An Analysis of Automobile Collisions by Ward in Washington, DC

Timothy Chen Allen & Kevin Bembridge

University of the District of Columbia

Dr Valbona Bejleri

20 March 2014

Abstract

This analysis considers automobile collision data for the Wards of the District of Columbia, using Exploratory Data Analysis (EDA) to find characteristics of the data, and then creating a 95% Simultaneous Confidence Interval (SCI) to determine the Ward that has the most collisions.

*Keywords: Automotive Collisions, Exploratory Data Analysis, Multiple Comparisons*

Data Description

The District of Columbia is organized into eight Wards (See Figure 1, DDOT, 2009). Traffic incident data are collected by the District Department of Transportation and periodically published (DDOT, 2009). An analysis of which Ward has the most collisions is of interest in terms of strategic resource allocation, such as investing in mitigation equipment, staging emergency response resources, and increasing public awareness.

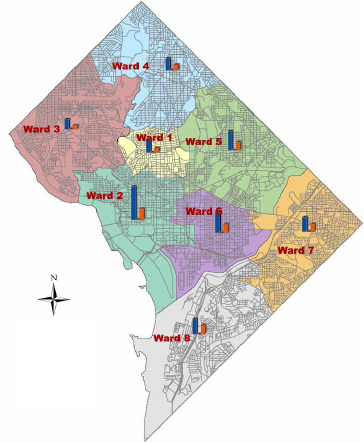


Figure 1. DC Wards System

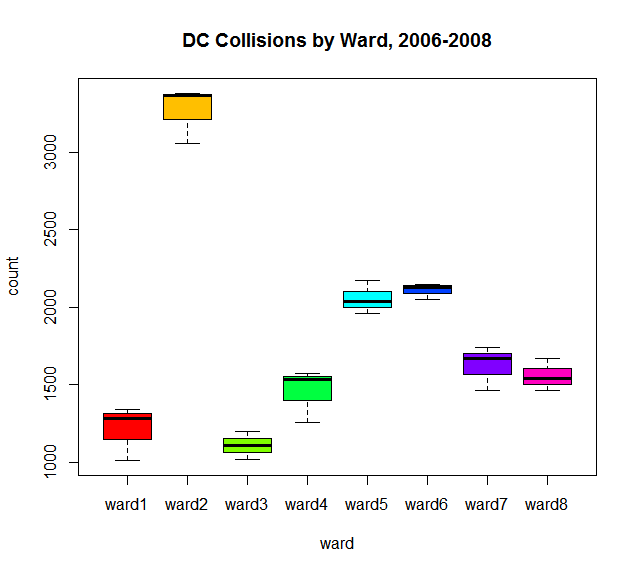
Data published by the DC Department of Transportation (DDOT, 2009) compares collisions reported to the DDOT in the eight districts between the years 2006 and 2008, including fatalities and injuries, as well as collisions occurring on borders and in unknown locations.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ward** | **2006 Collisions** | **2006 Fatalities** | **2006 Injuries** | **2007 Collisions** | **2007 Fatalities** | **2007 Injuries** | **2008 Collisions** | **2008 Fatalities** | **2008 Injuries** |
| 1 | 1,283 | 1 | 470 | 1,014 | 6 | 316 | 1,344 | 1 | 435 |
| 2 | 3,379 | 2 | 1,155 | 3,059 | 2 | 988 | 3,364 | 8 | 1,094 |
| 3 | 1,200 | 1 | 454 | 1,107 | 4 | 412 | 1,018 | 4 | 377 |
| 4 | 1,574 | 2 | 853 | 1,537 | 3 | 751 | 1,260 | 3 | 579 |
| 5 | 2,172 | 8 | 1,023 | 2,038 | 6 | 1,008 | 1,958 | 4 | 870 |
| 6 | 2,052 | 3 | 892 | 2,127 | 8 | 952 | 2,149 | 3 | 889 |
| 7 | 1,743 | 6 | 829 | 1,669 | 14 | 888 | 1,462 | 9 | 771 |
| 8 | 1,668 | 11 | 792 | 1,465 | 4 | 720 | 1,541 | 6 | 903 |
| Border | 1,133 | 7 | 593 | 1,090 | 7 | 536 | 1,472 | 1 | 627 |
| Unknown | 0 | NA | NA | 0 | NA | NA | 579 | 0 | 247 |
| Total | 16,204 | 41 | 7,061 | 15,106 | 54 | 6,571 | 16,147 | 39 | 6,792 |

