The Graphical Assessment of Multivariate Normality Using SPSS

Wan Nor Arifin

Unit of Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia

ABSTRACT

Multivariate analyses depend on multivariate normality assumption. Although the analyses are available in SPSS, it is not possible to assess the assumption from the basic package. Statistical assessment of the normality is available in a specialized package, SPSS Amos, in form of Mardia's multivariate kurtosis. However, graphical assessment of the normality by chi-square versus Mahalanobis distance plot is not available in both of the packages. The aim of this article is to present the steps to construct the plot in SPSS in a point-and-click manner as expected by most SPSS users.

CORRESPONDING AUTHOR: Dr Wan Nor Arifin, Unit of Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia.
Email: wnarifin@usm.my

Introduction

A multivariate analysis refers to analysis in which there are multiple dependent variables [1], with or without independent variable(s). Among commonly used multivariate analyses are exploratory factor analysis and multivariate analysis of variance (MANOVA). The analyses are dependent on the assumption of multivariate normality of relevant variables [1-3].

Statistical assessment of multivariate normality is available in a number of statistical packages, for examples in SPSS Amos (Mardia's multivariate kurtosis) and R (Mardia's, Royston's and Henze-Zirkler's multivariate normality tests via MVN package). Graphical assessment of the normality by chi-square versus Mahalanobis distance plot is not directly available in SPSS and SPSS Amos, although the packages are capable of running multivariate analyses. In order to obtain the plot in SPSS, a number of steps are