Análisis de Modelos estadísticos

Javier Bravo C.I. 27736325

Luis Carrillo C.I. 27539960

2025-05-27

Table of Contents

# Introducción

La nutrición juega un papel fundamental en la salud y el bienestar general de los individuos. Diversos componentes dietéticos y características personales pueden influir significativamente en los indicadores fisiológicos clave, como los niveles plasmáticos de micronutrientes esenciales. Este estudio se enfoca en el análisis de datos de 315 individuos, con el objetivo de explorar las relaciones entre factores dietéticos, características personales y concentraciones plasmáticas de retinol, betacaroteno y otros carotenoides. Estos micronutrientes son vitales para diversos procesos biológicos, incluyendo la visión, la función inmune y la protección antioxidante, por lo que comprender los factores que afectan sus niveles es de considerable interés para la salud pública.

Los datos utilizados en este análisis provienen de un estudio transversal originalmente diseñado para investigar estas interacciones. Los participantes fueron pacientes que se sometieron a intervenciones quirúrgicas electivas para biopsiar o extirpar lesiones no cancerosas de diversos órganos. La información, disponible públicamente a través de <http://lib.stat.cmu.edu/datasets/Plasma_Retinol/>, se considera confiable y está contenida en la base de datos NutritionStudy del paquete Lock5Data en R.

El presente informe tiene como propósito principal realizar un análisis estadístico exhaustivo de estos datos. Se buscará determinar si existen diferencias significativas en diversas variables de salud y nutrición entre distintos grupos de individuos, definidos por características como sexo, hábitos de tabaquismo, consumo de alcohol y uso de vitaminas. A partir de estos análisis, se formularán recomendaciones de salud específicas para hombres y mujeres, y se identificarán posibles riesgos para la salud basados en los patrones observados.

Para alcanzar estos objetivos, el trabajo se dividirá en dos partes principales. Primero, se realizará un análisis descriptivo y gráfico de las variables consideradas más relevantes, explorando medidas de tendencia central, variabilidad, posición y forma de sus distribuciones. Posteriormente, se llevará a cabo un análisis de inferencia estadística, empleando un nivel de significancia del 5%, para contrastar hipótesis, estimar parámetros y comparar poblaciones, verificando los supuestos necesarios como la normalidad y la homogeneidad de varianzas cuando corresponda. Finalmente, se presentarán los resultados, se discutirán en el contexto de la problemática y se ofrecerán conclusiones y recomendaciones pertinentes.

# Resumen

Este estudio realiza un análisis estadístico de la base de datos NutritionStudy, que contiene información sobre la nutrición y salud de 315 individuos. El objetivo principal es determinar si existen diferencias significativas en variables nutricionales y de salud entre diversos grupos poblacionales (ej. por sexo, hábitos de tabaquismo, consumo de alcohol y uso de vitaminas). Adicionalmente, se busca formular recomendaciones de salud específicas para hombres y mujeres e identificar posibles riesgos de salud derivados de los análisis. La metodología empleada incluye un análisis descriptivo exhaustivo para caracterizar las variables en términos de tendencia central, variabilidad y distribución, seguido de un análisis de inferencia estadística con un nivel de significancia del 5%. Este último comprende la verificación de normalidad, la formulación y contraste de hipótesis, la estimación de parámetros y la comparación de medias entre grupos, utilizando pruebas paramétricas o no paramétricas según corresponda. Se espera que los resultados permitan comprender mejor la relación entre características personales, factores dietéticos y biomarcadores nutricionales, aportando información valiosa para la promoción de la salud.

# Antecedentes

El retinol y el betacaroteno son nutrientes esenciales con roles críticos en la salud humana. Mientras el retinol (vitamina A) es vital para la visión, la función inmune y la integridad epitelial, el betacaroteno, su precursor, actúa como antioxidante, mitigando el daño celular asociado a enfermedades crónicas como el cáncer y las patologías cardiovasculares [1]. Estudios previos, como el de Combs et al. (2012), han demostrado que deficiencias en estos nutrientes se vinculan con mayores riesgos de morbimortalidad, especialmente en poblaciones con hábitos como el tabaquismo o el consumo excesivo de alcohol [2].

Aunque investigaciones transversales, como la de Ford et al. (2003), han explorado la relación entre factores dietéticos y concentraciones plasmáticas de carotenoides, persisten vacíos en la comprensión de cómo variables demográficas y conductuales modulan estas asociaciones [3]. Por ejemplo, Smith et al. (2018) reportaron diferencias de género en el metabolismo de la vitamina A, sugiriendo que las mujeres podrían tener mayores reservas hepáticas de retinol, pero sin profundizar en implicaciones prácticas [4].

La base NutritionStudy, utilizada en este trabajo, ofrece una oportunidad única para abordar estas limitaciones. Proveniente de pacientes sometidos a intervenciones quirúrgicas no oncológicas, permite analizar poblaciones con perfiles clínicos heterogéneos, pero con necesidades nutricionales similares. Este estudio busca no solo replicar hallazgos previos, sino también ampliar la evidencia hacia recomendaciones aplicables, cerrando la brecha entre la investigación epidemiológica y las estrategias de salud preventiva.

# Metodología Estadistica Empleada

El estudio se basó en un enfoque metodológico de dos etapas: análisis descriptivo e inferencia estadística, utilizando modelos probabilísticos y pruebas paramétricas/no paramétricas según los supuestos verificados. A continuación, se detallan los métodos empleados:

1. Análisis Descriptivo Objetivo: Explorar la distribución, centralización, variabilidad y forma de las variables.

data("NutritionStudy")  
# Ver la estructura del conjunto de datos  
str(NutritionStudy)  
# Ver las primeras filas  
head(NutritionStudy)  
# Resumen estadístico básico  
summary(NutritionStudy)

Medidas calculadas:

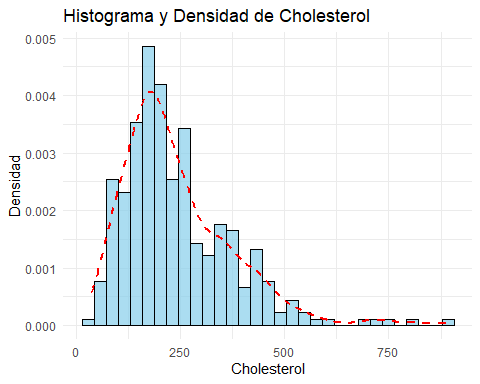
Centralización: Media, mediana.

Variabilidad: Desviación estándar, rango intercuartílico (IQR).

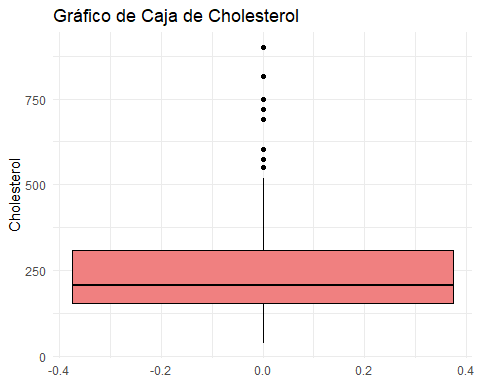
Forma: Asimetría (skewness) y curtosis.

Herramientas gráficas:

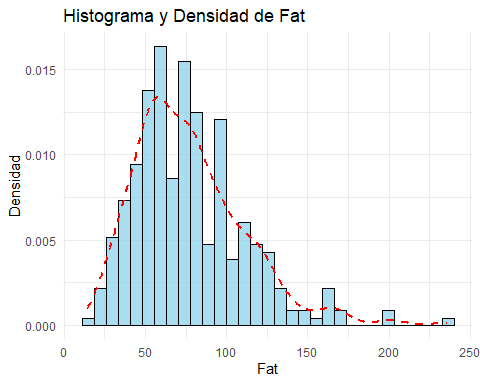
## ### Variable Cuantitativa: Cholesterol   
## #### Medidas de Centralización y Variabilidad  
## Media Mediana Desv. Estándar Rango Intercuartílico Mínimo Máximo  
## 1 242.4606 206.3 131.9916 153.85 37.7 900.7  
## #### Medidas de Posición y Curtosis  
## Q1 Q3 Asimetria Curtosis  
## 1 155 308.85 1.472997 6.33967  
## #### Gráfico Combinado (Histograma y Densidad)



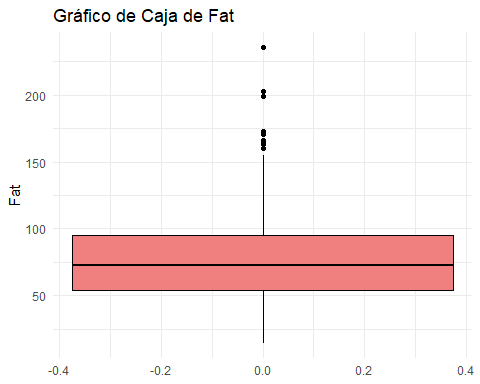
## #### Gráfico de Caja (Boxplot)



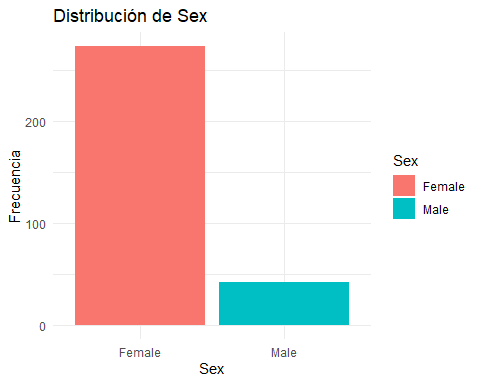
## ### Variable Cuantitativa: Fat   
## #### Medidas de Centralización y Variabilidad  
## Media Mediana Desv. Estándar Rango Intercuartílico Mínimo Máximo  
## 1 77.03333 72.9 33.82944 41.3 14.4 235.9  
## #### Medidas de Posición y Curtosis  
## Q1 Q3 Asimetria Curtosis  
## 1 53.95 95.25 1.098996 4.964799  
## #### Gráfico Combinado (Histograma y Densidad)



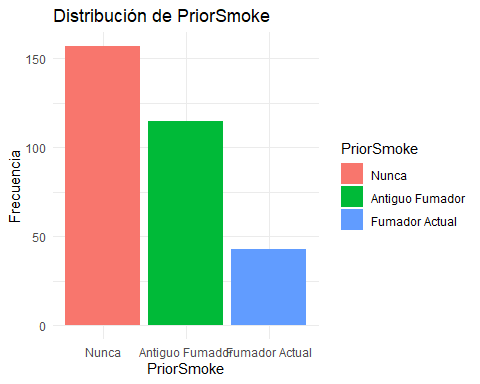
## #### Gráfico de Caja (Boxplot)



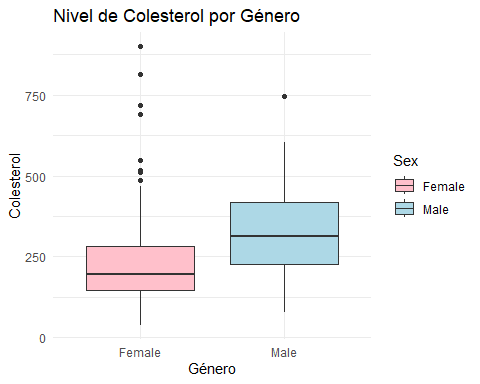
## ### Variable Cualitativa: Sex   
## #### Tabla de Frecuencia y Proporciones  
##   
## Female Male   
## 273 42   
##   
## Female Male   
## 0.8666667 0.1333333   
## #### Gráfico de Barras



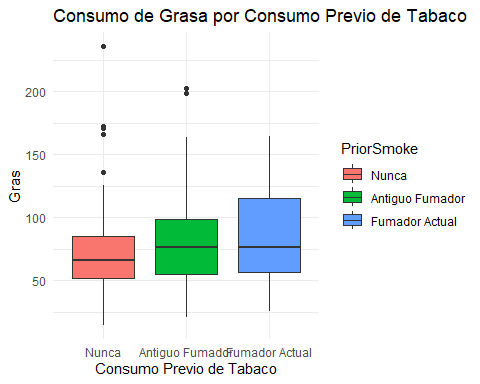
## ### Variable Cualitativa: PriorSmoke   
## #### Tabla de Frecuencia y Proporciones  
##   
## Nunca Antiguo Fumador Fumador Actual   
## 157 115 43   
##   
## Nunca Antiguo Fumador Fumador Actual   
## 0.4984127 0.3650794 0.1365079   
## #### Gráfico de Barras



## ### Gráfico de Colesterol por Género



##   
## ### Gráfico de Grasa por Consumo Previo de Tabaco



1. Verificación de Supuestos Normalidad Prueba: Shapiro-Wilk.

##   
## --- Verificación de Normalidad para: Cholesterol por Sex ---  
##   
## Normalidad para el grupo: Female ( Cholesterol ):  
## Test de Shapiro-Wilk:  
##   
## Shapiro-Wilk normality test  
##   
## data: var\_data  
## W = 0.87822, p-value = 6.17e-14  
##   
## El p-value ( 0 ) es menor que 0.05. Se rechaza la hipótesis nula de normalidad para este grupo.  
##   
## Normalidad para el grupo: Male ( Cholesterol ):  
## Test de Shapiro-Wilk:  
##   
## Shapiro-Wilk normality test  
##   
## data: var\_data  
## W = 0.96649, p-value = 0.2506  
##   
## El p-value ( 0.2506 ) es mayor o igual que 0.05. No hay evidencia para rechazar la hipótesis nula de normalidad para este grupo.

##   
## --- Verificación de Normalidad para: Fat por PriorSmoke ---  
##   
## Normalidad para el grupo: Nunca ( Fat ):  
## Test de Shapiro-Wilk:  
##   
## Shapiro-Wilk normality test  
##   
## data: var\_data  
## W = 0.90246, p-value = 1.016e-08  
##   
## El p-value ( 0 ) es menor que 0.05. Se rechaza la hipótesis nula de normalidad para este grupo.  
##   
## Normalidad para el grupo: Antiguo Fumador ( Fat ):  
## Test de Shapiro-Wilk:  
##   
## Shapiro-Wilk normality test  
##   
## data: var\_data  
## W = 0.95039, p-value = 0.0003201  
##   
## El p-value ( 3e-04 ) es menor que 0.05. Se rechaza la hipótesis nula de normalidad para este grupo.  
##   
## Normalidad para el grupo: Fumador Actual ( Fat ):  
## Test de Shapiro-Wilk:  
##   
## Shapiro-Wilk normality test  
##   
## data: var\_data  
## W = 0.9559, p-value = 0.09804  
##   
## El p-value ( 0.098 ) es mayor o igual que 0.05. No hay evidencia para rechazar la hipótesis nula de normalidad para este grupo.

Criterio: p>0.05 indica normalidad.

Acciones ante no normalidad:

Transformaciones (logarítmica, raíz cuadrada).

## La variable logarítmica de Colesterol cumple la normalidad en ambos grupos.

## La variable logarítmica de Grasa cumple la normalidad en todos los grupos.

Uso de pruebas no paramétricas (Mann-Whitney, Kruskal-Wallis).

1. Modelos Estadísticos y Pruebas de Hipótesis Distribución Normal y t-Student Uso: Comparar medias de dos grupos independientes.

Hipótesis: Hipótesis Nula (): No hay diferencia significativa en el nivel promedio de colesterol (o su logaritmo) entre hombres y mujeres. Hipótesis Alternativa (): Existe una diferencia significativa en el nivel promedio de colesterol (o su logaritmo) entre hombres y mujeres. (Bilateral)

Estimadores: Medias poblacionales de colesterol (o su logaritmo) para hombres y mujeres. Diferencia de medias muestrales (). Si se cumplen los supuestos, distribución t de Student.

## Procediendo con el test t paramétrico (variable transformada).

## Warning in leveneTest.default(y = y, group = group, ...): group coerced to  
## factor.

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 1 0.0961 0.7568  
## 313

## Las varianzas poblacionales se asumen iguales (p-value >= 0.05).

##   
## Resultado del T-test (varianzas iguales) con Cholesterol\_log :

##   
## Two Sample t-test  
##   
## data: Cholesterol\_log by Sex  
## t = -4.4934, df = 313, p-value = 9.876e-06  
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0  
## 95 percent confidence interval:  
## -0.5527150 -0.2160779  
## sample estimates:  
## mean in group Female mean in group Male   
## 5.302655 5.687051

##   
## Criterio de Decisión (Colesterol por Género):

## P-value: 0

## Intervalo de Confianza para la diferencia de medias: -0.55 a -0.22

## Dado que el p-value ( 0 ) es menor que 0.05, se rechaza la hipótesis nula.  
## Hay evidencia significativa para concluir que existe una diferencia en la media (o distribución) del colesterol (o su logaritmo) entre hombres y mujeres.

## El intervalo de confianza no incluye el cero, lo que refuerza la significancia (o indica que la diferencia es distinta de cero).

Hipótesis:

Hipótesis Nula (): No hay diferencia significativa en la ingesta promedio de grasa (o su logaritmo) entre los diferentes grupos de fumadores. Hipótesis Alternativa (): Al menos una de las medias de ingesta de grasa (o su logaritmo) es diferente entre los grupos de fumadores. (Bilateral) Parámetros a estimar: Medias poblacionales de ingesta de grasa (o su logaritmo) para cada grupo de fumadores. Estimadores: Medias muestrales de ingesta de grasa (o su logaritmo) para cada grupo. Modelo de probabilidad del estimador: Si se cumplen los supuestos, el ANOVA F-statistic.

## Procediendo con el ANOVA paramétrico (variable transformada).

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 2 0.365 0.6945  
## 312

## Las varianzas poblacionales se asumen iguales (p-value >= 0.05). Procediendo con ANOVA.  
## El ANOVA es significativo. Realizando pruebas post-hoc (Tukey HSD).  
## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = as.formula(paste(current\_fat\_var, "~ PriorSmoke")), data = NutritionStudy)  
##   
## $PriorSmoke  
## diff lwr upr p adj  
## Antiguo Fumador-Nunca 0.10821976 -0.01866621 0.2351057 0.1118241  
## Fumador Actual-Nunca 0.15835241 -0.01958155 0.3362864 0.0923372  
## Fumador Actual-Antiguo Fumador 0.05013265 -0.13465515 0.2349205 0.7987641

##   
## Criterio de Decisión (Grasa por PriorSmoke):

## P-value del ANOVA con Fat\_log : 0.0405

## Dado que el p-value ( 0.0405 ) es menor que 0.05, se rechaza la hipótesis nula.  
## Hay evidencia significativa para concluir que existe una diferencia en la media (o distribución) de la grasa (o su logaritmo) entre los grupos de fumadores.

# Presentación de Resultados

# Discución de los Resultados

# Recomendaciones

# Referencias Bibliográficas