



Prifysgol Cymru  
Y Drindod Dewi Sant  
University of Wales  
Trinity Saint David

200  
1822 - 2022

# FUNDAMENTALS in STATISTICS

**Describing Univariate Data**

*Man Geni Addysg Uwch yng Nghymru  
Where Higher Education Began in Wales*

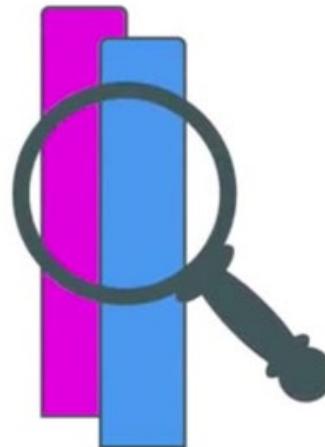
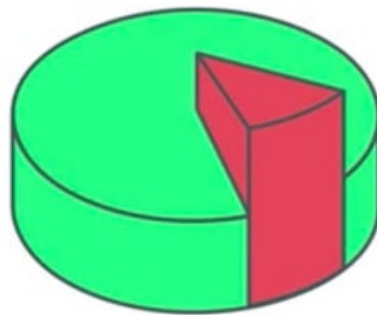
[www.uwtsd.ac.uk](http://www.uwtsd.ac.uk)

# Objectives

- Basic definition of Statistics
- Classification of Statistics
- Variables
- Types of Data
- Scales of Measurement
- Frequency distribution

## What are statistics?

Statistical analysis is the process of collecting, organizing, analyzing, and drawing conclusions from data to uncover patterns and relationships



Statistics

- extract meaningful insights from vast datasets
- Understanding statistical methods enables individuals to uncover patterns, trends, and relationships within data, leading to more informed insights.
- statistics plays a vital role in ensuring the reliability, validity, and generalizability of the findings from the clinical trial

# Classifications:

Statistics is broadly divided into two main areas:

- **Descriptive statistics**

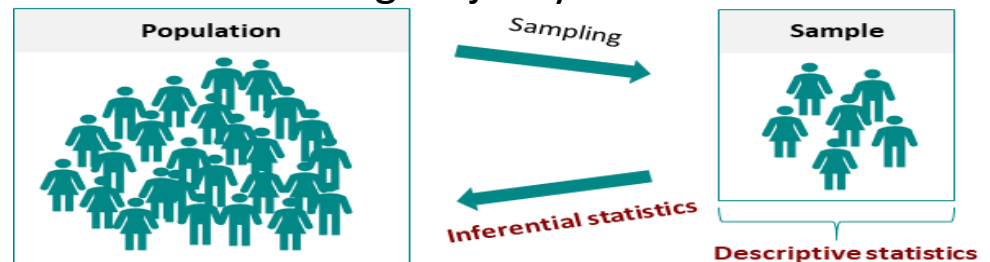
- involves methods for summarizing and describing the main features of a dataset.

Example: A graph showing the annual change in global temperature during the last 30 years;

- **Inferential statistics:**

- involves making inferences and predictions about a population based on a sample of data.

Example: a pollster's claim that an overwhelming majority of all U.K. voters favor stronger gun control laws;



## Activity:

Let's categorize each statement into either descriptive or inferential statistics

1. The average income of a sample of 500 employees in a company is £5,000.
2. Based on a survey of 1,000 customers, we estimate that 70% of our customers prefer online shopping
3. The majority of students in the class scored between 70% and 80% on the final exam
4. A study finds a significant correlation between regular exercise and improved mental health across a diverse population
5. Analysis of historical data shows a consistent pattern of increased sales during holiday seasons over the past decade.
6. After analyzing a representative sample of 1,000 voters, we predict with 95% confidence that the incumbent candidate is likely to win the upcoming election

# Variables

Variables represent things that differ between individuals or situations

- People:  
height, weight, blood pressure, body temperature, eye colour, shoe size, favourite colour etc...
- Computer systems:  
Processor, memory, chipset, network card, speed, graphics card, network speed etc...
- Businesses:  
number of employees, share price, sector, annual profit, profit margin, type, size etc...

