**2017 Subaru Impreza Fuelly App Data**

**Joseph Morales**

**Department of Computer Science, Stockton University**

**CSCI-3327: Probability and Applied Statistics**

**Professor Byron**

**December 11, 2024**

**Table of Contents**

* Chapter 1:
  + Page 6: Question 1.4 – Analyzing fuel efficiency records (MPG) for a Subaru Impreza Sport.
* Chapter 2:
  + Page 26: Question 2.8 – Fuel-ups and gas station brand analysis.
  + Page 32: Question 2.11 – MPG probability ranges and events.
  + Page 39: Question 2.28 – Selecting fuel-ups with high or low MPG.
  + Page 48: Question 2.35 – Combining fuel-ups from two regions.
  + Page 60: Question 2.94 – Union and complement probabilities for MPG ranges.
  + Page 73: Question 2.124 – Conditional probability for MPG exceeding 25 based on region.
  + Page 85: Question 2.179 – Comparing costs between two fuel-ups using binomial probability.
* Chapter 3:
  + Page 90: Question 3.1 – Distribution for conditions met by fuel-ups.
  + Page 119: Question 3.67 – Geometric probability for finding a specific fuel-up.
  + Page 123: Question 3.90 – Negative binomial probability for high-MPG fuel-ups.
  + Page 129: Question 3.104 – Hypergeometric probability for fuel packet selection.
  + Page 136: Question 3.128 – Poisson probability for cars achieving MPG > 20.
  + Page 148: Question 3.170 – Tchebysheff’s theorem for fuel-up costs.
* Chapter 4:
  + Page 167: Question 4.11 – Probability density and cumulative distribution for gallons.
  + Page 177: Question 4.49 – Random vehicle arrivals at a gas station.
  + Page 191: Question 4.94 – Exponential distribution for MPG performance.
* Chapter 5:
  + Page 233: Question 5.9 – Joint distribution for MPG and Cost/Gallon.
  + Page 243: Question 5.23 – Proportions of tank stocked and used.

**Page 6 Chapter 1**

**Question 1.4**

The dataset includes fuel efficiency records (MPG) for a Subaru Impreza Sport, that were recorded across multiple fuel-ups

Below are the first 40 MPG values recorded  
0.0, 18.7, 19.6, 18.4, 22.4, 23.0, 15.8, 20.3, 17.6, 21.1, 18.9, 19.3, 22.5, 24.2, 20.8, 19.7, 21.5, 23.1, 16.4, 15.6, 17.9, 20.6, 22.7, 18.3, 19.2, 21.9, 24.3, 19.0, 23.4, 20.0, 15.2, 16.7, 17.5, 19.8, 20.1, 22.8, 18.6, 21.3, 20.4, 19.5

1. Construct a relative frequency histogram to describe this data?
2. What proportion of fuel-ups recorded an MPG greater than 20 MPG?
3. If one Fuel-Up is selected randomly, what is the probability that its MPG was less than 25 MPG?

Solution

1. Construct a relative frequency histogram to describe this data?
2. What proportion of fuel-ups recorded an MPG greater than 20 MPG?

From the data set 24 out of the 40 fuel-ups recorded where above 20 MPG

24/40 = 0.6 = 60%

60% of fuel-ups recorded are likely to be over 20 MPG

1. If one Fuel-Up is selected randomly, what is the probability that its MPG was less than 25 MPG?

From the data set 39 out of the 40 fuel-ups recorded were less than 25 MPG

39/40 = 0.975 or 97%

The probability of randomly selecting a fuel-up with MPG less than 25 is 97.5%. \_

Report

While analyzing the MPG data for my Subaru Impreza Sport, I started by creating a relative frequency histogram to get visual representation of how the MPG values were distributed across 40 fuel-ups. Most of the trips seemed to fall in the 15-25 MPG range, with a few outliers. After that, I calculated that 60% of the fuel-ups had an MPG greater than 20, which shows that the car performs well on the majority of trips. Finally, I figured out that there was a 97.5% chance of randomly picking a fuel-up where the MPG was below 25. Overall, this analysis gave me a better understanding of my car's efficiency and how often it performs within certain range

**Page 26 Chapter 2**

**Question 2.8**

From a dataset of 60 recorded fuel-ups for a Subaru Impreza Sport:

9 fuel-ups were recorded at gas stations branded as "Brand A."

