

# **Vanderbilt University Medical Center: Elective Surgery Schedule**

Rashi Desai

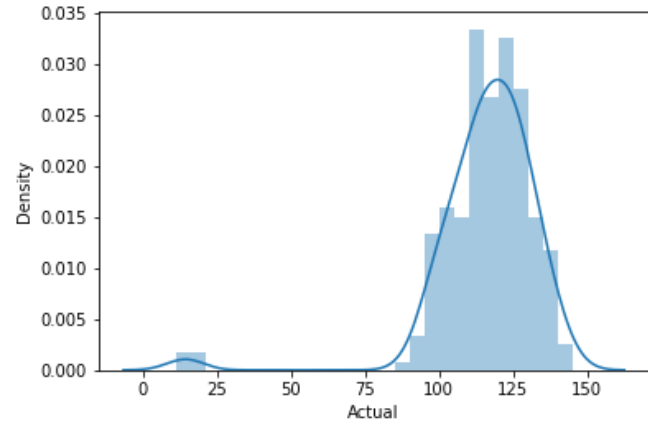
Vivek Kumar

# Exploratory Data Analysis

- The given data set is a 48-week surgery schedule for Vanderbilt Medical Center
- The data set has a schedule of surgery 28, 21, 14, 13....1 day prior to the actual surgery date and the count of actual surgeries on that day
- Data has been captured for weekdays from Oct 10, 2011 to Sept. 14, 2012
- 241 data records: 1 Datetime data type (SurgDate), 1 Object data type (DOW) and 17 Integer data types ('T-28', 'T-21', ....., 'T-1', 'Actual')

```
RangeIndex: 241 entries, 0 to 240
Data columns (total 19 columns):
#   Column      Non-Null Count  Dtype
---  -
0   SurgDate    241 non-null    datetime64[ns]
1   DOW         241 non-null    object
2   T - 28      241 non-null    int64
3   T - 21      241 non-null    int64
4   T - 14      241 non-null    int64
5   T - 13      241 non-null    int64
6   T - 12      241 non-null    int64
7   T - 11      241 non-null    int64
8   T - 10      241 non-null    int64
9   T - 9       241 non-null    int64
10  T - 8       241 non-null    int64
11  T - 7       241 non-null    int64
12  T - 6       241 non-null    int64
13  T - 5       241 non-null    int64
14  T - 4       241 non-null    int64
15  T - 3       241 non-null    int64
16  T - 2       241 non-null    int64
17  T - 1       241 non-null    int64
18  Actual      241 non-null    int64
dtypes: datetime64[ns](1), int64(17), object(1)
```

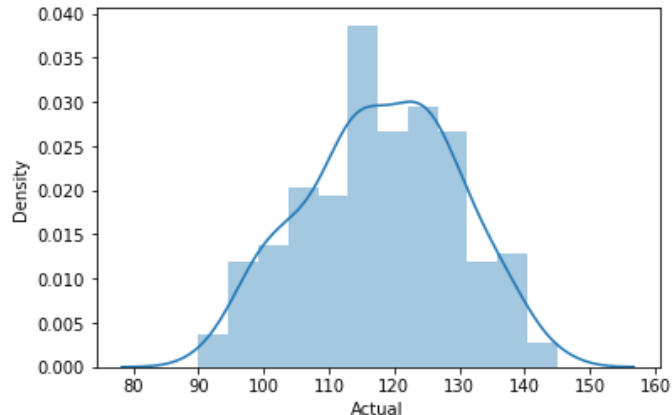
# Exploratory Data Analysis



Four outliers

11/25/2011 Fri      12/23/2011 Fri

12/26/2011 Mon    12/30/2011 Fri



- Average of actual surgeries is 116 with standard deviation of 17.63
- Average of actual surgeries (after removing outlier) is 118 with standard deviation of 11.67

