

5-1-2011

Using Finite Mixture Modeling to Deal with Systematic Measurement Error: A Case Study

Min Liu

University of Hawaii, minliu@hawaii.edu

Gregory R. Hancock

University of Maryland, ghancock@umd.edu

Jeffrey R. Harring

University of Maryland, harring@umd.edu

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Liu, Min; Hancock, Gregory R.; and Harring, Jeffrey R. (2011) "Using Finite Mixture Modeling to Deal with Systematic Measurement Error: A Case Study," *Journal of Modern Applied Statistical Methods*: Vol. 10 : Iss. 1 , Article 22.
DOI: 10.22237/jmasm/1304223660

Using Finite Mixture Modeling to Deal with Systematic Measurement Error: A Case Study

Min Liu
University of Hawaii,
Manoa, HI USA

Gregory R. Hancock Jeffrey R. Harring
University of Maryland,
College Park, MD USA

Conventional methods and analyses view measurement error as random. A scenario is presented where a variable was measured with systematic error. Mixture models with systematic parameter constraints were used to test hypotheses in the context of general linear models; this accommodated the heterogeneity arising due to systematic measurement error.

Key words: Finite mixture models, systematic measurement error, general linear model.

Introduction

In the social and behavioral sciences variables are frequently measured with error. A common approach is to treat measurement error as inherently unpredictable chance fluctuations, as opposed to something that can be caused by any factor that systematically affects measurement of a variable across the sample. This may be because systematic errors are – in general – difficult to detect, and because the estimation of the magnitude of such errors in practice is complex.

Overview of Finite Mixture Models

Finite mixture modeling is an analytical paradigm used to analyze data sampled from a heterogeneous population with a different

probability density function (PDF) for each component population. While population heterogeneity could arise from multiple unrelated probability distributions (e.g., Cauchy and Student's t distributions), the more typical viewpoint is to assume that data come from a composite (i.e., mixture) of two or more distributions from the same parametric family, with the stipulation that parameters are permitted to differ across the unobserved components (see Titterington, Smith & Makov, 1985). In the general case, for data in \mathbf{x} , a finite mixture of K densities can be formulated as

$$f(\mathbf{x} | \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k f_k(\mathbf{x} | \boldsymbol{\theta}_k), \quad (1)$$

where $\boldsymbol{\pi}$ contains mixing parameters π_k ($k = 1, \dots, K$) reflecting prior probabilities of sampling from the k^{th} mixture component (class), $f_k(\mathbf{x} | \boldsymbol{\theta}_k)$ represents the PDF for class k , and $\boldsymbol{\theta}_k$ is a parameter vector for class k . In addition, π_k values are restricted to be nonnegative and sum to 1 over all K classes. The likelihood for the general mixture in (1) can be written as:

$$L(\boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^N f(x_i | \boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^N \left[\sum_{k=1}^K \pi_k f_k(x_i | \boldsymbol{\pi}, \boldsymbol{\theta}_k) \right]. \quad (2)$$

Min Liu is an assistant professor in the Department of Educational Psychology in the University of Hawaii at Manoa. Her primary research interest is in latent variable mixture models. Email her at minliu@hawaii.edu. Gregory Hancock is a Professor and Chair of the Department of Measurement, Statistics and Evaluation. His primary research interest is in latent variable models. Email him at: ghancock@umd.edu. Jeffrey Harring is an Assistant Professor in the Department of Measurement, Statistics and Evaluation. Email him at: harring@umd.edu.

For finite mixtures, model parameter estimation is typically carried out by maximizing the logarithm of the likelihood function in (2) via the iterative expectation-maximization (EM) algorithm (Baum, Petrie, Soules & Weiss, 1970; Dempster, Laird & Rubin, 1977; Little & Rubin, 2002; McLachlan & Krishnan, 1997). In the current work, however, EM was not used in estimating parameters for the models of interest; instead, the natural logarithm of the likelihood in (2) was maximized directly using the Newton-Raphson algorithm (see Thisted, 1988).

This procedure uses a second-order Taylor series expansion of the gradient (first partial derivatives of the log-likelihood function) around the current estimate $\theta^{(w)}$ to produce the next estimate $\theta^{(w+1)}$. At each iteration the Newton-Raphson optimization scheme requires both first and second partial derivatives of the log-likelihood function with respect to the parameters, which can be computationally demanding. This challenge notwithstanding, Newton-Raphson and its many variants remain popular choices due to the algorithm's quick convergence near the solution, and the fact that standard errors of the estimated parameters can be computed directly at convergence.

Mixture Models with Functionally Related Parameters

Applications of mixture models in the natural and biological sciences typically involve the blending of different natural groupings, such as visually similar species (e.g., crabs from the Bay of Naples; Pearson, 1894) or within-species age clusters (e.g., fish; Summerfelt & Hall, 1987). Within the social and behavioral sciences, subgroups often result from differential responses to stimuli or treatments. This includes invoking different problem-solving strategies (e.g., in spatial rotation tasks; Mislevy & Verhelst, 1990), different responses to test speededness (Bolt, Cohen & Wollack, 2002), or different responses to individual test items (i.e., differential item functioning; Samuelsen, 2008). Ding (2008) recommended regression mixture models as useful tools for modeling population heterogeneity, thus improving the accuracy of the regression function as evidenced by the

much lower error variance within each class or component population.

