Multivariate normal distribution
and testing for means (see MKB Ch 3)

Where are we going?
  One-sample t-test (univariate) .............................................. 3
  Two-sample t-test (univariate) ............................................. 4
  One sample $T^2$-test (multivariate) ................................. 5
  Two sample $T^2$-test (multivariate) ................................. 6

Multivariate normal distribution 7
  Definition ................................................................. 8
  Why is it important in multivariate statistics? ...................... 9
  Why is it important in multivariate statistics? ................. 10
  Characterization .......................................................... 11
  Properties .............................................................. 12
  Properties .............................................................. 13
  Data matrices ............................................................ 14

Wishart distribution 15
  Quadratic forms of normal data matrices ......................... 16
  Definition ................................................................. 17
  Properties .............................................................. 18
  Properties .............................................................. 19
  Properties .............................................................. 20

Hotelling’s $T^2$ distribution 21
  Definition ................................................................. 22
  One-sample $T^2$ test ..................................................... 23
  Relationship with $F$ distribution .................................. 24
  Mahalanobis distance ................................................... 25
  Two-sample $T^2$ test ..................................................... 26

Final remarks 27
  Assumptions for one-sample $T^2$ test .............................. 28
  Assumptions for two-sample $T^2$ test .............................. 29
  Multivariate test versus univariate tests ....................... 30
One-sample t-test (univariate)

- Suppose that $x_1, \ldots, x_n$ are i.i.d. $N(\mu, \sigma^2)$. Then
  - $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \sim N(\mu, \sigma^2/n)$
  - $ns^2 = \sum_{i=1}^n (x_i - \bar{x})^2 \sim \sigma^2 \chi^2_{n-1}$
  - $\bar{x}$ and $s^2$ are independent

- Suppose that the mean $\mu$ is unknown. Then we can do a one-sample t-test:
  - $H_0: \mu = \mu_0$, $H_a: \mu \neq \mu_0$.
  - Test statistic: $t = \frac{\bar{x} - \mu_0}{s_u / \sqrt{n}}$, where $s_u$ is sample standard deviation with $n - 1$ in the denominator.
  - Under $H_0$, $t \sim t_{n-1}$, i.e., it has a student t-distribution with $n - 1$ degrees of freedom.
  - Compute p-value. If p-value $< 0.05$, we reject the null hypothesis.

Two-sample t-test (univariate)

- Suppose that $x_1, \ldots, x_n$ are i.i.d. $N(\mu_X, \sigma^2_X)$, and let $y_1, \ldots, y_m$ be i.i.d. $N(\mu_Y, \sigma^2_Y)$ (independent of $x_1, \ldots, x_n$).
- Suppose we want to test whether $\mu_X = \mu_Y$, under the assumption that $\sigma_X = \sigma_Y$. Then we can do a two-sample t-test:
  - $H_0: \mu_X = \mu_Y$, $H_a: \mu_X \neq \mu_Y$.
  - Test statistic: $t = \frac{\bar{x} - \bar{y}}{\sqrt{s_p(\frac{1}{n} + \frac{1}{m})}}$, where
    $$s_p^2 = \frac{1}{n + m - 2} \left( ns_X^2 + ms_Y^2 \right).$$
  - Under $H_0$, $t \sim t_{n+m-2}$.

One sample $T^2$-test (multivariate)

- Suppose that $x_1, \ldots, x_n$ are i.i.d. $N_p(\mu, \Sigma)$. Then
  - $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \sim N_p(\mu, \Sigma/n)$.
  - $nS \sim W_p(\Sigma, n-1)$, where $S$ is the sample covariance matrix and $W_p$ is a Wishart distribution.
  - $\bar{x}$ and $S$ are independent.

