Sample Report and Solution to

Diets Case and Study Habits Case

Prepared by: Isabela Santos

Date: 16 November 2024

Nova Southeastern University

H. Wayne Huizenga School of Business & Entrepreneurship

Assignment for Course: BIA 5485

Submitted to: Dr. Yulia Yurova

Submitted by: Isabela Santos

Date of Submission: 16 November 2024

Title of Assignment: Apply Cluster Analysis and Report on the Results

CERTIFICATION OF AUTHORSHIP: I certify that I am the author of this paper and that any assistance I received in its preparation is fully acknowledged and disclosed in the paper. I have also cited any sources from which I used data, ideas or words, either quoted directly or paraphrased. I also certify that this paper was prepared by me specifically for this course.

Student's Signature: _____IS_____

****

Instructor's Grade on Assignment:
Instructor's Comments:

Clustering Report on Diets

Executive Summary

This study analyzes the dietary patterns of 150 students based on their average daily intake of protein, fat and carbohydrates utilizing the JMP software. The dataset is analyzed employing Hierarchical Cluster analysis, identifying three different clusters of diverging dietary patterns: cluster 1 characterized by moderate protein intake, high fat intake, and moderate carbohydrate intake; cluster 2 characterized by low protein intake, high fat intake, and high carbohydrate intake; and cluster 3 characterized by high protein intake, low fat intake, and low carbohydrate intake. These differences were found to be statistically significant across the clusters through ANOVA and post hoc tests. The findings can inform educational initiatives targeted towards improving nutrition and eating habits addressing specific needs of different student groups. Further research could be conducted to explore the variables influencing differences across clusters to improve nutritional strategies further.

I. Background

Dietary habits among students may not only impact their health directly but also influence their performance in a demanding environment. Varied macronutrients pay different roles in making a diet adequate for students, and identifying trends in common dietary patterns can provide helpful insights in consumption trends and highlight areas of improvement. In this study, data was sourced for the average daily consumption of protein, fat, and carbohydrates in grams among 150 students.

This data may be properly assessed through cluster analysis, a method which can be applied to group data points with similar characteristics – in this case, students with similar dietary habits, allowing us to identify patterns in the multivariate data. Through exploring these clusters, dietary profiles from the sample can be uncovered, with quantifiable differences that allow for evaluation of the average daily consumption of macronutrients among students. All of the analyses in this study were conducted utilizing the JMP software and its tools.

II. Problem

Can distinct dietary patterns among the students sampled be spotted through clustering analysis of average daily consumption data on protein, fat, and carbohydrates?

III. Analysis

In order to analyze dietary patterns based on the consumption of protein, fat, and carbohydrates, cluster analysis was utilized to group data points with similar profiles in nutritional content. Among clustering techniques, Hierarchical Clustering was used because the data set is relatively small, and the technique is suitable for continuous variables such as those provided. Hierarchical Clustering also does not require the establishment of a predefined number of clusters making it a flexible technique to be employed in this case. The results of Hierarchical Clustering are displayed in a dendrogram in Figure 1, providing a comprehensive visualization of groupings in the data and displaying observations based on their similarity with no need for an assumption on cluster quantities.

The performance of the Hierarchical Clustering model was evaluated using the Cubic Clustering Criterion (CCC), which helps in assessing the quality of solutions provided by the technique by measuring the separation between clusters. The CCC that resulted in the highest CCC value for this study and thus optimal solution was achieved by dividing the data into three clusters (CCC = 10.049). This result aligned with the grouping suggested by the dendrogram, confirming that the resulting clusters provided the best way to segment the data in this case. The optimal solution identified consisted of separating the macronutrient consumption patterns into 3 different clusters, showcasing that there are 3 different dietary patterns resulting from the sample, visualized through the dendrogram in Figure 1 and the constellation plot in Figure 2.

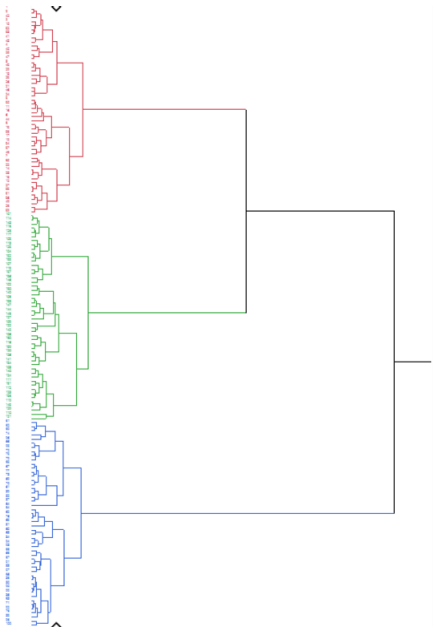

Figure 1: Dendrogram resulting from Hierarchical Clustering analysis

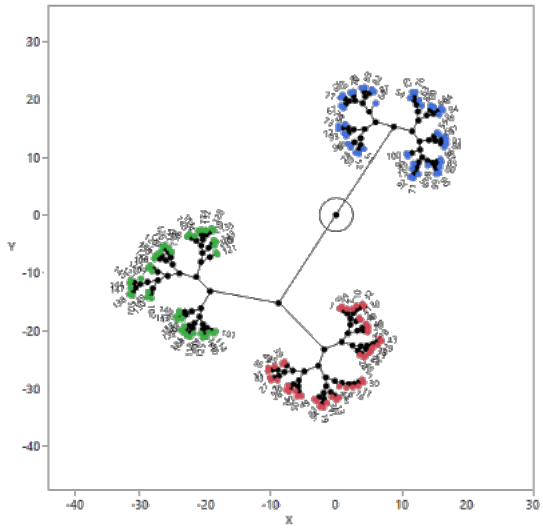

Figure 2: Constellation plot from Hierarchical Clustering analysis

The resulting cluster means which showcase the mean values for each variable in each cluster can be found in Table 1 below:

Cluster	Protein	Fat	Carbs
1 (n=50)	78.4g	55.3g	316.1g
2 (n=50)	49.6g	60g	349.4g
3 (n=50)	101.6	20.3g	299.9g

Table 1: Cluster means values for each resulting cluster

From the values of the cluster means, one can identify three distinct dietary patterns. Cluster 1 consists of a diet characterized of medium protein intake, high fat intake, and medium carbohydrate intake. Cluster 2 consists of low protein intake, high fat intake, and high carbohydrate intake. Cluster 3 consists of high protein intake, low fat intake, and low carbohydrate intake. The cluster means have been graphed below on Figure 3 for ease of visualization of each distinct characteristic between the three resulting clusters.

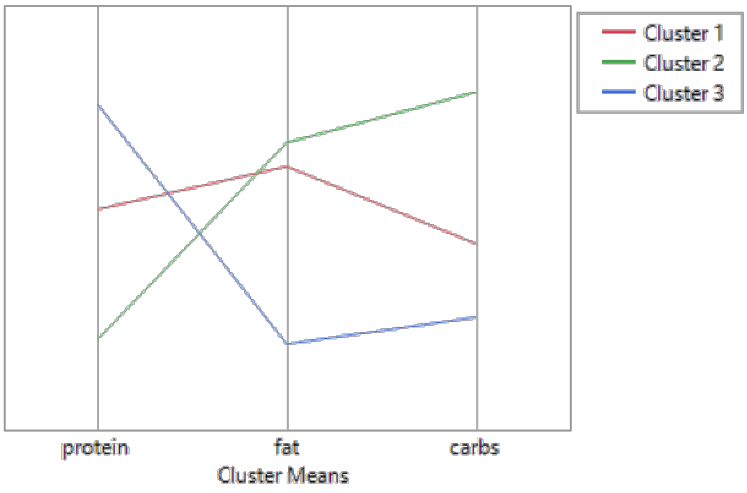

Figure 3: Graph of cluster means of protein, fat, and carbohydrates by cluster

The identification of three different clusters with diverging characteristics indicates that the average daily consumption of protein, fat, and carbs is not uniform across all study participants. To further explore whether there are statistically significant differences among diets, an ANOVA and post hoc tests were performed on the data set through Oneway analysis of the different variables (protein, fat, and carbohydrates) by the cluster. These tests all yielded *p*-values of <.0001, meaning that there are, in fact, statistically significant differences among diets.

IV. Conclusions and Recommendations

The analysis performed yielded three different clusters that represent different dietary patterns characterized by different consumption of protein, fat, and carbohydrates. Cluster 1 consists of a moderate intake of protein, high intake of fat, and moderate intake of carbohydrates; cluster 2 presents a low protein intake, a high fat intake, and a high carbohydrate intake; lastly, cluster 3 consists of high protein intake, low fat intake, and low carbohydrate intake. Further analysis confirmed there is, in fact, a statistically significant difference between the clusters' macronutrients consumption.

The division of this data in divergent clusters might assist dietary initiatives among students with different nutritional habits. Educational initiatives that promote balanced nutrition would greatly benefit from understanding the number of students belonging to each cluster and how that could shape programs targeted at improving nutrition amongst the students. Other surveys could in the future be conducted to better understand the causes of these differences in dietary habits.

Clustering Report on Study Habits

Executive Summary

This study explores time management and gender-related differences among 500 12th grade students in the US, more specifically exploring time spent on doing homework and watching TV. All analyses were performed using JMP software. Employing K-Means clustering analysis, three distinct clusters were identified: cluster 1, consisting of substantial hours spent on homework and few hours watching TV; cluster 2, consisting of moderate time spent on homework and substantial time spent on watching TV; and cluster 3, with few hours doing homework and few hours watching TV. Furthermore, while contingency analysis resulted in a statistically significant distribution of gender across the clusters, regression analysis through fit least squares revealed that gender does not significantly affect the pattern allocation of time to different activities. These findings suggest that gender is not a driver of time management habits, but rather behavioral differences are. Thus, policies and initiatives should not be gender-based but rather target behavioral patterns, which could be further explored through future surveys that include questions exploring different variables for a deeper comprehension of the diverging patterns revealed.

I. Background

Time management is a highly important factor that influences academic performance in high school students. Understanding how different factors affect this behavior, such as gender differences, may reflect broader trends in how individuals allocate their time and in school social dynamics. Through the examination of gender-based differences in time spent in doing homework and in watching TV, this analysis aims to bring to light potential differences that could influence academic results.

This study utilizes a sample collected in 2013 of 500 12th grade students from the United States, with the variables state, region, and survey responses. Cluster analysis is applied to explore overall differences and gender differences in average time spent on homework and watching TV, seeking to group students into clusters of similar habits and showcasing how time management may differ across genders. All the analyses conducted was performed in the JMP software.

II. Problem

Are there differences in students' average time spent doing homework and watching TV and can distinct patterns and gender-based variances be found in how time is allocated?

III. Analysis

To conduct an analysis on the time allocation between hours spent on homework and hours watching TV, cluster analysis was initially performed to group data points that indicated similar behavior amongst groups of students. K-Means clustering was selected in this case as it is suitable for slightly larger data sets as well as those presenting continuous variables (in this case, hours reported watching TV or doing homework). In order to select the optimal number of clusters, different CCC values, which determine the quality of separation between the resulting clusters, were analyzed through K-Means. In this case, the optimal CCC value of -3.0632 was a result of a three-cluster separation. The results of this clustering analysis are observed in Figure 4, which displays the scatterplot of data points grouped into the three diverging clusters. The red circle corresponds to cluster 1, green to cluster 2, and blue to cluster 3.

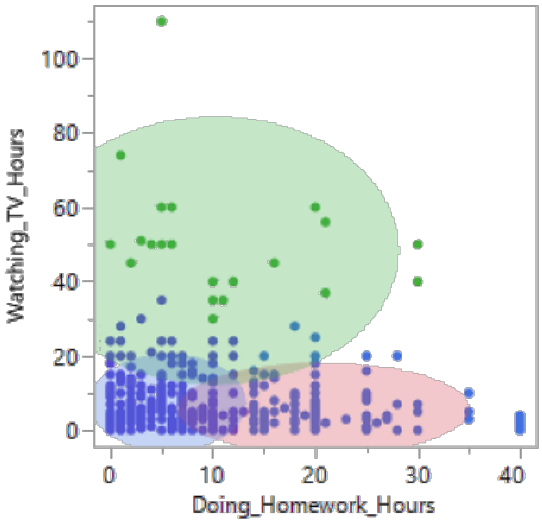

Figure 4: K-Means scatterplot of hours spent doing homework by hours spend watching TV

By observing Table 2 below, one may also classify different behaviors amongst the students grouped in each of the clusters. Cluster 1 englobes students with high number of hours on homework, and low number of hours watching TV. Cluster 2 presents moderate hours on homework and high hours watching TV. Lastly, cluster 3 presents low number of hours on homework as well as watching TV.

Cluster	Hours on Homework	Hours on TV
1 (n=115)	21 h	6 h
2 (n=24)	10 h	48 h
3 (n=314)	6 h	7 h

Table 2: Cluster means values for each resulting cluster

The three divergent resulting clusters indicate that students do not allocate equal hours on homework and watching TV and that there are different behavioral patterns in the data. To further investigate and comprehend the effects of gender in these behaviors, Contingency Analysis was performed.

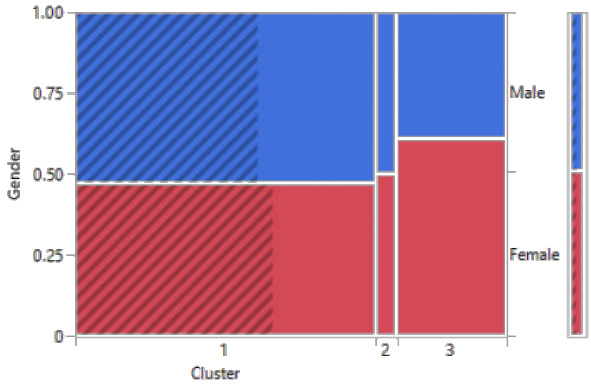

Figure 5: Mosaic plot resulting from contingency analysis of gender by cluster

The proportion of males and females in each of the three clusters is displayed the mosaic plot in Figure 5. Additionally, through a Chi-Square test, the relationship was statistically examined, yielding a *p*-value (*Prob>ChiSq*) of 0.0448, displaying a statistically significant difference between the distribution of males and females across the clusters. However, to then comprehend the combined effects of gender and cluster membership, a regression analysis through the Fit Least Squares technique was conducted.

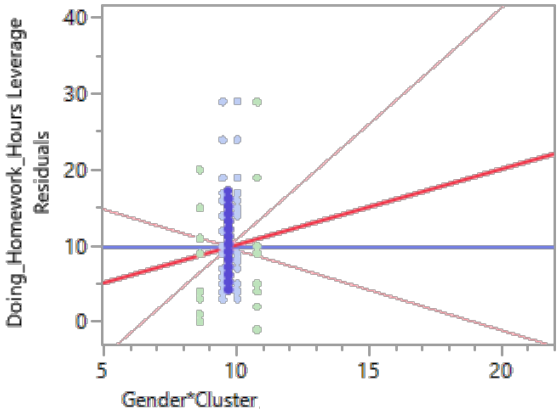

Figure 6: Leverage plot of Gender*Cluster regression through Fit Least Squares

The analysis of the interaction between gender and cluster yielded a *p*-value of 0.4933. This non-significant result suggests that gender differences within the clusters appear to be consistent. Further regressions were conducted plotting hours spent on homework by hours watching TV (*p*-value of <.0001), hours spent on homework by gender (*p*-value of 0.4667), and hours spent on homework by clusters (*p*-value of <.0001). These results indicate that the differences within clusters are only significantly significant when explained by behavioral differences, but not by gender. Thus, it can be concluded that the gender differences in average responses about hours spent on homework and on watching TV are not significant and that the habits of respondents remain consistent across all clusters regardless of gender and are explained instead by their behavior.

IV. Conclusions and Recommendations

This analysis identified three different groups within the student respondents: those who spend a substantial time on homework hours and low hours on watching TV; those who spend moderate time on doing homework and a significant time watching TV; and those who spend little time both on doing homework and on watching TV. These differences could be further explored by expanding the diversity of questions surveyed in order to comprehend where the differences in time management stems from for each cluster identified.

Moreover, while a statistically significant relationship between gender and belonging to a specific cluster was identified, gender differences for each cluster were found to not be statistically significant through regression. This result indicated that while gender might be relevant, it is not what drives the behavioral differences in the observed clusters. These findings indicate that if time management initiatives were to be implemented based on these results, they should not focus on gender-specific strategies, but rather, should focus on the behavioral differences themselves. As discussed, further drivers of these divergences could be identified by conducting future surveys that explore other possible factors that might influence clustering of these variables.

References

JMP®, Version *17*. SAS Institute Inc., Cary, NC, 1989–2024.