**random.R**

**I. Analyzing Contest Results**

**a.** Theory: Let P = {P\_1,...,P\_N}, where P is the set of N players s.t. any 9 players of this set sums to ~200+ fpts (i.e. earns decent payout). We suspect N is decreasing week over week b/c coaches are finding their most reliable players for the season, and these players get the most touches.

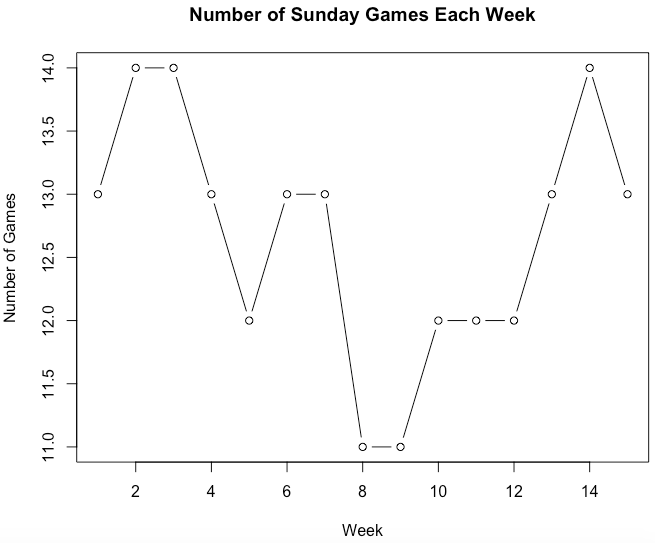
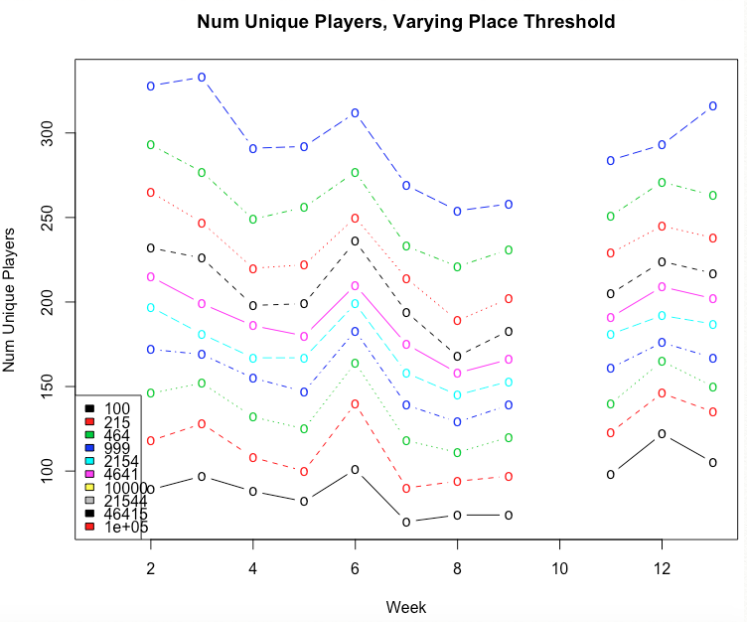


Figure 1 (note: colors in legend are incorrect; increasing threshold black (100) -> blue (10,000))

Doesn’t necessarily support our theory because set of unique players increases for all thresholds (top 100 lineups, top 215, etc). Interesting pattern of ~decreasing up to week 9 and then ~increasing afterwards. Let’s see if we observe this pattern for any relevant statistics.

