

Lab 14 - Bivariate Regression & Interpretation

Yana Chakalo

November 28, 2017

```
install.packages("tidyverse", repo = "https://CRAN.R-project.org/package=tidyverse")
```

```
## Warning: unable to access index for repository https://CRAN.R-project.org/package=tidyverse/src/contrib:
##   cannot open URL 'https://CRAN.R-project.org/package=tidyverse/src/contrib/PACKAGES'
```

```
## Warning: package 'tidyverse' is not available (for R version 3.4.1)
```

```
## Warning: unable to access index for repository https://CRAN.R-project.org/package=tidyverse/bin/macosx/el-capitan/contrib/3.4/PACKAGES
##   cannot open URL 'https://CRAN.R-project.org/package=tidyverse/bin/macosx/el-capitan/contrib/3.4/PACKAGES'
```

```
install.packages("readr", repo = "https://CRAN.R-project.org/package=readr")
```

```
## Warning: unable to access index for repository https://CRAN.R-project.org/package=readr/src/contrib:
##   cannot open URL 'https://CRAN.R-project.org/package=readr/src/contrib/PACKAGES'
```

```
## Warning: package 'readr' is not available (for R version 3.4.1)
```

```
## Warning: unable to access index for repository https://CRAN.R-project.org/package=readr/bin/macosx/el-capitan/contrib/3.4/PACKAGES
##   cannot open URL 'https://CRAN.R-project.org/package=readr/bin/macosx/el-capitan/contrib/3.4/PACKAGES'
```

```
data2 <- read_tsv("~/Downloads/Final.Porject.Labs/ICPSR_23625_2002/DS0002/23625-0002-Data.tsv")
```

```
## Error in read_tsv("~/Downloads/Final.Porject.Labs/ICPSR_23625_2002/DS0002/23625-0002-Data.tsv"): could not find function 'read_tsv'
```

```
data13 <- read_tsv("~/Downloads/Final.Porject.Labs/ICPSR_36118_2013/DS0002/36118-0002-Data.tsv")
```

```
## Error in read_tsv("~/Downloads/Final.Porject.Labs/ICPSR_36118_2013/DS0002/36118-0002-Data.tsv"): could not find function 'read_tsv'
```

```
data_subset2 <- data2 %>%
```

```
  select(CITY, STATECOD, POP1, TNUMVTMS, TNUMOFF, GOFFRAC, BIASMO1, VTYP_I1, LOCCOD1, VTYP_I1, LOCCOD1)
```

```
## Error in data2 %>% select(CITY, STATECOD, POP1, TNUMVTMS, TNUMOFF, GOFFRAC, : could not find function 'select'
```

```
data_subset13 <- data13 %>%
```

```
  select(CITY, STATECOD, POP1, TNUMVTMS, TNUMOFF, GOFFRAC, BIASMO1, VTYP_I1, LOCCOD1, VTYP_I1, LOCCOD1)
```

```
## Error in data13 %>% select(CITY, STATECOD, POP1, TNUMVTMS, TNUMOFF, GOFFRAC, : could not find function 'select'
```

Complete the following exercises below and include all code used to find the answers. Knit together the PDF document and commit both the Lab 14 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Select the main focal relationship you're interested in exploring for your poster project.

- Describe the response variable and the explanatory variable and the theoretical relationship you believe exists between these two variables.

The two variables that I want to look at is the bias motivation (BIASMO1) of the hate crime and the offenders race of the hate crime. The explanatory variable (x variable) is the offenders race and the response variable (y variable) is the bias motivation. I think these two exist in a relationship because hate crimes are a specific type of crime that focuses on someones identity so in theory the offenders race would thus determine what the bias motivation behind the hate crime was.

- Conduct a simple (bivariate) linear regression on your focal relationship and save the model object. Print out the full results by calling `summary()` on your model object.

2002

```
BIASMO1_GOFFRAC <- lm(formula = BIASMO1 ~ GOFFRAC, data = data_subset2)

## Error in is.data.frame(data): object 'data_subset2' not found
plot(BIASMO1_GOFFRAC)

## Error in plot(BIASMO1_GOFFRAC): object 'BIASMO1_GOFFRAC' not found
summary(BIASMO1_GOFFRAC)

## Error in summary(BIASMO1_GOFFRAC): object 'BIASMO1_GOFFRAC' not found
```

2013

```
BIASMO1_GOFFRAC <- lm(formula = BIASMO1 ~ GOFFRAC, data = data_subset13)

## Error in is.data.frame(data): object 'data_subset13' not found
plot(BIASMO1_GOFFRAC)

## Error in plot(BIASMO1_GOFFRAC): object 'BIASMO1_GOFFRAC' not found
summary(BIASMO1_GOFFRAC)

## Error in summary(BIASMO1_GOFFRAC): object 'BIASMO1_GOFFRAC' not found
```

- c. What is the direction, magnitude, and statistical significance of the bivariate association between the explanatory and response variables.

