Frequencies & Measures of Central Tendency

PSY 3307

Jonathan A. Pedroza, PhD

Cal Poly Pomona

2022-02-01

Agenda

- Frequencies
- Measures of Central Tendency
  - Mean
  - Median
  - Mode

Quick Note

The book goes over this kind of fast, so I'm going to use additional materials to hopefully better explain these topics

Frequencies

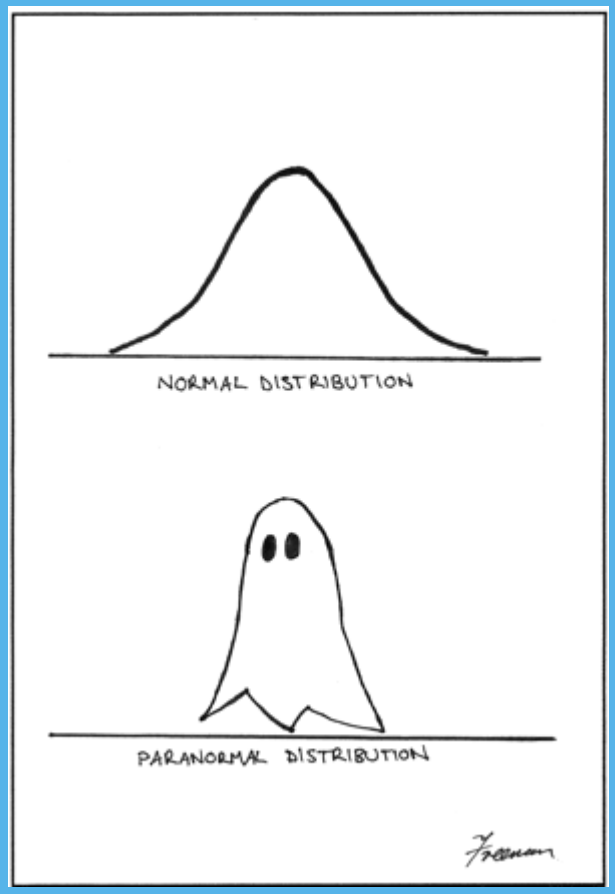
- **raw score** is the score given to a participant
- **Frequency** denoted as f; number of times a score occurs/is counted
    - *Note.* Not F, that is something completely different.
- **Frequency Distribution** is a distribution of each score and the number of times the score has occurred/is counted
    - often viewed through a **histogram**, or a plot that has the observations on the x-axis and the count/number of times a score has occurred on the y-axis

Frequency Distributions

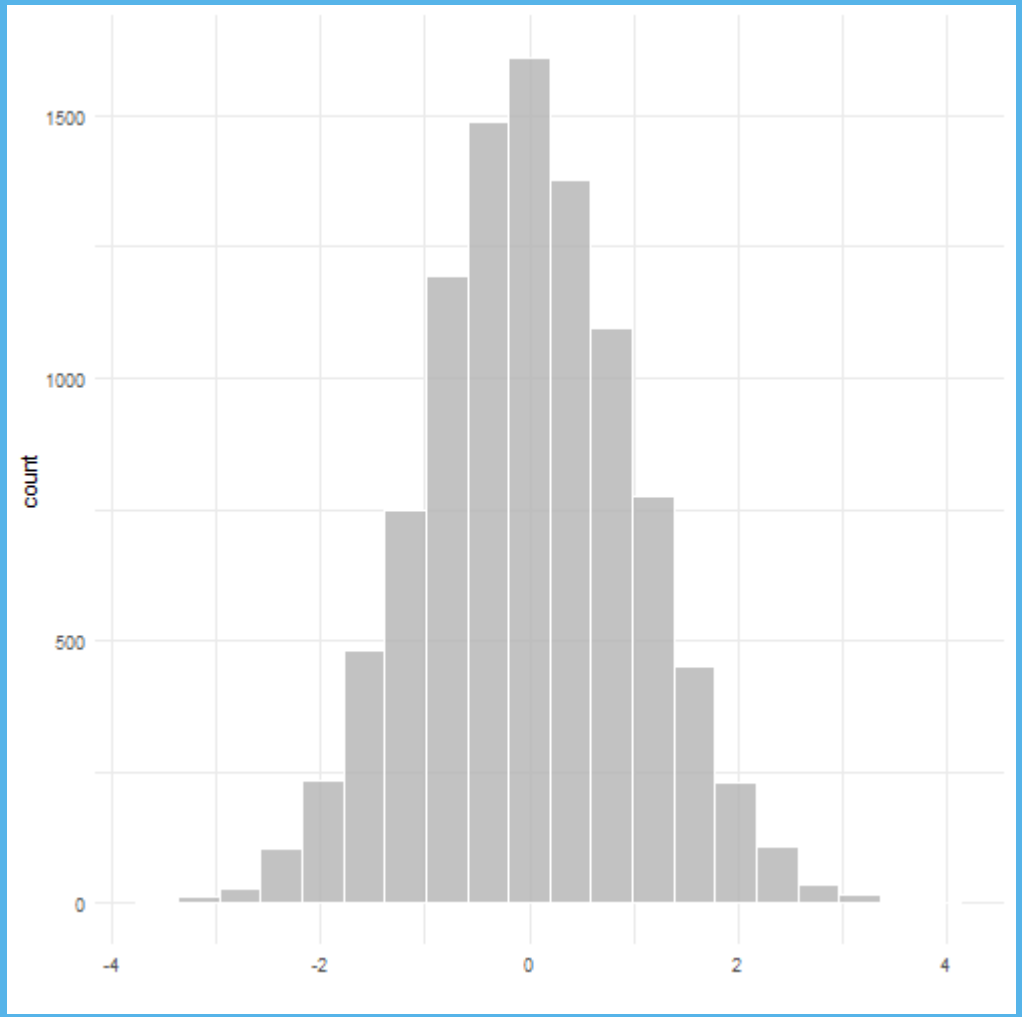
- the most common frequency distribution is the **normal distribution/normal curve**
  - where the number of scores are equal on both sides if a line down the middle is drawn

