Starbucks Calories Research Paper

By Nicole Wiechmann

Table of Contents

[**Introduction: 2**](#_53p9t9cx5nnj)

[**Chapter 1: 2**](#_wqfsq2l30t96)

[Mean, Median, and Mode: 2](#_1ymc1cg350qp)

[Standard Deviation and Variance: 3](#_ipb63chlt76y)

[**Chapter 2: 3**](#_nsd9rfl2zymw)

[Permutation Combinations: 3](#_woj9ithkm5yz)

[Conditional Probability: 4](#_vxn3vmdcqis7)

[Bayes: 4](#_u4md9fpd2reu)

[**Chapter 3: 5**](#_faauoqfi0afz)

[Binomial Distribution: 5](#_lq9ab9ljb6ub)

[Geometric Distribution: 5](#_7gkktw0iloc)

[Hypergeometric Distribution: 6](#_d3hsw5lpjosz)

[Negative Binomial Distribution: 6](#_3bfxwo9opjy)

[Poisson Distribution: 6](#_aojpn97jiari)

[Tchebysheffs Theorem: 7](#_7nml05bww76m)

[**Chapter 4: 7**](#_gh31wcmy4tu6)

[Uniform Distribution: 7](#_65qv6486uk69)

[**Chapter 5: 7**](#_cqzdsqxckk42)

[Joint Probability Mass Function & Joint Distribution Function: 7](#_az15jekldfgp)

[Marginal Density: 8](#_yi1itduzwgl5)

[Conditional Density: 8](#_h5e11z6f4llh)

[Independence: 8](#_jdgg4nlo7omd)

[**Conclusion: 8**](#_w2p2uin1bm1r)

[**Formulas: 9**](#_dq4b6eqjd3j)

# Introduction:

For my research paper, I decided to work with data that I already have some experience with, Starbucks drinks. When I originally found this dataset, I wanted to do my research on the prices of the drinks, as I personally believe them to be extremely overpriced and I wanted to find out just how overpriced they are. Once I had my dataset, however, I realized the dataset did not include the prices, and was more about calorie count and vitamins. Once I saw this, I decided to change my look at the data set and figure out just how many calories are in each drink, and to an extent find out just how unhealthy it would be if someone went to Starbucks every day.

# Chapter 1:

## *Mean, Median, and Mode*:

To start off my search on just how unhealthy Starbucks drinks are, I decided to look into the mean, median, and mode of all the drinks listed in the dataset. I decided for this category to keep to just using all 243 lines of drinks, even though the only difference between some were that they were iced vs hot. This, surprisingly, actually did make a decent impact on the calorie count, as if the drink was iced there would be less of the actual drink, where hot drinks didn't need room for ice so there would be more drink and more calories. For the mean, median, and mode I mostly used my StatsLibrary code, as I really did not want to add all 243 rows of calories by hand.

For the mean, I came up with an average of **193.87 calories**. This was actually a lot more than I expected, as I know there are some teas Starbucks has that are 0 calories, so I expected these to skew the data a little bit closer to 100. That being said, with the drinks with 0 calories and the average still being closer to 200 calories per drink, that should show the first clear sign as to how unhealthy the Starbucks drinks really are.

Next I focused on the median calorie count. With the median, I was hoping to see if the drinks with 0 calories really did skew the dataset, as I expected the median to be closer to 150 with the 5-10 drinks that were 0 calories. For the median, I came up with a median of **185 calories.** I was once again shocked to see the median was closer to 200 once again and genuinely did not expect this. At one point I actually thought maybe my program was not working so I used the built in median function in excel to double check and it was in fact 185 calories. To me this shows just how many drinks have calories well over 150-200 to the point the drinks with 0 calories barely affected the data. Should the drinks with 0 calories not have been included in the dataset, I believe the median would be closer to 210-240 range, which to me for a single drink is extremely unhealthy.

Lastly, I focused on the mode of the dataset, as this one should be the only value not affected by the extreme highs or extreme lows of the calorie counts. I ended up getting a mode of **150 calories**, which was very much what I originally expected. As unhealthy as I believe Starbucks drinks are, I felt that 150 calories is not nearly as bad as I originally feared it would be before I looked through the dataset and after watching how high the mean and median values were.