- Suppose that the mean $\mu$ is unknown. Then we can do a one-sample t-test:
  - $H_0: \mu = \mu_0$, $H_a: \mu \neq \mu_0$.
  - Test statistic: $T = n(\bar{x} - \mu_0)^T S_u^{-1}(\bar{x} - \mu_0)$, where $S_u$ is the sample covariance matrix with $n - 1$ in the denominator.
  - Under $H_0$, $T \sim T^2(p, n-1)$, i.e., it has Hotelling’s $T^2$ distribution with parameters $p$ and $n - 1$. 
Two sample $T^2$-test (multivariate)

- Suppose that $x_1, \ldots, x_n$ are i.i.d. $N_p(\mu_X, \Sigma_X)$, and let $y_1, \ldots, y_m$ be i.i.d. $N_p(\mu_Y, \Sigma_Y)$ (independent of $x_1, \ldots, x_n$).
- Suppose we want to test whether $\mu_X = \mu_Y$, under the assumption that $\Sigma_X = \Sigma_Y$. Then we can do a two-sample t-test:
  - $H_0 : \mu_X = \mu_Y$, $H_a : \mu_X \neq \mu_Y$.
  - Test statistic: $T = (nm/(n+m))(\bar{x} - \bar{y})'S_u^{-1}(\bar{x} - \bar{y})$, where $S_u = 1/(n+m-2)(nS_1 + mS_2)$.
  - Under $H_0$, $T \sim T^2(p, n + m - 2)$.

Multivariate normal distribution

Definition

- A random variable $x \in \mathbb{R}$ has a univariate normal distribution with mean $\mu$ and variance $\sigma^2$ (we write $x \sim N(\mu, \sigma^2)$) iff its density can be written as:
  $$ f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left( -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right) $$
  $$ = \{2\pi\sigma^2\}^{-1/2} \exp\left( -\frac{1}{2} (x-\mu)(\sigma^2)^{-1}(x-\mu) \right). $$
- A random vector $x \in \mathbb{R}^p$ has a $p$-variate normal distribution with mean vector $\mu$ and covariance matrix $\Sigma$ (we write $x \sim N_p(\mu, \Sigma)$) iff its density can be written as:
  $$ f(x) = |2\pi\Sigma|^{-1/2} \exp\left( -\frac{1}{2} (x-\mu)'\Sigma^{-1}(x-\mu) \right). $$

Why is it important in multivariate statistics?

- It is an easy generalization of the univariate normal distribution. Such generalizations are not obvious for all univariate distributions; sometimes there are several plausible ways to generalize them.
- The multivariate normal distribution is entirely defined by its first two moments. Hence, it has a sparse parametrization using only $p(p+3)/2$ parameters (see board).
- In the case of normal variables, zero correlation implies independence, and pairwise independence implies mutual independence. These properties do not hold for many other distributions.
Why is it important in multivariate statistics?

- Linear functions of a multivariate normal are univariate normal. This yields simple derivations.
- Even when the original data is not multivariate normal, certain functions like the sample mean will be approximately multivariate normal due to the central limit theorem.
- The multivariate normal distribution has a simple geometry. Its equiprobability contours are ellipsoids (see picture on slide).

Characterization

- $x$ is $p$-variate normal iff $a'x$ is univariate normal for all fixed vectors $a \in \mathbb{R}^p$ (To allow for $a = 0$, we regard constants as degenerate forms of the normal distribution.)

Geometric interpretation: $x$ is $p$-variate normal iff its projection on any univariate subspace is normal.

This characterization will allow us to derive many properties of the multivariate normal without writing down densities.

Properties

- (Th 3.1.1 of MKB) If $x$ is $p$-variate normal, and if $y = Ax + c$ where $A$ is any $q \times p$ matrix and $c$ is any $q$-vector, then $y$ is $q$-variate normal (see proof on board).

- (Cor 3.1.1.1 of MKB) Any subset of elements of a multivariate normal vector are multivariate normal. In particular, all individual elements are univariate normal (see proof on board).

Properties

- (Th 3.2.1 of MKB) If $x \sim N_p(\mu, \Sigma)$ and $y = Ax + c$, then $y \sim N_q(A\mu + c, A\Sigma A')$ (see proof on board).

- (Cor 3.2.1.1 of MKB) If $x \sim N_p(\mu, \Sigma)$ with $\Sigma > 0$, then $y = \Sigma^{-1/2}(x - \mu) \sim N_p(0, I)$ and $(x - \mu)'\Sigma^{-1}(x - \mu) = \sum_{i=1}^p y_i^2 \sim \chi^2_p$ (see proof on board).
### Data matrices

- Let \( x_1, \ldots, x_n \) be a random sample from \( N(\mu, \Sigma) \). Then we call \( X = (x_1, \ldots, x_n)' \) a data matrix from \( N(\mu, \Sigma) \) or a normal data matrix.