Normal Distribution

- **Normal curve** is often called the bell-shaped curve; is symmetrical
- represents the population because if you have enough data you will get a normal distribution (central limit theorem)
    - if your data looks like this in a histogram, you're in good shape
- **Distribution Tail** are the ends of a histogram/normal distribution
    - has two tails

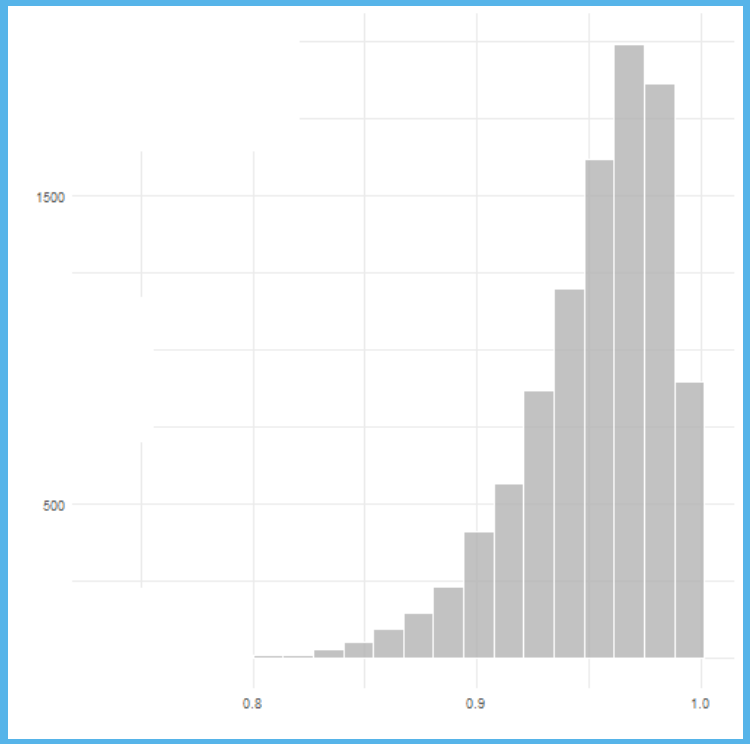


Normal Distribution



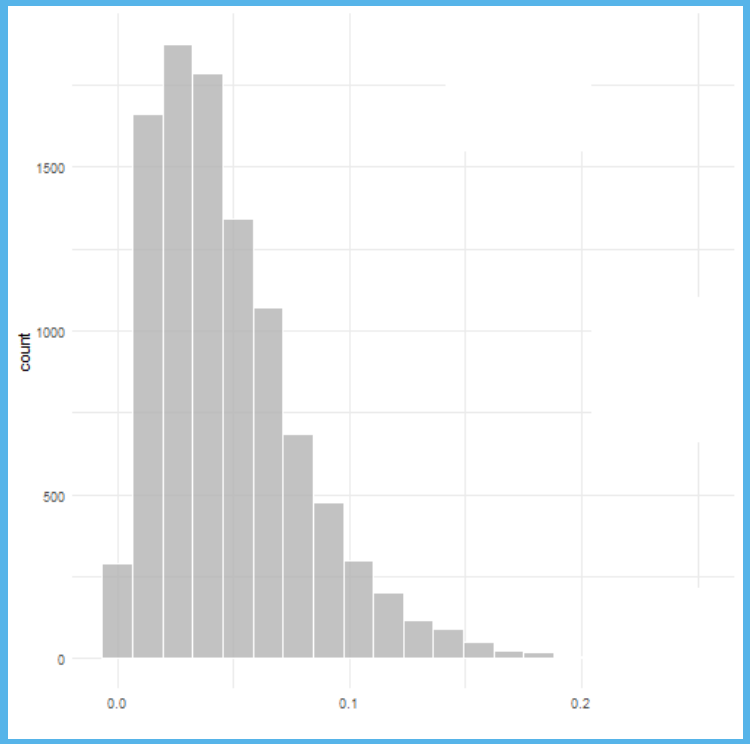
Skewed Distributions

- **Negative Skew** is like a finger pointing left
  - not normal and is asymmetrical
  - indicates higher frequency of middle and higher scores
  - no low frequency in higher scores



Skewed Distributions

- **Positive Skew** is like a finger pointing right
  - not normal and is asymmetrical
  - indicates higher frequency of low and middle scores
  - no low frequency in lower scores



