STATISTICS

A STATISTICAL INVESTIGATION

Turn on the TV or flip through a newspaper, and you'll often spot statistics in action. For example:

- Su averages 14.6 points per basketball game.

- Last year was the hottest on record since 1897.

Statistics is the science of gathering, organizing, analyzing, interpreting, and presenting data. It helps us make smart decisions in all kinds of areas. Check out these real-world examples:

- **Scientific Research**: Testing if a new medicine works by studying trial results.

- **Industrial Production**: Improving products by tracking defects and fixing processes.

- **Social Issues**: Figuring out what people think about new laws through surveys.

A statistical investigation follows these five steps:

- **Step 1: State the Problem**: Decide what you want to learn.
  Example: How has the average temperature changed over the last 100 years?

- **Step 2: Collect Data**: Gather the info you need.
  Example: Get temperature records from weather stations.

- **Step 3: Calculate Descriptive Statistics**: Summarize the data with tools like mean, median, or mode.
  Example: Find the average temperature for each decade.

- **Step 4: Organize and Display Data**: Put the data in order and show it with charts or graphs.
  Example: Make a graph of temperature changes over time.

- **Step 5: Interpret the Statistics**: Figure out what the data tells you.
  Example: Does the data show temperatures are rising significantly?

By following these steps, you can dig into data and use it to make solid decisions!

> **Definition Statistics**
>
> **Statistics** is all about collecting information, sorting it out, summarizing it, and figuring out what it means.

B STATING THE PROBLEM

When you start a statistical investigation, the first step is to ask a clear question. This keeps you focused on what you're trying to find out and who or what you're studying.

We call the group we're studying the **population**. It could be all the people in a country, every student in a school, all the animals of a species, or even every item made by a machine. The information we collect from this group is called **data**, and it can come in many forms—like numbers, words, or measurements.

> **Definition Problem**
>
> A **problem** in statistics is a question that guides us to the information we need to find.

Ex: Do girls like math more than boys?

> **Definition Population**
>
> A **population** is the whole group of people or things with something in common that we want to study.

Ex: The population is all the students in a college.

> **Definition Data**
>
> **Data** is the information we collect, like numbers, words, measurements, or observations.

Ex: For our math study, we collect:

- **Gender**: Is the student a boy or a girl?

- **Favorite Subject**: What subject do they like best (e.g., Math, Science, English)?

- **Math test score**: What was their grade on the last assessment??

> **Definition Types of Variables**
> - **Qualitative Variable (Categorical)**: Describes categories or groups that cannot be measured numerically.
> - **Quantitative Variable (Numerical)**: Represents measurable quantities with numerical values.

Ex: For our math study:

- **Qualitative Variables**: Gender and favorite subject.

- **Quantitative Variable**: Math test score.

C COLLECTING DATA

To collect data, we first decide who or what we're asking. We can either:

- Do a **census**: Ask every single member of the population.

- Do a **survey**: Ask just a part of the population (a sample).

Why choose a survey? A census takes a lot of time and money, especially for big groups. A survey is faster and cheaper, and if we pick the sample well, it can still tell us a lot about the whole population!

> **Definition Census**
> A **census** means collecting data from everyone in the population.

> **Definition Survey**
> A **survey** means collecting data from a smaller group (sample) of the population.

Ex: If you ask every student in the collège about their favorite subject, is it a census or a survey?

Answer: It's a census.

Ex: If you only ask the students who are in class math today, is it a census or a survey?

Answer: It's a survey.

> **Method Making Census/Survey**
> To do the census/survey, follow these steps:
> 1. Ask each student your question, like "What is your favorite pet?"
> 2. Write down what each student says. You can:
>   - Write their name next to their answer (e.g., "Emma: Dog").
>   - Or use tally bars to count the answers (e.g., draw a tally mark for each "Dog").

D DESCRIPTIVE STATISTICS

Descriptive statistics are numbers that help us summarize and understand data—like finding the average or the most common answer.

> **Definition A statistics**
> A **statistics** is a single value that sums up or describes a set of data.

Ex: The average score in a class is 85% is a statistics number because it tells us something about the whole group in one simple figure.

E DESCRIPTIVE STATISTICS: RELATIVE FREQUENCY

In statistics, it's important to understand the frequency of a category. This concept helps us analyze patterns and make predictions. It applies to everyday scenarios, such as gauging the popularity of a favorite food among friends or calculating how often a basketball player scores a shot. By studying relative frequencies, we gain valuable insights into data trends.

Ex: The data for favorite subject is: Maths: 15 students, Sciences: 12 students, English: 3 students.
Fill in the table:

Subject	Frequency	Relative frequency (%)
Maths		
Sciences		
English		
Total		100%

Answer:

Subject	Frequency	Relative frequency (%)
Maths	15	$\frac{15}{30} \times 100\% = 50\%$
Sciences	12	$\frac{12}{30} \times 100\% = 40\%$
English	3	$\frac{3}{30} \times 100\% = 10\%$
Total	30	100%

F DESCRIPTIVE STATISTICS: CENTRAL TENDENCY

In statistics, central tendency refers to a measure that identifies a single value as representative of the center or typical point of a dataset. Three key measures are commonly used to assess central tendency: the mode, the mean, and the median.

Definition **Mode**

The **mode** is the value that shows up most often in your data.

Ex: A group of students reported their last mark (out of 5) on a math exam as follows:

$$1, 4, 2, 3, 5, 4, 5, 4, 3$$

What is the mode of this dataset?

Answer: From the frequency table:

Mark	**Frequency**
1	1
2	1
3	2
4	4
5	2

The mode is 4 because it appears most frequently (4 times).

