

Lab 14 - Bivariate Regression & Interpretation

Your name here

November 28, 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.

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

I believe that the seriousness of the offense correlates with the punishment of individuals.

- 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(lm.beta)
# read in data
data <- read.csv("~/monitoring-federal-criminal-sentences/clean_data/merged_data/96-15.csv")

a <- lm(data = data, TOTPRISN ~ XFOLSOR)
summary(a)
```

```
##
## Call:
## lm(formula = TOTPRISN ~ XFOLSOR, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -177.89  -21.48   -3.70   17.59   898.28
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -49.310040   0.089846  -548.8   <2e-16 ***
## XFOLSOR       5.231743   0.004355  1201.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42.84 on 1219240 degrees of freedom
## (7825 observations deleted due to missingness)
## Multiple R-squared:  0.542, Adjusted R-squared:  0.542
## F-statistic: 1.443e+06 on 1 and 1219240 DF, p-value: < 2.2e-16
```

```
lm.beta(a)
```

```
##
## Call:
## lm(formula = TOTPRISN ~ XFOLSOR, data = data)
##
## Standardized Coefficients::
## (Intercept)      XFOLSOR
##  0.0000000    0.7362188
```

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

The seriousness of the crime has a p value of 0, and a positive 5.23 slope. Beta is .736, which is pretty positively strong.

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

For every 1 level of increase, the defendant receives on average of 5.23 extra months in sentence length

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

The multiple r-squared is 0.542, which is pretty good, this is calculated by the sd and how well it fits the line.

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

The association is consistent with my hypothesis because you expect more severe punishment with more serious crimes committed by the individual.

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 believe that the severity of the crime committed by the defendant has to do with the the type of the crime.

- 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.

```
b <- lm(data = data, XFOLSOR ~ factor(TYPE))
summary(b)
```

```
##
## Call:
## lm(formula = XFOLSOR ~ factor(TYPE), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.057  -5.057  -0.980   5.020  34.992
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.00794    0.01404  1068.6  <2e-16 ***
## factor(TYPE)1   9.04877    0.01790   505.5  <2e-16 ***
## factor(TYPE)2   4.97222    0.02609   190.6  <2e-16 ***
## factor(TYPE)3  -2.08297    0.01933  -107.8  <2e-16 ***
## factor(TYPE)4   7.64904    0.03742   204.4  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.485 on 1219039 degrees of freedom
## (8023 observations deleted due to missingness)
## Multiple R-squared:  0.2939, Adjusted R-squared:  0.2939
## F-statistic: 1.268e+05 on 4 and 1219039 DF, p-value: < 2.2e-16
```

```
lm.beta(b)
```

```
##  
## Call:  
## lm(formula = XFOLSOR ~ factor(TYPE), data = data)  
##  
## Standardized Coefficients:  
##      (Intercept) factor(TYPE)1 factor(TYPE)2 factor(TYPE)3 factor(TYPE)4  
##      0.0000000    0.4912859    0.1637408    -0.1026403     0.1646291
```

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

Drug crimes, gun violence, and violent crime are positive, whereas immigration is slightly negative but weak. However, they are all significant.

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

The model's intercept means that all crimes has an offense level of 15.

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

The multiple r-squared is 0.2939. Which means that the model explains 29.4% of the variability, not too bad for 1 single dependent variable.

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

No, it is so hard to determine what type of crime is more severe than the other, but this gives us a good indication of what our institution think of different types of crimes.