INDIAN INSTITUTE
OF
TECHNOLOGY
KHARAGPUR

NPTEL
National Programme
on
Technology Enhanced Learning

Applied Multivariate Statistical Modeling

Prof. J. Maiti
Department of Industrial Engineering and Management
IIT Kharagpur

Lecture – 33

Topic

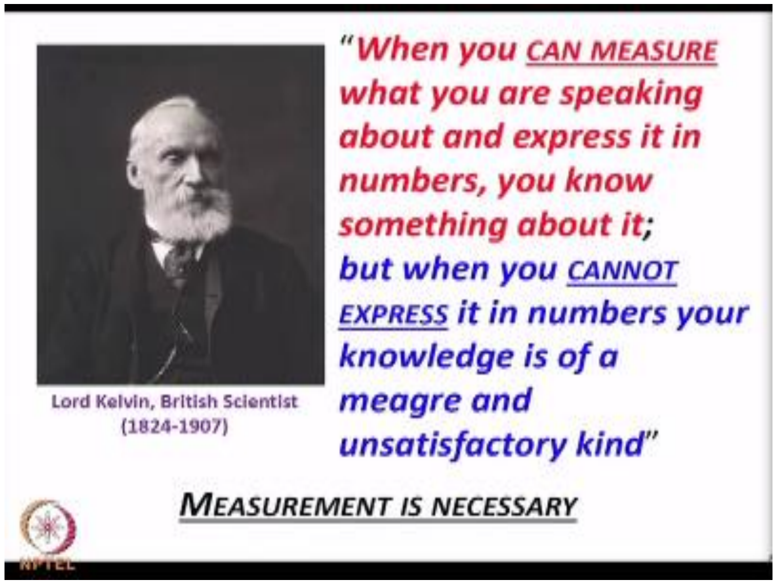
Factor Analysis

Good morning, today our discussion will be on factor analysis.

(Refer Slide Time: 00:24)

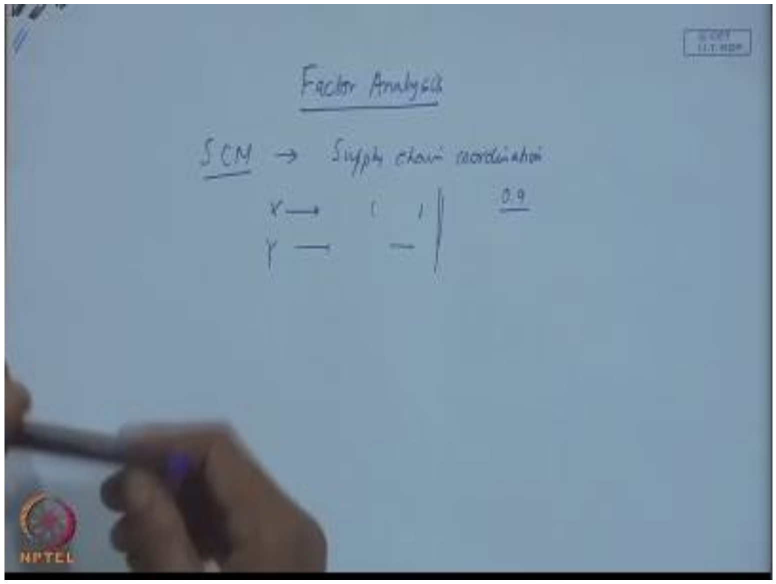
(Refer Slide Time: 00:35)

(Refer Slide Time: 00:36)



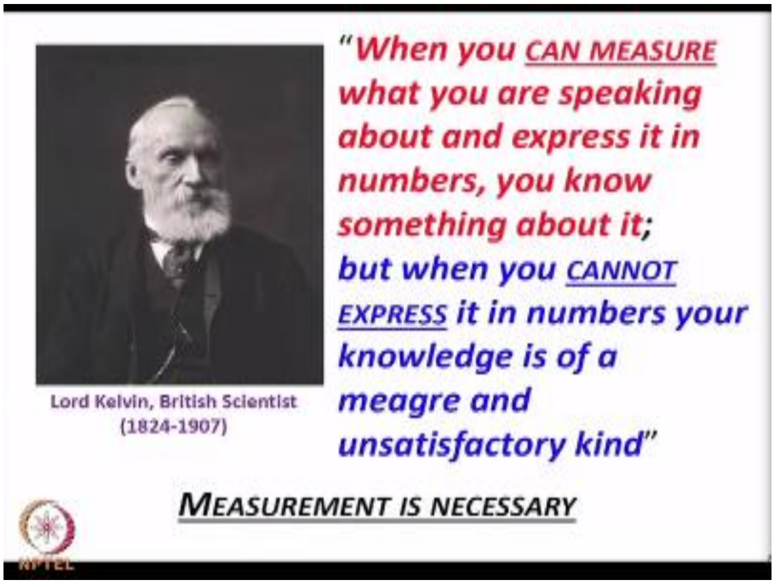
So, I will start with the famous quote of Lord Kelvin, the British scientist, what he said, "When you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot express it in numbers your knowledge is of a meager and unsatisfactory kind". The measurement is necessary that means, what he meant to say is suppose, whenever you talk about any system of its behavior then the behavior pattern you must be able to measure. You must able to give a number to it. For example, in our department there is one important area to work on.

(Refer Slide Time: 01:47)



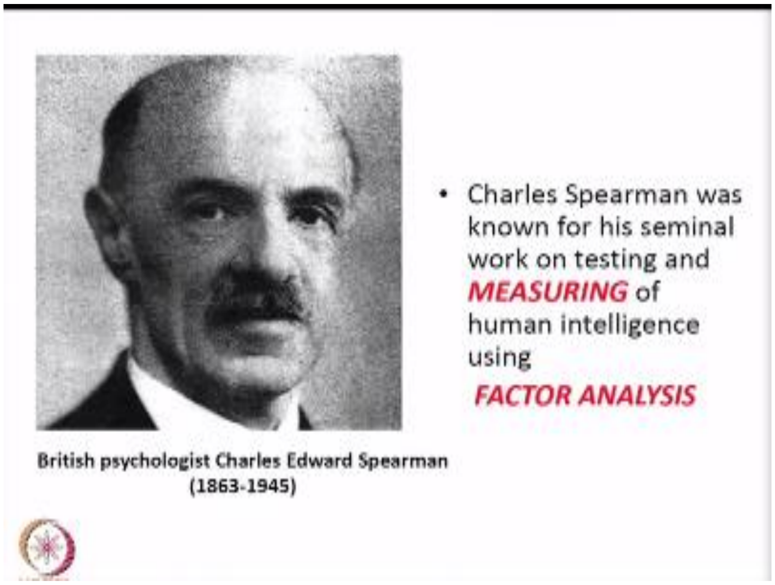
That is supply chain management and we talk much about supply chain coordination. So, we may say that X supply chain is coordination good, Y supply chain coordination bad like these when you talk up in this manner that this knowledge is not perfect. If I say that, yes suppose there is the particular food let it be paddy supply chain in Indian context. Then you are able to tell that okay, at the local level block level or at the state level, that the different agent how they are coordinated and the measure it 0.9. Suppose, something like this then this is a valuable knowledge that is what is important.

(Refer Slide Time: 02:55)



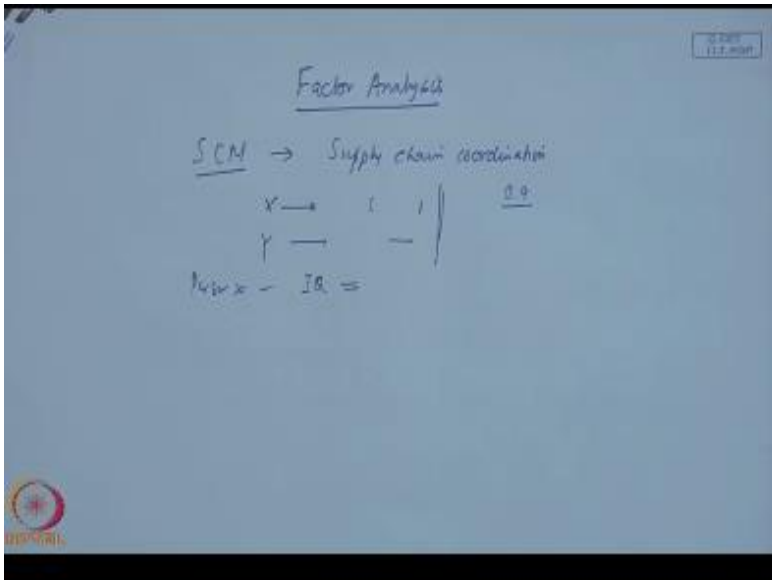
And that is said by Lord Kelvin.

(Refer Slide Time: 02:58)



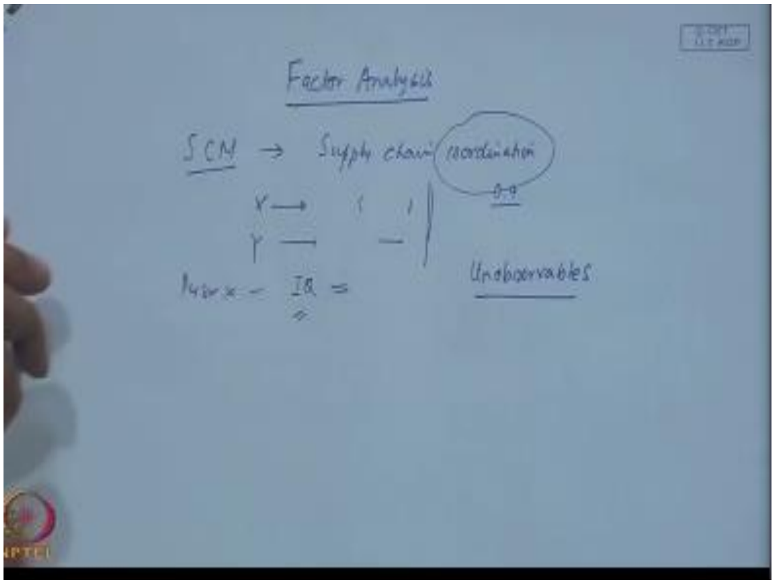
So second gentleman, I think he is one of the famous statistician Charles Edward Spearman. So, Charles Spearman he measured the human intelligence that means what I mean to say.

