PHP 2550: Assignment 1

Blain Morin
Group: Jessica Kaminsky and Derrick Yam

September 26, 2018

Problem #1

a.) Summarize the average pain score for each of the 7 days of the study.

First, we constructed a dataset with only the first observation for each individual. We then calculated the mean for each of the seven periods.

Here are the average pain measurements for each of the seven periods:

Mean_p1	Mean_p2	Mean_p3	Mean_p4	Mean_p5	Mean_p6	Mean_p7
8.949749	7.988924	7.490385	7.317771	6.788462	7.066667	6.471429

Note:
Means for each period are calculated with NA values removed.

From a visual inspection, it appears that the average pain score is decreasing over time.

b.) Regress each pain score on age and use the summary function to create a summary table for each regression. Then use the confint function to find the 95% confidence interval for the regression slopes and produce a table with the estimates, standard errors, p-values and confidence intervals of the 7 slopes and put these in a single table.

We first centered age so that the intercept would have an interpretation. We then regressed each of the pain scores on the centered age using a simple linear model. Tables 1-3 contain the estimates, standard errors, p-values and confidence intervals for each of the seven regressions:

Table 1:

	Dependent variable:		
	pain.1	pain.2	pain.3
	(1)	(2)	(3)
scale(age, scale = FALSE)	−0.024	−0.029	−0.035
	(−0.075, 0.026)	(−0.088, 0.030)	(−0.101, 0.032)
	p = 0.347	p = 0.340	p = 0.308
Constant	8.947	8.003	7.481
	(8.474, 9.421)	(7.441, 8.566)	(6.883, 8.078)
	p = 0.000***	p = 0.000***	p = 0.000***
Observations	199	158	156
R^2	0.005	0.006	0.007
Adjusted R^2	−0.001	−0.001	0.0003
Residual Std. Error	3.410 (df = 197)	3.603 (df = 156)	3.805 (df = 154)
F Statistic	0.891 (df = 1; 197)	0.919 (df = 1; 156)	1.047 (df = 1; 154)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2:

	Dependent variable:		
	pain.4	pain.5	pain.6
	(1)	(2)	(3)
scale(age, scale = FALSE)	−0.028	−0.058	−0.037
	(−0.090, 0.034)	(−0.121, 0.005)	(−0.100, 0.026)
	p = 0.379	p = 0.076*	p = 0.252
Constant	7.325	6.752	7.073
	(6.747, 7.903)	(6.174, 7.330)	(6.461, 7.685)
	p = 0.000***	p = 0.000***	p = 0.000***
Observations	166	156	150
R^2	0.005	0.020	0.009
Adjusted R^2	−0.001	0.014	0.002
Residual Std. Error	3.797 (df = 164)	3.674 (df = 154)	3.824 (df = 148)
F Statistic	0.779 (df = 1; 164)	3.203* (df = 1; 154)	1.327 (df = 1; 148)
Note:			*p<0.1; **p<0.05; ***p<0.01

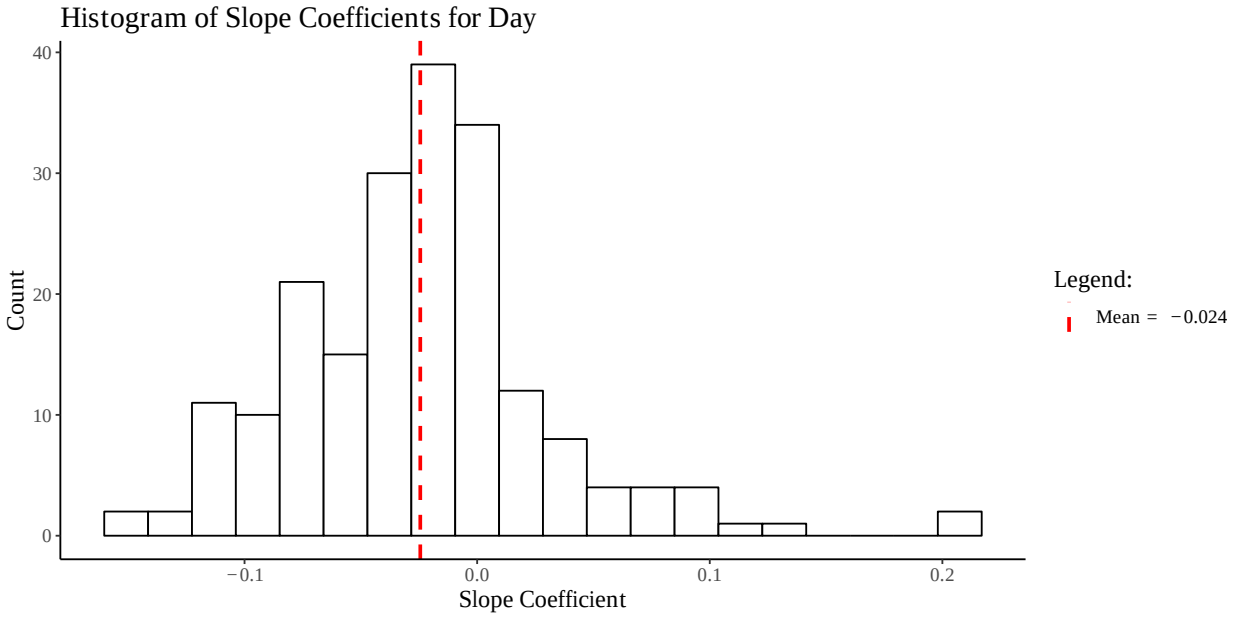
Table 3:

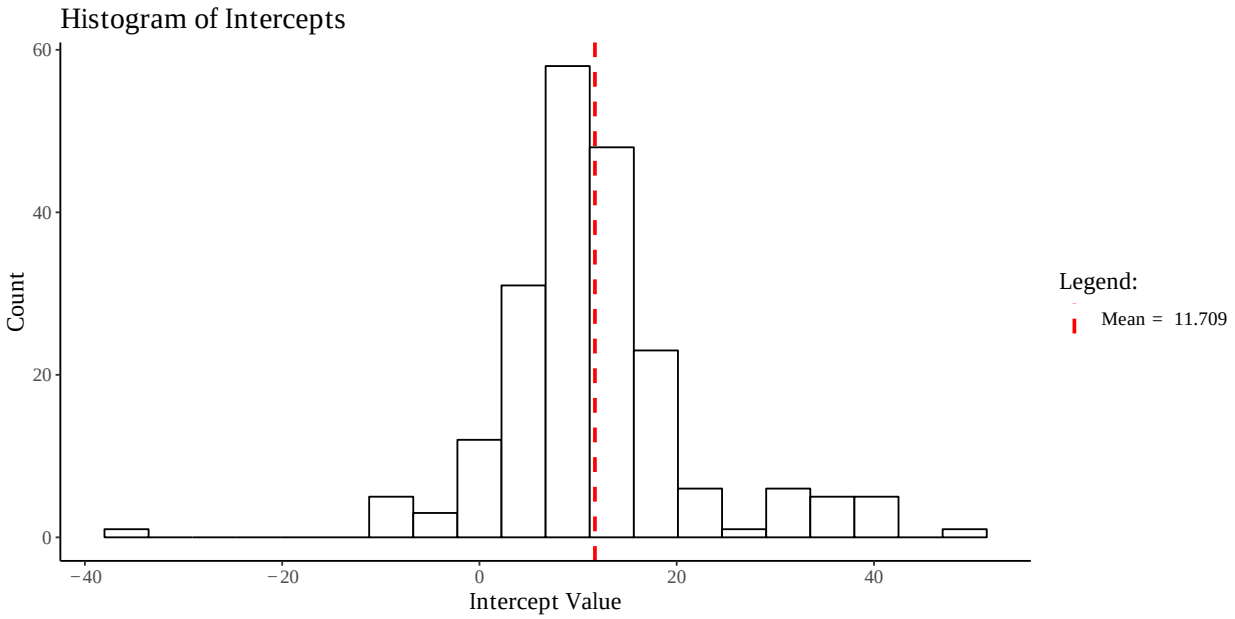
	Dependent variable:
	pain.7
scale(age, scale = FALSE)	−0.067
	(−0.139, 0.005)
	p = 0.072*
Constant	6.518
	(5.852, 7.184)
	p = 0.000***
Observations	140
R^2	0.023
Adjusted R^2	0.016
Residual Std. Error	4.011 (df = 138)
F Statistic	3.310* (df = 1; 138)
Note:	*p<0.1; **p<0.05; ***p<0.01

The constant is the average pain score for the mean age. Notice that this matches the values from part (a). The coefficient for age can be interpreted as the average decrease in pain for each additional year. The age coefficient is only significant at the 10% level for time period 5 and 7.

c.) For each individual fit a regression of pain on time. Summarize the slopes and intercepts produced.

First we put the data in long format. Then, we grouped it by individuals and ran individual regressions (pain regressed on day) for each person. The intercept is interpreted as the pain at baseline. The day coefficient is the average increase or decrease in pain per day. Here are the histograms for the intercepts and day coefficients:





We see that the mean slope coefficient for day is -.024 and that most of the slopes fall in the -.1 to .1 range. Since zero is in this range, this may suggest that there is no time trend in pain in our data. The mean intercept (baseline pain) is 11.71. Most of these intercepts are in the range of 0 to 20. The negative intercepts likely come from extrapolations from people with missing early observations.

d.) Are the slopes or intercepts related to any of the patient characteristics (age, race, income, treatment, sex, occupation, working status, use of NSAIDs)?

