



## Bivariate regression

The following example illustrates the case of two dependent variables  $y_1$  and  $y_2$ , and three explanatory variables  $x_1$ ,  $x_2$ , and  $x_3$ . Finn (1974) presents the data given in the table below.

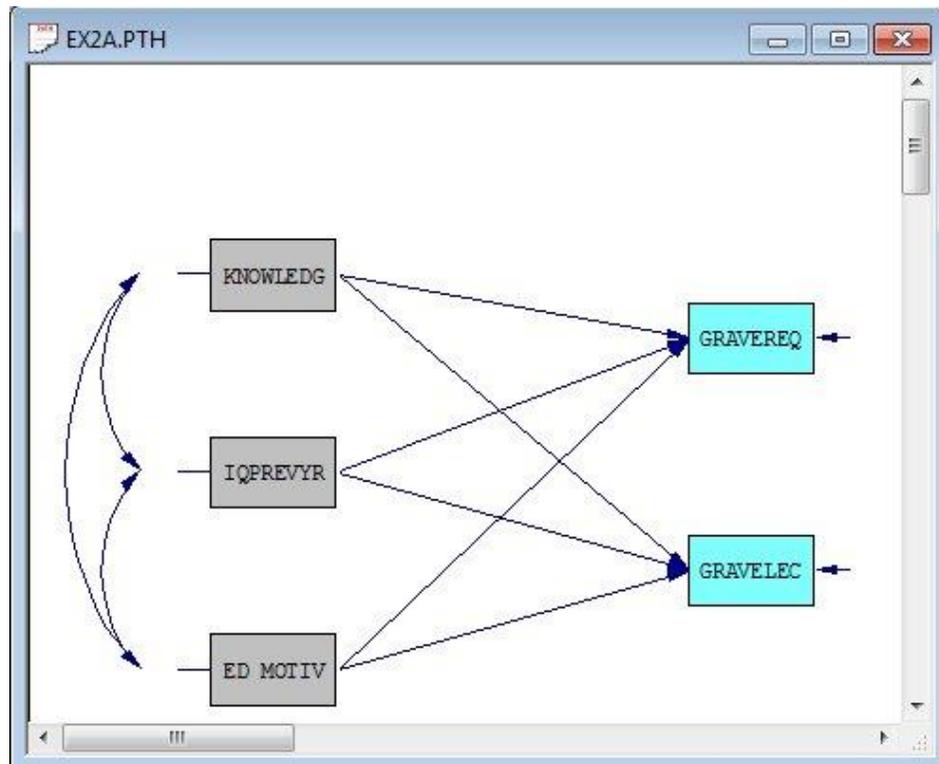
**Table 1: Scores for fifteen college freshmen on five educational measures**

Case	$y_1$	$y_2$	$x_1$	$x_2$	$x_3$
1	0.8	2.0	72	114	17.3
2	2.2	2.2	78	117	17.6
3	1.6	2.0	84	117	15.0
4	2.6	3.7	95	120	18.0
5	2.7	3.2	88	117	18.7
6	2.1	3.2	83	123	17.9
7	3.1	3.7	92	118	17.3
8	3.0	3.1	86	114	18.1
9	3.2	2.6	88	114	16.0
10	2.6	3.2	80	115	16.4
11	2.7	2.8	87	114	17.6
12	3.0	2.4	94	112	19.5
13	1.6	1.4	73	115	12.7
14	0.9	1.0	80	111	17.0
15	1.9	1.2	83	112	16.1

These data represent the scores of fifteen freshmen at a large midwestern university on five educational measures. The five measures are:

- $y_1$  = grade average for required courses taken (GRAVEREQ)
- $y_2$  = grade average for elective courses taken (GRAVELEC)
- $x_1$  = High school general knowledge test, taken previous year (KNOWLEDG)
- $x_2$  = IQ score from previous year (IQPREVYR)
- $x_3$  = educational motivation score from previous year (ED MOTIV)

We examine the predictive value of  $x_1$ ,  $x_2$ , and  $x_3$  in predicting the grade averages  $y_1$  and  $y_2$ . The path diagram of the model is shown below.



The SIMPLIS input for such a model is extremely simple:

```

EX2A.SPL
Prediction of Grade Averages
Observed Variables: GRAVEREQ GRAVELEC KNOWLEDG IQPREVYR 'ED MOTIV'
Sample Size: 15
Raw Data
0.8 2.0 72 114 17.3
2.2 2.2 78 117 17.6
1.6 2.0 84 117 15.0
2.6 3.7 95 120 18.0
2.7 3.2 88 117 18.7
2.1 3.2 83 123 17.9
3.1 3.7 92 118 17.3
3.0 3.1 86 114 18.1
3.2 2.6 88 114 16.0
2.6 3.2 80 115 16.4
2.7 2.8 87 114 17.6
3.0 2.4 94 112 19.5
1.6 1.4 73 115 12.7
0.9 1.0 80 111 17.0
1.9 1.2 83 112 16.1
Relationships
  GRAVEREQ = KNOWLEDG  IQPREVYR  'ED MOTIV'
  GRAVELEC = KNOWLEDG  IQPREVYR  'ED MOTIV'
Path Diagram
End of Problem

```

The variables are listed by labels. Note that since the label for the last variable includes a blank space, it must be included within single quotes.

