

# Untitled

Jyun-Yu Cheng, Haokun Zhang

2023-02-15

## Question 1: Saratoga house prices

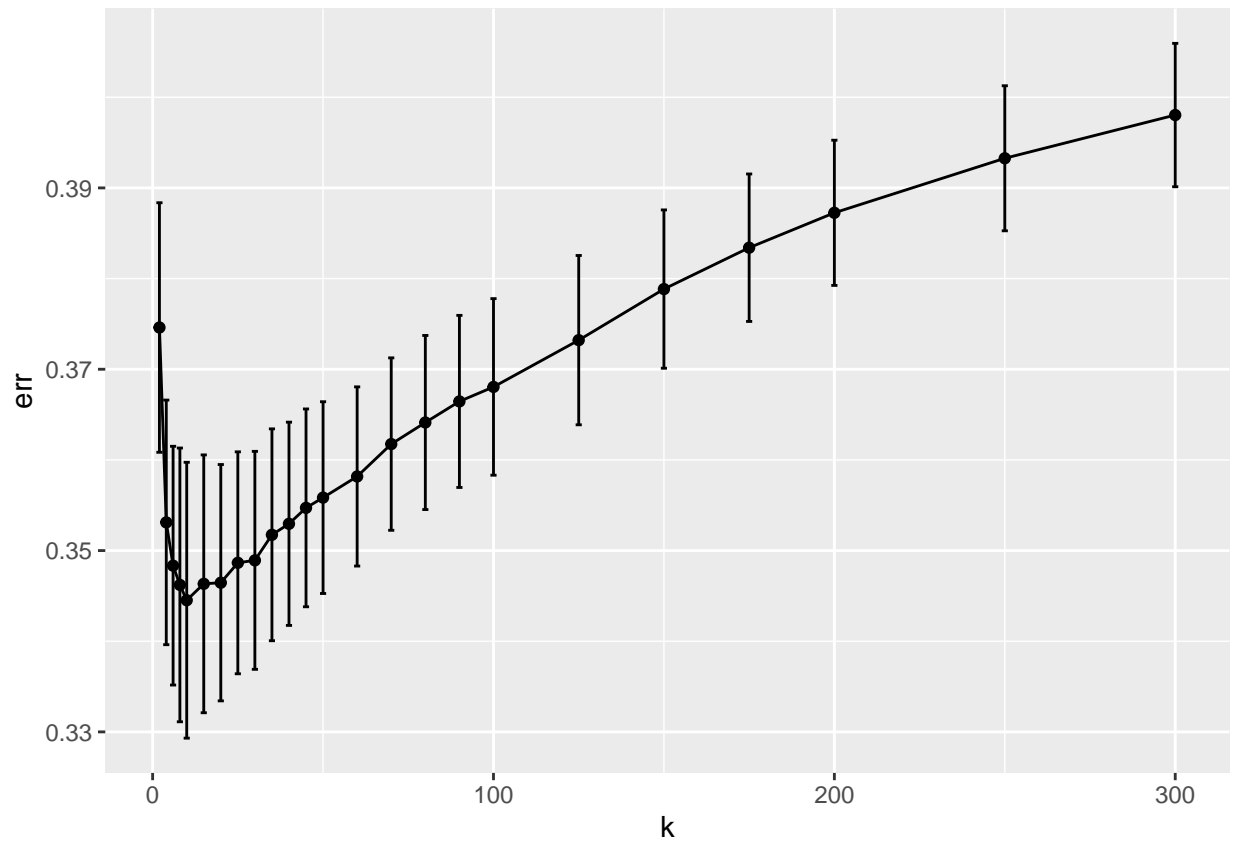
```
## [1] 66994.62
```

```
## [1] 64682.9
```

```
## [1] 64277.16
```

Since `lm3` has the least `rmse`, `lm3` is the best model among these above 3 `lm` models

```
##          k      err   std_err
## result.1  2 0.3746034 0.01376174
## result.2  4 0.3531099 0.01348960
## result.3  6 0.3483347 0.01316651
## result.4  8 0.3462131 0.01509775
## result.5 10 0.3445240 0.01521493
## result.6 15 0.3463334 0.01422008
```



## We can get the best k according to the following:

## [1] 10

Then we calculate the knn20 method RMSE

## [1] 0.3189704

Then averaging the estimate of out-of-sample RMSE over many different random train/test splits for the linear models, either randomly or by cross-validation.

