Team and School Characteristics that Affect the Average Attendance of NCAA Basketball Home Games

Jayce Davis

Written for ECON 4260; Dr. Kim Craft

Team and School Characteristics that Affect the Average Attendance of NCAA Basketball Home Games

Introduction

Within the schools with men's basketball teams categorized as Division 1 teams by the NCAA, average attendance per game in the 2015-16 season varied from 436 at the lowest (Bryant University, Northeast Conference) to 23,361 (University of Kentucky, Southeastern Conference). There are many characteristics of both the team and the university or college that cause this wide range in their average attendances. The objective of this project is to explore some of the differences between universities and colleges that cause the disparities in the average attendance of their men's basketball games at home. In this study, I analyze the effects of some variables that represent the team's current success, and some measure of their historical performance. I also include the size of the school, and other variables that I believe are good indicators of the types of things that influence potential attendees' choices.

Variables for Analysis

The theory I used in choosing variables that I felt would be important in modeling average attendance at collegiate basketball games was to identify the variables that determine demand in economic theory, such as price, cross price, number of buyers, expectations, and buyer tastes, and try to quantify them in a meaningful way. Most of those determinants of demand are difficult to use for this analysis the way I have structured it. The majority of the variables I have included in this model are those that I predicted would influence buyer taste, and the focus of my study is to find how each of these affect average game attendance.

The first variable that I included in the model was the population of the student body of the school. This student body is a large part of the pool from which school's basketball teams draw their potential fans and viewers. A larger student body should result in a higher average attendance, with more students to come to each game. A larger student body may also be associated with part of how people view the school. Larger, more well-known schools probably exhibit some “brand name” cachet that draws fans and therefore drives attendance. This variable was included in logarithmic form, as is usual for this type of data.

Percentage of games won at the end of the 2015-2016 season is included in my study. I predict that during seasons that the team is doing well, more fans and students will want to come out to see their teams, so this coefficient on this variable should be positive. I also included percentage of games won at the end of the previous season in my analysis. It seems reasonable to assume that teams coming off of winning seasons will have more viewership in the season following.

I also predicted that the NCAA conference a team is in will influence their average attendance. During the season teams play a large proportion of their games against other teams in their own conferences. Because better teams are brought into better conferences, stronger conferences will have games that cause more fan excitement. A team in a strong conference that is experiencing average success or worse is still going to draw more attention than another school from a more obscure conference.

In the first stages of my analysis, I attempted to use binary variables for all of the conferences in Division 1 basketball. The first problem with that approach was simply too many variables; there are 32 conferences, and Excel will only regress a maximum of 16 independent variables. Although I did use another statistical software program to do this regression initially, I felt that having that many variables was going to cause other problems that I didn’t want to deal with, such as a high degree of multicollinearity in the explanatory variables, and a lot of degrees of freedom being lost due to variables that I could probably consolidate in a meaningful way. I found two articles from last season that contained lists that the writers, who were sports analysts, said were the most competitive conferences that year; one article was found in Sports Illustrated and one from the NCAA's website. I then created a binary variable to indicate if the team was a part of one of the conferences that were rated most competitive in the NCAA article. Those conferences were the Big 12, Atlantic Coast Conference (ACC), Big Ten, Big East, Southeastern Conference (SEC), Pac-12, American Athletic Conference (AAC), West Coast Conference, and Missouri Valley Conference.

The final variable I included was recent appearances in the NCAA tournament. My theory in using this variable was that if the team has been able to play in the tournament in recent years, it would be a sign that the team has a tradition of winning that promotes fan loyalty. Making it into the tournament also a sign that the team is going to continue to do well in the immediate future, which will increase attendance.

There are a number of significant variables that ought to be included in a model such as this one, but I have left them out for various reasons. The most obvious one is price. Price is a major determinant of demand that, if it had been available, could possibly have enhanced the results of my model. Price data for this project would have been difficult at best or nearly impossible to find for several reasons. One is that in many college's stadiums, mainly those with larger attendance, ticket prices vary by seat, and occasionally may even vary from game to game if a particularly exciting match-up is in the team's home schedule. In addition, the initial price at which the tickets were distributed may not reflect accurately the true market price of attending a game. Calculating some kind of average to standardize these variations between teams would have been very hard. Finding the data to even do such a calculation would have been tedious and likely would have resulted in some kind of error in collecting and calculating, provided this data for each team even exists.

