Multivariate Multiple Regression with Applications to Powerlifting Data

Cassandra Quick
University of Minnesota Duluth
Department of Mathematics and Statistics

Advisor: Kang James

August 2013

Acknowledgements

I would like to express my deepest appreciation to Dr. Kang James for being a patient and wonderfully supportive advisor. Her encouragement and helpful critiques of my work are tremendously valued. I would also like to thank Dr. Yongcheng Qi and Dr. Marshall Hampton for serving on my committee. I am particularly grateful for the constant reassurance and endless support provided by my fellow graduate students. My special thanks should also be extended to the rest of the staff and faculty from the Department of Mathematics and Statistics at University of Minnesota Duluth. Collectively, they offered daily support, guidance, and knowledge that made for a great learning environment. Finally, I would like to acknowledge the continuous support and encouragement I have received from my closest friends and family throughout the progression of this project.

Abstract

Multivariate multiple linear regression has the potential to be a very powerful tool in many fields of work and research. This method is used when we have a problem consisting of two or more predictor variables *and* two or more response variables. In this paper, the multivariate multiple linear regression model is built using the foundations and properties of the multiple linear regression model. The multiple linear regression model concerns predicting or explaining values of *one* response variable based on values of more than one predictor. After building the multivariate model and developing its properties, an application problem relating to athletics will be addressed. Powerlifting is a relatively new sport and has been gaining popularity quickly over the past 15 years. There are no qualifying standards in place for a national meet in a certain powerlifting division. In order to ensure that the meet is competitive, the multivariate model and simultaneous prediction intervals can aid in establishing qualifying standards for this national meet.

Table of Contents

Acknowledgements **i**

Abstract **ii**

1 Introduction **1**

2 Multivariate Multiple Regression Model **4**

2.1 Model Construction *4*

2.2 Some Useful Properties *8*

3 Parameter Estimation **10**

3.1 Estimating $\boldsymbol{\beta}$ *10*

3.2 $\widehat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{\varepsilon}}$ *13*

3.3 Some Useful Properties *13*

3.4 Estimating $\boldsymbol{\Sigma}$ *17*

4 Making Predictions **20**

5 Application to Powerlifting Data **23**

5.1 The Problem *23*

5.2 The Data *25*

5.3 Estimated Parameters *27*

5.4 Predicting Means *27*

5.5 Prediction Intervals *32*

6 Final Conclusions **38**

7 Recommendations for Further Research **39**

8 References **41**

7 Appendix **42**

1 Introduction

Multivariate multiple linear regression is an extremely valuable concept and is applied practically in fields such as business, economics, politics, and medical research. The multivariate multiple regression model is an extension of the standard multiple linear regression model. Multiple linear regression concerns predicting or explaining values of *one* response variable based on values of a collection of two or more predictor variables. For example, we could explore the relationship between students' college GPAs (the *one* response variable) and their GPAs and attendance records in high school and ACT scores (the three predictor variables). Suppose that in addition to students' college GPAs, we are also interested in the number of years it takes them to graduate. Now we are concerned with predicting values of *two* response variables based on the values of the 3 predictors. We have a situation where we can use the multivariate multiple regression model. In multivariate linear regression, we take a collection of two or more predictor variables into account, and try to predict or explain values of *two or more* response variables. In this paper, the intricate multivariate multiple regression model is introduced in a clear manner, and its sampling properties are discussed and derived with precise detail. The estimation of model parameters and the discussion of simultaneous prediction intervals will also be featured in order to understand an application problem that will be introduced at the end of the paper.

A problem relating to athletics in which multivariate multiple regression can be applied will be presented.

In many athletic contests, participants are required to meet a predetermined qualifying standard in order to compete. Such qualifying criteria allow the most elite athletes to compete against each other in a specific event. A majority of athletic competitions that can implement the use of these qualifying standards have established marks that athletes strive to attain in order to be able to compete with the best in their sport. For example, if a woman long jumper from University of Minnesota Duluth wanted to compete in the 2013 NCAA Division II Outdoor Track and Field National Championships, she would need to jump at least 6.15 meters in any of the regular season competitions to qualify [8]. For track and field and other long-established sports, qualifying standards have been determined and are modified regularly to coincide with the overall performance trends set by athletes each year. Usually, improvements in athletes' performances occur collectively, year after year, due to better training regimens and new developments in equipment technology. These performance improvements are also occurring in the sport of powerlifting. While some powerlifting meets have qualifying standards that athletes try to reach in order to compete, there is an important powerlifting competition that occurs annually in which no qualifying standards have been implemented.

Emerging in the 1950's, the sport of powerlifting is relatively new [5]. There are two main types of powerlifting competitions: equipped (open) and unequipped (raw). In equipped meets, competitors are allowed to wear belts, supportive shirts, and special suits that allow them to lift more than if they weren't wearing any of the advantageous equipment. Lifters competing in raw meets aren't allowed to wear any of this special gear. This new concept of raw powerlifting has gained popularity quickly over the last fifteen years. However, it is still relatively new, so when it comes to organizing USA Powerifting's Raw Nationals (one of the biggest annual events for raw powerlifters), there is no limitation regarding who gets to compete. It would be a much more competitive meet if there were qualifying standards in place for lifters to strive to reach in order to compete.

Multivariate multiple linear regression can be used to make predictions about how much a person can lift in a certain powerlifting event depending on his or her body weight and age. We can also use a multivariate multiple regression model to help establish qualifying standards for the USAPL Raw National Meet by computing simultaneous prediction intervals for individual response variables based on a vector containing values of predictor variables. (Body weight and age are the predictors, and the amounts of weight lifted in three different events are the responses in this problem.) There is also plenty of room for further research on this topic, such as appropriately developing *better* age and weight class systems for Raw Nationals and determining the corresponding qualifying standards.

2 Multivariate Multiple Regression Model

2.1 Model Construction

We will introduce the multivariate multiple regression model by first considering the multiple linear regression model. In multiple linear regression, we are concerned with predicting values of one response variable based on a set of predictor variables. Table 2.1.1 represents a collection of multiple linear regression data. A collection of n observations, or trials, makes up the data set.

TABLE 2.1.1 Multiple Linear Regression Data

Observation, i	Response, y	Predictor, z_1	Predictor, z_2	$\cdots$	Predictor, z_r
1	y_1	z_{11}	z_{12}	$\cdots$	z_{1r}
2	y_2	z_{21}	z_{22}	$\cdots$	z_{2r}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	y_n	z_{n1}	z_{n2}	$\cdots$	z_{nr}

We can write a regression model for the i^{th} observation from the data set. The response for the i^{th} trial, y_i, can be written as a linear combination of the observed predictor variables, $z_{i1}, z_{i2}, \dots, z_{ir}$, with the constant β_0 term and a random error term, ε_i, added on [3].

$$y_i = [\beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \cdots + \beta_r z_{ir}] + \varepsilon_i$$

$$= [\beta_0 + \sum_{j=1}^{r} \beta_j z_{ij}] + \varepsilon_i$$

response on ith trial=[mean]+error

The model's constant term or intercept, β_0, can be interpreted as the mean for the response when the value of all of the predictor variables is zero. This intercept and the regression coefficients, $\beta_1, \dots, \beta_r$, are unknown parameters and can be estimated. We can interpret the values of the regression coefficients straightforwardly. The coefficient of the predictor variable in question represents the change in the mean response as that predictor changes by one unit, holding all other predictors constant. The mean of the i^{th} response is a linear function of the predictors, $z_{i1}, z_{i2}, \dots, z_{ir}$, and the unknown parameters $\beta_1, \dots,$ and β_r are their respective weights. The parameter estimation method for multiple linear regression and multivariate multiple regression will be discussed in the following section. The multiple linear regression model can also be written in matrix form.

$$\begin{bmatrix} y \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

We can simply write this model as $\boldsymbol{y} = \boldsymbol{Z\beta} + \boldsymbol{\varepsilon}$ [3].

When there is more than one response variable, we must use a multivariate multiple linear regression model. A data table containing multivariate multiple regression data would be organized like Table 2.1.2.

TABLE 2.1.2 Multivariate Multiple Regression Data

Observation, i	Response, Y_1	Response, Y_2	…	Response, Y_m	Predictor, z_1	Predictor, z_2	…	Predictor, z_r
1	Y_{11}	Y_{12}	$\cdots$	Y_{1m}	z_{11}	z_{12}	$\cdots$	z_{1r}
2	Y_{21}	Y_{22}	$\cdots$	Y_{2m}	z_{21}	z_{22}	$\cdots$	z_{2r}
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	Y_{n1}	Y_{n2}	$\cdots$	Y_{nm}	z_{n1}	z_{n2}	$\cdots$	z_{nr}

Now that we are considering a problem with multiple predictors *and* several response variables, we must be especially mindful of the notation. As an example for clarification, the Y_{21} entry in Table 2.2 represents the value of the first response variable (Y_1) observed on the second trial. The table contains m response variables and r predictor variables, and there were n trials performed (or n observations collected) for this experiment. Extending the multiple linear regression case to the multivariate multiple linear case, we can write a regression model for each response (there are m of them) on the i^{th} observation, where $i = 1, \dots, n$ [2].

$$Y_{i1} = [\beta_{01} + \beta_{11} z_{i1} + \beta_{21} z_{i2} + \cdots + \beta_{r1} z_{ir}] + \varepsilon_{i1}$$

$$Y_{i2} = [\beta_{02} + \beta_{12} z_{i1} + \beta_{22} z_{i2} + \cdots + \beta_{r2} z_{ir}] + \varepsilon_{i2}$$

$$\vdots$$

$$Y_{im} = [\beta_{0m} + \beta_{1m} z_{i1} + \beta_{2m} z_{i2} + \cdots + \beta_{rm} z_{ir}] + \varepsilon_{im}$$

Now we can construct each matrix component of our multivariate multiple regression model. We will consider the j^{th} response variable [2].

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nm} \end{bmatrix} = [\mathbf{Y}_{(1)} \quad \vdots\, \mathbf{Y}_{(2)} \quad \vdots \quad \cdots \quad \vdots \quad \mathbf{Y}_{(m)}], \quad \text{where } \mathbf{Y}_{(j)} = \begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{nj} \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{nr} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rm} \end{bmatrix} = [\boldsymbol{\beta}_{(1)} \quad \vdots\, \boldsymbol{\beta}_{(2)} \quad \vdots \quad \cdots \quad \vdots \quad \boldsymbol{\beta}_{(m)}], \quad \text{where } \boldsymbol{\beta}_{(j)} = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \\ \vdots \\ \beta_{rj} \end{bmatrix}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nm} \end{bmatrix} = [\boldsymbol{\varepsilon}_{(1)} \quad \vdots\, \boldsymbol{\varepsilon}_{(2)} \quad \vdots \quad \cdots \quad \vdots \quad \boldsymbol{\varepsilon}_{(m)}], \quad \text{where } \boldsymbol{\varepsilon}_{(j)} = \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{nj} \end{bmatrix}$$

The $\mathbf{Y}_{(j)}$ vectors are column vectors that contain the values of the j^{th} response variable for each of the n trials or observations. Similarly, each $\boldsymbol{\varepsilon}_{(j)}$ vector contains the random error terms obtained for each of the n trials when considering the j^{th} response variable. Each $\boldsymbol{\beta}_{(j)}$ vector is comprised of the unknown regression coefficients for the regression model obtained for the j^{th} response variable. Note that in the multivariate model, the design matrix, $\mathbf{Z}$, is the same as in the single response or multiple linear regression model. We can generalize a multiple linear regression model for each response using the notation above.

$$\mathbf{Y}_{(j)} = \mathbf{Z}\boldsymbol{\beta}_{(j)} + \boldsymbol{\varepsilon}_{(j)}, \qquad j = 1, 2, \dots, m$$

Combining each single response model, the following matrix model for multivariate linear regression can be constructed.

$$\begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nm} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r1} & \beta_{r2} & \cdots & \beta_{rm} \end{bmatrix} \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nm} \end{bmatrix}$$

For simplicity, we will write this model as $\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ [2]. The dimensions of all matrix components of regression models introduced up to this point can be found in Table 2.1.3.

TABLE 2.1.3 Matrix Demensions

Model	Matrix	Dimension
Multiple Linear Regression	$\boldsymbol{y}$	$n \times 1$
	$\boldsymbol{Z}$	$n \times (r+1)$
	$\boldsymbol{\beta}$	$(r+1) \times 1$
	$\boldsymbol{\varepsilon}$	$1 \times m$

Multivariate Multiple Linear Regression	$\boldsymbol{Y}$	$n \times m$
	$\boldsymbol{Z}$	$n \times (r+1)$
	$\boldsymbol{\beta}$	$(r+1) \times m$
	$\boldsymbol{\varepsilon}$	$n \times m$

2.2 Some Useful Properties

Some known properties of the multivariate multiple linear regression model will be discussed briefly below. These properties will be useful in various derivations discussed later.

