Session 6: Joint Normal Distribution, Multivariate Point Estimation, Generalized Variance, Hotelling's $T^2$ Test

Lecture Notes by: J. René van Dorp

www.seas.gwu.edu/~dorpjr

1 Department of Engineering Management and Systems Engineering, School of Engineering and Applied Science, The George Washington University, 1776 G Street, N.W., Suite 110, Washington D.C. 20052. E-mail: dorpj@gwu.edu.
Perhaps the joint normal distribution is a good model to describe the joint variation in height and weight of our multivariate sample of height and weight combinations of 20 women.
STATISTICAL REVIEW  

Joint Normal Distribution

- Probability density function of a bivariate normal distribution:

\[ X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MVN(\mu, \Sigma), \text{ Mean Vector: } \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \]

Covariance Matrix: \( \Sigma = \begin{pmatrix} \sigma_1^2 & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \sigma_2^2 \end{pmatrix} \)

\[
\begin{align*}
f(x, y) &= \frac{1}{\sqrt{2\pi|\Sigma|}} \exp \left[ (x - \mu)\Sigma^{-1}(x - \mu) \right] \\
\end{align*}
\]

- Note that:

\[
\Sigma = E[(X - \mu)(X - \mu)^T] = (n \times 1\text{-matrix}) \cdot (1 \times n\text{-matrix}) = (n \times n\text{-matrix})
\]

- How do we estimate the mean vector \( \mu \) and the variance covariance matrix \( \Sigma \)
• Let $X_1, X_2, \ldots, X_n$ be a random sample from a joint distribution with mean vector $\mu$ and covariance matrix $\Sigma$, where the vectors $X_i, i = 1, \ldots, n$ are independent. Then

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is an unbiased estimator of the mean vector $\mu$ and its variance covariance matrix is $\frac{1}{n} \Sigma$. In other words:

$$E[\overline{X}] = \mu, \quad Cov[\overline{X}] = E[(\overline{X} - \mu)(\overline{X} - \mu)^T] = \frac{1}{n} \Sigma$$

For convenience we shall from hereon write vectors only in bold font and not underline them anymore.

• Before we prove this result, recall the univariate case where $\overline{X}$ (obtained from an i.i.d. random sample $X_1, \ldots, X_n$ with mean $\mu$ and variance $\sigma^2$) is an unbiased estimator of the mean $\mu$ with variance $\sigma^2/n$
• Let \( \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n \) be a random sample from a joint distribution with mean vector \( \mathbf{\mu} \) and covariance matrix \( \Sigma \), where the vectors \( \mathbf{X}_i, i = 1, \ldots, n \) are independent. Then

\[
S = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_i - \mathbf{\bar{X}})(\mathbf{X}_i - \mathbf{\bar{X}})^T
\]

is an unbiased estimator of the variance covariance matrix \( \Sigma \), i.e. \( E[S] = \Sigma \)

• Recall the unbiased estimator for \( \sigma^2 \) in the univariate case

\[
S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2
\]

Example: Height-Weight data

\[
\mathbf{\bar{X}} = \begin{pmatrix} 62.85 \\ 123.60 \end{pmatrix}, \quad S = \begin{pmatrix} 10.87 & 44.52 \\ 44.52 & 242.46 \end{pmatrix}
\]
It is occasionally useful (especially when we are considering more than two variables) to obtain a single measure of linear dependence between a larger set of variables (more than 2). The most common measure for this purpose is the matrix determinant $|\Sigma|$ of the variance-covariance matrix $\Sigma$.

$$\Sigma = \begin{pmatrix} V[X_1] & Cov[X_1, X_2] \\ Cov[X_1, X_2] & V[X_2] \end{pmatrix}, \quad |\Sigma| = V[X_1]V[X_2] - Cov^2[X_1, X_2]$$

In the 2-dimensional case we use $Cov[X_1, X_2]$ to describe the linear dependence between $X_1$ and $X_2$. Note that:

$$|\Sigma| = V[X_1]V[X_2] \iff Cov[X_1, X_2] = 0$$

The highest value of $|\Sigma|$ is attained when $X_1$ and $X_2$ are uncorrelated and the smallest when $X_2 = aX_1 + b$ for some constants $a$ and $b$. Hence, the higher the value of $|\Sigma|$ the less linearly dependent $X_1$ and $X_2$. This interpretation carries over to more than two dimension as well. The measure $|\Sigma|$ is called the Generalized Variance. (which in my opinion is a misnomer).
Example: Height-Weight data

\[ S = \begin{pmatrix} 10.87 & 44.52 \\ 44.52 & 242.46 \end{pmatrix}, \]

\[ |S| = (10.87 \times 242.46) - (44.52 \times 44.52) \approx 654.17 \]
STATISTICAL REVIEW

Geometric Interpretation

- The EXCEL function MDETERM calculates $|\Sigma|$ for more than 2 variables.