Definition **Mean**

The **mean** is the average. Add up all the values and divide by how many there are:

$$\bar{x} = \frac{\text{sum of all values}}{\text{number of values}}$$
$$= \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Ex: Ratings: 1, 4, 2, 3, 5, 4, 5, 4, 4. What's the mean?

Answer:

$$\text{Mean} = \frac{1 + 4 + 2 + 3 + 5 + 4 + 5 + 4 + 4}{9}$$
$$= \frac{32}{9}$$
$$\approx 3.56$$

> **Definition Median**
>
> The **median** is the middle value when you arrange the data from smallest to largest:
>
> - If there is an **odd** number of values, the median is the value exactly in the middle.
>
> - If there is an **even** number of values, the median is the average of the two middle values.

Ex: Ratings:

$$1, \ 4, \ 2, \ 3, \ 5, \ 4, \ 5, \ 4, \ 4$$

What's the median?

Answer: Order the data:

$$1, 2, 3, 4, 4, 4, 5, 5$$

The middle value is 4, so the median is 4.

G DESCRIPTIVE STATISTICS: DISPERSION

When analyzing data, it's important not only to understand the **central tendency**—which refers to a typical value of a dataset (such as the mean, median, or mode)—but also to examine how much the data varies. This variation is called **dispersion**.

While measures of central tendency summarize the center of the data, measures of dispersion tell us how spread out the values are. To illustrate this, let's look at the test scores of two students:

- **Student A's scores:** 10, 50, 90

- **Student B's scores:** 45, 50, 55

Both students have the same mean score of 50. However, their scores are distributed very differently:

- **Student A's scores:** show a wide variation, ranging from 10 to 90.

- **Student B's scores:** are much more concentrated, between 45 and 55.

This example shows that even when two datasets have the same average, their distributions can be very different. Measures of dispersion, such as the range and interquartile range, help us better understand this variability.

> **Definition Range**
>
> The **range** is the difference between the maximum and minimum values in a dataset.
>
> $$\text{range} = \text{maximum} - \text{minimum}$$

Ex: Find the range for the following data: 1, 19, 10, 2, 18, 10, 5, 15, 10.

Answer: The minimum value is 1 and the maximum is 19.
So, the range is $19 - 1 = 18$.

> **Definition Quartile**
>
> **Quartiles** are values that divide an ordered dataset into four equal parts.

The median splits the data into two halves. The quartiles divide these halves again, giving us four equal parts.

> **Definition Interquartile Range**
>
> The **interquartile range (IQR)** is the difference between the upper quartile (Q_3) and the lower quartile (Q_1).
>
> $$\text{interquartile range} = Q_3 - Q_1$$

Ex: Find the quartiles and the interquartile range for the following data:
1, 19, 10, 2, 18, 10, 5, 15, 10

Answer:

- Order the data:

$$1, 2, 5, 10, 10, 10, 15, 18, 19$$

- The median (Q2) is 10.

- The lower half (before the median): $1, 2, 5, 10 \rightarrow Q_1 = \frac{2+5}{2} = 3.5$

- The upper half (after the median): $10, 15, 18, 19 \rightarrow Q_3 = \frac{15+18}{2} = 16.5$

- So, the interquartile range is $16.5 - 3.5 = 13$

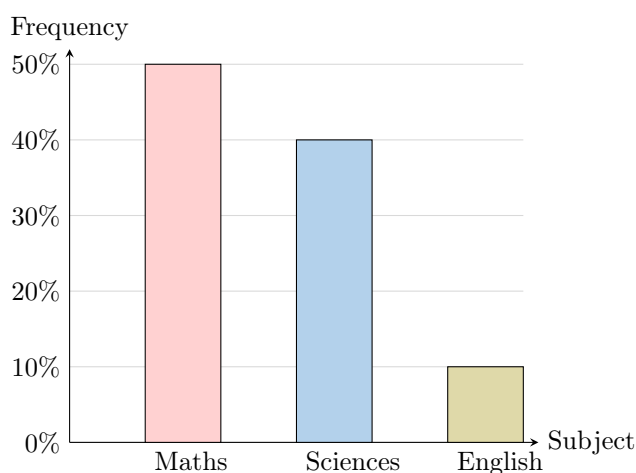# H ORGANIZING AND DISPLAYING DATA

> **Definition Bar Chart/Histogram**
>
> A **bar chart/histogram** shows data with bars:
>
> - Categories or values go on x-axis.
>
> - Frequencies go on y-axis.

Ex: Draw a bar chart for:

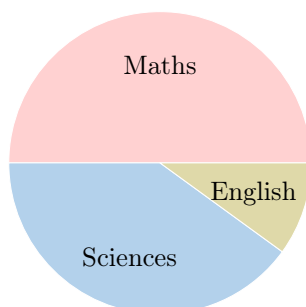| Subject	Relative frequency (%)
Maths	50%
Sciences	40%
English	10%

Answer:

> **Definition Pie Chart**
>
> A **pie chart** is a circle split into slices to show how data compares.

Ex: Draw the pie chart of the following data:

Subject	Frequency
Maths	15
Sciences	12
English	3
Total	30

Answer: Angles are :

- Maths : $\frac{15}{30} \times 360° = 180°$

- Sciences : $\frac{12}{30} \times 360° = 144°$

- English : $\frac{3}{30} \times 360° = 36°$

I INTERPRETING THE STATISTICS

Interpreting statistics means looking at the data to find out what it really tells us. We spot patterns, compare groups, and think about what the results mean in real life. The goal? Turn numbers into useful ideas for decisions or advice.

Ex: The girls' average score in math is 87 (B+), while the boys' average is 75 (C). Are girls better at math?

Answer: Yes, since $87 > 75$, on average, girls perform better than boys in math.