**Project- STAT 447**

**A Multivariate Approach to Investigate Motorcycle Crash Severity in Iowa**

**Submitted by:** Dimitrios Bilionis

 Mohammad Shaheed

**Scope:**

The scope of this project is to investigate the conditional events of different severity levels of motorcycle crashes in the state of Iowa under certain circumstances, such as the surface condition at the point of the crash, the prevailing weather conditions during the crash and the day of the week.

**I can see a number of valuable reasons for conducting such a study.**

**1.By knowing the conditions at a reported crash site you can better predict the level of medical attention that might be needed when you arrive at the scene.**

**2. By having prior probability information associated with your X variables, you can arrive at an annual prediction model for crashes of various levels of ‘cost’. This could allow you to estimate annual related expenses.**

**3. I like the idea of comparing numbers for the big cities in IA. That could allow those cities to address items 1 & 2 more accurately.**

**4. As a warning aid to bikers: e.g. crash prob. given rainy weather, or given poor road condition**

**I would recommend that your stated focus not be on conditional probabilities. Too mathematically sterile (and to most- uninteresting). I would recommend that you focus on items of a more practical nature (such as items 1-3 above).**

**You also have two X variables that may well be highly correlated: road condition and weather condition. Surely, independent of weather condition, the condition of the road structurally should be a factor. I would suggest that you look at these two variables jointly. It could well be that they are highly dependent in relation to weather. If you could somehow separate poor road condition associated with structural integrity and due to weather, then you could identify the role that crappy roads play in crashes.**

**Finally, do not forget: For every number you give, there was an action that it arose from. Hence, any number that is important should be associated with a random variable. And that random variable should be addressed using course concepts.**

**Data**

Crash data maintained by the Iowa Department of Transportation were used to extract motorcycle crash data from year 2001 to 2008. To address the scope of our problem in a clear manner, a number of random variables were defined to be used in the current study and they are to be analyzed by applying the knowledge gained form the course STAT 447.

The variables considered for our study includes the severity levels of motorcycle crashes in Iowa, weather conditions, surface conditions and day of the week during the crashes.

Specifically, following are the random variables that were used in our analysis:

Y = the act of recording the severity level of a motorcycle crash

The possible values that this variable can take (i.e. its sample space) are the following

Y = 1, for a fatal crash

 2, for a major injury crash

 3, for a minor injury crash

 4, for a potential injury crash

 5, for a property damage only crash

X1 = the act of recording the surface condition during the crash

The possible values that this variable can take (i.e. its sample space) are the following

X1 = 1, for dry condition

 2, for wet condition

 3, for sand/mud/dirt/oil or gravel on the surface

 4, for adverse conditions (i.e. ice, snow, slush, water on the surface)

 5, for other conditions (i.e. unknown or not reported)

X2 = the act of recording the weather conditions at the time of the crash

The possible values that this variable can take (i.e. its sample space) are the following

X2 = 1, for clear weather

 2, for party cloudy weather

 3, for cloudy weather

 4, for rainy weather

 5, for adverse weather (i.e. fog/smoke, mist, sleet, snow or severe winds)

 6, for other weather conditions (i.e. unknown or not reported)

X3 = the act of recording whether a crash happened on a weekday or weekend

The possible values that this variable can take (i.e. its sample space) are the following

X3 = 0, for weekdays

 1, for weekend

**Studying the variables**

Probabilities of all the possible events associated with the variables of our study.

