DISTRIBUTIONS

Analytics Primer

PROBABILITY DISTRIBUTION

Random Variables

- A **random variable** is a numerical description of the outcome of an experiment.
- They can be either discrete or continuous.
- A **discrete random variable** may assume either a finite number of values or an infinite sequence of values.

Random Variables

- A **random variable** is a numerical description of the outcome of an experiment.
- They can be either discrete or continuous.
- A **discrete random variable** may assume either a finite number of values or an infinite sequence of values.
  - Finite example: Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}

Random Variables

- A **random variable** is a numerical description of the outcome of an experiment.
- They can be either discrete or continuous.
- A **discrete random variable** may assume either a finite number of values or an infinite sequence of values.
  - Infinite example: Let x be the number of customers arriving in one day at a small department store where x can take the values of 0, 1, 2, …

Random Variables

- A **random variable** is a numerical description of the outcome of an experiment.
- They can be either discrete or continuous.
- A **discrete random variable** may assume either a finite number of values or an infinite sequence of values.
- A **continuous random variable** may assume any numerical value in an interval or collection of intervals.

Discrete vs. Continuous

- Discrete Example:
  - Let x be the number of individuals living in a home.

- Continuous Example:
  - Let x be the distance in miles from home to the store.

DESCRIBING DISTRIBUTIONS

Different Numerical Measures

- Center / Location
- Variability
- Distribution Shape
- Detecting Outliers

DESCRIBING DISTRIBUTIONS

Center / Location

Mean

- The **mean** of a data set is the average of all the data values.
- The sample mean $\bar{x}$ is the point estimator of the population mean μ.

Population Mean

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Sum of the values of the *N* observations

Number of the observations in the population

Sample Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Sample Mean

Sum of the values
of the *n* observations

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Sample Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Sum of the values of the *n* observations

Number of the observations in the sample

Median

- The **median** of a data set is the value in the middle when the data items are arranged in ascending order.
- Whenever a data set has extreme values, the median is the preferred measure of central location.
- For example, annual income and property value data.
- A few extreme large incomes or property values can inflate the mean, **but not the median**.

Median

- For an **odd number** of observations:

26	30	27	22	24	29	13

Original Data

Median

- For an **odd number** of observations:

26	30	27	22	24	29	13	Original Data
13	22	24	26	27	29	30	Ascending Order

Median

- For an **odd number** of observations:

26	30	27	22	24	29	13	Original Data
13	22	24	26	27	29	30	Ascending Order

Median

Median

- For an **even number** of observations:

26	30	27	22	24	29	13	27	Original Data

13	22	24	26	27	27	29	30	Ascending Order

$$\text{Median} = \frac{26+27}{2} = 26.5$$

Percentiles

- A percentile provides information about how the data are spread over the interval from the smallest values to the largest value.
- The ***p*^th^ percentile** of a data set is a value such that at least *p* percent of the items take on this value or less and at least (100 - *p*) percent of the items take on this value or more.
- For example, Bob's SAT score was in the 93rd percentile.

Percentiles Calculation

1. Arrange the data in ascending order.
2. Compute the position index, *i*, of the p^{th} **percentile** with:

$$i = \frac{p}{100} * n$$

3. Do one of the following:
   - If *i* is not an integer , then round up. The p^{th} **percentile** is the value in the i^{th} position.
   - If *i* is an integer the p^{th} **percentile** is the average of the values in the in the i^{th} position and $(i + 1)^{th}$ position.

NOTE: Quantiles are the same thing as percentiles, except quantiles are between 0 and 1 (so 93rd percentile is the 0.93 quantile).

Quartiles

- Quartiles are specific percentiles that are commonly used.
- The **first quartile, Q_1**, is the **25th percentile**.
- The **second quartile** is the **50th percentile**, which we already defined – the **median**.
- The **third quartile, Q_3**, is the **75th percentile**.

DESCRIBING DISTRIBUTIONS

Variability

Measures of Variability

- Location and center can only get you so far with describing a distribution.
- Typically, we also consider how spread out a data set is as well.
- This is called **variability** or dispersion.

Range

- The **range** of a data set is the difference between the largest and smallest values.
- Although this is an easy calculation, it can be very sensitive to observations with extreme values as it only focuses on the largest and smallest values in the data.

Interquartile Range (IQR)

- The **interquartile range** (**IQR**) of a data set if the difference between the third and first quartiles:

$$IQR = Q_3 - Q_1$$

- This is the **middle 50%** of the data set and is not bothered by extreme observations in the tails of the data set.

Variance

- **Variance** is a measure of dispersion around the mean of the data set.
- It is the average of the squared distances between each data value and mean.

Population Variance

- **Variance** is a measure of dispersion around the mean of the data set.
- It is the average of the squared distances between each data value and mean.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

Sample Variance

- **Variance** is a measure of dispersion around the mean of the data set.
- It is the average of the squared distances between each data value and mean.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Standard Deviation

- The problem with variance is that it is in terms of squared units of the data.
- To correct for this, we have the **standard deviation**, which is just the square root of the variance.

Standard Deviation

- The problem with variance is that it is in terms of squared units of the data.
- To correct for this, we have the **standard deviation**, which is just the square root of the variance.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{OR} \quad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Sample

Population

2 Characteristics of Variance

- Variance (and standard deviation) possess two common characteristics:
  1. If the variance equals zero, then all of the data in the data set has the same value.
  2. All measures of spread are positive (or nonnegative if zero spread) in value.

DESCRIBING DISTRIBUTIONS

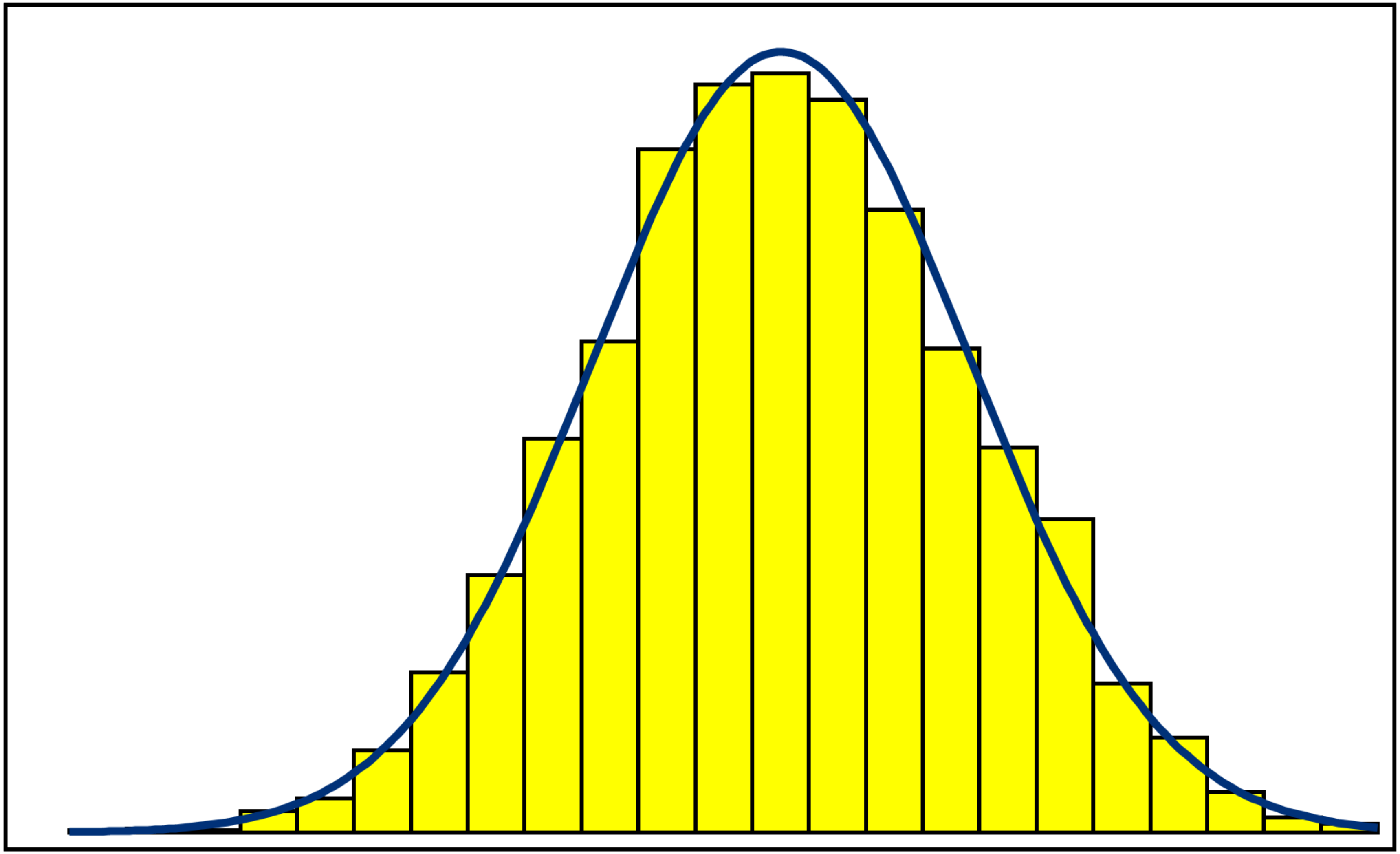
Shape

Skewness

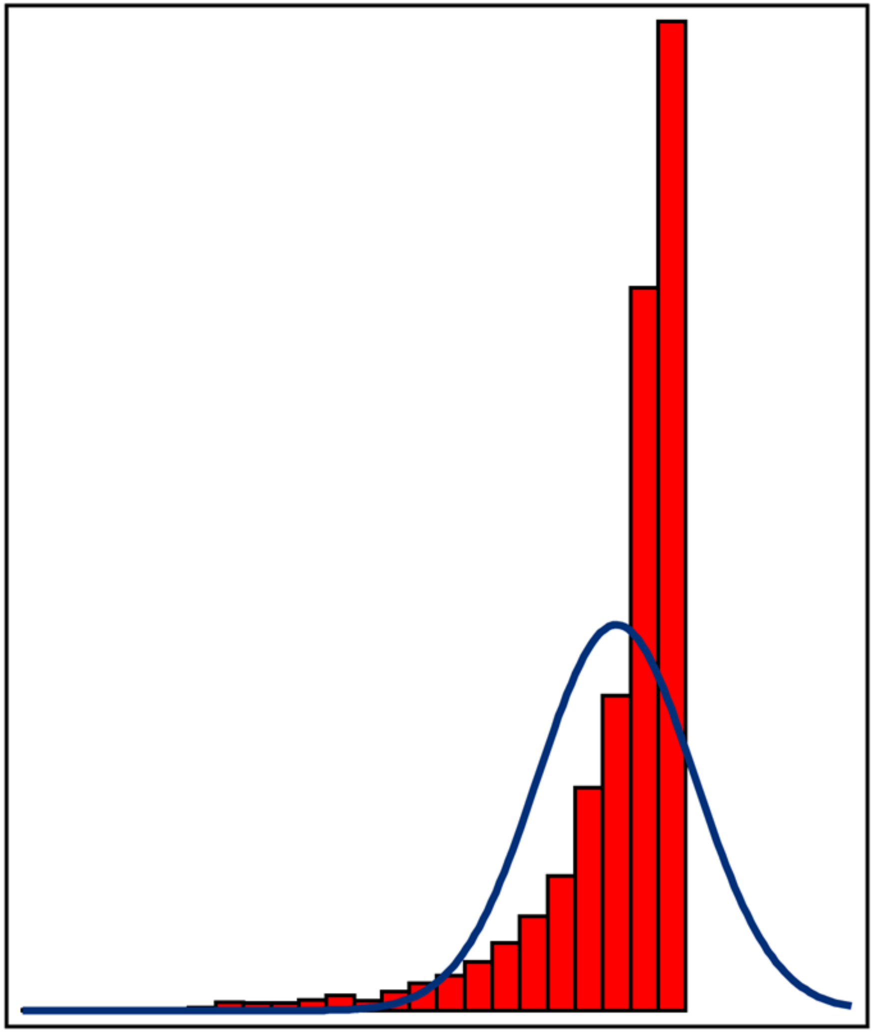
- **Skewness** of a distribution deals with the symmetry (or lack there of) that a distribution has.

