2006-10-27

A Logistic Regression Analysis of Utah Colleges Exit Poll Response Rates Using SAS Software

Clint W. Stevenson
Brigham Young University - Provo

A LOGISTIC REGRESSION ANALYSIS OF UTAH COLLEGES EXIT POLL RESPONSE RATES USING SAS® SOFTWARE

by

Clint Wesley Stevenson

A Project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

December 2006

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a Project submitted by

Clint Wesley Stevenson

This Project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Howard B. Christensen, Chair
Date	Scott D. Grimshaw
Date	Bruce J. Collings

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the Project of Clint Wesley Stevenson in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Howard B. Christensen
Chair, Graduate Committee

Accepted for the Department

Scott D. Grimshaw
Graduate Coordinator

Accepted for the College

Thomas W. Sederberg
Associate Dean, College of Physical and Mathematical Sciences

ABSTRACT

A LOGISTIC REGRESSION ANALYSIS OF UTAH COLLEGES EXIT POLL RESPONSE RATES USING SAS® SOFTWARE

Clint Wesley Stevenson

Department of Statistics

Master of Science

In this study I examine voter response at an interview level using a dataset of 7562 voter contacts (including responses and nonresponses) in the 2004 Utah Colleges Exit Poll. In 2004, 4908 of the 7562 voters approached responded to the exit poll for an overall response rate of 65 percent. Logistic regression is used to estimate factors that contribute to a success or failure of each interview attempt. This logistic regression model uses interviewer characteristics, voter characteristics (both respondents and nonrespondents), and exogenous factors as independent variables. Voter characteristics such as race, gender, and age are strongly associated with response. An interviewer's prior retail sales experience is associated with whether a voter will decide to respond to a questionnaire or not. The only exogenous factor that is associated with voter response is whether the interview occurred in the morning or afternoon.

Acknowledgements

Several people have contributed to the completion of this project. I would like to give a special thanks to my advisor Dr. Howard Christensen. He has been a great instructor providing wonderful insights and support. Dr. Scott Grimshaw who provided the direction to use and complete this project. The BYU Statistics department faculty who have given their support and encouragement in many ways. Dr. Kelly Patterson and Daniel Magleby for their involvement when the project first began. Dr. Quin Monson for accompanying me to the American Association for Public Opinion Research conference in Miami Beach, Florida. My wonderful wife, Andrea, who continually pushed me to complete this project. Most of all my Heavenly Father who provided me with the opportunity to start and complete this degree.

Contents

Chapter

Tables

Table

Chapter 1

Introduction

1.1 Overview

Exit polls have become a prominent feature of United States elections. National news organizations feature the results of such polls in their election coverage and academics examine the data from exit polls to understand the factors that shaped the election's outcome. Edison/Mitofsky currently provides national exit poll results to these news organizations. Exit polls are somewhat unique among polls because there are actual vote counts against which to compare the poll results. Therefore, the methods for gathering the data attract attention when the polls do not agree with the actual results.

Because of the relative infrequency of elections and the fact that elections are typically one day events, exit polls face unique challenges. Statewide and national exit polls must interact with a variety of political jurisdictions, all of which can affect the manner in which the polling takes place. Researchers must consider a variety of factors that can affect the probability of getting a voter to fill out a questionnaire. This paper seeks to contribute to the research available on

voter nonresponse in exit polls and specifically examines the ways in which voter and interviewer characteristics affect the probability of obtaining a response to an exit poll interview. This research primarily concentrates on the personal and exogenous dynamics that may lead a voter to agree to complete a questionnaire. The presence of nonresponse does not necessarily mean the survey is biased (exit poll accuracy is a topic of another discussion). This paper seeks to understand factors that influence a voter to initially respond to an exit poll questionnaire; therefore, item nonresponse is not addressed here.

In order to understand the dynamics of the interviewer and voter, the personal characteristics and exogenous factors related to each must be considered. To investigate the effect of these factors on voter response, a logistic regression model is formed using Utah Colleges Exit Poll data from 2004.

1.2 Utah Colleges Exit Poll History

The Utah Colleges Exit Poll (UCEP) began in 1982 under the direction of Dr. David Magleby of the Political Science Department and Dr. Howard Christensen of the Department of Statistics at Brigham Young University. Every two years since the first exit poll in 1982 students from these departments have worked together to conduct the exit poll in cooperation with participating colleges and universities from the state of Utah (Grimshaw, Christensen, Magleby, and Patterson 2004).

The first exit poll in 1982 utilized only one questionnaire with 34 ques-

tions. In 1986 the number of unique questionnaires increased to three. Since that time the number of questionnaire forms has ranged from three to five. Most recently, in 2004, four questionnaires were used with one questionnaire soliciting email addresses and basic demographic information to be used later for voluntary participation in an Internet survey (Mock, Christensen, Stevenson, Olsen, and Patterson 2006, p. 14).

In 2003 an off-year exit poll was conducted for several local races in Salt Lake City. This was the first time an interviewer questionnaire was implemented. The 2004 exit poll adopted a similar interviewer questionnaire to be administered to the 666 student interviewers attending the training meeting (542 actually participated on Election Day). Each interviewer attended a one hour training session that included a mixture of lecture, demonstration, and role-playing. After a brief introduction of exit poll methodology and procedure, three video clips were shown to demonstrate both proper and improper techniques. After each video clip trainers led a discussion about what the students thought the interviewer did or did not do well and asked for suggestions on how to improve the interviewing process. The demonstrations and discussions were followed by role-playing exercises in which students were given typical voter scenarios and asked to alternate playing the role of voter and interviewer. The interviewer questionnaire was administered at the end of the one hour training session. Training materials were made available on a web site (http://exitpoll.byu.edu) and students were encouraged to review these materials before Election Day. The interviewer training meetings were scheduled

at each of the participating colleges and universities. Having the materials available in electronic format facilitated training at multiple college campuses.

On Election Day students were divided into groups of three or four and given interviewing location assignments. Depending on their availability, interviewers were assigned full-day or half-day shifts. Therefore, it is common to have a total of eight interviewers at one polling place throughout Election Day. Because polling places were staffed with several interviewers, questionnaires could be distributed continually, so all voters designated to be in the sample were given the opportunity to participate. In other words, the 'miss rate' in the UCEP is zero. Edison Media Research and Mitofsky International discuss misses more fully in their evaluation of the 2004 National Exit Poll (Edison Media Research and Mitofsky International 2005). To handle problems, answer questions, and provide monitoring and direction, several roving teams of supervisors checked in with interviewers throughout the day in Salt Lake and Utah Counties. Two additional roving teams were sent to outer lying counties (Stevenson et al. 2004, p.48).

The goal of this paper is to determine ways to effectively improve a voter's inclination to respond to a questionnaire during the UCEP thus improving the overall efficiency of the exit poll. This paper will also provide recommendations on how to better train and position interviewers on Election Day. Chapter 5 will discuss these variables and describe methods to improve the response rate in the UCEP. Chapter 5 will also discuss factors that are associated with voters responding to a questionnaire that are impossible to change through administration. Voter

characteristics such as age, race, and sex are examples of factors that cannot (or should not) be changed.

Chapter 2

Literature Review

The Utah Colleges Exit Poll has been functioning for many years and has undergone many changes. As early as 1990 nonresponse adjustments were made in the UCEP to correct the sampling rate. Over the years several studies have been conducted to obtain better estimates of nonresponse and to better understand the dynamics of the voters. Mock, Christensen, Stevenson, Olsen, and Patterson (2006, Chap. 2) highlight these and other changes from 1982 to 2004.

The Utah Colleges Exit Poll, like any other survey, faces the challenge of getting people to participate. An official document published by the American Association for Public Opinion Research (AAPOR) states that to "maximize cooperation or response rates with the ethical treatment of human subjects" is among the "best practices" of survey research (AAPOR 1997). In a more recent document, AAPOR notes that the number of people that agree to participate in relation to the number of people invited to participate is the basis for calculating a response rate for any survey (AAPOR 2004). For the purposes of this paper, the focus will be on three categories thought to be associated with voter

response in surveys: first, the characteristics and attitudes of the person selected to participate; second, the characteristics and skill set of the interviewer; and finally, exogenous factors such as location of the polling place, time of day, and the interviewer location relative to the polling place.

The AAPOR Standard Definitions (AAPOR 2004) provides notation for calculating response rates. Slater presents this notation for use in an exit poll (Slater 2002, pp. 41–42). Using this notation, the response rate for the UCEP is defined to be

$$RR = \frac{I + P + B}{I + B + P + R + NC}. \tag{2.1}$$

RR=Response Rate

I=Completed Interview

P=Partial Interview

B=Break-off

R=Refusal

NC=Noncontact (or misses)

However, because the UCEP is administered using three to four interviewers per polling place at any given time, all eligible voters are contacted. Therefore $NC = 0$, resulting in

$$RR = \frac{I + P + B}{I + P + B + R}. \tag{2.2}$$

Research on nonresponse in surveys provides interesting insights into rea-

sons selected individuals choose not to participate. A selected individual may not participate because it is simply not convenient. Groves and Couper suggest that convenience for the respondent is the key to a successful interview (1998, p. 34). They maintain that most people do not decline to be interviewed because they are opposed to participation in a poll. Rather, selected individuals typically choose not to participate because they have something to do that they perceive as more important. There are costs associated with taking time to complete a lengthy survey, and very few benefits. If a voter perceives participation in a survey to be too costly, they will decline participation. Evidence of this cost benefit relationship can be seen in studies that indicate paying respondents before the survey for their time has had a positive impact on response (Groves and Couper 1998; Singer, Hoewyk, and Maher 2000, pp. 413–428). Dillman refers to this as "social exchange," where social norms of reciprocity enhance the likelihood of participation as a result of small rewards paid in advance (1978). By increasing the benefit of participation, selected individuals will be more likely to participate. Following this logic, a way to influence voter response is to reduce the costs of participation for the individuals invited to participate.

Trust is another factor that affects survey response rates. Different respondents assign varying levels of trust to the interviewers carrying out the poll and to the organizations they represent. Tourangeau, Rips, and Rasinski find that a positive rapport between an interviewer and respondent is an important factor in predicting participation in a poll (2002, Chap. 10). These dynamics between

selected participants and interviewers is the focus of this research.

Similarly, a familiarity with the organization the interviewer represents contributes to the likelihood of a successful interview. Groves and Couper claim that an assurance that the organization will maintain confidentiality of their response is an important factor in determining whether a recipient will participate (Groves and Couper 1998, p. 34). Notably, Keeter, Miller, Kohut, Groves, and Presser (2000, pp. 125–148) examined how different techniques influence response rates and the accuracy of telephone polls. In their study, Keeter and his colleagues carried out two surveys, one that they identified as standard, and one that they identified as rigorous. The methodology of each survey was similar, except the rigorous survey made "exhaustive efforts" to notify in advance and then find and survey individuals that the first wave of interviews missed. While the exhaustive efforts increased response rates, these efforts did not produce other notable differences between the two polls. In telephone surveys or personal interviews, polling organizations might be able to increase trust in their organization by sending letters to selected individuals ahead of the poll that explain the survey and what the gathered information will be used for. This will make the polling organization more recognizable for the respondent. If a person selected to participate in a survey recognizes the organization conducting the poll and views that organization as reputable and trustworthy, he or she will be more likely to participate in the survey. However, personalized contact in advance is not possible with exit polls except through advanced publications using the media. Nevertheless, a positive

rapport can be quickly established by clearly identifying the polling organization by using survey tables or booths on Election Day.

Singer, Frankel, and Glassman discuss how the interviewer affects data quality. For example, the interviewer's race, age, and/or gender could affect response rates. Even the interviewer's expectation about the response rate of a survey recipient could affect the rate of response. It follows from his discussion that interviewer characteristics may affect the probability of a successful interview (1983, pp. 68–83). Lavrakas suggests that, even over the telephone, interviewers are able to successfully identify a respondent's gender, race, and age for a majority of household level refusals (Bauman, Merkle, and Lavrakas 1992; Lavrakas 1993, Chap. 5). Conversely it would follow that a respondent will be able to accurately identify the interviewer's race, age and gender. During an in-person survey it may be possible for interviewer characteristics to influence how the respondent perceives the interviewer and thus how the selected individual responds to a request to participate. Groves and Couper state, "Interviewers are often demographically 'matched' with the race or ethnic composition of neighborhoods where fieldwork is to be done in hopes of minimizing social distance and increasing response rates" (1998, p. 34). Similarly, Edison and Mitofsky's report on their 2004 Exit Poll indicates that their final sample included a disproportionate number of younger voters as a result of an uncharacteristically younger interviewer corps hired to carry out the survey. Therefore, the increased number of younger interviewers was 'matched' with the younger voters (2005, pp. 42-46).

