UNIVERSITY OF SURREY

THE FACULTY OF ARTS AND SOCIAL SCIENCES

Surrey Business School

OPERATIONAL ANALYTICS COURSEWORK

Academic Year 2023-2024

Submission deadline: 1600 BST, Monday 13th May 2024

Student URN	Student Name	MSc Programme
6793260	ELDHOSE VARGHESE	BUSINESS ANALYTICS

Table of Contents

Introduction

Rapid DNA Forensic Process

The rapid DNA forensic process is divided into seven key steps beginning with sample collection and concluding with DNA matching. Initially, a crime scene investigator gathers the DNA sample at the crime scene and logs the sample information into a database using a tablet. Next, a courier service transports the collected DNA sample to the central forensic laboratory. At the lab, the samples are prepared for analysis. These prepared samples are then processed in a DNA sequencer for sequencing. After sequencing, a lab researcher checks all the results. Once this validation is complete, the findings are automatically uploaded to a UK-wide criminal DNA match database to search for potential matches.

Modelling

First crime is reported at the inter-arrival time given by the distribution $T_1 \sim RiskGamma(1{,}100)$ [average =100 min]. At this point the CSI officer starts his journey to the crime scene location taking time represented by the distribution $T_2 \sim RiskPert(40, 47.5, 70)$ [average $= 50$ min] to reach the destination. Here, Pert distribution is used because I am given with three-point estimates minimum, most likely and maximum. Then the CSI officer immediately starts sample collection process taking time $T_3 \sim RiskUniform(50, 70)$ [average $= 60$ min] to complete it. (Here, uniform distribution is used because I am given with only two-point estimates minimum and maximum.) At this point the courier staff collects sample from CSI officer and takes time $T_4 \sim RiskWeibull(7, 100)$ [average $= 93.54$ min] to deliver the sample to the forensic lab. Once the sample reaches the forensic lab it waits for $T_5 \sim RiskGamma(5, 20)$ minutes [average $= 100$ minutes] for the lab technician to finish 5 other jobs after which the sample preparation starts, and it takes $T_6 \sim RiskUniform(15{,}25)$ minutes [average $= 20$ minutes] to finish the preparation process. Sample one then waits for 7 other samples to be ready as the DNA sequencer is only available for a run once 8 samples are ready. The DNA sequencer is first run at the 8th sample preparation finish time and takes 120 minutes for sequencing. However, the samples then wait for $T_8 \sim RiskPert(75{,}97{,}135)$ $minutes$ [average $= 99.67$ minutes] for the senior lab researcher to start its validation. The lab researcher then validates the samples one by one taking $T_9 \sim RiskUniform(20{,}30)$ minutes [average $= 25$] for validating each sample. This process is repeated for one week. The relevant notations and equations used in the model are described below.

T_1 – Total minutes to report a crime

T_2 – Total minutes for a CSI to arrive at the crime scene

T_3 – Total minutes for CSI to collect sample and upload sample information to the database

T_4 – Total minutes for the courier company to deliver a sample to the forensic lab

T_5 - Total minutes for lab technician to complete 5 jobs before preparing sample

T_6 – Total minutes for preparing a sample

T_7 – Total minutes for DNA sequencing

T_8 – Total minutes for the lab researcher to start validation

T_9 – Total minutes for validation of sample

T_V - Validation finish time

T_R – Crime Reporting time

T_{DNA} – Total minutes to process a DNA sample

W_{CSI} – Wait time for CSI arrival

W_{SP} – Wait time for sample preparation

W_{SEQ} – Wait time for DNA sequencing

W_V – Wait time for sample validation

P_{match} - Probability that a sample will be matched to a profile on the DNA database

$T_1 \sim RiskGamma(1{,}100)\ min$

$T_2 \sim RiskPert(40, 47.5, 70)\ min$

[calculations – here I am given with mean value, so I need to find the most likely value.

$\mu = (a + 4b + c)/6$ (Wikipedia, 2022)

Where, a = minimum, b = most likely, c = max, μ = mean

So, $b = (6\mu - a - c)/4$

$b = \frac{6*50-40-70}{4} = 47.5$]

