Projection Matrices and Regression Sums of Squares

The projection matrices **P** and **M**

The multiple regression model for the population, expressed in matrix form, is $\mathbf{y} = \mathbf{X\beta} + \mathbf{u}$

Expressed in terms of the estimates, it is $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$ (Equation 1)

The solution is $\hat{\boldsymbol{\beta}} = \left(\mathbf{X'X}\right)^{-1}\mathbf{X'y}$

Equation 1 can be written as $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$, where $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$

$\hat{\mathbf{y}}$ can be rewritten as follows:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'y} = \left(\mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}\right)\mathbf{y}$$

Define $\mathbf{P} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}$, so $\hat{\mathbf{y}} = \mathbf{Py}$

Kutner et al. call this matrix **H**, the "hat matrix",
because it "puts the hat on" **y**.

Properties of the **P** matrix

P depends only on **X**, not on **y**. In some derivations, we may need different **P** matrices that depend on different sets of variables. In this case, the matrix **P** may be written with a subscript, indicating the **X** matrix that is depends on, as $\mathbf{P_X}$. In order to reduce the complexity of the notation, I won't do this unless I need to distinguish different **P** matrices.

If **X** has *n* rows and *k* columns, **P** is *n* by *n*, a square matrix. Note that *n* is typically much larger than *k*, so **P** is a large matrix. The **P** matrix has great theoretical importance, but it is not usually computed.

Property 1 – Multiplying by **P** leaves the **X** matrix unchanged, **P*X=X**

Proof:

$$\mathbf{PX} = \left(\mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}\right)\mathbf{X} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\left(\mathbf{X'X}\right) = \mathbf{X}$$

Property 2 – **P** is idempotent, meaning **P*P=P**.

Proof:

$$\mathbf{P}*\mathbf{P} = \left(\mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}\right)\left(\mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}\right) = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\left(\mathbf{X'X}\right)\left(\mathbf{X'X}\right)^{-1}\mathbf{X'}$$

$$= \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'} = \mathbf{P}$$

Property 3 **– P** is symmetric, **P' = P**

Proof:

$$\mathbf{P}' = \left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)' = \left(\mathbf{X}'\right)'\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]'\mathbf{X}' = \mathbf{X}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]\mathbf{X}' = \mathbf{P}$$

The central equality, $\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]' = \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]$ is true because **X'X** is symmetric, and the inverse of a symmetric matrix is symmetric.

The **M** matrix, the residualizer

In the formula $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$, I have showed a matrix expression for $\hat{\mathbf{y}}$ in terms of the matrix **P**.

There is also an expression for $\hat{\mathbf{u}}$.

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{P}\mathbf{y} = \left(\mathbf{I} - \mathbf{P}\right)\mathbf{y}$$

The matrix $\left(\mathbf{I} - \mathbf{P}\right)$ is called the residualizer, because it makes the residuals. It is usually given the symbol **M**. So now we have $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}} = \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y}$. This decomposes **y** into two chunks.

Properties of the **M** matrix

Property 1 – Multiplying **M** by **X** gives a **0** matrix, **MX=0**

Proof: $\mathbf{M}\mathbf{X} = \left(\mathbf{I} - \mathbf{P}\right)\mathbf{X} = \mathbf{X} - \mathbf{P}\mathbf{X} = \mathbf{X} - \mathbf{X} = \mathbf{0}$

Property 2 – like **P**, the matrix **M** is idempotent.

Proof: $\mathbf{M} * \mathbf{M} = \left(\mathbf{I} - \mathbf{P}\right)\left(\mathbf{I} - \mathbf{P}\right) = \mathbf{I} * \mathbf{I} - \mathbf{I} * \mathbf{P} - \mathbf{P} * \mathbf{I} + \mathbf{P} * \mathbf{P}$
$= \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P} = \mathbf{I} - \mathbf{P} = \mathbf{M}$

Property 3 – like **P**, the matrix **M** is symmetric

Proof: $\mathbf{M}' = \left(\mathbf{I} - \mathbf{P}\right)' = \left(\mathbf{I}' - \mathbf{P}'\right) = \left(\mathbf{I} - \mathbf{P}\right) = \mathbf{M}$

Property 4 – **P*M=0**, where **0** represents a matrix with all 0 elements.

