# کاربرد مدل جعبه خاکستری (Grey box model) در پیش بینی های مرتبط با بیماری COVID-19

Harvard Data Science Review • 1.1

# What is the Purpose of Statistical Modelling?

David Hand

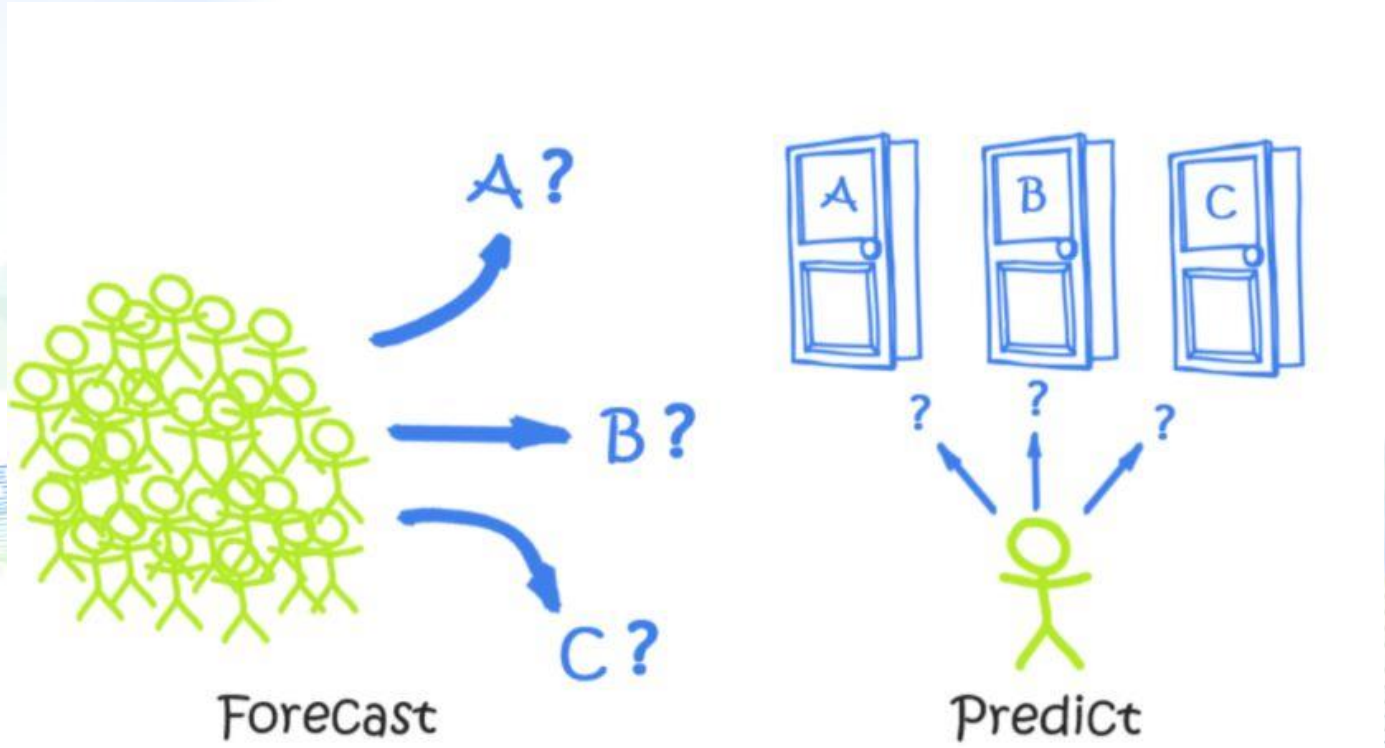
**Published on:** Jun 22, 2019

- بررسی ارتباط بین متغیرها در حضور سایر متغیرهای اندازه گیری شده
- پیش بینی (متغیر پاسخ برای مشاهده جدید یا روند آینده)

# Predict versus Forecast

**Prediction** is concerned with estimating the outcomes for unseen data.

**Forecasting** is a sub-discipline of prediction in which we are making predictions about the future, on the basis of time-series data.



*Statistical Science*
2010, Vol. 25, No. 3, 289–310
DOI: 10.1214/10-STS330


# To Explain or to Predict?

**Galit Shmueli**

*Abstract.* Statistical modeling is a powerful tool for developing and testing theories by way of causal explanation, prediction, and description. In many disciplines there is near-exclusive use of statistical modeling for causal explanation and the assumption that models with high explanatory power are inherently of high predictive power. Conflation between explanation and prediction is common, yet the distinction must be understood for progressing scientific knowledge. While this distinction has been recognized in the philosophy of science, the statistical literature lacks a thorough discussion of the many differences that arise in the process of modeling for an explanatory versus a predictive goal. The purpose of this article is to clarify the distinction between explanatory and predictive modeling, to discuss its sources, and to reveal the practical implications of the distinction to each step in the modeling process.