$T_3 \sim RiskUniform(50, 70)\ min$

$T_4 \sim RiskWeibull(7, 100)\ min$

$T_5 \sim RiskGamma(5, 20)\ min$

$T_6 \sim RiskUniform(15{,}25)\ min$

$T_7 = 120\ min$

$T_8 \sim RiskPert(75{,}97{,}135)\ min$

$T_9 \sim RiskUniform(20{,}30)$

$T_{DNA} = T_V - T_R$

T_{DNA} Can also be written as,

$T_{DNA} = W_{CSI} + T_3 + T_4 + W_{SP} + T_6 + W_{SEQ} + T_7 + W_V + T_9$

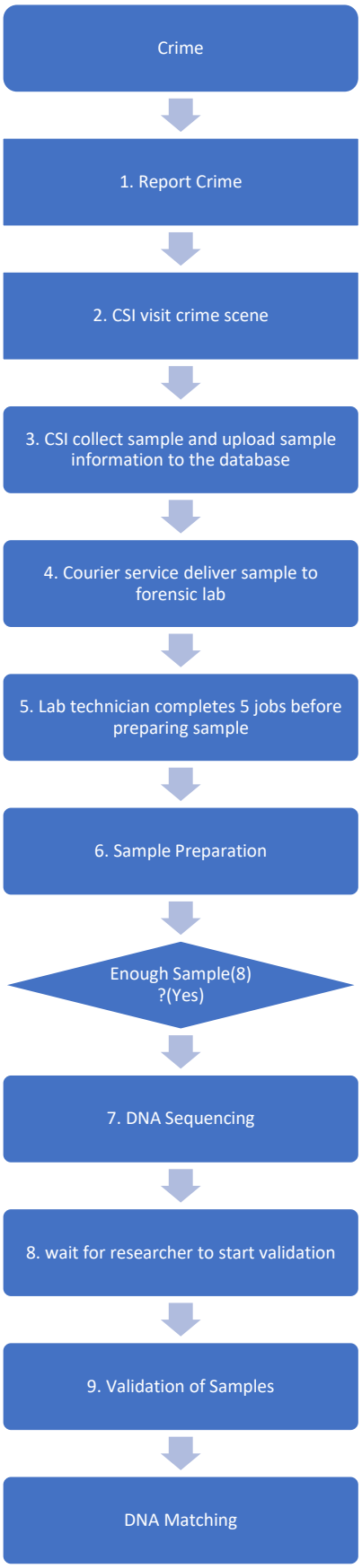

Figure 1: Detailed Flow Diagram for the Rapid DNA Forensic Process

Conditional aspects of the process

- The CSI officer cannot visit a new crime spot until any current crime sample collection is finished.
- The sample preparation can start only after the lab technician completes 5 jobs.
- The DNA sequencer machine becomes available for a run only when 8 samples are ready.
- The lab researcher can validate only one sample at a time.

Parameter estimation

Table 1: Summary of parameter estimates. All units in minutes.

Parameter	Description	PDF	Mean	5th %ile	95th %ile
T_1	Total minutes to report a crime	RiskGamma(1,100)	100.00	5.13	299.57
T_2	Total minutes for a CSI to arrive at the crime scene	RiskPert(40,47.5,70)	50.000	42.293	59.722
T_3	Total minutes for CSI to collect sample and upload sample information to the database	RiskUniform(50,70)	60.000	51.000	69.000
T_4	Total minutes for the courier company to deliver a sample to the forensic lab	RiskWeibull(7,100)	93.544	65.422	116.969
T_5	Total minutes for lab technician to complete 5 jobs before preparing sample	RiskGamma(5,20)	100.00	39.40	183.07
T_6	Total minutes for preparing a sample	RiskUniform(15,25)	20.000	15.500	24.500
T_8	Total minutes for the lab researcher to start validation	RiskPert(75,97,135)	99.667	82.442	119.055
T_9	Total minutes for validation of sample	RiskUniform(20,30)	25.000	20.500	29.500

Model Settings

Number of iterations = 10000

Random number seed = 3260

Results

Average solution

a. Number of samples processed in a week = 80 samples

b. Number of samples matched = 2 samples

c. Mean time to process a sample = 1436 minutes

d. Percentage of samples processed in less than one day = 51%

e. Under- resourced stages

Response Time for CSI Arrival – This phase is limited in resources because a CSI officer can only begin traveling to a subsequent crime scene after completing sample collection and uploading at their current location. This consecutive method restricts how swiftly CSIs can attend to new incidents, potentially causing holdups, particularly during times of frequent crime reports.

Sample Preparation – This stage is also under-resourced as the lab technicians has to do 5 other jobs as well before sample preparation. This may be due to limited number of staffs.

DNA Sequencing – This process is constrained as the sequencer is limited to processing just eight samples at a time, which may lead to delays.

Sample Validation – This stage faces resource limitations as well, since the lab researcher is only able to validate one sample at a time.

Monte Carlo simulation (Task 4)

b.

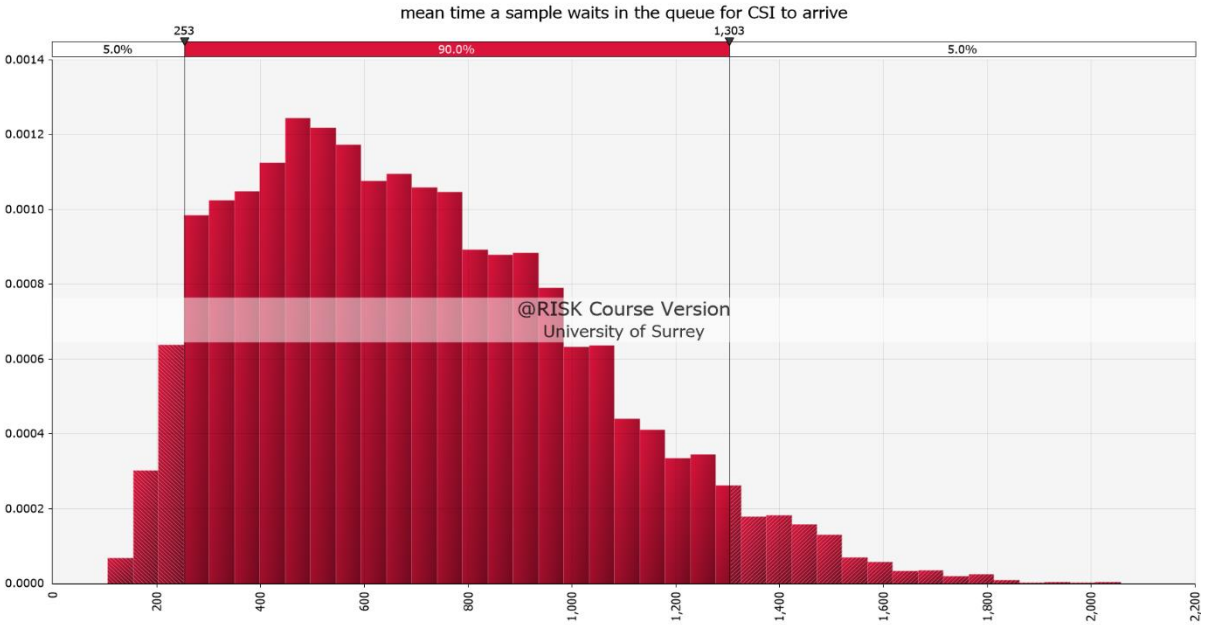

Figure 2: PDF of mean time a sample waits in the queue for CSI to arrive.

Figure 2 shows that the mean queueing time for CSI arrival is between 253 and 1303 minutes at 90% confidence level. The graph is skewed and contains extreme values.

