Open Access Journal of Behavioural Science & Psychology
ISSN: 2642-0856

Academic Strive
Researcher's perception is our reality

Research Article *Volume 1; Issue 2*

Multivariate Statistical Techniques to Manage Multiple Data in Psychology

Clarissa Ferrari[1]*, Ambra Macis[1], Roberta Rossi[2] and Michela Cameletti[3]

[1]Unit of Statistics, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Italy

[2]Unit of Psychiatry, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Italy

[3]Department of Management, Economics and Quantitative Methods, University of Bergamo, Italy

***Corresponding author:** Dr. Clarissa Ferrari, Unit of Statistics, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Via Pilastroni 4, Brescia, Italy, Tel: +390303501722; Email: cferrari@fatebenefratelli.eu

Received Date: August 31, 2018; **Published Date:** September 06, 2018

Abstract

Introduction: In big-data contexts, multivariate statistical techniques and machine learning methods play a crucial role for the assessment of the interrelations between and within sets of variables. In particular, in social and behavioural sciences, for which the exploration of patterns and mutual interrelation among subject features is needed, a proper use of this technique becomes paramount.

Methods: A series of multivariate techniques –clustering, decision trees, principal component, multiple correspondence, partial least discriminate analysis –was applied to a sample of patients with diagnosis of borderline personality disorder (BPD)and bipolar disorder (BD), in order to outline specific socio-demographic and clinical profiles for both the diagnoses.

Results: Although the BPD and BD patients are clinically blurred, some features appeared to well discriminate between the two diagnoses. BPD patients are more probably females who have shown self-harm behaviours and/or suicide attempts, while BD are more likely to be males who have never shown self-harm behaviours and have not attempted suicide. Moreover, the assessment variables with more discriminate power were BIS-11, SCL-90 and STAI-T. In particular, patients with SCL-90 total score <36 were more probably BD patients (probability p=87%); whereas patients with SCL-90 score ≥ 36 and a BIS-11 score ≥64 were more probably BPD patients (p=83%).

Conclusions: The application of multivariate statistical analyses and machine learning techniques allows the definition of specific clinical and diagnostic profiles that can be crucial for taking adequately charge of the patients in a context of precision medicine and an ad-hoc diagnostic and care pattern.

Keywords: Multivariate statistics; Clustering; Association; Discriminate analysis; Borderline disorder; Bipolar disorder; Data reduction; Machine learning

Abbreviations: BPD: Borderline Personality Disorder; BD: Bipolar Disorder; PSP: Personal and Social Performance Scale; SCL-90: Symptom Checklist-90-R; HAM-D: Hamilton Depression Rating Scale; TAS: Toronto Alexithymia Scale; BIS11: Barratt Impulsiveness Scale;

Citation: Clarissa Ferrari, et al. Multivariate Statistical Techniques to Manage Multiple Data in Psychology. OA J Behavioural Sci Psych 2018, 1(2): 180006.

STAI-T: State and Trait Anxiety Inventory- Trait; MCA: Multiple Correspondence Analysis; PCA: Principal Component Analysis; PC: Principal Component; PLS-DA: Partial Least Square Discriminate Analysis; PLS-R: Partial Least Square Regression; LDA: Linear Discriminate Analysis; CART: Classification and Regression Tree.

Introduction

In the last few years, due to the development of information technology and communication, huge quantities of data are being collected and stored, so that very large – both in volume and complexity- databases from observation and experimentation studies are now available in many contexts of the scientific research [1]. Consequently, attention is now being focused on new information and evidence that can be discovered from the already available huge databases. In a big-data context, multivariate statistical analyses and machine learning methods play a crucial role for an effective data analysis [2] in order to highlight the interrelationships between and within sets of variables. An adequate knowledge and use of such techniques in research disciplines, where often their application is still limited, can become a stimulus for further original and groundbreaking research, new discoveries and innovative applications.

Eligible contexts where the multivariate statistical methods can be applied fruitfully are the social and behavioural sciences [3], where the study designs usually aim to explore and discover presence of patterns and mutual (sometimes causal) interrelations among involved variables [4]. In psychology studies, where patients have undergone at a multi-assessment evaluation, the exploration of an internal structure among patients' features [5] could provide interesting tools for new data interpretation, with translational implication in the clinical practice as well. While psychologists generally think of multivariate statistics in terms of factor analysis [6] or in terms of MANOVA or multivariate regression models (and sometimes erroneously in terms of multiple regression), here we focus on multivariate statistics in terms of machine learning algorithms – including hierarchical clustering and decision tree methods- as well as data reduction and classification methods applied to different kinds (both continuous and categorical) of psychological data.

Although these methods require a slightly advanced statistical practice, their potential output in terms of, e.g., discovering of latent data structure, patients profile identifications, definition of potentially new diagnostic patterns useful for clinicians, are worth the efforts. Our aim is to make these methods as much as profitable and immediately usable by clinical researchers without a high-level statistical background. For this aim, we will provide practical applications of a series of multivariate techniques to psychological data in order to define socio-demographic and clinical profile of patients with both borderline personality disorder and bipolar disorder. Due to an overlap of many clinical and psychopathological features [7-9], such diagnoses are blurred and often single assessment instruments are not able to ensure a clear discrimination between them. The joint use and interpretation of such methods could contribute to overcome the misdiagnosis problem. In addition, in order to promote the use of such methods, detailed statistical R software code will be provided as well.

