# NHANES_FPED_EDA_Cont

April 10, 2021

### Objectives

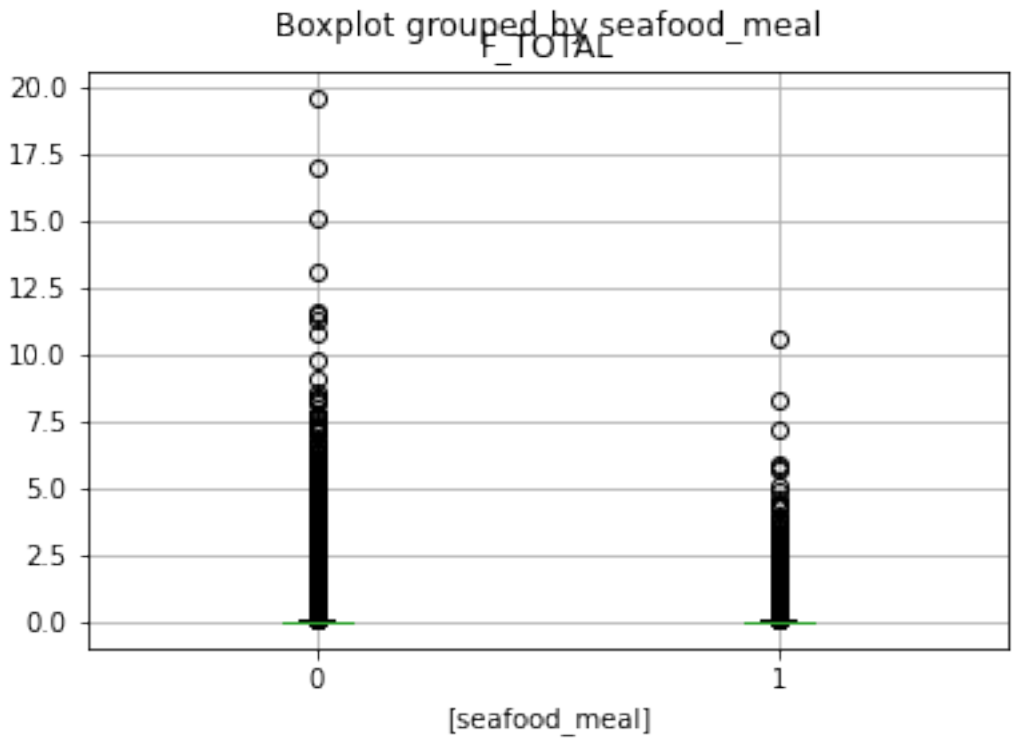
Provide statistical distribution plots of FPED components, for comparison on how they are distributed among the seafood vs non-seafood meals.

### Section 1: Fruits

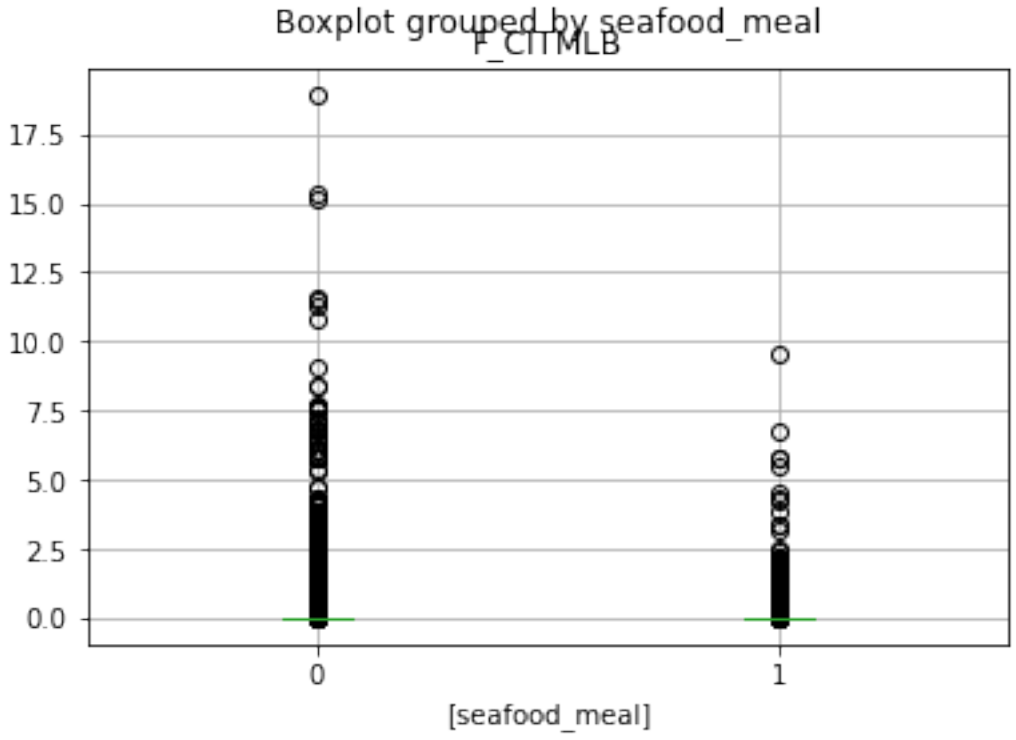
This section provides boxplots and density plots of the Fruit FPED components in the seafood meal and non seafood meal groups. The code for seafood meal is 1 if meal contains seafood, and 0 if meal does not contain seafood.

[56]:
```
import pandas as pd

#Read data frame and add plant pf total variable
df = pd.read_csv('../../Data/nhanes_full_pre_proc.csv')
```

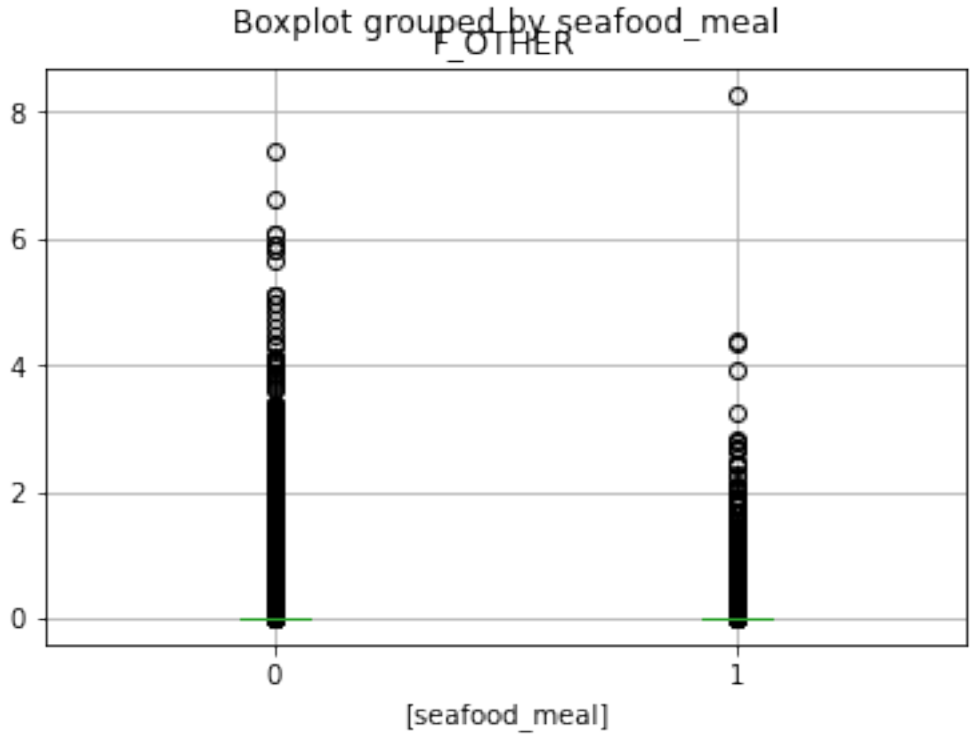
[55]:
```
import matplotlib.pyplot as plt

fruits = ['F_TOTAL', 'F_CITMLB', 'F_OTHER', 'F_JUICE']

for var in fruits:
    z = df.boxplot(column=var,by=['seafood_meal'])
    plt.show(z)
    plt.clf()
```



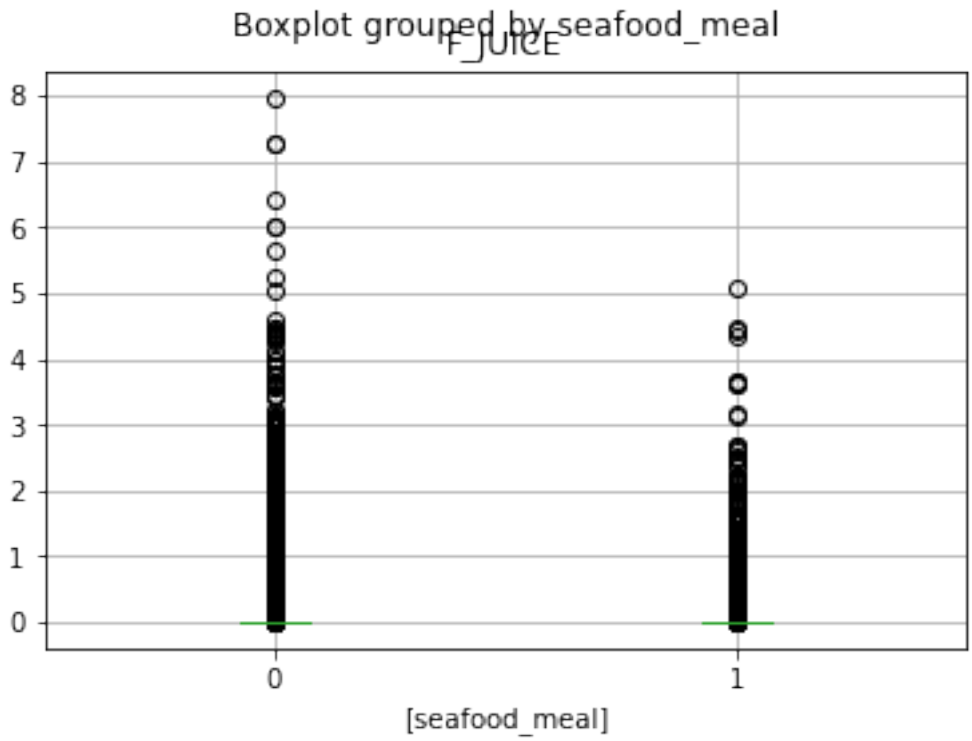

```
<Figure size 432x288 with 0 Axes>
```



```
<Figure size 432x288 with 0 Axes>
```

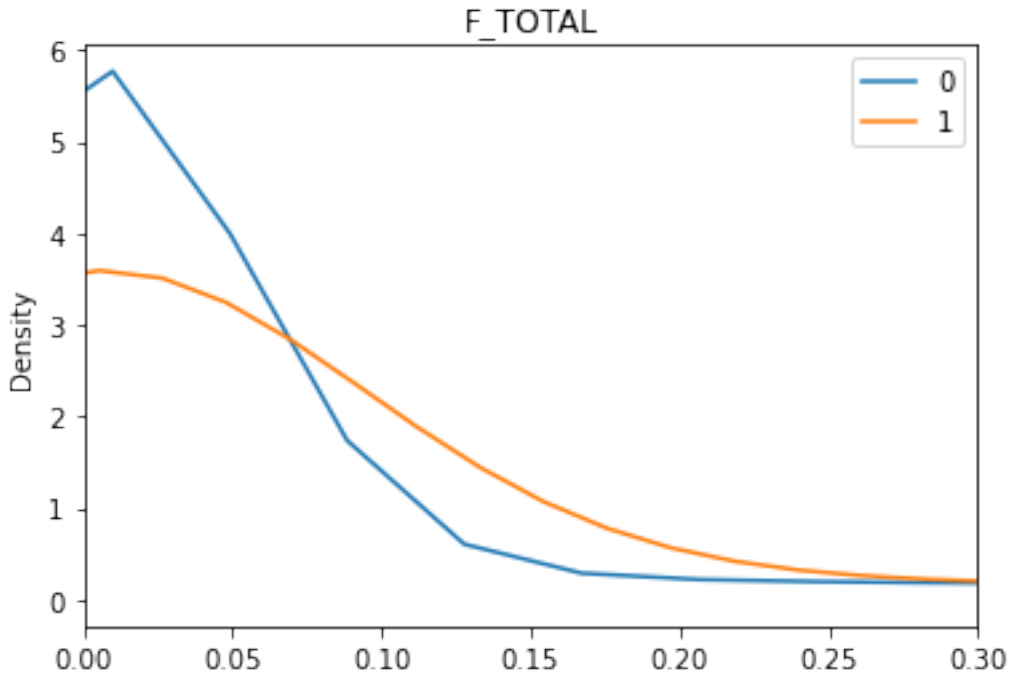


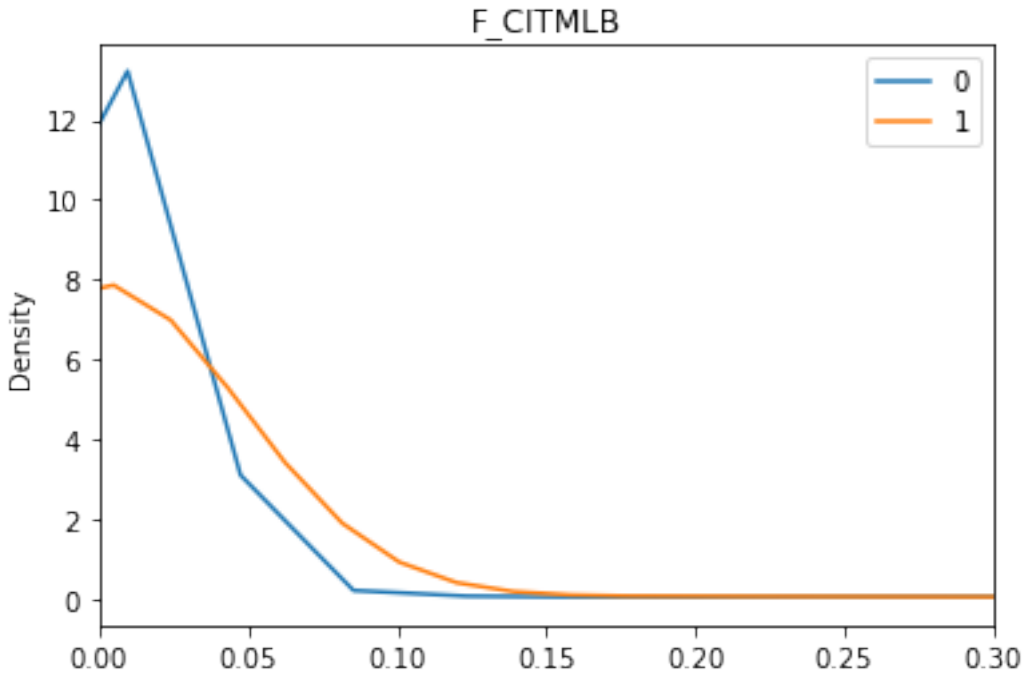

```
<Figure size 432x288 with 0 Axes>
```

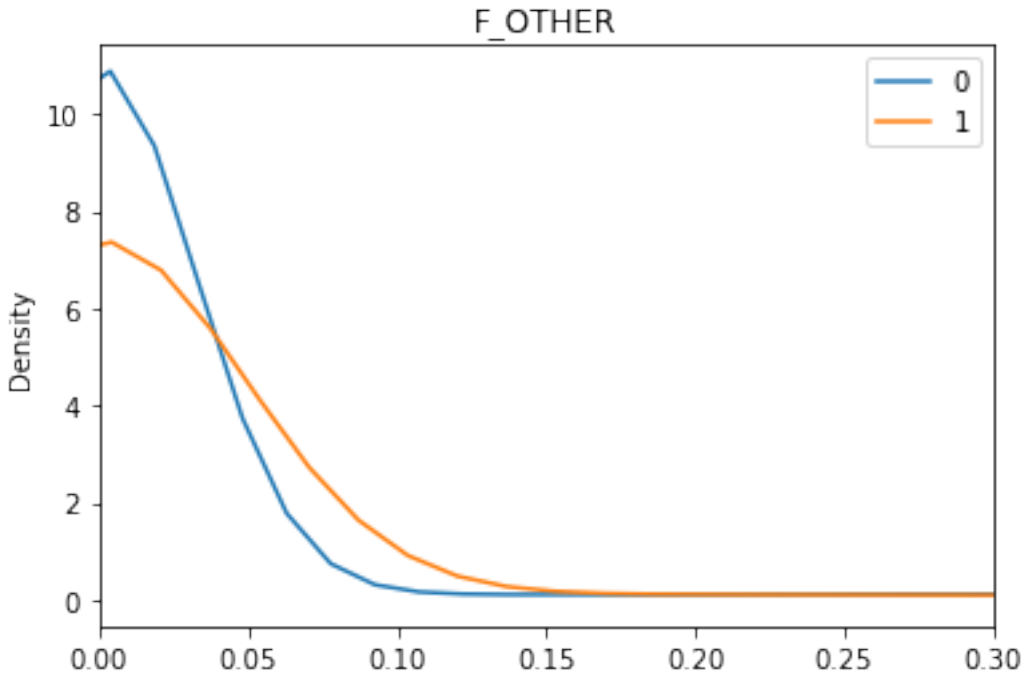


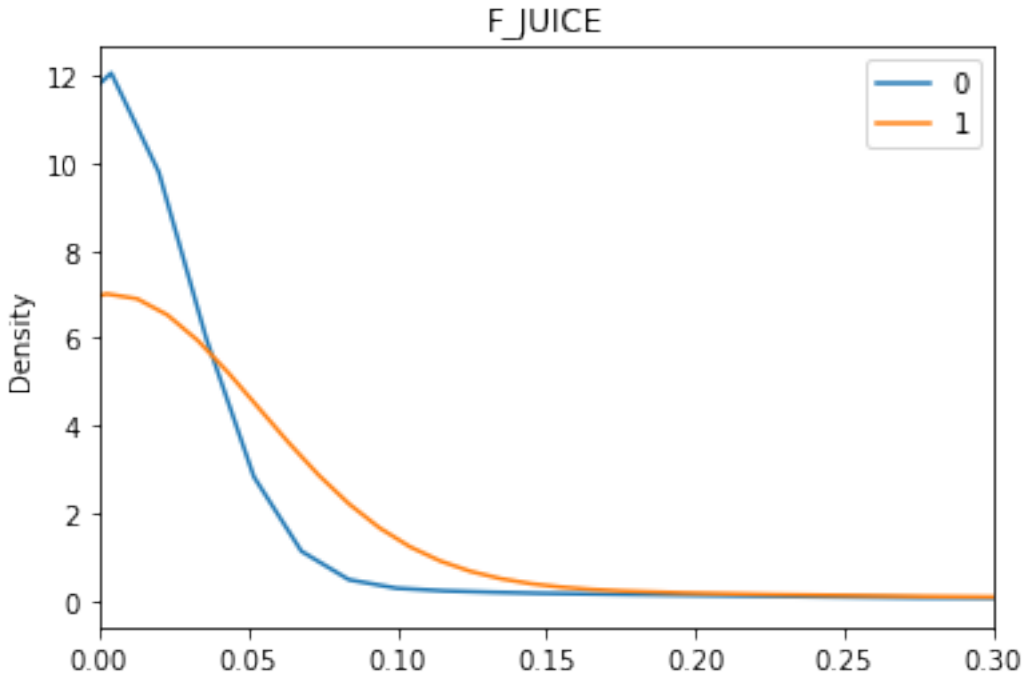

```
<Figure size 432x288 with 0 Axes>
```

