The Effect of Sociodemographic Factors on Attendance to Athletic Events by University of Mary Washington Students

Abstract

We do not wish to be considered for the opportunity to present our research at the end of the semester. We are researching what sociodemographic factors affect students' athletic event attendance at the University of Mary Washington. Using randomized data collection, SPSS and our knowledge gained in Applied Statistics and Research we were able to form hypotheses based on our input and output variables we felt played a role in student's ability to go to athletic events as well as testing variables that may or may not prove to hold value in this research. After randomly surveying a population of 100 UMW students we incorporated our collected data into SPSS and ran the appropriate tests based on the corresponding hypothesis. Although we did not find a strong amount of data to aid our research into this study, we did however prove that more athletes attend athletic events than non-athletes at UMW. This being said, there may be ways to use this information for later studies done by either future research groups, the athletic department and the department of student affairs to see why this proven data is true and use it to influence non-student athletes to support the athletic department and its events in the future.

Research Question and Justification

At the University of Mary Washington, attendance for the University's athletic events has never been high. Many say that our school is lacking spirit and enthusiasm compared to other schools such as James Madison University or University of Virginia. Since the University of Mary Washington is a DIII school, sports is not its number one priority, and many avid sports fans won't know who the athletes that the school is recruiting are. There won't be that excitement of going to see your favorite player, whom you've been following since their high school career play, like you would have at a DI school. While UMW may not fill a stadium, it does fill up a couple of rows and for our research project, we wanted to find out why this is. What affects a student's chance of attending athletic events. Since only a handful of students actually attend sports events, what are their sociodemographic factors? Are they all the same? Or do they come from different levels of involvement within the school. The Research question we posed to find this out is "What social demographic factors contribute to Mary Washington students level of attendance to on campus sports events." This research topic would not only help us understand what factors impact a student's chance of attending an athletic event but also help the athletic department better understand why the attendance of on campus athletic events are so low. Once the social demographic factors related to the support of athletic teams are identified they would be able to better promote to the students who attend less events than others.

Unit of analysis

The unit of analysis in research sciences is one of the most important parts of an experiment. The unit of analysis determines what we are going to be analyzing in our study. So for our research question; "What social demographic factors contribute to Mary Washington students level of attendance to on campus sports events" Our unit of analysis would be the Mary Washington student or better known as "the individual." We want to know what impacts an individual Mary Washington student to attend a sporting event at the University. If we chose Mary Washington as a whole to be our unit of analysis, then instead of studying what impacts an individual student, we would be studying what impacts Mary Washington as a whole institution to go to sports games. This is why the unit of analysis is the most important part of a research study, it determines what you will be analyzing.

Description of dataset:

We have 5 input variables in our research study. Two of these input variables are quantitative which means they are measurable. The first quantitative input variable is the "number of clubs a student on campus is involved in." This measures the construct of on campus student involvement. Our second quantitative input variable is "the number of hours a student logs at the gym in a week." We are using this input variable to measure the construct of the athletic appreciation, because we feel that the number of hours spent devoted to the gym correlates with the appreciation of others competing in athletic event. We have three categorical input variables. A categorical variable can not be measured. An example would be "gender" which is sorted into categories such as "female" or "male." Our categorical variables are whether the student is an athlete or not and would be answered in our survey by "athlete" or "non-athlete", whether the student is a "resident on campus" or a "commuter", and the students gender, "male" or "female". We have two output variables, which are the variables that change as a result of the input variables. We have one quantitative output variable, "Number of athletic events a student attends", and this measures the construct of school spirit. We also have one categorical output variable, "Whether or not the student considers themselves active on campus", and would be answered in our survey by "yes" or "no".

Input variables

Variable	**Nature**	**Construct Measured**
Number of clubs a student on campus is involved in	Quantitative	Campus student involvement
Number of hours logged at the gym in a week	Quantitative	Appreciation for the athletics
Whether the student is an athlete or not	Categorical	N/A
Whether the student is a commuter or resident	Categorical	N/A
Whether the student is a male or female	Categorical	N/A

Output variables

Variable	**Nature**	**Construct measured**
Number of athletic events a student attends	Quantitative	School spirit
Whether or not the student considers themselves active on campus	Categorical	N/A

Sampling method

Before we are able to find out what are sample size is, we need to determine what our target population is and the sample we need. Since our research question is asking about the average University of Mary Washington student, our population would be everyone that would fit into that demographic, which is approximately 4,000 undergraduate University of Mary Washington students. The sample size we chose to go with was 100 students that were randomly selected. To get the most accurate representation of the University of Mary Washington student population, we had to ensure that we were sampling students at complete randomness.