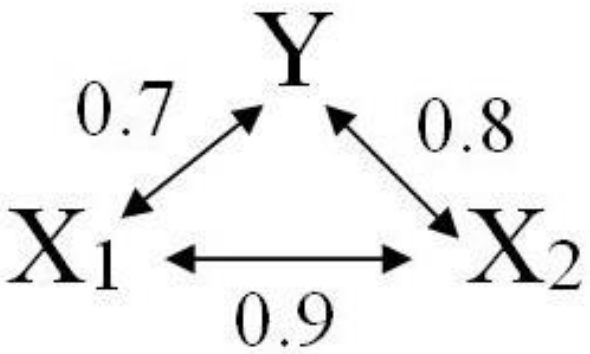
*Key words and phrases:* Explanatory modeling, causality, predictive modeling, predictive power, statistical strategy, data mining, scientific research.

Wrong model can sometimes predict better than the correct one.

Statistical significance plays a minor or no rule in assessing predictive performance. In fact, it is some times the case that removing inputs with small coefficients, *even if they are statistically significant*, results in improved prediction accuracy.

Greenberg & Parks, 1997; Wu, Harris & McAuley, 2007

# Greedy property




Stepwise regression
LASSO regression
LARS regression

Best subset regression
Ridge regression
Elastic-net regression

Adaptive boosting
Random forest

K-nearest neighbors

# Model selection or Variable selection

*Journal of Statistical Software*

*May 2010, Volume 34, Issue 12.* *http://www.jstatsoft.org/*

## glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models

Vincent Calcagno
McGill University

Claire de Mazancourt
McGill University

Journal of Machine Learning Research 15 (2014) 3133-3181 Submitted 11/13; Revised 4/14; Published 10/14

# Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?

**Manuel Fernández-Delgado** MANUEL.FERNANDEZ.DELGADO@USC.ES

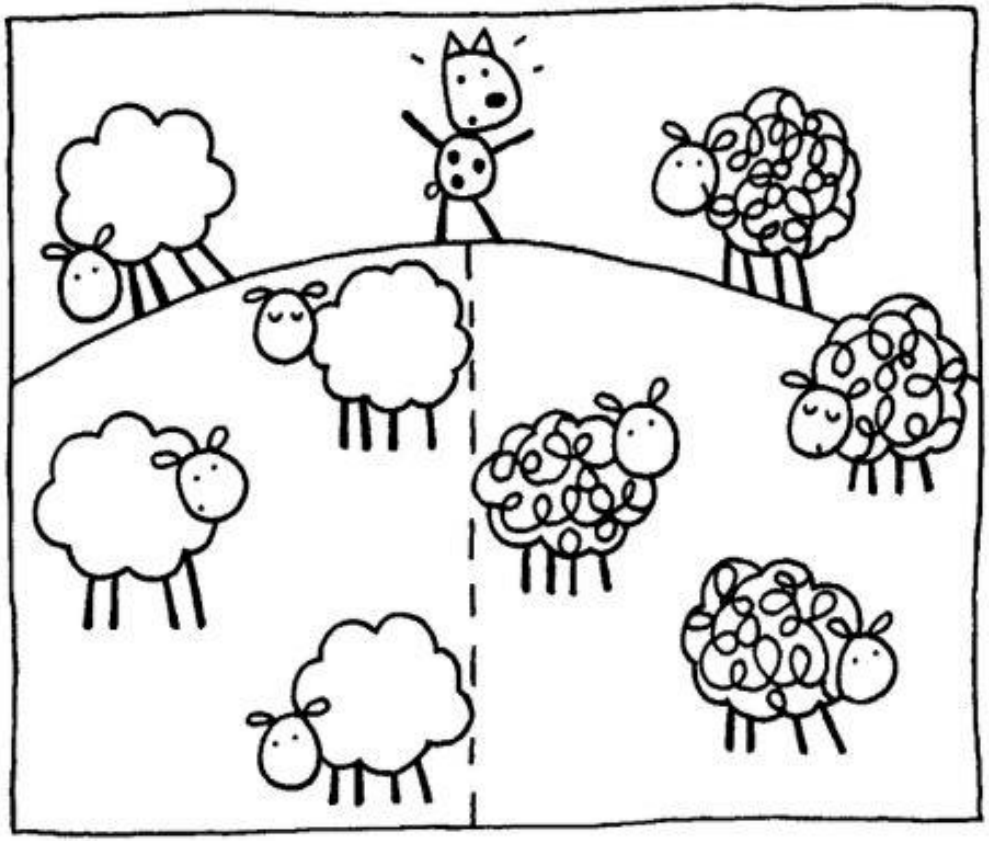
We evaluate **179 classifiers** arising from **17 families** (discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, boosting, bagging, stacking, random forests and other ensembles, generalized linear models, nearest-neighbors, partial least squares and principal component regression, logistic and multinomial regression, multiple adaptive regression splines and other methods), implemented in Weka, R (with and without the caret package), C and Matlab, including all the relevant classifiers available today. We use **121 data sets**, which represent **the whole UCI** data base (excluding the large-scale problems) and other own real problems, in order to achieve significant conclusions about the classifier behavior, not dependent on the data set col-


Forecast
Reality
"Prediction is very difficult, especially if it's about the future."

