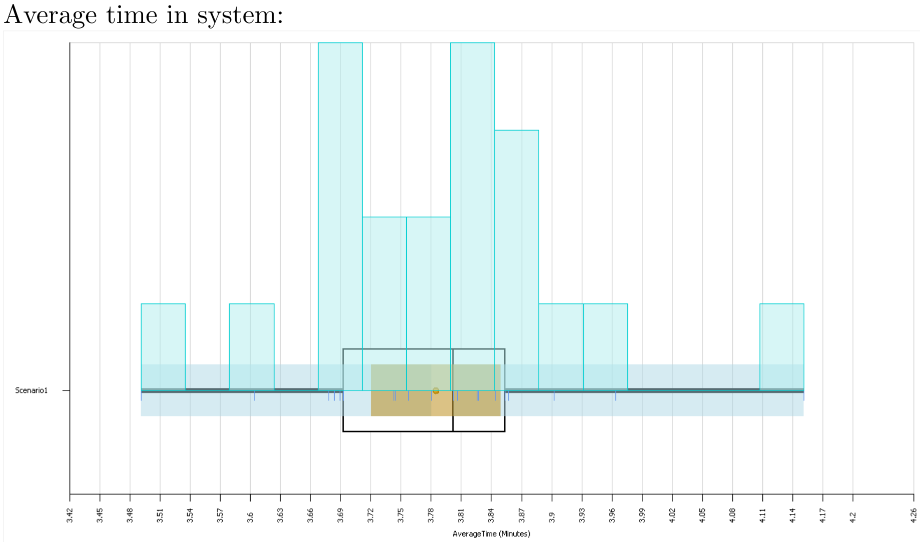
HW3 - ISYE 6644

Marcos Grillo

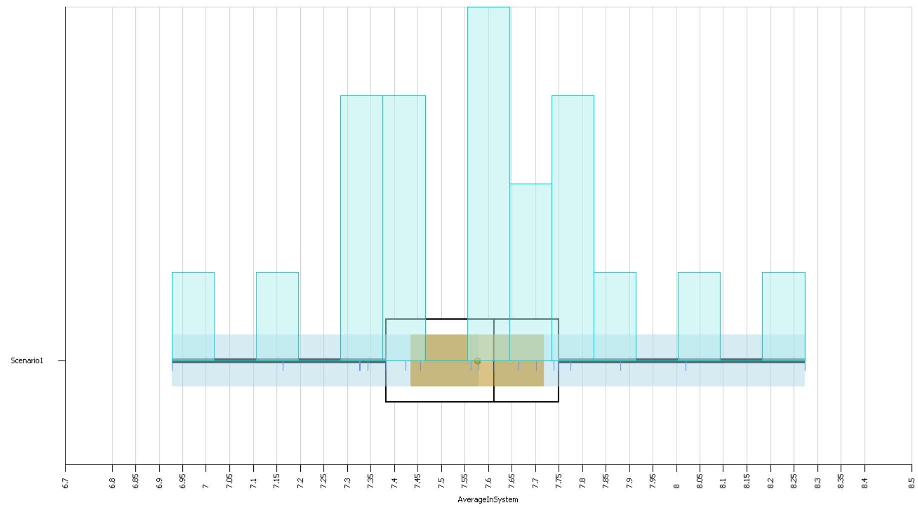
February 2023

1

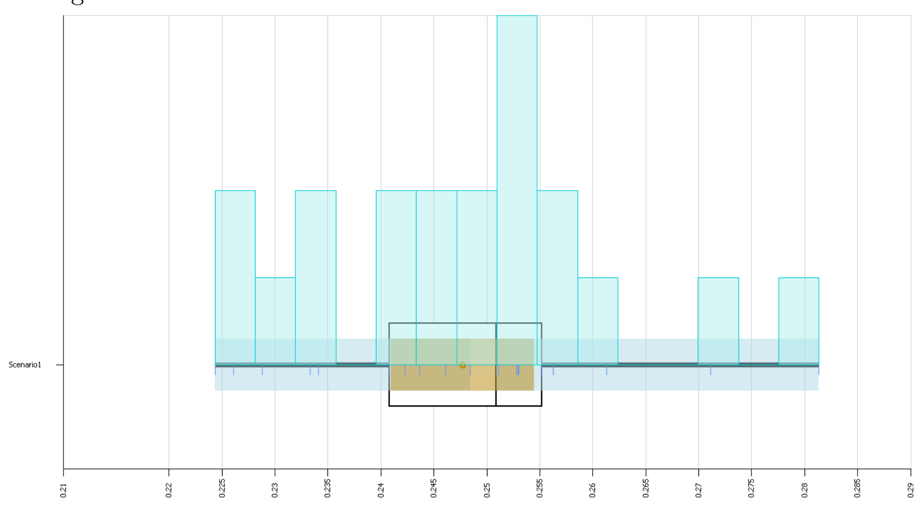
Attached are the (S)MORE plots:
Average time in system:



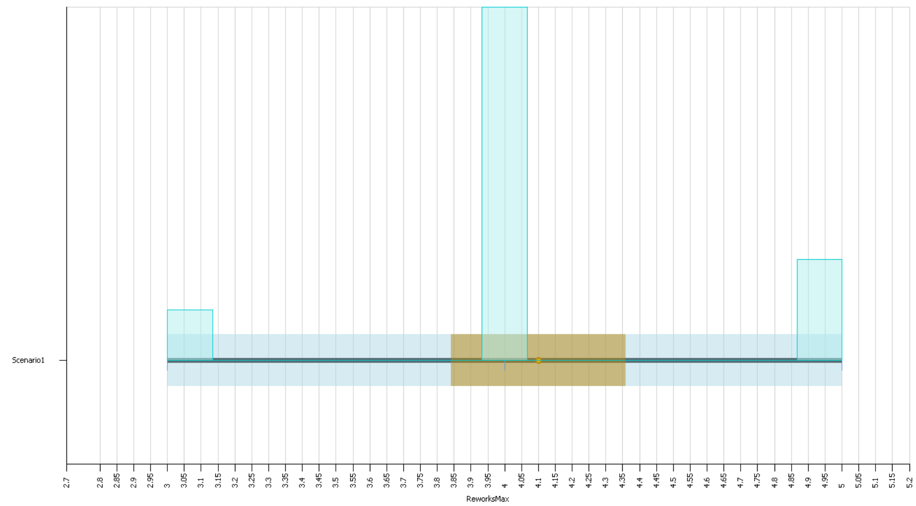
Average number in system:



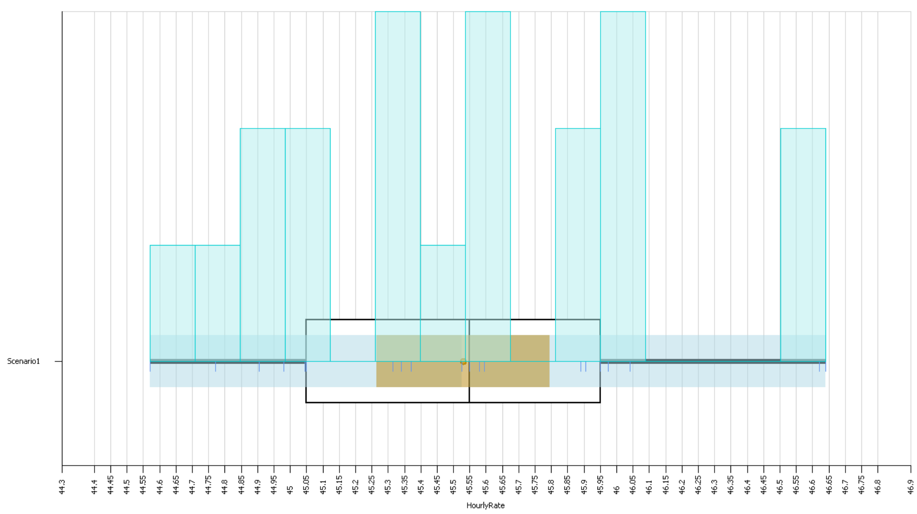
Average number of reworks:



Maximum reworks:



Hourly rate:



2

Please find the enhance model in the zip folder as "Q2.spfx"

