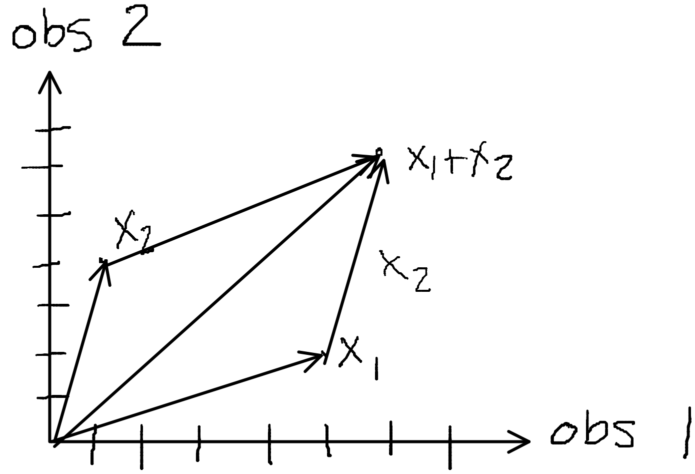
Graphical operations with vectors

Wednesday, February 08, 2012
5:43 PM

Plotting a vector in space

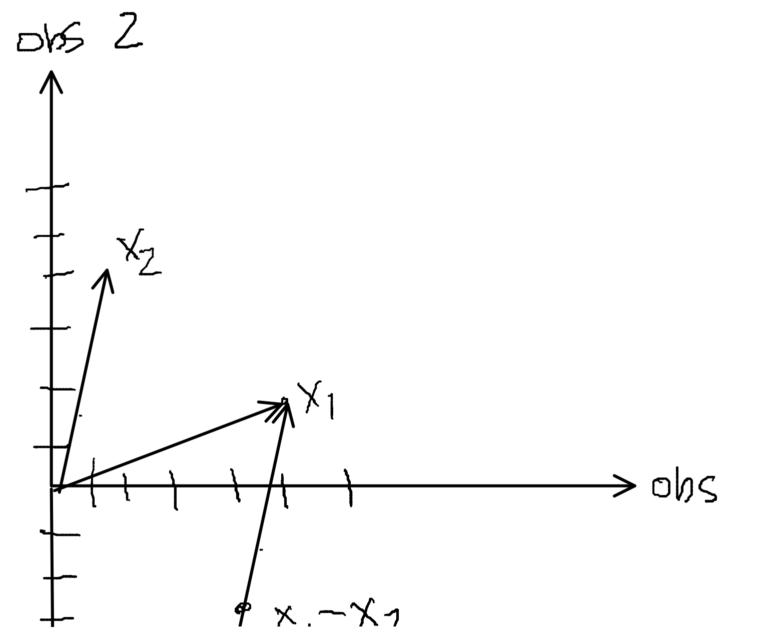
$$x_1 = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \quad x_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$



$$x_1 + x_2 = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

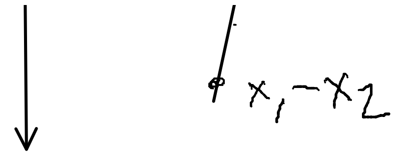
Adding two vectors : putting two vectors head to tail in a graphical sense is equivalent to adding vectors in an arithmetic sense.

$$x_1 - x_2 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$$



Subtracting two vectors

Subtracting two vectors



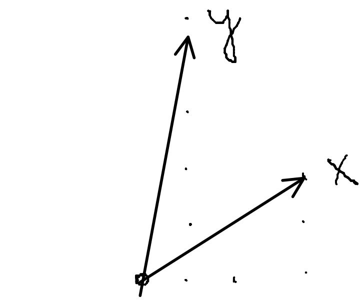
Vector Spaces

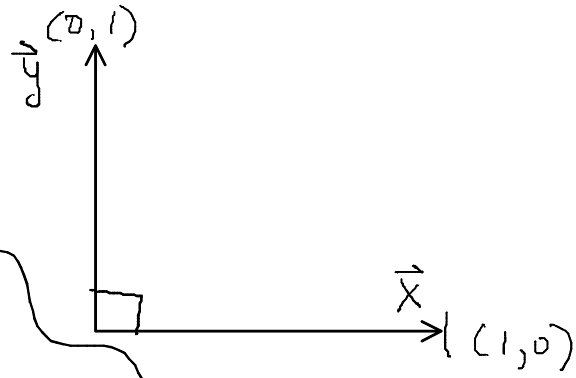
Wednesday, February 08, 2012
5:49 PM

Vectors form a subspace in $\mathbb{R}^k$

$$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$x = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

