The goal of the playoffPredictor website is to use statistical analysis in the absolute simplest terms to:

#1 Accurately Rate and Rank each FBS team, and

#2 Predict what the playoff committee will list as their top 4 teams each week

The largest overall fundamental premise to this method is the simplicity. The method aims to use no inputs or weights that are arbitrary. Many categories that on the surface would seem non-arbitrary are in fact arbitrary – examples include points scored, total yards gained, margin of victory, etc. The reason that these metrics are bad choices for a statistical method is because the method maker has to determine what constitutes "good" for those categories. To some people a 21-point margin of victory can be considered good, but to others it may not and this leads to subjectivity in any computer-based formula.

The only input that is not arbitrary is wins and losses. Football is a team sport and at the end of the day the only thing that is necessary and sufficient to rank all teams is wins and losses against the scheduled played. A winning outcome is the goal of the team and the only metric that need be considered to make a good model on ranking teams.

So how do we map this to college football? The best model is the one proposed by Pierre-Simon Laplace in 1814. Yes, that Laplace you studied in college math of Laplace transformation fame. Of course, Laplace did not apply his method to college football, since he preceded it by a couple hundred years. Instead Laplace sought to answer the question "Will the sun rise tomorrow?" Which turns out to be a perfect fit to answer the question "Will my football team win next week?". Confused? Read on to understand.

A football team winning or losing a single game is a binomial probability. In each trial (football game) there is only success (W) or failure (L). Will the sun rise tomorrow can also be modeled as a binomial – success (it will rise), or failure (it will not rise). Now, having no further insight except that the trial has to end in success or failure you can model this probability by

$$\mathbf{Pr}(\textbf{sun will rise tomorrow}) = \frac{d+1}{d+2}$$

where d represents the number of times the sun has risen in the past. So, on day 1 when God created Adam and Adam wondered if the sun would rise tomorrow, he would have computed the probability as ½ - having no prior data and no knowledge of the workings of gravity, etc – it's a 50/50 shot. On day 2 after a successful first sunrise the odds improve to 2/3, and by now with 3,000,000 days where humans have documented the sun did indeed rise yesterday, the odds for tomorrow improve substantially to

$$\frac{3{,}000{,}001}{3{,}000{,}002} = 0.999999667 \approx 1$$

Now of course Laplace had insight to say that if you understand the mechanics of gravity and planetary motion you can make a much better guess as to the true probability- but this is exactly where the tie in to college football comes in. Because college football plays a small sample size (a season is just ~12 games long), the s+1 / n+2 model fits the situation perfectly.

In general for n_w wins and n_l losses, the formula for rating at team is

$$r = \frac{1 + nw}{2 + nl + nw}$$

where nw can be rewritten as

$$nw = \frac{nw - nl}{2} + \frac{nw + nl}{2}$$

For the second term above effectively becomes the ratings of the teams opponents. Instead of using r=1/2 for all opponents, use their actual ratings and then

$$n_{w,i}^{eff} = (n_{w,i} - n_{\ell,i})/2 + \sum_{j=1}^{n_{tot,i}} r_j^i,$$

where r_{ij} is the rating of the j^{th} opponent of team i.

To figure out the r_{ij} which are inputs to the r_i , which are themselves other r_{ij} s an iterative scheme is used. The way it works is one first computes the ratings, as if all the opponents were r = 1/2 teams. Next, each team's strength of schedule is computed according to its opponents' ratings. The ratings are re-computed with the new schedule strengths, and then strengths of schedule are re-computed from the new ratings.

The end result will assign each team a probability of success in a random football game. This probability will be somewhere between 0 and 1 (mostly). We computer rank the teams from highest rating to lowest. This part of the method will not be explained in further detail here because Wes Colley has already documented it in great detail going back to the 1990s. You can read more on this part of the method at his site, www.colleyrankings.com

Ratings applied to probability theory

As an aside, once we have ratings for any given 2 teams, we can model the chances of success in a contest of those two teams by using simple statistics. Given team A has a probability of winning P(A) and a losing probability of P(A') = 1 – P(A) and using similar notation for team B there are four probability based outcomes:

Team A wins, Team B loses - P(A) * P(B')
Team A loses, Team B wins - P(A') * P(B)
Team A wins, Team B wins - P(A) * P(B)
Team A loses, Team B loses - P(A') * P(B')

Of course, under the rules of football scenarios 3 and 4 are not possible in a head-to-head match, so we are left with computing probabilities for the first two scenarios.

For example, if team A has a 0.9 rating [P(A}=0.9] and team B has a 0.4 rating [P(B)=0.4], then

$$P(A)\&\& P(B') = 0.9 * 0.6 = 0.36$$
$$P(A')\&\& P(B) = 0.1 * 0.4 = 0.04$$

scaling those 2 outcomes to 100% (because the other 2 scenarios are impossible)

$$\frac{P(A) * P(B')}{P(A) * P(B') + P(A') * P(B)} = \frac{0.36}{0.36 + 0.04} = \frac{0.36}{0.4} = 90\%$$

OK, so now we have our mathematical method to computer rate teams. We can use this method each week as we get more trial outcomes (games), and therefore assign a more correct winning probability to each team. Great! But this does not answer the question that needs to be answered, specifically what 4 teams will make the top 4 spots at the end of the season? In order to do that we have to recognize the committee is composed of humans and humans have biases. The question then becomes "how can we measure the biases of the committee?". Once we accurately know their biases we can add those back into the computer rankings – and the computer will spit out the predicted committee rankings for the subsequent week.

