Process Analytics: Final Project: Optimizing Resource Allocation

Case 9.8.5

Performance Analysis and Improvement of an Internet Ordering Process

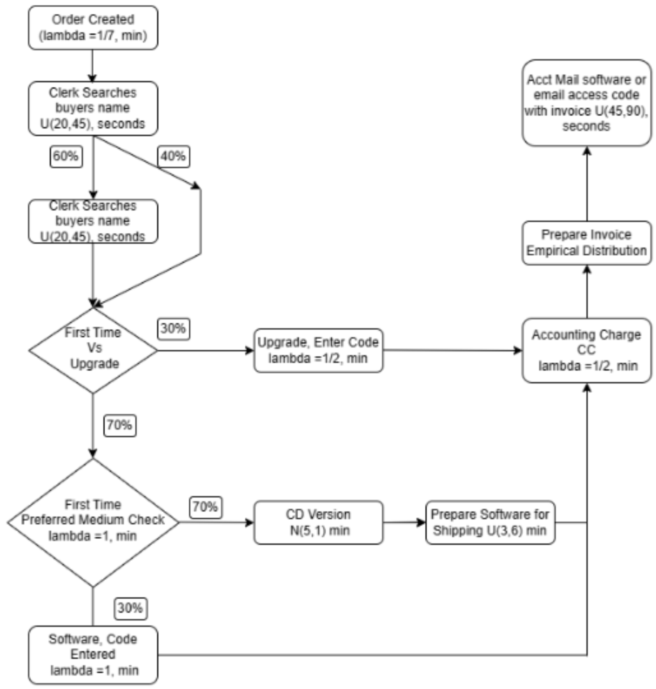
---

Executive Summary:

An analysis of the internet ordering process of a software company reveals that it is an unbalanced process leading to a high average cycle time of almost an hour and a half to process each request. The average wait time at the clerk's queue is 35+ times greater than that of the accountant–clearly a bottleneck. In order to balance this situation, we recommend adding one additional clerk to the process which results in a near 80% decrease in cycle time while also minimizing the total cost of additional resources. It is possible to compress the average cycle time even further; however, we feel the potential gains in cycle time minimization did not justify the potential cost of adding the number of resources it would require to create material gains.

Introduction:

At first glance, the case presented within 9.8.5 seems relatively straightforward – we are tasked with the analysis of a common business process: order taking and fulfillment. However, the nuances that present themselves at the intersection of distributions and probabilities present complications; to produce a useful model, each possible arc between nodes must be accurately accounted for. Additionally, one of the workflows (invoice preparation) has numerous entries on time, but lacks a specified distribution. In a reflection of the business world, we must extract generalizable features from in situ observations. In this case, that will entail the fitment of an appropriate distribution. From there, we can then move to precisely map the integrated workflows occurring between clerks and accountants. That said, in order to do so, we must first make use of a flowchart that establishes the relationships, as we understand them, between nodes.



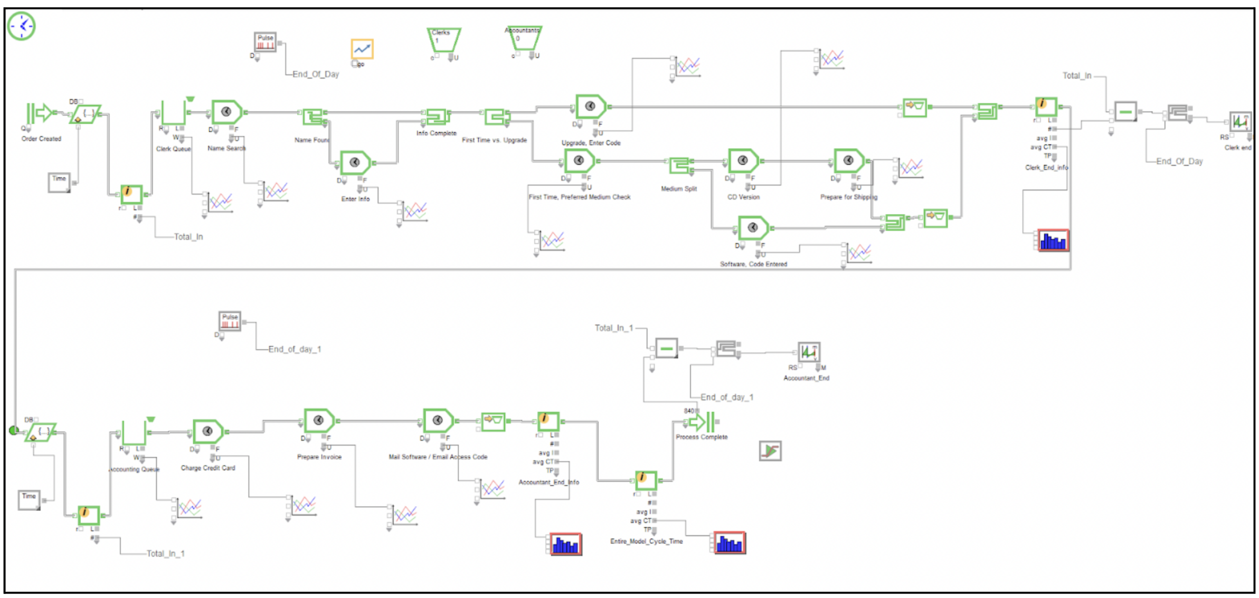
Motivation:

With the business context established, we must turn to the motivation driving this analysis: resource allocation. This is a problem of optimization – an effort to understand where, as a result of bottlenecks, additional staff will ameliorate the process the most. That said, before we can proceed further, we must describe the assumptions that will inform the findings produced: 1.) the business processes as described in the case are static and not subject to efficiency-creating alterations, 2.) the distributions that characterize activities/arrivals are static and not subject to change and, lastly, 3.) The business is, within the resource bounds specified, focused on minimizing cycle time. The third point bears some additional discussion as the business could, conceivably, be focused more on total costs rather than just waiting costs. The business could decide that, at some point, additional decreases in waiting costs are no longer worthwhile — the improvement does not offset the cost created by the worker's salary/benefits.

Before we could propose recommendations surrounding the allocation of additional staff, we had to first create a reference point; using a model and the distributions/data provided, we simulated the workflows described in case 9.8.5. From there, we were able to determine the important metrics (e.g cycle time, waiting time) that characterize the process's present performance. Next, we ran the simulation with either additional clerks or accountants – allowing us to better understand the bottleneck within the process. We were then able to extend this experiment by adding additional staff members at either position, noting the improvements to important metrics with each addition. Through the inclusion of graphing utilities and statistics blocks, we were able to, across many dimensions, describe performance. That said, our primary focus is minimizing total cycle time – we took into account utilization and other variables, but we were focused on finding a cycle time minimum within prescribed resource bounds.

