

Lab 2

Rowan Edwards, Lois Deschaux, Ceili DeMarais

1/22/2025

Lab 2 Instructions:

Working in a group of 2, 3, or 4 people (singletons allowed if you miss the in-person session and did not find a group), complete the following questions. Turn in a single word document for the group *with all group member names on it* after knitting this document with your answers “in-line” (after the questions).

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

- 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>

```
library(tidyverse)
data(TendonData) #Version of data set from catstats2
tendon <- TendonData

tendon <- 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)',
    HistoryPain = 'Hx Pain (Y=1)') %>%
  mutate_if(is.character, as.factor) %>%
  mutate(CurrPain = factor(CurrPain),
         Neovascularization = factor(Neovascularization),
         Sex = factor(Sex),
         HistoryPain = factor(HistoryPain))

tendonna1 <- tendon %>% drop_na(CSA)
tendonna2 <- tendon %>% drop_na()
```

For the following questions, use tendonna2:

- 1) The levels for Sex are less than ideal (and we stripped off the definition from the variable name - see code above). Uncomment and modify the following code to change the levels to be more explicit in the Sex factor variable:

```
levels(tendonna2$Sex)

## [1] "0" "1"

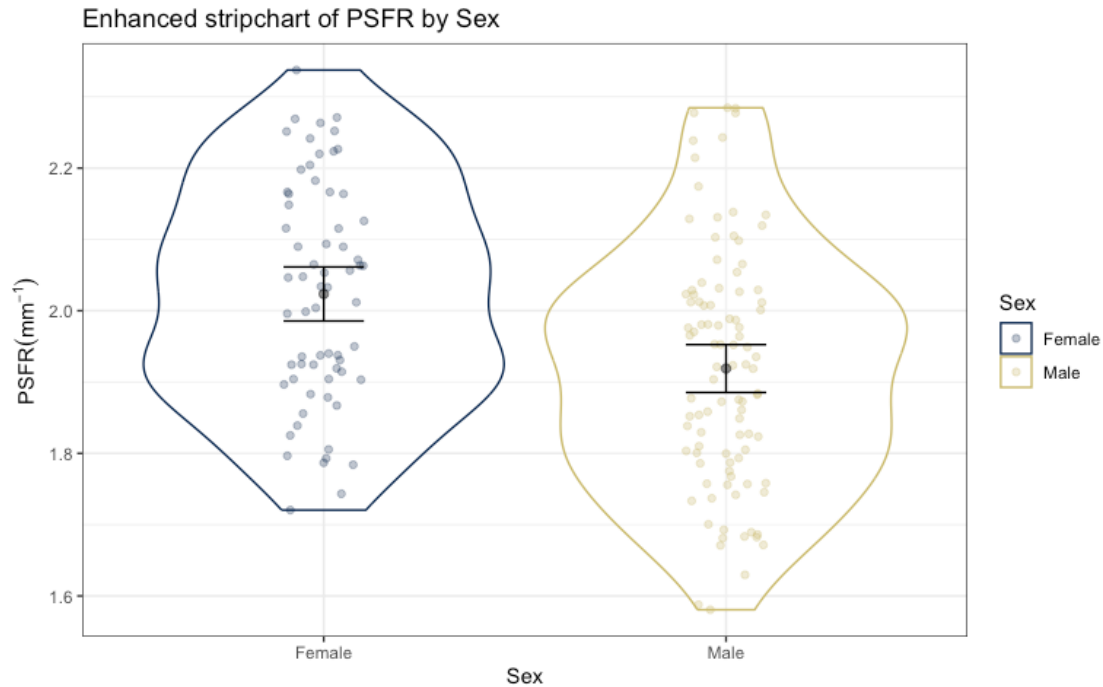
tendonna2 <- tendonna2 %>% mutate(
  Sex = fct_recode(Sex,
    "Female" = "0",
    "Male" = "1")
)

levels(tendonna2$Sex)

## [1] "Female" "Male"
```

