EMSE 4765: DATA ANALYSIS
For Engineers and Scientists

Session 7: Multivariate Inference, Matrix Determinant,
Multivariate Hotelling's T^2 Test for equality of Mean Vectors,
Mutlivariate Box's M-Test for equality of Covariance Matrices

Version: 2/22/2021

Lecture Notes by: J. René van Dorp[1]

www.seas.gwu.edu/~dorpjr

[1] Department of Engineering Management and Systems Egineering, School of Engineering and Applied Science, The George Washington University, 800 22nd Street, N.W., Suite 2800, Washington D.C. 20052. E-mail: dorpjr@gwu.edu.

Heights and Weights of 20 Individuals

X_1	X_2	X_{d1}	X_{d2}	X_{s1}	X_{s2}
57	93	-5.85	-30.60	-1.77427	-1.96516
58	110	-4.85	-13.60	-1.47098	-0.87341
60	99	-2.85	-24.60	-0.86439	-1.57984
59	111	-3.85	-12.60	-1.16768	-0.80918
61	115	-1.85	-8.60	-0.56109	-0.55230
60	122	-2.85	-1.60	-0.86439	-0.10275
62	110	-0.85	-13.60	-0.25780	-0.87341
61	116	-1.85	-7.60	-0.56109	-0.48808
62	122	-0.85	-1.60	-0.25780	-0.10275
63	128	0.15	4.40	0.04549	0.28257
62	134	-0.85	10.40	-0.25780	0.66790
64	117	1.15	-6.60	0.34879	-0.42386
63	123	0.15	-0.60	0.04549	-0.03853
65	129	2.15	5.40	0.65208	0.34679
64	135	1.15	11.40	0.34879	0.73212
66	128	3.15	4.40	0.95538	0.28257
67	135	4.15	11.40	1.25867	0.73212
66	148	3.15	24.40	0.95538	1.56699
68	142	5.15	18.40	1.56197	1.18167
69	155	6.15	31.40	1.86526	2.01654

X_1 Height
X_2 Weight
X_{d1} Height: mean centered
X_{d2} Weight: Mean Centered
X_{s1} Height: Standardized
X_{s2} Weight: Standardized

	X_1	X_2
Mean	62.85	123.60
St. Dev.	3.30	15.57

- We can first create two column vectors that take the values of **the mean centered variables for height and weight.**

- We can second create two column vectors that take the values of **the standardized variables for height and weight** (with mean 0 and standard deviation 1).

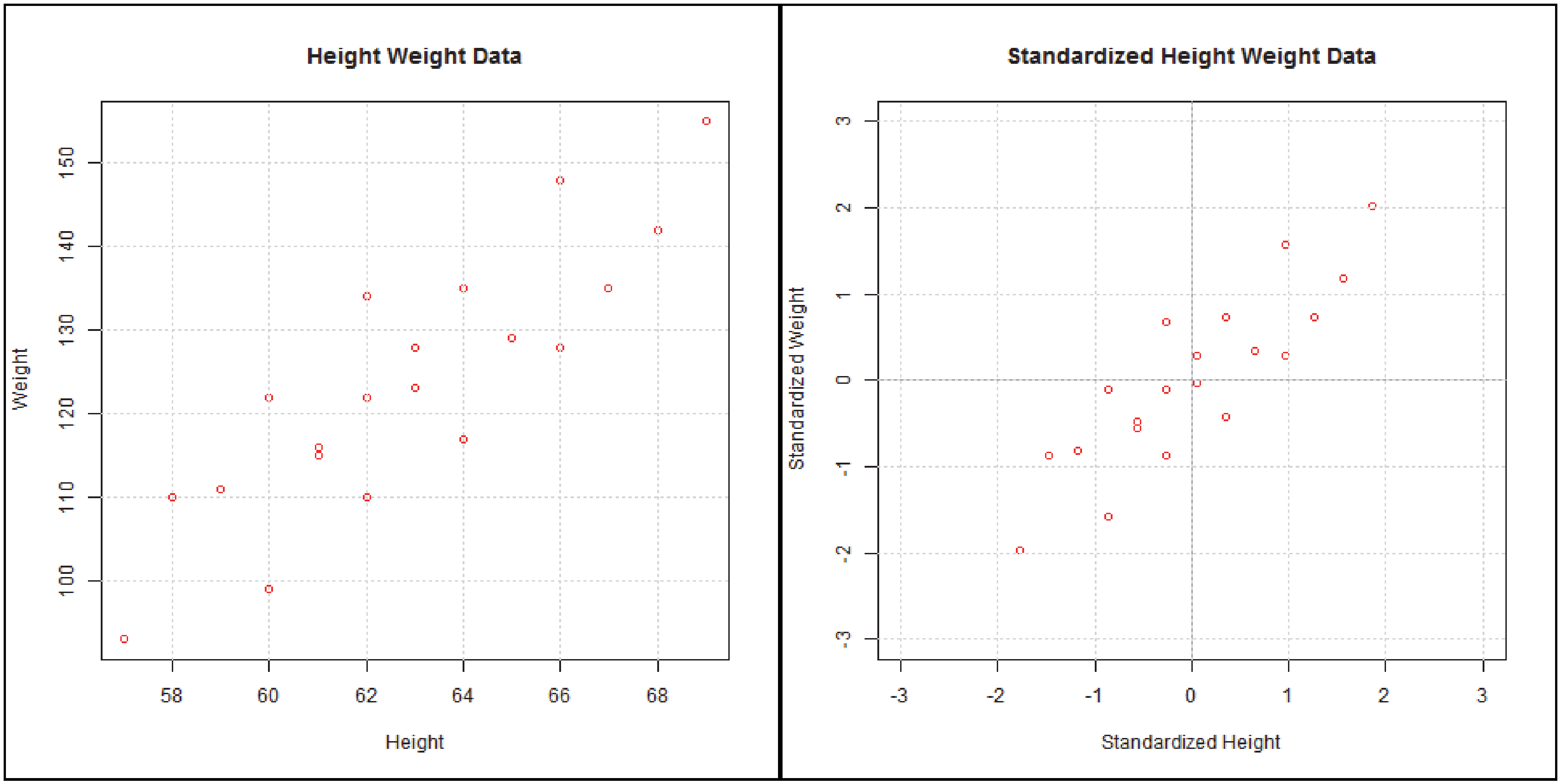
- Together, these two standardized column vectors form a matrix given by:

$$\boldsymbol{X}_s = \left(\underline{x}_{s1} \quad \underline{x}_{s2} \right)$$

- **Each row of this matrix corresponds to one object (person)** measured on **two different characteristics (height, weight).**

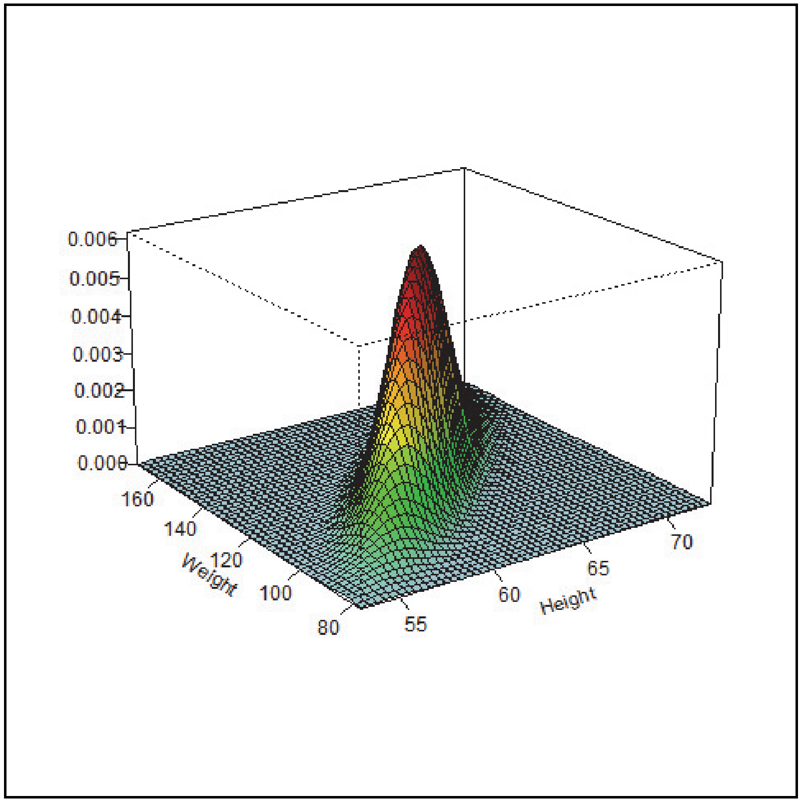
- By displaying all points in the same coordinate system, one can clearly visualize the pattern of observations and the position of each point relative to one another. **This type of representation is known as a scatter plot.**

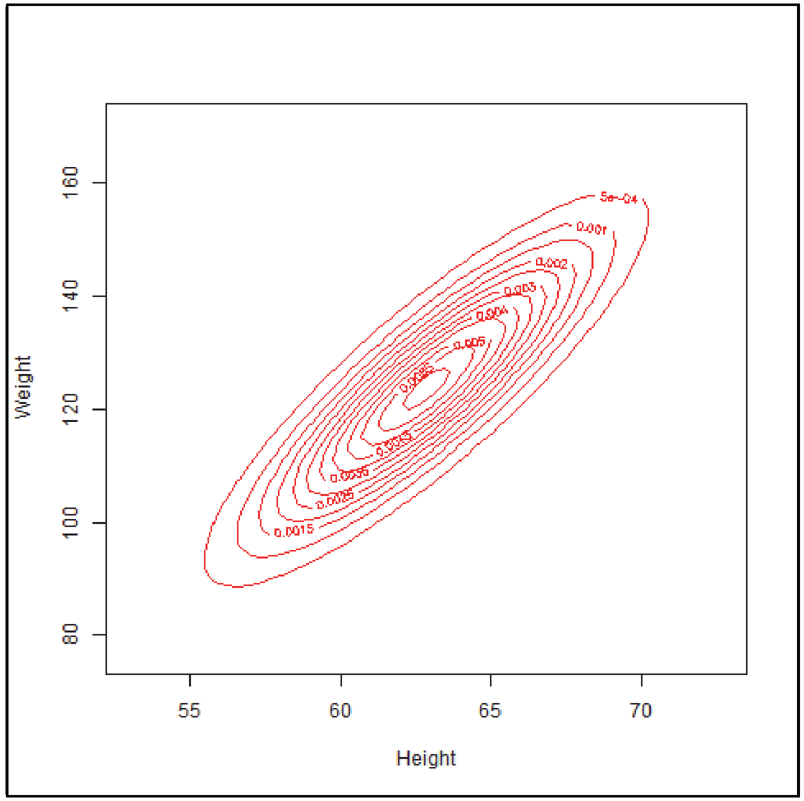
Scatter Plot of Height and Weight of 20 Individuals



