MTL766: Key Formulas and Concepts

A Quick Reference Guide

September 13, 2025

1 Descriptive Statistics for Multivariate Data

Let $\mathbf{X}$ be an $n \times p$ data matrix, where n is the number of samples and p is the number of variables. The i-th observation is $\mathbf{x}_i^T = [x_{i1}, x_{i2}, \dots, x_{ip}]$.

- **Sample Mean Vector ($\bar{\mathbf{x}}$):** The vector of sample means for each variable.

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$$

- **Sample Covariance Matrix (S):** A symmetric $p \times p$ matrix where the (j, k)-th element is the sample covariance between variable j and variable k.

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

  The diagonal elements s_{jj} are the sample variances, and off-diagonal elements s_{jk} are the sample covariances.

- **Sample Correlation Matrix (R):** A symmetric $p \times p$ matrix where the (j, k)-th element is the sample correlation coefficient r_{jk}.

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}}$$

$$\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2} \quad \text{where} \quad \mathbf{D} = \text{diag}(s_{11}, s_{22}, \dots, s_{pp})$$

- **Generalized Sample Variance:** The determinant of the sample covariance matrix, $|\mathbf{S}|$. It measures the overall spread of the data.

- **Total Sample Variance:** The trace of the sample covariance matrix, $\text{tr}(\mathbf{S}) = \sum_{j=1}^{p} s_{jj}$. It's the sum of the individual variances.

2 Geometric Concepts and Distances

- **Mahalanobis Distance:** The statistical distance of a point $\mathbf{x}$ from a group with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. It accounts for the correlation between variables.

$$D^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

- **Constant Density Ellipsoid:** For a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the points of constant probability density form an ellipsoid defined by:

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2$$

  The axes of the ellipsoid are in the direction of the eigenvectors of $\boldsymbol{\Sigma}$, and their lengths are proportional to the square roots of the corresponding eigenvalues.

3 The Multivariate Normal (MVN) Distribution

A random vector $\mathbf{X}$ follows an MVN distribution, denoted $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its probability density function is:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

Key Properties:

1. **Linear Combinations:** If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then for a constant matrix $\mathbf{A}$ $(q \times p)$ and vector $\mathbf{b}$ $(q \times 1)$:

$$\mathbf{A}\mathbf{X} + \mathbf{b} \sim N_q(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$$

2. **Quadratic Form:** The Mahalanobis distance from the mean follows a chi-square distribution:

$$(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2_p$$

3. **Conditional Distributions:** If $\mathbf{X}$ is partitioned as $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N_p \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$, the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is also normal:

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim N(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$$

   where

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$
$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

4 Sampling Distributions

- **Distribution of Sample Mean:** For a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the sample mean vector $\bar{\mathbf{X}}$ is also normally distributed:
$$\bar{\mathbf{X}} \sim N_p\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}\right)$$
- **Central Limit Theorem:** For a large sample size n from any population with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$:
$$\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma})$$
- **Wishart Distribution:** The distribution of the sample covariance matrix. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The matrix $\mathbf{A} = (n-1)\mathbf{S} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ follows a Wishart distribution with $n-1$ degrees of freedom, denoted $\mathbf{A} \sim W_p(n-1, \boldsymbol{\Sigma})$.
  - **Expectation:** $\mathbb{E}(\mathbf{A}) = (n-1)\boldsymbol{\Sigma}$.

5 Inference on the Mean Vector

- **Hotelling's T^2 Statistic (One-Sample):** Used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$.
$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$$
  Under H_0, the statistic has a scaled F-distribution:
$$\frac{n-p}{(n-1)p} T^2 \sim F_{p,n-p}$$
  Test Procedure:
  1. State hypotheses: $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.
  2. Calculate the T^2 value from the sample.
  3. Find the critical value $F_{p,n-p}(\alpha)$.
  4. Reject H_0 if $T^2 > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$.
- **100(1-α)% Confidence Ellipsoid for $\boldsymbol{\mu}$:** The set of all $\boldsymbol{\mu}$ vectors for which the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_{test}$ would not be rejected. It is defined by the inequality:
$$n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)$$

6 Maximum Likelihood Estimation (MLE)

For a random sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the likelihood function is:
$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^n f(\mathbf{x}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma})$$
The maximum likelihood estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are:

- $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$
- $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \frac{n-1}{n}\mathbf{S}$ (Note: This is a biased estimator of $\boldsymbol{\Sigma}$).

