Spring 2018

Athletic Event Attendance Breakdown

Joel Klein

Athletic Event Attendance Breakdown

St. Cloud State University

Joel Klein

Spring 2018

Table of Contents

Executive Summary:

St. Cloud State University (SCSU) is home to the Huskies and to the cardinal and black. St. Cloud State Athletics hosts 18 different varsity sports programs including division I men's and women's hockey, and division II men's and women's basketball, football, volleyball, and soccer. Husky Athletics, a staple of the central Minnesota culture, continues to be a main attraction for fans throughout the city of St. Cloud and surrounding areas. Husky Athletics also plays a major role in the student lifestyle at St. Cloud State. From "sieving" at the opposing goalie to enjoying the football tradition alongside the beautiful Mississippi River, students and fans serve as an essential factor to the success of Husky Athletics. In recent years however, much concern over budget costs have forced major cuts to SCSU athletics, namely the discontinuing of men's and women's tennis, women's Nordic skiing, men's cross-country and men's indoor/outdoor track and field in 2016. As reported in the *Star Tribune* in March 2016, "[St. Cloud State] is dealing with falling enrollment and nagging deficits. SCSU enrollment stands at 15,461, down from 18,650 in the fall of 2010 … It is currently battling a $6 million budget gap."

Athletic ticket sales and attendance for the major sports play a big role in attracting prospective students and eliminating the budget gap. This study mainly focuses on analyzing recent attendance trends for athletic events and better developing the SCSU Athletic, Student Life & Development (SLD), and Campus Involvement (DCI) Departments' understanding of the rates and frequencies students and fans are attending such events. This analysis examines the role factors such as opponent, weather, and other events play in both overall and student attendance levels. In addition, through a full breakdown of student attendance and registration records, this analysis compares the profiles of students who attend and do not attend athletic events.

Understanding which student attributes are associated with higher attendance levels influences predicting which students are likely to attend events. Altogether, this analysis provides strategic insights to effectively enhance current university advertising and marketing strategies aimed to acquire higher levels of student attendance in the future. The most important and influential insights from the analysis are:

- Soccer, volleyball, and men's and women's hockey average tickets sold per event increased since the 2012 – 2013 season.
- Average student attendance per event has declined for all sports since the 2014 – 2015 season.
- Overall fan attendance is not strongly correlated to student attendance.
- Only 25 percent of all undergraduate students attend an athletic event during the semester.
- Certain opponents affect both overall fan and student attendance for men's hockey and men's & women's basketball games.
- Rain and snow decrease student attendance for all sports.
- The variables with the biggest effects on the probability a student will attend an event are whether a student attended an event the previous semester and whether he/she lived on campus during his/her first term.

These insights will help influence important decisions regarding the continual improvement in attendance levels at St. Cloud State University athletic events.

Data and Demographics:

The SCSU Department of Analytics and Institutional Research provided multiple data queries for this analysis. There are three main extractions for use: general athletic ticket sales data for events spanning from the 2012 – 2013 season to 2016 – 2017 season (soccer, volleyball, football, men's & women's hockey, men's & women's basketball), student athletic attendance records for events spanning from the 2014 – 2015 season to 2017 – 2018 season (soccer, volleyball, football, men's & women's hockey, men's & women's basketball), and undergraduate enrollment information per semester from Fall 2014 to Fall 2018. These tables' contents allow for analysis and solutions to the main research questions listed above.

There are a total of 403 events listed in the general ticket sales data (2012 – 2013 to 2016 – 2017 season) and 273 events in the student attendance data (2014 – 2015 to 2017 – 2018 season). There were several events that were missing records from the years 2014 – 2017 in the student attendance data. These events are listed in a table in the appendix. A full list of the information these data sources contain is listed in the appendix.

Men's and women's basketball appear in the data as one sport. Typically, both the men's and the women's teams play on the same night one after another. Tickets to these games are sold for the entire night's events, meaning that a fan's ticket counts for both the men's and women's game. Throughout the analysis, men's and women's basketball will be analyzed together. However, there are instances where the men's or women's team host games separately. This will affect how a particular attendance levels are analyzed for some of the research questions.

The main demographics used to assess the attributes of the students who attend/do not attend events are: student cohort, gender, classification (freshman, sophomore, junior, senior), international/non-international, student of color/non-student of color, pell eligibility, gap range, GPA, qualitative points predicted (QPP), honors, location of hometown, term one residence (on/off Campus), and term credits attempted. This same data is also used in model building to predict the probability a student will attend an athletic event.

Research Problems:

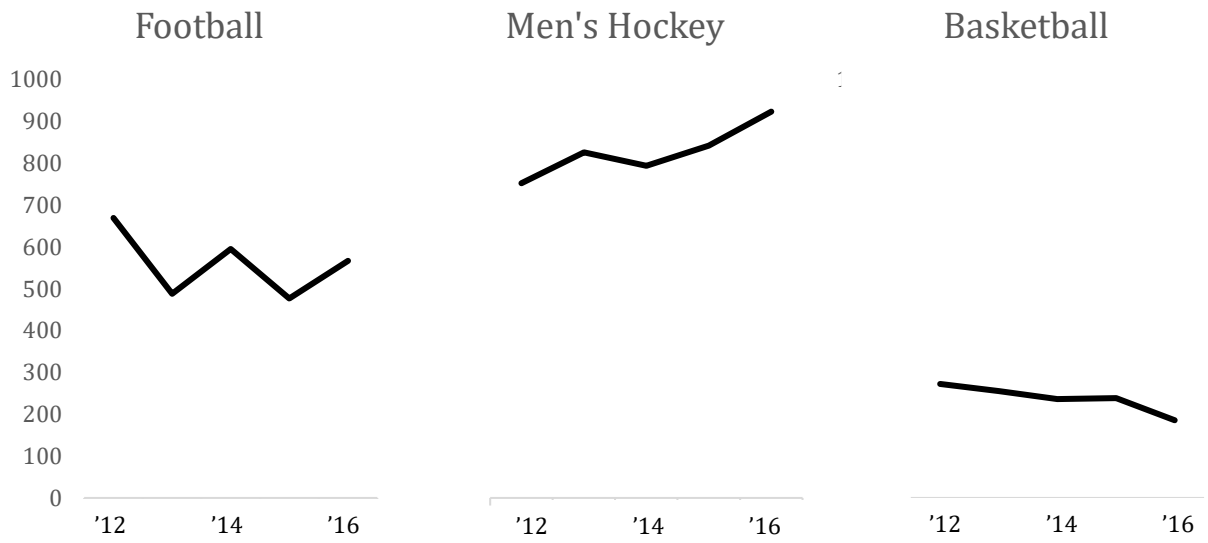
The SCSU Athletic, Student Life & Development (SLD), and Campus Involvement (DCI) departments are searching for insights to describe both the student and general attendance populations as it relates to athletic event attendance. The nine major questions of interest addressed in this report are:

- What is the general and student attendance history for each sport (soccer, volleyball, football, men's & women's hockey, men's & women's basketball) over the past several years?
- How does weather impact attendance? By how much/little?
- Does the opponent impact attendance? By how much/little?
- Is there correlation between student attendance and general (non-student) attendance levels?
- What proportion of the student crowd for each sport is male/female, international/non-international students, student-of-color (SOC)/non-student of color etc.? How does this compare with the overall student population at SCSU?
- What types of students do not attend sporting events? What proportion of students attend 0 events, attend 1 event, attend 2 events, etc.? How many students attend events from only one sport, from two sports, from three sports, etc.?
- How much of a difference is there in attendance for events during winter break?
- Do multiple events on the same date affect overall attendance levels?
- Can a statistical model be built to predict the number of events/if a student will attend an event during a given season?

Overall Attendance Breakdown

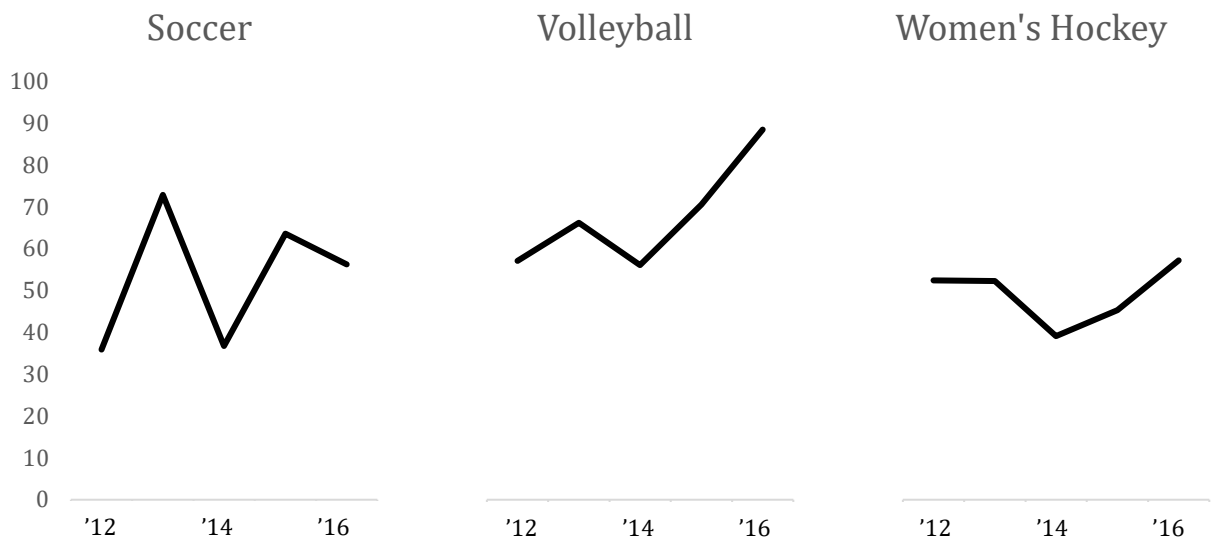
Findings:

Figure 1: *Average Tickets Sold per Event for Men's Sports (2012-2013 to 2016-2017 Seasons)*



A general trend analysis of the average tickets sold per event discovers whether attendance levels have increased or decreased in recent years. This leads insight into which sports attracted higher levels of attendance in the most recent past. Men's hockey is, as expected, the highest attended event of all sports on average over the last five full seasons ranging from 752 tickets sold per event in the 2012 – 2013 season to 923 in the 2016 – 2017 season, a 23 percent increase. It is the only men's sport whose average attendance is on the rise. Football and basketball decreased average attendance by 15 and 32 percent respectively since the 2012 – 2013 season. This is concerning and should be closer evaluated and monitored in the near future. This analysis does not directly address predicting why attendance levels have increased or decreased as it is not the main concern. However, with the summary analytics presented in this report, future student research may examine and develop a prediction model to forecast attendance levels.