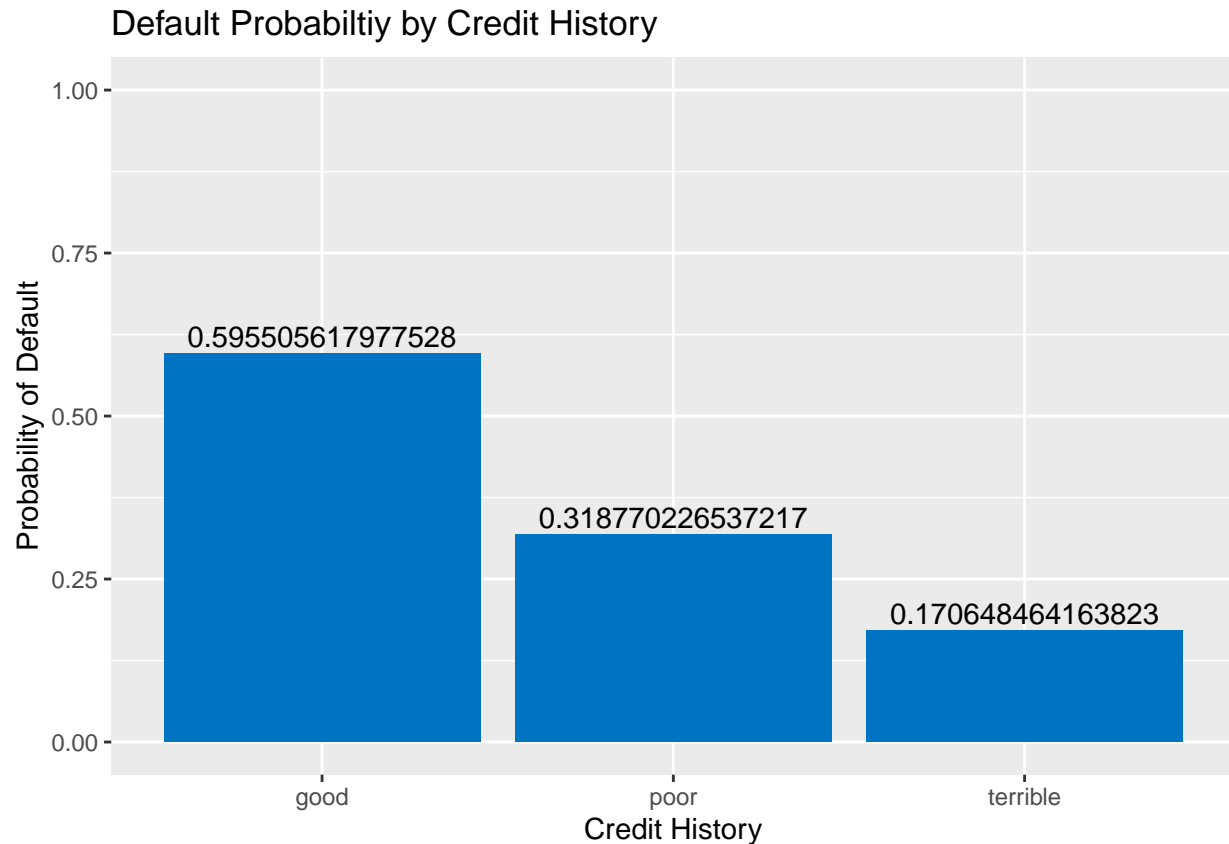
## [1] 62099.20 63608.94 58876.78

## [1] 0.3443018

## [1] 0.01226122

Since the error from knn model is smaller than that in lm3, I think knn model does a better prediction in the price of house in NY.

## Question 2: Classification and retrospective sampling



```
##
## Call:
## glm(formula = Default ~ +duration + amount + installment + age +
##      history + purpose + foreign, family = "binomial", data = credit_training)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1876  -0.8121  -0.5773   1.0454   2.4894
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -7.963e-01  5.171e-01  -1.540  0.12359
## duration       2.741e-02  9.283e-03   2.953  0.00315 **
## amount        7.758e-05  4.250e-05   1.825  0.06794 .
## installment    2.603e-01  8.625e-02   3.018  0.00255 **
## age           -2.455e-02  8.086e-03  -3.037  0.00239 **
## historypoor    -8.580e-01  2.732e-01  -3.141  0.00169 **
## historyterrible -1.665e+00  3.134e-01  -5.313 1.08e-07 ***
## purposeedu     7.783e-01  4.038e-01   1.928  0.05392 .
## purposegoods/repair -4.268e-02  2.844e-01  -0.150  0.88071
```

```
## purposenewcar      6.762e-01  3.093e-01  2.187  0.02876 *
## purposeusedcar    -5.355e-01  3.960e-01 -1.352  0.17630
## foreigngerman     -1.225e+00  6.353e-01 -1.928  0.05385 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 972.25  on 799  degrees of freedom
## Residual deviance: 859.53  on 788  degrees of freedom
## AIC: 883.53
##
## Number of Fisher Scoring iterations: 4
```

##	(Intercept)	duration	amount	installment
##	-0.80	0.03	0.00	0.26
##	age	historypoor	historyterrible	purposeedu
##	-0.02	-0.86	-1.66	0.78
##	purposegoods/repair	purposenewcar	purposeusedcar	foreigngerman
##	-0.04	0.68	-0.54	-1.22

```
## yhat
## y      0      1
## 0 127  10
## 1  41  22
```

## The coef for having poor and terrible credit history is -1.11 and -1.88. They are statistically significant. Having a poor or terrible credit history multiplies odds of default by 0.33 and 0.15, which counter-intuitively have negative effect on default. The retrospective, “case-control” method that the bank used to select samples resulted in a substantial oversampling of defaults, relative to a random sample of loans in the bank’s overall portfolio. If the purpose of the model is to screen prospective borrowers to classify them into “high” versus “low” probability of default, this data set is not appropriate for building a predictive model.

```
##
## 0 1
## 700 300
```

To improve the sampling scheme under limited resources to get the data on all subjects in the loans, it requires a method to reduce the oversampling of defaults. Previously, the set of non-defaulted loans (“controls”) is matched with similar set of defaulted loans, leaving “controls” less independent of the set of defaulted loans, the “controls” are not representative of the source population that produced the default cases either. Selection bias resulted. A better sampling scheme should adhere to fixing these two issues. A random sampling method will give defaulted loans and non-defaulted loans equal chance of being selected, and the “controls” is selected independently from the cases.

The second step is to increase the ratio between non-defaulted loan and defaulted loan. Since defaulted loans are rare cases and non-defaulted loans are plentiful, the statistical power of the study can be increased by enrolling more non-defaulted loans than the defaulted. However, ratio more than 4 controls over 1 case adds little impact on power. Therefore, if it is time-consuming or expensive to collect sets of controls, the ratio should not exceed 4:1. Alternatively, if it is not expensive, there is no reason to limit the number of non-defaulted loans. In this case of 300 defaulted loans, the number of non-defaulted loans can be increased from 700 to 1200 or even more. With these two steps, the sampling scheme will be improved.