$$g_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$$

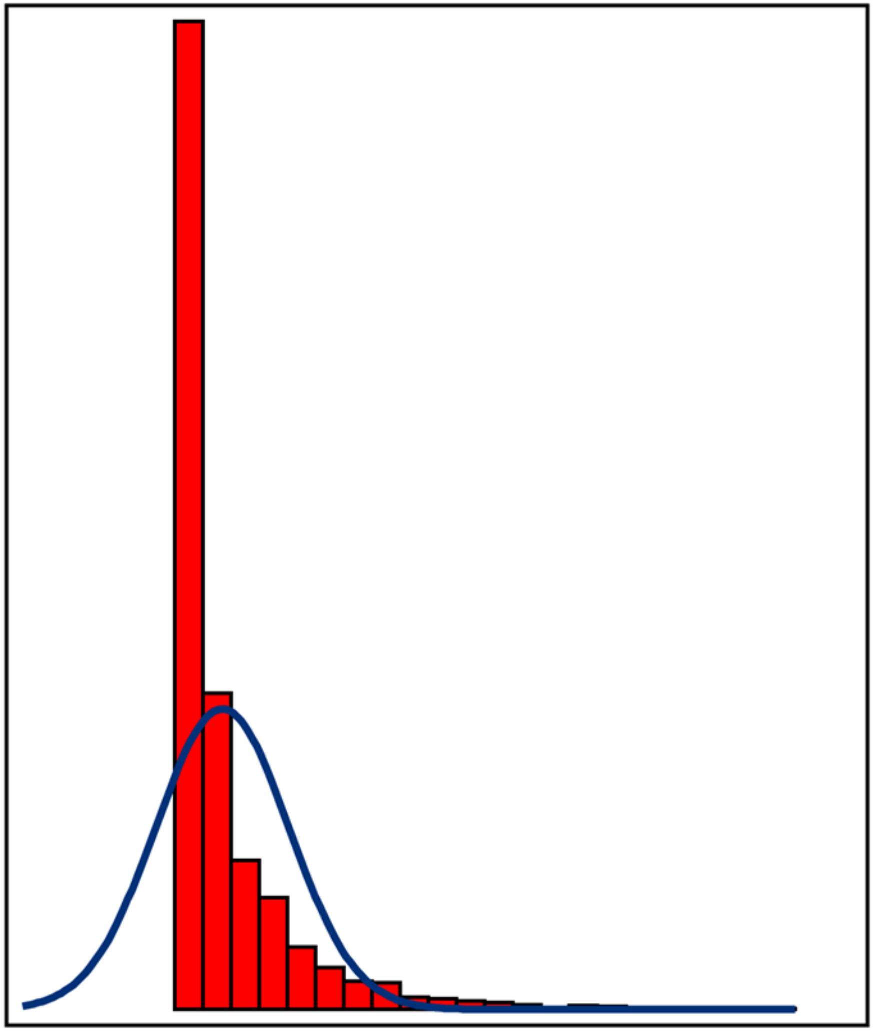
Symmetric



Left-skewed



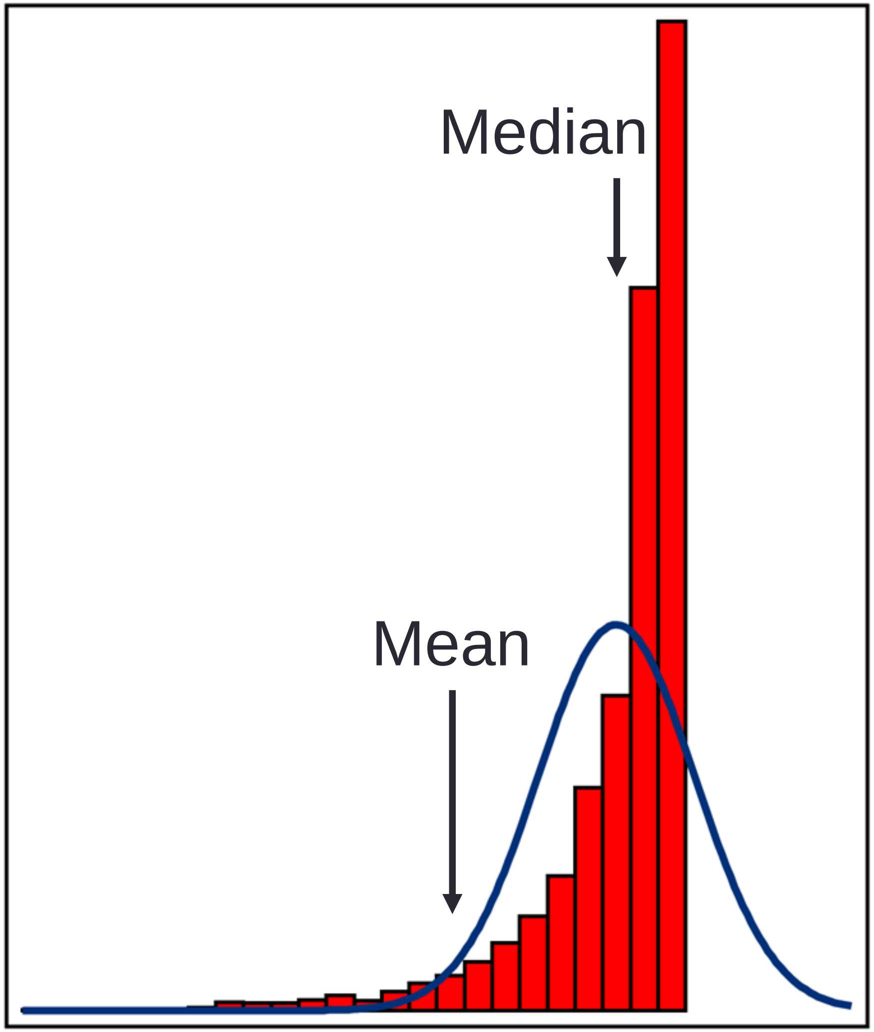
Right-skewed



Skewness

- **Skewness** of a distribution deals with the symmetry (or lack there of) that a distribution has.
- In symmetric distributions the mean and median are approximately the same (equal if perfectly symmetric).
- In left-skewed distributions the mean is smaller than the median.
- In right-skewed distributions the mean is larger than the median.

Left-skewed



Kurtosis

- **Kurtosis** of a distribution deals with the thickness of tails that a distribution has.

$$g_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4$$

Kurtosis

- **Kurtosis** of a distribution deals with the thickness of tails that a distribution has.

$$g_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4$$

- **Mesokurtic**: $g_2 = 3$ → baseline thickness (Normal distribution)
- **Platykurtic**: $g_2 < 3$ → thin-tailed (light tailed)
- **Leptokurtic**: $g_2 > 3$ → thick-tailed (heavy tailed)

Kurtosis

- **Kurtosis** of a distribution deals with the thickness of tails that a distribution has.

$$g_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4$$

- **Excess kurtosis** has a baseline measure of zero instead of three.

$$g_2^* = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4\right) - 3$$

DESCRIBING DISTRIBUTIONS

Detecting Outliers

Outliers

- An **outlier** is a data value that is unusually small or large in a data set.
- To understand how small or large is “unusual” we must first understand the relative placement of values in the data set.

Outlier Effects

- Means, variances, skewness, and kurtosis are influenced by outliers (extreme observations).
- Medians are not influenced by outliers.

13	22	24	26	27	29	30

Mean = 24.43

Variance = 32.95

Median = 26

Outlier Effects

- Means, variances, skewness, and kurtosis are influenced by outliers (extreme observations).
- Medians are not influenced by outliers.

13	22	24	26	27	29	30

Mean = 24.43
Variance = 32.95

Median = 26

13	22	24	26	27	29	300

Mean = 63
Variance = 10948.67

Median = 26

What to do with Outliers?

- Outliers might be one of the following:
  - Incorrectly recorded data value.
  - Valid data value that should not have been included in data set.
  - Valid data value that belongs in the data set.

What to do with Outliers?

- Outliers might be one of the following:
  - Incorrectly recorded data value.
  - Valid data value that should not have been included in data set.
  - **Valid data value that belongs in the data set.**

- Must investigate outliers!

Example

- A marketing firm collected data on annual household incomes for Outland, NC. They surveyed all 182 households in the small town of Outland. Now imagine that one of the marketing firm's vice presidents really liked the city after collecting data from them and moves to the city. Now the data set has 183 households. The vice president's household's annual income is $384,000. Re-calculate the mean, median, standard deviation, skewness, and kurtosis.

DISCRETE PROBABILITY DISTRIBUTIONS

Probability Distribution

- The **probability distribution** for a random variable describes how probabilities are distributed over the values of the random variable.

Notation

- **Frequency** – number of observations in each category in the data set
- **Relative Frequency** – proportion of total observations contained in a given category
- **Cumulative Frequency** – summary of data set *i* number of observations with values less than or equal to upper limit of the category
- **Cumulative Relative Frequency** – proportion of observations with value less than or equal to upper limit of the category

Probability Distribution

- The **probability distribution** for a random variable describes how probabilities are distributed over the values of the random variable.
- Relative frequencies can be used as estimates to the probability of an event occurring.
- Probability distributions for discrete random variables are best described with tables, graphs, or equations.

Discrete Probability Example

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90		
1	85		
2	70		
3	45		
4	50		
5	25		
	365		

Discrete Probability Example

- Let x be the number of TV’s sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let’s examine the past year of data.

TV’s Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90	90	90/365
1	85		
2	70		
3	45		
4	50		
5	25		
	365		

Discrete Probability Example

- Let x be the number of TV’s sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let’s examine the past year of data.

TV’s Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90	90	0.25
1	85	90+85	0.23
2	70		
3	45		
4	50		
5	25		
	365		

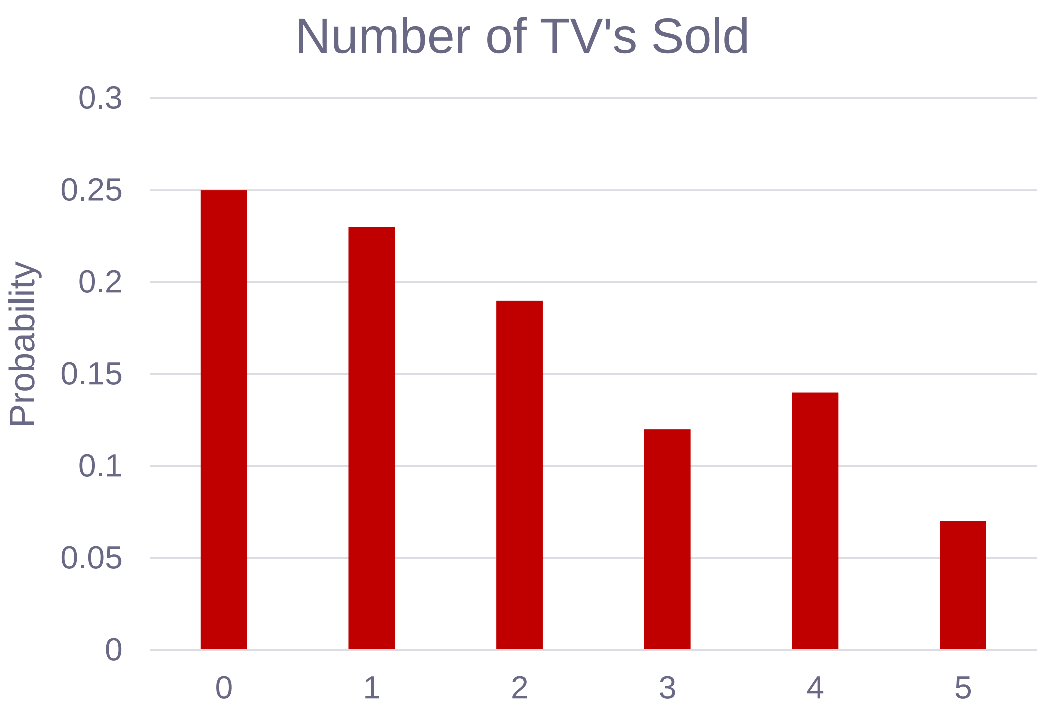
Discrete Probability Example

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.

TV's Sold	Number of Days (Freq)	Cumulative Frequency	Relative Frequency
0	90	90	0.25
1	85	175	0.23
2	70	245	0.19
3	45	290	0.12
4	50	340	0.14
5	25	365	0.07
	365		**1.00**

Discrete Probability Example

- Let x be the number of TV's sold at a small department store in one day where x can only take the values of {0, 1, 2, 3, 4, 5}
- Let's examine the past year of data.



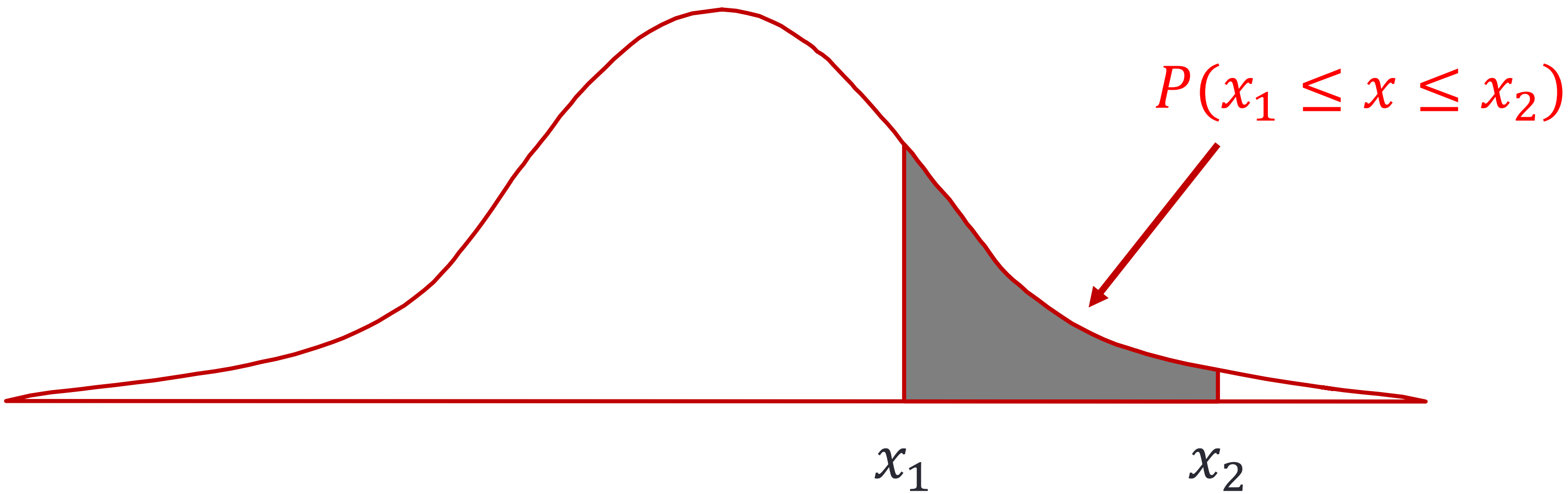
CONTINUOUS PROBABILITY DISTRIBUTIONS

Continuous Random Variables

- A **continuous random variable** can assume any value in an interval on the real line or in a collection of intervals on the real line.
- It is **not** possible to talk about the probability of the random variable assuming a particular value.
- Instead, we talk about the probability of the random variable assuming a value inside of a given interval.

Continuous Random Variables

- Instead, we talk about the probability of the random variable assuming a value inside of a given interval.
- The probability of the random variable assuming a value inside of a given interval from x_1 to x_2 is the **area under the graph** of the **probability density function** between x_1 and x_2.



CONTINUOUS PROBABILITY DISTRIBUTIONS

Normal Distribution

Importance

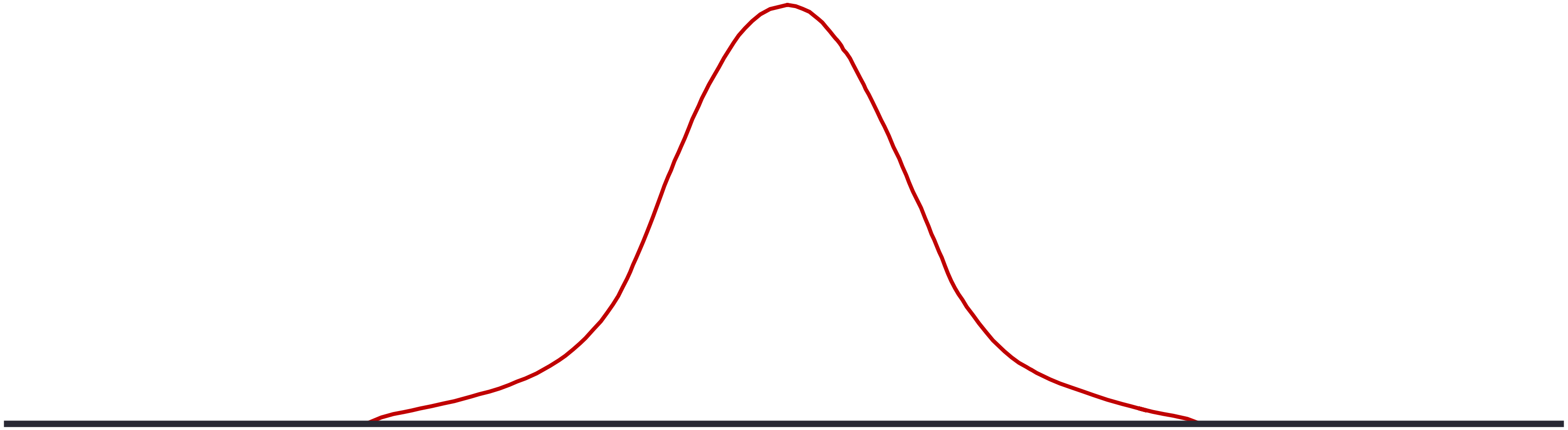
- The **Normal probability distribution** is one of the most common and important distributions for describing a continuous random variable.
- The Normal distribution is the foundation of statistical inference:
  - Hypothesis Testing
  - Confidence Intervals
  - Regression Analysis
- Appears in nature and real world data.

