Who Needs Health Care

Emeka Ume

ECON 453

Final Draft

Spring 2012

Table of Contents

Motivation

- Health Care is currently about 18% of the US GDP or about $2.7 Trillion. Since President Barack Obama signed the Patient Protection and Affordable Care Act in March 2010, the debate about the scope and legality of the bill has gone up to the US Supreme Court.

- Although the courts decision is up in the air, it is a good idea to create contingencies as to how to handle the bill being over turned. This includes gathering and analyzing information on the American work force incase the bill has to be re-written or re-modeled.

- One of the main quips about the Health Care bill is that it would raise costs of privately funded care. This would then cause employers to drop employees who would have to go get theirs regardless of costs due to the Individual Mandate The purpose of this project is purely informational. We would what percentage of Americans have health insurance. Of those who have health insurance, who purchases the service privately and who has it provided by employees.

Research Questions

- Primary Research Questions:
  - Who has private Insurance Coverage?
  - Who has employee provided/union provided coverage

- Secondary Questions: What are characteristics of people who have Private care and those who have employee or Union based care?
  - For instance, a minority who has 4 kids while working overtime and holding private care *might* be on the brink of the economic food chain.

Description of Others Findings

- According to Health and Human services(HHS) data, about 16% of Americans were uninsured in 2005. As you shall see in a few slides, this is up to 19% as of 2010 and growing.
- Of those who are uninsured, 25% of them have income below the Federal Poverty Line, while about 11% have income 5x the Federal poverty line.
- Nineteen percent of people 18-24 years of age are uninsured, 22% of people aged 25-34 are uninsured and 7% of people aged 55-64 are uninsured. A good explanation for the decline after age 35 is that people begin to have families, and thus the need for insurance comes up. Also, as people get older, they are more likely to get ill, so insurance is used as a financial safety net.
- Of those uninsured, 57% have no children. These are mostly young adults who have no incentive to purchase insurance since they already have kids.
- Most of the Uninsured are White(Non Hispanic), about 48% while they are 67% of the population. Blacks compromise 15% with 12% of the population and Hispanics make up 30% with 14% of the population. Native Americans make up 1% of the population with 2% of the uninsured, while pacific Islanders are 5% of the population and also 5% of the Uninsured.
- By Work Status, most uninsured are Full time workers– about 46% while being 55% of the total work force. Part time workers 28% are of the uninsured although they are 19% of the population. People who don't work make up 26% of the uninsured and 27% of the population.
- About 19% of uninsured people live in the South. The West Coast has 17% and the East and Northeast have 13 and 12% respectively.

http://aspe.hhs.gov/health/reports/2011/CPSHealthIns2011/ib.shtml

Description of Raw Dataset

- Source
  - American Community Survey(ACS) 2010 from The Integrated Public Use Micro-data Series (IPUMS)
    - This is a population information gathering database maintained by the University of Minnesota.
    - Data was requested and downloaded off IPUMS in .gz format, then extracted to get a .dat in other to allow SAS access.
- Description
  - Data has 3,061,692 Observations and 51 Variables.
  - Data was collected in 2010, and its preceding was collected in 2009.
  - Unit of Observation: Individualized.
  - Data is not censored. Military personnel are also included in the total dataset, although they will be skipped in this since the VA provides for them.

Dependent Variable

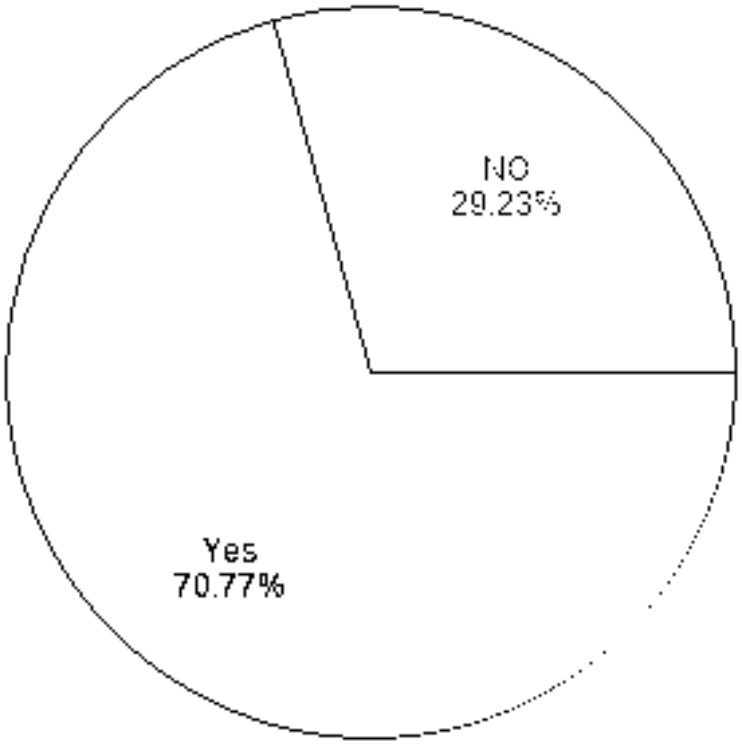
- Private health insurance
  - Did this individual have private health insurance coverage at the time of interview?
  - The US Census also includes people with Employee/Union healthcare as “private”. At the time of interview, persons had health insurance through a current employer, former employer, or union. Persons covered by another family member's current employer, former employer, or union included as well.
  - The Dependent variable “Private Health Insurance ” or PHI was selected in order to point out who-is-who amongst non private health insurance holders. The point is to see what the characteristics are of people who currently do not hold private health insurance.

DESCREPTIVE STATISTICS

Descriptive Statistics: Private Health Insurance (PHI)

- This is the Dependent Variable
- Majority of Americans, 70%, have private health insurance while 29% of Americans do not.
- The format for this was converted into a binary in order to build a regression model. 0 stands for No, 1 stands for Yes. For this project, we will be modeling “1” only, only those with care.
- There are no missing observations

Private health Insurance(PHI)				
HCOVPRIV	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NO	549074	29.23	549074	29.23
Yes	1329510	70.77	1878584	100.00



Independent Variables

- Age
- Sex
- Race
- Marital Status
- Number of Children in Household
- Classification of Work(e.g. Private Sector,)
- Education Attainment
- Hours Worked per Week

- In order to make the data reflect the reality on the ground, it had to be manipulated to show a sample of the actual US workforce. These are people aged 18 to 64. People are assumed to be in High School below age 18 and are eligible for Social Security Benefits at age 65.
  - The cleaned up data had a total of 1,878,584 Observations.

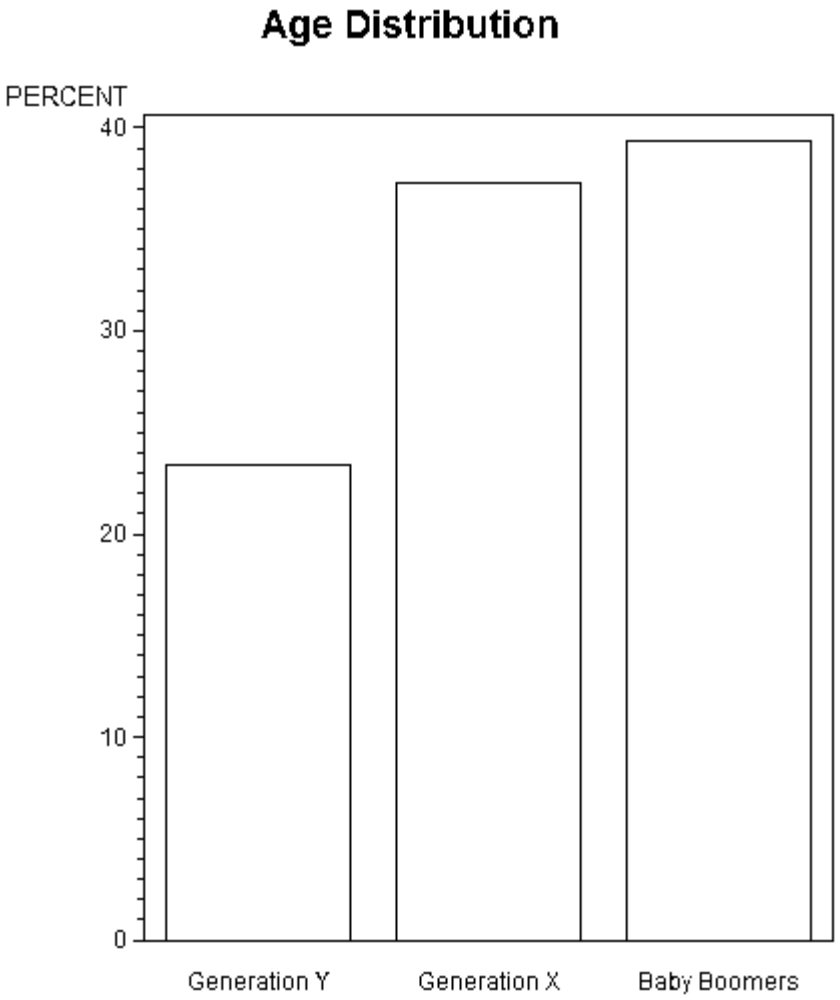
Regression Model

- Logistic Regression
  - The idea here is to look at certain social characteristics to see if there is a possibility of predicting or determining a person who is likely not to have private health insurance. Listed below are the Dependent and Independent variables.
    - Dependent Variable
      - Private health Insurance(PHI)
    - Independent Variables
      - Age
      - Sex
      - Race
      - Marital Status
      - Number of Children in Household
      - Classification of Work(e.g. Private Sector,)
      - Education Attainment
      - Hours Worked per Week

