

WU #17 - Outliers 2

Math 158 - Jo Hardin

Thursday 3/31/2022

Name: _____

Names of people you worked with: _____

Consider data taken on 301 fast food items with `calories` modeled on `sodium`, `total_carbs`, and `total_fat`. Below are some of the metrics discussed in class to assess the influence of particular points on the model. (See class notes for R code to calculate the metrics.) The second table reports columns of `dfbetas()`.

What would you do with the observations?

```
fastfood <- fastfood %>% drop_na()
ff_lm <- lm(calories ~ sodium + total_carb + total_fat, data=fastfood)
```

```
## # A tibble: 301 x 7
##   item      .resid  .hat  .cooksd  .std.resid  rstudent  dffits
##   <chr>      <dbl> <dbl> <dbl>      <dbl>      <dbl> <dbl>
## 1 20 piece But~ 213.  0.205  0.754      3.42      3.49  1.77
## 2 10 piece Swe~  37.2  0.0767  0.00638    0.554     0.554  0.160
## 3 Super Sonic ~  68.9  0.0718  0.0203     1.02     1.02  0.285
## 4 40 piece Chi~  62.4  0.0680  0.0156     0.926     0.926  0.250
## 5 12 piece But~ 123.  0.0675  0.0600     1.82     1.83  0.492
## 6 Super Sonic ~  62.2  0.0645  0.0146     0.921     0.921  0.242
## 7 Footlong Cor~  48.5  0.0640  0.00880    0.718     0.717  0.187
## 8 10 piece But~  83.8  0.0471  0.0187     1.23     1.23  0.274
## 9 Super Sonic ~  74.1  0.0427  0.0131     1.08     1.08  0.229
## 10 Footlong Swe~  75.2  0.0410  0.0129     1.10     1.10  0.227
## # ... with 291 more rows
```

```
## # A tibble: 301 x 5
##   item      `(Intercept)`  sodium  total_carb  total_fat
##   <chr>      <dbl> <dbl>      <dbl>      <dbl>
## 1 20 piece Buttermil~ -0.907  1.04    -0.765    0.438
## 2 10 piece Sweet N' ~ -0.112  0.0833  0.0386   -0.0370
## 3 Super Sonic Bacon ~ -0.0175 -0.126  -0.0173   0.274
## 4 40 piece Chicken M~ -0.130  -0.0224  0.0215   0.184
## 5 12 piece Buttermil~ -0.184  0.290   -0.259   0.125
## 6 Super Sonic Double~ -0.0110 -0.117  -0.00576  0.233
## 7 Footlong Corned Be~ -0.0623  0.169  -0.0463  -0.119
## 8 10 piece Buttermil~ -0.0750  0.185  -0.168   0.0382
## 9 Super Sonic Double~ -0.00604 -0.0972 -0.0170  0.216
## 10 Footlong Sweet Oni~ -0.0554 -0.0104  0.175   -0.106
## # ... with 291 more rows
```

Solution:

None of the observations are hugely influential. The most outlying of the observations impacts the beta values and their own predictions both just a bit. They don't seem to impact the other predictions much.

