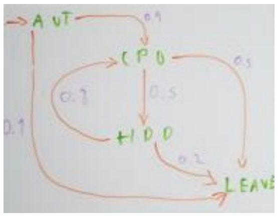
CS 350 – Fall 2020, Assignment 4

Answer 1

a) 

b) We know that the arrival of requests to the system follows a Poisson Distribution with a mean of 10 requests per second, then we have that the average arrival rate of requests in the system is 10/per second. If we couple this with the fact that we know that the probability of authentication is 0.9, then we have an arrival rate of 9 requests per second into the CPU via the Authenticator. This is however not the final answer, as we still have to account for requests that return from the HDD.

   As such, we know that, out of the 9/per second that reach the CPU, 50% go into the HDD, so we know that 4.5 requests per second reach the HDD initially, and of these, 80% return to the CPU, so we have a final average arrival rate into the CPU of 12.6 ($0.8 \times 4.5 + 9 = 12.6$).

c) The CPU will be the bottleneck. As the authenticator has a utilization of $\rho_a = 10 \times 0.02 = 0.2$, the CPU has a utilization $\rho_c = 12.6 \times 0.05 = 0.63$, meanwhile the HDD has a utilization of $\rho_h = 6.3 \times 0.1 = 0.45$.

d) The average number of requests present in the system at any point in time is the sum of the average number of requests present in each of the components of the system. As such we need to calculate this for each of the members: $q_{aut} = \frac{0.2}{1-0.2} = 0.25$, $q_{cpu} = \frac{0.63}{1-0.63} = 1.7$, $q_{hdd} = \frac{0.45}{1-0.45} = 0.82$ so the answer is $0.25 + 1.7 + 0.82 = 2.77$ requests in the system.
   Should be noted I assumed the utilizations calculated in c were correct when calculating this answer.

e) If I was told that the HDD had a limited queue, then it would become an M/M/1/K system, so I would have to account for the probability of losing requests in my calculations. Although in theory I could approach the problem in the same way, the utilization calculation would be different, and so would the calculation for the average number of requests in the HDD.

   The approach would be the exact same except the requests rate for the HDD, the utilization of the HDD and the number of requests in the HDD would have to be calculated differently using: $\lambda' = \lambda \times (1 - P(S_K))$, $\rho' = \lambda' \times T_s$, $q'_{hdd} = \frac{\rho'}{1-\rho'}$.

   In addition to this the rate of arrival in the CPU would change to reflect these changes, with the new rate being $\lambda_{cpu} = (0.8 \times \lambda'_{hdd} + 9)$, and once again all CPU calculations would need to be redone to reflect this.

Answer 2

A table with all the values obtained via the spreadsheet can be found at the bottom of the PS, labeled as TABLE 2 (Page 6).

a) If we perform average calculations on each of the PoCs, we get: $Avrg_1 = 52.57$, $Avrg_2 = 79.48$, $Avrg_3 = 76.57$. As such, the system with the best overall performance average would be PoC 2.

b) If we perform standard deviation calculations on each of the PoCs, we get: $Std_1 = 11.36$, $Std_2 = 32.27$, $Std_3 = 4.04$. As such, the system with the most deterministic behaviour would be PoC 3.

c) If we use x = 980 and y = 1000 then the % of confidence we are looking for is 98% or $\alpha = 1 - 0.98 = 0.02$, so applying this value to the equation $E = Z_{\alpha/2} \times \sqrt{\frac{S^2}{N}}$ to each of the three PoCs, we get:

   PoC 1: $E = 2.33 \times \sqrt{\frac{129.16}{21}} = 5.78$, [46.79, 58.35]
   PoC 2: $E = 2.33 \times \sqrt{\frac{1041.66}{21}} = 16.41$, [63,07, 95.89]
   PoC 3: $E = 2.33 \times \sqrt{\frac{16.36}{21}} = 2.06$, [74.51, 78.63]

   With that said, we cannot actually achieve the SLA with a=45 and b=60 with the previously stated x and y using any of these machines. However, PoC 1 is quite close to these values, despite not being exactly there.

