Analyzing the Revenues Generated by Big Sky Schools

Zachary Higa

Motivation

- Recently, it was announced that UCLA and USC were moving to the Big Ten Conference, causing a shift and discussions among other athletic conferences as well.
- The move might have been spurred due to the allure of more revenue generation for their athletic departments and sports programs.
- As all these discussions were interesting, as a University of Idaho student, I got interested in doing an analysis on the revenues generated by different schools within the athletic conference we participate in – the Big Sky Conference.
- Also, a recent article discussed that the University of Montana will be staying in the Big Sky Conference despite talks of moving up to the FBS division of football due to it being one of the highest revenue-generating schools in the Big Sky conference (Semb 2022).
- Thus for this project, I wanted to answer the research question of…

Research Question/Hypothesis

- Is there a significant difference in the revenue generated by various Big Sky conference schools for the year of 2018?

Hypotheses Tests of Major Interest

The null hypothesis of interest is (notation explained more later):

$$H_0\colon\ \alpha_1 = \cdots = \alpha_{10}$$

versus the alternative hypothesis of:

H_1: α_i is different between at least one of the groups

- In other words, the null hypothesis is that the treatment effects from each school on the revenue generated is the same, so there is no significant difference in revenue generated between schools.
- The alternative hypothesis is that there is at least one significant difference in revenue generated by the Big Sky schools, so there is at least one treatment effect that is different from another

Goals of the Project

- To see if there is a significant difference in revenues generated by schools in the Big Sky conference.
- If there is a significant difference in revenues, then try to see if we can find the specific schools in the Big Sky conference that have the significant differences in revenue.

Design of the Project – RCB Design

- For this project, we are using a **randomized complete block design (RCB)**.
- Our design is looking at analyzing if there is a difference between the mean revenue generated between the 10 full-time members of the Big Sky Conference (schools) in the year of 2018. Thus, the schools act as the treatment effect in this project.
- Blocking will be done on the sports that all the Big Sky schools sponsored during the 2018 season.

RCB Design – Why Blocking?

- Specifically, we are blocking on sports that all the Big Sky schools sponsored in the 2018 athletic season.
  - These sports are: Basketball, Track, Football, Women's Golf, Tennis, and Volleyball.
  - Blocking is used to control for the variation in the revenues generated by different sports. Also, we chose to block on these specific sports to control for the fact that some of these colleges sponsor sports that others do not sponsor.

RCB Design – Treatment Factors

- For this RCB design, each Big Sky school is acting as a treatment factor
- These treatments are fixed since we are purposefully selecting the 10 schools that are full-time members in the Big Sky athletic conference. There are other affiliate members of the Big Sky, but they only sponsor a few sports in the conference
- These 10 colleges are used because we want to examine if any school generates significantly more or less revenue than any of the other schools in the conference

RCB Design – Dependent Variable

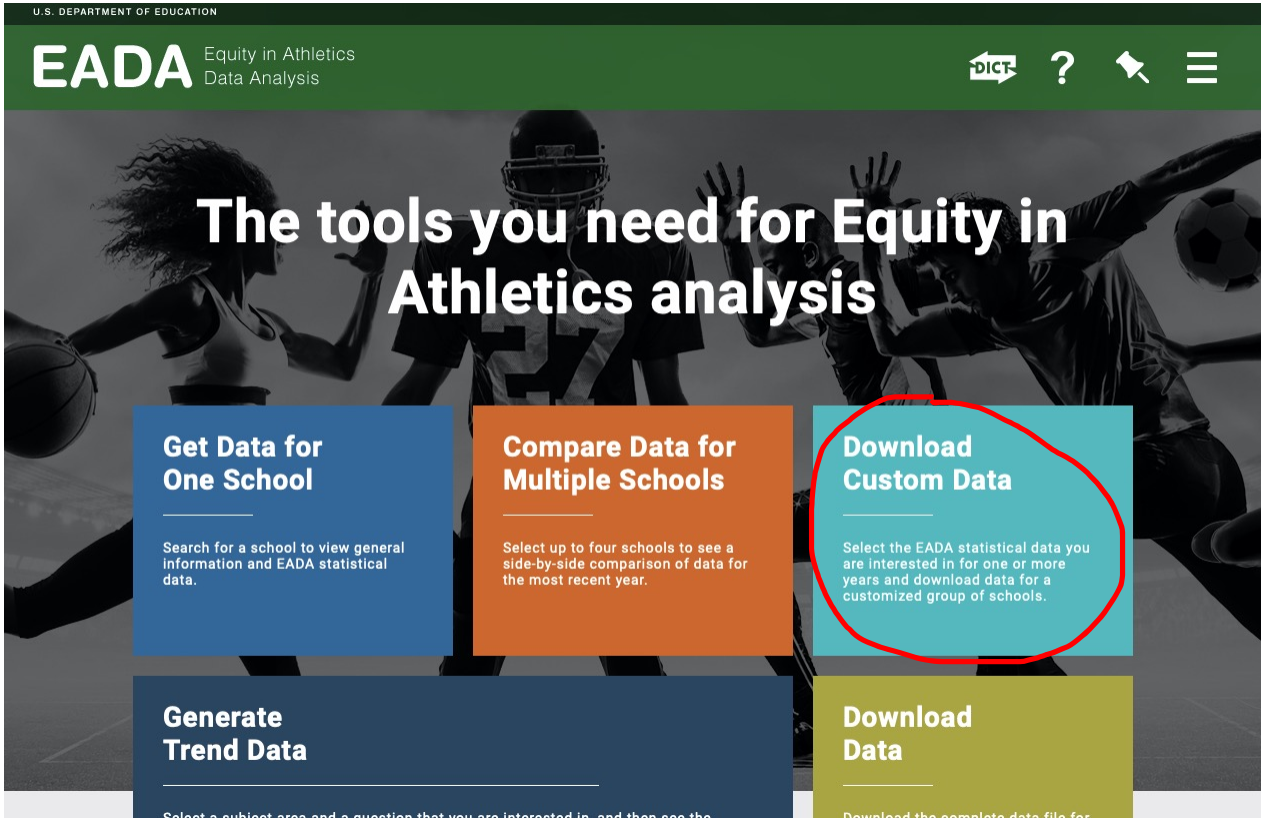
- The dependent variable that we are measuring in this project is the revenue generated by each school for the sports that we are blocking by.
- There is not much difficulty in measuring the dependent variable as the revenue generated by each school by a sport is clear cut. However, there is always a potential for clerical errors or numbers that are reported to be different than they actually are (though it is unlikely, or it would have most likely caused some scandal)

RCB Design – Why 2018?

