

Biodiversity Phenotypic and Genotypic Polymorphism Data Correlation Analysis Using SPSS 16.0 Software

Soleiman Jamshidi, and Nasser Mohebbalipour

Abstract—Biodiversity studies depend on phenotypic, geographic and molecular data collected from different sources of certain species. These data dependence and correlation analysis could give valuable information in terms to qualitative, quantitative traits, leading to reliable and meaningful facts about the species. Also, finding dependence between molecular and phenotypic traits will also help us to understand the genetic and phenotypic relationships in populations. To statistical analysis of these data we simply reviewed SPSS 16.0 application as a simple and strong statistic relevant software regarding polymorphism's data obtained from different phenotypic observations and genetic experiments. We also mentioned how the outputted data obtained from these analyses may be interpreted, also going examples.

Keywords—Correlation, molecular data stepwise multiple regression, quantitative and qualitative traits.

I. INTRODUCTION

IN statistics, dependence refers to any statistical relationship between two random variables or two sets of data. Correlation refers to any of a broad class of statistical relationships involving dependence [3]. In biodiversity as the degree of variation of life forms within a given species, ecosystem, biome, or an entire planet [5] finding a correlation among the studied traits might be very informative in understanding biodiversity phenomena. The correlations procedure calculates the correlation between variables and is used to measure the strength of linear association between two variables [3]. When one variable moves in the same direction, then it is called positive correlation. When one variable moves in a positive direction, and a second variable moves in a negative direction, then it is said to be negative correlation [2]. In such studies, statistically significant correlations will show how the traits could change depending on the others. Also, the estimates of genetic and phenotypic parameters are essential in determining the methods of selection to be employed and in assessing the genetic gains in performance traits [1]. Knowledge of relationship between genetic and phenotypic data is valuable when related traits are considered for selection. If genetic correlation among two traits is positive and high, the selection for one trait would result in an improvement of the other trait [6]. In statistics, stepwise

regression includes regression models in which the choice of predictive variables is carried out by an automatic procedure. It is an estimation of the linear relationship between a dependent variable and one or more independent variables or covariates [9]. SPSS (originally, Statistical Package for the Social Sciences) is a computer program used for survey authoring and deployment (IBM SPSS Data Collection), data mining (IBM SPSS Modeler), text analytics, statistical analysis, and collaboration and deployment [10]. It provides facilities for analysing and displaying information using a variety of techniques [2,10]. SPSS versions 16.0 and later run under Windows, Mac, and Linux [11]. It can be used to analyse statistics included bivariate statistics such as means, t-test, ANOVA, correlation (bivariate, partial, distances) and nonparametric tests also [2,11]. The paper aims how to use the software for correlation analysis of biodiversity various data obtained from different observation and experiments, including phenotypic and molecular.

II. PROCEDURE FOR PAPER SUBMISSION CORRELATION IN PHENOTYPIC DATA

A. Quantitative data set

A quantitative trait shows continued variation. This is because the trait is the sum of several small effects caused by the gene. An example of this is an animal's metabolism, which is under the influence of many different genes. The final products of the metabolism, as for instance milk yield or growth rate, are good examples of quantitative traits [5].

If there are two or more quantitative data sets, follow the steps below:

- Run the program SPSS 16.0 for Windows® after installation.
- Click on cancel button if “what would you like to do” dialog box is appeared.
- Click on “variable view” tab on bottom left side and define name for your variables and specify type, width, decimals and etc. if it is needed.
- Return to “data view” then copy and paste your data from your Excel data sheet.
- Then click “Bivariate” (Fig. 1).
- In “Bivariate Correlations” dialog box, transfer all data from the left to the right box (variables box) by selecting them and clicking on transfer button (Fig. 2).
- In “Correlation Coefficients” make sure the box beside “Pearson” has been ticked (Fig. 2).

Soleiman Jamshidi is with the Plant Protection Department, Islamic Azad University, Miyaneh Branch, Miyaneh, Iran

Nasser Mohebbalipour is with the Plant Protection Department, Islamic Azad University, Miyaneh Branch, Miyaneh, Iran

- Click Flag significant correlations to put a tick in the box. (It might already be ticked)
- Click “OK” and catch the correlation results as “output1”.

