Obesity Group Assignment

ROBERT GEORGE

3/11/2020

Due March 23 This assignment was orignally a group assignment. I have made it into an individual assignment as so many students were away from class the last two class periods.

Data

The combined and transformed data is in the file `obesity.csv`.

```
obesity = read_csv("obesity.csv") %>%
  mutate(zip = as.character(zip)) %>%
  mutate(obese_pop = round((obese/bmi) * pop)) %>%
  select(zip,obese,bmi,pop,sex,age,starts_with("pct_"))
```

Each row is for a single zip code / sex / age cohort, except that the education variables `pct_m_bach` and `pct_f_bach` are numerical values for the entire zipcode, not the sex/age cohort.

Recall these variable definitions.

- `zip` = Zip code
- `obese` = number of sampled obese individuals
- `bmi` = number of sampled individuals
- `pop` = population of zip/sex/age cohort according to ACS from 2013 - 2015
- `sex` = female or male
- `age` = one of five age ranges
- `pct_m_bach` = % of men aged 25+ with a bachelors degree
- `pct_f_bach` = % of men aged 25+ with a bachelors degree

Questions

1. Find the approximate fraction of all Wisconsin adults by sex aged 25+ that have a bachelors degree. Do this by excluding children, summing separately by sex the population (using `pop`) within each zip code, multiplying these population totals by the proportion of women or men, respectively, with at college degree, and summing. This calculation includes people aged 18-24 which might bias the results.

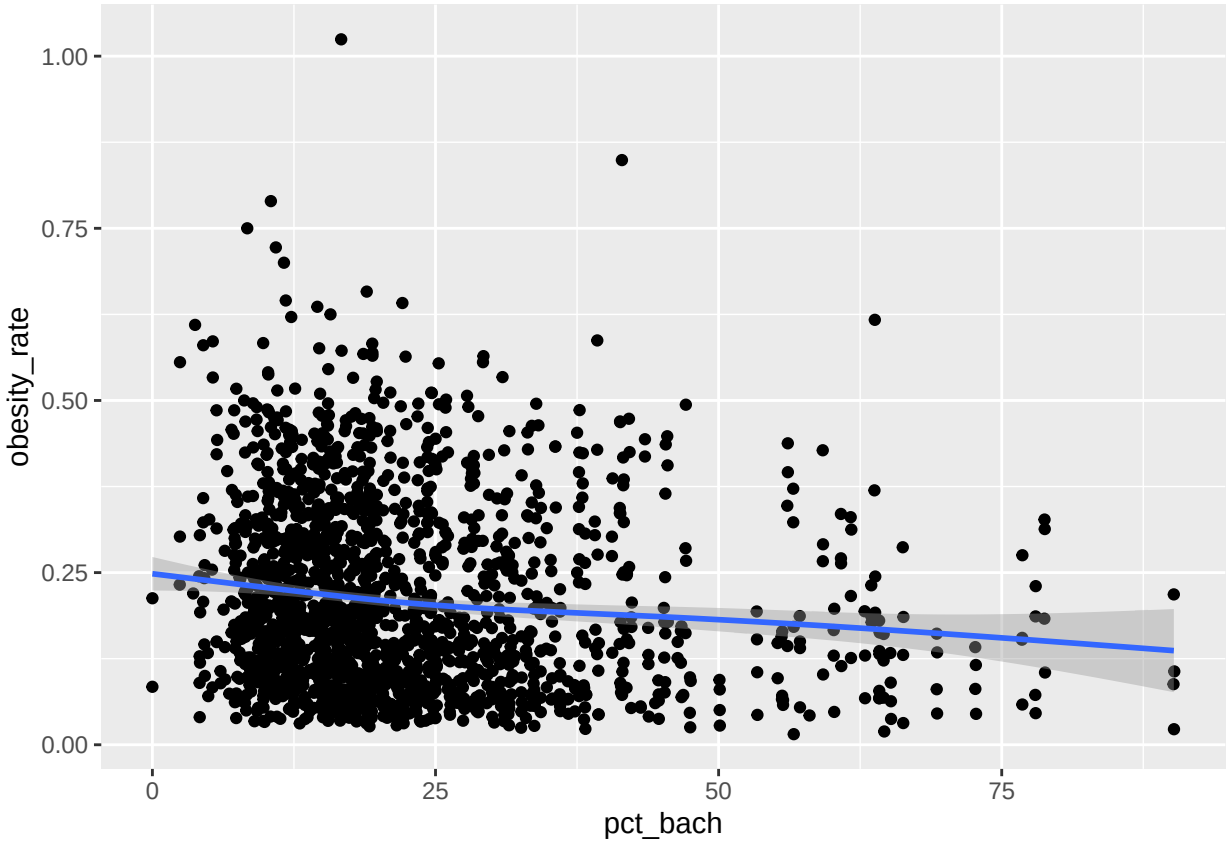
```
q_1 = obesity%>%
  select(pop, sex, zip, age, pct_f_bach, pct_m_bach, obese)%>%
  filter(age!="05-17")%>%
  group_by(zip, sex, pct_m_bach, pct_f_bach)%>%
  mutate(popzip= sum(pop),
         pct_bach = case_when(sex == "female" ~ pct_f_bach,
```

```
                              sex == "male" ~ pct_m_bach),
         bach_by_zip = popzip * (pct_bach/100),
         sum_bach = sum(bach_by_zip))%>%
  ungroup()

q_1%>%
  select(sum_bach)
```