Characteristics of Normal Distribution

- The Normal distribution has some useful characteristics:
  - Perfectly Symmetric (Skewness = 0)
  - Unimodal
  - Mean = Median = Mode
  - Asymptotic to x-axis (Can take any value from $-\infty$ to ∞)
  - Completely Defined by Mean and Standard Deviation
  - Excess kurtosis = 0 (kurtosis = 3)

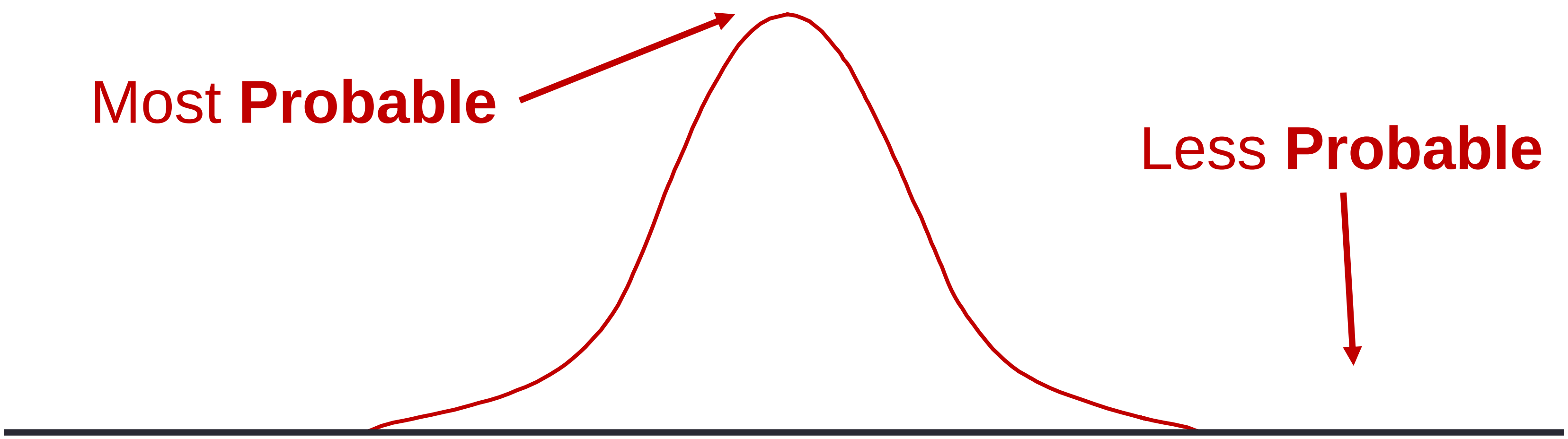
Characteristics of Normal Distribution

- The Normal distribution has some useful characteristics:
  - Perfectly Symmetric (Skewness = 0)
  - Unimodal
  - Mean = Median = Mode
  - Asymptotic to x-axis (Can take any value from $-\infty$ to ∞)
  - Completely Defined by Mean and Standard Deviation



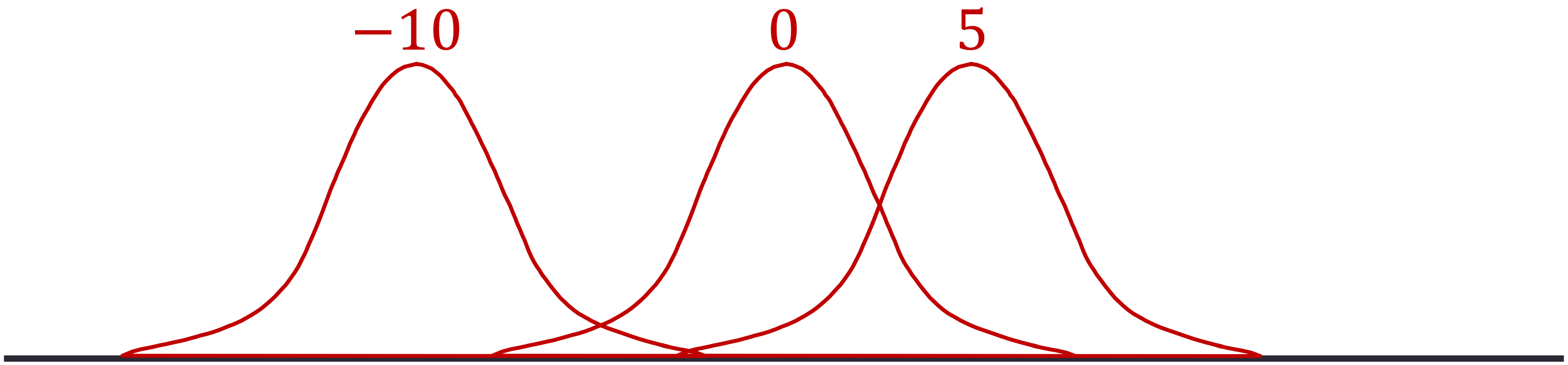
Characteristics of Normal Distribution

- The Normal distribution has some useful characteristics:
  - Perfectly Symmetric (Skewness = 0)
  - Unimodal
  - Mean = Median = Mode
  - Asymptotic to x-axis (Can take any value from $-\infty$ to ∞)
  - Completely Defined by Mean and Standard Deviation



Characteristics of Normal Distribution

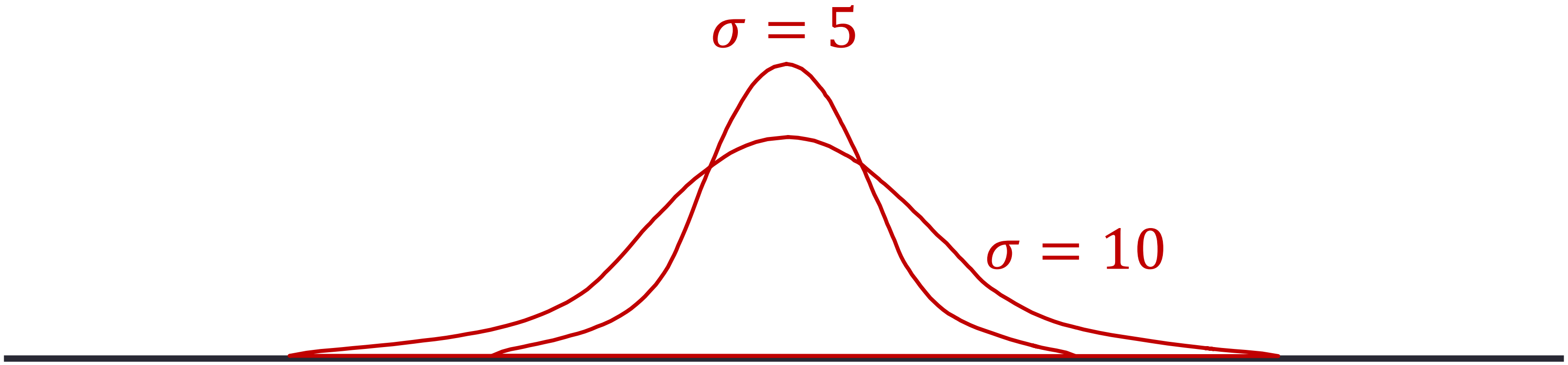
- The Normal distribution has some useful characteristics:
  - Perfectly Symmetric (Skewness = 0)
  - Unimodal
  - Mean = Median = Mode
  - Asymptotic to x-axis (Can take any value from $-\infty$ to ∞)
  - Completely Defined by Mean and Standard Deviation



Mean can take ANY value

Characteristics of Normal Distribution

- The Normal distribution has some useful characteristics:
  - Perfectly Symmetric (Skewness = 0)
  - Unimodal
  - Mean = Median = Mode
  - Asymptotic to x-axis (Can take any value from $-\infty$ to ∞)
  - Completely Defined by Mean and Standard Deviation



Standard Deviation controls the width

Probabilities

- The probabilities for the Normal random variable are determined by the area under the curve.
- The **total area under the curve = 1**.
- Since the Normal distribution is perfectly symmetric around the mean (and median), then the area of the curve below the mean = above the mean = 0.5.

Probability Density Function

- The probability density function for the Normal distribution is defined as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Probability Density Function

- The probability density function for the Normal distribution is defined as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Mean = μ

Probability Density Function

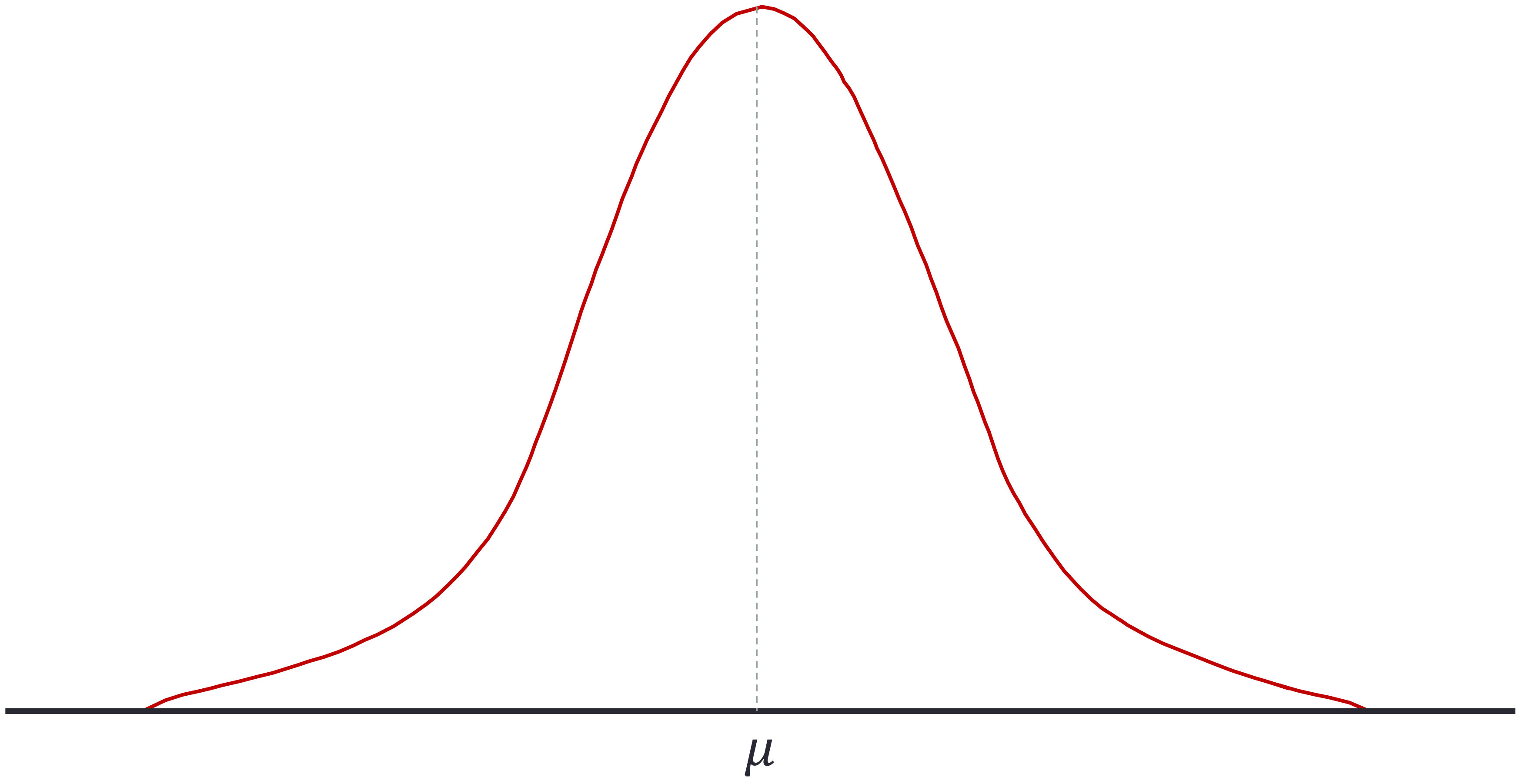
- The probability density function for the Normal distribution is defined as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