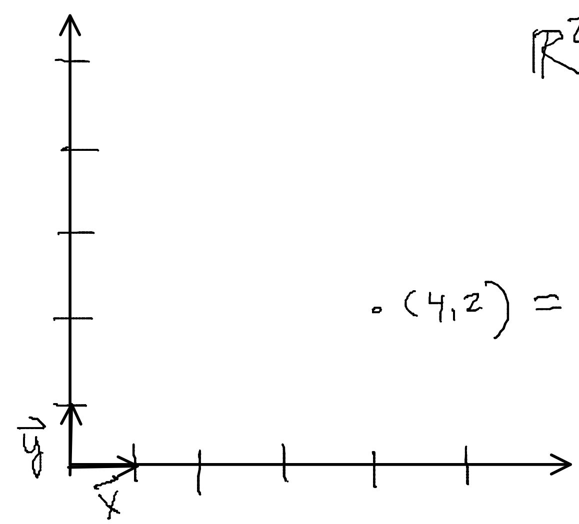




Any point in the space of k linearly independent column vectors $X_{n \times k}$ can be expressed as a combination of the column vectors in X.

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$\mathbb{R}^2$

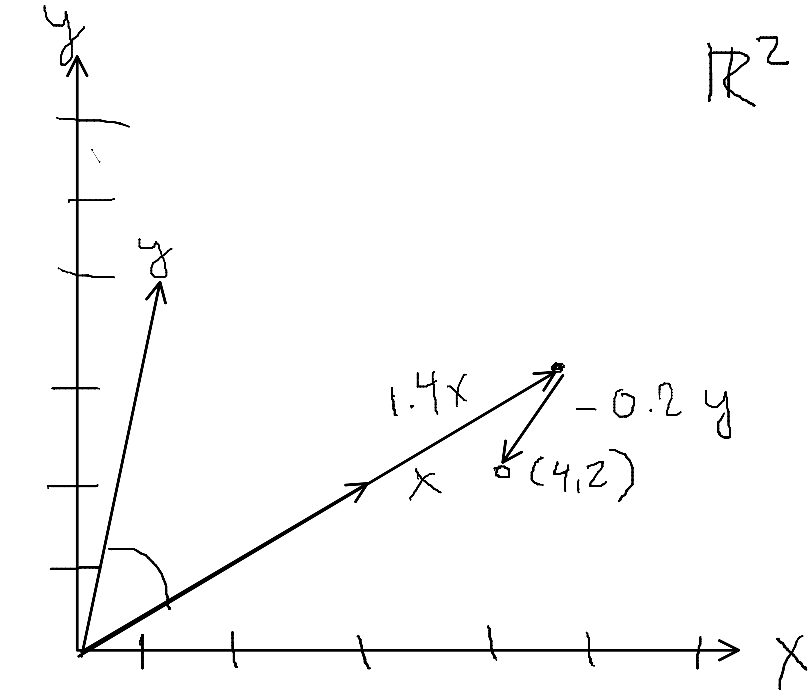


$(4, 2) = 4\vec{x} + 2\vec{y}$

y

$\mathbb{R}^2$

$\mathbb{R}^2$



$4\vec{x} + 2\vec{y}$

$$x = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$$ax + by = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \Rightarrow$$

