Statistics & Exploratory Analysis Study Guide

Joel Carreira Adam Jaworski

Contents

1 Descriptive Statistics

1.1 Summary Table

Measure	R function	Formula	Notes / Use Cases
Arithmetic Mean	`mean(x)`	$\frac{\sum x_i}{n}$	Symmetric distributions; sensitive to outliers.
Harmonic Mean	`psych::harmonic.mean(x)`	$H = \frac{n}{\sum \frac{1}{x_i}}$	Rates or ratios (speed, density); positive numbers only.
Geometric Mean	`psych::geometric.mean(x)`	$GM = \sqrt[n]{\prod x_i}$	Growth rates, log-normal data; positive numbers only.
Weighted Mean	`weighted.mean(x,w)`	$\frac{\sum w_i x_i}{\sum w_i}$	Values with different importance or frequency; sensitive to outliers.
Mid-range	—	$\frac{\min(x)+\max(x)}{2}$	Rough estimate of center; highly sensitive to outliers.
Trimean	—	$TM = \frac{Q_1+2Q_2+Q_3}{4}$	Robust measure using quartiles and median; resistant to outliers.
Trimmed Mean	`mean(x, trim=0.1)`	Mean after trimming fixed proportion from tails	Reduces effect of extreme values; resistant to outliers.
Winsorized Mean	`psych::winsor.mean(x)`	Extreme values replaced, then mean computed	Reduces influence of outliers without reducing sample size.
Mode	`table(x)`	Most frequent value	Useful for categorical data; may be multimodal in numerical data.
Median	`median(x)`	50th percentile	Resistant to outliers; good for skewed distributions.
Midmean	`mean(x, trim=0.25)`	Mean of middle 50%	Resistant to outliers; special case of trimmed mean.
Quartiles	`quantile(x, probs=c(0.25,0.5,0.75))`	Q_1, Q_2, Q_3	Describe spread and center; useful for boxplots.
Deciles	`quantile(x, probs=seq(0.1,0.9,0.1))`	10 equal parts	Fine-grained ranking (e.g., income levels).
Percentiles	`quantile(x, probs=seq(0.01,0.99,0.01))`	100 equal parts	Precise benchmarking (e.g., test scores).

1.2 Measures of Location

- **Sensitive to outliers:**
  - Arithmetic mean
  - Weighted mean
  - Midrange
- **Resistant to outliers:**
  - Harmonic mean
  - Geometric mean
  - Trimmed mean
  - Winsorized mean
  - Mode
  - Median
  - Midmean
  - Trimean
- **Quantiles / Percentiles:**
  - Quartiles
  - Deciles
  - Percentiles

1.3 Proportion of Observations within an Interval

Sometimes we want to know what fraction (or percentage) of our data falls between two values, a and b. There are two common ways to estimate this:

- **Counting from the data (empirical method):** Simply count how many observations are between a and b, then divide by the total number of observations.

$$\text{Proportion} \approx \frac{\#\text{ of values between } a \text{ and } b}{\text{total } \#\text{ of values}} \times 100\%$$

This gives the exact fraction based on your data.

- **Using the normal approximation (if data is roughly bell-shaped):**
  - Convert a and b into z-scores, which measure how many standard deviations they are from the mean:
  $$z_a = \frac{a - \bar{x}}{s}, \quad z_b = \frac{b - \bar{x}}{s}$$
  - Look up these z-scores in a standard normal table (or use a calculator) to find the probabilities.
  - Subtract the smaller probability from the larger to find the proportion of data between a and b:
  $$\text{Proportion} \approx \text{Probability up to } z_b - \text{Probability up to } z_a$$
  - This method works well when the data is approximately normal.

1.4 Measures of Shape

Skewness:

- Definition: Measures the asymmetry of a distribution around its mean.
- Interpretation:
  * $g_1 \approx 0$: distribution is approximately symmetric.
  * $g_1 > 0$: right-skewed / positively skewed (tail on the right).
  * $g_1 < 0$: left-skewed / negatively skewed (tail on the left).
- Exam relevance:
  * High skewness may violate assumptions of parametric tests like t-tests or ANOVA.
  * For small sample sizes, non-normality from skewness can reduce power or increase Type I error.

Kurtosis:

- Definition: Measures the "tailedness" or peak of a distribution relative to normal.
- Interpretation:
  * 0 (excess kurtosis = 0): normal-like tails (mesokurtic).
  * > 0 (leptokurtic): heavy tails, more extreme values, higher risk of outliers.
  * < 0 (platykurtic): light tails, fewer extreme values, flatter distribution.
- Exam relevance:
  * Heavy-tailed distributions increase the chance of extreme values affecting mean and variance.
  * Can influence the choice of robust or nonparametric tests.
  * Often considered alongside skewness to assess normality assumptions.

1.5 Outliers Detection

Methods:

- **IQR Rule (Interquartile Range):**

$$x < Q_1 - 1.5 \times IQR \quad \text{or} \quad x > Q_3 + 1.5 \times IQR$$

    * Q_1 and Q_3 are the first and third quartiles; $IQR = Q_3 - Q_1$.
    * Detects moderate outliers in skewed or non-normal distributions.
  - **Z-score Method:**

$$z_i = \frac{x_i - \bar{x}}{s}$$

    * s is the standard deviation.
    * Observation is flagged as an outlier if $|z_i| > 3$.
    * Works best for approximately normal distributions.
  - **Modified Z-score (robust version):**

$$M_i = \frac{0.6745(x_i - \text{median})}{MAD}, \quad MAD = \text{median}(|x_i - \text{median}|)$$

    * Uses median and median absolute deviation (MAD) instead of mean and SD.
    * Flag as outlier if $|M_i| > 3.5$.
    * More robust than the standard Z-score, especially in skewed distributions.

1.6 Graphical Summaries

- **Summarizing categorical variables:** Bar chart, Pie chart, Spine plot
- **Continuous variables:** Histogram, Density plot, Boxplot, Dot plots (Wilkinson, Cleveland)
- **Time-series data:** Run chart
- **relation between two or more variables:** Mosaic plots, scatter plot
- **other:** QQ plot (distribution), Violin plot (mix of boxplot and histogram)

1.7 Ordinal Variables and Measures of Center

Ordinal variables represent categories with a natural order, but the exact distances between categories are unknown. While it is possible to calculate a numerical mean, it is generally not meaningful. Instead, the median or mode are preferred as measures of central tendency, since they respect the inherent order of the data without assuming equation.