(Refer Slide Time: 03:24)



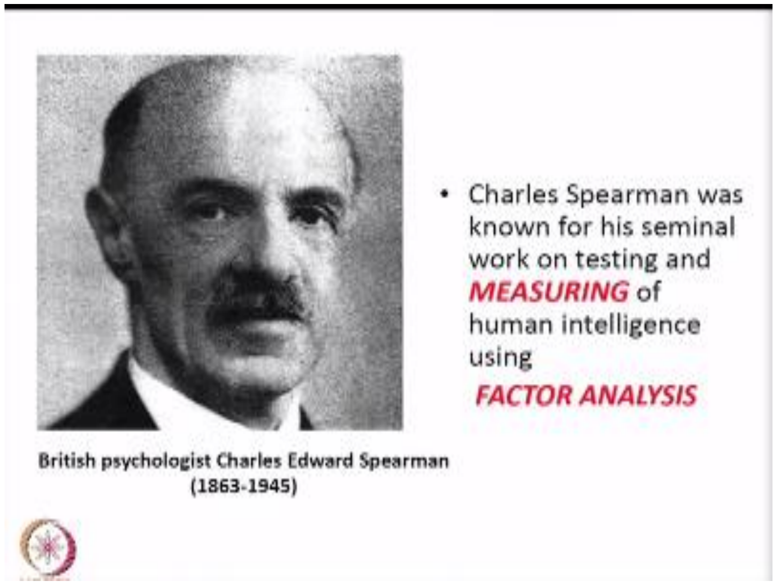
Suppose, person X intelligence level or IQ is some value person Y some value like this. So, human intelligence you can understand it is such a concept. It is not that 5 kg of rice or 25 kg of some other product. It is not like this, you cannot just take some this type of measurement and go. It is a soft thing for example, like supplies in coordination you will say soft thing your mental measurement of that intelligence, human intelligence. It is not visible, these all are basically unobservable things.

(Refer Slide Time: 04:12)



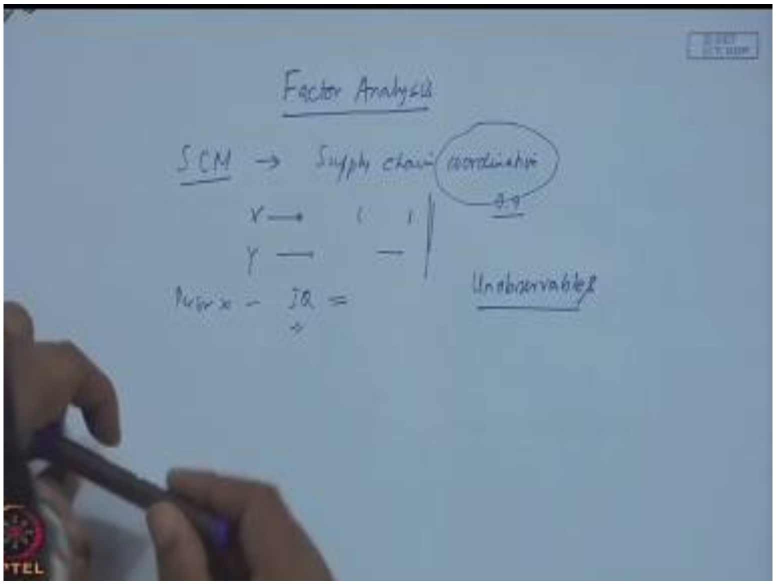
Unobservable I can say the coordination is unobservable. You cannot directly go and measure intelligence, also unobservable. Then question comes how do you measure such unobservable? So, there are certain techniques.

(Refer Slide Time: 04:35)



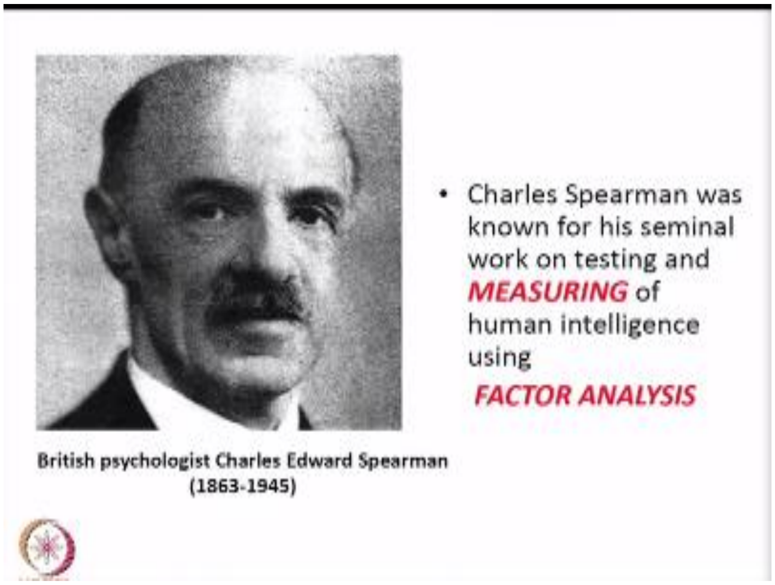
And the factor analysis is one of such techniques and probably the one of the best techniques.

(Refer Slide Time: 04:45)

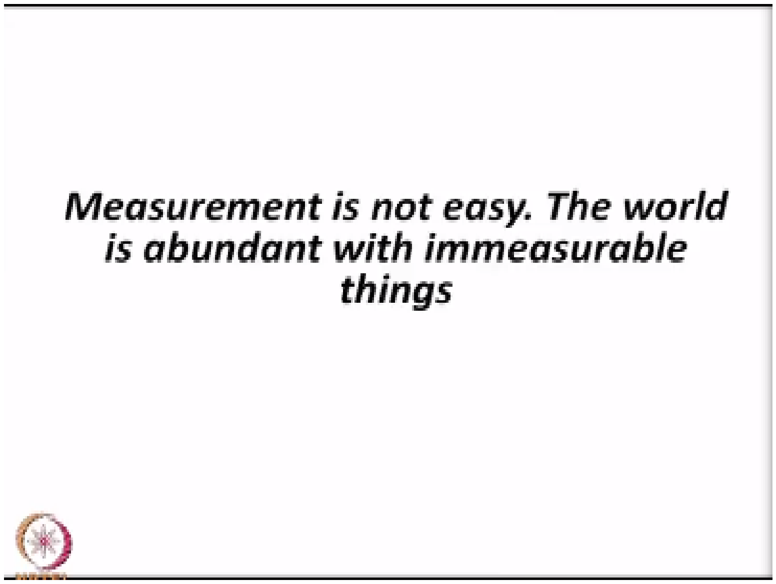


That is used in measuring unobservables okay.

(Refer Slide Time: 04:54)

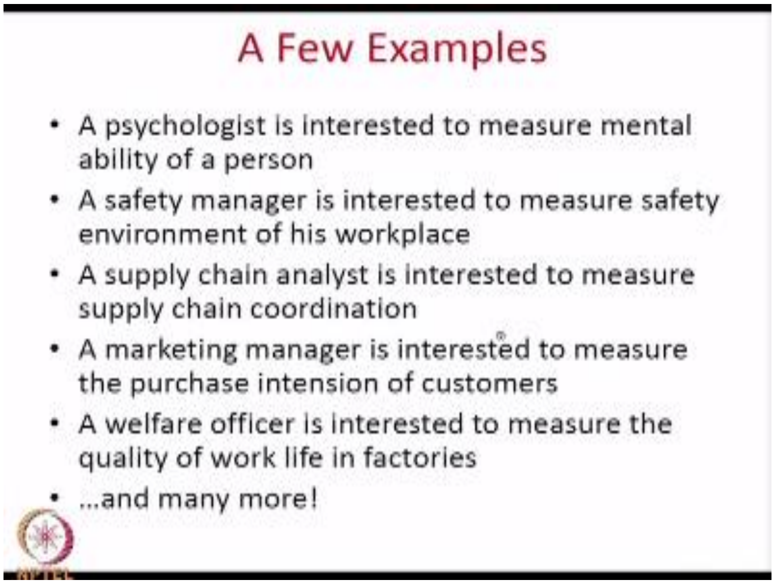


(Refer Slide Time: 04:55)



The key concept here is measurement is required, but measurement is not easy. The world is abundant with immeasurable things which cannot be measured. I will give you some example here.

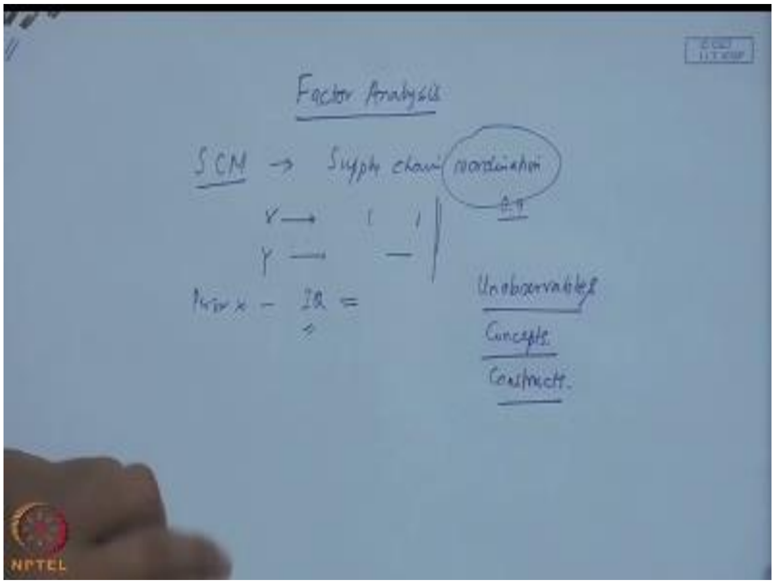
(Refer Slide Time: 05:13)



A psychologist is interested to measure the mental ability of a person. A safety manager is interested to measure safety environment of his workplace. A supply chain analyst is interested to measure supply chain coordination. A marketing manager is interested to measure the purchase intention of customers. A welfare officer is interested to measure the quality of work life in factories, and many more.

You can find out hundred such items examples, where you see that it is not there is one instrument hardware based instrument you will go and measure, it is not possible. We have a 90% of the cases take decisions based on this type of concepts.

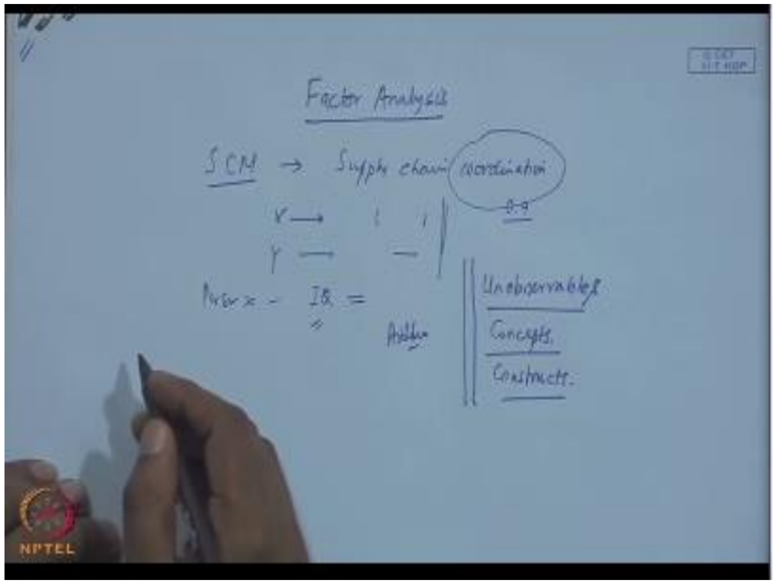
(Refer Slide Time: 06:07)



So, another word is I want to say that concepts many time, we say that constructs. So, by saying these we are basic unobservable concept construct, by saying this we are basically telling to you that there are certain things which are hidden unobservables that is not directly measureable. But those things are very, very important and because those things causes something to happen. For example, if there is not proper coordination in the supply chain, the supply chain performance will be poor okay.