Material and Methods

Study design and participants

Patients were enrolled in a period of 4 years (2006-2010) at the IRCCS San Giovanni di Dio-Fatebenefratelli as a part of a wide study aiming at describing brain morphological features psychiatric disorder. The sample is composed of 93 psychiatric patients, 68 with diagnosis of borderline personality disorder (BPD) and 25 with bipolar disorder (BD). The study was approved by the local ethics committee and written informed consent was obtained by all subjects. No compensation was provided for study participation.

Assessment instruments

The main socio-demographic and clinical information were collected for all patients (age, sex, history of alcohol and drugs abuse, presence of suicidal attempts and of self-injurious behaviours). All patients underwent a multidimensional assessment including: Personal and Social Performance Scale (PSP) to assess psychosocial functioning [10], the Symptom Checklist-90-R (SCL-90) to evaluate general psychopathology (e.g. somatization, obsessive-compulsive, depression, anxiety, hostility) [11], Hamilton Depression Rating Scale (HAM-D) assessing depressive symptoms [12], Toronto Alexithymia Scale (TAS) [13,14] to evaluate alexithymia (i.e. inability to identify and describe emotions), Barratt Impulsiveness Scale-11 (BIS-11) [15,16] and the State-Trait Anxiety Inventory (STAI) for the assessment of impulsivity and anxiety, respectively [17].

Statistical analysis

Socio-demographic and clinical differences between the two diagnostic groups were assessed through $\chi 2$ test for categorical data and t-test for continuous data, after

having checked normality assumption by visual inspection of variable distributions (QQ-plots and box plots) and through Shapiro-Wilk and Kolmogorov-Smirnov tests. All tests were two-tailed with statistically significance level set at p=0.05.

We focused on different multivariate statistical techniques chosen on the basis of the different variable types and provided output in order to outline a socio-demographic and clinical profile for BPD and BD patients. The Multiple Correspondence Analysis (MCA) was applied to analyze the association among different modalities (categories) of a series of categorical variables. The outcome of this method was represented in a unique two-dimensional space plot showing the relationships among categories, among subjects and among categories and subjects. Categories that are in the same quadrant or that are close enough suggest an association [18].

Considering the continuous variables, three different methods were applied: the cluster analysis, the principal component analysis (PCA) and the partial least-square discriminate analysis (PLS-DA). The first was performed to aggregate subjects (in our case patients) on the basis of their individual features (variables) [19]. The aim was to apply a data-driven (hierarchical clustering) or hypothesis-driven (k-means clustering) approach in order to detect groups of subjects similar in terms of target characteristics (socio-demographic and/or clinical) to help clinicians in patient discrimination. The PCA, although mainly used as a data-reduction technique [20], was applied to derive, through the *biplot* [21], a graphical representation of the associations between the variables and the subjects in a way similar to the MCA. The PLS-DA was carried out to define which variables, among the continuous ones, contribute mainly to discriminate between the categories of the response variable (in our case diagnosis: BPD vs BD) [22-24].

Finally, a machine learning method, that allows managing both categorical and continuous variables, was carried out in a decisional context. In detail, the classification and regression tree (CART) method was carried out on the diagnosis, as categorical dependent variable (to be predicted), depending on a set of socio-demographic and clinical variables (categorical and/or quantitative covariates) [25]. The output of the CART is given by different classification pathways (defined by estimated covariate cut-offs) and, for each of them, the probability, corresponding to the most likely dependent variable category, is given. Details and specifications for all the multivariate methods are reported in the Supplementary materials s- Methods.

All data were analyzed using R software [26]. R-packages and code for applying the multivariate techniques are reported in the Supplementary material s-Code.

Results

Sample description

Patients with BPD differed from patients with BD for age and sex (p=0.009, p=0.019). BPD and BD did not differ for what concerns alcohol and substance abuse and attempts of suicide or self-harm (although the differences for attempts of suicide and for self-harm shown a tendency toward significance), as reported in Table 1. Regarding the clinical assessment, the two diagnostic groups scored differently on the BIS-11 scale (p=0.001) and on the SCL-90 total score (p=0.005), showing that BPD patients were more impulsive and presented higher general psychopathology than BD. BPD and BD had a similar moderate degree of depression measured by HAM-D and presented similar levels of trait anxiety, as assessed by STAI-T. Furthermore, the two groups did not differ significantly in TAS scores (presenting, therefore, the same level of alexithymia) and in PSP, suggesting thus that BPD and BD were similar in terms of global functioning (Table 1).

	BPD (N=68)	**BD (N=25)**	***p-value***
Sex (Females)	45 (66.2%)	9 (36.0%)	0.009
Age	35.6 (10.0)	41.4 (9.6)	0.019
Alcohol abuse (Yes)	40 (60.6%)	14 (58.3%)	0.846
Substances abuse (Yes)	43 (66.2%)	12 (50.0%)	0.164
Suicide attempt (Yes)	36 (55.4%)	8 (33.3%)	0.065
Self-harm (Yes)	26 (41.9%)	5 (21.7%)	0.065
BIS-11	74.0 (9.1)	66.1 (9.4)	0.001
SCL-90	138.2 (72.4)	89.3 (65.5)	0.005
HAM-D	16.0 (10.8)	13.8 (10.4)	0.423
STAI-T	52.3 (11.4)	46.6 (15.6)	0.063
TAS	54.6 (14.1)	49.9 (15.0)	0.177
PSP	44.2 (14.3)	40.3 (12.1)	0.240

Table 1: Socio-demographic and clinical characteristics of 93 patients with borderline personality disorder (BPD) and bipolar disorder (BD). Mean (SD) for continuous variables and frequency (%) for categorical variables are reported.

