Stat 201: Introduction to Statistics

Standard 8: Numerical Summaries – Empirical Rule

Chapter Two

Summaries

From *Naked Statistics: Descriptive Statistics*

- "The standard deviation is the descriptive statistic that allows us to assign a single number to this dispersion around the mean."

- "The beauty of the normal distribution- its Michael Jordan power, finesse, and elegance – comes from the fact that we know by definition exactly what proportion off the observations in a normal distribution lie within one standard deviation of the mean (68%), within two standard deviations of the mean (95%), within three standard deviations of the mean (99.7%)."

The Empirical Rule

- About 68% of data fall within 1 standard deviation of the mean
- About 95% of data fall within 2 standard deviation of the mean
- About 99.7% of data fall within 3 standard deviation of the mean
- **The distribution must be symmetric and bell shaped to use this Rule**

The Empirical Rule with z-scores

- About 68% of data fall between z=-1 and z=1
- About 95% of data fall between z=-2 and z=2
- About 99.7% of data fall between z=-3 and z=3

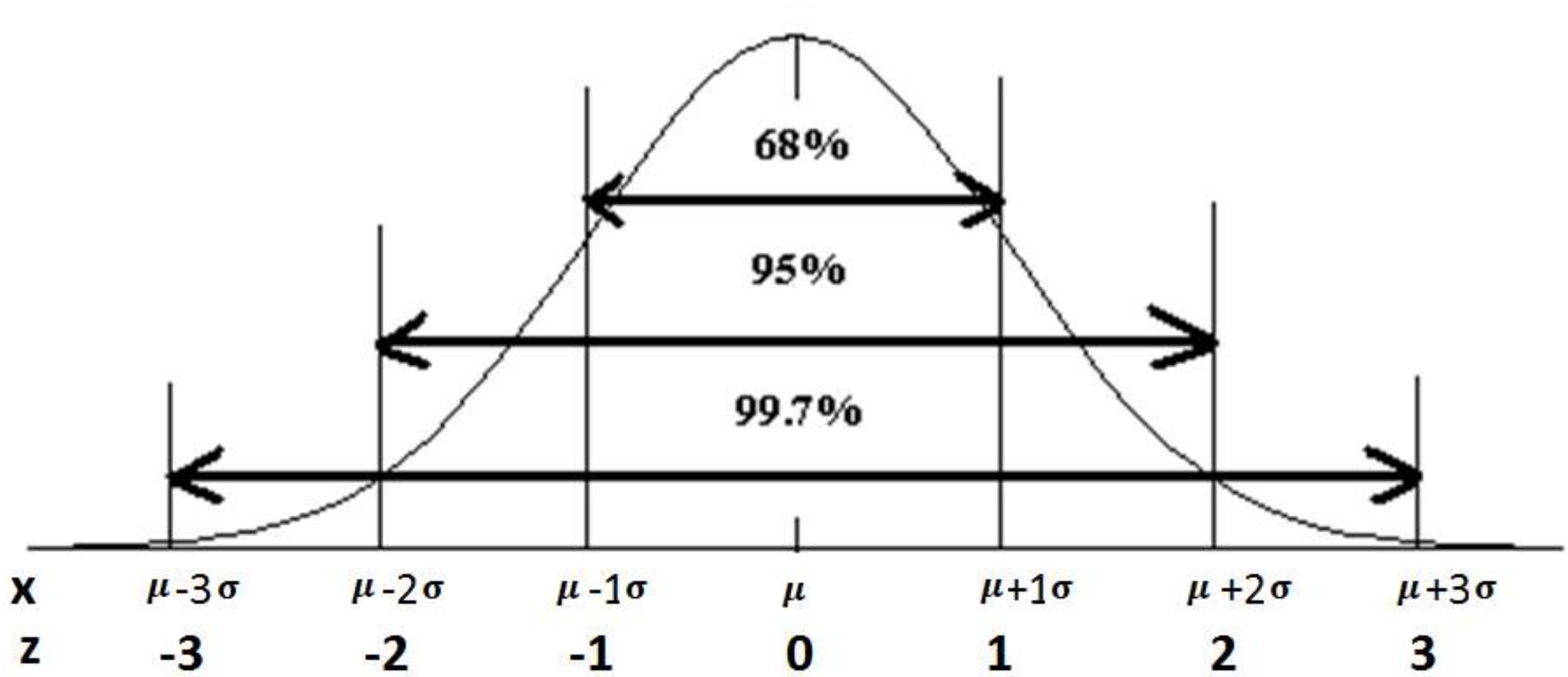
- **The distribution must be symmetric and bell shaped to use this Rule**

Z Score: How Do We Calculate It?