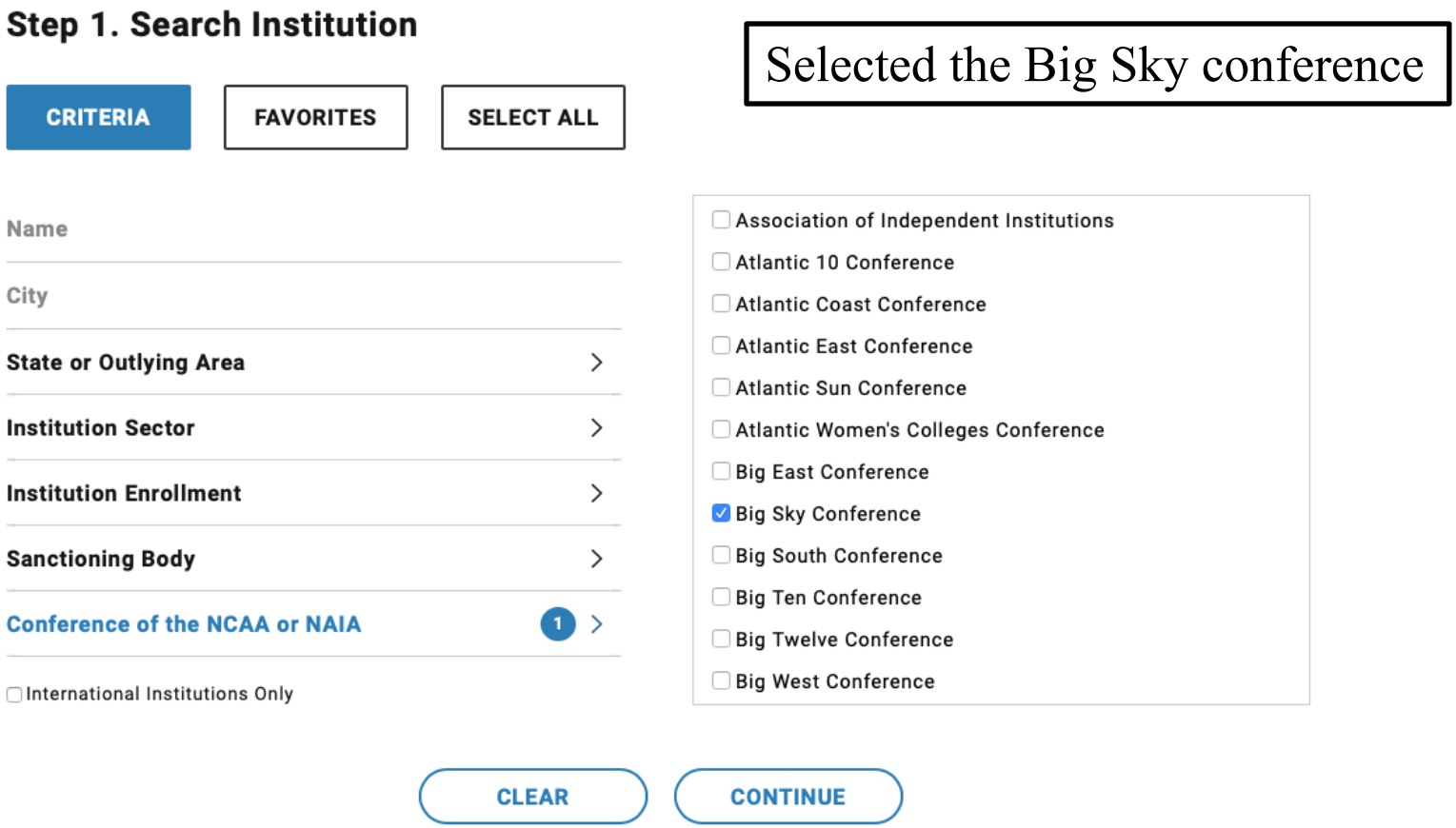
- The reason why I chose to analyze the revenues in the year of 2018 is because this was before Covid occurred.
- In the years after 2018, we can see that some schools did not sponsor all the sports that they may have sponsored in the years before, so there would only be two sports that every Big Sky school competed in.
- Thus, since I am blocking on the sports that all the Big Sky schools sponsor, the year 2018 was the most recent year where more than two common sports were sponsored by all the Big Sky schools.

