Unit-1

Multivariate Normal Distribution

Basics:

Let X be a set of *p* variables each has n values.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_p \end{bmatrix} \rightarrow X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ . & . & \dots & . \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

Col-wise Mean:

$$\boldsymbol{\mu} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} x_i$$

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ . \\ \boldsymbol{\mu}_p \end{bmatrix}$$

Variance

According to layman's words, the variance is a measure of how far a set of data are dispersed out from their mean or average value. It is denoted as 'σ^2'.

Properties of Variance

- It is always non-negative since each term in the variance sum is squared and therefore the result is either positive or zero.
- Variance always has squared units. For example, the variance of a set of weights estimated in kilograms will be given in kg squared. Since the population variance is squared, we cannot compare it directly with the mean or the data themselves.

Standard deviation

The spread of statistical data is measured by the standard deviation. Distribution measures the deviation of data from its mean or average position. The degree of dispersion is computed by the method of estimating the deviation of data points. It is denoted by the symbol, 'σ'.

Properties of Standard Deviation

- It describes the square root of the mean of the squares of all values in a data set and is also called the root-mean-square deviation.
- The smallest value of the standard deviation is 0 since it cannot be negative.
- When the data values of a group are similar, then the standard deviation will be very low or close to zero. But when the data values vary with each other, then the standard variation is high or far from zero.

Variance, Std.Div. calculation

	Population	Sample
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$	$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

The population variance formula is given by:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2$$

Here,

σ^2 = Population variance

N = Number of observations in population

X_i = ith observation in the population

μ = Population mean

The sample variance formula is given as:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

Here,

s^2 = Sample variance

n = Number of observations in sample

x_i = ith observation in the sample

$\overline{x}$ = Sample mean

The population standard deviation formula is given as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2}$$

Here,

σ = Population standard deviation

Similarly, the sample standard deviation formula is:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

Here,

s = Sample standard deviation

Covariance:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{12} & \dots & x_{2n} \end{bmatrix}$$

Where first row is x_1 and second row is x_2

$$cov(x_1, x_2) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1i} - \mu_1)(x_{2i} - \mu_2)$$

Subtraction mean from each value turn their distribution towards origin and we have new X. Hence, after the subtraction process, mean of x_1 (μ_1) and x_2 (μ_2) become 0.

Since we have to do matrix multiplication, consider transpose of X and the covariance is

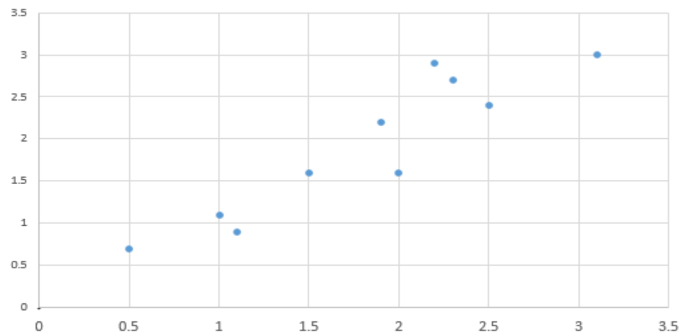
$$cov(x_1, x_2) = \frac{1}{n-1}X^T X$$

Covariance of two variable matrix and 3 variable matrix:

$$\begin{array}{c c} & \begin{array}{cc} x & y \end{array} \\ \begin{array}{c} x \\ y \end{array} & \begin{bmatrix} var(x) & cov(x,y) \\ cov(x,y) & var(y) \end{bmatrix} \end{array} \qquad \begin{array}{c c} & \begin{array}{ccc} x & y & z \end{array} \\ \begin{array}{c} x \\ y \\ z \end{array} & \begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(x,y) & var(y) & cov(y,z) \\ cov(x,z) & cov(y,z) & var(z) \end{bmatrix} \end{array}$$

Covariance example:

	X1	X2
1	2.5	2.4
2	0.5	0.7
,	2.2	2.9
,	1.9	2.2
,	3.1	3
,	2.3	2.7
,	2	1.6
,	1	1.1
,	1.5	1.6
10	1.1	0.9



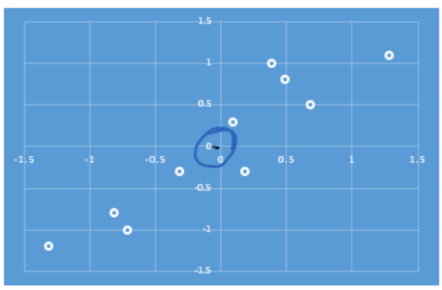
x1	x2
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

$\sum_{1 \text{ to } 10} / 10$

$\bar{x}_1$ μ_1 $\qquad$ $\bar{x}_2$ μ_2

x1=x1-1.81	x2=x2-1.91
2.5-1.81 0.69	2.4-1.91 0.49
0.5-1.81 -1.31	0.7-1.91 -1.21
2.2-1.81 0.39	0.99
0.09	0.29
1.29	1.09
0.49	0.79
0.19	-0.31
-0.81	-0.81
-0.31	-0.31
-0.71	-1.01

mean ≈ 0 $\qquad$ 0



11

$$X = \begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -0.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -0.01 \end{pmatrix}$$

μ_1 μ_2

- $\bar{X} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$C = \frac{1}{N-1}(X - \underline{1}\overline{X^T})^T(X - 1\overline{X^T})$$