- $z = \frac{observation - mean}{standard\ deviation}$

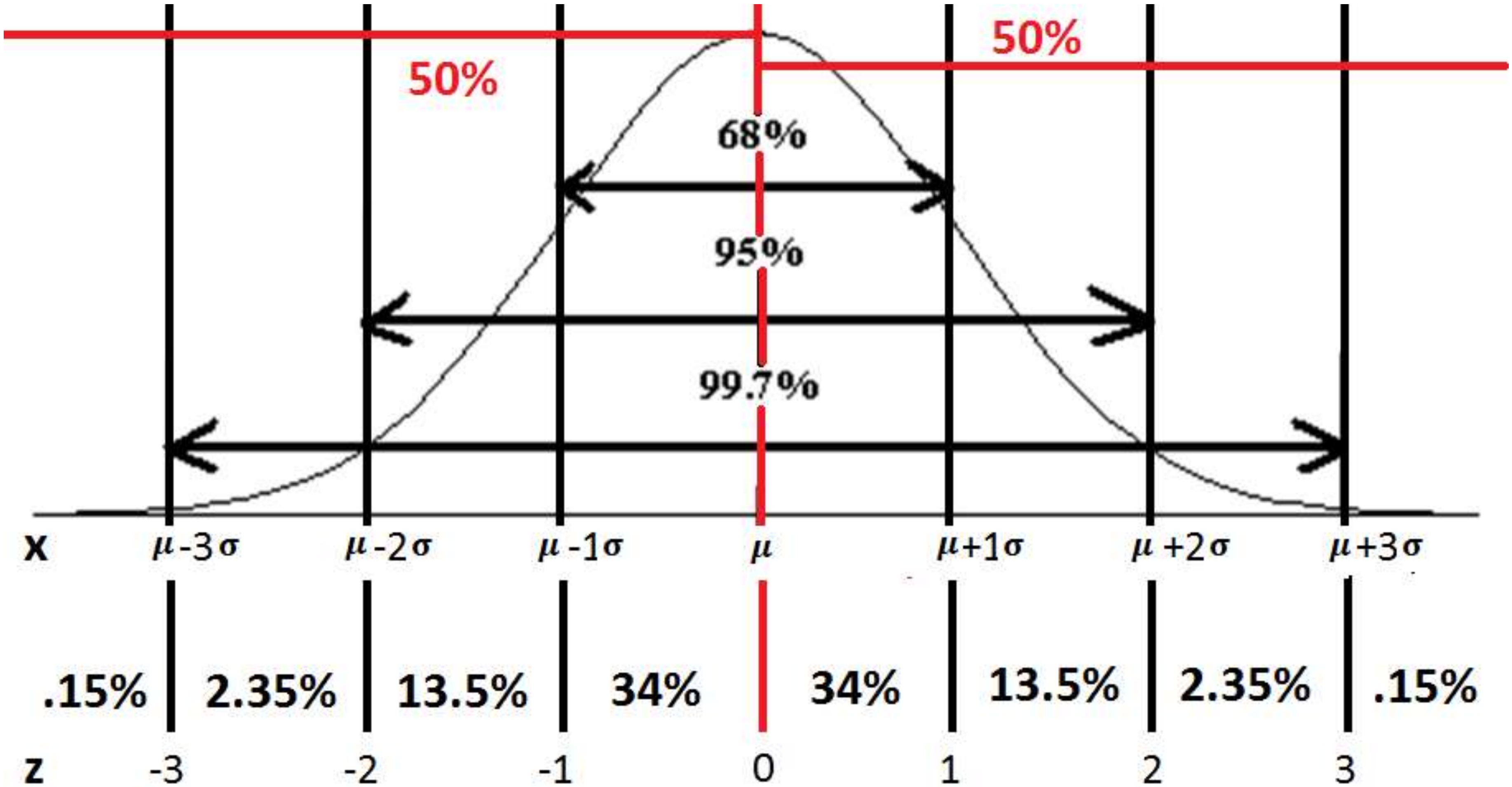
- This gives us the number of standard deviations from the mean the observation is

- **Note: we consider any observation with a Z score above 3 or below -3 an outlier**



