Optimizing number of substitute teachers to be hired given substitute pay rate and overtime teacher pay rate.

Robert Statham

1 Introduction

The problem presented is for us to find the optimal size of a given school district's pool of substitute teachers that minimizes cost to the district. A given school district will hire a certain number of substitutes, S, for a given day of the week. Also, if the school district does not hire enough substitutes on a given day, a regular teacher will have to fill that need but at a higher to the district. We are given data for a given day of the week and the number of substitute teachers demanded and the frequency at which that number of substitutes was demanded. Also, we were numbers: p and r which represent the amount of pay for substitutes and the daily overtime rate for regular teachers respectively. Letting x represent the number of substitutes needed on given day, we can represent the cost to the district, C, as the following:

$$C(x,S) = \begin{cases} pS & \text{if } x < S \\ pS + (x-S)r & \text{if } x \geq S \end{cases}$$

In this project, the given school district is Los Angeles County School District, the given day of the week is Tuesday, and we are assuming $p < r$. Data for Tuesdays is given:

Demand for substitute teachers on Tuesdays

Number of teachers	Relative percentage	Cumulative percentage
201–275	2.5	2.5
276–350	2.5	5.0
351–425	5.0	10.0
426–500	7.5	17.5
501–575	12.5	30.0
576–650	17.5	47.5
651–725	42.5	90.0
726–800	5.0	95.0
801–875	2.5	97.5
876–950	2.5	100.0

2 Methodology

We will be running N number of simulations determine the average cost to the district for each pool size and plotting cost vs. pool size. We will then visually determine the pool size that had the lowest cost and then take an interval around said value and repeat until we have selected an interval of pool sizes such that the difference between the costs of all elements of the interval are less than one $10.

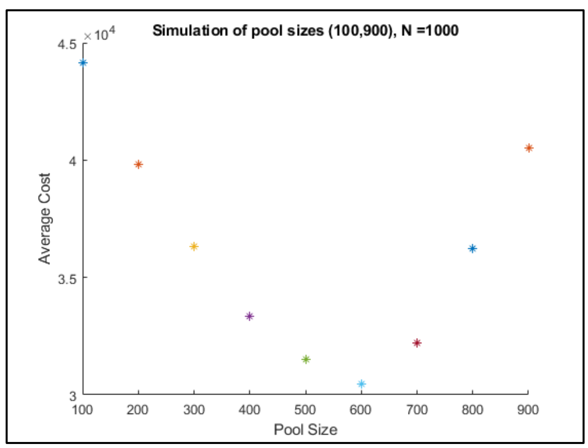
2.1 GIVEN INITIAL VALUES FOR P AND R, WE DETERMINE THE OPTIMAL NUMBER OF SUBSTITUTES.

We are given $p = \$45$, and $r = \$81$. We wrote an algorithm to calculate and plot the average cost per pool size as follows:

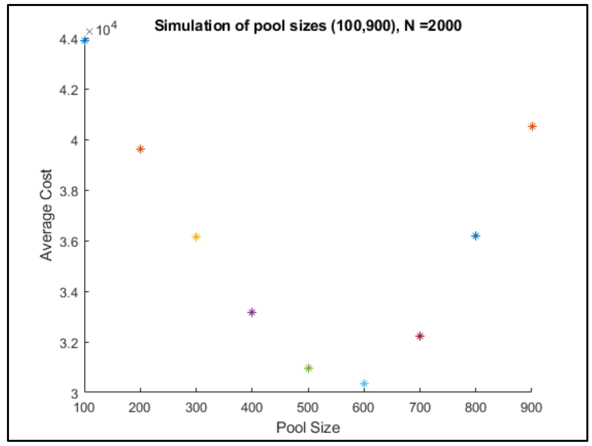
```
N = 1000;    %number of simulations
p = 45;
r = 81;

a = [201 275 350 425 500 575 650 725 800 875];  %left endpoints
b = [275 350 425 500 575 650 725 800 875 950];  %right endpoints
m = [0.025/37 0.025/75 0.05/75 0.075/75 0.125/75 0.175/75 0.425/75 0.05/75 0.025/75 0.025/112];
cp= [0.025 0.05 0.10 0.175 0.30 0.475 0.90 0.95 0.975 1];    %cumulative percentage
hold on;
for S=100:100:900   %different S sizes
    C = 0;
    for k = 1:N
         %total cost
        rd = rand;
        if rd<cp(1)
            x0 = (a(1)+b(1))/2; y0=cp(1); x = x0+(rd-y0)/m(1);
        elseif rd<cp(2)
            x0 = (a(2)+b(2))/2; y0=cp(2); x = x0+(rd-y0)/m(2);
        elseif rd<cp(3)
            x0 = (a(3)+b(3))/2; y0=cp(3); x = x0+(rd-y0)/m(3);
        elseif rd<cp(4)
            x0 = (a(4)+b(4))/2; y0=cp(4); x = x0+(rd-y0)/m(4);
        elseif rd<cp(5)
            x0 = (a(5)+b(5))/2; y0=cp(5); x = x0+(rd-y0)/m(5);
        elseif rd<cp(6)
            x0 = (a(6)+b(6))/2; y0=cp(6); x = x0+(rd-y0)/m(6);
        elseif rd<cp(7)
            x0 = (a(7)+b(7))/2; y0=cp(7); x = x0+(rd-y0)/m(7);
        elseif rd<cp(8)
            x0 = (a(8)+b(8))/2; y0=cp(8); x = x0+(rd-y0)/m(8);
        elseif rd<cp(9)
            x0 = (a(9)+b(9))/2; y0=cp(9); x = x0+(rd-y0)/m(9);
        else
            x0 = b(10); y0=cp(10); x = x0+(rd-y0)/m(10);
        end    %end if
        if x<S
            Cxs = p*S;
        else
            Cxs = p*S + (x-S)*r;
        end
        C = C + Cxs;    %adds up the cost for each simulation
    end
    C = C/N;    %takes the avg
    plot(S,C,'*');
    title(['Simulation of pool sizes (100,900), N =' num2str(N)])
    xlabel('Pool Size')
    ylabel('Average Cost')
end
```