2 Hypothesis Testing Fundamentals

2.1 General Process

1. Define research question
2. Choose appropriate test
3. State hypotheses:

$$H_0 : \text{Null hypothesis (default assumption)} H_1 : \text{Alternative hypothesis (claim being tested)}$$

4. Set significance level α (e.g., 0.05)
5. Check assumptions
6. Compute test statistic
7. Make decision:
    - H_1 if $p \leq \alpha$, Reject H_0
    - H_0 if $p > \alpha$, fail to reject H_0

2.2 Type I and Type II Errors

In hypothesis testing, errors can occur because decisions are made using sample data rather than the entire population.

- A **Type I error** occurs when we *reject a true null hypothesis.* This means we detect an effect or difference that does not actually exist.
- A **Type II error** occurs when we *fail to reject a false null hypothesis.* This means we miss a real effect or difference.

  Decision table:

Decision	H_0 **is true**	H_0 **is false**
Do not reject H_0	Correct decision	Type II error (β)
Reject H_0	Type I error (α)	Correct decision

  Key reminders:

- α is the probability of a Type I error (false positive).
- β is the probability of a Type II error (false negative).
- Reducing one type of error often increases the other.

3 One-Sample Tests

3.1 Choosing the Correct Test Variant and Interpreting p-values

When multiple versions of a test are available:

1. **Check assumptions:**
   - Normality: Shapiro-Wilk or visual inspection
   - Equality of variance: F-test or Levene's test
2. **Select test variant:**
   - Normal + equal variance: Student's t-test (var.equal = TRUE)
   - Normal + unequal variance: Welch's t-test (var.equal = FALSE)
   - Non-normal: Wilcoxon rank sum test
3. **Directional hypothesis:** Use one-sided p-value if hypothesis is *greater* or *less*, otherwise use two-sided p-value.

3.2 Directional vs Two-Sided Hypotheses

When performing hypothesis tests, it is important to choose the correct type of test based on the alternative hypothesis:

- **Two-sided (non-directional) test:** Use this when you are checking for *any difference* from the null hypothesis. Example: $H_1 : \mu \neq 10$ (the mean is not 10).
- **One-sided (directional) test:** Use this when you are testing for a difference in a specific direction. Examples:
  - $H_1 : \mu > 10$ (mean is greater than 10)
  - $H_1 : \mu < 10$ (mean is less than 10)
- The direction must be decided **before looking at the data**.
- Using a one-sided p-value without a pre-specified direction is invalid and can lead to misleading results.

Exam rule: If the alternative hypothesis does not explicitly state a direction, always use a **two-sided test**.

3.3 One-Sample t-Test (Parametric)

Purpose: The one-sample t-test is used to determine whether the mean of a single sample is significantly different from a known or hypothesized population mean.

When to Use:

- You have one continuous (interval/ratio) variable.
- You want to compare the sample mean to a theoretical or previously known population mean.
- Data are continuous and approximately normally distributed.
- Observations are independent of each other.
- Moderate skewness is acceptable if the sample size is large ($n > 30$).
- No extreme outliers that would heavily distort the mean.

Hypotheses:

$$H_0 : \mu = \mu_0 \quad \text{(the sample mean equals the population mean)}$$

$$H_1 : \mu \neq \mu_0 \quad \text{(the sample mean differs from the population mean)}$$

Interpretation of Results:

- H_0 **, fail to reject** H_0**:** There is no evidence that the sample mean differs from the population mean. The sample mean is consistent with the hypothesized value.
- H_1**, reject** H_0**:** There is evidence that the sample mean is significantly different from the population mean. The difference is unlikely due to random chance.

Practical Tips / Exam Notes:

- Check for normality first. For small samples ($n < 30$), use Shapiro-Wilk, Jarque-Bera or visual methods (histogram, Q-Q plot).
- If normality is violated or data is heavily skewed, consider using a nonparametric alternative: **Wilcoxon Signed-Rank Test**.
- Example: Testing if average daily sales differ from 100 units. If $p = 0.02$ and $\alpha = 0.05$, we reject H_0: average sales are significantly different from 100.

3.4 Wilcoxon Signed-Rank Test (Nonparametric)

Purpose: The Wilcoxon Signed-Rank Test is a nonparametric alternative to the one-sample t-test. It tests whether the median of a single sample differs from a hypothesized value.

When to Use:

- Data are ordinal, or interval/ratio but do not meet the normality assumption.
- Sample is relatively small, or distribution is not symmetric enough for the t-test.
- You want to test differences around the median rather than the mean.
- Observations are independent.
- The distribution is symmetric around the median (important if you want the test to reflect a difference in the median).

Hypotheses:

$$H_0 : \text{Median} = \text{ME} \quad \text{(the median equals the hypothesized value)}$$

$$H_1 : \text{Median} \neq \text{ME} \quad \text{(the median differs from the hypothesized value)}$$

Interpretation of Results:

- H_0**, fail to reject** H_0**:** There is no evidence that the median differs from the hypothesized value. The sample is consistent with the expected median.

- H_1, **reject** H_0**:** There is evidence that the median is significantly different from the hypothesized value. The difference is unlikely to be due to random chance.

Practical Tips / Exam Notes:

- Use this test when your data violates the normality assumption required for a t-test.
- Interpretation focuses on medians, not means, which is especially useful for skewed distributions.
- If your distribution is not symmetric, the test reflects a shift in overall distribution rather than strictly the median.
- Example: Testing whether the median customer satisfaction score differs from 7. If $p = 0.03$ and $\alpha = 0.05$, reject H_0: the median satisfaction score is significantly different from 7.

3.5 One-Sample Sign Test (Nonparametric)

Purpose: The One-Sample Sign Test is a simple nonparametric method to test whether the median of a single sample differs from a hypothesized value. Unlike the Wilcoxon Signed-Rank test, it does not consider the magnitude of the differences, only their direction.

When to Use:

- Data are ordinal, or interval/ratio but do not meet the normality assumption.
- Distribution may be asymmetric or skewed.
- You want to test whether the number of values above and below a hypothesized median is as expected by chance.
- Observations are independent.
- Data are at least ordinal.
- The test is robust to outliers and skewed distributions.

Hypotheses:

$$H_0 : \text{Median} = \text{ME} \quad \text{(the median equals the hypothesized value)}$$

$$H_1 : \text{Median} \neq \text{ME} \quad \text{(the median differs from the hypothesized value)}$$

Interpretation of Results:

- H_0, **fail to reject** H_0**:** There is no evidence that the median differs from the hypothesized value. The observed number of values above and below the median is consistent with chance.
- H_1, **reject** H_0**:** There is evidence that the median differs from the hypothesized value. The distribution is systematically shifted above or below the hypothesized median.

