Ministry of Higher Education and Scientific Research
Salahaddin University/Erbil
College of Science - Department of Mathematics

The Multivariate Normal Distribution

Final Year Research Project

Submitted by
Goran Khasro Sabah

Supervisor
Mrs. Namam J. Mahmoud

A Project submitted in partial fulfillment of the requirements for the Degree of B.Sc. in Mathematics

Abstract

This research project aims to explore the properties of the multivariate normal distribution and its applications, which is a statistical tool that can be used to analyze the correlation between two or more continuous variables. we used multivariate normal distribution, to analyze the correlation between Petal Width and (Petal Length, Sepal Width, Sepal Length), of the three iris species. The data were chosen in a stratified random manner from the population, so we took (50) plant from each flower spec. So, we get a sample of size (150).

Also in this study, Mathematical models will be developed to describe the joint distribution of variables, and the properties of the multivariate normal distribution will be analyzed in detail. The results of this research will provide insight into the relationship between Petal Width and (Petal Length, Sepal Width, Sepal Length), and will contribute to understanding how these variables are related.

This research will also provide a framework for future studies on the relationship between plant traits, length and width, with other traits, such as color, and with environmental characteristics, such as humidity. The results of this study can be used in various fields, such as agriculture, agronomy, and industry.

After processing the data statistically using the Statistical Bag for Social Sciences (SPSS) program, the results reached the following:

1. Correlation between petal width and sepal length is strong positive correlation.
2. Correlation between petal width and sepal width is moderate negative correlation.
3. Correlation between sepal length and sepal width is weak negative correlation.
4. Correlation between petal length and petal width is strong positive correlation.
5. Correlation between petal length and sepal length is strong positive correlation.
6. Correlation between petal length and sepal width is moderate negative correlation.

Keywords: multivariate normal distribution, positive definite, eigenvalues, Jacobian of the transformation, regression, R square.

Chapter One

1.1. Introduction:

The multivariate normal distribution is one of the most important of statistics distributions. It is powerful tool for understanding the relationships between different variables. In this Research we explored how to use it to model multiple variables related to each other. We also we proved that is a probability distribution function (p.d.f) using elementary matrix algebra. And discussed the mean vector, the covariance matrix, and some of their properties, and then we talked about their applications.

This project research consists of three chapters in chapter one we give all important definition that we need to find the probability density function of multivariate normal distribution and in chapter two there is two sections. Section 2.1 find multivariate normal distribution in two cases. Case one using moment generating function. Case two using theorem. Section 2.2 talked about that multivariate normal distribution is a probability distribution function (p.d.f) using seven stage. And finally in chapter three we talked about the applications of multivariate normal distribution, and by using multivariate normal distribution we modeled the relationship between Petal Width and (Petal Length, Sepal Width, Sepal Length) of the three iris species.

1.2. Review of some concepts:

Sample Space: The collection of every possible outcome of an experiment is known as the sample space of the experiment and is denoted by SS.

Random variable: Given a random experiment with a sample space. A function X, which assigns to each element C in sample space one and only one real number X(c) =X is called random variable.

Example:

SAMPLE	X_i
TTT	0
HTT,TTH,THT	1
THH,HTH,HHT	2
HHH	3

Symmetric Matrices: A real matrix A is symmetric if $A^T = A$. Equivalently, $A = (a_{ij})$ is symmetric if symmetric elements (mirror images in the diagonal) arc equal, i.e if each ($a_{ij} = a_{ji}$). (Note that A must be square in order for $A^T = A$).

Orthogonal Matrices: A real matrix A is said to be orthogonal if $AA^T = A^T A = I$.

Identity Matrix: The square matrix with 1's on the diagonal and $0's$ elsewhere, denoted by I_n or simply I, is called the identity (or unit) matrix.

Positive Definite: An n×n is a positive definite matrix mean if all eigenvalue are +ve.

Example: A=$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$

Solution:

$|A - \lambda I|$=0

$\left|\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right|$=$\begin{vmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{vmatrix}$ =$(2-\lambda)(3-\lambda) = 0$

From this

2-λ=0 and 3-λ=0

λ=2 and λ=3

Normal distribution: If x is a normal random variable with mean = μ and variance, = σ^2. Then the probability density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, -\infty < X < +\infty$$

Moment generating function:

$$M_X(t) = e^{ut+\frac{\sigma^2 t^2}{2}}$$

If $if\ \mu = 0\ and\ \sigma^2\ is\ called\ standard\ normal$

1.3. Transformation for variables of the continuous type:

A-Jacobian of the transformation for one continuous Random Variable: Let X be a continuous random variable having p.d.f defined by f(X) let A be the set defined by S.T f(x)≥ , $\forall_x \in A$, $consider\ y = u(x)$, where the set B defined by a 1-1 corresponding with A(i.e. A maps into B by a transformation) , let the inverse of y=$\mu(x)$ denoted by X=w(y) where X=$w(y)$ for each X$\in A$, if the derivative $\frac{dx}{dy}$ =w(y) is continuous and exist for all points y$\in B$,the p.d.f for y=w(y) can be defined through the distribution function for X as follows g(y) = $f[w(y)] * |w'(y)|$ where |w(y)| is called the Jacobian of the Transformation .

