

# VU LEARNING HUB

[vulearninghub.xyz](https://vulearninghub.xyz)

**HIGHLIGHTED HANDOUT**

## **STA633**

Multivariate Analysis

**STA633 by VU Learning HUB**

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

## STA633

# Multivariate Analysis

## Analysis of Multivariate Data — Study Handout

This handout covers STA633 Multivariate Analysis: the extension of univariate statistics to vectors of variables — multivariate normal distribution, Hotelling's  $T^2$ , MANOVA, principal components, factor analysis, discriminant analysis and clustering. Matrix notation is used throughout: bold  $\mathbf{x}$  is a  $p \times 1$  vector,  $\Sigma$  (sigma) is the covariance matrix.

Notation:  $\bar{\mathbf{x}}$  = sample mean vector,  $S$  = sample covariance matrix,  $\mu$  = population mean vector,  $\Sigma$  = population covariance matrix,  $|A|$  = determinant,  $\text{tr}(A)$  = trace. Work the small numerical examples by hand once — exams reward matrix-arithmetic fluency.

## Table of Contents

- Unit 1: Introduction: Multivariate Data and Matrix Tools ([Lectures 1–5](#))
- Unit 2: The Multivariate Normal Distribution ([Lectures 6–9](#))
- Unit 3: Hotelling's  $T^2$ : One-Sample and Two-Sample Tests ([Lectures 10–14](#))
- Unit 4: MANOVA: Comparing Several Mean Vectors ([Lectures 15–18](#))
- Unit 5: Principal Component Analysis ([Lectures 19–23](#))
- Unit 6: Factor Analysis ([Lectures 24–27](#))
- Unit 7: Discriminant Analysis and Classification ([Lectures 28–31](#))
- Unit 8: Cluster Analysis ([Lectures 32–35](#))
- Unit 9: Review: Choosing the Right Multivariate Method ([Lectures 36–38](#))

# Unit 1: Introduction: Multivariate Data and Matrix Tools

## Lectures 1–5

### Learning Objectives

- Explain why multivariate methods are needed
- Organize data in the  $n \times p$  data matrix
- Compute the mean vector and covariance matrix
- Review essential matrix algebra: inverse, determinant, eigenvalues

### Why Multivariate?

Real data comes in bundles: a patient's blood pressure, cholesterol and glucose; a customer's age, income and spending. Analyzing each variable separately **misses correlations** and inflates error rates (running 10 t-tests at 5% gives ~40% family-wise error). Multivariate methods analyze the **vector as a whole**, respecting how variables move together.

### The Data Matrix and Summaries

**Data matrix  $X$  ( $n \times p$ ):**  $n$  observations (rows),  $p$  variables (columns). —

**Mean vector  $\bar{x}$  ( $p \times 1$ ):** the  $p$  variable means stacked. —

**Sample covariance matrix  $S$  ( $p \times p$ ):**  $S_{jk}$  = covariance of variables  $j$  and  $k$ ; diagonal = variances.  $S$  is symmetric. —

**Correlation matrix  $R$ :**  $R_{jk} = S_{jk} / \sqrt{(S_{jj} \cdot S_{kk})}$ ; diagonal all 1s — scale-free, used when variables have different units.

### Matrix Essentials

- **Determinant  $|A|$ :** for  $2 \times 2$   $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ ,  $|A| = ad - bc$ .  $|A| = 0$  means singular (no inverse) — variables perfectly collinear.
- **Inverse:**  $A^{-1}A = I$ . For  $2 \times 2$ :  $(1/|A|)\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ .
- **Eigenvalues/eigenvectors:** solve  $|A - \lambda I| = 0$ . Eigenvectors give the principal axes of the data cloud; eigenvalues their variances — the engine of PCA.
- **Trace:** sum of diagonal = sum of eigenvalues = total variance.

*Worked:*  $X = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$  (covariance matrix).  $|X| = 16 - 4 = 12$ . Inverse =  $(1/12)\begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$ .

