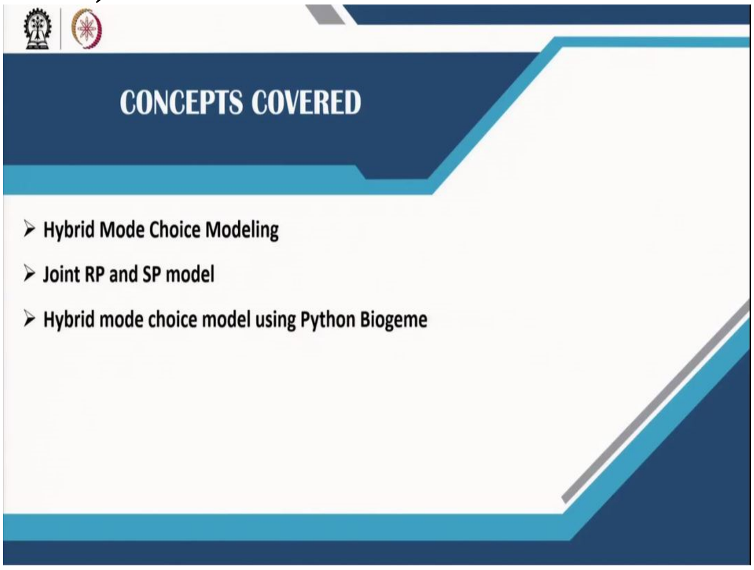
Urban Landuse and Transportation Planning
Prof. Debapratim Pandit
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Lecture - 39
Hybrid Mode Choice Model 2 (Joint RP SP Model)

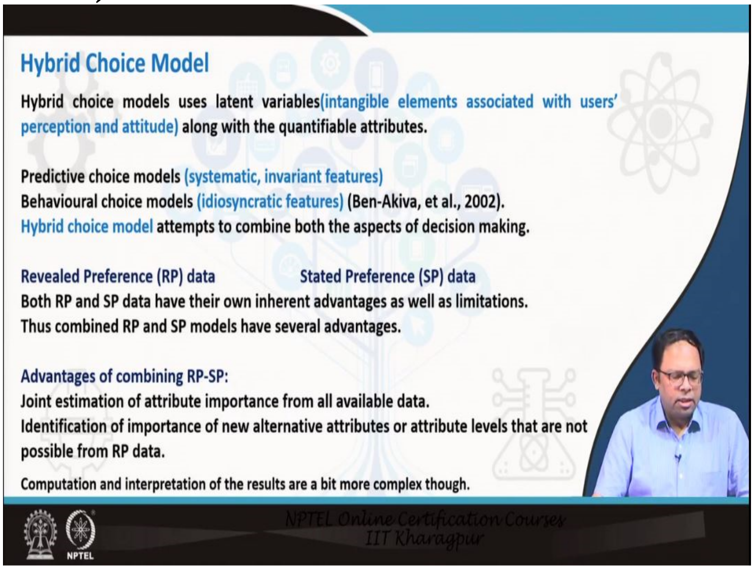
In this lecture, hybrid mode choice modelling and joint RP SP model is introduced.

(Refer Slide Time: 00:38)



The different concepts that has been covered are; hybrid mode choice modeling, joint RP and SP model; and hybrid mode choice model using Python Biogeme.

(Refer Slide Time: 00:51)



Hybrid Choice Model:

In hybrid mode choice model, many variables which are not quantifiable in the normal sense, are considered. These variables, also called latent variables, are determined through factor analysis (both exploratory and confirmatory factor analysis). Factor analysis can be used to reduce the number of independent variables associated with intangible aspects like users' perception, attitudes, etc., in the model. This is done by identifying the latent constructs that govern these intangible aspects. The factors are designed as weighted sum of the indicators (or the variables) they are associated with and hence the representation of the indicators are preserved in the factors. These factors can be directly used in a mode choice model which enhances the predicting power of the model.

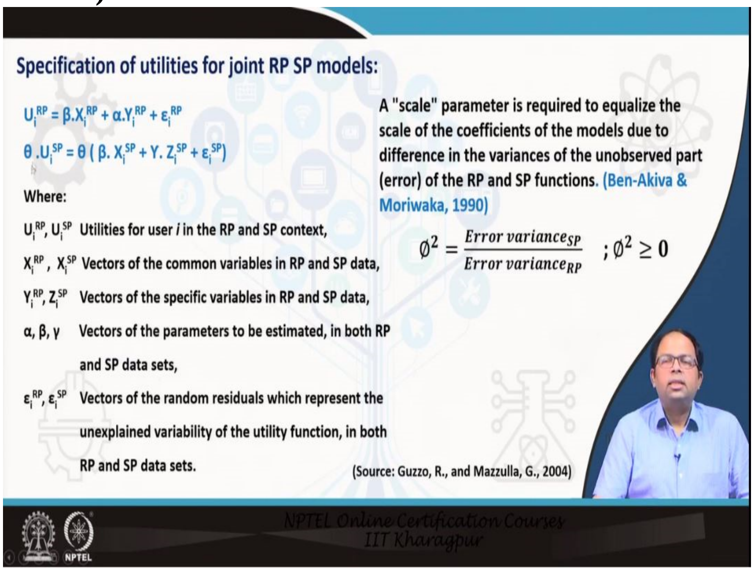
A mode choice model meant for prediction, called **predictive choice models**, are made using systematic variables like cost, time, etc. Another kind of models known as **behavioral choice models** are made using many idiosyncratic features like perception, attitudes, etc. **Hybrid choice model** combines both the kinds of decision making into one framework.

As discussed in previous lectures, models can be developed based on either SP data or RP data. But each has got their own strengths and weaknesses. It is always preferred to have a model which is based on a combination of both these kinds of data sets i.e., the parameter estimation is done in such a way that both the datasets are used. These models are known to be more robust as it gains the strength of both of these particular data sets, and cancels out the weaknesses of each other.

For example, during stated preference surveys, new modes, or new levels in a particular attribute can be added as an option which increases the variability of the data. In the case of a revealed preference survey, data is based on existing choices and those data sets allow including lots of parameters as well which are not possible in a stated preference survey. One limitation of stated preference data is that, there is a limitation related to parameters inclusion, since large number of parameters cannot be included in choice cards. Having many parameters in a scenario may make it difficult for the people to evaluate the scenario. But during the revealed preference survey, a surveyor can ask a person about their existing preferences or alternatively secondary source data about existing alternatives can be gathered, resulting in a comparatively large data set. Revealed preference data sets can be large but variability is less, whereas, in stated preference data sets, the attributes may be less in number, but the variability in the attributes is much higher.