Practical Tips / Exam Notes:

- Use this test when you cannot assume symmetry or normality and want a simple, robust method.
- The Sign Test has lower power than the Wilcoxon Signed-Rank Test because it ignores the magnitude of differences.
- Example: Testing whether the median daily number of customers is 100. If 8 of 10 days had more than 100 customers and 2 had fewer, you would calculate the p-value using a binomial model to see if this deviation is significant.

4 Two-Sample Tests

4.1 Independent Samples t-Test (Two-Sample t-Test)

Purpose: The Independent Samples t-Test is used to compare the means of a continuous dependent variable between two independent groups. It answers the question: *"Are the means* *of the two groups significantly different?"*

When to Use:

- You have two separate groups (e.g., men vs women, treatment vs control).
- Dependent variable is continuous (interval/ratio).
- Observations are independent (no pairing or repeated measures).
- Dependent variable is approximately normally distributed within each group.
- Variances of the two groups are equal (homoscedasticity) – test using an F-test.
- Moderate skewness is acceptable if distributions are unimodal and outlier-free.

Hypotheses:

$$H_0 : \mu_1 = \mu_2 \quad \text{(the group means are equal)}$$

$$H_1 : \mu_1 \neq \mu_2 \quad \text{(the group means are not equal)}$$

Interpretation of Results:

- **H_0, fail to reject H_0:** There is no evidence that the group means differ. Any observed difference could be due to random sampling variation.
- **H_1, reject H_0:** The difference in means is statistically significant. One group tends to have higher or lower values than the other.

Practical Exam Tips:

- Always check the equality of variances first. Using the wrong version of the t-test (pooled vs Welch) can lead to incorrect conclusions.
- This test is not appropriate for paired data (use paired t-test instead).
- If normality is violated, consider the nonparametric alternative: the Mann–Whitney U test (also called Wilcoxon rank-sum test).

4.2 Mann–Whitney U / Wilcoxon Rank-Sum Test (Nonparametric Two-Sample Test)

Purpose: This test is a nonparametric alternative to the independent samples t-test. It is used to compare two independent groups when the dependent variable is ordinal, not normally distributed, or when outliers may affect the t-test. It answers the question: *"Are the distributions (or medians) of the two groups different?"*

When to Use:

- Dependent variable is ordinal, interval, or ratio.
- Two independent groups (e.g., control vs treatment).
- Normality cannot be assumed or sample sizes are small.
- Observations are independent between groups.
- The distributions of the two groups have similar shape and spread if the goal is to test medians.
- Outliers can affect the interpretation if distributions are very different.

Hypotheses:

$$H_0 : \text{The medians of the two groups are equal (or distributions are similar)}$$

$$H_1 : \text{The medians of the two groups are not equal (or distributions differ)}$$

Interpretation of Results:

- **H_0, fail to reject H_0:** There is no evidence that the medians or distributions of the two groups differ.

- H_1, **reject** H_0**:** The two groups have significantly different medians, or the distributions differ systematically.

  Practical Exam Tips:

- Use this test when the t-test assumptions (normality, no outliers) are violated.
- If the distributions of the groups are very different in shape or spread, the test should be interpreted as a difference in distributions, not strictly medians.
- It is suitable for small or moderate sample sizes and robust to outliers compared to the t-test.

4.3 Paired t-Test (Parametric, Dependent Samples)

Purpose: The paired t-test is used to compare the means of two related groups. It answers the question: *"Is there a significant difference in the means of the same subjects measured under two conditions or at two time points?"*

When to Use:

- Measurements are paired or dependent (e.g., before-and-after studies, matched subjects).
- Dependent variable is continuous (interval/ratio).
- Differences between paired observations are approximately normally distributed.
- Observations within each pair are dependent, but pairs are independent of each other.
- The differences between paired measurements are approximately normally distributed.
- No significant outliers in the differences; moderate skewness is permissible if the sample is not very small.

Hypotheses:

$$H_0 : \mu_d = 0 \quad \text{(mean difference is zero, no effect)}$$

$$H_1 : \mu_d \neq 0 \quad \text{(mean difference is not zero, significant effect)}$$

Interpretation of Results:

- H_0, **fail to reject** H_0**:** No evidence of a significant difference between the paired measurements. The treatment or condition did not have a measurable effect.
- H_1 **reject** H_0**:** There is a statistically significant difference between the paired measurements. The treatment or condition had a measurable effect.

Practical Exam Tips:

- Ideal for pre-post designs, repeated measures on the same subject, or matched pairs experiments.
- Always check the distribution of differences before applying the t-test; if severely non-normal, consider the Wilcoxon signed-rank test.

4.4 Wilcoxon Paired Signed-Rank Test (Nonparametric, Paired Data)

Purpose: The Wilcoxon signed-rank test is the nonparametric alternative to the paired t-test. It tests whether the distribution of differences between paired observations is symmetric around zero. This is useful when the differences are not normally distributed or when the data are ordinal.

When to Use:

- Data are paired or dependent (e.g., pre-post measurements, matched subjects).
- Dependent variable is ordinal, interval, or ratio.
- Differences between pairs are roughly symmetric, but normality is not required.
- Small to moderate sample sizes, or when outliers may distort a parametric t-test.
- Observations are paired and differences are independent across pairs.

- Differences are symmetrically distributed around the median.
- No extreme outliers in the differences.

Hypotheses:

$$H_0 : \text{The distribution of differences is symmetric around 0}$$

$$H_1 : \text{The distribution of differences is not symmetric around 0}$$

Interpretation of Results:

- **H_0, fail to reject H_0:** No evidence of asymmetry; the differences are consistent with being centered around zero. Suggests no treatment effect or systematic change.
- **H_1, reject H_0:** Evidence that differences are not symmetric around zero; there is a significant effect of the treatment or condition.

Practical Exam Tips:

- Symmetrical data - no skewness, if there is skewness, consider the paired sign test instead

4.5 Paired Sign Test

Purpose: The Paired Sign Test is a simple nonparametric test used to determine whether the median difference between paired observations is zero. It is ideal when the assumptions of the paired t-test (normality of differences) are not met.

When to use:

- Data are paired or matched (e.g., before-and-after measurements on the same subjects).
- Dependent variable is ordinal or continuous.
- Differences between pairs may not be normally distributed or have outliers.
- You are interested in testing whether the typical change (median difference) is zero.

Hypotheses:

$$H_0 : \text{Median difference} = 0 \quad \text{vs.} \quad H_1 : \text{Median difference} \neq 0$$

Interpretation:

- **H_0, fail to reject H_0:** There is no evidence of a systematic median change.
- **H_1, reject H_0:** The median difference is significantly different from zero.

Tips / Exam Context:

- Use when the paired t-test is not appropriate due to non-normality or outliers.
- Does not assume symmetry of differences (unlike Wilcoxon paired signed-rank test, which is more powerful if symmetry holds).
- Focus on the direction of change rather than magnitude.

