TVET CERTIFICATE V

ADVANCED ANALYSIS AND STATISTICS

GENAM501

Apply Advanced analysis and statistics

Competence

Credits: 8

Learning hours: 80

Sector: All

Sub-sector: All

Module Note Issue date: June, 2020

Purpose statement

This General module gives essential knowledge in Mathematical Analysis and Statistics required to learners in order to perform very well in their engineering career.

Table of Contents

Learning Unit 1 – Apply fundamentals of integrals

LO 1.1 – Calculate the primitive functions

- Definition

Integration is the reverse process of differentiation. When we differentiate we start with an expression and proceed to find its derivative. When we want to find the primitive function (integral), we start with the derivative and then find the expression from which it has been derived.

For example, $\frac{d}{dx}(\sin x) = \cos x$. Therefore the primitive function of $\cos x$ with respect to x we know to be $\sin x$. This is written:

$$\int \cos x dx = \sin x$$

The symbols $\int f(x)dx$ denotes the **integral** of $f(x)$ with respect to the variable x, the expression $f(x)$ to be integrated is called the **integrand** and the differential dx is usefully there to assist in the evaluation of certain integrals.

Constant of integration

So $\frac{d}{dx}(x^2) = 2x \Leftrightarrow \int 2x dx = x^2$

Also $\frac{d}{dx}(x^2 + 3) = 2x \Leftrightarrow \int 2x dx = x^2 + 3$

And $\frac{d}{dx}(x^2 - 10) = 2x \Leftrightarrow \int 2x dx = x^2 - 10$

As we see in the above examples, any **constant term** in the original expression becomes zero in the derivative and all trace of it is lost.

The symbol $\int$ is an **integral sign**,

C is the **constant of integration**,

f(x) is the **integrand**.

Note that if **F(x)** is an integral of **f(x)** with respect to *x*, then **F(x) + c** is also such an integral, where ***c*** is any constant, since the derivative of any constant is zero.

- Properties of integrals

A. The derivative of the indefinite integral

$$\frac{d}{dx}\left[\int f(x)dx\right] = f(x)$$

B. Factor out constant function from integral sign

$$\int \lambda[f(x)]dx = \lambda\int f(x)dx = \lambda F(x) + C$$

C. The indefinite integral of the algebraic sum of two functions

$$\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$$

Example

$$\int [e^x + 7]dx = \int e^x dx + \int 7dx = e^x + c_1 + 7x + c_2 = e^x + 7x + c$$

Immediate primitive

01. $\int x^n dx = \frac{x^{n+1}}{n+1} + c$ $\{$ **provided** $n \neq -1$ $\}$

02. $\int \frac{1}{x} dx = \ln x + c$

03. $\int \frac{1}{x} dx = \ln x + c$

04. $\int e^x dx = e^x + c$

05. $\int e^{kx} dx = \frac{e^{kx}}{k} + c$

06. $\int a^x dx = \frac{a^x}{\ln a} + c$

07. $\int \sin x dx = -\cos x + c$

08. $\int \cos x dx = \sin x + c$

09. $\int \sec^2 x dx = \tan x + c$

10. $\int \frac{1}{x^2 + 1} dx = \arctan x + c$

11. $\int \frac{1}{x^2 - a^2} dx = \frac{1}{2}\ln\left|\frac{x-a}{x+a}\right| + c$

12. $\int \frac{1}{a^2 - x^2} dx = \ln\left|\frac{a+x}{a-x}\right| + c$

13. $\int \frac{1}{\sqrt{1-x^2}} dx = \arcsin x + c$

14. $\int \frac{1}{\sqrt{a^2 - x^2}} dx = \arcsin\left(\frac{x}{a}\right) + c$

- Techniques of integration

A. Integration by parts

We often need to integrate a product where either function is not the derivative of the other. For example, in the case of:

$\int x^2 \ln x dx$ lnx is not the derivative of x^2, the same x^2 is not the derivative of lnx. In such situation we have to find some other method of dealing with the integral. Let us establish the rule for such cases.

If u and v are functions of x, then we know that

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$$

By integrating both sides we get

$uv = \int u\frac{dv}{dx}dx + \int v\frac{du}{dx}dx$ **, and final we get**

$$\int udv = uv - \int vdu$$

For the example $\int x^2 \ln x dx$ mentioned above, we have $u = \ln x$, $du = \frac{1}{x}dx$ and $dv = x^2 dx$

so that $v = \frac{1}{3}x^3$ we omit the integration constant here because we are in the middle of evaluating the integral. The integration constant will come out eventually).

$$\int x^2 \ln x dx = \frac{x^3}{3}\ln x - \frac{1}{3}\int x^3 \frac{1}{x}dx = \frac{x^3}{3}\ln x - \frac{1}{3}\int x^2 dx = \frac{x^3}{3}\ln x - \frac{1}{3}\frac{x^3}{3} + c = \frac{x^3}{3}\left(\ln x - \frac{1}{3}\right) + c$$

Example2

$\int x \text{ Arctan} x dx$

Let u = Arctan$x$ and dv = xdx

Then $du = \frac{dx}{1+x^2}$ and $v = \int x\, dx = \frac{1}{2}x^2$

The integral becomes

$$\int x \text{ Arctan}x dx = \frac{1}{2}x^2 \text{ Arctan}x - \frac{1}{2}\int \frac{x^2 dx}{1+x^2}$$

$$\int \frac{x^2 dx}{1+x^2} = \frac{x^2+1-1}{1+x^2}dx = \int dx - \int \frac{dx}{1+x^2} = x - \text{Arctan}x$$

Therefore, $\int x \text{ Arctan}x dx = \frac{1}{2}x^2 \text{ Arctan}x dx - \frac{1}{2}x + \frac{1}{2}\text{Arctan}x + C$

$= \frac{1}{2}(x^2+1)\text{ Arctan}x - \frac{1}{2}x + C$

Notes

In order to be able to integrate by parts we have to be aware of the following priority order:

a) Inverse trigonometric functions ($\arcsin x, \arccos x, \arctan x, arc\cot anx, ...$)
b) Logarithmic function [ln(ux), log(ux),…]
c) Polynomial function (P(x))
d) Exponential function (e^{kx})
e) Trigonometric function [cos(ux), sin(ux), tan(ux), cotan(ux),…]

Remembering the priority order will save a lot of false starts.

Solved examples

1. Calculate

(a) $\int_{-1}^{0} xe^{-2x}\, dx$ (b) $\int_{1}^{2} x^2 \ln x dx$ (c) $\int e^{ex}\sin 3x dx$

(d) $\int x \text{ Arctan}x dx$ (e) $\int \text{Arcsin}x dx$ (f) $\int \frac{\text{Arccos}x dx}{\sqrt{1-x^4}}$

Solution

1. (a) Let $u = x$ and $dv = e^{-2x}dx$

Then $du = dx$ and $v = \int e^{-2x}dx = \frac{-1}{2} e^{-2x}$

The integral becomes;

$$\int_{-1}^{0} xe^{-2x}dx = \left[\frac{-1}{2} x e^{-2x}\right]_{-1}^{0} + \frac{1}{2}\int_{-1}^{0} e^{-2x}dx$$

$$= \left[-\frac{1}{2} xe^{-2x}\right]_{-1}^{0} - \frac{1}{4}\left[e^{-2x}\right]_{-1}^{0}$$

$$= \left[-\frac{1}{2} xe^{-2x} - \frac{1}{4} e^{-2x}\right]_{-1}^{0} = -\frac{1}{4} - \frac{1}{4} e^{2}$$

(b) $\int_{1}^{2} x^2 \ln x dx$

Let $u = \ln x$ and $dv = x^2dx$

Then $du = \frac{dx}{x}$ and $v = \int x^2dx = \frac{1}{3}x^3$

The integral becomes:

$$\int_{1}^{2} x^2 \ln x dx = \left[\frac{1}{3} x^3 \ln x\right]_{1}^{2} - \frac{1}{3}\int_{1}^{2} \frac{x^3dx}{x}$$

$$= \left[\frac{1}{3} x^3\ln x - \frac{1}{3}x^3\right]_{1}^{2} = \frac{8}{3}\text{In}2 - \frac{7}{3}$$

