1. Which of the following is one-variable data?

A. Height and weight of students

B. Time and distance of a journey

C. Marks scored by students in a test

D. Temperature and rainfall

---

2. What is the purpose of a histogram?

A. To show relationships between two variables

B. To display categorical data only

C. To show the distribution of continuous data

D. To calculate the mean directly

---

3. What does the median of a data set represent?

A. The most common value

B. The average value

C. The middle value when ordered

D. The difference between largest and smallest

1. C — Marks scored by students in a test
2. C — To show the distribution of continuous data
3. C — The middle value when ordered

- **Mean (平均数 / 平均值)** – the arithmetic average of the data.
- **Median (中位数)** – the middle value when data are ordered.
- **Mode (众数)** – the most frequent value in the data.
- **Range (极差)** – the difference between the maximum and minimum values.
- **Standard Deviation (标准差)** – a measure of how spread out the data are.

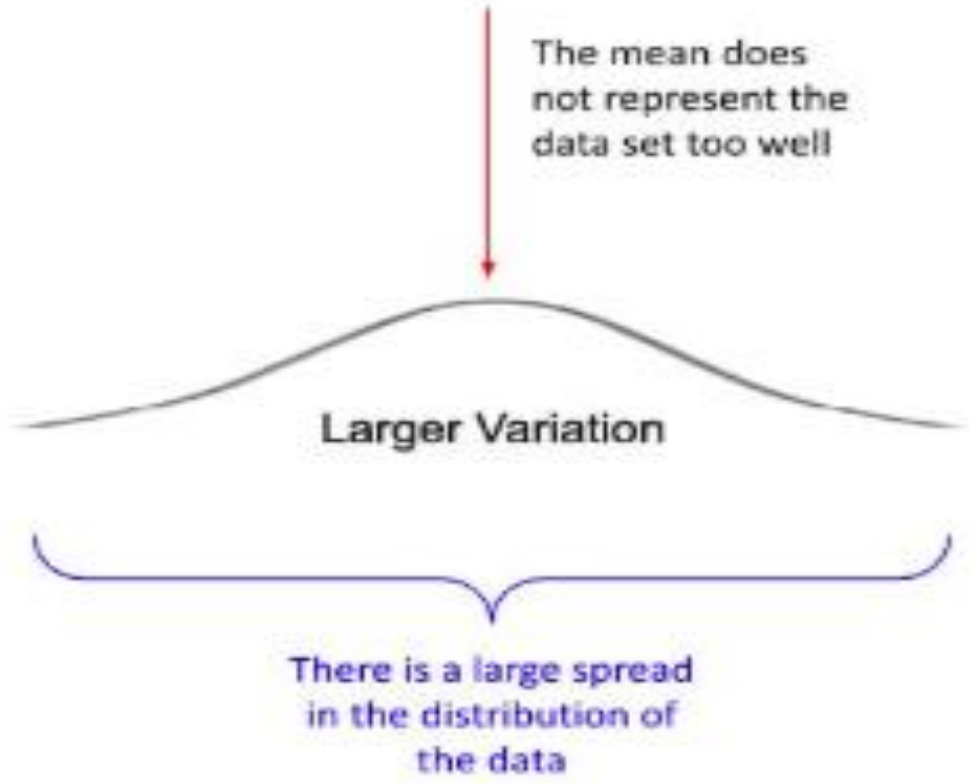
In **AP Statistics**, one-variable analysis focuses on describing a single quantitative variable.

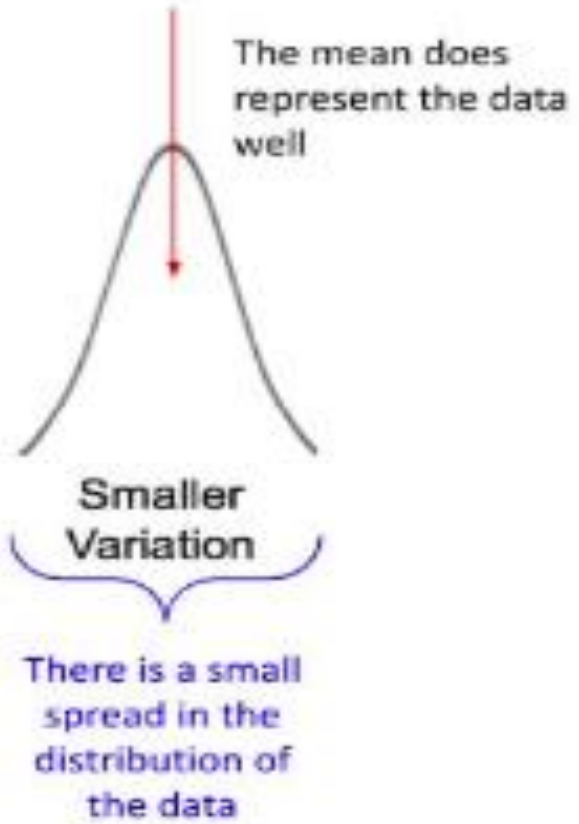
Students use graphs like dotplots, histograms, and boxplots to examine distributions.

They describe shape, center (mean/median), spread (IQR/standard deviation), and outliers. Percentiles, z-scores, and normal distributions help interpret relative standing.