68%
95%
99.7%
x
μ-3σ
μ-2σ
μ-1σ
μ
μ+1σ
μ+2σ
μ+3σ
z
-3
-2
-1
0
1
2
3

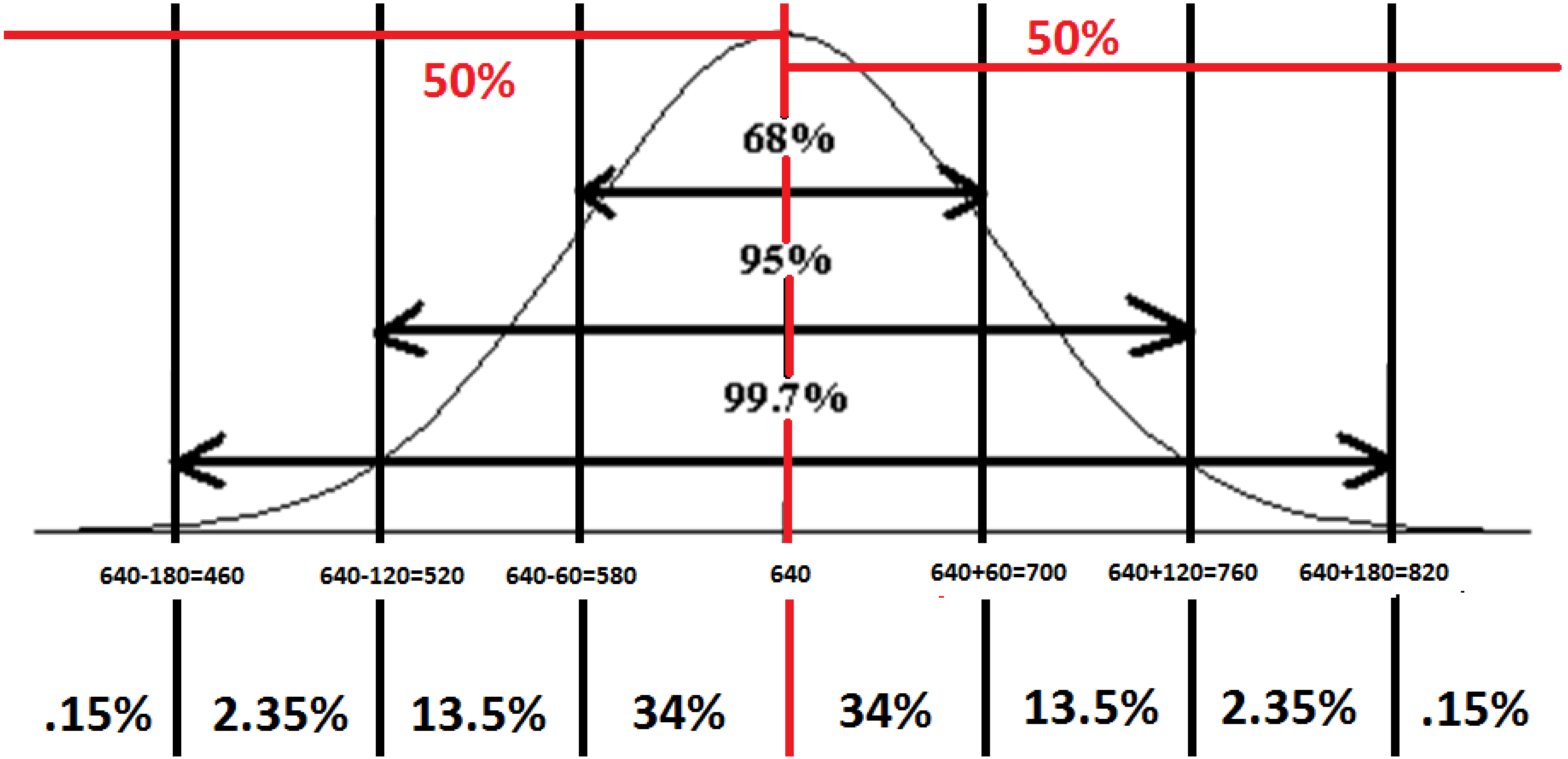
Empirical Rule



Walkthrough

The Empirical Rule: Example

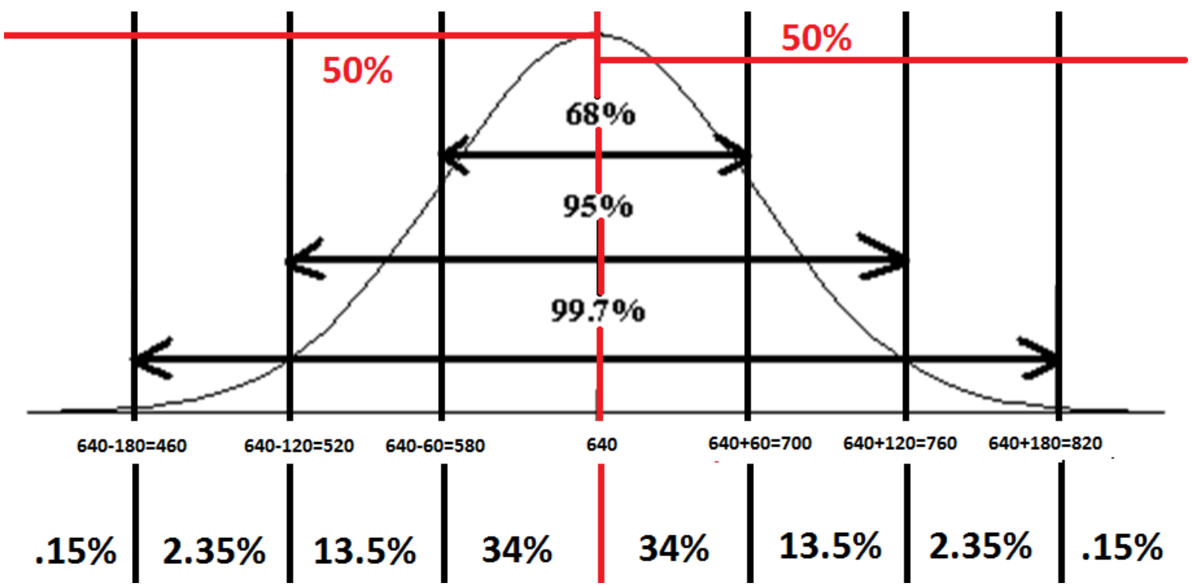
- The average college student consumes 640 cans of beer each year. Assume the distribution of cans of beers consumed per college student is **bell-shaped** with a **mean of 640 cans** and a **standard deviation of 60 cans**.



50%
50%
68%
95%
99.7%
640-180=460
640-120=520
640-60=580
640
640+60=700
640+120=760
640+180=820
.15%
2.35%
13.5%
34%
34%
13.5%
2.35%
.15%

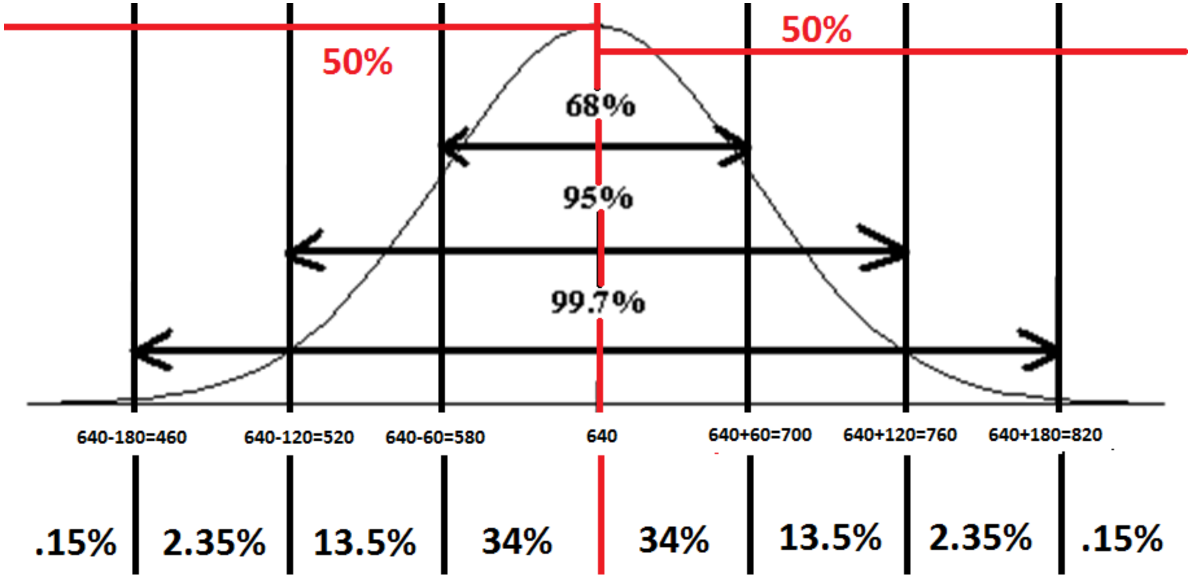
The Empirical Rule: Example

- What percent of students consume less than 700 cans of beer per year?



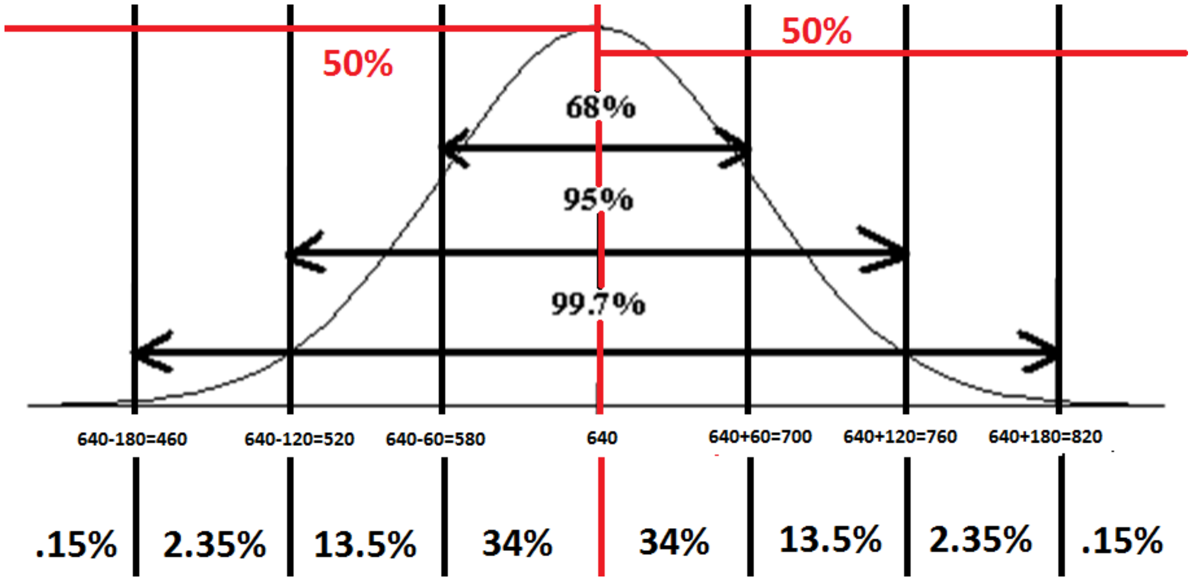
The Empirical Rule: Example

- What percent of students consume less than 700 cans of beer per year?
- We can add up the area under the curve as we go left
  2.5%+13.5%+34%+34%+.15%
  = 84%



