Descriptive Statistics

Basic Computations

What is **Descriptive Statistics**?

It is the term given to the analysis of data by using certain formulas or definition that ultimately helps describe, or summarize data in a meaningful way. It is commonly divided into **Central Tendency** and **Variability(Dispersion)**.

What are **Central Tendencies**?

Measures of central tendency include the **mean** , **median** and **mode**.

Finding Sample Mean (average)

What do we need to compute the **Sample Mean**?

- **Symbol**: $\bar{x}$
- **Sample Size**: n
- **Formula**: $\bar{x} = \dfrac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \dfrac{\sum x}{n}$

Example:

Find the mean of the sample $5, 7, 8, 5, 10, 4, 12,$ and 20 .

Solution:

$$\bar{x} = \frac{5 + 7 + 8 + 5 + +10 + 4 + 12 + 20}{8} = \frac{71}{8} = 8.875$$

Finding Sample Mode

What is the **Sample Mode**?

The **sample mode** is the most frequent observation that occurs in the data set.

- When no observation occurs the most, then data has no mode.
- When two observations occurs the most, then data is bimodal.
- When three observations occurs the most, then data is trimodal.

Example:

Find the mode of the sample $5, 7, 8, 5, 10, 4, 12,$ and 20 .

Solution:

The mode is 5 since it appeared the most.

Finding Sample Median

What is the **Sample Median**?

The **sample median** divides the bottom 50% of the sorted data from the top 50%.

How do we find the **Sample Median**?

- Arrange the data in ascending order.
- When the sample size n is odd, the median is the data element that lies in the $\frac{n+1}{2}$ position.
- When the sample size n is even, the median is the mean of the data elements that lie in the $\frac{n}{2}$ position and $\frac{n}{2}+1$ position.

Finding Sample Median

Example:

Find the median of the sample $62, 68, 71, 74, 77, 82, 84, 88, 90,$ and 98 .

Solution:

This data is already sorted and $n = 10$ is even, then we find the mean of the fifth $\left(\frac{n}{2} = \frac{10}{2} = 5\right)$ and sixth $\left(\frac{n}{2} + 1 = \frac{10}{2} + 1 = 6\right)$ data element.

$$\text{Median} = \frac{77 + 82}{2} = 79.5$$

Finding Sample Median

Example:

Find the median of the sample
$12, 15, 15, 17, 19, 19, 23, 25, 27, 30, 31, 33, 35, 40,$ and 50.

Solution:

This data is already sorted and $n = 15$ is odd, then the median is the eighth $\left(\frac{n+1}{2} = \frac{15+1}{2} = 8\right)$ data element.

$$\text{Median} = 25$$

What is **Descriptive Statistics**?

It is the term given to the analysis of data by using certain formulas or definition that ultimately helps describe, or summarize data in a meaningful way. It is commonly divided into **Central Tendency** and **Variability(Dispersion)**.

What is the measure of **Variability(Dispersion)**?

Measures of how data elements vary or dispersed with respect to the sample mean. This measure includes the **sample variance** , and **sample standard deviation**.

Finding Sample Variance

What do we need to find the Sample Variance?

- **Symbol**: S^2
- **Sample Size**: n
- **Sample Mean**: $\bar{x}$
- **Formula**: $S^2 = \dfrac{\sum (x - \bar{x})^2}{n - 1}$
- **Formula**: $S^2 = \dfrac{n \sum x^2 - (\sum x)^2}{n(n - 1)}$

While we can use technology to find the **sample variance** , it is a lot easier to use the second formula to find the **sample variance** .

Finding Sample Variance

Example:

Find the variance of the sample 8, 5, 10, 7, 5, 4, 8, and 6.

Solution:

We can begin this process by making a table.

x	8	5	10	7	5	4	8	6	$\sum x = 53$
x^2	64	25	100	49	25	16	64	36	$\sum x^2 = 379$

Using the second formula for the variance, we get

$$S^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{8 \cdot 379 - (53)^2}{8 \cdot (8-1)} = \frac{223}{56}$$

Finding Sample Standard Deviation

What is the **Sample Standard Deviation**?

