Biostatistics Made Easy

A Beginner's Guide with Worked Examples for Medical and Dental Students and Postgraduate Trainees

Prof. Zahid Mahmood · Professor of Surgery, Lahore Medical and Dental College · professorzahid.com

Nov 4, 2026 · @prof zahid mahmood

Preface

Biostatistics is the science of collecting, summarising and interpreting numbers about health and disease. Every time you read "p < 0.05" in a journal, biostatistics is speaking to you.

This book teaches it with almost no heavy mathematics. Each idea is explained in plain words, followed by a small worked example from surgical practice that you can solve with a simple calculator.

Who it is for: MBBS and BDS students, house officers, and postgraduate trainees preparing a synopsis, dissertation or exam.

How to use it

1. Read each chapter, then repeat the worked example yourself on paper.
2. Learn the "Remember" box at the end of each chapter; these points are asked most in vivas and MCQs.
3. Keep the formula sheet at the end beside you while analysing your own data.
4. Test yourself with the practice questions.

This book is a companion to *Research Methodology Made Easy*. That book explains how to plan a study; this one explains how to handle the numbers it produces.

Chapter 1: What Is Biostatistics? Data and Variables

Biostatistics has two jobs: to **describe** the patients we studied, and to **infer** (draw conclusions) about all similar patients.

- **Descriptive statistics:** summarise data — averages, percentages, tables, graphs.
- **Inferential statistics:** go from sample to population — p-values, confidence intervals, tests.

Population and sample

- **Population:** all the people of interest (all adults with inguinal hernia in Pakistan).

- **Sample:** the group actually studied (120 patients operated in your unit).
- **Parameter:** a value describing the population (usually unknown).
- **Statistic:** a value calculated from the sample, used to estimate the parameter.

Types of variables

Type	Scale	Meaning	Surgical example
Qualitative (categorical)	Nominal	Named groups, no order	Gender, blood group, wound infection yes/no
Qualitative (categorical)	Ordinal	Groups with a natural order	ASA grade I–IV, Clavien-Dindo grade, pain mild/moderate/severe
Quantitative (numerical)	Discrete	Counted, whole numbers	Number of pints transfused, number of polyps
Quantitative (numerical)	Continuous	Measured, any value	Age, haemoglobin, operating time, BMI

Continuous data are sometimes further divided into **interval** (zero is arbitrary, e.g. temperature in °C) and **ratio** (true zero, e.g. weight). For most clinical work, "continuous" is enough.

A binary (dichotomous) variable has only two values: dead/alive, infected/not infected. It is a special nominal variable and is very common in surgery.

Worked example. In a study of 50 patients after cholecystectomy you record: age (years), gender, ASA grade, operating time (minutes), conversion to open (yes/no), and number of ports.

Variable	Type
Age	Continuous
Gender	Nominal (binary)
ASA grade	Ordinal
Operating time	Continuous
Conversion to open	Nominal (binary)
Number of ports	Discrete

Converting variables. A continuous variable can be grouped into categories (age → < 40 / ≥ 40), but you lose information. Collect the exact number; you can always group it later.

> **Remember:** Identify the type of variable first. It decides how you summarise it, which graph you draw and which test you use.

Chapter 2: Presenting Data in Tables and Graphs

Raw data are a long list of numbers that nobody can read. A good table or graph shows the pattern at a glance.

The frequency table

It lists each category, how many patients fall in it (frequency) and the percentage.

Worked example. ASA grades of 40 patients undergoing elective surgery.

ASA grade	Frequency (n)	Percentage	Cumulative percentage
I	14	35%	35%
II	16	40%	75%
III	8	20%	95%
IV	2	5%	100%
Total	40	100%	—

Percentage = (frequency ÷ total) × 100. For ASA I: (14 ÷ 40) × 100 = 35%. Cumulative percentage adds each row to the ones above it; it tells us that 75% of patients were ASA I or II.

Rules for a good table

1. A clear title saying what, who, where and when
2. Units in column headings (minutes, kg, mmol/L)
3. Totals shown
4. Percentages with n, e.g. 8 (20%)
5. Understandable without reading the text

Choosing the right graph

Type of data	Graph	What it shows
Categorical	Bar chart	Frequency of each category; bars are separated by gaps

Type of data	Graph	What it shows
Categorical (parts of a whole)	Pie chart	Share of each category; use only with few categories
Continuous	Histogram	Shape of distribution; bars touch each other
Continuous, compare groups	Box-and-whisker plot	Median, IQR, range and outliers
Change over time	Line graph	Trend, e.g. monthly infection rate
Two continuous variables	Scatter plot	Relationship, e.g. BMI vs operating time
Survival over time	Kaplan-Meier curve	Proportion surviving at each time (Chapter 13)

Bar chart vs histogram. A bar chart is for categories, so the bars have gaps. A histogram is for continuous data divided into class intervals (e.g. age 20–29, 30–39), so the bars touch.

Common mistakes: 3-D pie charts, too many categories, axes that do not start at zero (exaggerating differences), and graphs repeating the same numbers already in a table.

> **Remember:** Categorical → bar or pie. Continuous → histogram or box plot. Two continuous → scatter. Over time → line.

Chapter 3: Measures of Central Tendency — Mean, Median, Mode

A measure of central tendency is one number that represents the "typical" value of the data.

Mean (average) = sum of all values ÷ number of values.

```
\bar{x} = \frac{\sum x}{n}
```

Median = the middle value when data are arranged in order. If the number of values is even, take the average of the two middle values.

