Lab 14 - Bivariate Regression & Interpretation

Snow Christensen

December 1, 2017

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.

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

I'm interested in looking at the relationship between the variables PLEASE and PRESSSI. The response variable would be PRESSSI which asks the respondent about whether or not they've been forced into sexual activity because of pressure. The explanatory variable would be PLEASE, which is about whether or not the respondent feels the need to please others. I am interested in looking at whether there is a correlation between people who want to please others and those who are targetted by sexual predators through pressure. My hypothesis is that there is a positive correlation between the variables PLEASE and PRESSSI.

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.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.4.3
```

```
library(ggplot2)

#install.packages("haven")
library(haven)

#load full dataset
data <- read_por("H:/Projects/sexual_violence_college-master/ICPSR_03212/DS0001/03212-0001-Data.por")

#data subset
data_ID <- data %>%
  dplyr::select(CODENUM, RACE, NOASSERT, PLEASE, PRESSSI, AUTHSI)

#convert data to factor
PRESSSI<- factor(data_ID$PRESSSI)
PLEASE <- factor(data_ID$PLEASE)
```

```r
data_ID <- as.data.frame(data_ID)

#conduct linear regression on data
LR_model <- lm(PLEASE ~ PRESSSI, data = data_ID)


#print out results of linear regression
summary(LR_model)
```

```
##
## Call:
## lm(formula = PLEASE ~ PRESSSI, data = data_ID)
##
## Residuals:
## <Labelled double>
##      Min       1Q   Median       3Q      Max
## -1.04638 -0.85352 -0.04638  0.14648  3.14648
##
## Labels:
##  value              label
##      0        No response
##      1 Not at all like me
##      2   A little like me
##      3   Somewhat like me
##      4     Mostly like me
##      5  Very much like me
##      9            Missing
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.66066    0.08032  20.674  < 2e-16 ***
## PRESSSI      0.19286    0.06233   3.094  0.00201 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.022 on 1565 degrees of freedom
##   (13 observations deleted due to missingness)
## Multiple R-squared:  0.006079,   Adjusted R-squared:  0.005444
## F-statistic: 9.572 on 1 and 1565 DF,  p-value: 0.00201
```

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

The direction is negative which we can see from the slope of the output which is -.019286. The magnitude is the absolute value of the slope which is 0.019286. The association is statistically significant because the p-value is only 0.00201 which means it is over 99% likely that the association is not due to chance.

d. What is the meaning of the model intercept? The model intercept is the expected response for PLEASE when when looking at the average of all responses for PRESSSI. This means that the average person in this dataset responded with the answer coded with two in regards to if it is like them to want to please people. The answer coded with 2 is 'A little like me.'

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

The bivariate model fits the data pretty well. We can judge this by seeing how evenly the values for the residuals are spread. All of the residual values for this regression are fairly close together ranging from around

-1 to 3.

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

No, this associate is opposite of my hypothesis. Originally, I had believed there would be a positive association between PLEASE and PRESSSI because I believed people who liked to please others would be targetted more often for sexual violence by people using pressure.

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.

I'm interested in looking at the relationship between the variables NOASSERT and AUTHSI. The response variable would be AUTHSI which asks the respondent about whether or not they've been forced into sexual activity because of authority. The explanatory variable would be NOASSERT, which is about whether or not the respondent feels the need to please others. I am interested in looking at whether there is a correlation between people who are not assertive and who is targetted by sexual predators in authoritative positions. My hypothesis is that there is a positive correlation between the variables NOASSERT and AUTHSI.

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.

```
#convert data to factor
AUTHSI <- factor(data_ID$AUTHSI)
NOASSERT <- factor(data_ID$NOASSERT)

#conduct linear regression on data
LR_model2 <- lm(NOASSERT ~ AUTHSI, data = data_ID)

#print out results of linear regression
summary(LR_model2)
```

```
##
## Call:
## lm(formula = NOASSERT ~ AUTHSI, data = data_ID)
##
## Residuals:
## <Labelled double>
##     Min      1Q  Median      3Q     Max
## -1.0769 -0.6667 -0.6667  0.3333  3.3333
##
## Labels:
##  value              label
##      0        No response
##      1 Not at all like me
##      2   A little like me
##      3   Somewhat like me
##      4     Mostly like me
##      5  Very much like me
##      9            Missing
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.2564     0.1842   6.820  1.3e-11 ***
## AUTHSI        0.4103     0.1798   2.282   0.0226 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9092 on 1566 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.003313,   Adjusted R-squared:  0.002677
## F-statistic: 5.206 on 1 and 1566 DF,  p-value: 0.02264
```

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

The direction is negative which we can see from the slope of the output which is -.4103. The magnitude is the absolute value of the slope which is 0.4103. The association is statistically significant because the p-value is only 0.02264 which means it is about 98% likely that the association is not due to chance.

d. What is the meaning of the model intercept? The model intercept is the expected response for AUTHSI when when looking at the average of all responses for NOASSERT. This means that the average person in this dataset responded with the answer coded with 2 in regards to if it is like them to not be assertive. The answer coded with 2 is 'A little like me.'

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

The bivariate model fits the data pretty well. We can judge this by seeing how evenly the values for the residuals are spread. All of the residual values for this regression are fairly close together ranging from around -1 to 3.

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

No, this associate is opposite of my hypothesis. Originally, I had believed there would be a positive association between NOASSERT and AUTHSI because I believed people who lacked assertiveness would be more likely to be chosen as victims of sexual violence from authority figures.