# Variables

| Category I | Category II           |
|------------|-----------------------|
| Discrete   | Independent variables |
| Continuous | Dependent variables   |

# Variables

- Classification into discrete and continuous is related to the nature of the data

## Examples:

- Number of students in a classroom (can't have a fraction of a student).
- Number of cars in a parking lot.
- Number of goals scored in a soccer match.

## Examples:

- Height of individuals (can be measured with high precision).
- Weight of objects.
- Temperature in degrees Celsius or Fahrenheit.

## Independent Variable

- This is the variable that is manipulated or changed by the researcher.
- It's called "independent" because its variation is not influenced by other variables in the experiment.
- It's often plotted on the x-axis of a graph.

### Example

In an experiment to test the effect of fertilizer on plant growth, the **amount of fertilizer applied** would be the independent variable.

## Dependent Variable

- This variable is what you measure in the experiment and is affected by the independent variable.
- It's called "dependent" because its value relies on the changes made to the independent variable.
- It's typically plotted on the y-axis of a graph.

### Example

- In the fertilizer experiment, the dependent variable would be the **height or growth of the plants**, as it depends on the amount of fertilizer used.

# THREE TYPES OF DATA

- Qualitative Data

A set of observations where any single observation is a word(Yes/No), letter(Y/N), or numerical code(0/1) that represents a class or category.

- Ranked Data

A set of observations where any single observation is a number that indicates relative standing.

(1st, 2nd, ... 40th place)

- Quantitative Data

A set of observations where any single observation is a number that represents an amount or a count.

(weights of 28, 17, ... 18 Kgs)

Indicate whether each of the following terms is **qualitative** (because it's a word, letter, or numerical code representing a class or category); **ranked** (because it's a number representing relative standing); or **quantitative** (because it's a number representing an amount or a count)

- ethnic group
- age
- family size
- academic major
- sexual preference
- IQ score
- net worth (dollars)
- third-place finish
- gender
- temperature

|                                    |                                |
|------------------------------------|--------------------------------|
| (a) ethnic group - Qualitative     |                                |
| (b) Age- Qualitative               | (h) third-place finish- ranked |
| (c) family size- Quantitative      | (i) gender- Qualitative        |
| (d) academic major-Qualitative     | (j) temperature-Quantitative   |
| (e) sexual preference-Qualitative  |                                |
| (f) IQ score-Quantitative          |                                |
| (g) net worth(Pounds)-Quantitative |                                |

# Scales of measurement

Specifies the extent to which a number (or word or letter) actually represents some attribute.

Three levels of measurement-nominal, ordinal, and interval/ratio

These levels are paired with qualitative, ranked, and quantitative data, respectively

## 1.Nominal Variables:

Categories with no inherent order or ranking.

### Example:

- Colors (red, blue, green)
- Gender (male, female, non-binary)
- Types of fruit (apple, orange, banana)

## 2. Ordinal Variables:

Categories with a meaningful order or ranking.

