

VU LEARNING HUB

vulearninghub.xyz

HANDOUT

STA100

General Mathematics and Biostatistics

STA100 by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

COMPREHENSIVE STUDY HANDOUT

STA100

General Mathematics and Biostatistics

Lecture-wise study notes prepared for Virtual University of Pakistan students. Covers all major topics with objectives, key definitions, summaries and exam pointers.

10 study units

Table of Contents

Placeholder for table of contents

0

Unit 1: Number Systems, Sets and Functions

Learning Objectives

- Objective 1:** Classify numbers and perform set operations

Objective 2: Work with functions: domain, range, composition

Objective 3: Apply basic counting principles

Mathematics begins with numbers: natural, integer, rational (ratios p/q), irrational, real, and complex. Rational numbers have terminating or repeating decimals; irrationals like π do not. Interval notation and absolute value (distance from zero) organize the real line; inequalities solve by sign analysis.

Sets collect objects: union, intersection, complement, difference, with Venn diagrams visualizing and De Morgan's laws (complement of union = intersection of complements) simplifying. A function assigns each domain element exactly one image; the vertical-line test checks graphs. Composition ($f \circ g$)(x) = $f(g(x))$ chains functions; inverses undo them where they exist.

Counting underlies probability: the multiplication principle (m ways then n ways = mn), permutations (ordered, $P(n,r) = n!/(n-r)!$) and combinations (unordered, $C(n,r) = n!/r!(n-r)!$). Very important: distinguishing permutations from combinations by whether order matters is the key skill; half of counting errors come from this single confusion.

Key Definitions

Function is defined as a rule assigning each domain element exactly one value.

Permutation is defined as an ordered arrangement; $P(n,r) = n!/(n-r)!$.

Combination is defined as an unordered selection; $C(n,r) = n!/r!(n-r)!$.

Key Points to Remember

- Number types: N , Z , Q (repeating decimals), irrationals, R ; absolute value = distance.
- Set ops: union, intersection, complement; De Morgan's laws; Venn diagrams.
- Functions: vertical-line test, composition, inverses.
- Counting: order matters $\rightarrow$ permutation; order irrelevant $\rightarrow$ combination.

Unit Summary

Numbers, sets and functions provide the language; counting principles (permutations vs combinations) open the door to probability.

Exam Pointer

Very important: solve counting problems choosing permutation vs combination correctly; apply De Morgan's laws.

Unit 2: Algebra: Equations, Inequalities and Matrices

Learning Objectives

Objective 1: Solve linear and quadratic equations

Objective 2: Work with matrices and determinants

Objective 3: Solve systems of equations

Linear equations solve by isolating the variable; quadratics by factoring, completing the square, or the quadratic formula $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. The discriminant $b^2 - 4ac$ classifies roots: positive (two real), zero (one repeated), negative (complex pair). Inequalities need sign charts, with direction flipping when multiplying by negatives.

Matrices organize linear algebra: addition, scalar multiplication, and multiplication (rows by columns). The determinant of a 2x2 matrix $ad - bc$ tests invertibility; larger ones expand by cofactors. Matrix inverses solve systems via $x = A^{-1}b$.

Systems solve by substitution, elimination, or matrix methods (Gaussian elimination, Cramer's rule). Applied problems (mixture, investment, diet formulations in biostatistics) translate words to systems. Very important: the discriminant's root classification and 2x2 determinant computation are rapid exam marks; matrix multiplication order matters ($AB \neq BA$).

Key Definitions

Discriminant is defined as $b^2 - 4ac$, classifying quadratic roots.

Determinant is defined as the scalar $ad - bc$ for 2x2 matrices, testing invertibility.

System of equations is defined as multiple equations sharing unknowns, solved simultaneously.

Key Points to Remember

- Quadratics: formula + discriminant classification (two/one/complex roots).
- Matrices: addition, multiplication (order matters), 2x2 determinant and inverse.
- Systems: substitution, elimination, Gaussian elimination, Cramer.
- Word problems: translate carefully, then solve mechanically.

Unit Summary

Algebraic fluency, quadratics, inequalities, matrices and systems, provides the computational base for everything that follows.

Exam Pointer

Very important: classify roots via discriminant; invert 2x2 matrices; solve systems by elimination.

Unit 3: Limits, Differentiation and Applications

Learning Objectives

Objective 1: Evaluate limits

Objective 2: Differentiate using standard rules

Objective 3: Apply derivatives to optimization and rates

Limits capture approaching behavior: direct substitution first, then factoring/rationalizing for 0/0 forms. Continuity means the limit equals the value; most applied functions are continuous. The derivative is the limit of difference quotients: instantaneous rate of change, the slope of the tangent.

