Stats 506, F18, Problem Set 2

Chen Xie, chenxie@umich.edu

October 15, 2018

Question 1

Aim: Use Stata to estimate the national totals for residential enrgy consumption.

And use the replicate weights to compute standard errors.

The table of estimates with 95% CI in Electricity Usage, Natural Gas Usage, Propane Usage, and Fuel Oil or Kerosene Usage is shown as below.

Table 1: **National Totals for Residential Energy Consumption**

Energy	Total Usage of Residential Energy (95% CI)
Electricity Usage, in kwh	1.267235e+12, (1.240325e+12, 1.294145e+12)
Fuel Oil Kerosene Usage, in gallons	3.380928e+09, (2.814850e+09, 3.947007e+09)
Natural Gas Usage, in hundres of cubic feet	3.962922e+10, (3.760638e+10, 4.165207e+10)
Propane Usagel, in gallons	3.951633e+09, (2.986881e+09, 4.916385e+09)

Question 2

Part a

In Stata, I used `import sasxport`, to read the data sets.

Part b

Aim: Fit the logistic regression to estimate the relationship between age and the probablity of individuals lose their primary upper right 2nd bicusppid.

Using this regression model, the estimated ages (in month) at 25%, 50%, and 75% of individuals losing their primary upper right 2nd bicuspid is (104, 120, 136).

And the range of representative age (in year) values with one year increments is (8, 9, 10, 11, 12).

Part c

The final model is the logistic regression between the probability and Age(in month)+Black(categorical)+Pir(poverty income ratio).

The regression table is shown below.

Table 2: **The Logistic Regression table of the Final Model**

Variables	Estimates	Standard Errors	Z Statistic	P.value	95% CI
Age, in month	0.0714	0.0027	26.37	0.000	(0.0661, 0.0767)
Black	0.4950	0.1489	3.32	0.001	(0.2031, 0.7869)
Poverty Income Ratio	-0.1191	0.0454	-2.62	0.009	(-0.208, -0.0301)
Intercept	-8.4603	0.3510	-24.10	0.000	(-9.1483, -7.7723)

The fitting process is:

The response variable is the probability of individuals losing primary right 2nd bicuspid.

And the smaller the BIC is, the better the model is.

First, we have the logistic regression between the Probability and Age(in month), which is our base model, and call it Model(Age). The BIC of this model is 1533.4068.

Second, we add the categorical variable, Gender, using the 'Female' as the reference, and call this Model(Age+Gender). The BIC of this model is 1542.0548, which is larger than the BIC of Model(Age). So we drop Gender, and still have the Model(Age).

Next, we create indicators of each race using the level 'White' as the reference, which is the largest group in race. Then we create indicators Black, Mexican, and Other.

```
First, we add the indicator Black, which represents the second largest group of race,
into our model, then we have Model(Age+Black), and the BIC of this model is 1529.2805.
Compared to the Model(Age), the BIC gets better, So we retain Black.

Then, we add the indicator Mexican, the third largest group of race, into our model,
then we have Model(Age+Black+Mexican), and the BIC of this model is 1533.1035.
Compared to the Model(Age+Black), we drop Mexican.

Last, we add the indicator Other, the smallest group of race, into our model,
then we have Model(Age+Black+Other), and the BIC of this model is 1536.1033.
Compared to the Model(Age+Black), we drop Other.

In final, we have the Model(Age+Black).
```

Finally, we add the Poverty Income Ratio into our model, and it becomes the Model(Age+Black+Pir). Compared to the Model(Age+Black), BIC is better, which is 1462.8945.

So our final model is Model(Age+Black+Pir).

Part d

(1)

The adjusted predictions at the mean at each of the representative ages can be shown in the table and figure below.