(c) $\int e^{2x}\sin 3x \, dx$

Let $u = \sin 3x$ and $dv = e^{2x}dx$

Then $du = 3\cos 3x dx$ and $v = \int e^{2}dx = \frac{1}{2}e^{2x}$

The integral becomes;

$$\int e^{2}\sin 3x dx = \frac{1}{2}e^{2x}\sin 3x - \frac{1}{2}\int 3e^{2x}\cos 3x dx$$

$$= \frac{1}{2}e^{2x}\sin 3x - \frac{3}{2}\int e^{2x}\cos 3x dx$$

Calculating $\int e^{2x}\cos 3x dx$:

Let $u = \cos 3x$ and $dv = e^{2x}dx$

Then $du = -3\sin 3x dx$ and $v = \int e^{2x}dx = \frac{1}{2}e^{2x}$

The original integral becomes;

$$\int e^{2x}\sin 3x dx = \frac{1}{2}e^{2x}\cos 3x - \frac{3}{2}\left[\frac{1}{2}e^{2x}\cos 3x + \frac{3}{2}\int e^{2x}\sin 3x dx\right]$$

$= \frac{1}{2}e^{2x}\sin 3x - \frac{3}{4}\left[\frac{1}{2}e^{2x}\cos 3x - \frac{9}{4}\right]e^{2x}\sin 3x dx$, which is equivalent to:

$$\int e^{2x}\sin 3x dx + \frac{3}{4}\int e^{2x}\sin 3x dx$$

$$= \frac{1}{2}e^{2x}\sin 3x - \frac{3}{4}e^{2x}\cos 3x \text{ ;}$$

$$\frac{13}{4}\int e^{2x}\sin 3x dx = \frac{1}{2}e^{2x}(\sin 3x - \frac{3}{2}\cos 3x)$$

Therefore, $\int e^{2x}\sin 3x dx = \frac{2}{13}e^{2x}(\sin 3x - \frac{3}{4}\cos 3x) + C$

(d) $\int x \operatorname{Arctan} x dx$

Let $u = \operatorname{Arctan} x$ and $dv = x dx$

Then $du = \frac{dx}{1+x^2}$ and $v = \int x dx = \frac{1}{2}x^2$

The integral becomes

$$\int x \operatorname{Arctan} x dx = \frac{1}{2}x^2 \operatorname{Arctan} x - \frac{1}{2}\int \frac{x^2 dx}{1+x^2}$$

$$\int \frac{x^2 dx}{1+x^2} = \frac{x^2+1-1}{1+x^2}dx = \int dx - \int \frac{dx}{1+x^2} = x - \operatorname{Arctan} x$$

Therefore, $\int x \operatorname{Arctan} x dx = \frac{1}{2}x^2 \operatorname{Arctan} x dx - \frac{1}{2}x + \frac{1}{2}\operatorname{Arctan} x + C$

$$= \frac{1}{2}(x^2+1)\operatorname{Arctan} x - \frac{1}{2}x + C$$

(e) $\int \text{Arcsin}x dx$

Let u = Arcsin$x$ and dv = dx

Then $du = \frac{dx}{\sqrt{1-x^2}}$ and v = x; $\int \text{Arcsin}x dx = x\text{Arcsin}x - \int \frac{x dx}{\sqrt{1-x^2}}$

$= \text{Arcsin}x + \sqrt{1-x^2+c}$

(f) $\int \frac{\text{Arccos}x dx}{\sqrt{1-x^2}} = (\text{Arccos}x)^2 - \int \frac{\text{Arccos}x dx}{\sqrt{1-x^2}}$

$2\int \frac{\text{Arccos}x dx}{\sqrt{1-x^2}} = -(\text{Arccos}x)^2;$

$\int \frac{\text{Arccos}x dx}{\sqrt{1-x^2}} = \frac{1}{2}(\text{Arccos}x)^2 + c$

Exercises

1. Evaluate:

(a) $\int_0^{\pi/2} e^x \cos 2x dx$ (b) $\int_{-1}^{0} \ln(1-x)dx$

2. Calculate:

(a) $\int x\cos^2 x dx$ (b) $\int \frac{x\text{Arctan}x dx}{1+x^2}$

3. Evaluate:

(a) $\int \frac{xe^2 dx}{(1+x)^2}$ (b) $\int x^2 e^x dx$

4. Evaluate:

(a) $\int (x+1)^2 e^x dx$ (b) $\int xe^{-3x} dx$

B. Integration by partial fractions

Suppose we have $\int \frac{x+1}{x^2-3x+2}dx$

Clearly this is not one of our standard types and in such case, we first of all express the algebraic fraction in terms of its partial fractions i.e. a number of simpler algebraic fractions which we shall most likely be able to integrate separately without difficulty.

$\frac{x+1}{x^2-3x+2}$ can, in fact, be expressed as $\frac{3}{x-2} - \frac{2}{x-1}$

Thus, $\int \frac{x+1}{x^2-3x+2}dx = \int \frac{3}{x-2}dx - \int \frac{2}{x-1}dx = 3\ln(x-2) - 2\ln(x-1) + c$

The rules of partial fractions

The rules of partial fractions are as follows:

i. The numerator of the given function must be of lower degree than that of the denominator. If it is not, then first of all divide out by long division

ii. Factorize the denominator into its prime factors. This is important, since the factors obtained determine the shape of the partial fractions.

iii. A linear factor $(ax+b)$ gives a partial fraction of the form $\frac{A}{ax+b}$

iv. Factors $(ax+b)^2$ give partial fractions $\frac{A}{ax+b} + \frac{B}{(ax+b)^2}$

v. Factors $(ax+b)^n$ give partial fractions $\frac{A}{ax+b} + \frac{B}{(ax+b)^2} + \frac{C}{(ax+b)^3} + \ldots\ldots \frac{K}{(ax+b)^{n-1}} + \frac{L}{(ax+b)^n}$

vi. A quadratic factor (ax^2+bx+c) gives a partial fraction $\frac{Ax+B}{ax^2+bx+c}$

Now for some examples

Example1

$\int \frac{x+1}{x^2-3x+2}dx$

$$\frac{x+1}{x^2-3x+2} = \frac{x+1}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2}$$

Multiplying both sides by the denominator we get

$x+1 = A(x-2) + B(x-1)$

Let's $(x-1)=0$ i.e substute x=1

$2 = A(-1) + B(0) \Rightarrow A = -2$

Let's $(x-2=0)$ **i.e substute x=2**

$3 = A(0) + B(1) \Rightarrow B = 3$

$$\frac{x+1}{x^2-3x+2} = \frac{x+1}{(x-1)(x-2)} = \frac{-2}{x-1} + \frac{3}{x-2}$$

$\int \frac{x+1}{x^2-3x+2}dx = \int \frac{3}{x-2}dx - \int \frac{2}{x-1}dx = 3\ln(x-2) - 2\ln(x-1) + c$ **(Do not forget the constant of integration!)**

Example2

To determine $\int \frac{x^2}{(x+1)(x-1)^2}dx$

$$\frac{x^2}{(x+1)(x-1)^2} = \frac{A}{x+1} + \frac{B}{x-1} + \frac{C}{(x-1)^2}$$

Clear the denominators: $x^2 = A(x-1)^2 + B(x+1)(x-1) + C(x+1)$

Put X=1 we get C=1/2

Put X= -1 we get A=1/4

When the substitution has come to an end, we can find the remaining constants (in this case, just

B) by equating coefficients. Choose the highest power involved, i.e X^2 in this example:

$[x^2] \quad 1 = A + B \Rightarrow B = 1 - A$

B=3/4

$$\frac{x^2}{(x+1)(x-1)^2} = \frac{1}{4}\frac{1}{x+1} + \frac{3}{4}\frac{1}{x-1} + \frac{1}{2}\frac{1}{(x-1)^2}$$

$$\int \frac{x^2}{(x+1)(x-1)^2}dx = \frac{1}{4}\int \frac{1}{x+1}dx + \frac{3}{4}\int \frac{1}{x-1}dx + \frac{1}{2}\int (x-1)^{-2}dx$$

……………………………………

$$\int \frac{x^2}{(x+1)(x-1)^2}dx \; \frac{1}{4}\ln(x+1) + \frac{3}{4}\ln(x-1) - \frac{1}{2(x-1)} + c$$

