POL SCI 231b (Spring 2017):

Problem Set 8

Prof. Thad Dunning/GSI Natalia Garbiras-Díaz

Dept. of Political Science

University of California, Berkeley

Suggested Solutions

1. **The bootstrap. An analyst assumes the following regression model:**

$$Y = X\beta + \epsilon, \tag{1}$$

where Y is an $n \times 1$ vector of observable random variables. Here, X is a fixed $n \times p$ matrix with a vector of 1's as the first column, and ϵ is mean-zero vector of i.i.d. random variables with $\text{var}(\epsilon_i) = \sigma^2$. The OLS estimator for this model is $\hat{\beta} = (X'X)^{-1}X'Y$. The residuals from the OLS fit are $e = Y - X\hat{\beta}$.

Suppose the analyst uses the procedure described by Freedman (2009, Chapter 8) to bootstrap the regression model. In particular, for the kth bootstrap replicate, she samples at random with replacement from the vector e to produce an $n \times 1$ vector of bootstrap errors, $\epsilon_{(k)} = \{\epsilon_{(k)1}, ..., \epsilon_{(k)n}\}'$. For each bootstrap replicate, she then constructs $Y_{(k)} = X\hat{\beta} + \epsilon_{(k)}$ and fits the OLS estimator, $\hat{\beta}_{(k)} = (X'X)^{-1}X'Y_{(k)}$. There are 1,000 bootstrap replicates.

Finally, let

$$\begin{aligned}
\hat{\epsilon}_{(k)} &= Y_{(k)} - X\hat{\beta}_{(k)}, \\
s_k^2 &= \frac{\hat{\epsilon}'_{(k)}\hat{\epsilon}_{(k)}}{n-p}, \\
\hat{\beta}_{\text{ave}} &= \frac{1}{1000}\sum_{k=1}^{1000}\hat{\beta}_{(k)}, \text{ and} \\
V &= \frac{1}{1000}\sum_{k=1}^{1000}[\hat{\beta}_{(k)} - \hat{\beta}_{\text{ave}}][\hat{\beta}_{(k)} - \hat{\beta}_{\text{ave}}]'.
\end{aligned}$$

Say whether the following statements are true or false, and most importantly, explain your answers (a correct answer with an incorrect explanation does not get full credit!):

(a) $E(\epsilon_{(k)}) = 0_{n\times 1}$.

True. The bootstrap errors, $\epsilon_{(k)}$, are an i.i.d. sample from a box with mean zero. In more detail, the original OLS residual vector e has mean zero because there is an intercept in the model, and $e \perp X$. We are sampling at random with replacement from this vector. Thus, the expected value of each observed bootstrap error is the mean of the box.

(b) $E(\hat{\beta}_{(k)}) = \hat{\beta}$.

True. By (a), the bootstrap errors have an expected value of zero, and they are independent of X (which here is held fixed). The bootstrap OLS estimator $\hat{\beta}_{(k)}$ is therefore an unbiased estimator of $\hat{\beta}$. In more detail,

$$\begin{aligned}
\hat{\beta}_{(k)} &= (X'X)^{-1}X'Y_{(k)} \\
&= (X'X)^{-1}X'(X\hat{\beta} + \epsilon_{(k)}) \\
&= \hat{\beta} + (X'X)^{-1}X'\epsilon_{(k)}
\end{aligned}$$

so $E(\epsilon_{(k)}) = 0_{n\times 1}$ implies that $E(\hat{\beta}_{(k)}) = \hat{\beta}$.

(c) $E(s_k^2) = \sigma^2$.

False. Here, s_k^2 is an unbiased estimator for the variance of the box from which the bootstrap errors are drawn–namely, the box of original residuals e. But the variance of the box of residuals is not σ^2, at least not exactly.

In more detail, s_k^2 is the bootstrap analogue of $\hat{\sigma}^2$, where σ^2 is the variance of the error term in the usual regression model. See Freedman (2009: Theorem 4, pp. 47-48) for a proof that $E(\hat{\sigma}^2|X) = \sigma^2$. For the same reason, s_k^2 is an unbiased estimator for the variance of the residuals from the original regression fit—that is, the variance of the box from which the bootstrapped errors are drawn.

You might reason that since s_k^2 is an unbiased estimator for the variance of the bootstrap population of errors—and since $\hat{\sigma}^2$ is an unbiased estimator for σ^2 under the regression model—s_k^2 must be an unbiased estimator of σ^2. But this is wrong for two reasons. First, the variance of the bootstrap population of errors is $\frac{1}{n}e'e$, not $\hat{\sigma}^2 = \frac{1}{n-p}e'e$ (see d and e). Second, we are focused here on the sampling process we can control, which is defined by the bootstrap. Because we draw the bootstrap errors as an i.i.d. sample from the residuals from the original fit, we know that s_k^2) is unbiased for the variance of the population of *bootstrap* errors. But we don't know what produced the original fit (though we can make assumptions as per the usual regression model).

(d) $E(s_k^2) = \frac{1}{n}e'e$**.**

True: see (c). The variance of the box of original residuals is $\frac{1}{n}e'e$, and s_k^2 is an unbiased estimator for the variance of this box. (Remember that the residuals have mean zero, so $(e-\bar{e})'(e-\bar{e}) = e'e$, where $\bar{e}$ is the average of the residuals).

(e) $E(s_k^2) = \frac{1}{n-p}e'e$**.**

False. The variance of the box of original residuals is $\frac{1}{n}e'e$, not $\frac{1}{n-p}e'e$: see (d).

(f) $E(V) = \sigma^2(X'X)^{-1}$

False. Under the OLS model, the theoretical variance-covariance matrix of $\hat{\beta}_{(k)}$ is the variance of the bootstrap "error term" times $(X'X)^{-1}$. Here, the error term is given by i.i.d. draws from the vector e—and *not* draws of the unobservable distribution of ϵ. Thus, in (f), σ^2 should be replaced by $\frac{1}{n}e'e$—that is, the theoretical variance of the bootstrap errors.

In more detail, the bootstrap errors are i.i.d. random variables, because we are drawing at random with replacement from the vector e. The variance of the vector e is $\frac{1}{n}e'e$ (see d). Thus, the variance of each bootstrap error term is $\frac{1}{n}e'e$. Finally, we are generating the bootstrap data according to the OLS model, and so the theoretical variance-covariance matrix of the bootstrap replicates is $\frac{1}{n}e'e(X'X)^{-1}$. [See also the NOTE to item g].

(g) **The square roots of the diagonal elements of V are the bootstrap standard errors.**

True. The variance-covariance matrix of $\hat{\beta}_{(k)}$ across all 1,000 bootstrap replicates is V, and the square roots of the diagonal elements are the bootstrap standard errors. [NOTE: There was a typo in the exam, egads, with 100 rather than 1000 in the denominator of $\hat{\beta}_{\text{ave}}$ as well as V, and with the summations taken across $k = 1, \ldots, 100$ rather than $k = 1, \ldots, 1000$. If you noted this typo as the reason that (g) is false, and otherwise gave a correct answer, you will get full credit; we will also make adjustments in (f) if you mentioned this].

(h) **The sample SD of the $\hat{\beta}_{(k)}$'s is a good approximation to the SE of $\hat{\beta}$.**

True. This is the bootstrap principle at work.

(i) $\hat{\epsilon}_{(k)} \perp X$ **for all** k**.**

True. The $\hat{\epsilon}_{(k)}$ are the residuals from the OLS fit to the kth bootstrapped data set. Mechanically, these residuals are orthogonal to X – that is what regression does.

(j) **The bootstrap can provide evidence that the original data were produced according to equation (1).**

False. This claim is essentially untestable. The bootstrap *assumes* the original data were produced according to equation (4), with i.i.d. errors, $E(\epsilon_i) = 0$ and $\text{var}(\epsilon_i) = \sigma^2$, and then evaluates the sampling distribution of estimators such as $\hat{\beta}$. That is where the bootstrap principle comes in.

(k) **The bootstrap can provide evidence that $E(\hat{\beta}) = \beta$ if the original data were produced according to equation (1), with i.i.d. errors, $E(\epsilon_i) = 0$ and $\text{var}(\epsilon_i) = \sigma^2$.**

True. See the solution to (j).

2. **Miguel and Fisman (2006: 1020) are interested in how cultural norms influence corruption. They write, "Until 2002, diplomatic immunity protected [United Nations] diplomats from parking enforcement actions, so diplomats' actions were constrained by cultural norms alone. We find a strong effect of corruption norms: diplomats from high-corruption countries (on the basis of existing survey-based indices) accumulated significantly more unpaid parking violations." They refer to this study as a "unique natural experiment."**

(a) **What is the treatment variable in this study? It is plausibly assigned as-if at random? Is this plausibly a natural experiment?**

The "treatment" variable is the extent of corruption in each country (as measured by the survey-based indices), or perhaps "corruption norms." It is hard to imagine how this is assigned as-if at random. For instance, corruption may be greater in poorer countries, or in those with certain kinds of electoral institutions. Variables like national income may be confounders, if they are also related to diplomats' behavior. Thus, this doesn't seem very plausibly to be a natural experiment.

(b) **Consider three threats to the substantive or theoretical relevance of the intervention here, as discussed in Dunning (2012: Chapter 10): external validity, idiosyncrasy, and bundling. Which of these do you consider to be the most important here and why?**

Each of these might be concerns, but bundling seems a first-order problem. For example, lots of elements are contained in a "corruption" treatment. Corruption might involve the payment of bribes to obtain diplomatic posts, or it might involve the extraction of rents from business people or ordinary citizens. It might be hard to know which of these influence diplomats' accumulation of unpaid parking tickets.

(c) **What if diplomats from richer countries tend to have paid parking spaces? What violation(s) of the natural-experimental setup would this imply?**

This is a violation of as-if random: see part (a). Here we have a confounding variable (wealth) that is associated with corruption and with diplomats' propensity

to accumulate unpaid parking violations—since diplomats who have their parking spaces paid (e.g. by their embassies) are less likely to get parking tickets. (Note that this is not a violation of excludability: i.e., this is not about an alternate channel other than corruption norms through which being from a high-corruption country influences propensity to accumulate parking violations).

3. **Regression Discontinuity Design**.

   (a) **Researcher A, analyzing a regression discontinuity design, uses all of the data inside of a given window to fit the following regression using OLS:**

$$E(Y_i|R_i, T_i) = \alpha_1 + \alpha_2 T_i + \alpha_3 R_i + \alpha_4 T_i * R_i, \tag{2}$$

   where T_i is the treatment indicator and R_i is the value of the running variable for i.

   Researcher B chooses the same window but splits the data into treatment and control group and runs two separate OLS regressions:

$$E(Y_i|R_i, T_i) = \beta_1 + \beta_2 R_i \quad \text{if } T_i = 1 \tag{3}$$

   and

$$E(Y_i|R_i, T_i) = \beta_3 + \beta_4 R_i \quad \text{if } T_i = 0 \tag{4}$$

   Write the coefficients in equation (2) in terms of the β_j coefficients in equations (3) and (4).

   Note that when $T_i = 0$, equation (2) reduces to

$$E(Y_i|R_i, T_i) = \alpha_1 + \alpha_3 R_i \tag{5}$$

   Thus, $\alpha_1 = \beta_3$ and $\alpha_3 = \beta_4$.

