California's Unlevel Playing Field

Emma Durler, Kylie Scharf, Nicole Yee, Curtis LeQuang

Table of contents

```r
library(tidyverse)
library(writexl)
#loading in data
tuesdata <- tidytuesdayR::tt_load('2022-03-29')
sports_dat <- tuesdata$sports


#subsetting and moving data
library(tidyverse)
sports_dat <- sports_dat |>
  filter(sports == "Baseball" | sports == "Softball") |>
  mutate(
    rev_men = ifelse(is.na(rev_men), 0, rev_men),
    rev_women = ifelse(is.na(rev_women), 0, rev_women),
    exp_men = ifelse(is.na(exp_men), 0, exp_men),
    exp_women = ifelse(is.na(exp_women), 0, exp_women),
  ) |>
  filter(state_cd == "CA")

sports_dat <- sports_dat |>
  rename(Sport = sports)


plot <- sports_dat |>
  group_by(Sport, year) |>
  filter(total_exp_menwomen != 0) |>
  summarise(mean_exp = mean(total_exp_menwomen), stan = sd(total_exp_menwomen)/sqrt(n())) 
  ggplot(aes(y = mean_exp, x = year, color = Sport, shape=Sport, fill=Sport)) +
   geom_ribbon(aes(ymin = mean_exp - stan, ymax = mean_exp + stan), alpha=.5, color=NA) +
```

```
  geom_point(size=4) +
  geom_line() +
  labs(x = "Year", subtitle = "Average Yearly Expense", y = "", title = "Fig 1. California
theme(plot.caption = element_text(hjust = 0)) +
scale_y_continuous(labels = scales::dollar)

plot
```

Fig 1. California College Baseball And Softball Expenses (
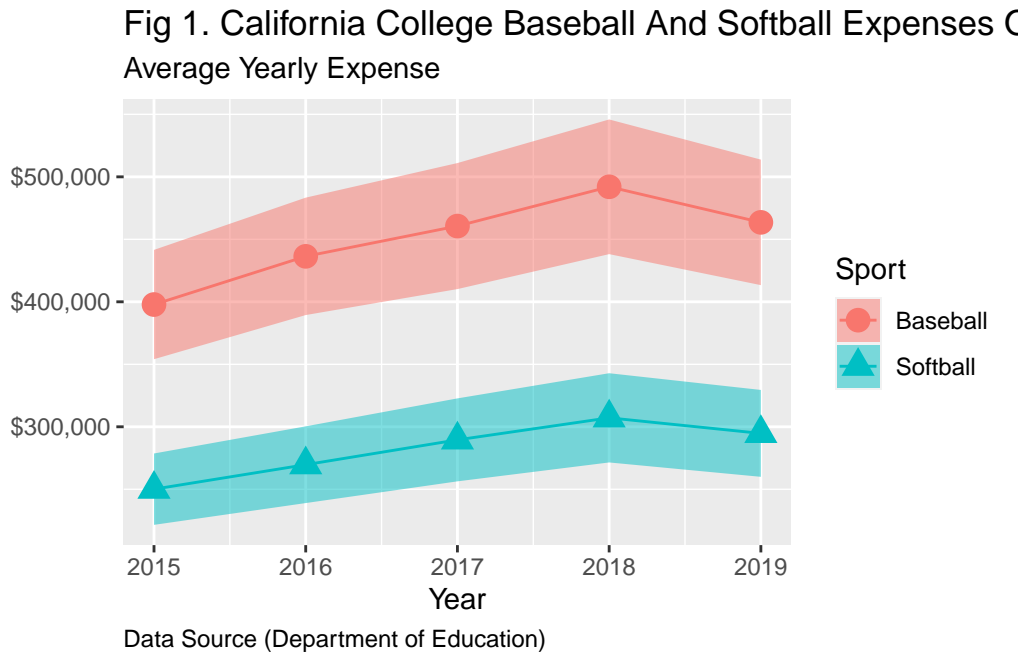Average Yearly Expense

Figure 1: Above entails an analysis of the differences in average expenses from 2015 to 2019
for collegiate baseball and softball in California. The Department of Education
collects data yearly on collegiate sports budgets in accordance to The Equity in
Althletics Disclosure Act. The graph above consists of such data subsetted to contain
soley softball and baseball teams from California in the years 2015 to 2019. The
expense for a particular year averaged over all California softball teams is plotted
as a blue triangle, wheras the expense averaged over all California baseball teams
in a particular year is plotted as a red circle. The highlighted region represents
adding or subtracting one standard error of the average from the plotted averages.
There appears to be more variability in the average expenses for baseball than for
softball in any particular year in California. The average expenses for baseball are
consistantly signifigantly larger than the expenses for women, although the expenses
for both sports tend to increase at a similiar rate over time.

```r
library(rstatix)
library(kableExtra)
table <- sports_dat |>
  group_by(Sport, year) |>
  filter(total_exp_menwomen != 0) |>
  summarise(mean_exp = mean(total_exp_menwomen)) |>
  pivot_wider(id_cols = year, names_from = Sport, values_from = mean_exp) |>
  rename(Year=year, Expenses_Men=Baseball, Expenses_Women=Softball)

kable(table,
      caption="CA Collegiate Baseball and Softball Expenses",
      col.names = c("Year",
                    "Expenses Men ($)",
                    "Expenses Women ($)")) |>
  kable_classic() |>
  kable_styling(full_width = FALSE, position = "center") |>
  footnote(general="Source: Department of Education")
```

Table 1: CA Collegiate Baseball and Softball Expenses

Year	Expenses Men ($)	Expenses Women ($)
2015	397818.2	250096.5
2016	436379.8	269675.2
2017	460608.8	289536.4
2018	492032.7	307159.0
2019	463564.6	294723.3

Note:

Source: Department of Education

```r
t_test(sports_dat, total_exp_menwomen~Sport, alternative="greater")
```

```
# A tibble: 1 x 8
  .y.                group1   group2     n1    n2 statistic    df        p
* <chr>              <chr>    <chr>   <int> <int>     <dbl> <dbl>    <dbl>
1 total_exp_menwomen Baseball Softball  775   693      6.36 1321. 1.35e-10
```