Prediction of hospital readmissions for patients with diabetes

Neeraj Kumar Reddy Panta

Ruchi Dilip Kukde

Outline

- Exploratory Analysis – Choice of Variable
- Graphical Description
  - Histogram
  - Box Plot
- Numerical Summary
  - Center
  - Spread
  - Shape of Distribution
- Findings and Implications

Exploratory Analysis

- First and critical step in data analysis process
- Performing initial investigations on data
  - To discover patterns
  - To spot anomalies
  - To test hypothesis
  - To check assumptions
- Description of distributions
  - summary statistics
  - graphical representations

Variable of Significance

- **Time in Hospital**
- Also called Length of Stay (LOS)
- It is a clinical metric that measures the length of time elapsed between a patient's hospital admittance and discharge

	A	B	C	D	E	F	G	H	I	J
1	encounter	patient_nl	race	gender	age	weight	admission	discharge_	admission	time_in_hospital
2	2278392	8222157	Caucasian	Female	[0-10)	?	6	25	1	1
3	149190	55629189	Caucasian	Female	[10-20)	?	1	1	7	3
4	64410	86047875	AfricanAm	Female	[20-30)	?	1	1	7	2
5	500364	82442376	Caucasian	Male	[30-40)	?	1	1	7	2
6	16680	42519267	Caucasian	Male	[40-50)	?	1	1	7	1
7	35754	82637451	Caucasian	Male	[50-60)	?	2	1	2	3
8	55842	84259809	Caucasian	Male	[60-70)	?	3	1	2	4
9	63768	1.15E+08	Caucasian	Male	[70-80)	?	1	1	7	5
10	12522	48330783	Caucasian	Female	[80-90)	?	2	1	4	13
11	15738	63555939	Caucasian	Female	[90-100)	?	3	3	4	12
12	28236	89869032	AfricanAm	Female	[40-50)	?	1	1	7	9
13	36900	77391171	AfricanAm	Male	[60-70)	?	2	1	4	7
14	40926	85504905	Caucasian	Female	[40-50)	?	1	3	7	7
15	42570	77586282	Caucasian	Male	[80-90)	?	1	6	7	10

Significance

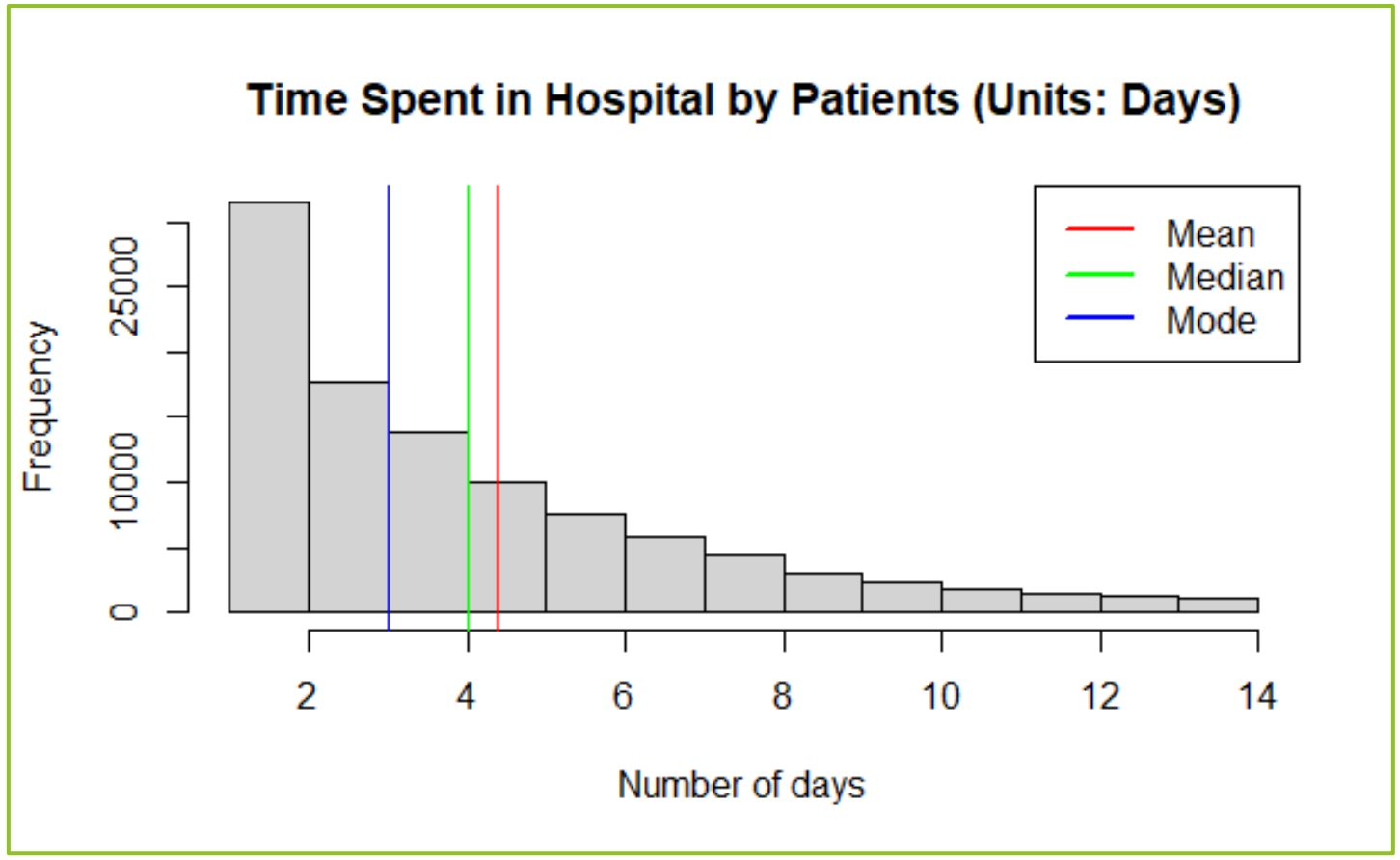
- The national average for a hospital stay is **4.5 days**, according to the Agency for Healthcare Research and Quality, at an average cost of $10,400 per day.