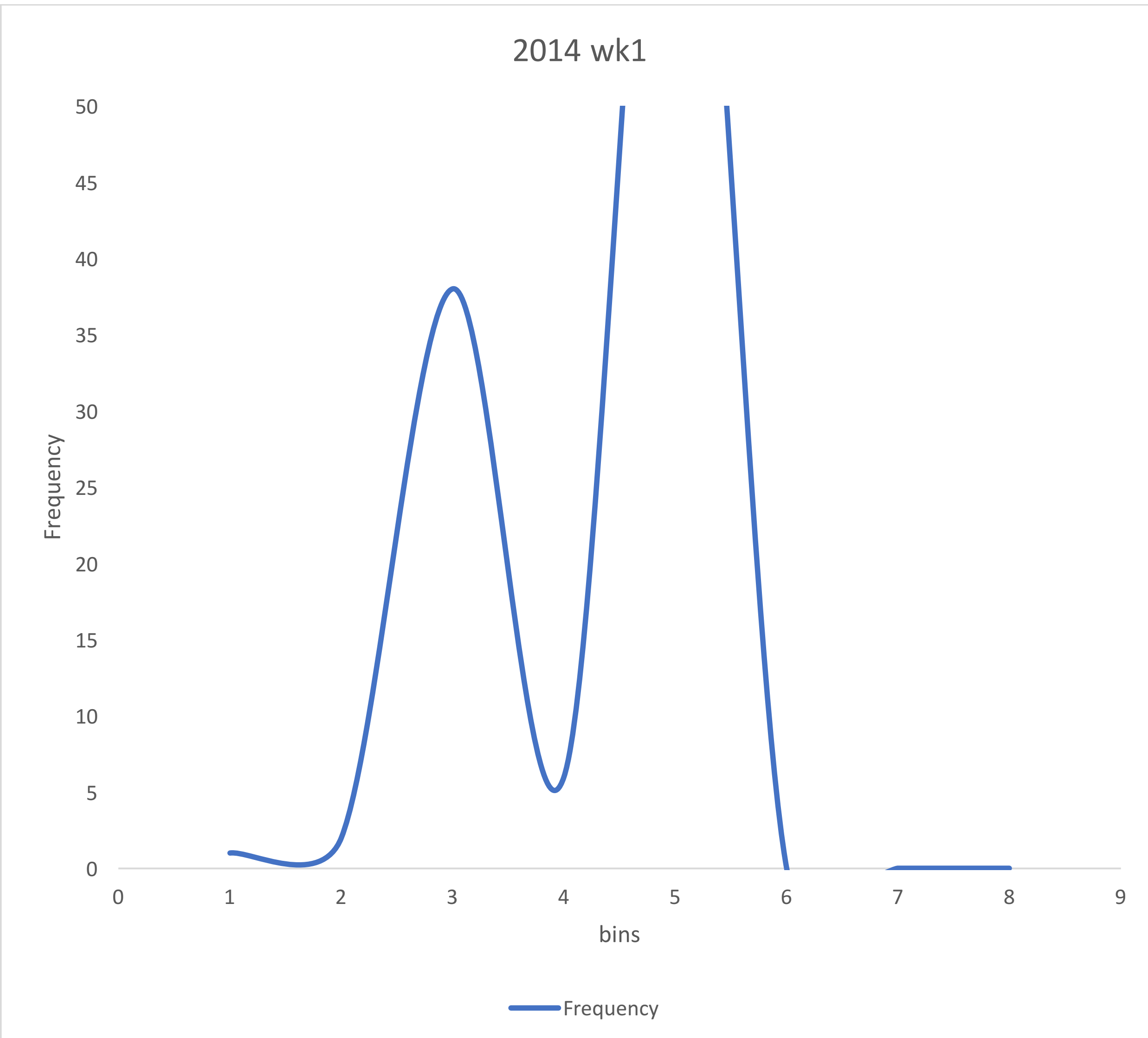
Bias calculation 2014-2016

So how do we go about measuring the committee's bias for a specific team in a simple, statistical way?

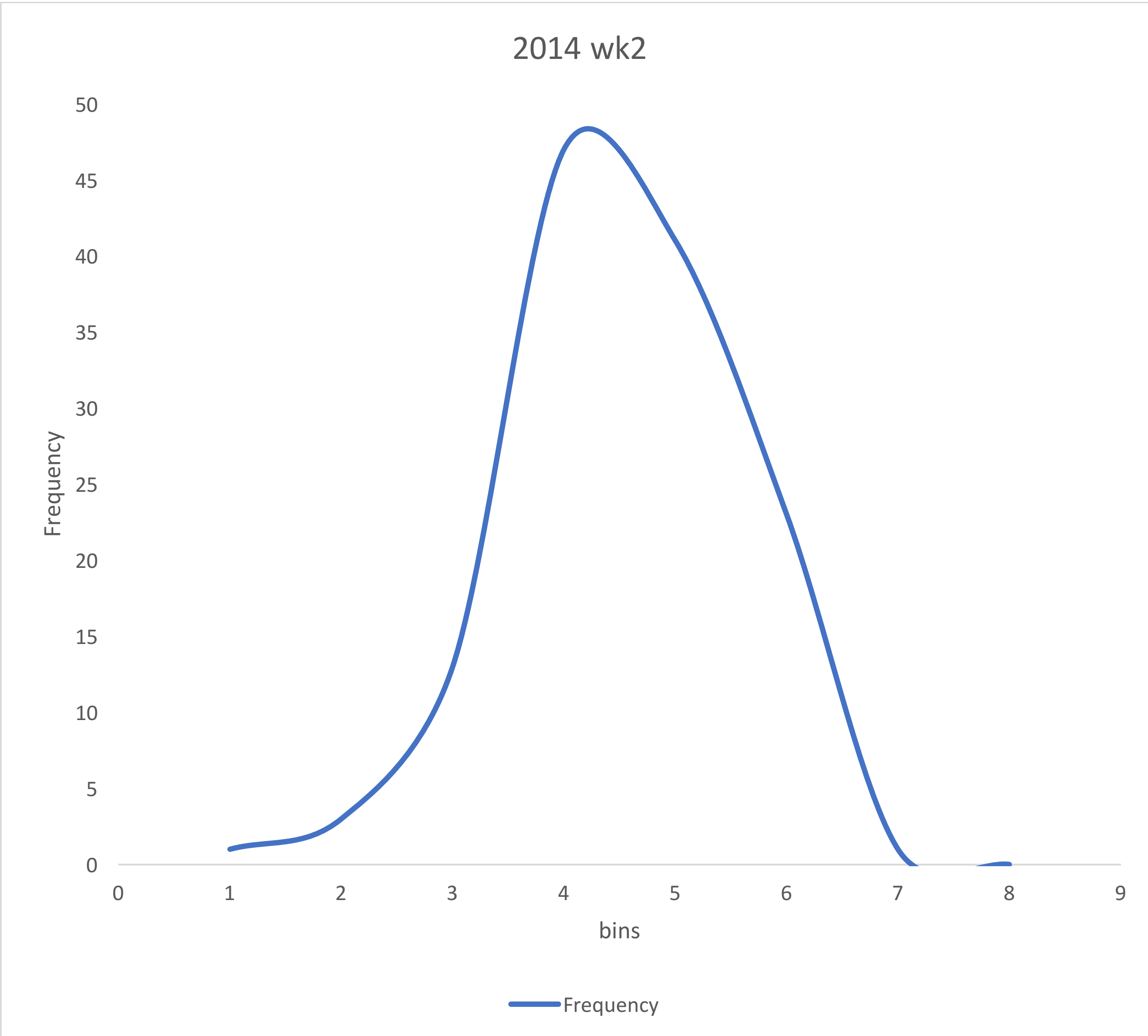
If you look at the data closely, you will notice that the ratings of teams are normally distributed (binomial distribution technically, but binomial -> normal as n gets large). We can visualize teams' ratings using histograms. Again, ratings will be between 0 and 1 (mostly), centered at 0.5 (nearly), and there are 129 teams in FBS (generally)*. First we use Sturges' formula to determine the number of bins for our histogram. Sturges' formula is derived from a binomial distribution