3

$$\begin{aligned} S_n^2 &= \frac{\sum_{i=1}^n [X_i - \bar{X}_n]^2}{n-1} \\ &= \frac{\sum_{i=1}^n [(X_i - \mu) - (\bar{X}_n - \mu)]^2}{n-1} \end{aligned} \tag{1}$$

$$\begin{aligned} E[S_n^2] = E[\frac{1}{n-1}\sum_{i=1}^n [(X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2]] \\ = \frac{1}{n-1} E[\sum_{i=1}^n (X_i-\mu)^2 - 2\sum_{i=1}^n ((X_i-\mu)(\bar{X}-\mu)) + \sum (\bar{X}-\mu)^2] \\ = \frac{1}{n-1} E[\sum_{i=1}^n (X_i-\mu)^2 - 2n\sum (X_i-\mu)(\bar{X}-\mu) + n(\bar{X}-\mu)^2] \\ = \frac{1}{n-1} E[\sum_{i=1}^n (X_i-\mu)^2 - 2n(\bar{X}-\mu)^2 + n(\bar{X}-\mu)^2] \\ = \frac{1}{n-1} E[\sum_{i=1}^n (X_i-\mu)^2 - n(\bar{X}-\mu)^2] \\ = \frac{1}{n-1(n\sigma^2 - nVar(\bar{X}))} \end{aligned} \tag{2}$$

$$Var(\bar{X}) = \frac{\sigma_X^2}{n}(1 + 2\sum(1-\frac{j}{n})p_j)$$

$$\begin{aligned} E(S_n^2) &= \frac{n\sigma_X^2 - \sigma_X^2(1 + 2\sum(1-\frac{j}{n})p_j)}{n-1} \\ &= \frac{\sigma_X^2(n - (1 + 2\sum(1-\frac{j}{n})p_j)}{n-1} \\ &= \sigma_X^2 \frac{1 + 2\sum(1-\frac{j}{n})p_j}{n-1} \end{aligned} \tag{3}$$

$$\begin{aligned} E(\frac{S_n^2}{n}) &= \frac{\sigma^2}{n}(1 - \frac{\frac{a_n - 1}{n-1}}{n}) \\ &= Var(\bar{X}_n)(\frac{\frac{n}{a_n} - 1}{n-1}) \end{aligned} \tag{4}$$

4

4.1

Since $E(\epsilon_i) = 0$, the marginal mean of X_i is given as follows:

$$a + 0 + 0.75(0) + 0.25(0) = a$$

Therefore, the marginal mean of X_i is a. For the variance:
The variance of a constant is 0, so we can ignore a.

$$Var(\epsilon_i) = 4^2 = 16$$

$$Var(X_i) = 16 + 0.75^2(16) + 0.25^2(16) = 26$$

$$\beta_0 = 1$$

$$\beta_1 = 0.75$$

$$\beta_2 = 0.25$$

4.2

$$\begin{aligned} Cov(X_i, X_{i-1}) &= Cov(\beta_0\epsilon_i + \beta_1\epsilon_{i-1} + \beta_2\epsilon_{i-2}, \beta_0\epsilon_{i-1} + \beta_1\epsilon_{i-2} + \beta_2\epsilon_{i-3}) \\ &= \beta_0\beta_1\sigma^2 + \beta_1\beta_2\sigma^2 \\ &= (\beta_0\beta_1 + \beta_1\beta_2)\sigma^2 \\ &= (0.75 + 0.75(0.25))16 \\ &= 15 \end{aligned} \tag{5}$$

$$\begin{aligned} Cov(X_i, X_{i-2}) &= Cov(\beta_0\epsilon_i + \beta_1\epsilon_{i-1} + \beta_2\epsilon_{i-2}, \beta_0\epsilon_{i-2} + \beta_1\epsilon_{i-3} + \beta_2\epsilon_{i-4}) \\ &= \beta_0\beta_2\sigma^2 \\ &= 0.25(16) \\ &= 4 \end{aligned} \tag{6}$$

If $i > 2$, the covariance is 0, since there are no matching pairs. We have shown thus that $Cov(X_i, X_{i-j})$ depends exclusively on the lag j. This implies weak stationarity, due to the aforementioned fact and the fact that $\sigma_X^2 = Var(X_i)$, and the mean stays constant.

4.3

Yes, because the covariances do not depend on time, but rather exclusively on lag, as shown above.

4.4

$$\begin{aligned}\rho_j &= Corr(X_i, X_{i-j}) \\ &= \frac{Cov(X_i, X_{i-j})}{\sqrt{Var(X_i)Var(X_{i-j})}}\end{aligned} \tag{7}$$

Thus, we need to find values here for $j = 1$ and $j = 2$, since for $j = 0$ we know the correlation will be 1, and for any values greater than 2 it will be 0.

$$\frac{Cov(X_i, X_{i-1})}{\sqrt{Var(X_i)Var(X_{i-1})}} = \frac{15}{26}$$

$$\frac{Cov(X_i, X_{i-2})}{\sqrt{Var(X_i)Var(X_{i-1})}} = \frac{4}{26} = \frac{2}{13}$$

Thus, $p_1 = \frac{15}{26}$ and $p_2 = \frac{2}{13}$.