$$\begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

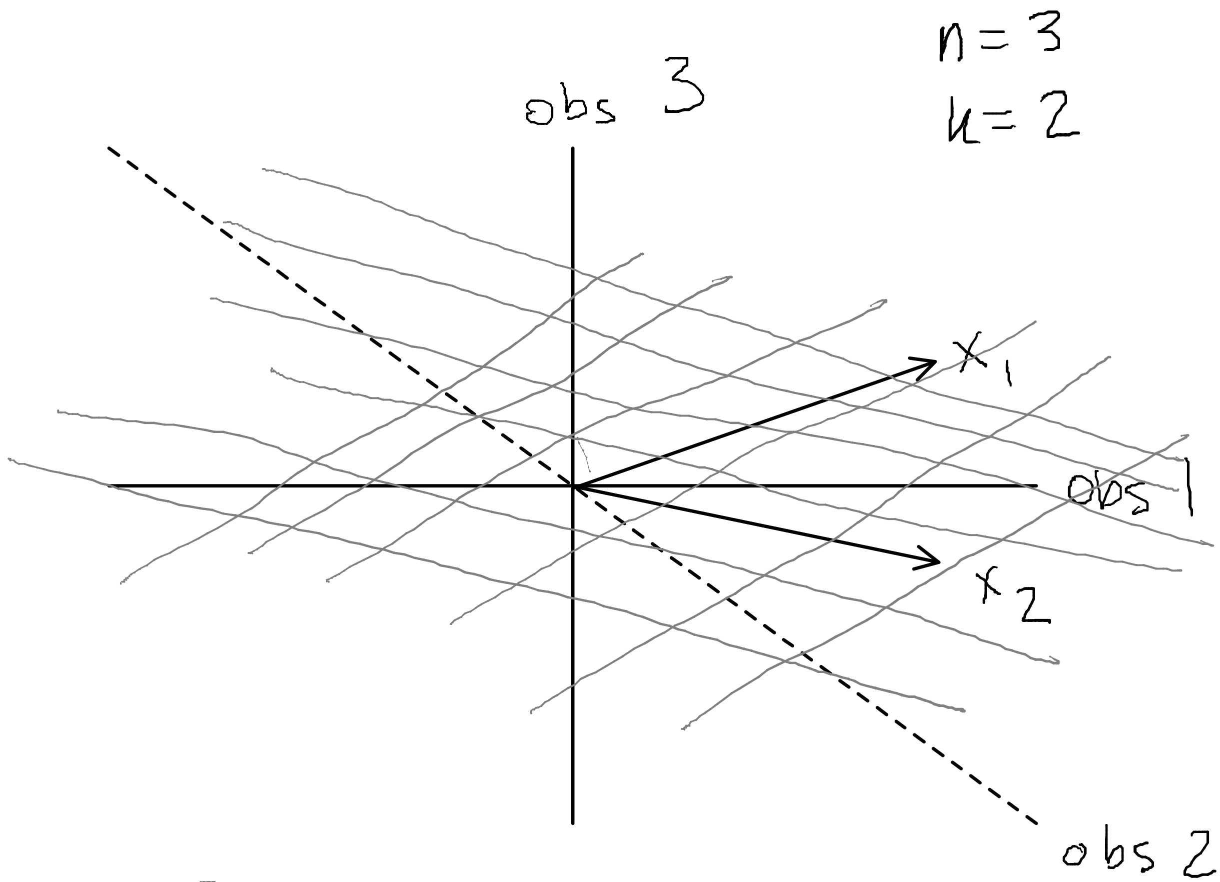
$$\begin{bmatrix} a \\ \end{bmatrix} = \begin{bmatrix} 1.4 \\ \end{bmatrix}$$

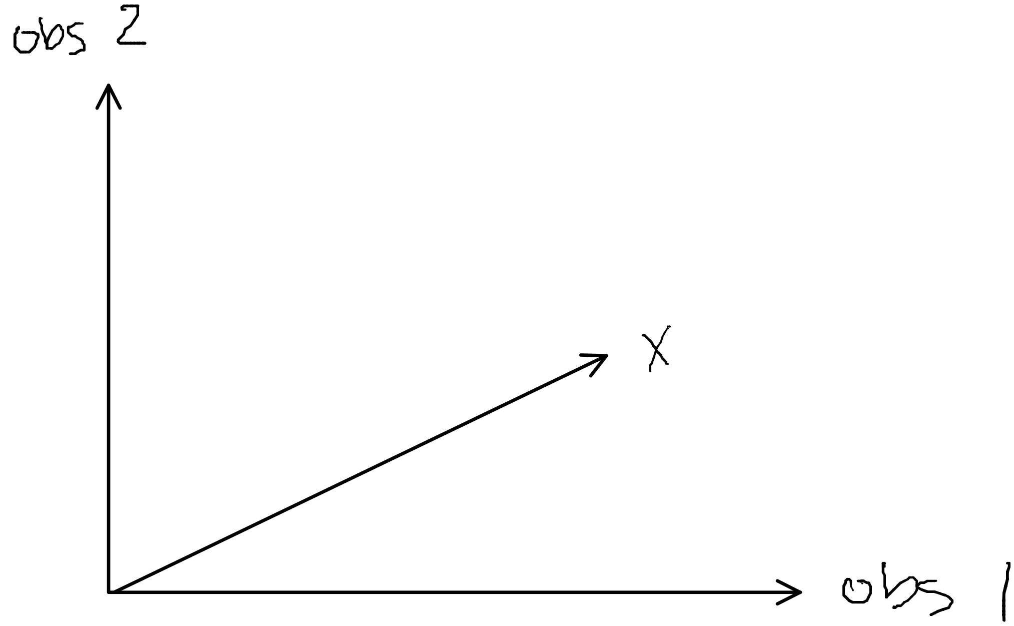
$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1.4 \\ -0.2 \end{bmatrix}$$

