Math 455 Class 6 February 3

- If there is a problem (or part) listed in the Gradescope outline of a homework, something should be "marked".
- Please write legibly! Or use LaTeX?!

---

There is a problem in the homework regarding the $t$- and $F$-distributions.

If $Y$ has the $t$-dist. with $\nu$ d.o.f. then $Y^2$ has the $F$-dist with 1 numerator dof and $\nu$ denominator d.o.f.

This follows from the definitions of the $t$- and $F$-dist. and Theorem 7.2.

This comes up in regression: with 1 explanatory variable, the $t$- and $F$-statistics are not separate statistics: one is determined by the other.

See Faraway p. 37.

| | Height | Blood Pressure | Cholesterol |
|---|---|---|---|
| Alice | | | |
| Bob | | | |
| Carol | | | |

In the "subject space," the coordinates are the experimental subjects and in the "variable space" the coordinates are the measured quantities.

Going back to our 2 vectors in $\mathbb{R}^3$, what linear algebra problem are we trying to solve?

$$y = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$$

$X$.

As before we can't solve this in general.

But we can give an approximate solution

First multiply by $X^T$.

$$X^T Y = X^T X \beta$$

Recall $X = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}$

If $X^T X$ is invertible, $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$

then $(X^T X)^{-1} X^T Y = \beta$.

and our approximate solution is

$$\hat{Y} = X\beta = X (X^T X)^{-1} X^T Y.$$

This is the matrix form of the least-squares solution: the vector of "fitted values"

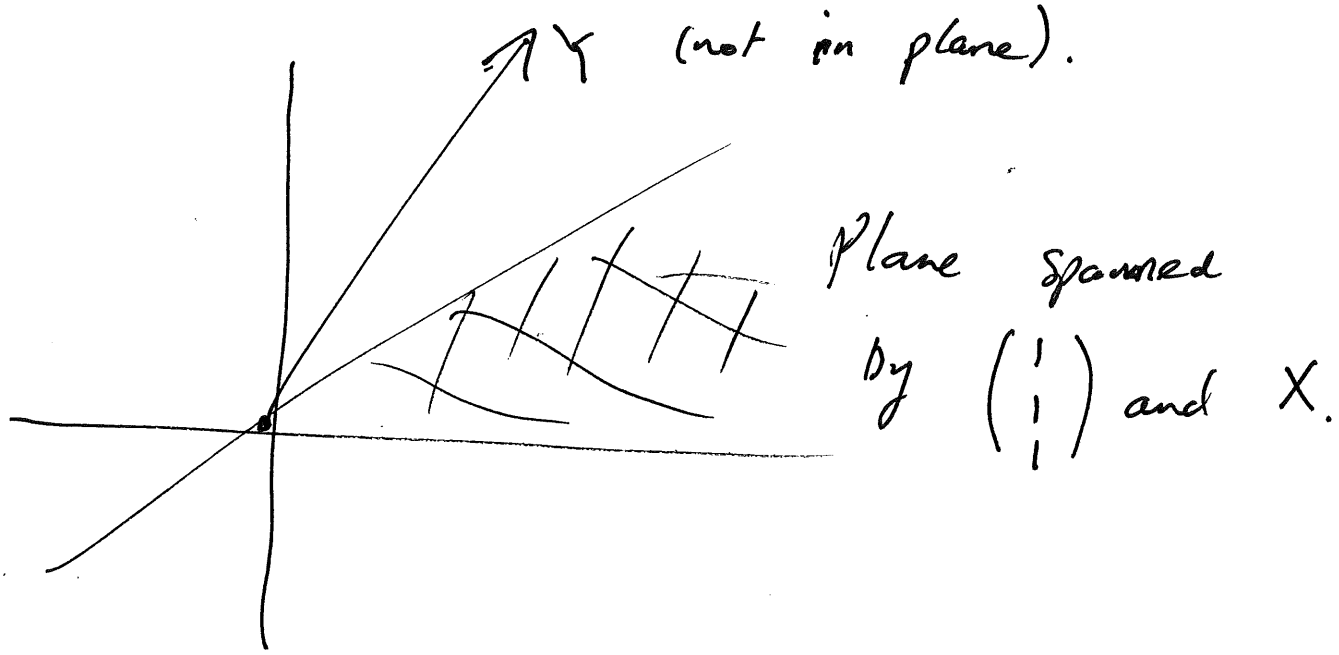
$$\hat{Y} = \hat{\beta}_0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \hat{\beta}_1 \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \quad \text{is} \quad X (X^T X)^{-1} X^T Y.$$

