

A



# APPROACH TO WOMEN'S HEALTH SCREENING ADHERENCE FOR BREAST CANCER AND CERVICAL CANCER



Written by Kelsey Pakkala

**A FUNCTIONAL DATA ANALYSIS APPROACH TO WOMEN'S HEALTH SCREENING ADHERENCE FOR  
BREAST CANCER AND CERVICAL CANCER**

The University of Michigan  
Statistics Department  
Honors College, Class of 2015  
HONORS THESIS BY: KELSEY PAKKALA  
ADVISOR: DR. EDWARD ROTHMAN

## **ACKNOWLEDGEMENTS**

I would like to thank Dr. Edward Rothman for his guidance and support over the past year as this project came together, even encouraging my random exploratory graphs that led to the finalized model in this project. I would also like to thank Allison Roach who introduced me to this topic and inspired me to continue research for women's health.

## TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION AND BACKGROUND INFORMATION .....	P. 04
CHAPTER 2 - STATISTICAL PRINCIPLES .....	P. 06
CHAPTER 3 - BRFSS DATA AND OPERATIONAL DEFINITIONS .....	P. 08
CHAPTER 4 - FUNCTIONAL DATA ANALYSIS METHODS .....	P. 14
CHAPTER 5 - RESULTS AND CONCLUSIONS .....	P. 22
CHAPTER 6 - CHALLENGES WITH ANALYSIS .....	P. 23
CHAPTER 7 - CODE INDEX .....	P. 24
CHAPTER 8 - BIBLIOGRAPHY .....	P. 77

## **ABSTRACT**

Breast cancer and cervical cancer are two prevalent cancers among North American women; mammograms and pap smears are two screening procedures used to detect early signs of these cancers. Despite recent changes in screening recommendations for pap smears and mammography made by the United States Preventive Task Force (USPSTF), personal beliefs and physician recommendations have perpetuated earlier screening practices. This study aims to look at lifestyle predictors in order to better understand the personal side in decision making for screening program frequency. This study uses data from the Center for Disease Controls' Behavioral Risk Factor Surveillance System to analyze lifestyle factors that influence adherence to the United States Preventative Services Task Force's recommendations for mammography and pap smears. Using functional data analysis and ordinal logistic regression, analysis was performed to understand the decision-making behavior behind deciding to adhere to routine screening recommendations. Family and household dynamics were of little importance in predicting screening behavior, these variables included children in the household, marital status, other adults in the household. Significant results were found for variables related to high-risk lifestyle behaviors, wealth, health care statistics and age such as seatbelt use, smoking, income, salary, type of primary physician and healthcare status.

## **CHAPTER 1: INTRODUCTION AND BACKGROUND INFORMATION**

Despite recent changes in screening recommendations for pap smears and mammography made by the United States Preventive Task Force (USPSTF), personal beliefs and physician recommendations have perpetuated out of date screening practices. This study aims to look at lifestyle predictors in order to better understand the personal side in decision making for screening program frequency.

Breast cancer is the second leading cause of cancer deaths for North American women (DeSantis, Siegel, Bandi, Jemal 2011). In 2011, about 230,487 new cases of breast cancer were detected and another 39,520 women died of breast cancer (DeSantis, Siegel, Bandi, Jemal 2011). Fortunately, when detected in early stages treatment is possible. Mammogram screening is used to detect possible early signs of breast cancer in the hopes of reducing rates of breast cancer related deaths.

Cervical cancer is another women-related cancer prevalent in North American women. In 2010 there were approximately 12,2000 new cases of cervical cancer and 4,210 cervical cancer related deaths (USPSTF). Similar to breast cancer, there are also preventative screening measures for women to detect early signs of cervical cancer. When detected early enough, women have a 91% five-year survival rate and if detected before surrounding tissue is infected, this survival rate becomes almost 100% (The Women's Health and Wellness Resource Network). Cervical cancer is unique because detected precancerous lesions that have not yet invaded surrounding tissue can be removed, which can possibly prevent the onset of the cancer (USPSTF).

The Papanicolaou test, known as the pap smear, is a screening test where cells taken from the cervix and vagina are examined for evidence of cancer and/or precancerous lesions (The Women's Health and Wellness Resource Network). An abnormal test result may be indicative of an infection, such as

human papillomavirus (HPV), inflammation, pre-cancerous cells and/or cervical cancer (The Women's Health and Wellness Resource Network). Despite popular belief it does not screen for ovarian or uterine cancer (The Women's Health and Wellness Resource Network).

## **SCREENING RECOMMENDATIONS**

The United States Preventative Services Task Force (USPSTF) is a government agency, which makes recommendations regarding medical screening programs for citizens in the United States. The Patient Protection and Affordable Care Act requires that health plans provide coverage for all services that the USPSTF rates as an "A" or "B" (USPSTF).

The USPSTF updated its recommendations for cervical cancer in 2012. The new recommendation is for a pap smear for women ages 21 – 65 every three years regardless of sexual history and gives this screening recommendation an "A" grade (USPSTF). This screening period may be extended to five years for women ages 30 to 65 when combined with HPV testing, this recommendation also received an "A" (USPSTF).

Since 1988, most medical societies have agreed that yearly testing for women ages 21 – 65 is unnecessary (USPSTF). Even more conservative medical organizations such as the American Cancer Society uphold the USPSTF three-year screening interval or the five-year screening with HPV testing recommendation (USPSTF). Screening for women under the age of 21 and over the age of 65 is not recommended by the USPSTF and received a "D" grade (USPSTF). The exception is for women over 65 years of age who have not previously had a pap smear, have not had at least three consecutive negative pap smears, or have a medical history that requires further screening.

The USPSTF also updated its screening recommendations for mammogram screening in 2009. They recommend mammograms every two years for women ages 50 to 74, and for women under 50 years of age the decision to have a mammogram should be based on individual history (USPSTF). The USPSTF gives these recommendations an "A" grade.

This change in the starting age for mammogram screenings from age 40 to age 50 has been met with much controversy. In fact, despite the official recommendation by the USPSTF, the Affordable Care Act makes an exception to its adherence with USPSTF screening recommendations grades in the situation of mammography. For mammography, insurers and Medicare are required to cover annual mammograms, without copay, for women ages 40 and older (Pace, He and Keating 2013).

## **ADHERENCE TO SCREENING RECOMMENDATIONS**

Despite these recently updated USPSTF screening recommendations, many recent studies and publications have shown that most women are not adhering to these recommendations. Women age 40 have begun mammogram screening even though the USPSTF does not recommend the beginning of screening until age 50.

For pap smears, there has been concern raised about misinformation surrounding the test. In a recent study, it was shown that only 15% of women surveyed had a good general understanding of what pap smears were testing for. Additionally, the same survey showed that 68% of those surveyed believed that lengthened screening times would increase their odds of getting cancer. These women were then asked how likely they were to adhere to the recommended three year screening interval, approximately 57% of women indicated unlikely (Hawkins, Benard, Greek, Roland, Manninen and Saraiya 2013).

One of the potential problems with lack of adherence and understanding regarding pap smears may be the environment in which they are performed. Women generally receive their pap smear during their annual gynecological exam in conjunction with multiple other screening tests for sexually transmitted infections and overall cervical and vaginal health. Due to the number of screenings performed together, it is possible that women misunderstand which tests apply to which diseases and may therefore be hesitant to change their screening habits. In another survey 55% of women reported yearly pap smears, whereas only 16% reported adherence to the three year USPSTF screening recommendation (Sirovich and Welch 2004).

Similarly, it was shown that mammography screening recommendations are also not being observed (Pace, He and Keating 2013). In 2009, the year when the USPSTF changed its mammography screening recommendations, 84% of family practitioners surveyed reported that they recommended mammograms to their female patients ages 40 to 49 despite the change in USPSTF recommendations (Pace, He and Keating 2013). Possibly influenced by physician opinion, 89% of women also reported that they believed that mammograms for women ages 40 to 49 were necessary (Pace, He and Keating 2013). There is also a fear among women, that longer screening time intervals can increase the odds of breast cancer, similar to the unfounded belief regarding cervical cancer.

## **CHAPTER 2: STATISTICAL PRINCIPLES**

### **FALSE POSITIVES**

Over-screening is unnecessary screening usually against medical recommendations. In this case, over-screening would mean a woman getting a pap smear more than once every three years. With regards to mammography, over-screening would mean getting a mammogram more than once every two years, or starting screening before the age of 50. Over-screening is dangerous because it can lead to false positives, that is a woman who does not in fact have cancer or precancerous lesions may have an abnormal or positive test result.

It is important to note that a positive result from a pap smear may indicate infection, swelling or inflammation, HPV or indicate precancerous lesions or cervical cancer. The pap smear has a sensitivity of 0.9 (The Women's Health and Wellness Resource Network).

$$P(\text{Positive} | \text{Disease}) = 0.9$$

This can be understood as, given that a person has the disease the test will be positive about 90% of the time. This is great support for routine screening however when a test result is positive this can lead to further testing and invasive procedures. The probability of a positive or abnormal result is about 5% (Sirovich and Welch 2004).

$$P(\text{Positive}) = 0.05$$

That is about five percent of all tests are abnormal. Furthermore among women who reported having had an abnormal pap smear result about 90% sought further testing and approximately 42% received treatment or even surgery (Sirovich and Welch 2004). Many of these abnormal results, about 93% of abnormal results, are atypical squamous cells or low-grade squamous intraepithelial lesions, many of which would resolve themselves over time without treatment (Sirovich and Welch 2004). Further testing and diagnostic procedures have serious side effects which include vaginal bleeding, pain, infection, infertility and more inconclusive results. There are also personal harms such as increased anxiety, distress and financial loss for unnecessary medical procedures (USPSTF).

Breast cancer also carries many risks associated with false positive results. It similarly can lead to personal distress, anxiety and financial loss for unnecessary medical procedures.

### **LENGTH BASED SAMPLING**

A common misconception shared among women is that an increased likelihood of getting cancer with longer screening time intervals is due to misunderstanding of length biased sampling. Length biased sampling is a selection biased where certain time periods are examined for an outcome. This means that events that occur over a longer period are more likely to be chosen over a shorter period.

Latency period is the time between development of cancerous cells and symptoms appearing. Screening programs can be thought of a sample at one time period. As latency periods vary person by person based on a number of individual factors, screening programs are more likely to detect those with a longer latency period versus a shorter latency period. This can explain many women's fear of extended screening intervals. They fear that development of precancerous cells and onset of symptoms will occur within the time between screenings. Furthermore, they fear that by starting screening programs later in life could lead to a later diagnosis and potentially negative outcome. It is important to note that once symptoms appear most women seek help, so we only consider the latency period for screening.

### **MISS-SPECIFIED EVENT**

Over-screening can be dangerous with regards to false positives and negative repercussions that come from them, but there is a benefit. Pap smears and mammograms are typically done in conjunction with other routine screening programs as well as yearly exams.

While the prevalence of these two diseases is low, meaning that the probability of any woman having cervical cancer or breast cancer is relatively low, the likelihood of having some disease is much larger.



Going to mammograms or pap smears more than recommended can encourage routine screening for other diseases. Thus while a woman risks a false positive with over-screening, she may benefit from receiving other necessary routine screenings for general health, breast health, sexually transmitted infections, cervical health, etc.

$$P(\text{Any Disease}) \neq P(\text{Breast Cancer})$$

and

$$P(\text{Any Disease}) \neq P(\text{Cervical Cancer})$$

There is a difference between these two events. The first event is that a person has a particular disease (breast cancer or cervical cancer) while the second event is that a person has some disease. Thus this over-screening behavior could catch some other disease as it encourages routine screening and doctors visits.

### **CHAPTER 3: BRFSS DATA AND OPERATIONAL DEFINITIONS**

#### **DATA**

The data used in this study comes from the Behavioral Risk Factor Surveillance System (BRFSS), a yearly telephone survey conducted by the Center for Disease Control (CDC) that began in 1984 (Morbidity and Mortality Weekly Report). The BRFSS is a coordinated survey of the 50 United States, the District of Columbia and all U.S. territories; it is the longest ongoing public health telephone survey (Morbidity and Mortality Weekly Report). Many changes have occurred in survey methodology in recent years. Beginning in 2011, cell phones were included along with landlines (Morbidity and Mortality Weekly Report). This came as a result of the high prevalence of cell phone users who had systematically been excluded in previous landline only telephone surveys creating a survey bias. This data is cross-sectional, as the participants are not the same in each year.

Additionally, BRFSS data was initially weighted using a postratification method. This is a standard method that adjusts respondent data using known population proportions of age, race, sex, geographical region and other US characteristics used from the U.S. census (Morbidity and Mortality Weekly Report). The limit with this weighting method is that it relies on knowing information about the population of interest already. This can be challenging for regions where information is limited or where we are hoping to predict such information about a specific population. In 2006, the CDC switched to a raking method (Morbidity and Mortality Weekly Report). Raking relies on making individual adjustments for each variable and then sample weights are representative of the population of interest (Morbidity and Mortality Weekly Report). The CDC presents many advantages that raking has over the previously used poststratification method: problems with regions where information is unknown are minimized, more demographic variables can be considered in the weighting, more flexible method allows for more representative estimates (Morbidity and Mortality Weekly Report).

My analysis relies on the published BRFSS data from the years 2003 through 2013, excluding the year 2011 where questions regarding mammography were not asked. Therefore all models using data from

2003 to 2005 rely on poststratification weighting and later models used the raking weighting. I only examined a subset of the population in each of those years: females between the ages of 30 and 70 who had responded to questions about both their mammogram and pap smear screening behaviors.

## **OPERATIONAL DEFINITIONS**

The following section provides an explanation for all variables chosen as well as explain their operational definitions in context of later analysis. I modified many BRFSS variables and created a few of my own variables in order to create uniformity over the years and simplify my later models.

### **GENDER**

While not used in the analysis, the gender variable called 'sex' was used in subsetting the data. As this study only focuses on female health screening programs, the population of interest was females and thus males were excluded. The respondents are provided with only two response options: male or female. This variable is available and used in every year of this analysis for the purposes of subsetting.

### **AGE**

In the BRFSS survey, age is used as a continuous variable for the years 1984 through 2012 and then changed to a categorical variable in 2013. In order to have continuity across the years of my analysis I created 8 indicator age variables called  $A_1, A_2, A_3, A_4, A_5, A_6, A_7$  and  $A_8$  where  $A_1$  corresponds to women ages 30 - 34,  $A_2$  to ages 35 - 39,  $A_3$  to ages 40 - 44,  $A_4$  to ages 45 - 49,  $A_5$  to ages 50 - 54,  $A_6$  to ages 55 - 59,  $A_7$  to ages 60 - 64 and  $A_8$  to ages 65 - 69. A given women will have a one for the variable with the interval that contains their age and a zero for the remaining age variables. For example, if a women is 42 then her age variables would look as follows:

$$\{A_1 = 0, A_2 = 0, A_3 = 1, A_4 = 0, A_5 = 0, A_6 = 0, A_7 = 0, A_8 = 0\}$$

and a woman who is 55 would have the following values:

$$\{A_1 = 0, A_2 = 0, A_3 = 0, A_4 = 0, A_5 = 0, A_6 = 1, A_7 = 0, A_8 = 0\}$$

This variable is available and is used in every year of this analysis.

### **SEATBELT USE**

Habitual seatbelt use is typically indicative of a cautious person, I therefore wanted to examine whether this type of cautious behavior was associated with higher adherence to screening recommendations. The survey question is asked as "How often do you use seatbelts when you drive or ride in a car? Would you say -" and then offers the following responses: "always", "nearly always", "sometimes", "seldom", "never", "don't know/not sure", "never drive or ride in a car" or "refused". Since there are many levels of responses, I simplified this variable by coding variables as 'yes' or 'no', where 'yes' is only those who responded "always" and 'no' includes all other responses. The seatbelt question is not asked every year and is only available in the years 2006, 2008, 2010, 2012 and 2013.

## MARITAL STATUS

The original survey question simply asked the respondent “Are you” and provided the following responses: “married”, “divorced”, “widowed”, “separated”, “never married”, “a member of an unmarried couple”, “refused”. In order to simplify the number of possibilities, I created an indicator variable called marital and coded all people who responded as ‘married’ or a ‘member of an unmarried couple’ as a ‘1’, and the rest of the population as a ‘0’. If for example a woman is divorced, her marital variable would be coded as ‘0’ and if for example, a woman is a married, her marital variable would be coded as ‘1’. This variable is available in all of the years of this study.

## NUMBER OF ADULTS IN THE HOUSEHOLD

The survey asks each respondent for the number of adults in their household and then records them as follows:

- 1 ..... 1 adult in the household
- 2 ..... 2 adults in the household
- 3 ..... 3 adults in the household
- 4 ..... 4 adults in the household
- 5 ..... 5 adults in the household
- 6 - 99 ..... 6 or more adults in the household

As this is very difficult to use, I created a new variable called ‘adulthh’ that was coded as a ‘1’ if the respondent lived with any adult other than herself and ‘0’ if the adult did not live with any other adults. This variable was available to use in all years of analysis.