4.5

$$\begin{aligned}Var(\bar{X}n) &= \sigma^2 \frac{[1 + 2\sum_{j=1}^{n-1}(1 - \frac{j}{n})p_j]}{n} \\ &= \sigma^2 \frac{1 + 2(\frac{15}{26}(1 - \frac{1}{n}) + \frac{2}{13}(1 - \frac{2}{n}))}{n} \\ &= \frac{\sigma^2}{n}(1 + 2(\frac{15}{26}\frac{15}{26n} + \frac{2}{13}\frac{4}{26n})) \\ &= \frac{\sigma^2}{n}(1 + \frac{38}{26} - \frac{46}{26n}) \\ &= \frac{\sigma^2}{n}(\frac{64n - 46}{26n}) \\ &= \frac{\sigma^2}{n}(\frac{32n - 23}{13n})\end{aligned} \tag{8}$$

4.6

$$\sigma^2 = lim(n- > \infty)nVar(\bar{X}n) = \sigma_X^2 + 2\sum_{n=1}^{\infty} C_k$$

$$\begin{aligned}\sigma_X^2 + 2\sum_{n\ =1}^{\infty} C_k &= 26 + 2(15) + 2(4) \\ &= 64\end{aligned} \tag{9}$$

5

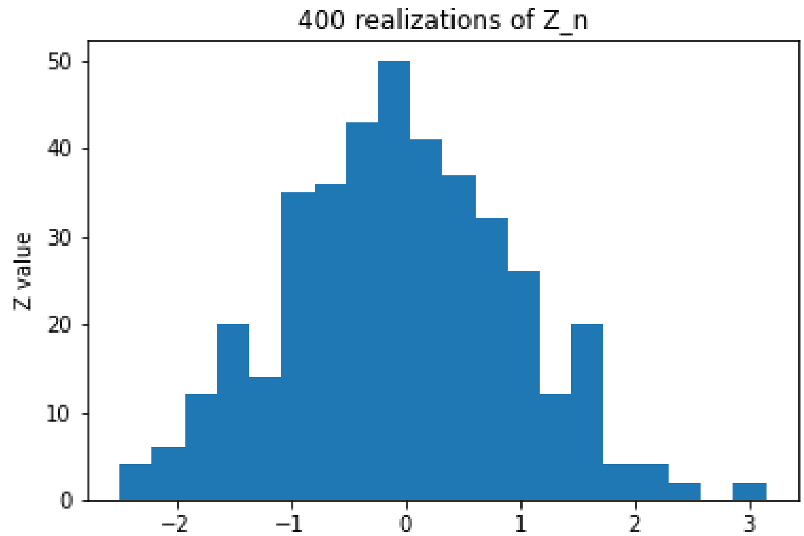
For exercises 5, 6 and 7, please refer to Q5.ipynb for code. I will just discuss outputs and findings here.

5.1

Below we can see the plot generated by the script used for this exercise.

5.2

Below is the histogram for the 400 realizations of Z_n:



An Anderson-Darling test fails to reject $H0$, meaning that we cannot say that the results of this experiment do not follow a normal distribution. Below is a screenshot of the code used.

```
anderson_b = st.anderson(Fiveb_resp)
print("Test statistic:", anderson_b[0])
print("Critical value at 0.05:", anderson_b[1][2])
if anderson_b[0] < anderson_b[1][2]:
    print("Since the test statistic is below the critical value, we fail to reject H0.")
```

```
Test statistic: 0.19181827728380085
Critical value at 0.05: 0.779
Since the test statistic is below the critical value, we fail to reject H0.
```

6

6.1

32.25% of the CI's were found to contain the true mean $\mu = 10$.

6.2

31% of the CI's were found to contain the true mean $\mu = 10$. The fraction seems to actually decrease in this instance. This is because the larger amount of iterations per run lowers the variance, which in turn shrinks the confidence intervals and makes it less likely for them to include the true mean.

6.3

93.75% of the CI's were found to contain the true mean $\mu = 10$. This can be explained by z-score not contemplating the variance of the sample, thus not being affected by it shrinking with sample size.

7

For both scenarios, 100% of the CI's contained the true mean $\mu = 10$.