### Question 3: Children and hotel reservations

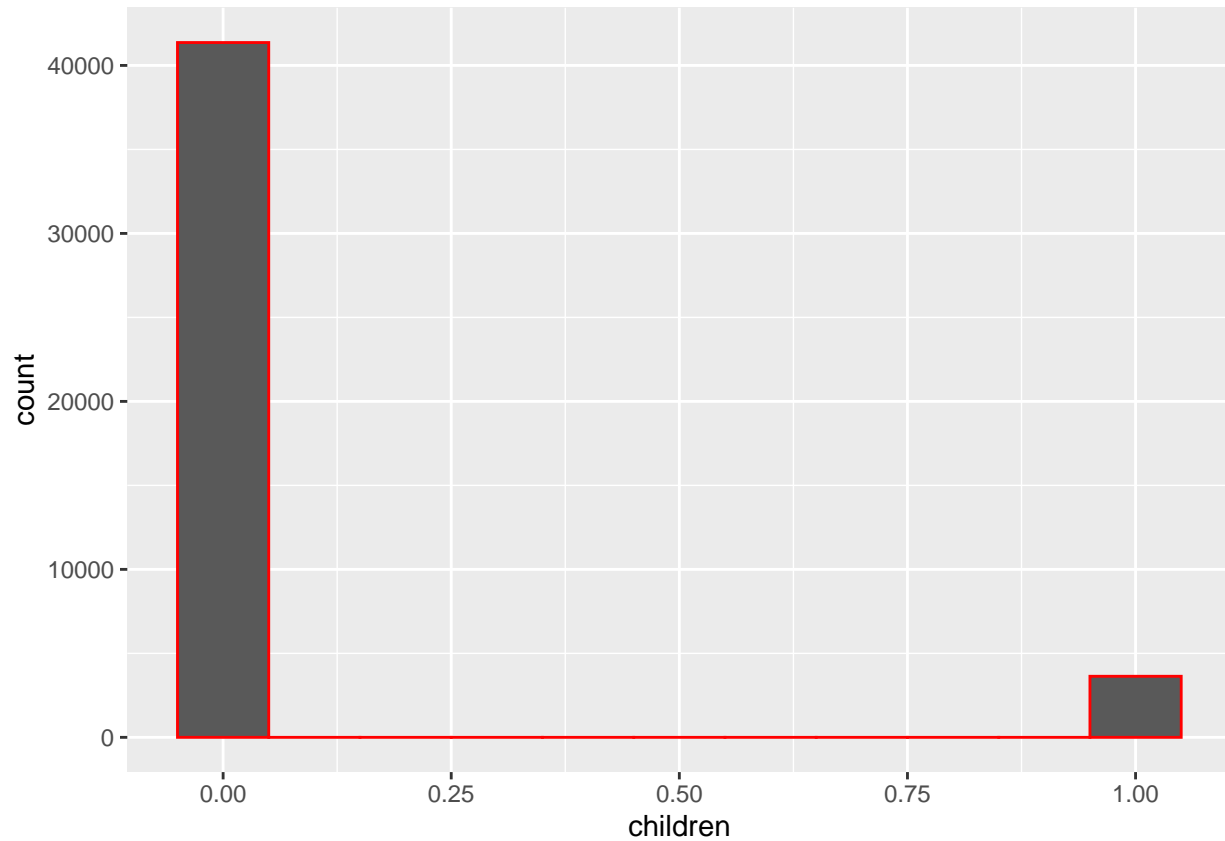
```
##          hotel lead_time stays_in_weekend_nights stays_in_week_nights adults
## 1   City_Hotel      217                1                3            2
## 2   City_Hotel        2                0                1            2
## 3 Resort_Hotel      95                2                5            2
## 4 Resort_Hotel     143                2                6            2
## 5 Resort_Hotel     136                1                4            2
## 6   City_Hotel      67                2                2            2
##  children meal market_segment distribution_channel is_repeated_guest
## 1         0   BB  Offline_TA/TO          TA/TO            0
## 2         0   BB       Direct          Direct            0
## 3         0   BB   Online_TA          TA/TO            0
## 4         0   HB   Online_TA          TA/TO            0
## 5         0   HB       Direct          Direct            0
## 6         0   SC   Online_TA          TA/TO            0
## previous_cancellations previous_bookings_not_canceled reserved_room_type
## 1                   0                   0                   A
## 2                   0                   0                   D
## 3                   0                   0                   A
## 4                   0                   0                   A
## 5                   0                   0                   F
## 6                   0                   0                   A
## assigned_room_type booking_changes deposit_type days_in_waiting_list
## 1                 A              0   No_Deposit            0
## 2                 K              0   No_Deposit            0
## 3                 A              2   No_Deposit            0
## 4                 A              0   No_Deposit            0
## 5                 F              0   No_Deposit            0
```

## 6	A	0	No_Deposit	0
##	customer_type	average_daily_rate	required_car_parking_spaces	
## 1	Transient-Party	80.75	none	
## 2	Transient	170.00	none	
## 3	Transient	8.00	none	
## 4	Transient	81.00	none	
## 5	Transient	157.60	none	
## 6	Transient	49.09	none	
##	total_of_special_requests	arrival_date		
## 1	1	2016-09-01		
## 2	3	2017-08-25		
## 3	2	2016-11-19		
## 4	1	2016-04-26		
## 5	4	2016-12-28		
## 6	1	2016-03-13		

