Sleep Efficiency

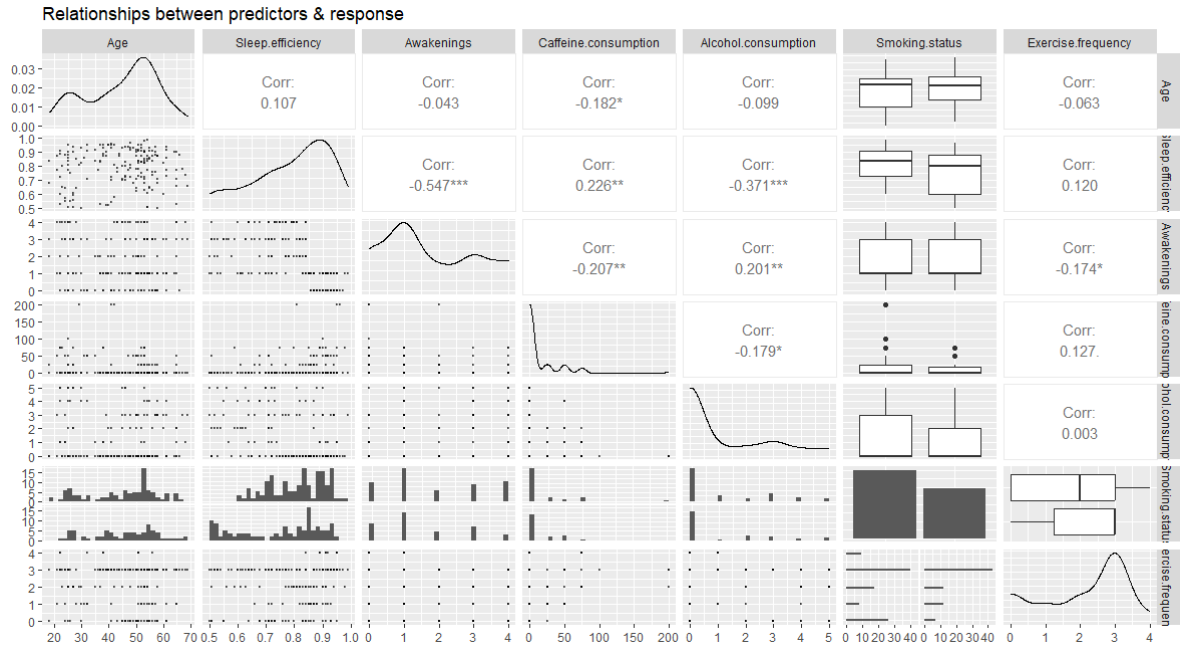
Andrea Violante

Introductory consideration

Dataset source: Kaggle

Dataset description: Sleep efficiency as a function of use and habits

Considerations for Conducting the Analysis: Filtering the Dataset



Qualitative analysis

Response correlations - covariates:

- Caffeine consumption: 0.225**
- Awakenings: - 0.543***
- Alcohol consumption: -0.371***

Correlations between covariates:

- Awakenings – caffeine: -0.207**
- Alcohol – caffeine: -0.209**

Summary of the model

1. Coefficient analysis
2. Variability explained

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.8864784  0.0344847  25.706  < 2e-16 ***
Age                  0.0008066  0.0005712   1.412    0.160
Awakenings          -0.0483596  0.0055173  -8.765 2.05e-15 ***
Caffeine.consumption 0.0001250  0.0002283   0.548    0.585
Alcohol.consumption -0.0220752  0.0045288  -4.874 2.52e-06 ***
Smoking.statusYes   -0.0901520  0.0147880  -6.096 7.29e-09 ***
Exercise.frequency   0.0069488  0.0058692   1.184    0.238
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09378 on 167 degrees of freedom
Multiple R-squared:  0.4951,    Adjusted R-squared:  0.4769
F-statistic: 27.29 on 6 and 167 DF,  p-value: < 2.2e-16
```

1. Analisi dei coefficienti

- Relevant coefficients: awakenings, alcohol consumption
- Coefficients not relevant: age, caffeine consumption, exercise frequency

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.8864784  0.0344847  25.706  < 2e-16 ***
Age                  0.0008066  0.0005712   1.412    0.160
Awakenings          -0.0483596  0.0055173  -8.765 2.05e-15 ***
Caffeine.consumption 0.0001250  0.0002283   0.548    0.585
Alcohol.consumption -0.0220752  0.0045288  -4.874 2.52e-06 ***
Smoking.statusYes   -0.0901520  0.0147880  -6.096 7.29e-09 ***
Exercise.frequency   0.0069488  0.0058692   1.184    0.238
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09378 on 167 degrees of freedom
Multiple R-squared:  0.4951,    Adjusted R-squared:  0.4769
F-statistic: 27.29 on 6 and 167 DF,  p-value: < 2.2e-16
```