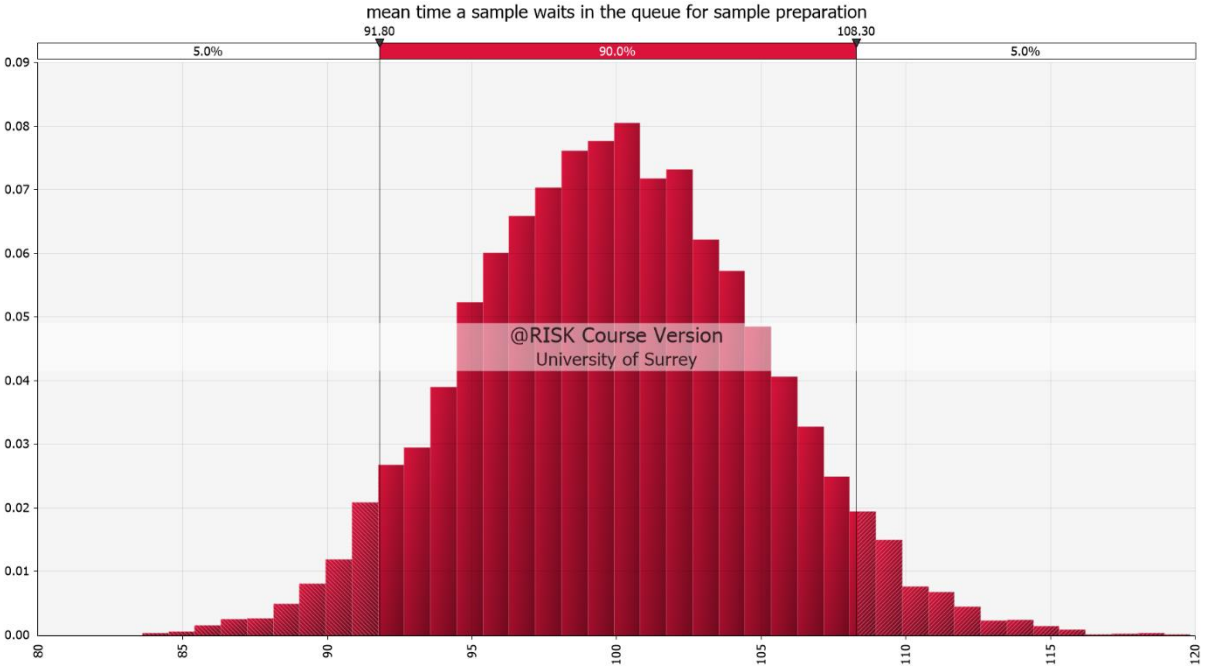

Figure 3: PDF of mean time a sample waits in the queue for sample preparation.

Figure 3 indicates that the average queueing time for sample preparation ranges from 91.8 to 108.3 minutes with a 90% confidence interval. The graph follows approximately a normal distribution.

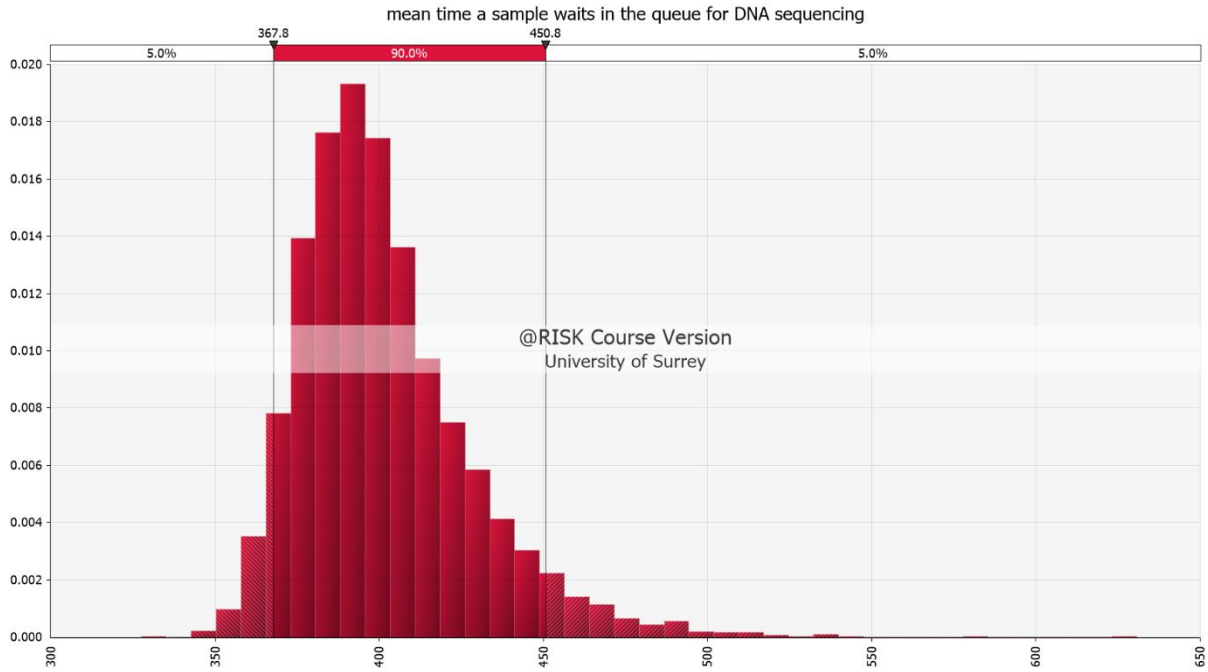

Figure 4: PDF of mean time a sample waits in the queue for DNA Sequencing.

Figure 4 shows that the average queueing time for DNA sequencing ranges from 367.8 to 450.8 minutes with a 90% confidence interval. The graph is right skewed.

