A Start to Reporting

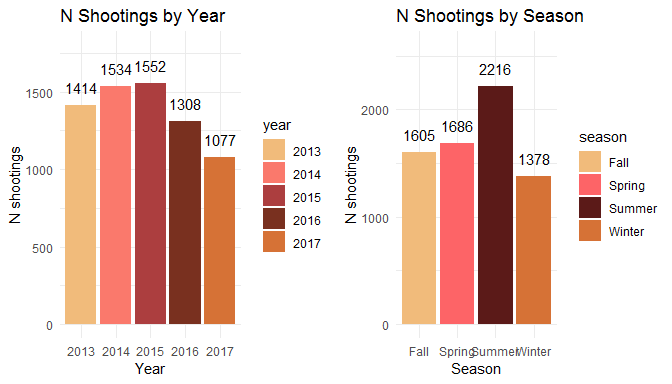
Rachel Weber

December 12, 2018

### Background

NYC shooting data is publicly available for access through <https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8>. The data are updated annually and contain data from January 1, 2013 to December 31, 2017–5 years of data. In this time 6885 shootings occured. Their date, time, precinct, borough, and location are available along with perpetrator and victim age, race, and sex. The weekend of october 12th, 2018 was the first shooting-free weekend the city has had in over a decade. This headline made national news and being familiar with the law of truly large numbers, the question arose asking, “Was this truly a newsworthy event of was it bound to happen?” Figure 1 below shows the number of shootings stratified by year (left) and by season (right). Shootings have been reportedly going down in recent years, so a shooting-free weekend may be becoming more likely. Additionally, shooting rates change with the season, peaking in summer, so the probability of a shooting-free weekend may change with the season.

|  |  |
| --- | --- |
|  | N = 6,885 |
|  | N(%) |
| **year** |  |
| 2013 | 1,414(21) |
| 2014 | 1,534(22) |
| 2015 | 1,552(23) |
| 2016 | 1,308(19) |
| 2017 | 1,077(16) |
| **season** |  |
| Fall | 1,605(23) |
| Spring | 1,686(24) |
| Summer | 2,216(32) |
| Winter | 1,378(20) |



### Analysis

One cannot assume that shootings on Saturday are completely independent of shootings on Friday. As such, we cannot multiply the days’ probabilities together to get our weekend probability. The definition of conditional probability though, does not necessitate independence and can be used here. The probability of a shooting free weekend was calculated as follows:

Therefore, the probability of our full weekend being shooting-free is:

The probability comes to be astoundingly small. The odds of a shooting free weekend given the last 5 years of data are about 7 is a billion. The odds of winning the Powerball jackpot are 1 in 292 million.

We calculated this assuming that Sundays shooting aren’t independent of Fridays shootings. But how affected is it really. In the table below, we explore how our odds change given previous days’ shooting totals.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Friday | Saturday | Sunday |
| **P(0)** | 1.62e-02 | 3.2e-03 | 9.22e-03 |
| **P(0|0)** |  | 3.2e-03 | 9.22e-03 |
| **P(0|1)** |  | 3.2e-03 | 9.22e-03 |
| **P(0|0,0)** |  |  | 1.5e-04 |
| **P(0|0,1)** |  |  | 9.07e-03 |
| **P(0|1,0)** |  |  | 1.5e-04 |
| **P(0|1,1)** |  |  | 9.07e-03 |

The odds of a shooting free Sunday when we’re looking at just Saturdays data–row 2–and when we’re also including friday are nearly identical. This would imply that our estimate for Sunday doesn’t depend of Friday’s shooting status. However, look at what happens when we compare odds of shooting-free Sundays given that Saturday had at least 1 shooting. Knowing that Friday was shooting-free increased our probability for Sunday a full order of magnitude. Is this difference significant? One could argue no. We’re comparing odds of .00922 to .00015. Both are vanishingly small. So can be assume independence of Sunday from Friday? Well, it depends. I’ll explore that later.

Let’s move into model building. Performing regression on our outcome of interest may shed more light than working with conditional probabilities. First I fit a cell means model using Day, Year, and Season as predictors of number of shootings in a given day. This value ranged from 0 to 19 in the 5 years of data used here.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| **Friday** | 3.598 | 0.2406 | 14.95 | 9.26e-48 |
| **Monday** | 3.462 | 0.2407 | 14.38 | 1.733e-44 |
| **Saturday** | 4.653 | 0.2406 | 19.33 | 7.118e-76 |
| **Sunday** | 4.824 | 0.2406 | 20.05 | 6.776e-81 |
| **Thursday** | 2.595 | 0.2406 | 10.78 | 2.558e-26 |
| **Tuesday** | 2.877 | 0.2403 | 11.97 | 7.741e-32 |
| **Wednesday** | 2.941 | 0.2406 | 12.22 | 4.522e-33 |
| **2014** | 0.3286 | 0.2033 | 1.616 | 0.1062 |
| **2015** | 0.3898 | 0.2033 | 1.918 | 0.05533 |
| **2016** | -0.3303 | 0.2032 | -1.626 | 0.1042 |
| **2017** | -0.9889 | 0.2033 | -4.864 | 1.247e-06 |
| **Spring** | 0.1255 | 0.1816 | 0.6909 | 0.4897 |
| **Summer** | 1.315 | 0.1816 | 7.245 | 6.383e-13 |
| **Winter** | -0.4975 | 0.1825 | -2.727 | 0.006461 |

The model suggests that knowing the day of the week is important in predicting how many shootings are likely to occur. It being summer is predicted to increase the number of shootings by just over 1 for the day and it being 2017 is predicted to lower the number of shootings by about 1 shooting.

But we aren’t that interested in how many shootings there were per day but whether or not there were none at all. We can model this with logistic regression with a binary outcome for shootings on a given day:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | z value | Pr(>|z|) |
| **Friday** | 3.003 | 0.3551 | 8.458 | 2.713e-17 |
| **Monday** | 2.505 | 0.321 | 7.804 | 5.982e-15 |
| **Saturday** | 4.444 | 0.5602 | 7.932 | 2.15e-15 |
| **Sunday** | 3.31 | 0.3834 | 8.633 | 5.954e-18 |
| **Thursday** | 1.96 | 0.2962 | 6.62 | 3.598e-11 |
| **Tuesday** | 2.229 | 0.3064 | 7.274 | 3.479e-13 |
| **Wednesday** | 2.339 | 0.3122 | 7.492 | 6.773e-14 |
| **2014** | -0.0008647 | 0.2917 | -0.002964 | 0.9976 |
| **2015** | 0.183 | 0.3034 | 0.6032 | 0.5464 |
| **2016** | -0.3716 | 0.273 | -1.361 | 0.1735 |
| **2017** | -0.7126 | 0.2603 | -2.737 | 0.006195 |
| **Spring** | -0.1292 | 0.2376 | -0.5439 | 0.5865 |
| **Summer** | 0.4334 | 0.2675 | 1.62 | 0.1052 |
| **Winter** | -0.3515 | 0.23 | -1.528 | 0.1264 |

Our results suggest that season does not impact whether a given day has a shooting and that having the day be in 2017 is associated with reduced odds in a shooting occuring.