$$C = \frac{1}{N-1} X^T X$$

$$cov(x_1, x_2) = \frac{1}{N-1} X^T X$$

X^T X

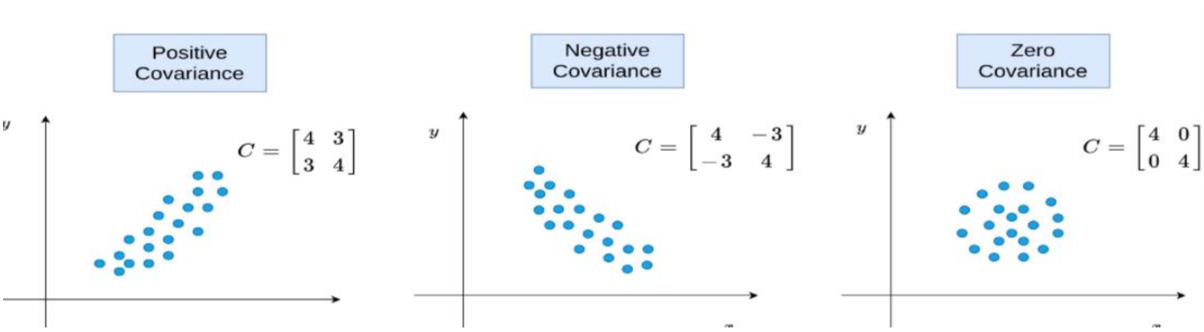
$$cov(x_1, x_2) = \frac{1}{10-1}\begin{pmatrix} 0.69 & -0.31 & 0.39 & 0.09 & 1.29 & 0.49 & 0.19 & -0.81 & -0.31 & -0.71 \\ 0.49 & -0.21 & 0.99 & 0.29 & 1.09 & 0.79 & -0.31 & -0.81 & -0.31 & -0.01 \end{pmatrix} \begin{pmatrix} 0.69 & 0.49 \\ -1.31 & -0.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -0.01 \end{pmatrix}$$

2×10 10×2

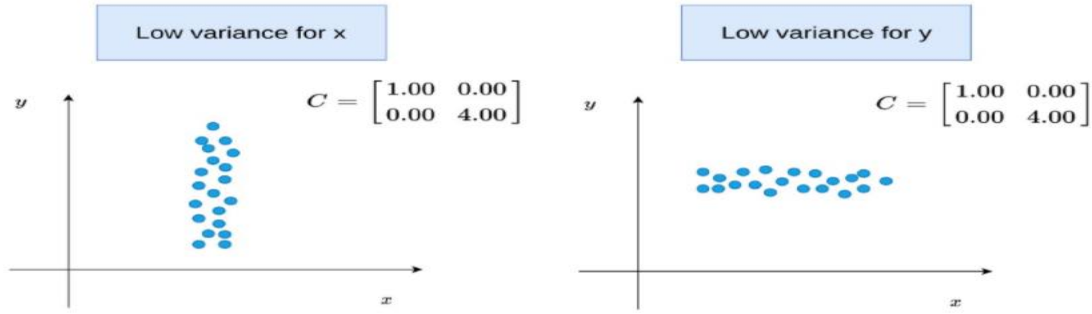
$$cov(x_1, x_2) = \begin{pmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{pmatrix}$$

Positive, Negative, and Zero States of The Covariance

It means variable X and variable Y variate in the same direction. In other words, if a value in variable X is higher, it is expected to be high in the corresponding value in variable Y too. In short, there is a positive relationship between them. If there is a negative covariance, this is interpreted right as the opposite. That is, there is a negative relationship between the two variables.

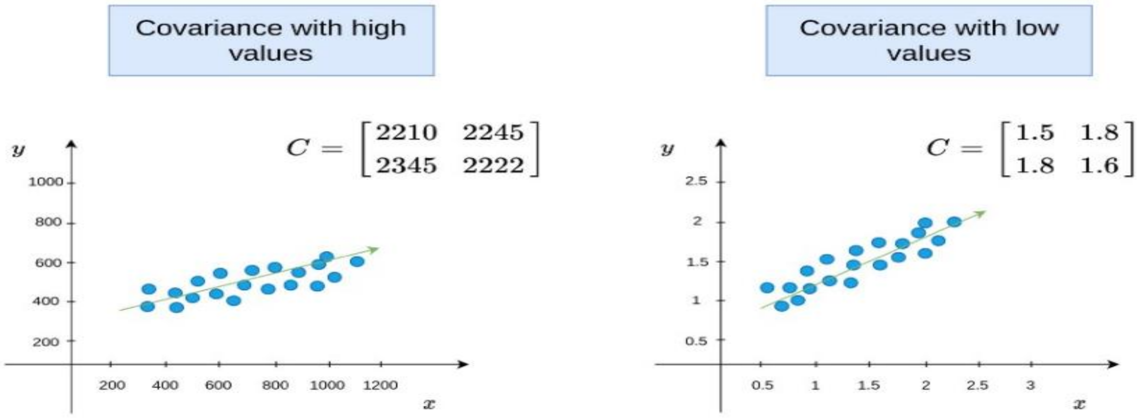


It happens while the covariance is near zero and the variance of variables are different



The size of covariance value

Unlike correlation, covariance values do not have a limit between -1 and 1. Therefore, it may be wrong to conclude that there might be a high relationship between variables when the covariance is high. The size of covariance values depends on the difference between values in variables. For instance, if the values are between 1000 and 2000 in the variable, it possible to have high covariance. However, if the values are between 1 and 2 in both variables, it is possible to have a low covariance. Therefore, we can't say the relationship in the first example is stronger than the second. The covariance stands for only the variation and relation direction between two variables.



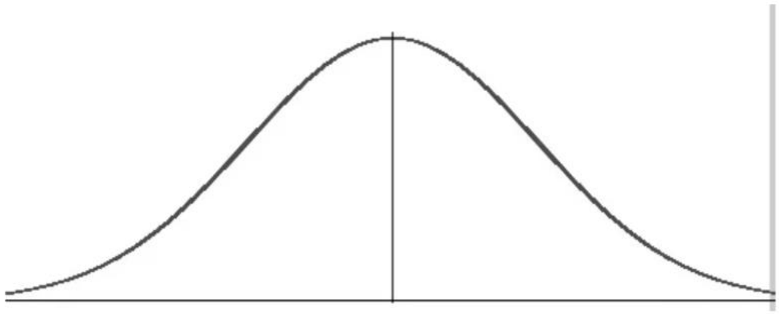
Correlation:

$$\rho_{x_1,x_1} = \frac{cov(x_1, x_1)}{\sigma x_1 \sigma x_2}$$

Number of variables in data= order of correlation / covariance matrix

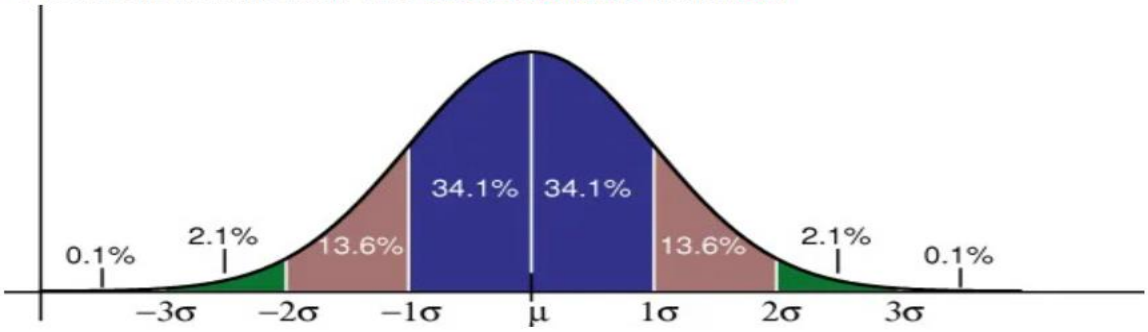
1.Normal Distribution:

