Group 2 - Brittany Fuller and Ian Parks
MA 331 – Intermediate Statistics
Project Part II

Does an NCAA women's basketball team have a home court advantage? A home court advantage is defined as the perceived "boost" in ability a team's players get when playing on their college's court in front of their own fans. We believe that there are two main ways to determine the presence and intensity of a home court advantage. One way is to analyze the difference between a team's home and away winning percentages; another way is to analyze point differentials. When using the first method we would determine the team with the largest difference to be the team with the greatest home court advantage and the team with the smallest difference to be the team with the least home court advantage. When using the method of point differentials we would determine the team with the highest difference between point differentials of their home and away games to be the team with the greatest home court advantage and the team with the lowest difference between point differentials between their home and away games to be the team with the worst home court advantage. We have collected a lot of data for various NCAA teams over a several year span. Through our analysis we have found answers to the following questions:

- Which teams in each of the ACC, NCC and E8 conferences we sampled have shown to have the greatest home court advantage, as judged by the difference in the winning percentages of their home and away conference games? Does each team have the highest winning percentage in the conference? Does each team have the highest home winning percentage in the conference?
- Is there any correlation between the teams that have recorded the highest winning percentage and the teams that have shown to have the greatest home court advantage in the ACC, NCC and E8? How about home winning percentage?
- Which team in the E8 has shown to have the greatest home court advantage, as judged by point differentials? Does this team have the highest winning percentage in the conference? Does this team have the highest home winning percentage in the conference?
- Which division (I, II or III) has shown to have the greatest home court advantage as judged by their home and away winning percentages?
- What role does the attendance of home games play in the presence and intensity of a team's home court advantage?

Samples of the data we have collected are listed below:

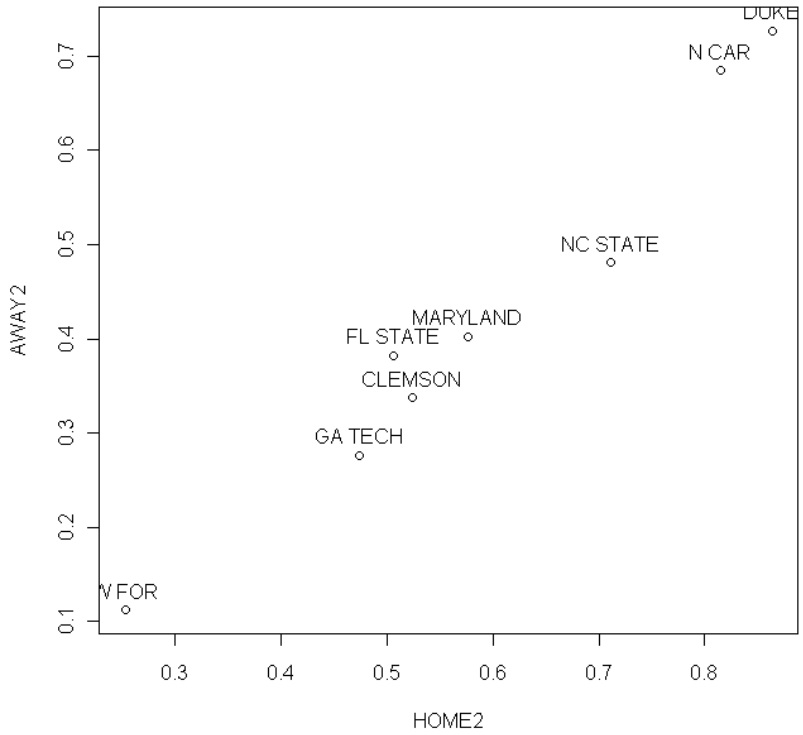
- **Twelve seasons of the Atlantic Coast Conference's (ACC), a division I conference, home and away win-loss records. (1996-2008)**

