Calories as a Functions of Other Nutrients
(“Situación Problema” MA1042.66)

BY:

Celeste Adriana Paz Hernández A01656010

Ana Fernanda Flores Gámez A01662498

Elid Sebastián López Pineda A01654141

Amy Vanesa Díaz García A01656419

Esther Karyme Chávez Flores A01654913

Abstract

In this evidence we will get the next formula for finding a model to predict calories.

$$Cal = 3.9584C+4.1011P+8.8663F$$

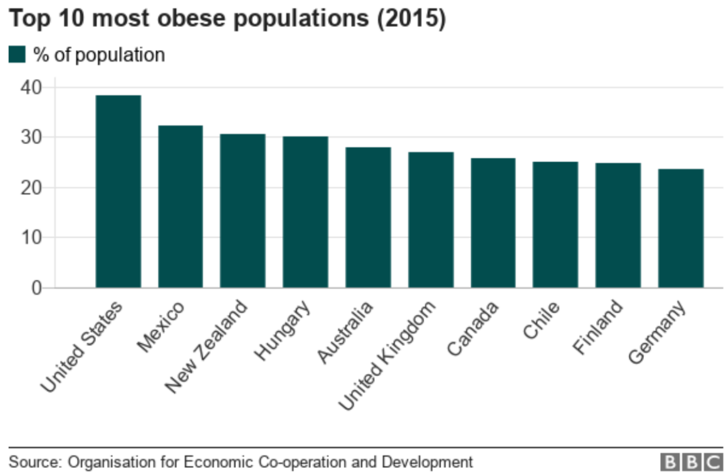
In the formula Cal means the amount of kilocalories of a food, C is the amount of carbohydrates, P the amount of proteins and F the amount of fats. The formula was based on a food binnacle of about 10 weeks of everyday documentation, in which we search about the sods, carbohydrates, proteins, fats and calories of each food through the Nutritionix web site (https://www.nutritionix.com/), all the recopilation to make three models of linear regression. In brief we will argue why we choose a model with hypothesis testing. In this document we use the Excel software, the Python programming language in Google Colab and the statistic system of Minitab.

Contents:

1. Introduction

According to the newspaper La Jornada, in Mexico overweight and poor nutrition predominate in 70% of adults. This article mentions that Mexico is in the first place among Latin American countries leading to a high rate of obesity and overweight and in the sale of processed products, like rich in sugars and fats such as fried foods or sugary drinks . The interesting fact is that several scientists have determined that obesity is one of the conditions of greatest vulnerability to Covid-19 (coronavirus), so people who are overweight or obese should be careful, in the case of contracting this virus are more likely to die from vulnerability .

According to the Organisation for Economic Co-operation and Development (OECD), made up of 35 major economies, we can see in the graph below that the US is the most obese country, where 38.2% of the population over the age of 15 is obese. Mexico is second with 32.4% of the population and New Zealand is third with 30.7%.



In Mexico, obesity rates are higher among women than men. The OECD average is 19.5%. The OECD works out the Body Mass Index (BMI). for different countries using estimates of height and weight from surveys and measured data based on estimates of height and weight from health examinations. Long term projections suggest obesity levels are expected to rise in Mexico until 2030, as they are in countries including the US, England, Canada, France and Spain.

2. Model assumptions

Throughout this paper, we will assume the existence of the model

$$Cal = \beta_0 + \beta_1 S + \beta_2 C + \beta_3 P + \beta_4 F + E$$

where C is the amount of kilocalories, S is the amount of sodium, C is the amount of carbohydrates, P is the amount of proteins, F is the amount of fats and E is a random variable with cero of measure constant variance and normal distribution.

We also will assume:

- We have a random sample
- Our data have diversity
- When there are not consume Sods, Prots, Fats, Carbs the calories will we cero
- We will assume the Parsimony principle
- we will take a significance level of $\alpha = 0.01$

3. The cleaning of our data

In our models we clean off the data from the ones that were atypical when we run the models, taking importance on normal graphics and the residuals of the third model

Nutritionix	Food	Sods	Carbs	Prots	Fats	Cals
https://www.n	bbq boneless	3610	89	50	60	1100
https://www.n	breakfast burri	930	27	15	370	370
https://www.n	Fajitas	1816	124	97	55	1401
https://www.n	ma po tofu	2290	57	57	59	920
https://www.n	boiled egg	62	78	6.3	5.3	78
https://www.n	Tortilla Wrap	280	15	5	1.5	50
https://www.n	Rendang	84	9.9	36	27	404
https://www.n	saudi banana b	345	79	8.7	13	438
https://www.n	chips tomato s	940	74	7	25	570
https://www.n	alfredo sauce p	65	25	1	2.5	25
https://www.n	chicken pesto p	1200	38	35	29	530
https://www.n	smores	370	85	8	31	610

Table 1: Atypical data

3.1 Correlation between our variables

	Sods	Fats	Carbs	Prots	Cals
Sods	1.000000	0.635082	0.385197	0.659620	0.654417
Fats	0.635082	1.000000	0.551058	0.621686	0.916818
Carbs	0.385197	0.551058	1.000000	0.276773	0.806184
Prots	0.659620	0.621686	0.276773	1.000000	0.673417
Cals	0.654417	0.916818	0.806184	0.673417	1.000000

Table 2 :Correlation

We make this correlation table in python where we also include a pairplot to indicate the correlation between the variables. We also can notice that there is a correlation between Sods, Fats and Prots, where fats have 0.635 and proteins 0.659, that maybe means that in the foods we choose the dishes were proteins with sods and fat.

4. Training with cleaning data

Previously of cleaning the data we train each model, we obtain the next result of linear regression for each model

```
################################################################################
MODEL1:With intercept, with sods
################################################################################
                            OLS Regression Results
==============================================================================
Dep. Variable:                   Cals   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.999
Method:                 Least Squares   F-statistic:                 3.829e+04
Date:                Fri, 03 Jun 2022   Prob (F-statistic):          6.68e-172
Time:                        00:28:19   Log-Likelihood:                -377.47
No. Observations:                 115   AIC:                             764.9
Df Residuals:                     110   BIC:                             778.7
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.7164      1.117     -0.641      0.523      -2.930       1.497
Sods          -0.0012      0.002     -0.735      0.464      -0.004       0.002
Fats           8.8791      0.066    135.158      0.000       8.749       9.009
Carbs          3.9694      0.027    146.335      0.000       3.916       4.023
Prots          4.1455      0.070     59.627      0.000       4.008       4.283
==============================================================================
Omnibus:                        6.201   Durbin-Watson:                   1.657
Prob(Omnibus):                  0.045   Jarque-Bera (JB):                5.687
Skew:                          -0.516   Prob(JB):                       0.0582
Kurtosis:                       3.347   Cond. No.                     1.59e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.59e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
################################################################################
```

Table 3: Results of model 1

This model just sets the model supposed with the data obtained from Nutritionix: This one has an intercept and also all the variables including Sodium. It is important to take into consideration that the condition value is too high.

```
################################################################################
MODEL2: No intercept, with sods
################################################################################
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   Cals   R-squared (uncentered):                   1.000
Model:                            OLS   Adj. R-squared (uncentered):              1.000
Method:                 Least Squares   F-statistic:                          1.081e+05
Date:                Fri, 03 Jun 2022   Prob (F-statistic):                   2.91e-198
Time:                        00:28:20   Log-Likelihood:                         -377.69
No. Observations:                 115   AIC:                                      763.4
Df Residuals:                     111   BIC:                                      774.4
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Sods          -0.0013      0.002     -0.860      0.392      -0.004       0.002
Fats           8.8822      0.065    135.951      0.000       8.753       9.012
Carbs          3.9627      0.025    159.046      0.000       3.913       4.012
Prots          4.1314      0.066     62.792      0.000       4.001       4.262
==============================================================================
Omnibus:                        4.507   Durbin-Watson:                   1.674
Prob(Omnibus):                  0.105   Jarque-Bera (JB):                3.988
Skew:                          -0.441   Prob(JB):                        0.136
Kurtosis:                       3.231   Cond. No.                         110.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
################################################################################
```

Table 4: Results of model 2. This model takes out the intercept and keeps the sodium variable. If our null hypothesis were true, which means if the sodium coefficient were 0 instead of -0.0013, then the probability of getting this data or more extreme is 0.392, which is bigger than alpha (ɑ=0.01). Therefore, we fail to reject the null hypothesis. In other words, this data is consistent with the assumption that the coefficient of sodium is 0.

```
################################################################################
MODEL3: No intercept, No sods
################################################################################
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   Cals   R-squared (uncentered):                   1.000
Model:                            OLS   Adj. R-squared (uncentered):              1.000
Method:                 Least Squares   F-statistic:                          1.445e+05
Date:                Fri, 03 Jun 2022   Prob (F-statistic):                   1.02e-200
Time:                        00:28:20   Log-Likelihood:                         -378.07
No. Observations:                 115   AIC:                                      762.1
Df Residuals:                     112   BIC:                                      770.4
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Fats           8.8664      0.063    141.641      0.000       8.742       8.990
Carbs          3.9585      0.024    162.184      0.000       3.910       4.007
Prots          4.1012      0.056     73.866      0.000       3.991       4.211
==============================================================================
Omnibus:                        3.843   Durbin-Watson:                   1.678
Prob(Omnibus):                  0.146   Jarque-Bera (JB):                3.370
Skew:                          -0.410   Prob(JB):                        0.185
Kurtosis:                       3.173   Cond. No.                         5.96
==============================================================================

-0.2673274944765053
```

Table 5: Results of model 3

In this model there is no intercept and no sods variable.

