Math 455 Class 5 January 28

Web page should now be fixed - please ask again if you can't see the notes.

Homework 2 is due today. The requirement that you use R Markdown only applies to homeworks from the Faraway text.

Q: Why the F-distribution?

A: The F-distribution appears in the usual regression summary. It is the statistic for a hypothesis test. The test is a test of the null hypothesis

$$\beta_1 = \beta_2 = \cdots = \beta_p = 0$$

against the alternative hypothesis

at least one $\beta_i \neq 0$.

For the model: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon$

response ... predictors

the null hypothesis of the test:

"none of the predictors have anything to do with the response"

the alternative hypothesis:

"at least one of the predictors has a linear relationship with the response."

In the 1-~~var~~ predictor-variable case, we saw that we could take a data set $(X_1, Y_1) \dots (X_n, Y_n)$ and use calculus to find the OLS estimators:

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2} \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Q: In writing down this formula, where did we use the 4 standard hypotheses of linear regression?

First we list the standard assumptions of linear regression.

1. Linearity: there is an approximate linear relationship $Y = \beta_0 + \beta_1 X + \varepsilon$

ε = error.

2. Normality: the errors ε have a normal distribution with mean 0.

3. Homoskedasticity: all errors have the same variance σ^2

4. Independence: the errors $\varepsilon_1, \ldots, \varepsilon_n$ are independent RVs.

A: We did NOT use the assumptions to get the formulas for $\hat{\beta}_1, \hat{\beta}_0$. Given any set of n points $(X_1, Y_1) \ldots (X_n, Y_n)$, we can find a line by minimizing

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_i \right)^2$$

Q: Where ARE we using the assumptions?

A: If we say $\hat{\beta}_0, \hat{\beta}_1$ are <u>also</u> the MLE (Max - Likelihood Est) of β_0, β_1, <u>then</u> we <u>do</u> need the axioms.

Other examples: (there are many).

We compute (for standard regression summary) something called an F-statistic. and other things we call t-statistics. We then say "these things have the t- (or F-) distributions, so I can compute a p-value."

When we compute this p-value we are using the t- or F-dist (with some parameters) and to know that the computed quantity has that distribution we need the axioms, because the t- and F-dist are cooked up out of normal RVs.

Other examples: we use (some of) the assumptions for the Gauss-Markov Theorem (OLS Est. are BLUE)

Errors vs. Residuals.

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \longleftarrow$ errors

$Y_i = \hat{\beta}_0 + \hat{\beta}_1 + X_i + e_i \longleftarrow$ residuals

What is the difference?

the residuals are defined using $\hat{\beta}_0, \hat{\beta}_1$:

y-dist from the regression line. $Y = \hat{\beta}_0 + \hat{\beta}_1 X$.

Why not use the errors instead?

In order to know the errors ε_i we need to know what β_0, β_1 are and we DON'T know this: β_0, β_1 are fixed but unknown parameters and the goal of our analysis is to estimate these parameters.

A property of the residuals:

$$\sum_{i=1}^{n} e_i = 0.$$

Q: Do the errors have this property?

Is it true that $\sum_{i=1}^{n} \varepsilon_i = 0$?

A: No. $\varepsilon_i \sim N(0, \sigma^2)$ are iid.

You learned in 447 (Wackerly, Th. 6.3) that a linear comb of indep normal RVs is normal.

So $\sum_{i=1}^{n} \varepsilon_i \sim N(0, n\sigma^2)$

<u>Remark on Notation:</u> Many notations for errors and residuals:

		Errors	Residuals.
Notation	1	ε_i	e_i
(Faraway) —	2	ε_i	$\hat{\varepsilon}_i$
	3	e_i	$\hat{e}_i$

Why is $\sum e_i = 0$?

Let's define $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

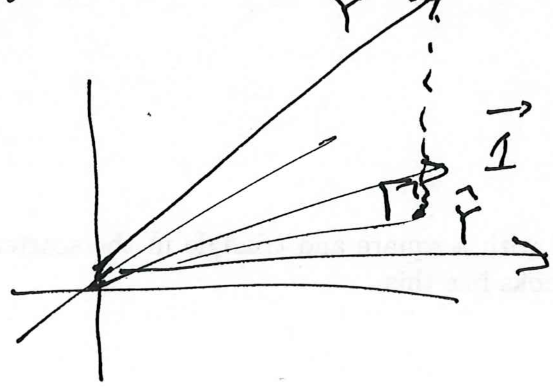
$\hat{Y}_i$ is our "estimate" of Y_i

$(X_i, \hat{Y}_i)$ it is the point on the regression line with X-coord X_i.

$$\hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \hat{\beta}_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \hat{\beta}_1 \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

$$Y = \hat{\beta}_0 \vec{1} + \hat{\beta}_1 X + e \qquad e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

Regression is orthogonal projection.