Table 3: **Adjusted Predictions at the Mean at each of the Representative Ages**

Age, in year	Estimates	Standard Errors	Z Statistic	P.value	95% CI
8	0.1459	0.0128	11.43	0	(0.1209, 0.1709)
9	0.2869	0.0167	17.23	0	(0.2542, 0.3195)
10	0.4865	0.0174	27.89	0	(0.4523, 0.5207)
11	0.6905	0.0155	44.67	0	(0.6602, 0.7208)
12	0.8401	0.0118	71.27	0	(0.817, 0.8632)

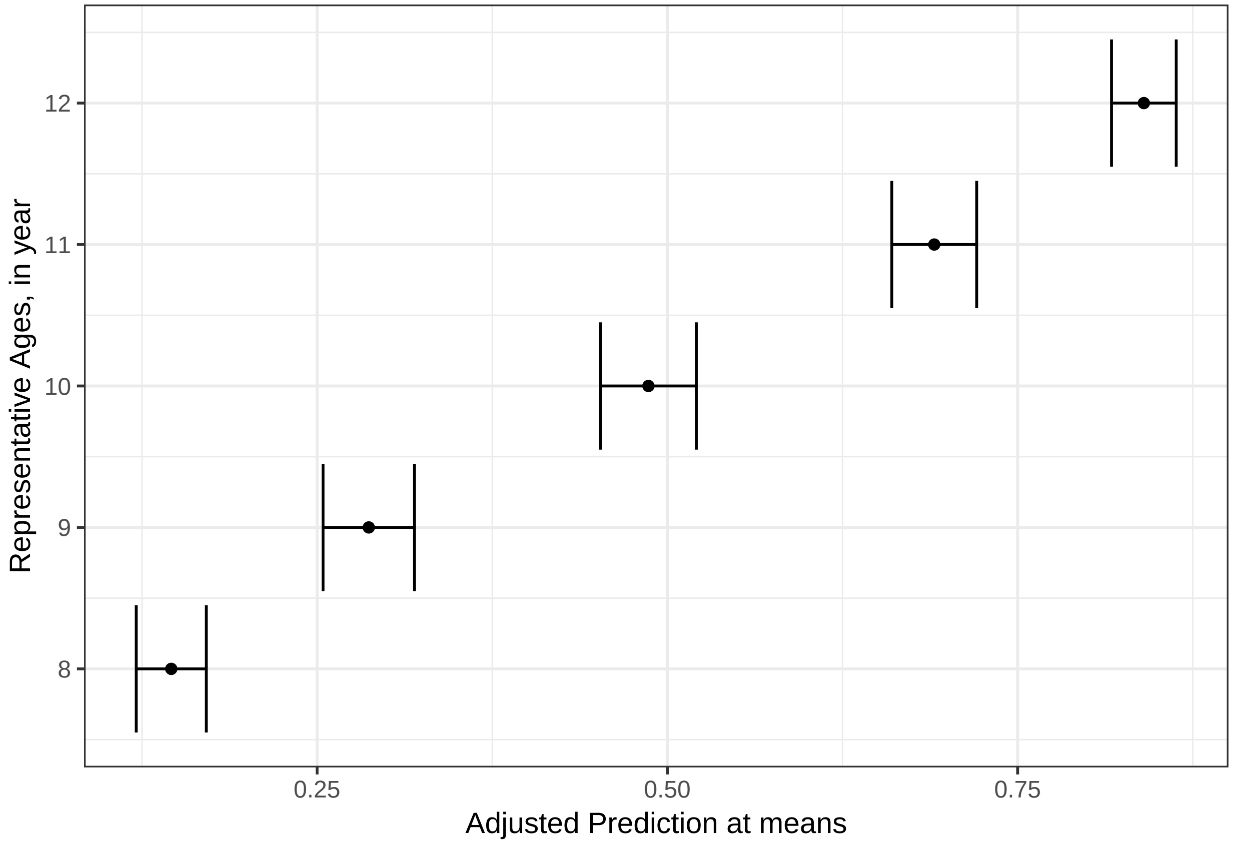

Figure 1: Adjusted Predictions(95%CI) at the Means at Each of the Representative Ages

(2)

The marginal effects at the means of black at each of the representative ages can be shown in the table and figure below.

