INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -30

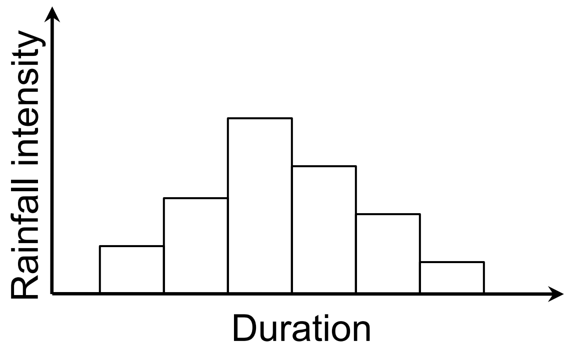
Course Instructor : Prof. P. P. MUJUMDAR
Department of Civil Engg., IISc.

Summary of the previous lecture

- IDF relationship
  - Procedure for creating IDF curves
  - Empirical equations for IDF relationships

IDF Curves

Design precipitation Hyetographs from IDF relationships:



Alternating block method :

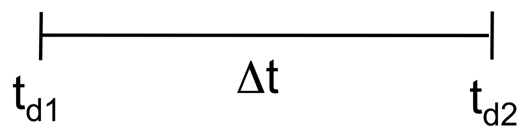
- Developing a design hyetograph from an IDF curve.
- Specifies the precipitation depth occurring in n successive time intervals of duration Δt over a total duration T_d.

IDF Curves

Procedure

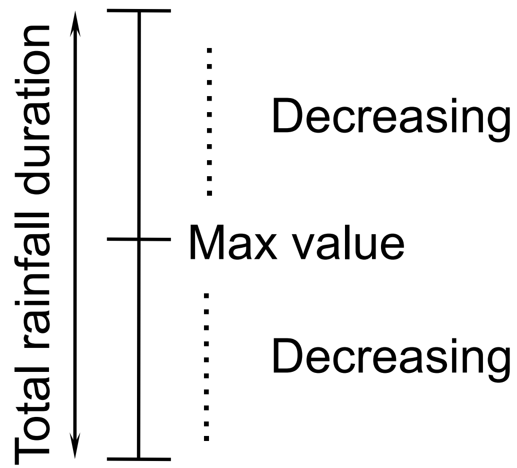
- Rainfall intensity (i) from the IDF curve for specified return period and duration(t_d) .
- Precipitation depth (P) = i x t_d
- The amount of precipitation to be added for each additional unit of time Δt.

$$P_{\Delta t} = P_{td2} - P_{td1}$$



IDF Curves

- The increments are rearranged into a time sequence with maximum intensity occurring at the center of the duration and the remaining blocks arranged in descending order alternatively to the right and left of the central block to form the design hyetograph.



Example – 1

Obtain the design precipitation hyetograph for a 2-hour storm in 10 minute increments in Bangalore with a 10 year return period.

Solution:

The 10 year return period design rainfall intensity for a given duration is calculated using IDF formula by Rambabu et. al. (1979)

$$i = \frac{KT^a}{\left(t+b\right)^n}$$

Example – 1 (Contd.)

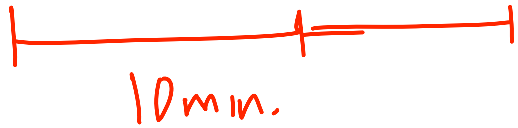
For Bangalore, the constants are

K = 6.275

a = 0.126

b = 0.5

n = 1.128



For T = 10 Year and duration, t = 10 min = 0.167 hr,

$$i = \frac{6.275 \times 10^{0.126}}{(0.167 + 0.5)^{1.128}} = 13.251$$ cms/hr

Example – 1 (Contd.)

- Similarly the values for other durations at interval of 10 minutes are calculated.
- The precipitation depth is obtained by multiplying the intensity with duration.