The graphical connection to OLS Regression

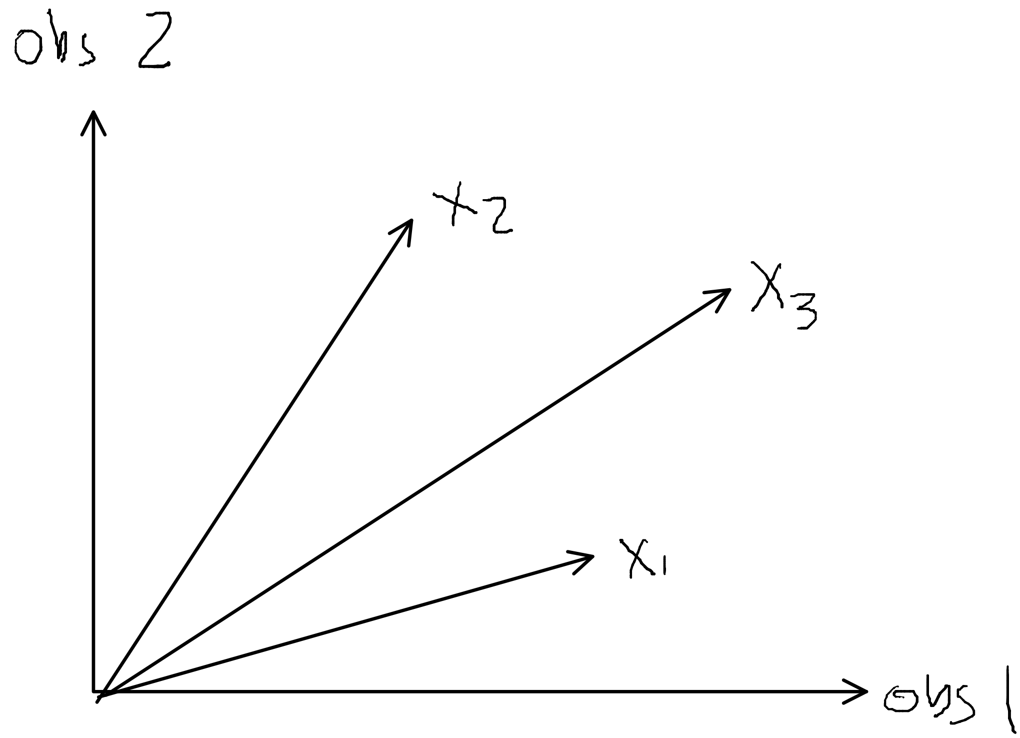
Wednesday, February 08, 2012
5:54 PM

In a linear regression, the columns of $X_{n \times k}$ form a space in $\mathbb{R}^k$. This space is a subspace of the n-dimensional space defined by observations.





Now, it should be clear why we need $n > k$ in order to run a regression!



Projection

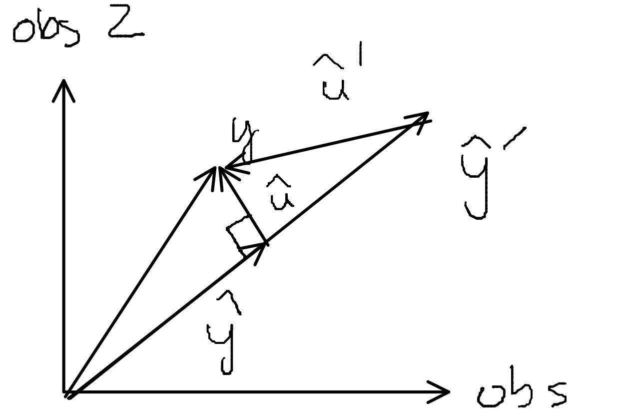
Wednesday, February 08, 2012
5:57 PM

We know that "something" is done to the y variable using the X matrix to produce predictions $\hat{y}$...

$$\hat{y} = X\hat{\beta} \qquad \hat{\beta} = (X'X)^{-1}X'y \qquad \hat{y} = X(X'X)^{-1}X'y$$

...and we know that $\hat{y} + \hat{u} = y$, so that by vector addition these two must form some kind of triangle in vector space.