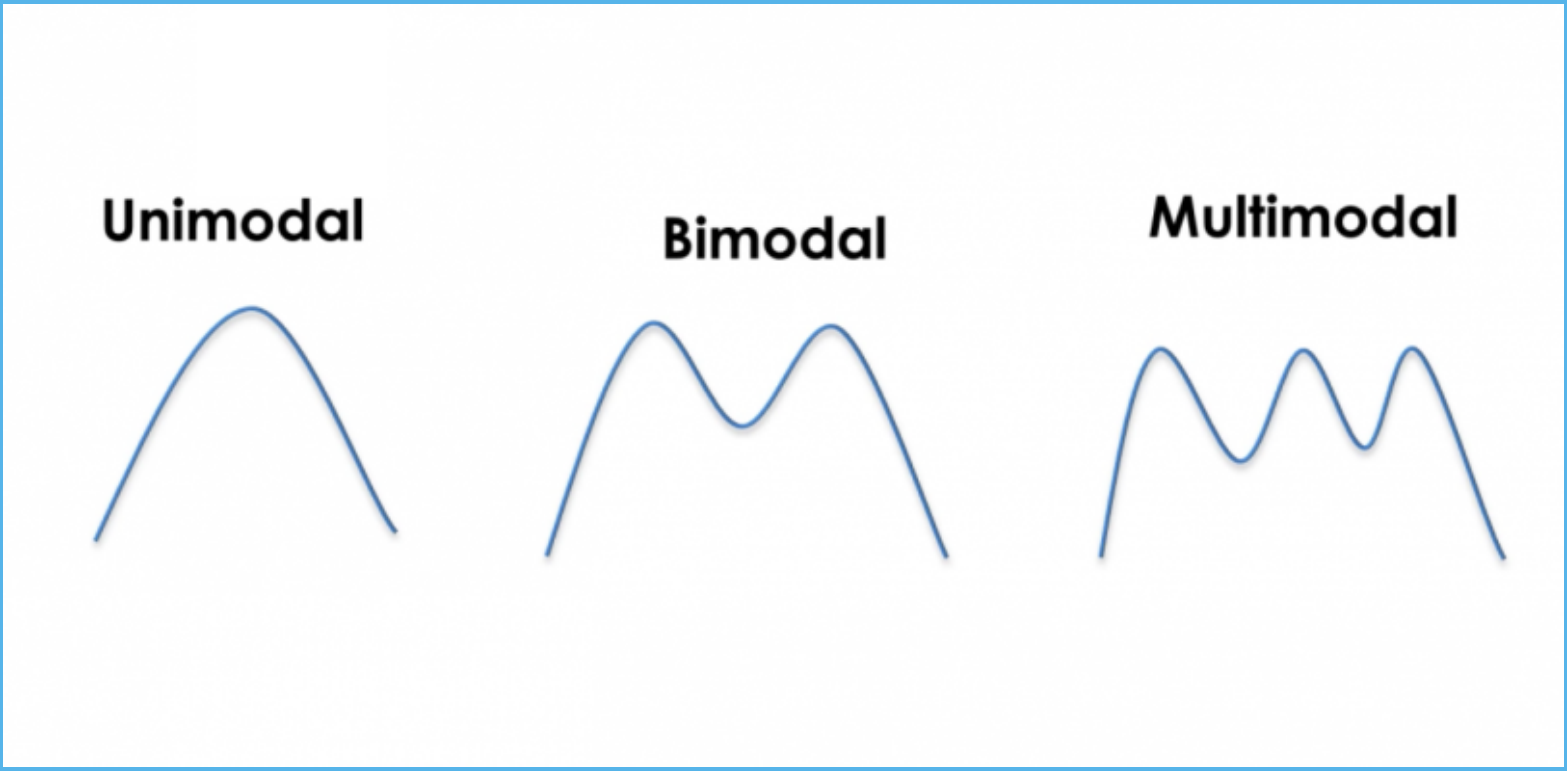
Kurtosis

- **Kurtosis** can be thought of as the pointyness of your distribution
- a **leptokurtic** distribution is reflective of positive kurtosis
  - extremely pointy
- a **platykurtic** distribution is reflective of negative kurtosis
  - extremely flat

Bimodal Distributions

- **Bimodal Distribution** is when your distribution has two modes (two highest points on a plot)
    - means that there are two areas of your distribution with higher frequencies of scores
- **Multimodal Distributions** are when your distribution has more than two modes (three or more highest points on a plot)
    - means that there are multiple areas of your distribution with higher frequencies of scores

Distribution Examples



Some Notes From JP

- Deciding what is positively and negatively skewed from visuals alone is not good enough
- You'll want to look into your descriptive statistics and look at skewness and kurtosis to make sure they are within +-3
- Some statistics can handle some skewness and kurtosis
- Other times you'll have to transform the variable using fancy methods that we will not talk about in this class.

Relative Frequency

- **Relative Frequency** the proportion of times a score occurred/is counted in the distribution

$$Relative\ frequency = f/N$$

- `f` is the frequency of one category of a categorical variable. `N` is the total number of observations for all the categories of that variable.
- **cumulative frequency** is the number of scores in the data at or below a particular score
    - add up the frequencies that make up that part of the histogram