7 Descriptive Statistics & Properties

Let $\mathbf{X}$ be an $n \times p$ data matrix with observations $\mathbf{x}_i^T = [x_{i1}, \ldots, x_{ip}]$.

- **Sample Mean Vector ($\bar{\mathbf{x}}$):** $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$
- **Sample Covariance Matrix (S):** $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$
  - $\mathbf{S}$ is positive definite if the mean-corrected data matrix ($\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}^T$) has linearly independent columns (requires $n > p$).
- **Sample Correlation Matrix (R):** $r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}s_{kk}}}$. The matrix form is $\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2}$, where $\mathbf{D} = \mathrm{diag}(s_{11}, \ldots, s_{pp})$.
- **Generalized Sample Variance:** $|\mathbf{S}|$.
  - **Geometric Meaning:** Proportional to the squared volume of the parallelepiped formed by the deviation vectors $(\mathbf{x}_i - \bar{\mathbf{x}})$.
  - **Property:** $|\mathbf{S}| = 0$ if and only if the deviation vectors are linearly dependent.
- **Total Sample Variance:** $\mathrm{tr}(\mathbf{S}) = \sum_{j=1}^{p} s_{jj}$.

8 Geometric Concepts & Distances

- **Mahalanobis Distance:** The distance from a point $\mathbf{x}$ to the center of a distribution $\boldsymbol{\mu}$, accounting for covariance $\boldsymbol{\Sigma}$.

  $$D^2 = (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

  This value is a unitless scalar and is non-negative since $\boldsymbol{\Sigma}$ is positive definite.
- **Constant Density Ellipsoid:** For $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the surface of constant density is an ellipsoid. The equation $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2$ defines this surface.
  - To find an ellipse containing $100(1-\alpha)\%$ of the probability, set $c^2 = \chi_p^2(\alpha)$.
  - The half-lengths of the axes are $\sqrt{\lambda_i c^2} = \sqrt{\lambda_i \chi_p^2(\alpha)}$, and their directions are given by the corresponding eigenvectors $\mathbf{e}_i$ of $\boldsymbol{\Sigma}$.
- **Constellation Graph:** A visualization technique for multivariate data.
  1. For each variable (e.g., subject), draw a ray from the origin, with equal angles between rays.
  2. For each observation (e.g., student), plot its value on the corresponding ray.
  3. Connect the points for a single observation to form a polygon (star). The shape and size of the star represent the student's profile.

9 The Multivariate Normal (MVN) Distribution

The PDF is $f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$.

- **Linear Combinations:** If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$:
  - For matrix $\mathbf{A}$, $\mathbf{AX} \sim N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}^T)$.
  - The covariance between two linear combinations $\mathbf{a}^T\mathbf{X}$ and $\mathbf{b}^T\mathbf{X}$ is $\mathbf{a}^T\mathbf{\Sigma b}$.
  - The covariance between a variable X_i and a linear combination $\mathbf{a}^T\mathbf{X}$ is the i-th element of the vector $\mathbf{\Sigma a}$.
- **Unbiased Estimator for Covariance of a Linear Combination:** To estimate $\text{Cov}(\mathbf{AX}) = \mathbf{A\Sigma A}^T$, use the unbiased estimator $\mathbf{ASA}^T$.
- **Conditional Distributions:** If $\mathbf{X} = \begin{pmatrix}\mathbf{X}_1 \\ \mathbf{X}_2\end{pmatrix}$, then the distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is normal with mean $\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ and covariance $\mathbf{\Sigma}_{1|2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$.