### Example:

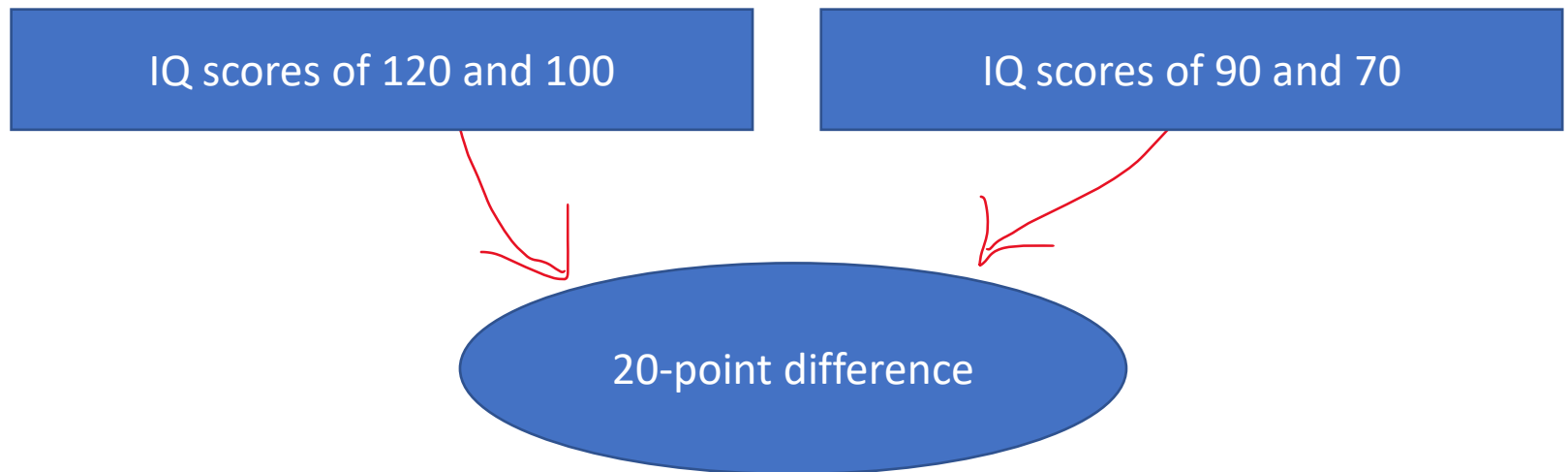
- Education levels (high school, bachelor's, master's).
- Customer satisfaction ratings (very dissatisfied, dissatisfied, neutral, satisfied, very satisfied)
- Socioeconomic status (low income, middle income, high income)

while the numbers assigned on an ordinal scale do reflect relative merit, the units of measurement are not equal

### 3. Interval Variables:

numbers that are assigned to variables reflect relative merit and **have equal units of measurement.**

#### Example



No true Zero point

**Example:** Temperature in Celsius or Fahrenheit.  
a temperature of  $0^{\circ}\text{C}$  or  $0^{\circ}\text{F}$  does not signify the complete absence of heat;

## 4. Ratio Variables:

Variables measured on an absolute scale **have a true zero point** and equal units of measurement.

**Example:** Age, income, weight.

### Implies:

- **Equal Intervals:** The difference between 2 kg and 3 kg is the same as the difference between 7 kg and 8 kg.
- **True Zero Point:** A weight of 0 kg implies the complete absence of weight.
- **Meaningful Ratios:** If one object weighs 6 kg and another weighs 3 kg, the ratio is 2:1, one object is twice as heavy as the other.
- **Arithmetic Operations:** All basic arithmetic operations are meaningful. For instance, you can add, subtract, multiply, or divide weights to perform various calculations.

Indicate the level of measurement-nominal, ordinal, or interval/ratio attained by the following sets of observations or data. Whenever appropriate, indicate that measurement is only approximately interval.

(a) height

(b) religious affiliation

(c) score for psychopathic tendency

(d) years of education

(e) military rank

(f) vocational goal

(g) GPA

(h) marital status

- |     |                                                  |
|-----|--------------------------------------------------|
| (a) | height-interval/ratio                            |
| (b) | religious affiliation-nominal                    |
| (c) | score for psychopathic tendency-approx. interval |
| (d) | years of education-interval/ratio                |
| (e) | military rank-ordinal                            |
| (f) | vocational goal-nominal                          |
| (g) | GPA-approx. interval                             |
| (h) | marital status-nominal                           |

# Frequency distributions

- collection of observations produced by sorting observations into classes and showing their frequency ( $f$ ) of occurrence in each class.

| WEIGHT | $f$ |
|--------|-----|
| 245    | 1   |
| 244    | 0   |
| 243    | 0   |
| 242    | 0   |
| *      |     |
| *      |     |
| *      |     |
| 161    | 0   |
| 160    | 4   |
| 159    | 1   |
| 158    | 2   |
| 157    | 3   |
| *      |     |
| *      |     |
| ↓      |     |

- A frequency distribution helps us to detect any pattern in the data
- Graphs of frequency distributions further aid our effort to detect data patterns and make sense out of the data