Differentiation rules: power, sum, product, quotient, and chain (for compositions). Higher derivatives give concavity via the second derivative. Standard derivatives (polynomials, exponentials, logs, trigonometric) combine through these rules; the chain rule is the most used and most erred.

Applications: optimization (set derivative to zero, test with second derivative), related rates, and marginal analysis in economics (marginal cost = derivative of cost). In biostatistics, growth models differentiate to find maximal growth rates. Very important: optimization word problems follow a fixed pipeline (model -> differentiate -> critical points -> verify); practice it.

Key Definitions

Limit is defined as the value a function approaches as its input approaches a point.

Derivative is defined as the instantaneous rate of change; limit of the difference quotient.

Critical point is defined as where the derivative is zero or undefined; candidate for extrema.

Key Points to Remember

- Limits: substitute, then resolve 0/0 by factoring/rationalizing.
- Rules: power, product, quotient, chain (most used); second derivative -> concavity.
- Optimization pipeline: model, differentiate, critical points, verify endpoints.
- Marginal analysis and growth-rate applications.

Unit Summary

Limits lead to derivatives, whose rules power optimization and rate analysis across science and economics.

Exam Pointer

Very important: differentiate composite functions (chain rule) and solve optimization word problems completely.

Unit 4: Integration and Its Uses

Learning Objectives

Objective 1: Integrate basic functions

Objective 2: Evaluate definite integrals

Objective 3: Apply integration to areas and totals

Integration reverses differentiation: the indefinite integral adds the constant C. Power rule (reversed), exponential, and $1/x \rightarrow \ln|x|$ cover most needs; substitution reverses the chain rule. The definite integral from a to b gives net accumulation: area under the curve when the function is positive.

The Fundamental Theorem connects the two: differentiate an accumulation and recover the function; evaluate definite integrals via antiderivatives at endpoints. Properties (linearity, additivity over

intervals, reversal flipping sign) simplify computation.

Applications: areas between curves, total from rate (integrate velocity for distance, marginal cost for total cost), and in biostatistics, probabilities as areas under density curves (previewing continuous distributions). Very important: substitution choice (let u be the inner function) is the key technique; area problems need correct top-minus-bottom setup.

Key Definitions

Antiderivative is defined as a function whose derivative is the given function; the indefinite integral.

Definite integral is defined as net accumulation from a to b ; area under positive curves.

Substitution is defined as reversing the chain rule by changing variables.

Key Points to Remember

- Indefinite: reverse the rules + C ; substitution for compositions.
- Definite: Fundamental Theorem, net accumulation/area.
- Setup matters: top minus bottom for areas between curves.
- Biostatistics link: probabilities as areas under curves.

Unit Summary

Integration accumulates: antiderivatives plus the Fundamental Theorem yield areas, totals and (later) probabilities.

Exam Pointer

Very important: integrate by substitution and set up area-between-curves integrals correctly.

Unit 5: Descriptive Biostatistics

Learning Objectives

- Objective 1:** Summarize data with tables and graphs
- Objective 2:** Compute measures of center and spread
- Objective 3:** Choose appropriate summaries for data types

Biostatistics applies statistics to biology and medicine. Data types matter: categorical (blood groups) versus numerical, discrete (counts) versus continuous (heights). Frequency tables, bar charts (categorical), histograms (continuous), and boxplots display distributions; choose the display matching the data type.

Center: mean (balance point, sensitive to outliers), median (middle, robust), mode (most frequent). Spread: range, interquartile range (IQR, robust), variance and standard deviation (average squared distance, then its root). The five-number summary (min, $Q1$, median, $Q3$, max) feeds boxplots; the $1.5 \times \text{IQR}$ rule flags outliers.

In medicine, summaries describe patient cohorts: mean blood pressure with standard deviation, median survival times (skewed data need medians). Very important: match the summary to the shape;

skewed data (income, survival) report median and IQR, symmetric data mean and SD. Examiners test this judgment.

Key Definitions

Standard deviation is defined as the root-mean-square distance of data from the mean.

Interquartile range is defined as $Q3 - Q1$; the spread of the middle 50%, robust to outliers.

Skewness is defined as asymmetry of a distribution; decides mean-vs-median choice.

Key Points to Remember

- Data types: categorical vs numerical, discrete vs continuous; match display to type.
- Center: mean (sensitive), median (robust), mode; spread: range, IQR, SD.
- Five-number summary -> boxplot; $1.5 \times \text{IQR}$ outlier rule.
- Skewed -> median/IQR; symmetric -> mean/SD (core judgment).

Unit Summary

Descriptive biostatistics matches displays and summaries to data types and shapes, with robustness guiding the mean/median choice.

Exam Pointer

Very important: choose mean vs median for skewed data; construct boxplots and interpret them.

