Analysis of Gender Expense Differences in the UH Manoa Athletic Program

University of Idaho
STAT 431 EO Final Project
Zachary Higa

Introduction

In this study, I will be conducting an analysis of a model examining the expense difference between the men's and women's athletic teams at the University of Hawaii at Manoa athletic department.

- I selected the University of Hawaii at Manoa as the athletic program that I wanted to analyze because it is my alma mater (Go Bows!), and I have been a huge fan of UH athletics since I was a child. Also, I originally wanted to choose the University of Idaho, but the data was harder to come across.
- The reason for selecting this topic is stemmed from the budget disparity that was on display in the 2021 NCAA Basketball tournament. A few female athletes took to social media to show the disparity in the various amenities, food, etc., between the male athletes and female athletes. As a result, this drove many to examine the disparity and inequality between male and female athletics throughout the country.
- Therefore, I thought it would be interesting to analyze the expense differences on a more local scale at the University level since I'm a fan of collegiate sports.

General Question/Hypothesis of Interest

Hence, the research question and general hypothesis of this study are:

Research Question: Has the expense differences between the men's and women's athletic teams at UH Manoa decreased over the years?

Therefore, the **null hypothesis** that we wish to reject is that the gender expense differences increase or remain the same over time ($H_0: \beta_1 \geq 0$). As a result, our **alternative hypothesis** is that the gender expense differences decrease over time ($H_a: \beta_1 < 0$).

Goals/Significant of this Study

- From this study, we wish to examine how the expenses between the men's and women's athletics teams change over the years. If we notice that there is a decreasing pattern, then this could indicate that there has been improvements to make things better. However, if it appears to be increasing over the years, this could indicate an issue developing at the UH Manoa athletics department.
- Also, I found many articles and research consisting of gender inequality within sports on a nationwide level, however, I could not find many research on gender inequality within sports in the state of Hawaii. As UH athletics are a big deal for many folks on the islands (with no professional sports teams), I think this study will provide more information on a local level.

Design of Study

In this study, the **dependent variable** of interest is the expense differences between the men's and women's athletics programs (rounded to the nearest dollar). Our **independent variable** is the years with recorded data.

Thus, our regression model for this study is:

$$\boldsymbol{y_i = \beta_0 + \beta_1 x_i + \epsilon_i}$$

where y_i = the expense difference for the ith observed year, x_i = the ith observed year, β_1 = the slope of the model (how the expense difference changes per year), β_0 = the y-intercept of model, ϵ_i = error of expense difference for the ith observed year.

Problems/Factors Involving Dependent Variable

As defined in the slide before, our dependent variable is the gender expense differences. For this study, we are just analyzing how this changes over time for this model, though there are other factors that could influence our dependent variable measured:

- Revenue/Profit - It is possible that income or profit from athletic teams from previous years may influence the expenses for teams in a following year, and the male and female teams generate different revenues.
- Number of Male vs. Female Athletes - The number of male/female athletes differ, which may have an influence on the gender expense differences (though from the dataset, I have seen that there are usually more female athletes than male athletes per year).

Though these factors may influence our dependent variable, I just want to focus on how the expenses change over time (and for simplicity's sake), but for future studies, the model can be adjusted to take these into account as well.

Sampling Procedure

In this study, our 'population' is the years of the UH athletic department where the expenses for the men's and women's sports are recorded. Thus, our 'subjects' selected from this are the years 2003-2019 because these were the years that had available data that I could find for the expenses.

- To get my sample, I downloaded the .csv file from the Equity in Athletics database for the University of Hawaii at Manoa.
- Then I edited the .csv file to include only the information on the observed year, institution, men's expenses, and women's expenses.
- From there, I calculated the difference between the men's expenses and women's expenses for each year, and I created a new column with these differences.
- After this, I read the data into R with the read.csv function (see code in Appendix), and I now have a data table (see Table 1 in Appendix) with the data I need to conduct my analysis.