1.3.1. Jacobian of the transformation for two continuous random variables:

Let X_1, X_2 be two continuous random variables having the joint p.d.f given as F(X_1, X_2) Where the set X_1, X_2 defined by all ordered pairs (x_1, x_2) is s.t F$(x_1, x_2) \geq 0\ \forall x_1,\ x_2 \in A$ define $y_1 = \mu_1(x_1, x_2)$and $y_2 = (x_1, x_2)$, if the set B defined by (y_1, y_2) s.t it defines as 1-1 transformation with A,and the inverse function $x_1 = w_1(y_{1,}y_2)$, $x_2 = w_2(y_1, y_2)$
Are defined and the Jacobin is defined as

$$j_1 = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial x_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Assuming that there first order partial derivatives are continuous, then the joint p.d.f for y_1, y_2 can be defined through the joint p.d.f for x_1, x_2 and as follows:
G$(y_1, y_2) = f[w_1(y_1, y_2)\ \ , w_2(y_1, y_2)] * |J|$
= 0 other wise.

1.3.2. Jacobian of the transformation for n continuous random variables:

Let $x_{1}, x_2, \cdots, x_n$ be n continuous random variable having the joint p.d.f given as f x_n where the set $x_1, x_2, \cdots x_n$ defined by all ordered $(x_1, x_2, \cdots, x_n)$ s.t $(x_1, x_2, \cdots, x_n) \geq 0\ \forall \in A$
Defined $y_1 = u_1(x_1, x_2, \cdots, x_n), y_2 = u_2(x_1, x_2, \cdots, x_n), \cdots, y_n = u_n(x_1, x_2, \cdots, x_n)$ if the set B defined by $(y_1, y_2, \cdots, y_n)$ s.t is defines as 1-1 transformation with A , and the inverse function $x_1 = w_1(y_1, y_2, \cdots, y_n), x_2 = w_2(y_1, y_2, \cdots, y_n), \cdots\ x_n = w_n(y_1, y_2, \cdots, y_n)$
Are defined and the Jacobian is defined as

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

Assuming that there first order partial derivative are Continuous, than the joint p.d.f for $y_1, y_2, \cdots, y_n$
Can be defined through the joint p.d.f for $(x_1, x_2, \cdots, x_n)$ and as follows $g(y_1, y_2, \cdots, y_n) = f[w_1(y_1, y_2, \cdots, y_n), w_2(y_1, y_2, \cdots, y_n) \cdots, w_n(y_1, y_2, \cdots, y_n)]|J|(y_1, y_2, \cdots, y_n) \in B$

$$= 0 \qquad other\ wise$$

Theorem: Let A be a square matrix of order n. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.

Proof: Let A is diagonalizable. ⇒ ∃ an invertible matrix P of order n such that $D = P'AP$ is a diagonal matrix.

Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be the diagonal entries of D, i.e.

$$\mathrm{D} = P'\mathrm{AP} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Let $v_1, v_{2,} \cdots, v_n$ be the columns of P.Then

$$AP = A[v_1 \;\; v_2 \;\; \cdots \;\; v_n] = \lambda \;\; [Av_1 \; Av_2 \; \cdots Av_n], \qquad \ldots(1)$$

And

$$PD = P\begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \; [\lambda_1 v_1 \, \lambda_2 v_2 \cdots \, \lambda_n v_n]. \quad \ldots(2)$$

But $D = \; P'AP \; \Rightarrow \; PD = AP$.

Hence from (1) and (2) we get

$$[Av_1 \; Av_2 \; \cdots Av_n] = [\lambda_1 v_1 \, \lambda_2 v_2 \cdots \, \lambda_n v_n]$$

Equating columns, we get

$$Av_i = \lambda_i v_i, \text{ i=1,2, … ,n.} \quad \ldots(3)$$

Since P is invertible, its columns must be LI and therefore each column v_i must be non-zero. The equation (3) shows that $\lambda_1 \;, \lambda_2, \cdots \lambda_n$ are eigenvalues and $v_1, v_2, \cdots, v_n$ are corresponding LI eigenvectors of A corresponding to the eigenvalues $\lambda_1 \;, \lambda_2, \cdots \lambda_n$ respectively. Now we construct a matrix P whose columns are $v_1 v_2, \cdots, v_n$; and we also construct a diagonal matrix D with diagonal entries as. Then $\lambda_1 \;, \lambda_2, \cdots \lambda_n$ from (1) – (2), we note that

AP = PD … (4)

Since the columns $v_1, v_2, \cdots, v_n$ of P are LI, therefore P is invertible.

Thus, equation (4) yields

$$A = PDP' \;\; \Rightarrow \text{A is diagonalizable}$$

Chapter two

2.1. Multivariate Distribution.

Let $Y = (Y_1, Y_2, Y_3, \ldots\ldots\ldots\ldots, Y_n)'$ be an n×1 vector of random variables .we say that Y is an n-dimensional random vector .The mean vector , E(Y), and covariance matrices ,COV(Y) , are defined by

$$E(Y)=[E(Y_1)\quad E(Y_2)\quad E(Y_3)\quad \ldots\ldots\ldots..E\,(Y_n)]'$$

$$Cov(Y) = \begin{bmatrix} var(Y_1) & \cdots & cov(Y_1, Y_n) \\ \vdots & \ddots & \vdots \\ cov(Y_n, Y_1) & \cdots & var(Y_n) \end{bmatrix}$$

That is $E(Y)$ is the $n \times 1$ vector whose ith component is the $E(Y)$, $Cov(Y)$ is the symmetric n×n matrics whose ith diagonal element is $var(Y_i)$ and whose $(i,j)th$ off-diagonal element is $Cov(Y_i, Y_j)$. Note that var (Yi) can be thought of as $cov(Yi,Yj)$.