We used linear regression to check if any of the patient characteristics were related to the slope and the intercept. For these regressions, we were careful to code the categorical variables as factors. Table 4 contains our regression results:

Table 4:

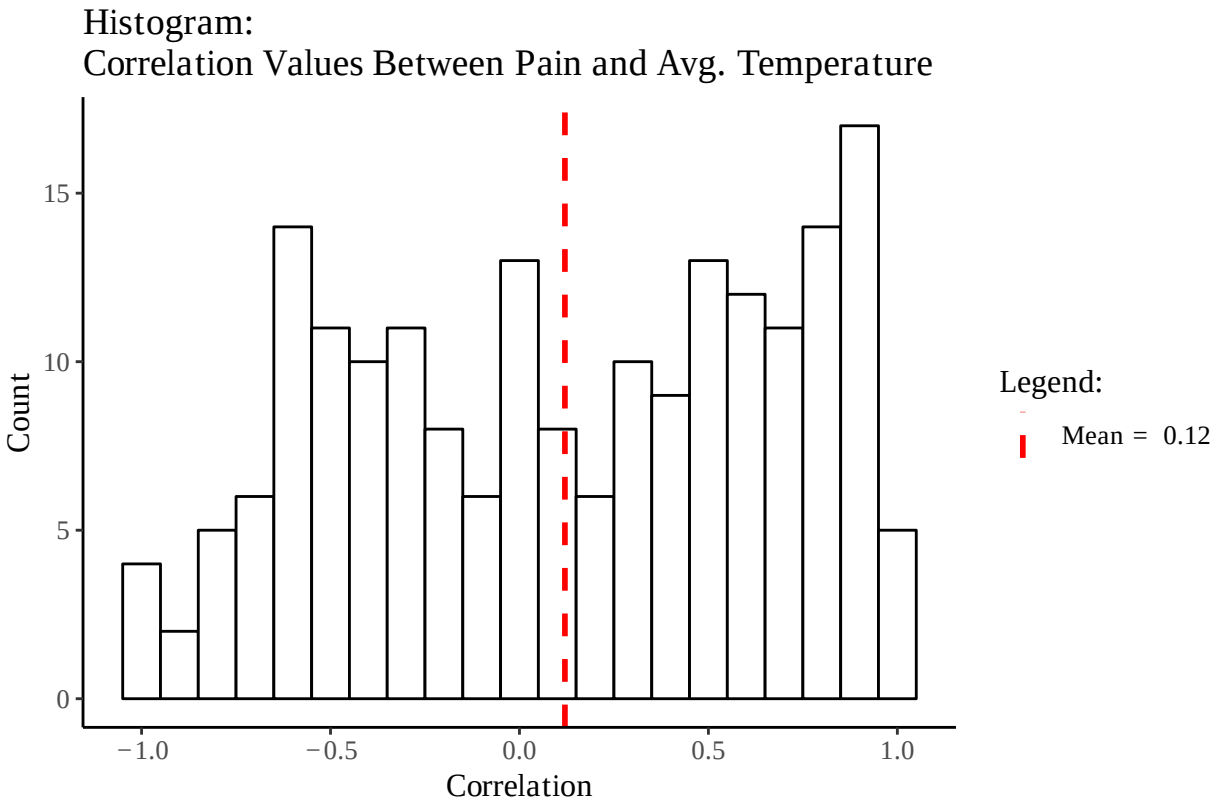
	Dependent variable:	
	slope	ints
	(1)	(2)
as.factor(Race)Native American	0.081	−4.454
as.factor(Race)Other	−0.022	−13.123
as.factor(Race)White	0.024	0.374
as.factor(Race)White Hispanic	0.033	−3.592
as.factor(Race)White Hispanic, Asian	0.085	−12.705
as.factor(Race)White Non-Hispanic	0.050	−6.911
age	0.002	−0.428**
as.factor(inccat)2	0.043**	−6.052
as.factor(inccat)3	0.025	−3.521
as.factor(inccat)4	0.023	−0.686
as.factor(inccat)5	0.032	−3.725
bmi	0.001	0.083
treat	0.002	−1.862
as.factor(sex)2	0.002	−1.397
nsaid	−0.026	−2.022
as.factor(retire)2	−0.023	4.453
opiate	0.042	−7.723
as.factor(severe2)2	0.001	−1.087
Constant	−0.206**	49.731**
Observations	90	91
R^2	0.150	0.165
Adjusted R^2	−0.065	−0.044
Residual Std. Error	0.056 (df = 71)	11.315 (df = 72)
F Statistic	0.697 (df = 18; 71)	0.789 (df = 18; 72)
Note:		*p<0.1; **p<0.05; ***p<0.01

The only variable that came back significant at the .05 level for the slope regression was income level two. Income level 2 represents incomes in the range of 15 to 35 thousand. The reference category is level 1. Thus, the interpretation for this coefficient is that the average increase in pain over time was .043 higher for those in income category 2 compared to those in category 1, all else equal.

