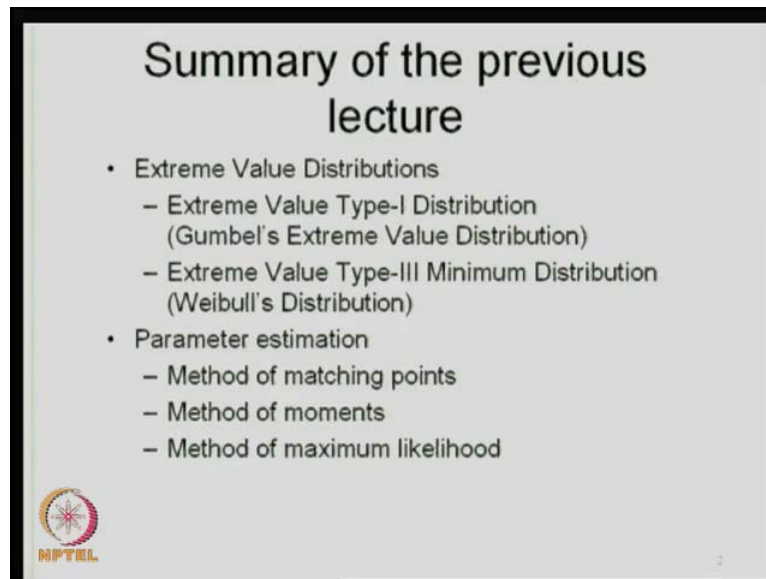


**Stochastic Hydrology**  
**Prof. P. P. Majumdar**  
**Department of Civil Engineering**  
**Indian Institute of Science, Bangalore**

**Lecture No. # 08**  
**Covariance and Correlation**

(Refer Slide Time: 00:23)



Good morning and welcome to this the eighth lecture in the course of stochastic hydrology or if you recall in the last class last lecture, we discuss the extreme value distributions specifically the extreme value type of one distribution, which is also called as the Gumbel's extreme value distribution. This is generally used for the maximum values for example, you may be interested in the peak flows maximum rainfall etcetera; then we also covered the extreme value type three distribution for minimum values, this is also called as the Weibull's distribution; recall that we use the Weibull's distributions specifically for low flows of minimum rainfall minimum water quality and so on.

Then we went on to discuss the topic of parameter estimation, where given a pdf we would be interested in getting the parameters from a given sample values. In this we discussed three methods, the method of matching points there from the available data observations we reduce certain probabilities for example, probability of  $X$  being greater than equal to a certain value is equal to let say 0.8 or some such thing and then

depending on the number of parameters that you have for the pdf, you deduce various such probabilities. Let say you if you have two parameters, you deduce two probabilities from the available observations and equate it to the theoretical probability that would result from the use of that specific probability distribution by equating, these you will be able to get the parameters of that particular distribution. The method of moments we take those many moments of the pdf as you have the number of parameters for example, if you have two parameters, you take consider the first two moments of the pdf, the first moment being the expected value of the random variable itself the second moment you take about the mean and therefore, you define the variants.

In the method of in the you generate those many moments as you have number of parameters solve those equations and then get the parameters from the sample in the method of maximum likelihood which will again review today you take you define what is called as the likelihood function, which is based on the sample values  $X_1, X_2, X_3$  etcetera,  $X_n$ ; and then you look at those parameter values of the pdf, which will maximize this likelihood. So, essentially the principle there is what is that set of parameters? We are looking for that particular set of parameters that will maximize the likelihood of obtaining the sample  $X_1, X_2$  etcetera  $X_n$  which has in fact, been realized that is the idea there.

(Refer Slide Time: 03:37)

**Method of Maximum Likelihood**

- The likelihood function is constructed as,  

$$L = f(x_1; \theta_1; \theta_2 \dots \theta_m) \times f(x_2; \theta_1; \theta_2 \dots \theta_m) \times f(x_n; \theta_1; \theta_2 \dots \theta_m)$$

$$= \prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_m)$$
- Maximize the likelihood function  

$$\frac{\partial L}{\partial \theta_i} = 0 \quad \forall i$$
- Solving the 'm' equations, the 'm' parameter estimated

MPTEL

So, in the maximum likelihood method we define the likelihood function as  $f$  of  $x_1$  theta 1 theta 2 etcetera, theta  $m$  where theta  $I$  are the parameters into  $f$  of  $x_2$  theta 1 theta 2 etcetera theta  $m$  and so on until  $x_n$ . So, this we define it as multiplication of  $f$  of  $x_i$  theta 1 theta 2 etcetera theta  $m$  then we look at those parameter values theta  $I$  which will maximize the likelihood function thus defined. So, we maximize the likelihood function with respect to theta  $I$  so, we take the first derivatives of the likelihood function with respect to theta  $I$  for all  $I$  equate them to 0, thus generating  $m$  equations we solve these  $m$  equations to get the associated theta  $I$  values that is the principle of the maximum likelihood function maximum likelihood method.

(Refer Slide Time: 04:38)

**Example-1**

Obtain the maximum likelihood estimates of the parameter ' $\beta$ ' in the pdf

$$f(x) = 2\beta \sqrt{\frac{\beta}{\pi}} x^2 e^{-\beta x^2} \quad -\infty < x < \infty$$

$$L(\beta) = 2\beta \sqrt{\frac{\beta}{\pi}} x_1^2 e^{-\beta x_1^2} \times 2\beta \sqrt{\frac{\beta}{\pi}} x_2^2 e^{-\beta x_2^2} \dots \dots \dots 2\beta \sqrt{\frac{\beta}{\pi}} x_n^2 e^{-\beta x_n^2}$$

$$= 2^n \beta^n \left(\frac{\beta}{\pi}\right)^{n/2} \left(\prod_{i=1}^n x_i^2\right) e^{-\sum_{i=1}^n \beta x_i^2}$$

$$= 2^n \beta^{(n+n/2)} \pi^{-n/2} \left(\prod_{i=1}^n x_i^2\right) e^{-\sum_{i=1}^n \beta x_i^2}$$

Let us consider an example in the last class, we discussed one simple example of the exponential distribution, which has a single parameter lambda. Now we will take another distribution which has a single parameter beta, the pdf is given by  $2\beta \sqrt{\beta/\pi} x^2 e^{-\beta x^2}$ , which is defined for  $x$  varying between minus infinity to plus infinity.