Figure 2: *Average Tickets Sold per Event for Women's Sports (2012-2013 to 2016-2017 Seasons)*



Despite decreases in two of the three men's sports, all three women's sport attendance per event, on average, increased from the 2012 – 2013 season. Volleyball increased the most at 55 percent, up from the increases of 36 and nine percent respectively for soccer and women's hockey. Although the average attendance increased over the past five seasons, there are relatively small attendance levels and more variability in the average attendance per year. Soccer since the 2012 – 2013 season year over year has increased, decreased, increased, and decreased. Again, although this study is not focused on forecasting general attendance levels, it would most likely be more difficult to accurately predict future attendance levels for women's sports because of the relative instability and variation from year to year.

Figure 3: *Average Tickets Sold by Opponent (Men's Hockey) vs. SCSU (2012-2013 to 2016-2017 Seasons)*

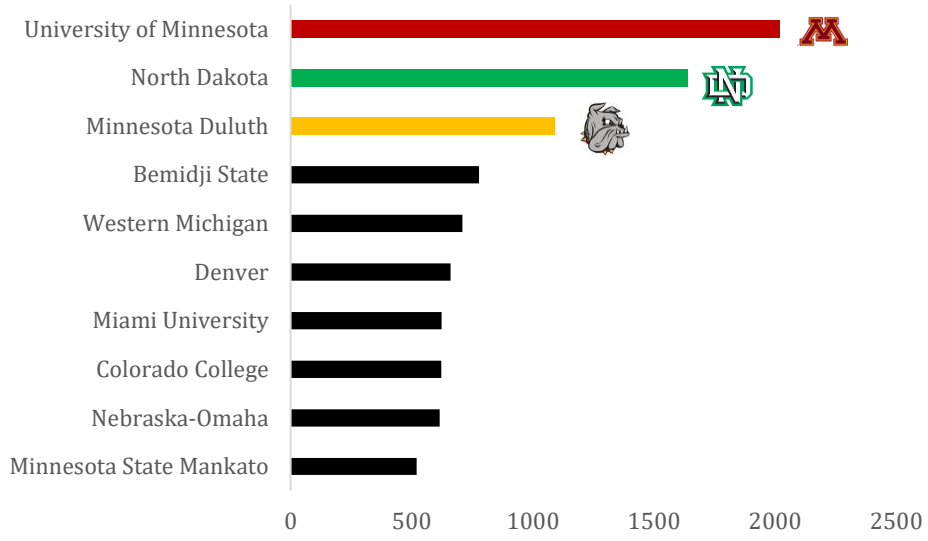


Table 1: *Average Tickets Sold by Opponent (Men's Hockey) vs. SCSU (2012-2013 to 2016-2017 Season)*

Opponent	Avg. Tickets Sold	N Events
University of Minnesota	2021	4
North Dakota	1640	10
Minnesota Duluth	1092	8
Bemidji State	778	2
Western Michigan	709	6
Denver	660	8
Miami University	624	10
Colorado College	623	8
Nebraska-Omaha	615	8
Minnesota State Mankato	520	4

** Note: This report only examines the average attendance levels by opponent for men's hockey and men's & women's basketball as they are the only two sports with large enough sample sizes and high enough attendance levels to decipher a practical difference against opponents. For men's hockey the analysis only considers NCHC opponents and opponents in the state of Minnesota (University of Minnesota, Minnesota State Mankato, Bemidji State). For men's & women's basketball, the analysis only considers NSIC opponents.*

As expected, the opponent effects the average attendance levels. Rival schools such as the University of Minnesota, University of North Dakota, and the University of Minnesota Duluth are the top three teams who attract higher attendance levels to men's hockey games. To much surprise, Minnesota State Mankato, another rival school, generated the lowest tickets sold (520 on average). Because Minnesota State Mankato, University of Minnesota, and Bemidji State are not in the NCHC, SCSU does not host many games against these teams. The average tickets sold per event for all men's hockey games was 824. The only two schools whose average attendance was significantly different from the overall average attendance at the five percent level were the University of Minnesota and the University of North Dakota. These small sample sizes may not lead to much statistical significance, but they do show practical significance to better decipher which opponents rank the highest in attracting fans.

Figure 4: *Average Tickets Sold by Opponent (Men's & Women's Basketball) vs. SCSU (2012-2013 to 2016-2017 Seasons)*

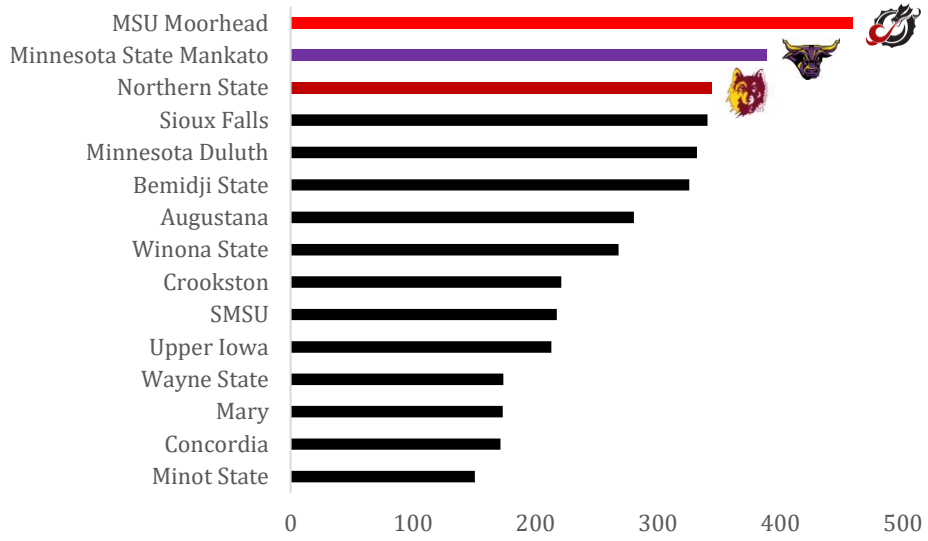


Table 2: *Average Tickets Sold by Opponent (Men's & Women's Basketball) vs. SCSU (2012-2013 to 2016-2017 Seasons)*

Opponent	Avg. Tickets Sold	N Events
MSU Moorhead	460	5
Minnesota State Mankato	389	3
Northern State	347	5
Sioux Falls	341	2
Minnesota Duluth	332	5
Bemidji State	326	5
Augustana	280	3
Winona State	268	3
Crookston	221	5
SMSU	218	2
Upper Iowa	213	3
Wayne State	174	3
Mary	173	5
Concordia	171	3
Minot State	151	5

The average tickets sold per event changes by the opponent men's & women's basketball faces. However, unlike the men's hockey schedule, the basketball schedule is structured differently. Men's & women's basketball hosts games against more schools. There are 8 teams in the NCHC and the men's hockey team hosts each team in the conference for two games each year. There are 16 teams in the NSIC conference and men's & women's basketball does not host games against each team each year. This explains why the sample sizes for each opponent are generally lower over the same time period. It is difficult to arrive at a reliable conclusion for whether the average tickets sold for a particular opponent is different from the average of all games due to the grouping of both men's & women's basketball events in the data set. Some games have relatively low attendance and might be due to only one of the two teams playing that day. This drives the true average of all games down and does not allow for an accurate statistical comparison. However, this analysis shows practical significance to better decipher which opponents rank higher in attracting fans. The most attended opponents are as expected. The top three rivals for basketball are Minnesota State Moorhead, Minnesota State Mankato, and Northern State and the average tickets sold reflects this (460, 389, 347 tickets respectively).

Figure 5: *Percentage Increase/Decrease in Average Tickets Sold when Raining (2012-2013 to 2016-2017 Seasons)*

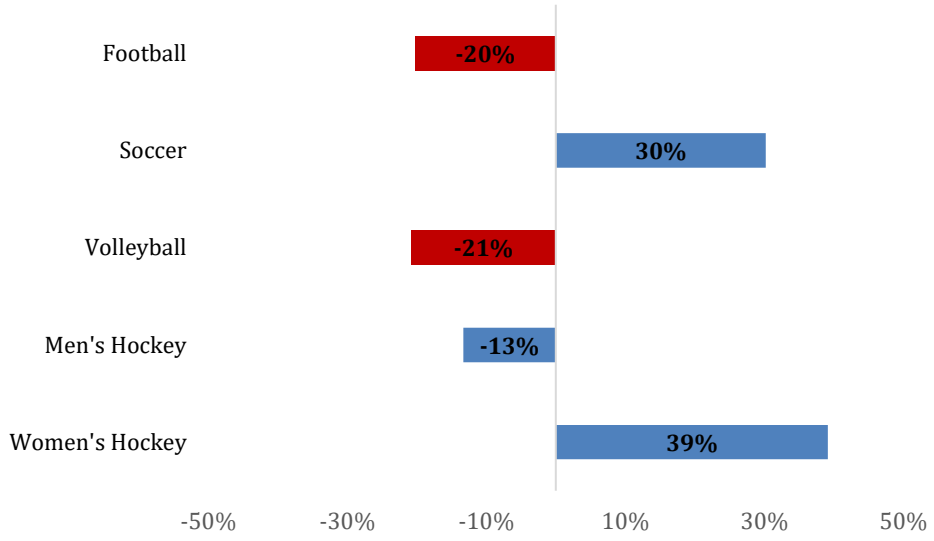


Table 3: *Percentage Increase/Decrease in Average Tickets Sold when Raining (2012-2013 to 2016-2017 Seasons)*

Sport	Avg. Tickets Sold w/ No Rain	Avg. Tickets Sold w/ Rain	N Events w/ No Rain	N Events w/ Rain	P-Value
Football	589	470	19	8	0.19
Soccer	51	66	76	15	0.80
Volleyball	73	57	38	8	0.03
Men's Hockey	843	730	34	16	0.19
Women's Hockey	46	64	70	16	0.93

** Note: Weather was extracted from Weather Underground's historic records from the St. Cloud Regional Airport from 2012 to 2018.*

It is hypothesized that bad weather may result in lower attendance levels. This may be due to poor driving conditions, poor viewing conditions, or other negative attributes associated with the disruptive Minnesota weather. This study mainly focuses on the effect of precipitation on average tickets sold per event. Only fall sports were considered for the rain analysis. The statistical tests performed tested if rain decreased average tickets sold per event. Only volleyball average tickets sold for events with rain was significantly less (21 percent) than the average for

events without rain. Although the difference in average ticket sales for football and men's hockey games is not large enough to detect statistical significance, there still is value to knowing that historically the average tickets sold decreased by almost 120 (20 percent) and 113 (13 percent) respectively.

