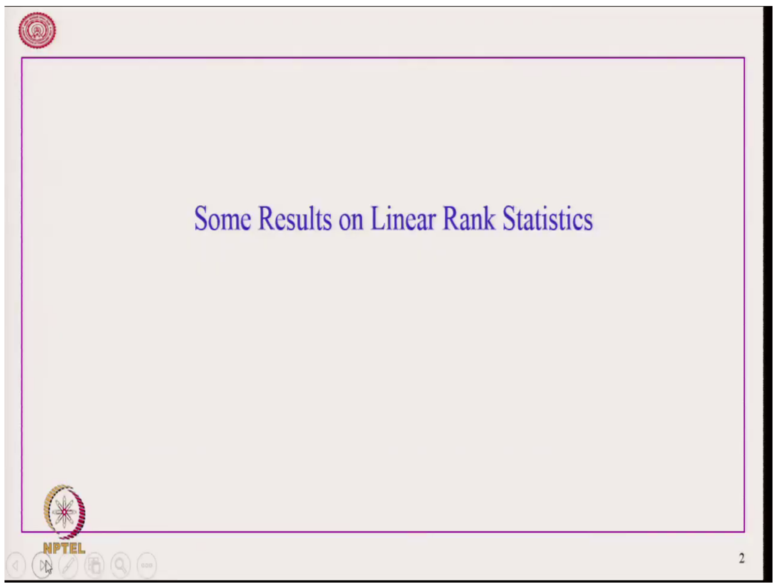
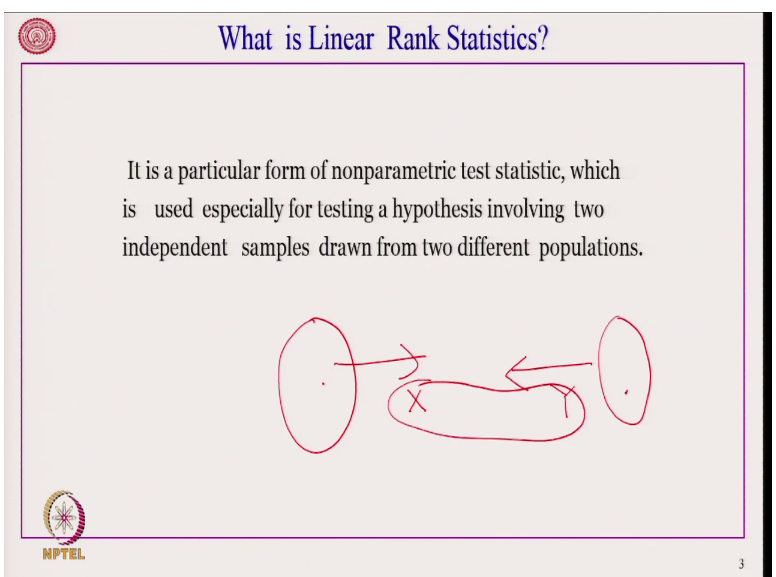
Nonparametric Statistical Inference
Professor Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi
Lecture 5
Nonparametric Statistical Inference

Welcome students to MOOCs series of lectures on nonparametric statistical inference, this is lecture number 5. As I said at the end of the last class, that in this class, we shall study some mathematical results on linear rank statistics.

(Refer Slide Time: 00:43)

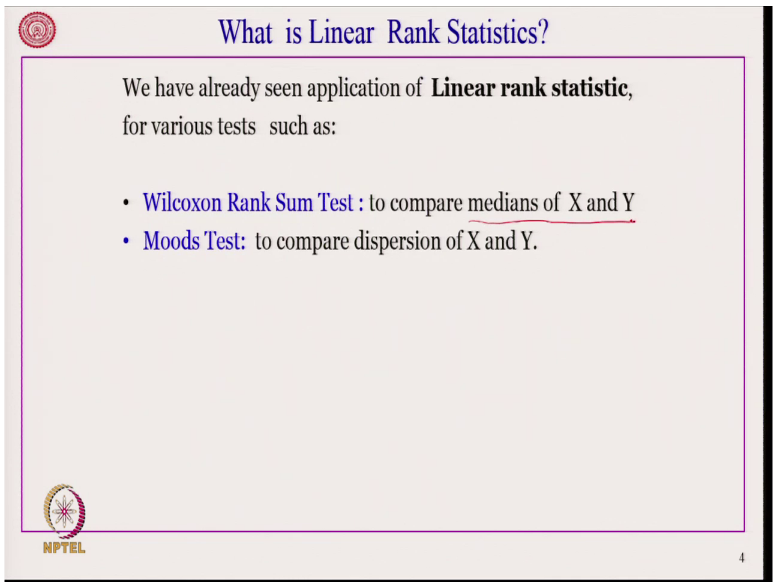




So, what is linear rank statistics? It is a particular form of nonparametric test statistic, which is used especially for testing a hypothesis involving 2 independent samples drawn from 2 different populations. So, consider 1 population from which we are drawing a sample X. There is another population from which you are drawing a sample Y.

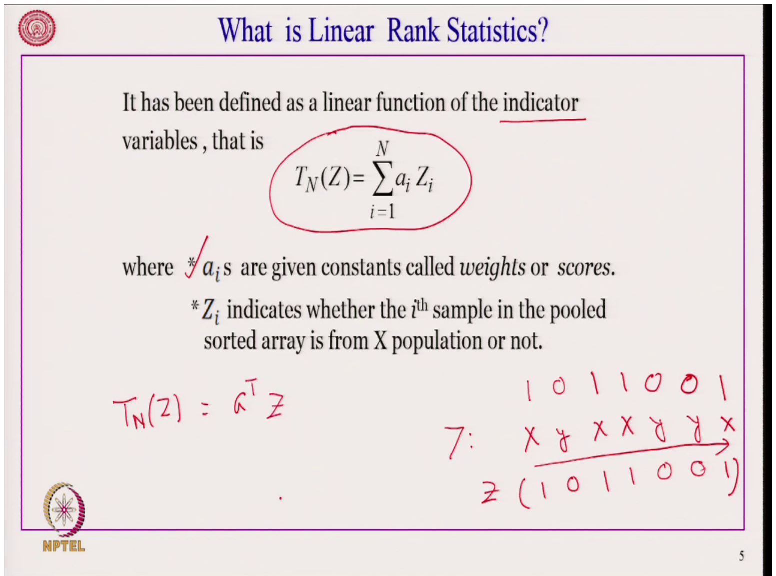
And we want to compare these 2 samples to understand the behavior or the properties of these 2 populations.

(Refer Slide Time: 01:29)



We have already seen some applications of linear rank statistics for various tests. In particular, you remember that we have discussed Wilcoxon Rank Sum Test to compare the medians of X and Y also we have done Moods test, **Freund Ansari test** and David Barton's test, they also have use of linear rank statistic.

(Refer Slide Time: 01:55)

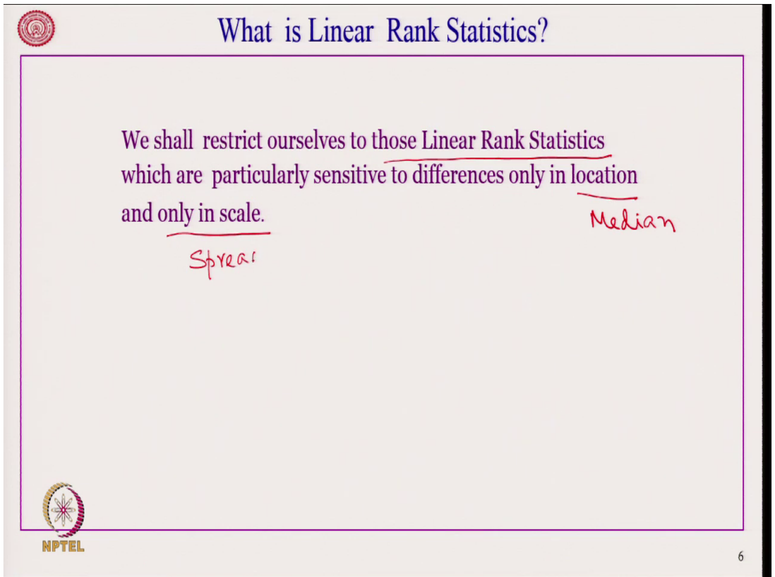


So, if we remember, the linear rank statistic is of the form, $T_N(Z) = \sum_{i=1}^{N} a_i Z_i$

where Zi is an indicator variable. In all our examples, we typically looked at the Zi value is equal to 1 for X and Zi value is equal to 0 for Y. Therefore, if we have 1 0 1 1 0 0 1 that means, there are 7 observations of which 4 are from X and 3 are from Y. And not only that, in the sorted order, they look like this.

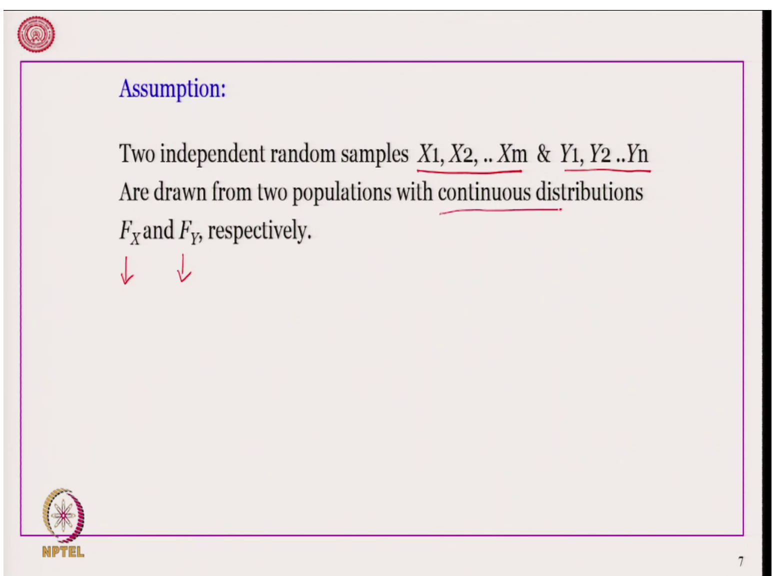
Therefore, Z vector will look like 1 0 1 1 0 0 1. Now, what is ai? ai is a coefficient depending upon the nature of the test we will use different ai's. As we will see later or we have already said therefore, Tn(Z) is equal to the dot product of a vector with Z written as a column.

(Refer Slide Time: 03:30)



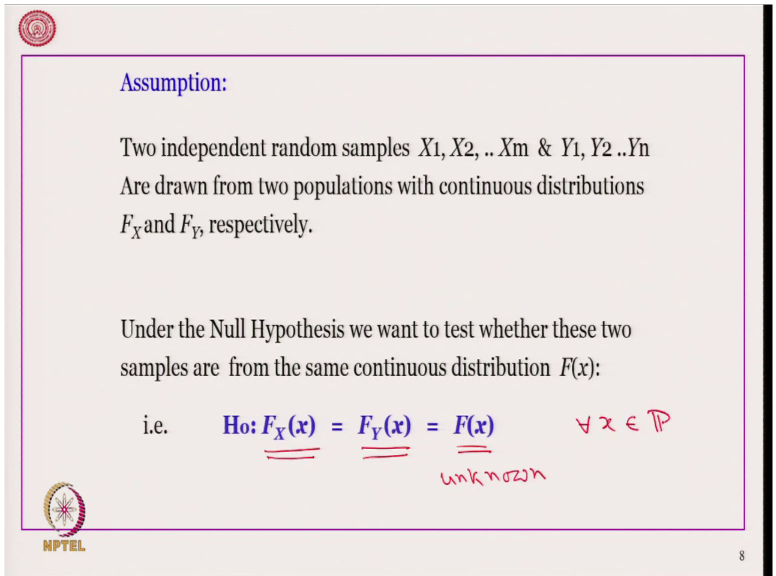
For this course, we shall restrict ourselves only to those linear rank statistic, which are particularly sensitive for differences in location that is median and only in scale that is spread.

(Refer Slide Time: 03:47)



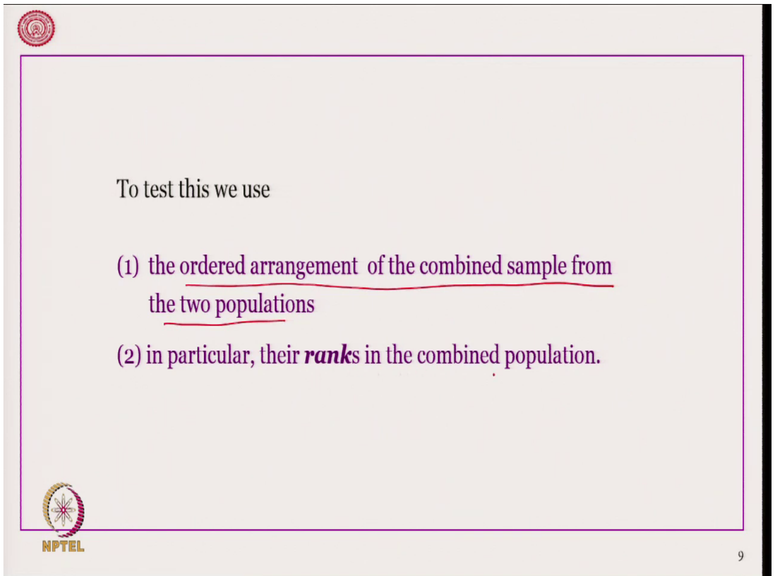
The assumptions is that we have taken 2 independent random samples, X1, X2, Xm from 1 population Y1, Y2, Yn from another population and essentially we want to see that the corresponding cumulative distribution functions that is Fx and Fy are same or not that is what we want to test. So, samples are taken like that from the Fx and Fy distribution only important assumption that they are continuous.

(Refer Slide Time: 04:31)



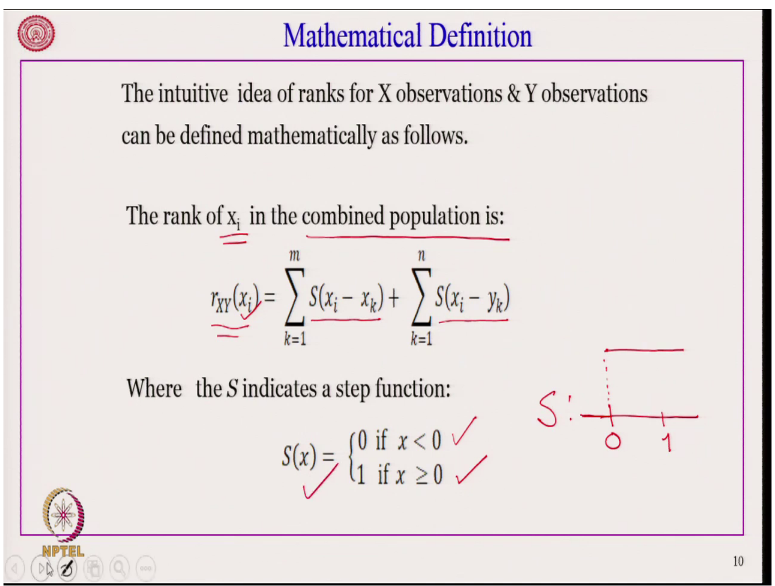
And as I said, under the null hypothesis we want to test whether they have the same distribution that is a Fx(x) is equal to $F_Y(x)$ is equal to F_X, which is unknown for all x belonging to R.