## Frequency distribution for ungrouped data

Progress Check: Students in a theater arts appreciation class rated the classic film 'The Wizard of Oz' on a 10-point scale, ranging from 1 (poor) to 10 (excellent), as follows:

|   |   |   |    |   |
|---|---|---|----|---|
| 3 | 7 | 2 | 7  | 8 |
| 3 | 1 | 4 | 10 | 3 |
| 2 | 5 | 3 | 5  | 8 |
| 9 | 7 | 6 | 3  | 7 |
| 8 | 9 | 7 | 3  | 6 |

| RATING | TALLY* | f  |
|--------|--------|----|
| 10     |        | 1  |
| 9      |        | 2  |
| 8      |        | 3  |
| 7      |        | 5  |
| 6      |        | 2  |
| 5      |        | 2  |
| 4      |        | 1  |
| 3      |        | 6  |
| 2      |        | 2  |
| 1      |        | 1  |
|        | Total  | 25 |

# Exercise

- Suppose you have an EQ test in which the students gained the following scores:

| Emotional Intelligence Test Scores |    |    |    |    |    |    |    |    |    |    |    |    |
|------------------------------------|----|----|----|----|----|----|----|----|----|----|----|----|
| 46                                 | 50 | 48 | 47 | 48 | 47 | 49 | 43 | 47 | 46 | 50 | 48 | 49 |
| 46                                 | 46 | 45 | 46 | 46 | 47 | 46 | 46 | 46 | 48 | 47 | 46 | 47 |
| 44                                 | 49 | 47 | 48 | 49 | 48 | 48 | 49 | 45 | 48 | 46 | 46 | 51 |
| 48                                 | 44 | 45 | 44 | 46 | 49 | 50 | 48 | 43 | 48 | 46 | 48 | 46 |
| 48                                 | 46 | 47 | 47 | 47 | 47 | 49 | 49 | 46 | 47 | 47 | 44 | 45 |
| 45                                 | 48 | 50 | 48 | 47 | 47 | 49 | 47 | 45 | 48 | 49 | 45 | 47 |
| 47                                 | 44 | 48 | 47 | 47 | 51 | 47 | 46 | 47 | 46 | 45 | 47 | 45 |
| 45                                 | 47 | 48 | 48 | 46 | 48 | 45 | 50 | 47 |    |    |    |    |

| Score | Tally | <i>f</i> |
|-------|-------|----------|
| 51    |       | 2        |
| 50    |       | 5        |
| 49    |       | 10       |
| 48    |       | 20       |
| 47    |       | 25       |
| 46    |       | 20       |
| 45    |       | 11       |
| 44    |       | 5        |
| 43    |       | 2        |

# Grouped frequency distributions

- observations are sorted into classes of more than one value
- Example:

Set of data representing the ages of students in a class. Construct a grouped Frequency distribution

18, 20, 21, 22, 23, 24, 25, 26, 27, 28,  
30, 31, 32, 33, 34, 35, 36, 38, 39, 40

Determine the range and decide on the class intervals (group sizes) that would suit the data.

# Grouped frequency distributions

Let's use a class interval of 5.

- The minimum value in the dataset is 18.
- The maximum value in the dataset is 40.
- The range is therefore  $40 - 18 = 22$ .

With a class interval of 5, create the following intervals:

- 15-19
- 20-24
- 25-29
- 30-34
- 35-39

Now, count how many values fall within each interval:

- 15-19: 1 (there's only one value, which is 18)
- 20-24: 6
- 25-29: 5
- 30-34: 5
- 35-39: 3

Solution:

| Age Range | Frequency |
|-----------|-----------|
| 15-19     | 1         |
| 20-24     | 6         |
| 25-29     | 5         |
| 30-34     | 5         |
| 35-39     | 3         |

# Exercise

- The IQ scores for a group of 35 high school dropouts are as follows:

|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 91  | 85  | 84  | 79  | 80  |
| 87  | 96  | 75  | 86  | 104 |
| 95  | 71  | 105 | 90  | 77  |
| 123 | 80  | 100 | 93  | 108 |
| 98  | 69  | 99  | 95  | 90  |
| 110 | 109 | 94  | 100 | 103 |
| 112 | 90  | 90  | 98  | 89  |

