

# PBA\_AS5

Yung Hsi Liu\_114078509

## Table of contents

<b>1 Question 1</b>	<b>2</b>
1.1 Rubric: Qmd file compiling . . . . .	2
1.2 Rubric: New variables . . . . .	2
1.3 Rubric: Correct <code>base_diff</code> . . . . .	2
1.4 Rubric: <code>base_p</code> . . . . .	2
1.5 Rubric: Reporting the difference/p-value and the rejection or not decision . . . . .	3
1.6 Rubric: Correct error type . . . . .	3
<b>2 Question 2</b>	<b>3</b>
2.1 Rubric: <code>base_diff_attend</code> . . . . .	3
2.2 Rubric: <code>base_p_attend</code> . . . . .	4
2.3 Rubric: Reporting the difference/p-value and the rejection or not decision . . . . .	4
2.4 Rubric: Correct error type . . . . .	4
2.5 Rubric: Correctly identifying what parameter controls this error . . . . .	4
<b>3 Question 3</b>	<b>4</b>
3.1 Rubric: Identifying which variable should be used . . . . .	4
3.2 Rubric: Justification . . . . .	4
<b>4 Question 4</b>	<b>5</b>
4.1 Rubric: <code>ate</code> . . . . .	5
4.2 Rubric: <code>ate_null_dist</code> . . . . .	5
4.3 Rubric: <code>ate_p</code> . . . . .	6
4.4 Rubric: Plot of the null distribution . . . . .	6
4.5 Rubric: Explanation of the null distribution and reporting if null is rejected . . . . .	7
<b>5 Question 5</b>	<b>7</b>
5.1 Rubric: <code>ate_lt</code> . . . . .	7
5.2 Rubric: <code>ate_lt_null_dist</code> . . . . .	7
5.3 Rubric: <code>ate_lt_p</code> . . . . .	8
5.4 Rubric: Plot of the null distribution . . . . .	8
5.5 Rubric: Reporting if null is rejected and describing the persistence over time. . . . .	9

# 1 Question 1

## 1.1 Rubric: Qmd file compiling

```
cbt <- read_csv("https://bit.ly/3KfrFtx")
glimpse(cbt)
```

```
Rows: 999
Columns: 10
$ partid      <dbl> 20001, 20002, 20003, 20005, 20006, 20007, 20008, 200~
$ cashassigned <dbl> 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1~
$ tpassigned   <dbl> 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1~
$ attend_80    <dbl> NA, 1, NA, 1, NA, NA, 1, 0, 0, NA, NA, 0, NA, 1, 0, ~
$ steals_baseline <dbl> NA, ~
$ homeless_baseline <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1~
$ year_born     <dbl> 1980, 1991, 1987, 1986, 1985, 1985, 1986, 1990, 1984~
$ fam_asb_st    <dbl> -0.65081871, -0.70169687, -0.66238648, -0.77098411, ~
$ carryweapon   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0~
$ fam_asb_lt    <dbl> -0.75319594, -0.71274430, -0.47705624, -0.79630595, ~
```

## 1.2 Rubric: New variables

```
set.seed(02138)
cbt <- cbt |>
  mutate(
    cbt_assigned = if_else(tpassigned == 1, "CBT", "No CBT"),
    unhoused_baseline = if_else(homeless_baseline == 1, "Unhoused", "Housed")
  )
```

## 1.3 Rubric: Correct base\_diff

```
base_diff <- cbt |>
  specify(unhoused_baseline ~ cbt_assigned, success="Unhoused") |>
  calculate(stat="diff in props", order=c("CBT", "No CBT"));base_diff
```

```
Response: unhoused_baseline (factor)
Explanatory: cbt_assigned (factor)
# A tibble: 1 x 1
  stat
  <dbl>
1 -0.00602
```

## 1.4 Rubric: base\_p

```

base_p <- cbt |>
  specify(unhoused_baseline ~ cbt_assigned, success="Unhoused") |>
  hypothesize(null="independence") |>
  generate(reps=1000, type="permute") |>
  calculate(stat="diff_in_props", order=c("CBT", "No CBT")) |>
  get_p_value(obs_stat=base_diff, direction="both");base_p

```

```

# A tibble: 1 x 1
  p_value
  <dbl>
1 0.856

```

## 1.5 Rubric: Reporting the difference/p-value and the rejection or not decision

The observed difference is  $-0.006$ . The two-sided p-value is more than 0.85, so we fail to reject the null hypothesis at the 0.05 level, indicating no significant difference between the two groups.

