Exploratory plots and tables

Swen Kuh

2022-08-29

Exploratory plots

High school grads have high inclusion probability in the urine sample, but also have large weights. Supposedly if you're more likely to be in the sample, you've smaller weights?

For intermediate sodium level intake, they've a high inclusion probability, but small mean and median weights

```
summary(nhdata)
```

```
##       SEQN           age              ethnicity       gender
##  Min.   :109266   18-29: 973   MexAme      : 835   male  :3291
##  1st Qu.:113288   30-39:1128   OtherHisp   : 720   female:3492
##  Median :117161   40-49:1176   NonHispWhite:2293
##  Mean   :117157   50-59:1215   NonHispBlack:1818
##  3rd Qu.:121078   60-69:1414   NonHispAsian: 781
##  Max.   :124821   70-80: 877   Other       : 336
##             educ                 marital_status high_bp  phys_act overweight
##  less9th      : 450   married/living tgt:4000   0:4304   0:5062   0:3989
##  9-11thgrade  : 705   widowed/separated :1391   1:2479   1:1721   1:2794
##  HSgrad       :1633   never married     :1390
##  SomecollegeAA:2296   77                :   1
##  Collegegrad  :1699   99                :   1
##
##             dep_severity        BMI_range         alc_exc           diabetes
##  mild             :2266   underweight:  93   no       :1838   no        :5594
##  moderate         :3090   healthy    :1583   moderate :2471   yes       : 996
##  moderately severe: 933   overweight :2096   excessive:2474   borderline: 193
##  severe           : 494   obeseI     :1500
##                           obeseII    : 814
##                           obeseIII   : 697
##  trb_sleep         pov_level    smk2_exp smk_tobcig        sodium_lvl
##  0:4786    low income   :2068   0:3469   0:5054     low         : 494
##  1:1997    middle income: 975   1:3314   1:1729     intermediate:1472
##            high income  :3478                       high        :2234
##            refused      :  54                       very high   :2583
##            dont know    : 208
##
##   potassium_lvl    inclusion
##  low     :3743   Min.   :0.0000
##  adequate:2161   1st Qu.:0.0000
##  high    : 879   Median :0.0000
##                  Mean   :0.3273
##                  3rd Qu.:1.0000
##                  Max.   :1.0000
```

```
summary(nhsub)
```

```
##       SEQN             wts             age             ethnicity      gender
##  Min.   :109266   Min.   :     0   18-29:302   MexAme      :283   male  :1081
##  1st Qu.:113222   1st Qu.: 29693   30-39:349   OtherHisp   :228   female:1143
##  Median :117048   Median : 52226   40-49:384   NonHispWhite:733
##  Mean   :117043   Mean   : 91641   50-59:411   NonHispBlack:608
##  3rd Qu.:120984   3rd Qu.:100539   60-69:498   NonHispAsian:256
##  Max.   :124821   Max.   :955677   70-80:280   Other       :116
##             educ                marital_status high_bp  phys_act overweight
##  less9th      :138   married/living tgt:1311   0:1385   0:1665   0:1321
##  9-11thgrade  :241   widowed/separated : 475   1: 839   1: 559   1: 903
##  HSgrad       :582   never married     : 436
##  SomecollegeAA:722   77                :   1
##  Collegegrad  :541   99                :   1
##
##             dep_severity       BMI_range        alc_exc          diabetes
##  mild             :765   underweight: 32   no       :589   no        :1827
##  moderate         :997   healthy    :514   moderate :810   yes       : 331
##  moderately severe:303   overweight :698   excessive:825   borderline:  66
##  severe           :159   obeseI     :479
##                          obeseII    :276
##                          obeseIII   :225
##  trb_sleep         pov_level    smk2_exp smk_tobcig        sodium_lvl
##  0:1544    low income   : 662   0:1129   0:1659     low         :162
##  1: 680    middle income: 318   1:1095   1: 565     intermediate:510
##            high income  :1151                       high        :681
##            refused      :  22                       very high   :871
##            dont know    :  71
##
##   potassium_lvl    inclusion
##  low     :1200   Min.   :0.0000
##  adequate: 707   1st Qu.:1.0000
##  high    : 317   Median :1.0000
##                  Mean   :0.9982
##                  3rd Qu.:1.0000
##                  Max.   :1.0000
```

```
# checking inclusino of education variable
# HS grad have higher inclusion
t1 = nhdata %>%
  group_by(educ) %>%
  summarise(mean_incl_fulldat = mean(inclusion))

t2 = nhsub %>%
  group_by(educ) %>%
  summarise(mean_weights_subdat = mean(wts),
            median_weights_subdat = median(wts)) %>%
  mutate(prob_incl = 1/median_weights_subdat)

t3 = nhdata %>%
  group_by(sodium_lvl) %>%
  summarise(mean_incl_fulldat = mean(inclusion))
