Moravian College

Bethlehem, PA

Nutrient Intake of Dancers:

A Measurement Error Analysis

Emily G. Bolger

Advisor: Dr. Brenna Curley

Liaison: Dr. Jeffrey Bush

2020

Abstract

Since diet plays an important role in an athlete's overall health, many studies seek to understand the relationship between an athlete's daily intake and performance quality. These studies are not as frequently conducted for performance-based activities, such as dance, where most of the prior research explores whether a difference in nutrient intake between dancers and non-dancers exists. However, these studies fail to account for the error in measuring long-term average intake. To account for this error, we propose using a measurement error model. As an application, we analyze the relationship between dietary intake, body composition, and energy levels of pre-professional contemporary dancers.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank Dr. Meghan Brown for allowing us to use her data for our study. Through analyzing this dataset, I was better able to understand the nutrition habits of other dancers like me.

Next, I would like to thank the Moravian College Honors Program for welcoming me to conduct this research for the past year. I am grateful for the amazing experiences that working on this project has provided me. Additionally, I would like to thank the Moravian College Mathematics Department for allowing me to complete research in this department as well as adequately prepare me to work on this project over the past four years.

To my friends and family, at Moravian and beyond, I want to thank you for supporting and believing in me throughout this entire process. Your continuous support helped to strengthen my own beliefs.

Most importantly, I would like to thank my faculty advisor, Dr. Brenna Curley, for agreeing to work with and guide me over the past year of research. This project has allowed me to get a taste of the world of statistics research and better prepare me for my next steps in life. Through your guidance and advice, I have become a better statistician and researcher.

CHAPTER 1

Introduction

To perform at their best, an athlete should understand how heavily their dietary choices can affect their performance quality. To better understand this relationship between daily intake and performance, numerous studies have been conducted for athletes participating in a variety of sports (Campbell et al., 2007; ACSM, 2000; Joseph and Carriquiry, 2010). However, these studies are not as frequently conducted for performance-based activities, such as dance (Brown et al., 2017; Frost Brown et al., 2017). Due to the pressure of maintaining a particular body image, the results of studies focused on athletes in non-performance based activities are not directly transferable to dancers.

Dance is a relatively high-intensity activity that requires a great deal of strength and endurance. Depending on the dancer, the style, and where they train, a typical dance schedule can vary drastically. Often, there are many sessions in a week with each session being multiple hours long. These sessions can consist of technique classes, performance rehearsals, or performances. The intensity of these sessions, particularly in the pre-professional and professional world, are comparable to that of an elite athlete (Twitchett et al., 2010). To be a strong performer, it is imperative that a dancer has a nutrient-rich diet to sustain their mind and body during these long rehearsal sessions. Previous research in the dance world has been focused on the nutrient requirements for ballet dancers, which is only one of the many styles that dancers can learn (Brown et al., 2017). So, while these results are important, they are not directly transferable to dancers of other styles (including modern, contemporary, ballroom, tap, etc.). Each style of dance has its own movement quality and these varying movement qualities often affect the physique of the dancer. For example, dancers in the contemporary or modern dance world are asked to complete more floorwork and be more grounded in nature. This requires a different muscle focus than dancers in the ballet world, who often are encouraged to dance in a lighter, more lifted manner.

However, across all styles of dance, there often is a common aesthetic, or expected appearance, which motivates most dancers to maintain a lean physique. This pressure can stem from the dancer individually, from their colleagues, from their dance instructors, or be a combination of all three. As the commitment and experience level increases, so can the demand for a certain appearance. Thus, dancers who have been training for many years, like most dancers at the collegiate level, may be more driven to resort to unhealthy eating habits. In order to maintain their desired physique, dancers may rely on a diet plan or restrict their food consumption. If a dancer continually follows these practices, they may be more likely to develop an eating disorder (Yannakoulia et al., 2000). These practices could further encourage dancers to focus on maintaining a low body fat percentage (the percentage of body weight that is made up of fat) by avoiding certain food groups that could actually be beneficial to their overall health. While restrictive dietary habits may serve their purpose of maintaining a lean physique, they often have negative repercussions on the dancer's overall health and performance quality. This ideology is common not only among dancers, but among all females athletes, and is typically known as the female athlete triad. This triad describes the relationship between reduced nutrient intake, menstrual irregularities, and decreased bone mineral density that typically occurs in female athletes (Friesen et al., 2011).

Although it is important for dancers to intake more calories overall, a higher nutrient intake is not sufficient enough to optimize a dancer's performance ability and overall health. Without the sufficient intake of required macronutrients and micronutrients (like protein, carbohydrates, healthy fats, calcium, iron, vitamin D, etc.), dancers can suffer from numerous injuries, both internal and external. Long strenuous hours of exercise can cause small tears in the muscle fibers and tissues. Adequate intake of protein and other amino acids are needed to properly repair these tears in the muscles and tissues as well as produce the enzymes necessary for metabolism. Additionally, a lack of carbohydrates could affect the dancer's metabolism. Without proper sustenance, the dancer's body will not be able to delay fatigue during long rehearsals or performances due to low glycogen levels. Further, dancers should work to maintain the recommended fat levels as this is needed, among other things, as an important fuel for muscles (Brown et al., 2017; Clarkson, 2005).

Through restrictive eating habits, dancers are in danger of deficiencies in many micronutrients, such as calcium, iron, and vitamin D. Multiple studies report dancers barely meeting the recommended levels for these nutrients (Brown et al., 2017; Clarkson, 2005; Friesen et al., 2011; Lim et al., 2015). Bone fractures or other musculoskeletal damage from a low bone mineral density can occur due to an inadequate intake of calcium. Even though weight-bearing activities, such as modern dance, can increase bone mineral density levels, Friesen et al. (2011) did not find a discernible difference in bone mineral density levels between collegiate dancers and non-dancers. Similarly, a lack of vitamin D can have an affect on bone health and immune function, while a lack of iron can fail to compensate for the high rate of hemolysis (the destruction of red blood cells). Thus, potential deficiencies in calcium, iron, and vitamin D can highly affect the dancer's ability to perform movement, shorten the longevity of their dance career, and have lasting impacts on their health (Lim et al., 2015; Sousa et al., 2013).

To avoid the effects that result from an inadequate amount of macronutrients and micronutrients, nutritionists advise that dancers eat a meal or snack with a high enough carbohydrate level to last a whole rehearsal – an average of 1 to 4g per kg a few hours prior to a rehearsal is suggested. In addition to food consumption, nutritionists recommend dancers also consume 5 to 7 mL per kg of water at least 4 hours prior to rehearsal. Depending on the length of the rehearsal, adequate replenishment of food and fluids should also be ingested during the rehearsal in order to sustain muscle and brain function (Sousa et al., 2013; Clarkson, 2005; Rodriguez et al., 2009). In order to rapidly replace muscle tissue after prolonged activity, proper levels of carbohydrates, protein, electrolytes, and fluids are required after the rehearsal session is completed. These levels will vary depending on the length of the activity and the size of the dancer. Further, these recommendations also vary between training sessions and performances. For example, performances are not only longer than rehearsals, but often more strenuous. During performances, the dancer may be more self aware of their body image, which could further alter their food choices. Whether during a rehearsal or a performance, a dancer must intake enough nutrients (specifically macronutrients) to sustain themselves throughout the whole rehearsal or performance (Sousa et al., 2013).

To gain a better understanding of how nutrient intake levels can effect a dancer's performance, most of the prior research conducted in the dance world explores whether there exists differences in nutrient intake between a dancer group and a non-dancer group; or, if there exists any difference in nutrient intake between days of consumption. For example, Brown et al. (2017) computed three separate paired sample t-tests to compare total energy intake (TEI) and total energy expenditure (TEE) across the whole week, the weekend, and the weekdays. TEI did not discernibly differ between the weekends and the weekdays, while energy balance (EB) and TEE did. Note, energy balance is the difference between total energy intake and total energy expenditure. Additionally, the dancers were found, on average, to have a negative EB – a common trend in dancers. Brown et al. (2017) also noted that the dancers narrowly achieved the recommended carbohydrate and protein intake levels.

Similarly, Lim et al. (2015) conducted a study to compare Korean dancers, ballet dancers, contemporary dancers, and non-dancers, where no difference in nutrient intake, body composition in mean lean tissue, bone mineral density, or isokinetic muscular function was found between the groups. Like the results seen in Brown et al. (2017), Lim et al. (2015) found that nutrient intake needs to be increased, across dancers and non-dancers alike.

Friesen et al. (2011) conducted a multivariate analysis of variance (ANOVA) to compare nutrient intake and body composition between a group of 31 female collegiate dance majors and 30 non-dancers from a large university. Unlike Brown et al. (2017) and Lim et al. (2015), this study focused on understanding the prevalence of the female triad in dancers. The study found that there was no significant difference in body composition, bone mineral density, or diet composition between the two groups. However, on average, the percentage of fat was higher in non-dancers. Overall, the study found that the collegiate dance majors had greater muscle strength, higher bone mass near the spine and hips, as well as a lower body fat percentage compared to non-dancers.

To better understand how differing levels of protein and carbohydrate intake affect body composition and performance, Frost Brown et al. (2017) conducted a one-way ANOVA on 30 collegiate female dancers from the Florida State University School of Dance. The study found that there were no significant differences in the dancer's characteristics, their body composition, or their performance quality. Similarly, there was no difference in their carbohydrate intake or fat intake, but their was a significant difference in their protein intake.

Another measure of body composition relevant to dancers is fat free mass (FFM) – the amount of an individual's body weight the does not contain fat. We note that FFM is inversely related to percent body fat by the relationship: $\text{FFM} = \text{Weight} * \left(1 - \frac{\%\ \text{body fat}}{100}\right)$. In their work, Yannakoulia et al. (2000) developed multiple predictive models for FFM based on measurements from the bioelectrical impedance analysis (BIA) method. This method was used to gather measures like resistance and reactance from 42 young, female dancers in a professional dance school in Athens, Greece. Four regression models were built to predict FFM based on the BIA method measurements as well as the height and weight of the dancers. Together, these four predictors explained more than 80% of the variance in FFM.

While these studies help us better understand the nutritional tendencies of dancers, they fail to account for the error in measuring their long-term average nutrient intake. With all common methods used to measure nutrient intake, including food diaries and 24-hour recall, there is an error associated with estimating the usual intake based on the observed data. This error is typically called measurement error and it occurs when we can not directly observe our variables of interest. For example, it is nearly impossible to know the true, long-term average amounts of various nutrients that an individual consumes overall based on just a few days of observed data. So, we must consider the observed intakes as the noisy measurements of the true, long-term average nutrient intakes. Depending on how large the measurement error is for a certain observation, the difference between the observed value and the true value could be very large. Therefore, simply using the observed values as being equivalent to the long-term averages in a model would result in an inaccurate representation of the true values and produce very misleading results.

In this work, we propose a linear measurement error model to describe the relationship between the nutrition information, body composition, and energy levels of 25 pre-professional contemporary dancers; data are from the Brown et al. (2017) study. Unlike past work, we provide new insight to the dance field by developing a regression model that accounts for the error in estimating usual intake. This is the first example of a measurement error model analysis for dancers, to the best of our knowledge. However, like previous work, the regression models we built consider important body composition variables like FFM as well as nutrition variables like TEI, protein, and carbohydrates. In Chapter 2, we outline the methodology for a measurement error model and provide consistent estimators for the parameters. We outline the methodology first for the simple regression model, then extend to the multiple regression case. Following the previous work in the field of dance as a guideline, in Chapter 3, we apply the measurement error model to the Brown et al. (2017) data to describe the dependency of fat free mass, total energy intake, and energy balance on various macronutrients like carbohydrates and protein. Finally, in Chapter 4, we draw conclusions about the research, explain some of the limitations we noticed in our model, and explore questions for future research.

CHAPTER 2

Methodology: Measurement Error Model (MEM)

In this chapter, we outline the parameter estimates for the simple and multiple measurement error regression models. In Section 2.1, we define the model (including introducing notation) and the consistent estimators for the simple linear regression measurement error model. At the end of Section 2.1, we emphasize the importance of using a measurement error model by explaining what happens to the parameter estimates when we fail to account for measurement error. In Section 2.2, we define the model and estimators for the multiple linear regression measurement error model.

2.1. Simple Linear Regression Measurement Error Model

2.1.1. The Model. For simplicity, we begin with a measurement error model with only one predictor, x_i, and one response, y_i, for $i = 1, 2, \ldots, n$ individuals. Following Fuller (2009), our overall regression model is,

$$y_i = \beta_0 + \beta_1 x_i + q_i, \tag{2.1}$$

where q_i's are distributed as $N(0, \sigma_{qq})$, are independent of the x_i's, and are known as the *error in the equation*.