   When $T_i = 1$, we can rewrite equation (2) as

$$E(Y_i|R_i, T_i) = (\alpha_1 + \alpha_2) + (\alpha_3 + \alpha_4) * R_i \tag{6}$$

   Thus, $\alpha_1 + \alpha_2 = \beta_1$ and $\alpha_3 + \alpha_4 = \beta_2$. Since $\alpha_1 = \beta_3$ and $\alpha_3 = \beta_4$, we have:

$$\begin{aligned} \alpha_1 + \alpha_2 &= \beta_1 \\ \beta_3 + \alpha_2 &= \beta_1 \qquad (7) \\ \alpha_2 &= \beta_1 - \beta_3 \qquad (8) \\ & \qquad (9) \end{aligned}$$

   and

$$\begin{aligned} \alpha_3 + \alpha_4 &= \beta_2 \\ \beta_4 + \alpha_4 &= \beta_2 \quad (10) \\ \alpha_4 &= \beta_2 - \beta_4 \quad (11) \\ & \quad (12) \end{aligned}$$

And so we can rewrite equation (2) as:

$$E(Y_i|R_i, T_i) = \beta_3 + (\beta_1 - \beta_3)T_i + \beta_4 R_i + (\beta_2 - \beta_4)T_i * R_i, \quad (13)$$

(b) **RD estimation function.** Modify Hidalgo's RD "estimate" function such that it calculates difference in means estimates using your own t-test function.

```
library(reshape2)
library(sandwich)

ttest <- function(y, x, two.tailed = TRUE) {

    # Calculating difference in means
    mean1 <- mean(y[x == 1], na.rm = T)
    mean0 <- mean(y[x == 0], na.rm = T)
    diff <- mean1 - mean0

    # Calculating SE of the difference
    N1 <- length(na.omit(y[x == 1]))
    N0 <- length(na.omit(y[x == 0]))
    var1 <- var(y[x == 1], na.rm = T)
    var0 <- var(y[x == 0], na.rm = T)
    varN1 <- var1/N1
    varN0 <- var0/N0
    se.diff <- sqrt(varN1 + varN0)

    # T-statistic
    t <- diff/se.diff

    # Degrees of freedom
    df.num <- ((varN1 + varN0)^2)
    df.den <- (varN1^2)/(N1 - 1) + (varN0^2)/(N0 -
        1)
    df <- df.num/df.den

    # P-value
    if (two.tailed == TRUE) {
        p <- 2 * pt(abs(t), df, lower.tail = F)
```

```
    }

    if (two.tailed == FALSE) {
        p <- pt(t, df, lower.tail = F)
    }

    # Preparing output
    res <- c(mean1, mean0, diff, se.diff,
        t, (N1 + N0), df, p)
    names(res) <- c("Mean 1", "Mean 0",
        "Difference", "SE Diff", "t-stat",
        "N", "df", "p-value")

    return(c(res))
}


rd.estim <- function(outcome.var, data) {
    library(car)
    data$outcomeVar <- data[, outcome.var]
    data$fv <- data$electorate.96 -
        40500
    data$wts <- ifelse(abs(data$fv/15000) <=
        1, 1 - abs(data$fv/15000), 0)
    ll.results <- lm(outcomeVar ~ comp.voting.98 *
        fv, data = data, weights = wts)

    ########################
    dm.results <- with(data[data$electorate.96 >
        35500 & data$electorate.96 <
        45500, ], ttest(outcomeVar,
        comp.voting.98))
    ########################

    fullsample.results <- lm(outcomeVar ~
        comp.voting.98, data = data)

    ll.out <- c(bandwidth = 15000, coef(ll.results)[[2]],
        sqrt(vcov(ll.results)[2, 2]),
        ll.results$df.residual + 2,
        coef(ll.results)[[1]])
    names(ll.out) <- c("Bandwidth",
        "RD Estimate", "Standard Error",
        "N", "Baseline")
```

```
    ########################
    dm.out <- c(dm.results[3], dm.results[4],
        dm.results[6], dm.results[2])
    names(dm.out) <- c("RD Estimate",
        "Standard Error", "N", "Baseline")
    ########################

    fullsample.out <- c(coef(fullsample.results)[[2]],
        sqrt(vcovHC(fullsample.results,
            type = "HC1")[2, 2]), fullsample.results$df.residual +
            2, coef(fullsample.results)[[1]])
    names(fullsample.out) <- c("RD Estimate",
        "Standard Error", "N", "Baseline")

    list(ll.results = ll.out, dm.results = dm.out,
        fullsample.results = fullsample.out,
        sd = sd(data$outcomeVar, na.rm = TRUE))
}
```

(c) **Balance tests/F-test.**

- Using your new function, produce a table replicating the individual balance tests in the paper.

```
## Balance
brazil.baldata <- brazil.baldata[brazil.baldata$remove !=
    1, ]
names(brazil.baldata)
##  [1] "uf"
##  [2] "muni"
##  [3] "ibge.municode"
##  [4] "tse.municode"
##  [5] "electorate.96"
##  [6] "all.comp.voting.98"
##  [7] "comp.voting.98"
##  [8] "region"
##  [9] "capital"
## [10] "remove"
## [11] "party_won_1996"
## [12] "gini_1991"
## [13] "pt_pres_1994"
## [14] "income_1991"
## [15] "psdb_1996"
## [16] "pfl_1996"
## [17] "pt_1996"
```

```
## [18] "latitude"
## [19] "longitude"
## [20] "poverty_high_1991"
## [21] "poverty_high_2000"
## [22] "frac_25_illit_1991"
## [23] "depfedbranco.94"
## [24] "depfednulo.94"
## [25] "depestbranco.94"
## [26] "depestnulo.94"
## [27] "totalvotes.94"
## [28] "missing.94"
## [29] "depfedPT.94"
## [30] "depfedPMDB.94"
## [31] "depfedPFL.94"
## [32] "depfedPSDB.94"
## [33] "totalvotes.82"
## [34] "depfedbranco.82"
## [35] "depfednulo.82"
## [36] "depestbranco.82"
## [37] "depestnulo.82"
## [38] "pref.blank.pct.96"
## [39] "pref.null.pct.96"
## [40] "ver.blank.pct.96"
## [41] "ver.null.pct.96"
## [42] "depfed.branconulo.94"
## [43] "depest.branconulo.94"
## [44] "SP"
## [45] "BA"
## [46] "MG"
## [47] "MA"
## [48] "RS"
```

```
br.bal.covar <- c("gini_1991", "pt_pres_1994",
    "income_1991", "psdb_1996", "pfl_1996",
    "poverty_high_1991", "frac_25_illit_1991",
    "depfednulo.94", "depestbranco.94",
    "depestnulo.94", "depfedPT.94",
    "depfedPMDB.94", "depfedPFL.94",
    "depfedPSDB.94", "ver.blank.pct.96",
    "ver.null.pct.96", "SP", "BA", "MG")
br.balance <- lapply(br.bal.covar, function(x) rd.estim(x,
    brazil.baldata))
names(br.balance) <- br.bal.covar
```

```
br.balance[[1]]
## $ll.results
##      Bandwidth    RD Estimate
##   1.500000e+04   1.240736e-02
## Standard Error              N
##   9.898765e-03   3.510000e+02
##       Baseline
##   5.562011e-01
##
## $dm.results
##    RD Estimate Standard Error
##   4.138879e-03   1.062609e-02
##              N       Baseline
##   1.140000e+02   5.560704e-01
##
## $fullsample.results
##    RD Estimate Standard Error
##   9.817923e-03   3.564537e-03
##              N       Baseline
##   5.238000e+03   5.245527e-01
##
## $sd
## [1] 0.05589197
# For the table, we will keep the
# results from the difference in
# means.

table <- matrix(NA, length(br.balance),
    4)

for (i in 1:length(br.balance)) {

    table[i, ] <- br.balance[[i]]$dm.results

}

rownames(table) <- br.bal.covar
colnames(table) <- names(br.balance[[1]]$dm.results)
table
##                     RD Estimate
## gini_1991           0.004138879
## pt_pres_1994        2.070160106
## income_1991        21.977536306
## psdb_1996           1.766638843
```

```
## pfl_1996             2.152656152
## poverty_high_1991   -4.718480967
## frac_25_illit_1991  -4.919452255
## depfednulo.94       -0.010634365
## depestbranco.94      0.001413563
## depestnulo.94        0.004798345
## depfedPT.94          0.035367793
## depfedPMDB.94       -0.009127465
## depfedPFL.94        -0.041635459
## depfedPSDB.94        0.039121623
## ver.blank.pct.96    -0.676297904
## ver.null.pct.96      0.270790861
## SP                   0.082541762
## BA                   0.055027841
## MG                   0.119882083
##                     Standard Error
## gini_1991              0.010626088
## pt_pres_1994           1.243874360
## income_1991           15.455934527
## psdb_1996              2.995355890
## pfl_1996               2.683217246
## poverty_high_1991      4.287302854
## frac_25_illit_1991     2.731015389
## depfednulo.94          0.013924671
## depestbranco.94        0.009130459
## depestnulo.94          0.012663530
## depfedPT.94            0.026578553
## depfedPMDB.94          0.038522408
## depfedPFL.94           0.035722181
## depfedPSDB.94          0.034660386
## ver.blank.pct.96       0.551703914
## ver.null.pct.96        0.535586368
## SP                     0.074307959
## BA                     0.062959598
## MG                     0.075350443
##                       N
## gini_1991           114
## pt_pres_1994        114
## income_1991         114
## psdb_1996           114
## pfl_1996            114
## poverty_high_1991   114
## frac_25_illit_1991  114
## depfednulo.94       114
```

```
## depestbranco.94    114
## depestnulo.94      114
## depfedPT.94        114
## depfedPMDB.94      114
## depfedPFL.94       114
## depfedPSDB.94      114
## ver.blank.pct.96   114
## ver.null.pct.96    114
## SP                 114
## BA                 114
## MG                 114
##                       Baseline
## gini_1991           0.55607042
## pt_pres_1994       16.30335258
## income_1991       165.80546446
## psdb_1996           9.90662296
## pfl_1996            7.51482395
## poverty_high_1991  55.58769025
## frac_25_illit_1991 29.01812673
## depfednulo.94       0.22611413
## depestbranco.94     0.14513861
## depestnulo.94       0.16780157
## depfedPT.94         0.09853285
## depfedPMDB.94       0.22678218
## depfedPFL.94        0.17370527
## depfedPSDB.94       0.11600526
## ver.blank.pct.96    9.37227465
## ver.null.pct.96     4.21623239
## SP                  0.12676056
## BA                  0.08450704
## MG                  0.11267606
```

- **The F-test.** We will use the F-test to evaluate whether the entire set of pre-treatment covariates can predict treatment assignment. The null hypothesis here is that the coefficients from a regression of treatment assignment on all of the pre-treatment covariates are all zero.
  i. Fit the big model (including all the pre-treatment covariates) and the small model by OLS and compute the sums of squares that are needed for the test: $\|e\|^2$, $\|X\hat{\beta}\|^2$, and $\|X\hat{\beta}^{(s)}\|^2$ using the matrix commands in R.
  The "big" model is the complete model, i.e. regressing the treatment vector on the entire set of covariates. For the "small" model, we regress the treatment vector on a constant only.
  Building the data set here is tricky. We have no real experiment, so we need to focus on a sample where we think that the claim of as-if random assignment applies, although that is also what we are testing here. For

the purpose of this exercise, we will subset the data to districts with the margin used in the paper to calculate the difference of means which is 5,000.
For this subset of data, we will need all the covariates that we used for the balance tests plus the treatment variable.

```
vars <- c(br.bal.covar, "comp.voting.98")

# keeping data within the 5,000
# window only the variables we need.
window.data <- brazil.baldata[brazil.baldata$electorate.96 >
    35500 & brazil.baldata$electorate.96 <
    45500, vars]
names(window.data)
##  [1] "gini_1991"
##  [2] "pt_pres_1994"
##  [3] "income_1991"
##  [4] "psdb_1996"
##  [5] "pfl_1996"
##  [6] "poverty_high_1991"
##  [7] "frac_25_illit_1991"
##  [8] "depfednulo.94"
##  [9] "depestbranco.94"
## [10] "depestnulo.94"
## [11] "depfedPT.94"
## [12] "depfedPMDB.94"
## [13] "depfedPFL.94"
## [14] "depfedPSDB.94"
## [15] "ver.blank.pct.96"
## [16] "ver.null.pct.96"
## [17] "SP"
## [18] "BA"
## [19] "MG"
## [20] "comp.voting.98"
Y <- window.data$comp.voting.98
# design matrix for 'complete' model
X <- as.matrix(cbind(1, window.data[,
    1:19]))
# design matrix for 'small' model
x <- as.matrix(rep(1, nrow(X)))

hatbetas_1 <- solve(t(X) %*% X) %*%
    (t(X) %*% Y)
hatbetas_2 <- solve(t(x) %*% x) %*%
    (t(x) %*% Y)
```

```
Xhatbeta1 <- X %*% hatbetas_1
xhatbeta2 <- x %*% hatbetas_2

res <- Y - Xhatbeta1
sum_res2 <- sum(res^2)

sumsq_Xhatbeta1 <- sum(Xhatbeta1^2)
sumsq_xhatbeta2 <- sum(xhatbeta2^2)
```

ii. Use your results to calculate the F-statistic.
Here, p_0 is 19: we are testing the null that the coefficients for the 19 covariates are zero.

```
Fstat <- ((sumsq_Xhatbeta1 - sumsq_xhatbeta2)/ncol(window.data[,
    1:19]))/(sum_res2/(nrow(window.data[,
    1:19]) - ncol(window.data[, 1:19])))
Fstat
## [1] 1.197479
```

To check our work, we can compare the results from the F-stat in the $lm()$ function.

```
summary(lm(comp.voting.98 ~ ., data = window.data))
##
## Call:
## lm(formula = comp.voting.98 ~ ., data = window.data)
##
## Residuals:
##     Min      1Q  Median      3Q
## -0.9019 -0.3594 -0.1163  0.4034
##     Max
##  0.8916
##
## Coefficients:
##                       Estimate
## (Intercept)          0.2366382
## gini_1991           -0.9473328
## pt_pres_1994         0.0144146
## income_1991          0.0009202
## psdb_1996           -0.0008839
## pfl_1996             0.0007003
## poverty_high_1991    0.0103589
## frac_25_illit_1991 -0.0120322
## depfednulo.94       -2.3782918
## depestbranco.94      2.4518628
```

```
## depestnulo.94       0.5029754
## depfedPT.94         0.2616539
## depfedPMDB.94       0.0434280
## depfedPFL.94        0.0641196
## depfedPSDB.94       0.4302337
## ver.blank.pct.96   -0.0227948
## ver.null.pct.96     0.0267611
## SP                  0.4223741
## BA                  0.2502070
## MG                  0.0945289
##                    Std. Error
## (Intercept)         0.9946649
## gini_1991           1.4563024
## pt_pres_1994        0.0099188
## income_1991         0.0029178
## psdb_1996           0.0035746
## pfl_1996            0.0039304
## poverty_high_1991   0.0125117
## frac_25_illit_1991  0.0082430
## depfednulo.94       1.4715671
## depestbranco.94     1.7802941
## depestnulo.94       1.6839311
## depfedPT.94         0.4809212
## depfedPMDB.94       0.2893097
## depfedPFL.94        0.3530692
## depfedPSDB.94       0.3815769
## ver.blank.pct.96    0.0280280
## ver.null.pct.96     0.0228648
## SP                  0.2254454
## BA                  0.1838685
## MG                  0.1838542
##                    t value
## (Intercept)          0.238
## gini_1991           -0.651
## pt_pres_1994         1.453
## income_1991          0.315
## psdb_1996           -0.247
## pfl_1996             0.178
## poverty_high_1991    0.828
## frac_25_illit_1991  -1.460
## depfednulo.94       -1.616
## depestbranco.94      1.377
## depestnulo.94        0.299
## depfedPT.94          0.544
```

```
## depfedPMDB.94         0.150
## depfedPFL.94          0.182
## depfedPSDB.94         1.128
## ver.blank.pct.96     -0.813
## ver.null.pct.96       1.170
## SP                    1.874
## BA                    1.361
## MG                    0.514
##                     Pr(>|t|)
## (Intercept)           0.8125
## gini_1991             0.5170
## pt_pres_1994          0.1495
## income_1991           0.7532
## psdb_1996             0.8052
## pfl_1996              0.8590
## poverty_high_1991     0.4098
## frac_25_illit_1991    0.1477
## depfednulo.94         0.1094
## depestbranco.94       0.1717
## depestnulo.94         0.7658
## depfedPT.94           0.5877
## depfedPMDB.94         0.8810
## depfedPFL.94          0.8563
## depfedPSDB.94         0.2624
## ver.blank.pct.96      0.4181
## ver.null.pct.96       0.2448
## SP                    0.0641 .
## BA                    0.1768
## MG                    0.6084
## ---
## Signif. codes:
##   0 '***' 0.001 '**' 0.01 '*'
##   0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4794 on 94 degrees of freedom
## Multiple R-squared:  0.1932,Adjusted R-squared:  0.03015
## F-statistic: 1.185 on 19 and 94 DF,  p-value: 0.2871
```

iii. Write your own $F - test$ function and calculate a p-value using randomization inference.

```
ftest <- function(Y, covariates) {

    # design matrix for 'complete' model
    X <- as.matrix(cbind(1, covariates))
```

```
    # design matrix for 'small' model
    x <- as.matrix(rep(1, nrow(X)))

    hatbetas_1 <- solve(t(X) %*% X) %*%
        (t(X) %*% Y)
    hatbetas_2 <- solve(t(x) %*% x) %*%
        (t(x) %*% Y)

    Xhatbeta1 <- X %*% hatbetas_1
    xhatbeta2 <- x %*% hatbetas_2

    res <- Y - Xhatbeta1
    sum_res2 <- sum(res^2)

    sumsq_Xhatbeta1 <- sum(Xhatbeta1^2)
    sumsq_xhatbeta2 <- sum(xhatbeta2^2)

    ((sumsq_Xhatbeta1 - sumsq_xhatbeta2)/ncol(covariates))/(sum_res2/(nr
        ncol(covariates)))
}

# Let's check it works
ftest(Y = window.data$comp.voting.98,
    covariates = window.data[, 1:19])  # works!
## [1] 1.197479
```

Now, to randomization inference. One way of doing this is to write a loop that produces random treatment vectors for each iteration calculates and saves the f-stat. We will use a similar strategy but writing a function and using *replicate*().

```
random_fs <- function(treat, covariates) {
    # this function takes the treatment
    # vector and covariates, generates a
    # new random treatment vector
    # sampling without replacement, and
    # outputs the f-stat

    treat <- sample(treat, length(treat),
        replace = FALSE)

    ftest(Y = treat, covariates = covariates)

}
```

```r
random_fs(treat = window.data$comp.voting.98,
    covariates = window.data[, 1:19])
```

```
## [1] 1.226852
```

```r
randomization_fstats <- replicate(1000,
    random_fs(treat = window.data$comp.voting.98,
        covariates = window.data[, 1:19]))

plot(density(randomization_fstats))
abline(v = ftest(Y = window.data$comp.voting.98,
    covariates = window.data[, 1:19]),
    col = "red", lwd = 3)

# Now to the randomization p-value:
sum(randomization_fstats >= ftest(Y = window.data$comp.voting.98,
    covariates = window.data[, 1:19]))/1000
```

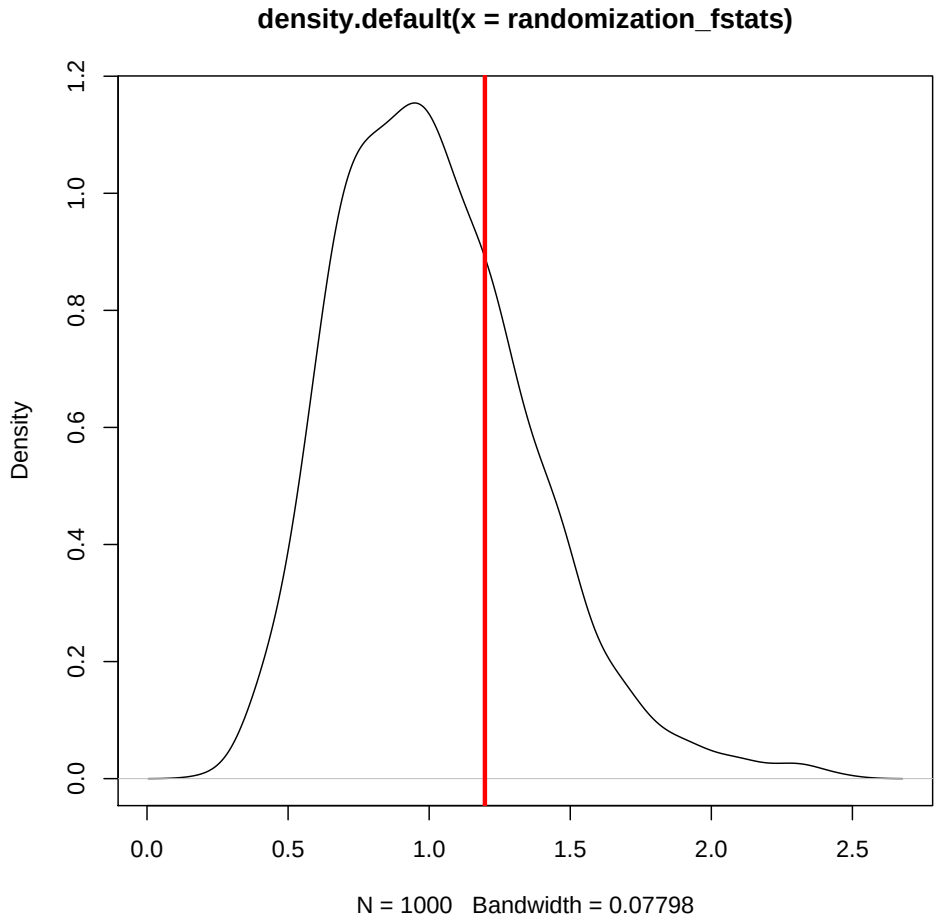
```
## [1] 0.276
```

```r
# Let's compare this to the one
# reported by lm() (which uses the
# chi-square distribution)
summary(lm(comp.voting.98 ~ ., data = window.data))
```

```
##
## Call:
## lm(formula = comp.voting.98 ~ ., data = window.data)
##
## Residuals:
##     Min      1Q  Median      3Q
## -0.9019 -0.3594 -0.1163  0.4034
##     Max
##  0.8916
##
## Coefficients:
##                      Estimate
## (Intercept)         0.2366382
## gini_1991          -0.9473328
## pt_pres_1994        0.0144146
## income_1991         0.0009202
## psdb_1996          -0.0008839
## pfl_1996            0.0007003
## poverty_high_1991   0.0103589
## frac_25_illit_1991 -0.0120322
## depfednulo.94      -2.3782918
## depestbranco.94     2.4518628
## depestnulo.94       0.5029754
```

```
## depfedPT.94          0.2616539
## depfedPMDB.94        0.0434280
## depfedPFL.94         0.0641196
## depfedPSDB.94        0.4302337
## ver.blank.pct.96    -0.0227948
## ver.null.pct.96      0.0267611
## SP                   0.4223741
## BA                   0.2502070
## MG                   0.0945289
##                     Std. Error
## (Intercept)          0.9946649
## gini_1991            1.4563024
## pt_pres_1994         0.0099188
## income_1991          0.0029178
## psdb_1996            0.0035746
## pfl_1996             0.0039304
## poverty_high_1991    0.0125117
## frac_25_illit_1991   0.0082430
## depfednulo.94        1.4715671
## depestbranco.94      1.7802941
## depestnulo.94        1.6839311
## depfedPT.94          0.4809212
## depfedPMDB.94        0.2893097
## depfedPFL.94         0.3530692
## depfedPSDB.94        0.3815769
## ver.blank.pct.96     0.0280280
## ver.null.pct.96      0.0228648
## SP                   0.2254454
## BA                   0.1838685
## MG                   0.1838542
##                     t value
## (Intercept)           0.238
## gini_1991            -0.651
## pt_pres_1994          1.453
## income_1991           0.315
## psdb_1996            -0.247
## pfl_1996              0.178
## poverty_high_1991     0.828
## frac_25_illit_1991   -1.460
## depfednulo.94        -1.616
## depestbranco.94       1.377
## depestnulo.94         0.299
## depfedPT.94           0.544
## depfedPMDB.94         0.150
```

```
## depfedPFL.94          0.182
## depfedPSDB.94         1.128
## ver.blank.pct.96     -0.813
## ver.null.pct.96       1.170
## SP                    1.874
## BA                    1.361
## MG                    0.514
##                     Pr(>|t|)
## (Intercept)           0.8125
## gini_1991             0.5170
## pt_pres_1994          0.1495
## income_1991           0.7532
## psdb_1996             0.8052
## pfl_1996              0.8590
## poverty_high_1991     0.4098
## frac_25_illit_1991    0.1477
## depfednulo.94         0.1094
## depestbranco.94       0.1717
## depestnulo.94         0.7658
## depfedPT.94           0.5877
## depfedPMDB.94         0.8810
## depfedPFL.94          0.8563
## depfedPSDB.94         0.2624
## ver.blank.pct.96      0.4181
## ver.null.pct.96       0.2448
## SP                    0.0641 .
## BA                    0.1768
## MG                    0.6084
## ---
## Signif. codes:
##   0 '***' 0.001 '**' 0.01 '*'
##   0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4794 on 94 degrees of freedom
## Multiple R-squared:  0.1932,Adjusted R-squared:  0.03015
## F-statistic: 1.185 on 19 and 94 DF,  p-value: 0.2871
```



iv. Is $\|Y\|^2 = \|X\hat{\beta}^{(s)}\|^2 + (\|X\hat{\beta}\|^2 - \|X\hat{\beta}^{(s)}\|^2) + \|e\|^2$? Coincidence or math fact?

```
sumsq_Y <- sum(Y^2)

sumsq_Y == sumsq_xhatbeta2 + (sumsq_Xhatbeta1 -
    sumsq_xhatbeta2) + sum_res2
## [1] FALSE
```

Math fact.

- **Optional:** Use the *DCdensity* function in the *rdd* package to test for sorting.[1]

(d) Download `rddata.Rda` file from bcourses. The file contains electoral outcomes for 2000 and 2004 for 6000 municipalities. You will estimate party incumbency advantage.

```
load("~/Dropbox/Academic/UC_Berkeley/GSI/PS_231B/Problem_sets/PS 8/rddata.Rda")
```

[1] You can see what the code needed for the test by checking the github repository

i. Use the 2000 electoral outcomes to set up an RDD where you keep the winner and the runner up parties in each municipality. Define the running variable as the margin of victory of the winner.
We want to keep just the winner and the runner up for each municipality and add the margin variable. We worked with this dataset in section and saw that there are 3 parties running in each municipality.

```
rd.data <- NA
muni <- unique(data$municipality)  # vector with unique muni numbers

for (i in 1:length(muni)) {

    temp <- data[data$municipality ==
        muni[i], ]
    # we discard the last candidate
    temp <- temp[-which(temp$vs_2000 ==
        min(temp$vs_2000)), ]
    # absolute margin between winner and
    # loser
    margin <- abs(temp$vs_2000[1] -
        temp$vs_2000[2])
    # loser gets negative margin
    temp$margin[which(temp$vs_2000 ==
        min(temp$vs_2000))] <- -margin
    # positive for the winner
    temp$margin[which(temp$vs_2000 ==
        max(temp$vs_2000))] <- margin

    # and we add the data for this muni
    # to our rd data
    rd.data <- rbind(rd.data, temp)

}

# let's check our new dataset
head(rd.data)
```

```
##   municipality  party  vs_2000
## 1           NA   <NA>       NA
## 2            1 center 41.72421
## 3            1   left 43.23756
## 4            2   left 33.82809
## 5            2  right 35.08161
## 7            3 center 34.91674
##    vs_2004 soc_spend    margin
## 1       NA        NA        NA
```

```
## 2 36.56648  40.67018 -1.513347
## 3 46.25497  48.98232  1.513347
## 4 37.67917  37.96897 -1.253524
## 5 37.52963  35.12098  1.253524
## 7 33.81395  37.22732 -3.415483
```

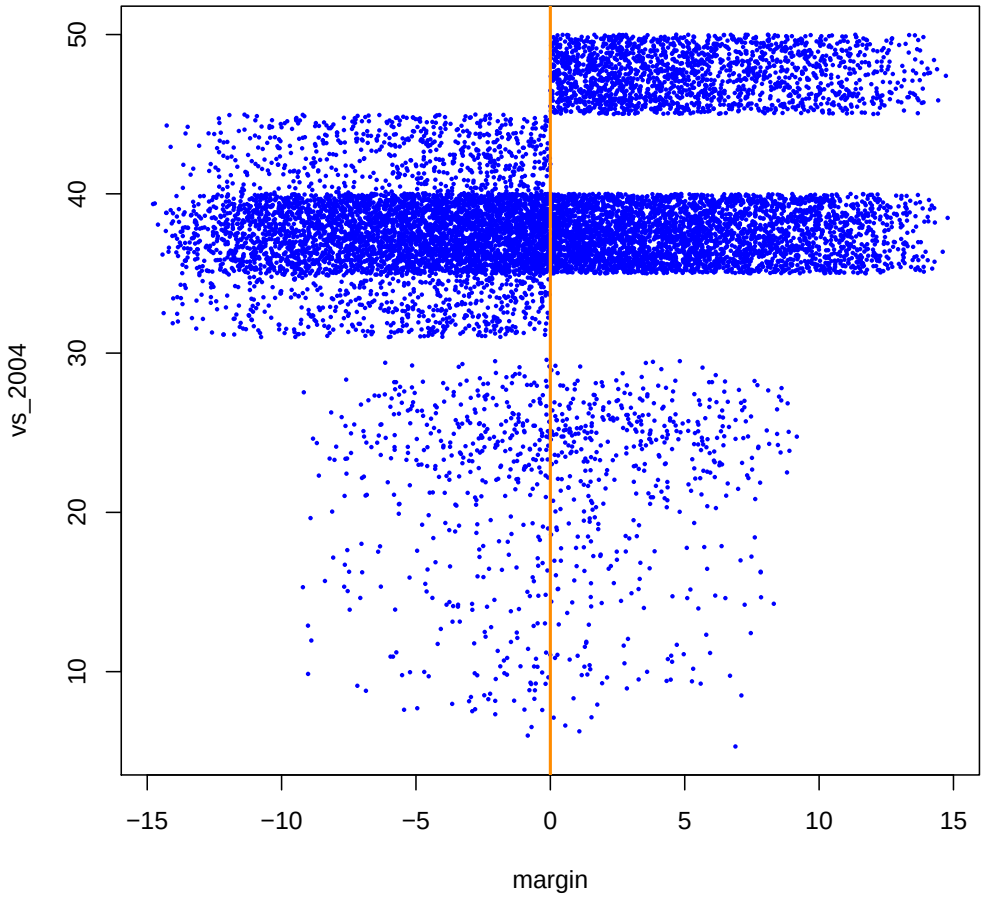
```
# drop the NA row
rd.data <- rd.data[-1, ]
```

And now let's add a treatment status indicator

```
rd.data$treat <- as.numeric(rd.data$margin >=
    0)
```

ii. Create an RDD plot with margin in the x axis and vote share in 2004 in the y axis. You can recycle Hidalgo's code to do this.
Here you could have created the plot with the binned means or a plot showing all the data:

```
with(rd.data, plot(margin, vs_2004,
    cex = 0.4, pch = 16, col = "blue"))
abline(v = 0, col = "darkorange", lwd = 2)
```



iii. Some of the outcome data is missing. Assess whether it is missing at random or systematically. Discuss a few strategy to deal with the missing data.
First let's compare the missing data among winners and losers.

```
table(is.na(rd.data$vs_2004), rd.data$treat)
```

```
##
##             0    1
##   FALSE 5786 6000
##   TRUE   214    0
```

And close to the cutoff

```
table((is.na(rd.data$vs_2004)[abs(rd.data$margin) <
    5]), ((rd.data$treat)[abs(rd.data$margin) <
    5]))
```

```
##
##             0    1
##   FALSE 3146 3299
```

```
##   TRUE   153    0
```

In both cases, missingness seems independent from treatent assignment. We can use a t-test for a formal test of this claim.

```
with(rd.data, t.test(is.na(vs_2004) ~
    treat))  # entire dataset
##
##  Welch Two Sample t-test
##
## data:  is.na(vs_2004) by treat
## t = 14.896, df = 5999, p-value
## < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.03097269 0.04036065
## sample estimates:
## mean in group 0 mean in group 1
##      0.03566667      0.00000000
with(rd.data[abs(rd.data$margin) < 5,
    ], t.test(is.na(vs_2004) ~ treat))  # close to the cutoff
##
##  Welch Two Sample t-test
##
## data:  is.na(vs_2004) by treat
## t = 12.665, df = 3298, p-value
## < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.03919768 0.05355770
## sample estimates:
## mean in group 0 mean in group 1
##      0.04637769      0.00000000
with(rd.data[abs(rd.data$margin) < 3,
    ], t.test(is.na(vs_2004) ~ treat))  # closer
##
##  Welch Two Sample t-test
##
## data:  is.na(vs_2004) by treat
## t = 10.345, df = 2151, p-value
## < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.03841300 0.05638254
## sample estimates:
```

```
## mean in group 0 mean in group 1
##      0.04739777      0.00000000
```

```
with(rd.data[abs(rd.data$margin) < 1,
    ], t.test(is.na(vs_2004) ~ treat))
```

```
##
##  Welch Two Sample t-test
##
## data:  is.na(vs_2004) by treat
## t = 6.3154, df = 783, p-value
## = 4.513e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.03340387 0.06353491
## sample estimates:
## mean in group 0 mean in group 1
##      0.04846939      0.00000000
```

```
with(rd.data[abs(rd.data$margin) < 0.5,
    ], t.test(is.na(vs_2004) ~ treat))
```

```
##
##  Welch Two Sample t-test
##
## data:  is.na(vs_2004) by treat
## t = 5.1504, df = 417, p-value
## = 4.016e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.03698248 0.08263475
## sample estimates:
## mean in group 0 mean in group 1
##      0.05980861      0.00000000
```