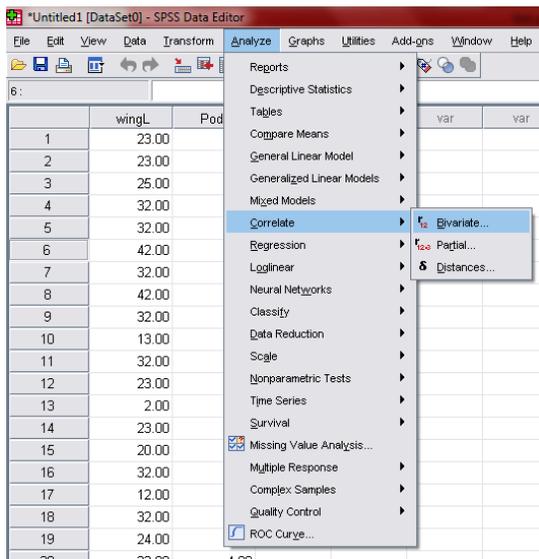


Fig. 1 Correlation data analysis procedure in SPSS 16.0

To interpret your data, consider that if you find superscripted one or two star (**/*) beside data in the table, the correlation is significant at 0.05 and 0.01 probability level ($p < 0.05$ and $p < 0.01$), respectively. If you found nothing, it means that there is no statistically significant correlation between two data corresponding sets. The correlation might be negative or positive.

The quantitative data could be morphological traits or geographic characteristics. For example, there might be negative correlation between altitude and horn length of an animal. It means animals with shorter horn are living in higher lands. The correlation might be found among phenotypic traits with ecosystem and ecological data too [3,4,5] and in some cases the predictions will be possible of the phenomena, too [4,5].

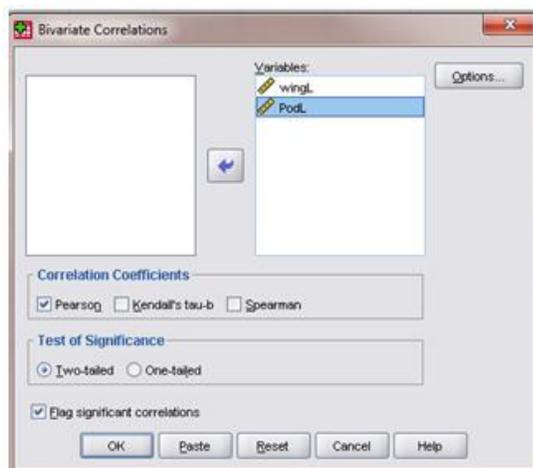


Fig. 2 Bivariate Correlation dialog box

B. Qualitative data set

Qualitative data come from quantitative traits. A qualitative trait is expressed qualitatively, which means that the phenotype falls into different categories. These categories do not necessarily have a certain order. The pattern of inheritance for a qualitative trait is typically monogenetic, which means that the trait is only influenced by a single gene. Inherited diseases caused by single mutations are good examples of qualitative traits. Another is blood type. The environment has very little influence on the phenotype of these traits [5].

To find possible correlation among qualitative data set, simply follow all steps mentioned in 2.1. section but simply tick “Kendall’s tau” in “Correlation Coefficients” box instead of “Pearson” (Fig. 2). The other steps and interpretation way would be the same.

➤ **Important 1:** Before analysis you should convert qualitative data to data similar to quantitative. For example if you have five colours between black and white, you may allocate 1 for white and 7 for black and convert all borderline colours to given numbers from 2 to 6.

Data interpreting is the same. For example, you may find a negative correlation between colour and dens of fur in polar bears. The meaning will depend on how you arrange qualitative data and convert it to quantitative one.

C. Between qualitative and quantitative data

To find any correlation between qualitative and quantitative data set, simply follow all steps mentioned in 2.1. section but just tick “Spearman” in “Correlation Coefficients” box instead of “Pearson”. The other steps and interpretation way would be the same.

➤ **Important 2:** According to SPSS user guide, if the variables in correlation analysis are ordinal in measurement scale, it would have been statistically more proper to use Spearman’s Rank correlation coefficients to measure the relationship.

➤ **Important 3:** As SPSS text book advices, if the two variables are continuous, the Pearson product moment correlation is an appropriate measure. If they are not continuous (that is, if they are discrete or categorical), it would be more appropriate to use Spearman’s rho or Kendall’s tau-b.

III. CORRELATION BETWEEN MOLECULAR AND PHENOTYPIC DATA

A genetic correlation is a description of the relationship between the sets of genes which are responsible for the genetic part of variance in the corresponding traits. Study of correlation between molecular and morphological has several applications. Technical application like morphological traits prediction using molecular data is one of the samples. Also, if we evaluate a germplasm from a species regarding important traits for special condition using molecular markers, we will be able to use these findings to generalize for other germplasm in certain condition. It will be possible to recognize special desire genome in laboratory condition [8].

