STAT 2183

Professor Landon

Min Yang

**Assignment 3 Report**

**Introduction**

To test what effect various categorical factors have on political ideology, we collect the data from a subset of the 2008 American National Election Study. The data set contains 840 observations with 6 variables—Region, Ideology, religious services, marital status, ethnicity and gender.

**Method**

First, we test the hypothesis of the proportion of different types of ideology: whether there is sixty percent of the population either moderate, slightly liberal, or slightly conservative, and twenty percent of the population either liberal or extremely liberal, and twenty percent of the population either conservative or extremely conservative.

We define the factor “ideology” with three levels: moderate, liberal, and conservative. Meanwhile, the expected cell counts are larger than 5. Then, we utilize the Chi-Square Goodness of Fit Test to see if the population proportion is corresponded to the hypothesis with significant level of 0.05.

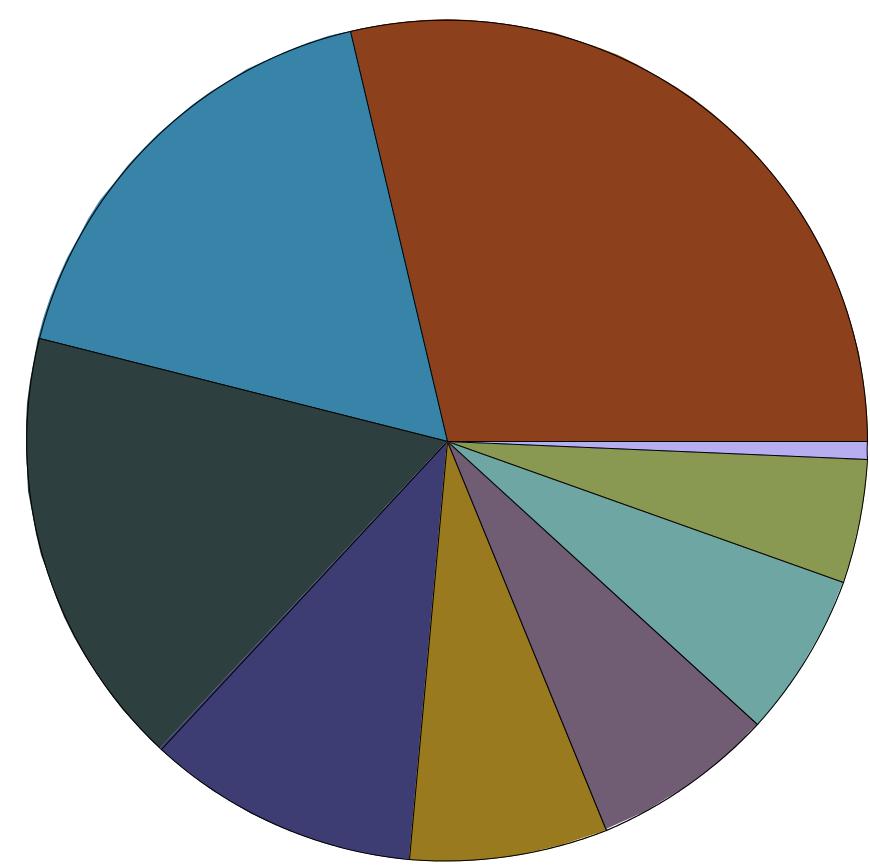
Secondly, we test two factors which might have effect on choosing political party—ethinicity and Marital status. We want to test the following two hypotheses:

* Married respondents are more likely than others to identify with the Republican Party.
* White people are less likely than others to identify with the Democratic Party.

Because each of the samples used in the test is large and the expected cell counts are larger than 5, we perform the Chi-Square Independence Test for both hypotheses. From there, we determine if the factors (gender, marital status) and political ideology are dependent or not by performing the test with significant level of 0.05.

The pie chart shows the factor ideology which is comprised of nine levels.