## NUMBER OF WOMEN IN THE HOUSEHOLD

Similar to the number of adults in the household variable discussed above, the number of women in the household has the following options for the participant:

- 1 ..... 1 adult woman in the household
- 2 ..... 2 adult women in the household

3 .....	3 adult women in the household
4 .....	4 adult women in the household
5 .....	5 adult women in the household
6 - 99 .....	6 or more adult women in the household

I created a new variable called “womenhh” which was coded as a ‘1’ if the respondent lived with at least one other adult women, or as a ‘0’ if the respondent was the only adult woman living in the household. This variable is available and used in every year of this analysis.

**PERSONAL HEALTH**

BRFSS provides an interesting variable where respondents are asked to rate their own personal health. This variable is called ‘general health’. The surveyor asks the respondent “Would you say that in general your health is” and provides the following responses: “excellent”, “very good”, “good”, “fair”, “poor”, “don’t know/not sure”, “refused”. As I was interested in the population of women who considered themselves in excellent health, I created a variable ‘G1’ where a woman would have a ‘1’ if they responded as saying they were in “excellent” general health and a ‘0’ otherwise. This variable is available and included in every year of this analysis.

**ROUTINE CHECKUP**

BRFSS includes a question regarding the length of time since the last routine checkup of the respondent. They ask “About how long has it been since you last visited a doctor for a routine checkup?”. They also clarify that a routine checkup is a general physical exam and is not for a specific injury, illness or condition. They provide the following responses: “within past year (anytime less than 12 months ago)”, “within past 2 years (1 year but less than 2 years ago)”, “within past 5 years (2 years but less than 5 years ago)”, “5 or more years ago”, “don’t know/not sure”, “never”, “refused”. Since a routine checkup is defined as within the past year, I created a variable ‘checkup’. If a respondent answered that their last checkup was “within past year (anytime less than 12 months ago)” then checkup is coded as a ‘1’ and for all other responses checkup is coded as ‘0’. This variable is available and used in all years of analysis excluding 2003 where the length of time since last routine checkup question was not asked.

**NUMBER OF CHILDREN IN THE HOUSEHOLD**

Another variable of interest was whether the respondent had children or not. I used the number of children in the household variable where the survey asks respondents “How many children less than 18 years of age live in your household”. The respondents can answer with any number between 1 and 87, 88 is coded as ‘none’ and 99 as ‘refused’. As I was only interested in whether or not the

respondent had children, I created a variable called 'childhh' which was coded as '1' if a respondent had any children and '0' if they had none. This variable is available and used every year.

## **EDUCATION LEVEL**

Education level is a BRFSS variable asked as "What is the highest grade or year of school you completed?". Respondents can choose from: "never attended school or only kindergarten", "grades 1 through 8 (elementary)", "grades 9 through 11 (some high school)", "grade 12 or GED (high school graduate)", "college 1 year to 3 years (some college or technical school)", "college 4 years or more (college graduate)", "refused". In order to simplify my model, I created an education indicator variable called 'hs'. I then assigned all those who had not completed high school as '0' and any respondent who completed high school and/or any other higher education as '1'. For example, if a women went to high school for 3 years but did not graduate, she would be coded as a '0'. If another women went to technical school, she would be coded as a '1'. This variable is available and used in every year of this study.

## **EMPLOYMENT STATUS**

The surveyor asks "Are you currently...?" and provides the following responses: "employed for wages", "self-employed", "out of work for 1 year or more", "out of work for less than 1 year", "a homemaker", "a student", "retired", "unable to work", "refused". Since this study focuses on medical screening behavior, unemployment is of most interest. Our initial hypothesis was that unemployment would mean no income and possible no health insurance, making those who were unemployed less likely to get screened. Because of this hypothesis I did not want to include homemakers, students, retirees and those unable to work as unemployed. I focused on only two of the categories "out of work for 1 year or more" and "out of work for less than 1 year". I created the variable 'employed' which was coded as a '0' if the respondent was "out of work for 1 year or more" or "out of work for less than 1 year" and as a '1' otherwise. This variable is asked by BRFSS in every year and included in all years of this analysis.

## **INCOME**

The United States defines the poverty line around \$20,000 depending on a number of factors: number of children, number of income earners, etc. I was interested in the relationship between poverty and medical screening. I used the 'income level' variable. Respondents were asked "Is your annual household income from all sources: " and gives the following interval choices: "less than \$10,000", "less than \$15,000 (\$10,000 to less than \$15,000)", "less than \$20,000 (\$15,000 to less than \$20,000)", "less than \$25,000 (\$20,000 to less than \$25,000)", "less than \$35,000 (\$25,000 to less than \$35,000)", "less than \$50,000 (\$35,000 to less than \$50,000)", "less than \$75,000 (\$50,000 to less than \$75,000)", "\$75,000 or more", "don't know/not sure", "refused". The closest cut off to approximate \$20,000 poverty line is the "less than \$20,000 (\$15,000 to less than \$20,000)" response. I created an indicator variable for income called 'income'. If a person responded with "less than \$10,000", "less than \$15,000 (\$10,000 to less than \$15,000)" or "less than \$20,000 (\$15,000 to less

than \$20,000)" 'income' was recorded as a '1' and for all others 'income' was recorded as a '0'. This variable is available and used in every year of this analysis.

## **PRIMARY DOCTOR**

Since doctors are a huge driver behind people getting screened, it is important to look at the type of primary doctor that each respondent sees. I used the 'CSRVDOC1' variable where the respondents are asked "What type of doctor provides the majority of your health care?". They are given many response options: "cancer surgeon", "family practitioner", "general surgeon", "gynecologic oncologist", "general practitioner, internist", "plastic surgeon, reconstructive surgeon", "medical oncologist", "radiation oncologist", "urologist", "other", "don't know/not sure", "refused". As this study focuses on routine cancer screening programs I focused on those who primarily saw a family practitioner (general doctor) or oncologist (grouping all different specialties together). I created three indicator variables for doctors: 'familypractitioner', 'oncologist', 'specialist/other'. The 'familypractitioner' variable is only coded as '1' if the respondent indicated that they received the majority of their care from a family practitioner, else it is coded as a '0'. The 'oncologist' variable is coded as a '1' if the primary care doctor a respondent sees is a gynecological oncologist, medical oncologist or radiation oncologist, else it is coded as a '0'. The 'specialist/other' variable is coded as a '1' if the respondent indicated as their primary care doctor a cancer surgeon, general surgeon, general practitioner, internist, plastic surgeon, reconstructive surgeon, urologist or other. If for example a women is seeing a medical oncologist for the majority of their care then her variables would be coded as follows:

$$\{ \text{familypractitioner} = 0, \text{oncologist} = 1, \text{specialist/other} = 0 \}$$

This variable is only available and used in the year 2012.

## **SALARY**

Another variable of interest is whether a person is paid hourly or by salary for their job, this was used as an indicator for blue collar versus white collar employees. To look at this variable, I used the BRFSS variable 'how are you generally paid for the work you do'. The respondent is asked "At your main job or business, how are you generally paid for the work you do. Are you:" and they are given the following options: "paid by salary", "paid by the hour", "paid by the job/task (e.g. commission, piecework)", "paid some other way", "don't know/not sure", "refused". I then created a variable called 'salary'. If a person was paid by salary the salary variable was coded as '1', else the variable was coded as '0'. This variable is only available in the years 2010 and 2012.

## **HEALTHCARE**

Healthcare is a major indicator of routine health screenings. I looked at the BRFSS health care coverage variable. This question was only asked to respondents ages 18 - 64, so the final age group 65 - 69 do not have this variable. Respondents are asked to indicate whether they "have health coverage", "do not have health coverage" or "don't know/not sure". I used this variable as it was asked. It is available and used in every year of this study.

## **OWN (RENT)**

Owning or renting a home may not be an indicator of wealth. This is because owning a trailer park is considered owning and thus it is possible that a large portion of the homeowners in this survey were mobile home owners. I therefore wanted to look at this independently of salary as it is possible that low salary earners could also be owners instead of renters. Respondents are asked "Do you own or rent your home? Home is defined as the place where you live most of the time/the majority of the year". The following responses are available: "own", "rent", "other arrangement", "don't know/not sure", "refused". I created a new variable called 'own' which was coded as '1' if the respondent owned their home and '0' otherwise. This variable is only available and used in 2012 and 2013 for the study.

## **SMOKE**

Smokers may be perceived as less concerned about their personal health and may therefore not adhere to the routine screening recommendations. I used a BRFSS computed smoking status variable which used four questions to decide whether the respondent was a "current smoker - now smokes every day", "current smoker - now smokes some days", "former smoker", "never smoked" or "don't know/refused/missing". I created my own computed 'smoker' variable that was coded as '1' if the respondent was a current smoker regardless of smoking every day or some days and '0' otherwise. This variable is available and used in every year of this analysis.

## **NUMBER OF TESTS**

As the response variable of interest in the following test relies on the number of screening programs to which a woman had done on a routine basis, a variable called 'routine' was created. This was a quantitative ordinal variable that was coded as a '0' if a respondent had not had a routine mammogram or a routine pap smears as defined by the USPSTF recommendations, as a '1' if a respondent had had either a routine mammogram or a routine pap smear as defined by the USPSTF recommendations, or as a '2' if the respondent had had both a routine mammogram and a routine pap smear as defined by the USPSTF recommendations. The BRFSS variables used in the calculation of this variable were 'HADMAM', 'HOWLONG', 'HADPAP' and 'LASTPAP'. For the first two variables ('HADMAM' and 'HOWLONG'), respondents were first asked "A mammogram is an x-ray of each breast to look for breast cancer. Have you ever had a mammogram?" and provided with the following responses "Yes", "No", "Don't know/not sure" and "Refused". The surveyor then followed up, if they answered "Yes", by asking "How long has it been since you had your last mammogram" with the following response choices: "Within the past year (anytime less than 12 months ago)", "Within the past two years (1 year but less than 2 years ago)", "Within the past 3 years (2 years but less than 3 years ago)", "Within the past 5 years (3 years but less than 5 years ago)", "5 or more years", "Don't know/not sure" or "Refused". For the pap smear questions, respondents were first asked "A Pap test is a test for cancer of the cervix. Have you ever had a pap test?" and provided with the following responses: "Yes", "No", "Don't know/not sure" or "Refused". If the respondent answered "Yes" to the

having had a pap smear, the surveyor asked “How long has it been since you had your last Pap test?” with the following response options: “Within the past year (anytime less than 12 months ago)”, “Within the past 2 years (1 year but less than 2 years ago)”, “Within the past 3 years (2 years but less than 3 years ago)”, “Within the past 5 years (3 years but less than 5 years ago)”, “Don’t know/not sure” or “Refused”.

The mammogram and pap smear variables from which this was computed are available for every year of this analysis and are used to compute this ‘routine’ response variable.

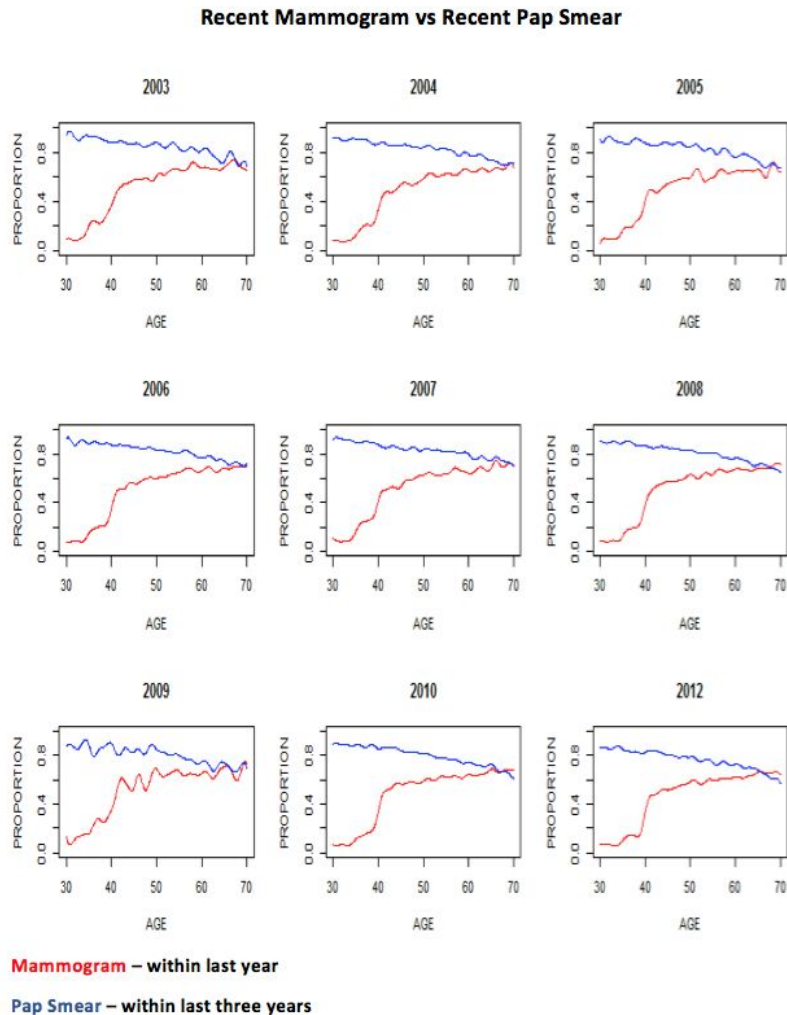
#### **CHAPTER 4: FUNCTIONAL DATA ANALYSIS METHODS**

##### **MODEL**

My analysis relied on the use of smoothing splines created using the ‘Functional Data Analysis’, abbreviated FDA, package in R. After calculating the proportion of women who had had a recent mammogram at ages 30 to 70 as well as the proportion of women who had had a recent pap smear at ages 30 to 70, I create a functional object that attempted to connect the proportions into a curve



while also penalizing for extremities using the raw curves own derivative.

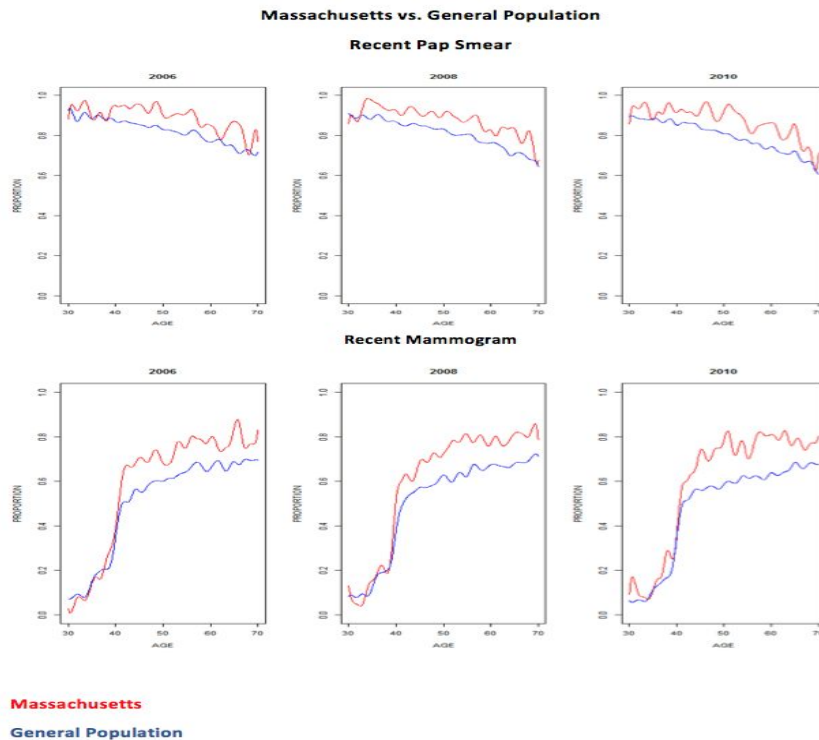


Graph 1: The graph above shows the proportion of women who have received a pap smear within the last 3 years by age with the blue line and the proportion of women who have received a mammogram within the last year by age with the red line for the years 2003 through 2012.

The image shown above is a plotted image of these functional objects from 2003 to 2012. The red line represents the proportion of women at a given age who have had a mammogram within the last year and the blue line represents the proportion of women who have had a pap smear within the three years. The proportion of women who received a routine pap smear appears to decrease with age, which indicates adherence to the USPSTF recommendations to stop screening at the age of 50. The proportion of women who received a routine mammogram appears to increase with age, which is consistent with the later start screening age for mammography versus a pap smear. However, it is important to note that even though the USPSTF recommendation is for first screening at the age of 50, all years show a sharp increase in screening at the age of 40 indicating the women are likely not adhering to these recommendations. As mentioned above the Affordable Care Act makes an exception to its adherence to USPSTF recommendations with regards to mammography, and

guarantees full coverage on mammograms for women 40 years and older. It therefore makes sense that there is a spike at 40 years of age despite official recommendations being for a later start.