Analysis of the Data

R Code Used:

```
expense_difference <- lm(hawaii_athletics$Total.Difference.in.Expenses~hawaii_athletics$Survey.Year)
summary(expense_difference)
```

Output:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-433380909	104560226	-4.145	0.000864 ***
Survey.Year	218962	51994	4.211	0.000755 ***

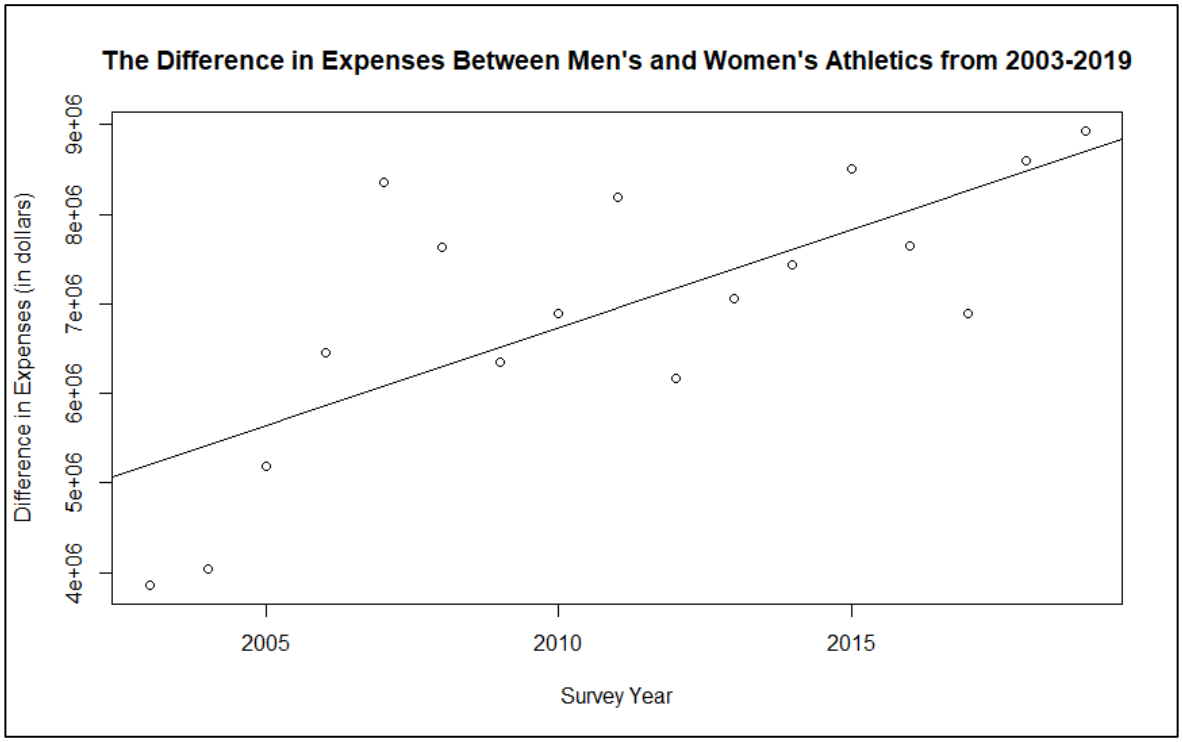
Therefore, from the values in the output above, the regression model for this study is:

$$\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i$$

$$\rightarrow y = -433380909 + 218962x$$

Analysis of the Data - Graph

Since we found our model, $y = -433380909 + 218962x$, we can plot the line along with our observations such that:



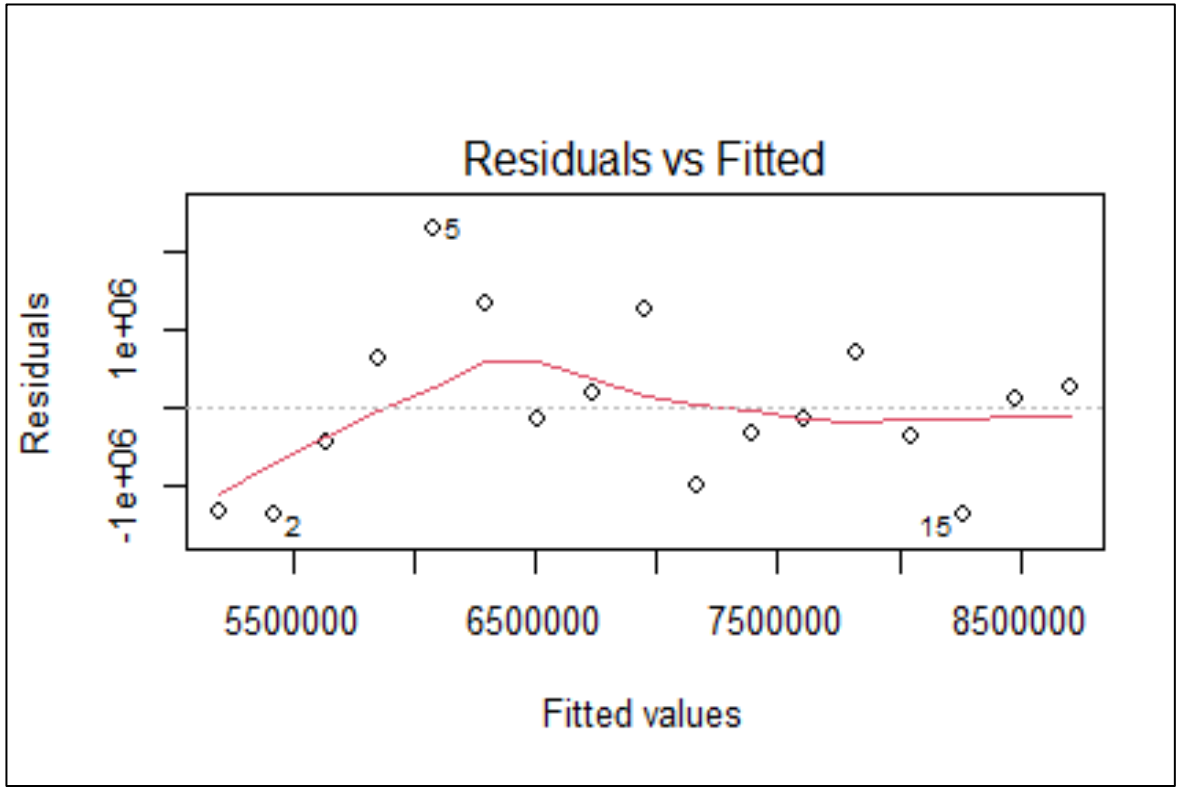
From this, we see that $\beta_1 > 0$, as the slope of the line from the model is positive, and we can check this using our test statistic as well.

