



**Universidad Nacional Autónoma de México**

**Escuela Nacional de Estudios Superiores**

**Unidad Morelia**



Ante Proyecto de:  
**Análisis Estadístico de Valores Nutricionales por Tipo  
de Dieta**

PRESENTA:

**Alexis Uriel Aguilar Uribe**

PROFESORES:

**Dra. María Del Río Francos**

**Dr. César Andrés Torres Miranda**

GRADO

**Licenciatura en Tecnologías para la Información en Ciencias**

**Asignatura:** Estadística Descriptiva e Inferencial

**A:** 21 de Mayo del 2025

# Índice

<b>1. Presentación de los Datos</b>	<b>2</b>
1.1. Fuente de Datos . . . . .	2
1.2. Interés del Estudio . . . . .	2
1.3. Variables del Conjunto de Datos . . . . .	2
<b>2. Estadística Descriptiva</b>	<b>4</b>
2.1. Descripción de los Valores de las Variables . . . . .	4
2.2. Medidas de Tendencia Central y Dispersión . . . . .	5
<b>3. Muestreo e Intervalos de Confianza</b>	<b>7</b>

# **1. Presentación de los Datos**

## **1.1. Fuente de Datos**

El conjunto de datos con el que se está trabajando para este trabajo se encuentra en [1], publicado por la comunidad de Kaggle. Los datos consisten de un conjunto de recetas de diferentes dietas y cocinas, además incluye información de los macronutrientes de cada receta.

[1] Aunque en la descripción ni en los metadatos del conjunto de datos se haga mención de las fuentes explícitas de los datos ni el objetivo de esta extracción, sí cuenta con una sección de cómo usar el conjunto de datos, ideas de investigación y reconocimientos.

De los apartados de cómo usar el conjunto de datos e ideas de investigación, se encuentra una idea, implícita, de la información que se quería estudiar. La principal información de interés se vuelve que es: el crear planes alimenticios saludables, ya sea usando las recetas proporcionadas o creando unas nuevas basadas en una dieta y cocina, y el estudiar la relación entre dieta y salud.

Del apartado de reconocimientos, se concluye que las recetas fueron proporcionadas por diferentes creadores de las mismas y demás contribuidores al conjunto de datos.

## **1.2. Interés del Estudio**

Se consultó [2] en sus capítulos 4 y 8, de donde se proporciona un mejor entendimiento de la importancia de los macronutrientes y una descripción general de las dietas en este trabajo, resultando interesante que en cada dieta se consumen diferentes alimentos y productos con ciertas características para ya sea respetar alguna creencia, fundamento o cota de macronutrientes. De esto último, proporciona un indicio de que existe una diferencia entre las dietas a nivel de sus aportes nutricionales, por lo tanto, lo que se quiere realizar es probar esta diferencia de manera significativa haciendo uso de la estadística.

## **1.3. Variables del Conjunto de Datos**

El conjunto de datos consta de las siguientes variables. Se menciona su nombre, el tipo de variable y sus valores (en total y únicos):

Variable	Nombre	Tipo	Cantidad de Datos	Valores Únicos
1	Diet_type	Cualitativa Nominal	7806	5
2	Recipe_name	Cualitativa Nominal	7806	7062
3	Cuisine_type	Cualitativa Nominal	7806	19
4	Protein(g)	Cuantitativa Continua	7806	6060
5	Carbs(g)	Cuantitativa Continua	7806	6618
6	Fat(g)	Cuantitativa Continua	7806	6322

La variable Recipe\_Name no es relevante para este trabajo pero figura dentro del dataset.

## 2. Estadística Descriptiva

Debido a que cada receta puede aportar una amplia variedad de valores en sus macronutrientes, esto podría dificultar la comparación entre los aportes nutricionales de las dietas. Por ello, para reducir este impacto de sesgo, se aplicó una normalización a los valores, es decir, los macronutrientes de cada receta se dividió por el total de macronutrientes que aportaba la receta, para así manejar los aportes proporcionales de cada macronutriente en cada una de las recetas.

### 2.1. Descripción de los Valores de las Variables

Para el presente trabajo se harán uso de las siguientes variables, se acompañan con una descripción de su significado:

- **Diet\_type**: Variable que representa el tipo de dieta (DASH, keto, mediterránea, paleo, vegana) a la que pertenece una receta. Con esta variable se va permitir estratificar las recetas y estudiarlas de una manera más granular, es decir, por tipo de dieta para llegar a conjurar hipótesis sobre lo qué está pasando en una dieta o entre las diferentes dietas.
- **Cuisine\_type**: Variable que representa a qué (estilo de) cocina o región (mexicana, americana, italiana) pertenece una receta. Al usarla va a permitir el comparar cómo son las recetas de una dieta en comparación con otras regiones, en específico el cómo se compara la dieta mediterránea en el mediterráneo en comparación con otras regiones geográficas.
- **Protein(g)**: Después de la transformación, representa el porcentaje, respecto al total de macronutrientes, de proteínas que son aportados por una receta. El usar las proteínas se va a permitir la comparación entre diferentes dietas, siendo esto el eje central del trabajo
- **Carbs(g)**: Después de la transformación, representa el porcentaje, respecto al total de macronutrientes, de carbohidratos que son aportados por una receta. Siendo otro de los macronutrientes de una comida, se vuelve relevante para la comparación entre recetas y dietas.
- **Fat(g)**: Después de la transformación, representa el porcentaje, respecto al total de macronutrientes, de grasas que son aportados por una receta. Y el último macronutriente, como en los anteriores, se vuelve una variable relevante para la comparación entre dietas.