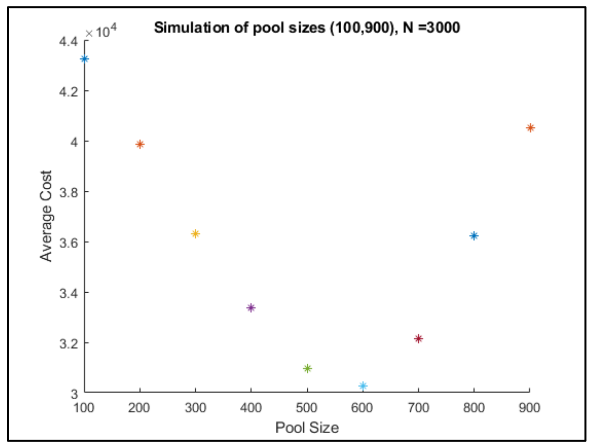
Following the guidelines for solving these particular problems, we examine the plot of the pool sizes: 100, 200, …, 900, starting with 1000 steps shown as follows:



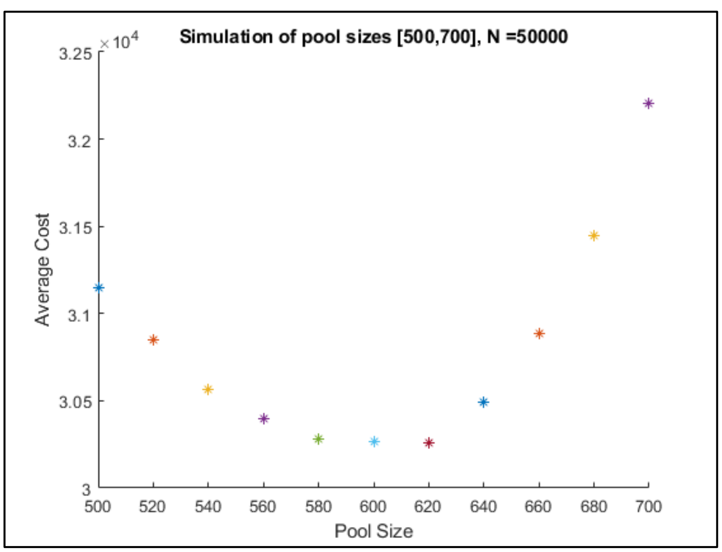
Reasonably smooth, trying *N=2000*:



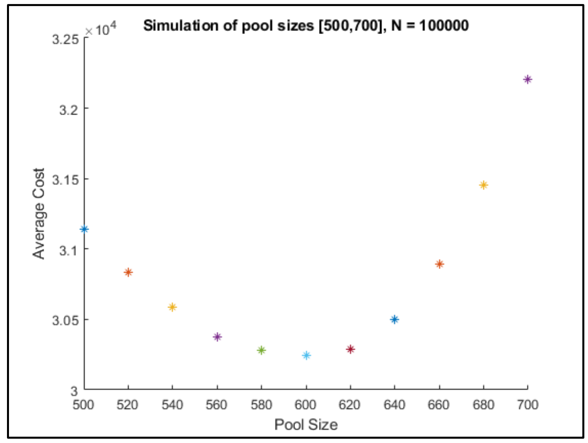
Pretty good, try *N=3000*:



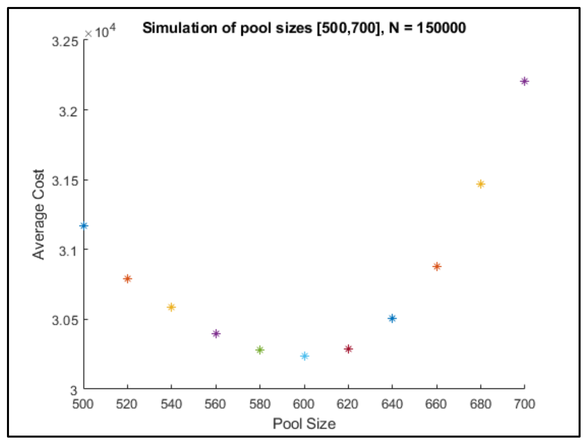
Very smooth, looking around *S=600* where cost seems to be lowest. Now we take the interval $S = [500{,}700]$ with steps of 20. Starting with $N = 50000$:



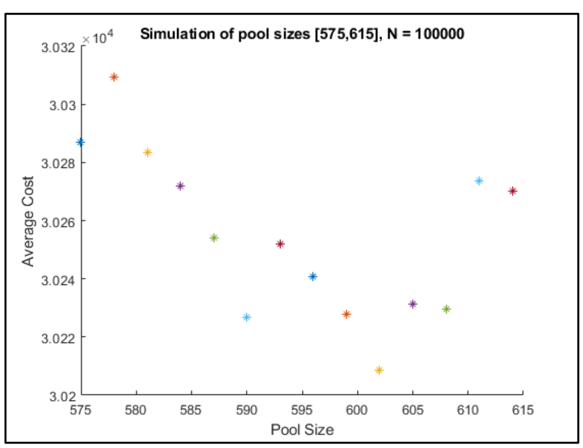
Again, reasonably smooth but could be better, try $N = 100000$:



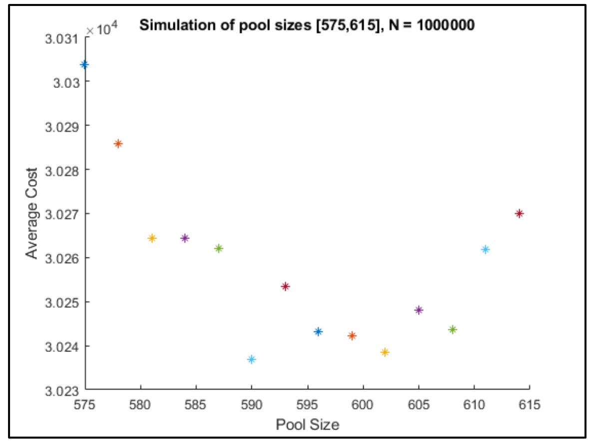
Pretty good, but again, could be better, try $N = 150000$:



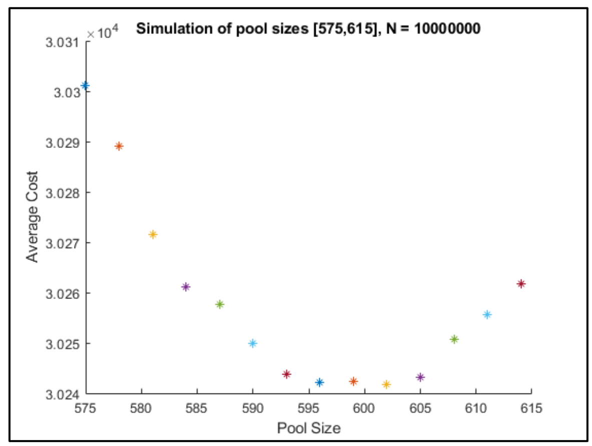
Very smooth, now we can see that at $S = 600$, cost is lowest again. Knowing this, we will now consider the interval $S = [575,615]$ with step sizes of 3. We will start with $N = 100000$:



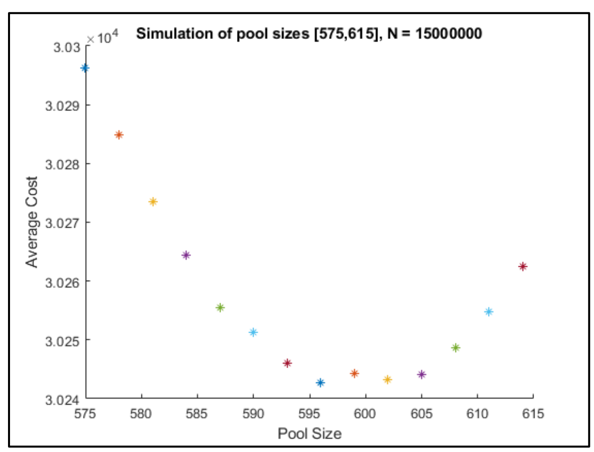
Not smooth at all, we will try to fix this by substantially increasing N to $N = 1000000$:



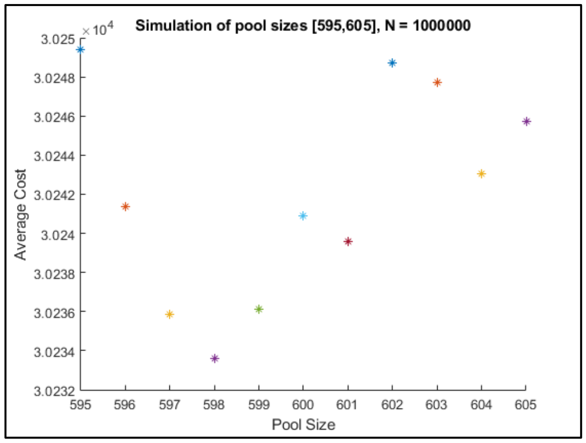
Again, not very smooth so we will go to $N = 10000000$:



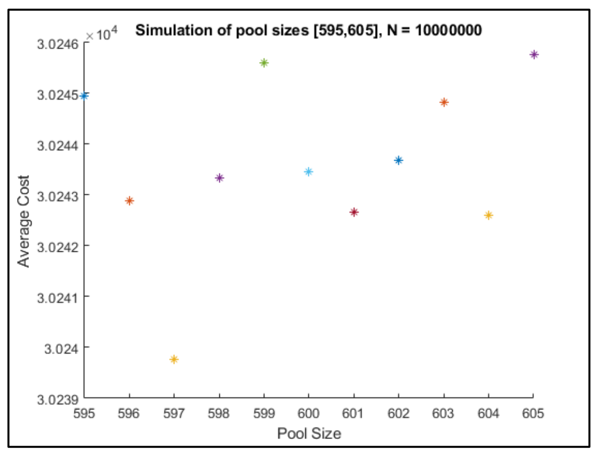
Still not very smooth, trying $N = 15000000$:



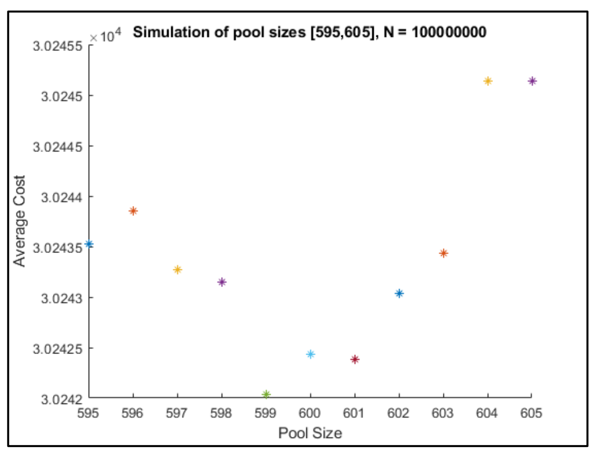
Looks like the minimum values can be found in the interval $S = [595,605]$, so we will now consider that interval with a step size of 1. Start with $N = 1000000$:



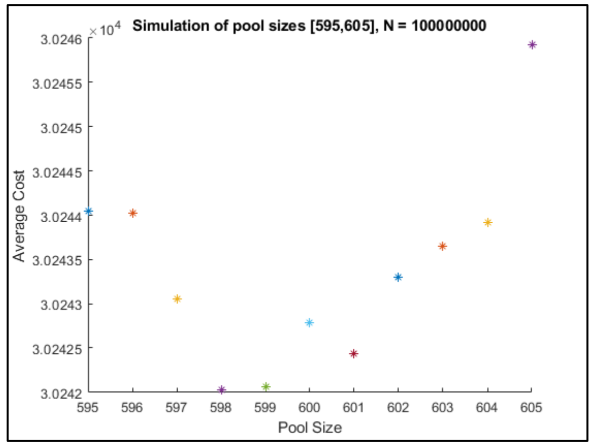
Not smooth at all, we will try $N = 10000000$:



Still not smooth, so we will try $N = 100000000$:



Not smooth again, we will try $N = 100000000$ again:



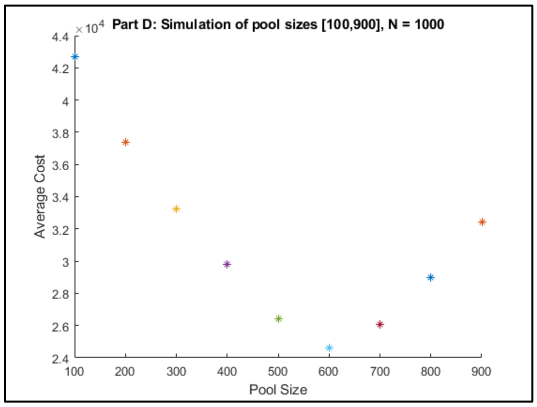
It seems that any selection of $S = [595{,}605]$ will suffice because the difference between the average cost of each is small. I.e., to the nearest cent, there is no difference. So, the Los Angeles County School district can hire anywhere between 595 and 605 substitute teachers on any given Tuesday and that will be the number with the lowest cost to the district. Here, the lowest cost to the district is around $30,243.

2.2 PART D)

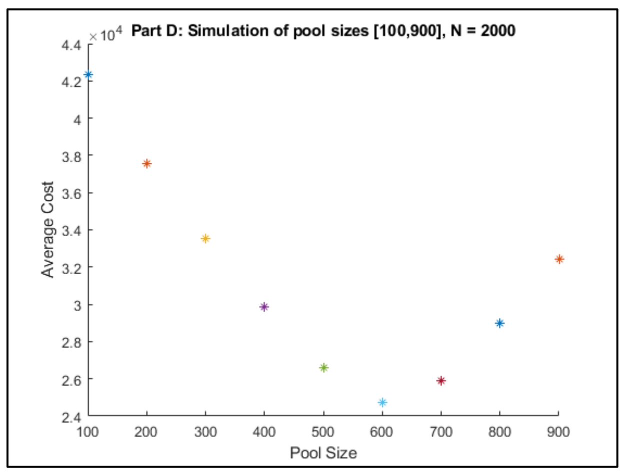
The code used to simulate the cost to the district is very similar to that of part C, except here $p = \$36$.

```
N = 1000;    %number of simulations
p = 36;     %changed from 45 to 36 per part D)
r = 81;
a = [201 275 350 425 500 575 650 725 800 875];  %left endpoints
b = [275 350 425 500 575 650 725 800 875 950];  %right endpoints
m = [0.025/37 0.025/75 0.05/75 0.075/75 0.125/75 0.175/75 0.425/75 0.05/75 0.025/75 0.025/112];
cp= [0.025 0.05 0.10 0.175 0.30 0.475 0.90 0.95 0.975 1];    %cumulative percentage
hold on;
for S=100:100:900   %different S sizes
    C = 0; %total cost
    for k = 1:N
        rd = rand;
        if rd<cp(1)
            x0 = (a(1)+b(1))/2; y0=cp(1); x = x0+(rd-y0)/m(1);
        elseif rd<cp(2)
            x0 = (a(2)+b(2))/2; y0=cp(2); x = x0+(rd-y0)/m(2);
        elseif rd<cp(3)
            x0 = (a(3)+b(3))/2; y0=cp(3); x = x0+(rd-y0)/m(3);
        elseif rd<cp(4)
            x0 = (a(4)+b(4))/2; y0=cp(4); x = x0+(rd-y0)/m(4);
        elseif rd<cp(5)
            x0 = (a(5)+b(5))/2; y0=cp(5); x = x0+(rd-y0)/m(5);
        elseif rd<cp(6)
            x0 = (a(6)+b(6))/2; y0=cp(6); x = x0+(rd-y0)/m(6);
        elseif rd<cp(7)
            x0 = (a(7)+b(7))/2; y0=cp(7); x = x0+(rd-y0)/m(7);
        elseif rd<cp(8)
            x0 = (a(8)+b(8))/2; y0=cp(8); x = x0+(rd-y0)/m(8);
        elseif rd<cp(9)
            x0 = (a(9)+b(9))/2; y0=cp(9); x = x0+(rd-y0)/m(9);
        else
            x0 = b(10); y0=cp(10); x = x0+(rd-y0)/m(10);
        end     %end if
        if x<S
            Cxs = p*S;
        else
            Cxs = p*S + (x-S)*r;
        end
        C = C + Cxs;    %adds up the cost for each simulation
    end
    C = C/N;    %takes the avg
    plot(S,C,'*');
    title(['Part D: Simulation of pool sizes [100,900], N = ' num2str(N)])
    xlabel('Pool Size')
    ylabel('Average Cost')
end
```