Model + Methods:

Though we will narrate important components of the model we created, we felt it best to start with a visual representation:



Though perhaps challenging to fit on a single page, we can nonetheless see that the simulation model is one that can, broadly, be divided into clerk (top half) and accountant (bottom half) responsibilities. For each team member, we then, with activities and queues, describe flows. Deploying select item-in and select item-out blocks, we separated and then re-integrated flows according to both the probabilities and workflows described. Building onto the framework now established, we incorporated our in-line information blocks and linked graph blocks; they allow us to understand the important aggregate statistics that characterize the simulation's runs. Taking that one step further, we then added pulse blocks and math blocks that, in conjunction with the flows captured by the info blocks, provide insight on resource usage at specific times (the end of the workday, for example).

Analysis + Results:

Once we had everything suitably configured to run, a natural question emerged – out of the clerk and accountant, which would be the bottlenecked resource? We felt the statistics surrounding the queues for each resource would most clearly answer this question. Turning to Figure 1, we can see that the clerk was the resounding bottleneck; it produced an average wait time that was 35+ times greater than that seen at the queue for the accountant. Furthermore, the average queue length for the clerk was 10.4 as opposed to 0.3 for the accountant. Further corroborating these findings, we can see, between Figure 2 and Figure 3, that the workday ending WIP for both clerks and accountants differed significantly — the constrained clerk consistently recorded double digit values compared to an average of ~1 for the accountant. Unsurprisingly, when tasked by the case to provide an additional team member, we choose to add an additional clerk. This single addition reduced average cycle time from 86.2 all the way to 16.2. This single additional clerk reduced average cycle time by over 80%. Turning to Figures 6 and 7, we can see that this had the further benefit of creating a more tightly centered distribution for cycle time.

A natural extension is then to ask what further improvements to average cycle times and distributions can be made through additional resources. In order to adequately answer this question, we needed to make use of the optimizer block. By doing so, we found a minimum that occurs through various combinations of clerks and accountants. Put differently, we were able to select any combination of clerks and accountants that would produce a staff count less than or equal to 8 – selecting the staff combination that produces a minimum in cost (in the form of average cycle time). Plugging these parameters into our optimizer block, we see that 5 clerks and 3 accountants yields a minimum for average cycle time: 12.98.

Recommendation:

The case does not explicitly reference the costs associated with additional staff members. That said, it is reasonable to assume that the cost their salaries represent is not negligible. Consequently, we must consider the incremental improvement associated with each additional staff member. Accordingly, we are going to recommend that the company only add one additional staff member – a clerk. The first additional staff member was able to produce a greater than 80% improvement to cycle time. Conversely, the next 5 staff members could only produce ~20-25% of improvement to cycle time (the third and fourth staff members produce larger improvements, but, on the whole, they only produce a ~4-5% improvement each). Therefore, even without hard data on staff costs, we are still confident that one additional staff member, 3 total, will produce a total cost (waiting costs + service costs) minimum. Inspiring further confidence for this 3 staff model, the cycle time distribution was sufficiently tight (+2 SD within 30% of mean) that surge risk does not warrant additional consideration. Nonetheless, the company could - if its staffing costs are low enough - explore the use of a 4th staff member for times that could produce higher volumes. They can then evaluate whether or not that staff member creates appreciable improvements to either average cycle time or cycle time distributions.

Case 10A.2

Emergency Room Staffing

Executive Summary:

An analysis of emergency room staffing reveals that doctors and emergency room nurses are the critical resources when it comes to minimizing the cycle time for category 1 patients. Our analysis revealed that it is possible to find a relatively good solution via trial-and-error; however, running an optimization model found that the ideal staffing for the emergency room will include 6 doctors, 5 emergency nurses, and one of every other staff type. This will minimize the category 1 patient cycle time of approximately 100 minutes while staying below the given threshold of 10 budget units. Our trial-and-error solution called for just 4 emergency nurses which resulted in slightly better budget allocation (9.6 vs. 9.9 BUs); however, that gave an average cycle time of about 121.81 minutes versus 100 minutes. We feel that the decrease in performance to save .3 BUs when performance could potentially mean life or death for these critical patients does not justify those savings; however, that is up to the hospital administrator to ultimately decide.

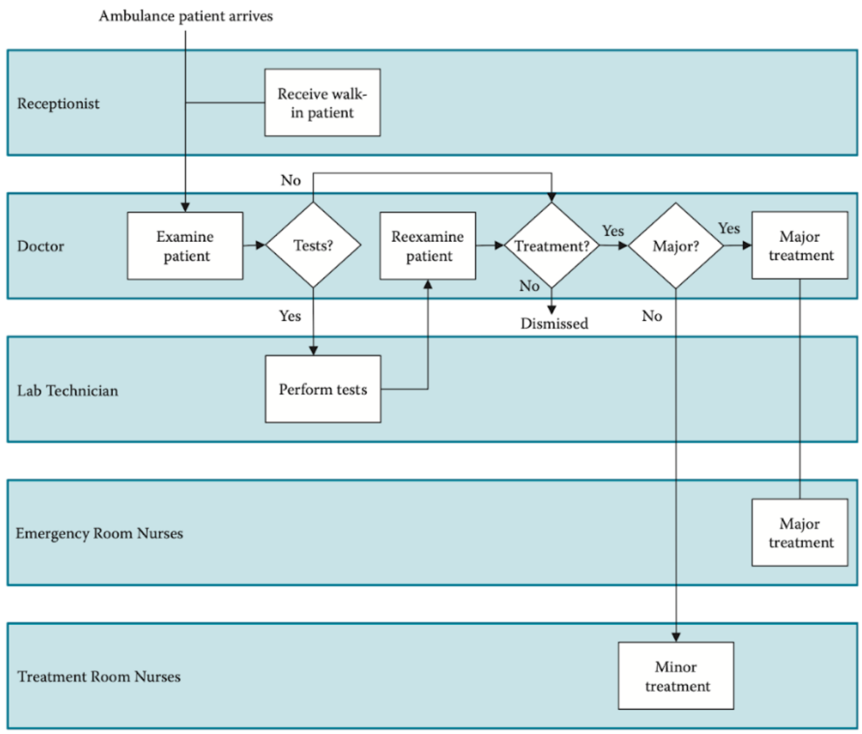
Introduction:

Emergency rooms and hospitals must contend with a continually stressful, high-stakes environment – triage is, rather than the exception, the norm. Patients, especially those in critical condition, must be quickly categorized, diagnosed and then treated. In order to effectively do this, hospitals must devise effective intake strategies that make the best use of scarce, expensive resources. Those patients in category 1 command special attention; they, more than budget constraints and general patient throughput, will determine the staff mix/allocation. Unlike other patients, they cannot afford to wait; it's been consistently shown that, for strokes, gunshot wounds and heart attacks, the first hour is critical – the interval between injury/health event and treatment is strongly predictive of survival. Though this is a primary consideration that must inform the decisions made, it cannot be the only one: the normal concerns surrounding efficient, profitable operation still apply.

