

ANALYZING ALCOHOL
BEHAVIOR IN
SAN LUIS OBISPO
COUNTY

Ariana Montes

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, Statistics

December 2014

TABLE OF CONTENTS

Methods	2
Part I: Original Plans	3
Part II: Grant Proposal	5
Part III: Demographics	6
Part IV: Analysis	12
Conclusion	13
Limitations	14
Acknowledgments	15

METHODS

The data was obtained from the Center for Disease Control (CDC)'s Behavioral Risk Factor Surveillance System (BRFSS). The annual survey data were downloaded in ASCII format. The files for 2005 to 2012 were individually read into SAS® 9.3 supplied by SAS Studio University Edition. Each dataset included 359 variables and approximately 350,000 to 500,000 observations from all 50 states. I narrowed this down to California first, and then to San Luis Obispo County. The primary predictors of interest included variables related to basic demographics characteristics, the outcome of alcohol behavior, and survey analysis-type variables (i.e. weight, strata, and cluster).

Various data manipulation took place in the creation of the data set. I created an age group variable of college-aged people (18-24) vs. not college-aged people (25+). The original reason for these categories was because I was interested in assessing alcohol behavior among college-aged people. Unfortunately, the younger age group was not as representative in the survey collecting, which is expressed in the demographic on page 7. A potential reason for this the lack of surveying cell phone users.

The demographic of marital status was put into the following groups: married, divorced, never married, and widowed. Employment status was categorized into: employed, not working, homemaker, student, and retired. The education level variable was categorized as high school graduate or below and college graduate. For income, the grouping is as follows: less than \$35000, \$35000 to \$50000, \$50000 to \$75000, and \$75000 or more. Race and ethnicity was divided into the following: White, Hispanic, and other.

The two main procedures that I used throughout my analysis were PROC SURVEYFREQ and PROC SURVEYLOGISTIC. For these procedures, the syntax included a WEIGHT, STRATA, and CLUSTER statement. The associated variables for these statements were found in the original data. For the weight variable in the data, there appeared to be a fault between the years 2005-2010 and 2011-2012. The names for the weights were different in these two groups, so in order to account for this glitch, I simply changed the name so that they would line up consistently.

Originally, I had stacked the eight years of data together and ran the analyses. However, this became an issue with the weighted frequencies. Since the sample sizes from the eight years added together, the total sizes were not representative of San Luis Obispo County's actual population sizes. I separated the years once again and was able to continue the analysis appropriately.

When deciding on how to assess the univariate variables for demographics, I decided that it would be more interesting to examine the demographics based on drinking behavior. For example, instead of assessing how many drinkers are married, the route I chose was to assess the following: "Of those who are married, what percentage are drinkers?" This decision explains why the frequencies do not add to one hundred percent.

To build the models found in the analysis starting on page 13, I initially ran univariate logistic regression models for each of the predictors of interest. With these, I made a list of which variables were potential multivariable predictors at the alpha .20 level. I then ran a saturated model with these potential predictor variables and removed the least significant variables one by one until I came upon a model with all variables with p-values significant at the .05 level.

PART I: ORIGINAL PLANS

In October of 2014, I was presented with the opportunity to work on a project with a nonprofit organization in San Luis Obispo. The organization was interested in understanding the general public's view on consumption of alcohol during pregnancy through the method of a countywide survey. Their goal was to use the survey data to present to the public as an example for teaching the harms of drinking alcohol during pregnancy. Unfortunately, this group wanted to conduct the survey themselves with no proper statistical knowledge to create the sampling frame. We felt that this would have been against the ethical guidelines of running a statistically sound survey. In an effort to relieve this predicament, we suggested the hiring of a local survey sampling company.

In an effort to decide on questions to put on the potential questionnaire, I researched previous surveys that had been conducted throughout the United States. After researching interesting questions and the results found from these surveys, I compiled the questions that invoked the most interesting results. The reason for this was to see if we could recreate these results in San Luis Obispo County and compare the populations of both groups that were questioned.

