Stats Modeling Project

Group 5: Xin Jin, Reid Ginoza, Heidi Lovejoy

11/18/2019

```
institution_data <- read_spss('InstLevel.sav') %>%
  select(institution_name, state_cd, ClassificationCode, Classification=classification_name, sector_cd,
  filter(grepl('4-year', sector_name))  %>% # only 4-year Schools
  filter(is.na(IL_PARTIC_COED_MEN) | IL_PARTIC_COED_MEN == 0 ) %>% # only schools with no male particip
  filter(is.na(IL_PARTIC_COED_WOMEN) | IL_PARTIC_COED_WOMEN == 0 ) %>% # only schools with no female pa
  select(-contains("COED")) %>% # ignore variables with the word "coed"
  filter(!is.na(HDCOACH_SALARY_MEN)) %>% # filter out NA
  filter(!is.na(HDCOACH_SALARY_WOMEN)) %>% # filter out NA
  type.convert()

ncaa.div.i_data <- institution_data %>%
  filter(grepl('NCAA Division I-', Classification)) # only NCAA division 1

men_data <- institution_data %>%
  select(-contains("WOMEN")) %>%
  rename(Salary = HDCOACH_SALARY_MEN, IL_PARTIC = IL_PARTIC_MEN)

women_data <- institution_data %>%
  select(-contains("_MEN")) %>%
  rename(Salary = HDCOACH_SALARY_WOMEN, IL_PARTIC = IL_PARTIC_WOMEN)

gender.separated_data <- bind_rows(
  men_data %>% add_column(Gender = "Men"),
  women_data %>% add_column(Gender = "Women")
)
```

Introduction

While it seems to be common knownledge that some of the highest paid public employees are college sports teams coaches, less is known about the salaries of college-level sports coaches in general. This project examined predictors of head coach salaries of sports teams at four-year colleges and universities in the United States. We used the institution-level data from the Equity in Athletics Survey, Year 2017-2018, from the U.S. Department of Education Office of Postsecondary Education (2018). Our analysis focused on the differences of head coach salaries between male and female teams, of classification, and of the number of student participants at an institution level. The classification consists of the athletic association, the division and subdivisions if applicable, such as NCAA Division I-FBS, NCAA Division I-FCS and NCAA Division I without football.

The full data set listed completed surveys for 2079 institutions. The 578 two-year colleges were excluded, as well as the 120 four-year schools with coed teams and 18 other institutions with missing values for head coach salaries. Our working data set had 1363 four-year institutions without coed sports teams.

Hypotheses

We examined whether the classifications and the gender of the team had effects on the head coach salaries. To do that, we first investigated the variances of head coach salaries of male teams compared with that of female teams. We checked for interaction effects between classification and gender of team. We also looked at whether the number of participants institution-wide predicted the salary for the head coaches.

Methods

Data was analyzed using R version 3.6.1 (2019-07-05) (R Core Team 2019). The data was described using counts (percentage) for categorical variables and mean (standard deviation) for continuous variables.

To examine the interactions of gender and classification, a two-way ANOVA was used. A Brown-Forsythe-Levene Test for homogeneity of variances was conducted on the variances of the classifications. After an examination of the residuals, a Kruskal-Wallis was used to test if there are differences between the classifications, and the posthoc Kruskal-Wallis procedure was used for pairwise comparisons.

Results

Description of Data

| Variable	Mean	SD
Head Coach–Men's Salary	107,801.45	230,718.91
Head Coach–Women's Salary	48,110.37	51,009.74
Head Coach–All Salary	77,955.91	169,698.1
Participation–Men	218.26	123.14
Participation–Women	169.02	94.32
Participation–Total	387.28	207.9

```
institution_data %>% group_by(Classification) %>% summarize(Count = n()) %>%
  mutate(Percentage = percent(Count / 1363)) %>% kable(booktabs = T)
```

Classification	Count	Percentage
Independent	4	0.3%
NAIA Division I	90	6.6%
NAIA Division II	102	7.5%
NCAA Division I without football	95	7.0%
NCAA Division I-FBS	117	8.6%
NCAA Division I-FCS	114	8.4%
NCAA Division II with football	160	11.7%
NCAA Division II without football	138	10.1%
NCAA Division III with football	223	16.4%
NCAA Division III without football	162	11.9%
NCCAA Division I	9	0.7%
NCCAA Division II	25	1.8%
NJCAA Division I	35	2.6%
NJCAA Division II	9	0.7%
NJCAA Division III	4	0.3%
NWAC	9	0.7%
Other	33	2.4%
USCAA	34	2.5%

Head Coach Salary

Analysis Results

```
classification.gender.interaction.results <- gender.separated_data %$% aov(Salary ~ Classification * Ge
```