(Refer Slide Time: 04:55)



And to test this, we use the ordered arrangement of the combined sample from the 2 populations in particular, their ranks in that combined population.

(Refer Slide Time: 05:06)



So, mathematically, we can define rank for X observations and Y observations in the following way. The rank of xi, the i th x observation in the combined population is

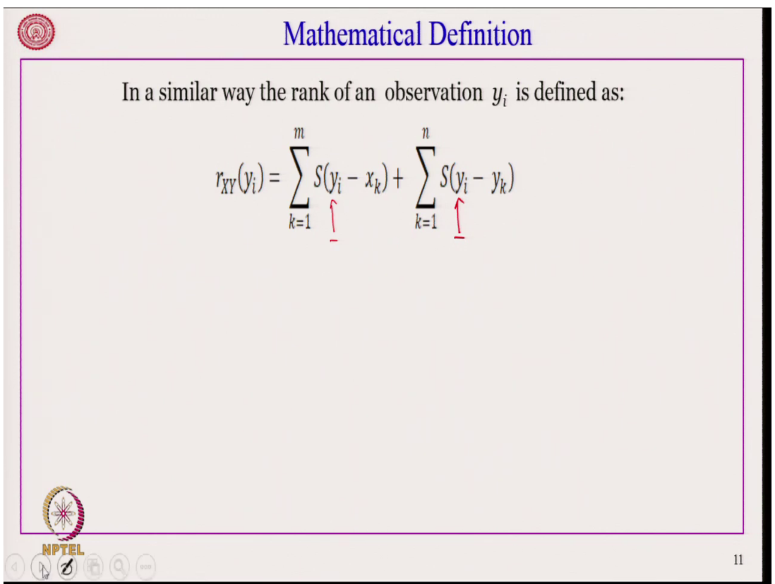
$$r_{XY}(x_i) = \sum_{k=1}^{m} S(x_i - x_k) + \sum_{k=1}^{n} S(x_i - y_k)$$

where S indicates a step function,

$$S(x) = \begin{cases} 0 \text{ if } x < 0 \\ 1 \text{ if } x \geq 0 \end{cases}$$

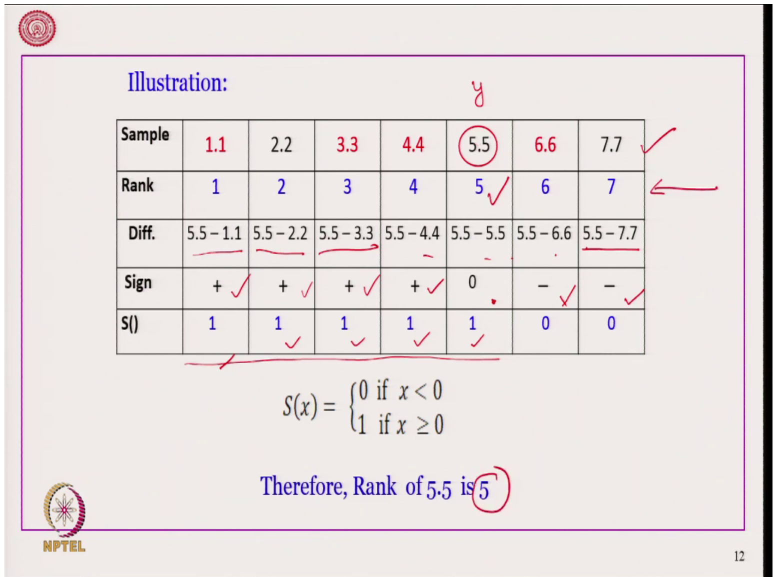
Therefore, S function looks like this it is 0, before 0, from 0 onwards it is 1. So, this is a type of step function that S(x) denotes.

(Refer Slide Time: 06:22)



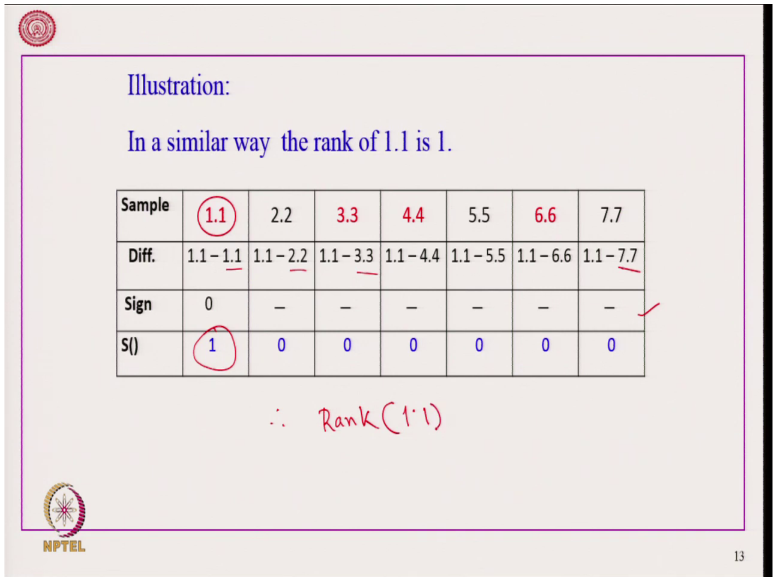
In a similar way, the rank of an Y observation yi can be defined as summation over k is equal to 1 to m S yi minus xk plus k is equal to 1 to n S of yi - yk. So, this is the 1 for which you are trying to find the rank and this we are subtracting all X values and Y values and based on that difference, we are trying to calculate the rank.

(Refer Slide Time: 06:58)



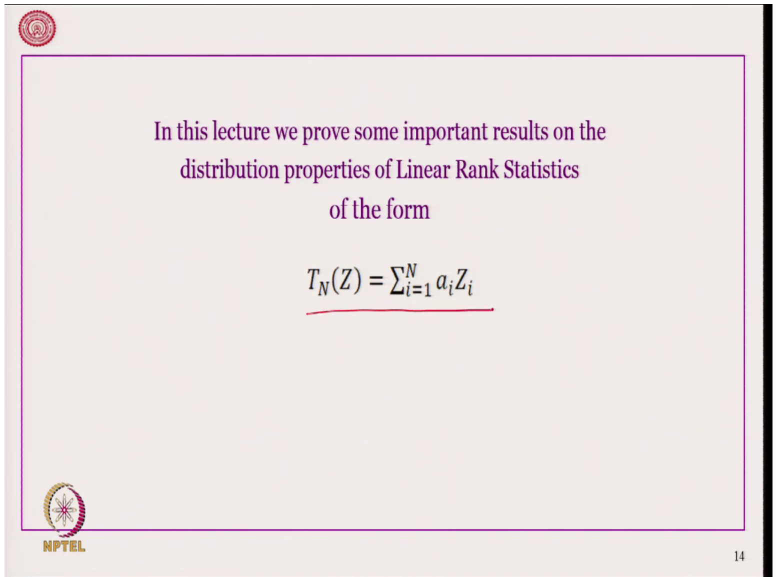
Let me illustrate with an example. Suppose my combined sample is 1.1, 2.2 3.3 like up to 7.7; what is the rank of 5.5? So, it is an Y observation let us assume, that blacks are Y's and X's are in red. So, its rank as we can easily see is 5, how to do that. So, you are subtracting all the observations from 5.5. Therefore, 5.5 minus 1.1, 5.5 minus 2.2, 5.5 minus 3.3, up to 5.5 minus 7.7. Quite naturally, these are all plus and this is 0 and these are minus. Therefore, S values is 1, 1, 1, 1, 1 and 0 here. So, how many ones are there? There are 5 ones. Therefore, rank of 5.5 is equal to 5 that in any case, we have tabulated here.

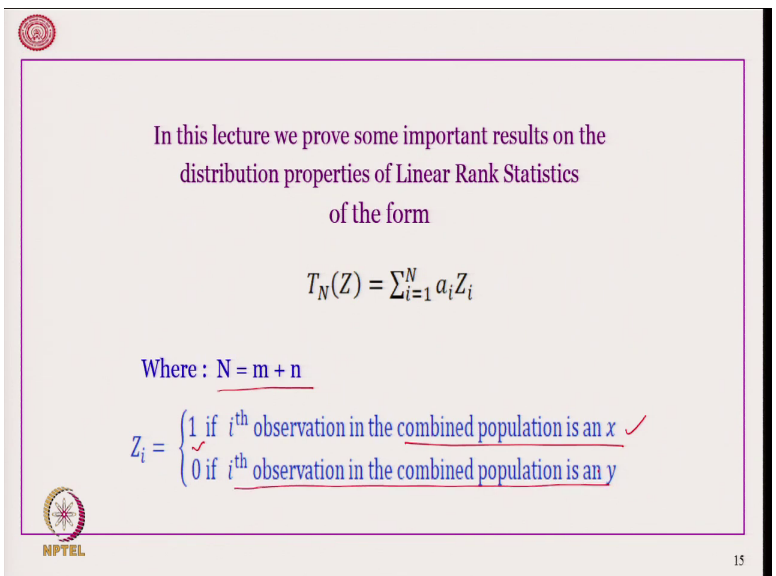
(Refer Slide Time: 08:18)



Another example what is going to be the rank of 1.1? So, what we are doing from 1.1 we are subtracting all the observations and all of them are minus except 0. So, in the S function, we get only 1, 1 and remaining all zeros. Therefore, rank of 1.1 is equal to 1.

(Refer Slide Time: 08:46)



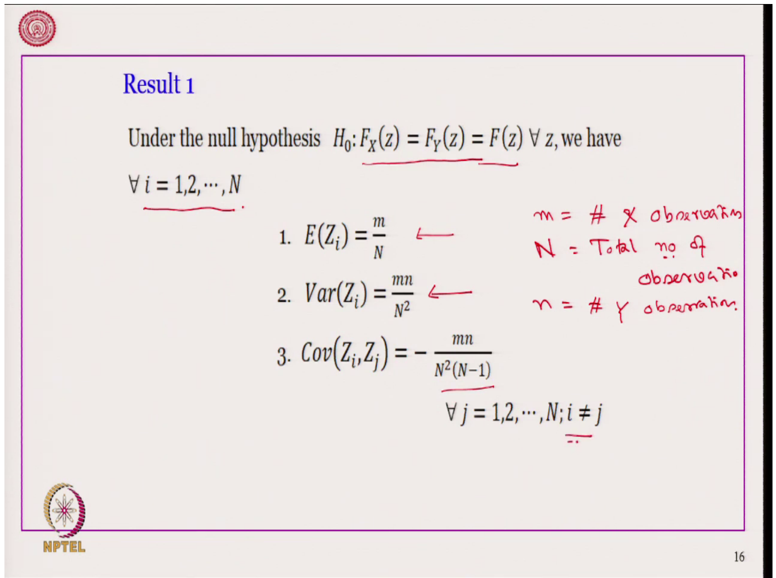


In this lecture, we prove some important results on the distribution properties of the linear rank statistic of the form

$$T_N(Z) = \sum_{i=1}^{N} a_i Z_i$$

As I said earlier that we will be focusing on X observations. Therefore, Zi is equal to 1 if i th observation in the combined population is an X and it is 0. If i th observation in the combined population is an y.

(Refer Slide Time: 09:18)



So, result 1 under the null hypothesis that is, both of them are from the same distribution,

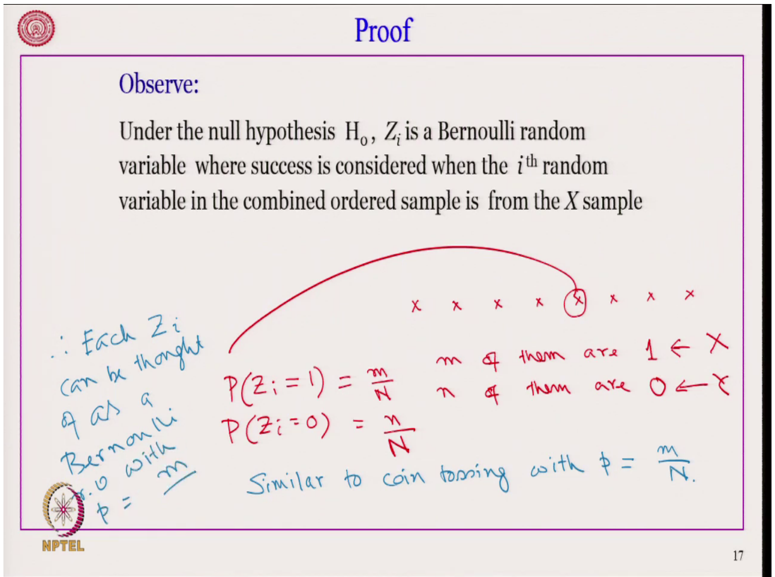
$$F_X(z) = F_Y(z) = F(z)\ \forall\ z, \quad \forall\ i = 1,2,\cdots,N$$

Expected value of Zi is equal to m by N, m is equal to number of X observations and N is total number of observations. Variance of Zi is equal to mn upon N square, where n is equal to number of y observations and covariance of Zi and Zj is equal to minus mn upon N square into N minus 1 when j is different from i.

1. $E(Z_i) = \frac{m}{N}$
2. $Var(Z_i) = \frac{mn}{N^2}$
3. $Cov(Z_i, Z_j) = -\frac{mn}{N^2(N-1)}$

$$\forall\ j = 1,2,\cdots,N; i \neq j$$

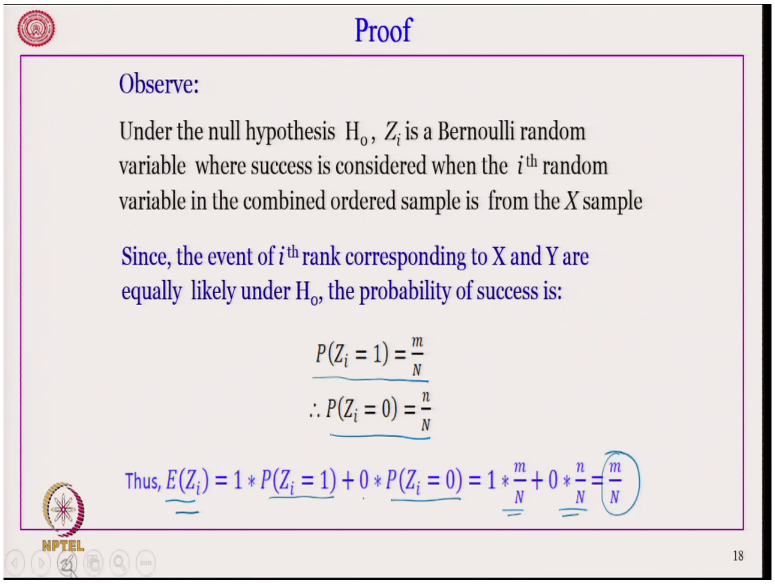
(Refer Slide Time: 10:35)



Under the null hypothesis Zi is a Bernoulli distribution or Bernoulli random variable where success is considered when the ith random variable in the combined ordered sample is from X. I hope you understand it, but let us assume that these are the n ordered observations, out of which m of them are 1, and n of them are 0.