The Minnesota Organization on Fetal Alcohol Syndrome (MOFAS) conducts a survey annually and assesses alcohol behavior among pregnant women. This website was a great resource for appropriate questions to ask on a potential survey. For example, one question that was from this survey and a consideration for the project was the following:

Would harm to a baby be more likely if a pregnant woman drank beer, wine, or liquor?

- All of these would cause the same amount of harm*
- Liquor would be more harmful*
- Wine would be more harmful*
- Beer would be more harmful*
- None of these would be harmful*

The reason for this question on the survey was because the study had found that one quarter of respondents thought that liquor would be more harmful than beer or wine.¹ I thought that if we used this question on the County Trax, it would be interesting to note possible similar results and inform the population about the misconceptions.

The Public Health Agency of Canada reported a survey that examined the knowledge and beliefs about alcohol use during pregnancy, awareness of FAS and FAE, and the expected behaviors of women and partners of women, during pregnancy.² The question that we decided to use from this survey is:

For women:

Thinking about yourself, if your spouse or partner continued to drink alcohol during your pregnancy,

- I would be more likely to drink alcohol*
- I would be less likely to drink more alcohol*
- It would not make a difference*
- I don't drink alcohol*
- Don't know/ Not sure*
- Don't have a spouse or partner*

The reason for this question on the survey was because the Canadian study found interesting results on social support for reduced alcohol consumption during pregnancy.

¹ <http://www.mofas.org/2013/10/alcohol-use-and-pregnancy/>

² <http://www.phac-aspc.gc.ca/publicat/fas-saf-natsurv-2000/index-eng.php>

A prenatal questionnaire conducted in Montana sought to describe the knowledge and attitudes of women of childbearing age regarding alcohol consumption during pregnancy and its effects on the fetus.³ The following question is an example of a question that we considered for the County Trax survey:

*When a woman is pregnant, how much alcohol do you think is safe for her to drink during her pregnancy?
(For the purposes of this survey, a “drink” is 12oz (can/bottle) of beer, 5oz (one glass) of wine, 1oz (shot, or mixed drink) of hard liquor)*

- Never; no amount of alcohol is safe*
- Once a month or less*
- 2 to 4 times per month*
- 2 to 3 times per week*
- More than 3 times per week*
- Don't know*

The reason for this proposed question was because the Montana survey found that respondents agreed that some amounts of alcohol were safe during pregnancy, as long as it was not too much.

These types of questions were combined into a list of potential questions for the survey and presented to the nonprofit organizations.

While this process was statistically sound, it was also quite expensive for the nonprofit organization. This expense was undesirable. The solution was to recruit more nonprofit organizations interested in surveying the county to share in the cost. Although this served as a successful alternative at first, it soon turned into a constant struggle with the members of the coalition and their understanding of how to conduct the survey properly and what sampling frame to use. From October until July, there was a lot of back and forth and eventually, the project came to an inevitable stall.

While the local survey became less and less likely, I continued to explore options for data to work with. Ultimately, I stumbled upon the BRFSS data sets and I decided to analyze alcohol behavior in general. My initial hypothesis of interest was to see if there was an association between college-aged people and drinking behaviors.

³ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2936428/>

PART II: GRANT PROPOSAL

In April of 2014, I received an email regarding the Warren J. Baker Endowment Grant. During the process of working with the nonprofits, we had agreed to put out a countywide survey and different organizations decided to pay for questions pertaining to their topics of interest. I thought that it would be interesting to use funds (up to \$2,500) to buy my own questions for the survey. I wanted to include questions on alcohol behavior of college-aged people. I expressed these ideas on my grant proposal.

The following were the objectives for the proposal:

- (1) To participate with nonprofit community partners to take part in a professional countywide survey.
- (2) To collect data on drinking behaviors in San Luis Obispo County with a focus on college-aged people.