- Construct a frequency distribution for grouped data.
- Assume class interval as 5

# Solution

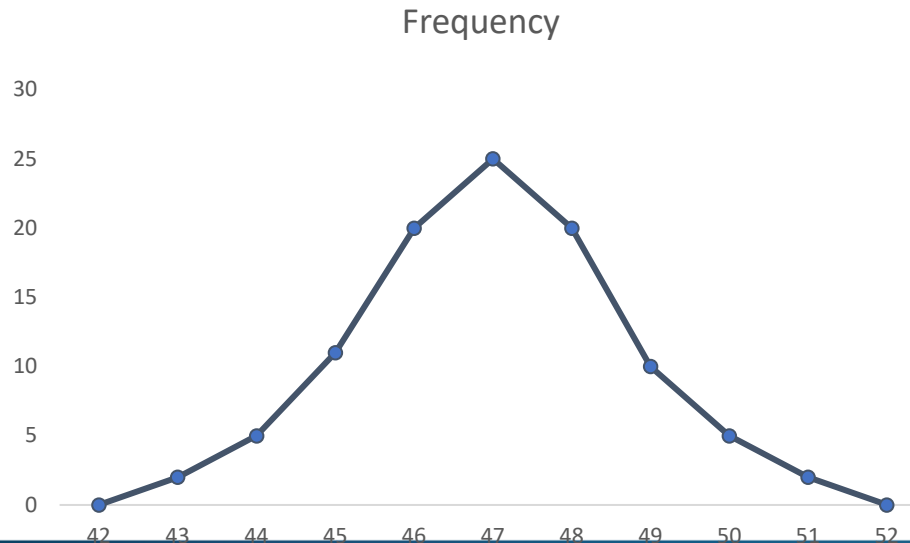
- The IQ scores for a group of 35 high school dropouts are as follows:

- Class interval:5

| IQ      | f |
|---------|---|
| 120-124 | 1 |
| 115-119 | 0 |
| 110-114 | 2 |
| 105-109 | 3 |
| 100-104 | 4 |
| 95-99   | 6 |
| 90-94   | 7 |
| 85-89   | 4 |
| 80-84   | 3 |
| 75-79   | 3 |
| 70-74   | 1 |
| 65-69   | 1 |

# Graphing Data

- Frequency distributions can also be displayed as bar charts (histograms) and frequency polygons
- The ***frequency polygon*** is one of the most commonly-used graphs in statistics
  - X-axis: Values or Intervals
  - Y-axis: Frequency



Data from the  
EQ test

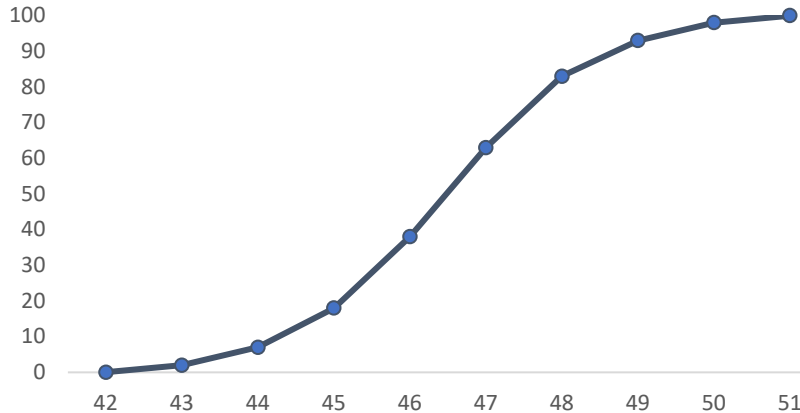


Microsoft Excel  
Worksheet

# Graphing Data

- Another graphical form is the ***cumulative frequency*** polygon
- This shows the number of scores occurring at and below a given point
  - It is also good for showing learning curves, where you might want to know how many trials it took to reach a particular level of expertise

