

# PS 531 Final Project - A Pre-Analysis Plan

Jane Betchley

2023-05-09

## Introduction

Framing is usually conceived of in the political science literature in terms of persuasion; that elites use frames to persuade individuals to adopt a particular point of view. However, given the increase in political sorting (Levendusky, 2009) and mass polarization (Abramowitz & Saunders, 2008; Abramowitz & Webster, 2016; Barber & McCarty, 2015; McCarty et al., 2006; Webster & Abramowitz, 2017) in recent years, if frames existed solely, or even primarily, for the purposes of persuasion, then we would expect to see a lot less of them in the contemporary context. In fact, highly partisan media ensures that frames are more prevalent than ever. Therefore, looking at frames only in the context of persuasion surely means we are missing a significant piece of the puzzle when it comes to understanding their effects.

In contrast, I will argue that framing's primary function in political communication is to convey messages in a way that reduces the odds of cognitive dissonance (Festinger, 1957) and readily meets the requirements of an individual's confirmation bias such that it can be absorbed without question. When that frame and the message content are accepted by an individual, their pre-existing political attitudes are reinforced, regardless of whether the information conveyed is novel or not.

In this paper, opinion reinforcement is operationalized as increased certainty. Certainty, "the degree to which an individual is certain that his or her attitude towards an object is correct." (Krosnick & Abelson 1992:180), is a very important but understudied attitude dimension. It should be made more distinct by political scientists, as if we mistake extremity or intensity of attitude for certainty, then it will be very difficult to engage in an open exchange of views and to alter our opinions.

I will test one hypothesis in this survey experiment; that attitude-congruent frames are associated with an increase in political issue attitude certainty. If issue certainty is increased among the treatment group relative to a control group which receives a neutrally framed message on the same topic, then I will conclude that the respondent's original opinion has been reinforced by the framed message. This expectation arises from previous work on confirmation bias (Taber & Lodge, 2006) and motivated reasoning (Kunda, 1990), which are the mechanisms through which this reinforcement effect is thought to occur. In addition, Turner (2007) found that neutral media is perceived as biased against the partisan in-group by partisans. This suggests that content received without a frame will be rejected and only congruent information will be accepted, thereby shoring up one's political views and underlying identity.

## Research Design and Data

The research design is a survey experiment. Participants will be randomly assigned to the treatment and control groups. Firstly, all participants will complete basic demographic and background political information questions such as on ideology, partisanship, media consumption, etc. They will

then answer a battery of policy issue questions taken from the Wilson-Patterson conservatism scale (Wilson & Patterson, 1968), and supplemented with more topical issues using questions from the most recent ANES survey (as well as items designed to match to post-treatment questions for ATE comparisons to be made, if these are needed in addition). As well as indicating their view on a policy matter, respondents will also rate their certainty about the issue, intensity (affect), extremity, the importance of the issue, and their level of knowledge about the topic. Then the experiment begins, with the control group receiving a neutral vignette and the treatment group receiving an opinion-congruent vignette based upon their partisan / ideological self-identification. Finally, they will be asked again for their issue opinion on the vignette topic, as well as some others to see if there is a general reinforcement effect or if only the one selected issue is affected (if it is).

I chose a survey experiment as this seems the best way to search for specific evidence that congruent frames do or do not reinforce political attitudes. I could have asked directly about it, but I suspect that framing effects are subtle and so people may not be aware of them occurring (if they do). As the frame is the sole difference between the treatment and control conditions, this is the purest possible test of whether a frame is having an effect on attitude certainty in the manner that is theorized.

The survey experiment is a good choice because the design allows for a direct test of the relationship of interest. Furthermore, such surveys can be carried out at a relatively low-cost online using a convenience sample. Although a nationally representative sample would be preferable, it is not realistic given my resources and so I intend to use a convenience sample.

