

Stats Modeling Project

Group 5: Xin Jin, Reid Ginoza, Heidi Lovejoy

11/18/2019

```
almost_sas <- function(aov.results){  
  par(mfrow=c(2,2))  
  plot(aov.results, which=1)  
  plot(aov.results, which=2)  
  aov_residuals <- residuals(aov.results)  
  plot(density(aov_residuals))  
  hist(aov_residuals)  
}
```

```
p.value.string = function(p.value){  
  p.value <- round(p.value, digits=4)  
  if (p.value == 0) {  
    return("p < 0.0001")  
  } else {  
    return(paste0("p = ", format(p.value, scientific = F)))  
  }  
}
```

```
institution_data <- as_tibble(read_spss('InstLevel.sav')) %>%  
  select( # TODO will change later when we decide what variables we want  
    -unitid, -addr1_txt, -addr2_txt, -city_txt, -zip_text, -sector_cd,  
    -ClassificationCode, -ClassificationOther  
  ) %>%  
  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"  
  mutate(Grand.Total.Profits = GRND_TOTAL_REVENUE - GRND_TOTAL_EXPENSE) %>%  
  filter(!is.na(HDCOACH_SALARY_MEN)) %>% # filter out NA  
  filter(!is.na(HDCOACH_SALARY_WOMEN)) # filter out NA  
ncaa.div.i_data <- institution_data %>%  
  filter(grepl('NCAA Division I-', classification_name)) # only NCAA division 1
```

Introduction

This project examines potential(?) salary differences between head coaches from public and private NCAA Division I institutions. We used the data from the Equity in Athletics Survey, Year 2017-2018, from the U.S. Department of Education Office of Postsecondary Education (2018).

Hypotheses

Divisions and Profits

Head Coaches of Men's Teams and School Type

H_0 :

H_1 :

```
wilcoxon.attempt <- institution_data %>%  
  select(classification_name, HDcoach_salary_men, HDcoach_salary_women) %>%  
  gather("Gender", "Salary", -classification_name)
```

```
## Warning: attributes are not identical across measure variables;  
## they will be dropped
```

```
wilcoxon.attempt$Gender %<>% recode(HDcoach_salary_men="Men", HDcoach_salary_women="Women")
```

```
wilcoxon.attempt
```

```
## # A tibble: 2,726 x 3  
##   classification_name      Gender Salary  
##   <chr>                <chr>   <dbl>  
## 1 NCAA Division I-FCS      Men    56920  
## 2 NCAA Division II without football Men    73267  
## 3 NCAA Division I-FCS      Men    84018  
## 4 NCAA Division I-FBS      Men   1705741  
## 5 NCAA Division II without football Men    37051  
## 6 NCAA Division I-FBS      Men   1360838  
## 7 NCAA Division III with football Men    44759  
## 8 NAIA Division I          Men    59471  
## 9 NCAA Division III with football Men    64314  
## 10 NCAA Division II with football Men    56806  
## # ... with 2,716 more rows
```

```
wilcoxon.attempt_results <- wilcox.test(Salary ~ Gender, alternative = 'two.sided', data = wilcoxon.att
```

```
wilcoxon.attempt_results
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: Salary by Gender  
## W = 1064179, p-value = 4.56e-11  
## alternative hypothesis: true location shift is not equal to 0
```

```
gender_wilcoxon_p.value <- wilcoxon.attempt_results$p.value %>% p.value.string
```

Hypotheses

$H_0: M_{male} = M_{female}$

$H_1: M_{male} \neq M_{female}$

Test Statistics

$$W = 1.064179 \times 10^6$$

***p*-value**

$$p < 0.0001$$

Rejection Region

Reject H_0 if $p < \alpha$, where $\alpha = 0.05$.

Conclusion

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/>.