

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)  
}
```

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

Introduction

This project examines NCAA Division I athletics. 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 :

Head Coaches of Men's Teams and Participation

H_0 :

H_1 :

Difference in Classifications

There are 18 different classifications, including NCAA Division I-FCS, NCAA II without Football, and so forth. This question looks at the means of the profits for each classification. $H_0: \mu_1 = \mu_2 = \dots = \mu_{18}$
 H_1 : At least one is different

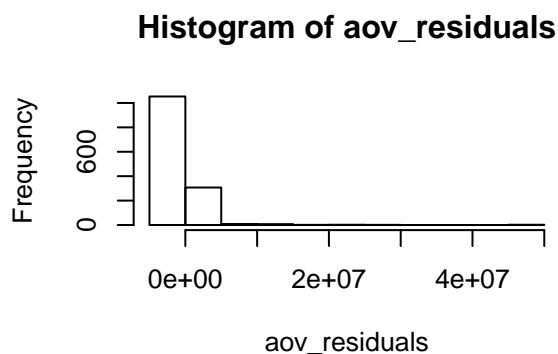
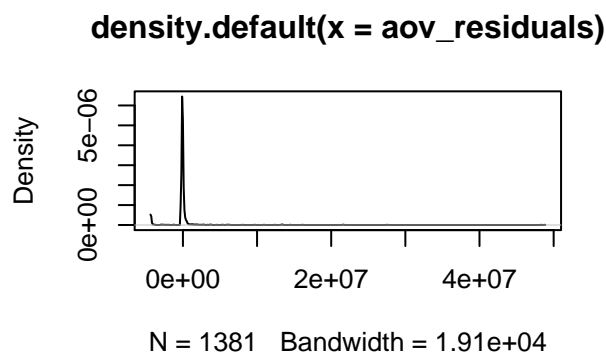
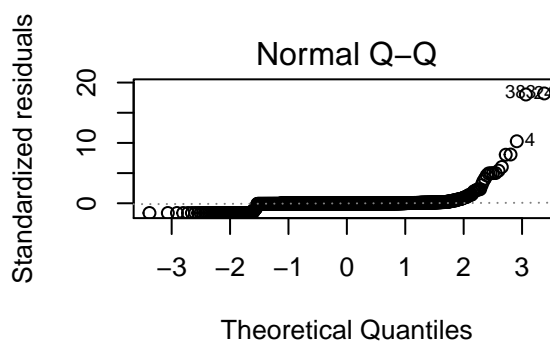
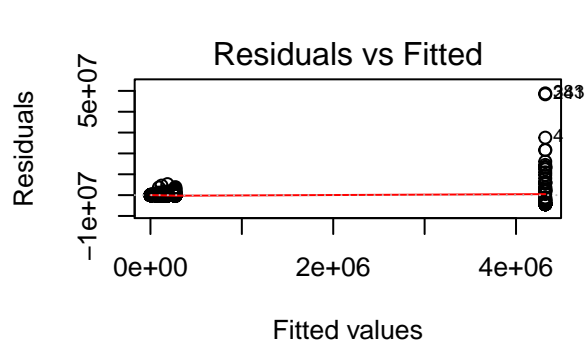
Difference in Classifications (Nonparametric)

Methods

Results

Results of Difference in Classification

```
classification.results <- aov(Grand.Total.Profits ~ classification_name, data = institution_data)
almost_sas(classification.results)
```