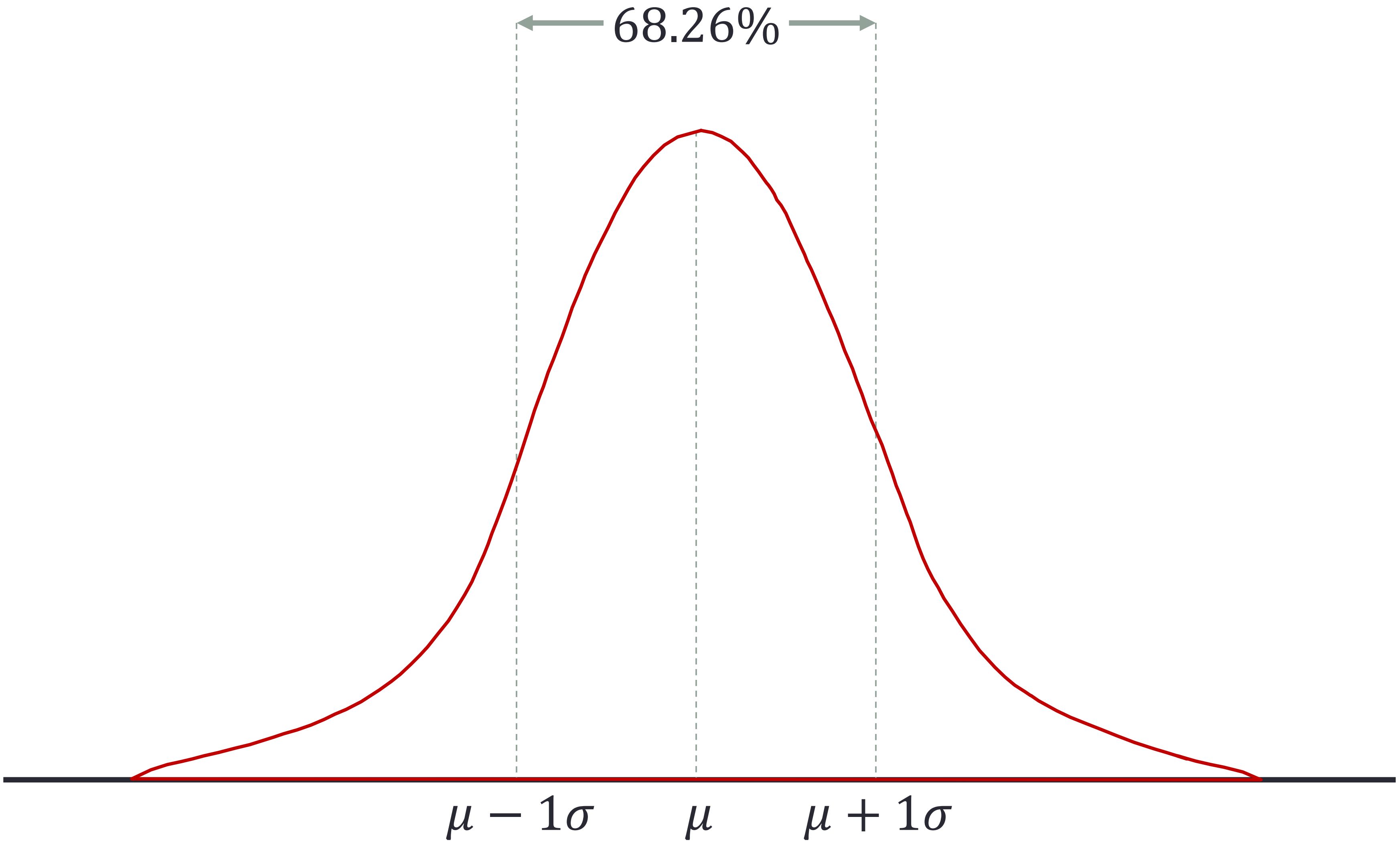
Mean = μ

Standard Deviation = σ

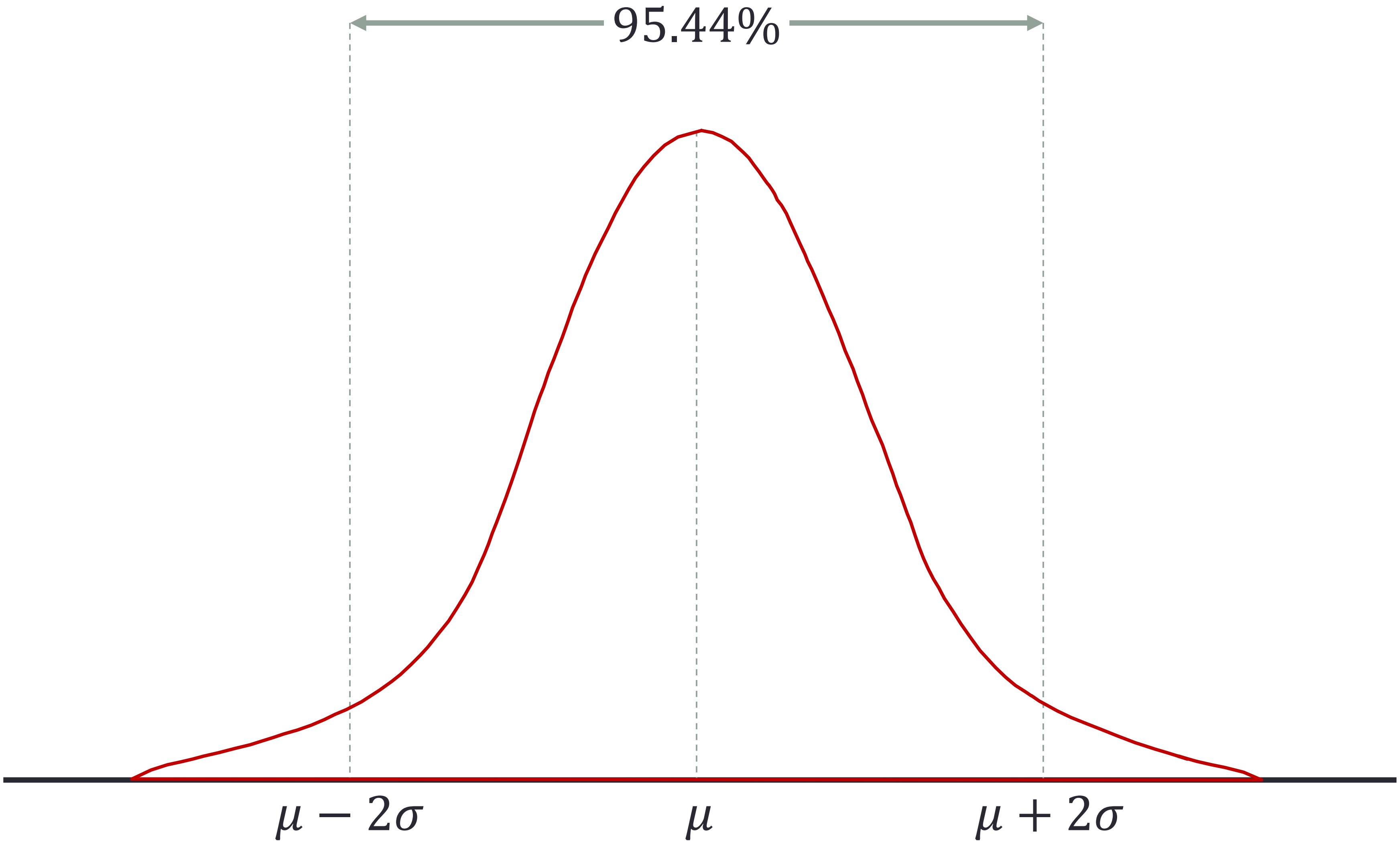
Empirical Rule



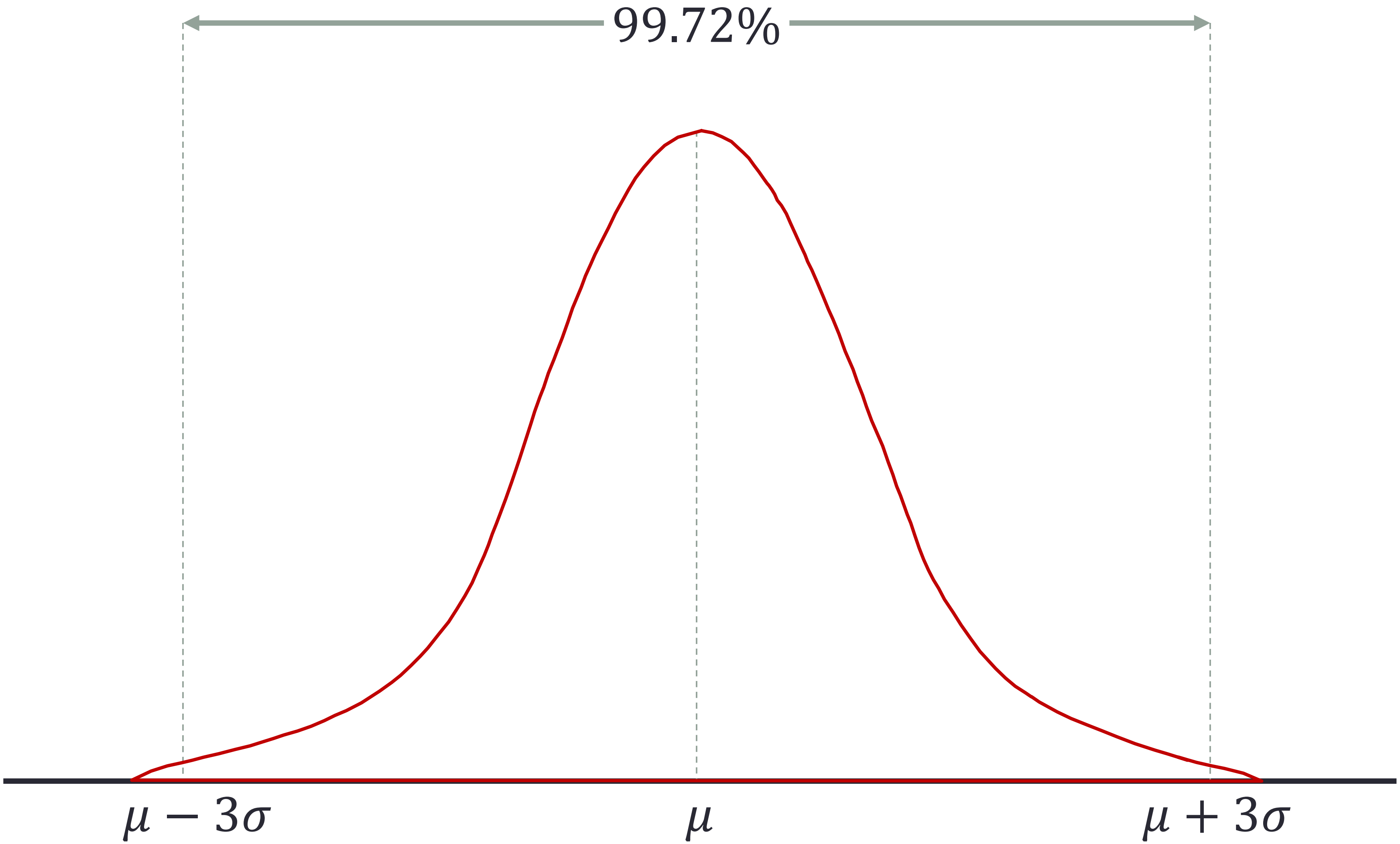
Empirical Rule



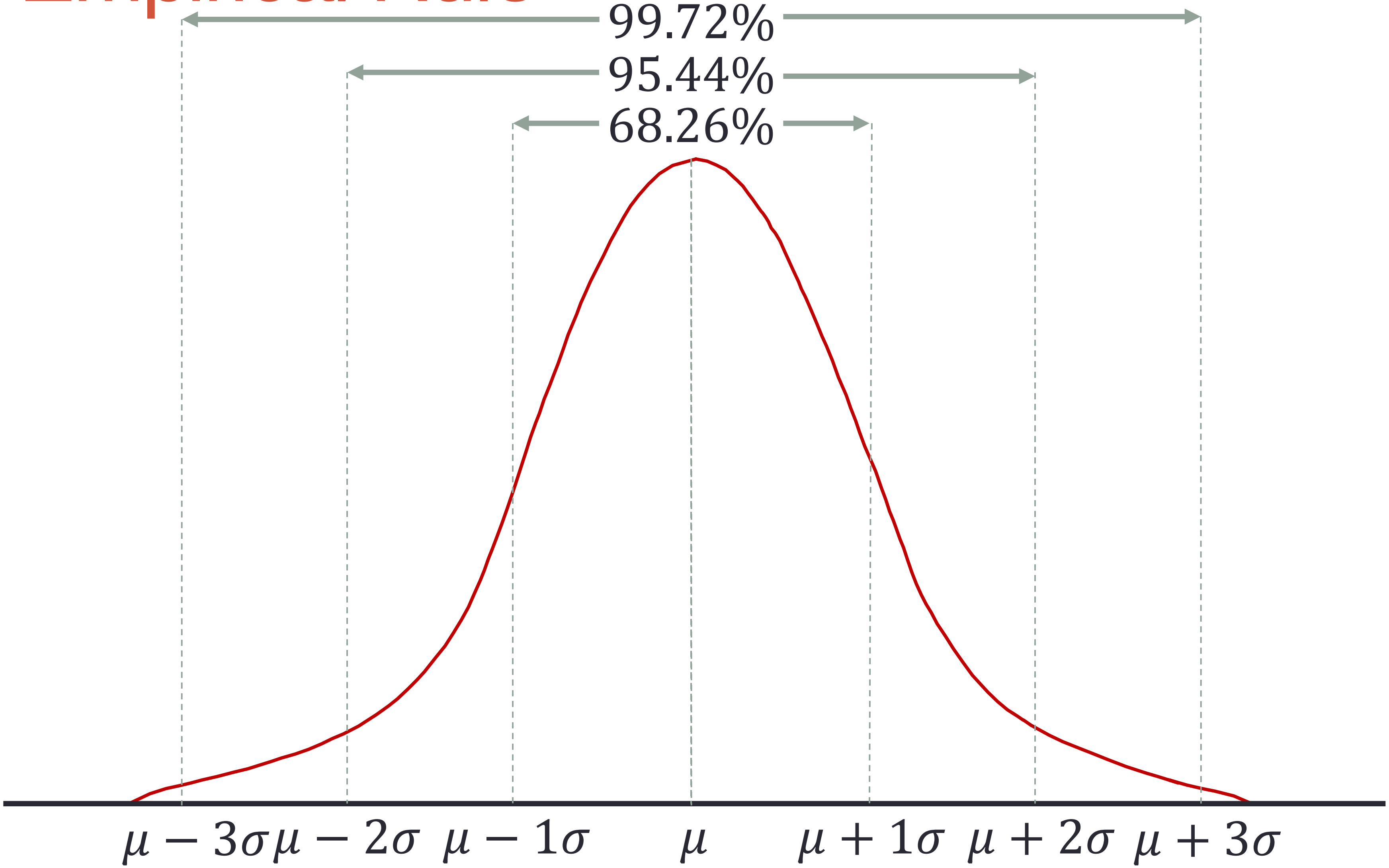
Empirical Rule



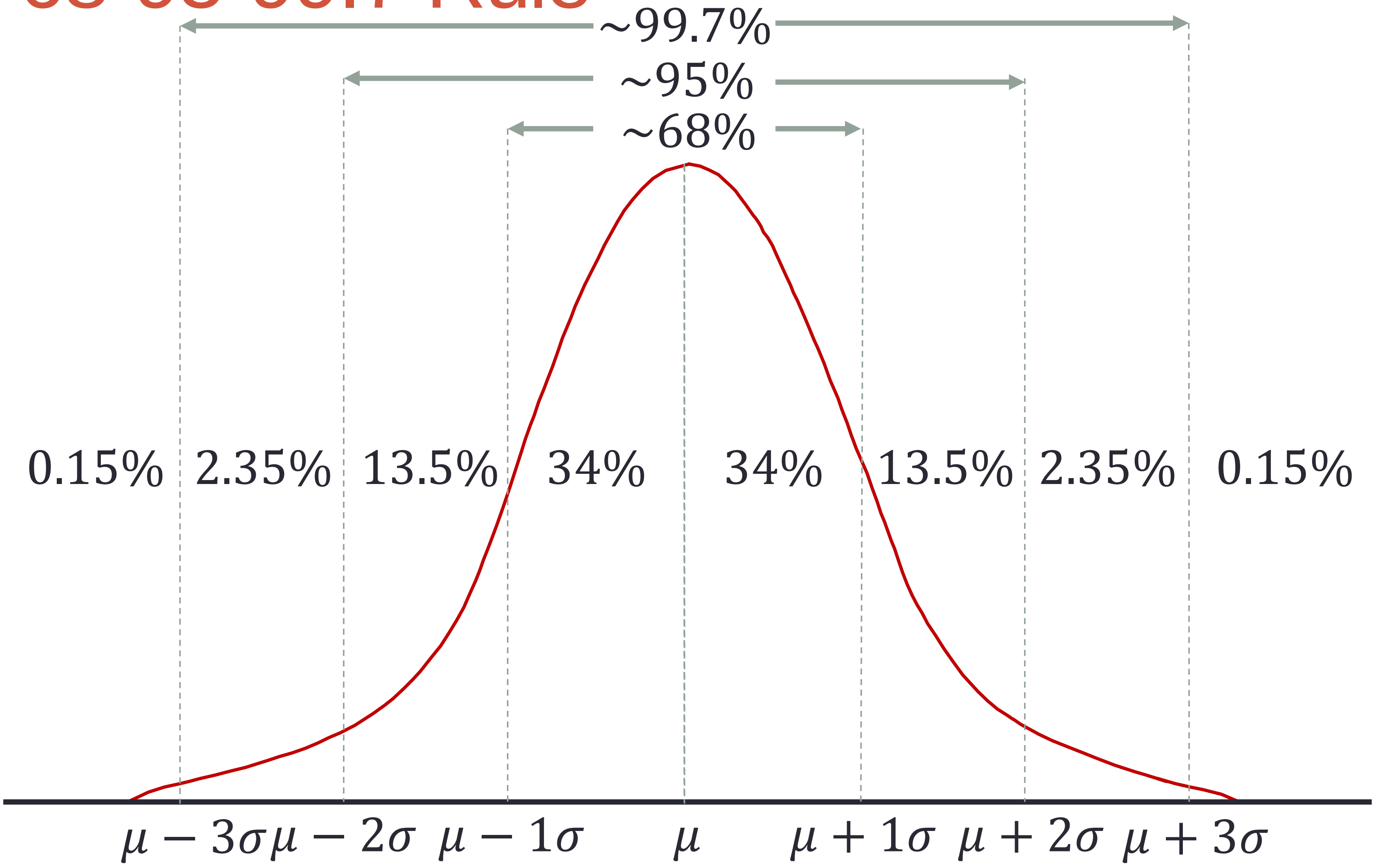
Empirical Rule



Empirical Rule



68-95-99.7 Rule

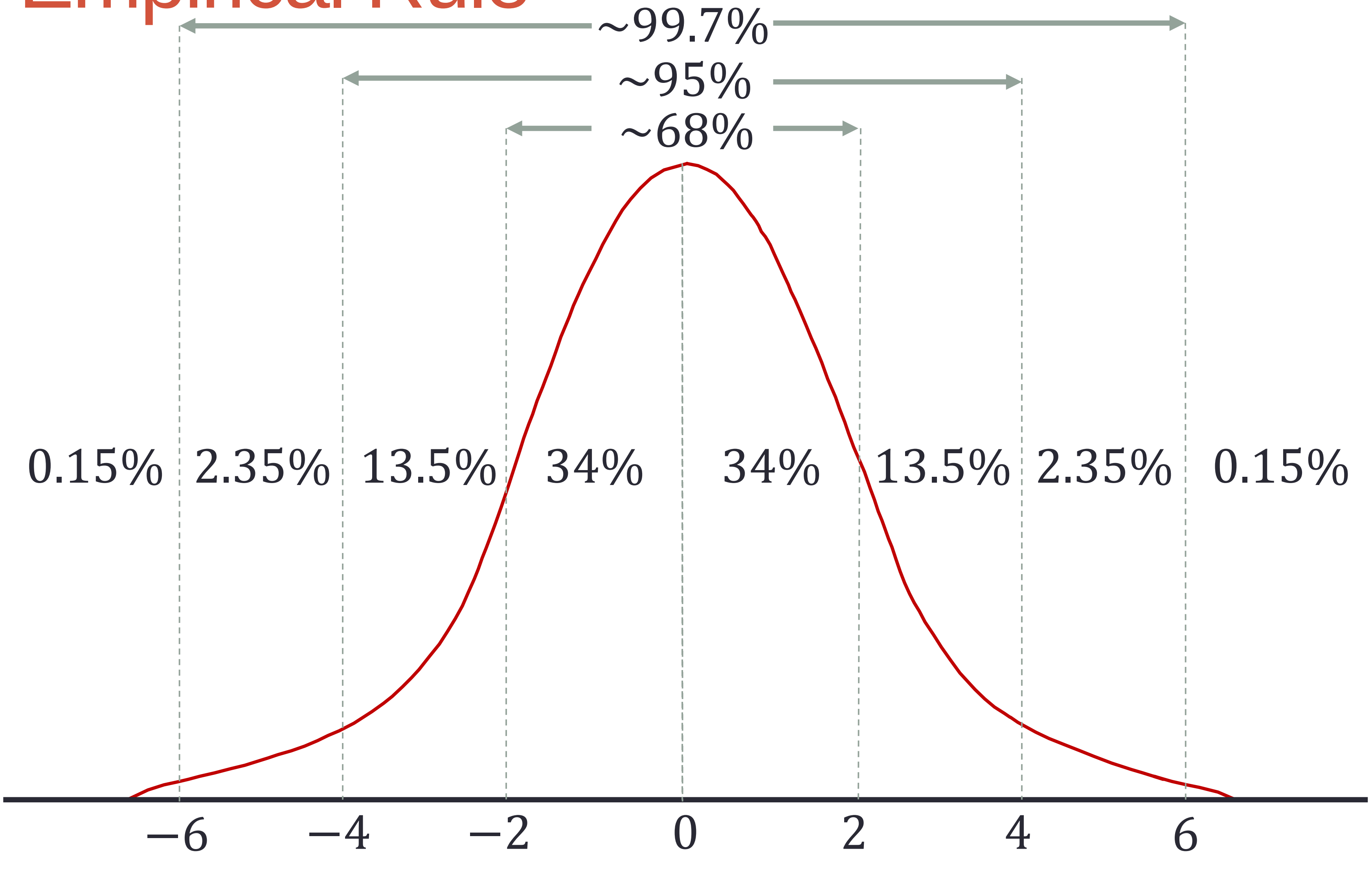


Example

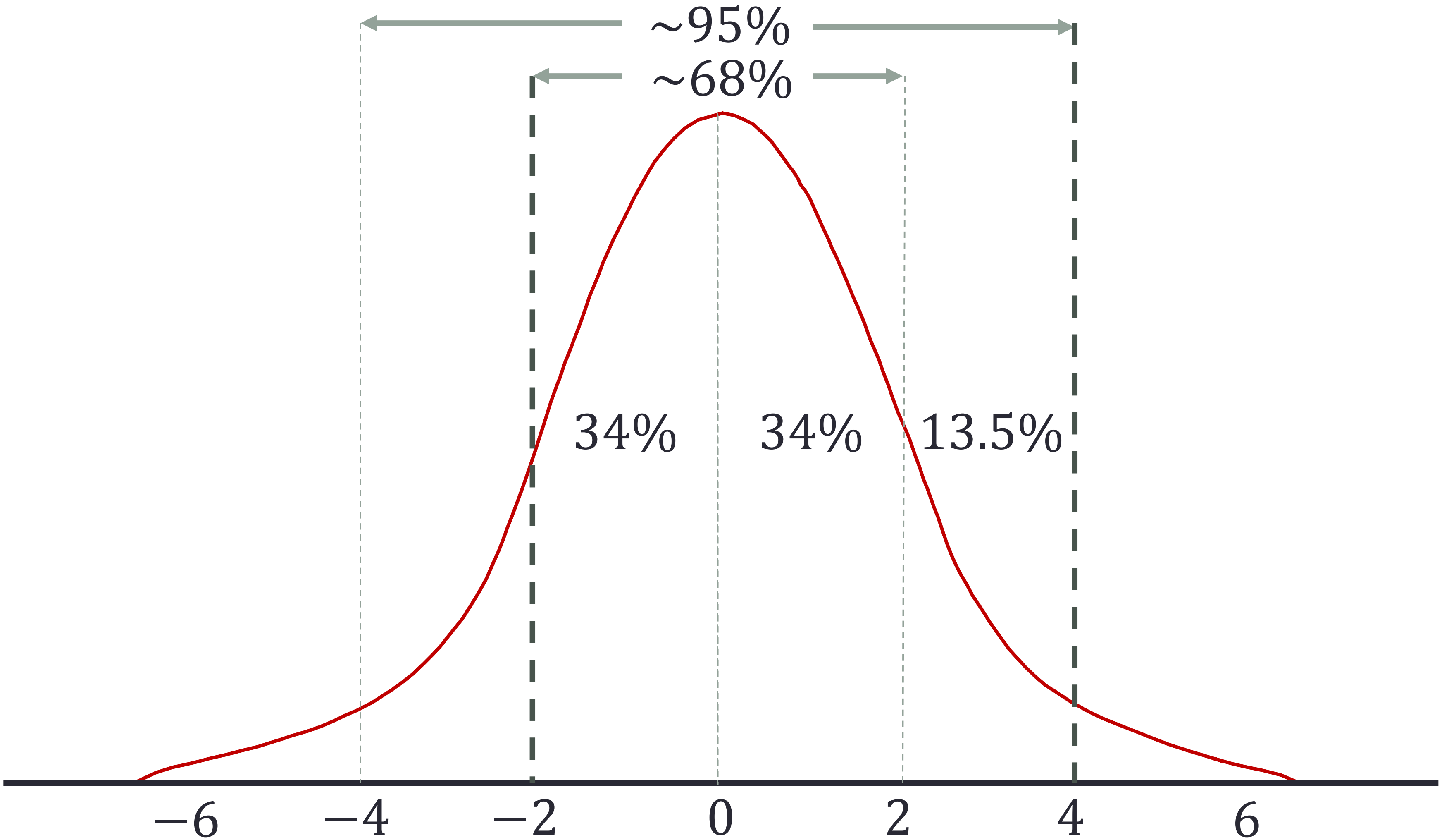
- What is the probability a random variable that follows a normal distribution with mean = 0 and s.d. = 2 takes a value between -2 and 4?

- Assume the weekly number of credit card transactions for customers of a major credit card follows a normal distribution with mean = 8.5 and s.d. = 1.5. What is probability randomly selected customer has between 7 and 13 transactions in a week?