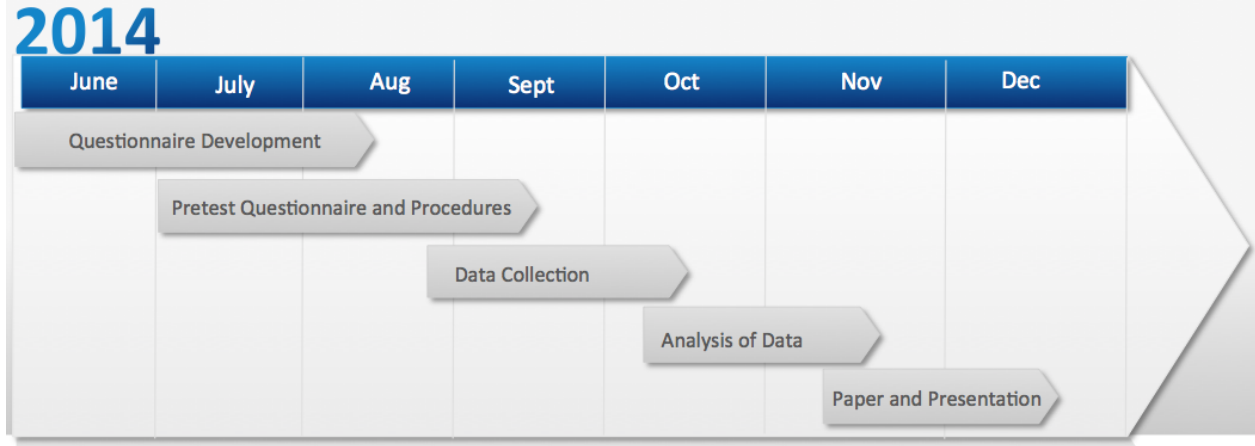


Figure 1. *Timeline for Baker Science Grant Proposal*

Figure 1 shows the proposed timeline for the project. I had projected that the project last seven months in total. The development of the questionnaire would take about two months, the pretest procedures would last about three months, data collection would last about two, analysis would take about one month, and the paper and presentation would take about a month as well.

The selection committee received over 36 proposals requesting over \$136,000 in total, when there was less than \$20,000 in available funds for the year. Unfortunately, I was not considered for the award; however, writing the grant was a positive experience in that it gave me an understanding for what the grant writing process is like.

PART III: DEMOGRAPHICS

i. Gender

Drinkers by Gender		
Year	Gender	Percent
2005	Male	42.44
	Female	25.10
2006	Male	22.17
	Female	39.66
2007	Male	30.31
	Female	29.25
2008	Male	34.32
	Female	28.63
2009	Male	23.93
	Female	25.88
2010	Male	27.54
	Female	27.86
2011	Male	32.48
	Female	32.96
2012	Male	28.40
	Female	28.91

Table 1A. *Drinkers by Gender*

Table 1 shows the distribution of drinkers by gender throughout the seven years of analysis. It shows that drinkers are pretty evenly distributed between males and females from 2007 thru 2012. The data from 2005 shows us that there were more males who responded saying they were drinkers (~42%) and in the next year, females had a higher proportion of responses to being drinkers (~40%).

ii. Age Group

Drinkers by Age Group		
Year	Gender	Percent
2005	18-24	10.94
	25+	56.49
2006	18-24	0.00
	25+	61.83
2007	18-24	9.60
	25+	49.96
2008	18-24	8.63
	25+	54.32
2009	18-24	4.79
	25+	44.79
2010	18-24	1.78
	25+	53.63
2011	18-24	7.26
	25+	58.18
2012	18-24	3.98
	25+	53.33

Table 2. *Drinkers by Age Group*

Table 2 shows the distribution of drinkers by age group. In 2006, none of the respondents under the age of 24 reported drinking. This pattern might be concerning because of the uneven distribution, but this variable was originally created to observe drinking behaviors of college-aged people.