# Exploratory Data Analysis

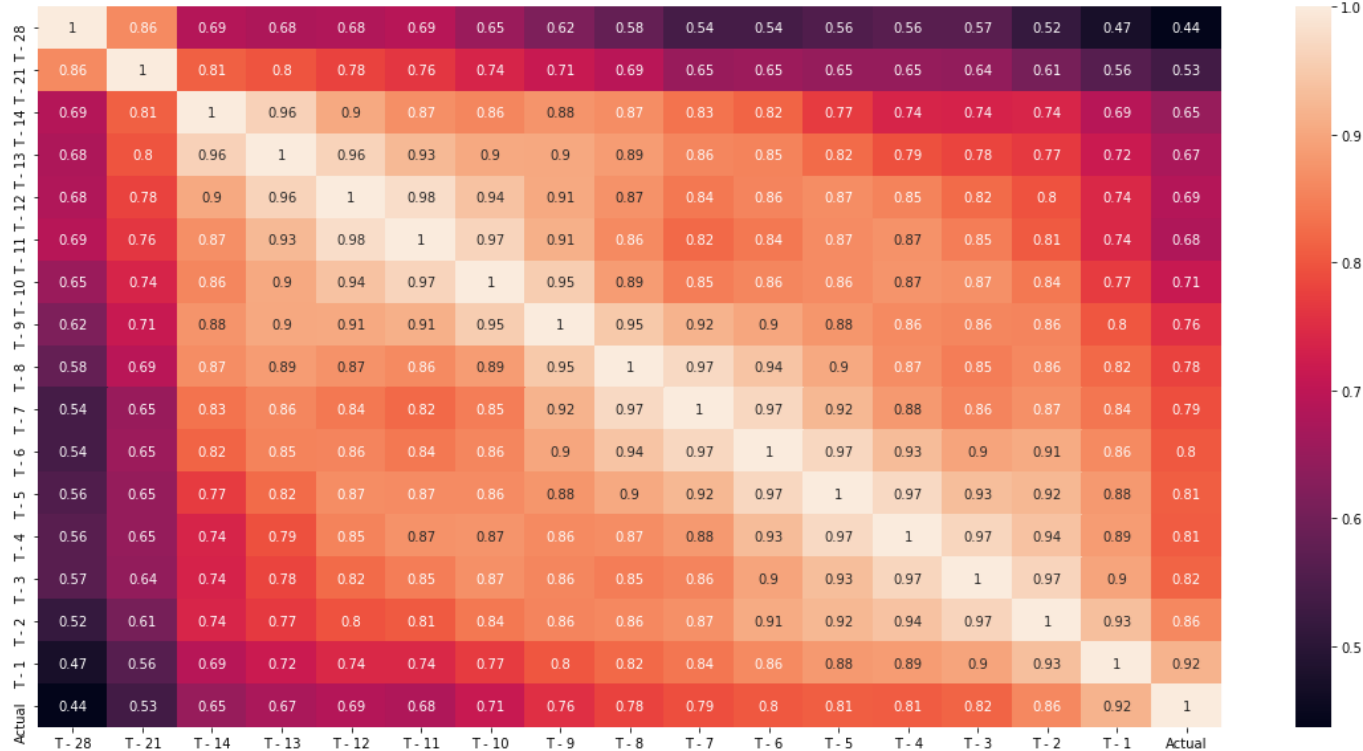
	Mean	Maximum	Standard Deviation
T - 28	34.245833	57	9.404107
T - 21	47.250000	73	11.343809
T - 14	64.458333	93	13.521033
T - 13	67.837500	99	14.227186
T - 12	70.525000	102	14.899924
T - 11	72.375000	106	15.001290
T - 10	74.954167	106	15.127699
T - 9	78.062500	112	15.082765
T - 8	82.375000	113	15.615711
T - 7	86.025000	118	16.118008
T - 6	89.291667	121	16.943863
T - 5	92.104167	121	17.405621
T - 4	94.691667	124	17.504713
T - 3	97.370833	127	17.626530
T - 2	101.170833	131	17.583512
T - 1	110.033333	139	17.817970

	Mean	Maximum	Standard Deviation
T - 28	34.835443	57	8.347481
T - 21	48.033755	73	9.606171
T - 14	65.518987	93	10.712571
T - 13	68.953586	99	11.268392
T - 12	71.683544	102	11.853138
T - 11	73.578059	106	11.785837
T - 10	76.198312	106	11.699573
T - 9	79.345992	112	11.292271
T - 8	83.713080	113	11.514635
T - 7	87.434599	118	11.776923
T - 6	90.759494	121	12.511908
T - 5	93.628692	121	12.798855
T - 4	96.270042	124	12.609641
T - 3	98.995781	127	12.464015
T - 2	102.831224	131	12.050257
T - 1	111.759494	139	11.658141

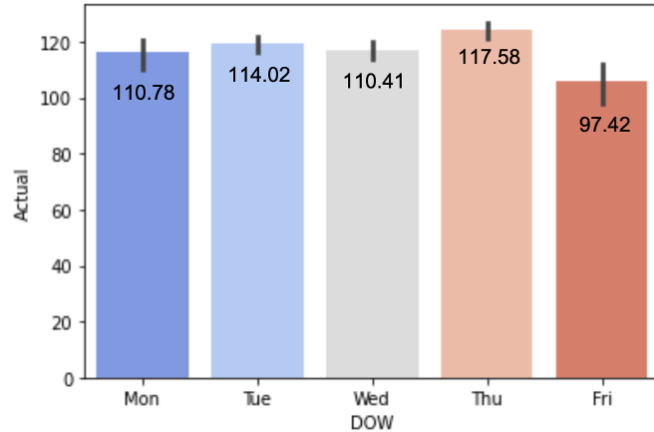
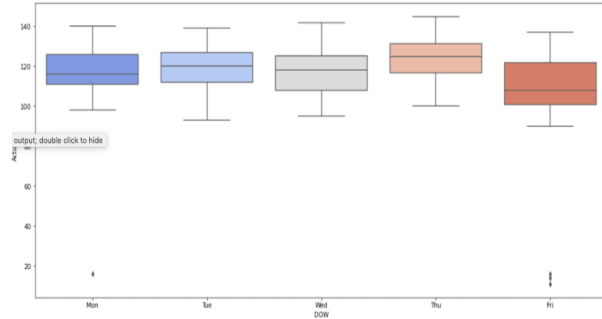
## Observations:

- The average number of surgery cases increase as we move closer to actual day
- Days nearer to the actual surgery date have the strongest correlation

# Exploratory Data Analysis - Correlation



# Exploratory Data Analysis



- Fridays have lowest number of surgeries
- Thursdays have highest number of surgeries
- Alternatively, the high number of surgeries on Thursday might be due to add-on (urgent) cases or fewer no. of cancellations for accommodating emergency cases

# Surgical Volume vs Day of Week

**Null Hypothesis:** Total surgical case volume does not differ based on the day of week

**Alternative Hypothesis:** Total surgical case volume differs based on the day of week

	sum_sq	df	F	PR(>F)
<b>DOW</b>	8909.054042	4.0	8.002734	0.000005
<b>Residual</b>	65681.825626	236.0	NaN	NaN

- By ANOVA test result, we can reject our null hypothesis with 99% confidence
- The p-value obtained from ANOVA analysis is  $< 0.05$
- We can conclude that there are significant differences among surgeries and day of week
- Thus, we reject the null hypothesis and accept alternative hypothesis that total surgical case volume differs by Day of Week

# Data Modelling

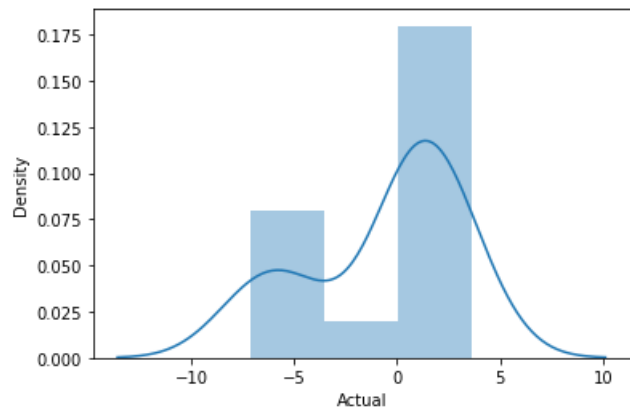
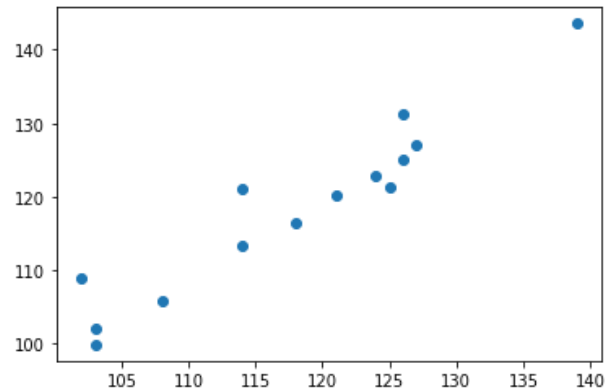
- Data Type: Continuous variables
- Dependent variable: Actual Surgery
- Independent variables: T - 28 to T - 1
- Regression Equation:

$$y = (a_1 * T - 28) + (a_2 * T - 21) + ..... + (a_{x-1} * T - 2) + (a_x * T - 1)$$

Coeffecients	
T - 28	-0.054687
T - 21	0.093760
T - 14	-0.107039
T - 13	0.108562
T - 12	0.071396
T - 11	-0.214555
T - 10	0.023739
T - 9	0.136674
T - 8	-0.012015
T - 7	0.131528
T - 6	-0.136356
T - 5	-0.050957
T - 4	0.109036
T - 3	-0.142896
T - 2	0.081252
T - 1	0.905446



# Model Performance



MAE: 2.779361985282982

MSE: 12.921687554798805

RMSE: 3.5946748886093727

SurgDate	Actual	Predicted	Error	Absolute Error
8/27/2012	127	126.96	0.04	0.04318403
8/28/2012	139	143.58	-4.58	4.58319612
8/29/2012	125	121.37	3.63	3.62693421
8/30/2012	126	131.22	-5.22	5.22316578
8/31/2012	124	122.89	1.11	1.11392976
9/4/2012	114	121.10	-7.10	7.09857747
9/5/2012	103	102.02	0.98	0.97750077
9/6/2012	126	125.00	1.00	1.00293919
9/7/2012	103	99.89	3.11	3.11119318
9/10/2012	118	116.36	1.64	1.63682478
9/11/2012	108	105.75	2.25	2.25328555
9/12/2012	121	120.23	0.77	0.77302749
9/13/2012	114	113.42	0.58	0.57501702
9/14/2012	102	108.89	-6.89	6.89229244
Mean	117.86	118.48		2.78