Another variable that I have omitted that is likely important is the population of the area in which the college is located. It would have also been difficult to find population data for each college's location that is consistent. A major consideration in attempting to include this variable would be the actual region from which the school draws local crowds. Comparing just the schools in Utah, we can assume that Southern Utah University pulls its basketball attendees from Cedar City only or Iron County generally. By contrast, Logan is only a small part of the area from which Utah State University's viewers may come. The 2010 census showed the population of the city of Logan to be only 48,174, but the Logan metropolitan area was 125,442 at the time, including residents from Cache County, Utah and Franklin County, Idaho. (United States Census Bureau, American Fact Finder) It would be difficult to go through each school to determine for which ones consideration of a larger metropolitan area would be appropriate, and what part of that area ought to be considered the pool from which potential attendees are drawn.

This table provides a summary of the variables used and their sources:

TotAttendance	The total attendance of home games for each team in the 2015-16 season. (This variable is not part of my analysis, but is provided in the table of summary statistics.)
Attend	Average attendance per game for each team in the 2015-16 season. (found on NCAA website at http://www.ncaa.org/championships/statistics/ncaa-mens-basketball-attendance)
ln(Attend)	The natural logarithm of a team's average attendance.
Pct16	The proportion (in percentages) of games won by each team in the 2015-16 season. (found on NCAA website at http://web1.ncaa.org/stats/StatsSrv/rankings?doWhat=archive&rpt=archive&sportCode=MBB)
Pct15	The proportion (in percentages) of games won by each team in the 2014-15 season. (found on NCAA website at http://web1.ncaa.org/stats/StatsSrv/rankings?doWhat=archive&rpt=archive&sportCode=MBB)

Appear2015, Appear2014…	A binary variable representing a team's appearance in the NCAA basketball tournament in 2015, 2014, and so on. (found in the NCAA "Final Four Record Book" at http://fs.ncaa.org/Docs/stats/m_final4/2017/TournField.pdf)
Appear(6yr)	The count of a team's appearances in the NCAA tournament for the last six years.
CompConf(NCAA9)	A binary variable signifying whether the team is part of a conference the article on the NCAA website listed as one of the most competitive conferences for the season. (article url: http://www.ncaa.com/news/basketball-men/article/2016-01-13/college-basketball-these-are-9-strongest-conferences-so-far)
ComfConf(SI5)	A binary variable signifying whether the team is part of a conference the article on the Sports Illustrated website listed as one of the most competitive conferences for the season (article url: http://www.si.com/college-basketball/2016/04/22/best-conference-acc-big-12-big-east-big-ten
Student Population	The 12-month student enrollment for each institution in the 2014-15 academic year. (from the National Center for Education Statistics at https://nces.ed.gov/ipeds/datacenter/Default.aspx)
ln(Pop)	The natural logarithm of Student Population.

The following is a table of summary statistics:

	TotAttendance	*Attend*	*Pct16*	*Pct15*	*Appear(6yr)*	*Student Population*
Average	71639.24855	4417.645	51.07168	51.1289	1.170520231	15541.93931
St.Dev.	76495.85487	4329.503	17.33497	17.20222	1.599763847	11554.66067
Median	35,874	2,501	12,931	1	52	52
Max	397148	23361	87.5	97.4	6	66137
Min	4834	436	12.5	6.7	0	1223
Range	392314	22925	75	90.7	6	64914

Statistical Analysis

I began my analysis with an OLS regression of average attendance as a function of *Pct16*, *ln(Pop)*, *CompConf(NCAA9)*, and *Appear* in the years 2010-2015[1].

$$Attend = \beta_0 + \beta_1 Pct16 + \beta_2 \ln(Pop) + \beta_3 CompConf(NCAA9) + \beta_4 Appear2015 + \beta_5 Appear2014 + \beta_6 Appear2013 + \beta_7 Appear2012 + \beta_8 Appear2011 + \beta_9 Appear2010 + u$$

