REPORT AND ANALYSIS

From Calories to Composition: Linking Intake, Activity, and Body Metrics

Nutritional Dietary data

Prepared By

Brylle Matthew A. Lupac

Naomi Christienne Tiama

1. Introduction

This dataset presents comprehensive nutritional and physiological information from 1000 individual patients, capturing a wide range of health-related variables that reflect diet, body composition, and lifestyle habits. For each patient, the dataset includes body fat percentage, muscle mass (in kilograms), Body Mass Index (BMI), weekly physical activity (in hours), and daily intake data for calories, protein, fat, carbohydrates, vitamin C, iron, and water. These variables are critical indicators of overall nutritional status and metabolic health. The data allows for the exploration of relationships between physical activity levels and macronutrient intake, as well as how dietary patterns may correlate with variations in body composition, such as higher muscle mass or elevated body fat percentage. The inclusion of micronutrients like vitamin C and iron adds depth to the analysis, offering opportunities to assess micronutrient sufficiency or deficiency within different activity levels or BMI categories. Additionally, water intake data supports hydration-related assessments, which are vital in nutrition and fitness studies. This dataset is well-suited for statistical analysis, nutritional modeling, or developing personalized health recommendations, particularly in the context of preventive health care, obesity research, or performance nutrition. Its breadth and detail make it valuable for understanding how diverse dietary profiles and lifestyle factors interact with body composition across a real-world sample of individuals.

2. Methods

2.1. Data Cleaning and Preparation

- The dataset titled **3_Nutritional_Dietary_data_Group_005.csv** was imported using the read.csv() function. Several essential R packages were loaded, including tidyverse, ggplot2, dplyr, readr, and psych, to enable efficient data manipulation, visualization, and statistical analysis.
- The structure of the dataset was checked using the str() and glimpse() functions to understand variable types and data consistency. The column names were manually renamed using the colnames() function to ensure clearer and more descriptive variable labels, such as “Body_Fat,” “BMI,” and “Daily_Calories.”
- To ensure data quality, the group checked for missing values using the is.na() function combined with sum(), which confirmed that the dataset had no missing entries.

2.2. Descriptive Statistical Analysis

- Descriptive statistics were generated for all continuous variables using the describe() function from the psych package. This provided detailed summary measures, including:
  - Mean – average value
  - Standard Deviation – measure of spread
  - Median – middle value
  - Minimum and Maximum – range
  - Skewness and Kurtosis – distribution shape indicators
- Variables included in this summary were: Body Fat, Muscle Mass, BMI, Physical Activity Hours, Daily Caloric Intake, and Nutrient Intake (Protein, Fat, Carbohydrates, Vitamin C, Iron, and Water in mL).

2.3. Data Visualization

- **Histogram:** A histogram of BMI was plotted using geom_histogram() to visualize the frequency distribution of BMI values.
- **Boxplot:** A boxplot was created to compare Body Fat percentages across different levels of Physical Activity Hours.
- **Scatter Plot:** A scatter plot with a linear regression line was constructed to examine the relationship between BMI and Body Fat. The plot used geom_point() and geom_smooth(method = "lm") to highlight any trends.
- **Bar Plot:** A bar plot was generated to show the distribution of daily water intake. Water intake values were binned into five categories (e.g., “0–1000 mL”, “1001–2000 mL”) using the cut() function. The frequency of individuals in each bin was then plotted using geom_bar() with a themed and angled x-axis for improved readability.

2.4. Advanced Statistical Test

- To assess potential differences in caloric intake based on fat and carbohydrate intake levels, the group performed two separate one-way ANOVA tests:
    - **Fat Intake ANOVA:** Individuals were grouped into “Low Fat,” “Medium Fat,” and “High Fat” categories using `cut()` based on fat intake tertiles. A one-way ANOVA was conducted using `aov()` to test if mean Daily Caloric Intake significantly differed across fat intake groups.
    - **Carbohydrate Intake ANOVA:** Similarly, carbohydrate intake was categorized into “Low Carb,” “Medium Carb,” and “High Carb.” A second one-way ANOVA tested whether caloric intake differed across these carbohydrate consumption groups.
- The results of both ANOVA tests were interpreted using the `summary()` function to assess p-values and statistical significance.

3. Results and Discussion

3.1. Summary Statistics

Table 1. Descriptive Statistics of Patient Health and Nutritional Variables

Variable	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
Body Fat (%)	35.48	14.32	33.69	7.03	72.81	65.78	0.41	-0.71	0.45
Muscle Mass (kg)	39.71	9.67	40.47	5.76	61.76	56.00	-0.31	-0.03	0.31
BMI	22.82	2.90	22.90	15.40	30.70	15.30	0.01	-0.52	0.09

Physical Activity (hrs/wk)	8.24	4.98	8.50	0.00	16.00	16.00	-0.08	-1.24	0.16
Daily Calories (kcal)	2445.5	823.54	2446.0	350.0	4556.0	4206.0	0.00	-0.48	26.04
Protein Intake (g)	125.37	66.69	127.00	10.00	240.00	230.00	-0.04	-1.20	2.11
Fat Intake (g)	101.76	57.51	103.00	5.00	200.00	195.00	0.00	-1.24	1.82
Carbohydrate (g)	257.06	142.06	250.50	20.00	500.00	480.00	0.03	-1.27	4.49
Vitamin C (mg)	729.70	691.25	500.00	50.00	2000.0	1950.0	0.89	-0.59	21.86
Iron (mg)	18.27	13.97	18.00	5.00	40.00	35.00	0.70	-1.17	0.44
Water Intake (ml)	2524.0	1119.2	3000.0	1000.0	4000.0	3000.0	-0.02	-1.36	35.39

- The average body fat percentage among the 1000 participants was 35.48%, with a standard deviation of 14.32%, ranging from 7.03% to 72.81%.
- The muscle mass had a mean of 39.71 kg, with values ranging from 5.76 kg to 61.76 kg.
- The average BMI was 22.82, indicating that most participants were in the normal weight range; values ranged from 15.4 to 30.7.
- Participants reported an average of 8.24 hours of physical activity per week, with a maximum of 16 hours. The daily caloric intake had a mean of 2445.54 kcal, ranging widely from 350 to 4556 kcal.
- The average protein intake was 125.37 g, with a wide variation (from 10 to 240 g).
- The fat intake averaged 101.76 g, with minimum and maximum values of 5 g and 200 g, respectively.
- The carbohydrate intake had a mean of 257.06 g, with a maximum of 500 g.
- For Vitamin C, the mean intake was 729.7 mg, with some participants reaching the maximum of 2000 mg.
- The average iron intake was 18.27 mg, with a range from 5 mg to 40 mg
- Water intake varied greatly among participants, averaging 2524 ml, with a range of 1000 ml to 4000 ml.

3.1. Data Visualizations

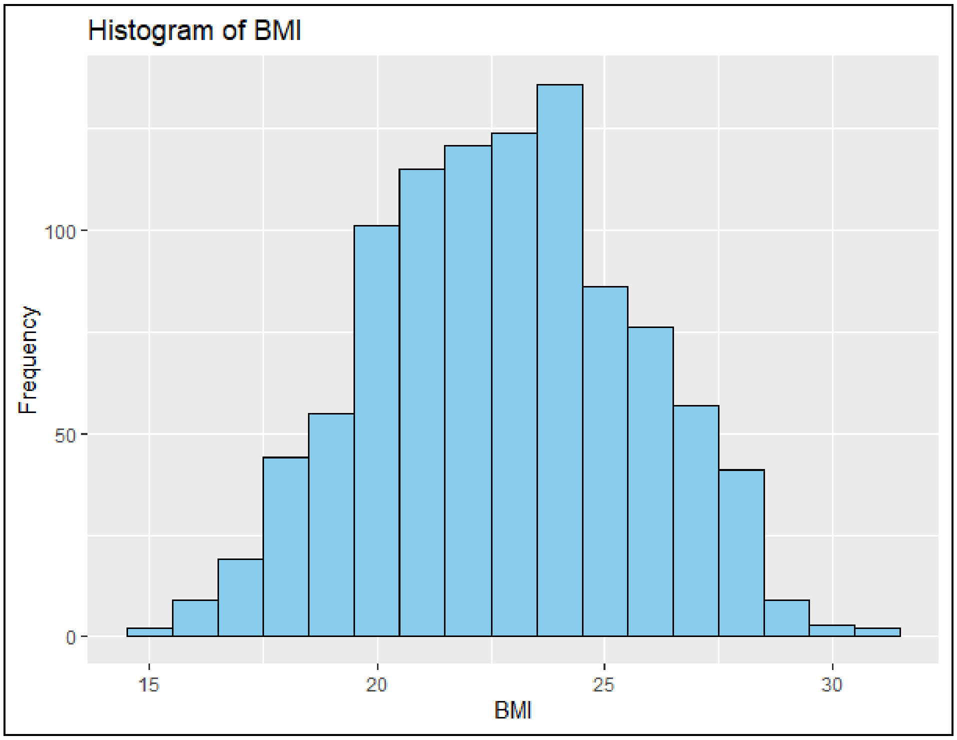

Figure 1. Histogram of BMI

- The histogram displays the frequency of individuals across different BMI (Body Mass Index) values, ranging approximately from 15.4 to 30.7.
- Most BMI values fall between 18 and 26, indicating that the majority of individuals are within the normal to slightly overweight range based on standard BMI classification.
- The peak frequencies occur at:
    - BMI 22 (22 individuals), BMI 20.6 (18 individuals), BMI 24.4 (18 individuals), and BMI 20.8 (17 individuals).
- Very low BMI values (below 17) and high BMI values (above 28) are rare, with only 1 to 4 individuals in each of those bins.
- The overall shape of the distribution appears to be roughly bell-shaped, but slightly right-skewed, meaning there are more individuals with higher BMI values than with very low BMI.

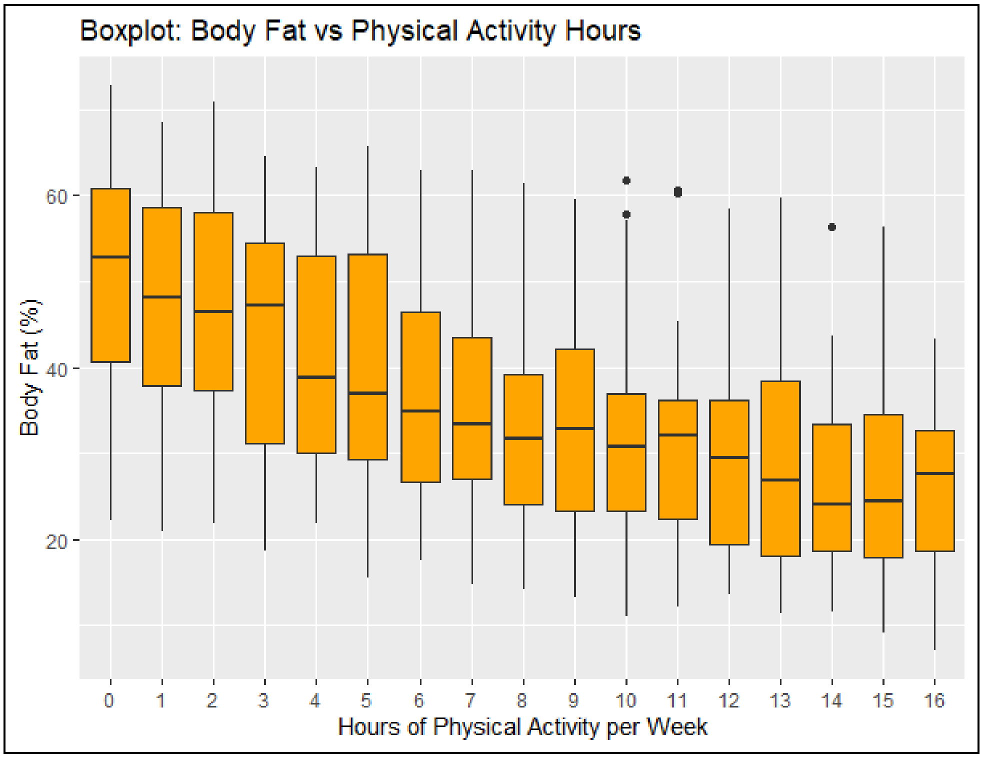

Figure 2. Boxplot: Body Fat by Physical Activity Hours

- The boxplot shows the relationship between Body Fat Percentage and the number of Physical Activity Hours per Week.
- In general, individuals who engage in more physical activity tend to have lower body fat percentages.
- The median body fat appears to decrease as physical activity increases, suggesting a negative association between the two variables.
- Groups with low physical activity (e.g., 0–2 hours) show higher median body fat, with wider interquartile ranges (IQRs), indicating greater variation in body fat.
- On the other hand, groups with higher activity levels (e.g., more than 5 hours/week) show lower and more consistent body fat percentages, with shorter boxes and fewer outliers.
- Outliers are present, especially in low-activity groups, showing that some individuals have unusually high or low body fat compared to others in the same activity range.

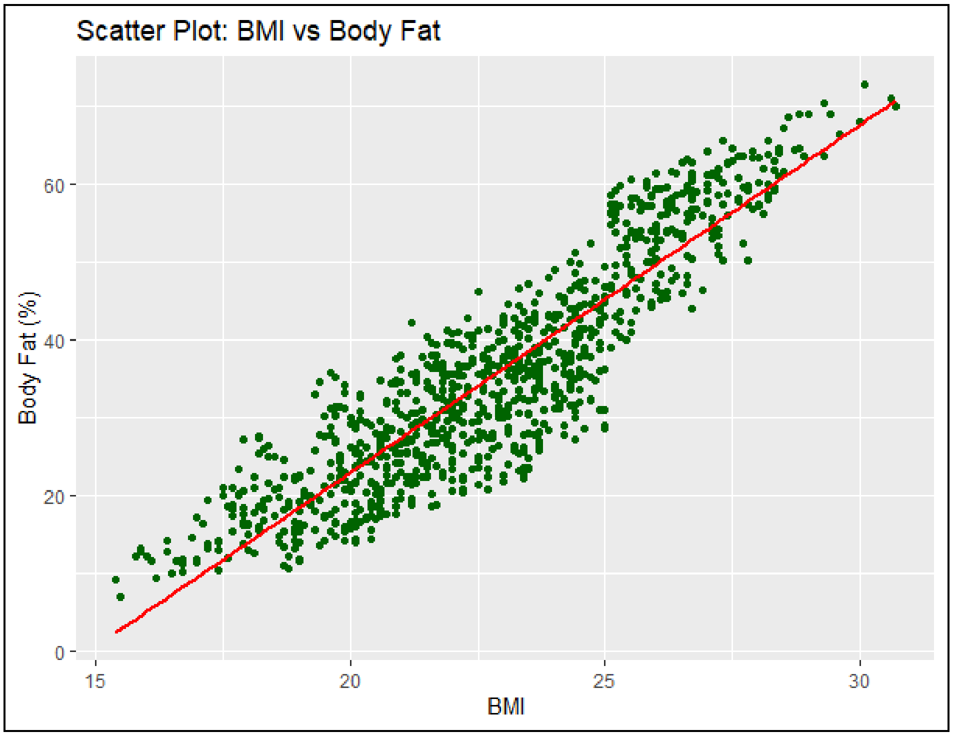

Figure 3. Scatter Plot: BMI vs Body Fat Percentage

- The scatter plot shows the relationship between a person's Body Mass Index (BMI) and their Body Fat Percentage.
- There is a strong positive linear trend: as BMI increases, Body Fat Percentage also tends to increase.
- This indicates that individuals with higher BMI values generally have higher body fat levels, which is expected since both are measures of body composition.
- The points are closely clustered along an upward-sloping line, suggesting a strong correlation between the two variables.
- However, there is some variability in body fat among individuals with the same BMI, meaning that BMI alone does not perfectly predict body fat.
- A few outliers are also visible—individuals with similar BMI values but unusually low or high body fat percentages.

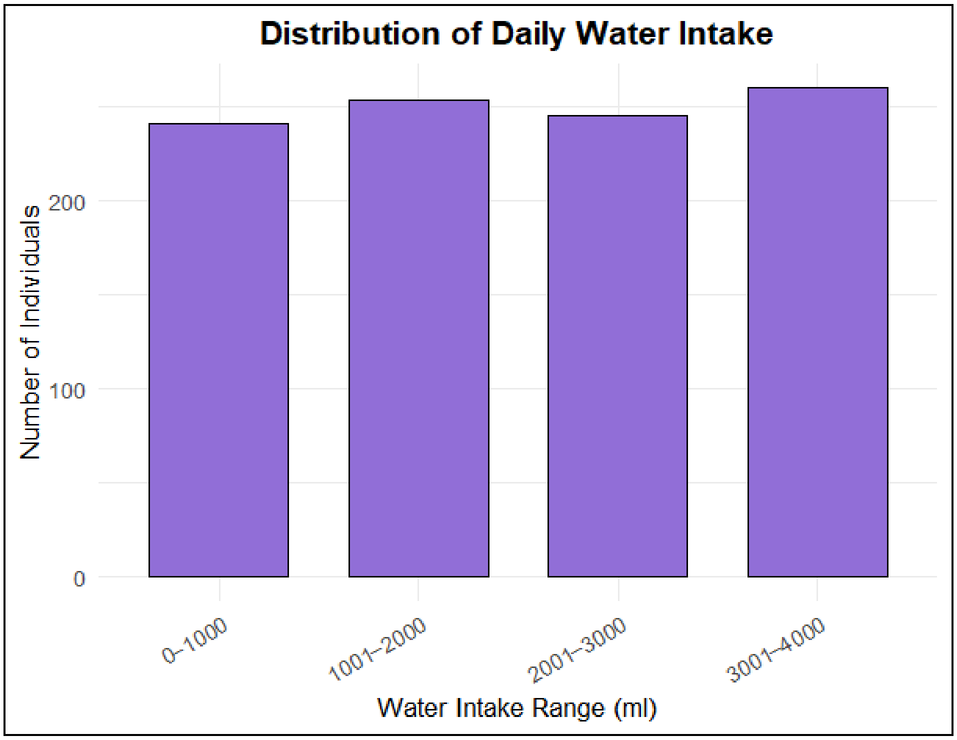

Figure 4. Bar Graph: Distribution of Daily Water Intake

- The group with the highest number of individuals is the 4000 ml and above range (260 individuals).
- The group with the lowest count is the 0–1000 ml range (241 individuals), but the difference across all groups is small.
- Overall, the distribution appears fairly even, suggesting that most individuals have varying water intake habits, and no single group dominates significantly.

2.3 Advanced Statistical Insight

Table 2. One-Way ANOVA: Daily Calories by Fat Intake Group

Source	Df	Sum of Squares	Mean Square	F value	Pr(>F)	Significance
Fat Group	2	239,973,679	119,986,839	273.4	< 2e-16	***
Residuals	997	437,558,960	438,876			

- The p-value was **less than 2e-16**, which means the result is **highly significant** and not due to random chance.
- The **F-value of 273.4** indicates a strong variation in caloric intake explained by fat intake grouping.
- This suggests that **fat consumption levels** have a **major effect** on how many calories individuals consume per day.

Table 3. One-Way ANOVA: Daily Calories by Carb Intake Group

Source	Df	Sum of Squares	Mean Square	F value	Pr(>F)	Significance
Carb Group	2	298,539,210	149,269,605	392.7	< 2e-16	***
Residuals	997	378,993,428	380,134			

- The **p-value < 2e-16** confirms a **strong statistical significance** in the differences observed.
- With an **F-value of 392.7**, the influence of carbohydrate intake grouping on daily caloric intake is even **stronger** than that of fat intake.
- This means that individuals grouped by **carbohydrate consumption** show clear and **distinct differences** in the amount of calories they consume daily.

4. Conclusion

This study utilized a detailed dataset of 1000 individuals to explore how nutritional intake and lifestyle behaviors influence body composition and energy consumption. Descriptive statistics revealed that the average participant maintained a normal BMI and muscle mass, yet body fat and nutrient intake levels varied widely. Visual analyses supported key relationships: BMI followed a slightly right-skewed distribution; increased physical activity was associated with lower body fat; and BMI showed a strong positive correlation with body fat percentage. Water intake habits varied but were relatively evenly distributed. Crucially, advanced statistical testing via one-way ANOVA demonstrated that both fat and carbohydrate intake groups had highly significant effects on daily caloric consumption ($p < 2e-16$), with carbohydrate intake explaining even greater variance in caloric intake than fat. These findings underscore the intertwined nature of diet and lifestyle in shaping health outcomes and suggest that both macronutrient composition and activity levels are critical for understanding energy balance and body composition. This analysis provides a strong foundation for developing personalized nutrition strategies in clinical or fitness settings.