One of the ways to find a correlation between molecular

data such as RAPD, SSR, and ISSR electrophoretic bands and phenotypic data is using the multiple stepwise regression method by SPSS 16.0. In this method, the phenotypic data will be considered as dependent variable and the molecular data will be heeled as independence one.

The procedure will be like steps below:

- Run the program SPSS 16.0 for Windows®.
- Click on “variable view” tab on bottom left side and define name for your variables and specify type, width, decimals and etc. if it is needed. You may consider the name of genotypes, isolates or individuals and from the second variable would allocate to phenotypic data and for each electrophonic band you should consider a variable.
- Return to “data view” and copy and paste your data from your Excel data sheet.
- Click “data view” after inserting variables and copy and paste the data from excel file (Fig. 3).
- **Important 4:** Each column should allocate to an electrophoretic band, you should notice the starting and ending column of each primer.
- Click “Analyze” tab and choose “Regression” then click “Linear” (Fig. 3).

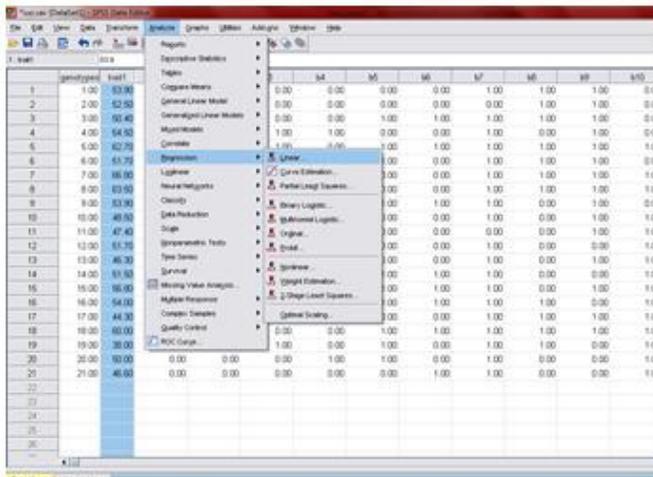


Fig. 3 Data view and multiple stepwise regression data analysis procedure in SPSS 16.0

- In “Linear Regression” dialog box, transfer first phenotypic variable to “Dependent” box and all molecular data to “Independent(s)” box by selecting it and clicking on transfer button.
- Click on options and in “Linear Regression Options” dialog box and in “Use probability of F” part insert 0.05 in “Entry” and 0.01 in “Removal” box and click “continue” (Fig. 4)
- Click “OK” and see the correlation results as output file.
- Repeat this procedure with all phenotypic traits one by one.

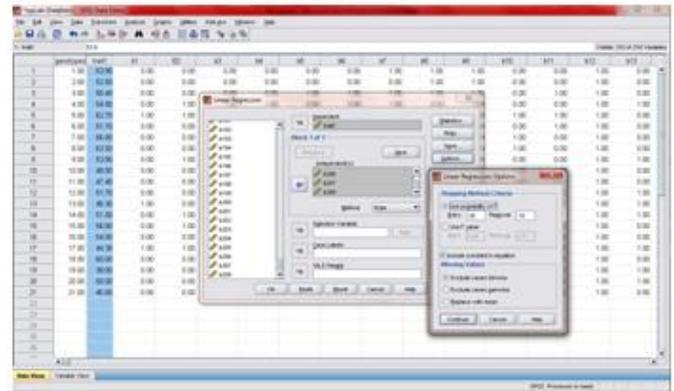


Fig. 4 Multiple stepwise regression data analysis procedure in SPSS 16.0

To interpret data you need to see which of electrophoretic bands had significant correlation with phenotypic traits. There will be easy to clarify which primers have the more correlation with each trait (important 4). Summarize data in a table like the sample table below (Table 1).

TABLE 1
REGRESSION COEFFICIENTS IN MULTIPLE REGRESSION BETWEEN PHENOTYPIC AND ISSR BANDS

Marker-band	Regression coefficient							
	Quantitative traits				Qualitative traits			
	LAL	LML	MAL	DAL	IGP	CC	ME	PE
Constant	156.9	17.5	6.6	16.9	2.0	2.8	0.50	-0.4
(GTG) _{S-2}		5.2						-0.2
(GTG) _{S-3}			0.433		-0.4			
(GTG) _{S-7}	22.0							
M13-12								0.4
M13-14								0.2
M13-16				-1.3				
M13-23					-0.4	-0.533		
Adjusted R ²	0.05	0.05	0.07	0.06	0.08	0.13	0.04	0.17
Informative markers no.	1	1	1	1	2	1	1	3

ACKNOWLEDGEMENTS

We deeply express our thanks to Dr. Reza Azizi-nejad for his fruitful advices in statistics.

