NYC Crime Analysis

Malcolm Hess

Wednesday, September 09, 2015

The purpose of this project is to look at crime in NYC and determine if there is a significant change, either decrease or increase in the amount of crime that takes place in the city of New York.

Crime data was collected from the FBI and aggregated into a total crime statistic. The original crime breakdown included Murder, Rape, Robberies, Assaults, Burglaries, Theft, and Auto Theft. The data was taken from an aggregated source here :http://www.city-data.com/crime/crime-New-York-New-York.html: but with original sourcing at the FBI site https://www.fbi.gov/stats-services/crimestats

Population was found through Google's open data software which has orignal sourcing at the US Census Bureau.

First steps are loading the data in.

```
nyccrime <- read.csv("NYC_Crime.csv")
nyccrime <- nyccrime[8:9,]
```

I did individual prop tests to check and see if the changes in crime were significant or not. I normally would have run this through a while loop instead of individually coded as it appears below however R does not print out the summary of these proportion tests when they are done in a while loop. The code will compare sequential years to see if there is a significant change in the amount of crime.

The proportion test will test wether or not the proportion of crimes to population between two years is significantly different or not (two sided). The Null hypothesis is that the proportion of crimes to population between two years is the same.

```
i <- 2
```

```r
prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 570.7151, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.001920896 0.002264519
## sample estimates:
##     prop 1     prop 2
## 0.03306259 0.03096989
```

```r
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 504.4693, df = 1, p-value < 2.2e-16
```

```
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.001737140 0.002069566
## sample estimates:
##     prop 1     prop 2
## 0.03096989 0.02906653
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 248.1098, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.001135318 0.001458270
## sample estimates:
##     prop 1     prop 2
## 0.02906653 0.02776974
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 276.6945, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.001177333 0.001492099
## sample estimates:
##     prop 1     prop 2
## 0.02776974 0.02643502
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 382.696, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.001372164 0.001677992
```

```
## sample estimates:
##     prop 1     prop 2
## 0.02643502 0.02490995
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 125.2757, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.0007012419 0.0009992304
## sample estimates:
##     prop 1     prop 2
## 0.02490995 0.02405971
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
```

```
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 14.7681, df = 1, p-value = 0.0001216
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.0001409780 0.0004347122
## sample estimates:
##     prop 1     prop 2
## 0.02405971 0.02377186
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 326.2766, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.001182690 0.001470848
## sample estimates:
##     prop 1     prop 2
```

```
## 0.02377186 0.02244510
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 49.9392, df = 1, p-value = 1.586e-12
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.0006606217 -0.0003735753
## sample estimates:
##     prop 1     prop 2
## 0.02244510 0.02296219
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
```

```
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 5.0203, df = 1, p-value = 0.02505
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -3.107741e-04 -2.074995e-05
## sample estimates:
##     prop 1     prop 2
## 0.02296219 0.02312796
```

```
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 13.4397, df = 1, p-value = 0.0002464
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.0004158763 -0.0001260451
## sample estimates:
##     prop 1     prop 2
## 0.02312796 0.02339892
```

```r
i <- i + 1

prop.test(c(nyccrime[1,i], nyccrime[1,(i+1)]), c(nyccrime[2,i], nyccrime[2,(i+1)]))
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(nyccrime[1, i], nyccrime[1, (i + 1)]) out of c(nyccrime[2, i], nyccrime[2, (i + 1)])
## X-squared = 24.8969, df = 1, p-value = 6.048e-07
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.0002225864 0.0005108279
## sample estimates:
##     prop 1     prop 2
## 0.02339892 0.02303221
```

As you can see that every one of the proportion tests has come out significant. Which means we can reject the null hypothesis that the proportion of crimes to population between any two sequential years are the same and accept the alternate hypothesis that every year since 2001 has seen a significant change in crime. This includes three years 2010, 2011 and 2012 that saw a significant increase in crime per population rate