36 fuel-ups achieved MPG greater than 20.

3 fuel-ups were both from "Brand A" and achieved MPG greater than 20.

Using this data, determine:

1. The number of fuel-ups that were from "Brand A," achieved MPG greater than 20, or both.
2. The number of fuel-ups that were **not** from "Brand A" but achieved MPG greater than 20.
3. The number of fuel-ups that were from "Brand A” did not achieve MPG greater than 20.

Solution

1. The number of fuel-ups that were from "Brand A," achieved MPG greater than 20, or both.

n(A) = 9: Fuel-Ups from “Brand A”

n(B) = B: Fuel-Ups with MPG > 20

n(A∩B) = 3: Fuel-ups that are both from "Brand A" and have MPG > 20

42 fuel-ups were from "Brand A," had MPG > 20, or both

1. The number of fuel-ups that were not from "Brand A" but achieved MPG greater than 20.

33 fuel-ups had the MPG > 20 but were not from “Brand A”

1. The number of fuel-ups that were from "Brand A” did not achieve MPG greater than 20.

6 fuel-ups were from “Brand A” but did not have the MPG > 20

Report

While analyzing the fuel-up data for my Subaru Impreza Sport, I looked at the connection between fuel efficiency and gas station brand. Using the inclusion-exclusion principle, I found that 42 fuel-ups either happened at "Brand A," had more than 20 MPG, or both. I also found that 33 fuel-ups had more than 20 MPG but were not from "Brand A," which shows that the car maintains better efficiency across different gas brands. Lastly, I determined that 6 fuel-ups from "Brand A" had an MPG below 20. This analysis helped me better understand how brand choice and MPG performance align and provided a better view on my car’s overall fuel efficiency trends.

**Page 32 Chapter 2**  
**Question 2.11**

The dataset includes records of MPG values for a Subaru Impreza Sport. Each fuel-up is classified into one of five events based on MPG ranges:

E1: MPG ≤15

E2: 15 < MPG ≤ 20

E3: 20 < MPG≤ 25

E4: MPG > 25

E5: MPG = 0

1. If P(E1) = P(E2) = 0.15, P(E3) = 0.4 and P(E4) = 2P(E5) , find the probabilities of E4 and E5.
2. If P(E1) = 3P(E2) = 0. find the probability of the remaining events (E3, E4, E5) assuming the remaining events are equally probable.

Solution

1. If P(E1) = P(E2) = 0.15, P(E3) = 0.4 and P(E4) = 2P(E5) , find the probabilities of E4 and E5.

Given that

P(E1) = P(E2) = 0.15

P(E3) = 0.4

P(E4) = 2P(E5)

Also

The probability that my MPG will be over 25 is 20%

The probability that my MPG will be 0 (Missed Fuel-up) is 10%

1. If P(E1) = 3P(E2) = 0. find the probability of the remaining events (E3, E4, E5) assuming the remaining events are equally probable.

Given that

P(E1) = 0.3

P(E2) = P(E1)/3 = 0.1

Report

While analyzing the MPG for my Subaru Impreza Sport, I grouped fuel-ups into five categories based on MPG ranges. For the first part, I used the probabilities provided and the relationship between P(E4) = 2P(E5) to calculate P(E4) = 0.2 and P(E5) = 0.1. This means 20% of fuel-ups had MPG over 25, and 10% had no MPG recorded. For the second part, I distributed the remaining probabilities equally among E3, E4 and E5, finding that they had a probability of 0.2 or 20%.

**Page 39 Chapter 2**  
**Question 2.28**

In a dataset of 40 recorded fuel-ups for a Subaru Impreza Sport, four specific fuel-ups were flagged for investigation: two with MPG above 25 and two with MPG below 15. Two of these fuel-ups will be selected at random for further analysis.

1. List the possible outcomes of this selection.
2. Assign reasonable probabilities to the sample points.
3. Find the probability that at least one of the fuel-ups selected has MPG above 25.