We first listed all the locations that generates the most foot traffic. Locations included the University Center, the library, the Hurley Convergence Center and the fountain, as well as various academic and residence hall locations. We then listed different sets of times, and different dates. We put the different locations, different times, and different dates into three separate buckets and each picked a location, time, and date. To get our sample of 100 Mary Washington students, we each had to sample 33 students.

Once at our location, we used a random number generator from excel that gave us a random set of 33 numbers and we used those numbers to choose which student to survey. For example; the first number generated was 5, so we surveyed the 5th student that walked out of Monroe, the next number was 2, so we waited till the 2nd person after the 5th person to walk out and surveyed them. This sampling method insured that we would get the most random set of students at the University of Mary Washington that could accurately represent the UMW population.

Hypotheses

Our first hypothesis chosen is a simple linear regression test. We used a quantitative input variable of "Number of clubs a student on campus is involved in" which measured the construct of student involvement on campus. We also used a quantitative output variable of "number of athletic events a student attends." This measured the construct of school spirit. Our hypothesis was built through logical expectation. It is expected that the more clubs you are a part of, the more you are involved on campus, and the more involved an individual is the more likely it is that they will attend a University of Mary Washington athletic event. Our null hypothesis was "There is no relationship between the number of clubs students at Mary Washington are involved in and the number of athletic events they attend" and our alternative hypothesis was "There is a relationship between the number of clubs students at Mary Washington are involved in and the number of athletic events they attend."

Our second hypothesis was a 2-sample T-Test. We had a categorical input variable of whether or not the student is an athlete or not and a quantitative output variable of the number of athletic events a student attends. Our quantitative output variable is measured by the construct of school spirit. This hypothesis was built on empirical observation. When you attend a sports game, it is easy to observe that most of the crowd in attendance is fellow athletes. Our null Hypothesis was "The true (population) mean of athletes that attend athletic events is equal to the true (population) mean of non-athletes that attend athletic events" and our alternative Hypothesis is “The true (population) mean of athletes that attend athletic events is greater than the true (population) mean of non-athletes that attend athletic events.”

Our third hypothesis we tested is also a 2-sample T-test. We had a categorical input variable which was whether or not the student is a commuter or resident. A resident being a student who lives in any of the University of Mary Washington housing options on campus, and a commuter being any student who lives anywhere that is not one of Mary Washington’s housing options. Our output variable was the number of athletic events a student attends within a school year (two semesters). This is a quantitative variable with the construct measured being school spirit. This hypothesis was built on logical expectation, because you would expect more residential students to attend more athletic events due to their proximity to the field rather than commuters who would most likely have to travel back to campus to attend the athletic events since most events take place in the evening. Our null hypothesis is “The true (population) mean of commuters that attend athletic events is equal to the true (population) mean of residents that attend athletic events” and our alternative hypothesis is “The true (population) mean of commuters that attend athletic events is less than the true (population) mean of residents that attend athletic events.”

Our fourth hypothesis test is a 2 proportion Z-Test. We had a categorical input variable with two selected categories. Our variable was whether or not the student is an athlete or not. The categories being “athlete” and “non-athlete” and we had an output categorical variable. The variable being whether or not the student considers themselves active on campus. This hypothesis was built on logical expectation. We expect that because if the student is an athlete they are more likely to consider themselves active on campus then a non-student-athlete.

Our null hypothesis is “The true (population) proportion of student athletes that consider themselves active on campus equals the true (population) proportion of non-student-athletes that consider themselves active on campus.” Our Alternative Hypothesis is “the true (population) proportion of student athletes that consider themselves active on campus is greater than the true (population) proportion of non-student-athletes that consider themselves active on campus.”

Our fifth hypothesis is the Chi Square Test. We had a categorical input variable, the variable being the student's gender. We also had a categorical output variable, the variable being whether or not the student considered themselves active on campus. One of the difference between the 2 proportion Z-Test and the Chi Square Test is that the categorical output variable in the Chi Square Test needs to have a minimum of 2 selected categories, unlike the Z test which only needed 1 selected category. The two categories being “active” and “not active.” This hypothesis was built through empirical observation, with more females being involved in more student clubs and organizations, we assume that the gender is related to whether or not they consider themselves active. Our null hypothesis is “A student's gender is not related to whether or not they consider themselves active on campus” and our alternative hypothesis is “A student’s gender is related to whether or not they consider themselves active on campus.”

