STAT 506 problem set 2 Meng-Ni Ho

Question 1: estimates, replicate standard error, 95% confidence interval for recs2015 data

1. Electricity usage in kilowatt hours: kwh
2. Natural gas usage, in hundreds of cubic feet: cufeetng
3. Propane usage, in gallons: gallonlp
4. Fuel oil or kerosene usage, in gallons: gallonfo

variable	estimate	variance	replicate_standard_error	lower_bound	upper_bound
kwh	10720.36	12950.79	1.06	10718.28	10722.44
cufeetng	355.25	1568.80	11.15	333.40	377.10
gallonlp	33.43	16.65	12.20	9.51	57.35
gallonfo	28.60	5.73	8.37	12.20	45.01

Question 2: fitting model using stata

representative age in part b is calculated using R

Question 3: fitting model using R

Part b:

- model1: null model, BIC = 1533.407

1. dependent variable: probability of individual lose their primary upper right 2nd bicuspid (outcome)
2. independent variable: age in months (ridagem)

```
##
## Call:  glm(formula = outcome ~ ridagemn, family = binomial(link = "logit"),
##     data = merge)
##
## Coefficients:
## (Intercept)     ridagemn
##    -8.35936      0.06968
##
## Degrees of Freedom: 7562 Total (i.e. Null);  7561 Residual
##   (119 observations deleted due to missingness)
## Null Deviance:       5763
## Residual Deviance: 1516  AIC: 1520
```

- Use fitted model to estimate ages (in months) in p = 0.25, 0.50, 0.75

1. when p = 0.25

```
## (Intercept)
##         104
```

2. when p = 0.50

```
## (Intercept)
##         120
```

3. when p = 0.75

```
## (Intercept)
##         136
```

- representative age (25%, 75%) in years:

```
## (Intercept) (Intercept)
##           8          12
```

Part C

- model2: BIC = 1542.055 add gender (riagendr) to model 1, since BIC increase, gender is excluded from the model
- model3: BIC = 1542.285 add race_1 (race = Mexican is coded 1, all other race coded 0) to the model1, since BIC increase, race_1 is excluded from the model
- model4: BIC = 1541.932 add race_2 (other hispanic and other race is coded 1, all other race coded 0) to model1, since BIC increase from model1, race_2 is excluded from the model
- model5: BIC = 1529.281 add race_3 (non-hispanic black is coded 1, all other race coded 0) to model4, since BIC decrease from model1, race_3 is retained in the model
- model6 (final model): BIC = 1462.895 add poverty income ratio (indfmpir) to model5, since BIC decrease from model5, poverty income ratio is retained in the model

final model:

```
##
## Call:  glm(formula = outcome ~ ridagemn + race_3 + indfmpir, family = binomial(link = "logit"),
##     data = merge)
##
## Coefficients:
## (Intercept)     ridagemn      race_31     indfmpir
##    -8.46029      0.07137      0.49498     -0.11907
##
## Degrees of Freedom: 7245 Total (i.e. Null);  7242 Residual
##   (436 observations deleted due to missingness)
## Null Deviance:       5535
## Residual Deviance: 1427  AIC: 1435
```

Part D

1. Adjusted predictions at the mean (for other values) at each of the representative ages determined in part b.

- adjsuted predictions at representative age = 8

```
## (Intercept)
##   0.2473362
```

- adjsuted predictions at representative age = 12

```
## (Intercept)
##   0.7533984
```

2. The marginal effects at the mean of any retained categorical variables at the same representative ages.

- marginal effect at representative age = 8

```
## (Intercept)
##  0.09341445
```

- marginal effect at representative age = 12

```
## (Intercept)
##  0.08793425
```

3. The average marginal effect of any retained categorical variables at the representative ages.

```
##   factor ridagemn    AME     SE      z      p  lower  upper
##  race_31 104.0000 0.0932 0.0293 3.1779 0.0015 0.0357 0.1507
##  race_31 136.0000 0.0878 0.0251 3.4980 0.0005 0.0386 0.1370
```