Merkle and Edelman, using Voter News Service (currently Edison/Mitofsky) data from 1992 and 1996, conclude that interviewer age and voter age are important factors influencing response. Multiple regression was implemented with the overall precinct-level response rate as the dependent variable. They showed that older interviewers produce higher response rates, whereas older voters have lower response rates (2002, pp. 243–257).

In the 1996 Voter News Service Exit Poll, Edelman and Merkle experimented with different factors to see how they contributed to precinct-level response rates. In some polling locations they distributed pens as an incentive to encourage sampled individuals to participate in the exit poll; in other locations they used a folder that contained instructions intended to standardize the approach made by interviewers. In some locations they used a combination of both pens and folders. By comparing these different factors against a control group, where interviewers used neither pens nor folders, they could examine the effect on response rates. They found that both the pen incentives and the instruction folders increased response rates. Notably, they also concluded that increased response rate did not contribute to increased accuracy of the exit poll results (2000).

Lavrakas concludes that an interviewer's skill set plays an important roll in response rates. An interviewer's skill set should include skills that would build a positive rapport and persuade people to respond to the questionnaire. To do this, interviewers should be warm, confident, and outgoing; in other words, interviewers should do those things that instill trust in potential respondents. He also indicates

that an interviewer's ability to tailor persuasion techniques to an individual respondent leads to a successful interview. General overviews for conducting survey research usually include some suggestions for screening and training interviewers on how to engage respondents (1993, Chap. 5). Groves suggests that if the recipient shows signs of active listening and the interviewer maintains this interaction, the response rate increases (2002, p. 132).

Finally, exogenous factors also affect the probability of obtaining a response. These factors would consist of anything that is not endogenous to the interaction between the interviewer and the recipient. Concerned principally with how nonresponse affects the accuracy of the poll, Merkle and Edelman found that certain environmental factors like interviewer location outside the voting place and weather conditions influence the likelihood of voter participation in a survey. They noted that the interviewer's location influences the authenticity that voters assign to the polling organization and to those conducting the interview; the closer interviewers were to the actual polling place the more likely they were to have a successful interview. Other factors such as weather on Election Day influence the rates at which people would participate in a survey that is typically conducted outside (2002, pp. 243–257).

Chapter 3

Methodology

3.1 Data Sources

The 2004 UCEP provides a substantial amount of data about the voter, the interviewer, and the polling location. During this exit poll, questionnaires were given to both the interviewers and voters to obtain this information. The data for this discussion comes from four sources: interviewer questionnaires filled out during interviewer training, exit poll questionnaires obtained on Election Day, polling place questionnaires obtained during the interviewers' check-in phone call, and voter nonresponse records obtained by interviewer observation after the voter declined to complete the questionnaire.

3.1.1 Interviewer Questionnaires

Interviewer information was gathered from a questionnaire that was administered during the training session conducted two to three weeks prior to Election Day. The questionnaires were given out as the interviewers entered the training room and were collected as the interviewers left. This questionnaire included

demographic questions such as gender, race, age, marital status, and university major. It also included questions regarding previous work experience (retail, door-to-door sales, waiter/waitress, and telemarketing experience), the reason for volunteering as an exit poll interviewer, as well as questions designed to gauge the interviewers' learning ability and comfort level in approaching strangers (see Appendix B.1).

3.1.2 Polling Place Questionnaires

When interviewers first arrived at their polling place on Election Day they were instructed to call in and report their arrival. During this first call, the telephone operators were instructed to ask the interviewers a series of questions relating to the polling place. Questions included the interviewer's proximity to the polling place, whether the interviewer was located inside or outside of the polling place, and the interviewer's access to a table (see Appendix B.2).

3.1.3 Exit Poll Questionnaires

Voters provided information on Election Day through exit poll questionnaires. This information includes candidate choices, opinions on important issues, and demographic characteristics. There were three versions of this questionnaire in 2004 printed on blue, yellow and pink paper. Roughly 40 questions were on each version, and approximately half of those questions were a standard set of demographic and candidate choice questions found on all three versions. The remaining

questions varied depending on the color of the questionnaire (see Appendix B.3). In 2004 an additional questionnaire was introduced known as 'the white form.' This questionnaire asked a few demographic questions, but primarily requested email addresses so the selected voters could later be contacted for participation in an Internet survey. Because the goal of the white form was different from the goal of the other three forms it will not be used in the analysis presented here.

3.1.4 Voter Nonresponse Records

Nonrespondent information was obtained by interviewer observation. When a voter declined to complete a questionnaire the interviewer assessed and recorded the voter's gender, race, and age category. The interviewer also recorded the time of day the interview took place to the nearest quarter hour. This information was recorded directly on the refused questionnaire (see Appendix B.3). Because each polling place is staffed with three to four interviewers, all voters designated to be in the sample were approached. Therefore, there are no 'misses' as defined by Edison and Mitofsky (2005, p. 8). When one interviewer is reporting the results to the call center another interviewer can continue to distribute the questionnaires to voters.

3.2 Combining Data Sources

The concept for this paper was developed using three categories that influence a voter's willingness to respond to a questionnaire in the 2004 UCEP. These

categories are interviewer characteristics, voter characteristics, and other exogenous characteristics. The first two categories are based on questionnaires given to the interviewer and the voter as detailed in sections 3.2.2 and 3.2.3. The third category is based on the physical surroundings of the interviewer and the polling place as detailed in Section 3.2.4. Exogenous factors are those that the interviewer generally cannot control. These data sources are then combined using the unique polling place ID and the unique interviewer ID.

3.2.1 Data Management

The data for this project is sourced from five different locations. All data is stored as a Comma Separated Value (CSV) file. The first data source is the design and predictions file (design_and_predictions.csv). This file contains all necessary information to handle the complex design and all preliminary voter turnout and nonresponse estimates. The second source is actual turnout and vote returns (actual_overall_smry.csv). This data originates from the State Elections Office. After the election is complete Utah generally makes the actual precinct-level data available for a nominal cost. Since county results are not formatted consistently from one county to the next this file must usually be formatted to use in any statistical software packages. The third source is the raw data (raw_data.por). It is made available through the exit poll data entry system. This data comes from the voter exit poll questionnaires. The fourth source is the interviewer questionnaire (interviewer_information.csv). This information is collected during the interviewer

training and entered into a Microsoft Excel file by volunteers and students enrolled in exit poll courses. Lastly, the polling location information is obtained during the interviewer's first call. This file also contains any additional research and experiments that students or faculty want to conduct (aux_info.csv).

The most powerful and cleanest way to handle the combining of these different datasets is to use SAS PROC SQL which joins/merges two separate datasets using a common variable relationship. With just a few lines of code all files can be intuitively joined together. See Appendix C for details regarding the SAS code.

3.2.2 Interviewer Characteristics

Interviewers play an important role in the exit poll process because they are the human contact the voter has with the exit poll. As a result it is believed that interviewers have varying abilities to influence whether a voter will initially respond to an exit poll questionnaire. The primary intention of the interviewer questionnaire, in both the 2003 and 2004 UCEP, was to help determine if interviewers will influence whether a voter will respond to the exit poll questionnaire or not. The interviewer questionnaire contains questions including demographics, reason for participating in the UCEP, and previous work experience. The interviewer questions are detailed specifically in Table 3.1. A new variable is also created to assess the interviewer's ability to accurately read instructions. This is accomplished by comparing the 'What is your sampling interval' question from

the interviewer's first call-in to the interval established prior to Election Day. This value is stored in the design and predictions file.

Table 3.1: Interviewer Questions

Topic	**Actual Question**	**Responses Values**
Sex	Are you:	1) Male 2) Female
Race	Are you:	1) Native American/Indian 2) Asian 3) African American/Black 4) Hispanic/Latino 5) White/Caucasian 6) Pacific Islander 7) Other
Age	What year were you born	(open)
Marital Status	Are you presently	1) Married 2) Single 3) Divorced 4) Engaged 5) Cohabiting with boyfriend/girlfriend 6) Widowed
Year in School	What is your year in school	1) Freshman 2) Sophomore 3) Junior 4) Senior 5) Graduate Student 6) Other 7) Don't Know
Home State	What state are you from	(open)
LDS Mission	Have you served and LDS mission	1) Yes 2) No
Height	How tall are you	Feet Inches
Reason for taking course	Are you in Political Science classes for (check all that apply)	1) Major Requirements 2) Minor Requirements 3) GE Requirement 4) Interest 5) Fun 6) Other
Outgoing Index	On a scale from 1 to 5, with 1 being very shy and 5 being very outgoing, where would you place yourself	1) Very Shy to 5) Very Outgoing
Previous Work	Have you ever worked in any of the following jobs (check all that apply)	1) Retail Sales 2) Door to Door Sales 3) Waiter/Waitress 4) Telemarketing/Surveys 5) None
Learning Ability	In a school or work setting, how long does it usually take you to master a new task, with 1 being very slow and 5 very fast	1) Very slow to 5) Very fast
Quality of Training	On a scale of 1 to 5 how helpful was the interviewer training session with 1 being not very helpful and 5 being very helpful	1) Not at all helpful to 5) Extremely helpful
Reason for Participation	What is the primary reason for participating in the Utah Colleges Exit Poll	1) My professor offered extra credit 2) To fulfill a course requirement 3) Interest in public opinion and polling 4) My friends are participating 5) A general interest in politics 6) Another reason not listed here
Confirming Polling Place Information	What is your sampling interval?	(open)
Interviewer Reliability	What time did you arrive at your polling place	(open)

3.2.3 Voter Characteristics

Voter characteristics will often influence the initial voter response. A large amount of information was collected on the respondents. Information collected on nonrespondents is limited to what an interviewer is able to observe. During the 2004 UCEP three items were obtained from both the respondents and nonrespondents as listed in Table 3.2.

Table 3.2: Voter Characteristics and Questions

Topic	**Actual Question**	**Responses Values**
Respondent Sex	Are you	1) Male 2) Female
Nonrespondent Sex	Gender	1) Male 2) Female
Respondent Race	Are you	1) Native American/Indian 2) Asian 3) African American/Black 4) Hispanic/Latino 5) White/Caucasian 6) Pacific Islander 7) Other
Nonrespondent Race	Race	1) White 2) Hispanic 3) Other
Respondent Age	What year were you born?	(open)
Nonrespondent Age	Age of nonrespondent	1) 18-35 2) 36-55 3) 56-75 4) 75+

In Table 3.2 the respondent race and age are recoded so that nonrespondents and respondents would have the same question-response scale. On the exit poll questionnaire the age question asked for the numeric age of the voter. However, on the nonresponse record the interviewer record was an age group of 18-35, 36-55, and 56+. Likewise, the respondent race question contained response levels of Native American, Asian, Black / African American, Hispanic / Latino, White / Caucasian, Pacific Islander, and other. The nonresponse record contained White, Hispanic, and other.

3.2.4 Exogenous Characteristics

In addition to voter and interviewer characteristics there are many other factors that have the potential to influence the response of a voter during an exit poll. Exogenous characteristics are those other than interviewer and voter characteristics. The exogenous characteristics investigated in this analysis are detailed in Table 3.3. The time of day variable was recorded when the interview occurred to the nearest 15 minutes. For the analysis presented here that variable is recoded as 1=PM and 0=AM.

Table 3.3: Exogenous Characteristics and Questions.