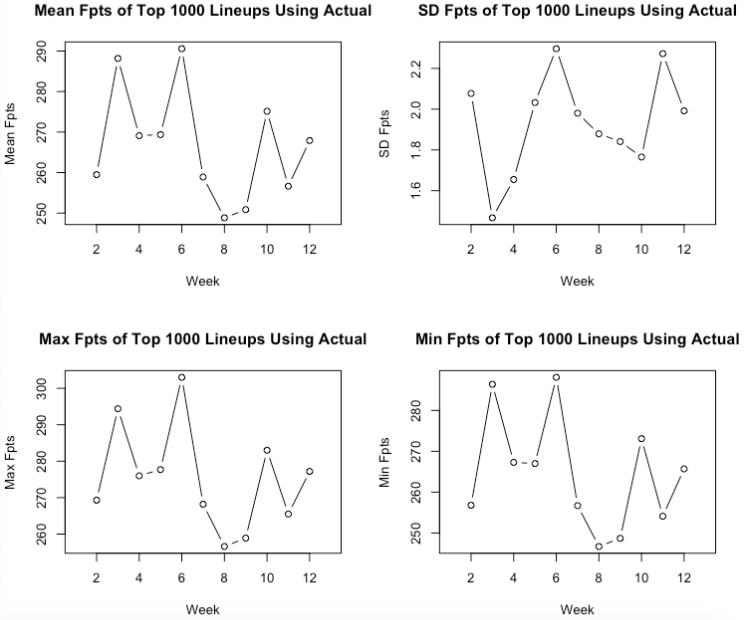


Figure 2

Looks like mean, max, and min decreased between weeks 6-8 just like in Figure 1. Not clear for other weeks. Let’s see if there’s a similar pattern for lineups generated by our formulations.

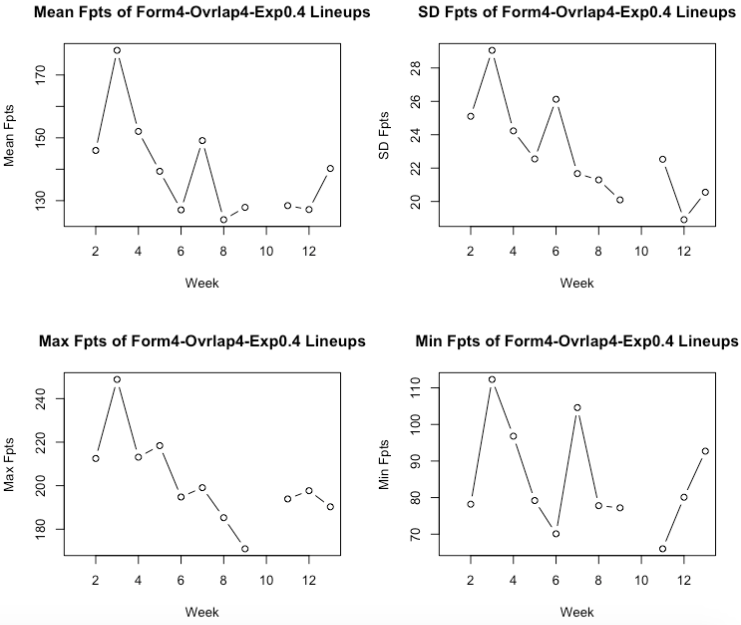


Figure 3

It seems that there’s also a correlation between the mean and max scores of lineups generated by Formulation4-Overlap4-Exposure0.4 (DFN predictions). We test this formulation because it wins $1M in week 3 (so it is of special interest) and, also, we theorize that it does well when the number of unique players among cashing lineups is high (since it doesn’t have any constraints that would restrict the set of players it chooses). Let’s quantify the correlations we’ve pointed out.

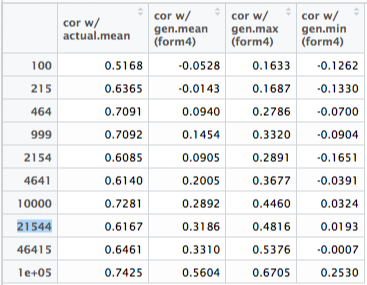


Figure 4

Indeed, there is some positive correlation between Figure 1 and Figure 2-Mean as well Figure 1 and Figure 3-Mean (max, too). Correlation between Figure 1 and Figure 2-Mean is especially high, but that isn’t particularly surprising. Note that the 21,544 threshold place is the approximate cutoff for cashing lineups. Not sure exactly what to make of this, but it seems like the positive correlation between Figure 1 and Figure 3-Mean supports the argument that Formulation4-Overlap4-Exposure0.4 does well with a higher number of unique players and worse with a lower number of unique players. We can confirm this checking if the correlation is high relative to the correlation with the mean score for another formulation, perhaps formulation 13 because it aims to reduce the set of players by constraining to top 3 target leaders.

