Technion – Israel Institute of Technology
Faculty of Industrial Engineering and management

Lecturer: Prof. Avishai Mandelbaum

Student no. : ________________
Name : ________________

Service systems engineering - 096324

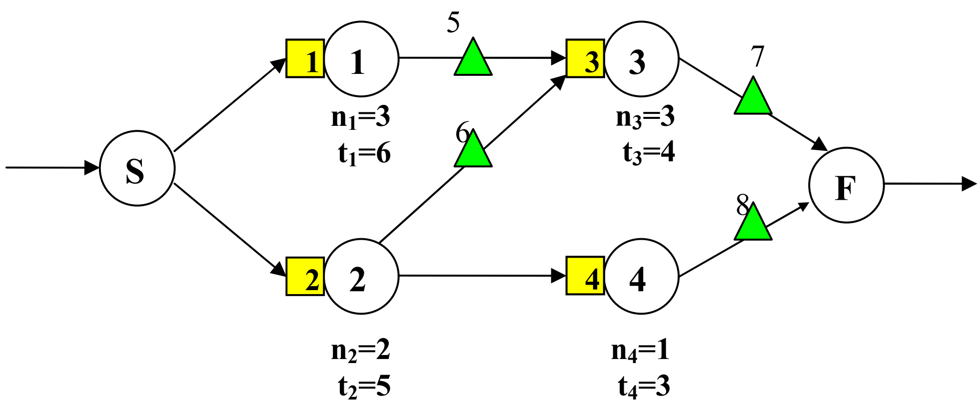
- The duration of the test is 3 hours. It is the examinee's responsibility to hand it in on time.
- The test is comprised of 28 pages, including this page.
- Fill in the details required at the head of the page clearly.
- The questions should be answered in the designated space, **which should be sufficient**. Unnecessary text will subtract from the score, rather than add to it.
- Explanations or demonstrations should be provided **only** when asked for specifically.
- This is a closed book test without the aid of a calculator.

Good Luck!

Question no. 1

A stochastic dynamic PERT/CPM network is given, as seen in Graph. 1 and as explained in class:
t_i is the mean service time in station i and n_i is the number of servers in station i, $i = 1,2,3,4$.
New projects are introduced to the system according to Poisson procedure with a mean time between arrivals of 3.5 days.

We assume the network is in stable condition.



Graph 1

1.1. The network has 8 queues, numbered 1 to 8. Which of the queues derive from synchronization gaps and which from limited resources?

Synchronization queues: 1, 2, 3, 4
Resources queues: 5, 6, 7, 8

1.2. What is the utilization rate of each of the 3 servers handling activity 3? Explain your answer in a sentence or two.

Server utilization: $4/(3.5*3)=0.3809$

Explanation: since the network is in stable condition the arrival rate for each station is 1/3.5 projects a day. Each project requires 4 workdays and the work is divided between 3 servers.

1.3. What is the minimal number of servers required to handle activity 3, so that the utilization of each of them will not exceed 33%?

Minimal number: $[4/(3.5*n_3)]\leq 0.33$ - The result is 4 servers.

1.4. T_{DS} shall mark the duration of performing a project arriving to a stable condition system.
Explain as precisely as you can, but in no more than 2 sentences, why $10<ET_{DS}$.

Explanation: The explanation includes 2 parts. Since the network is stochastic, we already know that $10<ET_{DS}$ (in the Homework we demonstrated that $E(\max X_i) > \max E(X_i)$). Also, the dynamics of the system creates queues that prolong even more the duration of the mean project.

Question no. 2

Graphs 2 and 3 on page 5 describe mean arrival rates and service times, as a function of the time of day. (This is a focus that was analyzed in class). Give two possible explanations to the fact that around 10:00 and around 15:00 an obvious increase occurs in both mean service time and arrival rates. As support to one of your explanations, use graphs 4 and 5 on page 6, including a short explanation of the graphs' content.

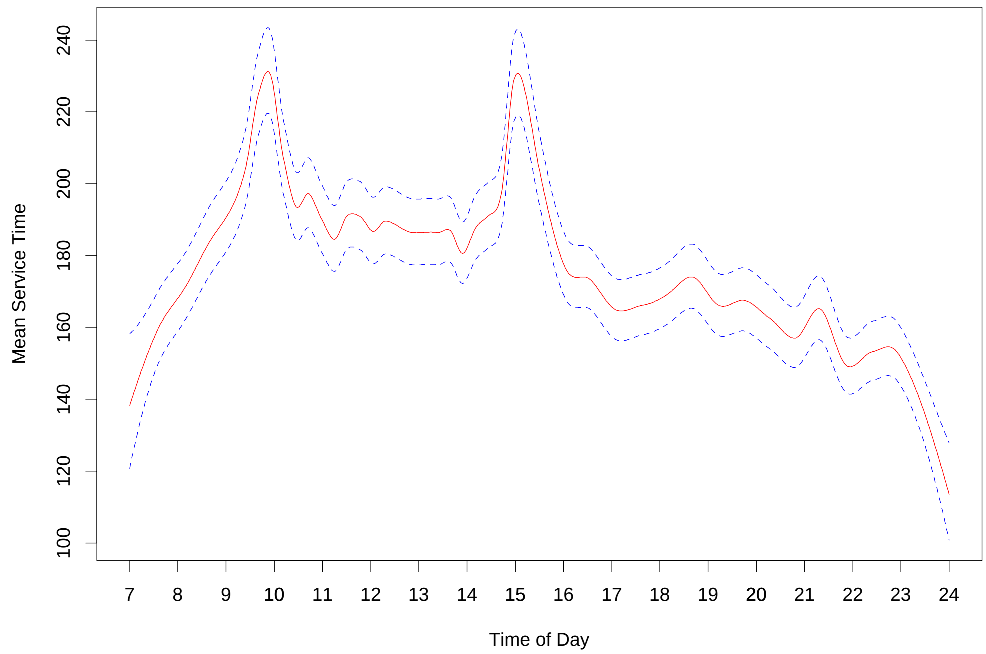
Explanation 1: The simple explanation is that since these are busy hours the customers who reach customer service waste time complaining and so the service time is lengthened.

Explanation 2: Graphs 2 and 3 show that the more patient customers are the ones who require a longer service time. Therefore, the customers with the longer service time are the ones that survive and the service time is lengthened.

Graph 4 shows: The survival function of 4 different service types. It can be seen from the graph that stocks customers, who are customers with higher priority, have longer service time than new and regular customers.

Graph 5 shows: The patience risk function of regular customers and high priority customers. Since the risk function of regular customers is higher, their mean patience is lower.