```

```
t4 = nhsub %>%
  group_by(sodium_lvl) %>%
  summarise(mean_weights_subdat = mean(wts),
            median_weights_subdat = median(wts)) %>%
  mutate(prob_incl = 1/median_weights_subdat)


t5 = nhdata %>%
  group_by(potassium_lvl) %>%
  summarise(mean_incl_fulldat = mean(inclusion))

t6 = nhsub %>%
  group_by(potassium_lvl) %>%
  summarise(mean_weights_subdat = mean(wts),
            median_weights_subdat = median(wts)) %>%
  mutate(prob_incl = 1/median_weights_subdat)

t7 = nhdata %>%
  group_by(age) %>%
  summarise(mean_incl_fulldat = mean(inclusion))

t8 = nhsub %>%
  group_by(age) %>%
  summarise(mean_weights_subdat = mean(wts),
            median_weights_subdat = median(wts)) %>%
  mutate(prob_incl = 1/median_weights_subdat)

t9 = nhdata %>%
  group_by(dep_severity) %>%
  summarise(mean_incl_fulldat = mean(inclusion))

t10 = nhsub %>%
  group_by(dep_severity) %>%
  summarise(mean_weights_subdat = mean(wts),
            median_weights_subdat = median(wts)) %>%
  mutate(prob_incl = 1/median_weights_subdat)


left_join(t1,t2) # educ
```

```
## # A tibble: 5 x 5
##   educ          mean_incl_fulldat mean_weights_sub~ median_weights_su~ prob_incl
##   <fct>                     <dbl>             <dbl>              <dbl>     <dbl>
## 1 less9th                   0.307             44471.             35004. 0.0000286
## 2 9-11thgrade               0.340             62951.             42814. 0.0000234
## 3 HSgrad                    0.356            102889.             65447. 0.0000153
## 4 SomecollegeAA             0.314             82955.             50673. 0.0000197
## 5 Collegegrad               0.317            115947.             56996. 0.0000175
```

```
left_join(t3,t4) # sodium_lvl
```

```
## # A tibble: 4 x 5
##   sodium_lvl   mean_incl_fulldat mean_weights_subd~ median_weights_su~ prob_incl
##   <fct>                    <dbl>              <dbl>              <dbl>     <dbl>
```

```
## 1 low                      0.328            71106.            48673. 0.0000205
## 2 intermediate             0.346            80437.            46744. 0.0000214
## 3 high                     0.304            93561.            51903. 0.0000193
## 4 very high                0.337           100520.            57082. 0.0000175
```

```
left_join(t5,t6) # potassium_lvl
```

```
## # A tibble: 3 x 5
##   potassium_lvl mean_incl_fulldat mean_weights_sub~ median_weights_su~ prob_incl
##   <fct>                     <dbl>            <dbl>              <dbl>     <dbl>
## 1 low                       0.320            87454.             49049. 0.0000204
## 2 adequate                  0.326            93468.             54609. 0.0000183
## 3 high                      0.361           103415.             59233. 0.0000169
```

```
left_join(t7,t8) # age
```

```
## # A tibble: 6 x 5
##   age   mean_incl_fulldat mean_weights_subdat median_weights_subdat prob_incl
##   <fct>             <dbl>               <dbl>                <dbl>     <dbl>
## 1 18-29             0.309             116347.               72241. 0.0000138
## 2 30-39             0.308             105242.               70180. 0.0000142
## 3 40-49             0.326              92954.               57686. 0.0000173
## 4 50-59             0.338              89673.               47694. 0.0000210
## 5 60-69             0.352              75345.               32562. 0.0000307
## 6 70-80             0.319              78114.               47757. 0.0000209
```

```
left_join(t9,t10) # dep_severity
```

```
## # A tibble: 4 x 5
##   dep_severity   mean_incl_fulldat mean_weights_sub~ median_weights_s~ prob_incl
##   <fct>                      <dbl>            <dbl>             <dbl>     <dbl>
## 1 mild                       0.337            88230.            48431. 0.0000206
## 2 moderate                   0.322            95162.            54725. 0.0000183
## 3 moderately se~             0.325            94809.            55162. 0.0000181
## 4 severe                     0.320            79935.            49015. 0.0000204
```

```
# checking for interaction effects
with(nhdata, tapply(as.numeric(high_bp), list(potassium_lvl, sodium_lvl), sum))
```

```
##          low intermediate high very high
## low      655         1715 1878       879
## adequate  10          313  986      1644
## high       1           40  183       958
```

```
with(nhdata, tapply(as.numeric(high_bp), list(gender, marital_status), sum))
```

```
##        married/living tgt widowed/separated never married 77 99
## male                 2938              775           819  2 NA
## female               2474             1338           914 NA  2
```

```
with(nhdata, tapply(as.numeric(high_bp), list(gender, dep_severity), sum))
```