The results of this model were basically how I expected; all the estimates for the coefficients were positive, meaning that each had a positive impact on average home game attendance. All were statistically significant but Appear2013 and Appear2011 at the .05 level of significance as well. At this point I tested this model for heteroskedasticity using the Bruesch-Pagan test (Wooldridge, pg 273). I regressed the squared residuals of my estimated model on each of the independent variables, as follows[2]:

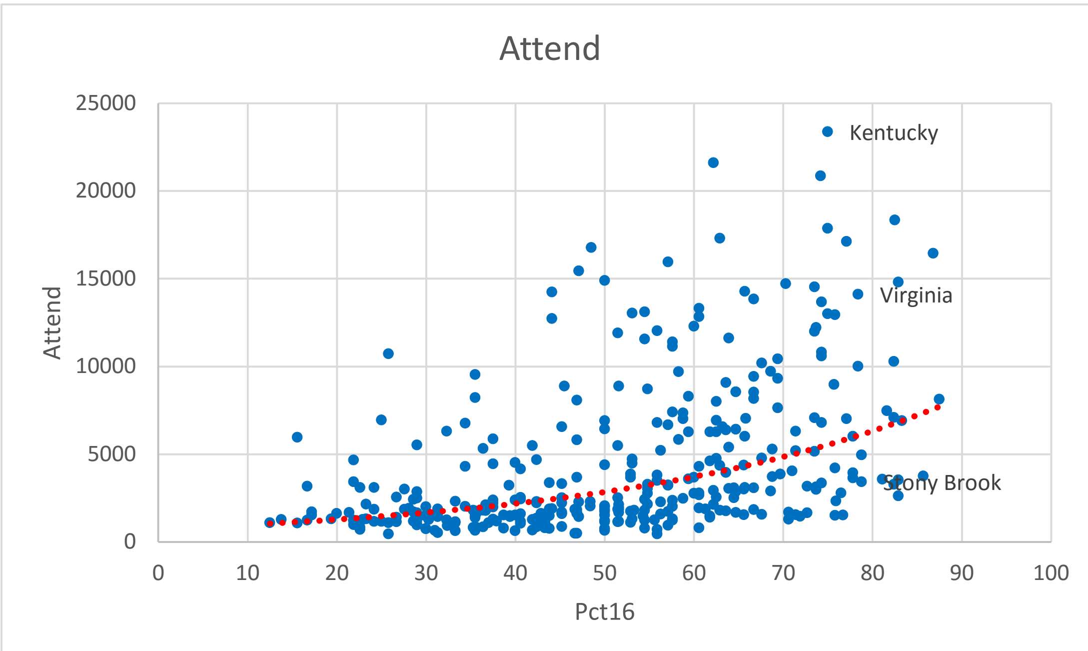
$$*\,\hat{u}^2 = \delta_0 + \delta_1 Pct16 + \delta_2 \ln(Pop) + \delta_3 CompConf(NCAA9) + \delta_4 Appear2015 + \delta_5 Appear2014 + \delta_6 Appear2013 + \delta_7 Appear2012 + \delta_8 Appear2011 + \delta_9 Appear2010 + error$$