Solution

1. List the possible outcomes of this selection.

A1 and A2 = (MPG > 25)

B1 and B2 = (MPG < 15)

1. Assign reasonable probabilities to the sample points.

Each of the pairs has a chance of being selected. Since there are 6 it would be

1. Find the probability that at least one of the fuel-ups selected has MPG above 25.

Select one of the fuel-ups with an MPG > 25. Then use complementary probability

The probability that the fuel-up will be > 25 is

Or 83%

Report

I looked at four specific fuel-ups from my data. Two with MPG > 25 and two with MPG < 15. I listed all possible pairs of two fuel-ups, which gave me six combinations. Since the selection was random, each pair had an equal chance of being chosen, which was about 16%. I then calculated the probability of selecting at least one fuel-up with MPG > 25. By figuring out the chance of not picking any high-efficiency fuel-ups, I found that there’s an 83% chance of picking at least one fuel-up with MPG > 25.

**Page 35 Chapter 2**  
**Question 2.2**  
Imagine analyzing the MPG data for a Subaru Impreza Sport where three possible MPG ranges are represented as follows:

G: MPG greater than 25 (a highly efficient trip).

D1: MPG between 15 and 25 (an average trip).

D2: MPG below 15 (a less efficient trip).

The dataset represents a situation where the MPG for a fuel-up falls into one of these three categories. Now, consider the following:

1. If the MPG range is chosen randomly, assign reasonable probabilities to the events and calculate the probability of selecting a highly efficient trip (G).
2. Suppose an event (a "dud" MPG range) is revealed, narrowing the possibilities.

i. If the contestant stays with their initial choice, what is the probability of selecting G?

ii. If the event is revealed to be D1, and the choice is switched, what is the probability of selecting G?

iii. What happens if the revealed event is D2, and the choice is switched?

iv. Calculate the probability of winning (G) when switching.

v. Based on your results, which strategy maximizes the probability of selecting G: staying or switching?

Solution

1. If the MPG range is chosen randomly, assign reasonable probabilities to the events and calculate the probability of selecting a highly efficient trip (G).

If the MPG range is chosen randomly, each range is equally likely:

1. Suppose an event (a "dud" MPG range) is revealed, narrowing the possibilities.

i. If the contestant stays with their initial choice, what is the probability of selecting G?

Or 33%

ii. If the event is revealed to be D1, and the choice is switched, what is the probability of selecting G?

If D1 is revealed, switching will result in G being selected if D2 was the first choice

iii. What happens if the revealed event is D2, and the choice is switched?

If D2 is revealed, switching will result in G being selected if D1 was the first choice

iv. Calculate the probability of winning (G) when switching.

Switching will have G being selected if the initial choice was a dud (D1 or D2):

Or 66%

v. Based on your results, which strategy maximizes the probability of selecting G: staying or switching?

The probability of winning when staying is ​, while the probability of winning when switching is ​. Making switching max the probability of selecting a high efficient trip

Report  
While analyzing MPG data for my Subaru Impreza Sport, I imagined a scenario where three possible MPG ranges: G (MPG > 25), D1 (15 < MPG ≤ 25), and D2 (MPG ≤ 15). At first each range had an equal chance of being chosen, so the probability of selecting G was ​. If one "dud" MPG range (D1 or D2) was shown, switching to the other unopened category improved the odds of selecting G. By calculating the probabilities, I found that staying with the initial choice had a chance of success ​, while switching increased the chance to ​ ​.

**Page 48 Chapter 2**  
**Question 2.35**

Imagine a dataset tracking 6 fuel-ups for a Subaru Impreza Sport in one region and 7 fuel-ups in another. Suppose we want to analyze the different combinations of fuel-ups that can occur over separate days.

1. How many different combinations of fuel-ups are possible if one fuel-up is selected from each region?
2. If the MPG values of each fuel-up are tracked, how many total MPG combinations are possible?

Solution

1. How many different combinations of fuel-ups are possible if one fuel-up is selected from each region?

Since there are 6 fuel-ups in one region and 7 in the other then it would be

1. If the MPG values of each fuel-up are tracked, how many total MPG combinations are possible?

42 different combinations

Report  
I looked at the possibilities of combining fuel-ups from two regions based on my Subaru Impreza Sport dataset. The first region had 6 fuel-ups recorded, while the second region had 7. For each fuel-up in the first region, there were 7 options from the second region. This resulted in a total of unique combinations of fuel-ups. Also, if MPG values were tracked for each fuel-up, the total number of MPG combinations would also be 42.