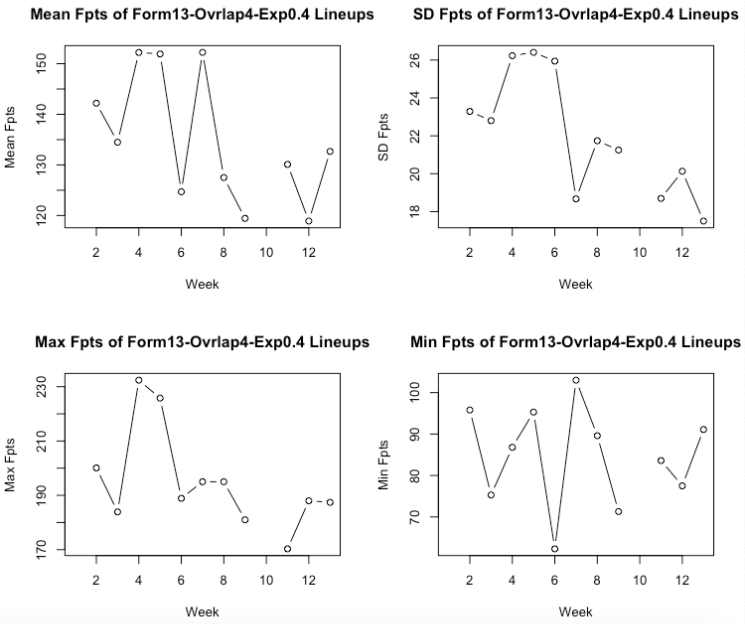


Figure 5

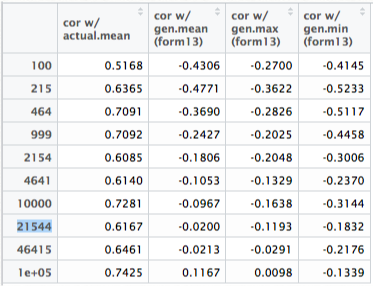


Figure 6

Notice the negative correlations. As we kind of expected, when the number of unique players decreases, the mean score using formulation 13 increases, and vice versa. This seems to make sense because formulation 13 reduces the set of players the algorithm chooses from, albeit reducing the set solely using a top 3 target leaders likely doesn’t yield the best set. Interestingly, the magnitude of the correlations are higher for lower threshold places and lower for higher thresholds, which is the opposite of what we observe in Figure 4. Furthermore, we notice a much higher correlation (in magnitude) with the min score, which also contrasts Figure 4. Perhaps this means that using the top 3 target leader constraint helps with reducing the number of terrible lineups (so the min is higher).

Anyhow, the correlations in Figure 5 appear to be quite high now that we have Figure 6 for comparison. This supports our understanding of when formulation 4 performs well (i.e. weeks where the number of unique players is high). Intuitively, formulation 4 should perform well relative to formulation 13 (or some other formulation following a similar strategy) in weeks with a high number of unique players because it only uses basic constraints that keep the set of available players quite large. (Note: Even though we may have confirmed this, we still need to figure out how to make money off of it. We don’t necessarily know when the number of unique players will be low. As mentioned earlier, our initial theory that the number of unique players decreases throughout the season doesn’t appear to hold.)

Just to confirm, we will compute the correlations for a few more formulations:

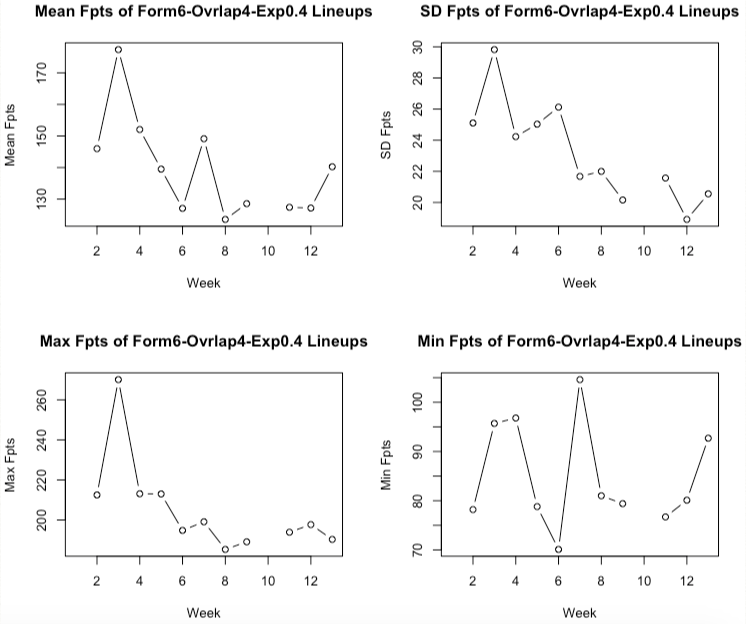


Figure 7

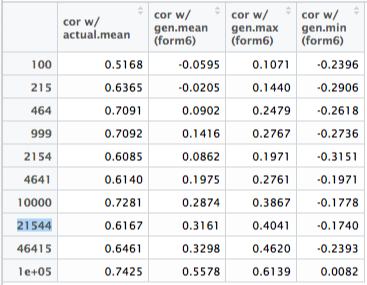


Figure 8

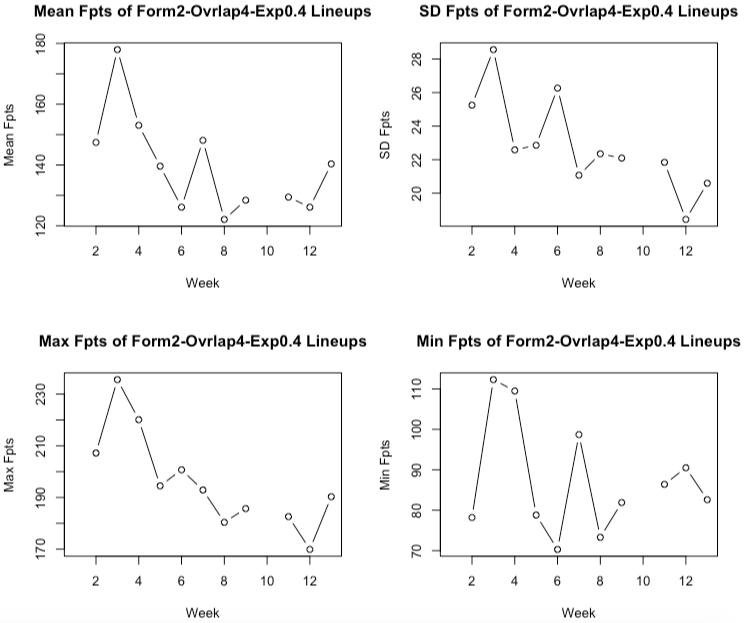


Figure 9

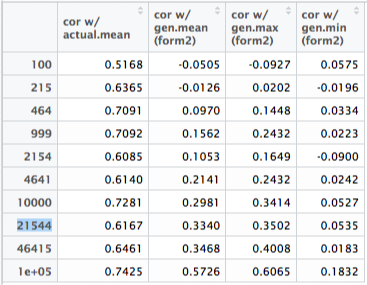


Figure 10

In Figures 8 and 10, notice the positive correlations. It appears that formulation 6 and formulation 2 are similar to formulation 4 in that the correlation with the number of unique players in positive. The negative correlations in Figure 6 suggest that formulation 13 differs significantly from formulations 4, 6, and 2 (in their connection to the number of unique players), which makes sense because among these four formulations, formulation 13 is the only one that dramatically reduces the set of available players (using the top 3 target leader constraint). Reducing the set of available players in a more astute manner may be worth looking into — if we (1) know what to expect for next week’s number of unique players (which we don’t have a good way to do at this point besides eyeballing the trend), and (2) find that the PnL of formulation 13 in weeks with a low number of unique players is significantly greater than the PnL of the other three formulations (to do).