5. From Model 1 to Model 2

The reason why we moved from model 1 to model 2, is because one of the assumptions from which our model is based is that by not consuming proteins, carbohydrates, sods or fats, the total calories would be 0 however in the first model apart from those variables also has an intercept.

$$Cal = \beta_0 + \beta_1 S + \beta_2 C + \beta_3 P + \beta_4 F + E$$

$$0 = \beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0) + \beta_4(0) + E$$

$$0 = \beta_0$$

So our demonstration is that the intercept is equal to 0. This is the reason why we incline for model 2.

From Model 2 to Model 3

In order to explain the reason we moved from model 2 to model 3, we are going to make use of hypothesis testing, and later we will apply ockham's razor.

Let

H_0: Sodium coefficient = 0

$H_a: Sodium\ coefficient \neq 0$

As mentioned over table 4, if H_0 were true, then the probability of getting this data or more extreme is $0.392 > \alpha = 0.01$. Therefore, we fail to reject the H_0

The meaning of these results is that, within reasonable doubt, our model is also consistent with a model that does not have sodium. Then, comparing R^2 from model 2=1.000 with R^2 from model 3=1.000 we notice they both explain our data exactly the same, but model 3 is simpler. Therefore, because of Ockham's razor principle, we incline for model 3.

7 Why no other Model?

To argue the reason we did not choose another model to describe our data, we will make hypothesis testing. Just as in the last step, where we assumed the coefficient of sodium was zero, we might just try the same with any other variable, for instance with Fats.

Let

H_0: Fats coefficient = 0

H_a: $Fats\ coefficient\ \neq 0$

Based on the information from model 3, if H_0 were true, then the probability of getting this data or more extreme is 0.000> ɑ = 0.01. Therefore, we reject the null hypothesis.

As it is visible on the table of model 3, the p value for the other 2 variables left (carbs and prots) is the same as fats, 0.000. That's why no other model was made.

8 Consistency of our model with our assumptions

- At this part of the document, we are going to check our data with respect to linear regression. Well, our first step is to check if our residuals are closer to cero, which is consistent with the assumption that *E* has a mean of *0.*

-0.2673274944765053

Table 6: Residual results of model 3

- Our second step would be to check our histogram, the shape of our histogram. As we can see, our histogram has a not so irregular bell shape. Thus it is consistent with the assumption that the error distribution is normal.

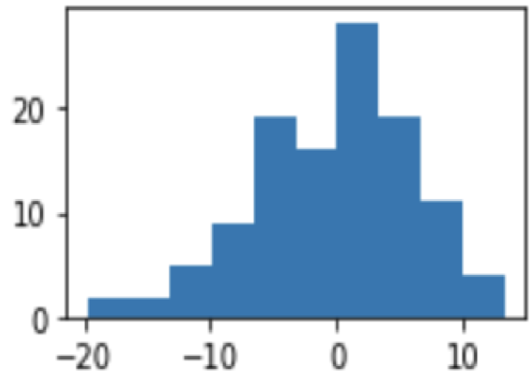


Figure 1: The histogram of residuals has a bell shape, which is consistent with the assumption that our errors have a normal distribution

- As we can see, in this *probplot* the red line have to be close to the blue dots, is consistent with the assumption and a normal distribution in our data

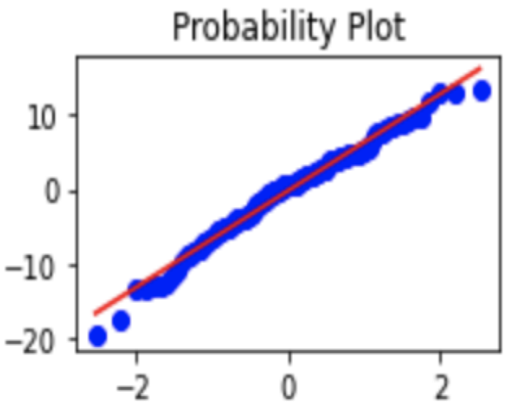


Figure 2: After separating our outlier data, our residuals data is very close to our theoretical values.

- Now, in the residuals vs order graphic, we have a constant width and we can't see clusters or patterns which defines that our choice of random samples is consistent and the error has a constant variance.

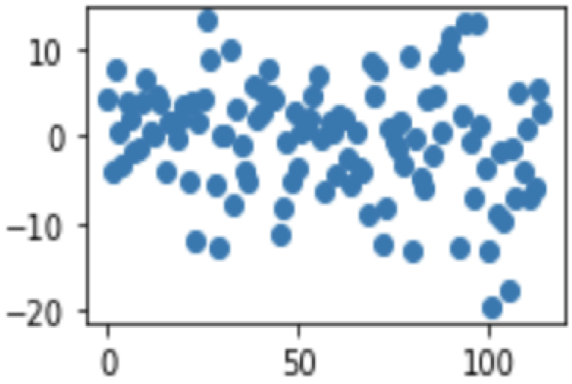


Figure 3: In the residual vs order plot, there are no patterns or clusters, which matches our random sample.

9 Data validation

Finally, we compare the results of our model with the following specified foods: salt, egg whites, oil and sugar. So, our final formula of our model is:

$$\text{Cal}=3.958C + 4.101P + 8.866F$$

Food	Sods	Carbs	Prots	Fats	Cals	Modelo 3
Salt	581	0	0	0	0	0
Egg whites	55	0.2	3.6	0.1	17	16.426
Oil	0	0	0	14	124	124.04
Sugar	0	4.2	0	0	16	16.38

Table 7: comparison of our model and Nutritionix data

In the table above, it can be seen that the results are very close to the original data, concluding that our model almost perfectly predicts the calories in our food.

References

Appendix 1: Python Code

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.formula.api as smf
import seaborn as sns
d=pd.read_csv('data nutrix.csv' , index_col=2)
d.drop_duplicates(inplace= True)
d =d.loc[:,['Sods','Fats','Carbs','Prots','Cals']]
d.dropna(inplace=True)
print(d.info())#all data types should be numbers, no objects
f, ax = plt.subplots(3,5, figsize=(8,5))

for ncol, col in enumerate(['Sods','Fats','Carbs','Prots','Cals']):
 data = d[col]
 order= np.arange(len(data))
 ax[0, ncol].hist(data)
 ax[1, ncol].boxplot(data)
 ax[0, ncol].set_title(col)
 ax[2, ncol].scatter(order, data)

plt.tight_layout()
plt.show()
display(d.describe())
```

```
print( d[d ['Sods']> 3000].index)
print( d[d ['Fats']> 200].index)
print( d[d ['Prots']> 50].index)
```

```
Index(['bbq boneless '], dtype='object', name='Food ')
Index(['breakfast burrito '], dtype='object', name='Food ')
Index(['Fajitas ', 'ma po tofu '], dtype='object', name='Food ')
```

```
d.drop(['bbq boneless ','breakfast burrito ','Fajitas ', 'ma po tofu '], inplace =True)
```

```
print("#"*80)
print("MODEL1:With intercept, with sods")
print("#"*80)
model1=smf.ols('Cals ~ Sods + Fats + Carbs +Prots',data=d)
results1=model1.fit()
print(results1.summary())

print("#"*80)
print("MODEL2: No intercept, with sods")
print("#"*80)
model2=smf.ols('Cals ~ 0 + Sods + Fats + Carbs +Prots',data=d)
results2=model2.fit()
print(results2.summary())

print("#"*80)
print("MODEL3: No intercept, No sods")
print("#"*80)
model3=smf.ols('Cals ~ 0 + Fats + Carbs +Prots',data=d)
results3=model3.fit()
print(results3.summary())
```

```
import scipy.stats as ss
residuals =results3.resid
fitted = results3.fittedvalues
order =np.arange(len(residuals))

f, ax = plt.subplots(2,2)

ax[0,0].hist(residuals)
ax[0,1].scatter(residuals, fitted)
ss.probplot(residuals, plot=ax[1,0])
ax[ 1,1].scatter(order, residuals )
plt.tight_layout()
plt.show()
```

```
residuals[(residuals < -20)| (residuals >20)].index
```

```
Index(['Tortilla Wrap', 'Rendang ', 'chinese food ', 'boiled egg',
       'korean ramen ', 'chips tomato salsa ',
       'alfredo sauce portion for 10 inch small crunchy thin crust pizza ',
       'chocolate milk shake ', 'bagel & cream cheese',
       'eggceptional skillet combo ', 'chicken pesto parm bowl ',
       'steakhouse sirloin salad', 'oatmeal apple pie', 'smores '],
      dtype='object', name='Food ')
```

+ Código + Texto

```
d.drop(['Tortilla Wrap', 'Rendang ', 'saudi banana bread mash',
       'chips tomato salsa ',
       'alfredo sauce portion for 10 inch small crunchy thin crust pizza ',
       'chicken pesto parm bowl ', 'smores '], inplace=True)
```