Table 4: **Marginal Effects at the Mean of Black at each of the Representative Ages**

Age, in year	Estimates	Standard Errors	Z Statistic	P.value	95% CI
8	0.0668	0.0217	3.08	0.002	(0.0244, 0.1093)
9	0.1057	0.0328	3.22	0.001	(0.0414, 0.1699)
10	0.1230	0.0365	3.37	0.001	(0.0514, 0.1946)
11	0.1008	0.0290	3.48	0.001	(0.044, 0.1576)
12	0.0616	0.0175	3.52	0.000	(0.0273, 0.096)

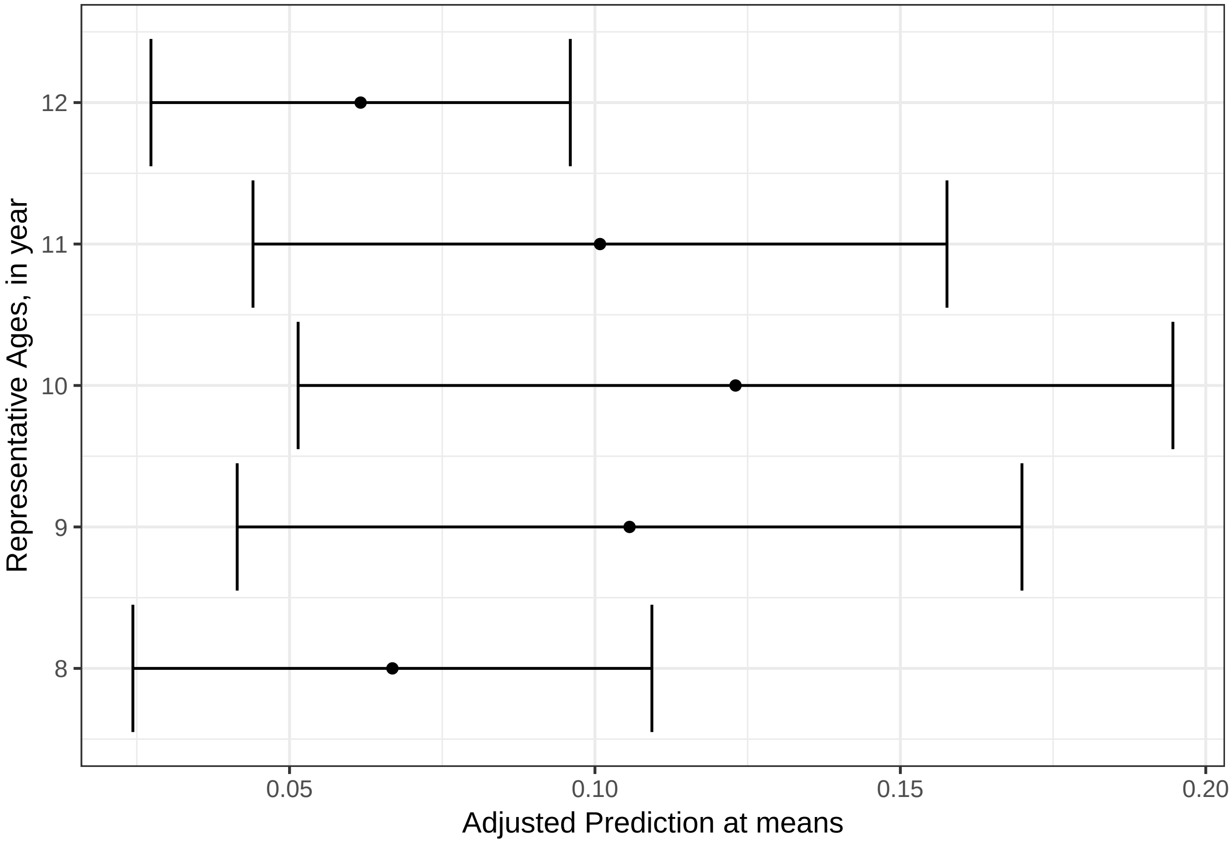

Figure 2: Marginal Effects at the Mean of Black at each of the Representative Ages

(3)

The average marginal effects of black at each of the representative ages can be shown in the table and figure below.

