

Problem Set 2

Israel Diego

October 19, 2018

Question 1

Below we show estimates of national totals for residential energy consumption using Stata:

- i. Electricity usage in kilowatt hours
- ii. Natural gas usage, in hundreds of cubic feet
- iii. Propane usage, in gallons
- iv. Fuel oil or kerosene usage, in gallons

Table 1: **National Totals for residential energy consumption**
2015 Results expressed in (Billions)

	Totals	BRR Std Err.	95% Low	95% Upper
KWH	1267.235	13.697	1240.043	1294.427
Natural Gas	39.629	1.030	37.584	41.674
Propane	3.952	0.492	2.976	4.928
Fuel/Kerosene	3.381	0.275	2.835	3.927

Question 2/3

For this question we use the 2005-2006 Nhanes oral health data and 2005-2006 demographic data. Analysis is done in Stata and in R.

Part a.

First we merge both data sets by participant id. Which drops some unusued observations in the demographic data set. Mainly going from 10348 observations in the demographic data to 8305 observations in the oral health data.

Part b.

Using logistic regression, we estimate the relationship between age (in months) and the probability that an individual has lost a primary upper right 2nd bicuspid. Using the fitted model we estimate the ages at which 25, 50, and 75% of individuals lose their primary upper right 2nd bicuspid.

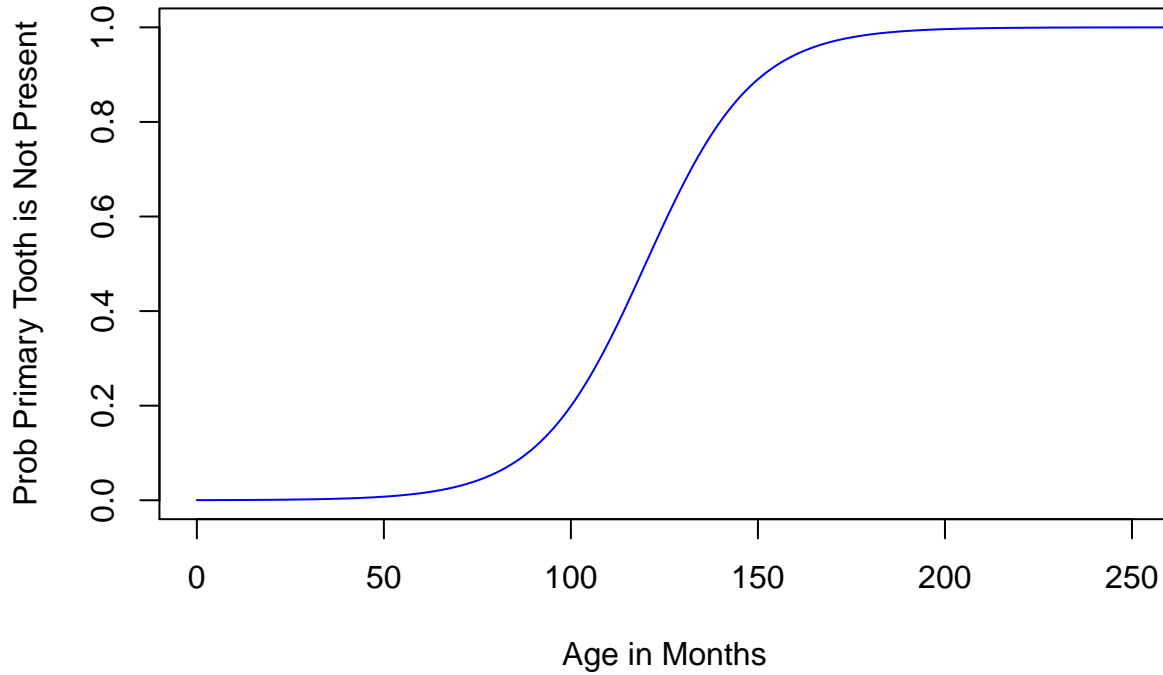
Below are the are the results of our logistic regression.

