Regression Model for GDP Growth from UN (United Nations 2018 DATASET)

1 author:

Rui Sarmento
Statsframe

Some of the authors of this publication are also working on these related projects:

R Packages View project

Business Bankruptcy Prediction View project

This Report regards only the following variables from UN (United Nations) datasets. This data was retrieved from Kaggle website and one of its contests. This dataset has been cleaned and prepared for analysis, for the following variables only. This report is also accompanied by other reports the reader might want to look at, like Inferential Reports, and Cluster Analysis.

Surface.area..km2.
Population.in.thousands..2017.
Population.density..per.km2..2017.
Sex.ratio..m.per.100.f..2017.
GDP..Gross.domestic.product..million.current.US..
GDP.growth.rate..annual....const..2005.prices.
GDP.per.capita..current.US..
Economy..Agriculture....of.GVA.
Economy..Industry....of.GVA.
Economy..Services.and.other.activity....of.GVA.
Employment..Agriculture....of.employed.
Employment..Industry....of.employed.
Employment..Services....of.employed.
Unemployment....of.labour.force.
Food.production.index..2004.2006.100.
International.trade..Exports..million.US..
International.trade..Imports..million.US..
International.trade..Balance..million.US..
Balance.of.payments..current.account..million.US..
Population.growth.rate..average.annual...
Urban.population....of.total.population.
Urban.population.growth.rate..average.annual...
Fertility.rate..total..live.births.per.woman.
Infant.mortality.rate..per.1000.live.births
Health..Total.expenditure....of.GDP.
Health..Physicians..per.1000.pop..
Education..Government.expenditure....of.GDP.
Mobile.cellular.subscriptions..per.100.inhabitants.
Individuals.using.the.Internet..per.100.inhabitants.
Threatened.species..number.
Pop..using.improved.sanitation.facilities..urban.rural....

1. Linear Regression Analysis:

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables when the focus is on the relationship between a dependent variable and one or more independent variables i.e. the predictors.

Please check the following table:

Multiple.R_Squared	Adjusted.R_Squared	F.Statistic	pvalue
1.000	1.000	256868071.554	0.000

The Adjusted R-squared value (100%) means the % of the total variability in "GDP.growth.rate..annual....const..2005.prices." that is explained by the independent variables used in the linear regression model.

The F-statistic is 256868071.554, and has a p-value equal to 0 ($p < 0.05$). This means that the null hypothesis should be rejected and, consequently, the model is highly significant.

However, it is important to know, if all the studied variables significantly contribute to the linear regression model.

Please check the following table:

term	estimate	std.error	statistic	p.value
(Intercept)	-1414.418	745.466	-1.897	0.060
Population.in.thousands..2017.	0.000	0.000	0.706	0.481
Population.density..per.km2..2017.	0.000	0.001	0.035	0.972
Sex.ratio..m.per.100.f..2017.	0.018	0.020	0.924	0.357
Economy..Agriculture....of.GVA.	7.167	0.137	52.126	0.000
Economy..Industry....of.GVA.	7.207	0.155	46.496	0.000
Economy..Services....of.GVA.	7.256	0.158	46.026	0.000
Employment..Agriculture....of.employed.	6.962	7.471	0.932	0.353
Employment..Industry....of.employed.	6.885	7.467	0.922	0.358
Employment..Services....of.employed.	6.933	7.479	0.927	0.356
Unemployment....of.labour.force.	-0.198	0.075	-2.643	0.009
International.trade..Exports..million.US..	-0.005	0.005	-0.973	0.333
International.trade..Imports..million.US..	0.005	0.005	0.972	0.333
International.trade..Balance..million.US..	0.005	0.005	0.972	0.333

The table presents the estimation of each variable's coefficient. Looking at the p-values, it is possible to verify that variables Economy..Agriculture....of.GVA.,

Economy..Industry....of.GVA., Economy..Services....of.GVA.,
Unemployment....of.labour.force. are significant (p<0.05).

Additionally, variables with positive coefficient (see "estimate" column in the table) are directly related to an increase in the dependent variable values. Variables with negative coefficient are directly related to a decrease in the dependent variable values.