Table 5: **Average Marginal Effects at the Mean of Black at each of the Representative Ages**

Age, in year	Estimates	Standard Errors	Z Statistic	P.value	95% CI
8	0.0671	0.0217	3.09	0.002	(0.0245, 0.1096)
9	0.1052	0.0326	3.22	0.001	(0.0412, 0.1691)
10	0.1219	0.0363	3.36	0.001	(0.0508, 0.1931)
11	0.1004	0.0289	3.47	0.001	(0.0437, 0.157)
12	0.0619	0.0176	3.52	0.000	(0.0274, 0.0963)

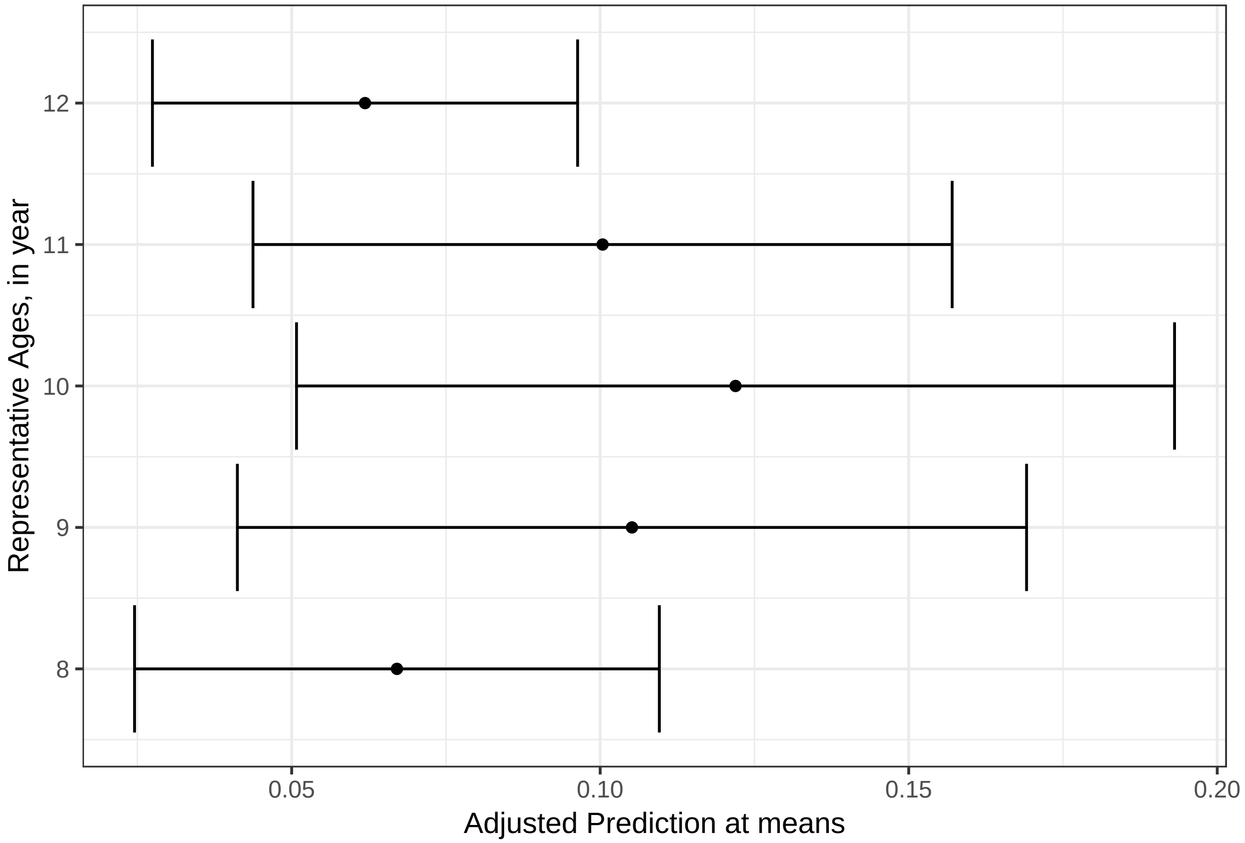

Figure 3: Average Marginal Effects at the Mean of Black at each of the Representative Ages

Part e

The refit regression table is shown as below.