We also have a bonus simple linear regression hypothesis. We had one last quantitative input variable of “Number of hours logged at the gym” and a quantitative output variable of “The number of athletic events at the University a student attends.” Our Null hypothesis is that “there is no relationship between the number of hours logged at the gym in a week, and the number of athletic events a student attends” Our alternative hypothesis is that “There is a relationship between the number of hours logged at the gym in a week and the number of athletic events a student attends.”

This is hypothesis was built through logical expectation. If a student spends many hours in the gym either shaping their own performance in a sport (such as pickup basketball) or simply going to feel healthy and accomplished; they will be more likely to appreciate the amount of hours a student-athlete spends perfecting their sport and showing their hard work paying off in competition.

Statistical tests

Our first and last statistical test we are applying in order to test two of our hypothesis is Simple Linear Regression. In simple linear regression, our input and output variables both have to be quantitative. For the first test we chose the input variable to be "the number of clubs a student is involved in", and our output variable, "the number of athletic events a student attends." Before we use linear regression, we have to make sure that certain conditions are met for the regression analysis to be valid. The conditions are; the observations must be independent, the relationship between X and Y should be linear, and the distribution of Y around its mean should be normally distributed. The last condition is going to have a Null hypothesis that states "The residuals are normally distributed" and an alternative hypothesis "The residuals are not normally distributed." If the P value of our tests exceeds the chosen significance level (α), we will have no reason to believe that the residuals are not normally distributed, and we will not reject the null hypothesis. A significance level (α) is the maximum amount of probability that is derived from a random sample that is just up to chance (also known as the P-Value) that you are willing to tolerate. If the P value is less than the chosen significance level, then you have enough evidence to reject the null hypothesis in favor of the alternative hypothesis. If all these conditions are met, we will be able to carry on the simple linear regression analysis that will tell us whether there is a relationship between the number of clubs a student is involved in, and the number of athletic events they attend.

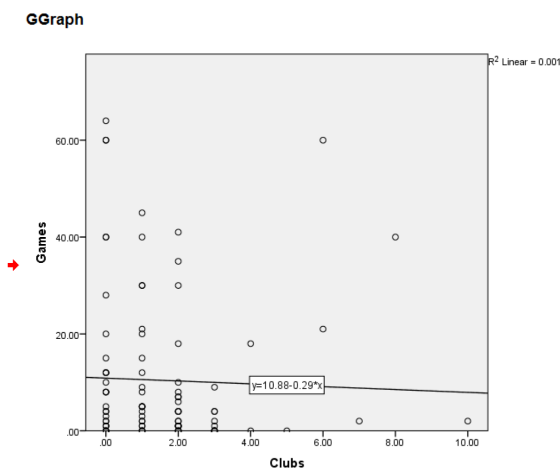
Our second statistical test is the 2 sample T-test. This is used to test our second and third hypothesis. The 2 sample t-test utilizes a categorical input variable with two selected categories. In our case, it would be whether the student is an athlete or not, and whether the student is a commuter or resident. It also utilizes a quantitative output variable. In both hypothesis, we used the output variable of “athletic events attended.” The 2 sample t-test is testing the difference of means between the two categories of out input variables.

Our third statistical test is the 2 proportion Z-test. This test is used to test our fourth hypothesis. The test utilizes categorical input and output variables. The input variable needs to have two selected categories. Our categories are student athlete, or non student athlete, with our output being whether they consider themselves active on campus. This test will tell us if one portion of the population (i.e student athletes) consider themselves more active than another portion of the student population (i.e non-student athletes.)

Our fourth and last statistical test we use to test our last hypothesis is the Chi-Square test. Just like the 2 proportion Z test, the Chi-Square test utilizes categorical input and output variables. The difference being that the input variable needs 0 to at least 2 selected categories, and the output needs a minimum of two selected categories. We used the input variable of student gender, whether they consider themselves active on campus. This test will help us to figure out if a student’s gender is in fact related to whether or not they consider themselves active on campus.

