Statistical Analysis of Automotive Fuel Economy

Written By: George Camac

Class: CSCI-3327- Probability and Applied Statistics

Professor: Byron Hoy

Table Of Content

1. Introduction (3)
2. Set Theory Applications (3-6)
   1. Chapter 2.3 page 3-4
   2. Chapter 2.4 page 4-5
   3. Chapter 2.5 page 5-6
3. Conditional Probability and Techniques (6-8)
   1. Chapter 2.6 page 6-7
   2. Chapter 2.7 page 7-8
4. Law of Total Probability and Bayes (8-9)
   1. Chapter 2.10 page 8-9
   2. Bayes Theorem 9
5. Random Variables and Distributions(10-14)
   1. Chapter 3.2 page 10
   2. Chapter 3.4 page 10-11
   3. Chapter 3.5 page 11
   4. Chapter 3.6 page 11-12
   5. Chapter 3.7 page 12-13
   6. Chapter 3.8 page 13-14
6. Continuous Random Variable Distributions(14-17)
   1. Chapter 4.2-4.4 page 14-15
   2. Chapter 4.6 page 15-16
   3. Chapter 4.10 page 16-17
7. Bivariate and Multivariate Distributions(17-18)
   1. Chapter 5.2-5.3 page 17-18
8. Conclusion

In my report I go and dive into a statistical analysis of automotive fuel economy specifically from cars the 1970s to the 1980s. Throughout the report you will find an array of different methods to try and find different statistics are cars, whether its for cars you are buying or just to get the information. By using different Statistical techniques we are able to uncover interesting statistics about cars, and it can definitely help when you are in the market for a car.

In todays day of age fuel economy is a very important thing as it impacts the environment and also importantly it impacts our wallets. It’s as simple as the less efficient a car is the more exhaust, and more money will be spent on that car. And with today’s gas prices we need to try and keep our usage to a minimum. The data set were looking at has 398 total cars, all of different make. However, we can derive information from it to observe some interesting information.

Whenever you’re buying a car you need to make sure you check the fuel economy of the car you want to buy as it’s important to know what you are getting yourself into. Whether you are checking the fuel economy for the purpose of trying to help the environment or in cases of trying to save money on the amount of gas used it is very important to check the statistics of the fuel you’re using. For Example…

SECTION 2.3

**Question:**

**Show using set notation. When there are 151 cars with low fuel efficiency, or less than 20 mpg and 92 cars with high fuel efficiency, or 30 or more mpg in a sample space of 398 cars.**

**A black and white rectangular with black circles and letters

Description automatically generatedSolution:**

**With cars with less than 20 mpg**

**cars with more than 30 mpg**

**the sample space of 398 cars**

**Question:**

**After separating cars into these categories, we found that a significant number of cars were unaccounted for. Find C which are the cars that have 20 mpg or better but have less than 30 mpg.**

**Set Union**

A yellow circles with black letters

Description automatically generated **Solution:**

Section 2.4

Now you can also find more information by comparing the horsepower to car cylinders. While some cars will have a higher horsepower output even though they less cylinders in their engine. There are 198 cars with 6 or more cylinders, and 250 cars with more than 100 horsepower. Compare the differences between…

**Question: Show this using Venn Diagram**

**Solution:**

**A diagram of a number and a number

Description automatically generated with medium confidence**

Section 2.5

**This shows that there are 18 cars with 6 cylinders that don’t have more than 100 horsepower as well as 70 cars with less than 6 cylinders have more than 100 horsepower. But we find 180 have both 100 horsepower and more than 6 cylinders**

**Question:**

Now if were looking to rent cars and there are 5 different cars in the lot. We need 2 cars from the lot, but they have 2 with less than 20 mpg which we consider to be defective due to the low efficiency. The cars are given at random, so what is the probability that we get 2 non-defective cars.

1. Show the sample space
2. Show Events
3. Find A, or the events where you will get 2 non-defective cars
4. Draw A Venn Diagram representing A
5. Find the probability of A

**Solution:**

1. ,
2. A green circle with black dots

   Description automatically generated

Showing you have 3/10 chance to get both cars with 20 or more mpg

Now when you are going to buy and check out cars you may want to test drive those cars since it is important to feel how to car actually drives before leaving the lot with it. This is extremely important especially if the car is a used car. But most dealerships won’t let you test drive as many cars as you want so its useful to know how many different ways you could use the allotted test-drives.