REFERENCES

- [1] A. Belisario and M. Hubbes. Genetic variability in *Gnomonia leptostyla*. Acta Horticulture. 1992, 442, pp. 385-387.
- [2] Anonymous. SPSS Step-by-Step, Tutorial: Part 2, For SPSS Version 11.5. 2004. pp. 31-34.
- [3] D.D. Mari and S. Kotz. Correlation and dependence. World Scientific Publishing Company. 2011, pp. 128-140.
- [4] L.T. Waser, M. Kuechler, M. Schwarz, S. Stofer, Ch. Scheidegger, E. Ivits, B. Koch. Correlation of remote sensing data with lichen field samples, prediction of biodiversity. The XXII Congress of the International Society for Photogrammetry and Remote Sensing. 25 August - 1 September. 2012. pp. 1-6.
- [5] M.W. Schwartz, C.A. Brigham, J.D. Hoeksema, K.G. Lyons, M.H. Mills and P.J. van Mantgem. Linking biodiversity to ecosystem function: implications for conservation ecology. Oecologia. 2000, 122, pp. 297-305.
<http://dx.doi.org/10.1007/s004420050035>
- [6] N.A. Naz, and M. Ahmad. Genetic and phenotypic correlations for some sexual maturity traits in nili ravi buffalo heifers. Pakistan Veterinary Journal. 2006, 26(3), pp. 141-143.
- [7] P. J. Folgarait. Ant biodiversity and its relationship to ecosystem functioning: a review. Biodiversity and Conservation, 1998, pp. 1221-1244.
<http://dx.doi.org/10.1023/A:1008891901953>

- [8] M. Shokr-pour, S.A. Mohammadi, M. Moghaddam, S.A. Ziaei and A. Javan-shir. Analysis of correlation between morphologic and biochemical and molecular AFLP markers in *Silybium marianum* L. Iranian Medicinal Plants Researches Journal, 2008. 24 (3), pp. 278-292.
- [9] SAS Institute Inc. SAS/STAT User's Guide, Ver. 6" 4th Edition, Vol. 2, Cary, NC: SAS Institute Inc. 1989.
- [10] University of Bristol. Introduction to SPSS (version 16) for Windows, Practical workbook. University of Bristol Information Services document spss16-2. Retrieved on 28 July 2012 from <<http://www.bristol.ac.uk/it-services/learning/documentation/spss16-2/spss16-2t.pdf>>. 2010. pp.55-56.
- [11] Wikipedia. SPSS. Retrieved on 28 July 2012 from <<http://en.wikipedia.org/wiki/SPSS>>.

Soleiman Jamshidi was born in Miyaneh, Iran in 21 March, 1974. He received his B.Sc. degree in plant protection from Tabriz University in Tabriz, Iran, M.Sc. degree in plant pathology from Ferdowsi University of Mashhad, Mashhad, Iran, and Ph.D. in plant pathology and fungal diseases filed from Islamic Azad University, Science and Research Branch of Tehran, Tehran, Iran.

He started working as academic staff since 2001 in Islamic Azad University, Miyaneh Branch and was Vice-President in Research Affairs in this university for 4 years starting from 2003 to 2007 and Vice-President in Education and Student Affairs of Agriculture and Natural Science Faculty in the same university in 2012. He is now a permanent Academic Staff and member of Research Affairs and Publication Councils and Young Research and Elite Club, Islamic Azad University, Miyaneh Branch's member. His publications include 25 journal papers, 22 international and 42 national conferences papers. His research interest is fungal phylogeny and plant disease management.

Dr. Jamshidi is member of Iranian Plant Pathologist and Iranian Mycologist Society member and senior member of the Asia-Pacific Chemical, Biological & Environmental Engineering Society (APCBEEES), ID: 100236].

Nasser Mohebalipour was born in Hashtroud, Iran in 21 September, 1980. He received his B.Sc. degree in agronomy and plant breeding from Mohaghegh Ardebili University in Ardebil, Iran, M.Sc. degree in plant breeding from University of Tabriz, Tabriz, Iran, and Ph.D. in plant breeding (molecular genetics and genetic engineering) from University of Tabriz, Tabriz, Iran.

He started working as academic staff since 2006 in Islamic Azad University, Miyaneh Branch. He is now a permanent Academic Staff and member of Young Research and Elite Club, Miyaneh Branch, Islamic Azad University. His publications include 11 journal papers, 1 international and 40 national conferences papers. His research interest is crop plant breeding, medicinal plant breeding and plant tissue culture.

Dr. Mohebalipour is member of Iranian Agronomy and Plant Breeding Society.