VFAST Transactions on Education and Social Sciences
http://vfast.org/journals/index.php/VTESS@ 2020 ISSN(e): 2309-3951;ISSN(p): 2411-0221
Volume 8, Number 1, January-December, 2020

pp. 01–06

CASE STUDIES ON REGRESSION ANALYSES DEALINGWITH HEALTH DATA: A SUMMARY OF LITERATURE

Hilal POLAT[1]
[1] Vocational School of Technical Sciences
hilal.polat@erdogan.edu.tr

ABSTRACT. *Computational science and computers are great significance for a lot of studies conducted today. Storing the data is necessary for continuation of computational science studies more quickly and safely.*
Regression Analysis being used from the 1800s to the present day has been a statistical method to examine the relationship between one dependent variable and one or more independent variables. Regression analysis enables future events in the unknown future to be estimated based on the known findings, which makes the method the most preferred one especially in the health studies. This method is very important to identify the situation of disease, treatment process, and other possible complications and to prevent these problems in the light of the existing information.
Keywords*:* Health, Regression Analysis, Computational Sciences, Computer Sciences

1. **Introduction.** Every day, we are more or less interested in many issues, and we search for the answers of many questions. For example; what will be the score of the match that will be played at weekend? What will be the most popular occupations in the future? What will be the extent of reduction in unemployment rate in the upcoming 15-20 years in our country? What is the amount of external debt ratio which is expected to be paid? Which political party receive how many votes in the upcoming elections?
We ask such questions with numerical answers and try to find answers to them. We may benefit from previous data and tables to find answers for such questions. However, in some instances, it might be necessary to make prospective calculations and estimations in consideration of unavailable data. For example; expected growth rate of a company can be estimated based on previous balance data, expected sales, expenses, and other costs.Finding accurate and reliable answers for the questions on which we have no or limited information depends on classification of scientific data, statistical analysis and interpretation regarding those questions.
There is no common view on the roots of the word “statistics”. Some linguists believe that the word was derived from Latin word “status” which refers to “the state” or “position”. Others believe that it was derived from the word “stato” or Italian word “istatista” which means “governmental state”. Another group of linguists stated that the word might be derived from the word “statizein” which refers to “observation” (Arıcı, H. 2001).
Statistics can be defined in three ways. These are as follows;

1.A group of figures and graphics suggested by numerical values calculated for certain issues. The figures and graphics showing afforestation areas and the amount spent for these areas in Turkey can be given as an example to this definition.

2.A branch of science based on the basis of positive sciences which are experiments, planning, observations, collecting data, organization of collected data, analysis, interpretation, making objective and accurate decisions and scientific techniques and methods relevant to these basis (Akar ve Şahinler, 1997).

3.The name attributed to various values such as average and variance obtained via chance or calculated based on the examples. Similarly, correlation and regression coefficients are statistics if they are calculated based on examples.

Nowadays, many statistical analysis methods with many principles and variables are easily implemented by researchers thanks to advanced computer and packet software. This enabled researchers to implement statistical methods and techniques and to interpret the obtained results thanks to these packet software (Kalaycı, Ş.).

The most debated issue for educators is whether this model will increase success or not from the aspect of using computers in education. Most of the educators are of the opinion that using computers for educational purposes will increase the success, and this increase will be easier compared to traditional methods and techniques. Results of some studies indicated that using computer in education enriches the instruction and it may create positive changes in terms of the quality of education. (Foreman, 2005; Nuño, 2005; Açıkalın and Duru 2005; Aydın, 2005; Bosco, 2004; 2003; Yiğit and Akdeniz 2003; Alkan, 1995; Mitzlaff and Wiederhold, 1992; Aşkar and Erden, 1986).

Statistics is one of the courses for which computer supported instruction is commonly studied. To Gardner (2007), the studies on statistics are conducted by psychologists, statistics instructors and mathematics instructors. While psychologists focus on statistical thinking and reasoning, mathematics instructors focus on mathematical or numerical skills which facilitates learning of statistics. However, the scope of study for statistics instructors is more extended. The effects of computer supported education on statistical success and the attitude towards statistics are some of the areas on which statistics instructors focus (p.26-27).