**SECTION 2.6**

**Question:**

**If you go to a dealership and want to test drive cars, the dealership only allows you to go on the 3 test drives. How many ways could you use these test drives, when you are given the selection of 6 cars?**

**Solution:**

**Showing there are 120 ways for you to test drive 6 cars with 3 test drives**

Now this method doesn’t give you the most accurate data, unless you plan on driving the same car multiple times. However, it could still be useful here we would be more likely to use the combinational method of this. For example…

**Question:**

**If you go to a dealership and want to test drive cars, the dealership only allows you to go on the 3 test drives. How many different ways could you use these test drives if order doesn’t matter, when you are given the selection of 6 cars?**

**Solution:**

So now you are able to figure out all the different ways you can test drive the cars. Time to see what the probability is that one of the cars you chose, if chosen at random, is fuel efficient, or 30 or more mpg.

SECTION 2.7

**Question:**

**If you go to a dealership and they have 398 cars available in the lot, 92 cars are fuel efficient 151 cars are inefficient. What is the probability of a car on the lot being fuel efficient, given that it isn’t classified as inefficient? How does this compare to the probability of getting a fuel efficient car of any given car on the lot?**

**Solution:**

So if you know you wont be shown a inefficient car you have about a 37.24% chance of finding a fuel efficient car. Which compared to the probability of any car on the lot including the inefficient cars you have about a 23.12% chance of finding a fuel-efficient car. This will help you to realize that you may want to talk to the dealership before going to look at cars. Now we can also use similar information to find if horsepower is directly dependent on if the car if fuel efficient.

**Question**

**Determine if a car’s fuel efficient decides the horsepower of the car. There are 250 cars with over 100 horsepower of 398. There are 92 cars that are fuel efficient, 60 of which are above 100 horsepower. Determine if these events are dependent or independent.**

**Solution:**

**This shows that these two events are independent**

Section 2.10

**Question:**

At a dealership they have 398 cars 149 of which are imported and 249 are domestic. Of those imported cars 32 are fuel efficient, while there are 60 fuel efficient domestic cars. What is the probability that a randomly selected car has 30 or more mpg, using the law of total probability?

**Solution:**

This can be helpful when finding the probability of getting aa fuel efficient car when given multiple groups whether it be used for things like domestic vs imported, or it could even be correlated to things like color of car.

Now similarly, you can also try to figure out the probability that the car is imported or domestic.If you go to the same dealership and you are shown a car but the only information you know about the car is that is it is fuel efficient, but you know how many fuel efficient cars are imported vs domestic you can have better guess if your car is imported or domestic

**Question:**

**You go to a dealership with 32 highly fuel efficient imported cars out of 149 imported and there. As well as 60 highly fuel efficient domestic cars out of 249. If you are given a car from the dealership with high fuel efficiency, what is the probability that it is imported.**

**Solution:**

**This shows that there is about a 65.2% chance the car is imported if it is fuel efficient**

Section 3.2  
Now you can also calculate when you are given a random car what the average probability of getting one of the 3 categories, inefficient fuel consumption, moderate consumption, and fuel efficient. This is done by using the random variable and finding the variance and expected values.

**Question:**

**You go to get a car from your local dealership, they tell you that they have 3 categories of cars. Category 1 is inefficient fuel consumption which is 38% of the cars at the dealership. Category 2 is moderate fuel efficiency which is 39% of the cars, lastly Category 3 is fuel efficient cars that make up 23% of the cars on the lot. When choosing a car what is the expected Category, as well as the spread?**

**Solution:**

**The expected value represents between 1-3 categories where does a randomly car selected fall on average. The Variance is the area in which the data can spread.**

Section 3.4

Another important thing to consider is when going somewhere and the show you a specific number of cars what is the probability they are desirable. For example..

**Question:**

You go to the car dealership and they select 10 cars for you to look at. 40% of the lot are fuel efficient. What is the probability that they select exactly 4 cars out of 10. What is the probability there are 6 or more cars that are fuel efficient.

**You will need**

**Solution:**

Section 3.5

Now what if you are trying to find how many iterations it will take before you find a car with fuel efficiency. So if the probability of finding a fuel efficient car is known you can find the probability you will find it on whatever iteration you want.

