INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NPTEL

NPTEL ONLINE CERTIFICATION COURSE

Marketing Research

Lec -28
SEM & CFA

Dr. Jogendra Kumar Nayak
Department of Management Students
Indian Institute of Technology Roorkee

Welcome friends to the section of marketing research and analysis where we have been discussing about structural equation modeling and the which talks about the interrelationship or the you know multiple relationships and what how one can you know derive inferences out of it in which we just started with what is a structural equation model and how it is connected with factor analysis and multiple regression and χ2 right.

Being some of the important tools used to understand structural equation modeling the we spoke abbot 2 important things one being the measurement model and the structural model so the measurement model which we say we start with when we talk about SEM CFA as 2 important parts you can understand that way.

(Refer Slide Time: 01:15)

So one is the measurement model right here we explain the you know the relationship of the variables with the factor right so a researcher tests it is a it is also called as confirmatory factor analysis right so where you are conforming as a researcher you are conforming the relationship between the variables and the factor right so this is already a priory you know it is something that the researcher as already defined the relationship right.

And now he is only conforming it so that is why it is a confirmatory otherwise it would have in a exploratory so you are not exploring anything you are only trying to conform right weather the relationship theoretically that you have found out or some experiment or some kind of a discussion some export opinion whatever so whatever you have found now is it coming true or not right are the variables loading on to the factor probably nicely or not that is what we do basically right and the second part once you have done it once you have done it second part is to test the structure.

The structure is like for example structural model so here this is something like a path analysis we say right visually visual deception would is called a path analysis path analysis so in a path analysis what we are doing is basically we are trying to formulate some relationships right A, B let us say is affecting let us say C okay so A and B are affecting C so C in this case is my dependent variable and A and B together are affecting my dependent variable.

So this is a case which we will measure right so this could be a one thing but this could this looks like a simple you know a multiple regression but what if the relationship becomes little more

complicated let us make it little more complicated right so it is become something like this now C is then affecting D and now this relationship has become more tan a multiple regression right so you can if you try to explain it through a multiple regression you can.

But then we will not get the you know you will not be able to understand the measure the error terms probably in this case right so let us understand and what is this CFA first and then we will go across.

(Refer Slide Time: 03:50)
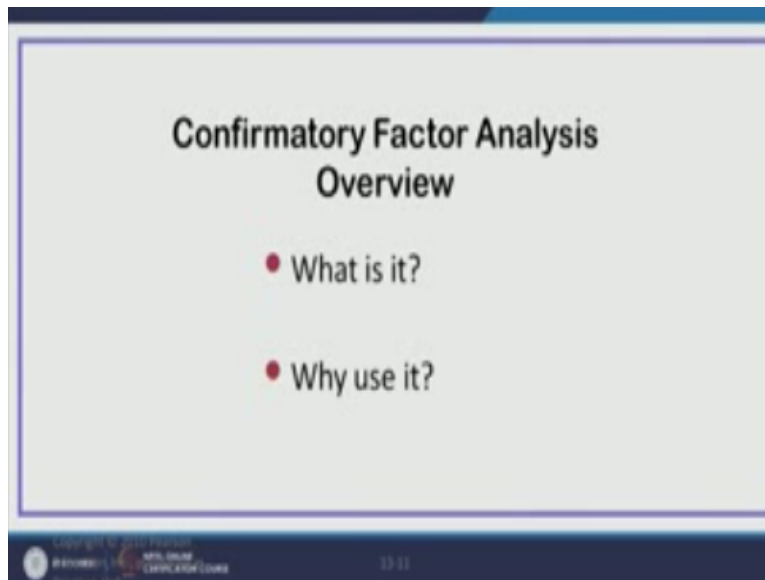
So confirmatory factor what is it? And why we use it?

(Refer Slide Time: 03:52)

DEFINITION

- Confirmatory Factor Analysis ... is similar to EFA in some respects, but philosophically it is quite different. With CFA, the researcher must specify both the number of factors that exist within a set of variables and which factor each variable will load highly on before results can be computed. So the technique does not assign variables to factors. Instead the researcher must be able to make this assignment before any results can be obtained. SEM is then applied to test the extent to which a researcher's a-priori pattern of factor loadings represents the actual data.
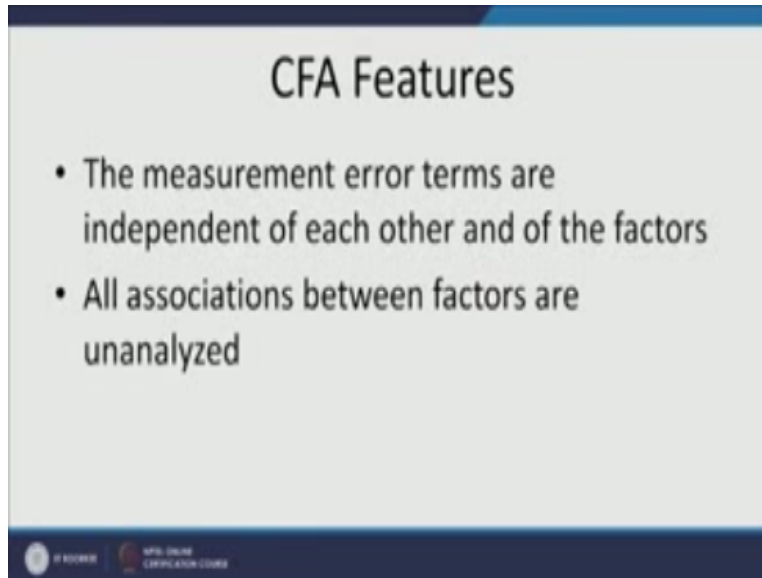
Let us say so the definition says it is similar to EFA is some respects but philosophically very different why it is philosophically different I had already explained that there is a already theoretical backing we see a researcher must be specify both the factors that exists within a set of variables and which factor each variable will load highly on before results can be computed that means the researcher tells it before hand okay what will be the relationship without even testing it right because the testing can be only be done later on.

