EMSE 4765: DATA ANALYSIS
For Engineers and Scientists

Session 6: Two Sample Hypothesis Tests, Joint Normal Distribution, Vectors and Matrices, Matrix Algebra, Linear Combinations, Geometric Interpretation

Version: 2/15/2021

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.

Let $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ be two random samples from **a normal distribution** with means μ_1 and μ_2 and **same variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$**, respectively. **(Y_j's independent of the X_i's).** Then we can construct the following T estimator :

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

which has a t distribution with $n + m - 2$ degrees of freedom

$$(S_p)^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

- $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

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

The **two-sample *t* test** for testing $H_0 : \mu_1 - \mu_2 = \Delta_0$ is a follows:

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.

Example 16: As the population ages, there is increasing concern about accident-related injuries to the elderly. The article "Age and Gender Differences in Single-Step Recovery from a Forward Fall", *Journal of Gerontology, 1999, M44-M50*, reported on an experiment in which **the maximum lean angle** — the furthest a person is able to lean and recover in one step – was determined for both a sample of younger females (21-29 years) and a sample of older females (67-81 years). The following observations are consistent with summary data given in the article:

YF: 29, 34, 33, 27, 28, 32, 31, 34, 32, 27 (Sample size $n = 10$).
OF: 18, 15, 23, 13, 12 (Sample size $n = 5$).

Does the data suggest that true average maximum lean angle is more than 10 degrees smaller than it is for younger females? State and test the relevant hypothesis at significance level .10 by obtaining a $p-$value.

Assumption: Let $(X_1, \ldots, X_n)$ be the YF and $(Y_1, \ldots, Y_m)$ be the OF random $i.i.d.$ sample from a **normal distribution** with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively.

Null-Hypothesis: $H_0 : \mu_1 - \mu_2 = 10$

Alternative-Hypothesis: $H_1 : \mu_1 - \mu_2 > 10$

Sample data: $\overline{x} \approx 30.7, n = 10,\ \overline{y} \approx 16.2, m = 5,\ s_1^2 \approx 7.6,\ s_2^2 \approx 19.7$

Pooled Variance Statistic estimate (assuming variances are the same) :

$$(s_p)^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} = 11.30$$

Statistic (assuming variances are the same) :

$$t_0 = \frac{\overline{x} - \overline{y} - 10}{s_p\sqrt{\frac{1}{10} + \frac{1}{5}}} \approx 2.44,$$

Degrees of freedom : $n + m - 2 = 13$

P-value: $Pr(T_{13} > t_0 | H_0) \approx 1.5\% < \alpha = 10\%$

Conclusion: Reject H_0

Let $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ be a random $i.i.d.$ sample from a **normal distribution** with means μ_1 and μ_2 and **different variances** $\boldsymbol{\sigma_1^2 \neq \sigma_2^2}$, respectively. **($\boldsymbol{Y_j}$'s independent of the $\boldsymbol{X_i}$'s).** Then we can construct the following T estimator :

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \sim t_v$$

which has approximately **a $\boldsymbol{t}$ distribution with $\boldsymbol{\nu}$ degrees of freedom**

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

(round ν down to the nearest integer).

- $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$:

$$(\overline{x}-\overline{y}) \pm t_{\nu,1-\alpha/2}\sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}}$$

The **two-sample *t* test** for testing $H_0 : \mu_1 - \mu_2 = \Delta_0$ is a follows:

Test statistic value:

$$t_0 = \frac{\overline{x}-\overline{y}-\Delta_0}{\sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}}}$$

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

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

Null-Hypothesis: $H_0 : \mu_1 - \mu_2 = 10$

Alternative-Hypothesis: $H_1 : \mu_1 - \mu_2 > 10$

Sample data: $\overline{x} \approx 30.7, n = 10,\ \overline{y} \approx 16.2, m = 5,\ s_1^2 \approx 7.6,\ s_2^2 \approx 19.7$

Statistic (assuming variances are the not the same) :

$$t_0 = \frac{\overline{x} - \overline{y} - 10}{\sqrt{\frac{s_1^2}{10} + \frac{s_2^2}{5}}} \approx 2.08$$

Degrees of freedom : $$\nu = \frac{\left[\frac{s_1^2}{10} + \frac{s_2^2}{5}\right]^2}{\frac{(s_1^2/10)^2}{9} + \frac{(s_1^2/5)^2}{4}} \approx 5.59 \Rightarrow \text{use } 5$$

P-value: $Pr(T_6 > t_0 | H_0) \approx 4.6\% < \alpha = 10\%$

Conclusion: Reject H_0

Two-Sample T-Test and CI: YF, OF

Method

μ_1: mean of YF
μ_2: mean of OF
Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
YF	10	30.70	2.75	0.87
OF	5	16.20	4.44	2.0

Estimation for Difference

Difference	Pooled StDev	95% Lower Bound for Difference
14.50	3.36	11.24

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 10$
Alternative hypothesis H_1: $\mu_1 - \mu_2 > 10$

T-Value	DF	P-Value
2.44	13	0.015

Two-Sample T-Test and CI: YF, OF

Method

μ_1: mean of YF
μ_2: mean of OF
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
YF	10	30.70	2.75	0.87
OF	5	16.20	4.44	2.0

Estimation for Difference

Difference	95% Lower Bound for Difference
14.50	10.13

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 10$
Alternative hypothesis H_1: $\mu_1 - \mu_2 > 10$

T-Value	DF	P-Value
2.08	5	0.046

STATISTICAL INFERENCE Two Sample Mean Hyp. Testing

R - Code

```r
# loading the readr package
library(readr)
LeanAngle <- read_csv("LeanAngle.csv")

# Assigning First Column to YF
YF <-LeanAngle[[1]]
# Assigning First Column to YF
OF <-LeanAngle[[2]]
OF <- OF[!is.na(OF)]
t.test(YF,OF,alternative = "greater",10,
       var.equal=TRUE)
```

R - Output

```
        Two Sample t-test

data:  YF and OF
t = 2.4441, df = 13, p-value = 0.01477
alternative hypothesis: true difference in means is greater than 10
95 percent confidence interval:
 11.23937      Inf
sample estimates:
mean of x mean of y 
     30.7      16.2 
```

Analysis in file "LeanAngle.R"

R - Code

```r
# loading the readr package
library(readr)
LeanAngle <- read_csv("LeanAngle.csv")

# Assigning First Column to YF
YF <-LeanAngle[[1]]
# Assigning First Column to YF
OF <-LeanAngle[[2]]
OF <- OF[!is.na(OF)]
t.test(YF,OF,alternative = "greater",
        10,var.equal=FALSE)
```

