CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

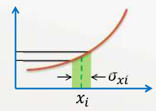
Lecture 31
Total Regression, part 2

Chris A. Mack
Adjunct Associate Professor

http://www.lithoguru.com/scientist/statistics/

Effective Variance Approximation

- We can simplify the regression for the case of small errors in x
  - Let $\hat{x}_i = x_i$
  - Define an effective variance in y using the model $\hat{y}_i = f(x_i)$



$$\sigma^2_{yi-eff} = \sigma^2_{yi} + \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma^2_{xi}$$

  - Use a weighted least-squares fit with these uncertainties
- Special case: $\sigma_{yi} = \sigma_{xi}$ = constant

Special Case: $\sigma_x = \sigma_y$

- The effective variance becomes

$$\sigma^2_{yi-eff} = \sigma^2_y\left[1 + \left(\frac{\partial f}{\partial x_i}\right)^2\right]$$

- For a straight-line model, $\sigma^2_{y-eff} \propto 1 + \beta_1^2$ and our chi-square becomes

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{\sqrt{1+\beta_1^2}}\right)^2$$

- Called Orthogonal Regression (or orthogonal distance regression)

Orthogonal Regression ($\sigma_x = \sigma_y$)

- The least-squares solution for this case is

$$b_1 = \frac{s_y^2 - s_x^2 + \sqrt{\left(s_y^2 - s_x^2\right)^2 + 4s_{xy}^2}}{2s_{xy}}$$

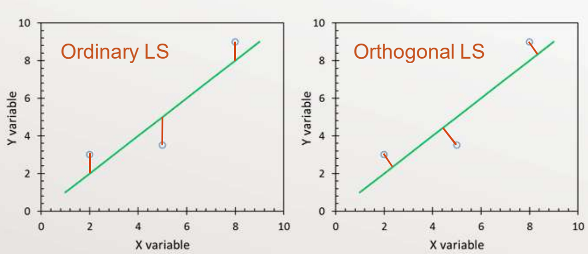
$$b_0 = \bar{y} - b_1\bar{x}$$

where s_x^2 = sample variance of x
s_y^2 = sample variance of y
s_{xy} = sample covariance of x and y

Standard error estimates will be discussed later

Orthogonal Regression ($\sigma_x = \sigma_y$)

- Very useful for tool matching, calibration curves

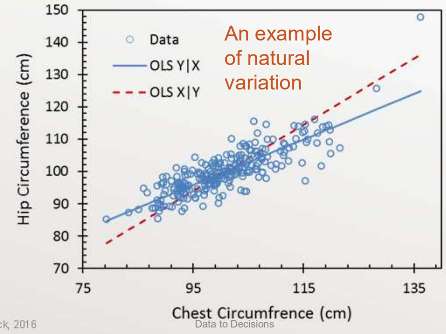


Regression: Y vs. X or X vs. Y?

- Sometimes, it is not obvious that one variable is a response and the other variable is a predictor
  - Natural variation (lurking variable): an unmeasured factor is controlling both variables
  - Calibration curve, tool matching
- How do we decide which variable to use as X (predictor) and which as Y (response)?
  - We often want a symmetric regression method

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Example: Body Fat Measurements



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Geometric Mean Regression

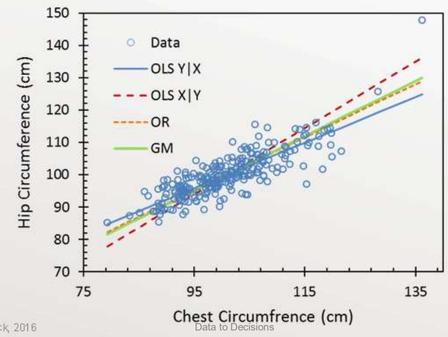
- Another proposal: estimate the slope as the geometric mean of the OLS slopes one obtains by regressing Y|X and X|Y
- After a small amount of algebra

$$b_0 = \bar{y} - b_1\bar{x}$$

$$OLS\ (Y|X):\ b_1 = \frac{s_{xy}}{s_x^2} \qquad GM:\ b_1 = \frac{s_y}{s_x}$$

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Example: Body Fat Measurements



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Method of Moments

- If we know σ_δ (the measurement error in x)
  - $x_i = \hat{x}_i + \delta_i$, $\delta \sim N(0, \sigma_\delta)$, s_δ = known

Then $$b_1 = \frac{b_{1(OLS}}{1 - \left(\frac{s_\delta}{s_x}\right)^2}$$ (only a reasonably good estimator for n > 50)

$$SE(b_1) = \frac{b_1}{\sqrt{n}}\sqrt{\frac{\left(s_x s_y\right)^2 + 2\left(b_1 s_\delta^2\right)^2}{\left(s_{xy}\right)^2} - 1}$$

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Consistency

- An estimator is consistent if its bias goes to zero as the sample size goes to infinity
- Let $b_{1(OLS)}$ be the OLS estimator of the slope of a straight-line model. We can show that

$$\lim_{n\to\infty} b_{1(OLS} = \frac{\beta_1}{1 + \left(\frac{\sigma_\delta}{\sigma_{\hat{x}}}\right)^2}$$

- The OLS estimator is not consistent when x has uncertainty (i.e., when σ_δ > 0)
  - We can predict how significant the bias will be before abandoning OLS

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

Lecture 31: What have we learned?

- What is orthogonal regression and when is it useful?
- What is geometric mean regression and when is it useful?
- What do the method of moments and the asymptotic behavior of OLS (as n → ∞) teach about the effect of x uncertainty on slope?