**survey of ideology (amount of people)**

****

146 241

6

40

143

53

59

88

64

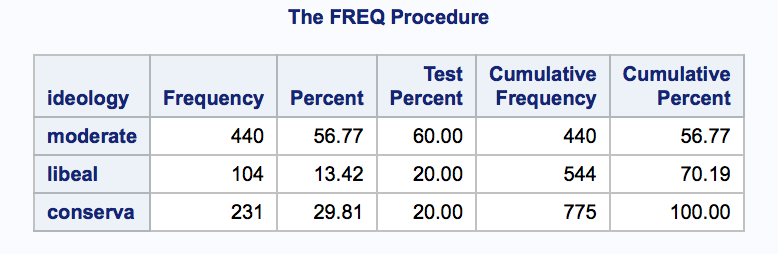
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ideology |  | conse |  | ex-conse |  | ex-liber |  | liberal |  | moderate |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  | refuse |  | sl-conse |  | sl-liber |  | no answers |  |  |
|  |  |  |  |  |  |  |

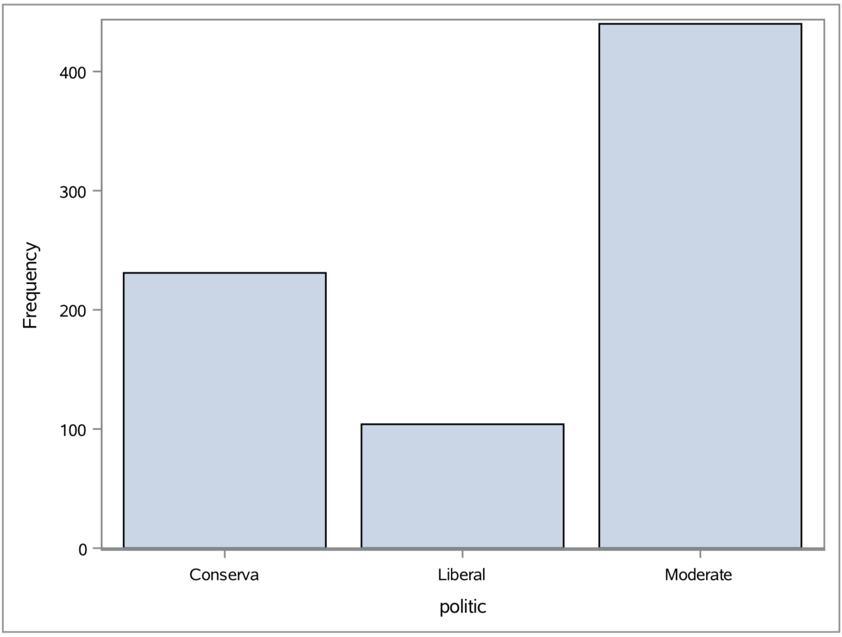


**Test for the Proportion:**

* 60% of the population is either moderate, slightly liberal, or slightly conservative.
* 20% of the population is either liberal or extremely liberal.
* 20% of the population is either conservative of extremely conservative.

Contingency Table





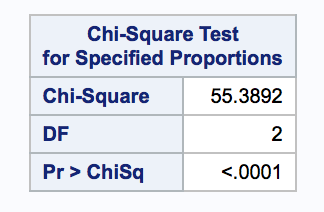
We can construct the hypothesis to test the proportion:

H0: 60% of the population is either moderate, slightly liberal, or slightly conservative

20% of the population is either liberal or extremely liberal

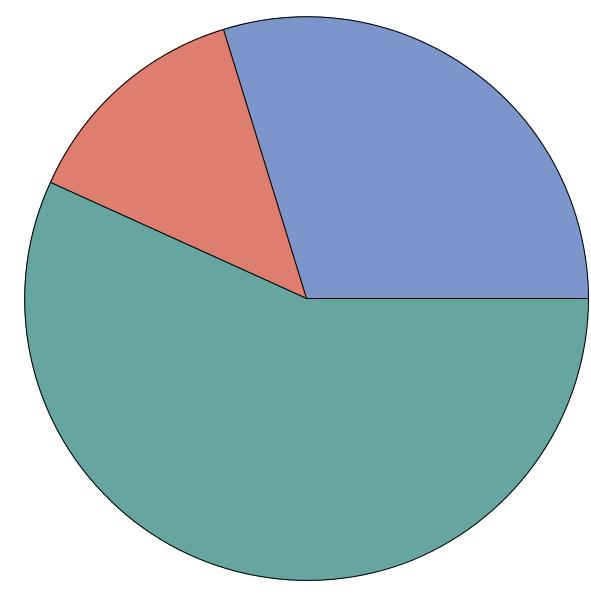
20% of the population is either conservative of extremely conservative.

Ha: at least one of the proportions differs from its hypothesis value.