As is suggested above, health coverage appears to be a strong predictor of screening adherence. Thus, we also created functional objects to compare screening rates among citizens in the United States versus citizens in Massachusetts for the years 2006, 2008 and 2010. These years were chosen as Massachusetts had a public healthcare reform passed in 2006, similar to the current federal Obamacare healthcare reform that began in 2012, that aimed to provide health coverage for citizens who could otherwise not afford it.



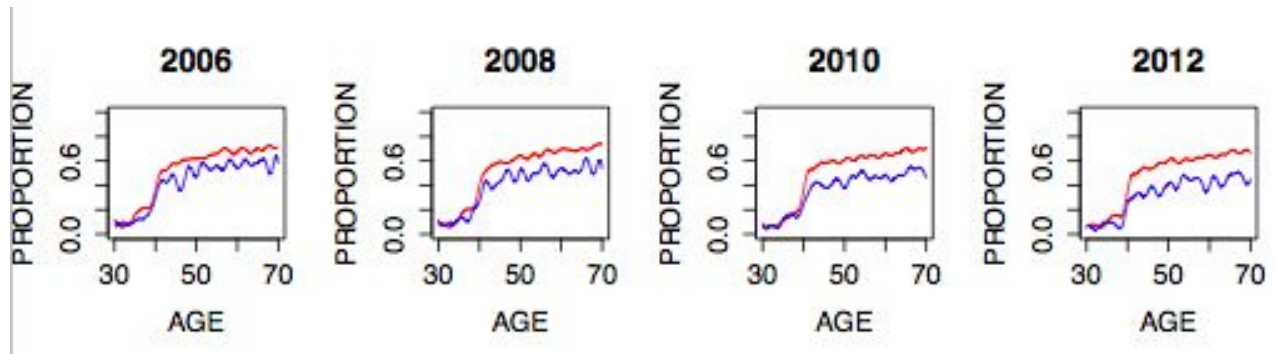
Graph 2: The graph above is a comparison between screening adherence for women in Massachusetts versus the rest of the general United States population in the years 2006, 2008 and 2010. The top three graphs are of proportion of women who have had a pap smear within the last three years by age and the bottom three graphs are for the proportion of women who have received a mammogram within the last year by age. The Massachusetts population is denoted by the red line and the rest of the United States general population is shown by the blue line.

In the graphs above, Massachusetts is denoted by the red curve and the rest of the United States general population by the blue curve. It is very clear that the proportion of those citizens getting routine screening for both mammography and pap smears is higher across all ages in Massachusetts versus the rest of the United States. This is very suggestive of health care increasing the likelihood of a woman adhering to routine screening recommendations.

Similar analysis using FDA smoothing splines was done to compare another subset of the population: seatbelt users versus non-seatbelt users. This was done in order to determine if other lifestyle predictors, such as high risk lifestyles (not wearing a seatbelt), could be associated with lower

screening rates. The graph below shows the results. Seatbelt users, as described above as those people who always wear seatbelts, are denoted with a red line and non-seatbelt users are denoted by the blue line.

**Routine Mammogram Screening Among Seatbelt Users vs Non-Seatbelt Users**



*Graph 3: The graph above shows the proportion of women who have had a mammogram in the past year by age. The red line denotes those who were recorded as always wearing a seatbelt and the blue line denotes those who did not answer that they always wore a seatbelt.*

There is clearly a higher proportion of women who wear seatbelts getting routine screening as compared to those in the non-seatbelt category. This is highly suggestive of certain risk based behaviors as potential predictors of lower screening rates. Thus smoking status and seatbelt use are included in later regression analysis as predictors of high risk behaviors.

To further explore this relationship, ordinal logistic regression was performed using STATA statistical software. Ordinal logistic regression creates a probability density curve for the response variable and then generates cut-off points. This model is difficult to express explicitly, and cut-off points are usually used in interpretation of results. These cut-off points represents the distinction between different levels. The response variable I used was the number of tests 'routine' variable, either '0' '1' or '2', described above in Chapter 3 'Operational Definitions' section. In this regression, there are two cut-off points generated. The area to the left of the first cut-off points represents the probability distribution of the first level, in this case 0 screening tests performed. The area between the two cut-off points represents the probability distribution of the second level, in this case 1 of the two screening tests performed. Finally, the area to the right of the second cut-off represents the probability distribution of the third level, in this case both of the screening tests are performed. the image below is a visualization of such a model.

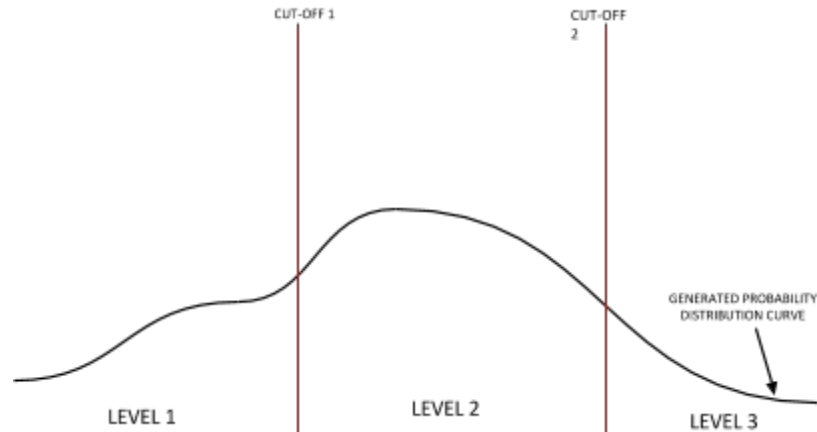


Figure 1: The above image is a visualize representation of the probability density curve generated by ordinal logistic regression as described above. This ordinal logistic regression has three levels and thus generates two cutoff points separating the probability density into the three distinct levels with the probability under the curve contained each area as the probability of a given level.

The predictors were: age, seatbelt use, marital status, number of adults in the household, number of women in the household, personal health, routine checkup, number of children in the household, education level, employment status, income, primary doctor, salary, healthcare, own and smoke (all described in the above 'Operational Definitions' section of Chapter 3). The results are shown below.

Variable / Odds Ratio	2003	2004	2005	2006	2007	2008	2009	2010	2012	2013
30-34	BASE LEVEL	BASE LEVEL	BASE LEVEL	BASE LEVEL	BASE LEVEL	BASE LEVEL	BASE LEVEL	BASE LEVEL	BASE LEVEL	BASE LEVEL
35-39	1.447	1.283	1.371	1.312	1.345	1.262	1.21	1.179	0.537	1.082
40-44	3.309	2.725	3.184	3.25	2.843	30.083	2.797	3.121	0.967	2.407
45-49	3.929	3.638	4.215	4.2562	3.843	3.869	3.522	3.768	1.265	2.648
50-54	4.818	4.263	4.109	4.129	4.102	4.032	4.326	3.656	0.818	2.514
55-59	4.876	4.078	3.987	4.271	3.944	3.75	3.359	3.271	1.771	2.43
60-64	3.39	4.157	3.801	3.841	3.535	3.451	3.085	3.042	1.148	2.168
65-69	4.112	3.608	2.882	3.186	3.408	2.96	2.739	2.675	1.287	1.783
Seatbelt	N/A	N/A	N/A	1.345	N/A	1.761	N/A	1.949	3.68	2.638
Marital Status	0.877	0.807	0.848	0.844	0.8442	0.835	0.901	0.854	0.594	0.822
Adults in Household	0.905	0.924	0.933	1.018	1.008	1.017	0.99	0.967	0.596	1.071
Women in Household	0.972	0.935	0.921	0.928	0.844	0.927	0.939	0.906	1.956	0.937
Excellent Health	1.064	1.029	1.146	1.108	1.1	1.159	0.934	1.169	1.64	1.231
Routine Checkup	N/A	0.976	4.127	4.249	4.055	3.861	3.531	3.653	2.39	3.184
Children in Household	0.983	0.976	0.992	0.977	0.982	0.97	0.992	0.971	1.013	1.018
High School Graduate	1.358	1.112	1.254	1.202	1.127	1.121	1.204	1.081	2.084	1.358
Employed	0.898	1.131	1.097	0.883	0.912	1.178	1.244	1.138	1.417	0.953
Income More than \$20,000	1.156	1.196	1.135	1.111	1.206	1.031	1.205	1.173	1.159	0.936
Family Practitioner	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	BASE LEVEL	N/A
Oncologist	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2.01	N/A
Specialist / Other	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.904	N/A
Salaried	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1.061	2.004	1.335
Healthcare	2.2106	2.354	2.2	1.936	1.881	1.856	1.8	2.057	2.144	2.004
Own Home	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.951	1.242
Smoker	1.492	1.517	0.667	0.643	0.672	0.616	0.632	0.613	0.669	1.546
Cutoff 1	0.84	0.893	1.618	1.649	1.255	1.986	1.633	2.446	2.508	7.93
Cutoff 2	11.751	12.24	25.13	25.877	18.725	27.967	17.662	34.27	20.895	51.064

Figure 2: The image above is of the output odds ratio for ordinal logistic regression model. P-values were not reported as the majority of the variables were reported as significant due to the large sample size used. Those variables highlighted in red are those variables that are significant predictors (differing greatly from 1 - no difference) over multiple years. The odds ratio can be interpreted as follows: in 2013 women who wore a seatbelt were 2.638 times more likely to get screened compared with non-seatbelt wearing women.

Values highlighted in red represent variables who have large odds ratios over multiple years. An odds ratio can be interpreted as follows: in 2013 women who wore a seatbelt were 2.638 times more likely to get screened as compared to those who did not.

The age variable odds ratio increases with age. This makes sense as older women are more likely to get screened as compared to their younger counterparts (women aged 30-34 were used as the base comparison). The odds ratio for women over 60 years of age tends to deviate from the pattern, however this can be explained by the smaller sample size of women in this age group. The year 2012 also deviates from this pattern. Many factors could explain this: Obamacare coming into effect, changes in USPSTF screening recommendations for pap smears, changes in sampling methods in the previous year, smaller sample size as this question is not asked in every region of every women but asked on a rotational basis.

Seatbelt usage had an average odds ratio around 2 for any given year that the variable was included in the model. This agrees with the results above that suggest that seatbelt users have a higher screening proportion as compared to those who do not wear seatbelt. Specifically, in 2012 women who wore seatbelts were 3.68 times more likely to get screened as compared to non-seatbelt users, this is the highest odds ratio for seatbelt users. The lowest odds ratio, 1.345, is in 2006.

Marital status hovers around 0.8 and 0.9 for all years of the study. This implies that there is little difference in the screening behaviors of women who are married versus those who are unmarried. This is also true for adults in the household and number of women in the household for which the odds ratio hovers around 1 for both variables. This suggests that having another adult in the household or having another female adult in the household does not increase the odds of getting screened. The only odds ratio to deviate from this pattern is in 2012, where the odds ratio falls to 0.594. As discussed above, there are many possible explanations for deviations in 2012.

Surprisingly women who rated themselves as in excellent health were slightly more likely than those who rated themselves as in less than excellent health to get screened. This is interesting, as this variable was expected to be lower than one. We had expected those who considered themselves in excellent health to be less preoccupied with routine screening as opposed to those who considered themselves as in less than excellent health.

As expected, the odds ratio for routine checkup was greater than one for all years after 2004. This implies that women who are going to yearly checkups are more likely to get screened routines as opposed who those who do not. This variable is also indicative of cautious behavior, like smoking and seatbelt use. This also supports the earlier discussion of mis-specified event. There is a benefit to screening as it encourages routine doctor visits and we are seeing that women getting routine checkups are more likely to get their routine mammogram and pap smear.

Another variable that was not a strong predictor of screening was number of children in the household. The odds ratio for this hovered around 1, meaning that having a child in the household did not increase the likelihood of getting screened.

High school education had an increased positive odds ratio ( $> 1$  in all years). This means that women who graduated from high school or received equivalent high school graduate status were more likely in all years to get screened as compared to women who did not complete their high school education. This could be because those with a higher level of education may earn more money or work in white collar and thus be able to afford to pay for visits and take time off of work. The largest odds ratio for education is in 2012, 2.084. This can be interpreted as women who received at least a high school education or equivalent were on average 2.084 times more likely to get screened as compared to those who did not complete a high school level education.

Employment status was another variable that we expected to have an odds ratio greater than one. This is because being employed is likely correlated with higher income and having healthcare. Yet despite these expectations the odds ratio for employment hovers around 1. The lowest odds ratio is 0.883 in 2006 and the highest is in 1.417 in 2012. There are multiple explanations for these deviations: the economic crisis of 2007 and 2008. As the economy crashed, routine screening probably became less of a priority or potentially missing work was not worth the benefits of screening and we see lower odds ratios during this time. Subsequent odds ratios begin to rise following recovery of the economy. Other possible explanations include the changes to sampling method and sampling weighting discussed above in Chapter 3 'BRFSS Data and Operational Definitions'.

Similarly, annual income was predicted to have increased odds ratio. This is due to the A. Maslow Hierarchy of Needs. We would expect those only those who have their basic needs met (food, shelter, etc.) to try to meet secondary or tertiary needs. As income is a means to purchase basic needs, we would expect those below the poverty line to focus on achieving their basic needs and likely neglect secondary needs such as routine screening. This relationship is shown by the odds ratios for income. Those living above the approximated poverty line (in this study-more than \$20,000 annually) were on average slightly more likely (on average 1.1 times more likely) to get screen as compared with those who had an annual income less than \$20,000. It is important to notice that the odds ratio is only slightly above one so it is not surprising or contradictory to have the odds ratio in 2013 to dip below 1 to 0.936.

As method of payment is another variable that could differentiate blue collar and white collar workers, the salaried variable was included in the model. A salaried person receives a set annual income and does not received compensation for overtime, however are more likely to have a better healthcare plan as compared to those paid hourly. Contrastly, those paid hourly are paid for overtime and thus have more set hours and possibly more time outside of work to go get screened. The salaried variable was only available and used 2010, 2012 and 2013 with odds ratios 1.061, 2.004 and 1.335 respectively. These odds ratios are all positive indicating an increased likelihood of screening for those who are salaried as compared to those paid hourly. It is difficult to determine if this odds ratio is significantly higher or not as the one of the three years is essentially 1 (no difference) and the following two years are significantly above 1.

Primary care physician was a variable of interest in this study as previous research indicated that physicians strongly influenced a patient's decision to get screened routinely or possibly over-screened. Unfortunately this variable is only available for analysis in 2012. Family care practitioner is used as the base and thus those seeing an oncologist of any type or a specialist are compared to those who were seeing a family practitioner as their primary care physician. As would be expected, women seeing an oncologist were twice as likely to get screened as compared to family practitioners. This makes sense as a person seeing an oncologist implies either a prior history of cancer or a higher risk for cancer. Those seeing a specialist or 'other' were 0.904 times as likely to get screened as compared to a family practitioner. This odds ratio is pretty close to 1 and indicates very little difference between screening practices in these two groups. This variable also includes many less commonly seen doctors and thus the small sample size could have been a barrier in analysis.

As discussed above in the comparison of Massachusetts with the general United States population, there is increased screening proportion for the years during which Massachusetts had public healthcare and the rest of the United States had not yet initiated Obamacare. The healthcare variable was associated with increased screening proportions in earlier analysis and thus included in the model. It was of no surprise that this variable had a large odds ratio across all years of analysis with the highest being 2.2106 in 2003 and the lowest being 1.8 in 2009. This can be understood as having health care increases the likelihood of getting screening 1.8 - 2.2 times across all years of the study.

Homeownership is typically associated with wealth, however homeowners includes those who own trailers and portable homes. This variable was included because I was interested in whether this would mean an increased odds ratio influenced by stereotypically wealthy home owners or if the trailer and portable homeowners would actually lead to a lower odds ratio for routine screening adherence. This variable was only available in the study for the years 2012 and 2013. Unfortunately the results are inconclusive as in 2012 the odds ratio was 0.951 (implying no change) and was 1.242 in 2013. Thus it is difficult to determine if homeownership has an increased or decreased odds ratio for screening. This does make sense because homeowners includes mobile homeowners and more permanent housing arrangements.

The final variable included in this analysis was smoking status. This was primarily due to the assumption that smokers were likely high risk individuals. Recent advertisements have made the unhealthy side effects of smoking common knowledge and those who currently smoke do so with full knowledge of the health repercussions. It would make sense that this type of person would be less concerned with following routine health screening procedures as they are already defying public health advertisements warning against smoking. This is clearly reflected in the odds ratios for 2005 - 2012 which hover around 0.6. This means that smokers are 60% as likely as nonsmokers to get screened. The deviations to this pattern occur in 2003, 2004 and 2013. As the values are not far from 1, this could be due to sampling bias (participants vary every year) or response bias (responses are self-reported and there is a taboo associated with smoking).