iii. Race/Ethnicity

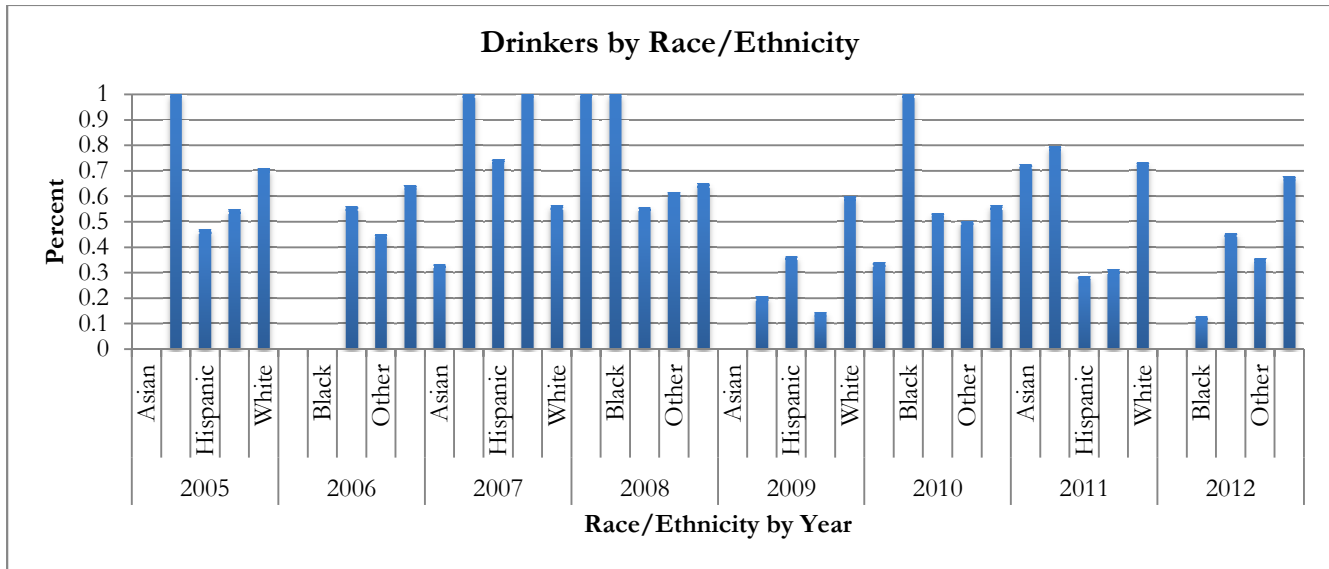


Figure 2. *Distribution of Drinkers by Race/Ethnicity*

Figure 2 shows the distributions of drinkers by racial group. These results are very scattered and appear inconsistent because the sample sizes for some groups within years were not large enough to be representative of the population. In an effort to visualize patterns in racial drinking behavior, I subset this analysis narrowing down to Hispanic and White only as seen in Figure 3 below.