Topic	**Actual Question**	**Responses Values**
Polling Place Location		1) Wasatch Front 2) Outside the Wasatch Front
Time of Interview	Time (nearest 15 min)	(open)
Location of Interviewer	Are you located inside the polling place	1) Yes 2) No
Exit Poll Banners		1) Yes 2) No
Tables	Do you have access to a table?	1) Yes 2) No
Election Judge Cooperation	Is the Election Judge allowing you to conduct the survey?	1) Yes 2) No

3.3 Sample Design

The Utah Colleges Exit Poll uses a complex sample design incorporating stratification and clustering along with probability proportionate to size (PPS) sampling. During the 2004 exit poll, Utah was stratified into 17 strata following county and Congressional District boundaries. Some strata in each congressional district were regarded as 'certainty' counties based on university and college par-

ticipation. The other strata were created after the certainty counties were established. Counties within these strata were then randomly selected with probability proportional to expected turnout. Other counties, for administrative reasons, were treated as individual strata.

Polling places, in many cases, contain more than one voting precinct. At these polling places it is difficult to determine the exact precinct of a voter. As a result, all information from precincts are brought together within a polling place. Therefore, UCEP defines the primary sampling unit within a county to be the polling place, not the precinct, except in the case where the precinct is the polling place. In the 2004 Utah election there were 1023 polling places and 469 of those were single-precinct polling places. Of all the polling places 90 were randomly selected proportional to the polling place's expected voter turnout. In 2004 the target was 100 completed questionnaires at each polling place. A fixed sampling interval of $\frac{et}{100/(1-nr)}$ was computed and a random start between one and the interval number was set, where nr is the estimated nonresponse and et is the estimated expected turnout at a given polling place (Stevenson et al. 2004, p. 25). Voters were then systematically sampled to participate in the survey at the specified rate. If a voter refused to participate then the interviewer continued with the interval and then selected that voter. Weights were calculated and assigned to the sampled voter based on the probability of selection of the said voter using estimated voter turnout; no other age, race or gender weight-class adjustments were made (however, for the 2004 exit poll a nonresponse adjustment was made

to the weights). The relationship between sample weights and the probability of selecting the i^{th} voter is $weight = \frac{1}{P(i)}$. Therefore, each sampled voter in the UCEP can be viewed as representing a specified number of voters in the population (See Section 3.3.2) (Mock et al. 2006, pp. 89–97).

3.3.1 Design Adjustments for Analysis

In 2004 there were four levels to the sample design: strata, county, polling place, and voter. At the county level there was often only one county within a stratum. However, in three of the 17 strata there were three counties in each stratum. These three strata are #5, #13, and #17. The other 14 strata had one county per stratum. This analysis will eliminate the county level boundaries of the design, resulting in only three levels of the design. Specifically it will assume that Morgan, Box Elder, and Summit counties (stratum 5) are simply the same county; Duchesne, Uintah and Emery (stratum 13) are the same county; and Beaver, Sevier, and Millard (stratum 17) are the same county (see (Mock et al. 2006, p. 23) for specific stratum and county identification coding). Eliminating the county boundaries is done for two reasons. First, the counties within the strata are, for the most part, politically and geographically similar. The counties' political affiliation tends toward very strong Republican with the exception of Summit county. Further investigation suggests that combining Summit with Morgan and Box Elder counties does not compromise candidate estimates or other summary statistics. Second, eliminating an unnecessary level of the design simplifies the

analysis while continuing to provide accurate results.

This design adjustment creates a stratified design with two stages, PPS sampling of the polling places (with replacement) and systematic sampling of the voter within the polling place. Simple random sampling theory is assumed for the estimation process of systematic sampling.

3.3.2 Sampling Weights

Each of the three levels of the design (after eliminating county boundaries within stratum) are taken into account so that the weight of any voter can be easily computed using the inverse of their probability of selection. Using notation presented by Mock et al. (2006), polling places are defined by the subscript i and the stratum is defined by the subscript h. The actual number of voters at a given polling place, within stratum h, is defined as M_{hi}, M_h is the actual total number of voters within a given stratum. The number of voters approached by an interviewer during the exit poll at any given polling place (both respondents and nonrespondents) is m_{hi} and n_h is the total number of polling places selected within the stratum. Each voter within a given polling place will have the same probability of selection. The sampling weights presented here are used in the analysis found in Section 4.2.3.

Prior to Election Day M_i and M_{hi} are unknown and they must be estimated by $\widehat{M_i}$ and $\widehat{M_{hi}}$ respectively. After Election Day these values are known and are made available by the State Elections Office. This paper proposes both weights

in the analysis for comparison purposes. One will be referred to as *predicted* and the other as *actual*; a third type will use weights of one for all voters and will be referred to as *unweighted*. The expression for the weight is:

$$w_{hi} = \frac{M_h}{m_{hi} \cdot n_h}. \tag{3.1}$$

The weight could also be seen as the average number of voters per polling place, M_h/n_h, divided by the number of people sampled within that polling place, m_{hi}. Because this is a probability sample it is important to recognize each observation's probability of selection.

3.4 Software Limitations and Considerations

Recently many statistical packages have begun to incorporate complex sampling techniques into their software. SAS and SUDAAN are two of the prominent statistical software packages that handle complex multistage probability sample designs. The analyses presented in this paper will use SAS version 9.1 to organize, manage, and analyze the data. The ideal software package is one that provides model selection functions and adjusts for a complex sample design. Currently neither SAS or SUDAAN software packages provide both functions in one procedure.

Up until recently SAS provided no functionality to calculate any complex survey sampling variance estimates. As a result SUDAAN has been used for analysis of complex sample designs in the UCEP. However, in version 7, SAS included three experimental routines for analyzing complex surveys with SURVEYMEANS,

SURVEYREG, and SURVEYSELECT procedures. These procedures entered production in SAS version 8 (An and Watts 1999). In version 9.1 SAS added the SURVEYFREQ and SURVEYLOGISTIC procedures. The SAS, PROC SURVEYLOGISTIC, is used to calculate both the coefficients and the standard error estimates of logistic regression under a complex sample design (SAS Institute Inc. 2006). This procedure will be used for the analysis in Chapter 4. This procedure also produces the standard error estimates based on the complex design in order to calculate the test statistics and test hypotheses about population parameters. With this procedure, independent variables can be tested to determine if they should be included in a logistic regression model.

To test the new SAS capabilities, SUDAAN was also used to calculate coefficient and standard error estimates. The coefficient estimates, and their associated p-values for the data analyzed were found to be the same as those obtained from SAS. Consequently, SAS alone is used to perform all analysis presented in the discussion and no analysis will be provided using SUDAAN.

One PROC LOGISTIC consideration is that the sample size is regarded to be equal to the sum of the weights. In the case of UCEP this will inflate the sample size to be either the estimated population size or the true population size, depending on which weighting scheme is used. Therefore it is important to apply normalized weights when using PROC LOGISTIC. No adjustment is required for the SURVEYLOGISTIC procedure because it already accounts for this in its computations.

3.5 Model and Analysis

The primary research purpose of this paper is to determine the factors that effectively predict a voter's inclination to respond to an exit poll questionnaire using data from the 2004 UCEP. This will be presented in Chapter 4; Chapter 5 will provide recommendations on ways to improve voter response rates.

A logistic regression model will be constructed using backward elimination to determine significant and meaningful independent variables. This strategy will eliminate variables from an overly complex model rather than adding variables to an overly simplistic model. Initially all interviewer characteristics, voter demographics characteristics, and exogenous characteristics without interactions are investigated in the first pass through the data. A total of 28 initial variables will be included. Of the 28 variables there are three that come from the voter characteristics, five from the exogenous characteristics and the remaining 20 are interviewer characteristics. Each of the 28 variables will be examined and will remain in the model if found to be significant at $\alpha = 0.05$. Insignificant factors will be removed from the model and the remaining factors will comprise the final, reduced model. Interactions will be tested once the number of factors has been reduced to a manageable number.

To ensure proper model fitting, three datasets are used for analysis. The first dataset is a *research* dataset, the second is the *validation* dataset, and the third is the *original data*. The original data is the union of the research and vali-

dation datasets. The original data is divided into two mutually exclusive groups with each observation randomly assigned to one of the two groups. This is done by randomly selecting roughly one-half of the exit poll observations from each polling place to be included in the research dataset. The other observations form the validation dataset. This creates two datasets that mirror the original data's sample design. The research and validation datasets are used to help ensure proper model fitting. However, all conclusions presented in Chapter 4 use the original dataset after model selection on the research and validation datasets.

Because voters are nested within a multistage sample design, assuming a simple random sample is not appropriate. Omitting the complexity of the sample design results in standard errors that are too small. The stratification and clustering of the polling places must be taken into consideration when calculating the standard error of the coefficients. Sharon Lohr concisely describes cluster sampling and the standard error, "Whereas stratification generally increases precision when compared with simple random sampling, cluster sampling generally decreases it" (1999, p. 132). By using a simple random sampling approach with the 2004 UCEP the standard error will be too small. Therefore, the potential exists that more variables will falsely be allowed to remain in the model.

Model selection options and the ability to accurately calculate standard errors based on a complex sample design are not available in one SAS procedure. The PROC LOGISTIC procedure has a convenient model selection function, but assumes a simple random sample. This is useful to quickly establish a model using

a large number of initial variables. However, because this procedure assumes a simple random sample it computes a smaller measure of standard error and will erroneously leave too many variables in the model when there is a complex sample design in place. Therefore, this procedure is useful but not entirely adequate for analysis on the UCEP data because it does not account for complex sample design. PROC SURVEYLOGISTIC will account for a complex sample design but, unfortunately, does not have a model selection function. Therefore, a combination of the two procedures is used to build an appropriate model. PROC SURVEYLOGISTIC is used to narrow the remaining variables left in the model from the LOGISTIC procedure. The variables in either of these procedures are removed when they do not meet the $\alpha = .05$ criterion.

Models will first be fit to the research and the validation data using the three different methods to calculate weights to decide which variables to include in the model. This will help establish variables that will be fit to the complete/original data.

The analyses presented in Chapter 4 incorporate normalized weights. This type of weight is useful because rather than having the weights sum to the population total of 867253 they will sum to the sample size, in this case 7562. This will preserve the relative adjustment for differing voter selection probabilities but keep the proper sample size. The process of normalizing can be found in Equations 3.2 and 3.3. This feature is particularly important when using PROC LOGISTIC because without normalized weights SAS will assume the population total to be

the sample size.

$$w_v^{norm} = \frac{7562}{\sum_{v=1}^{7562} w_v} \cdot w_v, \tag{3.2}$$

$$\sum_{v=1}^{7562} w_v^{norm} = 7562. \tag{3.3}$$

Once a preliminary model is established using both the research and validation datasets the model will be applied to the original/complete set of data using normalized weights. The logistic regression model selection for the UCEP dataset will be accomplished through a series of steps.

(1) Randomly divide the complete data into a research group and a validation group.

(2) Use PROC LOGISTIC and backward elimination on the 28 independent variables with normalized weights and apply this procedure to the research and validation dataset in order reduce the number of variables to be included in the final model.

(3) Construct a model, including interactions, using the variables from step 2 and fit using the PROC SURVEYLOGISTIC procedure.

(4) Repeat step 3 for all three weighting schemes.

(5) Compare the analyses for differences and similarities.

(6) Interpret and discuss conclusions.

Chapter 4

Results

4.1 Summary of Data

The discussion presented in this chapter will first provide a summary of the interviewer, voter, and exogenous categories. The logistic regression model-fitting process will then be discussed by describing the model as it proceeds through each of the steps as described in Section 3.5.

4.1.1 Interviewer Characteristics Summary

Interviews were conducted exclusively by undergraduates. Table 4.1.1 summarizes several key interviewer characteristics. This table describes the 542 active volunteer interviewers. However, there were a total of the 666 interviewers who attended the interviewer training. No further investigation was performed on the 124 interviewers who failed to participate as an interviewer. Table 4.1.1 describes the distribution of questionnaires on Election Day and the distribution of interviewers participating in the exit poll. For example, this table shows that even though Brigham Young University students represent 68.85 percent of the 542 in-

terviewers they accounted for only 53.60 percent of the total 7562 questionnaires. The remaining 31.15 percent of active interviewers were from Dixie State College, Southern Utah University, Snow College, Utah State University, Utah Valley State College, Weber State University, and Westminster College. Table 4.1.1 groups the interviewers into two different university categories. Because some of the schools only had a handful of students participating in the 2004 UCEP their sample sizes created small cell sizes resulting in test statistics that may be misleading. Because of this, all succeeding analyses will use two designations: Brigham Young University and Others.

