Dataset Research Report:

Automotive Sales Statistics

By Jake Cubernot

Table of Contents

The Story

A new car dealership is opening soon and like any business that offers a commodity, it is crucial to learn as much about the customers as possible so that one can offer the perfect variety of commodities. In this case, the dealership wants to perform multiple statistical problem scenarios with a dataset on sold cars throughout the United States. With these problem scenarios solved, the company hopes to offer the best vehicles for their customers and beat the local competition.

Chapter 1 Problems

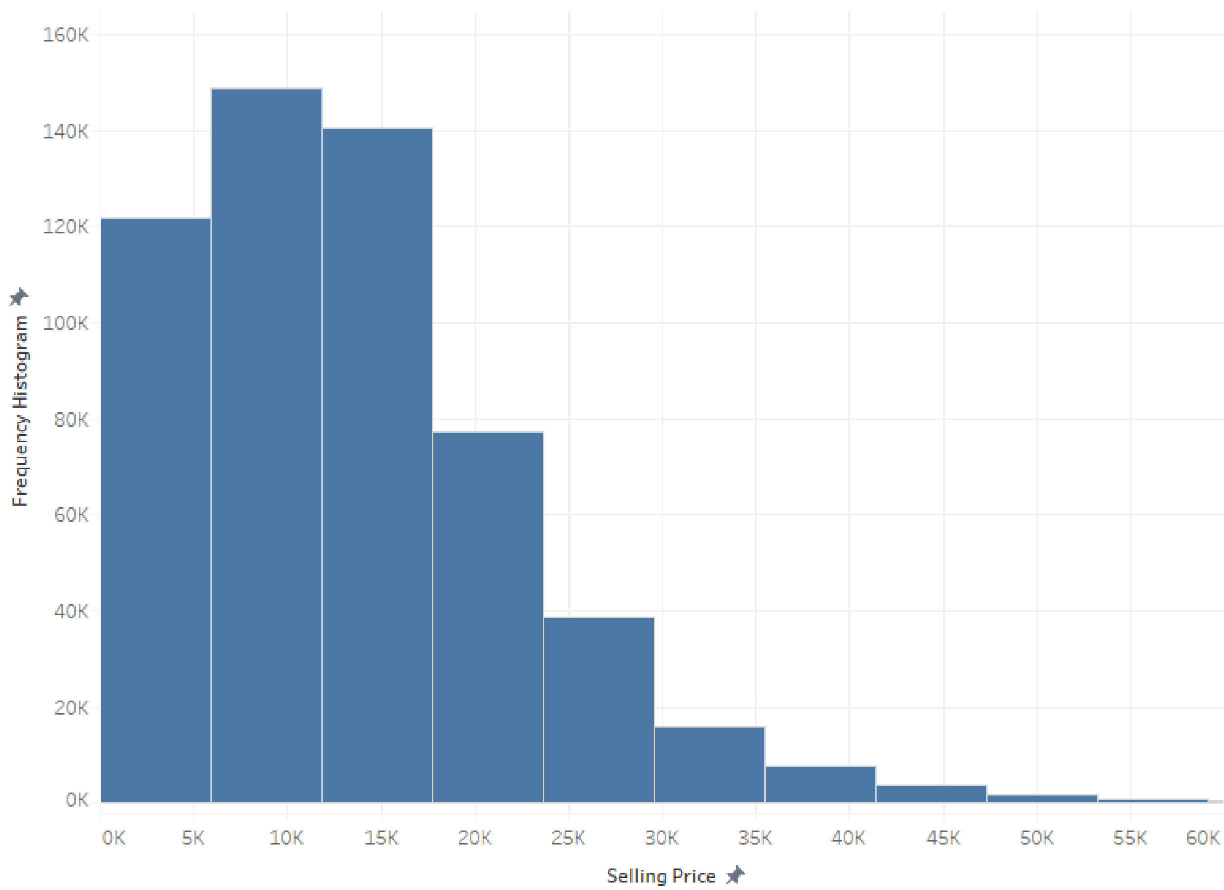
Chapter 1.1 Problem

A car dealership wants to estimate the average selling price of used vehicles based on vehicle sales and market trends in North America.