```
classification_kruskal.results <- kruskal.test(Grand.Total.Profits ~ classification_name, data = institution_data)
classification_post.hoc <- kruskalmc(Grand.Total.Profits ~ classification_name, data = institution_data)
```

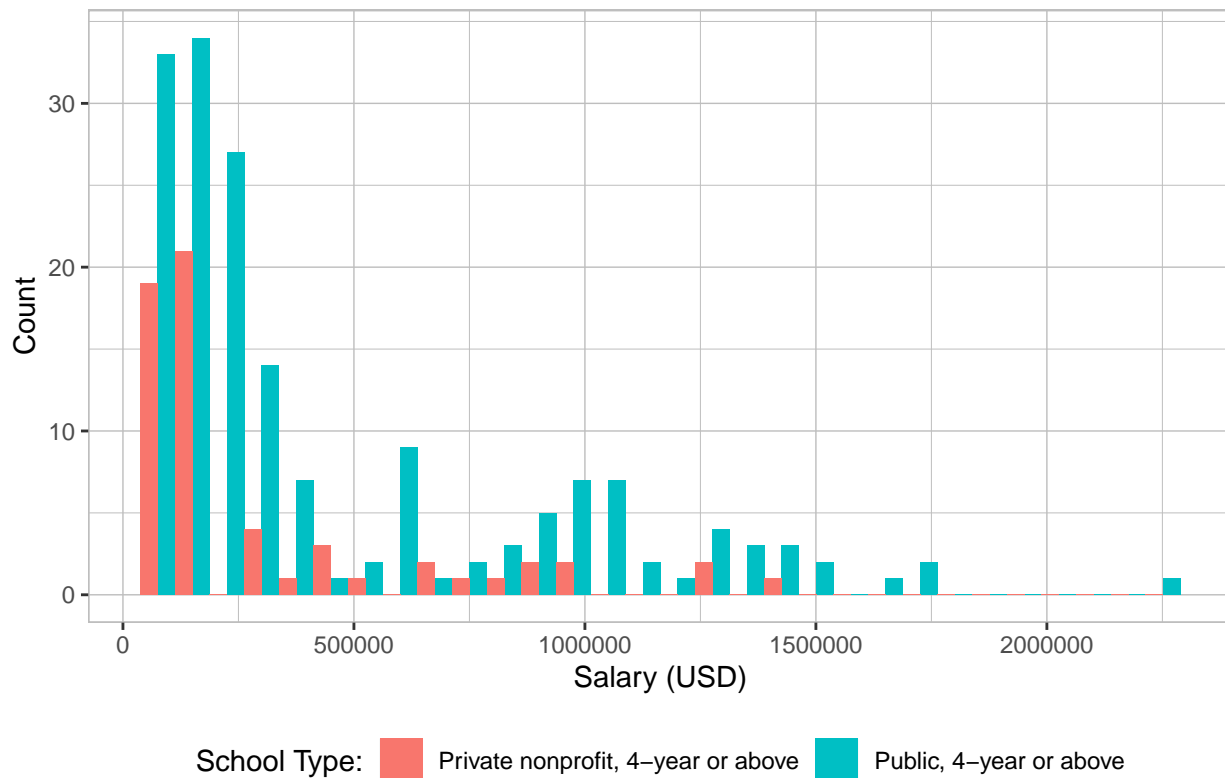
```
# pander(classification_post.hoc[[3]], style='rmarkdown') # TODO: shorten table!
```

Describing and Visualizing Head Coach data

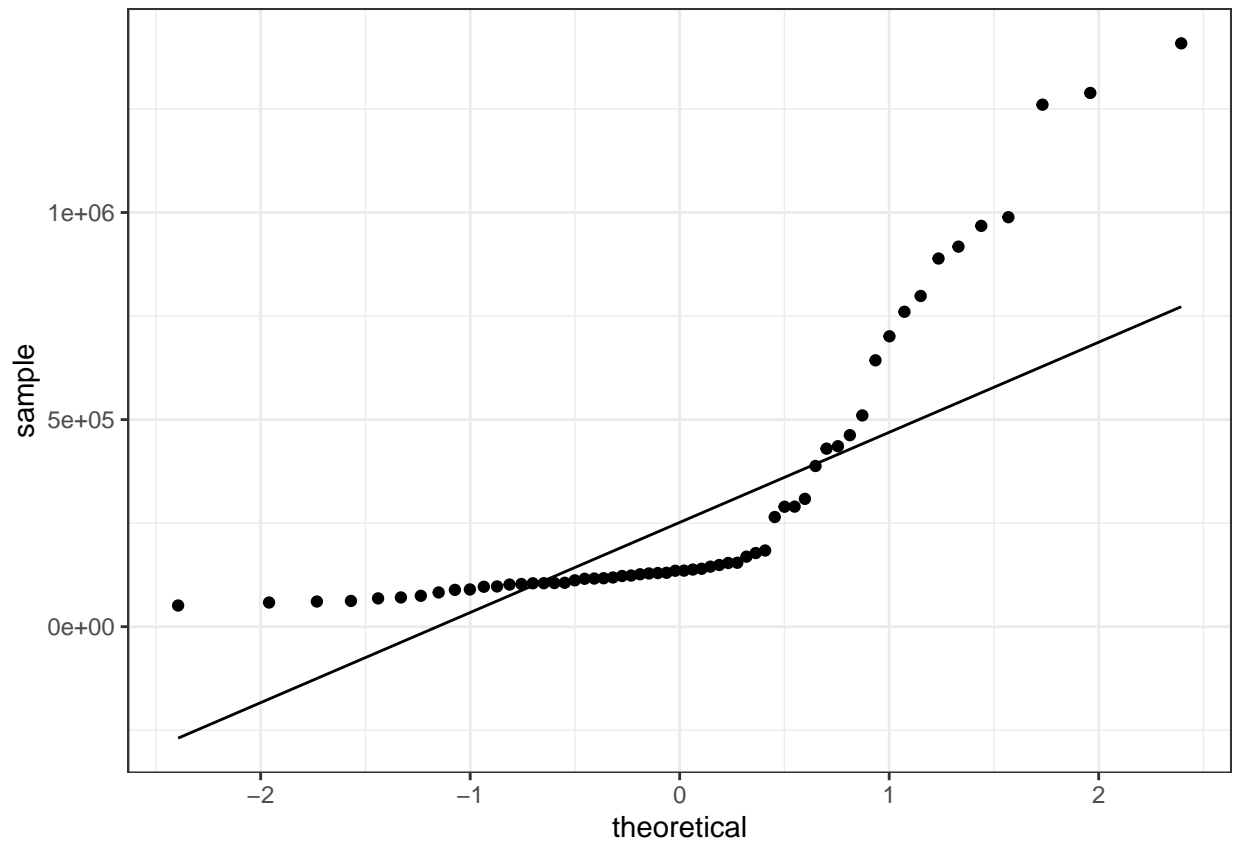
```
hd.coach.salary_data <- ncaa.div.i_data %>%
  select(sector_name, HDCOACH_SALARY_MEN)