[54]:
```
for var in fruits:
    z = df.groupby('seafood_meal')[var].plot.kde(title = var, legend='x')
    plt.show(z[0].set_xlim(0, 0.3))
    plt.clf()
```


F_TOTAL
Density
0
1
0
1
2
3
4
5
6
0.00
0.05
0.10
0.15
0.20
0.25
0.30


F_CITMLB
Density
0
1
0
2
4
6
8
10
12
0.00
0.05
0.10
0.15
0.20
0.25
0.30




```
<Figure size 432x288 with 0 Axes>
```

[114]:
```
for var in fruits:
    z = df.groupby('seafood_meal')[var].describe()
    print("Statistics for "+var+'\n')
    print(z)
    print('\n')
```

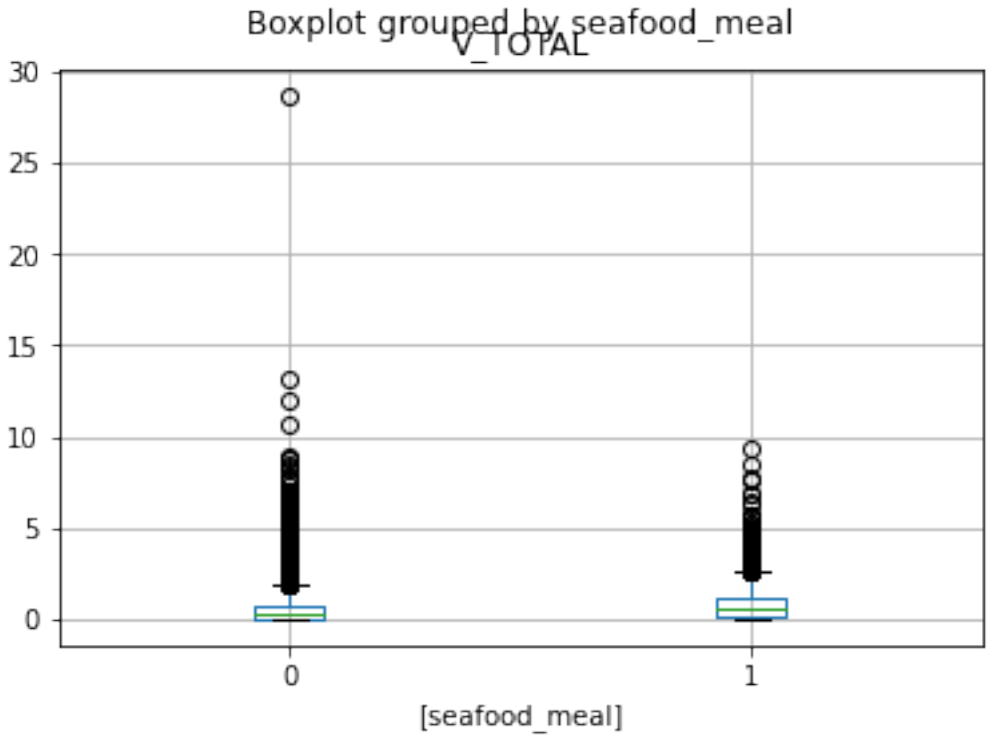
```
Statistics for F_TOTAL

                count      mean       std  min  25%  50%   75%    max
seafood_meal
0             94022.0  0.191101  0.517504  0.0  0.0  0.0  0.03  19.64
1              7709.0  0.188020  0.522423  0.0  0.0  0.0  0.05  10.64


Statistics for F_CITMLB

                count      mean       std  min  25%  50%  75%    max
seafood_meal
0             94022.0  0.035339  0.264883  0.0  0.0  0.0  0.0  18.94
1              7709.0  0.044862  0.284585  0.0  0.0  0.0  0.0   9.58


Statistics for F_OTHER

                count      mean       std  min  25%  50%  75%   max
seafood_meal
0             94022.0  0.085111  0.317481  0.0  0.0  0.0  0.0  7.39
1              7709.0  0.068436  0.291992  0.0  0.0  0.0  0.0  8.28


Statistics for F_JUICE

                count      mean      std  min  25%  50%  75%   max
seafood_meal
0             94022.0  0.070646  0.28669  0.0  0.0  0.0  0.0  7.97
1              7709.0  0.074720  0.29741  0.0  0.0  0.0  0.0  5.08
```

### Section 2: Vegetables

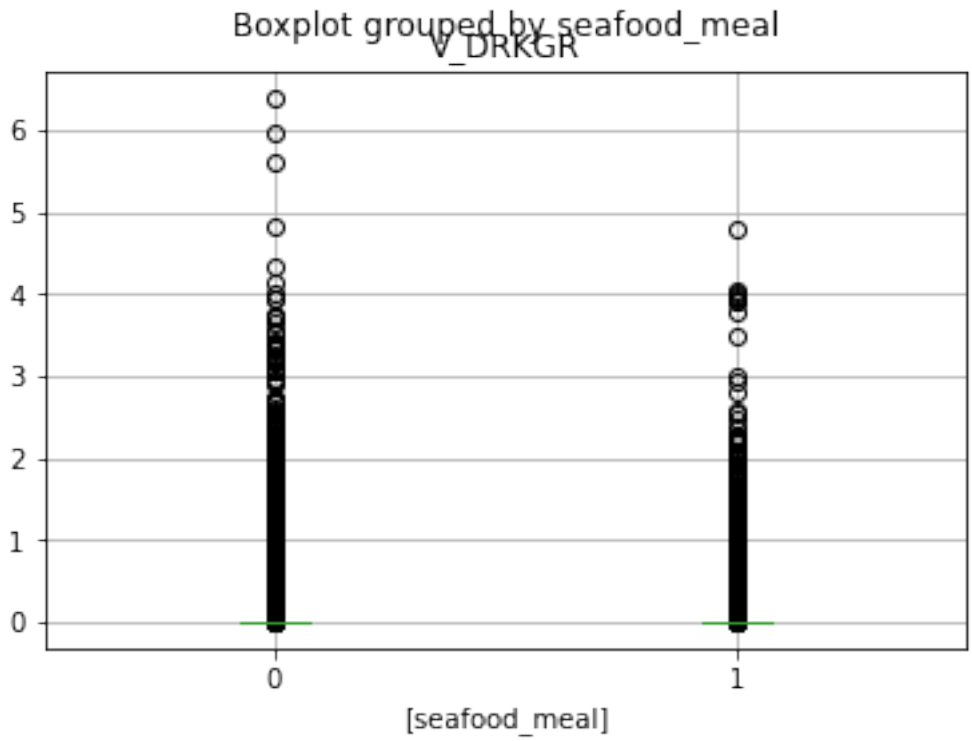
This section provides boxplots and density plots of the Vegetable FPED components in the seafood meal and non seafood meal groups. The code for seafood meal is 1 if meal contains seafood, and 0 if meal does not contain seafood.

[57]:
```
veggie = ['V_TOTAL', 'V_DRKGR', 'V_REDOR_TOMATO', 'V_REDOR_OTHER',␣
 ↪'V_STARCHY_POTATO',
                   'V_STARCHY_OTHER', 'V_OTHER', 'V_LEGUMES']
```

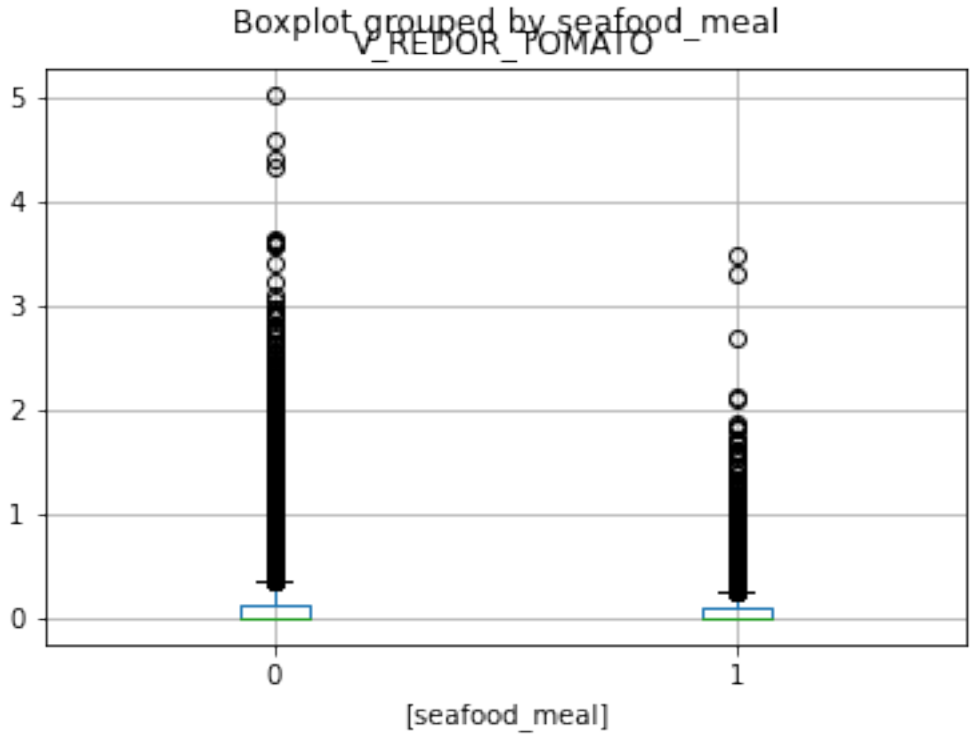
```
for var in veggie:
    z = df.boxplot(column=var,by=['seafood_meal'])
    plt.show(z)
    plt.clf()
```



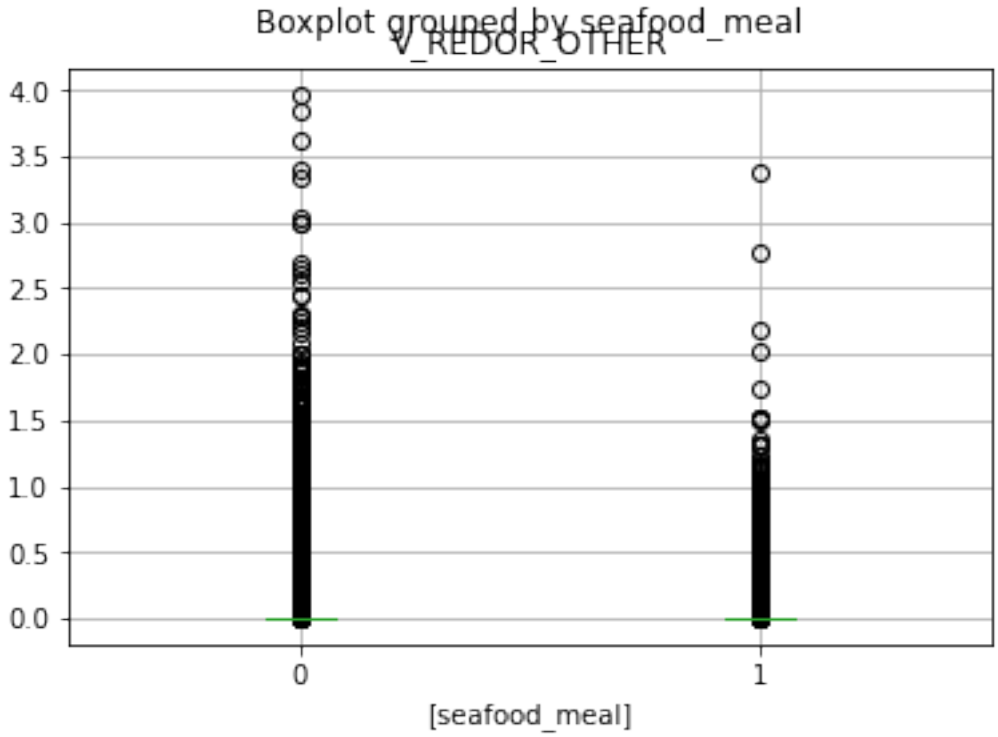

```
<Figure size 432x288 with 0 Axes>
```



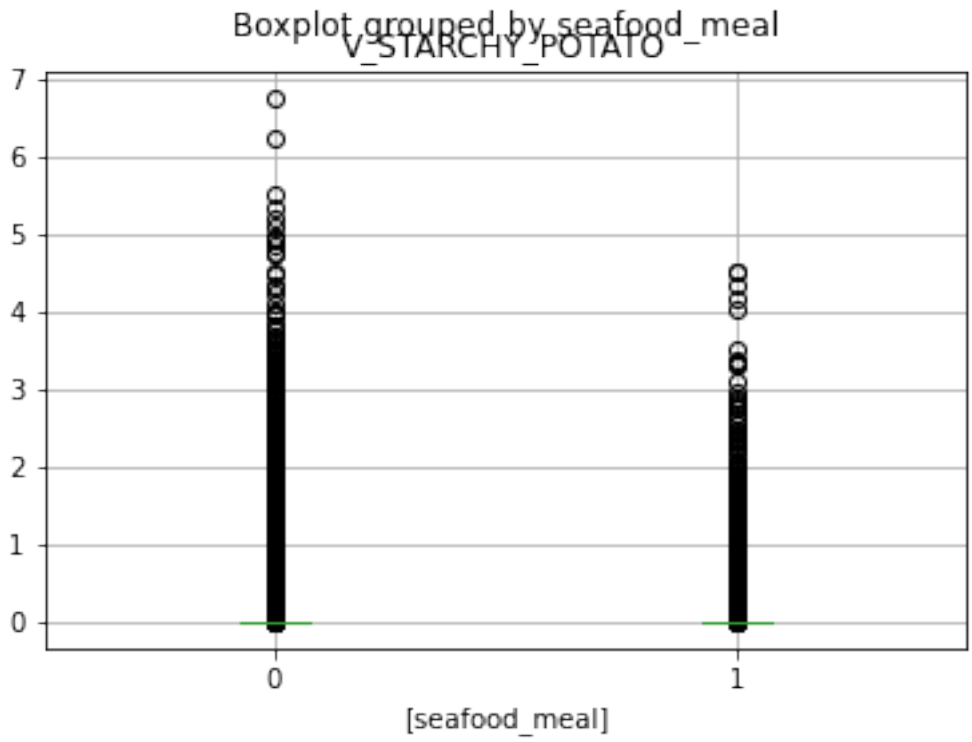

```
<Figure size 432x288 with 0 Axes>
```



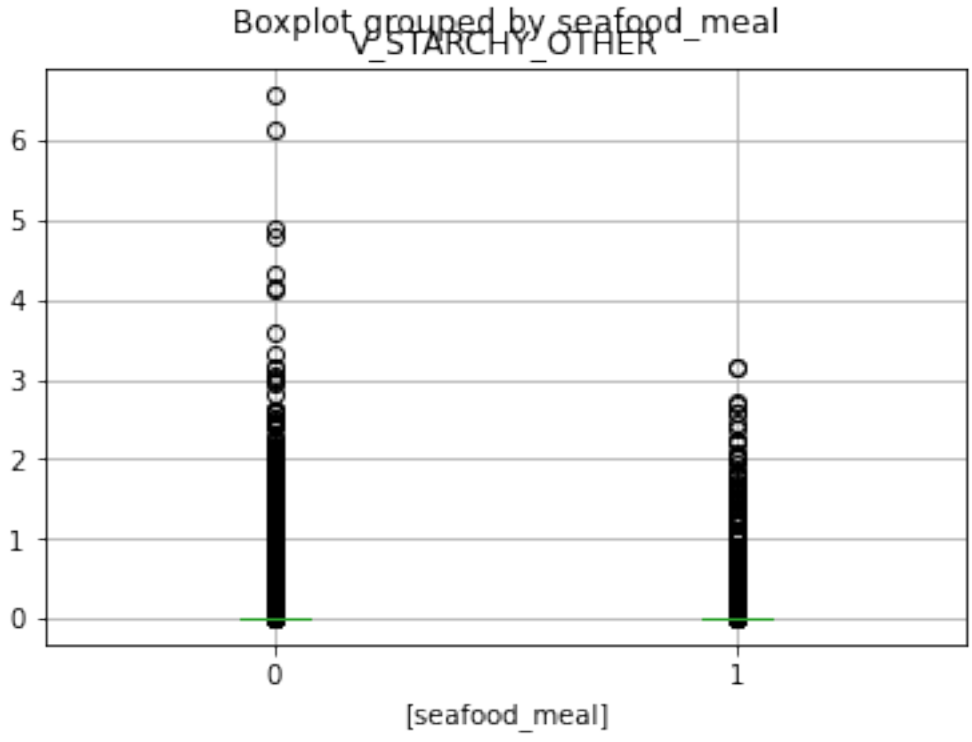
```
<Figure size 432x288 with 0 Axes>
```



