

California's Unlevel Playing Field

Emma Durler, Kylie Scharf, Nicole Yee, Curtis LeQuang

Table of contents

```
library(tidyverse)
library(readxl)
library(writexl)
library(kableExtra)

#loading in data
#tuesdata <- tidyuesdayR::tt_load('2022-03-29')
#sports_dat <- tuesdata$sports

#data set creation that doesn't call git hub
sports_dat <- read_excel("data/original_sub/original.xlsx")

#subsetting and moving data
sports_dat <- sports_dat |>
  filter(sports == "Baseball" | sports == "Softball", state_cd == "CA") |>
  rename(Sport = sports) |>
  group_by(Sport, year) |>
  pivot_wider(id_cols = year:sector_name, names_from = Sport, values_from = exp_men:exp_women) |>
  mutate(diff = exp_men_Baseball - exp_women_Softball) |>
  drop_na(exp_men_Baseball, exp_women_Softball, diff)

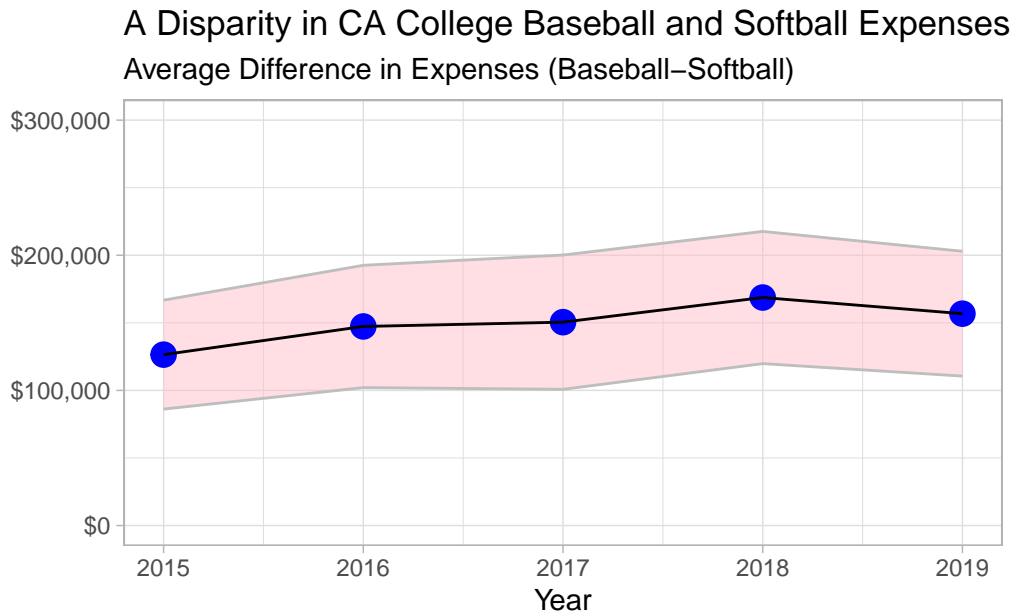
plot <- sports_dat |>
  group_by(year) |>
  summarize(mean_exp_men = mean(exp_men_Baseball),
            mean_exp_women = mean(exp_women_Softball),
            mean_diff = mean(diff),
            stan = sd(diff)/sqrt(n())
  ) |>
```

```

ggplot(aes(y = mean_diff, x = year)) +
  geom_ribbon(aes(ymin = mean_diff - 2*stan, ymax = mean_diff + 2*stan),
            alpha=.5,
            color="grey",
            fill = "pink") +
  theme_bw() +
  theme_light() +
  geom_point(size=4, color="blue") +
  geom_line() +
  labs(x = "Year",
       subtitle = "Average Difference in Expenses (Baseball-Softball)",
       y = "",
       title = "A Disparity in CA College Baseball and Softball Expenses",
       caption = "Source: Department of Education") +
  theme(plot.caption = element_text(hjust = 0)) +
  scale_y_continuous(labels = scales::dollar, limits=c(0,300000))

```

plot



Source: Department of Education

Figure 1: Above entails an analysis of the average differences in expenses from 2015 to 2019 for collegiate baseball and softball in California. The average difference in expenses for a particular year for all California teams is plotted as a blue circle. The pink highlighted region represents adding or subtracting two standard errors of the average difference from the plotted average differences. This shaded region therefore represents a roughly 95% confidence interval for the true average difference for all collegiate baseball and softball teams in California between 2015 to 2019 that meet the conditions required for data collection. The average difference in expenses are consistently greater than 0 along with the interval for said average differences, meaning the average expenses for baseball are consistently significantly larger than the expenses for softball.

```
#data table using kable
table <- sports_dat |>
  group_by(year) |>
  summarize(mean_exp_men = mean(exp_men_Baseball),
            mean_exp_women = mean(exp_women_Softball),
            mean_diff = mean(diff)
  )

kable(table,
      caption="Average expenses for baseball and softball and the average differences between them")
```