Population: used vehicles being sold in North America. Objective: to estimate the average selling price of used vehicles.

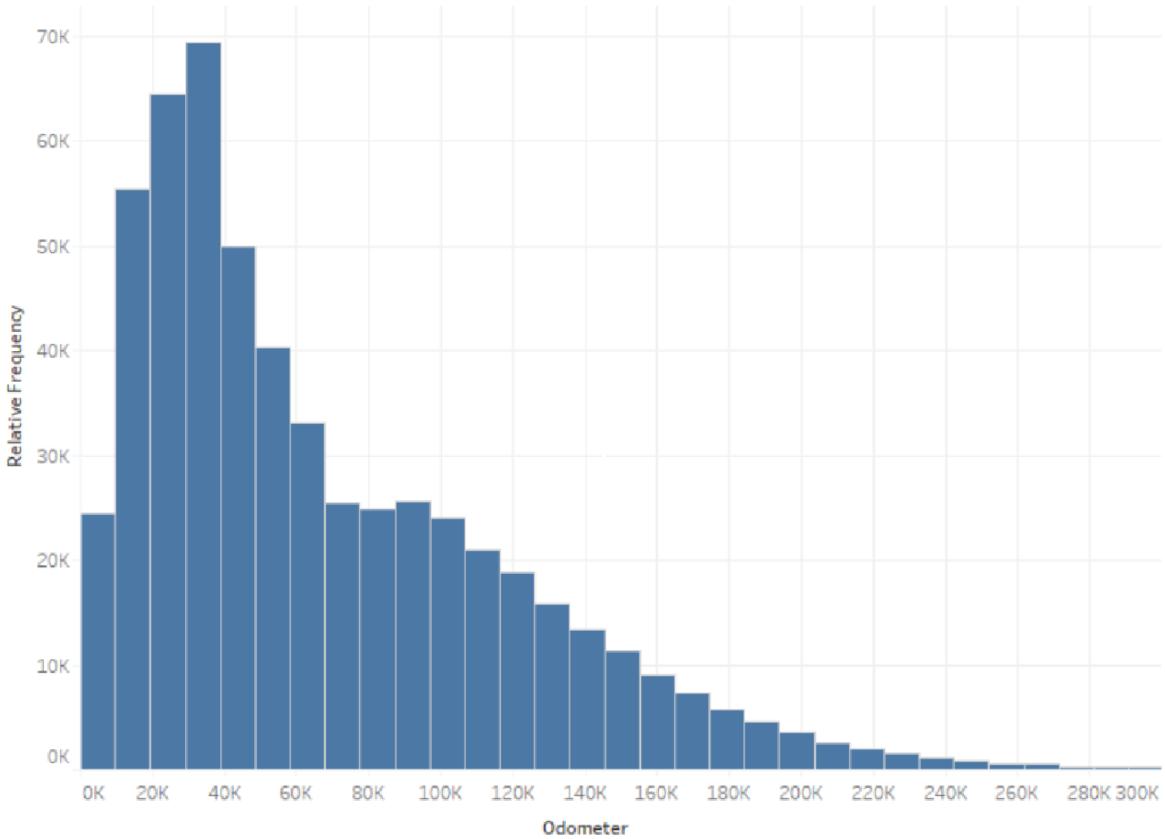
Chapter 1.3 Problem

A car dealership wants to learn which car of a certain price range sells the most. Cars ranging from $6K to $12K sold the most. The histogram is below



Chapter 1.5 Problem

Given here is the relative frequency histogram associated with sold cars' mileage in North America.



When observing the relative frequency histogram, the largest number of cars sold have a mileage ranging from 20K to 40K miles.