Table 6: **The Logistic Regression table of the Model in Survey Design**

Variables	Estimates	Linearized Standard Errors	Z Statistic	P.value	95% CI
Age, in month	0.0619	0.0072	8.57	0.000	(0.0465, 0.0774)
Black	0.5435	0.1462	3.72	0.002	(0.2319, 0.8551)
Poverty Income Ratio	-0.0812	0.0522	-1.56	0.141	(-0.1924, 0.0301)
Intercept	-7.5160	0.8616	-8.72	0.000	(-9.3524, -5.6796)

When the sample design is changed, and the estimated of coefficients are also changed. In the final model of part c, the estimates of Age, Black, PIR, Intercept is (0.0714, 0.4950, -0.1191, -8.4603), while the estimates in this model is (0.0619,0.5435,-0.0812,-7.5160)

The p-value of the variable ‘pir (poverty income ratio)’ in the final model of part c is 0.009, which indicates that ‘pir’ is an extremely significant predictor in that model. But in this model, the p-value of ‘pir’ becomes much larger, which is 0.141. So the variable pir is not significant in this model.

And the standard error of the intercept in this model is 0.8616, which is significantly larger than that of the final model of part c, which is 0.3510.

Question 3

Part a

In R, I used “sasxport”, which is in the package “Hmisc”, to read the data sets.

Part b

Aim: Fit the logistic regression to estimate the relationship between age and the probablity of individuals lose their primary upper right 2nd bicusppid.

The BIC of this regression fit is 1533.4067716.

The regression table is shown below.

Table 7: **The Logistic regression between the Probability of losing Primary right 2nd bicuspid and Age(in months)**

Parameters	Estimate	Standard Errors	Z Statistic	p.value	BIC
Intercept	-8.3594	0.3235	-25.8412	0	1533.407
Age	0.0697	0.0026	27.1569	0	1533.407

Using this regression model, the estimated ages (in month) at 25%, 50%, and 75% of individuals losing their primary upper right 2nd bicuspid is 104, 120, 136.

And the range of representative age (in year) values with one year increments is 8, 9, 10, 11, 12.

Part c

The final model is the logistic regression between the probability and Age(in month)+Black(categorical)+Pir(poverty income ratio).

Table 8: **The Logistic regression between the Probability of losing Primary right 2nd bicuspid and Age(in months) + Black + Poverty Income Ratio**

Parameters	Estimate	Standard Errors	Z Statistic	p.value	BIC
Intercept	-8.4603	0.3510	-24.1018	0.0000	1462.895
Age	0.0714	0.0027	26.3741	0.0000	1462.895
Black	0.4950	0.1489	3.3237	0.0009	1462.895
Poverty Income Ratio	-0.1191	0.0454	-2.6240	0.0087	1462.895

The fitting process is:

The response variable is the probability of individuals losing primary right 2nd bicuspid.

And the smaller the BIC is, the better the model is.

First, we have the logistic regression between the Probability and Age(in month), which is our base model, and call it Model(Age). The BIC of this model is 1533.4067716.

Second, we add the categorical variable, Gender, using the ‘Female’ as the reference, and call this Model(Age+Gender). The BIC of this model is 1542.0547957, which is larger than the BIC of Model(Age).

So we drop Gender, and still have the Model(Age).

Next, we create indicators of each race using the level 'White' as the reference, which is the largest group in race. Then we create indicators Black, Mexican, and Other.

```
First, we add the indicator Black, which represents the second largest group of race,
into our model, then we have Model(Age+Black), and the BIC of this model is 1529.2805061.
Compared to the Model(Age), the BIC gets better, So we retain Black.

Then, we add the indicator Mexican, the third largest group of race, into our model,
then we have Model(Age+Black+Mexican), and the BIC of this model is 1533.1034658.
Compared to the Model(Age+Black), we drop Mexican.

Last, we add the indicator Other, the smallest group of race, into our model,
then we have Model(Age+Black+Other), and the BIC of this model is 1536.1032843.
Compared to the Model(Age+Black), we drop Other.

In final, we have the Model(Age+Black).
```

Finally, we add the Poverty Income Ratio into our model, and it becomes the Model(Age+Black+Pir). Compared to the Model(Age+Black), BIC is better, which is 1462.8945194.