iv. Use your RD function to estimate a LATE. Is there a party incumbency advantage?
Because our RD function was written to analize Hidalgo's data, we need to make it more general in order to be able to use it in this setting.
To make it more general, let's just write it using the outcome, running variable (centered at zero) and treatment indicator as inputs. In addition, we need to indicate the window to be used.

```
rd.estim <- function(outcome, running,
    treat, window) {

    library(car)
```

```
    wts <- ifelse(abs(running/window) <=
        1, 1 - abs(running/window),
        0)

    ll.results <- lm(outcome ~ treat *
        running, weights = wts)
    ll.out <- c(bandwidth = 15000, coef(ll.results)[[2]],
        sqrt(vcov(ll.results)[2, 2]),
        ll.results$df.residual + 2,
        coef(ll.results)[[1]])
    names(ll.out) <- c("Bandwidth",
        "RD Estimate", "Standard Error",
        "N", "Baseline")

    dm.results <- ttest(outcome[abs(running) <=
        window], treat[abs(running) <=
        window])
    dm.out <- c(dm.results[3], dm.results[4],
        dm.results[6], dm.results[2])
    names(dm.out) <- c("RD Estimate",
        "Standard Error", "N", "Baseline")

    list(ll.results = ll.out, dm.results = dm.out,
        sd = sd(outcome, na.rm = TRUE))
}

rd.estim(outcome = rd.data$vs_2004[is.na(rd.data$vs_2004) ==
    FALSE], running = rd.data$margin[is.na(rd.data$vs_2004) ==
    FALSE], treat = rd.data$treat[is.na(rd.data$vs_2004) ==
    FALSE], window = 1.5)
```

```
## $ll.results
##      Bandwidth    RD Estimate
##   1.500000e+04   3.340490e+00
## Standard Error              N
##   5.177815e-01   2.208000e+03
##       Baseline
##   3.543153e+01
##
## $dm.results
##    RD Estimate Standard Error
##      3.0111884      0.3070076
##              N       Baseline
##   2210.0000000     35.5125278
##
```

```
## $sd
## [1] 6.426603
```

There seems to be a positive incumbency advantage: winning office in 2000 increases electoral support in 2004.

v. Does this effect vary by party? Calculate heterogeneous treatment effects. Interpret your results.

```
# left
with(rd.data[is.na(rd.data$vs_2004) ==
    FALSE & rd.data$party == "left",
    ], rd.estim(outcome = vs_2004, running = margin,
    treat = treat, window = 1.5))
## $ll.results
##      Bandwidth    RD Estimate
##   1.500000e+04   8.428246e+00
## Standard Error              N
##   9.804139e-01   7.460000e+02
##       Baseline
##   3.549524e+01
##
## $dm.results
##    RD Estimate Standard Error
##      7.6781126      0.6136944
##              N       Baseline
##    748.0000000     35.8735091
##
## $sd
## [1] 7.708961
# center
with(rd.data[is.na(rd.data$vs_2004) ==
    FALSE & rd.data$party == "center",
    ], rd.estim(outcome = vs_2004, running = margin,
    treat = treat, window = 1.5))
## $ll.results
##      Bandwidth    RD Estimate
##   1.500000e+04   2.117672e+00
## Standard Error              N
##   7.898396e-01   7.140000e+02
##       Baseline
##   3.411405e+01
##
## $dm.results
##    RD Estimate Standard Error
##      0.5985120      0.4625107
```

```
##               N       Baseline
##    716.0000000     35.3065945
##
## $sd
## [1] 5.004998
```

```
# right
with(rd.data[is.na(rd.data$vs_2004) ==
    FALSE & rd.data$party == "right",
    ], rd.estim(outcome = vs_2004, running = margin,
    treat = treat, window = 1.5))
```

```
## $ll.results
##      Bandwidth    RD Estimate
##  15000.0000000     -0.2012538
## Standard Error               N
##      0.7117246    744.0000000
##       Baseline
##     36.6748178
##
## $dm.results
##    RD Estimate Standard Error
##      1.0127981      0.4369031
##               N       Baseline
##    746.0000000     35.3218672
##
## $sd
## [1] 4.892163
```

Note that these are a heterogeneous effect but they are not causal: it might be the case that a particular type of party runs more often in a particular kind of district, for example.

4. **There is a study group of 10 subjects in a randomized controlled experiment, in which 7 of the subjects are assigned at random to treatment and 3 are assigned to the control group. In answering the questions below, you may use R as a calculator where appropriate (except where noted below), but discuss your work. Observed data on the response variable look as follows:**

   (a) **Construct a box model for this experiment, drawing on our discussion of the Neyman urn model. What is in the box?**

   The box has 10 tickets; each ticket has two values, $Y_i(1)$ and $Y_i(0)$. We will draw seven tickets at random without replacement and put those tickets in the assigned-to-treatment group; the other three tickets go in the control group. We will observe $Y_i(1)$ for the tickets assigned to treatment, and $Y_i(0)$ for the tickets assigned to control. In the table, for instance, we see seven observed values of $Y_i(1)$ (the first seven subjects listed) and three observed values of $Y_i(0)$.

Assigned to Treatment	Assigned to Control
3	–
2	–
5	–
6	–
3	–
4	–
5	–
–	2
–	4
–	3

(b) **Define the average causal effect in terms of the model you constructed in (a). Then estimate the average causal effect, using the data in the table.**

The average causal effect is the difference between two other parameters: (i) the average Y_i if we assigned all 10 subjects to the treatment group, and (ii) the average Y_i if we assigned all 10 subjects to the control group. (This is also called the "intention-to-treat parameter," since it is the average causal effect of treatment assignment, though here we are not distinguishing explicitly between treatment assignment and treatment receipt.)

We can estimate the average Y_i we would observe if we assigned all 10 subjects to the treatment group by the average of the seven Y_i's that we did assign at random to the treatment group. The average outcome for subjects assigned to treatment is $\frac{3+2+5+6+3+4+5}{7} = 4$. Similarly, we can estimate the average Y_i we would observe if we assigned all 10 subjects to the treatment group by the average of the three Y_i's that we did assign at random to the treatment group. This is $\frac{2+4+3}{3} = 3$. The estimated average causal effect is the difference, that is, $4 - 3 = 1$.

(c) **Estimate the standard error of your estimate in (b). To do this, use the "conservative variance formula" discussed in readings and lectures.**

The number 4 in (c) is a realized value of the average $Y_i(1)$ in the treatment group, which could have turned out differently in a different experiment. We can find the variance of this random variable as follows. The true variance of the mean is the variance of the $Y_i(1)$s in the box, divided by n. Here, n is seven. But we don't know the variance of the $Y_i(1)$s in the box. To estimate this variance, we can use the variance of the $Y_i(1)$s in our sample (that is, the treatment group). In the formula for the variance of the $Y_i(1)$s in our sample, we should divide by $n - 1$, not n, because we are using the sample mean (not the mean of the box) to find the variance of the $Y_i(1)$s in our sample. Dividing by $n - 1$, the variance of the $Y_i(1)$s in the treatment group an unbiased estimator for the $Y_i(1)$s in the box. Thus, the estimated variance of the $Y_i(1)$s in the box is the sum of the squared deviations from the average outcome of 4, for each subject assigned to the treatment group, divided by $n - 1$. Here, $n - 1 = 7 - 1 = 6$. For the first subject in the table,

whose outcome is 3, the squared deviation from average is $(3-4)^2 = (-1)^2 = 1$. The outcomes for the other treatment subjects are $2, 5, 6, 3, 4$, and 5. Thus, the estimated variance of the $Y_i(1)$s in the box is

$$\frac{(-1)^2 + (-2)^2 + (1)^2 + (2)^2 + (-1)^2 + (0)^2 + (1)^2}{6} = \frac{12}{6} = 2. \tag{14}$$

We now need to divide by the number of observations to get the estimated variance of the treatment group mean. This is $2/7 \doteq 0.29$

Similarly, the estimated variance of the $Y_i(0)$s in the box is the sum of the squared deviations from the average outcome of 3, for each subject assigned to the control group, divided by $n-1$. In the control group, $n-1 = 3-1 = 2$. So, our estimator of the variance in the box is

$$\frac{(2-3)^2 + (4-3)^2 + (3-3)^2}{2} = \frac{2}{2} = 1. \tag{15}$$

Dividing by the number of observations to get the estimated variance of the control group mean, we have $1/3 \doteq 0.33$.

Now, the variance of the average in the treatment group, minus the average in the control group, is just the sum of the variances. This is the "conservative variance formula," where we pretend that we have two independent samples from the study group; footnotes 10, 11, 14, and 21 on pages A-32 to A-36 in Freedman, Pisani, and Purves (2007) show why this is legitimate. Now, recall that if A is one random variable and B is another, and A and B are independent, then $\text{var}(A - B) = \text{var}(A) + \text{var}(B)$.

Thus, the sum of the variances is about $0.29 + 0.33 = 0.62$. The standard error is the square root of the variance, namely, $\sqrt{0.62} \doteq 0.79$.

(d) **Suppose you want to know whether the estimated effect in (b) is statistically significant. First, explain carefully what is meant by "statistically significant." Use the words "null hypothesis" in your answer. Explain what is the null hypothesis, in terms of your box model in (a).**

Under the null hypothesis, the average of the $Y_i(1)$s in the study group (a.k.a. the experimental population—that is, the tickets in the box) equals the average of the $Y_i(0)$s in the study group. This is the so-called "weak" form of the null hypothesis (which is contrasted from the "strong" or "strict" null, which says $Y_i(1) = Y_(0)$ for every i).

A statistical significance test assesses whether an observed difference between the mean of the treatment sample and the mean of the control sample could plausibly be due to chance (sampling) error, under this null hypothesis. The p-value provides us with an answer to the following question: if the null hypothesis holds, what is the probability that we would observe a difference of means as large as the one we observe? The conventional level of a significance test is $p < 0.05$, meaning that in 95 out of 100 randomizations, if the null hypothesis holds, we will observe a p-value greater than 0.05. So given the model, if we see a p-value less than

0.05, there are two possibilities: (1) something unusual happened; or (2) the null hypothesis is false.

(e) **Should you use a t-test to assess statistical significance in this study? Why or why not? Explain carefully.**

Here, the study group is small; so far, so good for the t-test. However, the potential outcomes are probably not normally distributed, which is required for the t-test unless the samples are large (in which case the t-distribution converges to the normal distribution, and normal approximations apply by the central limit theorems). For evidence that the potential outcomes are probably not normally distributed, look at our sample of potential outcomes under treatment (in the treatment group) and our sample of potential outcomes under control (in the control group). So, using a t-test is probably not a great idea.

(f) **Conduct the t-test. What is the p-value?**

The t-statistic is the ratio of the sample difference of means to its estimated standard error. Thus we have $t = \frac{1}{0.79} = 1.27$. For the p-value, we need the (Satterthwaite) degrees of freedom and then we can refer the t-statistic to the appropriate t distribution. You can implement the t-test as follows using your R function "ttest":

```
y <- c(3, 2, 5, 6, 3, 4, 5, 2, 4, 3)
group <- c(1, 1, 1, 1, 1, 1, 1, 0, 0,
    0)
data <- as.data.frame(cbind(group, y))
with(data, t.test(y, group))

##
##  Welch Two Sample t-test
##
## data:  y and group
## t = 6.6712, df = 11.309,
## p-value = 3.066e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.013521 3.986479
## sample estimates:
## mean of x mean of y
##       3.7       0.7
```

This gives us $t = 1.271, df = 5.541, p = 0.255$—not strong evidence against the null hypothesis.

(g) **Suppose you instead want to use randomization inference to assess statistical significance. What null hypothesis do you need to assume? How does this differ from the null hypothesis you defined in (d)? How else does randomization inference differ from the t-test in (e)-(f)? Is**

randomization inference more appropriate than a t-test, and why or why not?