required as presented by Burdenski [3] in form of SPSS syntax, which was based on MULTINOR program [4]. The aim of this article is to present the steps to construct chi-square versus Mahalanobis distance plot in SPSS in a point-and-click manner as expected by most SPSS users.

Chi-square versus Mahalanobis distance plot

For the purpose of following the steps as given in this article, the readers may download a dataset namely **Attitude_Statistics.sav** from this journal's website. The dataset consists of a variable namely **ID** that represents IDs of participants and 12 variables that represent responses for 12 questions on attitude towards statistics. The questions are rated on a Likert scale of 1 to 5. In total, there are 150 cases in the dataset.

Step 1. Obtain Mahalanobis distances for the cases. The values are obtained by running a multiple linear regression analysis with a dummy variable as the dependent variable and variables that we wish to check for multivariate normality as the independent variables. From menus, choose **Analyze → Regression → Linear**. In the **Linear Regression** dialog box, set **ID** as the dependent variable and **Q1** to **Q12** as the independent variables (Figure 1). As the dummy variable does not affect the values of the Mahalanobis distances, it can be replaced with any other variable. Click on **Save…** button top open up **Linear Regression: Save** dialog box. Select **Mahalanobis** option under **Distances** category followed by clicking on **Continue** button to return back to the main dialog box. Click **OK** in the **Linear Regression** dialog box. A new variable namely **MAH_1** will be created, consisting of Mahalanobis distances for the cases.
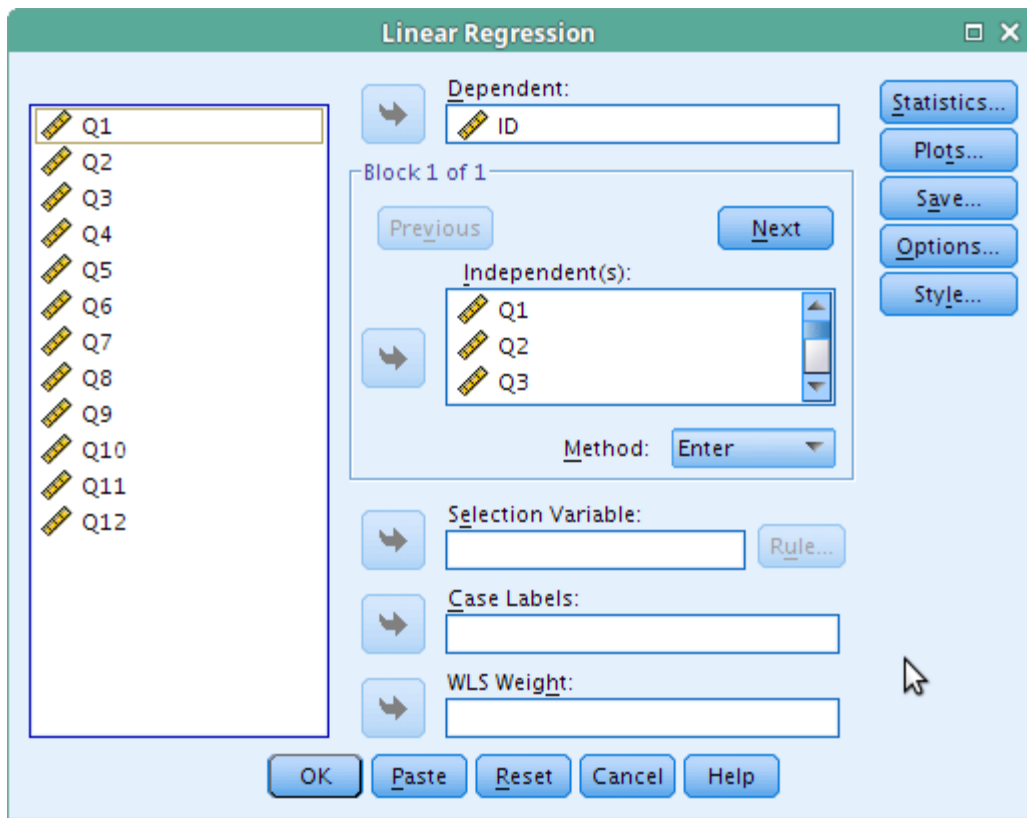