Descriptive Statistics: Age

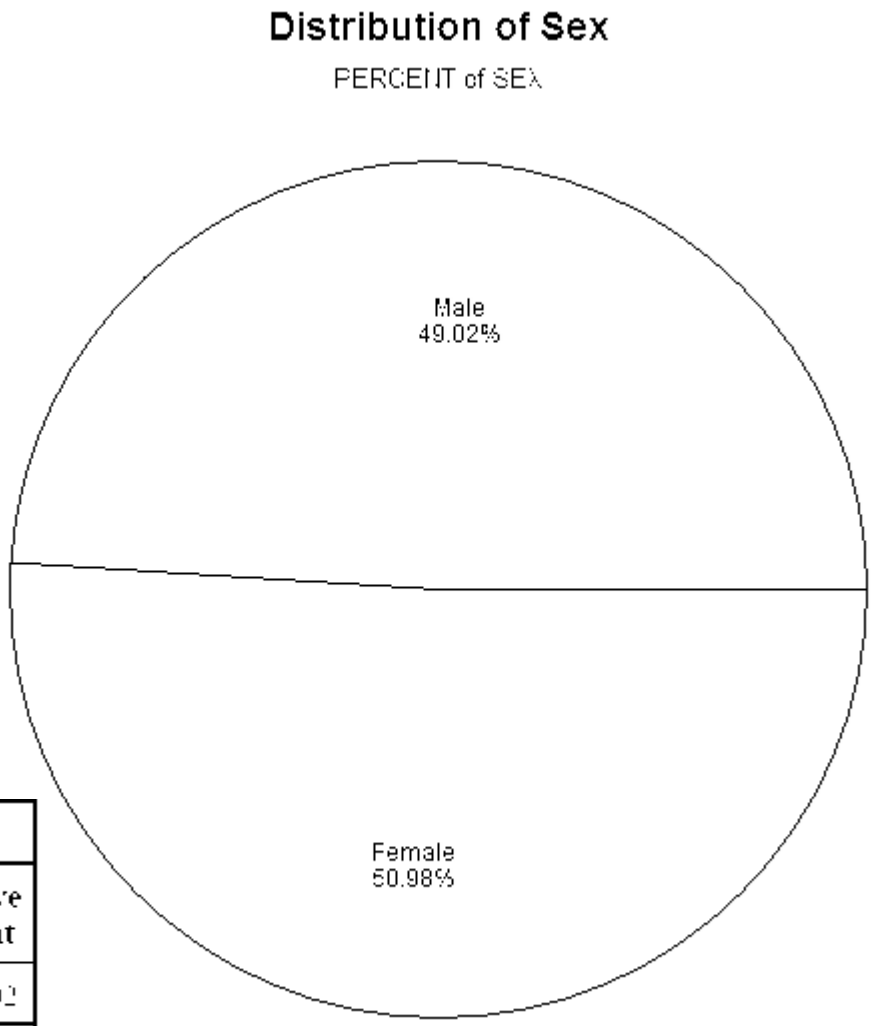
- We see that a majority of subjects in the sample are Baby Boomers, typically between the ages of 48 and 66. Next we have Generation X, people between ages 30 and 47. Finally, Generation Y, people between 18 and 29.
- Age range is from 18 to 64. The average age is just below 42.
- There are no missing Observations.

Analysis Variable : AGE Age		
Minimum	Maximum	Mean
18.0000000	64.0000000	41.8567511



Descriptive Statistics: Sex

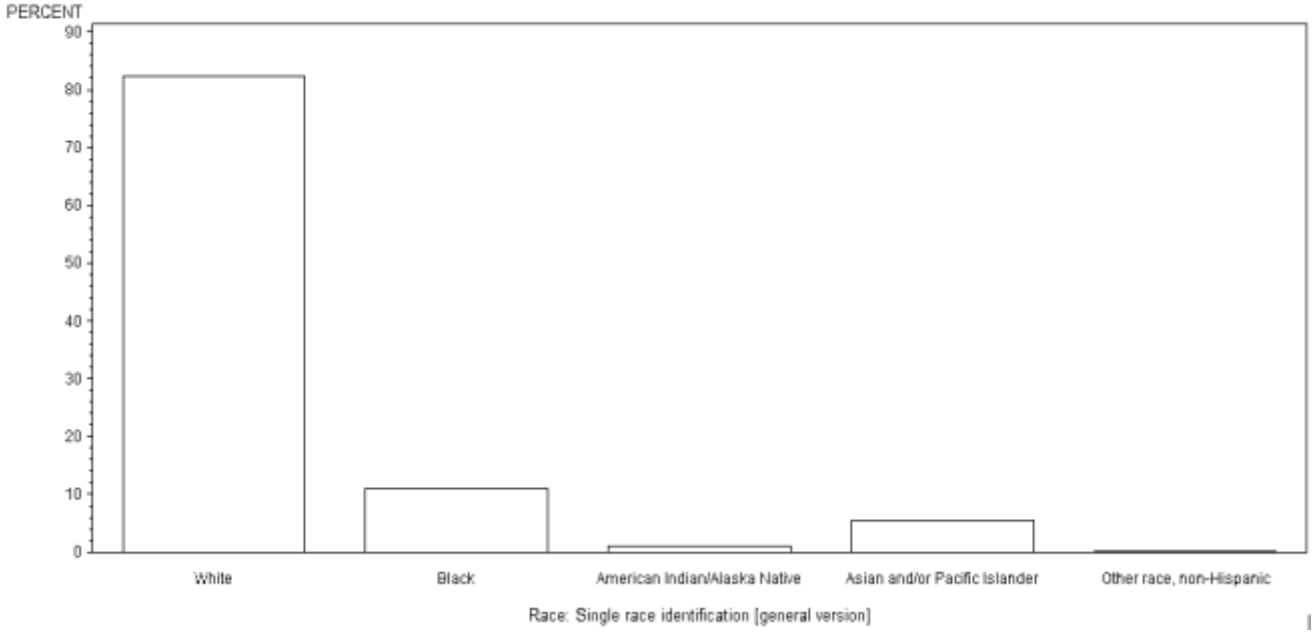
- Majority of the observations in our sample are Female, almost 51%, compared to men with about 49%
- There are no missing Observations.