The only variable that had a significant effect at the .05 level on the intercept was age. The direction of the effect is the opposite of what we would expect. However, the effect size is relatively small compared to the range of baseline pains. We can interpret this coefficient as follows: For each additional year of age, the average baseline pain measurement was .428 less, all else equal.

e.) For each individual, compute the correlation between pain scores and average temperature on the dates the pain scores were taken and constuct a graph to display these correlations. Discuss whether pain is correlated with temperature.

To answer this question, we put the data in long format and joined it with the average temperature for each individual on the day that the pain score was taken. Here is a histogram of the calculated correlations:



Note: individuals with only one observation are excluded (cannot calculate correlation)

We see that the mean correlation is .12, which suggests that there is no linear relationship between average temperature and pain. The values are pretty widely spread between -1 and 1. The values with perfect positive (1) and negative (-1) correlation represents the people with only two measurements.

Problem 2

In the paper by Wang et al. (paper #10 in your syllabus), reproduce Table 2 using just the outcome data without trying to fit the mixed model. In other words, the differences at each time in the two treatment groups should be calculated just as a difference in means with a corresponding confidence interval. It is easist if you fit a simple linear model to produce the effects at each time for each difference compared to baseline. You can then also find the difference of the differences to compare treated and control.

To calculate the values in the table, we first put the data in long format. For each of the outcomes, we ran three linear models. We used time as a categorical variable. To get column 1, we regress the outcome on time, for those in the Tai Chi group. To get column 2, we regressed the outcome on time, for those in the attention control group. To get the differences, we used an interaction term: we regressed the outcome on time + treatment + time * treatment.

Table 5: Changes in primary and secondary outcomes*

Variable	Improvement from baseline		Between-group differences, Tai Chi vs. attention control	P†
	Tai Chi (n = 20)	Attention Control (n = 20)		
Primary outcome: WOMAC pain: (range 0-500 mm)‡				
Week 12	-157.25 (-202.89,-111.61)	-38.45 (-104.97,28.07)	-118.8 (-198.83,-38.77)	0.00
Week 24	-131.55 (-177.19,-85.91)	-64.6 (-131.12,1.92)	-66.95 (-146.98,13.08)	0.10
Week 48	-115.35 (-160.99,-69.71)	-69.2 (-135.72,-2.68)	-46.15 (-126.18,33.88)	0.26
Secondary outcomes				
WOMAC physical function (range 0-1,700 mm)‡				
Week 12	-506.75 (-674.63,-338.87)	-182.15 (-399.96,35.66)	-324.6 (-597.4,-51.8)	0.02
Week 24	-440.5 (-608.38,-272.62)	-257.3 (-475.11,-39.49)	-183.2 (-456,89.6)	0.19
Week 48	-405.85 (-573.73,-237.97)	-300.55 (-518.36,-82.74)	-105.3 (-378.1,167.5)	0.45
WOMAC stiffness (range 0-200 mm)‡				
Week 12	-73.05 (-98.75,-47.35)	-50.15 (-83.25,-17.05)	-22.9 (-64.47,18.67)	0.28
Week 24	-65 (-90.7,-39.3)	-50.2 (-83.3,-17.1)	-14.8 (-56.37,26.77)	0.48
Week 48	-64.15 (-89.85,-38.45)	-60.5 (-93.6,-27.4)	-3.65 (-45.22,37.92)	0.86
Physician VAS (range 0-10 cm)‡				
Week 12	-3.15 (-4.14,-2.15)	-1.44 (-2.9,0.02)	-1.71 (-3.46,0.04)	0.06
Week 24	-2.6 (-3.59,-1.6)	-2.06 (-3.52,-0.6)	-0.53 (-2.29,1.22)	0.55
Week 48	-2.54 (-3.53,-1.54)	-1.55 (-3.03,-0.08)§	-0.98 (-2.74,0.78)	0.27
Patient global VAS (range 0-10 cm)‡				
Week 12	-2.98 (-4.14,-1.82)	-0.83 (-2.29,0.64)	-2.15 (-4,-0.3)	0.02
Week 24	-2.36 (-3.51,-1.2)	-1.71 (-3.17,-0.25)	-0.65 (-2.5,1.2)	0.49
Week 48	-1.66 (-2.81,-0.5)	-1.7 (-3.16,-0.23)	0.04 (-1.81,1.89)	0.97
6-minute walk test, yards¶				
Week 12	36.5 (-27.34,100.33)**	-1.42 (-57.3,54.47)	37.91 (-45.96,121.78)	0.37
Week 24	44.28 (-18.69,107.24)§	7.84 (-48.05,63.72)	36.44 (-46.84,119.72)	0.39
Week 48	28.75 (-34.22,91.71)§	20.69 (-35.93,77.3)§	8.06 (-75.75,91.87)	0.85
Balance score (range 0-5)¶				
Week 12	0.15 (-0.28,0.58)	0.25 (-0.25,0.75)	-0.1 (-0.76,0.56)	0.76
Week 24	0.15 (-0.28,0.58)	0.08 (-0.43,0.58)	0.07 (-0.58,0.73)	0.82
Week 48	0.35 (-0.08,0.78)	0.52 (0,1.03)§	-0.17 (-0.83,0.5)	0.62
Chair stand time, seconds‡				
Week 12	-12.02 (-19.41,-4.62)§	-0.84 (-5.68,4)§	-11.17 (-19.93,-2.42)	0.01
Week 24	-9.86 (-17.26,-2.46)§	-4.45 (-9.29,0.4)§	-5.41 (-14.17,3.34)	0.22
Week 48	-9.21 (-16.61,-1.81)§	-3.21 (-8.11,1.7)**	-6 (-14.81,2.81)	0.18
SF-36 MCS (range 0-100)¶				
Week 12	2.14 (-3.33,7.62)	1.93 (-5.4,9.27)	0.21 (-8.87,9.29)	0.96
Week 24	4.39 (-1.08,9.86)	4.5 (-2.83,11.83)	-0.11 (-9.19,8.97)	0.98
Week 48	5.8 (0.33,11.28)	1.04 (-6.3,8.37)	4.77 (-4.31,13.84)	0.30
SF-36 PCS (range 0-100)¶				
Week 12	11.57 (6.38,16.76)	4.14 (-1.74,10.02)	7.43 (-0.35,15.21)	0.06
Week 24	10.8 (5.61,15.99)	6.29 (0.4,12.17)	4.51 (-3.27,12.29)	0.25
Week 48	10.41 (5.22,15.6)	4.1 (-1.79,9.98)	6.32 (-1.46,14.1)	0.11
CES-D (range 0-60)‡				
Week 12	-7.4 (-12.92,-1.88)	-0.7 (-6.83,5.43)	-6.7 (-14.89,1.49)	0.11
Week 24	-6.4 (-11.92,-0.88)	-1.1 (-7.23,5.03)	-5.3 (-13.49,2.89)	0.20