Chapter 2 Problems

Chapter 2.3 Problem

Suppose we have a dataset containing information about vehicle sales, including the make, model, and condition of the vehicles. Let's define subsets based on the condition of the vehicles:

Subset A: Vehicles in excellent condition

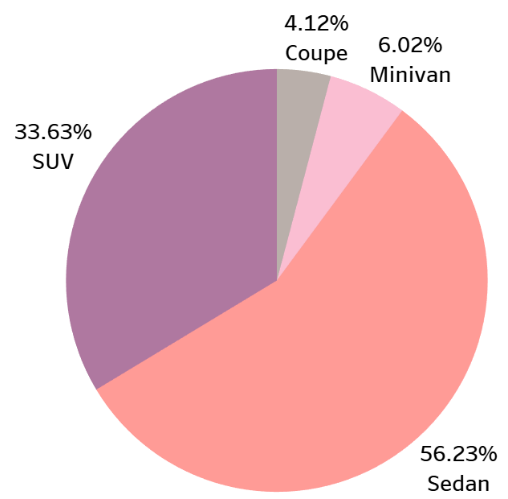
Subset B: Ford F-150s

Subset C: Vehicles with mileage less than 50,000 miles

A = {159,973 Vehicles}, B = {14,589 Vehicles}, C = {267,241 Vehicles}, A ∩ B = {5,112 Vehicles}, B ∩ C = {6,403 Vehicles}, A ∩ C = {98,580 Vehicles}

Chapter 2.4 Problem

When making a set of SUVs, Coupes, Minivans, and Sedans being sold in the US, the dealership found the percentages below for the total car bodies sold between the four types.



Denote the events as SUV (SU), Coupe (C), Minivan (M), and Sedan (SE).

P(someone is interested in buying a Minivan or an SUV) = P(SU) + P(M) = 39.65%

Chapter 2.5 Problem

The dealership considered a situation where a customer needs to choose a combination of their exterior and interior colors for their car, where the colors for both are either black, grey, or white.

The number of sample points within this sample space is S = {BB, BG, BW, GB, GG, GW, WB, WG, WW} = 9.

The probability that a customer chooses at least one black in their combination is a 5/9 or 55.56% chance.

The probability that a customer does not choose the same color for both the exterior and interior is a 6/9 or 66.67% chance.

Chapter 2.6 Problem

Twelve cars at the dealership are getting moved to other dealerships in the area. They will be sent to dealerships in such a way that four go to dealership A, five go to dealership B, and three go to dealership C.

In this situation, there are $\binom{12}{4}\binom{8}{5}\binom{3}{3} = 27,720$ distinct ways this be accomplished.

Chapter 2.7 Problem

The car dealership was able to determine that almost 34% of cars sold in a set were SUVs. They also found that there was close to a 22% chance of the cars sold being black in the set. Finally, within the set, they could determine that almost 8% of cars sold were black if the car was an SUB.

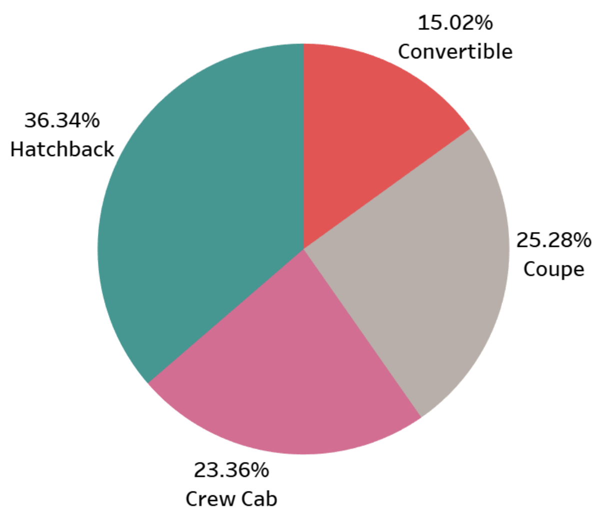
Subset A: vehicle is an SUV

Subset B: vehicle is black

P(A) = 34%, P(B) = 22%, and P(B | A) = 8%.

Chapter 2.8 Problem

The dealership creates a set of four car bodies below.



Subset A: vehicle is a hatchback

Subset B: vehicle is a crew cab

With the set, the dealership was able to determine the following probabilities:

P(A) = 36.34%

P(B) = 23.36%,

P(A ∩ B) = 8.49%

P(A ∪ B) = 59.70%

P($\overline{A}$) = 63.66%

P($\overline{B}$) = 76.64%

P($\overline{A}$ ∪ $\overline{B}$) = 40.30%

Chapter 2.9 Problem

The dealership needs to do inspections on three cars that will be going up for sale at their business. There is a 40% chance the cars will get rejected.