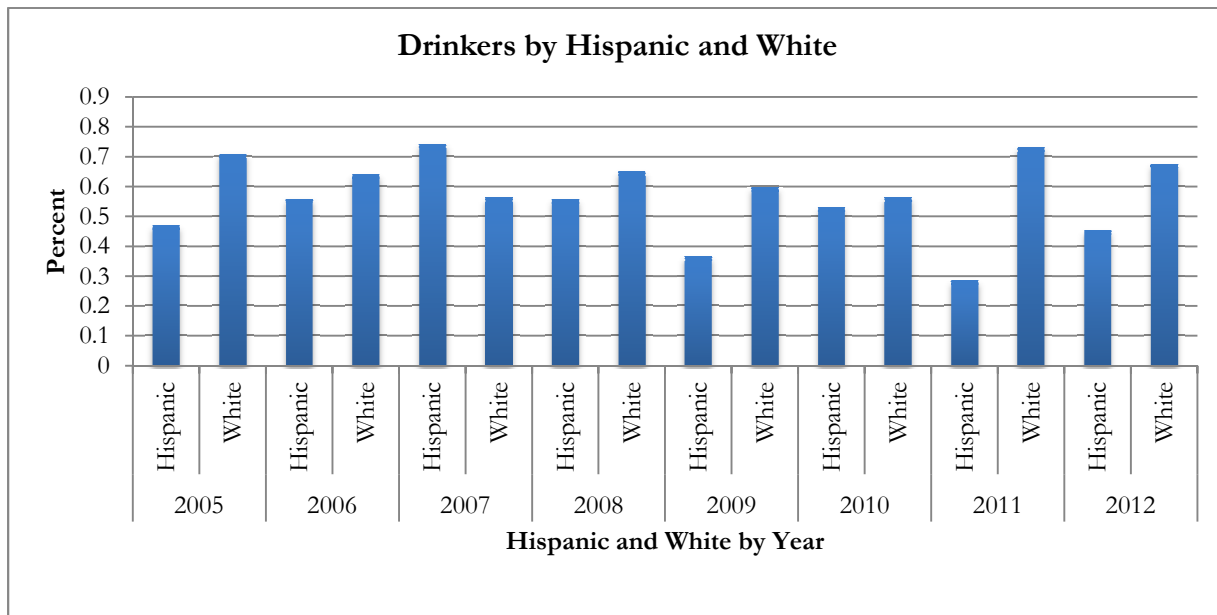


Figure 3. *Distribution of Drinkers by Race/Ethnicity (Hispanic and White only)*

Figure 3 shows the distribution of drinkers by Hispanic and White. We can see that Whites report drinking more than Hispanics seven out of the eight years.

iv. Marital Status

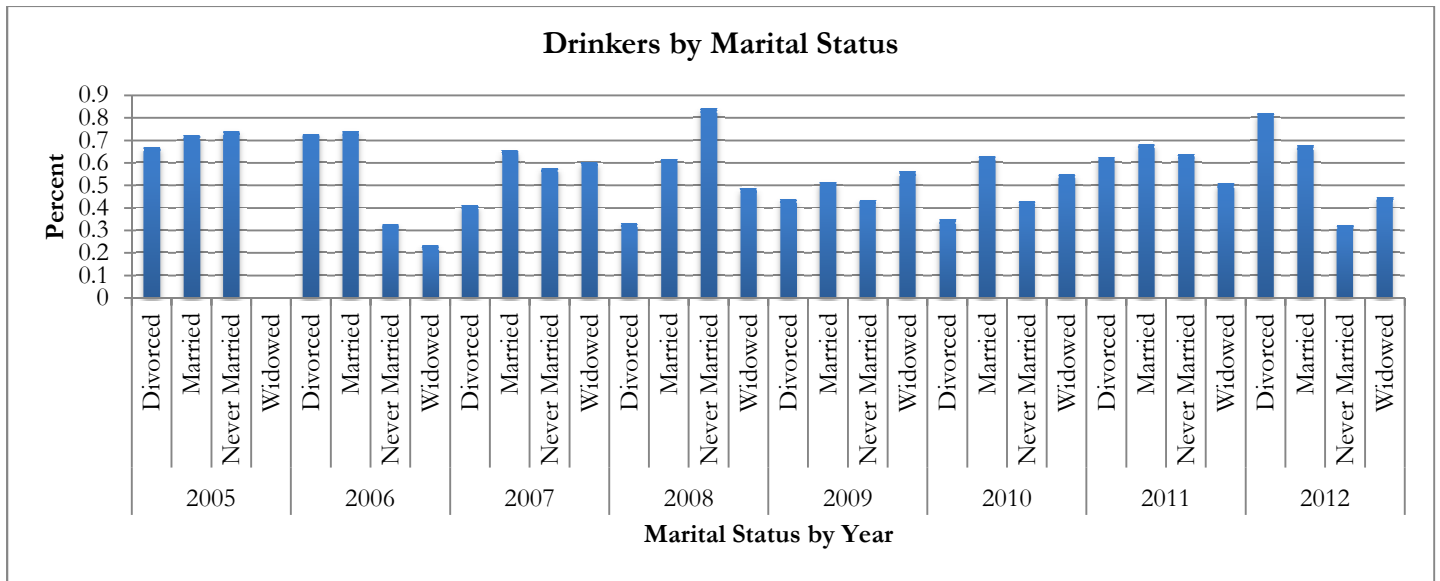


Figure 4. *Distribution of Drinkers by Marital Status*

Figure 4 shows the distributions of drinkers by reported marital status. We can see that more than 60% of married people reported being drinkers every year except for 2009. Those who were divorced or never married also appeared to report drinking more than those who were married or widowed.