So our final model is Model(Age+Black+Pir).

And the process can shown as the table below.

Table 9: **The Models in fitting process and their BIC**

Models	BIC
Age	1533.407
Age+Gender	1542.055
Age+Black	1529.281
Age+Black+Mexican	1533.103
Age+Black+Other	1536.103
Age+Black+PIR	1462.895

Part d

(1)

The adjusted predictions and 95%CI at the mean at each of the representative ages can be shown in the table and figure below.

Table 10: **Adjusted Predictions(95%CI) at the Means at Each of the Representative Ages**

Representative Ages, in year	Adjusted Prediction at the Mean (95% CI)
8	0.145906, (0.120383, 0.171429)
9	0.286881, (0.253575, 0.320187)
10	0.486482, (0.451598, 0.521365)
11	0.69049, (0.659572, 0.721408)
12	0.840091, (0.816516, 0.863666)

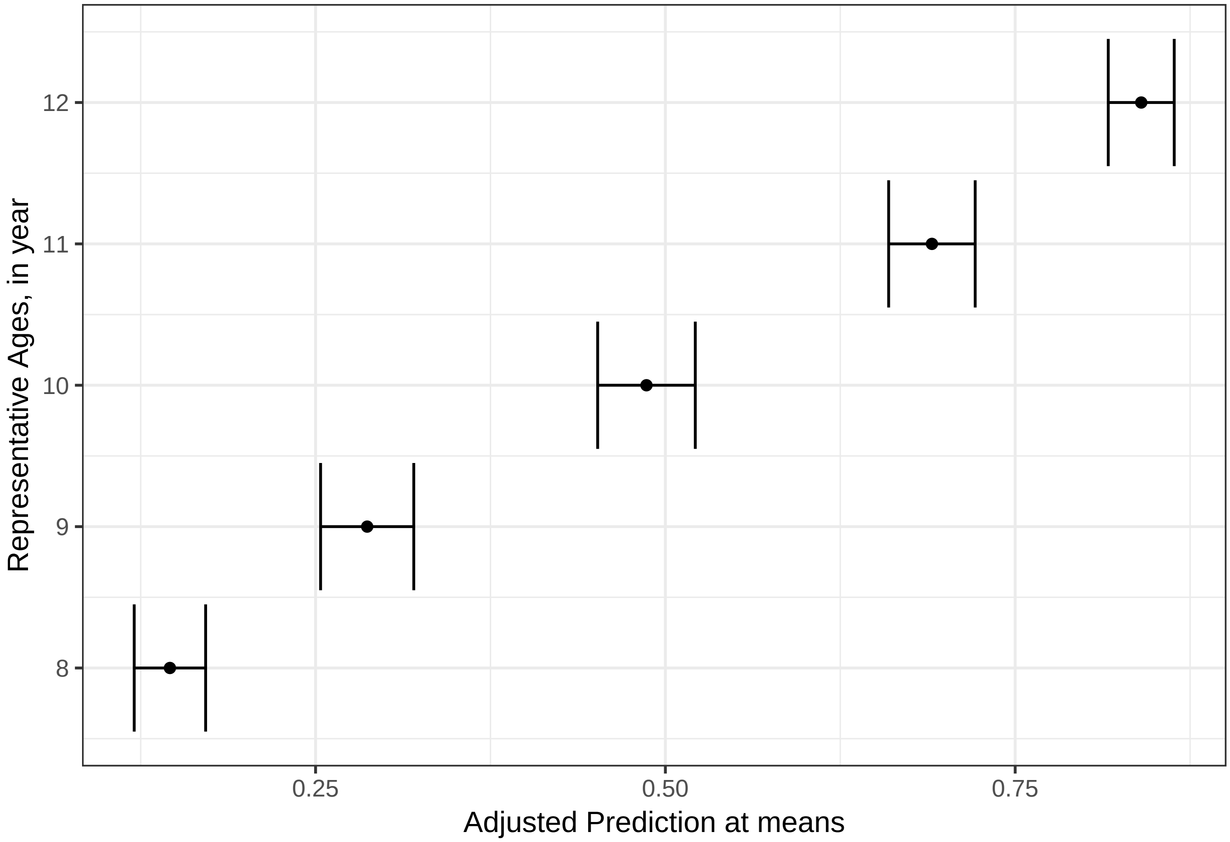

Figure 4: Adjusted Predictions(95%CI) at the Means at Each of the Representative Ages

(2)

The marginal effects at the means of black at each of the representative ages can be shown in the table and figure below.