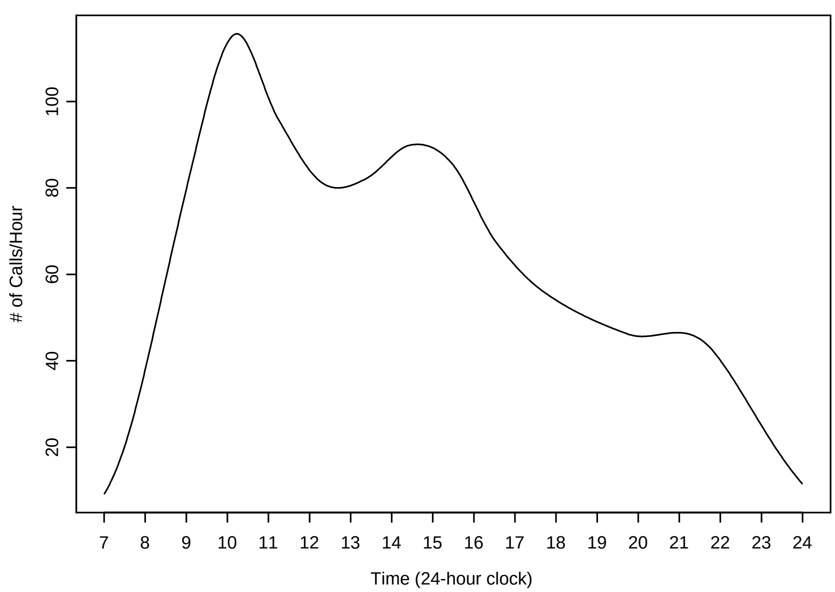
Figure 12: Mean Service Time (Regular) vs. Time-of-day (95% CI) ($n = 42613$)



Graph Number 2

Arrivals: Inhomogeneous Poisson

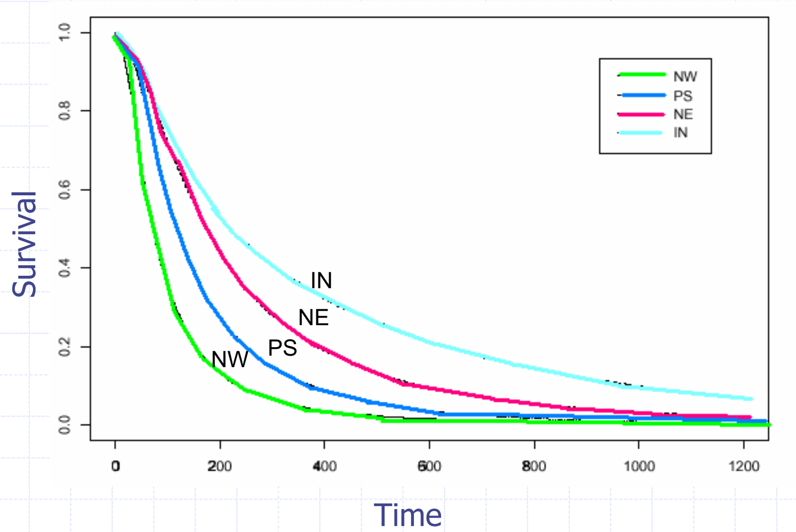
Figure 1: Arrivals (to queue or service) – "Regular" Calls



Graph Number 3

Service Time

Survival curve, by Types



Means (In Seconds)

NW (New) = 111

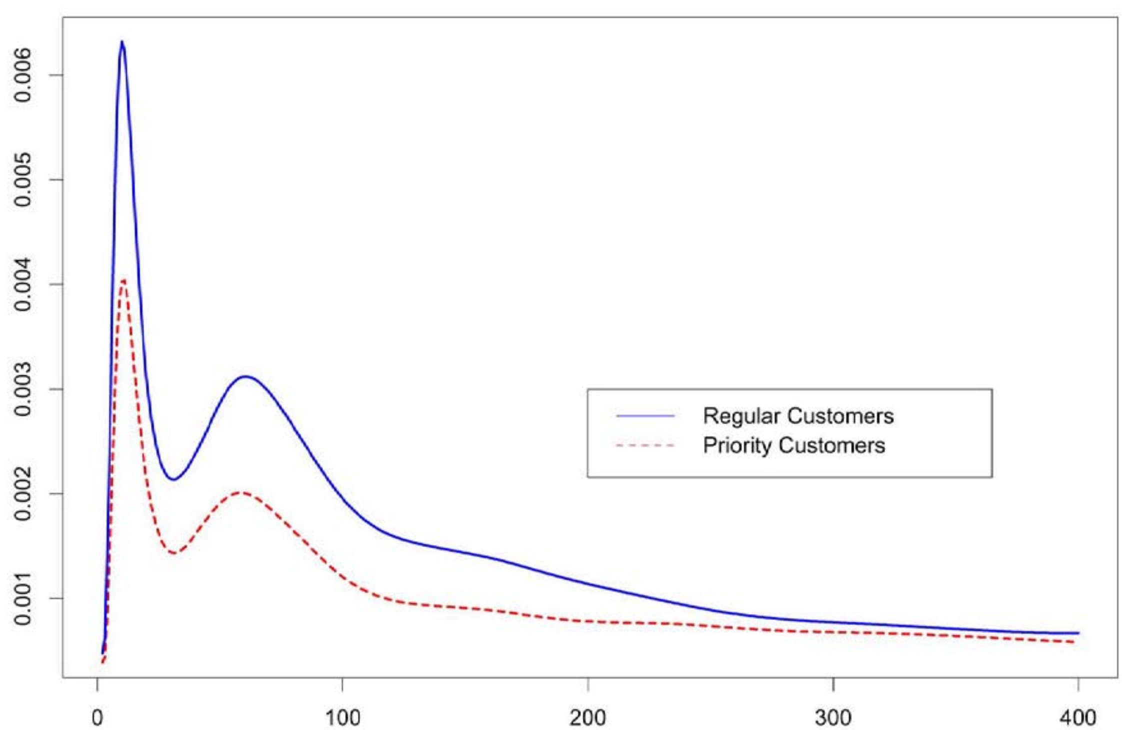
PS (Regular) = 181

NE (Stocks) = 269

IN (Internet) = 381

Graph Number 4

Hazard Rate: Empirical (Im)Patience



Graph Number 5

Question 3

It was said in class, and in other courses, that the behavior of the service providers is affected drastically by working conditions and incentives.

3.1. The graphs 6 and 7 on pages 8-9 support the above statement. Explain (one sentence for each graph).

Graph 6: Shows that giving incentives according to number of handled calls causes the service providers to hang up on customers.

Graph 7: On the afternoon surgeons receive payment per each surgery operation, so they postpone shorter operations to the afternoon.

3.2. What are the administrative consequences of the two phenomena seen in the graphs? In other words, what corrective measures, if any, are required to change the behavior described in the graphs? (One sentence for each phenomenon)

Graph 6: The phenomenon is problematic. It can be solved with tighter supervision or by implementing different service measures other than number of calls.

Graph 7: The phenomenon is positive in this case. More difficult operations are performed in the morning hours, when surgeons are more vigorous.

3.3. **Is the histogram in Graph 7 suitable for exponential distribution? Give 2 reasons (a sentence for each reason) to support your answer.**

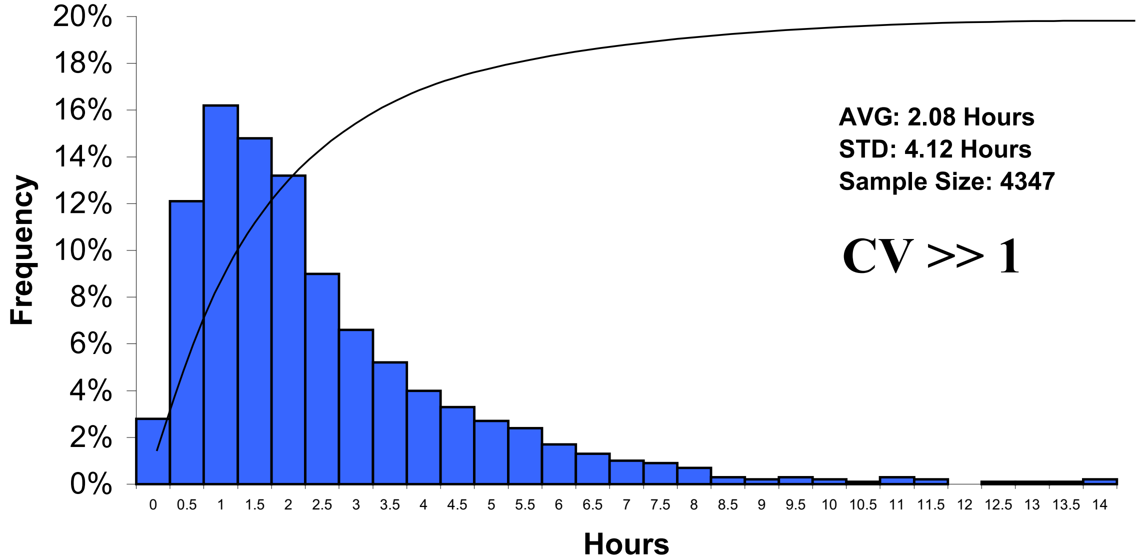
The histogram is **not** suitable for exponential distribution

Reason 1: CV>>1 in contrast to CV=1 for exponential distribution.

