

Elective Surgery Schedule Solution

Joel Shapiro
Clinical Associate Professor, MEDS

Northwestern | Kellogg

QUESTIONS

1. **Think of multiple ways that the creation of a predictive model can be helpful in this context.**
2. Conduct quant analysis to determine
 - a. Whether # scheduled cases is predictive of final case volume
 - b. Whether predictive power changes as surgery date nears
 - c. Is volume different across days of the week?
3. Most helpful predictive model
4. Sample dashboard

QUESTION 1: UTILITY OF MODEL

Operational improvement

- Improved efficiency (not over- or under-staffed)
- The right equipment and rooms available
- Revenue forecasting for budgeting purposes
- Surgical equipment demand and ordering cycles

Morale improvement

Labor Relations / Compliance

QUESTION 2+3

1. Think of multiple ways that the creation of a predictive model can be helpful in this context.
2. **Conduct quant analysis to determine**
 - a. **Whether # scheduled cases is predictive of final case volume**
 - b. **Whether predictive power changes as surgery date nears**
 - c. **Is volume different across days of the week?**
3. **Most helpful predictive model**
4. Sample dashboard

QUESTION 2

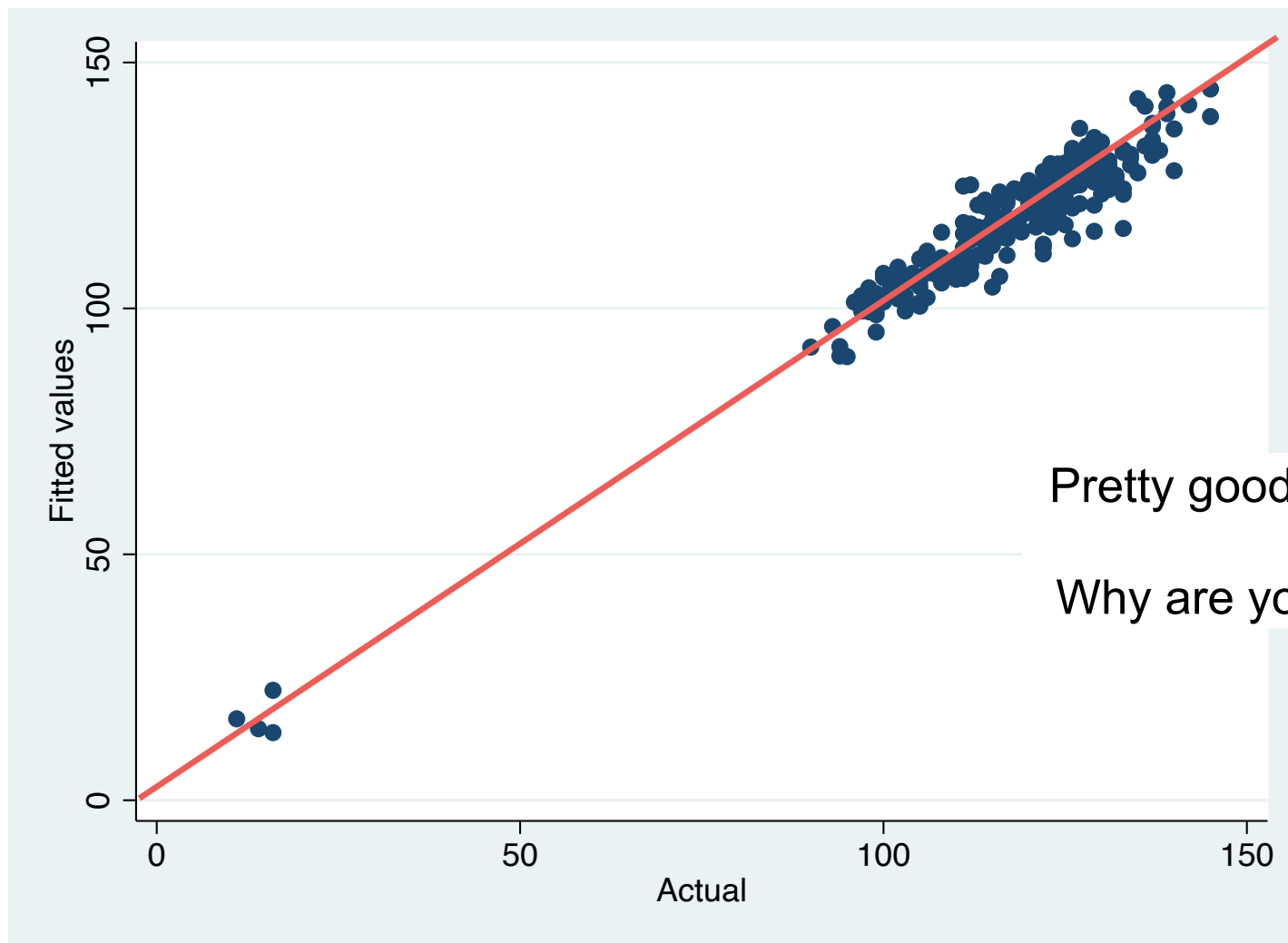
```
. reg actual t28 t21 t14 t13 t12 t11 t10 t9 t8 t7 t6 t5 t4 t3 t2 t1
```

Source	SS	df	MS	Number of obs	=	241
Model	69655.7965	16	4353.48728	F(16, 224)	=	197.60
Residual	4935.08319	224	22.0316214	Prob > F	=	0.0000
				R-squared	=	0.9338
				Adj R-squared	=	0.9291
Total	74590.8797	240	310.795332	Root MSE	=	4.6938

actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t28	-.0384949	.0757068	-0.51	0.612	-.1876836	.1106937
t21	.0889465	.0799785	1.11	0.267	-.0686601	.246553
t14	-.1000675	.1201409	-0.83	0.406	-.3368185	.1366835
t13	.099317	.1806372	0.55	0.583	-.2566486	.4552825
t12	.0466123	.2051393	0.23	0.820	-.3576375	.4508622
t11	-.1749222	.2014459	-0.87	0.386	-.5718937	.2220494
t10	.0169209	.1605497	0.11	0.916	-.29946	.3333018
t9	.1089836	.1431402	0.76	0.447	-.17309	.3910572
t8	.009978	.1451936	0.07	0.945	-.276142	.296098
t7	.1472039	.1708206	0.86	0.390	-.1894171	.4838249
t6	-.1387719	.1845997	-0.75	0.453	-.5025461	.2250024
t5	-.0474666	.1743055	-0.27	0.786	-.3909548	.2960217
t4	.0812244	.174596	0.47	0.642	-.2628363	.4252852
t3	-.162698	.1517318	-1.07	0.285	-.4617023	.1363063
t2	.1287132	.1314298	0.98	0.328	-.1302837	.3877101
t1	.9121406	.073297	12.44	0.000	.7677007	1.056581
_cons	10.87555	1.963297	5.54	0.000	7.006653	14.74444

```
. predict predsurgeries
(option xb assumed; fitted values)
```

