A structural equation model for the 2001 Monitoring the Future data

Contents

1. The data

The Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan has undertaken annual surveys designed to explore changes in important values, behaviors, and lifestyle orientations of contemporary American youth. The aims of these surveys are to provide a systematic, accurate description of the youth population of interest in a given year, and to explain relationships and trends observed over time. The Monitoring the Future surveys began in 1975. In the current example, data for 1608 respondents from the 2001 survey are used, and the focus is on relationships between the alcohol and marijuana use of respondents and traffic violations and/or accidents they were involved in.

Data for the first 10 participants on most of the variables used in this section are shown below in the form of a LSF named **select.LSF** which can be found in the **Complex survey sampling examples** folder.

select.lsf

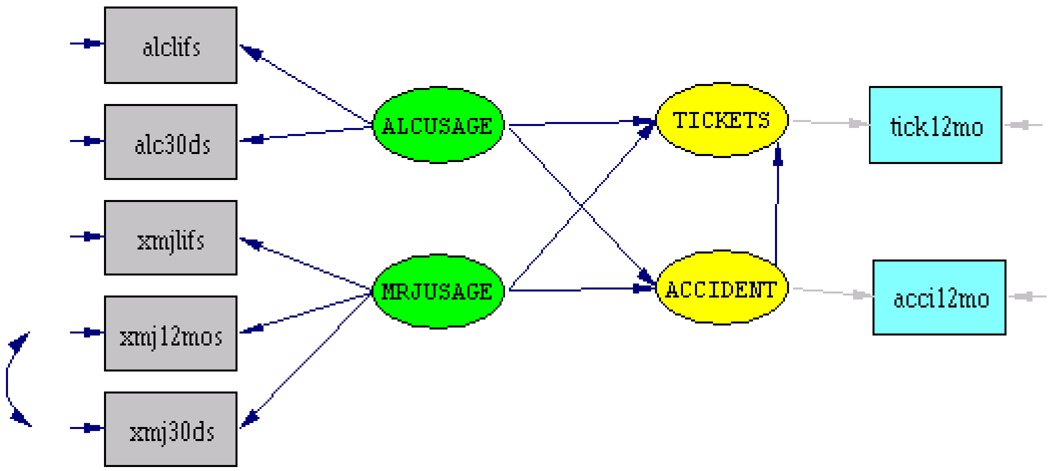
	school	region	alclifs	alcanns	alc30ds	xmjlifs	xmj12mos	xmj30ds
1	5.0	1.0	7.0	7.0	5.0	7.0	7.0	6.0
2	5.0	1.0	6.0	3.0	2.0	7.0	6.0	3.0
3	5.0	1.0	3.0	2.0	1.0	3.0	2.0	2.0
4	5.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
5	5.0	1.0	6.0	3.0	1.0	1.0	1.0	1.0
6	5.0	1.0	5.0	4.0	4.0	6.0	1.0	1.0
7	5.0	1.0	6.0	5.0	3.0	5.0	4.0	2.0
8	5.0	1.0	3.0	3.0	1.0	1.0	1.0	1.0
9	5.0	1.0	7.0	3.0	1.0	7.0	6.0	3.0
10	5.0	1.0	5.0	4.0	3.0	2.0	1.0	1.0

The following variables included in the LSF were selected from the survey data:

- school: This variable is used to indicate group membership of respondents within the 48 schools included in the survey.
- region: The 48 schools were drawn from 4 regions, and this variable indicates the region a school was drawn from.
- alclifs: The numerical response to the question "On how many occasions have you had alcoholic beverages to drink in your lifetime?"
- alc12mos: The numerical response to the question "On how many occasions have you had alcoholic beverages to drink in the past 12 months?"
- alc30ds: The numerical response to the question "On how many occasions have you had alcoholic beverages to drink in the past 30 days?"
- xmjlifs: The numerical response to the question "On how many occasions have you used marijuana in your lifetime?"
- xmj12mos: The numerical response to the question "On how many occasions have you used marijuana in the past 12 months?"
- xmj30ds: The numerical response to the question "On how many occasions have you used marijuana in the past 30 days?"
- tick12mo: The numerical response to the question "Within the last 12 months, how many times have you received a ticket (or been stopped and warned) for moving violations?"
- acci12mo: The numerical response to the question "Within the last 12 months, how many times you were involved in an accident while driving?"
- newwgt: The design weight of a student, computed as the inverse of the selection probability estimate of the region from which the student was selected. This selection probability estimate is merely the ratio of the sample frequency and the approximate population size of the region from which the student was selected.

2. The model

The five indicators or observed variables alclifs, alc30ds, xmjlifs, xmj12mos, and xmj30ds are modeled to measure the alcohol and marijuana usage. Alcohol and marijuana usage, represented by the latent variables ALCUSAGE and MRJUSAGE in our proposed model, are modeled as causes of the number of moving violations and accidents, as represented by the Eta variables ACCIDENT and TICKETS respectively. These two variables, in turn are measured without error by the two Y variables acci12mo and tick12mo. A path diagram of the model we intend fitting to the data is shown below.