**Page 60 Chapter 2**  
**Question 2.94**

Using the MPG data for my Subaru Impreza Sport, consider two conditions related to fuel efficiency:

Condition A: The probability of MPG greater than 20 (P(A)=0.6

Condition B: The probability of MPG less than 25 (P(B)=0.975

The probability that MPG falls into both categories (P(A∩B) is .6.

1. Find the probability that MPG is either greater than 20, <= 25, or both.
2. Find the probability that MPG is neither >= 20 or < 25.

Solution

1. Find the probability that MPG is either > 20, <= 25, or both.

The probability that the MPG is > 20, <= 25, or both is 97%

1. Find the probability that MPG is not >= 20 or < 25.

the probability that MPG is not >= 20 or < 25 is 2.5%

Report

While looking at the MPG for my Subaru Impreza Sport, I saw the probability of a fuel-up meeting two conditions: MPG > 20 or MPG< 25. Using the union formula, I found that 97.5% of the fuel-ups fit one or both the conditions. The remaining 2.5% was found where it was not >= 20 or < 25. Using the complement of .

**Page 73 Chapter 2**  
**Question 2.124**

The dataset tracks the MPG of a Subaru Impreza Sport and categorizes fuel-ups as belonging to one of two regions:

Region A: 40% of the fuel-ups.

Region B: 60% of the fuel-ups.

It is further reported that:

30% of the fuel-ups in Region A achieve MPG greater than 25.

70% of the fuel-ups in Region B achieve MPG greater than 25.

If a randomly selected fuel-up achieves MPG greater than 25, find the conditional probability that it came from Region B.

Solution

If a randomly selected fuel-up achieves MPG greater than 25, find the conditional probability that it came from Region B?

Let B be Region B and E be MPG > 25. Find

Bayes Theorem

Were

P(A)=0.4, P(B)=0.6P(B) = 0.6P(B)=0.6

P(E∣A)=0.3, P(E∣B)=0.7

Find P(B | B)

77% of random fuel-ups that are > 25MPG are from Region B

Report

I analyzed the MPG data for my Subaru Impreza Sport to see that a fuel-up from Region B given its MPG > 25. Using Bayes' theorem, I calculated that the probability is approximately 77.8% of the time. This shows that my most high MPG fuel-ups likely come from Region B.

**Page 85 Chapter 2**  
**Question 2.179**

From the dataset, consider two fuel-ups with the following Cost/Gallon values:

Fuel-Up A: $3.09 per gallon.

Fuel-Up B: $3.29 per gallon.

Assume the drivers compare costs randomly over six fuel-ups, where the drive of each comparison is determined randomly with an equal probability for each fuel-up.

1. What is the probability that the two drivers break even after six fuel-ups (three wins each)?
2. What is the probability that Fuel-Up B ($3.29) wins all six comparisons?

Solution

1. What is the probability that the two drivers break even after six fuel-ups (three wins each)?

To break even the driver must win all 3 rounds. Using Binomial Formula

The probability that they would break even after 6 fuel-ups is 31%

1. What is the probability that Fuel-Up B ($3.29) wins all six comparisons?

For Fuel-Up B to win all 6 Rounds

The probability of Fuel-Up B winning all 6 Rounds is 1.56%

Report  
I looked at the two specific fuel-ups from the dataset with costs of $3.09 and $3.29 per gallon. I used the binomial probability function to compare the fuel-ups. Comparing these fuel-ups randomly over six rounds, the probability of breaking even (three wins each) was 31.25%. Unlike the chance of Fuel-Up B ($3.29 per gallon) winning all six rounds was much smaller, at just 1.56%.

**Page 90 Chapter 3**  
**Question 3.1**

Using the dataset for fuel-ups, let’s consider two conditions related to fuel costs:

Condition A: Fuel cost per gallon exceeds $3.20.

Condition B: Fuel per gallon is below $3.00.

From the dataset, it was observed:

20% of the fuel-ups meet either condition.

40% meet Condition A.

50% meet Condition B.

Find the probability distribution for Y, the number of conditions met by a randomly chosen fuel-up.