Maryland	HOME WINS	AWAY WINS	ROAD WINS	ROAD LOSES	
1996-1997	6	2	3	5	
1997-1998	5	3	2	6	
1998-1999	2	6	1	7	
1999-2000	2	6	3	5	
2000-2001	4	4	4	4	
2001-2002	4	4	0	8	
2002-2003	2	6	2	6	
2003-2004	6	2	2	6	
2004-2005	4	3	3	4	
2005-2006	6	1	6	1	
2006-2007	5	2	5	2	
2007-2008	7	0	6	1	
	53	**39**	**37**	**55**	
		0.57608696		0.40217391	0.173913

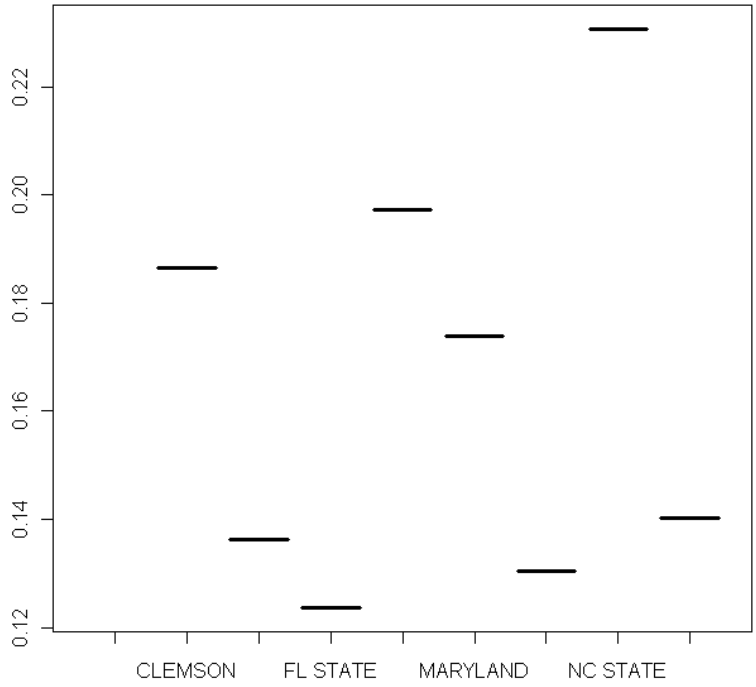
The data above is a sample of the data we have collected for the ACC. The sample of data is from the women's basketball program at Maryland. The data shows that over the past twelve seasons Maryland has won 53 games at home and 37 games on the road. Maryland has also lost 39 games at home and lost 55 games on the road. This would make Maryland's home winning percentage over the past 12 seasons 57.61% and Maryland's away winning percentage 40.22%. The difference between Maryland's home and away winning percentages is 17.39%. From this data we can also calculate the overall winning percentage over these seasons to be 48.913%.

TEAM	**HOME**	**AWAY**
CLEMSON	0.52381	0.337349
DUKE	0.863636	0.727273
FL STATE	0.505618	0.382022
GA TECH	0.473684	0.276316
MARYLAND	0.576087	0.402174
NC STATE	0.711538	0.480769
N CAR	0.815217	0.684783
W FOR	0.252632	0.11236

Above is a table showing the overall home and away winning percentages for the ACC teams.



The above R graph represents the plot of the average home and away winning percentages of the ACC schools for the past twelve seasons. This graph makes it easy to see that Duke has had the highest home and away winning percentages as well as the highest overall winning percentage over the past twelve seasons while Wake Forest has had the lowest home and away winning percentages as well as the lowest overall winning percentage over the past twelve seasons.



The R graph above represents the difference in the average home and away winning percentages of the ACC schools over the past twelve seasons. This graph makes it easy to see that NC State has had the greatest difference between their home and away winning percentage over the past twelve seasons, while Florida State has had the least. Therefore, NC State has proven to have the greatest home court advantage and Florida State has proven to have the least home court advantage. These two graphs make it easy to see that neither the team with the highest winning percentage nor the team with the highest home winning percentage over the past twelve seasons in the ACC is the same team that has the greatest home court advantage.