So the technique does not assign variables to factors instead the researcher must be able to make this assignment before any research can be obtained SEM is then applied to test the extent to which a researcher priory pattern of factor loading represents the actual data. So in this case what is he doing is trying to say whether what I have though is true or not true right features so what are the features the measurement error terms or independent.

(Refer Slide Time: 04:57)

Of each other and of the factors right so the error terms are independent of each other now what are the error terms so we are saying so this is a construct so and this is let say the error term right and this is another construct and this is it is error term let say right.

(Refer Slide Time: 05:14)

So now the error terms are independent of each other right and all association between the factors are unanalyzed so we will basically we say it is a covariance structure so covariance kind of a relationship it is a special correlation right, so and this is what exactly is happening right but there are cases to improve the model we will look into the measurement errors also we will see okay whether we can try to connect the will do some is there any relationship between the errors terms to and by doing that will the model significantly affect have a go positively will right.

(Refer Slide Time: 06:03)

EFA vs CFA

- CFA is confirmatory. The number of factors and the pattern of indicator factor loadings are specified in advance.
- CFA analyzes the variance-covariance matrix of unstandardized variables.
- The prespecified factor solution is evaluated in terms of how well it reproduces the sample covariance matrix of measured variables.

CFA confirmatory the number of factors and the pattern of an indicator factor loadings are specified in advance right, so for example we will say that variable one will have a larger varying or larger effect on the factor right, so we have defined already CFA analysis the variance covariance matrix of un standardized variables, now the last thing is says is it is pre specified factor solution is evaluated in terms of how well it represent reduces the sample covariance matrix of measured variables.

So now so what is CFA basically doing is in CFA we are trying to build a model and test this model okay whether as per our knowledge whether the model whether variables are loading properly and the construct is a valid construct or not this is what we are basically doing here right.

It fixes cross loadings to 0 that means what now let me draw it and show you what does how it would look like right generally I have seen I have said what I have said is now this is a construct and this is let say two or three variables let say and this is another construct and this has got three variables and this is another construct and now we are saying that these variables are independently fix to only certain factor, but what if one factor is loading into another construct let say one factor is loading to another construct so that is a cross loading case of a cross loading.

(Refer Slide Time: 07:43)

EFA vs CFA

- CFA models fix cross-loadings to zero.
- EFA models may involve cross-loadings of indicators.
- In EFA models errors are assumed to be uncorrelated
- In CFA models errors may be correlated.

EFA models mean involve cross loadings of indicators in EFA that provision is there but in CFA cross loadings are fix to 0 that means they treated as not existing right, okay and errors are assume to be uncorrelated this is very important now what did I say in EFA model the in this beginning of my last session also I had said this I believe I remember that in EFA the cross loading the error terms are uncorrelated, but in CFA the errors because a error is not an error by the word which we understand a error is nothing but the unexplained.

The unexplained variance right so why cannot there be a relationship between the unexplained variance right, so in case there is an relationship that would surely affect my model right, so whether it is used.

(Refer Slide Time: 08:46)
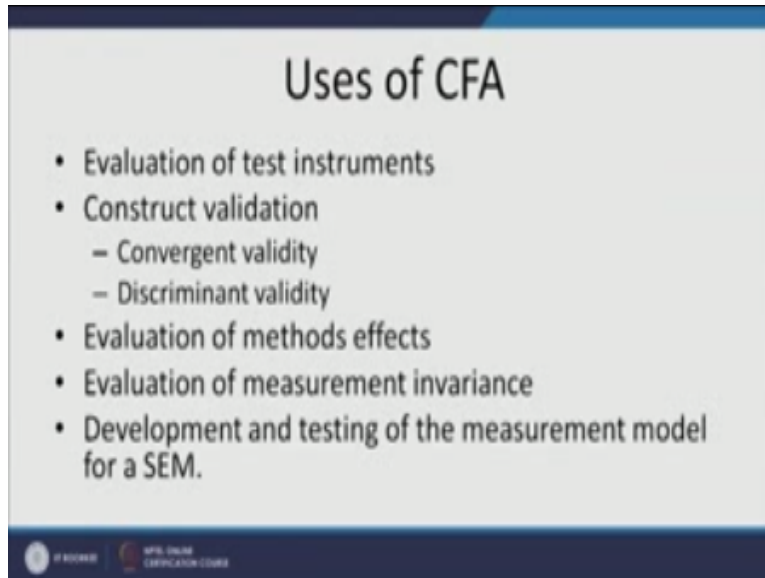
To evaluation of test instruments that means test instruments are my constructs to evaluate whether my instruments that I am using the construct that I am using are valid enough or not right the validity construct validation is done by convergent validity and discriminate validity now convergent validity as I said now suppose these are the variables right so this variables are my now suppose the variables will have a higher convergent validity of the inter item correlation that means.