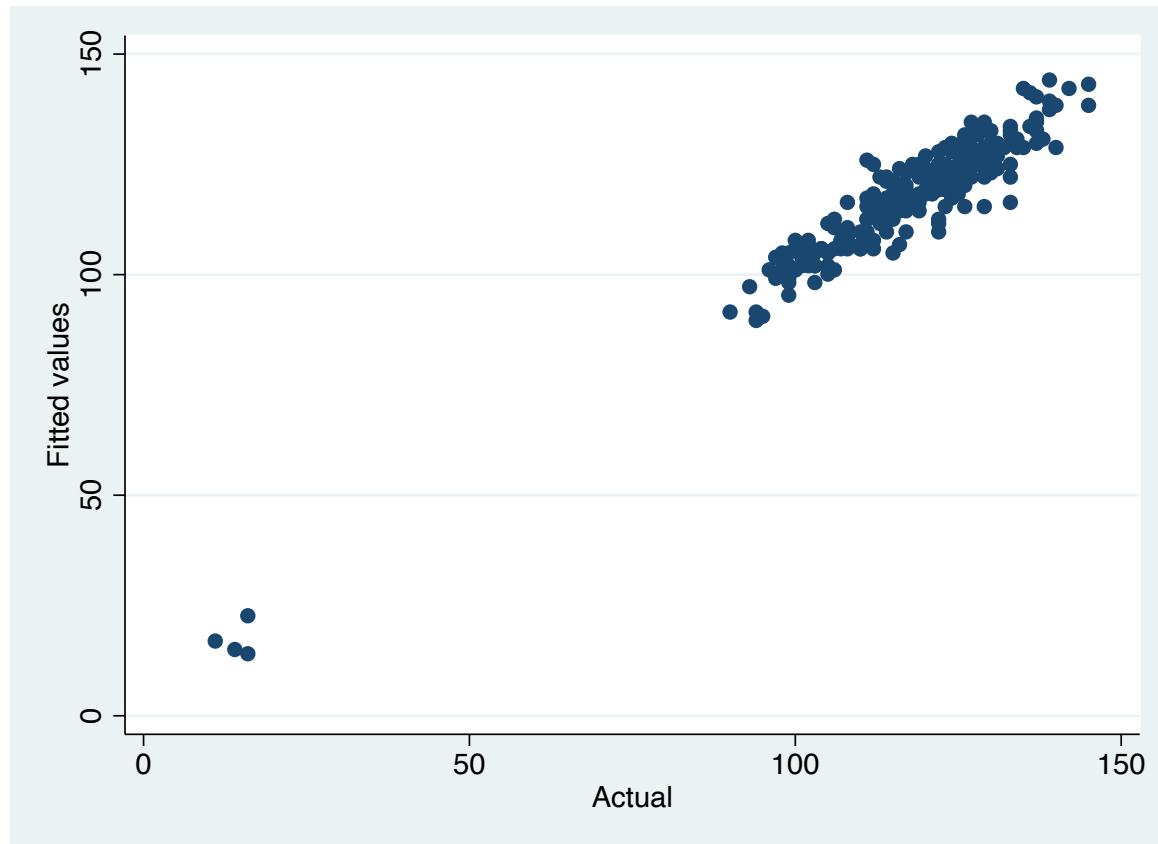
```
. scatter predsurgeries actual
```



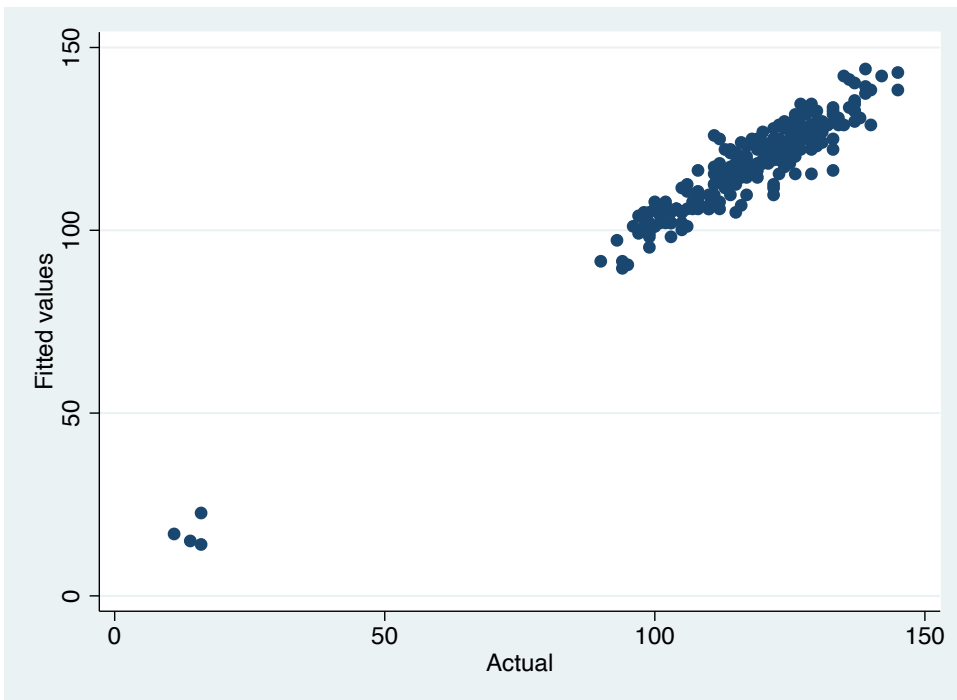
Pretty good prediction, huh?

Why are you not surprised?