Figure 6: *Percentage Increase/Decrease in Average Tickets Sold when Snowing (2012-2013 to 2016-2017 Seasons)*

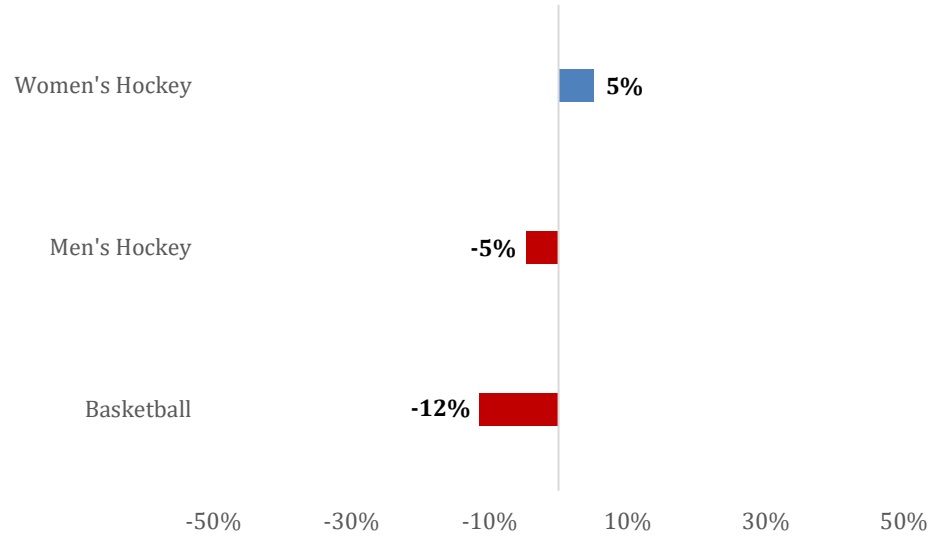


Table 4: *Percentage Increase/Decrease in Average Tickets Sold when Snowing (2012-2013 to 2016-2017 Seasons)*

Sport	Avg. Tickets Sold w/ No Snow	Avg. Tickets Sold w/ Snow	N Events w/ No Snow	N Events w/ Snow	P-Value
Women's Hockey	49	52	68	18	0.58
Men's Hockey	834	794	69	22	0.39
Basketball	242	214	58	19	0.24

Analysis similar to the effect of rain on fall sports was conducted to determine if snow affected the average tickets sold per event for winter sports. The statistical tests performed tested if snow decreased average tickets sold per event. Although no sport's difference in average tickets sold per event is large enough to detect statistical significance at the five percent level, there is value to knowing that historically the average tickets sold decreased for men's & women's basketball and men's hockey by 12 and five percent respectively.

Table 5: *Correlation between Tickets Sold and Student Attendance (2014-2015 to 2016-2017 Seasons)*

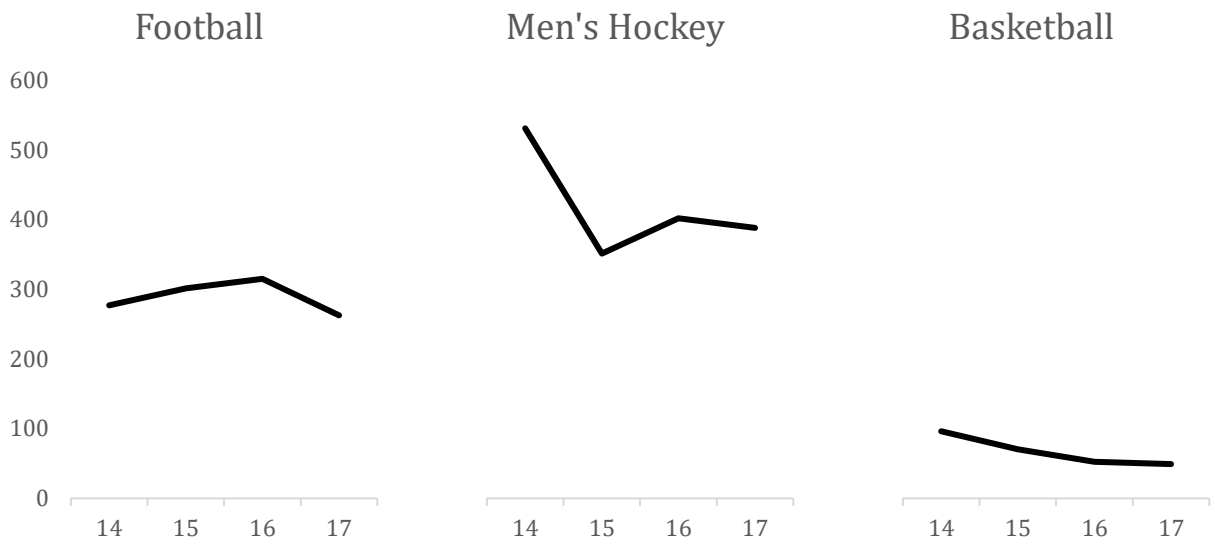
Sport	Correlation	Number of Events	P-value
Football	0.48	15	0.07
Men's Hockey	0.38	51	0.01
Women's Hockey	0.16	47	0.27
Basketball	-0.02	47	0.87
Volleyball	-0.07	27	0.72
Soccer	-0.11	22	0.61

The final overall attendance analysis focuses on discovering if overall tickets sold to events is positively correlated with student attendance. The analysis, performed for each sport, plotted tickets sold by student attendance. The correlation coefficient shows if a linear relationship exists. Generally, a correlation coefficient greater than 0.7 shows a strong positive relationship. The strongest relationship between overall and student attendance, for football, only produced a coefficient of 0.52, a relatively moderate positive relationship. In general, there is not a positive relationship between overall and student attendance at events.

Student Attendance Breakdown

Findings:

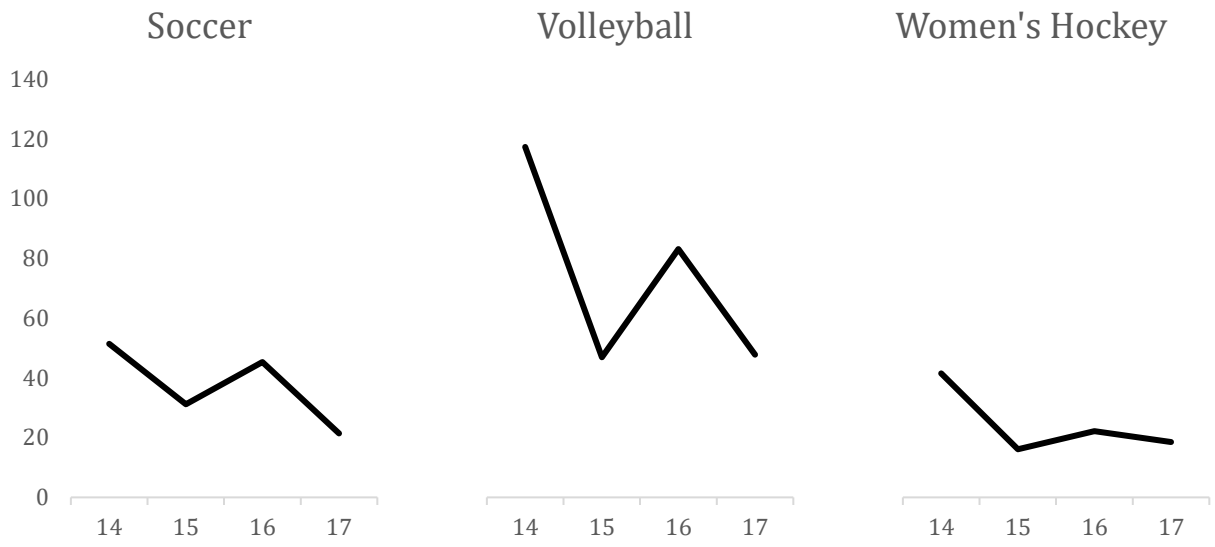
Figure 7: *Average Attendance per Event for Men's Sports (2014-2015 to 2017-2018 Seasons)*



A general trend analysis of the average student attendance per event shows whether attendance levels have increased or decreased in recent years. This gives insight into which sports attracted higher levels of student attendance in the years past. To no surprise, men's hockey is the highest attended event of all sports on average over the last five full seasons ranging from 530 students attending per event in the 2014 – 2015 season to 387 in the 2017 – 2018 season. All men's sport average student attendance fell since the 2014 – 2015 season. Undergraduate student enrollment levels have decreased from 13,752 in Fall 2014 to 13,236 in Fall 2016 (a 3.75 percent decrease). Football, men's hockey, and men's & women's basketball average student attendance per event decreased by 5, 27, and 49 percent since the 2014 – 2015 season respectively. This is concerning and should be closer evaluated and monitored in the near future as attendance levels are dropping faster than enrollment levels. This analysis does not directly address predicting why attendance levels have increased or decreased, as it is not the main concern. However, with the summary

analytics presented in this report, future student research may examine and develop a prediction model to forecast attendance levels.