Data Source/Methodology

https://ope.ed.gov/athletics/#/



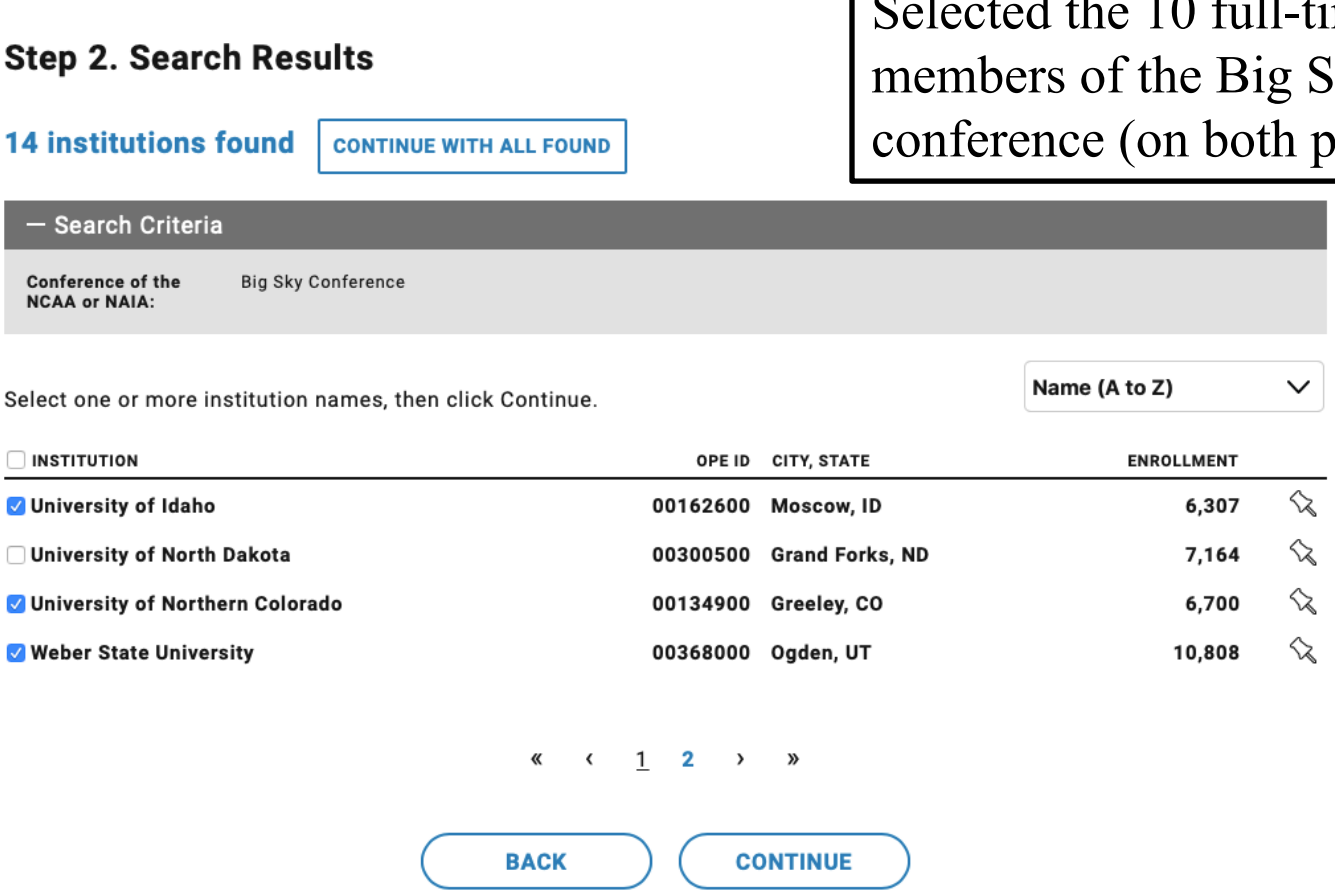
Data Source/Methodology



Selected the Big Sky conference

Data Source/Methodology

Selected the 10 full-time members of the Big Sky conference (on both pg. 1 and 2)



Step 2. Search Results

14 institutions found CONTINUE WITH ALL FOUND

— Search Criteria

Conference of the NCAA or NAIA: Big Sky Conference

Select one or more institution names, then click Continue.

Name (A to Z)

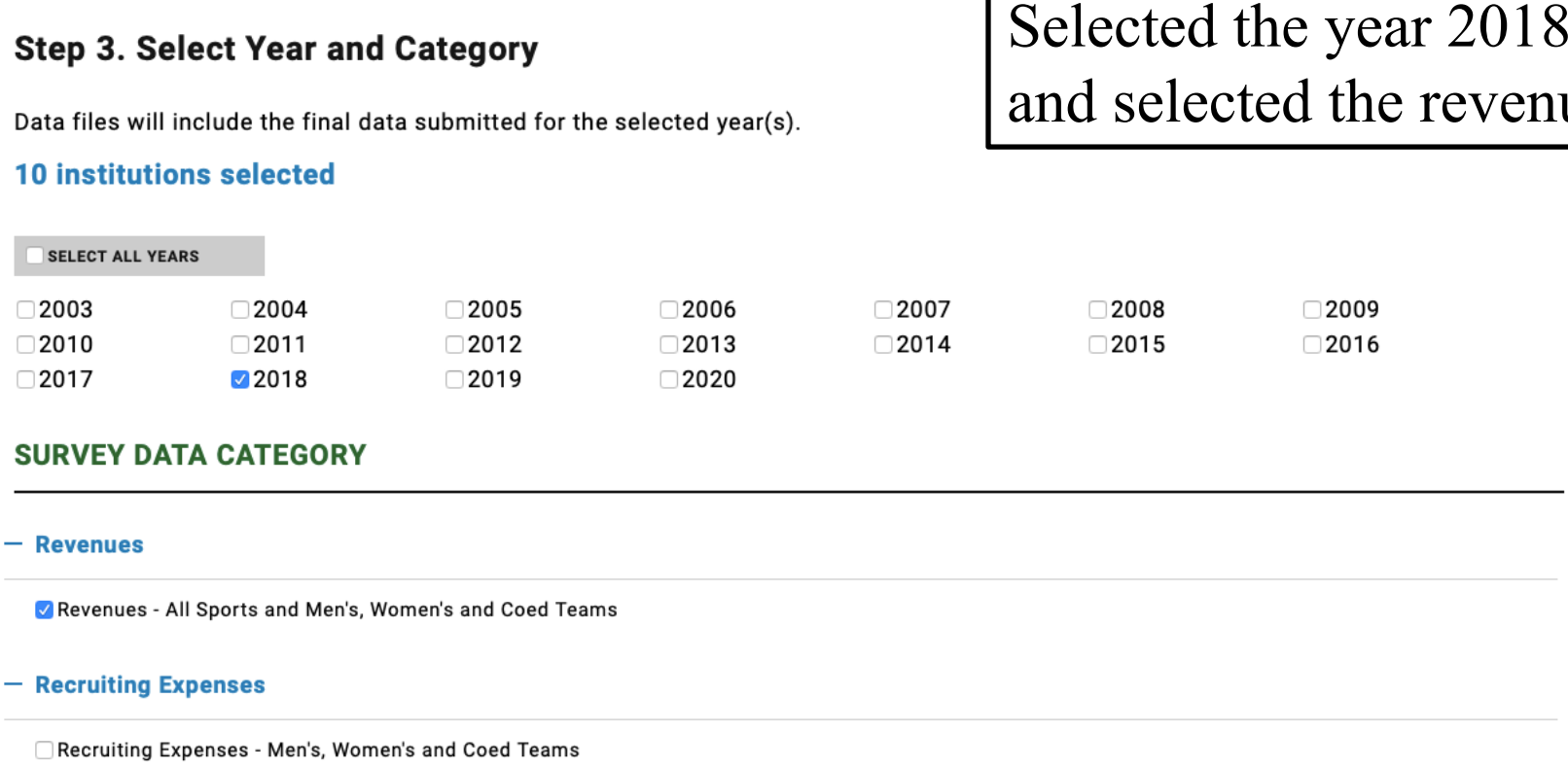
INSTITUTION	OPE ID	CITY, STATE	ENROLLMENT
University of Idaho	00162600	Moscow, ID	6,307
University of North Dakota	00300500	Grand Forks, ND	7,164
University of Northern Colorado	00134900	Greeley, CO	6,700
Weber State University	00368000	Ogden, UT	10,808

« ‹ 1 2 › »

BACK CONTINUE

Data Source/Methodology

Selected the year 2018 and selected the revenue



Methodology

- Once the data set is downloaded, I cleaned the data set to contain the revenues generated by the sports that I was blocking on for each school.
- Then once the data has been cleaned up, I loaded the data set into R, so I could get to analyzing the data set.