Finally, it is important to consider the meaning of the cutoff points. If two points are very close together, this would imply very few people are getting only one test and thus implying that the

response ('0' '1' or '2') may not be the best predictor. Fortunately, for this model this does not seem to be the case. The cutoffs are very different with the largest difference being in 2013 (7.93 to 51.064) and the smallest being in 2003 (0.84 to 11.751). This implies that there are many people who are only getting screened for one but not both of the tests. As shown above, screening decreases for women after age 50 for pap smears and increases for women at age 40 for mammograms (see Graph 1). Therefore we would expect many women to fall into the category of one test as they transition from the pap smear screening time frame (about 20 years of age to 50 years of age) to the mammogram screening time frame (about 40 years of age onward).

## **CHAPTER 5: RESULTS AND CONCLUSIONS**

Preliminary investigation revealed a downward trend by age for having a pap smear within the past three years over the years 2003 through 2012 and an upward trend by age for having a mammogram within the last year over the years 2003 through 2012. Further analysis showed that the proportion of women who consistently wore a seatbelt had higher screening proportions across all ages as compared to those who did not consistently wear a seatbelt. Analysis was also done to compare Massachusetts with the general United States population in the years 2006, 2008 and 2010. This was because public health care began in Massachusetts in 2006 as compared with Obamacare which began in 2012. This enabled us to see the effect of public health care on screening adherence and predict what will happen within the next few years with regards to Obamacare. This analysis revealed that women in Massachusetts in these years had higher screening rates across all ages for both tests as compared with the rest of the general United States population. This suggests that in the next few years screening rates for mammograms and pap smears are likely to increase nationwide as Obamacare comes into full effect.

Family and household dynamics turned out to be of least importance in terms of predicting screening behavior. In fact, marital status, homeownership status, living with another adult, living with another adult female or having children barely if at all changed the odds of getting screened compared to their counterparts. This is surprising as initial hypothesis were that having a family or spouse would increase likelihood to take care of oneself for the sake of the family's overall well being. Furthermore, we had expected that living with an adult or another adult female would create a strong influence to get screened, especially having another adult female who could relate to the benefits of getting routinely screened for female related diseases. With homeownership, there is a large dichotomy in the two types of persons in this group: wealthy homeowners and portable home owners. Additionally this question is limited to two years and thus could be influential if there was more data available and possible more information about the type of home owned.

High-risk lifestyle behaviors, wealth and health care characteristics appear to be the primary drivers behind health screening adherence. High-risk lifestyle behaviors include: non-seatbelt users and smokers; both of which had significant odds ratios for decreased likelihood for screening. Wealth variables included: high school education, employment status, annual income and salary. Again these variables had odds ratios that deviated from 1 and those who had received a high school education or equivalent, were employed with a salaried annual income greater than \$20,000 were much more



likely on average to get routine screening as compared with their counterparts. The final category, healthcare characteristics, includes self-evaluated as having excellent health, yearly check-up status, having health care coverage and type of primary care doctor. These again were highly suggestive of increased odds of getting screened. The most surprising of these variables was the self-evaluated health, as we had predicted that those who self evaluated themselves as in excellent health would be less concerned with routine screening as compared with those who had considered themselves in less than excellent health. The final variable with increased odds ratio for screening was age. This variable was not grouped with the others as it makes sense alone especially for breast cancer.

The distinct cutoffs tell us that there is a significant portion of the population only receiving one screening test: the area between the cutoffs is significant in size. This is consistent with the screening behavior shown in Graph 1 and the screening recommendations discussed in Chapter 1 'Screening Recommendations' section. The United States Preventative Services Task Force recommends pap smears beginning at age 21 through age 50 with stopping occurring after 3 consecutive negative pap smears. The United States Preventative Services Task Force recommends screening beginning at age 50 and onward, however as shown in the 'Chapter 5' screening onset actually begins at age 40.

Mammography and pap smears are two common tests that screen for early signs of breast and cervical cancer, two dangerous types of female related cancer. With early intervention both of these diseases have a high five year survival rate. Thus these tests provide many benefits. This study aimed to understand the decision making process behind a woman deciding to get screened through these two methods. Many individual variables were considered and in the end it was found that increase age, low-risk lifestyle factors, wealth and health care characteristics (such as having health care coverage) were strong predictors of screening behavior; whereas, family and home dynamics appeared less influential.

## **CHAPTER 6: CHALLENGES WITH ANALYSIS**

A drawback to this survey is the type of data available. Since the study is cross-sectional, the respondents are different each year and thus longitudinal conclusions cannot be made. If the data were longitudinal, it would be interesting to look at how a persons screening decisions change over time and see if certain lifestyle variables become more influential at certain ages as compared to other. Additionally, since pap smear screening decreases with age and mammogram screening increases with age, a look overtime would be better as to determine if a women was adhering to the screening time line as provided by the United States Preventative Services Task Force.

Furthermore, this data is self-reported. Interviewers ask respondents questions over the phone but there is not a way to verify if the information provided is truthful or not. In fact, a certain amount of response bias is expected. Questions regarding smoke status may be viewed as taboo by some and thus not be answered truthfully. Additionally, certain insurance companies provide incentives for healthy behavior and even though the survey is anonymous, a respondent may fear losing such incentives by providing an answer inconsistent with adherence to the proscribed healthy behavior.

The Behavioral Risk Factor Surveillance System also changes its survey design over the time periods included in this study. Cell phone recipients are not included in the survey until 2011. It is known that the prevalence of cell phone users increased dramatically up to this time and thus many cell phone only based households were systematically excluded up until 2011, creating a sampling bias. Additionally, sample weighting methods were modified in 2006 from a post-stratification method to a more flexible raking method. While this was deemed more robust by the Center for Disease Control, it calls into question possible challenges to analysis with the data weighted with the post-stratification method as the Center for Disease Control later claimed that this method was not rigorous enough.

Not included in this study was analysis on the people who only had one screening test done. It would be advantageous for future research to focus on this group and attempt to understand which of the tests is being done. Since pap smears decrease with age and mammograms increase with age, it would be interesting to observe which screening program was more prevalent among this one screening group level as this could indicate the type of women in the group (older versus younger).

## **CHAPTER 7: CODE INDEX**

### **R CODE**

#### **# Converting CDC Data from SAS to CSV File**

```
install.packages("foreign")
install.packages("Hmisc",dependencies=TRUE)
library("foreign")
library("grid")
library("lattice")
library("splines")
library("Formula")
library("Hmisc")
library("survival")
```

```
# 1987
```

```
> xpt = sasxport.get("CDBRFS87.XPT")
```

```
> write.csv(xpt,file="1987.csv")
```

```
# 1988
```

```
> xpt = sasxport.get("CDBRFS88.XPT")
```

```
> write.csv(xpt,file="1988.csv")
```

```
# 1989
```

```
> xpt = sasxport.get("CDBRFS89.XPT")
```

```
> write.csv(xpt,file="1989.csv")
```

```
# 1990
```

```
> xpt = sasxport.get("CDBRFS90.XPT")
```

```
> write.csv(xpt,file="1990.csv")
```

```
# 1991
```

```
> xpt = sasxport.get("CDBRFS91.XPT")
> write.csv(xpt,file="1991.csv")
# 1992
> xpt = sasxport.get("CDBRFS92.XPT")
> write.csv(xpt,file="1992.csv")
# 1993
> xpt = sasxport.get("CDBRFS93.XPT")
> write.csv(xpt,file="1993.csv")
# 1994
> xpt = sasxport.get("CDBRFS94.XPT")
> write.csv(xpt,file="1994.csv")
# 1995
> xpt = sasxport.get("CDBRFS95.XPT")
> write.csv(xpt,file="1995.csv")
# 1996
> xpt = sasxport.get("CDBRFS96.XPT")
> write.csv(xpt,file="1996.csv")
# 1997
> xpt = sasxport.get("CDBRFS97.XPT")
> write.csv(xpt,file="1997.csv")
> # 1998
> xpt = sasxport.get("CDBRFS98.XPT")
> write.csv(xpt,file="1998.csv")
# 1999
> xpt = sasxport.get("CDBRFS99.XPT")
> write.csv(xpt,file="1999.csv")
# 2000
> xpt = sasxport.get("CDBRFS00.XPT")
> write.csv(xpt,file="2000.csv")
# 2001
> xpt = sasxport.get("CDBRFS01.XPT")
> write.csv(xpt,file="2001.csv")
# 2002
> xpt = sasxport.get("CDBRFS02.XPT")
> write.csv(xpt,file="2002.csv")
# 2003
> xpt = sasxport.get("CDBRFS03.XPT")
> write.csv(xpt,file="2003.csv")
# 2004
> xpt = sasxport.get("CDBRFS04.XPT")
> write.csv(xpt,file="2004.csv")
# 2005
> xpt = sasxport.get("CDBRFS05.XPT")
```

```

> write.csv(xpt,file="2005.csv")
# 2006
> xpt = sasxport.get("CDBRFS06.XPT")
> write.csv(xpt,file="2006.csv")
# 2007
> xpt = sasxport.get("CDBRFS07.XPT")
> write.csv(xpt,file="2007.csv")
# 2008
> xpt = sasxport.get("CDBRFS08.XPT")
> write.csv(xpt,file="2008.csv")
# 2009
> xpt = sasxport.get("CDBRFS09.XPT")
> write.csv(xpt,file="2009.csv")
# 2010
> xpt = sasxport.get("CDBRFS10.XPT")
> write.csv(xpt,file="2010.csv")
# 2011
> xpt = sasxport.get("LLCP2011.XPT")
> write.csv(xpt,file="2011.csv")
# 2012
> xpt = sasxport.get("LLCP2012.XPT")
> write.csv(xpt,file="2012.csv")
# 2013
> xpt = sasxport.get("LLCP2013.XPT")
> write.csv(xpt,file="2013.csv")

```

## # Creating Graph 1

# Reading in and subsetting data variables

```

> dat = read.csv("2003.csv")
> dat_2003 = data.frame(WEIGHT=dat$finalwt, AGE = dat$age,SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap, LASTPAP = dat$lastpap,
>   STATE = dat$x.state);
> summary(dat_2003)
#
> dat = read.csv("2004.csv")
> dat_2004 = data.frame(WEIGHT=dat$finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2, LASTPAP = dat$lastpap2,
>   STATE = dat$x.state);
> summary(dat_2004)
#
> dat = read.csv("2005.csv")
> dat_2005 = data.frame(WEIGHT=dat$finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2, LASTPAP = dat$lastpap2,

```

```

> STATE = dat$x.state);
> summary(dat_2005)
#
> dat = read.csv("2006.csv")
> dat_2006 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2, LASTPAP = dat$lastpap2,
>   STATE = dat$x.state);
> summary(dat_2006)
#
> dat = read.csv("2007.csv")
> dat_2007 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2, LASTPAP = dat$lastpap2,
>   STATE = dat$x.state);
> summary(dat_2007)
#
> dat = read.csv("2008.csv")
> dat_2008 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2, LASTPAP = dat$lastpap2,
>   STATE = dat$x.state);
> summary(dat_2008)
#
> dat = read.csv("2009.csv")
> dat_2009 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2,
>   LASTPAP = dat$lastpap2, STATE = dat$x.state);
> summary(dat_2009)
#
> dat = read.csv("2010.csv")
> dat_2010 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2, LASTPAP = dat$lastpap2,
>   STATE = dat$x.state);
> summary(dat_2010)

> dat = read.csv("2012.csv")
> dat_2012 = data.frame(WEIGHT=dat$x.llcpwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2, LASTPAP = dat$lastpap2, STATE
>   = dat$x.state);
> summary(dat_2012)
#
> dat = read.csv("2013.csv")
> dat_2013 = data.frame(WEIGHT=dat$x.llcpwt, AGE = dat$x.age80, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, HADPAP = dat$hadpap2, LASTPAP = dat$lastpap2, STATE

```

```

>       = dat$x.state);
> summary(dat_2013)
#
> write.csv(dat_2003,"dat_2003.csv")
> write.csv(dat_2004,"dat_2004.csv")
> write.csv(dat_2005,"dat_2005.csv")
> write.csv(dat_2006,"dat_2006.csv")
> write.csv(dat_2007,"dat_2007.csv")
> write.csv(dat_2008,"dat_2008.csv")
> write.csv(dat_2009,"dat_2009.csv")
> write.csv(dat_2010,"dat_2010.csv")
> write.csv(dat_2012,"dat_2012.csv")
> write.csv(dat_2013,"dat_2013.csv")
# Finding proportions of screening for pap smears and mammograms
install.packages("survey",dep=TRUE)
library(survey)
#####
# 2003
#-----
# MAMMOGRAM General Population
> dat_2003 = read.csv("dat_2003.csv")
> dat_2003 = subset(dat_2003, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2003)
> des = svydesign(~1, weights=dat_2003$WEIGHT, data=dat_2003)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2003$HADMAM[which(dat_2003$AGE==age[i] & dat_2003$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2003[which(dat_2003$AGE==age[i] & dat_2003$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> GP_2003 = num/denom
> summary(GP_2003)
#-----
# PAPSMEAR Population
> dat_2003 = read.csv("dat_2003.csv")

```

```

> dat_2003 = subset(dat_2003, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2003)
> des = svydesign(~1, weights=dat_2003$WEIGHT, data=dat_2003)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2003$HADPAP[which(dat_2003$AGE==age[i] & dat_2003$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2003[which(dat_2003$AGE==age[i] & dat_2003$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2003 = num/denom
> summary(GP_PAP_2003)
#####
# 2004
#-----
# MAMMOGRAM General Population
> dat_2004 = read.csv("dat_2004.csv")
> dat_2004 = subset(dat_2004, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2004)
> des = svydesign(~1, weights=dat_2004$WEIGHT, data=dat_2004)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2004$HADPAP[which(dat_2004$AGE==age[i] & dat_2004$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2004[which(dat_2004$AGE==age[i] & dat_2004$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
> GP_2004 = num/denom
> summary(GP_2004)

```

```

#-----
# PAP Population
> dat_2004 = read.csv("dat_2004.csv")
> dat_2004 = subset(dat_2004, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2004)
> des = svydesign(~1, weights=dat_2004$WEIGHT, data=dat_2004)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2004$HADPAP[which(dat_2004$AGE==age[i] & dat_2004$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2004[which(dat_2004$AGE==age[i] & dat_2004$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2004 = num/denom
> summary(GP_PAP_2004)
#####
# 2005
#-----
# MAM Population
> dat_2005 = read.csv("dat_2005.csv")
> dat_2005 = subset(dat_2005, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2005)
> des = svydesign(~1, weights=dat_2005$WEIGHT, data=dat_2005)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2005$HADMAM[which(dat_2005$AGE==age[i] & dat_2005$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2005[which(dat_2005$AGE==age[i] & dat_2005$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])

```



```

> }
#
> GP_2005 = num/denom
> summary(GP_2005)
#-----
# PAP Population
> dat_2005 = read.csv("dat_2005.csv")
> dat_2005 = subset(dat_2005, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2005)
> des = svydesign(~1, weights=dat_2005$WEIGHT, data=dat_2005)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2005$HADPAP[which(dat_2005$AGE==age[i] & dat_2005$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2005[which(dat_2005$AGE==age[i] & dat_2005$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2005 = num/denom
> summary(GP_PAP_2005)
#####
# 2006
#-----
# MAM Population
> dat_2006 = read.csv("dat_2006.csv")
> dat_2006 = subset(dat_2006, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2006)
> des = svydesign(~1, weights=dat_2006$WEIGHT, data=dat_2006)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2006$HADPAP[which(dat_2006$AGE==age[i] & dat_2006$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)

```

```

> tempdat = dat_2006[which(dat_2006$AGE==age[i] & dat_2006$SEX==2),]
> tab2 = svytable(~tempdat$HOWLONG, design=des2)
> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> GP_2006 = num/denom
> summary(GP_2006)
#-----
# PAP Population
> dat_2006 = read.csv("dat_2006.csv")
> dat_2006 = subset(dat_2006, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2006)
> des = svydesign(~1, weights=dat_2006$WEIGHT, data=dat_2006)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2006$HADPAP[which(dat_2006$AGE==age[i] & dat_2006$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2006[which(dat_2006$AGE==age[i] & dat_2006$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2006 = num/denom
> summary(GP_PAP_2006)
#####
# 2007
#-----
# MAM Population
> dat_2007 = read.csv("dat_2007.csv")
> dat_2007 = subset(dat_2007, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2007)
> des = svydesign(~1, weights=dat_2007$WEIGHT, data=dat_2007)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){