If both these datasets are be combined, some parameters would be common to both RP and SP datasets, whereas, some variables or attributes are exclusive to either RP or SP data sets. There are several advantages when we combine these two datasets. First is the joint estimation of attribute importance. Since, this is a joint estimate, it is different from estimation of a multinomial logit model. The estimation is done following a similar structure as a nested logit model.

(Refer Slide Time: 05:36)



Specification of utilities for joint RP SP models:

The utilities for a user in real (RP) and hypothetical (SP) context can be given by the following equation respectively:

$$U_i^{RP} = \beta * X_i^{RP} + \alpha * Y_i^{RP} + \varepsilon_i^{RP}$$

$$\theta * U_i^{SP} = \theta(\beta * X_i^{SP} + \gamma * Z_i^{SP} + \varepsilon_i^{SP})$$

Both the utility equation has two components, which are the systematic portion of utility and error term. As discussed in earlier lectures, the systematic portion of utility corresponds to the information that is known to the analyst, whereas the error term represents randomness in the utility.

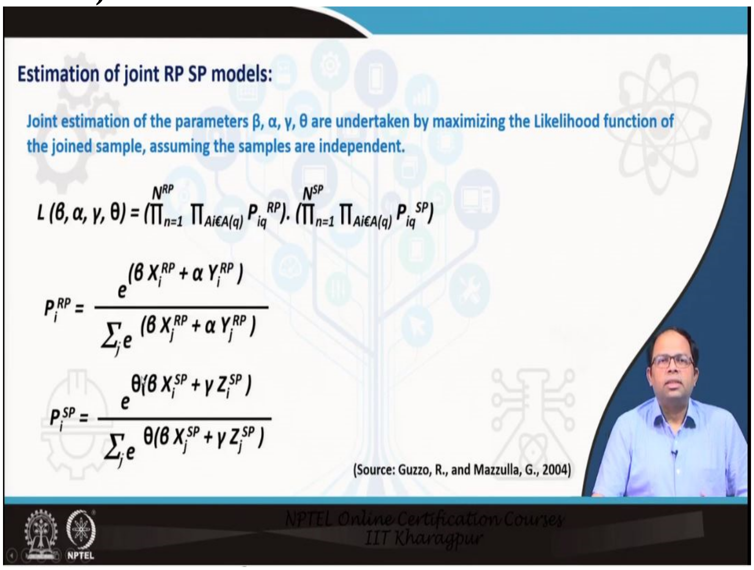
In utility equations, X_i^{RP} and X_i^{SP} are the vectors of the common variables in RP and SP data, therefore β coefficient vector is common in both the equations. On the other hand, Y_i^{RP} and Z_i^{SP} are the vectors of variables that are specific to RP and SP data respectively. So, α and γ are different vectors of parameters.

ε_i^{RP} and ε_i^{SP} are the error terms, which explains the unexplained variability of the utility in RP and SP data. Since RP and SP data are two different datasets, so there is a difference in variance of the unobserved part of the utility function of RP and SP data. Due to this difference, a scale parameter θ is introduced which equalizes the scale of the coefficients of the model. In most cases, the scale parameter θ is expressed as a function of ε_i^{RP} and ε_i^{SP} as follows:

$$\theta^2 = \frac{error\ variance_{SP}}{error\ variance_{RP}} \qquad where\ \theta^2 \geq 0$$

In the utility equation of SP data, parameter 'θ' scales all the variables and coefficients to account for the unequal variance. So, that is how both the utility equations are brought to the same platform. The next step is to estimate the parameters of both RP and SP utility function i.e. α, β, γ, and θ.

(Refer Slide Time: 09:02)



Estimation of joint RP SP model:

The joint estimation of the parameters β, α, γ, θ are undertaken by maximizing the likelihood function of the joined sample, assuming the samples are independent. The likelihood function is:

$$L(\beta,\alpha,\gamma,\theta) = \left(\prod_{n=1}^{N^{RP}} \prod_{A_i \epsilon A\,(q)} P_{iq}^{RP}\right) * \left(\prod_{n=1}^{N^{SP}} \prod_{A_i \epsilon A(q)} P_{iq}^{SP}\right)$$

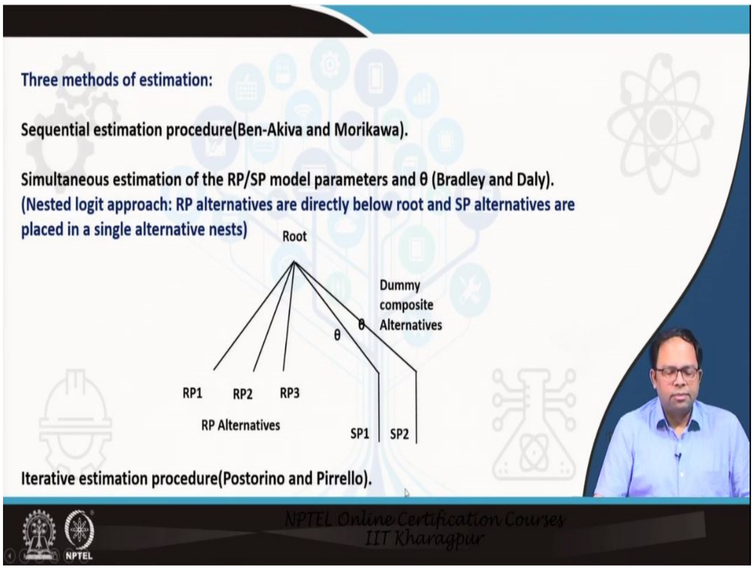
Where, P_{iq}^{RP} and P_{iq}^{SP} are probabilities for each choice and user, which are given by the following expression:

$$P_{iq}^{RP} = \frac{e^{\left(\beta X_i^{RP} + \alpha Y_i^{RP}\right)}}{\sum_j e^{\left(\beta X_j^{RP} + \alpha Y_j^{RP}\right)}}$$

$$P_{iq}^{SP} = \frac{e^{\theta\left(\beta X_i^{SP} + \alpha Y_i^{SP}\right)}}{\sum_j e^{\theta\left(\beta X_j^{SP} + \alpha Y_j^{SP}\right)}}$$

The parameters β, α, γ, θ are estimated by taking the first derivative of the likelihood function and equating it to zero. This task is generally done using software. So, once these parameters are estimated, the probabilities of choice in revealed preference, and stated preference context can be calculated. The probability equations for RP and SP are similar to the standard multinomial logit model format. The only difference is in SP data formulation, where the utility equation is multiplied by a scale parameter θ.