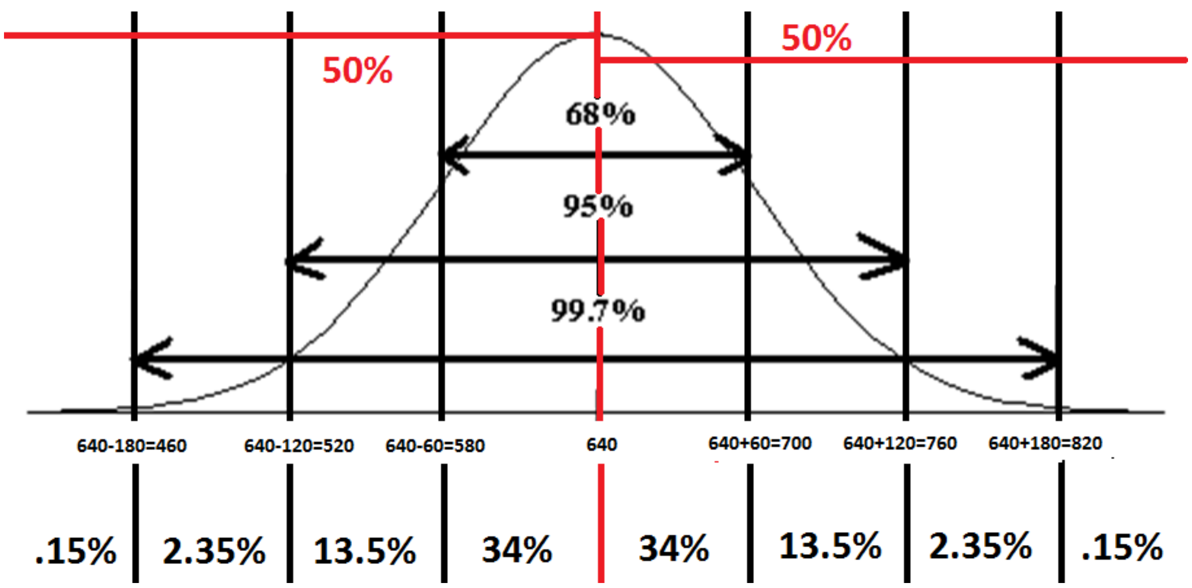
The Empirical Rule: Example

- What percent of students consume less than 700 cans of beer per year?
- We can subtract the area from 100% as we go right
  100%-13.5%-2.5%-.15%
  = 84%



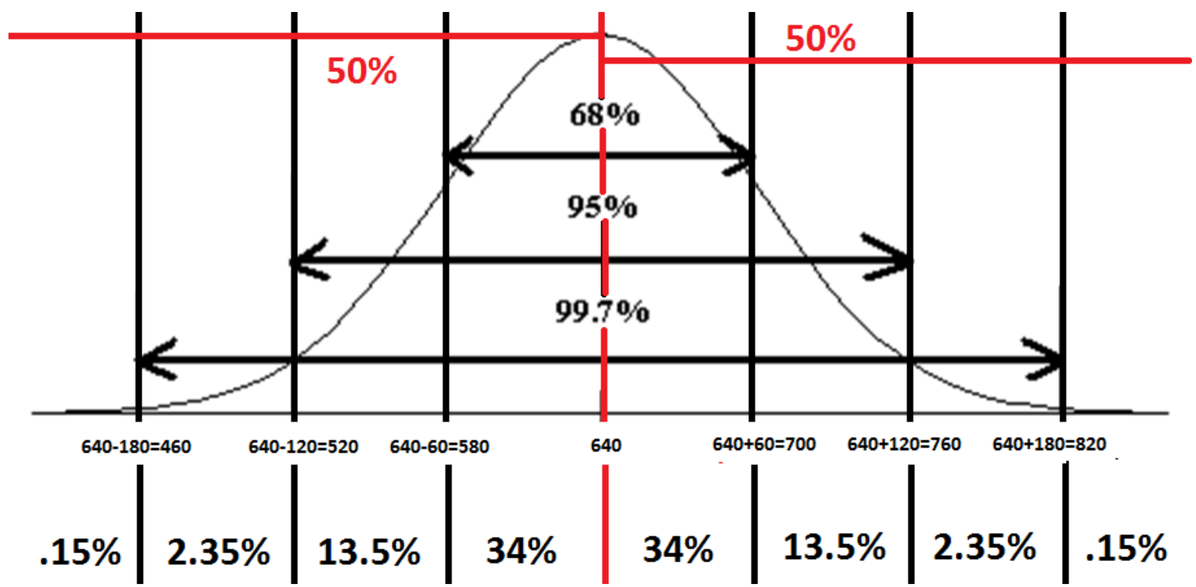
The Empirical Rule: Example

- What percent of students consume more than 700 cans of beer per year?