Lemma 1: Let x be a random vector with mean vector μ and covariance matrix ∑.
Let Y=AX +b, Then E(Y)=Aμ+b and COV(Y)=A∑ A′.

Proof: Y=AX +b

E(Y)=E(AX +b)
E(Y)=E(AX)+E(b)
E(Y)=AE(X)+b
COV(Y)=COV(AX +b)
COV(Y)=Aμ+b
COV(Y)=E[(Y-E(Y) $(Y - E(Y)\,')$]
COV(Y)=E[(AX +b-Aμ+b)$(AX + b - A\mu + b)'$]
COV(Y)= E[A (X-μ)$(X - \mu)'A'$]
COV(Y)=AE[(X-μ) $(X - \mu)'$]A'
COV(Y)=A∑A'.

Lemma 2:

a) Let X=AY +b Then $M_x(t)=e^{(b't)}M_y(A't)$
b) Let c∊ℝ. Let Z=cY. Then $M_z(t)= M_y(ct)$.

2.2. Finding The Multivariate Normal Distribution:

Case one:

Using moment generating function.

We now define the multivariate normal distribution and derive its basic properties. We want to allow the possibility of multivariate normal distributions whose covariance matrices is not necessarily positive definite. Therefore, we cannot define the distribution by its density function. Instead we define the distribution by its moment generating function. (The reader may wonder how a random vector can have a moment generating function if it has no density function. However, the moment generating function can be defined using more general types of integration). We find the density function for the case of positive definite covariance matrix in theorem

for the motivation, we start with Z_1,Z_2,……..,Z_n independent random variables such that Z_i~N (0,1). Let Z=(Z_1,Z_2,……..,Z_n)$'$. Then

$$E(Z)=0 \qquad COV(Z)=I \qquad M_Z(t)=\prod \exp\frac{t_i^{2}}{2}=\exp\frac{tt'}{2} \qquad (1.1)$$

Let μ be an n x 1 vector and A an n × n matrix. Let Y=AZ+μ.Then

$$E(Y)=\mu \qquad COV(Y)=AA' \qquad (1.2)$$

Let ∑=AA$'$. We now show shat the distribution of Y depends only on μ and ∑. The moment generating function M_Z(t) is given by

$$M_z(t)=\exp(\mu' t)\, M_z(A' t)=\exp\left[\ \mu' t+\frac{t'(AA')t}{2}\ \right]=\exp\left(\ \mu' t+\frac{t'\Sigma t}{2}\ \right) \qquad (1.3)$$

With this motivation in mind, Let μ be an n x 1 vector, and let ∑ be a nonnegative definite n × n matrix. Then we say that the n-dimensional random vector Y has an n-dimensional normal distribution with mean vector μ, and covariance matrix ∑, if Y has moment generating function

$$M_Z(t)=\exp\left(\ \mu t' + \frac{t'\Sigma t}{2}\ \right) \qquad (1.4)$$

We write Y ~ N_n (μ, ∑) .The following theorem summarize

some elementary facts about multivariate normal distribution.

Case two:

Theorem(1): Let Y ~ N_n (μ , ∑) ,with $\sum > 0$. Then Y has density $function$

$$f_y(y)= \frac{1}{(2\pi)^{n/2} \quad |\sum|^{1/2}} \exp\left(- \frac{1}{2} (y-\mu)' \sum^{-1} (y-\mu)\right) .$$

Proof: We could derive this by finding the moment generating function of this density and showing that it satisfied (1.4). We would also have to show that this function is a density function . we can avoid all that by starting with a random with a random vector whose distribution we know. Let

Z~ N_n(0 , 1) $Z \ = (Z_1, Z_2, \ldots \ldots .., Z_n)'$

Then the Z_i are independent and Z_i ~ N_1(0,1) by lemma 1.3 Therefore the joint density of the Z_i is

$$f_z(z) = \prod_{k=1}^{n} \frac{1}{(2\pi)^{1/2}} \ \exp(-\frac{1}{2} Z_i) = \frac{1}{(2\pi)^{n/2}} \ \exp(-\frac{1}{2} ZZ')$$

Let Y= $\sum^{-1/2}$Z+μ. By theorem 1.3 Y ~ N_n(μ,∑) .Also Z=$\sum^{1/2}$(Y-μ) ,and the transformation from Z to Y is therefore invertible . furthermore, the Jacobian of this inverse transformation is just $|\sum^{-1/2}|=|\sum|^{-1/2}$ .Hence the density of Y is

$$f_y(y)= \frac{1}{(2\pi)^{1/2} \quad |\sum|^{1/2}} \exp\left(- \frac{1}{2} (y-\mu)' \sum^{-1/2} (y-\mu)\right) .$$