Results

For our first hypothesis, we were trying to find out whether there is a relationship between the number of clubs a student is involved in, and the number of sports events a student attends. Since our statistical test was a simple linear regression, we have to make sure the conditions of regression are satisfied. The first condition is that the observations must be independent. This condition was met because the occurrence of number of clubs does not change the probability of number of athletic events. The second condition that we needed to meet was if the relationship between X and Y should be linear. To determine this, we generated a scatter plot and drew a best fit line.



We can conclude, based on the above graph that the relationship between X and Y is in fact not a linear relationship. The third condition is to find if the residuals are normally distributed.

Our Null hypothesis for this condition is "The residuals are normally distributed" and our alternative hypothesis "The residuals are not normally distributed."

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.247	99	.000	.698	99	.000

a. Lilliefors Significance Correction

There are two tests that we can use to test the normality of our residuals. If our sample size is less than 2000, we use the Shapiro-Wilk's test, and if it's over 2000, we use the Kolmogorov-Smirnov test. Since our sample size is only 100, we will be using the Shapiro-Wilk's test. Our P-value, under the Shapiro-Wilk's section is found under "sig" at .000. If our P-value exceeds our chosen significance level of .05, we have no reason to believe that our residuals are not normally distributed (i.e, we do not reject the null hypothesis.) Since our P-Value is less than our significance level, we can reject the null hypothesis and conclude that the residuals are not normally distributed. Since the second and third condition were failed to be met, the results of our regression analysis are not going to be valid, but we went ahead and did it anyway. What we wanted to find out with this test is if it is reasonable to believe that there is a true relationship between the number of club's students at Mary Washington are involved in and the number of athletic events they attend. If there is a true relationship, then the true slope of the regression best-fit line, should be non-zero. (i.e., $\beta 1^{1}$ 0)

Coefficients[a]

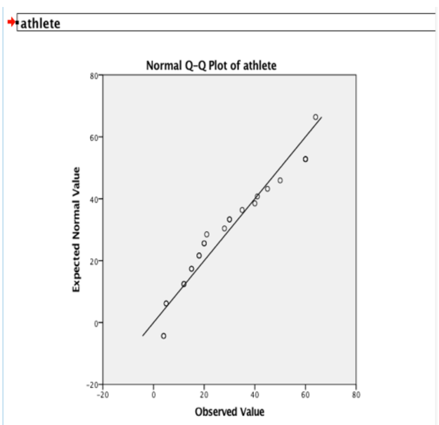
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10.879	2.125		5.118	.000
	clubs	-.294	.874	-.034	-.337	.737

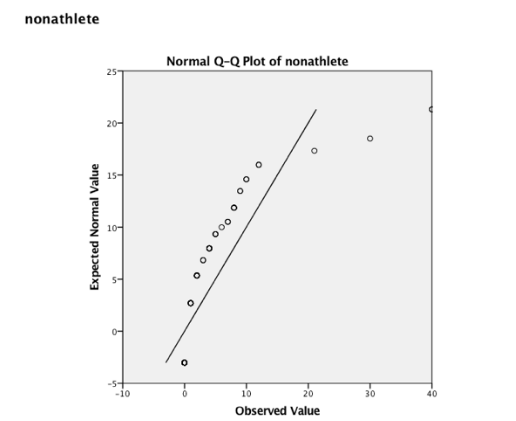
a. Dependent Variable: events

Under the column Sig, the bottom value of .737 is our P-value. Since our P-value of 73% is greater significance level of 5% we can not reject the null hypothesis in favor of the alternative hypothesis

For our second hypothesis, we were trying to find whether one portion of student population attends more athletic events then another portion of the student population so we chose the 2 sample T-Test. We want to test the difference in the population means between student-athletes and non student-athletes is going to be positive.

Our hypothesis that we are testing are: Null hypothesis: The true (population) mean of athletes that attend athletic events is equal to the true (population) mean of non-athletes that attend athletic events. Alternative Hypothesis: The true (population) mean of athletes that attend athletic events is greater than the true (population) mean of non-athletes that attend athletic events. Let $\mu 1$, $\mu 2$ be the population means for athletes and non athletes respectively. And the Significance level $\alpha = 0.05$ The hypothesis test in this case is right-tailed: $H0: \mu 1 - \mu 2 = 0$ $Ha: \mu 1 - \mu 2 > 0$. We first must test for normality of the data. If the points are somewhat aligned along the straight line drawn, we can assume normality and carry on with our hypothesis test.