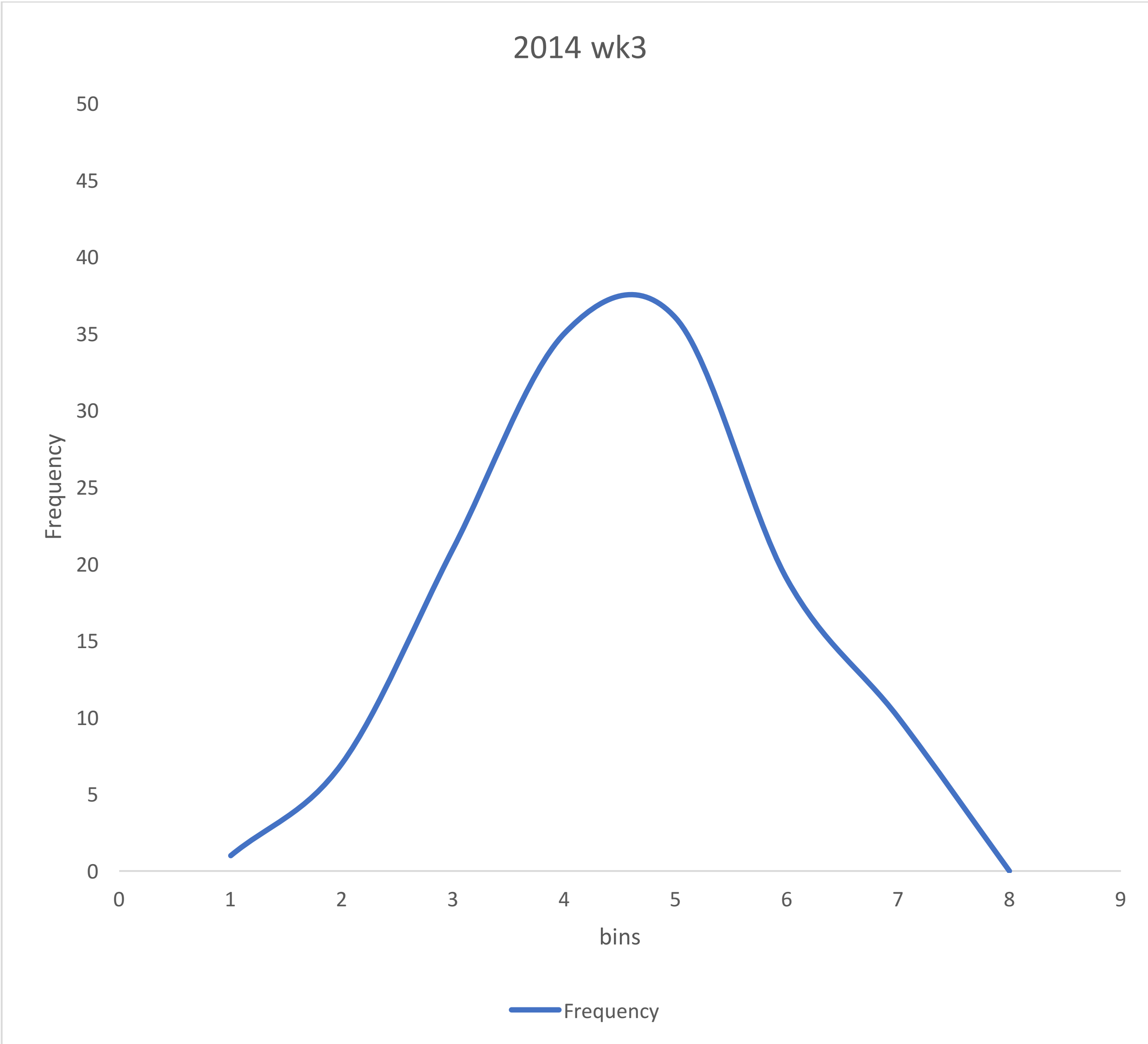
$$k = [log_2 n] + 1$$

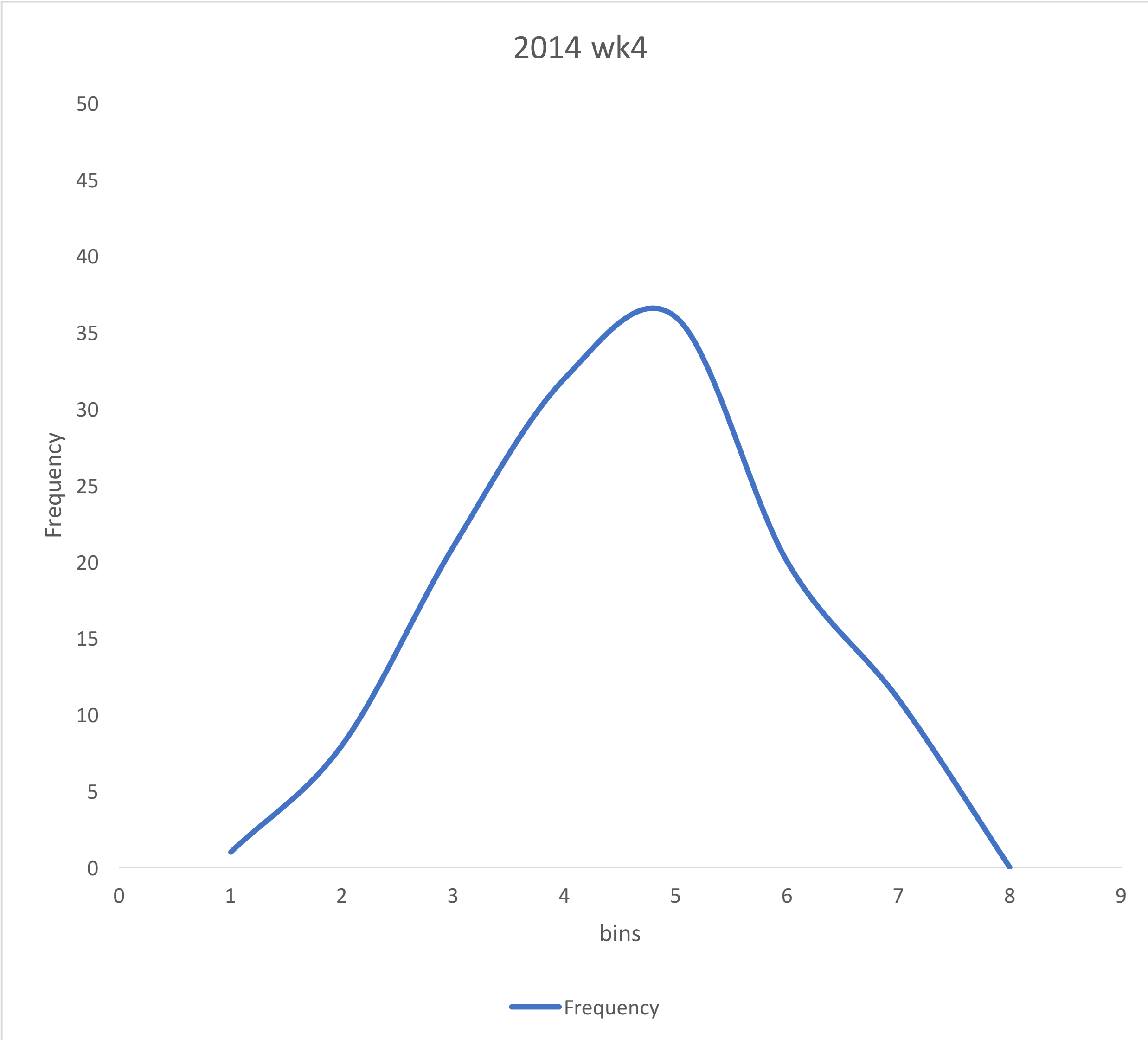
Using n=130 teams, we get k = 8.01, and we will use 8 bins. We expect the data to go from 0 to 1, so the bins are at 0, .125, .25, ..., 0.875. Plotting the data for the 2014 season week by week we get:

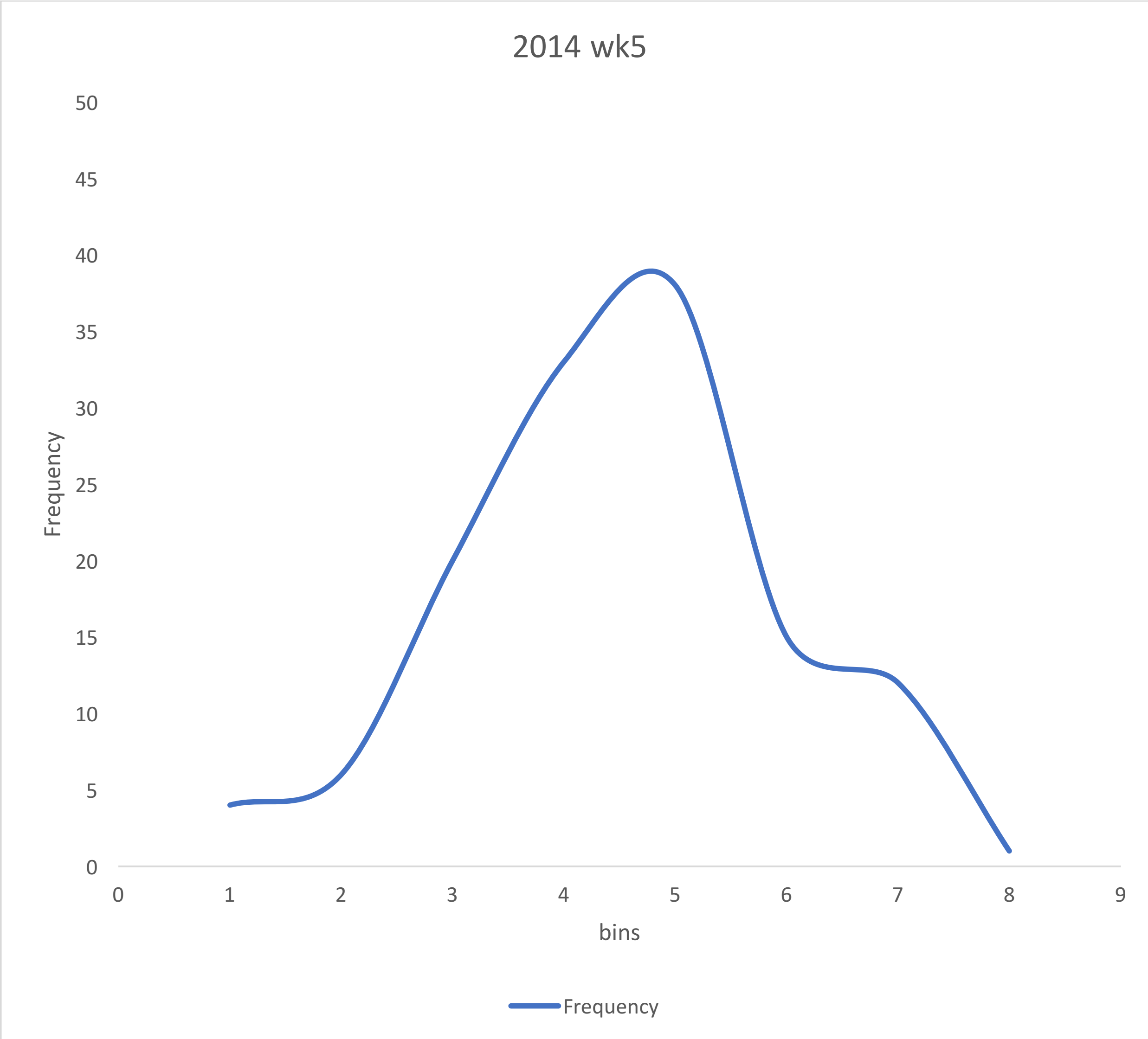


Week 1 – notice we have approximately half of the teams with a 1-0 record and a .625 rating, and another half with an 0-1 record and a .375 rating. The reason the data is skewed to .625 is many teams play FCS teams in week 1, and that 1 FCS team has a bad record / rating (2-40, 0.05).

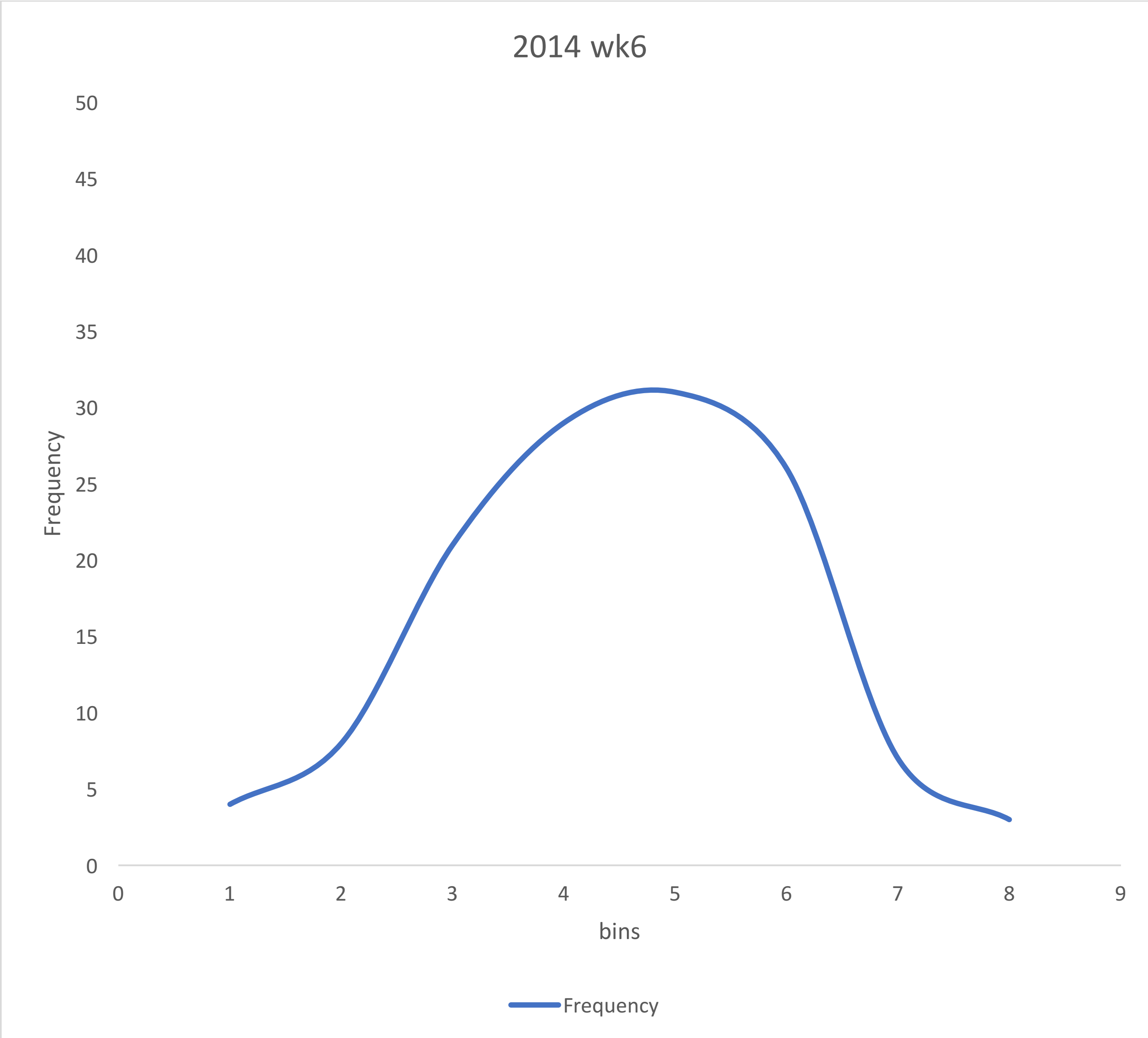


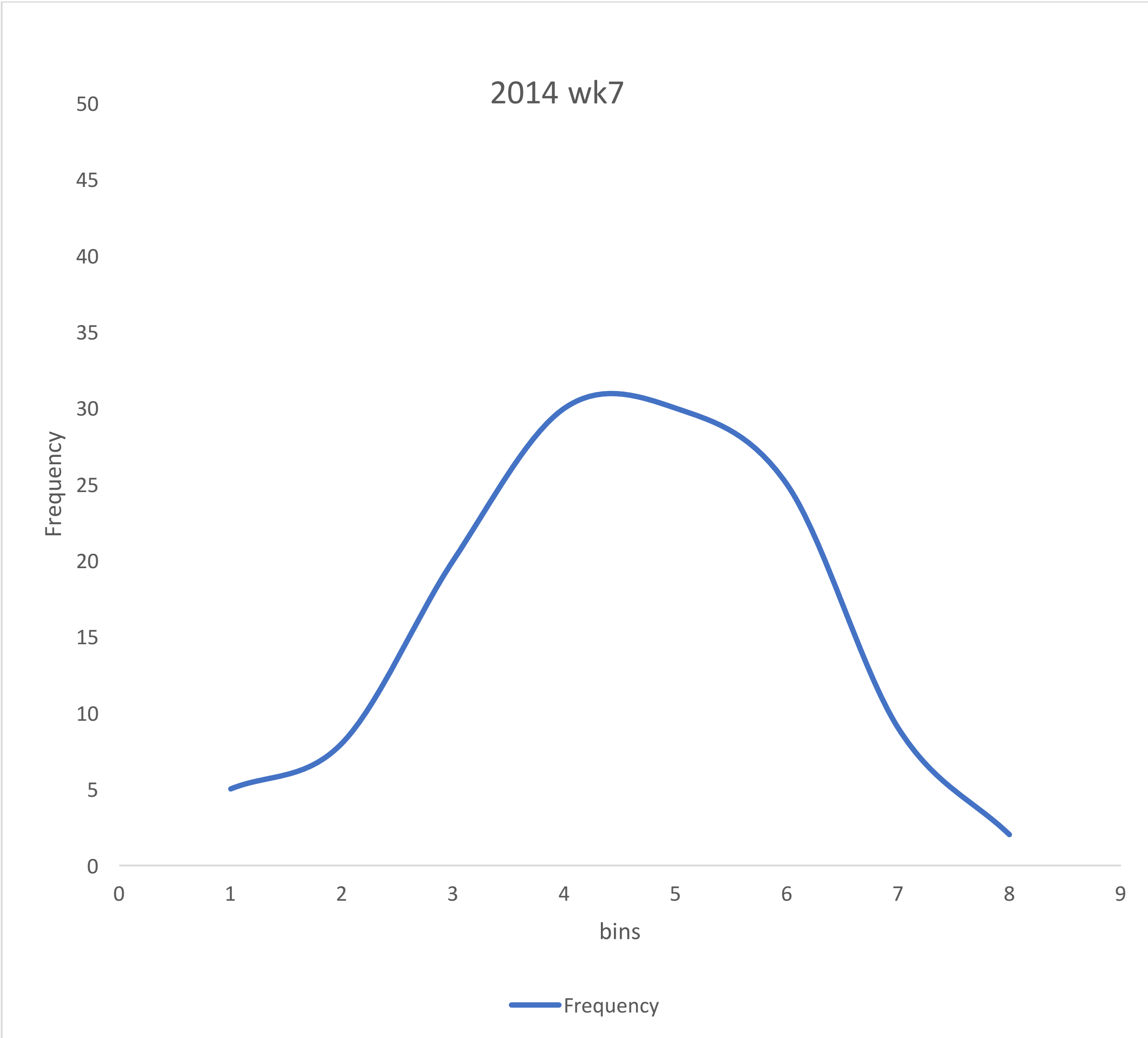

2014 wk3
Frequency
50
45
40
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6
7
8
9
bins
Frequency

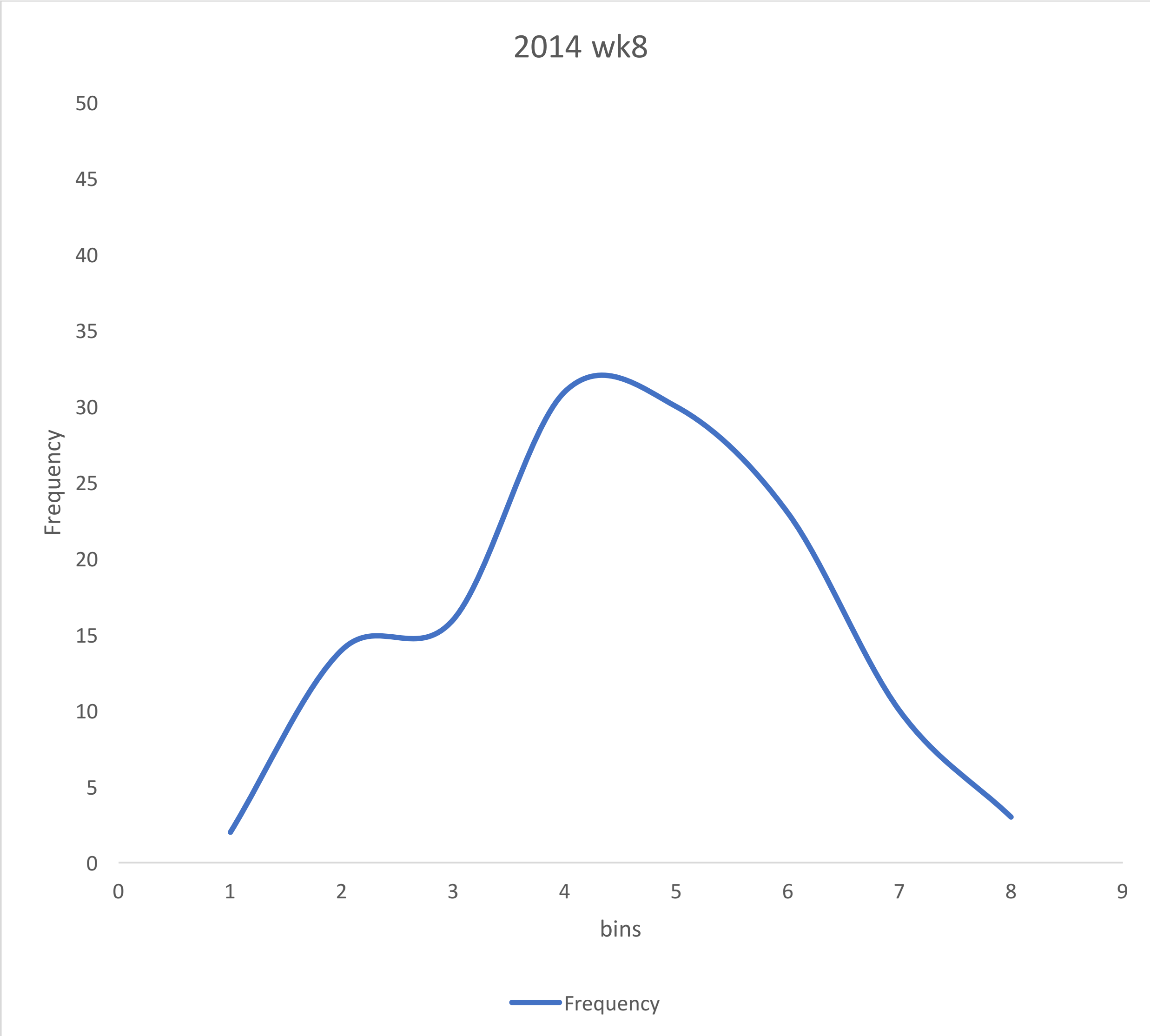

2014 wk4
Frequency
0
5
10
15
20
25
30
35
40
45
50
0 1 2 3 4 5 6 7 8 9
bins
Frequency

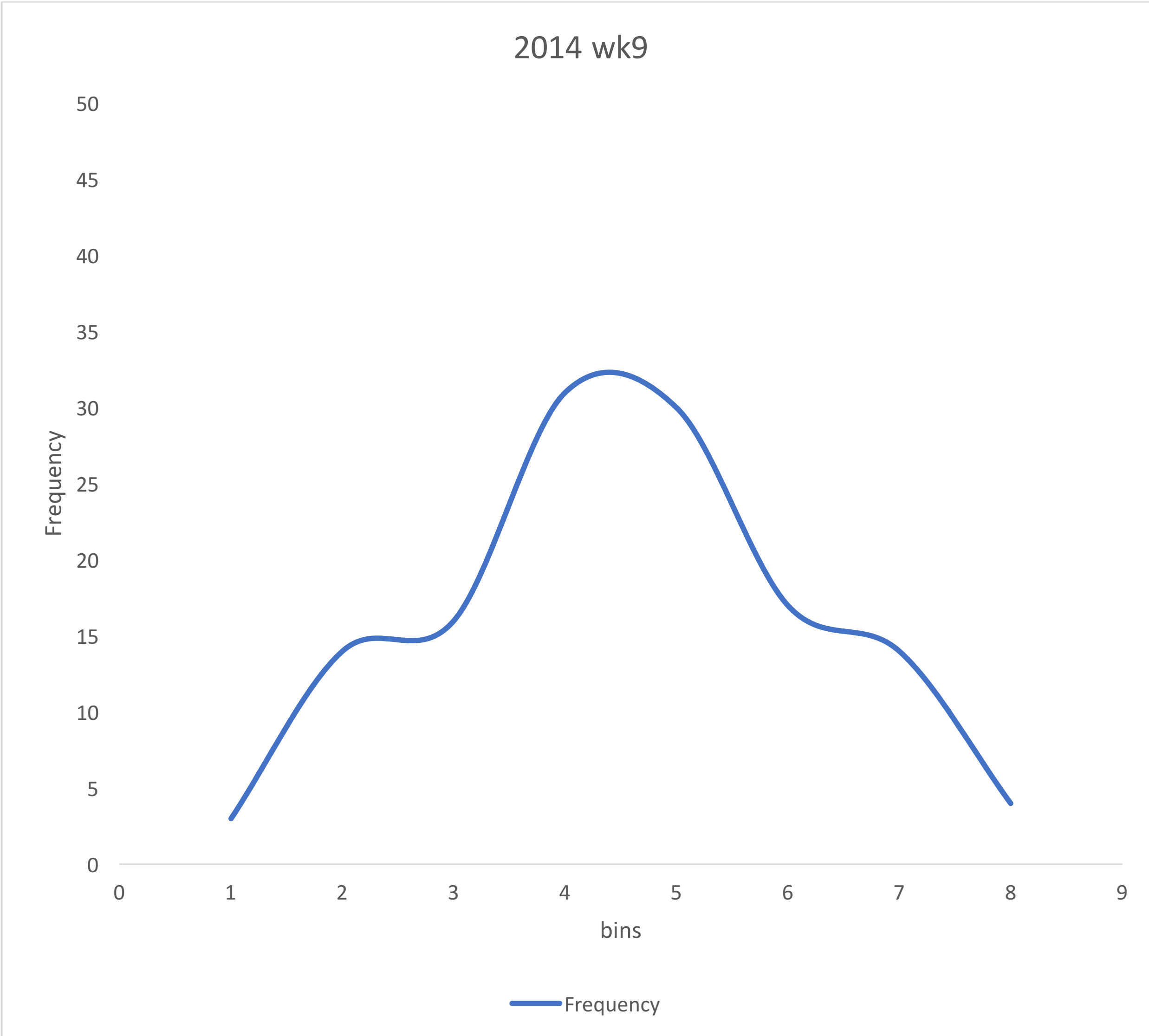


2014 wk5
Frequency
50
45
40
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6
7
8
9
bins
Frequency


2014 wk6
50
45
40
35
30
25
20
15
10
5
0
Frequency
0
1
2
3
4
5
6
7
8
9
bins
Frequency


2014 wk7
Frequency
50
45
40
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6
7
8
9
bins
Frequency


2014 wk8
Frequency
50
45
40
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6
7
8
9
bins
Frequency



with a mean of 0.5 and a standard deviation that progresses from .125 to .25 as the season goes from week 1 to week infinity (source, work here - .125 there are approx. 128 teams = 2^7 and 2 ^ -3 = .125, no that doesn't work...)

Using this method, the ratings in an ideal season of the top teams will be:

The measured ratings are 2008-2014 year historical Colley ratings for that rank number at that week

	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15
#1	0.973	0.987	0.971	0.963	0.972	0.988	1.000
#2	0.933	0.950	0.944	0.939	0.940	0.954	0.965
#3	0.918	0.920	0.927	0.910	0.921	0.922	0.932
#4	0.908	0.902	0.910	0.896	0.900	0.903	0.906
#5	0.893	0.894	0.890	0.890	0.891	0.892	0.891
#6	0.867	0.878	0.872	0.872	0.879	0.877	0.873
#7	0.851	0.854	0.857	0.864	0.865	0.868	0.866
#8	0.839	0.843	0.846	0.856	0.848	0.857	0.859
#9	0.833	0.828	0.841	0.846	0.842	0.845	0.853
#10	0.824	0.818	0.831	0.838	0.838	0.840	0.847
#11	0.819	0.813	0.817	0.826	0.831	0.834	0.836
#12	0.805	0.808	0.811	0.814	0.825	0.825	0.830
#13	0.797	0.802	0.804	0.807	0.818	0.814	0.814
#14	0.788	0.794	0.792	0.803	0.803	0.799	0.800
#15	0.781	0.789	0.782	0.794	0.798	0.795	0.784
#16	0.776	0.781	0.774	0.783	0.784	0.789	0.776
#17	0.767	0.770	0.764	0.776	0.777	0.776	0.767
#18	0.761	0.765	0.758	0.765	0.771	0.765	0.762
#19	0.754	0.757	0.753	0.756	0.763	0.753	0.752
#20	0.749	0.746	0.744	0.752	0.754	0.748	0.746
#21	0.740	0.733	0.741	0.747	0.748	0.735	0.738
#22	0.731	0.726	0.733	0.738	0.728	0.726	0.727
#23	0.727	0.723	0.726	0.732	0.721	0.719	0.717
#24	0.719	0.716	0.720	0.720	0.710	0.709	0.712
#25	0.715	0.707	0.714	0.715	0.703	0.706	0.707

So the bias would be to match each teams rank from the committee to the above table. For example, at week 10 if Auburn was ranked #3 by the committee, they would have an implied rating of 0.920. If their calculated rating was 0.900 , then their bias would be +0.020.

The bias is averaged over the course of a year and used to calculate the predicted committee rankings after games are played on Saturday, but before committee rankings are released on Tuesday.

Bias calculation 2017-

For the years 2017 onwards the method will assign a bias to a team based on the difference from the committee's ranking and the computers ranking, using the rating difference between those two rankings.

Using an example to illustrate, say the computer ranks Penn State #1 with a rating of .955 and the computer ranks Alabama #2 with a rating of .920. That same week if the committee ranks Alabama #1 and Penn State #2, then Alabama will be assigned a bias of .955-.920 (=+.035) and Penn State will be assigned a bias of .920-.955 (=-.035).

For a different example, assume the committee rankings are #1 Alabama, #2 Georgia, and #3 Penn State, while the computer rankings and ratings are #1 Penn State=0.955, #2 Alabama=0.920, and #3 Georgia=0.900 then the bias for each team would be:

- Alabama = 0.955-0.920 = +0.035
- Georgia = 0.920-0.900 = +0.020
- Penn State = 0.900 – 0.955 = -0.055

if in both the computer poll and committee poll Georgia is ranked #2, the bias is 0, regardless of what the computer rating of Georgia is.

Just like the 2014-2016 method, the bias is averaged over the course of a year and used to calculate the predicted committee rankings after games are played on Saturday, but before committee rankings are released on Tuesday. Bias is not carried over from year to year.

Notes on post 2017 method:

I have decided to move to this method because there is such a difference in computer rating between an undefeated team and a 1 loss top team. Because the committee only assigns rankings and not numerical ratings, it is very hard to capture this difference year to year. For example, there is a strong difference between 2016 Alabama at week 14, with a 13-0 record and ranked #1 and 2015 Michigan State at week 14, with a 12-1 record and also ranked #1. This method is the easiest way to capture that difference.