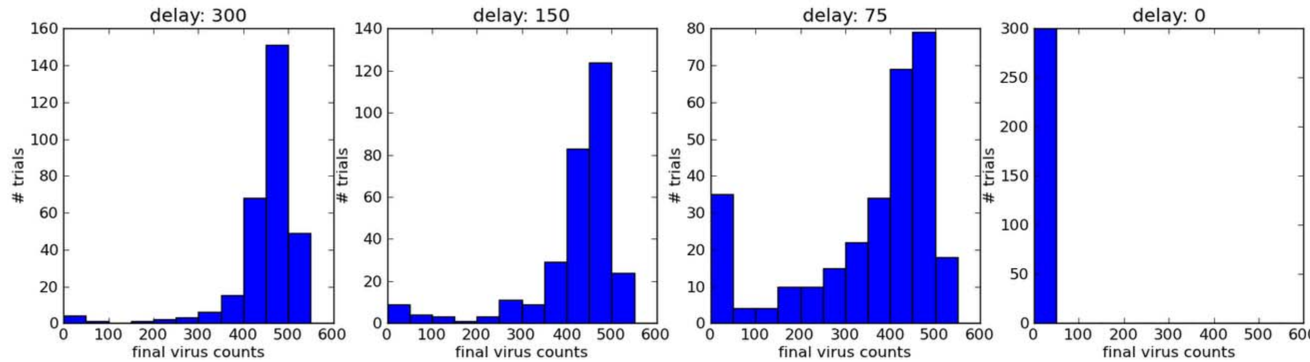
Problem 1: drug simulation with different delays

I ran 300 trials and got this histogram:



Basically the earlier the drug was applied the higher chance that the patient is cured. The longer the delay the higher chance that the patient will have more viruses in the system at the end. Notice how the size of the first histogram bin (0-50 final virus counts) decreases as delay increases, and at the same time the avg number of viruses in the patients increase.

To argue for why choosing 300 trials is good enough, students need to say something sensible, for instance the shape of the curve doesn't change much when more trials are run / the trend mentioned above is visible with the number of trials they ran / they ran multiple sets of N trials, recorded the number of cured patients and the number of patients with the peak number of viruses, and the numbers didn't change much (this is what I did to come up with the 300 number).

Note that arguing std dev decreases / mean stays the same with increasing number of trials does not make sense in this case, since these are not normal nor uni-modal distributions so the mean and std dev aren't really defined. As a case in point, here are the numbers that I got as I increased the number of trials:

for 10 trials:

delay	mean	std dev	coeff of var
0	0	0	n/a
75	304.8	165.75	0.543
150	378.4	143.28	0.378
300	448.1	33.69	0.07

for 100 trials:

delay	mean	std dev	coeff of var
0	0.02	0.14	7
75	353.59	147.94	0.418

150	387.9	134	0.345
300	454.2	66.13	0.145

for 300 trials:

delay	mean	std dev	coeff of var
0	0.05	0.224	4.21
75	349.8	156.61	0.447
150	417.76	106.7	0.255
300	450.63	75.29	0.167

As you can see there isn't much pattern as to the mean and std dev numbers.

As for % of patients cured:
300 ts delay: 5 / 300 = 1.6 %
150 ts delay: 8 / 300 = 2.6%
75 ts delay: 35 / 300 = 11.7%
0 ts delay: 100%