Random Independent Variables

STA302 Fall 2013

Don’t you think it’s strange?

- In the general linear regression model, the X matrix is supposed to be full of fixed constants.
- But in any non-experimental study, if you selected another sample, you’d get different X values, because of random sampling.
- So X should be random variables, not fixed.
- View the usual model as *conditional* on X=x.

Recall Double Expectation

$$E\{Y\} = E\{E\{Y|X\}\}$$

E{Y} is a constant. E{Y|X} is a random variable, a function of X.

$$E\{E\{Y|X\}\} = \int E\{Y|X = x\}\, f(x)\, dx$$

Beta-hat is (conditionally) unbiased

$$E\{\widehat{\boldsymbol{\beta}}|\mathbf{X}=\mathbf{x}\}=\boldsymbol{\beta}$$

Unbiased unconditionally, too

$$E\{\widehat{\boldsymbol{\beta}}\}=E\{E\{\widehat{\boldsymbol{\beta}}|\mathbf{X}\}\}=E\{\boldsymbol{\beta}\}=\boldsymbol{\beta}$$

Perhaps Clearer

$$
\begin{aligned}
E\{\widehat{\boldsymbol{\beta}}\} &= E\{E\{\widehat{\boldsymbol{\beta}}|\mathbf{X}\}\} \\
&= \int \cdots \int E\{\widehat{\boldsymbol{\beta}}|\mathbf{X} = \mathbf{x}\}\, f(\mathbf{x})\, d\mathbf{x} \\
&= \int \cdots \int \boldsymbol{\beta}\, f(\mathbf{x})\, d\mathbf{x} \\
&= \boldsymbol{\beta} \int \cdots \int f(\mathbf{x})\, d\mathbf{x} \\
&= \boldsymbol{\beta} \cdot 1 = \boldsymbol{\beta}.
\end{aligned}
$$

Conditional size α test, Critical region *A*

$$Pr\{F \in A | \mathbf{X} = \mathbf{x}\} = \alpha$$

$$\begin{aligned} Pr\{F \in A\} &= \int \cdots \int Pr\{F \in A | \mathbf{X} = \mathbf{x}\} f(\mathbf{x})\, d\mathbf{x} \\ &= \int \cdots \int \alpha f(\mathbf{x})\, d\mathbf{x} \\ &= \alpha \int \cdots \int f(\mathbf{x})\, d\mathbf{x} \\ &= \alpha \end{aligned}$$

The moral of the story

- Don't worry.
- Even though X variables are often random, we can apply the usual fixed-x model without fear.
- Estimators are still unbiased.
- Tests have the right significance level.
- And it's all *distribution-free* with respect to X.

Copyright Information