Statistical Analysis Project

SOC/ANT 3402

Britney Gomes

Part 1: Choosing your Topic and Variables

My dependent variable is health. This is an important topic for social researchers to identify social patterns. Aspects of health are influenced and determined by various factors across the spectrum. Understanding the condition of health through the lenses of the dependent variables picked below can tell us how certain demographics of people perceived their health condition. This can with further research allow us to evaluate what health disparities and inequities exist for those with higher rates of poor health.

Dependent Variable:
Health (ordinal)
1 Excellent
2 Good
3 Fair
4 Poor

Independent Variables:
Class (Ordinal)
1. Lower Class
2. Working Class
3. Middle Class
4. Upper Class
5. No Class

Race (Nominal)
1. White
2. Black
3. Native American
4. Asian
5. Pacific Islander
6. Other
7. Hispanic

Sex (nominal)
0. Male
1. Female

Marital Status (Nominal)
0. Not Married
1. Married

Research Hypotheses

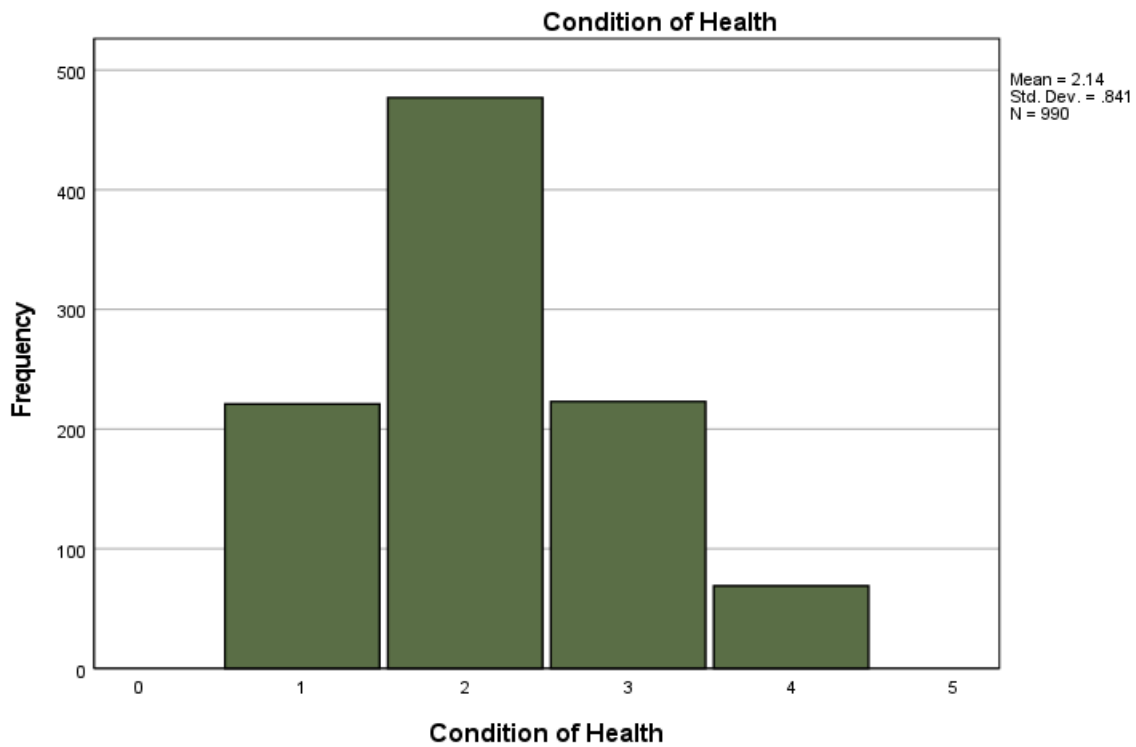H1: On average, men are more likely to have lower ranking health compared to women.

H2: On average, minority races will have lower-ranking health conditions, compared to white counterparts.