Conclusion: Height and Weight are positively correlated.

- 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 individuals.





- Hence now we need to estimate a mean vector and a covariance matrix.

- **Probability density function** of a bivariate normal distribution :

$$\underline{\boldsymbol{X}} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \textbf{ Mean Vector: } \underline{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

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

$$f(x,y) = \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}|}} exp\Big[(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\Big]$$

- **Definition Covariance Matrix**:

$$\boldsymbol{\Sigma} = E[(\underline{\boldsymbol{X}} - \underline{\boldsymbol{\mu}})(\underline{\boldsymbol{X}} - \underline{\boldsymbol{\mu}})^T] \ (\textbf{Recall: } Var(X) = E[(X-\mu)^2], \mu = E[X])$$

$$(n \times 1\text{-matrix}) \cdot (1 \times n\text{-matrix}) = (n \times n\text{-matrix})$$

- How do we estimate the mean vector $\underline{\boldsymbol{\mu}}$ and the variance covariance matrix $\boldsymbol{\Sigma}$

- Let $\underline{\boldsymbol{X}}_{\boldsymbol{1}}, \underline{\boldsymbol{X}}_{\boldsymbol{2}}, \ldots, \underline{\boldsymbol{X}}_{\boldsymbol{n}}$ be a random sample from **a joint distribution (with dimension m)** with mean vector $\underline{\boldsymbol{\mu}}$ and covariance matrix $\boldsymbol{\Sigma}$, where the m-dimensional vectors $\underline{\boldsymbol{X}}_{\boldsymbol{i}}, i = 1, \ldots, n$ are independent. Then

$$\overline{\underline{\boldsymbol{X}}} = \frac{1}{n}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \underline{\boldsymbol{X}}_i$$

is **an unbiased estimator of the mean vector $\underline{\boldsymbol{\mu}}$** and **its variance covariance matrix is $\frac{1}{n}\boldsymbol{\Sigma}$**. In other words: $E[\overline{\underline{\boldsymbol{X}}}] = \underline{\boldsymbol{\mu}}$. **Also:**

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

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

- **See the analogy** to the univariate case **with $E[\overline{X}] = \mu$ and $V[\overline{X}] = \sigma^2/n$** (Given random sample $(X_1, \ldots, X_n)$, $X_i \sim X$, $E[X] = \mu$, $V[X] = \sigma^2$).

- Let $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ be a random sample from **a joint distribution** with dimension m, $(m \times 1)$ mean vector $\boldsymbol{\mu}$ and $(m \times m)$ covariance matrix $\boldsymbol{\Sigma}$, where the $(m \times 1)$ vectors $\boldsymbol{X_i}, i = 1, \ldots, n$ are independent. Then

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X_i} - \overline{\boldsymbol{X}})(\boldsymbol{X_i} - \overline{\boldsymbol{X}})^T \text{ is an } (m \times m) \text{ matrix and}$$

is **an unbiased estimator of the variance covariance matrix** $\boldsymbol{\Sigma}$, i.e. $E[\boldsymbol{S}] = \boldsymbol{\Sigma}$, where the expectatios are taken elements wise.

- **See the analogy** to the unbiased estimator for σ^2 in the univariate case

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Example: Height-Weight data

$$\overline{\boldsymbol{X}} = \begin{pmatrix} 62.85 \\ 123.60 \end{pmatrix}, \quad \boldsymbol{S} = \begin{pmatrix} 10.87 & 44.52 \\ 44.52 & 242.46 \end{pmatrix}$$

- In the situation of **more than two variables** it is usefull to obtain **a single measure of linear dependence** between them. **The most common measure for this purpose** is **the matrix determinant** $|\mathbf{\Sigma}|$ of the variance-covariance matrix $\mathbf{\Sigma}$. To understand why, consider the 2-dimensional case:

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

- **Thus, in the 2-dimensional case the matrix determinant** is **directly linked to the covariance between X_1 and X_2, i.e. $Cov[X_1, X_2]$** and:

$$Cov[X_1, X_2] = 0 \; \Leftrightarrow |\mathbf{\Sigma}| = V[X_1]V[X_2]$$

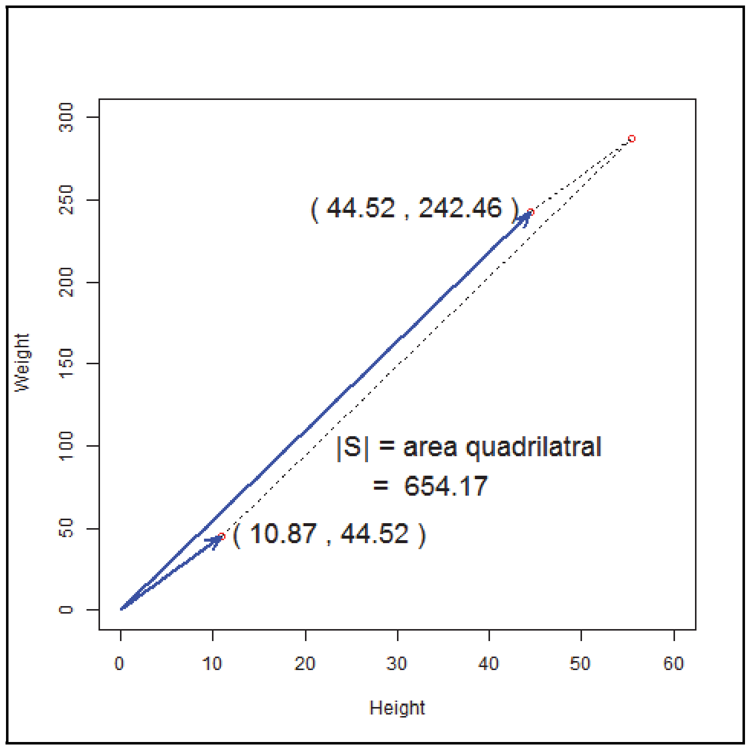
- **However,** the **highest value** of $|\mathbf{\Sigma}|$ is attained when $Cov[X_1, X_2] = 0$ and **the smallest value is attainend** when $X_2 = aX_1 + b$ for some constants a and b, i.e. when X_2 is a linear transformation of X_1.

- Hence, **the higher the value** of $|\mathbf{\Sigma}|$ the **less linearly dependent** X_1 and X_2 are.

Example: Height-Weight data

$$\boldsymbol{S} = \begin{pmatrix} 10.87 & 44.52 \\ 44.52 & 242.46 \end{pmatrix},$$

$$|\boldsymbol{S}| = (10.87 \times 242.46) - (44.52 \times 44.52) \approx 654.17$$

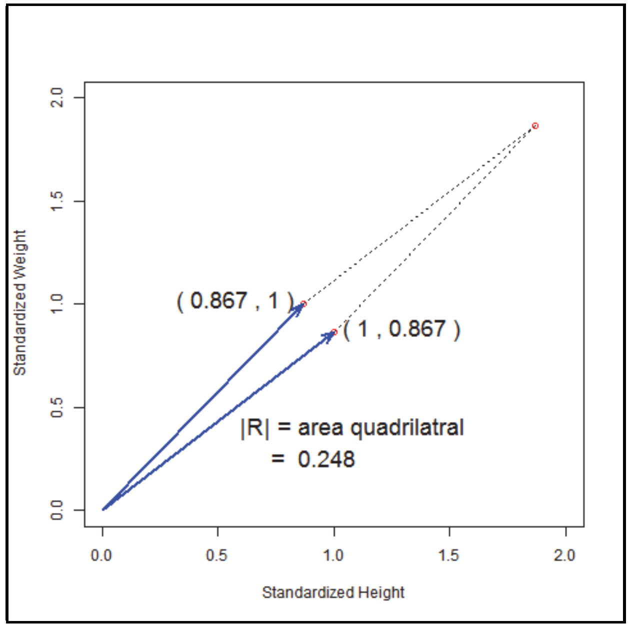


- The EXCEL function MDETERM calculates $|\boldsymbol{\Sigma}|$ for more than 2 variables.
- The value of $|\boldsymbol{\Sigma}|$ **may be dominated by one variable** due to **a difference in scale** between the two variables (leading to a comparatively large variance of one of the variables). **The same is true for the value of $Cov(X_1, X_2)$.**
- **To obtain a multivariate degree of dependence measure one first standardizes the data** and one calculates the correlation matrix $\boldsymbol{R}$ and its determinant $|\boldsymbol{R}|$. That matrix $\boldsymbol{R}$ is also **the correlation matrix of the original data set.** In two dimensions $\rho(X_1, X_2)$ (i.e. the standardized covariance) measures the degree of linear dependence.

Example: Height-Weight data

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

$$|\boldsymbol{R}| = 1 - (0.867)^2 \approx 0.248$$



- It can be shown that :

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

(which is why $|\boldsymbol{S}|$ also is referred to as the generalized variance.)

