Task

R Markdown

Part 0

Some weird entries were filtered out. (pred_lnlos ≤ 0, leave before arrive, leave = arrive)

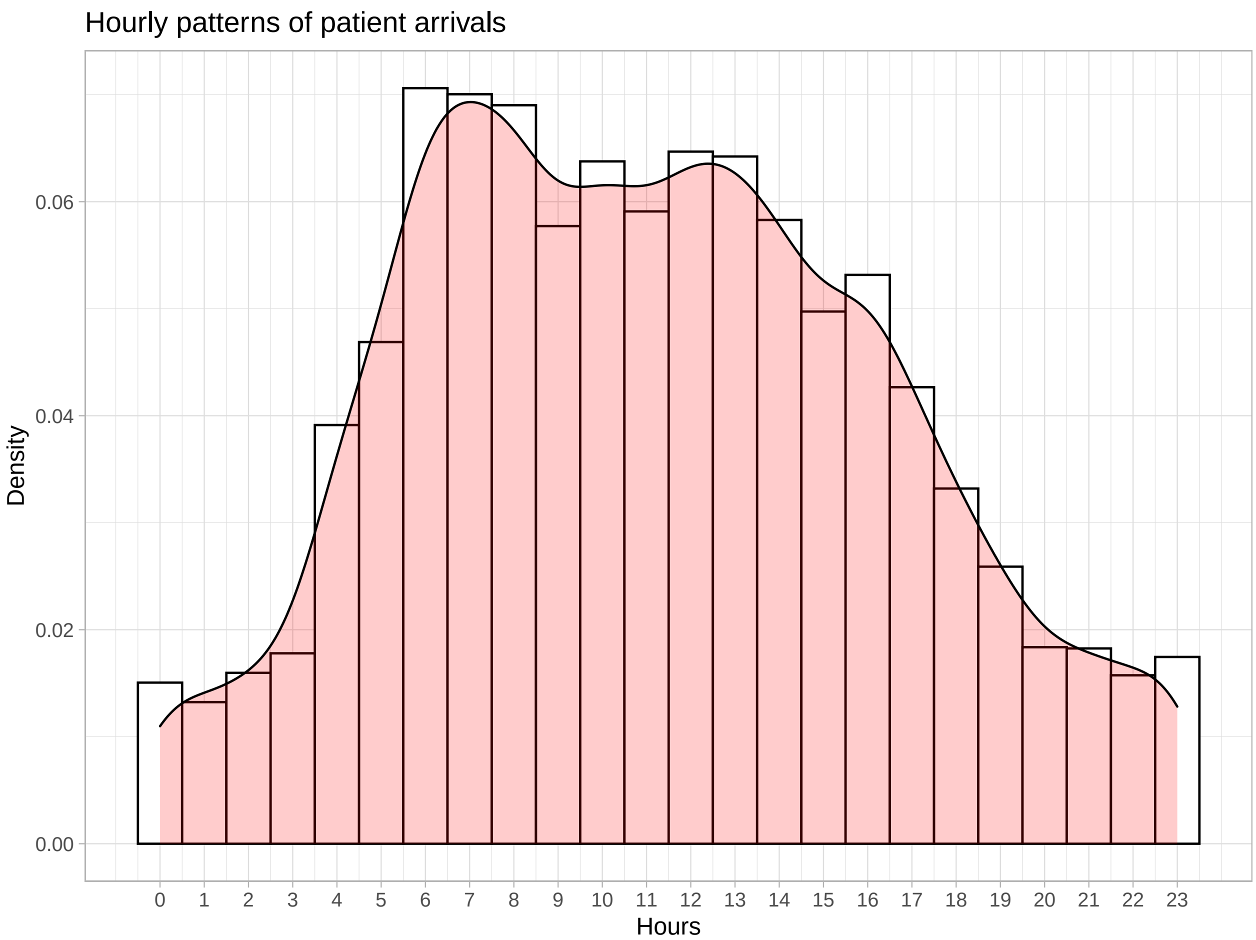
```
## [1] 0.07437828
```

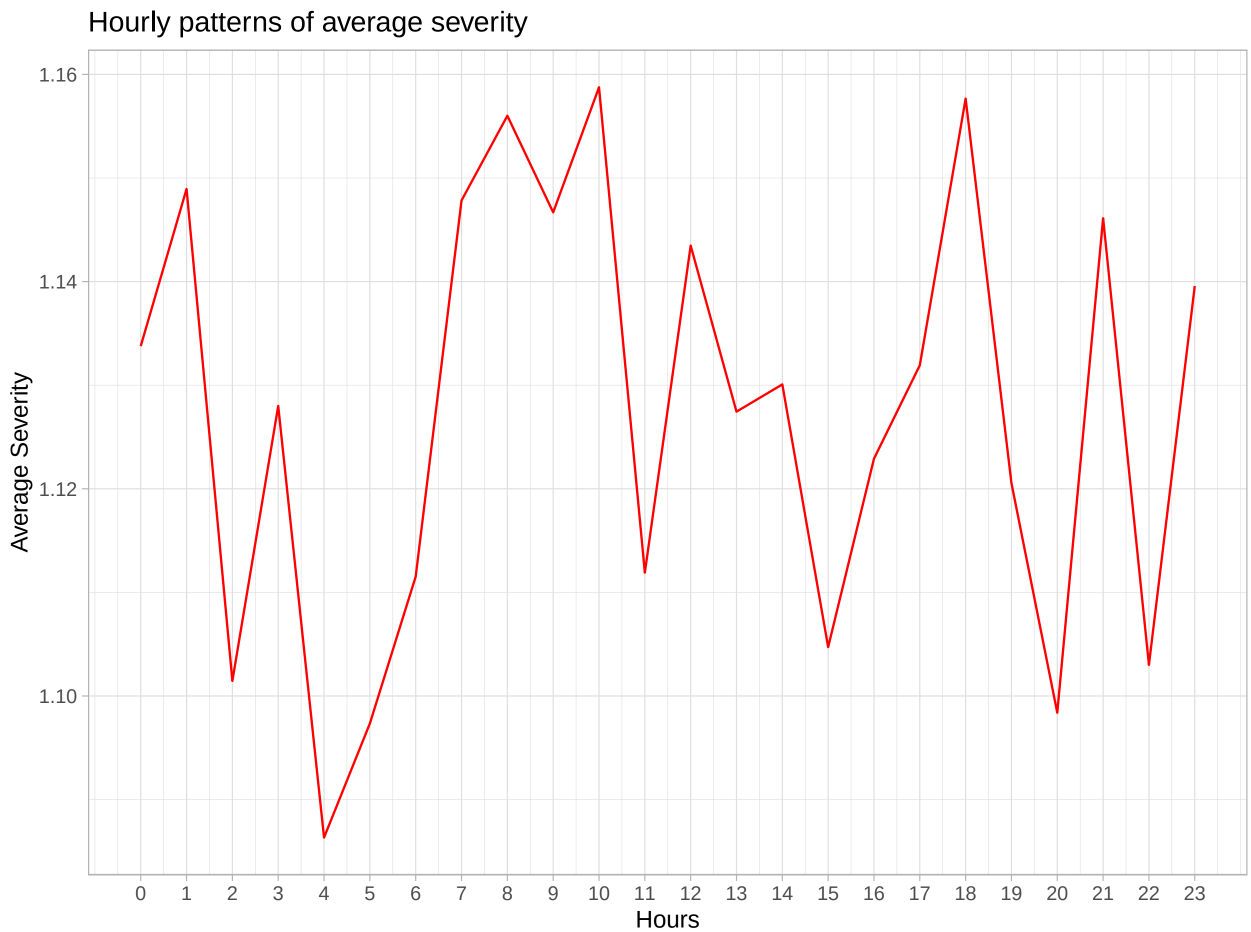
```
## [1] 0.1362081
```

Part 1

Approximately, 7.43% of the patients arrived before their physician's shift starts and 13.61% discharged after their physician's shift ends.

Part 2


Hourly patterns of patient arrivals
Density
0.06
0.04
0.02
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hours



From the hourly patterns of patient arrivals, We can see that patients come in most at 6 am and least at 1 am. There is a peak between 6am and 2pm, after which the number of patients keeps dropping until it starts to rise after 1am.

The hourly patterns of average severity shows that there are several peaks in patient severity throughout the day: 1 am, 7 am to 10 am, 12 noon, 6 pm and 9 pm.

We want to see whether patient severity is or is not predicted by hour of the day, let's add a regression line to the plot and the report of regression is shown below.

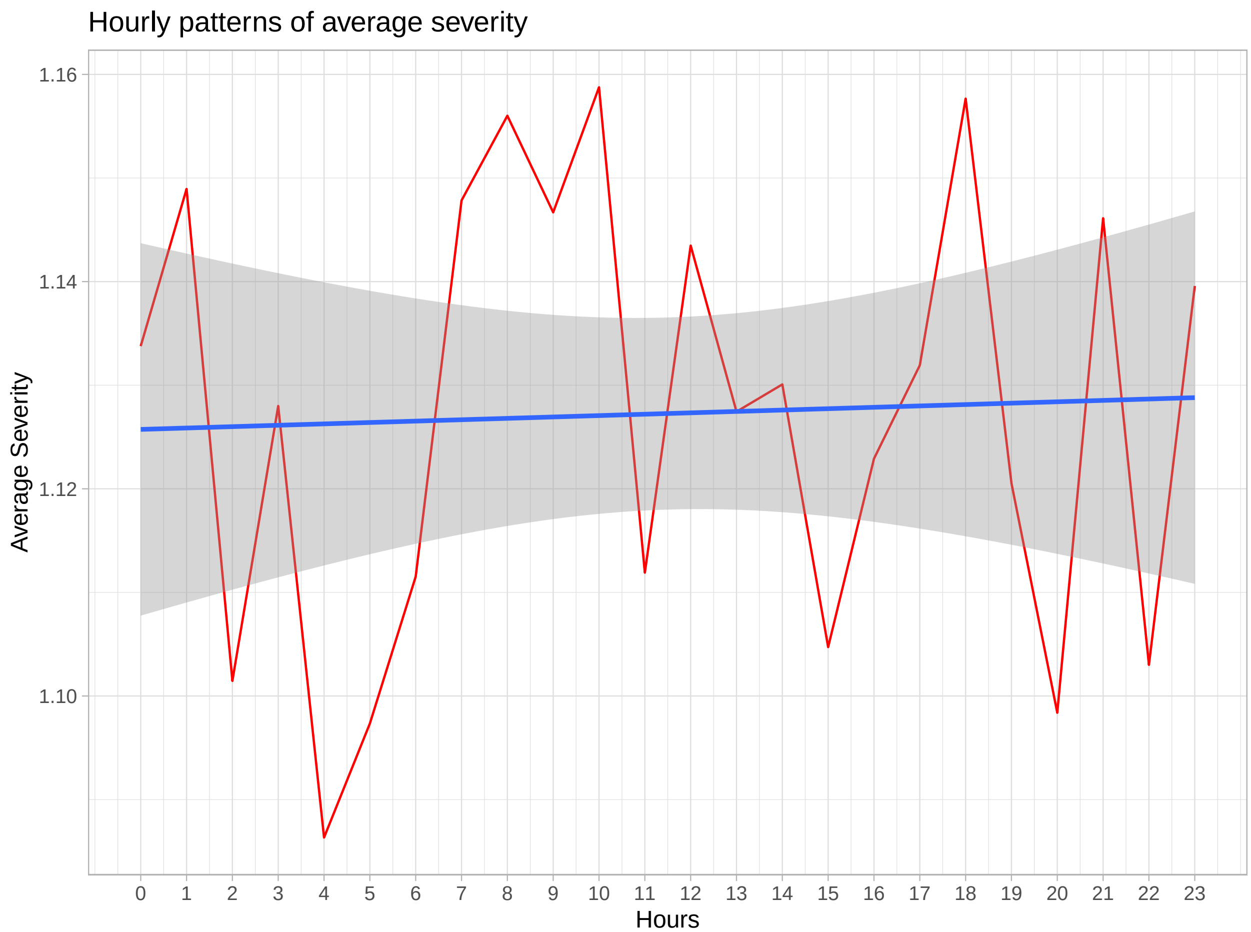