Table 2: **Logistic Regression of monthly age on presence of Primary Tooth Upper 2nd Bicuspid**

notPrimary	Coef.	Std. Err.	95% Low	95% Upper
ageMonths	0.07	0.00	0.06	0.07
constant	-8.36	0.32	-8.99	-7.73

On the next page we also plot our fitted model.

Logistic Regression of monthly age on presence of primary tooth



Below are the predicted ages in months and in years at which a primary tooth would be lost at the 25, 50, 75% levels.

Table 3: **Representative Ages at 25, 50, 75% levels according to our fitted model**

Age (Months) at 25%	Age (Months) at 50%	Age (Months) at 75%	Age (Years) at 25%	Age (Years) at 75%
104	120	136	8	12

Part c.

Continuing from the regression in part b, we now control for separate demographics, such as: Gender, Ethnicity, and Income Poverty Ratio.

Table 4: **Various Logistic Regressions on some demographic variables**

	Age	Age/Gender	Age/Mex	Age/Black	Age/Black/Other	Age/Black/InPovRatio
BIC	1533.41	1542.05	1542.28	1529.28	1536.1	1462.89

- i. We do not retain gender, because this caused an increase in BIC.
- ii. We do not retain Mexican, or Other categories of ethnicities, but we do retain Non-Hispanic Black.
- iii. We retain poverty income ratio because it drops BIC to 1462.895. As a final result we retain black ethnicity and Income Poverty Ratio to the regression which both improved BIC.

Part d.

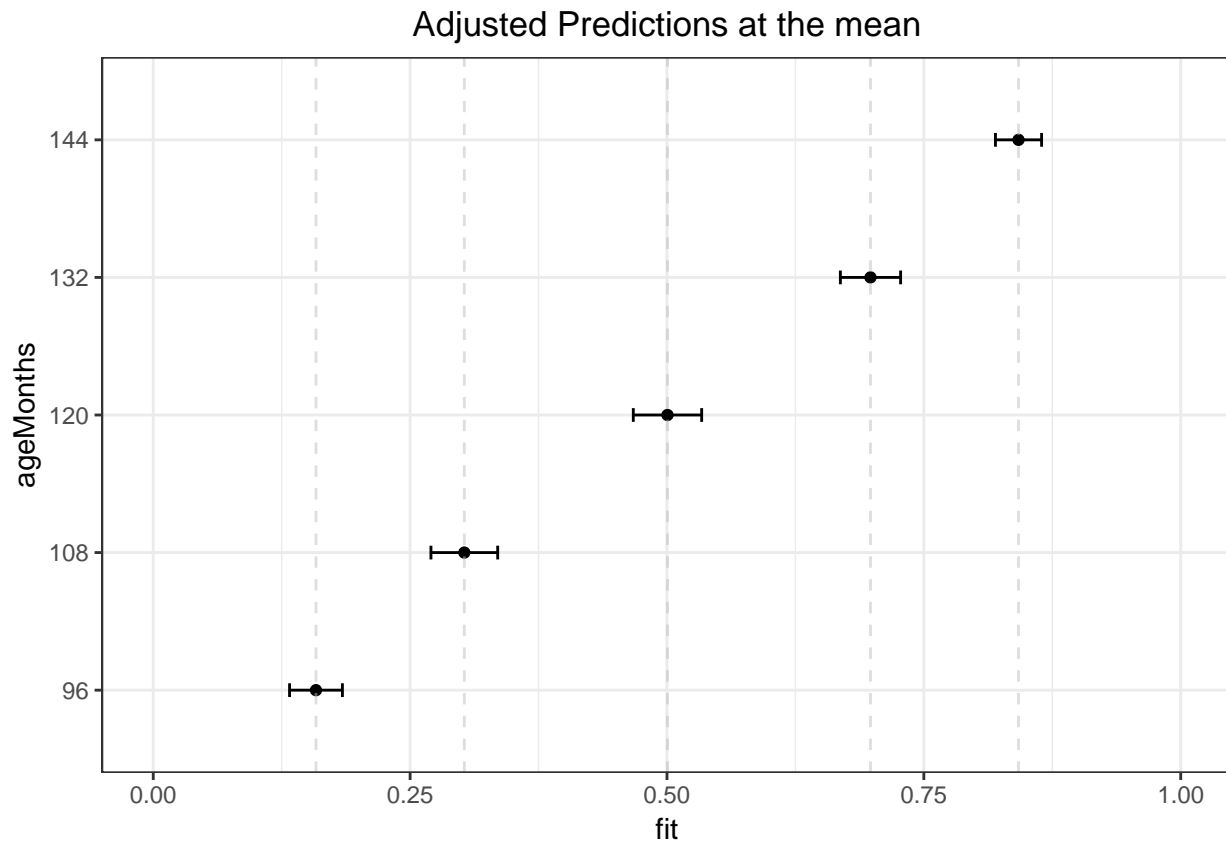
Now we use the margins command to compute the following:

- i. Adjusted predictions at the mean (for other values) at each of the representative ages (8 - 12 years old).

Table 5: **Adjusted predictions at the mean**

Age (Years)	Margin	Std. Err.	95% Low	95% Upper
8	0.158	0.013	0.133	0.184
9	0.303	0.016	0.271	0.335
10	0.500	0.017	0.468	0.533
11	0.698	0.015	0.669	0.727
12	0.842	0.011	0.820	0.864

Below we show a plot of the adjusted predictions which demonstrates that the representative ages are evenly spaced out.



- i. The marginal effects at the mean of Black and Income Poverty Ratio variables at the representative ages (8 - 12 years old).

Table 6: **Marginal Effects at the Mean**

Age (Years)	MEM	Std. Err.	95% Low	95% Upper
8	0.0617	0.0186	0.0253	0.0981
9	0.1013	0.0303	0.0419	0.1607
10	0.1237	0.0372	0.0508	0.1965
11	0.1058	0.0320	0.0430	0.1686
12	0.0665	0.0203	0.0267	0.1063

- i. The average marginal effect of Black and Income Poverty Ratio variables at the representative ages (8 - 12 years old).

Table 7: **Average Marginal Effects at Representative Ages**

Age (Years)	AME	Std. Err.	95% Low	95% Upper
8	0.0623	0.0189	0.0253	0.0994
9	0.1001	0.0295	0.0422	0.1579
10	0.1209	0.0355	0.0513	0.1904
11	0.1045	0.0313	0.0431	0.1659
12	0.0671	0.0207	0.0265	0.1077

Part e.

We refit the final model from part c using svyset and compare the new model to the old model.

Table 8: **Logit Regression with Before Survey Weighting**

notPrimary	Coef.	Std. Err.	p-value	95% Low	95% Upper
ageMonths	0.0714	0.0027	0.0000	0.0661	0.0767
black	0.4950	0.1489	0.0009	0.2031	0.7869
inPovRatio	-0.1191	0.0454	0.0087	-0.2080	-0.0301
constant	-8.4603	0.3510	0.0000	-9.1483	-7.7723

Table 9: **Logit Regression with Survey Weighting**

notPrimary	Coef.	Std. Err.	p-value	95% Low	95% Upper
ageMonths	0.0619	0.0072	0.0000	0.0465	0.0774
black	0.5435	0.1462	0.0021	0.2319	0.8551
inPovRatio	-0.0812	0.0522	0.1407	-0.1924	0.0301
constant	-7.5160	0.8616	0.0000	-9.3524	-5.6796

We notice that there are some slight changes in regression coefficients. The main result is that income poverty ratio is no longer significant as demonstrated by its p-value of 0.14. However, the age and black ethnicity predictors are still highly significant even though their standard errors increased a bit.