Mode = the value that occurs most often.

Worked example 1 — symmetrical data. Operating time (minutes) of 5 laparoscopic cholecystectomies: 40, 45, 50, 55, 60.

- Mean = (40 + 45 + 50 + 55 + 60) ÷ 5 = 250 ÷ 5 = **50 minutes**
- Median = middle (3rd) value = **50 minutes**

- Mode = none (no value repeats)

When data are symmetrical, mean and median are the same or close.

Worked example 2 — skewed data. Hospital stay (days) of 7 patients after laparotomy: 2, 3, 3, 4, 5, 6, 26.

- Mean = 49 ÷ 7 = **7 days**
- Median = 4th value = **4 days**
- Mode = **3 days**

One patient who stayed 26 days (an *outlier*) pulled the mean up to 7, although 6 of 7 patients stayed 6 days or less. The median (4 days) describes the typical patient much better.

Worked example 3 — even number of values. Stay of 6 patients: 2, 3, 3, 4, 5, 6. The two middle values are 3 and 4, so median = (3 + 4) ÷ 2 = **3.5 days**.

Which one to use?

Situation	Best measure
Continuous data, roughly symmetrical (normal)	Mean
Continuous data, skewed or with outliers	Median
Ordinal data (ASA grade, pain score)	Median
Nominal data (blood group)	Mode

Skewness in one line. In positively (right) skewed data, the tail is to the right and **mean > median > mode** (hospital stay, cost, CRP). In negatively (left) skewed data, the tail is to the left and **mean < median < mode** (e.g. age at death in a developed country).

> **Remember:** The mean is pulled by outliers; the median is not. Skewed data → report the median.

Chapter 4: Measures of Dispersion — How Spread Out Are the Data?

Two groups can have the same mean but very different spread. Surgeon A's operating times of 49, 50, 51 minutes and Surgeon B's of 20, 50, 80 minutes both average 50, yet Surgeon A is far more consistent. Measures of dispersion capture this.

Range = highest value − lowest value. Simple, but it depends on only two (often extreme) values.

Variance and standard deviation (SD). The SD is roughly the average distance of values from the mean. Variance is the SD squared.

```
SD = \sqrt{\frac{\sum (x - \bar{x})^{2}}{n - 1}}
```

We divide by n − 1 (not n) for a sample; this gives a better estimate of the population value.

Worked example. Operating times: 40, 45, 50, 55, 60 minutes (mean = 50).

Value (x)	x − mean	(x − mean)²
40	−10	100
45	−5	25
50	0	0
55	5	25
60	10	100
Total	0	**250**

1. Sum of squares = 250
2. Variance = 250 ÷ (5 − 1) = **62.5 min²**
3. SD = √62.5 = **7.9 minutes**
4. Range = 60 − 40 = **20 minutes**

Report as: operating time 50 ± 7.9 minutes (mean ± SD).

Interquartile range (IQR). Arrange data in order and split into four equal parts. Q1 (25th percentile) and Q3 (75th percentile) mark the middle 50% of patients. IQR = Q3 − Q1. Use it with the median for skewed data.

Worked example. Hospital stay (days) of 11 patients, in order: 1, 2, 2, 3, 3, **4**, 4, 5, 6, 8, 15.

- Median = 6th value = **4 days**
- Q1 = median of lower half (1, 2, 2, 3, 3) = **2 days**
- Q3 = median of upper half (4, 5, 6, 8, 15) = **6 days**
- IQR = 6 − 2 = **4 days**

Report as: median stay 4 days (IQR 2–6). Software may give slightly different quartiles because there are several calculation methods; that is acceptable.

Coefficient of variation (CV) = (SD ÷ mean) × 100. It compares spread of measurements on different scales. In our example, CV = (7.9 ÷ 50) × 100 = **15.8%**.

Central value	Matching spread	Use for
Mean	SD	Symmetrical continuous data
Median	IQR (or range)	Skewed or ordinal data

> **Remember:** Always pair them correctly — mean ± SD, or median (IQR). Never report a mean with an IQR.

Chapter 5: Probability and the Normal Distribution

Probability is the chance that an event happens, written from 0 (impossible) to 1 (certain), or 0% to 100%.

```
P(\text{event}) = \frac{\text{number of times the event occurs}}{\text{total number of trials}}
```

Example. If 12 of 120 patients develop wound infection, the probability of infection = 12 ÷ 120 = 0.10 (10%). The probability of *no* infection = 1 − 0.10 = 0.90.

Two simple rules

- **Addition rule (OR):** for events that cannot happen together, add the probabilities. If 40% of patients are blood group O and 30% are group B, P(O or B) = 0.40 + 0.30 = 0.70.
- **Multiplication rule (AND):** for independent events, multiply. If the infection risk is 0.10, the chance that two unrelated patients *both* get infected = 0.10 × 0.10 = 0.01 (1%).

The normal (Gaussian) distribution

Many biological measurements — height, haemoglobin, blood pressure in healthy adults — form a symmetrical bell-shaped curve when plotted. Its features:

- Mean = median = mode, all in the centre
- Symmetrical on both sides
- Defined completely by its mean and SD

The 68–95–99.7 rule

Range	Percentage of values inside
Mean ± 1 SD	About 68%
Mean ± 1.96 SD (≈ 2 SD)	95%
Mean ± 3 SD	About 99.7%

Worked example. Haemoglobin of healthy adult men has mean 15 g/dL and SD 1.2 g/dL.