Unit 6: Probability Foundations

Learning Objectives

Objective 1: Compute probabilities from sample spaces

Objective 2: Apply addition and multiplication rules

Objective 3: Use conditional probability and Bayes' theorem

Probability quantifies uncertainty: $P(\text{event}) = \text{favorable/possible}$ for equally likely outcomes, axiomatically between 0 and 1 with $P(S) = 1$. The addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ (subtract the overlap; mutually exclusive events simply add). Complements simplify: $P(\text{not } A) = 1 - P(A)$.

The multiplication rule: $P(A \text{ and } B) = P(A) \times P(B|A)$; independent events multiply directly. Conditional probability $P(B|A) = P(A \text{ and } B)/P(A)$ updates beliefs with evidence. Medical testing runs on this: sensitivity $P(\text{positive}|\text{disease})$, specificity $P(\text{negative}|\text{healthy})$.

Bayes' theorem inverts conditions: $P(\text{disease}|\text{positive})$ from $P(\text{positive}|\text{disease})$, using prevalence. The base-rate fallacy (ignoring prevalence) makes tests look more informative than they are; a 99%-accurate test for a rare disease still yields mostly false positives. Very important: Bayes numerals on medical testing are the signature application; set up the tree or table carefully.

Key Definitions

Conditional probability is defined as $P(B|A)$: probability of B given that A occurred.

Independence is defined as $P(A \text{ and } B) = P(A)P(B)$; occurrence of one does not affect the other.

Bayes' theorem is defined as inverting conditional probabilities using prior prevalence.

Key Points to Remember

- Addition rule subtracts overlap; complements simplify.
- Multiplication with conditioning; independence multiplies directly.
- Sensitivity/specificity are conditional probabilities.
- Bayes: posterior = likelihood x prior / evidence; beware base rates.

Unit Summary

Probability rules handle unions, intersections and conditioning; Bayes' theorem updates disease probabilities from test results against base rates.

Exam Pointer

Very important: solve medical-testing Bayes numericals; distinguish sensitivity from predictive value.

Unit 7: Probability Distributions

Learning Objectives

- Objective 1:** Work with binomial and Poisson distributions
- Objective 2:** Use the normal distribution and z-scores
- Objective 3:** Apply the central limit theorem

Discrete distributions model counts. The binomial counts successes in n fixed trials with constant p : $P(X=k) = C(n,k) p^k (1-p)^{n-k}$, mean np , variance $np(1-p)$. Requirements: fixed n , independence, constant p , binary outcomes. The Poisson counts rare events in time/space with mean λ , approximating the binomial when n is large and p small.

The normal distribution, bell-shaped and symmetric, is parameterized by mean μ and standard deviation σ ; the empirical rule (68-95-99.7%) gives quick probabilities. Z-scores $z = (x - \mu)/\sigma$ standardize any normal to $N(0,1)$ for table lookup. Biological measurements (heights, blood pressures) are often approximately normal.

The central limit theorem: sample means approach normality as n grows, whatever the population shape, with standard error $\sigma/\sqrt{n}$. This justifies inference from samples. Very important: z-score standardization and the empirical rule are rapid marks; CLT statements must include the $\sigma/\sqrt{n}$ standard error.

Key Definitions

Binomial distribution is defined as counts of successes in n independent trials with constant p .

Normal distribution is defined as the symmetric bell curve parameterized by mean and standard deviation.

Central limit theorem is defined as sample means become approximately normal with standard error $\sigma/\sqrt{n}$.

Key Points to Remember

- Binomial: fixed n , binary, constant p ; mean np . Poisson: rare events, mean λ .
- Normal: empirical 68-95-99.7 rule; $z = (x - \mu)/\sigma$ for table use.
- CLT: means $\rightarrow$ normal; standard error $\sigma/\sqrt{n}$; justifies sampling inference.
- Biological data often approximately normal.

Unit Summary

Binomial and Poisson model counts, the normal models measurements via z-scores, and the central limit theorem legitimizes inference from samples.

Exam Pointer

Very important: standardize with z-scores using the empirical rule; state the CLT with standard error.

Unit 8: Sampling, Estimation and Confidence Intervals

Learning Objectives

Objective 1: Distinguish sampling methods

Objective 2: Construct confidence intervals for means and proportions

Objective 3: Interpret confidence levels correctly

Good inference needs good samples: simple random (every subset equally likely), stratified (random within subgroups), systematic, and cluster sampling. Bias sources: selection bias, nonresponse, measurement bias. Sample size drives precision; representativeness drives validity. Medical trials randomize to kill confounding.

A confidence interval estimates a parameter with stated reliability: for a mean, $\bar{x} \pm z(\sigma/\sqrt{n})$ (large n) or with t for small samples; for a proportion, $\hat{p} \pm z\sqrt{\hat{p}(1-\hat{p})/n}$. Wider intervals (higher confidence, smaller n , more variability) trade precision for certainty.

