Multiple regression - Matrices

This handout will present various matrices which are substantively interesting and/or provide useful means of summarizing the data for analytical purposes. As we will see, means, standard deviations, and correlations are substantively interesting; other matrices are primarily useful for computational purposes.

Let $X_0 = 1$ for all cases. Let $A$ and $B$ be any two variables. We then define the following: (Note: To simplify the notation, when subscripting the $X$ variables, we will refer to them only by number, e.g. $s_{12}$ = sample covariance of $X_1$ and $X_2$.)

$$M_{AB} = \sum A \cdot B,$$

$$XP_{AB} = \sum (A_i - \bar{A})(B_i - \bar{B}) = M_{AB} - N\bar{A}\bar{B} = M_{AB} - M_{AB}M_{B0}/M_{00},$$

$$s_{AB} = \frac{XP_{AB}}{N - 1} = \frac{MP_{AB} - N\bar{A}\bar{B}}{N - 1} = \frac{\sum (A_i - \bar{A})(B_i - \bar{B})}{N - 1},$$

$$r_{AB} = \frac{s_{AB}}{\sqrt{s_{A}^2s_{B}^2}} = \frac{s_{AB}}{s_{A}^2s_{B}} = \frac{XP_{AB}}{\sqrt{XP_{AA}XP_{BB}}} = \frac{1}{N - 1} \sum \frac{(A_i - \bar{A})}{\sqrt{s_{A}^2}} \frac{(B_i - \bar{B})}{\sqrt{s_{B}^2}}.$$

Note that each of these are symmetric, e.g. $M_{1y} = M_{y1}$, $s_{12} = s_{21}$. We will discuss each of these in turn:

1. $M_{AB}$. Let us compute the values of $M$ for the variables $X_0$, $X_1$, $X_2$, $Y$.

<table>
<thead>
<tr>
<th>$M_{AB}$</th>
<th>Formula: $\Sigma (A \cdot B)$</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_{00}$</td>
<td>$\Sigma (X_0 \cdot X_0) = \Sigma 1 = N$</td>
<td>20.0</td>
</tr>
<tr>
<td>$M_{01} = M_{10}$</td>
<td>$\Sigma (X_0 \cdot X_1) = \Sigma X_1$</td>
<td>241.0</td>
</tr>
<tr>
<td>$M_{02} = M_{20}$</td>
<td>$\Sigma (X_0 \cdot X_2) = \Sigma X_2$</td>
<td>253.0</td>
</tr>
<tr>
<td>$M_{0y} = M_{y0}$</td>
<td>$\Sigma (X_0 \cdot Y) = \Sigma Y$</td>
<td>488.3</td>
</tr>
<tr>
<td>$M_{11}$</td>
<td>$\Sigma (X_1 \cdot X_1) = \Sigma X_1^2$</td>
<td>3,285.0</td>
</tr>
<tr>
<td>$M_{12} = M_{21}$</td>
<td>$\Sigma (X_1 \cdot X_2) = \Sigma X_1X_2$</td>
<td>2,999.0</td>
</tr>
<tr>
<td>$M_{1y} = M_{y1}$</td>
<td>$\Sigma (X_1 \cdot Y) = \Sigma X_1Y$</td>
<td>6,588.3</td>
</tr>
<tr>
<td>$M_{22}$</td>
<td>$\Sigma (X_2 \cdot X_2) = \Sigma X_2^2$</td>
<td>3,767.0</td>
</tr>
<tr>
<td>$M_{2y} = M_{y2}$</td>
<td>$\Sigma (X_2 \cdot Y) = \Sigma X_2Y$</td>
<td>6448.9</td>
</tr>
<tr>
<td>$M_{yy}$</td>
<td>$\Sigma (Y \cdot Y) = \Sigma Y^2$</td>
<td>13,742.3</td>
</tr>
</tbody>
</table>

Hence, to get the different possible values of $M_{AB}$, we multiply every variable (including $X_0$) by every variable. These numbers ought to look familiar. $M_{00} = N = 20$; The other numbers are the...
totals we got when we first presented the data. As we have seen, the different values of $M_{AB}$ contain all the information we need for calculating regression models.

It is often convenient to present the values of $M_{AB}$ in matrix form. We can write

$$
\begin{pmatrix}
X_0 & X_1 & X_2 & Y \\
X_0 & 20.0 & & \\
X_1 & 241.0 & 3,285.0 & \\
X_2 & 253.0 & 2,999.0 & 3,767.0 \\
Y & 488.3 & 6,588.3 & 6,448.9 & 13,742.27
\end{pmatrix}
$$

or,

$$
\mathbf{M} = \begin{bmatrix}
20.0 \\
241.0 & 3,285.0 \\
253.0 & 2,999.0 & 3,767.0 \\
488.3 & 6,588.3 & 6,448.9 & 13,742.27
\end{bmatrix}
$$

Additional comments:

a. For symmetric matrices, it is common to not present either the elements above or the elements below the diagonal, since they are redundant.

b. The Matrix $M/N$ (i.e. the $M$ matrix with all elements divided by the sample size $N$) is sometimes called the augmented moment matrix. If you exclude the rows and columns for $X_0$ from the $M$ matrix, then $M/N$ is called the matrix of moments about the origin. The reason for this name may be clearer after we look at the covariance matrix.

