International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 5 Issue 1, January 2018

ISSN: 2394-4404

Reasons For Polycystic Ovarian Syndrome Deficiency - A Data Mining Case Study

G. Nanthini @ Kalaiyarasi

PG Student, Dept of Computer Science, Jayaraj Annapackiam College for Women (Autonomous), Periyakulam, Theni Dt, TamilNadu, India

R. Gowsalya Devi

Asst.Professor, Dept of Computer Science, Jayaraj Annapackiam College for Women (Autonomous), Periyakulam, Theni Dt.

Abstract: ***The polycystic ovary syndrome (PCOS) is defined as a combination of hyperandrogenism and anovulation (infertility, and dysfunctional uterine bleeding), with or without the presence of polycystic ovaries on ultrasound. It represents the main endocrine disorder in the reproductive age, affecting 6% - 15% of women in men acme. It is the most common cause of infertility due to anovulation, and the main source of female infertility. Anxiety, depression, stress and personal dissatisfaction, frequently reported by women who live with PCOS may be aggravated by the image body change for consequence of weight gain. PCOS should be diagnosed and treated early in adolescence due to reproductive, metabolic and oncological complications which may be associated with it. Treatment options include drugs, diet and lifestyle improvement.***

Keywords: Significance, case study, Data Mining, Analysis, polycystic ovary syndrome (PCOS)

I. INTRODUCTION

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in women of reproductive age, affecting up to 10% of women. The clinical and biochemical features of the syndrome are heterogeneous, including menstrual irregularity and fertility problems, excess hair and acne. Women with PCOS are also more likely to be overweight and have an increased risk of metabolic syndrome, type 2 diabetes and cardiovascular disease. Although the exact aetiology of PCOS is unknown it involves a combination of genetic and environmental factors. Insulin resistance, with compensatory hyperinsulinaemia is considered a key factor in the development of symptoms. Hyperinsulinaemia is known to stimulate ovarian androgen production.

Significance of the study

- ✓ Irregular periods: In which the ovaries do not regularly release eggs (PCOS is the most common cause of infertility)
- ✓ Polycystic ovaries : In which the ovaries enlarge as more and more follicles develop but fail to release a mature egg
- ✓ High levels of androgens: Male sex hormones such as testosterone, which may cause physical signs such as excess facial or body hair

II. RESEARCH METHODOLOGY

A. METHODOLOGY

The case study will consist of deferent stages, roughly following the cross industry standard procedure CRISP-DM. Firstly, the business understanding phase has to be carried out. In this phase, the project objectives and requirements are stated and reined and the resulting data mining problem is formulated. The results of this phase are summarized in the previous sections. Although the collection of additional data results in a richer data set and is therefore likely to give better results, model acting on a data set that is already automatically kept-to-data is potentially a much useful tool.

B. ALGORITHM USED CLUSTER ANALYSIS

Cluster analysis is a multivariate analysis that attempts to form groups or "clusters" of objects (Sample Plots in our case) that are "similar" to each other but which differ among clusters. The exact definition of "similar" is variable among algorithms. But has a generic basis. The methods of forming clusters also vary, but follow a few general blueprints.

C. SIMILARITY, DISSIMILARITY AND DISTANCE

Similarity is characterization of the ratio of the number of attributes two objects share in common compared to the total list of attributes between them. Objects which have everything in common are identical, and have a similarity of 1.0. Objects which have nothing in common have a similarity of 0.0. As we have discussed previously, there is a large number of similarity indices proposed and employed, but the concepts are common to all.

Dissimilarity is the complement of similarity and is a characterization of the number of attributes two objects have uniquely compared to the total list of attribute between them. In general, dissimilarity can be calculated as 1-similarity.

D. K-MEANS CLUSTERING

The most common partitioning method is the K-means cluster analysis.