Empirical Rule



Empirical Rule



Example

- What is the probability a random variable that follows a normal distribution with mean = 0 and s.d. = 2 takes a value between -2 and 4?

0.815 or 81.5%

- Assume the weekly number of credit card transactions for customers of a major credit card follows a normal distribution with mean = 8.5 and s.d. = 1.5. What is probability randomly selected customer has between 7 and 13 transactions in a week?

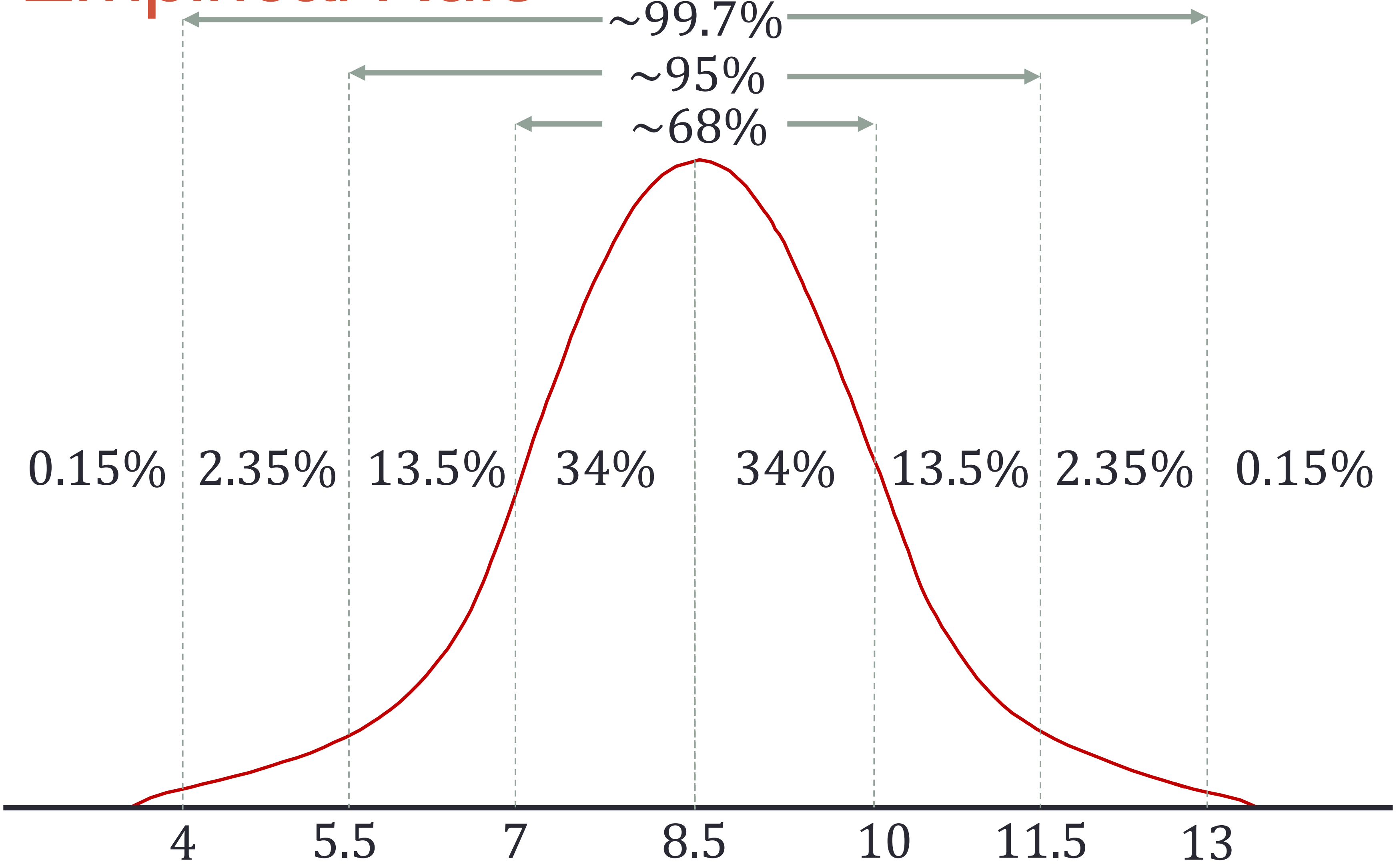
Example

- What is the probability a random variable that follows a normal distribution with mean = 0 and s.d. = 2 takes a value between -2 and 4?