Example3

Find $\int \frac{x^2}{(x-2)(x^2+1)}dx$

In this example, we have a quadratic factor which will not factorize any further

$$\frac{x^2}{(x-2)(x^2+1)} = \frac{A}{x-2} + \frac{Bx+C}{x^2+1} \Leftrightarrow A(x^2+1) + (x-2)(Bx+C) = X^2$$

Put $(x-2) = 0$ i.e. $x = 2$

$$5A + 0 = 4 \Rightarrow A = \frac{4}{5}$$

Equate coefficients:

$[x^2]$ $A + B = 1 \Rightarrow$ B=1-A, B=1/5

$[CT]$ $A - 2C = 0 \Rightarrow$ C=A/2, C=2/5

$$\frac{x^2}{(x-2)(x^2+1)} = \frac{4}{5}\frac{1}{x-2} + \frac{\frac{x}{5}+\frac{2}{5}}{x^2+1}$$

$$= \frac{4}{5}\frac{1}{x-2} + \frac{1}{5}\frac{x}{x^2+1} + \frac{2}{5}\frac{1}{x^2+1}$$

$$\int \frac{x^2}{(x-2)(x^2+1)}dx = \frac{4}{5}\ln(x-2) + \frac{1}{10}\ln(x^2+1) + \frac{2}{5}\tan^{-1}(x) + c$$

Solved examples

Evaluate the following integrals:

(a) $\int \frac{(x+3)dx}{x^2+3x+2}$ (b) $\int \frac{x+9}{x^2-2x+1}dx$

(c) $\int \frac{xdx}{x^4+6x^2+5}$ (d) $\int \frac{dx}{(x(x^2+1)^2}$

Solutions

(a) $\frac{x+3}{x^2+3x+2} = \frac{x+3}{(x+1)(x+2)} = \frac{A}{x+1} + \frac{B}{x+2} = \frac{A(x+3)+B(x+1)}{x^2+3x+2}$;

$x+3 = A(x+2) + B(x+1)$

if $x = -1$ then $A = 2$;

If $x = -2$ then $B = -1$;

$$\frac{x+3}{x^2+3x+2} = \frac{2}{x+1} - \frac{1}{x+2}$$

$$\int \frac{(x+3)dx}{x^2+3x+2} = 2\int \frac{dx}{x+1} - \int \frac{dx}{x+2}$$

$$= 2\ln|x+1| - \ln|x+2| + C$$

$$= \ln\left|\frac{(x+1)^2}{x+2}\right| + C$$

(b) $\frac{x+9}{x^2-2x+1} = \frac{x+9}{(x+1)^2} = \frac{A}{x-1} + \frac{B}{(x-1)^2}$

(b) $\frac{x+9}{x^2-2x+1} = \frac{x+9}{(x+1)^2} = \frac{A}{x-1} + \frac{B}{(x-1)^2}$

$$= \frac{A(x-1)+B}{x^2-2x+1}$$

$x+9 = A(x-1) + B$

If $x = 1$, then $B = 10$

$x+9 = A(x-1) + 10$;

$A(x-1) = x-1$; $A = 1$

$$\frac{x+9}{x^2-2x+1} = \frac{1}{x-1} + \frac{10}{(x-1)^2}$$

$$\int \frac{x+9}{x^2-2x+1}dx = \int \frac{dx}{x-1} + 10\,\frac{dx}{(x-1)^2}$$

$$= \ln\left|x-1\right| - \frac{10}{x-1} + C$$

(c) $\frac{x}{x^4-6x^2+5} = \frac{x}{(x^2+5)(x+1)}$

$$= \frac{Ax+B}{x^2+5} + \frac{Cx+D}{x^2+1}$$

$$= \frac{(Ax+B)\,(x^2+1) + (Cx+D)(x^2+5)}{x^4-6x^2+5}$$

$$x = (Ax+B)(x^2+1) + (Cx+D)(x^2+5)$$

$$x = (A+C)x^3 + (B+D)x^2 + (A+5C)x + (B+5D)$$

$$\begin{cases} A+C=0 \\ B+D=0 \\ A+5C=1 \\ B+5D=0 \end{cases} \longleftrightarrow \begin{cases} A=\frac{1}{4} \\ B=0 \\ C=\frac{1}{4} \\ D=0 \end{cases}$$

$$\frac{x}{x^4-6x^2+5} = -\frac{1}{4}\left(\frac{x}{x^2+5}\right) + \frac{1}{4}\left(\frac{x}{x^2+1}\right)$$

$$\int \frac{xdx}{x^4+6x^2+5} = -\frac{1}{4}\int \frac{xdx}{x^2+5} + \frac{1}{4}\int \frac{xdx}{x^2+1}$$

$$= -\frac{1}{8}\ln\,(x^2+5) + \frac{1}{8}\ln\,(x^2+1) + C$$

$$= \frac{1}{8}\ln\left(\frac{x^2+1}{x^2+5}\right) + C$$

(d) $\frac{1}{x(x^2+1)^2} = \frac{A}{x} + \frac{Bx+C}{x^2+1} + \frac{Dx+E}{(x^2+1)^2}$

$$\frac{A(x^2+1)^2\,(Bx+c)x + (^2)(x^2+1) + (Dx+E)x}{x(x^2+1)^2}$$

$$1 = A(x^2+1)^2 + (Bx+C)x\,(x^2+1) + (Dx+E)x$$

If $x=0$ then $A=1$

$$1 = (x^2+1)^2 + (Bx+C)x\,(x^2+1) + (Dx+E)x$$

$$x(Bx+C)(x^2+1)+x(Dx+E)=-x^2(x^2+2);$$
$$(Bx+C)(x^2+1)+(Dx+E)=-x(x^2+2)$$
$$Bx^3+Cx^2+(B+D)x+(C+E)=-x^3-2x$$

$$\begin{cases} B=-1 \\ C=0 \\ B+D=-2 \\ C+E=0 \end{cases} \Longleftrightarrow \begin{cases} A=1 \\ B=-1 \\ C=0 \\ D=-1 \\ E=0 \end{cases}$$

$$\frac{1}{x(x^2+1)^2}=\frac{1}{x}-\frac{x}{x^2+1}-\frac{1}{(x^2+1)^2}$$

$$\int\frac{dx}{x(x^2+1)^2}=\int\frac{dx}{x}-\int\frac{xdx}{x^2+1}-\int\frac{dx}{(x^2+1)^2}$$
$$=\ln|x|-\frac{1}{2}\ln(x^2+1)+\frac{1}{2(x^2+1)}+C.$$

Exercises

Find the antiderivative (integral) for the following:

1. $\int\frac{dx}{(x-4)^3}$
2. $\int\frac{dx}{x^2-3x+2}$
3. $\frac{(4x-2)dx}{x^3-x^2-2x}$
4. $\int\frac{x^2dx}{(x-1)^3}$
5. $\frac{(4x^2+6)dx}{x^3-3x}$

C. Integration by change of variables (substitution)

Formula

$\int f[g(x)]g(x)'dx=\int f(u)du$ **Where** $u=g(x)$

Strategy: The idea is to make the integral easier to compute by doing a change of variables.

1. Start by guessing what the appropriate change of variable u = g(x) should be. Usually you choose u to be the function that is \inside" the function.

2. Differentiate both sides of u = g(x) to conclude du = g'(x)dx.

3. Rewrite the integral by replacing all instances of x with the new variable and compute the integral.

4. If you computed the indefinite integral, then make sure to write your final answer back in terms of the original variables.

Example1

Find $\int \tan x dx$

Solution

Step1: we use the change of variable $u = \cos x$

$$\frac{du}{dx} = -\sin x \Rightarrow du = -\sin x dx$$

Step2: we can evaluate the integral under this change of variable

$$\int \tan x dx = \int \frac{\sin x}{\cos x} dx = -\int \frac{du}{u} = -\ln u + c = -\ln(\cos x) + c$$

More examples

Calculate:

(a) $\int \frac{e^x dx}{e^x + 3}$ (b) $\int_0^{\pi/2} \frac{\sin 6x \, dx}{1 + \cos^2 3x}$

(b) $\int \frac{1}{x^2} \sin \frac{1}{x} dx$ (c) $\int_0^1 xe^{x^2} dx$

Solution

(a) $\int \frac{e^x dx}{e^x + 3} =$

Let $e^x + 3 = t$. Then $e^x dx = dt$

The integral becomes $\int \frac{dt}{t} = \text{In}|t| + c$

Therefore, $\int \frac{e^x dx}{ex + 3} = \ln |e^x + 3| + c$

(b) Let $t = 1 + \cos^2 3x$.