Here, you need to assume the strict null hypothesis: see the solution to 4(d). With randomization inference, we do not need to make parametric assumptions about the distribution of potential outcomes—i.e., that they are normally distributed. Randomization inference is thus likely more appropriate than the t-test here. We do need to switch the null hypothesis from the weak to the strong (strict) form; this may be a feature but could be a bug.

(h) **Now, use randomization inference in R to assess the statistical significance of the estimated effect in (b). (Here, use code from section or problem sets if possible, rather than an R package such as `ri`. Set the seed to 12345 before you run your code). Is your p-value the same or different as in part (f)?**

```
set.seed(12345)
ri <- function(treat, outcome)
fake_t <- sample(treat, length(treat),
    replace = F)

mean(outcome[fake_t == 1]) - mean(outcome[fake_t ==
    0])

rand_out <- replicate(10000, ri(treat = group,
    outcome = y))
sum(abs(rand_out) >= ttest(y, group)[3])/10000
```

From this sampling-based approach to randomization inference, we get a p-value of 0.33. This is not dramatically different from the t-test p-value of 0.255 in this case (though in other examples, the two p-values may differ much more substantially—and if we were closer to the threshold for statistical significance, difference of 0.07 in the p-value could be meaningful indeed).

(i) **Now, suppose an investigator assumes the OLS model:**

$$Y_i = \alpha + \beta T_i + \epsilon_i, \tag{16}$$

where T_i is a 0-1 variable, with 1 indicating that a subject was assigned to treatment. Make a list of the "usual OLS assumptions."

Here are the usual OLS assumptions:

i. $Y_i = \alpha + \beta T_i + \epsilon_i$ (the statistical model holds).
ii. $T_i \perp\!\!\!\perp \epsilon_i$.[2]
iii. The ϵ_i are independent and identically distributed, with $E(\epsilon_i) = 0$ and $\text{var}(\epsilon_i) = \sigma^2$.

[2] One might denote the design matrix as X with typical row $[1 \;\; T_i]$; the OLS assumption is that $\epsilon_i \perp\!\!\!\perp X$. However, the random vector ϵ_i is trivially independent of the constant vector $(1, \ldots, 1)'$.

iv. The vector of T_i's is linearly independent of the constant vector.

If we're inferring causation, we also need a response schedule, which tells us how unit i responds to being assigned to treatment or control:

5. $Y_{i,T_i} = \alpha + \beta T_i + \epsilon_i$.

(j) **What are the differences between the Neyman urn model in part (a) and the OLS model in part (e)? What assumptions are shared by the models?**

For the causal assumptions: with the Neyman model, each unit has its own causal effect $Y_i(1) - Y_i(0)$, while in the regression model there is a single effect β which applies to all i: that is one major difference. For the similarities: both models assume that response of unit i depends only on assignment of unit i, not on the assignment of other units (this is called non-interference or sometimes SUTVA).

As for the statistical assumptions: with the Neyman urn model, we are sampling units into treatment and control without replacement; and the treatment groups are independent. Analysis under this model adjusts naturally for heteroskedasticity (unequal variances of the treatment and control group means), because we estimate the variance of the potential outcomes under treatment and under control separately, using the observed variances and sample sizes of the treatment and control groups. For the regression model, things are very different: for example, ϵ_i is assumed i.i.d.

(k) **Do you think the usual OLS assumptions are satisfied in this problem? Why or why not? Which assumptions are the most plausible? What assumptions are less plausible? Explain your answers carefully.**

You might quibble with the assumption that $Y_i = \alpha + \beta T_i + \epsilon_i$: for example, the existence of an additive random error term is not guaranteed by the experiment or the randomization. However, if you are willing to accept the equation, it seems plausible that $T_i \perp\!\!\!\perp \epsilon_i$; after all, subjects are assigned at random to treatment and control.

On the other hand, the distributional properties of the error term seem more suspect. For example, why is $\text{var}(\epsilon_i) = \sigma^2$ the same for all subjects? The formula for the variance of a difference-of-means adjusts automatically for heteroskedasticity–that is, unequal variances given T_i–in the treatment and control group. On the other hand, the usual OLS assumptions imply homoskedasticity. As we will see in 4(q), this can make a real difference to the estimated standard errors.

(A note on the usual full rank condition for OLS: since there is only one independent variable here, plus the constant, T_i will be linearly independent of the constant vector as long as there is some variance in T_i, which is guaranteed by the experimental design: seven subjects go to treatment and 3 to control. So the full rank condition is clearly satisfied).

(l) **Under the OLS model, what is $E(Y_i|T_i = 0)$? How about $E(Y_i|T_i = 1)$?**

Under the model, $E(Y_i|T_i = 0) = E(\alpha + \beta T_i + \epsilon_i|T_i = 0) = \alpha + E(\epsilon_i|T_i = 0) = \alpha$, because $T_I = 0$ so β drops out, and $E(\epsilon_i|T_i) = 0$. Similarly, $E(Y_i|T_i = 1) = \alpha + \beta$.

(m) **Denote the design matrix as X. What is a typical row of this matrix? What size is X? Denote the response variable as Y. What size is Y?**

The design matrix X is 10×2; it has typical row $[1 \;\; T_i]$. Y is 10×1.

(n) **Calculate $X'X$, $(X'X)^{-1}$, $X'Y$, and $(X'X)^{-1}X'Y$. Use $(X'X)^{-1}X'Y$ to estimate α and β. Here, you should work out the matrices by hand (e.g., don't use R commands such as `solve`).**

We'll write the second column of X with the seven subjects assigned to treatment stacked at the top, and the three subjects assigned to control at the bottom. The seven at the top thus have $T_i = 1$; the three at the bottom have $T_i = 0$. Thus,

$$(X'X) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \tag{17}$$

$$= \begin{pmatrix} 10 & 7 \\ 7 & 7 \end{pmatrix}. \tag{18}$$

Now, the determinant of $X'X$ is $(10)(7) - (7)(7) = 21$; the adjoint is

$$\begin{pmatrix} 7 & -7 \\ -7 & 10 \end{pmatrix}. \tag{19}$$

So we have

$$(X'X)^{-1} = \frac{1}{21} \begin{pmatrix} 7 & -7 \\ -7 & 10 \end{pmatrix}. \tag{20}$$

You can verify that $(X'X)^{-1}(X'X) = I_{2\times 2}$, as it should be. Now,

$$Y = \begin{pmatrix} 3 \\ 2 \\ 5 \\ 6 \\ 3 \\ 4 \\ 5 \\ 2 \\ 4 \\ 3 \end{pmatrix}. \tag{21}$$

(Remember, the treatment observations are stacked on top, and the control observations are stacked on the bottom). Thus,

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 5 \\ 6 \\ 3 \\ 4 \\ 5 \\ 2 \\ 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 37 \\ 28 \end{pmatrix} \tag{22}$$

Finally,

$$(X'X)^{-1}X'Y = \frac{1}{21} \begin{pmatrix} 7 & -7 \\ -7 & 10 \end{pmatrix} \begin{pmatrix} 37 \\ 28 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}. \tag{23}$$

Thus, the OLS estimates of a and b are $\hat{a} = 3$ and $\hat{b} = 1$.

(o) **Express $(\hat{Y}|T_i = 1) - (\hat{Y}|T_i = 0)$ in terms of your estimates $\hat{\alpha}$ and/or $\hat{\beta}$. How does this difference compare to your answer in (b)? Comment briefly.**

$(\hat{Y}|T_i = 1) = \hat{\alpha} + \hat{\beta}$ and $(\hat{Y}|T_i = 0) = \hat{\alpha}$. Thus, $(\hat{Y}|T_i = 1) - (\hat{Y}|T_i = 0) = \hat{\beta} = 1$. This is exactly the difference-of-means we calculated in (c).

Lesson: When we have a dummy (0-1) independent variable and no control variables, an OLS regression of the response on a constant and the dummy variable gives the difference of mean outcomes. Here, the estimated constant is the average response in the control group, that is, $(\hat{Y}|T_i = 0) = \hat{\alpha} = 3$. The average response in the treatment group is $(\hat{Y}|T_i = 1) = \hat{\alpha} + \hat{\beta} = 3 + 1 = 4$.

(p) **Now use the usual OLS formula to attach estimated standard errors to $\hat{\alpha}$ and $\hat{\beta}$. (Here, you can use R matrix commands such as `solve` if needed—but calculate terms explicitly and explain your work).**

The residual is the difference between the actual and fitted ("predicted") value. For example, for the first subject the actual value is 3, while the fitted value is $\hat{\alpha} + \hat{\beta} = 4$. (This subject was assigned to treatment). So the residual is $3 - 4 = -1$.

The vector of residuals is

$$e = \begin{pmatrix} -1 \\ -2 \\ 1 \\ 2 \\ -1 \\ 0 \\ 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \tag{24}$$

and the sum of squared residuals is

$$e'e = \begin{pmatrix} -1 & -2 & 1 & 2 & -1 & 0 & 1 & -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} -1 \\ -2 \\ 1 \\ 2 \\ -1 \\ 0 \\ 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} = 14. \tag{25}$$

The usual OLS formula for the estimated variance-covariance matrix of $\hat{\gamma} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$ is

$$\text{cov}(\hat{\gamma}) = \hat{\sigma}^2 (X'X)^{-1}, \tag{26}$$

where $\hat{\sigma}^2 = \frac{e'e}{n-p}$. Here, $n - p = 10 - 2 = 8$, so $\hat{\sigma}^2 = \frac{14}{8} = 1.75$.
Taking $(X'X)^{-1}$ from (h), we have

$$\hat{\sigma}^2 (X'X)^{-1} = \frac{1.75}{21} \begin{pmatrix} 7 & -7 \\ -7 & 10 \end{pmatrix} \doteq \begin{pmatrix} 0.583 & -0.583 \\ -0.583 & 0.833 \end{pmatrix} \tag{27}$$

The standard errors of $\hat{a}$ and $\hat{b}$ are the square roots of the diagonal elements of (27), namely, $\text{SE}_{\hat{a}} = \sqrt{0.583} = 0.764$ and $\text{SE}_{\hat{b}} = \sqrt{0.833} = 0.913$.

(q) **Attach a standard error to the difference $(\hat{Y}|T_i = 1) - (\hat{Y}|T_i = 0)$ you found in (o). How does this compare to your estimated standard error in (c)? Explain why they are the same or different.**

We showed in (0) that $(\hat{Y}|T_i = 1) - (\hat{Y}|T_i = 0) = \hat{\beta}$, so by (p) the standard error for the difference is 0.913. This is fairly different from the standard error we found in (c), namely, $\sqrt{3} = 0.79$.

Lesson: If the Neyman statistical model is right, the OLS nominal standard errors (those computed from the usual regression formulas) can be misleading. The problem is that the OLS formulas assume the error term is *i.i.d.* (in particular, they assume homoskedasticity), whereas the "conservative variance formula" adjusts naturally for unequal variances in the treatment and control groups (which may be due e.g. to their unequal sizes). The OLS formulas can be adjusted to account for unequal variances; indeed, "robust" (Huber-White) standard errors do the trick. But the ability to recover the estimates under the Neyman urn model does not mean that the regression model is the natural place to start; when things get more complicated, they may go even further off the rails.