- 68% lie between 15 ± 1.2 = **13.8 to 16.2 g/dL**
- 95% lie between 15 ± (1.96 × 1.2) = 15 ± 2.35 = **12.65 to 17.35 g/dL**

This is exactly how many laboratory **reference ranges** ("normal ranges") are made: the middle 95% of healthy people. It also means 5% of perfectly healthy people fall outside the "normal" range.

The Z-score (standard score) tells how many SDs a value lies from the mean.

```
Z = \frac{x - \text{mean}}{SD}
```

A man with Hb 12.6 g/dL: Z = (12.6 − 15) ÷ 1.2 = **−2.0**. He is 2 SDs below the mean; only about 2.3% of healthy men have a value this low or lower.

Other distributions you may hear about

- **Binomial:** number of "yes" outcomes in a fixed number of patients (how many of 20 develop a complication).
- **Poisson:** counts of rare events in a time period (number of bile duct injuries per year).
- **t-distribution:** like the normal curve but with fatter tails; used for small samples (Chapter 8).

Is my data normal? Look at a histogram (bell shape?), check that mean and median are close, or use the Shapiro-Wilk test (for small samples) or Kolmogorov-Smirnov test. A p-value > 0.05 on these tests suggests the data are compatible with normality.

> **Remember:** Mean ± 1.96 SD covers 95% of individuals in a normal distribution. This number, 1.96, appears again and again in statistics.

Chapter 6: Sampling, Standard Error and Confidence Intervals

If ten surgeons each studied a different sample of 100 hernia patients, they would get ten slightly different means. The standard error tells us how much a sample mean is likely to vary from the true population mean.

SD vs SE — the most confused pair in statistics

	Standard deviation (SD)	Standard error (SE)
Describes	Spread of individual patients	Precision of the sample mean
Gets smaller with bigger sample?	No	Yes
Use it to	Describe your patients	Build confidence intervals and tests

```
SE_{\text{mean}} = \frac{SD}{\sqrt{n}}
```

The central limit theorem (in simple words): even if individual values are not normal, the means of large samples (roughly n ⩾ 30) follow a normal distribution. This is why many tests work in practice.

Confidence interval (CI)

A 95% CI is the range within which we are 95% confident the true population value lies.

```
95\%\ CI = \text{mean} \pm 1.96 \times SE
```

Worked example 1 — mean. 100 patients after open hernia repair stayed a mean of 5 days, SD 2 days.

1. SE = 2 ÷ √100 = 2 ÷ 10 = **0.2 days**
2. 1.96 × 0.2 = 0.39
3. 95% CI = 5 ± 0.39 = **4.61 to 5.39 days**

We are 95% confident that the true mean stay of all such patients lies between about 4.6 and 5.4 days.

Worked example 2 — proportion. 20 of 200 patients developed wound infection (p = 0.10).

```
SE_{p} = \sqrt{\frac{p(1-p)}{n}}
```

1. SE = √(0.10 × 0.90 ÷ 200) = √0.00045 = **0.0212**
2. 1.96 × 0.0212 = 0.0416
3. 95% CI = 0.10 ± 0.042 = **5.8% to 14.2%**

What makes a CI narrower (more precise)?

- A larger sample size (n is under the square root)
- Less variability (smaller SD)
- A lower confidence level (90% CI is narrower than 95%, but less sure)

To halve the width of a CI you need about **four times** as many patients.

Reading a CI for significance

- For a **difference** (mean difference, risk difference): if the CI includes **0**, the result is not statistically significant.
- For a **ratio** (RR, OR, HR): if the CI includes **1**, the result is not statistically significant.

> **Remember:** SD describes patients; SE describes the accuracy of the mean. A CI gives more information than a p-value because it shows both significance and the size of the effect.

Chapter 7: Hypothesis Testing, p-Value, Errors and Power

Hypothesis testing asks one question: could the difference I found simply be due to chance?

A courtroom analogy. The accused (the null hypothesis) is presumed innocent until the evidence is strong enough to convict "beyond reasonable doubt". In statistics, "reasonable doubt" is usually set at 5%.

The five steps

1. **State the hypotheses.** H0: no difference between groups. H1: there is a difference.
2. **Set the significance level (α),** usually 0.05.
3. **Choose the right test** for the type of data (Chapters 8–10, 14).
4. **Calculate the test statistic and p-value** (software does this).
5. **Decide:** if $p < 0.05$, reject H0 ("statistically significant"); if $p \geq 0.05$, we fail to reject H0.

What the p-value really means

The p-value is the probability of getting a result as extreme as yours, or more, **if the null hypothesis were true**.

Example. Mean blood loss is 120 mL with harmonic scalpel and 160 mL with conventional diathermy; $p = 0.03$. Meaning: if there were truly no difference between the two methods, a difference this large would arise by chance only 3% of the time. So we reject H0.

What the p-value does NOT mean

- It is **not** the probability that H0 is true.

- It does **not** measure the size or clinical importance of the effect. With 10,000 patients, a 2 mL difference in blood loss could be "significant" yet meaningless.
- p = 0.06 does not prove "no difference"; it means the evidence was not strong enough. Absence of evidence is not evidence of absence.

Type I and type II errors

	H0 actually true (no real difference)	H0 actually false (real difference)
Study rejects H0	**Type I error (α)** — false positive	Correct (power)
Study does not reject H0	Correct	**Type II error (β)** — false negative

Memory aid: Type I = "crying wolf" (seeing something that is not there). Type II = missing the wolf that is really there.