BPD= Borderline Personality Disorder; BD= Bipolar Disorder; BIS-11= Barratt Impulsiveness Scale; SCL-90= Symptom Checklist-90-R; HAM-D= Hamilton Depression Rating Scale; STAI-T= State and Trait Anxiety Inventory- Trait; TAS= Toronto Alexithymia Scale; PSP= Personal and Social Performance Scale.

Socio-demographic and clinical profile identification for BPD and BD

Through MCA, two distinct patients' profiles were identified. People with BPD were more likely to be women and were more associated with suicide attempts and self-harm episodes, as shown in the lower right quadrant (blue circle) of the Figure 1, where these categories are close each other. Differently, patients with BD were more associated with male sex, no suicidal attempts and no self-injurious behaviours.

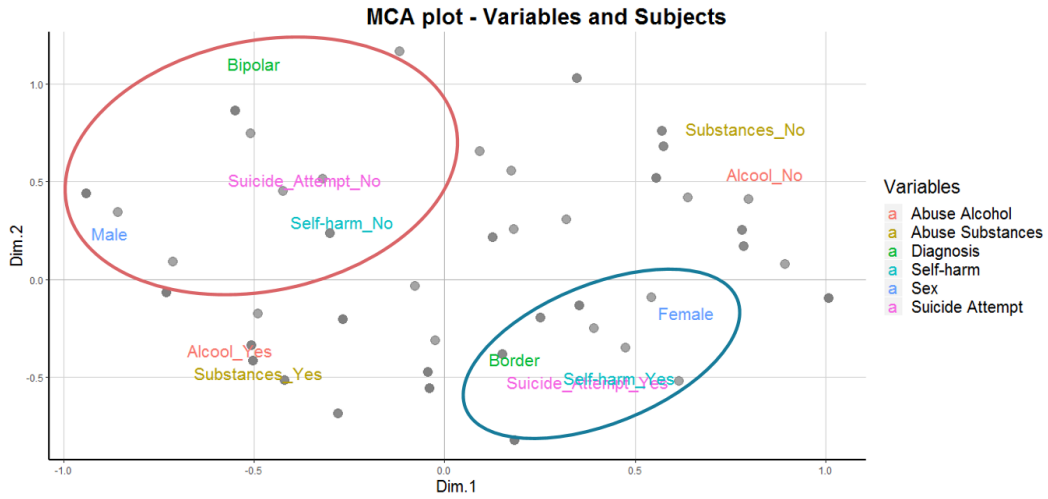
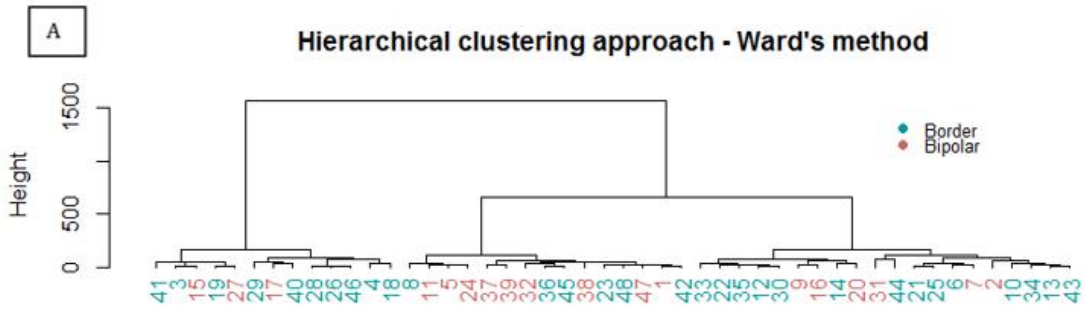


Figure 1: Multiple Correspondence analysis plot - Biplot- showing association between the categories of the following variables: abuse of alcohol, abuse of substances, diagnosis, self-harm behaviours, and sex and suicide attempts.

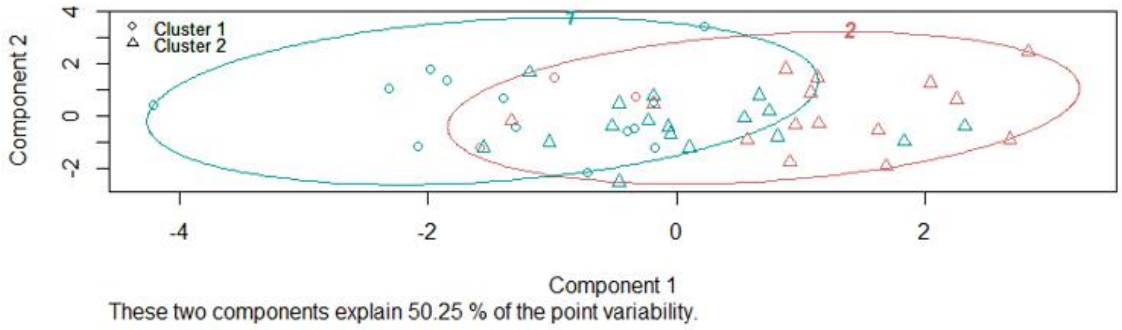
Points represent the patients while colored names represent the variables and corresponding categories. Red and Blue ellipses represent the categories more associated to Bipolar and Border patients, respectively.