Because these are coming from X and these are coming from Y. Therefore, when we randomly pick up 1 value, probability Zi is equal to 1 is equal to m by N and probability Zi is equal to 0 is equal to n by capital N. Therefore, it is similar to a coin tossing with p is equal to, m by N. Therefore, each Zi can be thought of, as a Bernoulli random variable with p is equal to, m by N.

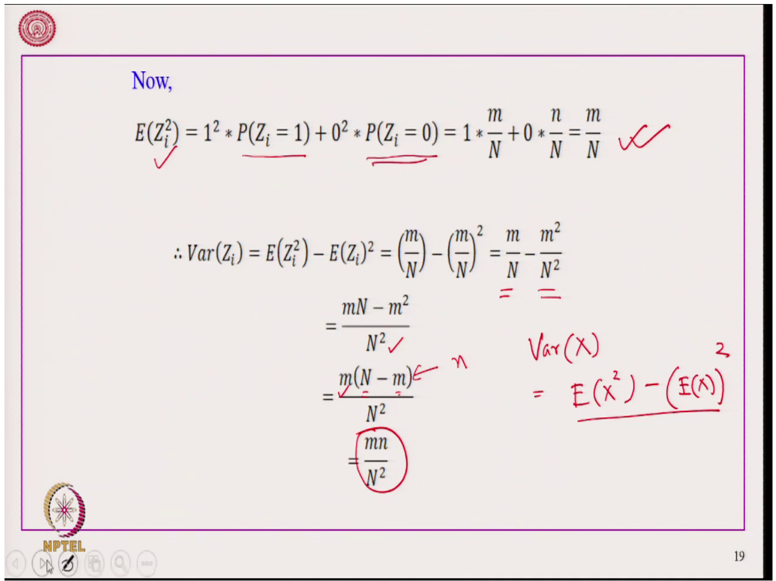
(Refer Slide Time: 12:46)



This is what we have written here. That probability Zi is equal to 1 is equal to m by N, Zi is equal to 0 is equal to n by N. Therefore, the expected value of Zi is equal to

$$1 * P(Z_i = 1) + 0 * P(Z_i = 0) = 1 * \frac{m}{N} + 0 * \frac{n}{N} = \frac{m}{N}$$

(Refer Slide Time: 13:14)



Now variance, we know that variance of a random variable X is equal to the expected value of X square minus expected value of X whole square. So, let us first calculate expected value of Zi square, this is equal to

$$E(Z_i^2) = 1^2 * P(Z_i = 1) + 0^2 * P(Z_i = 0) = 1 * \frac{m}{N} + 0 * \frac{n}{N} = \frac{m}{N}$$

Therefore, variance of Zi is equal to

$$E(Z_i^2) - E(Z_i)^2 = \left(\frac{m}{N}\right) - \left(\frac{m}{N}\right)^2 = \frac{m}{N} - \frac{m^2}{N^2}$$

$$= \frac{mN - m^2}{N^2}$$

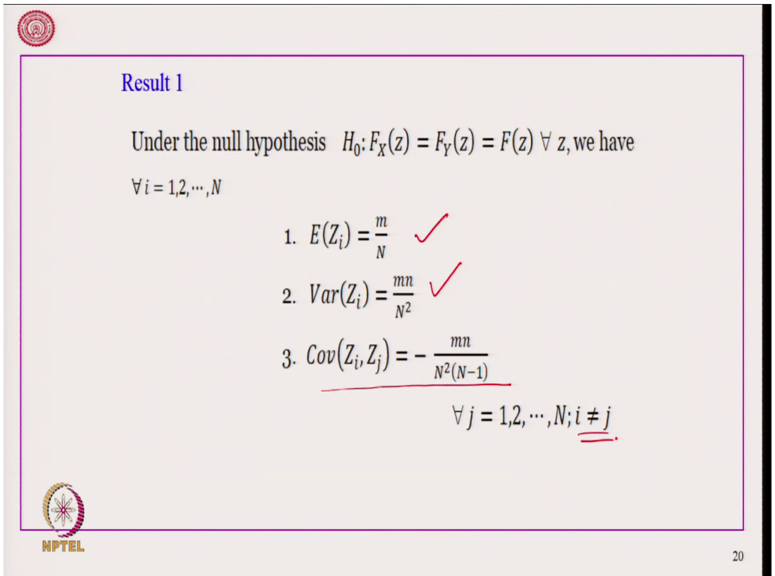
If we take m common, it is

$$\frac{m(N-m)}{N^2}$$

$$= \frac{mn}{N^2}$$

Therefore, mn upon capital N square is the variance of each Zi.

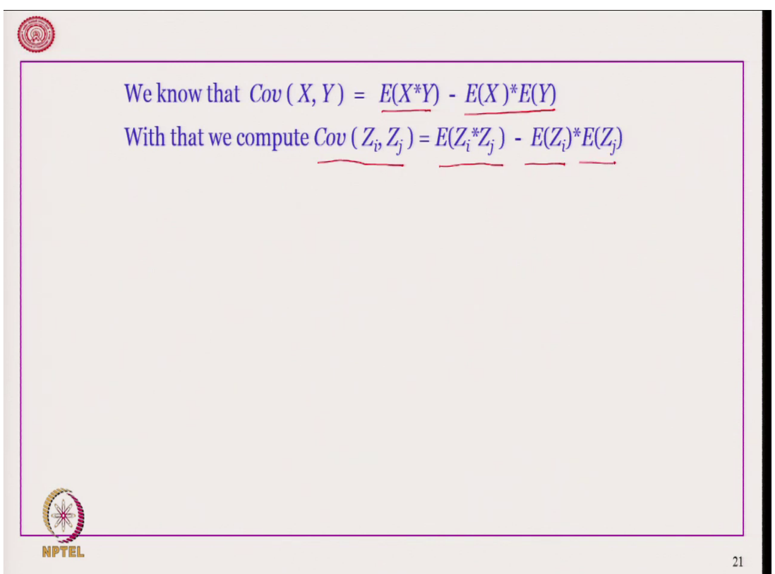
(Refer Slide Time: 14:26)



So, these 2 results have been proved. Now, we need to prove that covariance Zi, Zj is equal to

$$-\frac{mn}{N^2(N-1)}$$

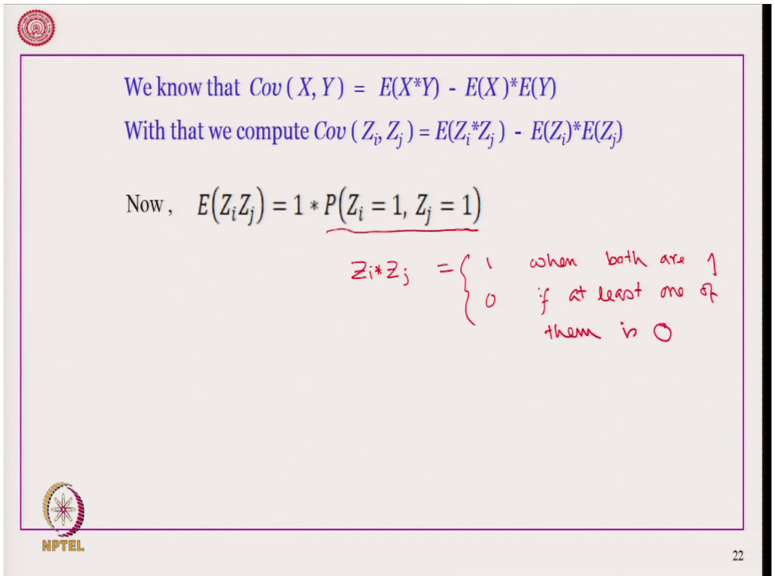
(Refer Slide Time: 14:43)



We know that covariance of X Y is equal to expected value of X Y minus the expected value of X times the expected value of Y. Therefore, when we are trying to compute covariance of Zi and

Zj this is going to be expected value of Zi Zj minus expected value of Zi times the expected value of Zj.

(Refer Slide Time: 15:11)

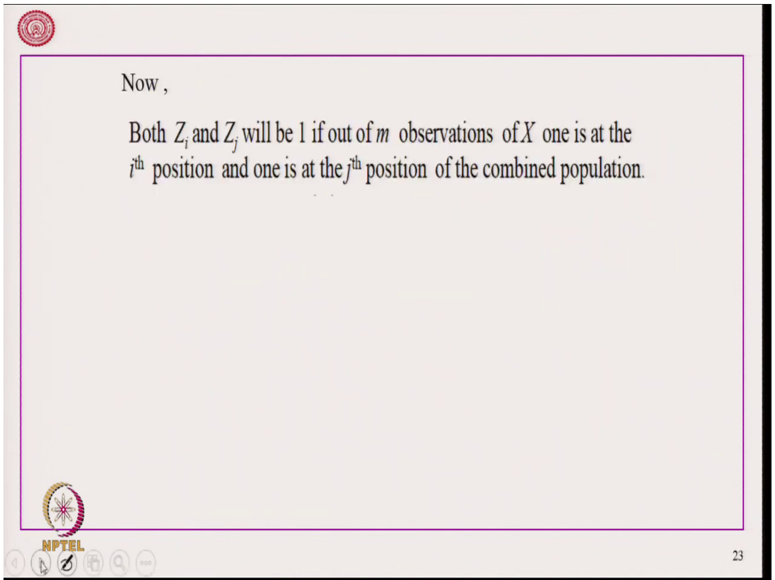


Now expected value of Zi Zj is equal to

$$1 * P(Z_i = 1, Z_j = 1)$$

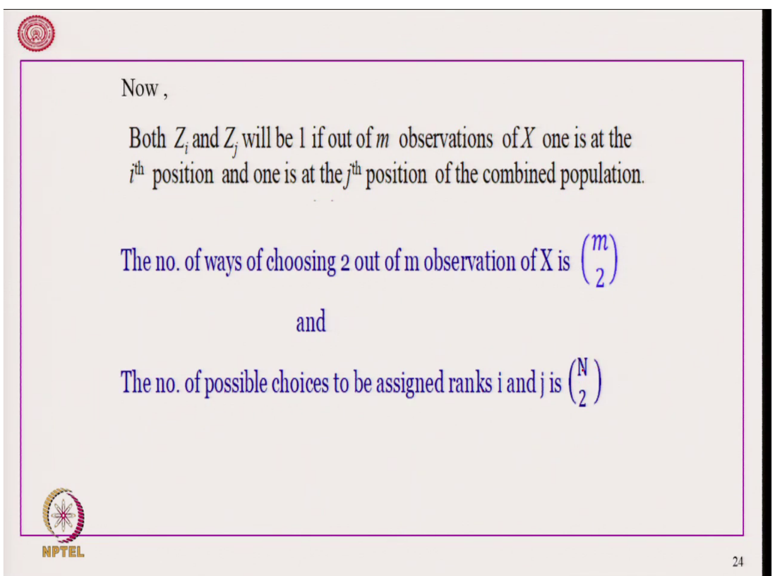
Why? Because Zi Zj, the product is going to be 1 when both are 1 and this is going to be 0 if at least 1 of them is 0.

(Refer Slide Time: 15:55)



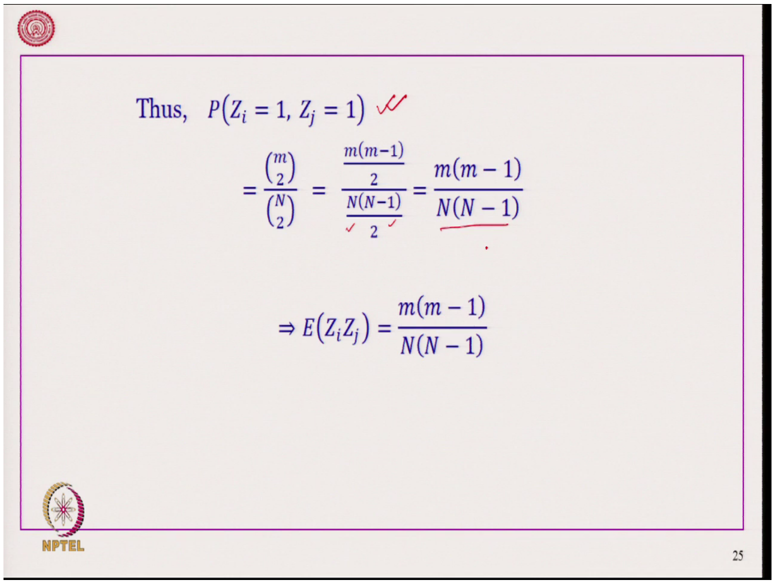
Now, both Zi and Zj will be 1 if out of the m of observations X 1 is at the i th position and 1 is at the j th position in the combined population.

(Refer Slide Time: 16:09)



Therefore, the number of ways of choosing 2 out of m observations of X is ${}^{m}C_2$ and the total number of possible choices is capital ${}^{N}C_2$.