hat matrix

This is exactly the orthogonal projection

There is no exact solution in general, because $\begin{pmatrix}1\\1\\1\end{pmatrix}$ and $\begin{pmatrix}0\\2\\1\end{pmatrix}$ are only 2 vectors they do not "span" $\mathbb{R}^3$.

What does this look like?




We want to give the best approximation to Y: a vector $\hat{\beta}_0 \begin{pmatrix}1\\1\\1\end{pmatrix} + \hat{\beta}_1 \begin{pmatrix}0\\2\\1\end{pmatrix}$

The solution is: orthogonal projection of Y onto the plane.

This solution $\hat{\beta}_0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \hat{\beta}_1 X$ is

exactly the same as the least-squares solution!

This solution is often written $\hat{Y}$ and there is a matrix formulation.

We are trying to solve the matrix equation $Y = X\beta$

where $X$ is now the matrix $\begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}$

and $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$

This is exactly the same as the equation.

$$Y = \beta_0 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$$

of $Y$ onto the space spanned by the columns of $X$.

Terminology: $X(X^TX)^{-1}X^T$ is called the "hat matrix" because applying it to $Y$ gives $\hat{Y}$. (It "puts a hat" on $Y$.)

This formula $X(X^TX)^{-1}X^T$ can't be made simpler. This is the formula for the orthogonal projection matrix onto the column space of $X$.

Q: We learned in linear algebra that $(AB)^{-1} = B^{-1}A^{-1}$ and $(C^T)^{-1} = (C^{-1})^T$.

So why can't we simplify:

$$X(X^TX)^{-1}X^T = X\,X^{-1}(X^T)^{-1}X^T = I\,?$$

A: By definition invertible matrices have to be square. Invertible means that there is a matrix $B$ such that $AB = BA = I$.

this is only possible if $A$ is square and our $X$ is NOT square.

Generally we will be in the setting where

$X$ has lots of rows (1000 or more!)

and few columns (5-10?)

There is a "high-dimensional" setting where we have lots of variables and relatively few rows (e.g. genetic data)

So in where I wrote "$X^{-1}$" this was already a mistake.

Note that if we have $p$ variables then $X^T X$ is $(p+1) \times (p+1)$

"+1" is for the constant term.

So for our $X = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$$

$p+1 = 1+1 = 2$ since our

data set only had 1 variable.

In the regression output there is a computed $R^2$ = "percentage of variance explained".

This can be interpreted in the geometric picture of orthogonal projection.

Definition of $R^2$ : Faraway p. 23

$$R^2 = 1 - \frac{\text{sum of squared residuals}}{\text{total sum of squares (corrected for mean)}}$$

$R^2$ is something people use to interpret the output of a regression.

Simplest point of view: high $R^2$ = good!

Notice: $0 \leq R^2 \leq 1$.

"High" means "close to 1".

so a large percentage of the variation in $Y$ (dependent variable) is "explained" by the values of the explanatory variables $X_1, \ldots, X_p$.

Q: How high is high enough?

A: This is domain-specific. It depends.

Example: you're trying to explain student test scores. $R^2 = 0.9$ could be <u>great</u>!
In a physics experiment, $R^2 = 0.9$ could be <u>awful</u>!

Q: Does a high $R^2$ imply a causal connection between $Y$ (dependent variable) and $X_1 \ldots X_p$ (explanatory variables)?

A: No: Ice cream sales (X) and shark attacks (Y).

There is a strong correlation, because X and Y both go up in good weather.