Proof: $\mathbf{P} * \mathbf{M} = \mathbf{P}\left(\mathbf{I} - \mathbf{P}\right) = \mathbf{P} - \mathbf{P} * \mathbf{P} = \mathbf{P} - \mathbf{P} = \mathbf{0}$

Property 4a – Property 4 means that for any vector **y**, **Py** is orthogonal to **My**.

Proof: Recall that orthogonal vectors have dot product = 0, and that the dot product can be expressed as a transpose product.

$$(\mathbf{Py})'(\mathbf{My}) = \mathbf{y}'\mathbf{P}'\mathbf{My} = \mathbf{y}'\mathbf{PMy} = \mathbf{y}'(\mathbf{0})\mathbf{y} = \mathbf{0}$$

Now $\mathbf{Py} = \hat{\mathbf{y}}$ and $\mathbf{My} = \hat{\mathbf{u}}$, so we have just proved that $\hat{\mathbf{y}} \bullet \hat{\mathbf{u}} = 0$. This is the algebraic equivalent of the regression geometry.

Sums of Squares

We can use the **P** and **M** matrices to prove facts about the regression sums of squares. We need some matrix definitions and facts.

The trace of a matrix

The trace of a matrix is the sum of its diagonal elements. In the matrix

$$\begin{pmatrix} 3 & 2 & 0 \\ 0 & -2 & 1 \\ -1 & 5 & 4 \end{pmatrix}$$, the trace is 3 + (-2) + 4 = 5.

I will use **tr** to indicate the trace of a matrix.

Trace property 1 – If **I** is an *n* by *n* identity matrix (sometimes denoted by $\mathbf{I}_n$), then $\text{tr}(\mathbf{I}_n) = n$.

Proof: $\mathbf{I}_n$ has *n* 1's along the diagonal.

Matrix fact: Note that in general, two matrices **A** and **B** may not be comformable for multiplication in both orders, i.e. **AB** and **BA** may not both exist. If they do, and if **A** is *n* by *m*, then **B** must be *m* by *n*. If $m \neq n$, the two products will have different dimensions, **AB** will be *n* by *n*, and **BA** will be *m* by *m*.

Trace property 2 – If **A** and **B** are comformable for multiplication in both orders then tr(**AB**)=tr(**BA**). This is true even if **AB** does not have the same dimensions as **BA**.

Proof: Suppose **A** is *n* by *m*, **B** is *m* by *n.*

$$\text{tr}(\mathbf{AB}) = \sum_{i=1}^{n} (\mathbf{AB})_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{m} \mathbf{A}_{ik}\mathbf{B}_{ki} \qquad \text{Formula 1}$$

$$\text{tr}(\mathbf{BA}) = \sum_{i=1}^{m} (\mathbf{BA})_{ii} = \sum_{i=1}^{m}\sum_{k=1}^{n} \mathbf{B}_{ik}\mathbf{A}_{ki} \qquad \text{Formula 2}$$

Looking at the 2 formulas, we can get formula 2 from formula 1 by a) switching the order of summation; b) switching the names of the indices *i* <--> *k*; and c) reversing the order of **A** and **B**. This last operation is permitted because the products are of individual elements of **A** and **B**, not the whole matrix, so multiplication is commutative. So Formula 1 = Formula 2, and tr(**AB**)=tr(**BA**).

Trace property 3 – If **A** and **B** are conformable for addition, tr(**A+B**) = tr(**A**) + tr(**B**) and tr(**A-B**) = tr(**A**) - tr(**B**).

Proof: These can be seen easily from the definitions of matrix addition and the trace.

