Hypothesis Testing

Jason Neumeyer

March 4, 2020

Load Data and Packages -

```
## Packages

##install.packages("here")
library(here)
##install.packages("readr")
library(readr)
library(tidyverse)
library(ggplot2)
##install.packages("expss")
library(expss)
##install.packages("broom")
library(broom)
##install.packages("purrr")
library(purrr)
##install.packages("stargazer")
library(stargazer)
library(lmtest)
library(MASS)
library(car)

setwd("C:/Users/jason/Desktop/UW-Milwaukee Graduate Year 2/Lab Meeting/Data")

dat <- read_csv("immigration_20191219_clean.csv")

## View(dat)
```

Preference Variable -

```
tv_prefer <- dat$tv_msnbc - dat$tv_fox
dat["tv_prefer"] <- tv_prefer
```

Hypothesis Testing -

H2a-c

```
libimm_data <- dat %>% dplyr::select(immig_increased, taxes_pos, jobs_pos, condition, sales_correct, em
libimm_data$condition <- as.factor(libimm_data$condition)
```

```r
libimm <- rowMeans(subset(libimm_data, select = c(immig_increased, taxes_pos, jobs_pos)), na.rm = TRUE)
libimm_data <- cbind(libimm_data, libimm)
libimm_data$libimm <- as.numeric(libimm_data$libimm)


libimm_data$condition <- relevel(libimm_data$condition, ref = "control")

##View(libimm_data)
h1a <- lm(libimm ~ condition, data = libimm_data)
summary(h1a)
```

```
##
## Call:
## lm(formula = libimm ~ condition, data = libimm_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60338 -0.13709  0.02919  0.19662  0.47844
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.52156    0.01734  30.085  < 2e-16 ***
## conditionassigned  0.08182    0.02427   3.371 0.000798 ***
## conditionchoice    0.05744    0.02421   2.372 0.018005 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2415 on 597 degrees of freedom
## Multiple R-squared:  0.01959,    Adjusted R-squared:  0.01631
## F-statistic: 5.966 on 2 and 597 DF,  p-value: 0.002721
```

```r
h1b <- lm(sales_correct ~ condition, data = libimm_data)
summary(h1b)
```

```
##
## Call:
## lm(formula = sales_correct ~ condition, data = libimm_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5245 -0.4604 -0.2680  0.4755  0.7320
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.26804    0.03466   7.734 4.46e-14 ***
## conditionassigned  0.19235    0.04853   3.964 8.27e-05 ***
## conditionchoice    0.25647    0.04841   5.298 1.65e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4827 on 597 degrees of freedom
## Multiple R-squared:  0.04815,    Adjusted R-squared:  0.04496
## F-statistic:  15.1 on 2 and 597 DF,  p-value: 4.007e-07
```

```r
h1c <- lm(employ_correct ~ condition, data = libimm_data)
summary(h1c)
```

```
##
## Call:
## lm(formula = employ_correct ~ condition, data = libimm_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4555 -0.4216 -0.1340  0.5445  0.8660
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.13402    0.03250   4.123 4.26e-05 ***
## conditionassigned  0.32142    0.04551   7.063 4.56e-12 ***
## conditionchoice    0.28755    0.04540   6.334 4.71e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4527 on 597 degrees of freedom
## Multiple R-squared:  0.09121,    Adjusted R-squared:  0.08816
## F-statistic: 29.96 on 2 and 597 DF,  p-value: 3.994e-13
```

```r
stargazer(h1a, h1b, h1c, dep.var.labels = c("Liberal Immigration Level", "Accurate Sale Numbers", "Accu
```

Table 1: Control vs. Free Choice vs. Assigned

	Dependent variable:		
	Liberal Immigration Level	Accurate Sale Numbers	Accurate Employment Numbers
	(1)	(2)	(3)
Assigned Condition	0.082***	0.192***	0.321***
	(0.024)	(0.049)	(0.046)
Free Choice Condition	0.057**	0.256***	0.288***
	(0.024)	(0.048)	(0.045)
Constant	0.522***	0.268***	0.134***
	(0.017)	(0.035)	(0.033)
Observations	600	600	600
R^2	0.020	0.048	0.091
Adjusted R^2	0.016	0.045	0.088
Residual Std. Error (df = 597)	0.241	0.483	0.453
F Statistic (df = 2; 597)	5.966***	15.100***	29.958***
Note:			*p<0.1; **p<0.05; ***p<0.01

```r
lh1 <- linearHypothesis(h1a, c("conditionassigned = conditionchoice"))

lh2 <- linearHypothesis(h1b, c("conditionassigned = conditionchoice"))

lh3 <- linearHypothesis(h1c, c("conditionassigned = conditionchoice"))
```

```
stargazer(lh1, header = FALSE, title = "Liberal Immigration Level")
```