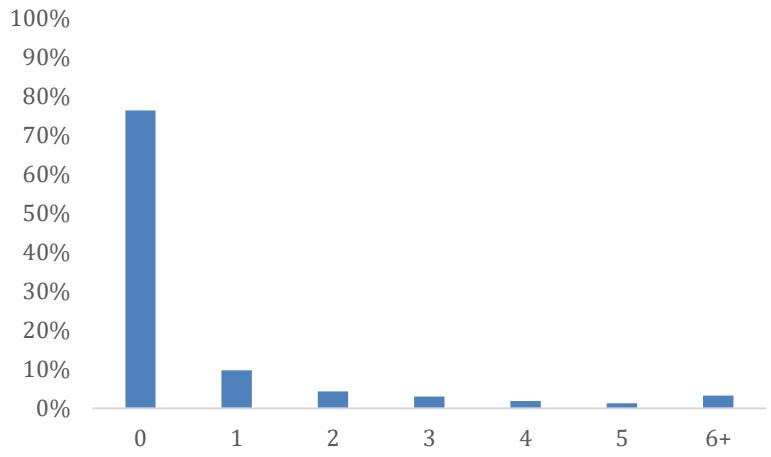
Figure 8: *Average Attendance per Event for Women's Sports (2014-2015 to 2017-2018 Seasons)*



Like all three men's sports, all three women's sports attendance per event, on average, decreased since the 2014 – 2015 season. Volleyball decreased the most at 59 percent, higher than the decreases of 58 and 55 percent respectively for soccer and women's hockey. Although the relative average attendance decrease over the past five seasons seems dramatic, there are small attendance levels so a small drop in the number of students attending will be a relatively large percent decrease. Volleyball, soccer, and women's hockey since the 2014 – 2015 season year over year has decreased, increased, and decreased.

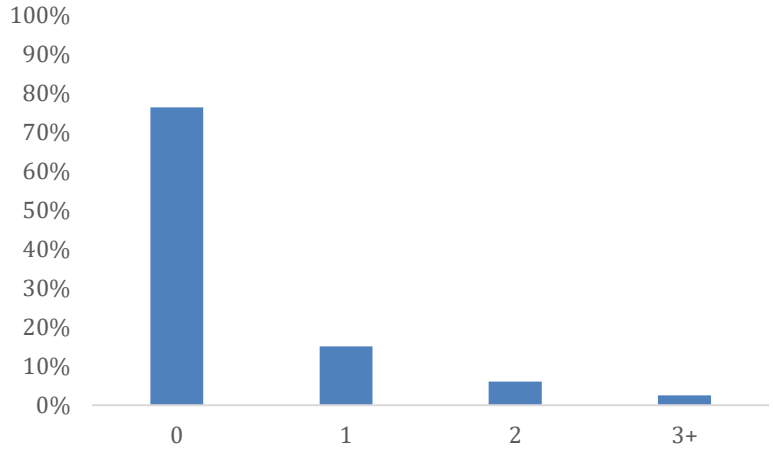
The decreasing student attendance levels are a cause for concern. The remainder of this analysis focuses on the student populations that are attending and the student populations that are not attending. This gives insight and summary information to further understand traits and behaviors of attending and non attending students. Based on such information and findings, staff may formulate more effective strategies to increase attendance levels.

Figure 9: *Percentage of Students who Attended 0, 1, 2, etc. Events in Fall 2017*



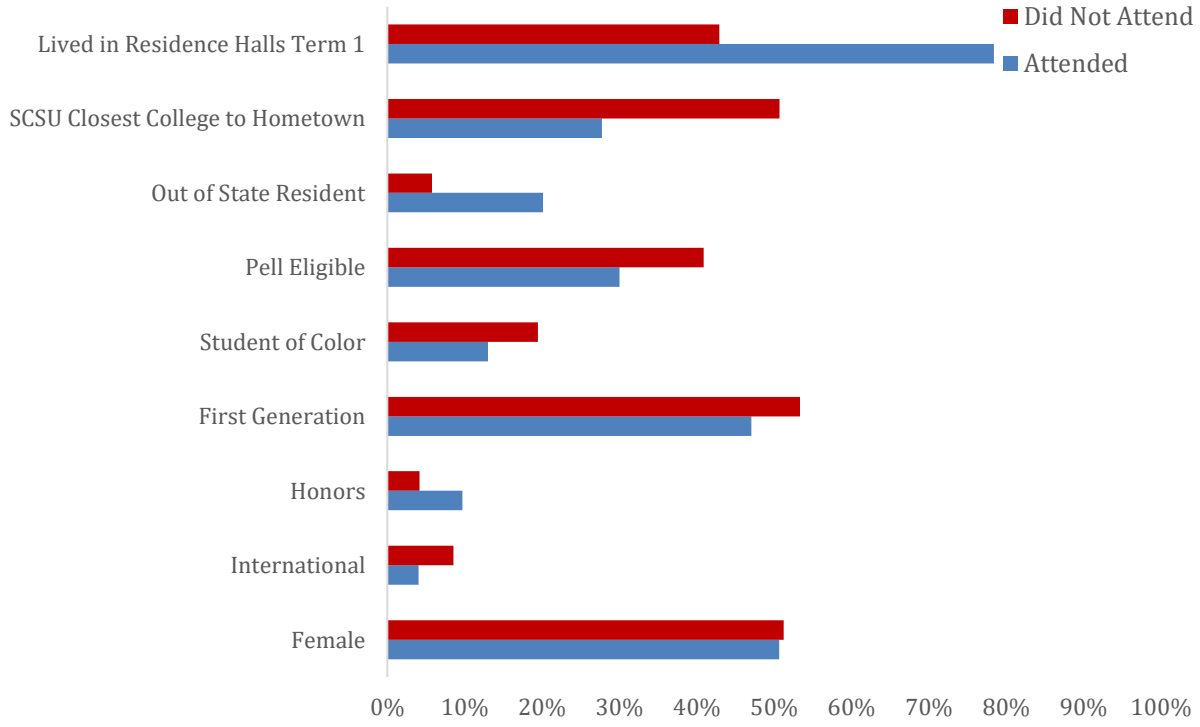
In the Fall 2017 semester 75 percent of undergraduate students did not attend a single athletic event. Of the remaining 25 percent, most only attended one or two events. The next analysis further analyzes the number of different sports students attend.

Figure 10: *Percentage of Students who Attended 0, 1, 2, etc. Different Sports in Fall 2017*



Again, in the Fall 2017 semester 75 percent of undergraduate students did not attend a single athletic event. 15 percent attended only one distinct sport. Six percent attended two different sports and only three percent of undergrads attended three or more different sports.

Figure 11: *Percentage of Fall 2014 – Fall 2017 Event Attending & Non-Attending Students of Demographic*



This analysis compares the percentage of students with a certain demographic if they attended an event or not. For example, of the students who attended an event, 19 percent are students of color. Of the students who do attend an event, 13 percent are students of color. This analysis allows for comparing the prevalence of a particular demographic for the students who attended an event and the prevalence of the demographic for students who did not attend. The biggest difference is the percentage of attending students who lived on campus their first term versus the percentage of non-attending students who lived on campus. 78 percent of the students who attended an event lived on campus their first term. Of the non-attending students, only 43 percent lived on campus their first term. The second large difference between the attending and non-attending students was the prevalence of students whose hometowns were close to St. Cloud State. Of the students who attended an event, 20 percent were out of state residents, compared to

only 8 percent of the non-attending students. Also, 28 percent of attending students grew up with SCSU as the closest university, compared to 51 percent of the non-attending students. Students who attend events are more likely from places further from St. Cloud State compared to the students who do not attend. This may result from a higher percentage of students from further away areas living in the dorms during their first term. Typically students living in the dorms more actively participate in campus activities.

Figure 12: *Summary Statistics of Fall 2014 – Fall 2017 Event Attending & Non-Attending Students*

Attended Event		***Did Not Attend an Event***
2.94	Avg. Semester GPA	2.86
2.96	Avg. Cumulative GPA	2.90
14.1	Avg. Semester Credits Attempted	12.6
51.8	Avg. Cumulative Credits Attempted	64.0
76.6	Avg. QPP	73.8

This analysis compares a handful of summary statistics for students who attended an event and those who did not attend an event. For example, the average semester GPA for students who attended an event is 2.94 compared to 2.86 for those who do not attend an event. Students who attend athletic events, on average, have a slightly higher semester GPA and cumulative GPA, attempt more credits in a semester, and have a higher QPP. Also note the average cumulative credits attempted is lower for attending students by roughly 12 credits. This may show that attending students are earlier in their program than non-attending.