Cumulative Frequency



Data from the EQ test

| Score | Tally | <i>f</i> |
|-------|-------|----------|
| 51    |       | 2        |
| 50    |       | 5        |
| 49    |       | 10       |
| 48    |       | 20       |
| 47    |       | 25       |
| 46    |       | 20       |
| 45    |       | 11       |
| 44    |       | 5        |
| 43    |       | 2        |

| Cumulative frequency |
|----------------------|
| 100                  |
| 98                   |
| 93                   |
| 83                   |
| 63                   |
| 38                   |
| 18                   |
| 7                    |
| 2                    |

# The Normal curve

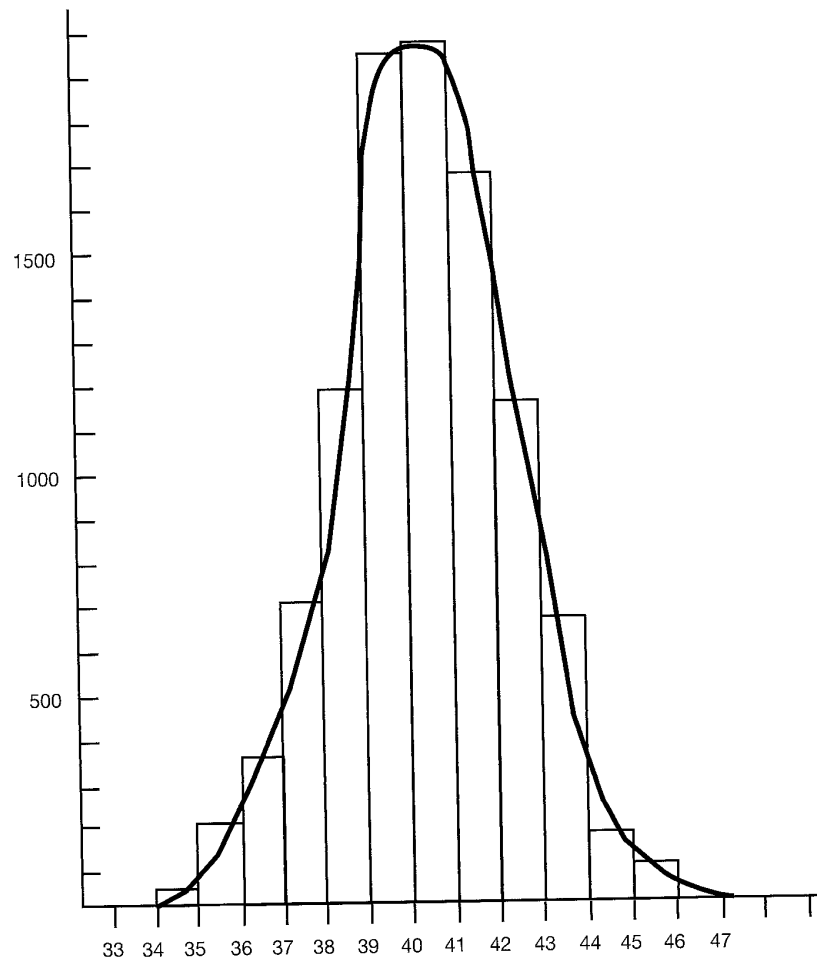
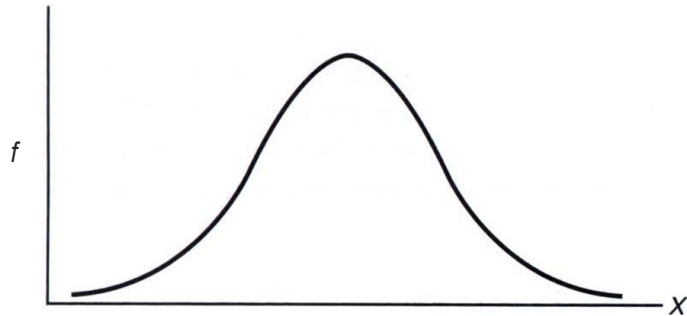


