Prueba de Hipótesis para Proyecto de Estadística Análisis de Valores Nutricionales por Tipo de Dieta

1 Introducción

Lo que haré es una prueba de hipótesis para cada tipo de dieta y otra que englobe a las cinco dietas (que sería para probar mi objetivo principal en este trabajo). Para ello, primero doy la versión general de lo que se quiere hacer, luego el procedimiento previo a verificar el hipótesis y por último de juego de hipótesis junto con la prueba de hipótesis que usaré.

Para las diferentes pruebas siempre se maneja el mismo nivel de significancia del $\alpha = 5\%$. Y la distribución de los macronutrientes, sin importar la dieta, siguen una distribución beta (su soporte es $[0, 1]$, los valores pertenecen a ese intervalo y su forma no es parecida a una distribución uniforme).

2 Dieta DASH

2.1 Visión General

Se quiere probar si la dieta está en balance nutricional, es decir, si se llega a consumir la misma proporción de macronutrientes de forma diaria.

2.2 Procedimiento

La cantidad diarias de comidas que se deberían que consumir son cinco, esto es equivalente a consumir cinco recetas. Por lo que para probar el balance bastaría con muestrear 50 días consumiendo comidas de la dieta DASH, es decir, muestrear de forma aleatoria 250 recetas y agruparlas en cinco en cinco recetas y obtener la proporción de macronutrientes que se consumen con las cinco recetas. Esto último es el conjunto de datos que se usará para la prueba de hipótesis.

2.3 Prueba de Hipótesis

Se tiene un conjunto de hipótesis por cada macronutriente, que lucen de la forma:

$$H_0 : \overline{x}_M = \frac{1}{3}$$
$$H_1 : \overline{x}_M \neq \frac{1}{3}$$

Donde M se refiere a uno de los tres macronutrientes con los que se cuenta. El signo de $\neq$ indica que la prueba será de dos colas debido a que se quiere capturar cualquier diferencia significativa por muy mínimo que sea.

Como el tamaño de la muestra es lo suficientemente grande, se puede suponer que la distribución de la media muestral es normal, por lo tanto la prueba que se va usar es la Prueba t según la homocedasticidad y se esperaría que diese resultados robustos.

3 Dieta Keto

3.1 Visión General

Se quiere probar si en todas las cocinas se respeta el principio de la dieta keto: El consumo de carbohidratos es menor que el consumo de grasas

3.2 Procedimiento

Lo único que se tiene que hacer es el separar las recetas de la dieta keto según el tipo de cocina y recuperar los datos acerca de los carbohidratos y grasas.

3.3 Prueba de Hipótesis

Se tiene un conjunto de hipótesis por cada tipo de cocina, que tienen la forma de:

$$H_0 : \overline{c}_C = \overline{f}_C$$
$$H_1 : \overline{c}_C < \overline{f}_C$$

Donde $\overline{c}$ y $\overline{f}$ indican las medias muestrales de carbohidratos y grasas reportadas por las recetas pertenecientes a la cocina C. El signo $<$ indica que la prueba será de una cola hacia la derecha debido a que se quiere probar si los carbohidratos son significativamente menores que las grasas.

Debido a que se cuenta con cocinas cuyo número de recetas son menores a 30 implica el uso inmediato de pruebas no paramétricas, es decir, usar la Prueba de Rangos de Wilcoxon. Para el caso cuando se cuenta con los suficientes datos, se usa la Prueba t dependiendo de la homocedasticidad de los datos.

4 Dieta Mediterránea

4.1 Visión Genera

Se quiere probar si existe una diferencia entre las recetas que provienen del mediterraneo en contraste con las recetas de los demás tipos de cocina.

4.2 Procedimiento

Se tiene primero que separar las recetas según si su tipo de cocina es del mediterraneo o no, y recuperar los valores referentes a los macronutrientes.

4.3 Prueba de hipótesis

Se tiene un conjunto de hipótesis por cada macronutriente, que lucen de la forma:

$$H_0 : F_m^M = F_m^O$$
$$H_1 : F_m^M \neq F_m^O$$

Donde F_m^M y F_m^O se refieren a la distribución del macronutriente m en las recetas del mediterraneo y en otras cocinas, respectivamente. El $\neq$ indica que será una prueba de dos colas debido a que se quiere capturar cualquier mínima diferencia.

Por como se definió, se tiene una prueba no paramétrica donde las distribuciones de los datos son no normales, por lo que la prueba que se usa es la Prueba Kolmogorov-Smirnov (aunado a que se cuentan con muestras de tamaño considerable).

5 Dieta Paleo

5.1 Visión Genera

Se quiere probar si existe una diferencia significativa en los aportes de proteínas en las diferentes cocinas.

5.2 Procedimiento

No es necesario realizar un procesamiento adicional a obtener las recetas de cada tipo de cocina y de recuperar las proteínas de ellas. Lo que si se realiza es un filtrado para eliminar los tipos de cocinas que tengan menos de 5 recetas, esto debido a la prueba que se usa.

5.3 Prueba de hipótesis

Debido a que las distribuciones son no normales y se podría suponer que los tipo de recetas no están relacionadas, se tiene que lo que se va a probar es que las medianas de los tipo de cocina son iguales. Por lo que el juego de hipótesis se puede poner como:

$$H_0 : \text{Todas las medianas son iguales}$$
$$H_1 : \text{Por al menos una mediana es diferente}$$

Como lo anterior luce como una versión no paramétricas de ANOVAS e investigando sobre pruebas para este estilo, la prueba que se usa es la Prueba de Kruskal-Wallis, que permite usar datos de cualquier tipo de distribución, junto con la Prueba de Dunn para determinar en qué parejas existe una diferencia significativa.

6 Dieta Vegana

6.1 Visión Genera

Se quiere probar si en todas las cocinas se verifica que la ingesta de carbohidratos es mayor que la de proteínas.

6.2 Procedimiento

Además de obtener las recetas por tipo de cocina y de recuperar los valores de carbohidratos y proteínas, también es necesario de eliminar los tipo de cocina donde se cuenta con registros (recetas) menos de cinco recetas.

6.3 Prueba de hipótesis

Para tipo de cocina tiene un conjunto de hipótesis que tienen la forma:

$$H_0 : \overline{p}_C = \overline{c}_C$$
$$H_1 : \overline{p}_C < \overline{c}_C$$

Donde $\overline{p}$ y $\overline{c}$ indican las medias muestrales de proteínas y carbohidratos reportadas por las recetas pertenecientes a la cocina C. El signo $<$ indica que la prueba será de una cola hacia la derecha debido a que se quiere probar si las proteínas son significativamente menores que los carbohidratos.

Debido a que se cuenta con cocinas cuyo número de recetas son menores a 30 implica el uso inmediato de pruebas no paramétricas, es decir, usar la Prueba de Rangos de Wilcoxon. Para el caso cuando se cuenta con los suficientes datos, se usa la Prueba t dependiendo de la homocedasticidad de los datos.

7 Entre todas las Dietas

7.1 Visión Genera

Se quiere probar si existe una diferencia significativa en los aportes nutricionales entre las diferentes dietas sin importar el origen de las recetas

7.2 Procedimiento

No es necesario hacer nada más que recuperar las recetas junto sus macronutrientes de cada dieta.

7.3 Prueba de hipótesis

Para cada tipo de dieta y macronutriente, lo que se prueba es lo siguiente: Si sus distribuciones son significativamente diferentes, en el sentido de que siguen distribuciones con diferentes parámetros, se usa la Prueba de Kolmogorov-Smirnov debido a que se quiere capturar cualquier mínima diferencia en las distribuciones. Se tiene el siguiente conjunto de hipótesis:

$$H_0 : F_m(x) = G_m(x)$$
$$H_1 : F_m(x) \neq G_m(x)$$

Donde F_m y G_m es la distribución del macronutriente m en dos dietas diferentes. De $\neq$ indica usar una Prueba de Kolmogorov-Smirnov de dos colas para detectar cualquier diferencia en las distribuciones.

Para poder dar un valor para saber qué tan diferentes son las dietas, lo que se hace es promediar los resultados de cada prueba donde 0 significa que no existe una diferencia significa y un 1 que si la hay. Por lo tanto, mientras más se aproxima a 1, menos parecidas son las dietas.