In the above examples there are typically no functional relations per se between the parameters governing distributions of the component populations, other than perhaps that one population should be higher on average than another (e.g., an older population of fish should be longer). In other cases, however, there might be a very specific relation between component populations' parameters. Oja, Koiranen and Rantakallio (1991), for example, examined birth weight data from Northern Finland for two cohorts: one from 1966 and one from 1985-1986. For the latter cohort the gestational age could be determined more accurately by reference to ultrasound measurements; for the earlier cohort, however, the gestational age could only be assumed based on each mother's self-report regarding her last menstrual period.

An examination of the data yielded an unexpected difference between the cohorts; specifically, the mean birth weight for the later cohort was higher than the earlier cohort, while its birth weight of (apparently) pre-term newborns was lower than those from the earlier cohort. To explain this difference, Oja, et al. (1991) hypothesized more frequent systematic measurement error in gestational assessment for the earlier cohort. As a result they modeled the two cohorts as being comprised of three subgroups: (1) newborns whose mothers correctly knew their last menstrual period, (2) newborns whose mothers mistakenly thought their last period was earlier, and (3) newborns whose mothers mistakenly thought their last period was later. The researchers successfully modeled both cohorts using these three classes, where relations among those classes' distributional parameters were constrained as a function of three gestational ages (i.e., correct age, 4 weeks earlier and 4 weeks later).

This last example demonstrates that, given specific knowledge or hypotheses about the origins of different classes, parameters may be functionally related across those classes. This study illustrates the case of such hypothesized relations across classes in the context of a systematic measurement problem as it relates to the accuracy of general linear model analyses.

Using Mixture Models to Accommodate Systematic Measurement Error

This study was inspired by a systematic measurement problem encountered when exploring and analyzing existing data for National Survey of Child and Adolescent Well-being (U.S. Department of Health and Human Services, Administration for Children, Youth and Families, 2003). Longitudinal data were collected from children who were subject to child abuse or neglect and included a wide variety of physiological and psychological variables.

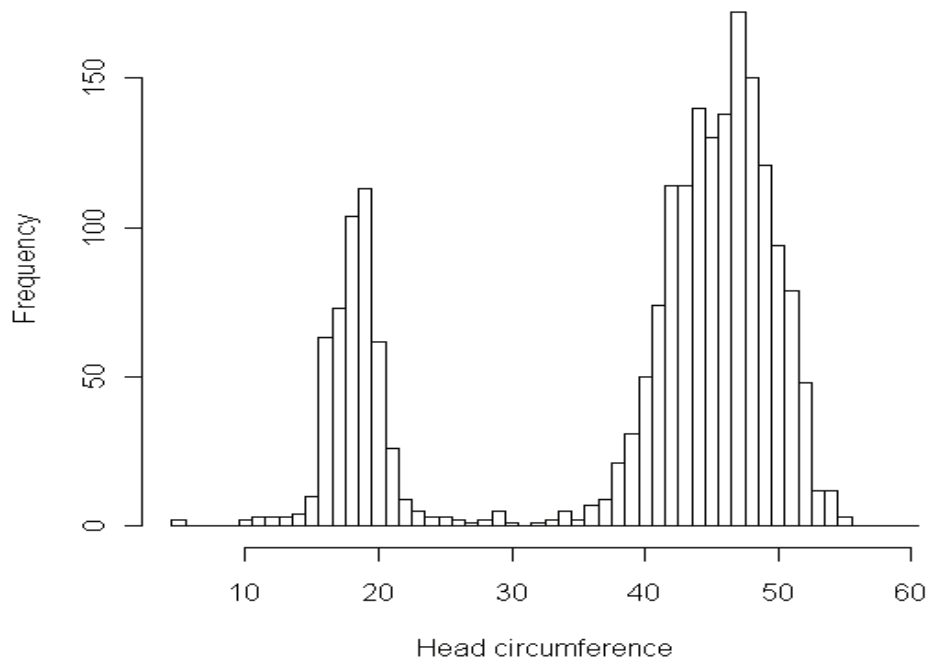
One key variable in the investigations of these data was head circumference for children up to 4 years old. Although members of the original research team had been instructed to measure head circumference in centimeters, an inspection of the data as shown in Figure 1 suggested that some of the researchers might have actually taken the measurements in inches (where 1 inch = 2.54 centimeters). [Unfortunately, the principal researchers were unable to contact individual field researchers to confirm our suspicions.] For example, the sample mean for the overall distribution was

38.99, which was inconsistent with prior published head circumference values for children in this age range (in either inches or centimeters).

Likewise, as is clear from Figure 1, the estimated standard deviation (12.29) was much larger than one would expect had only one set of units been employed. Thus, the apparently compromised first and second moments could not be used directly for even basic general linear model statistical analyses, such as a *t*-test or simple linear regression.