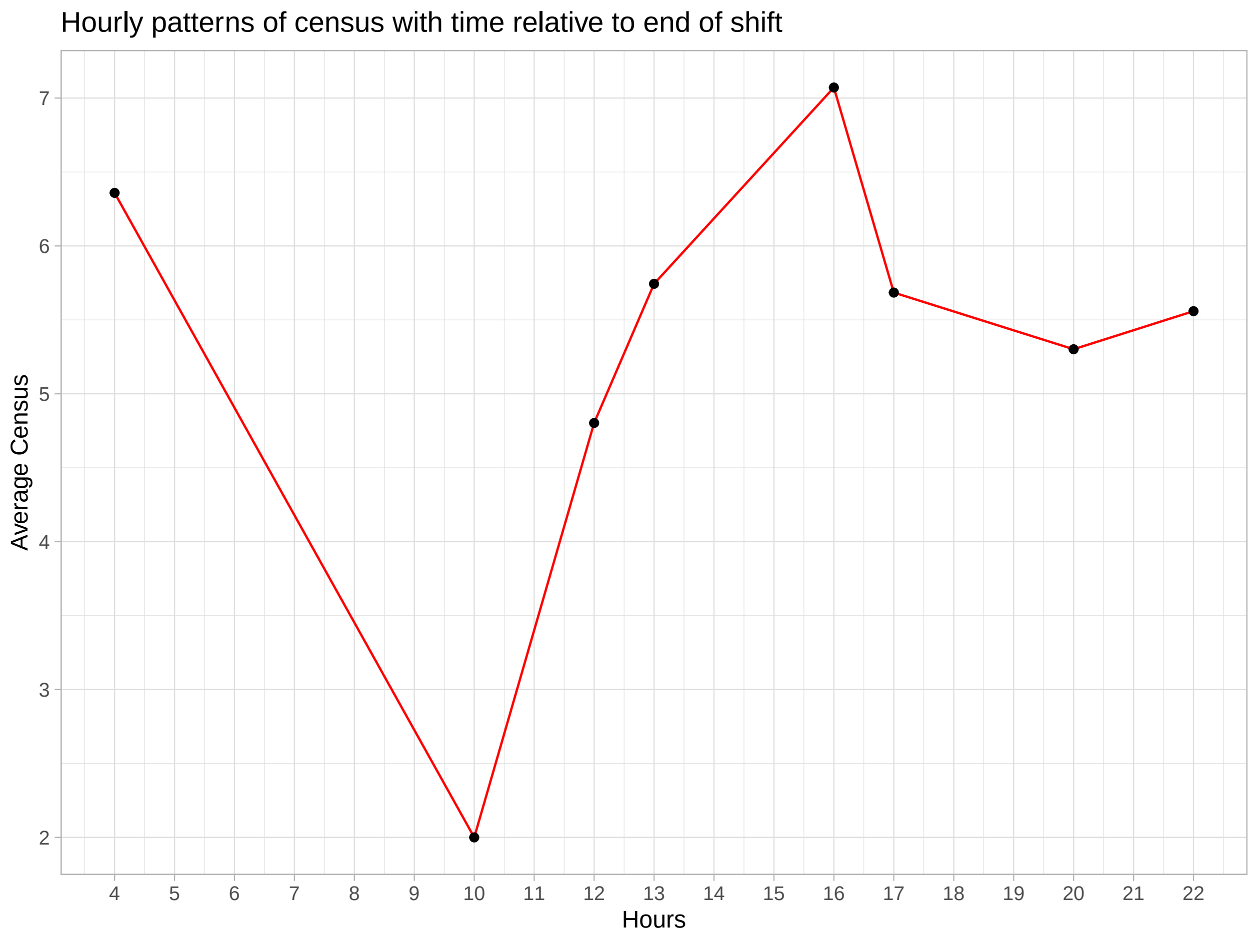