2. Spiegazione della variabilità

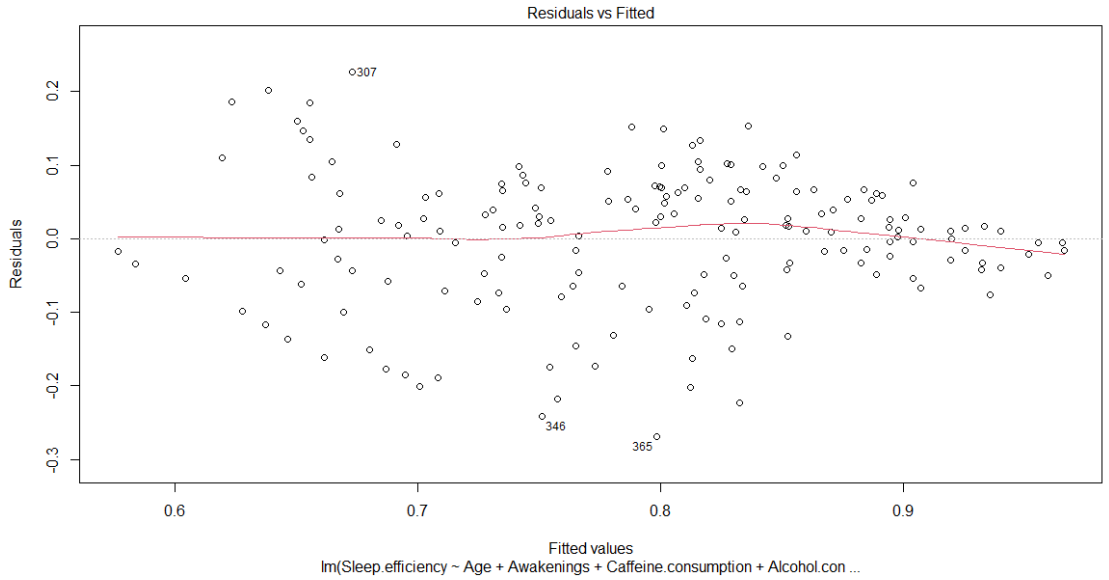
About 50% of the variability of the regression model data is well explained by covariates