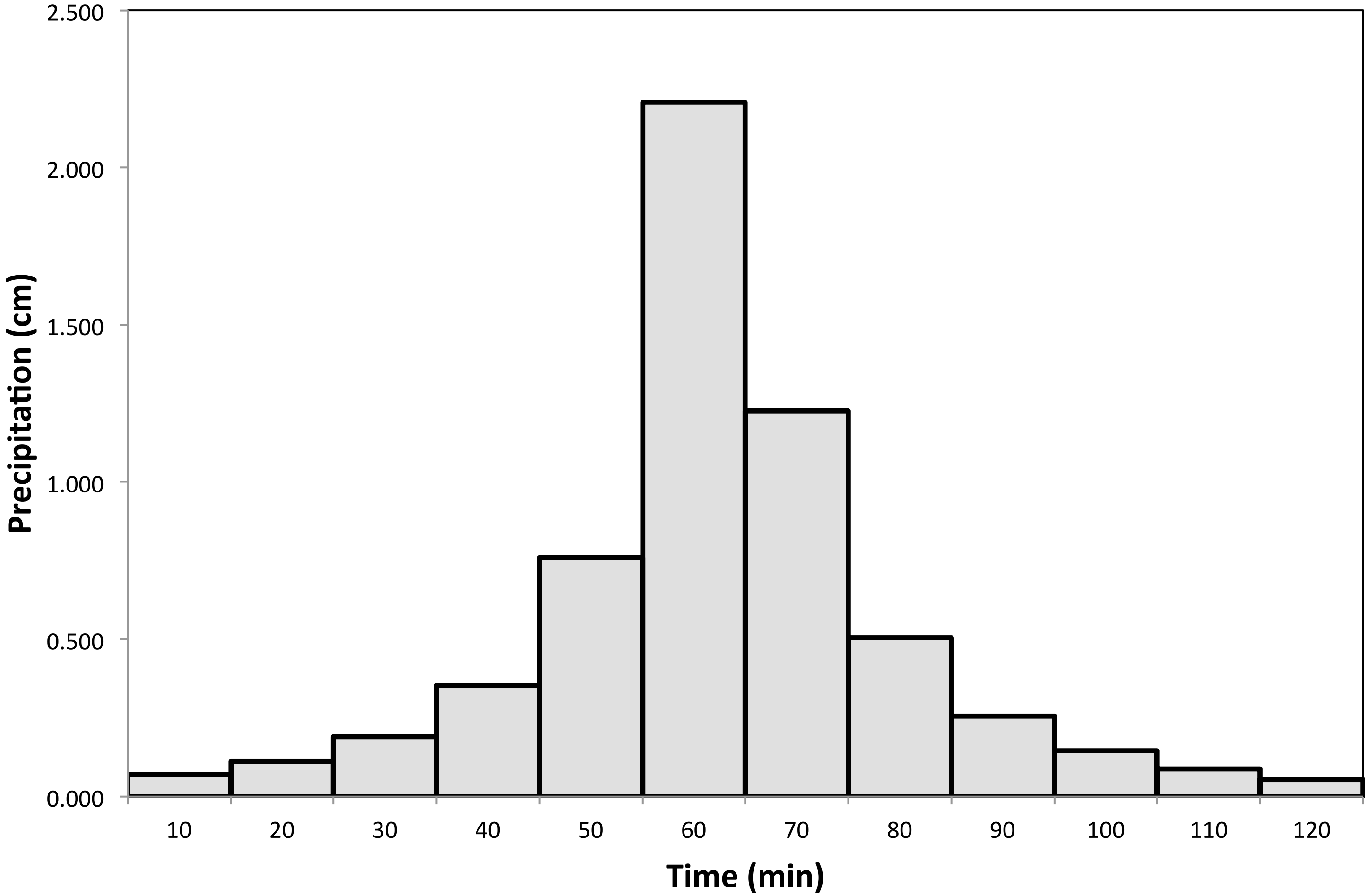
Precipitation = 13.251 * 0.167 = 2.208 cm

- The 10 minute precipitation depth is 2.208 cm compared with 3.434 cm for 20 minute duration, hence 2.208 cm will fall in 10 minutes, the remaining 1.226 (= 3.434 – 2.208) cm will fall in the remaining 10 minutes.
- Similarly the other values are calculated and tabulated

Example – 1 (Contd.)