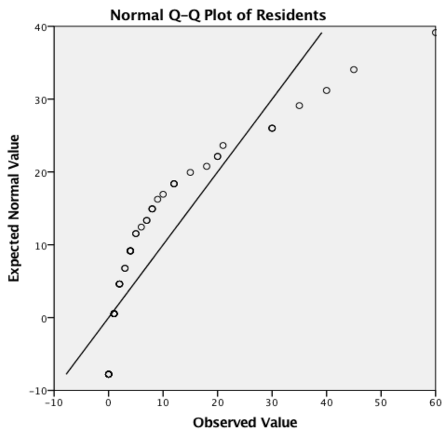
In both cases of athlete and non athlete, the Q-Q plot shows a linear pattern with the data falling along the straight line drawn. Since we can assume normality in both cases, we can continue to carry out our 2-sample t-test.

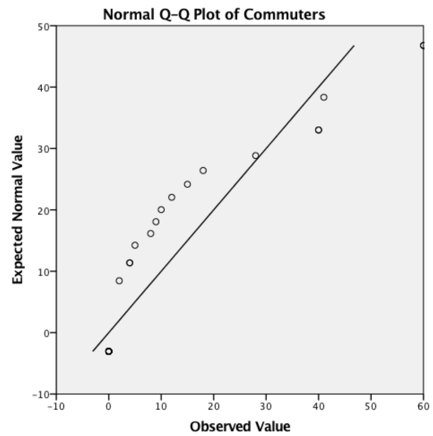
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
GamesAttended	Equal variances assumed	44.547	.000	-8.780	96	.000	-24.001	2.73353	-29.427	-18.575
	Equal variances not assumed			-6.049	26.754	.000	-24.001	3.96787	-32.145	-15.856

In the table above, under “Equal Variances Not Assumed” we have our t-statistic of -6.049. Our P-value is found under Sig (2-tailed) at .000. Because this is a right-tailed hypothesis test, and SPSS only carries two-tailed test. We need to divide our P-value by half. So our correct P-value is 0. If our P-value is less than our significance level, we have enough evidence to reject the null hypothesis. If our P-value is greater than our significance level, we fail to reject the null hypothesis because we do not have enough evidence to do so. Since our P-value is 0 and our significance level is .05 we have enough evidence to reject the null the hypothesis. At a 5% significance, we can conclude that the true population mean of student athletes that attend sports events is greater than non-student athletes that attend sports events.

For our third hypothesis, we were trying to find whether one portion of student population attends more athletic events then another portion of the student population so we chose the 2 sample T-Test. We want to test that the difference in the population means between commuters and residence is going to be positive. The hypothesis that we were testing are:
Null hypothesis: The true (population) mean of commuters that attend athletic events is equal to the true (population) mean of residents that attend athletic events. Alternative Hypothesis: The true (population) mean of commuters that attend athletic events is less than the true (population) mean of residents that attend athletic events. Let $\mu 1$, $\mu 2$ be the population means for commuters and residents respectively. And the Significance level $\alpha = 0.05$. The hypothesis test in this case is left-tailed: H0: $\mu 1 - \mu 2 = 0$ Ha: $\mu 1 - \mu 2 < 0$. We first must test for normality of the data. If the points are somewhat aligned along the straight line drawn, we can assume normality and carry on with our hypothesis test.





In both cases of athlete and non athlete, the Q-Q plot shows a linear pattern with the data falling along the straight line drawn. Since we can assume normality in both cases, we can continue to carry out our 2-sample t-test.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
AthleticEvents	Equal variances assumed	5.743	.019	1.308	78	.195	4.78545	3.65895	-2.4990	12.0699
	Equal variances not assumed			1.136	34.477	.264	4.78545	4.21184	-3.7697	13.3406

In the table above, under "Equal Variances not Assumed" we have our t-statistic of 1.136. Our P-value is found under Sig (2-tailed) at .264. Because this is a left-tailed hypothesis test, and SPSS only carries two-tailed tests, we need to divide our P-value by half. So our correct P-value is .132. If our P-value is less than our significance level, we have enough evidence to reject the null hypothesis. If our P-value is greater than our significance level, we fail to reject the null hypothesis because we do not have enough evidence to do so. Since our P-value is 13% and our significance level is 5% we do not have enough evidence to reject the null the hypothesis.

