PREDICT VIRALITY – EXERCISE

Basic Statistics:

n = 24

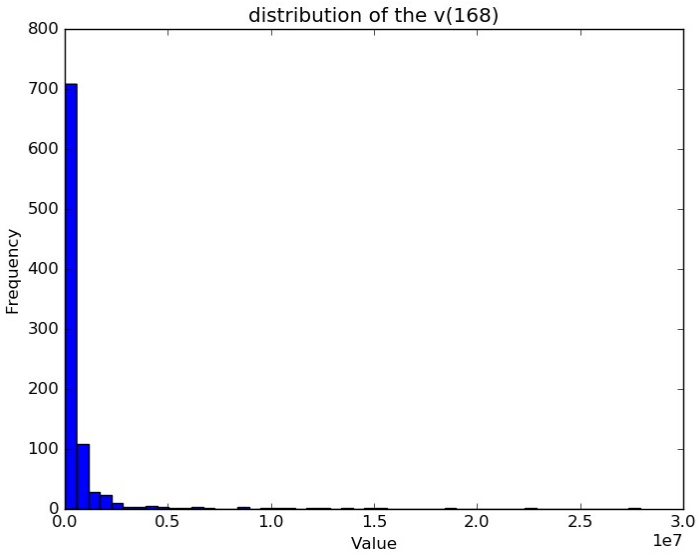
Mean value =	376765.517467
Median value =	194357.5
Standard deviation =	923142.428323
Range =	15263466
Kurtosis =	137.691960074

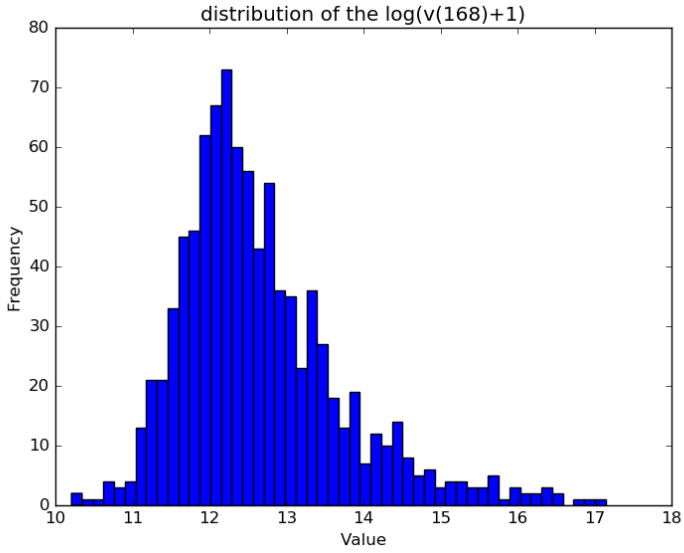
n = 72

Mean value =	613303.341703
Median value =	237417.5
Standard deviation =	1653066.35504
Range =	22890539
Kurtosis =	85.9188239299

n = 168

Mean value =	743209.837336
Median value =	252287.0
Standard deviation =	2006867.48178
Range =	27871098
Kurtosis =	70.5533867564

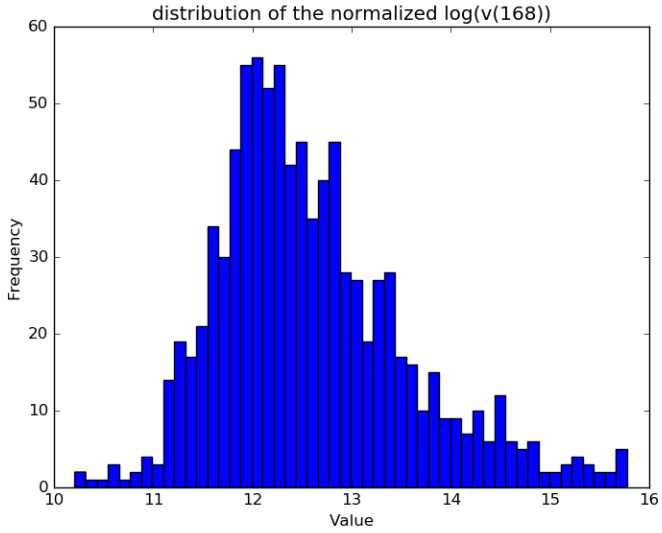




As we can see, distribution of v(168) is very irregular.
Distribution of log transformed looks more “Gaussian”.

