Available online at www.gi.sanu.ac.rs

J. Geogr. Inst. Cvijic. 63(2) (1-10)

Original scientific paper UDC: 551.58:519.2
DOI: 10.2298/IJGI1302001M

ABOUT HIDDEN INFLUENCE OF PREDICTOR VARIABLES -SUPPRESSOR AND MEDIATOR VARIABLES-

Boško Milovanović *[1]
* Geographical Institute "Jovan Cvijić" Serbian Academy of Sciences and Arts, Belgrade, Serbia

Received 03 April 2013; reviewed 18 April 2013; accepted 01 August 2013

Abstract: In this paper procedure for researching hidden influence of predictor variables in regression models and depicting suppressor variables and mediator variables is shown. It is also shown that detection of suppressor variables and mediator variables could provide refined information about the research problem. As an example for applying this procedure, relation between Atlantic atmospheric centers and air temperature and precipitation amount in Serbia is chosen.

Key words: Multiple linear regression, stepwise regression, suppressor variables, mediator variables, Serbia

List of abbreviations: R – precipitation amount in Serbia; T – air temperature in Serbia; LoAH – Longitude of Azores high; LaAH – Latitude of Azores high; IAH – Intensity of Azores high; IIL – Intensity of Iceland low; LoIL – Longitude of Iceland low; LaIL – Latitude of Iceland low; NAO – North Atlantic Oscillation index; R2 – coefficient of determination.

Introduction

Correlation analysis is often used in quantification of the relation between two variables x and y (this analysis could not prove cause and consequence relation) and in many cases as a measure of this correlation Pearson product-moment correlation coefficient could be used. When there is a need for researching relation of several independent or predictor variables $x_1...x_n$ and one dependent variable y, multiple linear regression could be used (under condition that predictors in the regression model has good theoretical background, residuals are normally distributed, linear relation between variables, large sample etc.)

General form of the multiple linear regression model is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i \tag{1}$$

[1] Correspondence to: b.milovanovic@gi.sanu.ac.rs

where:

- Y_i dependent variable
- X1i, X2i…Xki independent/predictor variables
- $\beta_0, \beta_1, \beta_2, \ldots \beta_k$ regression coefficients
- ε_i noise
- k number of independent variables

Desirable state in regression models is such that correlations between predictors are low, and that each of them has high correlation with the dependent variable. In situation like this, coefficient of determination is very high and contribution of each predictor for accounting variance of dependent variable is clear. On the other hand, there is often a case when there is a high correlation between predictors and dependent variable, but also high redundancy of predictors. In such a case, it is difficult to determine contributions of predictors, and including new ones in the model does not increase coefficient of determination.

It is interesting that sometimes there are predictor variables that show no correlation with dependent variable, but still contribute accounting for her variance. In situation like that, we are dealing with suppressor variables. Krus & Wilkinson (1986, p.21) quote: "A suppressor variable has a zero correlation with the criterion, but nevertheless contributes to the predictive validity of model. The current definition of a suppressor variable is that it is a variable which increases regression weights and, thus, increases the predictive validity of other variables in a regression equation." Pandey & Elliot (2010) describe suppressor variables as predictor variables that are not correlated (or correlated minimally) with the dependent variable, even though such predictor variables correlate with one or more of the independent variables. The correlations of suppressor variables with the independent variables increase the multiple correlations by partialling out invalid variance of the other predictors included in the regression equation. In other words, suppressor variables are predictors that in isolation correlate weakly (or zero) with the outcome variable but are strongly correlated with one or more predictors that are correlated with the outcome variable. This definition implies that the component of the predictor variable associated with the outcome variable has noise, that is, irrelevant variance; the suppressor variable suppresses or explains the part of the predictor variable that is irrelevant and not associated with the outcome variable.

Same authors divide suppressor variables into for types, in dependence of the sign of the relation between predictors and between predictors and dependent variable:

- Classic suppression - a suppressor and an outcome variable have a zero correlation, the prediction in the outcome variable increases when a suppressor variable is added to the equation simply because the suppressor variable is correlated with another predictor (or set of predictors) that are correlated with the outcome variable. In this case, the suppressor variable removes irrelevant predictive variance from the other predictor (or set of predictors) and increases the predictor's regression weight, thus increasing overall model predictability.
- Negative suppression - a negative suppressor works in a manner similar to that of a classic suppressor by removing irrelevant variance from a predictor (or set of predictors), increasing the predictor's regression weight, and increasing overall predictability of the regression equation. The difference between these two types of suppressors is the negative suppressor's positive zero-order correlation with other predictor variable (s) and with the outcome variable; however, when entered in multiple regressions, the negative suppressor has a negative beta weight.
- Reciprocal suppression both the predictor and the suppressor are positively correlated with the outcome variable but negatively correlated with each other.
- Absolute and relative suppression - absolute suppression is defined by the relationship between the predictor's weight in bivariate regression equation and its weight in multivariate equations. It exists whenever adding predictors increases the weight of the variable relative to its weight in the bivariate equation. On the other hand, if the regression weight of a predictor increases when a new variable is added to a regression equation, but the increase is not beyond the respective weight of the predictor in the bivariate mode, then the new variable is a relative suppressor (Tzelgov & Henik, 1991).