At a 5% significance, we can NOT conclude that the true population mean of commuter students that attend sports events is less than residential students that attend sports events.

For our fourth hypothesis we were trying to figure out whether a student athlete that considers themselves active on campus is greater than non student athlete that considers themselves active on campus. To do this we used a 2 proportion Z test. Our Hypothesis were: Null Hypothesis: The true (population) proportion, p1, true of student athletes that consider themselves active on campus equals the true (population) proportion, p2, true of non-student-athletes that consider themselves active on campus. Alternative Hypothesis: The true (population) proportion of student athletes that consider themselves active on campus is greater than the true (population) proportion of non-student-athletes that consider themselves active on campus. This hypothesis test in this case is right-tailed H0: p1,true = p2,true Ha: p1,true > p2,true. Unlike the 2 sample t-test, for the 2-proportion Z test, we do not need to test for normality.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.499[a]	1	.221		
Continuity Correction[b]	.742	1	.389		
Likelihood Ratio	1.525	1	.217		
Fisher's Exact Test				.289	.195
N of Valid Cases	32				

In the table above, under the column "exact sig. (2-sided)" we have our chi-squared P-value which is .289. But since it is a right-tailed test, and SPSS only does two-tailed, we have to divide our P-value by 2. So our adjusted P-value is 14%. With a P-value of 14% and a significance level of 5% we do not have enough evidence to reject the null the hypothesis. At a 5% significance level, we can NOT conclude student-athletes that consider themselves active on campus is greater than non student athletes.

Our Fifth hypothesis is testing whether a student's gender is related to how much they consider themselves active on campus. We tested this using the Chi-Square statistical test.

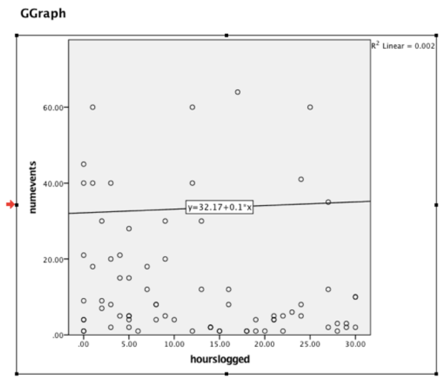
Our hypothesis were: Null Hypothesis: A student's gender is not related to whether or not they consider themselves active on campus Alternative hypothesis: A student's gender is related to whether or not they consider themselves active on campus. With our data loaded into SPSS, we get the following table:

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.214[a]	1	.644		
Continuity Correction[b]	.065	1	.799		
Likelihood Ratio	.213	1	.644		
Fisher's Exact Test				.681	.399
N of Valid Cases	99				

In the table above, under Asypm. Sig (2-sided) column in the Pearson Chi-Square row we have our chi-square P-value of .644. Since our P-value is more than our α of 5% we do not have enough evidence to reject the null hypothesis in favor of the alternative hypothesis. We can therefore conclude that a student's gender is not related to whether or not they consider themselves active on campus.

For our last bonus hypothesis, we were trying to find out whether there is a relationship between the number of hours logged at the gym, and the number of sports events a student attends. Since our statistical test was a simple linear regression, we have to make sure the conditions of regression are satisfied. The first condition is that the observations must be independent. This condition was met because the of number of hours logged in the gym does not change the probability of number of athletic events. The second condition that we needed to meet was if the relationship between X and Y should be linear. To determine this, we generated a scatter plot and drew a best fit line.



Based on the above graph, we can conclude that the relationship between X and Y is in fact not a linear relationship. There is too much scattering in the data points. but like our first hypothesis which was also simple linear regression with conditions not met, we decided to continue on with the tests even though our results will not be as valid. The third condition is to find if the residuals are normally distributed. Our Null hypothesis for this condition is "The residuals are normally distributed" and our alternative hypothesis "The residuals are not normally distributed."