Analysis of the Data - Test Statistic

- From the summary function output on the previous slide, we found that our t-statistic is $t = 4.211$.
- So, we reject our null hypothesis, $H_0: \beta_1 \geq 0$, if $t < -t_{\alpha,n-2}$.
- Let $\alpha = 0.05$, and from the data, we have $n = 17$ observations, so our tabled t-value is $-t_{0.05,\,15} = -1.753$.
- Therefore, since we see that $t = 4.211$ is not less than $-t_{0.05,15} = -1.753$, **we cannot reject our null hypothesis that $H_0: \beta_1 \geq 0$.**
- As a result, we see that we cannot conclude that the gender expense differences decrease over time, and from the positive slope seen on the graph, it may appear that the gender expense difference is increasing over time.

Analysis of the Data - Assumptions

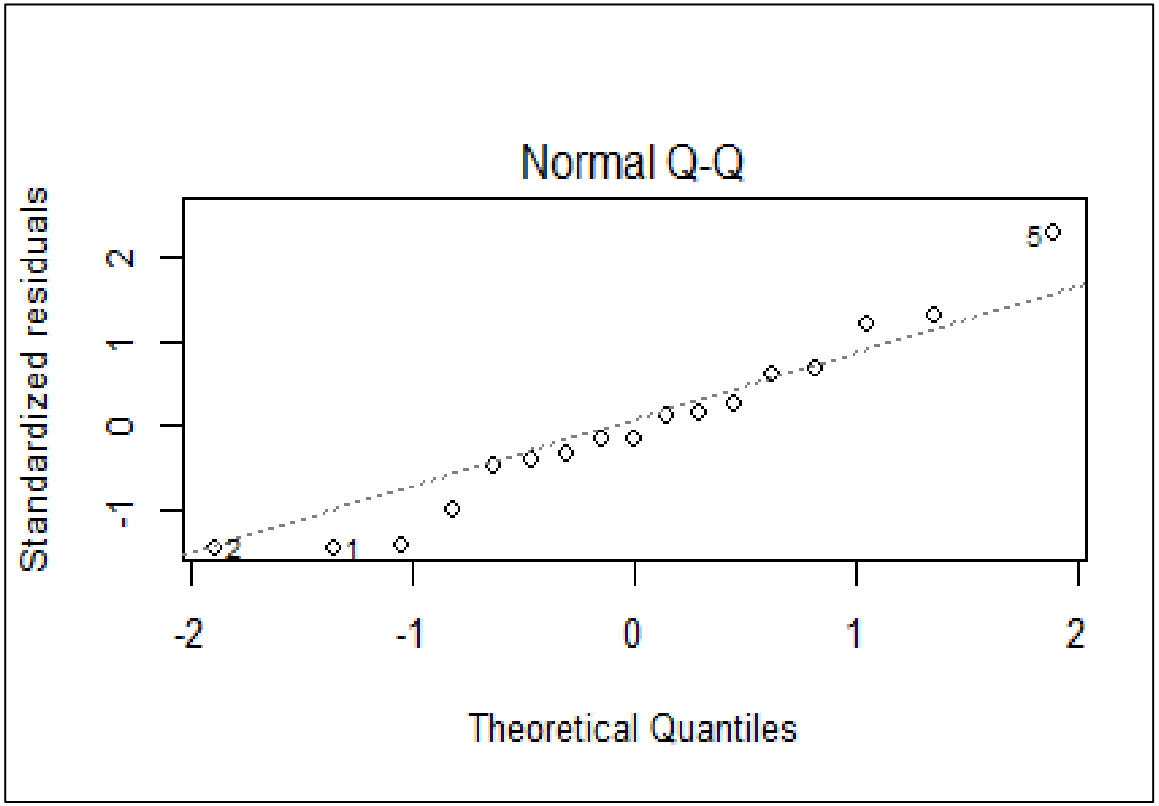
From R, I also plotted the residual by predicted value plot to check for the assumption of homogeneous variance:



From the plot, it does not appear that the variance is changing along the predicted values. However, the way the residuals are fluctuating up and down could indicate a lack of fit, so we may need to use a transformation or higher-order model to get a better fit.

Analysis of the Data - Assumptions

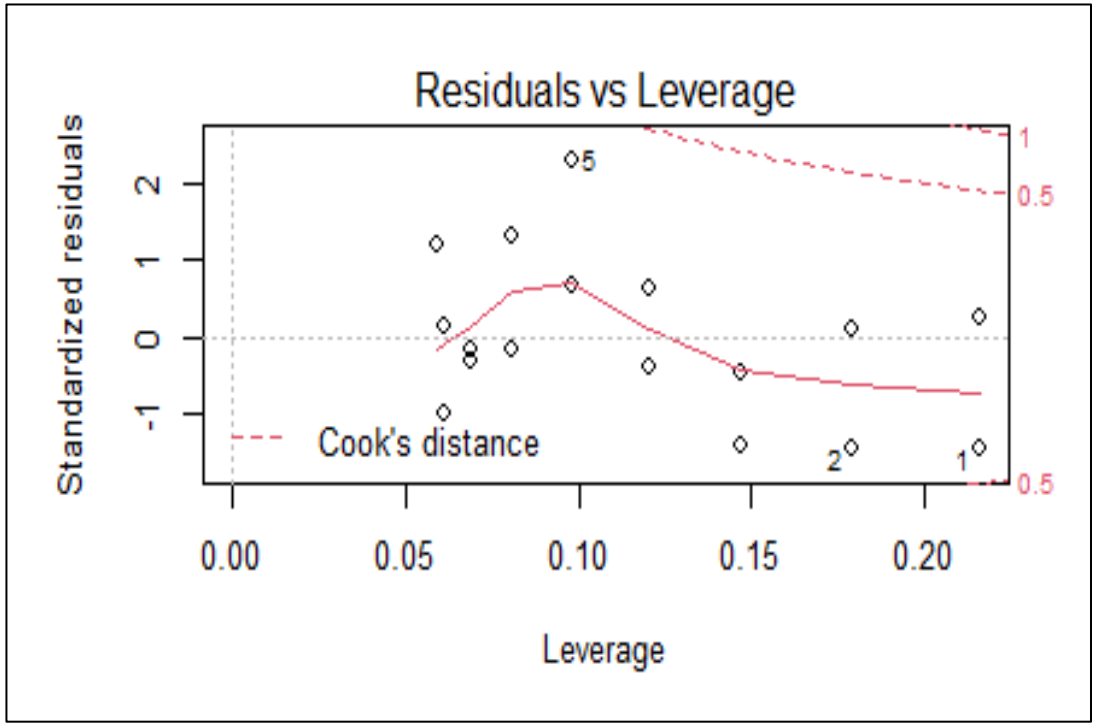
From R, I also plotted the Normal Q-Q plot to check for the assumption of normality:



From the plot, it appears that most points fit along the line, so it would seem that the normal assumption follows. However, there are a few points that do not follow along the line, so this Normal Q-Q plot can be better. If some points were removed due to being outliers, that could potentially help as well.