(Refer Slide Time: 16:24)



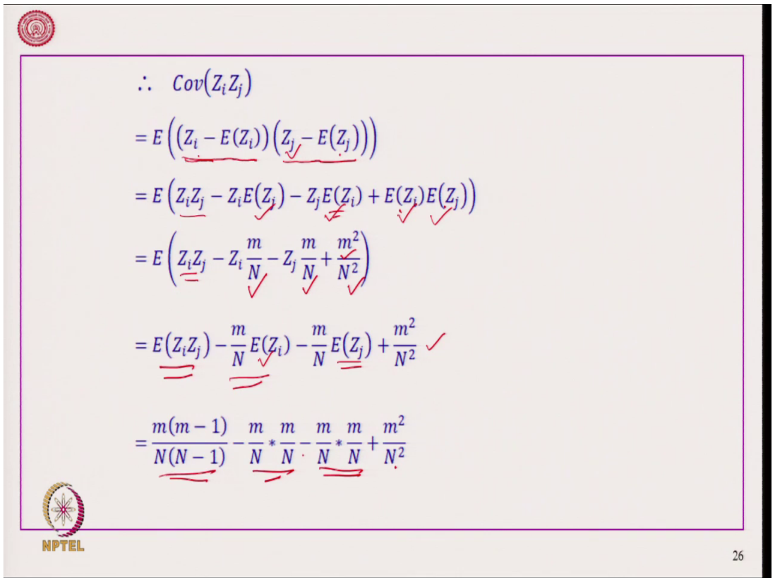
Hence, probabilities Zi is equal to 1 and Zj is equal to 1 is

$$\frac{\binom{m}{2}}{\binom{N}{2}} = \frac{\frac{m(m-1)}{2}}{\frac{N(N-1)}{2}} = \frac{m(m-1)}{N(N-1)}$$

Therefore, the expected value of Zi is Zj is equal to

$$\frac{m(m-1)}{N(N-1)}$$

(Refer Slide Time: 16:53)



Therefore, covariance of Zi Zj which is defined as

$$E\left(\left(Z_i - E(Z_i)\right)\left(Z_j - E(Z_j)\right)\right)$$

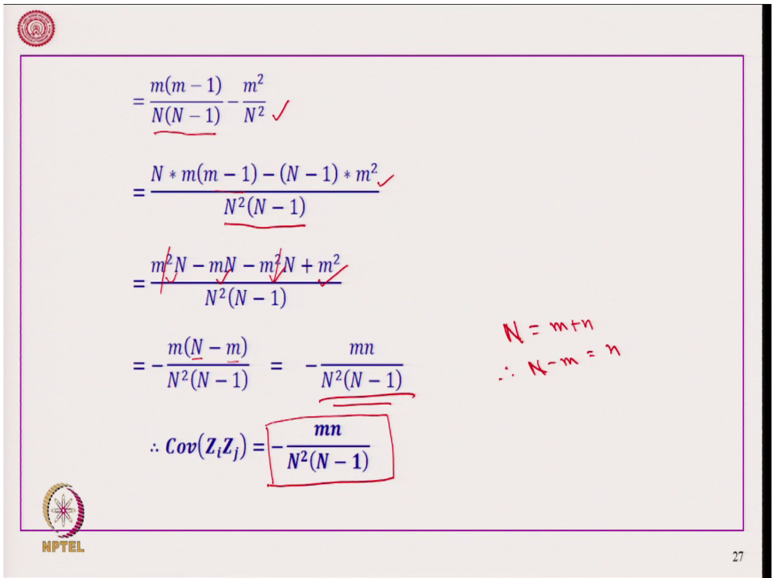
$$E\left(Z_i Z_j - Z_i E(Z_j) - Z_j E(Z_i) + E(Z_i)E(Z_j)\right)$$

Which comes out to be expected value of Zi is Zj minus Zi into m upon capital N, because for each Z the expected value is small m upon capital N. Similarly, expected value of this is also m upon capital N and both of them are giving me small m square upon capital N square. Now, by linearity of expectation with this is equal to

$$E(Z_i Z_j) - \frac{m}{N} E(Z_i) - \frac{m}{N} E(Z_j) + \frac{m^2}{N^2}$$

This we have already calculated. So, this is coming out to be m square upon capital N square with a minus sign plus m square upon capital N square.

(Refer Slide Time: 18:35)



Which is equal to

$$\frac{m(m-1)}{N(N-1)} - \frac{m^2}{N^2}$$

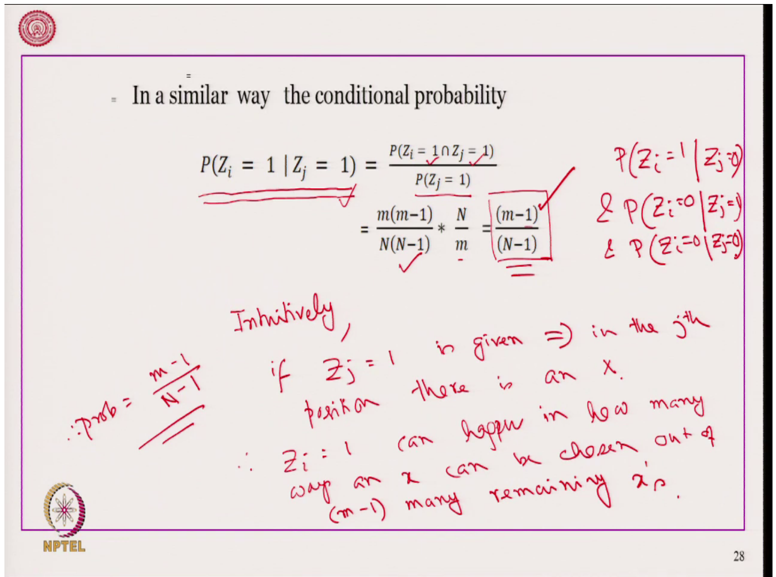
Let us now expand them. Therefore, we get in the denominator $N^2(N-1)$

numerator is coming out to be

$$N * m(m-1) - (N-1) * m^2$$

So, these 2 cancel out, if I take minus common, we are left with capital N minus small m and capital N is equal to m plus n. Therefore, N minus m is equal to small n and therefore, we get the result. So, this is proved.

(Refer Slide Time: 19:47)



In a similar way, let us prove another result, which is from conditional probability. What is the conditional probability that is Zi is equal to 1 given Zj is equal to 1. We know that from the definition of conditional probability probability Zi is equal to 1 given Zj is equal to 1 is equal to

$$\frac{P(Z_i = 1 \cap Z_j = 1)}{P(Z_j = 1)}$$

which is coming out to be,

$$\frac{m(m-1)}{N(N-1)} * \frac{N}{m}$$

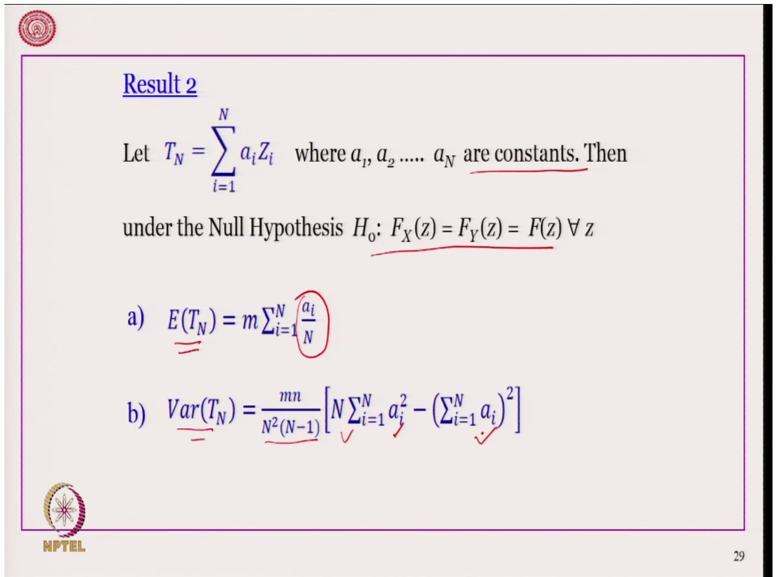
which is coming out to be

$$\frac{(m-1)}{(N-1)}$$

Intuitively it is clear. If Zj is equal to 1 is given implies that in the j th position there is an X. Therefore, Zi is equal to 1 can happen in how many ways an X can be chosen out of small m minus 1 many remaining X's. Thus, we get a minus 1 in the numerator in the denominator. It is very natural that out of the remaining capital N minus 1 observation anyone can come there. Therefore, the probability is going to be small m minus 1 upon capital N minus 1.

Try to prove the results, what is the probability Zi is equal to 1 given Zj is equal to 0 and probability Zi is equal to 0 given Zj is equal to 1 and probability Zi is equal to 0 given Zj is equal to 0. I will leave these as an exercise for you. This is simple combinatorial problems, but you will be able to do it.

(Refer Slide Time: 22:55)



Now, let us go to the result 2 . Let Tn is equal to

$$\sum_{i=1}^{N} a_i Z_i$$

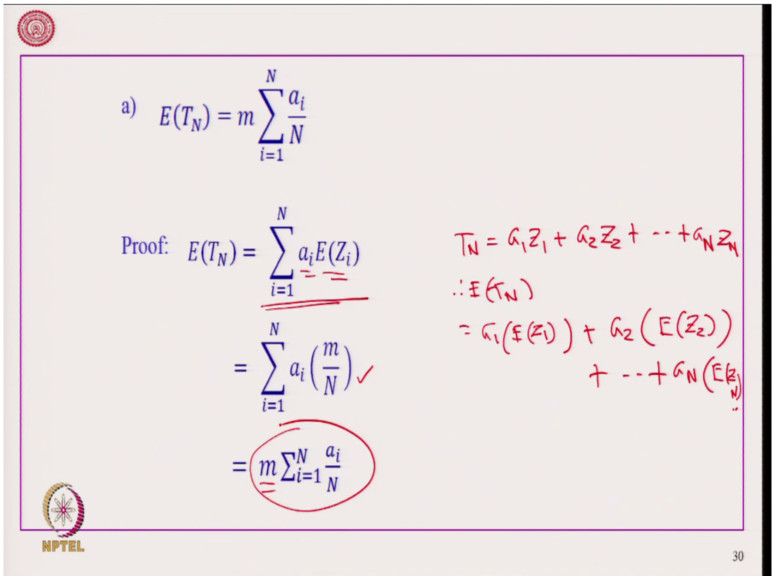
where a1, a2 an our constants. Then under the null hypothesis expected value of Tn is equal to

$$m \sum_{i=1}^{N} \frac{a_i}{N}$$

and variance of Tn is equal to

$$\frac{mn}{N^2(N-1)} \left[N \sum_{i=1}^{N} a_i^2 - \left(\sum_{i=1}^{N} a_i \right)^2 \right]$$

(Refer Slide Time: 23:31)



So, let us prove it the expected value of Tn is equal to

$$m\sum_{i=1}^{N}\frac{a_i}{N}$$

It is very straightforward, because Tn is equal to a1 times Z1 plus a2 times Z2 plus aN times ZN. Therefore, expected value of Tn is equal to

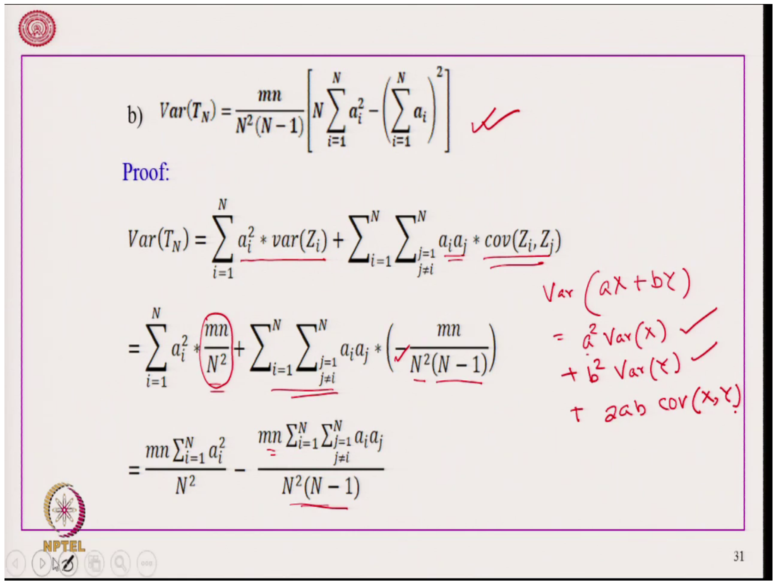
$$\sum_{i=1}^{N} a_i E(Z_i)$$

So, this is what we have written is equal to

$$\sum_{i=1}^{N} a_i\left(\frac{m}{N}\right)$$

If we take m out of the summation, we get the result. It is a very straightforward thing, but what about variance?

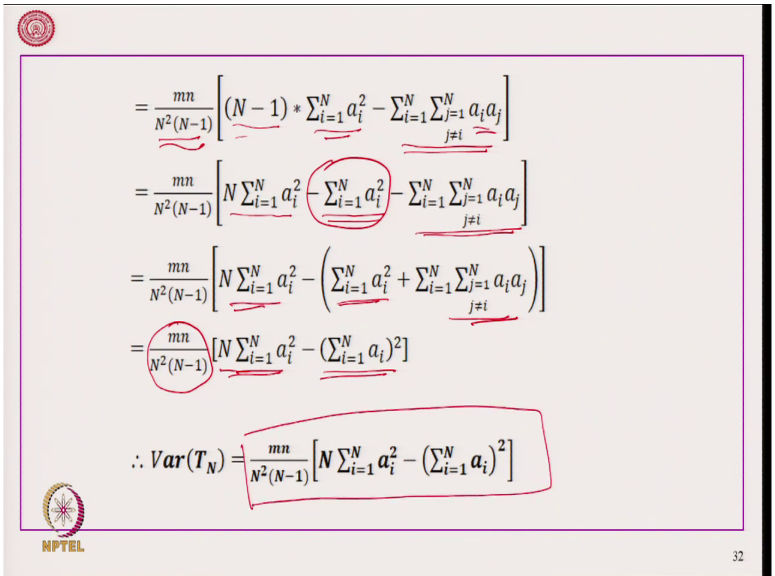
(Refer Slide Time: 24:44)



How do we show that variance of Tn is equal to this complicated quantity. So, let us start we know that variance of aX plus bY is equal to a square times variance of X plus b square times variance of Y plus 2ab covariance of X and Y. When we generalize it, we get variance of Tn is equal to ai square times variance of a Zi, i is equal to 1 to n plus, we are looking at all pairs and trying to find the covariance of Zi and Zj, this is equal to sigma ai square multiplied by mn upon N square.

This we have already found that variance of Zi is equal to this plus the double summation ai aj multiplied by minus mn upon N square into N minus 1, this we have already found out. Therefore, this is coming out to be a mn upon N square sigma ai square minus, mn comes out of the summation, therefore, we get double summation i and j, i not equal to j, ai aj divided by N square into N minus 1.

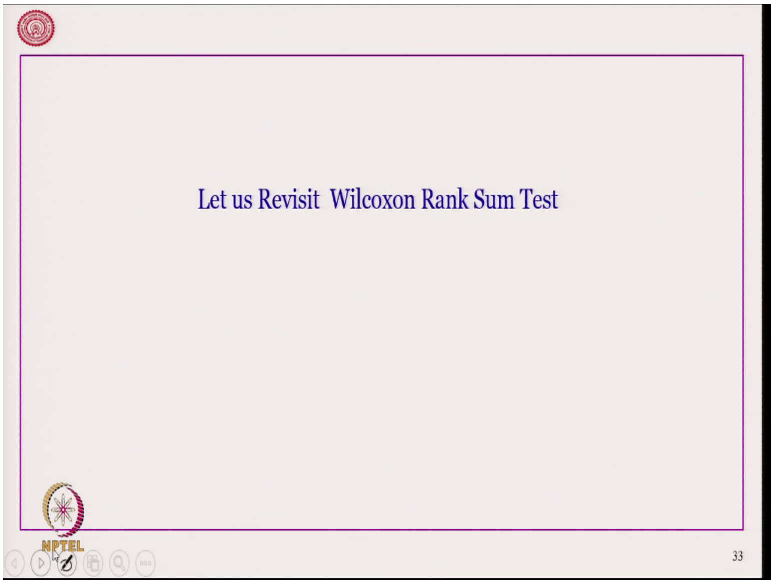
(Refer Slide Time: 26:36)

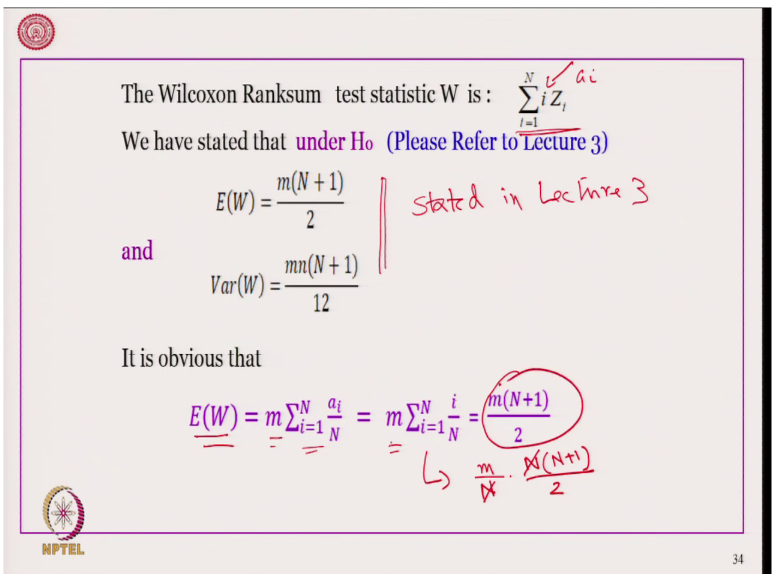


Now, if we take a mn upon N square into N minus 1 common. We have capital N minus 1 multiplied by sigma ai square minus this summation double summation i is equal to 1 to N, j is equal to 1 to N, ai aj, j not equal to i. Now, we break it as N into sigma ai square minus sigma ai square minus this double summation. Now, we put this inside the summation. Therefore, what we are getting N time sigma ai square minus sigma ai square and sigma sigma ai aj.

