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

Lecture 32
Total Regression, part 3

Chris A. Mack
Adjunct Associate Professor

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

Deming Regression

- Consider a straight-line model with errors in both variables
  - $y_i = \hat{y}_i + \varepsilon_i, \quad x_i = \hat{x}_i + \delta_i$
- If both variables are normally distributed with constant variances and we know their ratio $\lambda = \sigma_\varepsilon^2/\sigma_\delta^2$, then

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

(Note: λ need not be dimensionless, but usually is)

Deming Regression

- The Deming regression estimate of the slope is the MLE estimate if both X and Y are normally distributed with independent errors
  - $\varepsilon \sim N(0,\sigma_\varepsilon), \;\; \delta \sim N(0,\sigma_\delta), \;\; \mathrm{cov}[\varepsilon,\delta] = 0$
  - $\lambda = \sigma_\varepsilon^2/\sigma_\delta^2$ is known
- In general, λ is estimated from external knowledge
  - Prior measurement error analysis

Deming Regression

- Deming regression also allows an estimate for the true x-value, $\hat{x}_i$

$$\hat{x}_i = x_i + \frac{y_i - (b_0 + b_1 x_i)}{b_1 + \lambda/b_1}$$

  - An OLS regression of y_i vs $\hat{x}_i$ produces the same fit as a Deming regression of y_i vs x_i
- Deming (optimized) residuals:

$$e_i = sign(y_i - \hat{y}_i)\sqrt{(y_i - \hat{y}_i)^2/\lambda + (x_i - \hat{x}_i)^2}$$

Deming Regression vs. Geometric Mean

- Consider the Deming regression when

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2} = \frac{s_y^2}{s_x^2}$$

- For this case, the Deming regression slope equals the geometric mean slope = $\sqrt{\lambda}$
  - The GM approach can be justified if all of the variation in X and Y comes from randomness (i.e., we did not systematically vary X or Y)

Standard Error of Coefficients

- Calculating the standard error of the coefficients for a total regression is not trivial (generally, we use a jackknife or bootstrap method)
  - Let our statistical software perform the calculations
  - See M. A. Creasy, "Confidence Limits for the Gradient in the Linear Functional Relationship", *Journal of the Royal Statistical Society B*, **18**(1), 65-69 (1956).
- A Method of Moments estimator gives

$$SE(b_1) = \frac{b_1}{\sqrt{n}}\sqrt{\frac{\left(s_x s_y\right)^2}{\left(s_{xy}\right)^2} - 1} = \frac{b_1}{\sqrt{n}}\sqrt{\frac{1}{b_{1OL\quad|X)}\, b_{1OLS(X|Y)}} - 1}$$

WHAT STARTS HERE CHANGES THE WORLD

A Caution about Deming

- Deming and orthogonal regression assume no model error (sometime called equation error), so that all of the uncertainty in X and Y is measurement error
  - Just because we have measurement error in both X and Y doesn't mean that Deming regression is appropriate
  - In many cases, model/equation error (unexplained variance in Y) causes our estimate of λ to be way off
- If we know σ_ε^2 and σ_δ^2 separately to compute λ, compare Deming regression to Method of Moments estimate
  - If the two produce very different estimates of the slope, then probably model/equation error exists and Deming Regression may not be the most appropriate

WHAT STARTS HERE CHANGES THE WORLD

Using Replicates

- We often use repeated measurements to independently estimate σ_ε^2 and σ_δ^2 and their ratio
- Consider the case of exactly two replicates for each data i, x_{i1} and x_{i2} giving y_{i1} and y_{i2}

$$s_\delta^2 = \frac{\sum(x_{i1} - x_{i2})^2}{2n} \qquad s_\varepsilon^2 = \frac{\sum(y_{i1} - y_{i2})^2}{2n}$$

This assumes constant σ_δ^2 and σ_ε^2

WHAT STARTS HERE CHANGES THE WORLD

Total Regression Review

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

$$GM:\ b_1 = \frac{s_y}{s_x} \qquad (\lambda = s_y^2/s_x^2)$$

$$OR:\ b_1 = \frac{s_y^2 - s_x^2 + \sqrt{\left(s_y^2 - s_x^2\right)^2 + 4s_{xy}^2}}{2s_{xy}} \qquad (\lambda = 1)$$

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

WHAT STARTS HERE CHANGES THE WORLD

The Intercept

- For all cases of $\hat{y} = \beta_0 + \beta_1 x$

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

$$SE^2(b_0) = \frac{s_\varepsilon^2}{n} + SE^2(b_1)\bar{x}^2$$

WHAT STARTS HERE CHANGES THE WORLD

References on Total Regression

- C.-L. Cheng and J. W. Van Ness, Statistical Regression with Measurement Error, Arnold Publishers, London (1999).
- Wayne A. Fuller, Measurement Error Models, John Wiley & Sons, New York (1987 or 2006).
- Albert Madansky, " The Fitting of Straight Lines When both Variables are Subject to Error", *Journal of the American Statistical Association*, **54**(285), pp. 173-205 (Mar., 1959).

WHAT STARTS HERE CHANGES THE WORLD

Lecture 32: What have we learned?

- What is Deming regression and when is it useful?
- How does Deming regression relate to OLS, geometric mean regression, and orthogonal regression?
- How can we use replicates to estimate measurement uncertainty?
- When is it improper to use Deming regression when we now the uncertainty in x?