In Figure 2 *panel A,* a dendrogram (i.e. a tree diagram showing hierarchical relationship among subjects), obtained from a data-driven hierarchical clustering (Ward's mean method), is shown; whereas in Figure 2 *panel B* a classification from the hypothesis-driven k-means method is displayed. The results have been obtained on the basis of the following quantitative variables: age, BIS-11, SCL-90, HAM-D, STAI-T, TAS and PSP. The hierarchical method is recommended to be performed before the second clustering method in order to detect the best number of clusters: in our case, two subject groups have been well identified around the dendrogram height equal to 1000 (see Supplementary material for the definition of the number of clusters). The subsequent k- means method (Figure 2 panel B), thus, was applied by imposing a number of clusters equal to two, obtaining a patient classification represented by circles (and corresponding blue ellipse) and triangles (and corresponding red ellipse). The first cluster was composed for the major part by BPD patients (13 vs 3) while the second cluster included, in almost equal percentage, BPD and BD patients, as reported in the Table displayed in Figure 2 panel C. Thus, the continuous variables used for the clustering allowed to well define a subgroup of BPD patients (blue circles in the left ellipse) that reported high scores in alexithymia (TAS), levels of trait anxiety (STAI-T) and psychopathology (SCL-90); while the BPD patients with low scores in such variables appeared indistinguishable from BD.

A



B



C

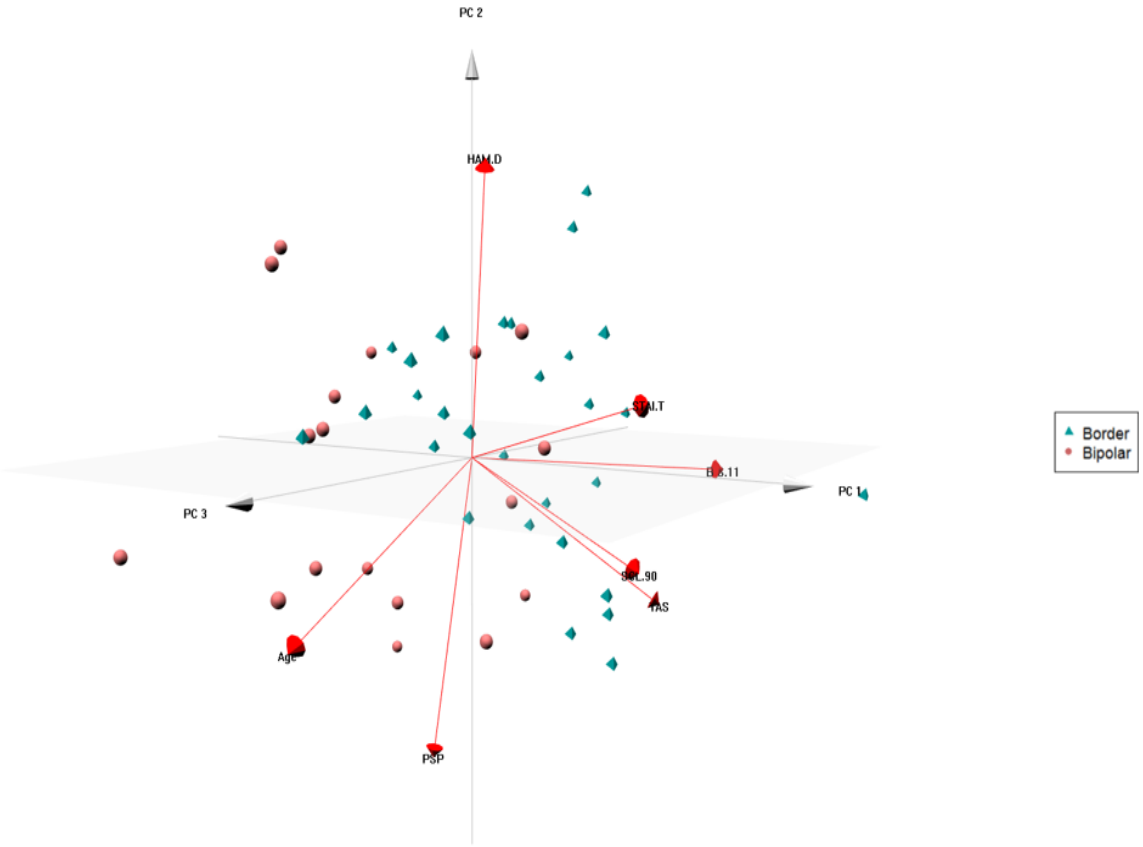
Cluster	Size	#BPD/#BD	AGE	HAM-D	PSP	TAS*	STAI-T**	BIS-11	SCL-90**
1	16	13/3	40.6	16.1	39.1	58.8	58.1	73.4	195.6
2	32	17/15	37.3	15.5	39.9	50.5	47.9	69.2	69.3