However, the survey experiment I am planning also has some drawbacks. Firstly, choosing an appropriate topic for the vignette is proving difficult, and then I have to write appropriate neutral and reinforcement vignette content which might be tricky to get right. For example, how will I know if my treatment condition is strong enough? Would I need to run a pilot test on that? Secondly, survey experiments suffer from a lack of external validity (Barabas & Jerit, 2010) and are therefore a stylized finding regarding a phenomenon that may or may not exist outside of the lab. If the effect does persist, then the experiment likely generates a stronger estimated effect than in reality when our attention is rarely focused on political issues, or our thinking about them is less intentional.

The independent variable is whether the respondent is assigned to the treatment or control condition. The dependent variable is the change in respondent's certainty of their answer post-treatment, as compared to pre-treatment. Certainty will be measured on a 1 - 10 self-report scale, so the range of difference is -9 to 9 points (calculated as post-treatment certainty score minus pre-treatment certainty score). Because I am expecting a net positive change in certainty, respondents who indicate absolute certainty pre-treatment will need to be dropped from the analysis, as there will be no room for their score to increase. I might analyze this group separately to check whether the treatment / control conditions reduced their certainty, but they cannot be included in the analysis of the primary hypothesis.

The random assignment of participants to the control or treatment conditions means that the proposed study has a randomized design. Survey questions can generate some random noise (Zaller & Feldman, 1992), and although a self-report measure for attitude certainty appears valid (Michael Alvarez 1996), it is unclear whether respondents will truly differentiate between certainty, intensity, and extremity of the attitude in question when evaluating their attitude positions. To overcome this problem, the survey will also ask for ratings of the other four relevant attitude dimensions identified by Krosnick and Abelson (1992), so that these can be incorporated into the analysis, most likely as a covariate used in matching respondents before analysis. Missing data is a fact of life with surveys, but the experimental questions are few, so the hope is that most people will complete them. The treatment vignette will be short to encourage compliance with the treatment, so a manipulation check is probably not necessary. It is possible that participants may work out the purpose of the study but since the survey will ask about multiple attitude characteristics, respondents are unlikely to know that certainty is the one of primary interest in this study. Furthermore, if an effect does occur, it is just as likely to appear in the control and treatment groups and therefore outcomes should be comparable, even if the effect sizes estimated are inflated by social desirability. Finally, I intend

to conduct a sensitivity analysis on the estimate generated by my experiment, as this will quantify the amount of unobserved covariate bias that would render my results insignificant. Hopefully the bias estimate will be low, meaning that unobserved covariates are unlikely to bias my findings.

## Testing Randomization

The data will come from a convenience sample and as sometimes even randomly assigned experiments can generate coincidentally unbalanced groups, I will use a balance test to check that observed covariates are balanced across the treatment and control groups. If they are not, then I will group respondents into strata using optimal matching based on the Mahalanobis distance, which captures the distance between respondents in a multivariate space, thereby allowing me to match on multiple covariates. Like treatment and control subjects would be placed into strata to make sure that the treatment effect is isolated from other potential confounding variables.