Table 1: The relationship between hours and patient severity.

	average_severity
arrive_hours	0.0001
	(0.001)
Constant	1.126***
	(0.009)
Observations	24
R^2	0.002
Adjusted R^2	−0.043
Residual Std. Error	0.022 (df = 22)
F Statistic	0.043 (df = 1; 22)
Note:	*p<0.1; **p<0.05; ***p<0.01

It seems patient severity is not predicted by hour of the day.

Part 3



How does the census vary with time relative to end of shift?

Above graph shows that the average census for each shift end time. We can see that for the shift that ends at 10 a.m., the average census is 1.5, while for the shift that ends at 4 p.m., the average census on shift is about 5.5. These results were predictable because there were fewer patients at nignt and more during the day.

Discuss conceptually how you construct censuses and address issues with discrete time.

First of all, I created empty tables and find all distinct ids and the shifts. In the first loop, I looped through all the distinct ids and shifts to subset the original data. Then, I separate the shifts into two cases: the same day shift and two day shift.

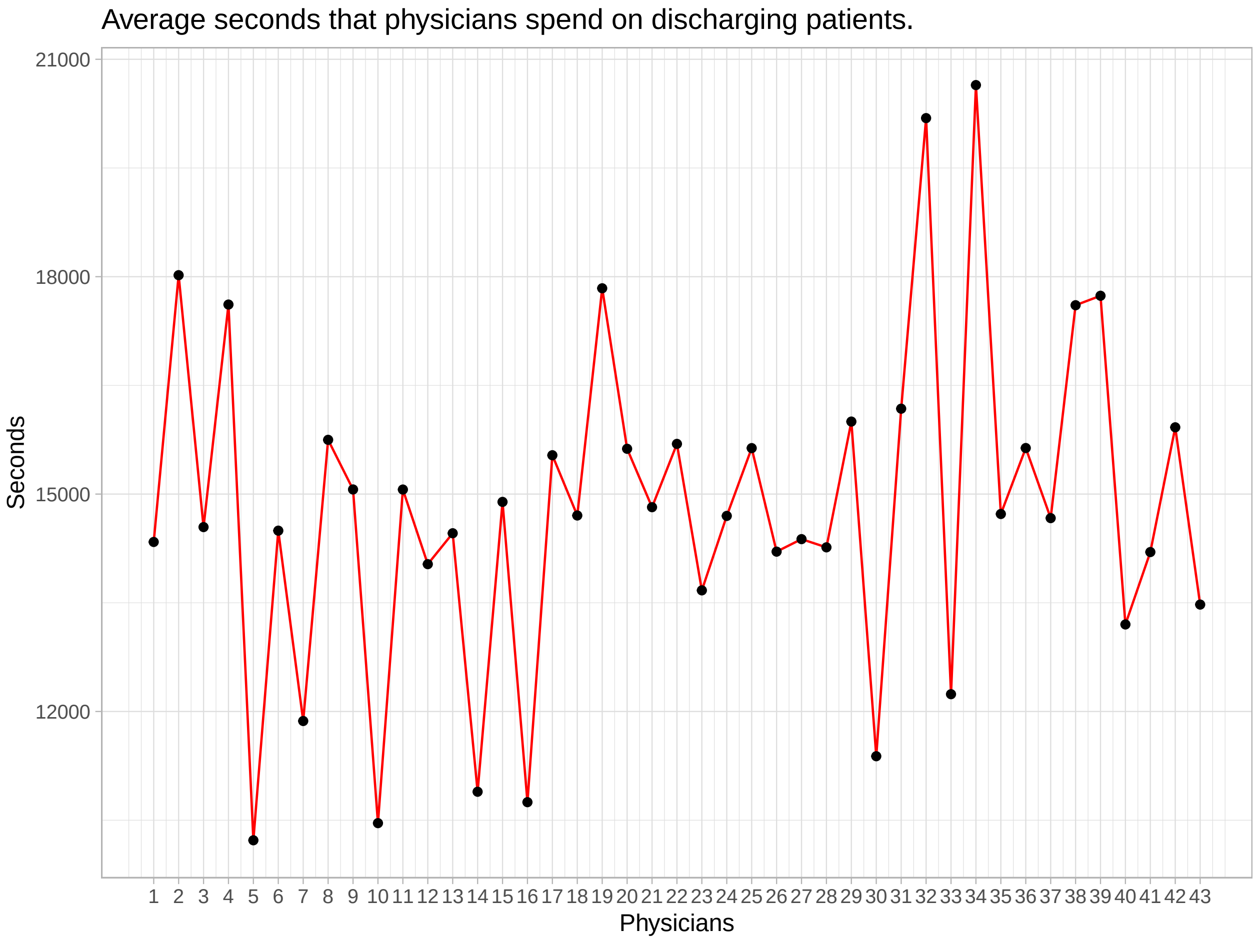
The same day shift is simple. I used big loop to loop throught the hours, then use the small loop to loop through the subseted data to compare hours, count, and enter it to the table.

The two day shift is bit complicated. Since the two day shifts always start after noon, I added 24 to the arrival time when necessary. For example, if the indicator that shows if the patient spent two days is true, I would add 24 hours to the hour of leave to compare with the hour of shift start + the nth hour of the shift.

When comparing hours, I used function `hours` to extract the hours of the arrival, leave, shift start and shift end. `hours` only cares about the hour term. If a patient leaves 1 minute after the first hour of the shift, I would still count it for the second hour of the shift.

