Hypothesis_Testing

Tanishq

2025-05-02

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.2     ✔ tibble    3.2.1
## ✔ lubridate 1.9.4     ✔ tidyr     1.3.1
## ✔ purrr     1.0.4
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
library(infer)
sat <- read_csv("SAT.csv")
```

```
## Rows: 460 Columns: 6
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (2): DBN, School Name
## dbl (4): Number of Test Takers, Critical Reading Mean, Mathematics Mean, Wri...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(sat)
```

```
## # A tibble: 6 × 6
##   DBN    `School Name`            Number of Test Taker…¹ Critical Reading Mea…²
##   <chr>  <chr>                                     <dbl>                  <dbl>
## 1 01M292 Henry Street School for …                    31                    391
## 2 01M448 University Neighborhood …                    60                    394
## 3 01M450 East Side Community High…                    69                    418
## 4 01M458 SATELLITE ACADEMY FORSYT…                    26                    385
## 5 01M509 CMSP HIGH SCHOOL                             NA                     NA
## 6 01M515 Lower East Side Preparat…                   154                    314
## # ℹ abbreviated names: ¹`Number of Test Takers`, ²`Critical Reading Mean`
## # ℹ 2 more variables: `Mathematics Mean` <dbl>, `Writing Mean` <dbl>
```

Data Cleaning

```
sat <- sat |> rename(reading = `Critical Reading Mean`,
                     math = `Mathematics Mean`,
                     writing = `Writing Mean`) |>
  filter(!is.na(reading), !is.na(math), !is.na(writing))

sat_aov <- sat |> select(reading,
                         math,
                         writing) |>
  pivot_longer(cols = c(reading,
                        math,
                        writing),
               names_to = "subject",
               values_to = "score")

  head(sat_aov)
```

```
## # A tibble: 6 × 2
##   subject score
##   <chr>   <dbl>
## 1 reading   391
## 2 math      425
## 3 writing   385
## 4 reading   394
## 5 math      419
## 6 writing   387
```

```
sat_math_reading <- sat |>
  select(math, reading) |>
  mutate(diff = math - reading)

sat_math_writing <- sat |>
  select(math, writing) |>
  mutate(diff = math - writing)

sat_writing_reading <- sat |>
  select(writing, reading) |>
  mutate(diff = writing - reading)
```

Anova Test

Significance level $\alpha = 0.05$ for all the tests.

H_0 : The mean SAT score is same across all the subject.

$\mu_{math} = \mu_{reading} = \mu_{writing}$

H_a : At least one of the mean SAT score is different from others.

At least one μ differs among math, reading and writing

```
anova_result <- aov(formula = score~subject, data = sat_aov)
summary(anova_result)
```

```
##               Df  Sum Sq Mean Sq F value  Pr(>F)   
## subject        2   45153   22576   6.278 0.00194 **
## Residuals   1155 4153518    3596                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova Conclusion

since p-value $= 0.00194 < 0.05$ We Reject Null Hypothesis.

At least one subject's mean SAT score (Math, Reading, Writing) is significantly different from others.

Pairwise T-Test

why? to find out which subject's mean SAT score differs from other.

```
pairwise.t.test(sat_aov$score,
                sat_aov$subject,
                p.adjust.method = "bonferroni")
```

```
## 
##  Pairwise comparisons using t tests with pooled SD 
## 
## data:  sat_aov$score and sat_aov$subject 
## 
##         math   reading
## reading 0.1332 -      
## writing 0.0013 0.3866 
## 
## P value adjustment method: bonferroni
```

