MEMORANDUM

To: Colleen Burns

From: Ben Laufer, Lamitr Dhir, Roee Morag, Jacob Perez

Date: May 15, 2024

Re: Body Composition and Anthropometric Data Analysis Recommendation

---

The purpose of this memo is to describe the statistical methods and findings from an analysis of your anthropometric and body composition data. We hope that this information helps you address your research question:

> *"What are the key health status differences between ethnic Greek/Christian minority Schools in Istanbul, Turkey: Zappeio Lyceum, Zographeio Lyceum, and Megali Scholi (Phanar Greek Orthodox Lyceum)."*

This memo is organized into four sections.

I. **Abstract of Key Findings** – an overview of key results from the analysis.
II. **Background and Data** – a summary of our understanding of your research questions and basic descriptive statistics to get an overview of your data including variables measured and how the data was collected.
III. **Statistical Methods** – a description of the models and methods.
IV. **Results and Discussion** – numerical and graphical summaries, interpretation of results, and limitations
V. **Technical Output** – varied computer output for reference

If you have any additional questions about this work following our consulting meeting today, please feel free to contact us at nlaufer@calpoly.edu so that we may set up another meeting to discuss your questions.

I. Abstract

We present an analysis of health status variables among a cohort of students across three separate schools in Istanbul, Turkey, using data collected in 2015. Our analysis aimed to understand how demographic factors (school) affect anthropometric and body composition measurements.

Utilizing Linear Mixed-Effects Models, we analyzed both the anthropometric and body composition datasets. For the first dataset (anthropometric), we found that arm lengths greatly differ across all three schools as well as the difference in waist measurements between waist 1 and 2. For the second dataset (body composition) we found significant differences in overall fat percent between Zographeion and Megali Sxol as well as a few other key differences in fat percent and total mass throughout individual ligament components and the trunk of students. On the other hand, from the use of Principal Component Analysis, we found that there aren't many other aspects that can be explained by the school effect of the data. Although there were a few significant differences (such as arm length) that could be explained by the school level, it's hard to say what other measurements can be explained by the school level due to the dependence between the variables.

However, the study's limitations—including sampling population, removal of data collected, and correlation of measurements—must be considered, as they could affect the interpretation and generalizability of these findings. Despite these constraints, the results offer valuable insights into the health status of ethnic Greek/Christian minority schools in Istanbul. Additionally, the normality assumption for our model is suboptimal and some transformations may be more helpful.

II. Background and Key Data

The nutritional epidemiological transition in Turkey has changed the way that their citizens eat and the food they are consuming. Our clients have been investigating this trend using anthropometric and body composition data from three Rum (ethnic Greek/Christian Orthodox) schools in Istanbul, hoping to draw future conclusions to alarming rates of cardiovascular disease (CVD) in Turkey.

With regards to this memorandum, we will not be focusing on the potential connection between the anthropometric makeup of the three Rum schools and CVD; instead we are looking to discover if there exists a difference in health status assessments, as quantified by both anthropometric and body composition variables, between the three schools. The three schools are Zappeio Lyceum, Zographeio Lyceum, and Megali Scholi.

The anthropometric data set consists of various measurements of the human individual. This includes height, weight, hip measurements, BMI, etc. Some variables of interest that aren't as commonplace include skinfolds and dynamometer. The main purpose of skinfold measurements is to estimate general fatness in individuals. The dynamometer variables are a measure of muscle strength.

The body composition data set includes different measurements that look more into proportions of different components that make up an individual's total weight. Impedance, measured with an omega, refers to the opposition that the body's tissues offer to the flow of an electrical current. There are additional segmental analyses that look at the specific muscle and fat measurements of different body parts.