- The value of $|\Sigma|$ may be dominated by one variable due to a difference in scale between the two variables (leading to a comparatively large variance of one of the variables). The same is true for the value of $\text{Cov}(X_1, X_2)$.

- To resolve this issue it is not uncommon to first standardize the data and next calculate the variance-covariance matrix $R$ and $|R|$ of the standardized data set. We denote this matrix by $R$ since $R$ is the correlation matrix of the original data set. Similarly, we look at $\rho(X_1, X_2)$ (i.e. the standardized covariance) to obtain a measure of linear dependence.

Example: Height-Weight data

$$R = \begin{pmatrix} 1 & 0.867 \\ 0.867 & 1 \end{pmatrix},$$

$$|R| = 1 - (0.867)^2 \approx 0.248$$
• It can be shown that:

\[ |\mathbf{S}| = \left\{ \prod_{i=1}^{n} s_{ii} \right\} |\mathbf{R}| = \left\{ \prod_{i=1}^{n} \text{Var}[X_i] \right\} |\mathbf{R}| \]

(which is possibly why they named $|\mathbf{S}|$ the generalized variance.)

Example: Height-Weight data

\[ |\mathbf{S}| = (10.87 \times 242.46) |\mathbf{R}| = (10.87 \times 242.46) \times 0.248 \approx 654.17 \]
Note that:

\[ R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \iff |R| = 1 \iff \rho(X_1, X_2) = 0 \]

The highest value of \(|R|\) is attained when \(X_1\) and \(X_2\) are **uncorrelated** and the smallest when \(X_2 = aX_1 + b\) for some constants \(a\) and \(b\). Hence, the higher the value of \(|R|\) the **less linearly dependent** \(X_1\) and \(X_2\) are. The above interpretation of \(|R|\) carries over to **more than two dimensions** as well.

**BE CAREFULL!** The **closer** the value of \(|R|\) to 0 the **higher** the degree of linear dependence (referred to as **collinearity** in dimensions higher than 2). When the value of \(|R| = 1\) there is **no collinearity** present within the data.

Observe that the direction of \(|R|\) is opposite to that of the correlation coefficient \(\rho(X_1, X_2)\).

\[ \rho(X_1, X_2) = \pm 1 \iff X_2 = aX_1 + b \text{ for some } a, b \in \mathbb{R} \]

Also \(|R| > 0\), whereas \(-1 \leq \rho(X_1, X_2) \leq 1\).
Recall univariate $t$-hypothesis test

$$H_0 : \mu = \mu_0, \ H_1 : \mu \neq \mu_0 \Rightarrow \text{Reject } H_0 \text{ when } |T| \geq t_{n-1,1-\alpha/2}$$
Reject $H_0$ when $|T| \geq t_{n-1,1-\frac{\alpha}{2}} \iff$ Reject $H_0$ when $T^2 \geq t_{n-1,1-\frac{\alpha}{2}}$

\[
T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \iff T^2 = n(\bar{x} - \mu_0)(s^2)^{-1}(\bar{x} - \mu_0)
\]

- Let $X_1, \ldots, X_n$ now be an i.i.d. $p$-dimensional sample from a $MVN(\mu, \Sigma)$. Note that the vectors are i.i.d. but the elements of the vectors are not. Then with

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad S = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T
\]

we can define the Hotelling $T^2$-statistic

\[
T^2 = n(\bar{X} - \mu_0)^T S^{-1}(\bar{X} - \mu_0)
\]

which is a direct generalization of the T-statistic of a univariate normal i.i.d. sample.
STATISTICAL REVIEW

Hotelling $T^2$ Test

Hotelling showed that:

$$\frac{n - p}{(n - 1)p} T^2 \sim F_{p,n-p}$$

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0, \ H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0 \Rightarrow \text{Reject } H_0 \text{ when } \frac{n - p}{(n - 1)p} T^2 \geq F_{p,n-p,1-\alpha}$$