(Refer Slide Time: 10:53)



There are three methods of estimation that are sequential, simultaneous, and iterative estimation procedure. The first method of estimation is the sequential estimation procedure. This method was proposed by Ben-Akiva and Morikawa. At first, the model is developed and the parameters are estimated for the first level. Then the estimated parameters are introduced in the model estimation of the next step. The disadvantage of the sequential method is that it does not consider all the information jointly.

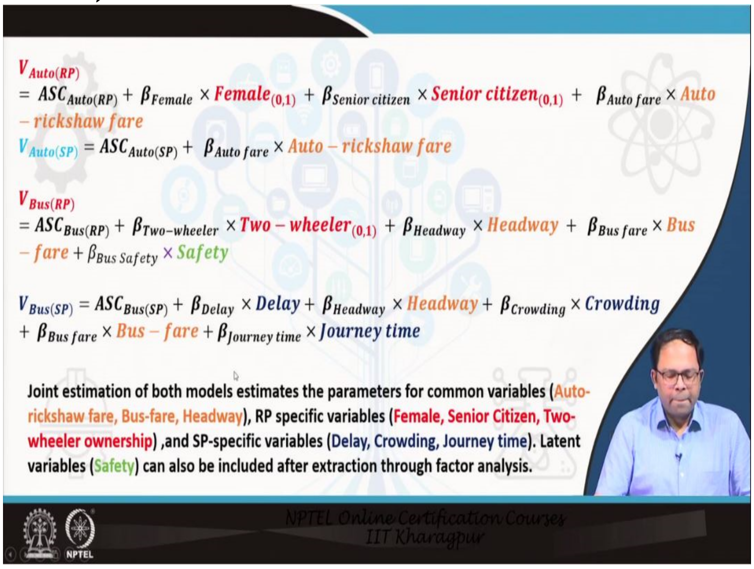
The second method of estimation is the simultaneous estimation procedure. It is similar to the simultaneous estimation procedure of nested logit model, which will be discussed in lecture 40. Similar to joint decision models, such as the joint decision of work location choice and residential location choice, or joint decision of intent to move and location choice, where both decisions are considered in a nested structure, the RP and SP alternatives are also considered in nested form. It means that a nest is created instead of taking all the alternatives at the same level.

In the given nest (adopted from Bradley and Dally), RP alternatives are kept just below the root, whereas SP alternatives are kept under single-alternative nest. There could be two or three categories within each nest, but in the present example, one SP alternative is considered under each nest. Also, at the same level of RP alternatives (below the root), a dummy composite alternatives is introduced. So, this structure is created to make sure that the scale parameter θ (or dummy composite alternative) can be estimated.

This procedure of estimation is similar to the nested logit model, where a similar kind of scale parameter is introduced. But instead of a dummy composite alternative, there is a composite alternative. For example, at first level, the mode choices are car, walk, two-wheeler, and transit. Within transit, there are bus and rail. Therefore, it becomes a two-alternative nest. So, this concept has been used to estimate the parameters simultaneously. This procedure is used in most of the mode choice models or most of the software.

Finally, the third method of estimation is called the iterative estimation procedure. This was proposed by Postorino and Pirrello. This is another method that has been used by other researchers.

(Refer Slide Time: 14:05)



In order to understand the formulation of utility equations, let us consider an example where the RP-SP model is to be developed for choice between auto-rickshaw and bus in a particular corridor. Suppose the auto-rickshaw choice in the RP context is influenced by factors such as gender (female), household category (senior citizen), and auto-rickshaw fare, whereas in the SP context it is influenced by auto-rickshaw fare only. Then, the utility equation for auto rickshaw in real (RP) and hypothetical (SP) context can be written as:

$$V_{Auto}^{RP} = ASC_{Auto}^{RP} + \beta_{Female} * Female_{(0,1)} + \beta_{Senior\ citizen} * senior\ citizen_{(0,1)} + \beta_{Auto\ fare} * Auto\ rickshaw\ fare$$

$$V_{Auto}^{SP} = ASC_{Auto}^{SP} + \beta_{Auto\ fare} * Auto\ rickshaw\ fare$$

Where V_{Auto}^{RP} and V_{Auto}^{SP} are utility of auto-rickshaw in RP and SP context, ASC_{Auto}^{RP} and ASC_{Auto}^{SP} are bias, β_{Female} and $\beta_{Senior\ citizen}$ are RP specific parameter coefficients, $\beta_{Auto\ fare}$ is a common parameter coefficient, and $Female_{(0,1)}$ and $senior\ citizen_{(0,1)}$ are dummy coded variables.

Similarly, consider the choice of bus in RP context is influenced by two-wheeler ownership, headway, safety, and bus fare. In the SP context, it is influenced by delay, headway, crowding level, bus fare, and journey time. So, the utility of bus for a user in RP and SP context can be given by:

$$V_{Bus}^{RP} = ASC_{Bus}^{RP} + \beta_{Two-wheeler} * Two\ wheeler_{(0,1)} + \beta_{Headway} * Headway + \beta_{Bus\ fare} * Bus\ fare + \beta_{Bus\ Safety} * Safety$$

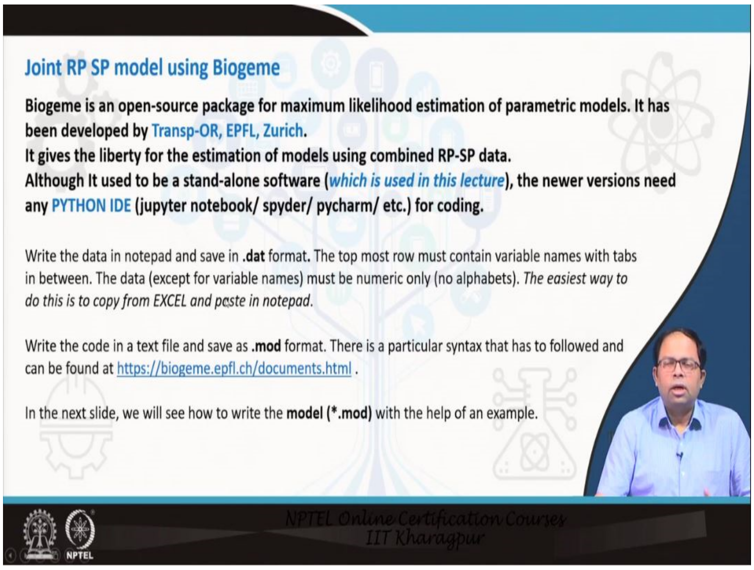
$$V_{Bus}^{SP} = ASC_{Bus}^{SP} + \beta_{Delay} * Delay + \beta_{Headway} * Headway + \beta_{Crowding} * Crowding + \beta_{Bus\ fare} * bus\ fare + \beta_{Journey\ time} * Journey\ time$$

where, V_{Bus}^{RP} and V_{Bus}^{SP} are utility of bus in RP and SP context, ASC_{Bus}^{RP} and ASC_{Bus}^{SP} are bias, $\beta_{Two-wheeler}$ and $\beta_{Bus\ Safety}$ are RP specific parameter coefficients, β_{Delay}, $\beta_{Crowding}$, and $\beta_{Journey\ time}$ are SP specific parameter coefficients, and $\beta_{Headway}$ and $\beta_{Bus\ fare}$ are common parameter coefficients in RP and SP data. In addition to these, variables such as two-wheeler ownership (0, 1) are also included.

In normal estimation of parameters, the models are developed for RP and SP data separately i.e., utility of bus and auto-rickshaw for RP context is considered together, and the utility of bus and auto-rickshaw for SP context is modelled together. The joint estimation of the RP and SP model estimates the parameters for common variables (auto-rickshaw fare, bus fare, and headway), RP specific variables (female, senior citizens, two-wheeler ownership, safety), and SP specific variables (delay, crowding, journey time) jointly.

The variables 'safety' is a latent variable that is included in the model after extraction through factor analysis. Finally, the scale parameter θ is multiplied in SP utility equations, which will give the true estimates of the parameters.

(Refer Slide Time: 18:33)

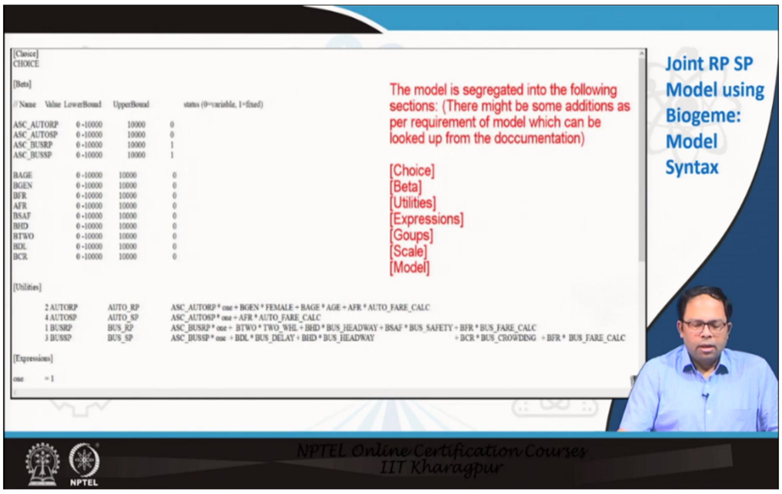


Joint RP-SP model using Biogeme:

There are many softwares that can be used to develop a joint RP-SP model. LIMDEP is a popular software used by researchers for this purpose. As LIMDEP is not freely available, and is costly, there are other alternatives for doing so. Python Biogeme is a one of the free option available to develop a joint RP-SP model. This is an open source package for maximum likelihood estimation of parametric models developed by Trans-OR, in EPFL Zurich. There are different versions of it, the newer versions are fully Python based, and available in Python IDE. The previous version was written in C++, and then in later versions, the internal code in Python, but the interface remained same as the old version. The version shown in the demonstration follows this version. All the versions can be downloaded from the Biogeme website. Rich documentation can also be found in the website on this software, which has many worked out examples of different kinds of models. The data file to be used in this software needs to be in a ***.dat** format, which can be written in a text file and saved with a ***.dat** extension. The model script needs to be in *.mod format, which can also be written in a text file, and saved with a ***.mod** extension.

(Refer Slide Time: 20:05)

(Video Starts: 20:05)



Biogeme syntax for joint RP-SP model:

In the demonstration, an example of a model script is shown for a joint RP-SP model. Scripts are basically some instructions that we have to type, which is done in other packages also like LIMDEP, R, etc. For any package, documentations are available that has instructions on ways and steps to specify a model using that particular package. Biogeme also has a documentation that has examples showing the syntax of writing a script for different types of models. The script for the demonstration model is given as follows:

<table>
<tr><th colspan="2">Biogeme script for a model file.</th></tr>
<tr><td>[Choice]
CHOICE</td><td>Specify the choice variable.</td></tr>
<tr><td>[Beta]

// Name Value LowerBound UpperBound status (0=variable, 1=fixed)

ASC_AUTORP 0 -10000 10000 0
ASC_AUTOSP 0 -10000 10000 0
ASC_BUSRP 0 -10000 10000 1
ASC_BUSSP 0 -10000 10000 1

BAGE 0 -10000 10000 0
BGEN 0 -10000 10000 0
BFR 0 -10000 10000 0
AFR 0 -10000 10000 0
BSAF 0 -10000 10000 0
BHD 0 -10000 10000 0
BTWO 0 -10000 10000 0
BDL 0 -10000 10000 0
BCR 0 -10000 10000 0</td><td>Define the coefficients for each of the variables.</td></tr>
<tr><td>[Utilities]

2 AUTORP AUTO_RP ASC_AUTORP * one + BGEN * FEMALE + BAGE * AGE + …….
4 AUTOSP AUTO_SP ASC_AUTOSP * one + AFR * AUTO_FARE_CALC
1 BUSRP BUS_RP ASC_BUSRP * one + BTWO * TWO_WHL + ………
3 BUSSP BUS_SP ASC_BUSSP * one + BDL * BUS_DELAY + ………..
[Expressions]</td><td>Specify the utility equations using the coefficients and the variable names mentioned in the data file.</td></tr>
</table>

one = 1 BUS_RP = SP == 0 BUS_SP = SP == 1 AUTO_RP = SP == 0 AUTO_SP = SP == 1 GROUP = SP	Specify: • Composite variable • Maker for SP and RP observations.
[Group] GROUP	
[Scale] 0 1 0.001 10000 1 1 1 0.001 10000 0	Variable to be used for grouping. Initializing scale parameter
[Model] $MNL	The model framework to be used.

The script in Biogeme has various compulsory and optional sections. The details can be seen in the documentation which has various examples exhibiting the use of various sections. In the demonstration model, an MNL model is estimated using a combined RP and SP dataset. The sections in the model are; **Choice**, **Beta**, **Utilities**, **Expressions**, **Group**, **Scale**, and **Model**. These are denoted by the respective names, written in a sentence case, within **square parenthesis** and they follow a particular sequence in the script.