$$\hat{y} + \hat{u} = y$$

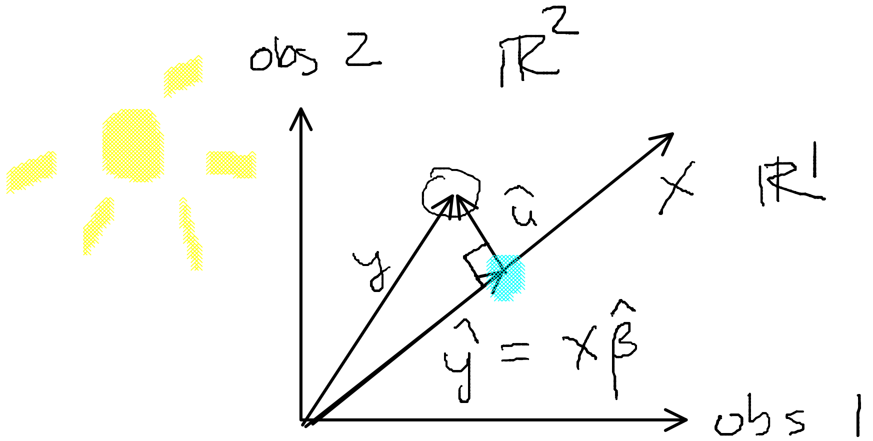


Furthermore, we know that $\hat{y}$ and $\hat{u}$ must be at right angles to each other... why?

if $\hat{y}$ and $\hat{u}$ are not perpendicular then $\hat{u}$'s length is not minimized.

But we know OLS determines $\hat{\beta}$ / $\hat{y}$ that results in minimized length of $\hat{u}$.

Linear regression represents a "projection" of the y-vector onto the space defined by $X_{n\times k}$. Let's see this graphically.



$$\hat{y} = \underbrace{\boxed{\underset{n\times k}{X}(\underset{k\times n}{X'}\underset{n\times k}{X})^{-1}\underset{k\times n}{X'}}}_{X_{n\times k}} y \longrightarrow P_X : \text{projection matrix (for } X).$$

$$\hat{y} = P_X y$$

We can also decompose the operations of regression mathematically to get a sense of what is going on here.

This final result is called the "projection matrix" P_X.

Orthogonality

Wednesday, February 08, 2012
6:04 PM

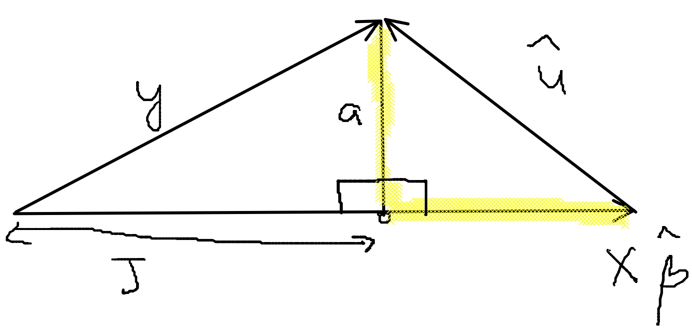
$X\hat{\beta}$ and $\hat{u}$ are orthogonal. By "orthogonal" we mean…

1. $\angle X\hat{\beta}\hat{u} = 90^{\circ}$
2. $\hat{u}$ cannot be represented by vectors in X
3. $X(X'X)^{-1}X'\hat{u} = 0$ (there is no projection of $\hat{u}$ onto X)
4. A regression of X on $\hat{u} \rightarrow \beta_{\mathrm{u}} = 0$.

$$P_X \hat{u} = 0 .$$

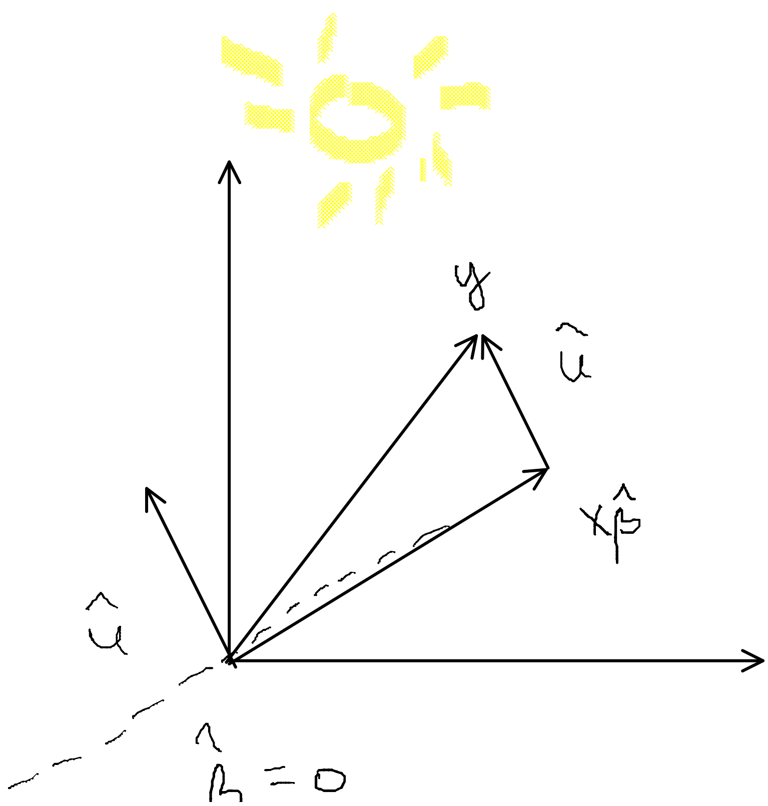
These are all properties of the ESTIMATES, not of the data generating process or the real world. It is not necessarily true that $X\beta$ and u are orthogonal! ANY regression of $\hat{y}$ on $\hat{u}$ will yield these results, no matter what the data generating process is.