Data

1	California State University-Sacramento	Basketball	2922440
2	Eastern Washington University	Basketball	3072193
3	Idaho State University	Basketball	2687947
4	Montana State University	Basketball	3133799
5	Northern Arizona University	Basketball	3252220
6	Portland State University	Basketball	3167684
7	The University of Montana	Basketball	3387584
8	University of Idaho	Basketball	3143063
9	University of Northern Colorado	Basketball	2790186
10	Weber State University	Basketball	2889290
11	California State University-Sacramento	Track	1405229
12	Eastern Washington University	Track	1441702
13	Idaho State University	Track	1122030
14	Montana State University	Track	1861983
15	Northern Arizona University	Track	2425029
16	Portland State University	Track	511314
17	The University of Montana	Track	1020364
18	University of Idaho	Track	1460530
19	University of Northern Colorado	Track	959427
20	Weber State University	Track	1426105
21	California State University-Sacramento	Football	5515779
22	Eastern Washington University	Football	6739277
23	Idaho State University	Football	3993358
24	Montana State University	Football	8709180
25	Northern Arizona University	Football	5002914
26	Portland State University	Football	4563174
27	The University of Montana	Football	8620428
28	University of Idaho	Football	5515778
29	University of Northern Colorado	Football	3882353
30	Weber State University	Football	4092794

	School	Sport	Revenue
31	California State University-Sacramento	Women's Golf	453544
32	Eastern Washington University	Women's Golf	296047
33	Idaho State University	Women's Golf	137204
34	Montana State University	Women's Golf	334753
35	Northern Arizona University	Women's Golf	389736
36	Portland State University	Women's Golf	476587
37	The University of Montana	Women's Golf	350275
38	University of Idaho	Women's Golf	439740
39	University of Northern Colorado	Women's Golf	273803
40	Weber State University	Women's Golf	297636
41	California State University-Sacramento	Tennis	730061
42	Eastern Washington University	Tennis	957361
43	Idaho State University	Tennis	547735
44	Montana State University	Tennis	826869
45	Northern Arizona University	Tennis	896472
46	Portland State University	Tennis	861057
47	The University of Montana	Tennis	702645
48	University of Idaho	Tennis	674926
49	University of Northern Colorado	Tennis	415057
50	Weber State University	Tennis	633700
51	California State University-Sacramento	Volleyball	857935
52	Eastern Washington University	Volleyball	764369
53	Idaho State University	Volleyball	675254
54	Montana State University	Volleyball	912567
55	Northern Arizona University	Volleyball	1019390
56	Portland State University	Volleyball	935322
57	The University of Montana	Volleyball	849895
58	University of Idaho	Volleyball	898131
59	University of Northern Colorado	Volleyball	894436
60	Weber State University	Volleyball	685564

Data Analysis – Model

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

where y_{ij} is the revenue generated by the ith school and the jth sport, μ is the grand mean revenue, α_i is the treatment effect of the ith school, β_j is the block effect of the jth sport, and ϵ_{ij} is the residual of the observation of the ith school and the jth sport.

Recall the hypotheses stated earlier…

Data Analysis – Recall Hypotheses

The null hypothesis of interest is:

$$H_0\colon\ \alpha_1 = \cdots = \alpha_{10}$$

versus the alternative hypothesis of:

$H_1\colon\ \alpha_i$ is different between at least one of the groups

Thus, we see that we are testing whether the treatment effects of the 10 schools on the revenue generated are equal as we now know that α_i is the treatment effect of the ith school

Data Analysis – ANOVA table

Generated from:
big_sky_revenue.lm <- lm(Revenue ~ Sport + School, data = big_sky_revenue)
anova(big_sky_revenue.lm)

Source	df	Sum Sq	Mean Sq	F value	pvalue
Sport	5	2.0608e+14	4.1216e+13	74.9742	< 2e-16
School	9	7.8993e+12	8.7771e+11	1.5966	0.1453
Error	45	2.4738e+13	5.4973e+11		
Total	59	2.3872e+14			

Data Analysis

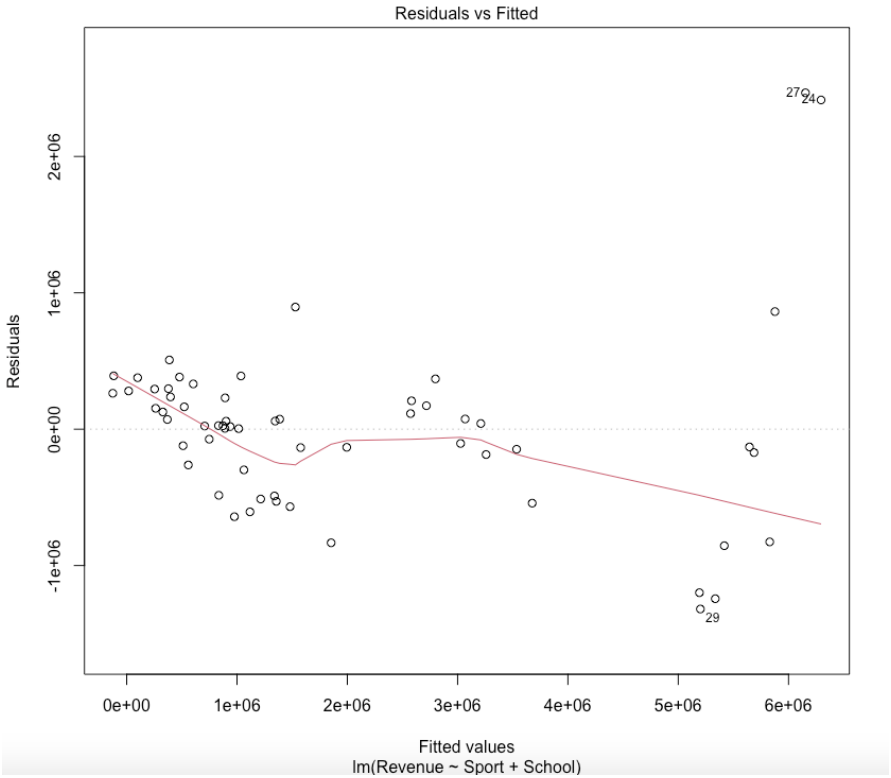
From the ANOVA table seen in the previous slide, we see that the pvalue for the treatment effect of the Big Sky schools is 0.1453, which is not less than 0.05. Therefore, we **cannot** reject the null hypothesis that the treatment effect of each school is the same, so there is **no significant differences** in the revenue generated between Big Sky Schools

