

Juanjo Carin

W241 (Field Experiments) – Problem Set #2 – MIDS Spring 2015

February 9, 2015

Contents

W241 – Problem Set #2	1
1. FE, exercise 3.6.	1
2. FE, exercise 3.7.	3
3. FE, exercise 3.8.	4
4. FE, exercise 3.11.	9
5. Online advertising.	10
6. Auction experiment.	12

W241 – Problem Set #2

1. FE, exercise 3.6.

```
lottery_table <- data.frame(ratings = seq(-12, 12),
                           control = c(rep(0, 3), 0.22, rep(0, 2), 0.45, 0,
                                         0.45, 0, 1.12, 1.56, 27.23, 18.30,
                                         24.33, 8.48, 5.80, 3.35, 3.79, 2.23,
                                         0.89, 0.22, 0.45, 0.67, 0.45),
                           treatment = c(0.2, rep(0, 3), 0.20, 0,
                                          rep(0.20, 2), 0.59, 0.20, 0.98,
                                          2.75, 18.63, 13.14, 25.29, 10.98,
                                          9.61, 3.92, 7.25, 2.55, 1.37,
                                          0.78, 0, 0.20, 0.98))

N_control <- 448
N_treatment <- 510
control <- numeric()
treatment <- control
for (i in 1:25) {
  control <- c(control, rep(lottery_table$ratings[i],
                           round(lottery_table$control[i]/100*N_control)))
  treatment <- c(treatment, rep(lottery_table$ratings[i],
                                round(lottery_table$treatment[i]/100*
                                      N_treatment)))
}
lottery <- data.frame(group = c(rep(0, N_control), rep(1, N_treatment)),
                     rating = c(control, treatment))
est.ate <- function(outcome, treat)
```

```

mean(outcome[treat==1]) - mean(outcome[treat==0])
ATE_lottery <- est.ate(lottery$rating, lottery$group)

distribSNH_lottery <- replicate(10e3,
                                est.ate(lottery$rating, sample(lottery$group)))
plot(density(distribSNH_lottery))
abline(v = ATE_lottery)

```

density.default(x = distribSNH_lottery)

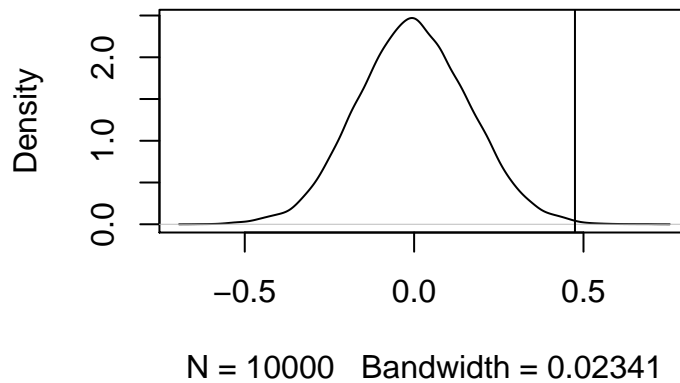


Figure 1: Lottery: Sampling distribution of the ATE under the SNH

```
(pvalue_lottery <- mean(ATE_lottery <= distribSNH_lottery))
```

```
## [1] 0.0019
```

```
sum(ATE_lottery <= distribSNH_lottery)
```

```
## [1] 19
```

```
sum(abs(ATE_lottery) <= abs(distribSNH_lottery))
```

```
## [1] 43
```

How many of the simulated random assignments generate an estimated ATE that is at least as large as the actual estimate of the ATE?

This number is calculated above: it is **19** (which is the p -value obtained above times the number of random assignments that were simulated).

What is the implied one-tailed p -value?

As mentioned, it is that number divided by 10,000 (the number of random assignments that were simulated): **0.0019.**

How many of the simulated random assignments generate an estimated ATE that is at least as large *in absolute value* as the actual estimate of the ATE?

This number is also calculated above: **43** (it is about twice the previous value because the sampling distribution is symmetric, not skewed; it is actually close to normal, due to the Central Limit Theorem).

What is the implied two-tailed p -value?