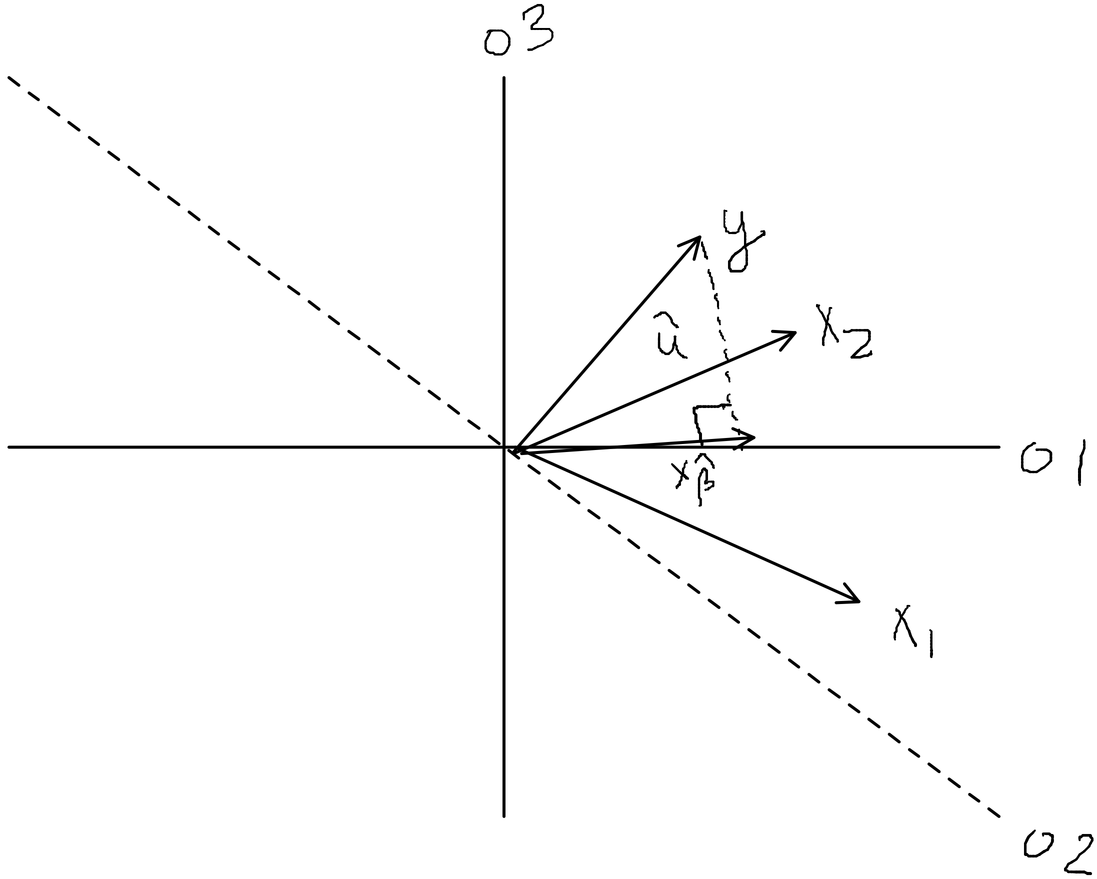
All of the points we've just made hold in > 2 dimensional space, of course (which is good--most data sets have more than two observations!). They're just hard to draw.



$$\hat{u}^2 = a^2 + (X\hat{\beta} - J)^2$$

$$\hat{u} = \sqrt{a^2 + (X\hat{\beta} - J)^2}$$




o3
y
$\hat{u}$
X_2
o1
$X\hat{\beta}$
X_1
o2

Variance decomposition and the sum of squares

Wednesday, February 08, 2012
6:13 PM

Recall: $X(X'X)^{-1}X'$ = projection matrix = P_X projects any $n \times 1$ vector onto the space defined by X.
There is an equivalent "residual matrix" M_X.

$$M_X = I_{n\times n} - P_{n\times n}$$

$$M_X y = \hat{u} \qquad M_X y + P_X y = y$$

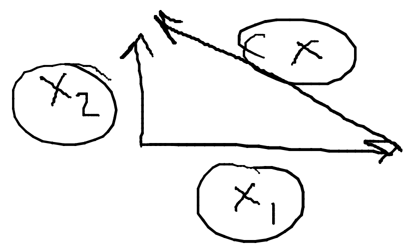
$$\hat{u} + \hat{y} = y$$

The Pythagorean theorem says that $\|X\beta\|^2 + \|\hat{u}\|^2 = \|y\|^2$.