This is nothing but sigma over ai whole square. Therefore, the whole thing boils down to mn, N into N square minus 1 multiplied by N into sigma ai square minus sigma whole square and thus we get the result for variance.

(Refer Slide Time: 27:48)





Now, let us revisit Wilcoxon Rank Sum Test. We have found out that in the Wilcoxon Rank Sum Test, this statistic is W, which is the sum of ranks of the X observations and therefore, we have written it as sigma i Zi. Therefore, the expected value of W is equal to

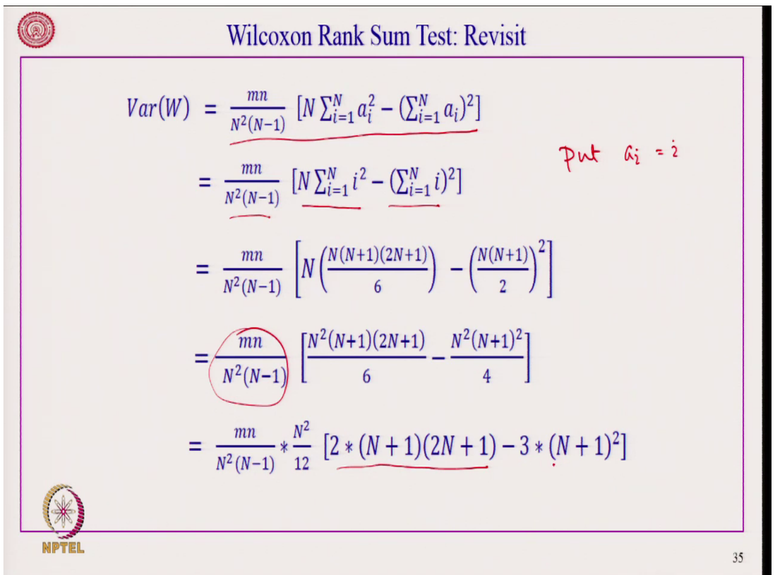
$$\frac{m(N+1)}{2}$$

and variance of W is equal to

$$\frac{mn(N+1)}{12}$$

This we have stated in lecture 3. Now, let us prove it the expected value of W is equal to m times sigma over i is equal to 1 to N, ai upon N. This we have just found out now replace ai with i because that is the value of ai in this summation that is ai. Therefore, we can write it as m times sigma over i 1 to N, i upon capital N and this is, is equal to. If I write, 1 step further m upon capital N into N into N plus 1 by 2, N gets cancelled and therefore, we get the result.

(Refer Slide Time: 29:17)

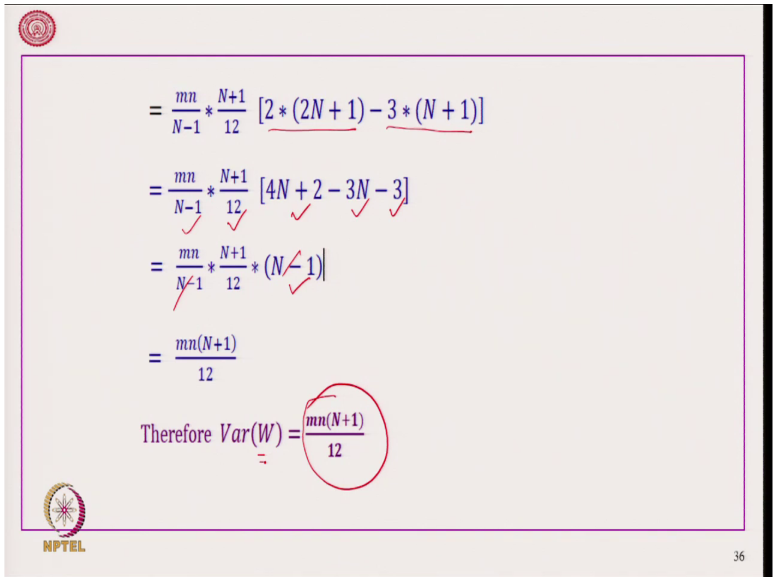


In a similar way, what is going to be the variance. Now, variance of W is equal to this term that we have already calculated. Now, let us put the value of ai. Therefore, this boils down to

$$\frac{mn}{N^2(N-1)} [N \sum_{i=1}^{N} a_i^2 - (\sum_{i=1}^{N} a_i)^2]$$

Now, summation over first N natural integer square is equal to an into n plus 1 into 2n plus 1 by 6 minus N into N plus 1 by 2 whole square multiplied by this quantity. So, let us take out N square upon 12 from this. Therefore, we are left with 2 into N plus 1 into 2 N plus 1 minus 3 into N plus 1 whole square.

(Refer Slide Time: 30:22)



Which is equal to

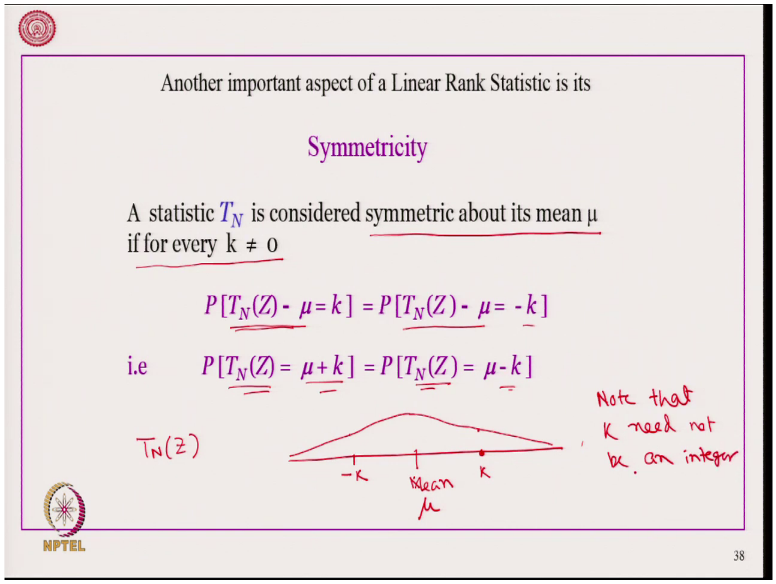
$$\frac{mn}{N-1} * \frac{N+1}{12} \; [4N + 2 - 3N - 3]$$

is equal to

$$\frac{mn}{N-1} * \frac{N+1}{12} \; [4N + 2 - 3N - 3]$$

This comes out to be N minus 1 which cancels with this. Therefore, we are left with mn into capital N plus 1 divided by 12. So, that is the variance of the Wilcoxon Rank Sum Test statistic W, we have used the result, but we proved it today.

(Refer Slide Time: 31:09)



Now, another important aspect of linear rank statistic is symmetricity that means, whether it is symmetric around this mean value. That means, a statistic Tn is considered symmetric about is mean mu. If for every k not equal to 0,

$$P[T_N(Z) - \mu = k] = P[T_N(Z) - \mu = -k]$$

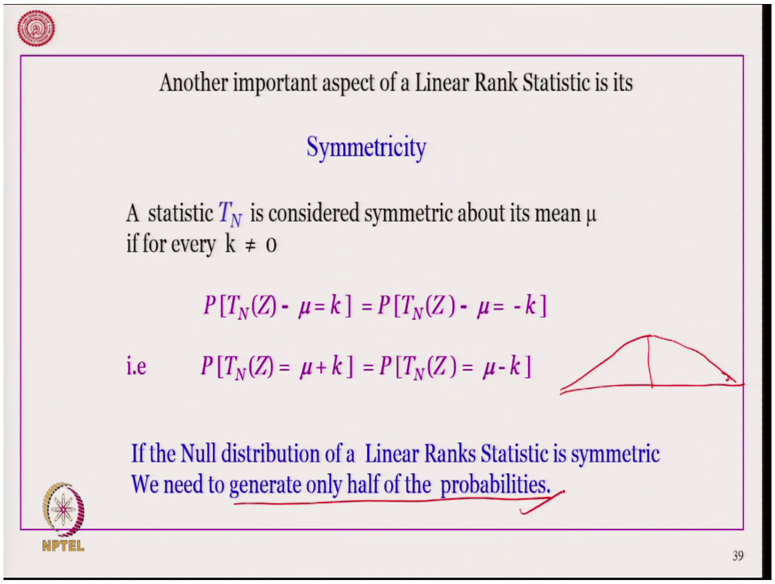
Or in other words,

$$P[T_N(Z) = \mu + k] = P[T_N(Z) = \mu - k]$$

So, that is what we call it symmetric. For illustration consider Tn Z. So, suppose this is the distribution, this is the mean value of Tn Z which we are calling mu.

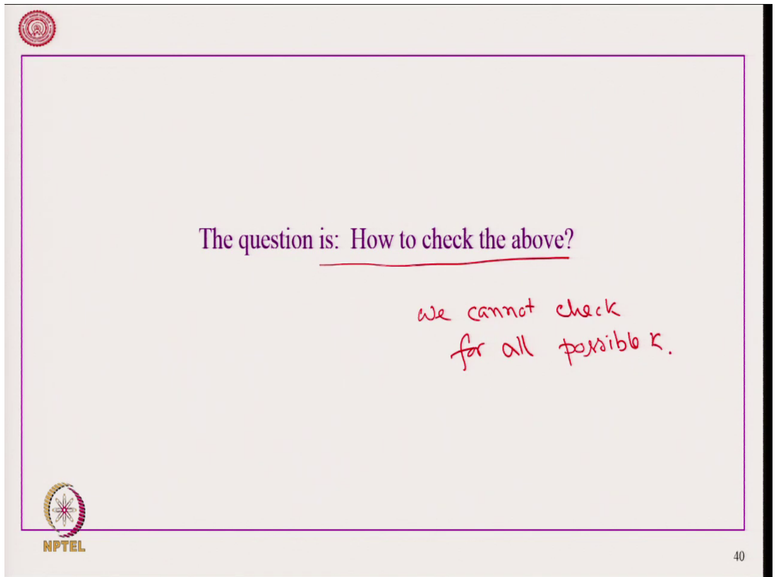
It is considered to be symmetric. If the value at a distance k from mu is same as the value at a distance minus k. That is what we are saying here. Note that k need not be an integer. It can be any real value.

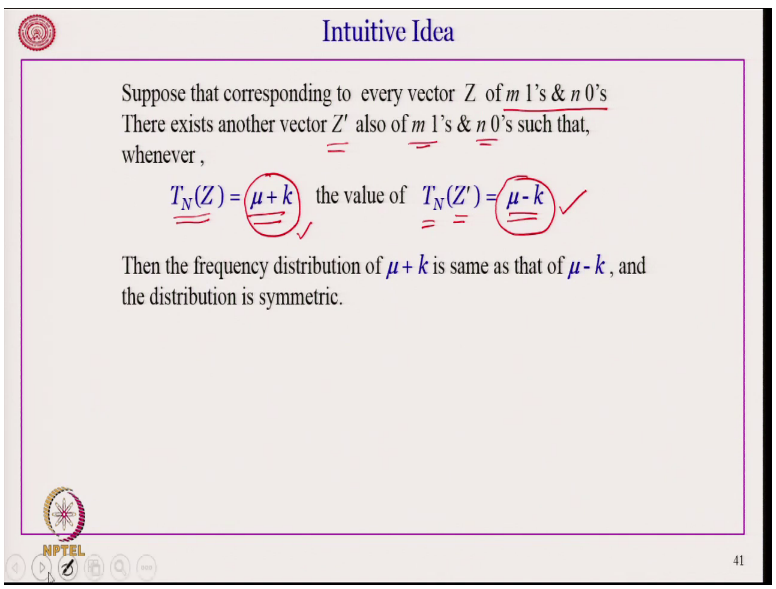
(Refer Slide Time: 32:50)



If the null distribution of linear rank statistic is symmetric, the advantage is that we need to generate only half of the probabilities as you can easily understand if we know the probability density for this part, we know what is the density for the other half. So, that is why symmetric distribution is very useful many times.

(Refer Slide Time: 33:20)

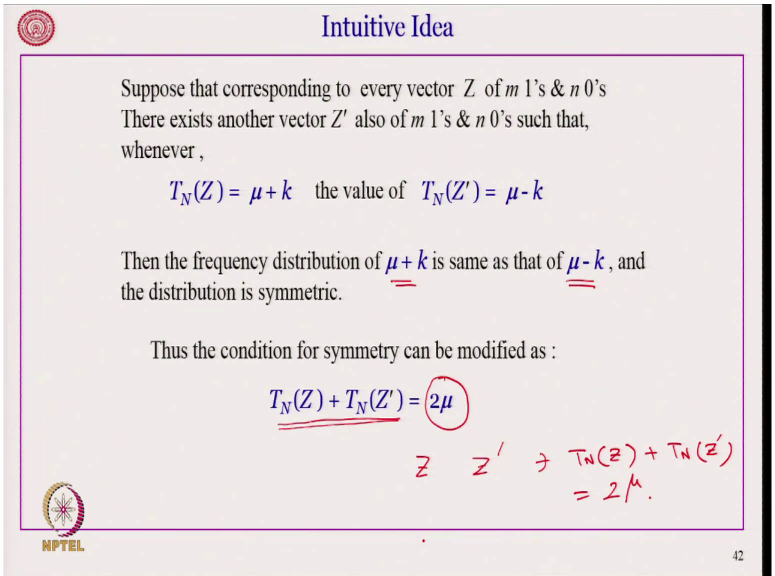


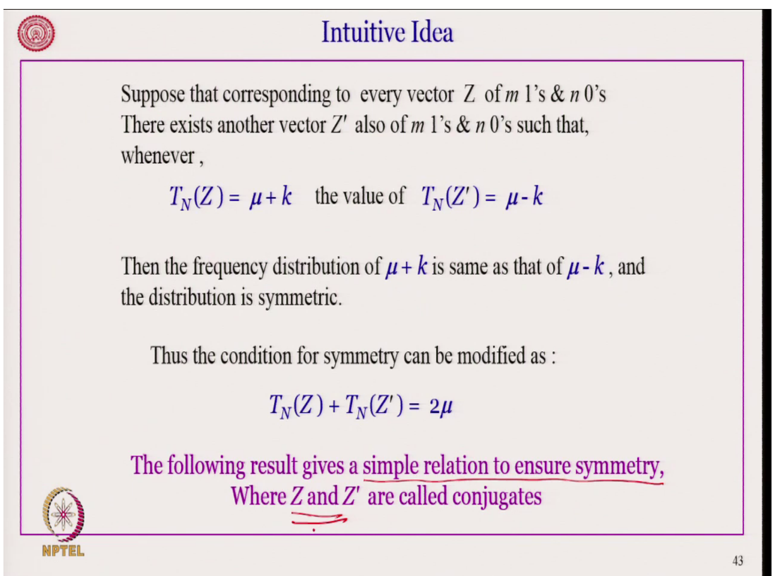


The question is how to check the above? How do you know whether a distribution is symmetric? We cannot check for all possible k therefore; we need to find out device. So, the intuitive idea is following suppose that corresponding to every vectors Z of m 1's and n 0's, there exists another vectors Z prime also of m 1's and n 0's such that Tn Z is equal to mu plus k.