Conceptually, the K-means algorithm:

- ✓ Selects K cancroids (K rows chosen at random)
- ✓ Assigns each data point to its closest centroid
- ✓ Recaculates the centroids as the average of all data points in a cluster(i.e.,the centriods are p-length mean vectors, where p is the number of variables)
- ✓ Assigns data points to their closest centroids
- ✓ Continues step 3 and 4 until the observations are not reassigned or the maximum number of iterations(R uses 10 as a default) is reached.

III. IMPLEMENTATION DETAILS FOR THIS APPROACH CAN VARY

R uses an efficient algorithm by Hartigan and Wong (1979) that partitions the observations into K groups such that the sum of squares of the observations to their assigned cluster centers is a minimum. This means that in steps 2 nad 4, each observation is assigned to the cluster with the smallest value of:

SS(K)=(Xij∑-Xkj)∑

Where K is the cluster, Xij is the value of the jth variabl for the ith observation, and XKj is the mean of the jth variable for the kth cluster.

K-means clustering can handle larger datasets than hierarchical cluster approaches. Additionally, observations are not permanently committee to a cluster. They are moved when doing so improves the overall solution. However, the use of means implies that all variables must be continuous and the approach can be severely affected by outliers. They also perform poorly in the presence of non-convex (e.g., U-Shaped) clusters.

The format of the K-means function in R is K-means(X, Clusters) where X is a numeric dataset (matrix or data frame 0 and centers is the number of clusters to extract. The function returns the cluster memberships, centroids, sum of squares (within, between, total) and cluster sizes.

Since K-means cluster analysis starts with k randomly chosen centroids, a different solution can be obtained each time the function is invoked. Use the set. Seed() function to guarantee that the results are reproducible. Additionally, this clustering can be sensitive to the initial selection of centroids. The Kmeans() function has an start option that attempts multiple initial configurations and reports on the best one. For example, adding nstart=25 will generate 25 initial configurations. This approach is often recommended.

Unlike hierarchical clustering, K means clustering requires that the number of clusters to extract be specified in advance. Again the NbClust package can be used as a guide. Additionally, a plot of the total within-groups sums of squares against the number of clusters in a K-means solution can be helpful. A bend in the graph can suggest the appropriate number of clusters.

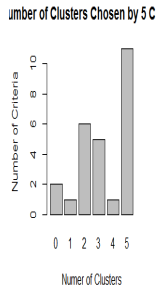
CLUSTER ANALYSIS IN R

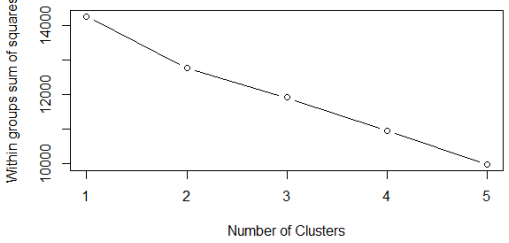
R has an amazing variety of function for cluster analysis. In this section, we use three of the many approaches: hierarchical agglomerative, partitioning, and model base

Data preparation: Prior to clustering data, you may want to remove or estimate missing data and rescale variables for comparability.

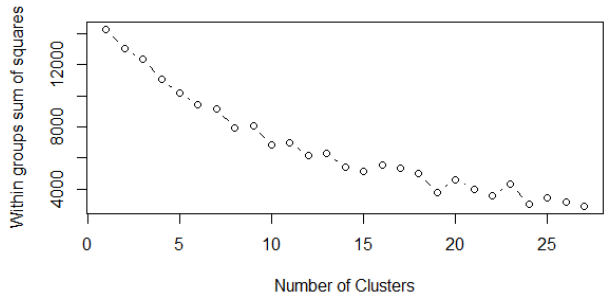
IV. FINDING, INTERPRETATION, RECOMMENDATION AND SUGGESTIONS