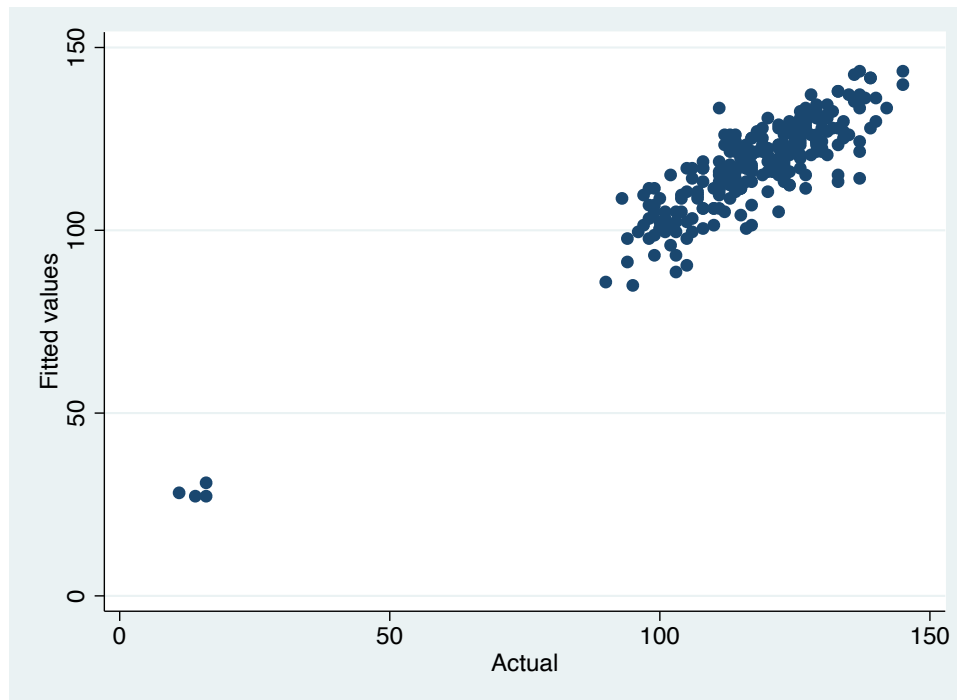
$$actual = a + b * t1$$



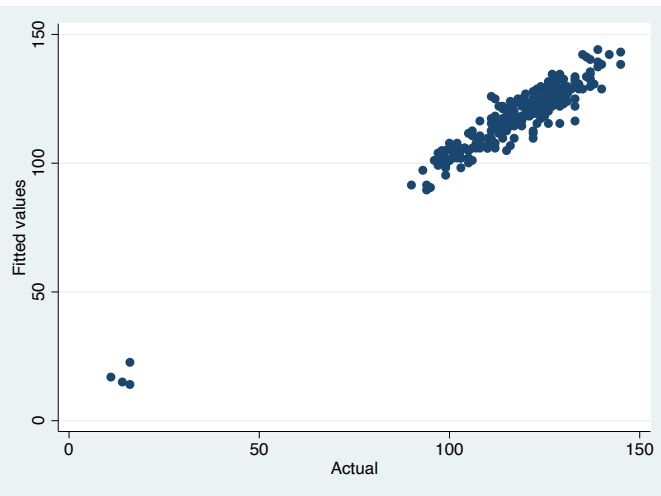
$$actual = a + b * t1$$



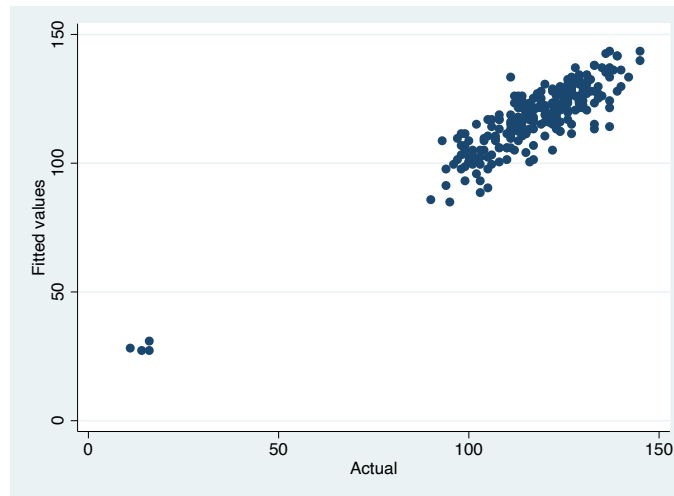
$$actual = a + b * t3$$



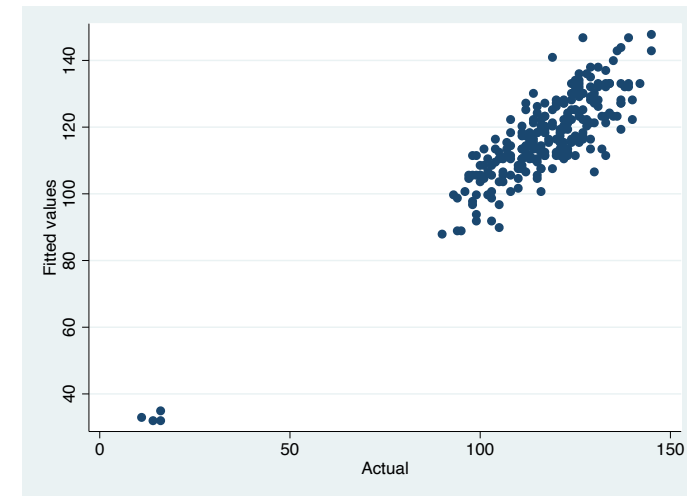
$$actual = a + b * t1$$



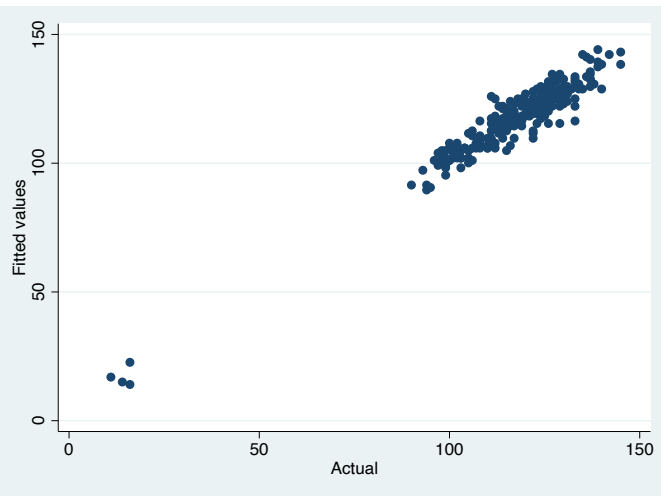
$$actual = a + b * t3$$



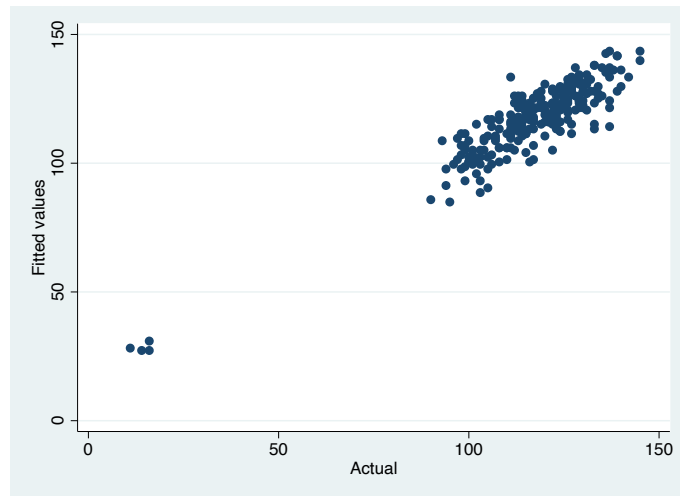
$$actual = a + b * t7$$