Table 11: **Marginal Effects at the Means of Black at Each of the Representative Ages**

Representative Ages, in year	Mariginal effects at the Means of Black
8	0.066838
9	0.105667
10	0.123012
11	0.100826
12	0.061634

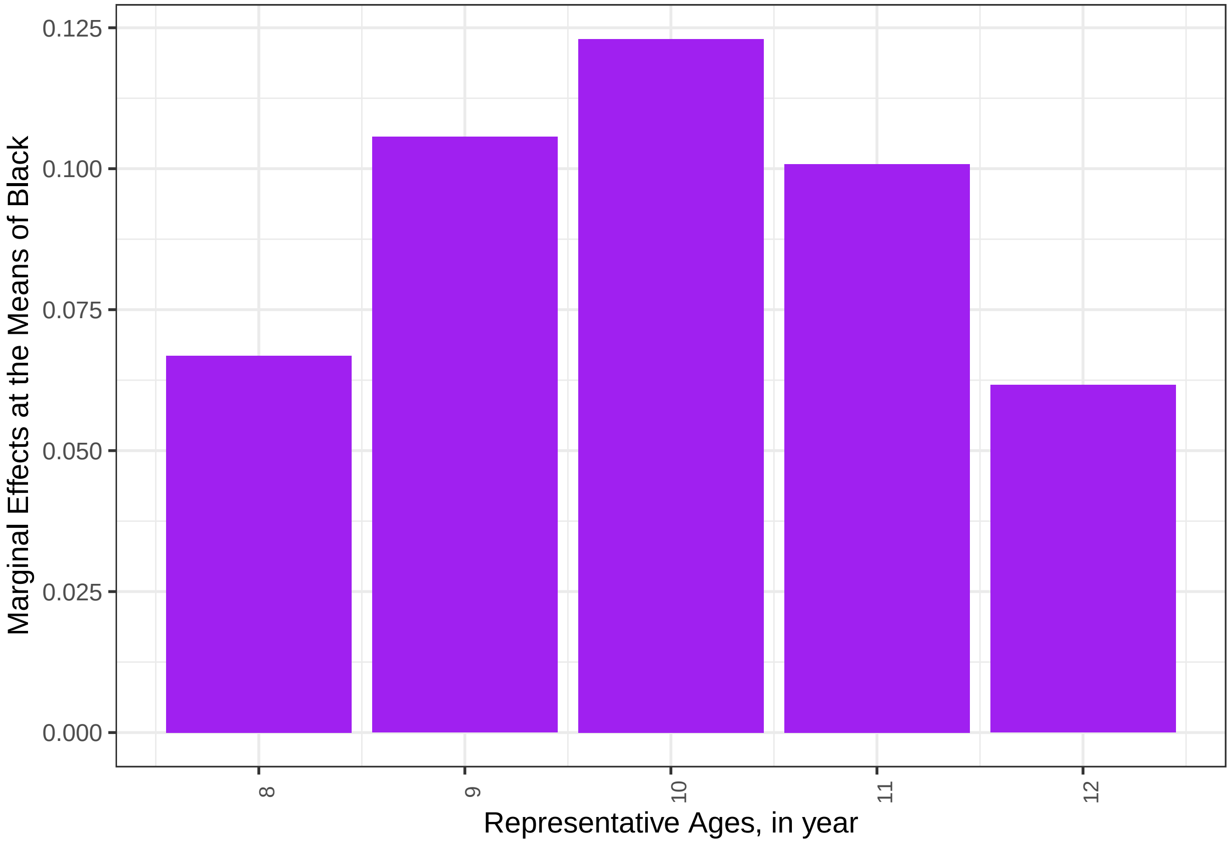

Figure 5: Marginal Effects at the Means of Black at Each of the Representative Ages

(3)

The average marginal effects of black at each of the representative ages can be shown in the table and figure below.