Figure 13: *Average Student Attendance by Opponent (Men's Hockey) vs. SCSU (2014-2015 to 2017-2018 Seasons)*

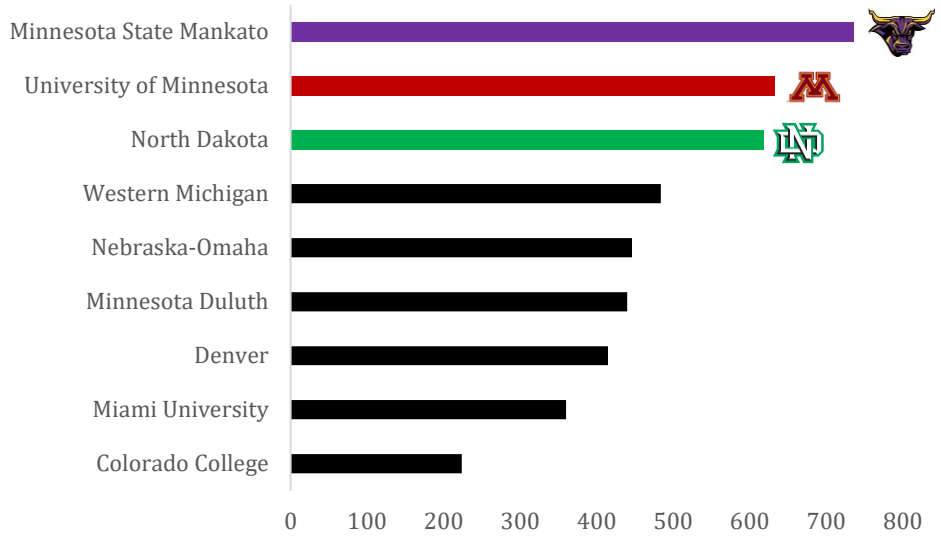


Table 6: *Average Student Attendance by Opponent (Men's Hockey) vs. SCSU (2014-2015 to 2017-2018 Seasons)*

Opponent	Avg. Attendance	N Events
Minnesota State Mankato	737	2
University of Minnesota	632	3
North Dakota	618	8
Western Michigan	484	4
Nebraska-Omaha	446	6
Minnesota Duluth	440	7
Denver	415	4
Miami University	360	6
Colorado College	224	6

** Note: This report only examines the average attendance levels by opponent for men's hockey and men's & women's basketball as they are the only two sports with large enough sample sizes and high enough attendance levels to decipher a practical difference against opponents. For men's hockey the analysis only considers NCHC opponents and opponents in the state of Minnesota (University of Minnesota, Minnesota State Mankato, Bemidji State). For men's & women's basketball, the analysis only considers NSIC opponents.*

As expected, the opponent affects the average attendance levels. Rival schools such as Minnesota State Mankato, University of Minnesota, and University of North Dakota are the top three teams who attract higher student attendance levels to men's hockey games. Because Minnesota State Mankato, and the University of Minnesota are not in the NCHC, SCSU does not host as many games against these teams, explaining the small sample size. The average attendance per event for all men's hockey games was 422. The only three schools whose average attendance was significantly different from the overall average attendance at the five percent level were Minnesota State Mankato, University of North Dakota, and Colorado College (lower). The small sample sizes may cause a lack in statistical significance. However, there still exists practical significance to better decipher which opponents rank the highest in attracting students.

Figure 14: *Average Student Attendance by Opponent (Men's & Women's Basketball) vs. SCSU (2014-2015 to 2017-2018 Seasons)*

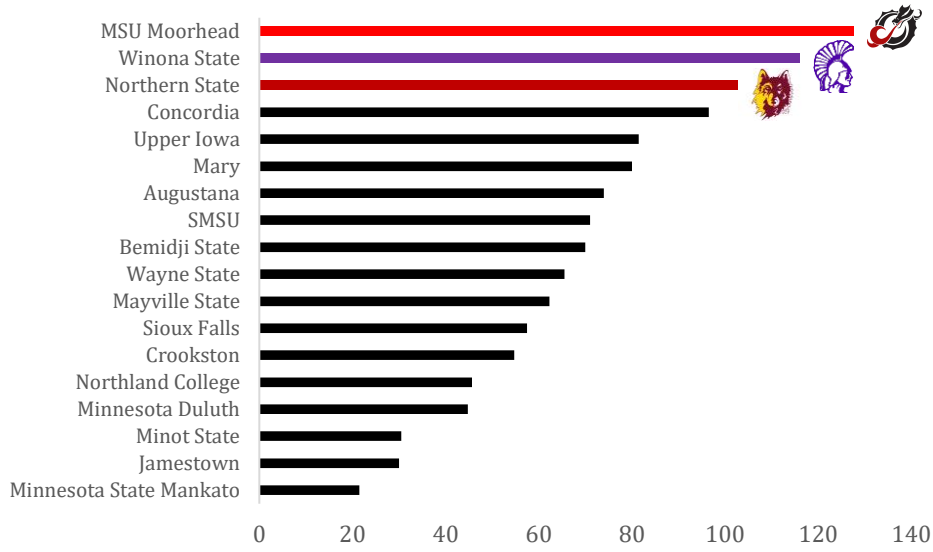


Table 7: *Average Student Attendance by Opponent (Men's & Women's Basketball) vs. SCSU (2014-2015 to 2017-2018 Seasons)*

Opponent	Avg. Attendance	N Events
MSU Moorhead	128	3
Winona State	116	2
Northern State	104	3
Concordia	97	2
Upper Iowa	82	2
Mary	80	3
Augustana	74	2
SMSU	71	2
Bemidji State	70	4
Wayne State	66	2
Mayville State	62	3
Sioux Falls	58	2
Crookston	55	4
Northland College	46	3
Minnesota Duluth	45	4
Minot State	31	2
Jamestown	30	3
Minnesota State Mankato	22	2

The average student attendance per event changes by the opponent men's & women's basketball faces. Again, unlike the men's hockey schedule, the basketball schedule is structured differently. Men's & women's basketball hosts games against more schools. This explains why the sample sizes for each opponent are generally lower over the same time period. It is difficult to arrive at a reliable conclusion for whether the average student attendance for a particular opponent is different from the average of all games. This is again due to the grouping of both men's & women's basketball events in the data set. Some games have really low attendance and might be due to only one of the two teams playing that day. This drives the true average of all games down and does not allow for an accurate statistical comparison. However, this analysis shows practical significance to better decipher which opponents rank higher in attracting fans. The most attended opponents are top rivals for basketball: Minnesota State Moorhead, Winona State, and Northern State. The average student attendances for these games are 128, 116, and 104 respectively. The surprisingly low average attendance for Mankato is likely due to the date the two events took place: January 2, 2016 and November 26, 2016.

Figure 15: *Percentage Increase/Decrease in Average Attendance when Raining (2014-2015 to 2017-2018 Seasons)*

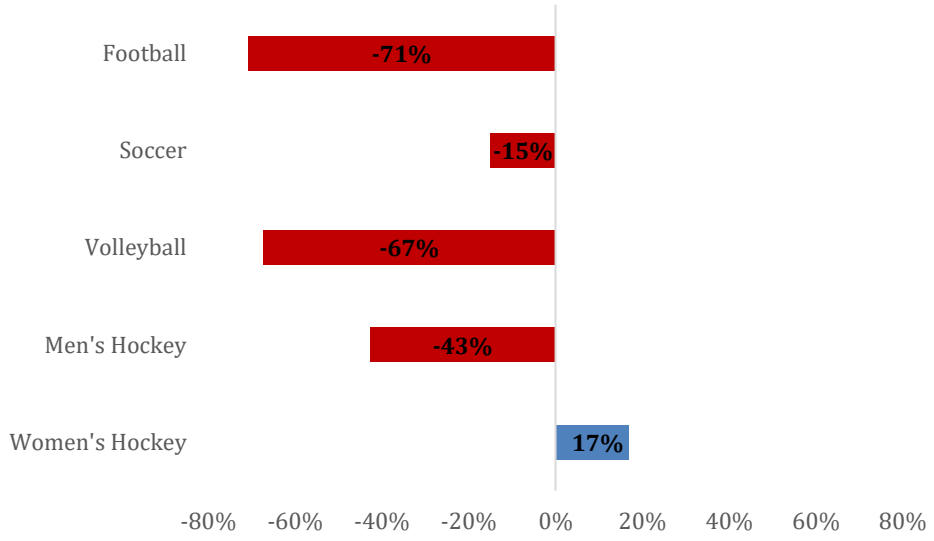


Table 8: *Percentage Increase/Decrease in Average Attendance when Raining (2014-2015 to 2017-2018 Seasons)*

Sport	Avg. Attendance w/ No Rain	Avg. Attendance w/ Rain	N Events w/ No Rain	N Events w/ Rain	P-Value
Football	336	98	16	4	<0.01
Soccer	44	14	22	9	<0.01
Volleyball	86	49	25	12	0.05
Men's Hockey	434	368	52	11	0.14
Women's Hockey	25	29	44	9	0.70

** Note: Weather was extracted from Weather Underground's historic records from the St. Cloud Regional Airport from 2012 to 2018.*

It was shown earlier in the report that bad weather results in lower overall attendance levels. Weather might also affect student levels. This study mainly focuses on the effect of precipitation on average student attendance. Only fall sports were considered for the rain analysis above. The statistical tests performed tested if rain decreased average student attendance. Football, volleyball, and soccer average student attendance for events with rain was significantly less than the average for events without rain at the five percent level. These sports experienced 71, 67, and

15 percent decreases in average student attendance respectively. Although the difference in student attendance for men's hockey games was not large enough to detect statistical significance, there is value to knowing that historically student attendance per event decreased by 66 (43 percent).