After estimating this equation, I tested the null hypothesis, that

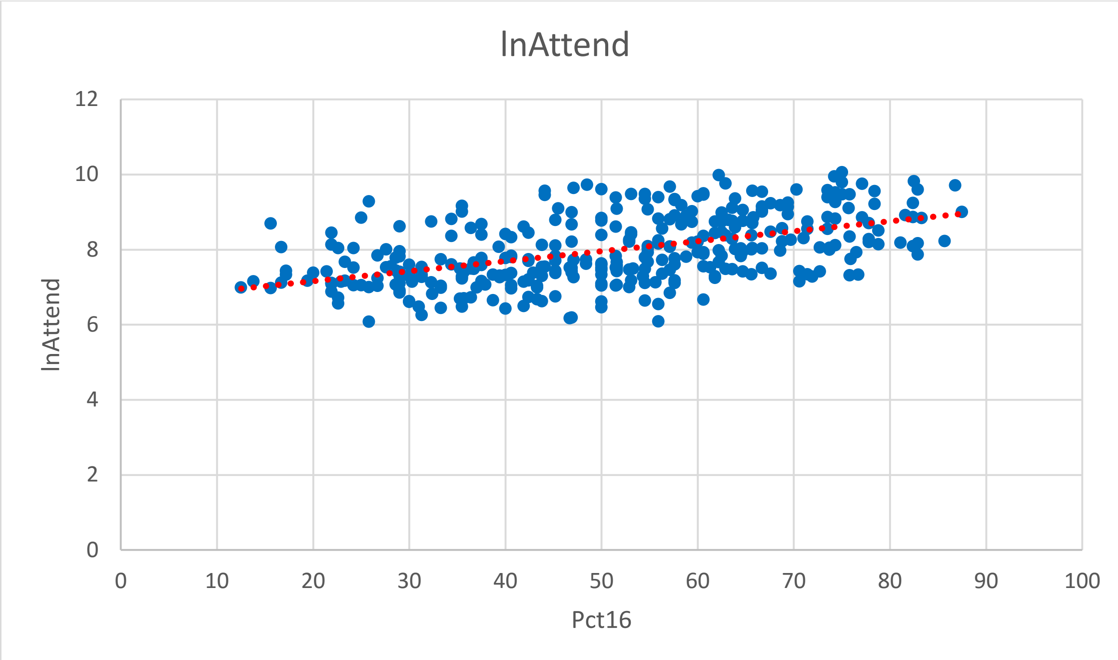
$$\delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = \delta_7 = \delta_8 = \delta_9 = 0$$

or, that the model is homoskedastic, that the distribution of the error term is not dependent on the independent variables.

I used an F statistic to test this hypothesis, where $F = \frac{R^2_{\hat{u}^2}/k}{(1-R^2_{\hat{u}^2})/(n-k-1)}$, *k* is the number of explanatory variables, *n* is the number of observations (in the case of my study is the 346 D1 schools), and $R^2_{\hat{u}^2}$ is the R square from the Bruesch-Pagan equation. The results of this regression are listed at the end of this paper. The F statistic for this test was 12.581, with df_1=9 and df_2=336. The p-value of F was 2.646E-17, so the null hypothesis of homoskedasticity was rejected at any usual level of confidence. To correct this problem, I reconsidered the functional form that would be most appropriate for this model. I found that average attendance tends to vary more as percentage of games won in the current season increases. Logically this would make sense, as teams less popular schools with less attendance will tend to play against one another, therefore some of these teams will have win percentage as high as some of the more popular teams who are winning a large portion of their games. This graph shows the relationship between *Attend* and *Pct16*, with three teams selected, University of Kentucky, University of Virginia, and Stony Brook University, each with 75, 78.4, 78.8 percent of games won in the 2016 season, respectively.



These data visually show what appears to be a nonlinear and heteroskedastic relationship. This problem was corrected by taking the natural log of the dependent variable, *Attend*. For my corrected model I use the dependent variable *ln(Attend)*. The relationship between this variable and *Pct16* is shown below:



I then estimated the log model[3]:

$$\ln(Attend) = \beta_0 + \beta_1 Pct16 + \beta_2 \ln(Pop) + \beta_3 CompConf(NCAA9) + \beta_4 Appear2015 + \beta_5 Appear2014 + \beta_6 Appear2013 + \beta_7 Appear2012 + \beta_8 Appear2011 + \beta_9 Appear2010 + u$$

and repeated the Bruesch-Pagan test for heteroskedasticity[4], and found the F-stat to be 1.163, with a p-value of .318, so I do not reject the null hypothesis that the error term is homoskedastic with respect to the explanatory variables for the logged model.