Motivation:

Now that we have a sense of the context specific stakes and particulars, we can motivate the tasks before us. We have the opportunity to present two different, acceptable solutions: A.) a solution that, after ensuring a sub 3.5 hour average cycle time for category 1 patients, minimizes costs and B.) a solution that within the bounds prescribed by 10 budget units, minimizes category 1 patient average cycle time. Accordingly, moral principles may help inform some of the tradeoffs and implicit tension between solutions of varieties A and B; increases/decreases to category 1 patient average cycle time will likely need to be priced in budget units (BUs). From that, a hospital CFO may have to countenance and tacitly acknowledge an unsavory dilemma – the fact that cost saving past a certain point may likely lead to increased patient deaths (less BUs → fewer staff → greater cycle time → lower survival rate → more deaths). Rather than being paralyzingly morbid, this discussion serves to acknowledge the context we should incorporate when comparing the solutions from part 1 and part 2.

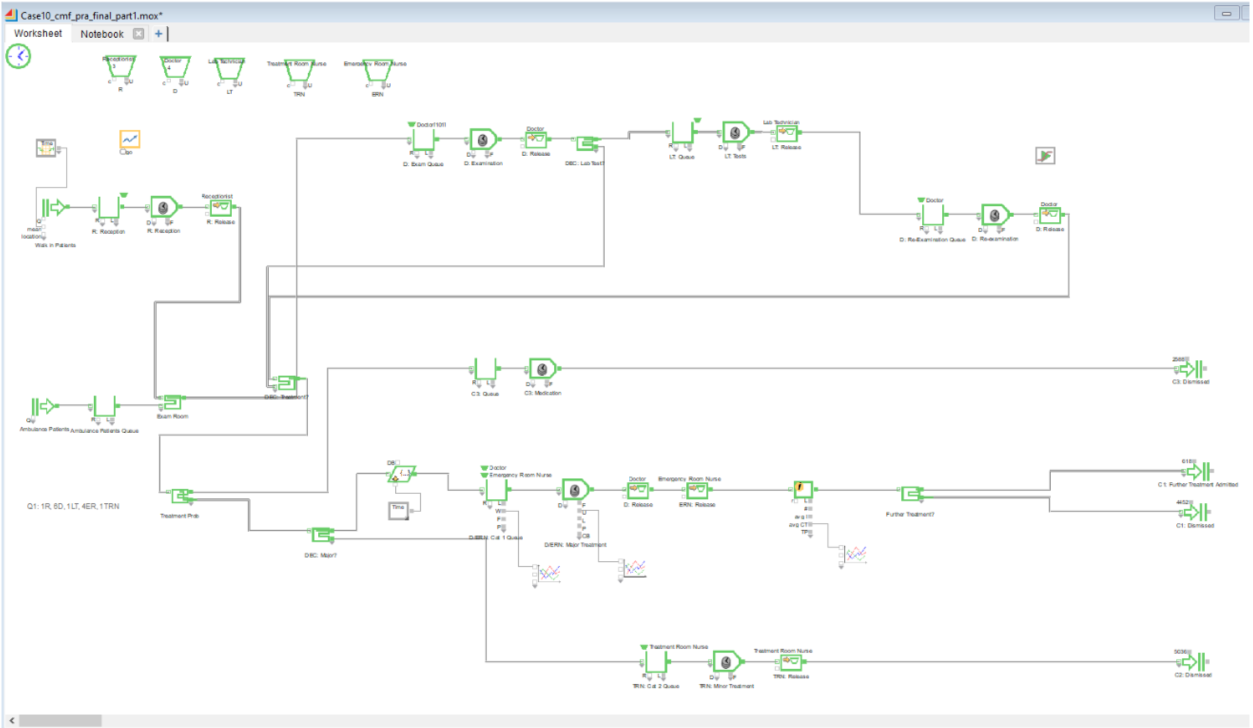
Before we can proceed to simulation models, we must cement our understanding of the workflows that describe this intake process. Turning to a visual representation, we have the following:



With our simulation of this process, we strove to both understand how to minimize cycle time for category 1 patients while also optimizing all activities secondary to the treatment of category 1 patients. Similar to Case 9.8.5, we hope to find obvious inflection points that can inform the basis for actionable insights.

Models + Methods:

Though this case has two parts with different objectives, the model used is, for illustrative purposes here, sufficiently similar. Part 2's diagram will be included in the appendix, but for the ensuing discussion, we can reference Part 1:



In a method similar to that followed in Case 9.8.5, we first sought to roughly describe the relationship between queues, activities and resources. With the help of select-out and select-in blocks we specified probabilities that, respectively, defined flow splits and integrations. From there, we incorporated the inline info blocks and linked graph blocks that allowed us to understand emergent, aggregate properties. This workflow diverges from the prior case in that it is not as linear; hospital staff do their best to separate out the critical patients, getting them into a process commensurate with their condition. Consequently, all processes condense coming into intake and diagnostics. From there, the critical patients can then be separated out and given the timely treatment their condition warrants.

Results + Analysis:

For the first part of the project we, as the prompt suggested, used trial and error to converge upon a satisfactory solution – one that would, with minimal resources, produce an acceptable cycle time for category 1 patients. Doctors were, like clerks in the previous case, the bottlenecked resource. To avoid exploding average cycle times, we would need to maximize the number of doctors (6). Furthermore, emergency room nurses were the next most critical resource as any number less than 4 produced significant deterioration in average cycle time for critical patients. This specific allocation (6 doctors, 4 emergency nurses, 1 receptionist, 1 lab tech and 1 treatment nurse) will require a total of 9.6 BUs. Consequently, it is a broadly acceptable solution – it is under budget and, per Figure 10, at 121.81 minutes average cycle time for critical patients, well under the 3.5 hour criteria. A marginally better, in budget solution is possible through the addition of an additional emergency room nurse (it would see average critical cycle time go from ~121 minutes to ~100 minutes). However, the objective was to find the cheapest solution that would satisfy the requirements.

With the trial and error approach only taking us so far, we implemented an optimizer block in order to more precisely converge upon the desired minimum (average cycle time for critical patients). Turning to Figure 11, we see what various combinations of staff produced as it relates to our target (average cycle time improvement). Interestingly, the optimizer block confirmed the hunch that we had regarding further improvement – looking at Figure 11, we can see that the addition of another nurse could significantly reduce cycle time without going over budget. Here the quandary described earlier will likely rear its head: a need to reconcile savings with less than optimal cycle times for category 1 patients. A careful study surrounding that 23 minute improvement would likely be necessary to make a fully informed decision; are there, for critical patients, frequent deaths occurring in that time interval (between 100 and 121 minutes) that could be avoided by more timely treatment? Though precepts and actual budget data will likely most influence the ultimate decision, we can also look at utilization data between the two solutions. Turning to the tables provided in Figures 13 and 14, we can see that the optimized solution produces marginally higher utilization across most activities (with the exception of minor treatment). Furthermore, there are not appreciable differences in either the maximums or averages for wait times and queue lengths.