**Prediction**

**Instance based (black box)**

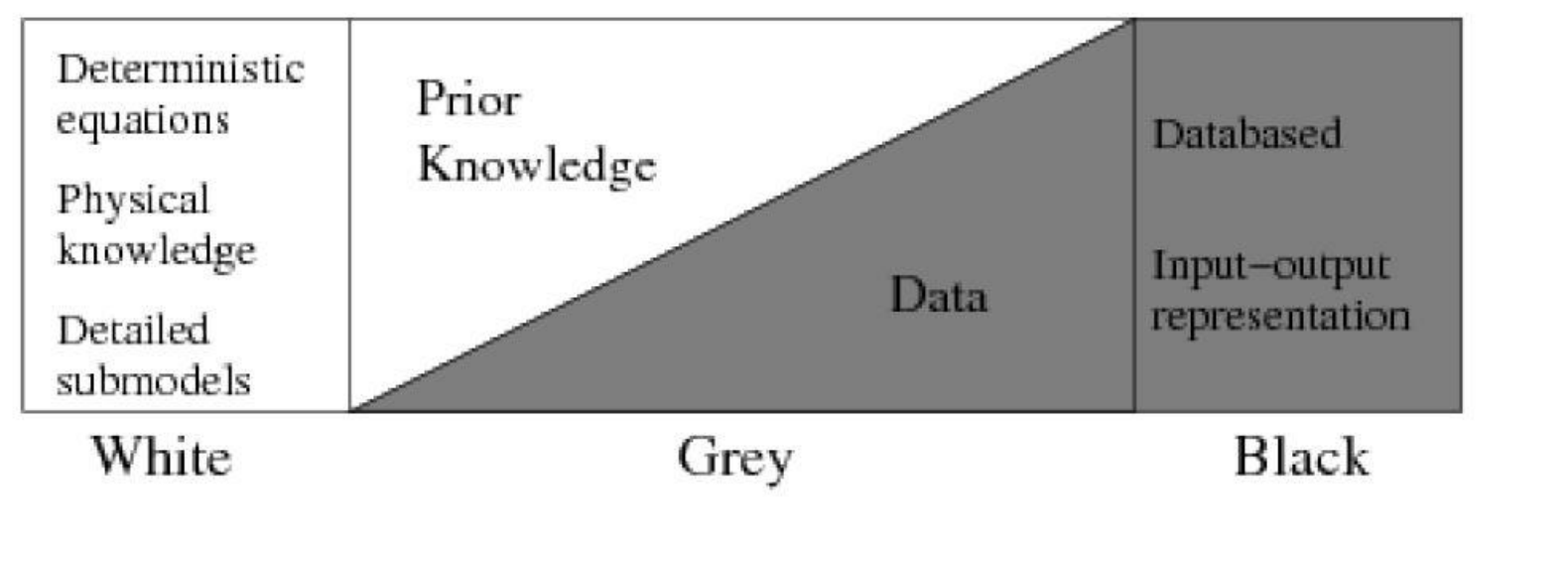
**Model based (white box)**



# Grey box model (GM)

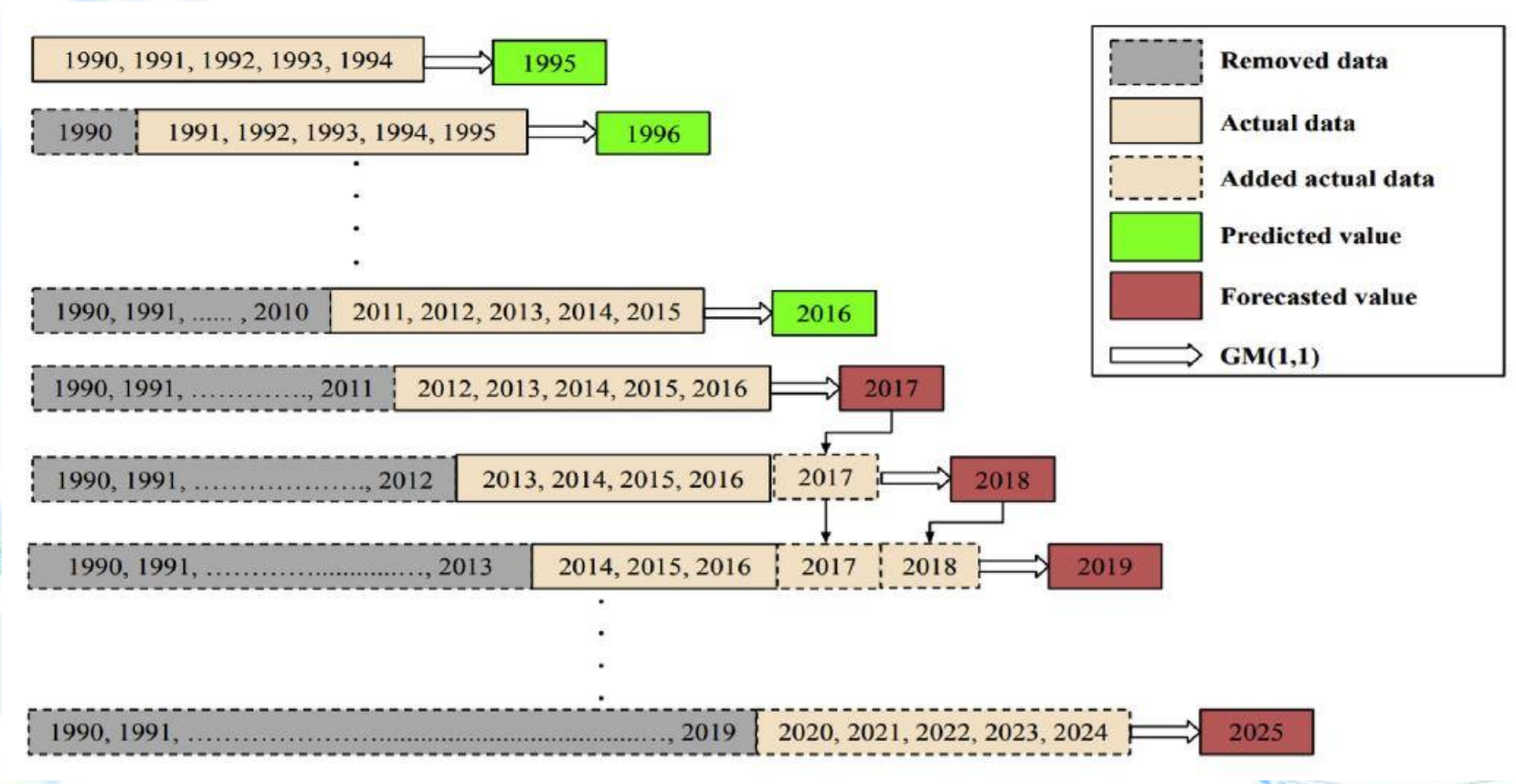
In the GM theory, “white”, “black” and “grey” defines the known information, unknown information and partially known information about the system, respectively (Deng, 1982).