Solution

Y = 0 Non are met

Y = 1: One condition is met

Y = 2: Both Conditions are met

Inclusive – exclusive

*P(A) = 0.40, P(B) = 0.50, and P(A∩B) = P(A) + P(B) – 1 + P(Y=0)*

P(Y = 2) = P(A∩B) = 0.10

P(Y = 0) = 0.20, P(Y = 1) = 0.70, P(Y = 2) = 0.10

Report

I used the fuel cost data from my data set to look at how often two conditions were met: costs above $3.20 (Condition A) and costs below $3.00 (Condition B). Using the inclusion-exclusion principle, I calculated the probability distribution for Y. I found that 20% of the fuel-ups didn’t meet either condition (Y = 0), 70% met exactly one of the conditions (Y = 1), and 10% met both conditions (Y = 2).

**Page 119 Chapter 3**  
**Question 3.67**

Using the fuel cost data from the dataset, suppose that 30% of the fuel-ups cost more than $3.20 per gallon. Fuel-ups are reviewed sequentially and selected at random.

What is the probability that the first fuel-up costing more than $3.20 per gallon is found on the fifth review?

Solutions

What is the probability that the first fuel-up costing more than $3.20 per gallon is found on the fifth review?

Since this problem is a geometric probability question, where the trials P = 0.3 and the failure is

K = 5

The probability that the first fuel-up will cost more than $3.20 is 7.2%

Report

I analyzed the fuel cost data to determine the likelihood of the first fuel-up costing > $3.20 per gallon on the fifth review. Using the geometric probability formula with a 30% chance of success per review, I calculated the probability to be 7.2%. This calculation showed how probabilities decrease with more failures before the first success. By looking at this pattern, I can estimate how often specific costs occur in my dataset.

**Page 123 Chapter 3**  
**Question 3.90**

Using the dataset for fuel-ups, assume the goal is to identify fuel-ups where the MPG exceeds 25. Suppose that 40% of the fuel-ups meet this condition. If the goal is to find three fuel-ups with MPG greater than 25, what is the probability that 10 fuel-ups must be reviewed to find the three that meet the condition?

Solution

If the goal is to find three fuel-ups with MPG greater than 25, what is the probability that 10 fuel-ups must be reviewed to find the three that meet the condition?

In this problem we are using the negative binomial distribution, where the success (p) is 0.40 and the number of successes (r) is 3

k = total trials (10)

r = number of success (3)

p = probability of success (0.40)

Now for

The probability that 10 fuel-ups need to be reviewed to find 3 fuel ups with > 25 MPG is 6.4%

Report

While analyzing the MPG data for my Subaru Impreza Sport, I looked at the likelihood of finding three fuel-ups with MPG greater than 25. If 40% of the fuel-ups met this condition and used the negative binomial formula to calculate the probability of needing for exactly 10 fuel-ups to find three that qualified ones. The result was a 6.46% probability. This means that while it’s possible to find three qualifying fuel-ups within 10 reviews.

**Page 129 Chapter 3**  
**Question 3.104**

The dataset includes information on fuel packets categorized by octane levels:

High-Octane (93): 15 packets.

Low-Octane (87): 5 packets.

Four packets were randomly selected, and all were found to be high-octane (93). Two additional packets were then selected from the remaining pool and sold.

What is the probability that the six packets selected are that the first four are high-octane (93), and the two sold are low octane (87)?

Solution

What is the probability that the six packets selected are that the first four are high-octane (93), and the two sold are low octane (87)?

Using the hypergeometric probability formula, we will file the probability

Total packets: N = 20 N = 20 N = 20

High-octane (93): K = 15 K = 15 K = 15

Low-octane (87): N−K = 5 N - K = 5N−K = 5.

First 4 packets: All high-octane (93).

Next 2 packets: All low-octane (87).

Hypergeometric Formula

Calculate the term

Find Probability

There is a 35.2% that the six packets selected are that the first four are high-octane (93), and the two sold are low octane (87)

Report

Using the octane data from my dataset, I analyzed the likelihood of selecting six packets that the four are high-octane (93) and the next two are low octane (87). Using the hypergeometric formula, I calculated the probability to be 35.2%. This shows that it is possible, it’s not guaranteed due to the amount of high and low-octane packets in the dataset.

**Page 136 Chapter 3**  
**Question 3.128**

Using the MPG data from the dataset, assume that the Subaru Impreza achieve MPG values according to a Poisson process with an average of 80 cars per hour reaching 20 MPG. If a driver pauses for a one-minute rest, what is the probability that at least one car achieves MPG greater than 20 during this time?