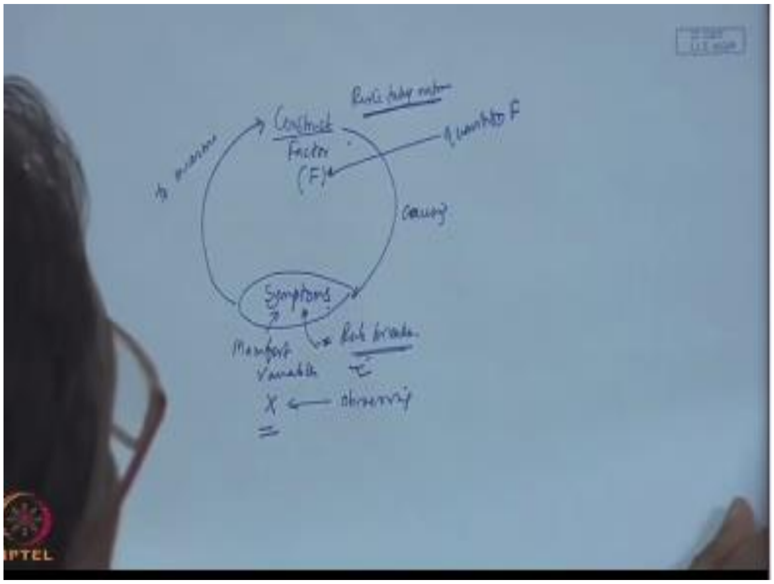
Suppose, the quality of what life is not good then the human productivity safety and all other decision accompany that will be poor.

(Refer Slide Time: 07:15)



So, that means although there are many things like these which are unobservable, which are concepts which constructs hidden. This cannot be directly measured, but their effects means because of the presence of these constructs or concept of unobservable things manifested into different symptoms. What I mean to say, I mean to say that.

(Refer Slide Time: 07:45)

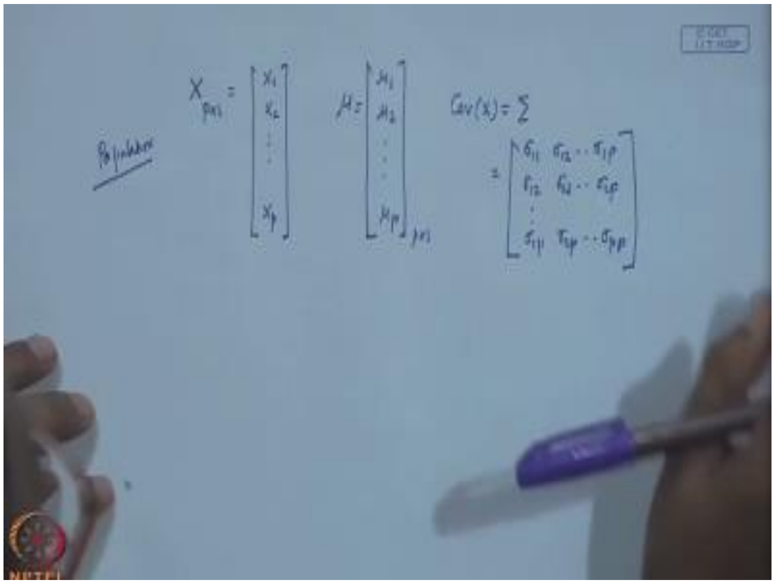


Suppose there is construct concept or we can say factor. So, these things manifested in terms of symptoms. So, because for example, someone is risk taking in nature, so what will be symptoms he will may be a rule breaker not follow the traffic rule. So, you are able to see this symptom that not following the rules may be rebellious, but it is so that risk taking that attitude. What is his personality that is the main reason, the cause of this rule breaking.

That is why we say that that sometime hard hidden on a immeasurable constructs or factors are there which are causing the symptoms, and these symptoms can be used to measure this factor or construct okay that mean factors or co-factoral construct causing the symptoms can be used to measure construct, getting me? So, essentially that concept is like this, if I say this factor is f and there are several symptoms which we can say the manifestation or manifest variable the symptoms are nothing, but manifestation of these hidden thing.

So, this manifest variable if we say X, correct, so, that mean what is available with you, you are observing this X and you want to quantify F, getting me? You observing this that means with this if we can say something like this.

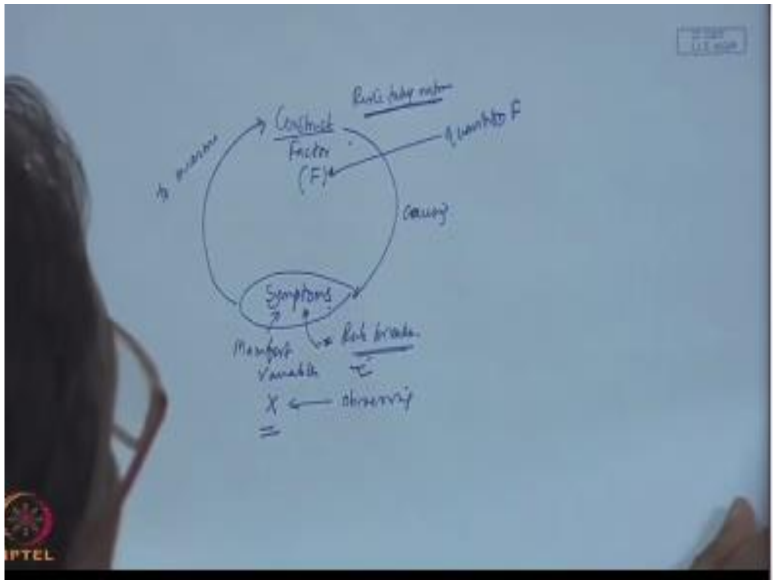
(Refer Slide Time: 10:41)



That you are observing X which is Px1 that means P number of this is a variable vector having P manifest variables okay that will be coming from a population. Now, suppose this X is having the mean vector like this, which will also be μ_1, μ_2, μ_P. This will be P x 1 vector, this X is also having a covariance matrix which is $\sum$ this one is you can write $\sum_{11}$, $\sum_{12}$, $\sum_{1P}$, $\sum_{12}$, $\sum_{22}$, $\sum_{2P}$, like this $\sum_{1P}$, $\sum_{2P}$, $\sum_{PP}$ okay.

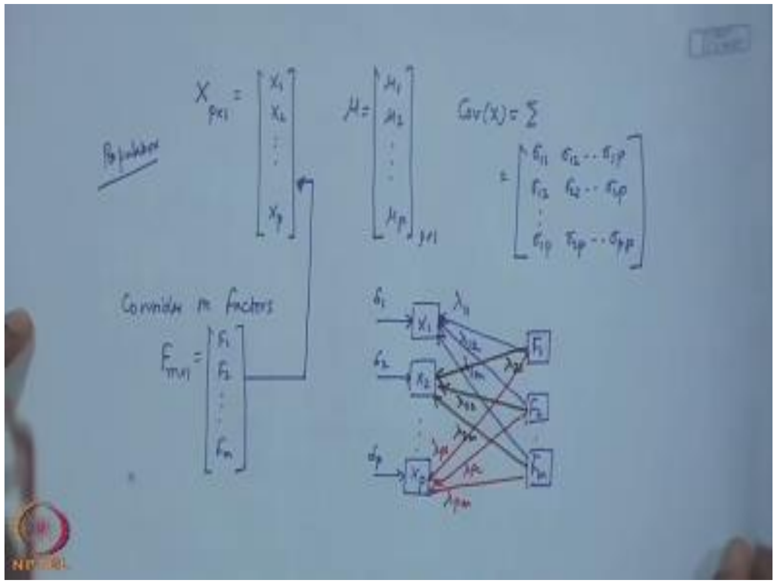
So, this is what is known about the population, that mean there are P variables, P mean vector, P plus C co-variance vectors. The mean that μ_1 to μ_P or $\sum_{11}$ to $\sum_P$, these values may not be known usually, not known from the population point of view okay.

(Refer Slide Time: 12:34)



Now, what we have described here? We said that in this place we described that this X you are observing, because there is certain underlined causes or hidden constructs or factors.

(Refer Slide Time: 12:48)



So, if this is the case, if we consider now consider m factors okay. So, then my F is the vector of factors whose is m x 1 this will be your F_1, F_2, like Fm and you are seeing that whatever you observe in X, this is because of the causal factor F, F is causing this, then we can write linear regression like equation also, but first I will pictorially represent this. So, what you can write? I have X_1, this is my one object variables. My second object variable is X_2, like this my last object variable is X_P.

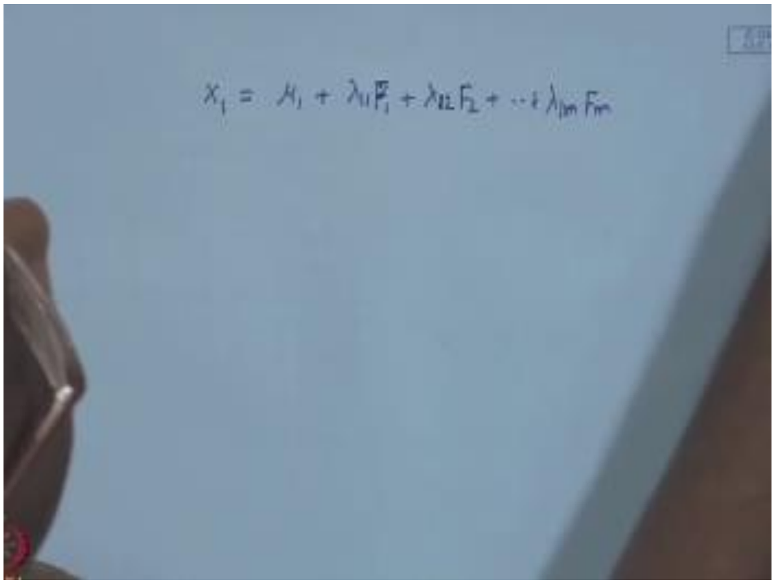
And we are saying that there is factor F_1, F_2 and Fm factors each are causing this X variables to occur to happen. So, in this case I can say X_1 is caused by F_1, X_1 also caused by F_2, like this Fm. What I am trying to say that there are m factors who is influencing this X_1, because of these factors X_1 you are able to observe. So, then if I write like this X_1 is caused by F_1, I can write λ_{11} for X_{11} F_{11} then λ_1 is caused by 2 λ_1 is caused by m okay.

So, what you can do? The same thing will happen to X_2. So, what we will write then this one will be λ_{21}, then λ_{22}, λ_{2m}, then another last one is your, correct? So, λP_1, λP_2, λP_m, when we are assuming what we are assuming this F causing X. So, these are the all influencing coefficient.

So, factor F_1 influenced on X_1 can be measured through λ_{11} like this. Now in addition to these what we assume that these factors collectively cannot explain everything about X_1.

