Estimated regressions

Finn (1974) presents the data shown in the table below. These data represent the scores of fifteen freshmen at a large Midwestern university in the United States, on five education measures:

- y_1 = grade average in required courses

- y_2 = grade average in elective courses

- x_1 = score on general knowledge test taken year before

- x_2 = score on IQ test from year before

- x_3 = score on educational motivation test from year before

Table: Scores for fifteen college freshmen on five educational measures

Observation	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

In this example (**EX4.PRL** in the **PRELIS Examples** folder), PRELIS estimates the regressions of y_1 and y_2 on x_1, x_2, and x_3.

```
EXAMPLE 4: PREDICTION OF GRADE AVERAGES
DA NI=5 NO=15;RA FI=DATA.EX4;CO ALL;RG 1 2 ON 3 4 5;OU MA=CM
```

The commands on the line above illustrate how compact input for this problem can be.

Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Freq.	Maximum	Freq.
VAR 1	2.267	0.771	-0.709	-0.514	0.800	1	3.200	1
VAR 2	2.513	0.868	-0.362	-0.936	1.000	1	3.700	2
VAR 3	84.200	6.889	-0.201	-0.500	72.000	1	95.000	1
VAR 4	115.533	3.204	0.836	0.728	111.000	1	123.000	1
VAR 5	17.013	1.641	-1.246	2.499	12.700	1	19.500	1

Covariance Matrix

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5
VAR 1	0.594				
VAR 2	0.483	0.754			
VAR 3	3.993	3.626	47.457		
VAR 4	0.426	1.757	4.100	10.267	
VAR 5	0.499	0.716	6.261	0.557	2.694

 Total Variance = 61.766 Generalized Variance = 28.391

 Largest Eigenvalue = 49.409 Smallest Eigenvalue = 0.076

 Condition Number = 25.511
Estimated Equations

```
   VAR 1 =  - 5.619 + 0.0854*VAR 3 + 0.00822*VAR 4 - 0.0149*VAR 5
Standerr     (5.614) (0.0270)        (0.0485)          (0.112)
t-values     -1.001   3.168           0.169            -0.134
P-values      0.337   0.008           0.868             0.896
```

$$+ \text{Error}, R^2 = 0.568$$

Error Variance = 0.327

```
   VAR 2 =  - 20.405 + 0.0472*VAR 3 + 0.145*VAR 4 + 0.126*VAR 5
Standerr     (5.398) (0.0259)        (0.0467)        (0.107)
t-values     -3.780   1.823           3.117           1.170
P-values      0.003   0.093           0.009           0.265
```

$$+ \text{Error}, R^2 = 0.685$$

```
Error Variance = 0.302


The following chi-squares test the hypothesis that all
regression coefficients are zero except the intercept.
Variable        -2lnL  Chi-square   df  Covariates
--------    ----------  ----------   --  ----------
   VAR 1       21.140      12.575     3  VAR 3 VAR 4 VAR 5
   VAR 2       19.961      17.338     3  VAR 3 VAR 4 VAR 5

Analysis of Variance Table
   Regression d.f.     Residual d.f.           F  Covariates
---------------     -------------           -  ----------
      4.718     3       3.595    11       4.812  VAR 3 VAR 4 VAR 5
      7.234     3       3.323    11       7.982  VAR 3 VAR 4 VAR 5
```

The *t*-values reveal immediately that only x_1 is a significant predictor of y_1, and only x_2 is a significant predictor of y_2. The variable x_3 is not significant for either purpose. It should be noted, however, that the sample size is too small to draw any safe conclusions.