Then $dt = -3\sin 6x dx$; $\sin 6 x dx = -\frac{1}{3} dt$

x	t
0	2
1	1

The integral becomes $-\frac{1}{3}\int_2^1 \frac{dt}{t} = -\frac{1}{3}[\ln|t|]_2^1 = \frac{1}{3}\ln 2$

Therefore, $\int_0^{\pi/2} \frac{\sin 6 x dx}{1 + \cos^2 3x} = \frac{1}{3}\ln 2$

(c) Let $t = \frac{1}{x}$, then $dt = -\frac{1}{x^2}dx$

The integral becomes $\int \frac{1}{x^2} \sin\frac{1}{x} dx = -\int \sin dt = \cos t + c$

Therefore, $\int \frac{1}{x^2} \sin\frac{1}{x} dx = \cos\frac{1}{x} + c$

(d) Let $t = x^2$, then $dt = 2xdx$; $xdx = \frac{1}{2}dt$

x	t
0	0
1	1

The integral becomes $\frac{1}{2}\int_0^1 e^t dt = \frac{1}{2}\left[e^t\right]_0^1 = \frac{1}{2}(e-1)$.

Exercises

Calculate:

1. $\int \sin x \ln(\cos x) dx$
2. $\int \frac{dx}{e^x + e^{-x}}$
3. $\int \frac{\ln x dx}{x(1-\ln x)}$
4. $\int \frac{xdx}{\sqrt{1-x^4}}$

LO 1.2 – Calculate definite integrals

- Definition

An integral with limits is called a definite integral.

With a definite integral, the constant of integration may be omitted, not because it is not there but because it occurs in both brackets and disappears in subsequent working.

So, to evaluate a definite integral:

a) Integrate the function (omitting the constant of integration) and enclose within square brackets with the limits at the right-hand end
b) Substitute the upper limit
c) Substitute the lower limit
d) Subtract the second result from the first result.

Formula

$$\int_a^b y\,dx = [F(x)]_a^b = F(b) - F(a)$$

Properties of definite integrals

Let f and g be continuous functions on $I = [a,b]$.

The following properties hold:

1. $\int_a^b f(x)dx = 0$
2. $\int_a^b f(x)dx = -\int_b^a f(x)dx$
3. $\int_a^b (f+g)\,(x)dx = \int_a^b f(x)dx + \int_a^b g(x)dx$
4. $\int_a^b kf(x)dx = k\int_a^b f(x)dx;$
5. $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$

- Methods of integration

A. **Integration of definite integrals by change of variable**

Example1

Evaluate $\int_0^{\frac{\pi}{2}} \frac{\sin 2x}{1+\cos^2 x}dx$ **by change of variables**

Solution

$u = 1+\cos^2 x \Rightarrow du = -2\sin x\cos x dx$

$\sin 2x dx = -du$

$$\int_0^{\frac{\pi}{2}} \frac{\sin 2x}{1+\cos^2 x}dx = \int \frac{-du}{u} = -\ln u + c = \left[-\ln(1+\cos^2 x)\right]_0^{\frac{\pi}{2}} = \left[-\ln(1+0)\right] - \left[-\ln(1+1\right] = \ln 2$$

B. **Integration of definite integrals by decomposition**

Example

$$\int_2^3 \frac{x^3-2x^2-4x-4}{x^2+x-2}dx$$

$$\frac{x^3-2x^2-4x-4}{x^2+x-2} = (x-3) + \frac{4}{x+2} -$$

$$\int_2^3 \frac{x^3-2x^2-4x-4}{x^2+x-2}dx = \int_2^3 \left[(x-3) + \frac{4}{x+2} - \frac{3}{x-1}\right]dx$$

$$= \left[\frac{x^2}{2} - 3x + 4\ln(x+2) - 3\ln(x-1)\right]_2^3 = \left[\frac{9}{2} - 9 + 4\ln 5 - 3\ln 2\right] - [2 - 6 + 4\ln 4 - 3\ln 1] = \text{-}1.687$$

C. **Integration of definite integrals by parts**

Evaluate $\int_1^2 xe^x dx = xe^x - \int e^x dx = xe^x - e^x + c$

$$\int_1^2 xe^x dx = \left[e^x(x-1)\right]_1^2 = e^2 - 0 = e^2$$

Exercises

Evaluate:

(a) $\int_0^1 (x^2 - 2x + 3)dx$ (b) $\int_0^2 (4x+1)^{\frac{1}{2}}dx$

(c) $\int_{-3}^{-2} t(t+1)^2 dt$ (d) $\int_1^3 (2\theta+1)(3-\theta)d\theta$

(e) $\int_a^{2a} (a+z)dz$ (f) $\int_1^8 (u^{\frac{1}{3}} - u^{\frac{1}{3}})du$

LO 1.3 – Apply definite integrals

- Calculation of area

The area under a curve between two points can be found by doing a definite integral between the two points. To find the area under the curve y = f(x) between x = a and x = b, integrate y = f(x) between the limits of a and b. Areas under the x-axis will come out negative and areas above the x-axis will be positive.

Consider the function $y = f(x)$ which is continuous, on interval $[a,b]$.

The area of the region enclosed between the curve $y = f(x)$, x-axis (y=0) and the vertical lines x=a, x=b is given by :

$Area = \int_a^b ydx$ if y>0 between a and b

$Area = -\int_a^b ydx$ if y<0 between a and b

Example1

Find the area of the region enclosed between curve $y = x^3 + 3x^2$, the x-axis and the line $x = 0$ and $x = 2$

Solution

x	$-\infty$		-3		0		$+\infty$
$x^2(x+3)$	$-$	$-$	0	$+$	0	$+$	$+$

Between $x = 0$ and $x = 2$, $y > 0$

The area is $\int_0^2 (x^3 + 3x^2)dx = \left[\frac{x^4}{4} + x^3\right]_0^2 = (4 + 8) - 0 = 12$

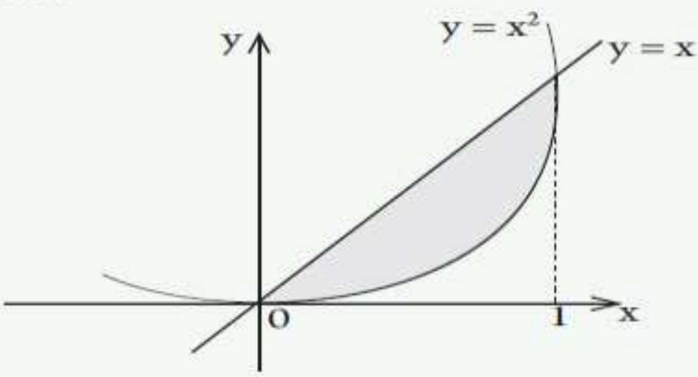
Area between two curves

Suppose that the area to be evaluated is between the curves $y_1 = f(x)$ and $y_2 = g(x)$ and the lines x=a and **x=b ,** $f(x) \le g(x)$**,** $a \le x \le b$ then the area is $A = \int_a^b [g(x) - f(x)]dx$

Example

Sketch the parabola $y = x^2$ and the line $y = x$ on the same figure and calculate the area of the region enclosed between them.