```

```

> des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
> tab = svytable(~dat_2007$HADMAM[which(dat_2007$AGE==age[i] & dat_2007$SEX==2)],
> design=des2)
> denom[i] = sum(tab)
> tempdat = dat_2007[which(dat_2007$AGE==age[i] & dat_2007$SEX==2),]
> tab2 = svytable(~tempdat$HOWLONG, design=des2)
> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> GP_2007 = num/denom
> summary(GP_2007)
#-----
# PAP Population
> dat_2007 = read.csv("dat_2007.csv")
> dat_2007 = subset(dat_2007, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2007)
> des = svydesign(~1, weights=dat_2007$WEIGHT, data=dat_2007)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
> des2 = subset(des, AGE==age[i] & SEX==2)
> tab = svytable(~dat_2007$HADPAP[which(dat_2007$AGE==age[i] & dat_2007$SEX==2)],
> design=des2)
> denom[i] = sum(tab)
> tempdat = dat_2007[which(dat_2007$AGE==age[i] & dat_2007$SEX==2),]
> tab2 = svytable(~tempdat$LASTPAP, design=des2)
> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
> tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2007 = num/denom
> summary(GP_PAP_2007)
#####
# 2008
#-----
# MAM Population
> dat_2008 = read.csv("dat_2008.csv")
> dat_2008 = subset(dat_2008, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2008)
> des = svydesign(~1, weights=dat_2008$WEIGHT, data=dat_2008)

```

```

> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2008$HADMAM[which(dat_2008$AGE==age[i] & dat_2008$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2008[which(dat_2008$AGE==age[i] & dat_2008$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> GP_2008 = num/denom
> summary(GP_2008)
#-----
# PAP Population
> dat_2008 = read.csv("dat_2008.csv")
> dat_2008 = subset(dat_2008, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2008)
> des = svydesign(~1, weights=dat_2008$WEIGHT, data=dat_2008)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2008$HADPAP[which(dat_2008$AGE==age[i] & dat_2008$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2008[which(dat_2008$AGE==age[i] & dat_2008$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2008 = num/denom
> summary(GP_PAP_2008)
#####
# 2009
#-----
# MAM Population

```

```

> dat_2009 = read.csv("dat_2009.csv")
> dat_2009 = subset(dat_2009, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2009)
> des = svydesign(~1, weights=dat_2009$WEIGHT, data=dat_2009)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2009$HADMAM[which(dat_2009$AGE==age[i] & dat_2009$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2009[which(dat_2009$AGE==age[i] & dat_2009$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> GP_2009 = num/denom
> summary(GP_2009)
#-----
# PAP Population
> dat_2009 = read.csv("dat_2009.csv")
> dat_2009 = subset(dat_2009, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2009)
> des = svydesign(~1, weights=dat_2009$WEIGHT, data=dat_2009)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2009$HADPAP[which(dat_2009$AGE==age[i] & dat_2009$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2009[which(dat_2009$AGE==age[i] & dat_2009$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2009 = num/denom
> summary(GP_PAP_2009)

```

```
#####
# 2010
#-----
# MAM Population
> dat_2010 = read.csv("dat_2010.csv")
> dat_2010 = subset(dat_2010, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2010)
> des = svydesign(~1, weights=dat_2010$WEIGHT, data=dat_2010)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2010$HADMAM[which(dat_2010$AGE==age[i] & dat_2010$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2010[which(dat_2010$AGE==age[i] & dat_2010$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> GP_2010 = num/denom
> summary(GP_2010)
#-----
# PAP Population
> dat_2010 = read.csv("dat_2010.csv")
> dat_2010 = subset(dat_2010, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2010)
> des = svydesign(~1, weights=dat_2010$WEIGHT, data=dat_2010)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2010$HADPAP[which(dat_2010$AGE==age[i] & dat_2010$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2010[which(dat_2010$AGE==age[i] & dat_2010$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
```

```

>           tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)))
> }
#
> GP_PAP_2010 = num/denom
> summary(GP_PAP_2010)
#####
# 2012
#-----
# MAM Population
> dat_2012 = read.csv("dat_2012.csv")
> dat_2012 = subset(dat_2012, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2012)
> des = svydesign(~1, weights=dat_2012$WEIGHT, data=dat_2012)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2012$HADMAM[which(dat_2012$AGE==age[i] & dat_2012$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2012[which(dat_2012$AGE==age[i] & dat_2012$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> GP_2012 = num/denom
> summary(GP_2012)
#-----
# PAP Population
> dat_2012 = read.csv("dat_2012.csv")
> dat_2012 = subset(dat_2012, SEX == 2 & HADMAM & HADPAP)
> summary(dat_2012)
> des = svydesign(~1, weights=dat_2012$WEIGHT, data=dat_2012)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2012$HADPAP[which(dat_2012$AGE==age[i] & dat_2012$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)

```

```

> tempdat = dat_2012[which(dat_2012$AGE==age[i] & dat_2012$SEX==2),]
> tab2 = svytable(~tempdat$LASTPAP, design=des2)
> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2012 = num/denom
> summary(GP_PAP_2012)
#
> GENERAL = data.frame(MAM03 = GP_2003, PAP03 = GP_PAP_2003,
>     MAM04 = GP_2004, PAP04 = GP_PAP_2004,
>     MAM05 = GP_2005, PAP05 = GP_PAP_2005,
>     MAM06 = GP_2006, PAP06 = GP_PAP_2006,
>     MAM07 = GP_2007, PAP07 = GP_PAP_2007,
>     MAM08 = GP_2008, PAP08 = GP_PAP_2008,
>     MAM09 = GP_2009, PAP09 = GP_PAP_2009,
>     MAM10 = GP_2010, PAP10 = GP_PAP_2010,
>     MAM12 = GP_2012, PAP12 = GP_PAP_2012)
> summary(GENERAL)
> write.csv(GENERAL, "GENERAL.csv")

```

# Creating FDA objects for Mammogram and Pap Smear Proportions

```

> install.packages("fda",dep=TRUE)
> library("fda")
#####
# General Population
> mam_props = read.csv("GENERAL.csv")
>
> lambda = 1e-3
> norder = 6
> samples = seq(from = 30,to = 70, by = 1)
> nbasis = length(samples) + norder-2
> mybasis = create.bspline.basis(c(30,70), nbasis, norder, samples)
> myfdPar = fdPar(mybasis, 4, lambda)
> myfd = smooth.basis(samples, as.matrix(mam_props[,c(-1,-1)]), myfdPar)$fd
>
> par(mfrow=c(3,3))
> plot(myfd[1], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2003",col='red')
> plot(myfd[2], ylim=c(0,1), col='blue',add=T)
> plot(myfd[3], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2004",col='red')
> plot(myfd[4], ylim=c(0,1), col='blue',add=T)

```



```

> plot(myfd[5], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2005",col='red')
> plot(myfd[6], ylim=c(0,1), col='blue',add=T)
> plot(myfd[7], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2006",col='red')
> plot(myfd[8], ylim=c(0,1), col='blue',add=T)
> plot(myfd[9], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2007",col='red')
> plot(myfd[10], ylim=c(0,1), col='blue',add=T)
> plot(myfd[11], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2008",col='red')
> plot(myfd[12], ylim=c(0,1), col='blue',add=T)
> plot(myfd[13], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2009",col='red')
> plot(myfd[14], ylim=c(0,1), col='blue',add=T)
> plot(myfd[15], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2010",col='red')
> plot(myfd[16], ylim=c(0,1), col='blue',add=T)
> plot(myfd[17], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2012",col='red')
> plot(myfd[18], ylim=c(0,1), col='blue',add=T)
#####
# Mass Compare
mam_props = read.csv("MASSCOMPARE.csv")

lambda = 1e-3
norder = 6
samples = seq(from = 30,to = 70, by = 1)
nbasis = length(samples) + norder-2
mybasis = create.bspline.basis(c(30,70), nbasis, norder, samples)
myfdPar = fdPar(mybasis, 4, lambda)
myfd = smooth.basis(samples, as.matrix(mam_props[,c(-1,-1)]), myfdPar)$fd

par(mfrow=c(1,3))
plot(myfd[1], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2006",col='red')
plot(myfd[2], ylim=c(0,1), col='blue',add=T)
plot(myfd[5], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2008",col='red')
plot(myfd[6], ylim=c(0,1), col='blue',add=T)
plot(myfd[9], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2010",col='red')
plot(myfd[10], ylim=c(0,1), col='blue',add=T)
par(mfrow=c(1,3))
plot(myfd[3], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2006",col='red')
plot(myfd[4], ylim=c(0,1), col='blue',add=T)
plot(myfd[7], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2008",col='red')
plot(myfd[8], ylim=c(0,1), col='blue',add=T)
plot(myfd[11], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2010",col='red')
plot(myfd[12], ylim=c(0,1), col='blue',add=T)
#title(main="Massachusetts vs General Screening",outer=T)

```

## # GRAPH 2

```
install.packages("survey",dep=TRUE)
library(survey)

#####
# 2006
#-----
# MASS Population
> dat_2006 = read.csv("dat_2006.csv")
> dat_2006 = subset(dat_2006, SEX == 2 & HADMAM & HADPAP & STATE == 25)
> summary(dat_2006)
> des = svydesign(~1, weights=dat_2006$WEIGHT, data=dat_2006)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2006$HADMMAM[which(dat_2006$AGE==age[i] & dat_2006$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2006[which(dat_2006$AGE==age[i] & dat_2006$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
> MASS_MAM_2006 = num/denom
> summary(MASS_MAM_2006)
# MAM Population
> dat_2006 = read.csv("dat_2006.csv")
> dat_2006 = subset(dat_2006, SEX == 2 & HADMAM & HADPAP & STATE != 25)
> summary(dat_2006)
> des = svydesign(~1, weights=dat_2006$WEIGHT, data=dat_2006)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2006$HADMMAM[which(dat_2006$AGE==age[i] & dat_2006$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2006[which(dat_2006$AGE==age[i] & dat_2006$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
```

```

> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> MAM_2006 = num/denom
> summary(MAM_2006)
#-----
# MASS Population
> dat_2006 = read.csv("dat_2006.csv")
> dat_2006 = subset(dat_2006, SEX == 2 & HADMAM & HADPAP & STATE == 25)
> summary(dat_2006)
> des = svydesign(~1, weights=dat_2006$WEIGHT, data=dat_2006)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2006$HADPAP[which(dat_2006$AGE==age[i] & dat_2006$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2006[which(dat_2006$AGE==age[i] & dat_2006$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
> MASS_PAP_2006 = num/denom
> summary(MASS_PAP_2006)
# PAP Population
> dat_2006 = read.csv("dat_2006.csv")
> dat_2006 = subset(dat_2006, SEX == 2 & HADMAM & HADPAP & STATE != 25)
> summary(dat_2006)
> des = svydesign(~1, weights=dat_2006$WEIGHT, data=dat_2006)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2006$HADPAP[which(dat_2006$AGE==age[i] & dat_2006$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2006[which(dat_2006$AGE==age[i] & dat_2006$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)

```

```

> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
> tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2006 = num/denom
> summary(GP_PAP_2006)
#####
#####
# 2008
#-----
# MASS Population
> dat_2008 = read.csv("dat_2008.csv")
> dat_2008 = subset(dat_2008, SEX == 2 & HADMAM & HADPAP & STATE == 25)
> summary(dat_2008)
> des = svydesign(~1, weights=dat_2008$WEIGHT, data=dat_2008)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2008$HADMAM[which(dat_2008$AGE==age[i] & dat_2008$SEX==2)],
>   design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2008[which(dat_2008$AGE==age[i] & dat_2008$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> MASS_MAM_2008 = num/denom
> summary(MASS_MAM_2008)
# MAM Population
> dat_2008 = read.csv("dat_2008.csv")
> dat_2008 = subset(dat_2008, SEX == 2 & HADMAM & HADPAP & STATE != 25)
> summary(dat_2008)
> des = svydesign(~1, weights=dat_2008$WEIGHT, data=dat_2008)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2008$HADMAM[which(dat_2008$AGE==age[i] & dat_2008$SEX==2)],

```

```

>           design=des2)
> denom[i] = sum(tab)
> tempdat = dat_2008[which(dat_2008$AGE==age[i] & dat_2008$SEX==2),]
> tab2 = svytable(~tempdat$HOWLONG, design=des2)
> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
> MAM_2008 = num/denom
> summary(MAM_2008)
#-----
# MASS Population
> dat_2008 = read.csv("dat_2008.csv")
> dat_2008 = subset(dat_2008, SEX == 2 & HADMAM & HADPAP & STATE == 25)
> summary(dat_2008)
> des = svydesign(~1, weights=dat_2008$WEIGHT, data=dat_2008)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2008$HADPAP[which(dat_2008$AGE==age[i] & dat_2008$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2008[which(dat_2008$AGE==age[i] & dat_2008$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
> MASS_PAP_2008 = num/denom
> summary(MASS_PAP_2008)
# PAP Population
> dat_2008 = read.csv("dat_2008.csv")
> dat_2008 = subset(dat_2008, SEX == 2 & HADMAM & HADPAP & STATE != 25)
> summary(dat_2008)
> des = svydesign(~1, weights=dat_2008$WEIGHT, data=dat_2008)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2008$HADPAP[which(dat_2008$AGE==age[i] & dat_2008$SEX==2)],
>     design=des2)

```

```

>   denom[i] = sum(tab)
>   tempdat = dat_2008[which(dat_2008$AGE==age[i] & dat_2008$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2008 = num/denom
> summary(GP_PAP_2008)
#####
# 2010
#-----
# MASS Population
> dat_2010 = read.csv("dat_2010.csv")
> dat_2010 = subset(dat_2010, SEX == 2 & HADMAM & HADPAP & STATE == 25)
> summary(dat_2010)
> des = svydesign(~1, weights=dat_2010$WEIGHT, data=dat_2010)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
>   tab = svytable(~dat_2010$HADMAM[which(dat_2010$AGE==age[i] & dat_2010$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2010[which(dat_2010$AGE==age[i] & dat_2010$SEX==2),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> MASS_MAM_2010 = num/denom
> summary(MASS_MAM_2010)
# MAM Population
> dat_2010 = read.csv("dat_2010.csv")
> dat_2010 = subset(dat_2010, SEX == 2 & HADMAM & HADPAP & STATE != 25)
> summary(dat_2010)
> des = svydesign(~1, weights=dat_2010$WEIGHT, data=dat_2010)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){

```

```

> des2 = subset(des, AGE==age[i] & SEX==2 & HADMAM & HADPAP)
> tab = svytable(~dat_2010$HADMAM[which(dat_2010$AGE==age[i] & dat_2010$SEX==2)],
> design=des2)
> denom[i] = sum(tab)
> tempdat = dat_2010[which(dat_2010$AGE==age[i] & dat_2010$SEX==2),]
> tab2 = svytable(~tempdat$HOWLONG, design=des2)
> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1)])
> }
#
> MAM_2010 = num/denom
> summary(MAM_2010)
#-----
# MASS Population
> dat_2010 = read.csv("dat_2010.csv")
> dat_2010 = subset(dat_2010, SEX == 2 & HADMAM & HADPAP & STATE == 25)
> summary(dat_2010)
> des = svydesign(~1, weights=dat_2010$WEIGHT, data=dat_2010)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
> des2 = subset(des, AGE==age[i] & SEX==2)
> tab = svytable(~dat_2010$HADPAP[which(dat_2010$AGE==age[i] & dat_2010$SEX==2)],
> design=des2)
> denom[i] = sum(tab)
> tempdat = dat_2010[which(dat_2010$AGE==age[i] & dat_2010$SEX==2),]
> tab2 = svytable(~tempdat$LASTPAP, design=des2)
> tab2new = as.data.frame(tab2)
> num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
> tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> MASS_PAP_2010 = num/denom
> summary(MASS_PAP_2010)
# PAP Population
> dat_2010 = read.csv("dat_2010.csv")
> dat_2010 = subset(dat_2010, SEX == 2 & HADMAM & HADPAP & STATE != 25)
> summary(dat_2010)
> des = svydesign(~1, weights=dat_2010$WEIGHT, data=dat_2010)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)

```

```

> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2)
>   tab = svytable(~dat_2010$HADPAP[which(dat_2010$AGE==age[i] & dat_2010$SEX==2)],
>     design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2010[which(dat_2010$AGE==age[i] & dat_2010$SEX==2),]
>   tab2 = svytable(~tempdat$LASTPAP, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.LASTPAP == 1 |
>     tab2new$tempdat.LASTPAP == 2 | tab2new$tempdat.LASTPAP == 3)])
> }
#
> GP_PAP_2010 = num/denom
> summary(GP_PAP_2010)
#
> MASSCOMPARE = data.frame(MassMam06 = MASS_MAM_2006, GenMam06 = MAM_2006,
>   MassPap06= MASS_PAP_2006, GenPap06 = GP_PAP_2006,
>   MassMam08 = MASS_MAM_2008, GenMam08 = MAM_2008,
>   MassPap08 = MASS_PAP_2008, GenPap08 = GP_PAP_2008,
>   MassMam10 = MASS_MAM_2010, GenMam10 = MAM_2010,
>   MassPap10= MASS_PAP_2010, GenPap10 = GP_PAP_2010)
> summary(MASSCOMPARE)
> write.csv(MASSCOMPARE,"MASSCOMPARE.CSV")
#
# Creating Functional Curves of Massachusetts Comparison
> install.packages("fda",dep=TRUE)
> library("fda")
#####
# General Population
> mam_props = read.csv("GENERAL.csv")
> lambda = 1e-3
> norder = 6
> samples = seq(from = 30,to = 70, by = 1)
> nbasis = length(samples) + norder-2
> mybasis = create.bspline.basis(c(30,70), nbasis, norder, samples)
> myfdPar = fdPar(mybasis, 4, lambda)
> myfd = smooth.basis(samples, as.matrix(mam_props[,c(-1,-1)]), myfdPar)$fd
> par(mfrow=c(3,3))
> plot(myfd[1], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2003",col='red')
> plot(myfd[2], ylim=c(0,1), col='blue',add=T)
> plot(myfd[3], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2004",col='red')
> plot(myfd[4], ylim=c(0,1), col='blue',add=T)
> plot(myfd[5], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2005",col='red')