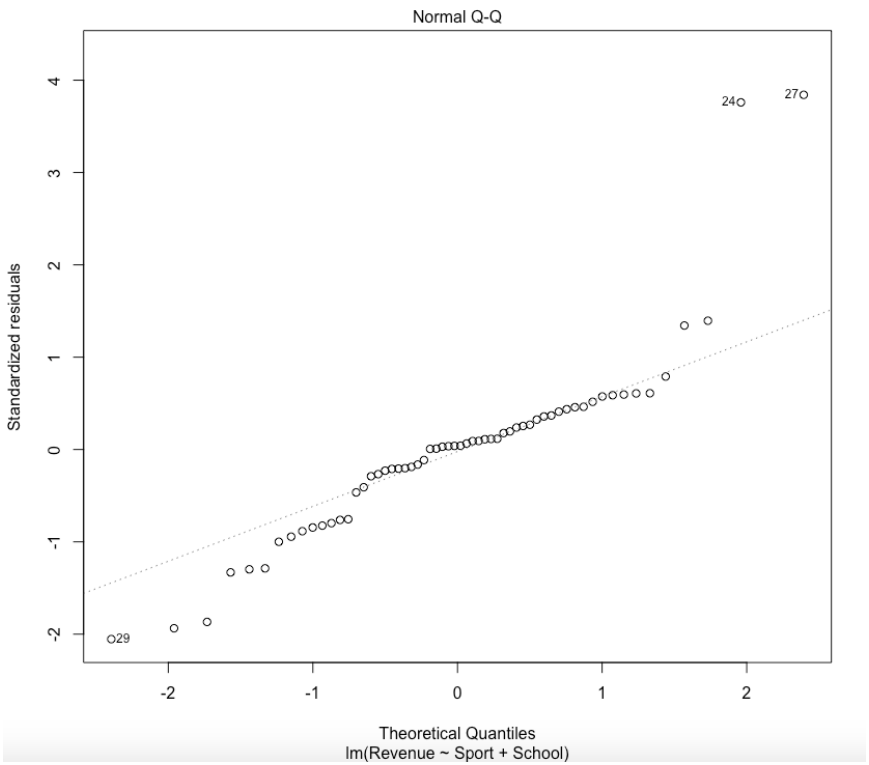
Data Analysis (cont.)

Additionally, we see that from the ANOVA table, that the blocking effect of Sports has a pvalue of < 2e-16, which is less than 0.05. Therefore, we see that the blocking effect of Sports is significant, indicating that the blocking by Sports is working as desired.

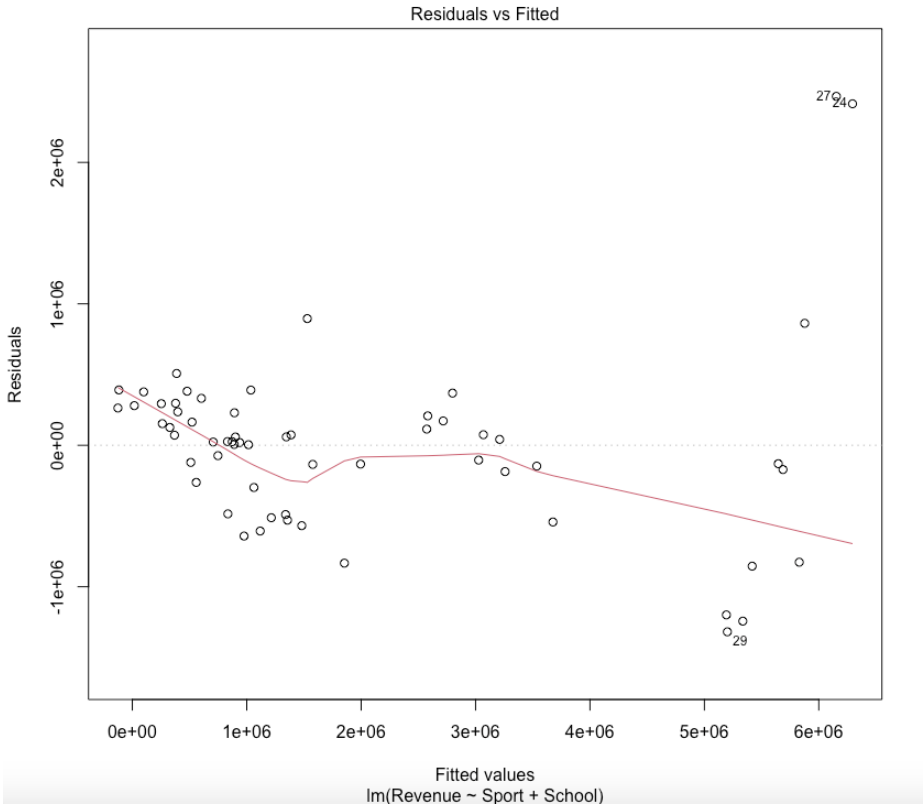
Data Analysis – Checking Assumptions

From: plot(big_sky_revenue.lm)