Mathematical Model

Measurement model

The measurement model for the latent variables ALCUSAGE, MRJUSAGE, ACCIDENT and TICKETS may be expressed as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_x \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{bmatrix}$$

where $\mathbf{y} = [\text{ tick21mo} \quad \text{acci12mo }]'$, $\mathbf{x} = [\text{ alclifs} \quad \text{alc30ds} \quad \text{xmjlifs} \quad \text{xmj12mos} \quad \text{xmj30ds }]'$, $\boldsymbol{\eta} = [\text{ TICKETS ACCIDENT }]'$, $\boldsymbol{\xi} = [\text{ ALCUSAGE MRJUSAGE }]'$, $\boldsymbol{\delta} = [\delta_1 \quad \delta_2 \quad \delta_3 \quad \delta_4 \quad \delta_5 \quad \delta_6]'$, $\boldsymbol{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2]'$,

$$\mathbf{\Lambda_y} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$\mathbf{\Lambda_x} = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \\ 0 & \lambda_5 \end{bmatrix}$$

where δ_1, δ_2, δ_3, δ_4, δ_5, δ_6, ε_1 and ε_2 denote measurement errors, and where λ_1, λ_2, λ_3, λ_4 and λ_5 denote unknown factor loadings.

Structural equation model

The structural equation model for the latent variables ALCUSAGE, MRJUSAGE, ACCIDENT and TICKETS is given by

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

where $\boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 & \zeta_2 \end{bmatrix}'$,

$$\mathbf{B} = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix}$$

and

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{bmatrix}$$

where β, γ_1, γ_2, γ_3 and γ_4 denote unknown regression weights, and ζ_1 and ζ_2 denote error terms.

The survey design variables school and region will be used as stratification and cluster variable respectively, while the design weight as represented by the variable newwgt will also be included in the specification of the analysis, as illustrated next.

3 Preparing the data and setting up the analysis

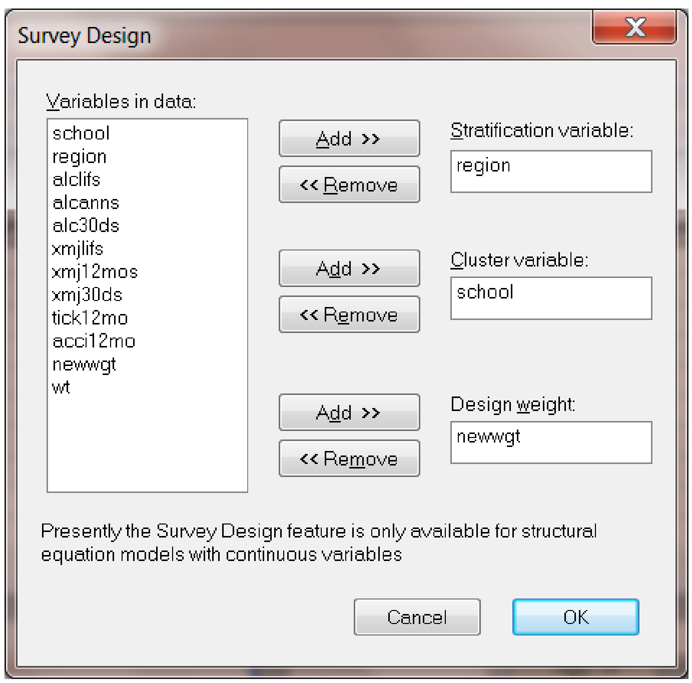
The model is fitted to the data in **select.LSF** by using the path diagram component (PTH window) of the LISREL GUI. After drawing the proposed model as a path diagram, SIMPLIS syntax is created and submitted. However, before we can fit the model, we need to specify the details of the complex survey design for the data in **select.LSF**.

The first step is to open the LSF, which is accomplished as follows:

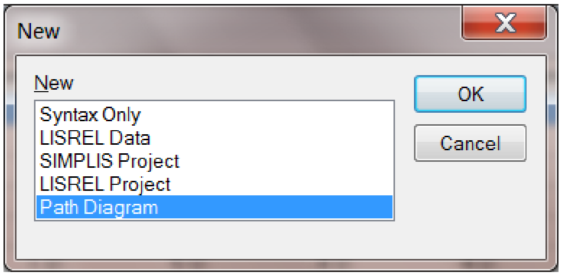
- Use the **File**, **Open** option to activate the display of an **Open** dialog box.
- Set the **Files of type** drop-down list box to **Lisrel Data (*.LSF)** and browse for the file **select.LSF**.
- Select the file and click the **Open** button to open the LSF in a LSF window.

