<u>Bonus Project</u>

<u>Sana Fathima</u>

<u>Motorcycle Fatalities</u>

<u>11/27/2015</u>

Overview:

It is said that the number one cause of death among young people between the ages of 15-29 is motor vehicle accidents. Nearly 44,868 vehicles were involved in fatal crashes in 2013. Roughly 10.6% of all these accidents came from motorcyclists' related collisions.

For this project, I have gathered information from a "Fatality Analysis Reporting System" (FARS) database about the number of fatal motorcycle accidents that occurred in the year 2013. After splitting the data up based on timing I was able to get a good representation of the amount of crashes in a certain time period.

Times	n	2013(%)	Num Of Crashes
12:00AM-5:59AM	4668	13.4	625
6:00AM-11:59PM	4668	15.9	738
12:00PM-5:59PM	4668	36.8	1716
6:00PM-11:59PM	4668	33.5	1562

This data is representation of total number of crashes that occurred during that year, and the percentages based on timings, and the total number of crashes during that time. However, in the data there were 27 crashes where the timing was unknown, this might have been because of bad reporting. Even though these 27 crashes were unknown, it is still incorporated in the number of fatal crashes(n), because motorcyclists' deaths still occurred.

My goal for this project was to test my hypotheses on the percentages that I had pre-determined. I believe that 12:00AM-5:59AM timing is one of the timings with a greater amount of fatal crashes. I predict that at least 20% of all deaths occur between midnight and 6:00AM, and I will test this hypothesis using a α value of .05.

1. Parameter of Interest: p
2. Null Hypothesis: $H_0 : p = .20$
3. Alternative Hypothesis: $H_1 : p > .20$

   $$\begin{cases} H_0 : p = .20 \\ H_1 : p > .20 \end{cases}$$

4. Test statistic: $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$
5. Reject H_0 if $Z_0 > Z_\alpha$
6. Computations:

   $$Z_0 = \frac{.134 - .20}{\sqrt{\dfrac{.20(1-.20)}{4668}}} = -11.27$$

7. Since -11.27 < 1.65 we cannot reject the null hypothesis. One thing to note is that when we accept the null hypothesis this is a weak conclusion. So we cannot fully say that p= .20, rather than we know for sure that p is not greater than .20; however it could be .20 or a less than that as well.

8. We can also calculate the p-value: $1-\phi(-11.27)=1-0=1$. This p-value>.05 and therefore we cannot reject the null hypothesis. This again agrees with our first computation above.

I will also compute a 100(1-z)% CI for a Population Proportion with a 95% confidence interval of this data to see what interval contains the **true** population proportion.

Using this formula: $\hat{p}-Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p}+Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ I will be able to compute the confidence interval.

$$.134-1.96\sqrt{\frac{.134(1-.134)}{4668}} \le p \le .134+1.96\sqrt{\frac{.134(1-.134)}{4668}}$$

$.124 \le \text{p} \le .144$

Given a confidence interval of p $\in$[.124, .144] we know that there is a 95% chance that the confidence interval calculated contains the true population parameter. The upper and lower bounds are about $\pm$.0098 from the true proportion of 13.4% or .134.

There is a weak conclusion that we accept the null hypothesis. So again the actual data could be equal to .20 or less. Regardless of the data, we see that there are not as many fatal crashes in the early morning hours.

Secondly, I believe that 12:00PM-5:59 PM timing is one of the timings with least amount of fatal crashes. I predict that less than 15% of all deaths occur between 12:00PM-5:59PM, and I will test this hypothesis using a α value of .05.

1. Parameter of Interest: p
2. Null Hypothesis: $p=.15$
3. Alternative Hypothesis: $p<.15$

$$\begin{cases} H_0 : p=.15 \\ H_1 : p<.15 \end{cases}$$

4. Test statistic: $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$
5. Reject H_0 if $Z_0 < -Z_\alpha$
6. Computations:

   $$Z_0 = \frac{.368 - .15}{\sqrt{\frac{.15(1-.15)}{4668}}} = 41.71$$
7. Since 41.71 > -1.65 we cannot reject the null hypothesis. Another thing to note is that when we accept the null hypothesis this is a weak conclusion. So we cannot fully say that p= .15, rather than we know for sure that p is not less than .15; however it could be .15 or a greater than that as well.
8. We can also calculate the p-value: $\phi(41.71)$ =1. This p-value>.05 and therefore we cannot reject the null hypothesis. This again agrees with our first computation above.

I will also compute a 100(1-z)% CI for a Population Proportion with a 95% confidence interval of this data to see what interval contains the **true** population proportion.

Using this formula: $\hat{p} - Z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ I will be able to compute the confidence interval.

$$.368 - 1.96\sqrt{\frac{.368(1-.368)}{4668}} \leq p \leq .368 + 1.96\sqrt{\frac{.368(1-.368)}{4668}}$$

$.354 \leq p \leq .382$

Given a confidence interval of p $\in$[.354, .382] we know that there is a 95% chance that the confidence interval calculated contains the true population parameter. The upper and lower bounds are about ±.01383 from the true proportion of 36.8% or .368.

Conclusion:

After computing the hypothesis testing, we can see that both my predictions were wrong. My reasoning behind these initial predictions is that during the nighttime it seems more dangerous for riders to be riding their motorcycle as some streets might not be light or the riders might be wearing dark clothing causing it to be hard to see them. Another reason I thought, that the early morning hours would have more fatal crashes is because some people might be drunk or have lack of sleep causing them to have accidents. I also thought that the 12:00PM-5:59PM times would have a fewer amount of crashes because again there is daylight and people are generally more awake and alert.

However this is not the case; after looking at the actual data and reviewing the hypothesis testing, we can see that there are more accidents that occur between 12:00PM-5:59PM. The main reason I believe that makes this statement true is that there are more cars on the road due to rush hour traffic between these times causing a higher probability of a crash occurring. This can be backed up by the data given from the guardian news source which states that "more fatalities occurred during the hours of 1pm and 7pm, the peak being 6.6% fatalities at 3pm". Also during the later half of this time period, more people might be rushing home and not paying attention as much causing more accidents. Also we see that the least amount of fatal crashes occur during the time period of 12:00AM-5:59AM. One explanation for this data is that because there is less riders on the road so there is a smaller probability that one of these riders will get in an accident. Again, after examining the data the time period of 12:00AM-5:59AM has the least amount of accidents

One way to make the data more representative is to maybe break it up into data from the 50 states, that way we will see if one state is skewing the results up or down. Also, it would be useful to look at history from about 10 years ago and see If on average the amount of deaths per time period increased or decreased.

Resources:

http://www.theguardian.com/news/datablog/interactive/2013/oct/27/interactive-safest-time-to-drive

http://www-fars.nhtsa.dot.gov/People/PeopleMotorcyclists.aspx