The **sample standard deviation** is a non–negative numerical value which shows the variation among all data elements with respect to the sample mean.

- When the value of the standard deviation is zero, then there in no deviation in the data set.
- When the value of the standard deviation is small, then data elements are close to the sample mean.
- When the value of the standard deviation is large, then data elements are not as close to the sample mean.

Finding Sample Standard Deviation

What do we need to find the **Sample Standard Deviation**?

- **Symbol**: S
- **Compute**: S^2
- **Formula**: $S = \sqrt{S^2}$

While we can find the value of the **sample standard deviation** by first finding the value of the **sample variance** , it is a lot easier and less time consuming to use technology to find **sample standard deviation**.

Finding Sample Mean, Variance, and Standard Deviation

Example:

Find the mean, variance, and standard deviation of the sample with $n = 15, \sum x = 303$ and $\sum x^2 = 6281$.

Solution:

Using the formulas that we have learned, we get

$$\bar{x} = \frac{\sum x}{n} = \frac{303}{15} = 20.2,$$

$$S^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{15 \cdot 6281 - (303)^2}{15 \cdot (15-1)} = \frac{401}{35},$$

$$S = \sqrt{S^2} = \sqrt{\frac{401}{35}} = 3.385.$$

Working With Grouped Data

How do we find the $\bar{x}$, S^2, and S for a grouped data?

- Compute all **Class Midpoints** which is the average of lower and upper class limits for each class and then update the frequency distribution table.
- Compute the sample size n by computing $\sum f$.
- Compute $\sum f \cdot x$, and $\sum f \cdot x^2$.
- Now we use the following formulas to complete this task:
  1. $\bar{x} = \dfrac{\sum f \cdot x}{n}$
  2. $S^2 = \dfrac{n \sum f \cdot x^2 - (\sum f \cdot x)^2}{n(n-1)}$
  3. $S = \sqrt{S^2}$

Example:

Use the frequency distribution table below,

Class Limits	Class Midpoints	Class Frequency
15 - 29		7
30 - 44		15
45 - 59		12
60 - 74		6

to find $\bar{x}$, S^2, and S.

Solution:

We first compute each class midpoint, and update the frequency distribution table.

Class Limits	Class Midpoints	Class Frequency
15 - 29	$\frac{15+29}{2} = \frac{44}{2} = 22$	7
30 - 44	$\frac{30+44}{2} = \frac{74}{2} = 37$	15
45 - 59	$\frac{45+59}{2} = \frac{84}{2} = 42$	12
60 - 74	$\frac{60+74}{2} = \frac{134}{2} = 67$	6

Solution Continued:

Now we start computing to complete the process.

- $n = \sum f = 7 + 15 + 12 + 6 = 40.$
- $\sum f \cdot x = 7 \cdot 22 + 15 \cdot 37 + 12 \cdot 42 + 6 \cdot 67 = 1615.$
- $\sum f \cdot x^2 = 7 \cdot 22^2 + 15 \cdot 37^2 + 12 \cdot 42^2 + 6 \cdot 67^2 = 72025.$
- $\bar{x} = \frac{\sum f \cdot x}{n} = \frac{1615}{40} = 40.375.$
- $S^2 = \frac{n \sum f \cdot x^2 - (\sum f \cdot x)^2}{n(n-1)} = \frac{40 \cdot 72025 - (1615)^2}{40(40-1)} = \frac{272775}{1560} = \frac{18185}{104}$
- $S = \sqrt{S^2} = \sqrt{\frac{18185}{104}} \approx 13.223$

Estimating Sample Standard Deviation

What is the **Range Rule–of–Thumb**?

The **Range Rule–of–Thumb** is a method to estimate the value of the **sample standard deviation** and is given by $S \approx \frac{\text{Range}}{4}$.