Power = 1 − β, the chance of detecting a real difference. Usually set at 80% (β = 0.20). Power increases with:

- Larger sample size
- Larger true difference between groups
- Less variability in the data
- Higher α (but this increases type I error)

Small studies are often "underpowered": they report "no significant difference" only because they had too few patients.

One-tailed vs two-tailed tests. A two-tailed test checks for a difference in either direction (A better *or* worse than B). Use two-tailed tests unless there is a very strong reason; this is what journals expect.

Multiple testing. If you perform 20 tests at α = 0.05, about one will be "significant" by chance alone. Decide your main outcome in advance, and use corrections (such as Bonferroni: divide 0.05 by the number of tests) when making many comparisons.

> **Remember:** p < 0.05 means "unlikely to be chance", not "important". Always report the effect size with its 95% CI alongside the p-value.

Chapter 8: Comparing Means — t-Tests, ANOVA and Non-Parametric Tests

When the outcome is a continuous number (stay, blood loss, pain score, Hb), we compare means between groups. Which test depends on how many groups there are and whether the groups are independent or paired.

Parametric tests assume roughly normal data and similar spread (variance) in the groups. **Non-parametric tests** make no such assumption; they work on ranks and compare medians or distributions.

Comparison	Parametric	Non-parametric equivalent
Two independent groups	Independent (unpaired) t-test	Mann-Whitney U test
Same patients measured twice	Paired t-test	Wilcoxon signed-rank test
Three or more independent groups	One-way ANOVA	Kruskal-Wallis test
Same patients measured three or more times	Repeated-measures ANOVA	Friedman test

1. Independent t-test — worked example

Hospital stay after laparoscopic vs open appendicectomy:

Group	n	Mean stay (days)	SD
Laparoscopic	50	2.0	0.8
Open	50	3.5	1.2

```
t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{SD_1^{2}}{n_1} + \dfrac{SD_2^{2}}{n_2}}}
```

1. Difference in means = 3.5 − 2.0 = **1.5 days**
2. SE of difference = √(0.8²/50 + 1.2²/50) = √(0.0128 + 0.0288) = √0.0416 = **0.204**
3. t = 1.5 ÷ 0.204 = **7.35**
4. With about 98 degrees of freedom, a t greater than about 1.98 is significant at 0.05. Our t is far larger: **p < 0.001**.
5. 95% CI of the difference = 1.5 ± (1.98 × 0.204) = **1.1 to 1.9 days**

Conclusion: laparoscopic patients went home 1.5 days earlier (95% CI 1.1 to 1.9; p < 0.001).

2. Paired t-test — worked example

Haemoglobin (g/dL) of 5 patients before and after hemicolectomy:

Patient	Before	After	Difference (d)
1	12	10	2
2	11	10	1
3	13	11	2
4	12	11	1
5	14	12	2

1. Mean difference = (2 + 1 + 2 + 1 + 2) ÷ 5 = **1.6 g/dL**
2. SD of differences = **0.55**
3. SE = 0.55 ÷ √5 = **0.245**
4. t = 1.6 ÷ 0.245 = **6.5**, with 4 degrees of freedom (n − 1)
5. The critical value for df = 4 is 2.78; 6.5 is larger, so **p < 0.01**

Conclusion: Hb fell significantly after surgery. Pairing removes differences between patients, so paired tests are more powerful when the same patients are measured twice.

3. One-way ANOVA

Used for three or more groups, e.g. mean operating time for three surgeons. ANOVA tells you only *whether* any difference exists. If p < 0.05, a **post-hoc test** (Tukey, Bonferroni) tells you *which* groups differ. Do not run many separate t-tests instead; that inflates type I error.

4. Non-parametric tests — when to use

- Data clearly skewed (hospital stay, cost, CRP)
- Ordinal outcome (pain score 0–10, satisfaction scale)
- Very small samples where normality cannot be checked

Example. Pain scores (0–10) on day 1 after Lichtenstein vs laparoscopic TEP hernia repair → compare with the **Mann-Whitney U test** and report medians (IQR).

> **Remember:** Two groups → t-test (or Mann-Whitney). Same patients twice → paired t-test (or Wilcoxon). Three or more groups → ANOVA (or Kruskal-Wallis).

Chapter 9: Comparing Proportions — Chi-Square and Fisher's Exact Test

When the outcome is categorical (infected / not infected, cured / not cured), we compare proportions, not means. The data are arranged in a **contingency table**.

The chi-square (χ^2) test compares the numbers we *observed* with the numbers we would *expect* if there were no difference between groups.

```
\chi^{2} = \sum \frac{(O - E)^{2}}{E}
```

O = observed number in a cell; E = expected number in that cell.

Worked example. Wound infection after laparoscopic vs open appendicectomy.

Group	Infected	Not infected	Total
Laparoscopic	6	94	100
Open	18	82	100
Total	24	176	200

Step 1 — expected values. E = (row total × column total) ÷ grand total.

- Infected, each group: (100 × 24) ÷ 200 = **12**
- Not infected, each group: (100 × 176) ÷ 200 = **88**

Step 2 — calculate χ^2.