Upper bound Census and Lower bound Census are generated by slightly changing the criterion of comparing. Please see code. (I'm not sure what are they tho)

Part 4



Physician 5 appears to be the fastest at discharging patients. Each patient on average, took him/her about 10223 seconds (170 minutes). But if we look at the data, the result doesn't look good.

Table 2: Record of Physician 5

visit_num	phys_id	shift_date	shift_start	shift_end	arrive_date	arrive_time	leave_date	leave_time	pre
810	5	1982-06-08	13:00:00	22:00:00	1982-06-08	17:31:04	1982-06-08	21:41:11	
811	5	1982-05-28	19:00:00	04:00:00	1982-05-29	00:19:31	1982-05-29	01:50:10	

we only have 2 observations for Physician 5. Let's look at the second lowest.

Table 3: Record of Physician 10

visit_num	phys_id	shift_date	shift_start	shift_end	arrive_date	arrive_time	leave_date	leave_time	pre
1542	10	1982-06-09	19:00:00	04:00:00	1982-06-09	21:57:11	1982-06-10	02:48:49	1
1543	10	1982-05-21	03:00:00	13:00:00	1982-05-21	04:17:30	1982-05-21	07:03:31	1
1544	10	1982-07-07	04:00:00	13:00:00	1982-07-07	06:50:32	1982-07-07	07:49:31	1
1545	10	1982-07-02	11:00:00	20:00:00	1982-07-02	12:07:04	1982-07-02	14:46:32	1
1546	10	1982-07-02	11:00:00	20:00:00	1982-07-02	14:21:25	1982-07-02	16:34:40	1
1547	10	1982-06-04	11:00:00	20:00:00	1982-06-04	13:54:11	1982-06-04	14:40:03	1
1548	10	1982-05-21	03:00:00	13:00:00	1982-05-21	06:26:23	1982-05-21	08:27:37	0
1549	10	1982-06-21	13:00:00	22:00:00	1982-06-21	14:21:22	1982-06-22	05:34:30	1
1550	10	1982-05-21	03:00:00	13:00:00	1982-05-21	04:17:30	1982-05-21	05:07:44	0
1551	10	1982-06-29	19:00:00	04:00:00	1982-06-29	21:40:35	1982-06-30	03:22:28	1

Physician 10 looks better, it has 158 observations.

Now let's do the regression:

Table 4: The relationship between Physician and log stay time

	log(diff)
phys_id	0.001
	(0.001)
Constant	9.209***
	(0.018)
Observations	8,766
R^2	0.0003
Adjusted R^2	0.0002
Residual Std. Error	0.836 (df = 8764)
F Statistic	2.372 (df = 1; 8764)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can't get much useful information. The coefficients only indicate that the latter Physicians will have longer patient stays, which implies Physicians 1 is the fastest, and it is not true.

What are potential threats to the validity of our assessment?

Table 5: Shifts of Physician 10

shift_end	observations	share
04:00:00	36	0.2278481
13:00:00	35	0.2215190
17:00:00	19	0.1202532
20:00:00	37	0.2341772
22:00:00	31	0.1962025

Let's define shifts that end at 4 am as empty shifts. If we look at the shifts of Physician 10, we can find that nearly 22.8% of the shifts are empty shifts, where the average is only about 12.1%. Thus the distribution of shifts could be one of the major threats to our assessment, so is patient severity.

Table 6: Shifts of all Physicians

shift_end	observations	share
04:00:00	1063	0.1212640
10:00:00	4	0.0004563
12:00:00	1512	0.1724846
13:00:00	1040	0.1186402
16:00:00	364	0.0415241
17:00:00	1746	0.1991786
20:00:00	1026	0.1170431
22:00:00	2011	0.2294091

Attempt:

Two new indicators are defined: `empty_shift` is true if the shift is an empty shift. `weekend` is true if the shift date is weekend.

Table 7: Physician effects to the log stay time

	log(diff)
phys_id	0.001*
	(0.001)
pred_lnlos	0.914***
	(0.020)
empty_shift	0.066***
	(0.023)
weekend	−0.020
	(0.017)
Constant	8.161***
	(0.028)
Observations	8,766
Residual Std. Error	0.640 (df = 8761)
Note:	*p<0.1; **p<0.05; ***p<0.01