Example:

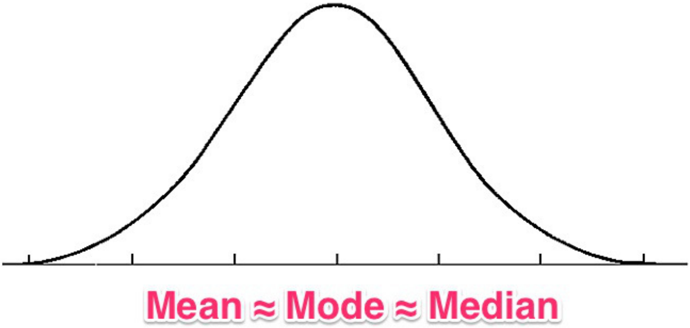
Estimate the value of the sample standard deviation of the sample with the minimum 54 and the maximum 97.

Solution:

$$S \approx \frac{\text{Range}}{4} = \frac{97 - 54}{4} = \frac{43}{4} = 10.75$$

What is a **Bell-Shaped Distribution**?

A data has a approximately **Bell-Shaped** distribution when the **mean** , **mode** , and **median** are equal or approximately equal.

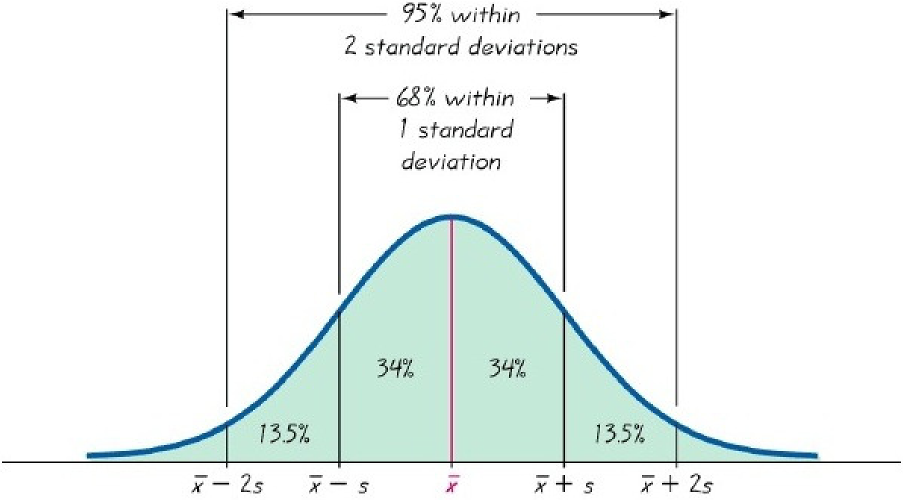


What is the **Empirical Rule**?

The **Empirical Rule** provides a quick estimate of the spread of data in a **Bell-Shaped** distribution given the mean and standard deviation.

What are the properties of the **Empirical Rule**?

- About 68% of all values fall within 1 standard deviation of the mean.
- About 95% of all values fall within 2 standard deviations of the mean.
- About 99.7% of all values fall within 3 standard deviations of the mean.


95% within
2 standard deviations
68% within
1 standard
deviation
34%
34%
13.5%
13.5%
$\bar{x} - 2s$
$\bar{x} - s$
$\bar{x}$
$\bar{x} + s$
$\bar{x} + 2s$

Example:

Find the 68% and 95% ranges of a bell-shaped distributed sample with the mean of 74 and standard deviation of 6.5.

Solution:

Since the data has a bell-shaped distribution, we can use the empirical rule to find the 68% and 95% ranges.

- For 68% range $\Rightarrow$ We compute $\bar{x} \pm s$.
  - $\bar{x} - s = 74 - 6.5 = 67.5$, and $\bar{x} + s = 74 + 6.5 = 80.5$.
  - So about 68% of the data falls within 67.5 and 80.5.
- For 95% range $\Rightarrow$ We compute $\bar{x} \pm 2s$.
  - $\bar{x} - 2s = 74 - 2(6.5) = 61$, and $\bar{x} + 2s = 74 + 2(6.5) = 87$.
  - So about 95% of the data falls within 61 and 87.