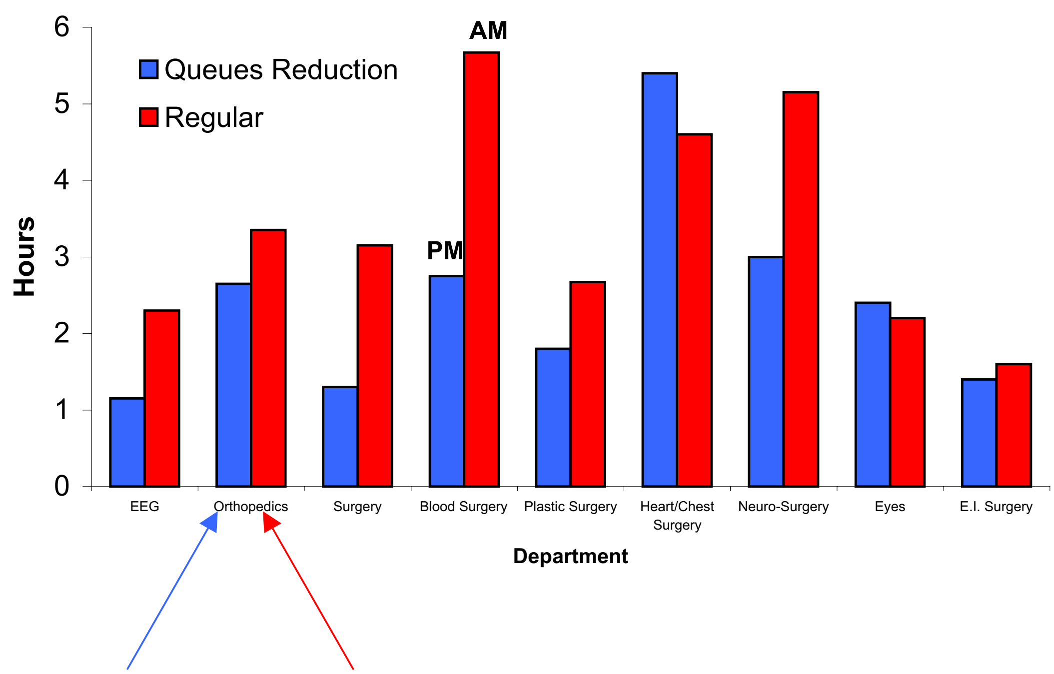
Reason 2: In exponential distribution the relation between sequential columns should be constant (exponential discrete proximity is geometrical). In particular, the histogram should be in monotonous decline.

Operations Time In a Hospital

Operations Time Histogram:



Operations Time - Morning vs. Afternoon:



Afternoon, by Case

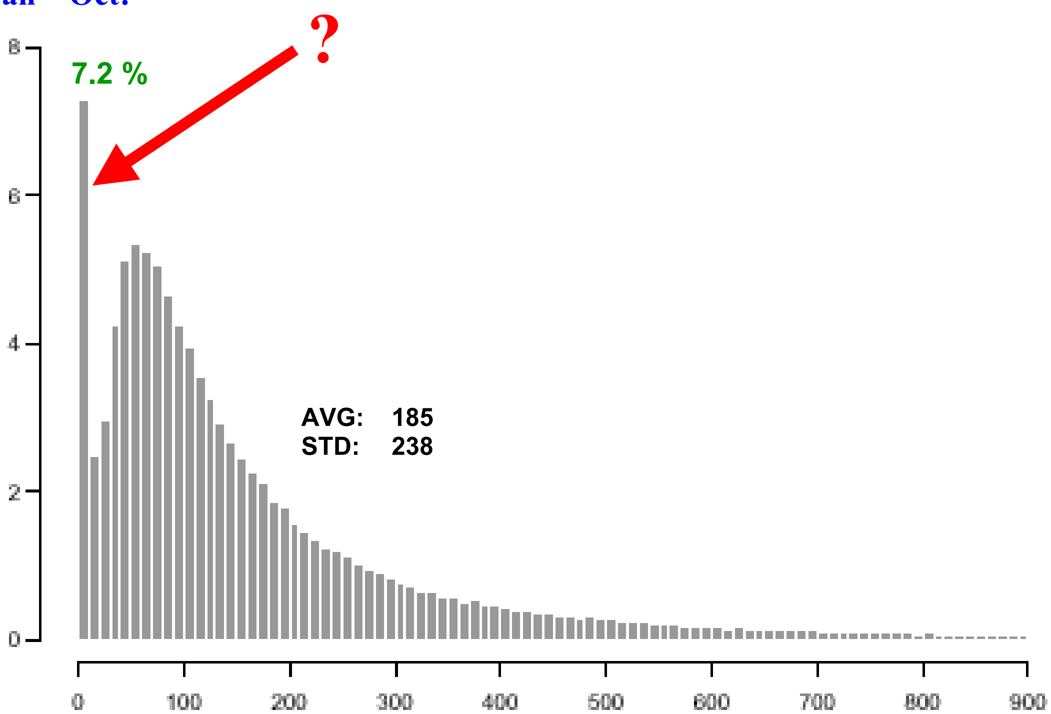
Morning, by Hour

Ethical?

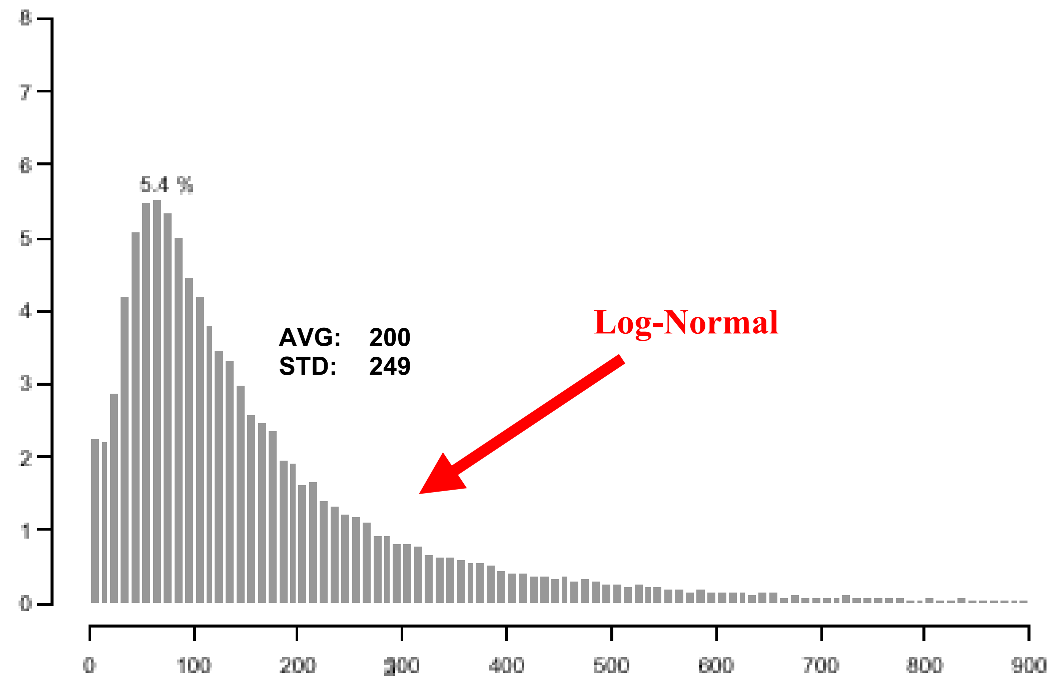
Even Doctors Can Manage!