## *Standard Deviation and Variance:*

When dealing with standard deviation and variance, I once again used my StatsLibrary program to find my values, as going though individually would take far too long. For variance, my program came up with a variance of 10537.136483163715, which at the time just seemed wrong. I decided to let my program continue and find the standard deviation, which came out to be **120 calories**. Once I saw this, I understood why my variance was such a large number, since I am working with numbers that are almost all over 100. To me, a standard deviation of 120 does make sense, and shows just how large the calorie counts get with the drinks.

# Chapter 2:

## *Permutation Combinations:*

For this section, I decided to look at the different types of beverages and look to see just how many have at least one version of the drink that is under 100 calories. As I will discuss further in the next section, I came up with 34 types of beverages, where 13 of those beverages had a version that was under 100 calories. I decided to experiment with 5 drinks and asked myself the question: “If 5 drinks were picked at random, what would the probability be of all 5 of them being a drink under 100 calories?” When I did the formula, please look below at the “formula” page should you want to see the work, I came up with a percentage of **.46%**. While I expected the percentage to be low, I did not expect it to be under 1%. To me, this shows how big the difference is when only 13 types of drinks have a version under 100 calories. Within this .46%, you must also keep in mind that these are only versions of the drinks, should someone get that same drink in a larger size or with whipped cream, the calories go back up to over 100. So in actuality, it is a .46% chance to get a drink that has a version that is under 100 calories, but not guaranteed. This can also be said for consumers that get at least 5 drinks a day or within a week, they are extremely unlikely to be getting a drink that is under that 100 calorie threshold every time, especially if you take into account that usually people will get the same drink every time they go to a store.

## *Conditional Probability:*

For conditional probability, I had to do a little bit more research into my dataset, as I needed to extract numbers for everything including the types of beverages, the calories under 100, and how many coffees total I had in the dataset. What I ended up with was a total of 243 coffees, 34 beverage types, a total of 37 calories under 100, and 13 types of beverages with a version that is under 100 calories. Once I found this information, I was able to start figuring out the conditional probability of getting a drink under 100 calories given that there are 34 types of beverages within the 243 coffees. What I ended up getting was that there was a **35.135%** chance of getting one of the coffees under 100 calories. While this is much higher than the previous section, it must be kept in mind that this means that there is only a 35% chance that someone can get a version that is under 100 calories out of all 243 coffees. That means the other 65% of the time you would be getting a drink that is either 100 calories or exceeds that threshold, showing how many drinks are much over 100 calories.

## *Bayes:*

For Bayes formula, I once again used the same numbers as I did with conditional probability. I also got around the same percentage for the probability of getting a dirk under 100 calories, that being **32.29%**. The fact that the percentage actually went down by using bayes is amazing to me, as I already thought that it was a pretty low percentage before. With this, you can see that the percentages for getting drinks with under 100 calories will continue to decrease, and it shows just how unhealthy Starbucks drinks are for someone getting them once or twice a day.

# Chapter 3:

## *Binomial Distribution:*

With Binomial distribution and all the distributions going forward, I once again started using my StatsLibrary program mixed with having to go into the CSV file to do more research. First, I had to find the exact percentage of beverages that have over 100 calories in them, which I found to be 84.77%. This helped me pretty much for every distribution going forward as well. For binomial distribution I decided to ask the question: “If there is a 84.77% chance of a beverage type being over 100 calories and 10 drinks are sampled, what is the probability that exactly 8 of them are over 100 calories? Once I did the calculations, I got the answer of **27.83%**. Now from a glance this does not seem to help the idea that Starbucks drinks have an unhealthy amount of calories in them, since the percentage is low, but it would be a mistake to think that. In this question, we are sampling 10 drinks with a percentage of 84.77% being over 100 calories. The fact that there is over a 20% chance of 8 out of the 10 drinks being over 100 calories is extremely bad for consumers, as that percentage should theoretically be at most 10%, but should be much less.