On the other hand, there is possibility for situation where in a simple matrix of correlations there are statistically significant correlation both between predictors and predictors and dependent variable. But, when introduced into regression model some of the predictors loose their importance. This is the situation where mediating variable could be detected.

According to Baron & Kenny (1986), variable has mediating function when it meets following criteria:

- variations in levels of the independent variable significantly account for variations in the presumed mediator.
- variations in the mediator significantly account for variations in the dependent variable.
- when the relations between predictor and mediator, and mediator and dependent variable are controlled, a previously significant relation between the independent and dependent variables is no longer significant, with the strongest demonstration of mediation occurring when the value of this relation is zero.

In regard to the last condition it may be added that when relation between predictor and dependent variable is not zero, this indicates the operation of multiple mediating factors. Preacher & Hayes (2008) quotes that in some cases just combination of few mediator variables could have mediating effect.

Data and methods

For the mean position of the fields with constantly high, or constantly low air pressure, usual term is center of action (established in 1883. by French meteorologist Teisserenc de Bort, Barry 2003, p. 3). Data about the centers of action above Atlantic ocean (Icelandic low and Azores high) consider their intensity (pressure on a sea level) and position (latitude and longitude) for the 1949–2004 period[2] (Machel, Kapala & Flohn 1998). Because of their dimensions, influence on frontal systems and moving of air masses, these permanent centers of action have great influence on the climate of Serbia (Radinović, 1981).

For the analysis of air temperature and precipitation amount in Serbia, data from 23 meteorological stations are used (period 1949-2004). On the territory of central part of Serbia and its northern province Vojvodina, there are 22 uniformly distributed meteorological stations. On the southern Serbian province of Kosovo and Metohia data from just one station (Prizren) were available. Interpolation of missing data in time series is done by reduction on the same period (Milosavljević, 1963). In the homogeneity check, we relied on the results

[2] Data about the Azores high and Iceland low are achieved in personal correspondence with Dr Hermann Mächel, KU21 Projekt: KLIDADIGI, Deutscher Wetterdienst Kaiserleistr. 44 63067 Offenbach am Main

of Tošić (2004). By the use of Standard Normal Homogeneity Test, this author analyzed the data from 30 meteorological stations located in Serbia and Montenegro (24 of them are in Serbia) and found that air temperature time series for the period 1951-2000. were homogeneous. Also, Unkašević, Tošić & Vujović (2004) analyzed the precipitation data from 35 meteorological stations in Serbia and Montenegro (29 of them are in Serbia) and noticed that just the data from meteorological station Prizren were inhomogeneous. That is why homogeneity check was repeated at this station and station Vršac[3]. We used software AnClim (Štěpánek, 2005).

According to Domonkos (2006) data from two or more stations are enough for creating referent time series. By usage of the weighted coefficients we formed referent time series (weighted coefficients were calculated as squared coefficient of correlation between test station and referent station). For the homogeneity check of the data from station Prizren, we used the data from stations Zlatibor and Vranje, and for the homogeneity check of the data from station Vršac we used the data from stations Zrenjanin and Veliko Gradište. Coefficients of correlation of first differences between Prizren and referent stations are from 0.91-0.94 for the air temperature, and between 0.61 and 0.72 for precipitation amount. Coefficients of correlation of first differences between Vršac and referent stations are 0.96-0.97 for the air temperature and 0.82-0.85 for the precipitation amount. On the station Prizren there are no inhomogeneities in the air temperature time series, but precipitation time series on this station show some inhomogeneities, so they were adjusted by multiplying with correction factor of 0.944. All time series from the station Vršac were homogeneous.

All the data were analyzed on the moving decadal level. Smoothing of time series is chosen to avoid interannual climate variations which could be caused by some other climate factors.

To determine the measure of relation between mentioned variables, on a first place Pearson product-moment correlation coefficient was calculated. To achieve information about the relation between intensity and position of centers of action on one side and air temperature/precipitation amount in Serbia on the other, multiple linear regression was used. Considering fact that there is a large number of predictors and possible combinations of predictors in centers of action, we used stepwise linear regression with forward selection approach. So, predictors could be in the regression model only if they show statistical

[3] Homogeneity of the data from this station was not tested in Tošić (2004).

significance. Software package Statsoft/Statistica 6.0. is used for processing data.

