Math 455 Class 5 January 31

2 points of view on regression:

Data Set

Y	X
y_1	x_1
$\vdots$	$\vdots$
y_n	x_n

two vectors in $\mathbb{R}^n$. $\longrightarrow$

$\longleftarrow$ n points (x_i, y_i) in $\mathbb{R}^2$.

We want to find the "best" β_0, β_1

to create a line $Y = \beta_0 + \beta_1 X$

fitting these points.

Of course the line will not go through all the points.

We define "best" to be "minimizing the sum of squared errors":

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 .$$

Q: Could we minimize something else?

A: YES. This is in Chapter 8.

Q: Why the sum of squared errors?

A: It has some optimality properties, e.g. Gauss-Markov Theorem. (Ch. 2)

We can see an optimality property from our other point of view.

We'd like to write $\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_0 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

We can't do this in general, because $\mathbb{R}^n$ is not spanned by the vectors $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

The next best thing: Write $\hat{y}$ = orthogonal projection of $\vec{y}$ onto the plane spanned by $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. This is the closest point in the plane to $\vec{y}$.

$$\text{Dist.}(\vec{y}, \hat{y}) = \sqrt{(y_1-\beta_0-\beta_1 x_1)^2 + \cdots + (y_n-\beta_0-\beta_1 x_n)^2}$$

Minimizing this distance is the same as minimizing the sum of squared errors.

The optimality properties of regression can often be explained using the optimality of orthogonal projection.

We defined the regression model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Linear relationship with error, where
- errors iid $N(0, \sigma^2)$
  - normality
  - independence
  - homoskedasticity (variances same)

Q: This model is wrong!

A: See later, transforming data, Ch. 7 and elsewhere.

But more importantly:

"All models are wrong, but some models are useful".

Notice that the model is CONDITIONAL: it is a model for Y in terms of X.

That is ε is a RV

So Y, which depends on ε is a RV

BUT X is not a RV.

X might be chosen by the experimenter.

Q: But we can solve the equation for X!

$$X = \frac{(Y - \varepsilon - \beta_0)}{\beta_1}$$

See! X depends on ε.

A: Sometimes equations in statistics don't have the same meaning as they do in pure math.

Q: Since $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$,

after we determine the optimal $\hat{\beta}_0, \hat{\beta}_1$

can't we just figure out what the ε_i are?

By plugging in $\hat{\beta}_0, \hat{\beta}_1$ and solving?

A: Remember the Math 448!

β_0, β_1 are fixed, but unknown, parameters

$\hat{\beta}_0, \hat{\beta}_1$ are our estimates of β_0, β_1. (not $=$).

These two things are not the same.

The difference. $y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = e_i$ (or $\hat{\varepsilon}_i$)

is the ~~error~~ residual (which we <u>can</u> observe).

Not the the <u>error</u> ε_i (which we can't see).

Q: What's the difference? (Between e_i and ε_i)

A1: We can observe e_i but not ε_i.

A2: The e_i satisfy $\sum_{i=1}^{n} e_i = 0$.

The ε_i need not satisfy this.

Q: What is $\sum_{i=1}^{n} \varepsilon_i$?

A: ε_i are iid $N(0, \sigma^2)$. Thus

$\sum_{i=1}^{n} \varepsilon_i$ is a RV, with distribution:

$N(0, n\sigma^2)$

Q: What do we need to know for the midterm?!

A: You need to know the basic theory of the regression model, including all facts mentioned today.

- Assumptions
- Formulas for $\hat{\beta}_0, \hat{\beta}_1$
- $\bar{X}, \bar{Y}$ is on the regression line
- $\sum e_i = 0$.
- Difference between errors & residuals.
- etc.

Q: When is Midterm 1?

A: Proposed Date: Wed Before Break.
Feb 26?
March 5?