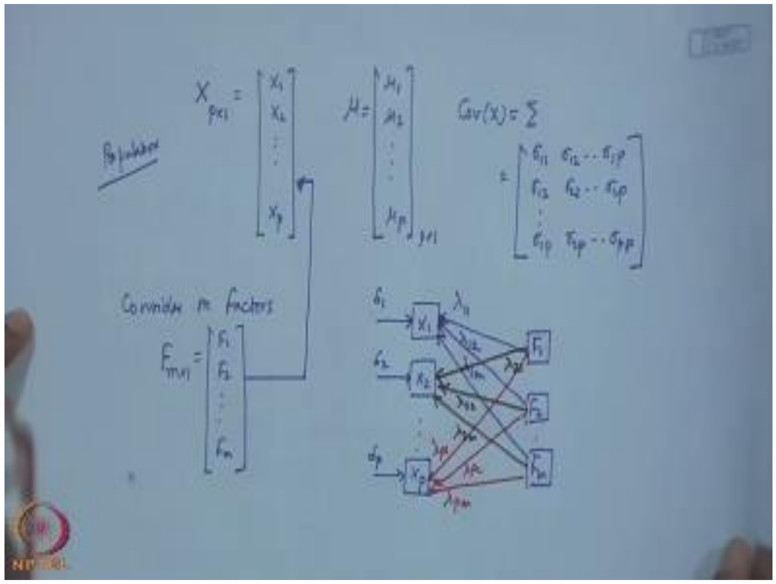
So, there will be one arrow term, correct? Now, using this, but please remember here we are not talking anything about the relationship between F_1, F_2 and F_m. When we are not saying anything here we are here, we are assuming that these factors are independent orthogonal. So, we will see the assumption later on.

(Refer Slide Time: 17:39)



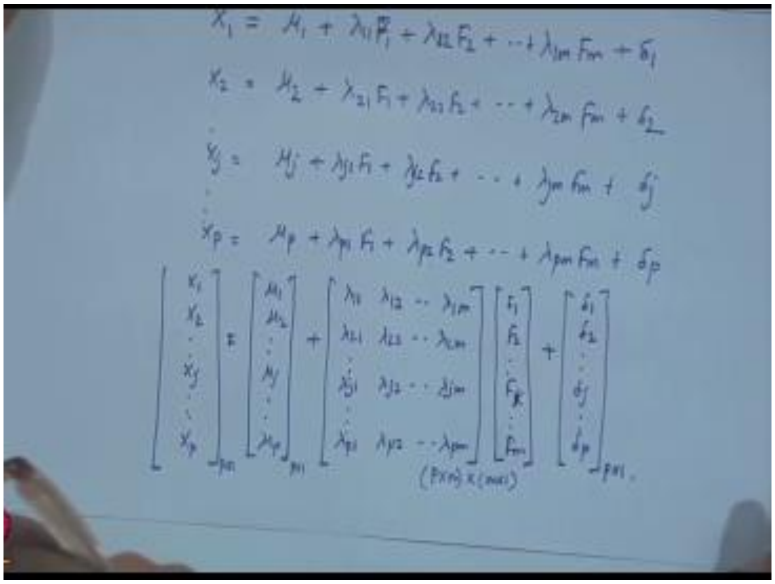
With these if I want to write down certain equation, can I not write down regression like equation $X_1 = \mu_1 + \lambda_{11}F_1 + \lambda_{12}F_2 \ldots \lambda_{1m}F_m$.

(Refer Slide Time: 18:06)



See here, what you are writing, $X_1 = \lambda_{11}F_1$, $\lambda_{12} F_2$, $\lambda_{1m}F_m$ what more is there? δ_1 is there.

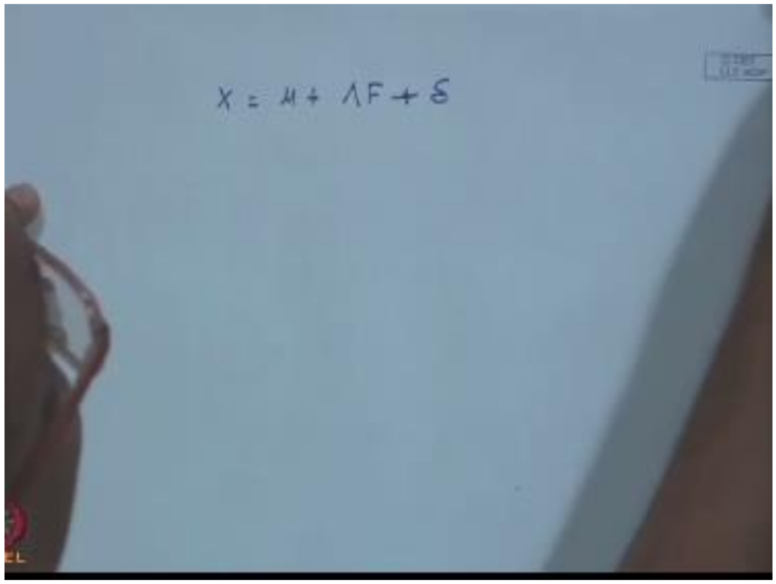
(Refer Slide Time: 18:19)



So + $\delta 1$, similarly, you can write $X_2 = \mu_2 + \lambda_2$ is affected by 1 F_1 $\lambda_{22}F_2$ like $\lambda_{2m}F_m + \delta_2$. So, in this manner you can write $X_j = m_j + \lambda_{j1}F_1$ $\lambda_{j2}F_2$ $\lambda_{jm}F_m + \delta_j$. And final one is $X_p = \mu p+\lambda P_1F_1+\lambda P_2F_2+\lambda P_mF_m+ \delta_p$ correct? So, that means in matrix form you can write X_1, X_2, Xj, X_p = μ_1, μ_2, μ_j, $\mu p+\lambda_{11}$, λ_{12}, λ_{1m}, λ_{21}, λ_{22}, λ_{2m} so, like this λj_1, λj_2, λ_{jm} in the same manner λP_1, λP_2, λP_m into your λ_f F_1 F_2, this will be you, let . Let it be F_j, then your, what will happen ultimately F1 to Fm.

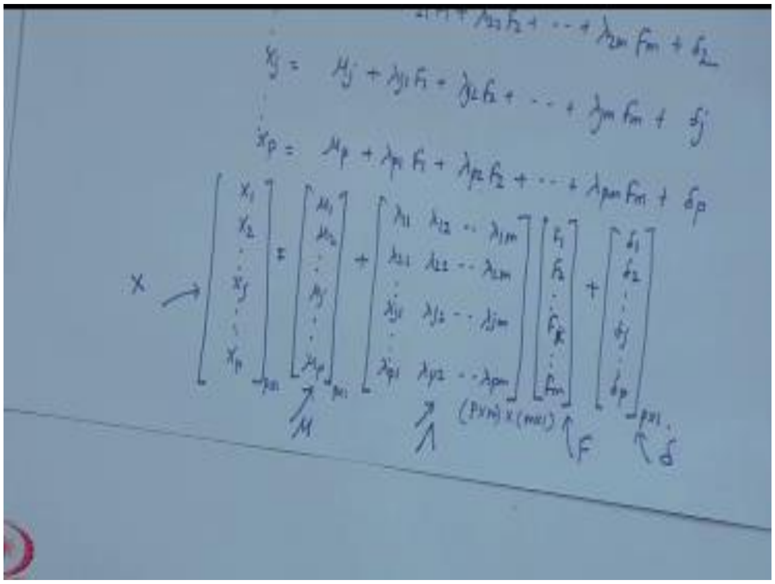
So, I am not giving F_j, let it be even F_k and Fm what is happening here? Then you have p x 1 = px1+ this is pxm x mx1 and one more thing is here that is δ. So, you will write δ_1, δ_2, δ_j and δ_p, okay.

(Refer Slide Time: 21:49)



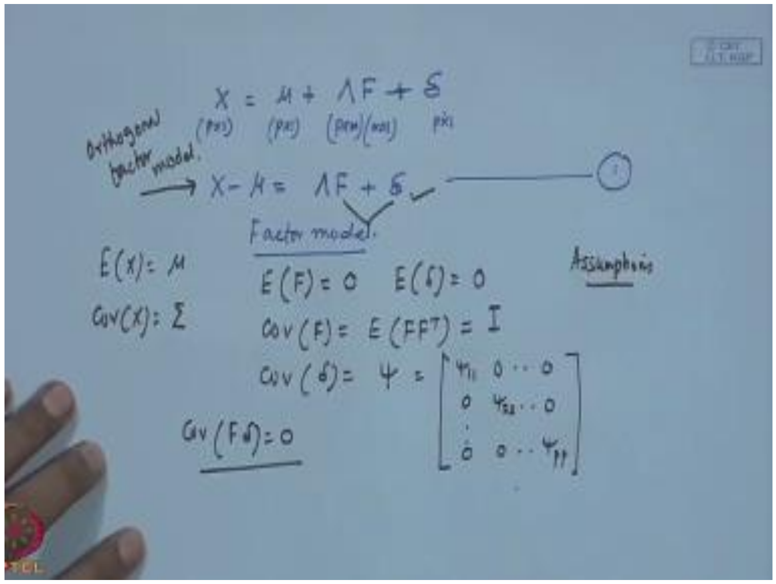
So now, I am writing this in matrix notation, X=μ+λ F+I can write this is δ. You can write like this. So, what will happen actually?

(Refer Slide Time: 22:10)



This one is your X, this one is your μ, this is your λ, this one is F and this one is δ.

(Refer Slide Time: 22:28)

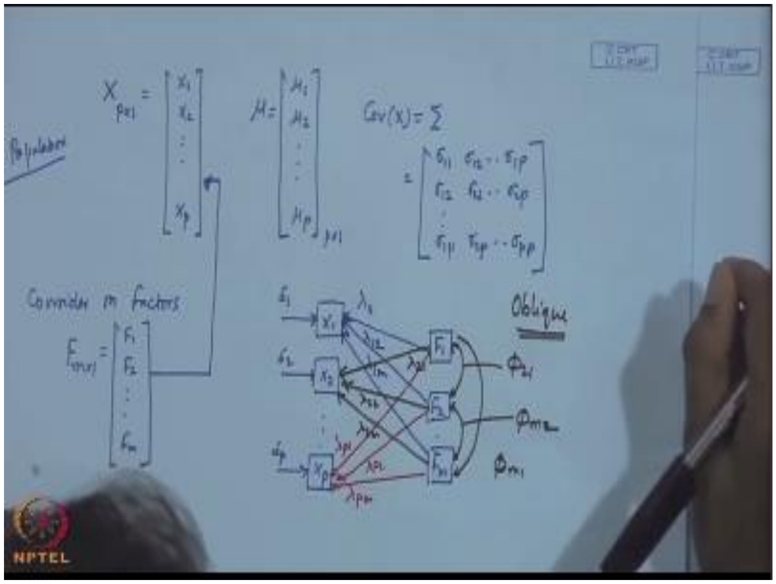


What is the result? Your dimension p x 1, p x 1, p x m, m x 1, p x 1 this is your desired vector. So, you can further simplify this $X - \mu\lambda F + \delta$. So, mean subtracted original variables is a function of the factors plus arrow. This equation if I say 1, this is the factor model okay. This equation is known as factor model, this equation will be further handled through several assumptions. I think all of you know that expected value of X is μ, covariance of X is ∑.