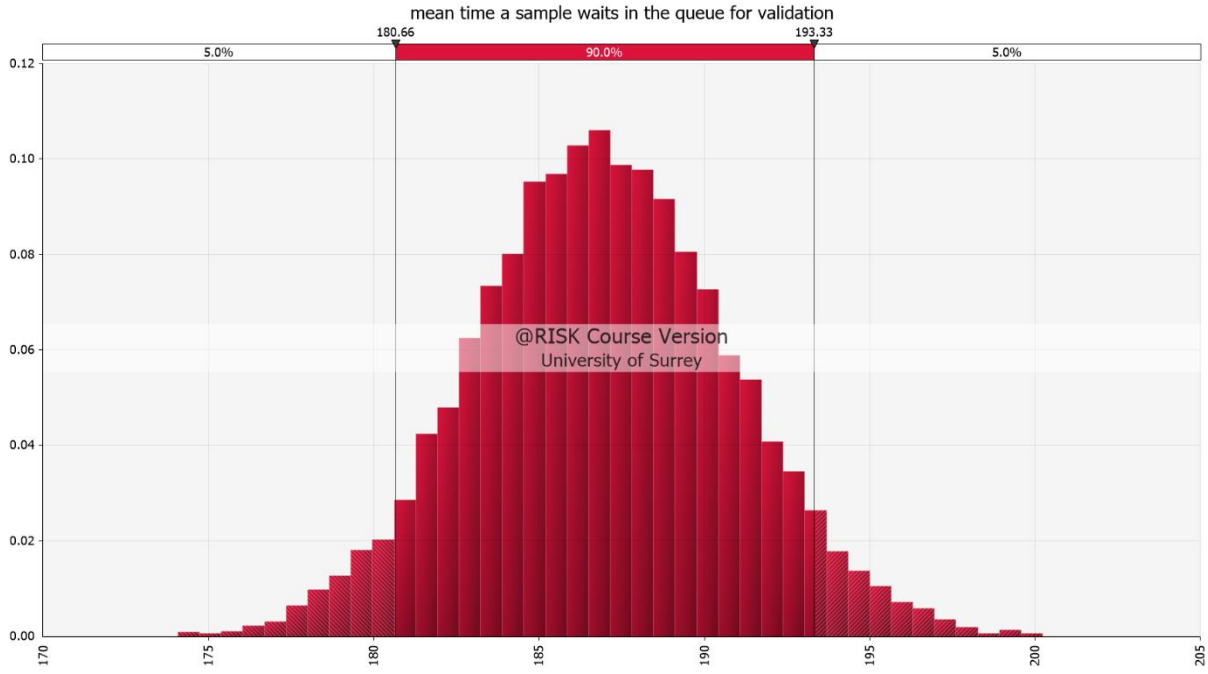

Figure 5: PDF of mean time a sample waits in the queue for Validation.

Figure 5 shows that the average queueing time for sample validation ranges from 180.66 to 193.33 minutes with a 90% confidence interval. The graph follows a normal distribution.

Table 2: Summary of times in each queue of the Rapid DNA forensics process. All units in minutes.

Stage of process	Mean	Median	5th %ile	95th %ile
Mean time to wait for CSI to arrive		659.00	253.35	1,303.04
Mean time to wait for sample preparation	99.979		91.803	108.301
Mean time in queue for DNA sequencer		396.82	367.79	450.77
Mean time to wait from end of sequencing until senior lab researcher is ready to validate samples	186.927		180.659	193.335

c.

Insights from the Queueing time distributions

Queueing time for CSI arrival – From Table 2 the median waiting time for a Crime Scene Investigator (CSI) to arrive is very high, at 659 minutes. The range of waiting times is rather large, with the 5th percentile being 253.35 minutes and the 95th percentile being 1303.04 minutes. This suggests a highly fluctuating and occasionally very long period of waiting. The wide range of variations and extreme scenarios is attributed to geographical obstacles, the accessibility of crime scene investigators, or simultaneous requests. Enhancing the scheduling or availability of CSIs greatly improve reaction times due to the large variability seen.

Queueing time for sample preparation - From Table 2 the mean wait time for sample preparation is 99.979 minutes, however, there is a significant variation in the duration, ranging from 91.803 minutes at the 5th percentile to 108.301 minutes at the 95th percentile. This implies that certain samples are promptly processed, while others have significant delays, due to the workload of lab technicians or the prioritization of activities. These delays are reduced by optimizing other laboratory tasks or increasing the number of technicians.

Queueing time for DNA Sequencing - From Table 2 the median wait time to use the DNA sequencer is 396.82 minutes, with a range from 367.79 to 450.77 minutes. However, waiting to fill the sequencer with eight samples before running it contributes to this wait. Adjusting this policy to start runs at less than full capacity during peak times could reduce waiting.

Queueing time for sample validation – Table 2 shows that the average wait time for sample validation by the senior lab researcher is 186.927 minutes, with a fairly tight distribution from 180.659 to 193.335 minutes. This suggests a more efficient and predictable process but still highlights a dependency on the availability of a single researcher. Increasing staffing or having a backup researcher during peak periods could further reduce this wait time.

d.

Mean number of samples processed in one week = (Mean: 78.163, 5th percentile: 72, 95th percentile: 80)

Mean number of samples matched to the DNA database in one week = (mean: 2, 5th percentile: 2, 95th percentile: 2)

e.

