Homework No.3. Empirical Models.

Description of data.

File Empdata.xls contains the data of this homework Specifically, it includes the "Summit" service type transactions of 3 days in 2001: May 25th, May 29th and May 30th, in worksheets "day2505", "day2905" and "day3005" respectively*.
In addition, for convenience, we have added the green column:

Sys_exit = time of exit from the call center of each customer (either exit from service for served customers or exit from queue for abandoned customers).

We also added a blue table consisting of the following 7 columns:

- **Time** = time epochs covering the period between 6:00 and 23:00 with 0:01 (1 minute) intervals.
- **Arr_sys** = cumulative number of arrivals to the call center until (not including) time specified in the column "**Time**", (referred to as Time in what follows).
- **Dep_sys** = cumulative number of departures from the call center until Time.
- **Arr_q**= cumulative number of arrivals to the queue until Time. (Note that Arr_q = Arr_sys)
- **Dep_q** = cumulative number of departures from the queue until Time.
- **Arr_ser** = cumulative number of arrivals to service until Time.
- **Dep_ser** = cumulative number of departures from service until Time.

Cleaning and correcting data.

We erased the column "vru_line", being irrelevant to the present homework.
We erased the following records with:

1. sys_entry between 23:00:00 and 6:00:00.
2. ser_start = ser_end & q_start = q_end.
3. outcome = Agent if the call was handled, and outcome=Abandon if the call was not handled (See Appendix).

Hold Time

The hold time is a part of the service time. Specifically, the agent places the service on hold in order to ask a supervisor a question about the service, while the customer is waiting on line.

Remarks: Using Excel's Pivot Table option in questions 6, 7, 8 and 12 is highly recommended.
Questions 10 and 11 have partial solutions on the course website.

* May 25th was a Friday, May 29th a Tuesday and 30th a Wednesday.

Questions.

Fluid model.

1. Consider customers that arrived to the queue on May 25th between 6:27 and 07:02. Calculate the average waiting time at this time interval. Now, assume "First Come – First Served" queue discipline. Using the fluid approach (i.e., using the functions Aq(t) and Dq(t) that are based on the real data, as in the recitation), calculate the average waiting time. Compare your results and comment on your findings.

2. Assume "First Come – First Served" queue discipline on May 30th. Assume that a "virtual customer" arrived to queue at 7:09. Using the fluid approach, find the customer **sojourn time** in the system (total time in the system), using the suitable cumulative arrivals and cumulative departures functions. Now take sojourn times of several "real customers" that arrived to the system around 7:09 (± 1 minute) and compare them with your answer. Can the standard fluid model be used in order to estimate the sojourn time of a specific customer? Explain your conclusions.

Empirical models: queues, staffing, arrivals, service times.

3. Draw the number of customers in the call center, as a function of time, for the three days under consideration. (Use 1-minute resolution.) Describe briefly the pattern of those graphs, characterizing common and different features of the three days. (It is helpful to also draw the 3 days on the same graph, for an easy comparison.)

4. Draw the number of customers in queue and the number of customers in service on May 25th. (Use 1-minute resolution.) Describe the graphs briefly. Can you identify any time moments with positive queue and no customers in service?

5. Our data does not include the staffing level (number of working servers) as it evolves in time. Find a way to approximate the staffing level on the three days. Compare staffing levels across days and times of day.
   Remark: If you can't run your algorithm automatically with the Excel's tools, explain in words the solution you propose, but note that you have to provide at least one algorithm to solve the problem. (A bonus may be awarded if you propose more than one solution.)
   Hint. You may assume that staffing levels remain constant at least over 1-hour periods.

6. Draw graphs of the arrival rates to the call center per hour for the three days. Compare the graphs.

7. Calculate the average service times per hour for May 29th. Explain shortly how you decide which calls to take into consideration. How would you calculate the daily average service time from the average service times per hour?

Waiting and abandonment.

8. Calculate average waiting time (E [*Wait*]) for the three days. Similarly, calculate average waiting times of the customers that had to wait for service (E [*Wait* / *Wait>0*]). Estimate daily probabilities that a customer has to wait in queue (P{*Wait>0*}). Deduce a relation between the three performance measures calculated above. Comment on your findings.

9. Consider two days, "busy" and "not busy". Denote by W_1 and W_2 their average waiting times respectively and by W_1^+ and W_2^+ the corresponding average waiting times of the customers that had to wait. What relation between $\frac{W_1}{W_2}$ and $\frac{W_1^+}{W_2^+}$ would you expect? Why?

10. Draw the graphs of average waiting times per hour for the three days. Describe their patterns briefly.

11. For each day calculate probabilities of abandonment, overall and per hour. (Draw the graphs per hour.) Describe your findings.