**Question:**

**You go to a dealership that has a probability of getting a fuel-efficient car of 23%. If the dealer shows you cars one by one what is the probability of finding a fuel-efficient car on the 4th car. Also, what is the expected number of cars the customer will need to see before finding a fuel-efficient car.**

**Solution:**

**This shows the probability a car shows on exactly the 4th car**

**This shows the expected number of cars selected before finding one that is fuel efficient**

Section 3.6

We are also able to use negative Binomials to generate information on if we want to find the probability of multiple events. Like this…

**Question:**

**You go to a dealership that has a probability of getting a fuel-efficient car of 23%. If the dealer shows you cars one by one what is the probability you find the 3rd fuel-efficient car on the 7th car shown? What is the expected number of cars the dealer needs to show to find 3 fuel efficient cars.**

**Solution:**

**You need**

**This means you would have a 6.43% chance of finding the 3rd fuel efficient car on the 7th car shown to you by the dealer.**

Section 3.7

Now lets say you are working for a company and you need to put a work order in for a large amount of cars at once. You may want to know the probability of getting a specific number of fuel-efficient cars. This may help you find out also if the company is purposely sending you non-efficient cars. For example,

**Question:**

**You work for a large company and the company has tasked you with buying 20 cars from a specific dealership that has 92 fuel-efficient cars and 306 inefficient cars. The company wants 5 fuel efficient cars but told you to randomly choose the cars. However, they do want to know if the dealership is purposely sending them inefficient cars. So What is the probability that after ordering 20 cars exactly 5 of them are efficient. And what is the expected number of fuel-efficient cars.**

**Solution:**

**You need**

**This means you have a 20.37% chance to get 5 fuel-efficient cars when ordering 20 cars**

**Showing us if you order 20 cars the average you get is 4.6**

Section 3.8

Another way of finding the probability of average vs desired would be with Poisson Distribution. Here’s an example using the dataset.

**Question:**

**The dataset indicates there is 398 cars and 249 of which are domestic cars. On average a random selection of 50 cars contains 31 cars that are domestic cars. What is the probability that exactly 35 of 50 cars are domestic? What is the probability that 25 or less cars of domestic?**

**Solution:**

**You need**

**Showing you have a 5.25% chance of getting 35 domestic cars out of 50**

**Showing that you have a 16.15% chance of getting 25 or less domestic cars out of 50**

Section 4.2-4.4

Another big factor in deciding whether or not the car is for you could be dependent on weight of the car which also effects emissions. For example, some companies may have a weight limit for vehicles if the vehicles need to be transported to another country.

**Question:**

**From our dataset we see that there is a range of weight from 1,613 lbs. to 5,140 lbs. Now assume that of these 398 cars the weigh is spread amongst them from 1,613 to 5,140.**

1. **Write a probability density function for the weight of a randomly selected car?**
2. **What is the probability that a randomly selected car weighs between 2,500 lbs. and 3,500 lbs.?**
3. **What is the expected weight of the randomly selected car?**
4. **What is the variance of the car wights**

**Solution:**

**For PDF you need**



Section 4.6

Sometimes you are told that cars have a certain lifespan, but as we know for humans too not all cars have a predetermined lifespan. Here in this example I will show you how you can use it to find the probability of how long the car can go without repair.

**Question:**

**When you go to a car dealership they tell you that average time it takes for their cars to need repair is 5 years. This follows an exponential distribution with a scale parameter of**

1. **Write the PDF for the time until a repair**
2. **What is the probability that a car requires repair in less than 3 years**
3. **What is the expected time and variance for the time until repair**

**Answer:**

**Using**

1. **PDF**
2. **CDF**

**showing a variance of 25 years and expected as 5 year**

Section 4.10

**Question:**

**When you go to the car dealership they give you the information about the weight of the vehicles. You find in the data set the average weight of the cars is 3,376.5 lbs. and the standard deviation is 1036.4 lbs. Using Tchebysheffs theorem find what the minimum proportion of weights that lie within 2 standard deviations? Calculate the range of weights in this interval.**

**Solution:**

**Use**

**saying at least 75% of the cars weigh within 2 standard deviations of the mean**

**,**

**showing a range of**

Section 5.2-5.3

**Question:**

**The data set characterizes cylinders and origin. We want you to assume these joint probabilities from the data set.**

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

1. **Calculate marginal probability**
2. **Calculate F(4,1) the cumulative probability**
3. **Show the conditional probability**

**Solution:**

**This shows that there is 60% of the cars have 4 cylinders and a 40% the car has 6**



**Showing there is a 35% chance its domestic if it has 4 cylinders**

**While analyzing automotive fuel economy with different statistical methods we were able to come across multiple different key factors when it comes to purchasing cars with this in mind. We focus mainly on weight, fuel efficiency, origins, and cylinders. This report is meant to demonstrate how you can use what you have learned in statistics to gather information with your statistical analysis. Understanding these concepts is important when it comes to buying a car, whether you’re looking at the likelihood for the need of a repair, or if you really are trying to protect the environment with putting less fumes with a more efficient car. Be sure to remember to check and run your statistics for all different kinds of aspects of cars and you will learn more about them extremely quickly.**