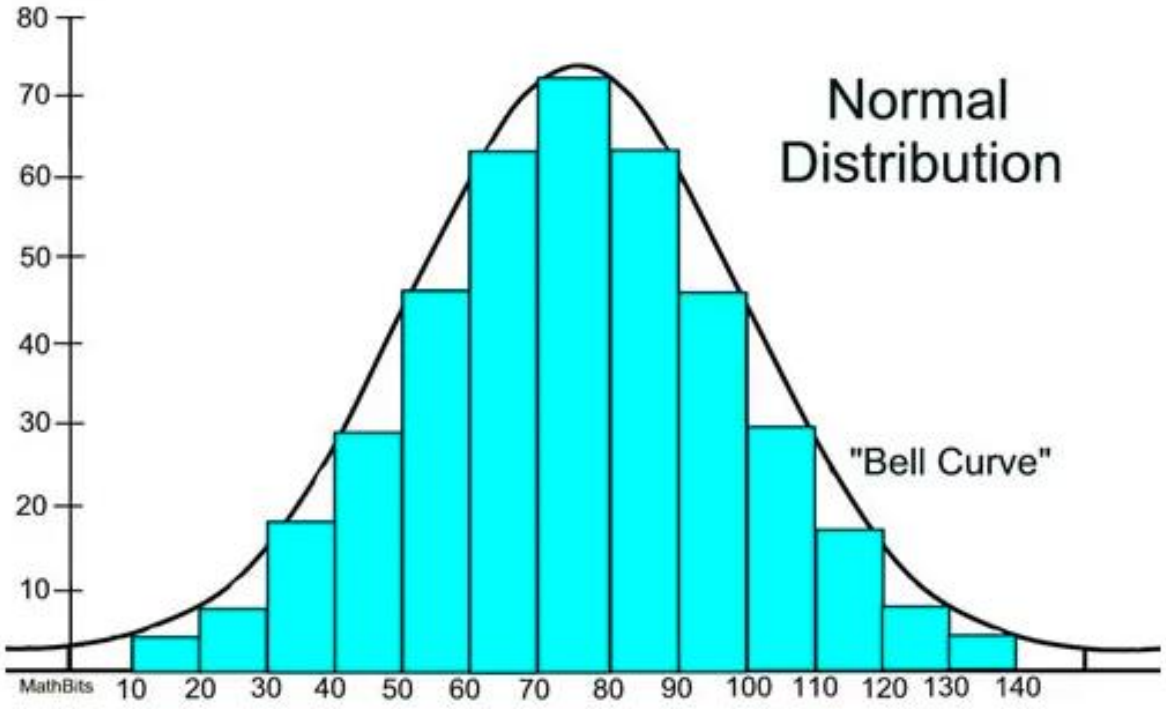
All conclusions must be explained clearly **in context** using correct statistical language.