R - Output

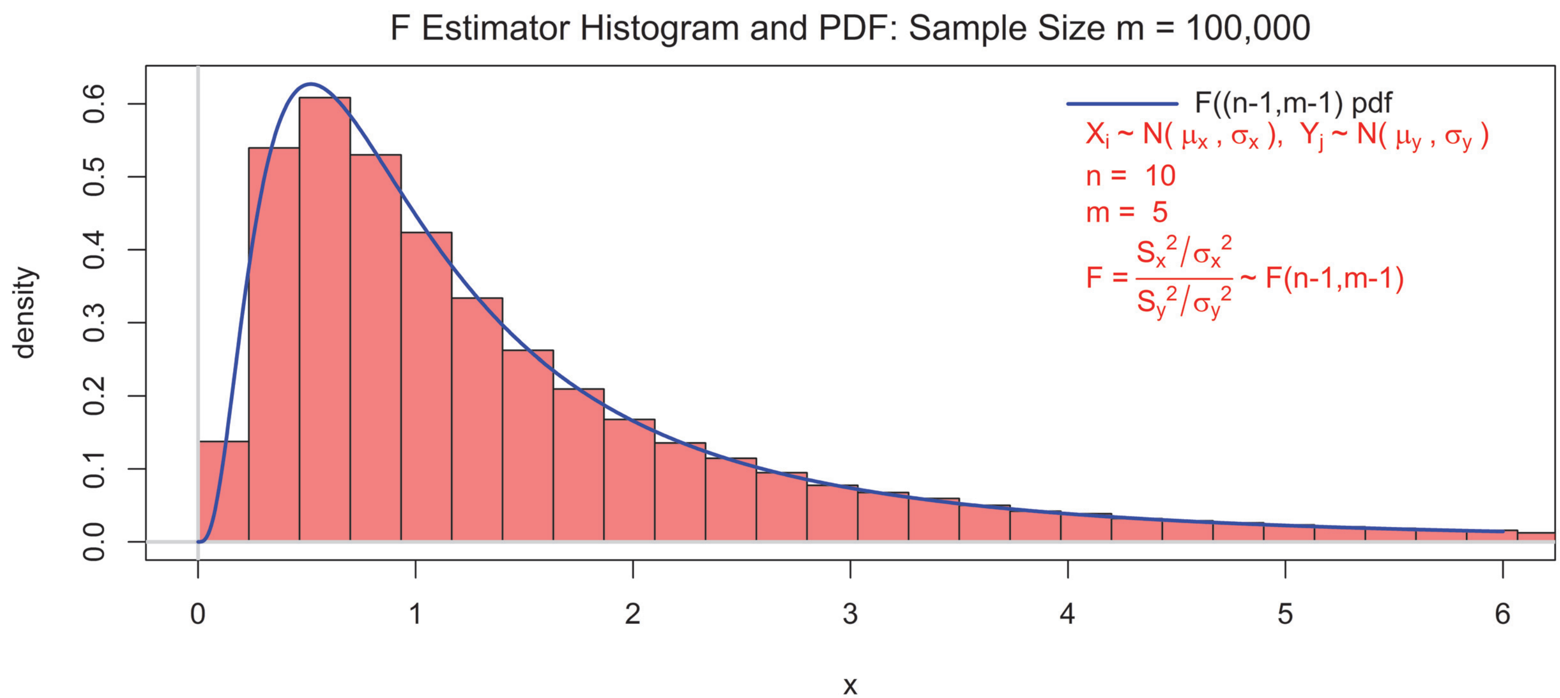
```
        Welch Two Sample t-test

data:  YF and OF
t = 2.0764, df = 5.5922, p-value = 0.04327
alternative hypothesis: true difference in means is greater than 10
95 percent confidence interval:
 10.23314      Inf
sample estimates:
mean of x mean of y 
     30.7      16.2 
```

Analysis in file "LeanAngle.R"

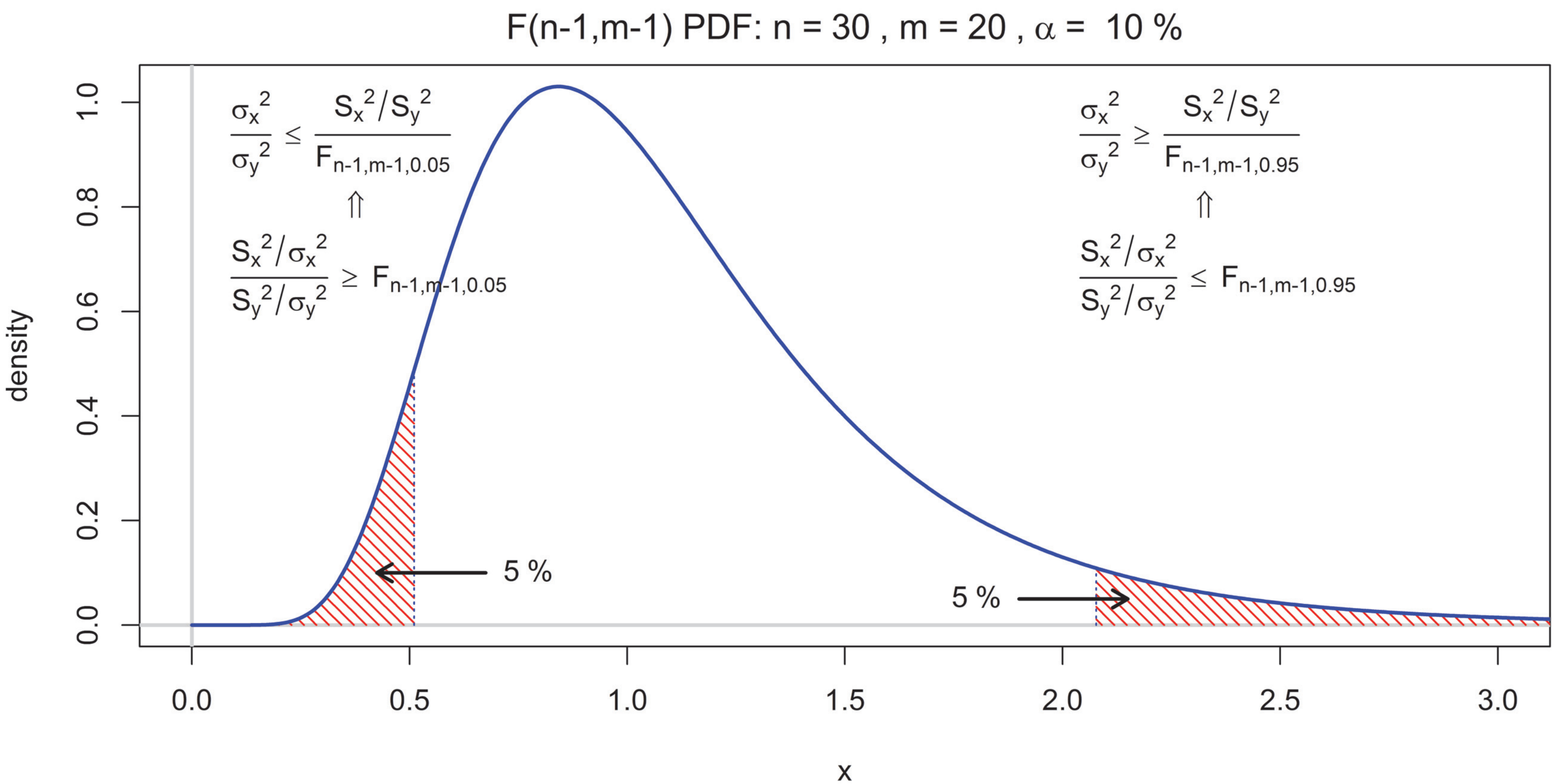
- **Recall:** Let $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ be a random $i.i.d.$ samples $X_i \sim N(\mu_x, \sigma_x)$, $Y_i \sim N(\mu_y, \sigma_y)$ (Y_j's independent of the X_i's) $\Rightarrow$

$$\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} = \left[\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma_x}\right)^2\right] \Big/ \left[\frac{1}{m-1}\sum_{i=1}^{m}\left(\frac{Y_i - \overline{Y}}{\sigma_y}\right)^2\right] \sim F_{n-1,m-1}$$

F distribution with $n-1$ and $m-1$ degrees of freedom



Analysis in "F_Estimator.R"