If two out of the three cars get rejected, there are $\binom{3}{2} = 3$ ways that the cars are evaluated with them all having a 40% chance of getting rejected, so the probability of three cars getting rejected is $3(.4)^3 = 0.192$ or a 19.20% chance.

Chapter 2.10 Problem

The car dealership was able to determine that almost 34% of cars sold in a set were SUVs. They also found that there was close to a 22% chance of the cars sold being black in the set. Finally, within the set, they could determine that almost 8% of cars sold were black if the car was an SUB.

Subset A: vehicle is an SUV

Subset B: vehicle is black

With this information, the dealership was now able to use Bayes' Theorem to determine the probability of cars sold were an SUV if they were black, which was

$P(A \mid B) = \frac{(.08)(.34)}{(.22)} = 0.1237$ or 12.37% chance.

Chapter 3 Problems

Chapter 3.2 Problem

The dealership has three cars that need to be temporarily moved to a storage garage. Each car is randomly placed into one of the 3 storage garages owned by the dealership. The company is curious to find out the probability distribution for a number of garages that do not get any of the three cars.

Y = the number of empty garages

There are 27 ways to place the 3 cars into the 3 garages.

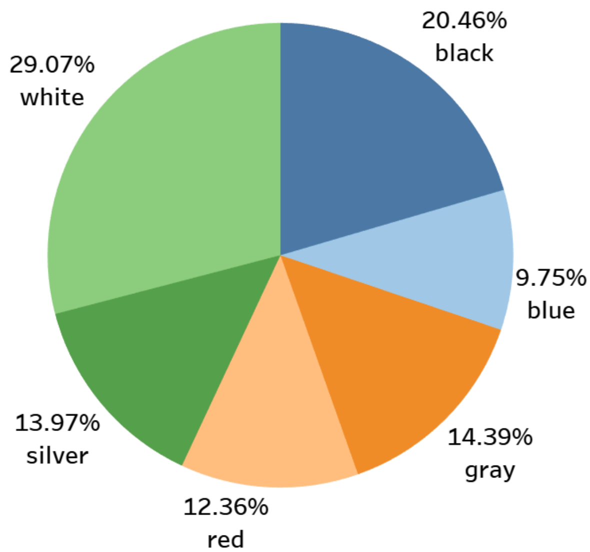
P(0 garages are empty) = $\frac{3!}{27} = \frac{6}{27} = 0.2223 = 22.23\%$

P(2 garages are empty) = $\frac{3}{27} = 0.1112 = 11.12\%$

P(1 garages are empty) = $1 - \frac{6}{27} - \frac{3}{27} = 0.6667 = 66.67\%$

Chapter 3.4 Problem

The dealership looked into the most sold Ford cars by color below. The dealership randomly selects 10 Ford vehicles and they want to find out the binomial distribution that at least one vehicle will be white.



The company found that there was a $\frac{10!}{(10-1)!\,1!} * 0.2907^{1} * (1 - 0.2907)^{10-1} = 0.1321 = 13.21\%$ chance of the Ford vehicles being white.

Chapter 3.5 Problem

The dealership determined that there is a 20.46% chance that a sold Ford vehicle is black. They want to now figure out the probability of selling a black Ford vehicle from their third client successful client. By using geometric distribution, the dealership concluded that there was a $(1 - 0.2046)^{3-1} * 0.2046 = 0.1294 = 12.94\%$ chance for that situation to happen.

Chapter 3.6 Problem

At the dealership, there is a 25% chance that they can sell a car. The company wants to find out the probability that they have exactly five failed sales before they have a second successfully sold car. They used the negative binomial probability distribution to find out that there is a $\binom{5-1}{2-1} * (.25)^{2} * (1 -.25)^{5-2} = 0.1055 = 10.55\%$ chance for this situation to happen.