It is that number divided by 10,000 (the number of random assignments that were simulated): **0.0043.** The associated hypothesis in this case would be that winning the visa lottery has an effect on the views toward people from other countries, either positive or negative.

2. FE, exercise 3.7.

Our hypothesis is that the treatment effect $\tau_i = 2 \forall i$. Hence, the null hypothesis is $\tau_i \leq 2 \forall i$ (and the alternate hypothesis is $\tau_i > 2 \forall i$).

```
diet <- data.frame(group = c(rep(1, 5), rep(0, 5)), loss = c(2, 11, 14, 0, 3,
                                                             1, 0, 0, 4, 3))
tau <- 2
```

So rather than trying to detect any difference > 0 between treatment and control groups, we aim to detect a difference of at least 2 pounds. So we subtract 2 from each outcome in the treatment group:

```
dietSNH <- diet
dietSNH$loss <- ifelse(dietSNH$group == 1, dietSNH$loss - tau, dietSNH$loss)
```

And then use randomization inference—simulating 200 possible random assignments, though the whole number of them would not be so large in this particular case: $\binom{20}{10} = 252$ —to calculate the p -value, i.e., how likely an observation like the one we have is, just by chance.

```
(ATE <- est.ate(dietSNH$loss, diet$group))
```

```
## [1] 2.4
```

```
distribSNH_diet <- replicate(200, est.ate(dietSNH$loss, sample(dietSNH$group)))
library(ri)
perms <- genperms(dietSNH$group, maxiter = choose(20, 10))
distribSNH_diet2 <- apply(perms, 2, function(x) est.ate(dietSNH$loss, x))
plot(density(distribSNH_diet))
abline(v = ATE)
```

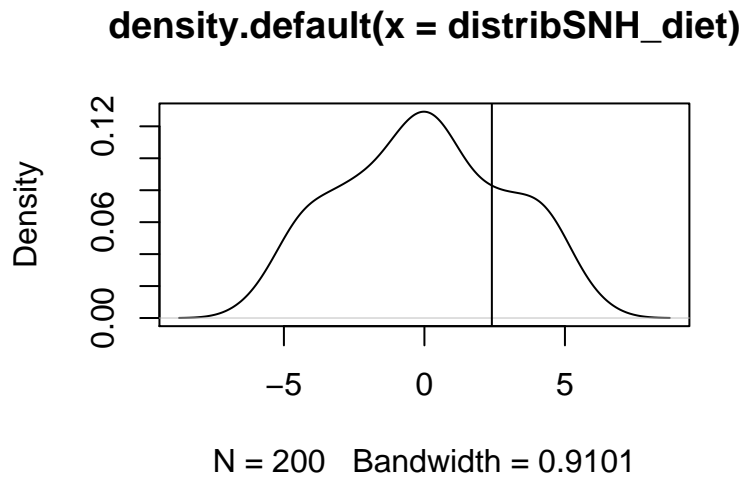


Figure 2: Diet: Sampling distribution of the ATE under the SNH

```
(pvalue_diet <- mean(ATE <= distribSNH_diet))
```

```
## [1] 0.27
```

We use a **one-sided test** (because our hypothesis is directional, i.e., the difference—with respect to $\tau_i = 2$ —is positive) and get a **p-value = 0.27** (had we use all possible random assignments, the value we would have obtained is 0.2380952). I.e., it would not be so unlikely (probability = 23.8%) to get the ATE we have observed—the diet program may have an effect, but we fail to reject the hypothesis that its effect is as large as more than 2 pounds of weight loss.

3. FE, exercise 3.8.

(a) For each state, estimate the effect of having a two-year term on the number of bills introduced.

```
library(foreign) # package allows R to read Stata datasets
senator <- read.dta("Titiunik_WorkingPaper_2010.csv.dta")
senator$state <- factor(senator$texas0_arkansas1, levels = 0:1,
                       labels = c("Texas", "Arkansas"))
senator$group <- factor(senator$term2year, levels = 0:1,
                       labels = c("Control", "Treatment"))
names(senator)[1:2] <- c("t2yr", "bills")
(ATE_senator_texas <- est.ate(senator$bills[senator$state == "Texas"],
                             senator$t2yr[senator$state == "Texas"]))
```

```
## [1] -16.74167
```

```
(ATE_senator_arkansas <- est.ate(senator$bills[senator$state == "Arkansas"],
                                senator$t2yr[senator$state == "Arkansas"]))
```

```
## [1] -10.09477
```

(b) For each state, estimate the standard error of the estimated ATE.