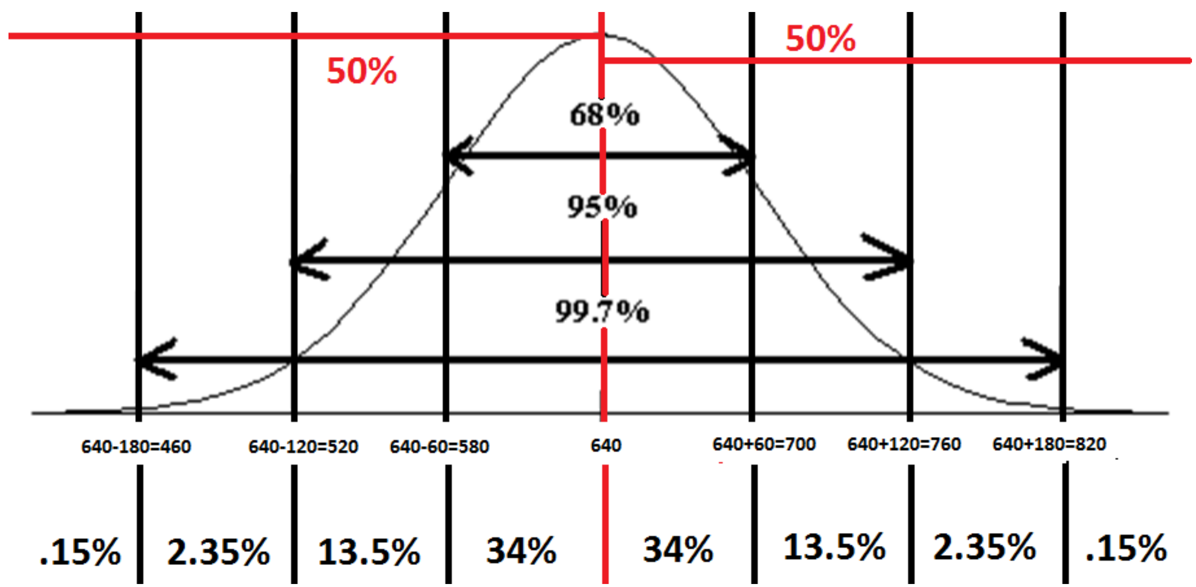
The Empirical Rule: Example

- What percent of students consume more than 700 cans of beer per year?
- We can add up the area under the curve as we go right
  13.5% + 2.35% + .15% = 16%



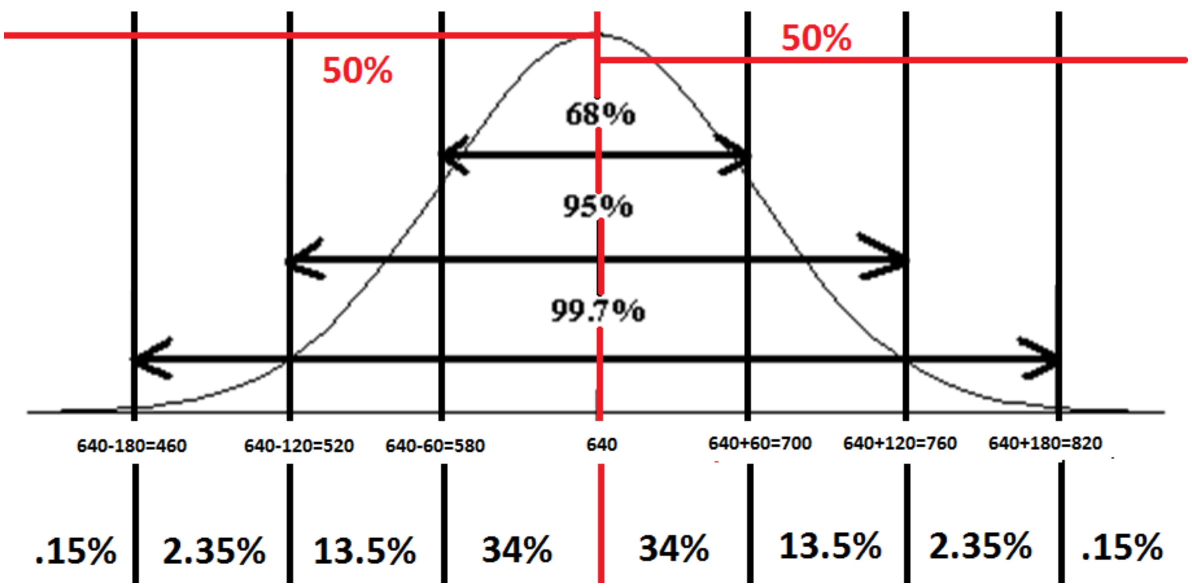
The Empirical Rule: Example

- What percent of students consume more than 700 cans of beer per year?
- We can subtract the area from 100% as we go left
  100%-34%-34%-13.5%-2.5%-.15%
  100%-84% (we know 84% from the last question)
  = 16%



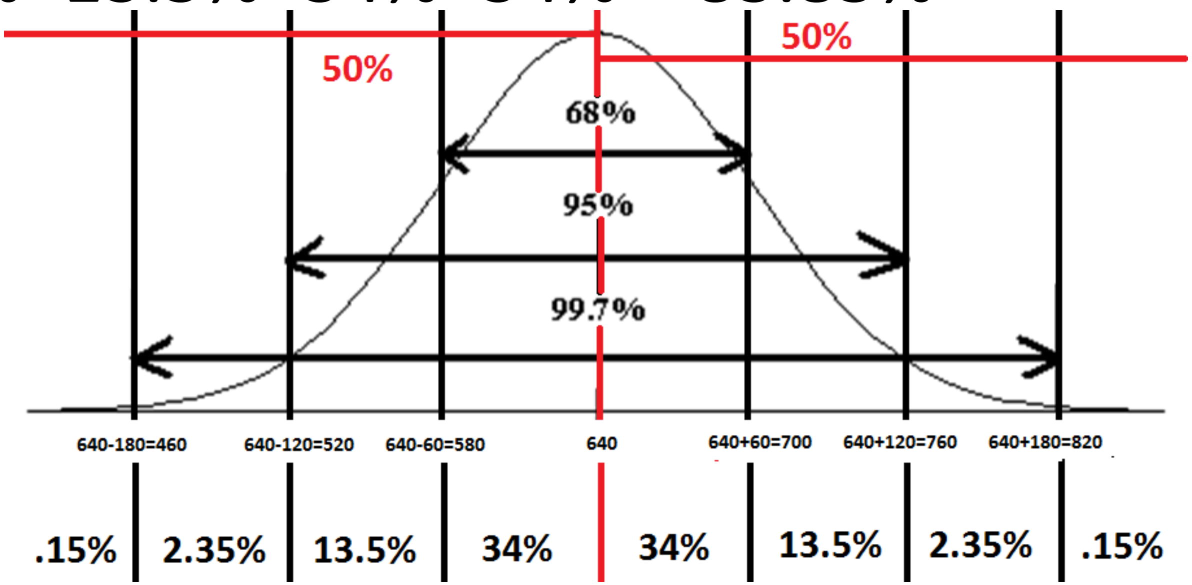
The Empirical Rule: Example

- What percent of students consume between 460 and 700 cans of beer per year?