As our interest was in comparing collisions between wards, we simplified this data set to include only the three years of collision data associated with Wards 1-8.

|  |  |  |  |
| --- | --- | --- | --- |
| **Ward** | **2006 Collisions** | **2007 Collisions** | **2008 Collisions** |
| 1 | 1,283 | 1,014 | 1,344 |
| 2 | 3,379 | 3,059 | 3,364 |
| 3 | 1,200 | 1,107 | 1,018 |
| 4 | 1,574 | 1,537 | 1,260 |
| 5 | 2,172 | 2,038 | 1,958 |
| 6 | 2,052 | 2,127 | 2,149 |
| 7 | 1,743 | 1,669 | 1,462 |
| 8 | 1,668 | 1,465 | 1,541 |

Initial Exploratory Data Analysis:



A box plot gives the initial impression that Ward 2 has a large number of collisions compared to the other wards. We would like to quantify this further.

Simultaneous Confidence Interval (SCI) comparison to best

We created a Simultaneous Confidence Interval (SCI) comparison to best (considering “best” as the ward with most collisions) (Keuhl, 2000).

1. ANOVA gives us MSE (which we will use as s2) and the corresponding df:

Analysis of Variance Table

Response: count

Df Sum Sq Mean Sq F value Pr(>F)

ward 7 10030292 1432899 77.9 3.6e-11 \*\*\*

Residuals 16 294363 18398

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

MSE = 18398, df = 16

Besides giving us input for the SCI, the ANOVA indicates to us that the mean number of collisions per ward is not equal for all wards; we suspected this from our Exploratory Data Analysis.

Calculate M with the inputs

α = 0.05, k = t – 1 = 7 comparisons, ν = 16 degrees of freedom. s2 = 18398, and r = 3

Appendix VI gives us a Bonferroni t of dα,k,ν = 2.56. So,

Calculate the 95% Simultaneous Confidence Interval and interpret:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Ward |  |  | Di | Di - M | Di + M | 95% SCI | Select? |
| 1 | 1213.7 | 3267.3 | -2053.6 | -2337.1 | -1770.1 | (-2337.1,0) | No |
| 2 | 3267.3 | 2109.3 | 1158 | 874.5 | 1441.5 | (0,1441.5) | Yes |
| 3 | 1108.3 | 3267.3 | -2159 | -2442.5 | -1875.5 | (-2442.5,0) | No |
| 4 | 1457 | 3267.3 | -1810.3 | -2093.8 | -1526.8 | (-2093.8,0) | No |
| 5 | 2056 | 3267.3 | -1211.3 | -1494.8 | -927.8 | (-1494.8,0) | No |
| 6 | 2109.3 | 3267.3 | -1158 | -1441.5 | -874.5 | (-1441.5,0) | No |
| 7 | 1624.7 | 3267.3 | -1642.6 | -1926.1 | -1359.1 | (-1926.1,0) | No |
| 8 | 1558 | 3267.3 | -1709.3 | -1992.8 | -1425.8 | (-1992.8,0) | No |

We select Ward 2 as having the most collisions, as it is the only ward with a 95% SCI with a lower limit of zero.

Further Investigation

We would like to continue the study by testing the hypothesis that collisions per ward are related to traffic flow per ward. We will seek traffic flow data and seek a correlation.

References

District Department of Transportation (2009). *Traffic safety report statistics*,retrieved from https://comp.ddot.dc.gov/Documents/Traffic%20Safety%20Report%20Statistics%202006-2008.pdf

Keuhl, R. (2000). *Design of experiments: statistical principles of research design and analysis*. Belmont: Duxworth