Statistical differences between cluster 1 and cluster 2: *p<0.1; ** p<0.05

Figure 2: A. Dendrogram obtained with a hierarchical approach. B. Clusters plot obtained with k-means method. C. Clusters details and corresponding clinical characteristics (centroids).
BPD= Borderline Personality Disorder; BD= Bipolar Disorder; HAM-D= Hamilton Depression Rating Scale; PSP= Personal and Social Performance Scale; TAS= Toronto Alexithymia Scale; STAI-T= State and Trait Anxiety Inventory-Trait; BIS-11= Barratt Impulsiveness Scale; SCL-90= Symptom Checklist-90-R.

The PCA output was reported through a 3d-biplot obtained by the first three component scores (PC1, PC2 and PC3) in Figure 3, where the variables (red arrows) and the subjects (coloured points) are displayed together. The major part of the BD patients (red points) were represented in the left part of the biplot (with low scores along the horizontal axis given by the PC1 and with quite low scores along the vertical axis PC2), whereas the most of the BPD patients were in the right part of the biplot represented by the blue triangles. The variables closer to

the BD were PSP and age while the ones more associated to BPD were STAI-T, TAS, SCL-90, BIS-11 and HAM-D.



PCA variables	PC1 loadings	PC2 loadings	PC3 loadings
Age	0.08	-0.25	0.76
HAM-D	0.17	0.60	0.20
PSP	-0.10	-0.66	0.02
TAS	0.41	-0.30	-0.19
STAIT-T	0.56	0.15	0.20
BIS-11	0.43	-0.05	-0.52
SCL-90	0.53	-0.16	0.17

Figure 3: 3d-representation of the first three Principal Components scores (PC1, PC2, PC3). Variables' loadings are represented by arrows and the subjects by points and triangles according to the diagnosis.
PC=Principal Component; HAM-D= Hamilton Depression Rating Scale; PSP= Personal and Social Performance Scale; TAS= Toronto Alexithymia Scale; SCL.90= Symptom Checklist-90-R; STAI.T= State and Trait Anxiety Inventory- Trait; BIS.11= Barratt Impulsiveness Scale.

These results have been confirmed also by the PLS-DA (Figure 4). The bar plot, representing the loadings of continuous variables in discriminating between the two diagnostic groups, showed that PSP and age were more associated to BD and the other variables were associated to BPD patients. In particular, BIS-11, STAI-T and SCL- 90 were the three variables that most contributed (with higher loadings) to identify patients with a borderline personality disorder characterized, as above shown, by high scores in impulsivity, levels of trait anxiety and psychopathology.

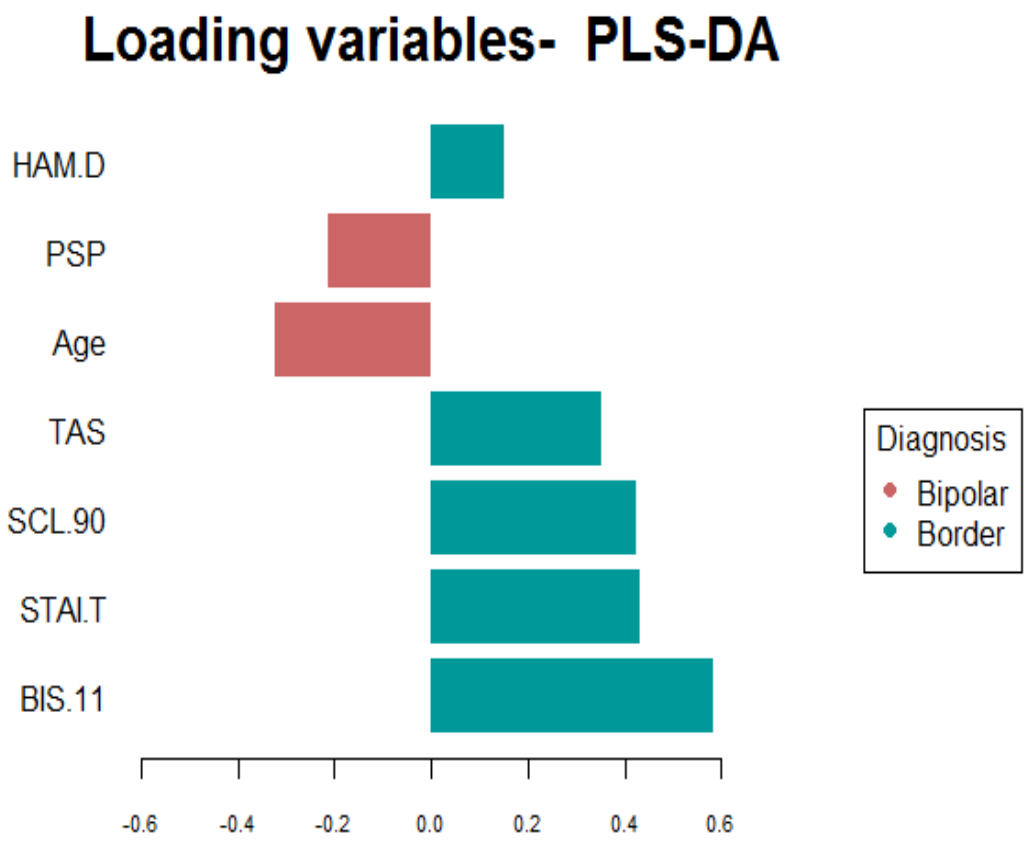


Figure 4: Plot of the variable loadings obtained with a Partial Least Squares Discriminant analysis.
HAM-D= Hamilton Depression Rating Scale; PSP= Personal and Social Performance Scale; TAS= Toronto Alexithymia Scale; SCL.90= Symptom Checklist-90-R; STAI.T= State and Trait Anxiety Inventory- Trait; BIS.11= Barratt Impulsiveness Scale.