The first possibility is to apply equation (3.6), page 61 of FE:

$$\widehat{SE} = \sqrt{\frac{\widehat{Var}(Y_i(0))}{N-m} + \frac{\widehat{Var}(Y_i(1))}{m}}$$

```
(SE_senator_texas <-
  sqrt(var(senator$bills[senator$state == "Texas" & senator$t2yr == 0]) /
        length(senator$bills[senator$state == "Texas" & senator$t2yr == 0]) +
        var(senator$bills[senator$state == "Texas" & senator$t2yr == 1]) /
        length(senator$bills[senator$state == "Texas" & senator$t2yr == 1])))
```

```
## [1] 9.345871
```

```
(SE_senator_arkansas <-
  sqrt(var(senator$bills[senator$state == "Arkansas" & senator$t2yr == 0]) /
        length(senator$bills[senator$state == "Arkansas" &
                                senator$t2yr == 0]) +
        var(senator$bills[senator$state == "Arkansas" & senator$t2yr == 1]) /
        length(senator$bills[senator$state == "Arkansas" &
                                senator$t2yr == 1])))
```

```
## [1] 3.395979
```

Another possibility is to obtain the SE given by the regression. The estimates will not be so accurate.

```
regression <- function(dataframe, ATE){
  reg <- summary(lm(dataframe[, 2] ~ dataframe[, 1]))
  estimate <- reg$coefficients[2,1]
  standard.error <- reg$coefficients[2,2]
  lower.bound <- estimate - standard.error * 1.96
  upper.bound <- estimate + standard.error * 1.96
  estimate.in.ci <- lower.bound < ATE & upper.bound > ATE
  return(SE = standard.error)
}
(SE_senator_texas2 <- regression(senator[senator$state == "Texas", ],
                                ATE_senator_texas))
```

```
## [1] 9.470159
```

```
(SE_senator_arkansas2 <- regression(senator[senator$state == "Arkansas", ],
                                    ATE_senator_arkansas))
```

```
## [1] 3.346872
```

And the third possibility is to estimate all the potential outcomes based on the estimated ATE, and simulate multiple random allocations. The values we obtain are closer to the first ones—using equation (3.6).

```
estimate.in.CI <- function(dataframe, ATE, num_iter){
  standard.error <- numeric(num_iter)
  for (i in 1:num_iter){
    observed_outcomes_control <- dataframe[dataframe[, 1] == 0, 2]
    observed_outcomes_treatment <- dataframe[dataframe[, 1] == 1, 2]
    potential_outcomes_control <- c(observed_outcomes_control,
                                     observed_outcomes_treatment - ATE)
    potential_outcomes_treatment <- c(observed_outcomes_control + ATE,
                                       observed_outcomes_treatment)

    treatment <- sample(dataframe[, 1])
    outcomes <- potential_outcomes_treatment * treatment +
                 potential_outcomes_control * (1 - treatment)
    standard.error[i] <- summary(lm(outcomes ~ treatment))$coefficients[2, 2]
  }
  SE <- mean(standard.error)
  return(SE = SE)
}
(SE_senator_texas3 <- estimate.in.CI(senator[senator$state == "Texas", ],
                                     ATE_senator_texas, 1e3))
```

```
## [1] 9.314177
```

```
(SE_senator_arkansas3 <- estimate.in.CI(senator[senator$state == "Arkansas", ],
                                         ATE_senator_arkansas, 1e3))
```

```
## [1] 3.292485
```

(c) Use equation (3.10) to estimate the overall ATE for both states combined.

```
(ATE_senator <- (dim(senator[senator$state == "Texas", ])[1] *
                 ATE_senator_texas +
                 dim(senator[senator$state == "Arkansas", ])[1] *
                 ATE_senator_arkansas) / dim(senator)[1])
```

```
## [1] -13.2168
```

(d) Explain why, in this study, simply pooling the data for the two states and comparing the average number of bills introduced by two-year senators to the average number of bills introduced by four-year senators leads to biased estimates of the overall ATE.

