Daily MRI Prediction Model

Shabnam Hajian

2019-02-11

Case summary and assumptions:

MRI services have seven different types based on the body part that needs to be scanned: Head (Brain), Head & Neck, Spine, Cardiac, Abdomen, Pelvis, and Extremities. In Ontario, patients for MRI are prioritized based on their acuity level and each level has a specific Waiting time target. These data are available as below table:

Priority level	Description	Waiting time target
1	Emergent	24 Hours
2	Inpatient or urgent	48 Hours
3	Semi-urgent	10 Days
4	Non-urgent	48 Days

*These targets have been set at the 90% significate level.

The hospital faced difficulty in meeting the waiting time targets.
Assumptions from the case:

1- Finding a Daily MRI demand is the first step to help the hospital to meet its target.
2- The arrivals on the same day in the previous week could be used to predict the arrivals of patients of each priority level the following week. (X_{T-7})
3- There is a relationship between the month of the year and the MRI demands (overall increase for MRI services during the warmer months from graph 1).
4- The seven-day forecast can accurately predict the weather at a 90% significate level (A five-day forecast can accurately predict the weather as high as 97 percent of the time).

Note: To deal with this data, we assume data is cleaned and there is no need for any further cleaning process.

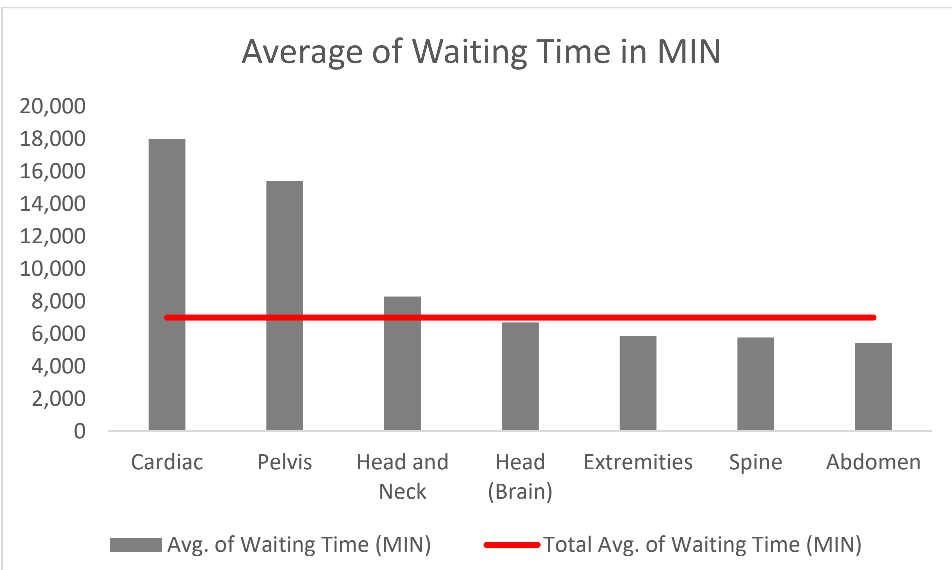
Exploring data:

Here we have data of MRI patients from targeted hospital A in two priority levels (1 and 2) from the year 2015 and 2016 and the related weather for the A hospital location in 2015.

Q1-a- For this question, we use time in Minutes and the average is based on minutes. To have a better understanding of the below table we need to consider that **every 24 hours is 1,440 minute**s and **every 48 hours is 2,880 minutes** which is the targeted waiting time for priorities 1 and 2. The average times for all service types are above 2,880 MIN, which means it is more than targets.

Service type	Average of Waiting Time in MIN
Cardiac	18,004
Pelvis	15,403
Head and Neck	8,290
Head (Brain)	6,703
Extremities	5,875
Spine	5,774
Abdomen	5,440
Grand Total	**6,997**

Q1-b- To create this graph, we could use the last question's table. In the below table the red line shows the total average and the bar charts are illustrate the average waiting times per service type. It could result that the average waiting time for "Cardiac", "Pelvis", and "Head" types of MRI are above the total average of waiting time.