2. $\mathbf{XP}_{AB}$. $\mathbf{XP}_{AB}$ gives the cross-product deviations from the means (which we have also referred to as SST and SP). In our current example:

<table>
<thead>
<tr>
<th>$\mathbf{XP}_{AB}$</th>
<th>Formula: $\Sigma (A - \bar{A})(B - \bar{B})$</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{XP}_{11}$</td>
<td>$\Sigma (X_1 - \bar{X}_1)(X_1 - \bar{X}<em>1)$ = SST</em>{1}</td>
<td>380.95</td>
</tr>
<tr>
<td>$\mathbf{XP}<em>{12} = \mathbf{XP}</em>{21}$</td>
<td>$\Sigma (X_1 - \bar{X}_1)(X_2 - \bar{X}<em>2)$ = SP</em>{12}</td>
<td>-49.65</td>
</tr>
<tr>
<td>$\mathbf{XP}<em>{1Y} = \mathbf{XP}</em>{Y1}$</td>
<td>$\Sigma (X_1 - \bar{X}<em>1)(Y - \bar{Y})$ = SP</em>{Y1}</td>
<td>704.29</td>
</tr>
<tr>
<td>$\mathbf{XP}_{22}$</td>
<td>$\Sigma (X_2 - \bar{X}_2)(X_2 - \bar{X}<em>2)$ = SST</em>{2}</td>
<td>566.55</td>
</tr>
<tr>
<td>$\mathbf{XP}<em>{2Y} = \mathbf{XP}</em>{Y2}$</td>
<td>$\Sigma (X_2 - \bar{X}<em>2)(Y - \bar{Y})$ = SP</em>{Y2}</td>
<td>271.91</td>
</tr>
<tr>
<td>$\mathbf{XP}_{YY}$</td>
<td>$\Sigma (Y - \bar{Y})(Y - \bar{Y})$ = SST_{Y}</td>
<td>1820.43</td>
</tr>
</tbody>
</table>
(We don't bother with $X_0$, since any cross-products involving it would equal 0). It is often convenient to present the $XP_{AB}$ values in Matrix form. We can write

\[
\begin{pmatrix}
X_1 & X_2 & Y \\
X_1 & 380.95 \\
X_2 & -49.65 & 566.55 \\
Y & 704.29 & 271.91 & 1820.43
\end{pmatrix}
\]

or,

\[
\mathbf{XP} = \begin{bmatrix}
380.95 \\
-49.65 & 566.55 \\
704.29 & 271.91 & 1820.43
\end{bmatrix}
\]

Incidentally, Hayes labels this as the SSCP matrix, for sums of squares and cross-products.

3. $s_{AB}$. $s_{AB}$ is the covariance of variables A and B. When $A = B$, $s_{AA} = \text{the sample variance of } A$. $s_{AA}$ can of course also be written as $s_A^2$. $s_{AB} = XP_{AB}/(N - 1)$. In matrix form, we can write

\[
\begin{pmatrix}
X_1 & X_2 & Y \\
X_1 & 20.05 \\
X_2 & -2.61 & 29.82 \\
Y & 37.07 & 14.31 & 95.81
\end{pmatrix}
\]

or,

\[
\mathbf{s} = \begin{bmatrix}
20.05 \\
-2.61 & 29.82 \\
37.07 & 14.31 & 95.81
\end{bmatrix}
\]

Additional comments:

a. The $s$ matrix is typically called the covariance matrix or the variance/covariance matrix. It is also sometimes called the Matrix of Moments about the mean, because the mean is subtracted from each variable.
b. The sample covariance gives us an indication of the association between two variables. If the covariance is zero, then there is no association. If the covariance is positive, that means that above-average values on one variable tend to be paired with above-average values on the other variable. The square roots of the diagonal elements of the $s$ matrix (i.e. the standard deviations) give an idea of how closely clustered cases are about the mean.

c. The values for the covariances are dependent on the metrics of the variables. For example, income here is measured in thousands of dollars; if instead, income were measured in dollars, $s_{YY}$ would be 1 million times larger, and all the other elements of the $s$ matrix would increase by a factor of 1,000.

d. Because the values of the $s$ matrix are so dependent on the metrics used, it is difficult to tell via just "eyeballing" how strong the association between variables is. For example, does the fact that $s_{Y1} = 37.07$ mean that there is a very strong link between education and income, or is it only a weak association? This difficulty in interpretation is one of the reasons that correlations between variables (discussed next) are often looked at.