Cell	O	E	$(O - E)^2 \div E$
Lap, infected	6	12	36 ÷ 12 = 3.00
Lap, not infected	94	88	36 ÷ 88 = 0.41
Open, infected	18	12	36 ÷ 12 = 3.00
Open, not infected	82	88	36 ÷ 88 = 0.41
Total χ^2			**6.82**

Step 3 — degrees of freedom = (rows − 1) × (columns − 1) = (2 − 1) × (2 − 1) = **1**.

Step 4 — decide. For df = 1, the critical value at p = 0.05 is **3.84**. Our χ^2 = 6.82 is larger, so **p < 0.05** (exactly, p ≈ 0.009).

Conclusion: wound infection was significantly lower after laparoscopic appendicectomy (6% vs 18%; p = 0.009).

Conditions for chi-square

- Data are counts (not percentages or means)
- Each patient appears in only one cell
- Expected value in each cell is at least 5 (in a 2 × 2 table)

Fisher's exact test. If any expected cell value is below 5 in a 2 × 2 table (common with rare complications or small samples), use Fisher's exact test instead. SPSS gives it automatically with the chi-square output.

Yates' continuity correction is sometimes applied to 2 × 2 tables to make chi-square more conservative; many statisticians now prefer simply reporting Fisher's test for small samples.

McNemar's test. For paired categorical data — the same patients assessed twice (e.g. pain present yes/no before and after a nerve block).

Chi-square for trend. When one variable is ordinal (age groups, ASA grade) and you want to see whether the outcome rises steadily across the categories.

> **Remember:** Categorical outcome, independent groups → chi-square; small expected numbers → Fisher's exact; same patients twice → McNemar.

Chapter 10: Correlation and Regression

Correlation asks: do two continuous variables move together, and how strongly? Regression asks: can one variable predict the other, and by how much?

Correlation coefficient (r) ranges from −1 to +1.

Value of r	Meaning
+1	Perfect positive: as one rises, the other rises
0	No linear relationship
−1	Perfect negative: as one rises, the other falls
0.0–0.3 (either sign)	Weak
0.3–0.7	Moderate
0.7–1.0	Strong

- **Pearson's r:** both variables continuous and roughly normal.
- **Spearman's rho (ρ):** skewed or ordinal data; based on ranks.

Worked example. BMI and operating time for 5 laparoscopic cholecystectomies (small, tidy numbers for learning):

BMI (x)	Time, min (y)	x − 30	y − 56	Product	$(x-30)^2$	$(y-56)^2$
20	40	−10	−16	160	100	256
25	50	−5	−6	30	25	36
30	55	0	−1	0	0	1
35	65	5	9	45	25	81
40	70	10	14	140	100	196
Total				**375**	**250**	**570**

Mean BMI = 30; mean time = 56.

```
r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^{2} \, \sum (y-\bar{y})^{2}}}
```

r = 375 ÷ √(250 × 570) = 375 ÷ 377.5 = **0.99** — a very strong positive correlation.

Coefficient of determination (r^2) = $0.99^2 \approx$ **0.98**: about 98% of the variation in operating time is explained by BMI in this (idealised) sample. Real clinical data are rarely this tidy.

Correlation is not causation. Ice-cream sales and drowning deaths are correlated — both rise in summer. A third factor (heat) explains both.

Simple linear regression draws the best straight line through the points:

```
y = a + b x
```

- **b (slope)** = $\Sigma(x-\bar{x})(y-\bar{y}) \div \Sigma(x-\bar{x})^2$ = 375 ÷ 250 = **1.5**. Each 1-unit rise in BMI adds 1.5 minutes.
- **a (intercept)** = $\bar{y} - b\,\bar{x}$ = 56 − (1.5 × 30) = **11**.
- **Equation:** Operating time = 11 + 1.5 × BMI.
- **Prediction:** for BMI 28, time = 11 + (1.5 × 28) = **53 minutes**.

Only predict within the range of your data (BMI 20–40 here). Predicting for BMI 60 is unsafe.

Multiple and logistic regression (concept only)

- **Multiple linear regression:** several predictors of a continuous outcome (operating time predicted by BMI, age and previous surgery). Each coefficient is "adjusted" for the others, which controls confounding.
- **Logistic regression:** predicts a yes/no outcome (wound infection) from several factors. Results are reported as **adjusted odds ratios** with 95% CI. Take a statistician's help for these.

> **Remember:** r measures strength of a straight-line relationship; regression gives the equation for prediction. Neither proves cause.

Chapter 11: Rates and Measures of Risk

These measures tell us how common a disease is and how strongly an exposure or treatment changes the risk.

Incidence and prevalence

- **Incidence:** *new* cases arising in a population during a period. Measures risk. "30 new cases of breast cancer per 100,000 women per year."
- **Prevalence:** *all* existing cases (old + new) at a point in time. Measures burden. "8% of adults have gallstones."
- Prevalence ≈ Incidence × average duration of disease. A chronic disease with long survival has high prevalence even with low incidence.

Other rates

- **Case fatality rate** = deaths from a disease ÷ people with that disease × 100
- **Mortality rate** = deaths in a period ÷ total population × (1,000 or 100,000)
- **Complication rate** = patients with a complication ÷ patients operated × 100

The standard 2 × 2 table

	Outcome present	Outcome absent
Exposed / treated	a	b
Not exposed / control	c	d

Worked example 1 — cohort study (relative risk). Wound dehiscence after laparotomy in smokers vs non-smokers.