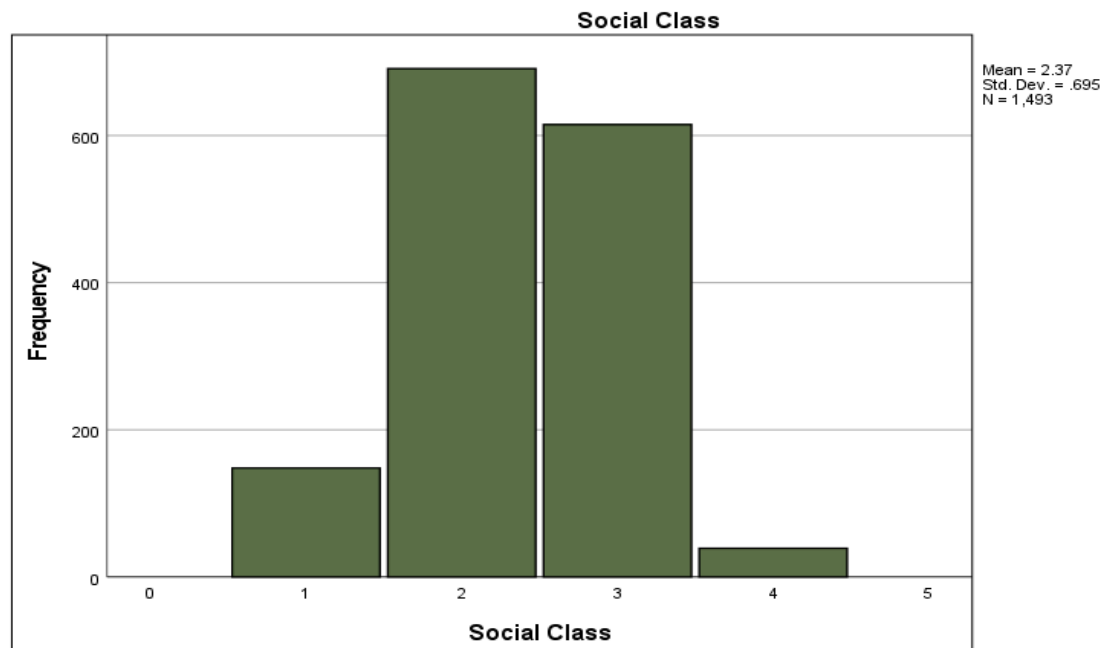H3: On average, high or middle-class groups will have higher rates of "excellent or good" conditions of health.

H4: On average, married groups will perceive better health conditions compared to unmarried groups.

Part 2: Descriptive Statistics

DV: Health

Condition of Health

Mean = 2.14
Std. Dev. = .841
N = 990

(Histogram of Condition of Health, Frequency on y-axis, Condition of Health on x-axis)

IV: Social Class

Social Class

Mean = 2.37
Std. Dev. = .695
N = 1,493

(Histogram of Social Class, Frequency on y-axis, Social Class on x-axis)