Trace property 4 – Suppose that matrix **X** is *n* by *k*, and matrix **X'X** has an inverse.

Then tr($\mathbf{P_X}$) = *k*.

Proof:

$\mathrm{tr}\left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right) = \mathrm{tr}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\right) = \mathrm{tr}\left(\mathbf{I}_k\right) = k$. Here $\mathbf{I}_k$ is the *k* by *k* identity matrix, which has *k* 1's along the diagonal.

Trace property 5 – tr(**M**) = *n* – *k*.

Proof:

$\mathrm{tr}\left(\mathbf{M}\right) = \mathrm{tr}\left(\mathbf{I}_n - \mathbf{P}\right) = \mathrm{tr}\left(\mathbf{I}_n\right) - \mathrm{tr}\left(\mathbf{P}\right) = n - k$

<u>Expectation of sum of squares of residuals</u>

Now we have the tools to show that if $\mathrm{var}\left(u\right) = \sigma^2$, then the expectation of the sum of squares of the residuals is $\left(n-k\right)\sigma^2$.

First, we'll look at the sum of squares of the residuals. This can be expressed as $\sum_{i=1}^{n} \hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}$, and rewritten as

$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \left(\mathbf{My}\right)'\mathbf{My} = \mathbf{y}'\mathbf{My} = \left(\mathbf{X\beta}+\mathbf{u}\right)'\mathbf{M}\left(\mathbf{X\beta}+\mathbf{u}\right)$

Recall that **MX=0**. In the last expression, there are 4 terms when we multiply it out:

$\left(\mathbf{X\beta}+\mathbf{u}\right)'\mathbf{M}\left(\mathbf{X\beta}+\mathbf{u}\right) = \boldsymbol{\beta}'\mathbf{X}'\mathbf{MX}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{Mu} + \mathbf{u}'\mathbf{MX}\boldsymbol{\beta} + \mathbf{u}'\mathbf{Mu}$.

All except the last term have either **MX** or **X'M** or both. But **X'M=(MX)'=0**, so all terms except the last drop out. Putting it together, $\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{Mu}$.

Now, **u** is a vector of the errors. We don't know the actual values of **u**, but we're still allowed to do algebra with the vector. In particular, $\mathrm{tr}\left(\mathbf{u}'\mathbf{Mu}\right) = \mathrm{tr}\left(\mathbf{Muu}'\right)$. Since we have showed the matrix expression, $\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{Mu}$, we must have $\mathrm{tr}\left(\hat{\mathbf{u}}'\hat{\mathbf{u}}\right) = \mathrm{tr}\left(\mathbf{u}'\mathbf{Mu}\right) = \mathrm{tr}\left(\mathbf{Muu}'\right)$.

If you look carefully at this last set of equations, you may notice something strange. On the left side is $\mathrm{tr}(\hat{\mathbf{u}}'\hat{\mathbf{u}})$. But $\hat{\mathbf{u}}$ is an *n* by 1 vector, so $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ is a 1 by 1 matrix, a scalar. Its trace is simply its value. On the other side, $\mathbf{Muu}'$ is an *n* by *n* matrix multiplied by a *n* by 1 vector multiplied by a 1 by *n* vector. The whole product is *n* by *n*. But we're not saying the matrices on the two sides are equal, just their traces. Now we're ready to show that the expectation of the sum of squares of the residuals is $(n-k)\sigma^2$.

We have $\mathrm{tr}(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = \mathrm{tr}(\mathbf{Muu}')$, so $E(\mathrm{tr}(\hat{\mathbf{u}}'\hat{\mathbf{u}})) = E(\mathrm{tr}(\mathbf{Muu}'))$. Now a trace is just a fancy way of doing a sum of terms, so we can move the expectation inside the trace operation. In addition, the matrix **M** is just a function of the **X** matrix, and we are treating **X** as non-random, so we can move the expectation inside the product with **M**, giving

$$E(\mathrm{tr}(\hat{\mathbf{u}}'\hat{\mathbf{u}})) = \mathrm{tr}(\mathbf{M} * E(\mathbf{uu}'))$$

Now we'll look at $E(\mathbf{uu}')$. Recall that $\mathbf{uu}'$ is an *n* by *n* square matrix. On the diagonal will be terms like u_i^2. The off diagonal elements will be terms like $u_i u_j$. Assuming that the errors are all independent and identically distributed, with variance σ^2, $\mathrm{E}(u_i^2) = \sigma^2$ and $\mathrm{E}(u_i u_j) = 0$.