Solution



Finding the points of intersection:

$$\begin{cases} y = x^2 \\ y = x \end{cases}$$

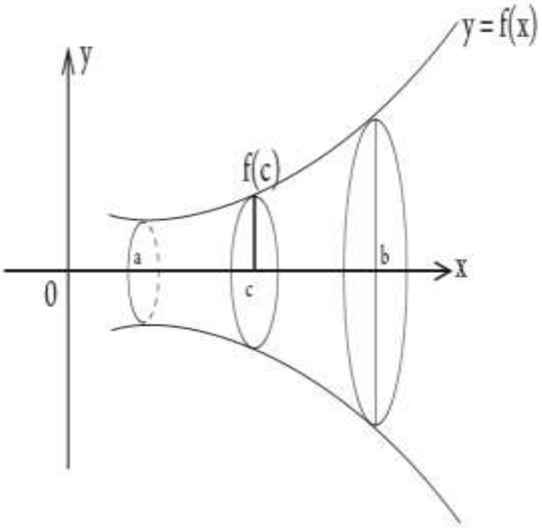
$y = x = x^2;$

$x(1 - x) = 0$; $x = 0$ or $x = 1$

$$\text{Area} = \int_0^1 (x - x^2)dx = \left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^1 = \frac{1}{6} \text{ square units}$$

- Calculation of volume

Consider the function $y = f(x)$ which is continuous, positive and strictly increasing over interval $[a,b]$. The region which is bounded by the curve $y = f(x)$, the x-axis and the ordinates x=a and x=b is revolved about the x-axis to form a solid of revolution.



x

Finally, the volume of the solid of revolution about the x-axis calculated from x = a to x = b is given by $V = \int_a^b \pi y^2 dx$ where y=f(x)

in the same way, the volume generated by rotating the region between curve $x = g(y)$, the y-axis and the horizontal lines, y=c, y=d about y-axis is

$$V = \int_a^b \pi x^2 dy$$ where x=g(y)

Example

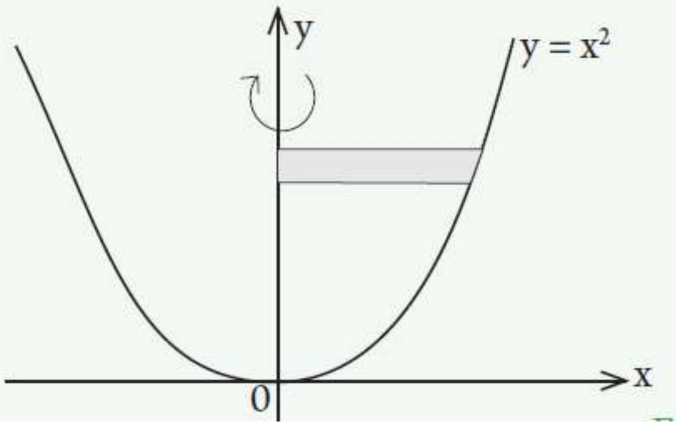
1. Find the volume of the solid generated by rotating about the y-axis the area in the first quadrant enclosed by $y = x^2$, $y = 1$, $y = 4$ and the y-axis.

Solution

Volume

$$V = \int_1^4 \pi x^2 dy$$

$$= \int_1^4 \pi y dy = \frac{\pi}{2}\left[y^2\right]_1^4 = \frac{15\pi}{2}$$



- Calculation of the length of curved surface

Let the function given by y = f(x) represent a smooth curve on the interval [a,b]. The length of f between a and b is

$$S = \int_a^b \sqrt{1 + \left(\frac{df}{dx}\right)^2} dx$$

Example

Find the arc length of the graph of $y = \frac{x^3}{6} + \frac{1}{2x}$ **on the interval** $\left[\frac{1}{2}, 2\right]$**.**

Solution

$$\frac{dy}{dx} = \frac{3x^2}{6} - \frac{1}{2x^2}$$

$$= \frac{3x^2}{6}\left(x^2 - \frac{1}{x^2}\right)$$

Arc length: $S = \int_{\frac{-1}{2}}^{1} \sqrt{1 + \left[\frac{1}{2}\left(x^2 - \frac{1}{x^2}\right)\right]^2} \quad dx$

$$= \int_{\frac{1}{2}}^{1} \sqrt{\frac{1}{4}\left(x^2 + 2 + \frac{1}{x^2}\right)}\, dx$$

$$= \int_{\frac{1}{2}}^{1} \frac{1}{2}\left(x^2 + \frac{1}{x^2}\right) dx$$

$$= \left[\frac{1}{2}\left(\frac{x^3}{3} - \frac{1}{x}\right)\right]_{\frac{1}{2}}^{2}$$

$$= \frac{1}{2}\left(\frac{13}{6} + \frac{47}{24}\right)$$

$$= \frac{33}{16}.$$

Exercises

Question1

Graph the parabolas $y = 4 - x^2$ and $y = x^2 - 2x$ on the same figure and calculate the area of the region enclosed between them.

Question2

Calculate the area of the region enclosed between curves $y = 15 - 2x - 2x^2$ and $y = 25 - 5x - 2x^2$

Question3

The area of the segment cut off by $y = 5$ from the curve $y = x^2 + 1$ is rotated about $y = 5$.

Find the volume generated.

Learning Unit 2 – Identify measures of dispersion and interpret bivariate data

LO 2.1 – Identify the measures of dispersion

- Definition

A measure of dispersion is the degrees of spread of observation in data. The common measures of dispersion are Range, Inter-quartiles range, the Standard deviation, and the Variance.

- Range

The range is defined as the difference between the largest value in the set of data and the smallest value in the set of data, XL – XS

Example

What is the range of the following data?

4 8 1 6 6 2 9 3 6 9

Solution

The largest score (X_L) is 9; the smallest score (X_S) is 1; the range is $X_L - X_S = 9 - 1 = 8$

- The standard deviation and the variance

The standard deviation σ is a very important and useful measure of spread. It gives a measure of the deviations of the readings from the mean $\overline{X}$. It is calculated using all the values in the distribution.

To calculate standard deviation σ:

(i) For each reading $\frac{(X-\overline{X})^2}{n}$ of x calculate $X - \overline{X}$, its deviation from the mean

(ii) Square this deviation to give $(X-\overline{X})^2$ and note that, irrespective of whether the deviation was positive or negative, this is now positive

(iii) Find $\Sigma (X-\overline{X})^2$ the sum of all these values,

(iv) Find the average by dividing the sum by n, the number of readings; this gives $\Sigma \frac{(X-\overline{X})^2}{n}$ which is known as **variance**

(v) Finally, take the positive square root of the variance to obtain the standard Deviation σ

Formula for variance $\sigma^2 = \Sigma \frac{(X - \overline{X})^2}{n}$

Formula for standard diviation $\sigma = \sqrt{\sum \frac{(x - \overline{x})^2}{n}}$

Example

Two machines A and B are used to pack biscuits. A random sample of ten packe was taken from each machine and the same mass of each packet was measur to the nearest gram and noted. Find the standard deviations of the masses the packets taken in the sample from each machine. Comment on your answ

Machine A (mass in g)	196, 198, 198, 199, 200, 200, 201, 201, 202, 20
Machine B (mass in g)	192, 194, 195, 198, 200, 201, 203, 204, 206, 20

Solution

Machine A: $\overline{x} = \frac{\Sigma x}{n} = \frac{2000}{10} = 200.$ Machine B: $\overline{x} = \frac{\Sigma x}{n} = \frac{2000}{10} = 200$

Since the mean mass for each machine is 200, $x - \overline{x} = x - 200$

To calculate standard deviation, s; put the data in a table:

Machine A		
x	x – 200	$(x - 200)^2$
196	–4	16
198	–2	4
198	–2	4
199	–1	1
200	0	0
200	0	0
201	1	1
201	1	1
202	2	4
205	5	25
		56

$s^2 = \frac{\Sigma(x - 200)^2}{10} = 5.6$

$s = \sqrt{5.6} = 2.37$

Machine A: s.d = 2.37g

Machine B		
x	x – 200	$(x-200)^2$
192	–8	64
194	–6	36
195	–5	25
198	–2	4
200	0	0
201	1	1
203	3	9
204	4	16
206	6	36
207	7	49
		240

$s^2 = \frac{\sum(x-200)^2}{10} = 24$

$s = \sqrt{24} = 4.980$ g

Machine B: s.d = 4.90 g

Machine A has less variation, indicating that it is more reliable than machine B.

<u>Alternative form of the formula for standard deviation</u>

The formula given above is sometimes difficult to use especially when x is not an integer; so an alternative form is often used. This is derived as follows:

$$s^2 = \frac{1}{n}\sum(x-\overline{x})^2 = \frac{1}{n}\sum(x^2 - 2\overline{x}\,x + \overline{x}^2) = \frac{1}{n}(\sum x^2 - 2\overline{x}\sum x + \sum \overline{x}^2)$$

since, $\frac{\sum x}{n} = \overline{x}$

$$s^2 = \frac{\sum x^2}{n} - 2\overline{x}\frac{\sum x}{n} + \frac{n\overline{x}^2}{n} = \frac{\sum x^2}{n} - 2\overline{x}\,(\overline{x}) + \overline{x}^2 = \frac{\sum x^2}{n} - 2\overline{x}^2 + \overline{x}^2$$

$$s^2 = \frac{\sum x^2}{n} - \overline{x}^2$$

$$s = \sqrt{\frac{\sum x^2}{n} - \overline{x}^2}$$

- Coefficient of variation

The coefficient of variation (CV) is defined as the ratio of the standard deviation σ to the mean μ.

$$C_v = \frac{\sigma}{\mu}$$

It shows the extent of variability in relation to the mean of the population. The coefficient of variation should be computed only for data measured on a ratio scale, as these are the measurements that can only take non-negative values.