```
print(d.info())#all data types should be numbers, no objects
f, ax = plt.subplots(3,5, figsize=(8,5))

for ncol, col in enumerate(['Sods','Fats','Carbs','Prots','Cals']):
 data = d[col]
 order= np.arange(len(data))
 ax[0, ncol].hist(data)
 ax[1, ncol].boxplot(data)
 ax[0, ncol].set_title(col)
 ax[2, ncol].scatter(order, data)

plt.tight_layout()
plt.show()
display(d.describe())
```

```
print("#"*80)
print("MODEL1:With intercept, with sods")
print("#"*80)
model1=smf.ols('Cals ~ Sods + Fats + Carbs +Prots',data=d)
results1=model1.fit()
print(results1.summary())

print("#"*80)
print("MODEL2: No intercept, with sods")
print("#"*80)
model2=smf.ols('Cals ~ 0 + Sods + Fats + Carbs +Prots',data=d)
results2=model2.fit()
print(results2.summary())

print("#"*80)
print("MODEL3: No intercept, No sods")
print("#"*80)
model3=smf.ols('Cals ~ 0 + Fats + Carbs +Prots',data=d)
results3=model3.fit()
print(results3.summary())
```

```
import scipy.stats as ss
residuals =results3.resid
fitted = results3.fittedvalues
order =np.arange(len(residuals))

f, ax = plt.subplots(2,2)

ax[0,0].hist(residuals)
ax[0,1].scatter(residuals, fitted)
ss.probplot(residuals, plot=ax[1,0])
ax[ 1,1].scatter(order, residuals )
plt.tight_layout()
plt.show()
```

```
d.drop(['boiled egg'],inplace=True)
```

```
print(d.info())#all data types should be numbers, no objects
f, ax = plt.subplots(3,5, figsize=(8,5))

for ncol, col in enumerate(['Sods','Fats','Carbs','Prots','Cals']):
 data = d[col]
 order= np.arange(len(data))
 ax[0, ncol].hist(data)
 ax[1, ncol].boxplot(data)
 ax[0, ncol].set_title(col)
 ax[2, ncol].scatter(order, data)

plt.tight_layout()
plt.show()
display(d.describe())
print("#"*80)
print("MODEL1:With intercept, with sods")
print("#"*80)
model1=smf.ols('Cals ~ Sods + Fats + Carbs +Prots',data=d)
results1=model1.fit()
print(results1.summary())

print("#"*80)
print("MODEL2: No intercept, with sods")
print("#"*80)
model2=smf.ols('Cals ~ 0 + Sods + Fats + Carbs +Prots',data=d)
results2=model2.fit()
print(results2.summary())

print("#"*80)
print("MODEL3: No intercept, No sods")
print("#"*80)
model3=smf.ols('Cals ~ 0 + Fats + Carbs +Prots',data=d)
results3=model3.fit()
print(results3.summary())
import scipy.stats as ss
residuals =results3.resid
fitted = results3.fittedvalues
order =np.arange(len(residuals))

f, ax = plt.subplots(2,2)

ax[0,0].hist(residuals)
ax[0,1].scatter(residuals, fitted)
ss.probplot(residuals, plot=ax[1,0])
ax[ 1,1].scatter(order, residuals )
plt.tight_layout()
plt.show()
print (residuals.mean())
```

```
ss.probplot(residuals, plot=ax[1,0])
ax[ 1,1].scatter(order, residuals )
plt.tight_layout()
plt.show()
```

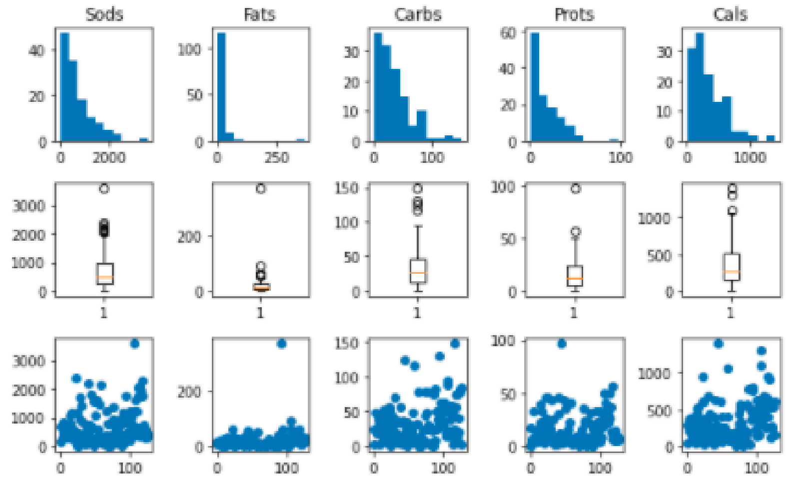
```
d.corr()
```

```
[ ] sns.pairplot(data=d)
    plt.show()
```

```
residuals.mean()
```

Appendix 2: Python Results

```
/usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.tes
  import pandas.util.testing as tm
<class 'pandas.core.frame.DataFrame'>
Index: 127 entries, Beef tacos to Hot dog
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Sods    127 non-null    float64
 1   Fats    127 non-null    float64
 2   Carbs   127 non-null    float64
 3   Prots   127 non-null    float64
 4   Cals    127 non-null    float64
dtypes: float64(5)
memory usage: 6.0+ KB
None
```

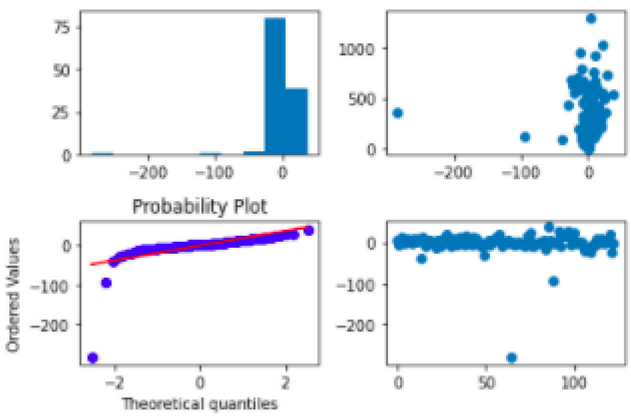


	Sods	Fats	Carbs	Prots	Cals
count	127.000000	127.000000	127.000000	127.000000	127.000000
mean	700.031496	19.203937	33.946457	16.640945	344.071654
std	633.896423	35.015825	29.329339	15.234898	269.552268
min	0.000000	0.000000	0.000000	0.300000	3.900000
25%	267.500000	6.050000	13.000000	5.550000	151.500000
50%	491.000000	12.000000	26.000000	12.000000	270.000000

```
Index(['bbq boneless '], dtype='object', name='Food ')
Index(['breakfast burrito '], dtype='object', name='Food ')
Index(['Fajitas ', 'ma po tofu '], dtype='object', name='Food ')
```

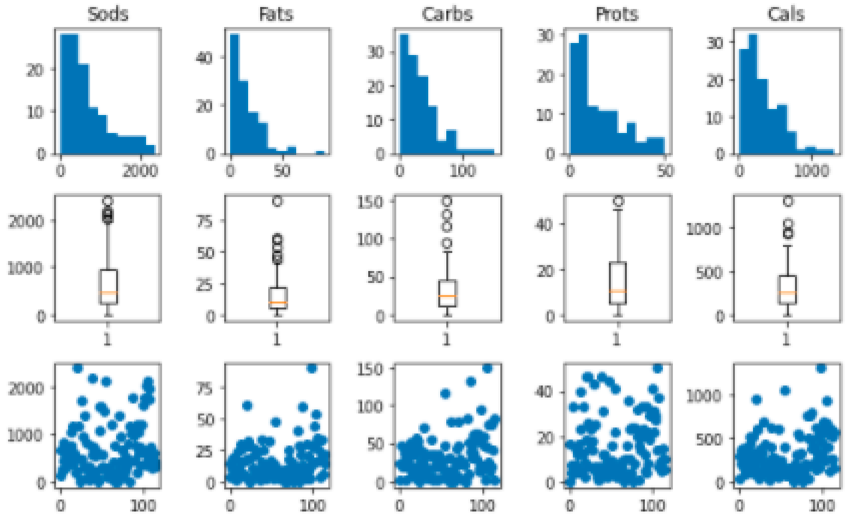
```
MODEL2: No intercept, with sods
################################################################################
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   Cals   R-squared (uncentered):                   0.995
Model:                            OLS   Adj. R-squared (uncentered):              0.995
Method:                 Least Squares   F-statistic:                              5702.
Date:                Fri, 03 Jun 2022   Prob (F-statistic):                   6.86e-135
Time:                        00:28:10   Log-Likelihood:                         -589.03
No. Observations:                 123   AIC:                                      1186.
Df Residuals:                     119   BIC:                                      1197.
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Sods           0.0073      0.007      1.094      0.276      -0.006       0.021
Fats           9.1973      0.288     31.974      0.000       8.628       9.767
Carbs          3.6500      0.103     35.378      0.000       3.446       3.854
Prots          3.9081      0.283     13.822      0.000       3.348       4.468
==============================================================================
Omnibus:                      214.930   Durbin-Watson:                   1.943
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            24367.763
Skew:                          -7.444   Prob(JB):                         0.00
Kurtosis:                      70.328   Cond. No.                         109.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
################################################################################
MODEL3: No intercept, No sods
################################################################################
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   Cals   R-squared (uncentered):                   0.995
Model:                            OLS   Adj. R-squared (uncentered):              0.995
Method:                 Least Squares   F-statistic:                              7589.
Date:                Fri, 03 Jun 2022   Prob (F-statistic):                   1.31e-136
Time:                        00:28:10   Log-Likelihood:                         -589.65
No. Observations:                 123   AIC:                                      1185.
Df Residuals:                     120   BIC:                                      1194.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Fats           9.2845      0.277     33.565      0.000       8.737       9.832
Carbs          3.6721      0.101     36.261      0.000       3.472       3.873
Prots          4.0679      0.242     16.793      0.000       3.588       4.548
==============================================================================
Omnibus:                      216.917   Durbin-Watson:                   1.971
```