- 2) Make a plot of PSFR (y-axis) by Sex from the tendonna2 data set using an enhanced_stripchart from the catstats2 package. Since the enhanced_stripchart is just a wrapper to a pile of ggplot2 code, you can use some of the modifications for ggplots to make it better. For example, you can modify labels using + labs(y = "better y-axis label"). Improve the y-axis label to include the units the authors mention for PSFR (see their Table 1, for example). I set up a spot in parentheses after the variable name where you can try out TeX code to get the units included. You can leave the title as it is since the function adds a decent one by default.

```
library(latex2exp)
enhanced_stripchart(PSFR ~ Sex, data = tendonna2) + labs(y = TeX("$PSFR
(mm^{-1})$"))
```



- 3) Generate descriptive statistics of PSFR by Sex groups using favstats from the mosaic package. If you can't get mosaic to work, you can use group_by as shown in the notes to find at least the min, max, and mean for each group.

```
favstats(PSFR~Sex, data = tendonna2)
```

```
##      Sex      min      Q1    median      Q3      max      mean      sd
## 1 Female 1.720478 1.906823 2.033386 2.160001 2.337499 2.023477 0.1541363
## 2   Male 1.580619 1.794935 1.922351 2.023034 2.284604 1.918896 0.1672685
## missing
## 1      0
## 2      0
```

- 4) Based on these results, what is the estimated difference in the mean PSFR levels between male and female subjects? Make sure you make it clear which one is larger/smaller in writing a sentence to report this (“*difference of ...*” is not read by different people the same way).
- The estimated difference in the mean PSFR levels between Male and Female subjects is -0.104581mm^{-1} (Male - Female). Male subjects had a mean PSFR level that was 0.104581mm^{-1} less than females.
- 5) Fit an lm with PSFR and as the response and Sex as the predictor. Generate a base R model summary, generate a nice table of results using `%>% tbl_regression(intercept = T) %>% add_global_p()`, and make an effects plot using the effects package (so something like `plot(allEffects(modelname))`). In

Table 3 in the paper, which level of Sex was their reference level? In your model, which one is the reference and how can you tell?

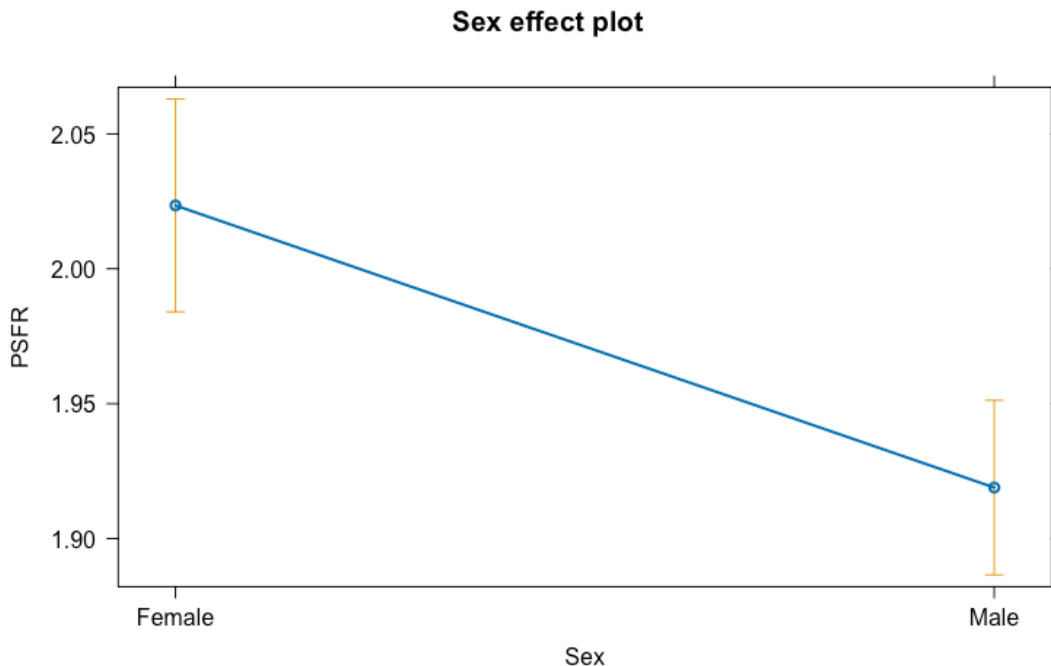
- In table 3 their reference level is ‘female’, which is the same in our model. We can see this in the regression table as the female level is blank and the male level has an estimated difference.

```
model <- lm(PSFR ~ Sex, data = tendonna2)
model %>% tbl_regression(intercept = T) %>% add_global_p()
```