The coefficient of variation (C.V) unlike the previous measures we have studied is a relative measure of dispersion. It is expressed as a percentage rather than in terms of the unit of the particular data. It is useful when comparing the variable of two or more batches of data. Those are expressed in different units of measurement.

$C.V = \frac{\delta}{\bar{x}} \times 100$ where δ is the standard deviation and $\bar{x}$ is the mean.

For example, given that $\delta = 6.26$ and $\bar{x} = 20$ then $CV = \frac{6.26}{20} \times 100 = 31.3\%$

That is for this sample the relative size of the average spread around the mean is 31.3%. The C.V is also very useful when comparing two or more sets of data which are measured in the same units.

Lo 2.2:Describe the measures of bivariate data

- Definition of bivariate data

When two variables are studied simultaneously to see how they are inter-related, the two sets of data to compare are said to be a **"bivariate"** or "**joint" distribution**. The quantities being studied are called "**Variables**" thus in a joint distribution or in a bivariate data we have two variables studied jointly.

Example

A doctor recording the **height** and the **weight** of a group of female patients in a hospital.

- Correlation

The word correlation is used in everyday life to denote some form of association. However, in statistical terms, we use **correlation** to denote association between two quantitative variables.

- Covariance

In probability theory and statistics, **covariance** is a measure of the joint variability of two random variables.

- **Positive covariance**: Indicates that two variables tend to move in the same direction.
- **Negative covariance**: Reveals that two variables tend to move in inverse directions.

The sign of the covariance therefore shows the tendency in the linear relationship between the variables.

Calculation of covariance

For n pairs of data $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, the covariance is denoted by Cov (x, y) or s_{xy} or σ_{xy} is given by

$$s_{xy} = \frac{1}{n}\Sigma\,(x - \bar{x})\,(y - \bar{y}).$$

The formula for covariance is simply an extension of the formula for variance, where

$$s_{xx} = \frac{1}{n}\Sigma\,(x - \bar{x})\,(x - \bar{x})$$

Final formula $S_{xy} = \frac{1}{n}\sum\left(x - \bar{x}\right)\left(x - \bar{x}\right)$

An alternative form of the formula for covariance can be

$$s_{xy} = \left(\frac{1}{n}\,\Sigma\,xy\right) - \bar{x}\,\bar{y}.$$

Example

For the following data

x	9	12	5	6	9	14	3	6	12	14
y	10	13	8	10	13	17	5	8	16	16

Calculate the covariance of the distribution.

Solution

$$s_{xy} = (\frac{1}{n} \Sigma xy) - \bar{x}\,\bar{y}$$

The means are $\bar{x} = \frac{9 + 12 + 5 + 6 + 9 + 14 + 3 + 6 + 12 + 14 = 9}{10} = 9$

and $\bar{y} = \frac{10 + 13 + 8 + 10 + 13 + 17 + 5 + 8 + 16 + 16}{10} = 11.6.$

x	y	xy
9	10	90
12	13	156
5	8	40
9	10	60
14	17	238
3	5	15
12	16	192
14	16	224
$\Sigma x = 90$ $\bar{x} = \frac{1}{n}\Sigma x$ $= \frac{90}{9}$ $= 9$	$\Sigma y = 116$ $\bar{y} = 11.6$	$\Sigma xy = 1180$ $\frac{1}{n}\Sigma xy$ $^{n}= 118$

The covariance is $s_{xy} = (\frac{1}{n} \Sigma xy) - \bar{x}\,\bar{y} = 118 - (9)\,(11.6) = 13\,.\,6.$

- Coefficient of correlation

The linear correlation coefficient, r, is the numerical value which indicates the degree of scatter.

For the bivariate data.

x_1	x_2	...	x_n
y_1	y_2	...	y_n

If $s_x^2 \neq 0$ and $s_y^2 \neq 0$, then the linear correlation coefficient is given by

$r = \frac{s_{xy}}{s_x s_y}$, where $s_{xy} = (\frac{1}{n}\Sigma xy) - \bar{x}\,\bar{y}$; $s_x^2 = (\frac{1}{n}\Sigma x^2) - \bar{x}^2$ and $s_y^2 = (\frac{1}{n}\Sigma y^2) - \bar{y}^2$.

Finally, $r = \frac{S_{XY}}{S_X S_Y}$

Example

For a sample of 100 paired items, the following data are known:

$\Sigma x = 36, \Sigma y = 25; \Sigma xy = 21; \Sigma x^2 = 1012.96$; and $\Sigma y^2 = 3366.25$.

Calculate the correlation coefficient between x and y:

Solution

$$\bar{x} = \frac{1}{n}\Sigma x = \frac{36}{100} = 0.36$$

$$\bar{y} = \frac{1}{n}\Sigma y = \frac{25}{100} = 0.25$$

$$s_{xy} = (\frac{1}{n}\Sigma xy) - \bar{x}\bar{y}$$

$$= \frac{21}{100} - (0.36)(0.25)$$

$$= 0.21 - 0.09 = 0.12$$

$$s_x^2 = (\frac{1}{n}\Sigma x^2) - \bar{x}^2 = 10.1296 - 0.1296 = 10$$

$$s_x = \sqrt{10}$$

$$s_y^2 = (\frac{1}{n}\Sigma y^2) - \bar{y}^2$$

$$= \frac{366.25}{100} - (0.25)^2$$

$$= 3.6625 - 0.0625 = 3.6$$

$$s_y = \sqrt{3.6}.$$

The correlation coefficient is $\Omega = \frac{sxy}{sx\,sy} = \frac{0.12}{\sqrt{10}\,\sqrt{3.6}} = \frac{0.12}{6} = 0.02$.

LO 2.3 – Determine the regression line

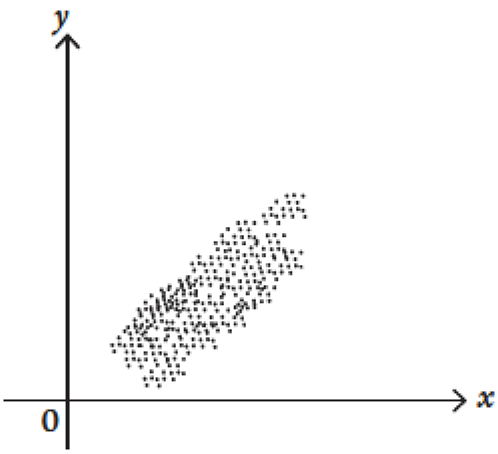
- Scatter diagram

A bivariate raw data is made of ordered pairs of values of the variables, recorded as they occurred.

x_1	x_2	x_3	...	x_n
y_1	y_2	y_3	...	y_n

If the variables are plotted in the Cartesian plane, the result is a "**scatter diagram**".

A scatter diagram can be as shown below:



Example

For the data below

Maths (x)	9	12	5	6	9	14	3	6	12	14
Physics (y)	10	13	8	10	13	17	5	8	16	16

a) Find the mean marks in:

i) Mathematics ii) Physics

b) Draw the scatter diagram for the bivariate distribution and plot $(\bar{x}, \bar{y})$:

Solution

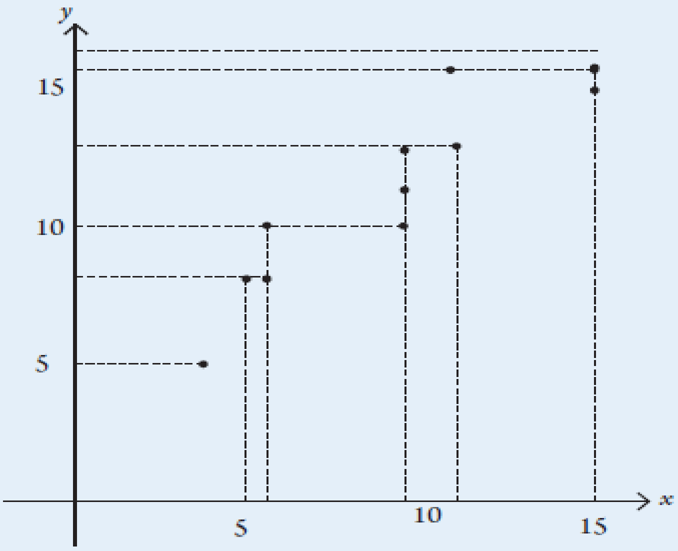
a) $\bar{x} = \frac{1}{n}\sum x = \frac{9+12+5+6+9+14+3+6+12+14}{10} = 9$

$\bar{y} = \frac{1}{n}\sum y = \frac{10+13+8+10+13+17+5+8+16+16}{10}$

$= 11.6$

The mean mark in mathematics is 9 and the mean mark in Physcis is 11.6.

b) Plotting the points (9, 10); (12, 13); (5, 8); (6, 10); (9, 13): (14, 17); (3, 5); (6, 8); (12, 16); (14, 16) and (9, 11.6) we have.