10 Key Sampling Distributions

- **Distribution of $\bar{\mathbf{X}}$:** For a sample from $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, $\bar{\mathbf{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\mathbf{\Sigma})$.
- **Central Limit Theorem:** For large n, $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$ is approximately $N_p(\mathbf{0}, \mathbf{\Sigma})$.
- **Chi-Square Distributions:**
  - For $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, $(\mathbf{X}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2_p$.
  - Asymptotically, for large n, $n(\bar{\mathbf{X}}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\mu}) \to \chi^2_p$.
- **Wishart Distribution ($W_p(m, \mathbf{\Sigma})$):**
  - **Definition:** If $\mathbf{Z}_1, \ldots, \mathbf{Z}_m$ are i.i.d. $N_p(\mathbf{0}, \mathbf{\Sigma})$, then $\mathbf{A} = \sum_{i=1}^m \mathbf{Z}_i\mathbf{Z}_i^T \sim W_p(m, \mathbf{\Sigma})$. The expectation is $\mathbb{E}(\mathbf{A}) = m\mathbf{\Sigma}$.
  - **From a Sample:** The matrix of sum of squares and cross-products $(n-1)\mathbf{S} = \sum_{i=1}^n(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ follows a $W_p(n-1, \mathbf{\Sigma})$ distribution.

11 Inference for the Mean Vector

- **One-Sample Hotelling's T^2 Test:**
  - For $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$: $T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$. Reject H_0 if $T^2 > \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$.
  - For Contrasts $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$: Use the test statistic $T^2 = n(\mathbf{C}\bar{\mathbf{x}})^T(\mathbf{CSC}^T)^{-1}(\mathbf{C}\bar{\mathbf{x}})$, where $\mathbf{C}$ is a $q \times p$ matrix of rank q. The critical value is based on the $F_{q,n-q}$ distribution.
- **Two-Sample Hotelling's T^2 Test (for $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$):**

- **Pooled Covariance:** $\mathbf{S}_{pooled} = \frac{(n_1-1)\mathbf{S}_1+(n_2-1)\mathbf{S}_2}{n_1+n_2-2}$.
- **Test Statistic:** $T^2 = \left(\frac{n_1 n_2}{n_1+n_2}\right)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}_{pooled}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$.
- **Distribution:** $\frac{n_1+n_2-p-1}{(n_1+n_2-2)p}T^2 \sim F_{p,n_1+n_2-p-1}$.

- **Confidence Regions:**
  - **Confidence Ellipsoid for $\boldsymbol{\mu}$:** The set of $\boldsymbol{\mu}$ satisfying $n(\bar{\mathbf{x}}-\boldsymbol{\mu})^T\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$.
  - **Simultaneous T^2 Confidence Intervals:** The intervals for all linear combinations $\mathbf{a}^T\boldsymbol{\mu}$ are given by:

$$\mathbf{a}^T\bar{\mathbf{x}} \pm \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\sqrt{\mathbf{a}^T\mathbf{S}\mathbf{a}}$$

    For a single component μ_i, $\mathbf{a}$ is a vector of zeros with a 1 in the i-th position.

12 Inference for the Covariance Matrix

- **Maximum Likelihood Estimators (MLEs):** For a sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:
  - $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$
  - $\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S}$ (This is biased).

- **Likelihood Ratio Test for $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$:**
  - **Test Statistic (large sample):** The test statistic is $-2\ln\Lambda = (n-1)[\text{tr}(\mathbf{S}\boldsymbol{\Sigma}_0^{-1}) - \ln(|\mathbf{S}\boldsymbol{\Sigma}_0^{-1}|) - p]$.
  - **Distribution:** Under H_0, this statistic is approximately distributed as $\chi^2_{p(p+1)/2}$.
  - **Procedure:** Reject H_0 if the calculated statistic is greater than the critical value $\chi^2_{p(p+1)/2}(\alpha)$.