Example: Height-Weight data

$$|\boldsymbol{R}| \approx 0.248, |\boldsymbol{S}| = (10.87 \times 242.46) \times 0.248 \approx 654.17$$

- Note that:

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \Leftrightarrow |\boldsymbol{R}| = 1 \Leftrightarrow \rho(X_1, X_2) = 0$$

The highest value of $|\boldsymbol{R}|$ is attained when $\rho(X_1, X_2) = 0$ and **the smallest when** $\boldsymbol{X_2 = aX_1 + b}$ for some constants a and b, i.e. when $\rho(X_1, X_2) = 1$. Hence, **the higher the value** of $|\boldsymbol{R}|$ the **less linearly dependent** X_1 and X_2 are. The above interpretation of $|\boldsymbol{R}|$ carries over to **more than 2 dimensions**.

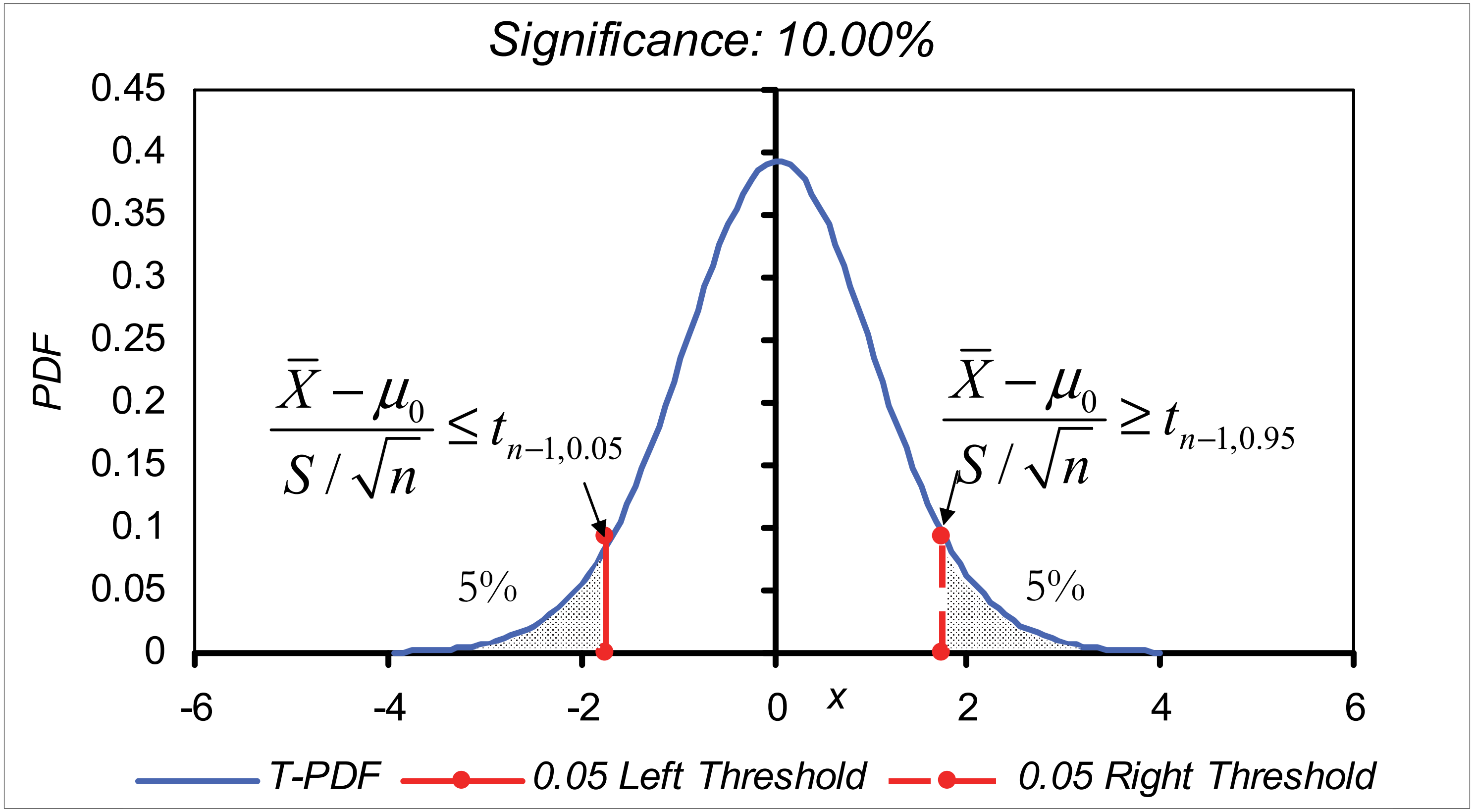
- **BE CAREFULL! The closer to 0** the value of $|\boldsymbol{R}|$ is, **the higher the degree of linear dependence** (referred to as **multi-colinearity in dimensions higher than 2**). When the value of $|\boldsymbol{R}| = 1$ there is **no colineariy** present within the data.

- Thus the direction of $|\boldsymbol{R}|$ is opposite to that of the correlation coefficient $\rho(X_1, X_2)$.

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

- Also $|\boldsymbol{R}| > 0$, whereas $-1 \leq \rho(X_1, X_2) \leq 1$.