- A normal distribution, sometimes called the bell curve
- Example: bell curve is seen in tests like the SAT and GRE.
- The bulk of students will score the average (C),
- while smaller numbers of students will score a B or D.
- An even smaller percentage of students score an F or an A.
- The bell curve is symmetrical.
- Half of the data will fall to the left of the mean; half will fall to the right.



A normal distribution.

- 68% of the data falls within one standard deviation of the mean.
- 95% of the data falls within two standard deviations of the mean.
- 99.7% of the data falls within three standard deviations of the mean.



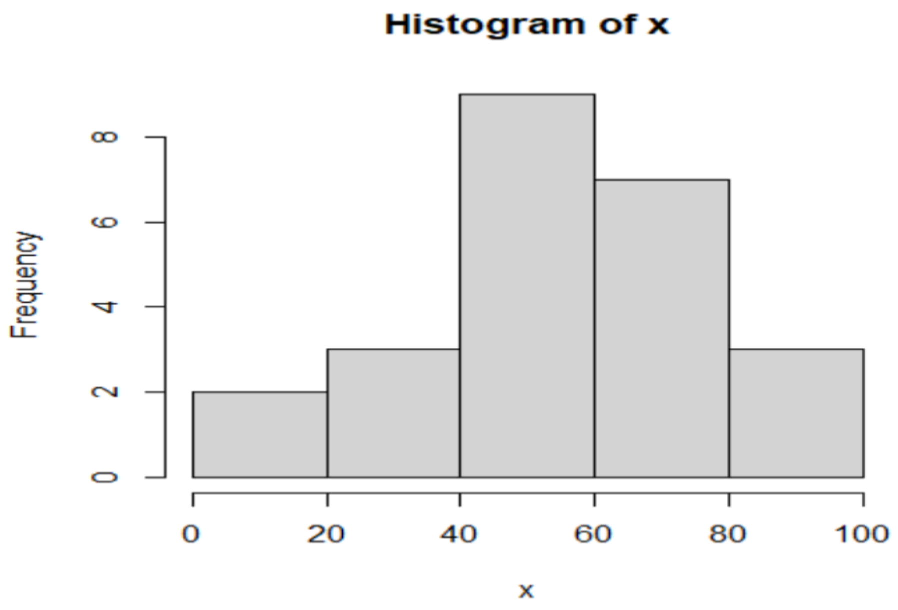
- A smaller standard deviation indicates that the data is tightly clustered around the mean; the normal distribution will be taller.
- A larger standard deviation indicates that the data is spread out around the mean; the normal distribution will be flatter and wider.

ND:Properties of a normal distribution

- The mean, mode and median are all equal.
- The curve is symmetric at the center (i.e. around the mean, μ).
- Exactly half of the values are to the left of center and exactly half the values are to the right.
- The total area under the curve is 1.

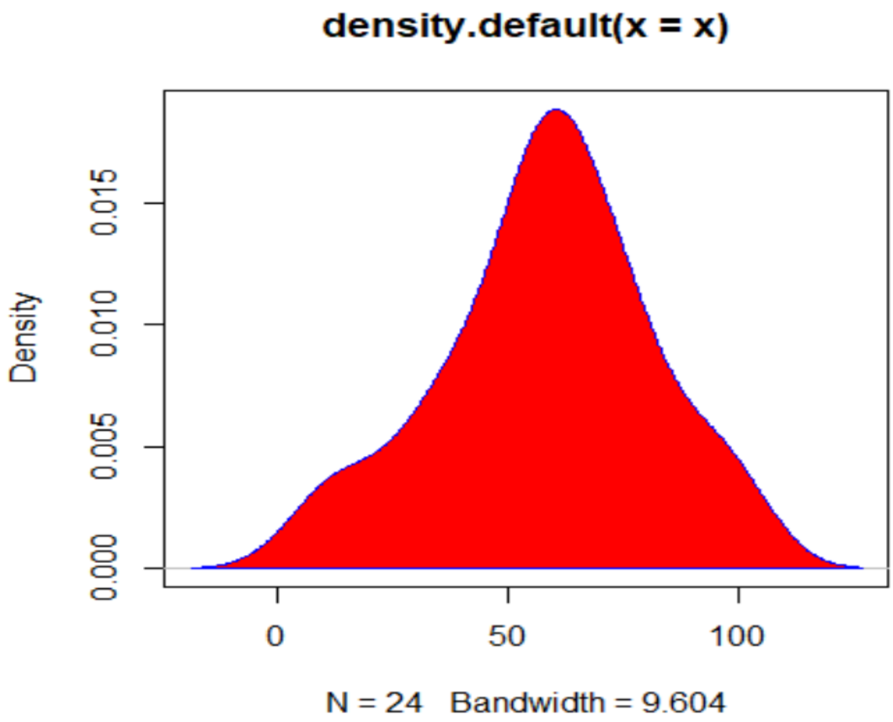
Example :R –code and output

X= c(10,15,38,33,44,48,50,58,59,58,60,61,65,58,30,74,70,74,75,79,88,98,98,58) hist(x)