So, we formulate the  $L$  of beta remember we have a sample  $x_1, x_2, x_3$  etcetera and we are numerating the pdf at those given values of  $x_i$  and then taking the product of these  $f$  of  $x$  defined over defined at that particular  $x_i$  and then we are calling that as a likelihood function. So, we define the likelihood function is equal to  $2\beta \sqrt{\beta/\pi} x_1^2 e^{-\beta x_1^2}$  into etcetera like this


every time we take the  $x$  value to  $x_1, x_2, x_3$  etcetera up to  $x_n$ , and then define the likelihood function. So, you have  $n$  terms here; so this would be  $2$  to the power  $n$  beta to the power  $n$  and beta by  $\pi$  to the power  $n$  by  $2$ , because it is a square root here and then you have  $n$  such terms so, beta by  $\pi$  to the power  $n$  by  $2$ , then look at the  $n$  values of  $x_i$ . So,  $x_1$  square into  $x_2$  square into  $x_3$  square etcetera, etcetera. So, I write that as  $\pi$  of  $I$  is equal to one to  $n$   $x_i$  square then  $e$  to the power minus beta  $x_1$  square plus  $x_2$  square etcetera. So, I write this as  $e$  to the power minus summation  $I$  is equal to  $1$  to  $n$  beta  $x_i$  square so, this in a simple form we write it as  $2$  to the power  $n$  beta to the power there is a  $n$  here and  $n$  by  $2$  here. So, I write it as beta to the power  $n$  plus  $n$  by  $2$ , then  $\pi$  to the power minus  $n$  by  $2$  corresponding to this term into the product  $I$  is equal to  $1$  to  $n$ , this term remains the same, and this term remains the same this is a likelihood function.

Now, we are looking for those values of beta which maximize this likelihood function recall that in the last class I mentioned about the log function the log of a log of  $m$  particular argument will have the maximum value at the same value corresponding to that argument where the argument itself would have had the maximum value, what I mean by that is log of a function is a monotonous function monotonic function and therefore, it will have the maximum value at the same point where the function itself would have the maximum value therefore it is sometimes advantageous to take the log of likelihood and then maximize the logarithm of that function specifically, when you have exponential functions like this. So, we take the log of the likelihood function from this you write the log of likelihood function as see we are looking at this point. So, we are taking the logarithm on both sides.

(Refer Slide Time: 08:04)

**Example-1 (contd.)**

$$\ln L(\beta) = n \ln 2 + (n + n/2) \ln \beta - \frac{n}{2} \ln \pi + \ln \left( \prod_{i=1}^n x_i^2 \right) - \beta \sum_{i=1}^n x_i^2$$
$$\frac{\partial \ln L(\beta)}{\partial \beta} = 0$$
$$(n + n/2) \frac{1}{\beta} - \sum_{i=1}^n x_i^2 = 0$$
$$\frac{3n}{2} = \sum_{i=1}^n x_i^2 \times \beta$$
$$\hat{\beta} = \frac{3n}{2 \sum_{i=1}^n x_i^2}$$

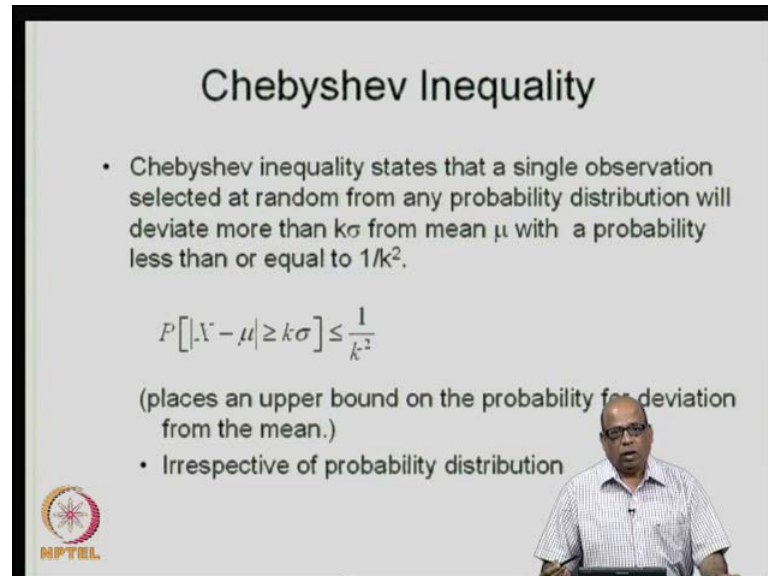


So, we write this as  $n \log 2$  plus  $n$  plus  $n$  by  $2 \log \beta$  minus  $n$  by  $2 \log \pi$  because we had a  $\pi$  of  $\pi$  to the power minus  $n$  by  $2$  plus logarithm of the product  $\prod_{i=1}^n x_i^2$  minus  $\beta \sum_{i=1}^n x_i^2$ .  $\beta$  is the only parameter, so we differentiate this function  $\log L$  of  $\beta$  with respect to  $\beta$  and equate to  $0$ . So, when you differentiate with respect to  $\beta$  the only terms containing  $\beta$  will appear here, so  $n$  plus  $n$  by  $2$  by  $1$  by  $\beta$ , you are differentiating with respect to  $\beta$  minus  $\sum_{i=1}^n x_i^2$ , you are again differentiating with respect to  $\beta$  those terms which do not contain  $\beta$  will vanish from here. And therefore, you get after simplification  $\hat{\beta}$  is equal to  $3n$  divided by  $2$  into summation  $i$  is equal to  $1$  to  $n$   $x_i^2$ .

So, essentially what we did is given a pdf? We formulate the likelihood function for the sample, which is realized which is  $x_1, x_2, x_3$  etcetera  $x_n$ , and then depending on the nature of the likelihood function sometimes we take the logarithm of the likelihood function and maximize the logarithm of the likelihood function in this particular case we did take the logarithm and because we are maximizing we take the first two differential the which is a necessary condition first differential with respect to the parameter equated to  $0$  and solve for that parameter. If you had more than one parameter seen this case what you would have done? You would have differentiated the logarithm of this likelihood function with respect to each of the parameters and therefore, and thus generating those

many equations as you have number of parameters solve all of them to get those parameters.

(Refer Slide Time: 10:24)




**Chebyshev Inequality**

- Chebyshev inequality states that a single observation selected at random from any probability distribution will deviate more than  $k\sigma$  from mean  $\mu$  with a probability less than or equal to  $1/k^2$ .

$$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

(places an upper bound on the probability for deviation from the mean.)

- Irrespective of probability distribution

 NPTEL

Now, we come to an interesting result called as a Chebyshev inequality see, what we did just now is to estimate the parameters. So, we had a probability distribution function given probability density function given and then we were estimating the parameters. Now, once we estimate the parameters for a given sample you have the complete description of a probability density function and therefore, the c d f cumulative distribution function in place and then you would have talked about various probabilities. So, so far we have been talking about probabilities associated with a given density function or a distribution function. Now, there will be situations, where you would be interested in not so much on the probability of a particular event itself as in the deviations of that particular random variable, and you do not have information on the probability underline probability density function or the probability distribution function.