Q1-c- To compare differences in the average waiting time between different types of MRI, we use a one-way ANOVA test with the following hypothesis:

H_0: There is no difference in the average waiting time between different MRI types

H_1: At least one of the average waiting times differs from the others

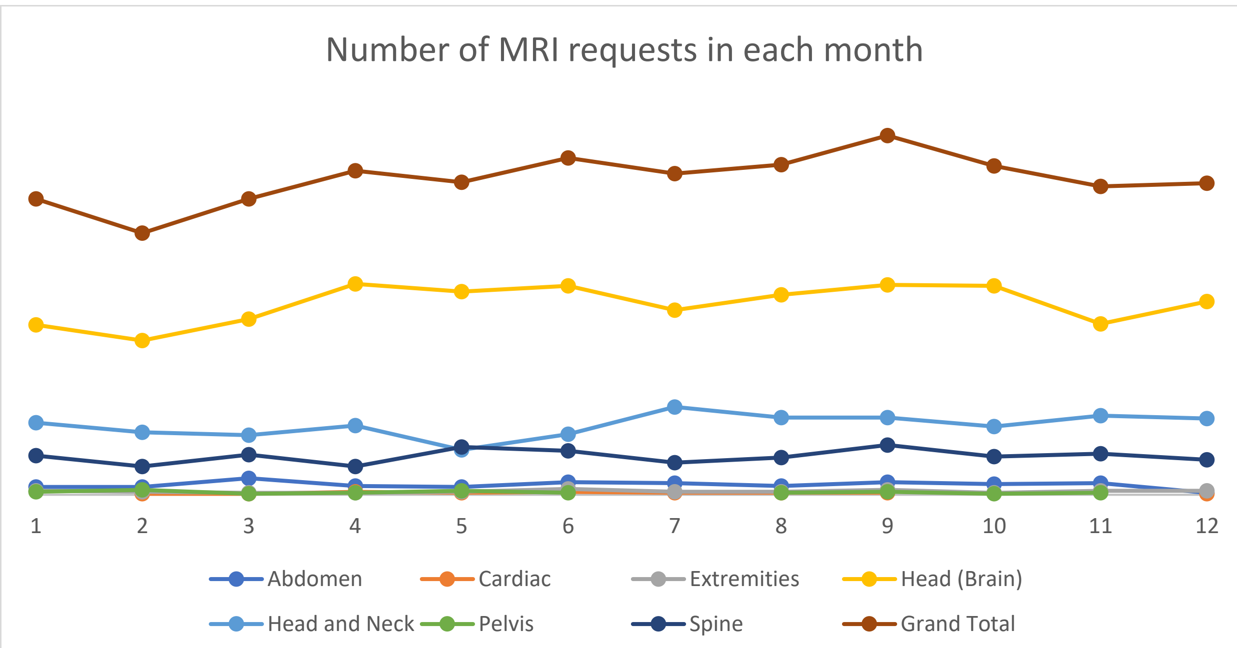
To use this test, we need to accept the following assumptions:

1- Normality – each sample is taken from a normally distributed population (large enough sample sizes)
2- Samples are independent –each sample has been drawn independently of the other samples

3- Variance Equality – The variance of data in the different groups should be the same
4- The dependent variable is continuous (Waiting time in minutes)

By using ANOVA from the "Data Analysis" and using **alpha=0.1** from excel we would have **p-value=0.028.** Since **p-value<alpha** we could reject the null hypothesis with a 90% significance level and conclude that "the average service waiting time in minutes is different between different types of MRI services".