5. **The *statistical power* of a test is the probability that it will reject the null hypothesis, given that the null hypothesis is false. In this question, you are asked to calculate statistical power in an experiment.**

   Thus, consider Tables 1 and 4 in Dunning and Harrison (2010). The first row of Table 4 compares respondents' evaluations of co-ethnic politicians who either are or are not their joking cousins. Dunning and Harrison report an estimated treatment effect of $\hat{\tau} = 0.49$ on their 1-7 scale, with an estimated standard error of 0.22. In the questions below, take 0.22 as your best guess of the true standard error of $\hat{\tau}$. Suppose that the Ns in the first row of Table 1 are large enough that a central limit theorem approximately applies. (You can use R freely if needed to answer any of the following questions).

   (a) **Fill in the following question marks with the correct answers: under the null hypothesis, the distribution of $\hat{\tau}$ is ?(?, ?).**

   Under the null hypothesis, the distribution of $\hat{\tau}$ is $N(0, 0.048)$. (Note that the variance is the square of the standard error: $0.22^2 \doteq 0.048$).

   (b) **Suppose that the true treatment effect is indeed $\tau = 0.49$. Fill in the question marks with the correct answers: under this alternative, the distribution of $\hat{\tau}$ is ?(?, ?).**

   Under this alternative, the distribution of $\hat{\tau}$ is $N(0.49, 0.048)$.

   (c) **Now suppose that we repeated Dunning and Harrison's experiment infinitely many times under identical conditions (just suppose). For each experiment, we use a z-score to conduct a test of the null hypothesis. For what fraction of these experiments would we reject the null hypothesis that $\tau = 0$, if in truth $\tau = 0.49$?**

   We reject the null hypothesis whenever $|z| > 1.96$. Now, under the null hypothesis,$z = 1.96$ corresponds to a realized $\hat{\tau}$ of $1.96 * 0.22 \doteq 0.43$.[3]

   We want to convert this value to standard units *relative to the mean under the alternative hypothesis*—because we want to know how often we will observe values

[3]We can safely ignore the lower tail of the distribution—the realizations of $\hat{\tau}$ such that $z < -1.96$—since these realizations are very unlikely under the alternative hypothesis: since 0 is $0 - 0.49/22 = -2.23$ standard errors below the mean of 0.49, realizations less than -1.96 are -2.23-1.96=-4.19 standard errors below the mean.

of $\hat{\tau}$ that lead us to reject the null hypothesis that $\tau = 0$, given that in fact $\tau = 0.49$. To convert to standard units, subtract the mean and divide by the standard error: $(0.43 - 0.49)/0.22 \doteq -0.27$. Thus, we reject the null hypothesis that $\tau = 0$ whenever the realization of $\hat{\tau}$ is greater than -0.27 standard errors *below* 0.49—the mean of the distribution under the alternative hypothesis.

How often will $\hat{\tau}$ be greater than -0.27 standard units below 0.49? The standard normal distribution answers this question:

$$1 - \Phi(-.27) = 0.61 \tag{28}$$

since $\Phi(-.27) = 0.39$. Thus, we would reject the null hypotheses that $\tau = 0$, if in truth $\tau = 0.49$, in about 0.61 of the experiments. The statistical power of the test is 0.61 or 61%.

(d) **Now suppose instead that the true treatment effect is 1.0. What is the power of the test? Compare your answer to (c). Do you have more power against smaller effects or bigger effects?**

If the true treatment effect is 1.0, then the realized value of $\hat{\tau}$ that leads us to reject the null—0.43—is now $(0.43 - 1)/0.22 \doteq -2.59$ standard errors below the mean of the normal distribution, under the alternative hypothesis. How often will $\hat{\tau}$ be greater than -2.59 standard units below 0.49? The answer is

$$1 - \Phi(-2.59) = 0.995, \tag{29}$$

since $\Phi(-.27) = 0.005$. Thus, we would reject the null hypotheses that $\tau = 0$, if in truth $\tau = 0.49$, in nearly 100% of the experiments.

Lesson: statistical power depends on the true effect size. It is easy to discern a very large effect, and much harder to discern smaller effects. If effects are smaller, you need a bigger N to compensate.

(e) **Now, suppose Dunning and Harrison are filling out a grant application before they do their study. The funders ask them to calculate the power of their test. They don't know the true effect size but they guess (based on the extensive previous experimental literature on cousinage) that the standard error of the estimated effect is 0.22. Then, they consider two alternative proposals: one in which they presume a true effect of 0.49, and another in which they presume an effect of 1.0. Which proposal do you think the donor would be more likely to fund, and why?**

It doesn't seem very likely that a grant agency would fund a study where the probability of detecting a significant effect is nearly 1. If that is true, why do the study in the first place?

(Note: we are implicitly fixing the study group size at the N in Dunning and Harrison (2010). A natural alternative is to decrease the size of the study group—say, until power against the null hypothesis reaches 80%, under the alternative that the true effect is 1.0. But here we are fixing the sample size and asking which

proposal would be more likely to be funded).

6. **We will now turn to an example in R in which we use simulations to evaluate the statistical power of an experiment, as a function of sample size. To do this, you will fix the N and the effect size and simulate 500 hundred experiments and calculate the proportion for which we can reject the null hypothesis. You will then vary the N to see how this proportion varies. Before running your, set the seed to 54321.**

   (a) **You first need to create your assumed "box" for the simulation. You will start by fixing N** $= 100$**. Now, use R to draw the control potential outcomes from a normal distribution with** mean $= 0$ **and** sd $= 1$**. To construct the potential outcomes under treatment, add** $.25$ **to each control potential outcome.**

   (b) **Now take the box as fixed and:**

      i. **generate a treatment vector by randomly assigning half of the units to the treatment group and the other half to control.**

      ii. **get the "observed" outcomes under that vector.**

      iii. **use your t-test function to test the hypothesis that the ATE is different from zero.**

      iv. **store an indicator of whether you reject the null hypothesis (take** $\alpha = 0.05$**).**

   (c) **Repeat the procedure in (b) 500 times. Indicate the power of your test.**

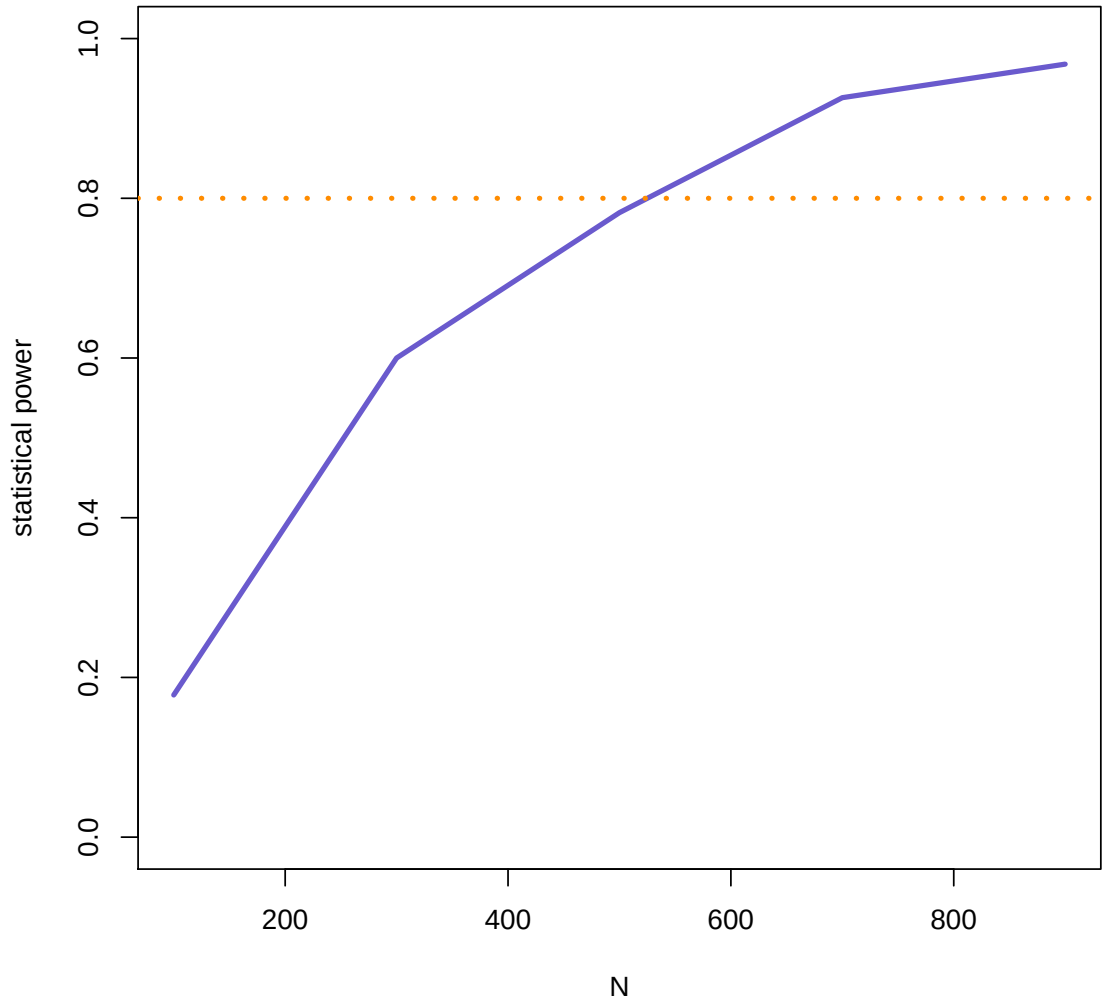
   (d) **Repeat (a)-(c) for N=** 300**, N=** 500**, N=** 700**, and N=** 900**.**

   (e) **Plot statistical power as a function of sample size, using results from (b)-(d).**

```r
set.seed(54321)

# The sample sizes we'll be
# considering
possible.ns <- c(100, 300, 500, 700,
    900)

# Empty object to collect simulation
# estimates
powers <- rep(NA, length(possible.ns))

# Standard significance level
alpha <- 0.05

# Number of simulations to conduct
```

```r
# for each N
sims <- 500

#### Outer loop to vary the number of
#### subjects ####
for (j in 1:length(possible.ns)) {

    N <- possible.ns[j]  # Pick the jth value for N

    # control potential outcomes
    Y0 <- rnorm(N, 0, 1)
    # treatment potential outcome
    Y1 <- Y0 + 0.25

    # Empty object to count significant
    # experiments
    significant.experiments <- rep(NA,
        sims)

    #### Inner loop to conduct experiments
    #### 'sims' times over for each N
    for (i in 1:sims) {

        # Do a random assignment
        Z.sim <- rbinom(n = N, size = 1,
            prob = 0.5)  # Observed outcomes
        Y.sim <- Y1 * Z.sim + Y0 * (1 -
            Z.sim)
        # Extract p-values
        p.value <- t.test(Y.sim ~ Z.sim)$p.value
        # Determine significance according
        # to p <= 0.05
        significant.experiments[i] <- (p.value <=
            alpha)

    }
    # store average success rate (power)
    # for each N
    powers[j] <- mean(significant.experiments)
}

powers


## [1] 0.178 0.600 0.782 0.926 0.968
```

```r
plot(possible.ns, powers, type = "l",
    col = "slateblue", lwd = 3, ylim = c(0,
        1), ylab = "statistical power",
    xlab = "N")
abline(h = 0.8, lty = 3, col = "darkorange",
    lwd = 3)
```



7. **Professor Smedley is interested in the effects of owning property on policy attitudes. Smedley theorizes that owning property will make people more supportive of market-based economic policies. After conducting extensive fieldwork in Guatemala, he believes he has found the perfect natural experiment. Following the passage of a new law in parliament, squatters in certain houses in Guatemala City were given formal titles to their land. The new**

law stated that all squatters could go to the municipal office and request property rights. According to Smedley's research, the officials would then grant property rights in a haphazard (as if random) manner. Smedley conducts a survey of some squatters. He only includes in his final dataset those squatters who applied for property rights. Smedley collects the following variables (which he has made publicly available in 'Smedley.RData").

```
x1 - Indicator for being awarded property rights at time t0
x2 - Indicator for if the applicant just ahead in line of the squatter
 was granted property rights at time t0.
x3 - Fraction of household income from agriculture at time t0
x4 - Indicator for minority status  at time t0
x5 - Indicator for TV ownership at time t0
x6 - Indicator for living in southern neighborhood at time t0
x7 - Indicator for voted for Mayor at time t0
y1 - A measure of support for liberal economic policies at time t1.
y2 - Neighbor's support for liberal economic policies at time t1.
```

Smedley assumes the Neyman model. To estimate the average causal effect, he regresses $y1$ on $x1$ and finds that being awarded property rights increases support for liberal economic policies by about 2.7 units. Note that `t1` is 2 months after `t0`.