Week 48	-7.25 (-12.77,-1.73)	1.65 (-4.48,7.78)	-8.9 (-17.09,-0.71)	0.03
Self-efficacy score (range 1-5)[¶]				
Week 12	0.6 (-0.07,1.27)	-0.11 (-0.66,0.44)	0.71 (-0.15,1.57)	0.10
Week 24	0.68 (0.01,1.35)	-0.17 (-0.72,0.38)	0.85 (-0.01,1.71)	0.05
Week 48	0.72 (0.05,1.39)	-0.24 (-0.79,0.31)	0.96 (0.1,1.82)	0.03

[*] Values are the mean (95 Confidence interval). WOMAC Western Ontario and McMaster Universities Osteoarthritis Index; VAS visual analog scale; SF-36 Short Form 36 health survey; MCS mental component summary; PCS physical component summary; CES-D Center for Epidemiologic Studies Depression Scale.
[†] P values were calculated by t-test for continuous variables.
[‡] Lower scores indicate improved state.
[§] N = 19
[¶] Higher scores indicate improved state.
[**] N = 18.

We see in the table that the many of the significant differences between the groups occurred at the first follow-up at week 12. The significant outcomes at the .05 level at week 12 were WOMAC pain, WOMAC physical function, patient global VAS range and chair stand time. The significant outcome at week 24 and week 48 was only self efficacy-score.

Appendix

Code for 1a:

```
### Load packages
library(knitr)
library(dplyr)
library(readr)
library(sjPlot)
library(kableExtra)
library(tidyr)
library(ggplot2)
library(gmodels)
library(grid)
library(stargazer)
library(gridExtra)
library(extrafont)
library(extrafontdb)

### Read in data
mcalindon = read_csv("McAlindon_Big.csv")


### Get only the first observation for each id
first.ob = mcalindon %>%
  group_by(ID) %>%
  slice(1) %>%
  ungroup()
```

```
### Take the mean of each pain column and put it in the table "ave.pains"
ave.pains = first.ob %>%
  summarise(Mean_p1 = mean(pain.1, na.rm = TRUE),
            Mean_p2 = mean(pain.2, na.rm = TRUE),
            Mean_p3 = mean(pain.3, na.rm = TRUE),
            Mean_p4 = mean(pain.4, na.rm = TRUE),
            Mean_p5 = mean(pain.5, na.rm = TRUE),
            Mean_p6 = mean(pain.6, na.rm = TRUE),
            Mean_p7 = mean(pain.7, na.rm = TRUE))

kable(ave.pains) %>%
  kable_styling() %>%
  footnote(general = "Means for each period are calculated with NA values removed.")
```