```



```

> plot(myfd[6], ylim=c(0,1), col='blue',add=T)
> plot(myfd[7], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2006",col='red')
> plot(myfd[8], ylim=c(0,1), col='blue',add=T)
> plot(myfd[9], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2007",col='red')
> plot(myfd[10], ylim=c(0,1), col='blue',add=T)
> plot(myfd[11], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2008",col='red')
> plot(myfd[12], ylim=c(0,1), col='blue',add=T)
> plot(myfd[13], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2009",col='red')
> plot(myfd[14], ylim=c(0,1), col='blue',add=T)
> plot(myfd[15], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2010",col='red')
> plot(myfd[16], ylim=c(0,1), col='blue',add=T)
> plot(myfd[17], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2012",col='red')
> plot(myfd[18], ylim=c(0,1), col='blue',add=T)
#####
# Mass Compare
> mam_props = read.csv("MASSCOMPARE.csv")
> lambda = 1e-3
> norder = 6
> samples = seq(from = 30,to = 70, by = 1)
> nbasis = length(samples) + norder-2
> mybasis = create.bspline.basis(c(30,70), nbasis, norder, samples)
> myfdPar = fdPar(mybasis, 4, lambda)
> myfd = smooth.basis(samples, as.matrix(mam_props[,c(-1,-1)]), myfdPar)$fd
> par(mfrow=c(1,3))
> plot(myfd[1], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2006",col='red')
> plot(myfd[2], ylim=c(0,1), col='blue',add=T)
> plot(myfd[5], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2008",col='red')
> plot(myfd[6], ylim=c(0,1), col='blue',add=T)
> plot(myfd[9], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2010",col='red')
> plot(myfd[10], ylim=c(0,1), col='blue',add=T)
> par(mfrow=c(1,3))
> plot(myfd[3], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2006",col='red')
> plot(myfd[4], ylim=c(0,1), col='blue',add=T)
> plot(myfd[7], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2008",col='red')
> plot(myfd[8], ylim=c(0,1), col='blue',add=T)
> plot(myfd[11], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2010",col='red')
> plot(myfd[12], ylim=c(0,1), col='blue',add=T)

# GRAPH 3
> dat = read.csv("1987.csv")
> dat_1987 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age,SEX = dat$sex,HADMAM =
>   dat$hadmam,HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1987)
> dat = read.csv("1988.csv")

```

```

> dat_1988 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1988)
> dat = read.csv("1989.csv")
> dat_1989 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1989)
> dat = read.csv("1990.csv")
> dat_1990 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1990)
> dat = read.csv("1991.csv")
> dat_1991 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age,SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1991)
> dat = read.csv("1992.csv")
> dat_1992 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1992)
> dat = read.csv("1993.csv")
> dat_1993 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age,SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1993)
> dat = read.csv("1994.csv")
> dat_1994 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age,SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1994)
> dat = read.csv("1995.csv")
> dat_1995 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1995)
> dat = read.csv("1996.csv")
> dat_1996 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age,SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1996)
> dat = read.csv("1997.csv")
> dat_1997 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1997)
> dat = read.csv("1998.csv")
> dat_1998 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_1998)
> dat = read.csv("2006.csv")
> dat_2006 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_2006)
> dat = read.csv("2008.csv")
> dat_2008 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_2008)
> dat = read.csv("2010.csv")
> dat_2010 = data.frame(WEIGHT=dat$x.finalwt, AGE = dat$age, SEX = dat$sex, HADMAM =
>   dat$hadmam, HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_2010)
> dat = read.csv("2012.csv")
> dat_2012 = data.frame(WEIGHT=dat$x.llcpwt, AGE = dat$age, SEX = dat$sex, HADMAM =

```

```

> dat$hadmam,HOWLONG = dat$howlong, SEATBELT = dat$seatbelt); summary(dat_2012)
install.packages("survey",dep=TRUE)
library(survey)
#####
# 2006
#-----
# Seatbelt
> des = svydesign(~1, weights=dat_2006$WEIGHT, data=dat_2006)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & SEATBELT == 1)
>   tab = svytable(~dat_2006$HADMAM[which(dat_2006$AGE==age[i] & dat_2006$SEX==2 &
>     dat_2006$SEATBELT == 1)], design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2006[which(dat_2006$AGE==age[i] & dat_2006$SEX==2 &
>     dat_2006$SEATBELT == 1),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1 |
>     tab2new$tempdat.HOWLONG==2)])
> }
> sb_prop06=num/denom
#-----
# No Seatbelt
> des = svydesign(~1, weights=dat_2006$WEIGHT, data=dat_2006)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & SEATBELT != 1)
>   tab = svytable(~dat_2006$HADMAM[which(dat_2006$AGE==age[i] & dat_2006$SEX==2 &
>     dat_2006$SEATBELT != 1)], design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2006[which(dat_2006$AGE==age[i] & dat_2006$SEX==2 &
>     dat_2006$SEATBELT != 1),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1 |
>     tab2new$tempdat.HOWLONG==2)])
> }
#

```

```

> nsb_prop06=num/denom
#####
# 2008
#-----
# Seatbelt
> des = svydesign(~1, weights=dat_2008$WEIGHT, data=dat_2008)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & SEATBELT == 1)
>   tab = svytable(~dat_2008$HADMAM[which(dat_2008$AGE==age[i] & dat_2008$SEX==2 &
>   dat_2008$SEATBELT == 1)], design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2008[which(dat_2008$AGE==age[i] & dat_2008$SEX==2 &
>   dat_2008$SEATBELT == 1),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1 |
>   tab2new$tempdat.HOWLONG==2)])
> }
#
> sb_prop08=num/denom
#-----
# No Seatbelt
> des = svydesign(~1, weights=dat_2008$WEIGHT, data=dat_2008)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & SEATBELT != 1)
>   tab = svytable(~dat_2008$HADMAM[which(dat_2008$AGE==age[i] & dat_2008$SEX==2 &
>   dat_2008$SEATBELT != 1)], design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2008[which(dat_2008$AGE==age[i] & dat_2008$SEX==2 &
>   dat_2008$SEATBELT != 1),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1 |
>   tab2new$tempdat.HOWLONG==2)])
> }
#
> nsb_prop08=num/denom

```

```

#####
# 2010
#-----
# Seatbelt
> des = svydesign(~1, weights=dat_2010$WEIGHT, data=dat_2010)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & SEATBELT == 1)
>   tab = svytable(~dat_2010$HADMAM[which(dat_2010$AGE==age[i] & dat_2010$SEX==2 &
>     dat_2010$SEATBELT == 1)], design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2010[which(dat_2010$AGE==age[i] & dat_2010$SEX==2 &
>     dat_2010$SEATBELT == 1),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1 |
>     tab2new$tempdat.HOWLONG==2)])
> }
> sb_prop10=num/denom
#-----
# No Seatbelt
> des = svydesign(~1, weights=dat_2010$WEIGHT, data=dat_2010)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & SEATBELT != 1)
>   tab = svytable(~dat_2010$HADMAM[which(dat_2010$AGE==age[i] & dat_2010$SEX==2 &
>     dat_2010$SEATBELT != 1)], design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2010[which(dat_2010$AGE==age[i] & dat_2010$SEX==2 &
>     dat_2010$SEATBELT != 1),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1 |
>     tab2new$tempdat.HOWLONG==2)])
> }
> nsb_prop10=num/denom
#####
# 2012
#-----

```

```

# Seatbelt
> des = svydesign(~1, weights=dat_2012$WEIGHT, data=dat_2012)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & SEATBELT == 1)
>   tab = svytable(~dat_2012$HADMMAM[which(dat_2012$AGE==age[i] & dat_2012$SEX==2 &
>   dat_2012$SEATBELT == 1)],design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2012[which(dat_2012$AGE==age[i] & dat_2012$SEX==2 &
>   dat_2012$SEATBELT == 1),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1 |
>   tab2new$tempdat.HOWLONG==2)])
> }
> sb_prop12=num/denom
#-----
# No Seatbelt
> des = svydesign(~1, weights=dat_2012$WEIGHT, data=dat_2012)
> denom=length(41)
> num=length(41)
> age=seq(30,70,by=1)
> for(i in 1:length(age)){
>   des2 = subset(des, AGE==age[i] & SEX==2 & SEATBELT != 1)
>   tab = svytable(~dat_2012$HADMMAM[which(dat_2012$AGE==age[i] & dat_2012$SEX==2 &
>   dat_2012$SEATBELT != 1)], design=des2)
>   denom[i] = sum(tab)
>   tempdat = dat_2012[which(dat_2012$AGE==age[i] & dat_2012$SEX==2 &
>   dat_2012$SEATBELT != 1),]
>   tab2 = svytable(~tempdat$HOWLONG, design=des2)
>   tab2new = as.data.frame(tab2)
>   num[i] = sum(tab2new$Freq[which(tab2new$tempdat.HOWLONG==1 |
>   tab2new$tempdat.HOWLONG==2)])
> }
> nsb_prop12=num/denom
#####
# Make Data Frame
>   age = seq(from = 30, to = 70, by = 1)
>   mam_props = data.frame(age,sb_prop87,nsb_prop87,sb_prop88,nsb_prop88,sb_prop89,
>   nsb_prop89, sb_prop90, nsb_prop90, sb_prop91, nsb_prop91, sb_prop92,
>   nsb_prop92, sb_prop93, nsb_prop93, sb_prop94, nsb_prop94, sb_prop95,

```

```

> nsb_prop95, sb_prop96, nsb_prop96, sb_prop97, nsb_prop97, sb_prop98,
> nsb_prop98, sb_prop06, nsb_prop06, sb_prop08, nsb_prop08, sb_prop10,
> nsb_prop10, sb_prop12, nsb_prop12)
# Write File
> write.csv(mam_props, "Mammogram Proportions by Seatbelt Use.csv")
#####
# Plot Data
> install.packages("fda", dep=TRUE)
> library("fda")
#
> mam_props = read.csv("Mammogram Proportions by Seatbelt Use.csv")
#
> lambda = 1e-12
> norder = 6
> samples = seq(from = 30, to = 70, by = 1)
> nbasis = length(samples) + norder - 2
> mybasis = create.bspline.basis(c(30, 70), nbasis, norder, samples)
> myfdPar = fdPar(mybasis, 4, lambda)
> myfd = smooth.basis(samples, as.matrix(mam_props[, c(-1, -2)]), myfdPar)$fd
> par(mfrow=c(4, 4))
> title(main="Mammogram In Last Two Years", outer=TRUE)
> plot(myfd[1], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1987", col='red')
> plot(myfd[2], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[3], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1988", col='red')
> plot(myfd[4], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[5], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1989", col='red')
> plot(myfd[6], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[7], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1990", col='red')
> plot(myfd[8], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[9], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1991", col='red')
> plot(myfd[10], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[11], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1992", col='red')
> plot(myfd[12], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[13], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1993", col='red')
> plot(myfd[14], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[15], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1994", col='red')
> plot(myfd[16], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[17], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1995", col='red')
> plot(myfd[18], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[19], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1996", col='red')
> plot(myfd[20], ylim=c(0, 1), col='blue', add=T)
> plot(myfd[21], ylim=c(0, 1), xlab="AGE", ylab="PROPORTION", main="1997", col='red')
> plot(myfd[22], ylim=c(0, 1), col='blue', add=T)

```

```

> plot(myfd[23], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="1998",col='red')
> plot(myfd[24], ylim=c(0,1), col='blue',add=T)
> plot(myfd[25], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2006",col='red')
> plot(myfd[26], ylim=c(0,1), col='blue',add=T)
> plot(myfd[27], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2008",col='red')
> plot(myfd[28], ylim=c(0,1), col='blue',add=T)
> plot(myfd[29], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2010",col='red')
> plot(myfd[30], ylim=c(0,1), col='blue',add=T)
> plot(myfd[31], ylim=c(0,1), xlab="AGE", ylab="PROPORTION", main="2012",col='red')
> plot(myfd[32], ylim=c(0,1), col='blue',add=T)

```

### # SUBSETTING DATA

```

> dat = read.csv("2003.csv")
> dataA = data.frame(weight = dat$x.finalwt, age = dat$age,sex = dat$sex,hadmam = >
>   dat$hadmam,howlong = dat$howlong, seatbelt = NA, marital = dat$marital, adultHH =
>   dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth,healthcov =
>   dat$hlthplan, checkup = NA, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ, income = dat$income2, rent_own = NA, smoke = dat$x.smoker2, hadpap =
>   dat$hadpap, lastpap = dat$lastpap, doctor = NA, paid = NA); summary(dataA)
#
> dataA = subset(dataA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(dataA)
#
> recentmam = as.numeric((dataA$hadmam == 1) & (dataA$howlong == 1)); summary(recentmam)
#
> A1 = as.numeric(dataA$age < 35); summary(A1)
> A2 = as.numeric(dataA$age > 34 & dataA$age <40); summary(A2)
> A3 = as.numeric(dataA$age > 39 & dataA$age < 45); summary(A3)
> A4 = as.numeric(dataA$age > 44 & dataA$age < 50); summary(A4)
> A5 = as.numeric(dataA$age > 49 & dataA$age < 55); summary(A5)
> A6 = as.numeric(dataA$age > 54 & dataA$age < 60); summary(A6)
> A7 = as.numeric(dataA$age > 59 & dataA$age < 65); summary(A7)
> A8 = as.numeric(dataA$age > 64); summary(A8)
#
> marital = as.numeric((dataA$marital != 1) & (dataA$marital != 6)); summary(marital)
#
> adults = as.numeric(dataA$adultHH > 1); summary(adults)
#
> women = as.numeric(dataA$womenHH > 1); summary(women)
#
> G1 = as.numeric(dataA$genhealth == 1); summary(G1)
> G2 = as.numeric(dataA$genhealth == 2); summary(G2)
> G3 = as.numeric(dataA$genhealth == 3); summary(G3)
> G4 = as.numeric(dataA$genhealth == 4); summary(G4)

```



```

> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
> if((datA$educa[i] < 4)) education[i] = 0
> if(datA$educa[i] == 4) education[i] = 1
> if(datA$educa[i] > 4) education[i] = 2
> if(is.na(dat$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
> if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
> if(datA$employ[i] == 1) employ[i] = 1
> if(datA$employ[i] == 2) employ[i] = 2
> if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){
> if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
> if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
> if((datA$income[i] == 7) | (datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)
> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric((datA$smoke != 1) & (datA$smoke != 2)); summary(smoke)
#

```

```

> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
#
> womenhealth = (recentmam + recentpap); summary(womenhealth)
#
> healthcov = as.numeric(datA$healthcov == 1); summary(healthcov)
#
> dat2003 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8, sex = datA$sex, recentmam = recentmam, marital = marital, adultHH =
>   adults, womenHH = women, G1 = G1, childHH = datA$childHH, education = E2 + E3,
>   healthplan = healthcov, semploy = E3, employ = E2, income = I2 + I3, recentpap = recentpap,
>   smoke = smoke, womenhealth = womenhealth); summary(dat2003)
> write.csv(dat2003, "dat2003.csv")
##### 2004
> dat = read.csv("2004.csv")
> dataA = data.frame(weight = dat$x.finalwt, age = dat$age, sex = dat$sex, hadmam =
>   dat$hadmam, howlong = dat$howlong, seatbelt = NA, marital = dat$marital, adultHH =
>   dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth, healthcov =
>   dat$hlthplan, checkup = NA, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ, income = dat$income2, rent_own = NA, smoke = dat$x.smoker2, hadpap =
>   dat$hadpap2, lastpap = dat$lastpap2, doctor = NA, paid = NA); summary(dataA)
#
> dataA = subset(dataA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(dataA)
#
> A1 = as.numeric(datA$age < 35); summary(A1)
> A2 = as.numeric(datA$age > 34 & datA$age < 40); summary(A2)
> A3 = as.numeric(datA$age > 39 & datA$age < 45); summary(A3)
> A4 = as.numeric(datA$age > 44 & datA$age < 50); summary(A4)
> A5 = as.numeric(datA$age > 49 & datA$age < 55); summary(A5)
> A6 = as.numeric(datA$age > 54 & datA$age < 60); summary(A6)
> A7 = as.numeric(datA$age > 59 & datA$age < 65); summary(A7)
> A8 = as.numeric(datA$age > 64); summary(A8)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$howlong == 1)); summary(recentmam)
#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)
#
> women = as.numeric(datA$womenHH > 1); summary(women)
#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
> G3 = as.numeric(datA$genhealth == 3); summary(G3)