So, the Chebyshev inequality lets first state it, and then see the significance of this the Chebyshev inequality states that a single observation selected at random from any probability distribution will deviate more than  $k$  sigma from the mean with a probability less than or equal to  $1$  by  $k$  square more formally we write this as probability of the absolute value of  $x$  minus  $\mu$  being greater than equal to a specified value  $k$  sigma  $k$  time sigma will be less than or equal to  $1$  by  $k$  square. So, here we are interested in what

is a maximum probability with which a given value of  $x$  will differ from its mean by more than  $k$  sigma, let us say more than 1 sigma 2 sigma 1.5 sigma etcetera. So, we are interested in how far is a deviation what is the maximum probability by which it will differ from a its mean on either side. So, it can be either  $x$  minus  $\mu$  may be positive or  $x$  minus  $\mu$  may be negative on either side what is the probability that it will deviate more than  $k$  sigma?

Now, the Chebyshev inequality places an upper bound on this probability and that is  $1/k^2$  remember this result is irrespective of the probability distribution and therefore, it becomes handy when we are when we want to place higher limits or the upper bounds on the probability of this particular deviation and this becomes quite handy in certain situations. For example you are talking about the stream flow stream flow and you have the mean value you do not have the information on the probability distribution and you would be interested in what is the probability that the stream flow will deviate from the mean by 1 sigma 2 sigma and. So, on and you will be interested in the maximum probability. So, that maximum probability is given by  $1/k^2$ . So, in many situations the Chebyshev inequality becomes is a very handy result to use in applications when you do not have information on the underlying probability distribution itself, but you would be interested in getting the maximum probabilities of the deviations from the mean.

(Refer Slide Time: 13:56)



### Example-2

The mean annual stream flow of a river is  $135 \text{ Mm}^3$  and standard deviation is  $23.8 \text{ Mm}^3$ . What is the maximum probability that the flow in a year will deviate more than  $45 \text{ Mm}^3$  from the mean.

Applying Chebyshev inequality,  $P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$

$k\sigma = 45$   
 $k \times 23.8 = 45$   
 $k = 1.891$

$P[|X - \mu| \geq 45] = P[|X - \mu| \geq 1.891\sigma] \leq \frac{1}{k^2}$   
 $\leq 1/1.89^2$   
 $\leq 0.28$

Let look at a simple example where we are considering the mean annual stream flow for river being given as 135 million cubic meters and it is a standard deviation is 23.8 million cubic meters remember these values we would have got from the sample. So, we would have estimated the mean to be 135 million cubic meters and the standard deviation to be 23.8 million cubic meters. Now we will be interested in what is a maximum probability that the flow in a year will deviate more than 45 million cubic meters from the mean these kind of questions should be of course, of practical relevance because knowing the mean we may be interested in a seeing how low the flow can go or what is a maximum probability that the flow will deviate by 45 million cubic meters. From this on either side, because you would like to plan for water resources utilization the stream flow utilization based from such information and therefore, you will be interested not so much in the probability density function or the probability distribution functions themselves, but you would be interested in what will be the maximum probability of such an event happening.

So, we will use the Chebyshev inequality which states that probability of the absolute value of the deviation  $x - \mu$  being greater than equal to  $k$  sigma will be less than equal to  $1/k^2$ . So, here we are saying that  $k$  sigma is equal to 45, because we are saying, what is the maximum probability that the flow in a year will deviate more than 45 million cubic meters. So, we are saying  $k$  sigma is equal to 45 so,  $k$  into 23.8 which is the sigma standard deviation will be equal to 45 and therefore,  $k$  will be equal to 1.891 in this particular expression and therefore, we write this as probability of the deviation the absolute value of the deviation being greater than equal to 45. I write this as probability of the absolute value of the deviation being greater than equal to 1.891 sigma and from the Chebyshev inequality this should be less than equal to  $1/k^2$ , this should be less than equal to  $1/1.891^2$  which is less than equal to 0.28 which means we are saying that the probability that the mean annual stream flow. The annual stream flow will deviate more than 45 million cubic meters from the mean is less than equal to 0.28 this is a result that we obtain from the Chebyshev inequality.

So, one is specifying a probability density function obtaining its parameters using any of the three methods that we discussed and then from the probability density function, which is thus defined completely by estimating the parameters from the sample you talk about various probabilities; whereas the chebyshev inequality does not bother about the



probability distribution from which the sample has been drawn, but you have the estimates of the mean and the standard deviation from which you talk about the maximum probabilities or the upper bounds on the probabilities of the deviation of  $x$  minus  $\mu$  on either side.

(Refer Slide Time: 17:45)



**Moments and Expectation –  
Jointly Distributed Random Variables**

$$\mu_n = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx \rightarrow n^{\text{th}} \text{ moment about mean}$$

... Single dimensional RV

X and Y are jointly distributed random variables;  
f(x,y) is joint pdf.  
r, s<sup>th</sup> moment of the two dimensional rv (X, Y) is

$$\mu_{r,s} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^r (y - \mu_y)^s f(x,y) dx dy$$

Now, we will look at another important topic where we are discussing about the joint variations of two random variables in our initial classes typically the second lecture, we discussed about the joint density function and subsequently the conditional density function and the marginal densities and so on. When we are talking about the two dimensional random vectors, we will now introduce the moments of the two dimensional random vectors with a specific purpose recall that the first the nth moment of a single dimensional random variable we define this as  $x$  minus  $\mu$  to the power  $n$  that is the integral minus infinity to plus infinity  $x$  minus  $\mu$  to the power  $n$   $f$  of  $x$   $dx$ , this is a nth moment about mean of the single dimensional variable and from this by putting  $n$  is equal to 2 what did we obtain we obtain the variants. So, sigma square was  $x$  minus  $\mu$  to the power 2  $f$  of  $x$   $dx$ .

Now, when you have two random variables  $x$  and  $y$  which are jointly distributed random variables with  $f$  of  $x$   $y$  as the joint p d f now we define analogous to the single dimension random variable we define the  $r, s$  moment of this two dimensional random variable as double integral minus infinity to plus infinity  $x$  minus  $\mu_x$  to the power  $r, y$  minus  $\mu_y$

to the power  $s$ ,  $f$  of  $x$   $y$   $dx$   $dy$ . So, this is the definition of the  $r, s$  moment of a two dimensional random variable analogous to the  $n$ th moment of the single dimension random variable. When  $r$  is equal to 1 and  $s$  is equal to 1, we call this moment as the covariance so, in the case of single dimension random variable when  $n$  is equal to 2 we called that as the variance. So, in the case of two dimensional random variables when  $r$  is equal to 1 and  $s$  is equal to 1 that is 1 1th moment is what we are talking about that is defined as the covariance.