First, we must note that, for the multiple linear regression model, $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $E(\boldsymbol{y}) = \boldsymbol{Z\beta}$, and $\text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ [3]. This means that when we are ready to construct our multivariate model using several multiple linear regression models, we have $E(\boldsymbol{\varepsilon}_{(j)}) = \mathbf{0}$, $E(\mathbf{Y}) = \boldsymbol{Z\beta}$, and $\text{Cov}(\boldsymbol{\varepsilon}_{(j)}) = \sigma_{jj}\mathbf{I}$ for the j^{th} response [2]. Note that the errors for different responses on the same trial may be correlated.

The unknown parameters for this model are σ^2 and $\boldsymbol{\beta}$ (or $\sigma^2{}_{jj}$ and $\boldsymbol{\beta}_{(j)}$ when we consider the model for each of our several responses separately).

When we consider the multivariate multiple regression model, we will observe the following properties: $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\varepsilon}_{(j)}, \boldsymbol{\varepsilon}_{(k)}) = \sigma_{jk}\mathbf{I} = \begin{bmatrix} \sigma_{jk} & 0 & \cdots & 0 \\ 0 & \sigma_{jk} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{jk} \end{bmatrix}$ (an $n \times n$ matrix), where $j, k = 1,2, \dots, m$ [3]. To get a better understanding of the meaning of this second property, let's write out the columns of $\boldsymbol{\varepsilon}$. We have

$\boldsymbol{\varepsilon}_{(1)} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{n1} \end{bmatrix}, \boldsymbol{\varepsilon}_{(2)} = \begin{bmatrix} \varepsilon_{12} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{n2} \end{bmatrix}, \dots, \boldsymbol{\varepsilon}_{(m)} = \begin{bmatrix} \varepsilon_{1m} \\ \varepsilon_{2m} \\ \vdots \\ \varepsilon_{nm} \end{bmatrix}$. The entries of $\boldsymbol{\varepsilon}_{(1)}$ $(\varepsilon_{11}, \varepsilon_{21}, \dots, \varepsilon_{n1})$ are independent, because they are observations from different trials (e.g. observations from different people). Similarly, the entries of $\boldsymbol{\varepsilon}_{(2)}$ are independent of each other. The entries of $\boldsymbol{\varepsilon}_{(3)}$ are independent, and so on. Thus, $\mathrm{Cov}(\varepsilon_{h1}, \varepsilon_{i1}) = 0$, $\mathrm{Cov}(\varepsilon_{h2}, \varepsilon_{i2}) = 0, \dots, \mathrm{Cov}(\varepsilon_{hm}, \varepsilon_{im}) = 0$ for $h, i = 1, \dots, n$. These account for the zero entries in the $\mathrm{Cov}(\boldsymbol{\varepsilon}_{(j)}, \boldsymbol{\varepsilon}_{(k)})$ matrix. On the other hand, let's look at the rows of $\boldsymbol{\varepsilon}$: $[\varepsilon_{11} \quad \varepsilon_{12} \quad \cdots \quad \varepsilon_{1m}], [\varepsilon_{21} \quad \varepsilon_{22} \quad \cdots \quad \varepsilon_{2m}], \dots, [\varepsilon_{n1} \quad \varepsilon_{n2} \quad \cdots \quad \varepsilon_{nm}]$. The entries in each row are the errors for responses from the same trial (e.g. m response variables are measured on the same person). Thus, the errors for different responses on the same trial can be correlated ($\mathrm{Cov}(\varepsilon_{1j}, \varepsilon_{1k}) \neq 0$, $\mathrm{Cov}(\varepsilon_{2j}, \varepsilon_{2k}) \neq 0, \dots, \mathrm{Cov}(\varepsilon_{nj}, \varepsilon_{nk}) \neq 0$, where $j, k = 1, \dots, m$). These account for the nonzero entries in the $\mathrm{Cov}(\boldsymbol{\varepsilon}_{(j)}, \boldsymbol{\varepsilon}_{(k)})$ matrix.

The unknown parameters for the multivariate model are $\mathbf{\Sigma}$ and $\boldsymbol{\beta}$. The structure of $\mathbf{\Sigma}$ is shown below [3].

$$\mathbf{\Sigma} = \{\sigma_{jk}\} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix}$$

Estimation of the parameters $\mathbf{\Sigma}$ and $\boldsymbol{\beta}$ will be discussed in the following section.

3 Parameter Estimation

3.1 Estimating $\boldsymbol{\beta}$

We will approach the task of estimating the $\boldsymbol{\beta}$ matrix from a multiple linear regression standpoint. Recall that multiple linear regression deals with predicting the value of only *one* response variable based on values of several predictors. When we are concerned with a model that includes many response variables, we can estimate the $\boldsymbol{\beta}_{(j)}$ vector for each single response model, $\mathbf{Y}_{(j)} = \mathbf{Z}\boldsymbol{\beta}_{(j)} + \boldsymbol{\varepsilon}_{(j)}$, where $j = 1,2, \dots, m$ [3]. Then we can combine each estimated $\boldsymbol{\beta}_{(j)}$ vector ($\widehat{\boldsymbol{\beta}}_{(j)}$) into a matrix that estimates $\boldsymbol{\beta}$. The construction of the matrix is shown below [3].

$$\widehat{\boldsymbol{\beta}} = \left[\widehat{\boldsymbol{\beta}}_{(1)} \quad \vdots\, \widehat{\boldsymbol{\beta}}_{(2)} \quad \vdots \quad \cdots \quad \vdots \quad \widehat{\boldsymbol{\beta}}_{(m)}\right]$$

To obtain each $\widehat{\boldsymbol{\beta}}_{(j)}$ estimate, we will use the least-squares method. This method yields estimates of the unknown regression coefficients using the given response values contained in $\mathbf{Y}_{(j)}$ and values of the predictors (organized in the $\mathbf{Z}$ matrix).

We will discuss some important ideas before introducing the proof of the least-squares method for finding $\widehat{\boldsymbol{\beta}}$. First, let's consider the vector $\mathbf{b}_{(j)} = \begin{bmatrix} \mathrm{b}_{0j} \\ \mathrm{b}_{1j} \\ \vdots \\ \mathrm{b}_{rj} \end{bmatrix}$, which contains trial values for $\boldsymbol{\beta}_{(j)}$. So we have a trial model:

$$\begin{bmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{nj} \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \mathrm{b}_{0j} \\ \mathrm{b}_{1j} \\ \vdots \\ \mathrm{b}_{rj} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{nj} \end{bmatrix}.$$

Next, observe that we can write a regression model for the j^{th} response on the i^{th} trial using the trial values, $\mathbf{b}_{(j)}$: $Y_{ij} = [\mathrm{b}_{0j} + \mathrm{b}_{1j}z_{i1} + \mathrm{b}_{2j}z_{i2} + \cdots + \mathrm{b}_{rj}z_{ir}] + \varepsilon_{ij}$, where $i = 1, \dots, n$ and $j = 1, \dots, m$. We want to focus on the difference between the observed value, Y_{ij}, and the predicted mean of Y_{ij} given by $\mathrm{b}_{0j} + \mathrm{b}_{1j}z_{i1} + \mathrm{b}_{2j}z_{i2} + \cdots + \mathrm{b}_{rj}z_{ir}$. We would like to minimize this difference, which would result in the error term taking on the smallest possible value. Ultimately, we will select the $\mathbf{b}_{(j)}$ vector that minimizes the sum of the squares of these differences. We call the resultant estimated parameter vector, $\widehat{\boldsymbol{\beta}}_{(j)} = \begin{bmatrix} \hat{\beta}_{0j} \\ \hat{\beta}_{1j} \\ \vdots \\ \hat{\beta}_{rj} \end{bmatrix}$. Using this least-squares method, we end up with $\widehat{\boldsymbol{\beta}}_{(j)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(j)}$ [2]. Next, we can collect all of the univariate $\widehat{\boldsymbol{\beta}}_{(j)}$ estimates and form the estimated parameter matrix for the multivariate case, $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_{(1)} \ \vdots\ \widehat{\boldsymbol{\beta}}_{(2)} \ \vdots\ \cdots \ \vdots\ \widehat{\boldsymbol{\beta}}_{(m)}]$. Our final result becomes $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\mathbf{Y}_{(1)} \ \vdots\ \mathbf{Y}_{(2)} \ \vdots\ \cdots \ \vdots\ \mathbf{Y}_{(m)}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ [2]. A detailed proof is constructed below.

THEOREM 1: $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ **(Johnson and Wichern, 2007)**

Proof:

Consider the model $\mathbf{Y}_{(j)} = \mathbf{Z}\boldsymbol{\beta}_{(j)} + \boldsymbol{\varepsilon}_{(j)}$.

$$\begin{aligned} \text{Let } S(\mathbf{b}_{(j)}) &= \textstyle\sum_{i=1}^{n}[Y_{ij} - (\mathrm{b}_{0j} + \mathrm{b}_{1j}z_{i1} + \mathrm{b}_{2j}z_{i2} + \cdots + \mathrm{b}_{rj}z_{ir})]^2 \\ &= (\mathbf{Y}_{(j)} - \mathbf{Z}\mathbf{b}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{Z}\mathbf{b}_{(j)}) \\ &= (\mathbf{Y}_{(j)}{}' - \mathbf{b}'_{(j)}\mathbf{Z}')(\mathbf{Y}_{(j)} - \mathbf{Z}\mathbf{b}_{(j)}) \end{aligned}$$

$$= \mathbf{Y}_{(j)}'\mathbf{Y}_{(j)} - \mathbf{b}_{(j)}'\mathbf{Z}'\mathbf{Y}_{(j)} - \mathbf{Y}_{(j)}'\mathbf{Z}\mathbf{b}_{(j)} + \mathbf{b}_{(j)}'\mathbf{Z}'\mathbf{Z}\mathbf{b}_{(j)}$$

$$= \mathbf{Y}_{(j)}'\mathbf{Y}_{(j)} - 2\mathbf{b}_{(j)}'\mathbf{Z}'\mathbf{Y}_{(j)} + \mathbf{b}_{(j)}'\mathbf{Z}'\mathbf{Z}\mathbf{b}_{(j)}$$

*Note: $\mathbf{b}_{(j)}'\mathbf{Z}'\mathbf{Y}_{(j)}$ is a scalar, and it's transpose, $\mathbf{Y}_{(j)}'\mathbf{Z}\mathbf{b}_{(j)}$, is the same scalar; this property allows us to go from the second-to-last step to the last step above.

To minimize $S(\mathbf{b}_{(j)})$, find its derivative with respect to $\mathbf{b}_{(j)}$ and evaluate it at $\widehat{\boldsymbol{\beta}}_{(j)}$, set it equal to zero and solve for $\widehat{\boldsymbol{\beta}}_{(j)}$. This will be our estimatior for the unknown parameter vector $\boldsymbol{\beta}_{(j)}$.

$$\frac{\partial S}{\partial \mathbf{b}_{(j)}}|_{\widehat{\boldsymbol{\beta}}_{(j)}} = -2\mathbf{Z}'\mathbf{Y}_{(j)} + 2\mathbf{Z}'\mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}$$

$$\frac{\partial S}{\partial \mathbf{b}_{(j)}}|_{\hat{\mathbf{b}}_{(j)}} = \mathbf{0} \Rightarrow -2\mathbf{Z}'\mathbf{Y}_{(j)} + 2\mathbf{Z}'\mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)} = \mathbf{0}$$

$$\Rightarrow \mathbf{Z}'\mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)} = \mathbf{Z}'\mathbf{Y}_{(j)}$$

$$\Rightarrow \widehat{\boldsymbol{\beta}}_{(j)} = (\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}'\mathbf{Y}_{(j)}$$

Now consider the multivariate case in which we are concerned with the matrix $S(\boldsymbol{B}) = (\mathbf{Y} - \mathbf{ZB})'(\mathbf{Y} - \mathbf{ZB})$ where $\mathbf{B} = [\mathbf{b}_{(1)} \ \vdots\ \mathbf{b}_{(2)} \ \vdots\ \cdots \ \vdots\ \mathbf{b}_{(m)}]$.