If the correlation among the variables sufficiently high right that means they are correlated with each other right on the other hand discriminate validity is something that is something that is used to measure that two different constructs are separate and distinctly separate from each other right how do we do it there are methods which I will be explaining in this that two popular methods which we use to do it, one is the basis on basis of the variance extracted and the other is to the chi square right I will be explain it so development and testing of the measurement model for an SEM.
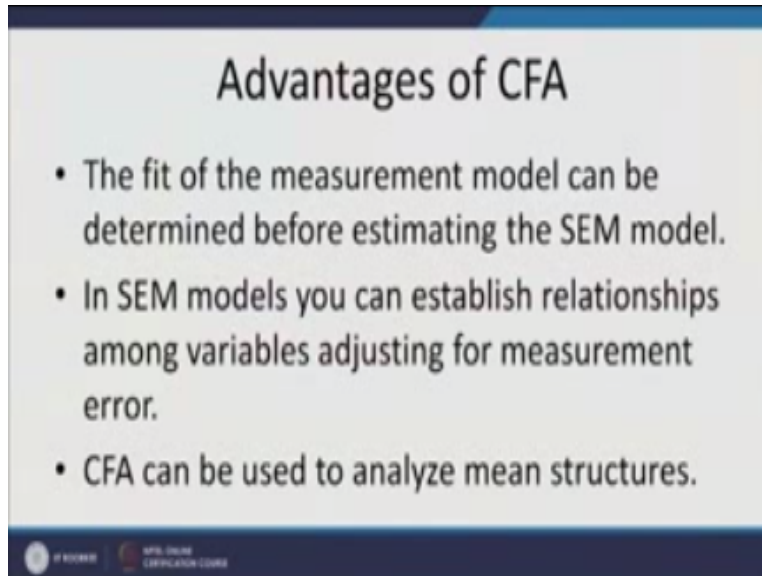
(Refer Slide Time: 10:00)

Uses of CFA

- Evaluation of test instruments
- Construct validation
  - Convergent validity
  - Discriminant validity
- Evaluation of methods effects
- Evaluation of measurement invariance
- Development and testing of the measurement model for a SEM.

For the first part is to see that the measurement model is proper or not, right.

(Refer Slide Time: 10:06)

Advantages of CFA

- The fit of the measurement model can be determined before estimating the SEM model.
- In SEM models you can establish relationships among variables adjusting for measurement error.
- CFA can be used to analyze mean structures.

What are the advantages of CFA, CFA the fit of the measurement model it helps in measuring in fit of the measurement model, before estimating the structural model that means before you get into the structural model or the relationship right, among the construct before that you can already confirm that my constructs are valid right, are justified relationship among variables with by adjusting for measurement errors one more important thing is in SEM models you can establish relationships by adjusting for the measurement error.

Now this is what I said we sometimes we will use the term called modification indices right, through the modification indices we try to reduce the measurement error right, by not reduce we adjust the measurement error by creating a relationship among the variables the error terms, right.

(Refer Slide Time: 11:03)

FIGURE 13-1 Visual Representation (Path Diagram) of a Measurement Theory
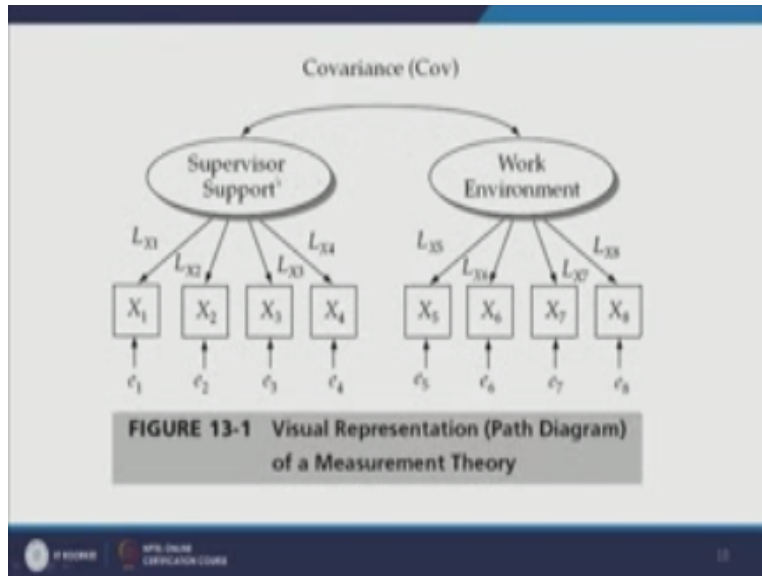
Now look at this, this is a simple case this is two things of the two constructs right, supervisors support work environment right, maybe this is the researcher is trying to is the measurement theory he is trying to find out relationship between these two constructs latent constructs right, supervisor support and work environment now these four are the variables for the first case and x5 to x8 are the variables for the second case work environment.

