Problem Set #1

BST 258: Causal Inference – Theory and Practice

Daniel Xu

2/11/24

Collaborator: Jenna Landy

Question 2

a) $P(A = 1) = m/n$

b) For two units $i \neq j$,

$$
\begin{aligned}
P(A_i = 1, A_j = 1) &= P(A_j = 1 | A_i = 1) P(A_i = 1) \\
&= \frac{(m-1)}{(n-1)} \frac{m}{n} \\
P(A_i = 1, A_j = 0) &= P(A_j = 0 | A_i = 1) P(A_i = 1) \\
&= \frac{(n-1-m)}{(n-1)} \frac{m}{n} \\
P(A_i = 0, A_j = 1) &= P(A_j = 1 | A_i = 0) P(A_i = 0) \\
&= \frac{m}{(n-1)} \frac{(n-m)}{n} \\
P(A_i = 0, A_j = 0) &= P(A_j = 0 | A_i = 0) P(A_i = 0) \\
&= \frac{(n-1-m)}{(n-1)} \frac{(n-m)}{n}
\end{aligned}
$$

c) For two units $i \neq j$,

$$
\begin{aligned}
Var(A_i) &= E(A_i^2) - E(A_i)^2 \\
&= P(A_i = 1)(1 - P(A_i = 1)) \\
&= \frac{m}{n}\left(1 - \frac{m}{n}\right) \\
Cov(A_i, A_j) &= E(A_i A_j) - E(A_i)E(A_j) \\
&= P(A_i = 1, A_j = 1) - P(A_i = 1)P(A_j = 1) \\
&= \frac{(m-1)m}{(n-1)n} - \left(\frac{m}{n}\right)^2
\end{aligned}
$$

d) Assuming $Y_i(1)$ and $Y_i(0)$ fixed,

$$
\begin{aligned}
E(\theta^{ATT}) &= E\left(\frac{1}{m}\sum_{i=1}^{n} A_i\left(Y_i(1) - Y_i(0)\right)\right) \\
&= \frac{1}{m}\sum_{i=1}^{n}\left(Y_i(1) - Y_i(0)\right)E(A_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i(1) - Y_i(0)\right) \\
&= \theta^{SATE}
\end{aligned}
$$

Question 3

Given that θ is fixed,

$$Var(Y_i(1)) = Var(Y_i(0) + \theta) = Var(Y_i(0))$$

and by linearity of covariances,

$$\begin{aligned}
\rho(Y_i(1), Y_i(0)) &= \frac{Cov(Y_i(1), Y_i(0))}{\sqrt{Var(Y_i(1))Var(Y_i(0))}} \\
&= \frac{Cov(Y_i(0) + \theta, Y_i(0))}{Var(Y_i(0))} \\
&= \frac{Var(Y_i(0)}{Var(Y_i(0))} \\
&= 1
\end{aligned}$$

Question 4

Assuming that you guess completely at random, let X be the number of correct guesses. Clearly, $X \sim Hypergeometric(N = 8, K = 4, n = 4)$ where N is the total number of cups, K is the total number of cups with tea poured first, and n is the number of cups you guess have tea poured first. Hence, we know the distribution of X and can easily calculate

$$P(X = 0) = \frac{1}{70}$$
$$P(X = 1) = \frac{16}{70}$$
$$P(X = 2) = \frac{36}{70}$$
$$P(X = 3) = \frac{16}{70}$$
$$P(X = 4) = \frac{1}{70}$$

Question 5

a) There is confounding of the treatment and success relationship by stone size. Large stones are harder to deal with (lead to worse outcomes) and physicians were more likely to assign patients with large stones to treatment A.

b) For small stones:

.	Treatment A	Treatment B
Male	94.7% (71/75)	95.0% (19/20)
Female	83.3% (10/12)	86.0% (215/250)

For large stones:

.	Treatment A	Treatment B
Male	74.5% (181/243)	75.0% (15/20)
Female	55.0% (11/20)	66.7% (40/60)

c) This phenomenon where a trend appears when stratified but then reverses when aggregated is known as Simpson's paradox. It can be problematic if the person analyzing the data attempts to blindly draw causal conclusions. For instance, in this case, without stratifying by stone size, the analyst may naively conclude that treatment B is more effective. However, especially because the data was not collected from a randomized experiment, we cannot in general attribute marginal associations to causal effects. In fact, here it seems that once we stratify by the confounding variable (stone size), we see that if anything, Treatment B seems more effective than Treatment A.