Chapter 3.7 Problem

When the dealership looked into Jeep vehicles sold, they found that there was a total number of 45 Jeeps. Out of the total, 10 Jeeps were sold in very poor condition.
The company wanted to determine the probability of selecting 20 of the total number of Jeeps at random without replacement and finding 2 very poorly conditioned Jeeps within the sample. They were able to use the hypergeometric distribution to find that there is a $\frac{\binom{10}{2}\binom{45-10}{20-2}}{\binom{45}{20}}$ = 0.0644 = 6.44% chance of that situation.

Chapter 3.8 Problem

The dealership has found that an average of 3 people come to the car lot every 10 minutes on the weekdays. The dealership wants to know the probability of getting 8 people on the car lot in 10 minutes. The company uses Poisson distribution to determine that there is a $e^{-3} * \frac{3^8}{8!} = 0.0081 = 0.81\%$ chance of that situation occurring.

Chapter 3.11 Problem

The dealership found that out of the 665 Lincoln vehicles being sold at a selling price between $6K and $12K, there was a standard deviation of 1,604 within the entire fleet of sold Lincoln vehicles. To find the probability of finding a Lincoln vehicle ranging from $6K to $12K, the dealership used Chebyshev's theorem to learn that there is a $1 - \frac{1}{1604^2} = 0.99999961 = 99.999961\%$ chance of this occurring.

Chapter 4 Problems

Chapter 4.2 Problem

The company wants to find the value c that makes a probability density function.

$$f(y) = \begin{cases} cy, & 0 \leq y \leq 2, \\ 0, & elsewhere. \end{cases}$$

$$\int_0^2 cydy = \left[\frac{cy^2}{2}\right] = (2c = 1) = (c = \frac{1}{2})$$

Chapter 4.4 Problem

There are 4 unique makes of cars, Ford, Jeep, Honda, and Lincoln, in the garage at the dealership. The company wants to learn the uniform probability distribution of pulling one of the 4 cars made out of the garage.

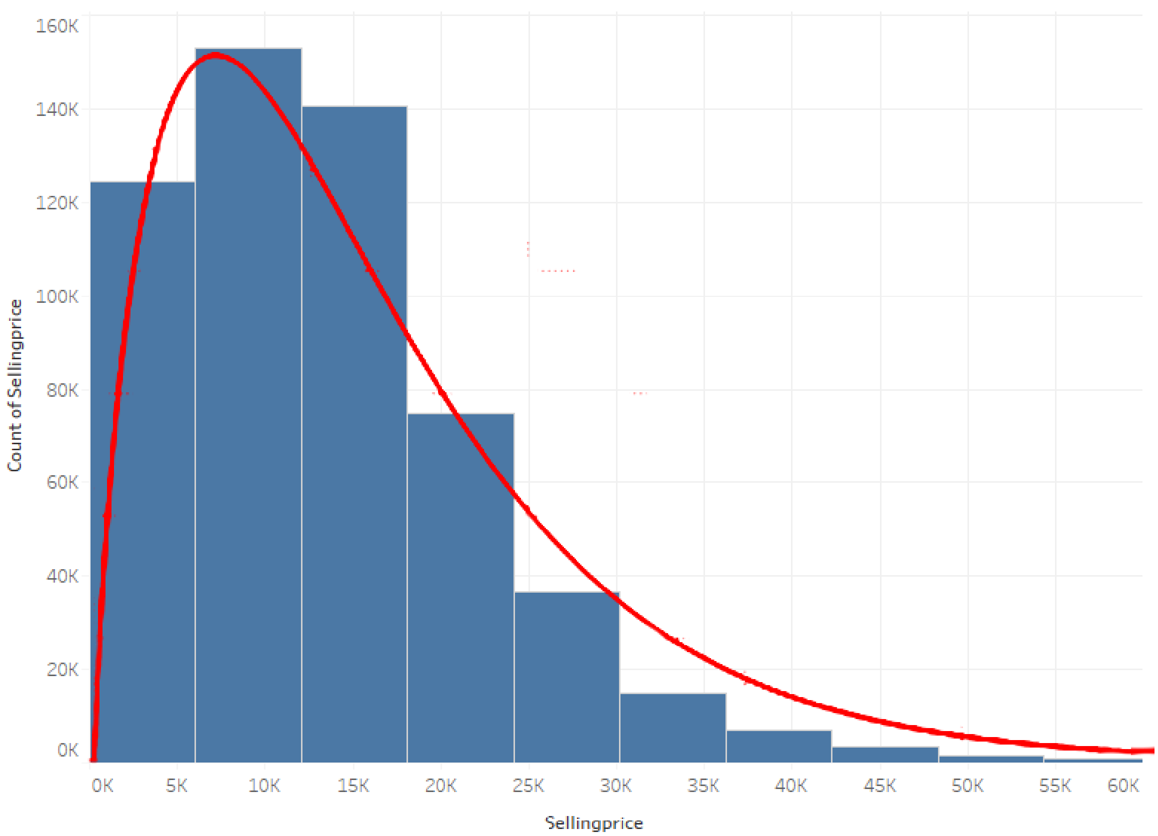
$P(0 \leq X \leq 1) = 1 = \frac{1}{4-0} = \frac{1}{4}$