Interpretation: 95% confidence means 95% of such intervals capture the true parameter; it is not the probability the parameter lies inside this particular interval. Very important: this correct interpretation is heavily examined because the misconception is so common; also know when t replaces z .

Key Definitions

Confidence interval is defined as a range estimating a parameter with stated long-run capture reliability.

Margin of error is defined as the $\pm$ term; half the interval width.

Sampling bias is defined as systematic distortion from unrepresentative sampling.

Key Points to Remember

- Methods: simple random, stratified, systematic, cluster; randomize to kill confounding.
- CI for mean: $\bar{x} \pm z \cdot \sigma / \sqrt{n}$ (t for small n); for proportion analogous.
- Width drivers: confidence level, n, variability.
- Correct interpretation: long-run capture rate, not probability about this interval.

Unit Summary

Random sampling enables inference; confidence intervals quantify estimate precision, interpreted as long-run capture rates.

Exam Pointer

Very important: construct CIs and interpret confidence correctly; choose z vs t.

Unit 9: Hypothesis Testing

Learning Objectives

- Objective 1:** Set up null and alternative hypotheses
- Objective 2:** Conduct z/t tests and interpret p-values
- Objective 3:** Explain errors, power and chi-square tests

Hypothesis testing asks whether data contradict a null hypothesis H_0 . Steps: state H_0 and H_1 , choose significance α , compute the test statistic (z or t), find the p-value (probability of data this extreme if H_0 true), reject H_0 if $p < \alpha$. The p-value is not $P(H_0 \text{ true})$; it assumes H_0 .

Errors: Type I (rejecting true H_0 , probability α), Type II (missing real effects, probability β); power = $1 - \beta$. Medical testing balances these: a false negative (Type II) may kill, a false positive (Type I) wastes resources. Larger samples raise power.

The chi-square test handles categorical data: goodness-of-fit (does the distribution match?) and independence (are two variables associated?), computed from observed vs expected counts. Very important: the five-step procedure and correct p-value interpretation are the exam core; chi-square suits genetics and survey data.

Key Definitions

P-value is defined as probability of data as extreme as observed, assuming H_0 true.

Type I / Type II error is defined as false positive / false negative relative to H_0 .

Chi-square test is defined as testing categorical fit or independence from observed vs expected counts.

Key Points to Remember

- Five steps: hypotheses, α , statistic, p-value, decision; p assumes H_0 .
- Type I = false alarm (α); Type II = miss (β); power = $1 - \beta$.
- Medical stakes decide which error matters more.
- Chi-square for categorical: goodness-of-fit and independence.

Unit Summary

Hypothesis testing weighs evidence against H_0 via p-values with controlled error rates; chi-square extends testing to categorical biomedical data.

Exam Pointer

Very important: run the five-step test, interpret p-values correctly, and apply chi-square to a contingency table.

Unit 10: Correlation, Regression and Vital Statistics

Learning Objectives

- Objective 1:** Compute and interpret correlation
- Objective 2:** Fit and use simple linear regression
- Objective 3:** Define basic vital and health statistics

Correlation r measures linear association strength and direction (-1 to 1); r^2 is the proportion of variance explained. Correlation is not causation: confounding, reverse causation and coincidence all produce correlations. Scatterplots reveal nonlinearity and outliers that r hides.

Simple linear regression fits $y = a + bx$ by least squares (minimizing squared residuals); $b = r(s_y/s_x)$ links slope to correlation. Use the line for prediction within the data range (extrapolation is risky); residuals check fit quality. In biostatistics, dose-response relationships are classic regressions.

Vital statistics describe populations: crude birth/death rates, infant mortality rate, maternal mortality ratio, life expectancy, incidence (new cases) versus prevalence (existing cases). Very important: incidence-vs-prevalence and correlation-vs-causation are the two distinctions examiners probe most; state them crisply.

Key Definitions

Correlation coefficient is defined as r in $[-1,1]$ measuring linear association; r^2 = variance explained.

Regression line is defined as the least-squares fit $y = a + bx$ minimizing squared residuals.

Incidence / prevalence is defined as new cases per period / existing cases at a time.

Key Points to Remember

- r : strength+direction of linear association; not causation.
- Regression: least squares, $b = r(s_y/s_x)$; predict within range; check residuals.
- Vital stats: birth/death rates, IMR, MMR, life expectancy.
- Incidence (new) vs prevalence (existing): crisp distinction.

Unit Summary

Correlation and regression quantify linear relationships without proving causation; vital statistics (incidence, prevalence, mortality rates) describe population health.

Exam Pointer

Very important: interpret r and r^2 ; distinguish incidence from prevalence; warn against causal claims from correlation.