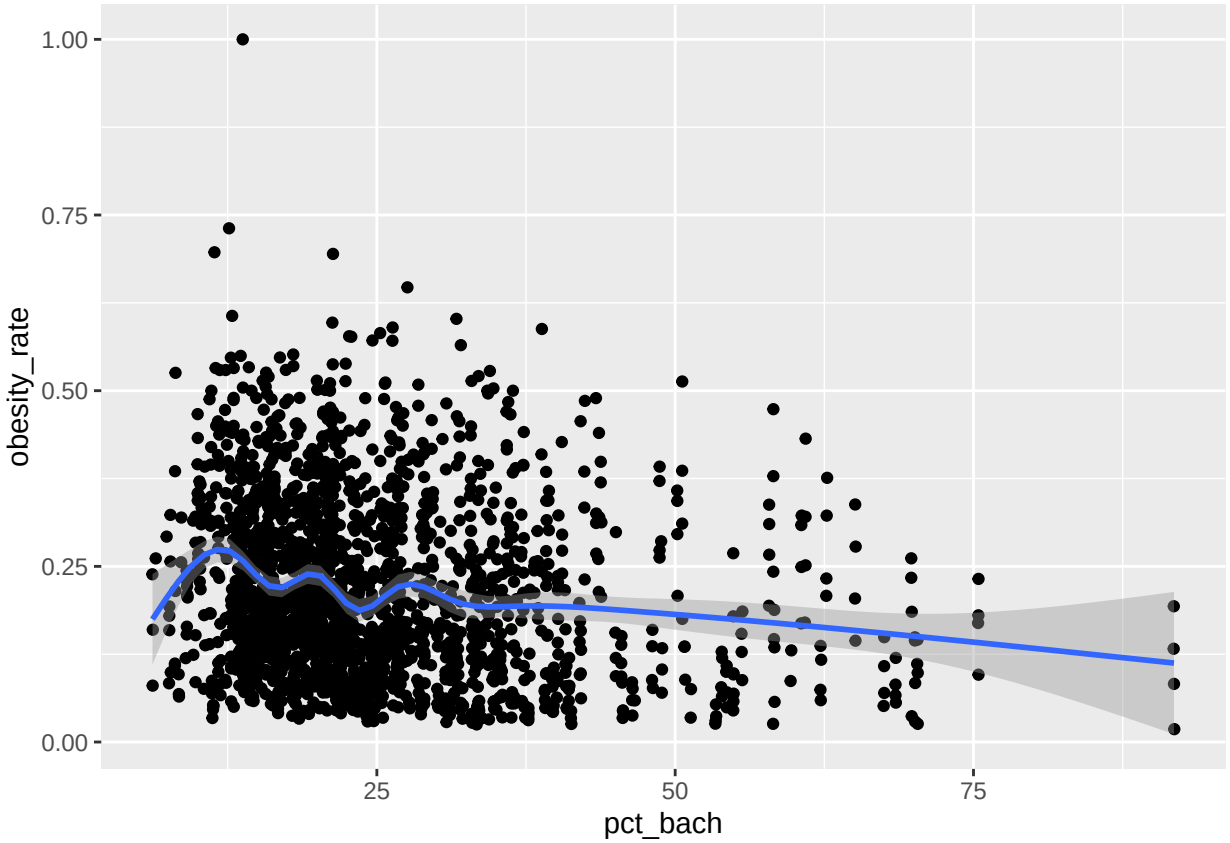
```
## # A tibble: 6,192 x 1
##    sum_bach
##       <dbl>
##  1   668.
##  2   993.
##  3     9.02
##  4  1334.
##  5 17332.
##  6   713.
##  7   833.
##  8  2094.
##  9   636.
## 10 13920.
## # ... with 6,182 more rows
```