The test-statistics is 55.3892, and the p-value is less than 0.0001. At the significant level α = 0.05, we reject the null hypothesis. So at least one of the proportion differs from its hypothesis value.

SUM of count by ideology



From the pie chart, we can see the proportion of group “moderate” is close to 60%, but the proportion of group “liberal” is less then 20% and the proportion of group “conservative” is more than 20%.

Conservative

liberal extreme conservative

extreme liberal 231

104

Moderate, slightly liberal,

Slightly conservative

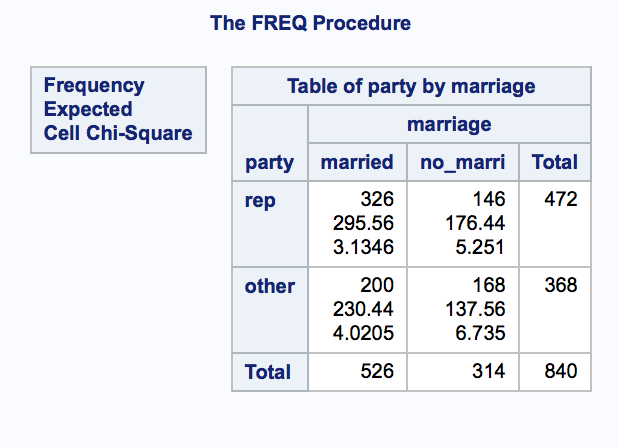
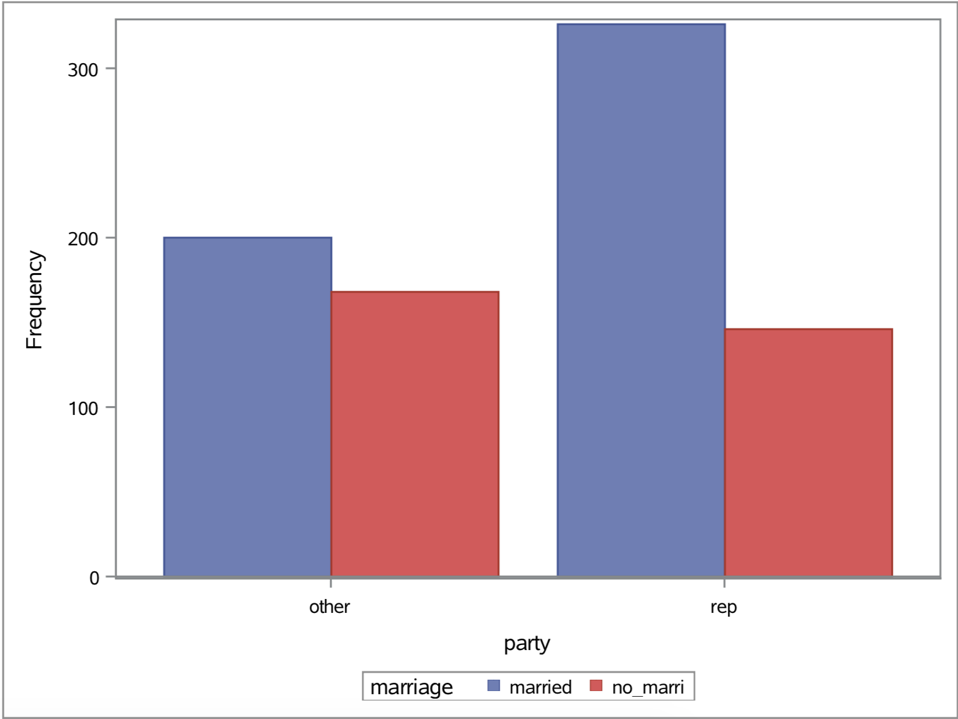
440

It is not true that sixty percent of people whose ideology is either moderate, slightly liberal, or slightly conservative; twenty percent of people whose ideology is either liberal, or extremely liberal; twenty percent of people whose ideology is either conservative or extremely conservative.

**Test 1:**

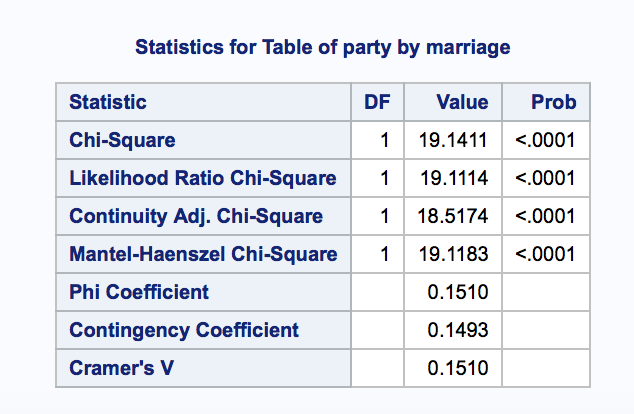
Married respondents are more likely than others to identify with the Republican Party.