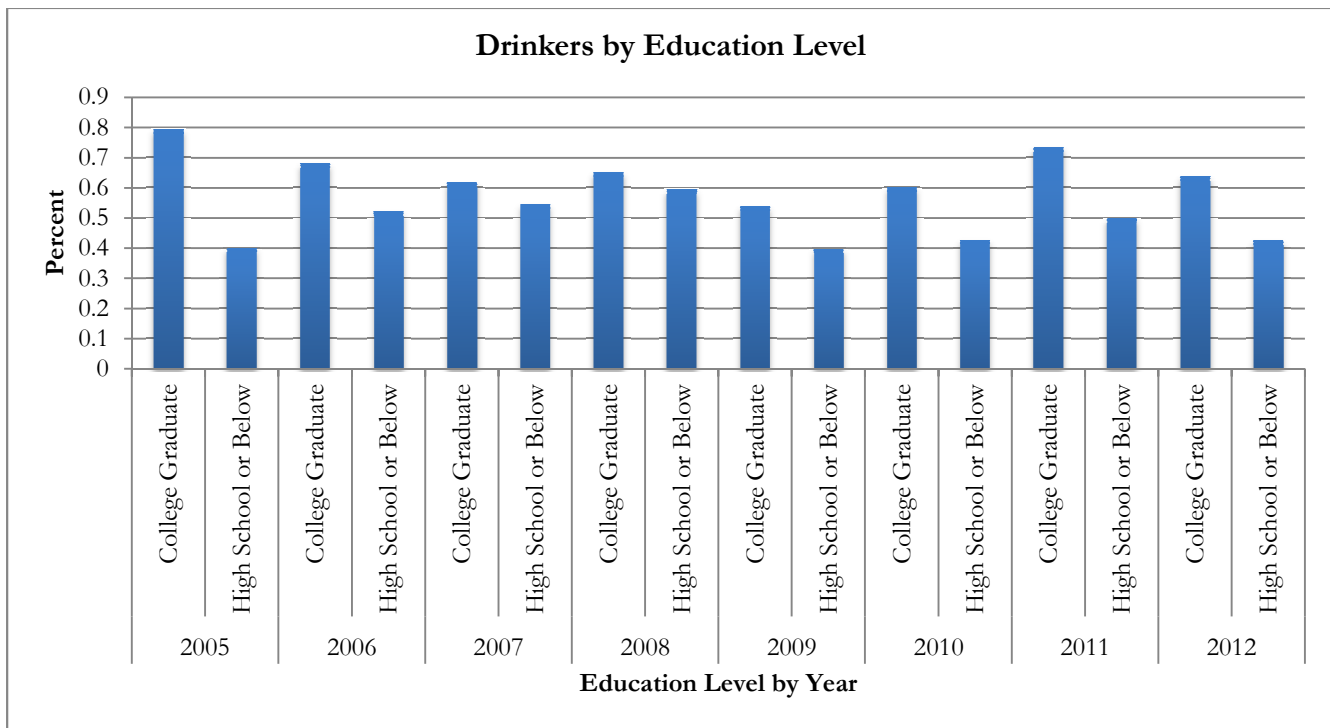


Figure 5. *Distribution of Drinkers by Education Level*

Figure 5 above shows the distributions of drinkers by education level. A surprising result to note from these graphs is that college graduates are much more likely to drink than high school graduates or below. This trend is consistent throughout the eight years.