VARIATION

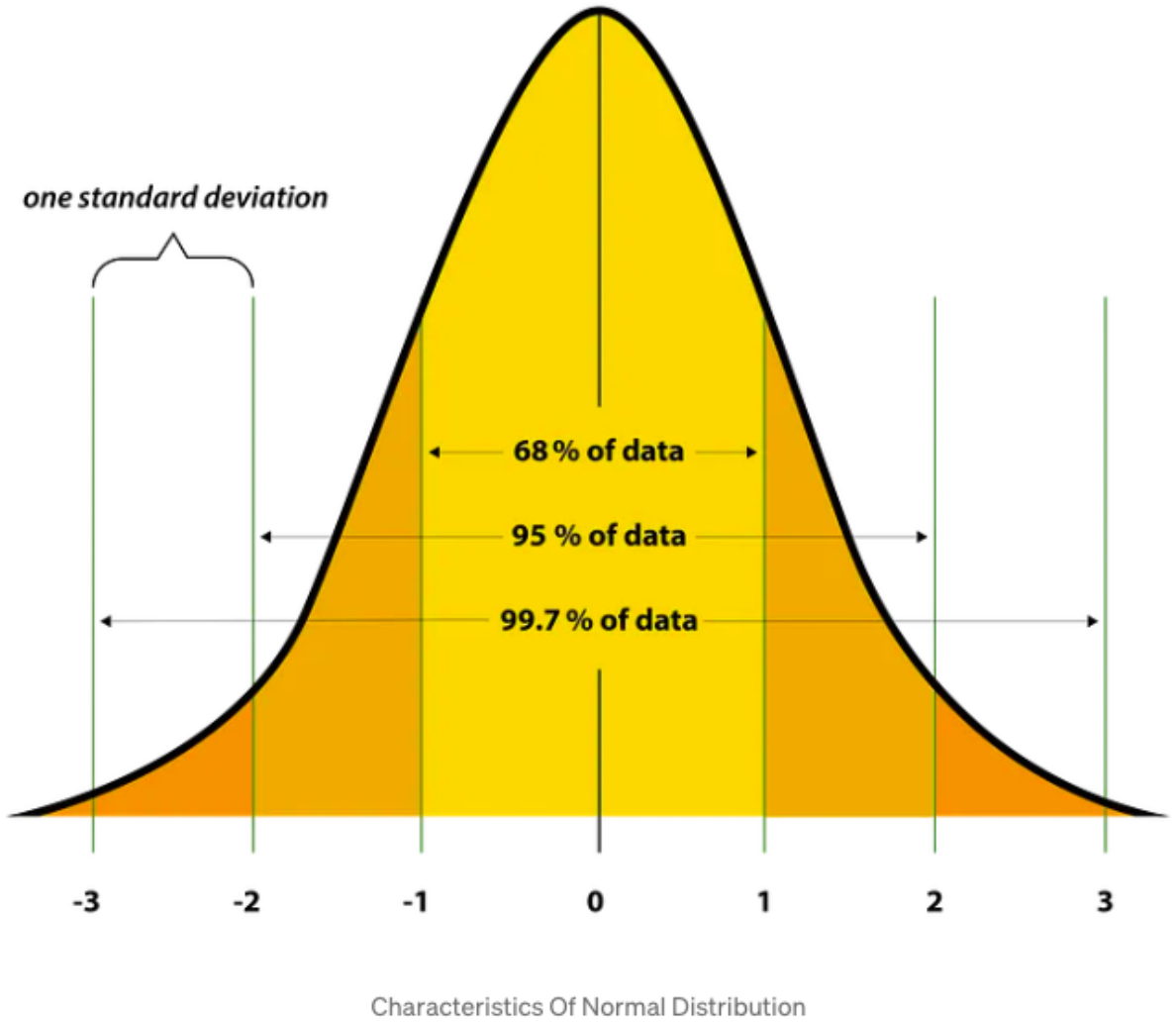

The mean does not represent the data set too well
Larger Variation
There is a large spread in the distribution of the data


The mean does represent the data well
Smaller Variation
There is a small spread in the distribution of the data