2.3. Probability Density Function of the Multivariate Normal Distribution.

Let A denote an $n \times n$ real symmetric matrix which is positive definite. Let μ denote the n x 1 matrix such that μ', the transpose of μ, is μ'=$[\mu_{1,}\mu_2, \cdots, \mu_n]$ where each μ_i is a real constant. Finally, let x denote the $n \times 1$ matrix such that $x' = [x_1, x_2, \cdots, x_n]$ We shall show that if C is an appropriately chosen positive constant, the nonnegative function

$$f(x_1 . x_2, \cdots, x_n) = c\ \exp[-\frac{(x-\mu)A(x-\mu)}{2}], \qquad -\infty < x_i < \infty, i = 1,2, \cdots n,$$

$$\text{where } c = \frac{1}{(2\pi)^{\frac{n}{2}} |A|^{\frac{1}{2}}}$$

is a joint p.d.f. of n random variables $X_1, X_{2,} \cdots, X_n$ that are of the Continuous type. Thus we need to show that

(1) $$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1 . x_2, \cdots, x_n) = 1.$$

Let t denote the n x 1 matrix such that $t' = [t_1, t_2, \cdots, t_n]$, where $t_1, t_2, \cdots, t_n$ are arbitrary real numbers. We shall evaluate the integral

(2) $$C\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} exp[t'X - \frac{(X-\mu)' A(X-\mu)}{2}] dx_1 \cdots dx_n$$

and then we shall subsequently set $t_1 = t_2 = \cdots = t_n$ = 0, and thus establish Equation (1). First we change the variables of integration in integral (2) from,x_1 , x_2, $\cdots, x_n$ to y_1, y_2,$\cdots, y_n$by writing x -μ =y, where y' =[$y_1, y_2, \cdots, y_n$]The Jacobian of the transformation is one and the n-dimensional x-space is mapped onto an n-dimensional y-space, so that integral (2) may be written as

(3) $$C \exp(t'\mu) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(t'y - \frac{y'Ay}{2}\right) dy_1 \cdots dy_n$$

Because the real symmetric matrix A is positive definite, the n characteristic numbers (proper values, latent roots, or eigenvalues) $a_1, a_2, \cdots, a_n$ of A are positive. There exists an appropriately chosen n x n real orthogonal matrix L (L' =L^{-1} , whereL^{-1} is the inverse of L) such that

$$L'AL = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}$$

for a suitable ordering of $a_1, a_2, \cdots, a_n$. We shall sometimes write L'AL= diag $[a_1, a_2, \cdots, a_n]$.In integral (3), we shall change the variables of integration from $y_1, y_2, \cdots, y_n$ to $z_1, z_2, \cdots, z_n$ by writing y = Lz, where z' = $[z_1, z_2, \cdots, z_n]$ .The Jacobian of the transformation is the determinant of the orthogonal matrix L. Since $L'L = I_n$ where I_n is the unit matrix of order *n*, we have the determinant $|L'L| = 1$.
And $|L|^2$. Thus the absolute value of the Jacobian is one. Moreover, the n-dimensional y-space is mapped onto an n-dimensional z-space.
The integral (3) becomes

(4) $$C \exp(t'\mu) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} exp\left[t'Lz - \frac{z'(L'AL)z}{2}\right] dz_1 \cdots dz_n.$$

It is computationally convenient to write, momentarily, t'L = w',
where w' = $[w_1, w_2, \cdots w_n]$. Then

$$exp[t'Lz] = exp[w'z] = exp(\textstyle\sum_1^n w_i z_i).$$

Moreover,

$$\exp\left[-\frac{z'(L'AL)z}{2}\right] = \exp\left[-\frac{\sum_1^n a_i z_i^2}{2}\right]$$

Then integral (4) may be written as the product of *n* integrals in the
Following manner:

(5) $$C \exp(w'L\mu) \prod_{t=1}^{n} \left[\int_{-\infty}^{\infty} exp\left(w_i z_i - \frac{a_i z_i^2}{2}\right) dz_i\right]$$

$$= C\, exp(w'L'\mu) \prod_{t=1}^{n} \left[\sqrt{\frac{2\pi}{a_i}} \int_{-\infty}^{\infty} \frac{exp\left(w_i z_i - \frac{a_i}{2}\right)}{\sqrt{\frac{2\pi}{a_i}}} dz_i\right]$$

The Integral. That involves $z_i$can be treated as the moment-generating function with the more familiar symbol *t* replaced by w_i of a distribution which IS n(0,1/a_i) .Thus the right-hand member of Equation (5) is equal to

(6) $$C \exp(W'L'\mu) \prod_{i=1}^{n} \left[\sqrt{\frac{2\pi}{a_i}}\ exp\left(\frac{{W_i}^2}{2_{a_i}}\right)\right]$$

$$= C \exp(W'L'\mu)\sqrt{\frac{(2\pi)^n}{a_1 a_2 \cdots a_n}}\ \exp\left(\sum_1^n \frac{{W_i}^2}{2a_i}\right).$$

Now, because $L^{-1} = L'$, we have

$$(L'AL)^{-1} = L'A^{-1}L = diag\left[\frac{1}{a_1}, \frac{1}{a_2}, \cdots, \frac{1}{a_n}\right].$$

Thus $$\sum_1^n \frac{{w_i}^2}{a_i} = w'(L'A^{-1}L)w = (Lw)'A^{-1}(Lw) = t'A^{-1}t.$$

Moreover, the determinant $|A^{-1}|$ of $A^{-1}\ is$

$$|A^{-1}| = |L'A^{-1}L| = \frac{1}{a_1 a_2 \cdots a_n}.$$

Accordingly the right-hand member of Equation (6), which is equal to Integral (2), may be written as