First the 2002 data does not follow a normal distribution because the points are very far off the linear line. In the 2013 data it follows the normal distribution better than the 2002 data but still is not following the line so this data too is not on a normal distribution. Statistical significance of 2002 data is shown through the p-value which is 0.1236. Which means that the p-value is high which my data are likely with a true null. Which means that I fail to reject the null hypothesis meaning my data is likely with a true null hypothesis. So the 2002 data statistical significance is showing there is little significance and no affect between these two variables. The 2013 p-value is 2.2×10^{-16} is a low p-value meaning data are unlikely with a true null. So this means that the 2013 data between these two variables has a high statistical significance and reliable. My 2002 and 2013 data intercept is a positive association.

- d. What is the meaning of the model intercept?

The model intercept for 2002 is about 20 and the 2013 data is about 23. This means that the mean of Y when all predictors equals zero. However, this data is only meaningful if X does have some values of zero but since my values are categorical there is no meaningful zeros so the slope intercept for this model has no meaning.

- e. How well does the bivariate model fit the data? How is this information calculated?

The bivariate model does not fit my data that much because both of my variables are categorical so when calculating a regression model, intercept, and means the numbers do not have any meaningful numeric value when calculated for a bivariate model.

- f. Is the observed association between the independent variable and dependent variable consistent with your hypothesis? Why or why not?

When looking at the p-value's of both the datasets I would say that for 2002 it does not make sense because the p-value is pretty big meaning there is not a significant association while the 2013 data with the same variables have a small p-value meaning the statistical significance is high. So maybe this means that back in

2002 the offenders race did not determine the bias motive of the hate crime but in 2013 the race does matter determining the bias motive. Or that the data is getting two different results because this bivariate model does not fit my data nor does it accurately represent this association.

2. Select a different focal relationship related to your project. This could be:

- A different response and a different explanatory variable
- A different response and the same explanatory variable
- The same response and a different explanatory variable

- a. Describe the response variable and the explanatory variable and the theoretical relationship you believe exists between these two variables.

The response variable is the location of the hate crime (LOCCOD1) and the explanatory variable is the bias motivation of the hate crime (BIASMO1). The location of the hate crime could explain what the bias motive behind the crime is because where crimes occur matters a lot in understanding what the circumstances of the crime and its people involved are in. So I can see that the specific locations of the crimes can help us understand what the motivation behind these crimes are and even who the offenders / victims were.

- b. Conduct a simple (bivariate) linear regression on your focal relationship and save the model object. Print out the full results by calling `summary()` on your model object.

2002

```
LOCCOD1_BIASMO1 <- lm(formula = LOCCOD1 ~ BIASMO1, data = data_subset2)

## Error in is.data.frame(data): object 'data_subset2' not found
plot(LOCCOD1_BIASMO1)

## Error in plot(LOCCOD1_BIASMO1): object 'LOCCOD1_BIASMO1' not found
summary(LOCCOD1_BIASMO1)

## Error in summary(LOCCOD1_BIASMO1): object 'LOCCOD1_BIASMO1' not found
```

2013

```
LOCCOD1_BIASMO1 <- lm(formula = LOCCOD1 ~ BIASMO1, data = data_subset13)

## Error in is.data.frame(data): object 'data_subset13' not found
plot(LOCCOD1_BIASMO1)

## Error in plot(LOCCOD1_BIASMO1): object 'LOCCOD1_BIASMO1' not found
summary(LOCCOD1_BIASMO1)

## Error in summary(LOCCOD1_BIASMO1): object 'LOCCOD1_BIASMO1' not found
```

- c. What is the direction, magnitude, and statistical significance of the bivariate association between the explanatory and response variables.

For the 2002 data the p-value is 0.4412 which is a high p-value meaning that the location of the crime committed is not statistically significant to determining the bias motive. For 2013 data the p-value is 0.2376 which is comparatively smaller than the 2002 data, however, that is still a high number meaning the two

variables are also not statistically significant to each other in 2013. The direction of the association for both datasets is positive.

d. What is the meaning of the model intercept?

Again for this data both variables are categorical so the intercept has no meaning in relation to my data.

e. How well does the bivariate model fit the data? How is this information calculated?

This data does not fit bivariate models very well because the variables are categorical. So the numbers are just a category placed for each level.

f. Is the observed association between the independent variable and dependent variable consistent with your hypothesis? Why or why not?

When looking at the p-value the statistical significance for both data was not that surprising because location of a crime can be more malleable and less meaningful since location type can be defined with many different labels. However, I was surprised at how high the p-values were and I thought that location would be associated a little more with bias motivation because the location of a crime can be very telling too. I would have to look at other models to run to find an association or not with more confidence.