

MDS26112 — UNIVERSITY OF BISHA

Biostatistics

Lecture 3

Measures of Central Tendency

Descriptive Statistics: Summarizing Data with a Single Value



Dr. Khalid Aboalbasher

03

Learning Outcomes

CLO 1.1

Define key terminology: **descriptive measures**, **central tendency**, **mean**, **median**, **mode**, **population vs. sample**, and **cumulative frequency**.

CLO 1.3

Explain how measures of central tendency help interpret medical data and support clinical decisions.

CLO 2.2

Calculate the **mean**, **median**, and **mode** for both untabulated and tabulated (grouped) data sets.

CLO 2.2

Determine which measure of central tendency is most appropriate for different data types and distributions.

What Are Descriptive Measures?

After organizing and presenting data (Lecture 2), the next step is **summarization** — calculating single values that capture essential information about the dataset. These values are called **descriptive measures**.



Measures of Central Tendency

Describe the **center** or **typical value** of a dataset. Include the mean, median, and mode.



Measures of Dispersion

Describe **how spread out** the data is. Include range, variance, and standard deviation.

The Three Measures of Central Tendency



Mean ($\bar{x}$)

The **arithmetic average**. Calculated by summing all values and dividing by n . Best for symmetric, continuous data.

Quantitative only

Affected by outliers



Median (M)

The **middle value** when data is sorted. Robust against outliers. Ideal for skewed distributions.

Quantitative only

Outlier-resistant



Mode (M_o)

The **most frequently occurring** value. Works for both quantitative and qualitative data. Can have multiple modes.

Quantitative & qualitative

Identifies common patterns

Why Central Tendency Matters in Medicine



Data summarization

Condense hundreds of lab results into a single typical value (e.g., average blood pressure = 130.8 mmHg).



Comparative analysis

Compare patient groups — does Treatment A yield lower average blood sugar than Treatment B?



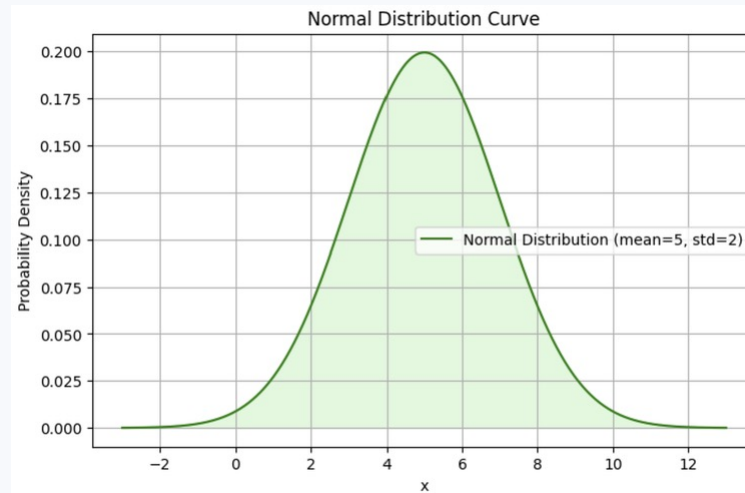
Foundation for advanced analysis

Mean and median are used in t-tests, ANOVA, and regression — the backbone of medical research.



Clinical decision-making

Reference ranges are built around central values; understanding them helps interpret whether a patient's result is normal.





ACTIVITY

Think-Pair-Share

CLO 1.3

A hospital reports that the average salary of its staff is 25,000 SAR per month. However, most employees earn between 8,000 and 12,000 SAR.

How is this possible?

1

THINK

30 sec

What single data characteristic could cause the mean to be much higher than most values?

2

PAIR

1 min

Discuss with your neighbor — which measure of central tendency would better represent the "typical" salary?

3

SHARE

—

Class discussion

01

The Mean

The Arithmetic Average

The most widely used measure of central tendency in medical research.