Figure 16: *Percentage Increase/Decrease in Average Attendance when Snowing (2014-2015 to 2017-2018 Seasons)*

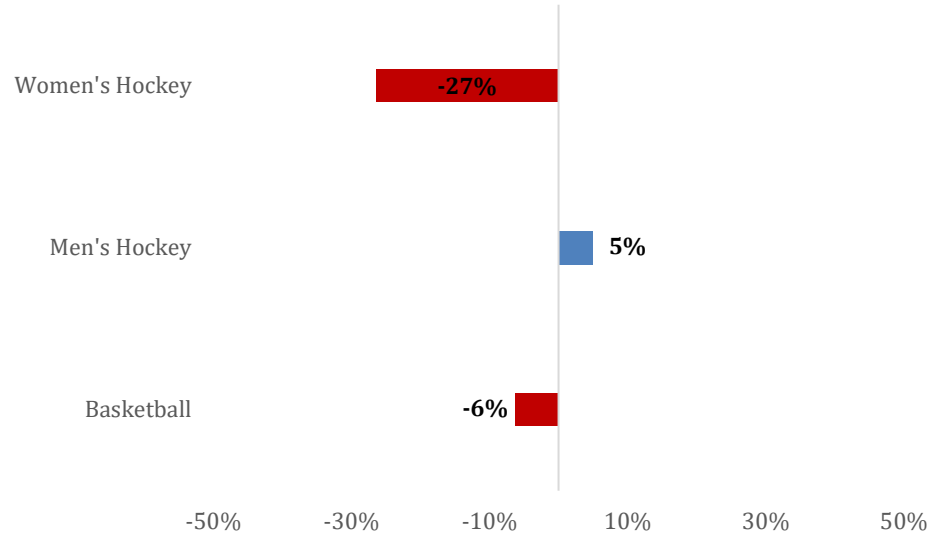


Table 9: *Percentage Increase/Decrease in Average Attendance when Snowing (2014-2015 to 2017-2018 Seasons)*

Sport	Avg. Attendance w/ No Snow	Avg. Attendance w/ Snow	N Events w/ No Snow	N Events w/ Snow	P-Value
Women's Hockey	26	19	46	7	0.11
Men's Hockey	419	439	51	12	0.64
Basketball	68	64	47	9	0.39

Analysis similar to the effect of rain on fall sports was conducted to determine if snow affected the average student attendance per event for winter sports. The statistical tests performed tested if snow decreased average student attendance per event. No sport's difference in average tickets sold per event was large enough to detect statistical significance at the five percent level. Average attendance levels for basketball and women's hockey were so low that the test does not say much. We are 95% confident that snow does not lower student attendance for winter sports.

Figure 17: *Average Student Attendance by Sport for Semester Events and Winter Break Events (2014-2015 to 2017-2018 Seasons)*

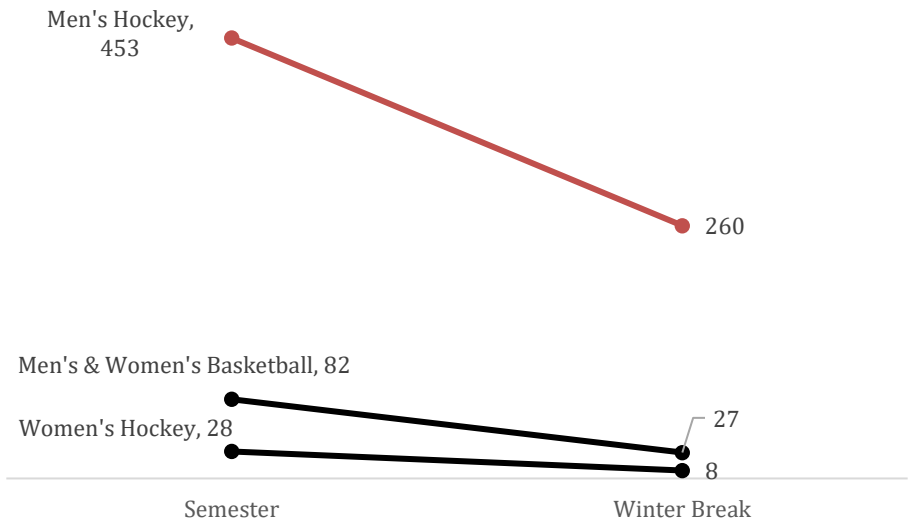


Table 10: *Average Student Attendance by Sport for Semester Events and Winter Break Events (2014-2015 to 2017-2018 Seasons)*

Sport	Semester Avg. Attendance	Winter Break Avg. Attendance	N Semester Events	N Winter Break Events	P-Value
Women's Hockey	28	8	47	6	<0.01
Men's Hockey	453	260	53	10	<0.01
Basketball	82	27	42	14	<0.01

It is hypothesized that events taking place during winter break may result in lower student attendance levels. This study mainly focuses on the effect of winter break on average student attendance. Only sports with events during winter break were considered for the analysis. The statistical tests performed tested if winter break decreased average student attendance. Basketball, men's hockey, and women's hockey average student attendance for events during winter break was significantly less than the average for events during the semester at the five percent level. These sports experienced strikingly large percent decreases in average student attendance: 66, 43, and 71 respectively.

Figure 18: *Average Student Attendance by Sport for Events with Multiple Events on the Same Day (2014-2015 to 2017-2018 Seasons)*

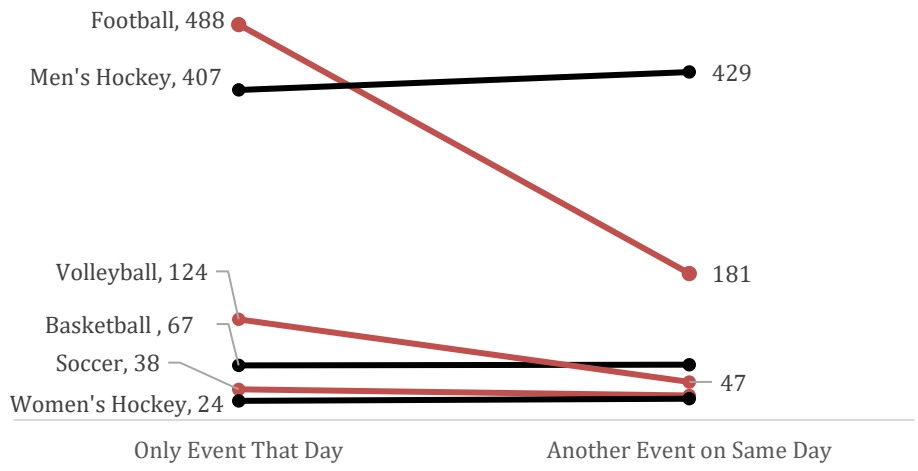


Table 11: *Average Student Attendance by Sport for Events with Multiple Events on the Same Day (2014-2015 to 2017-2018 Seasons)*

Sport	Avg. Attendance Only Event That Day	Avg. Attendance Another Event on Same Day	N Events w/ Only Event That Day	N Events w/ Another Event on Same Day	P-Value
Women's Hockey	24	26	13	40	0.75
Men's Hockey	407	429	20	43	0.70
Basketball	67	68	29	27	0.94
Volleyball	124	47	13	24	0.02
Soccer	38	30	20	11	0.40
Football	488	181	7	13	0.07

Analysis similar to the effect of winter break on student attendance levels was conducted to determine if multiple events held on the same day altered average student attendance per event. Only volleyball's difference in average student attendance was large enough to detect statistical significance at the five percent level. Although the result is not statistically significant, football student attendance levels historically decreased by 307 students (63 percent) when other events are held that same day.

The analysis and summary information discovered in the exploratory analysis above provides insight for building a predictive model to determine the probability a student will attend an athletic event during the fall semester. The model considers student demographic information, GPA, credits attempted, and QPP to determine the probability. The sample for the model includes a record for each student enrolled during the fall semesters from 2014 to 2017. A single student can be a part of the modeling sample several times if they are enrolled during multiple fall semesters. To build the model, the observed data was split into three groups: 60 percent to train the model, 30 percent to validate the model, and 10 percent to test the model.

Table 12: *Logistic Model to Predict the Probability a Student will Attend Athletic Event (2014-2015 to 2017-2018 Seasons)*

Term	Estimate	P- Value	Effect Size	Power (Decrease)	Power (Increase)
Intercept	-3.99	<0.01			
Attended Event Last Fall	1.83	<0.01	43%	70%	64%
Term 1 Campus Resident	1.39	<0.01	32%	94%	91%
Semester Attempted Credits	0.12	<0.01	25%	100%	100%
Attended Event Last Fall is Missing	0.93	<0.01	21%	89%	85%
From Out of State	0.85	<0.01	19%	72%	67%
Honors	0.35	<0.01	7%	49%	44%
Cumulative GPA	0.11	<0.01	4%	100%	100%
Female	-0.19	<0.01	-3%	97%	95%
Pell Eligible	-0.25	<0.01	-4%	95%	92%
Student of Color	-0.42	<0.01	-7%	77%	72%
Total Attempted Credits	-0.01	<0.01	-8%	100%	100%
Student of Color is Missing	-0.73	<0.01	-11%	9%	8%
New Incoming Freshman Missing GPA	-2.77	<0.01	-23%	4%	4%

** Note: The estimates are reported for a one unit change in the variable. The effect size is the change in probability of the response from the mean probability when the predictor is present. The effect sizes are calculated over the range of the predictor variable. The effect of semester-attempted credits (25%) is the effect of increasing the semester-attempted credits from 8 to 17. The effect of cumulative GPA (4%) is the effect of increasing the cumulative GPA 2 points. The effect of total attempted credits (-8%) is the effect of increasing total attempted credits by 100 credits. Power (Decrease/Increase) is the probability of correctly detecting a 2.5 percentage*

point decrease/increase in the average response probability for the given variable, with all other variables held constant.

Table 13: *Fit Statistics for Logistic Model*

Statistic	Training	Validation	Test
Entropy R Square	0.22	0.21	0.23
Area Under ROC Curve	0.82	0.81	0.82
Lift for Predicting Top 20%	2.50	2.50	2.50

The logistic model explanatory terms are listed in Table 12. All of the variables in the model are significant at the five percent level. There is also no need for p-value cutoff adjustment procedures to minimize the probability of committing a type one error. Although all add value in predicting the probability a student will attend an event, some have a larger effect relative to others. If a student attended an event last fall, if the student lived in the dorms during their first term, and the number of semester credits attempted have the largest effect on the outcome. Their respective effect sizes across their entire range were 43, 32, 25 percentage points. Females, pell eligible students, and students of color are all less likely to attend an event that males, non-pell eligible, and non-students of color.

The average probability a student attends an athletic event during the fall semester is 25 percent. This analysis is interested in detecting a change of at least 2.5 percentage points in either direction of the mean response for a particular demographic. Thus, if the retention is estimated to be 27.5 percent (or higher) or 22.5 percent (or lower), the demographic, with all other variables held constant, changes the probability at least 2.5 percentage points. The power (probability that the test will accurately detect such a difference) for each variable is reported in Table 12. For example, if a student attended an event last fall, the probability of correctly detecting a 2.5 percentage point increase in the probability a student attends an event is 64 percent. Because of

the small standard errors due to large sample sizes, the power of detecting a 2.5 percentage point difference for some predictor variables is 100 percent. In general, statisticians aim for a percentage of 80 percent or higher. Further descriptions of the types of statistical error, p-values, and power are summarized in Appendix C.