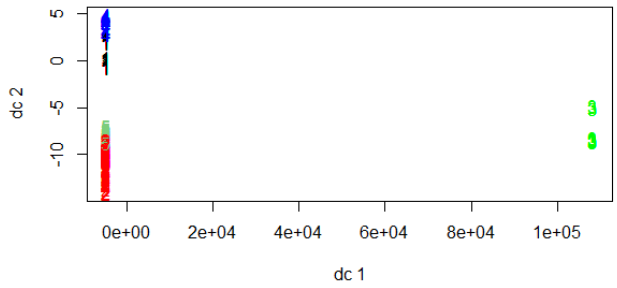
A. FINDING AND INTERPRETATIONS



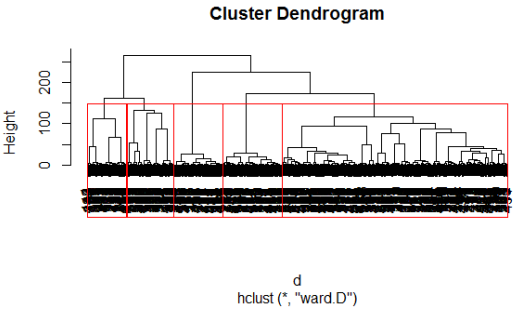


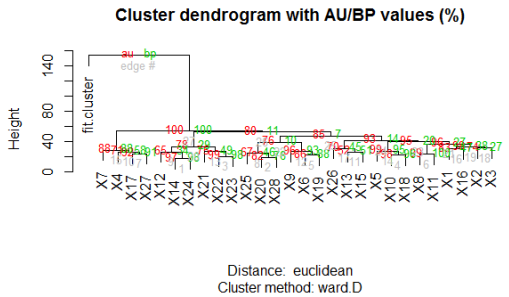
ISSN: 2394-4404



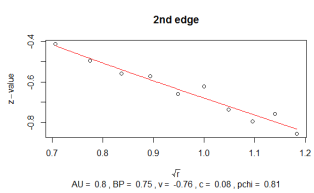


B. HIERARCHICAL AGGLOMERATIVE

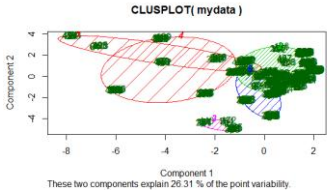




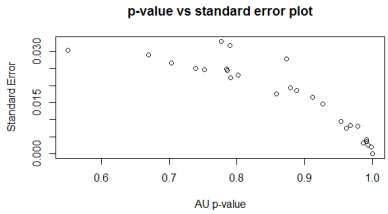
C. PLOTTING CLUSTER SOLUTION

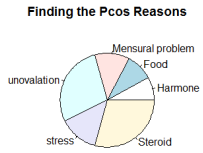


Finding Error Plot



PIE CHART





V. RECOMMENDATION

Since the origin of most symptoms associated with PCOS is thought to be related to insulin resistance, nutrition guidelines and lifestyle recommendations are centered on treating insulin resistance and its long term health effects.

The following recommendations have proven successful in treatment:

- ✓ Weight Loss
- ✓ Balanced Carbohydrate intake throughout day.
- ✓ Do not skip meals.
- ✓ Vitamin D Deficiency
- ✓ Include two to three servings of low fat dairy foods per day

VI. SUGGESTIONS

PCOS is recognized as the most common gynecological endocrinopathy, affecting 5% - 10% of women in childbearing age. It is also the most common cause of anovulatory infertility. However, there are some difficulties in its diagnosis due to different proposed diagnostic criteria. Women with PCOS have been pre- sented a greater risk of endometrial cancer. The overweight is also linked to this type of cancer and represents, therefore, an additional risk factor for endometrial cancer in these patients. A recent study found that obesity by itself was associated with the reduction of ovulatory rates, increasing in the abortion numbers of late pregnancy complications, and perhaps, therefore, increasing the infertility risk which was inherent to the syndrome. Anxiety, depression, stress and personal dissatisfaction, frequently reported by women who live with PCOS may be aggravated by the image body change for consequence of weight gain. PCOS should be diagnosed and treated already in adolescence due to reproductive, metabolic and onco- logical complications that probably are associated with it. The best way of prevention is through leading an adequate diet and a healthy lifestyle.