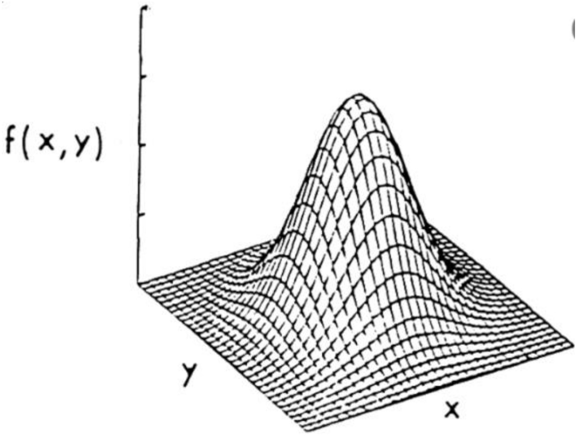
d<-density(x)

polygon(d, col="red", border="blue")




```
print(mean(x)
58.375
print(median(x)
58.5
getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
mm=getmode(x)
print(mm)
58
```

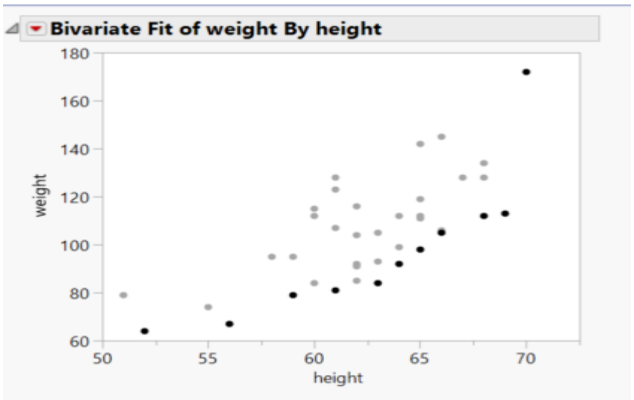
2. **Bivariate Normal Distribution**:



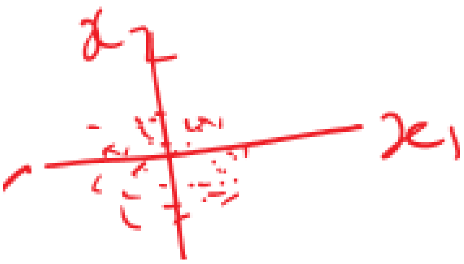
$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Example: dependent – correlation !=0



Independent: Corrlation (ρ_{12})=0; std.div(σ_{12})=0



2.1 Bivariate Normal Distribution- derivation

Let x_1, x_2 are independent with population density function N($\mu_1, {\sigma_1}^2$), N($\mu_2, {\sigma_2}^2$), σ_{12}=0

$$f(x_1) = \frac{1}{\sqrt{2\pi{\sigma_1}^2}} . e^{\frac{-1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2} \qquad -\propto < x_1 < \propto$$

$$f(x_2) = \frac{1}{\sqrt{2\pi{\sigma_2}^2}} . e^{\frac{-1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2} \qquad -\propto < x_2 < \propto$$

$$f(x_1, x_2) = f(x_1) \times f(x_2)$$

$$f(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1{}^2}} . e^{\frac{-1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2} \times \frac{1}{\sqrt{2\pi\sigma_2{}^2}} . e^{\frac{-1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}$$

$$f(x_1, x_2) = \frac{1}{(2\pi)^{\frac{2}{2}}(\sigma_1{}^2\sigma_2{}^2)^{\frac{1}{2}}} . e^{\frac{-1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2} \quad \text{--(1)}$$

Constant . Exponent

We need to derive these two parameters from population. Consider constant term,

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2{}^2 \end{bmatrix} \quad \text{--- (2)}$$

$|\Sigma| = \sigma_1{}^2\sigma_2{}^2 - \sigma_{12}{}^2$ (since σ_{12} *and* σ_{21} *are same*) ---(3)

Since x_1 and x_2 are independent, $\sigma_{12}{}^2 = 0$ -----(4)

$|\Sigma|^{\frac{1}{2}} = (\sigma_1{}^2\sigma_2{}^2)^{\frac{1}{2}}$ substituting this to constant term of equation (1),

$$f(x_1, x_2) = \frac{1}{(2\pi)^{\frac{2}{2}}|\Sigma|^{\frac{1}{2}}} \quad --(5)$$

Now, consider 3 variables,

Let, $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $\qquad \Sigma = \begin{bmatrix} \sigma_1{}^2 & 0 & 0 \\ 0 & \sigma_2{}^2 & 0 \\ 0 & 0 & \sigma_3{}^2 \end{bmatrix}$

$$f(x_1, x_2, x_3) = f(x_1) \times f(x_2) \times f(x_3)$$

The constant term for 3 variable is,

$$f(x_1, x_2, x_3) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma|^{\frac{1}{2}}}$$

The constant term for *P* variable is,

$$f(x_1, x_2, x_3) = \frac{1}{(2\pi)^{\frac{P}{2}}|\Sigma|^{\frac{1}{2}}}$$

We are going to derive the exponent term by considering Univariate p.d.f

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} . e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Exponent term:

$$\frac{-1}{2}(x-\mu)^2(\sigma^2)^{-1}$$

$$\frac{-1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)$$

$$\frac{-1}{2}(x-\mu)\Sigma^{-1}(x-\mu)$$

$$\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \quad \text{(note } X^TX=X^2\text{)}$$

From this, we can get bivariate exponent term:

$$-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\begin{bmatrix} {\sigma_1}^2 & 0 \\ 0 & {\sigma_2}^2 \end{bmatrix}^{-1}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \quad \text{----(6)}$$

From this, we can get multivariate exponent term:

$$\frac{-1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu) \quad \text{--(7)} \qquad \text{(note } X^TX=X^2\text{)}$$

Where X has set of variables.