Sex				
SEX	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	920951	49.02	920951	49.02
Female	957633	50.98	1878584	100.00

Descriptive Statistics: Race

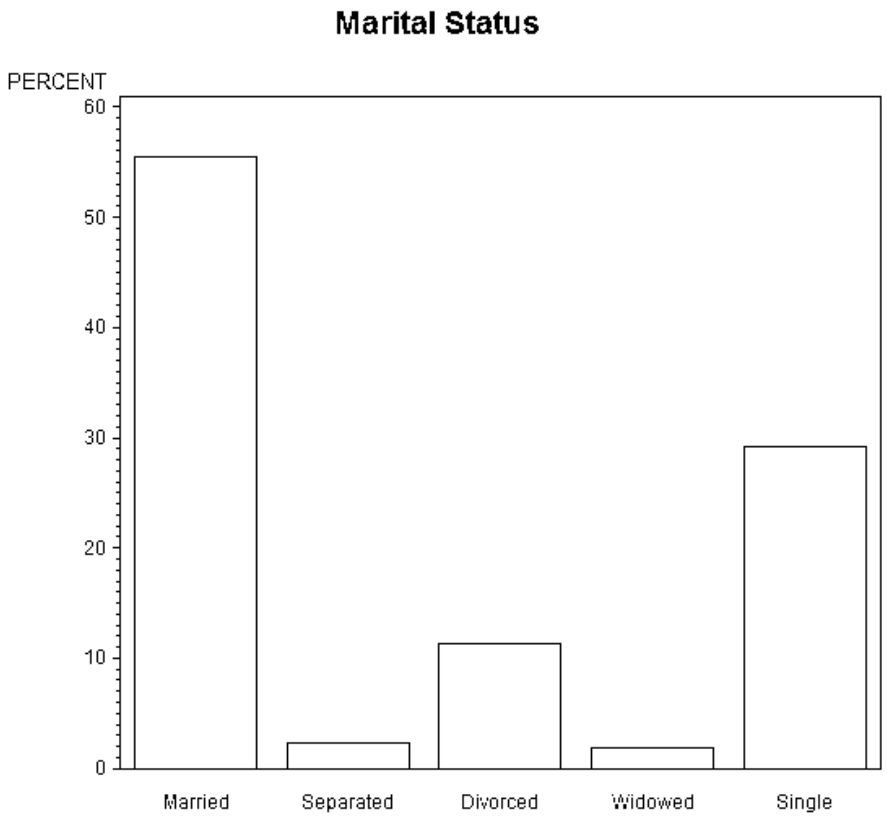
- We see that over 80% of the observations identified themselves as white, while just over 10% are black.
- Other races include Bi-racial people. Hispanic people are not included in the observation because Hispanic is an ethnicity, not a race.
- There are no missing Observations.



Racial identification				
RACESING	Frequency	Percent	Cumulative Frequency	Cumulative Percent
White	1547232	82.36	1547232	82.36
Black	207529	11.05	1754761	93.41
Native American	18843	1.00	1773604	94.41
Asian/Pacific Islander	102135	5.44	1875739	99.85
Other race	2845	0.15	1878584	100.00

Descriptive Statistics: Marital Status

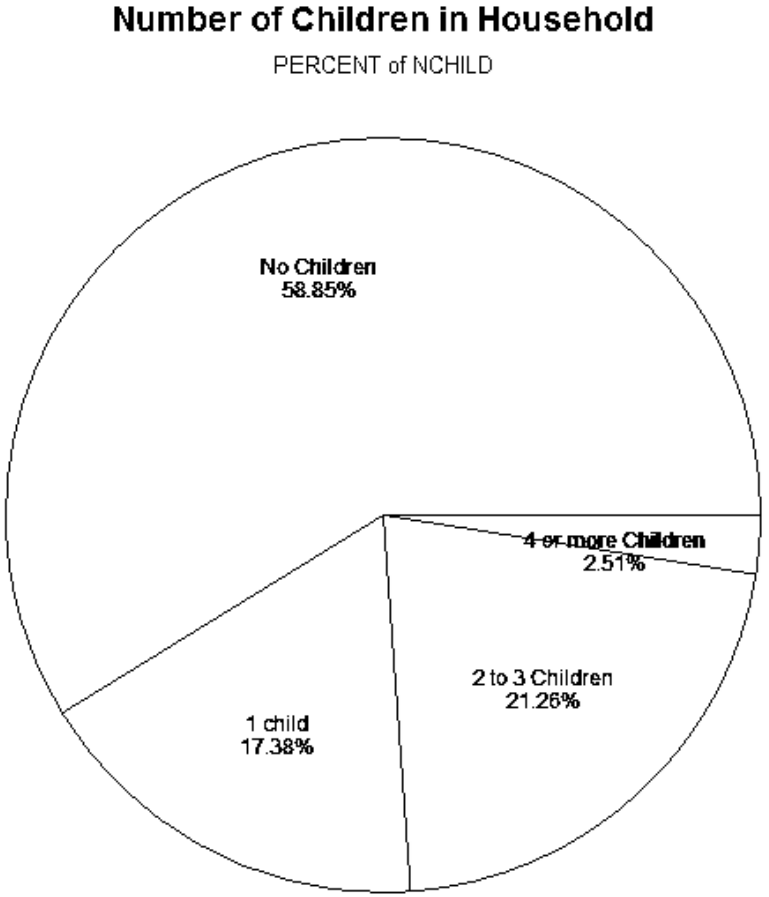
- Majority of our observations are married couples, over 50%. About 11% are divorced and 29% are single.
- There are no missing Observations.



Marital status				
MARST	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Married	1041846	55.46	1041846	55.46
Separated	42717	2.27	1084563	57.73
Divorced	212443	11.31	1297006	69.04
Widowed	33911	1.81	1330917	70.85
Single	547667	29.15	1878584	100.00

Descriptive Statistics: Number of children at home

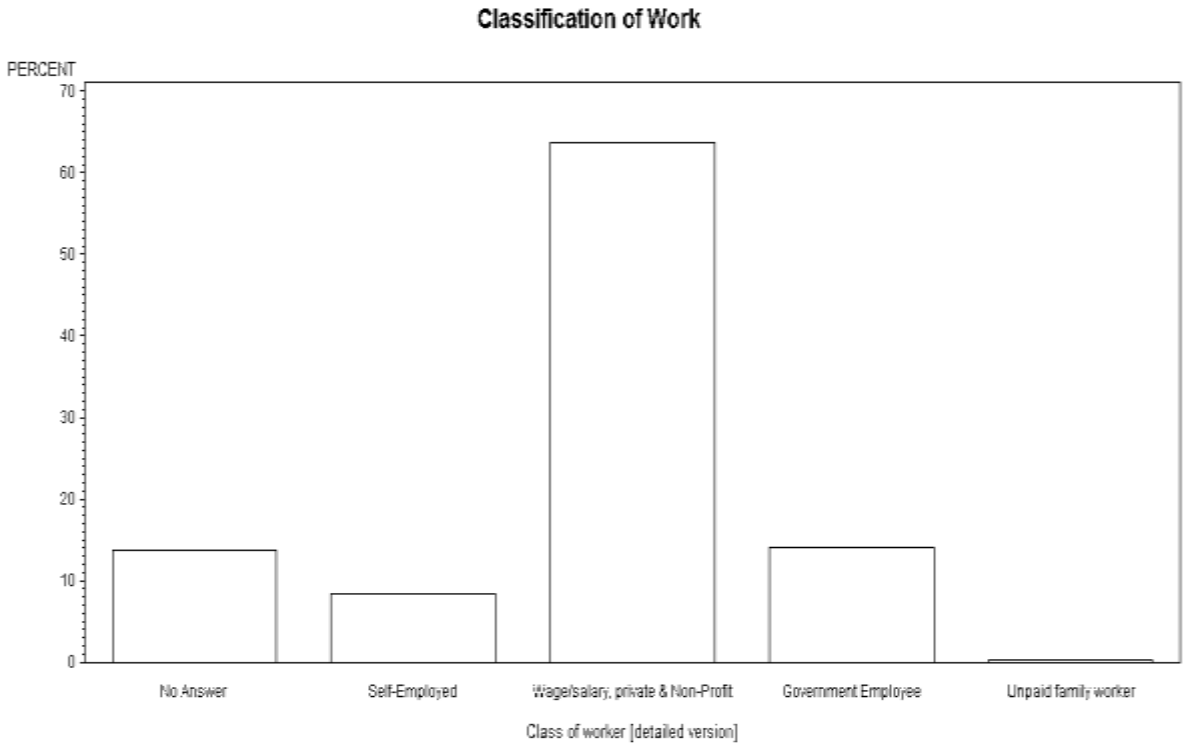
- In our raw dataset, we see that a majority of people have no children. 17% have one child and about 20% have 2 or 3 children.
- There are no missing Observations.



