

Final Exam Supplemental: Outlook on Life Survey

1 Summary

The 2012 Outlook Surveys, conducted by GfK Knowledge Networks on behalf of the University of California Irvine, were designed to study political and social attitudes in the United States. The project included two surveys fielded between August and December 2012 using a sample from an Internet panel. A total of 2,294 respondents participated in this study during Wave 1 and 1,601 were interviewed during Wave 2.

The target population was comprised of four groups: African American/Black males aged 18 and older, African American/Black females aged 18 and older, White/other race males aged 18 and older, and White/other race females aged 18 older, all non-institutionalized and residing in the United States.

The survey considered the ways in which social class, ethnicity, marital status, feminism, religiosity, political orientation, sexual behavior, and cultural beliefs or stereotypes influence opinion and behavior. Participants were asked an array of questions pertaining to voting preference, party identification, respondent perception of opportunity for success, and views on interracial dating. These variables and questions examine political and social attitudes in the United States. Additional questions addressed issues such as common fate, nationalism, equality, discrimination, and relations with law enforcement. Demographic variables include race ethnicity, age, gender, religious involvement, sexual orientation, citizenship, annual income, and education.

Study Purpose

The purpose of the 2012 Outlook Surveys were to study political and social attitudes in the United States. The specific purpose of the survey is to consider the ways in which social class, ethnicity, marital status, feminism, religiosity, political orientation, and cultural beliefs or stereotypes influence opinion and behavior.

Sample

Participants were drawn from the GfK Knowledge Network, a web panel designed to be representative of the United States population. Panel members are randomly recruited through probability-based sampling, and households are provided with access to the Internet and hardware if needed. Random-digit dialing and address-based sampling methodologies are used. The target population were non-institutionalized adults 18 years of age and older.

Mode of Data Collection & Response Rates

Web-based survey. Wave 1: 55.3 percent response rate. Wave 2: 75.1 percent response rate.

Description of Variables

The data include variables pertaining to social class, ethnicity, marital status, feminism, religiosity, and political orientation. In addition there are variables pertaining to gender, household composition and size, household income, employment status, education, and marital status.

2 Data Management

Demographics

```
pol$gender <- factor(pol$ppgender, labels=c("Male", "Female"))
pol$ppethm[pol$ppethm==5] <- 3
pol$eth <- factor(pol$ppethm, labels=c("W-NH", "B-NH", "O-NH", "H"))
pol$educat <- factor(pol$ppeducat, labels=c("Lt HS", "HS", "Some college", "BS +"))
pol$region <- factor(pol$ppreg4, labels=c("NE", "MW", "S", "W"))
```

Who did you vote for?

```
pol$w1_a5a[pol$w1_a5a %in% c(-1, 3)] <- NA
pol$whovotefor <- factor(pol$w1_a5a, labels=c("McCain", "Obama"))
```

How interested are you in what's going on in government and politics?

```
pol$w1_a1[pol$w1_a1 == -1] <- NA
pol$lv1_political_interest <- factor(pol$w1_a1,
  labels=c("Extremely", "Very", "Moderately", "Slightly", "Not at all"))
```

How many days in the past week did you watch national news programs on television or on the Internet?

```
pol$w1_a11[pol$w1_a11 == -1] <- NA
pol$nday_watch_news <- pol$w1_a11 - 1
```

Is anyone in your household currently unemployed?

```
pol$w1_p11[pol$w1_p11 == -1] <- NA
pol$any_unemp <- ifelse(pol$w1_p11 == 1, "Yes", "No")
```

People talk about social classes such as the poor, the working class, the middle class, the upper-middle class, and the upper class. Which of these classes would you say you belong to?

```
pol$w1_p2[pol$w1_p2 == -1] <- NA
pol$w1_p2[pol$w1_p2 == 5] <- 4
pol$class <- factor(pol$w1_p2, labels=c("Poor", "Working", "Middle", "Above Middle"))
```

A basic American belief has been that if you work hard you can get ahead and reach the goals you set and more. Is this true or false today? (Scale from 1-7, 1: Extremely true, 7: Extremely false)

```
pol$w1_f3[pol$w1_f3 == -1] <- NA
pol$work_hard <- pol$w1_f3 - 1
```