- (Th. 3.3.2 of MKB, without proof) If \( X(n \times p) \) is a normal data matrix from \( N_p(\mu, \Sigma) \) and if \( Y(m \times q) \) satisfies \( Y = AXB \), then \( Y \) is a normal matrix iff the following two properties hold:
  - \( A1 = \alpha 1 \) for some scalar \( \alpha \), or \( B'\mu = 0 \)
  - \( AA' = \beta I \) for some scalar \( \beta \), or \( B'\Sigma B = 0 \)

  When both these conditions are satisfied then \( Y \) is a normal data matrix from \( N_q(\alpha B'\mu, \beta B'\Sigma B) \).

Note: Pre-multiplication with \( A \) means that we take linear combinations of the rows. The conditions on \( A \) ensure that the new rows are independent. Post-multiplication with \( B \) means that we take linear combinations of the columns (variables).

### Wishart distribution

#### Quadratic forms of normal data matrices

- We now consider quadratic forms of normal data matrices, i.e., functions of the form \( X'CX \) for some symmetric matrix \( C \).

- A special case of such a quadratic form is the covariance matrix, which we obtain when \( C = n^{-1}(I - n^{-1}11') \) (see proof on board; \( H = I - n^{-1}11' \) is called the centering matrix).

#### Definition

- If \( M(p \times p) \) can be written as \( X'X \) where \( X(m \times p) \) is a data matrix from \( N_p(0, \Sigma) \), then \( M \) is said to have a \( p \)-variate Wishart distribution with scale matrix \( \Sigma \) and \( m \) degrees of freedom. We write \( M \sim W_p(\Sigma, m) \). When \( \Sigma = I_p \), then the distribution is said to be in standard form.

- Note: The Wishart distribution is a multivariate generalization of the \( \chi^2 \) distribution: when \( p = 1 \), the \( W_1(\sigma^2, m) \) distribution is given by \( x'x \), where \( x \in \mathbb{R}^m \) contains i.i.d. \( N_1(0, \sigma^2) \) variables. Hence, \( W_1(\sigma^2, m) = \sigma^2 \chi^2_m \).
Properties

- (Th 3.4.1 of MKB) If $M \sim W_p(\Sigma, m)$ and $B$ is a $p \times q$ matrix, then $B'MB \sim W_q(B'\Sigma B, m)$ (see proof on board).
- (Cor 3.4.1.1 of MKB) Diagonal submatrices of $M$ (square submatrices of $M$ whose diagonal corresponds to the diagonal of $M$) have a Wishart distribution.
- (Cor 3.4.1.2 of MKB) $\Sigma^{-1/2}M\Sigma^{-1/2} \sim W_p(I, m)$
- (Cor 3.4.1.3 of MKB) If $M \sim W_p(I, m)$ and $B(p \times q)$ satisfies $B'B = I_q$, then $B'MB \sim W_q(I, m)$
- (Cor 3.4.2.1 of MKB) The $i$th diagonal element of $M$, $m_{ii}$, has a $\sigma_i^2\chi^2_m$ distribution (where $\sigma_i^2$ is the $i$th diagonal element of $\Sigma$).

All these corollaries follow by choosing particular values of $B$ in Th 3.4.1.

Properties

- (Th 3.4.3 of MKB) If $M_1 \sim W_p(\Sigma, m_1)$ and $M_2 \sim W_p(\Sigma, m_2)$, and if $M_1$ and $M_2$ are independent, then $M_1 + M_2 \sim W_p(\Sigma, m_1 + m_2)$ (see proof on board).