**Variable Y:** Crash Severity Level

|  |  |  |  |
| --- | --- | --- | --- |
| **[Y=y]** | **Description** | **N[Y=y]** | **Pr[Y=y]** |
| **1** | Fatal Crash | 190 | 0.052 |
| **2** | Major Injury | 891 | 0.245 |
| **3** | Minor Injury | 1514 | 0.415 |
| **4** | Possible Injury | 704 | 0.193 |
| **5** | Property Damage Only | 345 | 0.095 |
| **Total** |   | 3644 | 1 |

**Variable X1:** Surface Condition during the crash

|  |  |  |  |
| --- | --- | --- | --- |
| **X1=x1** | **Description** | **N[X1=x1]** | **Pr[X1=x1]** |
| **1** | Dry Conditions | 3086 | 0.847 |
| **2** | Wet Conditions | 141 | 0.039 |
| **3** | Sand/Mind/Oil/Gravel | 226 | 0.062 |
| **4** | Adverse Conditions (Ice, Snow, Slush etc.) | 14 | 0.004 |
| **5** | Other (Unknown or Not Reported) | 177 | 0.049 |
| **Total** |   | 3644 | 1 |

**Variable X2:** Weather conditions at the time of the crash

|  |  |  |  |
| --- | --- | --- | --- |
| **X2=x2** | **Description** | **N[X2=x2]** | **Pr[X2=x2]** |
| **1** | Clear | 2496 | 0.685 |
| **2** | Partly cloudy | 659 | 0.181 |
| **3** | Cloudy | 230 | 0.063 |
| **4** | Rain | 81 | 0.022 |
| **5** | Adverse Conditions (Fog, Mist, Snow, Severe Winds etc.) | 71 | 0.019 |
| **6** | Other (Unknown or Not Reported) | 107 | 0.029 |
| **Total** |   | 3644 | 1 |

**Variable X3:** Recording whether a crash happened on a weekday or weekend

|  |  |  |  |
| --- | --- | --- | --- |
| **X3=x3** | **Description** | **N[X3=x3]** | **Pr[X3=x3]** |
| **0** | Weekday | 2067 | 0.567 |
| **1** | Weekend | 1577 | 0.433 |
| **Total** |   | 3644 | 1 |

The next step was to calculate the joint and conditional probabilities of the variable Y associated with the variables X1, X2, and X3.

This process revealed the following results, in terms of joint probabilities, which are given in a graphical format.

With the knowledge of the joint probabilities between the Y variable and the X variables, we can now calculate the conditional probabilities for variable Y associated with the different cases (values) of each of the Xs variables.

In order to find the conditional probabilities we used the simple formula:



In the following tables we present the results of this process

|  |  |
| --- | --- |
| **Pr[Y=y|X1=x1]** | **Surface Conditions X1** |
| **1** | **2** | **3** | **4** | **5** |
| **1** | 0.056 | 0.035 | 0.013 | 0.000 | 0.045 |
| **2** | 0.260 | 0.184 | 0.142 | 0.214 | 0.153 |
| **3** | 0.410 | 0.404 | 0.562 | 0.357 | 0.333 |
| **4** | 0.194 | 0.220 | 0.186 | 0.286 | 0.153 |
| 5 | 0.079 | 0.156 | 0.097 | 0.143 | 0.316 |

|  |  |
| --- | --- |
|  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|

|  |  |
| --- | --- |
| **Pr[Y=y|X2=x2]** | **Weather Conditions X2** |
| **1** | **2** | **3** | **4** | **5** | **6** |
| **1** | 0.049 | 0.058 | 0.057 | 0.025 | 0.056 | 0.093 |
| **2** | 0.243 | 0.261 | 0.291 | 0.123 | 0.268 | 0.150 |
| **3** | 0.424 | 0.425 | 0.400 | 0.469 | 0.239 | 0.262 |
| **4** | 0.200 | 0.184 | 0.170 | 0.210 | 0.225 | 0.121 |
| **5** | 0.084 | 0.073 | 0.083 | 0.173 | 0.211 | 0.374 |

 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **Pr[Y=y|X3=x3]** | **WEEKDAY VS WEEKEND** |
| **0** | **1** |
| **1** | 0.045 | 0.062 |
| **2** | 0.237 | 0.254 |
| **3** | 0.420 | 0.410 |
| **4** | 0.195 | 0.190 |
| **5** | 0.103 | 0.084 |