- Important indicator of efficiency of hospital management, patient quality of care, functional evaluation
- Shorter hospital stays reduce the burden of medical fees, increase the bed turnover rate
- This in turn increases the profit margin of hospitals, while lowering the overall social costs
- Important for further analysis: determination of impact of length of stay on readmission risk

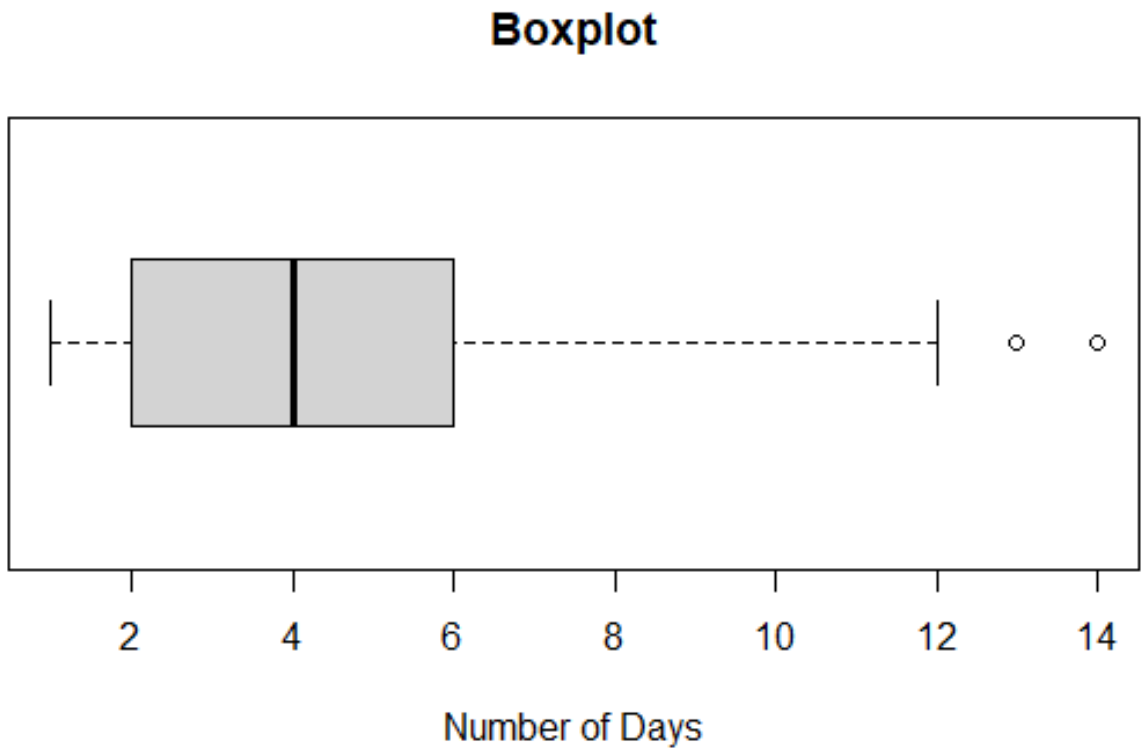
Variable: Time in Hospital

- Data Type: Numeric (specifically Integer)
- Units: days
- Observation: no null values exist for this variable
- Exploratory analysis is carried out on 101766 observations

Graphical Description: Histogram



Graphical Description: Box Plot



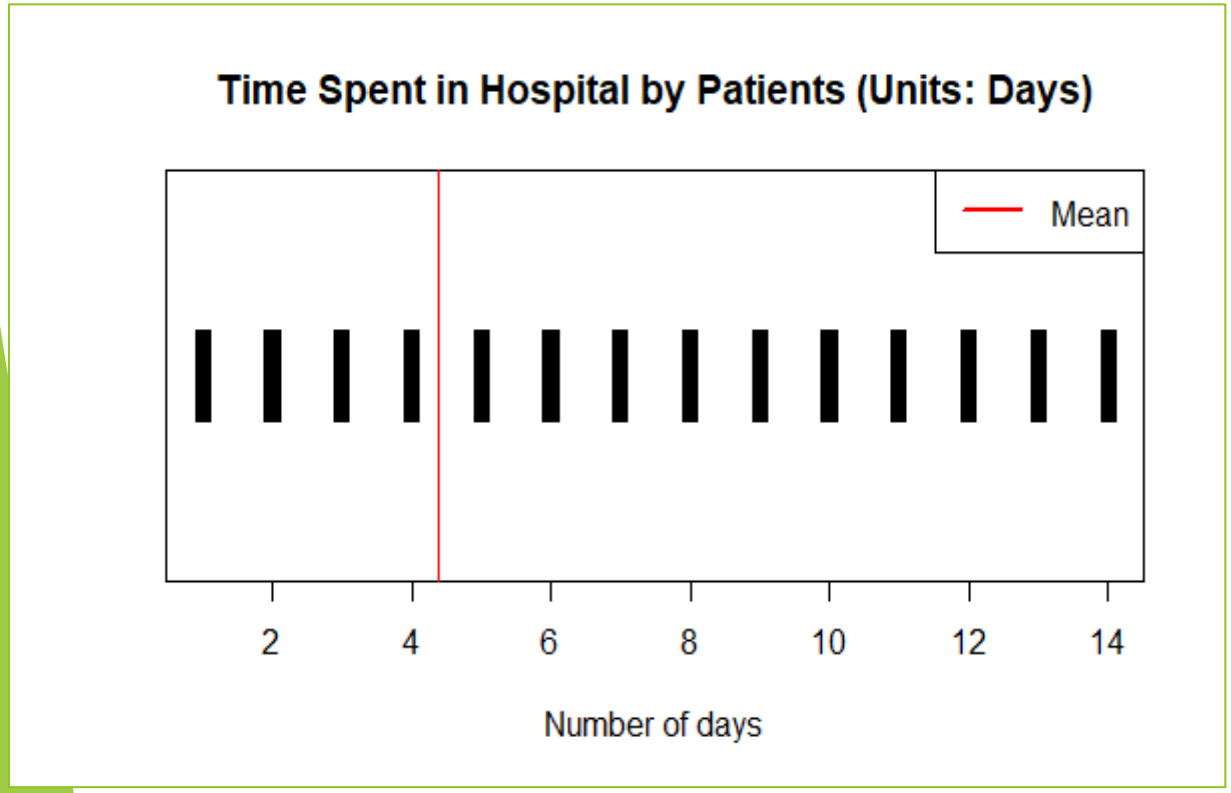
Min	Quartile 1	Median	Quartile 3	Max
1	2	4	6	14