- Regression lines

The scatter diagram for any real set of bivariate data does not have points lying exactly on a straight line.

If there is a linear correlation between the two variables *x* and *y* then it is possible to find the best straight lines that can be fitted to the points. There are two such lines.

– The regression line of *y* on *x*

– The regression line of *x* on *y*

- Calculation and plotting of regression lines

  A. Calculation of the coefficient of regression line of y on x

The equation of the regression line of ***y*** on ***x*** is

$$y - \bar{y} = a(x - \bar{x})$$

Where $a = \frac{\sum(x-\bar{x})(y-\bar{y})}{S_x^{\ 2}}$ is the coefficient of regression line.

B. Calculation of the regression line y on x

The equation of the regression line y on x is given by

$Y - \overline{Y} = a(X - \overline{X})$ Where,

$a = \frac{s_{xy}}{s_x^2}; \overline{x} = \frac{1}{n}\sum x$ and $\overline{y} = \frac{1}{n}\sum y.$

Example1

For the following data

x	0	1	2	3	4
y	2	3	5	4	6

Determine the equation of the regression line of y on x:

Solution

$$\overline{x} = \frac{1}{n}\sum x = 2$$

$$\overline{y} = \frac{1}{n}\sum y = 4$$

$$s_{xy} = \left(\frac{1}{n}\sum xy\right) - \overline{x}\,\overline{y} = 1.8$$

$$s_x^2 = \left(\frac{1}{n}\sum x^2\right) - \overline{x}^2$$

$$= 2$$

$$\frac{s_{xy}}{s_x^2} = \frac{1.8}{2} = 0.9.$$

The equation of the regression line of y on x is

$y - 4 = 0.9\,(x - 2).$

That is $y = 0.9x + 2.2.$

C. Calculation of the coefficient of regression line of x on y

The equation of regression line x on y is given by:

$(X - \overline{X}) = c(Y - \overline{Y})$

Where $c = \frac{S_{XY}}{S_Y^2}$ is the coefficient of the regression line x on y

$, \overline{x} = \frac{1}{n}\sum x, \overline{y} = \frac{1}{n}\sum y.$

Example 2

For a given set of data $\Sigma x = 21$, $\Sigma y = 33$, $\Sigma x^2 = 91$, $\Sigma y^2 = 205$, $\Sigma xy = 128$, $n = 6$.

Find the equation of the regression line of x on y:

Solution

$$\overline{x} = \frac{1}{n} \Sigma x = 3.5$$

$$\overline{y} = \frac{1}{n} \Sigma y = 5.5$$

$$s_{xy} = \left(\frac{1}{n} \Sigma xy \right) - \overline{x}\overline{y}$$

$$= 2.08$$

$$s_y^2 = \left(\frac{1}{n} \Sigma y^2 \right) - \overline{y}^2$$

$$= 3.9$$

The equation $x - \overline{x} = c(y - \overline{y})$

$$= x - 3.5 = \frac{2.08}{3.91} (y - 5.5), \text{ that is}$$

that is $x = 0.53y + 0.58$.

SOLVED EXERCISE ON SCATTER DIAGRAM AND REGRESSION LINE

Table 10.3 "Data on Age and Value of Used Automobiles of a Specific Make and Model" shows the age in years and the retail value in thousands of dollars of a random sample of ten automobiles of the same make and model.

a. Construct the scatter diagram.
b. Compute the linear correlation coefficient r. Interpret its value in the context of the problem.
c. Compute the least squares regression line. Plot it on the scatter diagram.
d. Interpret the meaning of the slope of the least squares regression line in the context of the problem.
e. Suppose a four-year-old automobile of this make and model is selected at random. Use the regression equation to predict its retail value.

TABLE 10.3 DATA ON AGE AND VALUE OF USED AUTOMOBILES OF A SPECIFIC MAKE AND MODEL

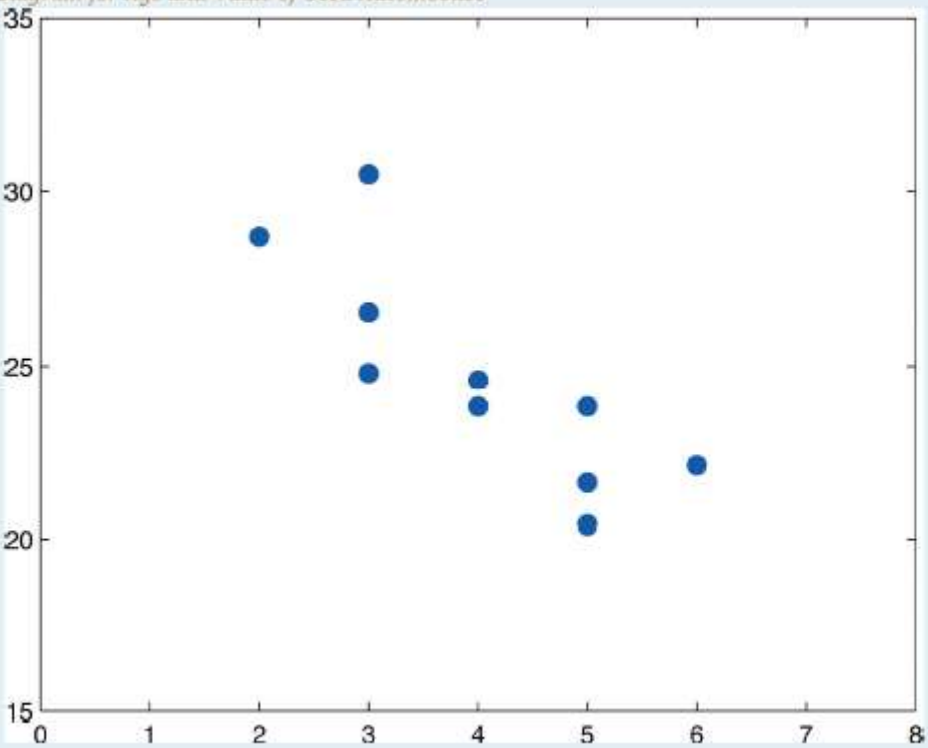
x	2	3	3	3	4	4	5	5	5	6
y	28.7	24.8	26.0	30.5	23.8	24.6	23.8	20.4	21.6	22.1

Solution:

a. The scatter diagram is shown in Figure 10.7 "Scatter Diagram for Age and Value of Used Automobiles".