*Eigenvalues:*  $|X - \lambda I| = (4 - \lambda)^2 - 4 = \lambda^2 - 8\lambda + 12 = 0 \rightarrow \lambda = 6, 2$ . Trace =  $8 = 6 + 2$  ✓. So total variance 8 splits into 6 along the first principal axis and 2 along the second.

### Unit Summary

- Multivariate analysis treats correlated variables jointly; separate univariate tests miss correlation and inflate errors.
- Summarize with  $\bar{x}$ ,  $S$  (covariances) and  $R$  (correlations).
- Determinant, inverse, eigenvalues/eigenvectors and trace are the working tools — PCA runs on eigen-decomposition.

**Exam Pointers** 

- ★ Compute  $S$  from a small data set and interpret one off-diagonal element.
- ★ Find eigenvalues, determinant and inverse of a  $2 \times 2$  covariance matrix.
- ★ Why can't ten separate t-tests replace one multivariate test?

## Unit 2: The Multivariate Normal Distribution

### Lectures 6–9

#### Learning Objectives

- Write the multivariate normal density
- Explain the role of  $\mu$  and  $\Sigma$
- Describe properties: marginals, conditionals, linear combinations
- Use Mahalanobis distance

#### Density and Parameters

**Density:**  $f(\mathbf{x}) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp(-\frac{1}{2}(\mathbf{x}-\mu)' \Sigma^{-1}(\mathbf{x}-\mu))$ . —

- $\mu$  ( $p \times 1$ ): centers the distribution.  $\Sigma$  ( $p \times p$ ): shapes it — variances on the diagonal, covariances off-diagonal; must be positive definite.
- Contours of equal density are **ellipsoids** centered at  $\mu$ , with axes along  $\Sigma$ 's eigenvectors and lengths  $\propto \sqrt{\text{eigenvalues}}$ .

#### Key Properties

- **Marginals** of a multivariate normal are normal; **conditionals** are normal with mean adjusted by regression on the known variables.
- **Linear combinations**  $\mathbf{a}'\mathbf{x}$  are univariate normal: mean  $\mathbf{a}'\mu$ , variance  $\mathbf{a}'\Sigma\mathbf{a}$ .
- If  $\Sigma$  is diagonal (zero correlations), the variables are **independent** — for normal data, uncorrelated  $\blacksquare$  independent.
- Sample  $\mathbf{x} \sim N_p(\mu, \Sigma/n)$ ; the scatter matrix  $(n-1)S \sim \text{Wishart}$  — the multivariate chi-square.

#### Mahalanobis Distance

$D^2 = (\mathbf{x} - \mu)' \Sigma^{-1}(\mathbf{x} - \mu)$  — distance corrected for variances and correlations.

- Unlike Euclidean distance, it down-weights high-variance directions and accounts for correlation — the right metric for outlier detection and classification.

*Worked:*  $\mu = (0,0)$ ,  $\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ . Point  $(4, 1)$ : Mahalanobis  $D^2 = 16/4 + 1/1 = 5$ . Euclidean distance<sup>2</sup> = 17. The point is 2 SDs out in  $x$  but only 1 in  $y$  — Mahalanobis correctly calls it less extreme than Euclidean suggests.

#### Unit Summary

- $N_p(\mu, \Sigma)$ :  $\mu$  centers,  $\Sigma$  shapes; density contours are ellipsoids.
- Marginals, conditionals and linear combinations stay normal; uncorrelated = independent.
- Mahalanobis  $D^2 = (\mathbf{x}-\mu)' \Sigma^{-1}(\mathbf{x}-\mu)$  is the correlation-aware distance.

#### Exam Pointers

- ★ Write the bivariate normal density and identify  $\mu$ ,  $\Sigma$ 's roles.
- ★ Compute Mahalanobis distance for a given point and compare with Euclidean.

- ★ State two properties that make the normal the default multivariate model.

## Unit 3: Hotelling's $T^2$ : One-Sample and Two-Sample Tests

### Lectures 10–14

#### Learning Objectives

- Construct Hotelling's  $T^2$  for one sample
- Convert  $T^2$  to an F statistic and test
- Extend to two independent samples
- Build simultaneous confidence intervals

#### One-Sample $T^2$

$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$  — the multivariate analogue of the t-statistic squared.