```
-0.2673274944765053
```



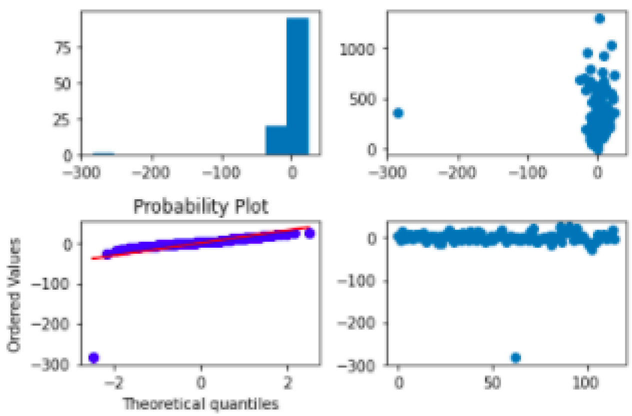
```
Index(['Tortilla Wrap', 'Rendang ', 'chinese food ', 'boiled egg',
       'korean ramen ', 'chips tomato salsa ',
       'alfredo sauce portion for 10 inch small crunchy thin crust pizza ',
       'chocolate milk shake ', 'bagel & cream cheese',
       'eggceptional skillet combo ', 'chicken pesto parm bowl ',
       'steakhouse sirloin salad', 'oatmeal apple pie', 'smores '],
      dtype='object', name='Food ')
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 116 entries, Beef tacos to Hot dog
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Sods    116 non-null    float64
 1   Fats    116 non-null    float64
 2   Carbs   116 non-null    float64
 3   Prots   116 non-null    float64
 4   Cals    116 non-null    float64
dtypes: float64(5)
memory usage: 5.4+ KB
None
```



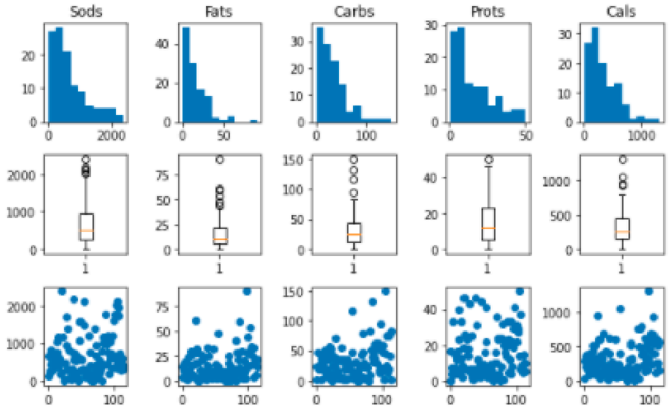
	Sods	Fats	Carbs	Prots	Cals
count	116.000000	116.000000	116.000000	116.000000	116.000000
mean	663.568966	15.223276	31.795690	15.462931	321.371552
std	564.760739	14.477511	27.796682	12.703745	241.683711
min	0.000000	0.000000	0.000000	0.300000	3.900000
25%	260.250000	6.000000	11.750000	5.475000	146.000000
50%	488.000000	11.000000	25.000000	11.000000	260.000000
75%	942.500000	21.000000	44.500000	23.000000	444.000000
max	2400.000000	90.000000	149.000000	50.000000	1310.000000

```
MODEL2: No intercept, with sods
###############################################################################
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   Cals   R-squared (uncentered):                   0.995
Model:                            OLS   Adj. R-squared (uncentered):              0.995
Method:                 Least Squares   F-statistic:                              5738.
Date:                Fri, 03 Jun 2022   Prob (F-statistic):                   1.54e-128
Time:                        00:28:15   Log-Likelihood:                         -551.04
No. Observations:                 116   AIC:                                      1110.
Df Residuals:                     112   BIC:                                      1121.
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Sods           0.0048      0.007      0.715      0.476      -0.009       0.018
Fats           9.1888      0.281     32.650      0.000       8.631       9.746
Carbs          3.6733      0.104     35.251      0.000       3.467       3.880
Prots          4.0148      0.285     14.098      0.000       3.451       4.579
==============================================================================
Omnibus:                      227.014   Durbin-Watson:                   1.946
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            37584.325
Skew:                          -8.824   Prob(JB):                         0.00
Kurtosis:                      89.398   Cond. No.                         109.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
###############################################################################
MODEL3: No intercept, No sods
###############################################################################
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   Cals   R-squared (uncentered):                   0.995
Model:                            OLS   Adj. R-squared (uncentered):              0.995
Method:                 Least Squares   F-statistic:                              7684.
Date:                Fri, 03 Jun 2022   Prob (F-statistic):                   2.08e-130
Time:                        00:28:15   Log-Likelihood:                         -551.30
No. Observations:                 116   AIC:                                      1109.
Df Residuals:                     113   BIC:                                      1117.
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Fats           9.2481      0.268     34.462      0.000       8.716       9.780
Carbs          3.6866      0.102     36.038      0.000       3.484       3.889
Prots          4.1235      0.240     17.163      0.000       3.648       4.599
==============================================================================
Omnibus:                      228.646   Durbin-Watson:                   1.960
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            38955.049
Skew:                          -8.939   Prob(JB):                         0.00
Kurtosis:                      90.977   Cond. No.                         5.99
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```




```
Index(['boiled egg'], dtype='object', name='Food ')
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 115 entries, Beef tacos to Hot dog
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Sods    115 non-null    float64
 1   Fats    115 non-null    float64
 2   Carbs   115 non-null    float64
 3   Prots   115 non-null    float64
 4   Cals    115 non-null    float64
dtypes: float64(5)
memory usage: 5.4+ KB
None
```

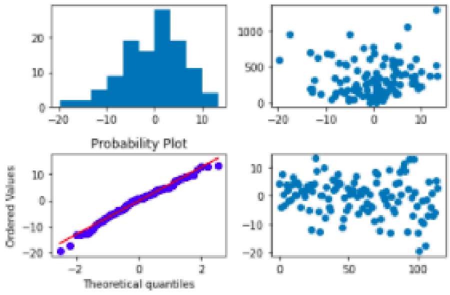


	Sods	Fats	Carbs	Prots	Cals
count	115.000000	115.000000	115.000000	115.000000	115.000000
mean	668.800000	15.309565	31.393913	15.542609	323.487826
std	564.402786	14.510879	27.577957	12.730197	241.659505
min	0.000000	0.000000	0.000000	0.300000	3.900000
25%	267.500000	6.050000	11.500000	5.450000	151.500000
50%	491.000000	11.000000	25.000000	12.000000	260.000000
75%	945.000000	21.000000	44.000000	23.000000	446.000000
max	2400.000000	90.000000	149.000000	50.000000	1310.000000