3.1 Preparing the data

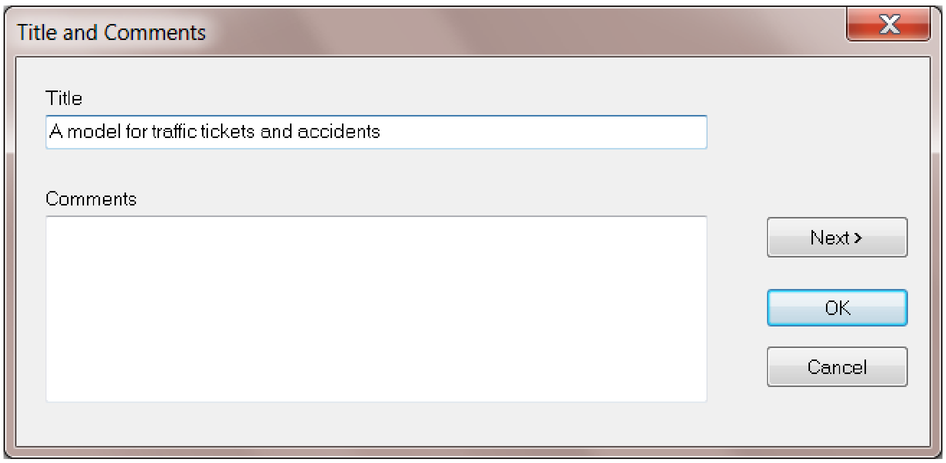
Click on the **Survey Design** option on the **Data** menu to load the **Survey Design** dialog box. Select the variable region from the **Variables in data:** list box and click on the **Add** button of the **Stratification variable** section. Next, select the variable school and add this variable to the **Cluster variable** section in a similar fashion. Finally, add the weight variable by selecting the variable newwgt from the **Variables in data:** list box and add this variable in the **Design weight** section. The completed **Survey Design** dialog box is shown below. Click on the **OK** button to return to the LSF window and click the **Save** option on the **File** menu.



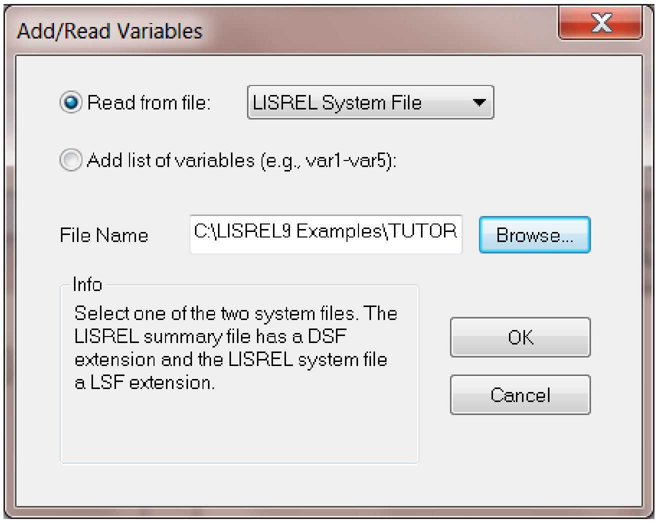
We now turn to creating a path diagram for the model to be fitted to these data. To open a new PTH window, select the **New** option on the **File** menu to load the **New** dialog box. Select the **Path Diagram** option from the list box on the **New** dialog box and provide a name for the path diagram, for example **select.pth**, in the **File name** string field of the **Save As** dialog box. Click on the **Save** button to open an empty PTH window.



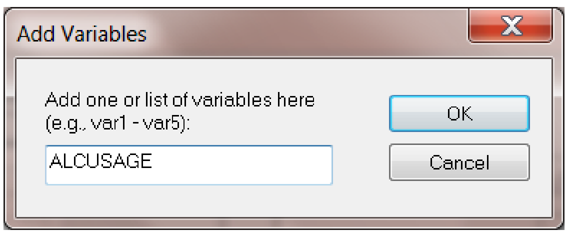
Select the **Title and Comments** option on the **Setup** menu to load the **Title and Comments** dialog box. Enter the title A model for traffic tickets and accidents in the **Title** string field and click on the **Next** button to load the **Group Names** dialog box.



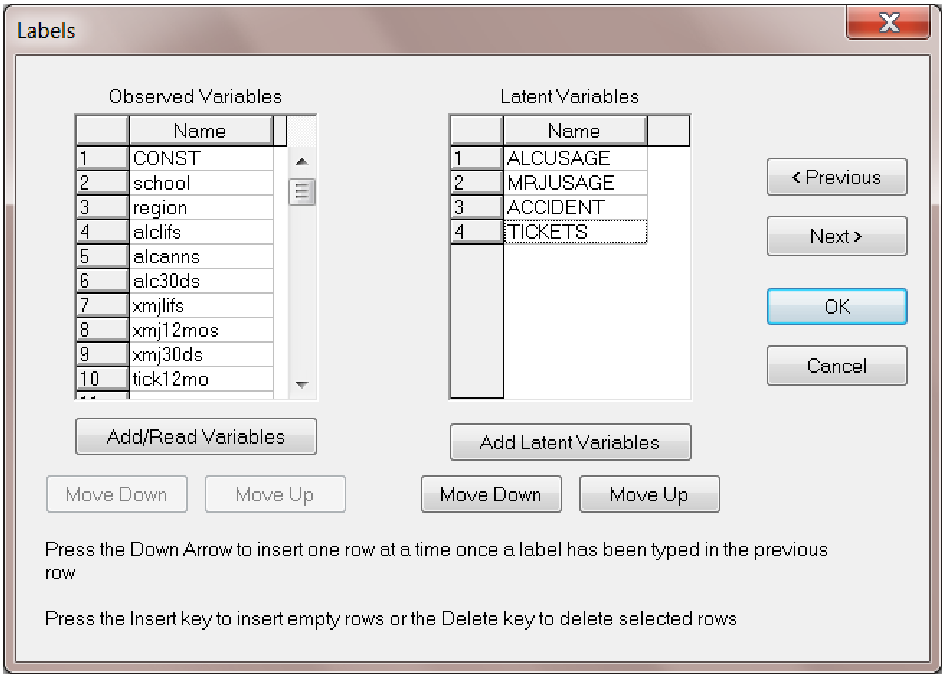
Click on the **Next** button to load the **Labels** dialog box. Click on the **Add/Read Variables** button to load the **Add/Read Variables** dialog box and select the **Lisrel System File** option in the **Read from file:** drop-down list box. Click on the **Browse** button to load the **Browse** dialog box and select the file **select.LSF** in the **TUTORIAL** folder. Click on the **OK** button to return to the **Labels** dialog box.