Test:  $F = [(n-p) / (p(n-1))] \times T^2 \sim F(p, n-p)$ . Reject  $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$  if F exceeds the critical value. —

- Intuition: squared Mahalanobis distance of  $\bar{\mathbf{x}}$  from  $\boldsymbol{\mu}_0$ , scaled by n — big distance ■ reject.

#### Two-Sample $T^2$

$T^2 = [n_1 n_2 / (n_1 + n_2)] (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{\text{pooled}}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ , where  $\mathbf{S}_{\text{pooled}} = [(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2] / (n_1 + n_2 - 2)$ . —

- Same F conversion with  $n = n_1 + n_2$ . Assumes equal covariance matrices (test with Box's M).

#### Simultaneous Confidence Intervals

After rejecting, find **which** variables differ: for any linear combination  $\mathbf{a}'\boldsymbol{\mu}$ , the interval  $\mathbf{a}'\bar{\mathbf{x}} \pm \sqrt{[p(n-1)/(n-p) \times F_{\alpha} \times \mathbf{a}'\mathbf{S}\mathbf{a}/n]}$  holds **simultaneously** for all  $\mathbf{a}$  at confidence  $1-\alpha$ . For single variables, set  $\mathbf{a}$  to unit vectors — wider than ordinary t-intervals, honestly accounting for multiplicity.

Worked ( $p=2$ ):  $n=20$ ,  $\bar{\mathbf{x}}=(5,8)$ ,  $\boldsymbol{\mu}_0=(4,7)$ ,  $\mathbf{S}=[[2,0.5],[0.5,3]]$ .  $\mathbf{d}=(1,1)$ .  $\mathbf{S}^{-1} = (1/5.75)[[3,-0.5],[-0.5,2]]$ .  $\mathbf{d}'\mathbf{S}^{-1}\mathbf{d} = (1/5.75)(3-1+2) = 4/5.75 = 0.696$ .  $T^2 = 20 \times 0.696 = 13.91$ .  $F = (18/(2 \times 19)) \times 13.91 = 6.59$ .  $F_{0.05,2,18} = 3.55 \rightarrow \text{reject } H_0$ .

#### Unit Summary

- $T^2 = n \cdot (\text{Mahalanobis distance of } \bar{\mathbf{x}} \text{ from } \boldsymbol{\mu}_0)^2$  — one test for the whole mean vector.
- Convert to  $F(p, n-p)$  for the p-value; two-sample version uses pooled S.
- Simultaneous CIs identify which variables drive rejection.

#### Exam Pointers

- ★ Compute one-sample  $T^2$  and the F test from given  $\bar{\mathbf{x}}$ ,  $\mathbf{S}$ ,  $n$ ,  $\boldsymbol{\mu}_0$ .
- ★ Write the two-sample  $T^2$  formula and state its assumption.
- ★ Why are simultaneous intervals wider than separate t-intervals?

## Unit 4: MANOVA: Comparing Several Mean Vectors

### Lectures 15–18

#### Learning Objectives

- Set up one-way MANOVA with H and E matrices
- Compute Wilks' lambda and test it
- Contrast MANOVA with separate ANOVAs
- Follow up significant results

#### H and E Matrices

One-way MANOVA with  $g$  groups: **H** (hypothesis SSCP) captures between-group variation; **E** (error SSCP) captures within-group variation — the multivariate analogues of SSB and SSW.  $H_0$ : all group mean vectors equal.

#### Wilks' Lambda

$\Lambda = |E| / |E + H|$ . Ranges  $0 \rightarrow 1$ ; small  $\Lambda$  = groups differ a lot ■ reject. —

- Convert to F (Rao's approximation) or chi-square for large  $n$ :  $-\frac{1}{2}[n-1-(p+g)/2] \cdot \ln \Lambda \sim \chi^2(p(g-1))$ .
- Alternatives: Pillai's trace, Hotelling–Lawley trace, Roy's largest root — Pillai is most robust to violations.