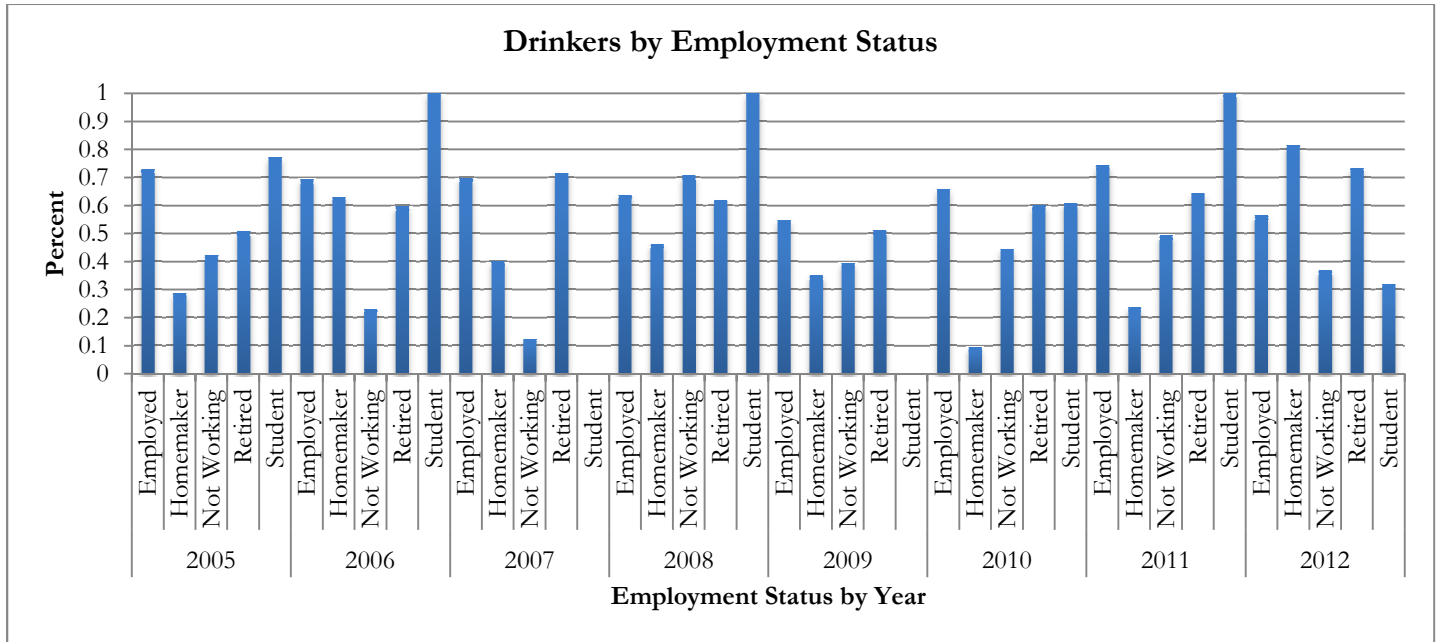


Figure 6. *Distribution of Drinkers by Employment Status*

Figure 6 displays the distribution of drinkers by reported employment status. It shows all of the surveyed students as reported drinkers in 2006, 2008, and 2011. However, 2007 and 2009 show the complete opposite. This apparent contradiction is because of the small samples in each year's data.

PART IV: ANALYSIS

From the demographics section, we can see that there was not too much of a difference in variables between the eight years. Because of this, I am focusing my multivariable analysis on responses from 2012, the most recent year.

Drinkers

Drinkers		
Effect	df	p-value
Marital Status	3	0.0503
Age Group	1	0.3791
Race	2	0.0088