5 ANOVA and Related Tests

5.1 One-Way ANOVA (Analysis of Variance)

Purpose: One-Way ANOVA is used to test whether there are significant differences between the means of three or more independent groups. It extends the independent samples t-test when there are more than two groups.

When to Use:

- Dependent variable (DV) is continuous (interval/ratio).
- Independent variable (IV) is categorical with 3 or more levels/groups.

- Comparing means across multiple groups, e.g., exam scores across 4 different teaching methods.
- Residuals (differences between observed values and group mean) are approximately normally distributed - shapiro wilk test.
- Homogeneity of variances (equal population variances across groups). Tests: **Levene's** (less sensitive to departures from normality), **Bartlett's** (The data must be normally distributed), or **Fligner-Killeen** (non-parametric test which is very robust against departures from normality); where H0: all populations variances are equal and H1: at least two of them differ.
- Observations are independent.

  Hypotheses:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

$$H_1 : \text{At least one group mean differs from the others}$$

Interpretation of Results:

- **H_0, fail to reject H_0:** No significant evidence that group means differ; observed differences could be due to random variation.
- **H_1, reject H_0:** At least one group mean differs significantly; further post-hoc analysis needed to identify which groups are different.

Post-Hoc Analysis (Multiple Comparisons):

- Used when H_0 is rejected to find which specific group means differ.
- Common methods: Tukey, Scheffé, Bonferroni, Holm, Sidak.
- Adjust p-values to control family-wise error rate (FWER) or false discovery rate (FDR).
- Example: If 4 groups are compared, Tukey's test will compare all 6 pairwise combinations with adjusted confidence intervals.

Practical Exam Tips:

- Always check assumptions first: normality of residuals (shapiro-wilk), equality of variances.
- ANOVA tells you if a difference exists but not where it is — post-hoc tests are crucial.
- If variances are unequal, consider Welch's ANOVA.
- If data are ordinal or assumptions are violated, use Kruskal–Wallis test instead.

5.2 One-Way ANOVA with Blocks

Purpose: One-way ANOVA with blocks is used to compare treatment means *while controlling for other sources of variability* that could affect the results. The "block" is a way to group similar experimental units so that differences within a block are minimized, allowing for a clearer comparison between treatments.

What is a block? A **block** is a set of experimental units that are similar in some way that might influence the outcome, but that you are not testing. By grouping similar units together, you "remove" this nuisance variability from the comparison of treatments.

Examples:

- In an agricultural experiment, plots of land might differ in soil quality. Each plot with similar soil can be a block.
- In a medical study, patients might be blocked by age or gender to account for differences in response.
- In a taste test, testers could be blocked by their prior experience to reduce variability in ratings.

How it works:

- Each block receives all treatments (or as many as possible).
- Differences between treatments are then compared within each block.
- Blocking reduces the "noise" caused by the variability between blocks, making it easier to detect real differences between treatments.

Key points:

- The main question is still about the *treatment effect*.
- Block effects are used to account for variability but are usually not interpreted themselves.
- Sometimes blocks are treated as *random* (random blocks), meaning the specific identity of the blocks is not important, only the fact that they help control variability.

5.3 Fixed, Random, and Mixed Effects

In ANOVA and related models, factors can be treated as **fixed** or **random** depending on how their levels are chosen and how the results are interpreted.

- **Fixed effects:** The levels included in the study are the *only ones of interest.* Conclusions apply only to these specific levels. *Examples:* specific cities, specific treatments, specific job categories.
- **Random effects:** The levels are viewed as a *random sample* from a larger population of possible levels. We are not interested in individual levels, only in how much variability they introduce. *Examples:* subjects, batches, machines, blocks.
- **Mixed effects:** The model contains both fixed and random effects. Typically, treatments are fixed and blocks or subjects are random.

Random blocks (conceptual explanation): When blocks are treated as random, we assume the observed blocks represent a random sample from a larger population of blocks. The goal is not to compare specific blocks, but to account for block-to-block variability.

Inference with random blocks:

- Individual block effects are *not* tested or interpreted.
- Blocks help reduce unexplained variability in the data.
- Conclusions about treatments generalize beyond the specific blocks observed.

Exam note: Random effects influence how variability is modeled and estimated; they do **not** change which means are being compared.

5.4 Types of Sum of Squares in ANOVA

When fitting an ANOVA model, different definitions of *sum of squares* answer slightly different questions about factor effects. The most common types are Type I, Type II, and Type III.

- **Type I (Sequential):** Effects are tested *in the order they are entered into the model.* Each effect explains whatever variation remains after the previous effects.
  - Results depend on the order of the factors.
  - In unbalanced designs, changing the order can change the conclusions.
  - Rarely used for final inference.
- **Type II:** Each *main effect* is tested after adjusting for the other main effects, but *ignoring interactions.*
  - Appropriate when interaction effects are absent or not of interest.
  - More powerful than Type III when no interactions exist.
  - Not valid if important interactions are present.
- **Type III:** Each effect is tested after adjusting for *all other effects*, including interactions.

– Does not depend on the order of factors.
– Standard choice when interactions are included.
– Common default in statistical software.

Exam rules:

- **Balanced design** ⇒ all sum of squares types give the same results.
- **Unbalanced design with interactions** ⇒ use **Type III**.
- **No interaction present** ⇒ Type II is often preferred.

5.5 ANOVA Assumptions Tests

Purpose: Before performing ANOVA, it is crucial to check whether its key assumptions are met. Violations can invalidate the test results or reduce power.

Assumption 1: Normality of residuals

- ANOVA assumes that the residuals (differences between observed values and group means) are normally distributed.
- Common tests:
  - **Shapiro-Wilk Test:** Tests if data comes from a normal distribution.
  - **Jarque-Bera Test:** Tests for skewness and kurtosis to assess normality.
- Practical guidance:
  - If residuals are severely non-normal, consider a nonparametric alternative (e.g., Kruskal-Wallis test).

Assumption 2: Equality of variances (Homoscedasticity)

- ANOVA assumes that the variances across groups are equal. Unequal variances can inflate Type I error.
- Common tests:
  - **Bartlett's Test:** Sensitive to departures from normality; best when data are normal.
  - **Levene's Test:** Less sensitive to non-normality; widely used in practice.
  - **Fligner-Killeen Test:** Nonparametric, robust against non-normality.
- Practical guidance:
  - If variances are unequal, consider Welch's ANOVA or a nonparametric test.