Whenever that happens that for Z, we get the value of Tn Z is equal to mu plus k, then for Z prime, the value of t and Z prime is going to be mu minus k that means, what for each case of mu plus k, there is a corresponding case of mu minus k. Therefore, the number of cases for this is same as the number of cases for this and therefore, their probabilities are also going to be the same.

(Refer Slide Time: 34:45)



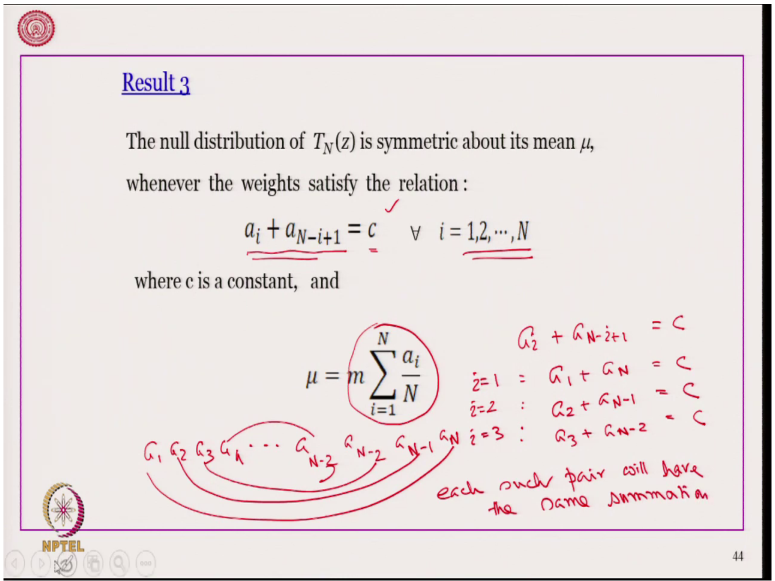


That is what we have written there. Then the frequency distribution of mu plus k is same as that of mu minus k and therefore, the distribution is going to be symmetric. Thus, condition for symmetric can be modified as

$$T_N(Z) + T_N(Z') = 2\mu$$

So, whenever we get, for each Z another Z prime, such that Tn Z plus Tn Z prime is equal to twice mu. And that happens for all Z and therefore, we can say that the distribution of Tn is symmetric. Now, there are several results available for that. So, we are stating some of the results, which are simple relations which ensure symmetry when Z and Z prime are called conjugates.

(Refer Slide Time: 35:50)



So, let us go to prove the result number 3, the null distribution of Tn Z is symmetric about its mean mu whenever the weight vector ai satisfies the following that

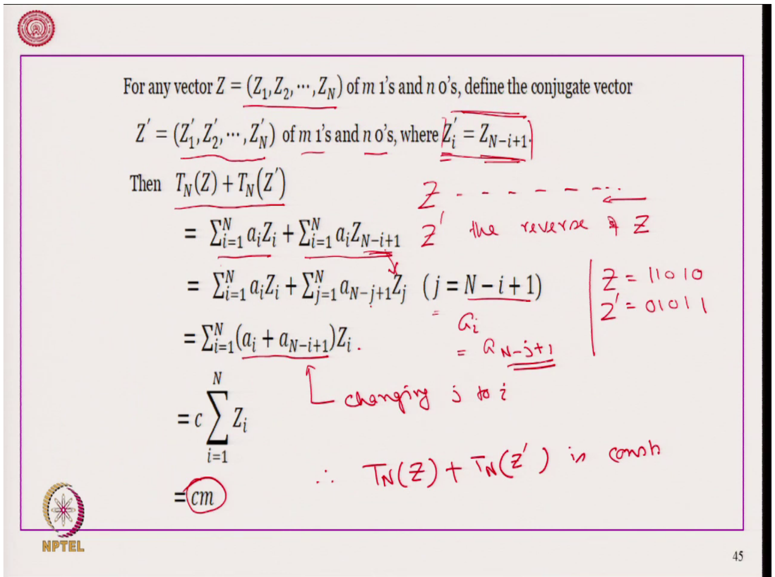
$$a_i + a_{N-i+1} = c \qquad \forall \quad i = 1,2,\cdots,N$$

where c is a constant and mu is equal to as you already know is

$$m\sum_{i=1}^{N}\frac{a_i}{N}$$

So, let us look at it very carefully, it is saying ai plus a an minus i plus 1 is equal to c. So, put i is equal to 1. So, therefore, it is saying a1 plus aN is equal to c put i is equal to 2 it is saying a2 plus aN minus 1 is equal to c, i is equal to 3 implies a3 plus aN minus 2 is equal to c. Therefore, what we are saying is that if a1, a2, a3, a4, aN minus 3, aN minus 2, aN minus 1 and aN if this is the vector then this plus this, this plus this, this plus this, this plus this, each such pair will have the same summation.

(Refer Slide Time: 37:56)



Let us first prove the result. So, corresponding to each Z of m 1's and n 0's we define the conjugate vectors Z1 prime, Z2 prime, Zn prime also have m 1's and n 0's, where Zi prime is equal to Z minus i plus 1. Therefore, if this is the Z vector, Z prime is going to be the reverse of that. That means we look at it from this side. Therefore, if Z is equal to 11010, Z prime is equal to 01011.

So, that is the basic property. In this case Tn Z plus Tn Z prime is equal to

$$\sum_{i=1}^{N} a_i Z_i + \sum_{i=1}^{N} a_i Z_{N-i+1}$$

because Zi Prime is equal to Z N minus i plus 1 is equal to sigma ai Zi plus. Now, we are making a change of variable we are calling N minus i plus 1 is equal to j. Therefore, Z N minus i plus 1 becomes Zj and ai is equal to a N minus j plus 1 because i is equal to Z N minus j plus 1 is equal to by change of variable j to i is equal to

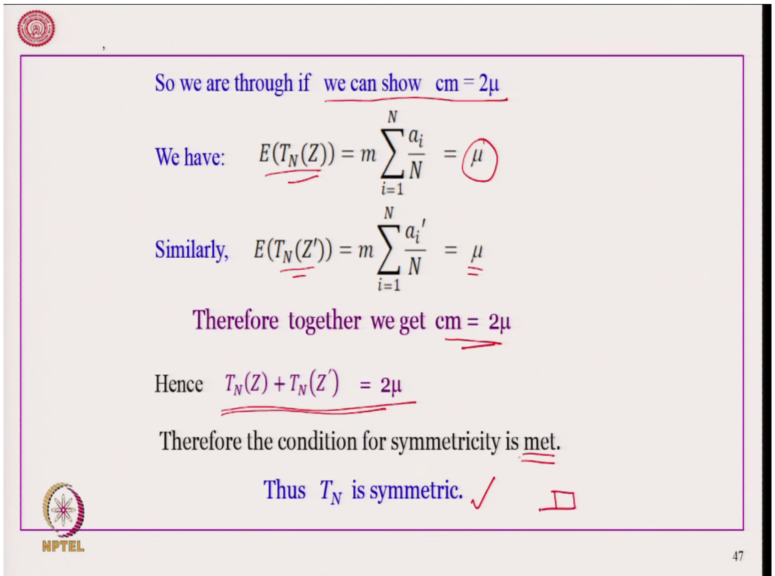
$$\sum_{i=1}^{N} (a_i + a_{N-i+1}) Z_i$$

$$c \sum_{i=1}^{N} Z_i$$

is equal to c times m.

Therefore, Tn Z plus Tn Z prime is a constant.

(Refer Slide Time: 40:19)



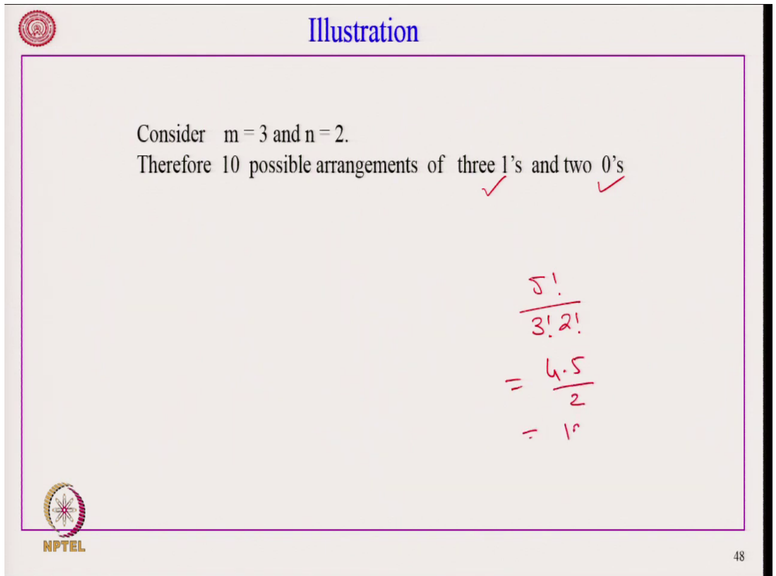
So, we are through if we can show that cm is equal to twice mu, because if you remember we said that distribution will be symmetric if Tn Z plus Tn Z prime is equal to twice mu. Now, the expected value of Tn Z is equal to m times sigma ai upon N is equal to mu. Similarly, expected value of Tn Z prime is equal to m times sigma ai prime upon N is equal to mu.

Where ai prime is the constant we defined for Z prime. Therefore, together we get cm is equal to twice mu. Therefore,

$$T_N(Z) + T_N(Z') = 2\mu$$

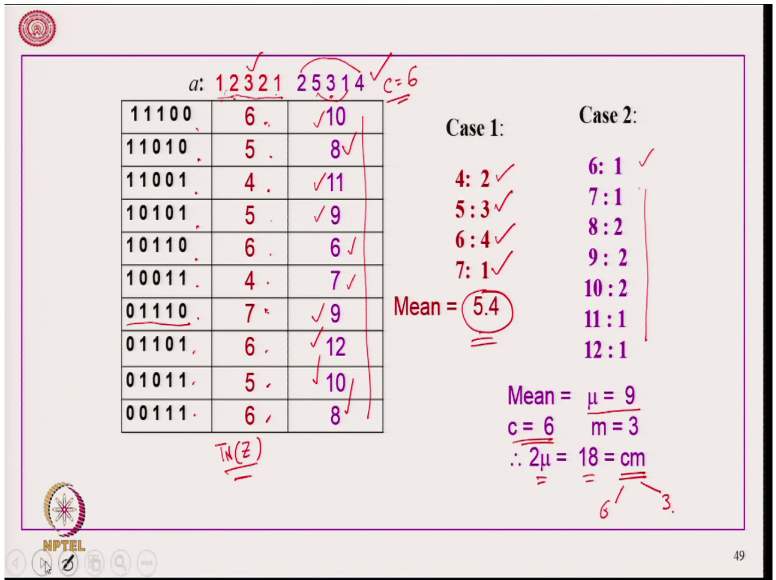
Therefore, the condition for symmetry is met with and therefore Tn is symmetric, proved.

(Refer Slide Time: 41:21)



Let me now illustrate this with a simple example m is equal to 3 and n is equal to 2. That means you are looking at Z vectors of 3 1's and 2 0's. And the number of possible ways of doing it is 10. Because this is factorial 5, upon factorial 3, factorial 2 is equal to 4 into 5 upon 2 is equal to 10.

(Refer Slide Time: 41:50)

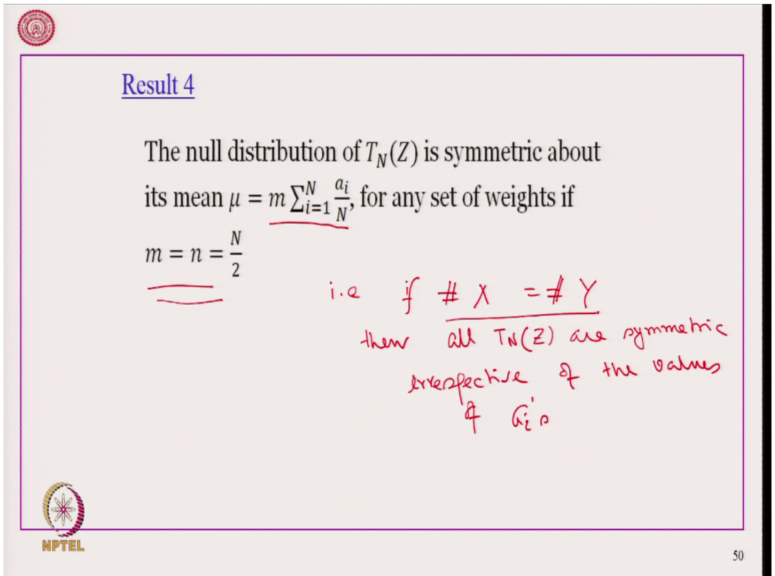


So, therefore, we need to consider 10 permutations, that is what we have done 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. Let us consider the a vector to be this and a vector like this. Note that here, the condition is not maintained that ai plus a N minus i plus 1 is constant, but in this case we can see that this is 2 plus 4, 5 plus 1 and 3 plus 3, all of them are going to be 6.

So, in this case c is equal to 6. Now, if we calculate the value of Tn Z; this is coming out to be, in this case 6, 5, 4, 5, 6, 4, 7, 6, 5, 6 this you can easily calculate for illustration, when it is 11100 its dot product with 1, 2, 3, 2, 1 is going to be 1 plus 2 plus 3 is equal to 6 on the other hands. If I consider 0110, then its dot product is going to be 2 plus 3 plus 2, therefore this is going to be 7. Therefore, what is going to be the distribution of Tn Z, we can see that it takes the value 4, two times 5, three times 6, four times and 7, one time.