Now these variables are already pre-defined as I have already said in many a times I am repeating using it has come out theory prior law knowledge, literature and they are pre-defined okay, and these you can see e1, e2,e3, e4 are all my unexplained the error terms, right. So each variable will have a error term obviously because their some part which is explained, there is some part which is not explained okay.

Now what is this Lx1, Lx2, Lx3 and Lx4 till Lx8 these are all my loadings that mean the correlation of the variable with the factor right, that particular variable with the factor. So Lx1 is a correlation between the x1 and the factors supervisor support, right.

(Refer Slide Time: 12:28)

So construct validity it one of the biggest advantages of CFA SEM is ability to assess the construct validity. Now construct validity means the construct that you are using that should be valid right, if somebody has used somewhere it is not necessary that it will valid in a new environment all together, right. So construct validity is the extent to which a set of measured items actually reflect the theoretical latent construct their design to measure.

Now please you can think about it again you can read and think about it, what he is saying construct validity is the extent to which a set of measured items the variables which I said actually reflect the theoretical hidden construct latent is hidden, hidden construct they are designed to me that means they are actually measuring what they are intended to measure, right.

So construct validity is measured through the factor loadings right, through a high factor loading is a good indicator the variance extracted now variance extracted means how much of variance the mean variance extracted the average variance extracted we say, the variance extracted in any study in any among the variables and the construct will help us to understand okay, how much of variance has been explained.

If it is we have a cut value of 0.5 the AV value average where is extracted is 0.5 so any value above 0.5 is good but value below 0.5 means that it is not explaining properly, right and the third is the reliability for which we use the in a case of generally in a exploratory factor analysis we were using the Chrome back alpha, the Chrome back alpha was a reliability value which was beside that it has to be more than 0.7 in social science.

But harden understand in their book they have been very clear and they have said okay, even in social science a 0.6 60% explanation is good enough, right and they something call a composite you know so composite reliability right, so composite reliability is a score which we do measure now what is a, how it is measured.

In a λ^2 which is a composite or reliability is equal to $\lambda^2/\lambda^2+1-\lambda^2$ so λ is nothing but the factor loading value. If you know this factor loading then you can obviously tell it from there this square of it and $1 - \lambda$ is the error right so you can measure it right so discriminate validity which I said we will see right so let us go in to this so standardize loadings should be 0.5 or higher in ideally 0.7 or higher right so what is that saying that means still loading that you see here these loadings the values right the estimates you can say in one way.

So this loadings this should not happens so I am deleting his part there should not be any cross loading so the loading should be 0.5 or higher if the loading is 0.5 what does it mean that means if my correlation is 0.5 I am explaining 25% if it is 0.7 I am explaining almost 50% right average variants exactly should be 0.5 which I have just said should be 0.5 or greater to suggest adequate convergent validity if the heavy score which is average variants extracted right which is also even it is λ2 / n basically.
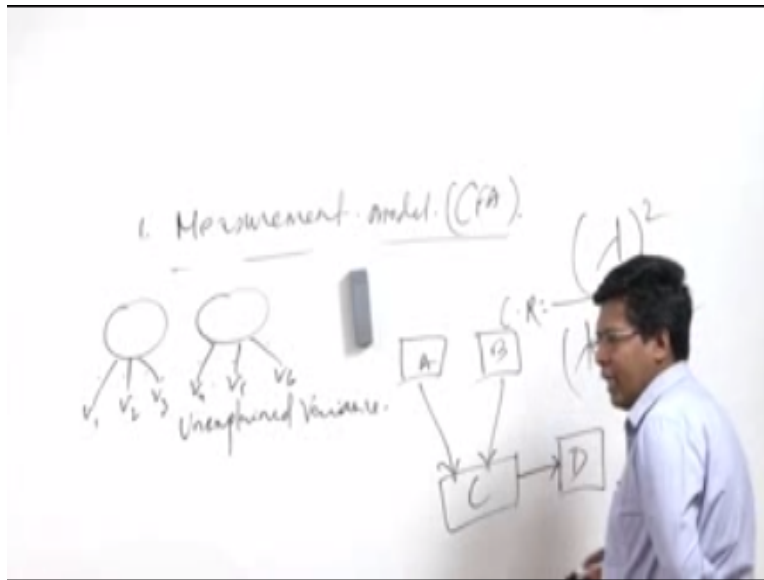
So if you measure this take out this heavy it should be more than 0.5 heavy estimates for two factors also should be greater than the square of the correlation now this is how you check the discriminate validity this part is an evidence of the discriminate validity. Now this is where students gets confused researcher get confused a lot now what does it mean? It means that is when there are two constructs right the heavy is score as it say the heavy estimates for two factors that means factor a f1 f2 let us say right so f1 f2.

Now the heavy estimates should be right suppose this is the correlation let us say R right so the heavy estimate should be greater than the square of the correlation that means heavy should be greater than the square of the correlation among them right so if this is happening then the research this is the very strong way this is very sufficiently strong way of even you can write in your research papers that if you have achieved this then the researcher has achieved a discriminate validity.

There is one more way of doing our discriminate validity also through the factor model where we looking in change in the chi square value which I will show which I will tell you later on. So construct validity is construct reliability the one should be 0.7 which the formula which I have written should be 0.7 or higher to indicate adequate convergent or internal consistency, so if this value is greater than 0.7 is greater t4hena 0.7 or 70% then you can assume that you are achieving a high internal consistency right.