Normal Distribution



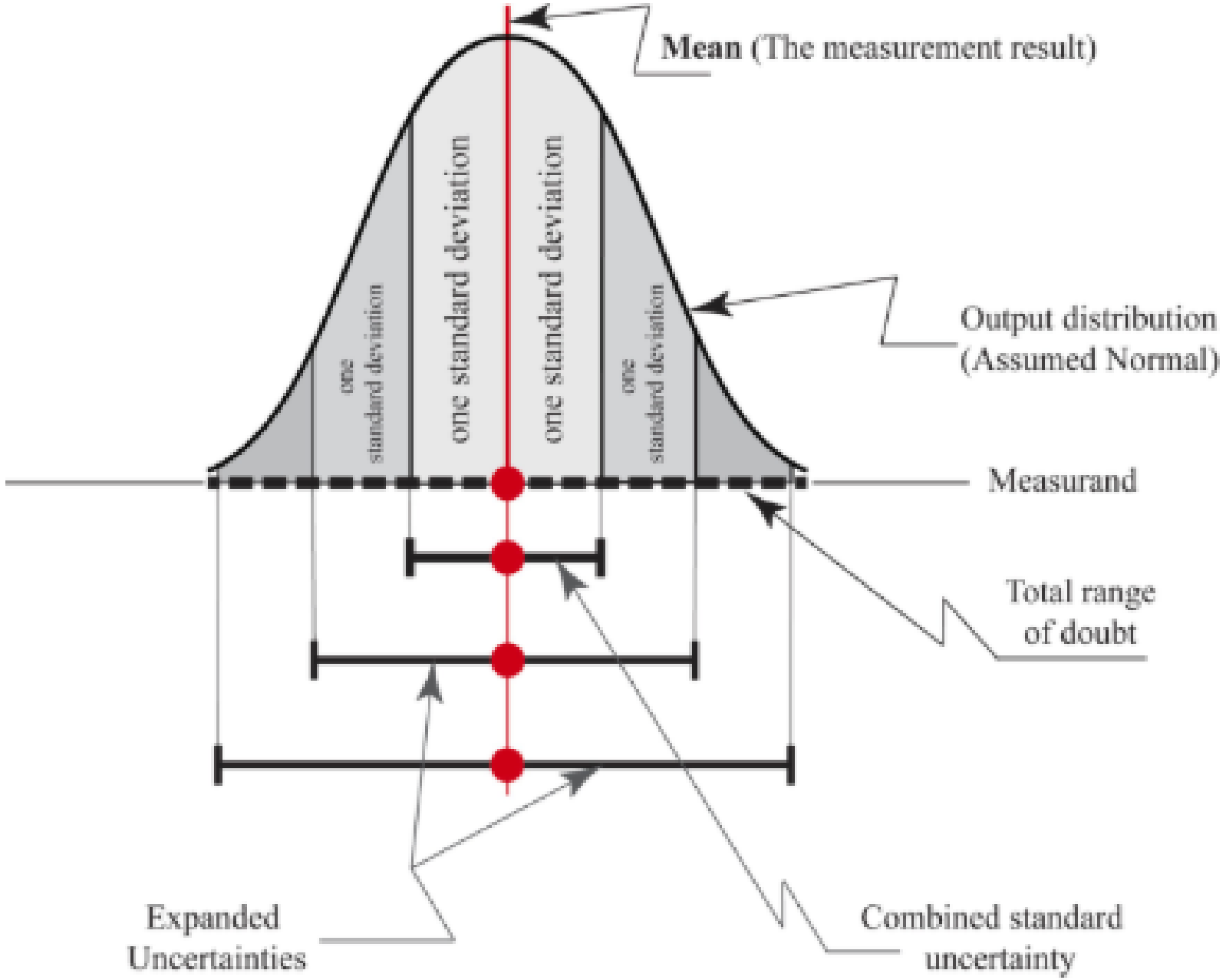
Normal DIstribution



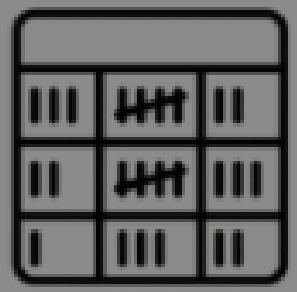
Characteristics Of Normal Distribution

About 68% of the data in a normal distribution falls within two standard deviations of the mean.

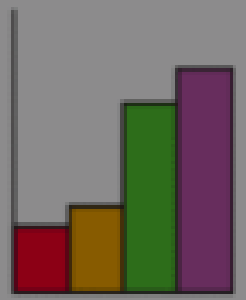
Uncertainty



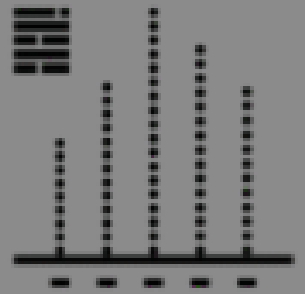
One-Variable Data: Distributions



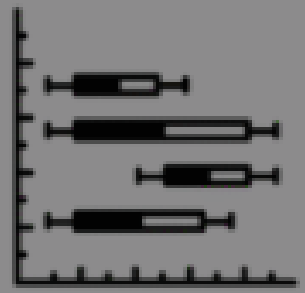
Frequency Distribution



Histograms



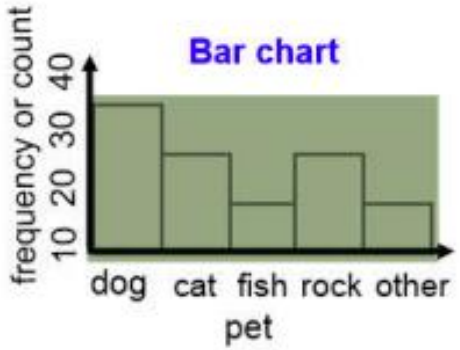
Dot Plots



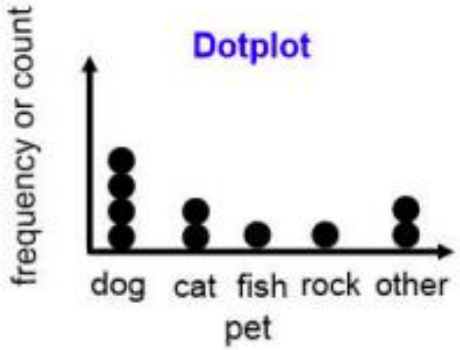
Box-and-Whisker Plots (Boxplots)

Categorical Data

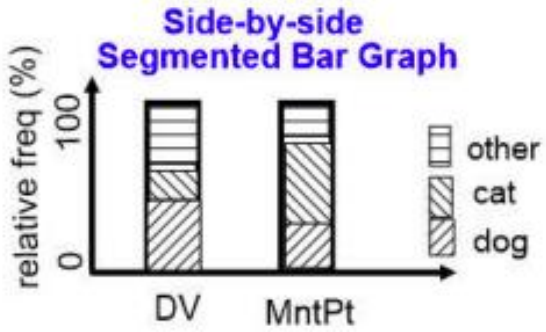
Bar chart



Dotplot

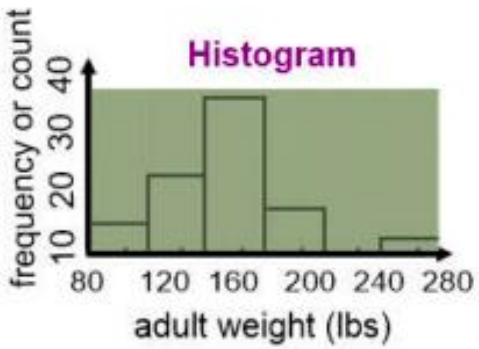


Side-by-side Segmented Bar Graph



Numerical Data

Histogram

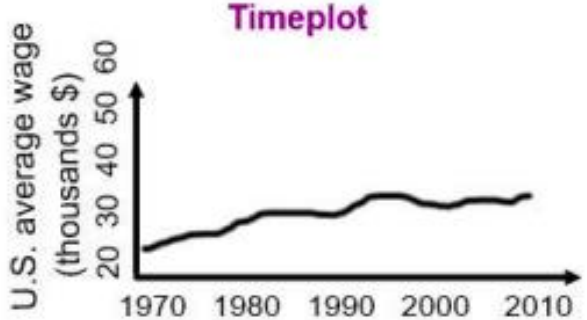


Stem/Leaf (Stemplot)