I have created a fake dataset for use in this paper by randomly sampling 300 respondents from the ANES 2020 study. I then randomly assigned treatment and control conditions to the participants. I assigned an average change in certainty score of +3 to the treatment group, and no change (0) to the control group, with a standard deviation of 0.5 in the scores (see code appendix at the close of the paper for more details on what is included in that mock dataset). Below is the result of a balance test on that dataset, which checks for balance across the treatment and control groups on age, gender, race, education level, income level, and partisanship:

```

strata():
stat      Treatment Control adj.diff std.diff  z
vars
age        52.4    50.2     2.23    0.13    1.16
white      0.733    0.747    -0.0133 -0.03   -0.26
male       0.420    0.447    -0.0267 -0.05   -0.47
educ       4.77     4.64     0.133   0.07    0.58
income     11.5     12.5    -0.987  -0.14   -1.24
pid7       3.79     3.97    -0.18   -0.08   -0.67
---Overall Test---
chisquare df p.value
--      4.68  6   0.586
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The result of the overall balance test is a p-value of 0.59. In substantive terms, this means that the distribution of covariates in the treatment and control groups in the fake dataset is better than would be expected in 59% of randomized experimental designs. That is to say, they are distributed as would be expected in a randomized sample of a population. This is a satisfactory outcome, and is not surprising as my sample was randomly drawn from a random but nationally representative US sample. We know from the theory of bootstrapping that if we re-sample a randomly drawn sample of the population, we can consider it representative of the parent population. The second table shows the results of the balance test by each covariate of interest, and we can see that they are sufficiently well balanced that none of the covariates requires adjustment before analysis.

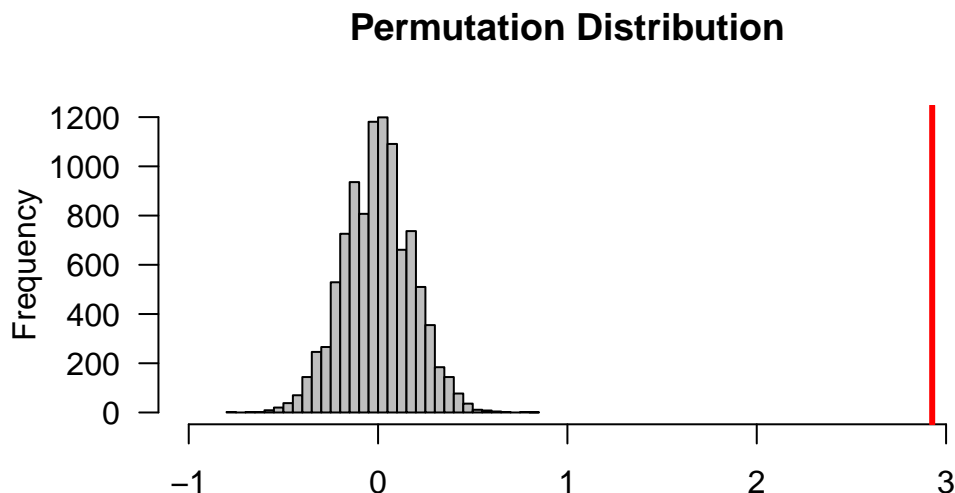
Extreme data points and covariate values will not be an issue as the questions will come with a scale for answers, hence a limited range of answer options is possible. Missing data could be an issue, though hopefully this will be minimal as the survey experiment will not be particularly long or arduous. Systematic “missingness” of data also seems unlikely, since there is no expectation that any particular group of people will be disadvantaged by taking the survey, or when taking it (due to content, comprehension ability, etc.). However, if missing data is a problem, that would be an

important thing to check for. I think it is more likely that I may have an issue with measuring treatment strength and determining if it is strong enough for any potential effect to come through in the data. Unfortunately, this can't be remedied after the fact, but I can run some treatment checks before I do the full experiment to try and stave off this possibility.

## Statistical Tests of the Hypothesis

I intend to use a permutation test to test the null hypothesis of no treatment effect in my data. I use this method as my treatment will be randomly assigned within the survey experiment, with all participants having an equal chance of being subjected to the treatment or control conditions. By permuting the treatment within matched groups, I will generate a distribution of the possible outcomes when the treatment has no relationship with the certainty outcome. I can then compare this distribution to the average treatment effect (ATE) result I estimate from my sample to gauge how likely my result would be if there was no true treatment effect.

Permutation tests are generally conservative in terms of false positive rates, meaning their size is often smaller than their level. I ran a permutation test on the data 10,000 times and created a distribution of outcomes such that would be generated if the null hypothesis were true:



The graph above shows the null hypothesis outcome distribution. On the right side, the red line indicates the effect size found by the fixed-effects matched model. It is clearly a great distance away from the null hypothesis and, therefore, I am confident that my ATE estimate is not a product of chance (indeed, I set it to be 3 points, so I know this is roughly correct).

## Alternative Testing Options

Permutation testing is my chosen testing method. However, there are other options that I could have equally chosen.

## ANOVA

Running an ANOVA analysis could have been another way to test whether the control and treatment group outcomes are meaningfully different. The output below shows the result of a one-way ANOVA test:

```
              Df Sum Sq Mean Sq F value Pr(>F)
group          1  642.4   642.4    1884 <2e-16 ***
Residuals     298  101.6     0.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
One-sample t test power calculation
```

```
      n = 150
      d = 0.4
sig.level = 0.05
  power = 0.9981767
alternative = two.sided
```

```
[1] 0.0476
```

The first output shows the results of the one-way ANOVA test, which is suitable for using on the dataset here as-is, as the data is well-balanced and did not require matching. It shows that the treatment and control groups' responses are different to a degree that would not be seen by chance. The power calculation output shows that With the current sample size, the power of the test is well over the 80% threshold. Using simulation, I find the size of the test to be 0.05, which matches the level of the test.

## T-Test (for randomized experimental samples only)

A t-test would also have been appropriate for a randomized experiment, such as the dataset is without matching:

```
Welch Two Sample t-test
```

```
data:  response by group
t = -43.4, df = 290.57, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -3.059388 -2.793946
sample estimates:
mean in group 0 mean in group 1
 -0.020000      2.906667
```

```
One-sample t test power calculation
```

```
      n = 150
      d = 0.4
sig.level = 0.05
  power = 0.9981767
alternative = two.sided
```

```
[1] 0.046
```

This test correctly finds that the difference in means is significant, and is between 2.79 - 3.06. It further estimates that the mean in control group is 0.0, and the mean in treatment group is 2.9, which is very accurate. The power of the test is well over the 80% threshold. Using simulation, I find the size of the test to be 0.05, which matches the level of the test.

## Estimation

I will use OLS regression to obtain an estimate of the ATE in this simulated experiment. This method is appropriate since my estimand is the simple difference of means in the change of certainty rating before and after the vignettes across the two groups. As the certainty score is rated on a scale of 1 to 10, the range of possible change answers is -9 to 9. Although discrete in measurement, this variable can therefore be thought of as continuous, and is suitable for use as the dependent variable in an OLS regression.

First, though this data is balanced, I use optimal matching to divide the sample into strata. The demographic information provided in the ANES survey data is used to match respondents before running regression analysis, to ensure that comparisons can be made with minimal confounding (Rosenbaum, 2010; Hansen & Bowers, 2008). This generates 144 strata, most of which are matched pairs (133), but with no strata containing more than four respondents. I also conduct an asymptotic independence test to check the assumptions of the balance test (IID-CLT), and it confirms the output of the balance test.

The balance test output below shows the differences between the control and treatment group means across all covariates:

	stat					
vars	Control	Treatment	std.diff	adj.diff	pooled.sd	z
age	50.0722222	52.4466667	0.14265596	2.374444444	16.6445515	3.2335547
white	0.7400000	0.7333333	-0.01514969	-0.006666667	0.4400529	-0.5773503
male	0.4533333	0.4200000	-0.06706696	-0.033333333	0.4970157	-2.2360680
educ	4.7122222	4.7733333	0.03090955	0.061111111	1.9770948	0.6153871
income	12.1666667	11.4866667	-0.09906496	-0.680000000	6.8641826	-2.1070206
pid7	3.8533333	3.7933333	-0.02565079	-0.060000000	2.3391092	-0.7595545
	stat					
vars	p					
age	0.007335594					
white	1.000000000					
male	0.126736593					
educ	1.000000000					
income	0.140463180					
pid7	1.000000000					

Age is the only covariate for which there is a meaningful statistical difference, with the average age of the control group being 50.0, compared to 52.4 in the treatment group. By chance, the treatment group is on average about 2.5 years older than their control counterparts. I do not consider this to be a problem, as I do not expect an age difference of ~2.5 years to have any effect on my experimental results. The control group skews slightly more male and a little wealthier, but these differences are not statistically meaningful, nor are race, education level, or partisanship.