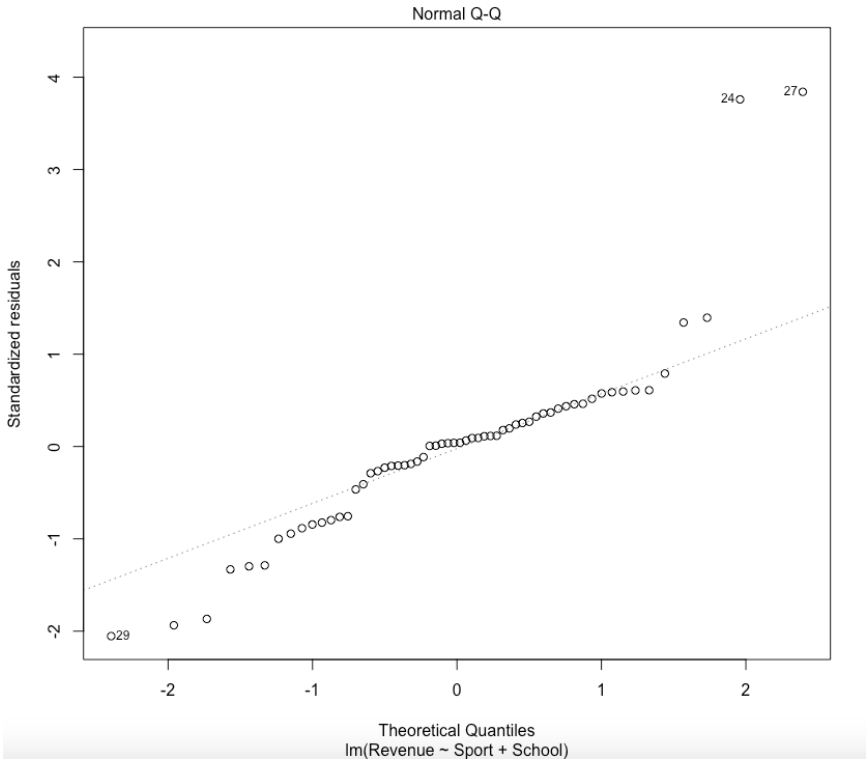


Data Analysis – Checking Assumptions



We see from the residual by predicted plot that there is a megaphone shape, as it appears as though the variance increases as the fitted values increase. Thus, the assumption of constant variance of residuals does not hold, and a transformation may need to be done.

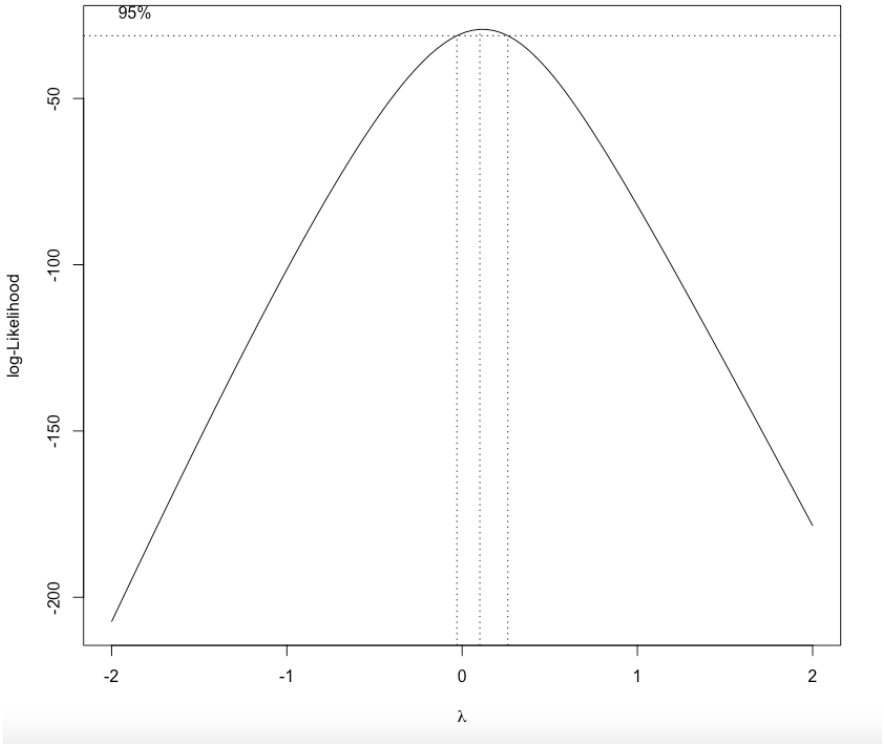
Data Analysis – Checking Assumptions



We see from the normal QQ plot, that towards the ends, data points seem to taper off of the line. Thus, it appears that the assumption of normality of residuals does not hold, and a transformation may need to be done.

Data Analysis – Box-Cox Transformation

```
boxcox(Revenue ~ Sport + School, data = big_sky_revenue, lambda = seq(-2.00, 2.00, length = 50))
```



Using the Box-Cox Transformation method, we see that we should use a lambda equal to 0, so we should take a log transformation of the revenue of the data set.

Data Analysis 2.0 – After Transformation

```
big_sky_revenue$logRevenue <- log10(big_sky_revenue$Revenue)
big_sky_revenue_log.lm <- lm(logRevenue ~ Sport + School, data = big_sky_revenue)
anova(big_sky_revenue_log.lm)
```

Source	df	Sum Sq	Mean Sq	F value	pvalue
Sport	5	9.8436	1.96873	170.4731	< 2.2e-16
School	9	0.3034	0.03371	2.9187	0.008243
Error	45	0.5197	0.01155		
Total	59	10.6667			