2. For each zip code and sex, calculate the adult obesity rate (across all adult age groups). Make a scatter plot with percentage of adults aged 25+ with a college degree on the x axis, the obesity rate on the y axis, and separate panels for females and males with one point for each zip code. Add a trend line to the plot.

```
p_2 = q_1%>%
  select(zip, sex, obese, pop, age, pct_bach)%>%
  filter(age!="05-17")%>%
  group_by(zip, sex)%>%
  mutate(obesity_rate = obese/pop)

p_2_f = p_2%>%
  select(age, pct_bach, sex, zip, obesity_rate)%>%
  filter(sex!="female")%>%
  group_by(zip)%>%
  drop_na()

p_2_m = p_2%>%
  select(age, pct_bach, sex, zip, obesity_rate)%>%
  filter(sex=="female")%>%
  group_by(zip)%>%
  drop_na()

  ggplot(data=p_2_f)+
  geom_jitter(mapping = aes(x=pct_bach, y=obesity_rate))+geom_smooth(mapping = aes(x=pct_bach, y=obesity
```




```
ggplot(data=p_2_m)+
geom_jitter(mapping = aes(x=pct_bach, y=obesity_rate))+geom_smooth(mapping = aes(x=pct_bach, y=obesity
```

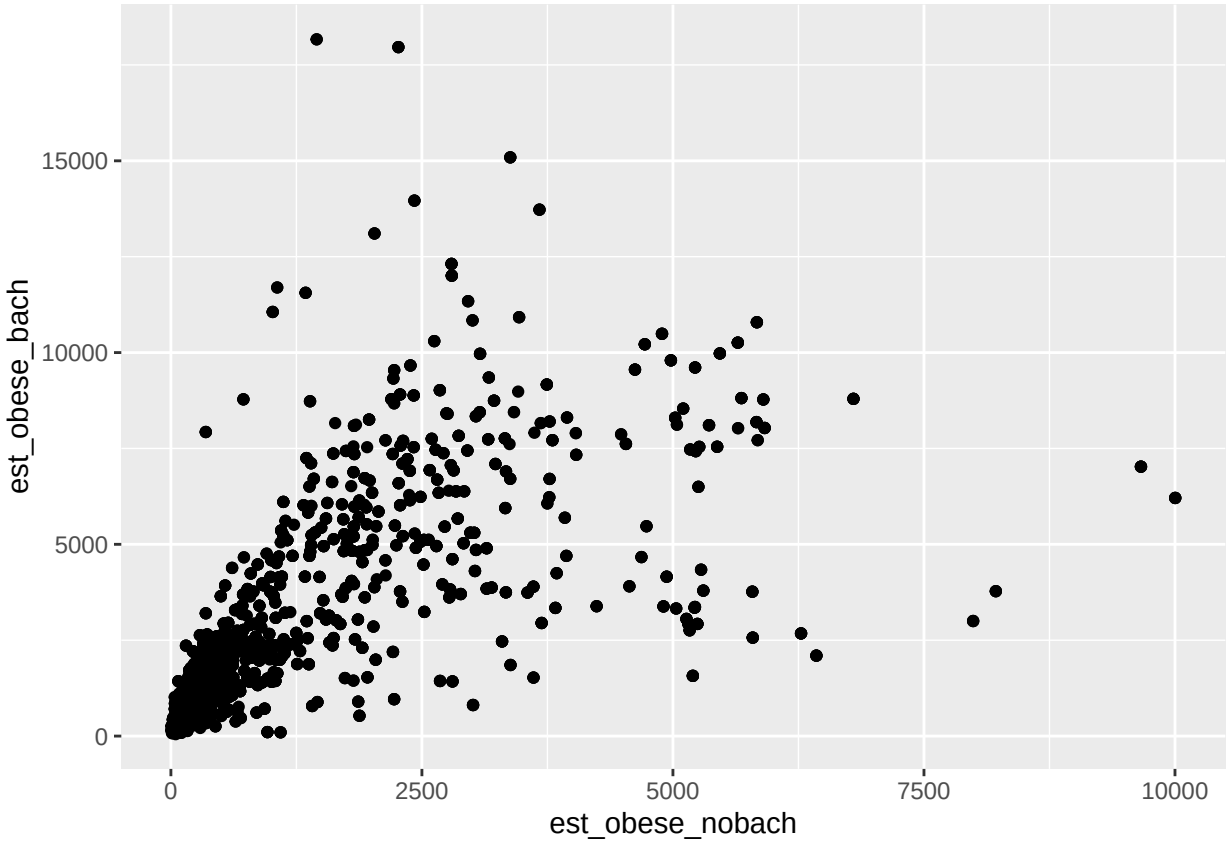


3. Make the assumption that within zip codes and within each sex, obesity prevalence and education level (at least bachelors or not) are independent (probably not a very accurate assumption). Then, for each zip code and sex, estimate the number of adults in each of the four categories: obese and bachelors, obese and no bachelors, not obese and bachelors, not obese and no bachelors. Total these values over zip codes for the entire state and make estimates of the obesity rates among each sex/education combination. Display this data in a table and in a plot.

Be careful how you treat missing data. Ignore the inclusion of people aged 18-24 in the calculations.

```
p_3 = obesity%>%
  select(pop, sex, zip, age, pct_f_bach, pct_m_bach, obese)%>%
  filter(age!="18-34")%>%
  group_by(zip, sex, pct_m_bach, pct_f_bach)%>%
  mutate(popzip= sum(pop),
         pct_bach = case_when(sex == "female" ~ pct_f_bach,
                              sex == "male" ~ pct_m_bach),
         bach_by_zip = popzip * (pct_bach/100),
         sum_bach = sum(bach_by_zip))