Standardizing Data

What is the **Z–Score**?

The number of **standard deviations** that a given data value is above or below the **mean** and can by computed by $Z = \dfrac{x - \bar{x}}{S}$. Round answers to 3–decimal places.

Example:

Lisa scored 82 on her exam. Find her Z–score if the class average was 73.4 with standard deviation of 5.3.

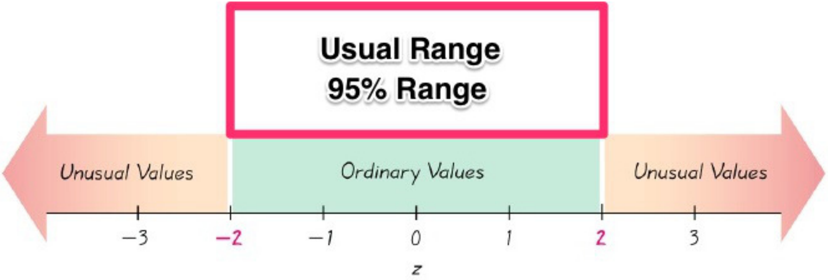
Solution:

$$Z = \frac{x - \bar{x}}{S} = \frac{82 - 73.4}{5.3} = \frac{8.6}{5.3} = 1.623$$

What are **Unusual** and **Ordinary** values?

Any data value that its **Z score** falls within -2 and 2 is considered an **ordinary** or **Usual** value.

The chart below clearly shows how to identify **Ordinary** and **Unusual** values.



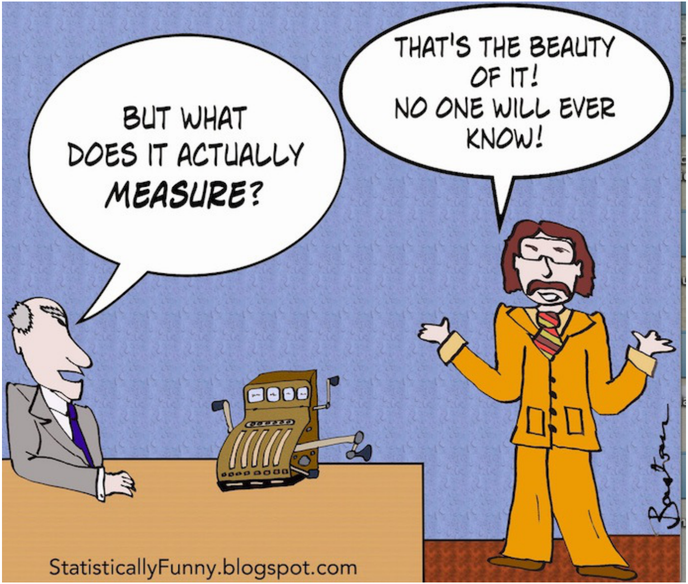
Example:

John makes a monthly salary of \$5750 as a nurse at the local hospital. The average salary for 25 randomly selected nurses was \$5275 with standard deviation of \$225. Find

1. Find the usual range of salaries according to the empirical rule.
2. Find the Z–score for John's salary.
3. Is John's salary considered to be ordinary or unusual?

Solution:

1. The usual range $\Rightarrow 5275 \pm 2(225) \Rightarrow 4825$ to 5725.
2. Z–score $\Rightarrow Z = \dfrac{x - \bar{x}}{S} = \dfrac{5750 - 5275}{225} = \dfrac{475}{225} = 2.111$
3. Ordinary or unusual? $\Rightarrow$ Unusual


BUT WHAT DOES IT ACTUALLY MEASURE?
THAT'S THE BEAUTY OF IT! NO ONE WILL EVER KNOW!
StatisticallyFunny.blogspot.com