Table 3. *Logistic Regression Output for Drinkers*

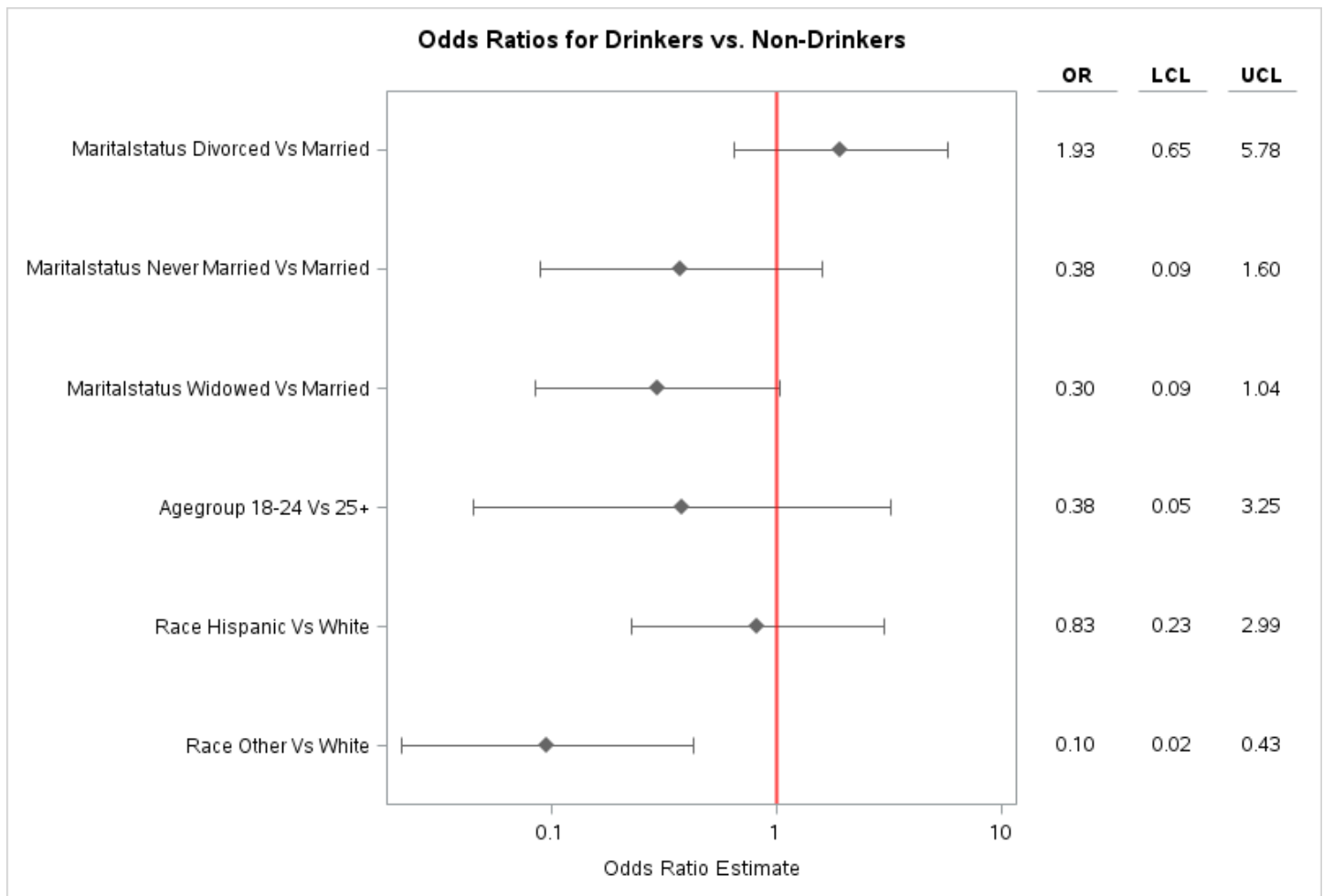


Figure 7. *Forest Plot of Odds Ratios from Logistic Regression in Table 3*

Table 3 and **Figure 7** are the results of the logistic regression for drinkers. We can see from **Table 3** that Race/Ethnicity is the only significant predictor ($\alpha=.05$). The forest plot in **Figure 7** shows that this significance occurs with Other vs. White. It shows that other races and ethnicities report drinking less than whites do.

CONCLUSION

From these results, I found interesting characteristics of San Luis Obispo County. I discovered some key features about drinking behaviors of the county's residents. In terms of gender, although there may be premonitions about one gender as more likely to drink than the other, my results show that the distribution of drinkers among the two are not very different. It was also discovered that Whites report drinking more than Hispanics. We also can note that college graduates report drinking more than those who graduated from high school or below. From the logistic regression, it was discovered that Whites drink significantly more than Hispanics or those of other races.

LIMITATIONS

One limitation of this project is the lack of respondents of the age group 18-24. This group was not representative compared to the age group 25 years or older. In order to resolve this discrepancy, the survey could be extended to cell phone numbers and not simply landlines. Those of the younger age group seem less likely to own their own house, and thus less likely to have a landline. Cell phones, however, are growing in popularity.

ACKNOWLEDGMENTS

I would like to thank my senior project advisors Rebecca Ottesen and Heather Smith for their support and assistance throughout the entirety of this project.