Table 2: Liberal Immigration Level

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Res.Df	2	597.500	0.707	597	597.2	597.8	598
RSS	2	34.839	0.043	34.809	34.824	34.854	34.869
Df	1	1.000		1.000	1.000	1.000	1.000
Sum of Sq	1	0.060		0.060	0.060	0.060	0.060
F	1	1.035		1.035	1.035	1.035	1.035
Pr(>F)	1	0.309		0.309	0.309	0.309	0.309

```
stargazer(lh2, header = FALSE, title = "Sales Correct")
```

Table 3: Sales Correct

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Res.Df	2	597.500	0.707	597	597.2	597.8	598
RSS	2	139.331	0.295	139.122	139.227	139.435	139.540
Df	1	1.000		1.000	1.000	1.000	1.000
Sum of Sq	1	0.417		0.417	0.417	0.417	0.417
F	1	1.790		1.790	1.790	1.790	1.790
Pr(>F)	1	0.181		0.181	0.181	0.181	0.181

```
stargazer(lh3, header = FALSE, title = "Employ Correct")
```

Table 4: Employ Correct

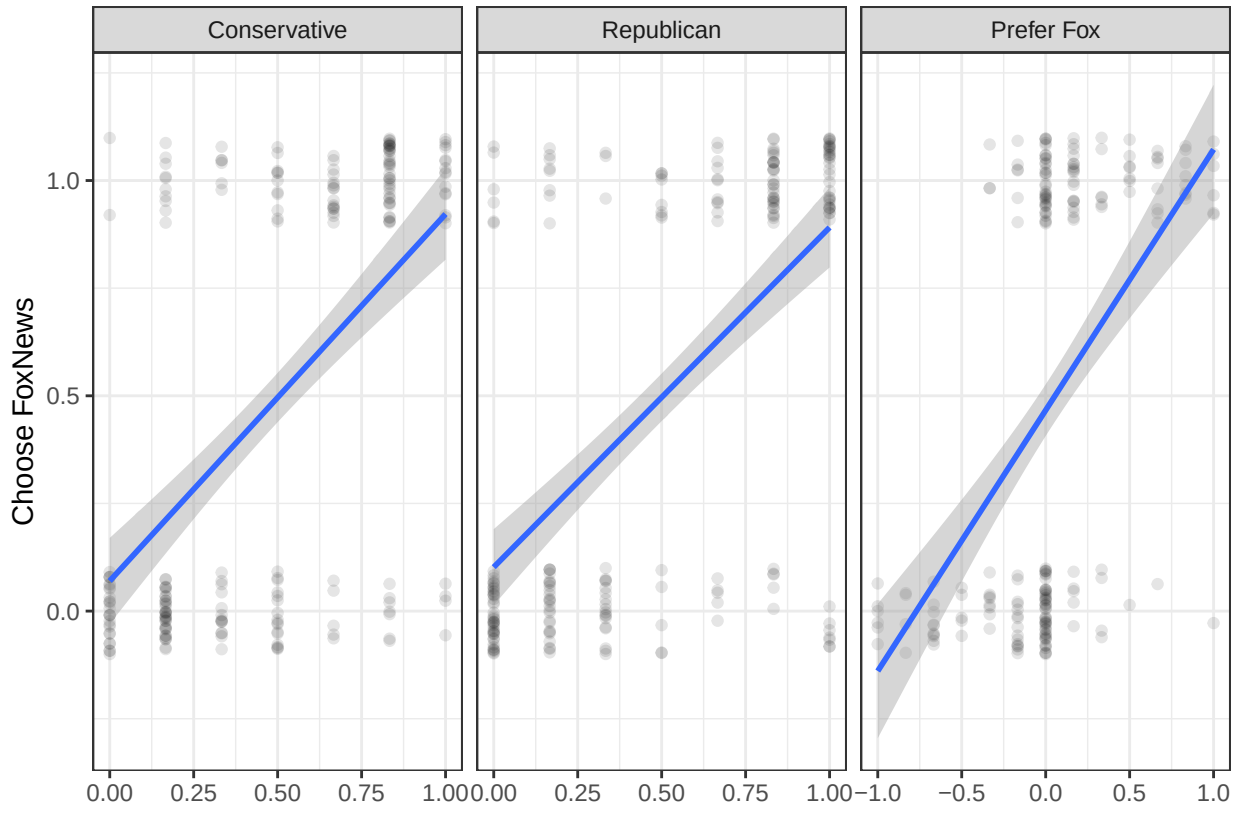
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Res.Df	2	597.500	0.707	597	597.2	597.8	598
RSS	2	122.418	0.082	122.360	122.389	122.447	122.476
Df	1	1.000		1.000	1.000	1.000	1.000
Sum of Sq	1	0.116		0.116	0.116	0.116	0.116
F	1	0.568		0.568	0.568	0.568	0.568
Pr(>F)	1	0.451		0.451	0.451	0.451	0.451