Residue Verification

Homoskedasticity

Normality

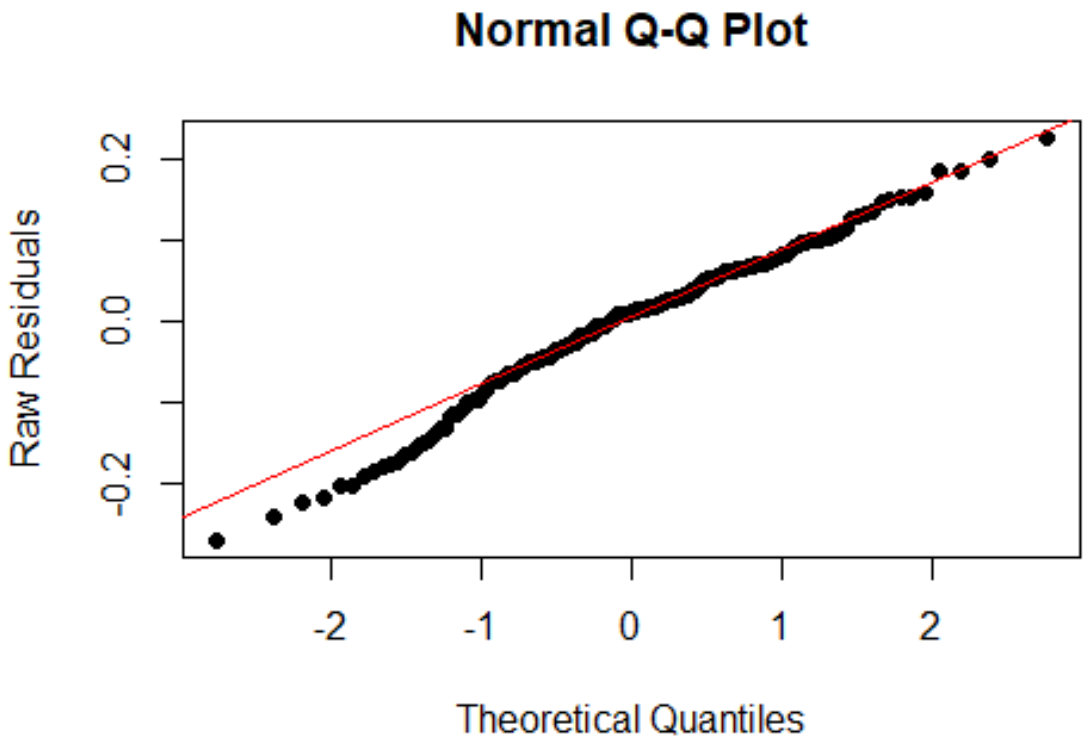
Homoskedasticity: graphic verification



The point cloud represented by the residuals is homogeneous around zero

Qualitative Symmetry Analysis (Normality)

Graphical Verification



Symmetry check (summary)

