STAT 346 Final Project
San Francisco AirBnB Analysis

Hailey Han, Josh Kang and Jihong Lee

Due Monday, April 27

Question 1.

(a) Frequemcy table for self-reported race variable.

```
# Read in brithwt data
library(MASS)
table(birthwt$race)
```

```
##
##  1  2  3
## 96 26 67
```

```
#create indicator variables
birthwt$race.cauc = ifelse(birthwt$race=="1",1,0)
birthwt$race.black = ifelse(birthwt$race=="2",1,0)
birthwt$race.other = ifelse(birthwt$race=="3",1,0)
```

(b)

```
summary(lm(bwt~age+lwt+smoke+race.black+race.other+ptl, data=birthwt))$coefficients
```

```
##                 Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 2853.8412272 321.131655  8.88682627 6.046071e-16
## age           -0.4701479   9.874866 -0.04761056 9.620788e-01
## lwt            3.7001244   1.751596  2.11243042 3.601351e-02
## smoke       -373.5910044 111.341333 -3.35536674 9.646749e-04
## race.black  -503.3695384 156.931424 -3.20757644 1.581704e-03
## race.other  -387.7938547 119.702151 -3.23965653 1.422769e-03
## ptl         -131.1114288 104.328347 -1.25671912 2.104659e-01
```

```
#anova(lm(bwt~age+lwt+smoke+race.black+race.other+ptl, data=birthwt))
```

Interpretation:

- Adjusted for other variables, having a self-reported race of `black` compared to the refrence group (`cauc`) decreases the expected baby's birth weight by 503.3695384 g
- Adjusted for other variables, being the race of `other` compared to the reference group, decreases expected baby weight by 387.7938547 g.
- Adjusted for other variables, being a smoker decreases the expected baby's weight by 373.5910044 g.

(c) Now we look at non-Caucasian vs Caucasian.

```r
# merge variables
birthwt$race.noncauc = ifelse(birthwt$race!="1",1,0)
summary(lm(bwt~age+lwt+smoke+race.noncauc+ptl, data=birthwt))$coefficients
```

```
##                   Estimate Std. Error     t value     Pr(>|t|)
## (Intercept)  2893.3394635 315.584655  9.16818805 9.916942e-17
## age             0.1148563   9.824741  0.01169052 9.906853e-01
## lwt             3.3412662   1.670932  1.99964281 4.701813e-02
## smoke        -385.7292216 109.798891 -3.51305207 5.581001e-04
## race.noncauc -424.1822476 107.419147 -3.94885137 1.119300e-04
## ptl          -129.8287367 104.163705 -1.24639131 2.142139e-01
```

Interpretation: adjusted for other variables, being a non-Caucaisn, aas compared to being a Caucasian, decreases the expected baby's birth weight by 424.1822476 grams.

This change affected all variables–the estimates have all changed, yet the p-values have not change signficantly, not overturning any interpretations of the results we obtained earlier.

(d) This is not a good idea–there is no "natural" ordering of the race categories, i.e. the differnece between having a self-reported race of Caucasian vs. Black is not equal to the difference between Black vs. Other. Unit increase in the `race` does not indicate anything meaningful.

(e) Now, we make previous premature labors a categorical variable.