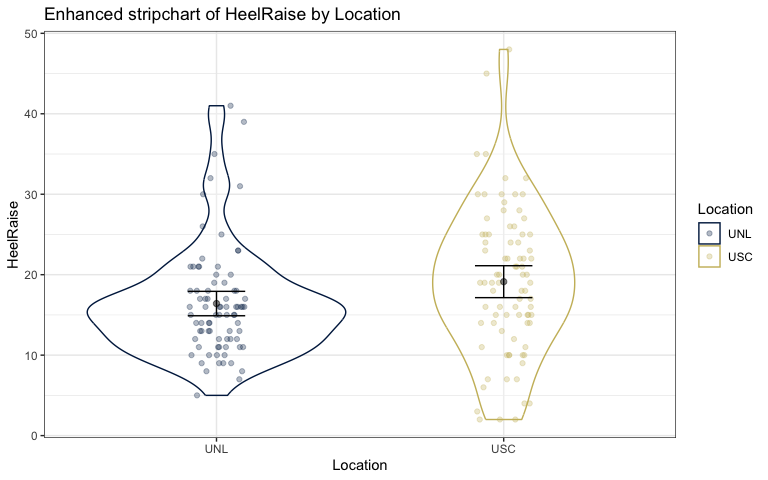
## PainCombs2 min Q1 median Q3 max mean sd n  
## 1 Neither 1.682439 1.856104 1.940357 2.078183 2.284604 1.963421 0.1566721 52  
## 2 Current 1.803465 1.847439 1.891414 1.935388 1.979363 1.891414 0.1243789 2  
## 3 History 1.587747 1.845178 2.011611 2.115495 2.337499 1.982401 0.1833098 79  
## 4 Both 1.580619 1.815095 1.935324 1.998517 2.251339 1.906802 0.1487741 31  
## missing  
## 1 0  
## 2 0  
## 3 0  
## 4 0

1. Generate an estimated mean from the previously estimated linear model for subjects with neither history of pain nor current pain. Show your work or explain how you found the result. Make sure you check your work versus the results from favstats in the previous question.
2. Generate an estimated mean from the model for subjects with current pain but no history of pain. Show your work.

## Heel Raise as a response variable

1. Make an enhanced stripchart of the HeelRaise variable based on the Location (university where the measurements were made) variable. No discussion and you do not need to modify the axes or title.

enhanced\_stripchart(HeelRaise ~ Location, data = tendonna2)



1. From their reference number 29, I found “The number of standing unilateral heel-raises above a level 5 cm from the floor was counted. The starting side for the test was the same as in the isokinetic tests. Each test was performed on both sides, followed by a 4 min rest… The leg that was being tested had to be kept straight. The test was stopped when the subject was unable to continue performing proper heel-raises due to fatigue.” Assuming that the testing protocol was the same in this study, what are the units of the measurements in HeelRaise? What is the smallest and largest observed value in the data set?

* The smallest measurement recorded was 2.0 heel raise repetitions and the largest was 48.0 heel raise repetitions.

summary(tendonna2$HeelRaise)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.0 12.0 16.0 17.8 22.0 48.0

1. Document any resources used outside of your fellow group members and course provided resources. If you do not use any, report “NONE” to get credit for this question.

* NONE