Recall univariate t-hypothesis test



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

$$\text{Reject } H_0 \text{ when } |T| \geq t_{n-1,1-\frac{\alpha}{2}} \Leftrightarrow \text{Reject } H_0 \text{ when } T^2 \geq t^2_{n-1,1-\frac{\alpha}{2}}$$

$$T = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \Leftrightarrow T^2 = \frac{(\overline{x} - \mu_0)^2}{s^2/n} \Leftrightarrow T^2 = n(\overline{x} - \mu_0)(s^2)^{-1}(\overline{x} - \mu_0)$$

- Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be an $i.i.d.$ **p-dimensional** sample from a $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Thus the vectors are independent but **the elements of the vectors are not.** Then with **sample mean vector** and **sample variance covariance matrix:**

$$\overline{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i, \quad \boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})^T$$

we can now define **the Hotelling T^2-statistic for a multivariate normal sample**

$$T^2 = n(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)^T\boldsymbol{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)$$

which is **a direct generalization of the T-statistic for a univariate normal $i.i.d.$ sample.**

Hotelling showed that: $$\frac{n-p}{(n-1)p} \times T^2 \sim F_{p,n-p}$$

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0,\ H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu_0} \Rightarrow \text{Reject } H_0 \text{ when } \frac{n-p}{(n-1)p} T^2 \geq F_{p,n-p,1-\alpha}$$

Example: Height-Weight data

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

$$\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0 = \begin{pmatrix} .85 \\ 3.60 \end{pmatrix}, T^2 = 20(\,.85 \quad 3.60\,)\begin{pmatrix} 10.87 & 44.52 \\ 44.52 & 242.46 \end{pmatrix}^{-1}\begin{pmatrix} .85 \\ 3.60 \end{pmatrix} \approx 1.33$$

Significance Level: $\alpha = 10\%$, $\frac{18}{19 \times 2}T^2 \approx 0.63187 < F_{2,18,0.9} \approx 2.624$

p-value: $Pr\Big(\frac{n-p}{(n-1)p} T^2 > 0.63817\Big) \approx 0.543 \Rightarrow$ Fail to Reject H_0: $\boldsymbol{\mu}_0 = \begin{pmatrix} 62 \\ 120 \end{pmatrix}$

- **Recall the $100(1-\alpha)\%$ univariate confidence interval** for $\mu_1 - \mu_2$:

$$(\overline{x}-\overline{y}) \pm t_{n+m-2,1-\alpha/2} \times S_p\sqrt{\frac{1}{n}+\frac{1}{m}},\ S_p = \sqrt{\frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}}$$

This **two-sample t-test** for testing $H_0 : \mu_1 - \mu_2 = \Delta_0$ **assumes that the variances of the univariate normal random samples $(X_1, \ldots, X_n)$ are $(Y_1, \ldots, Y_m)$ the same:**

Test statistic value:

$$t_0 = \frac{\overline{x}-\overline{y}-\Delta_0}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

Alternative Hypothesis	**Rejection Regions for significance α**	
$H_1 : \mu_1 - \mu_2 > \Delta_0$	$t_0 > t_{n+m-2,1-\alpha}$	(upper-tailed)
$H_1 : \mu_1 - \mu_2 < \Delta_0$	$t_0 < -t_{n+m-2,1-\alpha}$	(lower-tailed)
$H_1 : \mu_1 - \mu_2 \neq \Delta_0$	$t_0 > t_{n+m-2,1-\alpha/2}$ or $t_0 < -t_{n+m-2,1-\alpha/2}$	(two-tailed)

p-values can be constructed in a similar fashion as before.

Let $\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n}$ now be an $i.i.d.$ p-dimensional sample from **population 1**:

$$MVN(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1}).$$

Let $\boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2m}$ now be an $i.i.d.$ p-dimensional sample from **population 2**:

$$MVN(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2}).$$

Then with **sample mean vectors and sample variance covariance matrices**

$$\overline{\boldsymbol{X}}_{\boldsymbol{1}} = \frac{1}{n}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{X_{1i}}, \;\; \boldsymbol{S_1} = \frac{1}{\boldsymbol{n}-1}\sum_{i=1}^{n}(\boldsymbol{X_{1i}} - \overline{\boldsymbol{X}}_{\boldsymbol{1}})(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_{\boldsymbol{1}})^T$$

$$\overline{\boldsymbol{X}}_2 = \frac{1}{m}\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{X_{2i}}, \;\; \boldsymbol{S_2} = \frac{1}{m-1}\sum_{i=1}^{m}(\boldsymbol{X_{2i}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}})(\boldsymbol{X_i} - \overline{\boldsymbol{X}}_{\boldsymbol{2}})^T$$

(The first subscipt, 1 or 2, denotes the population)

We can now define another T^2-statistic with $\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n}$ and $\boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2m}$:

The **two-sample multivariate T^2 statistic** for testing $H_0 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} = \boldsymbol{\Delta}_0$ is:

$$T^2 = [\overline{\boldsymbol{X}}_{\mathbf{1}} - \overline{\boldsymbol{X}}_{\mathbf{2}} - \boldsymbol{\Delta}_0]^T \left[\left(\frac{1}{n} + \frac{1}{m} \right) \boldsymbol{S_{pooled}} \right]^{-1} [\overline{\boldsymbol{X}}_{\mathbf{1}} - \overline{\boldsymbol{X}}_{\mathbf{2}} - \boldsymbol{\Delta}_0],$$

where the pooled variance covariance matrix is:

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

This is **a direct generalization of the two-sample T-statistic for univariate normal $\boldsymbol{i.i.d.}$ samples with the assumption that the variances are the same. Thus here too the following assumption** is required to be able to to conduct this hypothesis tests (and this assumption should be tested) being:

$$\boldsymbol{\Sigma_1 = \Sigma_2}.$$

In that case:

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

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:

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

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

$$\boldsymbol{S}_{pooled} = \frac{(n-1)\boldsymbol{S}_1 + (m-1)\boldsymbol{S}_2}{n+m-2} = \begin{pmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{pmatrix}$$

- We want to test: $H_0 : (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T = (\,0 \quad 0\,)^T = \boldsymbol{\Delta}_0^{\boldsymbol{T}}$

$$T^2 = [\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}} - \boldsymbol{\Delta}_0]^T \left[\left(\frac{1}{n} + \frac{1}{m}\right)\boldsymbol{S_{pooled}}\right]^{-1} [\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}} - \boldsymbol{\Delta}_0]$$

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

$$= \begin{pmatrix} 74.4 \\ 201.6 \end{pmatrix}^T \begin{pmatrix} 443.0 & 868.9 \\ 868.9 & 2572.2 \end{pmatrix}^{-1} \begin{pmatrix} 74.4 \\ 201.6 \end{pmatrix} \approx 16.07$$

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

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:

$$T^2 = [\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 - \boldsymbol{\Delta}_0]^T \left[\left(\frac{\boldsymbol{S}_1}{n} + \frac{\boldsymbol{S}_2}{m}\right)\right]^{-1} [\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 - \boldsymbol{\Delta}_0]$$

and

$$T^2 \sim \chi^2_p$$

Example Wisconsin Power Data:

$$\overline{\boldsymbol{x}}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}, \boldsymbol{S}_1 = \begin{pmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{pmatrix}, n = 45$$

$$\overline{\boldsymbol{x}}_2 = \begin{pmatrix} 130.0 \\ 355.0 \end{pmatrix}, \boldsymbol{S}_2 = \begin{pmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{pmatrix}, m = 55$$

$$\frac{\boldsymbol{S}_1}{n} + \frac{\boldsymbol{S}_2}{m} = \begin{pmatrix} 464.2 & 886.1 \\ 886.1 & 2642.1 \end{pmatrix}$$

$$T^2 = [\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 - \boldsymbol{\Delta}_0]^T \left[\frac{\boldsymbol{S}_1}{n} + \frac{\boldsymbol{S}_2}{m}\right]^{-1} [\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 - \boldsymbol{\Delta}_0]$$

$$= \begin{pmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{pmatrix}^T \begin{pmatrix} 464.2 & 886.1 \\ 886.1 & 2642.1 \end{pmatrix}^{-1} \begin{pmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{pmatrix}$$

$$= \begin{pmatrix} 74.4 \\ 201.6 \end{pmatrix}^T \begin{pmatrix} 464.2 & 886.1 \\ 886.1 & 2642.1 \end{pmatrix}^{-1} \begin{pmatrix} 74.4 \\ 201.6 \end{pmatrix} \approx 15.66$$

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

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

Box's M test for equal covariance matrices when $n, m > 20,\ p < 5, k < 5$:

$$H_0 : \mathbf{\Sigma_1} = \mathbf{\Sigma_2}, \text{ in this example we have } k \equiv \#\text{ Matrices } = 2$$

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

$$M = (n-1)\ln(|\boldsymbol{S_1^{-1}}\boldsymbol{S_{pooled}}| + (m-1)\ln(|\boldsymbol{S_2^{-1}}\boldsymbol{S_{pooled}}|$$

$$\gamma = \frac{2p^2+3p-1}{6(p+1)(k-1)}\left[\frac{1}{n-1} + \frac{1}{m-1} - \frac{1}{n+m-k}\right],$$

$$(1-\gamma)M \sim \chi^2_{df}, \text{ where } df = p(p+1)(k-1)/2.$$

Example Wisconsin Power Data:

$$\boldsymbol{S}_1 = \begin{pmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{pmatrix},\ n = 45,\ \boldsymbol{S}_2 = \begin{pmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{pmatrix},\ m = 55$$

$$\boldsymbol{S_{pooled}} = \frac{(n-1)\boldsymbol{S_1} + (m-1)\boldsymbol{S_2}}{n+m-2} = \begin{pmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{pmatrix}$$

$$\boldsymbol{S_1^{-1} S_{pooled}} = \begin{pmatrix} 0.653 & 0.125 \\ 0.082 & 0.830 \end{pmatrix}, (n-1) \times \ln|S_1^{-1} S_{pooled}| = -27.283$$