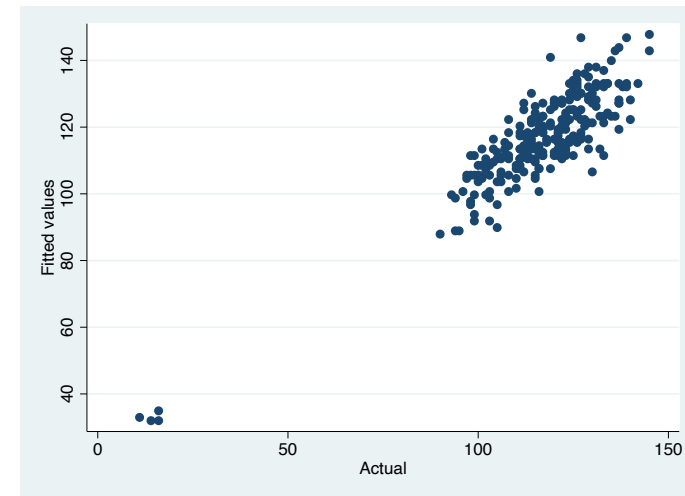
$$actual = a + b * t1$$



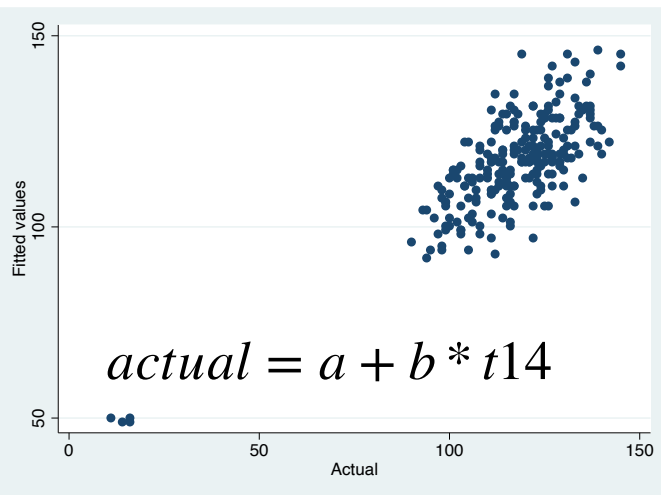
$$actual = a + b * t3$$



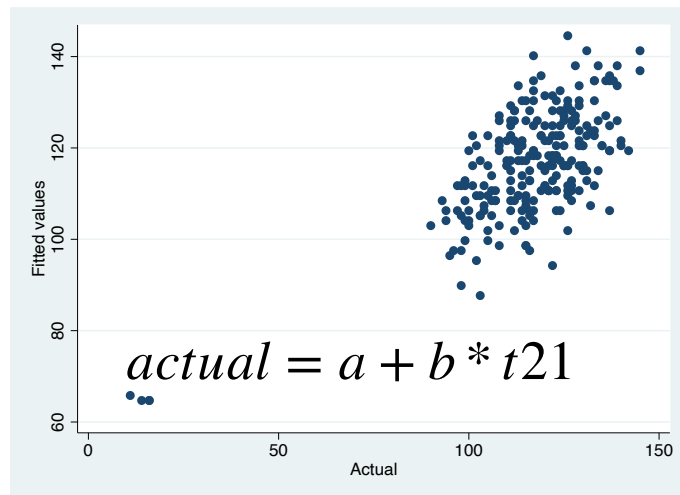
$$actual = a + b * t7$$



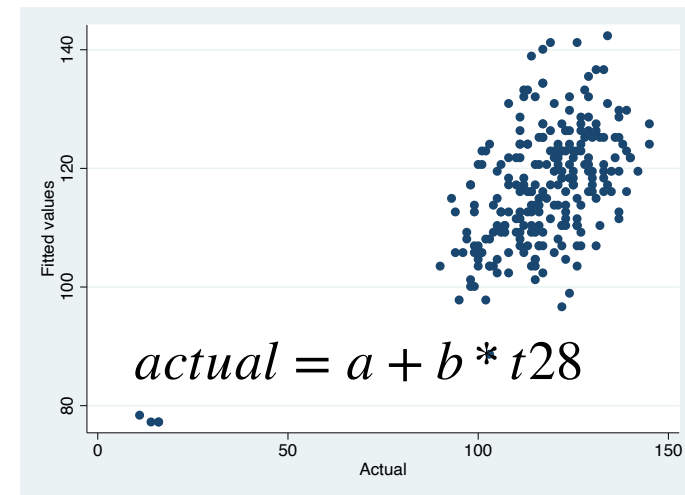
$$actual = a + b * t14$$



$$actual = a + b * t21$$



$$actual = a + b * t28$$



```
. reg actual t28 t21 t14 t13 t12 t11 t10 t9 t8 t7 t6 t5 t4 t3 t2 t1
```