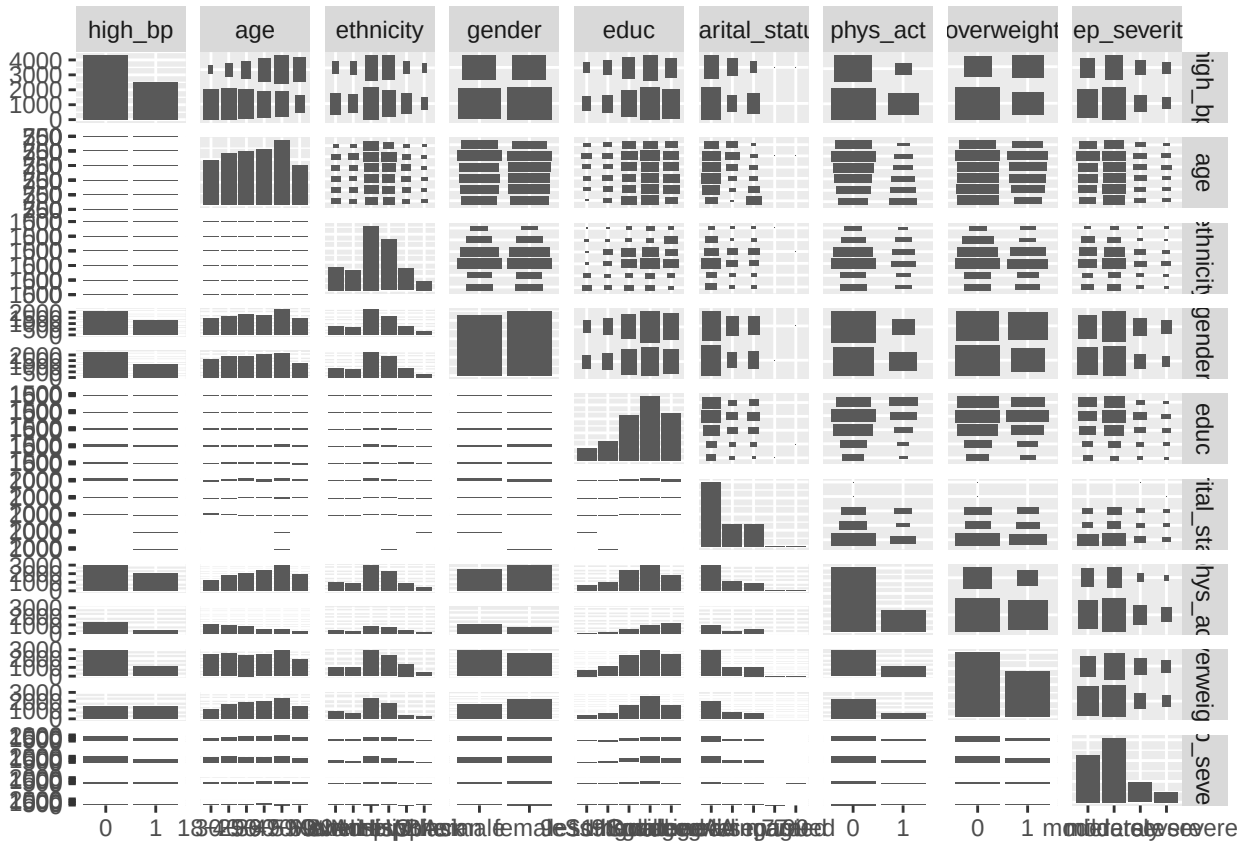
```
##        mild moderate moderately severe severe
## male   1703     1996               556    279
## female 1285     2205               781    457
```

```
summary(nhdata)
```