- **The overall conference records of teams of the ACC from 1996-2008**

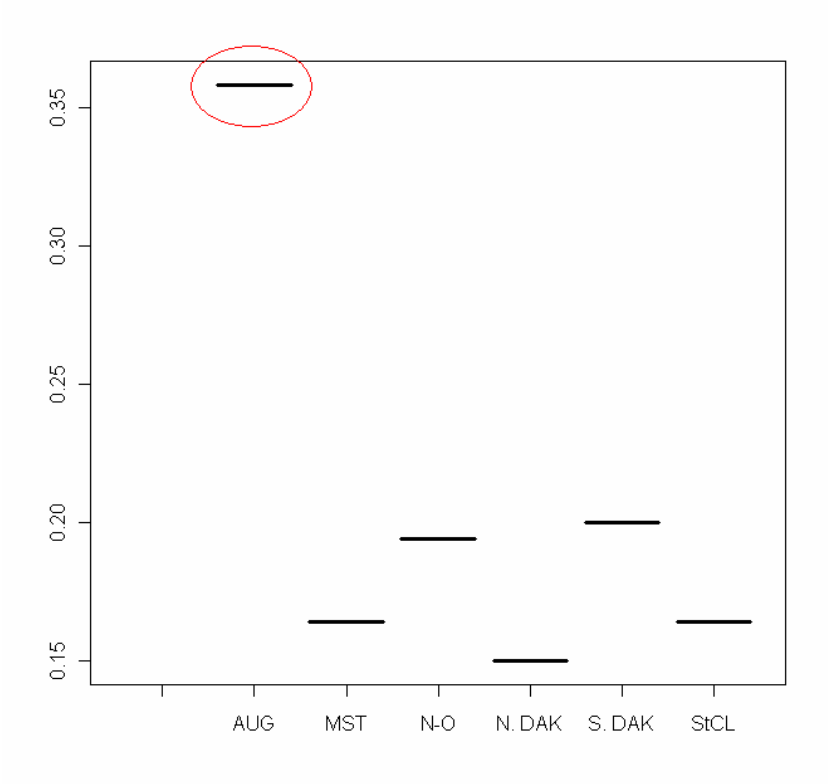
  It is easy to determine the overall conference records of the ACC teams from the data we have collected. According to the sample data for Maryland that we have included earlier in this report, Maryland's overall conference record for the past twelve seasons is 90-94 and their overall winning percentage is 48.913%.

- **Nine seasons of the North Central Conference's (NCC), a division II conference, home and away win-loss records. (1998-2008)**

We collected similar data for the NCC. For example, we will examine the data for Augustana. The data shows that Augustana has won 51 home games and 27 away games over the past nine seasons while losing 16 home games and 40 away games. This would make their home winning percentage 76.12% while their away winning percentage is only 40.30%. This makes the difference in their home and away winning percentages 35.82%. From this data we can calculate Augustana's overall winning percentage to be 58.209%.

TEAM	HOME	AWAY	
AUG	0.761194	0.402985	0.358209
MST	0.373134	0.208955	0.164179
N-O	0.402985	0.208955	0.194030
N. DAK	0.933333	0.783333	0.150000
S. DAK	0.533333	0.333333	0.200000
StCL	0.61194	0.447761	0.164179

The above table shows the home and away winning percentages for the NCC teams. The final column of the table is the difference between the home and away winning percentages for each team. As you can see, Augustana has the highest home court advantage by far, even though they don't have the highest home winning percentage.



The data shown in this graph for the NCC seems to be more interesting than the same graph for the ACC and the E8. This graph shows Augustana to have a significantly higher difference in their average home and away winning percentages than all of the other teams in the NCC. The difference between the average home and away winning percentages for Augustana for the past nine seasons is 35.82% and the average difference between the home and away winning percentages for all of the other teams in the NCC for the same nine years is 17.445%. If we were to add Augustana into this average it would increase the average up to 20.619%. For further calculations it might be good to treat Augustana as an outlier to this data set for this reason. We think that this information proves to be important later when we interpret which division has the highest home court advantage.