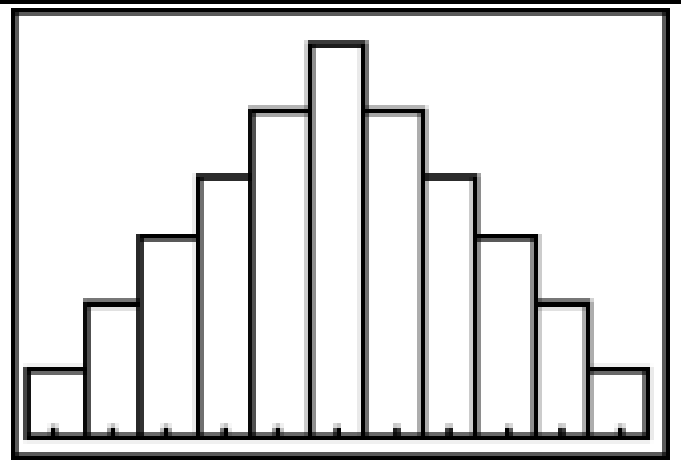
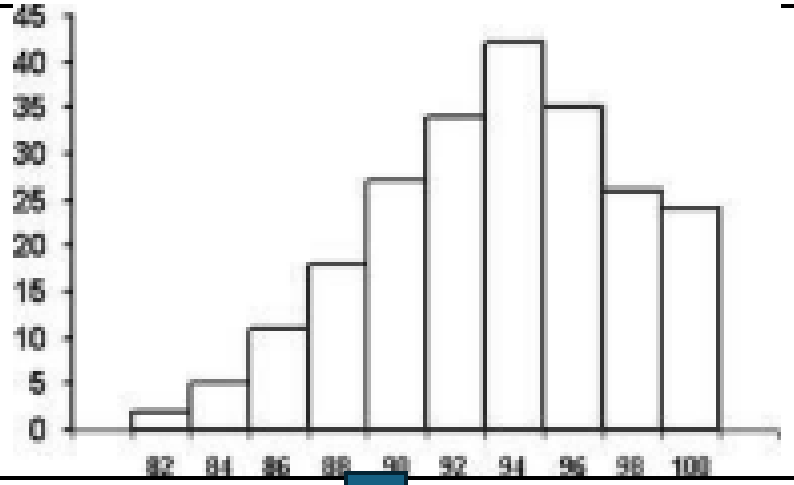
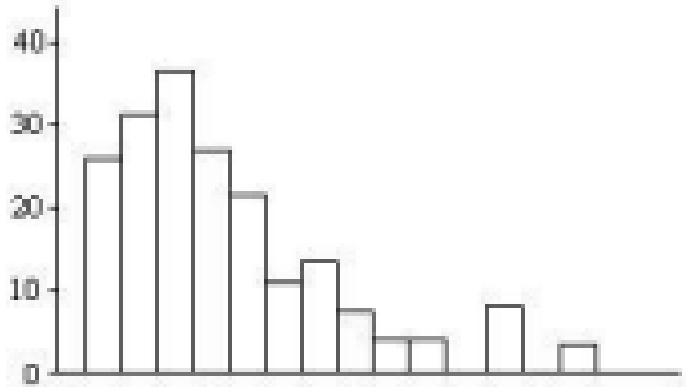
1	4	7		
2	4	6	2	
4	8	5	5	
5				
6	1	9	9	7

5 | 2 = 52

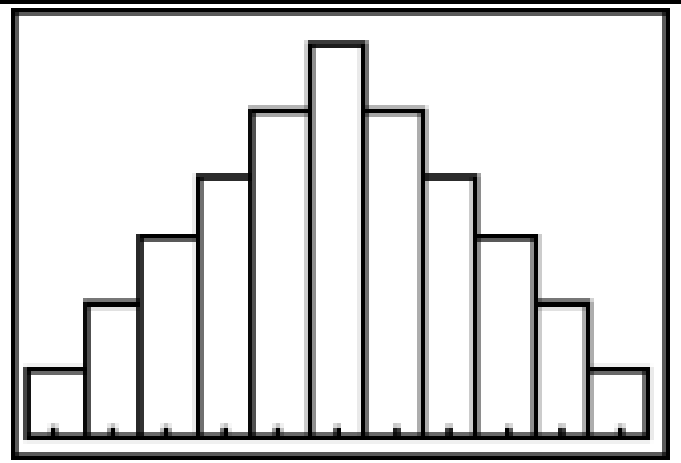
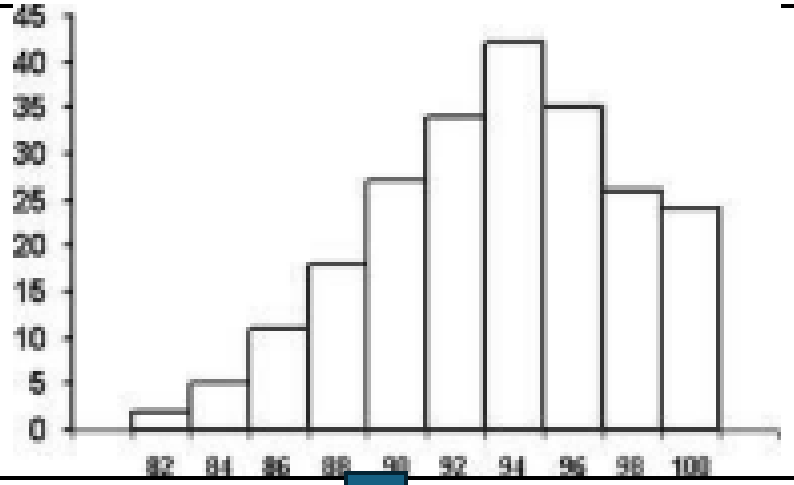
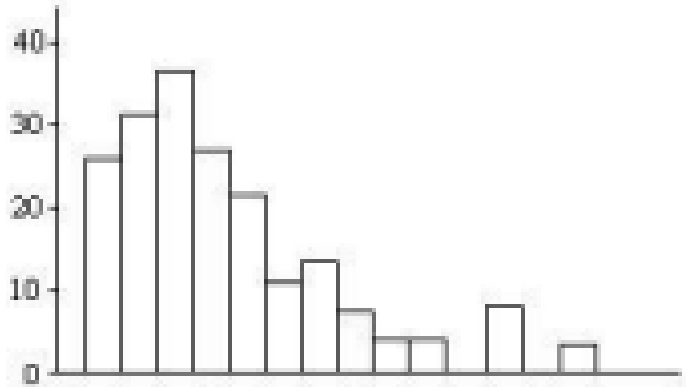
Timeplot