Now, what will be the, our assumption is expected value of F. This one is 0 the reason you say you have taken mean subtract it. So, expected value of F is 0 assumptions, expected value of δ that is 0 term that is 0 okay. Now, we want to make this type of things that covariance of F is expected value of FF^T because expected value if it is 0 this one, this will be I. And your covariance of another one, the δ that will be Ψ which is a diagonal matrix Ψ will be p x p and covariance of $F\delta$ that is 0 that mean F and δ this covariance is 0.

Now, these are what are the assumptions under, if this assumptions hold, then this factor model is known as exploratory factor model orthogonal definitely, exploratory plus orthogonal okay. So, here I think I should say little bit of as we are started with little bit of the types.

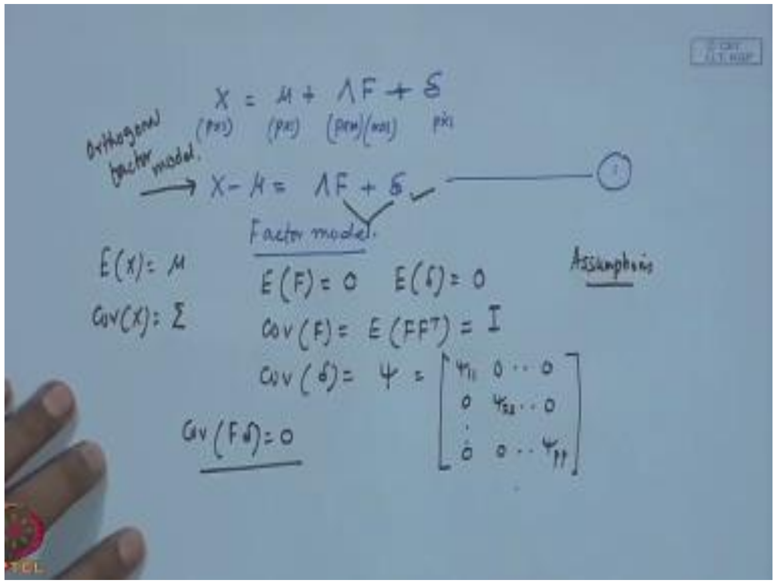
(Refer Slide Time: 26:12)



We are saying orthogonal because they are independent F1, F2, Fm they are independent, there will be there will be oblique factor model when you will say no they are not independent, they are each correlation between them, getting me? So, these correlation coefficient will be ϕ21, this will be your ϕm2, this will be your ϕm1. So, when you allow covariance component between the factors, then your factor model will be become oblique factor model.

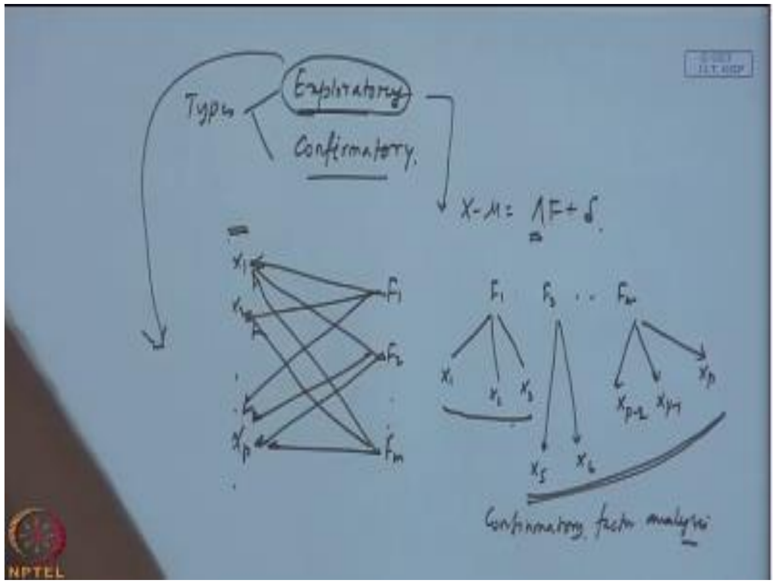
Now, this oblique factor model will be treated separately and we will discuss orthogonal factor model.

(Refer Slide Time: 27:04)



This is one.

(Refer Slide Time: 27:08)



Second one is important concept here is that we are saying the types of factor model types, one is exploratory factor model, another one is confirmatory factor model okay. Exploratory and confirmatory factor model what is exploratory and what is confirmatory? What is the difference between these two? See, in the first case we are saying that $X - \mu = \Lambda F + \delta$, here, we have taken all X and all F connected together, because of this δ matrix Λ, this capital Λ matrix, because this Λ matrix we have connected both.

That means what I mean to say, what I am trying to say that it is something like this. I have F1, F2 to Fm and I have X1, X2 to Xp. And what we are trying to say that each one of X, each one the X is dependent on F, all F. I do not know exactly which are the X values or which are the X variables that are basically coming out of F, which of the Fs I do not know. That is why what I am doing, I am allowing all the full sets of relationships here first hand.

And what you are saying? You are saying that okay this is my observed variable structure and you find out as many factors as possible. And then you say that okay what is this number of factors and what are those the factors that are linked to what type of variables. Under this condition you are exploring about the possibility of hidden factors, that is why it is exploratory factor analysis.

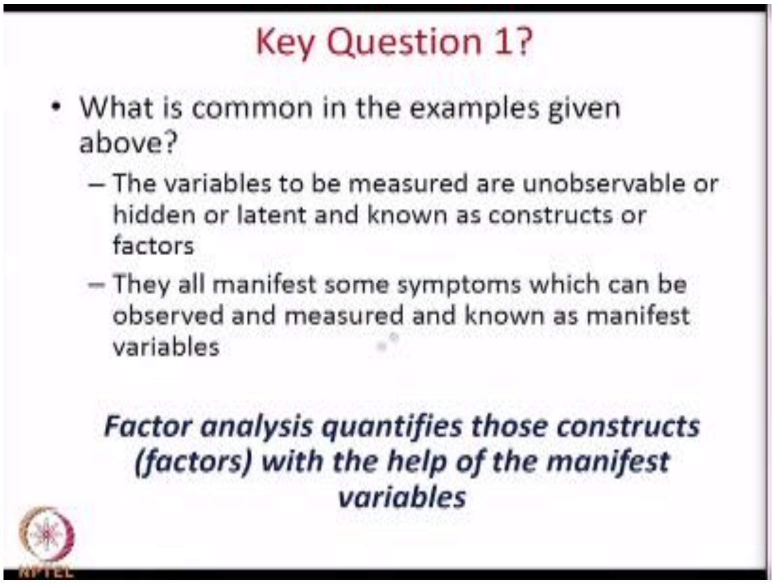
So, here also it will be linked this also will be linked to so everything is linked to I think this already done, getting me? Now, confirmatory factor analysis is different. What is then they are you say that F1, F2, suppose, m factors are there, but you clearly know what are the X1, X2 maybe X3, what are the variables x5, x6. So, let it be Xp – 2 Xp - 1 and Xp. You know clearly, you know that three variables are because of factor 1, another 2 manifest variable because of factor 2 this variable because of factor m.

So going in advance, because of your research work, your literature review and survey, all those things you know which are the manifest variable representing what type of which factor. So, essentially what is happening here? So, F1 to all other these variables the connectivity is not required, the relationship this is not required when it is coming from here only. Now this type of factor analysis is second, this factor analysis is confirmatory.

Why it is confirmatory? The reason is you are saying that there are suppose, if you manifest variable, let it be 3 or 5 related to F1 there are some other, some manifest variables related to F2. And you want to confirm whether you are correct or wrong. This is your hypothesis let it be then the first factor maybe something, let it be the supply chain coordination. So, there are certain indicators you are finding out and you are saying these indicator are measuring supply chain coordination, these indicators are measuring something else, something else like this.

Then through statistical route you are confirming that whether, what you have assumed or what you hypothesized what you have proposed, they are correct or false. That is your confirmatory factor analysis okay. So, we will stick to exploratory factor analysis now. Then I want to understand the purpose of exploratory factor analysis correct?

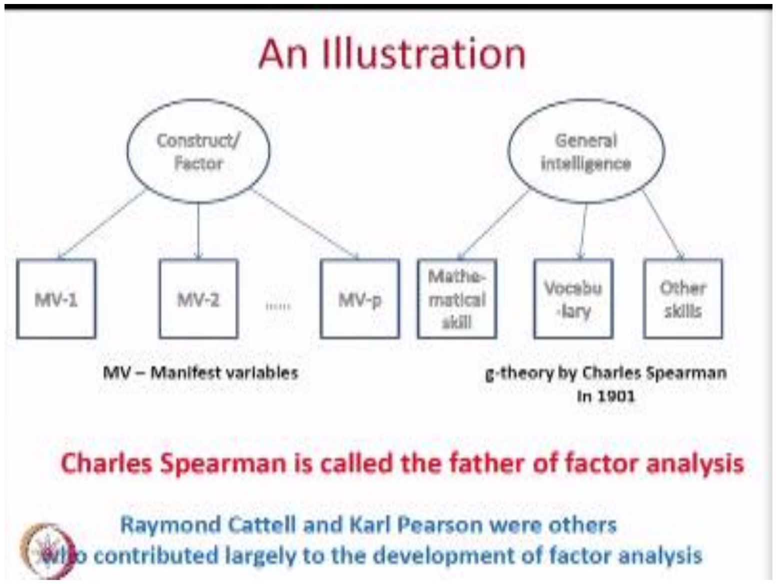
(Refer Slide Time: 32:36)



What is first question in factor analysis with respect to the examples given to you. So, I think around 7, 8 examples we have discussed. My question is, what is common in the examples given above? The variables to be measured are unobservable and you have seen that hidden or latent and known as constructs or factors. We have discussed, they all manifest some symptoms like in terms of X.

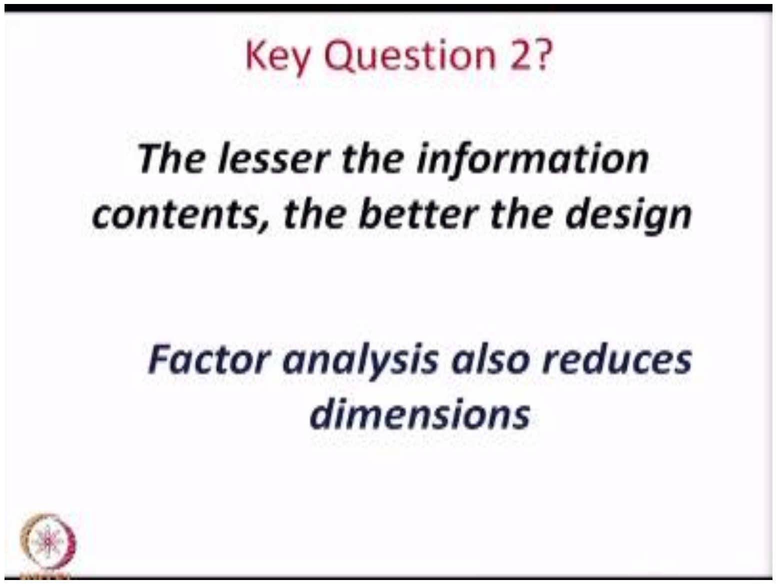
So, this is F, this one is X factor analysis quantifies those constructs or factors with the help of manifest variables. There is the first thing what we are doing in exploratory as well as in confirmatory also, you will be doing the same thing.