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.26866 -0.04929  0.01224  0.06256  0.22673
```

p-value Shapiro Test: **0,0214**

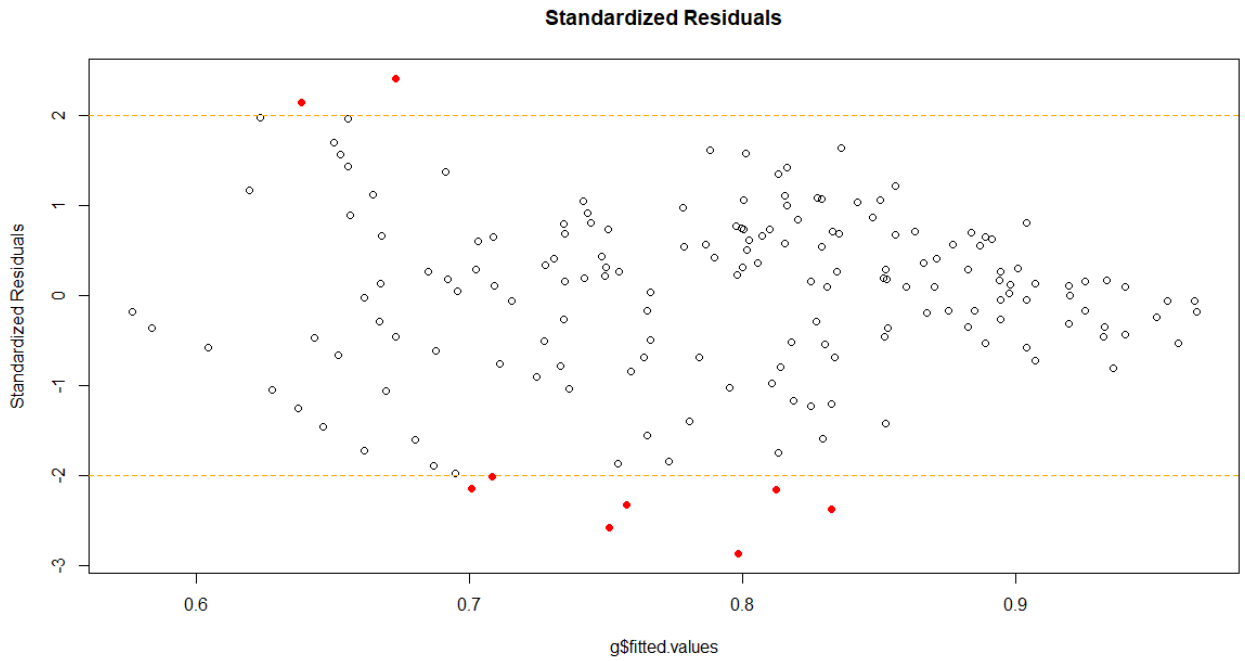
Research and analysis of influential points

Outliers

Leverage Points

Outliers

Analysis using standardised residues



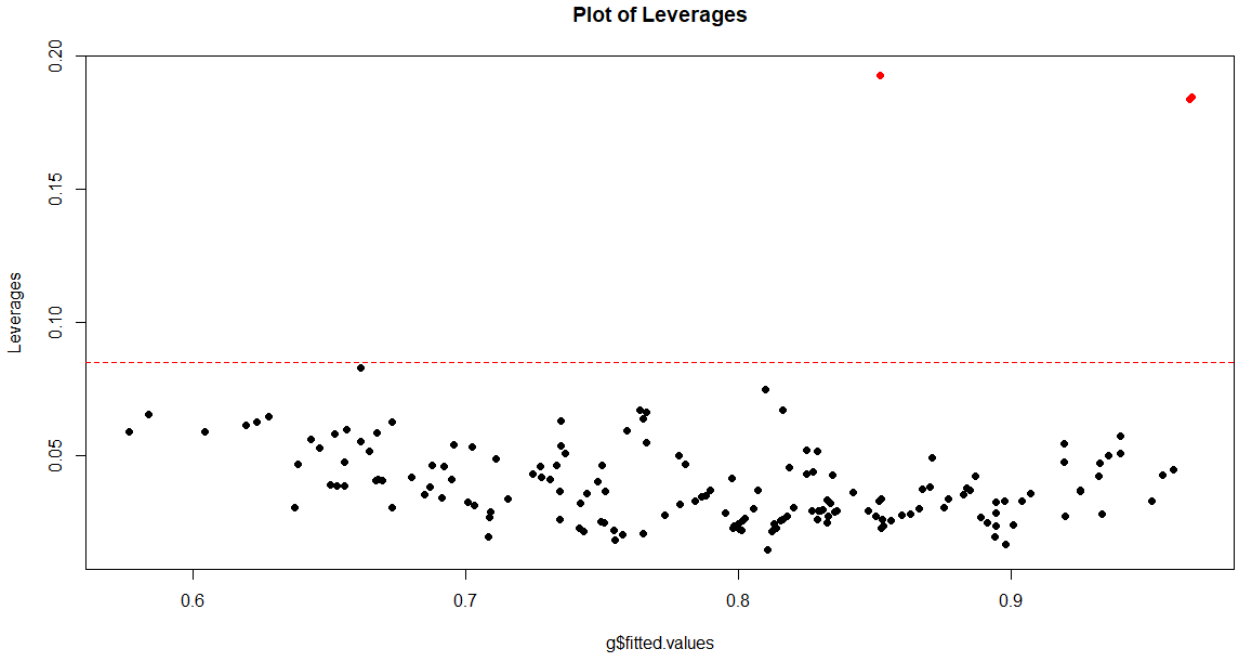
In detail:

	Age	Sleep.efficiency	Awakenings	Caffeine.consumption	Alcohol.consumption	Smoking.status	Exercise.frequency
67	25	0.84	4	0	4	No	2
70	40	0.52	2	0	2	Yes	3
215	27	0.61	1	25	2	No	2
259	54	0.61	1	0	0	Yes	3
273	58	0.50	2	0	3	Yes	3
307	27	0.90	1	0	5	Yes	2
346	25	0.51	1	50	2	Yes	3
359	41	0.54	1	0	2	Yes	3
365	37	0.53	1	0	0	Yes	3
Mean	44.78	0.793	1.713	15.37	1.126		2.086

p-value Shapiro Test: **0,0214**

Leverage Points

Leverages analysis

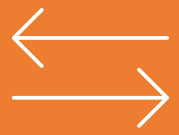


In detail:

	Age	Sleep.efficiency	Awakenings	Caffeine.consumption	Alcohol.consumption	Smoking.status	Exercise.frequency
85	50	0.64	4	0	3	No	0
100	65	0.77	4	0	0	No	3
169	35	0.92	0	50	1	No	3
Mean	44.78	0.793	1.713	15.37	1.126		2.086

p-value Shapiro Test: **0.03939**

Further analysis on the dataset

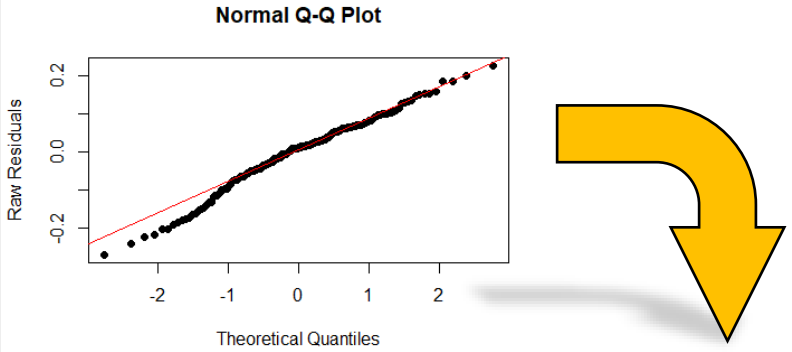


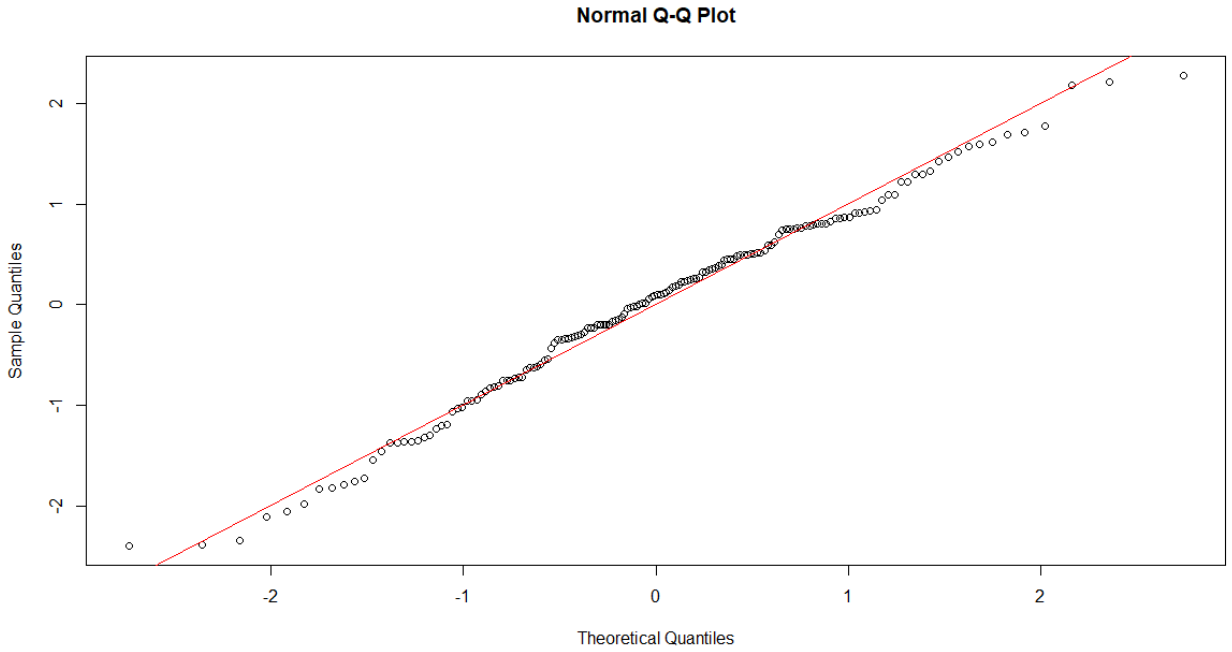
Box-Cox Transformation

Elimination of irrelevant covariates

Forecast

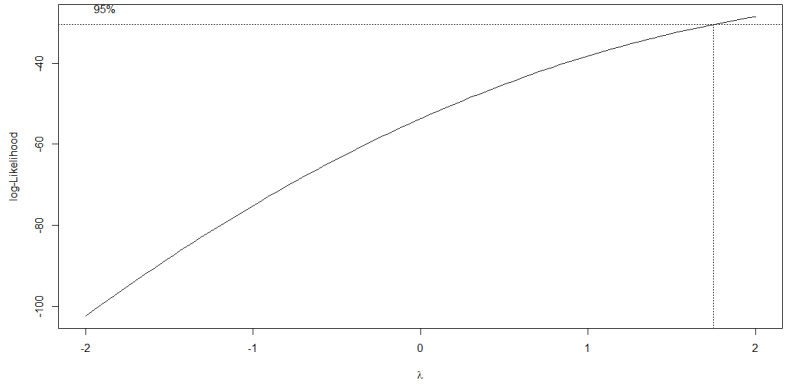
Graphical Verification





Box Transformation - Cox