Pairwise T-Test Conclusion

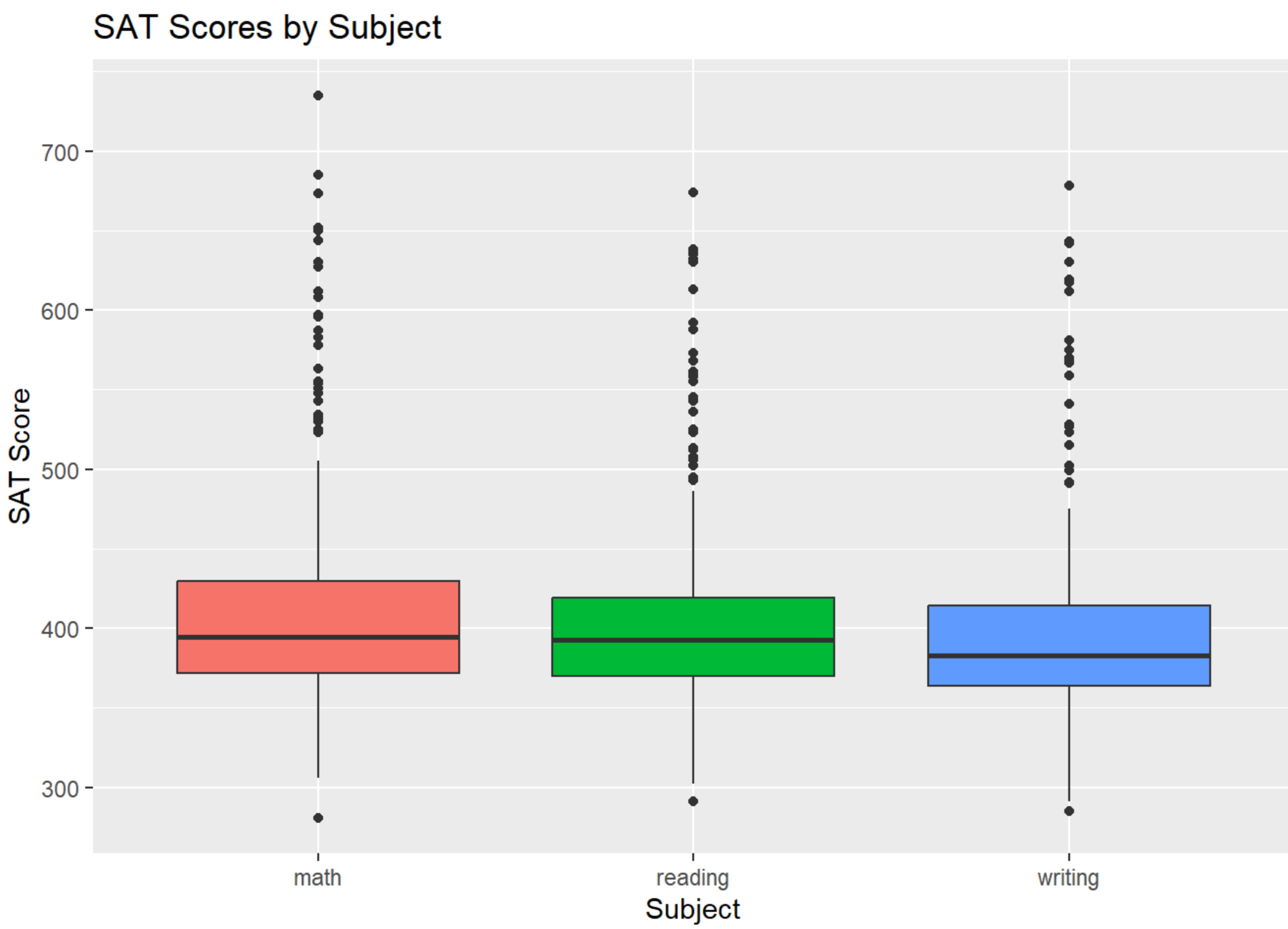
Comparison	Adjusted p-value	Result
Math vs Reading	0.1332	No Significant Difference
Math vs Writing	0.0013	Significant Difference
Writing vs Reading	0.3866	No Significant Difference

Only Math and Writing shows significant difference in mean SAT score.

Box Plot

To visualize difference between mean SAT scores across each subject.

```
ggplot(sat_aov, aes(x = subject, y = score, fill = subject)) +
  geom_boxplot() +
  labs(title = "SAT Scores by Subject",
       x = "Subject",
       y = "SAT Score") +
  theme(legend.position = "none")
```



Paired T-Test

Math vs Writing

H_0: There is no difference in Mathematics mean score and Writing Mean score across all schools

H_a: There is significant difference in Mathematics mean score and Writing Mean score across all schools

H_0 : diff $= 0$

H_a : diff $\neq 0$

Method 1: Using t.test (Assumes CLT condition are satisfied)

```
t.test(x = sat_math_writing$math,
       y = sat_math_writing$writing,
       alternative = "two.sided",
       paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  sat_math_writing$math and sat_math_writing$writing
## t = 10.71, df = 385, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  12.44724 18.04498
## sample estimates:
## mean difference
##        15.24611
```