4.1.2 Voter Characteristics Summary

The three voter characteristics collected in 2004 for exit poll voters are summarized in Table 4.2. The UCEP data shows that females represent 52.08 percent of voters. Additionally, the UCEP reports that 92.78 percent of voters responding to the questionnaire are white and the voting age is 35.42 percent for 18-35, 39.89 percent for 36-55, and 24.69 percent for those 56 and older.

4.1.3 Exogenous Characteristics Summary

Exogenous factors provide an additional dimension to the analysis of non-respondents. It provides insight into items not relating to voter or interviewer characteristics. Table 4.3 shows the five exogenous characteristic collected during the 2004 UCEP.

Table 4.1: Demographic characteristics of the exit poll interviewers. Active volunteers administered at least one questionnaire.

Demographic Characteristic	**Percent of Questionnaires Administered** (n=7562)	**Percent of Active Volunteers** (n=542)
School		
Brigham Young University	54.19	68.85
Other Universities	45.81	31.15
Major		
Political/Social Science	51.67	48.62
Physical & Mathematical Sciences/Business	10.18	13.58
Education/Fine Arts	7.90	8.44
Open	13.34	11.19
Other	16.91	18.17
Retail Sales Experience		
Yes	44.47	45.50
No	55.53	54.50
Door to Door Sales		
Yes	14.38	13.83
No	85.62	86.17
Worked as Waiter/Waitress		
Yes	21.55	23.31
No	78.45	76.69
Worked in Telemarketing or Surveys		
Yes	29.02	30.74
No	70.98	69.26
No Previous Employment as Listed Above		
No Employment	67.90	69.93
Employment	32.10	30.07
Did Not Know Sampling Interval at First Call-In		
Did Not Know	32.53	33.11
Did Know	67.47	66.89
Interviewer Training was Helpful		
Not Helpful (Response 1 & 2)	4.51	4.93
3	24.66	25.87
4	46.56	44.35
Extremely Helpful (Response 5)	24.27	24.85

Demographic Characteristic	**Percent of Questionnaires Administered** (n=7562)	**Percent of Active Active Volunteers** (n=542)
Served LDS Mission		
Yes	39.76	37.59
No	60.24	62.41
Marital Status		
Married	12.99	14.05
Unmarried	87.01	85.95
Speaks Spanish		
Yes	8.55	8.20
No	91.46	91.80
Identified as Outgoing		
Shy	41.31	40.36
Outgoing	58.69	59.64
Year in College		
Freshman	27.31	28.67
Sophomore	33.86	31.37
Junior	22.53	22.43
Senior	16.30	17.54
Master a New Task		
Very Slow (Response 1 & 2)	0.00	0.00
3	19.17	20.47
4	59.17	60.40
Very Fast	21.66	19.13
Gender		
Male	53.67	52.92
Female	46.33	47.08
Race		
White	91.08	89.36
Other	8.92	10.64

Table 4.2: Demographic characteristics of the exit poll voters based on both respondents and nonrespondents.

Voter Characteristic	**Percent of Questionnaires Administered** (n=7562)
Age	
18-35	35.42
36-55	39.89
56+	24.69
Race	
Other	7.22
White	92.78
Sex	
Female	52.08
Male	47.92

Table 4.3: Exogenous characteristics involving the polling place

Exogenous Factors	**Percent of Questionnaires Administered** (n=7562)
Time of Day	
Morning	57.38
Afternoon	42.62
Polling Place Location	
Wasatch Front	73.12
Outside Wasatch Front	26.88
Location of Interviewer	
Inside Polling Place	60.46
Outside Polling Place	39.54
Exit Poll Banners	
Yes	11.34
No	88.66
Table	
Yes	54.91
No	45.09

There were 60.46 percent of the **interviews** that occurred when the interviewer was located inside the polling place. Of the 90 polling places, 87.35 percent reported that they had access to a table. However, only 54.91 percent of questionnaires administered had a table at the polling place. The percent of polling places with banners was comparable to the percent of questionnaires that had a banner present during the interview (See Table 4.3 for details).

4.2 Logistic Regression Analysis

This section will focus on the results obtain by using the model-selection process described in Section 3.5. This section will primarily discuss the results in Tables 4.5 and Table 4.7.

4.2.1 Data Organization: Step 1

As described in Step 1 of Section 3.5 half the observations were randomly selected from the 2004 UCEP complete data (n=7562) and assigned to a research group; the other half were assigned to a validation group. Due to random selection the research group had 3711 observations and the validation group had 3851 observations.

4.2.2 Research and Validation Data – PROC LOGISTIC: Step 2

Next, backward elimination, assuming simple random sample theory, was applied to the research and validation data. Table 4.4 shows, with an asterisk (*),

the variables remaining after backward elimination was used on the research and validation datasets for the various weighting schemes. The variables remaining in the model when using unnormalized weights are not included in Table 4.4 because 23 of the 28 variables were declared significant. The variables from these models comprise the preliminary model and are applied to the complete data, accounting for the complex design.

The preliminary model consists of the significant variables produced from the research and validation datasets. This method eliminates variables that are definitely not associated with questionnaire nonresponse and does so by comparing two separate datasets. The variables listed in Table 4.4 are those found to be significant in either model and will be investigated using the complex sample design.

Four variables are significant in both the research and the validation datasets. These four will be included in the final model. The other seven variables that were significant in either the research or validation dataset will be tested again using the complete data, the union of the research and validation datasets, and applying the complex sample design.

4.2.3 Analysis using a Complex Sample Design – PROC SURVEYLOGISTIC: Steps 3 and 4

After reducing the number of variables using the resulting model from Section 4.2.2 a model is then applied to the complete data using these variables.

Table 4.4: Analysis of the research and validation datasets assuming a simple random sample using the three different weighting schemes. Significant variables for each weighting scheme are identified by an asterisk (*).

	Research Data			**Validation Data**		
FACTOR	Pred.	Act.	Unwgt.	Pred.	Act.	Unwgt.
Interviewer Training Helpful				*	*	*
Interviewer Major	*	*				
Interviewer Retail Sales	*	*		*	*	*
Interviewer Speaks Spanish					*	
Interviewer Birthday				*	*	
Interviewer Outgoing			*			
Time of Day	*	*		*	*	*
Table at Polling Place	*	*	*			
Voter Age	*	*	*	*	*	*
Voter Race	*	*	*	*	*	*
Voter Gender	*	*	*			

At this point variables are once again eliminated from the preliminary model based on an $\alpha = .05$ criterion. Interactions are also tested; they could not be tested earlier because SAS could not construct a 28-way interaction on the available data. The independent variables come from three categories: voter factors, interviewer factors, and exogenous factors. For the final model established here significant voter factors are the voter's gender, race, and age. The only significant interviewer factor is previous retail sales work experience and the only significant exogenous factor is the time of day. Initially variables were eliminated from the model due to their lack of significance when working with the research and validation data. No interaction was found to be significant using the $\alpha = .05$ criterion. In this section variables are eliminated when the variables in the complete data using a complex sample design do not meet the $\alpha = .05$ criterion.

Observing Table 4.5 suggests there is minimal difference in the coefficients which, as would be expected, are identical for the weighted models. The primary difference in this table is that the standard errors differ. Another item of note is that the model selection process used the same process to obtain the final model regardless of how the weights were constructed. Table 4.5 shows the sample design, the measure of voter turnout size used (predicted, actual, or unweighted), and whether the weights were normalized or not. The unnormalized simple random sample design (predicted and actual) are excluded from the comparison table because 23 of the 28 variables remained significant. Otherwise, regardless of weighting scheme, the models are very similar.

Table 4.5: Logistic Regression model comparison and standard errors. Weights are calculated using population (actual) voter turnout, predicted voter turnout, and unweighted.

Name	**Level**	**SRS Predicted Normalized (S.E.)**	**SRS Actual Normalized (S.E.)**	**Complex Actual Unnormalized (S.E.)**	**Complex Predicted Normalized (S.E.)**[a]	**Complex Actual Normalized (S.E.)**	**Complex — Unweighted (S.E.)**
Intercept		.9611	1.4413	1.4413	1.3394	1.4413	1.4410
Retail Sales Experience	Yes	0.0842 (0.0542)	0.1530 (0.0527)	0.1530 (0.0641)	0.1241 (0.0744)	0.1530 (0.0641)	0.1468 (0.0528)
Time of Day	PM	0.2720 (0.0535)	0.2325 (0.0535)	0.2325 (0.0608)	0.2708 (0.0576)	0.2325 (0.0608)	0.2383 (0.0549)
Voter Age	35-55	-0.5503 (0.0611)	-0.5179 (0.0612)	-0.5179 (0.0782)	-0.5343 (0.0878)	-0.5179 (0.0782)	-0.4750 (0.0706)
	56+	-0.7166 (0.0697)	-0.7882 (0.0691)	-0.7882 (0.0985)	-0.7042 (0.1074)	-0.7882 (0.0985)	-0.7552 (0.0959)
Voter Gender	Male	-0.2128 (0.0525)	-0.2440 (0.0524)	-0.2440 (0.0565)	-0.2067 (0.0579)	-0.2440 (0.0565)	-0.2230 (0.0485)
Voter Race	White	-0.5889 (0.1110)	-0.5908 (0.1142)	-0.5908 (0.1247)	-0.5739 (0.1262)	-0.5908 (0.1247)	-0.6466 (0.1111)
Interviewer Age	—	0.0222 (0.0095)	-[b] (-)	- (-)	- (-)	- (-)	- (-)
Interviewer Outgoing	Yes	0.1099 (0.0543)	- (-)	- (-)	- (-)	- (-)	- (-)

NOTE: [a] The complex sample design with unnormalized predicted turnout is identical to the normalized model.
[b] These terms are not significant and are therefore not included.

4.2.4 Predicted Probability of a Voter Responding

As shown in Table 4.6 there is a trend in the predicted response rates ranging from 45 to 86 percent. This table shows the probability of a voter accepting the initial invitation to complete the exit poll questionnaire. The table is sorted from lowest probability to highest. The last three columns show the predicted probability of a response using the three weighting schemes: actual, predicted, and unweighted. Regardless of weighting the probabilities are very similar. The predicted probabilities are produced using five variables in the model. These are listed in the first five columns of the table. This table shows that the lowest predicted response rate exists among interviewers with no retail sale experience, interviewing in the morning, with a white, male voter 56 years and older. This table is useful because it is simple to see what happens when one of the variable levels is modified. For example, when interviewers have retail sales experience the voter's predicted response rate, when using unweighted or actual weights, increases by about 2.8 percent $(76.5 - 73.7)$ compared to when interviewers don't have retail sales experience. The highest predicted response rate is achieved when the interviewer has previous retail sales experience and the interview is conducted in the afternoon with a nonwhite, female voter between 18 and 35. The predicted probability of a successful interview is $\hat{\pi}(x)$ and selected probabilities can be found in Table 4.6. The complete list of probabilities can be found in Appendix A.1.

Table 4.6: Selected predicted probabilities of voter response comparing unweighted, actual turnout weighted, and predicted turnout weighted logistic regression for a complex sample design.

Interviewer Retail Sales Sales	**Time of Day Day**	**Voter Race**	**Voter Gender**	**Voter Age**	**Unweighted Response Rate**	**Actual Weighted Response Rate**	**Predicted Weighted Response Rate**
No	AM	White	Male	56+	45.4%	45.5%	46.4%
No	AM	White	Male	36-55	52.4%	52.2%	50.6%
No	AM	White	Male	18-35	63.9%	64.7%	63.6%
No	AM	White	Female	18-35	68.9%	70.1%	68.3%
No	PM	White	Female	18-35	73.7%	74.7%	73.8%
Yes	PM	White	Female	18-35	76.5%	77.5%	76.1%
Yes	PM	Other	Female	18-35	86.1%	86.1%	85.0%

Unfortunately, with an exit poll, the characteristics associated with a higher probability of an initial questionnaire response cannot be consistently obtained. There are certain attributes of an interviewer-voter encounter that can be manipulated to encourage voters to complete an exit poll questionnaire and there are those that should not be manipulated. Voter characteristics and the exogenous characteristics such as the rural/urban balance are determined by the sample design and the probability sampling process; manipulating these variables could lead to a potential bias.