Let’s also compare to the correlation with the mean score for Formulation4-Overlap4-Exposure0.4 with use\_Freq\_Ind = true, since the purpose of using use\_Freq\_Ind is to decrease the set of players the algorithm can select from (Freq\_Ind is 1 if the player has done well historically in the season and 0 otherwise).

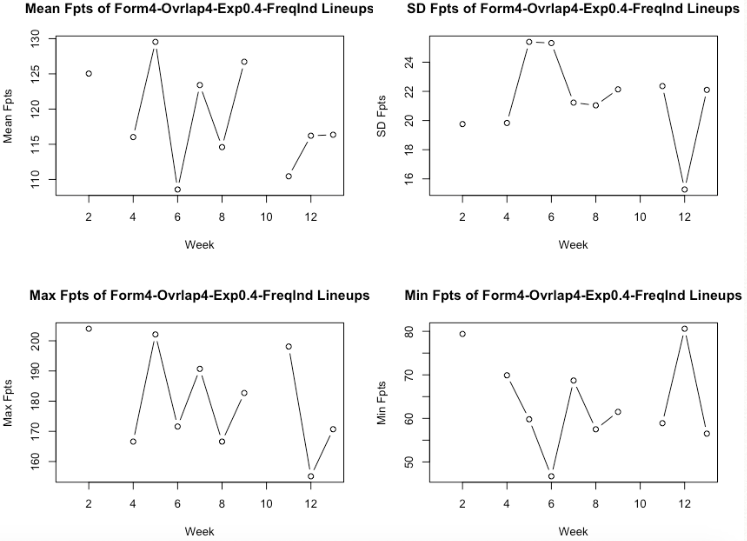


Figure 11

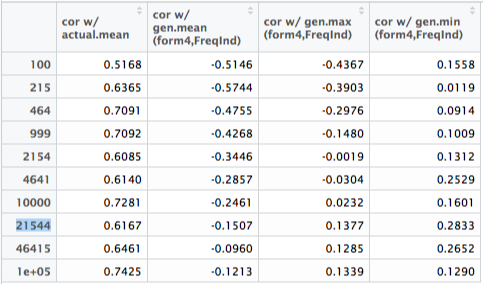


Figure 12

Notice the negative correlations between the number of unique players and the mean and max of lineups generated by Formulation4-Overlap4-Exposure0.4-FreqInd. Seems pretty clear that as the number of unique players decreases, the mean and max increase, and vice versa. Again, these findings support our claim that reducing the set of available players performs better when the number of unique players decreases. We observed this in Figure 6 (Formulation13-Overlap4-Exposure0.4) and Figure 12 (Formulation4-Overlap4-Exposure0.4-FreqInd). Given these correlation relationships, we theorize we should use formulation 13 or some formulation with FreqInd in weeks with a lower number of unique players in cashing lineups. Formulations 2, 4, and 6 (without FreqInd) are likely better for weeks with a higher number of unique players in cashing lineups. This theory needs to be confirmed by PnL analysis (next).

Side note: In weeks 2-9, there is a downtrend in the number of unique players in cashing lineups. This may be explain why Formulation2-Overlap4-Exposure0.4, Formulation4-Overlap4-Exposure0.4, and Formulation6-Overlap4-Exposure0.4 performed worse and worse week over week, while Formulation4-Overlap4-Exposure0.4\_FreqInd and Formulation13-Overlap4-Exposure0.4 perform better or stay about the same week over week (in terms of mean, min, and max fpts, as well as PnL).