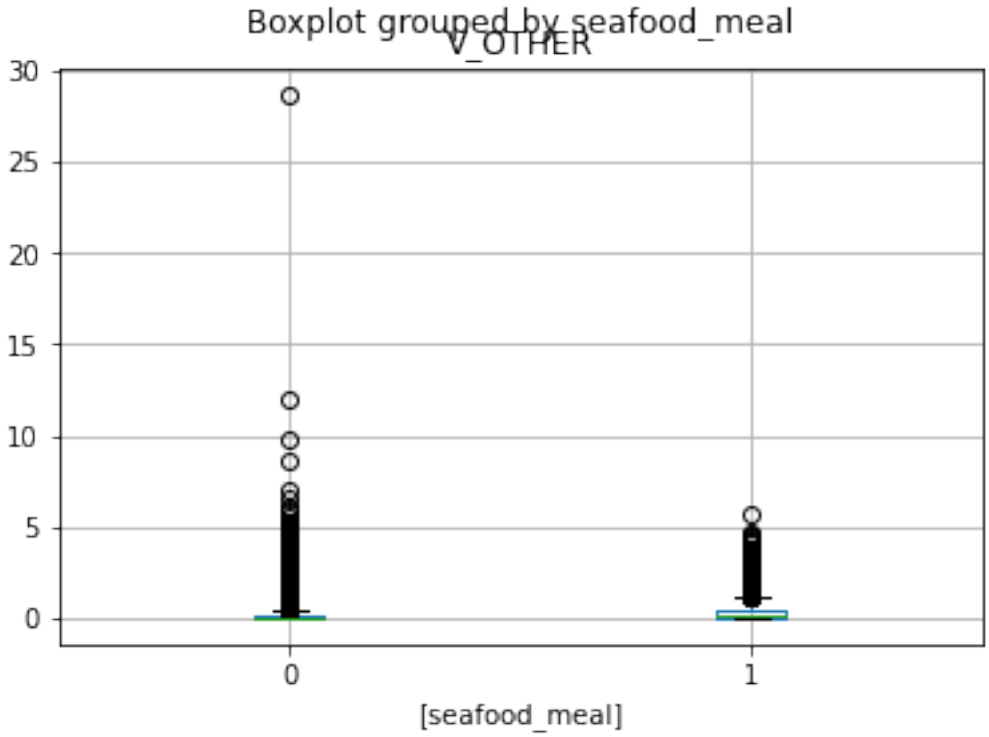

```
<Figure size 432x288 with 0 Axes>
```



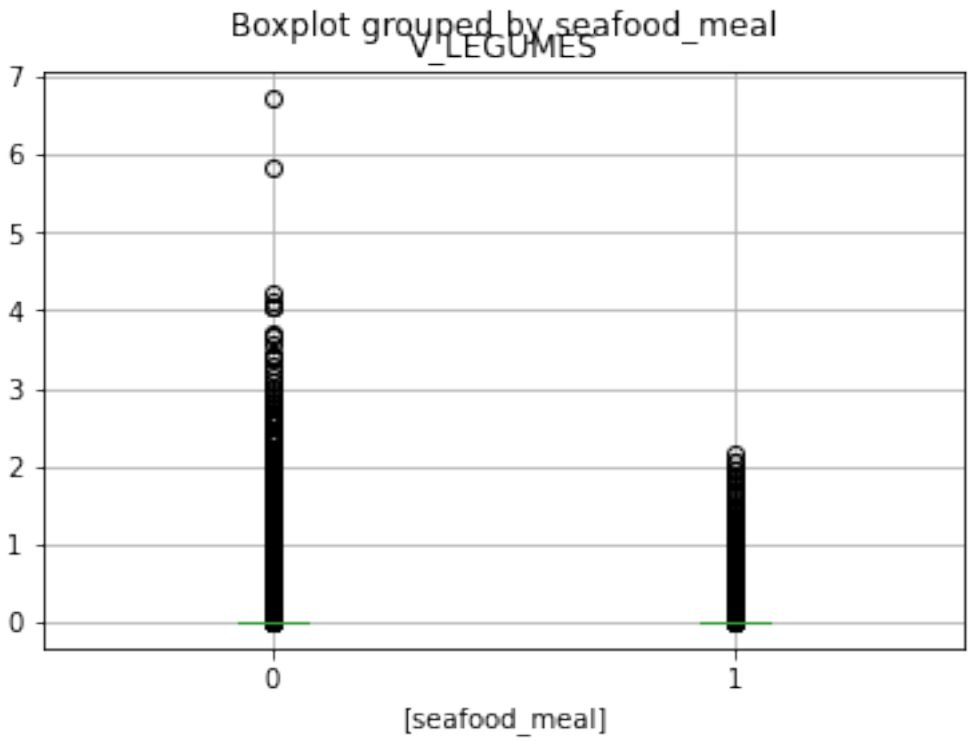

<Figure size 432x288 with 0 Axes>



```
<Figure size 432x288 with 0 Axes>
```

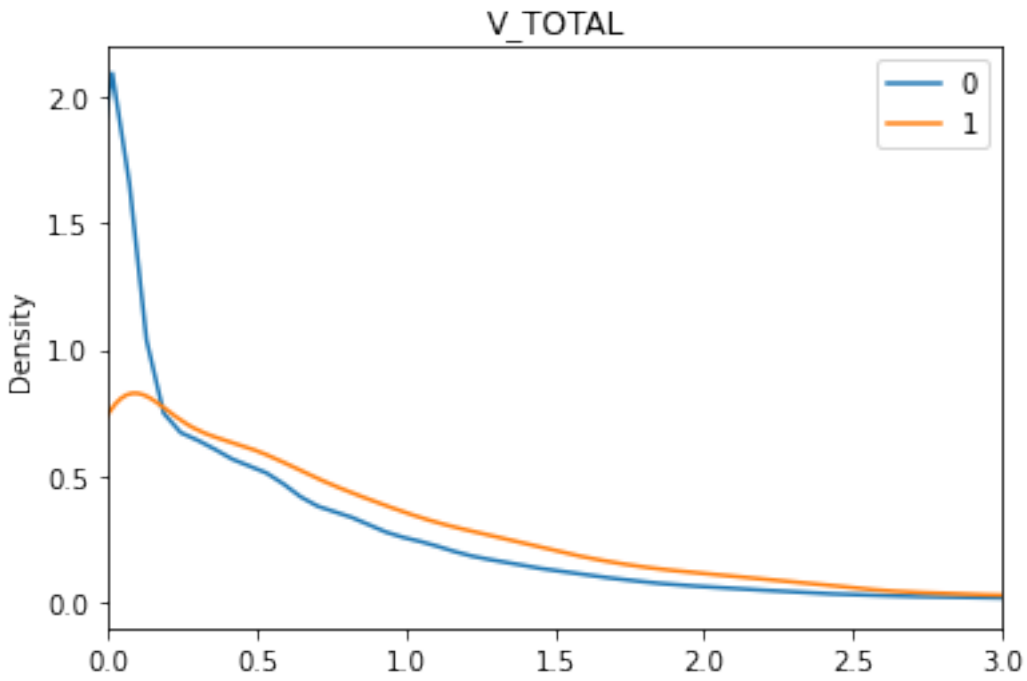



```
<Figure size 432x288 with 0 Axes>
```



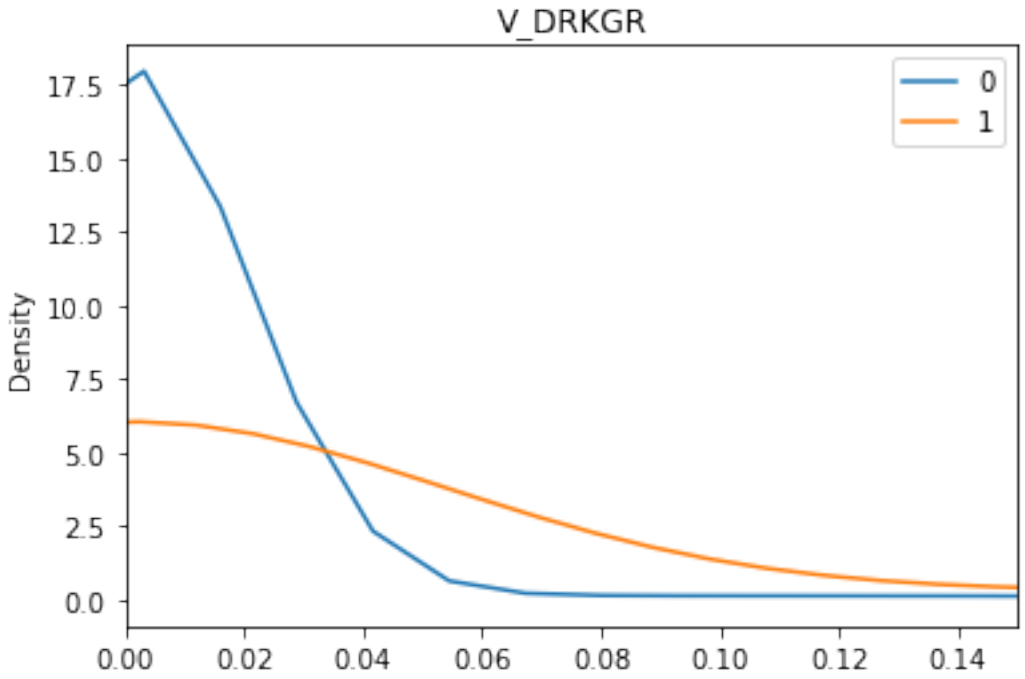

```
<Figure size 432x288 with 0 Axes>
```

[72]:
```
z = df.groupby('seafood_meal')[veggie[0]].plot.kde(title = veggie[0],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 3))
plt.clf()
```



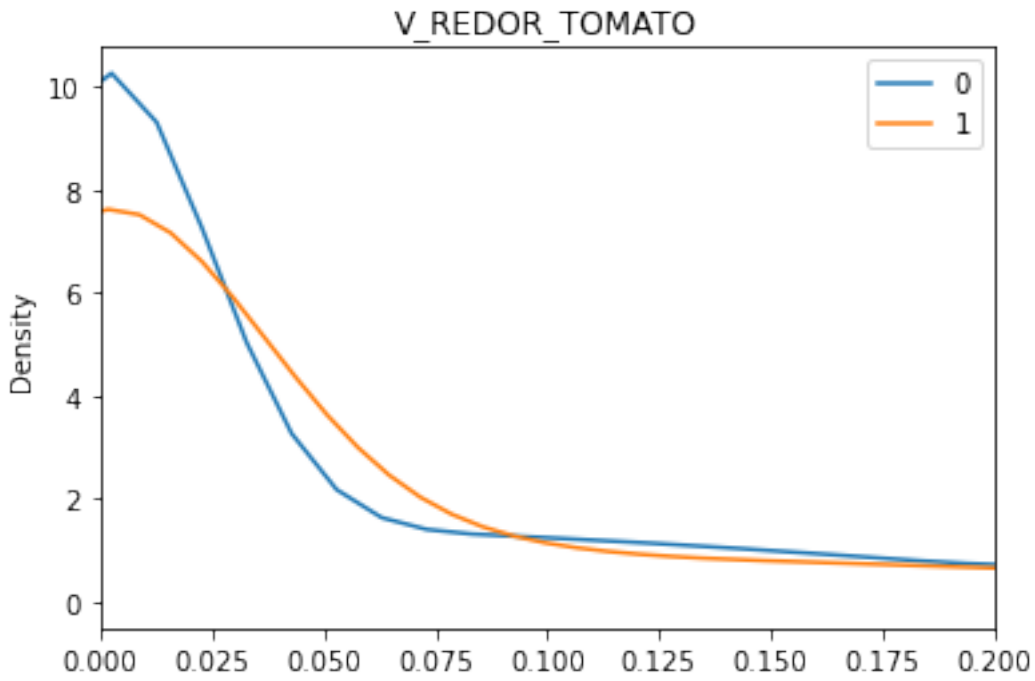

```
<Figure size 432x288 with 0 Axes>
```

[73]:
```
z = df.groupby('seafood_meal')[veggie[1]].plot.kde(title = veggie[1],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 0.15))
plt.clf()
```



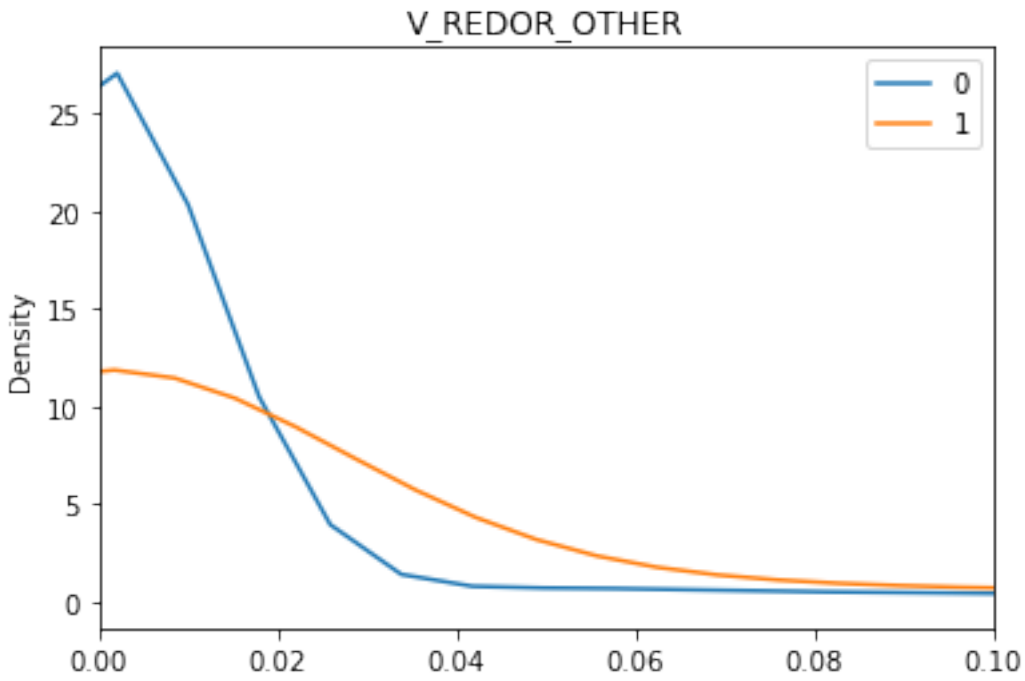

```
<Figure size 432x288 with 0 Axes>
```

[78]:
```
z = df.groupby('seafood_meal')[veggie[2]].plot.kde(title = veggie[2],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 0.2))
plt.clf()
```



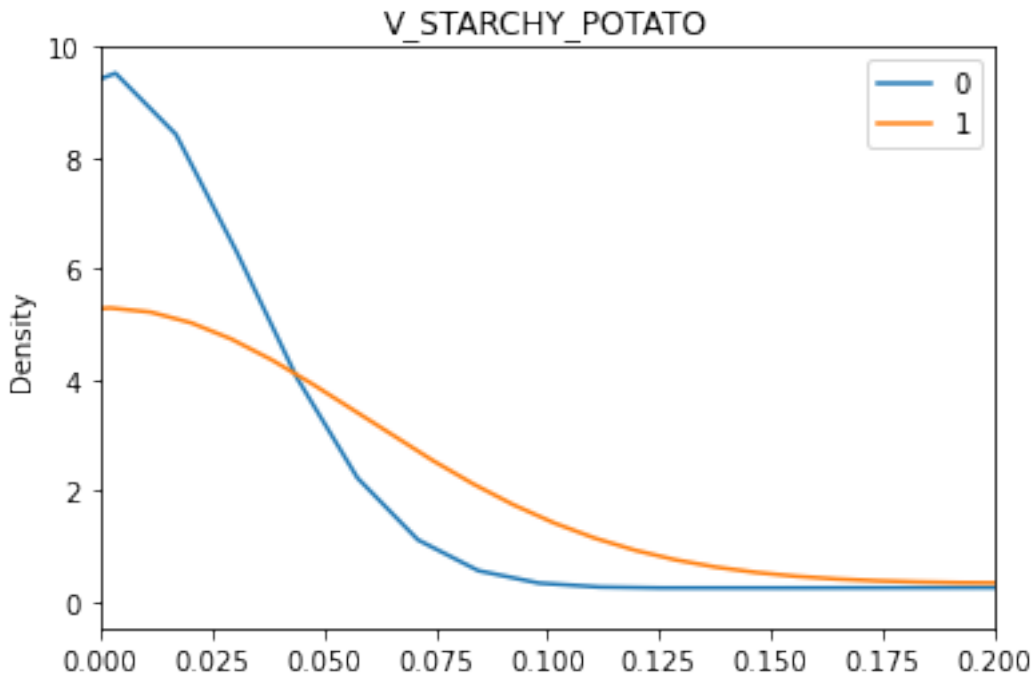

```
<Figure size 432x288 with 0 Axes>
```

[79]:
```
z = df.groupby('seafood_meal')[veggie[3]].plot.kde(title = veggie[3],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 0.1))
plt.clf()
```