hd.coach.salary_data %>%
  ggplot(aes(x = HDCOACH_SALARY_MEN, fill = sector_name)) +
  geom_histogram(position = "dodge") +
  labs(title="Histogram of NCAA Div. I Men's Head Coach Salary", x = "Salary (USD)",
       y = "Count", fill = "School Type:") +
  theme(
    legend.position="bottom",
    panel.background = element_rect(fill = 'white', color = 'grey'),
    panel.grid.major = element_line(size = 0.25, linetype = 'solid',
                                     color = "grey"),
    panel.grid.minor = element_line(size = 0.1, linetype = 'solid',
                                     color = "grey")
  )
)
```

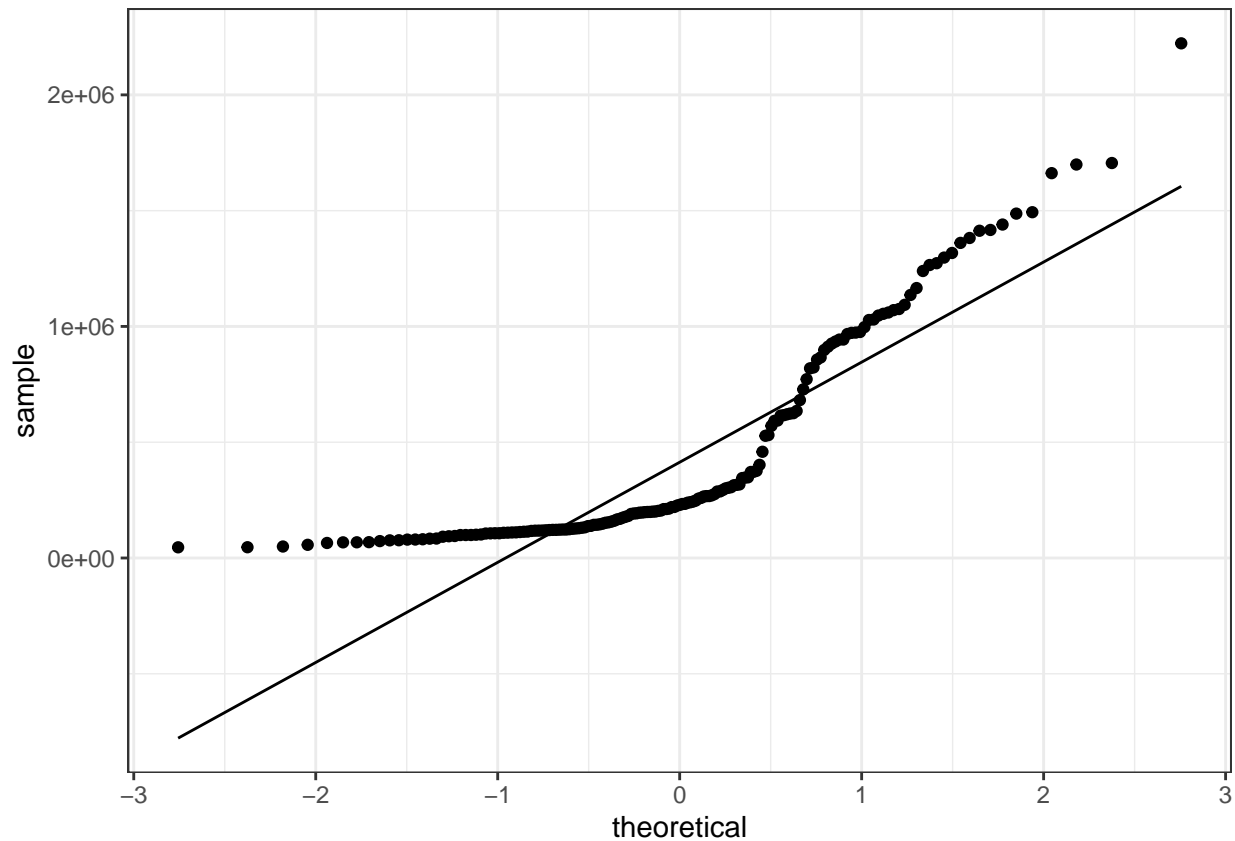
Histogram of NCAA Div. I Men's Head Coach Salary