H1a-c

```
## Choose Fox

labels_FOX <- c(ideol_con = "Conservative", pid_rep = "Republican", prefer_fox = "Prefer Fox")

choose_fox <- dat %>%
  filter(condition == "choice") %>%
  mutate(prefer_fox = tv_fox - tv_msnbc,
         choose_fox = as.numeric(tweet == "fox")) %>%
  dplyr::select(choose_fox, ideol_con, pid_rep, prefer_fox) %>%
  gather(variable, value, -choose_fox) %>%
```

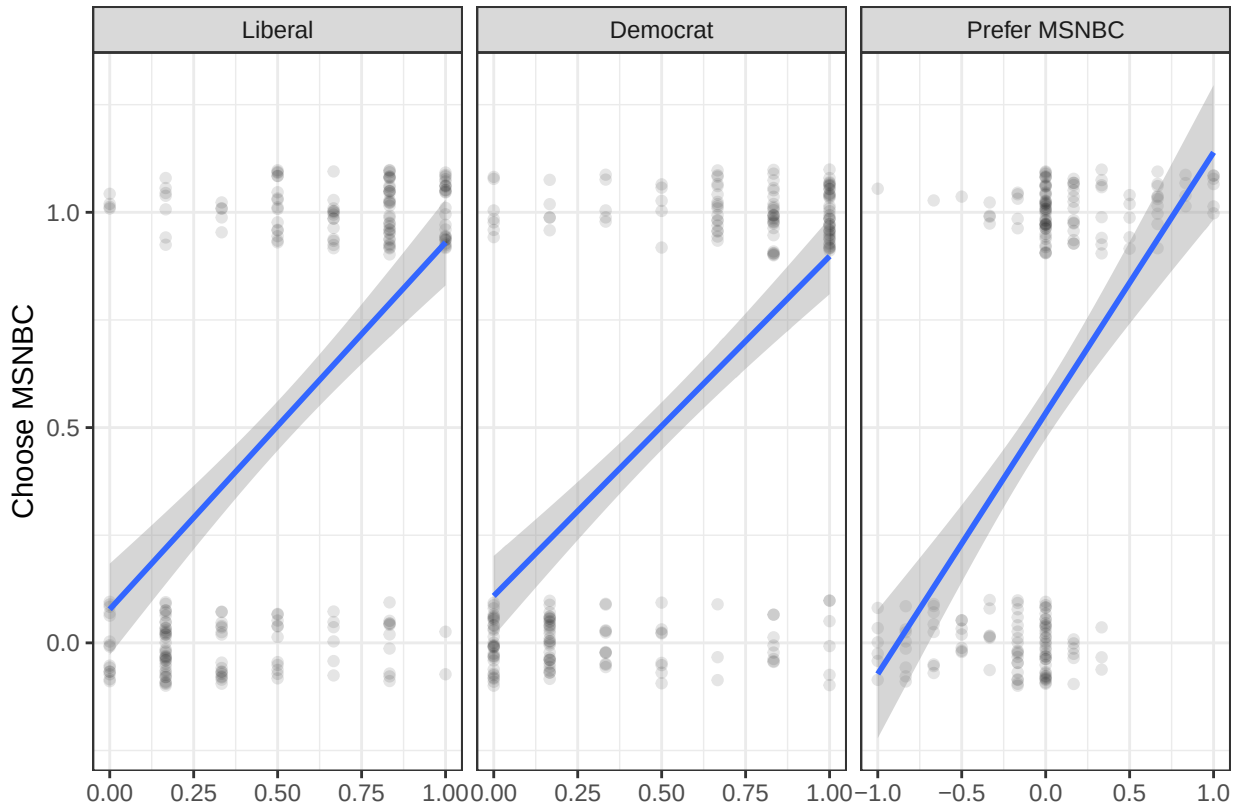
```
  ggplot(aes(x=value, y = choose_fox)) +
  geom_jitter(alpha = .1, height = .1) + geom_smooth(method = "lm") +
  facet_grid(~variable, scales = "free_x", labeller = labeller(variable = labels_FOX)) +
  theme_bw() + labs(y = "Choose FoxNews", x = "")

## Choose MSNBC

ideol_lib <- 1 - dat$ideol_con
dat["ideol_lib"] <- ideol_lib

pid_dem <- 1 - dat$pid_rep
dat["pid_dem"] <- pid_dem

labels_MSNBC <- c(ideol_lib = "Liberal", pid_dem = "Democrat", prefer_msnbc = "Prefer MSNBC")

choose_msnbc <- dat %>%
  filter(condition == "choice") %>%
  mutate(prefer_msnbc = tv_msnbc - tv_fox,
         choose_msnbc = as.numeric(tweet == "msnbc")) %>%
  dplyr::select(choose_msnbc, ideol_lib, pid_dem, prefer_msnbc) %>%
  gather(variable, value, -choose_msnbc) %>%
  ggplot(aes(x=value, y = choose_msnbc)) +
  geom_jitter(alpha = .1, height = .1) + geom_smooth(method = "lm") +
  facet_grid(~variable, scales = "free_x", labeller = labeller(variable = labels_MSNBC)) +
  theme_bw() + labs(y = "Choose MSNBC", x = "")

choose_fox
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 4 rows containing non-finite values (stat_smooth).

## Warning: Removed 4 rows containing missing values (geom_point).
```




```
choose_msnbc
```

```
## `geom_smooth()` using formula 'y ~ x'
## Warning: Removed 4 rows containing non-finite values (stat_smooth).
## Warning: Removed 4 rows containing missing values (geom_point).
```




```
## demographic model for choose FoxNews

choose_fox = as.numeric(dat$tweet == "fox")
prefer_fox <- dat$tv_fox - dat$tv_msnbc

test <- lm(choose_fox ~ prefer_fox + college + white + age, data = dat)
summary(test)
```

```
##
## Call:
## lm(formula = choose_fox ~ prefer_fox + college + white + age,
##     data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6866 -0.3425 -0.2461  0.5553  0.9524
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.437e-01  7.411e-02   4.638 4.34e-06 ***
## prefer_fox   3.045e-01  4.648e-02   6.551 1.25e-10 ***
## collegeTRUE  4.097e-02  3.784e-02   1.083    0.279
## whiteTRUE   -3.994e-02  4.710e-02  -0.848    0.397
## age         -4.949e-05  1.638e-03  -0.030    0.976
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.4557 on 585 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.06941,    Adjusted R-squared:  0.06305
