# The Value of Possession in College Football in Differing Conferences

Hanshi Zuo

Michigan State University

## Abstract

We compare differences in possession across different conferences in American collegiate football, specifically the Big12 and Big10. We estimate the value of possession on the field in 10 yard intervals and use that to determine the approximate value of picking up 10 yards in both conferences. We also show that there are significant differences in average points per drive that are especially noticeable in drives starting closer to the opponent's end. Finally, we observe differences in average drive lengths and use all the information we above to provide estimates for when a team should decide to go for a fourth down on their opponent's side of the 50.

# 1 Introduction

While research into American football often focuses on the professional level (see Romer), many could argue that there is much to explore at the collegiate level. For one, college level play features a wide diversity of teams and styles. There are about twice as many teams in the Power 5 conferences alone as there are NFL franchises. Furthermore, these differing play styles are often more popular in certain conferences than in others. Perhaps two of the most iconic conferences at this level of play are the Big12 and the Big10. Situated in the plains and upper mid-west respectively, teams in these two conferences are often assumed to feature stark differences in play style and philosophy. The Big12 is known for high powered offenses leading to high-scoring shootouts such as the 2020 Red River Rivalry (Oklahoma over Texas 53-45 4OT) and the classic 2008 Texas at Texas Tech match (39 - 33 Tech). The Big10 is often remembered for scores such as the 6-4 affair from the Iowa victory over Penn State in 2004. These differences form the basis for this article: we seek to see if there truly are differences between the two conferences, specifically in the value of possession and the points scored per drive.

The paper is sectioned up as follows: (2) provides a description of the methods and statistical tools used as well as our framework. (3) delivers our results and features some explanations and commentaries for these results. (4) discusses implications, including providing estimates for the crucial fourth down decision in "no man's land." Finally, in (5) we offer concluding remarks and questions that could be answered in future research.

# 2 Framework and Methods

We collect our data for the Big10 and Big12 conferences from cbssports.com. Specifically we look at all the play by play data from every intra-conference game from the 3 randomly selected weeks of 6,8,13 from the 2019 college football season. In our data, which can be found on Github(github.com/HanshiZ/499Paper), we record the quarter, the start and end of a drive, the points scored, and the possession of the ball. We also record how a drive ends by all the possible turnovers(interception, fumble lost, downs), punts, and kickoffs for a score. In the case of kickoffs, we record the end state as being a kickoff from the 35 (or wherever it takes place in the event of a penalty). To reduce the effects of end of half clock management, we only take the drives that start in either the third or first quarters of play.

From this information, we go on to determine the value of possession between ten yard intervals starting from the 0-9 all the way to the 90-99 with a method inspired by Romer. While it is easy enough to gather the estimates for the average amount of points obtained based on starting point, to estimate the actual value of possession within one of these ten yard strips, we take into account the value of eventually giving the ball back to the opposing team. Thus, the value of possession at any point is the average points gained minus the value of the opponent getting the ball back at any point multiplied by the chance of that occurring.We thus obtain the following set of equations:

$$v_1 = pts_1 - v_1 p_{1,1} - v_2 p_{1,2} ... - v_n p_{1,n} \tag{1}$$

$$v_2 = pts_2 - v_1 p_{2,1} - v_2 p_{2,2} ... - v_n p_{2,n} \tag{2}$$

$$...$$

$$v_n = pts_n - v_1 p_{n,1} - v_2 p_{n,2} ... - v_n p_{n,n} \tag{n}$$

Here, the $v_i$ represent having the ball in each possible interval from 1 to $n$ number of intervals, the $pts_i$ represent the average points for a team starting in the $i$th interval, and the $p_{i,j}$ represent the probability of starting from the $i$th interval and giving the ball to the opponent on the $j$th interval on the ensuing drive. As stated above, we divide our field into intervals of 10 yards, and as such we would expect to have to solve a system of 10 equations to determine the values. However, we only have data for the first 9 intervals so we settle for a system of 9 equations when computing our results. A slight disadvantage to this method is needing the start of the ensuing drive, therefore, the data used in computing these results in section 3 are a subset of the entire set as the last drive in each quarter is lost.

After obtaining the results for the various $v_i$, we use a linear regression to estimate the value gained from moving from one interval to another. This can be summed up as:

$$v_i = \beta_0 + \beta_1 x_i + \epsilon \tag{3}$$

where $v_i$ is the estimated value from the regression, $\beta_0$ the intercept (corresponding to starting within one's own endzone), $\beta_1$ the slope of the regression, $x_i$ the interval from $[(i-1)*10, i*10)$ yard lines, and the $\epsilon$ is the error term. Since we divided the field up by 10 yard intervals, the $\beta_1$ is the expected value gain from moving up an interval. This also roughly translates to the value of picking up a first down.

We also use Student's t - distribution in testing the difference of mean points and mean yards per drive. Since we do not know that the variances of the distributions before hand, we assume they are not equal and use a t-test for unequal variances for all of our tests. More commonly known as Welch's t-test, this test yields about the same results as a typical t-test when the variances are the same, but is far superior in cases of differing variances (Rice). However, in order to utilize a t- test, the sampling distribution must be approximately normal. We make no assumptions about the distribution of yards gained per down nor of the point averages (especially since football scores are not on a continuum). Instead, we consolidate our data into fewer groups and rely on the central limit theorem (CLT). Rice indicates that sample sizes of about $n = 25$ suffice for sampling distributions to become relatively normal in most cases. For our data that we use, $n = 24$ is the smallest sample that occurs, so we are confident that our usage of the t - test is justified.

The t test statistic for our test for samples with unequal variances is as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

with $\bar{x}_i$ being the $i$th sample mean, $s_i^2$ being the $i$th sample variance, and $n_i$ being the $i$th sample size. The degrees of freedom are based off of the following equation.

$$df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})}{(\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)})}$$

We round down to the nearest integer to provide a more conservative estimate.

## 3 Results

We begin by considering the average points scored in each conference based on the starting interval of a drive. In the following tables, a 1 indicates that the drive started between the zero and the nine yard lines, a 2 indicates that the drive started somewhere between the ten and nineteen yard lines, etc. We continue to measure from one's own goal-line, so that a 9 indicates that a team started a drive within their opponents twenty and eleven yard lines. We begin with the following table to compare the differences between the two conferences. Not surprisingly, the values of the Big12 are consistently higher than those of the

| Start | Avg. Points Big10 | Avg. Points Big12 |
|---|---|---|
| 1 | 1.27 | 2.33 |
| 2 | 0.923 | 1.73 |
| 3 | 1.74 | 1.82 |
| 4 | 2.35 | 2.55 |
| 5 | 2.25 | 4.77 |
| 6 | 1.17 | 2.33 |
| 7 | 2.5 | 8 |
| 8 | 2 | 7 |
| 9 | 5.67 | 5.67 |

Table 1: Average points and drive start (data where there is an ensuing drive)

Big10. However, due to our smaller sample size. we do not have observations

| $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ | $v_9$ | Avg. Points |
|---|---|---|---|---|---|---|---|---|---|
| 1.000 | 0.111 | 0.556 | 0.111 | 0.000 | 0.222 | 0.000 | 0.000 | 0.000 | 2.333 |
| 0.045 | 1.045 | 0.409 | 0.318 | 0.136 | 0.045 | 0.000 | 0.000 | 0.000 | 1.727 |
| 0.073 | 0.220 | 1.476 | 0.085 | 0.073 | 0.024 | 0.012 | 0.012 | 0.024 | 1.817 |
| 0.150 | 0.250 | 0.450 | 1.150 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 2.550 |
| 0.154 | 0.000 | 0.538 | 0.154 | 1.077 | 0.000 | 0.000 | 0.000 | 0.077 | 4.769 |
| 0.333 | 0.000 | 0.667 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 2.333 |
| 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 1.000 | 0.000 | 0.000 | 8.000 |
| 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 7.000 |
| 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 5.667 |
| | | | | | | | | | |
| 1.426 | 0.322 | 0.662 | 1.702 | 3.293 | 1.417 | 4.707 | 5.298 | 5.004 | |

Table 2: Big12 system of equations for the value of possession and the results

for drives starting within 10 yards of an opponent's endzone, and this issue extends to most regions on the "plus" side of the field. For example, on the only example of a drive starting between an opponents 40 and 30 yard lines in the Big12 ended in a rare score of 8 points, most likely reflecting the "chasing" of points in the third quarter (typically, going for two in football is ill-advisable, but during certain situations later in the game, teams will do it in order to get within a certain number of possessions or tie or to extend a 4 point lead into a 6 point lead as 3 is generally the smallest score increment). As we will show later, some of these average differences in points are indeed significant and meaningful.

We now consider the values of possessions in the two conferences. As stated above, we compute this using a system of equations, which considers both the probabilities of one's opponent receiving the ball at every possible interval along with the expected number of points obtained for having the ball together to obtain the value for that possession. We have the coefficients for the system for the Big12 first. Table 2 shows values that we would generally expect as there is a definite upward trend as the a team progresses down the field.

We consider a similar table for the Big10 in Table 3. Perhaps the most startling result that jumps out immediately is that the value of possession of the football within one's own 20 yard line is negative. The value never breaks above the 3 points of a field goal until a team is within the opponents 20 yard line. Furthermore, the rate at which the point values increase is noticeably smaller than those in the Big12. We can note that with the figures presented after the references section. The Big12 regression has an intercept of $-.439$ ($SE = .758$) and a slope of .617 ($SE = .134$) while the Big10 regression has an intercept of $-.894$ ($SE = .762$) and a slope of .436 ($SE = .135$). We emphasize here that an $x$ value of 0 here refers to a teams own endzone, and thus possession in the Big12 is never truly expected to be a negative thing as a team must at least start in the field of play (interval 1). However, the same is not true for the Big10, as its value of possession only starts to be positive in the third interval.

| $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ | $v_9$ | Avg. Points |
|---|---|---|---|---|---|---|---|---|---|
| 1.000 | 0.182 | 0.182 | 0.091 | 0.273 | 0.000 | 0.091 | 0.000 | 0.182 | 1.273 |
| 0.077 | 1.077 | 0.154 | 0.538 | 0.077 | 0.038 | 0.000 | 0.000 | 0.038 | 0.923 |
| 0.067 | 0.146 | 1.427 | 0.135 | 0.090 | 0.056 | 0.034 | 0.011 | 0.034 | 1.742 |
| 0.088 | 0.265 | 0.265 | 1.206 | 0.118 | 0.029 | 0.000 | 0.029 | 0.000 | 2.353 |
| 0.150 | 0.150 | 0.450 | 0.100 | 1.100 | 0.050 | 0.000 | 0.000 | 0.000 | 2.250 |
| 0.000 | 0.167 | 0.667 | 0.000 | 0.167 | 1.000 | 0.000 | 0.000 | 0.000 | 1.167 |
| 0.250 | 0.000 | 0.750 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 2.500 |
| 0.000 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 0.000 | 2.000 |
| 0.000 | 0.000 | 0.833 | 0.167 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | 5.667 |
| | | | | | | | | | |
| -0.446 | -0.371 | 0.834 | 1.684 | 1.644 | 0.398 | 1.986 | 1.166 | 4.691 | |

Table 3: Big10 system of equations for the value of possession and the results

The estimated values of moving from one interval to another (approximately the value of gaining a first down of ten yards) of around .62 and .44 indicate that moving the ball is slightly more valuable in the Big12. In fact, using drive length data collected (see Table 8) these estimates give that the average possesion is likely to yeild about .85 points more in the Big12. This is pretty similar to the difference in average points in all drives where we had data (.5, see Table 4).

These figures seem to demonstrate that there is a steady and likely significant difference between the value of possession between the two conferences. This seems to be especially true near an opposing team's endzone. However, due to our small sample size for values at the ends of the field and the fact that the values of possession that we have calculated use each other, we cannot accomplish significance testing without advance statistical techniques beyond the scope of this paper.

This, however, does not mean that we cannot test for the differences between the average points per drive between the conferences. Unlike our values of possession above, our average points do not depend on other averages. Moreover, average points per drive gives us a good estimation of the value possession. The data above suggests that the average points statistic is always about 1 to 2 points more than the actual value, as the value takes into account the opponent receiving the ball afterwards. However, we still need to account for the fact that in some regions, our sample sizes are relatively small. Therefore, for the following statistical tests, we re - consider the divisions of the football field by splitting the whole football field into three regions in order to get sufficiently large sample sizes $n$ to run typical statistical significance tests.

In football, there is a commonly used idea that a team can either start with good or bad field position. For our purposes, we consider the following three regions: good field position, which consists of team getting the ball between

| t-test $s_1^2 \neq s_2^2$ | $n$ | $\bar{x}$ | $s^2$ | $df$ | t-statistic | $p$ (1 -tail) |
|---|---|---|---|---|---|---|
| Big12 | 186 | 2.478 | 9.883 | | | |
| Big10 | 243 | 1.963 | 8.986 | | | |
| | | | | 388 | 1.717 | .044 |

Table 4: All drives t-test

their own 40 and the opponent's endzone, average field position, which consists of the area between one's own 20 and 39, and the rest to be bad starting field position. Using this framework, we utilize a $t$ - test for differing variances and sample sizes (Welch's t-test) to analyse the data.

We begin by looking at all drives in general with the null hypothesis being that the difference in points scored per drive is 0 and the alternative being that the average in the Big12 is higher. The results are evident in Table 4. The results for the one tail test results are significant at the $\alpha = .05$ level, but the difference of half a point per drive is not too large. The data becomes more interesting as we consider the individual regions. Tables 5,6, and 7 show the results for those tests.

| t-test $s_1^2 \neq s_2^2$ | $n$ | $\bar{x}$ | $s^2$ | $df$ | t-statistic | $p$ (1 -tail) |
|---|---|---|---|---|---|---|
| Big12 | 40 | 2.000 | 9.949 | | | |
| Big10 | 48 | .792 | 4.764 | | | |
| | | | | 67 | 2.048 | .022 |

Table 5: Between the 0 and 19 t-test

| t-test $s_1^2 \neq s_2^2$ | $n$ | $\bar{x}$ | $s^2$ | $df$ | t-statistic | $p$ (1 -tail) |
|---|---|---|---|---|---|---|
| Big12 | 122 | 2.131 | 8.726 | | | |
| Big10 | 152 | 2.151 | 9.626 | | | |
| | | | | 264 | -.055 | .48 |

Table 6: Between the 20 and 39 t-test

| t-test $s_1^2 \neq s_2^2$ | $n$ | $\bar{x}$ | $s^2$ | $df$ | t-statistic | $p$ (1 -tail) |
|---|---|---|---|---|---|---|
| Big12 | 24 | 5.042 | 8.824 | | | |
| Big10 | 43 | 2.605 | 9.721 | | | |
| | | | | 49 | 3.163 | .0013 |

Table 7: Between the 40 and 99 t-test

As expected from our earlier observations of the regression results, the differences in average points per drive when the drive starts in good or bad field position have significant differences between the Big12 and Big10. When teams start behind their own 20 in the Big10, they are expected to earn less than a point while their counterparts in the Big12 gain an average of about 2 points when starting in a similar situation. The most startling results occur in situations of good field position. Here, the results are not only quite significant but also very meaningful. The difference in the average amount of points scored from drives starting in this region (2.4) is almost the value of an entire field goal. This large difference in offensive performance most likely is the driving factor behind the average differences when looking over all the drives. When we see these results, one might expect that these differences are in part due to drive length differences. Thus, we conducted a test to see if this is indeed the case.

| t-test $s_1^2 \neq s_2^2$ | $n$ | $\bar{x}$ | $s^2$ | $df$ | t-statistic | $p$ (1 -tail) |
|---|---|---|---|---|---|---|
| Big12 | 186 | 35.930 | 839.7 | | | |
| Big10 | 243 | 28.975 | 809.3 | | | |
| | | | | 394 | 2.483 | .0068 |

Table 8: Drive length t-test with FG estimated end at 25 t-test

Since we do not have exact data for where a drive ending in a field goal occurs (this was marked as kickoff from the 35), we use the estimate of the opponent's 25 as an average end point for field goal drives (not surprisingly, the averages for setting this end to 35 are (34.8,28.2) with $p = .009$, to 15 are (37.1,29.8) $p = .005$, regardless, the differences are approximately the same and significant). The results seem to indicate that drives in the Big12 last on average 6 to 7 yards more than those in the Big10. While this might not seem to be a huge difference, this amount could easily make or break whether a team enters field goal range or scores a touchdown when it starts with excellent field position. We also note that these values are likely to slightly underestimate drive lengths as a drive must end at the opposing side's goal-line. From these values, a Big12 team that starts at their own 40 is expected to end up at the opponents 24, well within field goal range. A Big10 team in a similar situation finds themselves at the 31, most likely near the edge of most kickers' range (48 - 49 yard attempt).

# 4 Implications

The most immediate implication of this paper is that it seems to suggest that the common perception that the Big12 conference features more high-powered and offensively focused teams (or weaker defense) is at least partially grounded in reality. The average 0.5 extra points per drive gives about a touchdown more points per game in the Big12 on both sides of the ball if we assume each team gets roughly 3 drives a quarter. However, this is likely to be an underestimate of the point differentials in games as many Big10 teams feature a run heavy offense, resulting in drives taking less time in the Big12 and more in the Big10. In addition, the results in both estimating the values of possession the field position framework seem to indicate in general higher offensive efficiency (or worse defense) when points are on the line. This large differential in average points on the plus side of the field probably accounts for most of the differences in average points per drive.

From our data on drive distances, we also infer that it is more likely for Big10 games to fall into a “punt war” situation where teams end up punting back and forth. A vast number of drives start from one’s own 25 or 20 (touchback). Using our estimates for average drive lengths, a Big10 team that starts on their 25 is likely to end up in the vicinity of the opponents 45, “no man’s land”. Because of the high likelihood of a touchback, it is easy to see how this process can repeat. Even if a team ends up being pinned at their own 5, they are likely to advance the ball enough where an average punt one again puts their opponent somewhere near the opponent’s 25 or 30 yard lines (a 35 yard or 40 yard punt from one’s own 35), thus leading to the notorious low scoring affairs in the Big10. This is not to say this could not happen in a Big12 match, but because the average drive lasts longer in that conference, it seems that teams are more likely to escape this cycle.

The most intriguing implication from the results is whether a team should decide to punt or keep its offense on the field. At most levels of play, a team will punt on fourth down if it has the ball on its own side of the ball, despite this being potentially sub-optimal as shown by Romer. For our purposes we consider the fourth down decision when a team is in “no man’s land” a region where punting does not have a chance to massively impact field position but is outside the team’s kicker’s field goal range. This region is generally between the 50 and the opponent’s 35 yard lines. For simplicity we consider fourth down decisions between the 50 and the opponent’s 40 yard lines. For it to be worth it for a team to go for a fourth down conversion here, the value (or by proxy, the average point value) of the try must be greater than punting. We reasonably assume that a punt from this position of the field ends up with the opponent somewhere between the 20 (touchback) and the goal line. Using our framework of good, average, and bad field postion, this translates to:

$$g(p) - g(1-p) \geq -b \tag{4}$$

where $p$ represents the probability of success. For the Big12, this $p$ is about 0.30. For the Big10 this is about 0.35. If we use our values based system, this calculation becomes a little more difficult. We assume the average of $v_6$ and $v_7$ for the value of success, because if the conversion is successful, the offense is in the vicinity of the opponent’s 40 and the average of $v_1$ and $v_2$ for the value of punting. Thus, we obtain the following:

$$.5(v_6 + v_7)p - v_5(1-p) \geq -.5(v_1 + v_2) \tag{5}$$

Using the estimates from the regression (both actual $v_6$ seem to be low outliers), the values are .35 and .47 for the Big12 and Big10 respectively. This roughly indicates that a team in no man’s land in the Big12 should go for it on fourth down if they think their odds are at least 1 in 3. In the Big10, this is probably closer to 2 in 5. With the analytics of today, teams can easily use third down and distance statistics to estimate their chance of success on fourth down. Notice that our estimates do not include the distances to go for the first, this information comes into the decision when a team estimates its probability (ie 4th and 1 might have a .7 success rate while 4th and 7 might be a .1 success rate). The results also fit with the perception that Big10 coaches tend to be more conservative in this region of the field. We would like to indicate that these are rough estimates, however, they are almost certain to overestimate the $p$ needed to go for it as it is likely that too much weight is given to pinning an opponent deep and not enough to the more frequent outcome of punting in this region, a touchback.

# 5 Conclusion

This paper provides evidence to suggest that Big12 schools do produce more on offense than Big10 schools on college football. Specifically, possession of the ball is about a point or two more valuable in the Big12 near a team's own goal-line and about a field goal (3 points) more once a team reaches midfield or beyond. A consequence of this is that a 10 yard gain in field position is worth about .2 points more in the Big12. We also find evidence that suggests that the average Big12 drive goes for a longer distance (by around 6-7) yards than one in the Big10. We then use these estimations to estimate when a team should elect to keep its offense on the field on fourth down.

Nevertheless, more research into this field and college football in general is still warranted. One interesting question would be to look into the value of possession and scoring differentials across the entire college football landscape or at least among the Power 5 conferences (Big12, Big10, Pac12, SEC, ACC). A survey that incorporates larger data sets could certainly be helpful in gaining greater insight into the questions posed in this paper

There are also questions as to why there is such large difference between the two conferences that could use additional research. Perhaps one explanation for this is the effective field goal range difference between the Big12 and the Big10. Unlike the NFL, the vast majority of college games are played in outdoor stadiums. This potentially makes weather a factor in many kicking situations and limit the effective range of Big10 kickers, reducing the number of points scored on the plus side of the field. Regardless of whether this is the case, more research into this can hopefully shed light on this question.

Finally, one should be careful in extending our work to make conclusions about the likelihood of winning games. While it does seem to be the fact that Big12 teams generate more offense when they hold the ball, this makes points relatively less valuable in these matches. On the flip side, less points scored in the Big10 means that every point is relatively more valuable in these games.

Acknowledgments

The author would like to thank Professor Lawrence Martin at Michigan State University for his guidance and assistance throughout the process of writing this article.

# References

Rice, John, A. *Mathematical Statistics and Data Analysis.* Cengage Learning, 3rd ed., 2006.

Romer, David. 2002. "It's Fourth Down and What Does the Bellman Equation Say? A Dynamic-Programming Analysis of Football Strategy." Working Paper, National Bureau of Economic Research.

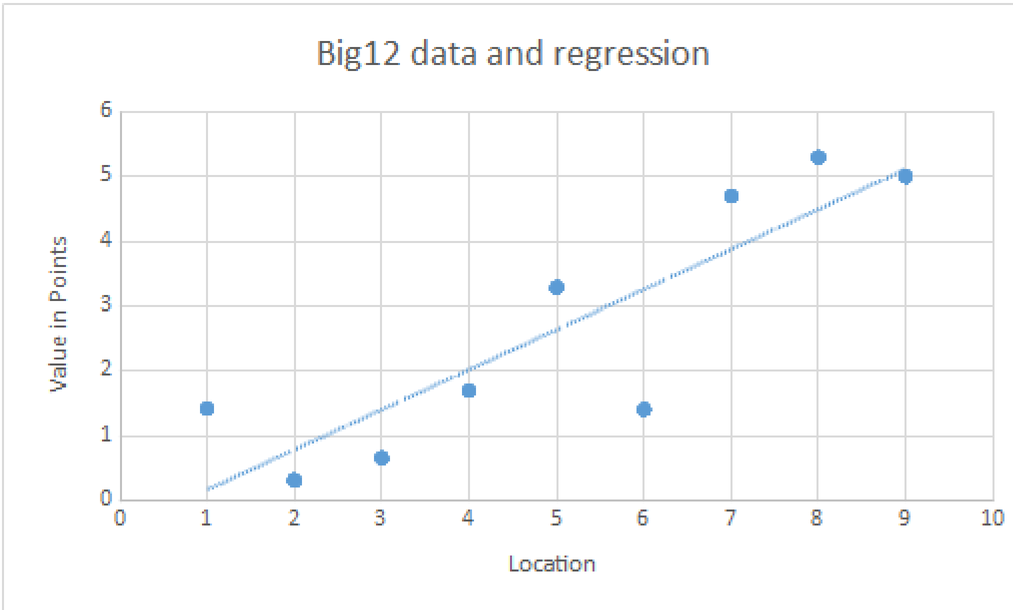


Figure 1: Regression and data for possession values in the Big12

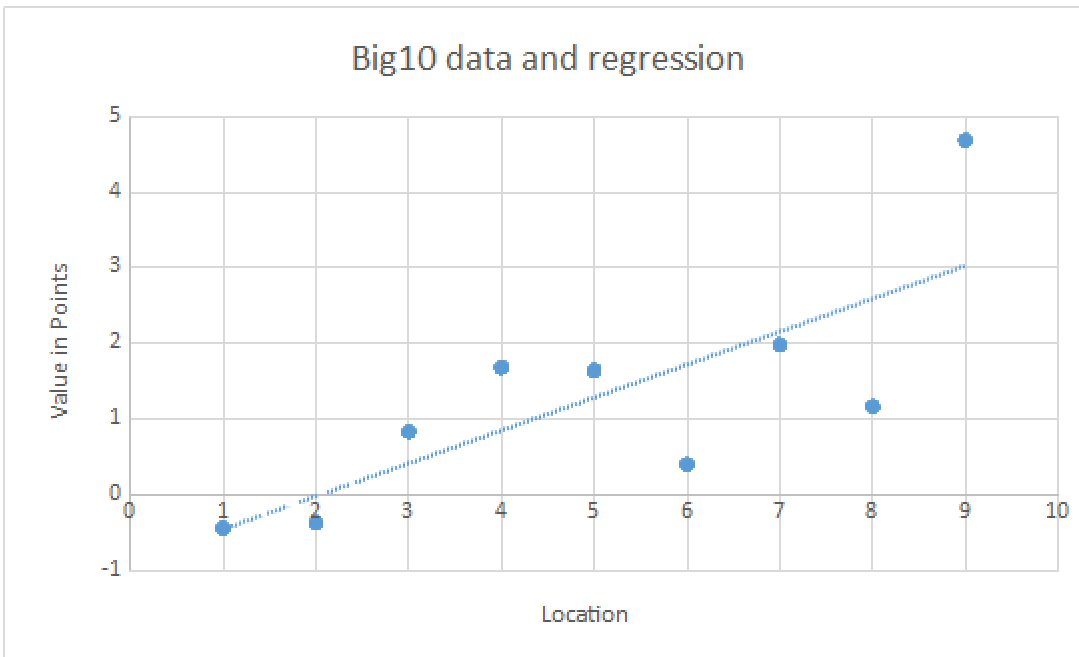


Figure 2: Regression and data for possession values in the Big10

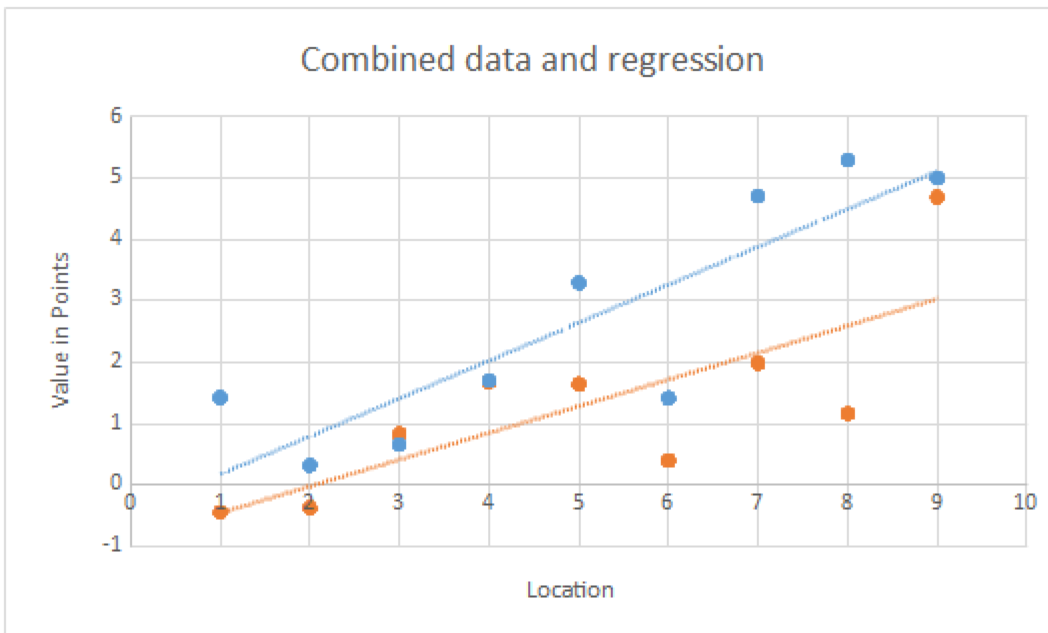


Figure 3: Previous figures displayed on shared axes