## *Geometric Distribution:*

For this distribution, I actually decided to ask two different theoretical questions. The first question being: “If 84.77% of beverage types are over 100 calories, what is the probability that the third drink is the first to be over 100 calories?” To this, I got the answer of **1.97%**. Once again, at a glance this seems to be screaming “Oh Starbucks drinks are not as unhealthy as it seems” but once again then would be a delusion. This says that there is a 1.97% chance that the ***third*** drink is the ***first*** to be over 100 calories, keywords being first drink. This means that there is a much higher chance that the first or second drink would be over 100 calories, and that it's very unlikely that someone’s third drink would be the first drink they’ve gotten that's over 100 calories.

The second question I asked was: “If 84.77% of beverage types are over 100 calories, what is the probability that the first drink over 100 calories is randomly selected on or before the 4th selection?” For this question I came to an answer of there being a **99.95%** chance that someone would randomly get a beverage over 100 calories. This was a big wake up call after all the other percentages I was getting, as this was the biggest indication of how high the calorie count is.

## *Hypergeometric Distribution:*

For hypergeometric distribution, I decided to look at the differences between hot and cold drinks instead. As stated above, hot drinks actually have more calories than their cold counterparts. Knowing this, I wanted to see what the probability of getting one over the other would be. I found the data values to be 34 beverage types, 17 hot drinks, and 17 cold drinks. I then asked the question: “If 7 drinks are picked out of the 34 types of drinks without replacement, what's the probability that exactly 5 drinks are cold?” I then got the answer of **15.64%** chance of all 5 drinks being cold. This shows that there is a low possibility of getting 5 cold drinks when randomly selecting your drinks, meaning there is a higher possibility of getting more hot drinks, and therefore more calories.

## *Negative Binomial Distribution:*

For negative binomial distribution, I went back to looking more at the calories themselves and less on the hot and cold debate. With this one I decided to ask myself: “If I randomly pick a drink with a .1523 probability of getting a drink under 100 calories, what's the probability that the second drink under 100 calories comes on the fourth drink ordered?” The answer I got was a flat **5%** chance. While this can mean a lot, such as maybe you can get two drinks in a row that were under 100 calories, to me this seems more like there is a low percent chance you will get a second drink that is under 100 calories.

## *Poisson Distribution:*

For poisson distribution, I wanted to focus more on the people that get multiple drinks a week. For this distribution I asked the question: “If a customer gets 1 drink every 4 days, what is the probability that they would get 2 drinks in 4 days?” I then got an answer of there being an **18.39%** chance that they would get more drinks within that 4 day period. This is actually the first distribution that does not help the point I am trying to convey in this paper, as it seems there is a low chance of people getting more than their average amount of drinks during a given period of time. The drinks they get during that time, however, have a high likelihood to be much more than the average 100 calories.

## *Tchebysheffs Theorem:*

For the last part of the chapter 3 distributions and theorems, I decided to ask the question: “What percentage of calories falls between 100 and 400 with a mean of 193 and standard deviation of 121? When I plugged this into my program, I got the answer of **53.72%**. This means that around 53% of the calories were within 100 and 400 calories, and knowing what I already did research on knowing there are only 37 drinks with under 100 calories, there are even more drinks that are around over 400 calories.

# Chapter 4:

## *Uniform Distribution:*

When getting into chapter 4 distributions, it gets a little bit more complicated. With my StatsLibrary program, however, it makes it much easier to understand the answers to my theoretical questions. For uniform distribution I asked the question: “if on average, a starbucks drink is between 153 & 210 calories, what is the probability the drink being ordered is in the 200’s?” I started this chapter with calories in the 200’s because I felt like I had made my point with calories over 100, and wanted to show just how many drinks are over 200 instead. To this question, I got the answer of there being a **17.54%** chance of a drink being over 200. This actually does make sense, as the majority of drinks are between 153 and 200, but this also shows there's a good amount of drinks that are around over the 200 calorie range.