Data Analysis 2.0 – After Transformation

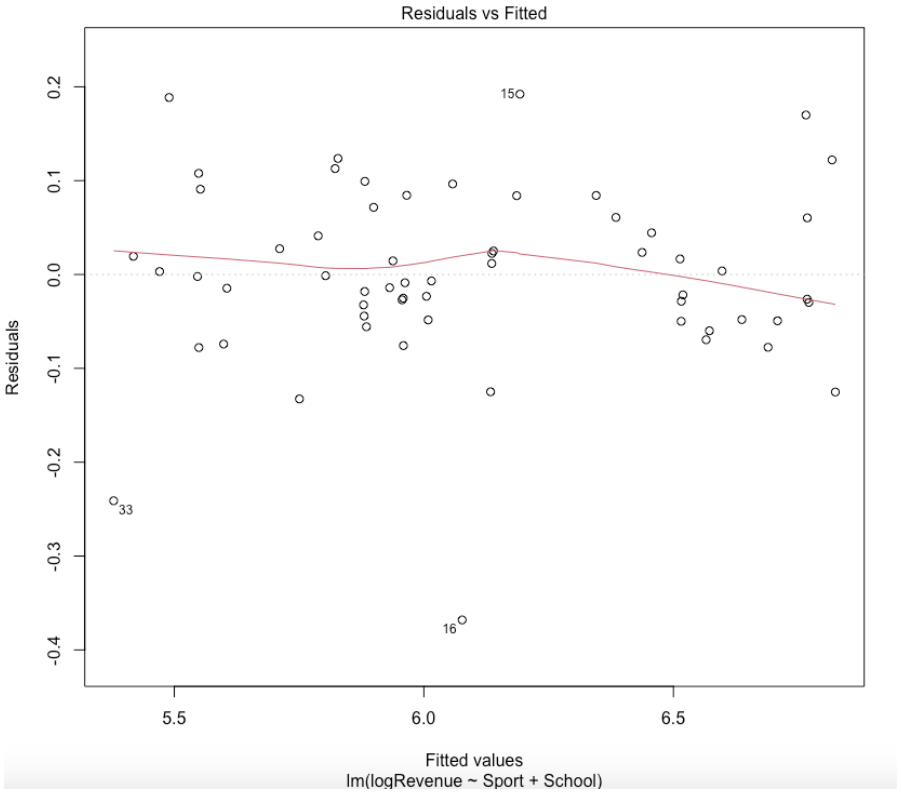
After doing the log transformation on the revenue for the Big Sky data set, from the ANOVA table seen in the previous slide, we see that the pvalue for the treatment effect of the Big Sky schools is 0.008243, which is now less than 0.05. Therefore, we **reject the null hypothesis** that the treatment effect of each school is the same, so there **is a significant differences** in the revenue generated between Big Sky Schools

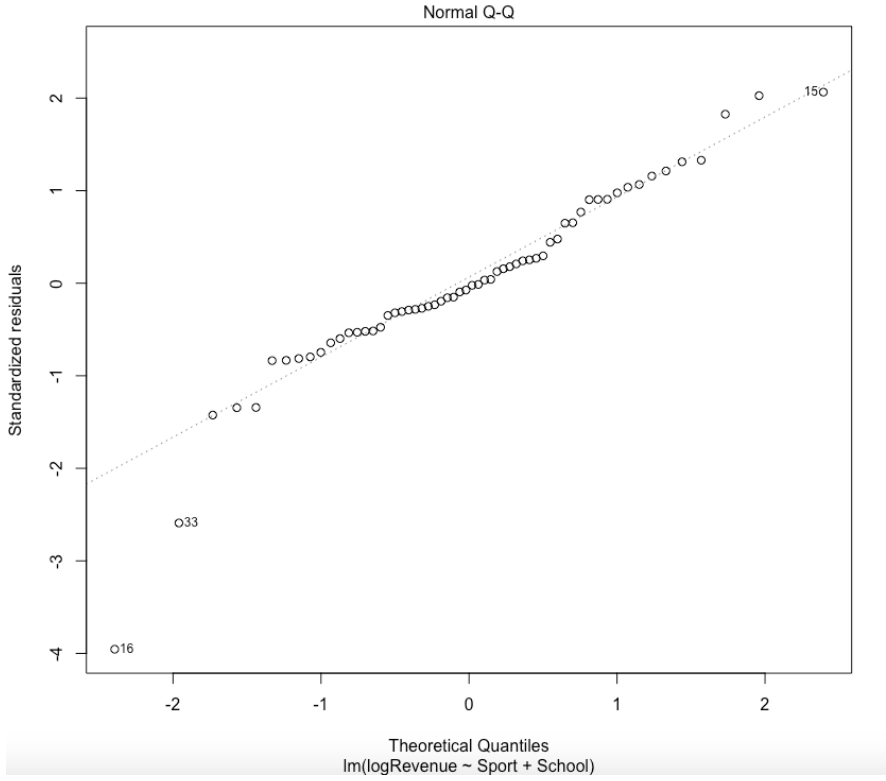
Data Analysis 2.0 – After Transformation

After doing the log transformation on the revenue for the Big Sky data set, we see that from the ANOVA table, that the blocking effect of Sports has a pvalue of < 2.2 e-16, which is less than 0.05. Therefore, we see that the blocking effect of Sports is significant, indicating that the blocking by Sports is working as desired, and it has remained significant from when we did the analysis before the transformation.

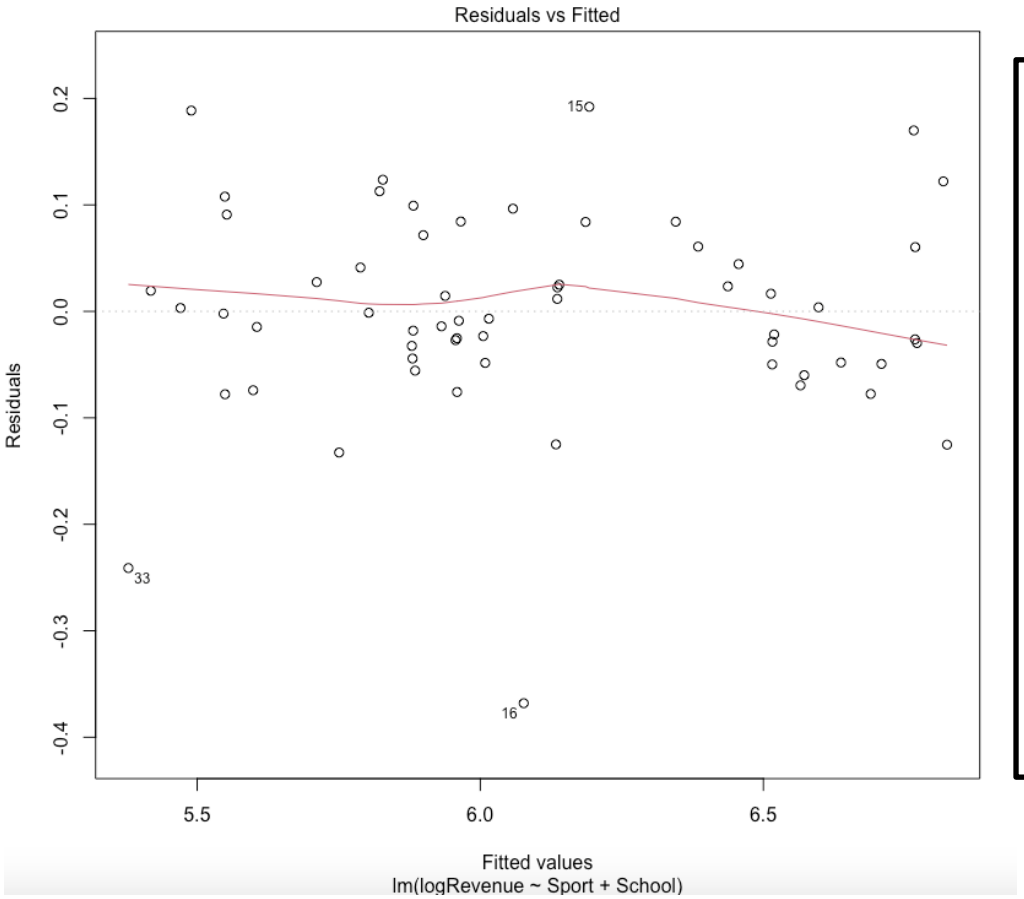
Data Analysis 2.0 – Assumptions?