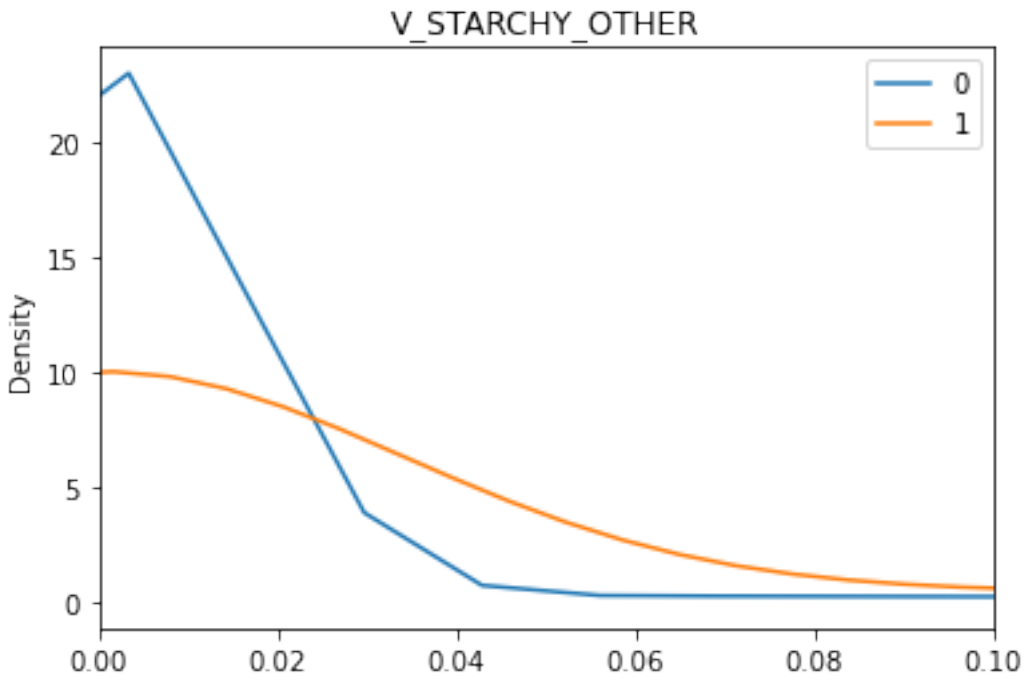

```
<Figure size 432x288 with 0 Axes>
```

[80]:
```
z = df.groupby('seafood_meal')[veggie[4]].plot.kde(title = veggie[4],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 0.2))
plt.clf()
```



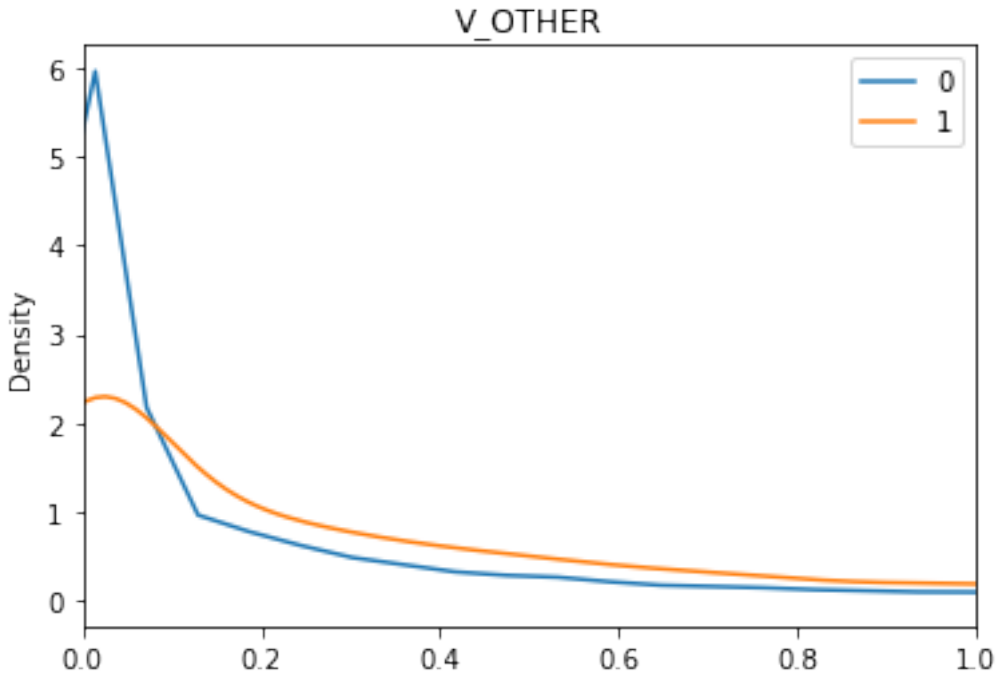

```
<Figure size 432x288 with 0 Axes>
```

[81]:
```
z = df.groupby('seafood_meal')[veggie[5]].plot.kde(title = veggie[5],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 0.1))
plt.clf()
```



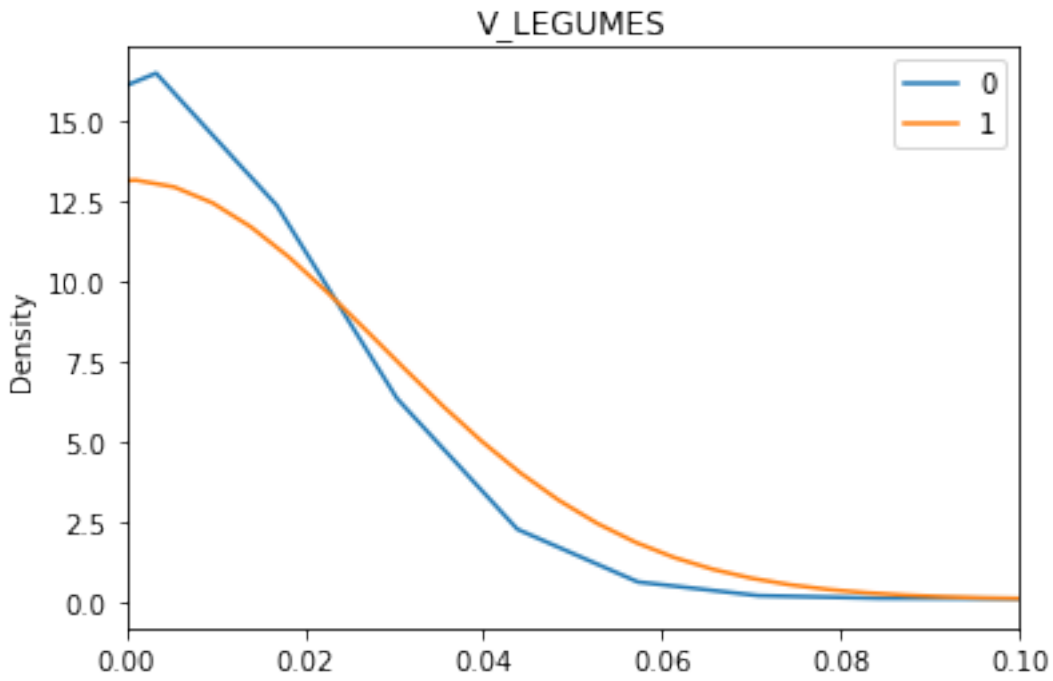

```
<Figure size 432x288 with 0 Axes>
```

[82]:
```
z = df.groupby('seafood_meal')[veggie[6]].plot.kde(title = veggie[6],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 1))
plt.clf()
```




```
<Figure size 432x288 with 0 Axes>
```

[84]:
```
z = df.groupby('seafood_meal')[veggie[7]].plot.kde(title = veggie[7],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 0.1))
plt.clf()
```




<Figure size 432x288 with 0 Axes>

[115]:
```
for var in veggie:
    z = df.groupby('seafood_meal')[var].describe()
    print("Statistics for "+var+'\n')
    print(z)
    print('\n')
```

Statistics for V_TOTAL

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.519475 | 0.714251 | 0.0 | 0.00 | 0.27 | 0.76 | 28.7 |
| 1 | 7709.0 | 0.779420 | 0.837273 | 0.0 | 0.16 | 0.54 | 1.13 | 9.4 |

Statistics for V_DRKGR

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.040892 | 0.199350 | 0.0 | 0.0 | 0.0 | 0.0 | 6.40 |
| 1 | 7709.0 | 0.109165 | 0.319705 | 0.0 | 0.0 | 0.0 | 0.0 | 4.78 |

Statistics for V_REDOR_TOMATO

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.114504 | 0.235472 | 0.0 | 0.0 | 0.0 | 0.14 | 5.03 |
| 1 | 7709.0 | 0.098916 | 0.213594 | 0.0 | 0.0 | 0.0 | 0.10 | 3.49 |

Statistics for V_REDOR_OTHER

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.030895 | 0.124817 | 0.0 | 0.0 | 0.0 | 0.0 | 3.97 |
| 1 | 7709.0 | 0.053303 | 0.159010 | 0.0 | 0.0 | 0.0 | 0.0 | 3.37 |

Statistics for V_STARCHY_POTATO

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.122740 | 0.324913 | 0.0 | 0.0 | 0.0 | 0.0 | 6.76 |
| 1 | 7709.0 | 0.133004 | 0.349785 | 0.0 | 0.0 | 0.0 | 0.0 | 4.51 |

Statistics for V_STARCHY_OTHER

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.032452 | 0.152372 | 0.0 | 0.0 | 0.0 | 0.0 | 6.58 |
| 1 | 7709.0 | 0.055058 | 0.202960 | 0.0 | 0.0 | 0.0 | 0.0 | 3.16 |

Statistics for V_OTHER

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.178000 | 0.390975 | 0.0 | 0.0 | 0.00 | 0.19 | 28.70 |
| 1 | 7709.0 | 0.329877 | 0.484034 | 0.0 | 0.0 | 0.14 | 0.47 | 5.65 |

Statistics for V_LEGUMES

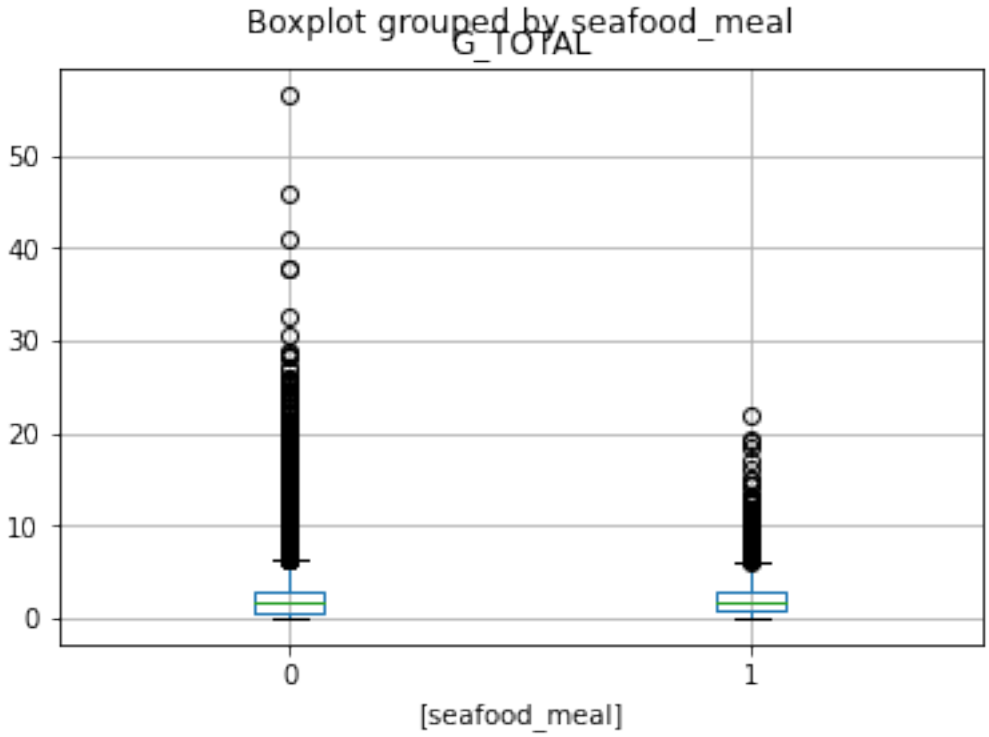
| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.052059 | 0.213986 | 0.0 | 0.0 | 0.0 | 0.0 | 6.73 |
| 1 | 7709.0 | 0.035462 | 0.169121 | 0.0 | 0.0 | 0.0 | 0.0 | 2.15 |

### Section 3: Grains

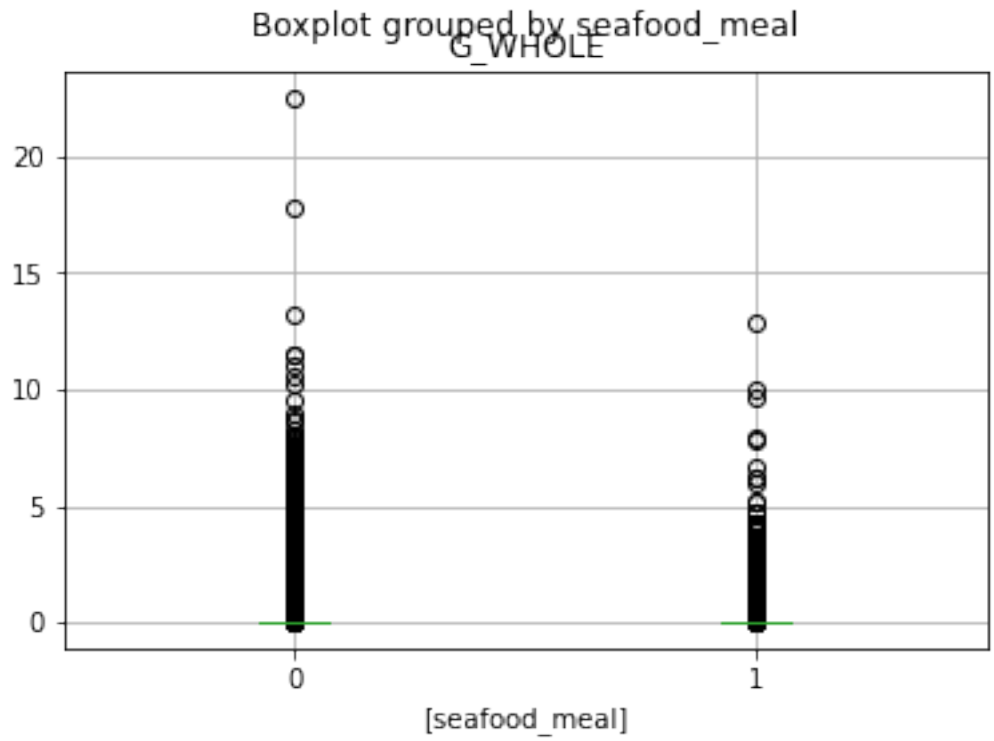
This section provides boxplots and density plots of the Grains FPED components in the seafood meal and non seafood meal groups. The code for seafood meal is 1 if meal contains seafood, and 0 if meal does not contain seafood.

Plot: Meal calories distribution with meals that are 0 KCAL removed, split by 'eathome' groups.

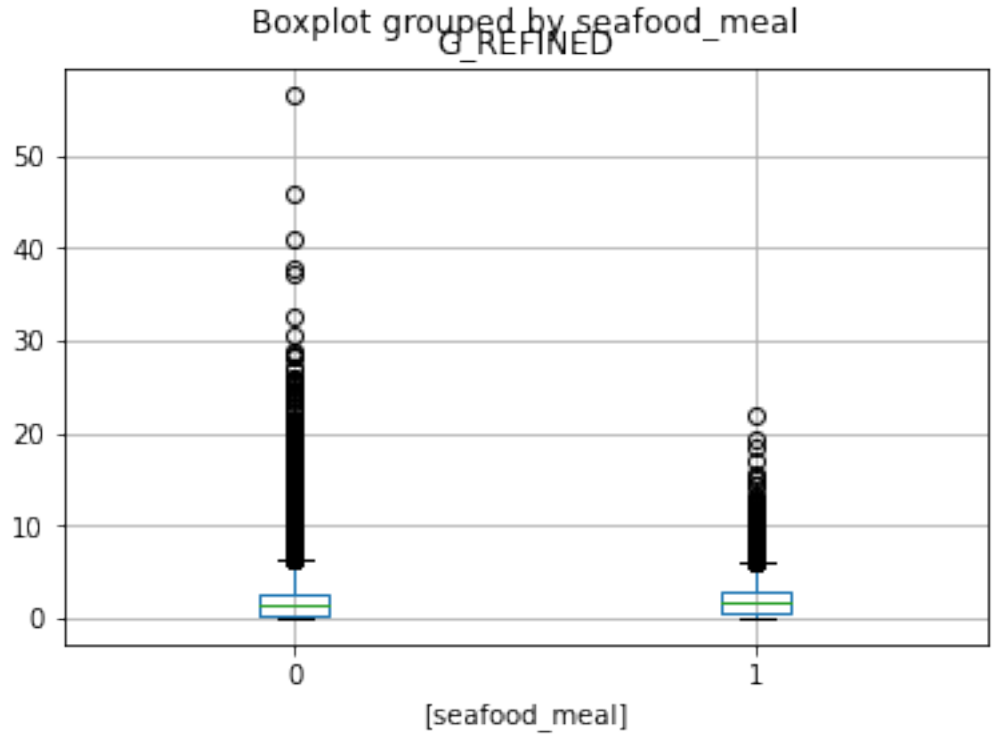
[61]:
```
grains = ['G_TOTAL', 'G_WHOLE','G_REFINED']

for var in grains:
    z = df.boxplot(column=var,by=['seafood_meal'])
    plt.show(z)
    plt.clf()
```




```
<Figure size 432x288 with 0 Axes>
```