Let’s look at the PnLs of each strategy week over week. Note that the “$50K to 1st” contest entry fee was $4 in week 10 and $3 in all other weeks. Also note that the “$1M to 1st” contest entry fee was $20 before week 10 and changed to $27 starting in week 10. We can’t think of a good way to standardize the PnL across entry fees — we can’t just compute the PnL for all weeks using a single entry fee because the payout structure differs. So, we simply compute all PnLs using the respective contest entry fee and note that the entry fees were not the same across all weeks. If necessary, we can attempt to evaluate results in a more standardized manner using the fantasy points — one metric we can compare across weeks is (mean score of our lineups) - (mean score of cashing lineups).

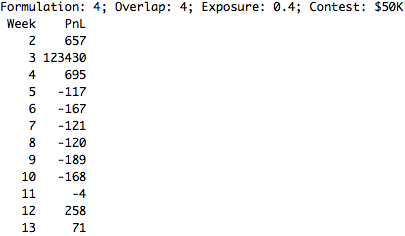


Figure 13

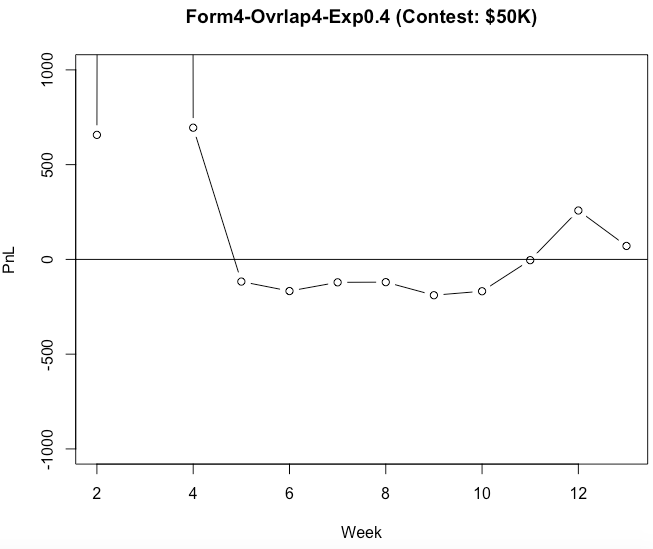


Figure 14

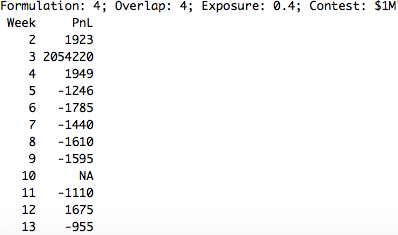


Figure 15

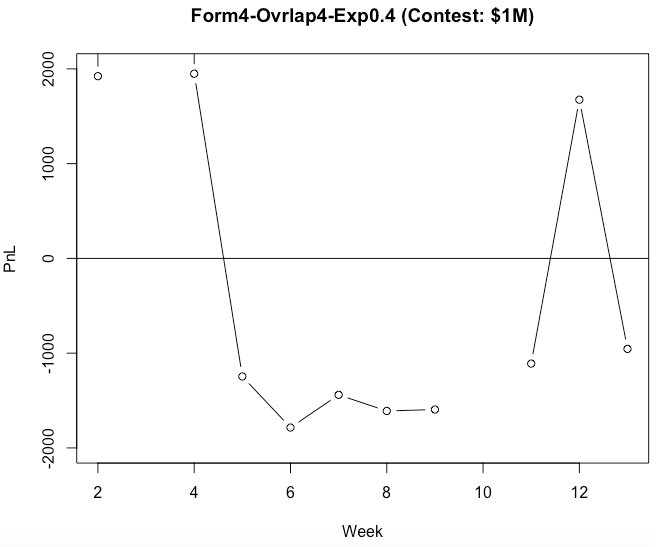


Figure 16