Figure 1. Running multiple linear regression.

Step 2. Sort Mahalanobis distances in ascending order. Right click on **MAH_1** column and choose **Sort Ascending** option.

Step 3. Obtain *P*-values of chi-square for the cases. The P-values equal case number after sorting minus 0.5, divided by number of cases or sample size. From menus, choose **Transform → Compute Variable**. In the **Compute Variable** dialog box, enter **P** as **Target Variable** and type ($casenum − 0.5)/150 in **Numeric Expression** text box (Figure 2). Click **OK** to obtain the *P*-values.
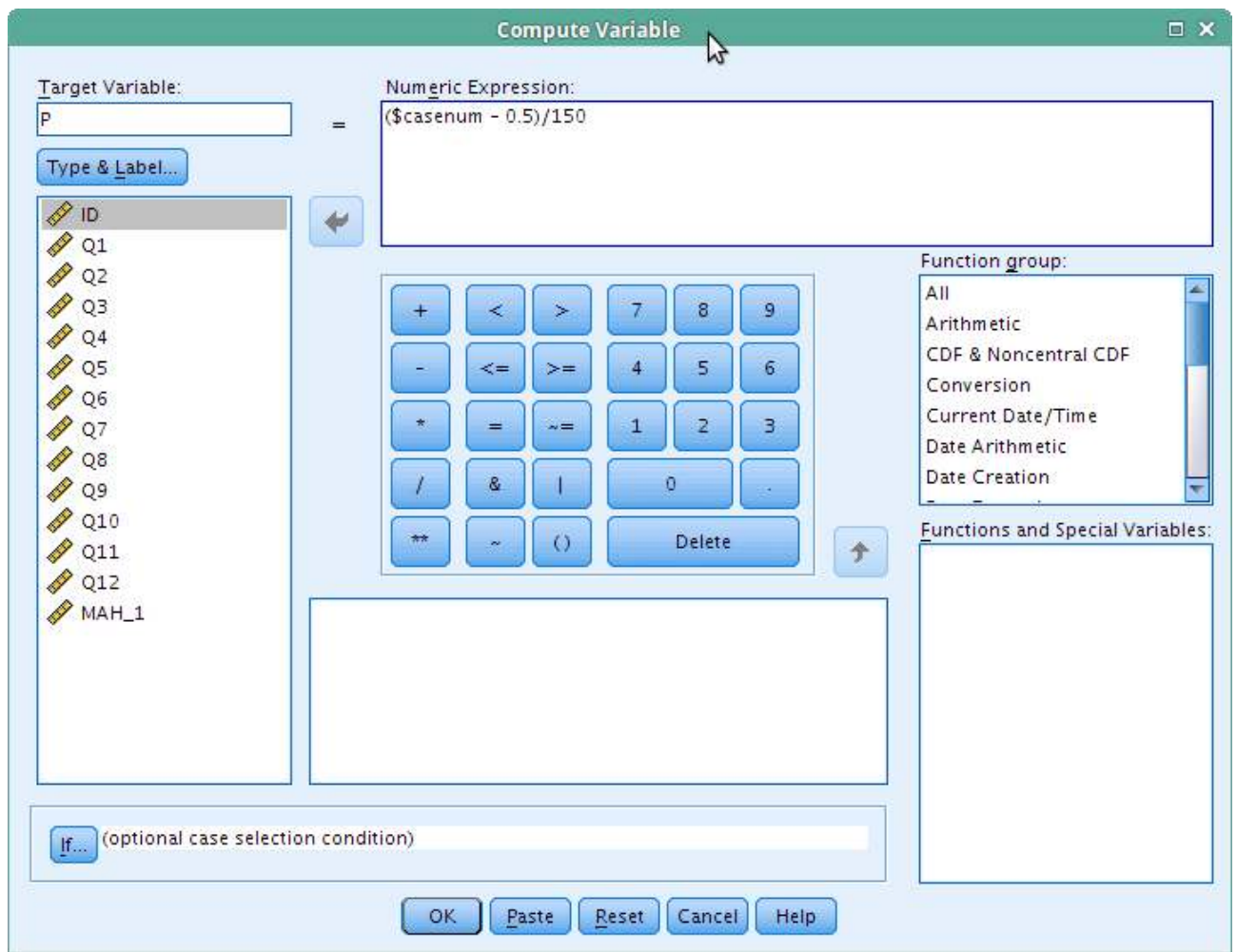
Figure 2. Creating *P*-values for chi-square.