Beyond Data Averages
Short Service Times

Jan – Oct:



Nov – Dec:



Graph Number 6

Question 4

An Erlang-A (M/M/N+M) model has 4 parameters:
λ = Arrival rate
μ = Service rate
θ = Individual abandonment rate
N = Number of servers

4.1. Under which values of the parameters is the model ergodic (reaches a stable state)?

Conditions: Since there are abandonments ($\theta > 0$) the model is ergodic under any variety of parameters.

4.2. P_{ab} **and** EW_q **shall mark the percent of abandonment and the mean time of waiting in queue in stable state, respectively. Demonstrate that:**

$$P_{ab} = \theta * EW_q$$

Proof:

$\lambda P_{ab} = \theta EL_q$
By Little's law
$EL_q = \lambda EW_q$
$\Downarrow$
$P_{ab} = \theta EW_q$

Daily (24 hrs.) calls data is presented in table 8 on page 12.

4.3. Using mathematical symbols, describe how, based on table 8, you will estimate parameter μ for the time period between 07:15 and 07:45.
(For instance, in order to estimate λ, we shall use A to mark the number of calls arrived at the call center between 07:15 and 07:45. Hence, estimate $\hat{\lambda}$ for λ is given by $\hat{\lambda} = 2A$, when λ^ units are the number of calls per hour).

$\hat{\mu}$ **- estimate for μ:** Denote by N_{ser} the number of customers served during the interval (not the number of customer arrivals), by S we define the total service time of customers served during the interval. Hence

$$\hat{\mu} = N_{ser} / S$$

4.4. Let $\hat{\theta}$ be an estimate for θ and let N_{ab} be the number of abandoners between 07:15 and 07:45. Define by W_{ab} the total waiting time of customers that abandoned during the interval.
A student who is yet to study Service System Engineering proposed N_{ab} / W_{ab} as an estimate for θ.
What is your opinion of his proposal? (Answer in a single sentence)
If you do not accept it, propose an alternative estimate for $\hat{\theta}$.

Opinion of the proposal: The student's proposal is certainly erroneous since it does not take into account the patience of all the customers which did not abandon. Consequently it is an incorrect estimate.

Alternative estimate $\hat{\theta}$, if relevant: W_{ser} will denote the total of waiting periods of the customers served during the interval. Hence

$$\hat{\theta} = N_{ab} / (W_{ab} + W_{ser})$$

customer_i	priority	type	date	vru_entry	vru_exit	vru_time	q_start	q_exit	q_time	outcome	ser_start	ser_exit	ser_time
0	0	PS	991109	7:03:23	7:03:34	11	0:00:00	0:00:00	0	AGENT	7:03:34	7:03:42	8
0	0	PS	991109	7:04:02	7:04:18	16	0:00:00	0:00:00	0	AGENT	7:04:17	7:05:19	62
0	0	PS	991109	7:04:45	7:04:56	11	0:00:00	0:00:00	0	AGENT	7:04:55	7:05:52	57
24512279	2	PS	991109	7:05:11	7:05:17	6	7:05:17	7:05:25	8	AGENT	7:05:24	7:07:09	105
0	0	PS	991109	7:06:23	7:06:33	10	0:00:00	0:00:00	0	AGENT	7:06:32	7:08:50	138
0	0	PS	991109	7:08:57	7:09:09	12	0:00:00	0:00:00	0	AGENT	7:09:07	7:13:49	282
0	0	PS	991109	7:11:16	7:11:26	10	0:00:00	0:00:00	0	AGENT	7:11:25	7:13:26	121
0	0	PS	991109	7:11:37	7:11:46	9	0:00:00	0:00:00	0	AGENT	7:11:46	7:15:34	228
29537602	1	PS	991109	7:12:31	7:12:37	6	7:12:37	7:13:31	54	AGENT	7:13:31	7:14:18	47
0	0	NW	991109	7:13:54	7:14:04	10	7:14:04	7:16:17	133	AGENT	7:16:17	7:22:25	368
0	0	PS	991109	7:16:02	7:16:14	12	0:00:00	0:00:00	0	AGENT	7:16:14	7:17:52	98
0	0	PS	991109	7:17:52	7:18:03	11	0:00:00	0:00:00	0	AGENT	7:18:03	7:18:41	38
0	0	NW	991109	7:18:01	7:18:11	10	7:18:11	7:21:00	169	HANG	0:00:00	0:00:00	0
0	0	PS	991109	7:19:41	7:19:57	16	0:00:00	0:00:00	0	AGENT	7:19:57	7:20:01	4
0	0	PS	991109	7:19:57	7:20:06	9	0:00:00	0:00:00	0	AGENT	7:20:06	7:20:57	51
0	0	PS	991109	7:21:22	7:21:34	12	0:00:00	0:00:00	0	AGENT	7:21:32	7:23:11	99
0	0	PS	991109	7:21:37	7:21:47	10	0:00:00	0:00:00	0	AGENT	7:21:47	7:22:59	72
0	0	PS	991109	7:22:58	7:23:08	10	0:00:00	0:00:00	0	AGENT	7:23:08	7:23:59	51
0	0	PS	991109	7:24:10	7:24:19	9	0:00:00	0:00:00	0	AGENT	7:24:18	7:27:04	166
55923791	2	PS	991109	7:24:43	7:24:49	6	7:24:49	7:25:50	61	AGENT	7:25:49	7:27:28	99
0	0	NW	991109	7:27:26	7:27:36	10	7:27:36	7:27:41	5	AGENT	7:27:40	7:28:35	55
0	0	PS	991109	7:27:39	7:27:49	10	0:00:00	0:00:00	0	AGENT	7:27:49	7:32:29	280
0	0	PS	991109	7:28:09	7:28:20	11	0:00:00	0:00:00	0	AGENT	7:28:19	7:28:24	5
0	0	PS	991109	7:29:00	7:29:10	10	0:00:00	0:00:00	0	AGENT	7:29:09	7:30:25	76
0	0	PS	991109	7:32:11	7:32:20	9	0:00:00	0:00:00	0	AGENT	7:32:19	7:33:33	74
0	0	PS	991109	7:35:36	7:35:46	10	0:00:00	0:00:00	0	AGENT	7:35:45	7:38:19	154
0	0	PS	991109	7:36:33	7:36:44	11	0:00:00	0:00:00	0	AGENT	7:36:43	7:38:48	125
0	0	PS	991109	7:37:32	7:37:40	8	0:00:00	0:00:00	0	AGENT	7:37:40	7:38:44	64
0	0	PS	991109	7:39:28	7:39:38	10	0:00:00	0:00:00	0	AGENT	7:39:38	7:40:32	54
0	0	PS	991109	7:41:32	7:41:41	9	0:00:00	0:00:00	0	AGENT	7:41:41	7:42:30	49
0	0	PS	991109	7:44:18	7:44:27	9	0:00:00	0:00:00	0	AGENT	7:44:26	7:49:27	301
0	0	PS	991109	7:45:28	7:45:37	9	0:00:00	0:00:00	0	AGENT	7:45:37	7:48:35	178
1489624	2	PS	991109	7:46:07	7:46:12	5	7:46:12	7:48:48	156	AGENT	7:48:48	7:50:00	72
53318796	2	PS	991109	7:46:12	7:46:17	5	7:46:17	7:47:01	44	HANG	0:00:00	0:00:00	0
16832883	1	PS	991109	7:46:37	7:46:43	6	7:46:43	7:48:24	101	HANG	0:00:00	0:00:00	0