Number of children in household				
NCHILD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No Children	1105543	58.85	1105543	58.85
1 child	326442	17.38	1431985	76.23
2 to 3 Children	399438	21.26	1831423	97.49
4 or more Children	47161	2.51	1878584	100.00

Descriptive Statistics: Classification of Work

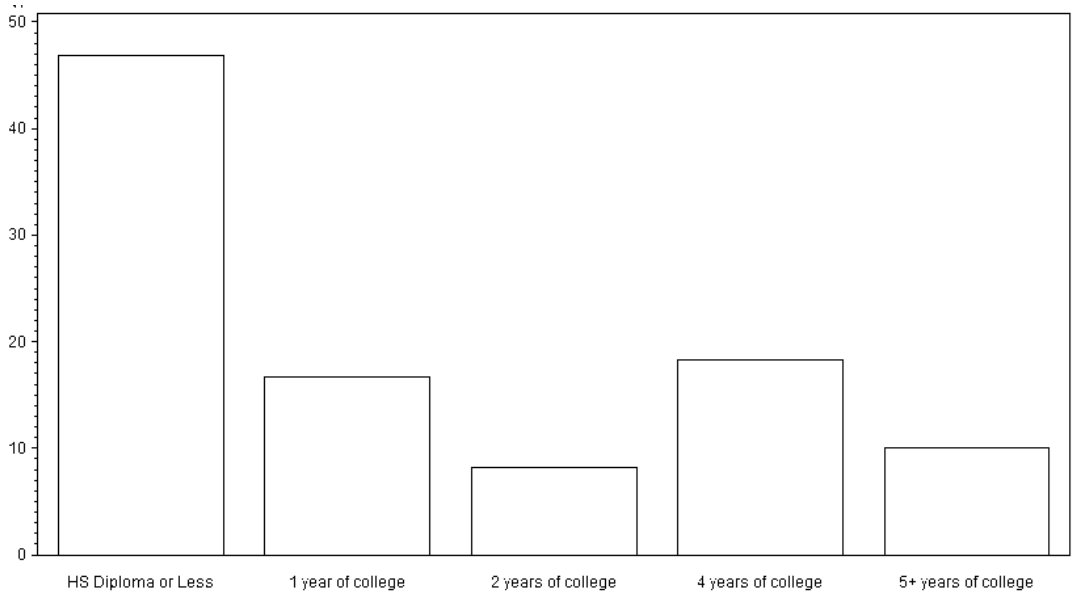
- Slightly over 60% of the modeling dataset work in the Wage/Salary sector. This includes both for profit and non-profit . When we add the Self-Employed group, this comprises the total private sector and the total comes up to about 70%. Just over 10% work with the Government(Federal, State and Local)
- There are no missing Observations.



Class of worker [detailed version]				
CLASSWKRD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Not Working	257810	13.72	257810	13.72
Self-Employed	160634	8.55	418444	22.27
Wage/salary, private & Non-Profit	1197118	63.72	1615562	86.00
Government Employee	263022	14.00	1878584	100.00

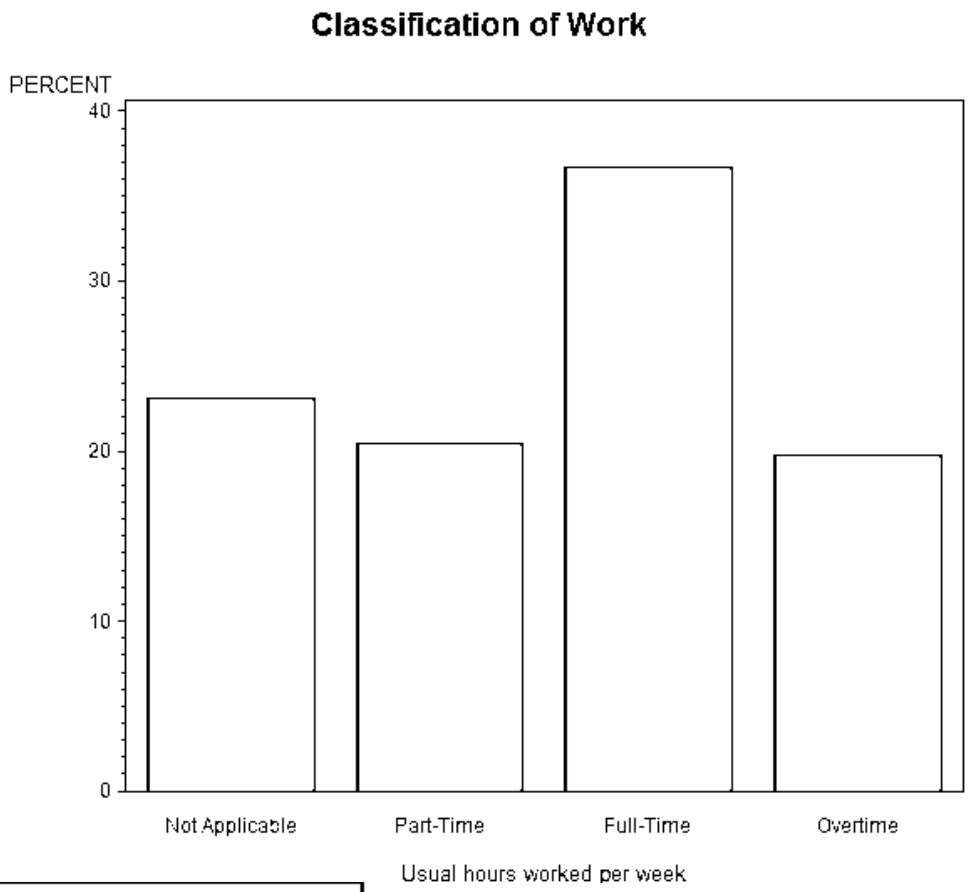
Descriptive Statistics: Education

- Just over 50% of the sample have at least a high school degree, but not an Associates(2 Year) degree.
- Almost 30% of the sample have at lease a bachelors degree. This includes Post Graduates.
- There are no missing Observations.



Educational attainment				
EDUC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
HS Diploma or Less	878961	46.79	878961	46.79
1 year of college	314264	16.73	1193225	63.52
2 years of college	152922	8.14	1346147	71.66
4 years of college	342797	18.25	1688944	89.91
5+ years of college	189640	10.09	1878584	100.00

Descriptive Statistics: Hours Worked per Week



- Just under 40% of the observation are full time workers. Full time is measured as between 36-40 hours per week. Any time below 36, but more than Zero is listed as Part-time, and people over 40 Hours per week are Overtime. People in the Non-Applicable spot are non workers and unemployed. Non-Workers for instance include full time college students who do not work. There are no missing observations.

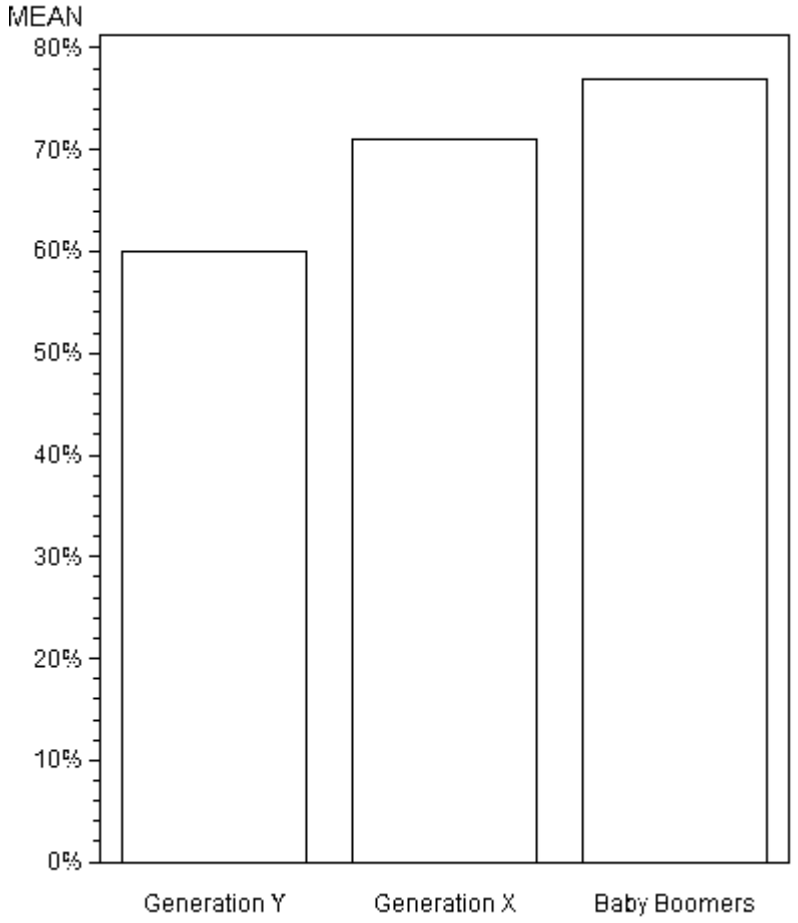
Usual hours worked per week				
UHRSWORK	Frequency	Percent	Cumulative Frequency	Cumulative Percent
None	433214	23.06	433214	23.06
Part-Time	384500	20.47	817714	43.53
Full-Time	689593	36.71	1507307	80.24
More than 40	371277	19.76	1878584	100.00