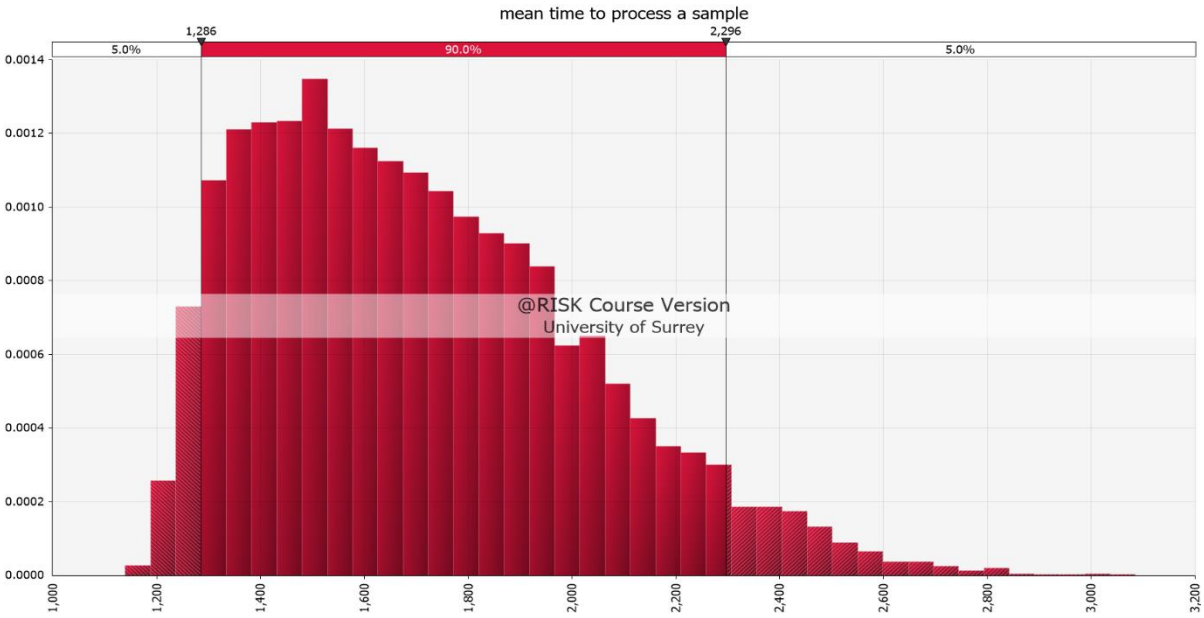

Figure 6: PDF of mean time to process a sample.

Figure 6 shows that the average time to process a sample range from 1286 to 2296 minutes with a 90% confidence interval. The distribution is skewed.

Mean time to process a sample = (median: 1,660.89, 5th percentile: 1,286.41, 95th percentile: 2,295.98)

f.

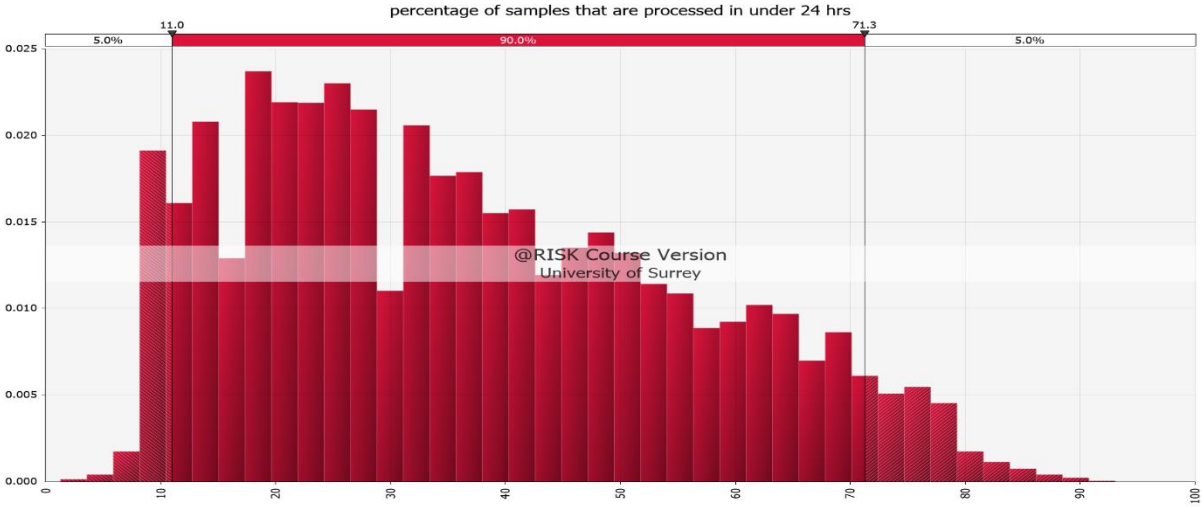

Figure 7: PDF of percentage of samples that are processed in under 24 hrs.

Figure 7 shows that the average percentage of samples that are processed in under 24 hrs range from 11% to 71.3% with a 90% confidence interval.

Percentage of samples that are processed in under 24 hours = (median: 33.750, 5th percentile: 10.976, 95th percentile: 71.250)

g.