(7) $$C\ e^{t'}\mu\ \sqrt{(\pi)^{\pi}|A^{-1}|}\ \exp(\frac{t'A^{-1}t}{2})$$

If, in this function we set $t_1 = t_2 = \cdots = t_n = 0, we\ have\ the\ value\ of\ the\ left - hand\ member\ of\ equation. Thus, we\ have$

$$C\sqrt{(2\pi)^n|A^{-1}|} = 1$$

Accordingly, the function

$f(x_1. x_2, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|A^{-1}|}} \exp\left[\frac{(X-\mu)'A(X-\mu)}{2}\right], \quad -\infty < x_i < \infty, i = 1,2, \cdots, n$

is a joint p.d.f of n random variables X_1,X_2, $\cdots$,X_n that are of the continuous type.Such a p.d.f is called a nonsingular multivariate normal p.d.f we have now proved that f($x_1. x_2, \cdots, x_n$) is a p.d.f. However, we have proved more than that. Because *f*($x_1. x_2, \cdots, x_n$) is a p.d.f, integral (2) is the moment- generating function M($t_1, t_2, \cdots, t_n$) of this joint distribution of probability. Since integral (2) is equal to function (7), the moment-generating function of the multivariate normal distribution is given by M($t_1, t_2, \cdots, t_n$) = $\exp\left(t'u + \frac{t'A^{-1}t}{2}\right)$.

Example: let X_1,X_2,⋯,X_n have a multivariate normal distribution with matrix μ of means and positive defi

Let the elements of the real, symmetric, and positive definite matrix A^{-1}be denoted by σ_{ij},i,j = 1,2,⋯ , n. Then

M$(0, \cdots, 0, t_i, 0, \cdots, 0)$ = exp$\left(t_i\mu_i + \frac{\sigma_{ii}t_i{}^2}{2}\right)$

Is the moment-generating function of $X_i, i = 1,2, \cdots, n$. Thus, X_i is $n(\mu_i, \sigma_{ii})i, 1,2, \cdots, n$. Moreover, with i≠j, we see that M$(0, \cdots, 0, t_i, 0, \cdots, 0)$,the moment-generating function of X_i and X_j, is equal to

$$exp\left(t_i\mu_i + t_j\mu_j + \frac{\sigma_{ii}t_i{}^2 + 2\sigma_{ij}t_it_j + \sigma_{ij}t_{jj}{}^2}{2}\right).$$

But this is the moment-generating function of a bivariate normal distribution,
so that σ_{ij} is the covariance of the random variables X_i and X_j.

Thus the matrix μ where μ'=$[\mu_1, \mu_2, \cdots, \mu_n]$, is the matrix of the means of the random variables $X_1, X_2, \cdots, X_n$.moreover,the elements on the principal diagonal of A^{-1} are,respectively,the variances σ_{ii}=$\sigma_i{}^2$, $i = 1,2, \cdots$,n, and the elements not on the principal diagonal of A^{-1} are, respectively, the covariances $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, i≠ j, of the random variables $X_1, X_2, \cdots, X_n$. We call the matrix A^{-1}, which is given by

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{bmatrix}$$

the covariance matrix of the multivariate normal distribution and
henceforth we shall denote this matrix by the symbol V. In terms of
the positive definite covariance matrix V, the multivariate normal
p.d.f. is written

$$\frac{-1}{(2\pi)^{\frac{n}{2}}\sqrt{|V|}} exp\left[\frac{(X-\mu)'V^{-1}(X-)}{2}\right], \quad -\infty < x_i < \infty\text{,i=1,2,}\cdots, n,$$

and the moment-generating function of this distribution is given by exp$\left(t'\mu + \frac{t'Vt}{2}\right)$

For all real values of t.

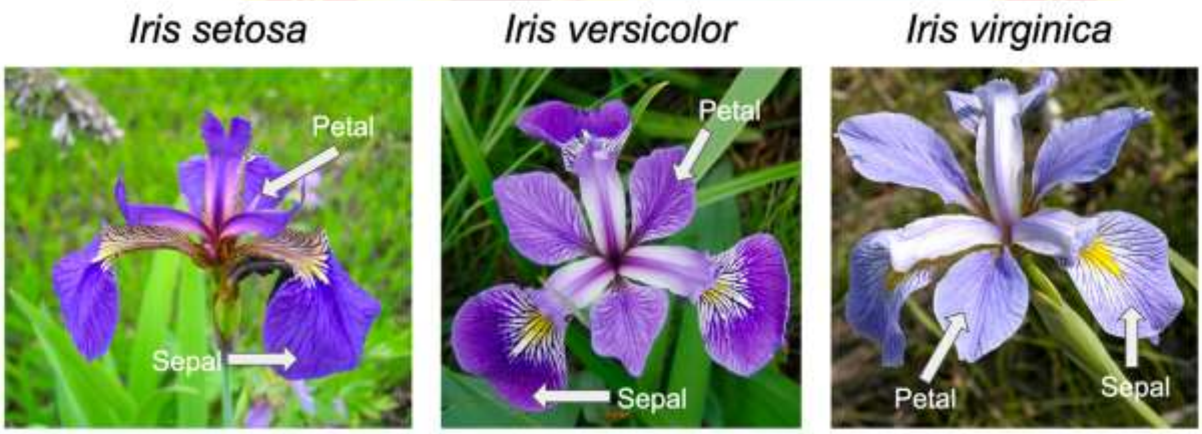
Chapter 3

3.1. Application: The Iris dataset was used in R.A. Fisher's classic 1936 paper, The Use of Multiple Measurements in Taxonomic Problems, and can also be found on the UCI Machine Learning Repository.
It includes each three iris species with 50 samples each as well as some properties about flower.