CONDITIONAL STATISTICS

Private Insurance holders based on Age Group

We see that a majority(50%+1) of all three generations in the observation have some form of Private Healthcare. The lowest is the Generation Y with 60%, these are the youngest generation, who may feel that there is no need for care since they are still relatively young with few or no children. On the other end, Baby Boomers, who are the more risk averse generation pull the most weight.,

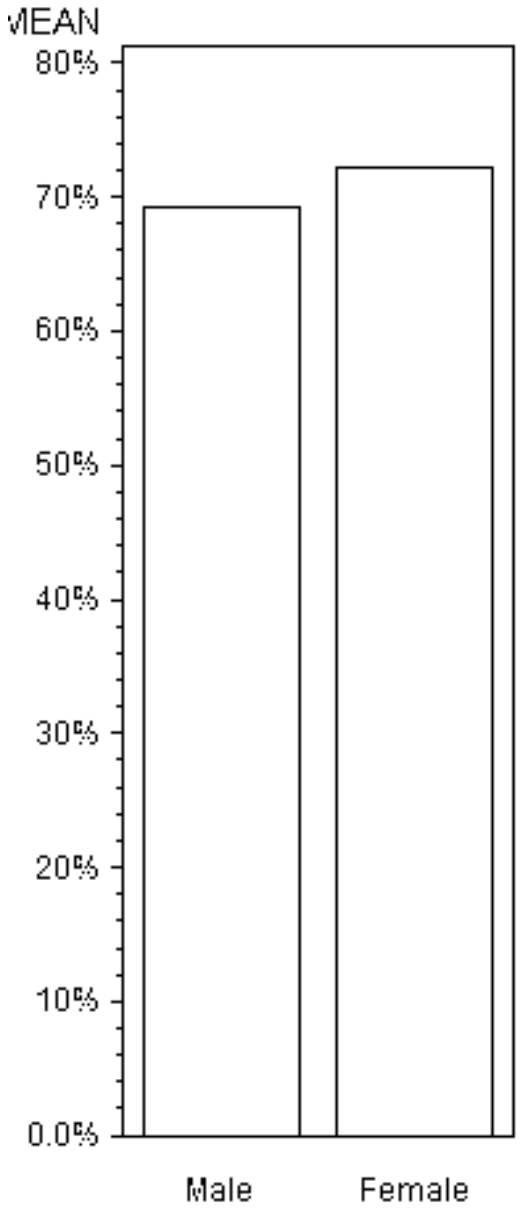
	care
	Percent
Age	
Generation Y	60.0%
Generation X	71.1%
Baby Boomers	76.9%
Overall	70.8%



Private Insurance holders based on Race

- Women are more likely to hold private health care than men. This may be due to the fact that women are more likely to take larger steps to look after their children than men. This includes purchasing more or better care.

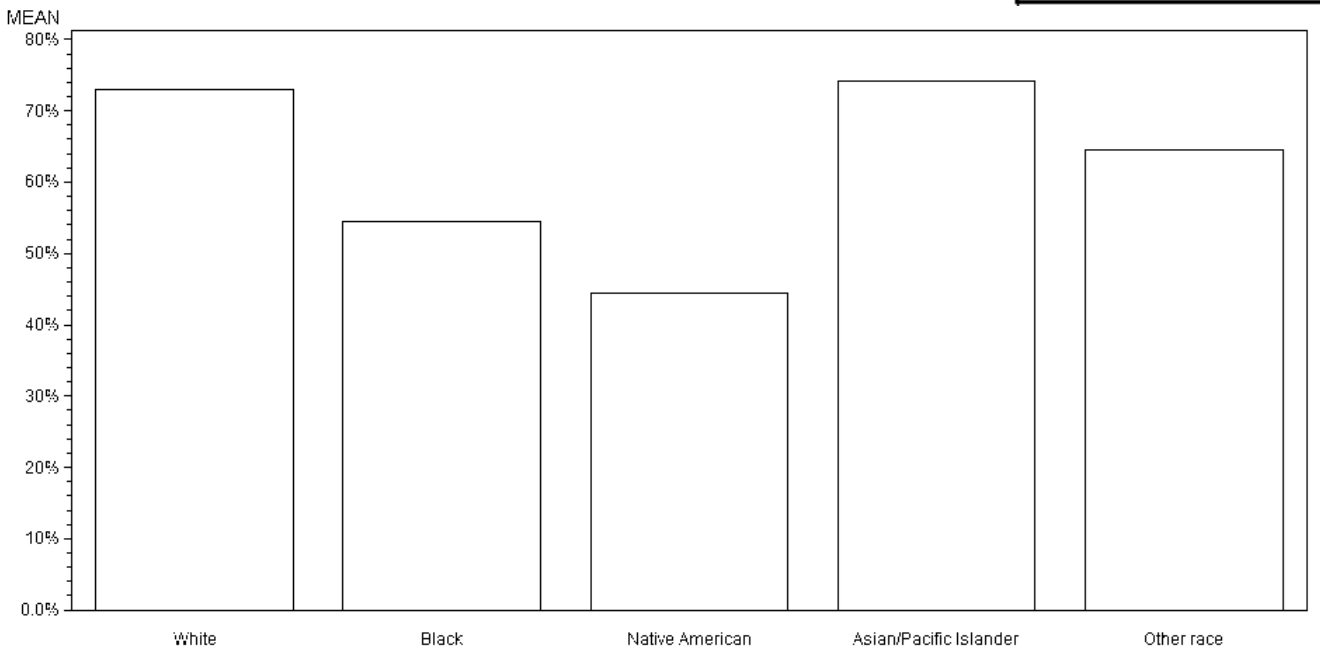
	care
	Percent
Sex	
Male	69.2%
Female	72.3%
Overall	70.8%



Private Insurance holders based on Race

- People of Asian roots are more likely than any other race, including Whites, to have private Health Insurance . This is contrary to popular conceptions that Whites dominate the health care ownership picture.
- A reason for this could be that Asian Americans have a larger percentage with college educations, and thus more disposable income than other races.

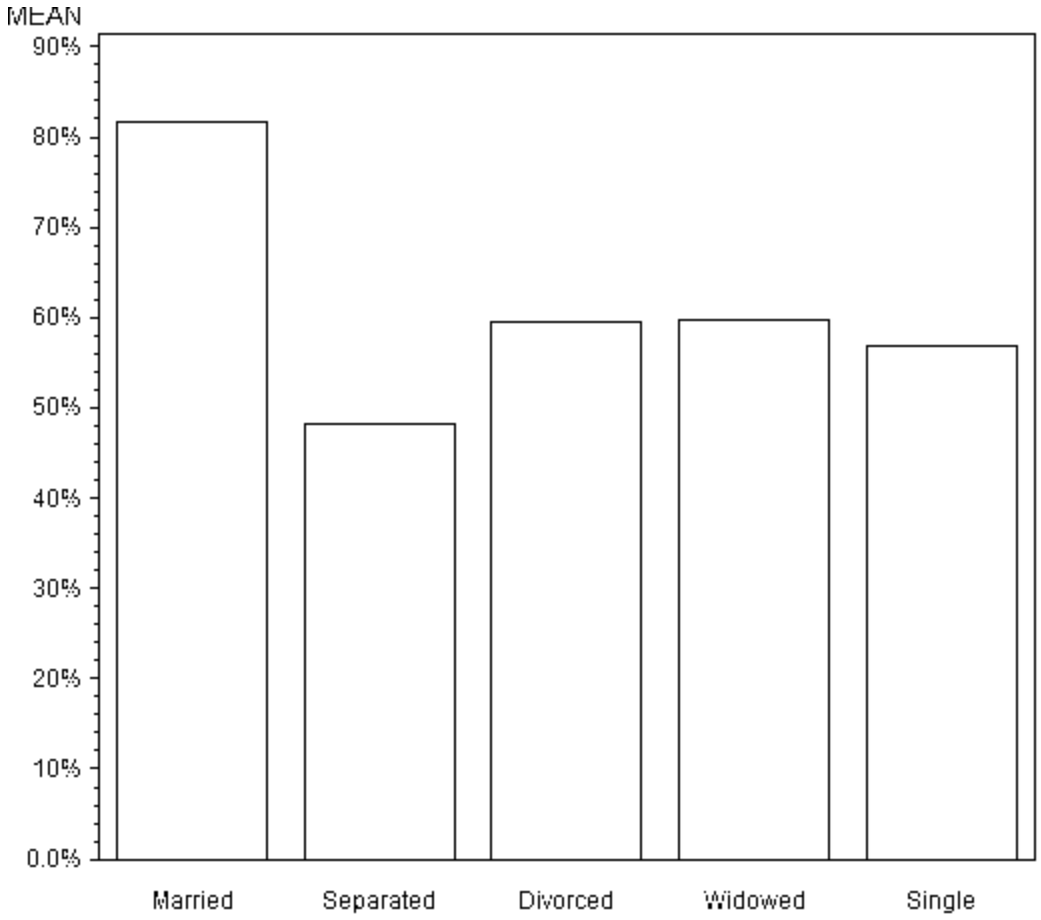
	care
	Percent
Racial identification	
White	73.1%
Black	54.5%
Native American	44.4%
Asian/Pacific Islander	74.1%
Other race	64.6%
Overall	70.8%



Private Insurance holders based on Marital Status

- 81% of married people in the observation have Private Coverage. Could be due to the fact that they have children to fend for.
- Separated people have the least percentage with private coverage. This may be due to financial reasons.