We are interested in a situation where we do not directly observe the predictor or the response. Instead, we observe a noisy measurement of our predictor and response, where,

$$Y_{ij} = y_i + w_{ij}, \tag{2.2}$$

$$X_{ij} = x_i + u_{ij}, \tag{2.3}$$

where $j = 1, 2, ..., m_i$ represents the number of replications for each individual i. Here, both Y_{ij} and X_{ij} represent the observed daily intakes, y_i and x_i represent the true, long-term average intake values, and w_{ij} and u_{ij} represent the errors in observing y_i and x_i. For estimation purposes, we assume that the response and predictor are independent and normally distributed: $y_i \sim N(\mu_y, \sigma_{yy})$ and $x_i \sim N(\mu_x, \sigma_{xx})$. Since the measurement errors between our response and predictors could be correlated, we define,

$$\begin{bmatrix} w_{ij} \\ u_{ij} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{ww}, \sigma_{wu} \\ \sigma_{wu}, \sigma_{uu} \end{bmatrix}\right). \tag{2.4}$$

Combining the regression model in (2.1) with the measurement error model for Y and X in (2.2) - (2.3) we have that,

$$Y_{ij} = \beta_0 + \beta_1 x_i + \epsilon_{ij}, \tag{2.5}$$

where $\epsilon_{ij} \sim N(0, \sigma_{ee})$ and $\epsilon_{ij} = w_{ij} + q_i$ is the sum of the error in measuring y_i, w_{ij}, and the error in the equation, q_i.

2.1.2. The Estimators. The maximum likelihood estimators for the parameters in the simple linear regression model in (2.5) are well established by Fuller (2009) and Buonaccorsi (2010). Note, the average value for a predictor for each individual is defined by,

$$\overline{X}_{i\cdot} = \frac{1}{m_i} \sum_{j=1}^{m_i} X_{ij}, \tag{2.6}$$

and similarly for a response,

$$\overline{Y}_{i\cdot} = \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij}. \tag{2.7}$$

Further, we define the average value of the predictor across all the individuals as,

$$\overline{X}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^{n} \overline{X}_{i\cdot} \tag{2.8}$$

and similarly for the response,

$$\overline{Y}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^{n} \overline{Y}_{i\cdot}. \tag{2.9}$$

We now estimate the coefficients of the measurement error model outlined in (2.5). To compute the estimators in (2.15) and (2.16), we will need to estimate the day-to-day variability and the overall variability for both the response and the predictors. The unbiased estimators for the within-person variance and covariance are,

$$\hat{\sigma}_{ww} = \frac{1}{n} \sum\sum \hat{\sigma}_{ww_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (Y_{ij} - \overline{Y}_{i\cdot})^2, \tag{2.10}$$

$$\hat{\sigma}_{uu} = \frac{1}{n} \sum\sum \hat{\sigma}_{uu_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (X_{ij} - \overline{X}_{i\cdot})^2, \tag{2.11}$$

$$\hat{\sigma}_{wu} = \frac{1}{n} \sum\sum \hat{\sigma}_{wu_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i - 1} \sum_{j=1}^{m_i} (X_{ij} - \overline{X}_{i\cdot})(Y_{ij} - \overline{Y}_{i\cdot}), \tag{2.12}$$

where $\hat{\sigma}_{ww}$ is the estimator for the variance in the response, $\hat{\sigma}_{uu}$ is the estimator for the variance in the predictors, and $\hat{\sigma}_{wu}$ is the estimator for the covariance between them.

The estimators for the between-person variances, or the overall variability, are,

$$m_{XX} = \frac{1}{n-1}\sum_{i=1}^{n}(\overline{X}_{i.} - \overline{X}_{..})^2, \tag{2.13}$$

and,

$$m_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(\overline{X}_{i.} - \overline{X}_{..})(\overline{Y}_{i.} - \overline{Y}_{..}). \tag{2.14}$$

Combining (2.10)-(2.14), we can define unbiased, consistent estimators for our regression coefficients by,

$$\hat{\beta}_1 = \frac{m_{XY} - \hat{\sigma}_{wu}}{m_{XX} - \hat{\sigma}_{uu}}, \tag{2.15}$$

and,

$$\hat{\beta}_0 = \overline{Y}_{..} - \overline{X}_{..}\hat{\beta}_1. \tag{2.16}$$

As a note for the reader, an estimator is considered consistent if it converges in probability to the true parameter value as n gets large.

The variance in $\hat{\beta}_1$ is estimated by,

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{1}{n}\left[\frac{s_{vv}}{m_{XX} - \hat{\sigma}_{uu}} + \frac{1}{(m_{XX} - \hat{\sigma}_{uu})^2}(\hat{\sigma}_{uu}s_{vv} + (\hat{\sigma}_{wu} - \hat{\beta}_1\hat{\sigma}_{uu}))^2\right] + \frac{1}{(m_{xx} - \hat{\sigma}_{uu})^2}\left[\hat{\sigma}_{uu}s_{rr} + (\hat{\sigma}_{wu} - \hat{\beta}_1\hat{\sigma}_{uu})^2\right] \tag{2.17}$$

where $s_{vv} = \frac{1}{n-1}\left(\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2\right)$ and $s_{rr} = \sigma_{ww} - 2(\hat{\beta}_1\sigma_{wu}) + \hat{\beta}_1^2\sigma_{uu}$.

We note that the estimator for β_0 in (2.16) is essentially identical to the parameter estimator for β_0 in the naive simple linear regression case. Further, there is a similarity between the m_{XX} value for the measurement error model and the parameter estimator for β_1 in the naive simple linear regression model. However, the difference for the β_1 estimator in the MEM case is that we account for noisiness in the data by accounting for the within-person variability. See Section 5.1 in the Appendix for a brief layout of the naive simple linear regression model.

2.1.3. Reliability Ratio. When using simple linear regression to model noisy data, the traditional least squares estimator does not account for the measurement error or the potential correlation between the measurement error in X and Y. In particular, when measurement error is present, the estimator is attenuated towards 0. An illustration of the effect on the regression coefficients when measurement error is ignored is seen in Figure 2.1. The data in Figure 2.1 representing the true, long-term averages are simulated from a normal distribution with mean and standard deviation similar to that of the observed protein intake seen in the Brown et al. (2017) dataset (see Section 3.1 in Chapter 3 for more detail). We also simulate the observed data by adding normally distributed errors to the true, long-term averages that have a mean and standard deviation of 0.5. We see that the error-prone, or observed, data (represented by the blue triangles) have much more variability around the regression line, while true long-term average values (represented by the pink circles) are much more tightly grouped to the best fit line. Thus, using a model which treats the noisy data as the truth (without accounting for additional measurement error) leads to bias in the estimated parameters of the regression line. This bias results in a regression coefficient that is attenuated towards zero in the simple linear regression case

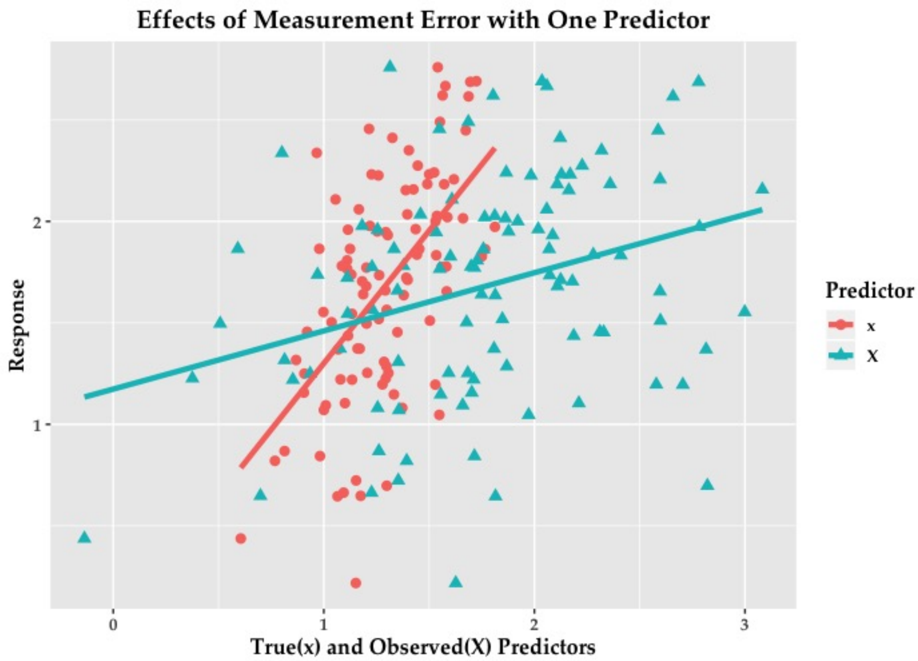

FIGURE 2.1. Depiction of the comparison between the true predictor values x_i and the observed X_{ij} values.

(Carroll et al., 2006). The attenuation factor is also referred to as the reliability ratio and provides a nice comparison between the $\hat{\beta}$ estimator for the naive simple linear naive regression model and the $\hat{\beta}$ estimator for a simple linear measurement error regression model. We define the reliability ratio by,

$$\kappa = (m_{XX})^{-1}(m_{XX} - \hat{\sigma}_{uu}), \tag{2.18}$$

for $\hat{\sigma}_{uu}$ defined in (2.11) and m_{xx} defined in (2.13) (Carroll et al., 2006; Buonaccorsi, 2010; Fuller, 2009). We can then determine the coefficient for the naive regression model, $\hat{\beta}_{SLR}$, by the relationship:

$$
\begin{aligned}
\hat{\beta}_{SLR} &= \kappa\hat{\beta}_{MEM} \\
&= (m_{XX})^{-1}(m_{XX} - \hat{\sigma}_{uu})\hat{\beta}_{MEM}. \qquad (2.19)
\end{aligned}
$$

In the simple linear regression case, we expect $|\hat{\beta}_{MEM}| \geq |\hat{\beta}_{SLR}|$. Referring to Figure 2.1, we see that the regression model for the observed data does indeed have a smaller slope. Note, the bias in the parameter estimates occurs in both the simple and multiple regression cases, however it is more clearly visualized in the simple case, where we see the bias parameter estimates attenuated towards zero.

2.2. Multiple Regression Measurement Error Model

We now extend the model in (2.1) - (2.5) to allow for multiple predictors.

2.2.1. The Model. For $i = 1, 2, ..., n$ individuals and $k = 1, 2, ..., p$ predictors, we define $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})'$ to be a vector of the true, long-term averages for each individual i and each predictor k. We then extend the regression model from (2.1) to be,

$$y_i = \beta_0 + \boldsymbol{\beta_1}\mathbf{x_i} + q_i \qquad (2.20)$$

where the error in the equation, q_i, is distributed as $N(0, \sigma_{qq})$ and independent of the x_i's and $\boldsymbol{\beta_1} = (\beta_1, \beta_2, ..., \beta_p)'$.

We define $\mathbf{X_{ij}} = (X_{ij1}, X_{ij2}, ..., X_{ijp})'$ to be a vector of observed intake values for each individual i across j replications for each predictor k. Similar to the simple linear case in (2.2) and (2.3), we cannot observe our response or predictors directly, so we have,

$$Y_{ij} = y_i + w_{ij}, \tag{2.21}$$

$$\mathbf{X_{ij}} = \mathbf{x_i} + \mathbf{u_{ij}}, \tag{2.22}$$

where $j = 1, 2, ..., m_i$ represents the replications collected for each individual i. We again see that Y_{ij} and $\mathbf{X_{ij}}$ represent the observed daily intakes, y_i and $\mathbf{x_i}$ represent the true, long-term average intake values with w_{ij} and $\mathbf{u_{ij}}$ as their respective errors.

We reach a similar final model as in (2.5) where,

$$Y_{ij} = \beta_0 + \boldsymbol{\beta_1} \mathbf{x_i} + \epsilon_{ij}, \tag{2.23}$$

for $\epsilon_{ij} = w_{ij} + q_i$, the sum of the error in the measuring y_i plus the error in the equation.

2.2.2. The Estimators. Following from Section 2.1.2, the average for each individual for each predictor is a $n \times p$ matrix, $\overline{\mathbf{X}}_{\mathbf{i}\cdot}$. The average value for the response is a $n \times 1$ matrix, $\overline{Y}_{i\cdot}$, the same as defined in Section 2.1.2. The average across all the individuals for each predictor is a $1 \times p$ vector, denoted by $\overline{\boldsymbol{X}}_{\cdot\cdot}$ which contains the mean for each predictor. Similarly, we average across all the individuals for the response and receive a single average value, $\overline{Y}_{\cdot\cdot}$. Note, this value is the same as in the simple case (2.9), since we only are considering a model with one response.

Extending the estimators defined in (2.10)-(2.12), we define the within-person variances measuring the day-to-day variability as,

$$\mathcal{S}_{ww} = \frac{1}{n}\sum\sum \mathcal{S}_{ww_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i - 1}\sum_{j=1}^{m_i}(Y_{ij} - \overline{Y}_{i\cdot})^2, \tag{2.24}$$

$$\mathcal{S}_{uu} = \frac{1}{n}\sum\sum \mathcal{S}_{uu_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i - 1}\sum_{j=1}^{m_i}(\mathbf{X_{ij}} - \overline{\mathbf{X}}_{\mathbf{i}\cdot})'(\mathbf{X_{ij}} - \overline{\mathbf{X}}_{\mathbf{i}\cdot}), \tag{2.25}$$

$$\mathcal{S}_{wu} = \frac{1}{n}\sum\sum \mathcal{S}_{wu_i} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i - 1}\sum_{j=1}^{m_i}(\mathbf{X_{ij}} - \overline{\mathbf{X}}_{\mathbf{i}\cdot})'(Y_{ij} - \overline{Y}_{i\cdot}). \tag{2.26}$$

The variances in (2.24) - (2.26) are similar to those in the simple linear regression case, except we now have another dimension to account for the multiple predictors. In other words, there are n, values for $\mathcal{S}_{ww_i}$ (overall dimension of $1{\times}1{\times}n$); n, $k{\times}k$ matrices for $\mathcal{S}_{uu_i}$ (overall dimension of $k{\times}k{\times}n$); and n, $k \times 1$ vectors of $\mathcal{S}_{wu_i}$ (overall dimension of $k \times 1 \times n$). When computing the estimates for $\mathcal{S}_{ww}$, $\mathcal{S}_{uu}$, and $\mathcal{S}_{wu}$ (see Appendix 5.3.1 for more details), we average across the first and second dimensions. We then end with a single value for $\mathcal{S}_{ww}$, a $k \times k$ matrix for $\mathcal{S}_{uu}$, and a $k \times 1$ vector for $\mathcal{S}_{wu}$. If there is no measurement error present in the predictor or response, these variances are zero.

