The Geometry of Least Squares & Multiple Linear Model

We have already seen that the least squares fit of a simple linear model can be viewed as the projection of the n-dimensional vector $\mathbf{y}' = (y_1, \ldots, y_n)$ onto the space spanned by the vector $\mathbf{1}' = (1, \ldots, 1)$ and $\mathbf{x}' = (x_1, \ldots, x_n)$.

Now, we consider the case where we have three variables on each subject, say (u_i, v_i, y_i), $i = 1, \ldots, n$, and we wish to find the best linear fit $a + bu + cv$ to y:

$$\sum_{i=1}^{n} (y_i - a - bu_i - cv_i)^2$$

If we consider our datat as three n-dimensional vectors $\mathbf{y}$, $\mathbf{u}$, and $\mathbf{v}$, then the minimization of the above quantity is equivalent to the following minimization:

$$|| \qquad\qquad\qquad\qquad ||^2$$

So we are projecting $\mathbf{y}$ into the space spanned by $\mathbf{1}$, $\mathbf{u}$, and $\mathbf{v}$.

What does this projection look like? What are the coefficients for **u** and **v**? To answer these questions, fill in the following derivation:

1. NOTATION:

$$\begin{aligned} P_1\mathbf{y} &= \bar{y}\mathbf{1} \\ &= b_{y(1)}\mathbf{1} \end{aligned}$$

$$\begin{aligned} P_{1,x}\mathbf{y} &= P_1\mathbf{y} + P_{x\cdot 1}\mathbf{y} \\ &= b_{y(1)}\mathbf{1} + b_{y(x\cdot 1)}(\mathbf{x} - \bar{x}\mathbf{1}) \end{aligned}$$

   That is, P stands for projection. The subscript tells us what space we are projecting on to. The subscript $x \cdot 1$ stands for that part of **x** which is orthogonal to **1**. The b represent the coefficients. For example, $b_{y(1)}$ is the cefficient for **1** that we get when we project **y** on to **1**, and $b_{y(x\cdot 1}$ is the coeffcient for $(\mathbf{x} - \bar{x}\mathbf{1})$.

2. Reexpress $\hat{a}$ and $\hat{b}$ from the simple linear regression in terms of these new coefficients.

3. Next consider the problem of projecting **y** on to the space spanned by **1**, **u**, and **v**. This space is equivalent to the space spanned by the orthogonal vectors:

   $\mathbf{1}$, $(\mathbf{u} - \bar{u}\mathbf{1})$, and ___________

   or equivalently the space spanned by

   $\mathbf{1}$, $(\mathbf{v} - \bar{v}\mathbf{1})$, and ___________

4. In terms of our projection notation,

$$\begin{aligned} P_{1,u,v}\mathbf{y} &= P_1\mathbf{y} + \_\_\_\_\_\_\_\_\_\_ + \_\_\_\_\_\_\_\_\_\_ \\ &= P_1\mathbf{y} + \_\_\_\_\_\_\_\_\_\_ + \_\_\_\_\_\_\_\_\_\_ \end{aligned}$$

5. Convert the above two equalities into the b notation to see what the coefficients would be:

$$\begin{aligned}\hat{\mathbf{y}} &= \bar{y}\mathbf{1} + b_{y(u\cdot 1)}(\mathbf{u} - \bar{u}\mathbf{1}) + b_{y(v\cdot 1,u)}(\_\_\_\_\_\_\_\_\_\_\_\_) \\ &= \bar{y}\mathbf{1} + b_{y(v\cdot 1)}(\_\_\_\_\_\_\_\_\_\_\_\_) \\ &\quad + \_\_\_\_\_\_\_\_\_\_(\_\_\_\_\_\_\_\_\_\_\_\_)\end{aligned}$$

6. Since the coefficents for **u** must be the same in these two equations (and likewise for **v**), then we can choose the expression from the simpler equation. This yields:

$$\hat{\mathbf{y}} = constant\mathbf{1} + b_{y(u\cdot 1,v)}\mathbf{u} + \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

We have found that the coefficient for **u** is that from the projection of **y** onto the part of **u** which is orthogonal to **v** and **1**. This means that when we consider the size of the coefficient and whether it is significantly different from 0, we must be keep in mind that it is the size when **v** and **1** are already in the equation and already being used to expalin the variability in **y**.

Consider the concrete example, where **y** is baby's birthweight, **u** is mother's height, and **v** is mother's weight. We find,

$\hat{y} =$ 35 + 1.2 **u** + 0.07 **v**
$\hat{y} =$ 27 + 1.4 **u**

For every increase in height of 1 inch, the average weight for the baby increases by 1.4 ounces. But, if we also know the mother's weight, then for those mother's of roughly the same weight, for every increase in height of 1 inch, the average weight for the baby increases by 1.2 ounces.

Why are these two coefficients different?

What about the residual sum of squares?

Fit	Residual sum of squares/n
birth weight on constant	337 oz^2
birth weight on constant, height	324 oz^2 = 337 - 13
birth weight on constant, weight	329 oz^2 = 337 - 8
birth weight on constant, weight,height	322 oz^2 = 337 - 15

Notice that the RSS for the two variable model is not $337 - 13 - 8$ – Why is this the case?

Fill in the table below for the two variable fit using smoking and height to explain birth weight.

Fit	Residual sum of squares/n
birth weight on constant	337 oz^2
birth weight on constant, height	324 oz^2 = 337 - 13
birth weight on constant, smoking	
birth weight on constant, height, smoking	

This time we do have that the RSS for the two variable model is $337 - 13 - 20$. Why is that the case?

What does it mean to fit birth weight to smoking anyway???