- Side note: what is ||·||?

$\|\cdot\|$ = "euclidean norm" : length of a vector.

$$\|X\| = \sqrt{\sum_{i=1}^{n} x^2}$$



This fact gives us some information about the degree to which X can explain the variance of y.

$$\|y\|^2 = \|X\hat{\beta}\|^2 + \|\hat{u}\|^2$$

$$\underbrace{y'y}_{TSS} = \underbrace{(X\hat{\beta})'(X\hat{\beta})} + \underbrace{\hat{u}'\hat{u}}_{RSS}$$

TSS =

if y has $\mu_y = 0$, then

$$var(y) = \sum_{i=1}^{n} (y_i - \mu_y)^2 = \sum_{i=1}^{n} y_i^2 = y'y$$

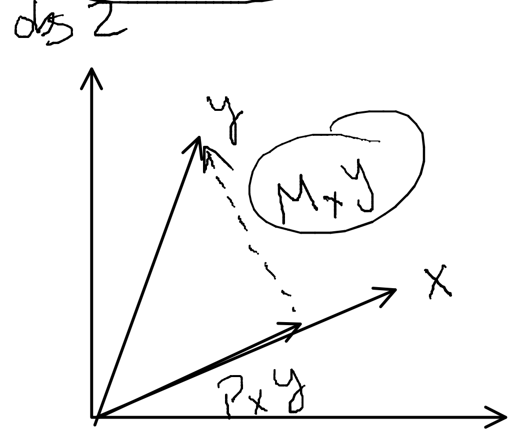
$$R^2 = \frac{ESS}{TSS}$$: Proportion of variation in y that is explained by X.

$$R^2 \in [0, 1]$$

Facts about projection matrices

Wednesday, February 08, 2012
6:19 PM

P_X is idempotent: $P_X P_X = P_X$. Why is this true intuitively? (Prove for homework!)



portion of y not representible in X

M_X is idempotent: $M_X M_X = M_X$. Why is this true intuitively? (Prove for homework!)

$M_X P_X = 0 = P_X M_X$. Why is this true intuitively? (Prove for homework!)

$P_X y$: portion of y representible in X.

$0 = M_X P_X y$: what portion of the above is NOT representible in X?

None of it!

$M_X = M_X'$ and $P_X = P_X'$. Let's prove this one formally to give you an example for your homework.