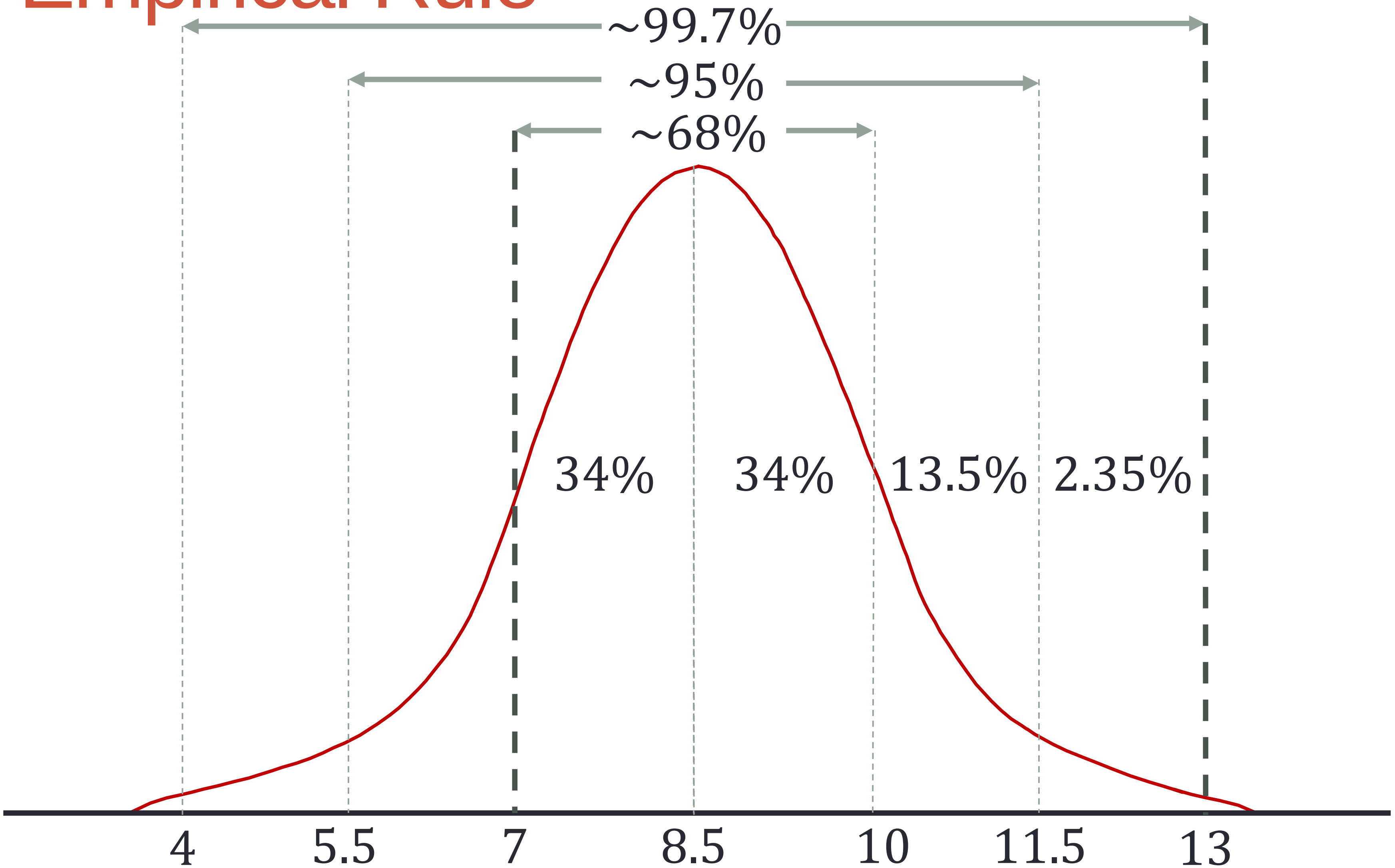
0.815 or 81.5%

- Assume the weekly number of credit card transactions for customers of a major credit card follows a normal distribution with mean = 8.5 and s.d. = 1.5. What is probability randomly selected customer has between 7 and 13 transactions in a week?

Empirical Rule



Empirical Rule



Example

- What is the probability a random variable that follows a normal distribution with mean = 0 and s.d. = 2 takes a value between -2 and 4?

0.815 or 81.5%

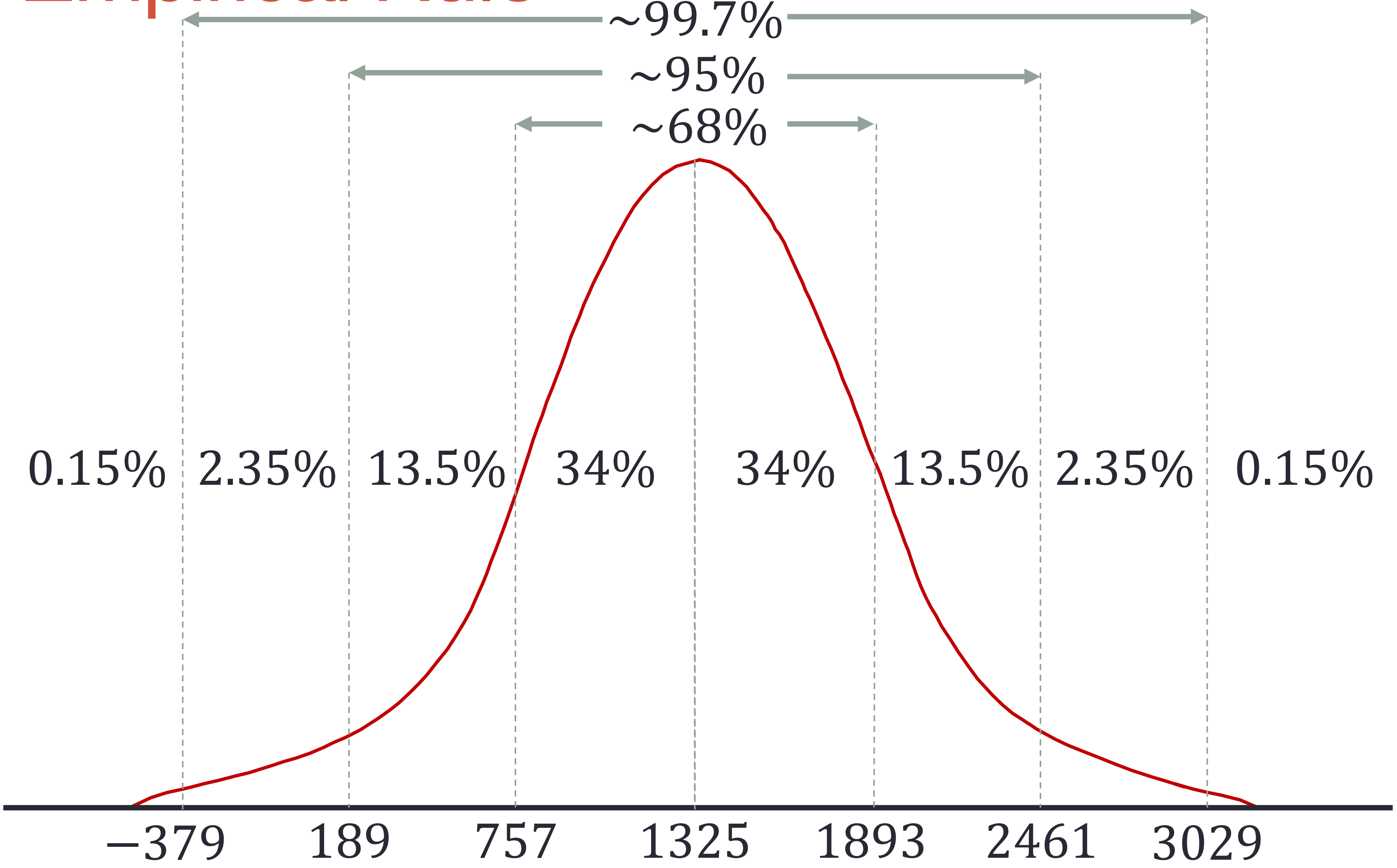
- Assume the weekly number of credit card transactions for customers of a major credit card follows a normal distribution with mean = 8.5 and s.d. = 1.5. What is probability randomly selected customer has between 7 and 13 transactions in a week?

0.8385 or 83.85%

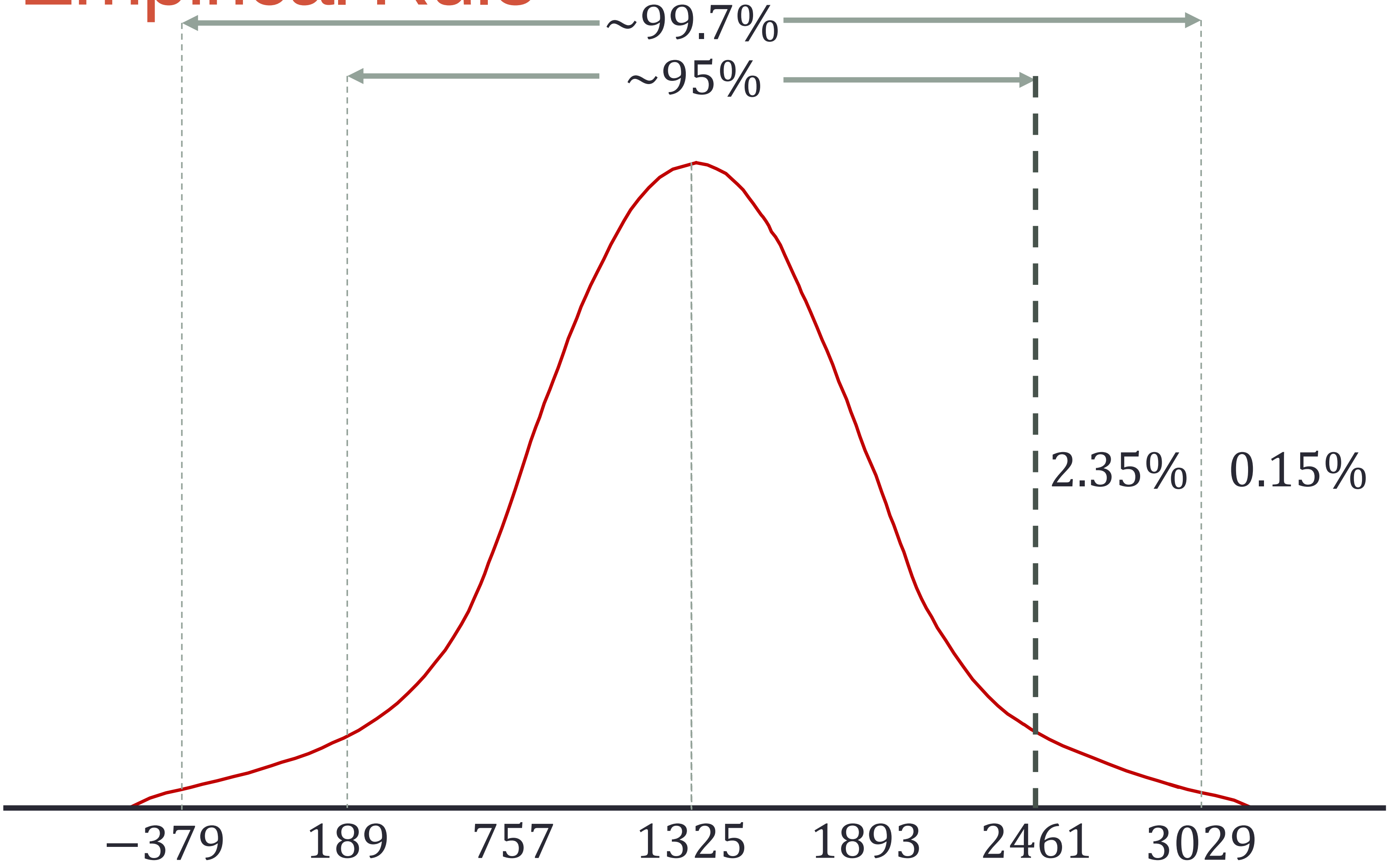
Example

- Assume the number of customers of a major television provider churn every month follows a normal distribution with mean = 1325 and s.d. = 568. What is probability randomly selected month has more than 2461 customers that churn?

Empirical Rule



Empirical Rule



Example

- Assume the number of customers of a major television provider churn every month follows a normal distribution with mean = 1325 and s.d. = 568. What is probability randomly selected month has more than 2461 customers that churn?

0.025 or 2.5%

Empirical Rule

- Good for quick, fast, rough analysis.
- Not good for exact analysis unless your interests are only in the integer standard deviations.
- What about fractions of standard deviations away from the mean?
- Need another way to quickly calculate area under the curve.

CONTINUOUS PROBABILITY DISTRIBUTIONS

Standard Normal Distribution

Conversion of Normal Distributions

- A random variable having a Normal distribution with a mean of 0 and a standard deviation of 1 is said to have a **standard Normal probability distribution**.
- All Normal distributions can be converted into standard Normal distributions for ease of computing probabilities under the curve.
- Standard Normal probability tables help calculate area under the curve.

Standard Normal Table

- The standard Normal table is an extension of the empirical rule where the area under the standard Normal curve to the left of any point is calculated up to two decimal points.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7517	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
.	.	.	.	.	.	.	.	.	.	.