Figure 1: Body Mass Index Compared Across Schools

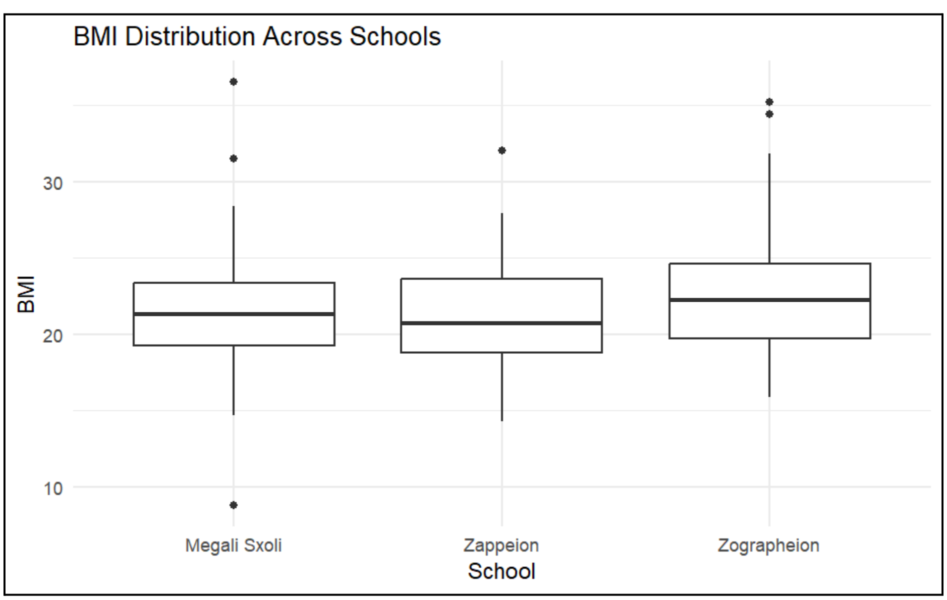
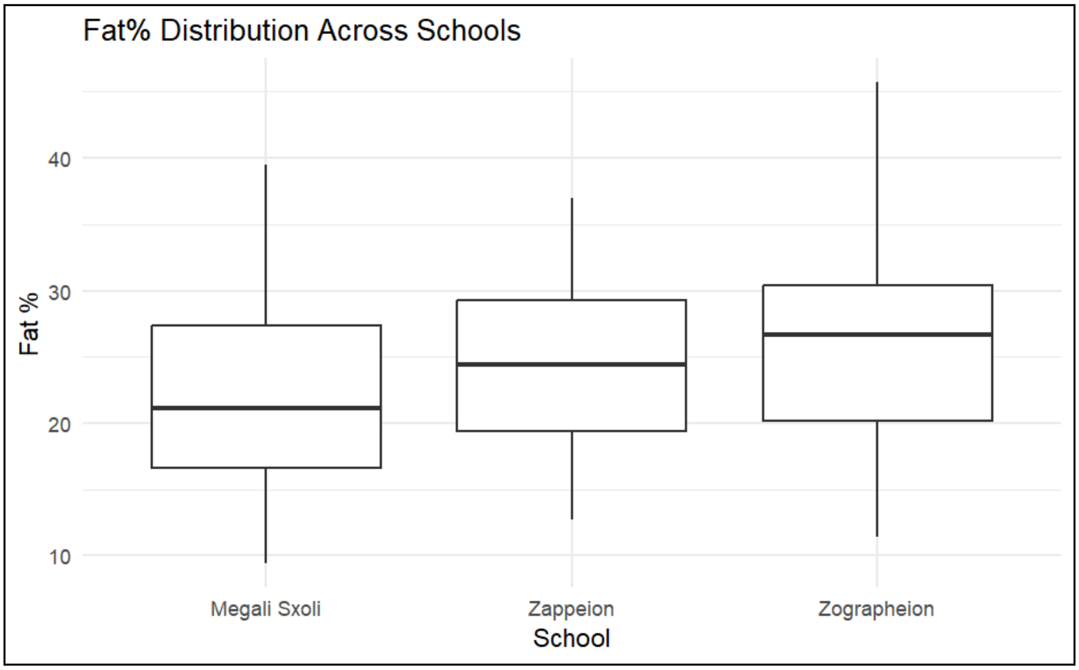