Extending the definitions of the estimators in (2.13) and (2.14), we define the between-person or overall variability as,

$$M_{XX} = \frac{1}{n}\sum_{i=1}^{n}(\overline{\mathbf{X}}_{\mathbf{i}\cdot} - \overline{\mathbf{X}}_{\cdot\cdot})'(\overline{\mathbf{X}}_{\mathbf{i}\cdot} - \overline{\mathbf{X}}_{\cdot\cdot}), \tag{2.27}$$

and,

$$M_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(\overline{\mathbf{X}}_{\mathbf{i}\cdot} - \overline{\mathbf{X}}_{\cdot\cdot})'(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}), \tag{2.28}$$

where M_{XX} is a $p \times p$ matrix and M_{XY} is a $p \times 1$ vector.

Similarly, following from (2.15) and (2.16), we define the unbiased, consistent estimators for our regression coefficient by,

$$\hat{\boldsymbol{\beta}}_{\mathbf{1}} = \frac{M_{XY} - \mathcal{S}_{wu}}{M_{XX} - \mathcal{S}_{uu}}, \tag{2.29}$$

and,

$$\hat{\beta}_0 = \overline{Y}_{\cdot\cdot} - \overline{\mathbf{X}}_{\cdot\cdot}\hat{\boldsymbol{\beta}}_{\mathbf{1}}, \tag{2.30}$$

where $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ is a $p \times 1$ vector and $\hat{\beta}_0$ is a single value.

The variance in $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$, a $p \times p$ matrix, is estimated as,

$$\begin{aligned}\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{\mathbf{1}}) = \frac{1}{n}\Big[\tilde{M}_{xx}^{-1}s_{vv} + (\tilde{M}_{xx})^{-1}(\mathcal{S}_{uu}s_{vv}\tilde{S}_{uv}'\tilde{S}_{uv}(\tilde{M}_{xx})^{-1})\Big] \\ +\tilde{M}_{xx}^{-1}\Big[\mathcal{S}_{uu}s_{rr} + \tilde{S}_{uv}'\tilde{S}_{uv}\Big](\tilde{M}_{xx})^{-1},\end{aligned} \tag{2.31}$$

where $s_{vv} = \frac{1}{n-p}\left(\sum_{i=1}^{n}(Y_i - \mathbf{X}_{\mathbf{i}}\hat{\boldsymbol{\beta}}_{\mathbf{1}})^2\right)$, $s_{rr} = \mathcal{S}_{ww} - 2(\hat{\boldsymbol{\beta}}_{\mathbf{1}}\mathcal{S}_{uu}) + (\hat{\boldsymbol{\beta}}_{\mathbf{1}}'\mathcal{S}_{uu}\hat{\boldsymbol{\beta}}_{\mathbf{1}})$, $\tilde{M}_{xx} = M_{XX} - S_{uu}$, and $\tilde{S}_{uv} = S_{uw} - S_{uu}\hat{\boldsymbol{\beta}}_{\mathbf{1}}$.

CHAPTER 3

Application

In this chapter, we investigate the relationship between the average nutrient intake, body composition, and energy levels of pre-professional, contemporary dancers. In Section 3.1, we explore the data shared with us by Brown et al. (2017). We describe how the data were collected and note how these data compare to current recommended dietary intake values. In Section 3.2, following the previous work completed in the dance world as a guideline, we present models to help explain the dependency of fat free mass, total energy intake, and energy balance on important nutrients. We conclude the chapter, in Section 3.3, by outlining the assumptions of the error terms in our model and use visual diagnostic tools to address whether these assumptions are reasonably met.

3.1. Description of the Data

The data we analyze in this work were collected as a part of a study published by Brown et al. (2017). The 25 pre-professional female undergraduate contemporary dance students in the study were participating in a three-year, full-time undergraduate dance program. Prior to the study, the dancers were asked to complete the Healthier Dance Practice National Survey which requested information on their dietary and dance history (Laws and Apps, 2005). In addition to the questionnaire, standard skinfold techniques were used to get measurements related to their body composition, such as waist to hip ratio, percentage body fat, and fat free mass (FFM). Also reported in the survey was demographic information on each dancer (including their height, weight, age, years of study, etc.). To calculate their total energy intake (TEI), the dancers were required to complete a 7-day weighed food diary that required them to weigh all food and drink at the time of consumption. For the food diary, the participants were instructed to provide clear descriptions on their food and fluid intake, including the time of consumption, how much they ate, and the brand names. To cross-reference and clarify any ambiguous information in the food diary, each dancer also participated in a 24-hour recall interview, where they were asked to verbally report what they ate in the past 24-hours to the researcher. The data collected from both the interview and the diary were used to calculate TEI per day for each dancer. Also reported in the data were the amount of each nutrient consumed each day (e.g., protein, carbohydrates, water, various sugars, various vitamins, etc.). Total energy expenditure (TEE) was estimated using a tri-axial accelerometer. The accelerometer measures the vibrations of the dancer's hip in three perpendicular directions to determine how fast and in which direction the dancer is moving. The accelerometer was worn on the dancer's right hip at all times during the 7-day period, excluding activities in which the device would be submerged in water or cause discomfort during sleep.

3.1.1. Characteristics of the Data. To gain an understanding of how the dancers in the Brown et al. (2017) study compare to the average dancer, we compare the nutrient intake of these dancers with the current dietary recommendations for the average dancer or the average young, female adult, depending on the recommendations available. We note that due to the active lifestyle and dietary choices of dancers, the recommended levels differ slightly than the recommended values for the average person or even the average athlete. By examining current literature, we can determine the recommended nutrient values specific to female dancers (Sousa et al., 2013; Clarkson, 2005; Chandler, 2018; Rodriguez et al., 2009; Campbell et al., 2007). For example, Sousa et al. (2013) gathered their recommended values from the American Diet Association for active and competitive adults. They then adjusted the recommended values slightly to account for the low caloric intake of the average dancer. Sousa et al. (2013), Clarkson (2005), and Rodriguez et al. (2009) each report the recommended intake of dancers for carbohydrates as 3 to 5 g/kg/day and the recommended intake for protein as 1.2-1.7g/kg/day. Additionally, a dancer's TEI must be composed of 20%-35% fat in order to avoid affecting performance quality. Further, dancers are advised to consume more than 2 liters of water per day in order to maintain hydration during long rehearsal sessions.

In Table 3.1, we see that, on average, the dancers in the Brown et al. (2017) study met or exceeded the recommended intake values for protein, carbohydrates, and water. In general, we see that the variability of intake in these macronutrients is small relative to the mean. While the dancers appear to meet the recommended intake value for protein, on average they are close to the lower bound of the recommended intake range. However, unlike protein, the dancers almost exceed the recommended intake level for carbohydrates, on average. Additionally, we see that the dancers in the

Brown et al. (2017) dataset, on average, exceed the recommended intake for water.

Table 3.1. Macronutrients Table

	Protein	**Carbohydrate**	**Water**
Current Recommendation	1.2 – 1.7 g/kg	3 – 5 g/kg	at least 2L
Minimum	0.84	2.95	2.18
1st Quartile	1.13	4.56	2.98
Median	1.27	5.05	3.35
3rd Quartile	1.46	5.66	4.61
Maximum	1.96	7.31	5.07
Mean	1.28	4.96	3.64
Standard Deviation	0.267	0.94	0.88

Although the dancers in the Brown et al. (2017) study appear to be meeting the requirements for macronutrients on average, we are also interested in assessing whether these dancers are meeting the recommendations for various micronutrients. Dancers, particularly those that resort to restrictive eating habits, are at a greater risk compared to the normal population for micronutrient deficiencies (Chandler, 2018; Sousa et al., 2013; Clarkson, 2005; Rodriguez et al., 2009). Some restrictive eating habits encourage removing certain food groups, which could lead to deficiencies in micronutrients such as iron, calcium, and vitamin D. For the recommended values of these micronutrients, we look at the recommendations for an average 19-30 year-old female from the National Institute of Health (NIH) Dietary Reference Intakes (DRI) (Del Valle et al., 2011). According to the NIH, the recommended value for iron is 18mg/day, for calcium is 1000 mg/day, and for vitamin D is 15μ g/day for 19-30 year-old females.

In Table 3.2, we see that, on average, the dancers of the Brown et al. (2017) study have met or exceeded the recommended micronutrient intake values, except for calcium where over 75% of the dancers did not meet the recommended level. On average, the dancers have well exceeded the recommended values for iron and vitamin D. Note, that there are a few larger, outlying intake values for iron and vitamin D that could be influencing the average intake across all dancers. We also see that the spread of the data for all three micronutrients is much larger than the spread of the macronutrients data, potentially due to a few large outliers in the dataset.

TABLE 3.2. Micronutrients Table

	Calcium	**Iron**	**Vitamin D**
Current Recommendation	1000mg	18mg	15μg
Minimum	148.30	2.59	1.22
1st Quartile	364.50	6.87	2.15
Median	673.10	12.43	2.77
3rd Quartile	757.30	16.20	5.39
Maximum	1060.90	154.01	1002.28
Mean	611.50	20.80	53.15
Standard Deviation	259.31	32.64	199.77

Ultimately, we are interested in how the dietary intake of the macronutrients and micronutrients relates to a dancer's body composition and energy levels. In Table 3.3, we look at values relating to the body composition and energy levels of the dancers in the study. Given the pressure for dancers to maintain a low body fat, we are interested in their fat free mass (FFM), which is the mass of all the body components except fat. We were unable to find a recommended value for FFM for dancers, however we see that the dancers in the Brown et al. (2017) study have an average FFM of around 45kg. Additionally, we see that the variability of the observed data for FFM is relatively small, with values ranging from about 39kg to just over 45kg.

Further, we are curious to see if the pressure of maintaining a low body fat has an effect on the total energy intake (TEI) of these dancers. In general, dancers who train more frequently are estimated to intake 45-50 calories per kilogram of body weight. To get an overall recommended range for TEI, we multiplied the overall average weight of the dancers (63.408 kg) by the recommended caloric intake per kilogram of body weight range (Clarkson, 2005). While TEI is based solely on nutrient consumption, total energy expenditure (TEE) is comprised of three general components: resting energy expenditure, dietary induced thermogenesis (the heat produced in response to the processing of food within the body), and exercise (Pacy et al., 1996; Health Engine, 2019). To the best of our knowledge, we have not found a recommended value for TEE (and thus no recommended value for energy balance (EB)), most likely due to the variability of TEE across individuals (Rodriguez et al., 2009; Health Engine, 2019). An average TEE level varies greatly based on the age, weight, and activity level of the individual. Further, the complexity of measuring TEE hinders the ability to receive an accurate recommended value. For example, in the Brown et al. (2017) study, TEE is measured through a tri-axial accelerometer. However, in other studies a calculation of Metabolic Equivalents (METs), a measure of physical activity based on oxygen consumption, was used to estimate TEE in calories (Health Engine, 2019). Yet even other studies used indirect calorimetry – a method which uses a ventilator to measure the volume and concentration of oxygen taken in and the carbon dioxide released by an individual (Pacy et al., 1996). There are benefits and disadvantages to each method. Some methods are more accurate, for example the indirect calorometry, while other methods are more practical for active individuals, such as the tri-axial accelerometer.

When we look at the average TEI and EB for the dancers in the Brown et al. (2017) study in Table 3.3, we see that the variability of observed TEI and EB are much larger than the variability of FFM. We also notice that the dancers in the Brown et al. (2017) study, on average, fall below the recommended intake values by over 500 calories per day. While low caloric intake is concerning, it does align with many previous studies (Lim et al., 2015; Friesen et al., 2011; Pacy et al., 1996). We also notice that the average EB for the dancers in the Brown et al. (2017) study is negative, indicating that they are expelling more energy than they are consuming. Since the dancers fall below the recommended nutrient intake values, it is not surprising that their average EB is negative. Again, while concerning, this was also seen in many previous dance studies (Lim et al., 2015; Friesen et al., 2011; Pacy et al., 1996).

Table 3.3. Body Composition and Energy Levels Table

	Fat Free Mass	**Energy Intake**	**Energy Balance**
Current Recommendation		2,853.36-3,170.40 (Kcal/day)	
Minimum	39.43	1630.00	-2036.75
1st Quartile	41.78	2083.00	-902.00
Median	45.91	2470.00	-119.46
3rd Quartile	49.51	2730.00	43.67
Maximum	52.00	3265.00	517.30
Mean	45.52	2428.00	-355.68
Standard Deviation	4.27	450.05	656.61

3.2. Estimating the Association Between Dietary Intake, Body Composition, and Energy Levels

In Chapter 2, we introduced a model to describe the dependency of a response on a predictor – both potentially observed with measurement error. In this section, we attempt to use the linear measurement error model in (2.23) to describe the dependency of a dancer's body composition and energy levels on their nutrient intake.