Question 6

We would expect that we would have higher power in rejecting the strong null vs. the weak null, but (perhaps) interestingly, this is not what we see. Let's think about this closer. Imagine the set of distributions of satisfying the strong null vs. the set of distributions satisfying the weak null. Clearly, the former lies entirely within the latter. Hence, if the true distribution satisfied the weak null but not the strong null, then we would surely expect higher power in rejecting the strong null vs. the weak null. This would be the case if for half the individuals there was a positive treatment effect and then for the other half there was a negative treatment effect such that the ATE was equal to 0. In this case, any reasonable test would be consistent (have asymptotic power = 1) for rejecting the strong null, but not necessarily for rejecting the weak null (as the weak null holds). In this scenario, we would expect to see the power against the strong null to be greater than the weak null (which aligns with our a priori expectations). However, this is the not the case we are dealing with. For this problem, the true distribution lies outside the distributions satisfying both the weak null and the strong null (as the ATE is non-zero), and hence we would expect any reasonable tests against the weak null and the strong null to be consistent, which is what we see in the simulations. Hence, any differences in power for a given sample size would be due to differences in the variability associated with the test statistic / the specific testing procedure. Given the simulation results, it seems that in this case, the variance associated with the difference in means estimator under the permutation distribution is greater than Neyman's variance estimator, which was unexpected.

```r
n1 <- c(10, 25, 50, 100, 250)
n0 <- n1
n <- n1 + n0
nsim <- 1000
B <- 10000

strong_null_reject <- matrix(data = NA, ncol = length(n1), nrow = nsim)
weak_null_reject <- matrix(data = NA, ncol = length(n1), nrow = nsim)
set.seed(999)
for (k in 1:length(n1)) {
  Y1 <- rnorm(n[k], mean = 1/10, sd = sqrt(1/16))
  Y0 <- rnorm(n[k], mean = 0, sd = sqrt(1/16))
  for (i in 1:nsim) {
    A <- sample(c(rep(1, times = n1[k]), rep(0, times = n0[k])),
                size = n[k], replace = FALSE)
    Y <- A * Y1 + (1 - A) * Y0
    dif_mean <- (1/n1[k]) * sum(A * Y) - (1/n0[k]) * sum((1 - A) * Y)
    var_dif_mean <- (1/n1[k]) * (1/(n1[k] - 1)) *
      sum(A*(Y - (1/n1[k]) * sum(A * Y))^2) +
      (1/n0[k]) * (1/(n0[k] - 1)) *
      sum((1-A)*(Y - (1/n0[k]) * sum((1 - A) * Y))^2)
    # generate null distribution
    dif_mean_null <- rep(NA, times = B)
    for (j in 1:B) {
      A <- sample(c(rep(1, times = n1[k]), rep(0, times = n0[k])),
                  size = n[k], replace = FALSE)
      dif_mean_null[j] <- (1/n1[k]) * sum(A * Y) -
        (1/n0[k]) * sum((1 - A) * Y)
    }
    strong_null_reject[i,k] <-
      mean(abs(dif_mean_null) >= abs(dif_mean)) <= .05
    weak_null_reject[i,k] <-
      abs(dif_mean) / sqrt(var_dif_mean) >= 1.96
  }
}

results <-
  data.frame(n,
             "strong" = apply(strong_null_reject, MARGIN = 2, FUN = mean),
             "weak" = apply(weak_null_reject, MARGIN = 2, FUN = mean))
```

```r
library(knitr)
kable(results, booktabs = T, caption = "Power for rejecting weak vs. strong null hypotheses f
```

Table 3: Power for rejecting weak vs. strong null hypotheses for 1,000 simulated datasets.

n	Power against strong null	Power against weak null
20	0.073	0.096
50	0.378	0.417
100	0.880	0.888
200	0.833	0.837
500	1.000	1.000

Question 7

a) We can set derivatives with respect to α and β to 0 and solve. Clearly

$$U(\alpha) = -\frac{1}{n}\sum_{i=1}^{n}(Y_i - \alpha - \beta A_i) = 0$$

which implies that

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\beta}A_i).$$

Likewise,

$$U(\beta) = -\frac{1}{n}\sum_{i=1}^{n}A_i(Y_i - \alpha - \beta A_i) = 0$$

which implies (by substitution) that

$$\frac{\beta}{n}\sum_{i=1}^{n}A_i = \frac{1}{n}\sum_{i=1}^{n}(A_iY_i - \alpha A_i)$$

$$\frac{\beta m}{n} = \frac{1}{n}\sum_{i=1}^{n}A_iY_i - \frac{\alpha m}{n}$$

$$\beta = \frac{1}{m}\sum_{i=1}^{n}A_iY_i - \alpha$$

$$\beta\left(\frac{n-m}{n}\right) = \bar{Y}_1 - \bar{(Y)}$$

$$\hat{\beta} = \frac{n}{n-m}\left(\bar{Y}_1 - \bar{Y}\right)$$

$$\hat{\beta} = \frac{n}{n-m}\left(\bar{Y}_1 - \frac{m}{n}\bar{Y}_1 - \frac{n-m}{n}\bar{Y}_0\right)$$

$$\hat{\beta} = \bar{Y}_1 - \bar{Y}_0$$

where

$$\bar{Y}_1 = \frac{1}{m}\sum_{i=1}^{n}A_iY_i, \qquad \bar{Y}_0 = \frac{1}{n-m}\sum_{i=1}^{n}(1 - A_i)Y_i$$

and we can resubstitute this expression for $\hat{\alpha}$ to obtain

$$\hat{\alpha} = \frac{m}{n}\bar{Y}_1 + \frac{n-m}{n}\bar{Y}_0 - \frac{m}{n}\left(\bar{Y}_1 - \bar{Y}_0\right)$$

$$= \bar{Y}_0$$

b) Clearly, we have that $\hat{\beta}$ is equivalent to our standard difference of means estimator, which as we have discussed in class is a "valid" estimator of the ATE.