(Refer Slide Time: 33:28)



So, this is a illustration.

(Refer Slide Time: 33:30)

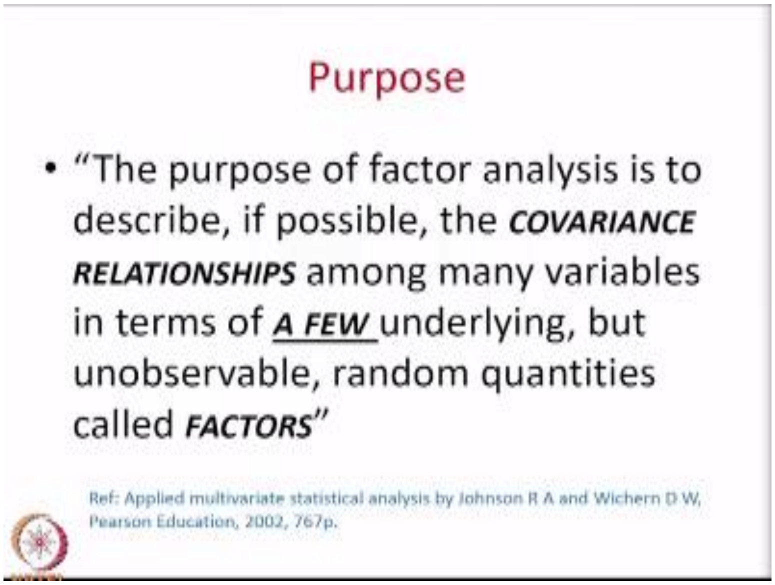


Now key question 2, I have a p number of variables, that mean p information, p dimensional information based. So I do not want this much of information. What I want? I want the reduced information, see if I require to know two dimensions only then I take decision. I act, I control system. All things will be much easier if there are five different dimensions, compared to that what will happen.

Suppose, you think of a pointer what we used to have in this pointer suppose, there are three buttons, one button to just point the lecture, one of the sentence or another one is just from you, just call from one slide to another slide and the third one is the switch off, this type of things are there okay. So, I must know these 3 important types information is known to me and I will do.

Now, suppose they are instead of these three, that n number of information available and second thing is thing is that if you required to go for slide change, you also required to do something else before, that is the dependent things are there also. So then what will happen? You will be confused. The same thing will happen if us, so, that is why what we are saying the lesser the information content, the better is the design, that means we want one inferent system or one recessive support system where our information will be at the reduced dimensions. So, factor analysis also do this far from this work.

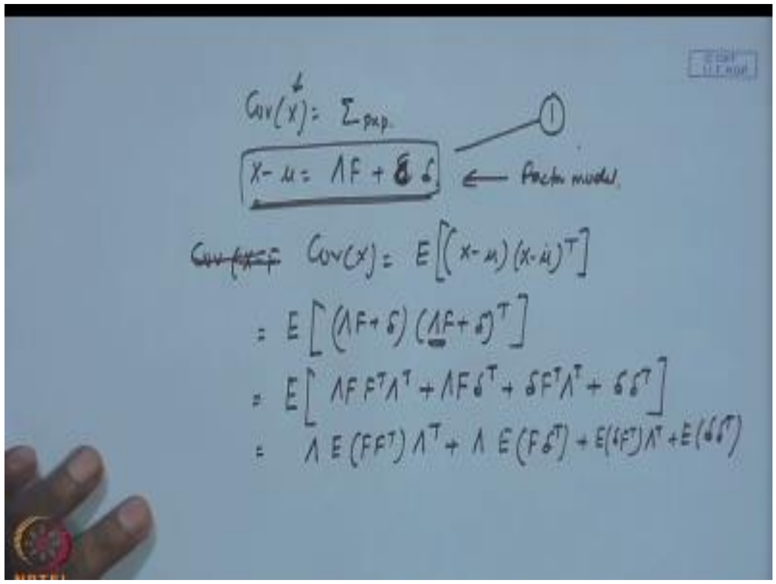
(Refer Slide Time: 35:30)



And what is information in factor your data set that is the covariance matrix. So factor analysis uses the covariance matrix. So, the purpose is the purpose of factor analysis is to describe if possible the covariance relationships among many variables in terms of a few underlying and but observable random quantities called factors that is the purpose.

So, we will, we are now in a position that we know what is construct or factor, what are manifest variables and what is the exploratory factor model and what are the different assumptions related to this factor model okay. So now, the crux of the matter is variability explanation.

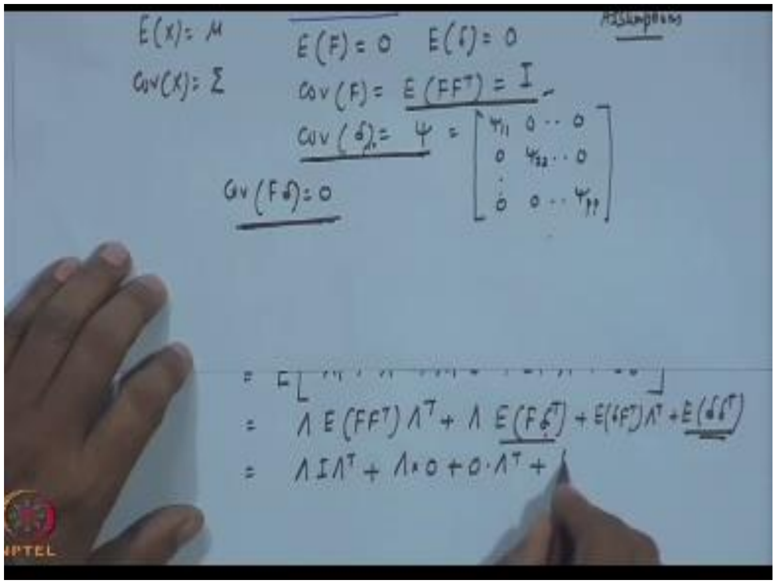
(Refer Slide Time: 36:36)



We are saying that information is covariance of X which is $\sum$. So, you are creating a factor model which is X – μ = ΛF+δ we are using ΛF+δ getting me? This is your factor model, I want to know that what is the variability you are able to explain using this model. What whose variability of X what is that, that is the covariance of Xp x p. So, if I want to understand covariance, can I not write like covariance of, sorry covariance of X equal to expected value of (X – μ) (X-μ^T).

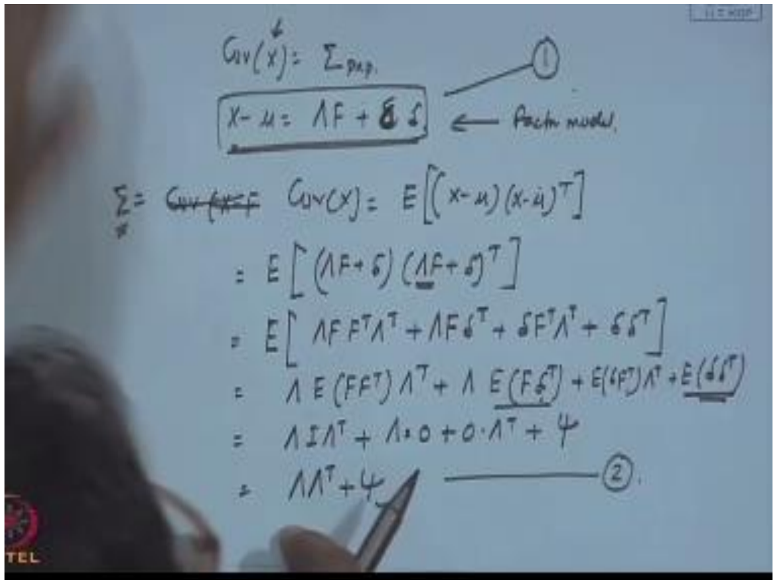
We have seen earlier also, that mean this quantity we can write like this X – μ, suppose, if I say this is my equation 1, from equation 1 I can write ΛF + δ x ΛF + δ^T, X-μ= ΛF+ δ, Λm- δ, so I am writing like this ΛF into this one. So, this one will be F^T Λ^T+ ΛFδ^T+ $\delta F^T \Lambda^T$+$\delta\delta^T$ okay. Now that Λ is the regression like constants, you have seen earlier. So, this constant will come, so, ΛE (FF^T) Λ^T+ΛE (Fδ^T) + E (δF^T) Λ^T + E ($\delta\delta^T$). Now, you have to see the covariate this assumptions okay.

(Refer Slide Time: 39:21)



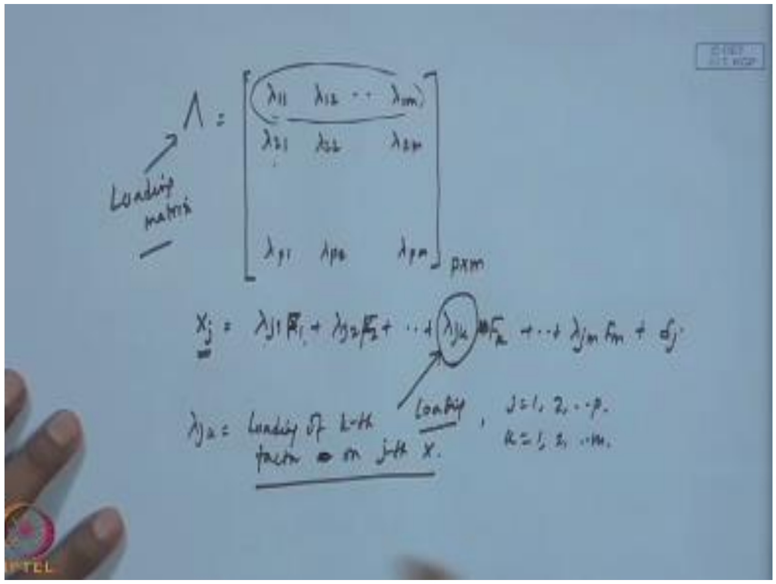
So, I am putting the assumptions parallely here. So, what we say assumption is expected value of FF^T is I. So, I will put like this ΛI, Λ^T+Λ expected value of F, and δ what we have written related to F and δ this one this is 0 so, this into 0. So similarly, this 0 into $\delta\Lambda^T$+ what is the covariance of delta, that is expected value of $\delta\delta^T$ so this one so, this is ϕ.

(Refer Slide Time: 40:15)



So then what is happening to the resultant matrix. Now, capital ΛI, IΛ is $\Lambda\Lambda^T$ + this what is this? Which is your ∑ value? Covariance of X is capital ∑, this is equation number 2 and this is the variability explanation point of view this equation. So, that means if you know this Λ and then Ψ, you know this, getting me? What is this Λ here?