The columns in this dataset are:

- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm
- Species

Figure (1) Species



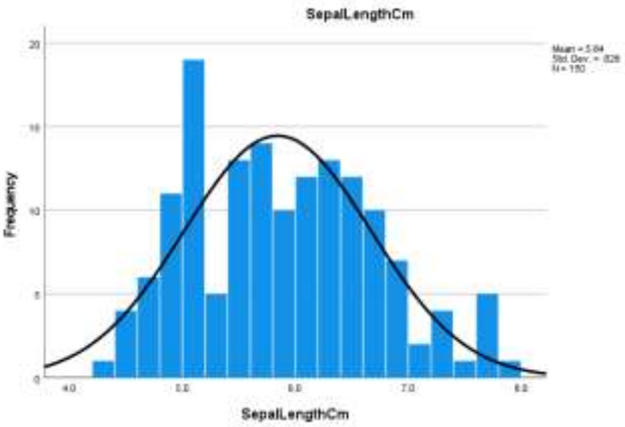
Definition: In multiple linear regression, there are p explanatory variables, and the relationship between the dependent variable and the explanatory variables is represented by the following equation

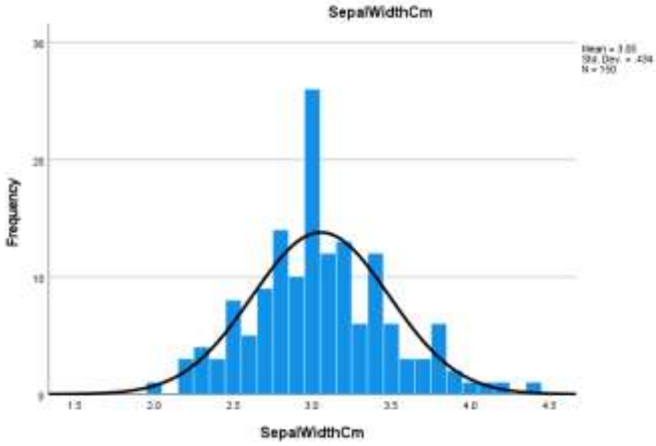
$$y_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\ldots\ldots+\beta_p x_{ip}+\varepsilon$$

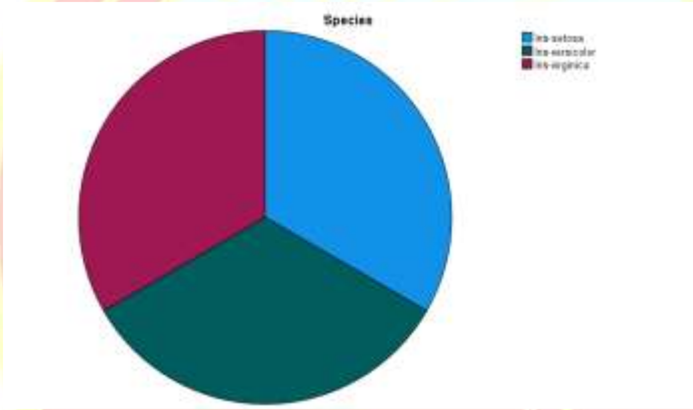
Where for i=n observations:

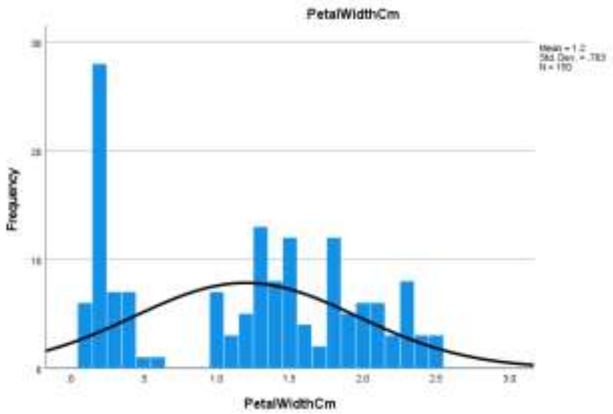
- β_0 is the constant term
- y_i=dependent variable
- x_i=explanatory variables
- β_p=slop coefficients for each explanatory variable
- ε =model error

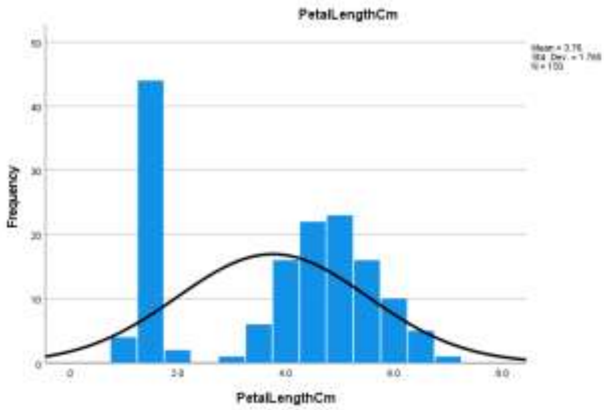
3.2. Frequency histograms:











3.3. The Statistical Measures that we used in this Research:

Mean: is the average of the numbers, a calculated “central” value of a set of numbers, and it is denoted by $\bar{x}$. $\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n}$

Variance: is the average of the squared differences form MEAN, and it is denoted by S^2.