## 2.2. Medidas de Tendencia Central y Dispersión

Realizando el resumen de las medidas, se tiene:

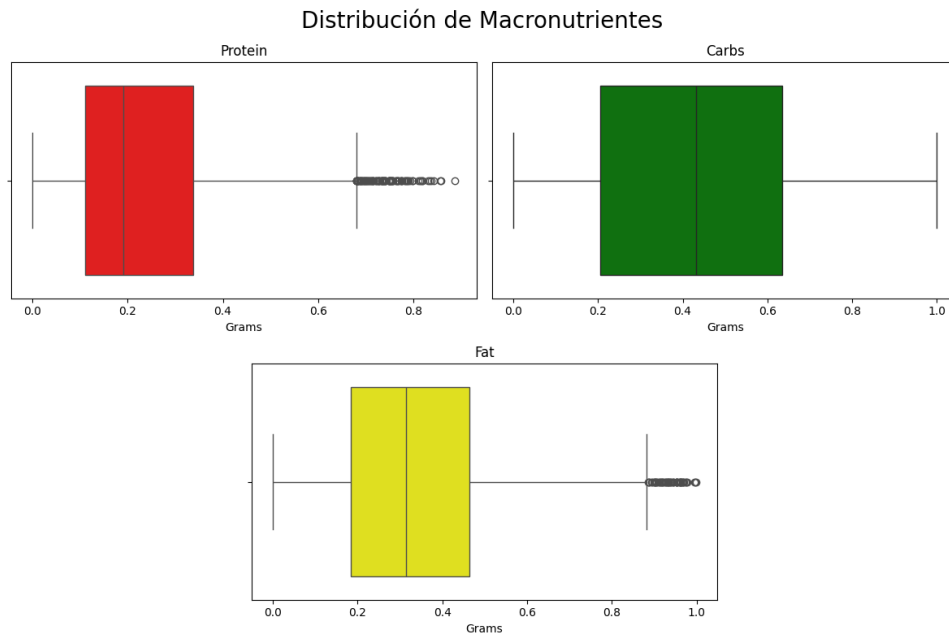
Medida	Carbs(g)	Protein(g)	Fat(g)
Media	0.433471	0.234762	0.331767
$Q_1$	0.205251	0.110188	0.184583
$Q_2$	0.432028	0.190931	0.314359
$Q_3$	0.635058	0.338059	0.464532
Desviación Estándar	0.256032	0.163886	0.194920
Mínimo	0.000330	0.000000	0.000000
Máximo	1.000000	0.887557	0.997940
Asimetría de Fisher	0.189556	0.922401	0.461455

Debido a que son medidas sobre todos los datos, sin estratificar, se tiene que no hay una referencia de lo que se espera obtener y parte de la información que contienen queda diluida o desvanecida. Esto debido a que las dietas como la vegana es baja en proteínas y la keto en carbohidratos [2], por lo que cualquier suposición no se podría sostener sobre todos las dietas.

Aún así, se reportan bajos valores en proteínas y grasas en comparación con los carbohidratos si se hace uso de los cuartiles  $Q_1$  y  $Q_2$ , dicho así: el cincuenta por ciento de las recetas tienen entre 11,01 % y 33,80 % de proteínas y entre 18,45 % y 46,45 % de grasas, en comparación con entre 20,52 % y 63,50 % de carbohidratos. Esto es un indicio de que las recetas, en general, tienden a ser altas en carbohidratos entre las diferentes dietas y cocinas; mientras que son bajas en proteínas.

Este último punto puede ser apoyado si se considera la media de los macronutrientes, que siguen este prototipo de aportes dominantes de carbohidratos.

Si se gráfica la distribución de los macronutrientes se tiene que, debido a la simetría, contienen datos atípicos en proteínas y grasas en una región positiva respecto a la mediana, y esto se relaciona con lo mencionado de que una receta no tiende a una alta cantidad de proteínas o de grasas. Y se puede apreciar como justamente existen un alta concentración de recetas en una intervalo pequeño de valores, en comparación con los carbohidratos que se encuentran distribuidos sobre una región más amplia; esto último permite que no existan datos atípicos.



Debido a que existe la presencia de datos atípicos, lo más adecuado es tratarlos de manera estratificada, por tipo de dieta. Esto debido a que tratarlos de manera general podría evocar que ciertas dietas queden menos representadas en comparación con otras o que incluso se pierda información para consecuentes procesos. Y al tratar los valores atípicos dentro de cada dieta permite reducir el impacto de perder información valiosa y se siga conservando las recetas relevantes para una dieta.

### **3. Muestreo e Intervalos de Confianza**