With the logged model, some of the hypothesis tests for the coefficients changed. *Pct16*, *ln(Pop)*, *CompConf(NCAA9)*, and *Appear* in the years 2015, 2014, and 2010 were all still positive and significant at the .05 level. The t-statistics for the estimates of $\beta_6, \beta_7,$ and β_8 had p-values of .052, .051, and .28, respectively. I wanted to test how far the positive effect of being in the NCAA tournament on average attendance of a team's home games would reach, so I performed an F test for joint significance of these variables. I ran a restricted regression[5] by excluding the *Appear* variables in the years 2010-13. The F-statistic from this test is found using the formula $F = \frac{(SSR_r - SSR_{ur})/q}{(SSR_{ur})/(n-k-1)}$, where SSR_r is the sum of squared residuals for the restricted model, SSR_{ur} is the sum of squared residuals for the unrestricted model, *q* is the number of restrictions (in this case 4), *k* is the number of independent variables in the unrestricted model, and *n* is the number of observations. It is used to test the null hypothesis that

$$\beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$$

The test resulted in an F-statistic of 7.610, with a p-value of 7.040E-06, therefore the null hypothesis is rejected. These variables, although they individually had low t-stats, have a significant effect on the overall model's usefulness in explaining variations in average attendance on NCAA Division 1 basketball games.

Another adjustment I made to the model to interpret its effect was to add *Pct15* as an independent variable[6]. I determined this was not a good addition to my model, due to the coefficients t-stat of -.42, and the adjusted R square (a measure of model fit that adjusts for degrees of freedom lost due to increased explanatory variables) dropping from 0.7135 to 0.7128.

Once I was satisfied with the form and inclusion of all the variables I had available, I tried using the Sports Illustrated article for my *CompConf* binary variable rather than the NCAA article[7]. Though the findings were all still significant, using the Sports Illustrated list caused the R square to drop, and the standard error to rise, leading to the conclusion that it made the model's overall fit worse.

Results and Conclusions

The results of the final model show a regression equation as follows:

$$\begin{aligned}\ln(\widehat{Attend}) = \underset{(.3083)}{4.6372} &+ \underset{(.0017)}{.0132}Pct16 + \underset{(.0330)}{.2379}\ln(Pop) + \underset{(.0650)}{.8539}CompConf(NCAA9) \\ &+ \underset{(.0788)}{.1667}Appear2015 + \underset{(.0772)}{.2117}Appear2014 + \underset{(.0809)}{.1579}Appear2013 \\ &+ \underset{(.0828)}{.1623}Appear2012 + \underset{(.0787)}{.0843}Appear2011 + \underset{(.0781)}{.1829}Appear2010\end{aligned}$$

We can use this model to answer some questions we may have about the factors that influence the attendance of collegiate basketball games. The interpretation of our estimates for each coefficient provide some useful insight. For example, the estimate of β_1 means that if a team increases its wins as a percentage of total season games by 1%, attendance should increase by 1.3%, holding all else constant. For β_2, the interpretation is that if the population of a school increases by 1%, attendance should increase by .24%. Both of these estimates are statistically significant.

According to this model, holding student body size, the teams win percentage, and their appearances in the NCAA tournament constant, a team in a stronger conference will see 85% higher average attendance. This is an explanation for the focus in collegiate athletics to get into higher conferences. Being in a conference like the Big East or the Pac-12 really puts a team on the map, and increases the attention received from the viewing community. There are obviously large returns to being in a competitive conference, although some of the interpretation from this coefficient may be that there some other characteristics about a team that gives them higher attendance and gets them invited into these conferences. We can conclude that being in a stronger conference is at least correlated with higher attendance.

To get an estimate of β_4, β_5, β_6, β_7, β_8, and β_9 that is easier to interpret, I combined the binary variables into another variable, *Appear(6yr)*, which represents a count of the number of times the team made it into the NCAA tournament in the six years previous to the one studied. The estimated model looks like this[8]:

$$\ln(\widehat{Attend}) = \underset{(.3058)}{4.6387} + \underset{(.0016)}{.0132}Pct16 + \underset{(.0327)}{.2378}\ln(Pop) + \underset{(.0645)}{.8511}CompConf(NCAA9) + \underset{(.0197)}{.1608}Appear(6yr)$$

The coefficient for the variable *Appear(6yr)* is interpreted to mean that, ceteris paribus, each additional appearance in the NCAA tournament in the previous six years causes attendance to be about 16% higher. I think that this is a meaningful finding to explain the difference in attendance between teams. All twenty-two of the teams that have at least five appearances in the tournament in the last six years also have attendance well above average, by contrast, only 24 of the 174 teams that never made it to the tournament from 2010-2015 had attendance above average.