Source	SS	df	MS	Number of obs	=	241
Model	69655.7965	16	4353.48728	F(16, 224)	=	197.60
Residual	4935.08319	224	22.0316214	Prob > F	=	0.0000
				R-squared	=	0.9338
				Adj R-squared	=	0.9291
Total	74590.8797	240	310.795332	Root MSE	=	4.6938

actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t28	-.0384949	.0757068	-0.51	0.612	-.1876836	.1106937
t21	.0889465	.0799785	1.11	0.267	-.0686601	.246553
t14	-.1000675	.1201409	-0.83	0.406	-.3368185	.1366835
t13	.099317	.1806372	0.55	0.583	-.2566486	.4552825
t12	.0466123	.2051393	0.23	0.820	-.3576375	.4508622
t11	-.1749222	.2014459	-0.87	0.386	-.5718937	.2220494
t10	.0169209	.1605497	0.11	0.916	-.29946	.3333018
t9	.1089836	.1431402	0.76	0.447	-.17309	.3910572
t8	.009978	.1451936	0.07	0.945	-.276142	.296098
t7	.1472039	.1708206	0.86	0.390	-.1894171	.4838249
t6	-.1387719	.1845997	-0.75	0.453	-.5025461	.2250024
t5	-.0474666	.1743055	-0.27	0.786	-.3909548	.2960217
t4	.0812244	.174596	0.47	0.642	-.2628363	.4252852
t3	-.162698	.1517318	-1.07	0.285	-.4617023	.1363063
t2	.1287132	.1314298	0.98	0.328	-.1302837	.3877101
t1	.9121406	.073297	12.44	0.000	.7677007	1.056581
_cons	10.87555	1.963297	5.54	0.000	7.006653	14.74444

```
. predict predsurgeries
(option xb assumed; fitted values)
```

```
. scatter predsurgeries actual
```

Strength: high predictive accuracy

Weakness: can't reasonably interpret interpret coefficients. *WHY?*

High correlations among independent variables = fine for prediction, bad for interpretation.

Strength: high predictive accuracy

Weakness: can't reasonably interpret interpret coefficients. *WHY?*

```
. corr t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t21 t28
(obs=241)
```

	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14	t21	t28
t1	1.0000															
t2	0.9701	1.0000														
t3	0.9509	0.9831	1.0000													
t4	0.9431	0.9688	0.9842	1.0000												
t5	0.9373	0.9597	0.9643	0.9849	1.0000											
t6	0.9280	0.9506	0.9466	0.9632	0.9840	1.0000										
t7	0.9181	0.9343	0.9255	0.9384	0.9600	0.9845	1.0000									
t8	0.9092	0.9277	0.9203	0.9301	0.9483	0.9692	0.9848	1.0000								
t9	0.8951	0.9229	0.9245	0.9258	0.9334	0.9457	0.9551	0.9715	1.0000							
t10	0.8712	0.9080	0.9262	0.9280	0.9222	0.9186	0.9122	0.9352	0.9733	1.0000						
t11	0.8519	0.8857	0.9089	0.9239	0.9203	0.9065	0.8965	0.9181	0.9478	0.9793	1.0000					
t12	0.8474	0.8770	0.8939	0.9110	0.9194	0.9128	0.9041	0.9222	0.9415	0.9621	0.9866	1.0000				
t13	0.8350	0.8627	0.8706	0.8783	0.8956	0.9120	0.9144	0.9311	0.9404	0.9416	0.9550	0.9773	1.0000			
t14	0.8215	0.8481	0.8457	0.8460	0.8635	0.8901	0.9005	0.9199	0.9248	0.9134	0.9188	0.9404	0.9756	1.0000		
t21	0.7184	0.7430	0.7637	0.7662	0.7668	0.7713	0.7693	0.7946	0.8074	0.8219	0.8397	0.8491	0.8625	0.8714	1.0000	
t28	0.6294	0.6550	0.6861	0.6855	0.6797	0.6694	0.6699	0.6979	0.7186	0.7443	0.7697	0.7643	0.7613	0.7670	0.8947	1.0000

. reg actual t1

Source	SS	df	MS	Number of obs	=	241
Model	69421.576	1	69421.576	F(1, 239)	=	3209.67
Residual	5169.30372	239	21.6288858	Prob > F	=	0.0000
Total	74590.8797	240	310.795332	R-squared	=	0.9307
				Adj R-squared	=	0.9304
				Root MSE	=	4.6507

actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t1	.9562828	.0168794	56.65	0.000	.9230315	.9895341
_cons	11.1827	1.880881	5.95	0.000	7.477477	14.88792

With only 1 IV, we can say... for every additional surgery that is scheduled the day prior, we see an additional 0.96 actual surgeries.

```
. reg actual t1 t2
```

Source	SS	df	MS	Number of obs	=	241
Model	69422.0074	2	34711.0037	F(2, 238)	=	1598.26
Residual	5168.87223	238	21.7179505	Prob > F	=	0.0000
				R-squared	=	0.9307
				Adj R-squared	=	0.9301
Total	74590.8797	240	310.795332	Root MSE	=	4.6603

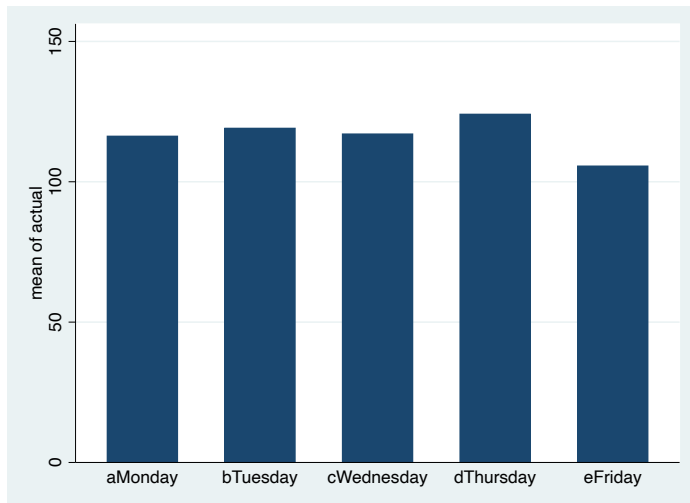
actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t1	.9467596	.0696475	13.59	0.000	.8095553	1.083964
t2	.0099502	.0705924	0.14	0.888	-.1291155	.149016
_cons	11.2237	1.907065	5.89	0.000	7.466817	14.98058

Here, with 2 highly correlated IVs...