(Refer Slide Time: 18:32)
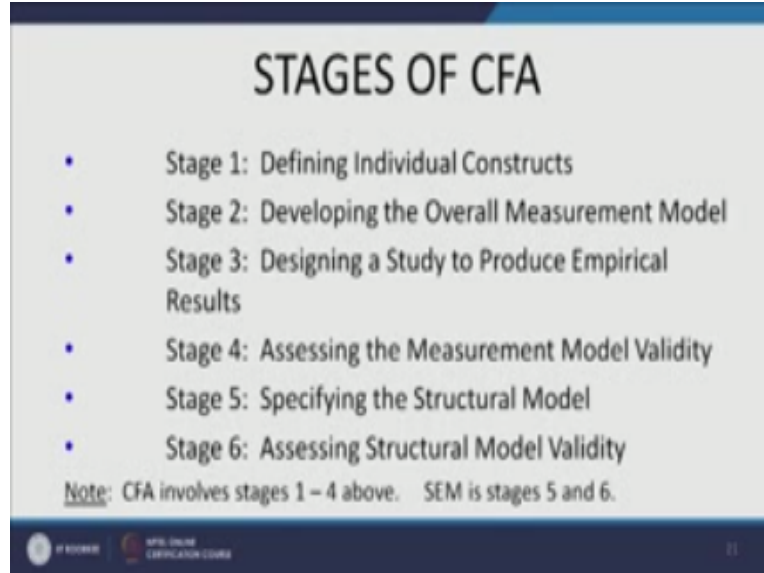
Now stages of CFA as it says it first define the individual constructs then develop the overall measurement model in just remember only difference between a measurement model and a structural model is that in the measurement model you do not have any directional relationships okay it is co variants you know model and every factor or every constructor is related to each other right.

So only let us say two factors are there so these two factors are correlated so in this we are not taking any dependent independent relationships okay, so designing a study to produce a imperial results s is the measurement model validity so if you can see as I have written here CFA involves stage one to 4 you know one to four individual construct definition measurement model then producing the empirical results.

Now how you get this empirical result I will just explain that then explain assess the measurement model validity and last two is what you do in SEM is in the stage five and six specify the structural model and assess the validity of the structural model. So if you do this much entire SEM is over right so let us see this first part right so define constructs.

(Refer Slide Time: 19:56)

So if you go back it said define the construct right so list constructs that will comprise the measurement model which constructs should be a part of them measurement model and determine if the existing scale are available are can be modified to test that means what if you have for example some scales which is already build by somebody some researcher you can use it.

It is much easier but suppose you this is sometimes this question comes to the mind of people right suppose you can I use my own yes you can if it has been you have tested it and it has tested with backing of some theory you can use it no issues right you can even modify you can modify the scale but remember there is no issue in doing all these things because ultimately you will have to check the validity of the scale.

Because if you make any changes then the validity or do not make you anyway you are checking the validity right so if you see that there is the valid instrument then all the problems are gone right third thing is if the existing scale are not available then you can develop new scales that is what I have been always emphasizing okay.

(Refer Slide Time: 21:18)

DEFINE CONSTRUCTS

- All constructs must display adequate construct validity, whether they are new scales or scales taken from previous research. Even previously established scales should be carefully checked for content validity.
- Content validity should be of primary importance and judged both qualitatively (e.g., expert's opinions) and empirically (e.g., unidimensionality and convergent validity).
- A pre-test should be used to purify measures prior to confirmatory testing.

So define it says it should display adequate construct validity which we discussed right and even previously so this is already discussed so I am not discussing further right content validity should be a primary importance see this is something which we should not aware neglect because research is a complete thing so if you by any chance feel something is less importance than the rest than we are sorry that is a wrong thing.

We should always go for face validity or content validity by checking it through experts with the opinion and all right pre test should be used to purify measures prior to confirmation testing now what is this say now before you test the relationships because if y6ou suppose you have brought something right some kind of a instrument that instrument souls be before actual production starts that instrument should be pre tested what happens if the you are once you are not tested the instrument and suddenly you try using it in the main production.

(Refer Slide Time: 22:42)

Key Issues ...

- Unidimensionality – no cross loadings
- Congeneric measurement model – no covariance between or within construct error variances
- Items per construct – identification

Suddenly the machine does not work so what will happen if the machine does not work then may be the whole process can come to a halt or it can give you results so all this things to avoid you should always pre test the instruments right so what should we see the key issues no cross loading should be there right covariance between or within construct error variances that mains that the and three things basically you should understand that there is no co variance between or within the construct error variances.

Please understand this so this are the suppose my this is the variable and this is the error one this is other variable this is error 2 similarly this is my error 3 so that should prone try to establish unnecessarily relationship between the error terms right and if required there are ways to do it which I have said already the modification which is used to do that but at the moment just understand this then is how many items are my using items per construct.