Below is the output of three linear regression models. Model 1 is a simple linear model containing just the independent and dependent variables. For comparison, Model 2 shows the simple linear model with the covariates of interest included as additive “controls”. Finally, Model 3 is a robust linear model with HC2 standard errors and fixed-effects applied by matched strata.

	LM	LM with Covariates	Robust LM, Matching FE
	(1)	(2)	(3)
Constant	-0.020 (0.048)	-0.168 (0.164)	
Treatment	2.927*** (0.067)	2.915*** (0.068)	2.922*** (0.071)
Age		0.001 (0.002)	
White (Race)		0.084 (0.082)	
Gender (Male)		-0.044 (0.069)	
Education		0.026 (0.018)	
Income		-0.006 (0.005)	
Partisan Identity		0.002 (0.015)	
Num.Obs.	300	300	300
R2	0.863	0.865	0.921
R2 Adj.	0.863	0.862	0.848
AIC	532.6	540.4	365.6
BIC	543.8	573.8	373.0
Log.Lik.	-263.320	-261.223	
F	1883.597	267.967	
RMSE	0.58	0.58	0.44

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

Model 3 is the best fit for the data according to the AIC and BIC values. It also has the lowest RMSE of the three, which indicates that it is the least biased, since its predicted values will be closer to the “true” values than the other two models. Therefore, I would choose Model 3 as my preferred model.

To check more carefully for bias, I then run a sensitivity analysis on Model 3, per Cinelli and Hazlett (2020). It indicates a coefficient estimate of a treatment effect of a 2.9 point increase in certainty. It also finds that the design is potentially sensitive only to unobserved confounders that could explain 91.4% of both the variance in treatment and control, as only unobserved confounders of this magnitude could render the result statistically indistinguishable from zero. Since my data is simulated, it is not surprising that the unobserved confounders would need to be so high in order to effect the conclusions drawn from the data. In reality, I would expect this result to be weaker, since real-life data is considerably more messy than its simulated counterpart.

## Interpreting the Results

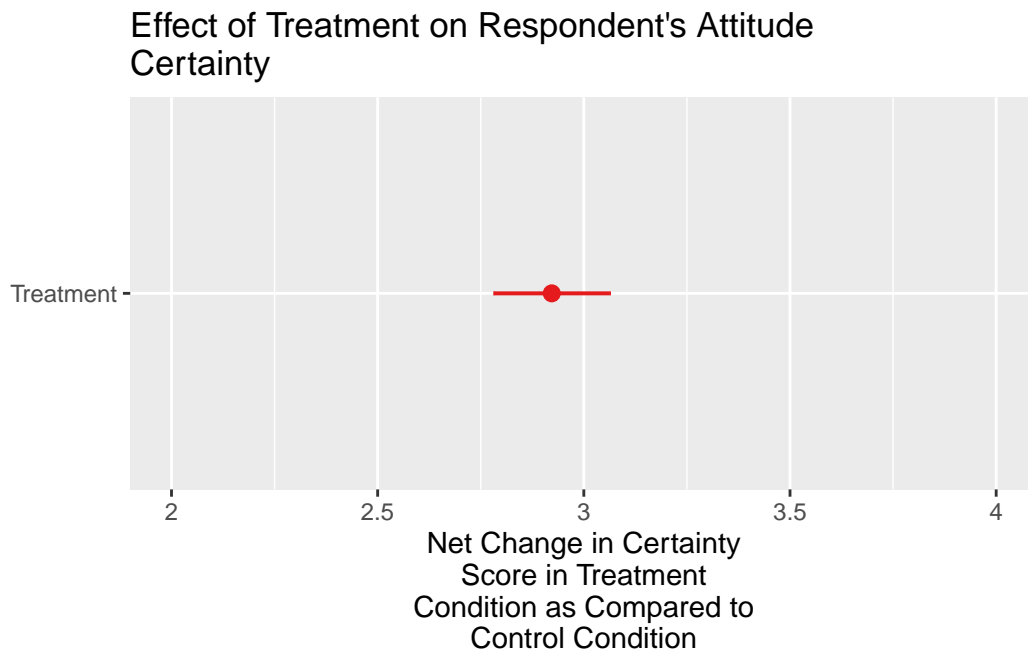
In the regression table presented above, I provide three OLS linear models, which show the difference in means between the outcome in the treatment and control groups for the change in certainty score before and after the vignette is received. The third linear model would be the one I present in a final paper, but I include the prior two here for comparison purposes.

Firstly, Model 1, the simple linear model, performs reasonably well in estimating the true effect (set at 3 for treatment group, and 0 for control group). Since the dataset was well-balanced due to

random sampling of the ANES data for mock respondents and the random treatment assignment, this is not surprising. For comparison, Model 2 shows the simple linear model with added “controls” for covariates. This model is not any better than the simple regression because the data was well-balanced to begin with, and so demographic factors have very little influence on the outcome coefficient of interest.

Model 3, the one I would include in the final paper, shows the output of a robust linear model with HC2 standard errors and fixed-effects applied by matched strata. Although matching was not strictly necessary as the dataset was well-balanced, running this type of model gives me greater confidence that my results are not being influenced by observable confounding, and the sensitivity analysis similarly allays concerns about unobservable bias. This regression estimates a treatment coefficient of 2.9, with a 95% confidence interval of 2.8 to 3.1 (not displayed on the table), i.e., the treatment effect results in a mean certainty score of around three points higher than the control group, comparing post to pre-treatment certainty scores (on a 1-10 scale). The “true” effect of treatment assigned in the fake dataset was 3.0. If this data were real, it would indicate that reinforcement framing can increase issue attitude certainty at the individual level. This would provide support for my theory that framing serves to reinforce political opinions.

In the graph below, I have visualized the treatment effect, with confidence intervals, as found in my simulated experimental data. (This graph is not ideal, as for some reason, I can’t get the control group on there, but their simulated change was 0, in comparison.)





## Link to GitHub Repository Containing Relevant Files

[https://github.com/janepb2/PS531\\_PreAnalysisPlan](https://github.com/janepb2/PS531_PreAnalysisPlan)

## Appendix: All code for this report

