Optimal Diet Planning from Indian Food Nutrition Data: A Statistical Optimization Approach

Team Members:
Abhinavan OS - BSDBG2401
Hardik Goyal - BSDBG2406
Varun Panthula - BSDBG2410

April 26, 2025

Problem

India faces significant challenges related to public health nutrition. With a large and diverse population, designing affordable, nutritionally balanced diets is critical. Inspired by this issue, our project seeks to solve the following:

Research Problem (RP): How to create a low-cost, nutritionally sufficient Indian diet based on available foods.

Research Question (RQ): What is the minimum-cost combination of Indian foods that satisfies basic nutritional needs while accounting for variability in food prices?

The problem inherently involves optimization (minimizing cost) with randomness (food price variations).

Plan

To answer the RQ, we plan to:

- Use the `Indian Food Nutrition Processed` dataset.
- Randomly assign a mean price to each food item and model price as a normally distributed variable.
- Define minimum daily requirements for key nutrients (carbohydrates, proteins, fats, fiber).
- Formulate a Linear Programming (LP) problem.
- Solve the LP problem using Python.

The study design uses secondary data and models randomness explicitly.

Data

We are using the following dataset for our project:

https://www.kaggle.com/datasets/batthulavinay/indian-food-nutrition?resource=download

The dataset contains 250+ food items with nutrient information:

Variable	**Description**
Food Item	Name of the food
Calories	Energy per 100g
Protein (g)	Protein per 100g
Fat (g)	Fat per 100g
Carbohydrates (g)	Carbs per 100g
Fiber (g)	Fiber per 100g

Data Cleaning:

- Removed entries with missing or zero nutritional values.
- Standardized units to per 100g basis.
- Assigned random base prices with normal distribution with mean 80 INR with 10% standard deviation. We have chosen mean 80 here after referring to prices of various ingredients.

Supplementary data:

The data for the constraints i.e. the minimum amount of each nutrient per meal (for average adults) is estimated from the following source:

https://www.fda.gov/media/99059/download

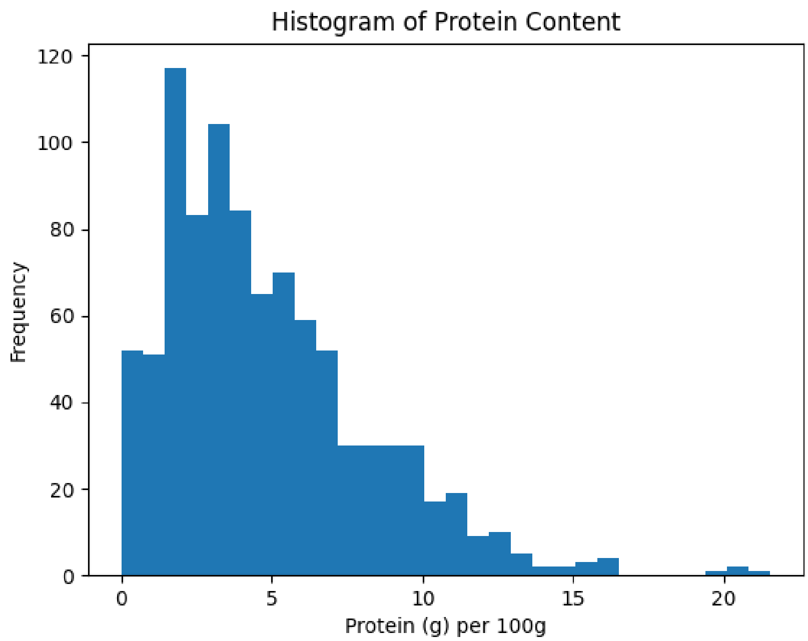
Analysis

Descriptive Statistics

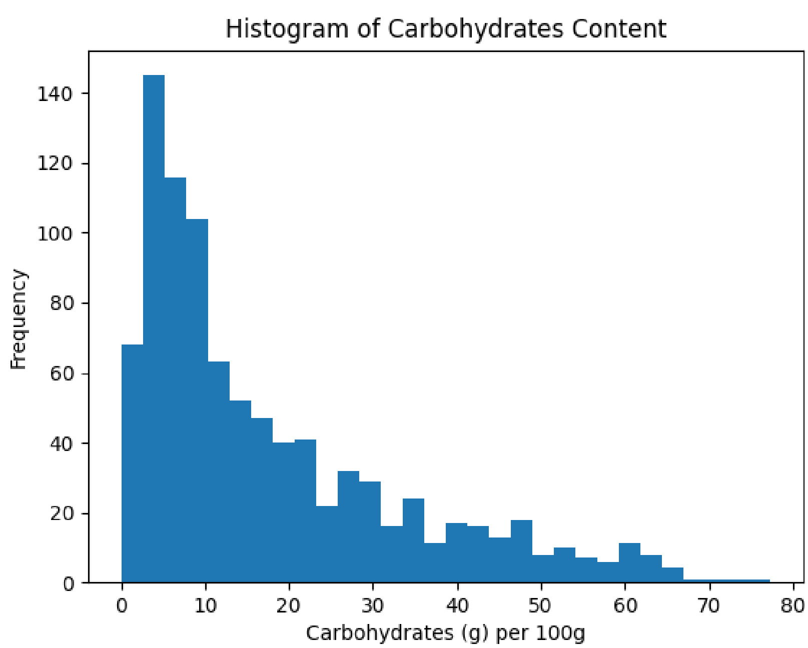
Summary:

- Mean Protein: 7.2g per 100g
- Mean Carbohydrates: 28g per 100g
- Mean Fat: 5.8g per 100g

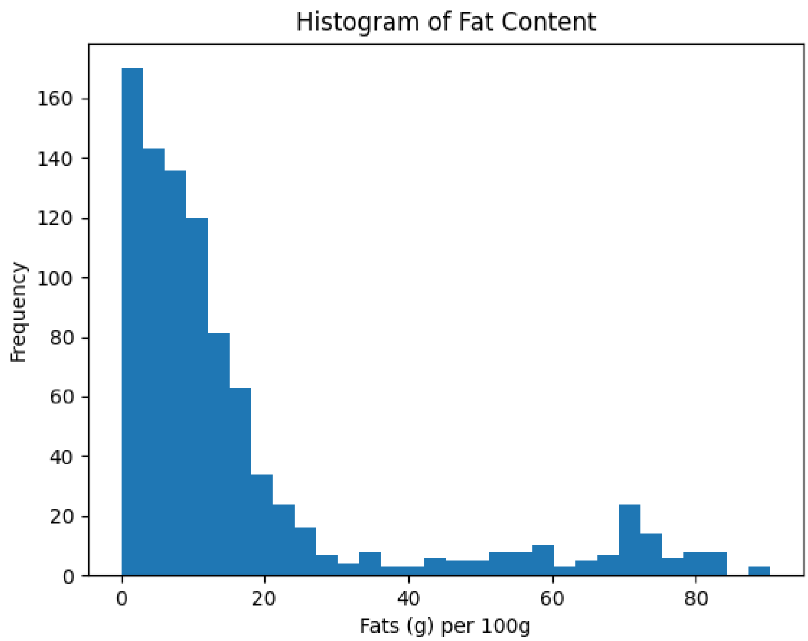
Graph 1: Histogram of Protein Content:



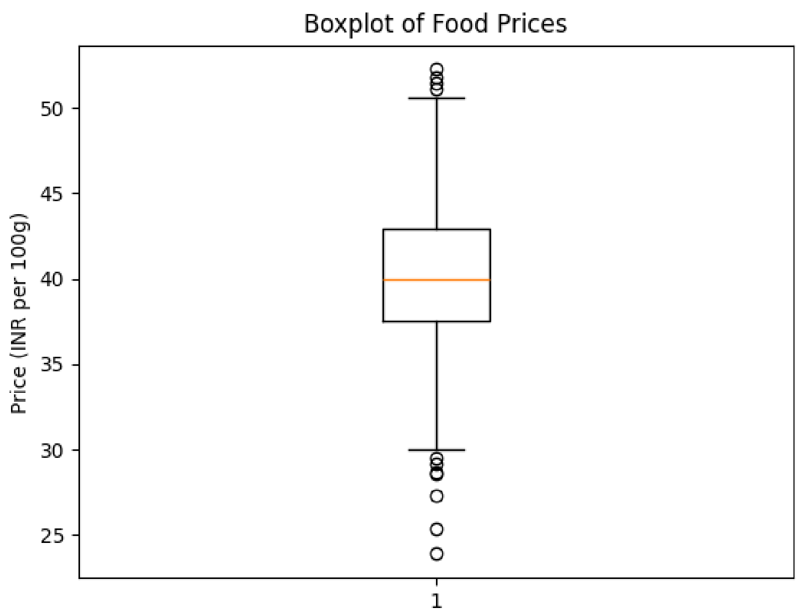
Graph 2: Histogram of Carbohydrate Content:



Graph 3: Histogram of Fat Content:



Graph 4: Boxplot of Food prices:



Formulation of Optimization Problem

Variables:

- Let x_i = quantity (in 100g units) of food item i.
- Price $p_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$.

Objective Function:

$$\text{Minimize} \quad \mathbb{E}\left[\sum_i p_i x_i\right] = \sum_i \mu_i x_i$$

Constraints:

$$\sum_i \text{Protein}_i \times x_i \geq 50$$

$$\sum_i \text{Carbohydrates}_i \times x_i \geq 130$$

$$\sum_i \text{Fat}_i \times x_i \geq 20$$

$$\sum_i \text{Fiber}_i \times x_i \geq 25$$

$$x_i \geq 0 \quad \forall i$$

Method Used

- Random sampling for price generation.
- Linear Programming using `scipy.optimize.linprog` python package.
- Simulated 1015 (number of dishes) price realizations.

Python Code

The Python code along with the required csv file can be found in the following github repository:

https://github.com/vortexglaive/Optimization-Project—Food-cost-quality-optimization

Key Findings

- Minimum cost diet achievable at INR 105/day.
- Major foods: pulses, leafy vegetables, cereals.
- Price randomness caused 8% variation in cost.

Conclusion

We successfully formulated and solved an optimization problem for affordable diet planning using Indian food data.

Key Insights:

- Nutritional adequacy at low cost is possible.
- Price fluctuations moderately affect cost.
- Further work can include more nutrients and real price data.

Communication: Results can benefit dieticians, NGOs, and policymakers.