Arithmetic Mean — Concept & Intuition

The arithmetic mean is the **sum of all data values divided by the total number of observations**. It represents the "balance point" of the data.



Imagine placing weights on a seesaw at each data value. The mean is the point where the seesaw balances perfectly.

- **Most important and most frequently used** measure of central tendency.
- Uses **every value** in the dataset — no information is lost.
- Denoted as $\bar{x}$ (x-bar) for a **sample**, and μ (mu) for a **population**.

Mean for Untabulated Data

$$\text{Mean} = \bar{x} = (x_1 + x_2 + x_3 + \dots + x_n) / n$$

Key Notation

$\bar{x}$ (x-bar) Sample mean

μ (mu) Population mean

n Number of observations (sample size)

$\sum x_i$ Sum of all values



Important

The mean can **ONLY** be calculated for **quantitative data**. It cannot be computed for qualitative (nominal or ordinal) categories.

Sum of all observations

$$\frac{\text{Sum of all observations}}{n}$$

Step-by-Step: Calculating the Mean

A doctor measures the systolic blood pressure of 8 patients. The readings are:

120, 130, 125, 135, 140, 132, 127, 138 mmHg

1 **Sum all values:**

$$120 + 130 + 125 + 135 + 140 + 132 + 127 + 138 = \mathbf{1047}$$

2 **Divide by $n = 8$**

3 **Calculate:**

$$\text{Mean} = 1047 / 8 = 130.875 \rightarrow \text{round to } \mathbf{130.9} \text{ mmHg}$$

Result

$$\bar{x} = \mathbf{130.9 \text{ mmHg}}$$

Average systolic BP for 8 patients



This value can be compared against the normal reference range (120 mmHg) to assess whether this patient group tends to have **elevated blood pressure**.

Mean for Grouped (Tabulated) Data

When data is organized in a frequency distribution table, we use **class midpoints** as representatives of all values in each class.

$$\bar{x} = (f_1x_1 + f_2x_2 + \dots + f_kx_k) / n$$

Where

x_i Class midpoint (Midpoint = (Lower + Upper) / 2)

f_i Frequency of the class

n Total frequency (sample size)

k Number of classes

$$\text{Mean} = \frac{\sum f_i x_i}{\sum f_i}$$

Steps

- 1 Calculate the midpoint (x_i) for each class
- 2 Multiply each midpoint by its frequency ($f_i \times x_i$)
- 3 Sum all products
- 4 Divide by total frequency n

Step-by-Step: Mean from Grouped Data

A nurse records the **length of hospital stay (days)** for 100 patients.

FREQUENCY TABLE

Class	f_i
4–8	6
8–12	19
12–16	25
16–20	30
20–24	20
Total	100

STEP 2 — EXTENDED TABLE

Class	f_i	Midpoint (x_i)	$f_i \times x_i$
4–8	6	6	36
8–12	19	10	190
12–16	25	14	350
16–20	30	18	540
20–24	20	22	440
Total	100	—	1,556

STEP 1 — MIDPOINTS

$$(4+8)/2 = 6$$

$$(8+12)/2 = 10$$

$$(12+16)/2 = 14$$

$$(16+20)/2 = 18$$

$$(20+24)/2 = 22$$

$$\text{Mean} = \frac{\sum(f \times x)}{\sum f} = \frac{1556}{100} = 15.56$$

CLO 2.2

CLO 1.3

Solution: Mean = 15.56 Days

Step 3 — Apply the formula

$$\bar{x} = \Sigma (f_i x_i) / n = 1,556 / 100 = 15.56 \text{ days}$$

Result

Average hospital stay:

15.56

days per patient



Clinical Interpretation

This average helps hospital administrators plan bed capacity and staffing. If the target average stay is 14 days, then 15.56 days suggests potential **inefficiencies in discharge planning** that warrant investigation.