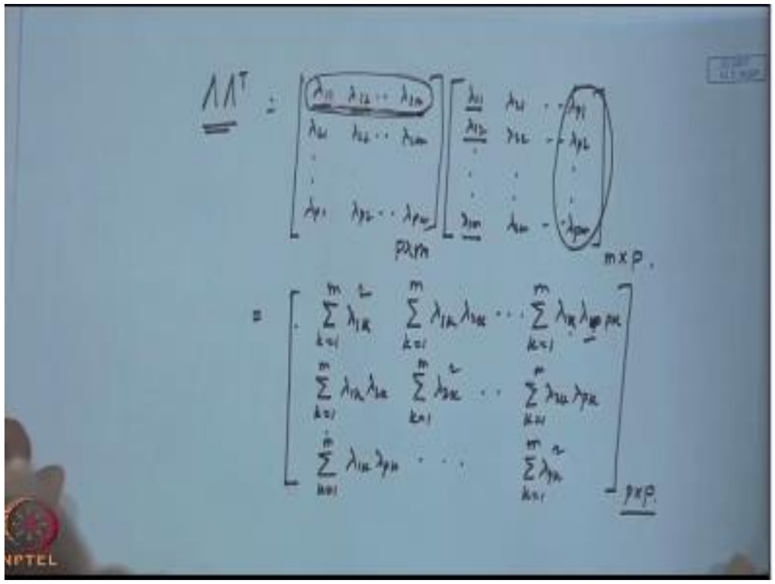
(Refer Slide Time: 41:07)



This one is nothing, but can you remember what we have given λ_{11} F_1, λ_{12} F_2, like this $\lambda_{1m}F_m$, $\lambda_{21}F_1$, $\lambda_{22}F_2$, $\lambda_{2m}F_m$, like this. So, this λ part I am writing now. So, λ_{p1}, λ_{p2}, λ_{pm}, this one is p x m matrix. What is this λ basically? Suppose, this one their loadings that λ is, so if you take a general formula like X_j, what you have written λj11121m, yes j1 X_1 λ_{j2} X_2 like this if I write λ_{jk}, your jk, sorry λX_j you write λ_{j1} F_1 λ_{j2} F_2, λ_{jk} F_k. So, similarly λ_{jm} F_m + δ_j, this is the original equation you have λ_{j1} F_1 like this.

So, see what is happening here this general one is λ_{jk} this is the general loading. So, it will be for j = 1, 2… p k = 1…m. Now, this one is known as loading matrix. So loading matrix mean this is really factor loading matrix and each λ_{jk} they are the loading of kth factor on jth variable X variable okay this one is λ_{jk} is known as loading of kth factor, on kth factor, on jth X, correct?

(Refer Slide Time: 44:00)

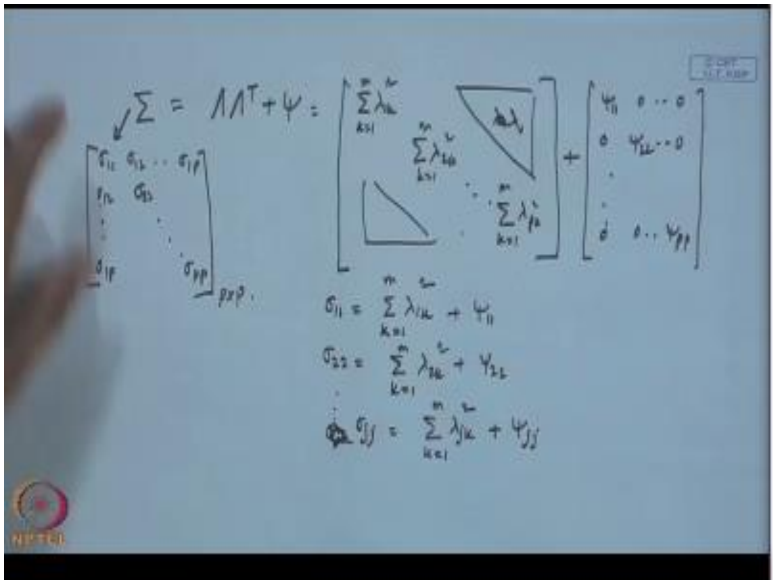


Now, what is your Λ^T so if you write down this, you have to write down $\lambda11$, $\lambda12$, $\lambda1m$, 21, 22, 2m, p1, p2, p m, then transpose 11, 12, 1m, 21, 22, 22, 2m, p1, p2, pm. This is your p x 1, this is now m x p and you must get p x p because this p x m, m x p okay. So, if you just multiply the matrix $\lambda11$ x $\lambda11$, $\lambda12$ x $\lambda12$, $\lambda1m$ x $\lambda1m$, what is changing here? This one remain constant, this one that let k I am giving k is basically 1 to m, the factor is changing k = 1….m, and the square $\lambda11^2 + \lambda12^2+\lambda1m^2$ like this.

Then second one will be again k = 1 to m, you see $\lambda11$ to 1m and 21 to 2 m. So, that mean $\lambda1k$, $\lambda2k$, the covariance part is coming here. So, same manner k = 1 to m $\lambda1k$ and λ, no, k is 1 to m. Now, it is 1, then 12 then 1p, this is p x p matrix, so $\lambda1k$, λpk okay. λ this row be this column, this is coming here. So, same manner now this will be a symmetric matrix, k = 1 to m, $\lambda1k$, $\lambda2$ k this will be $\lambda2k^2$, k = 1 to m the same manner, $\lambda2k$, λpk, k = 1 to m.

In the same way it will come k = 1 to m, $\lambda1k$, λpk, same manner you go it will be λpk^2, k = 1 to m. This will be p x p okay So, this diagonal element is the variance, component of diagonal element is the covariance component. And this diagonal element relates each of the variables also.

(Refer Slide Time: 47:57)

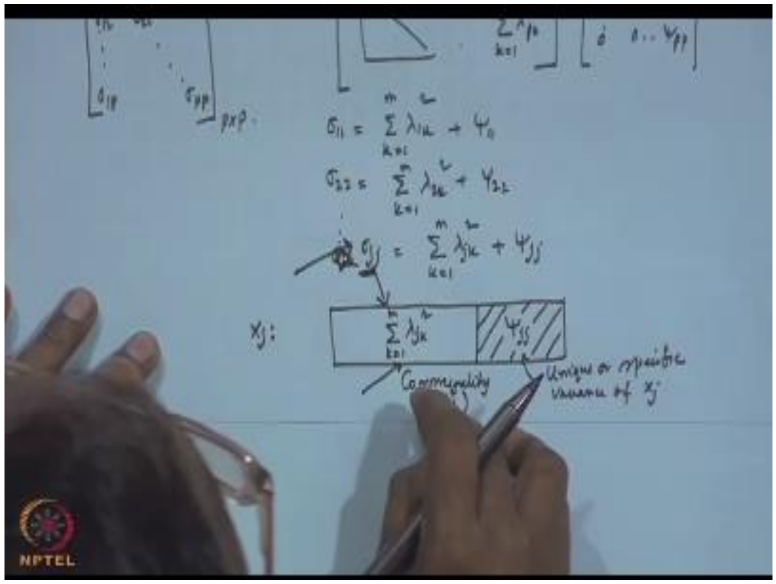


Then if I write down further I can straight away write that λ, λ^T + Ψ if I write this is nothing, but I am not writing the off diagonal element. Now, only diagonal element I am writing $1k^2$, k = 1 to m, $\lambda 2k^2$, k = 1 to m, λk = 1 to mpk^2. Then off diagonal elements, already I have written with λ1k, that λk and something like this. Already I have written this plus our Ψ is Ψ1100, 0Ψ220.

So, like this 00 Ψpp so this is what is your ∑. So, if I write this ∑ in the same manner, the way we have written in matrix. So, that means this one will be ∑11, ∑12, ∑ 1p, then ∑12, ∑1p. Now, ∑22, ∑pp, the ∑ is like this. So, this is p x p, then if you compare this versus this diagonal element what you will get? σ11 = sum total of $\lambda 1k^2$, k = 1 to m, correct + Ψ11, yes or no? So, similarly, you will be getting σ22 equal to k = 1 to m $\lambda 2k^2$+ Ψ22.

So, in the same manner if I want to understand here suppose, jj let it we will write now, jj because we are interested in the variance components jj. So, what will happen here, then k = 1 to m λjk^2+ Ψjj.

(Refer Slide Time: 50:39)



Then actually what is this? See, now try to understand the variability part. Our variable is X_j in this equation. Corresponding variable is X_j, it has σ_{jj} that is the total variability. So, I can say this total is σ_{jj}. Now, this is divided into two parts, if I say this is the case where this portion k = 1 to m, λ_j^2 and this portion is Ψ_{jj}. So, what I am trying to say then that σ_{jj} is partitioned into two parts, correct?

This one is what λ_{jj}, you see you are extracting m factors and $\lambda_{j1}^2 + \lambda_{j2}^2 + \lambda_{j3}^2$, like this up to λ_{jm}^2. So, this is the contribution by the m common factors, this is the contribution by the m common factors and this Ψ_{jj} is something which is not contributed by the common factors. That is why this one is known as communality and we denote it in terms of h_j^2 communality and this one is unique to x_j the common factors are not able to explain this much of variability, this is unexplained, this one is unexplained one, this is explained by the common factor this is unexplained.

So, we can say unique, we say unique or specific variance of X_j okay. So, it can be like this, so unique and specific. Again some you can argue that this portion which is related to X_j precisely related to X_j plus something which are, because you have taken m factors and some other maybe variable correlation, some other variable you have not considered, your data collection many things. So, error also maybe included here, but in factor analysis term, we say usually we denote like this.

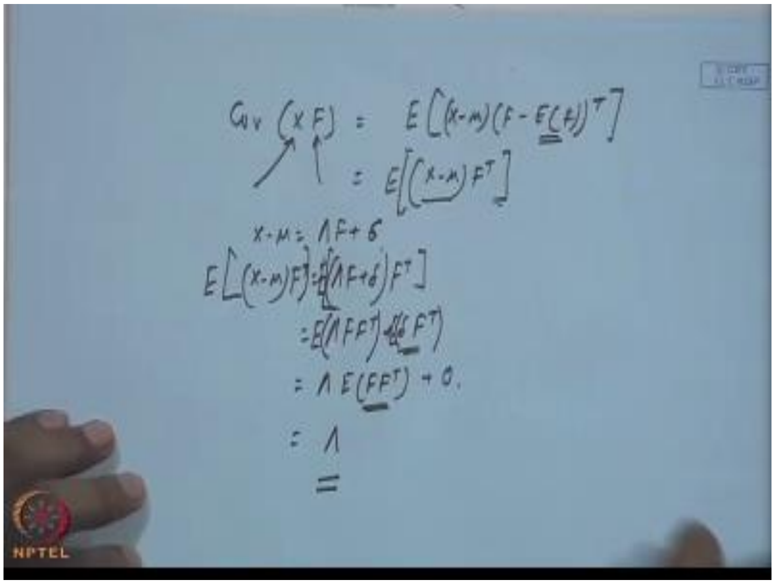
The variability of Xj is because of the communality and because of the uniqueness of that variable, correct? So, what you want then you definitely want to maximize the communality.