The overall ATE without considering blocked randomization is different from what we had previously calculated.

```
(ATE_senator_pooled <- est.ate(senator$bills, senator$t2yr))
```

```
## [1] -14.51515
```

This is a biased estimate because the probability of being assigned to the treatment group varies by block (state): in Texas this probability is $15/31 = 48.4\%$, while in Arkansas the probability of being assigned to the treatment group is $18/35 = 51.4\%$. Besides, the number of bills is lower on average in Arkansas, so the overall treatment effect calculated this way is larger than it actually is.

(e) Insert the estimated standard errors into equation (3.12) to estimate the standard error for the overall ATE.

```
(SE_senator <- sqrt(SE_senator_texas^2 *
  dim(senator[senator$state == "Texas", ])[1]^2 +
  SE_senator_arkansas^2 *
  dim(senator[senator$state == "Arkansas", ])[1]^2) /
dim(senator)[1])
```

```
## [1] 4.74478
```

(f) Use randomization inference to test the sharp null hypothesis that the treatment effect is zero for senators in both states.

```
distribSNH_senator_texas <-
  replicate(10e3, est.ate(senator$bills[senator$state == "Texas"],
    sample(senator$t2yr[senator$state == "Texas"])))
(pvalue_senator_texas <- mean(ATE_senator_texas >= distribSNH_senator_texas))
```

```
## [1] 0.0451
```

```
distribSNH_senator_arkansas <-
  replicate(10e3, est.ate(senator$bills[senator$state == "Arkansas"],
    sample(senator$t2yr[senator$state == "Arkansas"])))
(pvalue_senator_arkansas <- mean(ATE_senator_arkansas >=
  distribSNH_senator_arkansas))
```

```
## [1] 0.0029
```

```
distribSNH_senator <-
  replicate(10e3, est.ate(senator$bills,
    c(sample(senator$t2yr[senator$state == "Texas"]),
      sample(senator$t2yr[senator$state == "Arkansas"]))))
(pvalue_senator <- mean(ATE_senator >= distribSNH_senator))
```

```
## [1] 0.0069
```

```
df_distribSNH_senator <-
  rbind(data.frame(ATE = distribSNH_senator_texas,
    state = rep("Texas", 10e3),
    ATE_estimate = rep(ATE_senator_texas, 10e3)),
    data.frame(ATE = distribSNH_senator_arkansas,
      state = rep("Arkansas", 10e3),
      ATE_estimate = rep(ATE_senator_arkansas, 10e3)),
    data.frame(ATE = distribSNH_senator,
      state = rep("both", 10e3),
      ATE_estimate = rep(ATE_senator, 10e3)))

library(ggplot2)
density <- ggplot(df_distribSNH_senator, aes(ATE))
density + geom_density() + facet_grid(~ state) +
  geom_vline(aes(xintercept = ATE_estimate), color = "green") +
  theme(legend.position = "none")
```

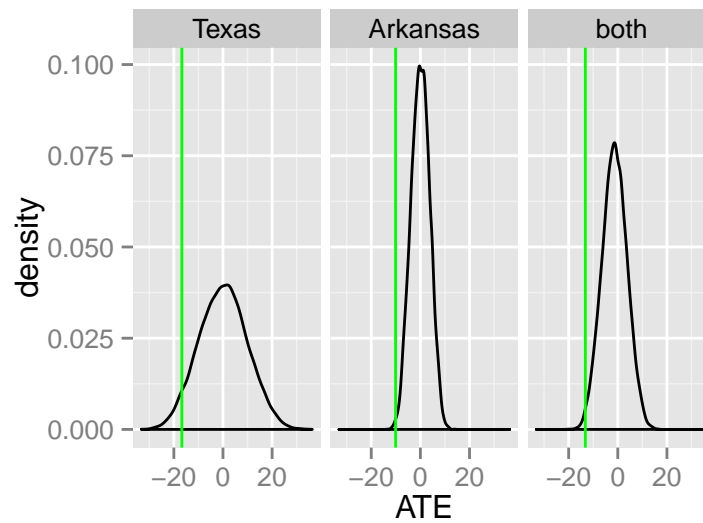


Figure 3: Senator: Sampling distribution of the ATE under the SNH in each state

The treatment effect is **statistically significant** in Texas, Arkansas, and both.