Table Number 8

<u>Question no. 5</u>

Graph 9 on page 14 shows a model of a queues system in which a predictable transformation plays a part too important to neglect. In other words, some of the model's parameters, appearing on page 14 as well, change over time.
$Q_1(t)$ shall mark the total number of customers in the system according to the *<u>Fluid Model.</u>*

5.1. Write a differential equation for Q_1.

Differential equation:

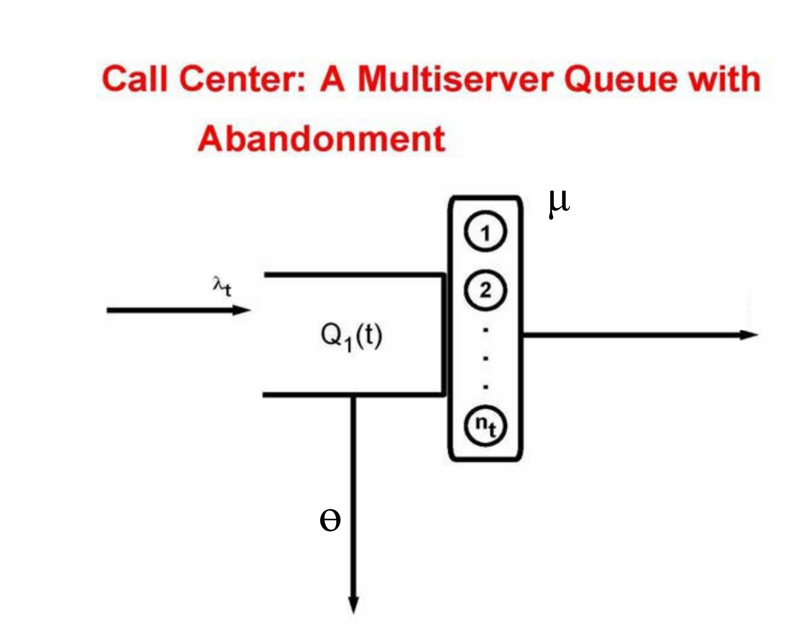
$$\frac{d}{dt}Q(t) = \lambda_t - \mu(n_t \wedge Q(t)) - \theta(Q(t) - n_t)^+$$

5.2. Explain shortly how the differential equation is solved using an electronic sheet.

Explanation:
We have to write a recursive difference equation, which is a discretezation of the real problem

$$Q(t+\Delta) = Q(t) + \Delta\lambda_t - \Delta\mu(n_t \wedge Q(t)) - \Delta\theta(Q(t) - n_t)^+$$

The recursive equation is then fed into the electronic sheet, when every line is a function of the previous line.



Graph Number 9

Primitives (Time-Varying Predictably)

λ_t exogenous arrival rate
e.g., continuously changing, sudden peak

μ service rate (of an individual server)

n_t number of servers

θ abandonment rate (for an individual customer)

<u>Question no. 6</u>

Table 10 on page 16 describes the hierarchy of activities beginning with forecast and ending with skill based routing, through staffing, shifts assignation and staff appointment. We will assume Erlang A is used to determine hourly staffing levels and accept that in any given hour 125 servers are required to staff the call center, in order to achieve a given level of service. Hence, #FTE's must be larger than 125.

6.1. Why is #FTE>125?

Answer: Service providers cannot be available for calls 100% of the time. Therefore more representatives than the number required should be appointed, in order to meet the manning goals.

6.2. What does the difference between #FTE and 125 depend on?

Answer: It depends on the breaks policy, labor agreements, etc.

Workforce Management: Hierarchical Operational View

Forecasting Customers: Statistics, Time-Series
Agents : HRM (Hire, Train; Incentives, Careers)

Staffing: Queueing Theory

Service Level, Costs

FTE's (Seats)
per unit of time

Shifts: IP, Combinatorial Optimization; LP

Union constraints, Costs

Shift structure

Rostering: Heuristics, AI (Complex)

Individual constraints

Agents Assignments

Skills-based Routing: Stochastic Control

<u>Question no. 7</u>

The graphs 11-12 in pages 18-19 display 2 density curves.

7.1. Explain in 2 sentences why the "line through dots" on graph 11 shows an advantage to size.

Explanation: The line shows an advantage to size since all along the line the amount of work per server for a time unit is constant. $\lambda / (N\mu)=Const$.
In spite of the constant load we can see that the waiting time shortens as the number of servers increases.

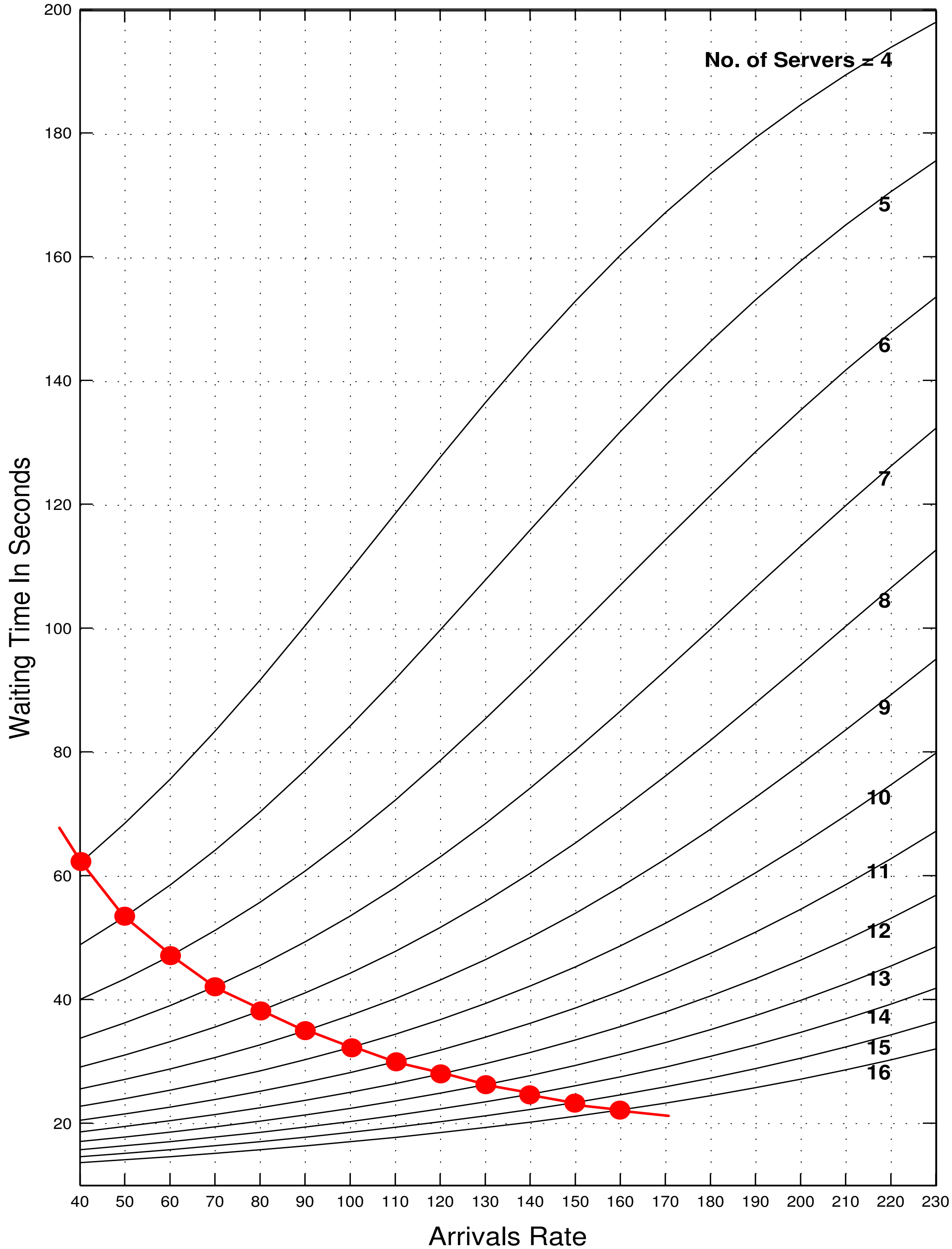
7.2. Draw an analogous "line through dots" in graph 12, namely a line that will show the advantage to size as expressed in graph 12.