Relative Frequency

```
## [1] 3 5 3 4 2 3
```

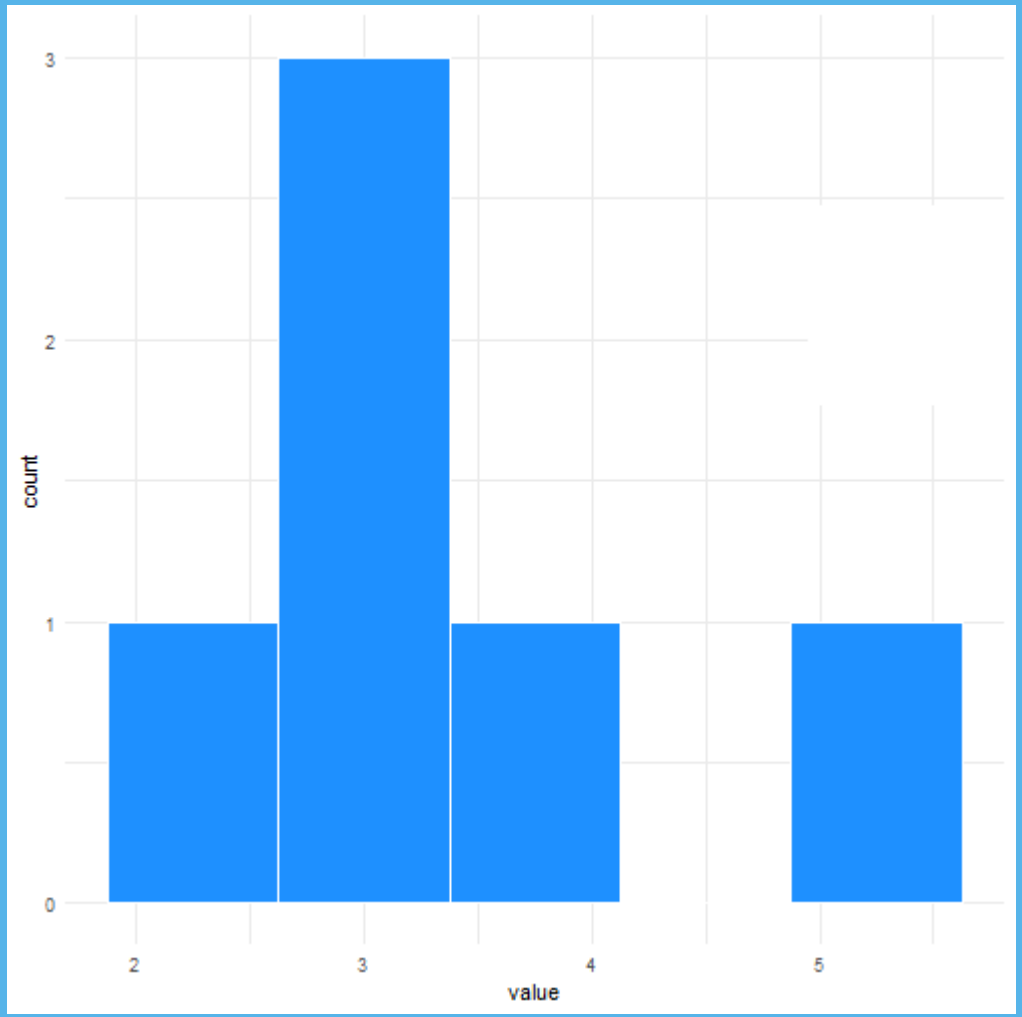
$$Count\ of\ 3 = f/N$$

$$Count\ of\ 3 = 3/6$$

$$Relative\ Frequency = 3/6 = .50$$

$$Percentage = 3/6 = .50 * 100 = 50\%$$

Histogram Visual



Terminology

$$\Sigma = Sum\ of\ scores$$

$$\Sigma X = Sum\ Of\ X$$

Measures of Central Tendency

- **Measures of Central Tendency** are statistics that summarize the location of a distribution on a variable by indicating where the center is
- A Normal distribution will have the central point right down the center
- A skewed distribution will have the central point where the frequency of scores is the highest

Measures of Central Tendency

- **mean**
  - an average score
  - sum of the scores divided by the number of observations
- **median**
  - middle value of a distribution when scores are ranked from lowest to highest
- **mode**
  - value(s) with the highest frequency

Mode

- Value/score with the highest frequency
- Not used in inferential statistics, mainly for descriptive statistics
- useful for telling whether a distribution is bimodal/multimodal or normal

Median

- will be the middle of the distribution
    - the 50% percentile
- median can be very useful for unusual data
- the symbol in literature is sometimes
- *Important Note* If you have an even number of scores/values, then you will add the two middle values and divide by 2

$$Mdn = Median$$

Median in Action