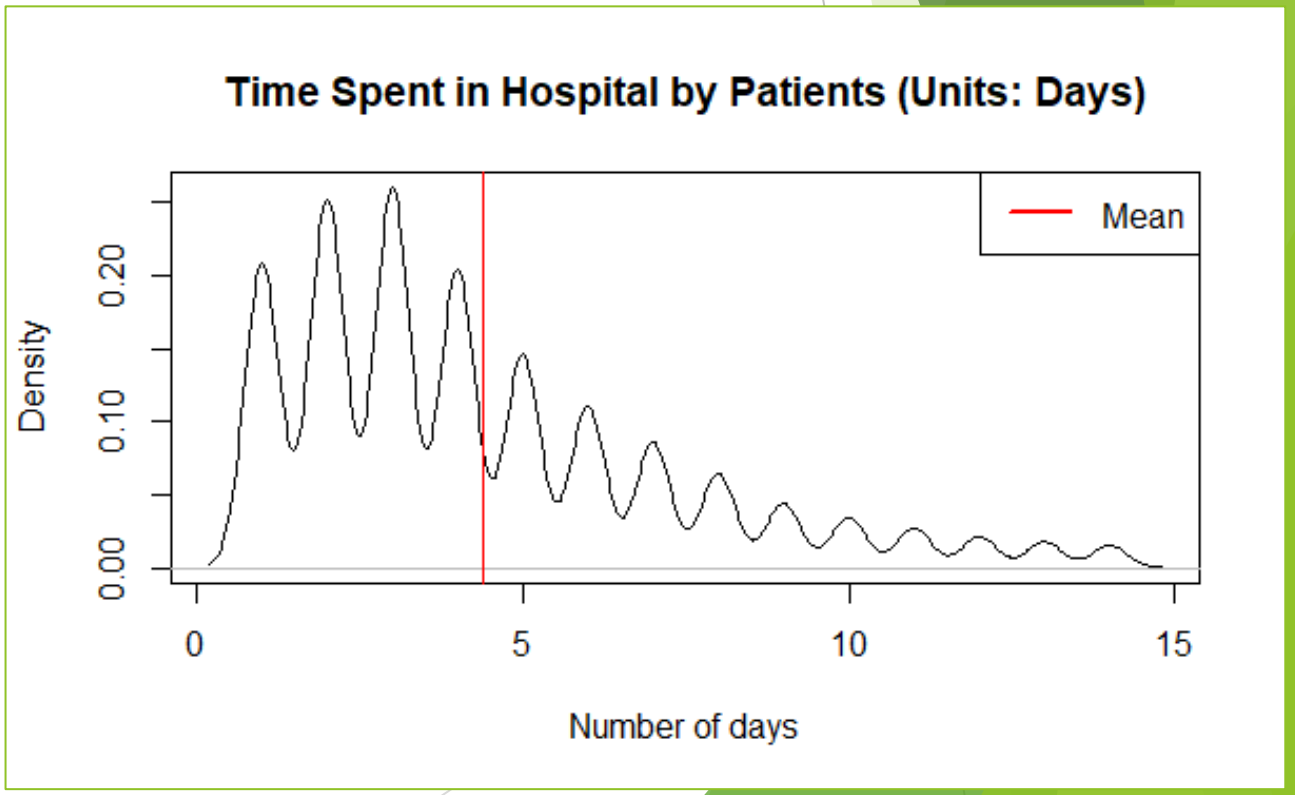
Numerical summary

Measures of Location	
Mean	4.395987
Median	4
Mode	3
Measures of spread	
Range	13
IQR	4
Variance	8.910868
Standard Deviation	2.985108

Shape of distribution

- Positively skewed distribution
- Mode < Median < Mean





Findings and Implications

- Mean of time in hospital (4.395 days) obtained for this dataset is close to the national average of 4.5 days
- Descriptive statistics indicate that time in hospital is a positively skewed distribution
- 50% of the patients spent 2-6 days in the hospital
- 2252 patients who stayed in hospital for 13 days and 14 days are identified as outliers

Findings and Implications

- Time in hospital can be analysed in association with other variables in the dataset from different perspectives
- It would be interesting to determine correlation of length of stay with readmissions and severity and type of diagnosis variables in the dataset
  - whether patients with a lower length of stays were readmitted more frequently to the hospital?
  - what were the diseases for such patients?
  - how many medications were administered?
  - were any lab procedures carried out?

Thank you