This equation shows the simple chance that pulling out one of the vehicles, like the Ford, has a 1 out of 4 chance of being picked.

Chapter 4.6 Problem

The dealership wanted to figure out the gamma distribution of the selling prices for vehicles sold in the U.S. Using a previous histogram of selling prices, the company was able to find a gamma distribution of:

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} \; for\ x, \alpha, \beta > 0,\ where\ \alpha = 1.8\ and\ \beta = 7$$



Chapter 5 Problems

Chapter 5.2 Problem

In one section of the dealership's car lot, there are 12 cars. They consist of 6 Ford vehicles, 4 Jeep vehicles, and 2 Lincoln vehicles. 3 of the 12 vehicles are selected for inspection. The dealership wants to find the probability that a Ford or a Jeep vehicle will be selected. They use the joint probability function to figure it out:

$Y_1 = the\ 6\ Ford\ vehicles$

$Y_2 = the\ 4\ Jeep\ vehicles$

$$p(y_1, y_2) = \frac{\binom{6}{y_1}\binom{4}{y_2}\binom{3}{4-y_1-y_2}}{\binom{12}{4}},\ where\ 0 \le y_1,\ 0 \le y_2,\ and\ y_1 + y_2 \le 3.$$

		y_1		
		0	1	2
	0	0	0	1/11
y_2	1	4/495	0	0
	2	2/55	0	0

Chapter 5.3 Problem

After the dealership's previous discovery with the joint probability function, they know want to find the marginal probability distribution of Y_1, the 6 Ford vehicles. The company learned that this is hypergeometric where N = 12, n = 4, r = 6. They also noted the marginal probability distribution of Y_2, the 6 Ford vehicles, is also hypergeometric where N = 12, n = 4, r = 4. This allows the company to find the probability of Y_1 if Y_2 is true.

$$P(Y_1 = 2 \mid Y_2 = 0) = \frac{P(Y_1 = 2, Y_2 = 0)}{P(Y_2 = 0)} = \frac{\binom{6}{2}\binom{4}{0}\binom{3}{4-2-0}}{\binom{12}{4}} \div \frac{\binom{4}{0}\binom{9}{4}}{\binom{12}{4}} = \frac{5}{14}$$

= 35.71%

Chapter 5.4 Problem

The dealership could determine that the Y_1 and Y_2 of the car lot, consisting of 12 cars, from the previous problems was dependent when

$0 \le y_1 \le 2,\ 0 \le y_2 \le 2,\ and\ 1 \le y_1 + y_2 \le 2$. This is because

$P(Y_1 = 2, Y_2 = 0) \ne P(Y_1 = 2)P(\ Y_2 = 0) = (\frac{14}{55})(\frac{42}{11})$

Sources

Dataset: Vehicle Sales Data: Vehicle/Car Sales Trends and Pricing Insights

Problems inspired by Wackerly, Dennis D., et al. Mathematical Statistics with Applications. 7th ed., Brooks/Cole Cengage Learning, 2008.

Graphs and tables made using Tableau Public and Excel