   It should be said that I utilized a spreadsheet to calculate the average, standard deviation and variance used above, and rounded them to the first two decimal cases, however even using more detailed values did not result in any change in this answer.

d) If we chose an x = 999 and y = 1000, then we need a confidence interval of 99.9% or $\alpha = 1 - 0.999 = 0.001$, so applying this value to the confidence interval equation (as used above), in PoC 2 we get:

   $E = 3.27 \times \sqrt{\frac{1041.66}{21}} = 23.03$, [56.45, 102.51]

   As such, in order not to be sued, we can only offer an SLA with $x = 999$ and $y = 1000$ if we make $a = 56.45$ and $b = 102.51$.

e) Since the confidence interval here is 98%, we already know that the boundaries for this interval in PoC 3 are [74.51, 78.63], as such, we cannot offer this SLA immediately.

Assuming that the average, standard deviation, and variance don't change with the number of samples, and the only thing that actually changes is the N representing said number of sample, then in order for us to be able to guarantee this SLA, we need to collect x more samples.

In order to get this value, we must first understand what the distance from the average that is being requested is, since we know the average for PoC 3 is 76.57, then we are being asked about a confidence interval that is at most 1.57 ms away ($76.56 - 75 = 1.57$, $78 - 76.56 = 1.43$), this added to our confidence interval of 98% allows us to calculate the number of samples x:

$E = Z_{a/2} \times \sqrt{\frac{S^2}{21+x}} \equiv 1.57 = 2.33 \times \sqrt{\frac{16.36}{21+x}} \equiv (\frac{1.57}{2.33})^2 = \frac{16.36}{21+x} \equiv 0.45 = \frac{16.36}{21+x} \equiv 21 + x = \frac{16.36}{0.45} \equiv 21 + x = 36.3 \equiv x = 15.3$

As such, and assuming that the standard deviation never changes, I need to ask my intern to collect at least 16 more samples.

Answer 3

A table with all the values obtained via the spreadsheet can be found at the bottom of the PS, labeled as TABLE 3.1 (Page 7).

a) Based purely on these samples, the average salary of someone pursuing a career in CS is 53222.22

b) If we work our way back using $E = Z_{\alpha/2} \times \sqrt{\frac{S^2}{N}}$:

$E = Z_{\alpha/2} \times \sqrt{\frac{S^2}{N}} \equiv 10000 = Z_{\alpha/2} \times \sqrt{\frac{1091211111}{10}} \equiv\equiv 10000 = Z_{\alpha/2} \times 10446.1 \equiv Z_{\alpha/2} = \frac{10000}{10446.1} \equiv Z_{\alpha/2} = 0.96$

From here we know that $1 - \frac{\alpha}{2} = 0.8315 \equiv 0.1685 = \frac{\alpha}{2} \equiv \alpha = 0.337$. So our confidence level that someone pursuing a career in CS will have a salary of +/- 10000 from 53222.22 is $x = 1 - 0.337 = 0.663$ or 66.3%.

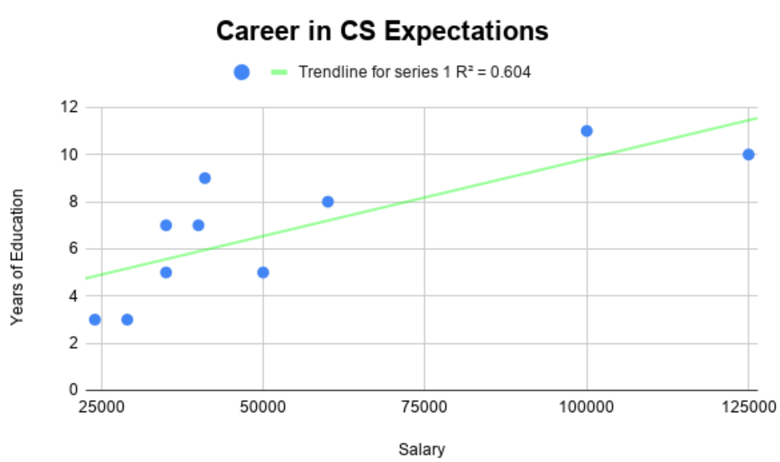
c) Assuming that by interviewing more people only the number of people being interview changes, and not the mean, standard deviation and variance, then we can work our way back from the equation used above.