Example: Height-Weight data

$$\overline{X} = \begin{pmatrix} 62.85 \\ 123.60 \end{pmatrix}, \ S = \begin{pmatrix} 10.87 & 44.52 \\ 44.52 & 242.46 \end{pmatrix}, \ \boldsymbol{\mu}_0 = \begin{pmatrix} 62 \\ 120 \end{pmatrix}, \ p = 2, n = 20$$

$$T^2 = 20 \begin{pmatrix} .85 & 3.60 \end{pmatrix} \begin{pmatrix} 10.87 & 44.52 \\ 44.52 & 242.46 \end{pmatrix}^{-1} \begin{pmatrix} .85 \\ 3.60 \end{pmatrix} \approx 1.33$$

Significance Level: $\alpha = 10\%$, \[ \frac{18}{19 \times 2} T^2 \approx 0.63 < F_{2,18,0.9} \approx 2.624 \]

$\Rightarrow$ Fail to Reject $H_0$
Recall Two Sample Mean Test

The **two-sample t test** for testing \( H_0 : \mu_1 - \mu_2 = \Delta_0 \) is as follows:

**Test statistic value:**

\[
t_0 = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}
\]

**Degrees of freedom** \( \nu \):

\[
\nu = \frac{\left[ \frac{s_1^2}{n} + \frac{s_2^2}{m} \right]^2}{\frac{(s_1^2/n)^2}{n-1} + \frac{(s_2^2/m)^2}{m-1}}
\]

<table>
<thead>
<tr>
<th>Alternative Hypothesis</th>
<th>Rejection Regions for significance ( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_1 : \mu_1 - \mu_2 &gt; \Delta_0 )</td>
<td>( t_0 &gt; t_{\nu,1-\alpha} ) (upper-tailed)</td>
</tr>
<tr>
<td>( H_1 : \mu_1 - \mu_2 &lt; \Delta_0 )</td>
<td>( t_0 &lt; - t_{\nu,1-\alpha} ) (lower-tailed)</td>
</tr>
<tr>
<td>( H_1 : \mu_1 - \mu_2 \neq \Delta_0 )</td>
<td>( t_0 &gt; t_{\nu,1-\alpha/2} ) or ( t_0 &lt; - t_{\nu,1-\alpha/2} ) (two-tailed)</td>
</tr>
</tbody>
</table>

\( p \)-values can be constructed in a similar fashion as before.
STATISTICAL REVIEW  Hotelling Two Sample Mean Test

Let \( X_{11}, \ldots, X_{1n} \) now be an i.i.d. \( p \)-dimensional sample from population 1:

\[
MVN(\mu_1, \Sigma_1).
\]

Let \( X_{21}, \ldots, X_{2m} \) now be an i.i.d. \( p \)-dimensional sample from population 2:

\[
MVN(\mu_2, \Sigma_2).
\]

Then with

\[
\bar{X}_1 = \frac{1}{n} \sum_{i=1}^{n} X_{1i}, \quad S_1 = \frac{1}{n-1} \sum_{i=1}^{n} (X_{1i} - \bar{X}_1)(X_i - \bar{X}_1)^T
\]

\[
\bar{X}_2 = \frac{1}{m} \sum_{i=1}^{m} X_{2i}, \quad S_2 = \frac{1}{m-1} \sum_{i=1}^{m} (X_{2i} - \bar{X}_2)(X_i - \bar{X}_2)^T
\]

(The first subscript \(-1\) or \(-2\) — denotes the population)

we can define another \( T^2 \)-statistic if \( X_{11}, \ldots, X_{1n} \) and \( X_{21}, \ldots, X_{2m} \) are i.d:
The **two-multivariate** $T^2$ test for testing $H_0 : \mu_1 - \mu_2 = \Delta_0$ is as follows:

$$ T^2 = \left[ \overline{X}_1 - \overline{X}_2 - \Delta_0 \right]^T \left[ \left( \frac{1}{n} + \frac{1}{m} \right) S_{\text{pooled}} \right]^{-1} \left[ \overline{X}_1 - \overline{X}_2 - \Delta_0 \right], $$

where

$$ S_{\text{pooled}} = \frac{(n - 1)S_1 + (m - 1)S_2}{n + m - 2}, $$

which is a direct generalization of the two sample T-statistic. If sample sizes $n$ and $m$ are small, one additional assumptions is needed to be able to execute calculation of confidence intervals and to conduct hypothesis tests, being

$$ \Sigma_1 = \Sigma_2. $$

In that case:

$$ \frac{(n + m - p - 1)}{(n + m - 2)p} T^2 \sim F_p, n + m - p - 1 $$
Example Wisconsin Power Data:
Samples of sizes $n = 45$ and $m = 55$ were taken of Wisconsin homeowners with and without airconditioning, respectively. (Data courtesy of Statistical Laboratory, University of Wisconsin). Two measurements of electrical usage (in kilowatt hours) were considered. The first is a measure of total on-peak consumptions ($X_1$) during July 1977 and the second is a measure of total off-peak consumption ($X_2$) during July 1977. (The off-peak consumption is higher than the on-peak consumption because there are more off-peak hours in a month). The resulting summary statistics are:

With AirCo: $\bar{x}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$, $S_1 = \begin{pmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{pmatrix}$, $n = 45$

Without AirCo: $\bar{x}_2 = \begin{pmatrix} 130.0 \\ 355.0 \end{pmatrix}$, $S_2 = \begin{pmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{pmatrix}$, $m = 55$

$$S_{pooled} = \frac{(n-1)S_1 + (m-1)S_2}{n + m - 2} = \begin{pmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{pmatrix}$$
STATISTICAL REVIEW  Hotelling Two Sample Mean Test

- We want to test: \( H_0 : (\mu_1 - \mu_2)^T = (0 \quad 0)^T = \Delta_0^T \)

\[
T^2 = [\bar{X}_1 - \bar{X}_2 - \Delta_0]^T \left[ \left( \frac{1}{n} + \frac{1}{m} \right) S_{pooled} \right]^{-1} [\bar{X}_1 - \bar{X}_2 - \Delta_0]
\]

\[
= \begin{pmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{pmatrix}^T \begin{pmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{pmatrix}^{-1} \begin{pmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{pmatrix}
\]

\[
= \begin{pmatrix} 74.4 \\ 201.6 \end{pmatrix}^T \begin{pmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{pmatrix}^{-1} \begin{pmatrix} 74.4 \\ 201.6 \end{pmatrix} \approx 16.07
\]

\[
\frac{(n + m - p - 1)}{(n + m - 2)p} T^2 = \frac{(45 + 55 - 3)}{(45 + 55 - 2)2} 16.07 \approx 7.95 > F_{2,97,0.95} \approx 3.09
\]

**Conclusion:** Reject \( H_0 \) (i.e. there is a difference between airconditioning and no airconditioning consumption).
In the case that sample sizes are large, the assumption $\Sigma_1 = \Sigma_2$ may be relaxed to allow for $\Sigma_1 \neq \Sigma_2$. In that case:

$$T^2 = [\bar{X}_1 - \bar{X}_2 - \Delta_0]^T \left[ \left( \frac{S_1}{n} + \frac{S_2}{m} \right) \right]^{-1} [\bar{X}_1 - \bar{X}_2 - \Delta_0]$$

and

$$T^2 \sim \chi_p^2$$

Example Wisconsin Power Data:

$$\bar{x}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}, \quad S_1 = \begin{pmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{pmatrix}, \quad n = 45$$

$$\bar{x}_2 = \begin{pmatrix} 130.0 \\ 355.0 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{pmatrix}, \quad m = 55$$

$$\frac{S_1}{n} + \frac{S_2}{m} = \begin{pmatrix} 443.0 & 868.9 \\ 868.9 & 2572.2 \end{pmatrix}$$
STATISTICAL REVIEW  Hotelling Two Sample Mean Test

\[ T^2 = [\overline{X}_1 - \overline{X}_2 - \Delta_0]^T \left[ \left( \frac{1}{n} + \frac{1}{m} \right) S_{pooled} \right]^{-1} [\overline{X}_1 - \overline{X}_2 - \Delta_0] \]

\[ = \begin{pmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{pmatrix}^T \begin{pmatrix} 443.0 & 868.9 \\ 868.9 & 2572.2 \end{pmatrix}^{-1} \begin{pmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{pmatrix} \]

\[ = \begin{pmatrix} 74.4 \\ 201.6 \end{pmatrix}^T \begin{pmatrix} 10^{-4} & 59.874 & -20.080 \\ -20.080 & 10.519 & 74.4 \\ 201.6 & \end{pmatrix} \approx 15.66 \]

\[ T^2 \approx 15.66 > \chi^2_{2,0.95} \approx 5.99 \]

Conclusion: Reject \( H_0 \) (i.e. there is a difference between airconditioning and no airconditioning consumption).