IV: Race

Race of Respondent

Mean = 1.58
Std. Dev. = 1.413
N = 1,487

IV: Sex

Female
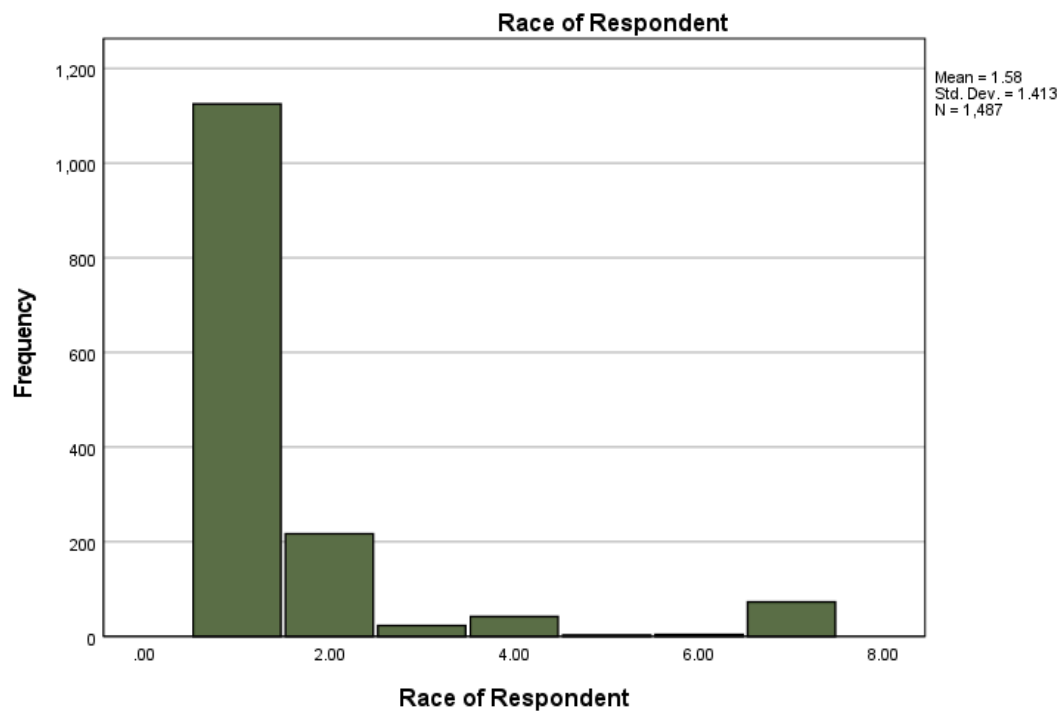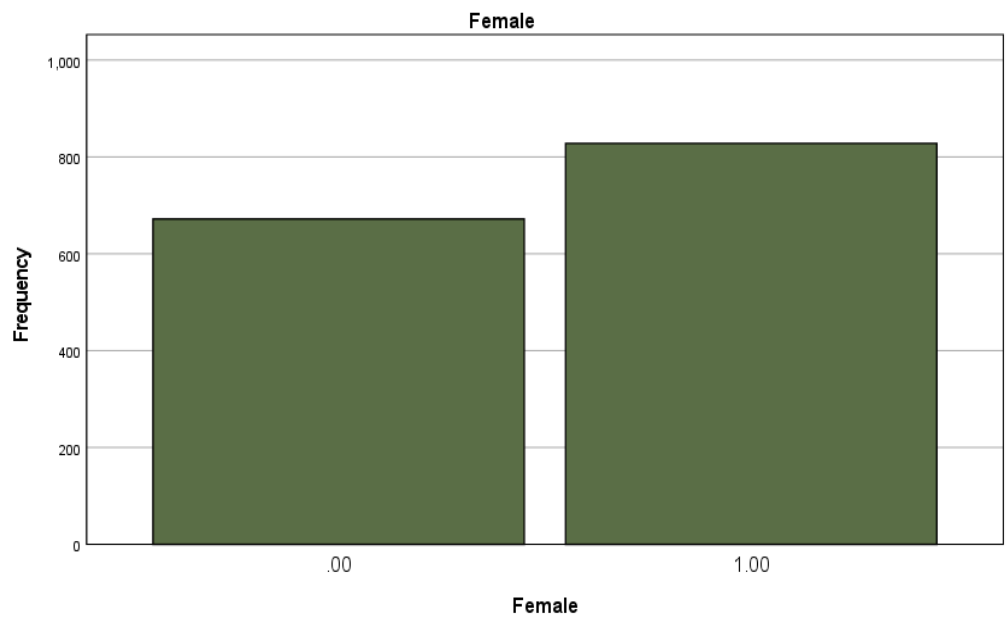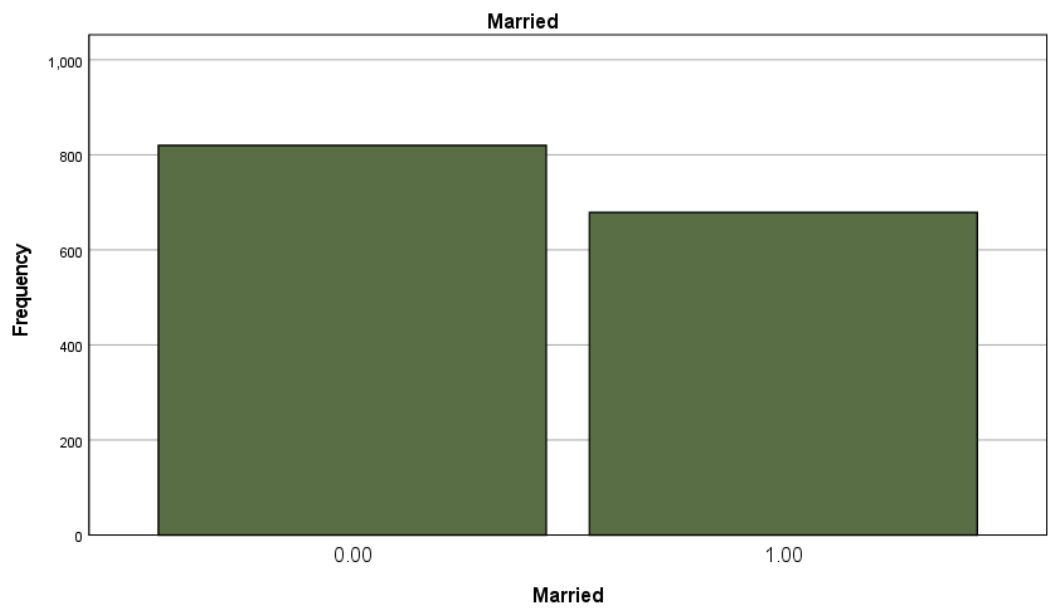
IV: Martial

