International Journal of Agriculture Sciences
ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 10, Issue 18, 2018, pp.-7145-7146.
Available online at https://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000217

Research Article

HIERARCHICAL POLYNOMIAL REGRESSION MODELS - CONSTRUCTIONS AND COMPARISONS

PRAMIT PANDIT*, MANJUNATH V. AND KRISHNAMURTHY K.N.
Department of Agricultural Statistics, Applied Mathematics and Computer Science, University of Agricultural Sciences, Bengaluru, Karnataka, 560 065, India
*Corresponding Author: Email - pramitpandit@gmail.com

Received: September 15, 2018; Revised: September 25, 2018; Accepted: September 26, 2018; Published: September 30, 2018

Abstract: A model is called to be a hierarchical polynomial regression model if all the lower order terms are present along with the highest order term(s). These models plays very significant role for the purpose of reparameterizations, Simplification in writing computer programs for polynomial model development and restricting our focus on few well-formulated models instead of all possible regression models. By the methods of stepwise regressions, backward elimination and forward selection, hierarchical polynomial regression models have been constructed.

Keywords: Construction of models, Hierarchical polynomial regression models, Forward selection method, Backward elimination method, Regression Models

Citation: Pramit Pandit, *et al*., (2018) Hierarchical Polynomial Regression Models - Constructions and Comparisons. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 10, Issue 18, pp.- 7145-7146.

Introduction

A regression model is said to be k^{th} order polynomial regression model if the highest power present among all predictors (or if they are in multiplicative form, then sum of their powers) in the polynomial regression model is k. This k is an integer and known as the order of the polynomial regression model [1]. If in a polynomial regression model, all the lower order terms are present along with the highest order term(s), then the model is called hierarchical polynomial regression model [1]. If not *i.e.* any of the lower order term is missing, then the model is considered as non-hierarchical. Hierarchical polynomial regression models are of prime importance in the field of regression analysis for the purpose of reparameterizations [2, 3], Simplification in writing computer programs for polynomial model development [4] and restricting our focus on few well-formulated [3] models instead of all possible regression models. Stepwise regressions well as one-step procedures like backward elimination and forward selection can be modified easily with a view to restrict the search for hierarchically well-formulated models [5, 6].

Construction Procedures of Hierarchical Polynomial Regression Models: Construction of hierarchical polynomial regression models can be done by three approaches: by means of forward selection method, backward elimination method and stepwise method. All three methods have been demonstrated for single variable and more than one variable, respectively.

By means of Forward Selection Method

By means of Forward Selection Method for Single Variable:

Step-1: No regressor is present in the model except intercept term.

Step-2: Check whether the partial F statistic value of the lowest order term is higher than the F-In value ($F_{\alpha,1,res\ df}$) or not. Include that lowest order term if partial F statistic value>F- In value. Otherwise, terminate the process by not including this term.

Step-3: If the lowest order term is included, run the same process for the next higher order term.

Step-4: Continue the process until the partial F statistic value of the term, which is not included in the model and order of that is just higher than the highest order term present in the model at that particular stage, does not exceed F- In($F_{\alpha,1,res\ df}$) value.

Illustration showing direction of variable selection:

$1 \longrightarrow X \longrightarrow X^2 \longrightarrow X^3 \longrightarrow \ldots$

By means of Forward Selection Method for More than One Variable:

Step-1: No regressor is present in the model except intercept term.

Step-2: All possible models with same lowest order terms are considered and partial F-statistic value for each term is computed. Among those terms, the term having highest partial F statistic value is included in the model if partial F statistic value> F- In($F_{\alpha,1,res\ df}$) value. Otherwise, terminate the process by not including that term.

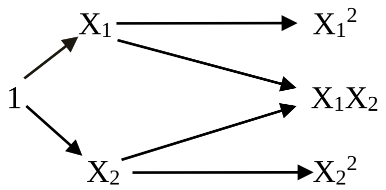
Step-3: Now, the partial F statistic values are computed for all the remaining lowest order terms in the presence of previously selected term and the term yielding highest partial F statistic value is included in the model if its partial F statistic value> F- In($F_{\alpha,1,res\ df}$) value. Otherwise, terminate the process by not including that term.

Step-4: Continue the process until the highest partial F statistic value of the term, which are not included in the model <F- In value or all the terms of lowest order are included in the model.

Step-5: If all the terms of lowest are already in the model, run the same process for all the terms of next higher order.

Step-6: Continue the process until the partial F statistic value of the term, which is not included in the model and order of that is just higher than the highest order term present in the model at a particular stage, does not exceed F- In ($F_{\alpha,1,res\ df}$) value.

Illustration showing direction of variable selection (Taking two variables, upto 2nd order):



By means of Backward Elimination Method

By means of Backward Elimination Method for Single Variable

Step-1: Start with full model *i.e.*, with all possible higher order terms. But is recommended not to include terms of too much higher order in order to avoid extreme ill conditioning of the X'X matrix due to high multicolinearity [2].

