csas data challenege

Bonos

2026-01-10

```r
library("tidyverse")
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.1.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library("fastDummies")
library("tidymodels")
```

```
## -- Attaching packages -------------------------------------- tidymodels 1.4.1 --
## v broom        1.0.9     v rsample      1.3.1
## v dials        1.4.2     v tailor       0.1.0
## v infer        1.0.9     v tune         2.0.0
## v modeldata    1.5.1     v workflows    1.3.0
## v parsnip      1.3.3     v workflowsets 1.1.1
## v recipes      1.3.1     v yardstick    1.3.2
## -- Conflicts ----------------------------------------- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
```

```r
library("ggforce")
library("ggplot2")

setwd("~/Desktop/R Working Directory/Data")
stones_pp <- read.csv("powerplay.csv") %>%
  as_tibble()

stones_pp
```

```
## # A tibble: 26,370 x 37
##    CompetitionID SessionID GameID EndID ShotID TeamID PlayerID  Task Handle
##            <int>     <int>  <int> <int>  <int>  <int>    <int> <int>  <int>
##  1             0         1      1     1      7     19        1     0      0
##  2             0         1      1     1      8     27        1     0      1
##  3             0         1      1     1      9     19        2     0      0
##  4             0         1      1     1     16     27        2     0      0
##  5             0         1      1     1     17     19        2     7      0
##  6             0         1      1     1     18     27        2     6      1
##  7             0         1      1     1     19     19        2     0      1
##  8             0         1      1     1     20     27        2     0      1
##  9             0         1      1     1     21     19        1     2      0
## 10             0         1      1     1     22     27        1     0      0
## # i 26,360 more rows
## # i 28 more variables: Points <int>, TimeOut <dbl>, stone_1_x <dbl>,
## #   stone_1_y <dbl>, stone_2_x <dbl>, stone_2_y <dbl>, stone_3_x <dbl>,
## #   stone_3_y <dbl>, stone_4_x <dbl>, stone_4_y <dbl>, stone_5_x <dbl>,
## #   stone_5_y <dbl>, stone_6_x <dbl>, stone_6_y <dbl>, stone_7_x <dbl>,
## #   stone_7_y <dbl>, stone_8_x <dbl>, stone_8_y <dbl>, stone_9_x <dbl>,
## #   stone_9_y <dbl>, stone_10_x <dbl>, stone_10_y <dbl>, stone_11_x <dbl>, ...
```

```r
library(dplyr)

# Load data
df <- read.csv("/Users/mbonos/Desktop/R Working Directory/Data/powerplay.csv")
end_df <- df %>%
  group_by(CompetitionID, SessionID, GameID, EndID, TeamID) %>%
  slice_max(ShotID, n = 1, with_ties = FALSE) %>%
  ungroup() %>%
  mutate(PowerPlayUsed = as.integer(!is.na(PowerPlay) & PowerPlay != "" & PowerPlay != 0))

# Create opponent result
end_df <- end_df %>%
  group_by(CompetitionID, SessionID, GameID, EndID) %>%
  mutate(OppResult = Result[3 - row_number()]) %>%
  ungroup()

# Run model
m3 <- lm(OppResult ~ PowerPlayUsed + EndID + factor(TeamID), data = end_df)
summary(m3)
```

```
##
## Call:
## lm(formula = OppResult ~ PowerPlayUsed + EndID + factor(TeamID),
##     data = end_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6281 -0.8892 -0.3899  0.4748  8.4081
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.098561   0.095679   1.030 0.303000
## PowerPlayUsed     -1.022693   0.069602 -14.693  < 2e-16 ***
## EndID             0.178336   0.009742  18.306  < 2e-16 ***
## factor(TeamID)11  0.128538   0.134669   0.954 0.339888    
## factor(TeamID)12  0.262122   0.207065   1.266 0.205608    
## factor(TeamID)13  0.924533   0.191497   4.828 1.42e-06 ***
## factor(TeamID)14  0.388468   0.130090   2.986 0.002838 ** 
## factor(TeamID)15  0.304891   0.136756   2.229 0.025826 *  
## factor(TeamID)16  0.100299   0.133438   0.752 0.452293    
## factor(TeamID)17  0.112959   0.119824   0.943 0.345874    
## factor(TeamID)18  0.122565   0.123287   0.994 0.320197    
## factor(TeamID)19  0.200939   0.122878   1.635 0.102052    
## factor(TeamID)20  0.093702   0.122378   0.766 0.443902    
## factor(TeamID)22  0.246964   0.123936   1.993 0.046349 *  
## factor(TeamID)23  0.714897   0.215501   3.317 0.000915 ***
## factor(TeamID)24  0.089294   0.122374   0.730 0.465619    
## factor(TeamID)25  0.355519   0.136945   2.596 0.009456 ** 
## factor(TeamID)27  0.088677   0.186094   0.477 0.633727    
## factor(TeamID)28  0.524480   0.199795   2.625 0.008688 ** 
## factor(TeamID)30  0.477594   0.158661   3.010 0.002624 ** 
## factor(TeamID)34  0.576257   0.137760   4.183 2.92e-05 ***
## factor(TeamID)36  0.328410   0.200930   1.634 0.102225    
## factor(TeamID)37  0.180226   0.127870   1.409 0.158763    
## factor(TeamID)43  0.377998   0.134489   2.811 0.004963 ** 
## factor(TeamID)44  0.488735   0.134847   3.624 0.000292 ***
## factor(TeamID)46  0.175403   0.123504   1.420 0.155602    
## factor(TeamID)51  0.448839   0.134858   3.328 0.000880 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.528 on 5247 degrees of freedom
## Multiple R-squared:  0.08614,    Adjusted R-squared:  0.08162 
## F-statistic: 19.02 on 26 and 5247 DF,  p-value: < 2.2e-16
```