Descriptive Statistics

	Cases	Cases Missing	Mode	Median	Mean	Standard Deviation	Range
Health	990	510	2	2	2.14	.841	3
Class	1493	7	2	2	2.37	.695	3
Race	1487	13	1	1	1.57	1.41	6
Sex	1500	0	0	-	-	-	-
Martial	1499	1	0	-	-	-	-

Sources: GSS14SSDS-A

Health:
Health (ordinal) variables coded 1 through 4, 1 "excellent", 2 "good", 3 "fair", 4 "poor". The median, mean, and mode are both 2 (good). The histogram of health shows a slight skewness (.432) dragged out on the right side. 22.2% of people self-reported their health as Excellent. 48.2% self-reported health as good. 22.5% self-reported health as fair and 7% indicated their health was poor.

Social Class:
Social class (ordinal) variable is coded 1 through 5. 1 is lower class, 2 working class, 3 middle class, and 4 upper class. The median and mode is 2 (working class) and the mean is 2.37. The distribution is relatively close to normal symmetrical distribution with very minimum skewness of (-.164).

Race:
Race (nominal) variable is coded 1through 7. 1 is coded as white, 2 is black or African American, 3 is American Indian or Alaska native, 4 is asian, 5 is pacfic islander, 6 is some other race, and 7 is hispanic. The median and mode is 1 (white) and the mean is 1.57. The highest reported race from respondents is white with 75% reporting. The next highest reported race is Black with 14.5%. The distribution has a positive skewed by (3.002).

Sex:
Sex (Nominal) variable is dichotomous. It is coded as 0 and 1. 0 is male and 1 is female. The mode is 0. 55.2% of respondents are male and 44.8% are female.

Martial:
Marital status (nominal) is a dichotomous variable. It is coded as 0 and 1. 0 is Non-married and 1 is married. The mode is 0. 54.7% of respondents are non-married and 45.3% of respondents are married.

Part 3: Estimation

Confidence Intervals

	Mean	Proportion	95% Confidence Interval	
			Lower	Upper
Health	2.14	-	2.09	2.19
Class	2.37	-	2.33	2.4
Black	-	0.1447	0.1280	0.1639
Female	-	0.552	0.5268	0.5772
Married	-	0.453	0.4277	0.4782

Sources: GSS14SSDS-A

Health: Based on the random sample of 990 respondents, I estimate at 95% confidence level that reported conditions of health average between 2.09 to 2.19 on a scale from poor health (4) to Excellent health (1)

Social Class: Based on the random sample of 1493 respondents, I estimate at 95% confidence level that social class average between a 2.33 and 2.4 on a scale from 1 (lower class) to 4 (Upper class).