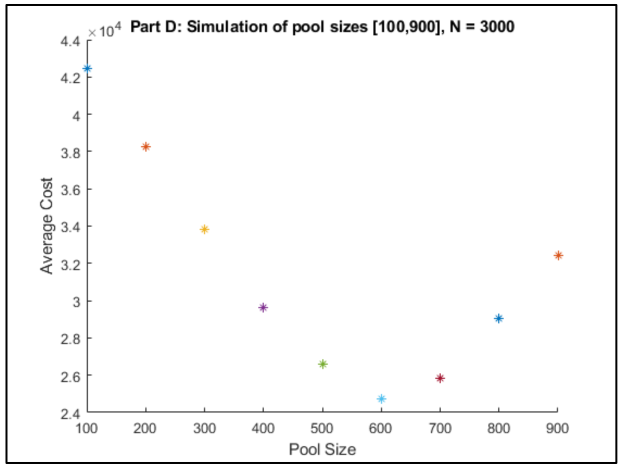
Similarly, to part C, we will first consider the interval: $S = [100{,}900]$ with a step size of 100 where $N = 1000$:



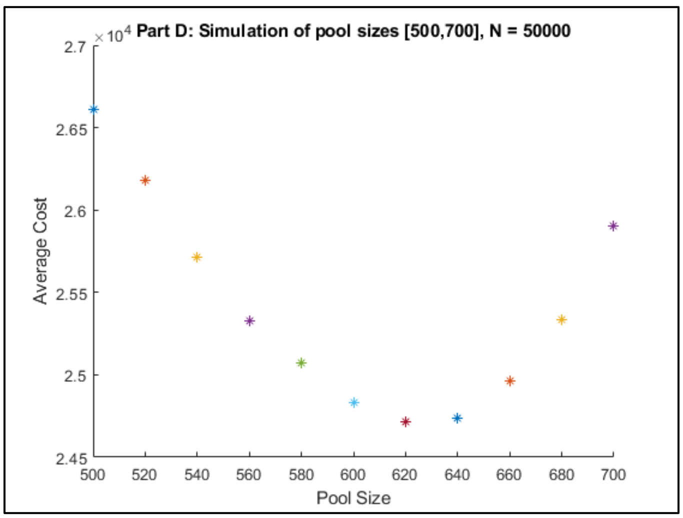
This is reasonably smooth, but we will try $N = 2000$:



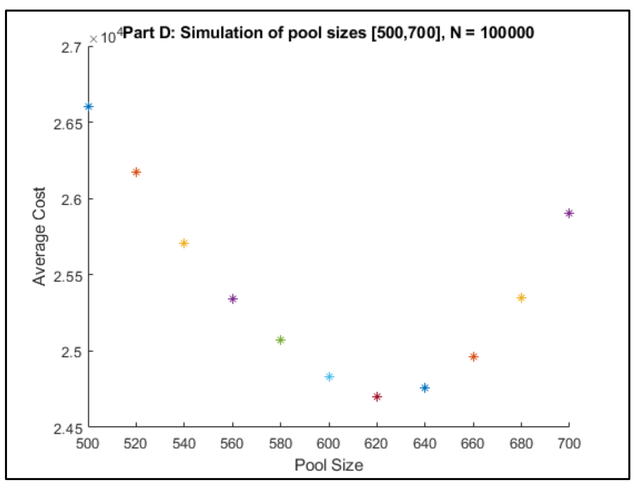
Again, this is reasonably smooth, but we will try $N = 3000$:



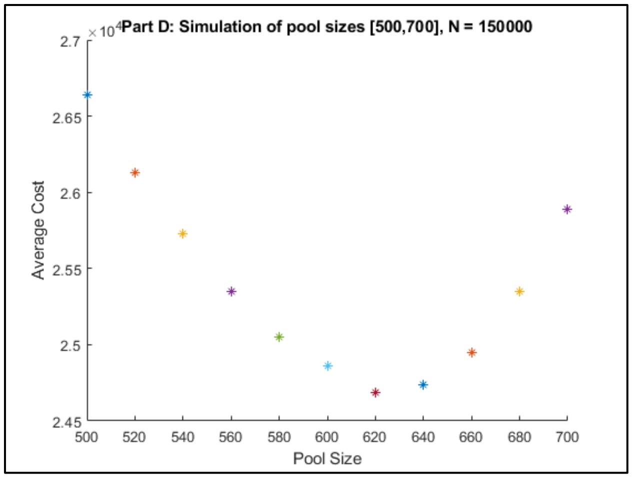
Now we can see that the minimum cost lies around $S = 600$ similarly to part C. So now we consider the interval $S = [500,700]$, with a step size of 20 and starting with $N = 50000$:



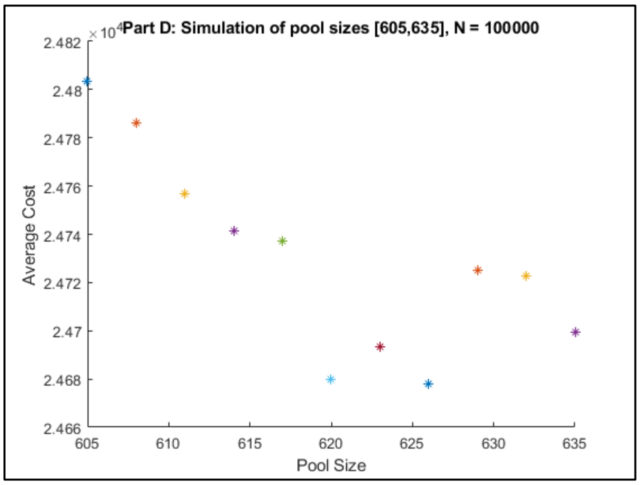
Reasonably smooth, but we can do better, so we are trying $N = 100000$:



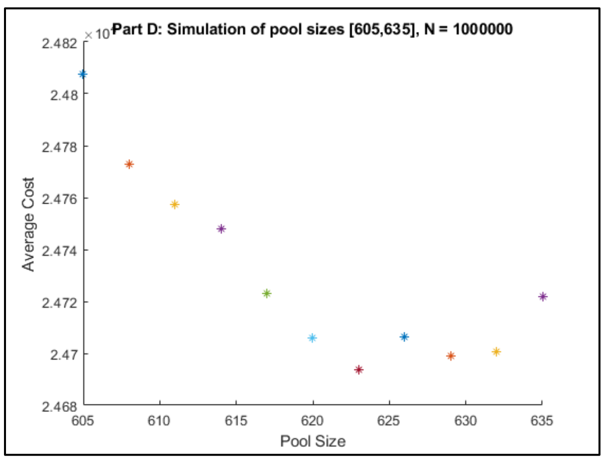
Pretty smooth, but we will try $N = 150000$:



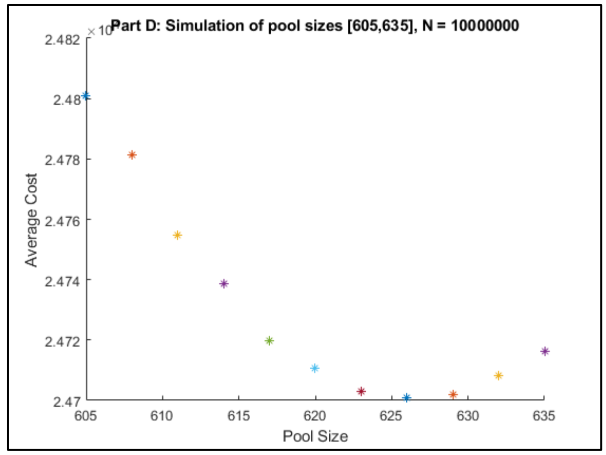
Looks very smooth and now we can see that the minimum cost seems to be around $S = 620$. So now we will look at the interval $S = [605{,}635]$, where our step size is 3 and starting with $N = 100000$:



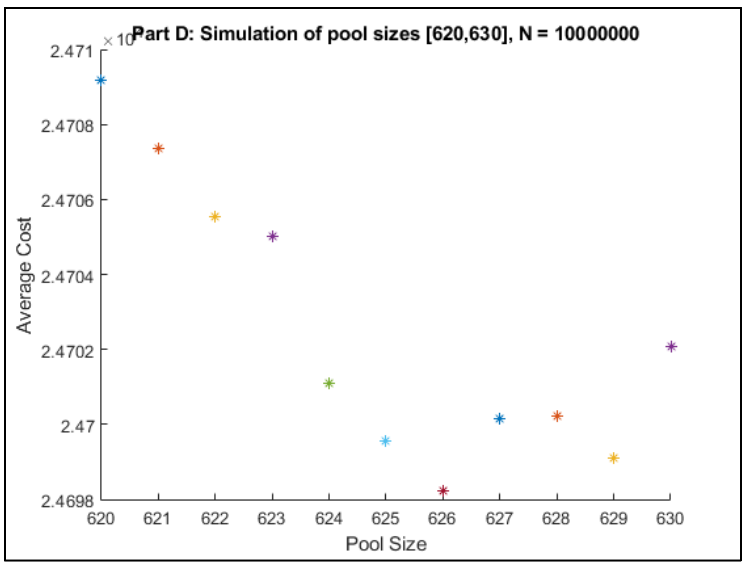
The above plot does not look smooth at all, so we will increase the number of simulations to $N = 1000000$:



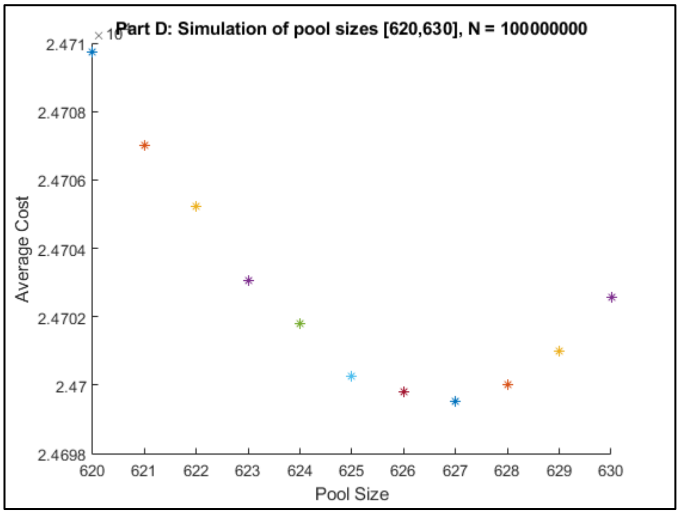
Looks smoother but we can still do better by taking $N = 10000000$:



Looks very smooth now. We will now consider the interval $S = [620,630]$ because that seems to be where the minimums of the cost lie. We will start with $N = 10000000$, simulations with a step size of 1:



The above does not look smooth enough to be able to determine the minimum value so we will take $N = 100000000$:



Now we have a smooth enough curve to say that the minimum cost values lie in the interval $S = [625,628]$. So, the Los Angeles County School District can hire 625, 626, 627, or 628 substitute teachers and their costs would be relatively similar at around $24,700. The absolute minimum cost in that interval was $24,699.5 at $S = 627$.

3 CONCLUSION

In the question of how many substitute teachers should a school district hire is not a question that can be answered simply by using a formula or by solving a linear system. The answer to this problem, and those like it, is simulation. Only recently through the invention of computers have humans had the means to simulate phenomena that would not take an absorbent amount of time. To solve the given problem of finding the optimal size of the pool of substitute teachers a school district should hire which minimizes cost, we had to formulate the appropriate algorithm that would allow us to accurately simulate the average cost for a certain number of substitute teachers.

In the cases provided, we were asked to find the minimizing pool size on Tuesdays for given pay rates for substitute and normal teachers given data provided by the county. In the first case (part C) we were given that the pay rate of a substitute teacher is $45 and the pay rate for a regular teacher substituting is $81. We found that the optimal number of substitute teachers to hire was anywhere between 595 and 605 and that the cost would be approximately $30,243. The average costs of every pool size between 595 and 605 were all similar to the nearest tenth of a cent, or there was only a difference of 10 cents between them all.

In the second case, we were asked to solve the same problem but where the pay rate of substitute teachers was lowered to $36. We found that this change did affect the outcomes of the simulations and the optimal pool size, however the change was not drastic. We found that the new optimal pool size is anywhere between 625 and 628 substitute teachers, and where the average price is around $24,700 where the differences were around 50¢ or $0.5.

Comparing our results from the two cases, we see that when the district lowers the pay rate of substitute teachers and leaves the pay rate of filling regular teachers the same, the district is able to hire more substitutes on a given day and decrease costs. This is because, as the number of substitutes you can hire goes up, it necessarily follows that you don’t then have to hire as many regular teachers which saves the district substantially.