##	hotel	lead_time	stays_in_weekend_nights	stays_in_week_nights	adults
## 1	Resort_Hotel	47	0	2	2
## 2	Resort_Hotel	46	0	2	2
## 3	City_Hotel	22	1	2	2
## 4	Resort_Hotel	209	2	5	2
## 5	City_Hotel	1	1	0	2
## 6	Resort_Hotel	171	2	5	2
##	children meal	market_segment	distribution_channel	is_repeated_guest	
## 1	1	BB	Direct	Direct	0
## 2	0	BB	Offline_TA/TO	TA/TO	0
## 3	0	BB	Online_TA	TA/TO	0
## 4	0	BB	Offline_TA/TO	TA/TO	0
## 5	0	SC	Online_TA	TA/TO	0
## 6	0	BB	Online_TA	TA/TO	0
##	previous_cancellations	previous_bookings_not_canceled	reserved_room_type		
## 1	0	0	C		
## 2	0	0	D		
## 3	0	0	D		
## 4	0	0	A		
## 5	0	0	A		
## 6	0	0	D		
##	assigned_room_type	booking_changes	deposit_type	days_in_waiting_list	
## 1	C	0	No_Deposit	0	
## 2	D	0	No_Deposit	0	
## 3	D	0	No_Deposit	0	
## 4	A	0	No_Deposit	0	
## 5	B	0	No_Deposit	0	
## 6	D	2	No_Deposit	0	
##	customer_type	average_daily_rate	required_car_parking_spaces		
## 1	Transient	289.00	none		
## 2	Transient	162.00	none		
## 3	Transient	121.33	none		
## 4	Transient	76.22	none		
## 5	Transient	98.00	none		
## 6	Transient-Party	182.86	none		
##	total_of_special_requests	arrival_date			
## 1	1	2017-08-23			
## 2	0	2016-12-30			

```
## 3          0  2017-03-13
## 4          0  2016-07-12
## 5          0  2016-08-07
## 6          1  2017-07-25
```



```
## fold 1,2,3,4,5,6,7,8,9,10,done.
```

```
## fold 1,2,3,4,5,6,7,8,9,10,done.
```

```
## 49 x 1 sparse Matrix of class "dgCMatrix"
##                               seg100
## intercept                    -4.4072735757
## hotelCity_Hotel                0.6619203886
## hotelResort_Hotel             -0.0009870392
## lead_time                      0.0005289516
## stays_in_weekend_nights        0.0354479056
## stays_in_week_nights          -0.0129234960
## adults                        -0.4684712563
## mealFB                        0.5604300676
## mealHB                        .
## mealSC                       -1.1082568362
## mealUndefined                  .
## market_segmentComplementary    0.2032275442
## market_segmentCorporate       -0.8699828146
## market_segmentDirect          .
```

```

## market_segmentGroups -1.1663374328
## market_segmentOffline_TA/T0 .
## market_segmentOnline_TA 0.0921315014
## distribution_channelDirect 0.1562537095
## distribution_channelGDS -1.5074854409
## distribution_channelTA/T0 .
## is_repeated_guest -0.7550961312
## previous_cancellations .
## previous_bookings_not_canceled -0.1396758073
## reserved_room_typeB 1.5989452595
## reserved_room_typeC 3.0119903865
## reserved_room_typeD -0.9918203415
## reserved_room_typeE .
## reserved_room_typeF 1.6625196302
## reserved_room_typeG 2.2748364708
## reserved_room_typeH 3.0992057797
## assigned_room_typeB 0.2134554211
## assigned_room_typeC 1.2848841992
## assigned_room_typeD 0.9488169705
## assigned_room_typeE 0.4539092747
## assigned_room_typeF 0.7129855276
## assigned_room_typeG 1.0131031814
## assigned_room_typeH 1.0513173398
## assigned_room_typeI 0.6516070930
## assigned_room_typeK 0.0928649665
## booking_changes 0.2343923602
## deposit_typeNon_Refund .
## deposit_typeRefundable .
## days_in_waiting_list -0.0029817785
## customer_typeGroup -0.0369586916
## customer_typeTransient 0.3176525385
## customer_typeTransient-Party -0.3960273810
## average_daily_rate 0.0100132113
## required_car_parking_spacesparking 0.0508729330
## total_of_special_requests 0.4439487125

## strong_interaction_name
## 1 reserved_room_typeB:assigned_room_typeI
## 2 market_segmentOnline_TA:assigned_room_typeG
## 3 reserved_room_typeD:assigned_room_typeB
## 4 market_segmentOffline_TA/T0:reserved_room_typeH
## 5 hotelResort_Hotel:reserved_room_typeB
## 6 market_segmentComplementary:reserved_room_typeF
## 7 mealHB:reserved_room_typeF
## 8 reserved_room_typeB:deposit_typeRefundable
## 9 mealUndefined:reserved_room_typeB
## 10 mealUndefined:assigned_room_typeD
## 11 mealSC:reserved_room_typeG
## 12 market_segmentOnline_TA:reserved_room_typeB
## 13 market_segmentComplementary:reserved_room_typeB
## 14 mealUndefined:assigned_room_typeG
## 15 reserved_room_typeB:assigned_room_typeG
## 16 is_repeated_guest:previous_bookings_not_canceled
## 17 reserved_room_typeE

```



```

## 18          hotelResort_Hotel:reserved_room_typeE
## 19          market_segmentDirect:reserved_room_typeB
## 20          market_segmentDirect:reserved_room_typeF
## 21          mealSC:reserved_room_typeF
## 22          reserved_room_typeH:assigned_room_typeD
## 23          market_segmentComplementary:reserved_room_typeC
## 24          market_segmentDirect:reserved_room_typeC
## 25          adults:previous_bookings_not_canceled
## 26          reserved_room_typeF:assigned_room_typeE
## 27          mealFB:assigned_room_typeK
## 28          reserved_room_typeG
## 29          reserved_room_typeF
## 30 market_segmentOnline_TA:previous_bookings_not_canceled
##      strong_interaction_beta abs_beta
## 1      4.67216577908299 4.672166
## 2      4.41374530734678 4.413745
## 3      4.28773253554402 4.287733
## 4      3.67245168706815 3.672452
## 5      3.67029476164202 3.670295
## 6      3.53898574572693 3.538986
## 7      3.30780968768956 3.307810
## 8      -3.05886640447598 3.058866
## 9      -3.04853580543665 3.048536
## 10     -3.03071781168826 3.030718
## 11     3.01345946115791 3.013459
## 12     -2.99685987353239 2.996860
## 13     2.9389174337881 2.938917
## 14     2.88893672318108 2.888937
## 15     2.69655506294253 2.696555
## 16     2.61587218243399 2.615872
## 17     2.39417758404882 2.394178
## 18     -2.3530034629503 2.353003
## 19     2.23526104059518 2.235261
## 20     2.09337309638563 2.093373
## 21     -2.08308963292592 2.083090
## 22     2.08259403708136 2.082594
## 23     2.05494170106681 2.054942
## 24     2.03025991940894 2.030260
## 25     -2.00527373152582 2.005274
## 26     1.94070350310936 1.940704
## 27     1.89409072282823 1.894091
## 28     1.88048455934398 1.880485
## 29     1.74546888484504 1.745469
## 30     1.71428881174543 1.714289

```

```

##          (Intercept)
##          -2.664833e+15
##          hotelResort_Hotel
##          -1.565505e+15
##          lead_time
##          2.061995e+12
##          stays_in_weekend_nights
##          9.427233e+13
##          stays_in_week_nights

```

```

##          -1.126774e+13
##          adults
##          -6.175660e+14
##          mealFB
##          2.204488e+15
##          mealHB
##          1.686359e+14
##          mealSC
##          -1.637716e+15
##          mealUndefined
##          2.219403e+15
##          market_segmentComplementary
##          2.790663e+15
##          market_segmentCorporate
##          3.615525e+15
##          market_segmentDirect
##          3.242009e+15
##          market_segmentGroups
##          1.388761e+15
##          market_segmentOffline_TA/T0
##          1.405347e+15
##          market_segmentOnline_TA
##          1.429228e+15
##          distribution_channelDirect
##          -5.212995e+14
##          distribution_channelGDS
##          -1.320679e+15
##          distribution_channelTA/T0
##          -1.051189e+15
##          is_repeated_guest
##          3.540267e+13
##          previous_cancellations
##          -3.640867e+14
##          previous_bookings_not_canceled
##          -4.118475e+13
##          reserved_room_typeB
##          3.830438e+14
##          reserved_room_typeC
##          -3.283321e+15
##          reserved_room_typeD
##          -1.168739e+15
##          reserved_room_typeE
##          2.728071e+15
##          reserved_room_typeF
##          2.135618e+15
##          reserved_room_typeG
##          -2.078262e+15
##          reserved_room_typeH
##          6.870820e+15
##          assigned_room_typeB
##          2.421047e+14
##          assigned_room_typeC
##          1.251675e+15
##          assigned_room_typeD