REFERENCES

[1] Norman, R.J., Dewailly, D., Legro, R.S. and Hickey, T.E. (2007) Polycystic Ovary Syndrome. Lancet, 370, 685-697. http://dx.doi.org/10.1016/S0140-6736(07)61345-2

[2] Doi, S.A. (2008) Neuroendocrine Dysfunction in PCOS: A Critique of Recent Reviews. Clinical Medicine &

Research, 6, 47-53. http://dx.doi.org/10.3121/cmr.2008.796

[3] Escobar-Morreale, H.F., Luque-Ramírez, M. and San Millán, J.L. (2005) The Molecular-Genetic Basis of Functional Hyperandrogenism and the Polycystic Ovary Syndrome. Endocrine Reviews, 26, 251-282. http://dx.doi.org/10.1210/er.2004-0004

[4] Xita, N. and Tsatsoulis, A. (2006) Review: Fetal Programming of Polycystic Ovary Syndrome by Androgen Excess: Evidence from Experimental, Clinical, and Genetic Association Studies. The Journal of Clinical Endocrinology & Me- tabolism, 91, 1660-1666. http://dx.doi.org/10.1210/jc.2005-2757

[5] Legro, R.S., Barnhar, T.H.X., Schlaff, W.D., Carr, B.R., Diamond, M.P., Carson, S.A., Steinkampf, M.P., Coutifaris, C., Mcgovern, P.G., Cataldo, N.A., Gosman, G.G., Nestler, J.E., Giudice, L.C., Leppert, P.C. and Myers, E.R., Coop- erative Multicenter Reproductive Medicine Network (2007) Clomiphene, Metformin, or Both for Infertility in the Po- lycystic Ovary Syndrome. The New England Journal of Medicine, 356, 551-566. http://dx.doi.org/10.1056/NEJMoa063971

[6] Ehrmann, D.A. (2005) Polycystic Ovary Syndrome. The New England Journal of Medicine, 352, 1223-1236. http://dx.doi.org/10.1056/NEJMra041536

[7] San Millán, J.L., Botella-Carretero, J.I., Álvarez-Blasco, F., Luque-Ramírez, M., Sancho, J., Moghetti, P., et al. (2005) A Study of the Hexose-6-Phosphate Dehydrogenase Gene R453Q and 11β-Hydroxysteroid Dehydrogenase Type 1 Gene 83557insA Polymorphism in the Polycystic Ovary Syndrome. The Journal of Clinical Endocrinology & Meta- bolism, 90, 4157-4162. http://dx.doi.org/10.1210/jc.2004-1523

[8] Yildiz, B.O. (2006) Diagnosis of Hyperandrogenism: Clinical Criteria. Best Practice & Research Clinical Endocrinol- ogy & Metabolism, 20, 167-176. http://dx.doi.org/10.1016/j.beem.2006.02.004

[9] Martin, K.A., Chang, R.J., Ehrmann, D.A., Ibanez, L., Lobo, R.A., Rosenfield, R.L., et al. (2008) Evaluation and Treatment of Hirsutism in Premenopausal Women: An Endocrine Society Clinical Practice Guideline. The Journal of Clinical Endocrinology & Metabolism, 93, 1105-1120. http://dx.doi.org/10.1210/jc.2007-2437

[10] Archer, J.S. and Chang, R.J. (2004) Hirsutism and Acne in Polycystic Ovary Syndrome. Best Practice & Research Clinical Obstetrics & Gynaecology, 18, 737-754. http://dx.doi.org/10.1016/j.bpobgyn.2004.05.007