In order to estimate the regression coefficients of the models that will be described in Section 3.2.1, we built a function to be used in the statistical software, R (version 3.4.3). The code can be found in Section 5.3.1 of the Appendix. As inputs, the function requires an $nm \times 1$ vector for the response variable, an $nm \times p$ matrix for the predictor variables, and an $nm \times 1$ vector containing a unique identifier for the participants. Note, for each individual $i = 1, \ldots, n$ there are m_i rows of data, depending on the number of days of data collected for each individual i. For example, in our dataset, we have 25 dancers each with 7 days of data, so we have a 175×1 vector for the response variable, a $175 \times p$ matrix for the p predictor variables, and a 175×1 vector of the 25 unique identifiers. As outputs, the function computes the statistics based on the estimators described in Section 2.2.2: a $px1$ $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ vector, a single value for $\hat{\beta}_0$, and a pxp $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_{\mathbf{1}})$ matrix.

3.2.1. Estimation in the Linear Measurement Error Model. In this section, we describe three different models that examine the possible relationships between the body composition, energy levels, and dietary intake of the female dancers in the Brown et al. (2017) study. Since previous research, such as Krieger et al. (2006) and Sousa et al. (2013), emphasize the impact of macronutrients on fat free mass and energy levels, we chose to focus on carbohydrates and protein as predictors, rather than the micronutrients. Further, preliminary analysis showed that models including water and the various micronutrients described in Section 3.1.1 did not meet the measurement error assumptions (see Section 5.2 in the Appendix for a more-detailed discussion). In the following Sections 3.2.1.1 - 3.2.1.3, we provide parameter estimates for both the measurement error model outlined in Section 2.2 as well as the naive estimates for a model assuming no error in estimating the long-term average intake (see Section 5.1). We will focus on the measurement error model estimates in this Chapter, however a discussion comparing the naive estimates with the measurement error estimates is in Chapter 4.

3.2.1.1. *Fat Free Mass.* Recall from Section 3.1.1, that FFM describes the amount in kilograms of an individual's body weight that does not contain fat. As indicated by Krieger et al. (2006), researchers have shown that a low carbohydrate and a high protein diet can retain FFM levels in an individual. With more protein, the individual would have an increased nitrogen balance which is essential to retaining FFM. Thus, it follows that the amount of fat in a dancer's body composition may be influenced by factors such as carbohydrates and proteins.

Given the findings from Krieger et al. (2006), we consider a model to investigate the dependency of FFM on the intake of protein and carbohydrates. Table 3.4 shows the test statistics, standard errors, and confidence intervals, for the parameter estimates of both the naive model and the measurement error model.

Table 3.4. Estimates for the FFM Model

X-Values	$\hat{\beta}_{SLR}$	$SE_{\hat{\beta}_{SLR}}$	$\hat{\beta}_{MEM}$	$SE_{\hat{\beta}_{MEM}}$	**t-stat**	**95% CI**
Carb	-1.33	0.37	1.36	1.57	0.87	(-1.87,4.60)
Protein	-0.61	1.31	0.71	4.40	0.16	(-8.39, 9.80)

Looking at the estimates in Table 3.4, we notice that the test statistic for both protein and carbohydrates are very small. These values, along with the confidence intervals, indicate that we do not have enough evidence to claim that protein or carbohydrates have an impact on FFM (while holding the other predictor constant) for the contemporary dancers in the Brown et al. (2017) study. In an attempt to control for the different weights of the individuals, we explored a model that included weight as a third predictor (similar to the TEI and EB models in Sections 3.2.1.2 and 3.2.1.3). However, we decided to not include weight in the final model, as it did not meet the model assumptions (see Chapter 4 for more details).

3.2.1.2. *Total Energy Intake.* In the next model, we look to predict total energy intake (TEI) based on protein, carbohydrates, and weight. As stated before in Section 3.1.1, it is crucial that dancers, as well as the general public, consume the proper amount of protein and carbohydrates to maintain a healthy lifestyle. Further, as mentioned in Chapter 1, it is critical that a dancer ingests enough protein to sustain their energy level, both physically and mentally, for an entirety of rehearsal. Dancers should also consume enough carbohydrates in order to delay fatigue during these long practice sessions. Thus, while other food groups might have an effect on TEI, one would expect that the most impactful for dancers are protein and carbohydrates. Also, as described by Chandler (2018) and other nutrition based websites (Health Engine, 2019; Sousa et al., 2013; Jodhun et al., 2017; Rodriguez et al., 2009), TEI is largely based on how frequently food is consumed by an individual as well as the weight or body mass index of that individual. Unlike the model for FFM, we include weight as a predictor for the TEI model to control the effect that weight has on TEI and gain a better understanding of how carbohydrates and protein impact TEI.

In order to understand how protein and carbohydrates may relate to TEI for the contemporary dancers in the Brown et al. (2017) study, we fit a model using carbohydrates, protein, and weight as predictors. Looking at the estimates in Table 3.5, we notice that the test statistic for protein is small, while the test statistics for carbohydrates and weight are large. Based on the parameter estimates and confidence intervals, we see that we have enough evidence to claim that carbohydrates and weight have a significant impact on TEI (while holding the other predictors constant), however, we do not have enough evidence to claim that protein has an impact on TEI.

Table 3.5. Estimates for the Total Energy Intake Model

X-Values	$\hat{\beta}_{SLR}$	$SE_{\hat{\beta}_{SLR}}$	$\hat{\beta}_{MEM}$	$SE_{\hat{\beta}_{MEM}}$	**t-stat**	**95% CI**
Carbs	262.15	22.21	390.55	134.20	2.91	(112.24, 668.86)
Protein	848.62	83.52	252.21	409.78	0.62	(-597.63, 1102.04)
Weight	45.17	3.07	40.09	10.62	3.77	(18.06, 62.13)

3.2.1.3. *Energy Balance.* After building the model for TEI, we were curious to see how carbohydrates and protein affect TEI, while accounting for TEE. Thus, following a similar structure to the TEI model, we built a model for energy balance (EB) using carbohydrates, protein, and weight as predictors. As described in Section 3.2.1.2, both protein and carbohydrates are expected to have a large impact on TEI and thus EB. Further, we include weight as a predictor again to control its effect on EB (Health Engine, 2019; Joseph and Carriquiry, 2010; Sousa et al., 2013; Rodriguez et al., 2009). Looking at Table 3.6, we see that the parameter estimate for carbohydrates has a large test statistic, while the estimates for both protein and weight have very small test statistics. Additionally, the confidence intervals for these parameter estimates indicate that we have enough evidence to claim that carbohydrates have a significant impact on EB (while holding protein and weight constant).

TABLE 3.6. Estimates for the Energy Balance Model

X-Values	$\hat{\beta}_{SLR}$	$SE_{\hat{\beta}_{SLR}}$	$\hat{\beta}_{MEM}$	$SE_{\hat{\beta}_{MEM}}$	**t-stat**	**95% CI**
Carb	181.08	48.53	467.10	202.28	2.31	(47.59, 886.60)
Protein	987.42	182.51	116.53	585.40	0.20	(-1097.53, 1330.58)
Weight	-6.36	6.71	-9.45	19.97	-0.47	(-50.86,31.96)

3.3. Model Diagnostics

3.3.1. Measurement Errors. As with any statistical model, we must make some assumptions about the errors in the model. One way to determine whether these assumptions are upheld are through visual diagnostic tools. The diagnostic tools used here are ones suggested by Carroll et al. (2006).

As stated in Section 2.1.1 and 2.2.1, we assume that the variance in the errors of our predictors, u_{ij}, are independent of the x_i's. To check this assumption, we plot the sample standard deviations of X_{ij} for each individual i against the respective sample mean, $\bar{X}_i$. If there are not any obvious trends in the plot, we can assume that the measurement error variance is independent of the x_i's. We check this assumption for our two predictors of intake: protein and carbohydrates (see Figure 5.2 in Appendix 5.2.1). Both plots show minimal trends, especially taking into account the small sample size of the Brown et al. (2017) dataset.

Similarly, we check that the measurement error variance does not depend on the predictors without measurement error. Since weight is only measured once for each individual over the course of the study, we would consider weight to be a variable measured without measurement error. To check this assumption, we plot the sample standard deviations of each X_{ij} for each individual i against the sample mean for weight. As seen in Figures 5.3 in Appendix 5.2.1, overall these plots are generally well-scattered, so we can reasonably assume that the measurement error variance does not depend on weight.

To estimate the model in 2.1.1 and 2.2.1, we also assume that the measurement errors are normally distributed. To check this assumption, we create a Normal Quantile-Quantile (QQ) plot for the measurement errors, u_{ij} (estimated by $\hat{u}_{ij} = X_{ij} - \bar{X}_i$), where we plot the quantiles of the measurement errors against the quantiles from a normal distribution, as this is the true (assumed) distribution of interest. As seen in Figure 5.4, most of the $\hat{u}_{ij}$ values follow closely to the line, except at the tails, not surprisingly, given the small sample size.

3.3.2. Residuals. As a reminder to the reader, the residuals are the difference between the observed and predicted response values. In simple linear regression, our residual values would be ϵ_i. However, we must account for the error in measuring x_i, so the residuals for the measurement error model are defined as:

$$\nu_i = \epsilon_i + \boldsymbol{\beta} u_i, \tag{3.1}$$

where ν_i is analogous to the error terms (or the deviations from the population regression line) in the naive regression model. Note, that ϵ_i and u_i, defined in (2.2)-(2.5), have constant variance, therefore ν_i also has constant variance. Additionally, since ϵ_i and u_i are independent of the x_i's, we know that the expected value of ν_i given x_i is 0. In an ordinary least squares regression, it is common to plot the residuals against the independent variables to provide clarity on nonlinearity in the regression and lack of homogeneity of the error variances. However, since the X_i's are dependent on the u_i values, we cannot plot the residual values against the X_i's. So, instead of using the observed sample average $\bar{X}_i$, we define an alternate estimator for the unobservable predictors, $\hat{x}_i$, to plot with our estimated residuals, $\hat{\nu}_i$ (Fuller, 2009).

3.3.2.1. *Best Linear Unbiased Predictor for Long-Term Average Intake.* In order to assess the assumptions of the errors in the equation, ϵ_i, we define the Best Linear Unbiased Predictor (BLUP) for x_i, denoted $\hat{x}_i$, following Carroll et al. (2006), Bertsekas and Tsitsiklis (2002), and Curley (2017). We define the BLUP for x_i to be the value of $E(x_i|X_i)$, where X_i is the average over each individual i. To determine this expected value, we define the joint distribution between x_i and X_i to be:

$$\begin{bmatrix} x_i \\ X_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_x + \mu_u \end{bmatrix}, \begin{bmatrix} \sigma_{x_i} & \sigma_{x_i X_i} \\ \sigma_{x_i X_i} & \sigma_{X_i} \end{bmatrix}\right). \tag{3.2}$$

Recall, from Chapter 2, we define the measurement errors as $u_{ij} \sim N(0, \sigma_{u_{ij}u_{ij}})$ and thus, the average over the measurement errors for each individual i is defined as $u_{i.} \sim N(0, \sigma_{u_{i.}u_{i.}})$. In the Brown et al. (2017) dataset, there are seven days of data for each individual i, so we can define $\sigma_{u_{i.}u_{i.}} = \frac{1}{7}\sigma_{u_{ij}u_{ij}}$.

Using properties of covariances, we further simplify the covariance between the long-term average intake (unobserved) and observed predictor. We can rewrite $\sigma_{x_iX_i}$ as,

$$\begin{aligned} Cov(x_i, X_i) &= Cov(x_i, x_i + u_{i}.) \\ &= E(x_i(x_i + u_{i}.)) - E(x_i)E(x_i + u_{i}.) \\ &= E(x_i^2 + x_i u_{i}.) - E(x_i)[E(x_i) + E(u_{i}.)] \\ &= E(x_i^2) + E(x_i u_{i}.) - E(x_i)^2 - E(x_i)E(u_{i}.) \\ &= E(x_i^2) - E(x_i)^2 \quad x_i\text{'s uncorrelated with } u_{i}. \ \& \ E(u_{i}.) = 0 \\ &= \sigma_{x_ix_i}. \end{aligned} \tag{3.3}$$

Following Carroll et al. (2006) and Bertsekas and Tsitsiklis (2002), we define $E(x_i|X_i)$ using properties of the conditional expectation of jointly normal random variables:

$$E(x_i|X_i) = \mu_{x_i} + \frac{\sigma_{x_ix_i}}{\sigma_{x_ix_i} + \sigma_{u_{i.}u_{i.}}}(X_i - (\mu_{x_i} + \mu_{u_{i.}})). \tag{3.4}$$

Using the BLUP, $\hat{x}_i$, we defined in (3.4), we can estimate the unobservable long-term average intake as:

$$\hat{x}_i = \hat{\mu}_{x_i} + \frac{\hat{\sigma}_{x_ix_i}}{\hat{\sigma}_{x_ix_i} + \hat{\sigma}_{u_{i.}u_{i.}}}(X_{ij} - \hat{\mu}_{x_i}), \tag{3.5}$$

noting that $\mu_{u_{i.}} = 0$.