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   Cals   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.999
Method:                 Least Squares   F-statistic:                 3.829e+04
Date:                Fri, 03 Jun 2022   Prob (F-statistic):          6.68e-172
Time:                        00:28:19   Log-Likelihood:                -377.47
No. Observations:                 115   AIC:                             764.9
Df Residuals:                     110   BIC:                             778.7
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.7164      1.117     -0.641      0.523      -2.930       1.497
Sods          -0.0012      0.002     -0.735      0.464      -0.004       0.002
Fats           8.8791      0.066    135.158      0.000       8.749       9.009
Carbs          3.9694      0.027    146.335      0.000       3.916       4.023
Prots          4.1455      0.070     59.627      0.000       4.008       4.283
==============================================================================
Omnibus:                        6.201   Durbin-Watson:                   1.657
Prob(Omnibus):                  0.045   Jarque-Bera (JB):                5.687
Skew:                          -0.516   Prob(JB):                       0.0582
Kurtosis:                       3.347   Cond. No.                     1.59e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.59e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
################################################################################
MODEL2: No intercept, with sods
################################################################################
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   Cals   R-squared (uncentered):                   1.000
Model:                            OLS   Adj. R-squared (uncentered):              1.000
Method:                 Least Squares   F-statistic:                          1.081e+05
Date:                Fri, 03 Jun 2022   Prob (F-statistic):                   2.91e-198
Time:                        00:28:20   Log-Likelihood:                         -377.69
No. Observations:                 115   AIC:                                      763.4
Df Residuals:                     111   BIC:                                      774.4
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Sods          -0.0013      0.002     -0.860      0.392      -0.004       0.002
Fats           8.8822      0.065    135.951      0.000       8.753       9.012
Carbs          3.9627      0.025    159.046      0.000       3.913       4.012
Prots          4.1314      0.066     62.792      0.000       4.001       4.262
```

	Sods	Fats	Carbs	Prots	Cals
Sods	1.000000	0.635082	0.385197	0.659620	0.654417
Fats	0.635082	1.000000	0.551058	0.621686	0.916818
Carbs	0.385197	0.551058	1.000000	0.276773	0.806184
Prots	0.659620	0.621686	0.276773	1.000000	0.673417
Cals	0.654417	0.916818	0.806184	0.673417	1.000000

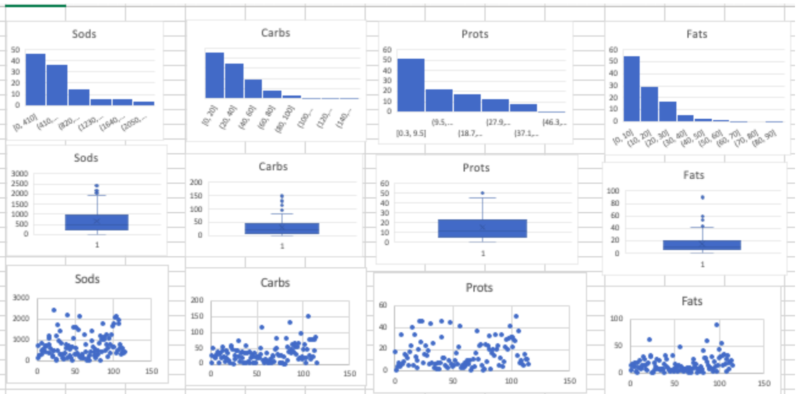
```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
################################################################################
MODEL3: No intercept, No sods
################################################################################
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                   Cals   R-squared (uncentered):                   1.000
Model:                            OLS   Adj. R-squared (uncentered):              1.000
Method:                 Least Squares   F-statistic:                          1.445e+05
Date:                Fri, 03 Jun 2022   Prob (F-statistic):                   1.02e-200
Time:                        00:28:20   Log-Likelihood:                         -378.07
No. Observations:                 115   AIC:                                      762.1
Df Residuals:                     112   BIC:                                      770.4
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Fats           8.8664      0.063    141.641      0.000       8.742       8.990
Carbs          3.9585      0.024    162.184      0.000       3.910       4.007
Prots          4.1012      0.056     73.866      0.000       3.991       4.211
==============================================================================
Omnibus:                        3.843   Durbin-Watson:                   1.678
Prob(Omnibus):                  0.146   Jarque-Bera (JB):                3.370
Skew:                          -0.410   Prob(JB):                        0.185
Kurtosis:                       3.173   Cond. No.                         5.96
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```



Appendix: Excel and Minitab Results

Descriptive Statistics

	Sods	Carbs	Prots	Fats	Cals
Mean	668.8	31.39391	15.54261	15.30957	323.4878
Standard Err	52.63083	2.571658	1.187097	1.353146	22.53487
Median	491	25	12	11	260
Mode	295	19	23	12	155
Standard De	564.4028	27.57796	12.7302	14.51088	241.6595
Sample Varia	318550.5	760.5437	162.0579	210.5656	58399.32
Kurtosis	0.746775	3.55268	-0.09182	6.398555	2.098115
Skewness	1.179822	1.582395	0.927414	2.080466	1.305227
Range	2400	149	49.7	90	1306.1
Maximum	2400	149	50	90	1310
Minimum	0	0	0.3	0	3.9
Sum	76912	3610.3	1787.4	1760.6	37201.1
Count	115	115	115	115	115
Geometric M	#¡NUM!	#¡NUM!	10.11389	#¡NUM!	232.8068
Harmonic M	#¡NUM!	#¡NUM!	4.705182	#¡NUM!	109.87
AAD	438.953	20.57552	10.52243	10.53372	189.1228
MAD	269	16	8	7	138
IQR	677.5	32.5	17.55	14.95	294.5



Regression Analysis

OVERALL FIT

Multiple R	0.9996411	AIC	438.5903
R Square	0.9992823	AICc	439.36808
Adjusted R Sq	0.9992562	SBC	452.31496
Standard Err	6.5907159		
Observations	115		

ANOVA — Alpha 0.5

	df	SS	MS	F	p-value	sig
Regression	4	6652744	1663186	38289.142	6.68E-172	yes
Residual	110	4778.1289	43.437536			
Total	114	6657522.1				

	coeff	std err	t stat	p-value	lower	upper	vif
Intercept	-0.716439	1.1169633	-0.641417	0.5225861	-1.472318	0.0394397	
Sods	-0.001164	0.0015844	-0.734677	0.4640993	-0.002236	-9.18E-05	2.0986124
Carbs	3.9694405	0.0271258	146.33451	7.04E-128	3.9510837	3.9877972	1.4686859
Prots	4.1455068	0.0695236	59.62736	1.369E-85	4.0984584	4.1925553	2.0557675
Fats	8.8790562	0.0656938	135.1583	4.19E-124	8.8345995	8.9235129	2.3849282

Regression Analysis

OVERALL FIT

Multiple R	0.9998717	AIC	437.019615
R Square	0.99974341	AICc	437.570073
Adjusted R S	0.99973417	BSC	447.999343
Standard Err	6.57321879		
Observations	115		

ANOVA — Alpha 0.05

	df	SS	MS	F	p-value	sig
Regression	4	18686829.1	4671707.26	108123.338	2.913E-198	yes
Residual	111	4795.99979	43.2072053			
Total	115	18691625.1				

	coeff	std err	t stat	p-value	lower	upper	vif
Sods	-0.0013381	0.00155679	-0.8595528	0.39188818	-0.004423	0.00174674	2.09861237
Carbs	3.96266051	0.02491523	159.045733	8.321E-133	3.91328933	4.0120317	1.46868592
Prots	4.13143399	0.0657957	62.7918507	1.4561E-88	4.00105541	4.26181257	2.05576753
Fats	8.88222487	0.06533384	135.951373	2.787E-125	8.75276151	9.01168823	2.38492823

Regression Analysis			
OVERALL FIT			
Multiple R	0.99987084	AIC	435.782534
R Square	0.99974171	AICc	436.14617
Adjusted R S	0.99973479	BSC	444.01733
Standard Err	6.56555039		
Observations	115		

ANOVA				Alpha	0.05	
	df	*SS*	*MS*	*F*	*p-value*	*sig*
Regression	3	18686797.1	6228932.38	144501.162	1.018E-200	yes
Residual	112	4827.92261	43.1064519			
Total	115	18691625.1				

	coeff	*std err*	*t stat*	*p-value*	*lower*	*upper*	*vif*
Carbs	3.95847956	0.02440731	162.184198	1.021E-134	3.91011961	4.00683951	1.45079851
Prots	4.10117584	0.05552184	73.8659909	7.5325E-97	3.99116643	4.21118526	1.64666763
Fats	8.86635363	0.06259723	141.641318	3.674E-128	8.74232525	8.99038201	2.1836142

DATA NUTRIX.CSV

Descriptive Statistics: Sods, Carbs, Prots, Fats, Cals

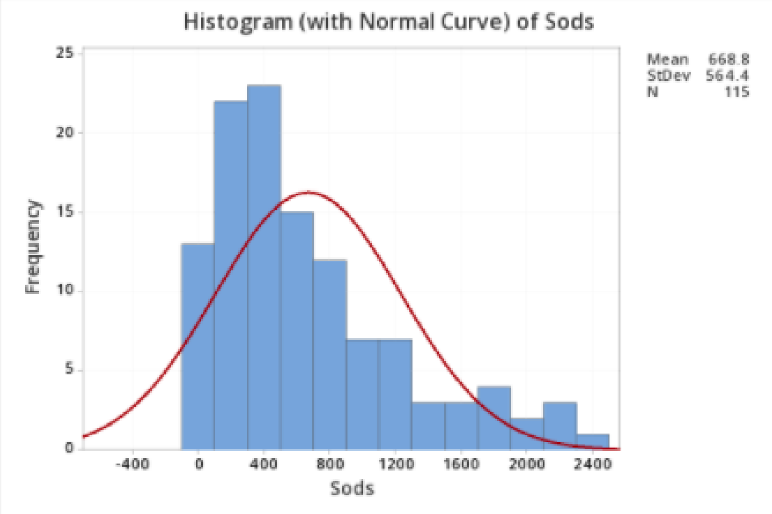
Statistics

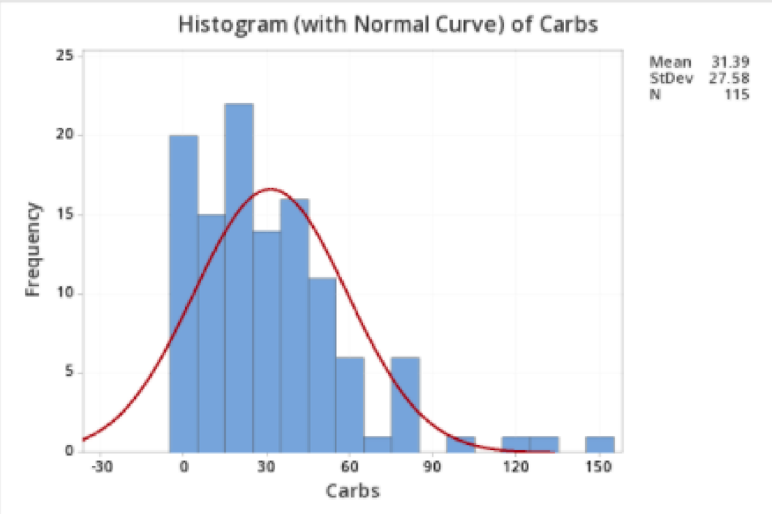
Variable	N	N*	Mean	SE Mean	StDev	Variance	Sum	Sum of
Sods	115	0	668.8	52.6	564.4	318550.5	76912.0	87
Carbs	115	0	31.39	2.57	27.58	760.54	3610.30	2
Prots	115	0	15.54	1.19	12.73	162.06	1787.40	
Fats	115	0	15.31	1.35	14.51	210.57	1760.60	
Cals	115	0	323.5	22.5	241.7	58399.3	37201.1	18