Duration (min)	Intensity (cm/hr)	Cumulative depth (cm)	Incremental depth (cm)	Time (min)	Precipitation (cm)
10	13.251	2.208	2.208	0 - 10	0.069
20	10.302	3.434	1.226	10 - 20	0.112
30	8.387	4.194	0.760	20 - 30	0.191
40	7.049	4.699	0.505	30 - 40	0.353
50	6.063	5.052	0.353	40 - 50	0.760
60	5.309	5.309	0.256	50 - 60	2.208
70	4.714	5.499	0.191	60 - 70	1.226
80	4.233	5.644	0.145	70 - 80	0.505
90	3.838	5.756	0.112	80 - 90	0.256
100	3.506	5.844	0.087	90 - 100	0.145
110	3.225	5.913	0.069	100 - 110	0.087
120	2.984	5.967	0.055	110 - 120	0.055

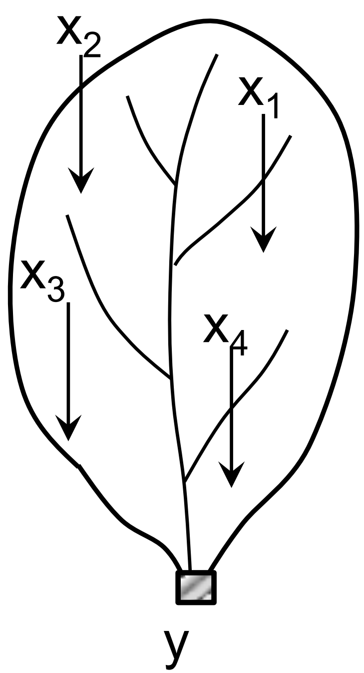
Example – 1 (Contd.)



MULTIPLE LINEAR REGRESSION

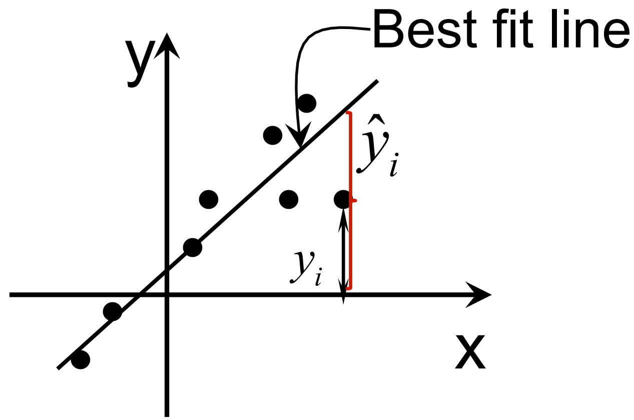
Multiple Linear Regression

- A variable (y) is dependent on many other independent variables, x_1, x_2, x_3, x_4 and so on.
- For example, the runoff from the water shed depends on many factors like rainfall, slope of catchment, area of catchment, moisture characteristics etc.



- Any model for predicting runoff should contain all these variables

Simple Linear Regression



(x_i, y_i) are observed values

$\hat{y}_i$ is predicted value of x_i

$$\hat{y}_i = a + bx_i$$

Error, $e_i = y_i - \hat{y}_i$

Estimate the parameters a, b such that the square error is minimum

Sum of square errors $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$

$$M = \sum_{i=1}^{n} \left\{ y_i - \left(a + bx_i \right) \right\}^2$$

Simple Linear Regression

$$M = \sum_{i=1}^{n} \left\{ y_i - a - bx_i \right\}^2$$

$$\frac{\partial M}{\partial a} = 0 \quad a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n}; \qquad a = \overline{y} - b\overline{x}$$

$$\frac{\partial M}{\partial b} = 0 \quad b = \frac{\sum x_i^{'} y_i^{'}}{\sum \left(x_i^{'} \right)^2} \quad \left(x_i - \overline{x} \right) = x_i^{'} \quad and \quad \left(y_i - \overline{y} \right) = y_i^{'}$$

$$\hat{y}_i = a + bx_i$$

Multiple Linear Regression

A general linear model of the form is

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots.. + \beta_p x_p$$

y is dependent variable,

$x_1, x_2, x_3, \ldots\ldots, x_p$ are independent variables and

$\beta_1, \beta_2, \beta_3, \ldots\ldots, \beta_p$ are unknown parameters

- ‘n’ observations are required on y with the corresponding ‘n’ observations on each of the ‘p’ independent variables.