Click on the **Add Latent Variables** button to load the **Add Variables** dialog box. Enter the label ALCUSAGE in the string field and click **OK**. Enter the labels MRJUSAGE, ACCIDENT, and TICKETS in the same way.

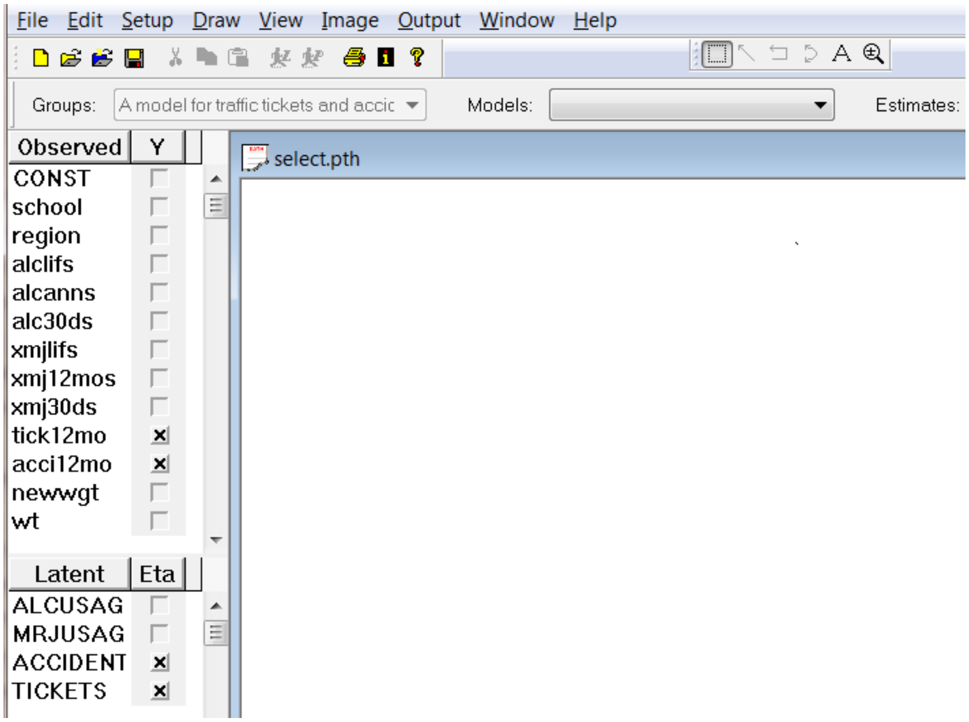


The completed **Labels** dialog box is shown below. Click on the **OK** button to return to the PTH window for **select.pth**.