Lastly, we can look to see if simulation duration will have meaningful impacts on either of the solutions described (trial and error, optimized). Specifically, we want to see if, on account of the simulation being too short, we failed to reach a steady state equilibrium. Looking at figure 10, we can see what that steady state convergence looks like over the course of 100 days for Part 1. Importantly, we can see that, at the 30 day mark, the system has not yet totally converged upon a steady state; closer to the 100 day mark, we can see the system almost asymptotically approach a ~116 minute average cycle time before ultimately recording a value of ~121. Consequently, we will want to perform the same longer run for Part 2 to see if, compared to the 30 day run, significant changes are recorded. From figure 11, we can see that, for Part 2, the longer run does produce a change: an increase to average cycle time, ~105 minutes. As a result, once both solutions reach a less variable, steady state, the differential between them narrows — there is only a ~13% decrease through the addition of the extra nurse. However, the differential in maximum cycle times is actually larger than that observed for the 30 day run (466 for Part 1 (Fig. 10) and 267 for Part 2 (Fig. 15)); put in more impactful terms, over the course of 100 days, one solution will necessitate almost 3.5 hours of additional suffering from someone in critical condition.

Recommendation:

We, for the reasons described in the motivation section of this case, recommend the optimized solution that suggests an additional nurse be provided. It is still in budget and doesn't expose the hospital to the moral hazard of having to equate human lives to surplus budget units. Furthermore, the incremental budget increase only represents a ~3% increase in expenditure. Additionally, the averages, for cycle time, do not tell the full story; through Figures 10 and 15 we can see the optimized solution, on a 100 day basis, produces a far superior maximum cycle time of 267 as opposed to the cheaper solution found in part 1 that produces a 467 maximum cycle time. That is a significant differential, one that, for 3% more in-budget money could be avoided. Since this represents an almost 3.5 hour additional interval that someone in desperate need of help could be suffering, we do not feel the 3% savings is a sufficient motivation. Finally, we recommend the hospital, in the interest of efficiency, perform an analysis that disaggregates some of the queues and activities within the broader workflow. Are there improvements to specific activities that can be made along the path for critical patients? Though we accepted that this could not be done in order to have a stable context for simulation building, that is not a true reflection of the actual process. Furthermore, doctors and emergency room nurses are the critical, bottleneck resources that minimize the acceptable solution space for staff allocation. Are there ways to transfer some of their responsibilities to less constrained resources/staff? Asking these questions will, with appropriate subsequent analysis, allow the hospital to converge upon a truly optimal solution that best balances budget costs and patient cycle times — we hope that, with our simulation, we've provided a good starting point for further analysis.

Appendix:

Figure 1

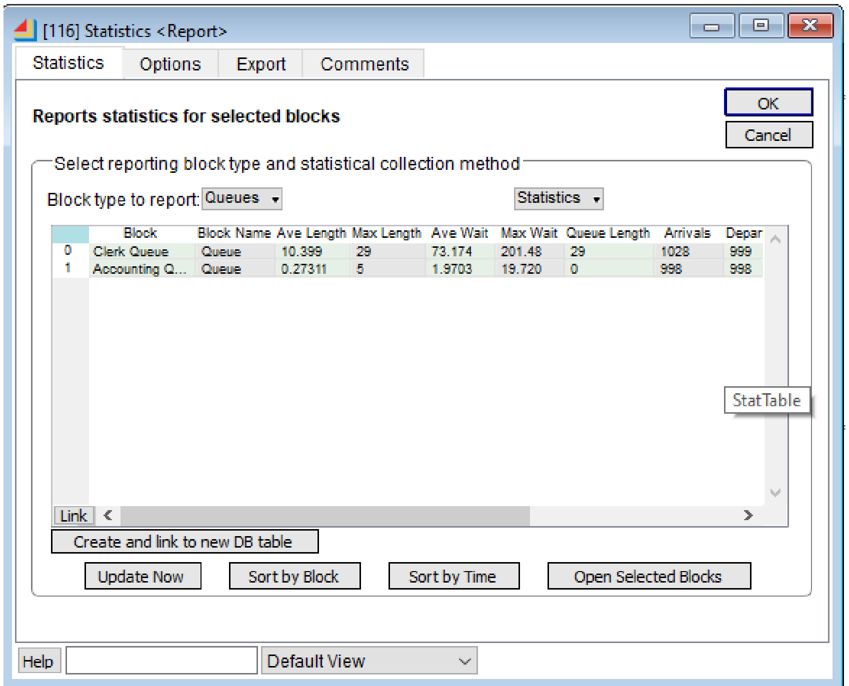


Figure 2: Day Ending WIP Clerk

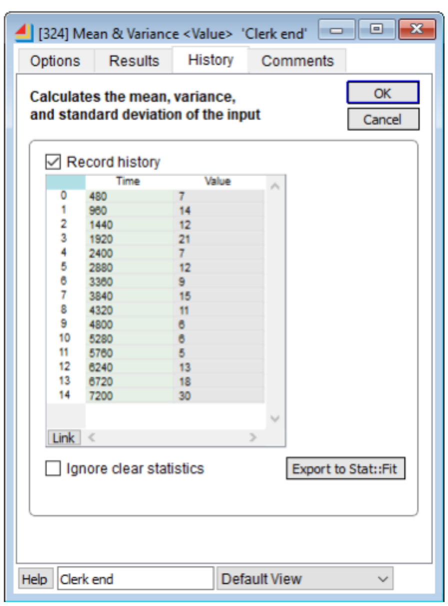


Figure 3: Day Ending WIP Accountant

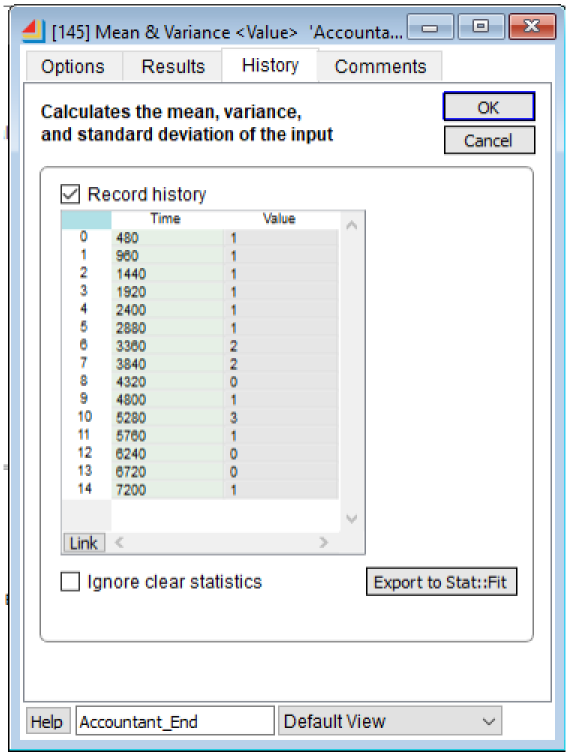


Figure 4: Wait Time, Single Clerk

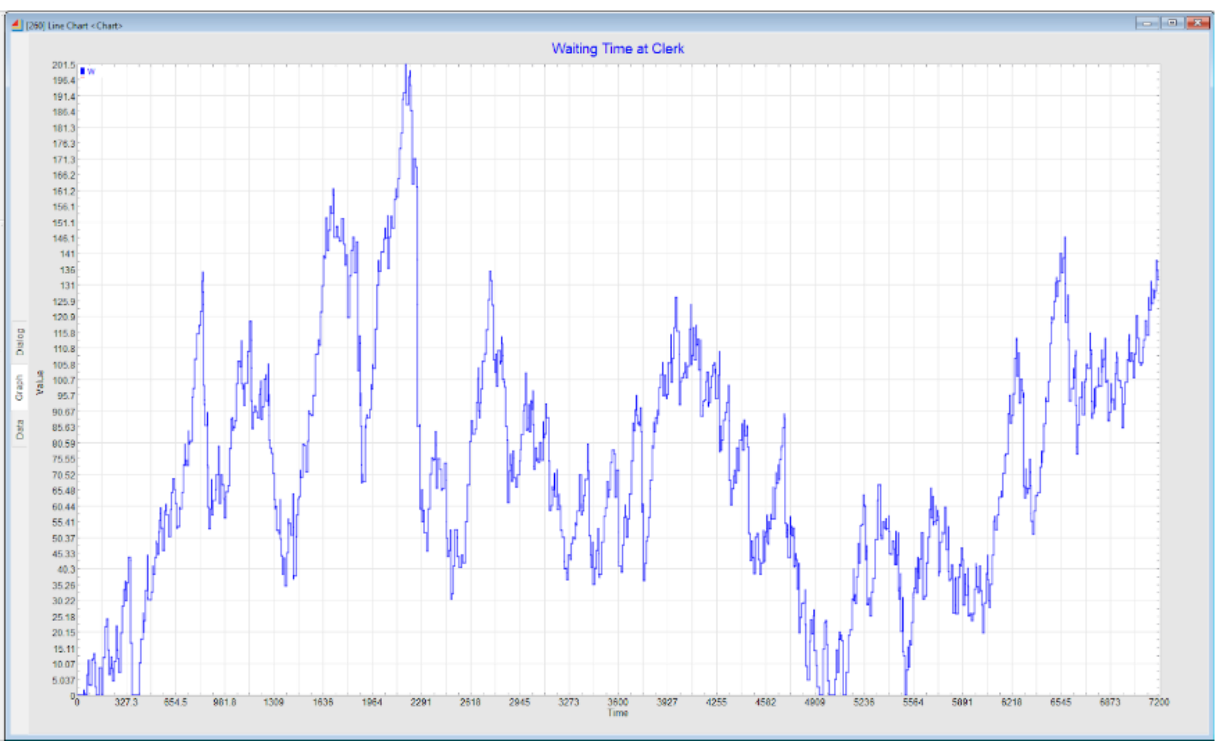


Figure 5: Wait Time, Two Clerks

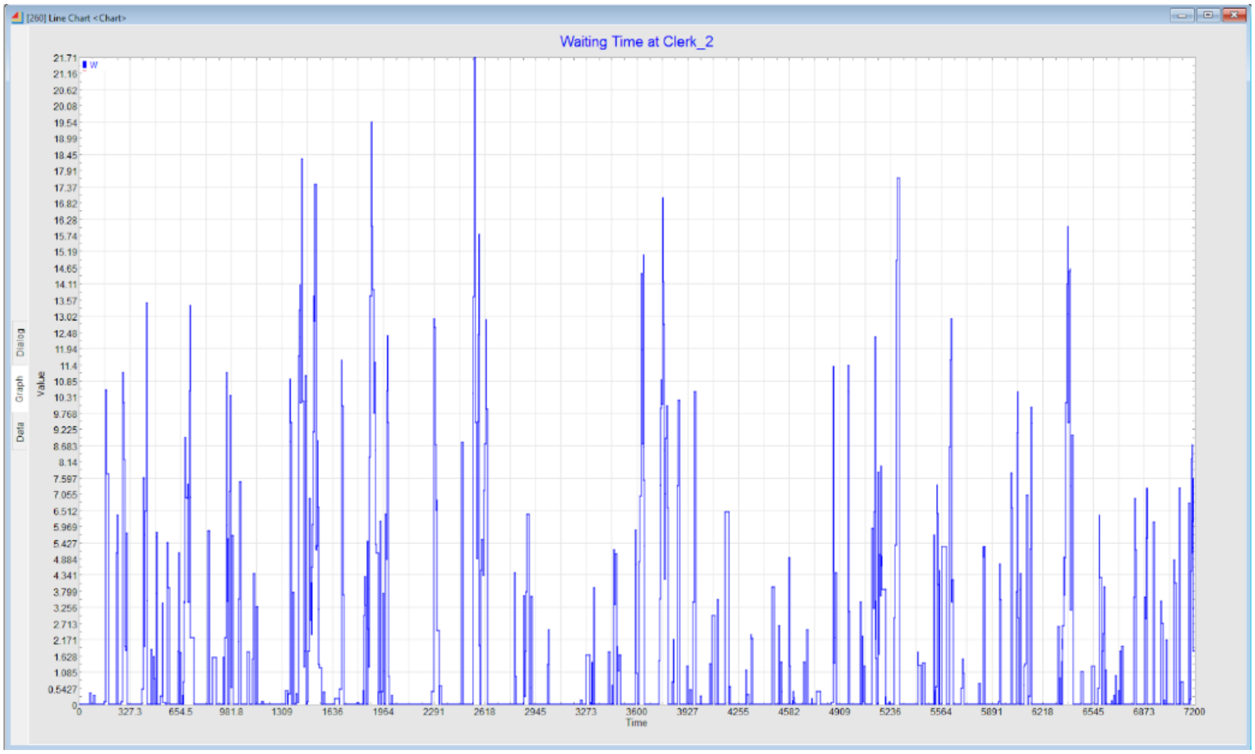