The Empirical Rule: Example

- What percent of students consume between 460 and 700 cans of beer each year?
- We can add up the area under the curve as we go from 460 to 700
  2.35%+13.5%+34%+34% = 83.85%



Z Score: If you don’t know what it is you can’t afford it.

- What happens when we’re interested in percentiles and x values that aren’t perfectly spaced according to the empirical rule?
- We note that in most scenarios the data we’re concerned with will fit this scenario.
- Later, in chapter 6, we will use the z-score to find these ‘in-between’ probabilities and percentiles so pay attention!
- **For now, we use z-scores to define outliers and to rewrite the empirical rule.**

Z Score: What are we doing here?

- What did we do with the empirical rule?
  - We looked at how many standard deviations away the data values were
- The idea here is to be able to find out how many standard deviations the data values we're looking at are from the mean but we allow fractional answers
  - answers outside of -3, -2,-1, 0 , 1, 2, 3 which the empirical rule covers

Z Score: How Do We Calculate It?

- $z = \frac{observation - mean}{standard\ deviation} = \frac{x - \mu_x}{\sigma_x}$

- This gives us the number of standard deviations from the mean the observation is

- **Note: we consider any observation with a Z score above 3 or below -3 an outlier**

Z Score: Example

- The average college student consumes 640 cans of beer per year. Assume the distribution of beers consumed per year per college student is **bell-shaped** with a **mean of 640 cans** and a **standard deviation of 60 cans**.

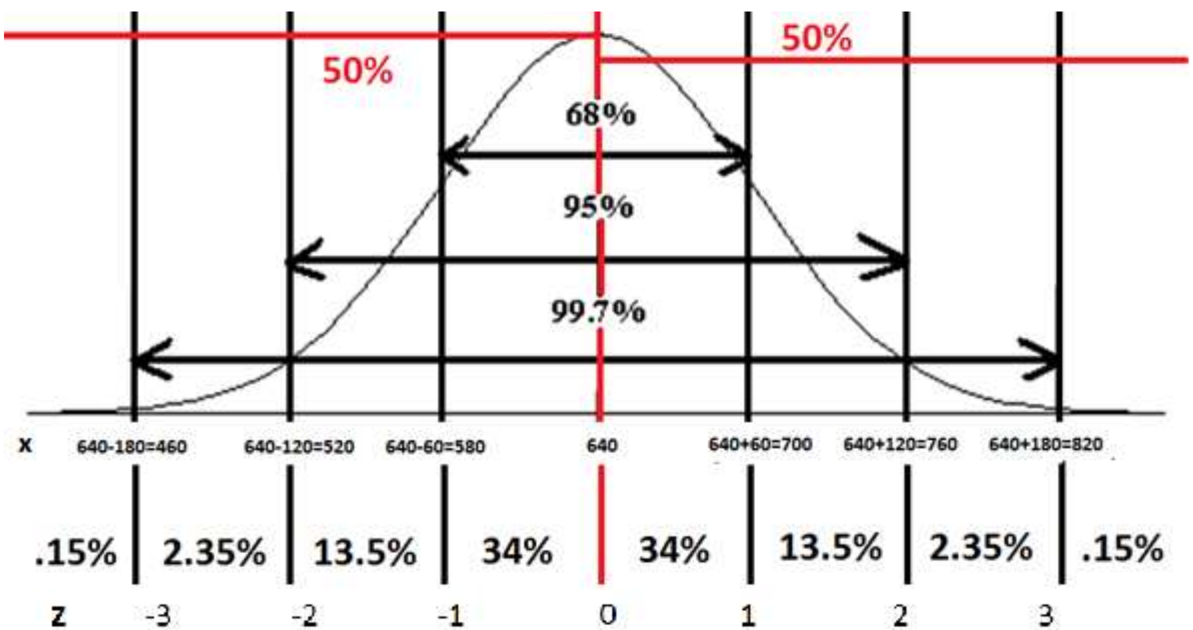
Z Score: Example

- $z_{460} = \dfrac{460 - 640}{60} = \dfrac{-180}{60} = -3$
- $z_{820} = \dfrac{820 - 640}{60} = \dfrac{180}{60} = 3$

- Note the Z score has given us the correct number of standard deviations from the mean for each case!

Z Score: Example

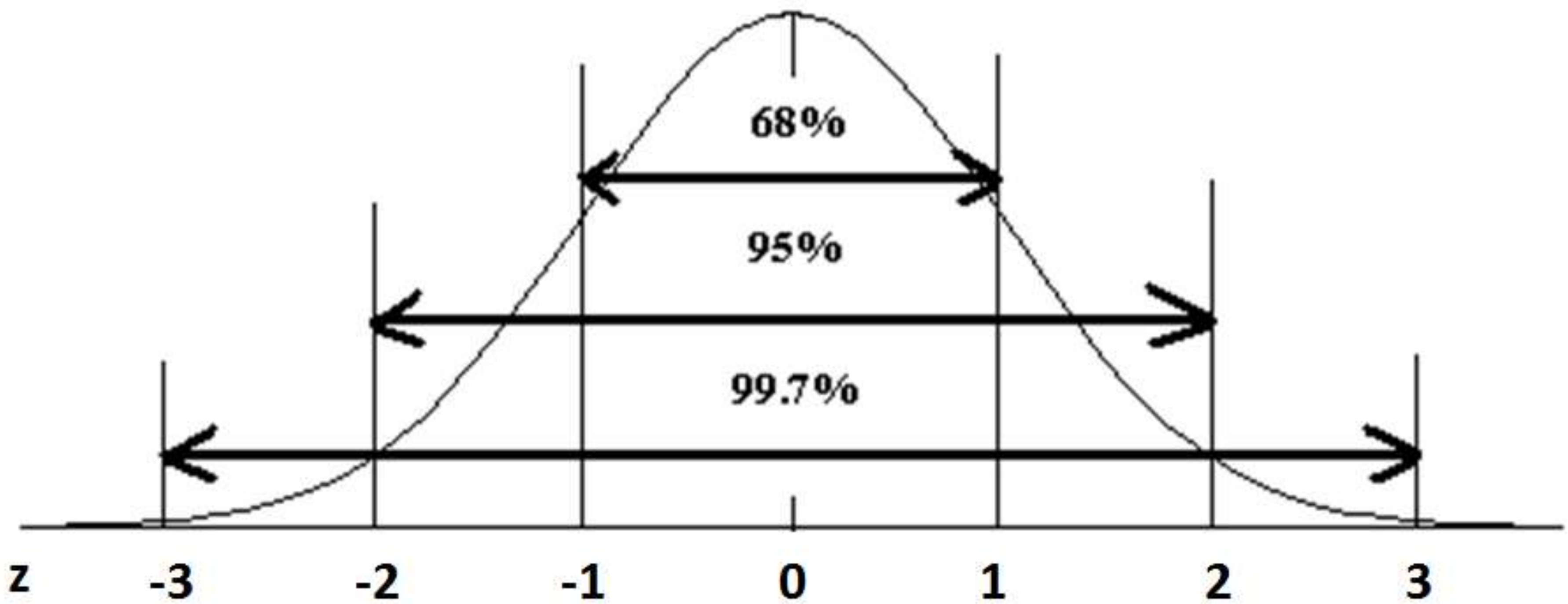
- Recall from the Empirical Rule that about 99.7% of college students consume between 460 and 820 cans of beer per year (+- 3 standard deviations)