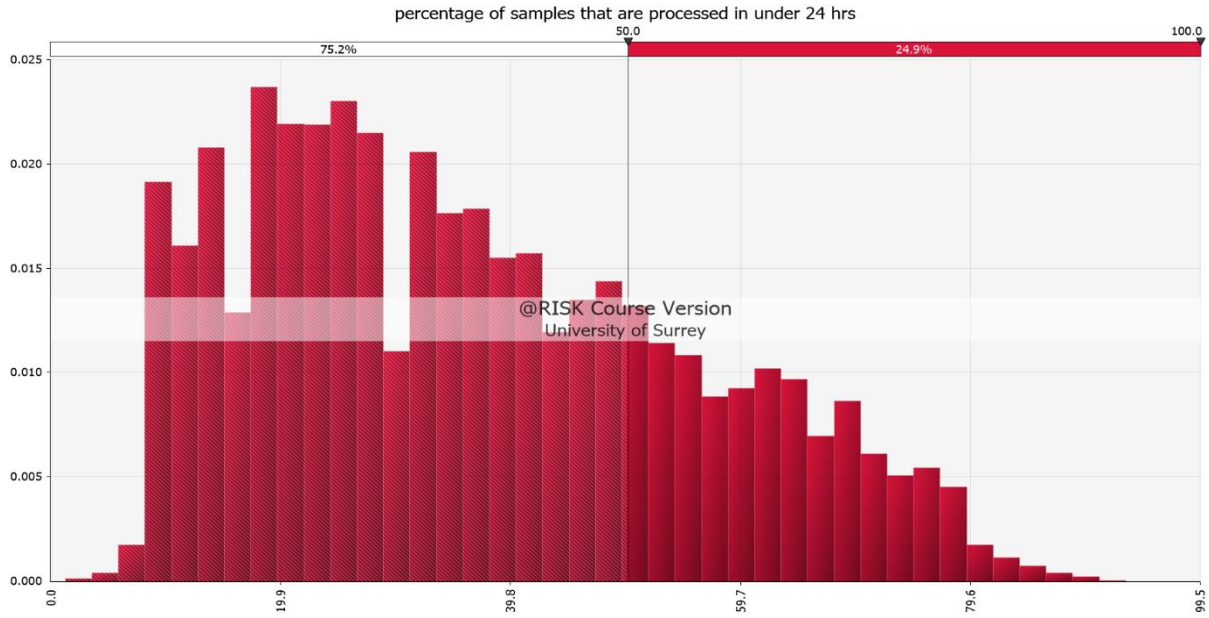

Figure 8: Confidence level of at least half of all samples in one week will be processed under 24 hrs.

From Figure 8 the confidence level of at least half of all samples in one week will be processed in under 24 hrs = 24.9%

h.

Sensitivity Analysis

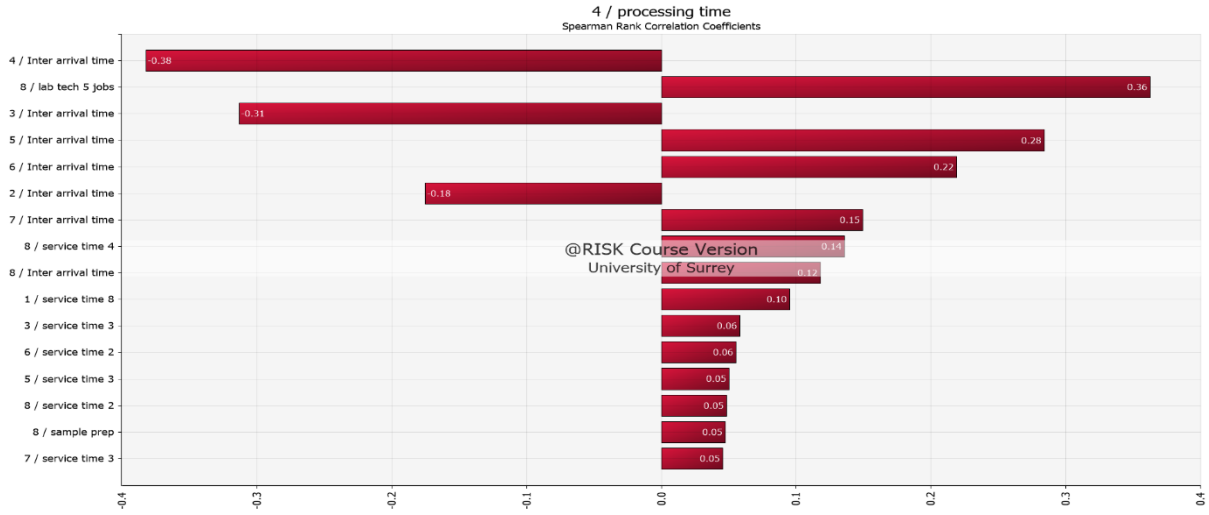

Figure 9: Sensitivity analysis of 4th sample.

Figure 9 shows the sensitivity analysis of 4th sample and indicates some of the key influencers for the overall processing time. However, in this case I cannot rely on the sensitivity analysis as the key influencers keeps on changing from sample to sample due to the randomness of the Queueing model. In other words, I cannot standardise the key influencers for decision making. Hence, I am conducting some Ad-hoc analysis to find the key influencers.

i.

Ad-hoc Analysis

From the queueing time distributions summarized in table 2, it is evident that the top two key influencers for sample processing time are the queueing time for CSI arrival and the queueing time for DNA sequencing, accounting for values 659 minutes and 396 minutes on average respectively.

time in system = wait time + service time

$$T_{DNA} = W_{CSI} + T_3 + T_4 + W_{SP} + T_6 + W_{SEQ} + T_7 + W_V + T_9$$

From the above equation it is evident that decreasing the key wait times by parameter optimization of processes contributing to wait times reduces the sample processing time. For example, parameter optimization of T3 decreases W_{CSI} which in turn reduces T_{DNA}.

What-if Scenario Modelling (Task 5)

What-if Scenario 1

What if the CSI officer takes between 30 and 50 minutes to complete the crime scene sample collection and upload the sample record to the database on the tablet.

The Ad-hoc analysis indicates that W_{CSI} is the most significant factor that can be reduced to improve the overall processing time. Parameter optimization of T3 reduces the queueing time for CSI arrival. Hence, I have considered what-if scenario 1.

Results

Table 3: Scenario 1 results. All units in minutes.

Stage of process	Mean	Median	5th%ile	95th%ile
Mean time a sample waits in the queue for CSI to arrive		250.87	134.08	630.68
Mean time a sample waits in the queue for sample preparation	99.966		92.125	108.140
Mean time a sample waits in the queue for DNA sequencing		358.76	313.70	434.73
Mean time a sample waits in the queue for validation	186.780		180.774	192.851
Mean number of samples processed in one week	87.486		72.000	97.000
Mean number of samples matched in one week	2.0316		2.0000	2.0000
Mean time to process a sample		1,203.85	1,106.98	1,545.74
Percentage of samples processed under 24 hrs		81.250	40.594	93.750

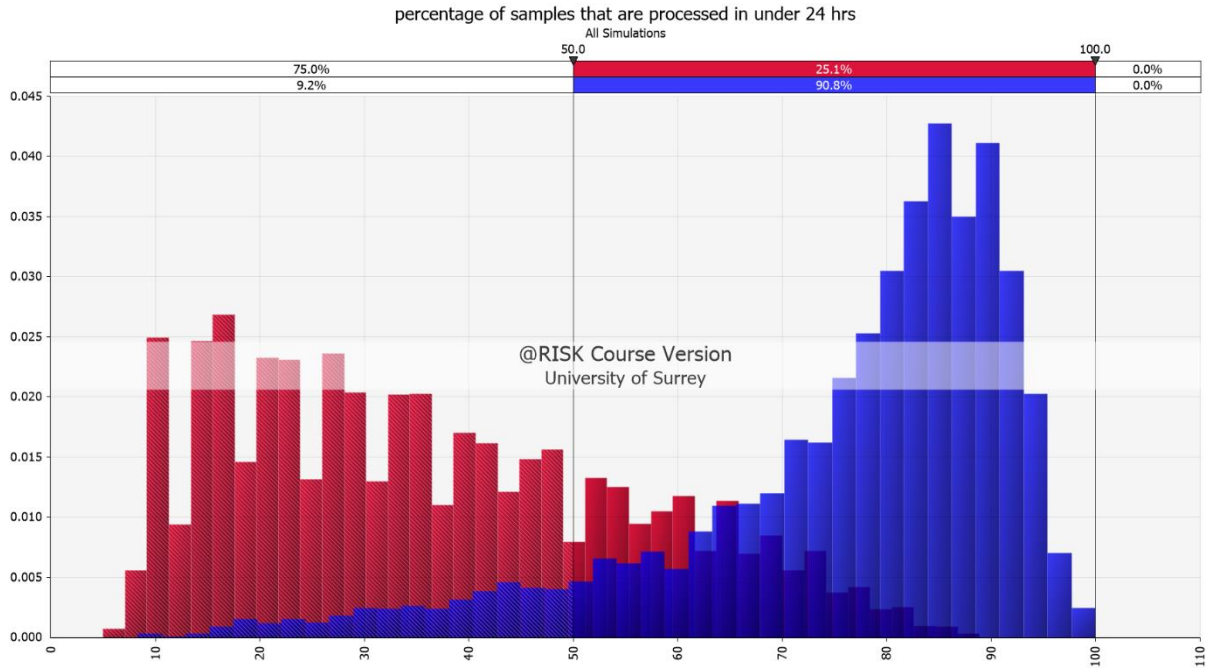

Figure 10: PDF of percentage of samples that are processed in under 24 hrs for scenario 1.

Table 3 shows that the median percentage of samples processed under 24 hours for scenario 1 is 81% which is 48% more compared to the initial scenario in task 4. Also, from figure 10 the likelihood of at least half of samples in one week is processed under one day has increased from 25% to 90%. Moreover, the median processing time has declined from 1660 minutes initially to 1203 minutes.

What-if Scenario 2

What if the DNA Sequencer machine is run at 75% of its capacity (6 samples at a time) instead of 100%.

From the PDF in figure 4 the average queueing time for DNA sequencing ranges from 367.8 to 450.8 minutes, which is the second largest among all queueing times. Optimizing this queueing time by modifying the machine run policy will reduce the overall processing time. Hence, I have considered Scenario 2.

Results

Table 4: Scenario 2 results. All units in minutes.

Stage of process	Mean	Median	5th%ile	95th%ile
Mean time a sample waits in the queue for CSI to arrive		663.94	255.65	1,327.50
Mean time a sample waits in the queue for sample preparation	99.958		91.947	108.201
Mean time a sample waits in the queue for DNA sequencing		284.19	260.83	322.44
Mean time a sample waits in the queue for validation	161.954		156.630	167.314
Mean number of samples processed in one week	79.989		72.000	84.000
Mean number of samples matched in one week	2.0000		2.0000	2.0000
Mean time to process a sample		1,529.45	1,140.09	2,187.70
Percentage of samples processed under 24 hrs		44.872	14.634	89.041

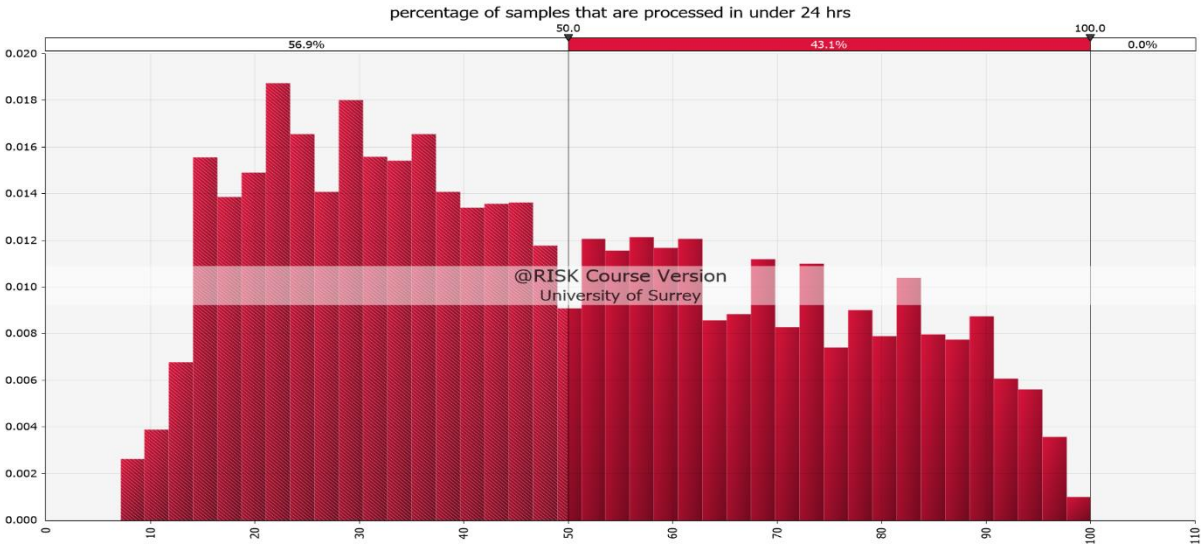

Figure 11: PDF of percentage of samples that are processed in under 24 hrs for scenario 2.