Multiple Linear Regression

- 'n' equations are written for each observation as

$$y_1 = \beta_1 x_{1,1} + \beta_2 x_{1,2} + \dots\dots\dots + \beta_p x_{1,p}$$
$$y_2 = \beta_1 x_{2,1} + \beta_2 x_{2,2} + \dots\dots\dots + \beta_p x_{2,p}$$
$$.$$
$$.$$
$$y_n = \beta_1 x_{n,1} + \beta_2 x_{n,2} + \dots\dots\dots + \beta_p x_{n,p}$$

- Solving 'n' equations for obtaining the 'p' parameters.
- 'n' must be equal to or greater than 'p', in practice 'n' must be at least 3 to 4 times large as 'p'.

Multiple Linear Regression

- If y_i is the i^{th} observation on y and $x_{i,j}$ is the i^{th} observation on the j^{th} independent variable, the generalized form of the equations can be written as

$$y_i = \sum_{j=1}^{p} \beta_j x_{i,j}$$

- The equation can be written in matrix notation as

$$Y_{(n\times 1)} = X_{(n\times p)} \times \mathrm{B}_{(p\times 1)}$$

Multiple Linear Regression

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ y_n \end{bmatrix}_{nx1} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & . & . & x_{1,p} \\ x_{2,1} & x_{2,2} & x_{2,3} & . & . & x_{2,p} \\ x_{3,1} & & & & & \\ . & & & & & \\ . & & & & & \\ x_{n,1} & x_{n,1} & & & & x_{n,p} \end{bmatrix}_{nxp} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ . \\ . \\ \beta_p \end{bmatrix}_{px1}$$

Y is an nx1 vector of observations on the dependent variable, X is an nxp matrix with n observations on each p independent variables, Β is a px1 vector of unknown parameters.

Multiple Linear Regression

- If $x_{i,1}=1$ for $\forall$ i, β_1 is the intercept
- Parameters β_j, j = 1….p are estimated by minimizing the sum of square errors (e_i)

$$e_i = y_i - \hat{y}_i$$

$$\hat{y}_i = \sum_{j=1}^{p} \beta_j x_{i,j}$$

Multiple Linear Regression

In matrix notation,

$$\sum e_i^2 = E^{'}E$$

$$= \left(Y - X\hat{\mathrm{B}}\right)^{'}\left(Y - X\hat{\mathrm{B}}\right)$$

$$\frac{d}{d\mathrm{B}}\sum e_i^2 = 0 \qquad \forall j$$

$$0 = -2X^{'}\left(Y - X\hat{\mathrm{B}}\right)$$

$$X^{'}Y = X^{'}X\hat{\mathrm{B}}$$

Multiple Linear Regression

- Premultiplying with $\left(X^{'}X\right)^{-1}$ on both the sides,

$$\left(X^{'}X\right)^{-1}X^{'}Y=\left(X^{'}X\right)^{-1}X^{'}X\hat{\mathrm{B}}$$

$$\left(X^{'}X\right)^{-1}X^{'}Y=\hat{\mathrm{B}}$$

or

$$\hat{\mathrm{B}}=\left(X^{'}X\right)^{-1}X^{'}Y$$

- $\left(X^{'}X\right)$ is a pxp matrix and rank must be p for it to be inverted.

Multiple Linear Regression

- Suppose if no. of regression coefficients are 3, then matrix $(X'X)$ is as follows

$$
(X'X) = \begin{bmatrix} \sum_{i=1}^{n} x_{i,1}^2 & \sum_{i=1}^{n} x_{i,2}x_{i,1} & \sum_{i=1}^{n} x_{i,3}x_{i,1} \\ \sum_{i=1}^{n} x_{i,1}x_{i,2} & \sum_{i=1}^{n} x_{i,2}^2 & \sum_{i=1}^{n} x_{i,3}x_{i,2} \\ \sum_{i=1}^{n} x_{i,1}x_{i,3} & \sum_{i=1}^{n} x_{i,2}x_{i,3} & \sum_{i=1}^{n} x_{i,3}^2 \end{bmatrix}
$$

Multiple Linear Regression

- A multiple coefficient of determination, R^2 (as in case of simple linear regression) is defined as

$$R^2 = \frac{Sum\ of\ squares\ due\ to\ regression}{Sum\ of\ squares\ about\ the\ mean}$$

$$= \frac{B' X' Y - n\bar{y}^2}{Y' Y - n\bar{y}^2}$$

Example – 2

In a watershed, the mean annual flood (Q) is considered to be dependent on area of watershed (A) and rainfall(R). The table gives the observations for 12 years. Obtain regression coefficients and R^2 value.

Q in cumec	0.44	0.24	2.41	2.97	0.7	0.11	0.05	0.51	0.25	0.23	0.1	0.054
A in hectares	324	226	1474	2142	420	45	38	363	77	84	46	38
Rainfall in cm	43	53	48	50	43	61	81	68	74	71	71	69

Example – 2 (Contd.)

The regression model is as follows

$$Q = \beta_1 + \beta_2 A + \beta_3 R$$

Where Q is the mean flood in m^3/sec,

A is the watershed area in hectares and

R is the average annual daily rainfall in mm

This is represented in matrix form as

$$Y_{(12\times1)} = X_{(12\times3)} \times \mathrm{B}_{(3\times1)}$$

Example – 2 (Contd.)

To obtain coefficients this equation is to be solved

$$\begin{bmatrix} 0.44 \\ 0.24 \\ 2.41 \\ 2.97 \\ 0.7 \\ 0.11 \\ 0.05 \\ 0.51 \\ 0.25 \\ 0.23 \\ 0.1 \\ 0.054 \end{bmatrix}_{12x1} = \begin{bmatrix} 1 & 324 & 43 \\ 1 & 226 & 53 \\ 1 & 1474 & 48 \\ 1 & 2142 & 50 \\ 1 & 420 & 43 \\ 1 & 45 & 61 \\ 1 & 38 & 81 \\ 1 & 363 & 68 \\ 1 & 77 & 74 \\ 1 & 84 & 71 \\ 1 & 46 & 71 \\ 1 & 38 & 69 \end{bmatrix}_{12x3} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}_{3x1}$$

Example – 2 (Contd.)

The coefficients are obtained from

$$\hat{B} = \left(X^{'}X\right)^{-1} X^{'}Y$$

$$\left(X^{'}X\right) = \begin{bmatrix} \sum_{i=1}^{n} x_{i,1}^{2} & \sum_{i=1}^{n} x_{i,2}x_{i,1} & \sum_{i=1}^{n} x_{i,3}x_{i,1} \\ \sum_{i=1}^{n} x_{i,1}x_{i,2} & \sum_{i=1}^{n} x_{i,2}^{2} & \sum_{i=1}^{n} x_{i,3}x_{i,2} \\ \sum_{i=1}^{n} x_{i,1}x_{i,3} & \sum_{i=1}^{n} x_{i,2}x_{i,3} & \sum_{i=1}^{n} x_{i,3}^{2} \end{bmatrix}$$

Example – 2 (Contd.)

$$\left(X^{'}X\right)=\begin{bmatrix} 12 & 5277 & 732 \\ 5277 & 7245075 & 269879 \\ 732 & 269879 & 46536 \end{bmatrix}$$

The inverse of this matrix is

$$\left(X^{'}X\right)^{-1}=\begin{bmatrix} 3.35 & -6.1\times10^{-4} & -0.05 \\ -6.1\times10^{-4} & 2.9\times10^{-7} & 7.9\times10^{-6} \\ -0.05 & 7.9\times10^{-6} & 7.5\times10^{-4} \end{bmatrix}$$

Example – 2 (Contd.)

$$\left(X^{'}Y\right) = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i,2} y_i \\ \sum_{i=1}^{n} x_{i,3} y_i \end{bmatrix} = \begin{bmatrix} 8.06 \\ 10642 \\ 417 \end{bmatrix}$$

Example – 2 (Contd.)

$$\hat{B} = \left(X^{'}X\right)^{-1} X^{'}Y$$

$$= \begin{bmatrix} 3.35 & -6.1\times10^{-4} & -0.05 \\ -6.1\times10^{-4} & 2.9\times10^{-7} & 7.9\times10^{-6} \\ -0.05 & 7.9\times10^{-6} & 7.5\times10^{-4} \end{bmatrix} \times \begin{bmatrix} 8.06 \\ 10642 \\ 417 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0351 \\ 0.0014 \\ 5.0135\times10^{-5} \end{bmatrix}$$

Example – 2 (Contd.)

Therefore the regression equation is as follows

$$Q = 0.0351 + 0.0014A + 5.0135*10^{-5}R$$

From this equation, the estimated Q and the corresponding errors are tabulated.

Example – 2 (Contd.)

Q	A	R	$\hat{Q}$	e
0.44	324	43	0.49	-0.05
0.24	226	53	0.35	-0.11
2.41	1474	48	2.10	0.31
2.97	2142	50	3.04	-0.07
0.7	420	43	0.63	0.07
0.11	45	61	0.10	0.01
0.05	38	81	0.09	-0.04
0.51	363	68	0.55	-0.04
0.25	77	74	0.15	0.10
0.23	84	71	0.16	0.07
0.1	46	71	0.10	0.00
0.054	38	69	0.09	-0.04

Example – 2 (Contd.)

Multiple coefficient of determination, R^2 :

$$R^2 = \frac{B'X'Y - n\bar{y}^2}{Y'Y - n\bar{y}^2}$$

$$= \frac{15.64 - 5.42}{15.77 - 5.42}$$

$$= 0.99$$

$$\bar{y} = 0.672,\ n = 12$$

$$B' = \begin{bmatrix} 0.0351 & 0.0014 & 5.0135\times10^{-5} \end{bmatrix}$$

$$\left(X'Y\right) = \begin{bmatrix} 8.06 \\ 10642 \\ 417 \end{bmatrix}$$

$$Y'Y = 15.77$$

PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis

- Powerful tool for analyzing data.
- PCA is a way of identifying patterns in the data and data is expressed in such a way that the similarities and differences are highlighted.
- Once the patterns are found in the data, it can be compressed (reduce the number of dimensions) without losing information.
- Eigenvectors and eigenvalues are discussed first to understand the process of PCA.

Matrix Algebra

Eigenvectors and Eigenvalues:

- Let A be a complex square matrix. If λ is a complex number and X a non–zero complex column vector satisfying $AX = \lambda X$, X is an eigenvector of A, while λ is called an eigenvalue of A.
- X is the eigenvector corresponding to the eigenvalue λ.
- Eigenvectors are possible only for square matrices.
- Eigenvectors of a matrix are orthogonal.

Matrix Algebra

- If λ is an eigenvalue of an $n \times n$ matrix A, with corresponding eigenvector X, then $(A - \lambda I)X = 0$, with $X \neq 0$, so $\det(A - \lambda I) = 0$ and there are at most n distinct eigenvalues of A.
- Conversely if $\det(A - \lambda I) = 0$, then $(A - \lambda I)X = 0$ has a non–trivial solution X and so λ is an eigenvalue of A with X a corresponding eigenvector.

Example – 3

Obtain the eigenvalues and eigenvectors for the matrix,

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

The eigenvalues are obtained as

$$\left| A - \lambda I \right| = 0$$

$$\begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = 0$$

Example – 3 (Contd.)

$$(1-\lambda)(1-\lambda)-4=0$$

$$\lambda^2-2\lambda-3=0$$

Solving the equation,

$$\lambda = 3, -1$$

Therefore the eigenvalues are 3 and -1 for matrix A.