	Dehiscence	No dehiscence	Total
Smokers	30	170	200
Non-smokers	20	380	400

1. Risk in smokers = 30 ÷ 200 = **0.15 (15%)**
2. Risk in non-smokers = 20 ÷ 400 = **0.05 (5%)**
3. **Relative risk (RR)** = 0.15 ÷ 0.05 = **3.0** — smokers have three times the risk.
4. **Attributable risk (risk difference)** = 0.15 − 0.05 = **0.10 (10%)**
5. **Number needed to harm (NNH)** = 1 ÷ 0.10 = **10** — for every 10 smokers operated, one extra dehiscence occurs.

Worked example 2 — case-control study (odds ratio). 100 patients with gastric cancer and 100 controls tested for H. pylori.

	Gastric cancer (cases)	No cancer (controls)
H. pylori positive	70 (a)	40 (b)
H. pylori negative	30 (c)	60 (d)

```
OR = \frac{a \times d}{b \times c}
```

OR = (70 × 60) ÷ (40 × 30) = 4,200 ÷ 1,200 = **3.5**. The odds of H. pylori infection were 3.5 times higher in cancer patients.

In case-control studies we cannot calculate risk (we chose how many cases and controls to include), so we use the OR. When the disease is rare, the OR approximates the RR.

Worked example 3 — treatment trial (NNT). From Chapter 9: wound infection 18% after open and 6% after laparoscopic appendicectomy.

1. **Relative risk** = 0.06 ÷ 0.18 = **0.33**
2. **Absolute risk reduction (ARR)** = 0.18 − 0.06 = **0.12 (12%)**
3. **Relative risk reduction (RRR)** = 0.12 ÷ 0.18 = **67%**
4. **Number needed to treat (NNT)** = 1 ÷ 0.12 = 8.3, rounded **up** to **9**

So 9 patients must have laparoscopic surgery to prevent one wound infection.

Relative vs absolute — why it matters. A drug that cuts a complication from 2% to 1% has an impressive RRR of 50%, but an ARR of only 1% (NNT = 100). Advertisements quote the RRR; patients deserve to hear the ARR and NNT.

Interpreting RR and OR

Value	Meaning
1	No association
More than 1	Increased risk (harmful exposure)
Less than 1	Reduced risk (protective exposure or effective treatment)

> **Remember:** Cohort and RCT → RR. Case-control → OR. Always round NNT up, and always report absolute as well as relative effects.

Chapter 12: Evaluating Diagnostic Tests

A diagnostic test is judged by comparing it with a **gold standard** (reference test) that tells the true diagnosis — for example, histopathology of the removed appendix.

The diagnostic 2 × 2 table

	Disease present	Disease absent
Test positive	True positive (TP)	False positive (FP)
Test negative	False negative (FN)	True negative (TN)

Formulas

Measure	Formula	Question it answers
Sensitivity	TP ÷ (TP + FN)	Of those WITH disease, how many test positive?
Specificity	TN ÷ (TN + FP)	Of those WITHOUT disease, how many test negative?
Positive predictive value (PPV)	TP ÷ (TP + FP)	If the test is positive, how likely is disease?
Negative predictive value (NPV)	TN ÷ (TN + FN)	If the test is negative, how likely is no disease?
Diagnostic accuracy	(TP + TN) ÷ total	How many results were correct overall?

Worked example. Ultrasound for acute appendicitis in 200 patients; histopathology is the gold standard.

	Appendicitis (histology)	No appendicitis	Total
Ultrasound positive	90 (TP)	10 (FP)	100
Ultrasound negative	20 (FN)	80 (TN)	100
Total	110	90	200

1. Sensitivity = 90 ÷ 110 = **81.8%**
2. Specificity = 80 ÷ 90 = **88.9%**
3. PPV = 90 ÷ 100 = **90%**
4. NPV = 80 ÷ 100 = **80%**
5. Accuracy = (90 + 80) ÷ 200 = **85%**

Memory aids

- **SnNout:** a highly **Sn**sitive test, when **N**egative, rules **out** disease (few false negatives). Good for screening.
- **SpPin:** a highly **Sp**ecific test, when **P**ositive, rules **in** disease (few false positives). Good for confirmation.

Predictive values depend on prevalence. Sensitivity and specificity belong to the test; PPV and NPV change with how common the disease is in the people tested.

The same ultrasound used in 1,000 patients where only 10% truly have appendicitis:

	Appendicitis	No appendicitis	Total
Ultrasound positive	82	100	182
Ultrasound negative	18	800	818
Total	100	900	1,000

Now PPV = 82 ÷ 182 = **45%** (it was 90%), while NPV = 800 ÷ 818 = **97.8%**. When a disease is rare, even a good test gives many false positives.

Likelihood ratios (LR) do not depend on prevalence and are useful at the bedside.

- **LR+** = sensitivity ÷ (1 − specificity) = 0.818 ÷ 0.111 = **7.4**
- **LR−** = (1 − sensitivity) ÷ specificity = 0.182 ÷ 0.889 = **0.20**

Rough guide: LR+ above 10 or LR− below 0.1 changes diagnosis strongly; LR+ 5–10 or LR− 0.1–0.2 moderately; values near 1 hardly at all.

ROC curve. For a test with a numerical result (Alvarado score, CRP, PSA), sensitivity is plotted against (1 − specificity) at every possible cut-off. The **area under the curve (AUC)** summarises performance:

AUC	Interpretation
0.5	No better than tossing a coin
0.7–0.8	Acceptable
0.8–0.9	Good / excellent
Above 0.9	Outstanding

Raising the cut-off increases specificity but lowers sensitivity; there is always a trade-off.