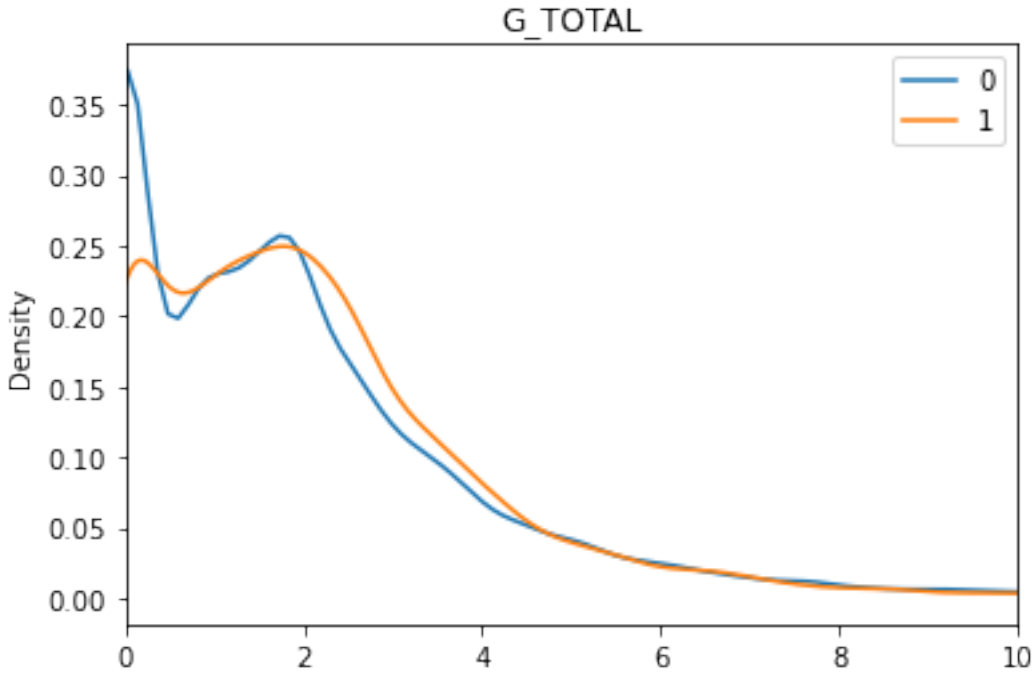



```
<Figure size 432x288 with 0 Axes>
```



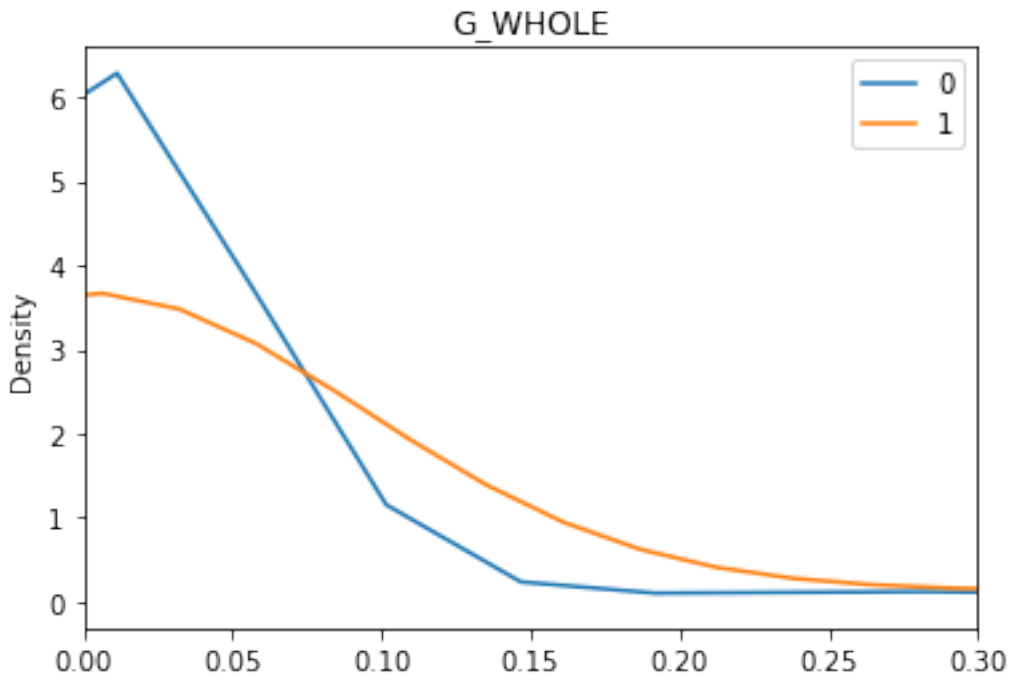
```
<Figure size 432x288 with 0 Axes>
```

[87]:
```
z = df.groupby('seafood_meal')[grains[0]].plot.kde(title = grains[0],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 10))
plt.clf()
```



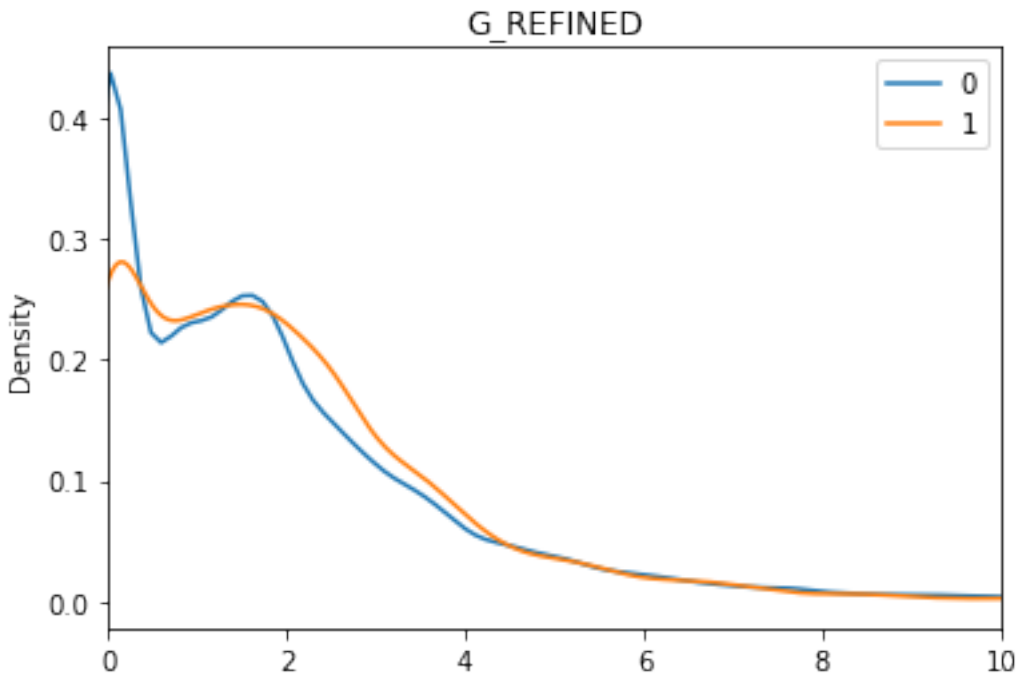

```
<Figure size 432x288 with 0 Axes>
```

[88]:
```
z = df.groupby('seafood_meal')[grains[1]].plot.kde(title = grains[1],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 0.3))
plt.clf()
```




```
<Figure size 432x288 with 0 Axes>
```

[89]:
```
z = df.groupby('seafood_meal')[grains[2]].plot.kde(title = grains[2],␣
 ↪legend='x')
plt.show(z[0].set_xlim(0, 10))
plt.clf()
```




<Figure size 432x288 with 0 Axes>

[116]:
```
for var in grains:
    z = df.groupby('seafood_meal')[var].describe()
    print("Statistics for "+var+'\n')
    print(z)
    print('\n')
```

Statistics for G_TOTAL

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 2.091975 | 2.239952 | 0.0 | 0.54 | 1.65 | 2.86 | 56.65 |
| 1 | 7709.0 | 2.147220 | 1.991164 | 0.0 | 0.76 | 1.81 | 2.94 | 21.82 |

Statistics for G_WHOLE

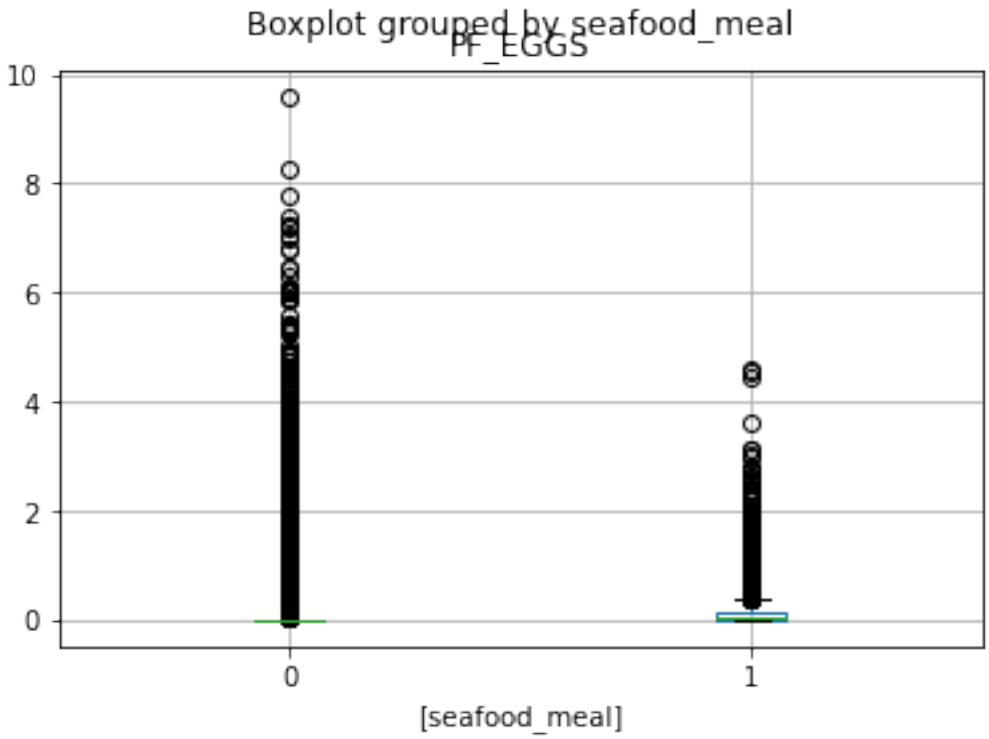
| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 0.142640 | 0.532072 | 0.0 | 0.0 | 0.0 | 0.0 | 22.55 |
| 1 | 7709.0 | 0.148072 | 0.561466 | 0.0 | 0.0 | 0.0 | 0.0 | 12.86 |

```
Statistics for G_REFINED

                count      mean       std  min   25%   50%   75%    max
seafood_meal
0             94022.0  1.949327  2.221885  0.0  0.34  1.43  2.70  56.65
1              7709.0  1.999153  1.955859  0.0  0.59  1.64  2.77  21.82
```

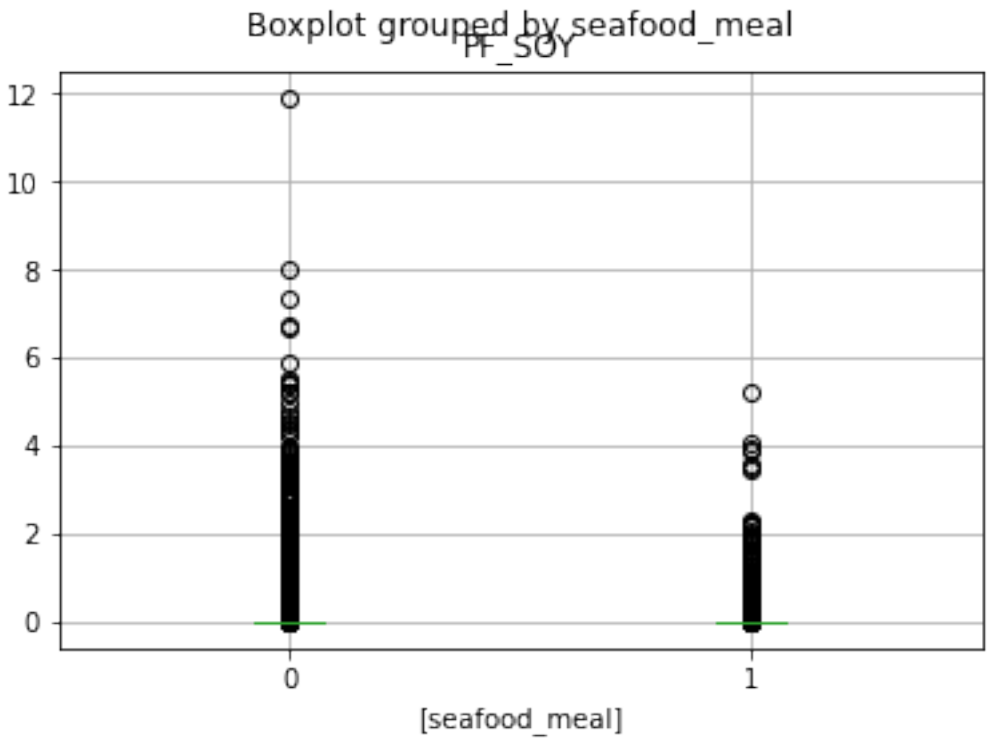
### Section 4: Non-Meat Proteins

This section provides boxplots and density plots of the non-meat protein FPED components in the seafood meal and non seafood meal groups. The code for seafood meal is 1 if meal contains seafood, and 0 if meal does not contain seafood.

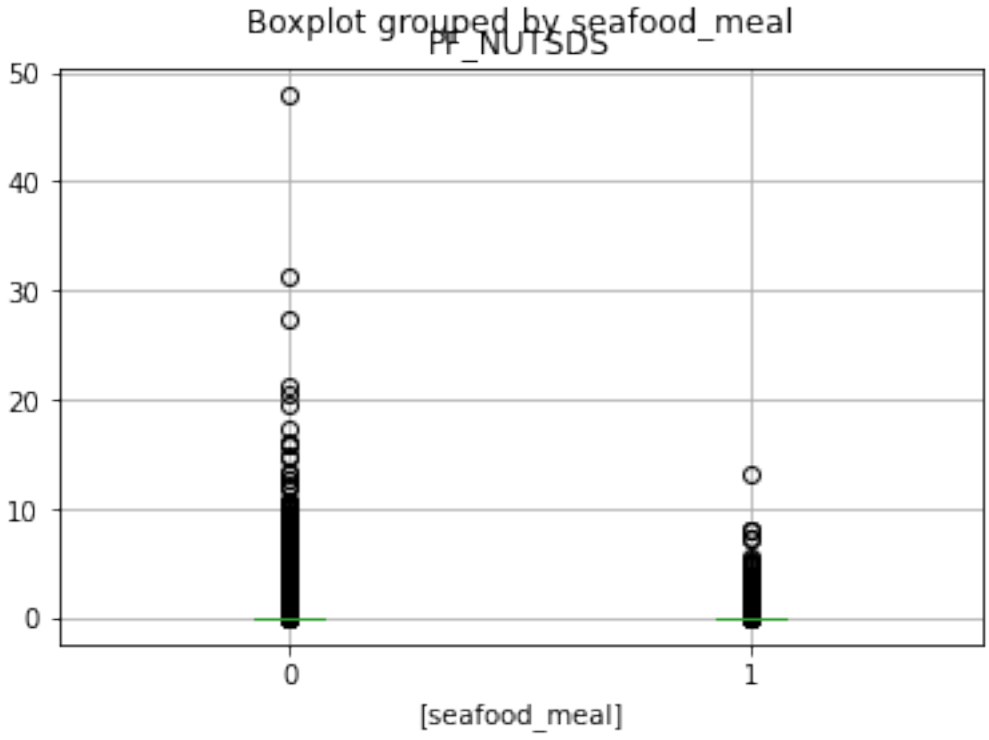
[63]:
```
non_meat_protein = ['PF_EGGS', 'PF_SOY', 'PF_NUTSDS', 'PF_LEGUMES']

for var in non_meat_protein:
    z = df.boxplot(column=var,by=['seafood_meal'])
    plt.show(z)
    plt.clf()
```



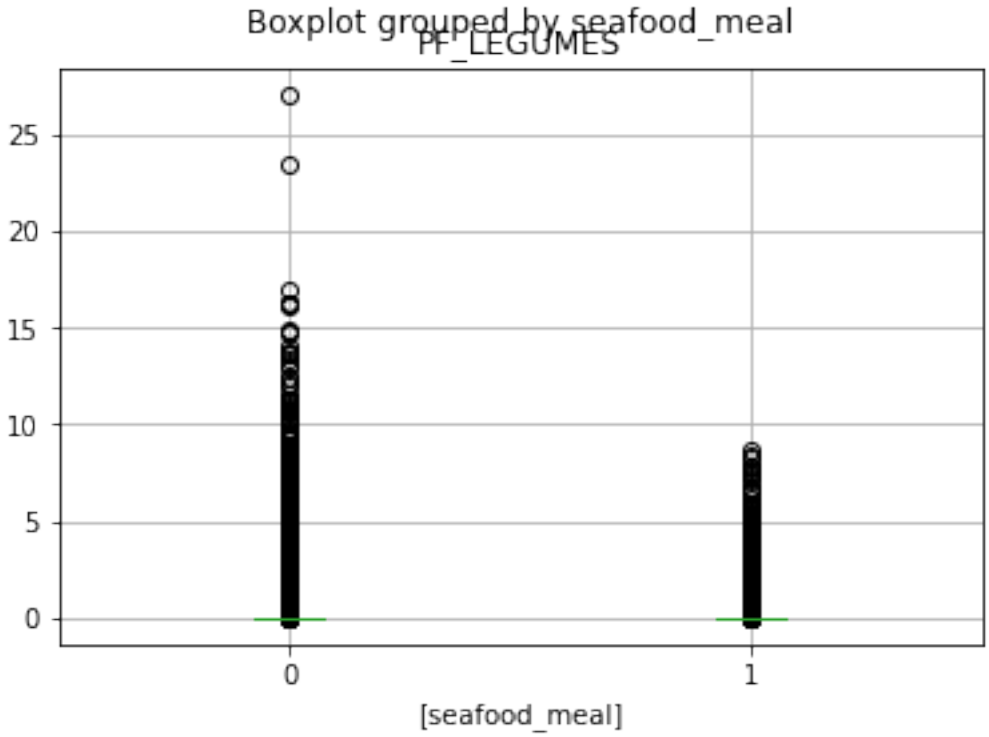

```
<Figure size 432x288 with 0 Axes>
```




```
<Figure size 432x288 with 0 Axes>
```