- **The overall conference win-loss records of teams in the NCC from 1998-2008**

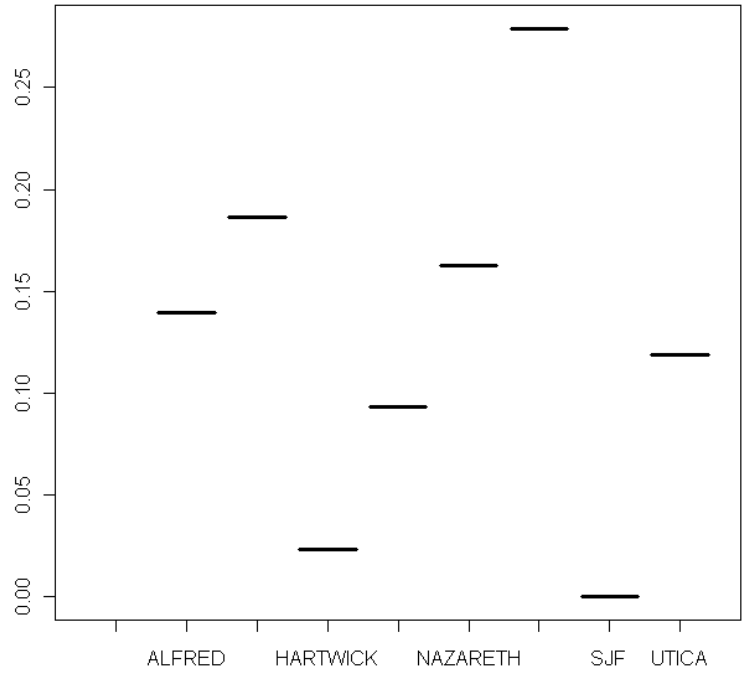
  It is easy to determine the overall conference records of the NCC teams from the data we have collected. According to the sample data for Augustana that we have included earlier in this report Augustana's overall conference record for the past nine seasons is 78-56 and their overall winning percentage is 58.209%.

- **Six seasons of the Empire 8 Conference's (E8), a division III conference, home and away win-loss records. (2002-2008)**

  We also collected similar data for the E8. For example, we will look at the data for Alfred. The data shows that Alfred has won 6 games at home and 0 games away over the past six seasons while losing 37 games at home and 43 games away. This would make their home winning percentage 13.95% and their away winning percentage 0. The difference between their home and away winning percentages is 13.95%. From the data we have collected it is easy to determine the overall winning percentage for Alfred over the past six seasons which is 6.977%.

TEAM	HOME	AWAY
ALFRED	0.139535	0
ELMIRA	0.581395	0.395349
HARTWICK	0.534884	0.511628
ITHACA	0.883721	0.790698
NAZARETH	0.44186	0.27907
RIT	0.55814	0.27907
SJF	0.738095	0.738095
UTICA	0.642857	0.52381

The above table shows the home and away winning percentages for the Empire 8 teams over the past six seasons.



The above R graph represents the difference in the average home and away winning percentages of the E8 schools over the past six years. This graph makes it easy to see that RIT has shown to have the greatest difference between home and away winning percentages while Saint John Fisher has had the smallest difference. Therefore, RIT has had the greatest home court advantage and Saint John Fisher has had the smallest home court advantage over the past six years. This graph makes it easy to see that the team with the highest overall and home winning percentages is not the same team as the team that has had the greatest home court advantage over the past six seasons.

- **The overall conference win-loss records of teams in the E8 from 2002-2008.**

  It is easy to determine the overall conference records of the E8 teams from the data we have collected. According to the sample data for Alfred that we have included earlier in this report Alfred's overall conference record for the past six seasons is 6-80 and their overall winning percentage is 6.977%.

  We performed linear regressions in R to test if there is a correlation between overall winning percentages and home court advantage for the ACC, NCC and E8. The linear regression for overall winning percentage and the difference between home and away winning percentages for the ACC has an r-squared value of 0.08288, which makes its r-value 0.2879. The linear regression for overall winning percentage and the difference between home and away winning percentages for the NCC has an r-squared value of 0.1215 which makes its r-value 0.3486. The linear regression for overall winning

percentage and the difference between home and away winning percentages for the E8 has an r-squared value of 0.2128 which makes its r-value 0.4613. Since the r-value is the correlation coefficient we can see that the E8 had the highest correlation between these variables and the ACC had the lowest correlation between these variables, while the NCC was somewhere in the middle. None of these variables are particularly strongly correlated.