Sample Mean vs. Population Mean



Sample Mean

$$\bar{x}$$

(x-bar)

Data source: A subset (sample) of the population

$$\bar{x} = \sum x_i / n$$

Use in: Most medical research (we rarely have data on ALL patients)



Population Mean

$$\mu$$

(Greek letter mu)

Data source: The entire population

$$\mu = \sum x_i / N$$

Use in: Census data or when the entire population is available

◆ Quick Quiz: Test Your Understanding

? Q: The hemoglobin levels (g/dL) of 6 patients are:

11.2, 14.5, 12.8, 13.1, 15.0, 12.4

Calculate the mean.

A 13.17 g/dL



B 13.00 g/dL

C 12.80 g/dL

D 14.50 g/dL

◆ Worked Solution

Step 1: Sum all values

$$11.2 + 14.5 + 12.8 + 13.1 + 15.0 + 12.4 = \mathbf{79.0}$$

Step 2: Divide by n

$$\text{Mean} = 79.0 / 6 = \mathbf{13.17 \text{ g/dL}}$$

CLO 2.2

The Median

The Middle Value



The Median — Why We Need It

The mean is **sensitive to extreme values** (outliers). When outliers are present, the mean can be **misleading**.



Definition

The median is the **middle value** in a dataset after arranging all observations in **ascending** (or descending) order.



When to Use the Median



Skewed distributions
Income data, hospital stay lengths



Datasets with extreme outliers
Values far from the center



Ordinal data
Where the mean is not meaningful

Median for Untabulated Data: When n is Odd



When n is odd, the median is the value at **position $(n + 1) / 2$** in the sorted array.

3	5	7	9	11	13	15
pos 1	pos 2	pos 3	pos 4	pos 5	pos 6	pos 7

$$\text{Position} = (7 + 1) / 2 = 4\text{th value} = 9$$

Key steps

- 1 Arrange data in **ascending order**
- 2 Calculate position = $(n + 1) / 2$
- 3 The value at that position is the **median**

Step-by-Step: Median (Odd n) — Blood Glucose

+ Clinical scenario: A healthcare provider records the fasting blood glucose levels (mg/dL) of 9 patients: 145, 92, 110, 155, 120, 130, 115, 150, 85

- 1 Arrange in ascending order:
85 92 110 115 **120** 130 145 150 155
- 2 $n = 9$ (odd). Position = $(9 + 1) / 2 = 5$ th value.
- 3 The 5th value in the sorted list is **120**.

Median = 120 mg/dL



Compare with the Mean

Mean of this data:

$$(85+92+110+115+120+130+145+150+155) / 9$$

$$\text{Mean} = \frac{1102}{9} \text{ mg/dL}$$

The mean (122.4) is **slightly higher** than the median (120) because the high value 155 pulls the mean up.

Median for Untabulated Data: When n is Even



When n is even, the median is the **MEAN** of the two middle values at positions $n/2$ and $(n/2) + 1$.

$$\text{Median} = (\text{Value at position } n/2 + \text{Value at position } n/2 + 1) / 2$$

2	4	5	7	9	11	14	16
pos 1	pos 2	pos 3	pos 4 = $n/2$	pos 5 = $n/2+1$	pos 6	pos 7	pos 8

Positions: $n/2 = 4$, $n/2 + 1 = 5 \rightarrow \text{Median} = (7 + 9) / 2 = 8$

Step-by-Step: Median (Even n) — Clinic Visits

✚ **Clinical scenario:** A doctor records the number of clinic visits made by 12 patients in the past year: 7, 5, 4, 11, 9, 6, 4, 10, 12, 2, 8, 15

- 1 Sort ascending:
2 4 4 5 6 **7** **8** 9 10 11 12 15
- 2 $n = 12$ (even). Position 1 = $12/2 = 6$ th value = **7**. Position 2 = $12/2 + 1 = 7$ th value = **8**.
- 3 Median = $(7 + 8) / 2 = \mathbf{7.5}$

Median = 7.5 clinic visits/year



Interpretation



The median of **7.5** means **half the patients** visited the clinic **more than 7.5 times** and **half visited fewer**.