After having identified the most significant variables that contributed to discriminate between the two groups, two classification trees were performed (one for continuous and one for categorical variables as predictors) to find a classification pathway defined through cut-offs of socio-demographic and clinical variables. In Figure 5 panel A, the two more predictive continuous variables (obtained after pruning procedure, see Supplementary materials) of the outcome (diagnosis) were BIS-11 and SCL-90. Patients with SCL-90 total score below 36 were more probably BD patients (probability p=87%); whereas patients with SCL-90 score higher or equal to 36 and a BIS-11 score higher or equal to 64 were more probably BPD patients (p=83%). For patients having SCL-90 score higher or equal to 36 and a BIS-11 score lower than 64 there was more uncertainty, showing that there is 56% of probability to be BD. The second CART was carried out by considering as predictors the following categorical variables: sex, alcohol abuse, substances abuse, suicide attempts and self-harm behaviours. The most predictive variables (obtained after pruning procedure) were sex, self-harm behaviours and suicide attempts, confirming the results obtained by the MCA, i.e. BPD patients are more probably females (p=83%) and, if male, they are more likely characterized by the presence of self-harm behaviours (p=89%) and suicide attempts (p=75%), while BD are more probably males who have never shown self-harm behaviours and who have not attempted suicide (p=59%) (Figure 5(A, B)).

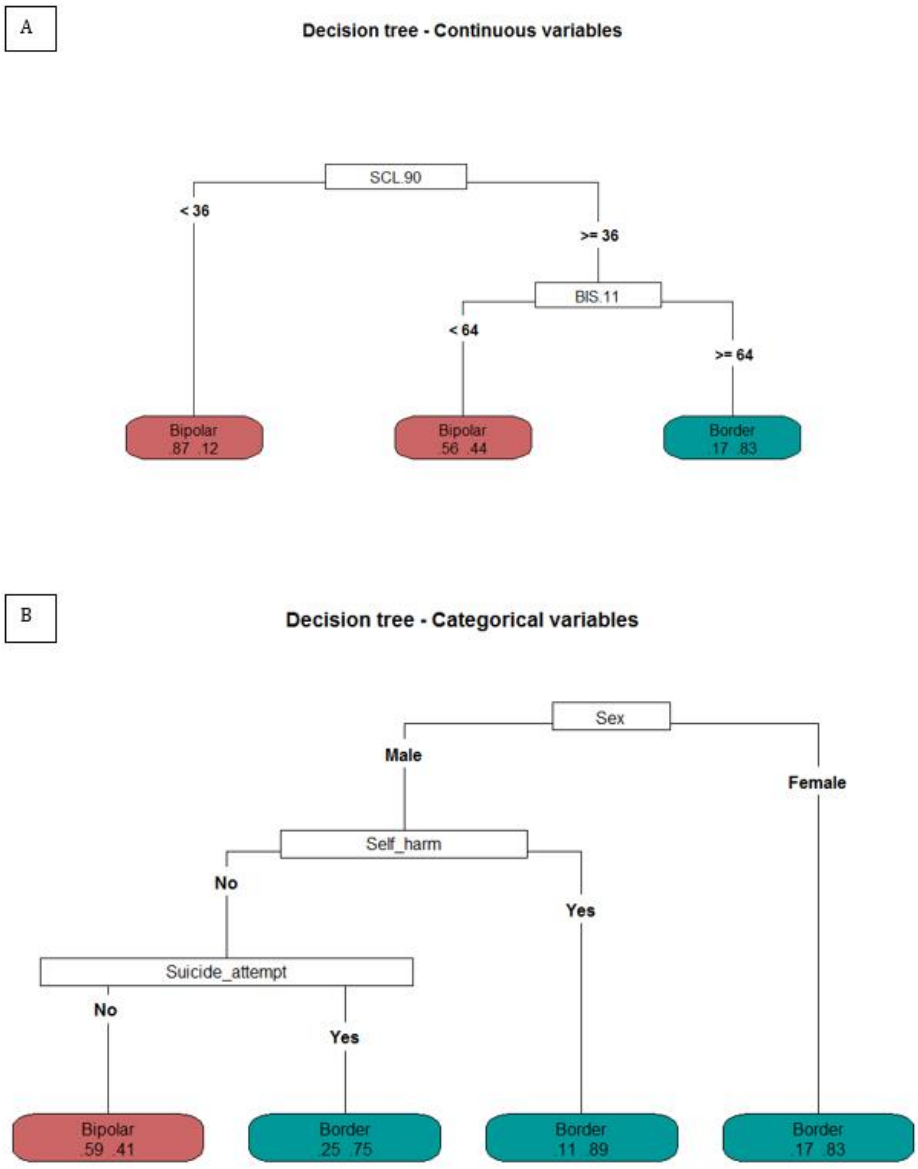


Figure 5: A) Decision tree that classifies patients with Borderline Personality Disorder and Bipolar Disorders on the basis of the most predictive continuous variables. B) Decision tree that classifies patients on the basis of the most predictive categorical variables.