> **Remember:** Sensitivity and specificity are fixed properties of the test; PPV and NPV change with prevalence.

Chapter 13: Survival Analysis Made Simple

Survival analysis studies *time until an event* — death, recurrence of hernia, anastomotic leak, or failure of a fistula repair. The "event" need not be death.

Why not just use percentages? Patients enter a study at different times and are followed for different lengths of time. Some are lost or the study ends before their event happens. These patients are **censored**: we know they were event-free up to a certain time, but not after. Survival methods use this partial information instead of throwing it away.

Kaplan-Meier method — worked example

10 patients followed after surgery for colon cancer.

Time (months)	Event	At risk just before	Deaths	Survival at this step	Cumulative survival
6	Death	10	1	9 ÷ 10 = 0.90	**0.90**
9	Lost to follow-up (censored)	9	0	—	0.90
12	2 deaths	8	2	6 ÷ 8 = 0.75	0.90 × 0.75 = **0.675**
18	Censored	6	0	—	0.675
24	Death	5	1	4 ÷ 5 = 0.80	0.675 × 0.80 = **0.54**

At 24 months, the estimated survival is **54%**; the remaining 4 patients were alive at the end of the study (censored).

Key points:

- Cumulative survival is calculated only when an event occurs, by multiplying the step survivals.
- Censored patients are removed from the "at risk" number but do not count as deaths.
- Plotted as a step-shaped **Kaplan-Meier curve**; ticks on the curve mark censored patients.

Median survival = the time at which cumulative survival falls to 50%. In our example it was not reached by 24 months.

Comparing two survival curves

- **Log-rank test:** tests whether two or more Kaplan-Meier curves differ significantly (e.g. survival after surgery with vs without adjuvant chemotherapy). It gives a p-value.
- **Cox proportional hazards regression:** gives the **hazard ratio (HR)** while adjusting for other factors (age, stage, grade).

Interpreting the hazard ratio. HR = 0.60 (95% CI 0.45–0.80) for chemotherapy means the risk of death at any point in time was 40% lower with chemotherapy, and the CI does not include 1, so it is significant. HR > 1 means higher hazard.

Common survival terms

- **Overall survival (OS):** time to death from any cause
- **Disease-free survival (DFS):** time to recurrence or death
- **5-year survival:** proportion alive at 5 years, read from the Kaplan-Meier curve

> **Remember:** Censored patients are not failures and not ignored; they contribute until they leave. Kaplan-Meier describes, log-rank compares, Cox adjusts.

Chapter 14: Choosing the Right Test and Running It in SPSS

Ask three questions and the test chooses itself: What type is my outcome? How many groups? Are the groups independent or paired?

Master table — which test?

Outcome type	Groups	Independent or paired	Test
Continuous, normal	1 group vs a known value	—	One-sample t-test
Continuous, normal	2	Independent	Independent t-test
Continuous, normal	2	Paired	Paired t-test
Continuous, normal	3 or more	Independent	One-way ANOVA (+ post-hoc)
Continuous skewed / ordinal	2	Independent	Mann-Whitney U
Continuous skewed / ordinal	2	Paired	Wilcoxon signed-rank
Continuous skewed / ordinal	3 or more	Independent	Kruskal-Wallis
Categorical	2 or more	Independent	Chi-square (Fisher's exact if expected < 5)
Categorical	2	Paired	McNemar
Two continuous variables	—	—	Pearson (normal) or Spearman (skewed) correlation

Outcome type	Groups	Independent or paired	Test
Yes/no outcome, several predictors	—	—	Logistic regression
Time to event	2 or more	Independent	Kaplan-Meier + log-rank; Cox regression

Setting up data in SPSS

1. **Variable View:** define each variable — name (no spaces), type, labels, values (1 = Male, 2 = Female) and measure (Scale, Ordinal, Nominal).
2. **Data View:** one row per patient, one column per variable.
3. Check for errors: Analyze → Descriptive Statistics → Frequencies, and look for impossible values (age 250).

Common SPSS menus

Task	Menu path
Frequencies and percentages	Analyze → Descriptive Statistics → Frequencies
Mean, SD, median	Analyze → Descriptive Statistics → Explore (also gives normality tests and box plots)
Chi-square / Fisher's	Analyze → Descriptive Statistics → Crosstabs → Statistics → Chi-square
Independent t-test	Analyze → Compare Means → Independent-Samples T Test
Paired t-test	Analyze → Compare Means → Paired-Samples T Test
One-way ANOVA	Analyze → Compare Means → One-Way ANOVA → Post Hoc
Mann-Whitney / Wilcoxon / Kruskal-Wallis	Analyze → Nonparametric Tests → Legacy Dialogs
Correlation	Analyze → Correlate → Bivariate
Logistic regression	Analyze → Regression → Binary Logistic
Kaplan-Meier	Analyze → Survival → Kaplan-Meier
ROC curve	Analyze → Classify → ROC Curve

Menu names can differ slightly between SPSS versions. Free alternatives with similar menus are **JASP** and **Jamovi**.

Reading the t-test output. SPSS first shows **Levene's test**. If its p > 0.05, read the row "Equal variances assumed"; if p ⩽ 0.05, read "Equal variances not assumed". Then report the mean difference, 95% CI and "Sig. (2-tailed)", which is the p-value.

