

olympic medal htest

```
library(tidyverse)
library(broom)
```

Feed in data

```
olympics <- read_csv("data/olympics.csv") |>
  drop_na(age)

## Rows: 271116 Columns: 15

## -- Column specification -----
## Delimiter: ","
## chr (10): name, sex, team, noc, games, season, city, sport, event, medal
## dbl (5): id, age, height, weight, year

##
## 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.
```

Data for test

```
olympics_df <- olympics |>
  select(age, medal) |>
  mutate(medal = case_when(medal == "Gold" ~ 1,
                           medal == "Silver" ~ 1,
                           medal == "Bronze" ~ 1,
                           TRUE ~ 0)) |>
  mutate(age = case_when(age < 25 ~ "Young",
                        age >= 25 ~ "Old"))
olympics_df
```

```
## # A tibble: 261,642 x 2
##   age    medal
##   <chr> <dbl>
## 1 Young    0
## 2 Young    0
## 3 Young    0
## 4 Old      1
## 5 Young    0
## 6 Young    0
## 7 Old      0
## 8 Old      0
## 9 Old      0
## 10 Old     0
## # ... with 261,632 more rows
```

Summary of data

```
olympics_summary <- olympics_df |>
  group_by(age) |>
  summarize(
    medal= sum(medal),
    n = n()
  ) |>
  arrange(desc(age))
olympics_summary
```

```
## # A tibble: 2 x 3
##   age  medal    n
##   <chr> <dbl> <int>
## 1 Young 17939 131134
## 2 Old   21112 130508
```

Prop test

```
test <- prop.test(olympics_summary$medal, olympics_summary$n,
  correct = FALSE, alternative = c("greater"))
result <- tidy(test)
result
```

```
## # A tibble: 1 x 9
##   estimate1 estimate2 statistic p.value parameter conf.low conf.high method
##   <dbl>      <dbl>      <dbl>  <dbl>      <dbl>      <dbl>      <dbl> <chr>
## 1    0.137    0.162    321.      1          1 -0.0273      1 2-sample t~
## # ... with 1 more variable: alternative <chr>
```

Given p-value > 0.05, we fail to reject the null hypothesis and conclude that the proportion of athletes younger than 25 that win a medal is equal to the proportion of athletes of age 25 or older that win a medal.