Step-2: Check whether the partial F statistic value of the highest order term is lower than the F-Out value ($F_{\alpha,1,res\ df}$) or not. Exclude that highest order term from the model if its partial F statistic value <F-Out value. Otherwise, terminate the process by keeping that term.

Step-3: If the highest order term is excluded, run the same process for the next lower order term.

Step-4: Continue the process until partial F statistic value of the highest order term present in the model at that particular stage>F-Out value or there is no term(except intercept) present in the model *i.e.* when the lowest order term is also excluded from the model.

Illustration showing direction of variable elimination:

$1 \longleftarrow X \longleftarrow X^2 \longleftarrow X^3 \longleftarrow \ldots$

By means of Backward Elimination Method for More than One Variable

Step-1: Start with full model *i.e.*, with all possible higher order terms of the variables. However, it is recommended not to include terms of too much higher order in order to avoid extreme ill conditioning of the X'X matrix due to high multicolinearity.

Step-2: Partial F-statistic value for all the terms of same highest order are computed. Among those terms, the term having lowest partial F statistic value is excluded from the model if its partial F statistic value < F-Out ($F_{\alpha,1,res\ df}$) value. Otherwise, terminate the process by keeping that term.

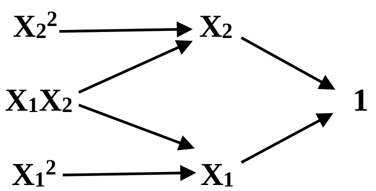
Step-3: Now, the partial F statistic values are computed for all the remaining highest order terms in the absence of previously excluded term and the term yielding lowest partial F statistic value is excluded from the model if its partial F statistic value< F-Out($F_{\alpha,1,res\ df}$) value. Otherwise, terminate the process by keeping that term.

Step-4: Continue the process until the lowest partial F statistic value of the terms present in the model >F-Out value or all the terms of highest order are excluded from the model.

Step-5: If all the terms of highest order are already excluded from model, run the same process for all the terms of next lower order.

Step-6: Continue the process until partial F statistic value of the highest order term present in the model at that particular stage>F-out value or there is no term (except intercept) present in the model *i.e.*, when all the lowest order terms are also excluded from the model.

Illustration showing direction of variable elimination (Taking two variable, from 2nd order) :



By means of Stepwise Regression Method

By means of Stepwise Regression Method for Single Variable: For a single variable, it is not possible to go for stepwise selection method. Either it will be like just as forward selection method of variables (with comparing unnecessarily partial F values with F-out values for exclusion of variables) or it may select one of those not well-formulated models as the best overall model after examining a larger number of models.

By means of Stepwise Regression Method for More than One Variable

For more than one variable case also, we cannot directly go for stepwise selection procedure of variables. We can meaningfully apply this algorithm only among the terms of same highest order. If all the variables of lower order terms are already included in the model, then only for all the terms of the highest order can be selected by stepwise selection method. Significance of all lower order terms can be directly checked by backward elimination method. If partial F statistic value of all those terms are higher than F-Out value, then only we go for checking the significance of the highest order terms.

Conclusion

Relative Comparison among the 3 Methods of Construction

When we want to include only few relevant terms, it will be wiser to go for forward method of selection by taking desirable F- In value. However, if we have some idea about the highest order terms of the polynomial from previous studies or graphs, it will be wiser to go for backward elimination method starting from the highest order term(s) of the variable(s).Here we can keep more number of informative term by choosing suitable F- Out value. When we know exactly the about the highest order of the polynomial, then among the terms of highest order, use of stepwise method will be wiser.

Application of research: Hierarchical polynomial regression models remain invariant under linear transformation. In the field of agriculture, models constructed hierarchically will be helpful in modelling crop and animal systems, which need linear transformations.

Research Category: Agricultural Statistics

Acknowledgement / Funding: Author thankful to Department of Agricultural Statistics, Applied Mathematics and Computer Science, University of Agricultural Sciences, Bengaluru, Karnataka, 560 065, India.

*** Research Guide or Chairperson of research: Dr K. N. Krishnamurthy**

University: University of Agricultural Sciences, Bengaluru, Karnataka, 560 065

Research project name or number: M.Sc. Research

Author Contributions: All author equally contributed

Author statement: All authors read, reviewed, agree and approved the final manuscript

Conflict of Interest: None declared

Ethical approval: This article does not contain any studies with human participants or animals performed by the author.

References

[1] Montgomery D.C., Peck E.A. and Vining G.G. (2017) *Introduction to Linear Regression Analysis (3rd ed.): Wiley India Pvt. Ltd*.

[2] Griepentrog G.L., Ryan J.M. and Smith L.D. (1982) *The American Statistician*, 36, 171- 174.

[3] Peixoto J.L. (1986) *Communications in Statistics-Theory and Methods*, 15, 1957-1973.

[4] Peixoto J. L. and Diaz J. (1966) *Journal of the Inter-American Statistical Institute*, 48(150-151),175-210.

[5] Peixoto J.L. (1987) *The American Statistician*, 41(4), 311-313.

[6] Peixoto J.L. (1990) *The American Statistician*, 44(1), 26-30.