```

```

> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(dat$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
>   if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
>   if(datA$employ[i] == 1) employ[i] = 1
>   if(datA$employ[i] == 2) employ[i] = 2
>   if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){
>   if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
>   if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
>   if((datA$income[i] == 7) |(datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)
> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric((datA$smoke != 1) & (datA$smoke != 2)); summary(smoke)

```

```

#
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
#
> womenhealth = (recentmam + recentpap); summary(womenhealth)
#
> healthcov = as.numeric(datA$healthcov == 1);summary(healthcov)
#
> dat2004 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8, sex = datA$sex, recentmam = recentmam, marital = marital, adultHH =
>   adults, womenHH = women, smoke = smoke, G1 = G1, childHH = datA$childHH, education =
>   E2 + E3, semploy = Em3, employ = Em2, income = I2 + I3, healthplan = healthcov, recentpap =
>   recentpap, womenhealth = womenhealth); summary(dat2004)
#
> write.csv(dat2004, "dat2004.csv")
##### 2005
> dat = read.csv("2005.csv")
> dataA = data.frame(weight = dat$x.finalwt, age = dat$age,sex = dat$sex,hadmam =
>   dat$hadmam,howlong = dat$howlong, seatbelt = NA, marital = dat$marital, adultHH =
>   dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth, healthcov =
>   dat$hlthplan, checkup = dat$checkup, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ, income = dat$income2, rent_own = NA, smoke = dat$x.rfsmok3, hadpap =
>   dat$hadpap2, lastpap = dat$lastpap2, doctor = NA, paid = NA); summary(dataA)
#
> dataA = subset(dataA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(dataA)
#
> A1 = as.numeric(datA$age < 35); summary(A1)
> A2 = as.numeric(datA$age > 34 & datA$age <40); summary(A2)
> A3 = as.numeric(datA$age > 39 & datA$age < 45); summary(A3)
> A4 = as.numeric(datA$age > 44 & datA$age < 50); summary(A4)
> A5 = as.numeric(datA$age > 49 & datA$age < 55); summary(A5)
> A6 = as.numeric(datA$age > 54 & datA$age < 60); summary(A6)
> A7 = as.numeric(datA$age > 59 & datA$age < 65); summary(A7)
> A8 = as.numeric(datA$age > 64); summary(A8)
#
> checkup = as.numeric(datA$checkup == 1); summary(checkup)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$howlong == 1)); summary(recentmam)
#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)
#
> women = as.numeric(datA$womenHH > 1); summary(women)

```

```

#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
> G3 = as.numeric(datA$genhealth == 3); summary(G3)
> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(dat$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
>   if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
>   if(datA$employ[i] == 1) employ[i] = 1
>   if(datA$employ[i] == 2) employ[i] = 2
>   if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){
>   if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
>   if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
>   if((datA$income[i] == 7) |(datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)

```

```

> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric(datA$smoke == 2); summary(smoke)
#
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
#
> womenhealth = (recentmam + recentpap); summary(womenhealth)
#
> healthcov = as.numeric(datA$healthcov == 1); summary(healthcov)
#
> dat2005 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8, smoke = smoke, sex = datA$sex, recentmam = recentmam, marital = marital,
>   adultHH = adults, womenHH = women, G1 = G1, checkup = checkup, healthplan = healthcov,
>   childHH = datA$childHH, education = E2 + E3, semploy = Em3, employ = Em2, income= I2 + I3,
>   recentpap = recentpap, womenhealth = womenhealth); summary(dat2005)
#
> write.csv(dat2005, "dat2005.csv")
##### 2006
> dat = read.csv("2006.csv")
> dataA = data.frame(weight = dat$x.finalwt, age = dat$age, sex = dat$sex, hadmam =
>   dat$hadmam, howlong = dat$howlong, seatbelt = dat$seatbelt, marital = dat$marital, adultHH
>   = dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth, healthcov =
>   dat$hlthplan, checkup = dat$checkup, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ, income = dat$income2, rent_own = NA, smoke = dat$x.rfsmok3, hadpap =
>   dat$hadpap2, lastpap = dat$lastpap2, doctor = NA, paid = NA); summary(dataA)
#
> dataA = subset(dataA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(dataA)
#
> A1 = as.numeric(datA$age < 35); summary(A1)
> A2 = as.numeric(datA$age > 34 & datA$age < 40); summary(A2)
> A3 = as.numeric(datA$age > 39 & datA$age < 45); summary(A3)
> A4 = as.numeric(datA$age > 44 & datA$age < 50); summary(A4)
> A5 = as.numeric(datA$age > 49 & datA$age < 55); summary(A5)
> A6 = as.numeric(datA$age > 54 & datA$age < 60); summary(A6)
> A7 = as.numeric(datA$age > 59 & datA$age < 65); summary(A7)
> A8 = as.numeric(datA$age > 64); summary(A8)
#
> checkup = as.numeric(datA$checkup == 1); summary(checkup)
#
> seatbelt = as.numeric(datA$seatbelt == 1); summary(seatbelt)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$howlong == 1)); summary(recentmam)

```

```

#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)
#
> women = as.numeric(datA$womenHH > 1); summary(women)
#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
> G3 = as.numeric(datA$genhealth == 3); summary(G3)
> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(dat$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
>   if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
>   if(datA$employ[i] == 1) employ[i] = 1
>   if(datA$employ[i] == 2) employ[i] = 2
>   if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){

```

```

> if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
> if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
> if((datA$income[i] == 7) |(datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)
> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric(datA$smoke == 2); summary(smoke)
#
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
#
> womenhealth = (recentmam + recentpap); summary(womenhealth)
#
> healthcov = as.numeric(datA$healthcov == 1); summary(healthcov)
#
> dat2006 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8, smoke = smoke, sex = datA$sex, recentmam = recentmam, seatbelt =
>   seatbelt, marital = marital, adultHH = adults, womenHH = women, G1 = G1, checkup =
>   checkup, childHH = datA$childHH, education = E2 + E3, semploy = E3, employ = E2, income =
>   I2 + I3, recentpap = recentpap,womenhealth = womenhealth); summary(dat2006)
#
> write.csv(dat2006, "dat2006.csv")
##### 2007
> dat = read.csv("2007.csv")
> dataA = data.frame(weight = dat$x.finalwt, age = dat$age,sex = dat$sex,hadmam =
>   dat$hadmam,howlong = dat$howlong, seatbelt = NA,marital = dat$marital, adultHH =
>   dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth, healthcov =
>   dat$hlthplan, checkup = dat$checkup, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ, income = dat$income2, rent_own = NA, smoke = dat$x.rfsmok3, hadpap =
>   dat$hadpap2, lastpap = dat$lastpap2, doctor = NA, paid = NA); summary(dataA)
#
> dataA = subset(dataA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(dataA)
#
> A1 = as.numeric(datA$age < 35); summary(A1)
> A2 = as.numeric(datA$age > 34 & datA$age <40); summary(A2)
> A3 = as.numeric(datA$age > 39 & datA$age < 45); summary(A3)
> A4 = as.numeric(datA$age > 44 & datA$age < 50); summary(A4)
> A5 = as.numeric(datA$age > 49 & datA$age < 55); summary(A5)
> A6 = as.numeric(datA$age > 54 & datA$age < 60); summary(A6)
> A7 = as.numeric(datA$age > 59 & datA$age < 65); summary(A7)
> A8 = as.numeric(datA$age > 64); summary(A8)

```



```

#
> checkup = as.numeric(datA$checkup == 1); summary(checkup)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$howlong == 1)); summary(recentmam)
#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)
#
> women = as.numeric(datA$womenHH > 1); summary(women)
#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
> G3 = as.numeric(datA$genhealth == 3); summary(G3)
> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(datA$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
>   if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
>   if(datA$employ[i] == 1) employ[i] = 1
>   if(datA$employ[i] == 2) employ[i] = 2
>   if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)

```

```

> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(length(datA$income))
> for(i in 1:length(income)){
>   if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
>   if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
>   if((datA$income[i] == 7) |(datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)
> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric(datA$smoke == 2); summary(smoke)
#
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
#
> womenhealth = (recentmam + recentpap); summary(womenhealth)
#
> healthcov = as.numeric(datA$healthcov == 1); summary(healthcov)
#
> dat2007 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8, sex = datA$sex, recentmam = recentmam, seatbelt = NA, marital = marital,
>   adultHH = adults,smoke = smoke, womenHH = women, G1 = G1, checkup = checkup, childHH =
>   datA$childHH, education = E2 + E3, semploy = E3, employ = E2, income = I2 + I3, healthplan
>   = healthcov, recentpap = recentpap, womenhealth = womenhealth); summary(dat2007)
#
> write.csv(dat2007, "dat2007.csv")
##### 2008
> dat = read.csv("2008.csv")
> dataA = data.frame(weight = dat$x.finalwt, age = dat$age,sex = dat$sex,hadmam =
>   dat$hadmam,howlong = dat$howlong, seatbelt = dat$seatbelt, marital = dat$marital, adultHH
>   = dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth, healthcov =
>   dat$hlthplan, checkup = dat$checkup1, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ, income = dat$income2, rent_own = NA, smoke = dat$x.rfsmok3, hadpap =
>   dat$hadpap2, lastpap = dat$lastpap2, doctor = NA, paid = NA); summary(dataA)
#
> dataA = subset(dataA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(dataA)
#
> A1 = as.numeric(datA$age < 35); summary(A1)
> A2 = as.numeric(datA$age > 34 & datA$age <40); summary(A2)
> A3 = as.numeric(datA$age > 39 & datA$age < 45); summary(A3)
> A4 = as.numeric(datA$age > 44 & datA$age < 50); summary(A4)

```

```

> A5 = as.numeric(datA$age > 49 & datA$age < 55); summary(A5)
> A6 = as.numeric(datA$age > 54 & datA$age < 60); summary(A6)
> A7 = as.numeric(datA$age > 59 & datA$age < 65); summary(A7)
> A8 = as.numeric(datA$age > 64); summary(A8)
#
> seatbelt = as.numeric(datA$seatbelt == 1); summary(seatbelt)
#
> checkup = as.numeric(datA$checkup == 1); summary(checkup)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$howlong == 1)); summary(recentmam)
#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)
#
> women = as.numeric(datA$womenHH > 1); summary(women)
#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
> G3 = as.numeric(datA$genhealth == 3); summary(G3)
> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(datA$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
>   if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
>   if(datA$employ[i] == 1) employ[i] = 1
>   if(datA$employ[i] == 2) employ[i] = 2

```

```

> if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){
>   if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
>   if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
>   if((datA$income[i] == 7) |(datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)
> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric(datA$smoke == 2); summary(smoke)
#
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
#
> womenhealth = as.numeric(recentpap + recentmam); summary(womenhealth)
#
> healthcov = as.numeric(datA$healthcov == 1); summary(healthcov)
#
> dat2008 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 =
>   A6, A7 = A7, A8 = A8, sex = datA$sex, recentmam = recentmam, seatbelt = seatbelt,
>   marital = marital, adultHH = adults, womenHH = women, G1 = G1, checkup = checkup,
>   childHH = datA$childHH, education = E3 + E2,
>   smoke = smoke, income = I2 + I3, healthplan = healthcov,
>   recentpap = recentpap, womenhealth = womenhealth)
> summary(dat2008)
#
> write.csv(dat2008, "dat2008.csv")
##### 2009
> dat = read.csv("2009.csv")
> datA = data.frame(weight = dat$x.finalwt, age = dat$age,sex = dat$sex,hadmam =
>   dat$hadmam,howlong = dat$howlong, seatbelt = NA, marital = dat$marital, adultHH =
>   dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth, healthcov =
>   dat$hlthplan, checkup = dat$checkup1, childHH = dat$x.chldcnt, educa = dat$educa,employ =
>   dat$employ, income = dat$income2, rent_own = dat$renthom1, smoke = dat$x.rfsmok3,

```

```

> hadpap = dat$hadpap2,lastpap = dat$lastpap2, doctor = dat$csrdoc, paid = dat$scntpaid);
> summary(datA)
#
> datA = subset(datA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(datA)
#
> A1 = as.numeric(datA$age < 35); summary(A1)
> A2 = as.numeric(datA$age > 34 & datA$age <40); summary(A2)
> A3 = as.numeric(datA$age > 39 & datA$age < 45); summary(A3)
> A4 = as.numeric(datA$age > 44 & datA$age < 50); summary(A4)
> A5 = as.numeric(datA$age > 49 & datA$age < 55); summary(A5)
> A6 = as.numeric(datA$age > 54 & datA$age < 60); summary(A6)
> A7 = as.numeric(datA$age > 59 & datA$age < 65); summary(A7)
> A8 = as.numeric(datA$age > 64); summary(A8)
#
> seatbelt = NA;
#
> paid = vector(, length(datA$paid))
> paid[which(datA$paid == 1)] = 1
> paid[which(datA$paid == 2)] = 2
> paid[which(datA$paid > 2)] = 3
> paid[which(is.na(paid))] = 3
> summary(paid)
> P1 = as.numeric(paid == 1); summary(P1)
> P2 = as.numeric(paid == 2); summary(P1)
> P3 = as.numeric(paid == 3); summary(P1)
#
> renthome = as.numeric(datA$rent_own == 1); summary(renthome)
#
> checkup = as.numeric(datA$checkup == 1); summary(checkup)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$howlong == 1)); summary(recentmam)
#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)
#
> women = as.numeric(datA$womenHH > 1); summary(women)
#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
> G3 = as.numeric(datA$genhealth == 3); summary(G3)
> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)

```

```

#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(dat$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
>   if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
>   if(datA$employ[i] == 1) employ[i] = 1
>   if(datA$employ[i] == 2) employ[i] = 2
>   if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){
>   if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
>   if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
>   if((datA$income[i] == 7) | (datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)
> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric(datA$smoke == 2); summary(smoke)
#
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)

```

```

#
> womenhealth = (recentmam + recentpap); summary(womenhealth)
#
> healthcov = as.numeric(datA$healthcov == 1); summary(healthcov)
#
> dat2009 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8, sex = datA$sex, recentmam = recentmam, seatbelt = seatbelt, marital =
>   marital, adultHH = adults, womenHH = women, G1 = G1, checkup = checkup, smoke = smoke,
>   childHH = datA$childHH, education = E2 + E3, semploy = Em3, employ = Em2, income = I2 + I3,
>   ownrent = renthome, recentpap = recentpap, healthplan = healthcov, womenhealth =
>   womenhealth); summary(dat2009)
> write.csv(dat2009, "dat2009.csv")
##### 2010
> dat = read.csv("2010.csv")
> dataA = data.frame(weight = dat$x.finalwt, age = dat$age,sex = dat$sex,hadmam =
>   dat$hadmam,howlong = dat$howlong, seatbelt = dat$seatbelt,marital = dat$marital, adultHH
>   = dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth, healthcov =
>   dat$hlthplan, checkup = dat$checkup1, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ, income = dat$income2, rent_own = dat$renthom1, smoke = dat$x.rfsmok3,
>   hadpap = dat$hadpap2, lastpap = dat$lastpap2, doctor = dat$csrvdoc, paid = dat$scntpaid);
>   summary(dataA)
#
> dataA = subset(dataA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(dataA)
#
> A1 = as.numeric(dataA$age < 35); summary(A1)
> A2 = as.numeric(dataA$age > 34 & dataA$age < 40); summary(A2)
> A3 = as.numeric(dataA$age > 39 & dataA$age < 45); summary(A3)
> A4 = as.numeric(dataA$age > 44 & dataA$age < 50); summary(A4)
> A5 = as.numeric(dataA$age > 49 & dataA$age < 55); summary(A5)
> A6 = as.numeric(dataA$age > 54 & dataA$age < 60); summary(A6)
> A7 = as.numeric(dataA$age > 59 & dataA$age < 65); summary(A7)
> A8 = as.numeric(dataA$age > 64); summary(A8)
#
> seatbelt = as.numeric(dataA$seatbelt == 1); summary(seatbelt)
#
> doctor = NA
#
> paid = vector(, length(dataA$paid))
> paid[which(dataA$paid == 1)] = 1
> paid[which(dataA$paid == 2)] = 2
> paid[which(dataA$paid > 2)] = 3
> paid[which(is.na(paid))] = 3
> summary(paid)