We also performed linear regressions in R to test if there is a correlation between home winning percentages and home court advantage for the ACC, NCC and E8. The linear regression for home winning percentage and the difference between home and away winning percentages for the ACC has an r-squared value of 0.002536, which makes its r-value 0.0504. The linear regression for home winning percentage and the difference between home and away winning percentages for the NCC has an r-squared value of 0.1872, which makes its r-value 0.4327. The linear regression for home winning percentage and the difference between home and away winning percentages for the E8 has an r-squared value of 0.08559, which makes its r-value 0.2925. Since the r-value is the correlation coefficient we can see that the NCC had the highest correlation, the ACC had the lowest correlation, and the E8 was somewhere in the middle. Again, none of the variables seem particularly strongly correlated.

- **The point differentials for all 72 E8 conference games in the 2007-2008 season.**

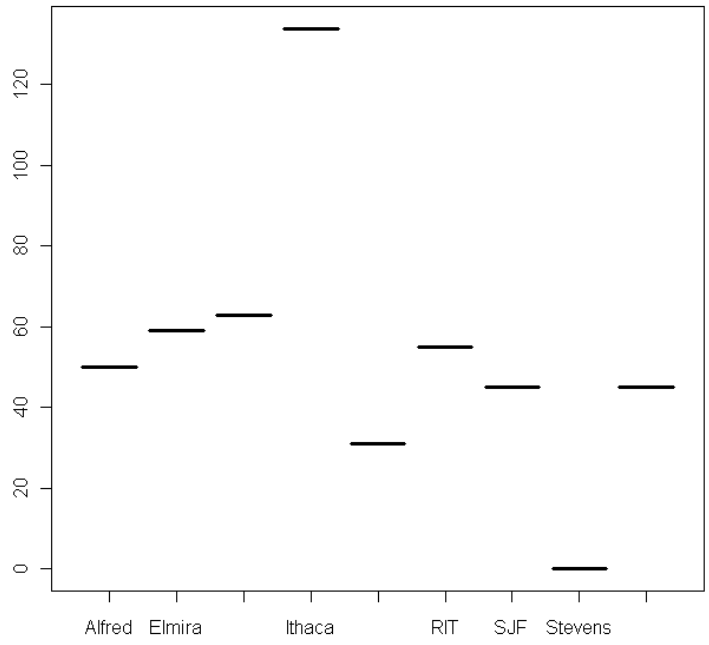
  We believe that point differentials would be another way to determine the presence and intensity of a team's home court advantage. In our research we took every conference game in the Empire 8 during the 2007-2008 season and determined the point differential. For example, the point differential for the two occasions when Alfred played Hartwick would be +4 for Alfred at home and -6 for Alfred on the road because Alfred beat Hartwick by 4 at home but lost to Hartwick by 6 on the road. When comparing those two games to determine home court advantage for Alfred we would say that the point differential would be -2 because Alfred lost to Hartwick by 2 more points on the road than they beat Hartwick by at home.

Alfred		Home	Away	
	Elmira	-14	-6	
	Hartwick	4	-6	
	Ithaca	-8	-47	
	Nazareth	-13	-27	
	RIT	-16	-16	
	SJF	-8	-20	
	Stevens	-54	-16	
	Utica	-23	-44	
		-132	-182	50

Above is a sample of the data for the point differentials for all of Alfred's conference games. For that season, Alfred posted a home winning percentage for the season of 12.50% and away winning percentage for the season of 0 with the difference between the two being 12.50%. The data shows that Alfred was outscored by their E8 opponents in the 2007-2008 season at home by a total of 132 points and that Alfred was outscored by their opponents in the 2007-2008 season for away games by a total of 182 points. Therefore Alfred scored 50 more points in conference games at home than away. Alfred also lost games by fewer points or beat a team by more points at home than they did when they played the same team on the road for 5 out of their 8 conference opponents.