**The next unknown data can be produced by a few number of past data**



# Grey box model (GM)

The GM(1,1), is the most widely used among other grey prediction models, implies a first-order single variable prediction model.



# Grey box model (GM)

## Package 'greybox'

May 20, 2020

**Type** Package

**Title** Toolbox for Model Building and Forecasting

**Version** 0.6.0

**Date** 2020-05-19

**URL** https://github.com/config-i1/greybox

**BugReports** https://github.com/config-i1/greybox/issues

**Language** en-GB

**Description** Implements functions and instruments for regression model building and its application to forecasting. The main scope of the package is in variables selection and models specification for cases of time series data. This includes promotional modelling, selection between different dynamic regressions with non-standard distributions of errors, selection based on cross validation, solutions to the fat regression model problem and more. Models developed in the package are tailored specifically for forecasting purposes. So as a results there are several methods that allow producing forecasts from these models and visualising them.

**License** GPL (>= 2)

**Depends** R (>= 3.0.2)

# Grey box model (GM)

1. dnorm - Normal distribution,
2. dlogis - Logistic Distribution,
3. dlaplace - Laplace distribution,
4. dllaplace - Log Laplace distribution,
5. dalaplace - Asymmetric Laplace distribution,
6. dt - T-distribution,
7. ds - S-distribution,
8. dls - Log S-distribution,
9. dfnorm - Folded normal distribution,
10. dlnorm - Log normal distribution,
11. dbcnorm - Box-Cox normal distribution,
12. dinvgauss - Inverse Gaussian distribution,
13. dbeta - Beta distribution,
14. dpois - Poisson Distribution,
15. dnbinom - Negative Binomial Distribution,
16. plogis - Cumulative Logistic Distribution,
17. pnorm - Cumulative Normal distribution.

# COVID-19 modeling

متغیر تحت بررسی؟؟

موارد جدید مبتلا

مرگ های جدید

موارد بستری

موارد بستری در بخش مراقبتهای ویژه

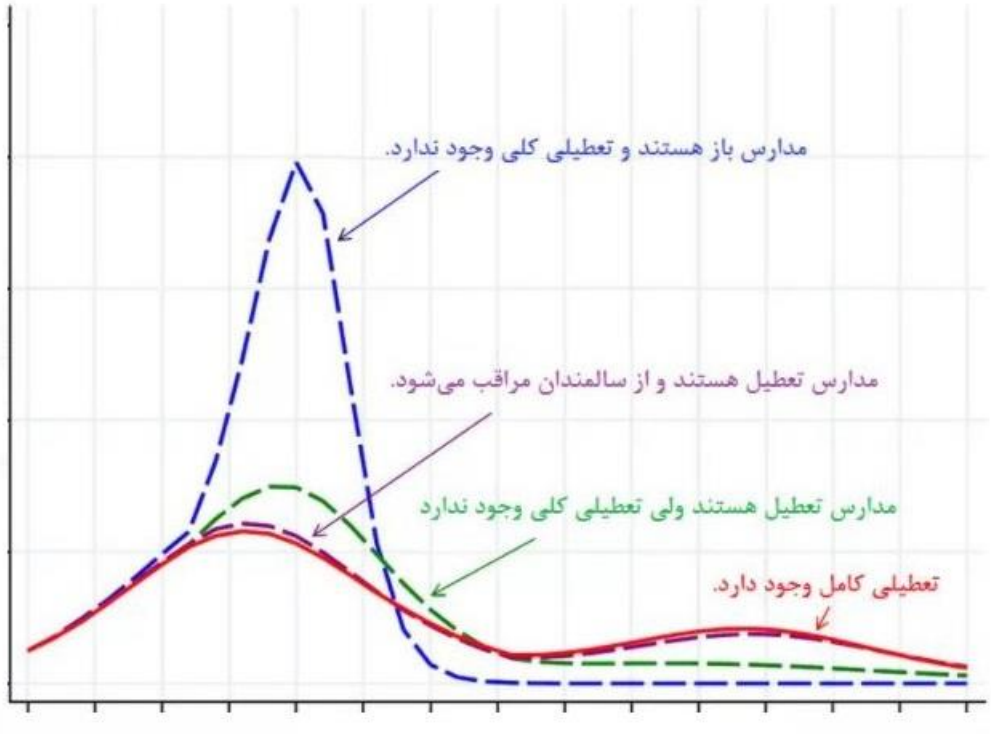
موارد مبتلایان فعال

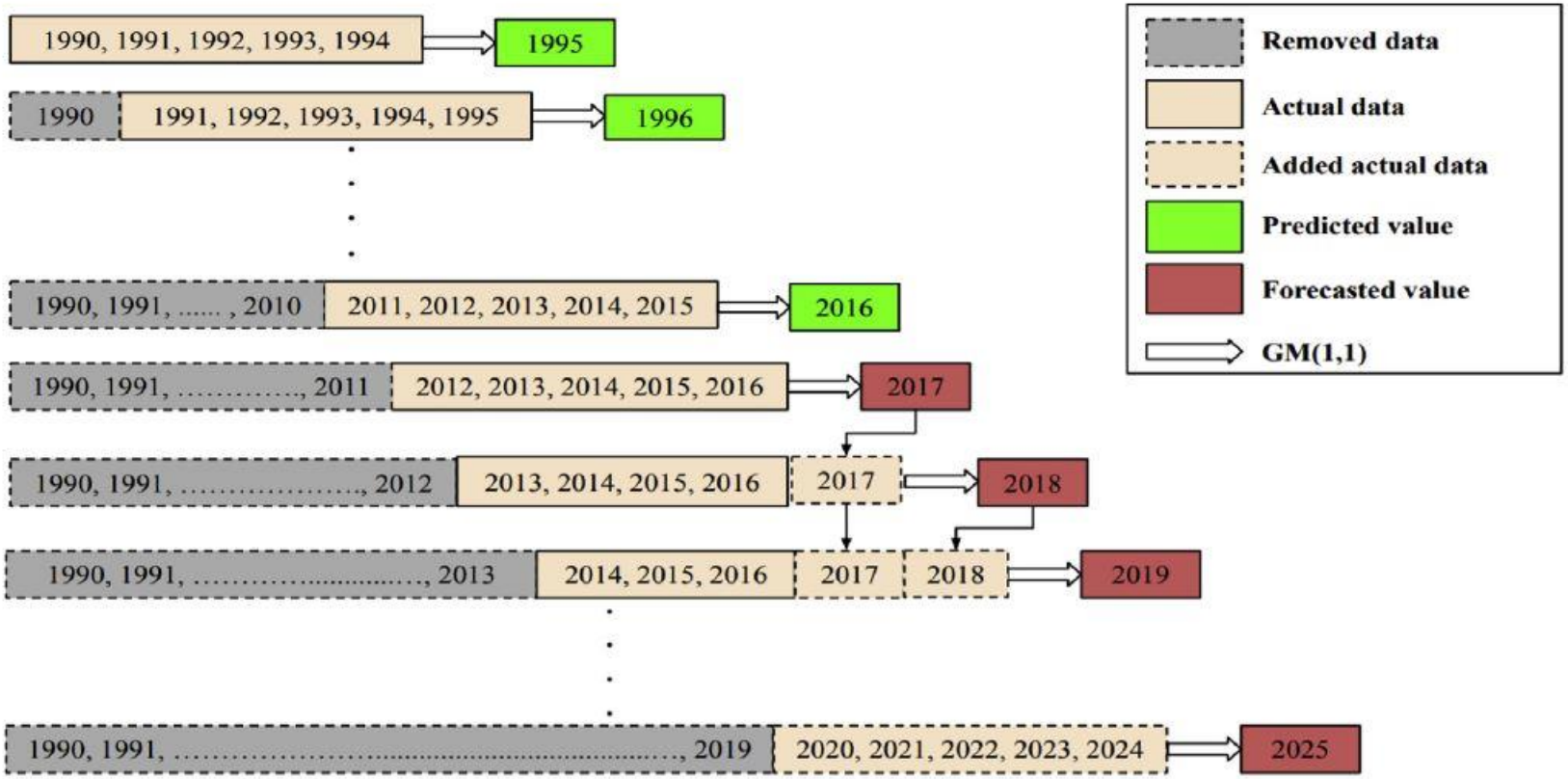
آیا علاوه بر متغیر تحت بررسی (پیامد)، متغیر دیگری در دسترس است؟

