Lecture 1
Introduction to Biostatistics

Prof. Dr Saad Abed Madhi

Introduction

- **Biostatistics** = application of statistical methods to biological and medical research.

- In **radiology**, biostatistics helps in:

- Analyzing imaging study data (CT, MRI, X-ray).

- Understanding patient demographics, disease prevalence.

- Evaluating diagnostic test performance.

- Supporting evidence-based decisions.

Important Terms in Biostatistics

- **Population**: Entire group under study (e.g., all patients with lung nodules).
- **Sample**: Subset of the population (e.g., 100 patients in a CT study).
- **Variable**: A characteristic measured (e.g., tumor size, age, CT density).
- **Parameter**: A numerical summary of the population.
- **Statistic**: A numerical summary of the sample.
- **Bias**: Systematic error in data collection or analysis.
- **Confounding**: A hidden factor affecting the relationship between variables.

Types of Data

- **Qualitative (Categorical Data)**

- *Nominal*: Categories without order (e.g., gender, blood type, scan type).

- *Ordinal*: Categories with order (e.g., tumor stage I–IV, pain scale).

- **Quantitative (Numerical Data)**

- *Discrete*: Countable (e.g., number of lesions).

- *Continuous*: Measured (e.g., lesion size in cm, Hounsfield units).

Data Presentation

- **Text/Tabular presentation**: Summary tables, frequency distributions.

- **Graphical presentation**: Histograms, pie charts, scatter plots, etc.

Frequency Table Presentation

- Summarizes data into intervals with counts.
- Example (Tumor Size in CT scans):

Tumor Size (cm)	Frequency
0–2	15
2–4	25
4–6	10
6–8	5

Two-Way Cross Classification Tables

•Used when studying relationship between two categorical variables.

•Example: CT scan findings vs. biopsy result

CT Findings \ Biopsy	Positive	Negative	Total
Positive CT	40	10	50
Negative CT	5	45	50
Total	45	55	100

Graphical Presentations

a) Histogram

- Bar graph of frequency distribution of continuous data.

- Example: Histogram of **lesion size (cm)** in patients.

b) Frequency Polygon

- Line graph connecting midpoints of histogram bars.

- Useful for comparing multiple datasets (e.g., lesion sizes in male vs. female)

c) Pie Chart

- Circular chart showing proportions.
- Example: Distribution of imaging modalities used:
  - CT = 40%
  - MRI = 30%
  - Ultrasound = 20%
  - X-ray = 10%

d) Scatter Plot

- Graph for relationship between two continuous variables.

- Example: Scatter plot of **age vs. tumor size**.

- Helps detect **correlation** (positive, negative, or none).

Application in Radiology

- Comparing diagnostic accuracy of CT vs. MRI.

- Analyzing dose-response relationships (radiation dose vs. image quality).

- Studying patient outcomes based on imaging findings.

- Visualizing trends in tumor growth or lesion size distribution.

Summary

- Biostatistics is essential for analyzing radiological data.

- Data can be qualitative or quantitative.

- Presentation includes tables and graphs (histograms, pie charts, scatter plots).

- Cross-tabulation is vital for diagnostic test evaluation.

Types of Data – Worked Example

Example: A radiology department records the following for 3 patients:

Patient	Gender	Tumor Stage	Lesion Size (cm)	Number of Lesions	Modality
1	M	II	2.5	1	CT
2	F	III	3.0	2	MRI
3	M	I	1.8	1	CT

- **Step 1 – Identify data types:**
- Gender → **Nominal (qualitative)**
- Tumor stage → **Ordinal (qualitative, ordered I–IV)**
- Lesion size → **Continuous quantitative**
- Number of lesions → **Discrete quantitative**
- Modality → **Nominal (categorical)**
- **Explanation:**
- Nominal: Categories with no order (Gender, Modality)
- Ordinal: Categories with a clear order (Tumor Stage)
- Discrete: Countable numbers (Number of lesions)
- Continuous: Measurable (Lesion size in cm)

Frequency Table – Worked Example

- **Objective:** Summarize lesion sizes for 15 patients into intervals.

- **Lesion sizes (cm):** 1.2, 2.5, 2.8, 3.0, 3.2, 3.5, 4.0, 4.2, 4.5, 5.0, 5.2, 5.5, 6.0, 6.2, 6.5

- **Step 1 – Define intervals:** 1–2, 2–3, 3–4, 4–5, 5–6, 6–7

Step 2 – Count frequencies:

Interval (cm)	Frequency
1–2	1
2–3	2
3–4	4
4–5	3
5–6	3
6–7	2

Step 3 – Explanation:
- Frequency table helps visualize **how lesion sizes are distributed** before plotting a histogram.

Two-Way Cross Classification Table – Worked Example

Objective: Compare CT findings to biopsy results in 50 patients.

Data:

Step 1 – Calculate total patients:
Total = 20 + 5 + 2 + 23 = 50

CT Findings	Biopsy Positive	Biopsy Negative
Positive	20	5
Negative	2	23

Step 2 – Add totals to table:

CT Findings	Biopsy Positive	Biopsy Negative	Total
Positive	20	5	25
Negative	2	23	25
Total	22	28	50

- **Step 3 – Compute diagnostic metrics:**
- **Sensitivity** = True Positives / (True Positives + False Negatives)
  = 20 / (20 + 2) = 0.909 → 90.9%
- **Specificity** = True Negatives / (True Negatives + False Positives)
  = 23 / (23 + 5) = 0.821 → 82.1%
- **Explanation:** Two-way tables are essential for evaluating diagnostic accuracy.

Histogram – Worked Example

Step 1: Use lesion size frequency table (from section 2).

Interval (cm)	Frequency
1–2	1
2–3	2
3–4	4
4–5	3
5–6	3
6–7	2

- **Step 2:** Plot histogram:
- X-axis: Lesion size intervals
- Y-axis: Frequency
- Draw bars for each interval with height = frequency
- **Step 3 – Interpretation:**
- Most lesions are **3–4 cm**, indicating typical size distribution.

Frequency Polygon – Worked Example

Step 1: Find midpoints of each interval:

Interval	Midpoint	Frequency
1–2	1.5	1
2–3	2.5	2
3–4	3.5	4
4–5	4.5	3
5–6	5.5	3
6–7	6.5	2

- **Step 2:** Plot points (midpoint, frequency) and connect with lines.

- **Step 3 – Interpretation:**

- Shape of the distribution is visible; can compare male vs female lesion sizes by plotting two lines.

Pie Chart – Worked Example

Objective: Show distribution of imaging modalities.

Modality	Count
CT	20
MRI	15
US	10
X-ray	5

- **Step 1 – Convert to percentages:**
- CT: 20/50 × 100 = 40%
- MRI: 15/50 × 100 = 30%
- US: 10/50 × 100 = 20%
- X-ray: 5/50 × 100 = 10%
- **Step 2 – Draw pie chart:**
- Sectors proportional to percentages.
- Label each sector.

Scatter Plot – Worked Example

Objective: Examine relationship between age and tumor size.

Patient	Age (years)	Tumor Size (cm)
1	25	1.5
2	30	2.0
3	35	2.5
4	40	3.0
5	45	3.5
6	50	4.0

- **Step 1 – Plot points:**
- X-axis: Age
- Y-axis: Tumor Size
- **Step 2 – Interpretation:**
- Positive trend: Older patients tend to have larger tumors (in this dataset).
- Can fit a **regression line** if needed.