Step 4. Obtain chi-square values for the respective *P*-values. In this step, we need to utilize inverse distribution function of chi-square distribution for specified degrees of freedom, which equal the number of variables checked for multivariate normality. Again, from menus, choose **Transform → Compute Variable**. In the **Compute Variable** dialog box, enter **CHISQ** as **Target Variable** and type IDF.CHISQ(P,12) in **Numeric Expression** text box (Figure 3). Click **OK** to obtain the chi-square values. You may also access the function from **Functions and Special Variables** list of functions.
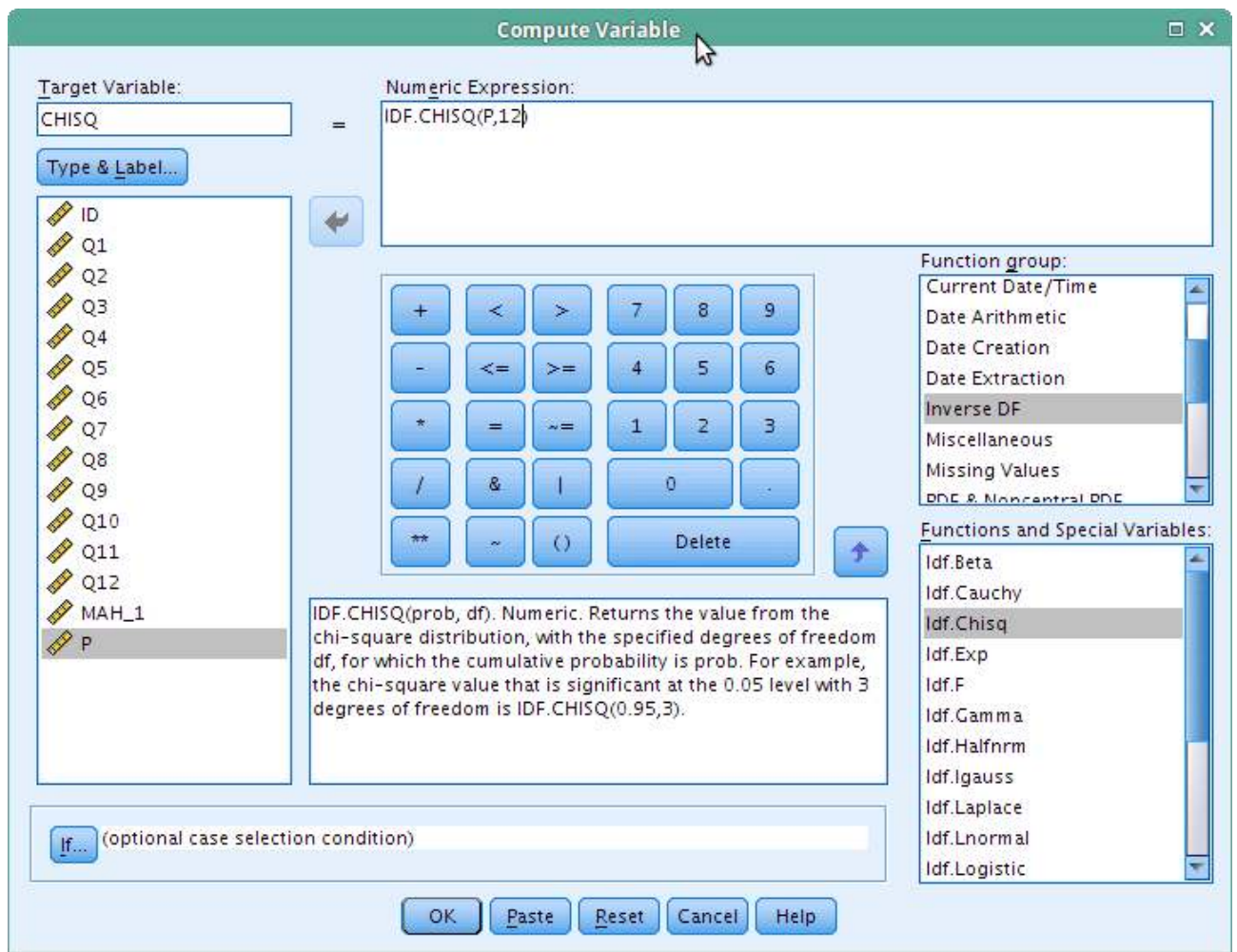
Figure 3. Obtaining chi-square values based on *P*-values.