```
library(stargazer)
library(optmatch)
library(RIttools)
library(dplyr)
library(estimatr)
# install.packages('modelsummary')
library(modelsummary)
# install.packages('lmtest')
library(lmtest)
library(sandwich)
library(gt)
library(ggplot2)
# install.packages('ggeffects')
library(ggeffects)
# install.packages('sjPlot')
library(sjPlot)
library(sjlabelled)
library(sjmisc)
# install.packages('pwr')
library(pwr)
library(sensemakr)
# install.packages('Matrix')
library(Matrix)
library(coin)
knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)

# Variables in this dataset age # range 18-80 (max) white #
# 1 white, 0 else male # 1 male, 0 female educ # range 1-8
# - 1 less than high school to 8 doctorate income # range
# 1-22 - 1 <$10k - 22 $250k+ pid7 # range 1-7 - 1 strong
# dem - 7 strong rep

# 300 randomly selected respondents from ANES 2020, with
# complete data on all of these variables (i.e., no NAs) -
# I compiled this from an existing file, by selecting these
# columns, removing rows with NAs, and randomly sampling
# 300 rows of data without replication.

# Now I will assign treatment and control groups, and then
# effects.

anes <- read.csv("https://raw.githubusercontent.com/janepb2/seinfelddata/main/quant2practice.csv")

# Assign treatment - although the assignment is to odd and
# even rows below, the assignment is still random, as the
# participants were drawn at random from the ANES dataset,
# and each one was equally likely to be placed into the
# control or treatment groups (odd or even rows,
```

```

# respectively).

# control group response change mean is 0, treatment group
# mean change is 3, sd = 0.5 scores can be -9 to 9 change
# ---> calculate new score - old score to get difference.

set.seed(820622)
response <- rnorm(n = 300, mean = c(0, 3), sd = 0.5)
response <- round(response, digits = 0)
group <- rep(0:1, length.out = 300)
sim <- data.frame(group, response)