4.2.5 Comparison of Models: Step 5

The weighted models provided in Table 4.7 use the actual voter turnout. The weights in this table are normalized. The unweighted model incorporates a complex design using weights of one for each observation. Even though the coefficients are consistently similar the measure of standard error for the unweighted model is uniformly smaller. The odds ratio here is particularly useful. When the odds ratio equals one the voter response variable is regarded as being independent of the predictor variable. The greater the value of the odds ratio, the greater the probability the 'Level' will result in a successful interview.

Table 4.7 shows the final unweighted and weighted (normalized) logistic regression models, including the associated p-value and odds ratio. The coefficients and the standard errors for these two models can also be found in the model comparison table (Table 4.5). Whether the model is weighted or unweighted the

coefficients have a fairly straightforward interpretation. The sign of the coefficient determines whether $\pi(x)$ is decreasing or increasing as the predictor variable increases. All independent variables are categorical and, excluding voter age, are either 0 or 1. The level coded as 1 is listed in Table 4.7 as the 'Level'.

Table 4.7: Unweighted and actual weighted logistic regression model showing coefficients, standard errors, p-values, and odds ratios for a complex sample design.

Name	**Level**	**Unweighted Coefficient (S.E.)**	**Weighted[a] Coefficient (S.E.)**	**P-Value Unweighted**	**P-Value Weighted**	**Odds Ratio Unweighted**	**Odds Ratio Weighted**
Retail Sales Experience	Yes	0.1468 (0.0528)	0.1530 (0.0641)	.0054	.0170	1.158	1.165
Time of Day	PM	0.2383 (0.0549)	0.2325 (0.0608)	$< .0001$	$< .0001$	1.269	1.262
Voter Age	36-55	-0.4750 (0.0706)	-0.5179 (0.0782)	$< .0001$	$< .0001$	0.622	0.596
	55+	-0.7552 (0.0959)	-0.7882 (0.0985)	$< .0001$	$< .0001$	0.470	0.455
Voter Gender	Male	-0.2230 (0.0485)	-0.2440 (0.0565)	$< .0001$	$< .0001$	0.800	0.783
Voter Race	White	-0.6466 (0.1111)	-0.5908 (0.1247)	$< .0001$	$< .0001$	0.524	0.554
NOTE: [a] for the complex sample design the weighted and unweighted models are identical.							

Chapter 5

Discussion and Summary

5.1 Interpretation

Five variables are found to have a reasonable degree of association with voter response as discussed in Chapter 4. These are voter age, voter race, voter gender, interviewer's previous retail sales experience, and the time of day. The reason for the association could vary widely and it would be necessary to conduct a formal experiment in order to determine a causal relationship for nonresponse in an exit poll. However, intuition seems to suggest a viable theory for many of these factors.

First, when the interviewers have retail sales experience the voter is more likely to respond. This seems intuitive because skills needed to sell products were used to sell the questionnaire.

Second, morning voters are less likely to respond to the exit poll questionnaire than afternoon voters. Morning voters may be in a hurry to get back to work or have other appointments whereas afternoon voters may have more time and do not have pressing engagements to attend.

Third, it is not surprising that voter characteristics are strong predictors of voter nonresponse. Younger voters, female voters, and nonwhite voters are all more likely to respond to an exit poll questionnaire. However, this discussion will not attempt to explain why this is the case.

5.2 Recommended Changes

Based on research presented here, there is primarily one way for us to reduce nonresponse, and that is through better training and/or better administration of the interviewing process. All other ways are simply infeasible due to the election process and the potential to introduce bias; an exit poll simply cannot be limited to women, younger voters, or nonwhites.

It has been shown here that interviewers with retail sales experience tend to achieve a higher response rate. If interviewers are given better pre-election training that teaches them how to persuade the recipient by 'selling' the questionnaire and the organization they represent, the voter should be more likely to respond to a questionnaire. Other administrative strategies that can be implemented are:

(1) Have interviewers with retail sales experience present the questionnaires to the voter while those without report the data to the call center.

(2) Spread interviewers with retail sales experience across more polling places rather than clustering them together at the same polling place.

Even if the voter characteristic bias is negligible it would not be appropriate to

limit the sample to any particular demographic variable simply to increase the response rate.

Listed in Table 5.1 are all the variables in the final weighted logistic regression model. Manipulating voter age, race, and gender as well as the time of day the interview took place are likely to result in biased estimates. However, using more interviewers with retail sales experience to conduct the interview with voters seems unlikely to result in a bias.

Table 5.1: List of variables they may be manipulated

Variables That May Be Manipulated	**Variables that Should Not Be Manipulated**
Interviewer Retail Sales Experience	Time of Day Voter Age Voter Race Voter Gender

5.3 Future Research

An underlying assumption not discussed here is that nonresponse bias is an issue of concern. The historical accuracy of the UCEP supports the position that nonresponse bias is not a concern. However, a more comprehensive study ought to be conducted that may include time of day and voter age, race, and gender characteristics. Research on the topic of exit poll bias has been conducted nationally by Merkle and Edelman (2002, pp. 243–257).

Up to date studies on time of day and interviewer characteristics should also be conducted. This will allow further confirmation of the results provided in

this discussion. Furthermore, it will establish the ability to measure interviewer characteristics and trends from one exit poll to the next.

Given available funds, future exit polls could include small incentives such as pens or pencils. Any future research on time of day and interviewer characteristics should be well planned to eliminate any bias. Other research using logistic regression may include an analysis of a year effect. A year effect using interviewer characteristics can be calculated after the 2006 exit poll because these characteristics were first collected in 2004. Additionally, real-time analysis could be conducted on Election Day to help assist administrators and roving teams identify polling places and interviewers with unexpectedly low or high response rates.

Appendix A

Complete List of Predicted Probabilities

Table A.1: Complete list of predicted probabilities of voter response comparing unweighted, actual turnout weighted, and predicted turnout weighted logistic regression for a complex sample design.

Interviewer Retail Sales Sales	**Time of Day Day**	**Voter Race**	**Voter Gender**	**Voter Age**	**Unweighted Response Rate**	**Actual Weighted Response Rate**	**Predicted Weighted Response Rate**
No	Morning	White	Male	56+	45.4%	45.5%	46.4%
Yes	Morning	White	Male	56+	49.1%	49.3%	49.5%
No	Morning	White	Male	36-55	51%	51.3%	50.6%
No	Morning	White	Female	56+	51.4%	51.6%	51.5%
No	Afternoon	White	Male	56+	52.4%	52.2%	53.1%
Yes	Morning	White	Male	36-55	54.6%	55.1%	53.7%
Yes	Morning	White	Female	56+	55%	55.4%	54.6%
No	Morning	White	Female	36-55	56%	56%	55.8%
Yes	Afternoon	White	Male	56+	56.9%	57.3%	56.2%
No	Afternoon	White	Male	36-55	57.9%	58%	57.3%
No	Afternoon	White	Female	56+	58.3%	58.2%	58.2%
Yes	Morning	White	Female	36-55	60.5%	60.1%	58.8%
Yes	Afternoon	White	Male	36-55	61.4%	61%	60.3%
No	Morning	Other	Male	56+	61.4%	61.6%	60.6%
Yes	Afternoon	White	Female	56+	61.8%	61.9%	61.2%
No	Afternoon	White	Female	36-55	63.6%	63.7%	62.3%
Yes	Morning	Other	Male	56+	63.9%	63.8%	63.5%
No	Morning	White	Male	18-35	64.8%	64.7%	63.6%
No	Morning	Other	Male	36-55	66.5%	65.5%	64.5%
Yes	Afternoon	White	Female	36-55	66.8%	65.8%	65.2%

Interviewer Retail Sales Sales	Time of Day Day	Voter Race	Voter Gender	Voter Age	Unweighted Response Rate	Actual Weighted Response Rate	Predicted Weighted Response Rate
No	Morning	Other	Female	56+	66.9%	66.4%	65.4%
Yes	Morning	White	Male	18-35	67.2%	67.2%	66.4%
No	Afternoon	Other	Male	56+	67.8%	68.1%	66.8%
Yes	Morning	Other	Male	36-55	68.9%	68.9%	67.3%
Yes	Morning	Other	Female	56+	69.2%	69.1%	68.1%
No	Morning	White	Female	18-35	69.7%	69.7%	68.3%
No	Morning	Other	Female	36-55	70%	69.8%	69.1%
Yes	Afternoon	Other	Male	56+	70.9%	70.1%	69.5%
No	Afternoon	White	Male	18-35	71.6%	70.8%	69.6%
No	Afternoon	Other	Male	36-55	71.9%	71.3%	70.5%
Yes	Morning	White	Female	18-35	72.2%	71.6%	70.9%
No	Afternoon	Other	Female	56+	72.4%	72.9%	71.2%
Yes	Morning	Other	Female	36-55	72.7%	73.2%	71.7%
Yes	Afternoon	White	Male	18-35	73.7%	73.9%	72.2%
Yes	Afternoon	Other	Male	36-55	74.5%	74.4%	73%
Yes	Afternoon	Other	Female	56+	75.3%	74.6%	73.7%
No	Afternoon	White	Female	18-35	75.5%	74.7%	73.8%
No	Afternoon	Other	Female	36-55	76.5%	76.1%	74.6%
No	Morning	Other	Male	18-35	76.9%	76.8%	75.6%
Yes	Afternoon	White	Female	18-35	77.2%	77.5%	76.1%

Interviewer Retail Sales Sales	Time of Day Day	Voter Race	Voter Gender	Voter Age	Unweighted Response Rate	Actual Weighted Response Rate	Predicted Weighted Response Rate
Yes	Afternoon	Other	Female	36-55	79.4%	78.7%	76.9%
Yes	Morning	Other	Male	18-35	79.7%	79.4%	77.8%
No	Morning	Other	Female	18-35	80.9%	80.7%	79.2%
No	Afternoon	Other	Male	18-35	81.1%	80.9%	80.3%
Yes	Morning	Other	Female	18-35	83%	83%	81.2%
Yes	Afternoon	Other	Male	18-35	83.2%	83.1%	82.2%
No	Afternoon	Other	Female	18-35	84.3%	84.2%	83.3%
Yes	Afternoon	Other	Female	18-35	86.1%	86.1%	85%

Appendix B

Questionnaires

B.1 Interviewer Questionnaire

This is for internal use only; your responses will remain completely confidential. We will use the information you provide on this questionnaire to help us better understand the exit polling process. Before performing any analysis of the information provided on the questionnaire, we will remove your name and any identifying information.