Fortunately, mixture modeling is a promising approach to accommodate systematic errors of this type. To this end, three studies were conducted using the problematic head circumference measure in the context of finite mixtures: a univariate analysis, a group means comparison and a simple regression. In each case competing models were constructed to analyze head circumference for a sample of 2,028 children. As noted, the Newton-Raphson algorithm was used and implemented using R to estimate all models (see Appendix for technical details; R code is available upon request from the first author).

Figure 1: Histogram of Head Circumference



MEASUREMENT ERROR MIXTURES

Methodology

Univariate Analysis Model 1: 1-Class Model

Based on the physiological literature (e.g., Fok, et al., 2003; World Health Organization, 2007), head circumference measurements were considered to approximate a normal distribution. A one-class model was used as a baseline with which to compare the results of other models. The PDF for each observation may be written as:

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\}.$$

Assuming there were N independent observations, the log-likelihood was:

$$LL = \ln L(\mu, \sigma^2) = \sum_{i=1}^N \ln f(x_i; \mu, \sigma^2).$$

The two parameters to be estimated in this model were the distribution mean μ and variance σ^2 .

Univariate Analysis Model 2: 2-Class Mixture Model without Constraints

The second model was a composite of PDFs assuming that measurements of head circumference arose from two distributions distinguished by which units (centimeters and inches):

$$\begin{aligned} f_{i_cent} &= f(x_i; \mu_{cent}, \sigma_{cent}^2) \\ &= \frac{1}{\sqrt{2\pi\sigma_{cent}^2}} \exp\left\{-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right\} \end{aligned}$$

and

$$\begin{aligned} f_{i_inch} &= f(x_i; \mu_{inch}, \sigma_{inch}^2) \\ &= \frac{1}{\sqrt{2\pi\sigma_{inch}^2}} \exp\left\{-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right\} \end{aligned}$$

The corresponding log-likelihood function for the mixture was formulated as:

$$\begin{aligned} LL &= \ln L(\mu_{cent}, \sigma_{cent}^2, \mu_{inch}, \sigma_{inch}^2, \pi_{inch}) \\ &= \sum_{i=1}^N \left\{ \ln \left[(1 - \pi_{inch}) f_{i_cent} + \pi_{inch} f_{i_inch} \right] \right\}. \end{aligned} \quad (3)$$

Five parameters were estimated in this model: two means μ_{cent} and μ_{inch} , two variances σ_{cent}^2 and σ_{inch}^2 , and the class mixing proportion parameter π_{inch} .

Univariate Analysis Model 3: 2-Class Mixture Model with Means Constrained

This model was identical to Model 2, with the exception that the means were constrained based on the fixed ratio of inch to centimeter ($\mu_{cent} = 2.54\mu_{inch}$). Given this constraint, only four parameters were estimated in this model: μ_{inch} , σ_{cent}^2 , σ_{inch}^2 , and π_{inch} .

Univariate Analysis Model 4: 2-Class Mixture Model with Variance Constrained

The fourth model hypothesized that the variances of two populations followed a fixed ratio of 6.45. This value came directly from the variance property

$$\begin{aligned} Var(cent) &= Var(2.54inch) \\ &= 2.54^2 Var(inch). \\ &= 6.45 Var(inch) \end{aligned}$$

Similar to the mean-constrained Model 3, the variance-constrained model also had four unknown parameters to be estimated: μ_{cent} , μ_{inch} , σ_{inch}^2 , and π_{inch} .

Univariate Analysis Model 5: 2-Class Mixture with Both Means and Variances Constrained

The final model assumed that the means and variances of the two distributions differed by a function of 2.54. With the additional equality constraints placed on the means and variances ($\mu_{cent} = 2.54\mu_{inch}$ and $\sigma_{cent}^2 = 6.45\sigma_{inch}^2$), only three unknown

parameters needed to be estimated: μ_{inch} , σ_{inch}^2 , and π_{inch} .

Model fit indices and corresponding parameter estimates for the univariate study are summarized in Tables 1 and 2, respectively. Three commonly used information criteria were determined for each analysis in this study. The Akaike's Information Criterion (AIC; Akaike, 1987) is based on the log-likelihood (LL) of the hypothesized model and the number of parameters (p), as follows:

$$AIC = -2LL + 2p.$$

An alternative to the AIC, the Bayesian information criterion (BIC; Schwarz, 1978), is obtained by modifying the penalty term based on sample size (N) as follows:

$$BIC = -2LL + p \ln(N).$$

In the context of finite mixture modeling, BIC has been recommended for its consistency (Haughton, 1988; Leroux, 1992), meaning that it tends to select the correct model more frequently as sample size increases. Sclove (1987) suggested a further sample size adjustment for BIC (S-BIC) where N is replaced by $N' = (N + 2) / 24$, and Yang (2006) advocated this index, citing better performance when the model has either a large number of parameters or a small sample size. In this study, all three information criteria were used to compare both non-nested and nested models; when models being compared were nested, Chi-square difference (likelihood ratio) tests were also conducted.

Table 1: Model Fit Indices from Univariate Analyses

Model	# Parameters	lnL	AIC	BIC	S-BIC
1. 1 Class	2	-7964.97	15933.94	15945.17	15938.82
2. 2-Class Mixture	5	-6501.70	13013.40	13041.47	13025.59
3. 2-Class Mixture Mean Constrained	4	-6504.55	13017.10	13039.56	13026.85
4. 2-Class Mixture Variance Constrained	4	-6561.07	13130.14	13152.60	13139.89
5. 2-Class Mixture Both Constrained	3	-6563.38	13132.76	13149.60	13140.07

Table 2: Maximum Likelihood Estimates from Univariate Analyses

Model	$\hat{\mu}_{cent}$	$\hat{\sigma}_{cent}^2$	$\hat{\mu}_{inch}$	$\hat{\sigma}_{inch}^2$	$\hat{\pi}_{inch}$
1. 1 Class	38.99	150.98	–	–	–
2. 2-Class Mixture	45.66	14.93	18.25	5.67	0.24
3. 2-Class Mixture Mean Constrained	45.73	15.01	18.00	5.63	0.24
4. 2-Class Mixture Variance Constrained	45.53	19.67	18.15	3.05	0.24
5. 2-Class Mixture Both Constrained	45.72	19.74	18.00	3.06	0.24

MEASUREMENT ERROR MIXTURES

As shown in Table 1, compared with Model 1 (which assumed population homogeneity), all other 2-class mixture models provided much better fit to the data; this is indicated by the larger (i.e., less negative) log-likelihood and the smaller fit indices (AIC, BIC and S-BIC). This finding provided initial support for the hypothesis that a systematic measurement error problem was present. Also noteworthy is the estimated mixing parameter (0.24), suggesting that roughly one-fourth of the subjects were improperly measured using inches rather than centimeters; this value was consistent across subsequent analyses as well.

Comparing the fit indices of the four mixture models, the unconstrained and mean-constrained models provided the best fit ($BIC_2=13041.47$, $BIC_3=13039.54$), whereas the last two models involving variance constraints did not fit as well. Looking at the two models without variance constraints shows that the estimated variances for the two classes did not follow the expected ratio of 6.45; this could be due to the different magnitudes of random measurement error caused by their different precisions of measurement.

Relatively speaking, measurement in centimeter units is more precise than that in inches; the latter thus introduces more random measurement error. Regarding the means, Table 1 shows the BIC favored the mean-constrained model (while the AIC and S-BIC were very close); Table 2 shows that the estimates from the mean-constrained and unconstrained models differed only at the decimal level. For these reasons, the functional relation of the two means was used in subsequent studies.

Group Means Comparison

The second study was a group means comparison that examined whether there was a gender difference in head circumference. Due to the apparent unit mixture problem, a typical t -test comparing males and females could not be used; similarly, because cases' membership into a centimeter or inch group was latent, separate male-female comparisons could not be directly conducted within each unit group. Thus, with the inch/centimeter mean constraint in place, the values of the log-likelihood were compared under two multi-sample mixture models: one

assuming a common mean for both genders and one estimating a separate mean for males (M) and females (F).

In the model assuming a gender difference, there were four normal PDFs crossing units and gender:

1. $f(x_i; \mu_{FemaleInch}, \sigma_{FemaleInch}^2)$ for females in inches with mixing proportion $\pi_{FemaleInch}$;
2. $f(x_i; 2.54\mu_{FemaleInch}, \sigma_{FemaleCent}^2)$ for females in centimeters with proportion of $1-\pi_{FemaleInch}$;
3. $f(x_i; \mu_{MaleInch}, \sigma_{MaleInch}^2)$ for males in inches with proportion of $\pi_{MaleInch}$; and
4. $f(x_i; 2.54\mu_{MaleInch}, \sigma_{MaleCent}^2)$ for males in centimeters with proportion of $1-\pi_{MaleInch}$.

It should be noted that a different mixing proportion was allowed for males and females ($\pi_{FemaleInch}$ and $\pi_{MaleInch}$), reflecting the possibility that researchers' erroneous use of inches rather than centimeters could have been related in some way to child gender. Overall, the log-likelihood for this model, which contained 8 parameters (4 per gender), could be written as:

$$LL = \ln L(\mu_{FemaleInch}, \sigma_{FemaleInch}^2, \sigma_{FemaleCent}^2, \pi_{FemaleInch}, \mu_{MaleInch}, \sigma_{MaleInch}^2, \sigma_{MaleCent}^2, \pi_{MaleInch}).$$

The second multi-sample mixture model, which assumed no gender difference in head circumference, included only one common mean for both gender groups to be determined. Thus, its log-likelihood, which contains 7 parameters, could be written as

$$LL = \ln L(\mu_{common}, \sigma_{FemaleInch}^2, \sigma_{FemaleCent}^2, \pi_{FemaleInch}, \sigma_{MaleInch}^2, \sigma_{MaleCent}^2, \pi_{MaleInch}).$$

Using the same estimation process previously described, model fit indices and estimated parameters for the mean comparison analysis are presented respectively in Tables 3

and 4. These results show that all model fit indices supported the model specifying a gender difference in head circumference over that which assumed gender equality ($BIC_1=13036.50$, $BIC_2=13057.74$). Additionally, because the second model assuming no gender difference was nested within the first model, a Chi-square difference test was used to compare the fit of the two models.