Race: Based on the random sample of 1500 respondents, I estimate at 95% confidence level that between 12.80% and 16.39% will be Black.

Sex:

Marital Status: Based on a random sample of 1499 respondents, I estimate at 95% confidence level that between 42.77% and 47.82% of respondents will be married

Sex: Based on a random sample of 1500 respondents, I estimate at 95% confidence level that between 52.68% and 57.72% of respondents will be female.

Part Four: Significance Testing

A. Two-Sample t-Test

Two Sample t-Test

Sex	N	Mean	t-score	sig.
Male	443	2.13	-.353	.799
Female	547	2.15	-.353	

Sources: GSS14SSDS-A

H0: There is no difference in conditions of health reporting between gender.
HA: There is a difference in conditions of health reporting between gender.

The mean respondents between gender reporting condition of health is 2.13 for Male and 2.15 for Females. The alpha used for this test was .95. The test is not statistically significant, therefore we accept the null hypothesis. We conclude that gender does not have a difference in the condition of health reporting.

B. Chi Square

Marital Cross Tabulation

		Marital Status		
		Nonmarried	Married	Total
Female 1.00	Count	481	346	827
	% within Marital	58.65%	50.96%	55.17%
Male .00	Count	339	333	672
	% within Marital	41.54%	49.04%	44.83%
Total	Count	820	679	1499
	% within Marital	100.0%	100.0%	100.0%

Sources: GSS14SSDS-A

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.907[a]	1	.003		
Continuity Correction[b]	8.598	1	.003		
Likelihood Ratio	8.908	1	.003		
Fisher's Exact Test				.003	.002
Linear-by-Linear Association	8.901	1	.003		
N of Valid Cases	1499				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 304.39.

b. Computed only for a 2x2 table

H0: There is no relationship between gender and marital status
HA: There is a relationship between gender and marital status

Gender and marital status are significantly correlated. We can reject the null hypothesis. According to the column percentages, women tend to be unmarried in status more than men. 58.65% of women are unmarried while 41.54% of men are unmarried.

C: ANOVA

Descriptives

CONDITION OF HEALTH

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LOWER CLASS	90	2.60	1.026	.108	2.39	2.81	1	4
WORKING CLASS	451	2.22	.793	.037	2.14	2.29	1	4
MIDDLE CLASS	419	1.97	.809	.040	1.89	2.05	1	4
UPPER CLASS	26	1.85	.613	.120	1.60	2.09	1	3
Total	986	2.14	.840	.027	2.09	2.19	1	4

ANOVA

CONDITION OF HEALTH

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	35.895	3	11.965	17.820	.000
Within Groups	659.346	982	.671		
Total	695.241	985			

H0: There is no difference between social class and reported condition of health.
Ha: There is a difference between social class and reported condition of health.

There is a significant difference between reported condition of health and social class. We can reject the null hypothesis because at least one group is significantly different from the others. The sample means show that reported conditions of health is around "Good" (coded as 2) for lower and working class but the trend follows as one moves up the social class hierarchy the mean is decreasing getting closer to the variable of "Excellent" (coded as 1). This indicates a statistically significant difference and can be concluded that the upper class has the best condition of health.

Part Five: Analyzing Strength of Relationships

Correlation Matrix

		Condition of Health	Class	Race	Sex	Marital Status
Condition of Health	Pearson's	1	-.222	.046	.011	-.115
	Sig.	-	.000	.149	.724	.000
	N	990	986	981	990	989
Class	Pearson's	-.222	1	-.101	-.009	.180
	Sig.	.000	-	.000	.720	.000
	N	986	1493	1480	1493	1492
Race	Pearson's	.046	-.101	1	.004	-.063
	Sig.	.149	.000	-	.866	.015
	N	981	1480	1487	1487	1486
Sex	Pearson's	.011	-.009	.004	1	-.077
	Sig.	.724	.720	.866	-	.003
	N	990	1493	1487	1500	1499
Marital	Pearson's	-0.115	.180	-.063	-.077	1
	Sig.	.000	.000	.015	.003	-
	N	989	1492	1486	1499	1499

Sources: GSS14SSDS-A

There is a significant correlation in the matrix above that is bolded. There is a significant positive relationship between marital status and social class. There is also a significant negative relationship between variables. Marital status and conditions of health, marital status and race, and maraital status and sex. From the matrix we also see negative relationships with race and class. Lastly, the condition of health and class also has a negative relationship. While none of these relationships are particularly strong in correlation, the strongest correlation is a negative relationship between class and conditions of health (-0.222).