To test the independence of political party and Marital status, we can construct the hypothesis to test the proportion:



H0 : Married respondents and others have the same possibility to identify with the Republican Party (marital status and political party is independent)

Ha : Married respondents are more likely than others to identify with the Republican Party. (marital status and political party is dependent)



The test-statistics is 19.1411, and the p-value is less then 0.0001. At the significant level α = 0.05, we reject null hypothesis.

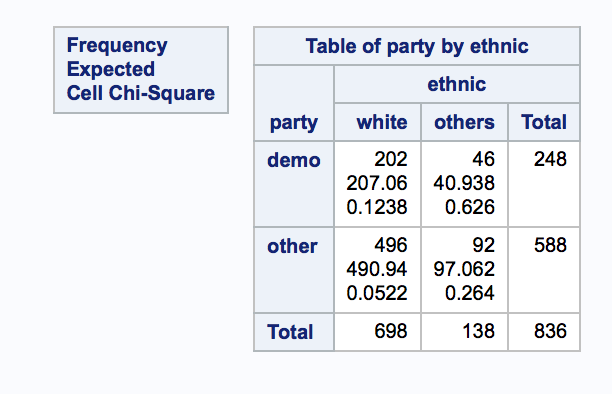
we can draw the conclusion that married respondents and others have different possibilities to identify with the Republican Party. To be specific, 295.56 married people are expected to be republican, but in fact there is 326 married people are republican; and for the same reason, the actual number of republican who are no-married is less than the expected number. So married people are more likely to be republican than those who are not married.

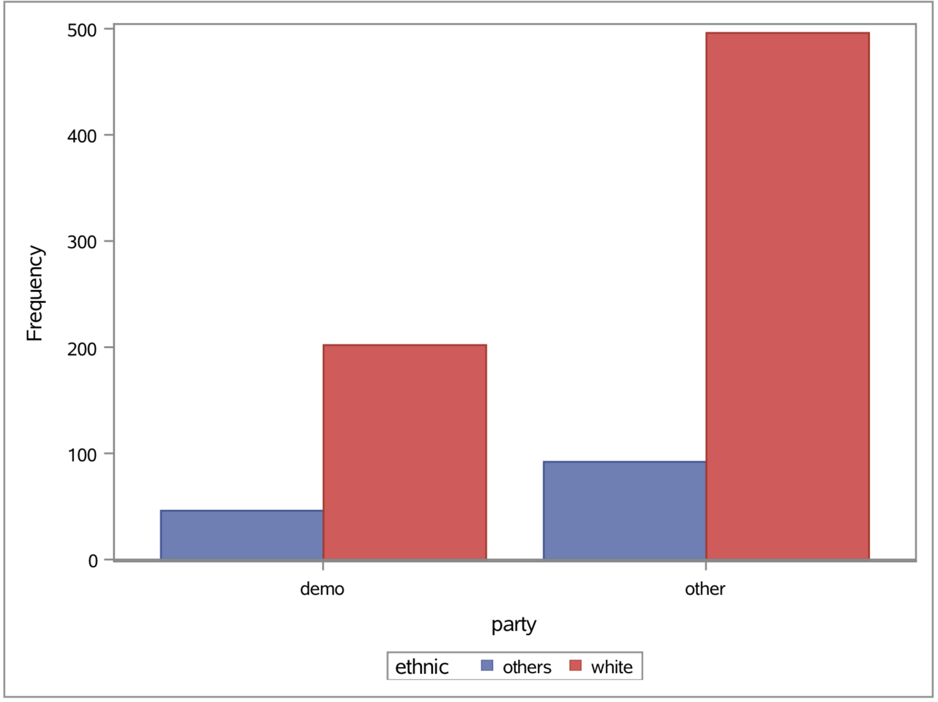
**Test 2:**

White people are less likely than others to identify with the Democratic Party

We assign the people with political ideology of slight liberal, liberal, extremely liberal to group “democratic.” Then, we assign people whose political ideology is moderate, slight conservative, conservative and extremely conservative, and those who did not or refused answer their ideology are assigned to the group “others.”

