Estimation of Measurement Error Properties: The Physical Activity Measurement Survey

by
Wayne A. Fuller, Bryan Stanfill, and David Osthus
Iowa State University
ITSEW
June 2013

Outline

Measurement error; introduction

Simple models

Physical Activity Measurement Study

- Procedure
- Model
- Estimates

Measurement

Conceptual quantity (Definition)

- Age
- Household
- Cultivated cropland

Measuring instrument

- Questionnaire
- Photo interpretation
- Physical measurement

Measurement Error and Definitions

Blood pressure is the conceptual average one would obtain if one could perform a large number of determinations with an accredited instrument

Unemployment rate is result obtained if one applies the Census Bureau procedure a large number of times to a large random sample

Types of Error

Device error
- Thermometer
- Questionnaire
- Photo interpretation

Natural variability
- Food intake
- Physical activity
- Lake size

Properties of Measurement Error

X = observation $\quad x$ = true

Distribution of X given x

$$E\{X \mid x\}$$

$$V\{X \mid x\}$$

Mean Properties

$$E\{X \mid x\} = h(x)$$

$$\text{Desirable: } E\{X \mid x\} = x \text{ for all } x$$

$$\text{Weaker: } E[E\{g(X) \mid x\}] = g(x)$$

$$E\left\{n^{-1} \sum_{i \in A} X_i\right\} = \mu_x$$

Simple Model

$$X_i = x_i + u_i,\ \ i = 1, 2, \ldots, m$$

$$\begin{pmatrix} x_i \\ u_i \end{pmatrix} \sim NI\left(\begin{pmatrix} \mu_x \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_u^2 \end{pmatrix} \right)$$

$$\begin{pmatrix} X_i \\ x_i \end{pmatrix} \sim NI\left(\begin{pmatrix} \mu_x \\ \mu_x \end{pmatrix}, \begin{pmatrix} \sigma_x^2 + \sigma_u^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 \end{pmatrix} \right)$$

$$E\{\overline{X}\} = E\{E(\overline{x} + \overline{u} \mid \overline{x})\} = E\{\overline{x}\} = \mu_x$$

$$V\{\overline{X}\} = n^{-1}(\sigma_x^2 + \sigma_u^2)$$

Simple Model (2)

$$X_i = x_i + u_i, \;\; i = 1, 2, \ldots, n$$

$$\theta = P\{x > C\} \qquad \hat{\theta} = n^{-1} \sum_{i \in A} \mathrm{I}\{X_i > C\}$$

$$E\{\hat{\theta}\} = P\{X > C\}$$

$$\text{Example}\,\{\sigma_x^2 = 1, \sigma_u^2 = 0.15, C = 1.645\}$$

$$E\{\hat{\theta}\} = 0.0625, \quad E\{\hat{\theta} - \theta\} = 0.0125$$

$$\text{At } n = 375, \quad (Bias)^2 = Variance$$

Estimating Error Properties

Additional data and Model, Replication

$$X_{ij} = x_i + u_{ij}$$

$$u_{ij} \sim ind\ (0, \sigma_u^2) \text{ ind of } x_i$$

$$\hat{\sigma}_u^2 = 0.5 \sum_{i=1}^{d} (X_{i2} - X_{i1})^2$$

Physical Activity Measurement Study (PAMS)

Principal Investigators:

Greg Welk, Kinesiology

Sarah Nusser, CSSM, Statistics

Alicia Carriquiry, Statistics

Data Collection: CSSM/SBRS (ISU)

Funding: NIH (R01 HL91024)

Physical Activity Measurement Study

Four Iowa counties, stratified, census tracts

Higher rates for minority tracts

List sample, Telephone screening, Age 21-70

Two years data collection; Sept. 2009-Sept. 2011

Time of year balance; 8 quarters

Monetary incentive

Sample

74% of telephone list contacted

33% of 74% agreed

88% of 33% completed two rounds

1347 of 1450 used in analysis

Data

Two rounds of data collection

Wear armband monitor 24 hours

24-hour Physical Activity Recall (PAR)

Telephone following day

Energy Expenditure Model

X_{ij} Monitor for person i on day j (log)

$X_{ij} = x_{ij} + u_{ij}$

$u_{ij} \sim ind(0, \sigma_u^2)$ ind. of x_{kt}

$x_{ij} = \mu + t_i + d_{ij}$ = true on day j

t_i = long run average for i

day effect = $d_{ij} \sim ind(0, \sigma_d^2)$ ind. of t_i

Energy Expenditure Model(2)

Y_{ij} Self report of person i for day j (log)

$$Y_{ij} = \beta_0^* + \beta_1 x_{ij} + r_i + e_{ij}$$

r_i = person reporting effect (long run)

e_{ij} = person day reporting effect

$r_i \sim ind(0, \sigma_r^2)$ ind. of (x_{kt}, t_k, d_{kt})

$e_{ij} \sim ind(0, \sigma_e^2)$ ind. of $(x_{kt}, t_k, d_{kt}, r_k)$

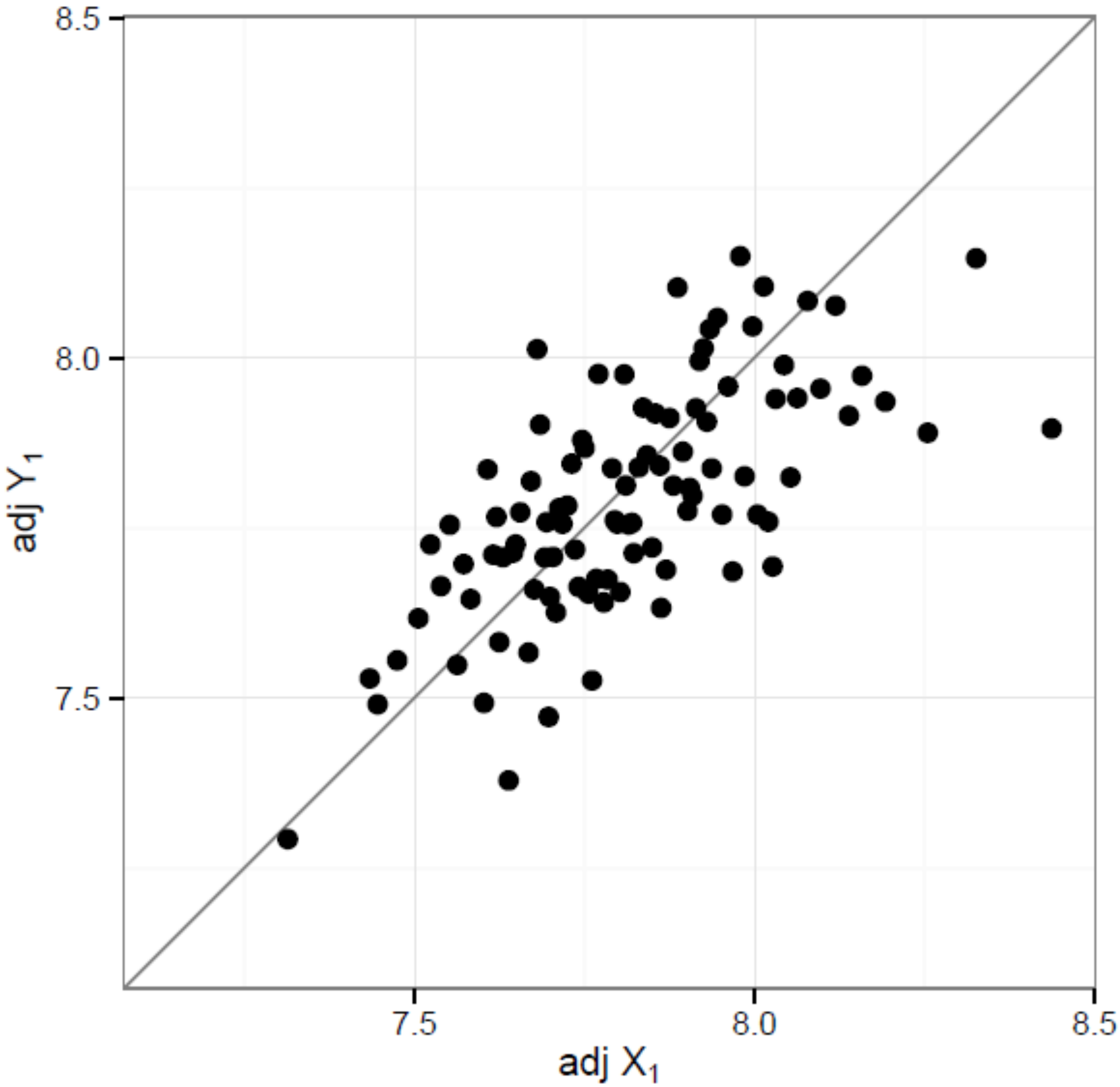
Model Assumptions

$$Y_{ij} = \mu + \beta_0 + \beta_1(t_i + d_{ij}) + r_i + e_{ij}$$

$$X_{ij} = \mu + t_i + d_{ij} + u_{ij} = x_{ij} + u_{ij}$$

$$(t_i, d_{ij}, r_i, e_{ij}, u_{ij}) \text{ mutually ind. } (\mathbf{0}, \mathbf{\Sigma})$$

$$\mathbf{\Sigma} = diag(\sigma_t^2, \sigma_d^2, \sigma_r^2, \sigma_e^2, \sigma_u^2)$$


8.5
8.0
7.5
adj Y_1
7.5
8.0
8.5
adj X_1

Method of Moments Estimation

$$V\{X_{i1} - X_{i2}\} = V\{d_{i1} - d_{i2} + u_{i1} - u_{i2}\} = 2(\sigma_d^2 + \sigma_u^2)$$

$$V\{0.5(X_{i1} + X_{i2})\} = V\{t_i + \overline{d}_{i\cdot} + \overline{u}_{i\cdot}\} = \sigma_t^2 + 0.5(\sigma_d^2 + \sigma_u^2)$$

$$[\hat{\sigma}_u^2, (\hat{\sigma}_u^2 + \hat{\sigma}_d^2)] / \hat{\sigma}_t^2 = [0.22\ (0.06), 0.40\ (0.03)]$$

$$[\hat{\sigma}_e^2, (\hat{\sigma}_e^2 + \hat{\sigma}_r^2)] / \hat{\sigma}_t^2 = [0.22\ (0.06), 0.44\ (0.04)]$$

$$\hat{\beta}_0 = -0.05(0.01)$$

$$\hat{\beta}_1 = 0.75\ (0.07)$$

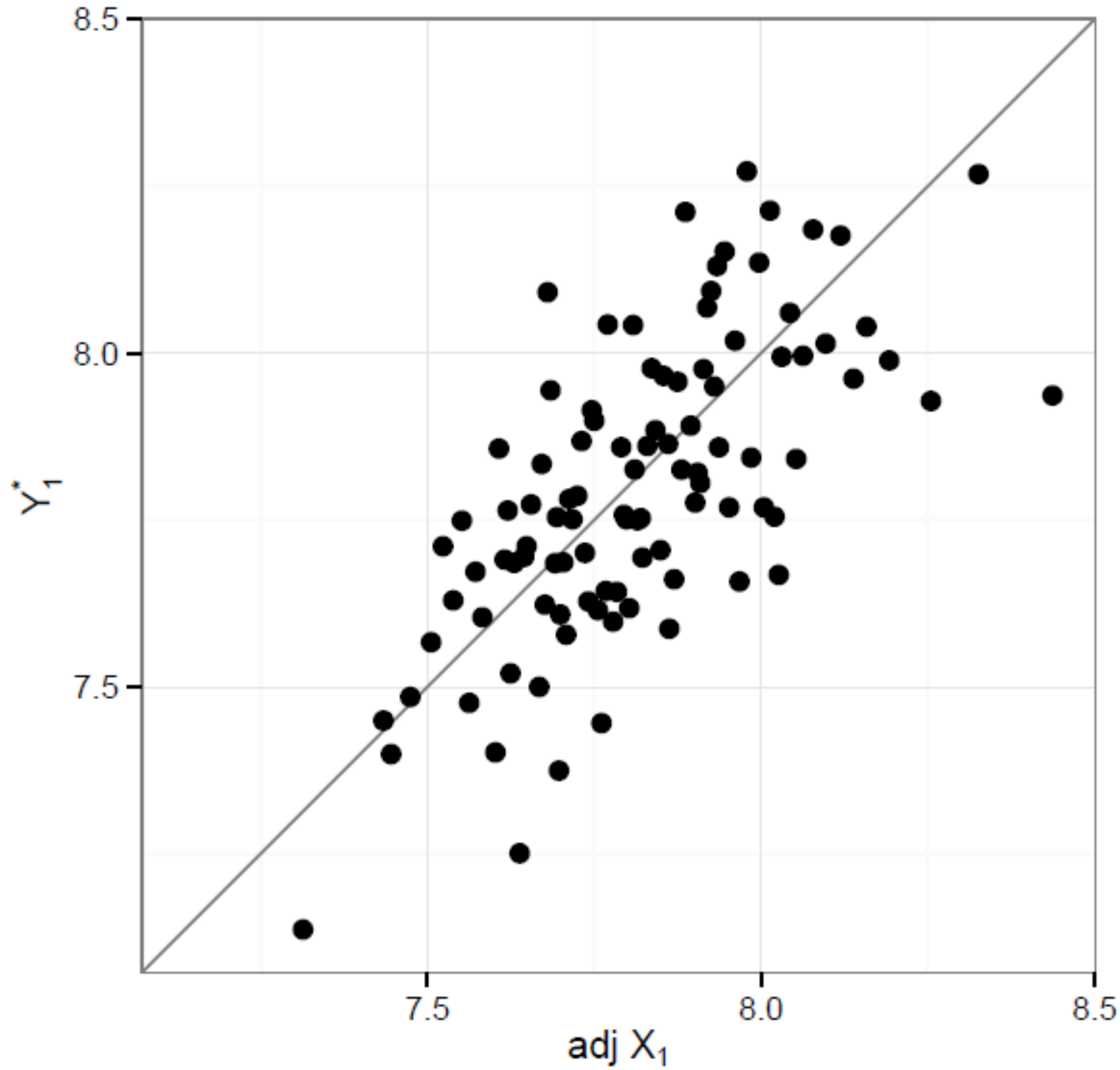
Calibrated Self Report

$$Y_{ij} = \mu + \beta_0 + \beta_1(t_i + d_{ij}) + r_i + e_{ij}$$

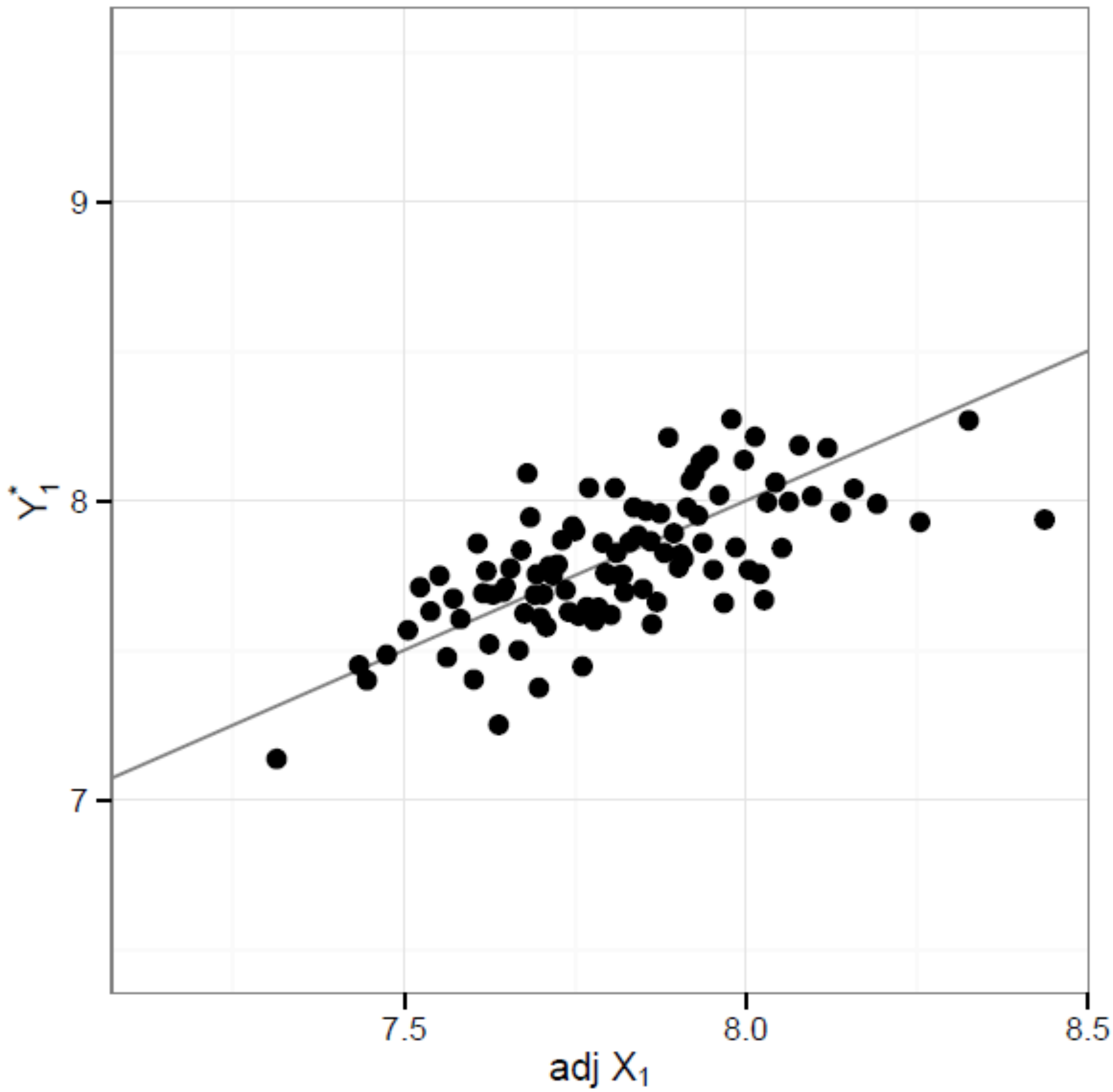
$$\beta_1^{-1}(Y_{ij} - \mu - \beta_0) + \mu = x_{ij} + \beta_1^{-1} r_i + \beta_1^{-1} e_{ij}$$

$$Y_{c,ij} = x_{ij} + r_{c,i} + e_{c,ij}$$

$$[\tilde{\sigma}_e^2, (\tilde{\sigma}_e^2 + \tilde{\sigma}_r^2), (\tilde{\sigma}_e^2 + \tilde{\sigma}_r^2, + \hat{\sigma}_d^2)]_c / \hat{\sigma}_t^2 = [\underset{(0.09)}{0.38}, \underset{(0.06)}{0.77}, \underset{(0.05)}{0.95}]$$



8.5
8.0
7.5
Y_1^*
7.5
8.0
8.5
adj X_1


9
8
7
Y_1^*
7.5
8.0
8.5
adj X_1

Variance Ratios

$$V\{\bar{Y}\}/V\{\bar{X}\} = 1.39$$

Regr. Coeff., $R^2 = 0.50$ Error Var. known

$$V\{\hat{\beta}_{\cdot|Y}\}/V\{\hat{\beta}_{\cdot|X}\} = 2.22$$

PAMS Study

Estimates of error variances

Estimate of calibration function

Internal Replication, (factor model)

$$Y_{2i} = \beta_{02} + x_i\beta_{12} + e_{2i}$$

$$Y_{1i} = \beta_{01} + x_i\beta_{11} + e_{1i}$$

$$X_i = x_i + u_i$$

$$(e_{2i}, e_{1i}, u_i) \sim (\mathbf{0}, diag(\sigma_{e2}^2, \sigma_{e1}^2, \sigma_u^2))$$

$$Cov\{Y_2(Y_2, Y_1, X)\} = (\beta_{12}^2\sigma_x^2, \beta_{12}\beta_{11}\sigma_x^2, \beta_{12}\sigma_x^2)$$

Estimating Error Properties

External information

Administrative, aggregate

- Census

Administrative, individual

- Medical
- Employer
- Government (Revenue Canada)

Money for Replicates

Estimate 5% point under normality

$$X_{ij} = x_i + u_{ij}$$

$$(x_i, u_{ij}) \sim N[(\mu_x, 0), diag(1, 0.15)]$$

$$\hat{\theta} = \overline{X} + 1.645(\hat{\sigma}_X^2 - \hat{\sigma}_u^2)^{0.5}$$

One X_i costs same as one d_f for $\hat{\sigma}_u^2$

Money for Replicates (2)

$$\hat{\theta} = \overline{X} + 1.645\hat{\sigma}_x = \overline{X} + 1.645(\hat{\sigma}_X^2 - \hat{\sigma}_u^2)^{0.5}$$

$$V(\hat{\theta}) = n^{-1}V(\overline{X}) + (1.645)^2 V(\hat{\sigma}_x)$$

$$V\{\hat{\sigma}_x\} \doteq 0.25\sigma_x^{-2}[2(n-1)^{-1}\sigma_X^4 + 2d_f^{-1}\sigma_u^4$$

$$\text{Budget} = 1000 \quad \text{equal cost}$$

$$d_f = 73 \qquad\qquad n = 927$$

Survey Estimation

Probability weights

Adjusted for nonreponse by strata

Post strata-Census age, gender, minority

Some collapsing of post strata