STAT 542 Example Solutions – Homework 4

Chapter 6, problem 6) a) The R code for part (a) produces this table:

> HELPfullsmall

 ID TIME DRUGRISK SEXRISK

1 1 0 0 4

2 1 6 0 1

3 1 18 0 1

4 1 24 0 3

5 2 0 0 7

6 2 6 0 0

7 3 0 20 2

8 3 6 13 4

9 3 24 19 4

b) For subject 3, the measurements were available for timepoints 0, 6, and 24.

c) The R code for part (c) produces this table:

> HELPfullwide

# A tibble: 3 × 4

 DRUGRISK\_0 DRUGRISK\_6 SEXRISK\_0 SEXRISK\_6

 <int> <int> <int> <int>

1 0 0 4 1

2 0 0 7 0

3 20 13 2 4

d) The cor() function in R produces these two values for the correlations between the 0-month and 6-month DRUGRISK and SEXRISK measurements:

cor(HELPfullall$DRUGRISK\_0,HELPfullall$DRUGRISK\_6,use="complete.obs")

# [1] 0.5991146

cor(HELPfullall$SEXRISK\_0,HELPfullall$SEXRISK\_6,use="complete.obs")

# [1] 0.5076615

Chapter 6, Problem 7) The R code the book provides does indeed work:

> all

# A tibble: 3 × 4

 id group vals diff

 <dbl> <chr> <dbl> <dbl>

1 1 T 4 -1

2 2 T 6 0

3 3 T 8 -2

Problems with first approach: If one observation is missing, this code may not subtract the paired values correctly. If the data set is not sorted in the right order, the measurement will also be paired incorrectly and the differences may be incorrect.

This better approach using pivot\_wider() is something like:

> ds2 <- ds1 %>%

+ pivot\_wider(names\_from=group, values\_from=vals) %>%

+ mutate(diff = T-C)

> ds2

# A tibble: 3 × 4

 id T C diff

 <dbl> <dbl> <dbl> <dbl>

1 1 4 5 -1

2 2 6 6 0

3 3 8 10 -2

Chapter 6, Problem 10) The R code should use pivot\_longer() and then produce a plot something like this:



Chapter 7, Problem 1)

 age anysubstatus cesd d1 daysanysub dayslink drugrisk e2b female i1 i2

1 35.65342 0.7723577 32.84768 3.059603 75.30738 255.6056 1.887168 2.504673 0.2362031 17.90728 24.54746

 id indtot linkstatus mcs pcs pss\_fr sexrisk avg\_drinks max\_drinks hospitalizations

1 233.4018 35.72848 0.3781903 31.67668 48.04854 6.706402 4.642384 17.90728 24.54746 3.059603

Chapter 7, Problem 4)

> map\_int(bk\_teams, count\_seasons)

[1] 1 4 6 1 68 1 2

>

> # Nicer display of results:

>

> Brooklyn\_count <- map\_int(bk\_teams, count\_seasons)

> print(data.frame(bk\_teams, Brooklyn\_count))

 bk\_teams Brooklyn\_count

1 BR1 1

2 BR2 4

3 BR3 6

4 BR4 1

5 BRO 68

6 BRP 1

7 BRF 2

Chapter 6, Problem 4) EXTRA CREDIT) Answer will vary, but should include scraping data from a table on a Wikipedia page, producing a figure, and making comments that interpret that figure.