TEAM	HOME	AWAY
ALFRED	-132	-182
ELMIRA	15	-44
HWICK	-44	-107
ITHACA	178	44
NAZ	-15	-46
RIT	16	-39
SJF	4	-41
STEVENS	128	128
UTICA	103	58

The above table represents the point differentials for the home and away games for each team in the E8 for the 2007-2008 season.



The above R graph represents the difference in home and away point differentials for the 2007-2008 season in the E8. The graph shows that Ithaca had the highest difference in point differentials while Stevens had the lowest. Therefore based on the point differentials Ithaca has had the greatest home court advantage and Stevens had the smallest home court advantage during the 2007-2008 seasons.

Judging by the difference in the team's winning percentage for home games and their winning percentage for away games in 2007-2008 one would conclude that RIT had the greatest home court advantage. Their difference in winning percentage of 37.50% is higher than any of the other teams'. However, if one would analyze the point differential data one would conclude that Ithaca had the greatest home court advantage. Ithaca outscored their opponents by 134 more points at home than away while RIT outscored their opponents by 16 points at home and was outscored by their opponents by 39 points away, making their point differential only 55 points. This analysis is possible for the Empire 8 conference but may not be possible for other conferences due to the fact that in the Empire 8 every team plays each other twice, once at home and once away. However, the other conferences do not schedule their games in this manner. In the division I conference, the ACC, for the 2007-2008 season, there were 11 teams and the teams only played 14 games against conference opponents. Therefore the teams only played three teams twice (one home and one away) and all the rest of the teams once (once home or once away). So depending on scheduling, a team may play all of the statistically best teams in the conference at home and all of the statistically worst teams away. Therefore we would expect that team to have a very low point differential. Another team could be scheduled exactly opposite so that one would expect them to have a very high point differential. Although from this data one might conclude that the second team had a greater home court advantage against conference opponents, this is not necessarily true.

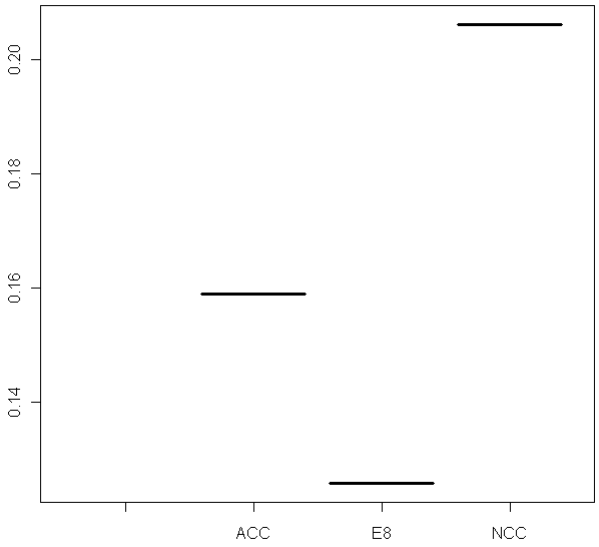
We believe that a study involving point differentials must take the team's schedule into account in order to be considered a superior method than analyzing home and away winning percentages. This is because the team in the conference with the worst winning percentage could lose every single conference game during the season. However, the worst team may only lose at home by a small margin on average and lose on the road by a large margin on average. Although the worst team lost all games, they still played teams to much closer games at home than away, which would show a home court advantage. However, if we were to just look at the difference between home and away winning percentages, the fact that they played the best team to a close game at home and were blown out on the road would not be shown. For this reason we think that it is important to note the point differentials whenever the proper data is available.

We performed a linear regression to test if there is a correlation between the difference in home and away winning percentages and the overall point differential for a season. The regression showed the r-squared value to be 0.1497, which makes the r-value 0.3869. We also performed a linear regression to test if there is a correlation between overall winning percentage and point differentials for a season. The regression showed the r-squared value to be 0.02677, which makes the r-value 0.1636. Sine the r-value is the correlation coefficient this data shows that the correlation between difference in home and away winning percentage and overall point differential is greater because it has a higher r-value than the correlation between overall winning percentage and point differentials.

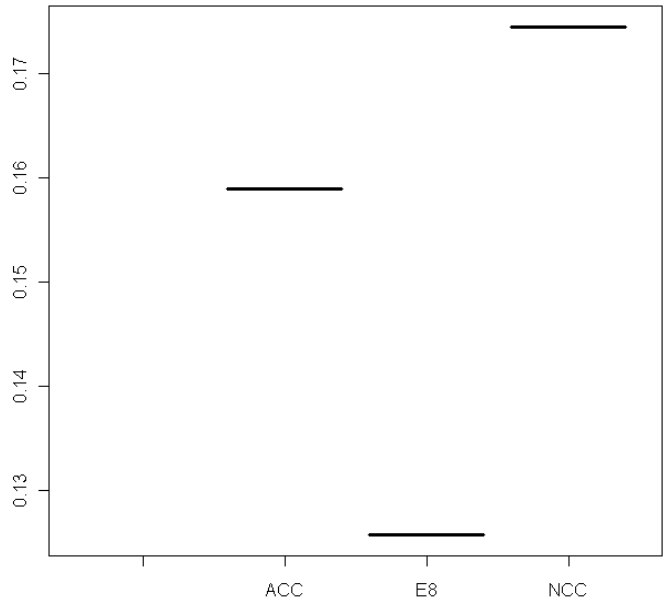
- **The average difference in home and away winning percentages for division I, II and III.**

CONF	HOME WINS	HOME LOSES	AWAY WINS	AWAY LOSES	HOME WINNING %	AWAY WINNING %	DIFF	DIFF2
ACC	352	272	250	367	0.564103	0.405186	0.158916	0.158916
NCC	232	156	152	236	0.597938	0.391753	0.206186	0.174455
E8	193	149	150	192	0.564327	0.438596	0.125731	0.125731

This table shows the difference in home and away winning percentages for the division I conference to be 15.892%, the division II conference to be 20.619%, and the division III conference to be 12.573%. Therefore the division II conference has had the highest home court advantage. However, when we determined the difference in the home and away winning percentages for the NCC we decided that Augustana should be treated as an outlier and eliminated from the data set. When we eliminated Augustana we found the difference in the division II home and away winning percentages to be 17.446%. This data still shows the division II conference to have had the greatest home court advantage but it shows the intensity of home court advantage for division I and II to be closer than when Augustana was not included in the data set.



The above R graph illustrates our data for the difference in home and away winning percentages for the ACC, NCC and E8, including Augustana for the NCC.



The above R graph illustrates our data for the difference in home and away winning percentages for the ACC, NCC and E8 when we remove Augustana from the data set.

- **The average attendance for the division I, division II, and division III for the 2007-2008 season.**

DIVISION I	1,629
DIVISION II	468
DIVISION III	239

This data shows that on average division I schools attract many more fans to their games than division II schools and that on average division II schools attract many more fans to their games than division III teams.

We wanted to determine if there is any correlation between attendance of home games and the idea of the home court advantage. As already stated above, we believe that the division with the highest difference in home and away winning percentages has the greatest home court advantage. To determine if there is a correlation between attendance and home court advantage we ran a linear regression in R of attendance and difference in home and away winning percentages for all three divisions. We found the r-squared value for the linear regression to be 0.002862, which makes the r-value 0.0535. This is a linear regression of the difference in winning percentage of the three divisions when we include Augustana in the data for division II. We also performed a linear regression between difference in home and away winning percentage of all three divisions and average attendance while treating Augustana as an outlier and removing it from our data. The r-squared value for these factors was 0.1243, which makes the r-value 0.3526. Since the values we have found for the difference in home and away winning percentage for all three divisions are pretty close in numbers, and the attendance numbers for the three divisions are extremely different, we can say that attendance is not a big factor in home court advantage. We actually expected attendance to play a bigger role in the intensity of the home court advantage. However, judging by the three conferences we chose, attendance does not appear to be very significant. For example, the NCC showed a greater home court advantage than the ACC, but the ACC had a much higher attendance rate than the NCC. A possible explanation for this could be that although division I schools have a higher attendance rate than division II schools, division I schools also draw fans from all over the country. So although most fans will be cheering for the home team, it is possible for away teams to draw large crowds at games too. This would presumably lessen the effect of fan cheering on the home team, but this would only have to be one small factor out of many.

In conclusion, many of our questions seem to have negative responses. We've seen that there's no apparent correlation between teams with the highest win rate and the greatest home court advantage. Neither is there a correlation between teams with the highest home win rate and teams with the greatest home court advantage. Finally, there does not seem to be any correlation between attendance and home court advantage.

FINAL ADDITIONS

We ran a multiple regression analysis to determine if there was any difference between the conferences. We compared the overall winning percentage with the home court advantage, using the different conferences as groups. Here is what we found:

```
summary(lm(wp~hca+confs))

>Call:

>lm(formula = wp ~ hca + confs)

>Residuals:

>    Min      1Q  Median      3Q     Max

>-163138  -58662  -13044   14597  688236

>Coefficients:

>            Estimate Std. Error t value Pr(>|t|)

>(Intercept)    80926     115596   0.700    0.493

>hca          -490766     583574  -0.841    0.411

>confsNCC      162785     100611   1.618    0.123

>confsE8       -19410      93474  -0.208    0.838

>Residual standard error: 181200 on 18 degrees of freedom

>Multiple R-squared:  0.16,      Adjusted R-squared: 0.01999

>F-statistic: 1.143 on 3 and 18 DF,  p-value: 0.3586

> anova(linreg2)

Analysis of Variance Table

Response: winning percentage

          Df     Sum Sq    Mean Sq F value Pr(>F)

hca        1 8.2049e+08 8.2049e+08  0.0250 0.8761

confs      2 1.1169e+11 5.5845e+10  1.7016 0.2104

Residuals 18 5.9074e+11 3.2819e+10
```

As you can see, neither of these tests is in any way significant. This means we can reject our hypothesis that there is a difference between the various conferences.

We then ran an analysis of variance on the winning percentage, using the conferences as categories. This will determine if there is any difference in the winning percentage between the conferences. The results were:

```
> summary(lm(wp~confs))

Call:

lm(formula = wp ~ confs)

Residuals:

       Min         1Q     Median         3Q        Max

-1.431e+05 -4.053e-01 -7.071e-02  8.621e-02  7.153e+05

Coefficients:

              Estimate Std. Error   t value Pr(>|t|)

(Intercept)  5.082e-01  6.355e+04  8.00e-06    1.000

confsNCC     1.431e+05  9.708e+04     1.474    0.157

confsE8     -5.797e-03  8.988e+04 -6.45e-08    1.000


Residual standard error: 179800 on 19 degrees of freedom

Multiple R-squared: 0.127,      Adjusted R-squared: 0.03509

F-statistic: 1.382 on 2 and 19 DF,  p-value: 0.2752

> anova(linreg4)

Analysis of Variance Table

Response: wp

          Df     Sum Sq    Mean Sq F value Pr(>F)

confs      2 8.9301e+10 4.4651e+10  1.3818 0.2752

Residuals 19 6.1395e+11 3.2313e+10
```

Again, we find that there is no significance in any of the tests. This indicates that we can reject the hypothesis that the means of the samples are the same.

We also wanted to know if there was any bearing of one season's homecourt advantage on another's. We compiled all the data for the E8 conference and organized it into groups based on season. We took the homecourt advantage for each team for that season and applied the analysis of variance test. Below are our results in R:

```
> anova(lm(HCA~season))

Analysis of Variance Table

Response: Homecourt Advantage

           Df  Sum Sq   Mean Sq  F value  Pr(>F)

season      5  0.29074  0.05815   1.9772  0.1019

Residuals  42  1.23517  0.02941
```

This tells us two important things. One is that the F-value is very close to 1. This implies that there is no group effect. That is, one year's homecourt advantage has absolutely no bearing on any others. Additionally, we find that the test is not significant. This allows us to reject the hypothesis that the group means are equal.