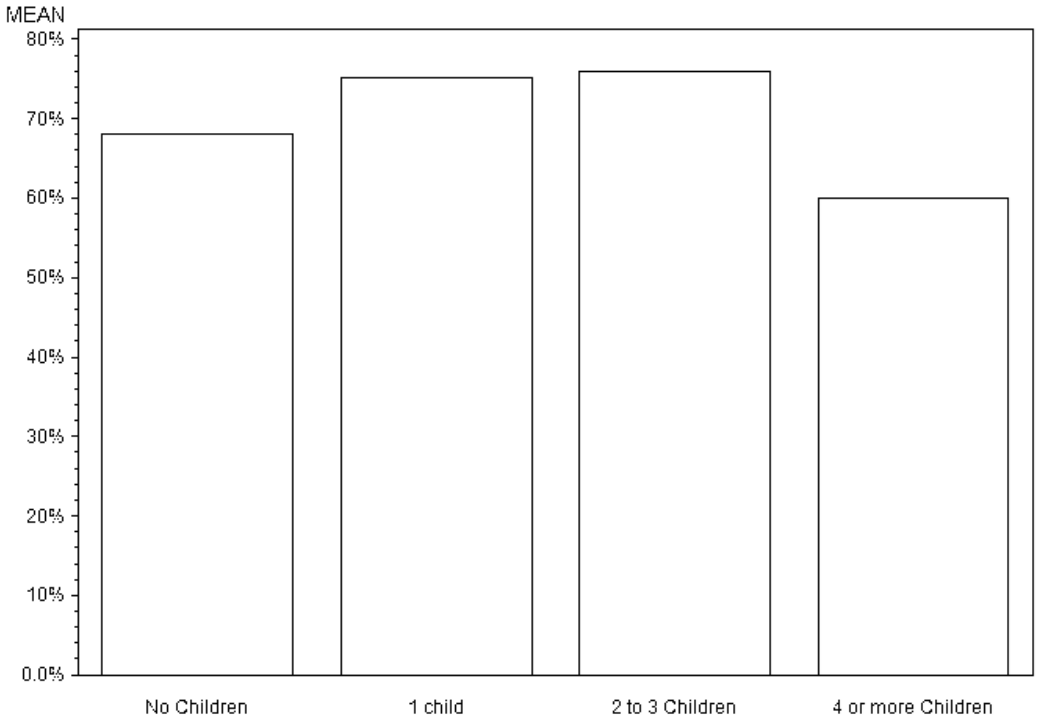
Marital status	
Married	81.7%
Separated	48.1%
Divorced	59.4%
Widowed	59.7%
Single	56.8%
Overall	70.8%



Private Insurance holders based on Number of Children

- People with 2 or three children are more likely than those with no children, 1 child or 4 children to have Private Care.
- It seemed like the more children you have, the more likely you are to have care, but after 3 children, the ratio begins to diminish. A possible explanation is that the cost of raising that many kids may be too costly to also hold private care.

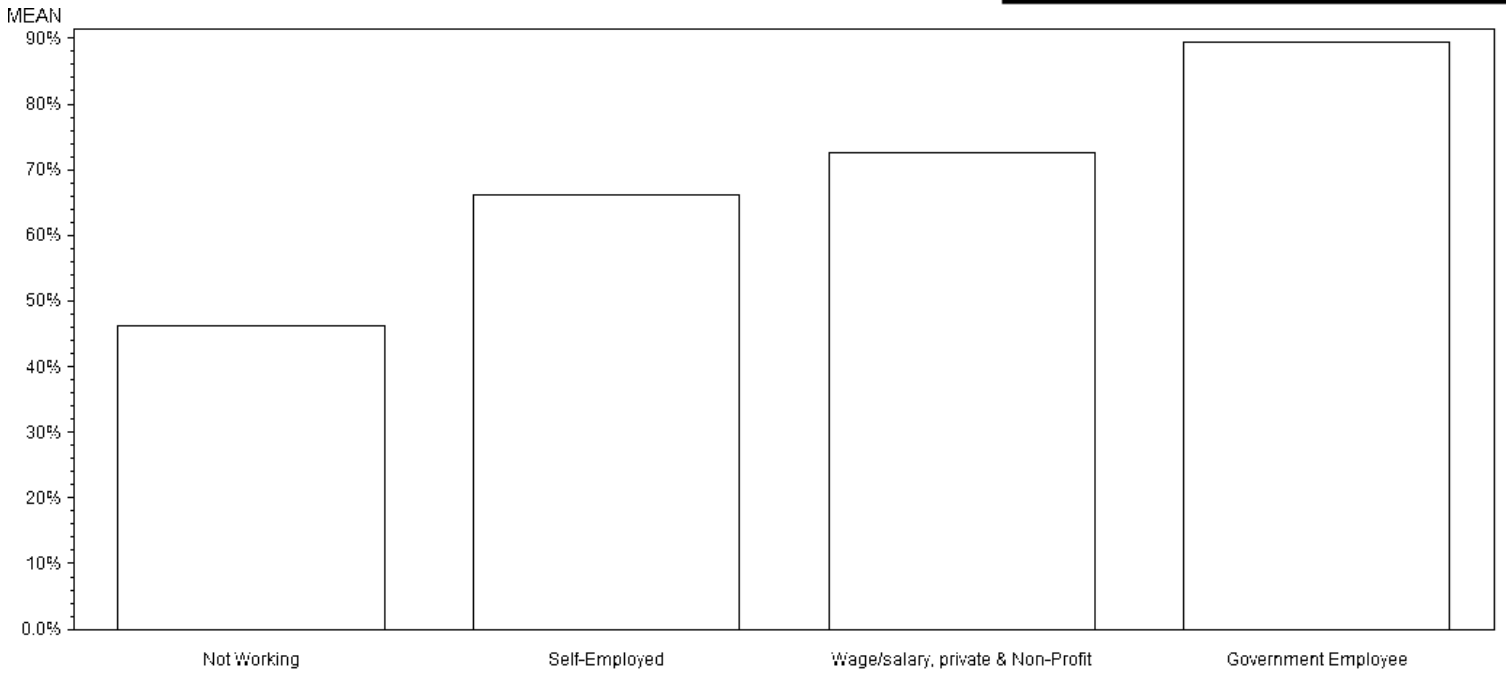
	care
	Percent
Number of children in household	
No Children	68.1%
1 child	75.1%
2 to 3 Children	76.0%
4 or more Children	60.0%
Overall	70.8%



Private Insurance holders based on Class of Work

- People in the public sector have almost a 90% Private Insurance overage. NOTE: Private care also includes employee and union provided care. Civil Servants are usually recipients of very generous government benefits, even after retirement.