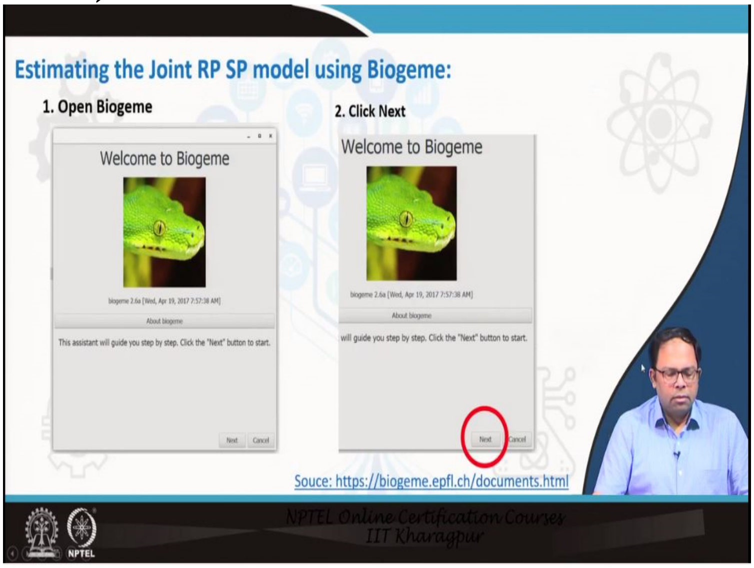
The name of the variable that holds the choice of alternatives is to be mentioned under **[Choice]**. The names should be mentioned exactly as written in the data file. In the **[Beta]** section, the coefficients for the variables to be used, are defined. The coefficients must match with the variables exactly in terms of their numbers i.e. there should not be any coefficient defined and not used in the utility equations. Corresponding to each coefficient; the lower bound; the upper bound; and status needs to be specified, which is kept as 0; 1000; and either ‘1’ if the coefficient is to be kept fixed (reference) or ‘0’, if it needs to be estimated. All the entities are separated by with a **SPACE** or a **TAB**. In the **[Utilities]** section, the code for a particular alternative in the choice column in the dataset is specified; followed by the name of the alternative (BUSRP, AUTORP, BUSSP, AUTOSP); followed by the availability variable (BUS_RP, AUTO_RP, BUS_SP, AUTO_SP); followed by the utility equation. A utility equation is estimated with the observations for which the availability variable returns ‘**1**’. In the equation, each term and operator must have a **SPACE** between each of them. In the **[Expressions]** section, any new variable used in utilities part, not present in the database, is specified. For example, the availability variables, **BUS_RP = SP==0** states that BUS_RP can be 1 only when SP variable is 0. This implies that BUS_RP is estimated only with RP data. Also GROUP is a variable, that has been used later in the script, that specifies the variable in the dataset based on which scale factor can be estimated for different group(s). In this case, as scale factor for SP needs to be estimated, GROUP is made based on SP. **[Group]** section required the expression to identify grouping variable to be specified. **[Scale]** section defines the scale parameters and initializes them. Similar to '**Beta**' section, the first column represents the name of the parameter, followed by the default value of starting the estimation (1 is used), followed by the lower bound and upper bound, and the status at last. Since in the demonstration, the SP scale parameter needs to be estimated, the status of RP scale parameter is kept as 1. In **[Model]** section, the model to be used is specified. In the demonstration an MNL framework is used, so **$MNL** is written. For cross nested logit model, $CNL is used. Details on the syntax of other components can be found at the given webpage.

https://transp-or.epfl.ch/pythonbiogeme/documentation/bisonsyntax/bisonsyntax.html

Similarly, a nested logit model can also be estimated. In the nested logit model, some additional steps are added where the nests are specified. The variables and their corresponding nests needs to be specified.

(Video Ends: 23:42)

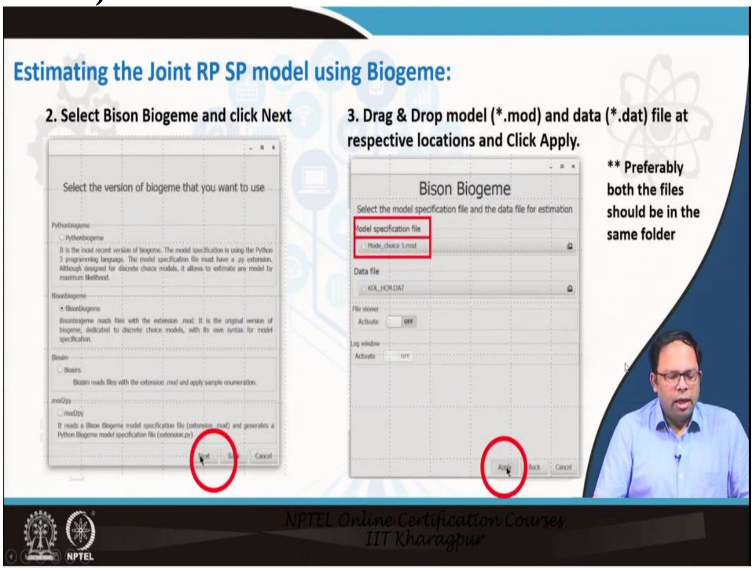
(Refer Slide Time: 23:45)



Estimating a Joint RP-SP model using Biogeme:

Once the code is written, and saved in a text file with ***.mod** extension, and the data file saved in another text file with ***.dat** extension, Biogeme can be opened. The distribution of Biogeme used in this demonstration is a C++ version of the software. The 'pandas' version, which is the latest one, is also available freely.

(Refer Slide Time: 23:57)



The steps to be followed in order to develop a combined RP-SP model in Biogeme are given below:

$\rightarrow$ *Click* ***Next*** *in the welcome page*

$\rightarrow$ *Select* ***Bison Biogeme*** *from the list of the versions*

$\rightarrow$ *Drag and drop the model file* $(*.mod)$ *in the* ***Model specification file*** *tab*

$\rightarrow$ *Drag and drop the data file* $(*.dat)$ *in the* ***Data file*** *tab*

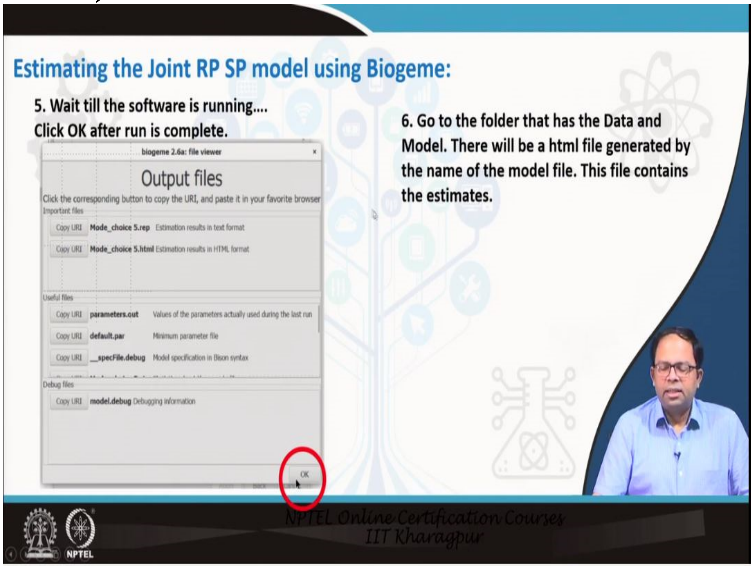
(*Preferably both the files should be in the same folder*)

$\rightarrow$ *Click* ***Apply*** *to start the estimation*

$\rightarrow$ *The result of the estimation is saved in html format in the same location as the data*

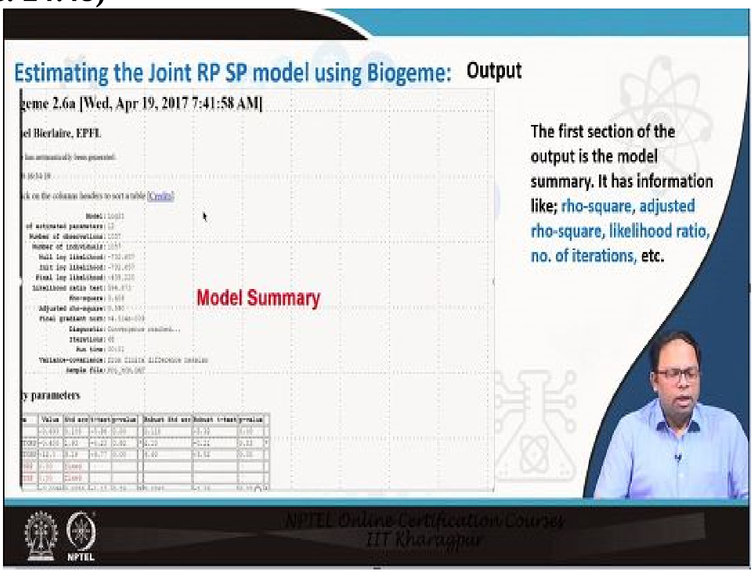
As the welcome screen of Biogeme pops up, '**Next**' has to be clicked to proceed. The next window is a list of versions of Biogeme that can be used for estimation. In the demonstration example, **Bison Biogeme** was selected. Other version like the Python Biogeme can also be used which has a different syntax than Bison Biogeme. The latest version of Biogeme can be used in a Python IDE, and has been enhanced using '**Pandas**' package of Python. In the '**Model specification file**' tab, the model file, which needs to be in '***.mod**' format, needs to be located (or simply dragged and dropped). Similarly, the data file, which needs to be in '***.dat**' format, needs to be dragged and dropped at the '**Data file**' tab. '**Apply**' needs to be clicked for estimation.

(Refer Slide Time: 24:28)



After the successful estimation of the model, the output is saved in a HTML file in the same location as the data file. The HTML file needs to be opened to understand the model estimates generated by the software.

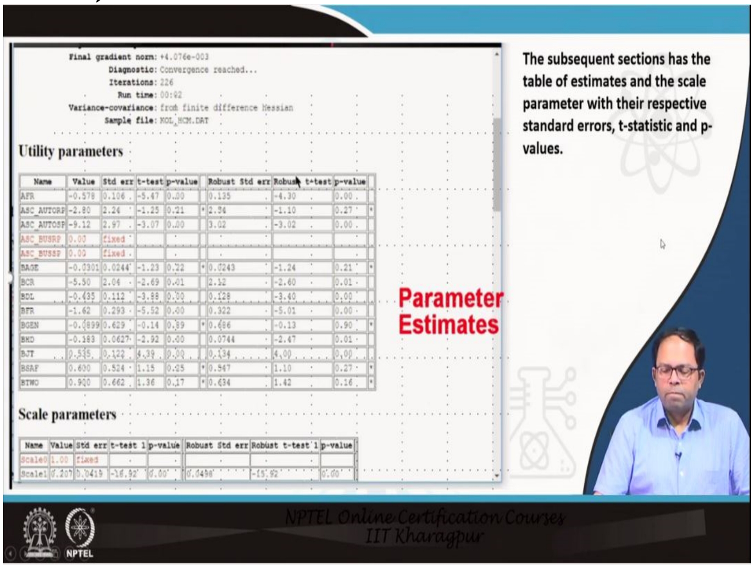
(Refer Slide Time: 24:48)



The image shows the top part of the HTML output generated by Biogeme. The first section is the model summary. It includes information about rho-square; adjusted rho-square; likelihood ratio; number of iterations; etc. For the demonstrated model, rho-square is 0.406; adjusted rho-square is

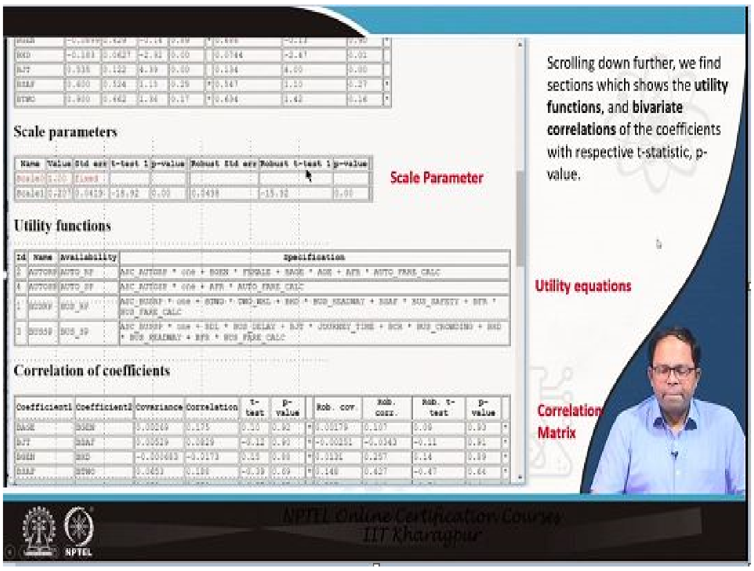
0.390; likelihood ratio is around 594, obtained as the double of the difference between final and initial log-likelihood; and number of iterations required to reach convergence is 66.