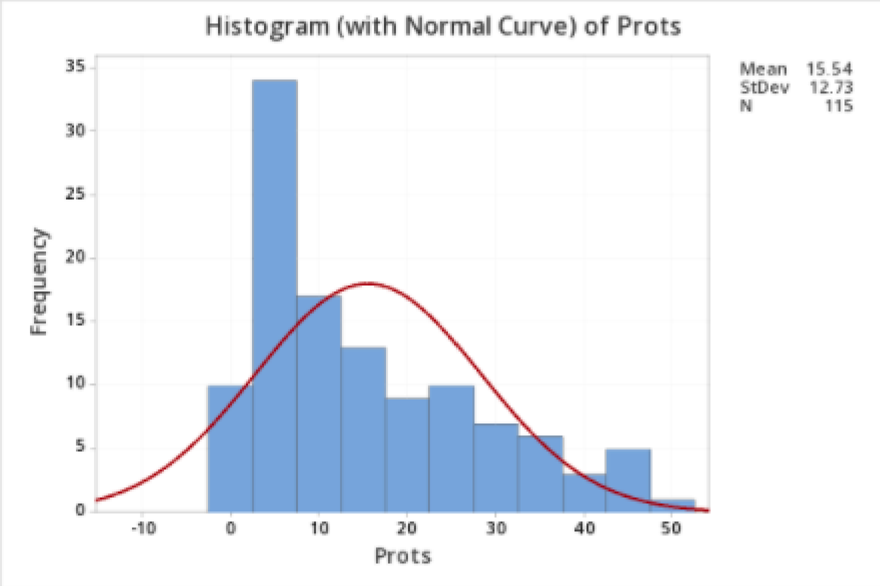
Variable	Median	Q3	Maximum	Range	IQR	Mode	N
Sods	491.0	950.0	2400.0	2400.0	685.0	226, 295, 360, 590	
Carbs	25.00	44.00	149.00	149.00	33.00	18, 19, 22, 35	
Prots	12.00	23.00	50.00	49.70	17.60	23	
Fats	11.00	21.00	90.00	90.00	15.00	12	
Cals	260.0	450.0	1310.0	1306.1	302.0	155, 168, 170, 530	

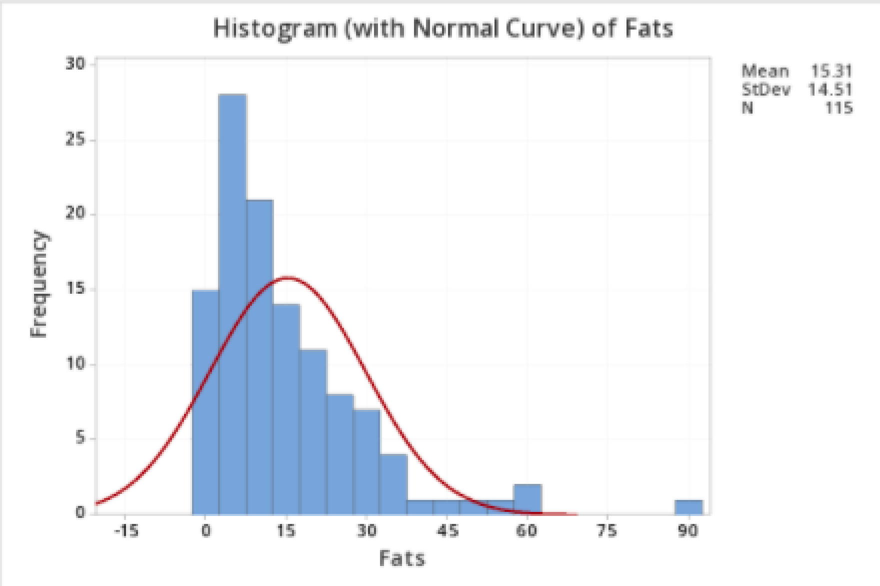
Variable	Kurtosis
Sods	0.75
Carbs	3.55
Prots	-0.09
Fats	6.40
Cals	2.10

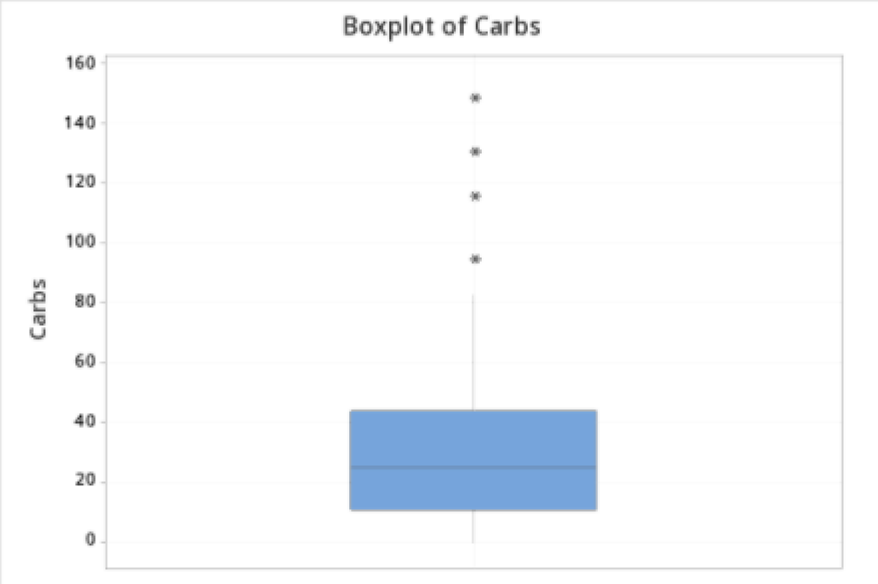
The data contain at least five mode values. Only the smallest four are shown.

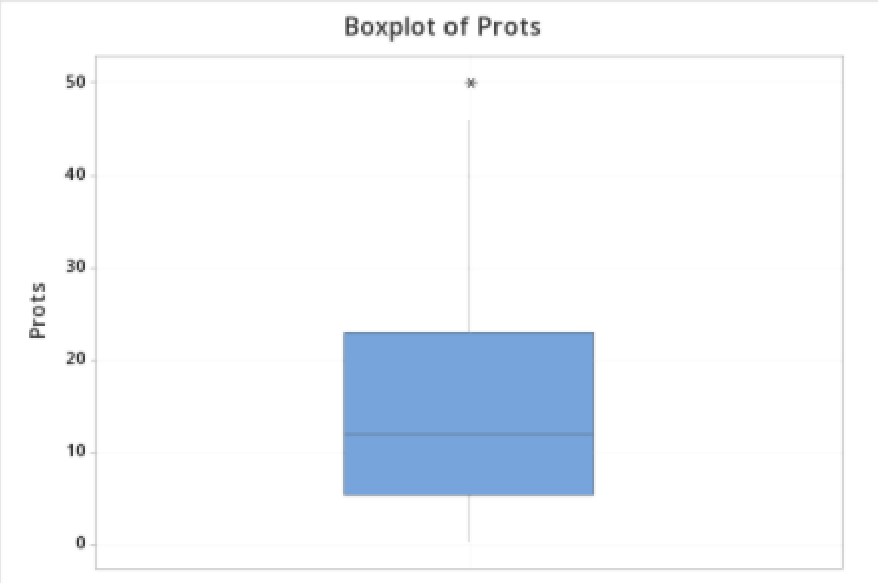

Histogram (with Normal Curve) of Sods
Mean 668.8
StDev 564.4
N 115
Frequency
Sods
0
5
10
15
20
25
-400
0
400
800
1200
1600
2000
2400