For all provided information, the model performs relatively well in predicting the outcome. Good accuracy measures for comparing logistic models are R-squared, misclassification rate, area under the receiver operating curve (ROC) curve, and lift. Generally, statisticians consider a model a 0.82 area under the ROC curve a good performing model. The model also more accurately selects the top 20 percent most likely to attend an event 2.5 times better than if students from the population were selected at random.

Conclusion:

St. Cloud State's vision is: *"Through active discovery, applied knowledge and creative interaction, we positively transform our students and the communities where they live and work."* This analysis leverages these pillars through student research to provide several departments throughout St. Cloud State (Athletic, Student Life & Development (SLD), and Campus Involvement (DCI)) with the insights necessary to further promote student attendance to sporting events. The main discoveries in this research are:

- Soccer, volleyball, and men's and women's hockey average tickets sold per event increased since the 2012 – 2013 season.
- Average student attendance per event has declined for all sports since the 2014 – 2015 season.
- Overall fan attendance is not strongly correlated to student attendance.
- Only 25 percent of all undergraduate students attend an athletic event during the semester.
- Certain opponents affect both overall fan and student attendance for men's hockey and men's & women's basketball games.
- Rain and snow decrease student attendance for all sports.
- The variables with the biggest effects on the probability a student will attend an event are whether a student attended an event the previous semester and whether he/she lived on campus during his/her first term.

Not only did this analysis determine which student attributes are associated with higher attendance levels, it also built a reliable predictive model to predict the probability a student will attend an athletic event. Based on these predictions, marketing campaigns can increase student

attendance levels by more effectively targeting students with a higher likelihood of attending an event. The analysis presented is a general overview for all the major sports. For future research, similar methods and procedures could be applied to perform a more in-depth analysis for attendance levels and trends for each sport.

Appendix A: Means of Analysis

Methods and Solution:

Problem: What is the general and student attendance history for each sport (soccer, volleyball, football, men's & women's hockey, men's & women's basketball) over the past several years?

- Trend analysis of general tickets sold per sport from the 2012 – 2013 season to the 2016 – 2017 season and student attendance per sport from the 2014 – 2015 season to the 2017 – 2018 season to examine the differences in attendance overtime and predict future attendance levels.

Problem: How does weather impact attendance? By how much/little?

- Comparative analysis of both general ticket sales and student event attendance by sport for events by average temperature and precipitation. This will show the size of the impact that temperature and precipitation play in overall attendance levels.

Problem: Does the opponent impact attendance? By how much/little?

- Comparative analysis of both general ticket sales and student event attendance by sport for events by the opposing school. This will show the size of the impact that certain rival schools play in overall attendance levels.

Problem: Is there correlation between student attendance and general attendance levels?

- Correlation analysis of the attendance by sporting event from the 2014 – 2015 season to the 2016 – 2017 season by sport to examine if there is a relationship between general ticket sales and student attendance.

Problem: What proportion of the student crowd for each sport is male/female, international/non-international students, student-of-color (SOC)/non-student of color etc.? How does this compare with the overall student population at SCSU?

- Comparative analysis to examine the demographic make-up of student crowds at athletic events to determine which types of students are under/over represented in comparison to the overall SCSU population by sport by year/season.

Problem: What types of students do not attend sporting events? What proportion of students attend 0 events, attend 1 event, attend 2 events, etc.? How many students attend events from only one sport, from two sports, from three sports, etc.?

- Question 1: Comparative analysis of only the student population who do not attend sporting events during a particular year/season to gain insight to which demographic characteristics are most prevalent in the non-attending students.
- Questions 2 & 3: Distribution analysis of all students enrolled to decipher the number of students who attend 0 events, attend 1 event, attend 2 events, etc. and for those attending multiple events, understanding if they are attending different sports. This will shed insight to give a broad sense of the proportion of students who are involved with athletic event activities.

Problem: How much of a difference is there in attendance for events during winter break?

- Comparative analysis of student event attendance by winter sports for events, which occur during winter break to those, which do not occur during winter break. This will show the size of the impact that winter break plays in overall attendance levels.

Problem: Do multiple events on the same date affect overall attendance levels?

- Comparative analysis of both general ticket sales and student event attendance by sport for events occurring on the same date as another event. This will show the size of the impact that multiple events have in overall attendance levels.

Problem: Can we build a statistical model to predict the number of events/if a student will attend an event during a given season?

- Prepare a predictive model using the Fall 2014 – Spring 2017 undergraduate enrollment information containing student demographic information and the student attendance data to predict the probability a student will attend an event in 2018.

Data Manipulation and Merging:

A vast series of data manipulation procedures are required in order to conduct the appropriate analysis. The manipulation, cleaning, and joining procedures are outlined by each table and their respective procedures.

General Ticket Sales Data:

1. Data tables are separated by season (i.e. 2012 – 2013, 2014 – 2015 are two different tables.)
2. Combine all tables into one table in Excel.
3. Remove Rows in Excel with Aggregate Totals for events for all years:
    a. Delete top two rows. Select whole sheet. Unmerge all cells. Sort Ascending by InvoiceDate. Remove observations with total ticket sales and revenue.
4. 2017-2018 Ticket Data are all duplicate records in the 2016-2017 Ticket Data. There was an obvious mistake in the extraction. Do not use the 2017 – 2018 ticket sales data in analysis.
5. Create a formula column in jmp to extract first to letters of event code: Name it SportID
6. Subset observations with SportID equal to MH, WH, FB, BB, SC, WR, VB (52950 Observations)
7. Remove Playoff Events that remain (850 Observations)
    a. 52100 Observations Remain
8. Reformat InvoiceDate and EventDate to Date Format in jmp.
9. Remove Husky Hoopster Classic Event Observations
    a. 51266 Observations Remain
10. Save file as: Athletic_Ticket_Data_2012-2017.

Student Attendance Data:

1. Similar to the general ticket sales data, because the tables are separated by season, concatenate all years of the student attendance data.
2. Create a formula column in jmp to extract first to letters of event code: Name it SportID
3. Subset observations with SportID equal to MH, WH, FB, BB, SC, WR, VB.
    a. 43719 Observations Remain
4. Reformat InvoiceDate and EventDate to Date Format
5. Remove 32 Observations where Total Tickets Sold is not equal to one.
    a. 43687 Observations Remain
6. Remove 157 Quick Sell Observations
7. Remove 130 Observations with EventName equal to Women's & Men's Basketball vs. NSIC 1st Round
    a. 43400 Observations Remain
8. Rename Ticket Account # to TechID

9. Create Formula Variable Named YRTR_of_Event using logical comparison of the date to the beginning dates of a given semester.
10. Create a variable named Season, combining the events occurring during a particular school year.
    a. An event in 20153 or 20155 would be assigned 2014 season.
11. Remove observations for the events occurring in 20185 YRTR
    a. 42360 Observations Remain
12. Remove Husky Hoopster Classic Event Observations
    a. 41750 observations remain
13. Save file as: Student_Ticket_Data_2014-2018

Creating Ticket Sales by Event Data Table:

1. With the Student_Ticket_Enrollment_Records_Join data, using jmp, click tables > summary.
2. Group by: Event Code, SportID, Performer Name, Event Name, Event Date, Opponent and summarize: Sum(Invoice Amount), Sum(Total Tickets Sold), Sum(Total Ticket Revenue), Sum(Ticket Fees)
3. Save this table as: Ticket_Sales_by_Event

Creating Student Attendance by Event Data Table:

1. With the Student_Ticket_Enrollment data, using jmp, click tables > summary.
2. Group by: Event Code, SportID, Performer Name, Event Name, Event Date, Opponent and summarize by N rows. This N rows represent the attendance for a particular EventCode.
3. Save this table as: Ticket_Sales_by_Event

Student Attendance Data joined with Undergraduate Enrollment Data:

Join Tables Student_Ticket_Data_2014-2018 and 20180202 Undergrads Enrolled 20153 to 20183 in JMP:

1. Left outer join of Undergrads onto Ticket by TechID, remove duplicates from the undergrads table, create a match flag. Delete the columns of Total_Att_Credits, YRTR_Att_Credits, YRTR_GPA, and Cumulative_GPA. Name the table Temporary1.
2. Left outer join of Undergrads onto Temporary1 by TechID and YRTR_of_Event = YRTR_Of_Enrollment, update information of matching columns, create a match flag. Name Temporary2.
3. Left outer join of Undergrads onto Temporary2 by TechID and YRTR_of_Event = YRTR_Of_Enrollment. Delete duplicate columns except for Total_Att_Credits, YRTR_Att_Credits, YRTR_GPA, and Cumulative_GPA. Remove the “of Temporary 2” from the duplicate variable names.
4. Name this final table: Student_Ticket_Enrollment_Records_Join.

Creating Student Events by Semester Data:

1. Using the student attendance data, create new dummy variables for each SportID, coded as 1 if the event was for the particular SportID, 0 if not. (i.e. for the new formula column BB, if(SportID = "BB", 1, 0).
2. In jmp, click tables > summarize.
3. Group by TechID, YRTR_of_Event, and student demographic information. Summarize by the sum of each of the new formula columns created in step 1.
4. Create Table showing number of times each student has attended a particular sport each semester. (Students_Events_by_semester)
5. 15007 Records in this data set.