Ref.work for understanding:

Let A be a matrix. $A^{-1} = \frac{Adj(A)}{det.A}$

Adj(A)=cofactor A^T

We know that $\Sigma = \begin{bmatrix} {\sigma_1}^2 & 0 \\ 0 & {\sigma_2}^2 \end{bmatrix}$ (from eq 2,3,4)

$$|\Sigma| = {\sigma_1}^2{\sigma_2}^2$$

$$cofactor\ \Sigma = \begin{bmatrix} (-1)^{1+1}{\sigma_2}^2 & 0 \\ 0 & (-1)^{2+2}{\sigma_1}^2 \end{bmatrix}$$

$$(cofactor\ \Sigma)^T = \begin{bmatrix} {\sigma_2}^2 & 0 \\ 0 & {\sigma_1}^2 \end{bmatrix}$$

Hence,

$$\Sigma^{-1} = \frac{1}{{\sigma_1}^2{\sigma_2}^2}\begin{bmatrix} {\sigma_2}^2 & 0 \\ 0 & {\sigma_1}^2 \end{bmatrix}$$

Substituting this in eq (6),

$$-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\frac{1}{{\sigma_1}^2{\sigma_2}^2}\begin{bmatrix} {\sigma_2}^2 & 0 \\ 0 & {\sigma_1}^2 \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \qquad \text{---(A)}$$

$$-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\frac{1}{{\sigma_1}^2{\sigma_2}^2}\begin{bmatrix} {\sigma_2}^2(x_1-\mu_1) \\ {\sigma_1}^2(x_2-\mu_2) \end{bmatrix}$$

$$-\frac{1}{2{\sigma_1}^2{\sigma_2}^2}\left[{\sigma_2}^2(x_1-\mu_1)^2+{\sigma_1}^2(x_2-\mu_2)^2\right]$$

$$-\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{{\sigma_1}^2}+\frac{(x_2-\mu_2)^2}{{\sigma_2}^2}\right] \qquad --(8)$$

Hence, for bivariate, from eq (5) and (8) equation (1) is derived,

$$f(x_1,x_2)=\frac{1}{(2\pi)^{\frac{2}{2}}({\sigma_1}^2{\sigma_2}^2)^{\frac{1}{2}}}\cdot e^{-\frac{1}{2}\left[\frac{(x_1-\mu_1)^2}{{\sigma_1}^2}+\frac{(x_2-\mu_2)^2}{{\sigma_2}^2}\right]} --(9)$$

From (9) and (7),

$$f(x_1,x_2)=\frac{1}{(2\pi)^{\frac{2}{2}}|\Sigma|^{\frac{1}{2}}}\cdot e^{\frac{-1}{2}\left[(X-\mu)^T\Sigma^{-1}(X-\mu)\right]} --(10)$$

For P variables,

$$f(x_1,x_2\ldots x_p)=\frac{1}{(2\pi)^{\frac{P}{2}}|\Sigma|^{\frac{1}{2}}}\cdot e^{\frac{-1}{2}\left[(X-\mu)^T\Sigma^{-1}(X-\mu)\right]}$$

$$-\propto< x_1 <\propto$$

2.2 Bivariate for dependent variables

$$X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mu=\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma=\begin{bmatrix} {\sigma_1}^2 & \sigma_{12} \\ \sigma_{21} & {\sigma_2}^2 \end{bmatrix}$$

$|\Sigma| = {\sigma_1}^2{\sigma_2}^2 - {\sigma_{12}}^2$ --- (11)

Note, Cor(x_1,x_2)= cov(x_1,x_2)/(σ x_1 . σ x_2)

$$\rho_{12}=\frac{\sigma_{12}}{\sigma_1\sigma_2}$$

$\sigma_{12} = \rho_{12}\ \sigma_1\sigma_2$ ----(12)

Substituting eq(12) in (11),

$$|\Sigma| = {\sigma_1}^2{\sigma_2}^2 - (\rho_{12}\ \sigma_1\sigma_2)^2$$

$$|\Sigma| = {\sigma_1}^2{\sigma_2}^2 - {\sigma_{12}}^2 \quad = {\sigma_1}^2{\sigma_2}^2(1-{\rho_{12}}^2) \qquad --(13)$$

We know,

$$\Sigma^{-1}=\frac{1}{{\sigma_1}^2{\sigma_2}^2-{\sigma_{12}}^2}\begin{bmatrix} {\sigma_2}^2 & -\sigma_{12} \\ -\sigma_{21} & {\sigma_1}^2 \end{bmatrix}$$ (Note: variables are dependent, so ${\sigma_{12}}^2 \neq 0$)

Consider exponent:

$$\frac{-1}{2}\left[(X-\mu)^T \Sigma^{-1}(X-\mu)\right]$$

$$-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\frac{1}{\sigma_1{}^2\sigma_2{}^2-\boldsymbol{\sigma_{12}}^2}\begin{bmatrix} \sigma_2{}^2 & -\boldsymbol{\sigma_{12}} \\ -\boldsymbol{\sigma_{21}} & \sigma_1{}^2 \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

(**Note how it differs from eq (A)**)

$$-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\frac{1}{\sigma_1{}^2\sigma_2{}^2-\boldsymbol{\sigma_{12}}^2}\begin{bmatrix} \sigma_2{}^2(x_1-\mu_1)-\boldsymbol{\sigma_{12}}(x_2-\mu_2) \\ -\boldsymbol{\sigma_{21}}(x_1-\mu_1)+\sigma_1{}^2(x_2-\mu_2) \end{bmatrix}$$

Order is 1×2 and 2×1, so result is 1×1,

$$-\frac{1}{2}\cdot\frac{1}{\sigma_1{}^2\sigma_2{}^2-\boldsymbol{\sigma_{12}}^2}\begin{bmatrix} \sigma_2{}^2(x_1-\mu_1)^2-\boldsymbol{\sigma_{12}}(x_2-\mu_2)(x_1-\mu_1)+ \\ -\boldsymbol{\sigma_{21}}(x_1-\mu_1)(x_2-\mu_2)+\sigma_1{}^2(x_2-\mu_2)^2 \end{bmatrix}$$

$$-\frac{1}{2}\cdot\frac{1}{\sigma_1{}^2\sigma_2{}^2-\boldsymbol{\sigma_{12}}^2}\left[\sigma_2{}^2(x_1-\mu_1)^2-\boldsymbol{2\sigma_{12}}(x_2-\mu_2)(x_1-\mu_1)+\sigma_1{}^2(x_2-\mu_2)^2\right]$$

$$-\frac{1}{2}\cdot\frac{\sigma_1{}^2\sigma_2{}^2}{\boldsymbol{\sigma_1{}^2\sigma_2{}^2-\sigma_{12}{}^2}}\left[\frac{(x_1-\mu_1)^2}{\sigma_1{}^2}-\frac{2\sigma_{12}}{\sigma_1{}^2\sigma_2{}^2}(x_2-\mu_2)(x_1-\mu_1)+\frac{(x_2-\mu_2)^2}{\sigma_2{}^2}\right]$$