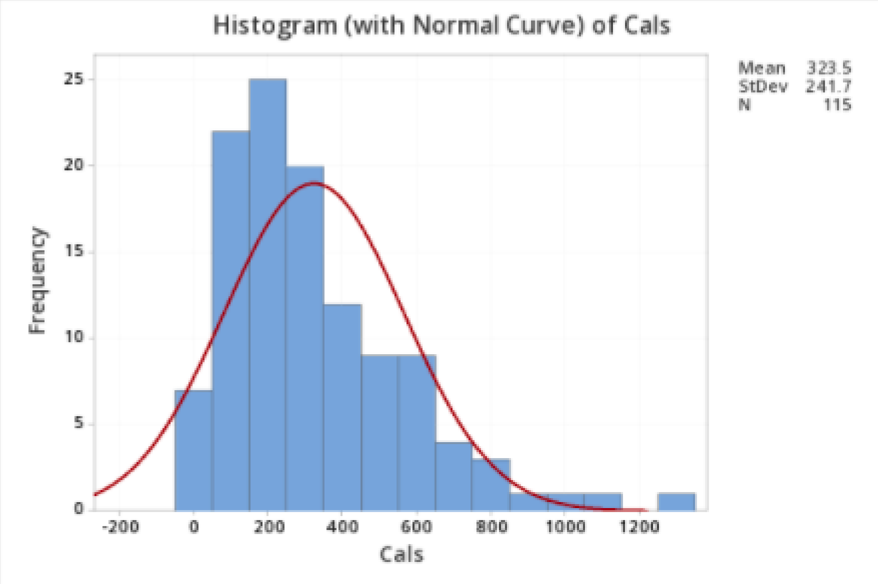

Histogram (with Normal Curve) of Carbs
Mean 31.39
StDev 27.58
N 115
Frequency
Carbs
0
5
10
15
20
25
-30
0
30
60
90
120
150

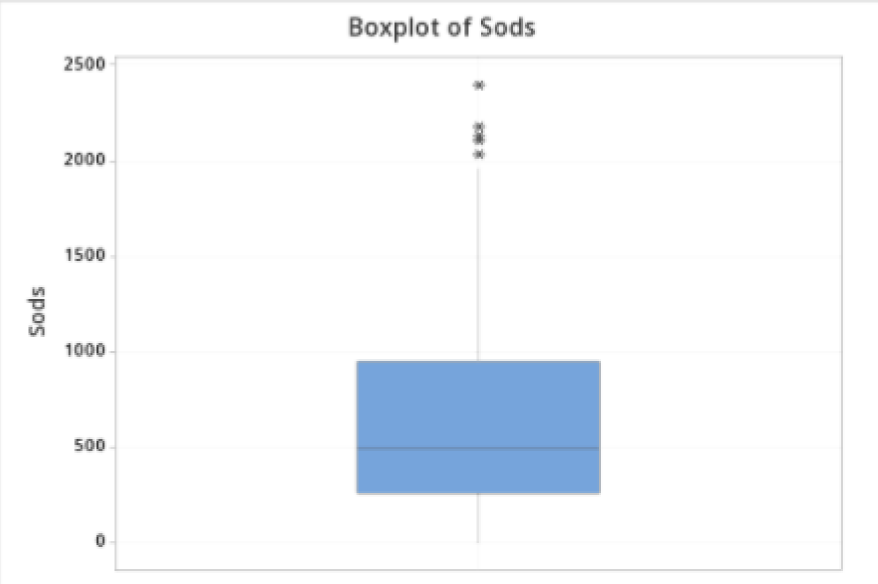

Histogram (with Normal Curve) of Prots
Mean 15.54
StDev 12.73
N 115
Frequency
Prots
35
30
25
20
15
10
5
0
-10
0
10
20
30
40
50

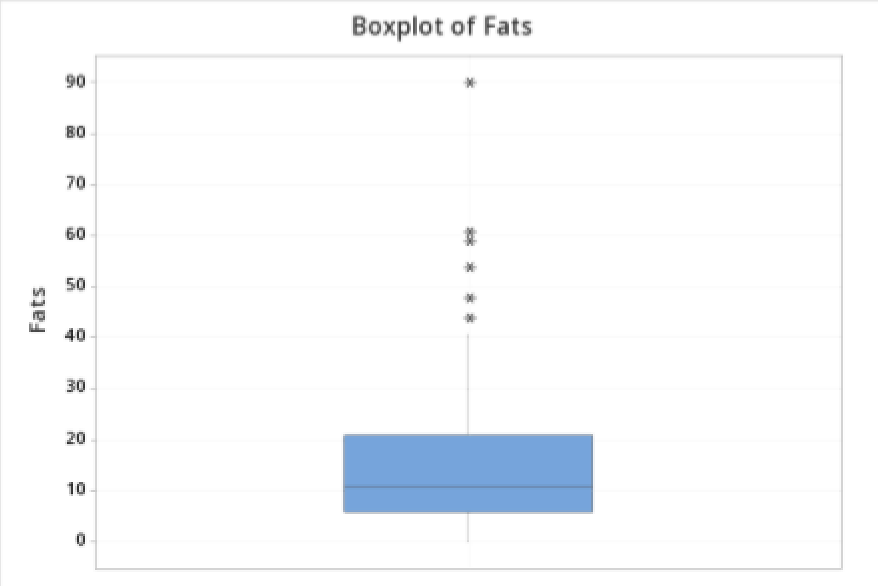

Histogram (with Normal Curve) of Fats
Mean 15.31
StDev 14.51
N 115
Frequency
Fats
30
25
20
15
10
5
0
-15
0
15
30
45
60
75
90

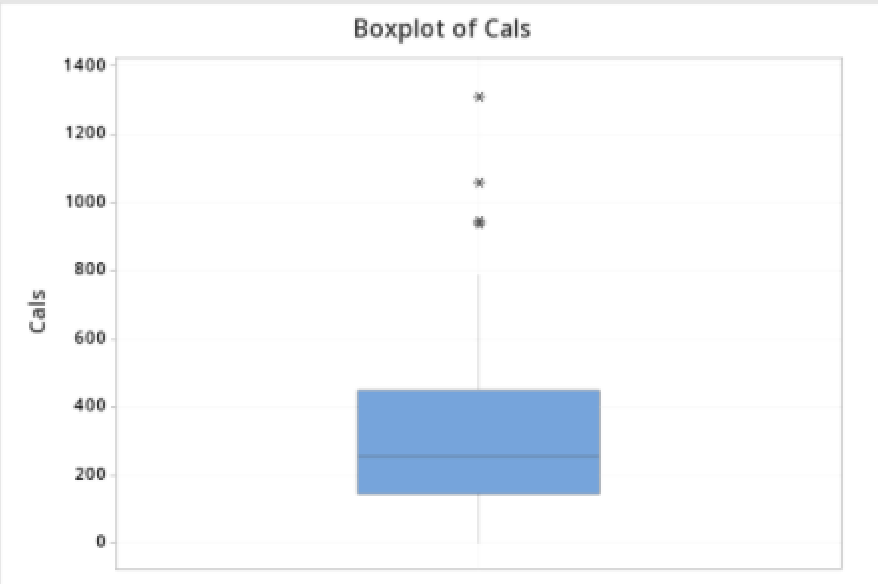

Boxplot of Carbs
160
140
120
100
80
60
40
20
0
Carbs

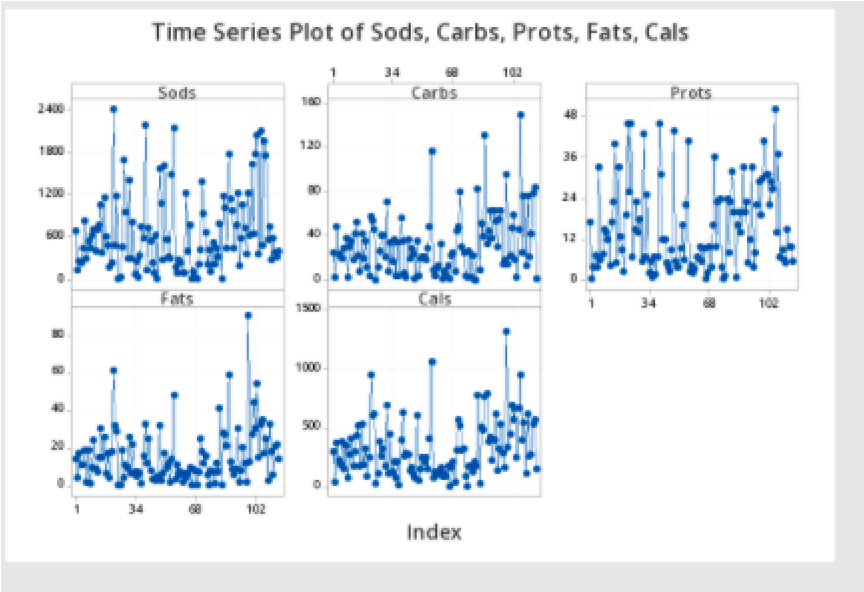

Boxplot of Prots
50
40
30
20
10
0
Prots


Histogram (with Normal Curve) of Cals
Mean 323.5
StDev 241.7
N 115
Frequency
Cals
25
20
15
10
5
0
-200
0
200
400
600
800
1000
1200


Boxplot of Sods
Sods
2500
2000
1500
1000
500
0


Boxplot of Fats
Fats
90
80
70
60
50
40
30
20
10
0


Boxplot of Cals
Cals
1400
1200
1000
800
600
400
200
0


Time Series Plot of Sods, Carbs, Prots, Fats, Cals
Sods
Carbs
Prots
Fats
Cals
Index