Characteristic	Beta	95% CI ¹	p-value
(Intercept)	2.0	2.0, 2.1	<0.001
Sex			<0.001
Female	—	—	
Male	-0.10	-0.16, -0.05	

¹CI = Confidence Interval

```
plot(allEffects(model))
```



6) Write out the estimated model, defining the indicator variable you used.

- $\hat{\mu}\{PSFR|Sex\} = 2 - 0.1I_{Sex=Male}$

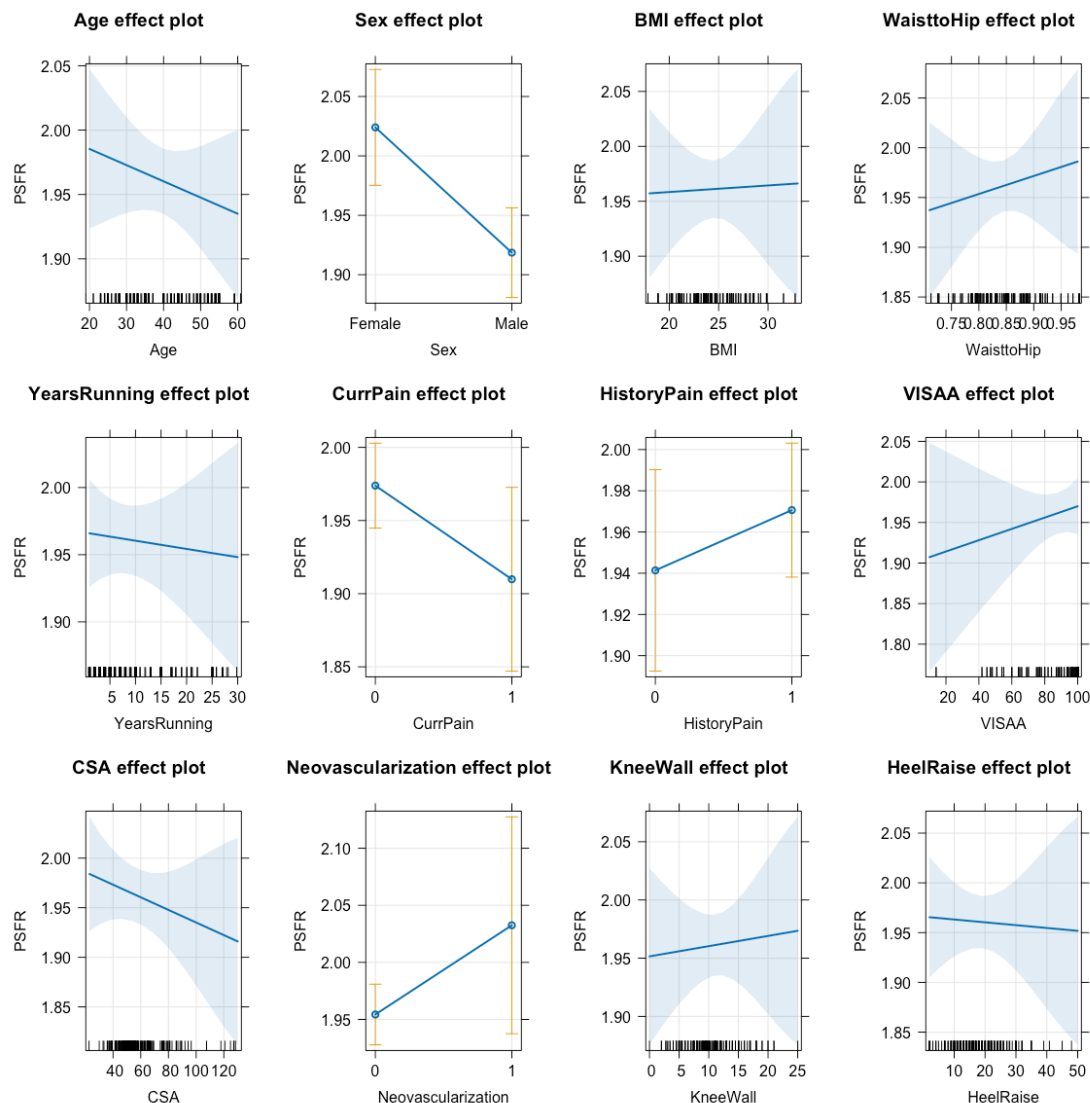
○ where $I_{Sex=Male}$ is 1 for a Male observation and 0 otherwise