The eigenvector is obtained by

$$(A-\lambda I)X=0$$

Example – 3 (Contd.)

For $\lambda_1 = 3$

$$\left(A - \lambda_1 I\right) X_1 = 0$$

$$\begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 0$$

$$-2x_1 + 2y_1 = 0$$

$$2x_1 - 2y_1 = 0$$

which has solution x = y, y arbitrary.

eigenvectors corresponding to λ = 3 are the vectors $\begin{bmatrix} y \\ y \end{bmatrix}$, with y ≠ 0.

Example – 3 (Contd.)

For $\lambda_1 = -1$

$$\left(A - \lambda_2 I\right) X_2 = 0$$

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = 0$$

$$2x_1 + 2y_1 = 0$$

$$2x_1 + 2y_1 = 0$$

which has solution x = -y, y arbitrary.

eigenvectors corresponding to λ = -1 are the vectors $\begin{bmatrix} -y \\ y \end{bmatrix}$, with y ≠ 0.

Principal Component Analysis

Principal Component Analysis (PCA):

- When data is collected on ‘p’ variables, these variables are correlated
- Correlation indicates information contained in one variable is also contained in some of the other p-1 variables.
- PCA transforms the ‘p’ original correlated variables into ‘p’ uncorrelated components (also called as orthogonal components)
- These components are linear functions of the original variables.

Principal Component Analysis

The transformation is written as

$$Z = X \times A$$

Where

X is nxp matrix of 'n' observations on p variables

Z is nxp matrix of 'n' values for each of p components

A is pxp matrix of coefficients defining the linear transformation

All X are assumed to be deviations from their respective means, hence X is a matrix of deviations from mean

Principal Component Analysis

Steps for PCA:

- Get the data for n observations on p variables.
- Form a matrix with deviations from mean.
- Calculate the covariance matrix
- Calculate the eigenvalues and eigenvectors of the covariance matrix.
- Choosing components and forming a feature vector.
- Deriving the new data set.

Principal Component Analysis

The procedure is explained with a simple data set of the yearly rainfall and the yearly runoff of a catchment for 15 years.

Year	1	2	3	4	5	6	7	8	9	10
Rainfall (cm)	105	115	103	94	95	104	120	121	127	79
Runoff (cm)	42	46	26	39	29	33	48	58	45	20

Year	11	12	13	14	15
Rainfall (cm)	133	111	127	108	85
Runoff (cm)	54	37	39	34	25

Principal Component Analysis

Step 2: Form a matrix with deviations from mean

Original matrix

$$\begin{bmatrix} 105 & 42 \\ 115 & 46 \\ 103 & 26 \\ 94 & 39 \\ 95 & 29 \\ 104 & 33 \\ 120 & 48 \\ 121 & 58 \\ 127 & 45 \\ 79 & 20 \end{bmatrix}$$

Matrix with deviations from mean

$$\begin{bmatrix} -1.3 & 3.4 \\ 8.7 & 7.4 \\ -3.3 & -12.6 \\ -12.3 & 0.4 \\ -11.3 & -9.3 \\ -2.3 & -5.6 \\ 13.7 & 9.4 \\ 14.7 & 19.4 \\ 20.7 & 6.4 \\ -27.3 & -18.6 \end{bmatrix}$$

Principal Component Analysis

Step 3: Calculate the covariance matrix

$$\mathrm{cov}(X,Y) = s_{X,Y} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{n-1}$$

$$\begin{bmatrix} \mathrm{cov}(X,X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \mathrm{cov}(Y,Y) \end{bmatrix} = \begin{bmatrix} 216.67 & 141.35 \\ 141.35 & 133.38 \end{bmatrix}$$

Principal Component Analysis

Step 4: Calculate the eigenvalues and eigenvectors of the covariance matrix

$$\begin{bmatrix} \operatorname{cov}(X,X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(Y,X) & \operatorname{cov}(Y,Y) \end{bmatrix} = \begin{bmatrix} 216.67 & 141.35 \\ 141.35 & 133.38 \end{bmatrix}$$

Principal Component Analysis

Step 5: Choosing components and forming a feature vector