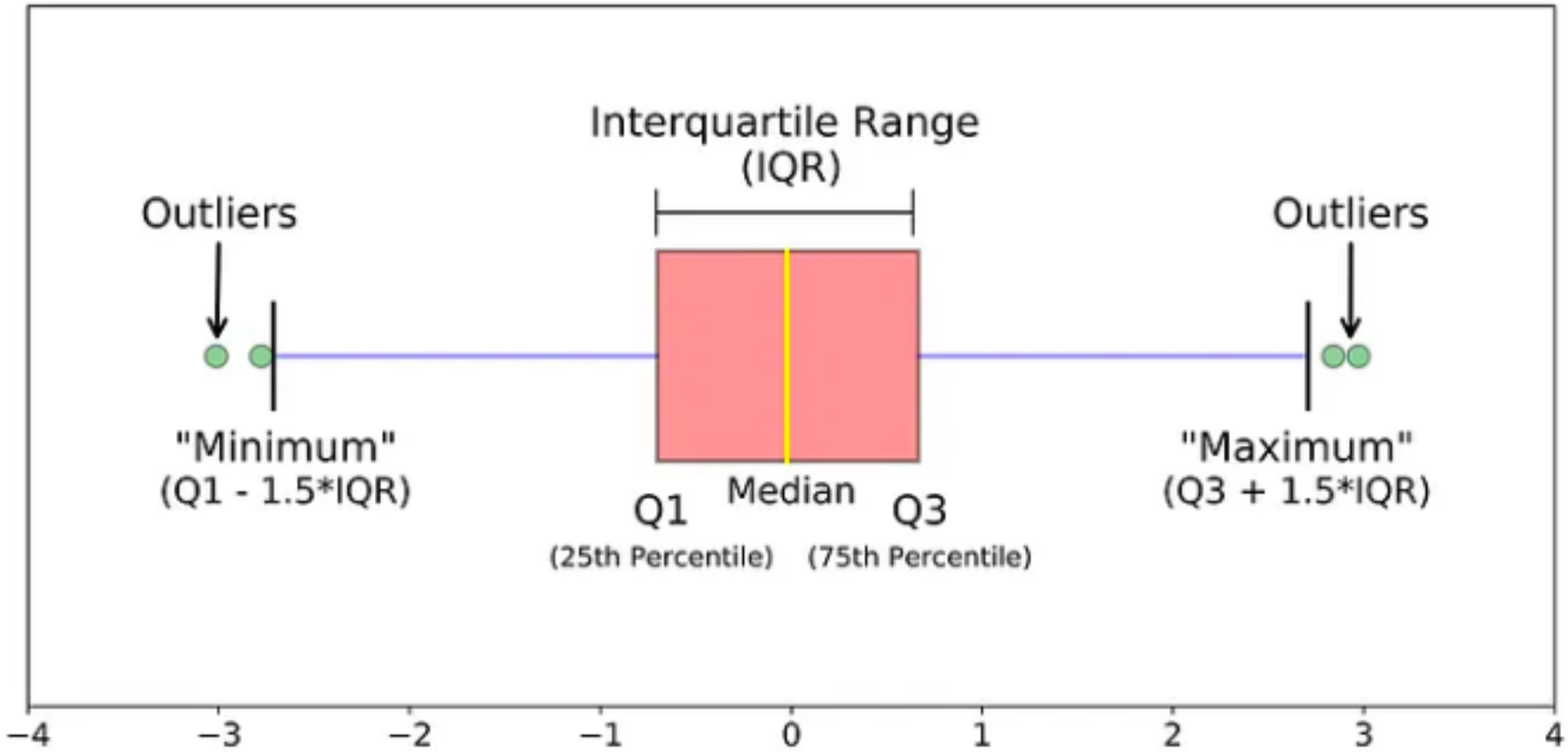


Describing the Shape of a Distribution

Symmetric	Skewed Left	Skewed Right
		
Mean ≈ Median	Mean < Median	Mean > Median

Describing the Shape of a Distribution

