



## Regression: predictor of grade averages

A regression model is a model of the form

$$\mathbf{y} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x} + \mathbf{u},$$

where  $\mathbf{y}' = (y_1, y_2, \dots, y_p)$  is a set of jointly dependent variables,  $\mathbf{x}' = (x_1, x_2, \dots, x_q)$  is a set of explanatory variables uncorrelated with the error terms  $\mathbf{u}' = (u_1, u_2, \dots, u_p)$ ,  $\boldsymbol{\alpha}$  is a vector of intercept terms and  $\boldsymbol{\Gamma}$  is a matrix of regression parameters.

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. These data (given in file **GRAV.RAW** in the **PRELIS Examples** folder) 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  $y_2$ .

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

To estimate the regression of  $y_1$  and  $y_2$  on  $x_1$ ,  $x_2$ , and  $x_3$  use the following PRELIS command file (see file **GRAV.PRL**).

#### Prediction of Grade Averages

```
DATA NI=5
LABELS
GRAVEREQ GRAVELEC KNOWLEDG IQPREVYR 'ED MOTIV'
RAWDATA=GRAV.RAW
CONTINUOUS ALL
RG GRAVEREQ GRAVELEC ON KNOWLEDG IQPREVYR 'ED MOTIV'
OUTPUT MATRIX=CM
```

The result is given in the output as

#### Estimated Equations

GRAVEREQ =	- 5.619	+ 0.0854*KNOWLEDG	+ 0.00822*IQPREVYR	- 0.0149*ED MOTIV
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

+ Error,  $R^2 = 0.568$

Error Variance = 0.327

GRAVELEC =	- 20.405	+ 0.0472*KNOWLEDG	+ 0.145*IQPREVYR	+ 0.126*ED MOTIV
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

+ Error,  $R^2 = 0.685$