Team	Attend	Appear(6yr)
Duke	9,314	6
Gonzaga	6,000	6
Kansas	16,436	6
Louisville	20,859	6
Michigan St.	14,797	6
Ohio St.	12,283	6
San Diego St.	12,209	6
Wisconsin	17,287	6
BYU	14,699	5
Cincinnati	9,415	5
Florida	9,686	5
Georgetown	8,879	5
Kansas St.	11,902	5
Kentucky	23,361	5
New Mexico St.	4,767	5
North Carolina	18,326	5
Notre Dame	8,517	5
Syracuse	21,592	5
Texas	12,828	5
VCU	7,637	5
Villanova	8,119	5
Xavier	10,281	5

Average	4417.644509
Standard Deviation	4329.502727

As a note to my methodology in combining the binary variables for appearance, initially I used the binary variables to determine how far back one should look to find the effect of tournament appearance on this year’s attendance. Once I had found that all six years I had included were significant, I combined them into one variable to make the model simpler. I felt this would be appropriate because it did not affect any measure of the model's fit, nor did it change the estimates for the other coefficients in any significant way.

One weakness of my model overall is a large standard error, .4786. (Because this is a log model, this can be interpreted as 47.86%.) This makes the model not very useful in prediction. This is likely due to a very large error term caused by exclusion of some of the variables I listed above, and a large part of the variation in average attendance being caused by unobservable characteristics of teams and schools. Because prediction was not really the focus of my project, but to estimate the effects of the characteristics that are measurable, these results do provide a pretty good framework for isolating what makes fans want to watch some NCAA men's basketball teams and not others.

References

2017 NCAA Men's Final Four Records. (2016). Retrieved December 1, 2016, from http://www.ncaa.org/championships/statistics/2017-ncaa-mens-final-four-records

Gardner, D. (2016, April 22). Villanova won title, but was Big East the best conference this year? Retrieved December 1, 2016, from http://www.si.com/college-basketball/2016/04/22/best-conference-acc-big-12-big-east-big-ten

IPEDS Data Center. (n.d.). Retrieved December 1, 2016, from https://nces.ed.gov/ipeds/datacenter/Default.aspx

NCAA Men's Basketball Attendance. (2016). Retrieved December 1, 2016, from http://www.ncaa.org/championships/statistics/ncaa-mens-basketball-attendance

NCAA.com, J. B. (2016, January 13). College basketball: These are the 9 strongest conferences so far this season. Retrieved December 1, 2016, from http://www.ncaa.com/news/basketball-men/article/2016-01-13/college-basketball-these-are-9-strongest-conferences-so-far

Population Estimates. (n.d.). Retrieved December 12, 2016, from http://www.census.gov/popest/data/metro/totals/2015/index.html

1.

SUMMARY OUTPUT: Attend=f(ln(Pop),Pct16,CompConf(NCAA9),Appear in 10-15)						
Regression Statistics						
Multiple R	0.843108822					
R Square	0.710832487					
Adjusted R Square	0.703086928					
Standard Error	2359.134363					
Observations	346					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	9	4596871862	510763540.2	91.77291685	5.02462E-85	
Residual	336	1870013021	5565514.945			
Total	345	6466884883				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-8430.206397	1511.588417	-5.577051468	5.0355E-08	-11403.57546	-5456.837334
Pct16	38.90947843	8.262443846	4.709197322	3.64386E-06	22.65684335	55.16211351
ln(Pop)	899.069087	161.7758671	5.557498179	5.58004E-08	580.8479672	1217.290207
CompConf(NCAA9)	3883.820213	318.6155754	12.18967468	1.49503E-28	3257.087642	4510.552785
Appear2015	1207.307413	386.4642832	3.123981867	0.001939483	447.1130827	1967.501743
Appear2014	1597.449702	378.6833563	4.218431245	3.168E-05	852.5608392	2342.338565
Appear2013	370.9037392	396.7099549	0.934949412	0.350485808	-409.4443332	1151.251812
Appear2012	1520.581618	405.7717118	3.747382021	0.000210276	722.4086218	2318.754613
Appear2011	758.5180411	385.7373631	1.966410604	0.05007402	-0.246401261	1517.282484
Appear2010	1046.32143	382.9009293	2.732616586	0.006615643	293.136393	1799.506467

2.