## F-statistic: 10.91 on 4 and 585 DF,  p-value: 1.549e-08
```

```
test2 <- lm(choose_fox ~ ideol_con + college + white + age, data = dat)
summary(test2)
```

```
##
## Call:
## lm(formula = choose_fox ~ ideol_con + college + white + age,
##     data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5515 -0.3260 -0.2244  0.5213  0.8459
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.700e-01  7.652e-02   2.221   0.0267 *
## ideol_con    3.430e-01  5.909e-02   5.805 1.05e-08 ***
## collegeTRUE  4.109e-02  3.801e-02   1.081   0.2801
## whiteTRUE   -1.259e-02  4.693e-02  -0.268   0.7886
## age         -8.422e-05  1.649e-03  -0.051   0.9593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4597 on 591 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.05539,    Adjusted R-squared:  0.049
## F-statistic: 8.664 on 4 and 591 DF,  p-value: 8.442e-07
```

```
test3 <- lm(choose_fox ~ pid_rep + college + white + age, data = dat)
summary(test3)
```

```
##
## Call:
## lm(formula = choose_fox ~ pid_rep + college + white + age, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5392 -0.3282 -0.2181  0.5204  0.8241
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.000e-01  7.518e-02   2.660  0.00801 **
## pid_rep      3.038e-01  5.060e-02   6.004 3.34e-09 ***
## collegeTRUE  3.124e-02  3.777e-02   0.827  0.40848
## whiteTRUE   -2.669e-02  4.699e-02  -0.568  0.57028
## age          8.277e-05  1.636e-03   0.051  0.95966
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4594 on 595 degrees of freedom
```

```
## Multiple R-squared:  0.05826,    Adjusted R-squared:  0.05192
## F-statistic: 9.202 on 4 and 595 DF,  p-value: 3.223e-07
```

```
stargazer(test, test2, test3, header = FALSE, single.row = T, column.sep.width = "0pt")
```

Table 5:

	Dependent variable:		
	choose_fox		
	(1)	(2)	(3)
prefer_fox	0.304*** (0.046)		
ideol_con		0.343*** (0.059)	
pid_rep			0.304*** (0.051)
college	0.041 (0.038)	0.041 (0.038)	0.031 (0.038)
white	−0.040 (0.047)	−0.013 (0.047)	−0.027 (0.047)
age	−0.00005 (0.002)	−0.0001 (0.002)	0.0001 (0.002)
Constant	0.344*** (0.074)	0.170** (0.077)	0.200*** (0.075)
Observations	590	596	600
R^2	0.069	0.055	0.058
Adjusted R^2	0.063	0.049	0.052
Residual Std. Error	0.456 (df = 585)	0.460 (df = 591)	0.459 (df = 595)
F Statistic	10.909*** (df = 4; 585)	8.664*** (df = 4; 591)	9.202*** (df = 4; 595)
Note:			*p<0.1; **p<0.05; ***p<0.01

H3a-c