Results and discussion

Considering coefficients of correlation shown in Table 1, it can be concluded that Azores high becoming more intense when it moves toward northeast. On the other hand, when Iceland low is moving toward southeast it becomes less intense. Relation between these two centers is like this. Pressure rise in Azores high is connected with pressure fall in Iceland low and its moving towards northwest. Decrease of latitude of the Iceland low is followed by move of Azores high toward south, and vice versa, increase of latitude of the Azores high is followed by move of the Iceland low toward north.

Table 1. Correlation matrix between moving decadal values of intensity and position of Atlantic centers of action, NAO index, air temperature and precipitation amount in Serbia[4]

	R	T	LoIL	LaIL	IIL	LoAH	LaAH	IAH	NAO
R	1	-0.454	0.097	0.257	0.617	-0.019	-0.053	-0.286	-0.608
T	-0.454	1	-0.097	-0.333	-0.274	-0.687	0.555	0.141	0.012
LoIL	0.097	-0.097	1	0.233	-0.511	-0.180	-0.034	0.421	0.348
LaIL	0.257	-0.333	0.233	1	-0.199	-0.133	0.426	0.621	0.064
IIL	0.617	-0.274	-0.511	-0.199	1	0.287	-0.216	-0.516	-0.881
LoAH	-0.019	-0.687	-0.180	-0.133	0.287	1	-0.858	-0.386	0.020
LaAH	-0.053	0.555	-0.034	0.426	-0.216	-0.858	1	0.544	-0.064
IAH	-0.286	0.141	0.421	0.621	-0.516	-0.386	0.544	1	0.262
NAO	-0.608	0.012	0.348	0.064	-0.881	0.020	-0.064	0.262	1

Table 1 also shows that, when precipitation amount in Serbia is in question, there is highest correlation with the intensity of Iceland low and NAO index respectively (increase of air pressure in Iceland low and weakening of NAO index is connected with increase of precipitation amount in Serbia). On the other hand, there is a highest (negative) correlation between air temperature in Serbia

[4] The bolded ones are statistically significant on the 95% confidence level

and latitude of Iceland low and position of Azores high (move of this center of action toward northeast is connected with increase of air temperature in Serbia).

The results shown here (concerning Iceland low) are not in compliance with findings of Milovanović, Radovanović & Ducić (2009). Mentioned authors noticed strongest connection between the latitude of Iceland low and air temperature in Serbia. Cause of this discrepancy may be in different data set which mentioned authors used (they used data for centers of action for period 1918-1994, based on an analysis of synoptic charts made by US Weather Service, German Weather Service, which were interpolated in 10°x5° grid, while we used NCEP-National Centers for Environmental Prediction / NCAR-National Center for Atmospheric Research data interpolated in 2.5° x 2.5° grid).

Table 2. Results of multiple linear regression for moving decadal values of Atlantic centers of action and precipitation amount in Serbia

Predictors	F statistic	t statistic and p level	R^2 (%)	Adjusted R^2 (%)
LoIL	80.454 p=0,001	t= 4.67495 p=0,001	92,524	91,374
LaIL		t= 11.33030 p=0,001		
IIL		t= 14.29447 p=0,001		
LoAH		t=-6.17389 p=0,001		
LaAH		t=-4.24611 p=0,001		
IAH		t=-6.14963 p=0,001		

When the dependent variable is amount of precipitation in Serbia, it is interesting to notice that all the elements of centers of action (their intensities, latitudes and longitudes) are incorporated in regression model (see Table 2). According to adjusted coefficient of determination[5] 91% of variance of moving decadal values of precipitation amount could be explained by changes in the mentioned elements of centers of action. Here is worth to mention that according to correlation matrix (Table 1) some predictors i.e. longitude of Iceland low, longitude and latitude of Azores high show no correlation with precipitation amount in Serbia. Still, when these predictors are incorporated in regression model, their contribution is significant. Obviously, they act like suppressor variables. It can be concluded that beside crucial influence of intensity and

[5] The adjusted coefficient of determination is coefficient of determination after adjusting for the degrees of freedom lost in calculation of regression coefficients.

latitude of Iceland low on precipitation amount in Serbia, there is an indirect effect of longitude of this center of action and also position and intensity of Azores high as well.

When the dependent variable is air temperature in Serbia (Table 3), by usage of stepwise regression we depicted three significant predictors (latitude and intensity of Iceland low, and latitude of Azores high). About 74% of air temperature variance could be explained by these three predictors. It is interesting to notice that longitude of Azores high was not incorporated in regression model, although it has the highest correlation coefficient with air temperature in Serbia. This fact indicates that some of the predictors are acting like mediator variables which affect relation between longitude of Azores high and air temperature in Serbia.