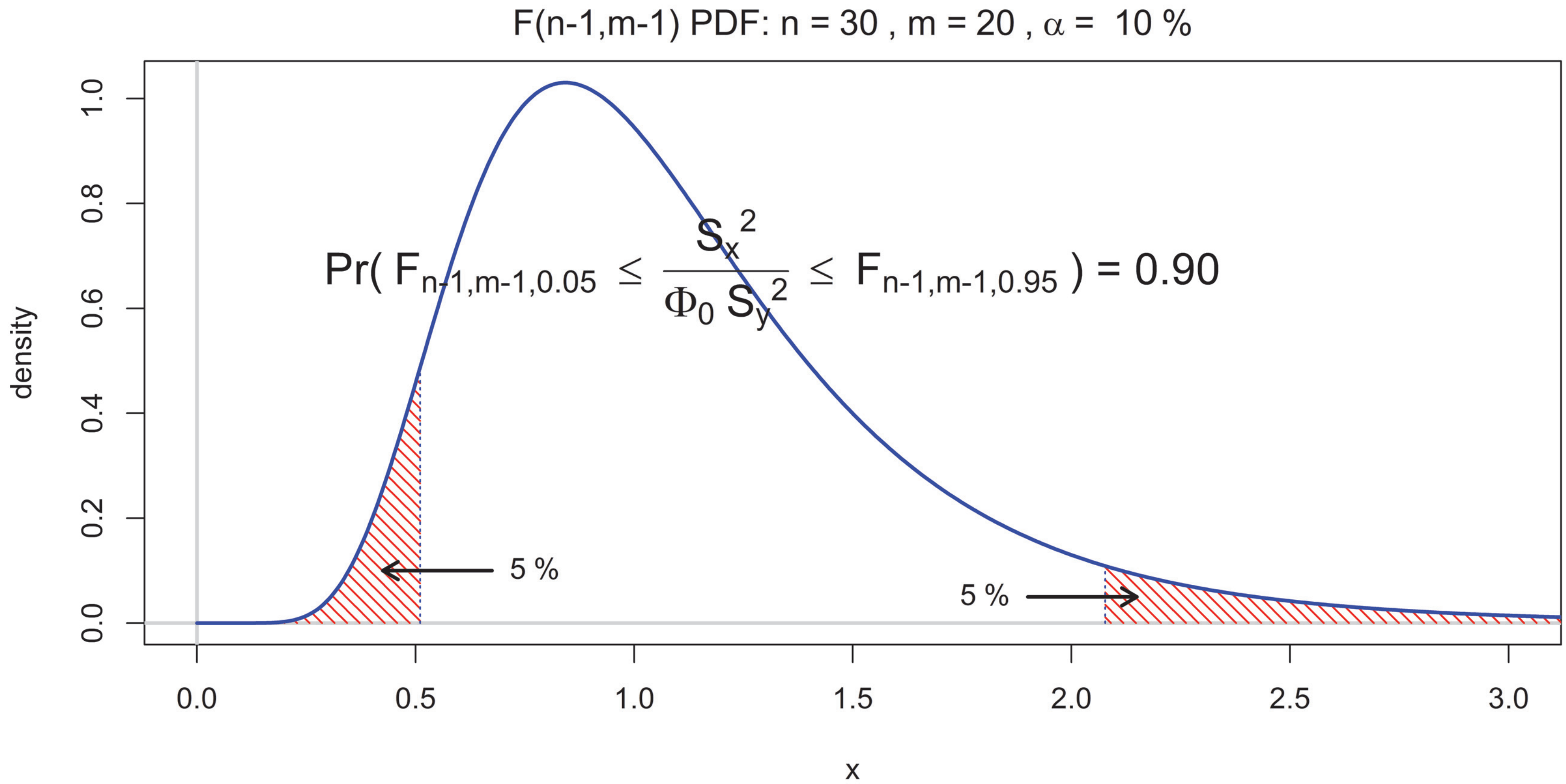


Two-sided $100(1-\alpha)\%$ confidence interval for the ratio σ_x^2/σ_y^2 :

$$\left(\frac{s_x^2/s_y^2}{F_{n-1,m-1,1-\alpha/2}}, \frac{s_x^2/s_y^2}{F_{n-1,m-1,\alpha/2}}\right)$$

The **two-sample *F* test** for testing $H_0 : \sigma_x^2/\sigma_y^2 = \Phi_0$. Assuming H_0 is true :

$$\sigma_x^2/\sigma_y^2 = \Phi_0 \Rightarrow F_0 = \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} = \frac{S_x^2}{S_y^2} \times \frac{\sigma_y^2}{\sigma_x^2} = \frac{S_x^2}{\Phi_0 S_y^2} \sim F_{n-1,m-1} \Rightarrow f_0 = \frac{s_x^2}{\Phi_0 s_y^2}$$



Alternative Hypothesis	Rejection Regions for significance α	
$H_1 : \sigma_x^2/\sigma_y^2 > \Phi_0$	$f_0 > F_{n-1,m-1,1-\alpha}$	(upper-tailed)
$H_1 : \sigma_x^2/\sigma_y^2 < \Phi_0$	$f_0 < F_{n-1,m-1,\alpha}$	(lower-tailed)
$H_1 : \sigma_x^2/\sigma_y^2 \neq \Phi_0$	$f_0 > F_{n-1,m-1,1-\alpha/2}$ or $f_0 < F_{n-1,m-1,\alpha/2}$	(two-tailed)

- The upper-tailed is frequently used in Analysis of Variance (ANOVA). The p-value of this test equals $Pr(F > f_0 | H_0)$.

Example 16 Continued: The article "Age and Gender Differences in Single-Step Recovery from a Forward Fall", *Journal of Gerontology, 1999, M44-M50*, reported on an experiment in which **the maximum lean angle** — the furthest a person is able to lean and recover in one step – was determined for both a sample of younger females (21-29 years) and a sample of older females (67-81 years). The following observations are consistent with summary data given in the article:

YF: 29, 34, 33, 27, 28, 32, 31, 34, 32, 27 (Sample size $n = 10$).
OF: 18, 15, 23, 13, 12 (Sample size $n = 5$).

Carry out a test at a significance level of $\alpha = 0.10$, whether the standard deviations for the two age group are different.

Null-Hypothesis: $$H_0 : \frac{\sigma_x^2}{\sigma_y^2} = \Phi_0 = 1$$

Alternative-Hypothesis: $$H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq \Phi_0 = 1$$

Sample data: $\overline{x} \approx 30.7, n = 10,\ \overline{y} \approx 16.2, m = 5,\ s_1^2 \approx 7.6,\ s_2^2 \approx 19.7$

Statistic : $$f_0 = \frac{s_x^2}{\Phi_0 s_y^2} \approx 0.384$$

Criticality Region: $(0, F_{9,4,0.05}) \cup (F_{9,4,0.95}, \infty) \approx (0, 0.27) \cup (6.00, \infty)$

P-Value: Cannot be calculated in this case

Conclusion : Fail to Reject H_0 $(0.38 \in [0.27, 6.00])$

Carry out a test at a significance level of 0.10, whether the standard deviations for the YF age group is less than that of the OF age group.

Null-Hypothesis: $H_0 : \sigma_x^2/\sigma_y^2 = 1$

Alternative-Hypothesis: $H_1 : \sigma_x^2/\sigma_y^2 < 1$

Sample data: $\overline{x} \approx 30.7, n = 10,\ \overline{y} \approx 16.2, m = 5,\ s_x^2 \approx 7.6,\ s_y^2 \approx 19.7$

Statistic : $$f_0 = \frac{s_x^2}{\Phi_0 s_y^2} \approx 0.384$$

Criticality Region: $(0, F_{9,4,0.10}) \approx (0, 0.371)$

P-Value: $Pr(F_{n-1,m-1} < f_0 | H_0) \approx 10.74\%$

Conclusion : Fail to Reject H_0 ($0.384 \in [0.371, \infty]$ and $10.74\% > 10\%$)

Test and CI for Two Variances: YF, OF

Method

σ_1: standard deviation of YF
σ_2: standard deviation of OF
Ratio: σ_1/σ_2
F method was used. This method is accurate for normal data only.