SUMMARY OUTPUT: Test for heteroskedasticity, linear model					
Regression Statistics			F-test		
Multiple R	0.497086305		F=	12.58055739	
R Square	0.247094795		p-value:	2.64575E-17	
Adjusted R Square	0.226927691				
Standard Error	9975499.319				
Observations	346				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	1.09732E+16	1.21924E+15	12.25236894	8.6754E-17
Residual	336	3.34356E+16	9.95106E+13		
Total	345	4.44087E+16			

3.

SUMMARY OUTPUT: ln(Attend)=f(ln(Pop),Pct16,CompConf(NCAA9),Appear in 10-15)						
Regression Statistics						
Multiple R	0.849130477					
R Square	0.721022568					
Adjusted R Square	0.713549958					
Standard Error	0.481144051					
Observations	346					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	9	201.0339028	22.33710031	96.48872178	1.27364E-87	
Residual	336	77.78386494	0.231499598			
Total	345	278.8177677				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.637198345	0.308287559	15.04179526	1.86358E-39	4.030781492	5.243615199
Pct16	0.013173021	0.001685121	7.817257753	7.01644E-14	0.009858306	0.016487737
ln(Pop)	0.237925105	0.032994092	7.211142685	3.69895E-12	0.173024097	0.302826113
CompConf(NCAA9)	0.853913046	0.064981457	13.14087253	4.09745E-32	0.726091311	0.981734782
Appear2015	0.166676063	0.078819161	2.114664271	0.035194568	0.011634882	0.321717245
Appear2014	0.211676774	0.077232245	2.740782333	0.006457021	0.059757133	0.363596415
Appear2013	0.157908568	0.08090876	1.951686927	0.051805902	-0.001242957	0.317060094
Appear2012	0.162269827	0.082756899	1.960801193	0.050727994	-0.000517081	0.325056735
Appear2011	0.084343176	0.078670906	1.072101243	0.284443988	-0.070406381	0.239092733
Appear2010	0.182871382	0.078092417	2.341730338	0.019778315	0.029259741	0.336483023

4.

SUMMARY OUTPUT: Test for heteroskedasticity, log model					
SUMMARY OUTPUT					
Regression Statistics			F-test		
Multiple R	0.173780535		F=	1.162563547	
R Square	0.030199674		p-value:	0.318306322	
Adjusted R Square	0.00422288				
Standard Error	0.319528858				
Observations	346				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	1.068265947	0.118696216	1.162563547	0.318306322
Residual	336	34.30516016	0.102098691		
Total	345	35.37342611			

5.

Test for Joint Significance

SUMMARY OUTPUT: Restricted Model

Regression Statistics			F-test	
Multiple R	0.834116129		F=	7.609646158
R Square	0.695749716		p-value:	7.04022E-06
Adjusted R Square	0.691275447			
Standard Error	0.499500883			
Observations	346			

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	193.9873827	38.79747653	155.5002021	1.38491E-85
Residual	340	84.83038505	0.249501132		
Total	345	278.8177677			

SUMMARY OUTPUT: Unrestricted Model

Regression Statistics	
Multiple R	0.849130477
R Square	0.721022568
Adjusted R Square	0.713549958
Standard Error	0.481144051
Observations	346

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	201.0339028	22.33710031	96.48872178	1.27364E-87
Residual	336	77.78386494	0.231499598		
Total	345	278.8177677			

6.