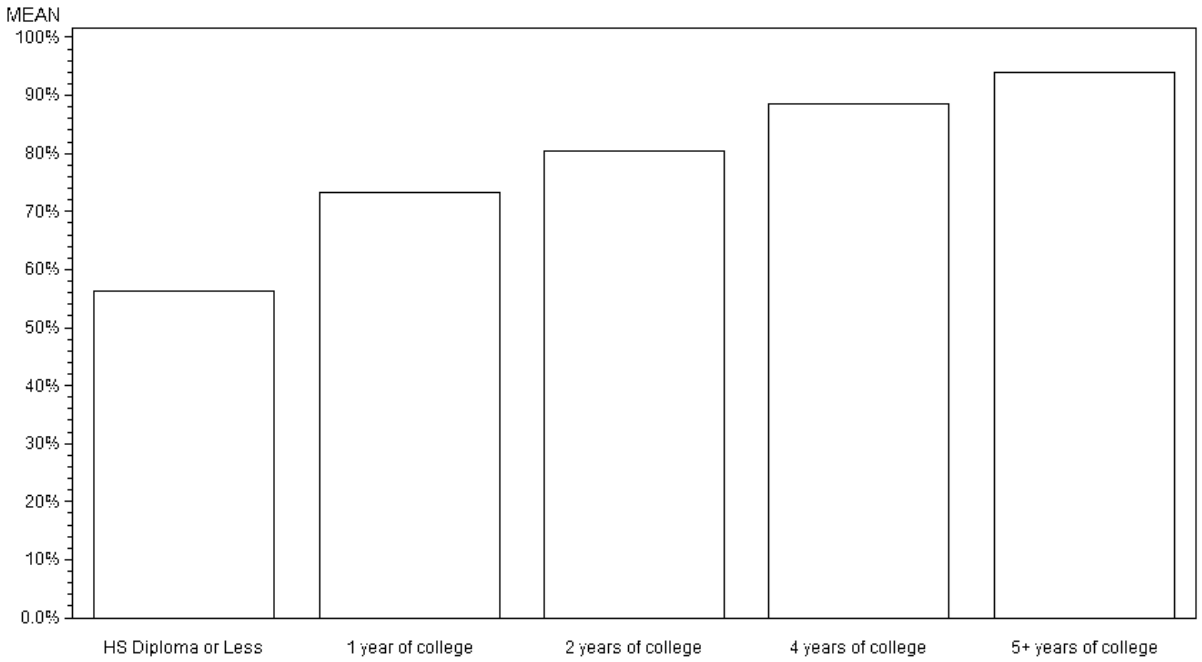
	care
	Percent
Class of worker [detailed version]	
Not Working	46.2%
Self-Employed	66.3%
Wage/salary, private & Non-Profit	72.6%
Government Employee	89.5%
Overall	70.8%



Private Insurance holders based on Education Level

- There is a clear linearity here between the amount of education acquired and the percentage f people with private care.
- People with 5+ more years would likely have a Doctorate, and thus more disposable income than someone with just 4 Years of College, not to talk of someone with a High school Diploma.

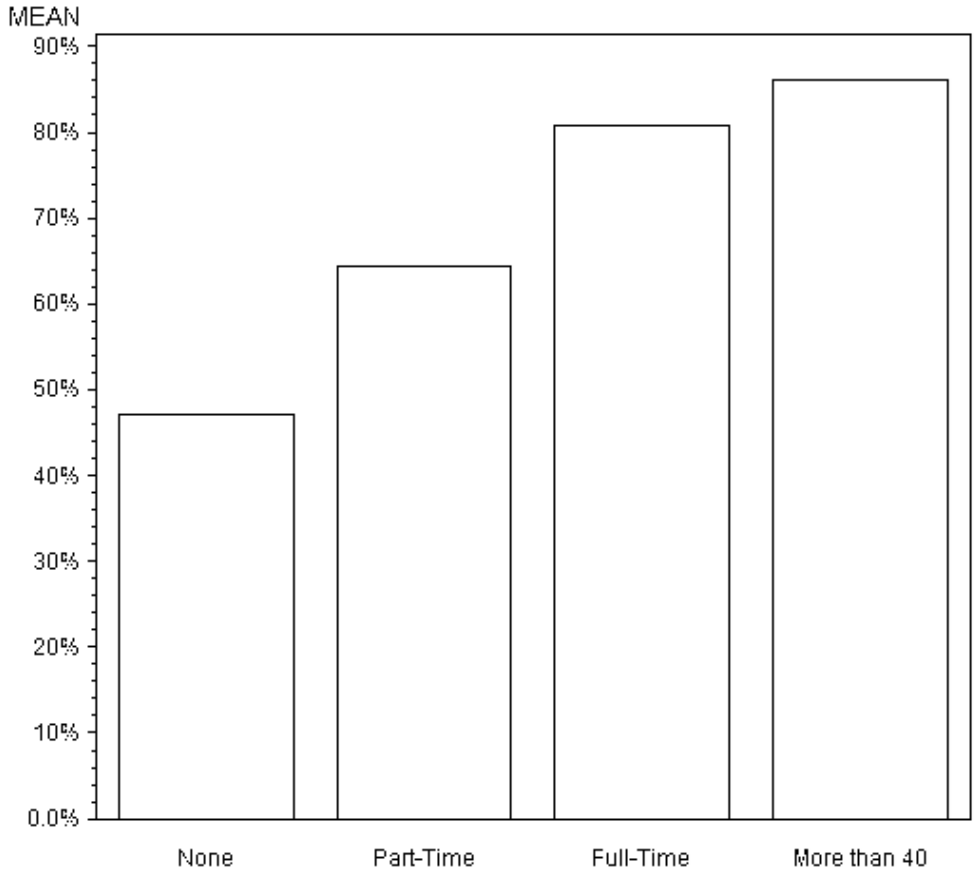
	care
	Percent
Educational attainment	
HS Diploma or Less	56.2%
1 year of college	73.2%
2 years of college	80.4%
4 years of college	88.6%
5+ years of college	94.1%
Overall	70.8%



Private Insurance holders based on Hours Worked Per week

- There is a Linear relationship between the amount of hours worked and the probability of having Private Insurance.
- People who answer could include students, and the unemployed.

	care
	Percent
Usual hours worked per week	
None	47.2%
Part-Time	64.4%
Full-Time	80.9%
More than 40	86.2%
Overall	70.8%



THE REGRESSION

The Set-up

- We have now seen all our variables, conditioned for “Care”.
- The next step to form regression. A regression seeks to inform us what the relationship is between a dependent and independent variable.
- To do this, we have to convert all the independent variables into binaries(dummies) 1 or 0.
- We also need to create baseline groups. The base line group is the group that does not get a dummy. After the binary-fication, the dummy is 0, while the others are 1.
- We would also have to impute missing values, so that it doesn’t ruin the regression. Imputation is substitution of some value for missing data. Luckily, there are no missing values in our observations.

Model: Dummy Variables

Variable	Baseline Group	Dummy Variables	Dummy Names	Imputations
Age	Baby Boomers	Generation X, Generation Y	generationx, generationy	N/A
Marital Status	Married	Separated, Divorced, Widowed, Single	marriagesing, marriagewid, marriagediv, Marriagesep	N/A
Education	HS Diploma	1 Year, 2 Years, 4 Years, 5+ Years of College	School5, school4, school2, school1	N/A
Class of Work	Private salaried Employment	Not Working, Self-Employed, Government Employed	Workclassgovt, workclassself, workclassNot	N/A
Weekly Work Hours	Full-Time	None, Part-time, Over Time	Hoursover, hourspart, hoursnone	N/A
Race	White	Black, Native American, Asian	raceAsn, racenatothr, raceblk	N/A
# of children	No Children	2 or 3, 4 to 9	Children4to9, Children2or3, Children1	N/A
Sex	Female	Male	genderm	N/A

The Logistic Regression

The LOGISTIC Procedure

Model Information

Data Set	WORK.TWO
Response Variable	care
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1878584
Number of Observations Used	1878584

Response Profile

Ordered Value	care	Total Frequency
1	1	1329510
2	0	549074

Probability modeled is care=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