Profile Plot of Salary by Classification and Gender
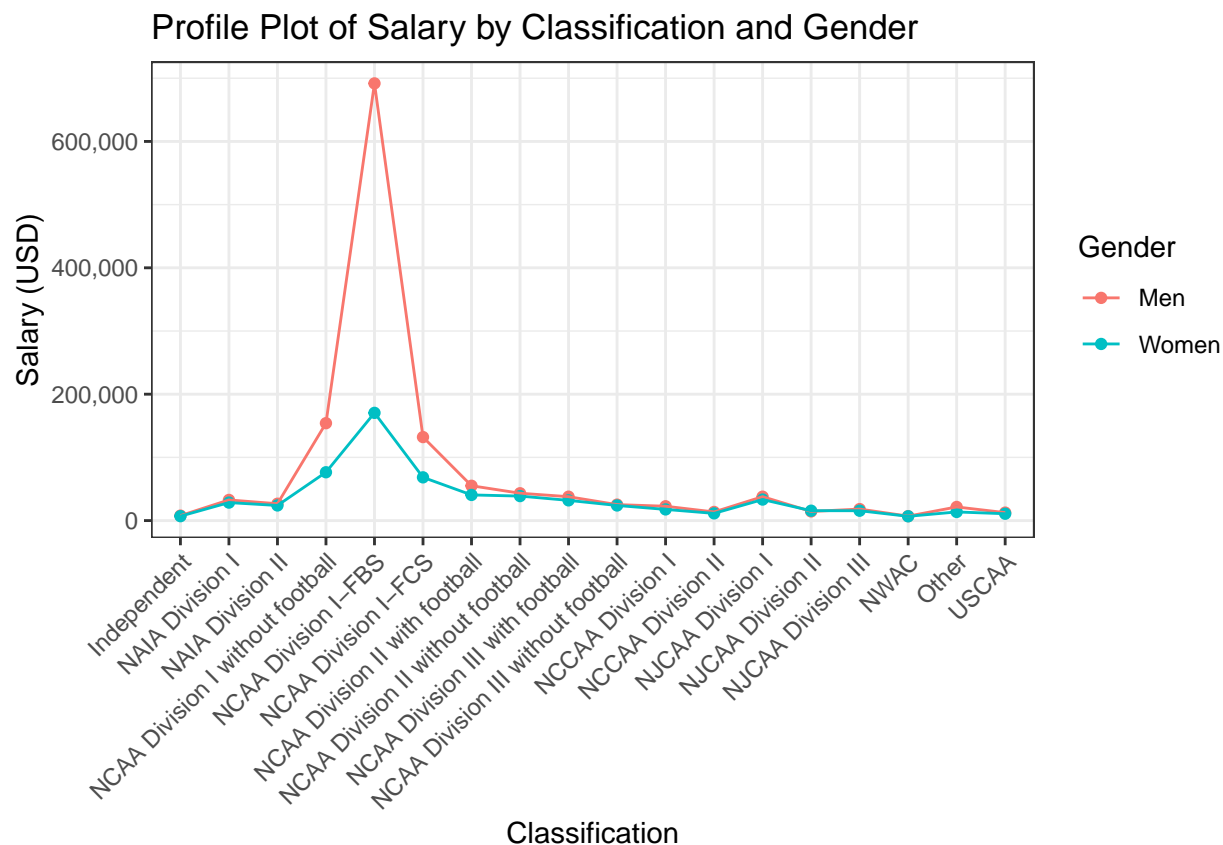
While this profile plot shows interaction effects, upon visual inspection, it seems to be limited to the three NCAA Division I classifications: NCAA Division I without football, NCAA Division I-FBS, and NCAA Division I-FCS. In those three cases, head coaches for men's teams have higher salaries, especially in the case of NCAA Division I-FBS.

To check the assumptions of the ANOVA model, we examined the variance across classifications to see if the variance was constant and the normality of the residuals.

```
gender.separated_data %$% leveneTest(Salary ~ Classification, center = median) -> classification.varian
```

The Brown-Forsythe-Levene Test for Homogeneity of Variances resulted in a test statistic of $F_0 = 77.03$ ($p < 0.0001$), so there is sufficient evidence to suggest that at least one of the classifications' salaries have a variance different from the others.