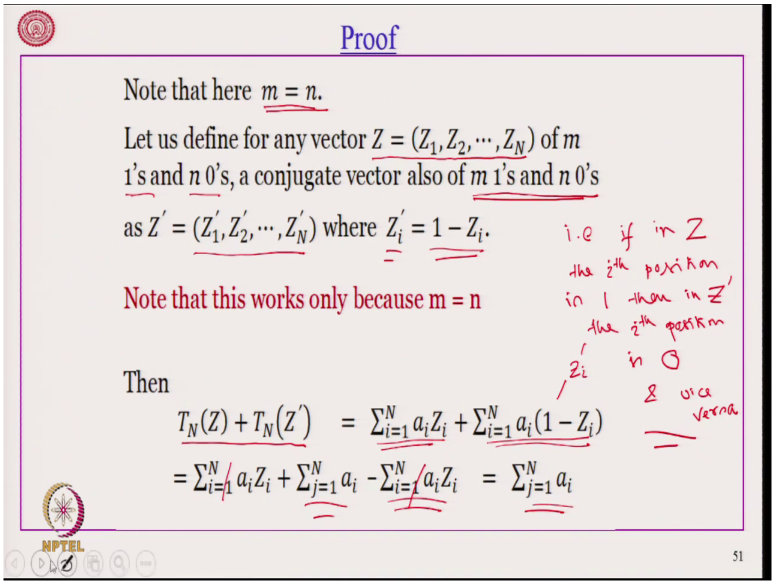
And the mean is coming out to be 5.4 because 6 plus 5, 11, plus 4, 15 plus 5, 20 plus 6, 26, 30, 37, 43, 48 plus 6 is equal to 54. Therefore, mean is equal to 5.4. In a similar way you calculate for this a vector, we will find that our distribution is coming out to be 6 once, 7 once 8 twice, 10 twice, 9 twice, 11 once and 12 once. Therefore, we can see that the mean of the distribution is coming out to be 9, c is equal to 6 we have already calculated, therefore, twice mu is equal to 18 is equal to cm, where c is equal to 6, a m is equal to number of x is equal to 3. And we can see that this is symmetric around the mean. It is not approved because proof we have already given this is only to visualize the result.

(Refer Slide Time: 44:41)



Now, let me give you another result, the null distribution of Tn Z is symmetric about its mean mu Which is, is equal to m into sigma ai upon N for any set of weights, if m is equal to capital N by 2, that is, if number of X is equal to number of Y, then all Tn Z's are symmetric; irrespective of the value of ai's. You consider any constant ai's, the distribution will remain symmetric, provided number of X is equal to number of Y.

(Refer Slide Time: 45:44)



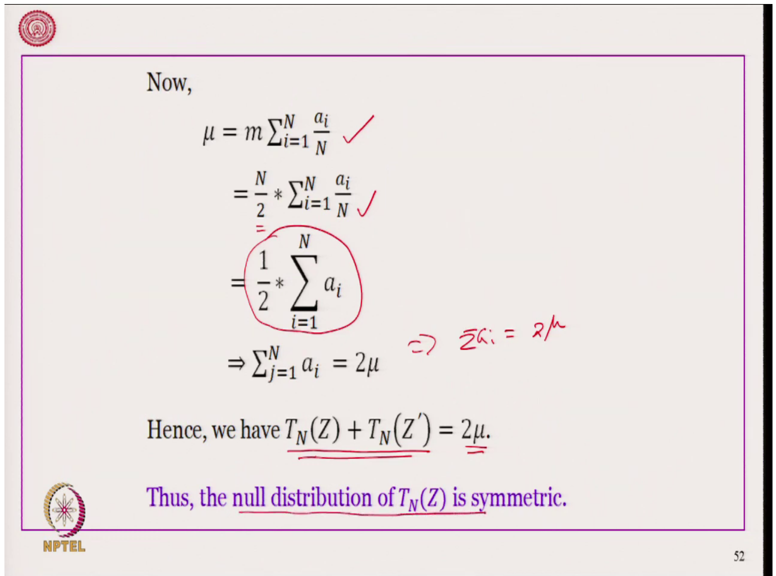
So, proof note that a m is equal to N, therefore, let us define for on any vector Z is equal to Z1, Z2, Zn of m 1's and n 0's, a conjugate vector also of m 1's and n 0's as Z prime is equal to Z1 prime, Z2 prime, Zn prime where Zi prime is equal to 1 minus Zi. That means, if in Z the i th position is 1, then in Z prime, the i th position is 0 and vice versa.

Because number of X is equal to number of Y that is a m is equal n. Therefore, by doing this transformation, we still maintain the number of 1's and 0's. Therefore,

$$\sum_{i=1}^{N} a_i Z_i + \sum_{i=1}^{N} a_i (1 - Z_i)$$

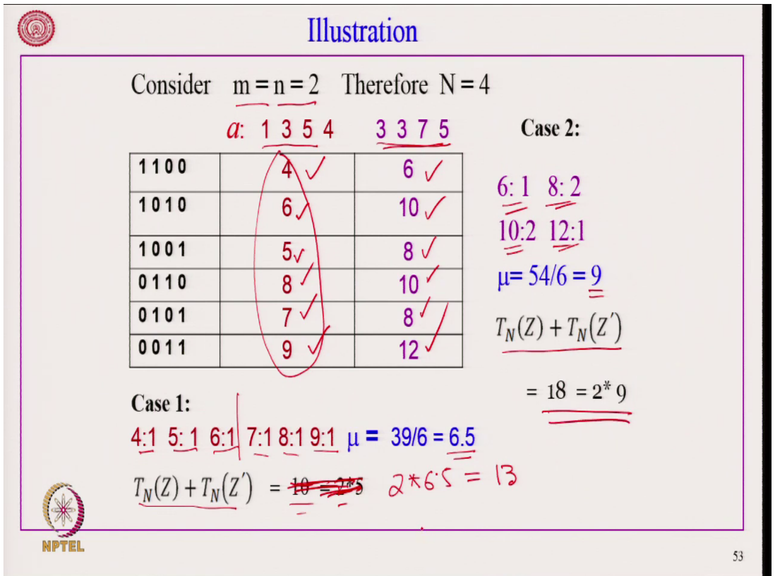
because this is coming from Zi prime is equal to sigma ai Zi plus sigma ai minus sigma ai Zi is equal to sigma ai, because, these 2 cancels.

(Refer Slide Time: 47:32)



Now, mu is equal to m by N sigma ai. Now, small m is equal to N by 2 times sigma ai by N. N cancels therefore, we have half times sigma ai implies that sigma ai is equal to twice mu. Therefore, Tn Z plus Tn Z prime is equal to twice mu. Therefore, by our condition of symmetricity, the null distribution of Tn Z is symmetric.

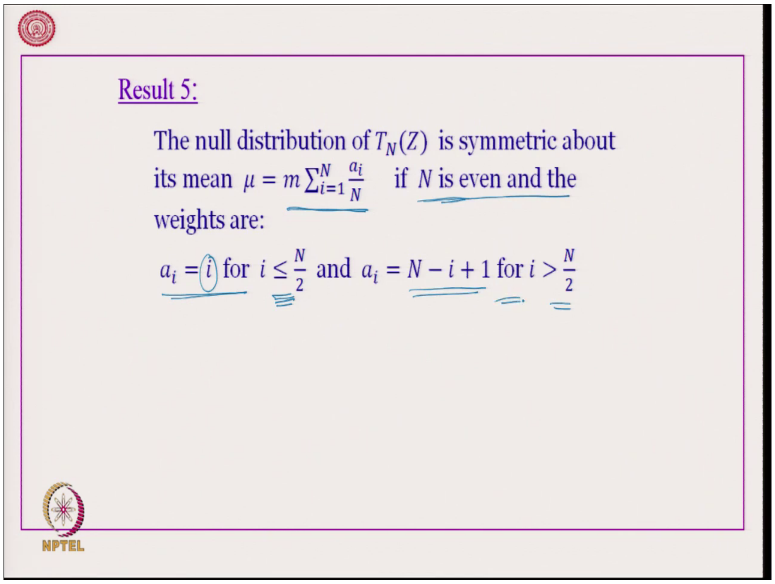
(Refer Slide Time: 48:10)

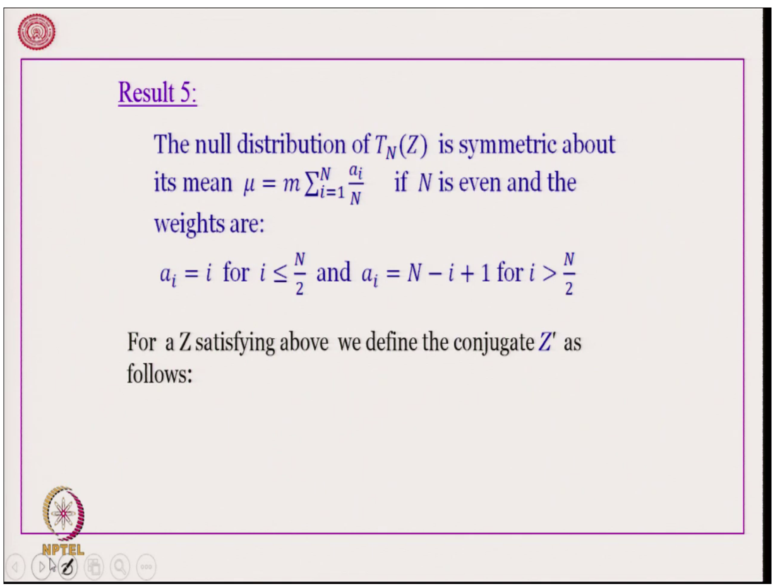


So, illustration, we have taken simple case, 2 m's and 2 n's, therefore, capital N is equal to 4. We have chosen 2 arbitrary weight vectors 1 3 5 4 and 3 3 7 5. In this case, this is the distribution 1 4, 1 5, 1 6, 1 7, 1 8 and 1 9; mu is equal to 6.5 and therefore, we can see that Tn Z plus Tn Z prime is equal to 2 into 6.5 is equal to 13. And as we can see here, that 9 plus 4 is equal to 13, 7 plus 6 is equal to 13. And 8 plus 5 is equal to 13. Similarly, for case 2, the distribution will come out to be 6 once, 8 twice, 10 twice, and 12 once.

The mu is coming out to be 9, and we can see that Tn plus Tn Z prime is equal to 12 plus 6 is equal to 18, 10 plus 8 is equal to 18 and 8 plus 10 is equal to 18. Therefore, we get that is equal to 2 times 9.

(Refer Slide Time: 49:30)

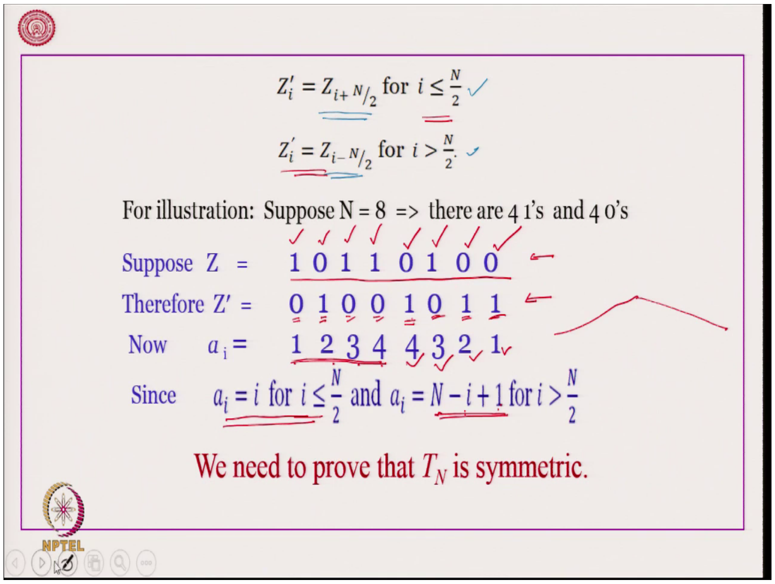




Now, let me give you another result, which is the result number 5. This is the last result that we are going to do in this class. The null distribution of Tn Z is symmetric about its mean mu, which we know is m into summation i is equal to 1 to N ai upon N, if N is even, and weights are ai is equal to i for i less than equal to N by 2, and ai is equal to N minus i plus 1 for i greater than N by 2.

That is, if the total number of observation is even, and the weights given our ai is equal to i for the first half of the values, and it is N minus i plus 1 for the other half of the values, then the distribution of Tn is going to be symmetric about its mean. For Z satisfying the above, we define a conjugate Z prime as follows.

(Refer Slide Time: 50:36)

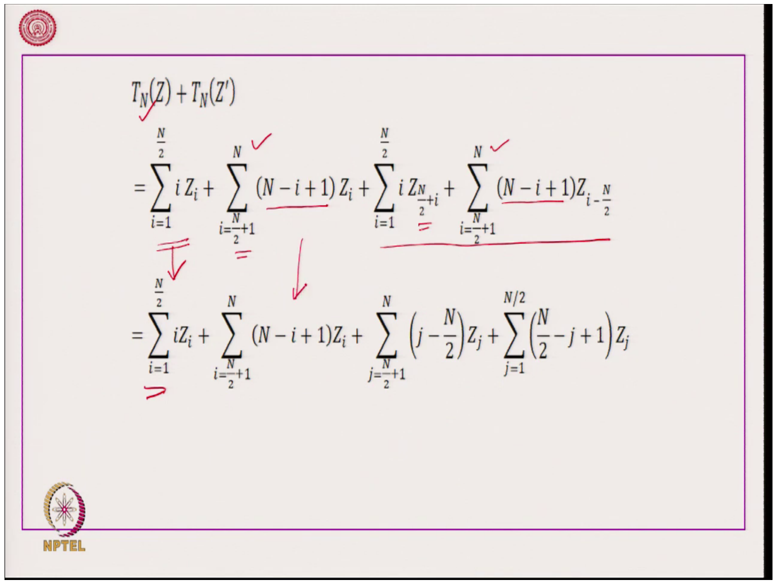


Z prime i is equal to Zi plus N by 2 for i less than equal to N by 2 and Zi prime is equal to Zi minus N by 2 for i is greater than N by 2. So, let me illustrate with an example. Suppose N is equal to 8 and there are 4 1's and 4 0's. So, consider a particular Z which is 1 0 1 1 0 1 0 0. Therefore, how do you construct a conjugate Z prime, for i is equal to 1 to N by 2, Zi prime is Z i plus N by 2. Therefore, Z1 prime is equal to 0, coming from here. Z2 prime is equal to 1 coming from here. Similarly, this is 3 from here, and this is 4 from here.

For i greater than N by 2 that means for 5, 6, 7, 8. The relationship is Zi prime is getting the value of Zi minus N by 2. Therefore, Z5 prime is getting the value of Z1, Z6 prime is getting the value of Z2, Z7 prime is getting the value of Z3 and Z8 prime is getting the value of Z4. So, like that, corresponding to each Z, we can construct and appropriate Z prime. And notice that this is also belonging to the same vector with 4 1's and 4 0's.