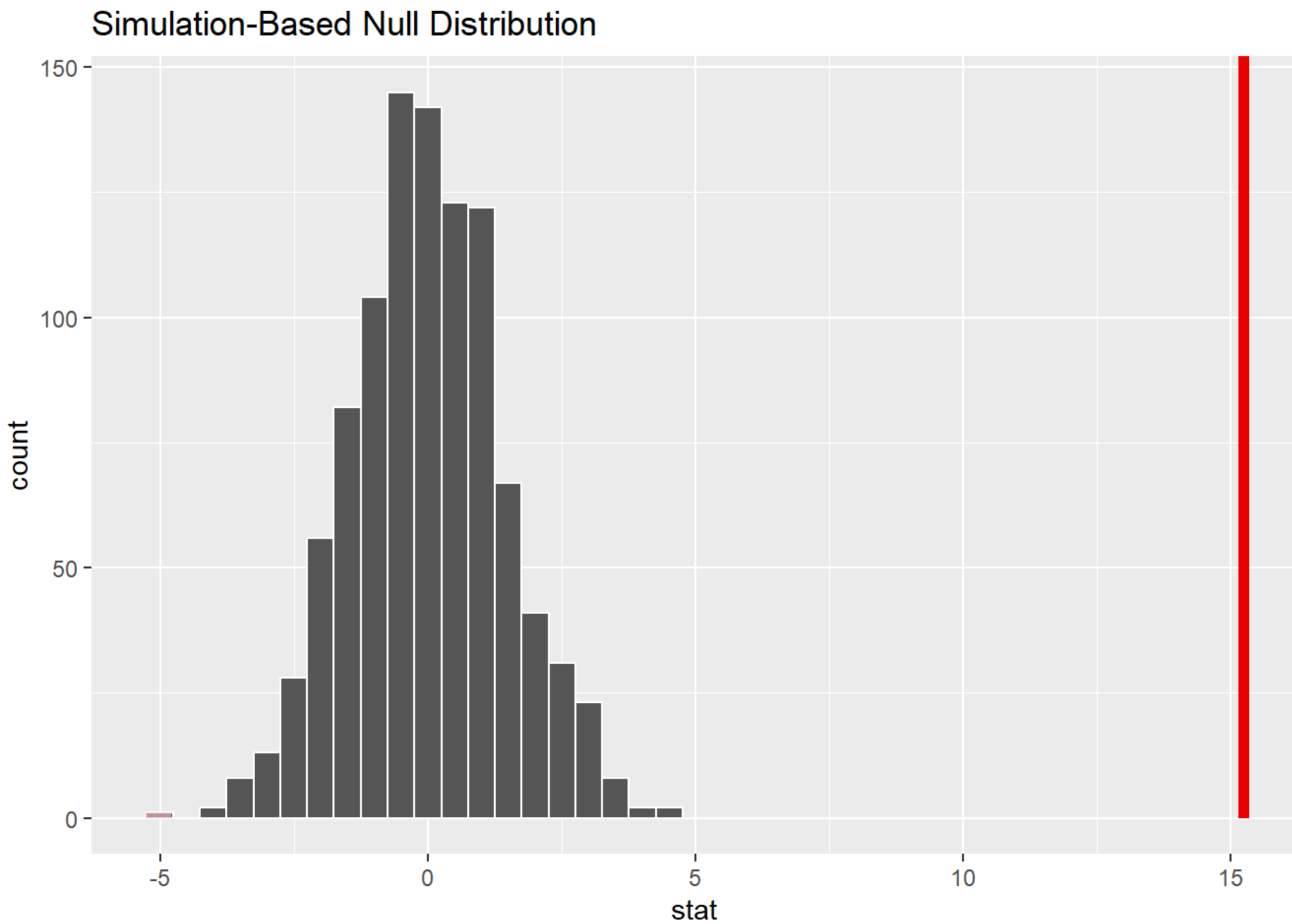
Method 2: Using Bootstrap re-samples

```
null_dist <- sat_math_writing |>
  specify(response = diff) |>
  hypothesize(null = "point", mu = 0) |>
  generate(reps = 1000, type = "bootstrap") |>
  calculate(stat = "mean")

obs_stat <- mean(sat_math_writing$diff)

get_p_value(null_dist, obs_stat, direction = "two.sided")
```

```
## 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 × 1
##   p_value
##     <dbl>
## 1       0
```

```
visualize(null_dist, bins = 20) +
  shade_p_value(obs_stat = obs_stat, direction = "two.sided")
```