Code for 1b:

```
### Regress each pain score on age
painreg1 = lm(pain.1 ~ scale(age, scale = FALSE), data = first.ob)
painreg2 = lm(pain.2 ~ scale(age, scale = FALSE), data = first.ob)
painreg3 = lm(pain.3 ~ scale(age, scale = FALSE), data = first.ob)
painreg4 = lm(pain.4 ~ scale(age, scale = FALSE), data = first.ob)
painreg5 = lm(pain.5 ~ scale(age, scale = FALSE), data = first.ob)
painreg6 = lm(pain.6 ~ scale(age, scale = FALSE), data = first.ob)
painreg7 = lm(pain.7 ~ scale(age, scale = FALSE), data = first.ob)


### Create summary tables using stargazer
stargazer(painreg1,
       painreg2,
       painreg3,
       type = "latex",
       ci = TRUE,
       header = FALSE,
       report = "vcsp*",
       table.placement = "H"
       )

stargazer(painreg4,
       painreg5,
       painreg6,
       ci = TRUE,
       header = FALSE,
       report = "vcsp*",
       table.placement = "H")

stargazer(painreg7,
       ci = TRUE,
       header = FALSE,
       report = "vcsp*",
       table.placement = "H")
```

Code for 1c:

```
### These functions put data in long format, then run regressions and pull out either the int
### or the slope
```

```
get.ints = function (p1, p2, p3, p4, p5, p6, p7, d1, d2, d3, d4, d5, d6, d7) {

  data.reg = data.frame(pain = c(p1,p2,p3,p4,p5,p6,p7),
                        day = c(d1,d2,d3,d4,d5,d6,d7))

  reg = lm(pain ~ day, data = data.reg)
  return(coef(reg)[1])

}

get.slopes = function (p1, p2, p3, p4, p5, p6, p7, d1, d2, d3, d4, d5, d6, d7) {

  data.reg = data.frame(pain = c(p1,p2,p3,p4,p5,p6,p7),
                        day = c(d1,d2,d3,d4,d5,d6,d7))

  reg = lm(pain ~ day, data = data.reg)
  return(coef(reg)[2])

}

indy.regs = first.ob %>% group_by(ID) %>%
  mutate(slope = get.slopes(p1 = pain.1,
                            p2 = pain.2,
                            p3 = pain.3,
                            p4 = pain.4,
                            p5 = pain.5,
                            p6 = pain.6,
                            p7 = pain.7,
                            d1 = lastdt1,
                            d2 = lastdt2,
                            d3 = lastdt3,
                            d4 = lastdt4,
                            d5 = lastdt5,
                            d6 = lastdt6,
                            d7 = lastdt7),
         ints = get.ints(p1 = pain.1,
                            p2 = pain.2,
                            p3 = pain.3,
                            p4 = pain.4,
                            p5 = pain.5,
                            p6 = pain.6,
                            p7 = pain.7,
                            d1 = lastdt1,
                            d2 = lastdt2,
                            d3 = lastdt3,
                            d4 = lastdt4,
                            d5 = lastdt5,
                            d6 = lastdt6,
                            d7 = lastdt7))


### ggplot histograms for the slopes and coefficients
```

```
hist.slopes = indy.regs %>%
  ggplot(aes(x = slope)) +
  geom_histogram(bins = 20, color = "black", fill = "white") +
  geom_vline(aes(xintercept=mean(indy.regs$slope, na.rm = TRUE),color="mean"),
             linetype="dashed",
             size=1) +
  xlab("Slope Coefficient") +
  ylab("Count") +
  ggtitle("Histogram of Slope Coefficients for Day") +
  theme_classic() +
  scale_color_manual( labels = c(paste("Mean = ", round(mean(indy.regs$slope, na.rm = TRUE), digits = 3)
                      name = "Legend: ", values = c(mean = "red")) +
  theme(legend.position = "right") +
  theme(text=element_text(size=14,  family="CM Sans"))

hist.ints = indy.regs %>%
  ggplot(aes(x = ints)) +
  geom_histogram(bins = 20, color = "black", fill = "white") +
  geom_vline(aes(xintercept=mean(indy.regs$ints, na.rm = TRUE), color="mean"),
             linetype="dashed",
             size=1) +
  xlab("Intercept Value") +
  ylab("Count") +
  ggtitle("Histogram of Intercepts") +
  theme_classic() +
  scale_color_manual( labels = c(paste("Mean = ", round(mean(indy.regs$ints, na.rm = TRUE), digits = 3))
                      name = "Legend: ", values = c(mean = "red")) +
  theme(legend.position = "right") +
  theme(text=element_text(size=14,  family="CM Sans"))


### Arrange the ggplots side by side
grid.arrange(hist.slopes, hist.ints, nrow = 2)
```