Figure 6: Total Cycle Time: 1 Clerks, 1 Accountant

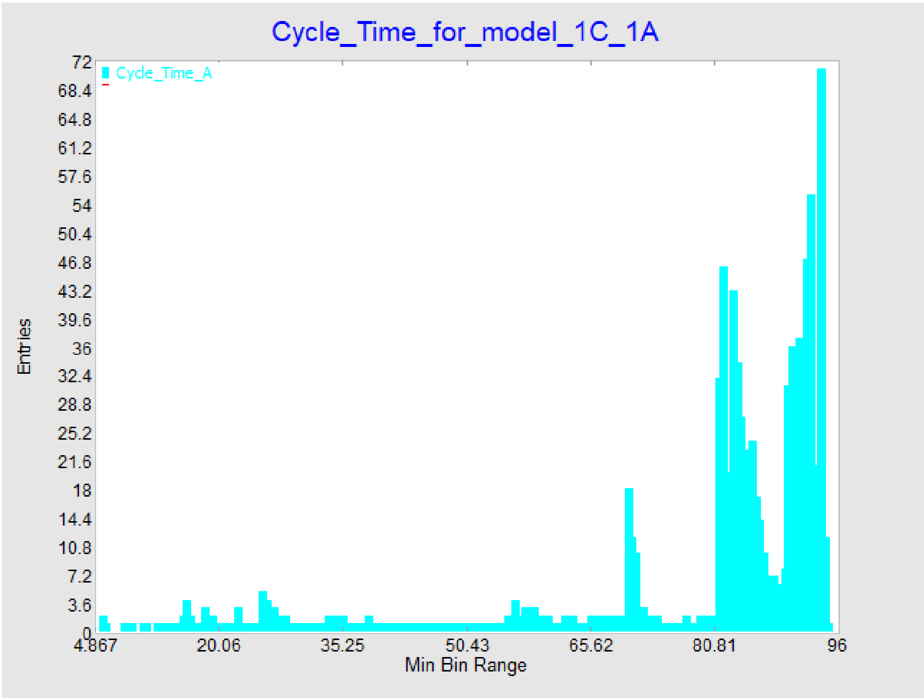


Figure 7: Total Cycle Time: 2 Clerks, 1 Accountant

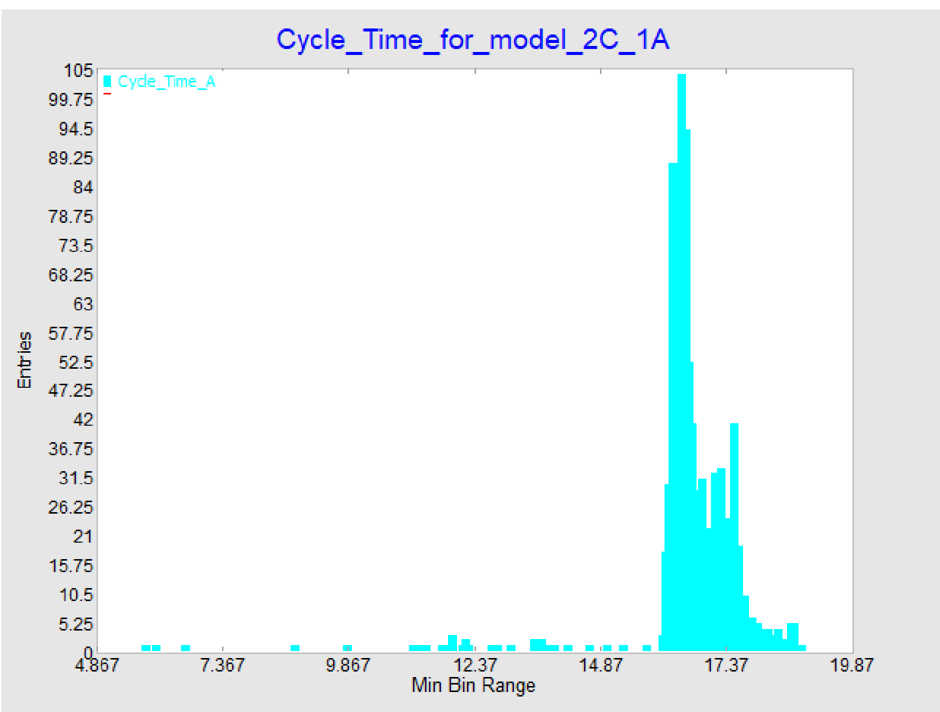


Figure 8: Case 10A.2 Part 2:

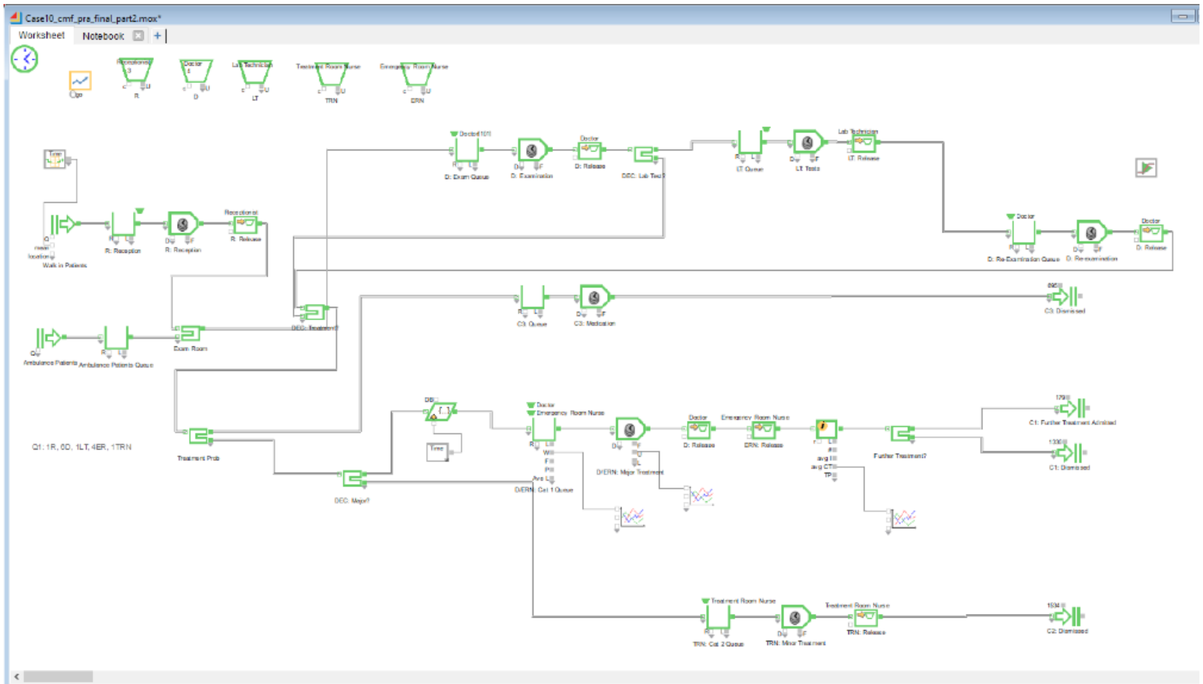


Figure 9: Case 10A.2 Part 1 Cycle Time (Critical): (100 Day Run)