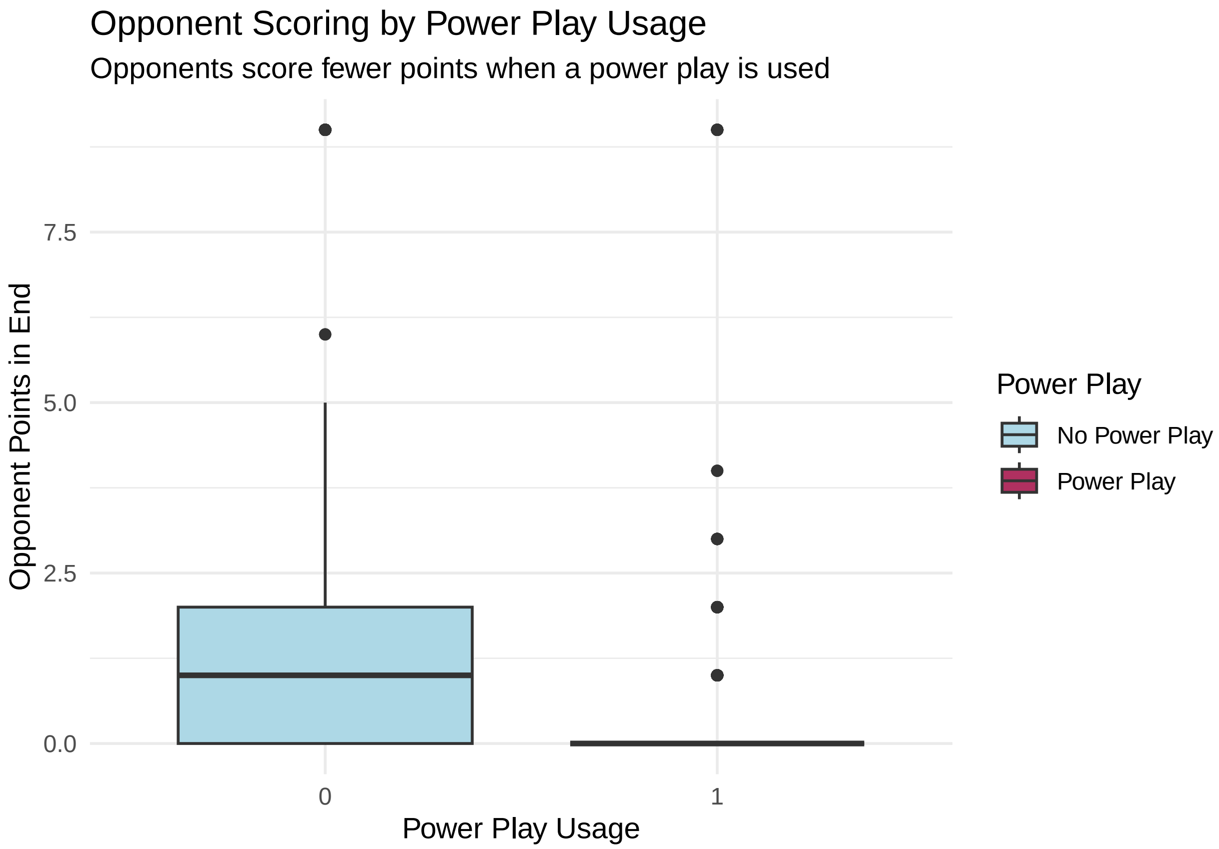
When a team uses a power play, the opponent scores about 1.02 fewer points in that end on average, holding end number and team constant.

Methodology. We construct an end-level dataset by collapsing the shot-level data to one observation per team per end, retaining the final board state and the points scored in that end. For each team–end, we identify whether the team used a power play and pair each observation with the opponent's score from the same end. We then estimate a linear regression with opponent points as the dependent variable and power play usage as the primary independent variable, controlling for end number and team fixed effects. This specification isolates the effect of power play usage on opponent scoring while accounting for strategic differences across ends and persistent differences in team quality.

Results and interpretation. The results show that using a power play reduces the opponent's scoring by approximately one point in the power play end, and this effect is highly statistically significant. In the context of curling, where most ends yield only zero to two points and matches are frequently decided by narrow margins, this represents a substantial defensive advantage. Rather than improving shot execution directly, the power play appears to reshape the end in a way that limits the opponent's ability to generate scoring opportunities, making it harder for them to place stones effectively or create multi-point ends. This finding suggests that the primary value of the power play lies in its strategic and positional impact, suppressing opponent scoring and thereby shifting the expected point differential in favor of the team that uses it.

```r
library(ggplot2)

ggplot(end_df, aes(x = factor(PowerPlayUsed), y = OppResult, fill = factor(PowerPlayUsed))) +
  geom_boxplot() +
  scale_fill_manual(
    values = c("0" = "lightblue", "1" = "maroon"),
    labels = c("No Power Play", "Power Play")
  ) +
  labs(
    x = "Power Play Usage",
    y = "Opponent Points in End",
    fill = "Power Play",
    title = "Opponent Scoring by Power Play Usage",
    subtitle = "Opponents score fewer points when a power play is used"
  ) +
  theme_minimal()
```


Opponent Scoring by Power Play Usage
Opponents score fewer points when a power play is used
Opponent Points in End
Power Play Usage
Power Play
No Power Play
Power Play