p_3=p_3%>%
  select(zip, sex, obese, pop, age, pct_bach, popzip)%>%
  filter(age!="18-34")%>%
  group_by(zip, sex)%>%
  mutate(obesity_rate = obese/pop)%>%
  drop_na()
```

```r
p_3 = p_3%>%
  select(popzip, zip, sex, obese, pop, age, pct_bach, obesity_rate, popzip)%>%
  group_by(zip, sex)%>%
  mutate(est_obese_bach = (pop*obesity_rate)*(pct_bach/100),
         est_obese_nobach = (pop*obesity_rate)*(1-(pct_bach/100)),
         est_noobese_bach = (pop - (pop*obesity_rate))*(pct_bach/100),
         est_noobese_nobach = (pop - (pop*obesity_rate))*(1-(pct_bach/100)),
         est_obese_bach_byZip = (popzip*obesity_rate)*(pct_bach/100),
         est_obese_nobach_byZip = (popzip*obesity_rate)*(1-(pct_bach/100)),
         est_noobese_bach_byZip = (popzip-(popzip*obesity_rate))*(pct_bach/100),
         est_noobese_nobach_byZip=(popzip-(popzip*obesity_rate))*(1-(pct_bach/100)))

p_3 = p_3%>%
  select(popzip, pct_bach, sex, zip, est_obese_bach, est_obese_nobach, est_noobese_bach, est_noobese_no
  group_by(sex, zip)%>%
  mutate(est_obese_nobach = popzip*(pct_bach/100),
         est_obese_bach = popzip*(1-(pct_bach/100))
         )




p_3_table=p_3%>%
  select(sex, zip, est_obese_nobach, est_obese_bach)%>%
  group_by(sex, zip)
ggplot(data = p_3_table)+geom_point(mapping=aes(x=est_obese_nobach, y=est_obese_bach))
```




```
p_3_table
```

```
## # A tibble: 3,526 x 4
## # Groups:   sex, zip [1,134]
##    sex    zip   est_obese_nobach est_obese_bach
##    <chr>  <chr>            <dbl>          <dbl>
##  1 female 53002             237.           695.
##  2 female 53010             528.          2355.
##  3 female 53012            3610.          3896.
##  4 female 53014             690.          2564.
##  5 female 53017            1041.          1437.
##  6 female 53021             373.          1336.
##  7 female 53022            3039.          4854.
##  8 female 53024            2781.          3824.
##  9 female 53027            2380.          6152.
## 10 female 53032             317.          1537.
## # ... with 3,516 more rows
```