The observed Chi-square difference statistic ($\chi^2_{df=1} = 28.86$, $p < 0.001$) indicated the first model with gender difference was statistically significantly better than the constrained model. Therefore, it was inferred that male and female children who experienced abuse or neglect did have different head circumferences, with males being larger (18.21 inches vs. 17.82 inches). It is worth noting that if a traditional t -test had been used directly with the original unit-compromised variable, a non-significant test result would have been obtained ($t_{(df = 2026)} = -1.706$, $p = .088$); thus, the model would have failed to detect any difference between males and females.

Regression Analysis

The third study investigated if age in months was a useful predictor of head circumference by applying systematically constrained mixture regression models to these two variables. After the first baseline model, four regression mixture models were investigated, assuming different regression functions for the two unknown groups with different measurement units.

Model 1: Simple Regression of Head Circumference (y) on Age (x)

The simple bivariate regression model could be specified as $y_i = \alpha + \beta x_i + e_i$, where α is the population intercept and β is the population slope. The residual e_i was assumed to be normally distributed, making the PDF for an individual observation

$$f(x_i, y_i; \alpha, \beta, \sigma_e^2) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left\{-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_e^2}\right\}.$$

Table 3: Model Fit Indices from Mean Comparison Analyses

Model	# Parameters	lnL	AIC	BIC	S-BIC
Assuming Gender Difference	8	-6487.79	12991.58	13036.50	13011.08
Assuming Gender Equality	7	-6502.22	13018.44	13057.74	13035.50

Table 4: Maximum Likelihood Estimates from Mean Comparison Analyses

Model	$\hat{\mu}_{FemaleInch}$	$\hat{\sigma}^2_{FemaleInch}$	$\hat{\sigma}^2_{FemaleCent}$	$\hat{\pi}_{FemaleInch}$	$\hat{\mu}_{MaleInch}$	$\hat{\sigma}^2_{MaleInch}$	$\hat{\sigma}^2_{MaleCent}$	$\hat{\pi}_{MaleInch}$
Assuming Gender Difference	17.82	4.91	14.67	0.246	18.21	7.85	13.64	0.244
Assuming Gender Equality	18.02	4.93	14.96	0.246	18.02	7.47	14.18	0.243

MEASUREMENT ERROR MIXTURES

The log-likelihood for the full sample was thus

$$LL = \ln L = \sum_{i=1}^N [\ln(f(x_i, y_i; \alpha, \beta, \sigma_e^2))].$$

As indicated by the log-likelihood function, 3 parameters needed to be determined.

Model 2: Mixture Regression Models without Any Constraints

In this model, all of the regression coefficients, residual variances, and class proportions were free to be estimated. Its equation could be expressed as $y_{ik} = \alpha_k + \beta_k x_i + e_i$ for each k^{th} class, with $k=1,2$ representing the inch and centimeter classes; thus, the two unit groups each had their own regression coefficients to be determined. Because each observation could be sampled from either of the two unit classes, there were two class-specific PDFs for an individual:

$$f_{i_inch} = f(x_i, y_i; \alpha_{inch}, \beta_{inch}, \sigma_{e_inch}^2) \\ = \frac{1}{\sqrt{2\pi\sigma_{e_inch}^2}} \exp\left\{-\frac{(y_i - \alpha_{inch} - \beta_{inch}x_i)^2}{2\sigma_{e_inch}^2}\right\}$$

and

$$f_{i_cent} = f(x_i, y_i; \alpha_{cent}, \beta_{cent}, \sigma_{e_cent}^2) \\ = \frac{1}{\sqrt{2\pi\sigma_{e_cent}^2}} \exp\left\{-\frac{(y_i - \alpha_{cent} - \beta_{cent}x_i)^2}{2\sigma_{e_cent}^2}\right\}.$$

The corresponding log-likelihood was

$$LL = \ln L \\ = \sum_{i=1}^N [\ln(\pi_{inch} f_{i_inch} + (1 - \pi_{inch}) f_{i_cent})];$$

in this case 7 parameters needed to be estimated.

Model 3: Mixture Regression Model with Slope Constraint

Given that the slope represents the degree of expected change in head

circumference per unit increase in age, the suspected unit problem would lead to the relation: $\beta_{cent} = 2.54\beta_{inch}$. Using this relation as a constraint, the PDF for centimeter group was

$$f_{i_cent} = f(x_i, y_i; \alpha_{cent}, \beta_{inch}, \sigma_{e_cent}^2) \\ = \frac{1}{\sqrt{2\pi\sigma_{e_cent}^2}} \exp\left\{-\frac{(y_i - \alpha_{cent} - 2.54\beta_{inch}x_i)^2}{2\sigma_{e_cent}^2}\right\}.$$

The resulting log-likelihood was

$$LL = \ln L = \sum_{i=1}^N \left[\ln(\pi_{inch} f_{i_inch}(\alpha_{inch}, \beta_{inch}, \sigma_{e_inch}^2)) \right. \\ \left. + (1 - \pi_{inch}) f_{i_cent}(\alpha_{cent}, \beta_{inch}, \sigma_{e_cent}^2) \right],$$

with 6 unknown parameters to estimate.