SUMMARY OUTPUT: ln(Attend)=f(ln(Pop),Pct16,CompConf(NCAA9),Appear in 10-15,Pct15)						
Regression Statistics						
Multiple R	0.849218664					
R Square	0.72117234					
Adjusted R Square	0.712849126					
Standard Error	0.481732278					
Observations	346					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	10	201.075662	20.1075662	86.64589944	1.17884E-86	
Residual	335	77.74210574	0.232065987			
Total	345	278.8177677				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.664642138	0.315371607	14.79093882	1.88752E-38	4.044283919	5.285000356
Pct16	0.013505441	0.001860288	7.259867342	2.72566E-12	0.009846124	0.017164758
Pct15	-0.000897617	0.002116024	-0.424199674	0.671692471	-0.005059986	0.003264752
ln(Pop)	0.237619189	0.0330423	7.191363495	4.21572E-12	0.172622652	0.302615726
CompConf(NCAA9)	0.853743152	0.065062134	13.12196674	5.01989E-32	0.725761341	0.981724963
Appear2015	0.181293245	0.086110575	2.105354019	0.036004589	0.011907664	0.350678826
Appear2014	0.213960957	0.077513923	2.760290665	0.006092846	0.061485598	0.366436316
Appear2013	0.160617073	0.081258916	1.976608625	0.048905707	0.000775049	0.320459098
Appear2012	0.161629109	0.08287184	1.950350186	0.051968084	-0.001385652	0.32464387
Appear2011	0.084024395	0.07877067	1.066696451	0.286877062	-0.070923076	0.238971865
Appear2010	0.18669721	0.078706335	2.372073476	0.018253129	0.031876291	0.341518129

SUMMARY OUTPUT: ln(Attend)=f(ln(Pop),Pct16,CompConf(SI5),Appear in 10-15)						
Regression Statistics						
Multiple R	0.807851356					
R Square	0.652623814					
Adjusted R Square	0.643319095					
Standard Error	0.53689654					
Observations	346					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	9	181.9631149	20.21812388	70.13901171	8.98798E-72	
Residual	336	96.85465277	0.288257895			
Total	345	278.8177677				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.382483622	0.342814521	12.7838331	9.12448E-31	3.708150529	5.056816715
Pct16	0.013443286	0.001880092	7.150333784	5.4393E-12	0.009745052	0.017141519
ln(Pop)	0.274757106	0.03654422	7.518483275	5.09014E-13	0.202872821	0.346641391
CompConf(SI5)	0.753481825	0.088478297	8.51600729	5.60779E-16	0.579440645	0.927523005
Appear2015	0.193834903	0.08809113	2.200390709	0.028459293	0.020555301	0.367114505
Appear2014	0.201165629	0.08665332	2.321499391	0.020857568	0.03071427	0.371616988
Appear2013	0.119338806	0.091019647	1.311132368	0.190708368	-0.059701336	0.298378947
Appear2012	0.243436419	0.092241247	2.639127582	0.008699241	0.061993331	0.424879507
Appear2011	0.156584983	0.087336832	1.792885995	0.073890505	-0.015210879	0.328380844
Appear2010	0.180857634	0.087649594	2.06341668	0.039840757	0.008446554	0.353268714

7.

SUMMARY OUTPUT: ln(Attend)=f(ln(Pop),Pct16,CompConf(NCAA9),Appear(6yr))						
Regression Statistics						
Multiple R	0.848429429					
R Square	0.719832497					
Adjusted R Square	0.716546074					
Standard Error	0.47862118					
Observations	346					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	4	200.7020898	50.17552246	219.032256	7.53717E-93	
Residual	341	78.11567788	0.229078234			
Total	345	278.8177677				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.638737005	0.305792693	15.16954822	4.4056E-40	4.037259556	5.240214453
Pct16	0.013173652	0.00164661	8.000470443	1.95669E-14	0.009934861	0.016412442
ln(Pop)	0.237846332	0.032682068	7.277579059	2.35843E-12	0.173562498	0.302130166
CompConf(NCAA9)	0.851101465	0.064455507	13.20448019	1.94145E-32	0.724321019	0.97788191
Appear(6yr)	0.160818103	0.019695914	8.165049071	6.30267E-15	0.122077321	0.199558885

8.