7) Replace the “...”s and make choices from the []s in the provided “size” interpretation for the slope coefficient for the Sex row in the model summary, finding and reporting a 95% confidence in parentheses as part of that single sentence. Hint: the confint function is an easy way to obtain a confidence interval or you can use the output from tbl_regression.

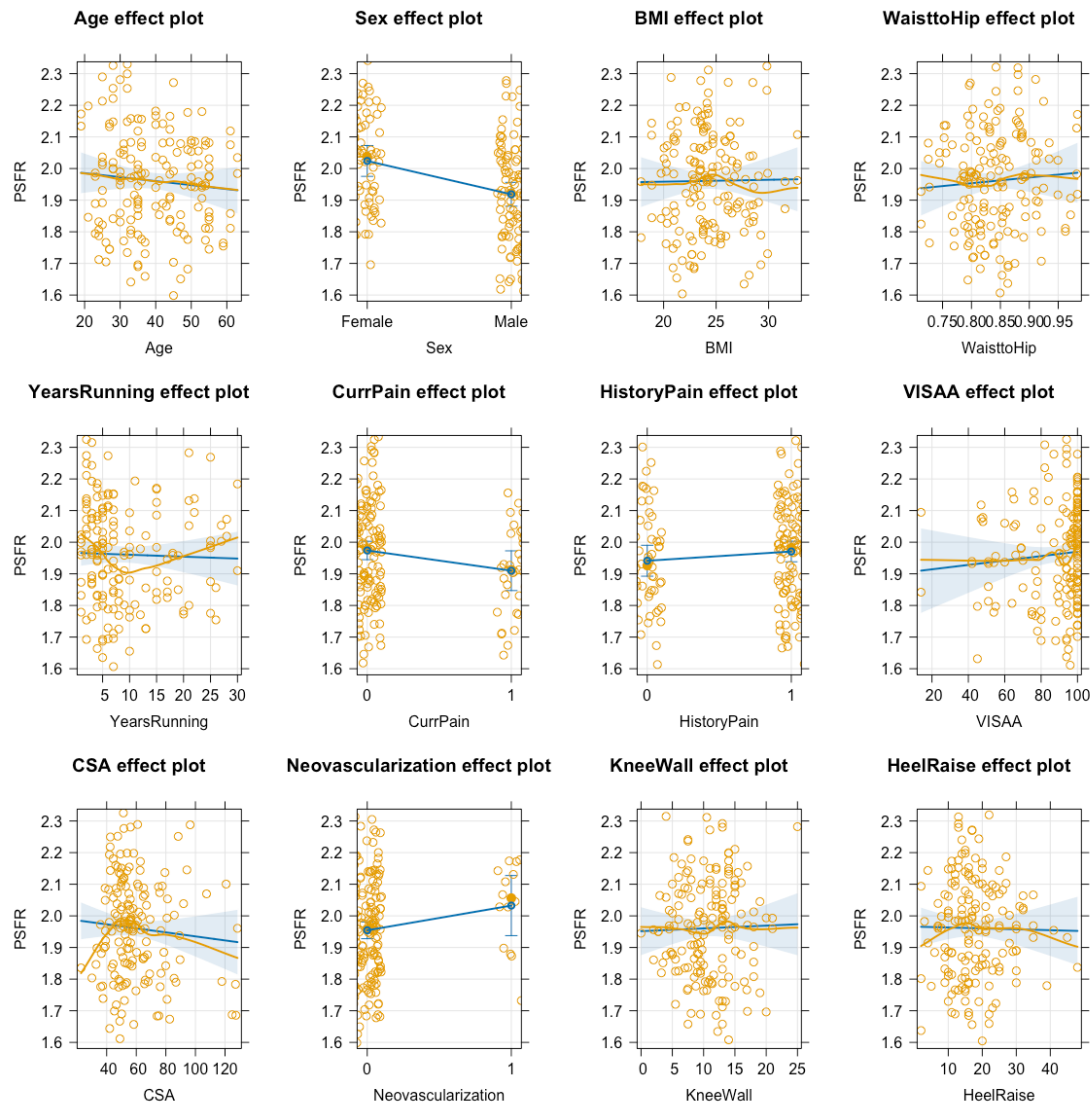
- For two otherwise similar subjects but that differ on the sex of the subjects, the estimated mean PSFR of male subjects is 0.01mm^{-1} lower than that of female subjects (95% CI from 0.05 to 0.16).
- 8) Now we want to start to explore their “multivariate” model. It contained Age, Sex, BMI, WaisttoHip, Years running, CurrPain, HistoryPain, VISA-A, CSA, Neovascularization, Knee-to-wall, and Heel raise as predictors (no interactions). Fit this model and make effects plots. Do these plots two ways, one with `plot(allEffects(modelname), grid = T)` and the other that adds partial residuals (we’ll discuss these more later) using `plot(allEffects(modelname, residuals = T), grid = T)`.

```
m1m <- lm(PSFR ~ Age + Sex + BMI + WaisttoHip + YearsRunning + CurrPain + HistoryPain + VISAA + CSA + Neovascularization + KneeWall + HeelRaise, data = tendonna2)
```

```
p1 <- plot(allEffects(m1m), grid = T)
```



```
p2 <- plot(allEffects(mlm, residuals = T), grid = T)
```