**FIGURE 4-3** Chest sizes of 5,738 soldiers.

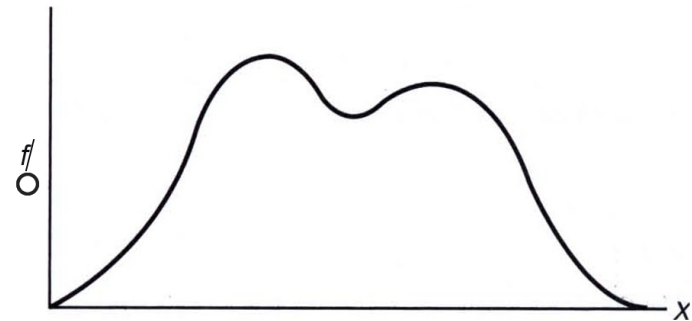
# Skewed distributions

- The Normal distribution is symmetrical, but some frequency curves are not
  - If a test is very easy most people get high scores
  - If a test is very difficult most people get low scores
- Depending on which way the bulk of the scores have shifted we can say if a curve is positively skewed or negatively skewed

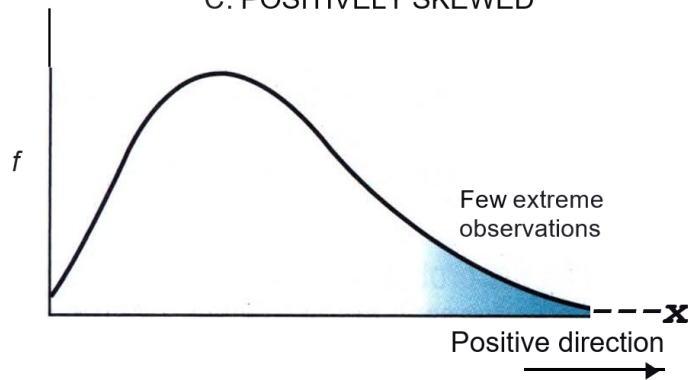
A. NORMAL



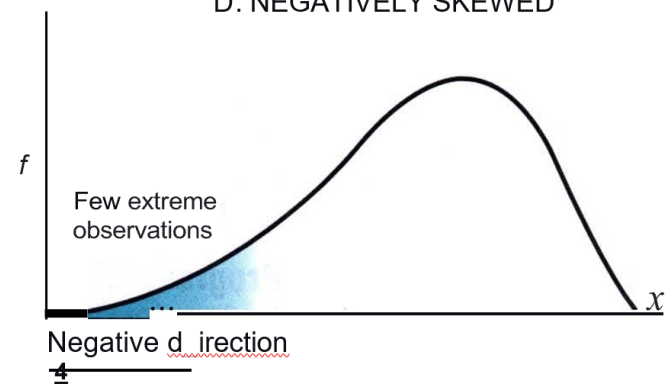
8. BIMODAL



C. POSITIVELY SKEWED



D. NEGATIVELY SKEWED



The following scores were obtained by third graders on a weekly spelling test (10 points possible):

| Scores |   |    |   |   |    |   |   |   |   |
|--------|---|----|---|---|----|---|---|---|---|
| 4      | 3 | 10 | 3 | 3 | 2  | 9 | 3 | 8 | 3 |
| 2      | 3 | 1  | 5 | 4 | 0  | 1 | 4 | 0 | 3 |
| 2      | 4 | 3  | 1 | 4 | 2  | 3 | 2 | 1 | 2 |
| 1      | 5 | 2  | 9 | 3 | 6  | 4 | 4 | 3 | 2 |
| 1      | 4 | 1  | 3 | 3 | 2  | 2 | 2 | 4 | 3 |
| 9      | 4 | 9  | 3 | 3 | 10 | 1 | 3 | 5 | 3 |
| 2      | 2 | 4  | 3 | 3 | 6  | 6 | 4 | 1 | 2 |
| 6      | 2 | 3  | 7 | 4 | 4  | 4 | 4 | 2 | 4 |

Find the frequency distribution and also cumulative frequency distribution

# Summary

- Basic of Statistics
- Classification of Statistics
- Variables
- Types of Data
- Scales of Measurement
- Frequency distribution
- Graphing data
- Normal Curve
- Next session we'll look at how to describe this data quantitatively rather than qualitatively

# Coming soon

- **Next session:** Descriptive statistics
- **Homework:** Read chapter 2 of *Statistics for the Terrified*.