The sum of “democratic” and “others” should be 840, but in the data set, there are 4 data missing for variable “ethnic.” So we only have 836 observations for this test.

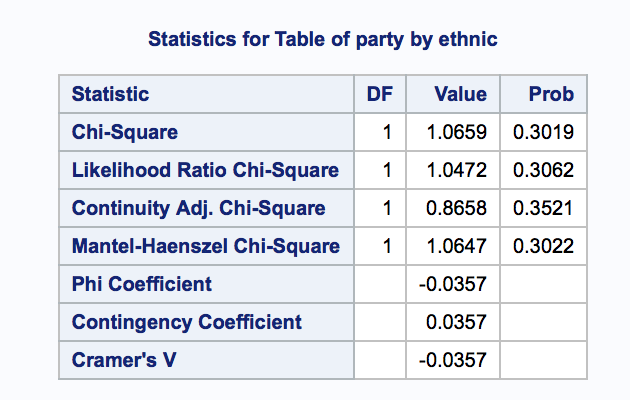




To test the independence of political party and ethnicity, we can construct the hypothesis to test the proportion:

H0 : White people and others have the same possibility to identify with the Democratic Party

Ha : White people are less likely than others to identify with the Democratic Party.



The test-statistics is 1.0659, and the p-value is 0.3019. At the significant level α = 0.05, we do not to reject null hypothesis.

From the test, we do not reject the null hypothesis, so white people and others have the same possibility to identify with the Democratic Party. We can also see that actual and expected numbers are very close for each index.

**Conclusion**

From the chi-square goodness of fit test, we reject the hypothesis and conclude that It is not true that sixty percent of people whose ideology is either moderate, slightly liberal, or slightly conservative; twenty percent of people whose ideology is either liberal, or extremely liberal; twenty percent of people whose ideology is either conservative or extremely conservative.

It is expected that marital status is dependent with political party, and married respondents are more likely than others to identify with the Republican Party. It is also expected that ethnicity is independent with political party, as white people and others have the same possibility to identify with the Democratic Party.

SAS code

data hw\_3;

input id region$ ideology$ relseve$ marstat$ ethnic$ gender$; datalines;

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 6 | 3 | 1 | 5 | 1 |
| 2 | 2 | 3 | 5 | 1 | 5 | 2 |
| ~~~~ |  |  |  |  |  |  |
| 1809 | 2 | 3 | 2 | 5 | 5 | 2 |
| run; |  |  |  |  |  |  |

data hw3\_1;

input ideology$ count;

datalines;

moderate 440

libeal 104

conservative 231

;

run;

ods rtf file='hw3\_1st\_final.rtf';

proc means data=hw\_3;

class ideology;

output;

run;

proc freq data=hw3\_1 order = data;

weight count;

tables ideology / testp=(.6, .2, .2); run;

proc gchart data = hw3\_1;

format count;

pie ideology / sumvar=count

coutline = black;

run;

/\*democratic 1+2+3 : 40+64+146=250\*/

/\*republican 7+6+5 : 88+143+241=472\*/

/\*others: 118\*/

/\*Married respondents are more likely than others to identify with the

Republican Party\*/

proc means data=hw\_3;

class marstat ideology;

output;

run;

data hw\_3\_5;

input party$ marriage$ number;

datalines;

rep married 326

rep no\_married 146

other married 200

other no\_married 168

run;

PROC SGPLOT DATA=hw\_3\_5;

VBAR party / freq=number group=marriage groupdisplay=cluster;

RUN;

proc freq data=hw\_3\_5 order=data;

weight number;

tables party\*marriage / chisq expected cellchi2 nocol norow nopercent; run;

/\*White people are less likely than others to identify with the

Democratic Party\*/

proc means data=hw\_3;

class ethnic ideology;

output;

run;

data hw\_3\_6;

input party$ ethnic$ number;

datalines;

demo white 202

demo others 46

other white 496

other others 92

run;

PROC SGPLOT DATA=hw\_3\_6;

VBAR party / freq=number group=ethnic groupdisplay=cluster;

RUN;

proc freq data=hw\_3\_6 order=data;

weight number;

tables party\*ethnic / chisq expected cellchi2 nocol norow nopercent; run;