*We **CAN** have a great prediction*

*We **can NOT** reasonably say “every additional surgery 2 days before only leads to an additional .0099 actual surgeries.”*

DOES DAY OF WEEK MATTER?



```
. tab dow1, sum(actual)
```

dow1	Summary of Actual		
	Mean	Std. Dev.	Freq.
aMonday	116.25532	18.456138	47
bTuesday	119.08163	10.864385	49
cWednesday	117.04167	11.240047	48
dThursday	124.08333	10.379672	48
eFriday	105.61224	26.357175	49
Total	116.38174	17.629388	241

```
. oneway actual dow1
```

Source	Analysis of Variance			F	Prob > F
	SS	df	MS		
Between groups	8909.05404	4	2227.26351	8.00	0.0000
Within groups	65681.8256	236	278.31282		
Total	74590.8797	240	310.795332		

Bartlett's test for equal variances: $\chi^2(4) = 69.1103$ Prob> $\chi^2 = 0.000$

We can confidently reject the null that all days are the same

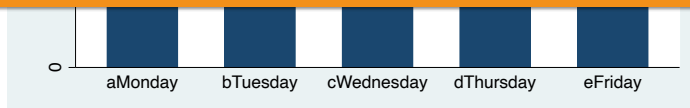
DOES DAY OF WEEK MATTER?



```
. tab dow1, sum(actual)
```

Summary of Actual

Day of the week matters - building it into our prediction will make for accurate predictions, but it's not trivially easy. Nor is the explanation going to be simple for decision-makers to process.



eFriday	105.61224	26.357175	49
Total	116.38174	17.629388	241

```
. oneway actual dow1
```

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	8909.05404	4	2227.26351	8.00	0.0000
Within groups	65681.8256	236	278.31282		
Total	74590.8797	240	310.795332		

Bartlett's test for equal variances: $\chi^2(4) = 69.1103$ Prob> $\chi^2 = 0.000$

We can confidently reject the null that all days are the same

OPTIONS FOR DEALING WITH DAY OF WEEK

Ignore: Create a regularly updating prediction that is based only on #days prior to date. By ignoring it, we sacrifice accuracy for explainability.

I've chosen t7 as an example b/c it is enough advance time for us to plan, but not so much so that it becomes too inaccurate. It FEELS like a decent balance.

. reg actual t7 ←

Source	SS	df	MS	Number of obs	=	241
Model	59853.1765	1	59853.1765	F(1, 239)	=	970.63
Residual	14737.7032	239	61.6640301	Prob > F	=	0.0000
Total	74590.8797	240	310.795332	R-squared	=	0.8024
				Adj R-squared	=	0.8016
				Root MSE	=	7.8526

actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t7	.9815375	.031505	31.15	0.000	.9194746	1.0436
_cons	31.96952	2.756242	11.60	0.000	26.53989	37.39915

Predicted = if no cases were scheduled at t7, we'd expect 32 cases. For every case at t7, we expect 0.98 new additional. In other words, prediction = 98% of cases scheduled at t7, plus another 32.

OPTIONS FOR DEALING WITH DAY OF WEEK

Use dummies for each day of week: Create a regularly updating prediction that is based only on #days prior to date. By ignoring it, we sacrifice accuracy for explainability.

$$\text{Surgeries} = a + b1 * t7 + b2 * \text{Tue} + b3 * \text{Wed} + b4 * \text{Thu} + b5 * \text{Fri}$$

```
. reg actual t7 tues wed thu fri
```

Source	SS	df	MS	Number of obs	=	241
Model	60997.5416	5	12199.5083	F(5, 235)	=	210.90
Residual	13593.3381	235	57.8439918	Prob > F	=	0.0000
				R-squared	=	0.8178
				Adj R-squared	=	0.8139
Total	74590.8797	240	310.795332	Root MSE	=	7.6055

actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t7	1.03106	.0343591	30.01	0.000	.9633685	1.098751
tues	-4.56168	1.572204	-2.90	0.004	-7.659095	-1.464264
wed	-3.13497	1.566168	-2.00	0.046	-6.220494	-.049447
thu	-7.134236	1.638417	-4.35	0.000	-10.3621	-3.906373
fri	-2.775591	1.574786	-1.76	0.079	-5.878092	.3269098
_cons	31.24773	3.042282	10.27	0.000	25.2541	37.24136

Predictions: 103% of cases scheduled at t7 plus...

If Mon: about 31 more
 If Tue: about 27 more
 If Wed: about 28 more
 If Thu: about 24 more
 If Fri: about 29 more

OPTIONS FOR DEALING WITH DAY OF WEEK

Use dummies + interactions: Create a regularly updating prediction that is based only on #days prior to date. By ignoring it, we sacrifice accuracy for explainability.

$$\begin{aligned} \text{Surgeries} = & a + b1 * t7 + b2 * Tue + b3 * Wed + b4 * Thu + b5 * Fri \\ & + b6 * Tue * t7 + b7 * Wed * t7 + b8 * Thu * t7 + b9 * Fri * t7 \end{aligned}$$

That is, does $t7$ matter differently depending on day of week? Does $t7$ predict final cases differently if it's a Monday or Tuesday or Wednesday...

```
. reg actual t7 tues wed thu fri tuext7 wedxt7 thuxt7 frixt7
```

Source	SS	df	MS	Number of obs	=	241
Model	62507.0822	9	6945.23136	F(9, 231)	=	132.77
Residual	12083.7974	231	52.3108114	Prob > F	=	0.0000
				R-squared	=	0.8380
				Adj R-squared	=	0.8317
Total	74590.8797	240	310.795332	Root MSE	=	7.2326

actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t7	1.11845	.0711491	15.72	0.000	.9782663	1.258635
tues	20.65563	11.47336	1.80	0.073	-1.950187	43.26144
wed	22.46584	10.10844	2.22	0.027	2.549323	42.38236
thu	29.82296	10.92592	2.73	0.007	8.295759	51.35016
fri	-5.854172	7.04852	-0.83	0.407	-19.74178	8.033434
tuext7	-.2883925	.1299931	-2.22	0.027	-.5445162	-.0322688
wedxt7	-.3006745	.1177971	-2.55	0.011	-.5327685	-.0685805
thuxt7	-.3942453	.1177535	-3.35	0.001	-.6262535	-.1622371
frixt7	.0500614	.0860285	0.58	0.561	-.1194395	.2195623
_cons	24.04265	5.960131	4.03	0.000	12.29948	35.78582

Monday
 24 = constant
 1.11 = marginal effect of t7
*Prediction = 24+1.11*t7*
 E.g., t7 = 100, then
 Predicted final volume =
24 + 1.11*100 = 135

```
. reg actual t7 tues wed thu fri tuext7 wedxt7 thuxt7 frixt7
```

Source	SS	df	MS	Number of obs	=	241
Model	62507.0822	9	6945.23136	F(9, 231)	=	132.77
Residual	12083.7974	231	52.3108114	Prob > F	=	0.0000
				R-squared	=	0.8380
				Adj R-squared	=	0.8317
Total	74590.8797	240	310.795332	Root MSE	=	7.2326

actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t7	1.11845	.0711491	15.72	0.000	.9782663	1.258635
tues	20.65563	11.47336	1.80	0.073	-1.950187	43.26144
wed	22.46584	10.10844	2.22	0.027	2.549323	42.38236
thu	29.82296	10.92592	2.73	0.007	8.295759	51.35016
fri	-5.854172	7.04852	-0.83	0.407	-19.74178	8.033434
tuext7	-.2883925	.1299931	-2.22	0.027	-.5445162	-.0322688
wedxt7	-.3006745	.1177971	-2.55	0.011	-.5327685	-.0685805
thuxt7	-.3942453	.1177535	-3.35	0.001	-.6262535	-.1622371
frixt7	.0500614	.0860285	0.58	0.561	-.1194395	.2195623
_cons	24.04265	5.960131	4.03	0.000	12.29948	35.78582

Monday

24 = constant
1.11 = marginal effect of t7

*Prediction = 24 + 1.11*t7*

E.g., t7 = 100, then
Predicted final volume =
24 + 1.11*100 = 135

Tuesday

24+20 = 44 constant
1.11-0.28=0.83 marginal effect

*Prediction = 44 + 0.83*t7*

E.g., t7 = 100, then
Predicted final volume =
44 + 0.83*100 = 127

In other words, 100 at t7 leads to a different prediction if a Monday versus Tuesday versus...

In other words, 100 at t7 leads to a different prediction if a Monday versus Tuesday versus...

24 = constant
1.11 = marginal effect of t7

24 = constant
1.11 = marginal effect of t7

24 = constant
1.11 = marginal effect of t7

24+20 = 44 constant
1.11-0.28=0.83 marginal effect

24+20 = 44 constant
1.11-0.28=0.83 marginal effect

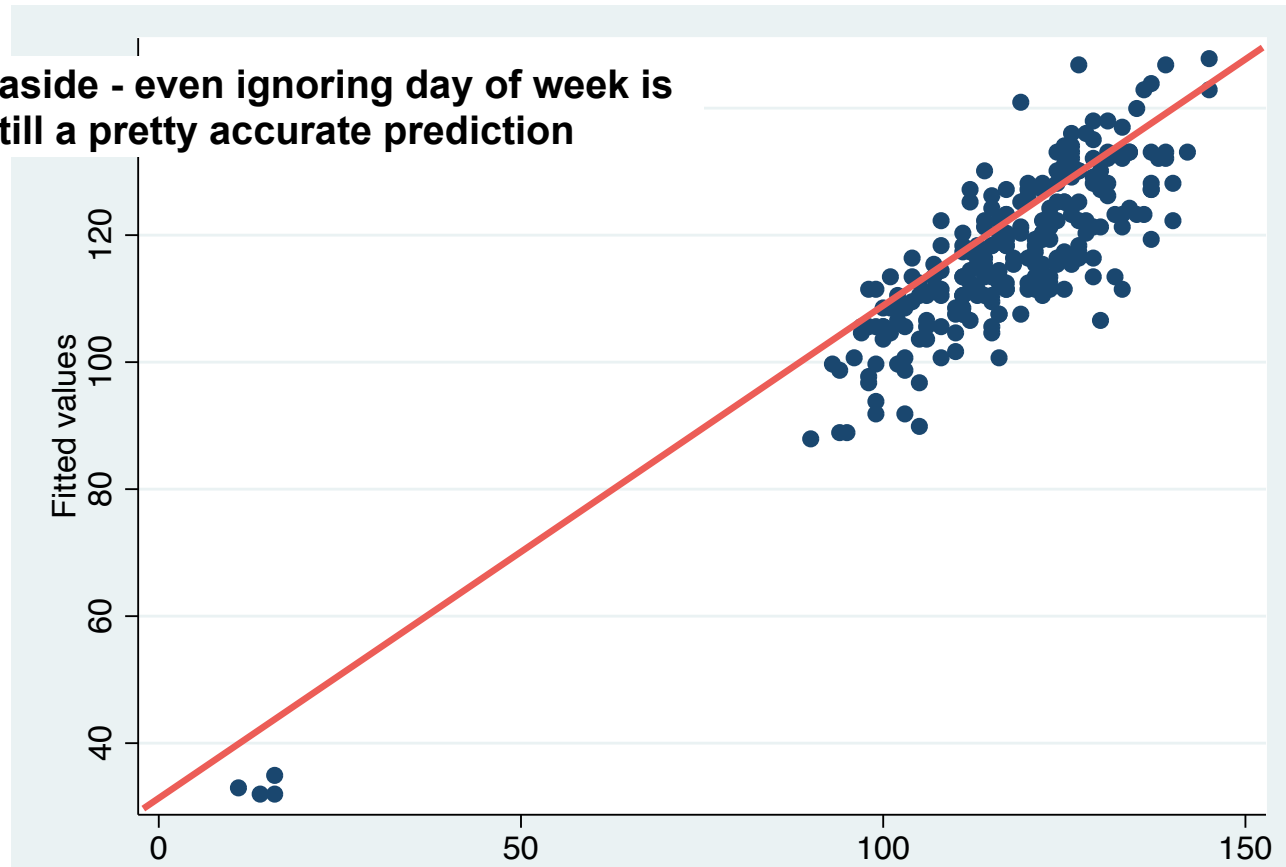
24+20 = 44 constant
1.11-0.28=0.83 marginal effect

24+20 = 44 constant
1.11-0.28=0.83 marginal effect

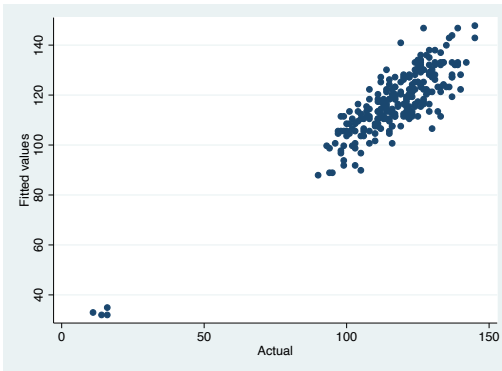
actual	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t7	1.11845	.0711491	15.72	0.000	.9782663	1.258635
tues	20.65563	11.47336	1.80	0.073	-1.950187	43.26144
wed	22.46584	10.10844	2.22	0.027	2.549323	42.38236
thu	29.82296	10.92592	2.73	0.007	8.295759	51.35016
fri	-5.854172	7.04852	-0.83	0.407	-19.74178	8.033434
tuext7	-.2883925	.1299931	-2.22	0.027	-.5445162	-.0322688
wedxt7	-.3006745	.1177971	-2.55	0.011	-.5327685	-.0685805
thuxt7	-.3942453	.1177535	-3.35	0.001	-.6262535	-.1622371
frixt7	.0500614	.0860285	0.58	0.561	-.1194395	.2195623
_cons	24.04265	5.960131	4.03	0.000	12.29948	35.78582

Quick aside - even ignoring day of week is still a pretty accurate prediction

**Predicted
Values**

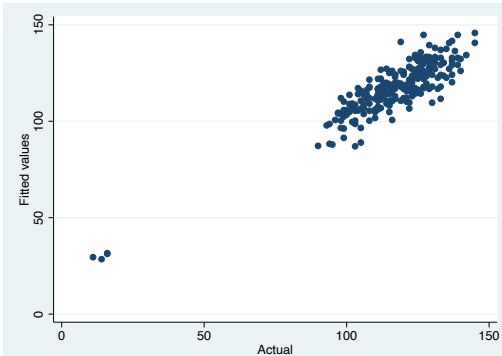


Actual Values



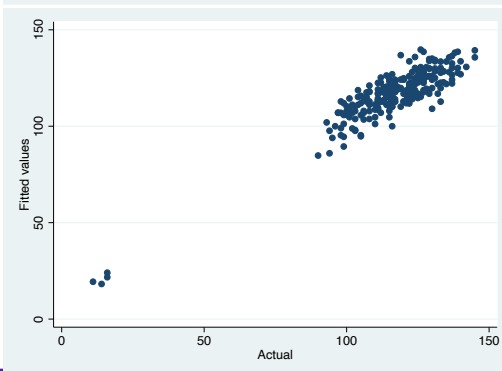
Create a regularly updating prediction that is based only on #days prior to date. Ignore day of week, sacrificing accuracy for explainability.

(At t7, $R^2 = 0.80$)



Build in day or week through dummies only

(At t7, $R^2 = 0.81$)



Build in day or week through dummies and interactions.

(At t7, $R^2 = 0.84$)

QUESTION 4

1. Think of multiple ways that the creation of a predictive model can be helpful in this context.
2. Conduct quant analysis to determine
 - a. Whether # scheduled cases is predictive of final case volume
 - b. Whether predictive power changes as surgery date nears
 - c. Is volume different across days of the week?
3. Most helpful predictive model
4. **Sample dashboard**

SAMPLE DASHBOARD

	Mon 7/14	Tues 7/15	Wed 7/16	Thu 7/17	Fri 7/18	Mon 7/21	Tues 7/22	Wed 7/23	Thu 7/24	Fri 7/25
Budgeted Volume	123	123	123	123	123	123	123	123	123	123
Prediction based on booked cases as of...	T-3	T-4	T-5	T-6	T-7	T-10	T-11	T-12	T-13	T-14
Booked cases to date	111	116	99	120	68	91	87	83	92	42
Predicted Volume	125	131	121	138	98	130	124	126	133	86
80% chance final will be between	118-132	124-138	114-128	131-146	89-107	121-139	115-134	117-135	123-142	75-98
Predicted Load	Medium	High	Medium	Very High	Low	Medium	Medium	Medium	High	Very Low

COOLEST SOLUTION I'VE SEEN SO FAR...

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2														
3														
4														
5	Enter data here													
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
26														
27														
28														
29														
30														
31														
32														
	Dashboard	OgData	DaySeparation	ErrorbyPredictionTime	PredictionData	DashboardData	+							

Future Caseload Expectation Dashboard

Day of the Week	Timeframe	Scheduled Cases
Monday	T-28	86

Today's Date		Prediction Date
11/14/22	28 days away	12/12/22 Mon
	21 days away	12/5/22 Mon
	14 days away	11/28/22 Mon
	7 days away	11/21/22 Mon

Expected Cases
172

Case Volume	
Very High	>144
High	126-144
Normal	107-125
Low	88-106
Very Low	<88

Timeframe	Avg MoE
T-28	8.83%
T-21	8.65%
T-14	7.70%
T-7	5.83%

Upcoming Schedule

Date	Expected Cases	MoE
2/7/22 (M)	120	5.83%
2/8/22 (Tu)	138	5.83%
2/9/22 (W)	90	7.70%
2/10/22 (Th)	111	7.70%
2/11/22 (F)	121	7.70%
2/12/22 (Sa)		
2/13/22 (Su)		
2/14/22 (M)	141	8.65%
2/15/22 (Tu)	145	8.65%
2/16/22 (W)	114	8.65%
2/17/22 (Th)	110	8.65%
2/18/22 (F)	14	8.65%
2/19/22 (Sa)		
2/20/22 (Su)		