$\hat{Y}$ is the orthogonal projection of Y onto the plane of $\vec{1}, X$.

So $Y - \hat{Y} = e$ is perp. to plane of $\vec{1}, X$.

Thus $e \cdot \vec{1} = 0$ and $e \cdot X = 0$.

Thus $\sum_{i=1}^{n} e_i = 0$ and $\sum_{i=1}^{n} e_i X_i = 0$.

Interview questions: Why is $\sum e_i = 0$?

Suppose we did linear regression where we constrained the β_0 to be 0, that is we try to find β_1 in $Y = \beta_1 X + \varepsilon$.

We estimate β_1, by $\hat{\beta}_1$ is the β_1 which minimizes

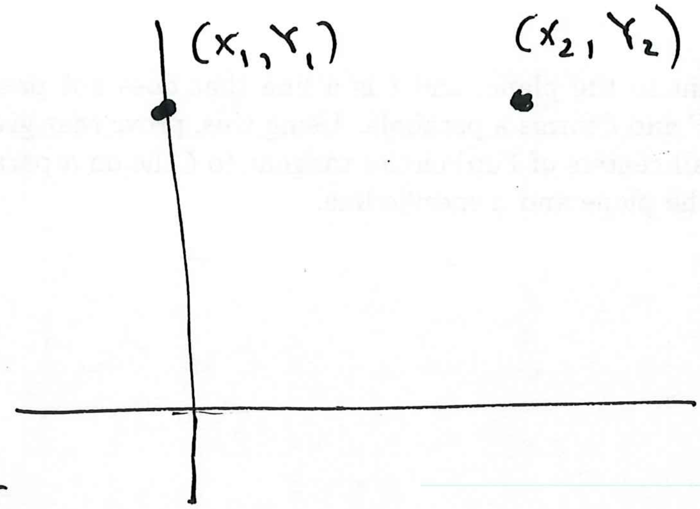
$$Q(\beta_1) = \sum_{i=1}^{n} (Y_i - \beta_1 X_i)^2$$

~~If~~ again we define $e_i = Y_i - \hat{\beta}_1 X_i$

Is it still true that $\sum e_i = 0$?

A: No.

Example:



n	X_1	Y_1
1	0	1
2	1	1

$$Y = \beta_1 X + \varepsilon$$

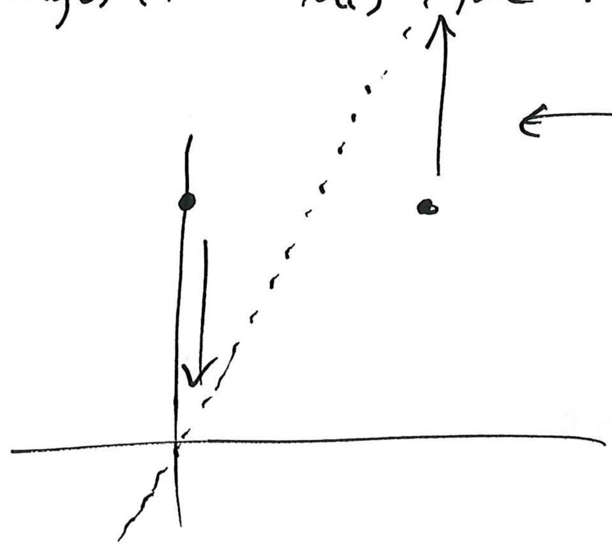
What is $\hat{\beta}_1$? Ans: 1

$\hat{\beta}_1$ minimizes $\quad Q(\beta_1) = \sum_{i=1}^{n} (Y_i - \beta_1 X_i)^2$

$\hat{\beta}_1 = 1. \quad Q(\beta_1) = (1 - \beta_1 0)^2 + (1 - \beta_1 1)^2$

$$= 1 + (1 - \beta_1)^2$$

Why isn't this the regression line



additional error makes $Q(\beta_1)$ bigger.