# run simple lm to check simulation
lmpractice <- lm(response ~ group, data = sim)
# as DV is numerical (and theoretically continuous,
# although bounded), this simple regression works
# reasonably well to capture the effect that has been
# specified in this simple setup. MSE looks correct as
# well. So simulation was successful.

stargazer(lmpractice, type = "text")

# merge the datasets
sim$X <- 1:300

mysample <- left_join(anes, sim, by = "X")

# age # 18-80 (max) white # 1 white, 0 else male # 1 male,
# 0 female educ # 1-8 1 under HS to 8 phd income # 1-22 1
# below 10k - 22 250k+ pid7 # 1-7 1 strong dem - 7 strong
# rep

bt1 <- balanceTest(group ~ age + white + male + educ + income +
  pid7, data = mysample)
bt1$overall
bt1
bt1

## Using a permutation distribution
set.seed(820622)

# Difference in means
original <- diff(tapply(response, group, mean))
mean(response[group == 1]) - mean(response[group == 0])
n <- 300
original
# Permutation test
permutation.test <- function(group, response, n) {
  distribution = c()
  result = 0
  for (i in 1:n) {
    distribution[i] = diff(by(response, sample(group, length(group),

```

```

        FALSE), mean))
    }
    result = sum(abs(distribution) >= abs(original))/(n)
    return(list(result, distribution))
}

test1 <- permutation.test(group, response, 10000)

permchart <- hist(test1[[2]], breaks = 50, col = "grey", main = "Permutation Distribution",
  las = 1, xlab = "", xlim = c(-1, 3))
abline(v = original, lwd = 3, col = "red")

# one way ANOVA can be run here as data is 'as-if random'
# being well balanced
one.way <- aov(response ~ group, data = mysample)

summary(one.way)

# test for power
pwr.t.test(n = 150, d = 0.4, sig.level = 0.05, type = "one.sample",
  alternative = "two.sided")

# test for size
sizeanova <- replicate(10000, summary(aov(response ~ group, data = mysample %>%
  mutate(group = sample(group, n(), replace = F))))[[1]][["Pr(>F)"]][1])

mean(sizeanova < 0.05)

# independent 2-group T-test
t.test(response ~ group, data = mysample)

# test for power
pwr.t.test(n = 150, d = 0.4, sig.level = 0.05, type = c("one.sample"))

# test for size
size_t <- replicate(10000, t.test(response ~ group, data = mysample %>%
  mutate(group = sample(group, n(), replace = F)))$p.value)

mean(size_t < 0.05)

# mahalanobis

mahal_dist <- match_on(group ~ age + white + male + educ + income +
  pid7, data = mysample, method = "mahalanobis")

fm <- fullmatch(mahal_dist, data = mysample)
summary(fm, min.controls = 0, max.controls = Inf)

bt2 <- balanceTest(group ~ age + white + male + educ + income +
  pid7 + strata(fm), data = mysample)

```

```

bt2$overall
bt2$results[, , "fm"]

# so within the strata, the chances of an experiment
# randomly matching these people is 0.001%, which is tiny.
# This means the matching went well. Overall balance test
# result remains the same as no data were dropped in the
# matching process.

bt2$overall["fm", ]

## Check validity (i.e. that normal distribution assumption
## is met) using coin:
overall_test_asymp <- independence_test(age + white + male +
  educ + income + pid7 ~ group | fm, data = mysample, teststat = "quadratic",
  distribution = asymptotic())

overall_test_asymp

# results are the same - looks good
bt2$results[, , "fm"]

simplelm <- lm(response ~ group, data = mysample)
lmwithcov <- lm(response ~ group + age + white + male + educ +
  income + pid7, data = mysample)

lmfe <- lm_robust(response ~ group, fixed_effects = ~fm, data = mysample)

modelsummary(list(simplelm, lmwithcov, lmfe))

cm <- c(`(Intercept)` = "Constant", group = "Treatment", age = "Age",
  white = "White (Race)", male = "Gender (Male)", educ = "Education",
  income = "Income", pid7 = "Partisan Identity")

regtable <- modelsummary(list(simplelm, lmwithcov, lmfe), stars = TRUE,
  output = "gt", title = "OLS Model Estimates of
  Treatment Effect of the Vignette",
  coef_map = cm)

regtable <- regtable %>%
  tab_spanner(label = "LM", columns = 2) %>%
  tab_spanner(label = "LM with Covariates", columns = 3) %>%
  tab_spanner(label = "Robust LM, Matching FE", columns = 4)