The LOGISTIC Procedure

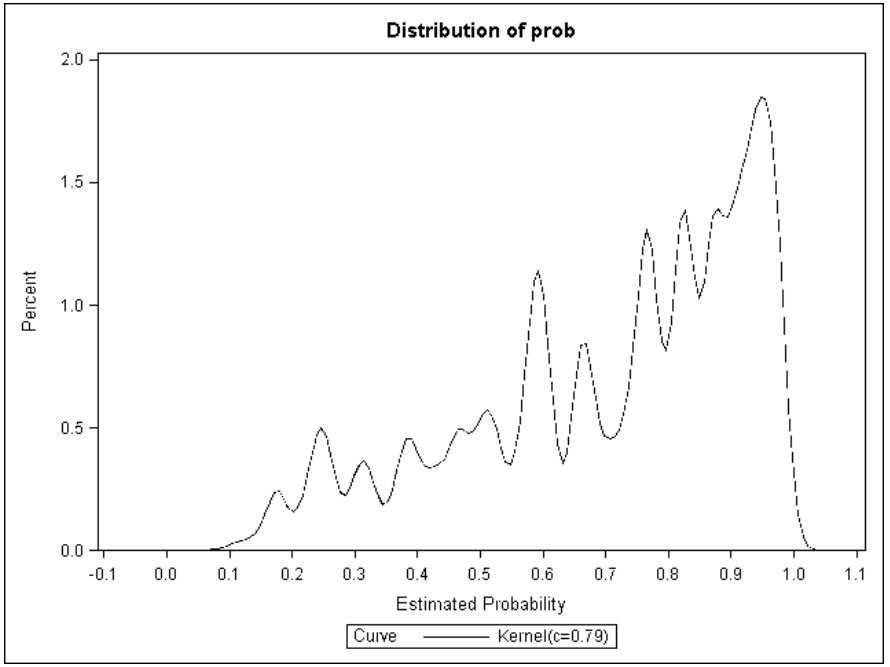
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.2552	0.0167	234.8732	<.0001
Children4to9	1	-1.5883	0.1199	175.5798	<.0001
Children2or3	1	-0.3078	0.00810	1442.7843	<.0001
Children1	1	0.3856	0.00456	7135.2039	<.0001
raceAsn	1	0.7000	0.0178	1548.4663	<.0001
racenatothr	1	0.9214	0.0159	3364.0224	<.0001
raceblk	1	0.4682	0.0166	798.3356	<.0001
hoursover	1	-0.3661	0.0615	35.4844	<.0001
hourspart	1	0.8490	0.00493	29658.6870	<.0001
hoursnone	1	-0.6676	0.00671	9912.0523	<.0001
workclassgovt	1	0.8357	0.0101	6837.3182	<.0001
workclassself	1	-0.3410	0.0376	82.2934	<.0001
workclassNot	1	0.0598	0.00694	74.2609	<.0001
genderm	1	-0.3236	0.00387	7006.4530	<.0001
school5	1	2.0543	0.0103	39738.9862	<.0001
school4	1	1.5041	0.00613	60194.6320	<.0001
school2	1	0.9470	0.00729	16871.7184	<.0001
school1	1	0.7868	0.00499	24810.5519	<.0001
marriagesing	1	-1.1929	0.00481	61465.0857	<.0001
marriagewid	1	-0.8841	0.0128	4752.3766	<.0001
marriagediv	1	-1.2473	0.00589	44800.3632	<.0001
marriagesep	1	-1.4562	0.0113	16720.6612	<.0001
generationx	1	0.0381	0.0124	9.3897	0.0022
generationy	1	-0.5391	0.0129	1744.3262	<.0001

The Univariate Procedure

- We see that the Kernel is skewed to the right, and the skewness quite linear. That is, there is a clear relationship between out variables and out data (remember we are trying to see what characteristics set care=1)

- Most of the of the observations have estimated probabilities over .7.



Analysis Variable : prob Estimated Probability				
N	Mean	Minimum	Median	Maximum
1878584	0.7077170	0.0284453	0.7670402	0.9920683

Statistical Significant Positive Estimates (Numerical Order)

Variables	Estimates
school5	2.0543
school4	1.5041
school2	0.9470
racenatothr	0.9214
hourspart	0.8490
workclassgovt	0.8357
school1	0.7868
raceAsn	0.7000
raceblk	0.4682
Children1	0.3856
workclassNot	0.0598
generationx	0.0381

- A positive estimate shows the intensity of a persons likelihood to have healthcare in this case.
- For instance, a hypothetical person with 5 years or more of Education(school5) , whose Asian (raceasn) and works part-time(hourspart) has a higher likelihood of having Private Health Insurance than supposedly someone with a different set of characteristics.

Statistical Significant Negative Estimates (Numerical Order)

Variables	Estimates
Children4to9	-1.5883
marriagesep	-1.4562
marriagediv	-1.2473
marriagesing	-1.1929
marriagewid	-0.8841
hoursnone	-0.6676
generationy	-0.5391
hoursover	-0.3661
workclassself	-0.3410
genderm	-0.3236
Children2or3	-0.3078

- A negative estimate shows the intensity of a persons likelihood NOT to have healthcare in this case.
- For instance, a hypothetical person with between 4 and 9 children(Children4to9), separated (marriagesep) and is between 18 and 29(generationy) is likely not to have private health insurance.

SIMULATIONS

- Simulations help us apply the new facts to real life. This is done by selecting individual characteristics like Age group, race, hours worked per seek…etc, and seeing the likelihood that an individual with such will have Private Care.

- We use 1s to identify specific characteristics that an individual may have, and 0s to identify characteristics that do not have. We then find the sumproduct of our 1s and our estimates.

- Because this is a Logarithm, we would need to convert it to a more understandable probability number. This is done using the exponential function in excel. (=EXP(Logarithm)).

- And then we get our Probability.

Simulation 1

Intercept	-0.2552	1	-0.2552				
Children4to9	-1.5883	0	0		Sum	Exponential	Probability
Children2or3	-0.3078	0	0		3.354	28.626	96.62%
Children1	0.3856	1	0.3856				
raceAsn	0.7000	1	0.7				
racenatothr	0.9214	0	0				
raceblk	0.4682	0	0				
hoursover	-0.3661	1	-0.3661				
hourspart	0.8490	0	0				
hoursnone	-0.6676	0	0				
workclassgovt	0.8357	1	0.8357				
workclassself	-0.3410	0	0				
workclassNot	0.0598	0	0				
genderm	-0.3236	0	0				
school5	2.0543	1	2.0543				
school4	1.5041	0	0				
school2	0.9470	0	0				
school1	0.7868	0	0				
marriagesing	-1.1929	0	0				
marriagewid	-0.8841	0	0				
marriagediv	-1.2473	0	0				
marriagesep	-1.4562	0	0				
generationx	0.0381	0	0				
generationy	-0.5391	0	0				

- A female babyboomer of Asian descent, with one child, works over time for the Government and has 5 years of school or more, has almost a 97% chance of having health care.

Simulation 2

Intercept	-0.2552	1	-0.2552			
Children4to9	-1.5883	0	0	Sum	Exponential	Probability
Children2or3	-0.3078	1	-0.3078	1.124	3.077	75.47%
Children1	0.3856	0	0			
raceAsn	0.7000	0	0			
racenatothr	0.9214	0	0			
raceblk	0.4682	1	0.4682			
hoursover	-0.3661	0	0			
hourspart	0.8490	0	0			
hoursnone	-0.6676	0	0			
workclassgovt	0.8357	0	0			
workclassself	-0.3410	0	0			
workclassNot	0.0598	0	0			
genderm	-0.3236	1	-0.3236			
school5	2.0543	0	0			
school4	1.5041	1	1.5041			
school2	0.9470	0	0			
school1	0.7868	0	0			
marriagesing	-1.1929	0	0			
marriagewid	-0.8841	0	0			
marriagediv	-1.2473	0	0			
marriagesep	-1.4562	0	0			
generationx	0.0381	1	0.0381			
generationy	-0.5391	0	0			

- A generationx married male of African descent with a 4 year degree that works full time for a private employer has a 75% probability of having private care..

Simulation 3

Intercept	-0.2552	1	-0.2552			
Children4to9	-1.5883	0	0	Sum	Exponential	Probability
Children2or3	-0.3078	1	-0.3078	-3.38	0.034	3.29%
Children1	0.3856	0	0			
raceAsn	0.7000	0	0			
racenatothr	0.9214	0	0			
raceblk	0.4682	0	0			
hoursover	-0.3661	1	-0.3661			
hourspart	0.8490	0	0			
hoursnone	-0.6676	0	0			
workclassgovt	0.8357	0	0			
workclassself	-0.3410	1	-0.341			
workclassNot	0.0598	0	0			
genderm	-0.3236	1	-0.3236			
school5	2.0543	0	0			
school4	1.5041	0	0			
school2	0.9470	0	0			
school1	0.7868	0	0			
marriagesing	-1.1929	0	0			
marriagewid	-0.8841	0	0			
marriagediv	-1.2473	1	-1.2473			
marriagesep	-1.4562	0	0			
generationx	0.0381	0	0			
generationy	-0.5391	1	-0.5391			

- A white male of generatioy who is self-employed and works over-time with no college education, divorced with 2 or 3 kids has only a 3.3% probability of having private health insurance.

CONCLUSION

- Overall, all variables seems to be important or significant in regards to the regression model. However, some extreme variables seems to stick out and have more of an influence than others.
  - School5: People with more education were more likely to have private care than people without.
  - Racenatothr
  - Babyboomers
  - Hourspart
  - Workclassgovt
  - Children4to9
  - Generationy

Information such as these can be very beneficial in helping develop public policy.