Now, therefore, the weights are going to be it is ai is equal to i for i less than equal to N by 2. Therefore, for the first 4 the weights are 1, 2, 3, 4 and ai is equal to N minus i plus 1 for i greater than N by 2. Therefore, when i is equal to 5, it is 8 minus 5 plus 1 therefore 4 when i is equal to 6, it is 8 minus 6 plus 1 that is 3. Similarly, 2 and 1. Therefore, we can see that the weights are decreasing in a symmetric manner as well. We need to prove that Tn is symmetric around this mean.

(Refer Slide Time: 53:06)

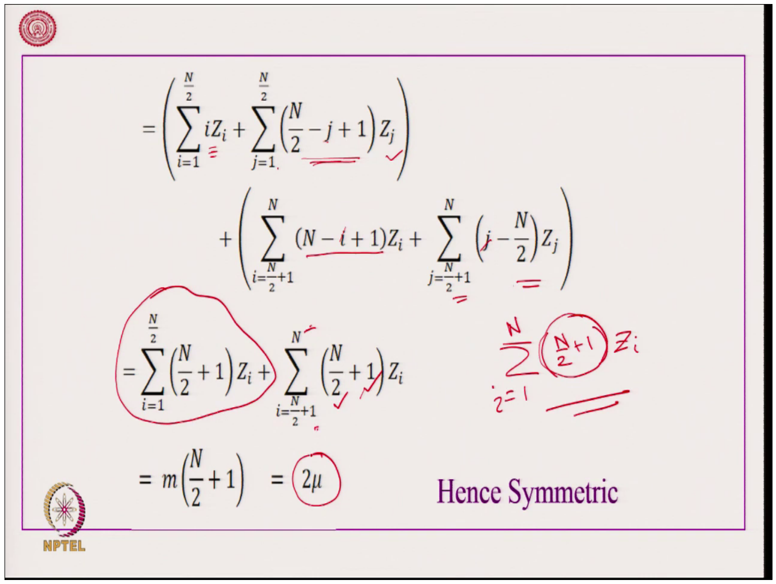


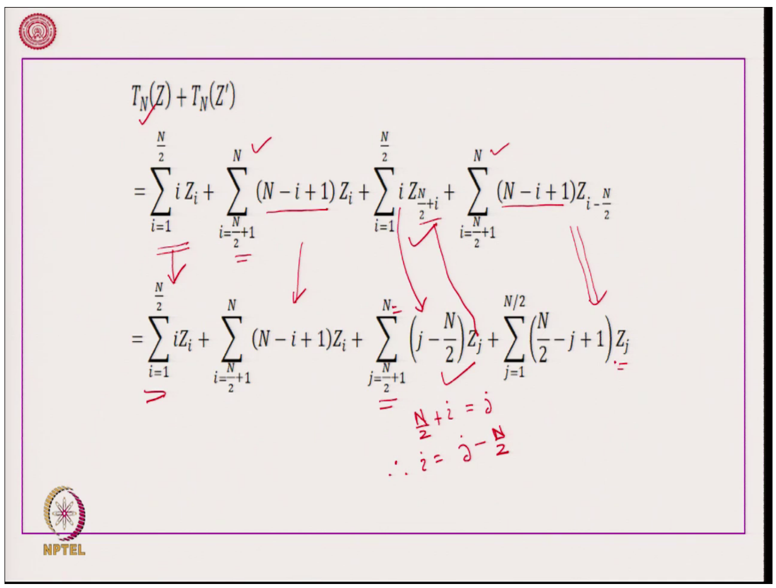
So, let us prove it mathematically, Tn Z plus Tn Z prime is equal to sigma over i is equal to 1 to N by 2 i Zi and i is equal to N by 2 plus 1 N it is N minus i plus 1 Zi. So, that is coming from Tn Z. In a similar way Tn Z prime is equal to i is equal to 1 to N by 2 Zn by 2 plus i, because that is how the Z prime has been allocated values and from i is equal to N by 2 plus 1 to N it is N minus i plus 1 into Zi minus N by 2, is equal to for i is equal to 1 to N by 2, i Zi plus i is equal to N by 2 plus 1, it is N minus i plus 1 Zi.

Now, here what we are doing, we are making a change of variable. So, on this side we make a change of variable N by 2 plus i is equal to j. Therefore, i is equal to j minus N by 2. Therefore, what we are doing? We are putting instead of i, j minus N by 2 and in place of N by 2 plus i we are putting j and since i is going from 1 to N by 2, j is going from N by 2 plus 1 to N.

In a similar way, we are again making a change of variable and what we are getting is that j is equal to 1 to N by 2, it is N by 2 minus j plus 1 multiplied by Zj,

(Refer Slide Time: 55:22)



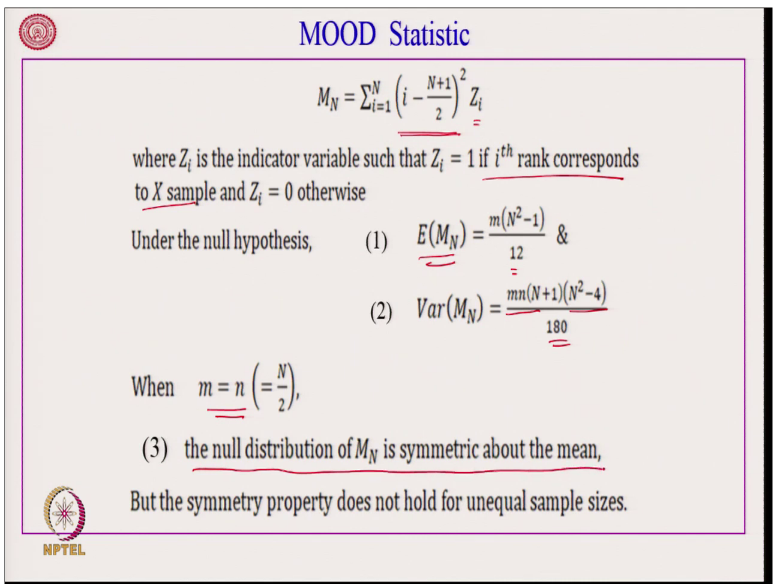


Which is equal to sigma i is equal to 1 to N by 2, i Zi plus j is equal to 1 to N by 2, N by 2 minus j plus 1 Zj that is, we are combining this and this. In a similar way, we combine these 2 and therefore, we get this following i is equal to N by 2 plus 1 to N, N minus i plus 1 Zi plus j is equal to N by 2 plus 1 to N, j minus n by 2 Zj.

Here we can make a change of variable instead of j if I call it i, then what we are getting is i Zi plus N by 2 minus i plus 1 Zi and when we combine then, i and minus i cancels therefore, we get N by 2 plus 1 times Zi i is equal to 1 to N by 2. In a similar way, when I am calling it instead of j equal to, we call it i, then we will get that it is equal to N by 2 plus 1 to N by 2 plus 1 to N. Now, this cancels with this.

Therefore, we get N minus N by 2 is equal to N by 2 plus 1 times Zi. So, together what we are getting? Together we are getting it is that, sigma i is equal to 1 to N, N by 2 plus 1 times Zi. Now, we can take it out of the summation and sigma Zi i is equal to 1 to N is equal to m. Therefore, what we are getting is twice mu. Therefore, it is symmetric.

(Refer Slide Time: 57:13)



With this background, we look at the Mood Test. In the Mood test, the statistic is a Mn is equal to i is equal to 1 to N, i minus N plus 1 by 2 whole square times Zi. Where Zi is equal to 1 if i th rank corresponding to an X sample, that means, we are looking at the ranks of X values. We want to show that expected value of a Mn is equal to m into N square minus 1 by 12. Variance of Mn is equal to mn into n plus 1 into n square minus 4 by 180. And if m is equal to n and both are equal to N by 2, then the null distribution of mn is symmetric about the mean.

(Refer Slide Time: 58:06)

Substituting this in the formula for the mean and variance of T_N are As follows:

$$E(M_N) = m\sum_{i=1}^{N}\frac{a_i}{N} = \frac{m}{N}\sum_{i=1}^{N}a_i = \frac{m}{N}\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)^2$$

$$= \frac{m}{N}\sum_{i=1}^{N}\left(i^2 - 2*\frac{N+1}{2}*i + \left(\frac{N+1}{2}\right)^2\right)$$

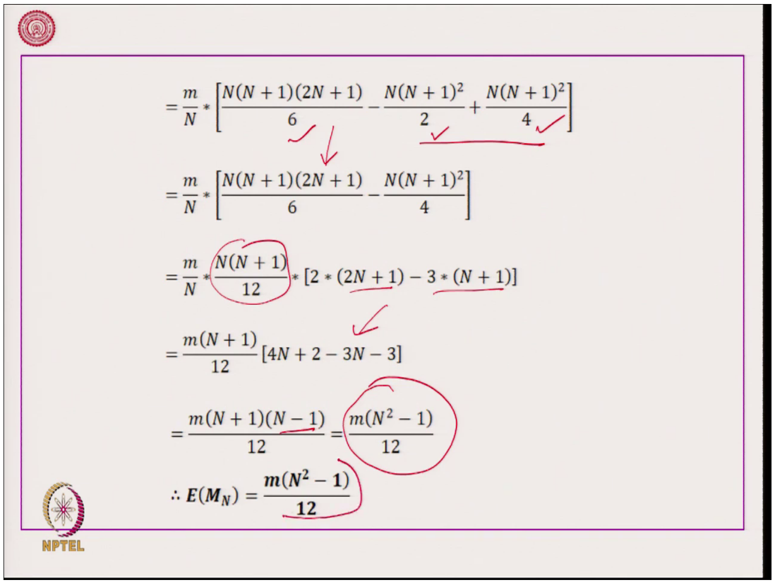
$$= \frac{m}{N}*\left[\sum_{i=1}^{N}i^2 - (N+1)*\sum_{i=1}^{N}i + \left(\frac{N+1}{2}\right)^2\sum_{i=1}^{N}1\right]$$

$$= \frac{m}{N}*\left[\frac{N(N+1)(2N+1)}{6} - (N+1)*\frac{N(N+1)}{2} + \left(\frac{N+1}{2}\right)^2 *N\right]$$

So, let us put the first result, expected value of a Mn is equal to m summation ai over N, this we all know is equal to this mn into sigma ai is equal to m by N sigma is equal to 1 to N, i minus N plus 1 by 2 whole square, is equal to m by N multiplied by summation of i square minus 2 into N plus 1 by 2 times i plus N plus 1 by 2 whole square by opening up this square.

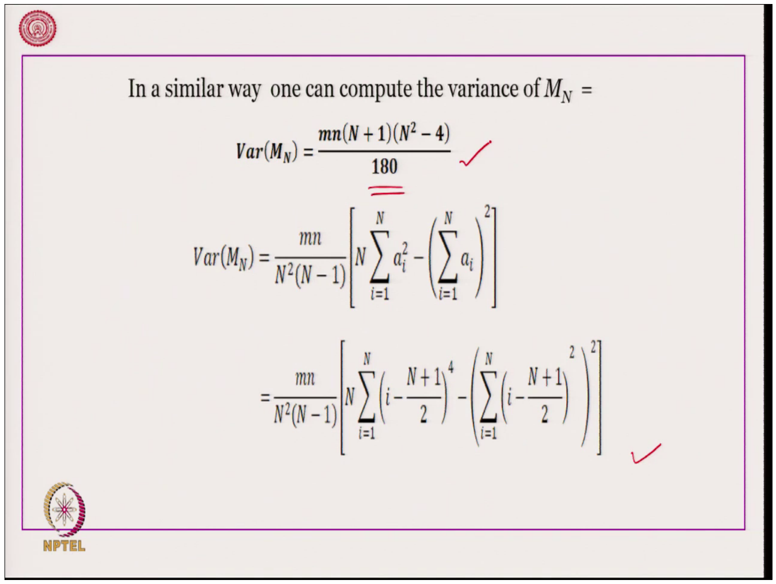
Is equal to m by N times sigma i square minus n plus 1 times sigma i plus n plus 1 by 2 whole square times sigma 1, because there is no variable involved is equal to N into N plus 1 into 2N plus 1 by 6 minus N plus 1 times N into N plus 1 by 2 summation over i plus N plus 1 by 2 whole square times N; this is straightforward.

(Refer Slide Time: 59:22)



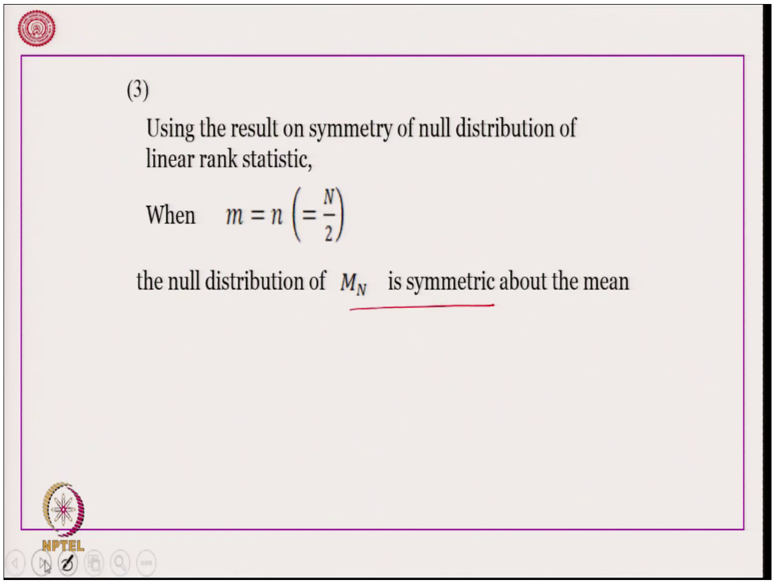
Now, let us simplify it is equal to N into N plus 1 into 2N plus 1 by 6 minus a N into N plus 1 whole square by 2 plus N into N plus 1 whose square by 4 is equal to the first term comes as it is these 2 gives us minus N into N plus 1 whose square by 4. Therefore, as you simply it, we take out N into N plus 1 by 2 as common, what we get is 2 into 2N plus 1 minus 3 into N plus 1. Which is further simplified as 4N plus 2 minus 3, which is equal to N minus 1. Therefore, the expected value comes out to be m into N square minus 1 by 12.

(Refer Slide Time: 60:13)



In a similar way 1 can compute the variance of Mn's. Mn into N plus 1 into N square minus 4 by 180. 1 has to go and expand it in this following way by putting the value of ai but, as you can understand that it is very laborious arithmetic, I am living it as an exercise for you. You please try and check that actually you could arrive at this value for the variance.

(Refer Slide Time: 60:43)



The third 1 was that if m is equal to n then the null distribution of MN is symmetric about the mean this already, we have proved that for equality of Mn, all the linear Rank statistics are going to be symmetric. Therefore, it comes naturally. Ok friends. I stop here today, in the next class. I shall describe a few tests on goodness of fit. Thank you very much.