Interpretation of Assumption Tests:

- **Normality Test:**
  - H_0**, fail to reject** H_0**:** residuals approximately normal → OK to proceed.
  - H_1**, reject** H_0**:** residuals not normal → consider transformation or nonparametric alternative.
- **Variance Equality Test:**
  - H_0**, fail to reject** H_0**:** variances approximately equal → ANOVA assumption met.
  - H_1**, reject** H_0**:** unequal variances → consider Welch's ANOVA or robust methods.

5.6 Post-Hoc Comparisons

Purpose: A significant ANOVA result tells us that *at least one group mean differs*, but it does not tell us *which* groups are different. Post-hoc tests are used **after rejecting the ANOVA null hypothesis** to identify the specific group differences, while controlling for errors caused by making many comparisons.

Why adjustments are needed: When many hypothesis tests are performed, the chance of finding a "significant" result just by chance increases. Post-hoc methods correct for this problem, known as **multiple testing**.

- **Family-Wise Error Rate (FWER):** Controls the probability of making *at least one* false positive across all comparisons.
- **False Discovery Rate (FDR):** Controls the *proportion* of false positives among all rejected hypotheses.

When to use post-hoc tests:

- Only after the overall ANOVA F-test is significant.
- When comparing pairs of group means or groups against a control.
- The choice of method depends on how conservative you want to be and how many comparisons are made.

Common Post-Hoc Methods:

- **Tukey's HSD (Honest Significant Difference):**
  - Designed for *all pairwise comparisons*.
  - Controls the family-wise error rate.
  - Assumes equal variances and independent observations.
  - Always two-sided.
- **Scheffé:**
  - Allows very general comparisons (not just pairwise).
  - Very conservative, especially with few groups.
  - Often used when comparisons were not planned in advance.
- **Bonferroni:**
  - Adjusts the significance level by dividing α by the number of comparisons.
  - Simple and widely applicable.
  - Can be overly conservative when many tests are performed.
- **Holm / Hochberg:**
  - Stepwise procedures that control the family-wise error rate.
  - Less conservative and more powerful than Bonferroni.
  - Commonly used when many comparisons are made.
- **Benjamini–Hochberg (FDR):**
  - Controls the false discovery rate instead of FWER.
  - Less conservative than FWER-based methods.
  - Especially useful when testing many hypotheses simultaneously.

Interpreting post-hoc results:

- Compare the *adjusted* p-values to the chosen significance level (usually 0.05).
- A significant result means there is evidence of a difference between the two groups being compared.

Exam tips:

- Post-hoc tests are only meaningful after a significant ANOVA.
- Always state which post-hoc method was used and why.
- Remember: Tukey and Bonferroni control FWER; Benjamini–Hochberg controls FDR.

5.7 Interpreting Compact Letter Displays (CLD)

Post-hoc test results are often summarized using a **Compact Letter Display (CLD)**. A CLD assigns letters to each group to quickly show which group means are statistically different.

How a CLD works:

- Each group is assigned one or more letters.
- **Groups that share at least one letter are not significantly different**.
- **Groups that share no letters are significantly different**.

Example CLD:

Group	Grouping Letter(s)
A	a
B	ab
C	b

How to interpret this example:

- Group A and Group C do **not** share a letter, so they are significantly different.
- Group B shares a letter with both A and C, so it is **not significantly different** from either.

Key reminder: Letters indicate *statistical similarity*, not the size or direction of differences.

5.8 Kruskal-Wallis Test (Nonparametric)

Purpose: Kruskal-Wallis is the nonparametric alternative to one-way ANOVA. It tests whether multiple independent groups come from the same distribution, focusing on medians when distributions are similarly shaped.

When to use:

- Dependent variable is ordinal, or interval/ratio but not normally distributed.
- Comparing 2 or more independent groups.
- Group distributions are similar in shape (otherwise the test reflects differences in distributions, not strictly medians).
- Observations are independent.
- Values can be ranked across all groups.
- For a test of medians specifically, distributions should have similar shape and spread.

Hypotheses:

$$H_0 : \text{Medians of all groups are equal}, \quad H_1 : \text{At least one group median differs}$$

Interpretation:

- H_1**, reject** H_0**:** there is a statistically significant difference in medians (or distributions) among groups.
- H_0**, fail to reject** H_0**:** no evidence of difference in medians.
- Note: If group distributions differ in spread or shape, significant results may reflect general distribution differences, not medians specifically.

Follow-up / Post-hoc Analysis:

- **Dunn test:** pairwise comparisons with p-value adjustments (e.g., Bonferroni, Holm).
- **Conover-Iman test:** more powerful alternative, only performed if Kruskal-Wallis is significant.
- Always adjust for multiple comparisons to control Type I error.

Exam Tips:

- State clearly whether you are testing medians or general distribution differences.
- Report both the overall H-statistic and pairwise post-hoc results if needed.

5.9 Mood's Median Test

Purpose: Mood's Median Test is a nonparametric method used to compare the *medians* of two or more independent groups. It tests whether the groups come from populations with the same median value.

When to use Mood's Median Test:

- The groups are independent.
- The data contain strong outliers or are heavily skewed.
- The primary interest is in comparing medians, not overall distribution shapes.

How the test works:

- Combine all observations and compute a single overall median.
- For each group, count how many observations fall above and below this median.
- Use a chi-square test to determine whether these counts differ across groups.

Hypotheses:

$$H_0 : \text{All population medians are equal}$$

$$H_1 : \text{At least one population median is different}$$

Important notes:

- Mood's Median Test is very robust to extreme outliers.
- It is generally less powerful than the Kruskal–Wallis test.
- Prefer this test when outliers make rank-based methods unreliable.

5.10 Friedman Test (Repeated Measures Nonparametric)

Purpose: The Friedman test is a nonparametric alternative to repeated measures ANOVA. It evaluates whether the distributions (or medians) of a dependent variable differ across multiple related groups or treatments.

When to use:

- The same subjects (or experimental units) are measured under multiple conditions (repeated measures).
- Dependent variable is ordinal, or continuous but not normally distributed.
- You have more than two conditions/treatments.
- Observations within each block (subject) are related, but blocks are independent of each other.
- Measurements can be ranked across treatments within each block.
- Distributions across treatments should have similar shapes if you interpret results as differences in medians.

Hypotheses:

$$H_0 : \text{Medians (or distributions) are equal across all treatments,}$$

$$H_1 : \text{At least one treatment differs}$$

Interpretation:

- **H_1, reject H_0:** at least one treatment differs significantly.
- **H_0, fail to reject H_0:** no evidence of differences among treatments.
- Be cautious: the test indicates differences exist but not which treatments differ.

Follow-up / Post-hoc Analysis:

- **Pairwise Sign Test:** compare treatments two at a time.
- Adjust p-values for multiple comparisons (e.g., Bonferroni, Holm) to control Type I error.