```

```

##          8.535043e+14
##      assigned_room_typeE
##          4.097027e+14
##      assigned_room_typeF
##          9.942336e+14
##      assigned_room_typeG
##          3.978732e+13
##      assigned_room_typeH
##          1.877007e+15
##      assigned_room_typeI
##          7.955229e+14
##      assigned_room_typeK
##          3.657954e+13
##      booking_changes
##          1.620566e+14
##      days_in_waiting_list
##          7.155654e+11
##      customer_typeGroup
##          9.675857e+14
##      customer_typeTransient
##          -1.909129e+15
##      customer_typeTransient-Party
##          6.577547e+14
##      average_daily_rate
##          7.282773e+12
##      required_car_parking_spacesparking
##          7.656592e+13
##      total_of_special_requests
##          3.972059e+14
##      mealFB:reserved_room_typeB
##          NA
##      mealHB:reserved_room_typeB
##          -4.160412e+15
##      mealSC:reserved_room_typeB
##          8.093419e+14
##      mealUndefined:reserved_room_typeB
##          NA
##      mealFB:reserved_room_typeC
##          -3.890726e+15
##      mealHB:reserved_room_typeC
##          -5.934811e+14
##      mealSC:reserved_room_typeC
##          3.362880e+15
##      mealUndefined:reserved_room_typeC
##          -3.171330e+15
##      mealFB:reserved_room_typeD
##          -2.041008e+15
##      mealHB:reserved_room_typeD
##          -2.555915e+14
##      mealSC:reserved_room_typeD
##          1.927008e+15
##      mealUndefined:reserved_room_typeD
##          -1.964990e+15
##      mealFB:reserved_room_typeE

```

```

## -3.281631e+15
## mealHB:reserved_room_typeE
## 1.196765e+14
## mealSC:reserved_room_typeE
## 2.442227e+14
## mealUndefined:reserved_room_typeE
## -1.452054e+15
## mealFB:reserved_room_typeF
## -4.833046e+15
## mealHB:reserved_room_typeF
## -4.937148e+14
## mealSC:reserved_room_typeF
## 3.505354e+15
## mealUndefined:reserved_room_typeF
## -3.440014e+15
## mealFB:reserved_room_typeG
## 4.580695e+15
## mealHB:reserved_room_typeG
## 7.433196e+13
## mealSC:reserved_room_typeG
## 1.294847e+15
## mealUndefined:reserved_room_typeG
## 2.834695e+15
## mealFB:reserved_room_typeH
## NA
## mealHB:reserved_room_typeH
## -2.875845e+14
## mealSC:reserved_room_typeH
## NA
## mealUndefined:reserved_room_typeH
## NA
## reserved_room_typeB:assigned_room_typeB
## 1.890661e+14
## reserved_room_typeC:assigned_room_typeB
## NA
## reserved_room_typeD:assigned_room_typeB
## -9.383219e+14
## reserved_room_typeE:assigned_room_typeB
## 6.223799e+15
## reserved_room_typeF:assigned_room_typeB
## 9.378604e+14
## reserved_room_typeG:assigned_room_typeB
## 7.417992e+15
## reserved_room_typeH:assigned_room_typeB
## NA
## reserved_room_typeB:assigned_room_typeC
## NA
## reserved_room_typeC:assigned_room_typeC
## 4.957232e+13
## reserved_room_typeD:assigned_room_typeC
## 1.368407e+15
## reserved_room_typeE:assigned_room_typeC
## -4.396179e+14
## reserved_room_typeF:assigned_room_typeC

```

```

##                                     NA
## reserved_room_typeG:assigned_room_typeC
##                                     2.569873e+15
## reserved_room_typeH:assigned_room_typeC
##                                     NA
## reserved_room_typeB:assigned_room_typeD
##                                     -9.635780e+14
## reserved_room_typeC:assigned_room_typeD
##                                     5.975163e+14
## reserved_room_typeD:assigned_room_typeD
##                                     5.579562e+14
## reserved_room_typeE:assigned_room_typeD
##                                     2.798638e+14
## reserved_room_typeF:assigned_room_typeD
##                                     NA
## reserved_room_typeG:assigned_room_typeD
##                                     NA
## reserved_room_typeH:assigned_room_typeD
##                                     -8.231169e+15
## reserved_room_typeB:assigned_room_typeE
##                                     4.597066e+14
## reserved_room_typeC:assigned_room_typeE
##                                     1.001841e+15
## reserved_room_typeD:assigned_room_typeE
##                                     1.027062e+15
## reserved_room_typeE:assigned_room_typeE
##                                     4.477376e+14
## reserved_room_typeF:assigned_room_typeE
##                                     -1.886082e+15
## reserved_room_typeG:assigned_room_typeE
##                                     6.211939e+15
## reserved_room_typeH:assigned_room_typeE
##                                     NA
## reserved_room_typeB:assigned_room_typeF
##                                     2.619371e+15
## reserved_room_typeC:assigned_room_typeF
##                                     NA
## reserved_room_typeD:assigned_room_typeF
##                                     5.960715e+14
## reserved_room_typeE:assigned_room_typeF
##                                     -3.043935e+14
## reserved_room_typeF:assigned_room_typeF
##                                     -1.869456e+15
## reserved_room_typeG:assigned_room_typeF
##                                     1.614420e+15
## reserved_room_typeH:assigned_room_typeF
##                                     NA
## reserved_room_typeB:assigned_room_typeG
##                                     2.264980e+15
## reserved_room_typeC:assigned_room_typeG
##                                     5.353233e+15
## reserved_room_typeD:assigned_room_typeG
##                                     5.839864e+14
## reserved_room_typeE:assigned_room_typeG