Symmetric	Skewed Left	Skewed Right
		
Mean Median	Mean Median	Mean Median

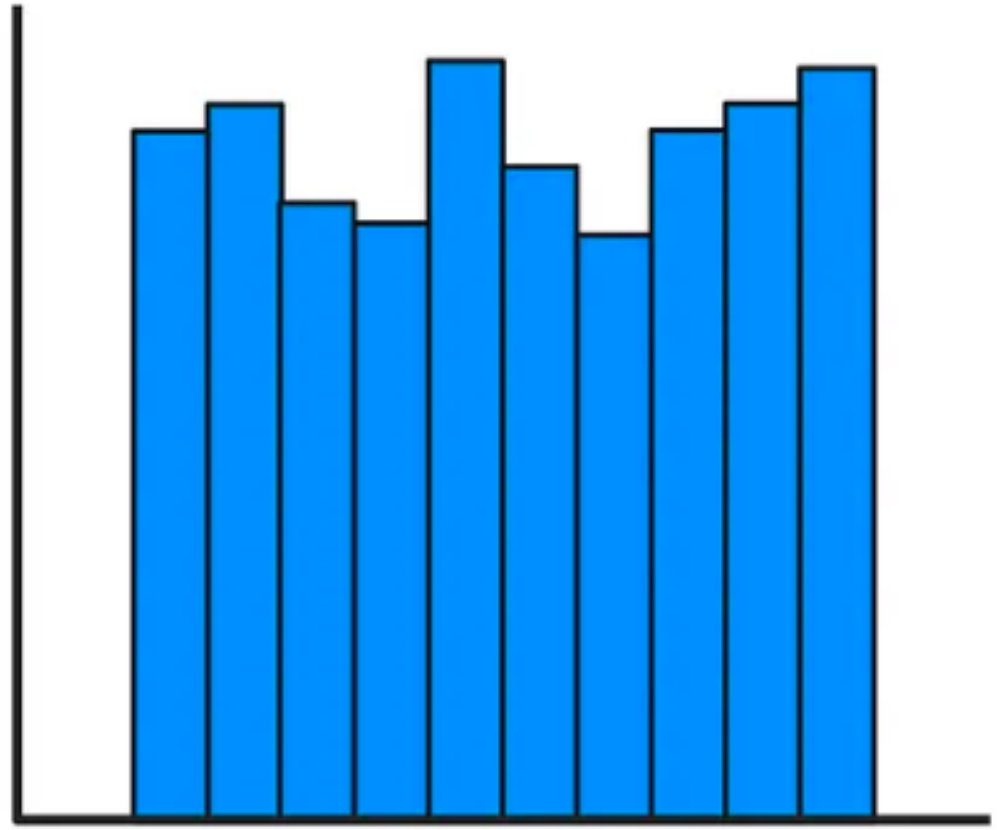


Box Plots

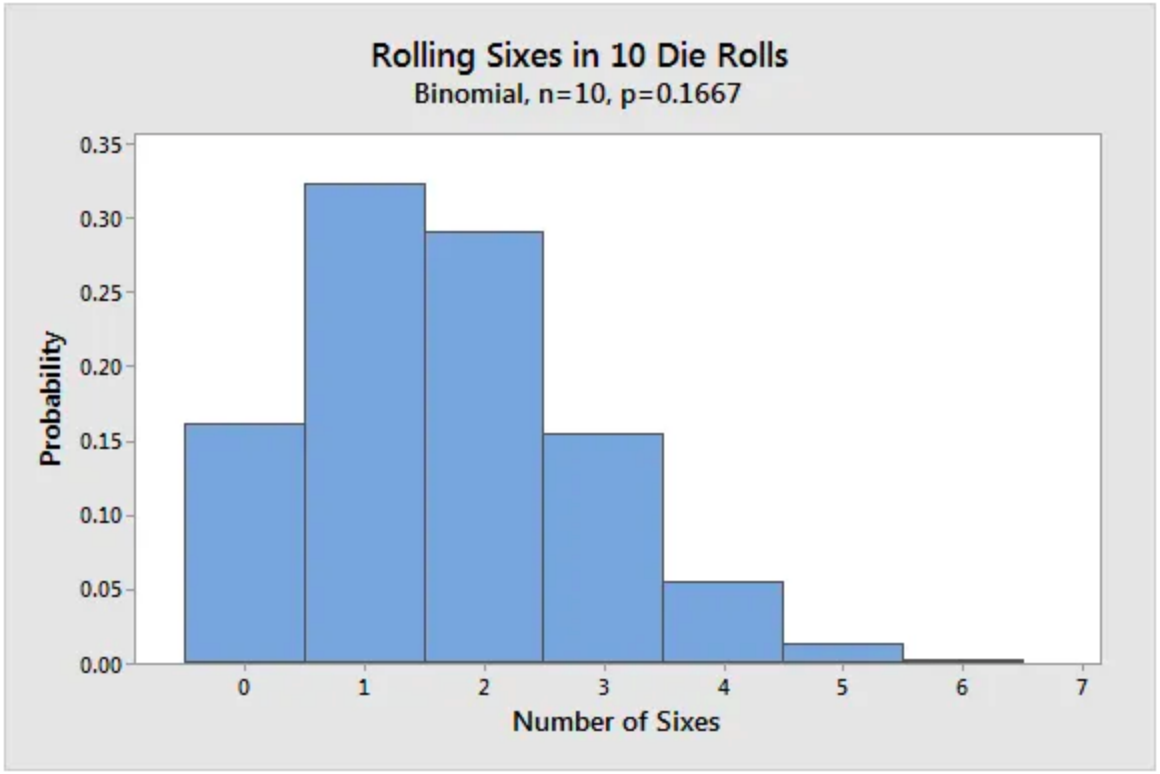
Box plots can display the spread and skewness of the data.