A. Are you:
- ❑ 1. Male
- ❑ 2. Female

B. Are you:
- ❑ 1. Native American/Indian
- ❑ 2. Asian
- ❑ 3. African American/Black
- ❑ 4. Hispanic/Latino
- ❑ 5. White/Caucasian
- ❑ 6. Pacific Islander
- ❑ 7. Other

C. What year were you born?

19_____

D. Are you presently:
- ❑ 1. Married
- ❑ 2. Single
- ❑ 3. Divorced
- ❑ 4. Engaged
- ❑ 5. Cohabiting with boyfriend/girlfriend
- ❑ 6. Widowed

E. What is your year in school?
- ❑ 1. Freshman
- ❑ 2. Sophomore
- ❑ 3. Junior
- ❑ 4. Senior
- ❑ 5. Graduate Student
- ❑ 6. Other
- ❑ 7. Don't Know

F. What state are you from?

_____

G. Have you served a full time LDS mission?
- ❑ 1. Yes
- ❑ 2. No/ Does not apply

H. What is your major?
- ❑ 1. Political Science
- ❑ 2. Social Science
- ❑ 3. Fine Arts/Humanities
- ❑ 4. Mathematics/Physics
- ❑ 5. Biology/Chemistry
- ❑ 6. Education
- ❑ 7. Business
- ❑ 8. Open
- ❑ 9. Other

I. How tall are you?

Feet _____ Inches_____

J. Are you in Political Science classes for (Check all that apply.):
- ❑ 1. Major Requirement
- ❑ 2. Minor Requirement
- ❑ 3. GE Requirement
- ❑ 4. Interest
- ❑ 5. Fun
- ❑ 6. Other

K. On a scale of 1 to 5, with 1 being very shy and 5 being very outgoing, where would you place yourself?
- ❑ 1. Very shy
- ❑ 2.
- ❑ 3.
- ❑ 4.
- ❑ 5. Very outgoing

L. Have you ever worked in any of the following jobs, check all that apply:
- ❑ 1. Retail Sales
- ❑ 2. Door to Door Sales
- ❑ 3. Waiter/Waitress
- ❑ 4. Telemarketing/Surveys
- ❑ 5. None

M. In a school or work setting, how long does it usually take you to master a new task, with 1 being very slow and 5 very fast?
- ❑ 1. Very slow
- ❑ 2.
- ❑ 3.
- ❑ 4.
- ❑ 5. Very fast

N. On a scale of 1 to 5 how helpful was the interviewer training session with 1 being not very helpful and 5 being very helpful?
- ❑ 1. Not at all helpful
- ❑ 2.
- ❑ 3.
- ❑ 4.
- ❑ 5. Extremely helpful

O. What is your primary reason for participating in the Utah Colleges Exit Poll?
- ❑ 1. My professor offered extra credit
- ❑ 2. To fulfill a course requirement
- ❑ 3. Interest in public opinion and polling
- ❑ 4. My friends are participating
- ❑ 5. A general interest in politics
- ❑ 6. Another reason not listed here

B.2 Polling Place Questionnaire From First Call

Questions for First Call from Each Polling Place

What is your polling place ID? _____ (located in your *packet instructions* or it's the 1st 3 digits of form ID in *Official Use Only* box)

What is your sampling interval? _____(located in your *packet instructions*)

What time did you arrive at your polling location? _____ a.m.

Is the election judge allowing you to conduct the survey? YES NO

Did you encounter any problems from the election judge in getting permission to conduct your survey?

Are there any other organizations also conducting surveys? YES NO

Are you located inside the polling place? YES NO

Do you have access to a table? YES NO

How long did it take you to reach this phone you are using? _____ minutes.

Is the phone you are using a pay phone? YES NO

If yes, is there another phone you could use? YES NO

Do you have any other difficulties?

********If you are calling from a cell phone or within Utah county, please call Data Entry at 801-378-4811 and Crisis at 801-378-6404**

B.3 Exit Poll Questionnaire

2004 KBYU-Utah Colleges Exit Poll **ORIGINAL** **BLUE**

[A] In today's election for U.S. president, did you just vote for:
- ❑ 1. George W. Bush / Dick Cheney, Republican
- ❑ 2. John F. Kerry / John Edwards, Democratic
- ❑ 3. Someone else
- ❑ 4. Don't remember / Did not vote for president

[B] In today's election for U.S. senator, did you just vote for:
- ❑ 1. Robert F. Bennett, Republican
- ❑ 2. R. Paul Van Dam, Democratic
- ❑ 3. Someone else
- ❑ 4. Don't remember / Did not vote for U.S. senator

[C] In today's election for U.S. House of Representatives, did you just vote for:
- ❑ 1. Congressman 1, Republican
- ❑ 2. Congressman 2, Democratic
- ❑ 3. Someone else
- ❑ 4. Don't remember / Did not vote for U.S. House

[D] In today's election for governor, did you just vote for:
- ❑ 1. Jon M. Huntsman, Jr. / Gary R. Herbert, Republican
- ❑ 2. Scott M. Matheson, Jr. / Karen Hale, Democratic
- ❑ 3. Someone else
- ❑ 4. Don't remember / Did not vote for governor

[E] In today's election for Utah attorney general, did you just vote for:
- ❑ 1. Mark Shurtleff, Republican
- ❑ 2. Gregory G. Skordas, Democratic
- ❑ 3. Someone else
- ❑ 4. Don't remember / Did not vote for attorney general

[F] <u>This question for Salt Lake County voters only</u>. In today's election for Salt Lake County mayor, did you just vote for:
- ❑ 1. Peter M. Corroon, Democratic
- ❑ 2. Ellis Ivory, Republican
- ❑ 3. Merrill Cook, unaffiliated
- ❑ 4. Someone else
- ❑ 5. Don't remember / Did not vote for county mayor

[G] In today's election, how did you just vote on Constitutional Amendment Number 3, regarding the legal definition of marriage?
- ❑ 1. For
- ❑ 2. Against
- ❑ 3. Don't remember / Did not vote on the amendment

[H] In today's election, how did you just vote on Initiative 1, the Open Spaces Initiative?
- ❑ 1. For
- ❑ 2. Against
- ❑ 3. Don't remember / Did not vote on the initiative

[I] Do you approve or disapprove of the job that George W. Bush is doing as president?

❑ 1. Approve ❑ 2. Disapprove ❑ 3. Don't know

[J] Do you approve or disapprove of the job that Orrin Hatch is doing as senator?

❑ 1. Approve ❑ 2. Disapprove ❑ 3. Don't know

[K] Do you approve or disapprove of the job that Robert Bennett is doing as senator?

❑ 1. Approve ❑ 2. Disapprove ❑ 3. Don't know

[L] Do you think that statewide ballot initiative elections:
- ❑ 1. Are a good thing for Utah
- ❑ 2. Are a bad thing for Utah
- ❑ 3. Don't make much difference
- ❑ 4. Don't know / No opinion

[M] Some people say that the current rules make it <u>too hard</u> for Utah citizens to put an initiative on the ballot. Some say that the current rules make it <u>too easy</u>. Others think the current rules are <u>about right</u>. Which comes closest to your view?
- ❑ 1. Too hard
- ❑ 2. Too easy
- ❑ 3. About right
- ❑ 4. Don't know / No opinion

[N] Do you agree or disagree with the following statements about Initiative 1, the Open Spaces Initiative?

	Strongly Disagree				Strongly Agree
a. The initiative will protect the environment	1	2	3	4	5
b. The initiative will hurt landowners	1	2	3	4	5
c. The initiative will help Utah's economy	1	2	3	4	5
d. The initiative will raise taxes	1	2	3	4	5
e. The initiative is too complicated or confusing	1	2	3	4	5

[O] Compared to four years ago, is your family's economic situation today:
- ❑ 1. Better
- ❑ 2. Worse
- ❑ 3. About the same
- ❑ 4. Don't know

[P] In general, how interested are you in politics?
- ❑ 1. Very interested
- ❑ 2. Interested
- ❑ 3. Not very interested
- ❑ 4. Not interested
- ❑ 5. Don't know

[Q] Which statement comes closest to your view?
- ❑ 1. By law, abortion should never be permitted.
- ❑ 2. The law should permit abortion only in case of rape, incest, or when the woman's life is in danger.
- ❑ 3. The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established.
- ❑ 4. By law, a woman should always be able to obtain an abortion as a matter of personal choice.

[R] Do you own a gun?

❑ 1. Yes ❑ 2. No

Please continue on other side →

2004 KBYU-Utah Colleges Exit Poll ORIGINAL BLUE

[S] Some people say the Patriot Act is a necessary tool in preventing terrorist attacks, while others say the act violates the civil liberties of average Americans. Which comes closest to your view?
- ❑ 1. It is a necessary tool
- ❑ 2. It violates civil liberties
- ❑ 3. Don't know / No opinion

[T] Thinking back on Mike Leavitt's service as governor, what grade would you give him? Circle one.

A A- B+ B B- C+ C C- D+ D D- F

[U] On a scale of 1 to 5, with 1 being not very influential and 5 being very influential, how influential was each of these sources in helping you decide how to vote? Circle one number per line.

	Not very influential				Very influential
a. Daily newspapers	1	2	3	4	5
b. Television	1	2	3	4	5
c. Radio	1	2	3	4	5
d. Internet	1	2	3	4	5
e. Friends, family, or co-workers	1	2	3	4	5

[V] Do you favor or oppose tax credits for parents who send their children to private schools?
- ❑ 1. Favor
- ❑ 2. Oppose
- ❑ 3. Don't know / No opinion

[W] Do you consider the area in which you live to be:

❑ 1. Urban ❑ 2. Suburban ❑ 3. Rural

[X] How often do you use the Internet?
- ❑ 1. Once or more a day
- ❑ 2. A few times a week
- ❑ 3. A few times a month
- ❑ 4. Hardly ever
- ❑ 5. I do not have access to the Internet

[Y] What year were you born? 19 ___ ___

[Z] Are you:

❑ 1. Male ❑ 2. Female

[AA] Generally speaking, do you consider yourself to be a(n):
- ❑ 1. Strong Democrat
- ❑ 2. Not so strong Democrat
- ❑ 3. Independent leaning Democrat
- ❑ 4. Independent
- ❑ 5. Independent leaning Republican
- ❑ 6. Not so strong Republican
- ❑ 7. Strong Republican
- ❑ 8. Other
- ❑ 9. Don't know

[BB] On most political matters do you consider yourself:
- ❑ 1. Strongly conservative
- ❑ 2. Moderately conservative
- ❑ 3. Neither, middle of the road
- ❑ 4. Moderately liberal
- ❑ 5. Strongly liberal
- ❑ 6. Don't know

[CC] What was the last year of school you completed?
- ❑ 1. Eighth grade or less
- ❑ 2. Some high school
- ❑ 3. High school graduate
- ❑ 4. Some college
- ❑ 5. College graduate
- ❑ 6. Post-graduate

[DD] What, if any, is your religious preference?
- ❑ 1. Protestant
- ❑ 2. Catholic
- ❑ 3. LDS / Mormon
- ❑ 4. Jewish
- ❑ 5. Other
- ❑ 6. No preference / No religious affiliation
- ❑ 7. Prefer not to say

[EE] Do you consider yourself to be a Born-Again or Evangelical Christian?

❑ 1. Yes ❑ 2. No

[FF] How active do you consider yourself in the practice of your religious preference?
- ❑ 1. Very active
- ❑ 2. Somewhat active
- ❑ 3. Not very active
- ❑ 4. Not active
- ❑ 5. Does not apply / Prefer not to say

[GG] What is your current employment status?
- ❑ 1. Self-employed
- ❑ 2. Employed by someone else
- ❑ 3. Unemployed
- ❑ 4. Home-maker
- ❑ 5. Retired
- ❑ 6. Student

[HH] Are you:
- ❑ 1. Native American
- ❑ 2. Asian
- ❑ 3. Black / African American
- ❑ 4. Hispanic / Latino
- ❑ 5. White / Caucasian
- ❑ 6. Pacific Islander
- ❑ 7. Other

[II] Are you presently:
- ❑ 1. Married
- ❑ 2. Divorced
- ❑ 3. Widowed
- ❑ 4. Single

[JJ] What do you expect your 2004 family income to be?
- ❑ 1. Under $25,000
- ❑ 2. $25,000 - $39,999
- ❑ 3. $40,000 - $49,999
- ❑ 4. $50,000 - $74,999
- ❑ 5. $75,000 - $99,999
- ❑ 6. Over $100,000

Thank you for your voluntary participation in this research.