Conclusion (Math vs Writing)

The both the method above show that p value is less than the significance level therefore, rejecting the null hypothesis and concluding that difference between math and writing mean score across school is significant

Math vs Reading

H_0: There is no difference in Mathematics mean score and Reading Mean score across all schools

H_a: There is significant difference in Mathematics mean score and Reading Mean score across all schools

H_0 : diff $= 0$

H_a : diff $\neq 0$

Method 1: Using t.test (Assumes CLT condition are satisfied)

```
t.test(x = sat_math_reading$math,
       y = sat_math_reading$reading,
       alternative = "two.sided",
       paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  sat_math_reading$math and sat_math_reading$reading
## t = 5.8569, df = 385, p-value = 1.012e-08
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   5.770501 11.602556
## sample estimates:
## mean difference
##        8.686528
```

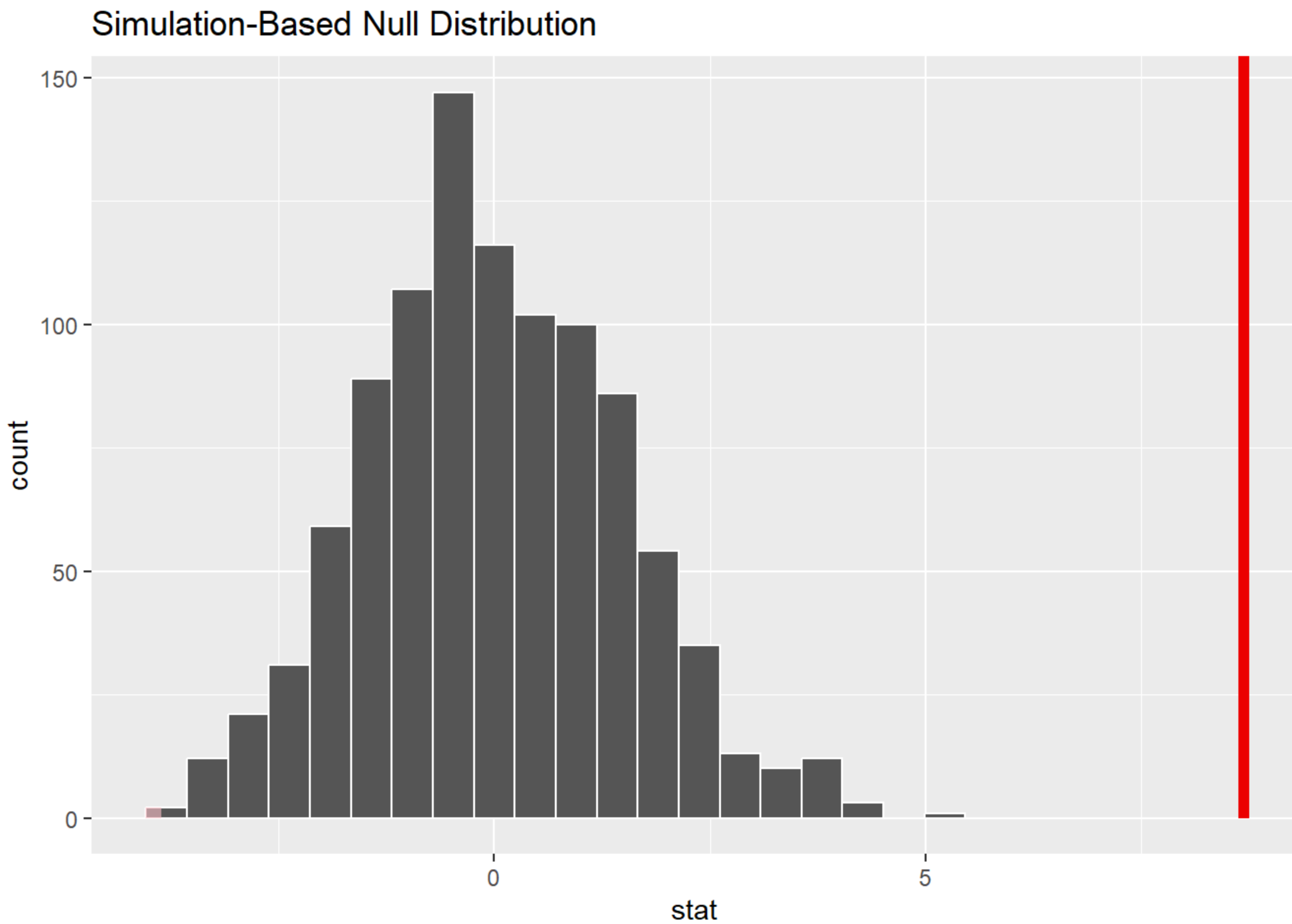
Method 2: Using Bootstrap re-samples

```
null_dist <- sat_math_reading |>
  specify(response = diff) |>
  hypothesize(null = "point", mu = 0) |>
  generate(reps = 1000, type = "bootstrap") |>
  calculate(stat = "mean")

obs_stat <- mean(sat_math_reading$diff)

get_p_value(null_dist, obs_stat, direction = "two.sided")
```

```
## 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 × 1
##   p_value
##     <dbl>
## 1       0
```

```
visualize(null_dist, bins = 20) +
  shade_p_value(obs_stat = obs_stat, direction = "two.sided")
```