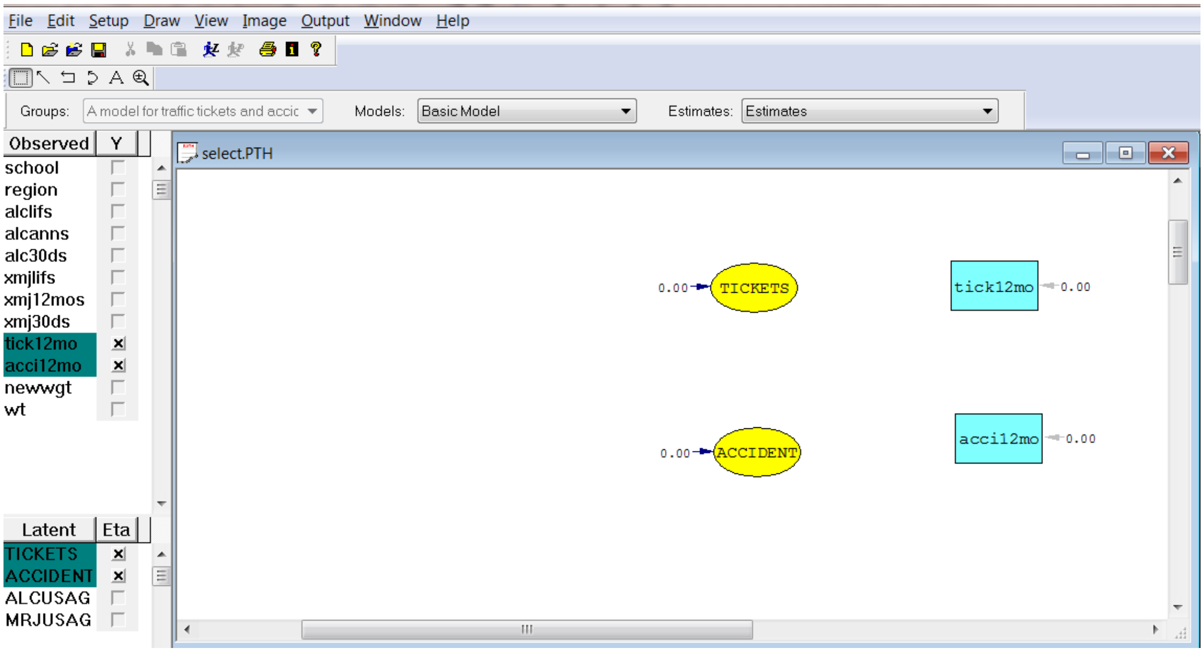


3.2 Setting up the analysis

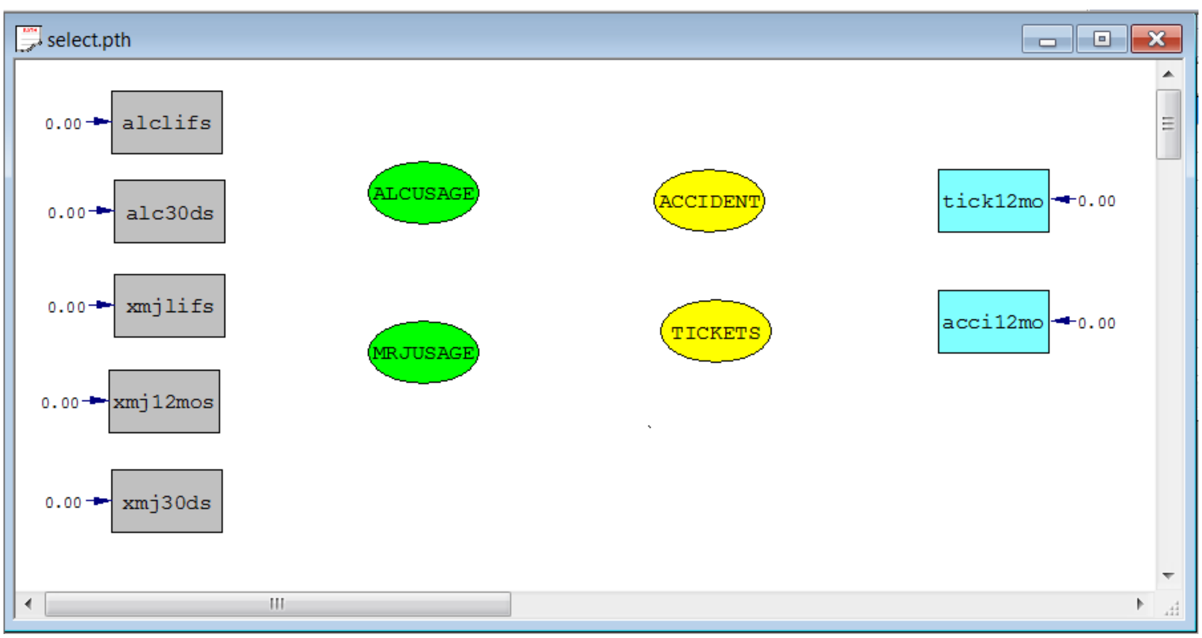
At this point, an empty PTH window is displayed, with variable names listed to the left. Check the **Y** check boxes of acci12mo and tick12mo respectively. Check the **Eta** check boxes of ACCIDENT and TICKETS respectively to obtain the window shown below.



Next, click, drag and drop the labels of the Y variables one at a time into the PTH window. Position these variables to the right of the PTH window. Click, drag and drop the labels of the latent variables ACCIDENT and TICKETS one at a time into the PTH window to obtain the window shown below. Note that labels of variables dragged to the PTH window are shown against a colored background.