Considering the current developments, it is possible to say that the software to be used in statistics instruction is increasing in number, computer technology and internet connections are more prevalent in quantitative terms and the possibility of learners to access these opportunities is higher (Bartz, 2001; Moore, 1997; Couch and Stoloff, 1989).

First of all,developing a hypothesis fitting to the purpose of the study and making a decision about the most appropriate statistical method are the most important steps in achieving the most accurate results. In the field of health, detecting a disease, other possible indications and deductions based on the available data are important to take steps for prevention.

One of the statistical methods which employs health data is regression analysis.

2.2. Regression Analysis

Regression analysis is carried out in order to reveal the relationship between two or more variables with cause-effect relationship, and to make estimations and predictions based on this relationship. It is possible to come across cause-effect relationship between many incidents in the nature.

In this type of analysis, a mathematical model is used to explain the relationship between two (simple regression) or more variables (multiple regression), and this model is called regression model.

Regression Analysis refers to the analysis which is used to transform the relationship between one criterion variable and one or more predictor variables into numerical values. Basically, regression analysis is intended to detect the quality of the relationship between the variables. If there is a single predictor variable, it is called simple regression while it is called multiple regression analysis if there is two or more predictor variables. The purpose is to detect the contribution of each predictor variable to the total change in criterion variable. Thus, based on linear combination value of predictor variable, criterion value can be estimated (19).

2.2.1. Simple Linear Regression Analysis

Y beingdependent (explained) and X beingindependent (explanatory), the method which reveals cause-effect relationship between two variables as mathematical model is called “regression analysis”. In order to implement this analysis, the variables must be divided into two as dependent and independent variables, and regression model must be established. For example; when the relationship between pH value (y) in saliva and daily number of cigarette consumed (x) is requested, pH value of saliva is a “dependent variable” since it is obtained after measured. The purpose of simple linear regression analysis here is to estimate y values of y=a+bx regression model within observation intervals, or whether it can be used to estimate values belonging to one or a couple of periods before and after the observation interval. In other words, its aim is to estimate how many cigarettes leads to how many units of increase in pH of saliva.

2.2.2. Logistic Regression Analysis

Logistic regression is a method to reveal cause-effect relationship via explanatory variables where answer variables are observed categorically and in double, triple and multiple categories. The effects of explanatory variables on dependent variable are obtained as probability rates and risk factors are turned into probability rates as well. It has no prerequisites like normal distribution assumption and continuity assumption12. For example; logistic regression analysis is carried out when parameters such as smoking (x1) gender, (x2) and family history (x3) are studies to reveal whether they are risk factors for contusion (y) (Gamgam H,1998; BaskanŞ,2004; Bernstein S.,1999; Baskan Ş. 2004; Boyacıoğlu H,2004).

Logistic regression analysis, which is currently being used frequently, is one of the three methods to appoint observations to groups (the other methods are clustering analysis and discriminant analysis). While the number of clusters to which observations will be appointed is unknown in clustering analysis, the number of groups is known in discriminant and logistic regression analyses. An awareness model is obtained based on available data, and thanks to this model, newly-added observations in data cluster can be appointed to groups.
The purpose of using Logistic Regression Analysis is the same with other model construction methods used in statistics. It is to construct a model, which is biologically acceptable andcapable of defining the relationships between dependent and independent variables, by using the least variable in return for the best concordance.
Logistic regression models are currently and commonly used in the fields of biology, medicine, economy, agriculture, veterinary and transportation.
The most distinct feature distinguishing logistic regression from linear regression is the fact that result variable is either double or multiple. The difference between logistic regression and linear regression is reflected on both parametric model selection and assumptions.
As it is the same with linear regression analysis, logistic regression tries to make estimations based on certain variable values. However, there are three important differences between these two methods (Elhan, A.H. 1997: 4)
The statistical method to be employed is determined according to obtained data.