Conclusion (Math vs Reading)

The both the method above show that p value is less than the significance level therefore, rejecting the null hypothesis and concluding that difference between math and reading mean score across school is significant

Writing vs Reading

H_0: There is no difference in Writing mean score and Reading Mean score across all schools

H_a: There is significant difference in Writing mean score and Reading Mean score across all schools

H_0 : diff $= 0$

H_a : diff $\neq 0$

Method 1: Using t.test (Assumes CLT condition are satisfied)

```
t.test(x = sat_writing_reading$writing,
       y = sat_writing_reading$reading,
       alternative = "two.sided",
       paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  sat_writing_reading$writing and sat_writing_reading$reading
## t = -10.791, df = 385, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -7.754799 -5.364372
## sample estimates:
## mean difference
##       -6.559585
```

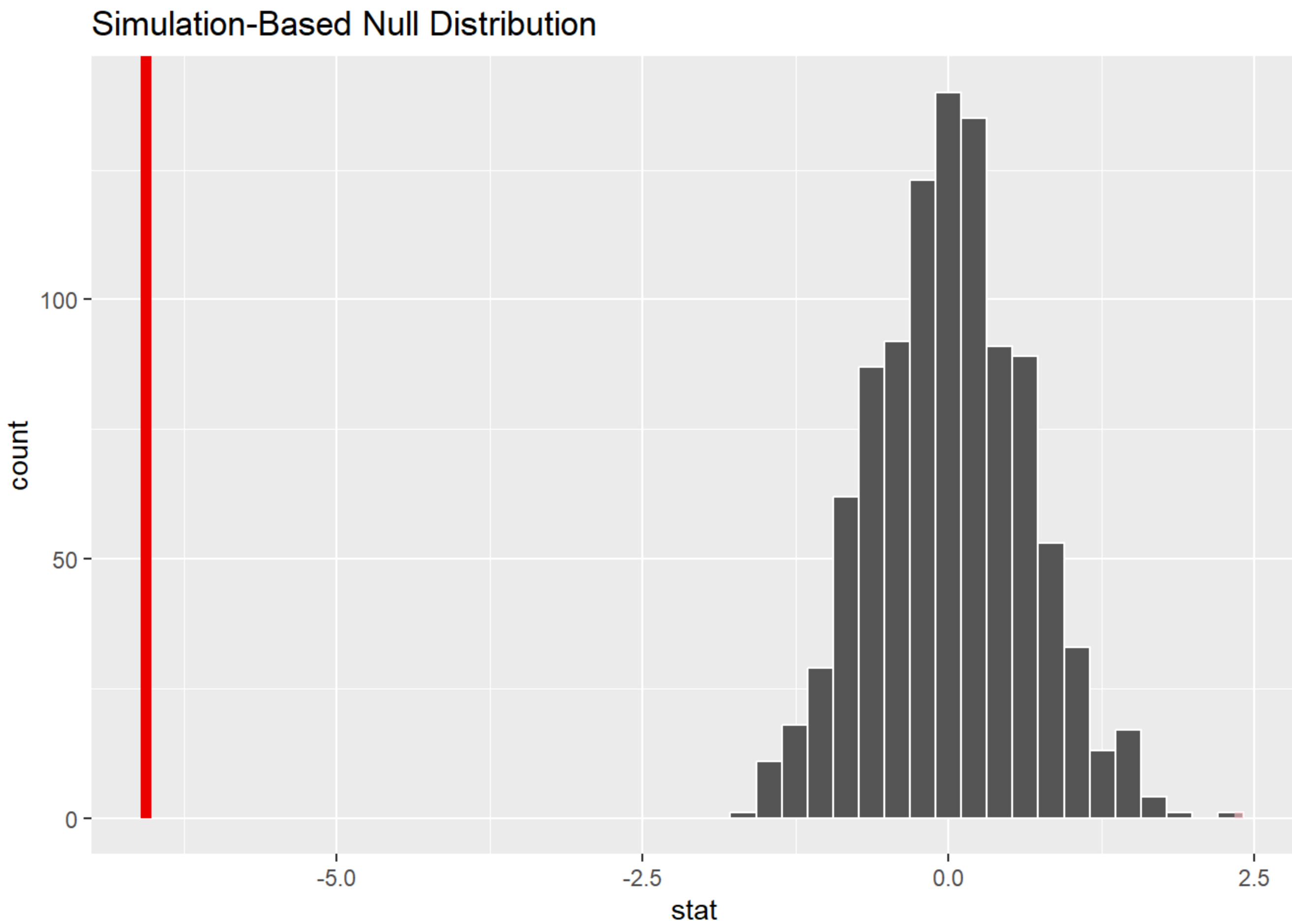
Method 2: Using Bootstrap re-samples

```
null_dist <- sat_writing_reading |>
  specify(response = diff) |>
  hypothesize(null = "point", mu = 0) |>
  generate(reps = 1000, type = "bootstrap") |>
  calculate(stat = "mean")

obs_stat <- mean(sat_writing_reading$diff)

get_p_value(null_dist, obs_stat, direction = "two.sided")
```

```
## 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 × 1
##   p_value
##     <dbl>
## 1       0
```

```
visualize(null_dist, bins = 20) +
  shade_p_value(obs_stat = obs_stat, direction = "two.sided")
```