How well does the U.S. Congress represent you?

```
pol$w1_i2[pol$w1_i2 == -1] <- NA
pol$congr_repr <- factor(pol$w1_i2, labels = c("Extremely", "Very", "Moderately",
  "A little", "Not at all"))
```

Please rate each [group/person] using the feeling thermometer (0-100, 0:Cold/unfavorable, 100:warm/favorable). Groups chosen: Obama (`rate_obama`), McCain (`rate_mccain`), Republican Party (`rate_rep`), Democrat Party (`rate_dem`), the Wealthiest 1% (`rate_onepct`), the unemployed (`rate_unemp`)

```
pol[,6:11][pol[,6:11] < 0 | pol[,6:11] > 100] <- NA
names(pol)[6:11] <- c("rate_obama", "rate_mccain", "rate_rep", "rate_dem",
  "rate_onepct", "rate_unemp")
```

3 Univariate Descriptions

Table 1: Summary statistics for numerical (and 7-level ordinal) variables

	N	mean	median	sd	min	max
nday_watch_news	2283.0	3.2	3.0	2.6	0.0	7.0
work_hard	2250.0	2.1	2.0	1.7	0.0	6.0
rate_obama	2190.0	67.5	85.0	33.4	0.0	100.0
rate_mccain	2096.0	31.4	30.0	29.2	0.0	100.0
rate_rep	2116.0	32.8	30.0	27.6	0.0	100.0
rate_dem	2145.0	61.8	70.0	29.2	0.0	100.0
rate_onepct	2033.0	40.6	50.0	27.9	0.0	100.0
rate_unemp	2092.0	60.3	60.0	23.9	0.0	100.0

Table 2: Distributions of categorical variables.

	N	%
Gender		
Male	1032	45%
Female	1262	55%
Education		
Lt HS	219	9.5%
HS	700	30.5%
Some college	682	29.7%
BS +	693	30.2%
Ethnicity		
White, Non-Hispanic (W-NH)	814	35.5%
Black, Non-Hispanic (B-NH)	1278	55.7%
Other, Non-Hispanic (O-NH)	79	3.4%
Hispanic (H)	123	5.4%
Region		
Northeast (NE)	398	17.3%
Midwest (MW)	499	21.8%
South (S)	994	43.3%
West (W)	403	17.6%
Voted for		
McCain	389	22.5%
Obama	1338	77.5%
Level of Political Interest		
Extremely Interested	389	17.1%
Very interested	679	29.8%
Moderately interested	658	28.9%
Slightly interested	354	15.5%
Not interested at all	198	8.7%
Any unemployed in house		
No	1365	61.2%
Yes	867	38.8%
Class		
Poor	258	11.7%
Working	755	34.2%
Middle	955	43.3%
Above Middle	239	10.8%

Figure 1: Number of days past week watched news.

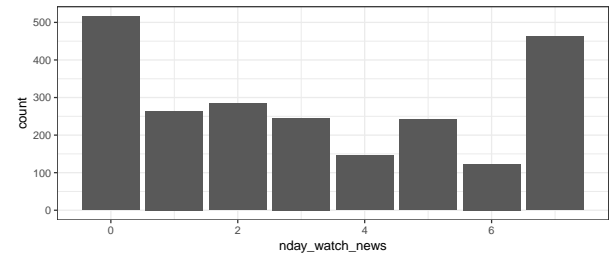


Figure 2: Believe scale about working hard lets you achieve goals.