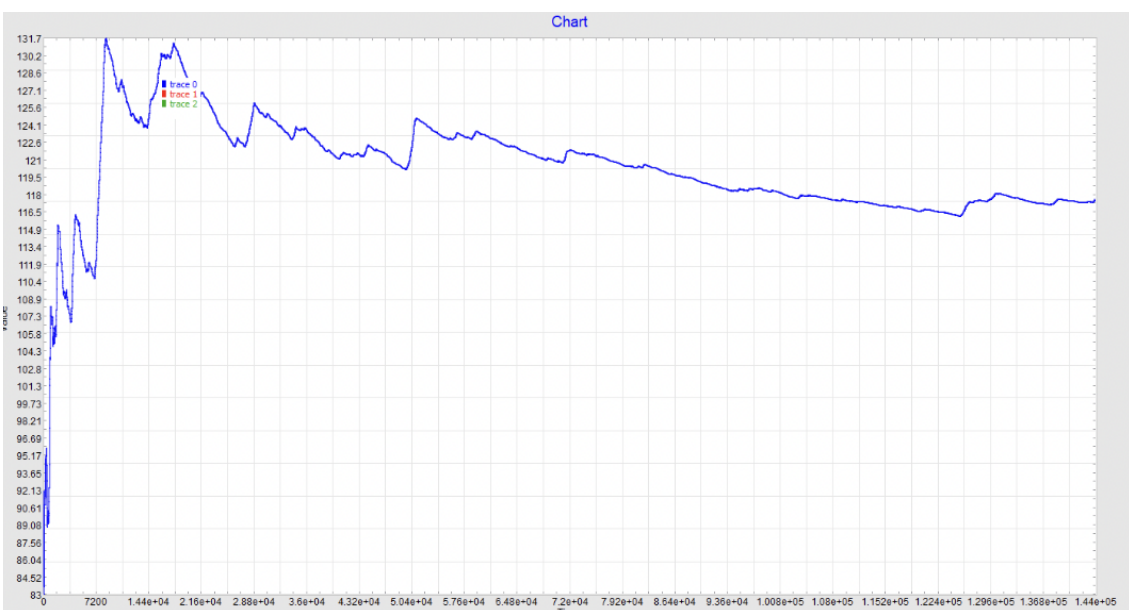


Figure 10: Case 10A.2 Part 1: Avg. Cycle Time Critical: (100 Day Run)

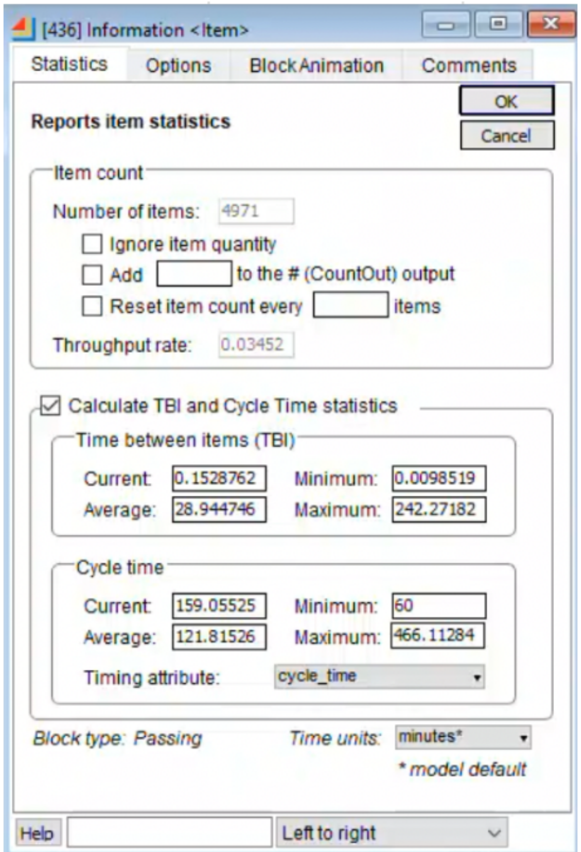


Figure 11: Case 10A.2 Part 2: Optimizing Staff Allocation: (30 Day Run)

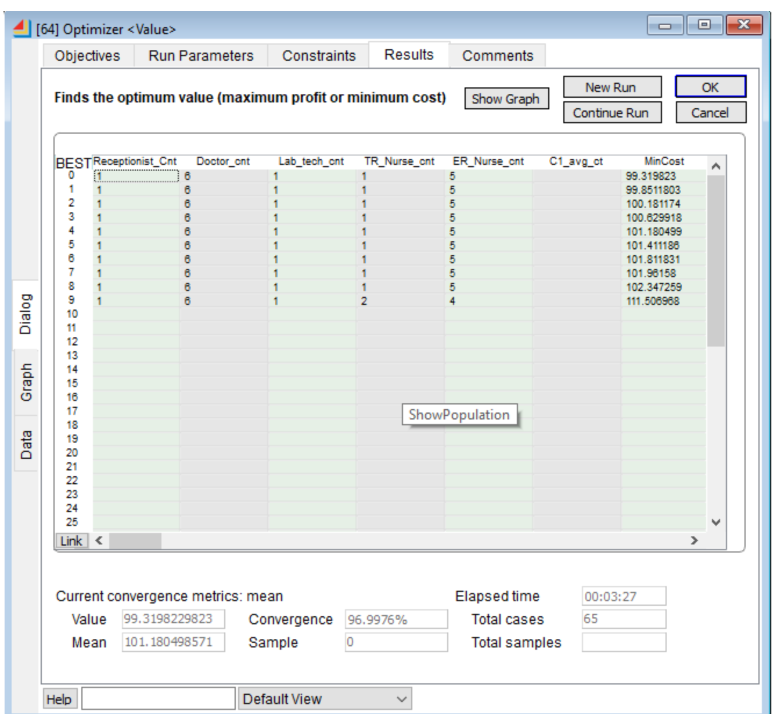


Figure 12: Case 10A.2 Part 2: Optimized Allocation: Cycle Time (30 Day Run)

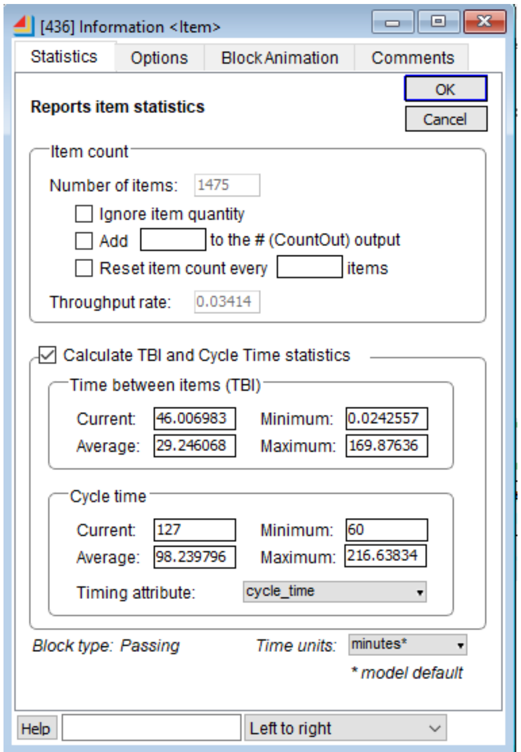


Figure 13: Case 10A.2 Part 2 Cycle Time (Critical): (30 Day Run)