3. Tools and Methods

The study was carried out based on the search made via Discovery Service and Google Academic search engines. The search covered studies conducted between the years 1999-2013 on patients who consulted to institutions of ministry of health. The studies which employed regression analysis were selected. The result covered 18 articles and a doctoral thesis which was written in Turkish language.
Literature review was made using “health data” and “regression analysis” as key words. Besides, seven studies which were presented in national congress proceedings belonging to years 1999-2007 were reviewed.
These studies were assessed in terms of setting, the number of the sample and their characteristics, methods, information source, the reasons as to why they were employed, perceived activity and the important results obtained at the end of the studies.

4. Findings

Reviewed studies are given in Table 1 in the order to author name and the date of the study.

- It was stated that majority of the studies were carried out with patients who were either in or out patients consulted to institutions which were affiliated with ministry of health.
- In all of the studies, the samples were comprised of either 250 or more patients.
- In all of the studies, the data was collected through survey method and/or face-to-face interview.
- The studies employed SPSS packet software and conducted statistical analyses. The method was Logistic regression methods in all of them.

Table 1.The Studies Dealing With Regression Analysis on Health Data

Year	Author Name	Method Employed	Language	Research
2009	ÖZKAN and et al.	loj reg	Turkish	Factors which may have effects on infant deaths
2000	Mehmet ESKiN	loj reg	Turkish	The prevalence of mental symptoms among high school adolescents and the possible relationship between these symptoms and suicidal behavior were analyzed.
1999	KARAKAş and et al.	doğrusal reg	Turkish	Neuropsychological tests for the assessment of cognitive processes in schizophrenia: measurement of memory and attention
1999	Yıldır Atakurt	loj reg	Turkish	Logistic regression analysis and an implementation in relation to its use in medicine
2010	Çetin and et al.	loj reg	Turkish	Health expenses and its effect on financial growth
2008	Çiftçi and et al.	loj reg	Turkish	The frequency of urinary incontinence
2012	Yargıç and et al.	loj reg	Turkish	The relationship between suicidal attempt, emotional abuse, self-destructive behavior and physical abuse
2004	Hüdaverdi Bircan	loj reg	Turkish	Revealing the extent to which birth weight is affected by risk factors that affect birth weight or lead to low birth weight
2006	Ayvaz and et al.	loj reg	Turkish	Studying the frequency of postpartum depression and risk factors during pregnancy

2013	Hıdıroğlu and et al.	loj reg	Turkish	Patient Knowledge, Attitude and Behaviors in Relation to Genetically Modified Organisms
2007	Kaya and et al.	loj reg	Turkish	The Prevalence of Depressive Symptoms, The Methods to Overcome the Stress and Influential Factors
2010	Vedat Bal	loj reg	Turkish	Detecting the amount of increase to occur in the revenue and the number of treatments provided by the hospital as a result of one-unit increase to take place in data processing procurement expenses
2001	Maral and et al.	loj reg	Turkish	The Prevalence of Depression and Risk Factors
2009	Ersan and et al.	loj reg	Turkish	Clinicopathologic progrnastic factors in patients with gastric cancer
2006	Çok and et al.	loj reg	Turkish	Comparing cisplatin-etoposide and mitomycn-iphosfamide-cisplatin combinations in advanced stage non-small cell lung carcinoma
2007	Pelitli Gürlü and et al.	loj reg	Turkish	The factors having effect on final visual acuity in open eyeball injuries
2007	Canbaz and et al.	loj reg	Turkish	Anxiety level of senior medical students and affecting factors
2008	Düzenli and et al.	loj reg	Turkish	The effect of long-term oxygen treatment on life-span in patients with chronic respiratory insufficiency

5. Conclusion

Regression analysis is commonly used for survey type data analysis. It also has the opportunity to be implemented in various fields though it is not commonly used in our country. The purpose of this study is to draw attention to the most commonly used method by conducting literature review.

Regression analysis is employed when the relationship between two variables is examined in the studies dealing with health. The fact that survey type data has discontinuous answers, and explanatory factors are also

frequently encountered. This prevents establishing regression model, when the aim is regression analysis, and discriminant function, when the aim is awareness. In such cases, Logistic Regression Analysis is used as an alternative.