Code for 1d:

```
### Regress slopes and intercepts on patient characteristics
slope.reg = lm(slope ~ as.factor(Race) + age + as.factor(inccat) + bmi + treat + as.factor(sex) + nsaid
               as.factor(retire) + opiate + as.factor(severe2), data = indy.regs)

int.reg = lm(ints ~ as.factor(Race) + age + as.factor(inccat) + bmi + treat + as.factor(sex) + nsaid +
                 as.factor(retire) + opiate + as.factor(severe2), data = indy.regs)

### Use stargazer to make table
stargazer(slope.reg, int.reg,
          font.size = "normalsize",
          single.row = TRUE,
          report = "vc*",
          header = FALSE,
          table.placement = "H")
```

Code for 1e:

```
### Split up the data so gather function doesnt get confused
weather = mcalindon %>%
```

```
  select(ID, WeatherDate, avgtemp)

pains = mcalindon %>%
  select(ID, pain.1, pain.2, pain.3, pain.4, pain.4, pain.5, pain.6, pain.7) %>%
  group_by(ID) %>%
  slice(1)


### Add an ID columns for joining
pains = pains %>%
  gather(key = pain.time, value = pain.score, pain.1:pain.7 ) %>%
  group_by(ID) %>%
  mutate( index = row_number(ID))

days = mcalindon %>%
  select(ID, lastdt1, lastdt2, lastdt3, lastdt4, lastdt5, lastdt6, lastdt7) %>%
  group_by(ID) %>%
  slice(1)

days = days %>%
  gather(key = time.name, value = day, lastdt1:lastdt7) %>%
  mutate(index = row_number(ID))


### Join the frams together
pain.w.days = pains %>%
  inner_join(days)

pain.day.temp = pain.w.days %>%
  rename(WeatherDate = day) %>%
  inner_join(weather)


### Calculate correlation
pain.day.temp = pain.day.temp %>%
  group_by(ID) %>%
  mutate(correlation = cor(pain.score, avgtemp, use = "complete.obs"))


only.cor = pain.day.temp %>%
  select(ID, correlation) %>%
  group_by(ID) %>%
  slice(1)


### ggplot histogram

hist.cor = only.cor %>%
  ggplot(aes(x = correlation)) +
  geom_histogram(binwidth = .1, color = "black", fill = "white") +
  geom_vline(aes(xintercept=mean(only.cor$correlation, na.rm = TRUE), color="mean"),
             linetype="dashed",
             size=1) +
```

```
  xlab("Correlation") +
  ylab("Count") +
  ggtitle("Histogram:\nCorrelation Values Between Pain and Avg. Temperature") +
  theme_classic() +
  scale_color_manual( labels = c(paste("Mean = ", round(mean(only.cor$correlation, na.rm = TRUE), digits
                      name = "Legend: ", values = c(mean = "red")) +
  theme(legend.position = "right") +
  labs(caption = "Note: individuals with only one observation are excluded (cannot calculate correlation
  theme(text=element_text(size=12,  family="CM Sans"))


hist.cor
```