Conclusion (Writing vs Reading)

The both the method above show that p value is less than the significance level therefore, rejecting the null hypothesis and concluding that difference between writing and reading mean score across school is significant.

Independent T-Test (Welch Two Sample T-Test)

Math vs Writing

$H_0 : \mu_{math} - \mu_{writing} = 0$

$H_a : \mu_{math} - \mu_{writing} \neq 0$

```
t.test(x = sat_math_writing$math,
       y = sat_math_writing$writing,
       alternative = "two.sided",
       paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  sat_math_writing$math and sat_math_writing$writing
## t = 3.4449, df = 759.54, p-value = 0.0006025
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.558154 23.934074
## sample estimates:
## mean of x mean of y
##  412.9352  397.6891
```

Conclusion (Math vs Writing)

p-values $= 0.006025 < 0.05$. Reject Null Hypothesis. Mean SAT score in math and writing have significant difference.

Math vs Reading

$H_0 : \mu_{math} - \mu_{reading} = 0$

$H_a : \mu_{math} - \mu_{reading} \neq 0$

```
t.test(x = sat_math_reading$math,
       y = sat_math_reading$reading,
       alternative = "two.sided",
       paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  sat_math_reading$math and sat_math_reading$reading
## t = 1.977, df = 756.49, p-value = 0.0484
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.06108013 17.31197686
## sample estimates:
## mean of x mean of y
##  412.9352  404.2487
```

Conclusion (Math vs Reading)

p-values $= 0.0484 < 0.05$. Reject Null Hypothesis. Mean SAT score in math and reading have significant difference.

Note: This result is different than pairwise t-test performed before why?

The result differs due to usage of Bonferroni correction made in pairwise t-test to reduce false positives.

current p_value $= 0.0484$

adjusted p_value $=$ current p_value $\cdot 3 = 0.1452$ which is closer to p_value in pairwise t-test for math vs reading, which was 0.1332

Writing vs Reading

$H_0 : \mu_{writing} - \mu_{reading} = 0$

$H_a : \mu_{writing} - \mu_{reading} \neq 0$

```
t.test(x = sat_writing_reading$writing,
       y = sat_writing_reading$reading,
       alternative = "two.sided",
       paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  sat_writing_reading$writing and sat_writing_reading$reading
## t = -1.5906, df = 769.79, p-value = 0.1121
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.65501   1.53584
## sample estimates:
## mean of x mean of y
##  397.6891  404.2487
```

Conclusion (Writing vs Reading)

p-values $= 0.1121 > 0.05$. Accept Null Hypothesis. Mean SAT score in writing and reading does not have significant difference.