Q1-d- To calculate the monthly average waiting time we need to have the number of days in each month. By using the sum of monthly waiting time and the number of days in each month we would have the average monthly waiting time in minutes for different service types.
To compare the average number of monthly requests we could use the number of patients and request month information and draw the below graph which shows the number of requests in each month:

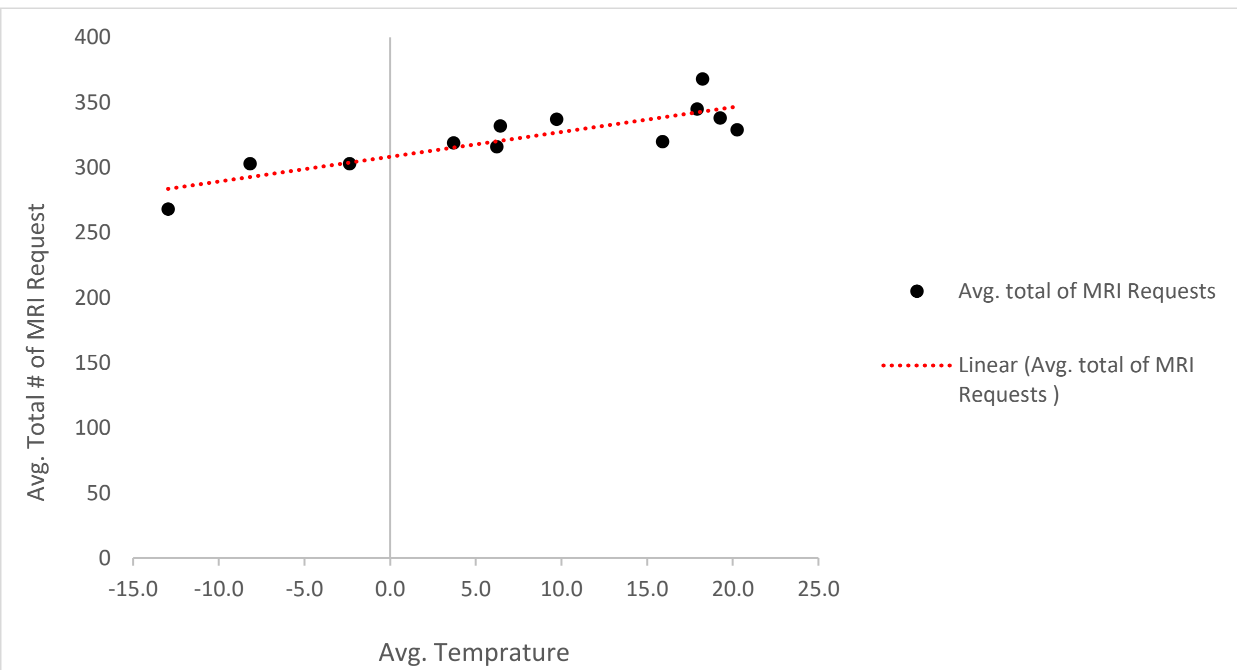


This graph illustrates that there is a relationship between the month of the year and the number of MRI requests. This relationship has a different pattern for different service types.

Q1-e- The calculate the 95% confidence interval for the average temperature of each month we need to calculate "average", "standard deviation" and the number of data that we have as a sample in the climate datasheet and by considering **$Z_{0.05}$ = 1.96** and following formula, calculate the confidence interval.
C.I = [Mean ± Z_α $S/\sqrt{n}$]
The below table shows the average temperature of each month and the lower bound and the upper bound of each month's average temperature by 95% confidence level.

Month	Average temperature	C.I. Lower bound	C.I. Upper bound
Jan.	-8.2	-8.6	-7.8
Feb	-13.0	-13.4	-12.5
Mar	-2.4	-2.8	-2.0
Apr	6.4	6.0	6.8
May	15.9	15.4	16.4
Jun	17.9	17.6	18.2
Jul	20.2	19.9	20.6
Aug	19.3	19.0	19.6
Sep	18.2	17.9	18.6
Oct	9.7	9.4	10.1
Nov	6.2	5.8	6.7
Dec	3.7	3.4	4.0

