# Highway Fatal Crash Statistics Data Analysis Project

I am conducting a series of scientific research projects looking at the effects of Mercury Retrograde. When considering issues such as on-time flight departures, the model is simple enough for me to run the numbers in Tableau, and the statistical analysis only involves evaluating the difference between two percentages. When considering data such as car crashes, I need a model that evaluates standard deviations and can provide the p value to determine the statistical significance. As I am not fluent in statistical analysis, this is far beyond my abilities.

I was working with another freelancer on this project, and while she was able to help me with other aspects of the data analysis, she was not able to properly model this data the way that I asked her to.

There are multiple data sets, and multiple permutations of the data, but the model for each analysis is consistent.

I need to be able to determine the number of standard deviations above or below the mean for each variable (number of crashes, number of vehicles, number of fatalities) for nine different speeds of Mercury. I also need to know the p value for each of these measurements to determine the statistical significance of the deviations.

I’m given to understand that p values are not easy to calculate in Tableau, so this may have to be an R project. I don’t need software to visualize the data — I need the results tables (in a format that I can easily import and edit in Excel). I will create the charts myself.

### Sample Demonstration

To be considered for this project, you will need to generate my required output for one data table (there are a total of 27 tables in this project). This is the only way that I can determine if a candidate is, in fact, capable of doing what I ask.

I’ve attached an Excel file with all of the data for this project. I would like the following analysis performed:

Using data from 1995 – 2004 ONLY (Average of 102–105 crashes per day)

Analyze the standard deviation from the mean for each of the nine Mercury speeds. Sort the speeds in this order:

xFD, FD, SD, xSDr, xSRr, SR, FR, xSRd, xSDd

There is great disparity in the number of days in each of these speeds, so you will need to evaluate the deviation from the mean for each speed from the deviation from the mean of the entire data set.

And let me be explicit about this, because this is where my original subcontractor failed:

**I need the reference mean to be the mean of the ENTIRE DATA SET**.

You cannot use one of the Mercury speeds as the reference point (which is what she did for a linear regression model — this is not what I need).

I need to evaluate **each speed** and know the standard deviation above or below the mean for the variable being considered (in this case, just the number of crashes). Each record/date in the database is a single crash incident.

I need the output as a .CSV file or Excel file in table format with the following data:

1. Mercury Speed

2. Standard Deviation (+ or -)

3. t Value

4. p Value

Use a 95% confidence level for the p value.

I do not expect there to be any statistically significant results in this table — however, I need to know that you’re able to model the data correctly and provide me with accurate results in a format that I can understand and analyze.