Only 1 out of 19 studies employed linear regression in the study. The rest 18 studies employed logistic regression analysis. Linear regression analysis investigates the relationship between two variables while more than one criteria were compared via logistic regression analysis in the other studies.

REFERENCES

[1] Sümbüloğlu K, Sümbüloğlu V. Sağlık Bilimlerinde Araştırma Yöntemleri, 6. Basım. Ankara: Hatipoğlu Yayınevi.
[2] Kul S. Klinik Araştırmalarda Örnek Genişliği Belirleme. Plevra Bülteni 2011;2:129-32.
[3] Arıcı, H., 2001**:** İstatistik Yöntemler ve Uygulamaları. Meteksan A.Ş. Ankara.
[4] Akar, M., Şahinler, S., 1997: İstatistik. Çukurova Ün. Ziraat Fakültesi Yayınları No:74, Ders Kitaplar No:17, Ç.Ü.Z.F.
[5] Foreman, K. K. (2005). Design and evaluation of computer-asssisted instruction in the health sciences.
[6] Unpublished Doctoral Dissertation. The University of Utah, ABD.
[7] Açıkalın, M. ve Duru, E. (2005). The use of computer technologies in the social studies calssroom. The Turkish Online Journal of Educational Technology – TOJET 4, 2, 18-26.
[8] Aydın, E. (2005). The Use of computers in mathematics education: A paradigm shift from "Computer Assisted Instruction" towards "Student Programming". The Turkish Online Journal of Educational Technology – TOJET 4, 2, 27-34.
[9] Bosco, A. (2004). ICT resources in the teaching of mathematics: between computer and school technologies. A case-study. The Curriculum Journal, 15,3, 265-280.
[10] Yiğit, N. ve Akdeniz, A. R. (2003). Fizik öğretiminde bilgisayar destekli etkinliklerin öğrenci kazanımları üzerine etkisi: elektrik devreleri örneği. GÜ, Gazi Eğitim Fakültesi Dergisi. 23, 3, 99-113.
[11] Alkan, C. (1995). Eğitim Teknolojisi. Dördüncü Baskı, Atilla Kitapevi: Ankara.
[12] Mitzlaff, H. and Wiederhold, K.A. "ilkokullarda Bilgisayar Uygulaması", Eğitim ve Bilim. (çev: Mualla Bilgin). Sayı:84, 1992.
[13] Aşkar, P. ve Erden, M. (1986). "Mikrobilgisayarların Okullarda Kullanımı", Eğitim ve Bilim. 11, 61.
[14] Gardner, K. D. (2007). Investigating secondary school students' experience of learning statistics. Unpublished Doctoral Dissertation. Georgia State University, Atlanta, ABD:
[15] Bartz, A. E. (2001). Computer and software use in teaching the beginning statistics course. Teaching of Psychology, 28, 147-149.
[16] Couch, J. V., & Stoloff, M. L. (1989). A national survey of microcomputer use by academic psychologists.
[17] Teaching of Psychology, 16, 145-147.
[18] Moore, D. S. (1997). New pedagogy and new content: The case of statistics. International Statistical Review, 65, 123-165.
[19] Kalaycı, Ş. (Ed.) (2006). *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*. Ankara: Asil Yayın Dağıtım.
[20] http://www.frekans.com.tr/tr_analizler.html
[21] Gamgam H. Parametrik Olmayan İstatistiksel Teknikler,Ankara, 1998, 18-20
[22] Bernstein S.,Bernstein R., Elements of Statistics II:Inferential Statistics, McGraw-Hill, 1999, 379-394
[23] BaskanŞ.Uygulamalı İstatistik,1.Baskı,İzmir,1993, 60-73 2004
[24] Boyacıoğlu H,Boyacıoğlu H,"Assessment of water quality by statisticalmethods", Journal of control of water pollution Su Kirlenmesi Kontrolü Dergisi, Cilt :14 ,Sayı:3,9-17
[25] Elhan, A.H. (1997), Lojistik Regresyon Analizinin İncelenmesi ve Tıpta Bir Uygulaması. (Biyoistatistik Yüksek Lisans Tezi) A.Ü.,4-29.