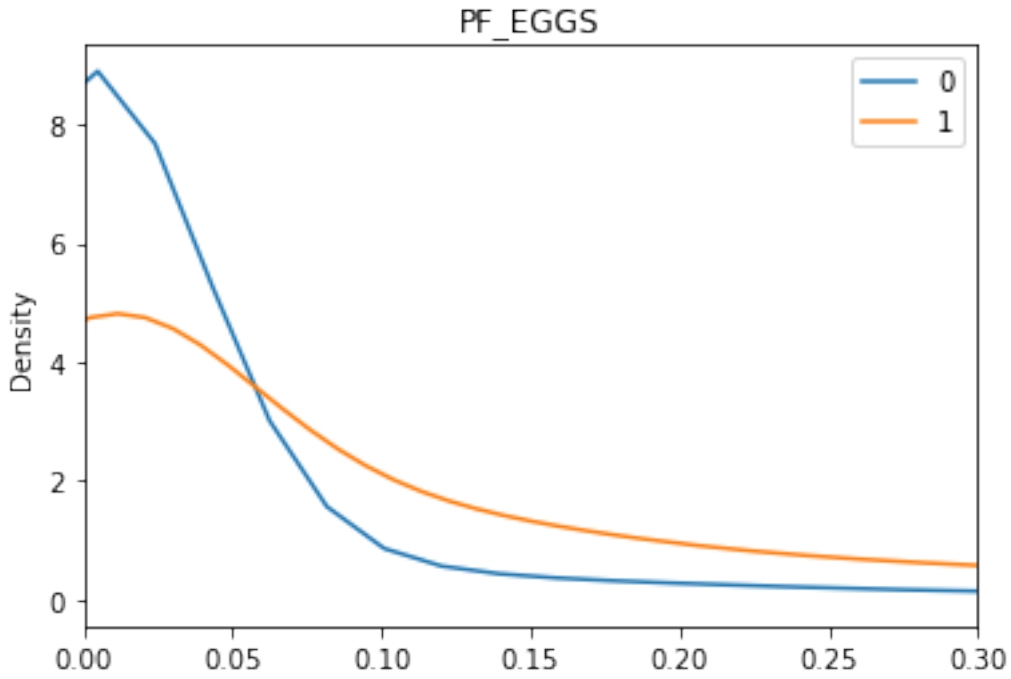


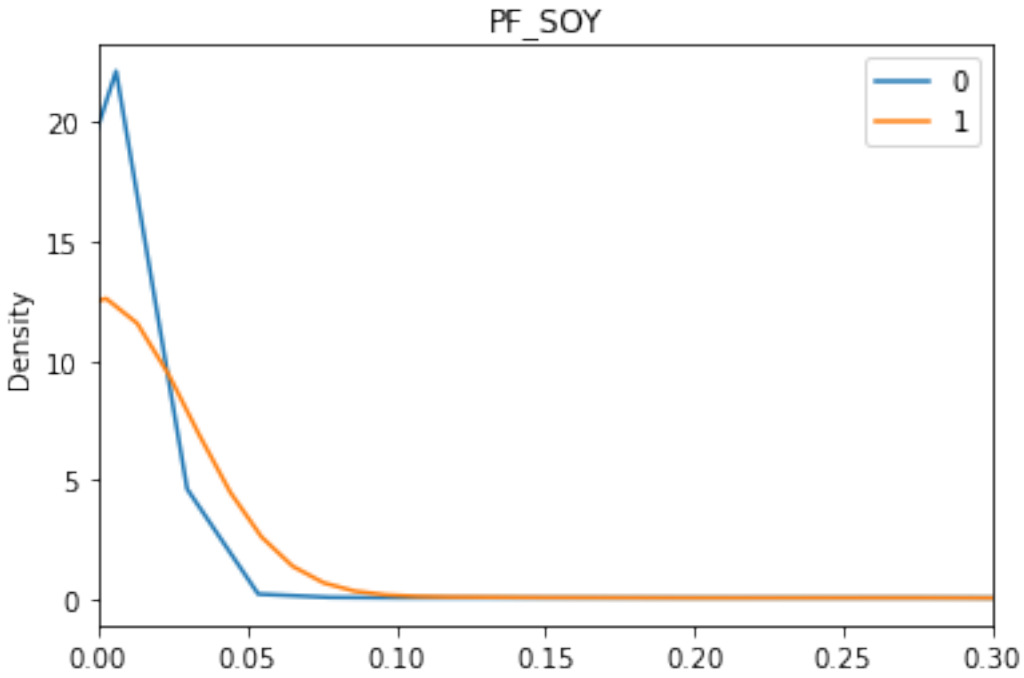
```
<Figure size 432x288 with 0 Axes>
```

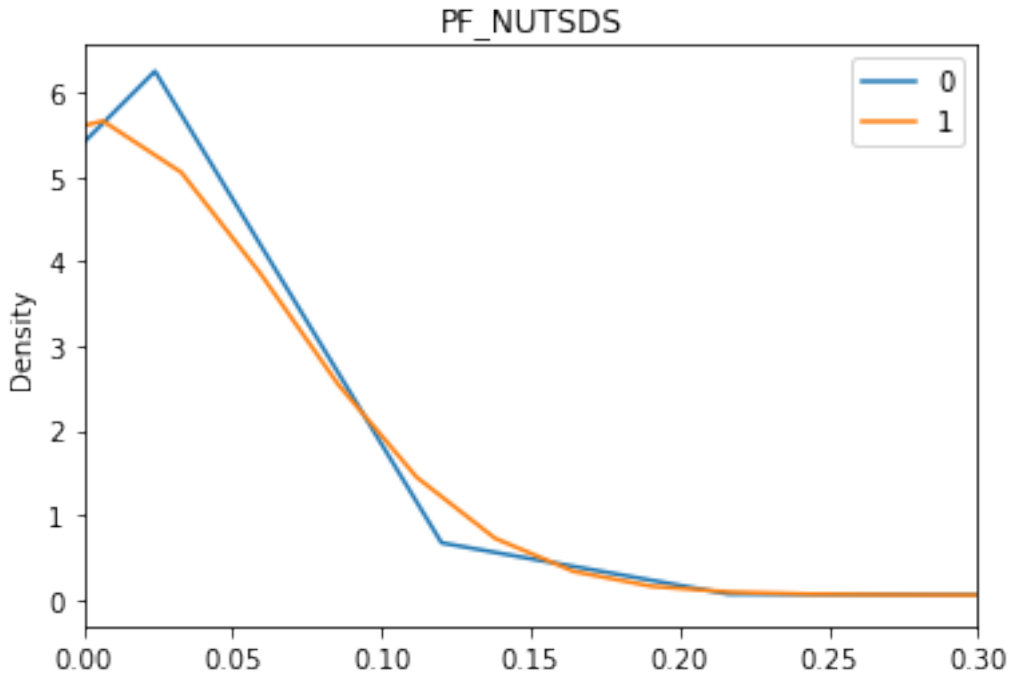


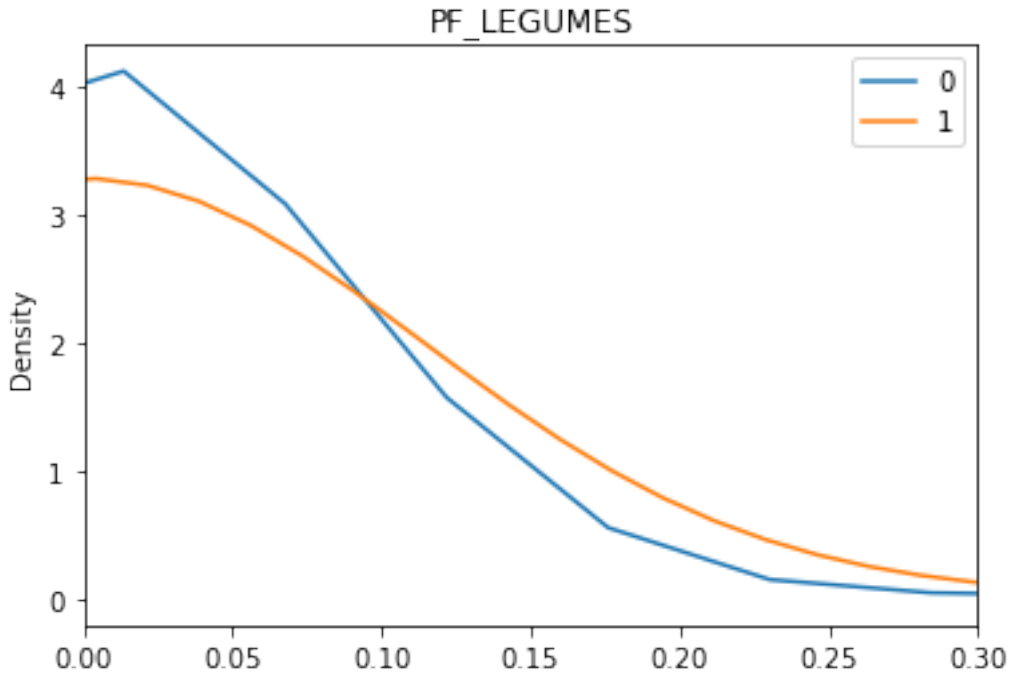

```
<Figure size 432x288 with 0 Axes>
```

[64]:
```
for var in non_meat_protein:
    z = df.groupby('seafood_meal')[var].plot.kde(title = var, legend='x')
    plt.show(z[0].set_xlim(0, 0.3))
    plt.clf()
```


PF_EGGS
0
1
Density
8
6
4
2
0
0.00
0.05
0.10
0.15
0.20
0.25
0.30


PF_SOY
0
1
Density
20
15
10
5
0
0.00
0.05
0.10
0.15
0.20
0.25
0.30




```
<Figure size 432x288 with 0 Axes>
```

[117]:
```
for var in non_meat_protein:
    z = df.groupby('seafood_meal')[var].describe()
    print("Statistics for "+var+'\n')
    print(z)
    print('\n')
```

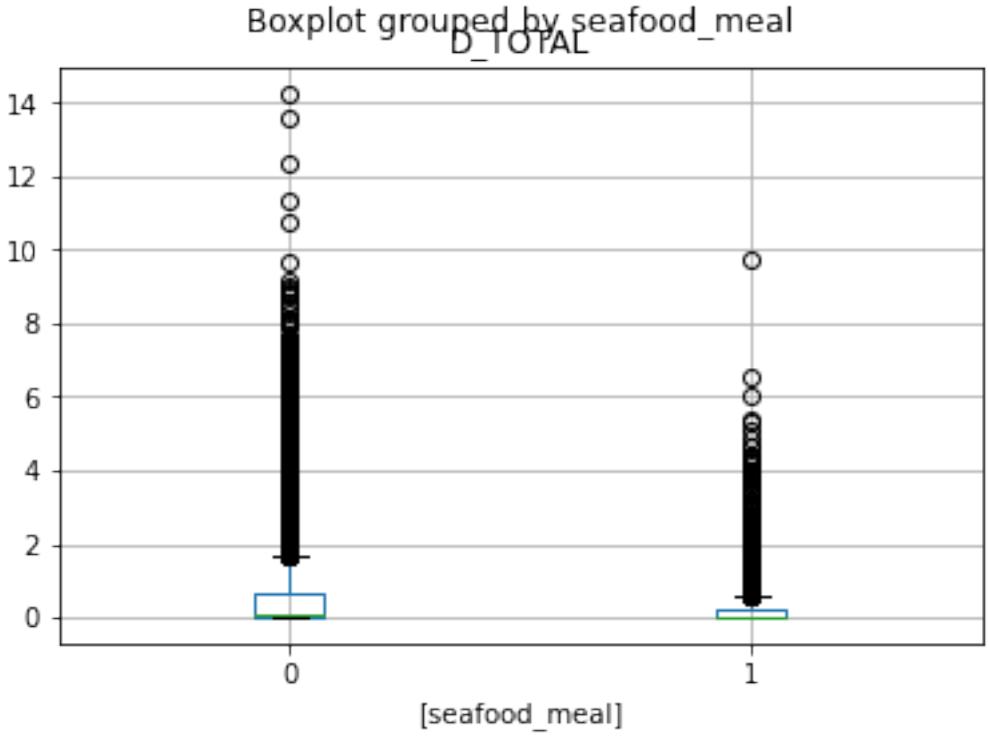
```
Statistics for PF_EGGS

                count      mean       std  min  25%   50%   75%   max
seafood_meal
0             94022.0  0.088269  0.367298  0.0  0.0  0.00  0.01  9.60
1              7709.0  0.143597  0.292593  0.0  0.0  0.03  0.16  4.62


Statistics for PF_SOY

                count      mean       std  min  25%  50%  75%    max
seafood_meal
0             94022.0  0.015055  0.161631  0.0  0.0  0.0  0.0  11.91
1              7709.0  0.021462  0.182076  0.0  0.0  0.0  0.0   5.21


Statistics for PF_NUTSDS

                count      mean       std  min  25%  50%  75%    max
seafood_meal
0             94022.0  0.074733  0.539219  0.0  0.0  0.0  0.0  48.00
1              7709.0  0.053675  0.398895  0.0  0.0  0.0  0.0  13.14


Statistics for PF_LEGUMES

                count      mean       std  min  25%  50%  75%    max
seafood_meal
0             94022.0  0.208430  0.856666  0.0  0.0  0.0  0.0  27.05
1              7709.0  0.142331  0.678301  0.0  0.0  0.0  0.0   8.63
```

### Section 5: Dairy

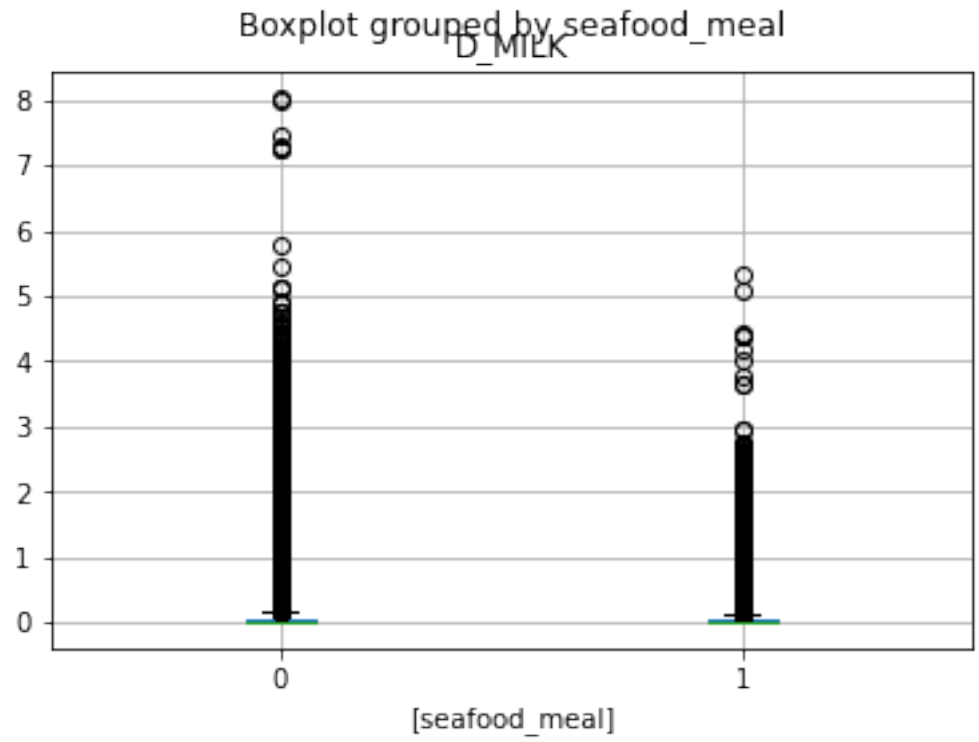
This section provides boxplots and density plots of the dairy FPED components in the seafood meal and non seafood meal groups. The code for seafood meal is 1 if meal contains seafood, and 0 if meal does not contain seafood.