Values

```
## [1] 1 3 5 4 2 1 6
```

```
## [1] 1 3 5 4 2 1 6 4
```

Sorted Values

```
## [1] 1 1 2 3 4 5 6
```

```
## [1] 1 1 2 3 4 4 5 6
```

Calculation for Even Number of Values

```
(3+4)/2
```

```
## [1] 3.5
```

Mean

- is the score located at the `mathematical` center of a distribution
- `Xbar` is often used for the mean
- two formula examples below
- Calculate the mean for interval and ratio scales
    - The mean of ordinal/ordered data makes no sense
- Basis for most `inferential statistics`

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{\Sigma X}{n}$$

Using the Mean

```
##  [1]  9.17  4.03  6.82  0.56  7.36  9.06  8.26  7.77  8.97 10.92
```

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{(9.17 + 4.03 + 6.82 + 0.56 + 7.36 + 9.06 + 8.26 + 7.77 + 8.97 + 10.92)}{10}$$

```
(9.17+4.03+6.82+0.56+7.36+9.06+8.26+7.77+8.97+10.92)
```

```
## [1] 72.92
```

```
# total is 72.92
```

$$\overline{X} = \frac{72.92}{10}$$

```
72.92/10
```

```
## [1] 7.292
```

```
# mean is 7.29
```

$$\overline{X} = 7.29$$

When The Median Shines

- Preferred for ordinal/ordered data
- Not the best option for normally distributed interval & ratio scores
- Is more reliable when dealing with skewed data
    - The mean will be pulled toward the extreme tail of the distribution

The Mean in Research

- Most statistics revolve around the mean
- Can't just trust descriptive statistics like the mean