Table 12: **Average Marginal Effects of Black at Each of the Representative Ages**

Representative Ages, in year	Average Mariginal effects at the of Black
8	0.067064
9	0.105153
10	0.121934
11	0.100388
12	0.061892

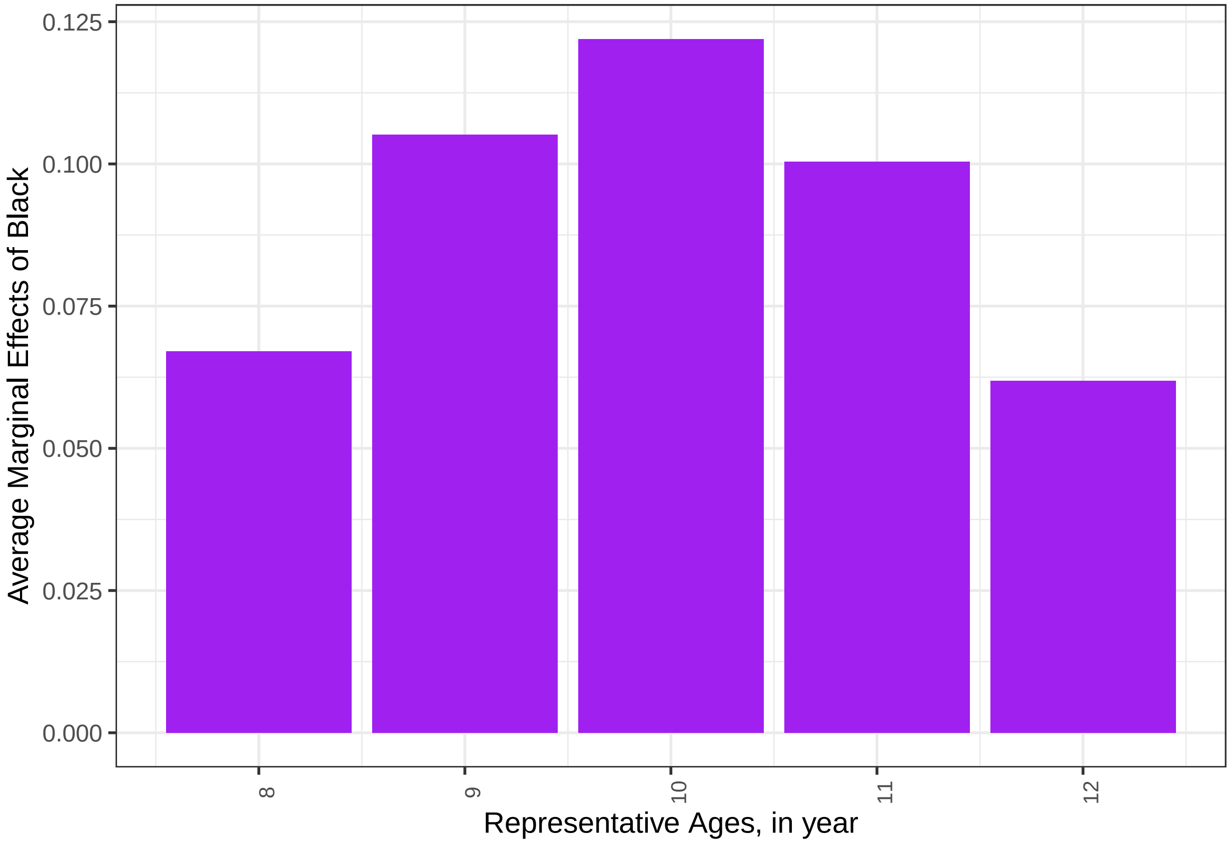

Figure 6: Average Marginal Effects of Black at Each of the Representative Ages