7.3. In class it was said that the measure on graph 12 is not used for call centers. That said, what is its importance?

Importance: this measure gives indication to the level of service the center provides. Using it can tell whether the system works within the QED range or within either ED or QD ranges.

Economies of Scale
Average Waiting Time - But Only of Those Who Wait

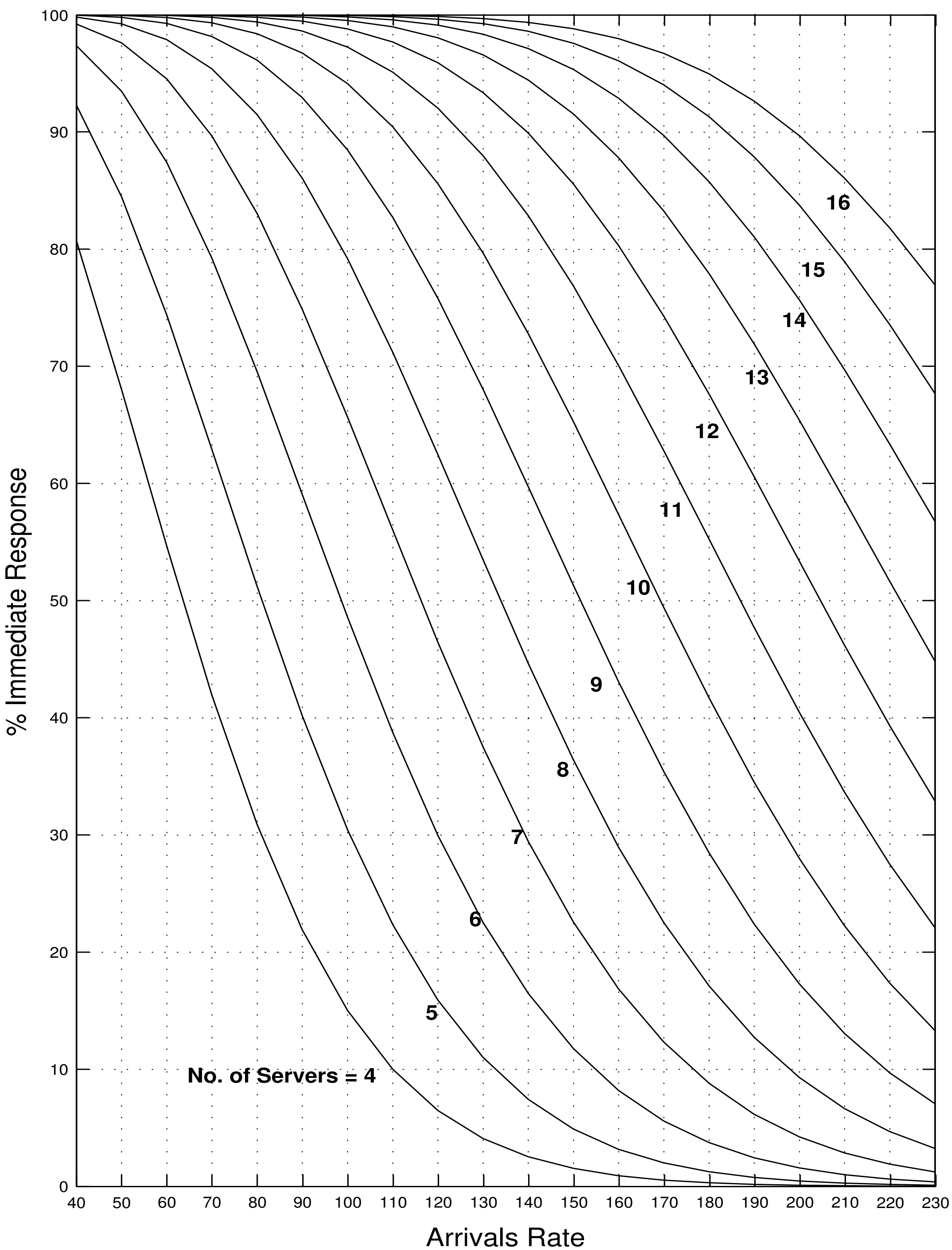
$E[W_q|W_q > 0]$ **(Load: 10 per server)**



Graph Number 11

% Immediate Response (Often Not Measured)

P(Wait = 0)



<u>Question no. 8</u>

Table 13 on page 21 displays ACD data of a call center in the USA (medium size in American standards).

Look at time intervals: 13:30-14:00, 14:30-15:00, 17:00-17:30.

Using 4CallCenter program we see that the mean patience is 3:30 minutes in all of the above intervals.

8.1. Describe a possible way, using the data in the table and the program, to receive the mean patience of 3:30 minutes.

Description: Using the Advanced Queries option it is possible to input the problem's parameters for each interval (arrivals, service time and number of servers), set in Goal the percent of abandonment, and perform a query on the mean patience.

8.2. Pay attention to the fact that the table is missing an important data which is the percent of customers forced to wait before admittance into service, namely ,

$$P\{Wait>0\}.$$

Estimate this measure roughly for each of the mentioned intervals, while using graphs 14 and 15 on pages 22-23.

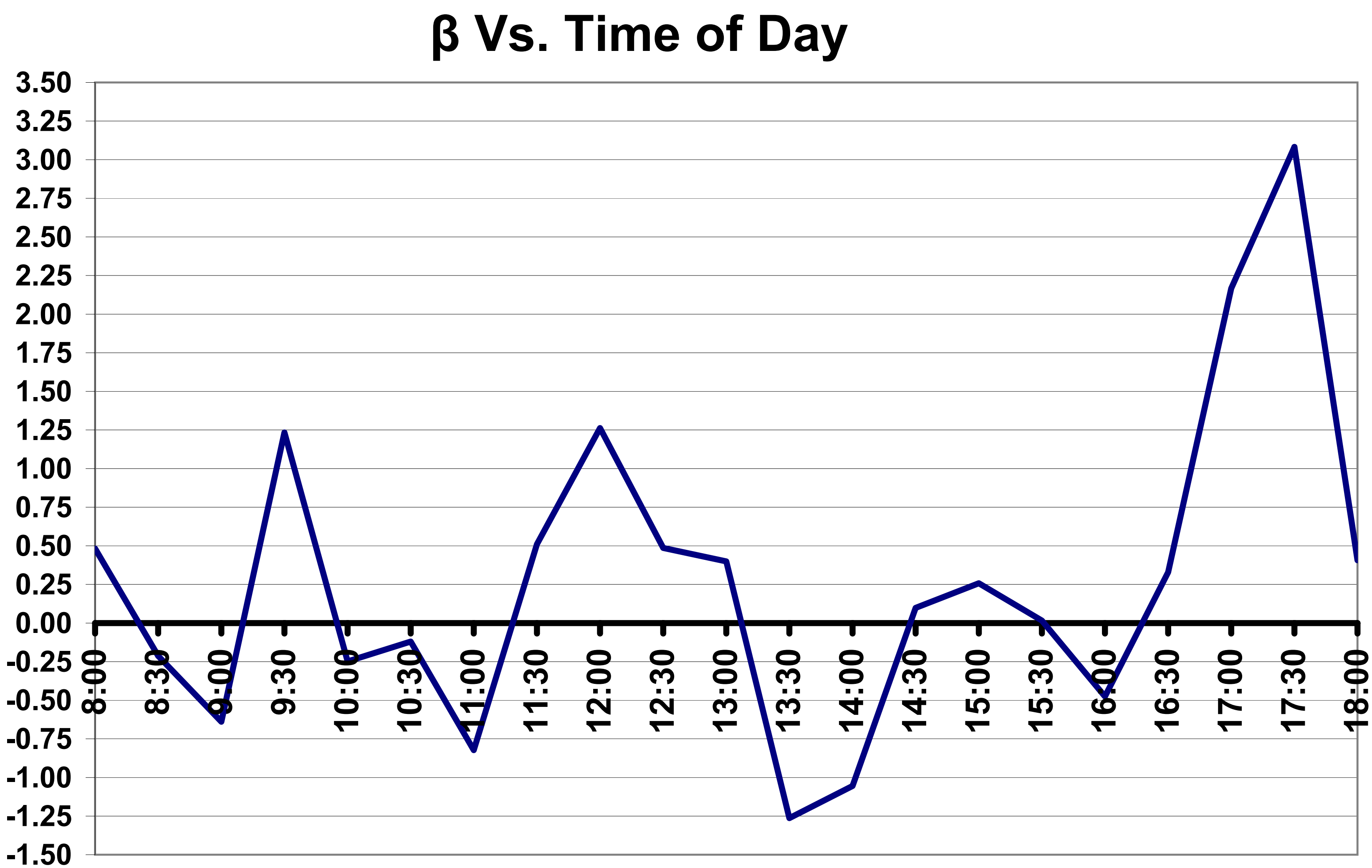
13:30-14:00 : $P\{Wait>0\}$ = <u>0.85</u>
14:30-15:00 : $P\{Wait>0\}$ = <u>0.45</u>
17:00-17:30 : $P\{Wait>0\}$ = <u>0.02</u>

Charlotte – Center

6/13/00 - Tue

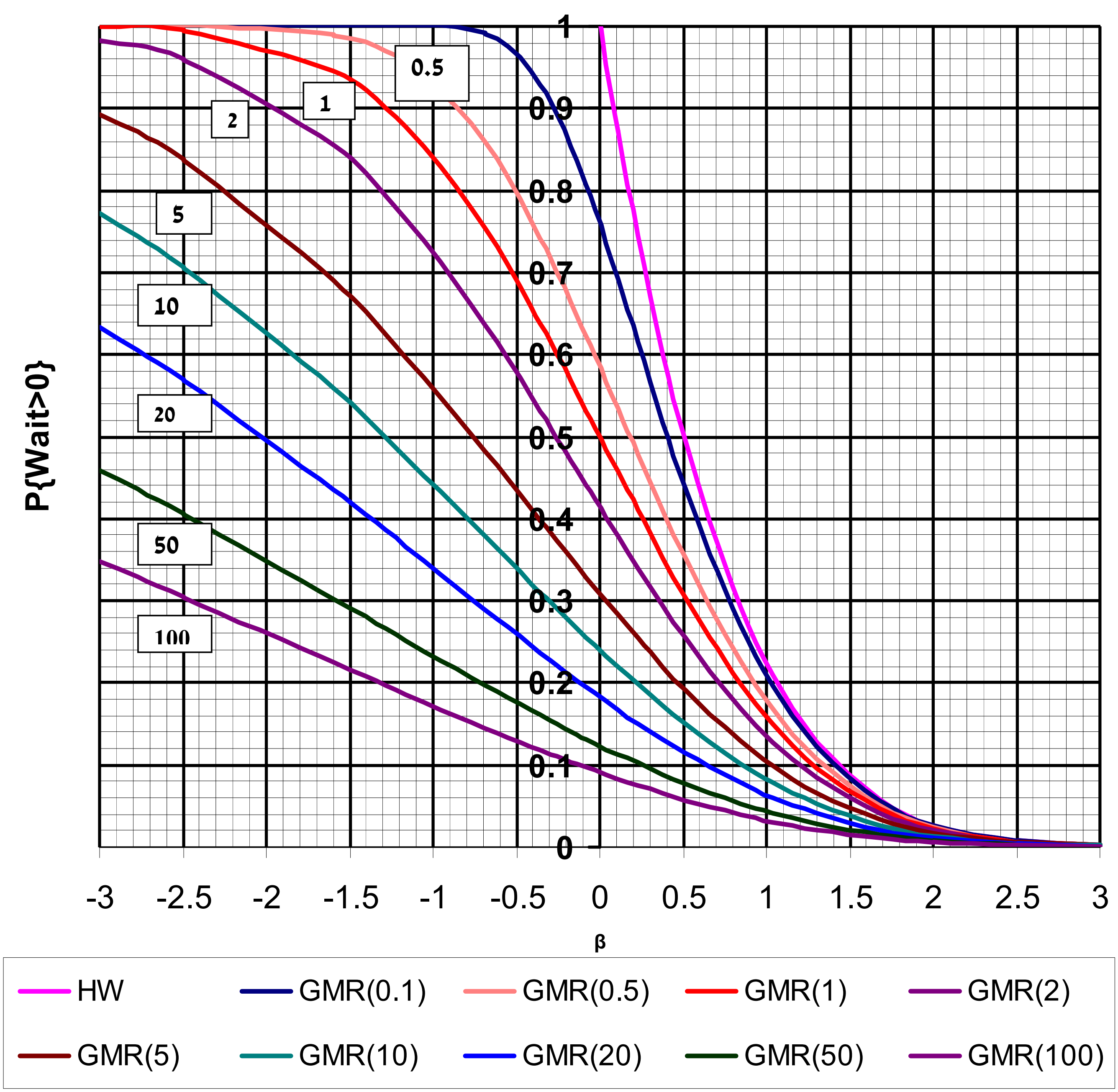
Time .1	Recvd	Answ	Abn %	ASA	AHT	Occ %	On Prod FTE	R	β= $(\boldsymbol{N}-\boldsymbol{R})/\sqrt{\boldsymbol{R}}$
0	**20,577**	**19,860**	**3.5%**	**30**	**307**	**95.1%**	**222.7**		
8:00	332	308	7.2%	27	302	87.1%	59.3	**55.70**	**0.48**
8:30	653	615	5.8%	58	293	96.1%	104.1	**106.29**	**-0.21**
9:00	866	796	8.1%	63	308	97.1%	140.4	**148.18**	**-0.64**
9:30	1,152	1,138	1.2%	218	303	90.8%	211.1	**193.92**	**1.23**
10:00	1,330	1.286	3.3%	22	307	98.4%	223.1	**226.84**	**-0.25**
10:30	1,364	1,338	1.9%	33	296	99.0%	222.5	**224.30**	**-0.12**
11:00	1,380	1,280	7.2%	34	306	98.2%	222.0	**234.60**	**-0.82**
11:30	1,272	1,247	2.0%	44	298	94.6%	218.0	**210.59**	**0.51**
12:00	1,179	1,177	0.2%	1	306	91.6%	218.3	**200.43**	**1.26**
12:30	1,174	1,160	1.2%	10	302	95.5%	203.8	**196.97**	**0.49**
13:00	1,018	999	1.9%	9	314	95.4%	182.9	**177.58**	**0.40**
13:30	1,061	961	9.4%	67	306	100.0%	163.4	**180.37**	**-1.26**
14:00	1,173	1,082	7.8%	78	313	99.5%	188.9	**203.97**	**-1.06**
14:30	1,212	1,179	2.7%	23	304	96.6%	206.1	**204.69**	**0.10**
15:00	1,137	1,122	1.3%	15	320	96.9%	205.8	**202.13**	**0.26**
15:30	1,169	1,137	2.7%	17	311	97.1%	202.2	**201.98**	**0.02**
16:00	1,107	1,059	4.3%	46	315	99.2%	187.1	**193.73**	**-0.48**
16:30	914	892	2.4%	22	307	95.2%	160.0	**155.89**	**0.33**
17:00	615	615	0.0%	2	328	83.0%	135.0	**112.07**	**2.17**
17:30	420	420	0.0%	0	328	73.8%	103.5	**76.53**	**3.08**
18:00	49	49	0.0%	14	180	84.2%	5.8	**4.90**	**0.41**