Speaker 2: The password is the variability with other components.

Common factors contributing towards explaining the object variables variability. So, that means what is the variance part from what is the contribution for the first factor $\lambda 11^2$. What is the contribution towards variability to the second factor, that is what we are saying that second factor λ1 that it is variable Xj, then $j1^2$ then $j2^2$, $j3^2$ λ that all λ^2, these are contribution from each of the factors in explaining to the variability of the Xj.

And something you will not be able to explain. You can explain if you suppose you extract p factors out of px variables okay everything will come under this communality, but that is not the purpose of factor analysis. Factor analysis, one end wants to produce the dimensions. Other end it will do in such a manner that the factors can be linked with the original variables and naming of factors can be possible, getting me?

(Refer Slide Time: 55:46)



Now, with respect the another one concept, here what will be the covariance of X and F, X is your original variable, F is the factor okay. So, your expected value of $X - \mu$ and then F - expected value of F^T that will be the case. Expected value of F is what, 0. So, ultimately this is expected value of $X - \mu$ F^T. Now, what is $X - \mu$, $X - \mu$ is $\lambda F + \delta$. So, then $X - \mu F^T$ is λF δ F^T so λFF $\delta +$ δF δ.

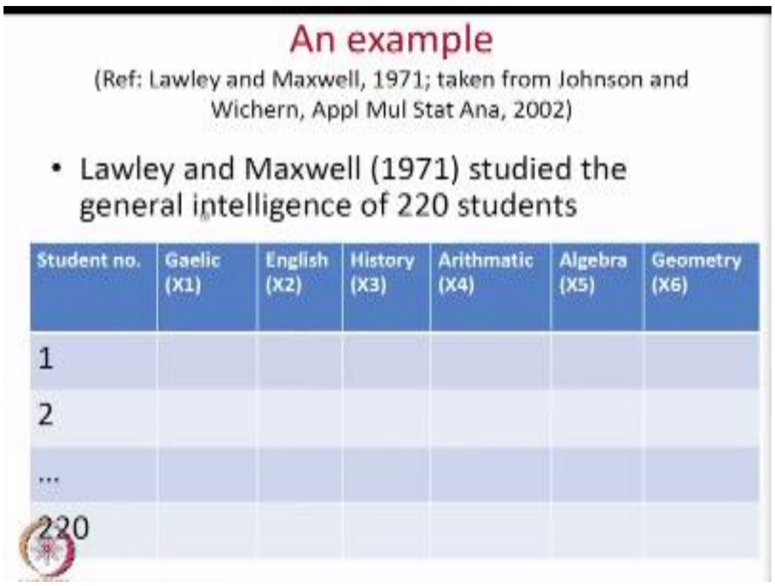
Now, if I take expected value of all those things. So, ultimately this equal to expected value of this equal to expected value of this, expected value of this, it will δ will λ will come expected value of FF^T + this one will be 0. Now, FF^T expectable is I loading, that loading matrix is coming into consideration okay. I think that to we have explained the primarily, the concept conceptual model of factor analysis.

This whatever we have discussed it is just the conceptual model, that what is factor analysis, what is exploratory factor analysis and how the variance of individual object variable have partitioned into communality and specific factors, and but what is required. Definitely our next topic will be estimation.

Speaker 2: Estimation of Λ.

Λ, that Λ matrix you do not know this. This Λ matrix is not known to you, you have to estimate. So there are different method of estimation okay. But I thought of giving you one example. Let us see one example, here.

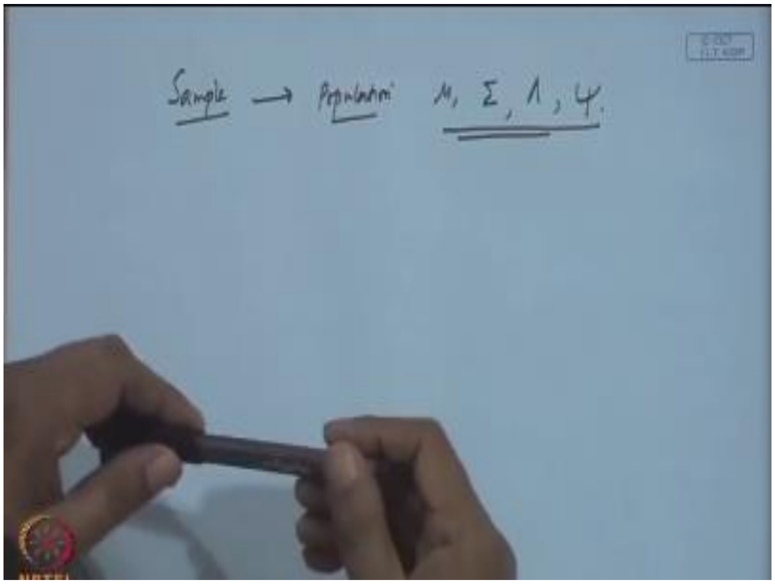
(Refer Slide Time: 58:35)



This example with reference to the factor analysis. This one is a study conducted by Lawley and Maxwell in 1971, the purpose is to measure the general intelligence. So, for students 220 students were considered and the variables that were taken, they are object variables. So, these are the verbal of first one Gaelic, English, history. These are basically the verbal subjects and arithmetic, algebra and geometry, these are your mathematical subject.

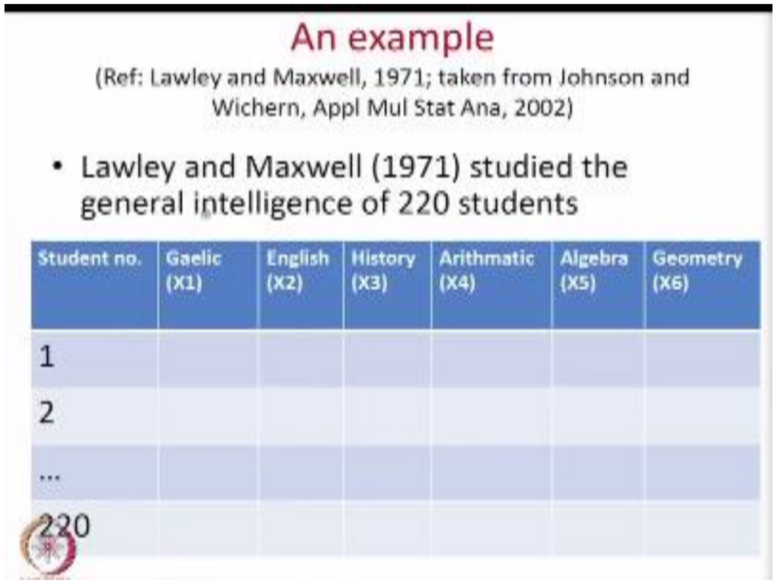
What happen you do not have to test the general intelligence of these many students. So, this six subject performance that mean performance of each student on the six subjects were what I guess is obtained, and then with these data set factor analysis was conducted. And to see that whether that the six X variables, what you are observing they are actually six or these can be combined into several reduced factors number of factors okay. So, that means what is happening here that we are talking about sample.

(Refer Slide Time: 01:00:10)



But so far what I have discussed? We have discussed for population, because we have taken μ, we have taken ∑, we talk about Λ, we talk about Ψ, all those things related to, they will be during estimation time we will see this sample part. But let us concentrate on this example now.

(Refer Slide Time: 01:00:41)



(Refer Slide Time: 01:00:43)

An example (contd.)

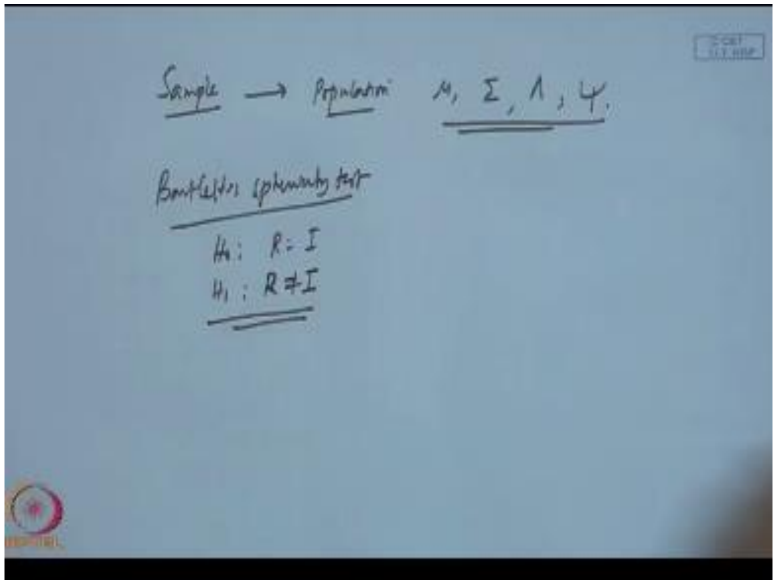
- The correlation matrix

	X1	X2	X3	X4	X5	X6
X1	1.00	0.44	0.41	0.29	0.33	0.25
X2		1.00	0.35	0.35	0.32	0.33
X3			1.00	0.16	0.19	0.18
X4				1.00	0.60	0.47
X5					1.00	0.46
X6						1.00

Then the immediate step you have to see the correlation matrix. You see that if I go by the correlation coefficient between the variables, there are few variables point with correlation more than 0.3. For example, X1, X2, 0.44, 0.41 like this 0.60 is the maximum. So, from correlation matrix also you can find out certain clues that whether you go for factor analysis or not, that factor ability criteria that mean the data matrix must be factorable.

If you have seen in principle component analysis, I say that Bartlett Sphericity Test what we have assumed there?

(Refer Slide Time: 01:01:43)



Bartlett Sphericity Test we assume H0, that R is I and H1, R not equal to I, that mean the correlation coefficient you have tested and combining you have tested this okay.

(Refer Slide Time: 01:02:09)

An example (contd.)

- The correlation matrix

	X1	X2	X3	X4	X5	X6
X1	1.00	0.44	0.41	0.29	0.33	0.25
X2		1.00	0.35	0.35	0.32	0.33
X3			1.00	0.16	0.19	0.18
X4				1.00	0.60	0.47
X5					1.00	0.46
X6						1.00

So, here I think we definitely require Bartlett Sphericity test to see that whether we will go for factor analysis or not, because if there is no, that not significant to much correlation what is the use. But factor analysis from correlation co coefficient point of view, if there is 0.3 or more correlation coefficients, a large number of correlation coefficients having 0.3 or more, then it is recommended, you can go for factor analysis and you will get certain benefits.

So, we will stop now, and next class I will explain our parameter estimation using that loading, how to estimate the loading matrix okay. Thank you very much.

NPTEL Video Recording Team

NPTEL Web Editing Team

Technical Superintendents

Computer Technicians

A IIT Kharagpur Production

www.nptel.iitm.ac.in