Now suppose you have two few constructs right two few items then there is a problem that can occur is that the you can say it is the under identified construct sometimes it is just the right one right and sometimes we have over identifies over identifies constructs are always better what is how do you understand this over identifier and under identifier
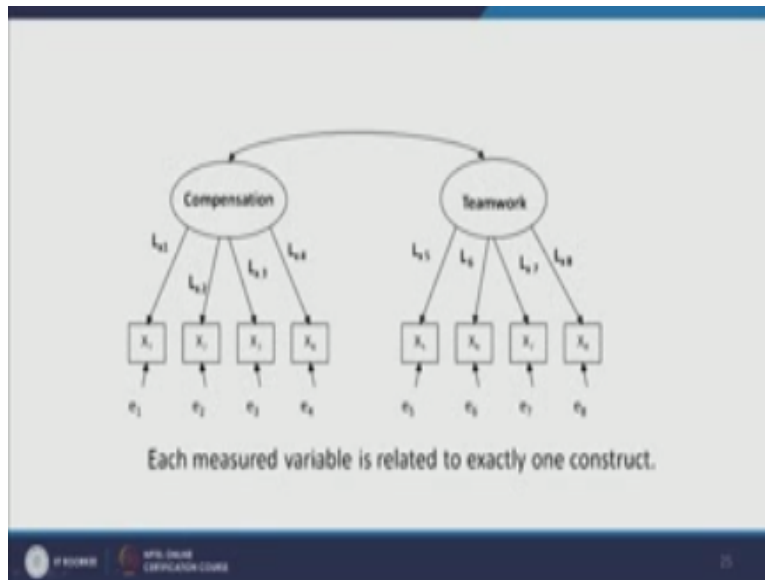
(Refer Slide Time: 24:08)
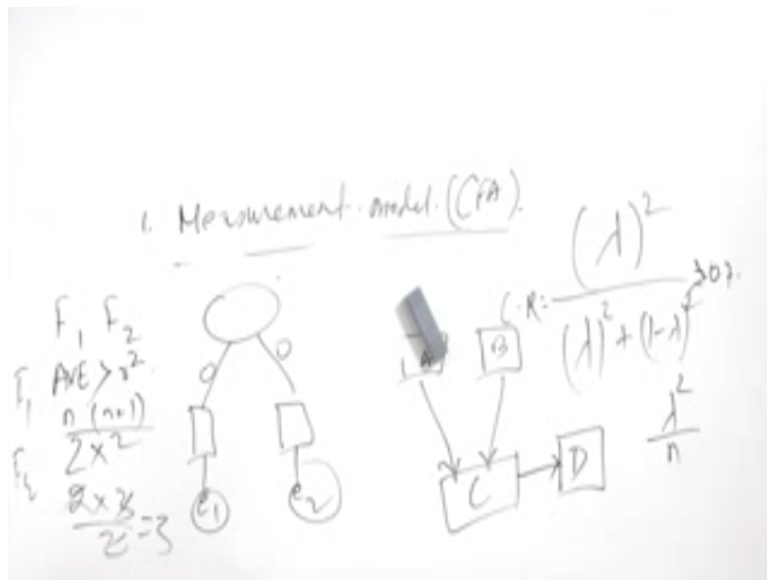
Suppose there are let say there is on variable it has got two one construct two variables let say and each construct has got a another term each variable is got term right so you are calculating two things one the error right and here the loading right so four things you are you know you are calculating but suppose how many actually variables you have two right.
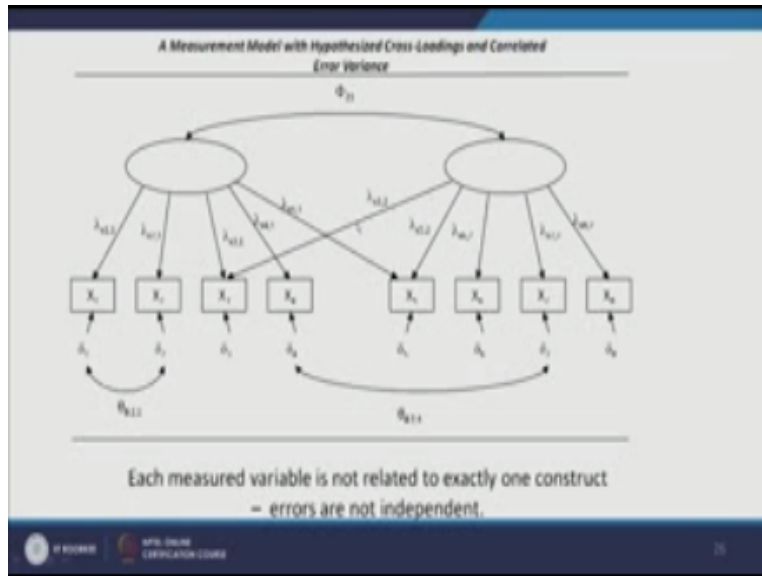
So two if I am having two variables so 2*n actually the formula goes n*n+1/2 so as part that we have 2*3/2, so 3 you got remains is 3 right but we need four things to calculate so that is why it the case which we said under identified in some cases if you have say 4 variables and you need 8 relationships, so 4 x 5/ 2, so already you have 10, so if you need to calculate 8 no issues with it right, so that is what you need to understand. So this is how it looks like.

(Refer Slide Time: 25:24)

Each measured variable is related to exactly one construct.

Each measured variable is related to exactly one construct, so no cross loadings okay.