From (**13**), and **(12)**

$$-\frac{1}{2}\cdot\frac{\sigma_1{}^2\sigma_2{}^2}{\boldsymbol{\sigma_1{}^2\sigma_2{}^2(1-\rho_{12}{}^2)}}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho_{12}\frac{(x_1-\mu_1)}{\sigma_1}\frac{(x_2-\mu_2)}{\sigma_2}+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$

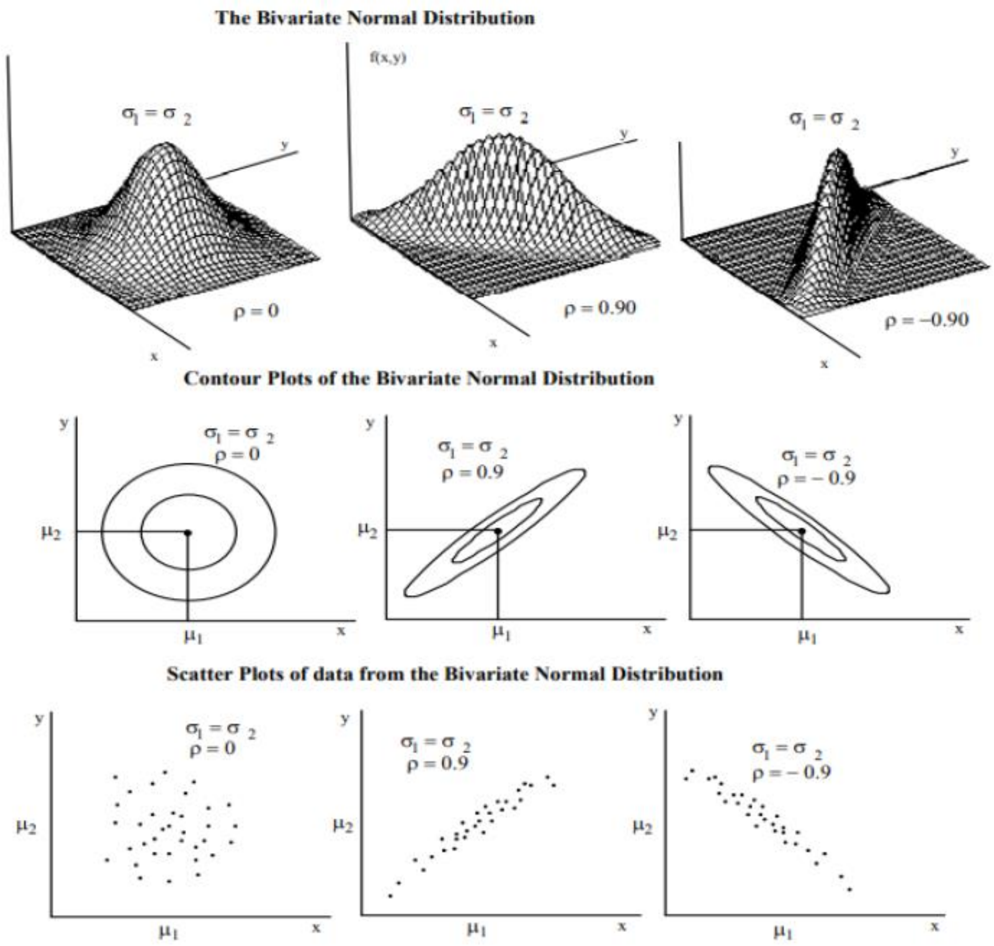
$$-\frac{1}{2}\cdot\frac{1}{\boldsymbol{(1-\rho_{12}{}^2)}}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho_{12}\frac{(x_1-\mu_1)}{\sigma_1}\frac{(x_2-\mu_2)}{\sigma_2}+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] \quad ---(14)$$

From eq(13), substitute $|\Sigma|$, and from eq (14) substitute exponent term

$$f(x_1, x_2)=\frac{1}{(2\pi)^{\frac{2}{2}}\sigma_1\sigma_2\sqrt{(1-\rho_{12}{}^2)}}\cdot e^{-\frac{1}{2}\cdot\frac{1}{(1-\rho_{12}{}^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho_{12}\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\ \ \sigma_2}+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]} \text{--(15)}$$

If $\boldsymbol{\rho_{12}} = 0$**, eq (14) becomes** ,

$$-\frac{1}{2}.\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]$$



Problem 1:

A process is X~N(μ,∑) is designed tp produce laminar aluminium sheet of length x_1 and breadth x_2 with the parameters. Obtain bivariate normal distribution.

$$\mu = \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} = \begin{bmatrix}100\\ 50\end{bmatrix} \qquad \Sigma = \begin{bmatrix}10 & 0\\ 0 & 5\end{bmatrix}$$

Solution: (ref. eq-9,10)

$$f\left(x_1,x_2\right) = \frac{1}{(2\pi)(\Sigma)^{\frac{1}{2}}}.e^{\frac{-1}{2}\left(\frac{x_1-100}{\sqrt{10}}\right)^2+\left(\frac{x_2-50}{\sqrt{5}}\right)^2}$$

3. MVND Properties

MVND Properties

1) If $X_{p\times 1}$ is distributed as $N_p(\mu, \Sigma)$

then x_j is also distributed as

$N(\mu_j, \sigma_j^2)$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{bmatrix}$$

Result 4.2. If $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any linear combination of variables $\mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$. Also, if $\mathbf{a}'\mathbf{X}$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for every $\mathbf{a}$, then $\mathbf{X}$ must be $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$a^T x = [1 \ 0 \ \cdots \ 0] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = x_1$$

$$a^T \mu = [1 \ 0 \ \cdots \ 0] \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu_1$$

$$a^T \Sigma a = \boxed{[1 \ 0 \ 0 \ \cdots \ 0]} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_{11}$$

x_1 is distributed as $N(\mu_1, \sigma_{11})$

x_i " $N(\mu_i, \sigma_{ii})$

Example 4.3 (The distribution of a linear combination of the components of a normal random vector) Consider the linear combination $\mathbf{a}'\mathbf{X}$ of a multivariate normal random vector determined by the choice $\mathbf{a}' = [1, 0, \ldots, 0]$. Since

$$\mathbf{a}'\mathbf{X} = [1, 0, \ldots, 0]\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = X_1$$

and

$$\mathbf{a}'\boldsymbol{\mu} = [1, 0, \ldots, 0]\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu_1$$

we have

$$\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = [1, 0, \ldots, 0]\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma_{11}$$

and it follows from Result 4.2 that X_1 is distributed as $N(\mu_1, \sigma_{11})$. More generally, the marginal distribution of any component X_i of $\mathbf{X}$ is $N(\mu_i, \sigma_{ii})$. ∎

The next result considers several linear combinations of a multivariate normal vector $\mathbf{X}$.