Exam Tips:

- Always check that your data are truly paired (repeated measures).

5.11 Quade Test

Purpose: The Quade test is a nonparametric alternative to a blocked (repeated-measures) ANOVA. It is used when treatments are compared across blocks and the size of block effects differs substantially. The test gives more weight to blocks where treatment differences are larger.

When to use the Quade test:

- Data are collected in blocks (e.g., subjects, judges, locations).
- The same treatments are observed within each block.
- Block-to-block variability is not uniform.
- A nonparametric method is needed.

How it differs from the Friedman test:

- Friedman treats all blocks as equally important.
- Quade weights blocks by how much variability they show.
- Quade is more powerful when treatment effects are stronger in some blocks than others.

Hypotheses:

$$H_0 : \text{All treatments have the same effect across blocks}$$

$$H_1 : \text{At least one treatment has a different effect across blocks}$$

Key note: The Quade test focuses on detecting treatment differences while accounting for unequal block importance, making it a stronger alternative to Friedman in some settings.

5.12 Two-Way ANOVA (Factorial)

Purpose: Two-way ANOVA is used to study the effect of *two independent categorical factors* on a continuous outcome. It also allows testing for *interaction effects*, i.e., whether the effect of one factor depends on the level of the other factor.

When to use:

- One continuous dependent variable (DV) and two categorical independent variables (factors), each with 2 or more levels.
- You want to test *main effects* of each factor and their *interaction effect*.
- Observations are independent across groups.
- Residuals are roughly normally distributed.
- Variances are similar across groups (homogeneity of variance).
- Balanced or reasonably balanced design is recommended for clear interaction interpretation.

Hypotheses:

- **Main effect of Factor A:**

$$H_0 : \text{All levels of A have the same mean}$$

$$H_1 : \text{At least one mean differs}$$

- **Main effect of Factor B:**

$$H_0: \text{All levels of B have the same mean}$$

$$H_1: \text{At least one mean differs}$$

- **Interaction effect (A $\times$ B):**

$$H_0: \text{The effect of A is the same at all levels of B}$$

$$H_1: \text{The effect of A depends on the level of B}$$

Interpretation:

- Rejecting H_0 for a main effect means that factor significantly affects the DV overall.
- Rejecting H_0 for interaction means that the effect of one factor *depends on the level of the other factor.*
- Significant interaction complicates main effect interpretation — consider visualizing with an interaction plot or testing *simple main effects* for clarity.

Follow-up / Post-hoc analysis:

- If a main effect is significant, use post-hoc tests such as Tukey, Bonferroni, or Holm to identify which levels differ.
- If the interaction is significant, focus on simple main effects rather than overall main effects.

5.13 Scheirer–Ray–Hare Test

Purpose: The Scheirer–Ray–Hare test is a nonparametric alternative to a two-way ANOVA. It is based on ranking the data and is useful when the assumptions of ANOVA (normality, equal variances) are not met.

What it tests:

- **Main effect of Factor A:** whether the levels of A differ in their central tendency.
- **Main effect of Factor B:** whether the levels of B differ in their central tendency.
- **Interaction (A $\times$ B):** whether the effect of one factor depends on the level of the other factor.

When to use:

- Two independent categorical factors and a continuous (or ordinal) outcome.
- Data violate ANOVA assumptions (non-normal, unequal variance, or outliers).
- Interested in both main effects and interaction effects.

Limitations:

- The interaction test has low statistical power — it may fail to detect true interactions.
- Main effects are generally more reliable than the interaction in this test.
- Interpret interaction results cautiously.

Interpretation:

- Significant main effect: the factor influences the outcome.
- Significant interaction: the effect of one factor may depend on the other factor, but confirm with caution or additional methods.

5.14 Coefficient of Determination

The coefficient of determination measures how well a model explains the variability in the outcome. It is a way to quantify model fit.

- **R^2:** The proportion of total variance in the dependent variable that is explained by the model. *Example:* $R^2 = 0.7$ *means the model explains 70% of the variation.*
- **Adjusted R^2:** Adjusts R^2 for the number of predictors in the model. Prevents overestimating fit when adding unnecessary variables. Useful for comparing models with different numbers of predictors.
- **Marginal R^2:** In mixed-effects models, measures the proportion of variance explained by *fixed effects only.*
- **Conditional R^2:** In mixed-effects models, measures the proportion of variance explained by *both fixed and random effects.*

Exam tip: Use **adjusted** R^2 when comparing models that have different numbers of predictors, because it accounts for model complexity.

5.15 Repeated Measures ANOVA

Purpose: Repeated Measures ANOVA is used to analyze data where the *same subjects are measured multiple times.* It allows you to test for differences between conditions or time points while accounting for the fact that measurements from the same subject are correlated.

When to use:

- You measure the same subjects under different conditions or at multiple time points.
- You want to compare means across conditions while controlling for individual differences.
- Observations within subjects are correlated; Repeated Measures ANOVA accounts for this correlation.
- The dependent variable is continuous and roughly normally distributed.

Why use it:

- Controls for variability due to individual differences, increasing statistical power.
- Reduces error compared to treating repeated measurements as independent.
- Allows testing for overall effects of time or condition (main effects) and interactions if there are multiple factors.

Hypotheses:

$$H_0 : \text{The mean outcome is the same across all conditions or time points}$$

$$H_1 : \text{At least one condition or time point has a different mean outcome.}$$

Key points for analysis:

- Check assumptions: sphericity (variance of differences between conditions is equal). Violations may require corrections (e.g., Greenhouse-Geisser).
- Significant results indicate overall differences, but post-hoc tests may be needed to see which conditions differ.
- Often visualized with line plots showing each subject's trajectory over time or across conditions.

5.16 Correlation Structures

In repeated measures or longitudinal data, observations from the same subject are correlated. Different correlation structures describe how these correlations are modeled:

- **Compound symmetry (CS):** All repeated measurements have the same correlation with each other. *Example:* all time points are equally correlated.
- **Autoregressive AR(1):** Correlation decreases as measurements are further apart in time. *Example:* measurements one day apart are more correlated than those a week apart.

- **Unstructured (UN):** Each pair of measurements has its own correlation; the most flexible approach. *Example:* no assumptions about how correlations change over time.

  Sphericity: An important assumption in repeated measures ANOVA.

- Assumes that the variances of differences between all pairs of repeated measurements are equal.
- Tested using **Mauchly's test of sphericity**.
- If sphericity is violated, adjustments like **Greenhouse–Geisser** or **Huynh–Feldt** are applied to correct the degrees of freedom.