Analysis of the Data - Leverage

From R, we can also look at this plot of the standardized residuals versus the leverage:

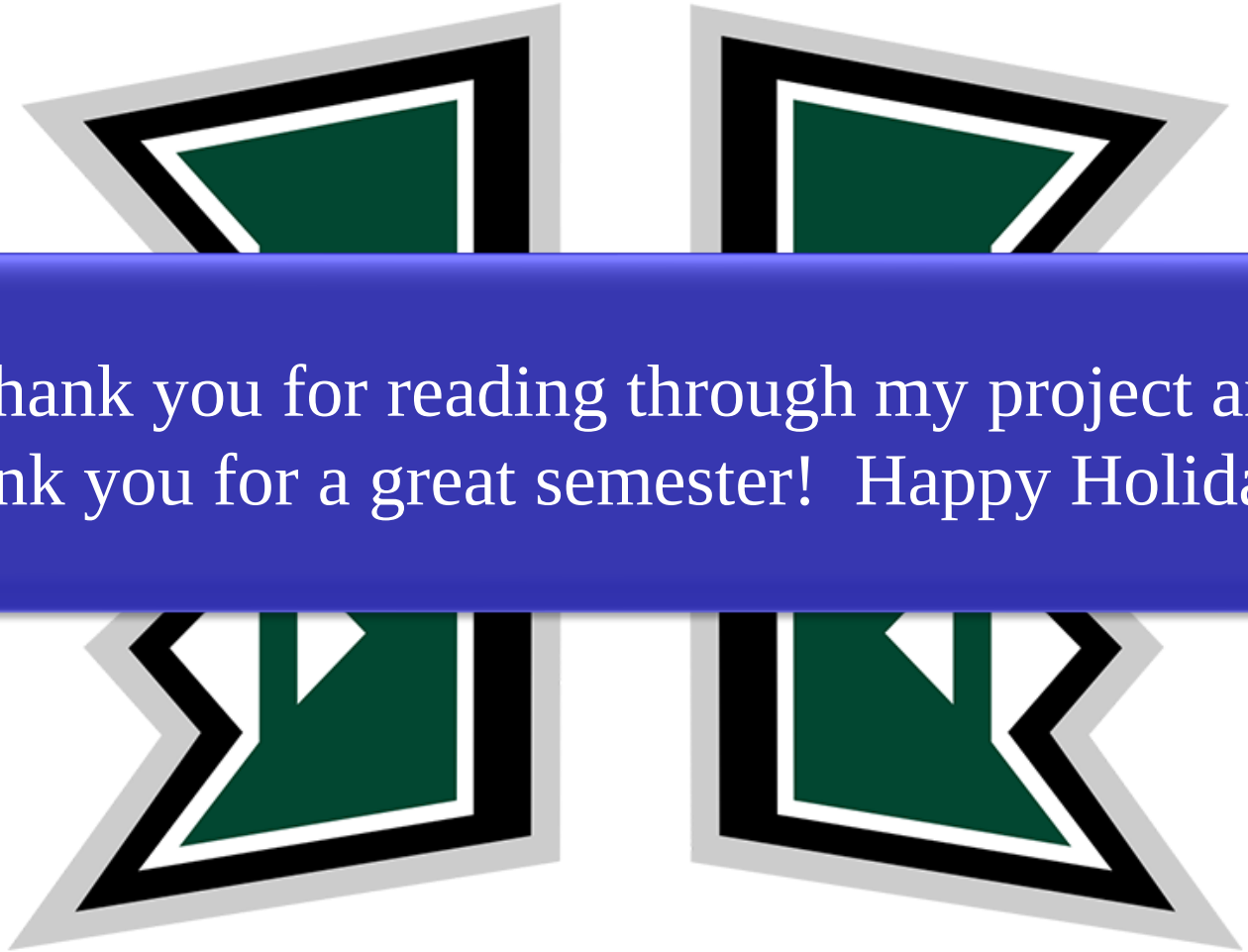


From this plot, we see that points 1 and 2 have high leverage (along with a couple other points), so these points may be high influence points, but they may also be outliers. Also, we see that point 5 (though a middle leverage) has a high standardized residual value, so that can be looked at as an outlier as well.

Conclusion

- Therefore, from our study, we were **not able to reject our null hypothesis**, so the expense differences between the men's and women's athletic teams at UH Manoa **would increase over time**.
- As our study showed an increase in expense differences, this is consistent with the recent news that argue about the presence of gender inequity evident in tournaments and athletic programs nationwide. Unfortunately, this means that even on a local level, this issue of gender inequality in athletics can still be prevalent.
- However, despite the implications of this particular study, steps can be taken to improve our results such as:
  - Involving the additional factors that could potentially affect the expense differences by introducing them in a model or adjusting the model if a higher-order/transformation is needed.
  - Including additional data points as more data is available and excluding any data points that may appear to be outliers.
- Though gender inequality is still an issue in athletics, hopefully, as more studies and people speak up about this issue, the more that will continue to be done to combat it.

Thank You!



Thank you for reading through my project and thank you for a great semester! Happy Holidays!