**Employing Matrix Operations for Statistical Procedures**

Consider the data set presented earlier, with a few more subjects.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | SAT | GPA | Self-Esteem | IQ |
| 1 | 560 | 3.0 | 11 | 112 |
| 2 | 780 | 3.9 | 10 | 143 |
| 3 | 620 | 2.9 | 19 | 124 |
| 4 | 600 | 2.7 | 7 | 129 |
| 5 | 720 | 3.7 | 18 | 130 |
| 6 | 380 | 2.4 | 13 | 82 |

The data matrix could be conceived of as 6 rows and 4 columns:

 X = 

Since so much of statistics is dependent on the use of deviation scores, we will compute a matrix of means, where each column corresponds to the mean of the corresponding column mean of X, and subtract that from the original matrix X.

D = X - M = - =

 -50 -.1 -2 -8

 170 .8 -3 23

 10 -.2 6 4

 -10 -.4 -6 9

 110 .6 5 10

 -230 -.7 0 -38

Using SPSS, the following commands produce the deviation matrix.

COMPUTE X = {560,3.0,11,112;780,3.9,10,143;620,2.9,19,124;

 600,2.7,7,129;720,3.7,18,130;380,2.4,13,82}.

COMPUTE ONES = MAKE(6,1,1).

COMPUTE M = ONES\*T(ONES)\*X\*(1/6).

COMPUTE D = X - M.

Because D is now the deviation matrix, D′D provides the sums of squares and cross products matrix, the SSCP matrix.

 SumSCP = 

*Note*: Because there is a SSCP function in SPSS, we cannot call a matrix “SSCP” so you will see it called “SumSCP.”

Using SPSS, the following commands produce the SumSCP matrix.

COMPUTE SumSCP = T(D)\*D.

*or*

COMPUTE SumSCP = SSCP(D).

SUMSCP

 10 \*\* 4 X

 9.660000000 .037000000 .026000000 1.410000000

 .037000000 .000170000 .000200000 .004740000

 .026000000 .000200000 .011000000 -.003300000

 1.410000000 .004740000 -.003300000 .223400000

*Note*: SPSS employs scientific notation, where each value is × 104 in this case.

Recall that the diagonal of the SumSCP matrix contains the sums of squares for each column (variable). The off-diagonal elements are the cross-products. To obtain the variances and covariances, we take the average sums of squares and average cross-products (when computed using deviation scores). We will call this matrix SIGMA.

 **Σ** = SIGMA = (1/n)SumSCP = 

COMPUTE SIGMA = 1/6\*SumSCP.

SIGMA

 10 \*\* 4 X

 1.610000000 .006166667 .004333333 .235000000

 .006166667 .000028333 .000033333 .000790000

 .004333333 .000033333 .001833333 -.000550000

 .235000000 .000790000 -.000550000 .037233333

We can convert SIGMA to a correlation matrix by standardizing, dividing each element by its corresponding standard deviation, which is equivalent to pre- and post-multiplying SIGMA by a scaling matrix where the diagonal contains the corresponding standard deviations.

Employing SPSS, we can obtain the scaling matrix by pulling out the diagonal of SIGMA and creating a diagonal matrix (S) that contains the inverse of each standard deviation (square root of the variances).

COMPUTE V = DIAG(SIGMA).

COMPUTE S = INV(MDIAG(SQRT(V))).

V

 10 \*\* 4 X

 1.610000000

 .000028333

 .001833333

 .037233333

S

 .007881104 .000000000 .000000000 .000000000

 .000000000 1.878672873 .000000000 .000000000

 .000000000 .000000000 .233549683 .000000000

 .000000000 .000000000 .000000000 .051824371

Once we have the SumSCP matrix and a scaling matrix (S), we can pre- and post-multiply SumSCP by S and obtain the correlation matrix.

COMPUTE R = S\*SIGMA\*S.

R

 1.000000000 .913037679 .079760605 .959818165

 .913037679 1.000000000 .146254485 .769152218

 .079760605 .146254485 1.000000000 -.066569610

 .959818165 .769152218 -.066569610 1.000000000

In full matrix notation, here are the procedures we employed:

|  |  |  |
| --- | --- | --- |
| *Operation* | *SPSS Syntax* | *Matrix Notation* |
| 1. Create matrix X | COMPUTE X = {…}. | X |
| 2. Compute a ones vector | COMPUTE ONES = Make (6,1,1). | 1 |
| 3. Compute the means matrix | COMPUTE M = ONES\*T(ONES)\*X\*(1/6). | 1 1′ X (1/16) |
| 4. Compute the deviation matrix | COMPUTE D = X - M.  | X – M |
| 5. Compute sums of squares and cross products matrix | COMPUTE SumSCP = T(D)\*D. | D′ D |
| 6. Compute Sigma, the variance/covariance matrix | COMPUTE SIGMA = 1/6\*SumSCP. | (1/6)SumSCP |
| 7. Obtain a diagonal matrix of variances | COMPUTE V = DIAG(SIGMA). | **V** |
| 8. Compute a scaling matrix, inverse of standard deviations | COMPUTE S = INV(MDIAG(SQRT(V))). | S |
| 9. Compute the correlation matrix | COMPUTE R = S\*SIGMA\*S. | R = S **Σ S** |
| ***alternatively*** |
| 10. Compute standardized scores | COMPUTE Z = D\*S. | Z = D S |
| 11. Compute the correlation matrix | COMPUTE R2 = (1/6)\*T(Z)\*Z. | R = (1/16) Z′ Z |