# when finished editing have to run:
gt::gtsave
regtable

sens_analysis_1 <- sensemakr(estimate = coef(lmfe)[["group"]],
  se = lmfe$std.error[["group"]], treatment = "group", dof = lmfe$df)

```

```

summary(sens_analysis_1)

# discrete x-axis

mysample <- mysample %>%
  mutate(Treatment = if_else(group == 1, 1, 0))

lmfe2 <- lm_robust(response ~ Treatment, fixed_effects = ~fm,
  data = mysample)

p <- plot_model(lmfe2, title = "Effect of Treatment on Respondent's Attitude Certainty",
  axis.title = "Net Change in Certainty Score in Treatment Condition as Compared to Control Con
p

```

## References

- Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? *Journal of Politics*, 70(2), 542–555. <https://doi.org/10.1017/S0022381608080493/ASSET/IMAGES/LARGE/FIG7G.JPEG>
- Abramowitz, A. I., & Webster, S. (2016). The Rise of Negative Partisanship and the Nationalization of U.S. Elections in the 21st Century. *Electoral Studies*, 41, 12–22. <https://doi.org/10.1016/j.electstud.2015.11.001>
- Barabas, J., & Jerit, J. (2010). Are Survey Experiments Externally Valid? *American Political Science Review*, 104(2), 226–242. <https://doi.org/10.1017/S0003055410000092>
- Barber, M., & McCarty, N. (2015). Causes and consequences of polarization. *Political Negotiation: A Handbook*, 37, 39–43.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press. <https://psycnet.apa.org/record/1993-97948-000>
- Cinelli, C., & Hazlett, C. (2020). Making Sense of Sensitivity: Extending Omitted Variable Bias. *Journal of the Royal Statistical Society*, 82(1), 39–67. <https://doi.org/10.1111/rssb.12348>
- Hansen, B. B., & Bowers, J. (2008). Covariate Balance in Simple, Stratified and Clustered Comparative Studies. *Statistical Science*, 23(2), 219–236. <https://www.jstor.org/stable/27645895>
- Krosnick, J. A., & Abelson, R. P. (1992). The Case for Measuring Attitude Strength in Surveys. In J. M. Tanur (Ed.), *Questions About Questions: Inquiries into the Cognitive Bases of Surveys* (pp. 177–203). Russell Sage Foundation.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Levendusky, M. (2009). *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. University of Chicago Press.
- McCarty, N. M., Poole, K. T., & Rosenthal, H. (2006). *Polarized America: the dance of ideology and unequal riches*. MIT Press.
- Michael Alvarez, R. (1996). *Survey Measures of Uncertainty: A Report to the National Election Studies Board on the Use of Certainty Questions to Measure Uncertainty About Candidate Traits and Issue Positions*.
- Rosenbaum, P. R. (2010). *Design of Observational Studies*. Springer.
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755–769. <https://doi.org/10.1111/J.1540-5907.2006.00214.X>
- Turner, J. (2007). The messenger overwhelming the message: Ideological cues and perceptions of bias in television news. *Political Behavior*, 29(4), 441–464. <https://doi.org/10.1007/S11109-007-9031-Z>
- Wilson, G. D., & Patterson, J. R. (1968). A New Measure of Conservatism. *British Journal of Clinical Psychology*, 7(4), 264–269.
- Zaller, J., & Feldman, S. (1992). A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science*, 36(3), 579. <https://doi.org/10.2307/2111583>