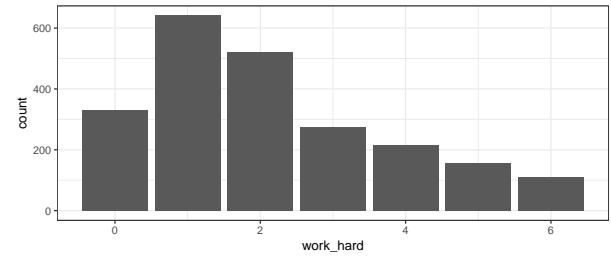
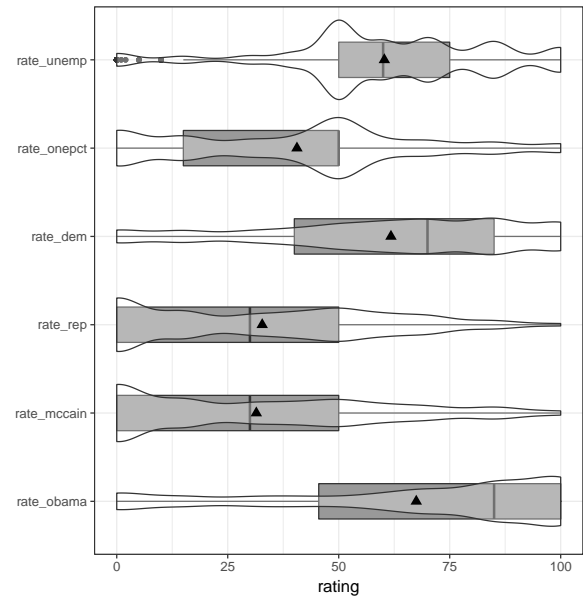


Figure 3: Distribution of Feeling Thermometer rating variables.



4 Analyses

4.1 Bivariate Comparisons

- by Ethnicity.

Table 3: Frequency table

	McCain	Obama	NA
W-NH	332	278	204
B-NH	21	977	280
O-NH	16	34	29
H	20	49	54

Table 4: Row proportions

	McCain	Obama	NA
W-NH	40.8	34.2	25.1
B-NH	1.6	76.4	21.9
O-NH	20.3	43.0	36.7
H	16.3	39.8	43.9

Pearson's Chi-squared test

```
data: table(pol$eth, pol$whovotefor)
X-squared = 598.44, df = 3, p-value < 2.2e-16
```

- by how well you feel Congress represents you.

Table 5: Frequency table

	McCain	Obama
Extremely	4	10
Very	4	36
Moderately	74	348
A little	145	408
Not at all	159	514

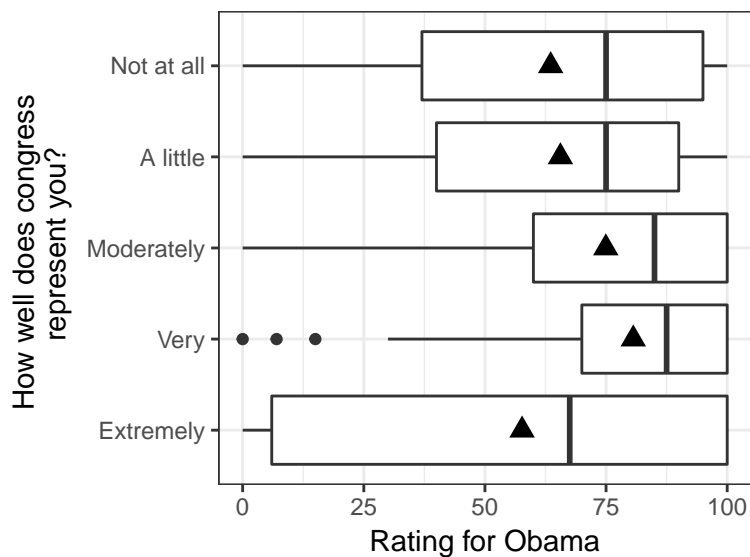
Table 6: Row proportions

	McCain	Obama
Extremely	28.6	71.4
Very	10.0	90.0
Moderately	17.5	82.5
A little	26.2	73.8
Not at all	23.6	76.4

Pearson's Chi-squared test

```
data: table(pol$congr_repr, pol$whovotefor)
X-squared = 14.61, df = 4, p-value = 0.005582
```

Figure 4: Congress vs. Obama



4.2 Modeling of the ratings for John McCain using a linear regression model

Table 7: Model 1

	Est	CI	pvalue
rate_onepct	0.11	(0.08,0.15)	<0.001
rate_rep	0.77	(0.74,0.81)	<0.001
rate_unemp	-0.02	(-0.05,0.02)	0.3649
any_unempYes	-1.33	(-3.03,0.36)	0.1238
genderFemale	-2.71	(-4.35,-1.07)	0.0012

Model 1 has a coefficient of determination of 0.6179.