A confidence level of 95% means an alpha of $\alpha = 1 - 0.95 = 0.05$, so using the equation above with x signifying the number of additional people we interviewed we get that:

$E = Z_{\alpha/2} \times \sqrt{\frac{S^2}{10+x}} \equiv 10000 = 1.96 \times \sqrt{\frac{1091211111}{10+x}} \equiv (\frac{10000}{1.96})^2 = \frac{1091211111}{10+x} \equiv 26030812.1616 = \frac{1091211111}{10+x} \equiv 10 + x = \frac{1091211111}{26030812.1616} \equiv x + 10 = 41.92 \equiv x = 31.92$

As such, and with the assumptions stated, we would need to interview 32 more people before we could increase our confidence level to 95%.

d) If we take a look at the chart below, we can see that there is a clear positive correlation between the number of years of education, and the salary.



e) If we decide to stay in school for five or more years, then our new dataset has an average salary of 60750, a standard deviation of 33661.13 and a variance of 1133071429. Based on this, we can calculate the confidence interval between [50000, 71500].

To begin with, we need to find what the E value being asked of us here is, since both values are exactly 10750 away from the mean, that will be the E, so now we can calculate:

$$E = Z_{\alpha/2} \times \sqrt{\frac{S^2}{N}} \equiv 10750 = Z_{\alpha/2} \times \sqrt{\frac{1133071429}{8}} \equiv 10750 = Z_{\alpha/2} \times 11901 \equiv Z_{\alpha/2} = \frac{10750}{11901} = 0.90$$

From here we know that $1 - \frac{\alpha}{2} = 0.8159 \equiv 0.1841 = \frac{\alpha}{2} \equiv \alpha = 0.3682$. So our confidence level that someone pursuing a career in CS will have a salary between 50000USD and 71500 USD, after completing 5 or more years of education is $x = 1 - 0.3682 = 0.6318$ or 63.18%.

The table with the values used for this subproblem can be found below under TABLE 3.2 (Page 7)

TABLE 2

Sample #	PoC 1	PoC 2	PoC 3
1	68	75	76
2	37	112	81
3	53	106	81
4	70	52	78
5	47	109	76
6	44	106	74
7	58	76	78
8	42	118	70
9	44	76	73
10	45	92	79
11	52	98	73
12	50	21	81
13	61	27	81
14	40	112	74
15	44	68	72
16	54	120	84
17	34	86	73
18	69	29	79
19	58	66	81
20	72	27	72
21	62	93	72
-	-	-	-
Avrg	52.57	79.48	76.57
Std	11.36	32.27	4.04
Var	129.16	1041.66	16.36

TABLE 3.1

Person #	Yearly Income ($)	Years of Education
1	125000	10
2	100000	11
3	40000	7
4	35000	7
5	41000	9
6	29000	3
7	35000	5
8	24000	3
9	50000	5
10	60000	8
-	-	-
Avrg	53222.22	6.67
Std	33033.48	2.78
Var	1091211111	7.73

TABLE 3.2

Person #	Yearly Income ($)	Years of Education
1	125000	10
2	100000	11
3	40000	7
4	35000	7
5	41000	9
6	35000	5
7	50000	5
8	60000	8
-	-	-
Avrg	60750	7.75
Std	33661.13	2.19
Var	1133071429	4.79

All table calculations preformed using Google Spreadhseets using the AVERAGE, STDEV, and VAR operators, and the ROUND(..., 2) to obtain the rounded results.

Spreadsheet can be found here: `https://bit.ly/2FrxwOz`