Name:

Stats in Practice, #4

Normal Distributions

# The Dataset

## Load the data into R

### Open RStudio and open a new “script”

### Load the dataset into R by typing:

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

# Variable: Crime per capita

### First, let’s create a variable “serious crime per capita” in R by typing:

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

### Create a histogram of this data by typing:

### hist(crimePerCapita, probability=TRUE)

### Find the mean and standard deviation of this variable. (Try to remember how to do this in R on your own).

###  Add a Normal density curve to your histogram by typing:

### curve(dnorm(x, mean, sd), add=TRUE, col=’red’)

### where “mean” and “sd” represent the mean and standard deviation that you found in part iii.

### Assume that crime per capita in a randomly selected US county is Normally distributed with the mean and standard deviation found in part iii. Use the Z-table to find the proportion of US counties which have a crime per capita rate *less than* 0.1.

### For the n=100 counties in our dataset, find the actual proportion which have a crime per capita rate less than 0.1. This can be done by typing:

### sum(crimePerCapita < 0.1)/length(crimePerCapita)

### How does this compare to your answer in part v.?

# Variable: Percentage of population below poverty

## The Normal distribution

### First, let’s create a variable in R by typing:

### percentPoverty <- cdi$PercentBelowPoverty

### Create a histogram of this variable by typing:

### hist(percentPoverty, probability=TRUE)

### Does this data appear to be normally distributed?

### Sometimes we can use a “transformation” to make the distribution of a variable more Normal. This can make the variable easier to work with. Let us create a new variable which is the square root of percentPoverty.

### sqrtPoverty <- sqrt(percentPoverty)

### Create a histogram of this new variable. Does it look more Normal?

### Now consider a new transformation:

### z <- (sqrtPoverty – mean(sqrtPoverty))/sd(sqrtPoverty)

### Create a histogram of the variable z and find its mean and variance. Explain what just happened. **Copy and paste this plot below.**