Standard Normal Table

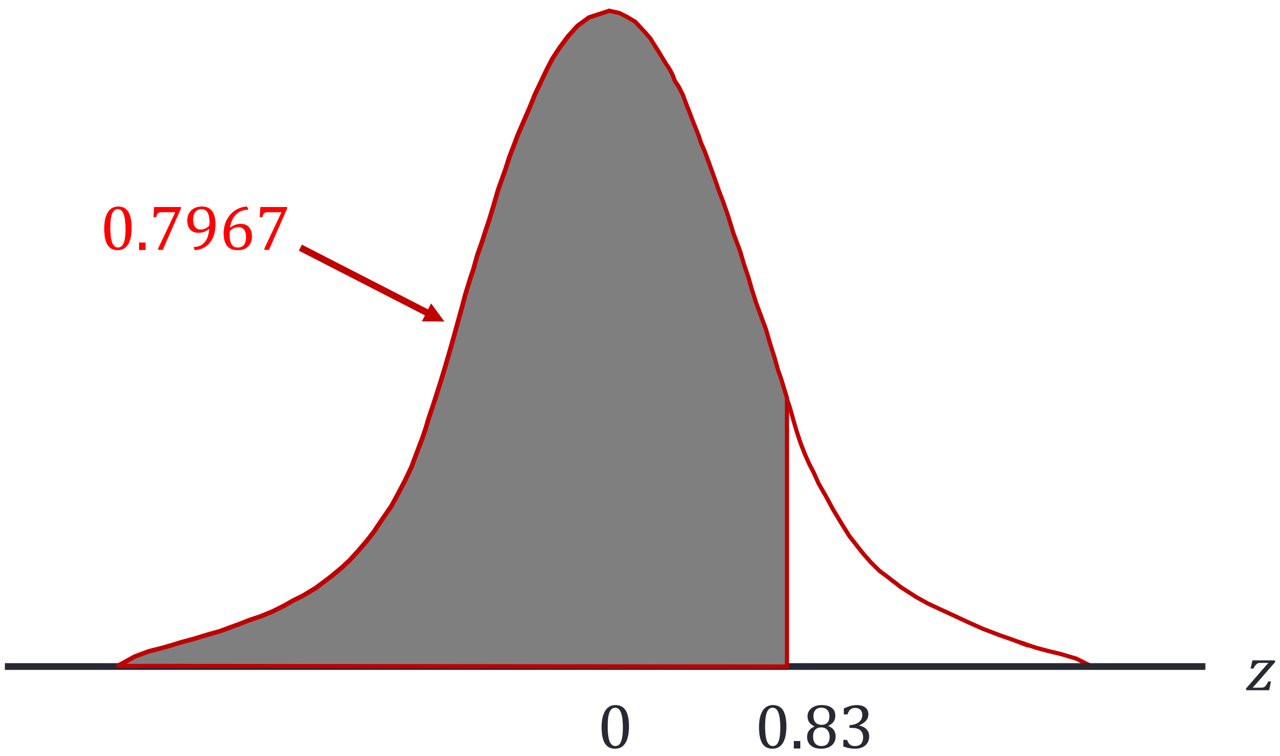
- The standard Normal table is an extension of the empirical rule where the area under the standard Normal curve to the left of any point is calculated up to two decimal points.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7517	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
.	.	.	.	.	.	.	.	.	.	.

$P(z \leq 0.83)$

Standard Normal Table

- The standard Normal table is an extension of the empirical rule where the area under the standard Normal curve to the **left** of any point is calculated up to two decimal points.



Using Software for Probabilities

- Most software (R, SAS, Python, Excel, etc) can calculate standard normal probabilities
- More accurate (and easier) to use software to calculate probabilities
- For example, in R to find $P(z \leq 0.83)$:

pnorm(0.83) = 0.7967

Standard Normal Table

- The standard Normal table is an extension of the empirical rule where the area under the standard Normal curve to the **left** of any point is calculated up to two decimal points.
- To calculate values to the **right** of any point, use the laws of probability:

$$
\begin{aligned}
P(z > 0.83) &= 1 - P(z \leq 0.83) \\
&= 1 - 0.7967 \\
&= 0.2033
\end{aligned}
$$

Example

- Assuming a normal distribution with mean = 0 and s.d. = 1, find the following probabilities:

1. $P(z > 2)$
2. $P(z \leq 1.12)$
3. $P(-1.33 \leq z \leq 1.33)$
4. $P(z \geq 3.02)$
5. $P(z \leq 6.87)$
6. $P(z \leq -6.87)$

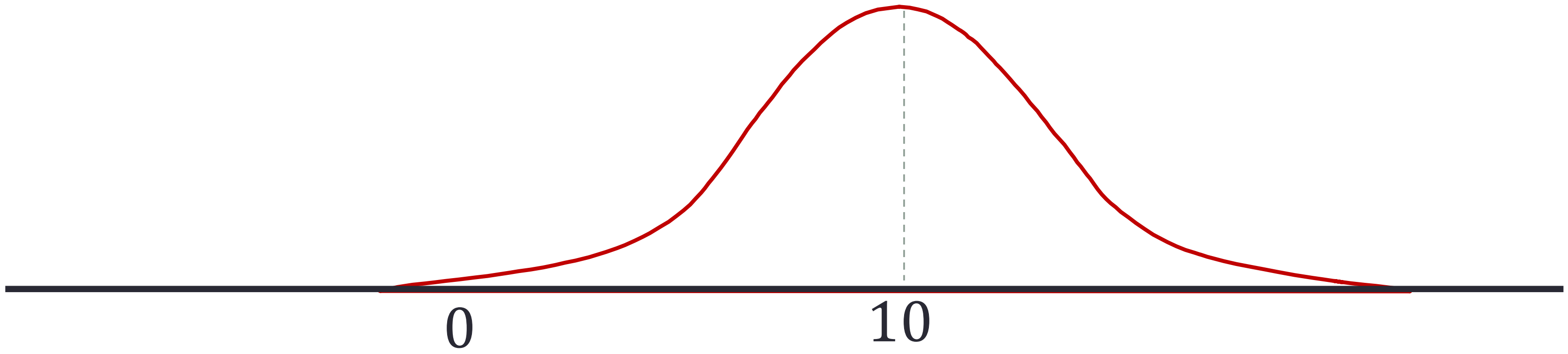
Examples

- Assuming a normal distribution with mean = 0 and s.d. = 1, find the following probabilities:

1. $P(z > 2) = 0.0228$
2. $P(z \leq 1.12) = 0.8686$
3. $P(-1.33 \leq z \leq 1.33) = 0.9082 - 0.0918 = 0.8164$
4. $P(z \geq 3.02) = 0.0013$
5. $P(z \leq 6.87) \approx 1$
6. $P(z \leq -6.87) \approx 0$

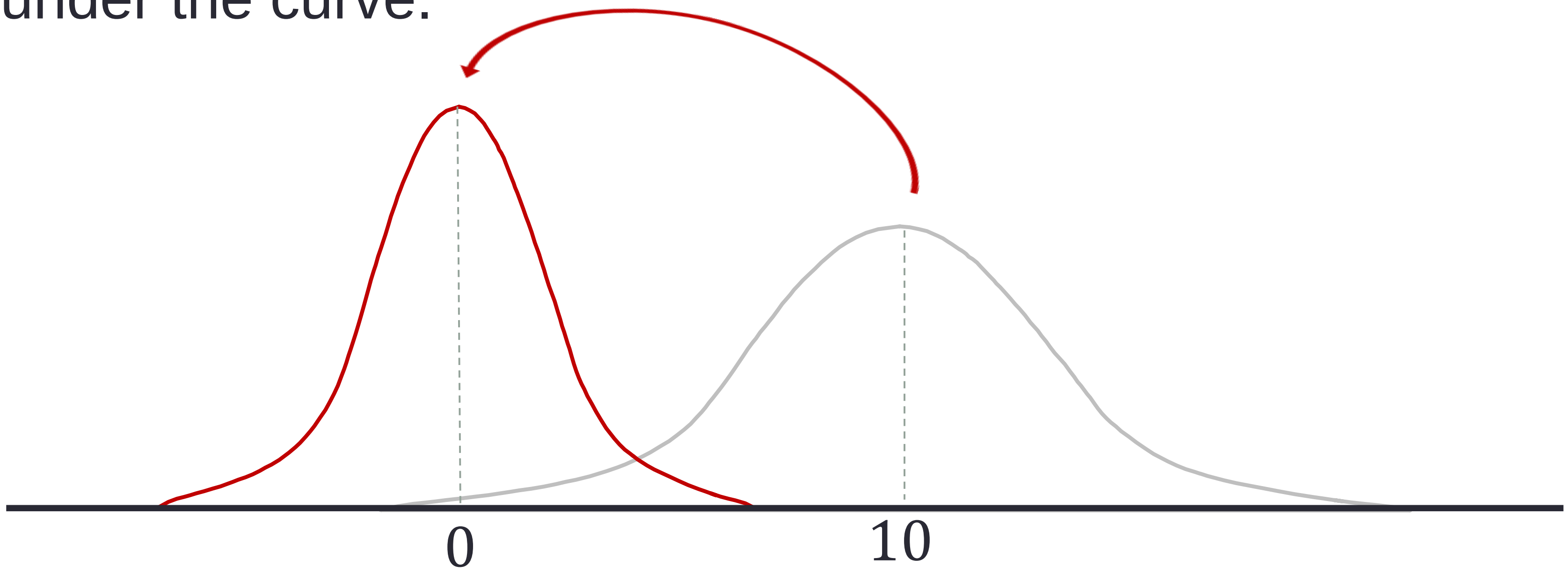
Conversion of Normal Distributions

- All Normal distributions can be converted into standard Normal distributions for ease of computing probabilities under the curve.



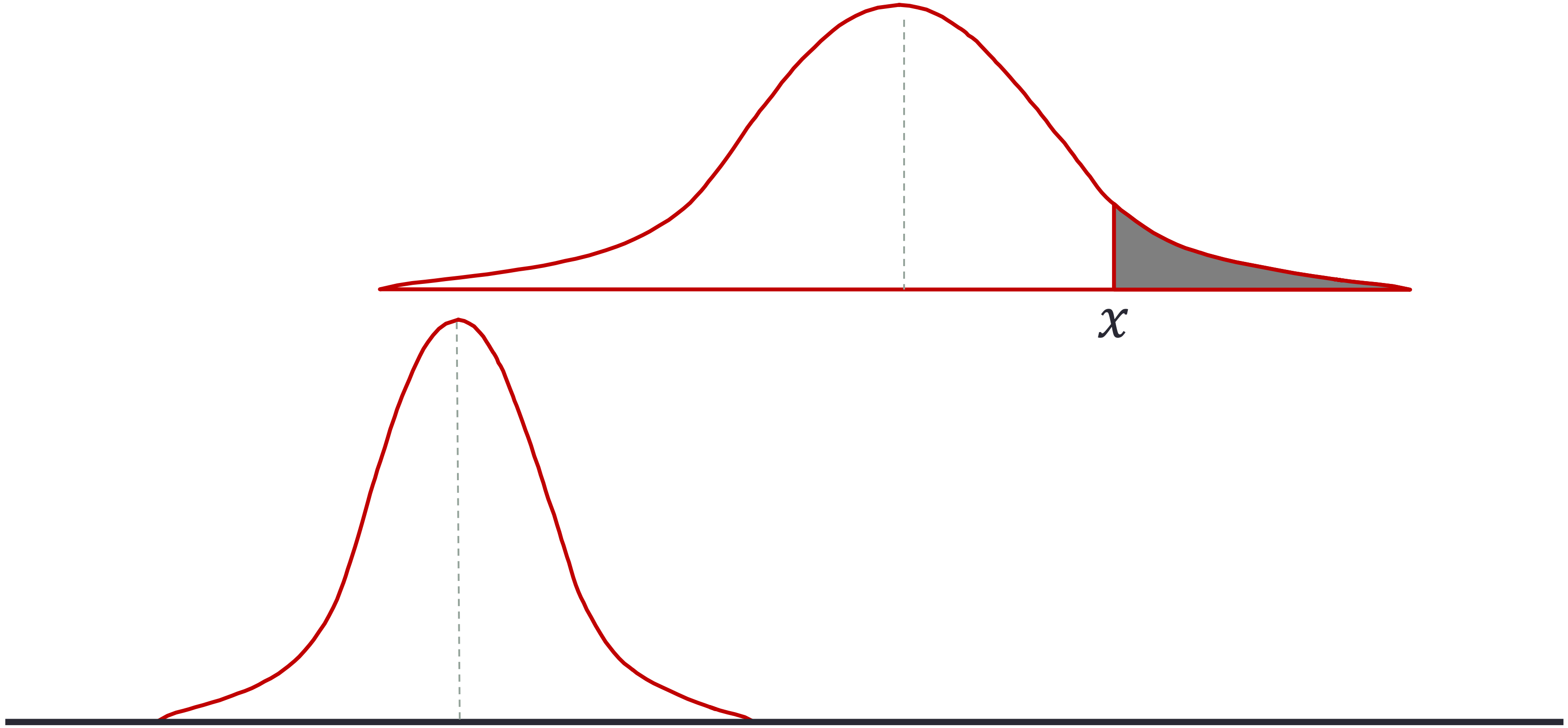
Conversion of Normal Distributions

- All Normal distributions can be converted into standard Normal distributions for ease of computing probabilities under the curve.