Solution

Using the Poisson probability we are finding the cars achieving > 20 MPG

Mean rate of cars achieving MPG > 20 per hour: λ=80.

Time interval: 1 minute ( of an hour).

Mean rate for the time interval:

Probability of no cars

Probability with one car

The probability that at least one car achieves MPG > 20 during the minute is 73.6%

Report

Using the MPG data for my Subaru Impreza Sport, I saw that at least one car had MPG > 20 during a one-minute interval. Assuming a Poisson process with an average rate of 80 cars per hour, I calculated the probability to be 73.6%. This questions demonstrates the use of Poisson distributions.

**Page 148 Chapter 3**  
**Question 3.170**

Using the Cost/Gallon data from the dataset, assume the average cost per gallon is $3.00 with a standard deviation of $0.10. Using Tchebysheff’s theorem, find a lower bound for the number of fuel-ups in a dataset of 400 that are expected to have a cost per gallon between $2.80 and $3.20.

Solution

The range 2.80 to 3.20 is on the mean of $3.00, with a distance of $0.20 on either side. This distance represents k standard deviations

Use Tchebysheffs theorem

This means at least 75% of the fuel-ups have a cost per gallon within the range $2.80 to $3.20.

Find Lower Bound

Making 300 fuel-ups are expected to have a cost per gallon of $2.80 and $3.20

Report

I looked at the Cost/Gallon data for my Subaru Impreza Sport, assuming an average cost of $3.00 and a standard deviation of $0.10. By using Tchebysheff’s theorem, I calculated that at least 75% of the data falls within two standard deviations of the mean, which corresponds to the range $2.80 to $3.20. For a dataset of 400 fuel-ups, this means at least 300 are expected to fall within this range.

**Page 167 Chapter 4**  
**Question 4.11**

Suppose the **Gallons** data from the dataset represents a variable YYY with the following probability density function:

where c is a constant.

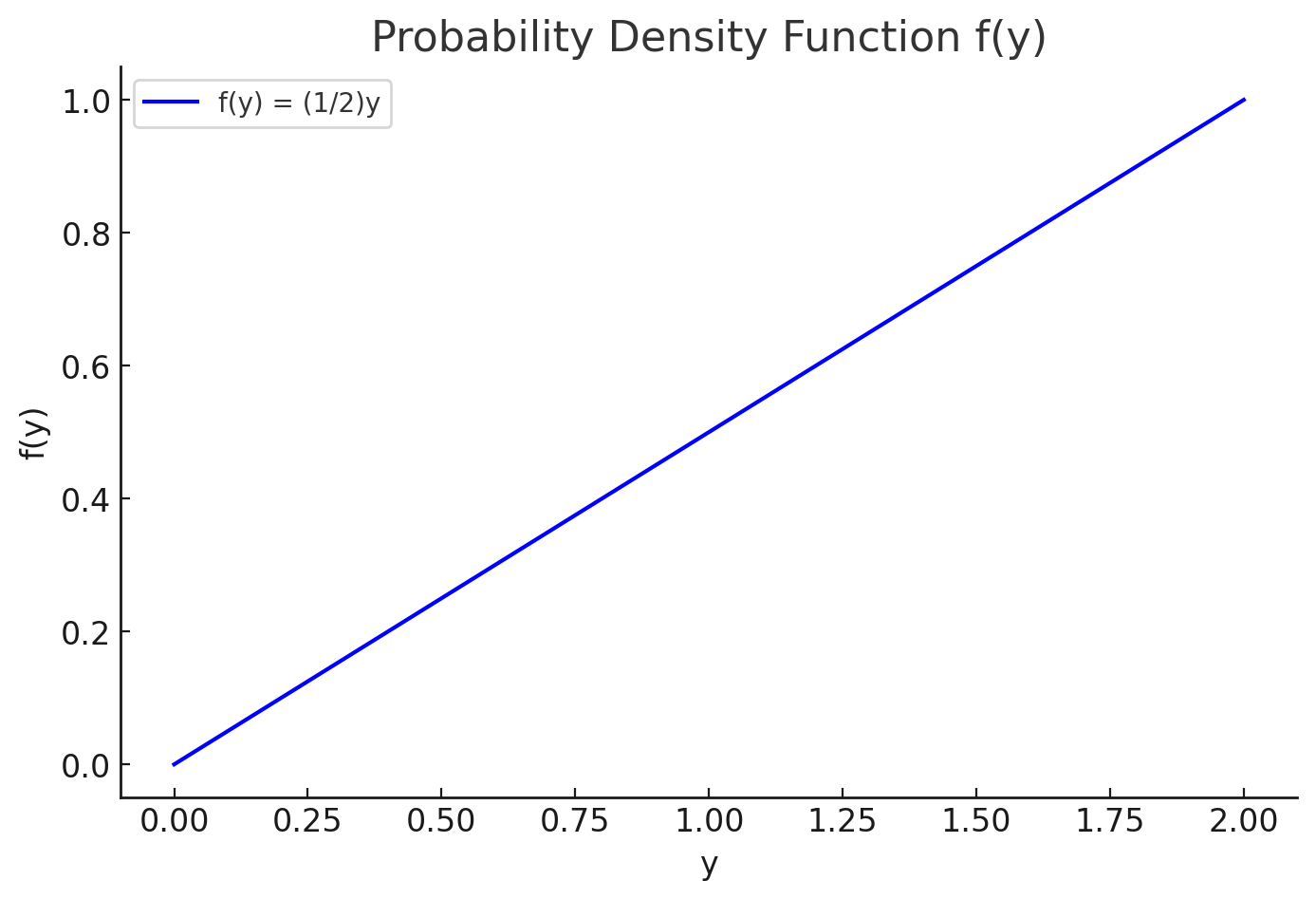
1. Find the value of C that makes f(y) a probability density function.
2. Find the cumulative distribution function F(y).
3. Graph f(y) and F(y)
4. Use F(y) to find P(1 ≤ Y ≤ 2)
5. Use f(y) and geometry to find P(1 ≤ Y ≤ 2)