4. $r_{AB}$. $r_{AB}$ is the correlation between variables A and B. As noted above, $r_{AB} = s_{AB}/s_As_B$. There are several properties about correlations worth noting:

a. $r$ can range from -1 to 1. The larger the absolute value of $r$ is, the stronger the association is between the two variables. Hence, $r$ provides a more intuitive means than $s$ for looking at association.

b. the correlation of any variable with itself is 1, e.g. $r_{AA} = s_{AA}/s_As_A = s_A^2/s_A^2 = 1$.

c. Another way of thinking of $r$ is that it is the covariance of the standardized variables. Let $A' = (A - \bar{A})/s_A$, $B' = (B - \bar{B})/s_B$ (Note that both $A'$ and $B'$ appear in one of our formulas for $r$). That is, let $A'$ and $B'$ be the z-score transformations of A and B. As we showed before, any Z score has a mean of zero and a variance of 1. Ergo,

$$r_{A'B'} = \frac{s_{A'B'}}{\sqrt{s_{A'A'}s_{B'B'}}} = s_{A'B'} = r_{AB}$$

Let us now compute the correlations of the variables. Keep in mind that $r_{11} = r_{22} = r_{YY} = 1$.

<table>
<thead>
<tr>
<th>$r_{AB}$</th>
<th>Formula: $s_{AB}/\sqrt{(s_{AA}s_{BB})}$</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{12} = r_{21}$</td>
<td>$-2.61/\sqrt{(20.05 \times 29.82)}$</td>
<td>-.1067</td>
</tr>
<tr>
<td>$r_{1Y} = r_{Y1}$</td>
<td>$37.07/\sqrt{(20.05 \times 95.81)}$</td>
<td>.8458</td>
</tr>
<tr>
<td>$r_{2Y} = r_{Y2}$</td>
<td>$14.31/\sqrt{(29.82 \times 95.81)}$</td>
<td>.2677</td>
</tr>
</tbody>
</table>
In matrix form, we can write this as

\[
\begin{bmatrix}
X_1 & X_2 & Y \\
X_1 & 1.00 \\
X_2 & -0.11 & 1.00 \\
Y & 0.85 & 0.27 & 1.00
\end{bmatrix}
\]

or,

\[
\mathbf{r} = \begin{bmatrix}
1.00 \\
-0.11 & 1.00 \\
0.85 & 0.27 & 1.00
\end{bmatrix}
\]

From the correlation matrix, it is clear that education \((X_1)\) is much more strongly correlated with income \((Y)\) than is job experience \((X_2)\).

5. **Alternative formulas.** It is very common for computer programs to report the correlations, standard deviations, and means for all the variables, along with the sample size. From this information, you can construct any of the matrices we have just talked about.

\[
s_{AB} = r_{AB} s_A s_B,
\]
\[
X_{PAB} = s_{AB} (N - 1) = r_{AB} s_A s_B (N - 1),
\]
\[
M_{AB} = X_{PAB} + N \overline{AB} = s_{AB} (N - 1) + N \overline{AB} = r_{AB} s_A s_B (N - 1) + N \overline{AB}
\]

Example. To illustrate this, let us compute \(s_{11}, s_{Y1}, X_{P12}, M_{Y2}\) using only the correlations, means, standard deviations, and sample size.

\[
s_{11} = r_{11} \cdot s_1 \cdot s_1 = 1 \cdot 4.478 \cdot 4.478 = 20.05,
\]
\[
s_{Y1} = r_{y1} \cdot s_Y \cdot s_1 = 0.8458 \cdot 9.788 \cdot 4.478 = 37.07,
\]
\[
X_{P12} = r_{12} \cdot s_1 \cdot s_2 \cdot (N - 1) = -0.1067 \cdot 4.478 \cdot 5.46 \cdot 19 = -49.57,
\]
\[
M_{Y2} = r_{y2} \cdot s_Y \cdot s_2 \cdot (N - 1) + N \overline{Y} \overline{X_2}
\]
\[
= 0.2677 \cdot 9.788 \cdot 5.46 \cdot 19 + 20 \cdot 24.415 \cdot 12.65 = 6448.82
\]
6. **Counting rules.** Let \( L = \# \text{ of } X \text{ variables} + \# \text{ of } Y \text{ variables} \) (Not counting \( X_0 \)). In this case, \( L = 3 \). Note that

A. The M matrix has \((L + 1)(L + 2)/2 = 10\) unique elements (i.e. diagonal and sub-diagonal elements).

B. The XP matrix has \((L)(L + 1)/2 = 6\) unique elements

C. The s matrix has \((L)(L + 1)/2 = 6\) unique elements

D. The correlation matrix has \((L)(L - 1) = 3\) unique (i.e. sub-diagonal) elements

E. There are L means (i.e. 3)

F. There are L standard deviations (i.e. 3)

Note further that, if you know

A. The correlations (3 pieces of information), means (3), standard deviations (3), and sample size (1 piece of information), (10 pieces of information altogether), or

B. The covariance matrix (6 elements), the means (3), and sample size (1) (10 pieces of information), or

C. The XP Matrix (6 elements), means (3), and sample size (1) (10 pieces of information altogether), or

D. The M matrix (10 elements)

THEN you have all the information you need to work any regression problem.