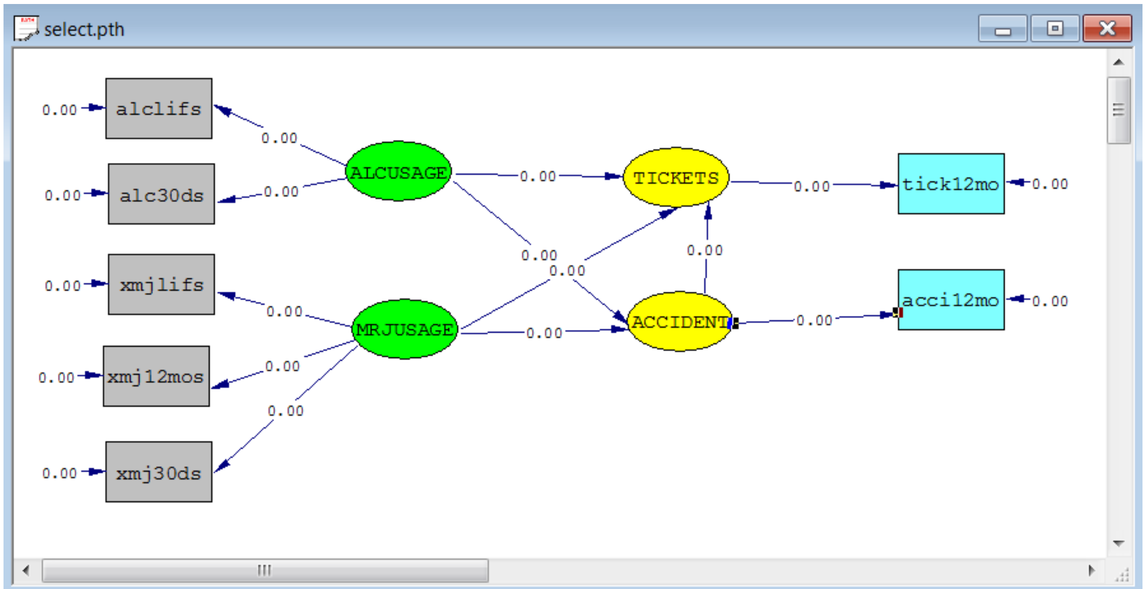


We now add the rest of the observed variables (alclifs, alc30ds, xmjlifs, xmj12mos, and xmj30ds) one at a time into the PTH window, positioning them to the left of the PTH window. The last variables to be added are the latent variables ALCUSAGE and MRJUSAGE.

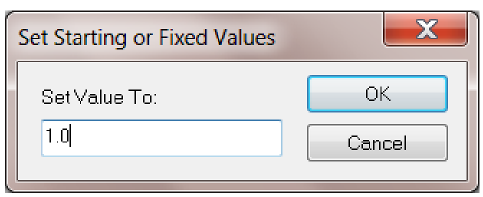


The next step is to add the paths between the variables dragged in the PTH window. Select the arrow icon on the **Drawing** toolbar and click and drag indicator paths from the latent variable ALCUSAGE to alclifs and alc30ds respectively. To do so, start by clicking inside the ellipse representing ALCUSAGE and do not release the mouse button before the cursor is inside the rectangle representing alclifs or alc30ds. Click and drag similar indicator paths from the latent variable MRJUSAGE to xmjlifs, xmj12mos and xmj30ds respectively.

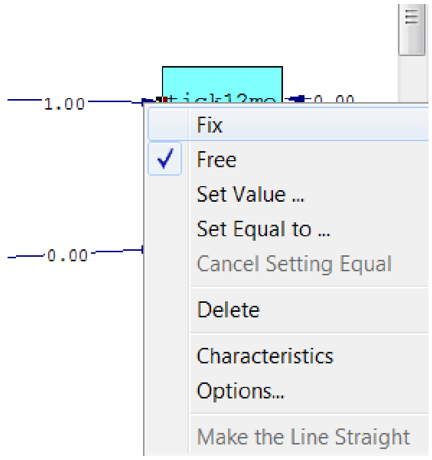
Structural paths from the latent variable ALCUSAGE to both ACCIDENT and TICKETS, and from the latent variable MRJUSAGE to ACCIDENT and TICKETS are added in the same way. Also add indicator paths from the latent variable ACCIDENT to acci12mo**,** from the latent variable TICKETS to tick12mo, and from TICKETS to ACCIDENT. The model should now look like the image below.



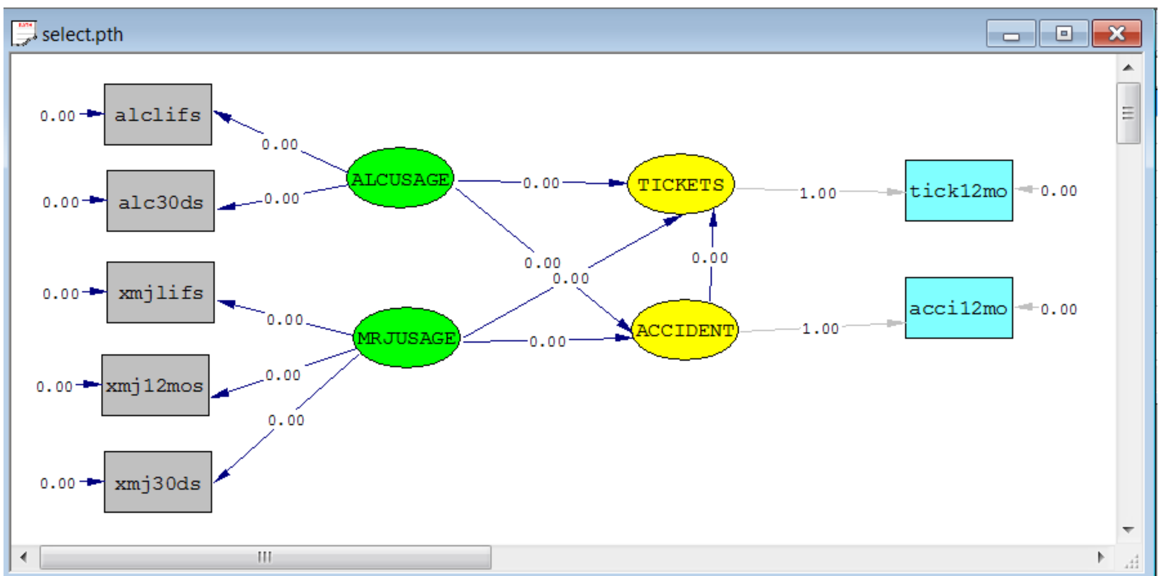
The two indicator paths ACCIDENT to acci12mo, and TICKETS to tick12mo have to be fixed to a value of 1.0. To do so, deselect the arrow icon on the **Drawing** toolbar by clicking on the selection icon to its left. Next, right click on the path between ACCIDENT to acci12mo and select the **Set Value** option from the pop-up menu that appears. Set the value to 1.0 and click **OK** to return to the PTH window.