12. For May 25th, check if the relation between E[Wait] and P{Abandon} is linear. What is the value of the slope (in seconds)? (We shall study later its meaning.)

Summary questions.

13. Describe the relations between arrival rate, waiting times, service times and probabilities of abandonment (e.g. when is waiting long?).

14. **Little's Law.** For May 25th calculate average number in queue and average number in service per hour. Check Little's Law for system = queue and system = service, per every hour and for the day overall. Describe your findings and explain reasons for discrepancies, if any.

15. **Capacity Analysis.** Consider May 25th. For hours number i, i=1,...,17, calculate
Offered load per server = (arrival rate * average service time) / (number of servers) and
Utilization = (arrival rate * P {customer was served} * average service time) / (number of servers).
(Use Question 5 to infer the number of servers.**)** Have you found any values greater than one? What would be the meaning of values greater than one for offered load per server and utilization?

16. We consider four possible reasons for the difference when compared across the 3 days, between the queues and customer waiting:
    - Variability of arrival rate: more arrivals on busy days.
    - Variability of service times: service times are longer on busy days.
    - Variability in the number of servers: fewer servers on busy days.
    - Inappropriate staffing: number of servers is not matched well with arrival load.

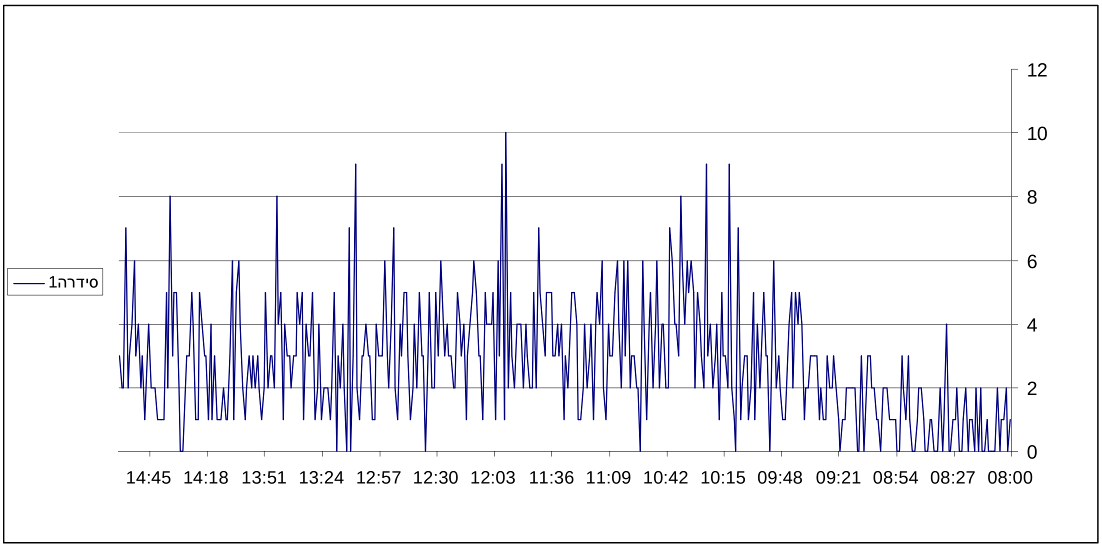
    Explain what causes the difference between the three days in consideration? (You are not restricted to the reasons mentioned above; that is, feel free to add reasons as you see fit.)

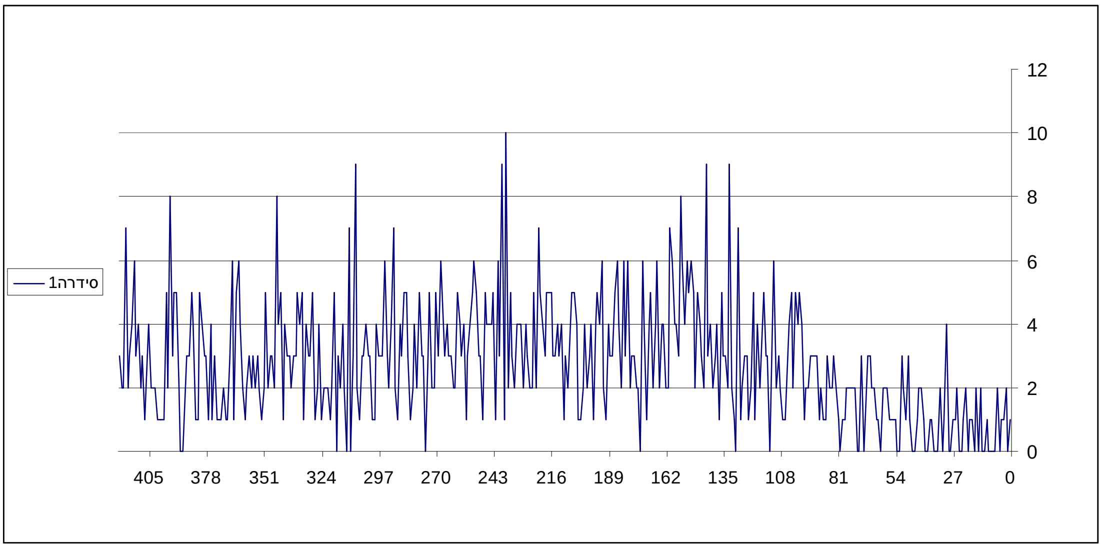
Graphs in Home-works

Once the graph has been created, take a minute or two to put the finishing touches to it.