  Exam tip: - Compound symmetry implies sphericity, but AR(1) or unstructured do not. - Always check Mauchly's test in repeated measures ANOVA and apply corrections if needed.

5.17 Repeated Measures vs Randomized Block ANOVA

Both repeated measures and randomized block designs deal with grouping or blocking, but they have important differences in purpose and modeling.

- **Randomized Block ANOVA:**
  - Blocks are considered *nuisance variables*—they introduce variability that we want to account for.
  - The main interest is in comparing *treatment effects*.
  - Block-to-block correlation is not explicitly modeled; blocks simply reduce unexplained variation.
- **Repeated Measures ANOVA:**
  - Subjects are treated as *random effects*.
  - Observations from the same subject are correlated, and this *within-subject dependence* is explicitly modeled.
  - Focus is still on treatment effects, but the design accounts for how repeated measurements are related.

  Key differences:

- Randomized block ANOVA controls variability from blocks but does not model correlations.
- Repeated measures ANOVA models correlations between repeated observations on the same subject.
- Use repeated measures when the same experimental units are measured multiple times.

6 Tests for Nominal Data

6.1 Goodness-of-Fit Tests (Chi-square, G-test)

Purpose: Goodness-of-Fit tests are used to determine whether the observed frequencies of a categorical variable match a theoretical or expected distribution. These tests help check if data behaves as expected under a certain model.

When to use:

- Your variable is nominal (categorical) with 2 or more categories.
- You know the expected proportions for each category (e.g., theoretical distribution, equal probability, or historical data).
- Observations are independent.

  Hypotheses:

$$H_0 : \text{Observed proportions match the expected proportions}$$

$$H_1 : \text{Observed proportions differ from the expected}$$

Test Options:

- **Chi-square goodness-of-fit:** Most common; compares observed and expected counts using
$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$
- **G-test (likelihood ratio test):** Alternative to chi-square; uses
$$G = 2 \sum O_i \ln \frac{O_i}{E_i}$$

Both approximate a chi-square distribution under H_0.

Interpretation:

- **H_1, reject H_0:** the observed distribution significantly differs from the expected.
- **H_0, fail to reject H_0:** no evidence that the observed distribution differs from expectations.

Tips / Exam Context:

- Good for testing "fairness" or uniformity, e.g., dice rolls, survey responses, or expected demographics.
- G-test and chi-square are asymptotic; with very small sample sizes, exact tests may be needed.

6.2 Bernoulli Model and Estimation of p

Purpose: A Bernoulli random variable models a single trial with two possible outcomes: success (1) or failure (0). The goal is to estimate the probability of success, p, from observed data.

Model: For a single trial X:

$$P(X = x) = p^x (1 - p)^{1-x}, \quad x \in \{0, 1\}$$

If we have n independent trials $x_1, x_2, \ldots, x_n$, the probability of observing the data (likelihood) is highest when:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Interpretation / Intuition:

- The *sample proportion* of successes is the **maximum likelihood estimator (MLE)** of p.
- This value of $\hat{p}$ is the one that makes the observed data most "likely" under the Bernoulli model.
- Using $\hat{p}$ in a goodness-of-fit test ensures the best possible fit of the model to your observed data.
- In practical terms: just count the successes and divide by the total number of trials.

6.3 Association Tests for Nominal Variables

Purpose: Association tests determine whether two categorical (nominal) variables are related. They check if the distribution of one variable depends on the levels of the other.

When to use:

- You have two categorical variables and want to test if they are independent or associated.
- Data can be *unpaired* (different subjects for each variable) or *paired* (same subjects measured twice or more).

Hypotheses:

$$H_0 : \text{The variables are independent}$$

$$H_1 : \text{There is an association between the variables}$$

Tests for unpaired nominal data:

- **Chi-square test of independence:** Compares observed counts in a contingency table to expected counts under independence. Suitable for moderate to large samples.
- **Fisher's Exact Test:** Use when sample sizes are small or expected counts are low (especially for 2×2 tables). Provides an exact p-value.

Tests for paired nominal data:

- **McNemar's test:** Used when the same subjects are measured twice (e.g., before/after treatment, matched pairs).
  - Checks whether changes in one direction are balanced by changes in the other direction.
  - H_0: the probability of change in either direction is equal.
  - H_1: there is a significant change or difference between paired responses.
- **Cochran's Q test:** Generalization of McNemar's test for more than two repeated measurements (k ¿ 2).
  - H_0: proportions are the same across all measurements.
  - H_1: at least one measurement differs in proportion.

Interpretation:

- Reject H_0 (H_1 is true): there is a significant association (unpaired) or change (paired).
- Fail to reject H_0: no evidence of association or change.

Key Decision Rule (Exam Critical):

- Same subjects measured twice or more ⇒ McNemar's or Cochran's Q test.
- Different subjects in each group ⇒ Chi-square or Fisher's Exact test.

Tips / Exam Context:

- Always determine whether the data are paired or unpaired before selecting a test.
- Chi-square provides an approximate p-value; Fisher's Exact is exact for small samples.
- McNemar's test is specifically for repeated measures on the same subjects; Cochran's Q extends this for multiple time points or conditions.

6.4 Stratified Association: Cochran–Mantel–Haenszel (CMH) Test

Purpose: The CMH test evaluates whether there is an association between two categorical variables while *controlling for a third stratifying variable*. It accounts for confounding by examining associations within each stratum and combining the results.

When to use:

- Both exposure and outcome are binary or categorical.
- Data are divided into *strata* by a potential confounder (e.g., gender, study site).
- Each stratum has the same contingency table structure.
- You want to test association *while adjusting for the stratifying variable*, rather than pooling all data.

Hypotheses:

$$H_0 : \text{The two variables are independent within each stratum}$$

$$H_1: \text{There is an association within at least one stratum}$$

Key intuition: - CMH combines information across strata to give an overall test of association, while accounting for stratification. - It is preferred over a simple pooled chi-square when a confounder might distort the association.

Exam note: - CMH tests *conditional independence* within strata. - Always check that table structures are consistent across strata before applying the test.

7 Association Measures

Purpose: Association measures quantify *how two variables are related.* They tell you:

- **Strength:** How strong the relationship is (effect size).
- **Direction:** In some cases, whether the relationship is positive, negative, or monotonic.

Why we use them:

- To summarize the relationship numerically.
- To help predict or explain one variable from another.
- To complement hypothesis tests: significance (p-value) tells us if a relationship exists, effect size tells us how meaningful it is.

Key Exam Rule: Strength vs Significance

- **Statistical significance (p-value):** *"Is there any detectable association?"*
- **Effect size:** *"How strong is the association?"*
- Always report **both**: even if $p > \alpha$, effect size shows practical relevance.