#### MANOVA vs Separate ANOVAs

- Separate ANOVAs inflate Type I error and ignore correlation between dependent variables.
- MANOVA can detect group differences in the **combination** of variables even when no single variable differs (the multivariate pattern).
- Assumptions: multivariate normality within groups, equal covariance matrices (Box's M test), independent observations.

*Follow-up plan after significant MANOVA: 1) simultaneous CIs / discriminant analysis to find separating variables, 2) univariate ANOVAs with Bonferroni correction as a descriptive aid (not a replacement), 3) plot group means with confidence ellipses.*

#### Unit Summary

- MANOVA: H (between) and E (within) SSCP matrices; Wilks'  $\Lambda = |E|/|E+H|$ , small = significant.
- Beats separate ANOVAs on error control and on detecting combined-variable effects.
- Assumptions: normality, equal covariances (Box's M), independence.

#### Exam Pointers

- ★ Define H and E and write Wilks' lambda.
- ★ Compute  $\Lambda$  from given  $|E|$ ,  $|H+E|$  and interpret.
- ★ Give two reasons MANOVA is preferred over multiple ANOVAs.

## Unit 5: Principal Component Analysis

### Lectures 19–23

#### Learning Objectives

- Derive principal components from eigen-decomposition
- Decide how many components to retain
- Interpret loadings and scores
- Choose between covariance and correlation matrix

#### The Idea

PCA finds **new uncorrelated variables** (components) that are linear combinations of the originals, ordered by variance captured. PC1 = the direction of maximum variance (first eigenvector of  $S$  or  $R$ ); PC2 = max variance orthogonal to PC1; and so on. If the first 2–3 PCs capture 80–90% of total variance,  $p$  variables collapse to 2–3 with little loss — the basis of dimension reduction and biplots.

#### Computation and Decisions

**Solve  $|S - \lambda I| = 0 \rightarrow$  eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ ;  $PC_i = e_i'x$ . Proportion of variance =  $\lambda_i / \sum \lambda_i$ . —**

- **How many to keep:** Kaiser's rule ( $\lambda > 1$ , for correlation matrix), scree plot elbow, cumulative 80–90% rule.
- **S or R?** Use  $R$  (correlation) when variables have different units/scales;  $S$  when units are commensurable and variance differences are meaningful.

#### Loadings and Scores

- **Loadings** (eigenvector elements, often scaled by  $\sqrt{\lambda}$ ): correlation of variable with component — name the component from its high-loading variables.
- **Scores:** each observation's coordinates on the PCs — used for plots, outlier detection and as inputs to regression/clustering.

*Worked:  $S = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$  (from Unit 1):  $\lambda = 6, 2$ ; PC1 explains  $6/8 = 75\%$ , PC2 25%. PC1 direction  $\propto (1,1)$ : the 'overall size' component; PC2  $\propto (1,-1)$ : the contrast. With 75% in one component, a 2-D problem is nearly 1-D.*

#### Unit Summary

- PCs = eigenvectors of  $S/R$  ordered by eigenvalue; each captures max remaining variance.
- Retain via Kaiser, scree elbow, or 80–90% cumulative variance.
- Loadings interpret components; scores locate observations; use  $R$  for mixed units.

#### Exam Pointers

- ★ Compute PCs, variance proportions for a  $2 \times 2$   $S$ .
- ★ State three retention rules.

- ★ When must PCA use the correlation matrix instead of covariance?

## Unit 6: Factor Analysis

Lectures 24–27

### Learning Objectives

- Distinguish factor analysis from PCA
- Write the orthogonal factor model
- Explain loadings, communalities and specific variances
- Describe rotation and factor interpretation

### PCA vs Factor Analysis

**PCA** is a data-reduction trick: components are exact linear combos capturing max variance. **Factor analysis** is a **model**: observed variables are generated by a few unobserved (latent) factors plus unique noise. PCA explains variance; FA explains **covariance/correlation structure**. Use FA when you hypothesize latent traits (e.g., 'intelligence' behind test scores).

### The Orthogonal Factor Model

$\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}$ , where  $\mathbf{L}$  ( $p \times m$ ) = loadings,  $\mathbf{F}$  ( $m \times 1$ ) = common factors,  $\boldsymbol{\varepsilon}$  = specific factors. —

**Covariance structure:**  $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$ , with  $\Psi$  diagonal (specific variances). —

- **Communality**  $h^2_j$  = sum of squared loadings of variable  $j$  = variance explained by common factors.
- **Specific variance**  $\psi_j = 1 - h^2_j$  (using R) = unique + error variance.

### Estimation and Rotation

- **Estimation:** principal-component method (eigen-decomposition of R) or maximum likelihood (assumes normality; gives goodness-of-fit test).
- **Rotation problem:** L is not unique — rotate (orthogonal: varimax; oblique: promax) to get simple structure where each variable loads highly on one factor only.
- Interpret factors from the rotated loading pattern; validate with the chi-square fit test (ML method).

*Example: 6 exam scores → 2 factors: 'quantitative ability' (loads on math, physics) and 'verbal ability' (loads on Urdu, English), after varimax rotation. Communalities ~0.7 mean the two factors explain 70% of each test's variance; the rest is test-specific.*

### Unit Summary

- FA models  $\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon}$ : latent factors explain correlations;  $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$ .
- Communality = variance explained by factors; specific variance = the rest.
- Rotate (varimax) for interpretability; ML estimation adds a fit test.

### Exam Pointers

- ★ Contrast PCA and factor analysis (purpose and model).

- ★ Write the factor model and define communality.
- ★ Why rotate? What is simple structure?

## Unit 7: Discriminant Analysis and Classification

### Lectures 28–31

#### Learning Objectives

- Build Fisher's linear discriminant function
- Classify observations via Mahalanobis distance
- Extend to multiple groups (canonical discriminants)
- Evaluate classification with confusion matrices

#### Fisher's Linear Discriminant (Two Groups)

Find the linear combination  $y = a'x$  that **maximizes between-group variance relative to within-group variance**:  $a = S_{\text{pooled}}^{-1}(x_{\text{group 1}} - x_{\text{group 2}})$ . Project all data onto this line; classify by which group mean is nearer. This is the direction that best separates the groups — closely related to the two-sample  $T^2$  and to MANOVA.

#### Classification Rule

**Allocate**  $x$  to **group 1** if  $(x_{\text{group 1}} - x_{\text{group 2}})' S_{\text{pooled}}^{-1} (x - (x_{\text{group 1}} + x_{\text{group 2}})/2) > \ln[(c_{\text{group 1}} p_{\text{group 1}})/(c_{\text{group 2}} p_{\text{group 2}})]$ -ish threshold; with equal costs and priors, the midpoint rule: nearer Mahalanobis distance wins. —

- With unequal priors or misclassification costs, shift the cutoff toward the cheaper/more likely group.
- **Quadratic discriminant analysis (QDA)**: allow each group its own covariance — curved boundaries; needs more data.

#### Multiple Groups and Evaluation

- **Canonical discriminant functions**: up to  $\min(g-1, p)$  uncorrelated directions separating  $g$  groups; plot the first two for a visual map.
- **Confusion matrix**: cross-tabulate true vs predicted group; hit rate = diagonal/total. Validate honestly: split-sample or **leave-one-out cross-validation** (resubstitution flatters).

*Example: bank loan applicants (income, debt ratio) classified good/bad. Fisher's function weights income positively, debt ratio negatively; cutoff set from historical defaults. Confusion matrix on holdout data: 85% correct — but check the costly error: misclassifying a defaulter as good costs far more than the reverse, so adjust the threshold.*

#### Unit Summary

- Fisher's rule: project on  $a = S_{\text{pooled}}^{-1}(x_{\text{group 1}} - x_{\text{group 2}})$ ; classify by nearest (cost/prior-adjusted) center.
- QDA for unequal covariances; canonical functions for  $g > 2$  groups.
- Report cross-validated confusion matrices, and weight errors by cost.

#### Exam Pointers

- ★ Derive/interpret Fisher's linear discriminant direction.

- ★ Classify a new observation by Mahalanobis distance (numbers given).
- ★ Why is resubstitution accuracy optimistic? Name two honest alternatives.

## Unit 8: Cluster Analysis

### Lectures 32–35

#### Learning Objectives

- Distinguish hierarchical and non-hierarchical clustering
- Apply linkage methods and read a dendrogram
- Run k-means logic and choose k
- Validate and profile clusters

#### The Clustering Task

Unlike discriminant analysis, clustering is **unsupervised**: no group labels exist — we discover them. Steps: choose a **distance** (usually Euclidean on standardized variables, or Mahalanobis), choose an **algorithm**, decide **how many clusters**, then **profile** them (what makes each cluster distinctive?).

#### Hierarchical Clustering

- **Agglomerative**: start with n clusters, repeatedly merge the closest pair; draw the **dendrogram** and cut it at a chosen height.
- **Linkage rules** define cluster distance: single (nearest neighbors — can chain), complete (farthest — compact balls), average, Ward's (minimizes within-cluster variance increase — often best).

#### K-Means

Fix k; assign each point to the nearest centroid; recompute centroids; repeat until stable. Fast for large n. Weaknesses: must pre-specify k, sensitive to starting seeds and outliers, assumes spherical clusters. Choose k via the **elbow** in within-cluster sum of squares, silhouette scores, or practical interpretability.

- **Standardize first**: variables with big units dominate distances otherwise.

*Example: segmenting 500 bank customers on (balance, transactions/month, loan usage). Standardize → Ward's hierarchical suggests 4 clusters → k-means refines → profiles: 'dormant savers', 'active transactors', 'borrowers', 'premium low-activity'. Marketing then targets each differently — the business payoff of clustering.*

#### Unit Summary

- Clustering is unsupervised group discovery: distance + algorithm + k + profiling.
- Hierarchical: merge stepwise, read the dendrogram; linkage choice shapes clusters (Ward's often best).
- K-means: iterate assign/update; pick k by elbow/silhouette; always standardize first.

#### Exam Pointers

- ★ Contrast hierarchical and k-means clustering (4 points).

- ★ Explain single vs complete vs Ward's linkage.
- ★ Why standardize variables before clustering?

## Unit 9: Review: Choosing the Right Multivariate Method

### Lectures 36–38

#### Learning Objectives

- Map research questions to methods
- State the assumptions checklist for each method
- Avoid common multivariate pitfalls
- Plan a complete multivariate analysis

#### Method Selection Map

Compare one mean vector to a standard → Hotelling's one-sample  $T^2$ . —

Compare two groups' mean vectors → two-sample  $T^2$ . —

Compare 3+ groups' mean vectors → MANOVA (+ discriminant follow-up). —

Reduce dimensions / find latent structure → PCA (reduction) or factor analysis (latent model). —

Classify into known groups → discriminant analysis (QDA if covariances differ). —

Discover unknown groups → cluster analysis. —

#### Assumption Checklist

- Multivariate normality (check marginals + Mahalanobis  $D^2$  chi-square plot).
- Equal covariance matrices for  $T^2$ /MANOVA/LDA (Box's  $M$ ).
- No extreme multicollinearity (singular  $S$  breaks inverses) — drop or combine redundant variables.
- $n$  comfortably exceeds  $p$  (rule of thumb:  $n > 20p$  for stable covariance estimates).

#### Pitfalls and Workflow

- Running many univariate tests instead of one multivariate test (error inflation).
- Interpreting PCs/factors without rotation or validation.
- Reporting resubstitution accuracy for classifiers.
- Forgetting to standardize before PCA-on-R decisions, clustering or regularized methods.

Sensible workflow: 1) clean and screen data (outliers via  $D^2$ ), 2) check assumptions, 3) pick method by question type, 4) fit, 5) validate (cross-validation, fit tests), 6) interpret in subject-matter terms.

*Capstone mini-case: A bank has 8 financial ratios for 300 firms, half defaulted. Task: build an early-warning model. Plan: screen outliers ( $D^2$ ), check normality/equal covariances, run LDA (or QDA if Box's  $M$  rejects), validate with leave-one-out, report the confusion matrix with cost-weighted errors, and interpret the discriminant weights as the risk factors. Every method from this course appears in one pipeline.*

### Unit Summary

- Match the question to the method:  $T^2$  (1–2 groups), MANOVA (3+), PCA/FA (structure), LDA (classification), clustering (discovery).
- Check normality, equal covariances, no singularity, adequate n.
- Workflow: screen → assume-check → fit → validate → interpret.

### Exam Pointers

- ★ Given five research scenarios, name the method for each.
- ★ List the assumption checks before running MANOVA.
- ★ Outline a full analysis plan for the bank default-prediction case.

## Key Terms Glossary

*All key terms from this handout in one place — use for quick revision.*

**Data matrix  $X$  ( $n \times p$ ):**  $n$  observations (rows),  $p$  variables (columns). —

**Mean vector  $\bar{x}$  ( $p \times 1$ ):** the  $p$  variable means stacked. —

**Sample covariance matrix  $S$  ( $p \times p$ ):**  $S_{jk}$  = covariance of variables  $j$  and  $k$ ; diagonal = variances.  $S$  is symmetric. —

**Correlation matrix  $R$ :**  $R_{jk} = S_{jk} / \sqrt{(S_{jj} \cdot S_{kk})}$ ; diagonal all 1s — scale-free, used when variables have different units.

**Density:**  $f(x) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu))$ . —

$D^2 = (x - \mu)'\Sigma^{-1}(x - \mu)$  — distance corrected for variances and correlations.

$T^2 = n(x - \bar{x})'S^{-1}(x - \bar{x})$  — the multivariate analogue of the t-statistic squared.

**Test:**  $F = [(n-p) / (p(n-1))] \times T^2 \sim F(p, n-p)$ . Reject  $H_0: \mu = \mu_0$  if  $F$  exceeds the critical value. —

$T^2 = [n_1 n_2 / (n_1 + n_2)] (x_1 - x_2)' S_{\text{pooled}}^{-1} (x_1 - x_2)$ , where  $S_{\text{pooled}} = [(n_1 - 1)S_1 + (n_2 - 1)S_2] / (n_1 + n_2 - 2)$ . —

$\Lambda = |E| / |E + H|$ . Ranges  $0 \rightarrow 1$ ; small  $\Lambda$  = groups differ a lot ■ reject. —

Solve  $|S - \lambda I| = 0 \rightarrow$  eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ ;  $PC_i = e_i'x$ . Proportion of variance =  $\lambda_i / \sum \lambda$ . —

$x = \mu + LF + \varepsilon$ , where  $L$  ( $p \times m$ ) = loadings,  $F$  ( $m \times 1$ ) = common factors,  $\varepsilon$  = specific factors. —

**Covariance structure:**  $\Sigma = LL' + \Psi$ , with  $\Psi$  diagonal (specific variances). —

Allocate  $x$  to group 1 if  $(x - \bar{x}_1)' S_{\text{pooled}}^{-1} (x - (\bar{x}_1 + \bar{x}_2)/2) > \ln[(c_1 p_1) / (c_2 p_2)]$ -ish threshold; with equal costs and priors, the midpoint rule: nearer Mahalanobis distance wins. —

Compare one mean vector to a standard  $\rightarrow$  Hotelling's one-sample  $T^2$ . —

Compare two groups' mean vectors  $\rightarrow$  two-sample  $T^2$ . —

Compare 3+ groups' mean vectors  $\rightarrow$  MANOVA (+ discriminant follow-up). —

Reduce dimensions / find latent structure  $\rightarrow$  PCA (reduction) or factor analysis (latent model). —

Classify into known groups  $\rightarrow$  discriminant analysis (QDA if covariances differ). —

Discover unknown groups  $\rightarrow$  cluster analysis. —