(Refer Slide Time: 20:15)

**Covariance**

- Covariance of X and Y

$$\mu_{1,1} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy$$

$$= E[(x - \mu_x)(y - \mu_y)]$$

- Also denoted as  $\sigma_{X,Y}$  or  $\text{Cov}(X, Y)$
- $\sigma_{X,Y} = \text{Cov}(X, Y) = 0$ , if X and Y are independent
- The converse may not be necessarily true
- Sample estimate for population covariance is given by

$$s_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

So, we write the covariance of X and Y as double integral minus infinity to plus infinity  $x$  minus  $\mu_x$ ,  $y$  minus  $\mu_y$ ,  $f$  of  $x$   $y$   $dx$   $dy$  we call that what is your expected value? expected value of let say I write this as  $g$  of  $x$  comma  $y$  in a two dimensional random variable case this is nothing, but minus infinity to plus infinity the function  $g$  of  $x$  comma  $y$  into the  $p$   $d$   $f$  the joint pdf  $x$   $f$  of  $x$   $y$   $d$   $x$   $d$   $y$  this is how we define the expected value of a function we use that definition, and then our function here is  $x$  minus  $\mu_x$  into  $y$  minus  $\mu_y$  this is a function of  $x$  and  $y$  and multiplied by the joint density function  $f$  of  $x$   $y$  and with respect to  $d$   $x$  with respect to  $x$  and  $y$  we are integrating that with respect to  $x$  and  $y$ .

So, and therefore, from this result we write this as the expected value of  $x$  minus  $\mu_x$  into  $y$  minus  $\mu_y$  so, the covariance of  $x$  and  $y$  covariance of X, Y is given by the expected value of  $x$  minus  $\mu_x$  into  $y$  minus  $\mu_y$  the covariance is denoted as  $\sigma_x$

y or simply covariance of X, Y. So, this is how we denote the covariance remember from your earlier single dimension random variables we can show that  $\sigma_{xy}$  is equal to 0, if x and y are independent how do we show that, let say we will come back to this problem that is we are saying that covariance of x comma y or  $\sigma_{xy}$  how did we define this  $\sigma_{xy}$  we define as  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy$ .

This is how we defined our covariance now if x and y are independent that is these are independent random variables what is a property of stochastic independence? Recall that when x and y are independent your joint density function f of x y will be equal to the product of some marginal density function that is d of x into h of y. So, we use this result and then write f of x y is equal to g of x y x into h of y in this expression so, what do we write this will be  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy$  x will keep it as it is y minus  $\mu_y$  will keep it as it is and in place of f of x y i will write this as g of x into h of y, where g of x is the marginal density of x of x and h of y is marginal density of y. So, from this see here from this we write this as  $\sigma_{xy}$  that is covariance of x, y as  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) g(x) h(y) dx dy$ , because that g of x is a function of x alone and h of y is a function of y alone and therefore, I write this as  $\int_{-\infty}^{\infty} (x - \mu_x) g(x) dx \int_{-\infty}^{\infty} (y - \mu_y) h(y) dy$  this is with respect to x and this is with respect to y.

Can you recall this integral for example, I will write this as  $\int_{-\infty}^{\infty} (x - \mu_x) g(x) dx$  what is  $\int_{-\infty}^{\infty} (x - \mu_x) g(x) dx$  that is  $\mu_x$  itself minus  $\mu_x$  into  $\int_{-\infty}^{\infty} g(x) dx$  what is that integral  $\int_{-\infty}^{\infty} g(x) dx$  x will be equal to 1 because g of x is the probability density function and therefore, this should be  $\mu_x$  minus  $\mu_x$  itself that will be equal to 0. Similarly, this will be  $\mu_y$  into  $\mu_y$  itself that will be 0. So, this should be  $\sigma_{xy}$  will be equal to 0 so,  $\sigma_{xy}$  will be equal to 0, if x and y are independent you must remember; however, that the converse is not in general true that is you may have covariance of x, y as 0, but that does not necessarily mean that x and y are independent. So, we state  $\sigma_{xy}$  is equal to 0 if x and y are independent.

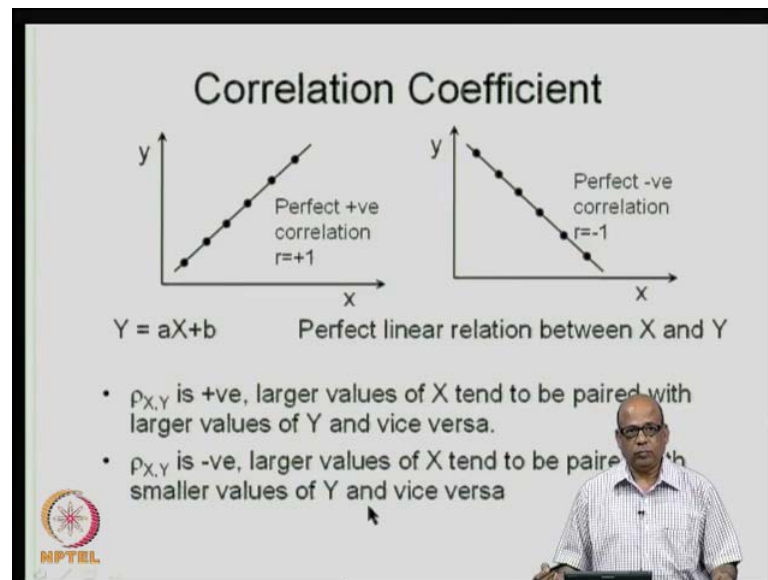
However the converse may not be necessarily true and from this again we write the sample estimate of the covariance is given by  $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

minus 1. So, you have  $n$  values of the  $n$  observe value of  $x_i$  and concurrent values of  $y_i$  so, you have  $n$  values of  $x_i$  and  $n$  values of  $y_i$  and in which case you estimate the covariance as given by this expression for example, you may have rainfall values let say annual rainfall values for a 50 years and the concurrent runoff values at a particular location generated by this particular rainfall for the same 50 years. So, you have the 50 values of rainfall which had generated the 50 values of runoff and then you are relating these two, and you are talking about the covariance between rainfall and runoff. So, that is how you estimate you estimate from the sample the covariance of this.

Now, from the covariance we move on to an important concept called as the correlation so, the correlation is actually a major of a degree the degree of association between two random variables  $x$  and  $y$  as you can see from the definition we define the covariance correlation  $\rho_{xy}$  as the covariance  $\sigma_{xy}$  divided by the standard deviation of  $x$  multiplied by the standard deviation of  $y$  this is a normalized covariance. So, we are normalizing the covariance  $\sigma_{xy}$  with respect to the standard deviation of  $x$  and standard deviation of  $y$  as you can see  $\sigma_{xy}$  had the units of  $x$  multiplied by the units of  $y$ , if your rainfall was in millimeters and runoff was also in millimeters then the covariance between rainfall a covariance of rainfall and runoff would be having units of millimeters square and the standard deviation of the rainfall  $x$  would be in millimeters. Standard deviation of  $y$  will be in millimeters and therefore, the  $\rho_{xy}$  that you get, which is a correlation is a unit less parameter and it is thus a normalized covariance, presently we will show that a  $\rho_{xy}$  in fact, varies between minus 1 and plus 1 and  $\rho_{xy}$  is equal to 0, if  $x$  and  $y$  are independent, because your  $\sigma_{xy}$  will be 0 as we just showed. So,  $\sigma_{xy}$  which is a covariance between  $x$  and  $y$  covariance of  $x$  and  $y$  will be 0, if  $x$  and  $y$  are independent and therefore, the correlation will be 0.