```
hd.coach.salary_data %>%
  filter(grepl('Private', sector_name)) %>%
  ggplot(aes(sample = HDCOACH_SALARY_MEN)) +
  stat_qq() + stat_qq_line() + theme_bw()
```



```
hd.coach.salary_data %>%  
  filter(grepl('Public', sector_name)) %>%  
  ggplot(aes(sample = HDcoach_salary_men)) +  
  stat_qq() + stat_qq_line() + theme_bw()
```

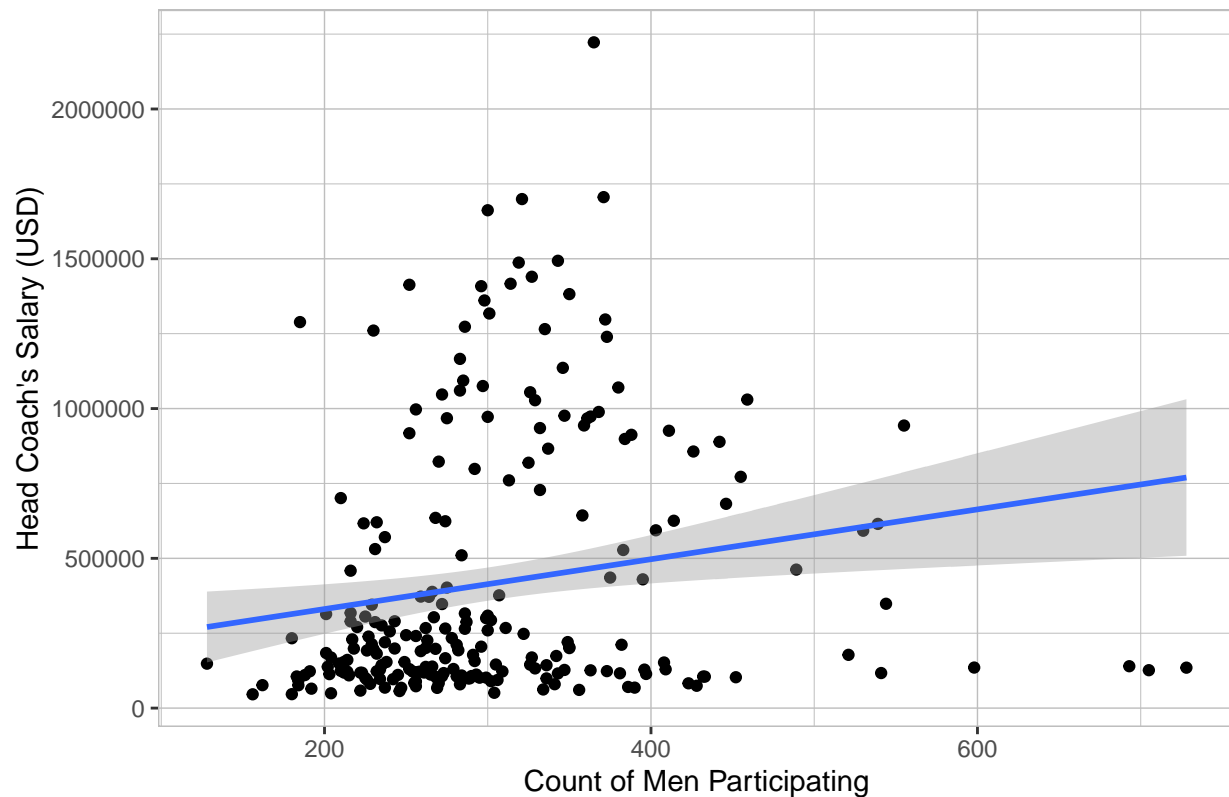


Describing and Visualizing Participation of Men and Head Coach Salaries