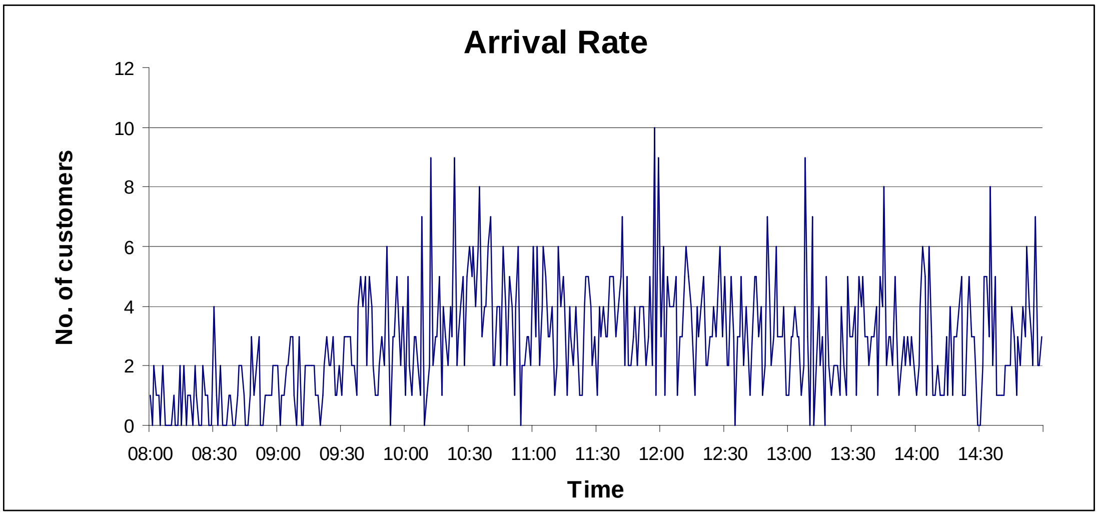
1. Make sure that the graph and the axes have titles. You should always place units under your axis heading. For instance, cm, m, km for distances, and minutes, hours for time, etc.
2. While creating a graph of Y as a function of time you should use the times during the day and do not make any transformation of it, which might reduce information. For instance, presenting the number of minutes from the beginning of the day makes it difficult to understand when the peaks, lows, etc. occur.
3. If the scaling is shown improperly, for example 07:11, 07:41, 08:11 and so on, open "Format axis" by double clicking on the X-axis, then choose the "Scale" tab and modify the minimum, maximum and the major unit values. For instance, for data of arrival rate during every minute between 07:00 and 15:00, if you choose the first category as minimum, the last one as maximum and major unit as 30 the scaling you will get is 07:00, 07:30, … , 15:00.
4. Present the Y-axis in the left side of the X-axis: go to Format axis (X) -> Scale and remove the v-mark from the "Values in reverse order".
5. With one set of data on the graph, the legend box is not useful. You should delete it.
6. Choose dark colors for your graphs so it will be possible to see it on a black and white printing.

Example of "bad" graphs:





Example of a "good" graph:



Technical Appendix - Call Outcome

(not relevant for the present analysis)

In our original data file, the call outcome is indicated by a code, which define the call termination. The causes of call termination are Handled/ Transferred/ Abandoned/ Undefined. In normal situations, the calls terminates when a caller hangs up after receiving the service (code 1), or when an agent hangs up after providing service (code 2). The abandon short termination (code 11) occurs when a caller or agent abandons, within an abandon threshold time, without reconnecting to a Call Center resource, or a caller hangs up during delay, queue, or ring time. The abandon termination (code 12) occurs in the same situations as described before, but after the abandon threshold time.
Note: in this homework, we assume that the "Other Unhandled" outcome is a third type of abandon and that "Agent Transfer" is another type of after service outcome.

Outcome field

Code	Name
1	Caller Termination
2	Agent Termination
3	Undetermined Termination
11	Abandoned
12	Abandoned Short
13	Other Unhandled
22	Agent Transfer (to other call center)

In the file Empdata.xls, we classify the calls into two classes:

1. Agent - calls that ended after receiving a service by an agent.
2. Abandon – calls that ended by abandonment.