Removing outsiders:

Mean value: 12.6547238564
Standard deviation: 1.06419725854
3-sigma values: [9.4621320808015525, 15.847315632044715]

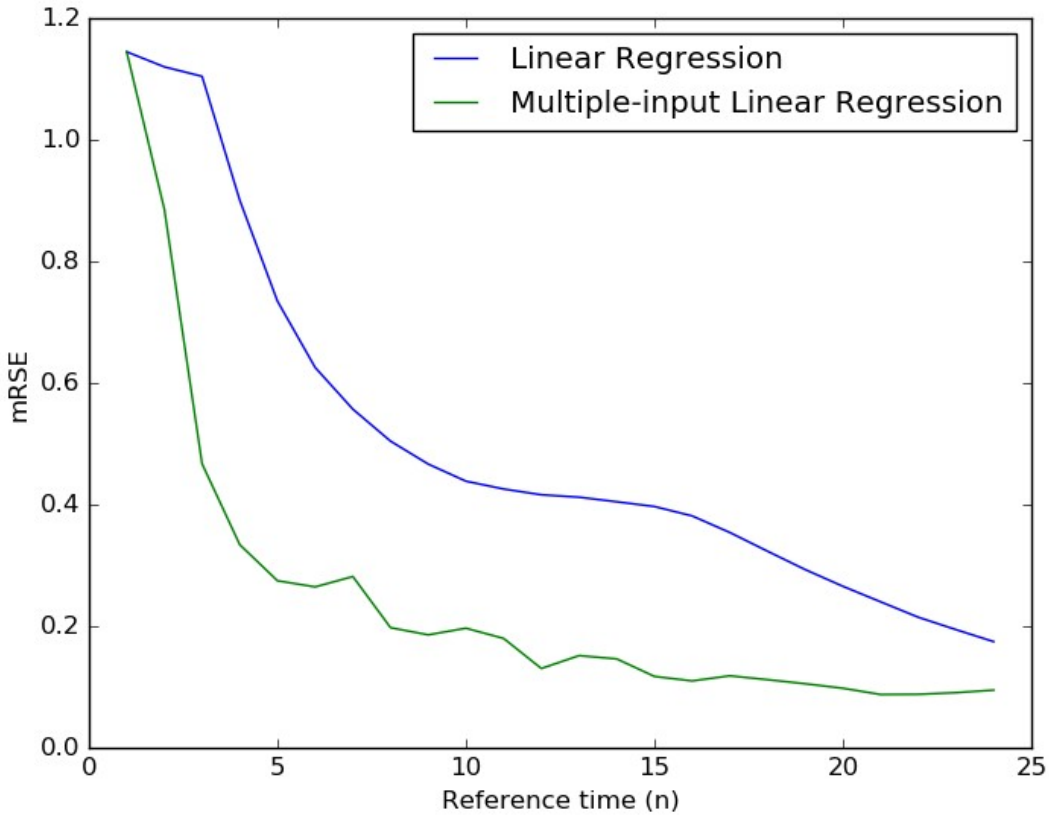


Correlation coefficients:

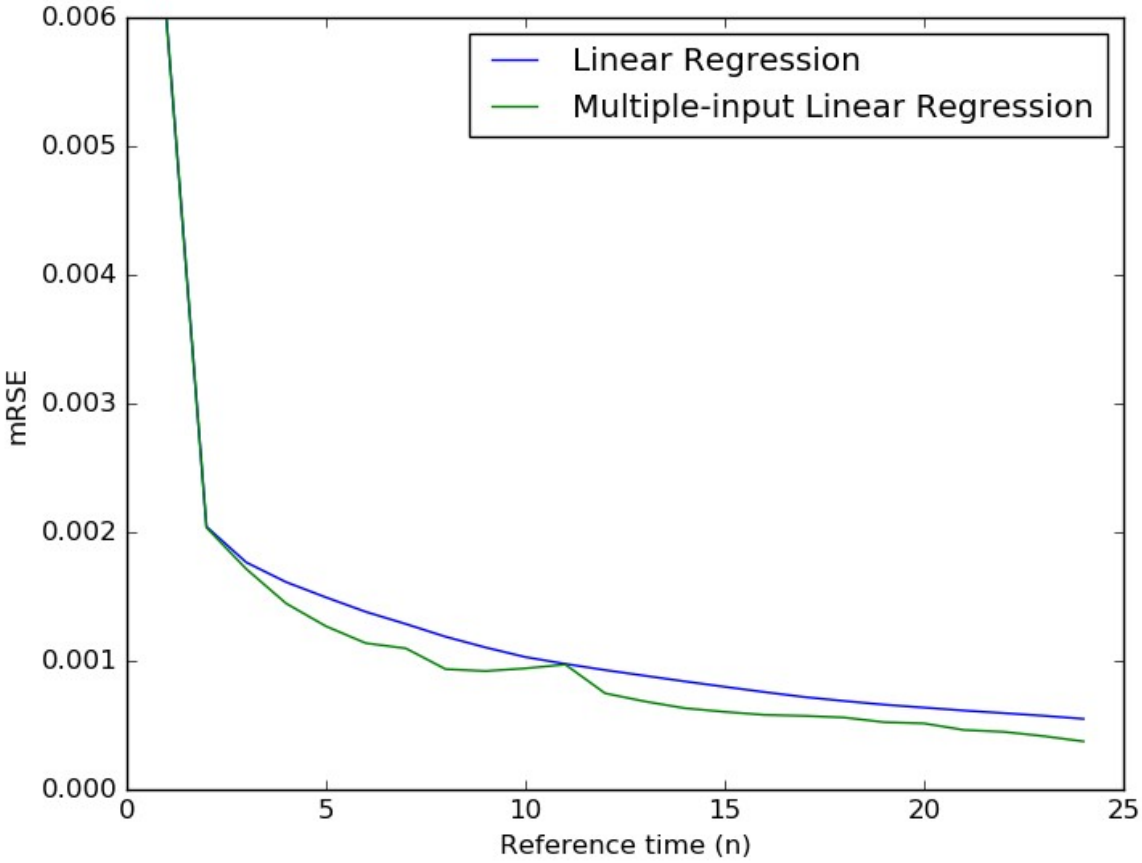
n = 1 : 0.254537927904
n = 2 : 0.784769091433
n = 3 : 0.839769859523
n = 4 : 0.855707613377
n = 5 : 0.868397449686
n = 6 : 0.880234466249
n = 7 : 0.890877743533
n = 8 : 0.901029037145
n = 9 : 0.909913787351
n = 10 : 0.916680721227
n = 11 : 0.921825217879
n = 12 : 0.926514229853
n = 13 : 0.930555346821
n = 14 : 0.934296861281
n = 15 : 0.937993613195
n = 16 : 0.941416070544
n = 17 : 0.944743309964
n = 18 : 0.947513604639
n = 19 : 0.949909681318
n = 20 : 0.952153630818
n = 21 : 0.954276447604
n = 22 : 0.956258069182
n = 23 : 0.958027343346
n = 24 : 0.959683838288

Plot the mRSE values for n ∈ (1, 24) computed on the test dataset:

- raw views:



- log transformed:


Linear Regression
Multiple-input Linear Regression
mRSE
Reference time (n)
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
5
10
15
20
25