In addition to the problems in the book, plot histograms for both the treatment and control groups in each state (for 4 histograms in total).

```
library(ggplot2)
hist <- ggplot(senator, aes(bills))
hist + geom_histogram(binwidth = 5) + facet_grid(group ~ state) +
  theme(legend.position = "none")
```

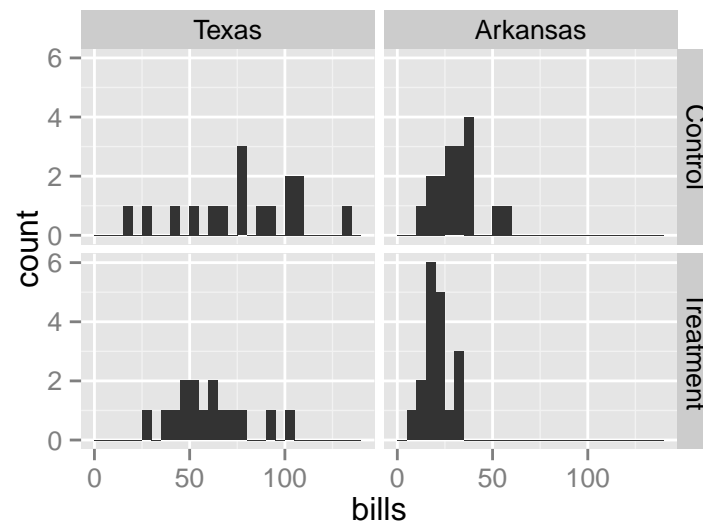


Figure 4: Senator: Histograms for both the treatment and control groups in each state

As we can see in the histograms below, the variance of the untreated outcomes in Texas is higher, so the

estimated standard error of the ATE in that state may be overestimated. In Arkansas the variance of both treated and untreated outcomes is lower, so the treatment effect is more visible (the center of the histogram of the treatment group is moved to the left).

4. FE, exercise 3.11.

Note: assume 3 clusters in treatment and 4 in control.

(a) Suppose that clusters are formed by grouping observations {1,2}, {3,4}, {5,6}, . . . , {13,14}. Use equation (3.22) to calculate the standard error assuming half of the clusters are randomly assigned to treatment.

Since all the potential outcomes are known in this example, I've calculated the standard error using both the population and sample variances and covariance.

```
public_works <- read.csv("GerberGreenBook_Chapter3_Table_3_3.csv")

SE_cluster <- function(df, cluster, clusters_in_treatment){
  k <- length(unique(cluster))
  N <- dim(df)[1]
  m <- clusters_in_treatment * 2
  df[, 5] <- cluster
  mean_Y0 <- unlist(lapply(1:k, function(x)
    mean(df[, 2][df[, 5] == x])))
  mean_Y1 <- unlist(lapply(1:k, function(x)
    mean(df[, 3][df[, 5] == x])))
  SE_ATE <- sqrt(1/(k-1)*(m/(N-m)*var(mean_Y0) +
    (N-m)/m*var(mean_Y1) +
    2*cov(mean_Y0, mean_Y1)))
  SE_ATE_pop <- sqrt(1/(k-1)*(m/(N-m)*var(mean_Y0)*(k-1)/k +
    (N-m)/m*var(mean_Y1)*(k-1)/k +
    2*cov(mean_Y0, mean_Y1)*(k-1)/k))
  return(list(SE_ATE = SE_ATE, SE_ATE_pop = SE_ATE_pop))
}

cluster <- unlist(lapply(1:7, function(x) rep(x, 2)))
unlist(SE_cluster(public_works, cluster, 3))
```

```
##      SE_ATE SE_ATE_pop
## 4.918953  4.554065
```