Your task is to evaluate the strength of Smedley's design using balance and placebo tests. Do the assumptions of Smedley's analysis hold?

Under the Neyman model, the assumptions are as-if random treatment assignment, the exclusion restriction, and non-interference. As usual, we can never prove these assumptions, but we can test them. The data Smedley collected allows us to test as-if random and non-interference.

To test as-if random, we can first start by doing individual balance tests. For this, we will run individual t-tests using all pre-treatment covariates.

```r
# Individual balance tests
with(data, t.test(x2 ~ x1))
```

```
##
##  Welch Two Sample t-test
##
## data:  x2 by x1
## t = -1.1883, df = 297.65,
## p-value = 0.2357
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.17707467  0.04374134
```

```
## sample estimates:
## mean in group 0 mean in group 1
##       0.5866667       0.6533333

with(data, t.test(x3 ~ x1))

##
##  Welch Two Sample t-test
##
## data:  x3 by x1
## t = -0.71345, df = 296.69,
## p-value = 0.4761
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.09099086  0.04257089
## sample estimates:
## mean in group 0 mean in group 1
##       0.4798949       0.5041049

with(data, t.test(x4 ~ x1))  # This one is imbalanced

##
##  Welch Two Sample t-test
##
## data:  x4 by x1
## t = -2.3837, df = 291.22,
## p-value = 0.01778
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.21907853 -0.02092147
## sample estimates:
## mean in group 0 mean in group 1
##            0.20            0.32

with(data, t.test(x5 ~ x1))

##
##  Welch Two Sample t-test
##
## data:  x5 by x1
## t = -0.23634, df = 297.99,
## p-value = 0.8133
## alternative hypothesis: true difference in means is not equal to 0
```

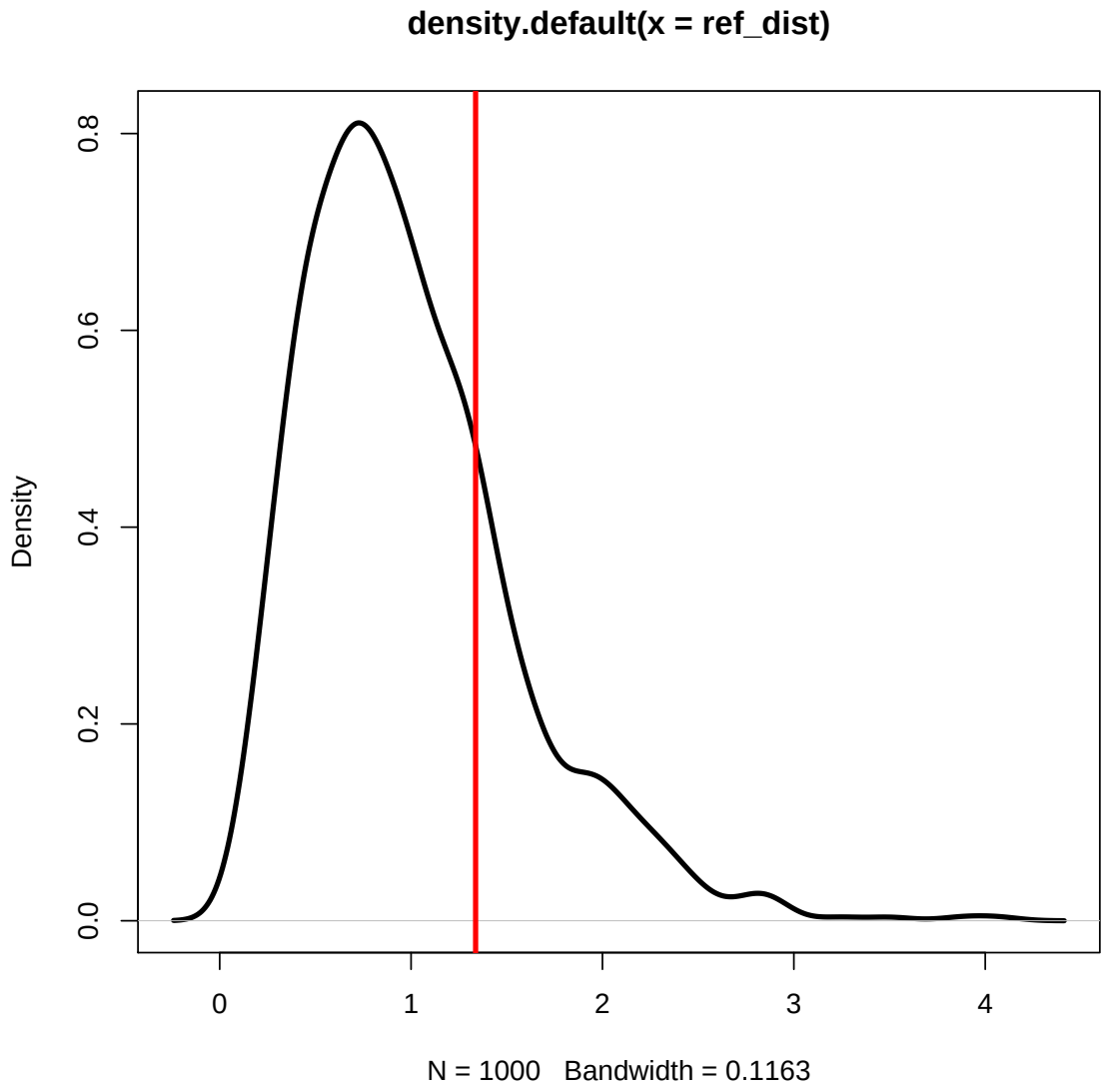
```
## 95 percent confidence interval:
##  -0.12435644  0.09768978
## sample estimates:
## mean in group 0 mean in group 1
##       0.6066667       0.6200000

with(data, t.test(x6 ~ x1))

##
##  Welch Two Sample t-test
##
## data:  x6 by x1
## t = 1.0429, df = 296.79,
## p-value = 0.2979
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.04731083  0.15397750
## sample estimates:
## mean in group 0 mean in group 1
##       0.7600000       0.7066667

with(data, t.test(x7 ~ x1))

##
##  Welch Two Sample t-test
##
## data:  x7 by x1
## t = 0.25452, df = 297.94,
## p-value = 0.7993
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.08975876  0.11642543
## sample estimates:
## mean in group 0 mean in group 1
##       0.2933333       0.2800000
```

One of the variables–$x4$, the indicator for minority status at t_0–is imbalanced. We know that we can expect 1 out of every 20 covariates to be imbalanced out of random chance. Yet this is a setting with only 6 covariates. As a complementary analysis, we can do an omnibus test, and see if the entire set of pre-treatment covariates can predict treatment assignment. For this, we can run a regression to see if the entire set of pre-treatment covariates can predict treatment assignment. We can use the F-test to assess this claim.

```
summary(with(data, lm(x1 ~ x2 + x3 +
    x4 + x5 + x6 + x7)))

##
## Call:
## lm(formula = x1 ~ x2 + x3 + x4 + x5 + x6 + x7)
##
## Residuals:
##      Min       1Q   Median
## -0.69953 -0.46627 -0.04116
##       3Q      Max
##  0.51849  0.61360
##
## Coefficients:
##             Estimate Std. Error
## (Intercept)  0.451262   0.090872
## x2           0.057035   0.060098
## x3           0.047080   0.098980
## x4           0.148706   0.066156
## x5           0.007537   0.059464
## x6          -0.067538   0.065401
## x7          -0.012366   0.064248
##             t value Pr(>|t|)
## (Intercept)   4.966 1.16e-06 ***
## x2            0.949   0.3434
## x3            0.476   0.6347
## x4            2.248   0.0253 *
## x5            0.127   0.8992
## x6           -1.033   0.3026
## x7           -0.192   0.8475
## ---
## Signif. codes:
##   0 '***' 0.001 '**' 0.01 '*'
##   0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4991 on 293 degrees of freedom
## Multiple R-squared:  0.02665,Adjusted R-squared:  0.006722
## F-statistic: 1.337 on 6 and 293 DF,  p-value: 0.2404
```

The p-value for the F-test using the chi-square distribution is .24, which means we cannot reject the null hypothesis that pre-treatment covariates cannot predict treatment assignment. We can complement this analysis finding the p-value of the permutation F-test which makes less distributional assumptions.

```r
# Permutation F-test
observed <- summary(with(data, lm(x1 ~
    x2 + x3 + x4 + x5 + x6 + x7)))$fstatistic[1]

# Reference distribution:
perm_ftest <- function(data) {
    data$x1 <- sample(data$x1, length(data$x1),
        replace = F)
    summary(with(data, lm(x1 ~ x2 +
        x3 + x4 + x5 + x6 + x7)))$fstatistic[1]
}
ref_dist <- replicate(1000, perm_ftest(data))

plot(density(ref_dist), lwd = 3)
abline(v = observed, col = "red", lwd = 3)

# p-value
sum(ref_dist >= observed)/length(ref_dist)

## [1] 0.223
```



In this case, the p-value for the permutation F-test is .25, very close from the one we got looking at the "table" for the chi-squared distribution. Altogether, the analysis so far provides evidence that supports the as-if random assumption. Note: some of you chose to exclude x_2 from the F-test which is also correct. In that case the F-statistic is 1.425 and its p-value is around .21.

There are two pos-treatment variables, y_1, which is Smedley's main outcome, and y_2 a measure of neighbor's support for liberal economic policies in time t_1. We can do two types of placebo tests. First, we can use x_2, whether the individual right before in line got property rights, as a placebo treatment and evaluate whether it has an effect on y_1. This is both a placebo test of as-if random and of spillover effects. Finally, we can use y_2 as a placebo outcome, since a treatment effect on this variable would be evidence of a SUTVA violation.

```r
# test 1
with(data, t.test(y1 ~ x2))

##
##  Welch Two Sample t-test
##
## data:  y1 by x2
## t = -0.47719, df = 230.01,
## p-value = 0.6337
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5118636  0.3122700
## sample estimates:
## mean in group 0 mean in group 1
##        1.867115        1.966911

# test 2
with(data, t.test(y2 ~ x1))

##
##  Welch Two Sample t-test
##
## data:  y2 by x1
## t = -7.859, df = 296.21,
## p-value = 7.225e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.2481669 -0.7482373
## sample estimates:
## mean in group 0 mean in group 1
##       0.5925087       1.5907108
```

Whereas the design survives the first placebo test, the second test suggests there are some spillover effects and thus provides evidence against the assumption of non-interference. Putting all the evidence together, we can conclude that although there is supporting evidence for the hypothesis that treatment assignment was as-if random, the estimated effect will be biased due to spillover.