[65]:
```
dairy = ['D_TOTAL', 'D_MILK', 'D_YOGURT', 'D_CHEESE']

for var in dairy:
```

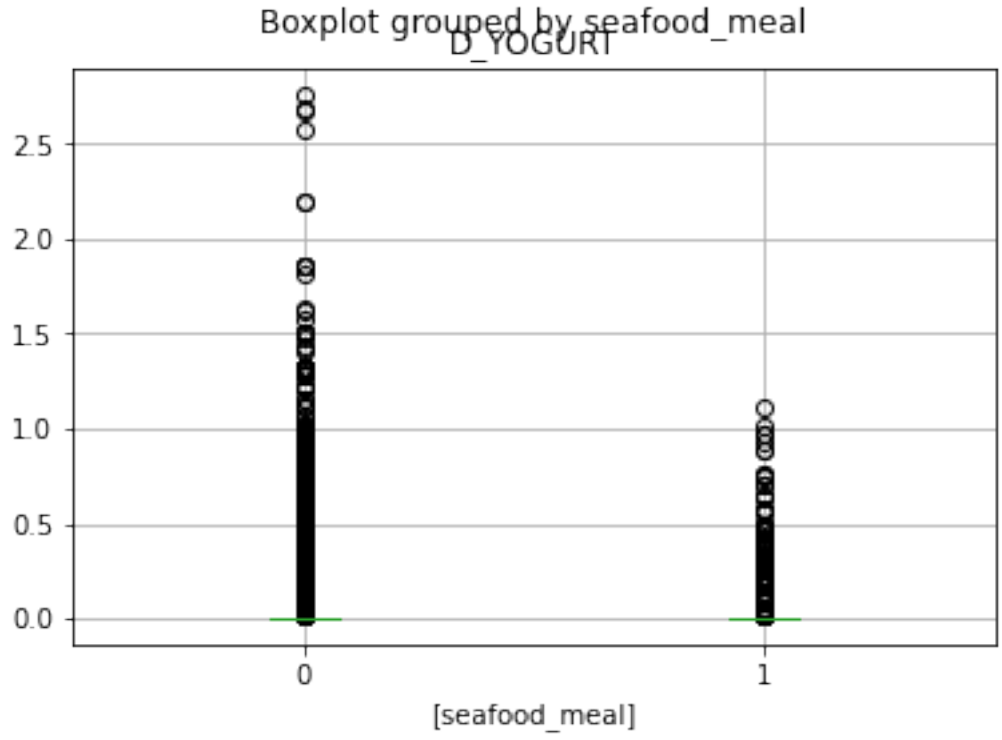
```
    z = df.boxplot(column=var,by=['seafood_meal'])
    plt.show(z)
    plt.clf()
```



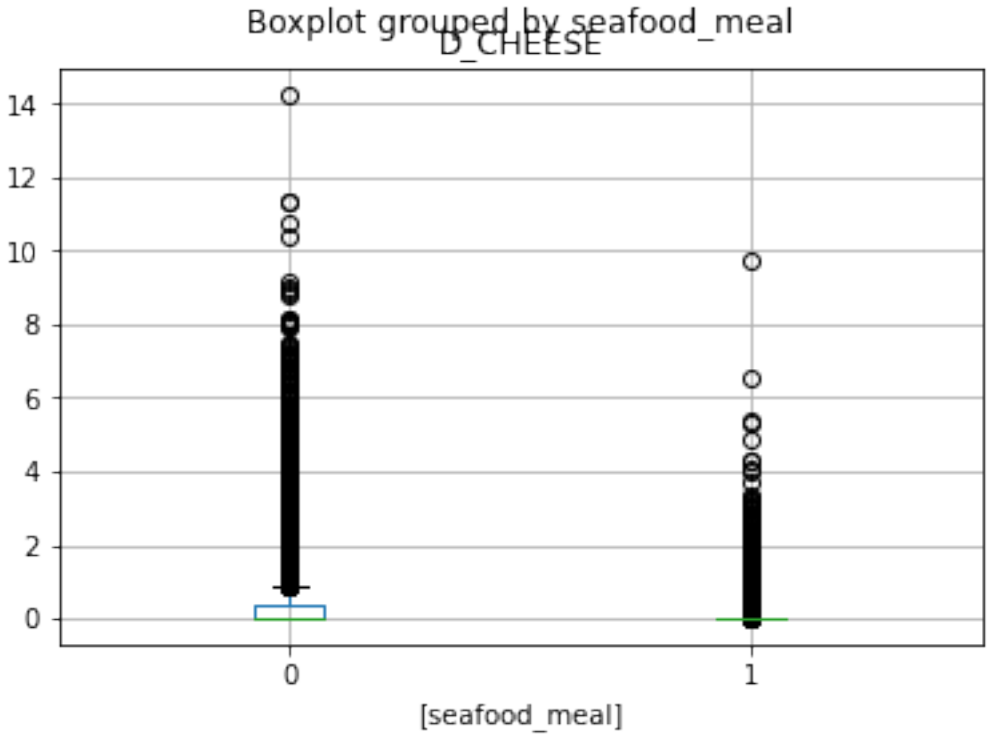

```
<Figure size 432x288 with 0 Axes>
```




```
<Figure size 432x288 with 0 Axes>
```

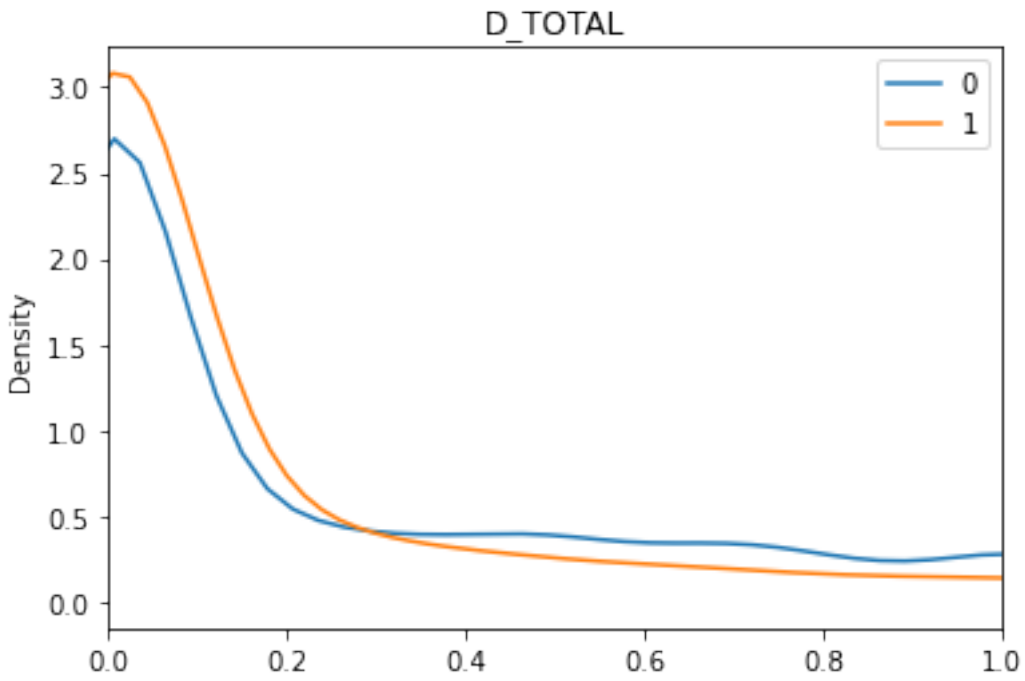


```
<Figure size 432x288 with 0 Axes>
```



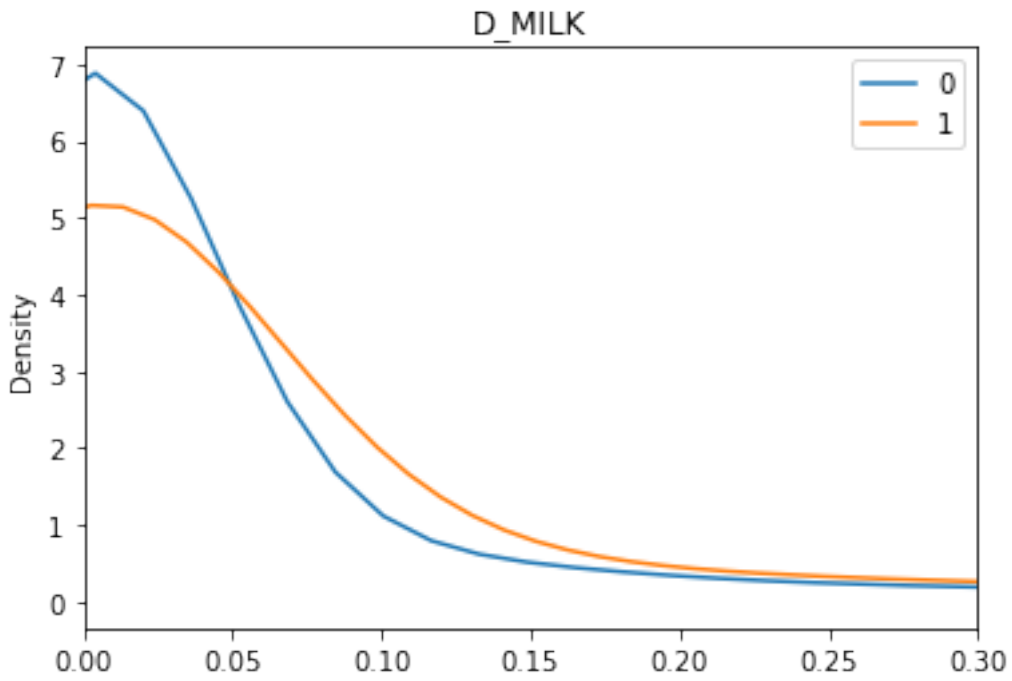

```
<Figure size 432x288 with 0 Axes>
```

[90]:
```
z = df.groupby('seafood_meal')[dairy[0]].plot.kde(title = dairy[0], legend='x')
plt.show(z[0].set_xlim(0, 1))
plt.clf()
```



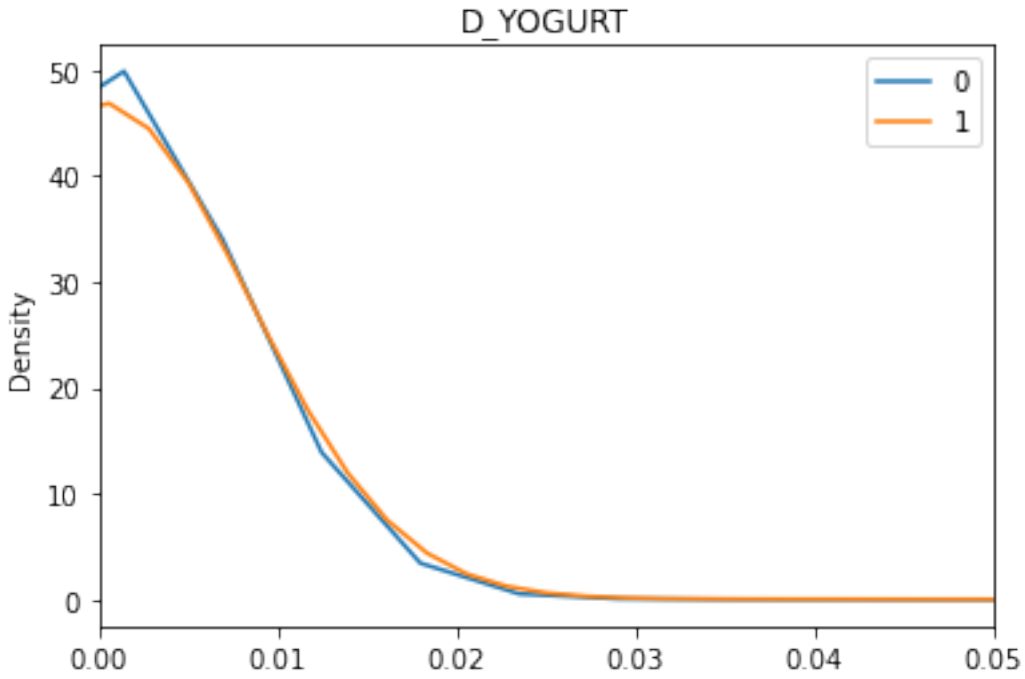

```
<Figure size 432x288 with 0 Axes>
```

[91]:
```
z = df.groupby('seafood_meal')[dairy[1]].plot.kde(title = dairy[1], legend='x')
plt.show(z[0].set_xlim(0, 0.3))
plt.clf()
```



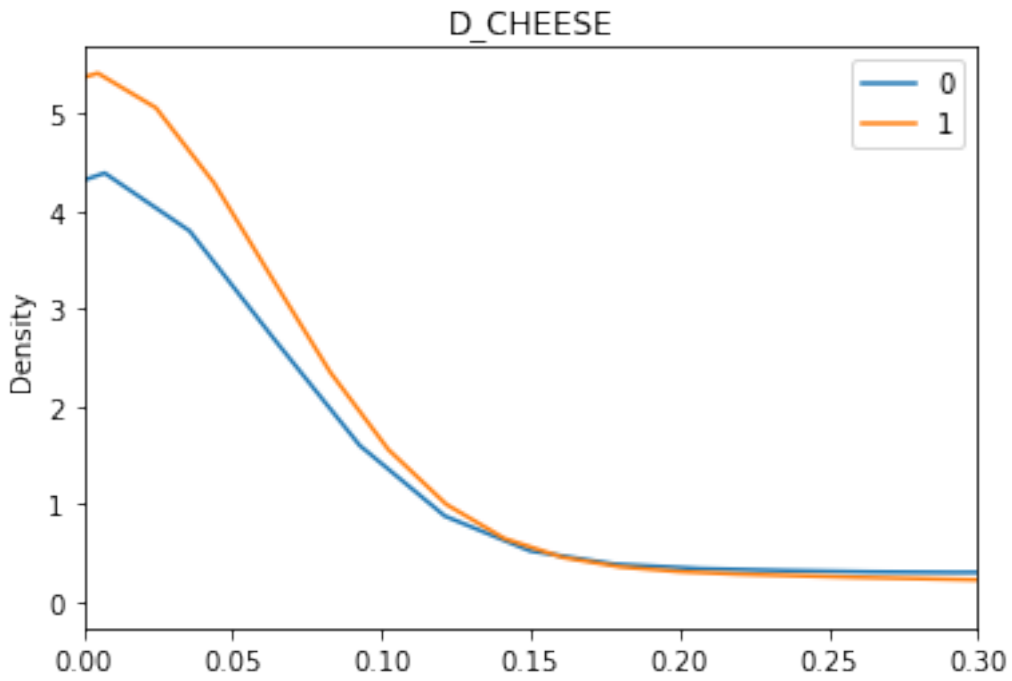

```
<Figure size 432x288 with 0 Axes>
```

[92]:
```
z = df.groupby('seafood_meal')[dairy[2]].plot.kde(title = dairy[2], legend='x')
plt.show(z[0].set_xlim(0, 0.05))
plt.clf()
```




```
<Figure size 432x288 with 0 Axes>
```

[93]:
```
z = df.groupby('seafood_meal')[dairy[3]].plot.kde(title = dairy[3], legend='x')
plt.show(z[0].set_xlim(0, 0.3))
plt.clf()
```




```
<Figure size 432x288 with 0 Axes>
```

[118]:
```
for var in dairy:
    z = df.groupby('seafood_meal')[var].describe()
    print("Statistics for "+var+'\n')
    print(z)
    print('\n')
```

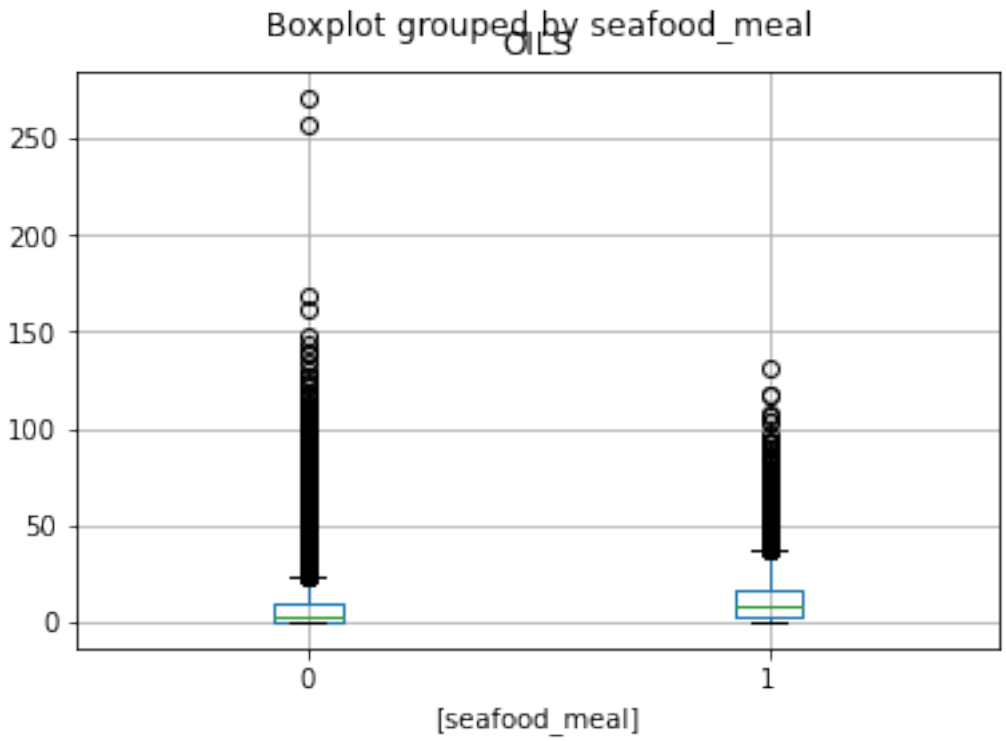
```
Statistics for D_TOTAL

                count      mean       std  min  25%   50%   75%    max
seafood_meal
0             94022.0  0.456065  0.728999  0.0  0.0  0.09  0.68  14.25
1              7709.0  0.238925  0.524462  0.0  0.0  0.00  0.23   9.76


Statistics for D_MILK

                count      mean       std  min  25%  50%   75%   max
seafood_meal
0             94022.0  0.166192  0.421444  0.0  0.0  0.0  0.06  8.04
1              7709.0  0.123777  0.356548  0.0  0.0  0.0  0.05  5.32
```

```
Statistics for D_YOGURT

                count      mean       std  min  25%  50%  75%   max
seafood_meal
0             94022.0  0.008638  0.076202  0.0  0.0  0.0  0.0  2.76
1              7709.0  0.004691  0.050101  0.0  0.0  0.0  0.0  1.11


Statistics for D_CHEESE

                count      mean       std  min  25%  50%   75%    max
seafood_meal
0             94022.0  0.278217  0.592725  0.0  0.0  0.0  0.36  14.25
1              7709.0  0.108280  0.364528  0.0  0.0  0.0  0.00   9.73
```

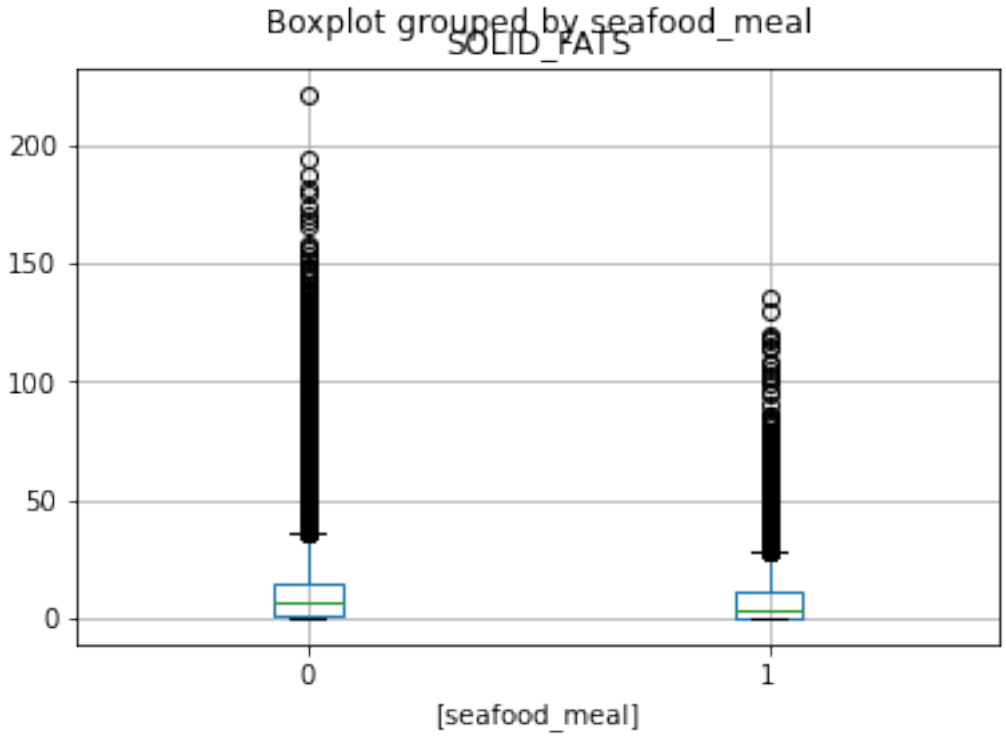
### Section 6: Others

This section provides boxplots and density plots of the other FPED components in the seafood meal and non seafood meal groups. The code for seafood meal is 1 if meal contains seafood, and 0 if meal does not contain seafood.