(Refer Slide Time: 25:26)



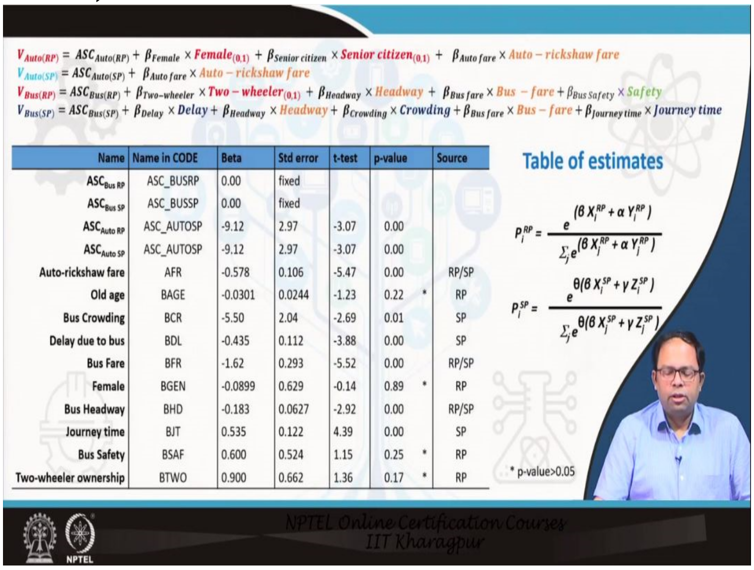
Scrolling down a bit, the estimated parameters are shown. Corresponding to each of the coefficients defined in biogeme, for each variable in the utility equation, their regression coefficients, standard error, t-statistic, and p-value can be seen. In the table of estimates, ‘**ASC_BUSRP**’ and ‘**ASC_BUSSP**’ are zero, which means BUS has been taken as the reference both in case of RP data and SP data. This is because, at the end of the day, this is essentially a binomial logit model, where probability of one mode is calculated with reference to another. So, as the alternative specific coefficient (ASC) of bus is turned to 0, the estimate for ASC_AUTO is estimated for both RP and SP data. Also, a scale parameter is estimated , which has been found to be 0.207. So, when the probability estimation is done, the scale parameter is to be added as discussed earlier.

(Refer Slide Time: 26:39)



Scrolling down further, the utility equations can be seen, as specified in the model specification file. A table of correlation of coefficients is also estimated, which shows the correlation coefficient for pair-wise correlation among variables, and the significance level of the correlation. This is used to verify if there is a significantly large correlation among variables in the same utility equation. Otherwise, such variables are removed.

(Refer Slide Time: 27:06)



Name	Name in CODE	Beta	Std error	t-test	p-value		Source
$ASC_{Bus\ RP}$	ASC_BUSRP	0.00	fixed				
$ASC_{Bus\ SP}$	ASC_BUSSP	0.00	fixed				
$ASC_{Auto\ RP}$	ASC_AUTOSP	-9.12	2.97	-3.07	0.00		
$ASC_{Auto\ SP}$	ASC_AUTOSP	-9.12	2.97	-3.07	0.00		
Auto-rickshaw fare	AFR	-0.578	0.106	-5.47	0.00		RP/SP
Old age	BAGE	-0.0301	0.0244	-1.23	0.22	*	RP
Bus Crowding	BCR	-5.50	2.04	-2.69	0.01		SP
Delay due to bus	BDL	-0.435	0.112	-3.88	0.00		SP
Bus Fare	BFR	-1.62	0.293	-5.52	0.00		RP/SP
Female	BGEN	-0.0899	0.629	-0.14	0.89	*	RP
Bus Headway	BHD	-0.183	0.0627	-2.92	0.00		RP/SP
Journey time	BJT	0.535	0.122	4.39	0.00		SP
Bus Safety	BSAF	0.600	0.524	1.15	0.25	*	RP
Two-wheeler ownership	BTWO	0.900	0.662	1.36	0.17	*	RP

$$\boldsymbol{V_{Auto(RP)} = ASC_{Auto(RP)} + \beta_{Female} \times Female_{(0,1)} + \beta_{Senior\ citizen} \times Senior\ citizen_{(0,1)} + \beta_{Auto\ fare} \times Auto - rickshaw\ fare}$$

$$\boldsymbol{V_{Auto(SP)} = ASC_{Auto(SP)} + \beta_{Auto\ fare} \times Auto - rickshaw\ fare}$$

$$\boldsymbol{V_{Bus(RP)} = ASC_{Bus(RP)} + \beta_{Two-wheeler} \times Two - wheeler_{(0,1)} + \beta_{Headway} \times Headway + \beta_{Bus\ fare} \times Bus - fare} + \beta_{Bus\ Safety} \times \boldsymbol{Safety}$$

$$\boldsymbol{V_{Bus(SP)} = ASC_{Bus(SP)} + \beta_{Delay} \times Delay + \beta_{Headway} \times Headway + \beta_{Crowding} \times Crowding + \beta_{Bus\ fare} \times Bus - fare + \beta_{Journey\ time} \times Journey\ time}$$

Name	**Name in CODE**	**Beta**	**Std error**	**t-test**	**p-value**	**Source**
$ASC_{Bus\ RP}$	ASC_BUSRP	0.00	fixed			
$ASC_{Bus\ SP}$	ASC_BUSSP	0.00	fixed			
$ASC_{Auto\ RP}$	ASC_AUTOSP	-2.80	2.97	-3.07	0.00	
$ASC_{Auto\ SP}$	ASC_AUTOSP	-9.12	2.97	-3.07	0.00	
Auto-rickshaw fare	AFR	-0.578	0.106	-5.47	0.00	RP/SP
Old age	BAGE	-0.0301	0.0244	-1.23	0.22	RP
Bus Crowding	BCR	-5.50	2.04	-2.69	0.01	SP
Delay due to bus	BDL	-0.435	0.112	-3.88	0.00	SP
Bus Fare	BFR	-1.62	0.293	-5.52	0.00	RP/SP
Female	BGEN	-0.0899	0.629	-0.14	0.89	RP
Bus Headway	BHD	-0.183	0.0627	-2.92	0.00	RP/SP
Journey time	BJT	0.535	0.122	4.39	0.00	SP
Bus Safety	BSAF	0.600	0.524	1.15	0.25	RP
Two-wheeler ownership	BTWO	0.900	0.662	1.36	0.17	RP

The above table shows the utility equations for both the modes, for RP data as well as SP data. The table shows the estimates for the coefficients of the variables included in the utility equation, generated by Biogeme for the demonstration model. As discussed above, the ASC_{Bus} is 0 for both RP and SP, but the same for auto-rickshaw has been estimated. In order to find the probability for each of the modes, the following formula can be used:

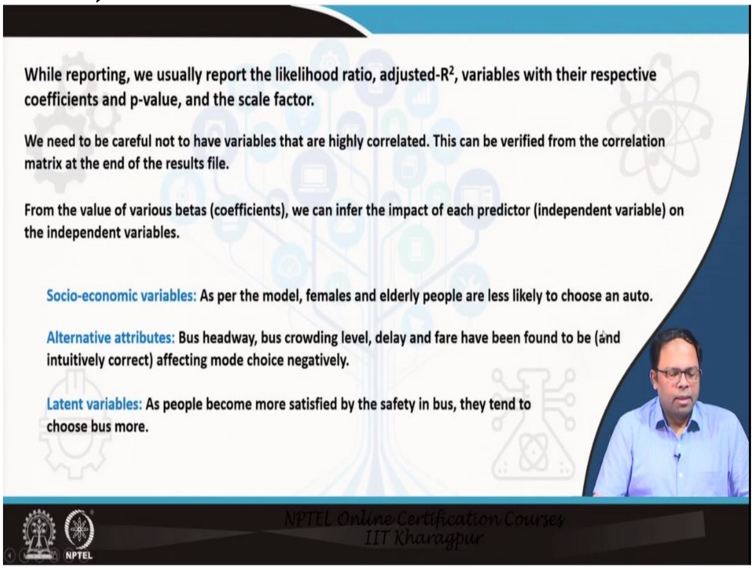
$$P_i^{RP} = \frac{e^{(\beta X_i^{RP} + \alpha Y_i^{RP})}}{\sum_j e^{(\beta X_j^{RP} + \alpha Y_j^{RP})}} \tag{1}$$

$$P_i^{SP} = \frac{e^{\theta.(\beta X_i^{SP} + \alpha Y_i^{SP})}}{\sum_j e^{\theta.(\beta X_j^{SP} + \alpha Y_j^{SP})}} \tag{2}$$

Same as the standard MNL model, the probability of selection of '**i^{th}**' alternative is estimated as the ratio of the exponential of the utility of '**i^{th}**' alternative, to the exponential of utility summed over all the alternatives. In the given case, probability of either of the modes (bus, auto-rickshaw) will be the exponential of the respective utility, divided by the sum of exponential of utility of bus and auto-rickshaw. The only difference between the probability estimation for the RP dataset from the SP dataset is that, a scale parameter (**Ө**) is multiplied with the utility equations of modes in SP.

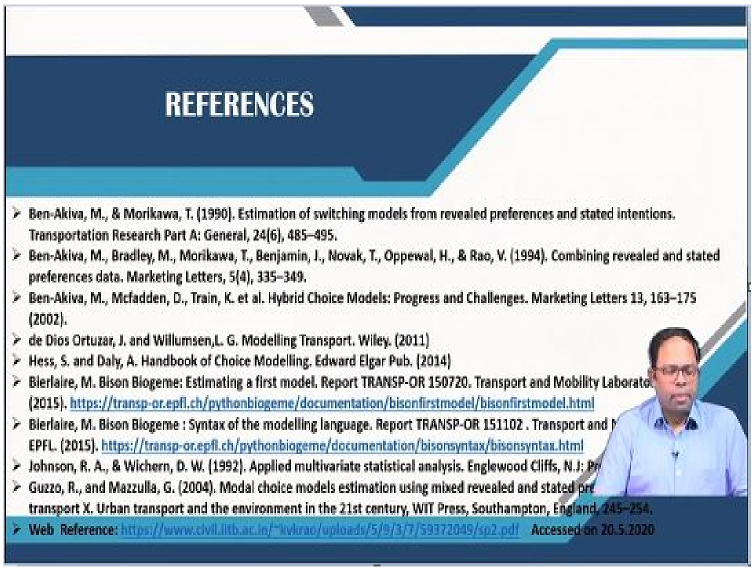
This is done to account for the different variability in both the datasets. **ϴ** has been estimated to be **0.207** for the demonstration model.

(Refer Slide Time: 28:35)



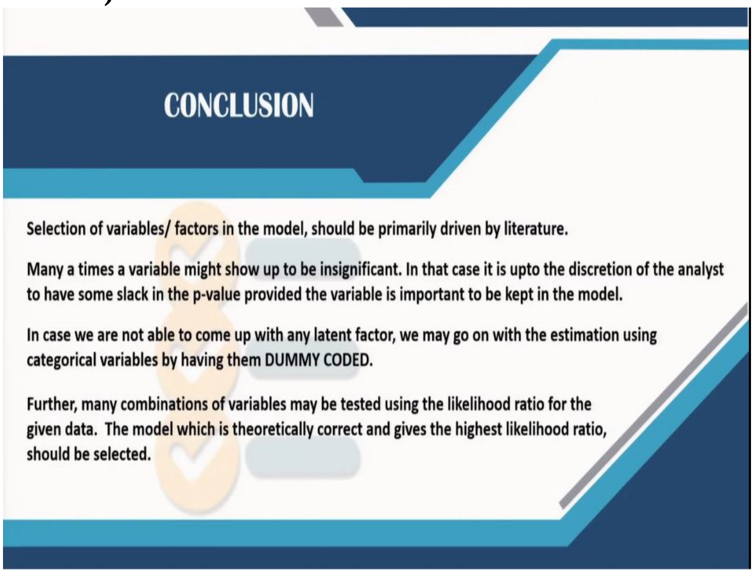
While reporting the model, there are certain values and statistics that should be present. Usually, the likelihood ratio; adjusted R square; coefficients and p value of the variables and the scale factor are reported. As already discussed, correlation should be checked carefully. The coefficients are just estimates, and in order to make the model more readable and informative, sensible inferences must be made using these coefficients. For example, as per the model, it was found that females and elderly people are less likely to choose an auto-rickshaw; with the increase in the value of alternative attributes like bus headway, bus crowding level, delay, and fare, bus ridership decreases; As people become more satisfied with the safety of the bus, they tend to choose bus over auto-rickshaw. So, this is how a joint RP-SP model is estimated.

(Refer Slide Time: 29:58)



Some references are given which can be referred for further reading.

(Refer Slide Time: 30:04)



In the conclusion, it can be said that selection of variables and factors in the model should be primarily driven by literature. Many times variables show up to be insignificant, but it is upto the discretion of the analyst to have some slack in the p-value, provided the variable is important to be kept in the model. For example, if a variable is very important it can be allowed if it is significant at 90%. In cases where no latent factors can be found, categorical variables can be used directly by having them dummy coded. Before selecting a final model, many combination of variables are tested using the likelihood ratio, and the underlying theory.