Official Use Only

Form ID: **Interviewer ID** _____

Time (nearest 15 min.) _____ AM PM ❑ Called In

Continue only if Non-Response

Age of Non-respondent: ❑ 18-35 ❑ 36-55 ❑ 56-75 ❑ 75+

Gender: ❑ M ❑ F **Race:** ❑ White ❑ Hispanic ❑ Other

Appendix C

SAS Code

```
options ls=120 FORMDLIM="*";

filename intdata "h:\Masters Project 2004 Analysis\interviewer_information.csv";
filename design "h:\Masters Project 2004 Analysis\design_and_predictions.csv";
filename ppi "h:\Masters Project 2004 Analysis\aux_info.csv";
filename actual "h:\Masters Project 2004 Analysis\actual_overall_smry.csv";
libname raw spss "h:\Masters Project 2004 Analysis\raw_data.por";
libname exportme "h:\Masters Project 2004 Analysis\";

/*************************************************
Read in each of the datasets;
*************************************************/
data intdata;
  infile intdata firstobs=2 dsd;
  input intvwrid name$ spanish school$ drive intsex intrace bday marriage grade state$ mission
      major height reason1 reason2 reason3 reason4 reason5 reason6 outg jobs1 jobs2 jobs3
      jobs4 jobs5 learning helpful particip;
  if school in ('Dixie','SUU','Snow','USU','UVSC','Weber','Westmins') then intbinaryschool = 0;
  else if school='BYU' then intbinaryschool=1;
  else intbinaryschool=.;

  if major in (1,2) then major5=1;
  else if major in (4,5,7) then major5=2;
  else if major in (3,6) then major5=3;
  else if major =8 then major5=4;
  else if major =9 then major5=5;
  else major=.;

  if intrace=5 then intracetmp=1;
  else intracetmp=0;

run;
data ppi;
  infile ppi firstobs=2 dsd;
  input ppid TABLE BANNER SampInt TimeArrive$ JudgeAllow OtherProb OtherOrg$ LocatedInside selftable
Time2Phone Phone OtherPhone$;
run;
data raw;
  set raw.raw_data;
  ppid = placeid;
  if ppid=165 or ppid=353 or ppid=582 or ppid=721 then delete;
run;
data design;
infile design firstobs=2 dsd missover;
input COUNTY_NM$ DISTRICT STRATUM COUNTY_NUM PPID PRED_VOTE PRED_NONRESPONSE_RATE INTERVAL wgt1
wgt2;
drop wgt2;
run;
data actual;
  infile actual firstobs=2 dsd missover;
  input YEAR UID DISTRICT_CD COUNTY_NM$ COUNTY_CD PRECINCT_ID$ SAMPLED STRATUM PPID PP_NM$ ADDRESS$
CITY$ ZIP ABSENTEE PRES_REP PRES_DEM PRES_OTH SEN_REP SEN_DEM SEN_OTH H1_REP H1_DEM H1_OTH H2_REP
H2_DEM H2_OTH H3_REP H3_DEM H3_OTH GOV_REP GOV_DEM GOV_OTH AG_REP AG_DEM AG_OTH P3_FOR P3_AGNST
I1_FOR I1_AGNST SL_CNTY_REP_MAYOR SL_CNTY_DEM_MAYOR SL_CNTY_OTH_MAYOR;
  x=1;
  run;



/****** Rolls up the actual results data into the polling place level ********/
proc sql;
/************************************
                    Original Estimates
************************************/
      CREATE TABLE og_cnts AS
            SELECT count(presvote) AS og_pres_cnts, count(senvote) AS og_sen_cnts, count(govvote) AS
og_gov_cnts, ppid FROM raw GROUP BY ppid;

/*************************************************
```

```
         New Estimates with Fused Counties & New Weights
*****************************************************/
      CREATE TABLE actual_d AS
            SELECT ppid, sum(pres_rep) AS pres_rep, sum(pres_dem) as pres_dem, sum(pres_oth) as
pres_oth,
                  sum(sen_rep) as sen_rep, sum(sen_dem) as sen_dem, sum(sen_oth) as sen_oth,
                  sum(gov_rep) as gov_rep, sum(gov_dem) as gov_dem, sum(gov_oth) as gov_oth
            FROM actual
            WHERE ppid ne .
            GROUP BY ppid;

/*****************************************************
Exit Poll Results at the Polling Place Level
*****************************************************/
      CREATE TABLE pp_xp_smry AS
            SELECT raw.ppid, count(raw.response) AS pp_response_cnts, count(raw.PRESVOTE) AS
pp_pres_voter_cnts, count(raw.SENVOTE) AS pp_sen_voter_cnts, count(raw.GOVVOTE) AS
pp_gov_voter_cnts,
                  count(raw.PRESVOTE)/count(raw.response) AS pp_pres_rate,
count(raw.SENVOTE)/count(raw.response) AS pp_sen_rate, count(raw.GOVVOTE)/count(raw.response) AS
pp_gov_rate
            FROM raw
            JOIN design ON design.ppid = raw.ppid
            GROUP BY design.stratum, raw.ppid;

/*****************************************************
Exit Poll Results at the Strata Level
*****************************************************/
      CREATE TABLE str_xp_smry AS
            SELECT design.stratum, COUNT(raw.PRESVOTE) AS str_pres_voter_cnts, COUNT(raw.SENVOTE) AS
str_sen_voter_cnts,
                  COUNT(distinct raw.ppid) AS cnt_pp
            FROM raw
            JOIN design ON design.ppid = raw.ppid
            GROUP BY design.stratum;

/*****************************************************
Interviewer Counts & INFORMATION
*****************************************************/
      CREATE TABLE int_xp_smry_responses AS
            SELECT count(raw.ppid) AS respond_IntCnts, raw.intvwrid
            FROM raw
            JOIN intdata ON intdata.intvwrid=raw.intvwrid
            WHERE raw.response=1
            GROUP BY raw.intvwrid;
      CREATE TABLE int_xp_smry_all AS
            SELECT count(raw.ppid) AS all_IntCnts, raw.intvwrid
            FROM raw
            JOIN intdata ON intdata.intvwrid=raw.intvwrid
            GROUP BY raw.intvwrid;
            ;
      CREATE TABLE int_xp_smry AS
            SELECT CASE WHEN int_xp_smry_responses.respond_IntCnts=. THEN 0 ELSE
int_xp_smry_responses.respond_IntCnts END AS respond_IntCnts,
                  CASE WHEN int_xp_smry_responses.respond_IntCnts=. THEN 0 ELSE
int_xp_smry_responses.respond_IntCnts END/int_xp_smry_all.all_IntCnts AS int_RespRate,
                  int_xp_smry_all.all_IntCnts, int_xp_smry_all.intvwrid
            FROM int_xp_smry_all
            LEFT JOIN int_xp_smry_responses ON
int_xp_smry_responses.intvwrid=int_xp_smry_all.intvwrid
            ;

/*****************************************************
Actual Strata Results
*****************************************************/
      CREATE TABLE str_actual_smry AS
            SELECT actual.stratum, count(actual.ppid) AS str_precinct_cnts,
                  sum(actual.PRES_DEM+actual.PRES_REP+actual.PRES_OTH) AS str_pres_tot,
                  count(distinct actual.ppid) AS str_pp_cnts
```

```
            FROM actual
            GROUP BY actual.stratum;
/*****************************************************
Actual Polling Place Results
*****************************************************/
      CREATE TABLE pp_actual_smry AS
            SELECT ppid, stratum, county_nm, county_cd, sum(PRES_DEM) AS PRES_DEM, sum(PRES_REP) AS
PRES_REP, sum(PRES_OTH) AS PRES_OTH,
                  sum(SEN_DEM) AS SEN_DEM, sum(SEN_REP) AS SEN_REP, sum(SEN_OTH) AS SEN_OTH,
                  sum(GOV_DEM) AS GOV_DEM, sum(GOV_REP) AS GOV_REP, sum(GOV_OTH) AS GOV_OTH
            FROM actual GROUP BY actual.stratum, county_nm, county_cd, actual.ppid
            ORDER BY ppid;

            CREATE TABLE str_pp_sampled AS SELECT stratum, count(distinct ppid) AS
str_pp_sampled_cnt FROM actual where sampled=1 GROUP BY stratum;

/*****************************************************
CREATE THE FINAL DATASET USED FOR ANALYSIS
*****************************************************/
            CREATE TABLE final AS
            SELECT  1 AS ones, design.district AS CD, design.county_num,
                  pp_actual_smry.stratum,
                   str_actual_smry.str_pres_tot, str_actual_smry.str_pp_cnts,
                  raw.ppid, raw.formid,
                  pp_actual_smry.PRES_DEM AS pp_pres_dem, pp_actual_smry.PRES_REP AS pp_pres_rep,
pp_actual_smry.PRES_OTH AS pp_pres_oth,
                  pp_actual_smry.PRES_DEM+pp_actual_smry.PRES_REP+pp_actual_smry.PRES_OTH AS
pp_pres_tot,

      pp_actual_smry.PRES_DEM/(pp_actual_smry.PRES_DEM+pp_actual_smry.PRES_REP+pp_actual_smry.PRES_O
TH) AS pp_pres_dem_perc,

      pp_actual_smry.PRES_REP/(pp_actual_smry.PRES_DEM+pp_actual_smry.PRES_REP+pp_actual_smry.PRES_O
TH) AS pp_pres_rep_perc,
                  str_pp_sampled.str_pp_sampled_cnt,
                  pp_xp_smry.pp_pres_voter_cnts, pp_xp_smry.pp_sen_voter_cnts,
pp_xp_smry.pp_gov_voter_cnts,

/*************************************************************
                        Calculate Weights Base On Voter Population
*************************************************************/

      str_actual_smry.str_pres_tot/(pp_xp_smry.pp_response_cnts*str_pp_sampled.str_pp_sampled_cnt)
AS weight_nr_forced,
/*************************************************************
                        Calculate Weights Base On Predicted Turnout
*************************************************************/
                  CASE WHEN raw.response=1 THEN design.wgt1/pp_xp_smry.pp_pres_voter_cnts ELSE . END
AS weight_pres,
                  CASE WHEN raw.response=1 THEN design.wgt1/pp_xp_smry.pp_sen_voter_cnts  ELSE . END
AS weight_sen,
                  CASE WHEN raw.response=1 THEN design.wgt1/pp_xp_smry.pp_gov_voter_cnts  ELSE . END
AS weight_gov,
                  design.wgt1/pp_xp_smry.pp_response_cnts AS weight_nr,

/************************************
                     Voter Information
************************************/
                  raw.presvote, raw.senvote, raw.govvote, raw.rep1vote, raw.rep2vote, raw.rep3vote,
raw.agvote,
                  CASE WHEN raw.response=1 THEN (CASE WHEN raw.race=. THEN . WHEN raw.race=5 THEN 1
ELSE 0 END) WHEN raw.response=2 THEN (CASE WHEN raw.nrrace=. THEN . WHEN raw.nrrace=1 THEN 1 ELSE 0
END) ELSE . END AS voterrace,
                  CASE WHEN raw.response=1 THEN (CASE WHEN raw.sex=. THEN . WHEN raw.sex=1 THEN 1
ELSE 0 END) WHEN raw.response=2 THEN (CASE WHEN raw.nrsex=. THEN . WHEN raw.nrsex=1 THEN 1 ELSE 0
END) ELSE . END AS votersex,
                  CASE WHEN raw.response=1 THEN (CASE WHEN 2004-raw.yearborn BETWEEN 18 AND 35 THEN
1 WHEN 2004-raw.yearborn BETWEEN 36 AND 55 THEN 2 WHEN 2004-raw.yearborn >= 56 THEN 3 ELSE . END)
```

```
WHEN raw.response=2 THEN (CASE WHEN raw.nrage IN (3,4) THEN 3 ELSE raw.nrage END) ELSE . END AS
voterage,

/************************************
    Polling Place/Exogenous Information
************************************/
                CASE WHEN design.stratum IN (2,3,4,9,12,14) THEN 1 WHEN design.stratum IN
(1,5,6,7,8,10,11,13,15) THEN 0 ELSE . END AS wasatch,
                CASE WHEN ppi.TABLE=1 OR ppi.selftable=1 THEN 1 WHEN (ppi.TABLE=. AND
ppi.selftable=.) THEN . ELSE 0 END AS PP_TABLE, ppi.BANNER, ppi.SampInt, ppi.LocatedInside,
                design.interval,
                CASE WHEN raw.timeampm=1 THEN 1 WHEN raw.timeampm=2 THEN 0 ELSE . END AS
MorningAfternoon,

/************************************
               Interviewer Information
************************************/
                intdata.intvwrid, intdata.jobs1 AS intretailsales, intdata.jobs2 AS intdoortodoor,
                CASE WHEN intdata.jobs3 IN (0,1) THEN intdata.jobs3 ELSE . END AS intwaiter,
                CASE WHEN intdata.jobs4 IN (0,1) THEN intdata.jobs4 ELSE . END AS inttelesurvey,
                CASE WHEN intdata.jobs5 IN (0,1) THEN intdata.jobs5 ELSE . END AS intnojobs,
                CASE WHEN intdata.learning = 0 THEN . WHEN intdata.learning IN (1,2,3) THEN 3 WHEN
intdata.learning IN (4,5) THEN intdata.learning ELSE . END AS intlearning,
                CASE WHEN intdata.intsex = 2 THEN 0 WHEN intdata.intsex=1 THEN 1 ELSE . END AS
intsex,
                CASE WHEN intdata.intrace = 5 THEN 1 WHEN intdata.intrace IN (1,2,3,4,6,7) THEN 0
ELSE . END AS intrace,
                CASE WHEN intdata.mission = 0 THEN . ELSE intdata.mission END AS intmission,
intdata.spanish,
                intdata.school AS intschool,
                /* CASE WHEN intdata.school NOT IN
('BYU','Dixie','SUU','Snow','USU','UVSC','Weber','Westmins') THEN . WHEN intdata.school='BYU' THEN
'BYU' ELSE intdata.school END AS school, */
                intdata.intbinaryschool,
                CASE WHEN intdata.bday >=2004 THEN . ELSE 2004-intdata.bday END AS bday,
                CASE WHEN intdata.marriage=. THEN . WHEN intdata.marriage =1 THEN 1 ELSE 0 END AS
intmarriage,
                CASE WHEN intdata.grade IN (1,2,3,4,5) THEN intdata.grade WHEN intdata.grade IN
(6,7) THEN . ELSE . END AS intyearinschool,
                intdata.height AS intheight,
                CASE WHEN intdata.helpful=0 THEN . WHEN intdata.helpful IN (1,2) THEN 1 ELSE
intdata.helpful END AS inthelpful,
                CASE WHEN intdata.outg IN (1,2,3) THEN 0 WHEN intdata.outg IN (4,5) THEN 1 ELSE .
END AS intoutg,
                CASE WHEN intdata.major IN (1,2) THEN 1 WHEN intdata.major IN (4,5,7) THEN 2 WHEN
intdata.major IN (3,6) THEN 3 WHEN intdata.major = 8 THEN 4 WHEN intdata.major=9 THEN 5 ELSE . END
AS intmajor, intdata.major,
                int_xp_smry.respond_IntCnts, int_xp_smry.all_IntCnts, int_xp_smry.int_RespRate,
                CASE WHEN ppi.sampint ^= design.interval AND ppi.sampint ^= . THEN 1 ELSE 0 END AS
IntDontKnowInterval,

                design.pred_vote, design.pred_nonresponse_rate,
                CASE WHEN raw.response=2 THEN 0 WHEN raw.response=1 THEN 1 ELSE 99 END AS response
            FROM raw
            JOIN design ON design.ppid=raw.ppid
            JOIN og_cnts ON og_cnts.ppid=raw.ppid
            JOIN pp_actual_smry ON raw.ppid=pp_actual_smry.ppid
            JOIN str_actual_smry ON str_actual_smry.stratum=pp_actual_smry.stratum
            JOIN str_pp_sampled ON str_pp_sampled.stratum=str_actual_smry.stratum
            JOIN pp_xp_smry ON pp_xp_smry.ppid=design.ppid
            JOIN ppi ON ppi.ppid=design.ppid
            LEFT JOIN intdata ON intdata.intvwrid=raw.intvwrid
            LEFT JOIN int_xp_smry ON int_xp_smry.intvwrid=intdata.intvwrid
            ORDER BY raw.ppid
            ;

quit;
```

```
/*****************************************
END OF SETTING UP THE DATA.
ANALYSIS OF DATA
*****************************************/

options ls=150 FORMDLIM="*";
libname read "h:\Masters Project 2004 Analysis\";
data read.holdout_tmp read.validate_tmp;
set read.ep2004;
decide=ranuni(12345);
do i=100 to 1000;
      if ppid=i then do;
            if decide ge .5 then output read.holdout_tmp;
            else output read.validate_tmp;
      end;
end;
run;
data tmp2;
set read.ep2004;
run;
proc sql;
create table blahb AS select distinct ppid, count(weight_nr)/sum(weight_nr)*weight_nr AS wpred,
count(weight_nr_forced)/sum(weight_nr_forced)*weight_nr_forced AS wactual from read.ep2004;
create table tmp AS select * from tmp2 JOIN blahb ON tmp2.ppid = blahb.ppid;
quit;

proc logistic data=read.ep2004 order=data desc noprint;
class voterage (ref='18-35') voterrace (ref='Other') votersex (ref='Female') morningafternoon
(ref='Morning') intretailsales (ref='No')
/*intoutg (ref='Shy')*/
/ param=ref
;
model response (ref='Nonresponse') = intretailsales morningafternoon voterage votersex voterrace
/* intoutg bday */
/ aggregate scale=none clparm=both clodds=both lackfit link=glogit;
weight weight_nr_forced / norm;
output out=preds p=p lower=lower upper=upper;
run;
proc sort data=preds;
by p;
run;
proc sql;
select distinct p, intretailsales, morningafternoon, voterrace, votersex, voterage
from preds ORDER BY p;

select distinct p, intretailsales, morningafternoon, voterrace, votersex, voterage
from preds where intretailsales=0 and morningafternoon=1 and voterage=3 and voterrace=1 and
votersex=1;

select distinct p, intretailsales, morningafternoon, voterrace, votersex, voterage
from preds where intretailsales=0 and morningafternoon=1 and voterage=2 and voterrace=1 and
votersex=1;

select distinct p, intretailsales, morningafternoon, voterrace, votersex, voterage
from preds where intretailsales=0 and morningafternoon=1 and voterage=1 and voterrace=1 and
votersex=1;

select distinct p, intretailsales, morningafternoon, voterrace, votersex, voterage
from preds where intretailsales=0 and morningafternoon=1 and voterage=1 and voterrace=1 and
votersex=0;

select distinct p, intretailsales, morningafternoon, voterrace, votersex, voterage
from preds where intretailsales=0 and morningafternoon=0 and voterage=1 and voterrace=1 and
votersex=0;

select distinct p, intretailsales, morningafternoon, voterrace, votersex, voterage
from preds where intretailsales=1 and morningafternoon=0 and voterage=1 and voterrace=1 and
votersex=0;
```

```
select distinct p, intretailsales, morningafternoon, voterrace, votersex, voterage
from preds where intretailsales=1 and morningafternoon=0 and voterage=1 and voterrace=0 and
votersex=0;
quit;

proc surveylogistic data=tmp order=data;
strata stratum;
cluster ppid;
class voterage (ref='18-35') voterrace (ref='Other') votersex (ref='Female') morningafternoon
(ref='Morning') intretailsales (ref='No')
/ param=ref;
model response (order=data) = intretailsales morningafternoon voterage votersex voterrace;
*weight wactual;
*weight weight_nr_forced;
run;
proc logistic data=read.validate order=data descending ;
class IntDontKnowInterval intdoortodoor inthelpful
intmarriage intrace intsex intyearinschool intlearning intmajor intmission intnojobs intoutg
(ref='Outgoing')
intretailsales (ref='No') intbinaryschool spanish inttelesurvey intwaiter banner locatedinside
morningafternoon (ref='Afternoon') pp_table wasatch (ref='Wasatch Front')
voterage (ref='56+') voterrace (ref='Other') votersex (ref='Female') / param=ref;
model response (order=data) =
IntDontKnowInterval bday intdoortodoor inthelpful intheight
intmarriage intrace intsex intyearinschool intlearning
intmajor intmission intnojobs intoutg intretailsales
intbinaryschool spanish inttelesurvey intwaiter banner
interval locatedinside morningafternoon pp_table wasatch
voterage voterrace votersex / selection=backward alpha=.05;
*weight wactual;
*weight weight_nr_forced / norm;
run;

proc sort data=read.ep2004;
by stratum ppid;
PROC RLOGIST DATA=read.ep2004;
  *WEIGHT _ONE_;
  *WEIGHT weight_nr_forced;
  nest Stratum PPID / missunit;
  CLASS       intretailsales morningafternoon intoutg voterage votersex voterrace;
  *LEVELS         2 2 2 3 2 2;
  MODEL response = intretailsales votersex intretailsales * votersex voterage voterrace
morningafternoon intoutg;
  RECODE intoutg=(0 1) intretailsales=(0 1) morningafternoon=(0 1) voterrace=(0 1) votersex=(0 1);
  *REFLEV votersex=2 voterace=2 voterage=4 jobs1=2 timeampm=2 wasatch=2;

  *RECODE because 0 in SUDAAN represents a missing value;

run;
```