Write $S(\boldsymbol{B})$ in matrix form.

$$S(\boldsymbol{B}) = (\mathbf{Y} - \mathbf{ZB})'(\mathbf{Y} - \mathbf{ZB})$$

$$= \begin{bmatrix} (\mathbf{Y}_{(1)} - \mathbf{Zb}_{(1)})'(\mathbf{Y}_{(1)} - \mathbf{Zb}_{(1)}) & (\mathbf{Y}_{(1)} - \mathbf{Zb}_{(1)})'(\mathbf{Y}_{(2)} - \mathbf{Zb}_{(2)}) & \cdots & (\mathbf{Y}_{(1)} - \mathbf{Zb}_{(1)})'(\mathbf{Y}_{(m)} - \mathbf{Zb}_{(m)}) \\ (\mathbf{Y}_{(2)} - \mathbf{Zb}_{(2)})'(\mathbf{Y}_{(1)} - \mathbf{Zb}_{(1)}) & (\mathbf{Y}_{(2)} - \mathbf{Zb}_{(2)})'(\mathbf{Y}_{(2)} - \mathbf{Zb}_{(2)}) & \cdots & (\mathbf{Y}_{(2)} - \mathbf{Zb}_{(2)})'(\mathbf{Y}_{(m)} - \mathbf{Zb}_{(m)}) \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{Y}_{(m)} - \mathbf{Zb}_{(m)})'(\mathbf{Y}_{(1)} - \mathbf{Zb}_{(1)}) & (\mathbf{Y}_{(m)} - \mathbf{Zb}_{(m)})'(\mathbf{Y}_{(2)} - \mathbf{Zb}_{(2)}) & \cdots & (\mathbf{Y}_{(m)} - \mathbf{Zb}_{(m)})'(\mathbf{Y}_{(m)} - \mathbf{Zb}_{(m)}) \end{bmatrix}$$

* Note: $S(\boldsymbol{B})$ is the sum of squares and cross products (SSCP) matrix; the diagonal entries are the squared errors, and the off-diagonal entries are cross products.

Minimize the trace of $S(\boldsymbol{B})$ to minimize the sum of the squared error terms.

$$\mathrm{Tr}[S(\boldsymbol{B})] = \left(\mathbf{Y}_{(1)} - \mathbf{Zb}_{(1)}\right)'\left(\mathbf{Y}_{(1)} - \mathbf{Zb}_{(1)}\right) + \left(\mathbf{Y}_{(2)} - \mathbf{Zb}_{(2)}\right)'\left(\mathbf{Y}_{(2)} - \mathbf{Zb}_{(2)}\right) + \cdots + \left(\mathbf{Y}_{(m)} - \mathbf{Zb}_{(m)}\right)'(\mathbf{Y}_{(m)} - \mathbf{Zb}_{(m)})$$

Minimizing each term separately will minimize $\mathrm{Tr}[S(\boldsymbol{B})]$.

We already determined that $\widehat{\boldsymbol{\beta}}_{(j)} = (\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}'\mathbf{Y}_{(j)}$ will minimize $\left(\mathbf{Y}_{(j)} - \mathbf{Zb}_{(j)}\right)'(\mathbf{Y}_{(j)} - \mathbf{Zb}_{(j)})$. So, we can collect and place all univariate $\widehat{\boldsymbol{\beta}}_{(j)}$ vectors into a $\widehat{\boldsymbol{\beta}}$ matrix that will estimate $\boldsymbol{\beta}$ in the multivariate multiple regression case.

Now we have $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_{(1)} \quad \vdots\, \widehat{\boldsymbol{\beta}}_{(2)} \quad \vdots \quad \cdots \quad \vdots \quad \widehat{\boldsymbol{\beta}}_{(m)}]$
$= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\mathbf{Y}_{(1)} \quad \vdots\, \mathbf{Y}_{(2)} \quad \vdots \quad \cdots \quad \vdots \quad \mathbf{Y}_{(m)}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$, as desired.

3.2 $\widehat{\boldsymbol{Y}}$ and $\hat{\boldsymbol{\varepsilon}}$

Once we determine $\widehat{\boldsymbol{\beta}}$, we are able to obtain the $\widehat{\mathbf{Y}}$ matrix. This is a matrix of fitted or predicted responses, generated based upon a known matrix of predictor variables, $\mathbf{Z}$, and the estimated parameter matrix, $\widehat{\boldsymbol{\beta}}$. Entries are generated by the matrix equation $\widehat{\mathbf{Y}} = \mathbf{Z}\widehat{\boldsymbol{\beta}}$ [2]. Once the $\widehat{\mathbf{Y}}$ matrix is computed, the predicted error matrix, $\hat{\boldsymbol{\varepsilon}}$, can be obtained by $\mathbf{Y} - \widehat{\mathbf{Y}}$ [2]. This gives us the distances between the observed Y_{ij} values and the predicted Y_{ij} values.

3.3 Some Useful Properties

$\widehat{\boldsymbol{\beta}}$ is an unbiased estimator and the maximum likelihood estimator for $\boldsymbol{\beta}$ [3]. A maximum likelihood estimator has parametric values that make the observed results most probable given the model. See below for the proofs of these results.

LEMMA 1: $\widehat{\boldsymbol{\beta}}$ is unbiased **(Johnson and Wichern, 2007)**

<u>Proof</u>:

$$\begin{aligned}
\mathrm{E}(\widehat{\boldsymbol{\beta}}) &= \mathrm{E}((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}) \\
&= \mathrm{E}\left((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right) \\
&= \mathrm{E}((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}\boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathrm{E}(\boldsymbol{\varepsilon}) && [(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{I} \text{ and } (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \text{ is a matrix of constants}] \\
&= \boldsymbol{\beta}, \text{ as desired.} && [\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}]
\end{aligned}$$

LEMMA 2: $\widehat{\boldsymbol{\beta}}$ is the maximum likelihood estimator (when errors are normally distributed) **(Johnson and Wichern, 2007)**

<u>**Proof**</u>:

Using properties mentioned in section 2.2, we can see that $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma_{jj}^2)$, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

So we can write the normal density function for each ε_{ij}:

$$f(\varepsilon_{ij}) = \frac{1}{\sigma_{jj}(2\pi)^{1/2}} \exp[-\frac{1}{2\sigma_{jj}^2}\varepsilon_{ij}{}^2]$$

Next, we can construct a likelihood function – the joint density of $\varepsilon_{1j}, \varepsilon_{2j}, \ldots, \varepsilon_{nj}$ for the j^{th} response variable:

$$\mathrm{L}(\boldsymbol{\beta}_{(j)}, \sigma_{jj}^2) = \prod_{i=1}^{n} f(\varepsilon_{ij})$$

$$= \frac{1}{(2\pi)^{n/2}\sigma_{jj}{}^n} \exp[-\frac{1}{2\sigma_{jj}^2}\boldsymbol{\varepsilon}_{(j)}{}'\boldsymbol{\varepsilon}_{(j)}]$$

$$= \frac{1}{(2\pi)^{n/2}\sigma_{jj}{}^n} \exp[-\frac{1}{2\sigma_{jj}^2}(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})]$$

Taking the natural log of both sides, we obtain:

$$\ln \mathrm{L}(\boldsymbol{\beta}_{(j)}, \sigma_{jj}^2) = -\frac{n}{2}\ln 2\pi - n\ln\sigma_{jj} - \frac{1}{2\sigma_{jj}^2}(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})$$

Maximizing $\ln \mathrm{L}(\boldsymbol{\beta}_{(j)}, \sigma_{jj}^2)$ has the same effect of maximizing $\mathrm{L}(\boldsymbol{\beta}_{(j)}, \sigma_{jj}^2)$. For fixed σ_{jj}, to maximize $\ln \mathrm{L}(\boldsymbol{\beta}_{(j)}, \sigma_{jj}^2)$, $(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})$ must be minimized. Recall the proof of the result $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$. We minimized each $(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})$ term using the least-squares method. Minimizing $(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})$ in this context gives us the same result: $\widehat{\boldsymbol{\beta}}_{(j)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(j)}$. Collecting all of the single response results, we have $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$. Thus, $\widehat{\boldsymbol{\beta}}$ is a maximum likelihood estimator, as desired.

Another useful sampling property involving $\widehat{\boldsymbol{\beta}}$ with a proof that follows can be observed: $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{(j)}, \widehat{\boldsymbol{\beta}}_{(k)}) = \sigma_{jk}(\mathbf{Z}'\mathbf{Z})^{-1}$, where $j, k = 1, \ldots, m$ [2].

THEOREM 2: $\text{Cov}(\hat{\boldsymbol{\beta}}_{(j)}, \hat{\boldsymbol{\beta}}_{(k)}) = \sigma_{jk}(\mathbf{Z}'\mathbf{Z})^{-1}$, where $j, k = 1, \ldots, m$ **(Johnson and Wichern, 2007)**

__Proof__:

$\text{Cov}(\hat{\boldsymbol{\beta}}_{(j)}, \hat{\boldsymbol{\beta}}_{(k)}) = \text{Cov}((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(j)}, (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(k)})$

Let $\mathbf{c} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Now,
$$\begin{aligned}
\text{Cov}(\mathbf{cY}_{(j)}, \mathbf{cY}_{(k)}) &= \text{E}[\left(\mathbf{cY}_{(j)} - \text{E}(\mathbf{cY}_{(j)})\right)\left(\mathbf{cY}_{(k)} - \text{E}(\mathbf{cY}_{(k)})\right)'] \\
&= \text{E}[\left(\mathbf{cY}_{(j)} - \text{E}(\mathbf{cY}_{(j)})\right)\left(\mathbf{Y}_{(k)}{}'\mathbf{c}' - \left(E(\boldsymbol{Y}_{(k)})\right)'\mathbf{c}'\right)] \\
&= \text{E}[\mathbf{c}\left(\mathbf{Y}_{(j)} - \text{E}(\mathbf{Y}_{(j)})\right)\left(\mathbf{Y}_{(k)} - E(\boldsymbol{Y}_{(k)})\right)'\mathbf{c}'] \\
&= \mathbf{c}\text{E}[\left(\mathbf{Y}_{(j)} - \text{E}(\mathbf{Y}_{(j)})\right)\left(\mathbf{Y}_{(k)} - E(\boldsymbol{Y}_{(k)})\right)']\mathbf{c}' \\
&= \mathbf{c}\text{E}[(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)})']\mathbf{c}' \qquad [\text{E}(\mathbf{Y}_{(j)}) = \mathbf{Z}\boldsymbol{\beta}_{(j)}] \\
&= \mathbf{c}\text{E}[\boldsymbol{\varepsilon}_{(j)}\boldsymbol{\varepsilon}_{(k)}{}']\,\mathbf{c}' \\
&= \mathbf{c}\text{E}[\left(\boldsymbol{\varepsilon}_{(j)} - \text{E}(\boldsymbol{\varepsilon}_{(j)})\right)\left(\boldsymbol{\varepsilon}_{(k)} - E(\boldsymbol{\varepsilon}_{(k)})\right)']\mathbf{c}' \qquad [\text{E}(\boldsymbol{\varepsilon}_{(j)}) = \mathbf{0}] \\
&= \mathbf{c}\text{Cov}(\boldsymbol{\varepsilon}_{(j)}, \boldsymbol{\varepsilon}_{(k)})\,\mathbf{c}' \\
&= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\sigma_{jk}\mathbf{I})((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')' \\
&\underbrace{=}_{*} (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\sigma_{jk}\mathbf{I})(\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}) \\
&= \sigma_{jk}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\,\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \\
&= \sigma_{jk}(\mathbf{Z}'\mathbf{Z})^{-1}\text{, as desired.}
\end{aligned}$$

$$* \; ((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')' = \mathbf{Z}((\mathbf{Z}'\mathbf{Z})^{-1})' = \mathbf{Z}((\mathbf{Z}'\mathbf{Z})')^{-1} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$$

Two useful properties regarding the $\hat{\boldsymbol{\varepsilon}}_{(j)}$ vectors should also be noted: $\text{E}(\hat{\boldsymbol{\varepsilon}}_{(j)}) = \mathbf{0}$, which implies that $\text{E}(\hat{\boldsymbol{\varepsilon}}) = \mathbf{0}$, and $\text{E}(\hat{\boldsymbol{\varepsilon}}_{(j)}{}'\hat{\boldsymbol{\varepsilon}}_{(k)}) = \sigma_{jk}(n - r - 1)$ [2]. Recall that n is the number of trials performed or observations taken and r is the number of predictor variables. The latter result will help us obtain an unbiased estimator for $\mathbf{\Sigma}$, which will be discussed in the following section. Proofs for these properties are constructed below.

THEOREM 3: $\mathrm{E}(\hat{\boldsymbol{\varepsilon}}_{(j)}) = \mathbf{0}$ **(Johnson and Wichern, 2007)**

<u>Proof</u>:

$\hat{\boldsymbol{\varepsilon}}_{(j)} = \mathbf{Y}_{(j)} - \widehat{\mathbf{Y}}_{(j)} = \mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}$ by definition.

$$\begin{aligned}
\text{So, } \mathrm{E}(\hat{\boldsymbol{\varepsilon}}_{(j)}) &= \mathrm{E}(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}) \\
&= \mathrm{E}(\mathbf{Y}_{(j)}) - \mathrm{E}(\mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}) \\
&= \mathbf{Z}\boldsymbol{\beta}_{(j)} - \mathbf{Z}\mathrm{E}(\widehat{\boldsymbol{\beta}}_{(j)}) \\
&= \mathbf{Z}\boldsymbol{\beta}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)} \qquad \text{[Since } \widehat{\boldsymbol{\beta}}_{(j)} \text{ is an unbiased estimator]} \\
&= \mathbf{0}\text{, as desired.}
\end{aligned}$$

THEOREM 4: $\mathrm{E}(\hat{\boldsymbol{\varepsilon}}_{(j)}{}'\hat{\boldsymbol{\varepsilon}}_{(k)}) = \sigma_{jk}(n - r - 1)$ **(Johnson and Wichern, 2007)**

<u>Proof</u>:

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\varepsilon}}_{(j)}{}'\hat{\boldsymbol{\varepsilon}}_{(k)}) &= \mathrm{E}\left[\left(\mathbf{Y}_{(j)} - \widehat{\mathbf{Y}}_{(j)}\right)'\left(\mathbf{Y}_{(k)} - \widehat{\mathbf{Y}}_{(k)}\right)\right] \\
&= \mathrm{E}\left[\left(\mathbf{Y}_{(j)} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(j)}\right)'\left(\mathbf{Y}_{(k)} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(k)}\right)\right] \\
&= \mathrm{E}\left[\left((\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(j)}\right)'\left((\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)}\right)\right] \\
&= \mathrm{E}\left[\mathbf{Y}_{(j)}{}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)}\right] \\
&= \mathrm{E}\left[\mathbf{Y}_{(j)}{}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)}\right] \quad {}^*[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \text{ is symmetric and idempotent}] \\
&= \mathrm{E}\Big[(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\left(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)}\right) + \left(\mathbf{Z}\boldsymbol{\beta}_{(j)}\right)'(\mathbf{I} \\
&\qquad - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)}\Big]
\end{aligned}$$

*In the preceding step, by adding the $\left(\mathbf{Z}\boldsymbol{\beta}_{(j)}\right)'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)}$ term, we correct for subtracting $\mathbf{Z}\boldsymbol{\beta}_{(j)}$ from $\mathbf{Y}_{(j)}$ and $\mathbf{Z}\boldsymbol{\beta}_{(k)}$ from $\mathbf{Y}_{(k)}$.

We will continue and focus on the correction term next.

$$\begin{aligned}
\left(\mathbf{Z}\boldsymbol{\beta}_{(j)}\right)'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)} &= \boldsymbol{\beta}_{(j)}{}'\mathbf{Z}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)} \\
&= \left(\boldsymbol{\beta}_{(j)}{}'\mathbf{Z}' - \boldsymbol{\beta}_{(j)}{}'\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\right)\mathbf{Y}_{(k)} \qquad [\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} = \mathbf{I}] \\
&= \left(\boldsymbol{\beta}_{(j)}{}'\mathbf{Z}' - \boldsymbol{\beta}_{(j)}{}'\mathbf{Z}'\right)\mathbf{Y}_{(k)} \\
&= \mathbf{0}
\end{aligned}$$

So now we are just working with $\mathrm{E}\left[(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\left(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)}\right)\right] =$ $\mathrm{E}\left[\boldsymbol{\varepsilon}_{(j)}{}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\boldsymbol{\varepsilon}_{(k)}\right]$.

We will utilize the property that says $\mathbf{x}'\mathbf{A}\mathbf{y} = \mathrm{tr}(\mathbf{x}'\mathbf{A}\mathbf{y}) = \mathrm{tr}(\mathbf{A}\mathbf{y}\mathbf{x}')$ when $\mathbf{A}$ is a symmetric matrix and $\mathbf{x}$ and $\mathbf{y}$ are vectors.

Since $\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is symmetric, we have

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{\varepsilon}_{(j)}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\boldsymbol{\varepsilon}_{(k)}\right] &= \mathrm{E}\left[\mathrm{tr}\left[(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\boldsymbol{\varepsilon}_{(k)}\boldsymbol{\varepsilon}_{(j)}'\right]\right] \\
&= \mathrm{tr}[(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathrm{E}[\boldsymbol{\varepsilon}_{(k)}\boldsymbol{\varepsilon}_{(j)}']] \\
&= \mathrm{tr}[(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathrm{E}[\left(\boldsymbol{\varepsilon}_{(k)} - \mathrm{E}\left(\boldsymbol{\varepsilon}_{(k)}\right)\right)\left(\boldsymbol{\varepsilon}_{(j)} - \right. \\
&\quad \left. E\left(\boldsymbol{\varepsilon}_{(j)}\right)\right)']] \\
&= \mathrm{tr}\left[(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\ \mathrm{Cov}\left(\boldsymbol{\varepsilon}_{(k)}, \boldsymbol{\varepsilon}_{(j)}\right)\right] \\
&= \mathrm{tr}\left[(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\sigma_{jk}\ \right] \\
&= \sigma_{jk}\mathrm{tr}[(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\ ] \\
&= \sigma_{jk}[\mathrm{tr}(\mathbf{I}) - \mathrm{tr}(\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')] \\
&= \sigma_{jk}\left[n - \underbrace{(r+1)}_{*}\right] = \sigma_{jk}(n - r - 1), \text{ as desired.}
\end{aligned}$$

*Please refer to the appendix to see how these results are obtained.

The proof of the fact that $\widehat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\varepsilon}}$ are uncorrelated involves a method similar to one used in the proof above. We would start with $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{(j)}, \hat{\boldsymbol{\varepsilon}}_{(k)})$ and then use the definition of covariance to write the next step. We need to use the fact that $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator and the definitions of $\hat{\boldsymbol{\varepsilon}}$ and $\widehat{\mathbf{Y}}$ in the following steps. Matrix algebra and the purposeful subtraction of terms and the addition/subtraction of the corresponding correction terms will get us to the end result: $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{(j)}, \hat{\boldsymbol{\varepsilon}}_{(k)}) = \mathbf{0}$ [2]. The complete proof is located in the appendix.

3.4 Estimating $\mathbf{\Sigma}$

In the case where we determined an estimate for the $\boldsymbol{\beta}$ matrix, we were able to conclude that our estimator, $\widehat{\boldsymbol{\beta}}$, was both unbiased and a maximum likelihood estimator. When we consider finding an estimator for $\mathbf{\Sigma}$, we will discover that there is not an estimator, $\widehat{\mathbf{\Sigma}}$, that is both unbiased and the maximum likelihood estimator. One of the properties proved in section 3.3 will easily help us find an unbiased estimator for $\mathbf{\Sigma}$. Let's recall that $\mathrm{E}(\hat{\boldsymbol{\varepsilon}}_{(j)}{}'\hat{\boldsymbol{\varepsilon}}_{(k)}) = \sigma_{jk}(n-r-1)$. We can extend this result to the multivariate case by adapting the meaning of $\mathrm{E}(\hat{\boldsymbol{\varepsilon}}_{(j)}{}'\hat{\boldsymbol{\varepsilon}}_{(k)}) = \sigma_{jk}(n-r-1)$. We will focus on the predicted error matrix, $\hat{\boldsymbol{\varepsilon}}$, which corresponds to a multivariate multiple linear regression problem. Using the previously mentioned result, we can construct a new one corresponding to the multivariate case:

$$\mathrm{E}(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) = (n-r-1)\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{bmatrix} = (n-r-1)\mathbf{\Sigma} \text{ [3]}.$$

Now we can easily see how to choose an unbiased estimator for $\mathbf{\Sigma}$. See below for the proof that $\widehat{\mathbf{\Sigma}} = \frac{1}{n-r-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ is an unbiased estimator for $\mathbf{\Sigma}$ [2].

LEMMA 3: $\widehat{\mathbf{\Sigma}} = \frac{1}{n-r-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$ is unbiased **(Johnson and Wichern, 2007)**

Proof:

$$\begin{aligned}\mathrm{E}(\widehat{\mathbf{\Sigma}}) &= \mathrm{E}\left(\frac{1}{n-r-1}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}\right) \\ &= \frac{1}{n-r-1}\mathrm{E}(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) \\ &= \frac{1}{n-r-1}(n-r-1)\mathbf{\Sigma} \\ &= \mathbf{\Sigma}\text{, as desired.}\end{aligned}$$

To find the maximum likelihood estimator (MLE) for $\mathbf{\Sigma}$, we will work with the single response model first and extend the resulting estimator to the multivariate case. We will consider the natural log of the likelihood function discussed in the proof that $\widehat{\boldsymbol{\beta}}$ is the maximum likelihood estimator for $\boldsymbol{\beta}$. For the j^{th} response, we have $\ln \mathrm{L}\left(\boldsymbol{\beta}_{(j)}, \sigma_{jj}^2\right) = -\frac{n}{2}\ln 2\pi - n\ln\sigma_{jj} - \frac{1}{2\sigma_{jj}^2}\left(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)}\right)'(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})$. Taking the derivative of this log likelihood function with respect to σ_{jj} and evaluating σ_{jj} at $\hat{\sigma}_{jj}$ and $\boldsymbol{\beta}_{(j)}$ at $\widehat{\boldsymbol{\beta}}_{(j)}$, we obtain $\frac{d}{d\sigma_{jj}}\ln \mathrm{L}\left(\boldsymbol{\beta}_{(j)}, \sigma_{jj}^2\right)|_{\sigma_{jj}=\hat{\sigma}_{jj},\, \boldsymbol{\beta}_{(j)}=\widehat{\boldsymbol{\beta}}_{(j)}} = -\frac{n}{\hat{\sigma}_{jj}} + \frac{1}{\hat{\sigma}_{jj}^3}\left(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}\right)'(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)})$. Setting the derivative equal to zero and solving for $\hat{\sigma}_{jj}^2$ will give us the value at which the log likelihood function (and the original likelihood function) is maximized. This will be the MLE for σ_{jj}^2. Continuing, we have $\frac{1}{\hat{\sigma}_{jj}}\left(-n + \frac{1}{\hat{\sigma}_{jj}^2}\left(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}\right)'\left(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}\right)\right) = 0$ which implies that $-n + \frac{1}{\hat{\sigma}_{jj}^2}\left(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}\right)'\left(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}\right) = 0$. Solving for $\hat{\sigma}_{jj}^2$, we obtain the MLE for σ_{jj}^2: $\hat{\sigma}_{jj}^2 = \frac{1}{n}\left(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}\right)'\left(\mathbf{Y}_{(j)} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_{(j)}\right)$. The MLE for $\mathbf{\Sigma}$ is developed similarly, and the resulting estimator is $\widehat{\mathbf{\Sigma}} = \frac{1}{n}\left(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}\right)$ [2].

4 Making Predictions

Using multivariate multiple linear regression to make predictions is a valuable concept that will be discussed in the following paragraphs. We will place fixed values of predictor variables in a row vector called $\boldsymbol{z}_0$. Writing out $\boldsymbol{z}_0$, we have $\boldsymbol{z}_0 = [1, z_{01}, \dots, z_{0r}]$. We will think about the estimation of the mean vector given this $\boldsymbol{z}_0$ vector. Multiplying $\boldsymbol{z}_0$ by the estimated parameter matrix $\widehat{\boldsymbol{\beta}}$ will result in an estimated vector of response variables, $\widehat{\mathbf{Y}}_0 = [\mathrm{Y}_{01}, \mathrm{Y}_{02}, \dots, \mathrm{Y}_{0m}] = \boldsymbol{z}_0 \widehat{\boldsymbol{\beta}}$ [2].

To clarify, let's look at an example. We will concentrate on the powerlifting problem considered in the introduction. We want to predict the amount of weight a 23-year-old woman who weighs 62 kilograms can squat, bench press and deadlift (the three lifts in a powerlifting competition). The vector of predictors will look like this: $\boldsymbol{z}_0 = [1 \;\; 62 \;\; 23]$. To obtain the vector of predicted responses, we have $\widehat{\mathbf{Y}}_0 = \boldsymbol{z}_0 \widehat{\boldsymbol{\beta}} =$

$$[1 \;\; 62 \;\; 23] \begin{bmatrix} 117.6379 & 71.7612 & 139.3978 \\ 0.3552 & 0.1214 & 0.4731 \\ -1.0857 & -0.4387 & -0.9772 \end{bmatrix} = [137.16 \;\; 69.20 \;\; 146.25].$$

Interpreting the resultant $\widehat{\mathbf{Y}}_0$ vector will require the definition of the three response variables. We have $\mathrm{Y}_1 =$ amount of weight squatted, $\mathrm{Y}_2 =$ amount of weight bench pressed, and $\mathrm{Y}_3 =$ amount of weight deadlifted. So, based on this woman's age and body weight, the predicted mean responses take on the following values: $\mathrm{Y}_{01} = 137.16\ kg$, $\mathrm{Y}_{02} = 69.20\ kg$, and $\mathrm{Y}_{03} = 146.25\ kg$. She would be predicted to squat $137.16\ kg$, bench press $69.20\ kg$, and deadlift $146.25\ kg$.

Next, it will be helpful to construct the distribution of the mean response, $\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}$, in order to understand some ideas that will be discussed later. Recall that $\mathrm{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{(j)}, \widehat{\boldsymbol{\beta}}_{(k)}) = \sigma_{jk}(\mathbf{Z}'\mathbf{Z})^{-1}$. To find the mean of $\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}$, we need to find it's expected value, $\mathrm{E}(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}})$. Since $\boldsymbol{z}_0$ is a vector of fixed values, $\mathrm{E}(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}) = \boldsymbol{z}_0\ \mathrm{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{z}_0\boldsymbol{\beta}$. Next, $\mathrm{Cov}(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)}, \boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(k)}) = \mathrm{E}\left[\left(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)} - \mathrm{E}(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)})\right)\left(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(k)} - \mathrm{E}(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(k)})\right)\right] = \mathrm{E}\left[(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{z}_0\boldsymbol{\beta}_{(j)})(\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{z}_0\boldsymbol{\beta}_{(k)})'\right] = \mathrm{E}\left[\boldsymbol{z}_0(\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta}_{(j)})(\boldsymbol{z}_0(\widehat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)}))'\right] = \mathrm{E}[\boldsymbol{z}_0(\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta}_{(j)})(\widehat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)})'\boldsymbol{z}_0{}'] = \boldsymbol{z}_0\mathrm{E}[(\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta}_{(j)})(\widehat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)})']\boldsymbol{z}_0{}' = \boldsymbol{z}_0\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{(j)}, \widehat{\boldsymbol{\beta}}_{(k)})\boldsymbol{z}_0{}' = \boldsymbol{z}_0(\sigma_{jk}(\mathbf{Z}'\mathbf{Z})^{-1})\boldsymbol{z}_0{}' = (\boldsymbol{z}_0(\mathbf{Z}'\mathbf{Z})^{-1}\boldsymbol{z}_0{}')\sigma_{jk}$. If we want the distribution of $\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}$, we will use the former result and extend the latter to the multivariate case to obtain $\boldsymbol{z}_0\widehat{\boldsymbol{\beta}} \backsim N_m(\boldsymbol{z}_0\boldsymbol{\beta}, (\boldsymbol{z}_0(\mathbf{Z}'\mathbf{Z})^{-1}\boldsymbol{z}_0{}')\boldsymbol{\Sigma})$ [2]. This distribution and the fact that $n\widehat{\boldsymbol{\Sigma}} \backsim \mathrm{W}_{m,n-r-1}(\boldsymbol{\Sigma})$ will help us write $100(1-\alpha)\%$ simultaneous prediction intervals for individual responses, Y_{0j} [2]. A prediction interval will give us an estimated range in which future responses will fall based on the $\boldsymbol{z}_0$ vector of predictor variable values. Given the distribution of the mean response, $\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)} \sim N(\boldsymbol{z}_0\boldsymbol{\beta}_{(j)}, (\boldsymbol{z}_0(\mathbf{Z}'\mathbf{Z})^{-1}\boldsymbol{z}_0{}')\sigma_{jj}^2)$, we can find the distribution of the difference of the mean response and the future observation Y_{0j}. The forecasted response has the distribution $\mathrm{Y}_{0j} \sim N(\boldsymbol{z}_0\boldsymbol{\beta}_{(j)}, \sigma_{jj}^2)$ [2]. The mean of the difference distribution becomes $\mathrm{E}(\mathrm{Y}_{0j} - \boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)}) = \mathrm{E}(\boldsymbol{z}_0\boldsymbol{\beta}_{(j)} - \boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)}) = 0$. The variance of the difference of two normally distributed variables is the sum of their respective variances: $\mathrm{Var}\ (\mathrm{Y}_{0j} - \boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)}) = \sigma_{jj}^2 + (\boldsymbol{z}_0(\mathbf{Z}'\mathbf{Z})^{-1}\boldsymbol{z}_0{}')\sigma_{jj}^2 = (1 + \boldsymbol{z}_0(\mathbf{Z}'\mathbf{Z})^{-1}\boldsymbol{z}_0{}')\sigma_{jj}^2$. Extending these results to the multivariate case, we have

$\mathbf{Y_0} - \boldsymbol{z}_0\widehat{\boldsymbol{\beta}} \sim N_m(\mathbf{0}, (1 + \boldsymbol{z}_0(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\boldsymbol{z}_0{}')\boldsymbol{\Sigma})$ [2].

Using the distribution of $\mathbf{Y_0} - \boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)}$, a simultaneous prediction interval can be written if $j = 1, \dots, m$, $F_{m,n-r-m}(\alpha)$ is the upper $100\alpha^{\text{th}}$ percentile of an F-distribution with m and $n - r - m$ degrees of freedom, $\widehat{\boldsymbol{\beta}}_{(j)}$ is the j^{th} column of $\widehat{\boldsymbol{\beta}}$, and $\hat{\sigma}_{jj}$ is the i^{th} diagonal element of the MLE of $\boldsymbol{\Sigma}$, $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$:

$$\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m}(\alpha)}\sqrt{(1 + \boldsymbol{z}_0(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\boldsymbol{z}_0{}')\left(\frac{n}{n-r-1}\hat{\sigma}_{jj}\right)} \text{ [2].}$$

These simultaneous prediction intervals will be very useful when we approach the application problem in the following section. These will help us determine where we can expect the next data point to fall. Suppose we calculate 90% prediction intervals. If we sample one more value from the population many times, we would expect the value to lie within the prediction interval in 90% of the samples. This concept is much more useful in the context of the application problem than computing confidence intervals where we are concerned with predicting a range in which a mean will fall.

5 Application To Powerlifting Data

Now it is time to use ideas from multivariate multiple regression to solve a practical problem relating to athletics. We will focus on the sport of powerlifting. Powerlifting is a relatively new sport in which competitors train to compete in three different weightlifting events. Similar to Olympic lifting (the weightlifting competition featured in the Olympics), powerlifting consists of three lifts. The three powerlifts are squat, bench press, and deadlift. In a powerlifting meet, competitors get three attempts in each event, and the best performance of the three is recorded as his or her score. “Scores” are recorded as amount of weight lifted, typically in measured in kilograms. In powerlifting meets, lifters are placed in age and weight classes, based on their age (in years) and bodyweight (in kilograms). They are judged against other athletes of the same gender, weight class, and age class. There are two types of powerlifting. In equipped powerlifting, competitors are allowed to wear special equipment, such as belts and bench shirts (worn while bench pressing) that are made of extra supportive material that will aid in lifting more weight. Unequipped or “raw” powerlifting events require participants to be totally unequipped while competing. Raw powerlifting is a newer idea that has been quickly gaining popularity among lifters over the past 15 years [1] [3] [7].

5.1 The Problem

In many individual athletic events, there are standards that one must reach in order to get to the next level of competition. For example, in the sport of track and field, a runner must run faster than a set standard, or a thrower must hit a specific pre-determined distance in his or her event to qualify for a certain prestigious meet. This allows for the most elite athletes in a sport to compete with other athletes that have an equally superior skill level. The concept of qualifying standards makes so many athletic contests exciting for fans to watch and can also help explain why athletes invest so much time, energy, and heart into their particular sport. Although many athletes thrive on achieving personal bests, being the best among others is also a driving force behind why many athletes compete in their respective sport in the first place.

This is where we encounter a problem concerning the organization of the national meet for raw powerlifters. This annual event, referred to as USAPL Raw Nationals, is not limited in terms of who can register to compete. (USAPL stands for USA Powerlifting, one of the biggest powerlifting organizations in the world.) Anyone who competes in a meet sanctioned by USAPL in a certain year "qualifies" to compete at Raw Nationals. Consequently, Raw Nationals currently isn't an event consisting of only the most elite powerlifters. Since the raw division of powerlifting is relatively new, this organizational structure works for now. USAPL representatives want the sport to gain recognition, and they want to promote participation as much as possible. Eventually, when there are a lot more powerlifters throughout the country, holding a national meet in which the best are competing against the best is going to be a good way to *keep* people interested. So, in the following sections, I will develop and propose a set of qualifying standards that could be used to regulate which athletes are able to compete in a national powerlifting meet for unequipped lifters. This is done from a multivariate multiple regression framework. We will use two predictor variables (body weight and age) and three response variables (amount of weight squatted, amount of weight bench pressed, and amount of weight deadlifted) to build a multivariate multiple regression model. (It is multivariate since we are concerned with predicting values of more than one response.) This model will give us information about how to predict values of the three response variables based on fixed values of the predictor variables.

5.2 The Data

In order to build the model, powerlifting data was collected from USAPL's website [7]. Men's and women's data was available for each of the five Raw National meets that have taken place since the inaugural year of 2008. For each observation, the information we have includes the following: age, body weight, amount of weight squatted, amount of weight benched, and amount of weight deadlifted. Age and body weight are the two predictor variables. Using notation introduced in the prior discussion of the model, we have $r = 2$. The amounts of weight squatted, bench pressed, and deadlifted are the three response variables, so this means we have $m = 3$. There were a few hurdles to jump over when compiling the data.

First, when the data for each year was combined to form one women's data set and one men's data set, within each data set, the observations weren't all independent of each other. A lot of the same lifters compete at Raw Nationals each year, so when the yearly data was combined, there were multiple observations for many of the competitors. Independent trials are required when taking a sample to perform multivariate multiple regression on. Without independent observations, we could have autocorrelated errors. To correct for this problem, the observations were sorted by name (person), and if there was more than one set of data for a single person, the observations for this person were averaged. The final data set used to do the regression consisted of the observations from people who only competed once between 2008 and 2012 and the averaged observations of people who competed more than once from 2008 to 2012.

The second problem that was encountered occurred after the first model was built using the modified data sets (containing averaged values) described above. It was noted that the responses weren't quite linear with respect to age. Data was collected for teens as young as 13 and adults as old as 80. Think about how the performance of this 13 year old would change each year. Barring some strange circumstances, he or she will most likely be able to lift more as he or she grows and gets stronger. On the other hand, adults approaching older ages are going to most likely see a decrease in their weightlifting performance, as their bodies gradually get naturally weaker with age. So the general trend seen was that performance continually got better with age until about the mid-20s. To account for this phenomenon, the observations from participants in the teen and junior age classes (ages 13-23) were disregarded. The final data set included observations from adults aged 24 and older. The sample size for the men's data turned out to be $n = 296$, and for the women's data, $n = 92$.

5.3 Estimated Parameters

To find the estimated values of the model parameters, data was imported into SAS, and the multivariate multiple regression was done using the general linear models procedure. The resulting $\widehat{\boldsymbol{\beta}}$ matrices are shown below.

$$\widehat{\boldsymbol{\beta}}_W = \begin{bmatrix} 117.6379 & 71.7612 & 139.3978 \\ 0.3552 & 0.1214 & 0.4731 \\ -1.0857 & -0.4387 & -0.9772 \end{bmatrix}$$

$$\widehat{\boldsymbol{\beta}}_M = \begin{bmatrix} 140.1218 & 89.7520 & 189.3326 \\ 1.4093 & 0.9870 & 1.1262 \\ -1.8468 & -0.9769 & -1.5782 \end{bmatrix}$$

We also obtained the $\widehat{\boldsymbol{\Sigma}}$ matrices by dividing each entry in the sum of squares and cross product (SSCP) matrices obtained from SAS output by the corresponding n value. (For women $n = 92$, and for men $n = 296$.) The resulting $\widehat{\boldsymbol{\Sigma}}$ matrices are shown below.

$$\widehat{\boldsymbol{\Sigma}}_W = \begin{bmatrix} 431.5993 & 196.4048 & 340.0263 \\ 196.4048 & 228.5138 & 214.1768 \\ 340.0263 & 214.1768 & 478.8697 \end{bmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_M = \begin{bmatrix} 1352.5093 & 472.1065 & 975.9308 \\ 472.1065 & 632.7332 & 397.1339 \\ 975.9308 & 397.1339 & 1248.3534 \end{bmatrix}$$

5.4 Predicting Means

Predictions of mean responses were calculated for each combination of age and weight class for men and women. Two tables are shown for each response variable corresponding to each gender. The first (larger) table gives the predicted mean responses using the current age classes enforced by USAPL. The second (smaller) table gives the predicted mean response using three new suggested age classes. USAPL currently groups all adults of ages 24-39 together in one age class. I would recommend that they split this age class up into three smaller age classes, and the results using these three new age classes have been included. For the 84+ kilogram and 120+ kilogram weight classes, I used the average of the weights of all observations in the corresponding weight class, since there is no upper bound for these weight classes. For the women's 84+ kilogram weight class, 103 kilograms was used to make predictions, and for the men's 120+ kilogram weight class, 133 kilograms was used in making predictions.

The mean responses were predicted by $\widehat{\mathbf{Y}} = \mathbf{Z}\widehat{\boldsymbol{\beta}}$, where $\mathbf{Z}$ is the matrix of predictor variable values and $\widehat{\boldsymbol{\beta}}$ is the estimated parameter matrix obtained from SAS output. For example, if we wanted to predict mean responses for all age classes of women in the 43 kilogram weight class, we would use:

$$\widehat{\mathbf{Y}} = \mathbf{Z}\widehat{\boldsymbol{\beta}}_W = \begin{bmatrix} 1 & 43 & 31.5 \\ 1 & 43 & 42 \\ 1 & 43 & 47 \\ 1 & 43 & 52 \\ 1 & 43 & 57 \\ 1 & 43 & 62 \\ 1 & 43 & 67 \\ 1 & 43 & 72 \\ 1 & 43 & 77 \\ 1 & 43 & 82 \\ 1 & 43 & 27 \\ 1 & 43 & 32 \\ 1 & 43 & 37 \end{bmatrix} \begin{bmatrix} 117.6379 & 71.7612 & 139.3978 \\ 0.3552 & 0.1214 & 0.4731 \\ -1.0857 & -0.4387 & -0.9772 \end{bmatrix}$$

All of these calculations were done using Mathematica.

TABLE 5.4.1 Predictions for mean women's squat. Columns represent weight classes (kg) and the rows are age classes (years).

	43	47	52	57	63	72	84	84+
31.5	98.7	100.1	101.9	103.7	105.8	109.0	113.3	120.0
42	87.3	88.7	90.5	92.3	94.4	97.6	101.9	108.6
47	81.9	83.3	85.1	86.9	89.0	92.2	96.4	103.2
52	76.5	77.9	79.7	81.4	83.6	86.8	91.0	97.8
57	71.0	72.4	74.2	76.0	78.1	81.3	85.6	92.3
62	65.6	67.0	68.8	70.6	72.7	75.9	80.2	86.9
67	60.2	61.6	63.4	65.1	67.3	70.5	74.7	81.5
72	54.7	56.2	57.9	59.7	61.8	65.0	69.3	76.1
77	49.3	50.7	52.5	54.3	56.4	59.6	63.9	70.6
82	43.9	45.3	47.1	48.9	51.0	54.2	58.4	65.2

TABLE 5.4.2 Predictions for mean women's squat. Columns represent weight classes (kg) and the rows are age classes (years).

	43	47	52	57	63	72	84	84+
27	103.6	105.0	106.8	108.6	110.7	113.9	118.2	124.9
32	98.2	99.6	101.4	103.1	105.3	108.5	112.7	119.5
37	92.7	94.2	95.9	97.7	99.8	103.0	107.3	114.1

TABLE 5.4.3 Predictions for mean women's bench. Columns represent weight classes (kg) and the rows are age classes (years).

	43	47	52	57	63	72	84	84+
31.5	63.2	63.6	64.3	64.9	65.6	66.7	68.1	70.4
42	58.6	59.0	59.6	30.3	61.0	62.1	63.5	65.8
47	56.4	56.8	57.5	58.1	58.8	59.9	61.3	63.6
52	54.2	54.7	55.3	55.9	56.6	57.7	59.1	61.5
57	52.0	52.5	53.1	53.5	54.4	55.5	57.0	59.3
62	49.8	50.3	50.9	51.5	52.2	53.3	54.8	57.1
67	47.6	48.1	48.7	49.3	50.0	51.1	52.6	54.9
72	45.4	45.9	46.5	47.1	47.8	48.9	50.4	52.7
77	43.2	43.7	44.3	44.9	45.6	46.7	48.2	50.5
82	41.0	41.5	42.1	42.7	43.4	44.5	46.0	48.3

TABLE 5.4.4 Predictions for mean women's bench. Columns represent weight classes (kg) and the rows are age classes (years).

	43	47	52	57	63	72	84	84+
27	65.1	65.6	66.2	66.8	67.6	68.7	70.1	72.4
32	62.9	63.4	64.0	64.6	65.3	66.5	67.9	70.2
37	60.7	61.2	61.8	62.4	63.2	64.3	65.7	68.0

TABLE 5.4.5 Predictions for mean women's deadlift. Columns represent weight classes (kg) and the rows are age classes (years).

	43	47	52	57	63	72	84	84+
31.5	129.0	130.9	133.2	135.6	138.4	142.7	148.4	157.3
42	118.7	120.6	123.0	125.3	128.2	132.4	138.1	147.1
47	113.8	115.7	118.0	120.4	123.3	127.5	133.2	142.2
52	108.9	110.8	113.2	115.6	118.4	122.6	128.3	137.3
57	104.0	105.9	108.3	110.7	113.5	117.7	123.4	132.4
62	99.2	101.0	103.4	105.8	108.6	112.9	118.6	127.5
67	94.3	96.2	98.5	100.9	103.7	108.0	113.7	122.7
72	89.4	91.3	93.6	96.0	98.8	103.1	108.8	117.8
77	84.5	86.4	88.8	91.1	94.0	98.2	103.9	112.9
82	79.6	81.5	83.7	86.2	89.1	93.3	99.0	108.0

TABLE 5.4.6 Predictions for mean women's deadlift. Columns represent weight classes (kg) and the rows are age classes (years).

	43	47	52	57	63	72	84	84+
27	133.4	135.2	137.6	140.0	142.8	147.0	152.8	161.7
32	128.5	130.4	132.7	135.1	137.9	142.2	147.9	156.9
37	123.6	125.5	127.8	130.2	133.0	137.3	143.0	152.0

TABLE 5.4.7 Predictions for mean men's squat. Columns represent weight classes (kg) and the rows are age classes (years).

	53	59	66	74	83	93	105	120	120+
31.5	156.6	165.1	175.0	186.2	198.9	213.0	229.9	251.1	269.4
42	137.2	145.7	155.6	166.8	179.5	193.6	210.5	231.7	250.0
47	128.0	136.5	146.3	157.6	170.3	184.4	201.3	222.4	240.8
52	118.8	127.2	137.1	148.4	161.0	175.2	192.1	213.2	231.5
57	109.5	118.0	127.9	139.1	151.8	165.9	182.8	204.0	222.3
62	100.3	108.8	118.6	129.9	142.6	156.7	173.6	194.7	213.1
67	91.1	99.5	109.4	120.7	133.4	147.5	164.4	185.5	203.8
72	81.8	90.3	100.2	111.4	124.1	138.2	155.1	176.3	194.6
77	72.6	81.1	90.9	102.2	114.9	129.0	145.9	167.0	185.4
82	633	71.8	81.7	93.0	105.7	119.7	136.7	157.8	176.1

TABLE 5.4.8 Predictions for mean men's squat. Columns represent weight classes (kg) and the rows are age classes (years).

	53	59	66	74	83	93	105	120	120+
27	165.0	173.4	183.3	194.5	207.2	221.3	238.2	259.3	277.7
32	155.7	164.1	174.0	185.3	198.0	212.1	229.0	250.1	268.5
37	146.5	154.9	164.8	176.1	188.8	202.9	219.8	240.9	259.2

TABLE 5.4.9 Predictions for mean men's bench. Columns represent weight classes (kg) and the rows are age classes (years).

	53	59	66	74	83	93	105	120	120+
31.5	111.3	117.2	124.1	132.0	140.9	150.7	162.6	177.4	190.3
42	101.0	107.0	197.4	121.8	130.6	140.5	152.4	167.2	180.0
47	96.1	102.0	189.5	116.9	125.8	135.6	147.5	162.3	175.1
52	91.3	97.2	181.6	112.0	120.9	130.7	142.6	157.4	170.2
57	86.4	92.3	173.7	107.1	116.0	125.9	137.7	152.5	165.3
62	81.5	87.4	165.8	102.2	111.1	121.0	132.8	147.6	160.5
67	76.6	82.5	157.9	97.3	106.2	116.1	127.9	142.7	155.6
72	71.7	77.6	150.0	92.5	101.3	111.2	123.0	137.9	150.7
77	66.8	72.8	142.1	87.6	96.5	106.3	118.2	133.0	145.8
82	62.0	67.9	134.2	82.7	91.6	101.4	113.3	128.1	140.9

TABLE 5.4.10 Predictions for mean men's bench. Columns represent weight classes (kg) and the rows are age classes (years).

	53	59	66	74	83	93	105	120	120+
27	115.7	121.6	128.5	136.4	145.3	155.2	167.0	181.8	194.6
32	110.8	116.7	123.6	131.5	140.4	150.3	162.1	176.9	189.8
37	105.9	111.8	118.7	126.6	135.5	145.4	157.2	172.0	184.9

TABLE 5.4.11 Predictions for mean men's deadlift. Columns represent weight classes (kg) and the rows are age classes (years).

	53	59	66	74	83	93	105	120	120+
31.5	199.3	206.1	213.9	223.0	233.1	244.4	257.9	274.8	289.4
42	182.7	189.5	197.3	206.4	216.5	227.8	241.3	258.2	272.8
47	174.8	181.6	189.5	198.5	208.6	219.9	233.4	250.3	264.9
52	167.0	173.7	181.6	190.6	200.7	212.0	225.5	242.4	257.1
57	159.1	165.8	173.7	182.7	192.9	204.1	217.6	234.5	249.2
62	151.2	157.9	165.8	174.8	185.0	196.2	209.7	226.6	241.3
67	143.3	150.0	157.9	166.9	177.1	188.3	201.8	218.7	233.4
72	135.4	142.1	150.0	159.0	169.2	180.4	194.0	210.8	225.5
77	127.5	134.3	142.1	151.2	161.3	172.5	186.1	203.0	217.6
82	119.6	126.4	134.2	143.3	153.4	164.7	178.1	195.1	209.7

TABLE 5.4.12 Predictions for mean men's deadlift. Columns represent weight classes (kg) and the rows are age classes (years).

	53	59	66	74	83	93	105	120	120+
27	206.4	213.2	221.1	230.1	240.2	251.5	265.0	281.9	296.5
32	198.5	205.3	213.2	222.2	232.3	243.6	257.1	274.0	288.6
37	190.6	197.4	205.3	214.3	224.4	235.7	249.2	266.1	280.7

5.5 Prediction Intervals

Instead of predicting a range in which the mean of a certain predicted response is expected to fall, as we do when we construct confidence intervals, for this problem, we are interested in constructing an estimated interval in which future responses will fall, based on fixed values of the predictors. Let's say we want to calculate a 95% simultaneous prediction interval for a certain response variable. The interval would represent a range in which future observations will fall 95% of the time if the experiment is repeated many times.

This concept will help us set standards for a national powerlifting meet for raw competitors. The upper bound of 50% simultaneous prediction intervals have been calculated for all combinations of age and weight classes for both men and women. The actual age values used to calculate intervals were the average of the range of values in a particular age class. On the other hand, the body weight values used to calculate the intervals were the exact values of the weight class cutoffs. Usually, lifters try to be at the top end of his or her weight class, because, generally, one lifts more when he or she weighs more. To demonstrate the calculation of a prediction interval, imagine a 43-year-old man who weighs 98 kilograms. He falls into the 40-

44 year age class and the 105 kilogram weight class. Constructing a 50% simultaneous prediction interval for amount of weight squatted, using the men's data with values of predictors fixed at 42 years and 105 kilograms, we will get a range of values in which a future observation will fall 50% of the time. So, a future observation will be greater than the upper bound of the prediction interval 25% of the time. The choice to use 50% prediction intervals was subjective; the results obtained using these intervals can be used as standards in an inaugural competition, and based on the number of entrants, the standards can be raised or lowered according to what the organization deems appropriate. For a more prestigious meet, standards should be raised. (Use a wider interval – its upper bound will be greater.) For a less competitive meet, standards should be lowered. (Use a smaller interval – its upper bound will be lower.)

The prediction interval formula from section 4 was used to calculate the upper bounds:

$$\boldsymbol{z}_0\widehat{\boldsymbol{\beta}}_{(j)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right)F_{m,n-r-m}(\alpha)}\sqrt{(1+\boldsymbol{z}_0(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}\boldsymbol{z}_0{}')\left(\frac{n}{n-r-1}\hat{\sigma}_{jj}\right)} \quad [2]$$

The $\boldsymbol{z}_0$ vector takes the form $\boldsymbol{z}_0 = [1, z_{01}, z_{02}]$, where z_{01}, and z_{02} are body weight and age values respectively. Next, the $\widehat{\boldsymbol{\beta}}_{(j)}$ vector is the j^{th} column of the $\widehat{\boldsymbol{\beta}}$ matrix corresponding to the j^{th} response variable. The number of response variables is m, the number of predictor variables is r and n is the number of trials. Using our data, we have $m = 3, r = 2$, and $n = 296$ for men and $n = 92$ for women. The $F_{m,n-r-m}(\alpha)$ value is the upper $(100\alpha)^{\text{th}}$ percentile of an F distribution with

m and $n - r - m$ degrees of freedom [3]. In our case, $\alpha = 0.5$. For the men, we have $df_1 = 3$ and $df_2 = 291$, and for the women, $df_1 = 3$ and $df_2 = 87$. The respective F values are 0.7905 for men's data and 0.7949 for the women's. The $(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}$ is a $(r+1)\times(r+1)$ matrix. Our 3×3 $(\mathbf{Z}'\mathbf{Z})^{-\mathbf{1}}$ matrices were obtained from SAS output. Lastly, the $\hat{\sigma}_{jj}$ comes from the variance/covariance matrix estimated by $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\left(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}\right)$, where $\left(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}\right)$ is the sum of squares and cross products matrix obtained from SAS output.

The upper bounds of these 50% prediction intervals could be used as qualifying standards for a national meet. The tables presented below show the upper bound of a 50% prediction interval using all combinations of age and weight classes for a particular response variable, calculated using a function built in Mathematica. There are two tables for each response variable for each gender. The first and larger table gives the upper bound using the current age classes determined by USAPL. The second and smaller table gives the upper bound using three new suggested age classes. Currently, USAPL groups all adults of ages 24-39 together in one age class. I would propose that they split this age class up into three smaller age classes, and I have included results using these suggested age classes.

TABLE 5.5.1 Upper bounds of 50% simultaneous prediction intervals for women's squat. Columns represent weight classes (kg) and the rows are age classes (years).

	43	**47**	**52**	**57**	**63**	**72**	**84**	**84+**
31.5	106.1	107.5	109.2	111.0	113.1	116.3	120.6	127.6
42	94.71	96.1	97.8	99.6	101.7	104.9	109.2	116.1
47	89.3	90.7	92.4	94.2	96.3	99.5	103.8	110.4

52	83.9	85.3	87.0	88.8	90.9	94.1	98.4	105.3
57	78.6	79.9	81.7	83.4	85.5	88.7	93.0	99.9
62	73.2	74.6	76.3	78.1	80.2	83.4	87.7	94.5
67	67.9	69.2	71.0	72.7	74.8	78.0	82.3	89.2
72	62.5	63.95	65.6	67.4	69.5	72.7	77.0	83.8
77	57.2	58.6	60.3	62.1	64.2	67.4	71.6	78.5
82	51.9	53.3	55.0	56.8	58.9	62.0	66.3	73.2

TABLE 5.5.2 Upper bounds of 50% simultaneous prediction intervals for women's squat. Columns represent weight classes (kg) and the rows are age classes (years).

	43	**47**	**52**	**57**	**63**	**72**	**84**	**84+**
27	111.0	112.4	114.2	115.9	118.0	121.2	125.6	132.5
32	105.6	107.0	108.7	110.5	112.6	115.8	120.1	127.0
37	100.1	101.5	103.3	105.0	107.1	110.3	114.6	121.6

TABLE 5.5.3 Upper bounds of 50% simultaneous prediction intervals for women's bench. Columns represent weight classes (kg) and the rows are age classes (years).

	43	**47**	**52**	**57**	**63**	**72**	**84**	**84+**
31.5	69.2	69.7	70.3	70.9	71.6	72.7	74.2	76.7
42	64.6	65.1	65.7	66.3	67.0	68.1	69.6	72.0
47	62.5	62.9	63.5	64.1	64.8	65.9	67.4	69.8
52	60.3	60.8	61.3	61.9	62.6	63.7	65.2	67.7
57	58.2	58.6	59.2	59.8	60.5	61.6	63.1	65.5
62	56.0	56.5	57.1	57.6	58.3	59.4	60.9	63.3
67	53.9	54.4	54.9	55.5	56.2	57.3	58.8	61.2
72	51.8	52.2	52.8	53.4	54.1	55.2	56.7	59.1
77	49.7	50.1	50.7	51.3	52.0	53.1	54.6	57.0
82	47.6	48.1	48.6	49.2	49.9	51.0	52.5	54.9

TABLE 5.5.4 Upper bounds of 50% simultaneous prediction intervals for women's bench. Columns represent weight classes (kg) and the rows are age classes (years).

	43	**47**	**52**	**57**	**63**	**72**	**84**	**84+**
27	71.2	71.7	72.3	72.9	73.6	74.7	76.2	78.7
32	69.0	69.5	70.1	70.7	71.4	72.5	74.0	76.4
37	66.8	67.3	67.9	68.5	69.2	70.3	71.8	74.2

TABLE 5.5.5 Upper bounds of 50% simultaneous prediction intervals for women's deadlift. Columns represent weight classes (kg) and the rows are age classes (years).

	43	**47**	**52**	**57**	**63**	**72**	**84**	**84+**
31.5	136.6	138.4	140.8	143.1	145.9	150.2	155.9	165.1
42	126.3	128.2	130.5	132.8	135.7	139.9	145.6	154.8
47	121.4	123.3	125.6	128.0	130.8	135.0	140.8	149.9

52	116.6	118.4	120.8	123.1	125.9	130.2	135.9	145.1
57	111.8	113.6	115.9	118.3	121.1	125.4	131.1	140.2
62	107.0	108.8	111.1	113.5	116.3	120.5	126.2	135.4
67	102.2	104.0	106.3	108.7	111.5	115.7	121.4	130.6
72	97.4	99.2	101.6	103.9	106.7	110.9	116.6	125.8
77	92.6	94.5	96.8	99.1	101.9	106.2	111.9	121.0
82	97.8	89.7	92.0	94.4	97.2	101.4	107.1	116.2

TABLE 5.5.6 Upper bounds of 50% simultaneous prediction intervals for women's deadlift. Columns represent weight classes (kg) and the rows are age classes (years).

	43	**47**	**52**	**57**	**63**	**72**	**84**	**84+**
27	141.0	142.8	145.2	147.5	150.3	154.6	160.3	169.5
32	136.1	137.9	140.3	142.6	145.4	149.7	155.4	164.6
37	131.2	133.0	135.4	137.7	140.5	144.8	150.5	159.7

TABLE 5.5.7 Upper bounds of 50% simultaneous prediction intervals for men's squat. Columns represent weight classes (kg) and the rows are age classes (years).

	53	**59**	**66**	**74**	**83**	**93**	**105**	**120**	**120+**
31.5	166.1	174.6	184.4	195.7	208.3	222.4	239.3	260.5	278.8
42	146.7	155.2	165.0	176.3	188.9	203.0	219.9	241.1	259.4
47	137.5	145.9	155.8	167.0	179.7	193.8	210.7	231.8	250.2
52	128.3	136.7	146.5	157.8	170.5	184.5	201.5	222.6	241.0
57	119.0	127.5	137.3	148.6	161.2	175.3	192.2	213.4	231.8
62	109.8	118.2	128.1	139.3	152.0	166.1	183.0	204.2	222.5
67	100.6	109.0	118.9	130.1	142.8	156.9	173.8	195.0	213.3
72	91.4	99.8	109.7	120.9	133.6	147.7	164.6	185.8	204.1
77	82.2	90.6	100.5	111.7	124.4	138.5	155.4	176.6	194.9
82	73.0	81.4	91.2	102.5	115.2	129.3	146.2	167.4	185.7

TABLE 5.5.8 Upper bounds of 50% simultaneous prediction intervals for men's squat. Columns represent weight classes (kg) and the rows are age classes (years).

	53	**59**	**66**	**74**	**83**	**93**	**105**	**120**	**120+**
27	174.5	182.9	192.7	204.0	216.7	230.7	247.7	268.8	287.2
32	165.2	173.6	183.5	194.7	207.4	221.5	238.4	259.6	277.9
37	156.0	164.4	174.2	185.5	198.2	212.2	229.2	250.3	268.7

TABLE 5.5.9 Upper bounds of 50% simultaneous prediction intervals for men's bench. Columns represent weight classes (kg) and the rows are age classes (years).

	53	**59**	**66**	**74**	**83**	**93**	**105**	**120**	**120+**
31.5	119.1	125.0	131.9	139.8	148.7	158.5	170.4	185.2	198.1
42	108.9	114.8	121.7	129.5	138.4	148.3	160.1	174.9	187.8
47	104.0	109.9	116.8	124.7	133.5	143.4	155.2	170.0	182.9

52	99.1	105.0	111.9	119.8	128.6	138.5	150.4	165.2	178.0
57	94.2	100.1	107.0	114.9	123.8	133.6	145.5	160.3	173.2
62	89.3	95.2	102.1	110.0	118.9	128.8	140.6	155.4	168.3
67	84.5	90.4	97.3	101.2	114.0	123.9	135.7	150.6	163.4
72	79.6	85.5	92.4	100.3	109.2	119.0	130.9	145.7	158.6
77	74.7	80.6	87.5	95.4	104.3	114.2	126.0	140.8	153.7
82	69.9	75.8	82.7	90.6	99.4	109.3	121.2	136.0	148.8

TABLE 5.5.10 Upper bounds of 50% simultaneous prediction intervals for men's bench. Columns represent weight classes (kg) and the rows are age classes (years).

	53	**59**	**66**	**74**	**83**	**93**	**105**	**120**	**120+**
27	123.5	129.5	136.3	144.2	153.1	163.0	174.8	189.6	202.5
32	118.6	124.6	131.4	139.3	148.2	158.1	169.9	184.7	197.6
37	113.8	119.7	126.6	134.4	143.3	153.2	165.0	179.8	192.7

TABLE 5.5.11 Upper bounds of 50% simultaneous prediction intervals for men's deadlift. Columns represent weight classes (kg) and the rows are age classes (years).

	53	**59**	**66**	**74**	**83**	**93**	**105**	**120**	**120+**
31.5	208.6	215.3	223.2	232.2	242.3	253.6	267.1	284.0	298.7
42	192.0	198.8	206.6	215.6	225.7	237.0	250.5	267.4	282.1
47	184.1	190.9	198.7	207.7	217.8	229.1	242.6	259.5	274.2
52	176.2	183.0	190.8	199.8	210.0	221.2	234.7	251.6	266.3
57	168.4	175.1	183.0	191.9	202.1	213.3	226.8	243.8	258.4
62	160.5	167.2	175.1	184.1	194.2	205.5	219.0	235.9	250.6
67	152.6	159.3	167.2	176.2	186.3	197.6	211.1	228.0	242.7
72	144.7	151.5	159.3	168.3	178.5	189.7	203.2	220.1	234.8
77	136.9	143.3	151.5	160.5	170.6	181.8	195.3	212.3	227.0
82	129.0	135.7	143.6	152.6	162.7	174.0	187.5	204.4	219.1

TABLE 5.5.12 Upper bounds of 50% simultaneous prediction intervals for men's deadlift. Columns represent weight classes (kg) and the rows are age classes (years).

	53	**59**	**66**	**74**	**83**	**93**	**105**	**120**	**120+**
27	215.7	222.5	230.3	239.3	249.4	260.7	274.2	291.1	305.8
32	207.8	214.6	222.4	231.4	241.5	252.8	266.3	283.2	297.9
37	199.9	206.7	214.5	223.5	233.6	244.9	258.4	275.3	290.0

6 Final Conclusions

To establish a competitive inaugural national meet for raw powerlifters, I would recommend using the qualifying standards obtained by calculating the upper bounds of 50% simultaneous prediction intervals. Calculating larger or smaller $100(1-\alpha)\%$ prediction intervals would aid in establishing standards for more or less prestigious competitions, respectively. These values would be guidelines in which to start from and could be modified depending on how well they worked after being used.

7 Recommendations for Further Research

Multivariate multiple regression has many applications relating to athletics. There are lots of situations in which it would be helpful to build a multivariate regression model to help establish new or modify existing qualifying standards for a particular sport. It would be interesting to use the prediction interval method to obtain standards for a certain event with qualifying standards already in place and compare the results from the multivariate regression with the existing standards. Using the upper bound of a prediction interval doesn't have to be a method that provides standards that are set in stone. Rather, they could act as guidelines to work from that could be changed on a case-by-case basis.

Regarding our particular powerlifting problem, it would have been nice to have a larger sample size. Obtaining data from powerlifting meets other than Raw Nationals would improve our model. By collecting data from smaller meets across the country, we would have a larger sample size, and we would have a better sample from the population. As was mentioned before, anyone is allowed to compete at Raw Nationals, but there are factors involved in determining which lifters actually compete each year. Location is an important determining factor, as lifters closer to the meet location will most likely be more likely to choose to compete. Travel costs may play a role in controlling whether or not a lifter will enter Raw Nationals. It may also be the case that there are more upper-skill level powerlifters than lower-skill level lifters at Raw Nationals each year. Some of the lifters that aren't as strong may be intimidated by a national meet, even though anyone is allowed to compete. If this is the case, our standards may have been set too high. Multivariate multiple regression could be used yearly on powerlifting data to analyze and keep up with performance trends.

A lot additional time could be allocated to look into these sorts of situations. The problem in this paper was addressed appropriately based on the limited time and resources available.

References

[1] *International Powerlifting Federation*. Web. June 2013. <http://www.powerlifting-ipf.com/Downloads.55.0.html>

[2] Johnson, Richard A. and Wichern, Dean W. *Applied Multivariate Statistical Analysis*. Upper Saddle River: Pearson Education Inc. 2007.

[3] Montgomery, Douglas C., Peck, Elizabeth A., and Vining, Geoffrey, G. *Introduction to Linear Regression Analysis.* Hoboken: John Wiley & Sons, Inc. 2006.

[4] *Multivariate Analysis with SAS*. Karl L. Wuensch. 23 Aug. 2012. East Carolina University Department of Psychology. June 2012. <http://core.ecu.edu/psyc/wuenschk/sas/sas-mv.htm>

[5] "Powerlifting." *Wikipedia, The Free Encyclopedia*. Wikimedia Foundation, Inc. Web. July 2013. <http://en.wikipedia.org/wiki/Powerlifting>

[6] *SAS Data Analysis Examples*. UCLA Academic Technology Services. Web. March 2013. <http://www.ats.ucla.edu/stat/sas/dae/mvreg.htm>

[7] *USA Powerlifting*. Tony Cardella. USAPL/ADFPA. Web. August 2012. <http://www.usapowerlifting.com/>

[8] *USA Track and Field*. USA Track and Field, Inc. Web. June 2013. <http://www.usatf.org/Home.aspx>

Appendix

A.1 LEMMA 4: $\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is symmetric [3]

Observe: $(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')' = \mathbf{I}' - (\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')' = \mathbf{I} - (\mathbf{Z}')'((\mathbf{Z}'\mathbf{Z})^{-1})'\mathbf{Z}' = \mathbf{I} - \mathbf{Z}((\mathbf{Z}'\mathbf{Z})')^{-1}\mathbf{Z}' = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'(\mathbf{Z}')')^{-1}\mathbf{Z}' = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Since $(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')' = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is symmetric, as desired.

A.2 LEMMA 5: $\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is idempotent [3]

Observe: $(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}') = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{I} - 2\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ (since $\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} = \mathbf{I}$).

Now we have $\mathbf{I} - 2\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Since $(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}') = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, $\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is idempotent, as desired.

A.3 Obtaining the $\left(\mathbf{Z}\boldsymbol{\beta}_{(j)}\right)'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)}$ correction term

Start with $\mathbf{Y}_{(j)}{}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Y}_{(k)}$.

Let $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Now observe: $\mathbf{Y}_{(j)}{}'(\mathbf{I} - \mathbf{H})\mathbf{Y}_{(k)} = (\mathbf{Y}_{(j)}{}' - \mathbf{Y}_{(j)}{}'\mathbf{H})\mathbf{Y}_{(k)} = \mathbf{Y}_{(j)}{}'\mathbf{Y}_{(k)} - \mathbf{Y}_{(j)}{}'\mathbf{H}\mathbf{Y}_{(k)}$.

Next we have $(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{H})\left(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)}\right) = (\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'\left(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)} - \mathbf{H}\mathbf{Y}_{(k)} + \mathbf{H}\mathbf{Z}\boldsymbol{\beta}_{(k)}\right) = (\mathbf{Y}_{(j)}{}' - \left(\mathbf{Z}\boldsymbol{\beta}_{(j)}\right)')\left(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)} - \mathbf{H}\mathbf{Y}_{(k)} + (\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Z}\boldsymbol{\beta}_{(k)}\right) =$

$(\mathbf{Y}_{(j)}{}' - (\mathbf{Z}\boldsymbol{\beta}_{(j)})')(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)} - \mathbf{H}\mathbf{Y}_{(k)} + \mathbf{Z}\boldsymbol{\beta}_{(k)}) = (\mathbf{Y}_{(j)}{}' - (\mathbf{Z}\boldsymbol{\beta}_{(j)})')(\mathbf{Y}_{(k)} - \mathbf{H}\mathbf{Y}_{(k)}) = \mathbf{Y}_{(j)}{}'\mathbf{Y}_{(k)} - \mathbf{Y}_{(j)}{}'\mathbf{H}\mathbf{Y}_{(k)} - (\mathbf{Z}\boldsymbol{\beta}_{(j)})'\mathbf{Y}_{(k)} + (\mathbf{Z}\boldsymbol{\beta}_{(j)})'\mathbf{H}\mathbf{Y}_{(k)} = \mathbf{Y}_{(j)}{}'\mathbf{Y}_{(k)} - \mathbf{Y}_{(j)}{}'\mathbf{H}\mathbf{Y}_{(k)} - (\mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(k)} - \mathbf{H}\mathbf{Y}_{(k)}) = \mathbf{Y}_{(j)}{}'\mathbf{Y}_{(k)} - \mathbf{Y}_{(j)}{}'\mathbf{H}\mathbf{Y}_{(k)} - (\mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{H})\mathbf{Y}_{(k)}$.

Since $\mathbf{Y}_{(j)}{}'(\mathbf{I} - \mathbf{H})\mathbf{Y}_{(k)} = \mathbf{Y}_{(j)}{}'\mathbf{Y}_{(k)} - \mathbf{Y}_{(j)}{}'\mathbf{H}\mathbf{Y}_{(k)}$, $(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{H})(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)}) = \mathbf{Y}_{(j)}{}'(\mathbf{I} - \mathbf{H})\mathbf{Y}_{(k)} - (\mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{H})\mathbf{Y}_{(k)}$ or $(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{H})(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)}) + (\mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{H})\mathbf{Y}_{(k)} = \mathbf{Y}_{(j)}{}'(\mathbf{I} - \mathbf{H})\mathbf{Y}_{(k)}$.

So when we subtract $\mathbf{Z}\boldsymbol{\beta}_{(j)}$ from $\mathbf{Y}_{(j)}$ and $\mathbf{Z}\boldsymbol{\beta}_{(k)}$ from $\mathbf{Y}_{(k)}$, we must add on the correction term $(\mathbf{Z}\boldsymbol{\beta}_{(j)})'(\mathbf{I} - \mathbf{H})\mathbf{Y}_{(k)}$.

A.4 LEMMA 6: $\mathrm{Tr}[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = r + 1$ [3]

First, observe that $(\mathbf{Z}'\mathbf{Z})^{-1}$ is symmetric: $((\mathbf{Z}'\mathbf{Z})^{-1})' = ((\mathbf{Z}'\mathbf{Z})')^{-1} = (\mathbf{Z}'(\mathbf{Z}')')^{-1} = (\mathbf{Z}'\mathbf{Z})^{-1}$.

Also note that since $\mathbf{Z}$ is an $n\times(r+1)$ matrix, $(\mathbf{Z}'\mathbf{Z})^{-1}$ has dimension $(r+1)\times(r+1)$.

To show that $\mathrm{tr}[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = r + 1$, we will use the property that says $\mathbf{X}'\mathbf{A}\mathbf{Y} = \mathrm{tr}(\mathbf{X}'\mathbf{A}\mathbf{Y}) = \mathrm{tr}(\mathbf{A}\mathbf{Y}\mathbf{X}')$ for a symmetric matrix $\mathbf{A}$ and matrices $\mathbf{X}$ and $\mathbf{Y}$.

So for the symmetric matrix $(\mathbf{Z}'\mathbf{Z})^{-1}$,

$\mathrm{tr}[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = \mathrm{tr}[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}] = \mathrm{tr}[\mathbf{I}_{(r+1)\times(r+1)}] = r + 1$, as desired.

A.5 Proof of THEOREM 5: $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{(j)}, \hat{\boldsymbol{\varepsilon}}_{(k)}) = \mathbf{0}$ [5]

Observe: $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{(j)}, \hat{\boldsymbol{\varepsilon}}_{(k)}) = \mathrm{E}\left(\left(\widehat{\boldsymbol{\beta}}_{(j)} - \mathbf{E}(\widehat{\boldsymbol{\beta}}_{(j)})\right)\left(\hat{\boldsymbol{\varepsilon}}_{(k)} - \mathbf{E}(\hat{\boldsymbol{\varepsilon}}_{(k)})\right)'\right) = \mathrm{E}\left((\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta}_{(j)})(\hat{\boldsymbol{\varepsilon}}_{(k)} - \mathbf{0})'\right) = \mathrm{E}\left((\widehat{\boldsymbol{\beta}}_{(j)} - \boldsymbol{\beta}_{(j)})(\mathbf{Y}_{(k)} - \widehat{\mathbf{Y}}_{(k)})'\right) =$
$\mathrm{E}\left(((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(j)} - \boldsymbol{\beta}_{(j)})(\mathbf{Y}_{(k)} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(k)})'\right)$

Next, let $\mathbf{c} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$.

Now we have

$\mathrm{E}\left((\mathbf{cY}_{(j)} - \boldsymbol{\beta}_{(j)})(\mathbf{Y}_{(k)} - \mathbf{ZcY}_{(k)})'\right) = \mathrm{E}\left((\mathbf{cY}_{(j)} - \boldsymbol{\beta}_{(j)})((\mathbf{I} - \mathbf{Zc})\mathbf{Y}_{(k)})'\right) = \mathrm{E}\left((\mathbf{cY}_{(j)} - \boldsymbol{\beta}_{(j)})(\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})')\right) = \mathrm{E}\left((\mathbf{cY}_{(j)} - \boldsymbol{\beta}_{(j)})\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right)$ since $(\mathbf{I} - \mathbf{Zc})$ is idempotent (see A.1).

Continuing, $\mathrm{E}\left((\mathbf{cY}_{(j)} - \boldsymbol{\beta}_{(j)})\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) = \mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc}) - \boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) = \mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) - \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right)$.

Observe that $\mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) = \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}' - \boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}'\mathbf{Zc}\right) = \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}' - \boldsymbol{\beta}_{(j)}(\mathbf{Z}\boldsymbol{\beta}_{(k)})'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\right) = \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}' - \boldsymbol{\beta}_{(j)}\boldsymbol{\beta}_{(k)}{}'\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\right) = \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}' - \boldsymbol{\beta}_{(j)}\boldsymbol{\beta}_{(k)}{}'\mathbf{Z}'\right) = \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}' - \boldsymbol{\beta}_{(j)}(\mathbf{Z}\boldsymbol{\beta}_{(k)})'\right) = \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}' - \boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}'\right) = \mathbf{0}$.

So now we are concerned with $\mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right)$.

Take $\mathrm{E}\left(\mathbf{c}(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)})'(\mathbf{I} - \mathbf{Zc})\right)$

$$\begin{aligned}
&= \mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) - \mathbf{c}\mathrm{E}\left(\mathbf{Y}_{(j)}\boldsymbol{\beta}_{(k)}{}'\right)\mathbf{Z}'(\mathbf{I} - \mathbf{Zc}) - \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\mathbf{Y}_{(k)}{}'\right)(\mathbf{I} - \mathbf{Zc}) \\
&\qquad + \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\boldsymbol{\beta}_{(k)}{}'\right)\mathbf{Z}'(\mathbf{I} - \mathbf{Zc}) \\
&= \mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) - \mathbf{c}\mathrm{E}\left(\mathbf{Y}_{(j)}\boldsymbol{\beta}_{(k)}{}'\right)(\mathbf{Z}' - \mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}') - \mathbf{0} \\
&\qquad + \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\boldsymbol{\beta}_{(k)}{}'\right)(\mathbf{Z}' - \mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}') \\
&= \mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) - \mathbf{c}\mathrm{E}\left(\mathbf{Y}_{(j)}\boldsymbol{\beta}_{(k)}{}'\right)(\mathbf{Z}' - \mathbf{Z}') \\
&\qquad + \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\boldsymbol{\beta}_{(k)}{}'\right)(\mathbf{Z}' - \mathbf{Z}') \\
&= \mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) - \mathbf{c}\mathrm{E}\left(\mathbf{Y}_{(j)}\boldsymbol{\beta}_{(k)}{}'\right)(\mathbf{0}) + \mathrm{E}\left(\boldsymbol{\beta}_{(j)}\boldsymbol{\beta}_{(k)}{}'\right)(\mathbf{0}) \\
&= \mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right)
\end{aligned}$$

So, $\mathrm{E}\left(\mathbf{cY}_{(j)}\mathbf{Y}_{(k)}{}'(\mathbf{I} - \mathbf{Zc})\right) = \mathrm{E}\left(\mathbf{c}(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)})'(\mathbf{I} - \mathbf{Zc})\right)$ with no correction terms needed.

Continue with $\mathrm{E}\left(\mathbf{c}(\mathbf{Y}_{(j)} - \mathbf{Z}\boldsymbol{\beta}_{(j)})(\mathbf{Y}_{(k)} - \mathbf{Z}\boldsymbol{\beta}_{(k)})'(\mathbf{I} - \mathbf{Zc})\right) = \mathbf{c}\mathrm{E}\left(\boldsymbol{\varepsilon}_{(j)}\boldsymbol{\varepsilon}_{(k)}{}'\right)(\mathbf{I} - \mathbf{Zc}) = \mathbf{c}(\mathbf{0})(\mathbf{I} - \mathbf{Zc}) = \mathbf{0}$, as desired.