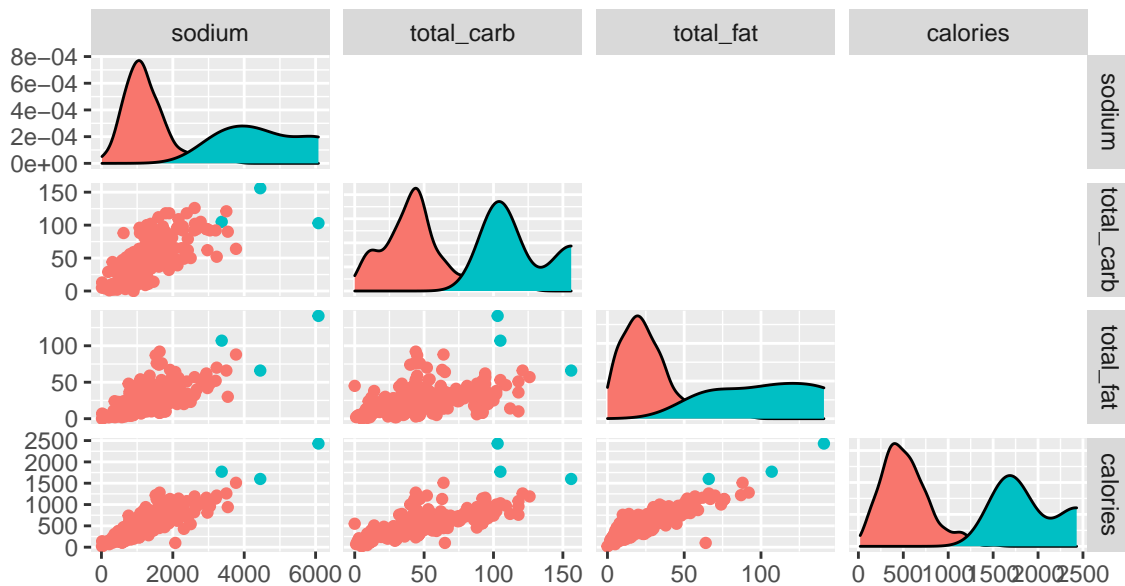
I'd be hesitant to take out the variables without a good external reason. Although, in looking at the dataset, I might subset the variables to be only those which can be thought of as one meal for one person (i.e., not 40 Piece Chicken McNuggets). But that decision would certainly be very subjective!

Outliers are:

20 piece Buttermilk Crispy Chicken Tenders

10 piece Sweet N' Spicy Honey BBQ Glazed Tenders

40 piece Chicken McNuggets



Note that if other variables are added into the model, the particular observations become less influential.

```
ff2_lm <- lm(calories ~ sodium + total_carb + total_fat +  
             protein + cholesterol + protein, data=fastfood)
```

```
## # A tibble: 301 x 7
```

```
##   item      .resid  .hat .cooksd .std.resid rstudent  dffits  
##   <chr>    <dbl> <dbl> <dbl>    <dbl>    <dbl>  <dbl>  
## 1 20 piece Bu~ 28.5 0.324 3.38e-2  0.651    0.650  0.450  
## 2 10 piece Sw~ -15.5 0.125 2.30e-3 -0.311   -0.310 -0.117  
## 3 12 piece Bu~ 16.3 0.111 2.18e-3  0.324    0.324  0.114  
## 4 Premium Bac~  3.34 0.0948 7.56e-5  0.0658   0.0657  0.0213  
## 5 Footlong Ve~ 39.6 0.0834 9.14e-3  0.776    0.776  0.234  
## 6 Super Sonic~ 18.3 0.0813 1.89e-3  0.358    0.358  0.106  
## 7 Super Sonic~ 14.6 0.0737 1.07e-3  0.284    0.284  0.0800  
## 8 40 piece Ch~ 16.3 0.0714 1.29e-3  0.317    0.317  0.0878  
## 9 10 piece Bu~  6.18 0.0713 1.85e-4  0.120    0.120  0.0333  
## 10 Footlong Co~ -11.9 0.0707 6.80e-4 -0.232   -0.231 -0.0638  
## # ... with 291 more rows
```

```
## # A tibble: 301 x 7
```

```
##   item      `(Intercept)`  sodium total_carb total_fat  protein  
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>    <dbl>  
## 1 20 piece~      -0.196    0.0327   -0.218    0.121    0.256
```

```
## 2 10 piece~      0.0721 -0.0448 -0.0577  0.0686  0.0551
## 3 12 piece~     -0.0363  0.00718 -0.0635  0.0333  0.0676
## 4 Premium ~      0.00559 -0.00122 -0.0161  0.00865  0.0188
## 5 Footlong~     -0.0280 -0.0925  0.0203  0.0637  0.179
## 6 Super So~     -0.0118 -0.0425  0.0116  0.0343 -0.0155
## 7 Super So~     -0.00772 -0.0339  0.0118  0.0250 -0.0130
## 8 40 piece~     -0.0469 -0.0163  0.00780  0.0335  0.00805
## 9 10 piece~     -0.00820  0.00532 -0.0207  0.00789  0.0186
## 10 Footlong~    0.0232 -0.0375  0.00734  0.0358 -0.00125
## # ... with 291 more rows, and 1 more variable: cholesterol <dbl>
```