## 1.6 Rubric: Correct error type

Since the null hypothesis is not rejected, the main concern is a Type II error, where a true difference exists but the test fails to detect it.

# 2 Question 2

## 2.1 Rubric: base\_diff\_attend

```

set.seed(02138)

cbt <- cbt |>
  mutate(
    cbt_attended = if_else(attend_80 == 1, "Attended CBT", "Not Attended"),
  )

base_diff_attend <- cbt |>
  filter(!is.na(attend_80)) |>
  specify(unhoused_baseline ~ cbt_attended, success="Unhoused") |>
  calculate(stat="diff_in_props", order=c("Attended CBT", "Not Attended"));base_diff_attend

```

```

Response: unhoused_baseline (factor)
Explanatory: cbt_attended (factor)
# A tibble: 1 x 1
  stat
  <dbl>
1 -0.147

```

## 2.2 Rubric: base\_p\_attend

```
base_p_attend <- cbt |>
  filter(!is.na(attend_80)) |>
  specify(unhoused_baseline ~ cbt_attended, success="Unhoused") |>
  hypothesize(null="independence") |>
  generate(reps=1000, type="permute") |>
  calculate(stat="diff in props", order=c("Attended CBT", "Not Attended")) |>
  get_p_value(obs_stat=base_diff_attend, direction="both");base_p_attend
```

Warning: Please be cautious in reporting a p-value of 0. This result is an approximation based on the number of `reps` chosen in the `generate()` step.  
i See `get\_p\_value()` (`?infer::get\_p\_value()`) for more information.

```
# A tibble: 1 x 1
  p_value
  <dbl>
1      0
```

## 2.3 Rubric: Reporting the difference/p-value and the rejection or not decision

The observed difference in baseline proportions is  $-0.147$ . The two-sided permutation test yields a p-value of 0, so we reject the null hypothesis at the 0.05 level, indicating a significant difference between those who attended CBT and those who did not.

## 2.4 Rubric: Correct error type

Since the null hypothesis is rejected, the main concern is a Type I error, where we incorrectly conclude that a difference exists when it does not.

## 2.5 Rubric: Correctly identifying what parameter controls this error

The Type I error is controlled by the significance level alpha.

# 3 Question 3

## 3.1 Rubric: Identifying which variable should be used

The CBT assignment should be used as the treatment variable.

## 3.2 Rubric: Justification

This is because CBT assignment is randomized and balanced at baseline, while actual attendance is not random and would introduce bias.

## 4 Question 4

### 4.1 Rubric: ate

```
set.seed(02138)

ate <- cbt |>
  specify(fam_asb_st ~ cbt_assigned) |>
  calculate(stat="diff in means", order=c("CBT", "No CBT"));ate
```

Warning: Removed 77 rows containing missing values.

```
Response: fam_asb_st (numeric)
Explanatory: cbt_assigned (factor)
# A tibble: 1 x 1
  stat
  <dbl>
1 -0.225
```

### 4.2 Rubric: ate\_null\_dist

```
ate_null_dist <- cbt |>
  specify(fam_asb_st ~ cbt_assigned) |>
  hypothesize(null="independence") |>
  generate(reps=1000, type="permute") |>
  calculate(stat="diff in means", order=c("CBT", "No CBT"));ate_null_dist
```

Warning: Removed 77 rows containing missing values.

```
Response: fam_asb_st (numeric)
Explanatory: cbt_assigned (factor)
Null Hypoth...
# A tibble: 1,000 x 2
  replicate    stat
  <int>    <dbl>
1       1 -0.0253
2       2  0.0662
3       3 -0.133 
4       4 -0.0315
5       5  0.0359
6       6 -0.0325
7       7  0.0113
8       8 -0.0830
9       9 -0.0248
10      10  0.0222
# i 990 more rows
```

### 4.3 Rubric: ate\_p

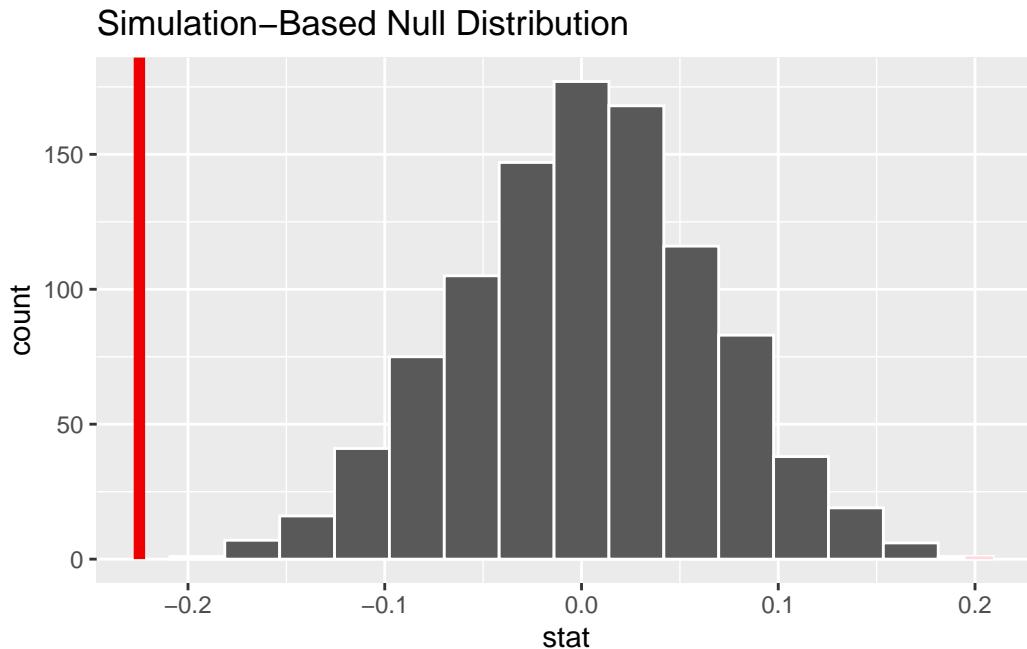
```
ate_p <- ate_null_dist |>
  get_p_value(obs_stat=ate, direction="both"); ate_p
```

Warning: Please be cautious in reporting a p-value of 0. This result is an approximation based on the number of `reps` chosen in the `generate()` step.  
i See `get\_p\_value()` (`.?infer::get\_p\_value()`) for more information.

```
# A tibble: 1 x 1
  p_value
  <dbl>
1      0
```

### 4.4 Rubric: Plot of the null distribution

```
ate_null_dist |>
  visualize() +
  shade_p_value(obs_stat=ate, direction="both"); ate_p
```



```
# A tibble: 1 x 1
  p_value
  <dbl>
1      0
```

## 4.5 Rubric: Explanation of the null distribution and reporting if null is rejected

The null distribution shows the differences in means we would expect if CBT had no effect. Since the observed ATE (-0.22) lies in the extreme tail and the p-value is 0, we reject the null hypothesis at the 0.05 level.

## 5 Question 5

### 5.1 Rubric: ate\_lt

```
set.seed(02138)

ate_lt <- cbt |>
  specify(fam_asb_lt ~ cbt_assigned) |>
  calculate(stat="diff in means", order=c("CBT", "No CBT"));ate_lt
```

Warning: Removed 52 rows containing missing values.

```
Response: fam_asb_lt (numeric)
Explanatory: cbt_assigned (factor)
# A tibble: 1 x 1
  stat
  <dbl>
1 -0.185
```

### 5.2 Rubric: ate\_lt\_null\_dist

```
ate_lt_null_dist <- cbt |>
  specify(fam_asb_lt ~ cbt_assigned) |>
  hypothesize(null="independence") |>
  generate(reps=1000, type="permute") |>
  calculate(stat="diff in means", order=c("CBT", "No CBT"));ate_lt_null_dist
```

Warning: Removed 52 rows containing missing values.

```
Response: fam_asb_lt (numeric)
Explanatory: cbt_assigned (factor)
Null Hypoth...
# A tibble: 1,000 x 2
  replicate   stat
  <int>     <dbl>
1       1  0.0374
2       2  0.0306
3       3  0.0120
4       4 -0.0116
```

```

5      5  0.0140
6      6 -0.0246
7      7 -0.104
8      8  0.0229
9      9 -0.0498
10     10  0.00512
# i 990 more rows

```

### 5.3 Rubric: ate\_lt\_p

```

ate_lt_p <- ate_lt_null_dist |>
  get_p_value(obs_stat=ate_lt, direction="both");ate_lt_p

```

Warning: Please be cautious in reporting a p-value of 0. This result is an approximation based on the number of `reps` chosen in the `generate()` step.  
 i See `get\_p\_value()` (`?infer::get\_p\_value()`) for more information.

```

# A tibble: 1 x 1
  p_value
  <dbl>
1      0

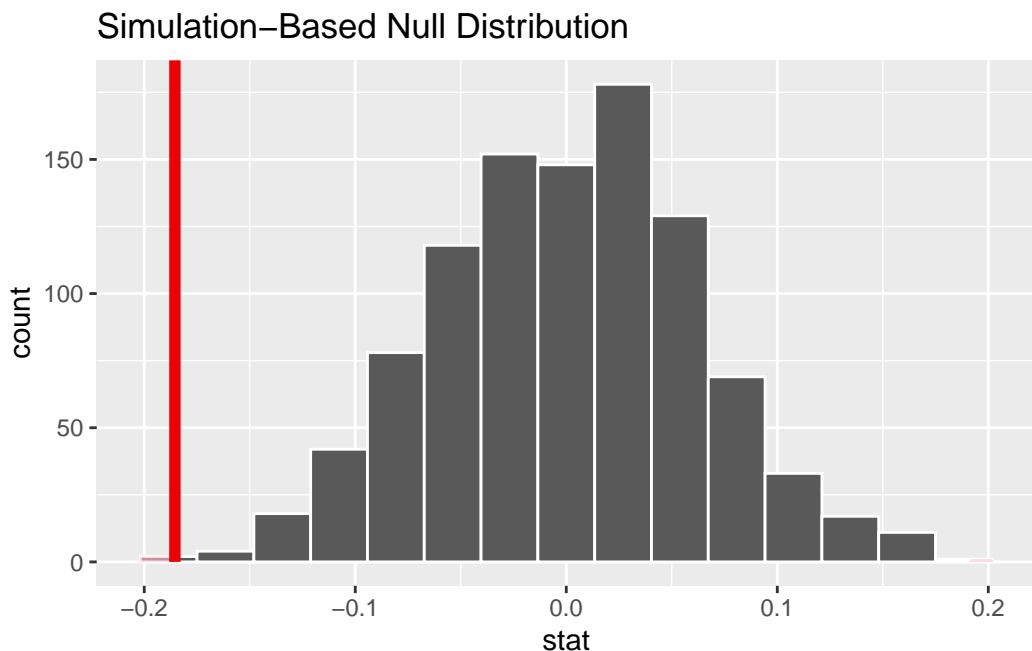
```

### 5.4 Rubric: Plot of the null distribution

```

ate_lt_null_dist |>
  visualize()+
  shade_p_value(obs_stat=ate_lt, direction="both");ate_lt_p

```



```
# A tibble: 1 x 1
  p_value
  <dbl>
1      0
```

### **5.5 Rubric: Reporting if null is rejected and describing the persistence over time.**

The long-term ATE shows that the CBT group still has lower anti-social behavior, with a p-value of 0, so we reject the null hypothesis at the 0.05 level. Compared to Question 4, the effect persists over time but is smaller in the long term, suggesting that the impact of CBT diminishes but does not disappear.