# Basics of Regression from Geometry

*Georges Monette*

*2019-03-28*

Some basic theorems

## The Projection Theorem

THis says that if we express $Y$ as a sum of two orthogonal components, then we have expressed it as $\hat{Y} + e$.

**Theorem:** *Projection theorem:* Let $X$ be a matrix of full column rank and let

$$Y = Xb + e$$

where $e$ is orthogonal to $\mathcal{L}(X)$.

Then $b$ is the vector of estimated coefficients for the least-squares regression of $Y$ on $X$, $e$ is the residual, $||e||^2 = e'e$\$ is the residual of the regression and the sum of squares for error, $SSE$, is equal to $e'e$.

*Proof:*

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'Xb + (X'X)^{-1}X'e \\ &= b + (X'X)^{-1}0 \quad \text{since } e \perp \mathcal{L}(X) \\ &= b \end{aligned}$$

and thus $e = Y - X\hat{\beta}$. *QED*

Our next theorem is the "Added Variable Plot Theorem" more formally known as the Frisch-Waugh-Lovell Theorem. It took some decades to prove it but here's an easy proof.

The AVP for the regression of $Y$ on $X_1$ controlling for $X_2$ is the regression of the residual of $Y$ regressed on $X_2, on the residuals of the regression of $X_1$ regressed on $X_2$.

**Theorem:** Consider the regression of $Y$ on two blocks of predictors given by full column rank matrices $X_1$ and $X_2$, where, moreover, the partitioned matrix, $[X_1 X_2]$ is of full rank. Suppose

$$Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + e$$

Then the residual of $Y$ regressed on $X_2$ regressed on the residual of $X_1$ on $X_2$ has regression coefficient $\hat{\beta}_1$ and $SSE = e'e$.

*Proof:* The residual of $Y$ in the regression on $X_2$ is obtained by pre-multiplying $Y$ by

$$Q_2 = I - P_2 = I - X_2(X_2'X_2)^{-1}X_2'$$

and similarly for $X_1$. We obtain

$$Q_2Y = Q_2X_1\hat{\beta}_1 + Q_2X_2\hat{\beta}_2 + Q_2e$$

Now, $Q_2X_2 = 0$, so that $Q_2X_2\hat{\beta}_2 = 0$. Also, since $e \perp X_2$ it follows that $Q_2e = e$. Thus

$$Q_2Y = Q_2X_1\hat{\beta}_1 + 0 + e$$

Moreover, $e'Q_2X_1\hat{\beta}_2 = e'X_1\hat{\beta}_2 = 0'\hat{\beta}_2 = 0$ so that, by the Projection Theorem, $\hat{\beta}_1$ is the regression coefficient of $Q_2Y$ on $Q_2X_1$ and has $SSE = e'e$. *QED*

Finally we show that, the partial coefficient for the regression of $Y$ on $X_1$ adjusting for $X_2$ is the same as the partial coefficient for the regression of $Y$ on $X_1$ adjusting for the predictor of $X_1$ based on $X_2$, i.e. $P_2X_1 = X_2(X_2'X_2)^{-1}X_2'X_1$. However, the $SSE$ of this regresion is larger than that of the full multiple regression which is equal to that of the AVP regression.

**Theorem:** Consider the regression of $Y$ on two blocks of predictors given by full column rank matrices $X_1$ and $X_2$, where, moreover, the partitioned matrix, $[X_1X_2]$ is of full rank. Suppose

$$Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + e$$

Consider, also, the regression of $Y$ on $X_1$ and the predicted value of $X_1$ based on $X_2$.

Then the regression coefficients on $X_1$ are the same for both regressions. The $SSE$ for the second regression is at least as large as that of the first regression and is equal to

$$ZZZ + e'e$$

.

*Proof:* Observe that the predicted value of $X_1$ on $X_2$ is $P_2X_1$ where $P_2 = X_2(X_2'X_2)^{-1}X_2'$ so the resduals from the regression on $P_2X_1$ are obtained by premultiplying by

$$\begin{aligned} Q &= I - P_2X_1(X_1'P_2'P_2X_1)^{-1}X_1'P_2 \\ &= I - P_2X_1(X_1'P_2X_1)^{-1}X_1'P_2 \end{aligned}$$

added variable plot for the second regression is obtained by premultiplying by ....