(b) Suppose that clusters are formed by grouping observations {1,14}, {2,13}, {3,12}, . . . , {7,8}. Use equation (3.22) to calculate the standard error assuming half of the clusters are randomly assigned to treatment.

```
cluster <- c(1:7, 7:1)
unlist(SE_cluster(public_works, cluster, 3))
```

```
##      SE_ATE SE_ATE_pop
## 1.264924  1.171092
```

(c) Why do the two methods of forming clusters lead to different standard errors? What are the implications for the design of cluster randomized experiments?

Because untreated outcomes and most of the treated outcomes grow with the village index. That caused the cluster-level means of the first group of clusters to have a high variance, which led to a high standard error. The second group of clusters have more similar average potential outcomes, so the variances of $\bar{Y}_j(0)$ and $\bar{Y}_j(1)$ were much lower.

The conclusion is that clusters should be formed trying to reduce the variability of the cluster-level means. For instance, subjects with similar potential outcomes should be placed in different clusters, so the cluster-level means are as similar as possible to each other.

5. Online advertising.

a. By how much does the ad campaign need to increase the probability of purchase in order to be “worth it” and a positive ROI (supposing there are no long-run effects and all the effects are measured within that week)?

In the absence of any advertising, Apple sells an iPhone to 5,000 users ($0.5\% \times 1,000,000$ users), making a profit of \$500,000 ($\$100 \times 5,000$ users). The cost of the ad campaign would be \$100,000 ($\$0.10 \times 1,000,000$ users). So in order to have a positive ROI, Apple should sell 1,000 iPhones ($\$100,000 / \100) more—i.e., a **0.1 percentage point increase** ($1,000 / 1,000,000$)—making a total of 6,000 iPhones ($5,000 + 1,000$), and hence increasing the probability of selling an iPhone from 0.5% to 0.6%, i.e., a **20% increase** ($\frac{0.6-0.5}{0.5}$).

b. Assume the measured effect is 0.2 percentage points. If users are split 50:50 between the treatment group (exposed to iPhone ads) and control group (exposed to unrelated advertising or nothing; something you can assume has no effect), what will be the confidence interval of your estimate on whether people purchase the phone?

The standard error for a two-sample proportion test is

$$\sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

where $p = \frac{x_1 + x_2}{n_1 + n_2}$, where x and n refer to the number of “successes” (here, purchases) over the number of “trials” (here, site visits). The length of each tail of a 95% confidence interval is calculated by multiplying the standard error by 1.96.

```
n1 <- 500e3
n2 <- 500e3
x1 <- 0.005*n1
x2 <- 0.007*n2

p <- (x1 + x2) / (n1 + n2)
SE_apple <- sqrt(p*(1-p)*(1/n1+1/n2))
(CI_apple <- c(2e-3 - 1.96*SE_apple, 2e-3 + 1.96*SE_apple))

## [1] 0.00169727 0.00230273
```

Therefore, the confidence interval is [0.17%, 0.23%].

Another way to calculate the standard error would be:

```
control_apple <- c(rep(1, 5e3), rep(0, 995e3))
treatment_apple <- c(rep(1, 7e3), rep(0, 993e3))
(SE_apple <- sqrt(var(control_apple)/500e3 + var(treatment_apple)/500e3))
```

```
## [1] 0.000154441
```

c. Is this confidence interval precise enough that you would recommend running this experiment? Why or why not?