Bivariate Regression: Conditions of Health and Social Class

Variable	b	SE	Beta	P-value
(Constant)	2.786	0.094		0.000
Social Class	-0.272	0.038	-0.222	0.000
R- Squared	0.049			
N	986			

Sources: GSS14SSDS-A

Bivariate Regression: Condition of Health and Female

Variable	b	SE	Beta	P-value
(Constant)	2.131	0.040		0.000
Female	0.019	0.054	0.011	0.724
R-Squared	0.000			
N	1500			

Sources: GSS14SSDS-A

Bivariate Regression: Condition of Health and Married

Variable	b	SE	Beta	P-value
(Constant)	0.233	0.037		0.000
Married	-0.194	0.053	-0.115	0.000
R-squared	0.013			
N	1499			

Sources: GSS14SSDS-A

Bivariate Regression: Conditions of Health and Race

Variable	b	SE	Beta	P-value
(Constant)	2.130	0.031		0.000
Black	.017	.077	.007	0.828
Native American	-0.05	0.213	-0.001	0.980
Asian	-0.096	-0.159	-0.019	0.548
Pacific Islander	-0.130	0.842	-0.005	0.877
Other	0.370	0.596	0.020	0.535
Hispanic	0.225	0.129	0.056	0.081
R-squared	0.004			
N	1487			

Sources: GSS14SSDS-A

There is not a significant correlation between health and female. N0r is there a significant correlation between health and race from p-value. There is a significant negative correlation between health and marriage. The beta of (-0.115) indicates that as one becomes married the association of having better reported conditions of health are likely. We also see a similar significant negative correlation between conditions of health and social class. Beta of (-0.222) indicates that individuals with higher social class hierarchy, have higher reported conditions of health is towards excellence.

*** Reminder that conditions of health is coded as 1 "Excellent" 2 "Good" 3- "Fair" 4 "Poor" that contributes to the negative significant correlation in married and social class **

Multivariate Regression

Variable	b	SE	Beta	P-value
(Constant)	2.798	0.104		0.000
Female	-0.005	0.053	-0.003	0.924
Married	**-0.128**	**0.055**	**-0.076**	**0.019**
Social Class	**-0.249**	**.039**	**-0.203**	**0.000**
Black	-0.062	0.077	-0.026	0.417
R-Squared	0.055			
N	1500			

Sources: GSS14SSDS-A

In a multivariate regression, the significance of the variables remained the same however the strength of the significant correlation slightly decreased. The bolded variables represent the significance. Combined, these variables account for 5.5% of the variation in condition of health.

Conclusion

After analyzing gender, marital status, social class, and race with the relationship to condition of health there were significant statistical findings. Although my analysis regarding race and gender was insignificant, my findings supported my third and fourth research hypotheses on marital and social class variables. H3: On average, high or middle-class groups will have higher rates of "excellent or good health" and H4: On average, those that are married will perceive health conditions as good. These findings were significant through the statistical analysis with chi-square, ANOVA, bivariate and multivariate regressions. The findings prove and disprove some associated factors regarding conditions of health. I think it is highly interesting the statistical findings of marriage and conditions of health. Marriage serves as a protective factor in health and can help boost healthy living habits that prevent or reduce the risk of chronic diseases. It was also surprising to see that no significant statistical findings in race and condition of health were present. I think with an in-depth analysis of specified health variables such as weight, diet, teen pregnancies, tobacco usage, and domestic violence for example, we can identify which ethnic and racial demographics have a higher or lower relationships. However, I think it is important to note that the variable of health is self-reported and thus could reflect not the true conditions of health in individuals.