Undergraduate Enrollment Data joined with Student Events by Semester Data:

1. Left Outer Join Students_Events_by_semester onto Undergrads by Tech_ID and YRTR_of_Event = YRTR_of_Enrollment
2. 90826 observations in new data set.
    a. 13761 observations matched.
3. 14300 TechIDs in Undergrad Data missing demographic information on: QPP2, QPP2BaselineRetention, Honors, IntlFlag, SOCFlag, Female, ClosestToSCSU, MilesToSCSU, OutOfStateFlag, PellFlag, FirstGeneration, T1_Gap, T1_ResidenceHallFlag.
    a. These account for 23017 records in the Undergrad data from 20153 to 20183. These are either HS or non-degree seeking students. Exclude these from your analysis.
4. Int'l and SOC flags do not appear until Cohort 20113 students in this data set.

Appendix B: Events Without Student Attendance Records

Event	Date
Soccer vs. University of Sioux Falls	10/5/14
Soccer vs. Concordia University	10/24/14
Soccer vs. University of Minnesota-Crookston	10/31/14
Volleyball vs. Augustana	10/31/14
Women's Hockey vs. Wisconsin	11/22/14
Women's Hockey vs. Minnesota State	2/16/15
Women's & Men's Basketball vs. Minot State University	2/20/15
Soccer vs. St. Olaf (Exhibition)	8/21/15
Volleyball vs. Bemidji State University	9/17/15
Volleyball vs. Minot State University	10/2/15
Women's Basketball vs. South Dakota School of Mines	11/25/15
Men's Hockey vs. Minnesota-Duluth	2/27/16
Soccer vs. Wayne State College	10/23/16
Football vs. Minnesota Duluth	11/12/16
Women’s Hockey vs. Ohio State University	11/12/16

Appendix C: Types of Statistical Error, P-Values, and Power

Types of Statistical Error:

When statistician performs a hypothesis test, a decision is made whether to reject or not reject the null hypothesis in favor of an alternative. In either instance, there is potential the conclusion is incorrect. There are two types of incorrect conclusions a hypothesis test is susceptible to: false positive and false negative.

In the instance of a false positive, the null hypothesis is wrongly rejected in favor of the alternative. This is known in statistics as a type one error. The probability of making a type one error (α level) is predetermined before the test and is the p-value cutoff the statistician uses as a guideline whether to reject the null. To avoid type one errors, a statistician is advised to reduce the α level for a more conservative test. For instance, typically the α level is set low during medical trials involving the effectiveness of a particular drug with high side affects. This is to assure that the probability of incorrectly observing this result is low.

The second type of error is a false negative conclusion, which occurs if the null hypothesis is not rejected when the alternative is true. This false negative is referred to as a type two error. Unlike the probability of a type one error, the probability of a type two error cannot generally be calculated without knowing the population parameter of interest. The probability of a type two error is merely 1 minus the power. Typically the probability of performing a type two error decreases as the sample size increases. In the case of testing for effective drugs, a larger sample size will lower the probability of falsely not detecting a difference. This will further assure that the truly effective drugs will not go undetected.

It is important for statisticians to understand the impact a wrong decision will have on the business, health, etc. This is why determining and understanding type I and type II errors is essential before conducting analysis and reporting results.

P-Values:

P-values are often misused and misinterpreted, leading to many inaccurate results. The definition of a p-value is simply the probability, under the assumption that there is no difference, of collecting data that shows a difference equal to or more extreme than what was observed. A p-value is calculated under that assumption that there is no difference. The p-value only shows the probability the data is more extreme than the null hypothesis. A p-value does not show any meaning of effect size. For instance, in statistical hypothesis tests, p-values and the specified cutoff only tells whether or not there is a likely difference in means, proportions, etc. It does not tell how big the difference is, or give any other insight as to how to practically use the information.

P-values are misused in a variety of other ways such as determining significance in multiple comparison tests, selecting explanatory variables in models, and determining the false discovery rate of an experiment. Statisticians can follow a number of best practices to minimize misinterpretations and misuse of the p-value.

When performing a lot of multiple comparison tests for significance of regressors, procedures such as Bonferoni's method and the Benjamani-Hochberg method can adjust the p-value threshold to assure the false discovery rate does not increase. Another practice when determining

explanatory variables in models is to incorporate the effect size of a variable through the selection process. Although a p-value may be less than .05, the size of the effect could still explain a good chunk of the variability in the response. There is also the possibility of committing type II error if a variable is excluded from a model. This is also why determining power is important. It was shown during multiple presentations that p-values are very susceptible to change as the sample size varies. A statistician does not want to make the mistake of eliminating a predictor with a large effect size in the early stages of the analysis due to simply evaluating a p-value. As a general practice, statisticians should always report the effect size and the power of a test whenever reporting p-values.

Power:

Statistical power is inversely related to the probability of a type two error. A type two error is a false negative, or incorrectly failing to reject the null in favor of the alternative. Power is equal to 1 minus the probability of a type two error. Power is defined to be the probability of correctly rejecting the null when the alternative is true. In other words, if there is a true difference in means, proportions, etc., power is the probability we detect it. Power is an important calculation because it determines the probability we correctly detect the alternative hypothesis when it is true. Although power cannot be directly chosen like that of a p-value, it is possible to calculate using mathematics. Power calculation is dependent on three parameters: effect size, sample size, and the alpha level. Greater effect sizes are easier to detect and larger sample sizes decrease the test sensitivity. If any three out of these four parameters are known, the fourth can be calculated.

There are a couple of reasons why a power analysis is useful. The most frequent use of power is to determine the number of trials or the sample size to find a certain sized effect (if the power is

known). This is important as it tells the statistician how many trials must be run to avoid conducting a type two error. The second use is to actually calculate the power of a test. If you only have a certain sample size to work with what will the probability be that you actually detect the true alternative. If the power in this instance is low and the null is rejected, there is uncertainty that the result is true.

Calculating power is fairly difficult and often requires software. Because most of the data is already collected for analysis in STAT 381, the power analysis will be useful to determining the power of a given statistical test. The desired power level is often 80%.

Appendix D: Variable List of Original Data

General Ticket Sales Data

Variable	Description
Invoice Date	Date of Ticket Purchase
Account #	
Invoice Amount	
Performer Name	Name of Athletics Program
Event Code	Unique Identifier of Event
SportID	Abbreviation for Sport
Event Name	Name of the Athletic Event
Event Date	Date of the Event
Total Tickets Sold	
Total Ticket Revenue	
Ticket Fees	
Item Seat Summary	Summary of Seat Information
Ship Last Name	… of Buyer
Ship First Name	… of Buyer
Email Address	… of Buyer
Ship Address1	… of Buyer
Ship Address2	… of Buyer
Ship City	… of Buyer
Ship State Prov	… of Buyer
Ship Postal Code	… of Buyer
Ship Country Code	… of Buyer
Ship Phone 1	… of Buyer
Customer Comments	… of Buyer
Card Number Masked	… of Buyer
Opponent	Team Playing Against

Undergraduate Enrollment Data

Variable	Description
CohortYRTR	Year Term of First Semester at SCSU
TechID	8 Digit Student Identifier
Classification	FR, SO, JR, SR Identifier
YRTR_Of_Enrollment	Year Term of Semester

NEF	New-Entering Freshman Dummy Variable
QPP2	QPP of Term 2
QPP2BaselineRetention	QPP Baseline Retention of Term 2
Honors	Honors Student Indicator Dummy Variable
Age	Student Age during First Term
IntlFlag	International Student Indicator Dummy Variable
SOCFlag	Student of Color Indicator Dummy Variable
Female	Female Indicator Dummy Variable
ClosestToSCSU	Students Hometown Closest to SCSU of other University? Indicator
MilesToSCSU	Students Hometown Miles Away from SCSU
OutOfStateFlag	Out of State student Indicator Dummy Variable
PellFlag	Pell Eligible Indicator Dummy Variable
FirstGeneration	First Generation Indicator Dummy Variable
T1_Gap	Term 1 Financial Gap
T1_ResidenceHallFlag	Term 1 Residence Hall Indicator Dummy Variable
Total_Att_Credits	Total Attemtped Credits at SCSU
YRTR_Att_Credits	Year Term Total Attempted Credits at SCSU
YRTR_GPA	Year Term GPA
Cumulative_GPA	Cumulative GPA

Ticket Sales per Event Data

Variable	**Description**
Event Code	Unique Identifier of Event
SportID	Abbreviation for Sport
Performer Name	Name of Athletics Program
Event Name	Name of the Athletic Event
Event Date	Date of the Event
Opponent	Team Playing Against
N Rows	Number of Ticket Purchases
Sum(Invoice Amount)	Total Invoice Amount per Event
Sum(Total Tickets Sold)	Total Tickets Sold per Event
Sum(Total Ticket Revenue)	Total Ticket Revenue per Event
Sum(Ticket Fees)	Total Ticket Fees per Event
Season	Year of the Athletic Event Season

Student Attendance per Event Data

Variable	Description
YRTR_of_Event	Year & Term of the Semester
TechID	8 Digit Student Identifier
First/Last Name	… of student
BB	Number of Basketball Games Attended in semester
FB	Number of Football Games Attended in semester
MH	Number of Men's Hockey Games Attended in semester
SC	Number of Soccor Games Attended in semester
VB	Number of Volleyball Games Attended in semester
WH	Number of Womens Hockey Games Attended in semester
WR	Number of Wrestling Matches Attended in semester
Total Events	Number of Total Events Attended in semester

Weather Underground Data from St. Cloud Regional Airport

Variable	Description
Temp. (°F)-high	High Temperature on Day of Event
Temp. (°F)-avg	Average Temperature on Day of Event
Temp. (°F)-low	Low Temperature on Day of Event
Dew Point (°F)-high	High Dew Point on Day of Event
Precip. (in)-sum	Amount of Precipitation on Day of Event
Events	Snow, Rain, Fog, or Thunderstorm on Day of Event?
Snow	Dummy Variable if Snowed on Day of Event
Rain	Dummy Variable if Rained on Day of Event
Thunderstorm	Dummy Variable if Thunderstormed on Day of Event

Resources:

Walsh, P. (2016, March 3). In all-sport meeting, St. Cloud State athletes told 6 programs being eliminated. *Star Tribune.* Retrieved from http://www.startribune.com/st-cloud-state-dropping-6-sports-programs/370796061/