$$\boldsymbol{S_2^{-1} S_{pooled}} = \begin{pmatrix} 1.951 & -0.461 \\ -0.300 & 1.299 \end{pmatrix}, (m-1) \times \ln|S_2^{-1} S_{pooled}| = 47.191$$

$$M = (47.191 - 27.283) = 19.367,$$

$$\gamma = \frac{2 \cdot 2^2 + 3 \cdot 2 - 1}{6 \cdot (2+1) \cdot (2-1)} \left[\frac{1}{44} + \frac{1}{55} - \frac{1}{98} \right] \approx 0.022$$

$$(1-\gamma)M = (1 - 0.022) \times 19.367 \approx 18.933,$$

$$df = \{2(2+1)(2-1)/2\} = 3,$$

$$p\text{-value} = Pr(\chi_3^2 > 18.933) \approx 0.028\%$$

Conclusion: Reject the null hypothesis that the covariance matrices are the same.

Box's M test for equal covariance matrices when sample sizes are small:

$$H_0: \boldsymbol{\Sigma_1} = \boldsymbol{\Sigma_2}, \text{ in this example we have } k \equiv \#\text{ Matrices } = 2$$

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

$$M = \left[(n-1)\ln(|\boldsymbol{S_1^{-1}S_{pooled}}| + (m-1)\ln(|\boldsymbol{S_2^{-1}S_{pooled}}|\right]$$

$$\gamma = \frac{2p^2+3p-1}{6(p+1)(k-1)}\left[\frac{1}{n-1} + \frac{1}{m-1} - \frac{1}{n+m-k}\right],$$

$$\xi = \frac{(p-1)(p+2)}{6(k-1)}\left[\frac{1}{(n-1)^2} + \frac{1}{(m-1)^2} - \frac{1}{(n+m-k)^2}\right],$$

$$df_1 = p(p+1)(k-1)/2,\ df_2 = (df_1+2)/|\xi - \gamma^2|$$

$$a^+ = df_1/(1-\gamma-df_1/df_2), \qquad F^+ = M/a^+$$

$$a^- = df_2/(1-\gamma+2/df_2), \qquad F^- = (df_2/df_1)\times[M/(a^- - M)]$$

$$F = \begin{cases} F^+ & \xi-\gamma^2 > 0 \\ F^- & \xi-\gamma^2 < 0 \end{cases} \Rightarrow F \sim F_{df_1, df_2}$$

Example Wisconsin Power Data:

$$\boldsymbol{S}_1 = \begin{pmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{pmatrix}, \ n = 45, \ \boldsymbol{S}_2 = \begin{pmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{pmatrix}, \ m = 55$$

$$\boldsymbol{S_{pooled}} = \frac{(n-1)\boldsymbol{S_1} + (m-1)\boldsymbol{S_2}}{n+m-2} = \begin{pmatrix} 10963.7 & 21505.5 \\ 21505.5 & 63661.3 \end{pmatrix}$$

$$\boldsymbol{S_1^{-1} S_{pooled}} = \begin{pmatrix} 0.653 & 0.125 \\ 0.082 & 0.830 \end{pmatrix}, \ (n-1) \times \ln |S_1^{-1} S_{pooled}| = -27.283$$

$$\boldsymbol{S_2^{-1} S_{pooled}} = \begin{pmatrix} 1.951 & -0.461 \\ -0.300 & 1.299 \end{pmatrix}, (m-1) \times \ln |S_2^{-1} S_{pooled}| = 47.191$$

$$M = (47.191 - 27.283) = 19.367$$

$$\gamma = \frac{2 \cdot 2^2 + 3 \cdot 2 - 1}{6 \cdot (2+1) \cdot (2-1)} \left[\frac{1}{44} + \frac{1}{55} - \frac{1}{98} \right] \approx 0.022$$

$$df_1 = \{2(2+1)(2-1)/2\} = 3$$

$$\xi = \frac{(2-1)\cdot(2+1)}{6\cdot(2-1)}\left[\frac{1}{44^2}+\frac{1}{55^2}-\frac{1}{98^2}\right] \approx 5.036\times 10^{-4},$$

$$\xi - \gamma^2 = 9.5\times 10^{-7} > 0$$

$$df_2 = (3+2)/|\xi-\gamma^2| \approx 5270593.4,$$

$$a^+ = 3/(1-0.022-5.7\times 10^{-7}) \approx 3.069,$$

$$F^+ = 19.367/3.069 \approx 6.311$$

$$a^- = 5270593.4/(1-0.022+2/5270593.4 \approx 5.4\times 10^6,$$

$$F^- = (5270593.4/3)\times[19.367/(5.4\times 10^6 - 19.367)] \approx 6.311$$

$F = F^+ \approx 6.311$ since $\xi - \gamma^2 = 9.5\times 10^{-7} > 0$

p-value $= Pr(F_{df_1,df_2} > 6.311) \approx 0.028\%$

Conclusion: Reject the null hypothesis that the covariance matrices are the same.
Typically α is set at 0.1% due to sensitivity of the Box's M test.