(Refer Slide Time: 25:29)

A Measurement Model with Hypothesized Cross-Loadings and Correlated Error Variance

Each measured variable is not related to exactly one construct = errors are not independent.

So this is the case of cross loading, I think this is clear and here there is another cross loading, so this is not to be allowed and errors are not independent, which is wrong. One should not be trying to do it until there is highly serious requirements right. One should avoid the relationship between the error terms.

(Refer Slide Time: 26:03)

So developing the overall measurements model, in standard CFA application testing and the latent constructs should be indicated by at least three measurable variables, probably 4 or more is also preferable right and they should be statistically identified. Now this is something like you know two things that is important, formative factors and reflective factors. Now formative factors are slightly different to reflective factors, what do they mean?

I think it is not mentioned here, let us take a case sometimes, I think I explained in my last session also, the relationship between 2 variables could be let say this is the constructed variable, C could be effecting the or it could be that C is getting effected by V. So this is the case which we generally talk about, so this is the case of the reflective construct and this is the formative construct. In the reflective construct what we are trying is that the construct is effecting the variable right. In other case we are saying that the variable is affecting the construct, now it is written over here.

(Refer Slide Time: 27:27)

Reflective Vs Formative Constructs

- Formative factors are not latent and are not validated as are conventional reflective factors. Internal consistency and reliability are not important. The variables that make up a formative factor should explain the largest portion of variation in the formative construct itself and should relate highly to other constructs that are conceptually related (minimum correlation of .5):
  o Formative factors present greater difficulties with statistical identification.
  o Additional variables or constructs must be included along with a formative construct in order to achieve an over-identified model.
  o A formative factor should be represented by the entire population of items that form it. Therefore, items should not be dropped because of a low loading.
  o With reflective models, any item that is not expected to correlate highly with the other indicators of a factor should be deleted.
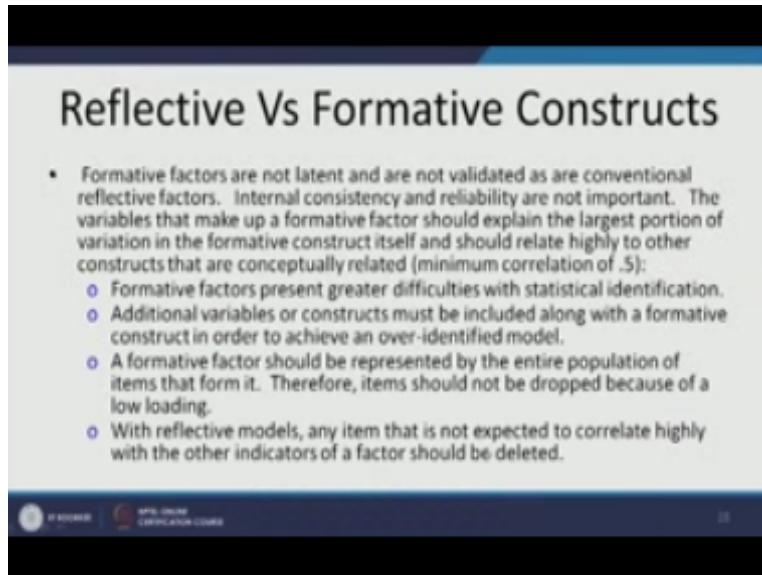
You can see formative construct are not latent and they are not validated right, suppose in this case, this is the formative that means the construct is affecting the variables which is actually a classical suggestion but in this case the variables is effecting. Now I suppose as I said this is the social class, social class is getting explained let say the religion, something else, so this is the case, these are two different cases right.

We will not get to much into this because still what is it saying is formative constructs latent and not validated as conventional reflective factors okay. Internal consistency and reliability are not important in the case of the formative construct. The variables that make up a formative factor should explain the largest portion of variance in the formative construct itself and should relate highly into other constructs are conceptually related.

So what he is saying this is formative construct although there has to be certain correlation among the constructs among each other or themselves, and the minimum correlation value should be 0.5. Formative factors present the greater difficulties with the statistical identification, obviously most of the time what we use is reflective constructs most of the studies and we use less the formative construct okay.

Items this is the beautiful part that in the formative construct is that in the formative construct we do not drop an item, just because of it is poor loading but in case of the reflective construct what we do is when an item or a variable, does not justify a sufficient loading on to the factor, then this item is an very good case for delusion once you delete it right with reflective models.