Figure 10.7

Scatter Diagram for Age and Value of Used Automobiles



a. We must first compute SS_{xx}, SS_{xy}, SS_{yy}, which means computing Σx, Σy, Σx^2, Σy^2, and Σxy. Using a computing device we obtain

$$\Sigma x = 40 \quad \Sigma y = 246.3 \quad \Sigma x^2 = 174 \quad \Sigma y^2 = 6154.15 \quad \Sigma xy = 956.5$$

$$SS_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 174 - \frac{1}{10}(40)^2 = 14$$

$$SS_{xy} = \Sigma xy - \frac{1}{n}(\Sigma x)(\Sigma y) = 956.5 - \frac{1}{10}(40)(246.3) = -28.7$$

$$SS_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 6154.15 - \frac{1}{10}(246.3)^2 = 87.781$$

so that

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \cdot SS_{yy}}} = \frac{-28.7}{\sqrt{(14)(87.781)}} = -0.819$$

The age and value of this make and model automobile are moderately strongly negatively correlated. As the age increases, the value of the automobile tends to decrease.

b. Using the values of Σx and Σy computed in part (b),

$$\bar{x} = \frac{\Sigma x}{n} = \frac{40}{10} = 4 \quad \text{and} \quad \bar{y} = \frac{\Sigma y}{n} = \frac{246.3}{10} = 24.63$$

Thus using the values of SS_{xx} and SS_{xy} from part (b),

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-28.7}{14} = -2.05 \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 24.63 -$$

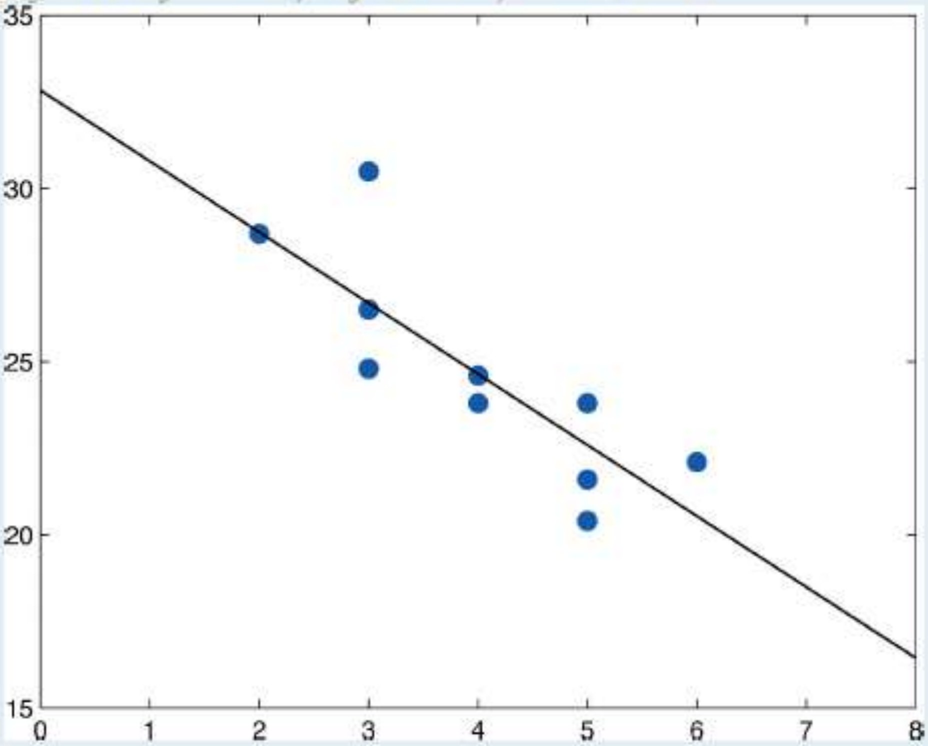
The equation $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0$ of the least squares regression line for these sample data is

$$\hat{y} = -2.05x + 32.83$$

Figure 10.8 "Scatter Diagram and Regression Line for Age and Value of Used Automobiles" shows the scatter diagram with the graph of the least squares regression line superimposed.

Figure 10.8

Scatter Diagram and Regression Line for Age and Value of Used Automobiles



a. The slope −2.05 means that for each unit increase in x (additional year of age) the average value of this make and model vehicle decreases by about 2.05 units (about $2,050).

b. Since we know nothing about the automobile other than its age, we assume that it is of about average value and use the average value of all four-year-old vehicles of this make and model as our estimate. The average value is simply the value of $\hat{y}$ obtained when the number 4 is inserted for x in the least squares regression equation:

$$\hat{y} = -2.05\,(4) + 32.83 = 24.63$$

which corresponds to $24,630.

The linear regression line of y on x and the linear regression line of x on y can be obtained using the equations known as **normal equations.**

Here are the normal equations y on x

$$\left(\sum_{i=1}^{n} x_i^2\right)a + \left(\sum_{i=1}^{n} x_i\right)b = \sum_{i=1}^{n} x_i y_i$$

$$\left(\sum_{i=1}^{n} x_i\right)a + nb = \sum_{i=1}^{n} y_i$$

Here are the normal equations x on y

$$\begin{cases} \left(\sum_{i=1}^{n} y_i\right)c + nd = \sum_{i=1}^{n} x_i \\ \left(\sum_{i=1}^{n} y_i^2\right)c + \left(\sum_{i=1}^{n} y_i\right)d = \sum_{i=1}^{n} x_i y_i \end{cases}$$

or simply $\begin{cases} (\Sigma y)c + nd = \Sigma x \\ (\Sigma y^2)c + (\Sigma y)d = \Sigma xy \end{cases}$.

Finding the regression line using normal equations

Worked example

Problem 1. In an experiment to determine the relationship between frequency and the inductive reactance of an electrical circuit, the following results were obtained:

Frequency (Hz)	Inductive reactance (ohms)
50	30
100	65
150	90
200	130
250	150
300	190
350	200

Determine the equation of the regression line of inductive reactance on frequency, assuming a linear relationship.

Since the regression line of inductive reactance on frequency is required, the frequency is the independent variable, X, and the inductive reactance is the dependent variable, Y. The equation of the regression line of Y on X is:

$$Y = a_0 + a_1 X$$

and the regression coefficients a_0 and a_1 are obtained by using the normal equations

$$\sum Y = a_0 N + a_1 \sum X$$

and $$\sum XY = a_0 \sum X + a_1 \sum X^2$$

(from equations (1) and (2))

A tabular approach is used to determine the summed quantities.

Frequency, X	Inductive reactance, Y	X^2
50	30	2500
100	65	10000
150	90	22500
200	130	40000
250	150	62500
300	190	90000
350	200	122500
$\sum X = 1400$	$\sum Y = 855$	$\sum X^2 = 350000$

XY	Y^2
1500	900
6500	4225
13500	8100
26000	16900
37500	22500
57000	36100
70000	40000
$\sum XY = 212000$	$\sum Y^2 = 128725$

The number of co-ordinate values given, N is 7. Substituting in the normal equations gives:

$$855 = 7a_0 + 1400a_1 \qquad (1)$$

$$212000 = 1400a_0 + 350000a_1 \qquad (2)$$

$1400 \times (1)$ gives:

$$1197000 = 9800a_0 + 1960000a_1 \qquad (3)$$

$7 \times (2)$ gives:

$$1484000 = 9800a_0 + 2450000a_1 \qquad (4)$$

$(4) - (3)$ gives:

$$287000 = 0 + 490000a_1$$

from which, $a_1 = \dfrac{287000}{490000} = 0.586$

Substituting $a_1 = 0.586$ in equation (1) gives:

$$855 = 7a_0 + 1400(0.586)$$

i.e. $$a_0 = \frac{855 - 820.4}{7} = 4.94$$

Thus the equation of the regression line of inductive reactance on frequency is:

$$\mathbf{Y = 4.94 + 0.586\,X}$$

Exercises

Question1

Find the equations of the regression lines of x on y, and y on x for the bivariate data in the table below:

x	2	3	4	5	6
y	5	2	5	3	1

Question2

Use normal equation to find the the equation of the regression line of y on x for the data in the table below.

x	1	2	4	6	7	8	10
y	10	14	12	13	15	12	13

Question3

For a given set of data, the regression line of y on x is $y = 0.4 + 1.3x$ and the regression line of x on y is $x = -0.1 + 0.7y$. Find:

a) The linear correlation coefficient.

b) The means $\bar{x}$ and $\bar{y}$.

Reference(s):

1. Ngezahayo, E. P. (2017). *Advanced Mathematics for Rwanda Secondary Schools. Learners' Book Senior Six*. Kigali: Fountain.
2. Ngezahayo, E. P. (2017). *Advanced Mathematics for Rwanda Secondary Schools. Learners' Book Senior Five*. Kigali: Fountain.
3. Ngezahayo, E. P. (2016). *Advanced Mathematics for Rwanda Secondary Schools*. Learners' Book Senior Four. Kigali: Fountain.
4. Dexter, J.B(2002). Engineering Mathematics: 5^{th} Edition
5. John, B. (2006). *Higher Engineering Mathematics*:5^{th} Copyright _c 2006.
6. Bivariate data, (2020) *Wikipedia.* Available at: https://en.wikipedia.org/wiki/Bivariate data (Accessed: 8 June 2020).