Model 4: Mixture Regression Model with Intercept Constraint

Based on the population relation for the intercept $\alpha_k = \mu_y - \beta_k \mu_x$, and assuming comparable ages for the two classes, it was not unreasonable to expect the familiar functional relation between the two classes' intercept terms $\alpha_{cent} = 2.54\alpha_{inch}$. After this replacement, the PDF for the centimeter group was

$$f_{i_cent} = f(x_i, y_i; \alpha_{inch}, \beta_{cent}, \sigma_{e_cent}^2) \\ = \frac{1}{\sqrt{2\pi\sigma_{e_cent}^2}} \exp\left\{-\frac{(y_i - 2.54\alpha_{inch} - \beta_{cent}x_i)^2}{2\sigma_{e_cent}^2}\right\}.$$

The resulting log-likelihood is

$$LL = \ln L = \sum_{i=1}^N \left[\ln(\pi_{inch} f_{i_inch}(\alpha_{inch}, \beta_{inch}, \sigma_{e_inch}^2)) \right. \\ \left. + (1 - \pi_{inch}) f_{i_cent}(\alpha_{inch}, \beta_{cent}, \sigma_{e_cent}^2) \right],$$

with 6 parameters to be estimated.

Model 5: Mixture Regression Model with Constraints on Both Intercepts and Slopes

Based on the rationales provided for Models 3 and 4, both relations $\beta_{cent} = 2.54\beta_{inch}$ and $\alpha_{cent} = 2.54\alpha_{inch}$ were applied in this

model. From these constraints, the PDF for the centimeter group was specified as

$$f_{i_cent} = f(x_i, y_i; \alpha_{inch}, \beta_{inch}, \sigma_{e_cent}^2) = \frac{1}{\sqrt{2\pi\sigma_{e_cent}^2}} \exp\left\{-\frac{(y_i - 2.54\alpha_{inch} - 2.54\beta_{inch}x_i)^2}{2\sigma_{e_cent}^2}\right\}$$

The corresponding log-likelihood function was

$$LL = \ln L = \sum_{i=1}^N \left[\ln(\pi_{inch} f_{i_inch}(\alpha_{inch}, \beta_{inch}, \sigma_{e_inch}^2)) + (1 - \pi_{inch}) f_{i_cent}(\alpha_{inch}, \beta_{inch}, \sigma_{e_cent}^2) \right]$$

with 5 parameters to be estimated in total.

Results

All estimated results and corresponding model fit indices are summarized in Tables 5 and 6, respectively. Table 5 shows that the mixture models (Models 2 - 5) were similar in fit and were strongly favored over the simple regression model (Model 1). Among the four mixture models, the choice depended on the information criterion measure used; specifically, while all mixture models were fairly close in information criteria values, the AIC and S-BIC favored the model with only the slope constrained (Model 3) while the BIC favored the model with intercept and slope constraints (Model 5).

A comparison of these two models using

a Chi-square difference test yielded a statistically significant difference ($\chi_{df=1}^2 = 6.08, p < 0.01$), thus favoring the model without the intercept constraint. A potential reason for this constraint's failure is that it rests upon the assumption that the mean age for the inch and centimeter classes is the same (i.e., that a researcher's mistaken decision to use inches was unrelated to the child's age); however, the estimated mean age for the two classes (based on a posterior probability-based group assignment) was 16.77 months for the inch class and 15.65 for the centimeter class. This mean difference notwithstanding, the slope and intercept parameter estimates for all regression mixture models, constrained or unconstrained, are practically close in value. Thus, to assess the linear relation between age in months and head circumference, the slope would lead to an estimate roughly one-fifth (0.20) of a centimeter increase in head circumference for each one month increase in age.

These results are corroborated by the plot in Figure 2, where circles represent bivariate age/head circumference observations. It may be observed that the solid simple regression line does not capture the relation between the two variables, while the two dashed regression lines generated from the mixture model provided a much better approximation to the plotted observations.

Table 5: Model Fit Indices from Regression Analyses

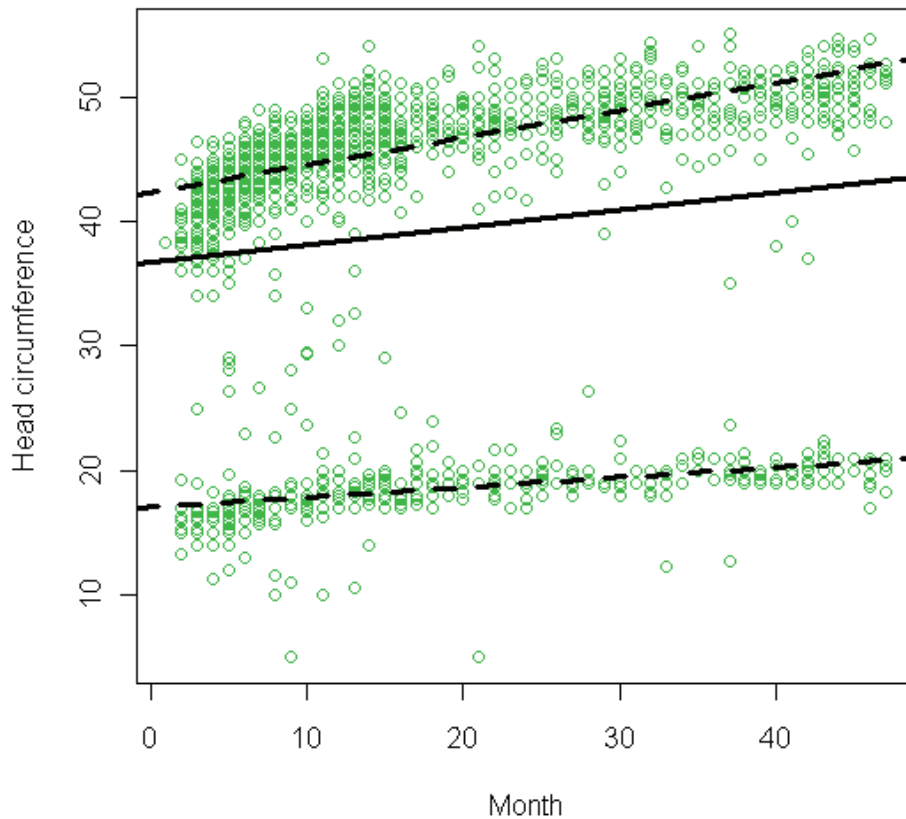
Model	# Parameters	lnL	AIC	BIC	S-BIC
1. Simple Regression	3	-7942.91	15908.66	15899.13	15899.13
2. 2-Class Regression Mixture No Constraint	7	-5918.70	11851.40	11890.70	11868.46
3. 2-Class Regression Mixture Constrained Slope	6	-5918.90	11849.80	11883.49	11864.43
4. 2-Class Regression Mixture Constrained Intercept	6	-5920.70	11853.40	11887.09	11868.03
5. 2-Class Regression Mixture Constrained Intercept and Slope	5	-5921.94	11853.88	11881.95	11866.07

MEASUREMENT ERROR MIXTURES

Table 6: Maximum Likelihood Estimates from Regression Analyses

Model	$\hat{\alpha}_{inch}$	$\hat{\beta}_{inch}$	$\hat{\sigma}_{e_inch}^2$	$\hat{\pi}_{inch}$	$\hat{\alpha}_{cent}$	$\hat{\beta}_{cent}$	$\hat{\sigma}_{e_cent}^2$
1. Simple Regression	36.70	0.14	147.73	–	–	–	–
2. 2-Class Regression Mixture No Constraint	17.04	0.08	5.88	0.25	42.34	0.22	6.83
3. 2-Class Regression Mixture Constrained Slope	16.95	0.08	5.89	0.25	42.35	0.20	6.83
4. 2-Class Regression Mixture Constrained Intercept	16.69	0.09	5.92	0.25	42.39	0.21	6.83
5. 2-Class Regression Mixture Constrained Intercept and Slope	16.69	0.08	5.93	0.25	42.39	0.20	6.84

Figure 2: Simple Regression Line (Solid) vs. Unconstrained Mixture Regression Lines (Dashed)



Conclusion

Based on the series of analyses examined in this study, mixture modeling appears to be an effective tool for investigating data consisting of variables with systematic measurement error. Systematic measurement errors have the potential to render data virtually useless: the

implementation of mixture models thus has the potential to salvage information regarding the key population relations in the data thereby avoiding the otherwise tremendous waste of time and expense associated with gathering such problematic data.

The studies presented may be viewed as a tip of the analytical iceberg. Although the only analyses demonstrated regarded mean comparisons and simple regression, extensions exist for the multivariate general(ized) linear model, as well as for latent variable models (e.g., structural equation models). The key to the implementation of such models is to have a specific hypothesis about the nature of the systematic measurement error and then translate that hypothesis into model constraints. It is hoped that the current case study has provided a useful illustration of the accommodation and adjustment for such measurement errors thereby bringing meaning to otherwise compromised data.

References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317-332.
- Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, *41*, 164-171.
- Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Applications of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, *39*, 331-348.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society*, *B39*, 1-38.
- Ding, C. S. (2008). Enhancing accuracy in research using regression mixture analysis. In J. Osborne (Ed.), *Best Practices in Quantitative Methods*, 410-421. Thousand Oaks, CA: Sage Publications
- Fok, T. F., et al. (2003). Updated gestational age specific birth weight, crown-heel length, and head circumference of Chinese newborns. *Archives of Disease in Childhood – Fetal and Neonatal Edition*, *88*, 229-236.
- Haughton, D. (1988). On the choice of a model to fit data from an exponential family. *The Annals of Statistics*, *16*, 342-355.
- Leroux, B. G. (1992). Consistent estimation of a mixing distribution. *Annals of Statistics*, *20*, 1350-1360.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data (2nd Ed.)*. New York: John Wiley.
- McLachlan, G., & Krishnan, T. (1997). *The EM algorithm and extensions. Wiley series in probability and statistics*. New York: John Wiley & Sons.
- Mislevy, R. J., & Verhelst, N. (1990). Modeling item response when different subjects employ different solution strategies. *Psychometrika*, *55*, 195-215.
- Oja, H., Koiraenen, M., & Rantakallio, P. (1991). Fitting mixture models to birth weight data: A case study. *Biometrics*, *47*, 883-897.
- Pearson, K. (1894). Contributions to the theory of mathematical evolution. *Philosophical Transactions of the Royal Society of London*, *A185*, 71-110.
- Samuelsen, K. M. (2008). Examining differential item function from a latent class perspective. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in Latent Variable Mixture Models*, 177-197. Charlotte, NC: Information Age Publishing, Inc.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.
- Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*, 333-343.
- Summerfelt, R. C., & Hall, G. E. (1987). *The age and growth of fish*. Ames, IA: Iowa State University Press.
- Thisted, R. A. (1988). *Elements of statistical computing: Numerical computation*. New York, NY: Chapman & Hall.
- Titterington, D. M., Smith, A. F. M., & Makow, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
- U. S. Department of Health and Human Services, Administration for Children, Youth and Families (2003). *National Survey of Child and Adolescent Well-Being (NSCAW)*, Restricted Release, Waves 1-4. Washington, DC.

World Health Organization (2007). *WHO child growth standards: Head circumference-for-age, arm circumference-for-age, tricep skinfold-for-age and subscapula skinfold-for-age*. Retrieved from http://www.who.int/childgrowth/standards/second_set/technical_report_2.pdf.

Yang, C. C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis*, 50, 1090-1104.

Appendix

The Newton-Raphson is a general optimization strategy based on a quadratic Taylor series expansion of the gradient (first partial derivatives of the log-likelihood function with respect to model parameters). To fashion the Newton-Raphson update, this quadratic function is maximized with respect to $\boldsymbol{\theta}$ in order to generate the next iterate. Setting the gradient expression equal to zero and solving for $\boldsymbol{\theta}$ provides the update

$$\boldsymbol{\theta}^{w+1} = \boldsymbol{\theta}^w - \mathbf{H}^{-1}(\boldsymbol{\theta}^w) \mathbf{g}(\boldsymbol{\theta}^w)$$

where \mathbf{H} is the Hessian, the matrix of partial second derivatives of the log-likelihood with respect to the parameter vector, and \mathbf{g} is the gradient.

Recall the log-likelihood function from Equation (3):

$$LL = \sum_{i=1}^N \left\{ \ln \left[\begin{aligned} &(1 - \pi_{inch}) f_{i_cent}(\mu_{cent}, \sigma_{cent}^2) \\ &+ \pi_{inch} f_{i_inch}(\mu_{inch}, \sigma_{inch}^2) \end{aligned} \right] \right\}$$

The gradient is defined as

$$\mathbf{g} = \begin{pmatrix} \frac{\partial LL}{\partial \mu_{inch}} \\ \frac{\partial LL}{\partial \sigma_{inch}^2} \\ \frac{\partial LL}{\partial \mu_{cent}} \\ \frac{\partial LL}{\partial \sigma_{cent}^2} \\ \frac{\partial LL}{\partial \pi_{inch}} \end{pmatrix},$$

with elements defined as shown in Appendix Figures. After elements of the gradient have been computed analytically, subsequent elements of the Hessian matrix can be obtained numerically.

Appendix Figures: Elements of Gradient

$$\frac{\partial LL}{\partial \mu_{inch}} = \sum_{i=1}^N \frac{\frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right) \left(\frac{x_i - \mu_{inch}}{\sigma_{inch}^2}\right)}{\left(\frac{1 - \pi_{inch}}{\sigma_{cent}}\right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

$$\frac{\partial LL}{\partial \sigma_{inch}^2} = \sum_{i=1}^N \frac{\frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right) \frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^4} - \frac{\pi_{inch}}{2\sigma_{inch}^3} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}{\left(\frac{1 - \pi_{inch}}{\sigma_{cent}}\right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

Appendix Figures: Elements of Gradient (continued)

$$\frac{\partial LL}{\partial \mu_{cent}} = \sum_{i=1}^N \frac{\left(\frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) \left(\frac{x_i - \mu_{cent}}{\sigma_{cent}^2} \right)}{\left(\frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

$$\frac{\partial LL}{\partial \sigma_{cent}^2} = \sum_{i=1}^N \frac{\left(\frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) \frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^4} - \left(\frac{1-\pi_{inch}}{2\sigma_{cent}^3} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right)}{\left(\frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$

$$\frac{\partial LL}{\partial \pi_{inch}} = \sum_{i=1}^N \frac{\frac{-1}{\sigma_{cent}} \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{1}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}{\left(\frac{1-\pi_{inch}}{\sigma_{cent}} \right) \exp\left(-\frac{(x_i - \mu_{cent})^2}{2\sigma_{cent}^2}\right) + \frac{\pi_{inch}}{\sigma_{inch}} \exp\left(-\frac{(x_i - \mu_{inch})^2}{2\sigma_{inch}^2}\right)}$$