Solution

1. Find the value of C that makes f(y) a probability density function.
2. Find the cumulative distribution function F(y).

For f(y) = and 0

1. Graph f(y) and F(y)



A green line graph with numbers and a white background

Description automatically generated

1. Use F(y) to find P(1 ≤ Y ≤ 2)

*P(1 ≤ Y ≤ 2)=F (2) − F(1)*

1. Use f(y) and geometry to find P(1 ≤ Y ≤ 2)

Report

I viewed the distribution of the Gallons for fuel-ups using a given probability density function. I found c = ​ to find the function and get the cumulative distribution function F(y) = ​. Using F(y), I calculated P (1 ≤ Y ≤ 2) = 0.75 or 75% which was confirmed using the function. These calculations showed how the density functions and cumulative functions relate to probabilities in real-world data with my Cost/Gallon.

**Page 177 Chapter 4**  
**Question 4.49**

My vehicle arrives at a gas station at random within a one-hour interval. The station was fully occupied for the first 15 minutes of the hour. What is the probability that the vehicle arrived when the station was not fully occupied?

Solution

Uniformed Probability problem

Total interval: 60 minutes

Fully occupied: 15 minutes

Unoccupied Intervals: 60 – 15 = 45 minutes

The probability that the Subaru will arrive when the station is not fully occupied is 75%

Report

By using the Time data, I looked at the probability of a my vehicle arriving at a gas station during a random one-hour interval. I Assumed the station was fully occupied for the first 15 minutes, I calculated the probability of arriving when the station was not fully occupied to be 75%.

**Page 191 Chapter 4**  
**Question 4.94**

The MPG data for a Subaru Impreza Sport is modeled using an exponential distribution. Initially, the mean MPG is 25.2.

1. What is the probability that a randomly selected one-hour MPG exceeds 40 MPG?
2. If a driving strategy reduced the mean MPG to 20.0, what is the probability that the one-hour MPG exceeds 40 MPG?

Solution

1. What is the probability that a randomly selected one-hour MPG exceeds 40 MPG?

Using the probability density function

MPG > 40 when the mean is 25.2

Use =

The probability that the MPG will be > 40 when the mean is 25.2 is 20.4%

1. If a driving strategy reduced the mean MPG to 20.0, what is the probability that the one-hour MPG exceeds 40 MPG?

Use =

The probability that the MPG will be > 40 when the mean is 20 is 13.5%

Report

I analyzed the MPG data from my dataset and used it with exponential distribution. At first, the mean MPG was 25.2, and I calculated that the probability of > 40 MPG in a one-hour period was 20.4%. Then trying to reduce the mean to MPG to 20.0, this probability decreased to 13.5%.