Figure 2: Fat Percentage Compared Across Schools



III. Statistical Methods

Linear Mixed-Effects Models

Linear mixed-effects modeling is the primary statistical method used to analyze both the body composition and anthropic datasets. The primary uses for this analysis are when working with clustered data, which is essentially data with groups, in the context of our study, each school is a cluster. The analysis also allows us to keep our results focused and concise, while factoring in all observations. Additionally, linear mixed-effects modeling also lets us treat some variables as "random effects", which means that those variables, although in the data, are not our primary variable of interest and that we can treat them as random. This ensures that despite the difference in the variable(s) that are being treated as random effects, we can still judge the variability between each individual. This is also why we chose to address this analysis through linear mixed-effects models rather than MANOVA (ANOVA for multiple dependent variables), as using random effects with linear modeling is simpler.

Due to the clustering by school in our data, for this memorandum, we chose to apply the same linear mixed-effects model for both the body composition and anthropic datasets, as it was the best fit. For our model, and to explore the differences in the 3 schools, we set school as the sole explanatory variable, while setting age and gender as random effects. We used this model for all response variables. The following assumptions have been met for this model: linearity, independence, homoscedasticity, and normality.

$Y_{ij} = \beta_0 + \beta_1 x_1 + u_{0i} + u_{1j} + \varepsilon$
x_1 = school
u_0 = random effect for age
u_1 = random effect for gender
ε = residual error

Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method used to reduce the dimensionality of large data sets by identifying its directions, called principal components, that explain the most variance in the data. By transforming the data into a lower-dimensional space defined by these components, PCA simplifies complex datasets while retaining important information about the variation present in the original variables.

For this memorandum, we implemented principle component analysis in order to visualize the differences, or lack thereof, in the anthropometric and body composition makeup between the three schools.

IV. Results and Discussion

To analyze the differences in anthropometric variables and body demographic variables between demographic groups, namely the three schools Megali Scholi, Zappeio Lyceum, and Zographeio Lyceum, we ran two separate analyses. Both analyses were done for the anthropometric variables as well as for the body composition variables.

<u>Analysis # 1 (Linear Mixed-Effects Model)</u>

This initial analysis showcases the results of our mixed-effects model, using random slopes for both gender and age. Using random slopes allows us to, in a sense, ignore differences between ages and genders that we may see in the anthropometric and body composition data and focus on the differences between the schools. The random slopes treat both age and gender as random effects that can differ at each school. Treating these effects on the anthropometrics and body composition as random helps to find differences

Table 1: Significant Anthropometric Differences Between Schools

	Anthropometrics	Created Variables
School	Arm length (cm)	Waist Diff
Zappeion	1.6923**	-0.7064
Zographeion	2.5484**	-1.6739**
Meg. Sholi	Ref. Group	--

** Values are Significant

Looking first at the anthropometric variables table some of our key findings include:

- **Arm Length:** a significant difference in arm lengths between the schools when compared to Megali Scholi
    - Zographeio Lyceum has a significantly higher arm length than Megali Scholi as evidenced by the fixed effect of 2.548
    - Zappeio Lyceum has a significantly higher arm length than Megali Scholi as evidenced by the fixed effect of 1.692
- **Waist Difference:** a significant difference in waist difference was found when compared to Megali Scholi

- Zographeio Lyceum has a significantly lower waist difference than Megali Scholi as evidenced by the negative fixed effect of -1.674
- **Notable Variables:**
    - No differences were found between **Height, BMI, Weight, Skinfolds, and Dynamometers**

Possible differences between arm length and our created variable waist difference indicate a difference between these schools for the listed anthropometrics.

We then ran the same analysis on the body composition data and created the following table of our results:

Table 2: Significant Differences in Body Composition Between Schools

			Impedance (Ω)		Right Leg	Right Arm	Left Arm	Trunk	
School	Fat %	Fat Mass (kg)	Right Arm (Ω)	Left Arm (Ω)	Fat%	Fat%	Fat%	Fat%	Fat Mass (kg)
Zappeion	2.678	1.542	33.91**	40.35**	0.552	2.556	3.165	3.90**	1.1471
Zographeion	4.341**	3.64**	24.23	27.68	3.71**	4.357**	5.31**	4.686**	1.9379**
Meg. Sholi	Ref. Group	--	--	--	--	--	--	--	--

** Values are Significant

Looking at our body composition table we have found more differences in the schools when compared to our anthropometrics table. Some of the notable differences highlighted in the table are:

- **Zographeio vs. Megali:** the body composition between Zographeio Lyceum and Megali Scholi differed in several major categories
    - **Fat %:** Zographeio Lyceum was found to have a significantly higher fat percentage when compared to Megali Scholi, as displayed by the fixed effect of 4.341 from the table
    - **Fat Mass:** Again Zographio Lyceum was found to have a significantly higher fat mass in the population than Megali Scholi (fixed effect of 3.643)

        - Overall we can see that comparing total fat % and fat mass across the whole body between Zographeio Lyceum and Megali Scholi, Zographeio is higher
    - In the rest of the table, we can see that Zographeio corresponds to a higher:
        - **Fat % in Right Leg**
        - **Fat % in Right Arm**
        - **Fat % Right Arm**
        - **Fat % Left Arm**
        - **Fat % Trunk**
        - **Fat Mass Trunk**
    - This is not too surprising seeing as the overall fat percentage and fat mass for the whole body were found to be significantly different however, something of note as there are clear differences we are observing in fat percentage and mass between the schools
- **Right Arm and Left Arm Impedance:** for both the right and left arm Zappeio Lyceum is associated with a higher impedance measured in omega
    - The right arm impedance of Zappeio Lyceum has a fixed effect of 33.91 when compared to Megali Scholi
    - The left arm impedance of Zappeio Lyceum has a fixed effect of 40.35 when compared to Megali Scholi

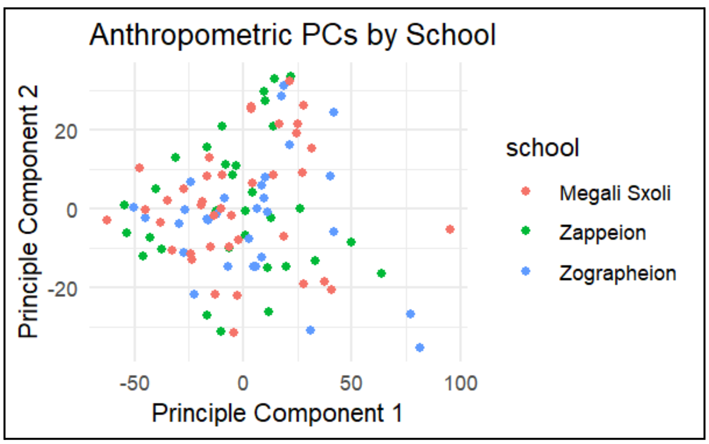
The most notable differences between demographics came from our primary analysis of body composition. We can see clear differences between the fat percentage and fat mass of Zographeio Lyceum and Megali Scholi. This may indicate differences in how the diets have affected both schools however we cannot make this conclusion based on our analysis. There was also an observed difference between the impedance levels of Zappeio Lyceum and Megali Scholi in both arms. This may also indicate a difference in fat percentages in both arms however we cannot conclude with our analysis.

Finally, from our linear mixed model effects results we believe there is a difference in the fat composition in Megali Scholi that would be worth further investigating or inferring.

Analysis # 2 (Principal Component Analysis)

In our second analysis, we implemented principle component analysis on the two data sets separately, similarly to how we ran our linear mixed effect model. We first ran principle component analysis on the anthropometric data and then separately on the body composition data.

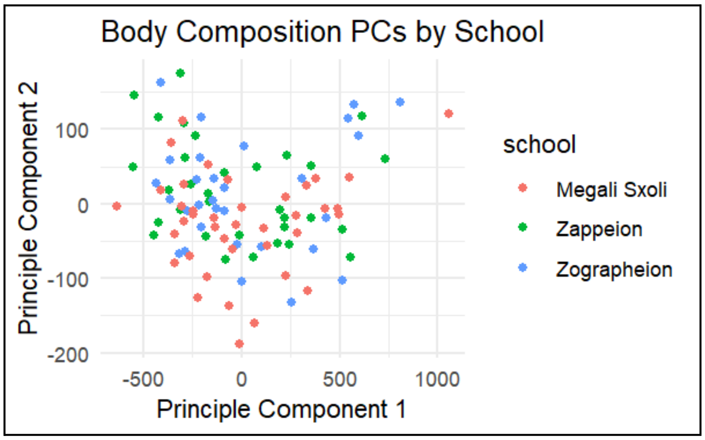
For the anthropometric dataset, the first two principal components account for 90.4% of the variance in the original data. The figure below is a scatter plot of the first two principal components, with the points colored by school displaying how the components capture the variation between the schools. Ideally, there are clear clusters between the groups that the principal component analysis can discover. Yet, looking at our figure below we cannot distinguish any clear clusters or groups that may be school-related.



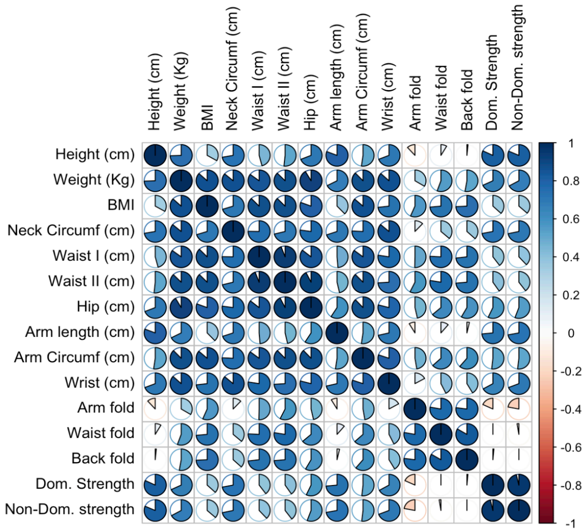
We could not find any differences between the anthropometric variables between the schools as a result of our principal component analysis on the anthropometric data set.

When running our principal component analysis on the body composition dataset we found similar results to our findings of the anthropometric dataset. Again we have provided a scatterplot below displaying the components of the body composition that are captured by the schools. No distinct grouping can be seen from this analysis of principal component analysis as well, but we can see there is a group of red, representing Megali Scholi, near the bottom left of the middle group. We believe this to be consistent with our linear mixed model effects analysis of the body composition of the schools. One of the biggest findings was the lower fat composition of Megali Scholi which seems to be represented slightly in the clustering of red at the bottom of the scatterplot.

Similarly to our first analysis, there seems to be more to explore in relation to the body composition in Megali Scholi when compared to the other schools. Our analysis using PCA did not help us to find any other relationships we had not already found using our linear mixed models.



Finally, one of our reasons for running both PCA and the linear mixed model effects was because of the high correlation and possible dependence of some of the variables in the study. We created a correlation plot to closer observe the relationship between the variables in the anthropometric dataset. This high correlation between variables may have been a reason why we could not find many differences between the schools when working with the anthropometric dataset.



This graph shows the correlation between variables based on their intersected square. A pie that is more full means the two variables are more extremely correlated to 1 or -1 and the pies that are less full indicate less correlation. Finally, the blue circles indicate a positive correlation between variables and the red circles indicate a negative correlation.

V. Technical Output

Models: **STAT466 Project 2 School Differences Using Models**

Data