The confidence interval is precise enough, but assuming the effect is so high (0.2%), a much simpler (and cheaper) experiment could be conducted, allocating much less subjects to the treatment group. For example, if only 20,000 out of the 1,000,000 users are shown the ad, the lower bound of the confidence interval would still be above 0.1%.

```
n2 <- 20e3
n1 <- 1e6 - n2
x1 <- 0.005*n1
x2 <- 0.007*n2

p <- (x1 + x2) / (n1 + n2)
SE_apple <- sqrt(p*(1-p)*(1/n1+1/n2))
(CI_apple <- c(2e-3 - 1.96*SE_apple, 2e-3 + 1.96*SE_apple))
```

```
## [1] 0.001008606 0.002991394
```

d. Your boss at the newspaper, worried about potential loss of revenue, says he is not willing to hold back a control group any larger than 1% of users. What would be the width of the confidence interval for this experiment if only 1% of users were placed in the control group?

```
n1 <- 0.01*1e6
n2 <- 1e6 - n1
x1 <- 0.005*n1
x2 <- 0.007*n2

p <- (x1 + x2) / (n1 + n2)
SE_apple <- sqrt(p*(1-p)*(1/n1+1/n2))
(CI_apple <- c(2e-3 - 1.96*SE_apple, 2e-3 + 1.96*SE_apple))
```

```
## [1] 0.000359995 0.003640005
```

The width of the interval (from lower to upper bound) would now be 0.33 percentage points, so we would not be completely certain about the positive impact of the ads (in terms of positive ROI).

6. Auction experiment.

- a. Compute a 95% confidence interval for the difference between the treatment mean and the control mean, using analytic formulas for a two-sample t-test from your earlier statistics course.

We just need to know the size, mean and standard deviation of each group, and hence the degrees of freedom of our sample.

```
auction <- read.csv("PS 2 data - list_luckingreiley_auction_data.csv")
```

```
(n1 <- length(auction$bid[auction$uniform_price_auction == 0]))
```

```
## [1] 34
```

```
(n2 <- length(auction$bid[auction$uniform_price_auction == 1]))
```

```
## [1] 34
```

```
(s1 <- sd(auction$bid[auction$uniform_price_auction == 0]))
```

```
## [1] 19.98101
```

```
(s2 <- sd(auction$bid[auction$uniform_price_auction == 1]))
```

```
## [1] 15.40172
```

```
(DF = (s1^2/n1 + s2^2/n2)^2 /  
      (((s1^2 / n1)^2 / (n1 - 1)) + ((s2^2 / n2)^2 / (n2 - 1))))
```

```
## [1] 61.98287
```

```
mean1 <- mean(auction$bid[auction$uniform_price_auction == 0])  
mean2 <- mean(auction$bid[auction$uniform_price_auction == 1])  
(difference <- mean2 - mean1)
```

```
## [1] -12.20588
```

```
StdError <- sqrt(s1^2 / n1 + s2^2 / n2)  
alpha = 1 - 0.95  
p = 1 - alpha/2  
(critical_value <- qt(p, DF))
```

```
## [1] 1.998983
```

```
(margin_error <- critical_value * StdError)
```

```
## [1] 8.648742
```

```
(CI <- c(difference - margin_error, difference + margin_error))
```

```
## [1] -20.854624 -3.557141
```

Therefore, the confidence interval is [-20.85%, -3.56%].

b. In plain language, what does this confidence interval mean?

If we conducted this experiment multiple times, 95% of them the difference-in-means would be bracketed within this confidence interval (and since it does not cross zero, we would detect a treatment effect—that the treatment auction format produces lower bids—at least in 95% of the experiments).

c. Regression on a binary treatment variable turns out to give one the same answer as the standard analytic formula you just used. Demonstrate this by regressing the bid on a binary variable equal to 0 for the control auction and 1 for the treatment auction. Calculate the 95% confidence interval you get from the regression.

```
(regression <- summary(lm(auction$bid ~ auction$uniform_price_auction)))
```

```
##
## Call:
## lm(formula = auction$bid ~ auction$uniform_price_auction)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.824 -11.618  -3.221   8.382  58.382
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      28.824      3.059   9.421 7.81e-14 ***
## auction$uniform_price_auction -12.206      4.327  -2.821 0.00631 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.84 on 66 degrees of freedom
## Multiple R-squared:  0.1076, Adjusted R-squared:  0.09409
## F-statistic: 7.959 on 1 and 66 DF, p-value: 0.006315
```

```
difference_reg <- regression$coefficients[2,1]
StdError_reg <- regression$coefficients[2,2]
margin_error_reg <- critical_value * StdError_reg
(CI_reg <- c(difference_reg - margin_error_reg, difference + margin_error_reg))
```

```
## [1] -20.854624 -3.557141
```

Again, the confidence interval is [-20.85%, -3.56%].

d. On to p-values. What p-value does the regression report?