Descriptive Statistics

Variable	N	StDev	Variance	95% CI for σ
YF	10	2.751	7.567	(1.892, 5.022)
OF	5	4.438	19.700	(2.659, 12.754)

Ratio of Standard Deviations

Estimated Ratio	95% CI for Ratio using F
0.619754	(0.208, 1.346)

Test

Null hypothesis $H_0: \sigma_1 / \sigma_2 = 1$
Alternative hypothesis $H_1: \sigma_1 / \sigma_2 \neq 1$
Significance level $\alpha = 0.05$

Method	Test Statistic	DF1	DF2	P-Value
F	0.38	9	4	0.215

Test and CI for Two Variances: YF, OF

Method

σ_1: standard deviation of YF
σ_2: standard deviation of OF
Ratio: σ_1/σ_2
F method was used. This method is accurate for normal data only.

Descriptive Statistics

Variable	N	StDev	Variance	95% Upper Bound for σ
YF	10	2.751	7.567	4.526
OF	5	4.438	19.700	10.530

Ratio of Standard Deviations

Estimated Ratio	95% Upper Bound for Ratio using F
0.619754	1.181

Test

Null hypothesis $H_0: \sigma_1 / \sigma_2 = 1$
Alternative hypothesis $H_1: \sigma_1 / \sigma_2 < 1$
Significance level $\alpha = 0.05$

Method	Test Statistic	DF1	DF2	P-Value
F	0.38	9	4	0.107

R - Code

```
# loading the readr package
library(readr)
LeanAngle <- read_csv("LeanAngle.csv")

# Assigning First Column to YF
YF <-LeanAngle[[1]]
# Assigning First Column to YF
OF <-LeanAngle[[2]]
OF <- OF[!is.na(OF)]
var.test(YF,OF,ratio=1,alternative = "two.sided",
         conf.level=0.90)
```

R - Output

```
        F test to compare two variances

data:  YF and OF
F = 0.38409, num df = 9, denom df = 4, p-value = 0.2147
alternative hypothesis: true ratio of variances is not equal to 1
90 percent confidence interval:
 0.06402882 1.39545024
sample estimates:
ratio of variances
         0.3840948
```

Analysis in file "LeanAngle.R"

R - Code

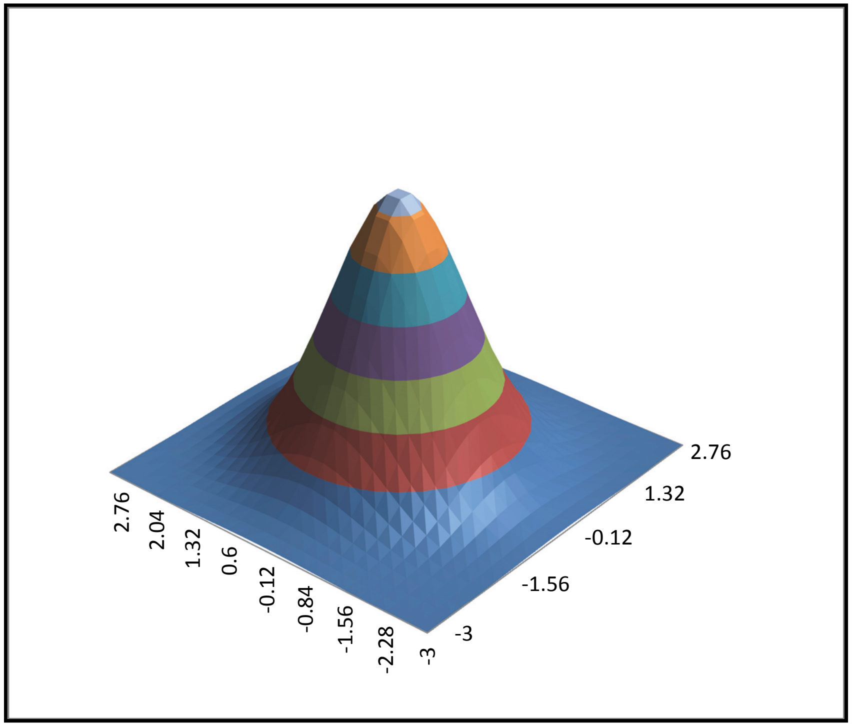
```
# loading the readr package
library(readr)
LeanAngle <- read_csv("LeanAngle.csv")

# Assigning First Column to YF
YF <-LeanAngle[[1]]
# Assigning First Column to YF
OF <-LeanAngle[[2]]
OF <- OF[!is.na(OF)]
var.test(YF,OF,ratio=1,alternative = "less",
         conf.level=0.90)
```

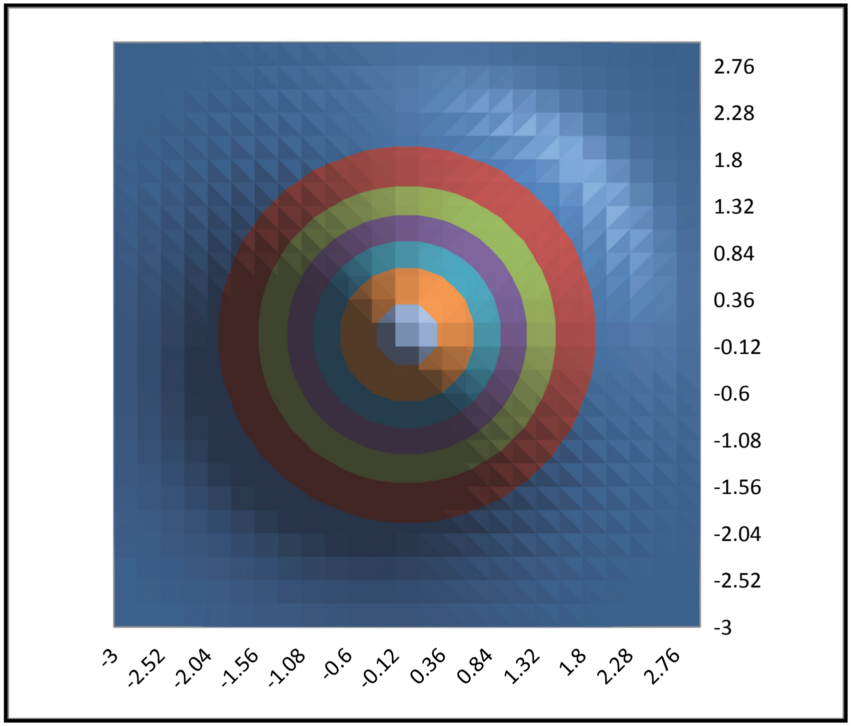
R - Output

```
        F test to compare two variances

data:  YF and OF
F = 0.38409, num df = 9, denom df = 4, p-value = 0.1074
alternative hypothesis: true ratio of variances is less than 1
90 percent confidence interval:
 0.000000 1.034244
sample estimates:
ratio of variances
         0.3840948
```

Analysis in file "LeanAngle.R"

The assumption of the two-sample hypothesis tests is that the $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ be a random $i.i.d.$ sample from **a normal distribution** with means μ_1 and μ_2 and the same variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$ and **Y_j's independent of the X_i's** implies that the distribution of (X, Y) is a **bivariate normal distribution.**