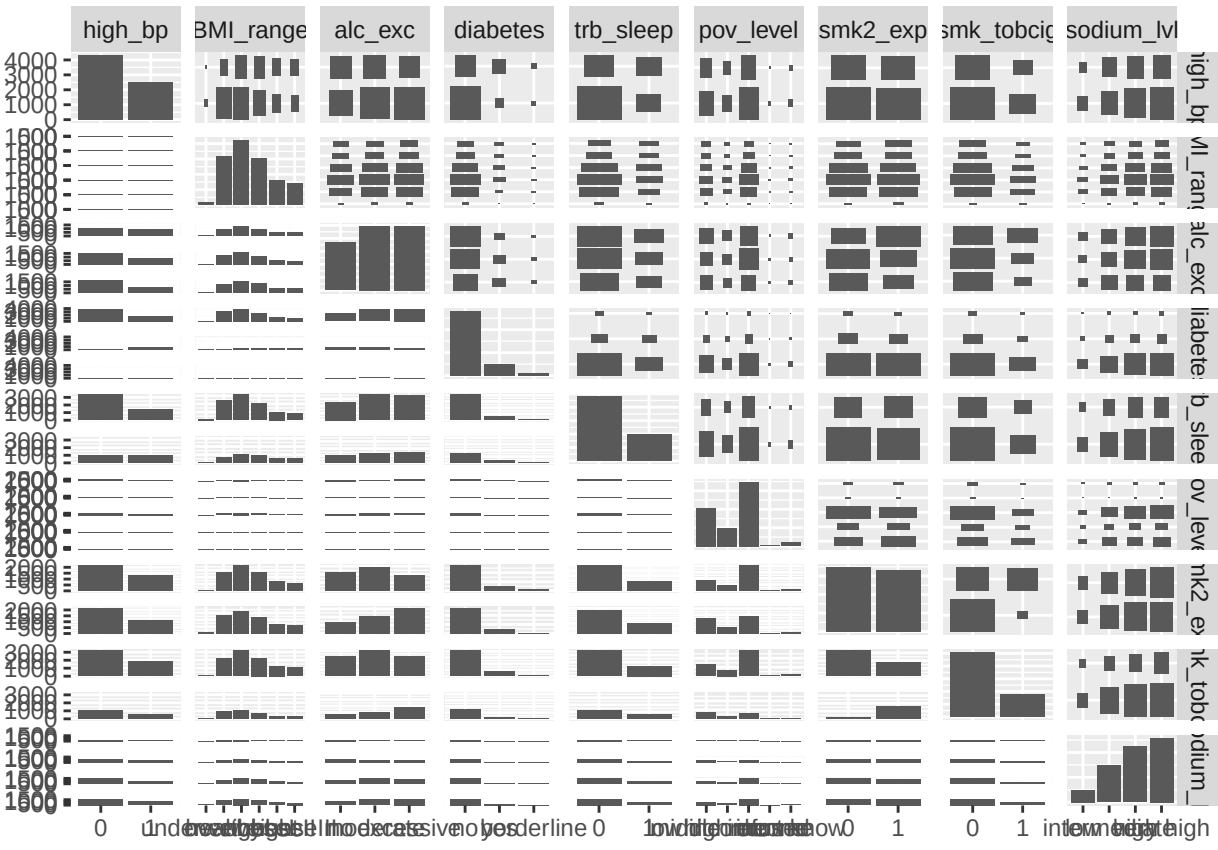
```
##       SEQN           age              ethnicity       gender
```

```
##  Min.   :109266   18-29: 973   MexAme      : 835   male  :3291
##  1st Qu.:113288   30-39:1128   OtherHisp   : 720   female:3492
##  Median :117161   40-49:1176   NonHispWhite:2293
##  Mean   :117157   50-59:1215   NonHispBlack:1818
##  3rd Qu.:121078   60-69:1414   NonHispAsian: 781
##  Max.   :124821   70-80: 877   Other       : 336
##             educ                  marital_status high_bp  phys_act overweight
##  less9th      : 450   married/living tgt:4000   0:4304   0:5062   0:3989
##  9-11thgrade  : 705   widowed/separated :1391   1:2479   1:1721   1:2794
##  HSgrad       :1633   never married     :1390
##  SomecollegeAA:2296   77                :   1
##  Collegegrad  :1699   99                :   1
##
##             dep_severity        BMI_range         alc_exc           diabetes
##  mild             :2266   underweight:  93   no       :1838   no         :5594
##  moderate         :3090   healthy    :1583   moderate :2471   yes        : 996
##  moderately severe: 933   overweight :2096   excessive:2474   borderline: 193
##  severe           : 494   obeseI     :1500
##                           obeseII    : 814
##                           obeseIII   : 697
##  trb_sleep         pov_level    smk2_exp smk_tobcig        sodium_lvl
##  0:4786    low income   :2068   0:3469   0:5054     low         : 494
##  1:1997    middle income: 975   1:3314   1:1729     intermediate:1472
##            high income  :3478                       high        :2234
##            refused      :  54                       very high   :2583
##            dont know    : 208
##
##   potassium_lvl    inclusion
##  low     :3743   Min.   :0.0000
##  adequate:2161   1st Qu.:0.0000
##  high    : 879   Median :0.0000
##                  Mean   :0.3273
##                  3rd Qu.:1.0000
##                  Max.   :1.0000
```

```
ggpairs(nhdata[,c(7,2:6,8:10)])
```




```
ggpairs(nhdata[,c(7,11:18)])
```


high_bp
BMI_range
alc_exc
diabetes
trb_sleep
pov_level
smk2_exp
smk_tobcig
sodium_lvl
0
1
underweight
normal
overweight
obese
Inadequate
Adequate
excessive
no
yes
borderline
low
high
intermediate