SCL.90= Symptom Checklist-90-R; BIS.11= Barratt Impulsiveness Scale. In the down level of the tree the most likely outcome (dependent variable category: BD vs BPD) is reported for each pathway. The variables displayed are the ones obtained after pruning procedure for detecting the most predictive features. The values under the category are the probabilities of Bipolar and Border diagnoses respectively.

Discussion

Most of the multivariate statistical techniques for the association and classification analysis have been known since the half of the last century, although their application in many applied research contexts, such as medicine, social and behavioural sciences, is still limited [27]. This is probably due to the difficulty in applying and

interpreting these advanced statistical methods and perhaps the lack of knowledge of their potential and uses. This manuscript aims to help in overcoming these difficulties by presenting a practical application of some multivariate statistical techniques to a psychological dataset. In particular, detailed explanations of the methodologies and R software code are provided together with an exhaustive interpretation of the results. Five different multivariate techniques were applied for the identification of patient profiles that might help for the definition of potential new diagnostic patterns useful for clinicians.

The chosen study sample offered us the possibility to highlight the potential capability of the multivariate techniques in helping to solve the controversial hypothesis for which a clinical overlap between BPD and mood disorders (in particular bipolar disorder) exists since the introduction of BPD in Diagnostica and Statistical Manual of Mental Disorders (DSM-III) [28]. BPD is a chronic condition characterized by impulsivity, instability of social interactions and aggressive and parasuicidal behaviours; and the impulsivity, particularly when accompanied by substance abuse, together with suicide attempts, are common features also for BD patients [9]. Moreover, many of the clinical assessments used for evaluating the impulsivity and other social and psychological instabilities are not updated and fail in taking into account very peculiar psychopathologic features.

In such unclear clinical pattern, the use of advanced and more informative techniques (with respect to the more commonly used linear models) becomes paramount [29]. In addition, considering the different aspects and specific purposes of each of the multivariate techniques, it is worth noting that it is often advisable to apply more than one of such methods in order to obtain valuable, robust and coherent results. On our sample the applied methods provided consistent results. Both the methods for categorical data (MCA and CART) detected sex (female), presence of self-harm episodes and suicide attempts (in this order) as the variables more associated with the borderline personality disorder. Coherence was found also among the multivariate technique applied to continuous variables of the dataset. Cluster analysis, PCA and PLS- DA highlighted the assessment variables BIS-11, SCL-90, STAI-T and TAS (in this order) as the most prominent in predicting a borderline personality disorder, suggesting that high level of impulsivity, psychopathology and anxiety (even accompanied by self-harm episodes and suicide attempts) should favour toward a BPD instead of BD diagnosis.

All these results are in line with clinical knowledge. It is known, in fact, that BPD is diagnosed predominantly (about 75%) in females [30] and that BPD patients are usually characterized by unstable self-image, negative affectivity and disinhibition resulting therefore more associated with excessive self-criticism, depressivity, impulsivity and risk taking with consequent self-harming behaviors under emotional distress and potentially self-damaging activities [31]. On the other hand, BD patients usually present clear changes in mood that can range from periods of extremely up-behaviours (manic episodes) to very down-behaviours (depressive episodes) [32]. Such symptoms and behaviours are coherent with the hypothesis of a strict overlap of these two disorders on a same spectrum this might justify the results of our cluster analysis where many BPD patients were indistinct from BD patients [33]. The results of PCA, PLS-DA and CART thus, may be paramount to help the clinicians to better diagnose and discriminate patients with BPD from patients with BD, by focusing on impulsivity, psychopathology, anxiety and functioning.

Some limitations of the present study need to be acknowledged. The statistical techniques here presented find a fruitful application especially in big-data context. Our limited sample (both in terms of number of patients and number of variables analyzed) do not allow exploring all the potentiality of such methodologies. However, our aim was to provide a practical guide with an easy example to show and explain the application, the use and the interpretations of the results of these techniques. Another limitation regards the lack of validation for the decision tree method. Commonly, CART should be run first on a training set and then validated on a test set of data but the small sample did not allow such a procedure. Nevertheless, the choice to apply more than one of such methods allow us to obtain valuable and robust results that overcome partially the missing of the validation procedure.

References

1. Liang Y, Kelemen A(2016) Big Data Science and Its Applications in Health and Medical Research: Challenges and Opportunities. J Biom Biostat 7: 307.

2. James G, Witten D, Hastie T, Tibshirani R (2013) An Introduction to Statistical Learning with Applications in R. (1st edn), Springer Text in Statistics, New York, USA.

3. Bartholomew DJ, Steele F, Galbraith J, Moustaki I (2008) Analysis of multivariate social sciences data. (

2nd edn), Chapman and Hall/CRC, Boca Raton, FL, USA.