Properties

- (Th 3.4.4 of MKB) If $X(n \times p)$ is a data matrix from $N_p(0, \Sigma)$ and $C(n \times n)$ is a symmetric matrix, then
  - $X'CX$ has the same distribution as a weighted sum of independent $W_p(\Sigma, 1)$ matrices, where the weights are eigenvalues of $C$;
  - $X'CX$ has a Wishart distribution if $C$ is idempotent. In this case $X'CX \sim W_p(\Sigma, r)$ where $r = tr(C) = rank(C)$;
  - If $S = n^{-1}X'HX$ is the sample covariance matrix, then $nS \sim W_p(\Sigma, n-1)$.

(See proof on board)

Hotelling’s $T^2$ distribution

Definition

- If $\alpha$ can be written as $md'M^{-1}d$ where $d$ and $M$ are independently distributed as $N_p(0, I)$ and $W_p(I, m)$, then we say that $\alpha$ has the Hotelling $T^2$ distribution with parameters $p$ and $m$. We write $\alpha \sim T^2(p, m)$.

- Hotelling’s $T^2$ distribution is a generalization of the student $t$-distribution. If $x \sim t_m$, then $x^2 \sim T^2(1, m)$. 
One-sample $T^2$ test

- (Th 3.5.1 of MKB) If $x$ and $M$ are independently distributed as $N_p(\mu, \Sigma)$ and $W_p(\Sigma, m)$ then $m(x - \mu)'M^{-1}(x - \mu) \sim T^2(p, m)$ (see proof on board).

- (Cor 3.5.1.1 of MKB) If $\bar{x}$ and $S$ are the mean vector and covariance matrix of a sample of size $n$ from $N_p(\mu, \Sigma)$, and $S_u = (n/(n-1))S$, then

$$T^2_1 = (n-1)(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) = n(\bar{x} - \mu)S_u^{-1}(\bar{x} - \mu)$$

has a $T^2(p, n-1)$ distribution (see proof on board).

Relationship with $F$ distribution

- The $T^2$ distribution is not readily available in R. But the $T^2$ distribution is closely related to the $F$-distribution:

(Th 3.5.2 of MKB, without proof)

$$T^2(p, m) = \left\{ mp/(m - p + 1) \right\} F_{p, m-p+1}.$$ 

- (Cor 3.5.2.1 of MKB) If $\bar{x}$ and $S$ are the mean and covariance of a sample of size $n$ from $N_p(\mu, \Sigma)$, then $\frac{n-p}{(n-1)p}T^2_1$ has a $F_{p, n-p}$ distribution.

Mahalonobis distance

- The so-called Mahalonobis distance between two populations with means $\mu_1$ and $\mu_2$ and common covariance matrix $\Sigma$ is given by $\Delta$, where

$$\Delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$$

In other words, $D$ is the euclidian distance between the re-scaled vectors $\Sigma^{-1/2}\mu_1$ and $\Sigma^{-1/2}\mu_2$.

- The sample version of the Mahalonobis distance, $D$, is defined by

$$D^2 = (\bar{x}_1 - \bar{x}_2)'S_u^{-1}(\bar{x}_1 - \bar{x}_2),$$

where $S_u = (n_1S_1 + n_2S_2)/(n - 2)$, $\bar{x}_i$ is the sample mean of sample $i$, $n_i$ is the sample size of sample $i$, and $n = n_1 + n_2$. 

23 / 30

24 / 30

25 / 30
Two-sample $T^2$ test

- (Th 3.6.1 of MKB, without proof) If $X_1$ and $X_2$ are independent data matrices, and if the $n_i$ rows of $X_i$ are i.i.d. $N_p(\mu_i, \Sigma_i)$, $i = 1, 2$, then when $\mu_1 = \mu_2$ and $\Sigma_1 = \Sigma_2$,

$$T^2 = \frac{n_1 n_2}{n} D^2$$

has a $T^2(p, n - 2)$ distribution.

- Corollary: $\frac{n_1 n_2}{(n - 2)p} T^2$ has a $F_{p, n - 1 - p}$ distribution.

Final remarks

Assumptions for one-sample $T^2$ test

- We have a simple random sample from the population
- The population has a multivariate normal distribution

Assumptions for two-sample $T^2$ test

- We have a simple random sample from each population
- In each population the variables have a multivariate normal distribution
- The two populations have the same covariance matrix

Multivariate test versus univariate tests

See board