Reading the chi-square output. Use "Pearson Chi-Square" for larger samples. If SPSS notes that a cell has expected count less than 5, report "Fisher's Exact Test" instead.

How to write results

- *Continuous:* "Mean hospital stay was shorter after laparoscopic surgery (2.0 ± 0.8 vs 3.5 ± 1.2 days; mean difference 1.5 days, 95% CI 1.1–1.9; p < 0.001)."
- *Categorical:* "Wound infection occurred in 6 (6%) laparoscopic and 18 (18%) open patients (p = 0.009)."
- Report exact p-values (p = 0.03) rather than "p < 0.05"; write p < 0.001 for very small values. Never write p = 0.000.

> **Remember:** Decide the test before collecting data, write it in the synopsis, and get a statistician to check your plan.

Formula Sheet

Measure	Formula
Mean	Σx ÷ n
Standard deviation	√[Σ(x − mean)² ÷ (n − 1)]
Coefficient of variation	(SD ÷ mean) × 100
Z-score	(x − mean) ÷ SD
Standard error of mean	SD ÷ √n
Standard error of proportion	√[p(1 − p) ÷ n]
95% confidence interval	estimate ± 1.96 × SE
Chi-square	Σ (O − E)² ÷ E; E = row total × column total ÷ grand total
Degrees of freedom (chi-square)	(rows − 1) × (columns − 1)
Relative risk	[a ÷ (a + b)] ÷ [c ÷ (c + d)]

Measure	Formula
Odds ratio	(a × d) ÷ (b × c)
Absolute risk reduction	risk in control − risk in treated
Relative risk reduction	ARR ÷ risk in control
NNT	1 ÷ ARR (round up)
Sensitivity	TP ÷ (TP + FN)
Specificity	TN ÷ (TN + FP)
PPV	TP ÷ (TP + FP)
NPV	TN ÷ (TN + FN)
LR+	sensitivity ÷ (1 − specificity)
LR−	(1 − sensitivity) ÷ specificity
Regression line	y = a + bx

Glossary

Term	Simple meaning
Censored	Patient whose event was not observed during follow-up
Confidence interval	Range likely to contain the true population value
Degrees of freedom	Number of values free to vary; used to read test tables
Hazard ratio	Ratio of event rates over time between two groups
Interquartile range	Spread of the middle 50% of values
Non-parametric test	Test that does not assume normal distribution
Null hypothesis	Statement that there is no difference or association
Outlier	Value far away from the rest
p-value	Probability of the observed result if the null hypothesis were true

Term	Simple meaning
Power	Probability of detecting a real difference
Skewed data	Data with a long tail on one side
Standard error	Precision of a sample estimate

Practice Questions

Choose the single best answer.

1. Hospital stays of five patients are 3, 5, 7, 9 and 26 days. The best measure of the typical stay, and its value, is: A. Mean, 10 days B. Median, 7 days C. Mode, 3 days D. Mean, 7 days E. Range, 23 days
2. A sample of 100 patients has a mean Hb of 12 g/dL with SD 1.0 g/dL. The standard error of the mean is: A. 1.0 g/dL B. 0.5 g/dL C. 0.1 g/dL D. 0.01 g/dL E. 10 g/dL
3. Of 50 patients with proven cholangitis, 40 have a raised bilirubin. The sensitivity of bilirubin is: A. 20% B. 40% C. 50% D. 80% E. 90%
4. Infection occurs in 20% of exposed and 10% of unexposed patients in a cohort study. The relative risk is: A. 0.5 B. 1.0 C. 2.0 D. 10 E. 20
5. A new antibiotic reduces wound infection from 30% to 20%. The number needed to treat is: A. 3 B. 5 C. 10 D. 20 E. 50
6. Postoperative pain scores (0–10 scale) are compared between two independent groups. The most suitable test is: A. Paired t-test B. Mann-Whitney U test C. Chi-square test D. Pearson correlation E. McNemar test
7. A study reports an odds ratio of 1.2 (95% CI 0.8–1.6). This means: A. Significant increased risk B. Significant decreased risk C. No statistically significant association D. The study is biased E. The sample was too large
8. A correlation coefficient of r = −0.8 between age and wound healing rate indicates: A. Weak positive B. Strong positive C. No relationship D. Weak negative E. Strong negative
9. A 2 × 2 table of a rare complication has an expected count of 3 in one cell. The appropriate test is: A. Chi-square B. Fisher's exact test C. Independent t-test D. ANOVA E. Kruskal-Wallis
10. A screening test is moved from a specialist clinic (high prevalence) to the general population (low prevalence). Which value will fall? A. Sensitivity B. Specificity C. Positive predictive value D. Negative predictive value E. Likelihood ratio

Answers

Q	Answer	Explanation
1	B	Data are skewed by the 26-day outlier, so the median (7) is best; the mean is 50 ÷ 5 = 10
2	C	SE = 1.0 ÷ √100 = 0.1 g/dL
3	D	40 ÷ 50 = 80%
4	C	0.20 ÷ 0.10 = 2.0
5	C	ARR = 0.10; NNT = 1 ÷ 0.10 = 10
6	B	Ordinal data in two independent groups
7	C	The CI of a ratio includes 1
8	E	Negative sign, magnitude above 0.7
9	B	Expected count below 5
10	C	PPV falls as prevalence falls; sensitivity and specificity stay the same

---

JazakAllah. Prof. Zahid Mahmood, Professor of Surgery, Lahore Medical and Dental College · professorzahid.com