- ◆ Notice that **7.5** is **not** an actual data value — it is the *average* of the two middle values. This is expected when n is even.

Median for Grouped (Tabulated) Data

When data is in a frequency distribution table, we use **cumulative frequencies** to locate the median class.

$$\text{Median} = A + \frac{\left[\frac{\sum f_i}{2}\right] - \sum f_1}{f} \times H$$

Where:

A Lower limit of the median class

$\sum f_i$ Total frequency

$\sum f_1$ Cumulative freq. of class
BEFORE median class

f Frequency of the median class

H Class width



Steps

- 1 Calculate median rank
median rank = $\sum f_i / 2$
- 2 Build cumulative frequency column
Add frequencies in ascending order
- 3 Find the median class
First class where $\sum f_1 \geq$ median rank
- 4 Apply the formula

Step-by-Step: Median (Grouped) — ED Wait Times

+ Clinical scenario: A healthcare provider records the number of hours 500 patients spent in the emergency department.

Step 1: Median rank = $\sum f_i / 2 = 500/2 = 250$

Step 2: Build cumulative frequency table

Class (hours)	f_i	Less than	CF
2–4	21	<4	21
4–6	35	<6	56
6–8	52	<8	108
8–10	73	<10	181
10–12	91	<12	272
12–14	84	<14	356
14–16	69	<16	425
16–18	45	<18	470
18–20	30	<20	500
Total	500	—	—

Step 3: Identify parameters

Median class = **10–12**

First class where $\sum f^1$ (272) ≥ 250

Formula values:

A (lower limit of median class)	10
$\sum f^1$	181
	91
H (class width)	2

Solution: Median = 11.52 Hours

4 Apply the formula

$$\text{Median} = A + \frac{\left[\frac{\sum f_i}{2}\right] - \sum f_1}{f} \times H$$

$$= 10 + \frac{250-181}{91} \times 2$$

$$= 10 + [69 / 91] \times 2$$

$$= 10 + 0.758 \times 2$$

$$= 10 + 1.516$$

$$= 11.516 \text{ hours} \approx 11.52 \text{ hours}$$



The median ED wait time is

11.52 hours



Clinical Interpretation

This means **half** of the 500 patients waited **less than 11.52 hours**, and half waited **more**.

⚠ If the hospital target is **under 10 hours**, this median reveals that **performance improvement is needed**.

The Mode — The Most Common Value

Identifying the most frequent observation in a dataset.

03

The Mode — Concept & Types

① Definition

The mode is the value that occurs most frequently in a dataset. It can be calculated for both **quantitative** and **qualitative** data.

★ Key Difference

Unlike the mean and median, the mode works with **Categorical data** — making it the only central tendency measure usable for nominal data.

Unimodal

One mode

e.g., most common blood type

Bimodal

Two modes

e.g., two equally common diagnoses

Multimodal

More than two modes

Multiple frequent values

Non-modal

None

All values occur equally



When to Use

Categorical/nominal data (most common blood type, most frequent diagnosis) · Identifying the typical outcome in a dataset

CLO 1.1

Mode for Untabulated Data: Examples

◆ Example 1 — Unimodal

3, 5, 7, 9, 2, 8, 10, 9, 10, 11, 9, 9,
7, 10, 9, 8, 7, 6, 5

Counts: 9 → 5×, 10 → 3×, 7 → 3×, 5 → 2×, others → 1×

Result Mode = 9 (unimodal)

◆ Example 2 — Bimodal

9, 7, 10, 9, 8, 7, 6, 5, 7, 9, 10, 8

Counts: 7 → 3×, 9 → 3×, 8 & 10 → 2× each

Result Modes = 7 and 9 (bimodal)



Clinical Example

In a ward of 30 patients, the most common blood type is **Type O** (appearing 14 times).

MODE = TYPE O This tells the blood bank which type to stock most.

Mode for Grouped (Tabulated) Data

CLO 2.2

$$\text{Mode} = A + \frac{f - f_1}{2f - (f_1 + f_2)} \times H$$

A — Lower limit of the modal class
(class with highest frequency)

f — Frequency of the modal class

f₁ — Frequency of the class BEFORE
the modal class

f₂ — Frequency of the class AFTER
the modal class

H — Class width

◆ Key note

The modal class is simply the class with the **HIGHEST frequency** — unlike the median class which requires cumulative frequencies.

Steps

- 1 Identify modal class (highest f)
- 2 Read A, f, f₁, f₂, and H
- 3 Apply the formula

Step-by-Step: Mode (Grouped) — Hypertension Ages

CLO 2.2

CLO 1.3

- ✚ A hospital analyzes the age distribution of patients diagnosed with hypertension over a month.

Data Table

Classes	f_i
60–64	5
64–68	18
68–72 ★	42
72–76	27
76–80	8
Total	100

Step 1

Modal class = **68–72** (highest $f = 42$)

Step 2

$A = 68$, $f = 42$, $f_1 = 18$, $f_2 = 27$, $H = 4$

Step 3 — Apply Formula

$$\begin{aligned}\text{Mode} &= 68 + [42 / (2 \times 42 - 18 - 27)] \times 4 \\ &= 68 + [42 / (84 - 45)] \times 4 \\ &= 68 + [42 / 39] \times 4 \\ &= 68 + 1.077 \times 4 \\ &= 68 + 4.308 \\ &= 72.31\end{aligned}$$

RESULT

The modal age for hypertension diagnosis is approximately **72.3 years**.

◆ Clinical Note

This suggests that hypertension is most commonly diagnosed around age 72 in this hospital population.

Putting It All Together

Comparing the three measures and choosing the right one for clinical data.

04

Mean vs. Median vs. Mode — Comprehensive Comparison

CLO 1.1

Feature	Mean	Median	Mode
Definition	Arithmetic average	Middle value	Most frequent value
Data Type	Quantitative only	Quantitative only	Quantitative AND qualitative
Outlier Sensitivity	HIGH (affected)	LOW (robust)	LOW
Best For	Symmetric distributions	Skewed distributions	Categorical data, common patterns
Uses All Values	Yes	No (position-based)	No (frequency-based)
Always Exists	Yes	Yes	No (can be no mode)
Further Analysis	Essential (t-tests, ANOVA)	Limited	Limited

- ◆ In medical research, the **mean** is the most commonly reported measure, but the **median** is preferred when data is skewed.

◆ Activity: Which Measure Should the Doctor Use?

CLO 2.2

1 Blood Type Study

A researcher wants to report the typical blood type in a sample of 200 patients.

2 Hospital Stay Comparison

A hospital wants to compare average length of stay across 3 departments. The data has some patients who stayed over 90 days.

3 Cholesterol Trial

A clinical trial measures cholesterol levels (normally distributed) for a treatment and control group.

4 Pain Score Survey

A survey asks patients to rate their pain on a scale of 1–10.

👤 Discuss with your neighbor (2 minutes)

For each scenario, identify the most appropriate measure and explain why.

1 → Mode (nominal)

2 → Median (outliers)

3 → Mean (symmetric)

4 → Median (ordinal)



Practice Questions

Q1

The following are the ages of 10 patients: 35, 42, 28, 55, 31, 48, 33, 61, 29, 38. Calculate the mean, median, and mode. Which measure best represents the center? Justify your answer.

Q3

Explain why the median is preferred over the mean for reporting household income data in health surveys.

Q2

Using the grouped data from the hospital stay example (Lecture 2, slide 20), calculate the median length of stay using the cumulative frequency method.

Q4

For the hypertension age data (classes 60–64 to 76–80, $f = 5, 18, 42, 27, 8$), verify the mode calculation by showing all steps. Is the data unimodal?

▼ These questions align with **CLOs 1.1, 1.3, and 2.2.**