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

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

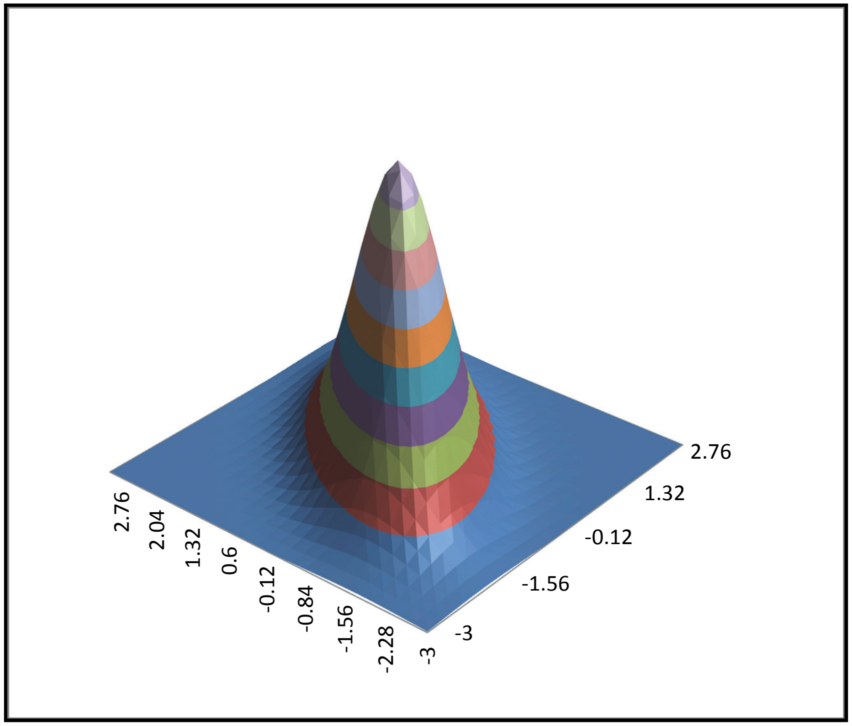
$$\text{Covariance Matrix: } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & Cov(X_1, X_2) \\ Cov(X_1, X_2) & \sigma_2^2 \end{pmatrix}$$

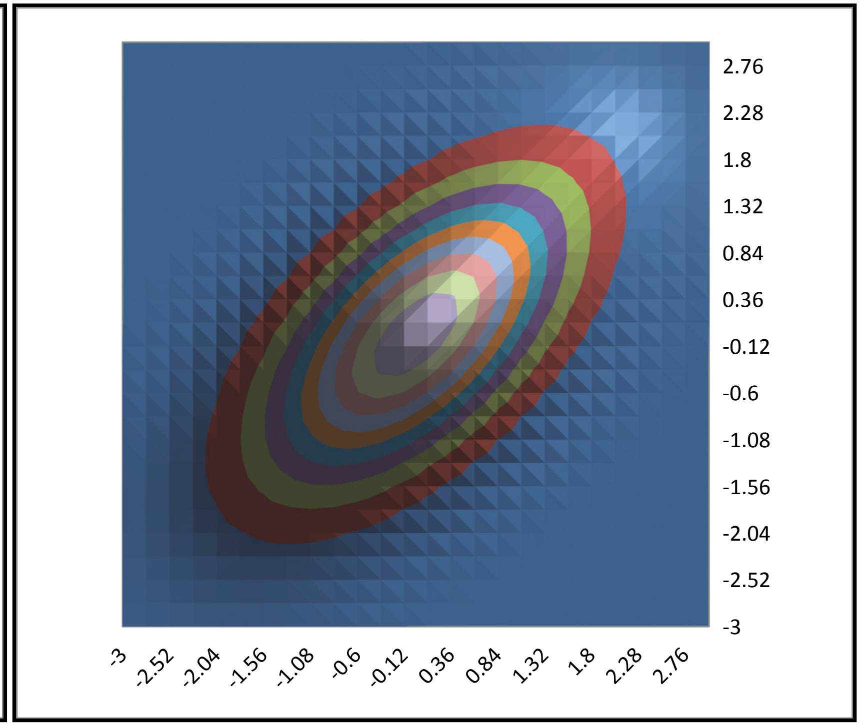
$$f(x, y) = \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}|}} exp\Big[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\boldsymbol{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \Big]$$

- Independence and same variance in case of the bivariate normal distribution:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}, \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/\sigma^2 \end{pmatrix}$$

- What is the shape of the pdf in case of **positive dependence between two normal marginals** with **different variances $\sigma_1^2 \neq \sigma_2^2$?**





To be able to perform **Statistical Inference** in the case that the $(X_1, \ldots, X_n)$ and the $(Y_1, \ldots, Y_m)$ are samples from **a multivariate normal distribution** with **a non-diagonal covariance matrix Σ** requires knowledge of **multivariate analysis techniques.**

- A review of **Matrix Algebra** involving **scalars, vectors and matrices.**
- **Hotellings T^2 test: A multivariate hypothesis test** for the **mean values of a vector**, where **the variance covariance matrix is not unit diagonal.**
- **Regression Analysis: A multivariate analysis methodology** to investigate the relationship between **a single dependence variable** and **multiple explanatory variables.** The variance covariance matrix of the explanatory variables in most cases in not a diagonal matrix.
- **Analysis of Variance: A statistical analysis technique** to **evaluate the effect of one or more factors on a response variable.**

- **Typical convention** is that **vectors are written as columns**. The i-th element of a vector is indicated by x_i. Hence, an n-dimensional vector is:

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

- **Convention: Underline** to indicate a vector or write them in **a bold font**.

- **An $\boldsymbol{m} \times \boldsymbol{n}$-matrix $\boldsymbol{A}$** may be viewed as **$\boldsymbol{n}$ columns** each of **dimension $\boldsymbol{m}$**. Its elements are indicated by a_{ij} where the index i refers to the **row number** and the index j refers to the **column number**.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & & \dots & a_{mn} \end{pmatrix}$$

- **Convention:** Use **capital letters for matrices** and often they are written in a bold font.
- If $m = n$, then the matrix is called a **square matrix**.
- If $a_{ij} = a_{ji}$ for all elements of **a square matrix**, then the matrix is called **symmetric**.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 5 & 6 & 7 \\ 5 & 2 & 8 & 9 \\ 6 & 8 & 3 & 10 \\ 7 & 9 & 10 & 4 \end{pmatrix}$$

- If $a_{ij} = 0$ for all off-diagonal elements of **a square matrix**, then the matrix is called **a diagonal matrix**.
- If $a_{ii} = 1$ for all on-diagonal elements of **a diagonal matrix**, then the matrix is called **the identity matrix** and is usually denoted by $\boldsymbol{I}$.
- An **n-dimensional vector** can be thought of as **an $n \times 1$ matrix**.