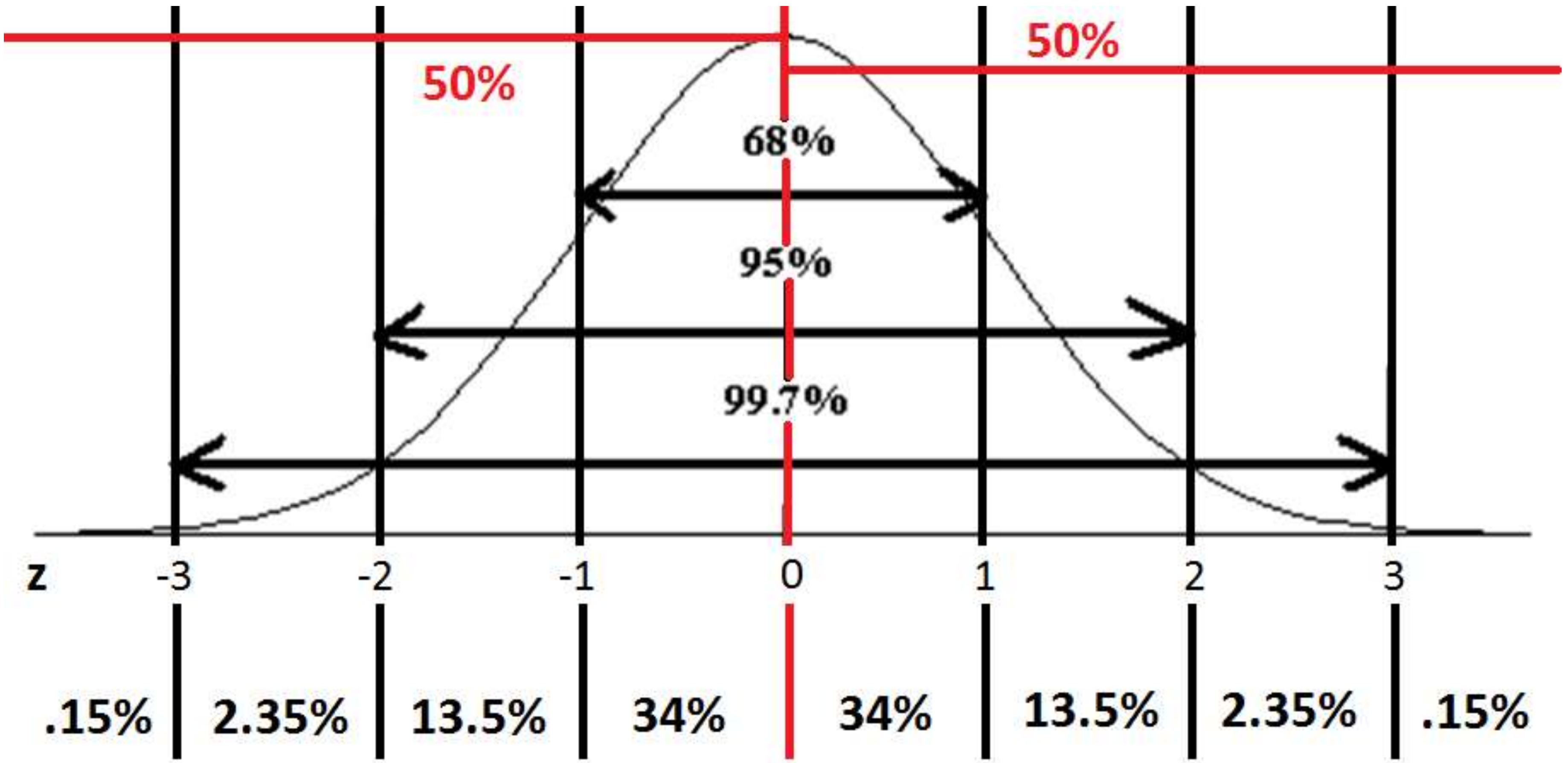
The Empirical Rule with z-scores

- About 68% of data fall between z=-1 and z=1
- About 95% of data fall between z=-2 and z=2
- About 99.7% of data fall between z=-3 and z=3