Result 4.3. If $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the q linear combinations

$$\underset{(q\times p)}{\mathbf{A}}\underset{(p\times 1)}{\mathbf{X}} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix} \qquad \mathbf{Y}_{q\times 1} = \mathbf{A}_{q\times p}\mathbf{X} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q1} & a_{q2} & \cdots & a_{qp} \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

are distributed as $N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$. Also, $\underset{(p\times 1)}{\mathbf{X}} + \underset{(p\times 1)}{\mathbf{d}}$, where $\mathbf{d}$ is a vector of constants, is distributed as $N_p(\boldsymbol{\mu} + \mathbf{d}, \boldsymbol{\Sigma})$.

Problem 2:

Numerical Example with Multiple Combinations

$$\mathbf{X} \sim \mathcal{N}_2\left(\begin{pmatrix} 5 \\ 10 \end{pmatrix}\begin{pmatrix} 16 & 12 \\ 12 & 36 \end{pmatrix}\right)$$

$$\begin{matrix} Y_1 = X_1 + X_2 \\ Y_2 = X_1 - X_2 \end{matrix} \quad \text{so} \quad \mathbf{A}_{2\times 2} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\mu_Y = \mathbf{A}\mu = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \begin{pmatrix} 15 \\ -5 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_Y = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 16 & 12 \\ 12 & 36 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 76 & -20 \\ -20 & 28 \end{pmatrix}$$

So

$$\mathbf{Y} \sim \mathcal{N}_2\left(\begin{pmatrix} 15 \\ -5 \end{pmatrix}, \begin{pmatrix} 76 & -20 \\ -20 & 28 \end{pmatrix}\right)$$

Result 4.4. All subsets of $\mathbf{X}$ are normally distributed. If we respectively partition $\mathbf{X}$, its mean vector $\boldsymbol{\mu}$, and its covariance matrix $\boldsymbol{\Sigma}$ as

$$\underset{(p\times 1)}{\mathbf{X}} = \begin{bmatrix} \underset{(q\times 1)}{\mathbf{X}_1} \\ \hline \underset{((p-q)\times 1)}{\mathbf{X}_2} \end{bmatrix} \qquad \underset{(p\times 1)}{\boldsymbol{\mu}} = \begin{bmatrix} \underset{(q\times 1)}{\boldsymbol{\mu}_1} \\ \hline \underset{((p-q)\times 1)}{\boldsymbol{\mu}_2} \end{bmatrix}$$

and

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \left[\begin{array}{c|c} \underset{(q\times q)}{\boldsymbol{\Sigma}_{11}} & \underset{(q\times(p-q))}{\boldsymbol{\Sigma}_{12}} \\ \hline \underset{((p-q)\times q)}{\boldsymbol{\Sigma}_{21}} & \underset{((p-q)\times(p-q))}{\boldsymbol{\Sigma}_{22}} \end{array}\right]$$

then $\mathbf{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Proof. Set $\underset{(q\times p)}{\mathbf{A}} = \left[\underset{(q\times q)}{\mathbf{I}} \;\vdots\; \underset{(q\times(p-q))}{\mathbf{0}}\right]$ in Result 4.3, and the conclusion follows. To apply Result 4.4 to an *arbitrary* subset of the components of $\mathbf{X}$, we simply relabel the subset of interest as $\mathbf{X}_1$ and select the corresponding component means and covariances as $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_{11}$, respectively. ■

Example 4.6 (The equivalence of zero covariance and independence for normal variables) Let $\underset{(3\times 1)}{\mathbf{X}}$ be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Are X_1 and X_2 independent? What about (X_1, X_2) and X_3?

Since X_1 and X_2 have covariance $\sigma_{12} = 1$, they are not independent. However, partitioning $\mathbf{X}$ and $\boldsymbol{\Sigma}$ as

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \hline X_3 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \left[\begin{array}{cc|c} 4 & 1 & 0 \\ 1 & 3 & 0 \\ \hline 0 & 0 & 2 \end{array}\right] = \left[\begin{array}{c|c} \underset{(2\times 2)}{\boldsymbol{\Sigma}_{11}} & \underset{(2\times 1)}{\boldsymbol{\Sigma}_{12}} \\ \hline \underset{(1\times 2)}{\boldsymbol{\Sigma}_{21}} & \underset{(1\times 1)}{\boldsymbol{\Sigma}_{22}} \end{array}\right]$$

we see that $\mathbf{X}_1 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and X_3 have covariance matrix $\boldsymbol{\Sigma}_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore, (X_1, X_2) and X_3 are independent by Result 4.5. This implies X_3 is independent of X_1 and also of X_2. ■

Problem 3:

A statistics class takes two exams X (Exam 1) and Y (Exam 2) where the scores follow a bivariate normal distribution with parameters:

- $\mu_X = 70$ and $\mu_Y = 60$ are the marginal means
- $\sigma_X = 10$ and $\sigma_Y = 15$ are the marginal standard deviations
- $\rho = 0.6$ is the correlation coefficient

Suppose we select a student at random. What is the probability that. . .

(a) the student scores over 75 on Exam 2?

(b) the student scores over 75 on Exam 2, given that the student scored $X = 80$ on Exam 1?

(c) the sum of his/her Exam 1 and Exam 2 scores is over 150?

(d) the student did better on Exam 1 than Exam 2?

(e) $P(5X - 4Y > 150)$?