The data for this example is raw data which is read in free format with one line per case. The sample size need not be specified. The program determines the sample size by counting the number of lines of data. In applications here the raw data contain many cases, it is recommended that the raw data be stored in a data files rather than in the input file.

The relationships specify that there are two equations to be estimated, one for each dependent variable. Each equation is specified by listing the dependent variable first and then the three independent variables.

The three right-hand variables in each equation are consecutive variables, so one can replace the second right-hand variable by a hyphen and write the two relationships:

```
GRAVEREQ = KNOWLEDG - 'ED MOTIV'
GRAVELEC = KNOWLEDG - 'ED MOTIV'
```

The two relationships can be simplified even further by recognizing that the two left-hand variables are also consecutive variables and the same right-hand variables are involved in each relationship. The two relationships can therefore be specified on a single line:

```
GRAVEREQ - GRAVELEC = KNOWLEDG - 'ED MOTIV'
```

This is interpreted by the program as "Each of the variables from GRAVEREQ to GRAVELEC depends on all of the variables from KNOWLEDG to ED MOTIV".

Alternatively, one can specify the model by paths:

```
KNOWLEDG - 'ED MOTIV' -> GRAVEREQ - GRAVELEC
```

which is short for:

```
KNOWLEDG - 'ED MOTIV' -> GRAVEREQ
KNOWLEDG - 'ED MOTIV' -> GRAVELEC
```

The two relationships are estimated as

#### Structural Equations

```
GRAVEREQ = 0.0854*KNOWLEDG + 0.00822*IQPREVYR - 0.0149*ED MOTIV, Errorvar.= 0.257 ,
R2 = 0.568
Standerr (0.0258)          (0.0465)          (0.107)          (0.105)
Z-values  3.309           0.177           -0.140          2.449
P-values  0.001           0.860           0.889           0.014
```

GRAVELEC = 0.0472\*KNOWLEDG + 0.145\*IQPREVYR + 0.126\*ED MOTIV, Errorvar.= 0.237 , R<sup>2</sup> = 0.685

Standerr	(0.0248)	(0.0447)	(0.103)	(0.0969)
Z-values	1.904	3.255	1.222	2.449
P-values	0.057	0.001	0.222	0.014

The *t*-values show that only KNOWLED is a significant predictor of GRAVEREQ, and only IQPREVYR is a significant predictor of GRAVELEC. ED MOTIV is not significant for either purpose. It should be emphasized, however, that the sample size is too small to draw any safe conclusions.

The output also contains a section called GOODNESS OF FIT STATISTICS, the first line of which is:

Degrees of Freedom for (C1)-(C2)	1
Maximum Likelihood Ratio Chi-Square (C1)	9.521 (P = 0.0020)

LISREL assumes by default that the two error terms  $z_1$  and  $z_2$  are uncorrelated. This assumption means that  $y_1$  and  $y_2$  would be uncorrelated after the effects of  $x_1$ ,  $x_2$ , and  $x_3$  are removed. The chi-square value given in the output file is a formal test of this assumption. Since it is significant, one may be interested in estimating the error covariance, i.e., the covariance between  $z_1$  and  $z_2$ . This can be done by including the following line in the input file (see **EX2B.SPL** in the **Simplis Examples** folder):

Let the Error Terms on GRAVEREQ and GRAVELEC be Correlated

Alternatively, this may be specified as:

Set the Error Covariance between GRAVEREQ and GRAVELEC Free

The output file for the model with correlated error terms shows the same estimate relationships and then gives the estimated covariance matrix of the two error terms  $z_1$  and  $z_2$ :

Error Covariance for GRAVELEC and GRAVEREQ =	0.169
	(0.0903)
	1.875

The error covariance is 0.17. The partial correlation between  $y_1$  and  $y_2$  for given  $x_1$ ,  $x_2$ , and  $x_3$  is  $0.17 / \sqrt{(0.26 \times 0.24)} = 0.68$ .

Suppose we want to test the hypothesis that ED MOTIV has no effect on either GRAVEREQ nor GRAVELEC. One might think that this is just a matter of deleting ED MOTIV from the two relationships in the input file **EX2B.SPL**. However, this will not work because the resulting model is still saturated and fits the data perfectly. So no test of the hypothesis is provided by the program.

To test the hypothesis, one should instead specify the relationships as

```
GRAVEREQ = KNOWLEDG IQPREVYR 0*'ED MOTIV'  
GRAVELEC = KNOWLEDG IQPREVYR 0*'ED MOTIV'
```

The resulting output gives a chi-square of 3.40 with 2 degrees of freedom. So the hypothesis is not rejected.

With the SIMPLIS command language, selection of variables is automatic in the sense that only the variables involved in the model are used even if there are more variables in the data. Thus, when ED MOTIV is deleted from the relationships, the program uses only the first four variables and estimates the two regression of GRAVEREQ and GRAVELEC on KNOWLEDG and IQPREVYR, and with correlated error terms, this is a saturated model. When the relationships are specified as above, ED MOTIV is still in the model but its effect on GRAVEREQ and GRAVELEC is forced to be zero.