- **The distribution must be symmetric and bell shaped to use this Rule**


68%
95%
99.7%
z
-3
-2
-1
0
1
2
3

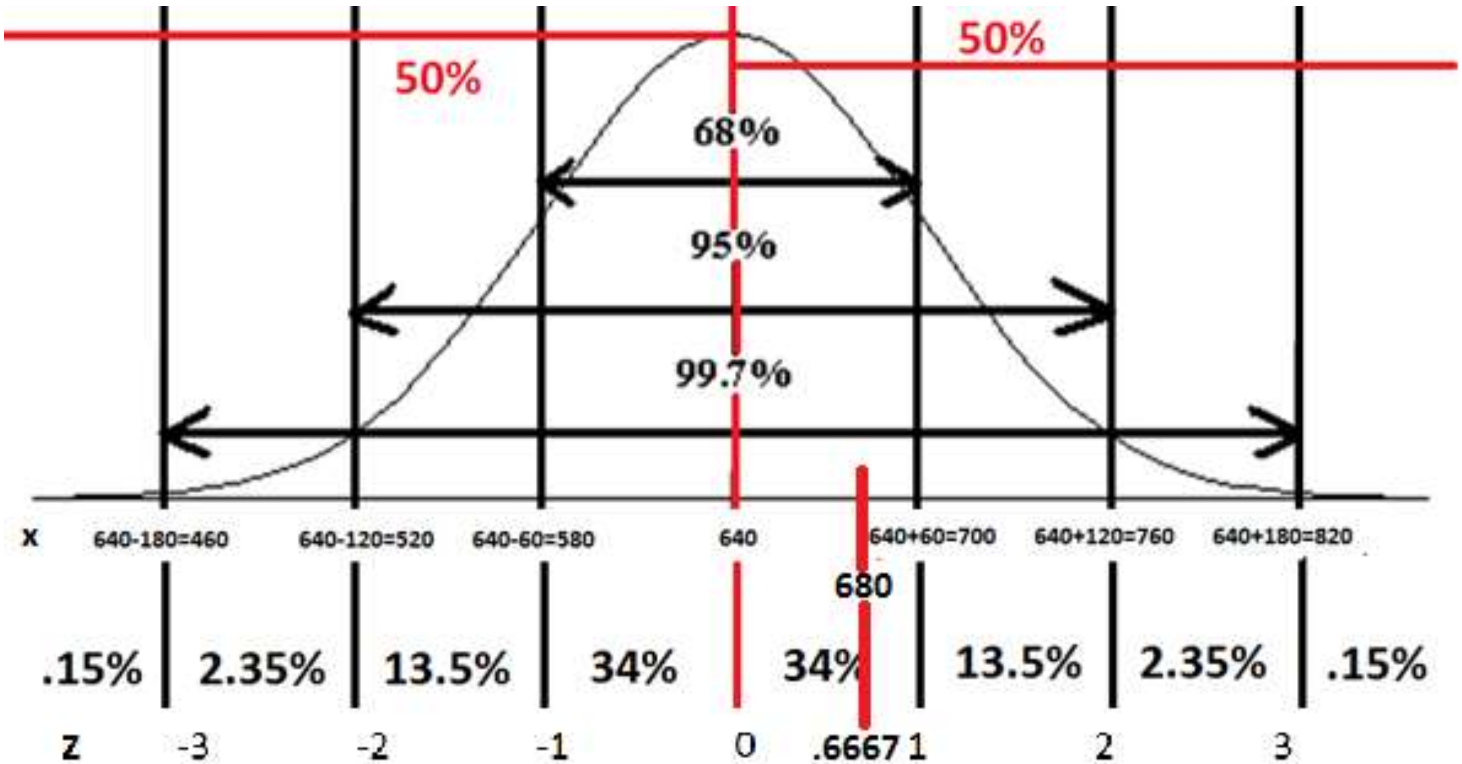
Empirical Rule



Z Score: Example 2

- Let’s consider an observation of 680 cans of beer.
  - 680 is not 1, 2, or 3 standard deviations away
  - $z = \frac{680-640}{60} = .6667$
    - X=680 is .6667 standard deviations above the mean
    - .6667 indicates this observation is not an outlier because .6667<3 and .6667>-3
    - We will be able to find these percentages in chapter 6 so don’t forget z-scores!

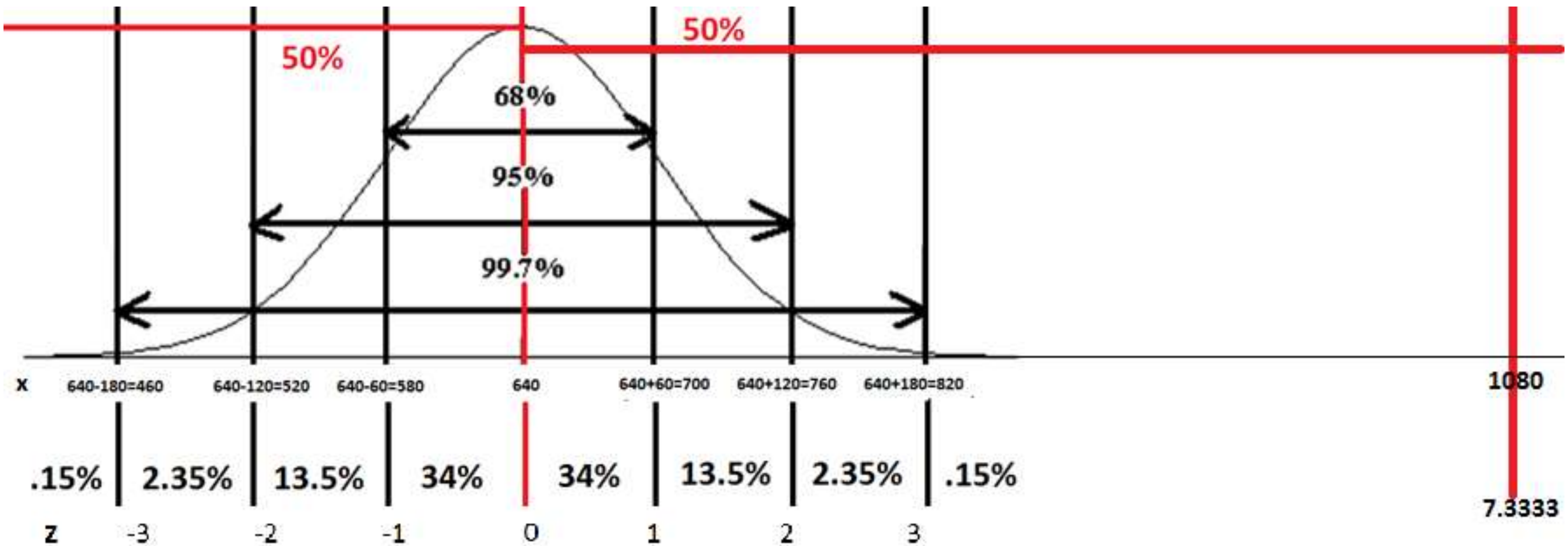
Z Score: Example 2



Z Score: Example 3

- Let's consider an observation of 1080 cans of beer.
  - 1080 is not 1, 2, or 3 standard deviations away
  - $z = \frac{1080-640}{60} = 7.3333$
    - X=1080 is 7.3333 standard deviations above the mean
    - Z=.7333 indicates this observation is an outlier because 7.3333>3

Z Score: Example 3



Z Score: Example 4

- Let's consider an observation of 500 cans of beer.
  - 500 is not 1, 2, or 3 standard deviations away
  - $z = \frac{500-640}{60} = -2.3333$
    - X=500 is 2.3333 standard deviations below the mean
    - -2.3333 indicates this observation isn't an outlier because -2.3333<3 and -2.3333>-3

Z Score: Example 3