```

```

> P1 = as.numeric(paid == 1); summary(P1)
> P2 = as.numeric(paid == 2); summary(P2)
> P3 = as.numeric(paid == 3); summary(P3)
#
> renthome = as.numeric(datA$rent_own == 1); summary(renthome)
#
> checkup = as.numeric(datA$checkup == 1); summary(checkup)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$howlong == 1)); summary(recentmam)
#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)
#
> women = as.numeric(datA$womenHH > 1); summary(women)
#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
> G3 = as.numeric(datA$genhealth == 3); summary(G3)
> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(dat$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
>   if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
>   if(datA$employ[i] == 1) employ[i] = 1
>   if(datA$employ[i] == 2) employ[i] = 2
>   if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3

```



```

> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){
>   if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
>   if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
>   if((datA$income[i] == 7) |(datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)
> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric(datA$smoke == 2); summary(smoke)
#
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
#
> womenhealth = as.numeric(recentpap + recentmam)
> summary(womenhealth)
#
> healthcov = as.numeric(datA$healthcov == 1);summary(healthcov)
#
> dat2010 = data.frame(weight = datA$weight,A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8, sex = datA$sex, recentmam = recentmam, seatbelt = seatbelt, marital =
>   marital, adultHH = adults, womenHH = women, G1= G1, checkup = checkup, childHH =
>   datA$childHH, education = E2 + E3, semploy = Em3, employ = Em2, smoke = smoke,income =
>   I2 + I3, ownrent = renthome, recentpap = recentpap, paid = P1, healthplan = healthcov,
>   womenhealth = womenhealth)
> summary(dat2010)
#
> write.csv(dat2010, "dat2010.csv")
##### 2012
> dat = read.csv("2012.csv")
> datA = data.frame(weight = dat$x.llcpwt, age = dat$age,sex = dat$sex,hadmam =
>   dat$hadmam,howlong = dat$howlong, seatbelt = dat$seatbelt, marital = dat$marital, adultHH
>   = dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth,healthcov =
>   dat$hlthpln1, checkup = dat$checkup1, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ, income = dat$income2, rent_own = dat$renthom1, smoke = dat$x.rfsmok3,

```

```

> hadpap = dat$hadpap2, lastpap = dat$lastpap2, doctor = dat$csrvdoc1, paid = dat$scntpaid);
> summary(datA)
#
> datA = subset(datA, (sex == 2) & hadmam & hadpap & age >= 30 & age < 70); summary(datA)
#
> A1 = as.numeric(datA$age < 35); summary(A1)
> A2 = as.numeric(datA$age > 34 & datA$age < 40); summary(A2)
> A3 = as.numeric(datA$age > 39 & datA$age < 45); summary(A3)
> A4 = as.numeric(datA$age > 44 & datA$age < 50); summary(A4)
> A5 = as.numeric(datA$age > 49 & datA$age < 55); summary(A5)
> A6 = as.numeric(datA$age > 54 & datA$age < 60); summary(A6)
> A7 = as.numeric(datA$age > 59 & datA$age < 65); summary(A7)
> A8 = as.numeric(datA$age > 64); summary(A8)
#
> seatbelt = as.numeric(datA$seatbelt == 1); summary(seatbelt)
> summary(datA$doctor)
> D1 = as.numeric(datA$doctor == 2); summary(D1)
> D2 = as.numeric(datA$doctor == 1 | datA$doctor == 4 | datA$doctor == 7
> | datA$doctor == 8 | datA$doctor == 9); summary(D2)
> D3 = as.numeric(datA$doctor == 3 | datA$doctor == 5 | datA$doctor == 6 |
> datA$doctor == 10); summary(D3)
#
> paid = vector(, length(datA$paid))
> paid[which(datA$paid == 1)] = 1
> paid[which(datA$paid == 2)] = 2
> paid[which(datA$paid > 2)] = 3
> paid[which(is.na(paid))] = 3
> summary(paid)
> P1 = as.numeric(paid == 1); summary(P1)
> P2 = as.numeric(paid == 2); summary(P1)
> P3 = as.numeric(paid == 3); summary(P1)
#
> renthome = as.numeric(datA$rent_own == 1); summary(renthome)
#
> checkup = as.numeric(datA$checkup == 1); summary(checkup)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$howlong == 1)); summary(recentmam)
#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)
#
> women = as.numeric(datA$womenHH > 1); summary(women)

```

```

#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
> G3 = as.numeric(datA$genhealth == 3); summary(G3)
> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(dat$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){
>   if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>     employ[i] = 0
>   if(datA$employ[i] == 1) employ[i] = 1
>   if(datA$employ[i] == 2) employ[i] = 2
>   if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){
>   if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
>   if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
>   if((datA$income[i] == 7) | (datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)

```

```

> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric(datA$smoke == 2); summary(smoke)
#
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
#
> womenhealth = as.numeric(recentpap + recentmam)
> summary(womenhealth)
> healthcov = as.numeric(datA$healthcov == 1); summary(healthcov)
> dat2012 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 = A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8,sex = datA$sex, recentmam = recentmam, seatbelt = seatbelt, healthplan =
>   healthcov, marital = marital, adultHH = adults, womenHH = women, G1 = G1, checkup =
>   checkup, childHH = datA$childHH, education = E2 + E3, semploy = Em3, employ = Em2, income
>   = I2 + I3, smoke = smoke, ownrent = renthome, recentpap = recentpap, oncologist = D2,
>   specialist = D3, paid = P1, healthplan = healthcov, womenhealth = womenhealth);
>   summary(dat2012)

write.csv(dat2012, "dat2012.csv")
# 2013
> dat = read.csv("2013.csv")
> dataA = data.frame(weight = dat$x.lcpwt2, age = dat$x.ageg5yr, sex = dat$sex,hadmam =
>   dat$hadmam, howlong = dat$howlong, seatbelt = dat$seatbelt,marital = dat$marital, adultHH
>   = dat$numadult, womenHH = dat$numwomen, genhealth = dat$genhlth,healthcov =
>   dat$hlthpln1, checkup = dat$checkup1, childHH = dat$x.chldcnt, educa = dat$educa, employ =
>   dat$employ1, income = dat$income2, rent_own = dat$renthom1, smoke = dat$x.rfsmok3,
>   hadpap = dat$hadpap2, lastpap = dat$lastpap2, doctor = NA, paid = dat$scntpaid);
>   summary(dataA)
>
> summdatA = subset(dataA, (sex == 2) & hadmam & hadpap & age >= 3 & age < 11); summary(dataA)
#
> A1 = as.numeric(datA$age == 3); summary(A1)
> A2 = as.numeric(datA$age == 4); summary(A2)
> A3 = as.numeric(datA$age == 5); summary(A3)
> A4 = as.numeric(datA$age == 6); summary(A4)
> A5 = as.numeric(datA$age == 7); summary(A5)
> A6 = as.numeric(datA$age == 8); summary(A6)
> A7 = as.numeric(datA$age == 9); summary(A7)
> A8 = as.numeric(datA$age == 10); summary(A8)
#
> seatbelt = as.numeric(datA$seatbelt == 1); summary(seatbelt)
#
>paid = vector(, length(datA$paid))

```

```

>paid[which(datA$paid == 1)] = 1
> paid[which(datA$paid == 2)] = 2
>paid[which(datA$paid > 2)] = 3
> paid[which(is.na(paid))] = 3
> summary(paid)
> P1 = as.numeric(paid == 1); summary(P1)
> P2 = as.numeric(paid == 2); summary(P1)
> P3 = as.numeric(paid == 3); summary(P1)
#
>renthome = as.numeric(datA$rent_own == 1); summary(renthome)
#
>checkup = as.numeric(datA$checkup == 1); summary(checkup)
#
> recentmam = as.numeric((datA$hadmam == 1) & (datA$showlong == 1)); summary(recentmam)
#
> marital = as.numeric((datA$marital != 1) & (datA$marital != 6)); summary(marital)
#
> adults = as.numeric(datA$adultHH > 1); summary(adults)

> women = as.numeric(datA$womenHH > 1); summary(women)
#
> G1 = as.numeric(datA$genhealth == 1); summary(G1)
> G2 = as.numeric(datA$genhealth == 2); summary(G2)
>G3 = as.numeric(datA$genhealth == 3); summary(G3)
> G4 = as.numeric(datA$genhealth == 4); summary(G4)
> G5 = as.numeric(datA$genhealth == 5); summary(G5)
#
> coverage = as.numeric(datA$healthcov == 1); summary(coverage)
#
> education = vector(,length(datA$educa))
> for(i in 1:length(education)){
>   if((datA$educa[i] < 4)) education[i] = 0
>   if(datA$educa[i] == 4) education[i] = 1
>   if(datA$educa[i] > 4) education[i] = 2
>   if(is.na(dat$educa[i])) education[i] = 0
> }
> summary(education)
> E1 = as.numeric(education == 0); summary(E1)
> E2 = as.numeric(education == 1); summary(E2)
> E3 = as.numeric(education == 2); summary(E3)
#
> employ = vector(,length(datA$employ))
> for(i in 1:length(employ)){

```

```

> if((datA$employ[i] == 3) | (datA$employ[i] == 4) | (datA$employ[i] == 9) | (is.na(datA$employ[i])))
>   employ[i] = 0
> if(datA$employ[i] == 1) employ[i] = 1
> if(datA$employ[i] == 2) employ[i] = 2
> if((datA$employ[i] > 4) & datA$employ[i] < 9) employ[i] = 3
> }
> summary(employ)
> Em1 = as.numeric(employ == 0); summary(Em1)
> Em2 = as.numeric(employ == 1); summary(Em2)
> Em3 = as.numeric(employ == 2); summary(Em3)
> Em4 = as.numeric(employ == 3); summary(Em4)
#
> income = vector(,length(datA$income))
> for(i in 1:length(income)){
>   if((datA$income[i] < 4) | (datA$income[i] > 8)) income[i] = 0
>   if((datA$income[i] >= 4) & (datA$income[i] < 7)) income[i] = 1
>   if((datA$income[i] == 7) | (datA$income[i] == 8)) income[i] = 3
> }
> summary(income)
> I1 = as.numeric(income == 0); summary(I1)
> I2 = as.numeric(income == 1); summary(I2)
> I3 = as.numeric(income == 3); summary(I3)
#
> smoke = as.numeric(datA$smoke != 2); summary(smoke)
> recentpap = as.numeric((datA$hadpap == 1) & (datA$lastpap < 4)); summary(recentpap)
> womenhealth = as.numeric(recentmam + recentpap); summary(womenhealth)
> healthcov = as.numeric(datA$healthcov == 1); summary(healthcov)
> dat2013 = data.frame(weight = datA$weight, A1 = A1, A2 = A2, A3 =A3, A4 = A4, A5 = A5, A6 = A6,
>   A7 = A7, A8 = A8, sex = datA$sex, recentmam = recentmam, seatbelt = seatbelt, marital =
>   marital, adultHH = adults, womenHH = women, G1 = G1, smoke = smoke, healthplan =
>   healthcov, checkup = checkup, childHH = datA$childHH, education = E2 + E3, semploy = Em3,
>   employ = Em2, income = I2 + I3, ownrent = renthome, recentpap = recentpap, womenhealth
>   = womenhealth, paid = P1); summary(dat2013)
> write.csv(dat2013, "dat2013.csv")

```

### STATA CODE for Ordinal Logistic Regression

```
# It is important that any NA values are changed to '' before running analysis
```

```
# 2003
```

```
logit womenhealth a2 a3 a4 a5 a6 a7 a8 marital adulthh womenhh g1 childhh education healthplan
employ income smoke [pweight=weight]
```

```
# 2004
```

```
logit womenhealth a2 a3 a4 a5 a6 a7 a8 marital adulthh womenhh smoke g1 childhh education
employ income healthplan [pweight=weight]
```

# 2005

logit womenhealth a2 a3 a4 a5 a6 a7 a8 smoke marital adulthh womenhh g1 checkup healthplan  
childhh education employ [pweight=weight]

# 2006

logit womenhealth a2 a3 a4 a5 a6 a7 a8 smoke seatbelt marital adulthh womenhh g1 checkup childhh  
education employ income [pweight=weight]

# 2007

logit womenhealth a2 a3 a4 a5 a6 a7 a8 marital adulthh womenhh g1 checkup childhh education  
employ income healthplan [pweight=weight]

# 2008

logit womenhealth a2 a3 a4 a5 a6 a7 a8 seatbelt marital adulthh womenhh g1 checkup smoke income  
healthplan [pweight=weight]

# 2009

logit womenhealth a2 a3 a4 a5 a6 a7 a8 marital adulthh womenhh g1 checkup smoke childhh  
education employ income ownrent healthplan [pweight=weight]

# 2010

logit womenhealth a2 a3 a4 a5 a6 a7 a8 seatbelt marital adulthh womenhh g1 checkup childhh  
education employ smoke income ownrent paid healthplan [pweight=weight]

# 2012

logit womenhealth a2 a3 a4 a5 a6 a7 a8 seatbelt healthplan marital adulthh womenhh g1 checkup  
childhh education employ income smoke ownrent oncologist specialist paid healthplan  
[pweight=weight]

# 2013

logit womenhealth a2 a3 a4 a5 a6 a7 a8 seatbelt marital adulthh womenhh g1 smoke healthplan  
checkup childhh education employ income ownrent [pweight=weight]

## CHAPTER 8: BIBLIOGRAPHY

"About the Behavioral Risk Factor Surveillance System." *Center for Disease Control and Prevention*. 19  
Mar. 2013. Web. 12 Jan. 2015.

Carlos, Ruth C., A. Mark Fendrick, Stephanie Patterson, and Steven J. Bernstein. "Associations in  
Breast and Colon Cancer Screening Behavior in Women." *Academic Radiology* 1 Apr. 2005: 451-58.  
Print.

Carlos, Ruth C., Willie Underwood III, A Mark Fendrick, and Stephen J. Bernstein. "Behavioral  
Associations Between Prostate and Colon Cancer Screening." *Journal of the American College of  
Surgeons* 1 Feb. 2005: 216-23. Print.

"Chronological History of ACS Recommendations for the Early Detection of Cancer in People Without  
Cancer Symptoms." 28 Feb. 2014. Print.

DeSantis, Carol, Rebecca Siegel, Priti Bandi, and Ahmedin Jemal. "Breast Cancer Statistics, 2011." *A Cancer Journal for Clinicians* 1 Jan. 2011: 409-18. Print.

Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2014). Hmisc: Harrell Miscellaneous. R package version 3.14-4. <http://CRAN.R-project.org/package=Hmisc>

Hawkins, Nikki A, Vicki B. Benard, April Greek, Katherine B. Roland, Diane Manninen, and Mona Saraiya. "Patient Knowledge and Beliefs as Barriers to Extending Cervical Cancer Screening Intervals in Federally Qualified Health Centers." *Preventive Medicine* 5 Sept. 2013: 641-45. Print.

J. O. Ramsay, Hadley Wickham, Spencer Graves and Giles Hooker (2014). fda: Functional Data Analysis. R package version 2.4.4. <http://CRAN.R-project.org/package=fda>

"Methodologic Changes in the Behavioral Risk Factor Surveillance System in 2011 and Potential Effects on Prevalence Estimates." *Morbidity and Mortality Weekly Report* 8 June 2012: 410-13. Print.

"New Cervical Cancer Screening Guidelines Announced." *United States Preventive Services Task Force* 1 Jan. 2012. Print.

Nystrom, Lennarth, Ingvar Andersson, Nils Bjurstam, Jan Frisell, and Bo Nordenskjold. "Long-term Effects of Mammography Screening: Updated Overview of the Swedish Randomised Trials." *Lancet* 16 Mar. 2002: 909-19. Print.

Pace, Lydia E., Yulei He, and Nancy L. Keating. "Trends in Mammography Screening Rates After Publication of the 2009 US Preventive Services Task Force Recommendations." *Cancer* 15 July 2013: 2518-523. Print.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

"RomneyCare: What Is RomneyCare?" *Obamacare Facts*. Web. 10 Jan. 2015.  
<<http://obamacarefacts.com/romneycare-romneyhealthcare/>>.

"Screening for Breast Cancer Final Research Plan BACKGROUND." *United States Preventive Services Task Force* 1 Jan. 2009. Print.

Sirovich, Brenda E, and H Gilbert Welch. "The Frequency of Pap Smear Screening in the United States." *Journal of General Internal Medicine* 19.3 (2004): 243-250. PMC. Web. 18 Feb. 2015.

Soni, Anita. "Use of the Pap Test as a Cancer Screening Tool among Women Age 18-64, U.S. Noninstitutionalized Population, 2005." *Medical Expenditure Panel Survey* 1 June 2007. Print.

StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.



T. Lumley (2014) "survey: analysis of complex survey samples". R package version 3.30.

*U.S. Preventive Services Task Force Grade Definitions After May 2007*. May 2008.

<http://www.uspreventiveservicestaskforce.org/uspstf/gradespost.htm>

"What Is a Pap Smear and Why Is It Performed?" *Women's Health and Wellness Resource Network* 1  
Jan. 2014. Print.