Specifically for the Brown et al. (2017) dataset, since $\sigma_{u_{i.}u_{i.}} = \frac{1}{7}\sigma_{u_{ij}u_{ij}}$, we predict the unobservable long-term average intake x_i by,

$$\hat{x}_i = \hat{\mu}_{x_i} + \frac{\hat{\sigma}_{x_i x_i}}{\hat{\sigma}_{x_i x_i} + \frac{1}{7}\hat{\sigma}_{u_{ij}u_{ij}}}(X_{ij} - \hat{\mu}_{x_i}). \tag{3.6}$$

Following the estimators outlined in Section 2.2.2, we define σ_{X_i} to be M_{XX}, where M_{XX} is estimated by $\hat{\sigma}_{x_i x_i} + \frac{1}{7}\hat{\sigma}_{u_{ij}u_{ij}}$. Incorporating this idea, we can define $\hat{x}_i$ to be:

$$\hat{x}_i = \mu_{x_i} + \frac{M_{XX} - \frac{1}{7}\hat{\sigma}_{u_{ij}u_{ij}}}{M_{XX}}(X_{ij} - \hat{\mu}_{x_i}). \tag{3.7}$$

3.3.2.2. *Estimated Residuals.* As mentioned previously, we are not able to observe v_i directly and therefore must estimate it. We define an estimator for v_i, the residuals, to be

$$\hat{v}_i = Y_i - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}. \tag{3.8}$$

As mentioned in the beginning of the section, it is common to plot the residuals against the independent variables to better understand the presence of homoskedasticity in the errors in a measurement error model. Since we do not observe v_i and x_i directly, we plot $\hat{v}_i$ against $\hat{x}_i$ for each of the predictors of all three models (see Section 3.2.1 for more details) in Figures 5.5-5.7 in the Appendix . Overall, the residual plots for each of the models have decent scatter and lack any significant patterns.

CHAPTER 4

Discussion

As discussed in Chapter 1, due to the pressure of maintaining a desired physique, dancers often resort to restrictive eating practices that focus on maintaining a low body fat. For this reason, it is important to analyze the relationship between nutrient intake and body composition in dancers. However, as with all nutrient intake data, we must account for the noisy measurements in estimating long-term average intake. Following Fuller (2009), in this study, we proposed a measurement error linear regression model to investigate the impact on the nutrient intake of dancers. In Chapter 1, we outlined the previous research that has been performed in the dance world to provide insight on the nutrition of dancers at the collegiate level. In Chapter 2, we explained the methodology of the measurement error model, describing the parameter estimators for both the simple and multiple regression models. We also discussed the importance of using this method to analyze noisy data and the need to account for the within-person variability. In Chapter 3, we conducted an exploratory analysis of the data from a study conducted by Brown et al. (2017). To estimate the parameters in the linear measurement error model, we built a function in the statistical software, R. In our application, we looked at the relationship between nutrient intake, body composition, and energy levels of just 25 collegiate level contemporary dancers. However, the function in R can easily be extended to explore relationships in more comprehensive datasets for any response and predictors measured with error.

In order to investigate the relationship between the nutrient information, body composition, and energy levels of the dancers, we looked at three different models. In Section 3.2, we described how fat free mass (FFM), total energy intake (TEI), and energy balance (EB) depend on protein and carbohydrates. In Table 3.4, we saw that both protein and carbohydrates were not significant in predicting FFM. This was unexpected based on the findings from Krieger et al. (2006), as discussed in Chapter 3.2, which indicated that a high protein and low carbohydrate diet can retain FFM. Unlike the models for TEI and EB, we did not include weight as a third predictor in the FFM model. In the residual plots for the FFM model with weight included, we saw a clear upward trend in the protein residual plot which violates our assumptions about the error terms. Ideally, with more data and the ability to potentially add more predictors, we would no longer see a trend in the protein residual plot and thus be able to control for the effect weight has on FFM by including it as a predictor. For both TEI and EB, we saw that protein did not have an impact on the response, while there was enough evidence to suggest that carbohydrates did have an impact, as seen in Tables 3.5 and 3.6. Since dancers are recommended to intake a lot of protein and carbohydrates, as suggested by Sousa et al. (2013), Clarkson (2005), and Rodriguez et al. (2009), we were surprised that their protein intake did not have an impact on TEI and EB. We see a couple explanations for this: the small sample size leading to large variability of the parameter estimates (the Brown et al.

(2017) study only included data on 25 contemporary dancers), contemporary dancers may be a different population than that of other studies, or, by accounting for the measurement error, we were better able to detect true relationships between dietary intake, body composition, and energy levels. We now discuss a few of these possibilities further.

To emphasize the importance of accounting for the measurement error that occurs when measuring long-term average nutrient intake, we compare the naive regression model parameter estimates in Tables 4.1 - 4.3 with the measurement error model parameter estimates in Tables 3.4 - 3.6. For example, in Table 4.1, unlike the measurement error model where neither protein nor carbohydrates were significant in explaining FFM, we notice that in the naive model, carbohydrates has a small p-value. Thus, indicating that we have enough evidence to claim that carbohydrates has an effect on FFM. Further, looking at Table 4.2, we see that holding the other variables constant, there is enough evidence to claim that each of the predictors have an impact on TEI. Recall, from the measurement error regression model parameter estimates in Table 3.4 that only carbohydrates and weight had an impact on TEI. Table 4.3 shows the naive estimators for explaining the relationship between the dancer's nutrient intake and their EB. Unlike the measurement error regression model estimates, we have enough evidence to claim that carbohydrates and protein have an impact on EB (holding weight constant). However, similar to the measurement error model, we see that there is not enough evidence to claim that weight has a significant impact on EB, holding carbohydrates and protein constant. Thus, in each of the naive models, we saw that there was enough evidence to suggest that almost all of the predictors were significant in explaining the respective responses, holding the other variables in the model constant. This difference in the estimates and significance of the estimates between the naive and measurement error models is another illustration of the importance in accounting for the noisiness of the observed data.

Table 4.1. Estimates for the FFM Naive Model

X-Values	$\hat{\beta}_{SLR}$	$SE_{\hat{\beta}_{SLR}}$	**t-stat**	**p-value**
Carb	-1.33	0.37	-3.60	0.0004
Protein	-0.61	1.31	-0.47	0.64

Table 4.2. Estimates for the Energy Intake Naive Model

X-Values	$\hat{\beta}_{SLR}$	$SE_{\hat{\beta}_{SLR}}$	**t-stat**	**p-value**
Carb	262.14	22.21	11.80	<2e-16
Protein	848.62	83.512	10.16	<2e-16
Weight	45.17	3.07	14.70	<2e-16

Table 4.3. Estimates for the Energy Balance Naive Model

X-Values	$\hat{\beta}_{SLR}$	$SE_{\hat{\beta}_{SLR}}$	**t-stat**	**p-value**
Carbs	181.08	48.53	3.73	0.0003
Protein	987.42	182.51	5.41	2.1e-07
Weight	-6.36	6.71	-0.95	0.34

In Chapter 3, we discussed the assumptions that we made about the model, highlighting the visual diagnostic tools that we used to assess these assumptions. While the assumptions were not perfectly met, the plots only indicated some concern. One of the challenges we encountered was the small sample size of the Brown et al. (2017) dataset. Despite having seven days of data, we still only had 25 participants. This made it extremely difficult to have any more than 2-3 predictors in our models as it would decrease the degrees of freedom for that model. We remind the reader that degrees of freedom are the number of values in a statistic that are free to vary. Thus, for our models specifically, if we used too many predictors to explain our response, we would begin to affect the parameter estimates and test-statistics when analyzing the Brown et al. (2017) data. With a larger dataset, we would have a larger degrees of freedom and therefore we would have the ability to add more predictors (for example the micronutrients) to our model to help explain our response. Additionally, with a larger dataset, we would expect the standard errors of our parameter estimates to be smaller and therefore increase the precision of the estimates.

As discussed in Chapter 2, we assume that the days of collected data are independent for each dancer. In other words, the nutrient intake of a dancer for one day does not influence what they ingest on the following day. Normally, to avoid violating this assumption, the data are collected a few days apart. However, in the Brown et al. (2017) dataset, there are seven days of consecutive data. While the start of the data collection varied by day of the week for each dancer, there is still some concern that an association between the days of nutrient intake exists which could influence the parameter estimates for our models. To check this assumption, we created an order plot for each predictor by plotting the unobservable value for the seven days of data for each individual. We estimate $\hat{u}_{ij}$ by $X_{ij} - \bar{x}_i$, where $\bar{x}_i$ is an estimate for x_i and calculated as an overall average for individual i. If daily intakes for an individual are independent, we would expect there to be no patterns in the plot, meaning that the points are randomly disbursed around zero. We looked at the order plots for every individual for each potential predictor and did not see anything that was too concerning. The order plots for protein and carbohydrates of a few individuals in the dataset are shown in Section 5.2.3 of the Appendix.

As mentioned earlier, we developed a function that estimates the regression coefficients for any size dataset (see Appendix 5.3 for details). With this function, in future work, we could analyze the relationship of the body composition, energy levels, and dietary intake of dancers in a larger study. For example, we briefly began exploring similar relationships with data from the 2003-2004 National Health and Nutrition Examination Survey (NHANES) – a stratified, multistage survey that collects dietary and health data from thousands of individuals in the United States. Since dietary data are collected for 1-2 days, each at least 3 days apart, the data may reasonably be considered independent. Although these data do not have information on pre-professional collegiate level contemporary dancers specifically, we could focus on a different dancer population – individuals that chose dance as a recreational sport.

Results of a preliminary analysis of these data, using the same models as defined in Section 3.2 for the Brown et al. (2017) dataset, may be found in Tables 4.4 and 4.5. Note, we were unable to recreate the model with energy balance as a response, since the NHANES dataset does not provide energy expenditure values. For both models, we found similar results for recreational dancers as we did for the collegiate level, contemporary dancers in the Brown et al. (2017) study. We found it intriguing that we saw the same results using a larger, more comprehensive dataset. While these preliminary results are helpful in understanding the recreational dancer population, we still would need to conduct further research to both assess model fit and account for the unequal probability of selection of participants by incorporating survey weights for the estimates. Further, given the larger sample size, we would like to explore models with more predictors. We would also be interested in comparing how various nutrients impact FFM and TEI for individuals who dance recreationally versus those who do not.

Table 4.4. Estimates for NHANES Fat Free Mass Model

X-Values	β_{SLR}	$SE_{\beta_{SLR}}$	$\hat{\beta}_{MEM}$	$SE_{\hat{\beta}_{MEM}}$	**t-stat**	**95% CI**
Carb	-2.18	0.25	-1.86	3.29	-0.57	(-8.33, 4.61)
Protein	1.04	0.85	-16.49	12.55	-1.31	(-41.15, 8.17)

Table 4.5. Estimates for NHANES Energy Intake Model

X-Values	β_{SLR}	$SE_{\beta_{SLR}}$	$\hat{\beta}_{MEM}$	$SE_{\hat{\beta}_{MEM}}$	**t-stat**	**95% CI**
Carb	281.42	12.59	438.52	111.66	3.93	(219.30, 657.73)
Protein	676.58	41.08	107.79	349.21	0.31	(-577.81, 793.39)
Weight	24.14	0.97	25.43	5.21	4.88	(15.20, 35.66)

Despite the challenges of working with a dataset that had a small sample size and initial concerns around independence, the research discussed within this paper is novel in its approach to understanding a dancer's nutrition in relation to their FFM, TEI, and EB. As discussed in Chapter 1, collegiate level dancers often struggle with the pressures of maintaining a particular body image. As a result, many dancers have a decreased nutrient intake and subsequently suffer from various deficiencies and injuries. Further, studies that seek to better understand the nutrition of collegiate level dancers are not as frequently conducted in comparison to similar studies for athletes in other sports. Thus, the goal of our research was to better understand the factors that influence a dancer's dietary intake and what this could potentially mean for their overall health. In this work, we described the relationship between the body composition and energy levels with the nutrition of pre-professional collegiate level contemporary dancers. To account for the nuisance day-to-day variability when estimating long-term average intake, we propose a measurement error linear regression model. While other researchers have studied the nutrition of dancers before (Brown et al., 2017; Frost Brown et al., 2017; Friesen et al., 2011; Lim et al., 2015; Yannakoulia et al., 2000), they fail to account for the noisiness of the observed data. Our work has made positive steps towards better understanding the nutrition habits of contemporary dancers at the collegiate level. While we do not address the impact that micronutrients have on the dancer's body composition and nutritional intake, we provide a clear explanation of the relationship between protein and carbohydrates with fat free mass, total energy intake, and energy balance. We hope that this research can be used by dancers at the collegiate level to gain a better understanding of the nutritional factors that influence their body composition and energy levels.

CHAPTER 5

Appendix

5.1. Simple Linear Regression Model: Naive Estimators

In this section, we discuss the naive estimators for simple linear regression. We briefly derive them here to provide a clearer understanding of where the estimators come from as well as make note of the connection between the naive regression model estimators and the measurement error regression model estimators.

We define the simple linear regression model by,

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \tag{5.1}$$

for each observation $i = 1, \ldots, n$. The estimators for our regression coefficients are found by minimizing the sum of the squared errors (SSE): $\sum \epsilon_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$, where y_i is our observed value and $\hat{y}_i$ is our predicted value.

To minimize the SSE, we set the first derivative of the SEE, with respect to the intercept, equal to zero:

$$
\begin{aligned}
\frac{d}{d\beta_0}\left(\sum(y_i - (\beta_0 + \beta_1 x_i))^2\right) &= 0 \\
\sum\left(\frac{d}{d\beta_0}(y_i - (\beta_0 + \beta_1 x_i)\right)^2 &= 0 \\
2\sum(y_i - (\beta_0 + \beta_1 x_i))(-1) &= 0 \\
\sum(y_i - (\beta_0 + \beta_1 x_i)) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\beta_0 - \sum_{i=1}^{n}\beta_1 x_i &= 0 \\
n\bar{y} - n\beta_0 - n\beta_1\bar{x} &= 0 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x}. \qquad (5.2)
\end{aligned}
$$

Similarly, we can find the estimator for β_1:

$$
\begin{aligned}
\frac{d}{d\beta_1}\left(\sum(y_i - (\beta_0 + \beta_1 x_i))^2\right) &= 0 \\
\sum\left(\frac{d}{d\beta_1}(y_i - (\beta_0 + \beta_1 x_i))\right)^2 &= 0 \\
2\sum(y_i - (\beta_0 + \beta_1 x_i))(-x_i) &= 0 \\
\sum y_i x_i - \beta_0\sum x_i - \beta_1\sum x_i^2 &= 0 \\
\sum y_i x_i - (\bar{y} - \beta_1\bar{x})\sum x_i - \beta_1\sum x_i^2 &= 0 \\
\sum y_i x_i - \bar{y}\sum x_i + \beta_1\bar{x}\sum x_i - \beta_1\sum x_i^2 &= 0 \\
\sum x_i y_i - n\bar{x}\bar{y} &= \beta_1(\sum x_i^2 - n\bar{x}^2) \\
\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \qquad (5.3)
\end{aligned}
$$

Looking back in Chapter 2, there is a clear relationship between the calculation of $\hat{\beta}_0$ in both the naive and measurement error regression models (see equations (5.2) and (2.16)). Additionally, we see that $\hat{\beta}_1$ in the naive regression model in (5.3) and m_{XX} in the measurement error regression model in (2.15) are essentially the same. However, in the measurement error model, we account for the within-person variability to account for the noisiness of the observed data.

5.2. Additional Model Diagnostics

In this section, we provide more details regarding the assumptions for the errors of the measurement error model used in our application described in Chapter 3. As noted in Section 3.2.1, we focused on the macronutrients as predictors in our models as the micronutrients did not meet the assumptions for the our measurement errors. Overall, the assumption plots for the micronutrients often showed patterns or trends, thus violating the assumptions that the measurement error variances of our predictors are independent of the x_i's for predictors measured with and without error. Additionally, the Normal Quantile-Quantile plots for the micronutrients did not follow the line closely, especially at the tails, and therefore violated the assumption that the measurement errors should approximately follow a normal distribution. Further, many of the residuals plots for models with micronutrients as predictors lacked random scatter. As an example of these violations, in Figure 5.1, we provide the residual plot and Normal Quantile-Quantile plot for iron in a model where we predict TEI from iron and vitamin D.

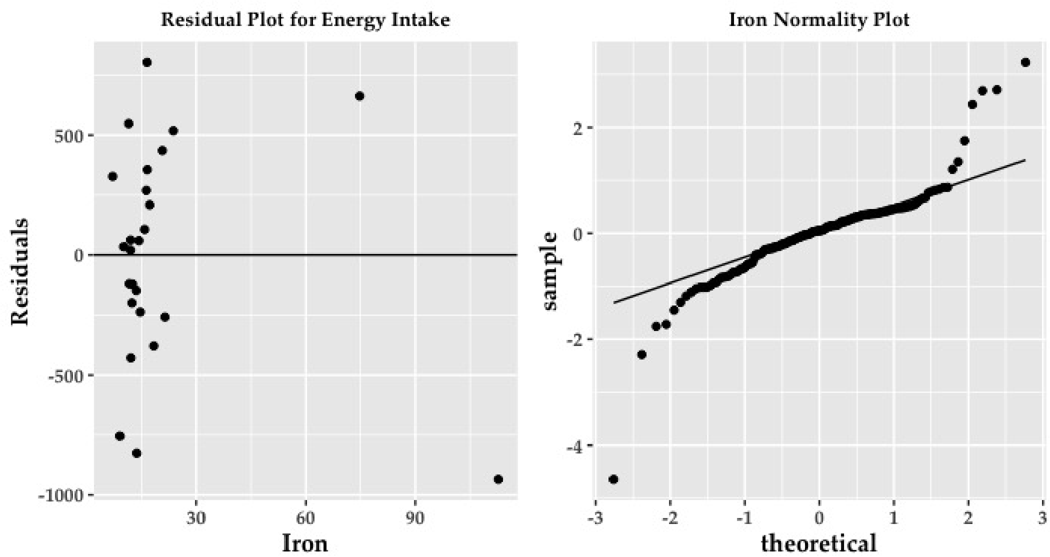

Figure 5.1. Residual and Normal Quantile-Quantile plot for iron as a predictor with vitamin D for TEI.

Given the evidence that the model assumptions were violated, and the emphasis on macronutrients from prior research in the dance field, we focused our work to examine models with only the macronutrients as predictors. We now provide some details around the models described in Sections 3.2.1.1-3.2.1.3.

5.2.1. Visual Diagnostics for the Measurement Errors. To check whether the measurement error variances can reasonably be assumed to be independent of the x_i's, we plot the sample standard deviation for X_{ij} for each individual i against the respective sample mean, $\bar{X}_i$; the plots for protein and carbohydrates are in Figure 5.2. In these plots, there appear to be a few outliers for larger intake values of protein and carbohydrates. Due to the small sample size of the Brown et al. (2017) study, we are not surprised to see some outliers and larger variability in our observations.

Given that there is no strong trend overall, we are willing to assume that the measurement error variance is independent of the x_i's.

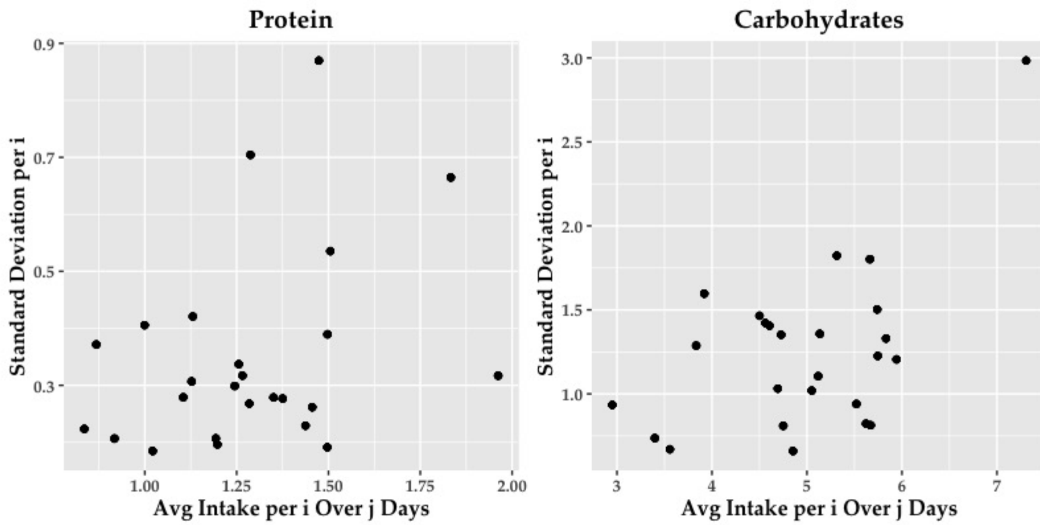

FIGURE 5.2. Checking Model Assumption 1: Sample standard deviation for X_{ij} for each individual i against the respective sample mean, $\bar{X}_i$.

Similarly, to check whether the measurement error variance is independent of the predictors measured without error, we plot the standard deviation of X_{ij} for each individual i against the sample mean, $\bar{X}_i$; the plot for weight is in Figure 5.3. Again, we see that there are a few outliers at the smaller weight values, but given that there is not a strong trend overall, we are willing to assume that the measurement error variance is independent is of the predictors measured without error.

To check whether we can reasonably assumed that the errors follow a normal distribution, we use a Normal Quantile-Quantile Plots for the measurement errors, u_{ij} (estimated by $\hat{u}_{ij} = X_{ij} - \bar{X}_i$). If the points follow closely to the line, it is fair to assume that the data is normally distributed. As seen in Figure 5.4, most of the $\hat{u}_{ij}$ values follow relatively closely to the line, except at the tails. Recalling the outliers from the previous plots, this deviation at the tails is somewhat expected. While we would like to have the data follow along the line more closely, for estimation purposes, we will continue to assume that the measurement errors are normally distributed.

5.2.2. Residual Plots. To check for nonlinearity in the regression and lack of homogeneity of the error variances, we look at the residual plots for each predictor in the three models we built (as seen in Figures 5.5 - 5.7)

In Figures 5.5-5.7, we look at the residual plots for the predictors of the FFM, TEI, and EB models described in Section 3.2.1.1. Overall, we notice that the residual plots are fairly randomly-scattered. While there appear to be a few outliers, we are willing to assume that each of the models accurately describe the relationship between their predictors and response.

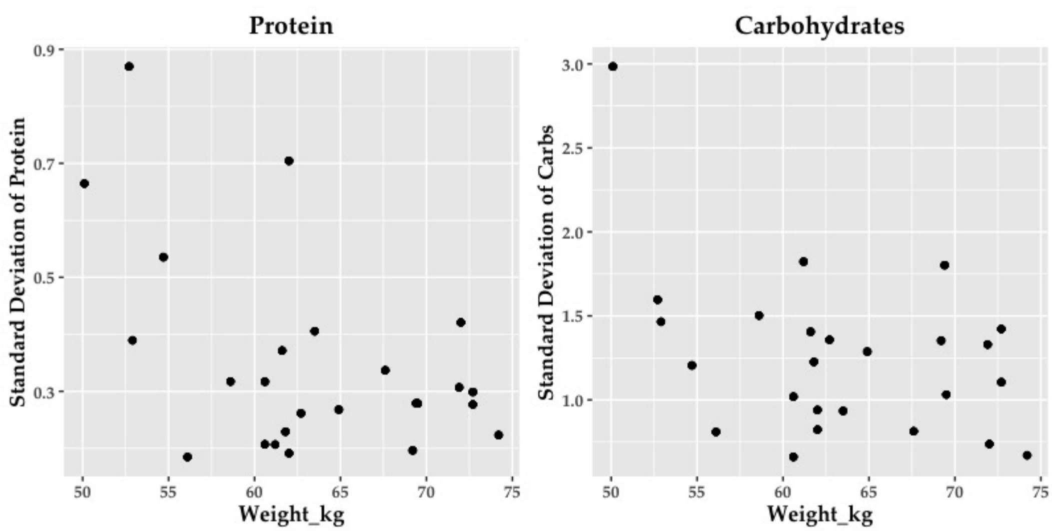

Figure 5.3. Checking Model Assumption 2: Sample standard deviation for X_{ij} for each individual i against the sample mean, for weight.

5.2.3. Order Plots. As discussed in Chapter 4, we assume that the days of collected data are independent for each dancer. In Figures 5.8 and 5.9,

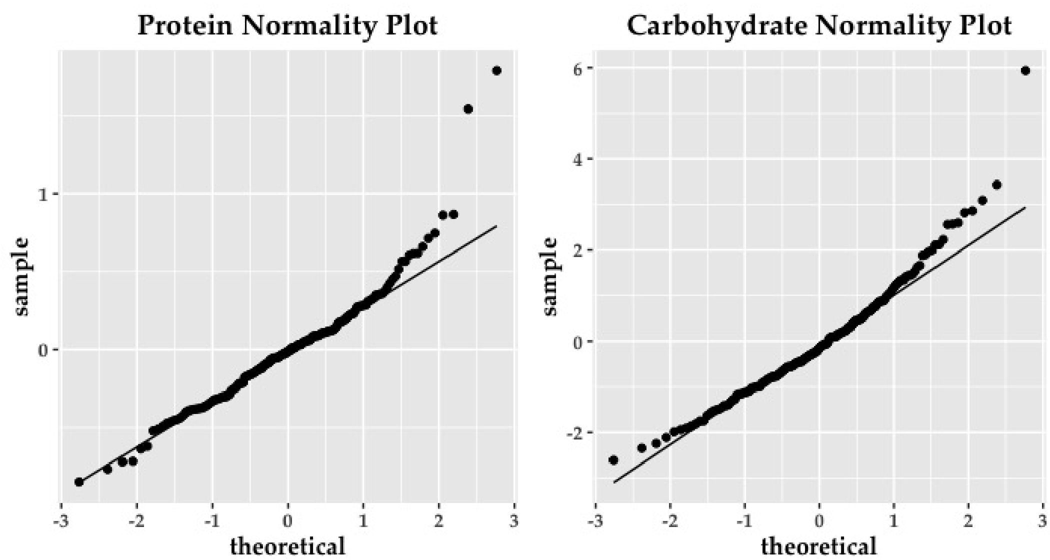

FIGURE 5.4. Checking Assumption 3: Normality QQ Plots for the measurement errors, u_{ij}.

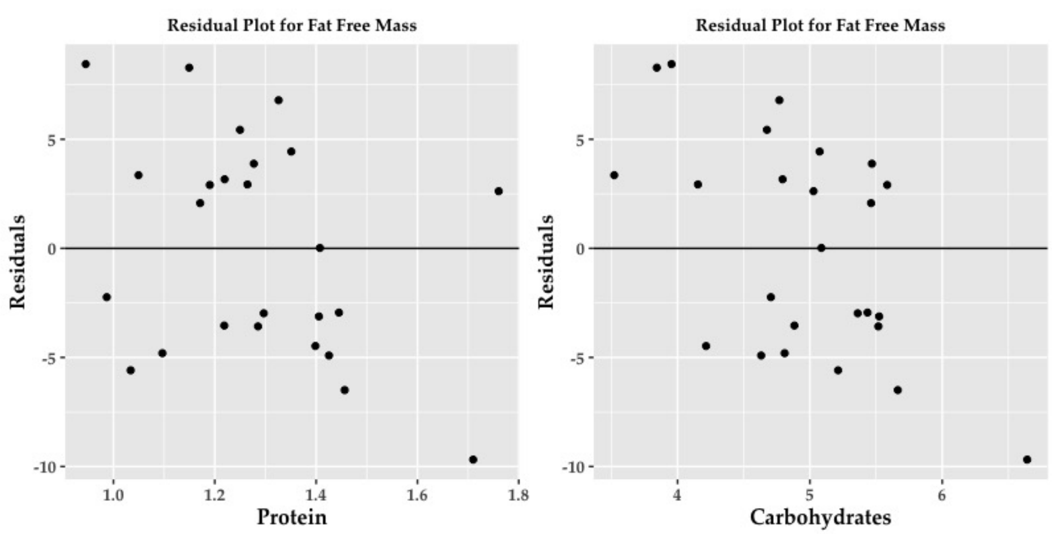

FIGURE 5.5. Residual plots for the chosen predictors of fat free mass: protein and carbohydrates.

we show the order plots to check this assumption for the protein and carbohydrates intake of the dancers in the Brown et al. (2017). We plot the $\hat{u}_{ij}$ (estimated by $\hat{u}_{ij} = X_{ij} - \bar{X}_i$) for protein and carbohydrates for each dancer across the seven days of nutrient intake. We look at participant 17 through participant 21's intake of protein and participant 6 through participant 10's intake of carbohydrates in the order plots. In both plots, we see that the points are relatively scattered around zero for each participant, with little to no pattern. Although not a perfect measure, the minimal pattern in the

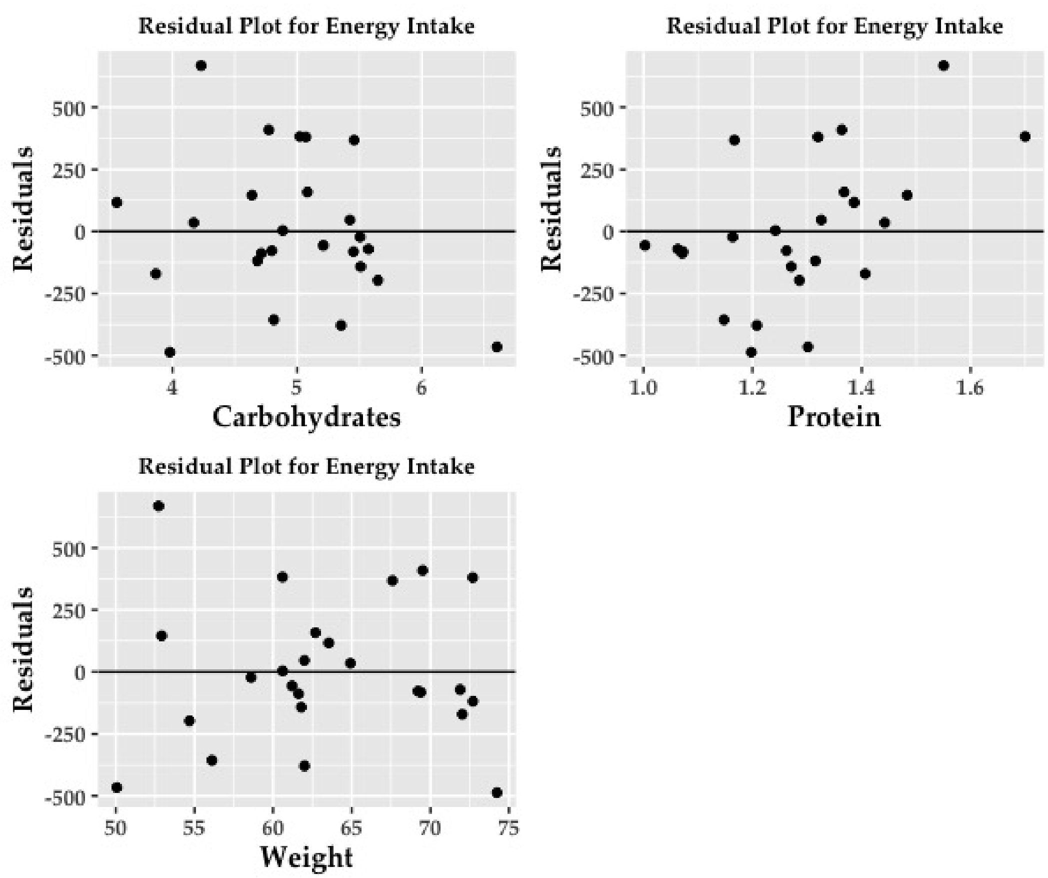

FIGURE 5.6. Residual plots for the chosen predictors of total energy intake: protein, carbohydrates, and weight.

plots help support that we can reasonably assume that the dancer's days of intake are independent.

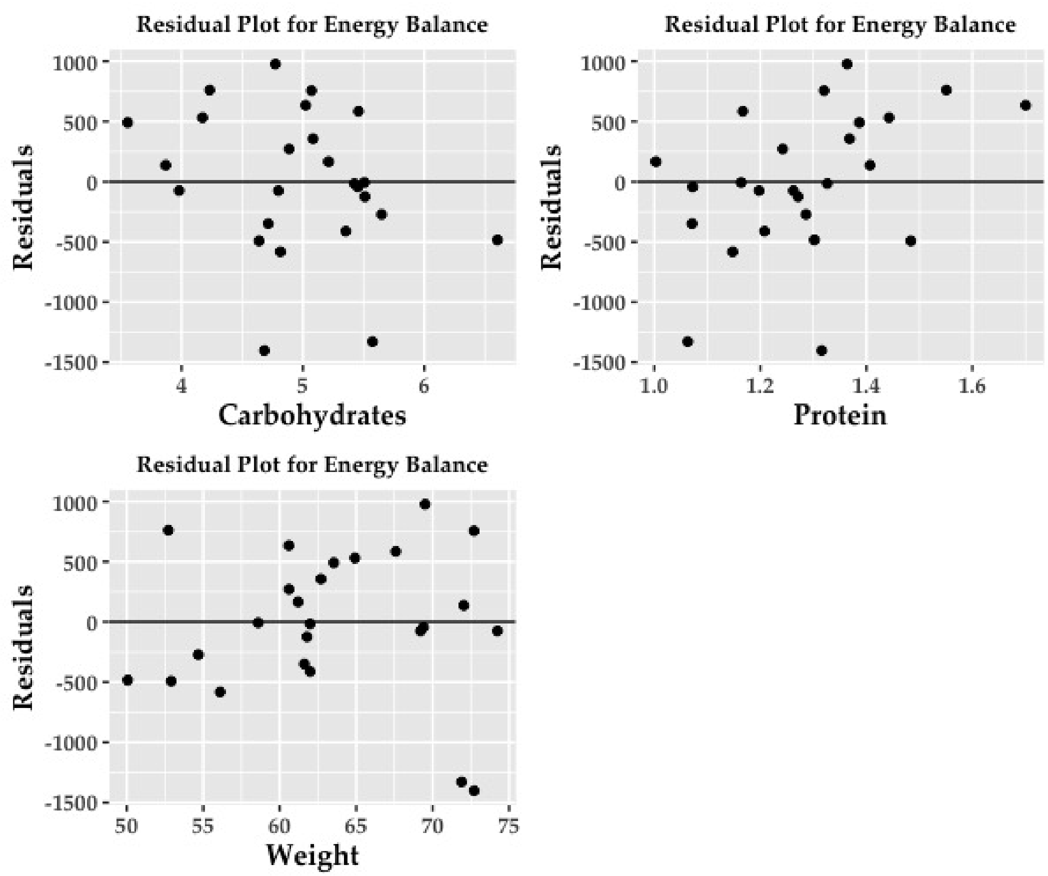

FIGURE 5.7. Residual plots for the chosen predictors of energy balance: protein, carbohydrates, and weight.

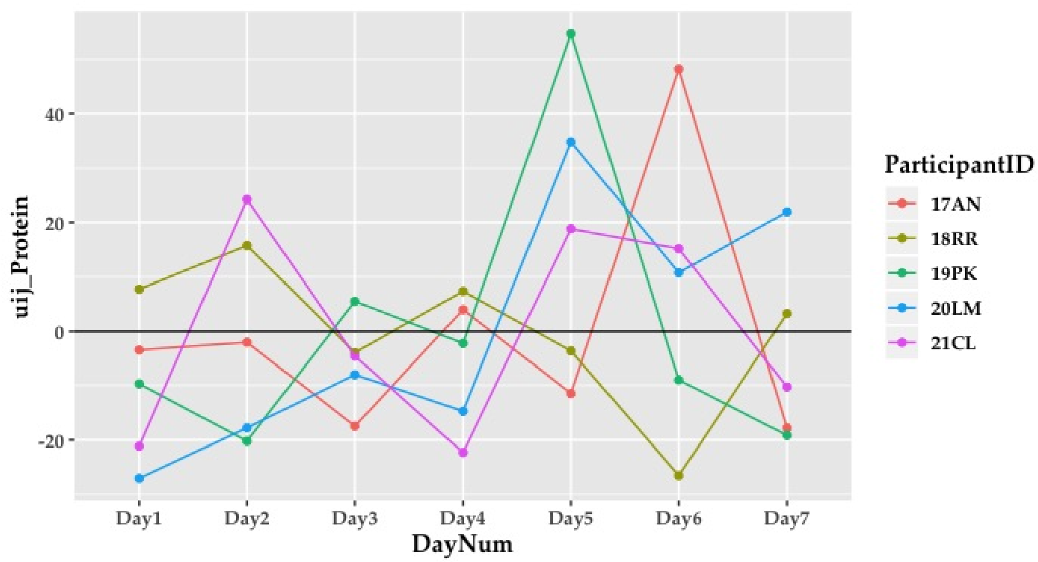

Figure 5.8. Order plot of participants 17 through 21's protein intake.

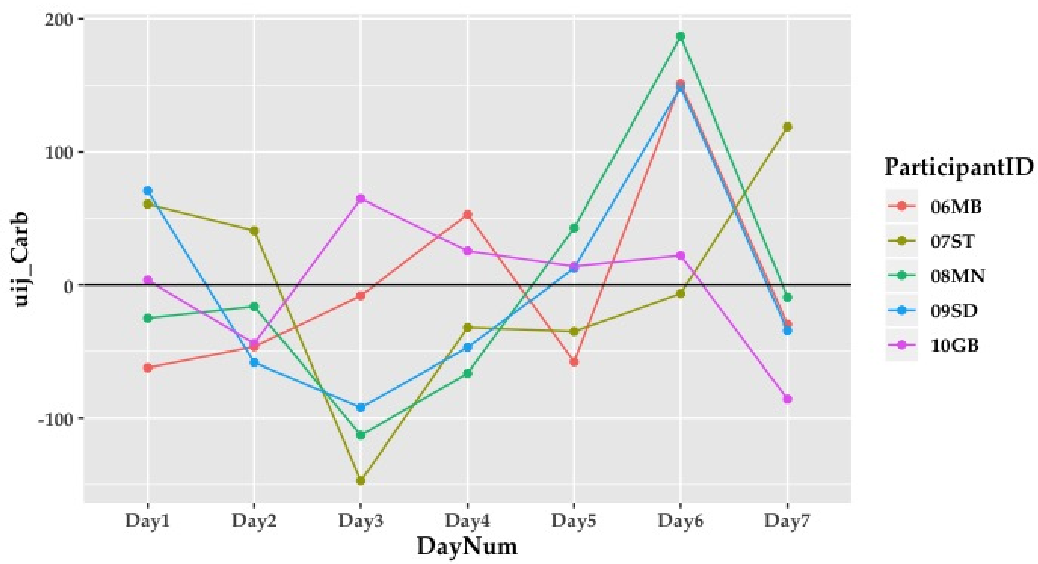

Figure 5.9. Order plot of participants 6 through 10's carbohydrate intake.

5.3. The R Code

5.3.1. Measurement Error Regression Model Function. In this section, we share the R code used to calculate the unbiased parameter estimates for the measurement error regression model outlined in Sections 2.1.2 and 2.2.2. As inputs, the function requires an $nmx1$ vector for the response, an $nmxp$ matrix for the predictors, and an $nmx1$ vector of unique identifiers. The function is written to provide parameter estimates for a single response and one or multiple predictors. The function returns a $px1$ vector for $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$, a single value for $\hat{\beta}_0$, and a pxp $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_{\mathbf{1}})$ matrix.

```
MEM_functionMult <- function(y2=NULL, x2=NULL, id2=NULL){
  #Adding all inputs to a dataframe
  dat2<-x2
  dat2$y<-y2
  dat2$id2 <- as.numeric(as.factor(id2))

  #Transforming x,y to matrices to use in for loops later
  x2<-as.matrix(x2)
  y2<-as.matrix(y2)

  #Calculating number of days, number of participants, number
     of predictors, number of responses
  numdaystable <- table(dat2$id2)
  numdays <- numdaystable[[1]]
  numpart <- length(unique(id2))
  numpred <- ncol(x2)
  numresp <- ncol(y2)

  #Calculation for the average across the total number of days
      for each individual
  x_bardotmatrix<-matrix(ncol = numpred, nrow = numpart)
```

```
for(j in 1:numpred){
  for(i in 1:length(unique(id2))){
    x_bardotmatrix[i,j] <- mean(x2[as.numeric(as.factor(id2))
        ==i,j])
  }
}

#Calculation for average of the averages above, ie overall
   average per predictor
x_dotdotmatrix<-matrix(ncol = numpred, nrow = 1)
for (k in 1:numpred){
  x_dotdotmatrix[,k] <- mean(x_bardotmatrix[,k])
}

#Calculation for average response for each individual
y_bardotmatrix<-matrix(ncol = numresp, nrow = numpart)
for(j in 1:numresp){
  for(i in 1:length(unique(id2))){
    y_bardotmatrix[i,j] <- mean(y2[as.numeric(as.factor(id2))
        ==i,j])
  }
}

#Calculation for overall average for each response
y_dotdotmatrix<-matrix(ncol = numresp, nrow = 1)
for (k in 1:numresp){
  y_dotdotmatrix[,k] <- mean(y_bardotmatrix[,k])
}

#Calculating the between person or overall variability for
   each predictor
Mxx_sdxt <- matrix(nrow=numpred, ncol = numpart)
Mxx_sdx <- matrix(nrow=numpart, ncol = numpred)
Mxx_sdmatrix<- matrix(nrow = numpred, ncol = numpred)
```

```
for (j in 1:numpred){
  for (i in 1:numpart){
    Mxx_sdx[i,j] <- x_bardotmatrix[i,j] - x_dotdotmatrix[,j]
    Mxx_sdxt <- t(Mxx_sdx)
    Mxx_sdmatrix <- Mxx_sdxt %*% Mxx_sdx
  }
}
M_xxmatrix <- (1/(numpart-1))*Mxx_sdmatrix



#Calculating the between person or overall variability
   between the predictor and response
Mxy_sdy <- matrix(nrow=numpart, ncol = numresp)
Mxy_sdmatrix<- matrix(nrow = numpred, ncol = numpred)
for (j in 1:ncol(Mxy_sdy)){
  for (i in 1:numpart){
      Mxy_sdy[i,j] <- y_bardotmatrix[i,j] - y_dotdotmatrix[,j
          ]
      Mxx_sdxt <- t(Mxx_sdx)
      Mxy_sdmatrix <- Mxx_sdxt %*% Mxy_sdy
  }
}
M_xymatrix <- (1/(numpart-1))*Mxy_sdmatrix

#Calculation for the within person or day to day variability
     for each predictor
siguu_i<-array(dim = c(numpred, numpred, numpart))
for (i in 1:numpart){
  #Compute the variance for each individual's predictor value
  siguu_i[,,i]<-var(x2[dat2$id2 ==i,])
}
#Average across first and second dimensions
Sig_uu <- apply(siguu_i, c(1,2), mean, na.rm=T)
```

```
#Calculation for the within person or day to day variability
    for the response
sigww_i<-array(dim = c(numresp, numresp, numpart))
for (i in 1:numpart){
  sigww_i[,,i]<-var(y2[dat2$id2 ==i,])
}
Sig_ww <- apply(sigww_i, c(1,2), mean, na.rm=T)

#Calculation for the within person or day to day variability
    for the covariance of the predictor and reponse
sigwu_i<-array(dim=c(numpred,numresp,numpart))
for (i in 1:numpart){
  sigwu_i[,,i] <- cov(y2[dat2$id2 == i], x2[dat2$id2 ==i,])
}
Sig_wu <- apply(sigwu_i, c(1,2), mean, na.rm=T)

#Calculating the matrix of beta values and beta_0
beta_matrix<- ginv(M_xxmatrix-Sig_uu)%*%(M_xymatrix-Sig_wu)
beta_0mult<- y_dotdotmatrix - (x_dotdotmatrix %*%
   beta_matrix)

#Calculating the variance of error in the equation
s_vvmult<-(1/(numpart-numpred))*((t(y_bardotmatrix - rep(
   beta_0mult,numpart) - (x_bardotmatrix%*%beta_matrix)))%*%
    (y_bardotmatrix - rep(beta_0mult,numpart) - (
   x_bardotmatrix%*%beta_matrix)))
s_rrmult <- Sig_ww - (2*(t(beta_matrix)%*%(Sig_wu))) + (t(
   beta_matrix)%*%(Sig_uu)%*%(beta_matrix))
s_vvmult<-as.numeric(s_vvmult)
s_rrmult<-as.numeric(s_rrmult)
Sig_qq<-s_vvmult - s_rrmult

#Error in e is independent of x and u
Sig_eu <-0
```

```
  #Calculation for the variance of error in the response
  Sig_ee <- M_yymatrix - 2*M_xymatrix%*%beta_matrix + t(
     beta_matrix)%*%M_xxmatrix%*%beta_matrix + 2*Sig_eu%*%
     beta_matrix - t(beta_matrix)%*%Sig_uu%*%beta_matrix

  #Calculating for the variance for the beta estimates
  var_betamult<-
    (1/numpart)*((ginv(M_xxmatrix-Sig_uu)*s_vvmult)

               + (ginv(M_xxmatrix-Sig_uu)%*%(Sig_uu*s_vvmult+((
                  Sig_wu - Sig_uu%*%beta_matrix)%*%(t(Sig_wu -
                   Sig_uu%*%beta_matrix))))%*%ginv(M_xxmatrix-
                  Sig_uu))) +

    (1/(numpart-numpred))*(ginv(M_xxmatrix-Sig_uu)) %*% (Sig_uu
       *s_rrmult +((Sig_wu - Sig_uu%*%beta_matrix)%*%(t(Sig_wu
        - Sig_uu%*%beta_matrix)))) %*%(ginv(M_xxmatrix-Sig_uu)
       )



  res <- list(x_bardotmatrix = x_bardotmatrix, x_dotdotmatrix
     = x_dotdotmatrix, y_bardotmatrix= y_bardotmatrix,
     y_dotdotmatrix = y_dotdotmatrix, M_xx = M_xxmatrix, M_xy
     = M_xymatrix, Sig_uu = Sig_uu, Sig_wu = Sig_wu, Sig_ww =
     Sig_ww, beta_matrix = beta_matrix, beta_0mult =
     beta_0mult, s_vvmult=s_vvmult, s_rrmult= s_rrmult, Sig_qq
      = Sig_qq, Sig_ee = Sig_ee, var_betamult = var_betamult,
     numpart = numpart, numpred = numpred)
  return(res)

}
```

5.3.2. Test Statistic, Standard Error, Confidence Interval Functions. In this section, we share the functions we built in R that calculate the test statistic, the standard error, and the confidence interval for each of the estimated coefficients in the measurement error regression model.

The test statistic function requires an $px1$ vector for $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ and an pxp $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_{\mathbf{1}})$ matrix and returns the test statistic for each parameter estimate. The standard error function requires an pxp $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_{\mathbf{1}})$ matrix and returns the standard error for each parameter estimate. Additionally, the confidence interval function requires an $px1$ vector for $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$, an pxp $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_{\mathbf{1}})$ matrix, the number of participants in the study, and the confidence level. The function returns the confidence level for each parameter estimate.

```
#Calculating the t-statistic for each beta value
tstat_function<- function(betamatrix=NULL, var_betamatrix=NULL
   ){
  diag_variance<- diag(var_betamatrix)
  tstat_matrix <- matrix(nrow = nrow(betamatrix), ncol = 1)
  for (i in 1:nrow(betamatrix)){
    tstat_matrix[i,1] <- betamatrix[i,1]/ sqrt(diag_variance[i
       ])
  }
  return(tstat_matrix)
}

#Function calculating the standard error for each beta value
SE_function <- function(var_betamatrix = NULL){
  diag_variance<- diag(var_betamatrix)
  SE <-matrix(nrow = nrow(var_betamatrix), 1)
  for (i in 1:nrow(var_betamatrix)){
    SE[i,] <- sqrt(diag_variance[i])
  }
```

```
  return(SE)
}

#Calculating the confidence interval for each beta value
CI_function<- function(betamatrix=NULL, var_betamatrix=NULL,
    numpart = NULL, CIlevel = NULL){
  diag_variance<- diag(var_betamatrix)
  CI_matrix <- array(dim = c(2, 1, nrow = nrow(betamatrix)))
  for (i in 1:nrow(betamatrix)){
    left<-betamatrix[i,1]-qt(CIlevel,df=numpart-nrow(betamatrix
        ))*sqrt(diag_variance[i])
    right<-betamatrix[i,1]+qt(CIlevel,df=numpart-nrow(
        betamatrix))*sqrt(diag_variance[i])
    CI<- matrix(c(left, right))
    CI_matrix[,,i] <- CI
  }
  return(CI_matrix)
}
```

Bibliography

ACSM. Joint position statement: nutrition and athletic performance. american college of sports medicine, american dietetic association, and dietitians of canada. Medicine and Science in Sports and Exercise, 32(12): 2130, 2000.

Dimitri P Bertsekas and John N Tsitsiklis. Introduction to probability vol. 1. 2002.

Meghan A Brown, Glyn Howatson, Edel Quin, Emma Redding, and Emma J Stevenson. Energy intake and energy expenditure of pre-professional female contemporary dancers. PloS one, 12(2):e0171998, 2017.

John P Buonaccorsi. Measurement error: models, methods, and applications. Chapman and Hall/CRC, 2010.

Bill Campbell, Richard B Kreider, Tim Ziegenfuss, Paul La Bounty, Mike Roberts, Darren Burke, Jamie Landis, Hector Lopez, and Jose Antonio. International society of sports nutrition position stand: protein and exercise. Journal of the International Society of Sports Nutrition, 4(1):8, 2007.

Raymond J Carroll, David Ruppert, Leonard A Stefanski, and Ciprian M Crainiceanu. Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC, 2006.

Brynne Chandler. A healthy meal plan for dancers, 2018. URL https://www.livestrong.com/article/162703-healthy-diets-for-dancers/.

P Clarkson. Nutrition fact sheet: fueling the dancer. Resource Paper. International Association for Dance Medicine, 2005.

Brenna Curley. Nonlinear models with measurement error: Application to vitamin d. 2017. doi: 10.31274/etd-180810-5746.

Heather B Del Valle, Ann L Yaktine, Christine L Taylor, A Catharine Ross, et al. Dietary reference intakes for calcium and vitamin D. National Academies Press, 2011.

Karlie J Friesen, Ralph Rozenek, Karen Clippinger, Kathy Gunter, Albert C Russo, and Susan E Sklar. Bone mineral density and body composition of collegiate modern dancers. Journal of Dance Medicine & Science, 15(1): 31–36, 2011.

Ann Frost Brown, Christopher William Bach, Giuliana De Almeida, Shawn Leonard, Tom Welsh, and Michael James Ormsbee. Body composition and performance capabilities based on protein intake in collegiate dancers: A pilot study. Journal of Sports Science, 5:189–197, 2017.

Wayne A Fuller. Measurement error models, volume 305. John Wiley & Sons, 2009.

Health Engine. Energy expenditure: How the body burns calories. HealthEngine, Mar 2019. URL https://healthengine.com.au/info/energy-expenditure-how-the-body-burns-calories.

Bibi Mushirah Jodhun, Dhandevi Pem, and Rajesh Jeewon. A systematic review of factors affecting energy intake of adolescent girls. African Health Sciences, 16(4):910, Jul 2017. doi: 10.4314/ahs.v16i4.5.

Maria L Joseph and Alicia Carriquiry. A measurement error approach to assess the association between dietary diversity, nutrient intake, and mean probability of adequacy. The Journal of nutrition, 140(11):2094S–2101S, 2010.

James W Krieger, Harry S Sitren, Michael J Daniels, and Bobbi Langkamp-Henken. Effects of variation in protein and carbohydrate intake on body mass and composition during energy restriction: a meta-regression. The American journal of clinical nutrition, 83(2):260–274, 2006.

Helen Laws and Joanna Apps. Fit to Dance 2: Report of the second national inquiry into dancers' health and injury in the UK. Dance UK, 2005.

Se-Na Lim, Joo-Hee Chai, Jong Kook Song, Myong-Won Seo, and Hyun-Bae Kim. Comparison of nutritional intake, body composition, bone mineral density, and isokinetic strength in collegiate female dancers. Journal of exercise rehabilitation, 11(6):356, 2015.

Paul Pacy, Magita Khalouha, and Yiannis Koutedakis. Body composition, weight control and nutrition in dancers. Dance Research, 14(2):93–105, 1996.

Nancy R Rodriguez, NM Marco Di, and Susie Langley. American college of sports medicine position stand. nutrition and athletic performance. Medicine and science in sports and exercise, 41(3):709–731, 2009.

Mónica Sousa, Pedro Carvalho, Pedro Moreira, and Vítor Hugo Teixeira. Nutrition and nutritional issues for dancers. Med Probl Perform Art, 28

(3):119–23, 2013.

Emily Twitchett, Manuela Angioi, Yiannis Koutedakis, and Matthew Wyon. The demands of a working day among female professional ballet dancers. Journal of dance medicine & science, 14(4):127–132, 2010.

Mary Yannakoulia, Antonios Keramopoulos, Nikolaos Tsakalakos, and M Antonia-Leda. Body composition in dancers: the bioelectrical impedance method. Medicine and science in sports and exercise, 32(1): 228–234, 2000.