Table 8: Model 2

	Est	CI	pvalue
rate_onepct	0.11	(0.08,0.14)	<0.001
rate_rep	0.77	(0.74,0.8)	<0.001
rate_unemp	-0.02	(-0.05,0.02)	0.3290
any_unempYes	-0.91	(-2.67,0.84)	0.3072
genderFemale	-2.56	(-4.22,-0.91)	0.0024
classWorking	1.95	(-1.07,4.98)	0.2061
classMiddle	2.58	(-0.41,5.56)	0.0906
classAbove Middle	4.75	(1.06,8.43)	0.0117

Model 2 has a coefficient of determination of 0.6174.
The Wald test for the categorical variable **class** resulted in a $F_{3,1838} = 2.26$, $p = 0.079187$.

Table 9: Model 3

	Est	CI	pvalue
rate_onepct	0.13	(0.1,0.16)	<0.001
rate_rep	0.71	(0.68,0.75)	<0.001
rate_unemp	0.00	(-0.04,0.04)	0.9897
any_unempYes	-0.60	(-2.27,1.08)	0.4858
genderFemale	-2.35	(-3.97,-0.74)	0.0043
ethB-NH	-7.94	(-9.83,-6.06)	<0.001
ethO-NH	0.40	(-4.28,5.07)	0.8679
ethH	-3.36	(-7.35,0.63)	0.0984

Model 3 has a coefficient of determination of 0.6317.
The Wald test for the categorical variable **eth** resulted in a $F_{3,1851} = 24.2$, $p = < 0.001$.

4.3 Logistic Regression

Table 10: β estimates, their standard errors, test statistics and p-values for each coefficient in the logistic regression model on the independent variable `whovotefor`.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.11	0.27	4.15	0.00
rate_onepct	-0.03	0.00	-10.68	0.00
rate_unemp	0.02	0.00	6.67	0.00
nday_watch_news	-0.01	0.03	-0.30	0.76
lvl_political_interestVery	0.16	0.17	0.91	0.37
lvl_political_interestModerately	0.38	0.19	1.96	0.05
lvl_political_interestSlightly	0.37	0.25	1.44	0.15
lvl_political_interestNot at all	0.22	0.37	0.60	0.55

5 Examples of R code and output for SPSS Users

These are not on the exam, but presented as examples of how modeling output from R appears (for cases where a nice table is not easily created). For these I am using a data set on characters in the Star Wars universe.

```
starwars$bald <- ifelse(starwars$hair_color == "none", "no hair", "has hair")
starwars$isbald <- ifelse(starwars$hair_color=="none", 1, 0)
```

Two sample T-tests for a mean Droids tend to both be tall, and have no hair. On average, are individuals with no hair taller than those with hair?

```
t.test(height~bald, data=starwars)
```

Welch Two Sample t-test

```
data: height by bald
t = -1.065, df = 64.31, p-value = 0.2909
alternative hypothesis: true difference in means
95 percent confidence interval:
 -23.907438  7.279661
sample estimates:
mean in group has hair mean in group no hair
172.5750 180.8889
```

Two sample tests for a difference in proportions Is the proportion of bald individuals higher for males? – Still can use a *t.test*, but the outcome must be a binary 0/1 indicator variable.

```
t.test(isbald~gender, data=starwars)
```

Welch Two Sample t-test

```
data: isbald by gender
t = -1.2556, df = 31.491, p-value = 0.2185
alternative hypothesis: true difference in means is not
95 percent confidence interval:
 -0.41872574  0.09948502
sample estimates:
mean in group female mean in group male
0.3157895 0.4754098
```

Correlation Is there a significant correlation between height and mass? The point estimate is at the bottom under *cor*: $r = 0.13$

```
cor.test(starwars$height, starwars$mass)
```

Pearson's product-moment correlation

```
data: starwars$height and starwars$mass
t = 1.02, df = 57, p-value = 0.312
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1265364  0.3770395
sample estimates:
cor
0.1338842
```