واحد زمانی اطلاعات موجود بر اساس روز، هفته یا ماه هستند؟

## How new cases are changing by day

## روند میزان درصد رعایت پروتکل های بهداشتی در کشور

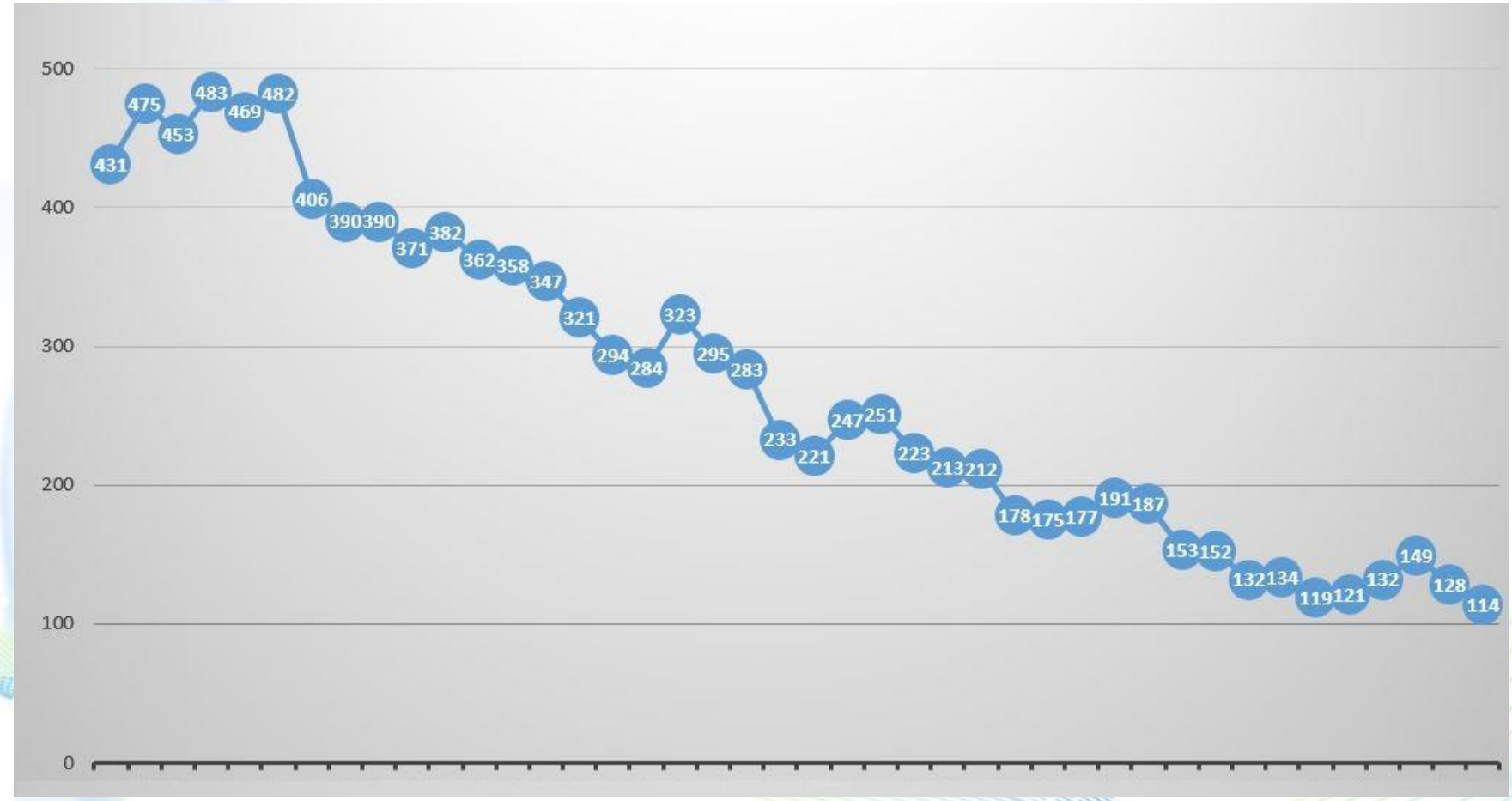






$$APE\ (\%) = \left|\frac{u(i) - \hat{u}(i)}{u(i)}\right| x100$$

$$MAPE\ (\%) = \sum_{i=1}^{n} \left|\frac{u(i) - \hat{u}(i)}{u(i)}\right| x\frac{100}{n}$$