As previously shown, it is 0.0063148. So the difference-in-means is statistically significant. (The p -value is based on a two-tailed test. A one-tailed test would be more appropriate since the treatment auction format was theoretically predicted to produce *lower* bids than the control auction format.)

e. Now compute the same p-value using randomization inference.

```
ATE_auction <- est.ate(auction$bid, auction$uniform_price_auction)
distribSNH_auction <- replicate(10e3,
                                est.ate(auction$bid,
                                          sample(auction$uniform_price_auction)))
(pvalue_auction <- mean(abs(distribSNH_auction) >= abs(ATE_auction)))

## [1] 0.0057
```

The p-value calculated this way (0.0057) is just slightly different.

f. Compute the same p-value again using analytic formulas for a two-sample t-test from your earlier statistics course-

```
(t <- (difference) / sqrt(s1^2 / n1 + s2^2 / n2))

## [1] -2.821144

(sig <- 2*(1-pt(abs(t), DF)))

## [1] 0.006420778
```

In this case the p -value is 0.0064208.

Everything could also be done using the *t.test* function in R:

```
t.test(auction$bid ~ auction$uniform_price_auction, alternative = "two.sided")

##
## Welch Two Sample t-test
##
## data: auction$bid by auction$uniform_price_auction
## t = 2.8211, df = 61.983, p-value = 0.006421
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 3.557141 20.854624
## sample estimates:
## mean in group 0 mean in group 1
## 28.82353 16.61765
```

g. Compare the two p-values in parts (e) and (f). Are they much different? Why or why not? How might your answer to this question change if the sample size were different?

The last two p -values are similar, but not equal. The reason is that the sampling distribution of the ATE under the sharp null hypothesis is only close to normal. So the p -value given by the t -test is not so accurate.

```
plot(density(distribSNH_auction))
```

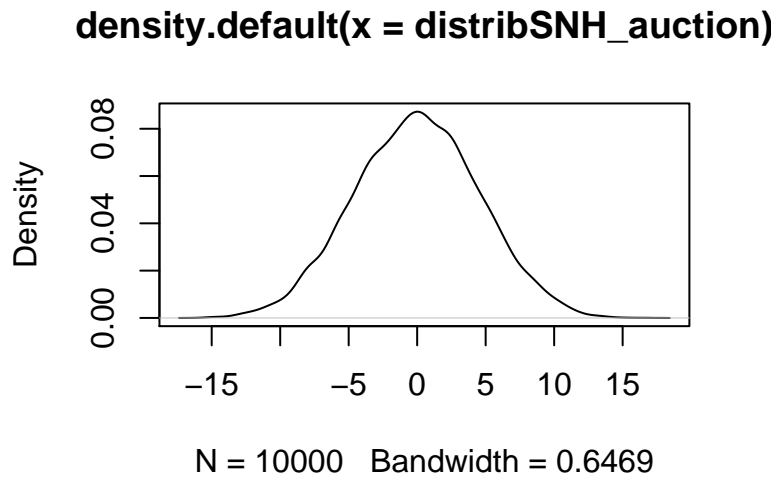


Figure 5: Auction: Sampling distribution of the ATE under the SNH

```
shapiro.test(distribSNH_auction[1:5000])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  distribSNH_auction[1:5000]
## W = 0.9994, p-value = 0.1424
```

If the sampling distribution were even far from normality (e.g., because one group contains an outlier that causes the outcome histogram to be skewed, and the sampling distribution to be bimodal), both p -values would have been much more different (and, as mentioned, the most accurate would have been the one calculated with randomization inference). In any case, increasing the sample size can lead the sampling distribution of the ATE to be closer to normality (e.g., by reducing the effect of outliers), and the two p -values to be more similar.