- **Vector-Scalar multiplication:**

$$\lambda \underline{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}, \text{ e.g. } 2 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix}$$

- **Transposed column n-vector** becomes **a row n-vector**:

$$\underline{x}^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$$

- **Conventions:** Write $\underline{x}^T$, $\underline{x}^t$ or $\underline{x}'$ to indicate a transposed vector. A transposed column vector becomes a row vector **and vice versa.**

- **An $m \times n$-matrix** may also be viewed also as **m row vectors** each of **dimension n.**

- **Matrix-Vector multiplication:**

$$\boldsymbol{A}\underline{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \vdots \\ & & \ddots & a_{m-1,n} \\ a_{m1} & & a_{m,n-1} & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix}$$

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

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

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 \\ 2 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 \end{pmatrix} = \begin{pmatrix} 20 \\ 36 \end{pmatrix}$$

- **Vector-Matrix multiplication:**

$$\underline{x}^T \boldsymbol{A} = (\, x_1 \quad x_2 \quad \cdots \quad x_m \,) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \vdots \\ & & \ddots & a_{m-1,n} \\ a_{m1} & & a_{m,n-1} & a_{mn} \end{pmatrix} =$$

$$\left(\sum_{i=1}^{m} x_i a_{i1} \quad \sum_{i=1}^{m} x_i a_{i2} \quad \cdots \quad \sum_{i=1}^{m} x_i a_{in} \right)$$

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

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

Example:

$$(\, 4 \quad 5 \,) \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \end{pmatrix} =$$

$$(\, 4 \cdot 1 + 5 \cdot 2 \quad 4 \cdot 2 + 5 \cdot 4 \quad 4 \cdot 3 + 5 \cdot 5 \,) = (\, 14 \quad 28 \quad 37 \,)$$

- **Matrix-Matrix multiplication:**

$$\boldsymbol{AB} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \vdots \\ & & \ddots & a_{m-1,n} \\ a_{m1} & & a_{m,n-1} & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & & b_{2p} \\ \vdots & & & \vdots \\ & & \ddots & b_{n-1,p} \\ b_{n1} & & b_{n,p-1} & b_{np} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{n} a_{1j}b_{j1} & \sum_{j=1}^{n} a_{1j}b_{j2} & \dots & \sum_{j=1}^{n} a_{1j}b_{jp} \\ \sum_{j=1}^{n} a_{2j}b_{j1} & \sum_{j=1}^{n} a_{2j}b_{j2} & & \sum_{j=1}^{n} a_{2j}b_{jp} \\ \vdots & & & \vdots \\ & & \ddots & \sum_{j=1}^{n} a_{m-1,j}b_{jp} \\ \sum_{j=1}^{n} a_{mj}b_{j1} & & \sum_{j=1}^{n} a_{mj}b_{j,p-1} & \sum_{j=1}^{n} a_{mj}b_{jp} \end{pmatrix}$$

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

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} =$$

$$\begin{pmatrix} 1 \cdot 1 + 2 \cdot 3 + 3 \cdot 5 & 1 \cdot 2 + 2 \cdot 4 + 3 \cdot 6 \\ 2 \cdot 1 + 4 \cdot 3 + 5 \cdot 5 & 2 \cdot 2 + 4 \cdot 4 + 5 \cdot 6 \end{pmatrix} = \begin{pmatrix} 22 & 28 \\ 39 & 50 \end{pmatrix}$$

- **Matrix-Matrix multiplication is Non-Commutative.**
  First of all we have to consider square matrices in this case (why?). But even when we consider square matrices the following is not true in general!

$$AB \neq BA$$

Example:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, B = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 2 & 1 \cdot 3 + 2 \cdot 4 \\ 3 \cdot 1 + 4 \cdot 2 & 3 \cdot 3 + 4 \cdot 4 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 11 & 25 \end{pmatrix}$$

$$BA = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 3 \cdot 3 & 1 \cdot 2 + 3 \cdot 4 \\ 2 \cdot 1 + 4 \cdot 3 & 2 \cdot 2 + 4 \cdot 4 \end{pmatrix} = \begin{pmatrix} 10 & 14 \\ 14 & 20 \end{pmatrix}$$

- **Transpose of a matrix:**

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \vdots \\ & & \ddots & a_{m-1,n} \\ a_{m1} & & a_{m,n-1} & a_{mn} \end{pmatrix} \boldsymbol{A}^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & & a_{m2} \\ \vdots & & & \vdots \\ & & \ddots & a_{m,n-1} \\ a_{1n} & & a_{m-1,n} & a_{mn} \end{pmatrix}$$

Example:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \boldsymbol{A}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

$$(m \times n\text{-matrix})^T = (n \times m\text{-matrix})$$

- **Transpose of a matrix product:**

$$(\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T$$

$$\begin{aligned} &[(m \times n\text{-matrix}) \cdot (n \times p\text{-matrix})]^T = \\ &(n \times p\text{-matrix})^T \cdot (m \times n\text{-matrix})^T = \\ &(p \times n\text{-matrix}) \cdot (n \times m\text{-matrix})]^T = (p \times m\text{-matrix}) \end{aligned}$$

Example:

$$\begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 44 & 56 \\ 98 & 128 \end{pmatrix}$$

$$\left[\begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \right]^T = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^T \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix}^T =$$

$$\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} 2 & 8 \\ 4 & 10 \\ 6 & 12 \end{pmatrix} = \begin{pmatrix} 44 & 98 \\ 56 & 128 \end{pmatrix} = \begin{pmatrix} 44 & 56 \\ 98 & 128 \end{pmatrix}^T$$

- **The inverse of a square matrix** is defined such that

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = I$$

where I is **the identity matrix.**

Example:

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \text{ where } |\boldsymbol{A}| = ad - bc$$

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}, \boldsymbol{A}^{-1} = \frac{1}{5}\begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix}$$

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}\frac{1}{5}\begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \frac{1}{5}\begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

- **The inverse of a matrix product:**

$$(AB)^{-1} = B^{-1}A^{-1}$$

Example:

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}, \boldsymbol{A}^{-1} = \frac{1}{5}\begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix}$$

$$\boldsymbol{B} = \begin{pmatrix} 6 & 7 \\ 3 & 5 \end{pmatrix}, \boldsymbol{B}^{-1} = \frac{1}{9}\begin{pmatrix} 5 & -7 \\ -3 & 6 \end{pmatrix}$$

$$\boldsymbol{AB} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 6 & 7 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 15 & 19 \\ 30 & 41 \end{pmatrix},$$

$$(\boldsymbol{AB})^{-1} = \frac{1}{45}\begin{pmatrix} 41 & -19 \\ -30 & 15 \end{pmatrix}$$

$$\boldsymbol{B}^{-1}\boldsymbol{A}^{-1} = \frac{1}{9}\begin{pmatrix} 5 & -7 \\ -3 & 6 \end{pmatrix}\frac{1}{5}\begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix} = \frac{1}{45}\begin{pmatrix} 41 & -19 \\ -30 & 15 \end{pmatrix}$$

Definition: Given **a collection of random variables** $X_1, \ldots, X_n$ and n numerical **constants/coefficients** $a_1, \ldots, a_n$, the rv

$$Y = a_1 X_1 + \ldots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

is called a **linear** combination of the X_i's. Hence we can identify two vectors:

$$\underline{\boldsymbol{a}}^T = (\, a_1 \quad \ldots \quad a_n \,),\ \underline{\boldsymbol{X}}^{\boldsymbol{T}} = (\, X_1 \quad \ldots \quad X_n \,)$$

and write

$$Y = \underline{\boldsymbol{a}}^T \underline{\boldsymbol{X}}$$

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

- Let $E[\underline{\boldsymbol{X}}] = \underline{\boldsymbol{\mu}}$, where $\underline{\boldsymbol{\mu}}^{\boldsymbol{T}} = (\, \mu_1 \quad \ldots \quad \mu_n \,)$ then

$$E[\,\underline{\boldsymbol{a}}^T \underline{\boldsymbol{X}}] = E[Y] = a_1\mu_1 + \ldots + a_n\mu_n = \underline{\boldsymbol{a}}^T \underline{\boldsymbol{\mu}} = \underline{\boldsymbol{a}}^T E[\underline{\boldsymbol{X}}]$$

Recall in single dimension the following holds: $\boldsymbol{E[a\mathrm{X}] = aE[X]}$

- Recall, that if X and Y are dependent random variables

$$V[aX + bY] = a^2V[X] + b^2V[Y] + 2abCov[X, Y]$$

- Generalization to n RV's X_i's that are mutually dependent

$$V[Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j COV[X_i, X_j] \text{ (Note: } COV[X, X] = V[X])$$