The simulation results shows that the median percentage of samples processed under 24 hours has surged from 33% initially to 44% in scenario 2. Also, from figure 11 the likelihood of at least half of samples in one week is processed under one day has raised from 25% to 43%. Furthermore, the median processing time has dropped from 1660 minutes initially to 1,529.45 minutes.

What-if Scenario 3 (scenario 1 and scenario 2 combined)

- What if the CSI officer takes between 30 and 50 minutes to complete the crime scene sample collection and upload the sample record to the database on the tablet and
- What if the DNA Sequencer machine is run at 75% of its capacity (6 samples at a time) instead of 100%

Optimizing the two major bottlenecks together reduces the overall processing time. Hence considered scenario 3.

Results

Table 5: Scenario 3 results. All units in minutes.

Stage of process	Mean	Median	5th%ile	95th%ile
Mean time a sample waits in the queue for CSI to arrive		252.37	134.03	628.68
Mean time a sample waits in the queue for sample preparation	99.963		92.192	107.985
Mean time a sample waits in the queue for DNA sequencing		256.37	222.17	311.06
Mean time a sample waits in the queue for validation	161.959		156.969	167.069
Mean number of samples processed in one week	88.764		76.000	102.000
Mean number of samples matched in one week	2.0819		2.0000	3.0000
Mean time to process a sample		1,073.48	976.12	1,425.99
Percentage of samples processed under 24 hrs		93.590	51.111	100.000

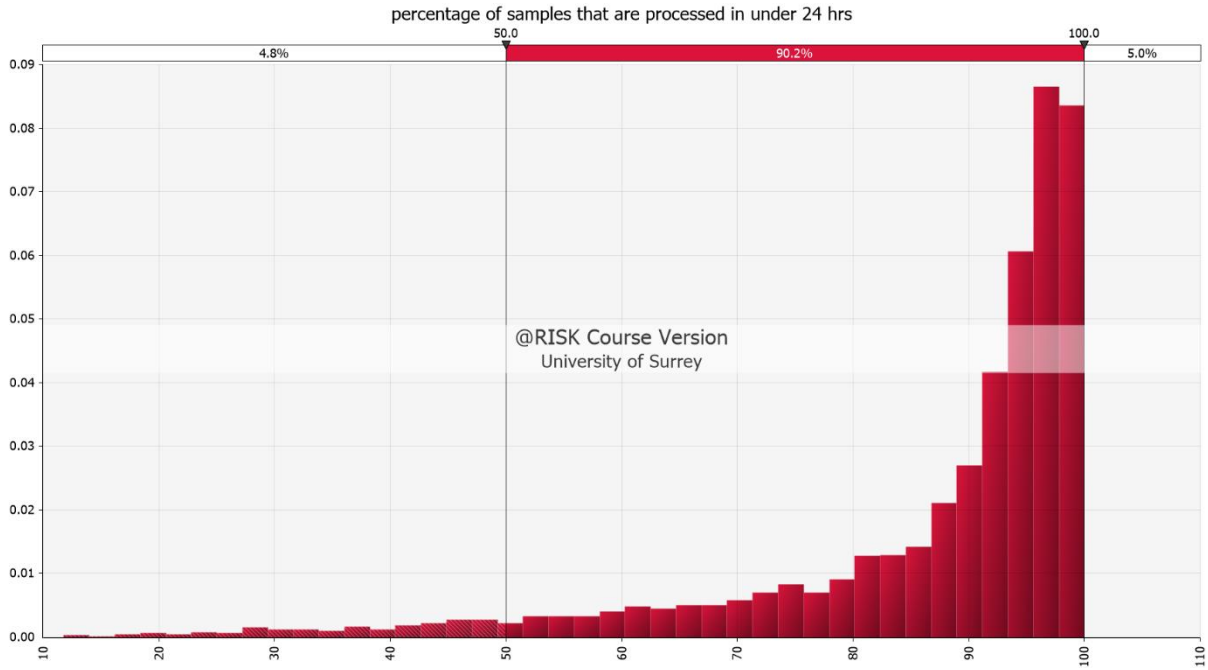

Figure 12: PDF of percentage of samples that are processed in under 24 hrs for scenario 3.

In scenario 3 the median percentage of samples processed under 24 hours is 93.5% which is 59.8% more when compared to the original scenario. Also, from figure 12 the probability of at least half of samples in one week is processed under one day equals 90.2% which is 65.2% more compared to the original scenario in task 4. Furthermore, the median processing time has declined from 1660 minutes initially to 1,073.48 minutes.

Insights and recommendations

Risk analysis of the current rapid DNA forensic process indicates some significant bottlenecks at various stages of the process namely CSI arrival stage, sample preparation stage, DNA Sequencing stage and sample validation stage. The key factor for this bottleneck is the huge waiting time or queueing time for these processes, especially for the CSI arrival and DNA sequencing accounting for 659 minutes and 396 minutes on average respectively. This huge queueing time leads to increased processing time, limiting the number of DNA samples processed in one week which in turn reduces the number of samples matched.

From the what-if scenario analysis conducted, it is evident that optimizing the processes as per scenario 3 gives the best results. However, despite good results, the high cost of running the DNA sequencer at 75% of its capacity is not economical. Moreover, scenario 2 is costly and less efficient as the percentage to samples processed under 24 hrs is relatively low. Hence to keep a balance between cost and outcome, I would recommend modifying the rapid DNA process as per scenario 1, by optimizing the sample collection and uploading process by employing some additional staffs to assist CSI officers.

References

PERT distribution. (2022, May 29). Wikipedia. https://en.wikipedia.org/wiki/PERT_distribution

Hill. (2022, January 28). *Excercise sheet 2 - Solutions*. https://surreylearn.surrey.ac.uk/d2l/le/lessons/254754/lessons/2745241.