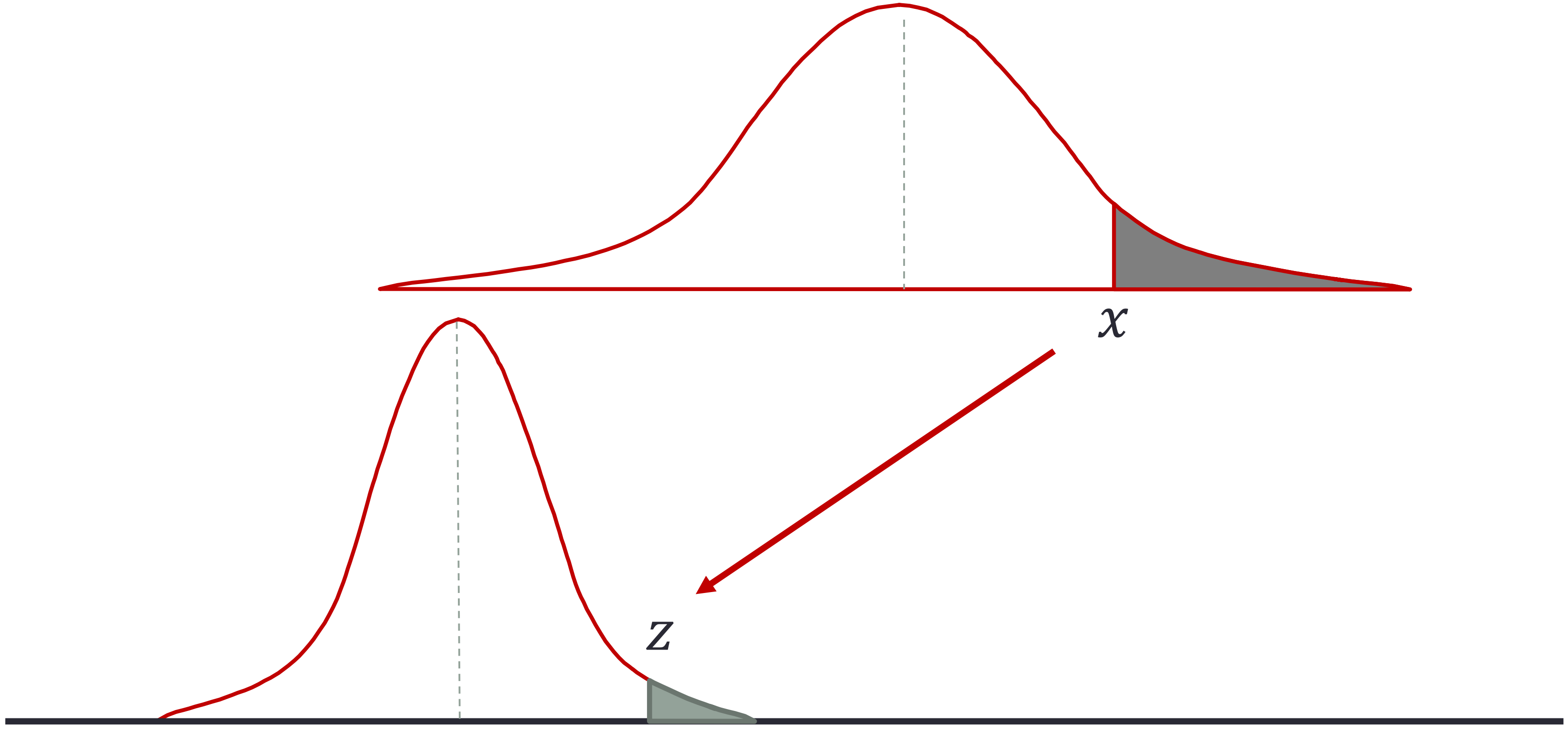
Conversion of Normal Distributions

- All Normal distributions can be converted into standard Normal distributions for ease of computing probabilities under the curve.



Conversion of Normal Distributions

- All Normal distributions can be converted into standard Normal distributions for ease of computing probabilities under the curve.



Z-Scores

- Converting from any point on any Normal distribution (with mean μ and standard deviation σ) to the corresponding point on the standard Normal distribution can be done through **z-scores**.

$$z = \frac{x - \mu}{\sigma}$$

Z-Scores Example

- You manage a credit card company. The weekly number of transactions your customers have follows a Normal distribution with a mean of 20.5 with a standard deviation of 9. What is the probability that a random customer uses their credit card more than 28 times in a week (4/day average)?

Z-Scores Example

- You manage a credit card company. The weekly number of transactions your customers have follows a Normal distribution with a mean of 20.5 with a standard deviation of 9. What is the probability that a random customer uses their credit card more than 28 times in a week (4/day average)?

$$z = \frac{x - \mu}{\sigma} = \frac{28 - 20.5}{9} = 0.83$$

Z-Scores Example

$$z = \frac{x - \mu}{\sigma} = \frac{28 - 20.5}{9} = 0.83$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.		.	.	.	.	.	.
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7517	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
.	.	.	.	.	.	.	.	.	.	.

$P(z \leq 0.83)$

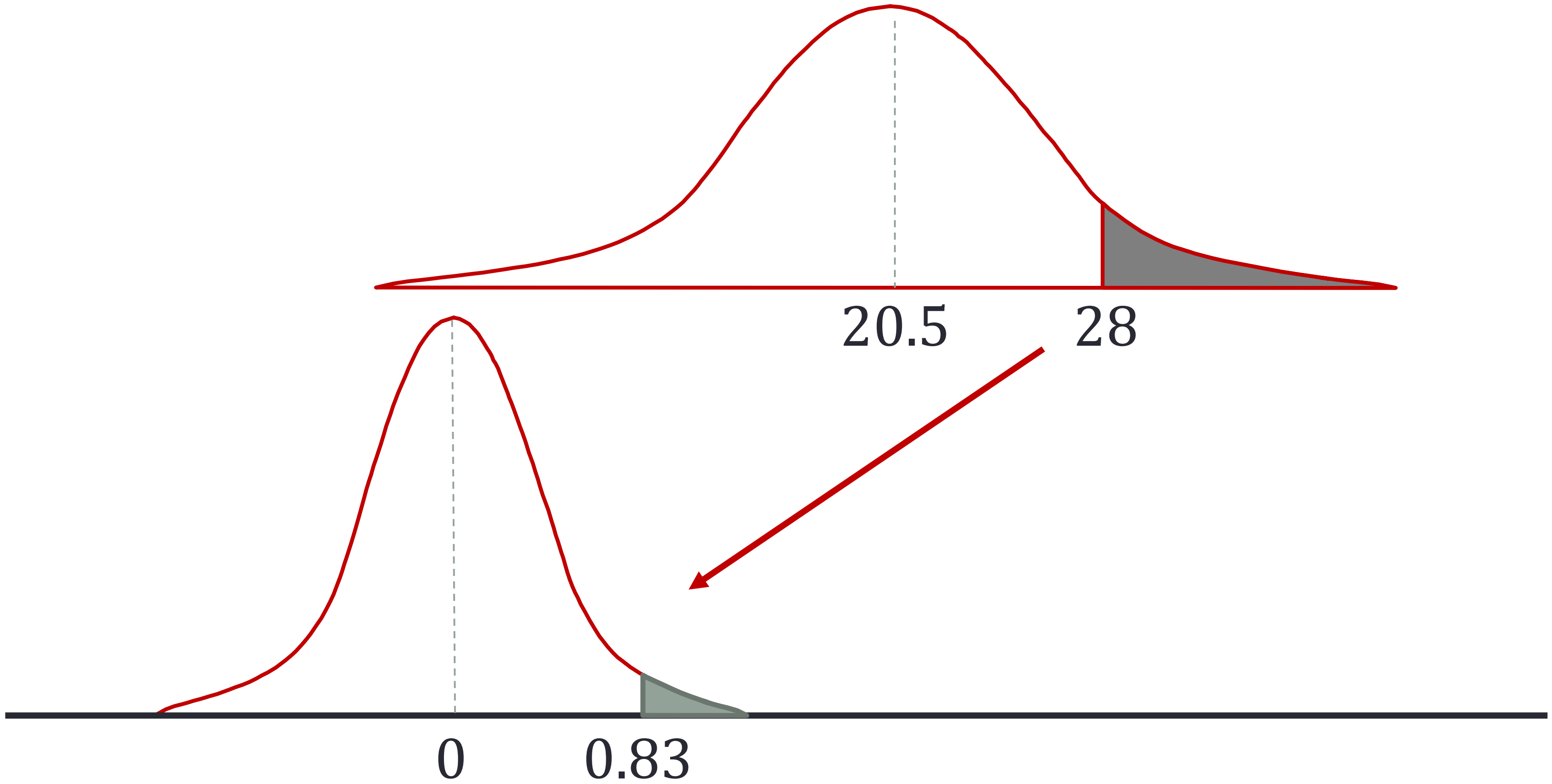
Z-Scores Example

- You manage a credit card company. The weekly number of transactions your customers have follows a Normal distribution with a mean of 20.5 with a standard deviation of 9. What is the probability that a random customer uses their credit card more than 28 times in a week (4/day average)?

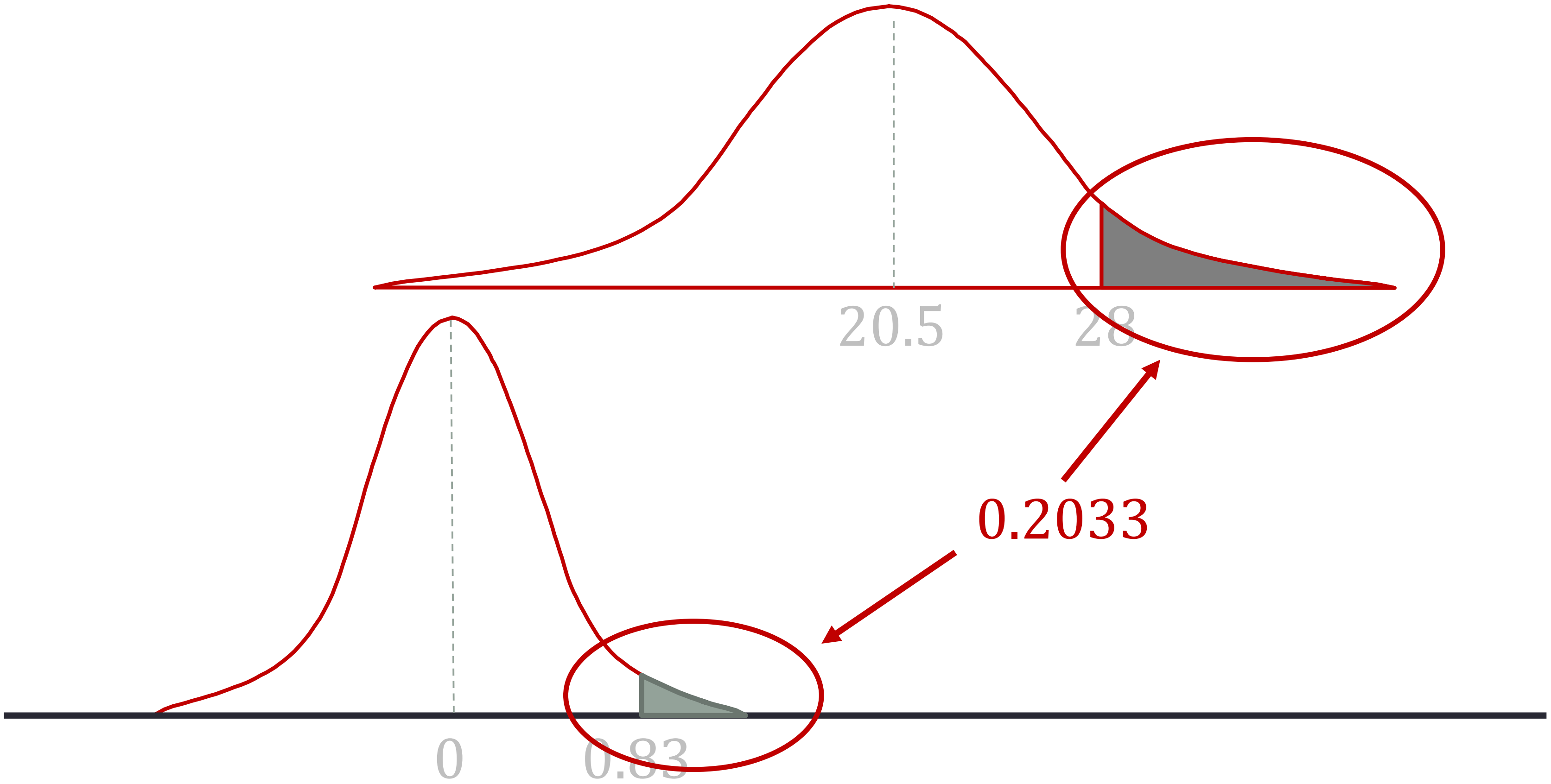
$$z = \frac{x - \mu}{\sigma} = \frac{28 - 20.5}{9} = 0.83$$

$$\begin{aligned} P(z > 0.83) &= 1 - P(z \leq 0.83) \\ &= 1 - 0.7967 \\ &= 0.2033 \end{aligned}$$

Z-Scores Example



Z-Scores Example



Z-Scores Example

- You manage a credit card company. The weekly number of transactions your customers have follows a Normal distribution with a mean of 20.5 with a standard deviation of 9. What is the number of transactions for the lower 10% of customers?

Z-Scores Example

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
.	.	.	.	.	.	.	.	.	.	.

$$P(z \leq ?) = 0.1$$

Z-Scores Example

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.	.	.	.	.	.	.	.	.	.	.
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
.	.	.	.	.	.	.	.	.	.	.

$P(z \leq ?) = 0.1$ or qnorm(0.1) = -1.28

Z-Scores Example

- You manage a credit card company. The weekly number of transactions your customers have follows a Normal distribution with a mean of 20.5 with a standard deviation of 9. What is the number of transactions for the lower 10% of customers?

$$z = \frac{x - \mu}{\sigma} = -1.28 = \frac{x - 20.5}{9}$$

$$x = 8.98$$

More Examples

- Assume new employees at a company have previous years of professional experience that follow a normal distribution, find the following probabilities if the mean is 5 and the s.d. is 2.5:

1. New employee has more than 5 years of previous experience.

2. New employee has less than 2 years of previous experience.

More Examples

- Assume new employees at a company have previous years of professional experience that follow a normal distribution, find the following probabilities if the mean is 5 and the s.d. is 2.5:

1. New employee has more than 5 years of previous experience.

$$z = \frac{x - \mu}{\sigma} = \frac{5 - 5}{2.5} = 0 \to 0.5$$

2. New employee has less than 2 years of previous experience.

$$z = \frac{x - \mu}{\sigma} = \frac{2 - 5}{2.5} = -1.2 \to 0.1151$$

More Examples

- Assume new employees at a company have previous years of professional experience that follow a normal distribution, find the following probabilities if the mean is 5 and the s.d. is 2.5:

3. New employee has between 1 and 7.5 years of previous experience.

4. What is the 90th percentile of employee experience? What is the 10th percentile?

More Examples

- Assume new employees at a company have previous years of professional experience that follow a normal distribution, find the following probabilities if the mean is 5 and the s.d. is 2.5:

3. New employee has between 1 and 7.5 years of previous experience.

$$z = \frac{7.5 - 5}{2.5} = 1 \rightarrow 0.8413 \qquad z = \frac{1 - 5}{2.5} = -1.6 \rightarrow 0.0548$$

$$0.8413 - 0.0548 = 0.7865$$

4. What is the 90th percentile of employee experience? What is the 10th percentile?

$$1.28 = \frac{x - 5}{2.5} \rightarrow x = 8.2 \qquad -1.28 = \frac{x - 5}{2.5} \rightarrow x = 1.8$$