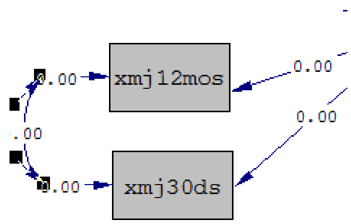
Right click on this path again and select the **Fix** option from the pop-up menu. Note that the color representing the path has changed in the PTH window. Set the path between TICKETS to tick12mo to 1.0 in the same way.



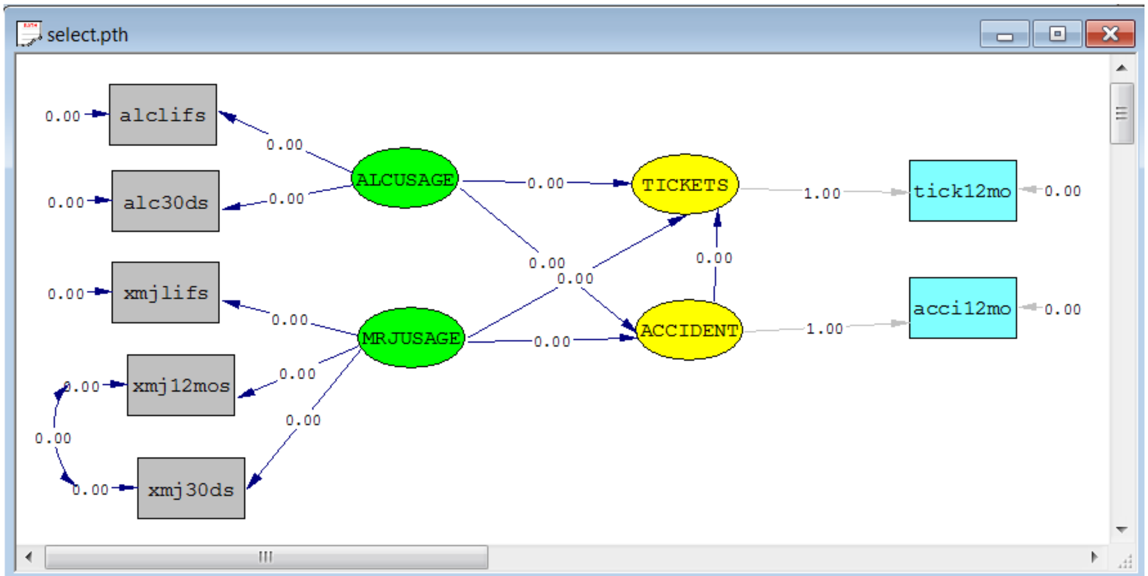
Finally, set the error variances of acci12mo and tick12mo to zero by right-clicking the error arrows and selecting the **Fix** option from the pop-up menu.



The last paths to be added to the path diagram are the covariance between the measurement errors of xmj12mos and xmj30ds. Select the double arrow icon on the **Drawing** toolbar and click and drag a path between the error arrows of xmj12mos and xmj30ds. Be sure to position the cursor over each arrow before activating and releasing the mouse button.



The path diagram should look like the following image.

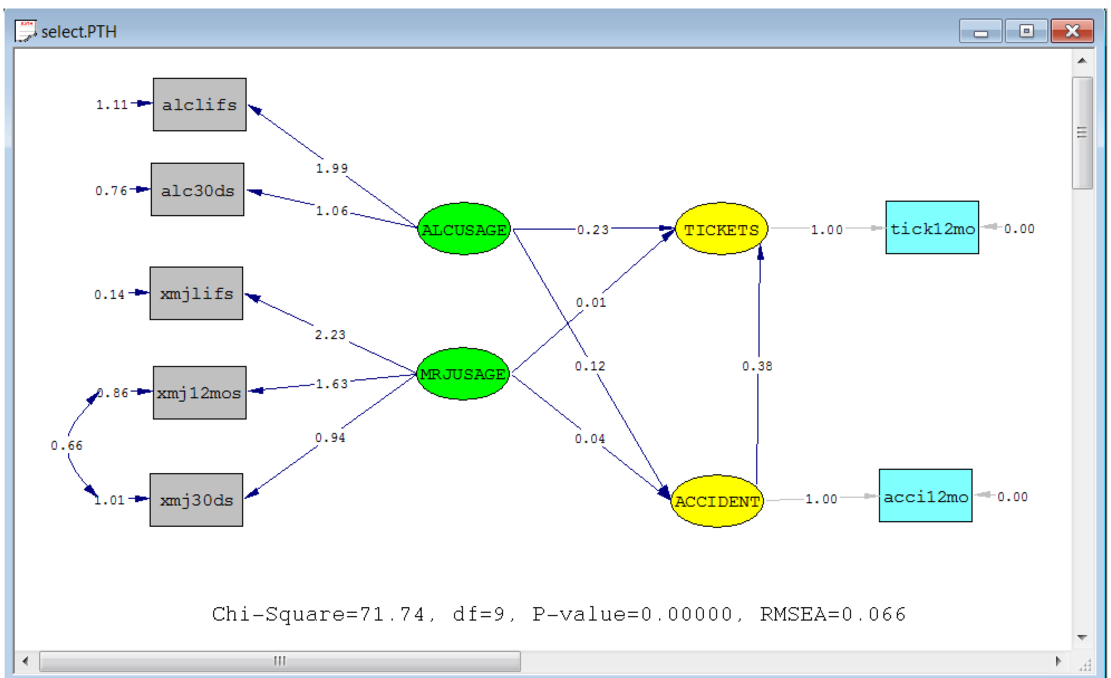


Select the **Build SIMPLIS Syntax** option on the **Setup** menu. The generated syntax is automatically displayed in a SPJ window, as shown below.