Table Number 13



Graph Number 14

Erlang-A: P{Wait>0}=α vs. β (N=R+β√R)



GMR(x) describes the asymptotic probability of delay as a function of β when $\theta/\mu = x$. Here, θ and μ are the abandonment and service rate, respectively.

Graph Number 15

Question no. 9

The following data has been collected for each judge in labor court in Haifa:

Number of cases that were being handled at the end of every month: $L_1, L_2, .., L_{12}$
Number of cases closed every month: $\lambda_1, \lambda_2, .., \lambda_{12}$

The president of the court wishes to estimate the mean time a case stays in court: for each judge separately and in total.

Write answers to these questions while using only the above given measures.

Let L_{ij} denote the number of cases that were being handled at the end of month i by judge j and λ_{ij} denote the number of cases closed in month i by judge j. Assume there are N judges. Hence

For judge j: $$\sum_{i=1}^{12} L_{ij} \Big/ \sum_{i=1}^{12} \lambda_{ij}$$

Total mean: $$\sum_{i=1}^{12} \sum_{j=1}^{N} L_{ij} \Big/ \sum_{i=1}^{12} \sum_{j=1}^{N} \lambda_{ij}$$

Question No. 10

We compare two M/G/1 systems. The arrivals into the two systems are assumed Poissonian with a rate of 20 customers an hour.

The server in system A is slow but consistent:
The mean service time is 2.1 minutes and the standard deviation is 0.1 minutes. It can be shown that the mean queue length in system A is 49/60.

The server in system B is fast but less consistent:
The mean service time is 2 minutes and the standard deviation is 1 minute.

Which system has a shorter mean queue length? Demonstrate your calculation. In which system is the mean number of people in the system higher?

For the purpose of your answer, you are supplied with the Kingman Density Law for a G/G/1 system which states:

$$\frac{1}{E(S)}W_Q \approx \begin{cases} 0 & w.p.\ 1-\rho \\ \exp[\ mean = \dfrac{1}{1-\rho}\dfrac{C_a^2 + C_S^2}{2}\ ] \quad , & otherwise \end{cases}$$

When ≈ denotes distribution approximation

The system with the shorter mean queue: A

Method of calculation:

$$E(L_Q^2) = \lambda E(W_Q^2) = \frac{\rho^2}{1-\rho}\frac{1+C_S^2}{2} = \frac{50}{60}$$

The system with a higher mean number of people: A

$$E(L^1) = \rho_1 + E(L_Q^1) = 91/60$$
$$E(L^2) = \rho_2 + E(L_Q^2) = 90/60$$

<u>Question no. 11</u>

Students in the Technion's Faculty of Industrial Engineering and Management conducted a research into schedule-accuracy of public transportation.

In a random sample of waiting periods of passengers for bus number 31 in the bus station near the faculty the students came up with a mean waiting period of 20 minutes.

The service department of buses company, Egged, reported that the mean waiting period should be 10 minutes.

Explain (in 2 sentences) the difference between the students' results and Egged's report.

Explanation: biased estimation. Egged knows that a bus leaves every 20 minutes. They assume that because people arrive randomly to the station they would have a mean waiting period of half that time. Of course, they are wrong and do not pay attention to the phenomenon of biased estimation that is realized in the students' results.

Question no. 12

We shall inspect an M/M/N/B+M system with the following parameters: the arrival rate is 150 customers a minute, the service time is 1 minute and the mean patience is 1 minute as well. The number of lines, B, is 175 and the number of servers, N, is 150.

12.1. How would an addition of a server affect the probability of blocking (the percent of customers which encounter a busy signal)?
Explain your answer as accurately as you can.

Answer: No effect.

Explanation: When $\mu = \theta$, the transfer rates in the status diagram are equal for any number of servers. Therefore, the probability of blocking (namely, the probability that the process of birth-death will be in status B) is equal for any number of servers and will not change with the addition of a server.

In all of the following subsections assume that the number of lines, B, is infinite:

12.2. For staffing of exactly 150 servers (and an infinite number of lines) server utilization equals 96.7%. What is the percent of abandonment? Demonstrate your calculation.

Percent of abandonment: 3.3%

Calculation: $\rho = \dfrac{\lambda_{eff}}{N\mu} = \dfrac{\lambda(1-P_{ab})}{N\mu}, \quad \Rightarrow \quad P_{ab} = 1 - \dfrac{\rho N\mu}{\lambda} = 1 - 0.967$

12.3. By using 4CallCenters we see that in order to assure waiting customers percentage of approximately 25%, the system should be staffed with 159 servers. Under such staffing level the percent of abandonment is 1.1.

Assuming that the arrival rate is multiplied by 9, what is the staffing level that assures the same percent of waiting customers? Calculate roughly the percent of abandonment in this case.

Staffing level: N=150*9+9*3=1377

Abandonment rate: $\%_{ab}$=1.1/3=0.3666%

12.4. What, in one word, is the distribution of the number of customers in the system.

Deviation: Poisson.

12.5. Assume the arrival rate is 150, the mean service time is 1 minute, the mean patience is 1 minute, but the number of servers is now 120.

What is, roughly, the probability to wait? Explain shortly.
What is, roughly, the percent of abandonment?

Probability to wait: 1

Explanation: The system works under heavy load (in the ED range) *(β=-2.7386)* and therefore the probability to wait is approximately 1. (There is no need to calculate β in order to reach this conclusion).

Probability of abandonment: $P_{ab} \cong 0.2$

$$P_{ab} \cong 1 - \frac{1}{1 \vee \rho}, \qquad where \qquad \rho = \frac{\lambda}{N\mu}$$