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

Standard deviation: is a quantity calculated to indicate the extent of deviation for a group as a whole. In addition, it is denoted by $\boldsymbol{s}$, standard deviation is square root of variance.

Multiple Correlation Coefficient (R):

Is a statistical measure of the strength and direction of the linear relationship between the dependent variable and the set of independent variables in a multiple regression model. Its value ranges between -1 and 1, where -1 represents a perfect negative relationship, 1 represents a perfect positive relationship, and 0 represents no linear relationship.

$$R = \sqrt{\frac{\text{the regression sum of squares(SSR)}}{\text{the total sum of squares(SST)}}} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

where:

Y is the observed value of the dependent variable.

$\hat{\mathbf{Y}}$ is the predicted value of the dependent variable from the regression model.

Figure 2

$\bar{\boldsymbol{Y}}$ the mean of the dependent variable.

Coefficient of Determination (R-square):

Is a statistical measure of the proportion of the total variance in the dependent variable that is explained by the independent variables in the regression model. Its value also ranges between 0 and 1, where 0 indicates that the independent variables do not explain any variance, and 1 indicates that the independent variables explain all the variance in the dependent variable.

$$R^2 = \frac{\text{the regression sum of squares(SSR)}}{\text{the total sum of squares(SST)}} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \; or \; 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Adjusted R-square:

Is a statistical measure that is an adjusted version of the R-squared (R^2) value and provides a more conservative estimate of the model's goodness of fit by taking into account the number of predictors **k** and the sample size **n**, its value ranges from negative infinity to 1. Like R-square, a higher Adjusted R-square value indicates a better fit of the regression model.

$${R^2}_{adj} = 1 - \frac{(1 - R^2)\,(\mathrm{n} - 1)}{n - k - 1}$$

Standard Error of the Estimate:

Is a statistical measure that represents the standard deviation of the residuals. It indicates the typical amount of error or variability in the dependent variable that is not accounted for by the regression model. A smaller Std. Error of the Estimate indicates that the regression model provides a better fit to the data. Conversely, a larger Std. Error of the Estimate suggests that there is more variability in the dependent variable that cannot be explained by the regression model.

$$\text{Std. Error of the Estimate} = \sqrt{\frac{\textit{the}\text{ sum of squares of the residues}(SSE)}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1}}$$

The pearson correlation coefficient: r, measures the strength of the relation between two variables. The range of r is between -1 to 1. A value of 0 indicates no correlation between the variables, while a negative value of r indicates that as one variable increase, the other decreases. A positive correlation coefficient indicates that as one variable increases, so does the other variables. If r>0.7, the correlation is considered as strong correlation between to variables, if 0.7>r>0.3 it is considered as moderate correlation, if r≤0.3 it is considered as weak correlation between two variables. If 0<r<-0.3 it is considered as weak negative correlation between two variables. If -0.7<r<-0.3 it is considered as moderate negative correlation between two variables. If r<-0.7 it is considered as strong negative correlation between two variables.

3.4. Data Analysis:

Table (1):

Descriptive Statistics

	Mean	Std. Deviation	N
PetalWidthCm	1.199	.7632	150
SepalLengthCm	5.843	.8281	150
SepalWidthCm	3.054	.4336	150
PetalLengthCm	3.758	1.7653	150

In table (1) descriptive statistics table which shows the **mean** of Petal Width=1.199, **mean** of sepal length=5.843, **mean** of sepal Width=3.054, **mean** of Petal length=3.758 and **standard deviation** of petal width=0.7632, **standard deviation** of sepal length=0.8281, **standard deviation** of sepal width=0.4336, **standard deviation** of petal length=1.7653 for 150 flowers

Table (2):

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.969[a]	.938	.937	.1918

Table (2) shows that:
$R = 0.969$ close to 1, which means that the relationship between Petal Width and (Petal Length, Sepal Width, Sepal Length) is strong positive linear relationship.
$R^2 = 0.938$ close to 1, which means that, (Petal Length, Sepal Width, Sepal Length) explain approximately all the variance in Petal Width.
${R^2}_{adj} = 0.937$ close to 1, like R^2 it means that, this high Adjusted R-square value indicates a good fit of the regression model.
Std. Error of the Estimate= 0.1918, this small Std. Error of the Estimate indicates that the regression model provides a good fit to the data.

Table (3)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	81.410	3	27.137	737.761	.000[b]
	Residual	5.370	146	.037		
	Total	86.780	149			

The specified significance level= 0.05
In table (3) shows that, the sig. value is equal to 0.000 < 0.05, Therefore, we can infer that there are statistically significant differences between the Petal Width and Petal Length, Sepal Width, Sepal Length.