Bibliography

AAPOR (1997), *Best practices for survey and public opinion research and survey practices that AAPOR condemns.*, Ann Arbor, MI: AAPOR.

— (2004), *Standard definitions: Final disposition of case codes and outcome rates for surveys. 3rd ed.*, Lenexa, KS: AAPOR.

An, A. and Watts, D. (1999), "New SAS Procedures for Analysis of Sample Survey Data," http://support.sas.com/rnd/app/papers/survey.pdf.

Bauman, S., Merkle, D., and Lavrakas, P. (1992), "Interviewer estimates of refusers' gender, age, and race in telephone surveys." *Paper presented at the Midwest Association for Public Opinion Research Conference, Chicago.*

Dillman, D. A. (1978), *Mail and telephone surveys: the total design method.*, New York: Wiley.

Edelman, M. and Merkle, D. (2000), *A review of the 1996 Voter News Service Exit Polls from a total survey error perspective. In Election Polls, the news media, and democracy. Paul J. Lavrakas and Michael W. Traugott, eds.*, New York: Seven Bridges Press.

Edison Media Research and Mitofsky International (2005), "Evaluation of Edison/Mitofsky Election System 2004," Http://www.exit-poll.net/election-night/EvaluationJan19 2005.pdf.

Grimshaw, S. D., Christensen, H. B., Magleby, D. B., and Patterson, K. D. (2004), "The Utah Colleges Exit Poll: Learning by Doing," *Chance*, 17, 32–38.

Groves, R. and Couper, M. (1998), *Nonresponse in Household interview surveys*, New York: Wiley, p. 34.

— (2002), *Nonresponse in Household interview surveys*, New York: Wiley.

Keeter, S., Miller, C., Kohut, A., Groves, R., and Presser, S. (2000), "Consequences of reducing nonresponse in a national telephone survey," *Public Opinion Quartley*, 64, 125–148.

Lavrakas, P. J. (1993), *Telephone survey methods: Sampling, selection, and supervision, 2nd ed.*, Newbury Park, CA: Sage Publications.

Lohr, S. (1999), *Sampling: Design and Analysis*, Pacific Grove, CA: Brooks/Cole.

Merkle, D. and Edelman, M. (2002), *Nonresponse in exit polls: A comprehensive analysis. In Survey Nonresponse, Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderck J. A. Little, eds.*, New York: Wiley-Interscience, pp. 243–257.

Mock, D., Christensen, H. B., Stevenson, C. W., Olsen, J. A., and Patterson, K. D. (2006), "Utah Colleges Exit Poll: Sample Design and Methodology," *Brigham Young University*, unpublished.

SAS Institute Inc. (2006), *SAS OnlineDoc® 9.1.3*, Cary, NC: SAS Institute Inc., http://support.sas.com/onlinedoc/913/docMainpage.jsp.

Singer, E., Frankel, M. R., and Glassman, M. B. (1983), "The effect of Interviewer Characteristics and Expectations on Response," *Public Opinion Quartley*, 47, 68–83.

Singer, E., Hoewyk, J. V., and Maher, M. P. (2000), "Experiments with incentives for survey participation in telephone surveys," *Public Opinion Quartley*, 64, 413–428.

Slater, M. C. (2002), "Applying AAPOR's final Disposition Codes and Outcome Rates to the 2000 Utah Colleges' Exit Poll," Master's thesis, Brigham Young University.

Stevenson, C. W., Redding, M., Stevenson, A., Mah, K., Steed, S., Larsen, R., Poulton, M., Rajeev, D., Johnson, P., Li, J., and Wunderlich, J. (2004), "Exit Poll Report 2004," *Brigham Young University*, 37, unpublished.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2002), *The Psychology of Survey Response*, Cambridge: Cambridge University Press.