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.137	1031	.000	.928	1031	.000

a. Lilliefors Significance Correction

In this case, we would want a high P-value to be able to fail to reject the null hypothesis because we do want our residuals to be normally distributed. Since our sample size is lower than 2000, we will be using the Shapiro-Wilk's test. Our P-value in this case is .000 which is lower than our alpha, so we can reject the null hypothesis. This means that are residuals are not normally distributed. Since our sample size is lower than 2000, we will be using the Shapiro-Wilk's test. Our P-value in this case is .000 which is lower than our alpha, so we can reject the null hypothesis. This means that are residuals are not normally distributed. Our alternative hypothesis is that "there is a relationship between number of hours logged at the gym, and number of sporting events attended." For this to be true, we need our P-value to be lower than our significance level of .05.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	32.168	.926		34.730	.000
	hourslogged	.096	.066	.045	1.452	.147

a. Dependent Variable: numevents

Based on the table above, under the "Sig" column, our P-value is greater than .05, at .147. Since our P-value is greater than our significance level, we fail to reject the null hypothesis that there no relationship between number of hours logged at the gym and sports events attended.

Analysis and discussion

Our findings represent that the factors influencing whether or not a student at Mary Washington attends an athletic event may not be as predictable as we were expecting. While forming our hypothesis we expected many factors to play a role in higher or lower attendance of athletic events such as involvement on campus, athlete or non-athlete, commuter or resident and the number of games available for a student to attend. After our research we found that these factors did not play as big of a role as we were expecting except that athletes are more likely to attend than non-athletes. This would be an important finding for the athletic department to look at as well as student affairs to evaluate what students are attending the events and how to better gauge advertisements to those who are not attending. These departments want to bring out as many students to the games as possible not only for support but to increase concession sales and have fundraisers be successful. Since we concluded that non-athletes are less likely to attend games than athletes, the athletic department and student affairs should use this research to reach more non-athletes through advertisements, promotions and increased "game day" environment to raise the attendance numbers at home games. Our least relevant finding was number of clubs a student is involved in, having the highest p-value of .737 and then gender having a p-value of

.644. The third and fourth least relevant findings were student-athletes' perception of their involvement on campus versus non-athletes' perception of their involvement on campus, having a p-value of .14 and then whether a student is a commuter or resident, having a p-value of .13. Our most relevant finding was whether or not a student is an athlete, having a p-value of 0. Even though we did not conclude that being a commuter or resident had an affect on whether or not a student attends an athletic event is was our second most relevant and we feel that it plays the second biggest role in attendance of athletic events based on the mere proximity of a student to the event.

Conclusion

After conducting our research, we were able to prove that the number of student-athletes that attend athletic events at UMW was greater than the number of non-student athletes attending the events. We were not surprised by this because as a group we felt that it would be more likely for athletes to support their fellow athletes and to stay at an event following one of their own than it would be for non-athletes to take time out of their academic schedule to support the athletic program. This being said, another factor we considered was the greater knowledge of games prior to the actual game day for student-athletes via fellow student-athletes, coaches, and the athletic department as a whole. One of the more interesting results that surprised us was the fact that we could not prove that residential students attend more games than commuter students. This stood out to us because we believed there would be a greater attendance of residential students at home games given that their commute to a game would be much shorter than that of a student commuting. One of the outliers though that did not go into consideration when we were testing between residential and commuter students is that some of the commuter students are athletes as well so they would have a more likely chance to attend a sporting event than a non-student athlete that is also a resident. Overall, while we were not able to reject all the null hypothesis, we are still satisfied with the results we collected.

Assumptions, limitations and future research

There are many assumptions and limitations to this research project. As a group we can assume that your average UMW student is willing and able to attend at least one athletic event due to a variety of limitations. We are also assuming that the student is upholding the Mary Washington honor code and not lying while taking the survey. The sampling we have chosen poses the greatest reducer of limitations possibly faced during this study, but there are still errors or lurking variables that could happen during the experiment that could potentially impact the data and our research. Students not willing to answer to stop and answer the survey questions is one of our limitations. Another possible limitation of our study would be the mere popularity of certain sports over other sports. For example, UMW students are more likely to go to the well advertised basketball games opposed to the rugby games that very little people hear about. We would need to take into consideration the average amount of people that attend a particular athletic event when determining how populated it is. Another possible limitation we feel we should consider is the availability of home games for athletic events. If there are few home games more students might attend because there will be limited times they could go, but if there are several home games the attendance to games may be more spread out meaning lower numbers of students at athletic events. If we had more time, we would have pooled a larger sample size. We were limited to very little time and resources and had to limit our sample to only a 100 students. Ideally, we would have sampled approximately 3,000 students since our population is around 4,000 and this could potentially bring our results to a more reliable standard.