Tip: Think of p as "is there a signal?" and effect size as "how loud is the signal?"

7.1 Nominal–Nominal Variables

Purpose: Measure the association between two categorical variables.

- **2×2 tables: Phi coefficient (ϕ)**
  - Range: 0–1; 0 = no association, 1 = perfect association.
  - Measures the strength of the relationship between two binary variables.
- **n×m tables:**
  - **Cramér's V:** Generalizes ϕ for larger tables; 0 = no association, 1 = strong association.
  - **Goodman–Kruskal Lambda:** Measures predictive power; shows how much knowing the independent variable reduces errors in predicting the dependent variable.
- **Statistical tests:**
  - **Chi-square test of independence:** Standard test for association.
  - **Fisher's exact test:** Use when expected counts are small (especially in 2×2 tables).
- **Usage:** Both variables must be categorical. Always report **strength (effect size)** and **statistical significance**.

7.2 Nominal–Ordinal Variables

Purpose: Assess the association between a nominal variable (categories without order) and an ordinal variable (ordered categories).

- **Freeman's Theta:** Measures the strength of association between a nominal variable and an ordinal variable. Useful for understanding how well the nominal variable explains differences in the ordinal variable.

- **Cochran–Armitage Test for Trend:** Tests whether proportions of the nominal variable show a consistent *increasing or decreasing trend* across the levels of the ordinal variable. Best for 2×k tables (two nominal categories across multiple ordinal levels).
- **Interpretation:**
  - Significant test → there is a consistent trend (increasing or decreasing) across the ordinal levels.
  - Non-significant test → no clear trend.

Exam Tip: Always check which variable is ordinal and which is nominal; the trend test only works if the ordinal variable is ordered correctly.

7.3 Ordinal–Ordinal Variables

Purpose: Assess the strength and direction of the relationship between two ordinal variables (both with natural order).

- **Kendall's Tau-b:** Measures rank concordance while handling ties.
  - Range: [-1, 1]
  - Positive → more concordant pairs (agreement in ranks)
  - Negative → more discordant pairs (opposite ranks)
- **Goodman–Kruskal Gamma:** Ignores ties; often produces a stronger association estimate than Tau-b.
- **Somers' D:** Asymmetric measure: evaluates how well X predicts Y (directional).
- **Tests:** Linear-by-linear association test, permutation or exact tests for small samples.

Trend Interpretation:

- Positive statistic → increasing trend between variables
- Negative statistic → decreasing trend between variables
- Significant linear-by-linear association → a monotonic trend exists across the ordinal levels

Exam Tip: Always check for ties; choose Kendall Tau-b if ties are present, Gamma if you want a simpler estimate ignoring ties, and Somers' D when predicting one variable from another.

7.4 Continuous–Continuous Variables

Purpose: Assess the strength and direction of the relationship between two continuous variables.

- **Pearson correlation (r):** Measures *linear association.*
  - Sensitive to outliers.
  - Assumes approximate bivariate normality.
  - Range: [-1, 1]; positive → variables increase together, negative → one increases while the other decreases.
- **Spearman rank correlation (ρ):** Measures *monotonic association* using ranks.
  - Robust to outliers.
  - Can detect non-linear but monotonic relationships.
- **Kendall Tau:** Measures rank concordance.
  - Handles ties better than Spearman.
  - Robust to outliers.
  - Also range [-1, 1]; positive = concordant ranks, negative = discordant ranks.

- **Hypotheses:**

$$H_0 : \text{No association } (\rho = 0), \quad H_1 : \text{Association exists } (\rho \neq 0)$$

- **Interpretation:**
  - Sign → direction of association
  - Magnitude → strength of association
  - Always check scatterplots for outliers or non-linear patterns before interpreting.

Exam Tip:

- Use Pearson for linear, normally distributed data with few outliers.
- Use Spearman or Kendall for monotonic, non-linear, or skewed data.
- Always report both effect size (correlation coefficient) and p-value.

7.5 Weak or Non-Significant Correlations

Definition: A correlation is considered weak or non-significant when there is little evidence of a meaningful relationship between two variables.

- If $|r| < 0.2$ **and** $p > \alpha$, the association is negligible.
- Report as: "No statistically significant association detected."
- The direction of the correlation (positive/negative) is not meaningful when the association is non-significant.

Exam Tips:

- Always identify variable types first: nominal, ordinal, or continuous.
- Choose the correlation measure based on scale and study design.
- Report both **effect size** (strength) and **significance** (p-value).
- Use nonparametric measures (Spearman or Kendall) for skewed data, outliers, or tied ranks.

Correlation Selection Hierarchy:

- **Pearson:** Linear relationship, few outliers, approximate normality.
- **Spearman:** Monotonic but non-linear relationship, outliers present.
- **Kendall:** Many tied ranks or small sample size.

Variable 1	**Variable 2**	Association Measure / Test	Notes / Use
Continuous	**Continuous**	Pearson correlation (linear, normal) Spearman ρ (monotonic, non-normal) Kendall τ (monotonic, handles ties)	Measures strength & direction of relationship Pearson for linear & normal Spearman/Kendall for monotonic or skewed data
Continuous	**Ordinal**	Spearman ρ Kendall τ	Treat continuous as ranks if non-normal monotonic relationship
Continuous	**Nominal**	Point-biserial correlation (2 groups) t-test (2 groups) ANOVA (¿2 groups)	Nominal variable = grouping variable; compares means
Ordinal	**Continuous**	Spearman ρ Kendall τ	Monotonic relationship; ranks can be used for both
Ordinal	**Ordinal**	Spearman ρ Kendall τ-b Gamma / Somers' D Linear-by-linear association test	Measures monotonic association use Kendall τ-b for ties Somers' D for directional prediction
Ordinal	**Nominal**	Cochran–Armitage test for trend (2 nominal $\times$ k ordinal) Linear-by-linear association test Mann–Whitney U (2 groups) Kruskal–Wallis (¿2 groups)	Ordinal must be correctly ordered trend tests detect monotonic increase/decrease
Nominal	**Continuous**	t-test (2 groups) ANOVA (¿2 groups) Point-biserial correlation	Nominal variable = grouping variable
Nominal	**Ordinal**	Cochran–Armitage test for trend Linear-by-linear association test Mann–Whitney / Kruskal–Wallis	Ordinal = outcome variable; trend tests require order
Nominal	**Nominal**	Chi-square test of independence (unpaired) Fisher's exact test (small n, unpaired) McNemar test (paired) Cochran's Q test (k ¿ 2 repeated measures)	Check if data are **paired or unpaired** 2×2 vs n×m tables always report effect size (Phi, Cramér's V, Odds Ratio)

Table 1: Association measures and tests based on variable types