So $E(\mathbf{uu}')$ will be a matrix with σ^2 for each diagonal element and 0 for each off-diagonal element. In other words, $E(\mathbf{uu}') = \sigma^2\mathbf{I}_n$.

Finally, $\mathbf{M} * E(\mathbf{uu}') = \mathbf{M}(\sigma^2\mathbf{I}_n) = \sigma^2\mathbf{M}$, so $\mathrm{tr}(\mathbf{M} * E(\mathbf{uu}')) = \mathrm{tr}(\sigma^2\mathbf{M}) = \sigma^2\,\mathrm{tr}(\mathbf{M}) = (n-k)\sigma^2$.

If you go back over the steps, you'll see that we've showed that E(SSR) = $(n-k)\sigma^2$, where SSR is the sum of squares of the residuals. That means that SSR/($n - k$) is an unbiased estimator of σ^2.

<u>Expectation of model sum of squares</u>

Consider the regression model $\mathbf{y} = \mathbf{X\beta} + \mathbf{u}$. Now assume that the true $\boldsymbol{\beta}$ is $\mathbf{0}$ (a vector where all components are 0). Then all that's left of the regression model is $\mathbf{y} = \mathbf{u}$. In this case, what's the expectation of the model (regression) sum of squares?

First, estimate the model and get $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$. Note that although $\boldsymbol{\beta} = \mathbf{0}$, in general $\hat{\boldsymbol{\beta}} \neq \mathbf{0}$.

Second, calculate the model sum of squares:

Model sum of squares = $\sum_{i=1}^{n} \hat{y}_i^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}}$.

Next, plug in $\hat{\mathbf{y}} = \mathbf{Py}$. Model sum of squares = $\mathbf{y'P'Py} = \mathbf{y'Py} = \mathbf{u'Pu}$.

Note that **u'u** is the sum of squares of the errors, which has expectation $n\sigma^2$.

Also, $\mathbf{u'u} = \mathbf{u'Iu} = \mathbf{u'}\left(\mathbf{P}+\mathbf{M}\right)\mathbf{u} = \mathbf{u'Pu} + \mathbf{u'Mu}$.

So $\mathbf{u'Pu} = \mathbf{u'u} - \mathbf{u'Mu}$, and $\mathrm{E}\left(\mathbf{u'Pu}\right) = \mathrm{E}\left(\mathbf{u'u}\right) - \left(\mathbf{u'Mu}\right) = n\sigma^2 - \left(n-k\right)\sigma^2 = k\sigma^2$

In words, the expectation of the model sum of squares is the number of regressors times the variance of the errors, if the complete null hypothesis, $\boldsymbol{\beta}=\mathbf{0}$, is true.

Sums of squares when β_0 is not 0

The derivation given above assumes that all the β's are 0, including β_0. More realistically, we are interested in the situation where β_0 may be non-zero, and want to test all the other β's. Note that the derivation of the expectation of the residual sum of squares did not assume anything about the value of the β's, so the expectation of the residual sum of squares is still $\left(n-k\right)\sigma^2$, where *k* is the total number of regressors, including the constant. With some additional work (not shown here), we can show that the effect of β_0 can be eliminated if we make the model sum of squares be the sum of deviations of the predicted values from the overall mean, rather than simply the sum of squares of predicted values. In this case, the expectation of the model sum of squares is $\left(k-1\right)\sigma^2$. Notice that $k-1$ is simply the number of regressors, excluding the constant.