# Grey box model (GM)

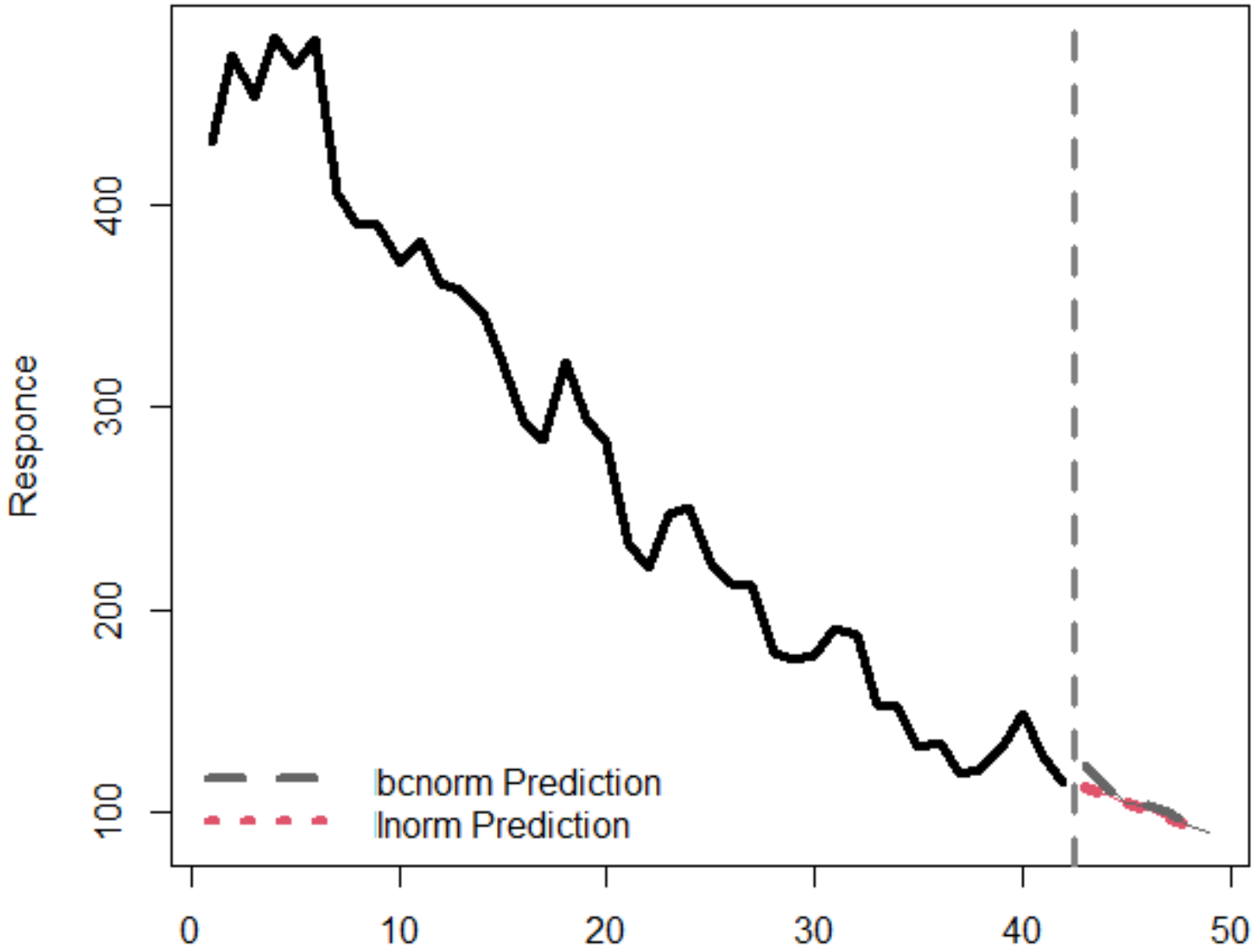


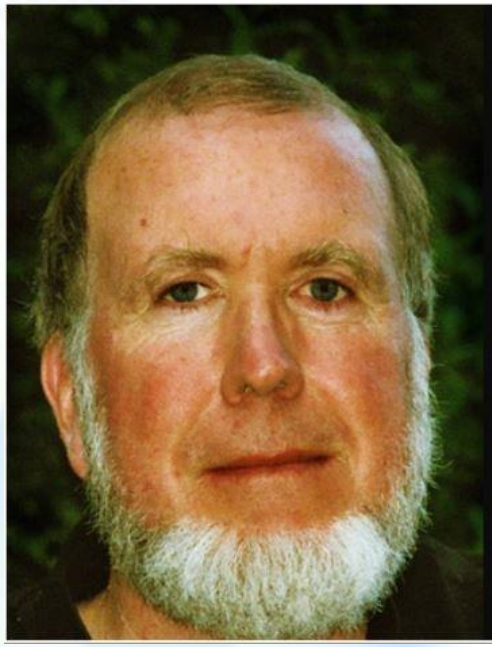
R Console

```
> library(greybox)
> ####################
> h.b5=1
> dist=c('dnorm','dlogis','dlaplace','dalaplace','ds','dt','dfnorm','dlnorm','dl$
> o=matrix(,length(dist),5,dimnames=list(dist,c('AIC','AICc','BIC','BICc','MAPE'$
> for(i in 1:length(dist))
+ {
+  ml=alm(y~x,df,dist=dist[i])
+  o[i,1:4]=summary(ml)$ICs
+
+  df.p=cbind(df,p=NA,APE=NA)
+  for(j in 1:(n-4))
+  {
+   x.b5=df[j:(j+4),'x']
+   df.b5=cbind(x.b5,y.b5=df[j:(j+4),'y'])
+
+   xn.b5=(max(x.b5)+1):(max(x.b5)+h.b5)
+   dn.b5=cbind(x.b5=xn.b5,y.b5=NA)
+   in.b5=which(df.p[,'x']==xn.b5)
+   df.p[in.b5,'p']=predict(alm(y.b5~x.b5,df.b5,dist=dist[i]),newdata=dn.b5)$mean
+   df.p[in.b5,'APE']=abs((df.p[in.b5,'y']-df.p[in.b5,'p'])/df.p[in.b5,'y'])*100
+  }
+ o[i,5]=mean(df.p[,'APE'],na.rm=T)
+ }
> o
                AIC      AICc       BIC      BICc      MAPE
dnorm      394.5574 395.1890 399.7704 400.9507 10.211539
dlogis     394.2396 394.8711 399.4526 400.6329 10.296051
dlaplace   392.8836 393.5152 398.0966 399.2769 10.833072
dalaplace  394.5328 395.6138 401.4834 403.5038 10.698800
ds         398.3371 398.9686 403.5501 404.7304  9.962392
dt         591.1379 591.7694 596.3509 597.5312 10.087543
dfnorm     394.5610 395.1926 399.7740 400.9543 10.197174
dlnorm     376.6297 377.2613 381.8427 383.0230  9.600206
dllaplace  383.5674 384.1989 388.7804 389.9607  9.675757
dls        394.7923 395.4239 400.0053 401.1856  9.650170
dbcnorm    374.0273 374.6588 379.2403 380.4206  9.643768
dinvgauss  376.6238 377.2554 381.8368 383.0171  9.629683
dpois      383.6653 384.2969 388.8783 390.0586  9.743713
> |
```