«Best Lambda»

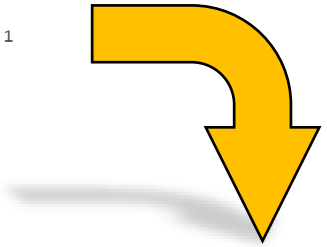


p-value Shapiro Test: **0.2076**

Summary

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -1.023e-01  2.290e-02  -4.467 1.52e-05 ***
Age                  4.805e-04  3.816e-04   1.259   0.2099
Awakenings          -3.994e-02  3.653e-03 -10.933  < 2e-16 ***
Caffeine.consumption  8.409e-05  1.476e-04   0.570   0.5696
Alcohol.consumption  -1.703e-02  3.027e-03  -5.625 8.43e-08 ***
Smoking.statusYes    -5.443e-02  9.837e-03  -5.533 1.31e-07 ***
Exercise.frequency    8.626e-03  3.840e-03   2.246   0.0261 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06022 on 155 degrees of freedom
Multiple R-squared:  0.5954,    Adjusted R-squared:  0.5797
F-statistic: 38.02 on 6 and 155 DF,  p-value: < 2.2e-16
```



```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -0.0982238  0.0217169  -4.523 1.20e-05 ***
Age                  0.0004385  0.0003736   1.174   0.2423
Awakenings          -0.0403055  0.0035874 -11.235  < 2e-16 ***
Alcohol.consumption -0.0173262  0.0029738  -5.826 3.14e-08 ***
Smoking.statusYes   -0.0553198  0.0096909  -5.708 5.59e-08 ***
Exercise.frequency   0.0088602  0.0038096   2.326   0.0213 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06009 on 156 degrees of freedom
Multiple R-squared:  0.5946,    Adjusted R-squared:  0.5816
F-statistic: 45.75 on 5 and 156 DF,  p-value: < 2.2e-16
```

Elimination of irrelevant covariates

Best model (R^2_{adj})

1,2,4,5,6	1,2,3,4,5,6	2,4,5,6	2,3,4,5,6	1,2,4,5
0.566	0.564	0.563	0.560	0.555

Prediction on the TEST SET

	fit	lwr	upr
414	0.8427256	0.8221625	0.8627987
416	0.7920239	0.7561476	0.8263440
418	0.5985463	0.5543128	0.6397285
419	0.8356796	0.8089628	0.8615683
420	0.8538435	0.8192980	0.8870447
422	0.7208977	0.6857033	0.7544522
425	0.7787205	0.7339185	0.8210816
428	0.8944891	0.8678990	0.9203113
429	0.7364774	0.6948156	0.7759054
430	0.9162005	0.8924198	0.9393794
433	0.7477664	0.7187347	0.7757124
437	0.8020433	0.7751739	0.8280412
438	0.8954690	0.8683488	0.9217916
439	0.8487501	0.8241339	0.8726722
440	0.5844620	0.5331073	0.6316553
442	0.6461349	0.6070989	0.6829433
445	0.7510283	0.7105553	0.7894290
449	0.7614919	0.7278831	0.7936789
451	0.8089144	0.7845822	0.8325357
452	0.8697394	0.8384813	0.8999124

Forecast

Standard Error (Level of significance5%)

414	416	418	419	420	422	425
0.008665876	0.014059363	0.012908239	0.011122327	0.014630673	0.012530974	0.017154291
428	429	430	433	437	438	439
0.011862082	0.015094056	0.010887081	0.010776938	0.010727240	0.012108388	0.010423804
440	442	445	449	451	452	
0.014527635	0.012383324	0.014973684	0.012670605	0.009814577	0.013515944	