**Page 233 Chapter 5**  
**Question 5.9**

Let Y1 (MPG) and Y2​ (Cost/Gallon) have the joint probability density function:

1. Find the value of k that makes f(y1,y2) a probability density function.
2. Find P (Y1 ≤ 3/4, Y2 ≥ 1/2)

Solution

1. Find the value of k that makes f(y1,y2) a probability density function.

Find K

For y1

for y2

1. Find P (Y1 ≤ 3/4, Y2 ≥ 1/2)

P(Y1​ ≤ 3/4, Y2 ​≥ 1/2) =

Case 1 (1/2 ≤ y2 ​≤ 3/4)

Case 2 (3/4 ≤ y2 ​≤ 1)

]

Report

I analyzed the joint distribution of my MPG and Cost/Gallon, I used my data with a custom joint probability density function. I calculated k = 6 to create the function. Using integration, I found that P(Y1 ≤ 3/4, Y2 ≥ 1/2) = 0.969, showing a high probability of the outcomes. These calculations demonstrate how joint distributions provide insight into relationships between my vehicles.

**Page 243 Chapter 5**  
**Question 5.23**

Using the dataset, let Y1​ represent the proportion of the tank that is filled at the beginning of a fuel-up and Y2​ represent the proportion of the tank that is used by the end of the fuel-up. Their joint density is given by:

1. Find the marginal density function for Y2​.
2. For what values of y2​ is the conditional density f(y1 ∣ y2) defined?
3. What is the probability that more than half a tank is sold (Y2>0.5) given that three-fourths of a tank is stocked (Y1 = 0.75)?

Solutions

1. Find the marginal density function for Y2​.

Use

1. For what values of y2​ is the conditional density f(y1 ∣ y2) defined?

The conditional density f(y1∣y2) is defined wherever the joint density f(y1,y2) is valid, which occurs when 0≤ y2 ≤ y1 ≤ 1. Then, f(y1∣y2) is

1. What is the probability that more than half a tank is sold (Y2 > 0.5) given that three-fourths of a tank is stocked (Y1 = 0.75)?

The conditional density formula

For Y1 = 0.75

The probability P(Y2​ > 0.5∣Y1​ = 0.75) is

Report

I saw the relationship between Gallons (Y1​) and Total Cost (Y2​) using a given joint probability density function. I found that the density for Y2​ is fY2(y2) = 3/2(1−y), for 0 ≤ y2 ≤ 10. The conditional density f(y1 ∣ y2) is defined for y2 ≤ y1 ≤ 1. Finally, I calculated that the probability of selling more than half a tank (Y2 > 0.5) given that three-fourths of a tank (Y1 = 0.75) was stocked is approximately 66.67%. This highlights how the relationship between stocked and sold proportions can be modeled using joint distributions.

**Report Summary**

This project was about reviewing my vehicle's fuel-up data that I have been tracking for around 2 years. By using information like MPG, Cost/Gallon, and other details from the dataset, I worked through a series of textbook questions that included multiple different formulas.

The process started with organizing the data and matching it to the questions. I worked on finding probabilities, like the chances of getting a certain MPG range or finding how fuel costs might vary in areas. Some questions involved probability distributions like binomial or exponential, which helped model events like random MPG values or fuel costs. This was great for my own insight of my car.

I also explored relationships between variables, such as MPG and Cost/Gallon, using joint and conditional probabilities. This helped me understand how one variable could change into another. One example was I calculated probabilities for tank usage based on how much was filled, using techniques like integration and geometry.

For most questions I had to adapt my data to fit the question as well. For example, I used Tchebysheff’s theorem to estimate how many fuel-ups would fall within a cost range and used hypergeometric probabilities to look at packet selection scenarios.

Overall, this project helped me apply statistical ideas to my real-life data. It showed me how to use formulas and methods to answer questions about patterns and probabilities, while also giving me a better understanding of my car’s MPG and stations to visit.