STAT 3320

Date:**4 / 13 / 2017**

Title: **Relation of Gas Price to Proximity to Other Fueling Stations**

Project Group Members:

**Stephanie Loya (sloya4@miners.utep.edu)**

**Stephanie Figueroa (sfigueroa2@miners.utep.edu)**

**Victor Fernandez(vmfernandez4@miners.utep.edu)**

**Joseph Gutierrez (jagutierrez21@miners.utep.edu)**

The above students have submitted a proposal to conduct this research as part of their Stat 3320 class requirements.

If you have any questions, please do not hesitate to contact: (choose your instructor)

Mr. Desmond Koomson at dkoomson@utep.edu or 915-747-7004

**Corrections to proposal are needed: YES / NO** \_\_\_\_\_ (instructor initial)

Proposal Approved By: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, on \_\_\_\_ / \_\_\_\_ / \_\_\_\_

Department of Mathematical Sciences

University of Texas at El Paso

# 

# **STAT 3320 Proposal Summary**

|  |  |
| --- | --- |
| **Title** | Relation of Gas Price to Proximity to Other Fueling Stations |
| **Research Question** | Is the proximity to other gasoline stations a factor in thesale price of gasoline at that location? |
| **Population** | This observational study will be conducted on a statewide sample, including those in different states close enough in proximity to a given station. This provides a more accurate representative value of gasoline stations that are near each other and further away. |
| **Sampling Method** | This is a convenience sample because we will consider data for any gasoline station that we can find within the defined area or limit inside the state of choice. |
| **Main Variables** | **Explanatory:**  A gas station’s proximity to other gas stations  **Response:**  The price of gas at that particular location |
| **Secondary Variables** | Because we are averaging all of the values, there are no secondary variables. |
| **Lurking Variables** | Stations that claim to provide better quality gas, do so at a premium. Stations that are parts of chains have more buying power. Club stores sell gas for cheaper |
| **Statistical Method** | We will collect average values for price over a specified time and the average distance to the station’s five closest gas stations. Then attain the mean, median, standard deviations, and perform a simple linear regression |
| **Select one Lurking**  **Variable and explain how it will be controlled** | In regard to gasoline stations offering better quality fuel, we will only consider the lowest grade fuel they offer. The governing organizations that regulate and control gasoline sales, have strict guidelines on the lowest grade of fuel that can be commercially sold. So this should ensure more uniformity, rather than looking at higher end fuel. |
| **Justify if sample represents population** | By using a state, it is a sample of convenience; therefore, it should not be an approximately representative value for the trends in other states or trends in the United States as a whole. |

**STAT 3320 Research Proposal**

**Relation of Gas Price to Proximity to Other Fueling Stations**

**Statement of the Problem:**

The purpose of this project is to examine the linear association of gas prices and the proximity to other gas stations. In theory, a gas station that is closer to others has more competition so it would need to keep its prices lower in order to directly compete with its neighboring stations. On the other hand though, an argument could be made that because stations that are more remote are generally located closer to highways and more commercially accessible roadways, that the cost of getting gasoline to these locations is lower and therefore they should have the lower prices. Either way, this study intends to find whether or not this correlation exists.

**Significance:**

This research will help us understand how gas prices are different from one another depending on how many other gas stations are around. This is helpful to anyone who is trying to save more money on gas. There’s a possibility that a gas station that is surrounded by more gas stations has better prices, because they are constantly competing with those in the area, than a gas station that has no surrounding gas stations within miles. Being able to tell where gas prices are generally cheaper based solely on the gas station’s proximity to others is useful for anyone in the community.

**Background:**

For anyone who has to commute to work or school, the price of gasoline is a major and very influential factor in their finances. This study tries to focus on the underlying reasons for fluctuations in gas prices. Location is a hugely important factor for any kind of business. It stands to reason, therefore, that gasoline prices should also be based on the location from which it is being sold.

**Population & Sampling:**

This observational study will be conducted on a statewide sample and its surrounding areas in order to get a good representative value of gasoline stations that are near each other and further away. In order to attain a good representative sample, we will consider data for any gasoline station that we can find within the defined area or limits. The price of gas differs greatly from one part of the country to another, for this reason, we are choosing a more local and relevant measure. This should be able to give a generalized idea for average gas prices in areas with similar public transportation and rural and urban areas. Due to this change in gas prices from region to region, we will be using a sample of convenience and only considering a single state.

**Independent Variable:**

The quantitative average distance of a fueling station to its five nearest neighboring fueling stations measured in kilometers. This value is as an average of the nearest five because most small towns and or rest stops have at least two fueling stations to avoid monopolization. By taking the average distance from several fueling stations we can better gauge how close it is to other fueling stations. In this way, averaging also helps us to ensure that we get representative values for each data point.

**Dependent Variable:**

The quantitative average price, over several days, of lowest grade fuel for that fueling station. This variable is based on the lowest grade fuel because by law, the lowest grade fuel that is commercially available is E85, however, many high end fueling chains offer very expensive fuel that contains additives that raise the price of fuel in an uneven way. It is based on the an average over a series of days because gasoline prices can fluctuate several times in a given day and even more over several days. Taking an average of a set time frame ensures that we get a representative value for the data.

**Lurking Variables:**

Gas station companies may create a whole host of lurking variables. Some companies claim to sell better quality gas than others, therefore their prices may be higher. For this reason, we are only considering the lowest grade gasoline available. Some companies that belong to large chains, like Walmart, have more buying power and can provide cheaper gas. These stations are often closer to other stations in more populous areas and will have similar distances between them. Additionally, some gas stations belong to “club stores”, where you must purchase a membership in order to buy their gas, like Sam’s or Costco. For our purposes, we will completely disregard these locations because their prices are likely to be independent of distance to other gas stations. A final lurking variable would be the amount of public transportation available in a city of interest. If there public transportation is more readily available, people are more likely to use it and this makes distance to fueling stations not as pertinent as a factor.

**Data Analysis Strategies:**

For this particular application, much of the data gathering will be done computationally by programs specifically designed for this study. In order to arrive at a final answer, the following will occur.

1. A sample space (geographic area of interest) will be defined, with help from the professor.
2. All gasoline stations within the bounds of the area of interest will be identified and their geographic location will be found (in units of latitude and longitude).
3. An algorithm will determine how far away each gas station is from its 5 nearest neighbors. (For this step, other gasoline stations outside of our area of interest may be considered for fuel stations on the outskirts of the bounds, but no further data will be collected for them, they will only be used as indicators of proximity for the fuel stations we are considering)
4. We will remove invalid gasoline stations from our “watch list” set. (Club stores or any provider of gasoline that has an agreement with its patrons that is either exclusive or based on a previous transaction.)
5. The prices of the gasoline stations still being considered in the “watch list” will be recorded over a span of days.
6. For any fuel stations in our valid data set, we will compute the simple average for the distance to its five closest neighbors and the simple average value for the price of gasoline then input it into a table similar to the one found in the appendix.
7. The table’s data will be analyzed for trends, using representative values, like mean, median, standard deviation, variance, and quartiles. We will also attempt to perform a simple linear regression on the data in order to be able to come up with a generalized formula to predict average gas prices based on average distance.

|  |  |  |  |
| --- | --- | --- | --- |
| Gas Station ID | Average Price of Gasoline | Average Distance from 5 Nearest Neighbors | Ratio of Average Price to Average Distance |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |