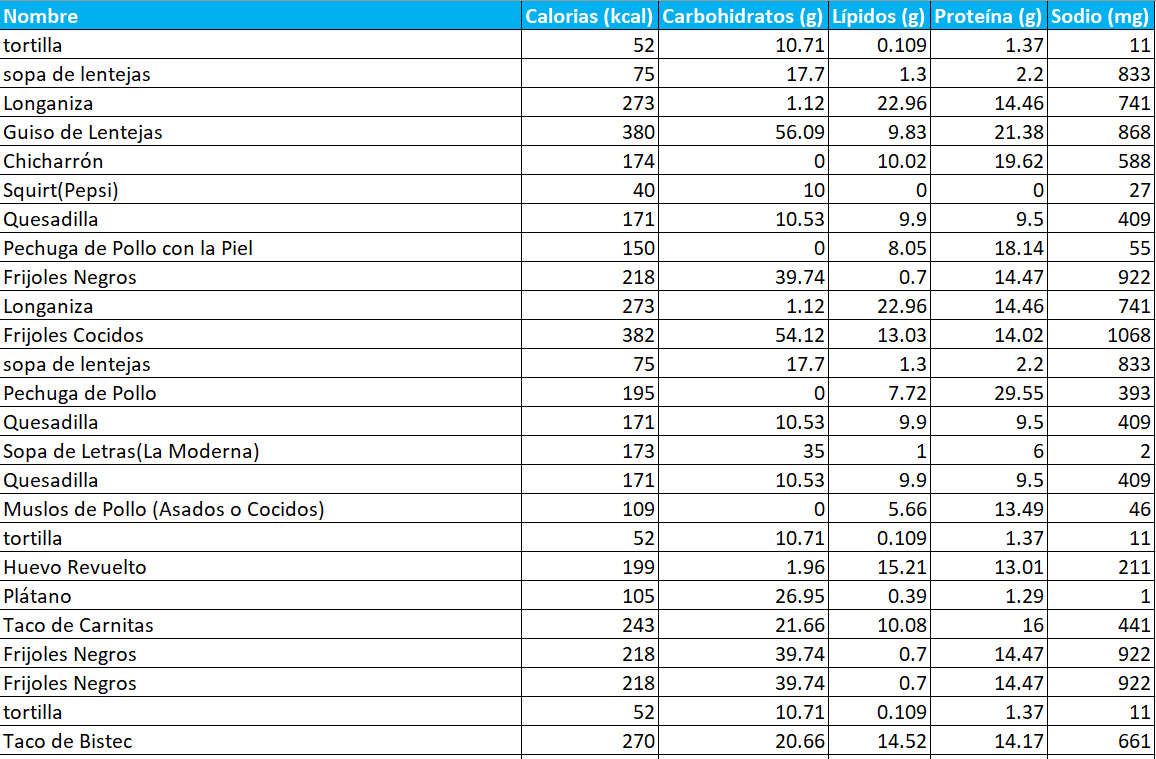
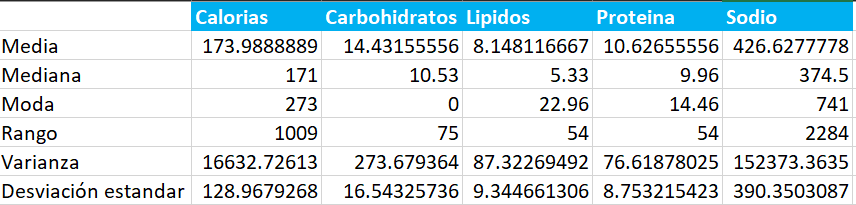
El presente trabajo describe la forma en como se realizó el procedimiento para obtener la ecuación de regresión para describir el comportamiento que hay entre la relación de los carbohidratos, los lípidos, las proteínas y el sodio con las calorías por lo que el procedimiento se describe a continuación.

Primero se llevo a cabo un seguimiento del consumo nutricional durante cinco semanas como se muestra en la imagen de abajo

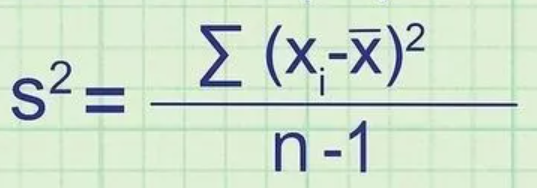


Los datos presentes son registros de los alimentos ingeridos durante las cinco semanas donde también están presentes bebidas, estos datos fueron registrados con en la aplicación de Samsung Health diariamente y donde automáticamente rellenaba las filas de las columnas con los datos equivalentes a la cantidad de consumo por alimento.

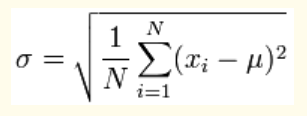
A continuación, se calcularon: la media, la mediana, la moda, el rango, la varianza, la desviación estándar dando como resultado los siguientes datos:



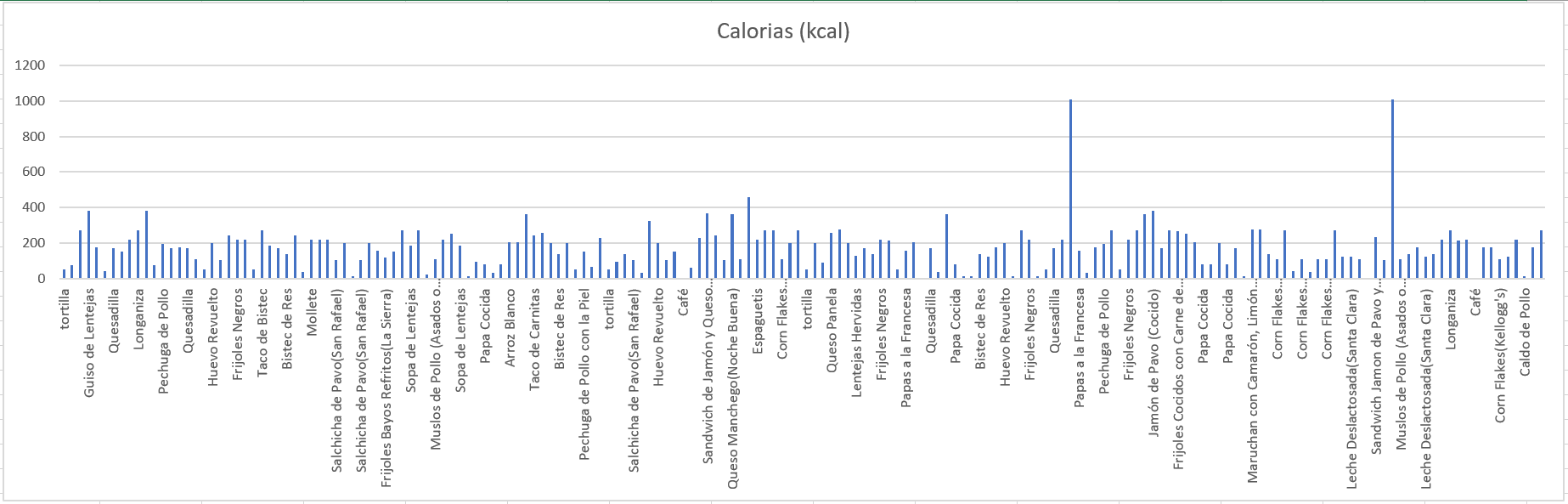
La formula para calcular la media es la suma de todos los datos dividido entre el número de datos, para calcular la mediana primero se ordenan los datos de menor a mayor y luego se calcula el valor que está en medio, para calcular la moda primero se ordenan los datos y se identifica cual se repite mas y ese es la moda, para calcular el rango se resta el numero más grande menos el número más pequeño, la varianza se calcula con la siguiente formula



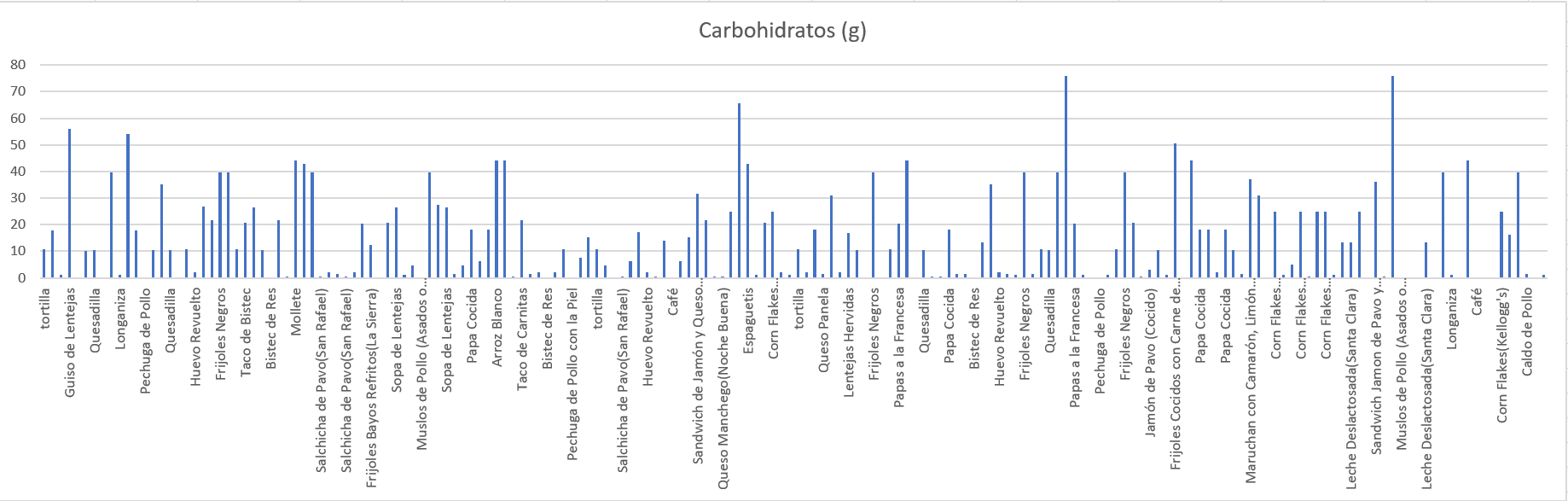
Para calcular la desviación estándar se calcula con la siguiente formula:



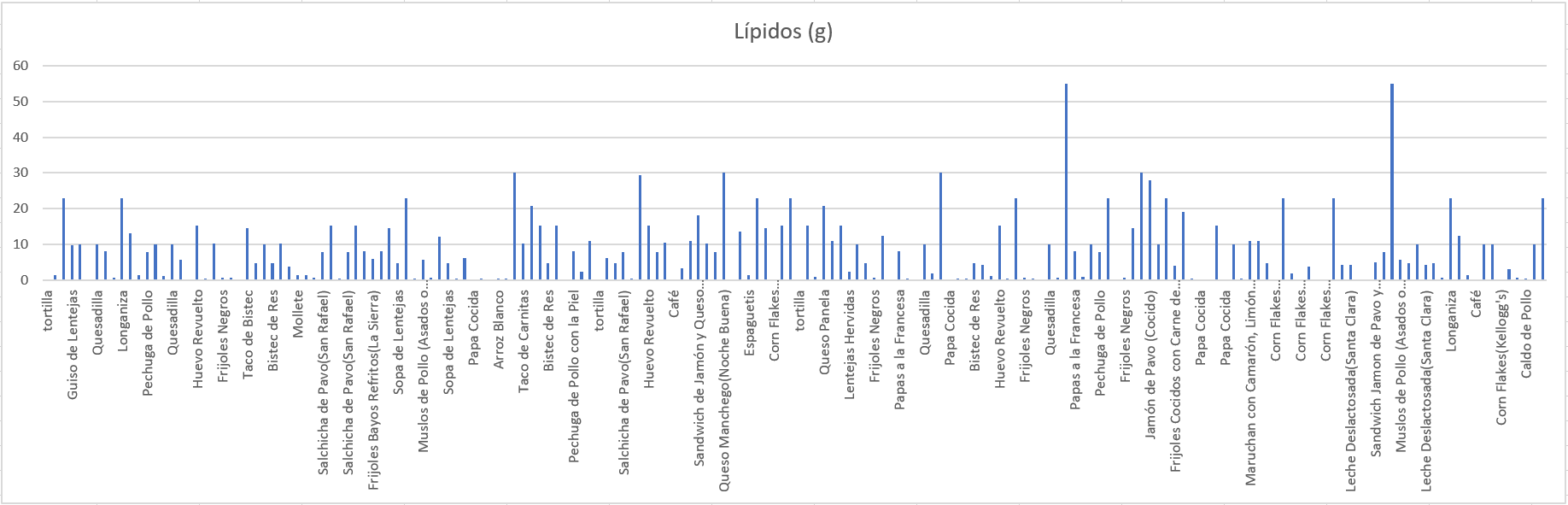
Luego se realizó un histograma de los datos obtenidos de cada variable (calorías, carbohidratos, los lípidos, las proteínas y el sodio):



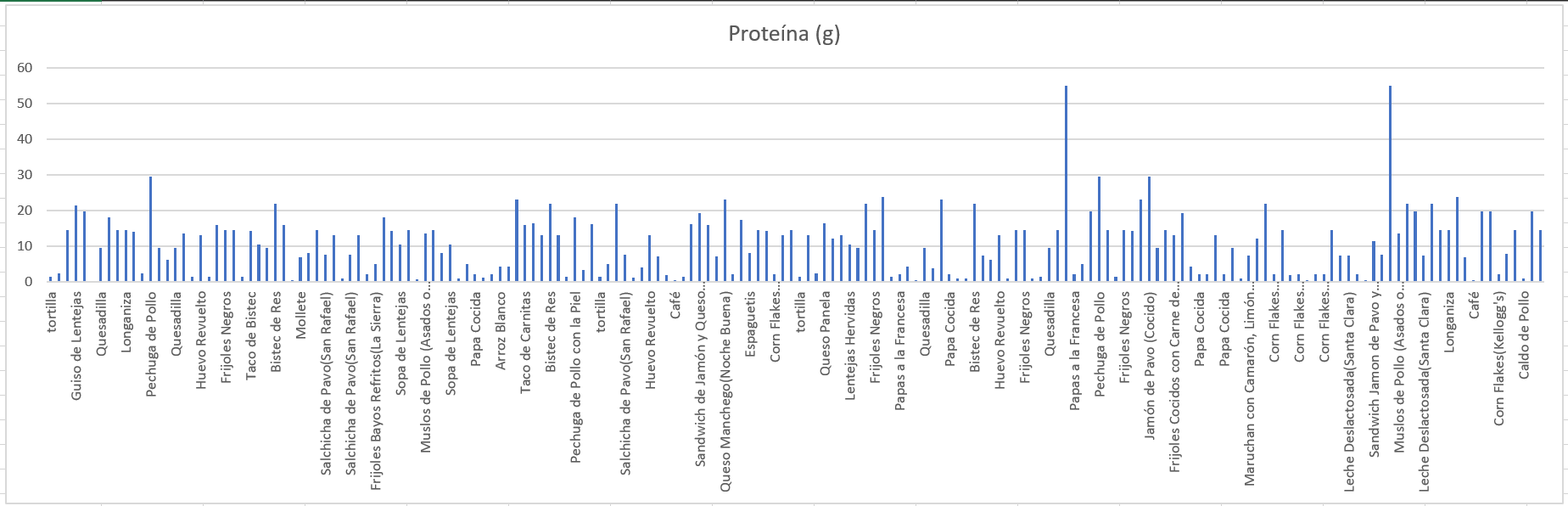
En la grafica nos podemos dar cuenta de que hay dos alimentos que resaltan con un mayor nuero de calorías estos alimentos que resaltan son de comida rápida y solo fueron ingeridos dos veces en todo el tiempo que se realizo el estudio, el resto de los alimentos no pasan del 400 por lo que mi ingesta calórica por alimento es baja.



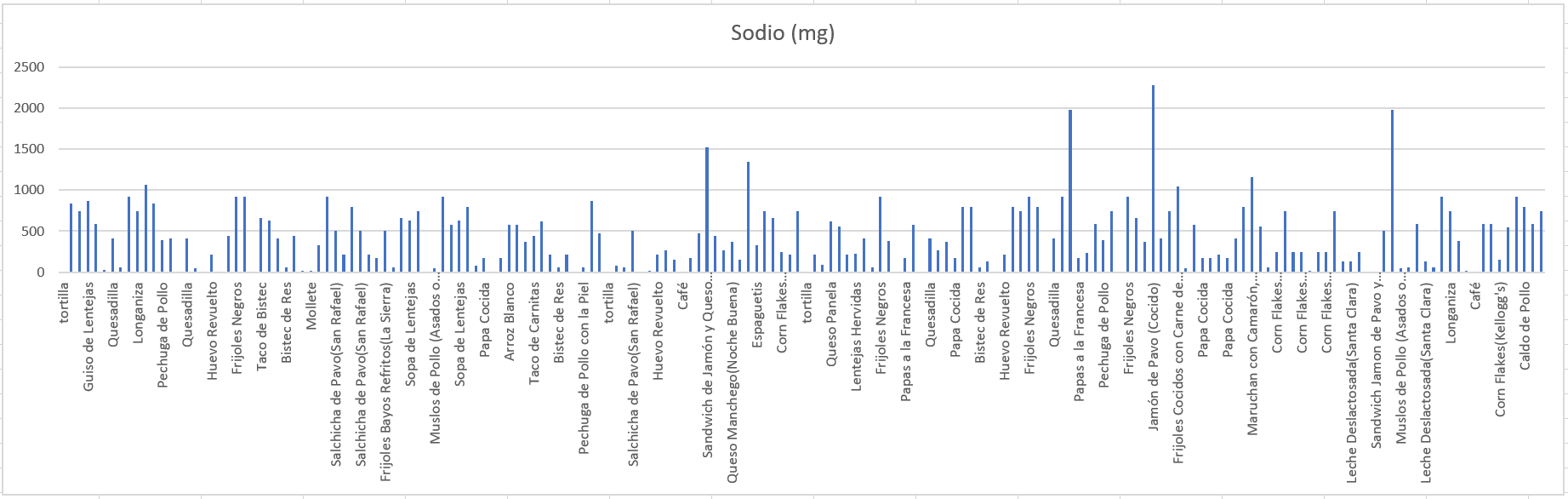
En la grafica de los carbohidratos podemos darnos cuenta de que la ingesta no pasa de los 80 gramos y que es son muy grandes la mayoría de los valores dentro de esta grafica donde las pastas y las legumbres y panes son los que más resaltan en la grafica por lo que se puede decir que son muy ricos en carbohidratos.



En esta grafica ahora podemos ver como hay dos alimentos que destacan y otros que también son notorios como los alimentos de origen animal como ciertos embutidos y como es el pollo, nos podemos dar cuenta que en esta grafica no pasa de los 60 gramos.



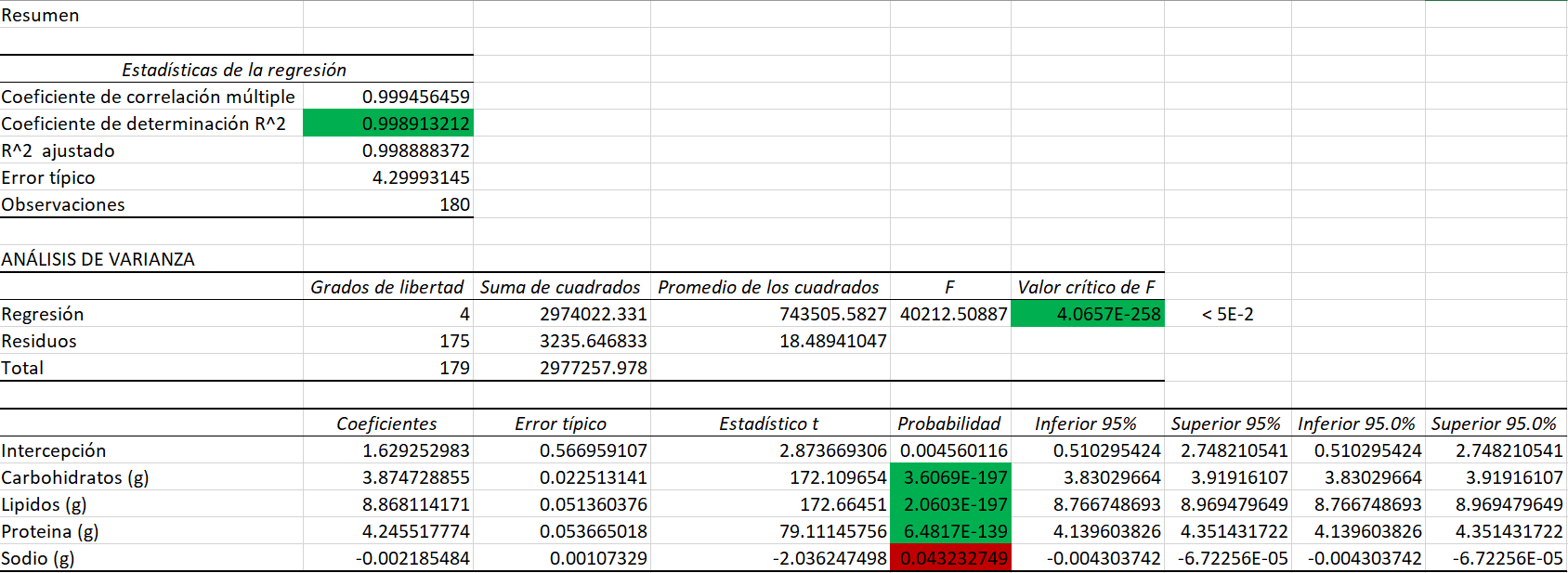
En la gráfica de proteína los datos no pasan de los 60 gramos donde los alimentos que más destacan son los de origen animal siendo así que los alimentos de origen animal tienen más proteína que los algunos de origen vegetal



En la gráfica de sodio podemos percatarnos que son las legumbres y los cereales los que destacan con un mayor número de sodio que el resto de los alimentos, también nos damos cuenta de que son números muy grandes no pasando de los 2500.

Después de haber realizado todo el procedimiento anterior se continuo con el inicio del análisis del modelo de regresión.

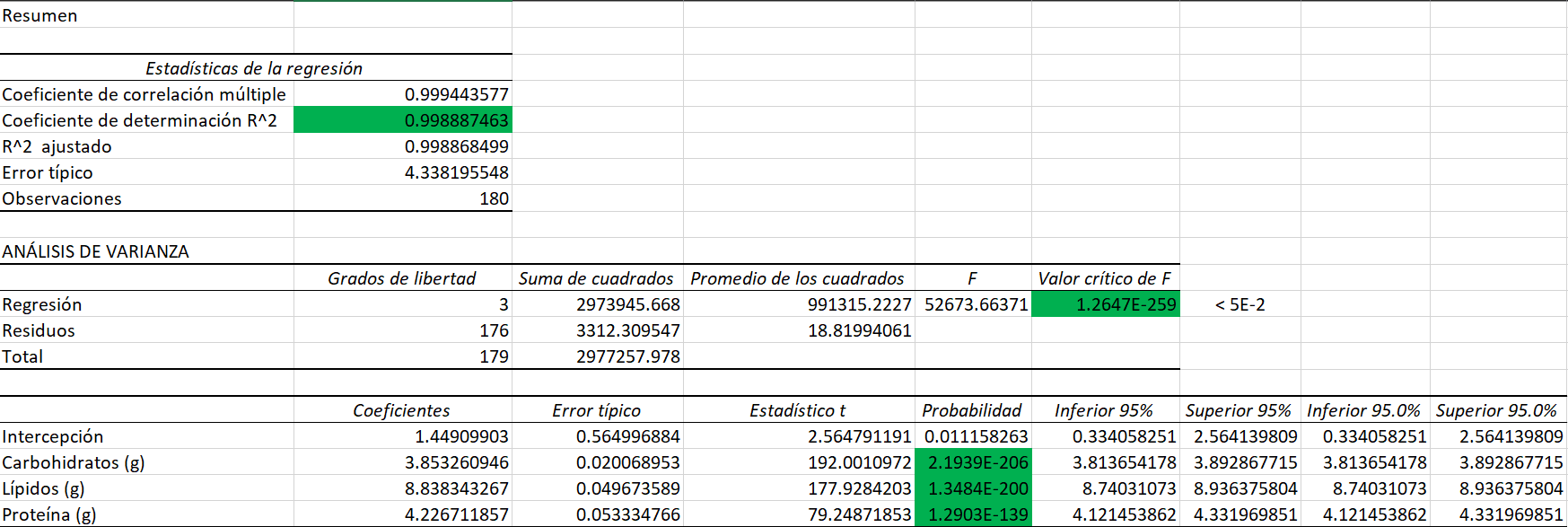
En esta etapa se procedió en identificar primero cuales son las variables que hay las cuales quedaron como: calorías como variable dependiente, carbohidratos, lípidos, proteína, sodio como variable independiente. Una vez identificado esto se buscó la realización de la tabla ANOVA dentro de nuestro análisis de regresión dando como resultado:



Donde nos pasamos a identificar el coeficiente de determinación R^2 cuyo valor es de 0.998913212 lo que es un número muy cercano a 1 donde el 1 seria que el modelo es exacto. También se identificó el valor critico que es de 4.0657E-258 cuyo valor debe ser menor a 5E-3 por lo que si hubiera pasado de este valor la hipótesis planteada hubiera sido desechada.

Se paso a revisar los resultados de probabilidad de la donde todo valor grande marca que este debe ser descartado como una variable independiente que afecte a la función que nos permitirá predecir el valor de calorías.

Al ser el sodio muy grande pasa a ser descartado y se vuelve a realizar el análisis de regresión sin el sodio porque era un valor muy grande, ahora el resultado de los valores de interés para esta practica son cada vez más exactos considerando que este análisis de regresión es correcto con una mayor exactitud que la anterior como se muestra a continuación:

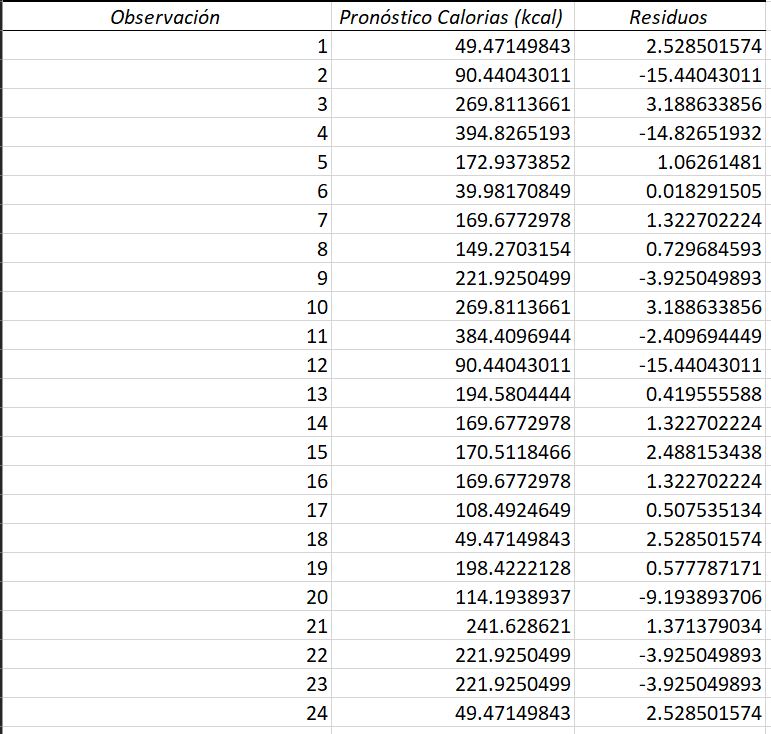


Por lo que plantea que nuestra ecuación de regresión es la siguiente:

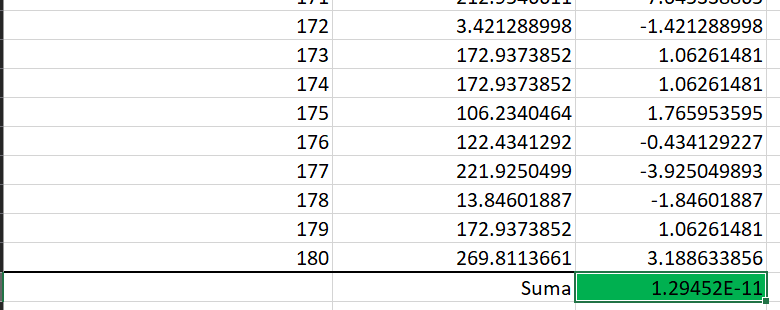
y = x1\*carbohidratos + x2\*lípidos + x3\*proteína

donde ‘y’ es el valor a predecir y las ‘x#’ son las variables independientes donde se va a registrar los valores de un alimento multiplicado respectivamente por los coeficionetes.

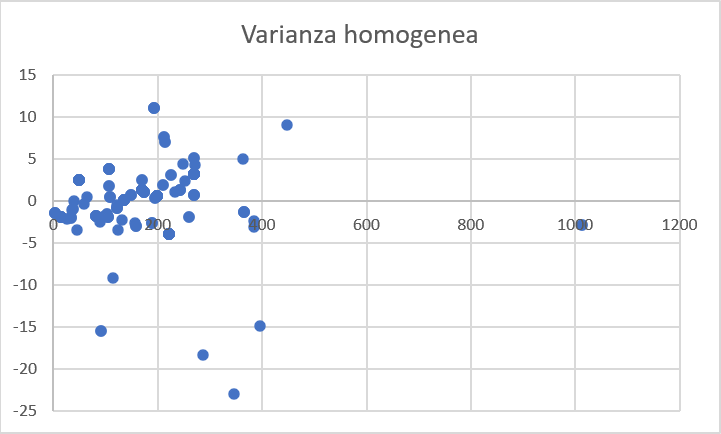
El tener todos los valores dentro de lo esperado se procede a comprobar el modelo haciendo un análisis de residuos como se muestran en la siguiente tabla:



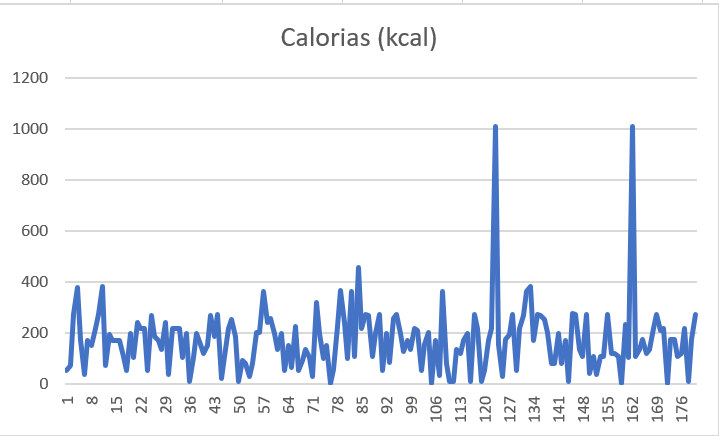
Los datos obtenidos se suman y la suma de esos datos debería dar cero en un caso perfecto, pero en este caso debe de dar un valor muy pequeño casi cercano al cero.

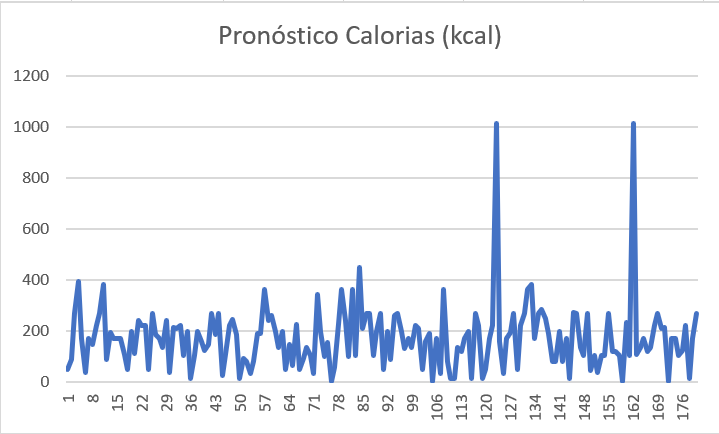


Este resultado nos permite saber que nuestro modelo de regresión es bueno al permitir defender lo planteado. Luego se continua a realizar las gráficas:

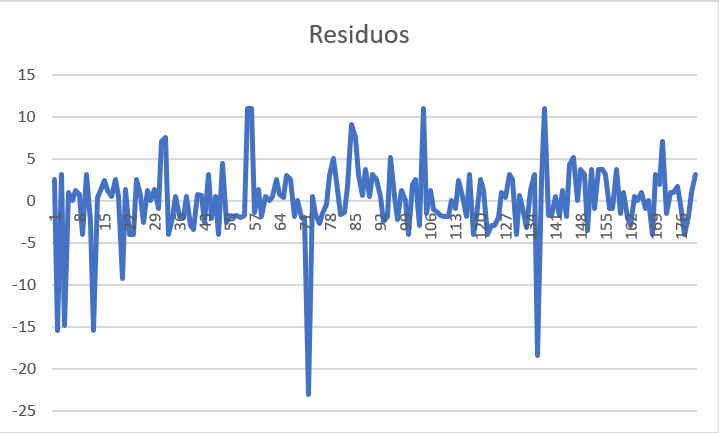


Aquí la grafica de la varianza homogénea nos permite ver que las calorías están muy cercanas al cero por lo que el consumo de calorías es bajo, pero el valor que está más alejado muestra que hubo un alimento que sobre pasó el consumo común de calorías.

Esta grafica nos permite ver el consumo real de las calorías durante las semanas en que se llevó acabo el registro.

Esta grafica nos permite ver lo que se pronostica usando la ecuación de regresión planteada.

Podemos darnos cuenta de que es muy parecida a la grafica con los valores reales por lo que se puede decir que es un buen modelo.

Esta es la gráfica de residuos que muestra la prueba de independencia de los datos calculados como residuos.

Ahora, en esta parte se viene a comprobar la eficiencia de la ecuación de regresión obtenida usando tres alimentos registrados y tres alimentos no registrados para luego compararlos con los valores reales de calorías de cada alimento:

Alimentos registrados:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nombre | Calorías | Carbohidratos | Lípidos | Proteína | Sodio |
| Frijoles Cocidos | 382 | 54.12 | 13.03 | 14.02 | 1068 |
| Quesadilla | 171 | 10.53 | 9.9 | 9.5 | 409 |
| Taco de Carnitas | 243 | 21.66 | 10.08 | 16 | 441 |

Se ingresan en la formula y = x1\*carbohidratos + x2\*lípidos + x3\*proteína

Donde x1 = 192.001097188719, x2 = 8.83834326692382 y x3 = 4.22671185655216

Dando como resultado:

Frijoles = 382.9605954

Quesadilla = 168.2281987

Tacos de carnitas = 240.1795219

Comparando los números obtenidos con los reales nos podemos dar cuenta que están muy cerca de ser iguales con un margen de error no mayor a 3

Alimentos no registrados:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nombre | Calorías | Carbohidratos | Lípidos | Proteína | Sodio |
| Ensalada mixta | 135 | 5.5 | 11.6 | 3.6 | 291 |
| Tacos de barbacoa | 196 | 16.1 | 10.1 | 9.7 | 348 |
| Taco de Canasta | 129 | 16.3 | 5.6 | 4 | 136 |

Se ingresan en la formula y = x1\*carbohidratos + x2\*lípidos + x3\*proteína

Donde x1 = 192.001097188719, x2 = 8.83834326692382 y x3 = 4.22671185655216

Dando como resultado:

Ensalada mixta = 138.9338798

Tacos de barbacoa = 192.3038732

Tacos de canasta = 129.2097231

Comparando los números obtenidos con los reales nos podemos dar cuenta que están muy cerca de ser iguales con un margen de error no mayor a 3