DATA NUTRIX.CSV

Regression Analysis: Cals versus Sods, Carbs, Prots, Fats

Regression Equation

Cals = -0.72 - 0.00116 Sods + 3.9694 Carbs + 4.1455 Prots + 8.8791 Fats

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.72	1.12	-0.64	0.523	
Sods	-0.00116	0.00158	-0.73	0.464	2.10
Carbs	3.9694	0.0271	146.33	0.000	1.47
Prots	4.1455	0.0695	59.63	0.000	2.06
Fats	8.8791	0.0657	135.16	0.000	2.38

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
6.59072	99.93%	99.93%	99.92%

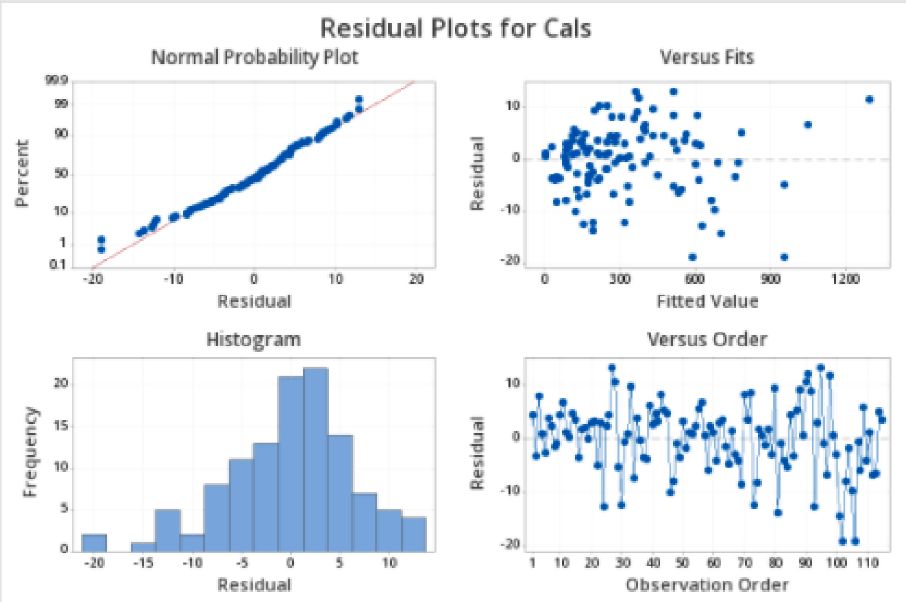
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	6652744	1663186	38289.14	0.000
Sods	1	23	23	0.54	0.464
Carbs	1	930162	930162	21413.79	0.000
Prots	1	154439	154439	3555.42	0.000
Fats	1	793507	793507	18267.76	0.000
Error	110	4778	43		
Lack-of-Fit	109	4778	44	*	*
Pure Error	1	0	0		
Total	114	6657522			

Fits and Diagnostics for Unusual Observations

Obs	Cals	Fit	Resid	Std Resid		
81	180.00	193.73	-13.73	-2.16	R	
86	760.00	763.28	-3.28	-0.55		X
87	790.00	784.90	5.10	0.83		X
95	530.00	516.92	13.08	2.01	R	
98	1310.00	1298.45	11.55	2.16	R	X
101	690.00	704.28	-14.28	-2.25	R	
102	570.00	588.98	-18.98	-2.96	R	
103	660.00	668.05	-8.05	-1.33		X
106	940.00	958.98	-18.98	-3.24	R	X

R Large residual
X Unusual X



Appendix: Cleaning data table

Food	Sods	Carbs	Prots	Fats	Cals
Beef tacos	682	24	17	14	293
Guacamole	127	2.6	0.6	4.1	45
French Fries	246	48	4	17	365
French toast	237	23	7.3	11	219

Avocado toast	439	20	3.8	11	189
Tuna salad	824	19	33	19	383
Sandwich	295	28	6.2	1.9	155
Sushi	537	38	7.8	19	349
Ceviche	436	2.6	15	1.3	82
Chicket enchilada	622	30	14	9.9	268
pasta salad	708	35	12	24	407
Margherita pizza	398	18	4.5	9	170
Vegan Burger	711	42	17	7	298
Tuna Wrap	760	52	23	15	438
Meat cake	1056	22	40	30	521
Greek Salad	370	7.4	5.1	15	179
Chicken tacos	1160	41	33	26	530
grilled chicken tacos	601	19	13	6.2	185
sopes	491	35	9	17	326
pancakes	176	11	2.6	3.9	91
chicken salad	226	3.3	19	18	254
Chilaquiles	2400	57	46	61	950
Lomes Saltado	485	53	26	32	604
Chicken teriyaki bowl	1178	46	46	29	616
Tuna Sashimi	13	0	6.9	0.1	31
Nigiri	26	12	15	0.4	117

Spicy tuna roll	463	26	23	19	379
tortilla soup	1700	40	14	4	260
Indian food	950	40	18	11	324
Mole	306	21	5.5	9.6	178
Fish and chips	1403	71	43	26	688
chicken tenders	296	7.1	7	6.2	112
Paella	812	34	25	22	442
Ratatouille	81	16	2.1	7.1	127
arepas	265	36	5.4	5.3	215
Falafel	30	3.7	1.1	7.3	83
Bosrch	329	34	7	6.5	217
kimchi	747	3.6	1.7	0.8	23
scones	582	56	7	16	398
Hot pot	2180	35	46	33	622
tandoori chicken	132	6.1	31	12	263
Bratwurst	719	2.4	12	25	283
Nasi goreng	549	36	12	8.4	274
Dim sum	225	19	5.2	3.7	134
Masala dosa	94	29	3.9	3.7	168
Som tam	629	23	2.6	2.3	111
tofu	3.6	1.1	9.1	4.8	76
lasagne	1576	35	44	32	602
miso soup	1071	3.5	5.8	3	59
spring rolls	270	19	3.7	6.2	148
gazpacho	1599	20	4.1	17	243
tacos	571	21	9.4	9.9	210

fish tacos	326	18	16	12	244
sandwich	295	28	6.2	1.9	155
ham sandwich	1479	48	22	14	409
chinese food	2134	116	41	48	1060
oat pancakes	181	9.4	2.5	3.8	81
keto pancakes	93	3.8	4.3	11	124
potato panckakes	283	10	2.2	5.5	99
pecan pancakes	226	13	3.5	7.4	132
yogurt blueberry	90	32	6.8	2	168
fried egg	95	0.4	6.3	6.8	90
fish cake soup	1211	7.9	9.7	4.5	113
sogogi guk	406	3	5.6	6.6	91
shourbat shufan	769	13	9.5	9.6	170
arabic coffee	5.4	0.3	0.3	0.1	3.9
kleicha date cookies	112	22	2.2	8.5	168
jarish	29	25	10	8	207

jute potherb cooked boiled drained no salt	11	7.3	3.7	0.2	37
mandi	213	44	16	7.3	313
kabsa	416	48	36	25	564
korean ramen	1388	79	10	18	521
korean spicy chicken	940	30	23	12	307
korean braised chiken with vegetables	670	24	24	16	327
korean bbq pork ribs	205	3.8	4.1	6.1	87
chamoy sauce	419	1.7	0.3	0	7.2
frosted sugar cookie	110	26	1.2	8.2	179
barbacoa	530	2	24	7	170
black beans	210	22	8	1.5	130
carnitas	450	0	23	12	210
chiken	310	0	32	7	180

chips con queso	790	82	20	41	770
mandarin orange	0	9	1	0	35
big fish sandwich	1180	51	16	28	510
nuggets 10 pc	1010	39	20	27	480
chocolate milk shake	430	131	14	21	760
egg croissan	1780	32	33	59	790
6 inch B.L.T	1130	44	20	13	380
6 INCH APPLEWOOD PULLED PORK	980	62	23	9	420
gluten free toast	440	37	5	6	230
bagel & cream cheese	760	63	12	9	390
shakshuka	1220	53	33	30	620
cinnamon raisin toast	190	30	4	2	140

gluten free pancakes	590	54	8	8	320
eggceptional skillet combo	1050	63	22	20	530
Chicken caesar salad	730	14	29	12	280
thai chicken crunch salad	360	18	19	2	160
fettuccine alfredo	1210	95	30	90	1310
teriyaki chicken	630	14	41	13	340
buffalo cicken bowl	1620	22	31	27	450
fish taco bowl	650	47	31	44	690
thai bowl	1780	59	22	30	570
steakhouse sirloin salad	2040	18	29	54	660
chicken kebab	360	2	27	15	250
beef with broccoli	2110	46	50	33	670

banana spring rolls	480	149	14	35	940
edamame	1960	25	37	17	400
4 veggie spring rolls	1750	75	7	25	540
wonton soup	570	13	9	3	120
carrot cake	740	75	6	33	610
pesto pasta salad	280	20	5	18	260
lentil soup	590	41	15	6	270
churro	380	78	10	21	530
oatmeal apple pie	320	83	10	22	570
Hot dog	409	1.3	5.6	14	155

Appendix: Full data binnacle link

https://docs.google.com/spreadsheets/d/1ks85TEmR69NWRTbxmilvZxmgziuJPbJlQT7Q7Ff2DKI/edit?usp=sharing