Upon a visual examination of the residuals and the q-q plot, it is also clear that the normality assumptions of the residuals was not satisfied.

```
classification.gender.interaction.results %>% almost_sas()
```
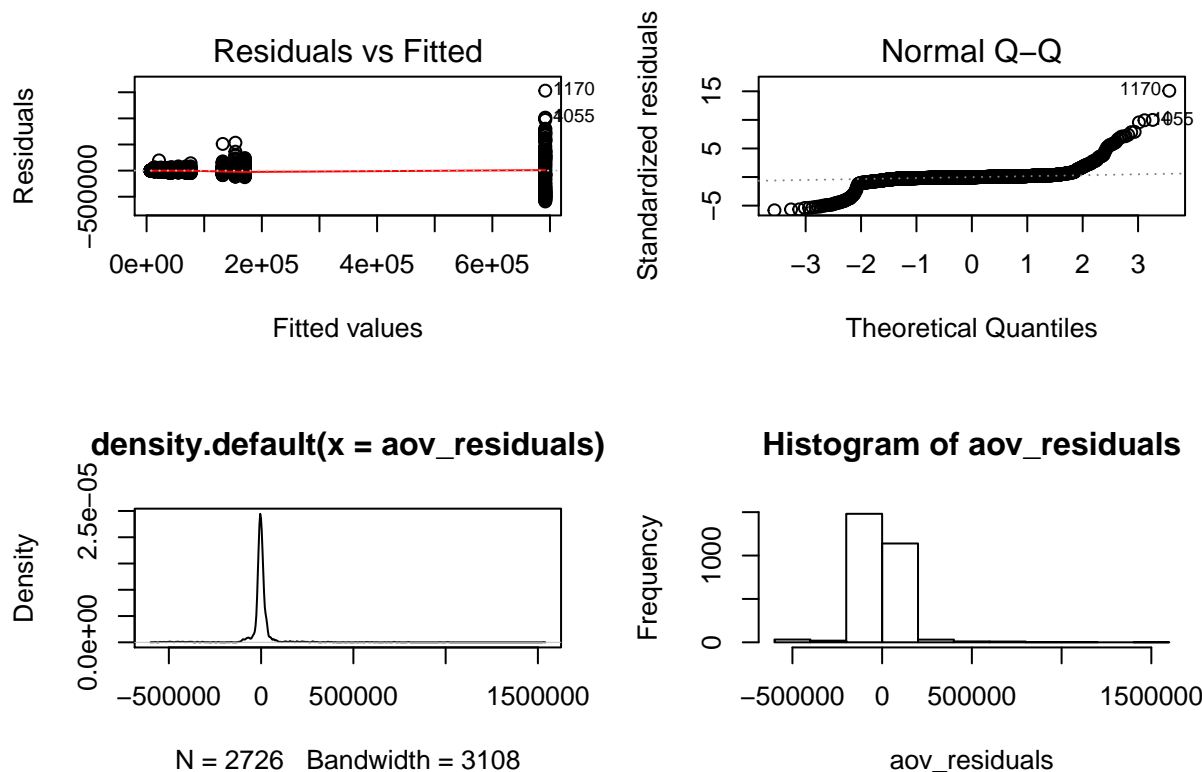
```
classification.kw.test <- gender.separated_data %$% kruskal.test(Salary ~ Classification) %>% tidy()
```

The Kruskal-Wallis test on the salaries grouped by classifications resulted in a test statistic $H = 1752.51$ ($p < 0.0001$), so there is sufficient evidence to suggest that at least one classification is different from the others. Thus, the posthoc procedure for the Kruskal-Wallis test was completed to find pairwise differences between the classifications.

```
classification.kw.posthoc <- gender.separated_data %$% kruskalmc(Salary ~ Classification)
```

From the 153 possible pairwise comparisons, 91 resulted in significant difference. Complete results are not included, but these are the summaries regarding the NCAA Division I: NCAA Division I-FBS had a significant difference from the 17 other classifications, NCAA Division I-FCS had a significant different from 16 other classifications, NCAA Division I without football the only exception, and similarly with NCAA Division I without football, the only non-significant difference was with NCAA Division I-FCS.

Conclusion

Examining the interaction effects between classification and gender, reveleated that the NCAA Division I had the most differences between male and female head coaches. Furthermore, it appears that the NCAA Division I, with its different subdivisions (FBS, FCS, without football) were each different from all of the remaining classifications. This makes sense as there are different funding allowances between these subdivisions, but only between football teams. A future study might investigate the head coach salaries of collegiate football head coaches separately from the other sports or compare head coaches across classification and sport. Perhaps the sport of football is where the greatest gender disparity of salary exists.

Our data set contained only the averages of head coach salaries of a given institution, so we were unable to capture any variance within an institution. Additionally, this study only used the surveys from one year and future studies should investigate the salaries over multiple years.

This was an initial study of salaries of head coaches of collegiate sports teams. While there is more work to be done, we believe we have paved the way in identifying important factors of their salaries, and we hope others will continue the investigation.

References

Office of Postsecondary Education. 2018. "Equity in Athletics Data Analysis." U.S. Department of Education. https://ope.ed.gov/athletics/#/datafile/list.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.