- 9) Then choose one slope coefficient from this model other than Sex and report an evidence sentence for it, filling in the needed parts of the provided evidence reporting template.

```
summary(mlm)

##
## Call:
## lm(formula = PSFR ~ Age + Sex + BMI + WaisttoHip + YearsRunning +
##     CurrPain + HistoryPain + VISAA + CSA + Neovascularization +
##     KneeWall + HeelRaise, data = tendonna2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35579 -0.12186  0.00352  0.11466  0.35956
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.8715017   0.2890251    6.475 1.26e-09
## Age           -0.0012581   0.0014721   -0.855  0.39410
## SexMale       -0.1054296   0.0353705   -2.981  0.00335
## BMI           0.0005986   0.0058474    0.102  0.91860
## WaisttoHip     0.1809880   0.3244423    0.558  0.57778
## YearsRunning  -0.0006114   0.0019574   -0.312  0.75522
## CurrPain1     -0.0639763   0.0364711   -1.754  0.08143
## HistoryPain1   0.0292259   0.0316341    0.924  0.35703
## VISAA         0.0006968   0.0009074    0.768  0.44374
## CSA          -0.0006358   0.0007238   -0.878  0.38113
## Neovascularization1 0.0781951   0.0507096    1.542  0.12516
## KneeWall      0.0008782   0.0033575    0.262  0.79401
## HeelRaise     -0.0002848   0.0017567   -0.162  0.87143
##
## Residual standard error: 0.163 on 151 degrees of freedom
## Multiple R-squared:  0.1448, Adjusted R-squared:  0.07689
## F-statistic: 2.131 on 12 and 151 DF,  p-value: 0.01797
```

- There is little to no evidence against the null hypothesis of no interaction between Years Running and PSFR ($t_{151} = -0.312$, two-sided p-value = 0.75522), controlling for Age, Sex, BMI, WaisttoHip, CurrPain, HistoryPain, VISA-A, CSA, Neovascularization, Knee-to-wall, and Heel raise, so we conclude that there is not a relationship between PSFR and Years Running and would drop the term from the model.**
- 10) One more check of your version of R. It should say 4.4.2 in your compiled (knitted) word document:
- R version (short form): 4.4.2
- 11) 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