Introducing the variance-covariance matrix $\mathbf{\Sigma}$ of $\underline{\boldsymbol{X}}$:

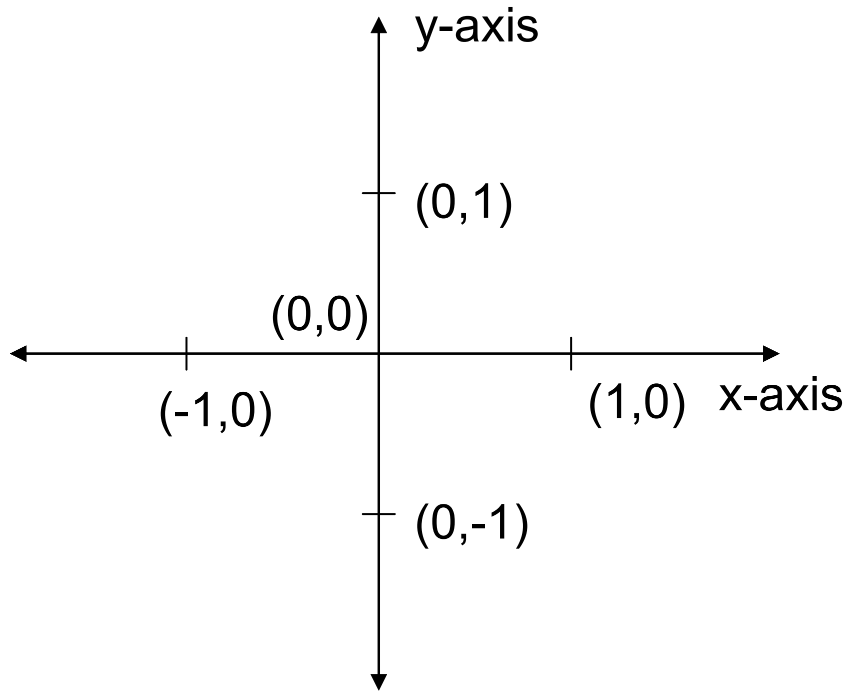
$$\mathbf{\Sigma} = \begin{pmatrix} V[X_1] & Cov(X_1, X_2) & \dots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & V[X_2] & & \\ \vdots & & \ddots & \vdots \\ Cov(X_n, X_1) & \dots & & V[X_n] \end{pmatrix}$$

we can rewrite $V[Y]$ in vector-matrix notation, which is much more concise:

$$V[Y] = \underline{\boldsymbol{a}}^T \mathbf{\Sigma} \underline{\boldsymbol{a}}$$

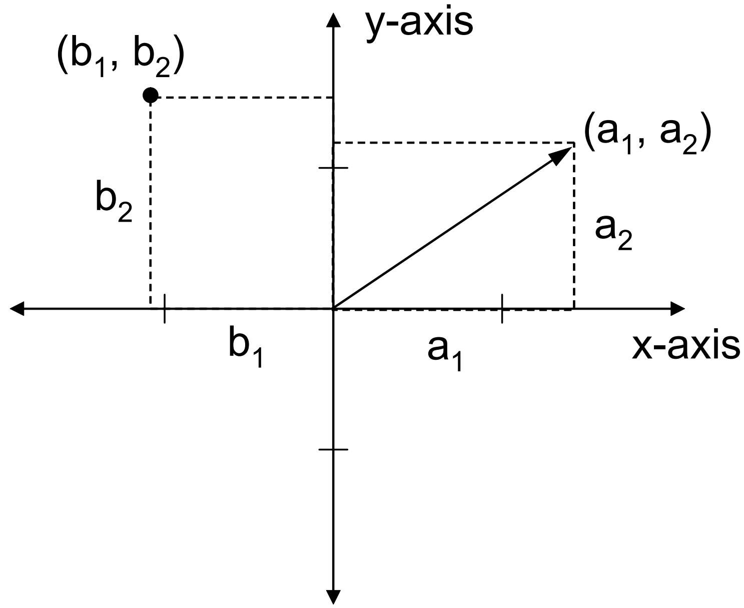
- **The variance-covariance matrix** is **a square symmetric matrix** (why?) and is **a positive definite matrix**, i.e. $\underline{\boldsymbol{x}}^T \mathbf{\Sigma} \underline{\boldsymbol{x}} > 0$ for all vectors $\underline{\boldsymbol{x}} \neq \underline{\mathbf{0}}$ **(Recall: V[X] > 0 in single dimension)**

For **a coordinate system** one needs three things:



1. An **origin** $(0,0)$
2. Two lines, called **coordinate axes**, that go through the origin. In the system above each line is **perpendicular**, which makes it a **Cartesian System**.
3. One point (other than the origin) on each axis to establish scale. These points identify the **standard base vectors** $\underline{\boldsymbol{e}}_1^T = (\mathbf{1} \;\; \mathbf{0})$ **and** $\underline{\boldsymbol{e}}_2^T = (\mathbf{0} \;\; \mathbf{1})$.

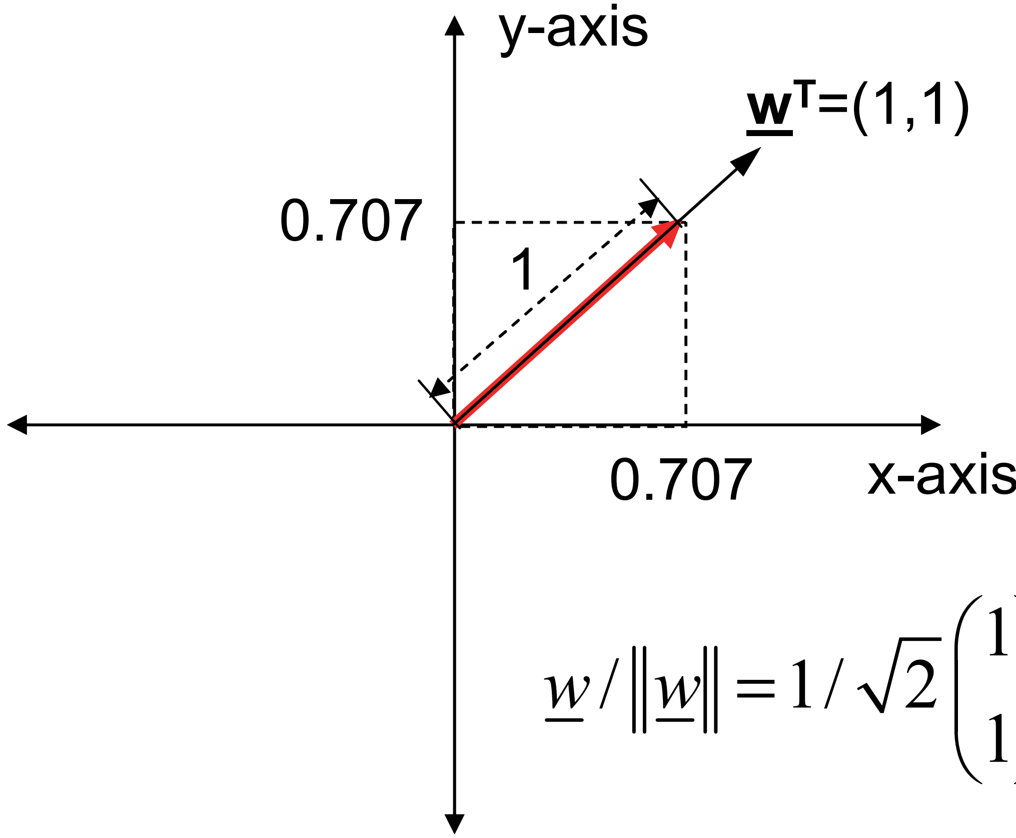
- Having a coordinate system allows us to **assign to each point in the system its coordinates (a_1, a_2).**