```

```

##                                8.859179e+14
##    reserved_room_typeF:assigned_room_typeG
##                                -8.785389e+14
##    reserved_room_typeG:assigned_room_typeG
##                                2.967371e+15
##    reserved_room_typeH:assigned_room_typeG
##                                -1.251870e+15
##    reserved_room_typeB:assigned_room_typeH
##                                NA
##    reserved_room_typeC:assigned_room_typeH
##                                9.584243e+14
##    reserved_room_typeD:assigned_room_typeH
##                                -9.244121e+14
##    reserved_room_typeE:assigned_room_typeH
##                                -3.887790e+15
##    reserved_room_typeF:assigned_room_typeH
##                                -6.001326e+15
##    reserved_room_typeG:assigned_room_typeH
##                                -9.278101e+12
##    reserved_room_typeH:assigned_room_typeH
##                                -4.240702e+15
##    reserved_room_typeB:assigned_room_typeI
##                                NA
##    reserved_room_typeC:assigned_room_typeI
##                                1.794075e+15
##    reserved_room_typeD:assigned_room_typeI
##                                8.439337e+14
##    reserved_room_typeE:assigned_room_typeI
##                                -8.817335e+14
##    reserved_room_typeF:assigned_room_typeI
##                                -1.728613e+15
##    reserved_room_typeG:assigned_room_typeI
##                                1.779065e+15
##    reserved_room_typeH:assigned_room_typeI
##                                NA
##    reserved_room_typeB:assigned_room_typeK
##                                -1.251350e+15
##    reserved_room_typeC:assigned_room_typeK
##                                NA
##    reserved_room_typeD:assigned_room_typeK
##                                -1.478985e+15
##    reserved_room_typeE:assigned_room_typeK
##                                -9.676576e+14
##    reserved_room_typeF:assigned_room_typeK
##                                -7.843752e+14
##    reserved_room_typeG:assigned_room_typeK
##                                8.353700e+14
##    reserved_room_typeH:assigned_room_typeK
##                                NA
##    hotelResort_Hotel:reserved_room_typeB
##                                2.653255e+15
##    hotelResort_Hotel:reserved_room_typeC
##                                4.889688e+15
##    hotelResort_Hotel:reserved_room_typeD

```

```

##                                1.582587e+15
##          hotelResort_Hotel:reserved_room_typeE
##                                -1.924889e+14
##          hotelResort_Hotel:reserved_room_typeF
##                                7.675983e+14
##          hotelResort_Hotel:reserved_room_typeG
##                                2.069165e+15
##          hotelResort_Hotel:reserved_room_typeH
##                                NA
## market_segmentComplementary:reserved_room_typeB
##                                5.984467e+14
##          market_segmentCorporate:reserved_room_typeB
##                                4.996704e+15
##          market_segmentDirect:reserved_room_typeB
##                                1.006134e+15
##          market_segmentGroups:reserved_room_typeB
##                                3.054308e+15
## market_segmentOffline_TA/TO:reserved_room_typeB
##                                1.162550e+14
##          market_segmentOnline_TA:reserved_room_typeB
##                                NA
## market_segmentComplementary:reserved_room_typeC
##                                -1.502124e+15
##          market_segmentCorporate:reserved_room_typeC
##                                1.163812e+15
##          market_segmentDirect:reserved_room_typeC
##                                1.093626e+15
##          market_segmentGroups:reserved_room_typeC
##                                2.953259e+15
## market_segmentOffline_TA/TO:reserved_room_typeC
##                                6.494507e+14
##          market_segmentOnline_TA:reserved_room_typeC
##                                NA
## market_segmentComplementary:reserved_room_typeD
##                                4.085767e+14
##          market_segmentCorporate:reserved_room_typeD
##                                -1.006103e+15
##          market_segmentDirect:reserved_room_typeD
##                                3.713477e+14
##          market_segmentGroups:reserved_room_typeD
##                                1.420930e+15
## market_segmentOffline_TA/TO:reserved_room_typeD
##                                5.145085e+12
##          market_segmentOnline_TA:reserved_room_typeD
##                                -2.038087e+14
## market_segmentComplementary:reserved_room_typeE
##                                -2.381563e+15
##          market_segmentCorporate:reserved_room_typeE
##                                -5.414092e+15
##          market_segmentDirect:reserved_room_typeE
##                                -1.822841e+15
##          market_segmentGroups:reserved_room_typeE
##                                -3.181045e+15
## market_segmentOffline_TA/TO:reserved_room_typeE

```

```

## -2.382786e+15
## market_segmentOnline_TA:reserved_room_typeE
## -3.630564e+15
## market_segmentComplementary:reserved_room_typeF
## -1.709018e+14
## market_segmentCorporate:reserved_room_typeF
## -2.804142e+14
## market_segmentDirect:reserved_room_typeF
## 4.753674e+14
## market_segmentGroups:reserved_room_typeF
## 1.902550e+15
## market_segmentOffline_TA/TO:reserved_room_typeF
## -3.611553e+14
## market_segmentOnline_TA:reserved_room_typeF
## NA
## market_segmentComplementary:reserved_room_typeG
## -6.590371e+14
## market_segmentCorporate:reserved_room_typeG
## -1.086727e+15
## market_segmentDirect:reserved_room_typeG
## -1.756681e+14
## market_segmentGroups:reserved_room_typeG
## -1.449117e+15
## market_segmentOffline_TA/TO:reserved_room_typeG
## -4.586353e+14
## market_segmentOnline_TA:reserved_room_typeG
## NA
## market_segmentComplementary:reserved_room_typeH
## NA
## market_segmentCorporate:reserved_room_typeH
## NA
## market_segmentDirect:reserved_room_typeH
## -4.011209e+14
## market_segmentGroups:reserved_room_typeH
## NA
## market_segmentOffline_TA/TO:reserved_room_typeH
## NA
## market_segmentOnline_TA:reserved_room_typeH
## NA
## mealFB:is_repeated_guest
## -4.288143e+15
## mealHB:is_repeated_guest
## -3.791467e+14
## mealSC:is_repeated_guest
## -7.659743e+12
## mealUndefined:is_repeated_guest
## 7.433266e+14
## adults:previous_bookings_not_canceled
## 7.470193e+12
## mealFB:previous_bookings_not_canceled
## -1.615756e+15
## mealHB:previous_bookings_not_canceled
## -5.613039e+13
## mealSC:previous_bookings_not_canceled

```



```

## -1.226196e+14
## mealUndefined:previous_bookings_not_canceled
## 1.674770e+14
## market_segmentComplementary:customer_typeGroup
## -3.784171e+15
## market_segmentCorporate:customer_typeGroup
## -3.171323e+15
## market_segmentDirect:customer_typeGroup
## -3.525703e+15
## market_segmentGroups:customer_typeGroup
## -6.190621e+14
## market_segmentOffline_TA/TO:customer_typeGroup
## -4.397259e+14
## market_segmentOnline_TA:customer_typeGroup
## NA
## market_segmentComplementary:customer_typeTransient
## 1.334188e+15
## market_segmentCorporate:customer_typeTransient
## 1.316617e+14
## market_segmentDirect:customer_typeTransient
## -3.182418e+14
## market_segmentGroups:customer_typeTransient
## 6.552677e+14
## market_segmentOffline_TA/TO:customer_typeTransient
## 2.992854e+15
## market_segmentOnline_TA:customer_typeTransient
## 2.746038e+15
## market_segmentComplementary:customer_typeTransient-Party
## -2.125420e+15
## market_segmentCorporate:customer_typeTransient-Party
## -3.024303e+15
## market_segmentDirect:customer_typeTransient-Party
## -3.250910e+15
## market_segmentGroups:customer_typeTransient-Party
## -2.019519e+15
## market_segmentOffline_TA/TO:customer_typeTransient-Party
## -4.828031e+14
## market_segmentOnline_TA:customer_typeTransient-Party
## NA
## is_repeated_guest:assigned_room_typeB
## -2.864685e+14
## is_repeated_guest:assigned_room_typeC
## -1.179412e+15
## is_repeated_guest:assigned_room_typeD
## -1.101011e+15
## is_repeated_guest:assigned_room_typeE
## -7.854344e+14
## is_repeated_guest:assigned_room_typeF
## -2.697365e+14
## is_repeated_guest:assigned_room_typeG
## 2.029264e+13
## is_repeated_guest:assigned_room_typeH
## -2.373260e+14
## is_repeated_guest:assigned_room_typeI

```

```

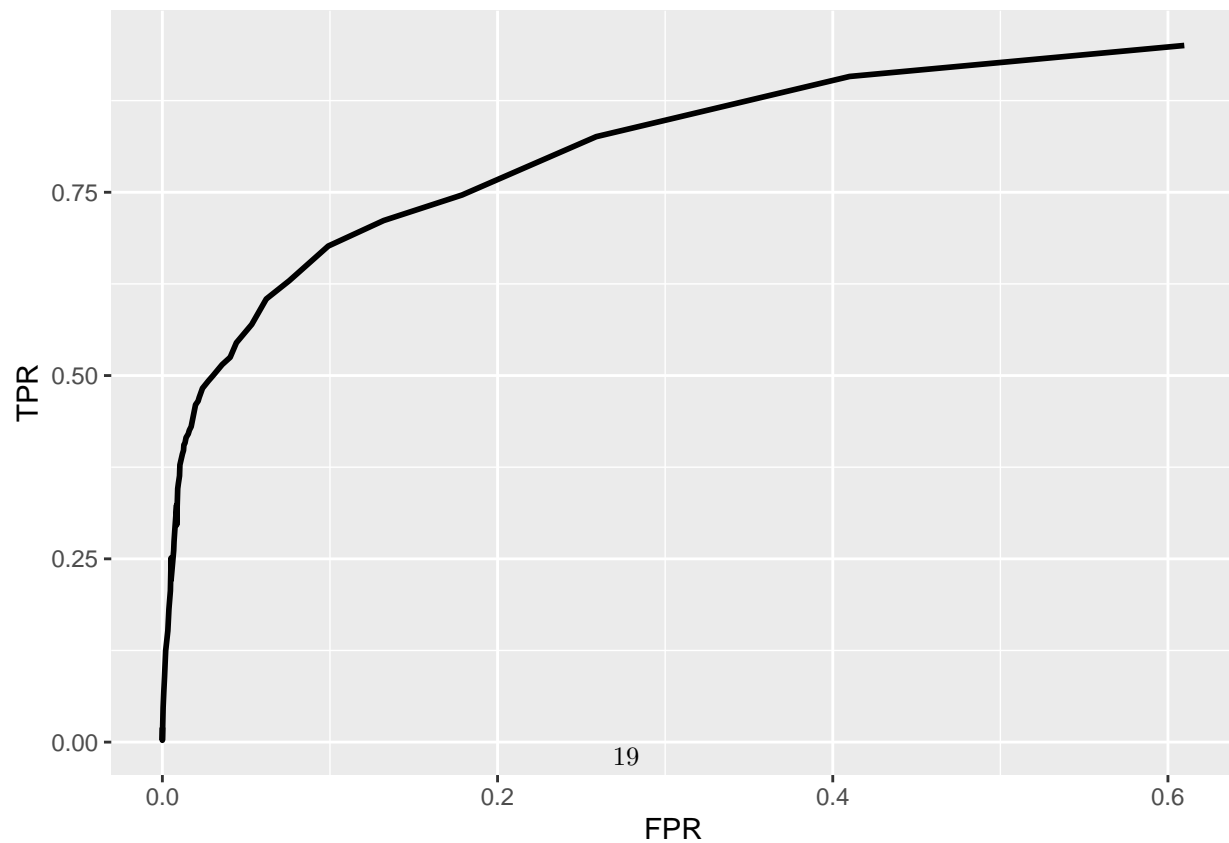
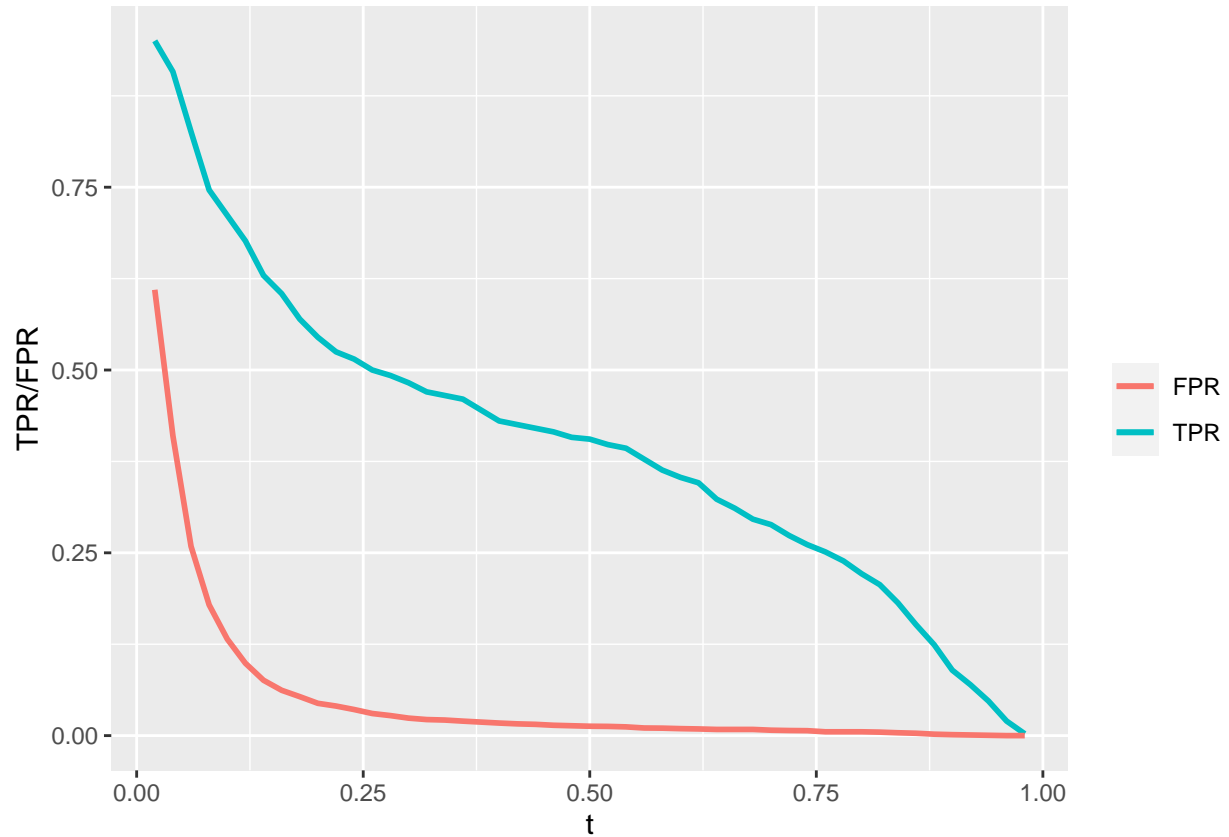
##                                     1.002340e+15
##                               is_repeated_guest:assigned_room_typeK
##                                     1.051444e+15
##   assigned_room_typeB:required_car_parking_spacesparking
##                                     -1.474422e+15
##   assigned_room_typeC:required_car_parking_spacesparking
##                                     -1.738664e+14
##   assigned_room_typeD:required_car_parking_spacesparking
##                                     -3.928812e+13
##   assigned_room_typeE:required_car_parking_spacesparking
##                                     1.983718e+14
##   assigned_room_typeF:required_car_parking_spacesparking
##                                     -1.775681e+13
##   assigned_room_typeG:required_car_parking_spacesparking
##                                     1.347968e+14
##   assigned_room_typeH:required_car_parking_spacesparking
##                                     -6.710884e+13
##   assigned_room_typeI:required_car_parking_spacesparking
##                                     -8.112400e+14
##   assigned_room_typeK:required_car_parking_spacesparking
##                                     9.806670e+14

##           [,1]      [,2]      [,3]      [,4]
## measurement    "Deviance" "TPR"    "FPR"    "FDR"
## eval_baseline1  "5027.604" "0"      "0"      "NaN"
## eval_baseline2  "3450.847" "0.358" "0.011" "0.262"
## eval_lasso_selected "3444.603" "0.383" "0.012" "0.254"

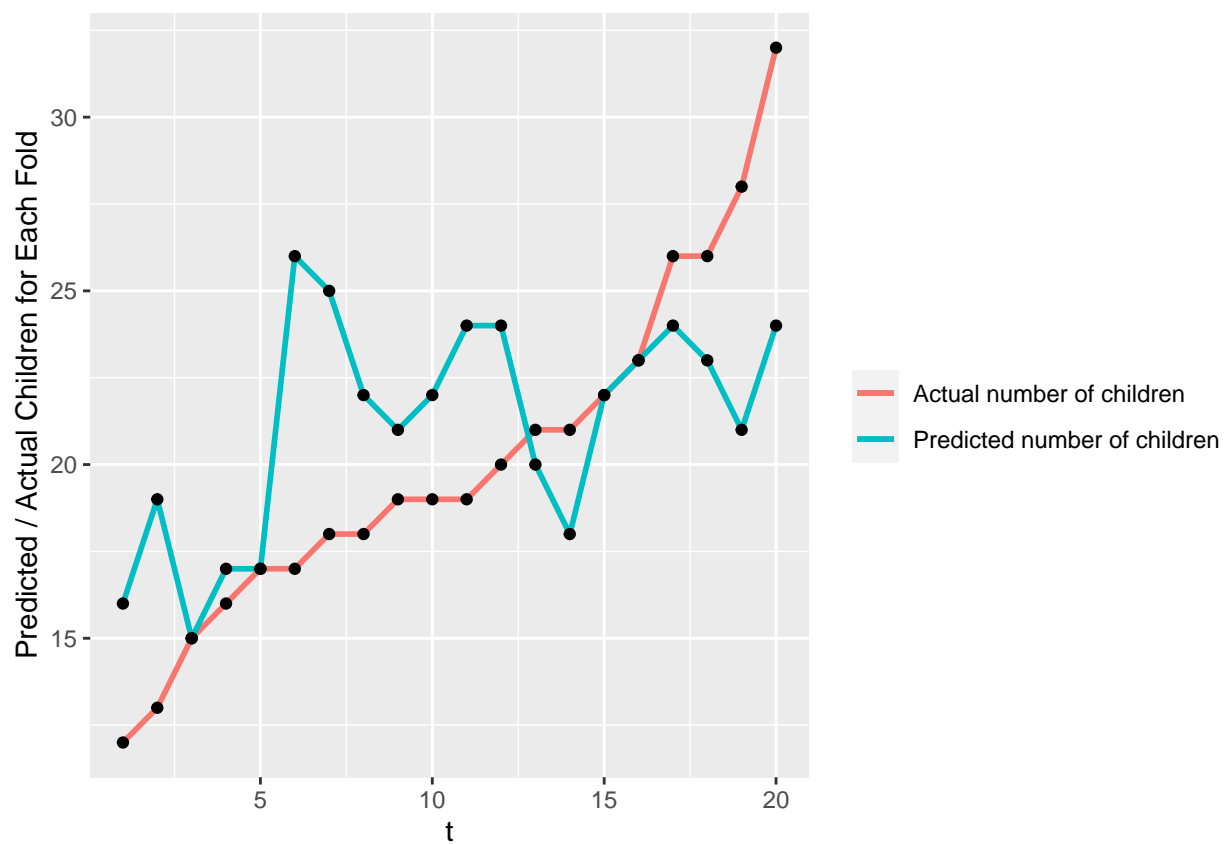
```

We can see that the lasso model has the smallest deviance, the highest TPR(higher is better)

and the lowest FPR & FDR (Both 2 are “lower is better”)



## The ROC curve of our best model



## We can see that our model doesn't perform well ## The prediction isn't accurate. The predict numbers wiggle more than actual number.