```
men.participation.data <- ncaa.div.i_data %>%
  select(IL_PARTIC_MEN, HDcoach_SALARY_MEN)

men.participation.data %>% ggplot(aes(x=IL_PARTIC_MEN, y=HDcoach_SALARY_MEN)) +
  geom_point() +
  geom_smooth(method = lm) +
  labs(title="Scatterplot of Men's Head Coach Salary by Men Participation Counts", x = "Count of Men Participation",
        y = "Head Coach's Salary (USD)") +
  theme(
    panel.background = element_rect(fill = 'white', color = 'grey'),
    panel.grid.major = element_line(size = 0.25, linetype = 'solid',
                                     color = "grey"),
    panel.grid.minor = element_line(size = 0.1, linetype = 'solid',
                                     color = "grey")
  )
```

Scatterplot of Men's Head Coach Salary by Men Participation Counts



Results of Head Coach data

Results of Participation of Men and Head Coach Salaries

Conclusion

Session Info

```
sessionInfo()
```

```
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
##  [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
```

```
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] haven_2.1.1      magrittr_1.5      kableExtra_1.1.0  knitr_1.24
## [5] forcats_0.4.0    stringr_1.4.0     dplyr_0.8.3       purrr_0.3.2
## [9] readr_1.3.1      tidyr_0.8.3       tibble_2.1.3      ggplot2_3.2.1
## [13] tidyverse_1.2.1  pander_0.6.3      pgirmess_1.6.9
##
## loaded via a namespace (and not attached):
## [1] httr_1.4.1        viridisLite_0.3.0  jsonlite_1.6
## [4] splines_3.6.1     modelr_0.1.5       gtools_3.8.1
## [7] assertthat_0.2.1  expm_0.999-4       sp_1.3-1
## [10] cellranger_1.1.0  yaml_2.2.0         LearnBayes_2.15.1
## [13] pillar_1.4.2      backports_1.1.4    lattice_0.20-38
## [16] glue_1.3.1        digest_0.6.20      rvest_0.3.4
## [19] colorspace_1.4-1  htmltools_0.4.0    Matrix_1.2-17
## [22] pkgconfig_2.0.2   broom_0.5.2        gmodels_2.18.1
## [25] webshot_0.5.1     scales_1.0.0       gdata_2.18.0
## [28] generics_0.0.2    withr_2.1.2        lazyeval_0.2.2
## [31] cli_1.1.0         crayon_1.3.4       readxl_1.3.1
## [34] deldir_0.1-23     maptools_0.9-8     evaluate_0.14
## [37] nlme_3.1-140      MASS_7.3-51.4      xml2_1.2.2
## [40] foreign_0.8-71    class_7.3-15       tools_3.6.1
## [43] hms_0.5.1         munsell_0.5.0      compiler_3.6.1
## [46] e1071_1.7-2       rlang_0.4.0        classInt_0.4-2
## [49] units_0.6-5       grid_3.6.1         rstudioapi_0.10
## [52] labeling_0.3      rmarkdown_1.15     boot_1.3-22
## [55] gtable_0.3.0      DBI_1.0.0          R6_2.4.0
## [58] splancs_2.01-40   lubridate_1.7.4    rgdal_1.4-6
## [61] rgeos_0.5-2       zeallot_0.1.0      spdep_1.1-3
## [64] KernSmooth_2.23-15 stringi_1.4.3       Rcpp_1.0.2
## [67] vctrs_0.2.0       sf_0.8-0           spData_0.3.2
## [70] tidyselect_0.2.5  xfun_0.9           coda_0.19-3
```

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/>.