select.SPJ

```
Raw Data from file 'C:\LISREL9 Examples\TUTORIAL\select.lsf'
Latent Variables  TICKETS ACCIDENT ALCUSAGE MRJUSAGE
Relationships
tick12mo = 1.00*TICKETS
acci12mo = 1.00*ACCIDENT
alclifs = ALCUSAGE
alc30ds = ALCUSAGE
xmjlifs = MRJUSAGE
xmj12mos = MRJUSAGE
xmj30ds = MRJUSAGE
TICKETS = ACCIDENT
TICKETS = ALCUSAGE MRJUSAGE
ACCIDENT = ALCUSAGE MRJUSAGE
Set the Error Variance of tick12mo to 0.00
Set the Error Variance of acci12mo to 0.00
Set the Error Covariance of xmj30ds and xmj12mos Free
LISREL OUTPUT: ND=3
Path Diagram
End of Problem
```

Click on the **Run LISREL** icon on the main toolbar to produce the following PTH window.



4. Discussion of results

Portions of the output file **select.out** are shown below.

From the results, it is evident that the five factor loadings are statistically significant if a 1% level of significance is used. In addition, the error covariance for xmj12mos and xmj30ds is also significant at a 1% level of significance. In other words, the results do not indicate any misspecifications in the measurement model of the latent variables ALCUSAGE and MRJUSAGE.

Since $\hat{\beta} = 0.38 \quad (t = 11.03)$, it follows that a student's number of accidents exerts a significant positive influence $\left(p < 0.01 \right)$ on his/her number of traffic tickets. Thus an increase in the number of accidents corresponds to an increase in the number of traffic tickets. Similarly, it follows that the alcohol usage of a student is a significant antecedent $\left(p < 0.01 \right)$ of both the number of accidents and the number of traffic tickets of the student. On the other hand, the marijuana usage is not a statistical significant antecedent of both the student's number of accidents and traffic tickets.

LAMBDA-X

	ALCUSAGE	MRJUSAGE
alclifs	1.99	- -
	(0.05)	
	39.21	
alc30ds	1.06	- -
	(0.04)	
	28.72	
xmjlifs	- -	2.23
		(0.05)
		47.85
xmj12mos	- -	1.63
		(0.05)
		32.32
xmj30ds	- -	0.94
		(0.07)
		14.31

THETA-DELTA

	alclifs	alc30ds	xmjlifs	xmj12mos	xmj30ds
alclifs	1.11				
	(0.16)				
	6.72				
alc30ds	- -	0.76			
		(0.06)			
		12.37			
xmjlifs	- -	- -	0.14		
			(0.08)		
			1.76		
xmj12mos	- -	- -	- -	0.86	
				(0.08)	
				10.75	
xmj30ds	- -	- -	- -	0.66	1.01
				(0.04)	(0.11)
				17.07	9.45

The R^2 value for the number of accidents follows as 1 – 0.44 = 0.56. In other words, the alcohol usage and marijuana usage of a student explains approximately 56% of the variation in the student's number of accidents. Similarly, it follows that they explain approximately 27% of the variation in the number of traffic accidents of the student.

```
         BETA

            ACCIDENT    TICKETS
            --------   --------
ACCIDENT       - -        - -

 TICKETS      0.38        - -
            (0.03)
             11.03

         GAMMA

            ALCUSAGE   MRJUSAGE
            --------   --------
ACCIDENT      0.12       0.04
            (0.03)     (0.02)
              4.70       1.93

 TICKETS      0.23       0.01
            (0.05)     (0.04)
              4.90       0.17

         PSI
         Note: This matrix is diagonal.

            ACCIDENT    TICKETS
            --------   --------
              0.44       0.73
            (0.03)     (0.05)
             16.10      14.42
```

select.OUT

```
             Global Goodness of Fit Statistics, FIML case

             Chi-Square Scale Factor =      0.73392
             Number of Strata    =             4
             Number of Clusters =             28


               -2ln(L) for the saturated model =      33407.994
               -2ln(L) for the fitted model    =      33505.738

 Degrees of Freedom = 9
 Full Information ML Chi-Square                      71.737 (P = 0.0000)
 Root Mean Square Error of Approximation (RMSEA)     0.0658
 90 Percent Confidence Interval for RMSEA            (0.0522 ; 0.0804)
 P-Value for Test of Close Fit (RMSEA < 0.05)        0.0290
```

From the results above, it is evident that the χ^2 test statistic value for the null hypothesis of a perfect fit is significant if a 1% level of significance is used. There is sufficient evidence that the theoretical model does not fit the data perfectly. However, the RMSEA point estimate of 0.0658 indicates that the model does provide a reasonable approximation to the data (Browne & Cudeck, 1993).