Any item that is not accepted to correlate highly with other indicators of the factor should be deleted what you are saying any item that is not expected to core relate authentically should be automatically deleted okay.
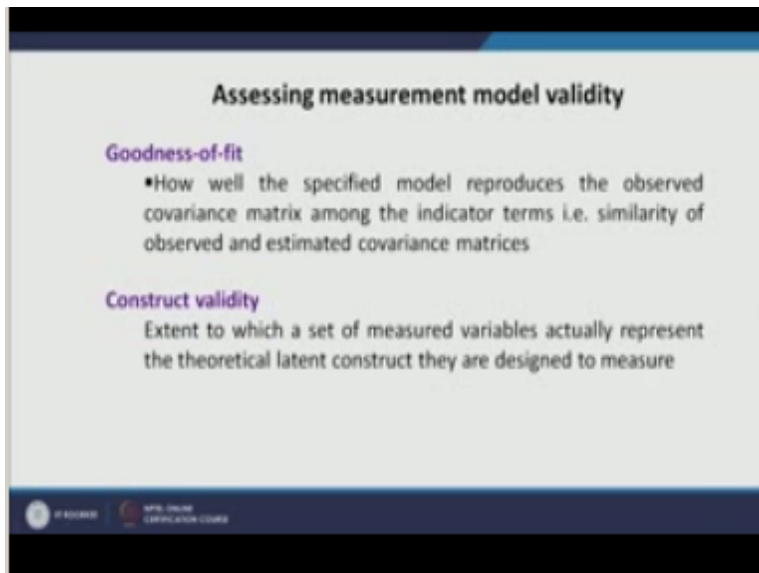
Now how do you check the models fitness now you have done through SEM now after doing CFA you need to check whether the model is valid or not valid right to check the validity of a model I said there are certain you know ways of doing it this is some test which are highly used the goodness of fit test right the goodness of fit test is a test which says that a specified model reproduces the observed covariance matrix if you remember I have said that there is observed matrix and there is an expected matrix.

And the difference between the observed and the expected matrix should be very low now how it is possible when the value should be less that means when the value suppose if you have two models rights and if you checking both the models the difference in the models if it is low then it is good situation right so there is a it means that it is that what x^2 does is goodness of fit test basically right so this is what it says the second thing is the construct validity so they extend to which a set of measure variables actually represent the theoretical latent construct they are designed to measure.

So we have discussed about it also right so this composite reliability how to calculate and all these are some of the before that let me go to the values right so the different values if you do not understand if you just want to remember it remember this much there are certain values call the absolute fit right during any confirmatory factor analysis or then later on the structural equation further you know the structural model also for that the measurement of structural both for the both things the same fitness induces are used and in this case.

When the fitness induces are that you are using they need to be sufficiently large above point 0.9 or more right but there is a small but here in case you have an extremely large sample size then it is researcher should know that even a fit induces of little less than 0.9 around 0.88 or yo8u know just 9 is not bad enough right it is still adequate and good right and what is this root mean square error.

(Refer Slide Time: 32:51)
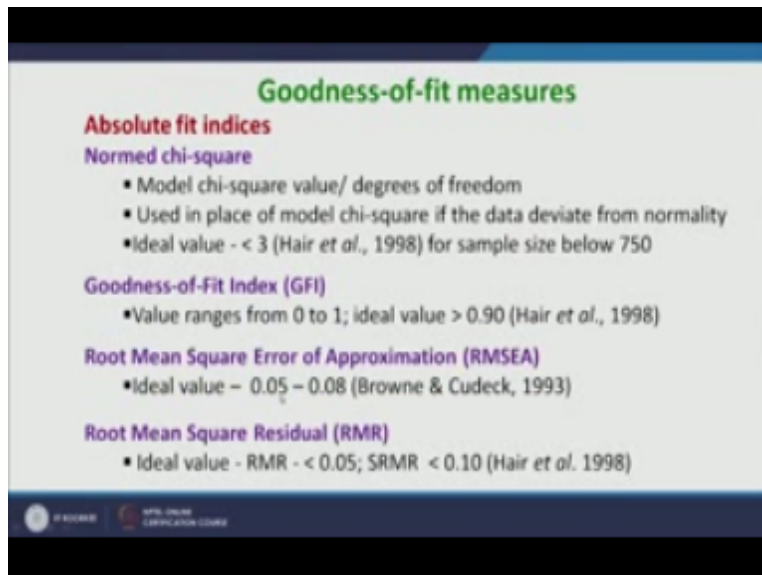
 This is the value which is been idle value has to be given to be 0.08 anything below 0.8 is good enough right so this is some of the index the values which you will find while doing a test and if this values are if you are getting within the limits then you are model is fit right so the best way to find you are model fit is just find out the calculate the χ^2/dt this value if it anything less than 3right then your model is fit although researcher say anything till 5 also I s good even sometimes they set n but I would always advice you to go for χ^2/dt value of maximum of up to 3 right anything below 3 that means your model is fitting right.

So this is how you test the fitness of the model and there are some of the values everything all this values have to be around 0.9 right and then I explain the descriptive validity I explain through the average variable extension and the correlation the √ave >square correlation when you do this you find the distinct validity.

So this is all you do to check validate and explain the constructs right and then once you have done your constructs are valid in nature then the next step comes is to test the relationships

through a structural model we will take it up in the next session that is what this is what we have discussed today and thank you very much for this session.

For Further Details Contact

Coordinator, Educational Technology Cell
Indian Institute of Technology Roorkee
Roorkee 247 667
E-Mail Etcellitrke@gmail.com etcell@itr.ernet.in
Website: www.itr.ac.in/centers/ETC. www.nptel.ac.in

Production Team	**Camera**
Sarath Koovery	**Arun. S**
Mohan Raj. S	
Jithin. K	**Online Editing**
Pankaj saini	**Arun.S**
Graphics	**Video Editing**
Binoy. V.P	**Arun.S**

NPTEL Cooridinator
Prof B.K Gandhi

An Educational Technology Cell
IIT Roorkee Production