From: plot(big_sky_revenue_log.lm)



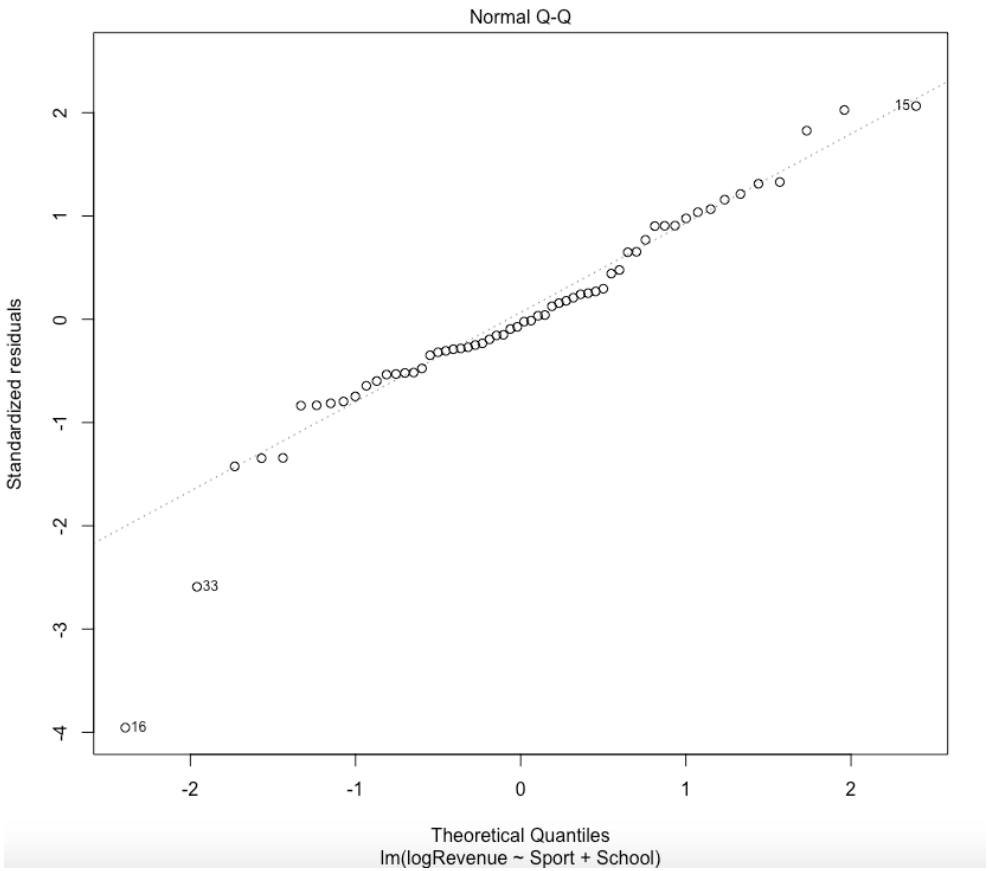


Data Analysis 2.0 – Assumptions?



We see from the residual by predicted plot that there now appears to be a constant band. Thus, the assumption of constant variance of residuals holds, and we may just need to remove some outliers (like observations 16 and 33).

Data Analysis 2.0 – Assumptions?



We see from the normal QQ plot, most of the points fall upon the line, other than a couple of points. However, if we removed the outliers of observations 16 and 33, then most of the points are on the line, so the assumption of normality of the residuals holds.

Data Analysis 2.0 – Tukey's Test

From: summary(glht(big_sky_revenue_log.lm, linfct = mcp(School = "Tukey")))

```
Montana State University - Idaho State University == 0    0.2203804  0.0620446   3.552   0.0280 *
Northern Arizona University - Idaho State University == 0 0.2269333  0.0620446   3.658   0.0206 *
```

From Tukey's Multiple Comparison Test, we see that there are significant difference in the revenue generated between Montana State University and Idaho State University, and a significant difference in revenue generated between Northern Arizona University and Idaho State University. When looking at the data sets (using tapply() to look at the means), we see that Montana State University and Northern Arizona University generate significantly more revenue than Idaho State University.

Discussion

- Though we at first found that the revenues generated were not significantly different among Big Sky schools, after performing a log transformation of the data to allow for assumptions to hold, we found that there was a significant difference in the revenues generated between Big Sky schools.
- Conducting Tukey's Multiple Comparison test on the Big Sky schools, we found that Montana State University and Northern Arizona University generate significantly more revenue than Idaho State University in the year of 2018.

Implications/Future Studies

- Thus, we may want to investigate some of the things that Montana State and Northern Arizona did differently to generate more revenue than Idaho State in the year of 2018.
- It may also be relevant to analyze the expenses for the Big Sky schools in the year of 2018 as more expenses going towards their programs may be the reason why they generate more revenue. So, it may be worth looking at profits instead of revenues in a future project.
- Also, it may be worth investigating and analyzing the revenues or profits generated by a certain school with years as the treatment effect, so a school's athletic department can reflect on what worked well for that particular year compared to others.

References

- "Data Equity in Athletics." *Equity in Athletics*, https://ope.ed.gov/athletics/#/.
- Semb, Lucas. "Montana Grizzlies Are Staying in the Big Sky - for Now, and Ad Kent Haslam Explains Why." *406 MT SPORTS*, 10 Dec. 2022, https://406mtsports.com/college/big-sky-conference/university-of-montana/montana-grizzlies-are-staying-in-the-big-sky-for-now-and-ad-kent-haslam-explains/article_539a1396-77f2-11ed-b132-9b9ce3fd41bc.html.

Thank you for a great semester!!!