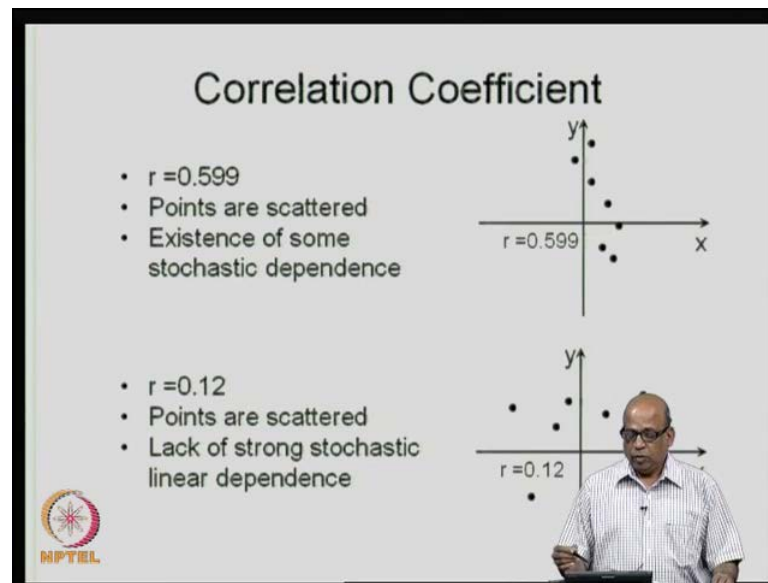
Remember if  $x$  and  $y$  are independent the correlation coefficient is 0, but often we confuse if correlation is 0, we often confuse it with  $x$  and  $y$  being independent it is not necessarily true. So, the fact that correlation is 0 should not directly employ that  $x$  and  $y$  are independent however if  $x$  and  $y$  are independent then correlation will be necessarily 0. So, from this definition we get the sample estimate of the correlation as  $r_{xy}$  is equal to the sample estimate of the covariance which is  $s_{xy}$  by the sample estimate of the standard deviation  $f_x$  and sample estimate of the standard deviation  $s_y$ .

(Refer Slide Time: 30:45)



If we have a perfect correlation; that means, if your  $r$  is equal to 1 or the row  $x$   $y$  between two variables  $x$  and  $y$  the correlation is one this indicates that all the values all the observed values will lie on a perfect straight line if your correlation is minus one then it will lie on a straight line something like this which means the higher the values of  $x$  the lower will be the values of  $y$ . So, row  $x$   $y$  is positive indicates that the large larger values of  $x$  tend to be paid with larger values of  $y$  and vice versa that is if row  $x$   $y$  is positive that is we say the higher the value of  $x$  the higher will be the value of  $y$  if row  $x$   $y$  is positive, if row  $x$   $y$  is negative the higher the value of  $x$  the lower will be the value of  $y$  and therefore, the larger values of  $x$  tend to be paid with smaller values of  $y$  and vice versa, if row  $x$   $y$  is negative remember also that the correlation coefficient is a measure of linear dependence. We will show that presently that the way we have defined row  $x$   $y$  it indicates the linear dependence between  $x$  and  $y$  and that is why in fact, the fact that row  $x$   $y$  is equal to 0 does not necessarily mean that there is a complete independence between  $x$  and  $y$  there may be a non-linear dependence between  $x$  and  $y$  which the correlation coefficient cannot capture.

(Refer Slide Time: 32:37)



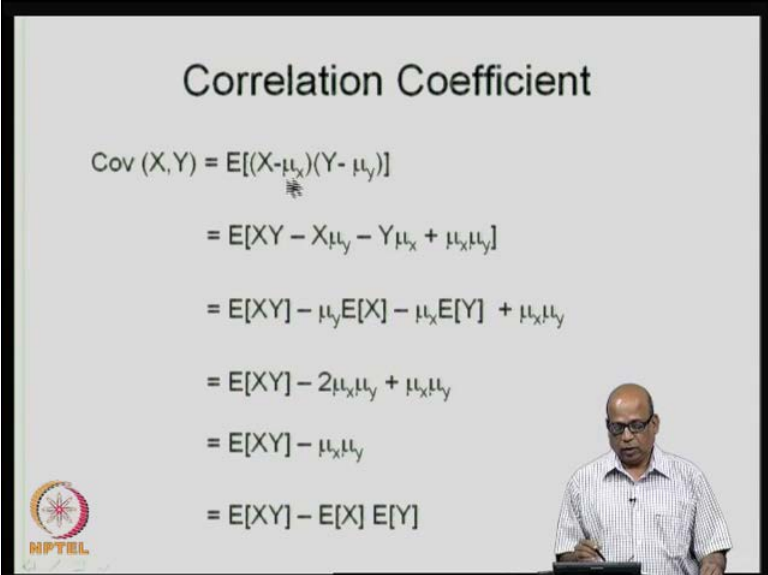
So, we will take some examples now you have observed values of  $x$  and  $y$  and these are the observed values, let say and you have a certain amount of correlation between them. So, there is certain stochastic dependence between these two so, it indicates that there is a certain linear dependence between  $x$  and  $y$  and if you take such a scatter here and you calculate the correlation coefficient you may get a much lower correlation coefficient corresponding to this. So, points are scattered to a greater extent than they were here and therefore, there is a smaller correlation coefficient. So, if you try to fit a straight line for this you may get a better straight line fit compared to this straight line the straight line that you may fit for this type of scatter and therefore, the correlation coefficient is much lower compared to the correlation coefficient here.

You look at this example here now this is a may be a parabolic type of this points lie on a parabola and you get a very high value of correlation  $r$  is equal to 0.949 although the relationship there is a functional relationship between  $x$  and  $y$  here the relationship was non-linear, but you can fit a straight line with much smaller scatter compared to the previous case here. So, you may be able to fit a straight line with a much smaller scatter and therefore, the correlation which indicates the linear dependence between  $x$  and  $y$  is much larger in this particular case you take the case of points which are lying around lying on the periphery of a circle on this on the circle. So, this is there is a perfect functional relationship between  $x$  and  $y$  defined by the equation of a circle here; however, the correlation coefficient will be 0 here. In this particular case, because there

is a non-linear functional relationship and the coefficient the correlation coefficient captures only the linear relationship.

(Refer Slide Time: 35:26)

### Correlation Coefficient

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - \mu_x)(Y - \mu_y)] \\ &= E[XY - X\mu_y - Y\mu_x + \mu_x\mu_y] \\ &= E[XY] - \mu_y E[X] - \mu_x E[Y] + \mu_x\mu_y \\ &= E[XY] - 2\mu_x\mu_y + \mu_x\mu_y \\ &= E[XY] - \mu_x\mu_y \\ &= E[XY] - E[X]E[Y] \end{aligned}$$


So,  $r$  will be equal to 0 although there is a perfect functional relationship between  $x$  and  $y$ . So, the point you must remember is that the correlation coefficient indicates or captures the degree of linear dependence between  $x$  and  $y$  we will just see what does correlation coefficient being for a perfectly linear relationship indicate for this we will express this covariance as covariance of  $x$  and  $y$  in a slightly more convenient fashion which this is equal to expected value of  $x$  minus  $\mu_x$   $y$  minus  $\mu_y$ . So, this we write it as expected value of  $x$   $y$  we simply simplify this  $x$   $y$  minus  $x$  into  $\mu_y$  minus  $y$  into  $\mu_x$  plus  $\mu_x$  into  $\mu_y$ . So, from this after you simplify this you get it as equal to expected value of  $x$   $y$  minus expected value of  $x$  expected value of  $y$ . So, this is how we express covariance of  $x$ ,  $y$  and then we revisit the definition of correlation coefficient this is  $\text{row } x \text{ } y$  as is defined as  $\sigma_{xy}$  by  $\sigma_x \sigma_y$ .

(Refer Slide Time: 36:19)

**Correlation Coefficient**

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

Consider  $Y = aX + b$ ; perfect linear relation

$$\rho_{X,Y}^2 = \frac{(\sigma_{X,Y})^2}{\sigma_X^2 \sigma_Y^2}$$
$$= \frac{(E[XY] - E[X]E[Y])^2}{\sigma_X^2 \sigma_Y^2}$$

Substitute  $Y = aX + b$

$$= \frac{(E[aX^2 + bX] - E[X]E[aX + b])^2}{\sigma_X^2 \sigma_Y^2}$$

Now, we consider a perfectly linear relationship between  $x$  and  $y$ , let say that  $y$  is defined as  $a x$  plus  $b$  which is the perfect linear relationship. Now we consider  $\rho_{x y}$  square which is a correlation coefficient square we write this as  $\sigma_{x y}$  square divided by  $\sigma_x$  square  $\sigma_y$  square and what is  $\sigma_{x y}$ ?  $\sigma_{x y}$  is expected value of  $x y$  minus expected value of  $x$  expected value of  $y$  and we are squaring the whole term and we will retain this  $\sigma_x$  square  $\sigma_y$  square as they are now and then substitute  $y$  is equal to  $a x$  plus  $b$ . So, wherever  $y$  is there I substitute  $a x$  plus  $b$  and simplify this as expected value of  $a x$  square, because  $y$  is equal to  $x$  plus  $b$  plus  $b x$  minus expected value of  $x$  and expected value of  $y$  here which is expected value of  $a x$  plus  $b$  the denominator will keep it as it is and then when we simplify this what do we get a square expected value of  $x$  square minus expected value of  $x$  the whole square can you recognize this term within the bracket it is I in fact,  $\sigma_x$  square which is the variance of  $x$ .

So, we write this as a square  $\sigma_x$  square the whole square. So, we get a square  $\sigma_x$  to the power 4 and we have  $\sigma_y$  square here and  $y$  is  $a x$  plus  $b$ . So, when  $y$  is  $a x$  plus  $b$  you recall that  $\sigma_y$  square is a square into  $\sigma_x$  square and therefore, we wrote this  $\sigma_x$  square and  $\sigma_x$  square gets cancelled here a square gets cancelled here. So, this turns out to be one therefore,  $\rho$  is equal to plus minus 1 plus or minus 1, if there is a perfect linear relationship, because we got  $\rho$  square is equal to 1 here. So, we started with  $\rho_{x y}$  square.



So, that a transfer to be 1 and therefore, row is equal to plus or minus 1, if there is a perfect linear relationship between x and y. So, the correlation coefficient is in fact, a measure of linear dependence. So, if correlation coefficient has any other value than 1 plus or minus 1, it indicates that there is a lesser degree of linear dependence between x and y compare to the perfect relationship which would have yielded a correlation coefficient of plus or minus 1.

(Refer Slide Time: 39:18)


**Example-3**

Obtain the correlation coefficient for 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



Let do an simple example, now let us say you have rainfall at a particular location and the associated runoff we have 15 observed values may be 15 years of observed values the rainfall values are in centimeters and the runoff values are also in centimeters. This runoff that we are talking about is in fact, generated by this rainfall. So, in the first year you have a 105 centimeters of rainfall that has generated 42 centimeters of runoff and so on. Now, we are trying to see, what is the relationship? What is the dependence of runoff on rainfall? So, we use the correlation coefficient as a measure of linear dependence and examine how much of dependence exist how much of linear dependence exists between runoff and rainfall. So, we simply use this expression that we have for the sample estimates of row x y this is  $s_{xy}$  by  $s_x s_y$  and how do we estimate  $s_{xy}$   $s_{xy}$  is simply summation of  $x$  minus  $\bar{x}$  into  $y$  minus  $\bar{y}$  summed over  $i$  is equal 1 to  $n$  divided by  $n$  minus 1 this is how we estimate the covariance between covariance of  $x$  and  $y$ .

(Refer Slide Time: 41:15)

### Example-3 (contd.)



$$\text{Mean, } \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\sum_{i=1}^n x_i = 1627$$

Therefore mean,  $\bar{x} = 1627/15$   
 $= 108.5 \text{ cm}$

$$\text{Variance, } s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{3499.73}{15-1} = 250$$



Standard deviation,  $s_x = 15.811 \text{ cm}$

So, we use these expressions and then calculate the correlation coefficient. So, first we obtain the mean which turns out to be a 1627 is a summation and mean of rainfall is 108.5 and similarly the variance  $s_x^2$  comes out to be 250 and the standard duration comes out to be 15.811. Similarly, for  $y$  which is the runoff we estimate  $\bar{y}$  as 38.33 centimeters and the variance as 117.5 centimeter square and therefore, the  $s_y$  comes out to be 10.841 centimeters. So, once we have the mean and the standard duration we open out columns like this the rainfall values are given here and the concurrent values of runoff are given here.

(Refer Slide Time: 41:39)

Year	Rainfall cm ( $x_i$ )	Runoff cm ( $y_i$ )	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	105	42	-3.47	3.67	12.02	13.44	-12.71
2	115	46	6.53	7.67	42.68	58.78	50.09
3	103	26	-5.47	-12.33	29.88	152.11	67.42
4	94	39	-14.47	0.67	209.28	0.44	-9.64
5	95	29	-13.47	-9.33	181.35	87.11	125.69
6	104	33	-4.47	-5.33	19.95	28.44	23.82
7	120	48	11.53	9.67	133.02	93.44	111.49
8	121	58	12.53	19.67	157.08	386.78	246.49
9	127	45	18.53	5.67	343.48	44.44	123.56
10	79	20	-29.47	-18.33	868.28	336.11	0.22
11	133	54	24.53	15.67	601.38	245.44	143.36
12	111	37	2.53	-1.33	42	1.78	-3.38
13	127	39	18.53	0.67	343.48	0.44	0.56
14	108	34	-0.47	-4.33	0.22	18.78	-2.05
15	85	25	-23.47	-13.33	550.68	177.78	310.11
<b>Total</b>	<b>1627</b>	<b>575</b>	<b>0</b>	<b>0</b>	<b>3499.73</b>	<b>1177.5</b>	<b>177.5</b>

Then we obtain  $x_i - \bar{x}$  and  $y_i - \bar{y}$   $(x_i - \bar{x})^2$   $(y_i - \bar{y})^2$  and then  $(x_i - \bar{x})(y_i - \bar{y})$ . So, your  $\bar{x}$  was given as 138 or something here  $\bar{x}$  is 108.5 and  $\bar{y}$  is 38.33. So, what we do is  $x_i - \bar{x}$  here and  $y_i - \bar{y}$  here remember because you are taking the first deviation it can be negative. So, similarly here first deviations can be negative when you square they will all be positive here and then you get  $(x_i - \bar{x})(y_i - \bar{y})$ . So, essentially you're multiplying this term with this term to obtain this term  $(x_i - \bar{x})(y_i - \bar{y})$  you will get this and this 1 is  $(x_i - \bar{x})^2$ . So, this term the whole square will give you this term. So, like this you calculate these for all of these terms here until 15 year.

So, for all the 15 terms you calculate these and you have the associated sums here. So, all these sums are available on the last row here now this you can do readily on any spreadsheet programs like Microsoft Excel and so on. So, very easily you can do all these calculations and then use these calculations for estimating your covariance first. So,  $s_{xy}$  you estimate it as  $(x_i - \bar{x})(y_i - \bar{y})$  summation of that, which is available in this column  $(x_i - \bar{x})(y_i - \bar{y})$  which is available in this column by 15 minus 1; so, 1974.67 by 15 minus 1, which turns out to be 141.05. What will be the units of this will have units of rainfall into runoff, which means centimeter square then from this you estimate the correlation coefficient  $s_{xy}$  by  $s_x s_y$ . So,  $s_{xy}$  is 141.05 and  $s_x$  which is a standard deviation of  $x$  calculated here as 15.811.

And similarly, standard deviation of  $y$  is 10.841. So, we use those values and get the correlation coefficient as 0.823, which is quite a high correlation coefficient indicating that there is a good linear dependence of runoff on the rainfall one the associated rainfall; however, how significant is the correlation is a different story all together we should explore whether the correlation coefficient that we got just now in fact, statistically significant that we will study slightly later. But you must remember that the correlation coefficient value, let us say 0.8, 0.7, 0.6 etcetera when you get correlation coefficients values like this you should not make a judgment just based on those values whether there is a strong linear relationship between the two variables that you are considered or not. You must also examine how significant is this correlation coefficient value for example, in the last example that we numerical example that we considered, we had 15 values and we obtained the correlation coefficient of let say 0.84 or something.

If we had a let us say instead of 15 values we had 100 values and we obtained a correlations of a 0.53 or something then we say that the 0.53 correlation indicates that there is a much lesser linear dependence of y on x compare to the 0.84 or something that you obtained for the 15 values we will not be able to say this and that is where we have to check whether the correlation coefficients we just obtained are in fact, statistically significant or not next we will go on to the next topic which is the dependent which is a continuation of the discussion of correlation where we are interested in obtaining a linear relationship. Let say between x and y we have the observed values of let us say rainfall and concurrent value of runoff or rainfall and let us say ground water recharge rainfall and evaporations and so on. So, we have concurrently observed values of two variables and we want to obtain a functional relationship between these two variables.

So, we what do we have we have just the observed value so, we have a scatter point scatter plot. So, if you plot x verses y simply you get scatter of x verses y now, with these scatter we would like to established a functional relationship and specifically a linear relationship to begin with between x and y what is the use of this let us say you have 50 observed value of rainfall and the concurrent 50 values of runoff observed. If you can fit a linear relationship between these two, you can use the linear relationship for estimating or predicting the value of runoff for any given value x that is the rainfall. So, for a given value of rainfall we should be able to estimate what will be the value of runoff. So, we can use these function relationship, how do we obtain these functional relationship you have the observed values of rainfall and runoff, and we have the scatter point as you can see from this figure you have these observed values.

(Refer Slide Time: 47:58)

### Simple Linear Regression



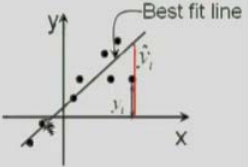
$(x_i, y_i)$  are observed values

$\hat{y}_i$  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  $\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

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


So, you have this scatter points like this now we want to obtain a linear relationship between these scatter points this procedure is called as simple linear regression simple because we are considering only two variables x and y, and y is the dependent variable, x is the independent variable and therefore, it is a simple regression linear, because we are fitting a linear relationship between x and y, if you have more independent variables. Let us say that y depends on x 1, x 2, x 3 and so one for example, runoff is y and it depends on rainfall x 1 it depends on let us say the temperature x 2, it depends on the catchment slope x 3 and so on. So, if the dependent variable depends on more than one independent variable then it is called as a multiple regression. And similarly if you are fitting a non-linear relationship it can be a non-linear regression multiple non-linear regression simple non regression and so on. So, first we begin with simple linear regression for what is our purpose.

We have these observed values each given by  $x_i, y_i$ . So, these are the actual observed values. So, when you plot them as scatter plots you will get these black dots. So, these black dots are the observed values, and we want to have the best fit line for which will represents these scatter values. So, if you look at a particular point  $y_i$  for a given  $x_i$  we have the  $y_i$ , which is the observed value and if you fit this best fit line you would have predicted the point to be at this location and this we denoted as  $\hat{y}_i$ . So,  $\hat{y}_i$  is the predicted value  $y_i$  is you observed value. So, like this for every value you have the observed value as well as the associated predicted value predicted value is on the line

which we would have use to predict that. So,  $\hat{y}_i$  is the predicted value of  $x_i$  and we write  $\hat{y}_i$  as  $a + bx_i$ , because we are fitting a straight line. So, we write this as  $a + bx_i$ .

What is the error your actual observed value is  $y_i$ , but your predicted value is  $\hat{y}_i$  therefore, error  $e_i$  or the point  $x_i, y_i$  is given by  $y_i - \hat{y}_i$ . So, there are two parameters here  $a$  and  $b$ . So, once we estimate these parameter  $a$  and  $b$  your straight line is completely defined we want to estimate these line such that the error are the sum of squared errors is a minimum that is over all these observed points the sum of the squared errors should be minimum. So, we will consider the sum of the squared errors  $i$  is equal to 1 to  $n$  the error square, which is  $(y_i - \hat{y}_i)^2$  and that we defined it as  $(y_i - a - bx_i)^2$  the whole square and we simplify that we are interested in getting those parameters  $a$  and  $b$  which will minimize this sum of square errors. So, we are looking at that particular straight line which will minimize the sum of square errors.

(Refer Slide Time: 51:42)

**Simple Linear Regression**

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

$$\frac{\partial M}{\partial a} = 0 \quad -2 \sum_{i=1}^n \{y_i - a - bx_i\} = 0$$

$$\sum_{i=1}^n \{y_i - a - bx_i\} = 0$$

$$\sum_{i=1}^n y_i - na - b \sum_{i=1}^n x_i = 0$$

$$a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n}$$


$$a = \bar{y} - b\bar{x}$$

So, we take the differential with respect to  $a$  and  $b$  both of this sum of the squared errors  $M$  and simplify that we first get  $a$  is equal to  $\bar{y} - b\bar{x}$  you get a  $\sum y_i$  term here and  $\sum x_i$  term here and divided by  $n$ . So, you get  $a$  is equal to  $\bar{y} - b\bar{x}$ . Similarly, you differentiate with respect to  $b$  now,  $\frac{dM}{db}$  is equal to 0 you get this expression all the details are here it is a simple arithmetic here and then you get this

term, but we write this in a slightly more relevant form by taking  $x_i$  minus  $\bar{x}$  which is a deviation of this particular point  $x_i$  from its mean  $\bar{x}$  as  $x_i - \bar{x}$  and  $y_i$  minus  $\bar{y}$  we write it as  $y_i - \bar{y}$ .

(Refer Slide Time: 52:59)


### Simple Linear Regression

$$\begin{aligned} \sum x_i' y_i' &= \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \sum \left( x_i y_i - \frac{\sum y_i}{n} x_i - \frac{\sum x_i}{n} y_i + \frac{\sum x_i}{n} \frac{\sum y_i}{n} \right) \\ &= \sum x_i y_i - \frac{\sum y_i}{n} \sum x_i - \frac{\sum x_i}{n} \sum y_i + n \left( \frac{\sum x_i \sum y_i}{n^2} \right) \\ &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} - \frac{\sum x_i \sum y_i}{n} + \frac{\sum x_i \sum y_i}{n} \\ &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} \end{aligned}$$


Using this we write the expression for  $b$  in more elegant form as this is all of this will give you the simplification what we said is  $x_i$  minus  $\bar{x}$  equal to  $x_i - \bar{x}$  and  $y_i$  minus  $\bar{y}$  is equal to  $y_i - \bar{y}$ . So, we consider this  $x_i - \bar{x}$   $y_i - \bar{y}$  which is  $x_i$  minus  $\bar{x}$   $y_i$  minus  $\bar{y}$  summation of that and that can be written as summation of  $x_i y_i$  minus summation  $x_i$  summation  $y_i$  by  $n$ , which means essentially we are writing this is equal to this expression and then we write this summation  $x_i - \bar{x}$  square which is nothing, but sigma of  $x_i$  minus  $\bar{x}$  the whole square that turns out to be sigma of  $x_i$  bar square minus sigma  $x_i$  the whole square by  $n$ .

(Refer Slide Time: 53:38)


### Simple Linear Regression

$$\begin{aligned}
 \sum (x'_i)^2 &= \sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
 &= \sum \left( x_i^2 - 2 \frac{\sum x_i}{n} x_i + \left\{ \frac{\sum x_i}{n} \right\}^2 \right) \\
 &= \sum x_i^2 - 2 \frac{\sum x_i}{n} \sum x_i + n \left\{ \frac{\sum x_i}{n} \right\}^2 \\
 &= \sum x_i^2 - 2 \frac{(\sum x_i)^2}{n} + \frac{(\sum x_i)^2}{n} \\
 &= \sum x_i^2 - \frac{(\sum x_i)^2}{n}
 \end{aligned}$$


Using these two expressions we write b in a more useful and more elegant form as  $\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$  which is  $\frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$ . So, you can get b once you get  $\sum x_i y_i$  and  $\sum x_i^2$ . So, b we obtain from the relationship that we just discussed and from that you know  $\bar{y}$  which is a mean of y and you know  $\bar{x}$  and therefore, you get a once you get a and b your relationship is completely defined  $\bar{y} = a + b\bar{x}$ .

(Refer Slide Time: 53:59)

### Simple Linear Regression

$$\begin{aligned}
 b &= \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \\
 &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\
 &= \frac{\sum x'_i y'_i}{\sum (x'_i)^2}
 \end{aligned}$$




So, this is how we obtain the best fit line for a scatter of points we will revisit the example that we just did in which we calculated the covariance and the correlation coefficient we will take the same example, because all the columns are available now with us.

(Refer Slide Time: 55:01)

**Example-4**

Consider the previous example, obtain the regression equation between rainfall (X) and runoff (Y)

$$\bar{x} = 108.5$$

$$\bar{y} = 38.33$$

$$\sum x_i' y_i' = 1974.67$$

$$\sum (x_i')^2 = 3499.73$$

$$b = \frac{\sum x_i' y_i'}{\sum (x_i')^2} = \frac{1974.67}{3499.73} = 0.564235$$

$$a = \bar{y} - b\bar{x} = 38.33 - (0.564235 \times 108.5) = -22.889$$

Therefore the equation is

$$Y = 0.564235X - 22.8895$$

So, we obtain x bar is equal to 108.5 and y bar is equal to 38.33, look at this columns here. So, you got x i minus x bar into y i minus y bar you got also x i minus x bar what is x i minus x bar in our notation now? It is x i dash, similarly this is y i dash and the summations are available at the end. So, we get sigma of x i dash into y i dash which is available in this column x i dash y i dash which is 1974.67 and, similarly x i dash the whole square. So, summation x i dash the whole square which is available from this column x i minus x bar the whole square. So, from this you get first b, b turns out to be 0.56423 and then from once you get b you get a which is y bar minus b, x bar, which turns out to be minus 23.889 and therefore, y is equal to a plus b x which is 0.564, which is b x minus 22.8295. So, you define the line completely now by using the data that was available with you and thus you are able to say that for a given x my predicted value of y will be as given by that line. So, you specify any value of x you should be able to get the value of y.

Now, these can be used for several occasions; that means, you have observed the values and from the observed values you are converting the observed values into a mathematical

model which is a simple straight line and then you can use this straight line to represent the runoff from that location then you will be able to answer several questions if my rainfall is so, much what will be my runoff and so on, and this can also be used to some extent for data extension that is, if you have 50 years of data then you can extend it to let say 60 years or 70 years using this particular relationship of course there are other issues involved there the issues concerned with outliers and so on will at this point of time we will not worry too much about this. So, in this lecture, we have we started with the method of maximum likelihood we solved one example and then we stated the Chebyshev inequality which places an upper bound on the deviations of a particular random variable from its mean.

Then we went on to define the covariance between two variables and define the correlation coefficient which is in fact, a linear which is in fact, a measure of linear dependence between  $x$  and  $y$  and from that we moved on to simple linear regression where we defined a straight line between the variables  $y$  and  $x$ ,  $y$  is a dependent variable and  $x$  is the independent variable. So, we will continue this discussion in the next lecture, thank you for your attention.