By experimenting with a number of multiple linear regression models it is concluded that connection between longitude of Azores high and air temperature in Serbia depends on combination of latitudes of Iceland low and Azores high.

Table 3. Results of multiple linear regression for moving decadal values of Atlantic centers of action and air temperature in Serbia

Predictors	F statistic	t statistic and p level	R^2 (%)	Adjusted R^2 (%)
LaIL	44.098 p=0,001	t= -8.61990 p=0,001	75,903	74,181
IIL		t= -3.11556 p=0.003		
LaAH		t=9.58950 p=0,001		

Conclusions

Analysis of the latitude, longitude and intensity of Atlantic centers of action data showed that:

Air pressure rise in Azores high when it moves toward northeast;

When Iceland low is moving toward southeast it becomes less intense;

Pressure rise in Azores high is connected with pressure fall in Iceland low and its moving toward northwest;

Decrease of latitude of the Iceland low is followed by move of Azores high toward south - increase of latitude of the Azores high is followed by move of the Iceland low toward north.

The highest correlation exists between precipitation amount in Serbia and intensity of Iceland low and NAO index. With increase of pressure in Iceland low, weakening of NAO index, precipitation amount in Serbia rises. All the elements of centers of action (their intensities, latitudes and longitudes) are incorporated in regression model and contributing significantly in explaining 91% of variance of moving decadal values of precipitation amount in Serbia. This indicates that beside crucial influence of the intensity of Iceland low on precipitation amount in Serbia, other predictors (latitude and longitude of Iceland low, and position and intensity of Azores high) have indirect influence as well.

On the other hand, the highest correlation exists between position of Azores high and air temperature in Serbia (moving of Azores high toward northeast is connected with air temperature increase). Air temperature also shows high correlation with the latitude of Iceland low. However, it is shown that correlation between air temperature in Serbia and longitude of Azores high is in dependence of latitudes of Iceland low and Azores high.

Acknowledgements
This paper is the result of the Project 47007 supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia

References

Baron, R. M., Kenny, D. A. (1986). The Moderator-Mediator Distinction in Social Psychological Research: Conceptual, Strategic and Statistical Considerations. *Journal of personality and Social Psychology*, Vol 51, No 6, pp. 1173-1182.

Barry, R., Chorley R. (2003). *Atmosphere, Weather and Climate*. Eight Edition Routledge, London. pp. 1-421.

Domonkos, P. (2006). Application of objective homogenization methods: Inhomogeneities in time series of temperature and precipitation. *Időjárás*, vol. 110, no. 1: 63-87.

Krus, D. J., Wilkinson, S. M. (1986). Demonstration of properties of a suppressor variable. *Behavior Research Methods, Instruments and Computers*, 18, pp. 21-24.

Machel, H., Kapala, A., Flohn, H. (1998). Behaviour of the centres of action above the Atlantic since 1881. Part I: Characteristics of seasonal and interannual variability. *Int. J. Climatology, 18, pp.* 1-22.

Milovanović, B., Radovanović M., Ducić V. (2009). Udvojeni sistem okean-atmosfera – povezanost temperature vode subpolarnog Atlantika, Islandskog minimuma i temperature vazduha u Srbiji. *Glasnik Srpskog geografskog društva, LXXXIX*, str.165-175.

Milosavljević, M. (1963). *Klimatologija.* Naučna knjiga, Beograd, 1-259.

Pandey, S., Elliot, W. (2010). Suppressor Variables in Social Work Research: Ways to Identify in Multiple Regression Models. *Journal of the Society for Social Work and Research*, Vol 1, No 1, pp. 28-40.

Preacher, K. J., Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods,* 40/3, 879-891.

Radinović Đ. (1981). Vreme i klima Jugoslavije. Građevinska knjiga, Beograd, str. 1-423.

Štěpánek, P. (2005). AnClim - software for time series analysis [computer software]. Dept. of Geography, Fac.of Natural Sciences, MU, Brno.

Tošić, I. (2004). *Analiza vremenskih serija temperature i padavina*. Doktorska disertacija. Fizički fakultet – Institut za meteorologiju, Univerzitet u Beogradu, str. 1-163.

Tzelgov, J., Henik, A. (1991). Suppression Situations in Psychological Research: Definitions, Implications, and Applications. *Psychological Bulletin,* Vol. 109, No. 3, 524-536.

Unkašević, M., Tošić., I, Vujović, D. (2004). Variability and probability of annual and extreme precipitation over Serbia and Montenegro. *Theor. Appl. Climatol.* 79, pp.103–109.