# Chapter 5:

## *Joint Probability Mass Function & Joint Distribution Function:*

Once into chapter 5, the theoretical questions become much harder to come up with. For joint probability mass function and distribution, I decided to figure out the probability if there were 3 different beverage choices available and 2 customers walk in and pick a drink at random. I wanted to find the total probability if the customers picked drink 1 and drink 2. The answer I got was about an **88.88%** probability that drinks one and two were chosen. Now thinking about it logically, what if it was set up so that drink 1 was a drink between 100-200 calories, drink 2 was a drink over 200 calories, and drink 3 was a drink under 100 calories. Thinking about it like this, there is also a likely possibility that the customers would pick drink 2 and 3, but overall this does not affect the likelihood of how unhealthy the drink is.

## *Marginal Density:*

For marginal density, I decided to ask the question: “What would the marginal density be if x = # of calories in a hot drink and y = # of calories in a cold drink for ?” As you can probably tell, this is when coming up with the questions was extremely difficult. The answer I came up with for this was that there is a marginal density of **5y and 5x**. To me, this means that there is an equal chance of getting the same number of hot drinks as cold drinks.

## *Conditional Density:*

For conditional density, I decided to try and find the conditional density of x given y = y for the same equation above for the hot vs cold drink debate. The answer I got was that for hot drinks, the conditional density would be about **2x**. To me, this means that for every y there will be 2x, so for every cold drink ordered, two hot drinks would be ordered, meaning a higher calorie intake.

## *Independence:*

Lastly, I decided to see if the number of hot drinks were independent of the number of cold drinks, while excluding the frozen drinks. I decided to exclude the frozen drinks because I wanted to see just how much of a difference the frozen drinks had on the hot vs cold drink debate. They ended up being very **dependent** on each other, but once frozen drinks were added in they became independent of each other.

# Conclusion:

In conclusion, I learned a lot about just how unhealthy an amount of calories are within a good portion of the drinks Starbucks gives out. Whether it has to do with the amount of sugar within the drinks, whether they are hot or cold or frozen, or how big of a size you get, they all have much too many calories within them. This has been very important knowledge for me because over the past couple days since starting this research paper and project, I have increased my intake of Starbucks drinks. Now that I know exactly how much the difference is between the “healthy” drinks and the drinks well over 100 to 200 calories each, I as well as many others should consider holding back on ordering more Starbucks in the future.

# Formulas:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Permutation Combinations: if 5 different types of beverages are picked, whats the probability that they are all under 100 calories

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Conditional Probability: 243 coffees, 34 beverages, 37 calories under 100, 13 beverages with calories under 100

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Bayes: 243 coffees, 34 beverages, 37 calories under 100, 13 beverages with calories under 100

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Binomial Distribution: 84.77% of beverages are over 100 calories. If 10 are sampled, what's the probability that exactly 8 are above 100 calories?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Geometric Distribution: 84.77% of all beverages are over 100 calories. What’s the probability of the third drink being the first to be over 100 calories?

Probability that the first drink over 100 calories is randomly selected on or before the 4th selection?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Hypergeometric Distribution: 34 beverage types, 17 hot & 17 cold. If 7 drinks are picked without replacement, what is the probability of exactly 5 picked are cold?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Negative Binomial Distribution: Randomly order a drink with a .1523 probability of getting a drink under 100 calories. Find the probability that the second drink under 100 calories comes on the fourth drink ordered

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Poisson Distribution: Customer receives about 1 drink every 4 days. Whats the probability they will get 2 drinks in 4 days?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tchebysheffs Theorem: What percentage values fall between 100 and 400 with mean 193 and standard deviation of 121?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Uniform Distribution: On average, a starbucks drink is between 153 & 210 calories. Find the probability the drink being ordered is in the 200’s.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Joint Probability Mass Function & Joint Distribution Function: 3 different beverage choices are available and 2 customers walk in and pick a drink at random. Find F(1,2).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Marginal Density: Let x = # of calories in an hot drink and y = # of calories in a cold drink.

If , find the marginal density.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Conditional Density: Find the conditional density of x given y = y of:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Independence: Is the # of hot starbucks drinks independent of the # of iced starbucks drinks, not including frozen drinks?

X = hot starbucks drinks =

Y = iced starbucks drinks =

They are dependent.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_