Lab 02 – Data Analysis

Melissa Juarez

March 7, 2022

Q1: Run a simple regression, and interpret your results. (Did the results fit your expectations? Why? Why not?) How does the a respondent's trust in their neighbors affect their sense of community?

In this problem, I look at the association between the following variables:

- `localcommunity`/`V213`: this is a ordinal categorical variable where 1 indicates that a respondent strongly agrees to seeing themselves as part of their local community and 4 indicates strongly disagreeing. *However, I will be recoding this variable for 1 to mean strongly disagree and for 4 to mean strongly agree. This is so that the wording of the analysis is easier to understand, where an increase in levels corresponds with an increase in feeling part of the community.*

- `neighbortrust`/`V103`: this is an ordinal categorical variable where 1 indicates that a respondent completely trusts people in their neighborhood and 4 indicates that they don't trust at all. *However, I will be recoding this variable for 0 to no trust at all and for 3 to trust completely. This is so that the wording of the analysis is easier to understand, where an increase in level corresponds with an increase in trust.*

I expect that as trust in your neighbors increases, respondents will feel more like they are part of the community. My reasoning is that it is easier to forge connections and maintain them (in order to build community) with those you trust.

```r
## need to rename variables and recode
wvs <- wvs %>%
  mutate(V213 = case_when(
    V213 == 1 ~ 4, #Strongly agree becomes highest number
    V213 == 2 ~ 3,
    V213 == 3 ~ 2,
    V213 == 4 ~ 1, #Strongly disagree becomes lowest number
  ), V103 = case_when(
    V103 == 1 ~ 3, #Trust Completely becomes highest number
    V103 == 2 ~ 2,
    V103 == 3 ~ 1,
    V103 == 4 ~ 0, #No trust at all becomes 0
  )) %>%
  dplyr::rename(c(localcommunity = V213, neighbortrust = V103)
  )


## create linreg model
m1 <- lm(localcommunity ~ neighbortrust, data = wvs)
summary(m1)

##
## Call:
## lm(formula = localcommunity ~ neighbortrust, data = wvs)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.31748 -0.19711 -0.07674  0.80289  1.04363
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.956371   0.007636  387.18   <2e-16 ***
## neighbortrust 0.120370   0.003736   32.22   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8113 on 70062 degrees of freedom
##   (3978 observations deleted due to missingness)
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.01458
## F-statistic:  1038 on 1 and 70062 DF,  p-value: < 2.2e-16
```

In the output above, we have the estimated model

$$localcommunity = 2.956 + 0.120 neighbortrust$$

. We can interpret the model in the following way:

- *Slope:* Given two respondents, where one respondent has a 1 level higher of trust in their neighbors as the other, the respondent with the highest trust is expected to report 0.120 levels higher for feeling like a part of their local community. In other words, for a 1 level increase in trust in your neighbors, the respondent is expected to have a 0.120 level increase in feeling like a part of their local community.
- *Intercept:* When a respondent no trust in their neighbors ($neighbortrust = 0$), the respondent is expected to feel at a 2.956 level (almost reaching 'agree') of feeling part of their community.
- R^2: It is given that the Multiple R-squared = 0.01458. We can interpret this by saying that 1.458% of the variation in feeling part of your local community is explained by its linear relationship with the amount of trust in your neighbors. This is not a great model.

Q2: Add an interaction term to that model that you think might moderate the original relationship between X1 and X2. Explain why you think an interaction might be present and in what direction it would work. Explain your results. Did it work out? Yes? No? In this problem, I look at the interaction effect of the following variable:

- `sex`/`V240`: this is a nominal categorical variable where 1 indicates that a respondent is a Male and 2 indicates that the respondent is female. I plan to include this variable to see whether the interaction between trusting your neighbors and sex is significant. I expect that there will be a interaction between sex and neighborhood trust, where the effect of `neighbortrust` on feeling part of the local community will decrease when the respondent is a woman. In other words, the slope of `neighortrust` will be lower than for males.

```
## need to rename variables and recode
wvs <- wvs %>%
  mutate(V240 = case_when(
    V240 == 1 ~ "Male",
    V240 == 2 ~ "Female",
  )) %>%
  dplyr::rename(sex = V240)

wvs$sex <- factor(wvs$sex, levels = c("Male", "Female")) #there is no reason to
# level these, but they are often already leveled with male first & because of
# that, it is easier for me to interpret if male is the baseline.
# forgive me!
```

```
## female will be a dummy variable 'u'
m2 <- lm(localcommunity ~ neighbortrust * sex, data = wvs)
summary(m2)
```

```
##
## Call:
## lm(formula = localcommunity ~ neighbortrust * sex, data = wvs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.36712 -0.21665 -0.08502  0.78335  1.08431
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.915692   0.011307 257.867  < 2e-16 ***
## neighbortrust            0.150477   0.005476  27.479  < 2e-16 ***
## sexFemale                0.076358   0.015325   4.982 6.29e-07 ***
## neighbortrust:sexFemale -0.057510   0.007490  -7.678 1.64e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8109 on 70045 degrees of freedom
##   (3993 observations deleted due to missingness)
## Multiple R-squared:  0.01579,    Adjusted R-squared:  0.01574
## F-statistic: 374.5 on 3 and 70045 DF,  p-value: < 2.2e-16
```

In the output above, we have the estimated model

$$localcommunity = 2.765 + 0.150 neighbortrust + 0.113 u_{female} - 0.058 neighbortrust * u_{female}$$

. All of the terms are significant with near zero p-values. We have two equations for the relationship between local community attitudes and trusting neighbors for males and for females. Each value has both different intercept and slope, showing that there is an interaction effect between neighborhood trust and sex: - males: $localcommunity = 2.765 + 0.150 neighbortrust$ - females: $localcommunity = 2.899 + 0.093 neighbortrust$

We can interpret the equation as follows:

- When sex == male, respondents have a lower baseline for feeling like a part of their community, with a 2.77 level of feeling like a part of their community, when they have no trust in their neighbors. When sex == female, respondents have a higher baseline for feeling like a part of their community, with a 2.77 level of feeling like a part of their community, when they have no trust in their neighbors.

- There is a significant interaction term which explains the differences in the effects of `neighbortrust` on the dependent variable. When sex == female, the effect of `neighbortrust` decreases by 0.057 compared to the `neighbortrust` effect for males, meaning. For an increase in 1 level of trusting in your neighbor, females are expected to show 0.057 lower levels of feeling like a part of their community than males. Specifically, for a 1 level increase of trusting your neighbor, females are expected to have a 0.093 level increase in feeling part of their community, whereas males are expected to have a 0.150 level increase in feeling part of their community. This works out as I expected, although the R^2 is still quite low.