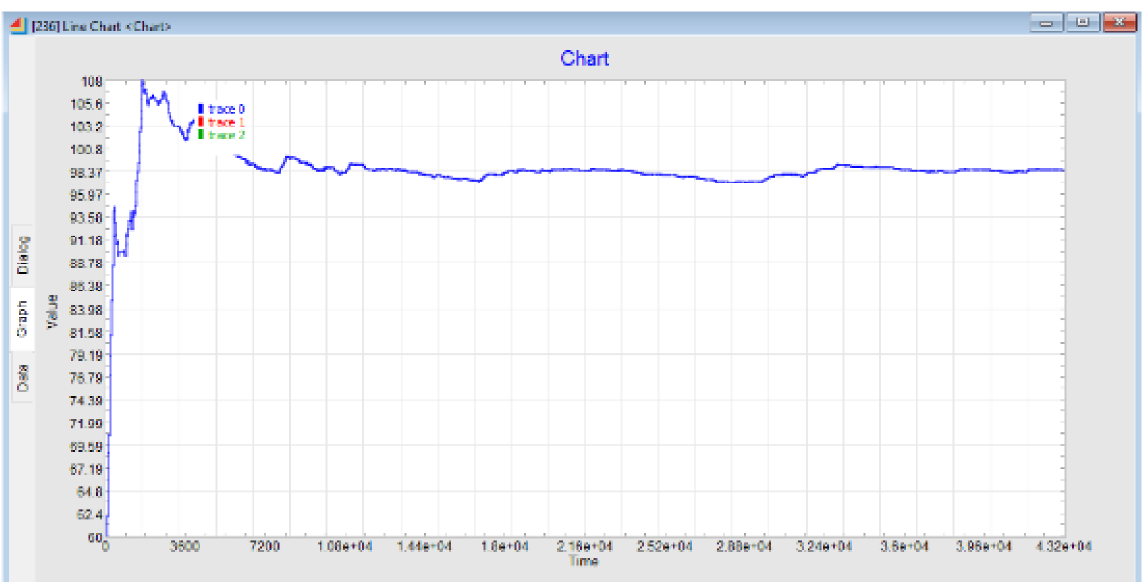


Figure 14: Case 10A.2 Part 3: Optimizing Staff Allocation: (100 Day Run)

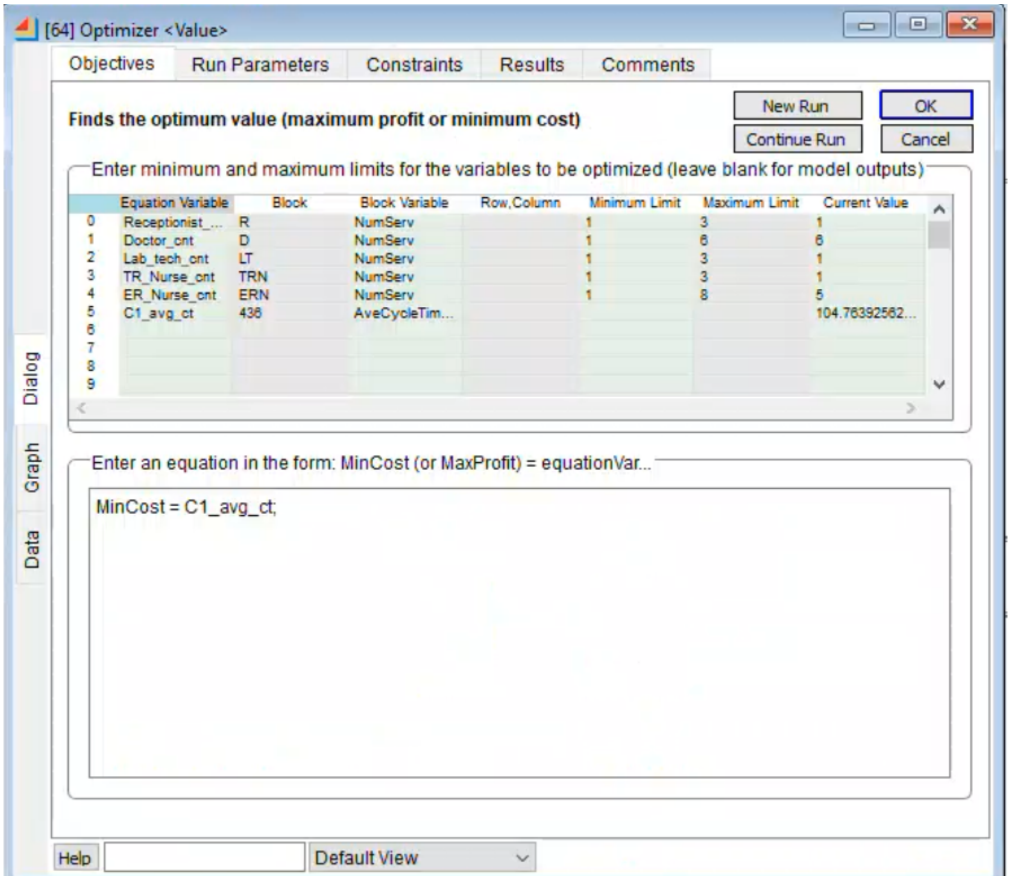


Figure 15: Case 10A.2 Part 3: Optimized Allocation: Cycle Time (100 Day Run)

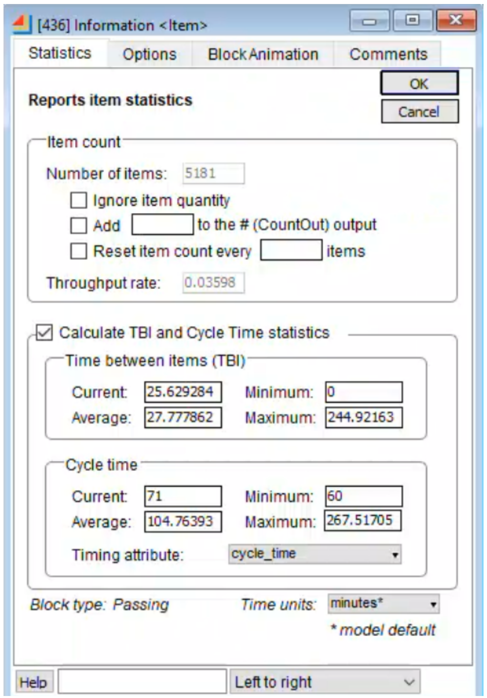


Figure 16: Table for Trial and Error Part 1 Solution: (100 Day run)

	Utilization	Avg. Length	Max Length	Avg. Wait	Max Wait
R: Reception	0.041516	0.41516	1	7.5218	10
D: Examination	0.13259	1.3259	6	15.036	20
LT: Tests	0.042814	0.85627	1	19.384	24.14
D: Re-examinati	0.042029	0.42029	3	9.5145	12
TRN: Minor Trea	0.087551	0.87551	1	25.034	30
C3: Medication	0.0017972	0.017972	2	1	1
D/ERN: Major Tr	0.31813	3.1813	4	90.341	120

Figure 17: Table for Optimized Part 3 Solution: (100 Day Run)

	Utilization	Avg. Length	Max Length	Avg. Wait	Max Wait
R: Reception	0.042103	0.42103	1	7.4963	10
D: Examination	0.13377	1.3377	6	15.009	20
LT: Tests	0.043518	0.87035	1	19.493	24.141
D: Re-examinati	0.041903	0.41903	3	9.389	12
TRN: Minor Trea	0.090526	0.90526	1	25.098	30
C3: Medication	0.0017847	0.017847	2	1	1
D/ERN: Major Tr	0.31452	3.1452	5	90.006	120