- **We may also write (a_1, a_2) as a vector** as follows:

$$\underline{\boldsymbol{a}} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

- **Vector-Scalar multiplication**



Length of a Vector:

$$\|\underline{w}\| = \left[\sum_{i=1}^{n} w_i^2\right]^{1/2}$$

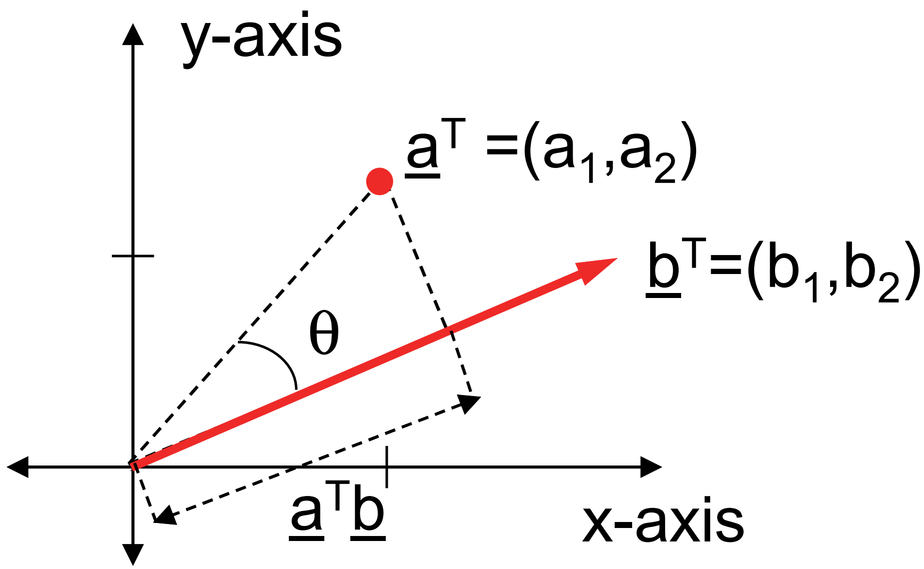
$$\underline{w}/\|\underline{w}\| = 1/\sqrt{2}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix} \approx \begin{pmatrix}0.707\\0.707\end{pmatrix}$$

- **The effect of vector-scalar multiplication** is **stretching or shrinking the length of a vector while maintaining its direction.**

- **Vector-Vector multiplication**

$$\underline{a}^T\underline{b} = \begin{pmatrix} a_1 & a_2 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = a_1b_1 + a_2b_2$$

$$\underline{a}^T\underline{b} = \|\underline{a}\|\|\underline{b}\|\cos(\theta)$$



- **When** $-90° < \theta < 90° \Rightarrow \cos(\theta) > 0$ **: In that case,** $\underline{a}^T\underline{b} =$ **distance from the origin to the perpendicular projection of the point** $\underline{a} = (a_1, a_2)$ onto **the line spanned by the vector** $\underline{b}$ **with** $||b|| = 1$.

- **When** $90° < \theta < 270° \Rightarrow \cos(\theta) < 0$ **: In that case,** $\underline{a}^T\underline{b} =$ **negative distance from the origin to the perpendicular projection of the point** $\underline{a} = (a_1, a_2)$ onto **the line spanned by the vector** $\underline{b}$ **with** $||b|| = 1$.

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 now create two column vectors $\underline{x}_{s1}$ and $\underline{x}_{s2}$ that take the values of the standardized variables for height and weight.

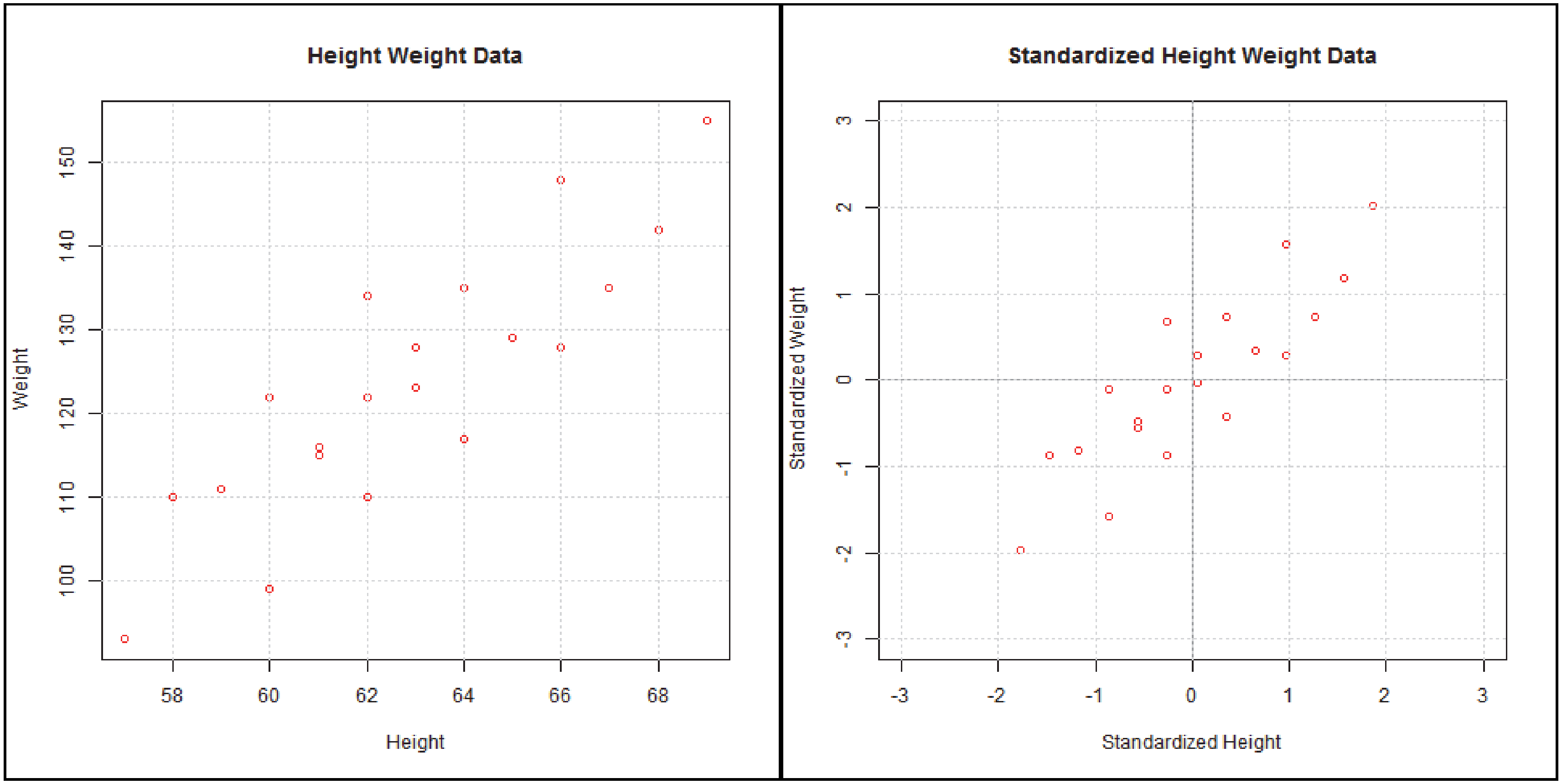
- Together, these two column vectors form a (20×2) matrix given by:

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

- **Each row of these matrices corresponds to one object (person)** measured on **each of 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.