Step 5. Graph a scatter plot of chi-square versus Mahalanobis distance. From menus, choose **Graphs → Legacy Dialogs → Scatter/Dot**. In the dialog box, choose **Simple Scatter** followed by **Define** button. Enter **CHISQ** under **Y Axis** and **MAH_1** under **X Axis**. Click **OK**. The resulting plot is shown in Figure 5. If the variables form a multivariate normal distribution, the points will form a straight line [3]. From the plot, the variables do not form a clear cut straight line and noticeably there is an outlier at upper right part of the plot. We may conclude that the variables do not form a multivariate normal distribution.
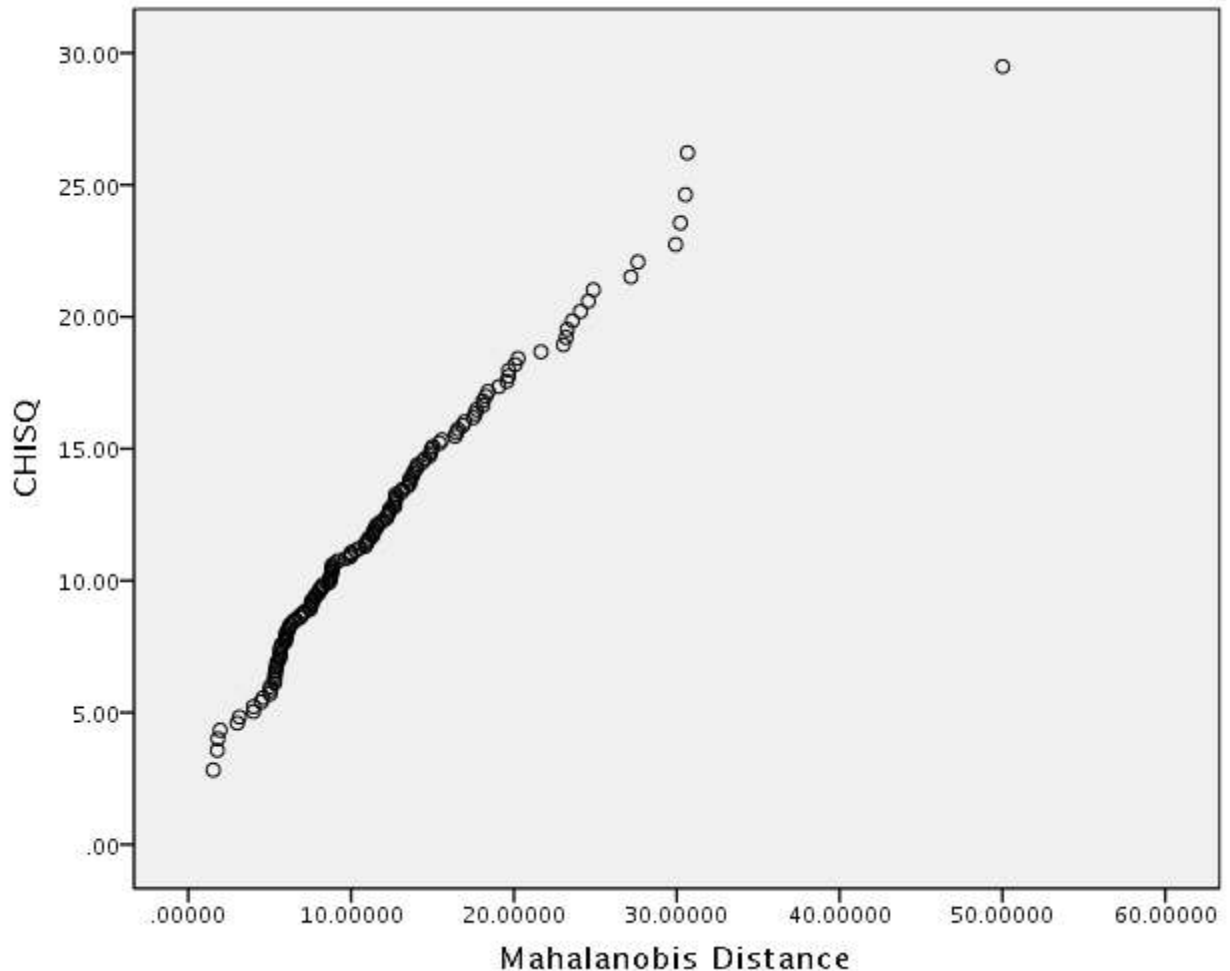
Figure 4. Chi-square versus Mahalanobis distance plot for Q1 to Q12.

Conclusion

In this article, I have shown the steps to obtain a chi-square versus Mahalanobis distance plot for graphical assessment of multivariate normality assumption in SPSS. It allows simple assessment of the normality in a familiar statistical package in the absence of a specialized package. In addition, for interested readers, you may also obtain the similar plot in R via MVN package provided that you are willing to leave your comfort zone in SPSS environment.

Reference

1. Tabachnick BG, Fidell LS. Using Multivariate Statistics, 5th ed. Boston: Pearson Education; 2007.
2. Hair Jr JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis, 7th ed. Upper Saddle River, NJ: Pearson Prentice-Hall; 2009.
3. Burdenski T. Evaluating univariate, bivariate, and multivariate normality using graphical and statistical procedures. Multiple Linear Regression Viewpoints 2000; 26(2):15-28.
4. Thompson, B. Multinor: A Fortran program that assists in evaluating multivariate normality. Educational and Psychological Measurement 1990; 50(4): 845-848.