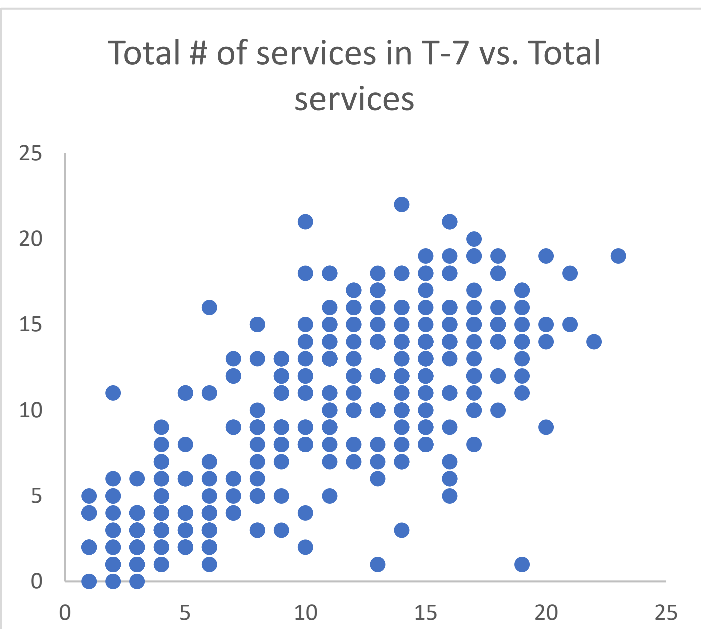
Q1-f- To have a picture of the relationship between the average temperature and the average of total requests that the hospital received, we could use the below graph.

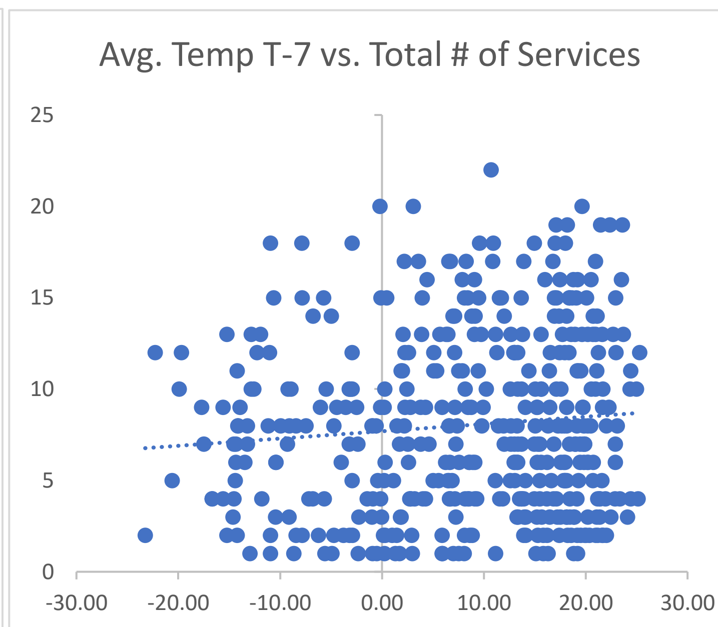


The graph shows that there is a relationship between the average temperature of the location of the hospital and the average of total requests that the hospital received. Also, by using regression tool from the analysis data package we could have the estimated linear relationship as follow:

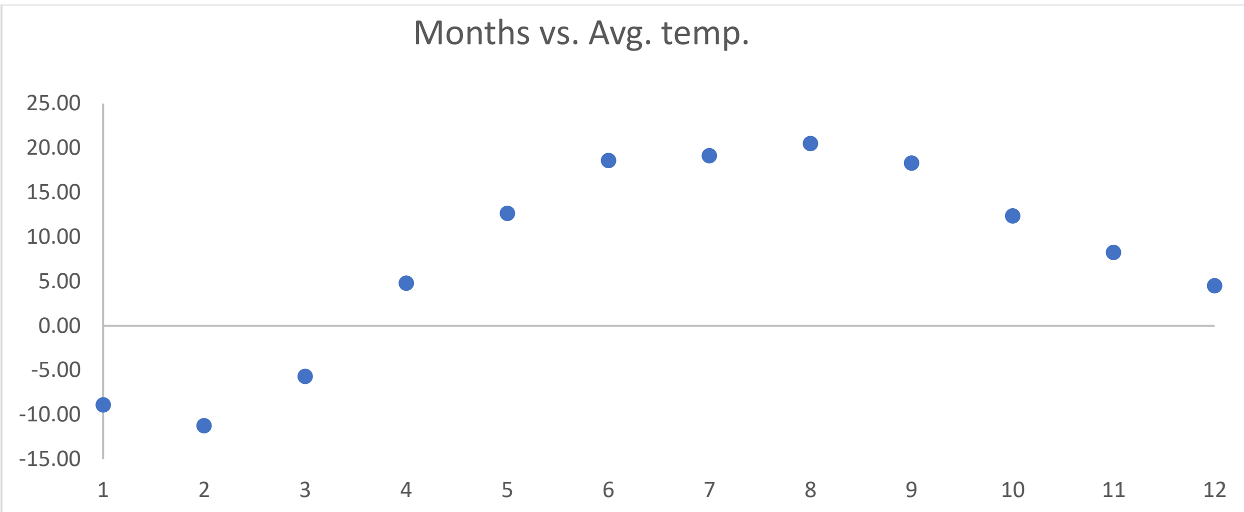
Y = 308.2 + 1.9 **X**

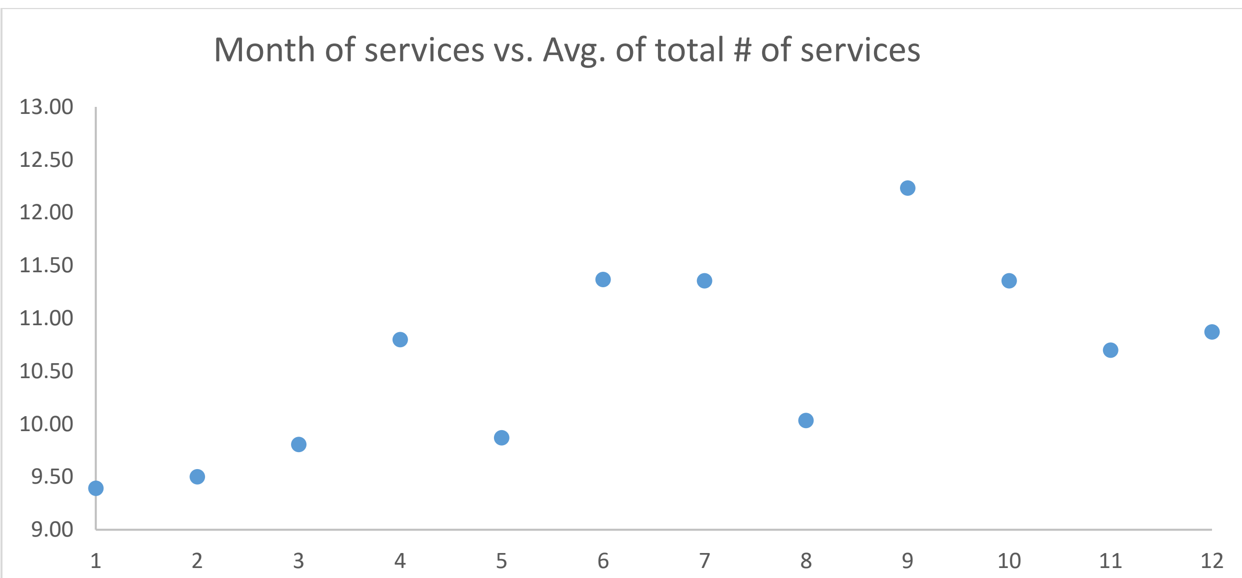
Q2-a- To start using linear modelling in this question, we need to make some changes in our data. First, we should create a new variable as the total number of services seven days before (named as total # of services in T-7). Also, to investigate the effect of average temperature, we should create a summary table from the climate data sheet that shows the average daily temperature and use this information to create two new variables as the “Avg. Temp“ and the “Avg. Temp T-7”. Next, we need to mix these two data sets. In this step, we need to clean data based on the available information. We know that there is no T-7 data for the first 7 days, also the climate data is only for the year 2015 and we don’t have any information about the climate in 2016. Hence, we clear data of the first days and days in the year 2016. As a result, we would have 3489 observations as a sample. (Sheet “Q2-a-BaseData”)
Based on the information in this case, we know that the total services in each day as an outcome variable has a relationship with two independent variables as the temperature and the number of services in seven days before that date. Also, we know that the temperature of seven days before could be used to predict each day’s temperature. (The first table in relationship sheet “Q2-a-Relationships” shows this relationship)
To continue with this data, first we should find out more about the relationships between these data. As an initial step, we need to make sure that “Avg. Temp T-7” and “Total # of services in T-7” are good tools to predict the total number of services of each day. To do so, we graph the data for “Avg. Temp T-7” with “total # of services” and “Total # of services in T-7” with “total # of services” (sheet “Q2-a-Relationships”). These graphs would be as below:





From the below graph, we could easily interpret the linear relationship between “Total # of services in T-7” and “total # of services” but it is not easy to have this conclusion about the “Avg. Temp T-7” from this graph. To have a better understanding of the data we need to look more wisely at the relationship between each month and the average temperature. The below plot shows the relationship between “service month” and the “avg. temp.”, and we could easily see that there is a relationship between average temperature and the month of the year. Moreover, by plotting the relationship between the total number of services in different months, we could conclude that there is a relationship between different months and the total number of services (sheet “Q2-a-Relationships”). As a result of these two plots, generally, we could conclude that the total number of services is more in warmer (higher temperature) months(this relationship was mentioned in the case details too). Thus, although we know that a linear relationship between average temperature and the total number of services is not the best description for this relationship, for simplicity we assume that this relationship is linear and use it in our model.





The above plot between the month of service and the total number of services also shows the differences between the average of total services in different months. As an instance, comparing average temperature in months 7 and 8 and the average of total services in these two months. This means that to have a more accurate model, we should consider different monthly averages in our model.

On the other hand, since we confirm "Total # of services in T-7" and "Avg. Temp T-7" as two predictor variables, we need to make sure about the independency between these two variables; to do so, we use regression and the small amount of regression makes us sure to be able to have a linear regression model.

	Avg. Temp T-7	*# of services in T-7*
Avg. Temp T-7	1	
# of services in T-7	0.063266	1

To model, as we said before we separate each month and create a linear model for each month separately. The final model will be a total of 12 models, each model for each month of the year. This kind of model helps us to see the amount of error for each month (Q2-a-Models Data). The final intercepts and coefficients would be as below table:

Month of services	# of Obs.	Intercept	Coefficient of Avg. Temp. in T-7	Coefficient of # of services in T-7
Jan	27	2.16	0.04	0.82
Feb	33	1.90	-0.04	0.66
Mar	37	0.95	-0.04	0.80
Apr	37	2.88	0.02	0.67
May	33	0.17	0.10	0.82
Jun	41	-1.55	0.18	0.74
Jul	39	1.06	0.01	0.89
Aug	33	8.52	-0.35	0.88
Sep	40	-1.17	0.16	0.79
Oct	43	1.84	0.01	0.76
Nov	38	2.88	-0.06	0.70
Dec	34	2.08	0.27	0.62
Overall	435	1.84	0.01	0.76

Important notes:

1- According to our first plots, we knew that the relationship between average temperature and the total number of services is not clear. From the results, the 95% confidence interval for the coefficient of average temperature, contains zero which means that this predictor variable could not be considered in these models.

2- By having more data, for example, data of two or more years, we would have a better picture of the relationship between average temperature and the total amount of services.

3- By having more data, we could do more investigation about different MRI types and involve them in our model to make it more accurate. It seems the Head (Brain), Head and Neck, and Spine are more sensitive about the month of the year and these sensitivities should be considered in the model.