Assignment 8

Matt Lawrence

12/2/21

```
library(tidyverse)
library(huxtable)
midd_survey <- read.csv("https://raw.githubusercontent.com/mjclawrence/soci385_f21/main/data/m
```

1. Create a new binary variable identifying respondents who think they are most likely to find a marriage or long-term partner at Middlebury with a 1 and everyone else with a 0. What is the mean of this new variable, and what does that value represent?

```
midd_survey <- mutate(midd_survey,
                      find_partner_midd =
                           ifelse(find_partner=="Middlebury",1,0))
```

```
mean(midd_survey$find_partner_midd)
```

```
## [1] 0.05076142
```

```
table(midd_survey$find_partner, midd_survey$gender)
```

```
##
##                   Man Other Woman
##   Family            2     1    15
##   Friends         201     6   349
##   Graduate School  26     1    36
##   Job/Career       87     0   117
##   Middlebury       28     0    22
##   None             19     4    25
##   Other            23     1    22
```

The mean of this binary variable is the proportion of respondents who think they are most likely to find a marriage or long-term partner at Middlebury.

How does this vary by gender?

```
prop.table(table(midd_survey$gender, midd_survey$find_partner_midd),1)
```

```
##
##                   0          1
##   Man    0.92746114 0.07253886
##   Other  1.00000000 0.00000000
##   Woman  0.96245734 0.03754266
```

No respondents in the "Other" gender category think they will find their partner at Middlebury. Let's take them out of the dataset so the rest of the models are easier to interpret.

```
midd_survey <- midd_survey |>
     filter(gender!="Other") |>
     droplevels()
```

2. Regress the binary variable you created in #1 on gender, and interpret the coefficients.

```
model1 <-
      lm(find_partner_midd ~ gender,
      data = midd_survey)

summary(model1)
```

```
##
## Call:
## lm(formula = find_partner_midd ~ gender, data = midd_survey)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07254 -0.07254 -0.03754 -0.03754  0.96246
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.07254    0.01122   6.465 1.61e-10 ***
## genderWoman -0.03500    0.01445  -2.422   0.0156 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2205 on 970 degrees of freedom
## Multiple R-squared:  0.006009,   Adjusted R-squared:  0.004985
## F-statistic: 5.864 on 1 and 970 DF,  p-value: 0.01563
```

On average, women are 3.5 percentage points less likely than men to think they will find a partner at Middlebury. This difference is significant.

3. Add type of housing as a control variable to the model, and interpret the coefficients.

```
model2 <-
      lm(find_partner_midd ~ gender + housing,
      data = midd_survey)

summary(model2)
```

```
##
## Call:
## lm(formula = find_partner_midd ~ gender + housing, data = midd_survey)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.07733 -0.07219 -0.03719 -0.03719  0.96281
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.072194   0.011711   6.165 1.03e-09 ***
## genderWoman        -0.035008   0.014488  -2.416   0.0159 *
## housingHouse        0.005140   0.019964   0.257   0.7969
## housingOff Campus -0.005831   0.027595  -0.211   0.8327
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2207 on 968 degrees of freedom
## Multiple R-squared:  0.006138,   Adjusted R-squared:  0.003058
## F-statistic: 1.993 on 3 and 968 DF,  p-value: 0.1134
```

Controlling for housing type, there is still a significant difference of 3.5 percentage points between men and women on average. Or, controlling for gender, students living in dorms, in houses and off campus are equally likely to expect to find a partner at Midd, on average.

4. Add an interaction between gender and type of housing to the model, and interpret the coefficients.

```
model3 <-
      lm(find_partner_midd ~ gender * housing,
      data = midd_survey)

summary(model3)
```

```
##
## Call:
## lm(formula = find_partner_midd ~ gender * housing, data = midd_survey)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.13462 -0.05788 -0.04719 -0.04719  0.98936
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     0.05788    0.01246   4.646 3.85e-06 ***
## genderWoman                    -0.01069    0.01624  -0.658  0.51059
## housingHouse                    0.07674    0.03291   2.331  0.01993 *
## housingOff Campus               0.07256    0.04747   1.528  0.12674
## genderWoman:housingHouse       -0.11329    0.04129  -2.743  0.00619 **
## genderWoman:housingOff Campus -0.11975    0.05821  -2.057  0.03995 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2197 on 966 degrees of freedom
## Multiple R-squared:  0.01697,    Adjusted R-squared:  0.01189
## F-statistic: 3.336 on 5 and 966 DF,  p-value: 0.005407
```

The differences between men and women vary across types of housing. The average differences between men and women are significantly larger for those who live in houses rather than dorms and for those who live off campus rather than in dorms.

5. Save the predicted values from the model in #4. Create a table with `group_by()` and `summarize()` showing how the predicted values differ by gender and housing type. Interpret any interesting results.

```
midd_survey$pred_find_partner <- fitted(model3)

find_partner_predictions <- midd_survey |>
     group_by(gender, housing) |>
     summarize(prob_find_partner =
      round(mean(pred_find_partner, na.rm=TRUE),3))

huxtable(find_partner_predictions)
```

gender	housing	prob_find_partner
Man	Dorm	0.058
Man	House	0.135
Man	Off Campus	0.13
Woman	Dorm	0.047
Woman	House	0.011
Woman	Off Campus	0