Histogram

Uniform distribution



Binomial distribution – rolling a six



A teacher recorded the scores of 10 students on a quiz:

72, 85, 90, 78, 85, 88, 91, 76, 84, 85

(a) Find the **mean** and **median** of the scores.

(b) Identify the **mode**.

(c) Calculate the **range**.

(d) Explain what the **standard deviation** tells you about these scores.

(a) Mean = 83.4, Median = 85

(b) Mode = 85

(c) Range = 19

(d) Standard deviation: moderate spread around the mean

STRETCH and CHALLENGE

1. Interpreting data distribution

The marks of 11 students are:
2, 3, 3, 4, 5, 5, 5, 6, 7, 9, 20

(a) Which measure of central tendency (mean, median, or mode) is most appropriate and why?
(b) What does this suggest about the data distribution?

---

2. Comparing measures of spread

Two classes have the following ranges of test scores:

- Class A: 45, 50, 52, 53, 55
- Class B: 20, 40, 55, 70, 90

(a) Calculate the range for each class.
(b) Which class shows greater consistency? Explain your reasoning.

---

3. Histogram reasoning

A histogram shows that most data values are clustered on the left with a long tail to the right.

(a) What type of skew is shown?
(b) How would the mean compare to the median, and why?

---

4. Real-life interpretation

A teacher records the number of hours students study per week. The data is positively skewed.

(a) Suggest a reason why the data might be skewed.
(b) Which measure (mean or median) gives a better "typical" value? Explain.

1. Interpreting data distribution

(a) **Median** is most appropriate because there is an outlier (20) that would distort the mean.
(b) The data is **positively skewed (right-skewed)** due to the high outlier.

---

2. Comparing measures of spread

(a)

- Class A range = $55 - 45 = 10$
- Class B range = $90 - 20 = 70$

(b) **Class A** shows greater consistency because it has a much smaller range, meaning scores are more clustered.

---

3. Histogram reasoning

(a) The distribution is **positively skewed (right-skewed)**.
(b) The **mean is greater than the median** because the long right tail pulls the mean upwards.

---

4. Real-life interpretation

(a) Possible reason: Some students study a lot more than others, creating a few high values (outliers).
(b) The **median** is better because it is not affected by extreme values and gives a more typical study time.

Descriptive Statistics – One Variable Analysis

1. Data — 数据
2. Variable — 变量
3. Frequency — 频数
4. Mean — 平均数
5. Median — 中位数
6. Mode — 众数
7. Range — 极差
8. Outlier — 异常值
9. Distribution — 分布
10. Histogram — 直方图

1. One-variable data involves **only one characteristic or variable** being studied.
2. Data is often organised using a **frequency table** to make patterns clearer.
3. The **mean** is the average of all values and is sensitive to extreme values.
4. The **median** is the middle value when data is ordered and is resistant to outliers.
5. The **mode** is the most frequently occurring value in a data set.
6. The **range** measures spread using: max value − min value.
7. An **outlier** is a value that is much higher or lower than the rest of the data.
8. Data distributions can be **symmetrical or skewed** (positive or negative skew).
9. A **histogram** is used to represent continuous data distributions.
10. Statistical results must always be **interpreted in context**, not just calculated.

The times (in minutes) taken by 9 students to complete a test are:

$$12,\ 15,\ 14,\ 18,\ 20,\ 22,\ 22,\ 25,\ 40$$

(a)

Calculate the:

- mean
- median
- mode
- range

(b)

A teacher says the class performance is "typical and consistent."

Do you agree? Justify your answer using statistical evidence.

(c)

Identify whether the data is **symmetrical, positively skewed, or negatively skewed**. Explain your reasoning.

(d)

Explain which measure of central tendency (mean or median) is more appropriate for representing this data set, and why.

(a)

- Mean = **20.89**
- Median = **20**
- Mode = **22**
- Range = **28**

(b)

Disagree. Data is not fully consistent due to the **outlier (40)** and a **large range (28)**.

(c)

Positively skewed because of the high outlier (40) and a right tail.

(d)

Median is better because it is not affected by the outlier and better represents the typical value.

TRUE or FALSE

- One-variable data involves only one characteristic being measured.

TRUE

TRUE or FALSE

- The mean is always the best measure of central tendency.

FALSE

TRUE or FALSE

- The median is resistant to extreme values (outliers).

TRUE