| APE (%) | تعداد پیش بینی شده | تعداد مرگها | روز |
|---|---|---|---|
| NA | NA | 431 | 1 آذرماه |
| NA | NA | 475 | 2 آذرماه |
| NA | NA | 453 | 3 آذرماه |
| NA | NA | 483 | 4 آذرماه |
| NA | NA | 469 | 5 آذرماه |
| 1.3 | 488 | 482 | 6 آذرماه |
| 18.6 | 481 | 406 | 7 آذرماه |
| 10.0 | 429 | 390 | 8 آذرماه |
| 4.1 | 374 | 390 | 9 آذرماه |
| 3.8 | 357 | 371 | 10 آذرماه |
| 10.4 | 342 | 382 | 11 آذرماه |
| 1.7 | 368 | 362 | 12 آذرماه |
| 0.6 | 360 | 358 | 13 آذرماه |
| 1.2 | 351 | 347 | 14 آذرماه |
| 6.8 | 343 | 321 | 15 آذرماه |
| 6.9 | 314 | 294 | 16 آذرماه |
| 0.8 | 286 | 284 | 17 آذرماه |
| 18.2 | 264 | 323 | 18 آذرماه |
| 2.0 | 289 | 295 | 19 آذرماه |
| 4.7 | 296 | 283 | 20 آذرماه |
| 25.2 | 292 | 233 | 21 آذرماه |
| 9.0 | 241 | 221 | 22 آذرماه |
| 19.7 | 198 | 247 | 23 آذرماه |
| 15.6 | 212 | 251 | 24 آذرماه |
| 4.5 | 233 | 223 | 25 آذرماه |

| | | | |
|---|---|---|---|
| 26 آذرماه | 213 | 238 | 11.5 |
| 27 آذرماه | 212 | 219 | 3.1 |
| 28 آذرماه | 178 | 198 | 11.4 |
| 29 آذرماه | 175 | 171 | 2.1 |
| 30 آذرماه | 177 | 163 | 7.9 |
| 1 دیماه | 191 | 160 | 16.0 |
| 2 دیماه | 187 | 174 | 6.7 |
| 3 دیماه | 153 | 192 | 25.4 |
| 4 دیماه | 152 | 165 | 8.8 |
| 5 دیماه | 132 | 146 | 10.7 |
| 6 دیماه | 134 | 121 | 9.6 |
| 7 دیماه | 119 | 117 | 1.4 |
| 8 دیماه | 121 | 114 | 6.1 |
| 9 دیماه | 132 | 111 | 16.2 |
| 10 دیماه | 149 | 124 | 17.0 |
| 11 دیماه | 128 | 144 | 12.2 |
| 12 دیماه | 114 | 143 | 25.8 |
| میانگین خطای نسبی (MAPE): | | | %9.6 |
| روز | تعداد مرگها | تعداد پیش بینی شده | |
| 13 دیماه | | 123 | |
| 14 دیماه | ؟ | 113 | |
| 15 دیماه | ؟ | 104 | |
| 16 دیماه | ؟ | 103 | |
| 17 دیماه | ؟ | 99 | |
| 18 دیماه | ؟ | 92 | |
| 19 دیماه | ؟ | 89 | |

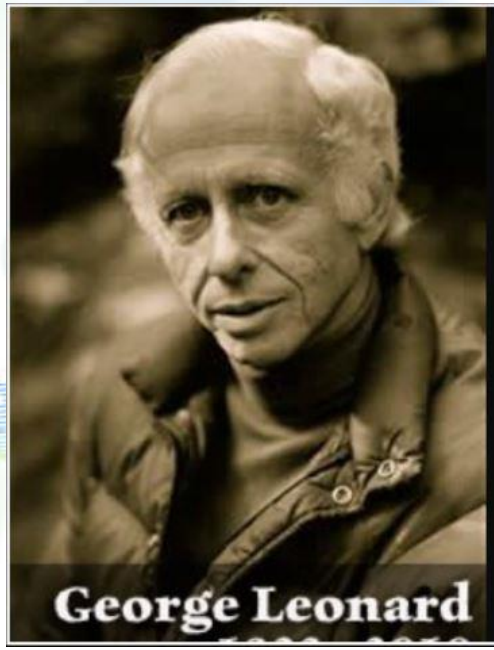

Responce
100
200
300
400
0
10
20
30
40
50
bcnorm Prediction
lnorm Prediction



Any believable prediction will be wrong. Any correct prediction will be unbelievable.

— Kevin Kelly —

AZ QUOTES




Perhaps the safest prediction we can make about the future is that it will surprise us.

— George Leonard —

AZ QUOTES


THANK YOU