Name:

Stats in practice #5

Scatterplots and Correlation

# The Dataset

### cdi <- read.csv(‘http://math.unm.edu/~knrumsey/cdi\_sample.csv’)

###

# Lurking Variables

## Crime and doctors

### Consider the variables, Total serious crime and Number of Physicians.

### totalCrimes <- cdi$TotalSeriousCrimes

### numPhysicians <- cdi$NumberOfPhysicians

### Create a scatterplot and calculate the correlation of these variables by typing:

### plot(totalCrimes, numPhysicians)

### cor(totalCrimes, numPhysicians)

### is the relationship stronger than you expected? What lurking variable explains this?

### We can account for the lurking variable by comparing these variables *per capita.* Create two new variables by dividing the old ones by “cdi$population”. Create a scatterplot and calculate the correlation of these variabels. **Explain what happens and copy/paste your plot below.**

# Correlation

### Consider 4 different variables

#### X1 = percent of adult population with a HS diploma

### x1 <- cdi$PercentHSGraduates

#### X2 = percent of adult population with a Bachelors degree

### x2 <- cdi$PercentWithBachelors

#### X3 = serious crimes per capita

### x3 <- cdi$TotalSeriousCrimes/cdi$Population

#### X4 = percent of population below poverty level

### x4 <- cdi$PercentBelowPoverty

### Think carefully about the relationship between each of these variables. Fill in the following table by guessing what you think the correlation between each variable will be. Primarily, focus on the *direction* and *strength* of the relationship.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1  | X2 | X3 | X4 | X5 |
| X1 |  |  |  |  |  |
| X2 |  |  |  |  |  |
| X3 |  |  |  |  |  |
| X4 |  |  |  |  |  |

### Calculate the correlations in R. (*You can save time by remembering that* *Cor(X1, X2) = Cor(X2, X1)).* How close is this to your answers in part ii.?

### Choose any pair of the variables above that you find most interesting and create a scatterplot. The title of your plot should include the correlation. The following code requires adjustment wherever you see a “?”.

### plot(?, ?, pch=?, col=?, main=paste(‘Correlation is ‘, round(cor(?, ?), 3), xlab=’?’, ylab=’?’)

### **Copy and paste this plot below.**