Table (4)

Correlations

		PetalWidthCm	SepalLengthCm	SepalWidthCm
Pearson Correlation	PetalWidthCm	1.000	.818	-.357
	SepalLengthCm	.818	1.000	-.109
	SepalWidthCm	-.357	-.109	1.000
	PetalLengthCm	.963	.872	-.420
Sig. (1-tailed)	PetalWidthCm	.	.000	.000
	SepalLengthCm	.000	.	.091
	SepalWidthCm	.000	.091	.
	PetalLengthCm	.000	.000	.000
N	PetalWidthCm	150	150	150
	SepalLengthCm	150	150	150
	SepalWidthCm	150	150	150
	PetalLengthCm	150	150	150

since significant level=0.001.
In table (4), we can see the following results:

1. pearson correlation for petal width and sepal length is 0.818 so r>0.7 which means correlation between petal width and sepal length is strong positive correlation, and correlation significant.
2. pearson correlation for petal width and sepal width is -0.357 so -0.7<r<-0.3 which means correlation between petal width and sepal width is moderate negative correlation, and correlation significant.
3. pearson correlation for sepal length and sepal width is -0.109 so 0<r<-0.3 which means correlation between sepal length and sepal width is weak negative correlation, and correlation significant.
4. pearson correlation for petal length and petal width is 0.963 so r>0.7 which means correlation between petal length and petal width is strong positivecorrelation, and correlation significant.
5. pearson correlation for petal length and sepal length is 0.872 so r>0.7 which means correlation between petal length and sepal length is strong positive correlation, and correlation significant.
6. pearson correlation for petal length and sepal width is -0.420 so -0.7<r<-0.3 which means correlation between petal length and sepal width is moderate negative correlation, and correlation significant.

Table (5)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.244	.178		-1.372	.172
	SepalLengthCm	-.211	.048	-.229	-4.442	.000
	SepalWidthCm	.229	.049	.130	4.673	.000
	PetalLengthCm	.526	.024	1.217	21.552	.000

In table(5) we can see that:

1. Constant: The constant term (intercept) is -0.244. This means that when all the predictor variables (Sepal Length, Sepal Width, and Petal Length) are zero, the predicted value of the dependent variable (Petal Width) is -0.244.
2. Sepal Length (Cm): The coefficient for Sepal Length is -0.211. This indicates that a one-unit increase in Sepal Length is associated with a decrease of -0.211 units in Petal Width, while holding other variables constant.
3. Sepal Width (Cm): The coefficient for Sepal Width is 0.229. This means that a one-unit increase in Sepal Width is associated with an increase of 0.229 units in Petal Width, while holding other variables constant.
4. Petal Length (Cm): The coefficient for Petal Length is 0.526. This indicates that a one-unit increase in Petal Length is associated with an increase of 0.526 units in Petal Width, while holding other variables constant.
5. which is show that the Estimated Regression Equation is
   $\widehat{petal\ width}$ =-0.244-0.211(sepal length) +0.229(sepal width) +0.526(petal length)
   $\hat{y}$=-0.244-0.211x_1+0.229x_2+0.526x_3

Discussion:

In this research project, the results of the correlation analysis provide valuable insights into the relationship between Petal Width and three variables, namely Petal Length, Sepal Width, and Sepal Length, across three species of iris flowers. The correlations were assessed using a sample size of 150 for each variable, and a significance level of 0.001 was used to determine statistical significance.
The correlation coefficients reveal interesting patterns:

1. Petal Width shows a strong positive correlation with Petal Length ($r = 0.963$, $p < 0.001$). This suggests that as the width of the petal increases, so does its length. This finding implies a proportional relationship between petal size dimensions within the same iris species.
2. Petal Width exhibits a strong positive correlation with Sepal Length ($r = 0.818$, $p < 0.001$). This indicates that as the width of the petal increases, the length of the sepal tends to increase as well. This finding suggests a potential relationship between the size of the petal and the overall size of the flower structure.
3. Petal Width displays a moderate negative correlation with Sepal Width ($r = -0.357$, $p < 0.001$). This implies that as the width of the petal increases, the width of the sepal tends to decrease. This observation suggests a contrasting relationship between the dimensions of the petal and sepal within the iris species.

It is worth noting that the correlation between Sepal Length and Sepal Width is not statistically significant at the 0.001 level ($r = -0.109$, $p = 0.091$). This indicates that there may not be a strong linear association between these two variables. Further research is encouraged to explore these relationships in more depth, accounting for other potential factors that may influence the observed correlations.

Conclusion:

In this research project we explained the concept of multivariate normal distribution, and we found the probability density function of multivariate normal distribution in two cases. Using moment generating function. and using theorem. Also, we investigate relationship between Petal Width and (Petal Length, Sepal Width, Sepal Length), of the three iris species. based on the data we collected of our sample. Although the data we collected was limited, our results are statistically significand and can be considered reliable. Understanding the relationships between these floral traits can have implications across various fields. For instance, in botany and plant biology, these findings contribute to our understanding of the morphological characteristics and development of irises. In horticulture and breeding programs, the identified relationships can inform selection criteria for desired floral traits. Additionally, in ecology and evolutionary biology, these correlations may provide insights into the adaptive significance and ecological functions of the measured traits within the iris species.

References:

[1] Probability and Statistics for Engineering and Sciences, Jay L. Devore,8th Edition.

[2] INTRODUCTION TO STATISTICS K.M.AL-RAWI ,1971 Baghdad

[3] SCHAUM’S OUTLINE OF THEORY AND PROBLEMS OF LINEAR ALGEBRA Second Edition

[4] Robert V. Hogg, Joseph W.McKean, Allen T. Craig, Introduction to mathematical statistics, 2013, 2005, 1995 Pearson Education, Inc. /. – 7th ed.

[5] Probability and Statistics for Engineering and Sciences, Jay L. Devore,8th Edition.

[6] https://www.kaggle.com/