4. Lanfredi M, Candini V, Buizza C, Ferrari C, Boero ME, et al. (2014) The effect of service satisfaction and spiritual well-being on the quality of life of patients with schizophrenia. Psychiatry Res 216(2): 185-191.

5. Fertonani A, Ferrari C, Miniussi C (2015) What do you feel if I apply transcranial electric stimulation? Safety, sensations and secondary induced effects. Clin Neurophysiol 126(11): 2181-2188.

6. Pedrini L, Ferrari C, Ghilardi A (2018) Psychometric properties of the Italian Perceived Maternal Parenting Self-Efficacy (PMP S-E). J Clin Psychol Med Settings. doi: 10.1007/s10880-018-9578-3.

7. Coulston CM, Tanious M, Mulder RT, Porter RJ, Malhi GS (2012) Bordering on bipolar: the overlap between borderline personality and bipolarity. Aust N Z J Psychiatry 46 (6): 506-521.

8. Rossi R, Lanfredi M, Pievani M, Boccardi M, Beneduce R, et al. (2012) Volumetric and topographic differences in hippocampal subdivisions in borderline personality and bipolar disorders. Psychiatry Res 203(2-3): 132-138.

9. Rossi R, Pievani M, Lorenzi M, Boccardi M, Beneduce R, et al. (2013) Structural brain features of borderline personality and bipolar disorders. Psychiatry Res 213(2): 83-91.

10. Gigantesco A, Vittorielli M, Pioli R, Falloon IR, Rossi G, et al. (2006) The VADO approach in psychiatric rehabilitation: a randomized controlled trial. Psychiatr Serv 57(12): 1778–1783.

11. Derogatis LR (1994) SCL-90-R: Symptom Checklist-90-R: Administration, Scoring and Procedures Manual. (3rd edn), NCS Pearson, Minneapolis, USA.

12. Hamilton M (1960) A rating scale for depression. J Neurol Neurosurg Psychiatry 23: 56-62.

13. Bagby RM, Taylor GJ, Parker JDA (1994) The Twenty-item Toronto Alexithymia Scale-II. Convergent, discriminant, and concurrent validity. Journal of Psychosomatic Research 38(1): 33–40.

14. Bressi C, Taylor G, Parker J, Bressi S, Brambilla V, et al. (1996) Cross validation of the factor structure of the 20-Item Toronto Alexithymia Scale: An Italian multicenter study. J Psychosom Res 41(6): 551-559.

15. Patton JH, Stanford MS, Barratt ES (1995) Factor structure of the Barratt impulsiveness scale. J Clin Psychol 51(6): 768-774.

16. Fossati A, DiCeglie A, Acquarini E, Barratt ES (2001) Psychometric properties of an Italian version of the Barratt Impulsiveness Scale-11 (BIS-11) in nonclinical subjects. J Clin Psychol 57(6): 815-828.

17. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs AG (1983) State - Trait Anxiety Inventory for Adults Sampler Set. Manual for the State-Trait Anxiety Inventory (Form Y), Consulting Psychologists Press, Inc, Palo Alto, USA.

18. Rencher AC (2003) Methods of Multivariate Analysis. (2nd edn), Wiley Series in Probability and Statistics, John Wiley & Sons, Inc, USA.

19. Everitt BS, Landau S, Leese M, Stahl D (2011) Cluster Analysis. (5th edn), Wiley Series in Probability and Statistics, John Wiley & Sons, Inc, USA.

20. Joliffe IT (2002) Principal Component Analysis. (2nd edn), Springer Series in Statics. Springer-Verlag, New York, USA.

21. Flom PL (2008) An introduction to biplots. NDRI Statistics Support Group.

22. Boulesteix AL, Strimmer K (2006) Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. Brief Bioinform 8(1): 32-44.

23. Barker M, Rayens W (2003) Partial least squares for discrimination. Journal of Chemometrics 17(3): 166-173.

24. Brereton RG, Lloyd GR (2014) Partial least squares discriminant analysis: taking the magic away. Journal of Chemometrics 28(4): 213-225.

25. Rokach L, Maimon O (2008) Data mining with decision trees: theory and applications. World Scientific Publishing Co, Inc. River Edge, NJ, USA.

26. R Development Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

27. Thiese MS, Arnold ZC, Walker SD (2015) The misuse and abuse of statistics in biomedical research. Biochem Med (Zagreb) 25(1): 5-11.

28. American Psychiatric Association (1980) Diagnostic and Statistical Manual of Mental Disorders. (3rd edn), American Psychiatric Press, Washington DC, USA.

29. DJ Hand (1992) Statistical Methods in diagnosis. Statistical Methods in Medical Research 1(1): 49-67.

30. Skodol AE, Bender DS (2003) Why are women diagnosed borderline more than men? Psychiatric Quarterly 74(4): 349-360.

31. American Psychiatric Association (2012) Diagnostic and Statistical Manual of Mental Disorders. (5th edn), American Psychiatric Press, Washington DC, USA.

32. The National Institute of Mental Health (2016) Bipolar Disorder. The National Institute of Mental Health Information Resource Center.

33. Ruggero CJ, Zimmerman M, Chelminski I, Young D (2010) Borderline personality disorder and the misdiagnosis of bipolar disorder. J Psychiatr Res 44(6): 405-408.