$$P_X = P_X'$$

$$X(X'X)^{-1}X' = [\underbrace{X(X'X)^{-1}}_{A}\ \underbrace{X'}_{B}]'$$

$$= X(\underbrace{X}_{A}\ \underbrace{(X'X)^{-1}}_{B})'$$

$$= X(((X'X)^{-1})'X')$$

$$= X([(X'X)']^{-1}X') = X(X'X)^{-1}X'$$

$(AB)' = B'A'$

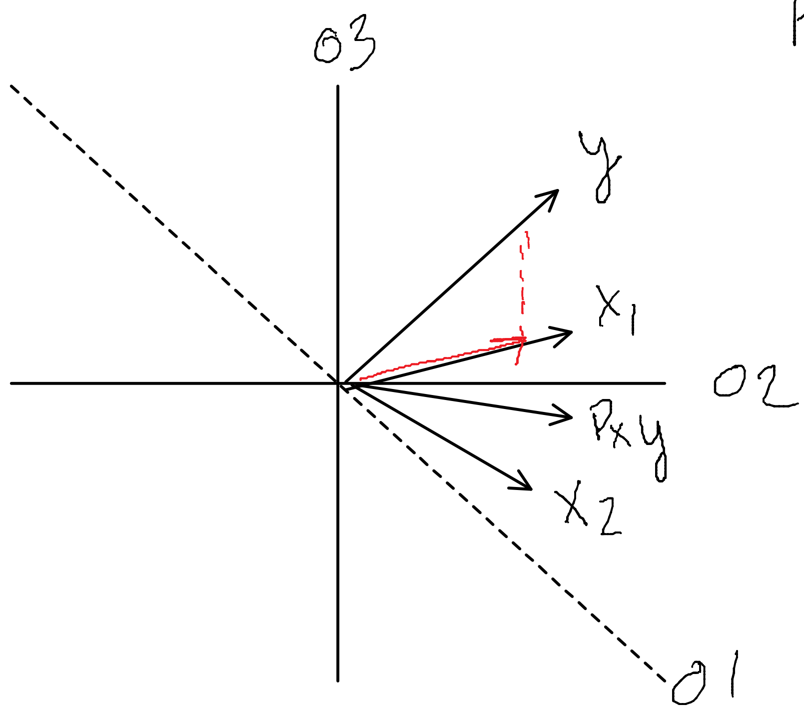
$(A^{-1})' = (A')^{-1}$

P_1: projection matrix for X_1

Consider $X = [X_1 \;\; X_2]$. We can show that $P_1 P_X = P_X P_1 = P_1$ and $M_1 M_X = M_X M_1 = M_X$.

$$X = [\underbrace{x_1 \;\; x_2}_{X_1} \;\; \underbrace{x_3 \;\; \cdots \;\; x_k}_{X_2}] = [\underset{n \times k_1}{X_1} \quad \underset{n \times k_2}{X_2}]$$

$P_X y$



$$P_X P_1 = \boxed{P_1 P_X} = P_1$$

$$P_X X_1 = X_1$$

$$P_X X_1 (X_1' X_1)^{-1} X_1'$$

$$X_1 (X_1' X_1)^{-1} X_1' = P_1$$

$$X_1 (X_1' X_1)^{-1} X_1' P_X$$

$$X_1' P_X = (P_X' X_1)'$$

$$(AB)' = B'A'$$

$$= (P_X X_1)'$$

$$(X_1)' = X_1'$$

$$X_1 (X_1' X_1)^{-1} X_1' = P_1$$

$$\blacksquare$$

Frisch-Waugh-Lovell Theorem

Wednesday, February 08, 2012
6:27 PM

Consider two regressions:

1) $y = X\beta + u$
2) $M_1 y = M_1 X_2 \beta_2 + residuals$

where $X = [X_1\ X_2]$ (a partitioned matrix of variables) and $\beta = \begin{vmatrix} \beta_1 \\ \beta_2 \end{vmatrix}$.

The Frisch-Waugh-Lovell (FWL) theorem says that the estimates of β_2 are the same for regressions (1) and (2).

Proof:

Run a regression on $M_1 y$ using $M_1 X_2$, reg (2).

$\underbrace{(X'X)^{-1}X'y}_{\hat{\beta}}$

residuals of a regression: lm$(X_2 \sim X_1)$.

$$\hat{\beta}_2 = \left[(M_1 X_2)' M_1 X_2\right]^{-1} (M_1 X_2)' M_1 y$$

$$= \left[X_2' M_1' M_1 X_2\right]^{-1} X_2' M_1' M_1 y$$

$$= \left[X_2' M_1 M_1 X_2\right]^{-1} X_2' M_1 M_1 y$$

$$\hat{\beta}_2 = \left[X_2' M_1 X_2\right]^{-1} X_2' M_1 y$$

lm$(y \sim X)$. $\hat{y} + \hat{u}$

$$y = P_X y + M_X y$$

$$= X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + M_X y$$

pre multiplied by $X_2' M_1$.

$$X_2' M_1 y = \cancel{X_2' M_1 X_1 \hat{\beta}_1} + X_2' M_1 X_2 \hat{\beta}_2 + X_2' \underline{M_1 M_X} y$$

$$X_2' M_1 y = X_2' M_1 X_2 \hat{\beta}_2 + \underbrace{\cancel{X_2' M_X y}}_{(M_X X_2)' y \, = 0}$$

$$X_2' M_1 y = \underline{X_2' M_1 X_2} \hat{\beta}_2$$

$$\underline{(X_2' M_1 X_2)^{-1} X_2' M_1 y} = \hat{\beta}_2$$

Uses of the FWL Theorem

Wednesday, February 08, 2012
6:31 PM

1) Scatterplots can be created that correct for spurious correlation. This is a great way to give a visually compelling but statistically honest reporting of the relationship between two variables. (Show example in R.)

2) Standardizing variables (e.g., by centering them around their mean) does not change our estimates of $\hat{\beta}$.

$$\mu_y = \text{lm}(y \sim 1) \qquad \iota = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix}$$

$$(i'i)^{-1} i'y = \mu_y \qquad y - \mu_y$$

$$M_i y = M_i X \hat{\beta} \qquad M_i y$$

$$y = X\hat{\beta}$$

3) Influential observations analysis.