Code for 2:

```
### Vector of outcome names for looping over

wang <- read.csv("Wang.csv")

outcomes <- c("womac.pain.", "womac.phys.func.", "womac.stiff.", "physician.vas.", "pt.global.vas.", "wa

variables <- c("womac.pain.1", "womac.pain.2", "womac.pain.3", "womac.pain.4")

tai_chi <- c()
attention <- c()
between <- c()

### Get means and CIs for all outcomes over time

for(i in outcomes) {

  variables <- c(paste(i, 1, sep = ""), paste(i, 2, sep = ""), paste(i, 3, sep = ""), paste(i, 4, sep =

  long2 <- gather(wang, key = "time", value = "score", variables)

  long2$time = as.factor(long2$time)

  treated_model <- lm(score ~ time, data = long2[long2$group == 1, ])
  untreated_model <- lm(score ~ time, data = long2[long2$group == 0, ])
  difference_model <- lm(score ~ time + group + time:group, data = long2)

  # estimate + CI for treated group

  tai_chi <- rbind(tai_chi, cbind(treated_model$coefficients[-1], confint(treated_model)[-1,]))

  # estimate + CI for untreated group

  attention <- rbind(attention, cbind(untreated_model$coefficients[-1], confint(untreated_model)[-1,]))

  # estimate, CI, p-value for difference of differences

  between <- rbind(between, cbind(difference_model$coefficients[6:8], confint(difference_model)[6:8,], a

}
```

```
### Fancy kable styling
CompleteTable <- cbind(tai_chi, attention, between)
CompleteTable <- as_tibble(CompleteTable)
names(CompleteTable) <- c("TMean", "TLower", "TUpper", "CMean", "CLower", "CUpper", "DMean", "DLower",
options(knitr.table.format = "latex")

FormattedTable <- CompleteTable %>%
  mutate_at(vars(1:ncol(CompleteTable)), funs(round(., 2)))

FormattedTable <- FormattedTable %>%
  mutate(CCI = paste0(CMean, " ", "(", CLower, ",", CUpper, ")"),
         TCI = paste0(TMean, " ", "(", TLower, ",", TUpper, ")"),
         DCI = paste0(DMean, " ", "(", DLower, ",", DUpper, ")"))

FormattedTable <- FormattedTable %>% mutate(Variable = rep(c("Week 12", "Week 24" , "Week 48"), 12))
FormattedTable <- FormattedTable %>% select(Variable, TCI, CCI, DCI, Pvalue)

##Add all of the footnote symbols to the variables.
FormattedTable$TCI[16] <- paste0(FormattedTable$TCI[16], footnote_marker_symbol(6, double_escape = F))#
#FormattedTable$TCI[c(17,18, 22, 23, 24)] <- #§
FormattedTable$TCI[17] <- paste0(FormattedTable$TCI[17], footnote_marker_symbol(4, double_escape = F))
FormattedTable$TCI[18] <- paste0(FormattedTable$TCI[18], footnote_marker_symbol(4, double_escape = F))
FormattedTable$TCI[22] <- paste0(FormattedTable$TCI[22], footnote_marker_symbol(4, double_escape = F))
FormattedTable$TCI[23] <- paste0(FormattedTable$TCI[23], footnote_marker_symbol(4, double_escape = F))
FormattedTable$TCI[24] <- paste0(FormattedTable$TCI[24], footnote_marker_symbol(4, double_escape = F))

FormattedTable$CCI[24] <- paste0(FormattedTable$CCI[24], footnote_marker_symbol(6, double_escape = F))#
#FormattedTable$CCI[c(12, 18, 21, 22, 23)] <- #§
FormattedTable$CCI[12] <- paste0(FormattedTable$CCI[12], footnote_marker_symbol(4, double_escape = F))
FormattedTable$CCI[18] <- paste0(FormattedTable$CCI[18], footnote_marker_symbol(4, double_escape = F))
FormattedTable$CCI[21] <- paste0(FormattedTable$CCI[21], footnote_marker_symbol(4, double_escape = F))
FormattedTable$CCI[22] <- paste0(FormattedTable$CCI[22], footnote_marker_symbol(4, double_escape = F))
FormattedTable$CCI[23] <- paste0(FormattedTable$CCI[23], footnote_marker_symbol(4, double_escape = F))

##Group Rows Length and Names Vector
grouprowslength <- c(3,3,3,3,3,3,3,3,3,3,3,3,3)
grouprowsnames <- linebreak(c(paste0("Primary outcome: WOMAC pain: (range 0-500 mm)", footnote_marker_sy
                      paste0("Secondary outcomes \n WOMAC physical function (range 0-1,700 mm)", footn
                      paste0("WOMAC stiffness (range 0-200 mm)", footnote_marker_symbol(3, double_escap
                      paste0("Physician VAS (range 0-10 cm)", footnote_marker_symbol(3, double_escape =
                      paste0("Patient global VAS (range 0-10 cm)", footnote_marker_symbol(3, double_es
                      paste0("6-minute walk test, yards", footnote_marker_symbol(5, double_escape = T)
                      paste0("Balance score (range 0-5)", footnote_marker_symbol(5, double_escape = T)
                      paste0("Chair stand time, seconds", footnote_marker_symbol(3, double_escape = T)
                      #paste0("Body mass index, kg/m2"),
                      paste0("SF-36 MCS (range 0-100)", footnote_marker_symbol(5, double_escape = T)),
                      paste0("SF-36 PCS (range 0-100)", footnote_marker_symbol(5, double_escape = T)),
                      paste0("CES-D (range 0-60)", footnote_marker_symbol(3, double_escape = T)),
                      paste0("Self-efficacy score (range 1-5)", footnote_marker_symbol(5, double_escape
names(grouprowslength) <- grouprowsnames

firstheader <- c(1,1,1,1,1)
```

```
names(firstheader) <- linebreak(c("Variable", "Tai Chi\n(n = 20)", "Attention Control\n(n = 20)",
                     "Between-group\ndifferences, Tai Chi\nvs. attention control", paste0("P", footnote_
kabcomplete <- kable(FormattedTable, "latex", digits = 5, col.names = NULL,
      booktabs = T, longtable = T, linesep = "", escape = F,
      caption = paste0("Changes in primary and secondary outcomes",  footnote_marker_symbol(1))) %>%
  kable_styling(bootstrap_options = "condensed", full_width = F, latex_options = "hold_position") %>%
  add_header_above(firstheader, escape = F) %>%
  add_header_above(c(" " = 1, "Improvement from baseline" = 2,
                     " " = 2)) %>%
  footnote(symbol = c("Values are the mean (95 Confidence interval). WOMAC  Western Ontario and McMaster
                      scale; SF-36  Short Form 36 health survey; MCS  mental component summary; PCS  phy
                      Studies Depression Scale.", "P values were calculated by t-test for continuous var
                      "N = 19", "Higher scores indicate improved state.", "N = 18."), escape = F,
           threeparttable = T)


group_rows(kabcomplete, index = grouprowslength, escape = F)
```