Answer for 1(a):

Note that $Y \sim \mathrm{N}(60, 15^2)$, so the probability that the student scores over 75 on Exam 2 is

$$\begin{aligned} P(Y > 75) &= P\left(Z > \frac{75 - 60}{15}\right) \\ &= P(Z > 1) \\ &= 1 - P(Z < 1) \\ &= 1 - \Phi(1) \\ &= 1 - 0.8413447 \\ &= 0.1586553 \end{aligned}$$

where $\Phi(x) = \int_{-\infty}^{x} f(z)\mathrm{d}z$ with $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ denoting the standard normal pdf (see R code for use of `pnorm` to calculate this quantity).

Answer for 1(b):

Note that $(Y|X = 80) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 60 + (0.6)(15/10)(80 - 70) = 69$

$\sigma_*^2 = \sigma_Y^2(1 - \rho^2) = 15^2(1 - 0.6^2) = 144$

If a student scored $X = 80$ on Exam 1, the probability that the student scores over 75 on Exam 2 is

$$\begin{aligned} P(Y > 75|X = 80) &= P\left(Z > \frac{75 - 69}{12}\right) \\ &= P(Z > 0.5) \\ &= 1 - \Phi(0.5) \\ &= 1 - 0.6914625 \\ &= 0.3085375 \end{aligned}$$

Answer for 1(c):

Note that $(X + Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where

$\mu_* = \mu_X + \mu_Y = 70 + 60 = 130$

$\sigma_*^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y = 10^2 + 15^2 + 2(0.6)(10)(15) = 505$

The probability that the sum of Exam 1 and Exam 2 is above 150 is

$$\begin{aligned} P(X + Y > 150) &= P\left(Z > \frac{150 - 130}{\sqrt{505}}\right) \\ &= P(Z > 0.8899883) \\ &= 1 - \Phi(0.8899883) \\ &= 1 - 0.8132639 \\ &= 0.1867361 \end{aligned}$$

Answer for 1(d):
Note that $(X - Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where
$\mu_* = \mu_X - \mu_Y = 70 - 60 = 10$
$\sigma_*^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y = 10^2 + 15^2 - 2(0.6)(10)(15) = 145$

The probability that the student did better on Exam 1 than Exam 2 is

$$\begin{aligned} P(X > Y) &= P(X - Y > 0) \\ &= P\left(Z > \frac{0 - 10}{\sqrt{145}}\right) \\ &= P(Z > -0.8304548) \\ &= 1 - \Phi(-0.8304548) \\ &= 1 - 0.2031408 \\ &= 0.7968592 \end{aligned}$$

Answer for 1(e):
Note that $(5X - 4Y) \sim \mathrm{N}(\mu_*, \sigma_*^2)$ where
$\mu_* = 5\mu_X - 4\mu_Y = 5(70) - 4(60) = 110$
$\sigma_*^2 = 5^2\sigma_X^2 + (-4)^2\sigma_Y^2 + 2(5)(-4)\rho\sigma_X\sigma_Y =$
$25(10^2) + 16(15^2) - 2(20)(0.6)(10)(15) = 2500$

Thus, the needed probability can be obtained using

$$\begin{aligned} P(5X - 4Y > 150) &= P\left(Z > \frac{150 - 110}{\sqrt{2500}}\right) \\ &= P(Z > 0.8) \\ &= 1 - \Phi(0.8) \\ &= 1 - 0.7881446 \\ &= 0.2118554 \end{aligned}$$

4.Conditinal Distribution

Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and for $1 \leq m < n$ define

$$\boldsymbol{X}_1 = (X_1, \ldots, X_m)^T, \qquad \boldsymbol{X}_2 = (X_{m+1}, \ldots, X_n)^T$$

We know that $\boldsymbol{X}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{X}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ where $\boldsymbol{\mu}_i = E(\boldsymbol{X}_i)$, $\boldsymbol{\Sigma}_{ii} = \mathrm{Cov}(\boldsymbol{X}_i)$, $i = 1, 2$.

Then $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be partitioned as

$$\boldsymbol{\mu} = \left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{array}\right], \qquad \boldsymbol{\Sigma} = \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]$$

where $\boldsymbol{\Sigma}_{ij} = \mathrm{Cov}(\boldsymbol{X}_i, \boldsymbol{X}_j)$, $i, j = 1, 2$. Note that $\boldsymbol{\Sigma}_{11}$ is $m \times m$, $\boldsymbol{\Sigma}_{22}$ is $(n - m) \times (n - m)$, $\boldsymbol{\Sigma}_{12}$ is $m \times (n - m)$, and $\boldsymbol{\Sigma}_{21}$ is $(n - m) \times m$. Also, $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^T$.

We assume that $\boldsymbol{\Sigma}_{11}$ is *nonsingular* and we want to determine the conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$.

$$f(x|y) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma_x\sqrt{(1-\rho^2)}} \cdot e^{-\frac{\left(x_1-\mu_1-\frac{\sigma_1}{\sigma_2}\rho(x_2-\mu_2)\right)^2}{2(1-\rho_{12}{}^2)\sigma_1^2}} \quad \text{---(16)}$$

Conditional Distribution

The conditional distribution of $\mathbf{X}_1$ given known values for $\mathbf{X}_2 = \mathbf{x}_2$ is a multivariate normal with:

$$\text{mean vector} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)$$
$$\text{covariance matrix} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Bivariate Case

Suppose that we have p = 2 variables with a multivariate normal distribution. The conditional distribution of X_1 given knowledge of x_2 is a normal distribution with

$$\text{Mean} = \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2)$$
$$\text{Variance} = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$$

Example 6-1: Conditional Distribution of Weight Given Height for College Men

Suppose that the weights (lbs) and heights (inches) of undergraduate college men have a multivariate normal distribution with mean vector $\mu = \begin{pmatrix}175\\71\end{pmatrix}$ and covariance matrix $\Sigma = \begin{pmatrix}550 & 40\\40 & 8\end{pmatrix}$.

The conditional distribution of X_1 weight given x_2 = height is a normal distribution with

$$\begin{aligned}\text{Mean} &= \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2)\\ &= 175 + \frac{40}{8}(x_2 - 71)\\ &= -180 + 5x_2\end{aligned}$$

$$\begin{aligned}\text{Variance} &= \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\\ &= 550 - \frac{40^2}{8}\\ &= 350\end{aligned}$$

For instance, for men with height = 70, weights are normally distributed with mean = -180 + 5(70) = 170 pounds and variance = 350. (So standard deviation $\sqrt{350} = 18.71$ = pounds)