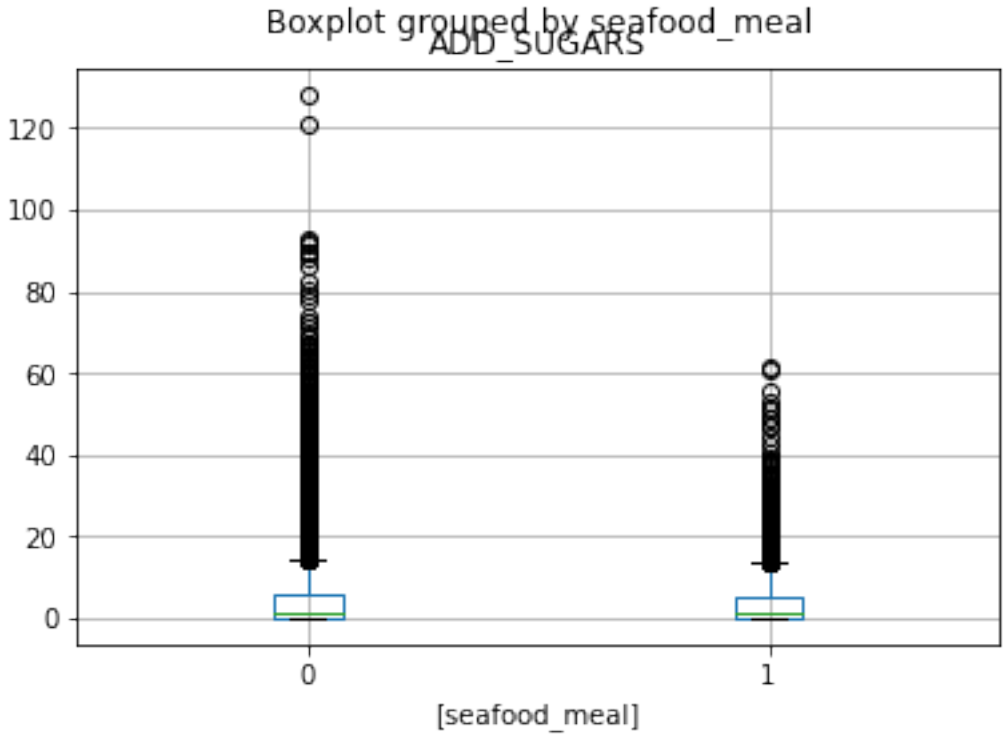
[67]:
```
other = ['OILS', 'SOLID_FATS', 'ADD_SUGARS', 'A_DRINKS']

for var in other:
    z = df.boxplot(column=var,by=['seafood_meal'])
    plt.show(z)
    plt.clf()
```



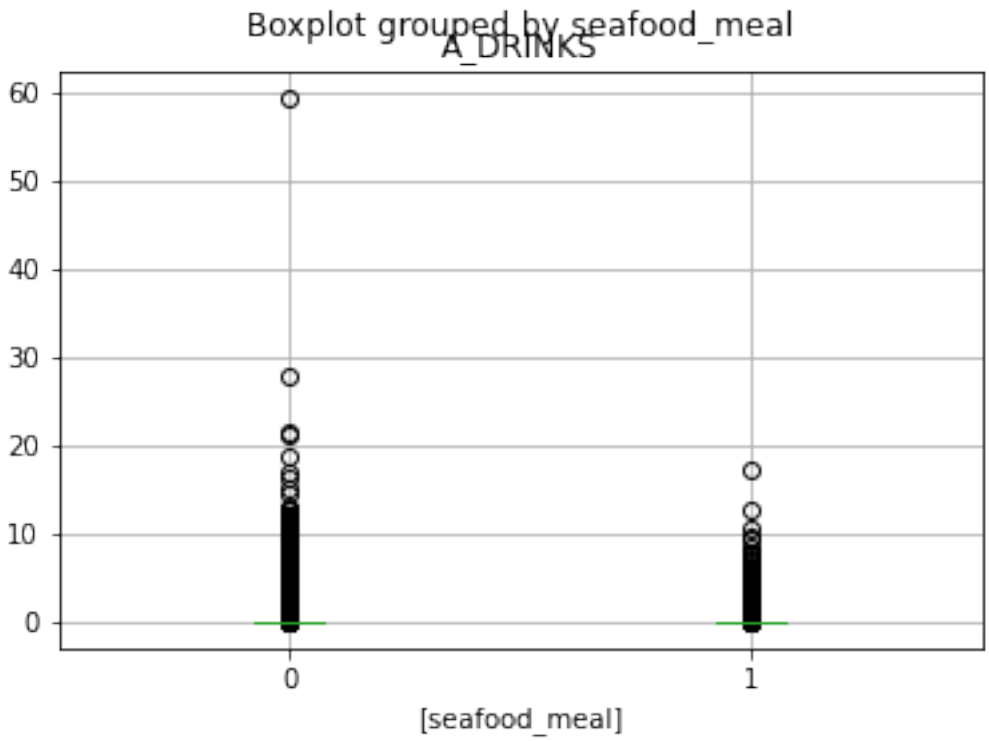

```
<Figure size 432x288 with 0 Axes>
```



```
<Figure size 432x288 with 0 Axes>
```

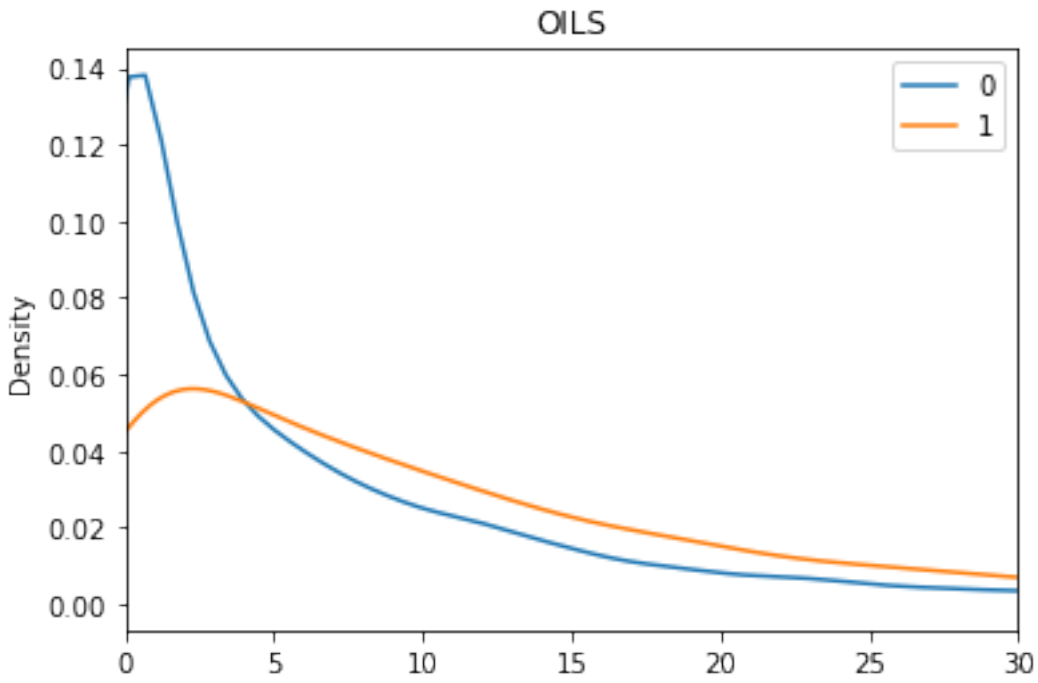



```
<Figure size 432x288 with 0 Axes>
```



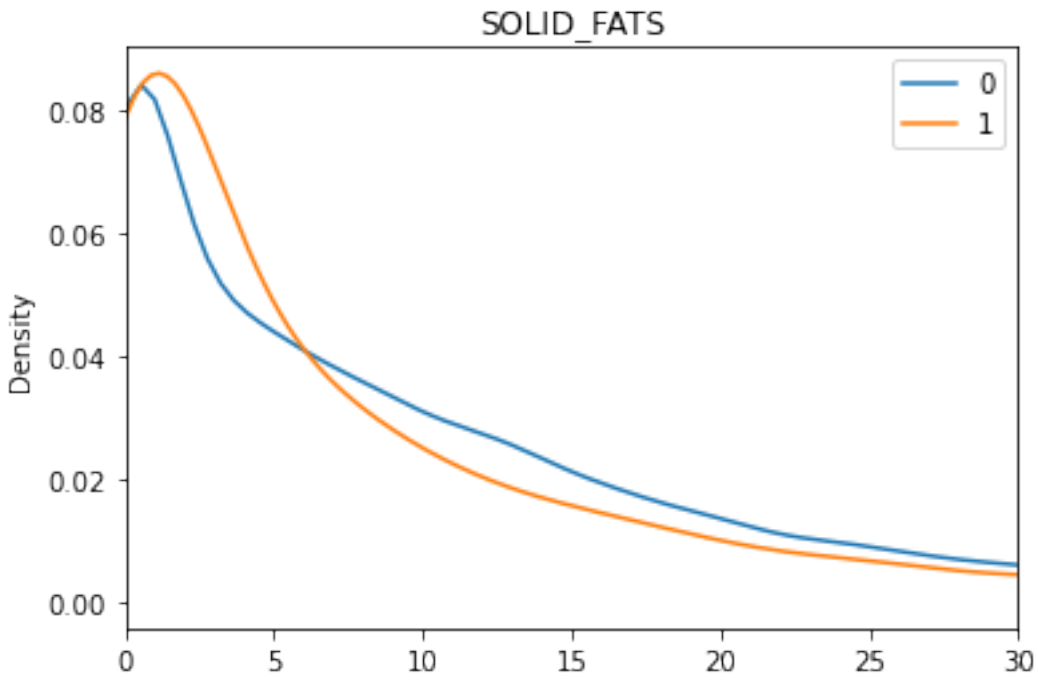

```
<Figure size 432x288 with 0 Axes>
```

[97]:
```
z = df.groupby('seafood_meal')[other[0]].plot.kde(title = other[0], legend='x')
plt.show(z[0].set_xlim(0, 30))
plt.clf()
```



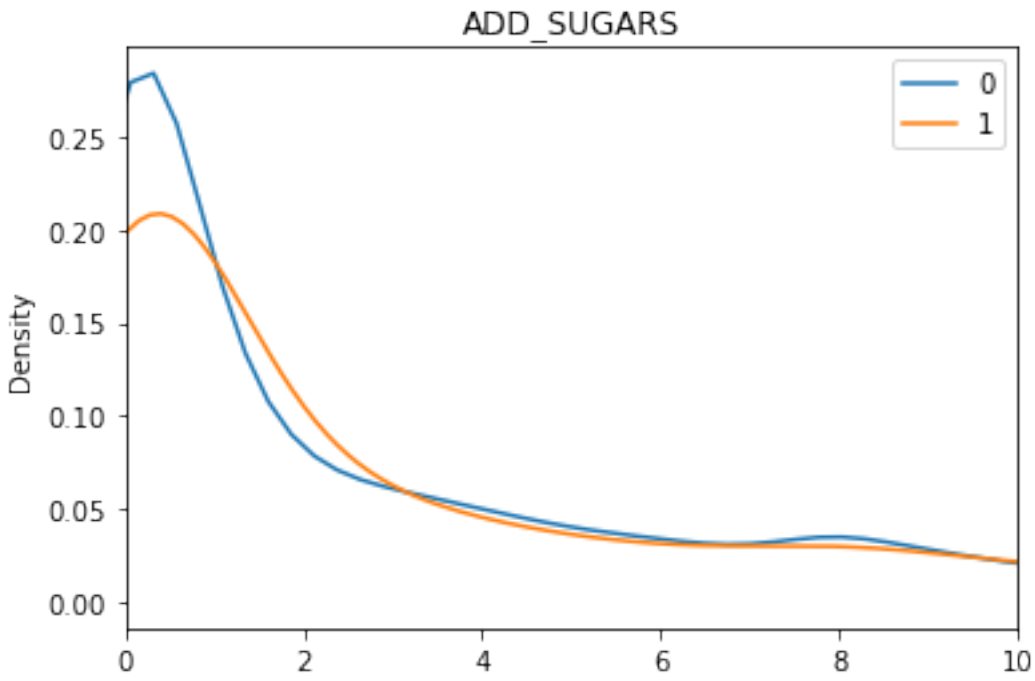

```
<Figure size 432x288 with 0 Axes>
```

[98]:
```
z = df.groupby('seafood_meal')[other[1]].plot.kde(title = other[1], legend='x')
plt.show(z[0].set_xlim(0, 30))
plt.clf()
```



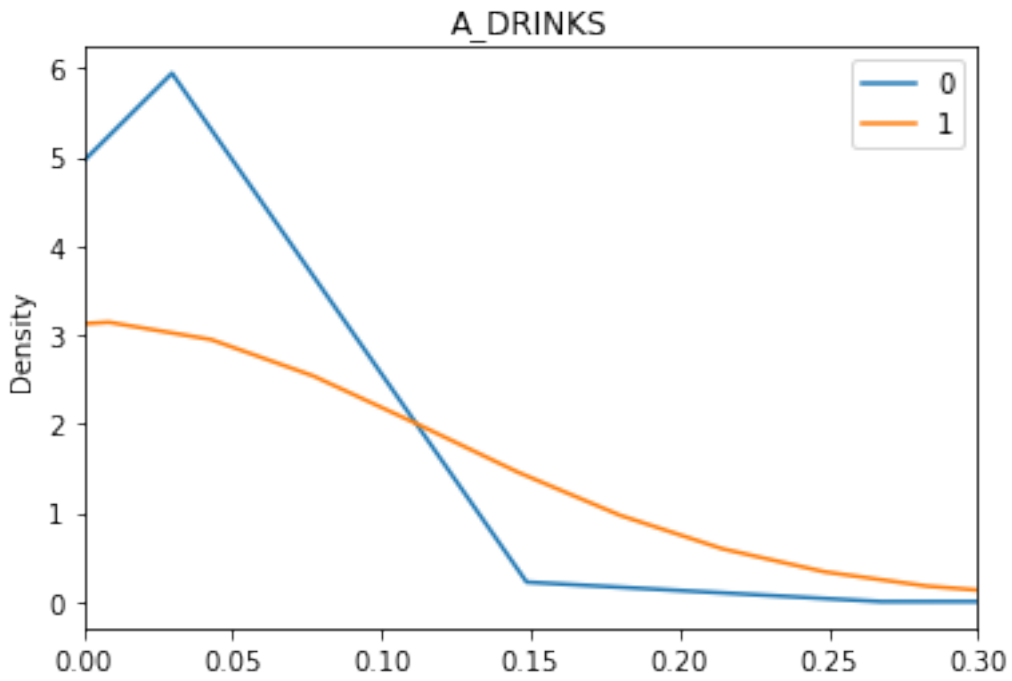

```
<Figure size 432x288 with 0 Axes>
```

[103]:
```
z = df.groupby('seafood_meal')[other[2]].plot.kde(title = other[2], legend='x')
plt.show(z[0].set_xlim(0, 10))
plt.clf()
```




```
<Figure size 432x288 with 0 Axes>
```

[104]:
```
z = df.groupby('seafood_meal')[other[3]].plot.kde(title = other[3], legend='x')
plt.show(z[0].set_xlim(0, 0.3))
plt.clf()
```




<Figure size 432x288 with 0 Axes>

[119]:
```
for var in other:
    z = df.groupby('seafood_meal')[var].describe()
    print("Statistics for "+var+'\n')
    print(z)
    print('\n')
```

Statistics for OILS

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 7.073091 | 10.301582 | 0.0 | 0.15 | 3.23 | 9.76 | 271.03 |
| 1 | 7709.0 | 11.971891 | 13.250389 | 0.0 | 2.77 | 7.98 | 16.77 | 131.06 |

Statistics for SOLID_FATS

| seafood_meal | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| 0 | 94022.0 | 10.777561 | 13.322028 | 0.0 | 1.32 | 6.66 | 15.15 | 221.47 |
| 1 | 7709.0 | 8.641013 | 12.728439 | 0.0 | 0.54 | 3.89 | 11.54 | 135.45 |

```
Statistics for ADD_SUGARS

                count      mean       std  min   25%   50%   75%     max
seafood_meal
0             94022.0  3.899550  5.858774  0.0  0.02  1.33  5.72  128.28
1              7709.0  3.862788  5.920019  0.0  0.05  1.17  5.55   61.64


Statistics for A_DRINKS

                count      mean       std  min  25%  50%  75%    max
seafood_meal
0             94022.0  0.069034  0.560395  0.0  0.0  0.0  0.0  59.40
1              7709.0  0.151065  0.701752  0.0  0.0  0.0  0.0  17.13
```

[ ]: