

RESEARCH ARTICLE

A Web-Based Software Developed for Bayesian Tests and an Application in Medicine

İpek Balıkçı Çiçek¹, Şeyma Yaşar¹, Zeynep Tunç¹, Cemil Çolak¹

¹Department of Biostatistics and Medical Informatics, Faculty of Medicine, Inonu University, Malatya, Turkey

Received: 12 June 2020, Accepted: 24 August 2020, Published online: 31 August 2020
© Ordu University Institute of Health Sciences, Turkey, 2020

Abstract

Objective: In this study, it is aimed to develop a new user-friendly web-based software to easily carry out Bayesian tests, which are becoming more and more common, instead of using the classical approach, which is generally preferred in analysis from statistical modeling.

Method: Shiny, an open-source R package, is used to develop the recommended web software. In the developed software, by selecting “the Specify Sample Number” tab, the number of samples presented as “Single”, “Two” options is selected, and analyzes are made by selecting the appropriate data set from the file upload menu.

Results: The data set “ulcer recurrence” was used to examine the way the developed web-based software works and to evaluate its output. To test whether there is a difference in age variable in terms of result variable, “Two Independent Sample Bayes Tests” were selected and analyzes were performed. According to the results obtained, statistically “little evidence for Ho” was found for the age variable in terms of the result variable. With the evidence obtained, it is said that no statistically significant difference was obtained for the dependent variable according to the independent variable.

Conclusion: The developed software is a new user-friendly web-based software that can be used to easily use Bayesian tests used as an alternative to the classical approach.

Key words: Bayesian approach, Bayesian tests, Classical approach, Web-based software

Suggested Citation: Balıkçı Cicek I, Yasar S, Tunc Z, Colak C. A Web-Based Software Developed for Bayesian Tests, and an Application in Medicine. Middle Black Sea Journal of Health Science, 2020; 6(2):212-219.

Address for correspondence/reprints:

İpek Balıkçı Çiçek,

Telephone number: +90 422 341 0660/1281

ORCID-ID 0000-0002-3805-9214

E-mail: ipek.balikci@inonu.edu.tr

DOI: 10.19127/mbsjohs.752102

Introduction

It is known that in the development process of statistical science from the past to the present, it was influenced by two approaches known as the “Classic Approach” and the “Bayesian Approach”. The bases of the Bayes approach are based on Bayes' theorem, and this approach was introduced by Thomas Bayes. This approach has been left behind by the intensity of classical approach studies and the effect of those who support the classical approach. However, in addition to F.P. Ramsey's experiment called Reality and Probability in the early 20th century, it has been brought to the agenda again in recent years as it responded to the results that some researchers could

not achieve in their studies with a classical approach (Demirci, 2006).

There are some situations where the Bayesian approach is more useful and superior to the classical approach. Some of these situations are as follows. While the classical approach is based on many theorems, the Bayesian approach is based only on the Bayesian theorem. Therefore, the Bayesian approach makes making inferences easier. In the classical approach, the presence of prior information is neglected for the prediction of the parameters of the statistical models, while the prediction of the Bayesian approach is made using the presence of the prior information. The Bayesian approach at this point provides superiority. In the classical approach, range estimation is obtained for the parameter prediction, while in the Bayesian approach, the probability of the parameter prediction is obtained by benefiting the posterior distribution. Therefore, in the Bayesian approach direct calculation of the estimate provides an advantage in practice (Press 1989, Demirhan 2004). Estimating the parameters of models that are complicated in the classical approach is rather difficult. Since the marginal posterior distribution is used in the Bayesian approach, estimation of parameters of models is easier. In the classical approach, when the parameters are estimated with small samples, the reliability of the predictions may be lost. However, more reliable results can be obtained with small samples in the Bayesian approach. While making the parameter estimation with the classical approach, the assumptions such as

unbiased and consistency must be fulfilled, whereas in the Bayesian approach, the parameter estimation is a possibility, so there is no need to provide these assumptions. Also, since the variance obtained in the Bayesian approach is found using a priori distributions, it is smaller than the variance obtained with the classical approach, so that the interval estimates are obtained narrower. Thus, more accurate predictions can be achieved with the Bayesian approach. For these reasons, Bayesian analysis is preferred over the classical approach (Press 1989, Demirhan 2004). The purpose of this research is to develop a new user-friendly web-based software that uses the Bayesian approach as an alternative to the Classic approach to enable users to obtain more accurate results in their studies.

Methods

Data set

The data set “ulcer_recurrence” has been obtained from the link https://www.ibm.com/support/knowledgecenter/SSLVMB_24.0.0/spss/tutorials/data_files.html to be able to analyze how the web-based software developed in this study works and evaluate its output. This data set contains partial information from a study designed to compare the effectiveness of the two therapies to prevent relapse of ulcers. Age, duration, treatment, time, and result variables are given in the data set. The descriptive properties of the variables used in the data set are given in Table 1.

Table 1. Descriptive property table of variables

Variables	Variable Type	The Role of the Variable	Explanation
Age	Numerical	Independent/Predictive	Patient's age at the time of diagnosis
Duration	Categorical	Independent/Predictive	Duration of the disease (Less than five years / Five years and more)
Treatment	Categorical	Independent/Predictive	Treatment group (A / B)
Time	Numerical	Independent/Predictive	Duration of the last visit to the hospital in months
Result	Categorical	Dependent / Target	Result (Ulcer detected - No ulcer detected)

Bayesian Approach

Bayes theorem is one of the important theorems of statistics. This theorem: When modeling any situation, it aims to produce results using universal lines and observations. The use of observations and subjective opinions for the prediction of noninformative information is seen as the most important feature that distinguishes this approach from classical statistical methods (Akar and Gundogdu 2014). In the classical approach, the parameters of statistical models are defined as fixed, while in the Bayesian approach, these parameters are defined depending on probability. Therefore, there is

a distribution of each parameter. These possibilities are defined as the degrees of belief. In other words, parameters are accepted as random variables in the Bayesian approach (Bolstad 2007, Ibrahim et al, 2014).

The Bayesian approach is very advantageous compared to the classical approach, as it uses previously acquired knowledge (prior knowledge) in the modeling of complex data (Yin and Ibrahim 2006). Obtaining prior information plays an important role in the Bayesian inference. Prior knowledge is derived from past experiences and is based on a subjective interpretation of information

from previously conducted studies. By combining the results obtained from this prior information, the posterior distribution is formed (Congdon 2007, Wong, Lam and Lo 2005). The most important problem in obtaining posterior information is the difficulty of the calculation in the distribution. The difficulty of analytic solutions in complex distributions accelerated the development of computer software and thus facilitated finding solutions (Bolstad 2007).

Prior and Posterior Distribution

The preliminary distribution of a parameter is the probability distribution that enables us to obtain noninformative information for the parameter before starting the analysis of the data. Prior distribution: are the distributions obtained from the researcher's opinions, expert opinions, and similar studies. The posterior distribution of the parameter is obtained by multiplying the prior distribution with the likelihood function. All inferences of the parameter can be made by making use of the posterior distribution. Bayesian inference or modeling cannot be done without using a prior distribution. The Bayesian probability is considered to be the measure of the degree of belief of a random event. Based on this definition, the probability mentioned is quite subjective (informative). Therefore, all prior distributions are classified as subjective a priori and objective prior (Cengiz et al. 2012, Berger 2006). Objective averages have a minimal effect on the posterior distribution and are called uncertain or flat a priori. These prior distributions are used by most researchers because they seem more objective. However, it is not always appropriate to give the total uncertainty for the parameter with an objective a priori distribution. Therefore, objective prior distributions cause the researchers to obtain the posterior distribution incorrectly (Kass and Wasserman 1996).

Subjective prior distributions are the most preferred and used prior distributions in practice. Among the subjective prior distributions, conjugate a priori and Jeffreys' prior are the most preferred and known preliminary distributions. Conjugate a priori distribution is preferred because the posterior distribution is easily obtained and provides convenience in subsequent calculations. The preliminary distribution of Jeffreys was proposed by Jeffreys in 1961 and can be easily calculated. This a priori has uniform distribution properties and does not include large values outside the range in which the parameters are defined (Jeffreys 1998, Ashby and Smith 2000).

In practice, one of the difficulties encountered for posterior distributions in the Bayesian approach is that the integrals required for statistical inferences are difficult to solve and even there are no analytical solutions for these integrals. Some models are used to make the truth understandable in science, technology, management, and many other fields. With simulation studies, comments can be made using models to create specific situations and to examine the reaction against these situations. Some simulation methods have been developed to overcome the difficulties in the Bayesian approach and these methods are used in the solution of integrals (Altuntas 2011).

Markov Chain Monte Carlo Method

Solutions of integrals to be used in the calculation of posterior distributions in the Bayesian approach are either too difficult or impossible. In these cases, approaches are given to derive the Markov chain and achieve convergence properties and posterior distribution (Berg 2005). Thus, using the Markov chain derivation and simulation techniques, approaches called chain data replication were developed. In general, these approaches are examined through the Markov chain Monte Carlo (MCMC). With the help of the Monte Carlo method, a set of simulation values is generated from the desired probability distribution independent of each other. In other words, many values are drawn randomly from the posterior distribution. The MCMC method, on the other hand, generates a chain value where each simulation value will depend on the previous value. The purpose of the MCMC approach is to converge to the target posterior distribution randomly in the θ parameter space. That is, the $\theta^{(j+1)}$ random variable derived to converge to the posterior distribution in MCMC approaches depends on the random variable θ^j . For this purpose, sampling is done from $q(\theta^j | \theta^{(j-1)})$ Markov transition distribution (Cengiz et al. 2009).

Metropolis Algorithm and Gibbs sampling are the most used algorithms in MCMC. With these methods, it is possible to draw samples from complex posterior distributions and obtain posterior statistics.

Metropolis Algorithm

This algorithm was discovered by the American physicist and computer scientist C. Metropolis. The algorithm is very practical and easy to use. Estimates are made to obtain random samples from arbitrary complex target distribution of any size known as the normalization constant. Samples are to be created from a univariate distribution, with probability density function $f(\theta|y)$. Suppose the t th sample obtained with f is θ^t . To start using the Metropolis algorithm, θ^0 is taken as the initial value. Also, there is a need for a symmetrical $q(\theta^{(t+1)}|\theta^{(t)})$ target density function. The application of the algorithm is the process of deriving a sample from the $q(\cdot|\cdot)$ density function for the $\theta^{(t)}$ the parameter in the case of $(t + 1)$ iteration and deciding whether to accept or reject the new sample. If the new sample is accepted, the algorithm starts with the new sample and repeats itself. If the new sample is to be rejected, the algorithm remains at the current point and starts repeating. The algorithm repeats itself until necessary. In practice, the user has the advantage of stopping sampling and deciding the total number of samples after sufficient iteration has been completed.

We can summarize this algorithm as follows.

- 1) With $f(\theta^0|y) > 0$, $t = 0$ and the starting point θ^0 are selected.
- 2) With the help of the proposed $q(\cdot|\theta^{(t)})$ distribution, θ_{new} is produced.
- 3) $r = \min\{(f(\theta_{new}|y))/(f(\theta^t|y)), 1\}$ value is calculated.
- 4) U is generated from $U(0,1)$.
- 5) If $u < r$, it becomes $\theta^{(t+1)} = \theta_{new}$. Otherwise, $t = t + 1$ is taken. Return to Step 2 if T is the desired sample number and $t < T$. Otherwise, the process is completed (Gilks, Richardson and Spiegelhalter).

Gibbs Sampling:

Gibbs sampling is a special case of the Metropolis algorithm. For the solution to the problems encountered in the Bayesian approach, Gibbs sampling was first used by Gelfand et al (1990) (Gelfand et al. 1990). The main purpose of Gibbs sampling is to run the prediction process periodically through conditional relationships and to reach the compound distributions of the parameters. The expression $\pi(\theta_i | \{\theta_j, i \neq j\}, y)$ for the $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ parameter vector, $p(y|\theta)$ likelihood function, and $\pi(\theta)$ a priori distribution can be written as follows.

$$\pi(\theta_i | \{\theta_j, i \neq j\}, y) \propto p(y|\theta)\pi(\theta) \tag{1}$$

The algorithm of Gibbs sampling under the above explanations is as follows.

- 1) $t=0$ is taken and chosen to be an arbitrary initial value $\theta^{(0)} = \{\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)}\}$.
 - 2) Each component of θ
 - From $\pi(\theta_1 | \theta_2^{(t)}, \dots, \theta_k^{(t)}, y)$ to $\theta_1^{(t+1)}$
 - From $\pi(\theta_2 | \theta_1^{(t)}, \dots, \theta_k^{(t)}, y)$ to $\theta_2^{(t+1)}$
 -
 - From $\pi(\theta_k | \theta_1^{(t+1)}, \dots, \theta_{(k-1)}^{(t+1)}, y)$ to $\theta_k^{(t+1)}$
- is obtained.
- $t = t + 1$ is taken. Return to Step 2 if T is the desired sample number and $t < T$. Otherwise, the process is completed.

When Gibbs sampling is run with long enough iteration, all loops in the algorithm will be able to derive a sampling for all components of θ using conditional distributions. However, this sampling does not work when conditional distributions of x parameters are not easily available (Gelfand et al. 1990).

Developed Web-Based Software:

To create this web-based application, the Shiny library is used, which allows the design of interactive web-based applications based on the R programming language (Chang et al. 2016). The main and submenus of the software are described below.

Upload a File:

In the first stage of this web-based application, the file containing the data set is loaded. In data analysis, loading is done with MS Excel (.xls / .xlsx) and SPSS (.sav) file types, which are the most widely used file types with different extensions. Besides, this menu includes the 'Determine the type and role of the variables' tab, which will enable determining the type and role of the variables in the uploaded file. The process of determining the roles of variables is required for analysis.

Bayesian Analysis:

In this menu, the variable type must first be determined to decide on the test to be used in the web-based application. After deciding on the type of variable, the sample size must be determined. When the Bayesian analysis menu is opened, two options are offered, namely "Quantitative Variables" and "Qualitative Variables" from the "Determine Variable Type" tab. After the variable type is decided, the sample size is selected from the options presented

as "One" and "Two" from the "Determine Sample Size" tab.

Here, "One Sample Bayes Tests" for "One" sample size, "Two Independent Sample Bayes Tests" and "Two Dependent Sample Bayes Tests" tabs for "Two" sample size are opened. After loading the data set from the file upload menu, the sample size suitable for the loaded data set is selected. Then, analysis is made by making test selections by the sample size. Figure 1 shows the Bayesian Analysis menu.

Accessibility and citation of the improved interactive web application:

The developed interactive web-based software can be accessed free of charge at <http://biostatapps.inonu.edu.tr/BTY/>. Information on how to show the software as a source in scientific studies is available in the "Citation" menu. Figure 2 shows the "Citation" menu."Shiny" (Chang et al. 2016), "BayesFactor" (Morey et al. 2015) package was used in the development of this software.

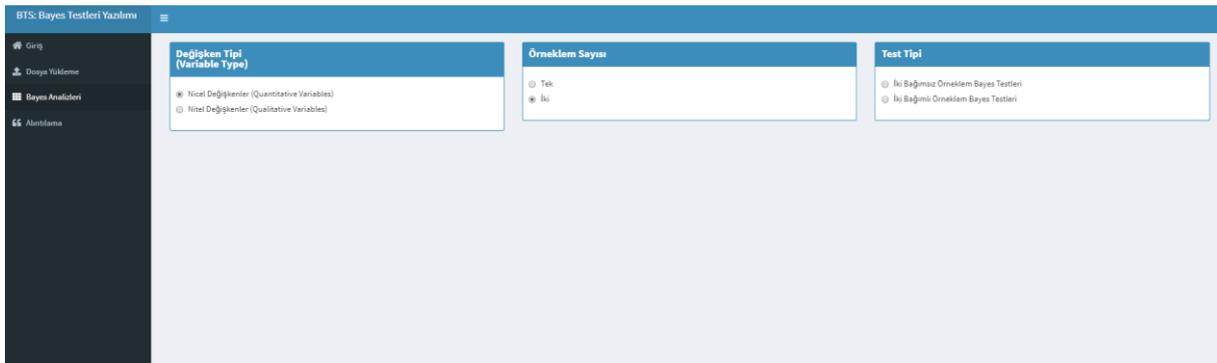


Figure 1. Bayesian Analysis menu

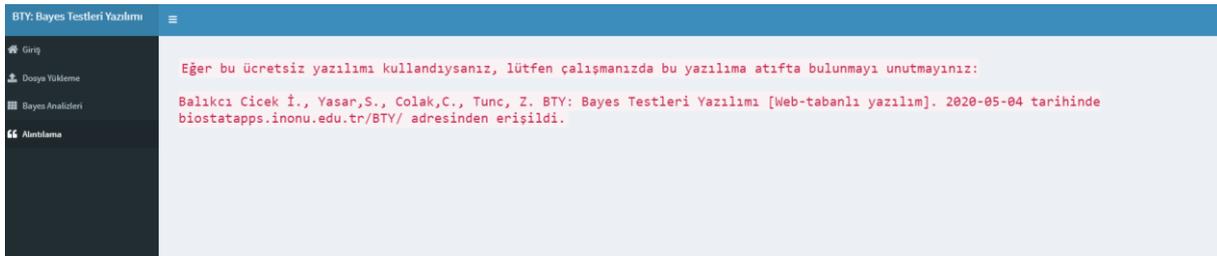


Figure 2. Citation Menu

Results

In this study, the "ulcer recurrence" dataset was used to analyze how this web-based software works and evaluate its output. The user interface of the software is shown in Figure 3.

The descriptive statistics table for age and time variables in the data set are given in Table 2.

The distribution table for the target variable in the data set is given in Table 3.

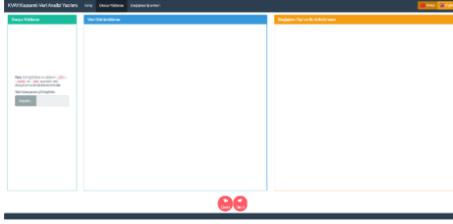
Firstly, the data set is loaded to the file upload menu in the software. Figure 4 shows the images of the uploaded file.

After the file is uploaded, the test selection is made from the "Bayesian Analysis" tab by the purpose of

the study and the data set. To determine whether there is a difference in age variable in terms of result variable, "Two Independent Sample Bayes Tests" were selected and analyzes were performed. Figure 5 shows the images of the results of "Two Independent Sample Bayes Tests" for the variables selected.

According to this output, statistically "little evidence for Ho" was found in the age variable in terms of the result variable. According to the result obtained, it is said that there is no statistically significant difference in terms of the dependent variable compared to the independent variable.

Bayes Testleri Yazılımı



Klasik yaklaşım ile parametre tahmini yapılırken yansızlık ve tutarlılık gibi varsayımların yerine getirilmesi gerekirken, Bayes yaklaşımında parametre tahmini bir olasılık olduğu için bu varsayımların sağlanmasına gerek yoktur. Ayrıca Bayes yaklaşımında elde edilen varyans, önsel dağılımlar kullanılarak bulunduğu için klasik yaklaşımla elde edilen varyanstan daha küçüktür ve dolayısıyla aralık tahminleri daha dar elde edilir. Böylece Bayes yaklaşımıyla daha doğru tahminlere ulaşılabilmektedir. Bu nedenlerle Bayes analizi Klasik yaklaşıma tercih edilmektedir. Bu açık kaynak web-tabanlı yazılım, Anabilim dalımız tarafından geliştirilen R uygulamalarını birer web uygulaması olarak oluşturmayı

sağlayan bir RStudio projesi olan Shiny kütüphanesi ile web ortamına aktarılmıştır.

[Yazılıma erişim için tıklayınız](#)

Figure 3: Software Interface

Table 2. Descriptive Statistics table for Age and Time variables

Variables	Mean	Standard deviation	Median	Minimum	Maximum
Age	50.98	16.42	52	23	76
Time	10.42	2.91	12	2	12

Table 3. Distribution table of the variable Result

Ulcer detected		No ulcer detected	
Count	Percentage	Count	Percentage
11	25.6	32	74.4

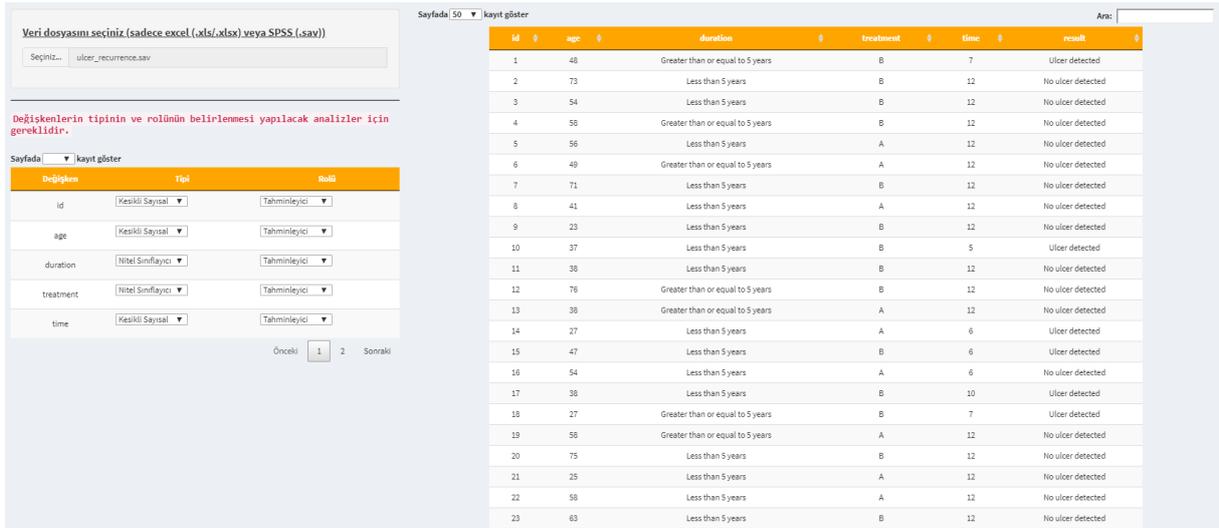


Figure 4. Image of the uploaded file

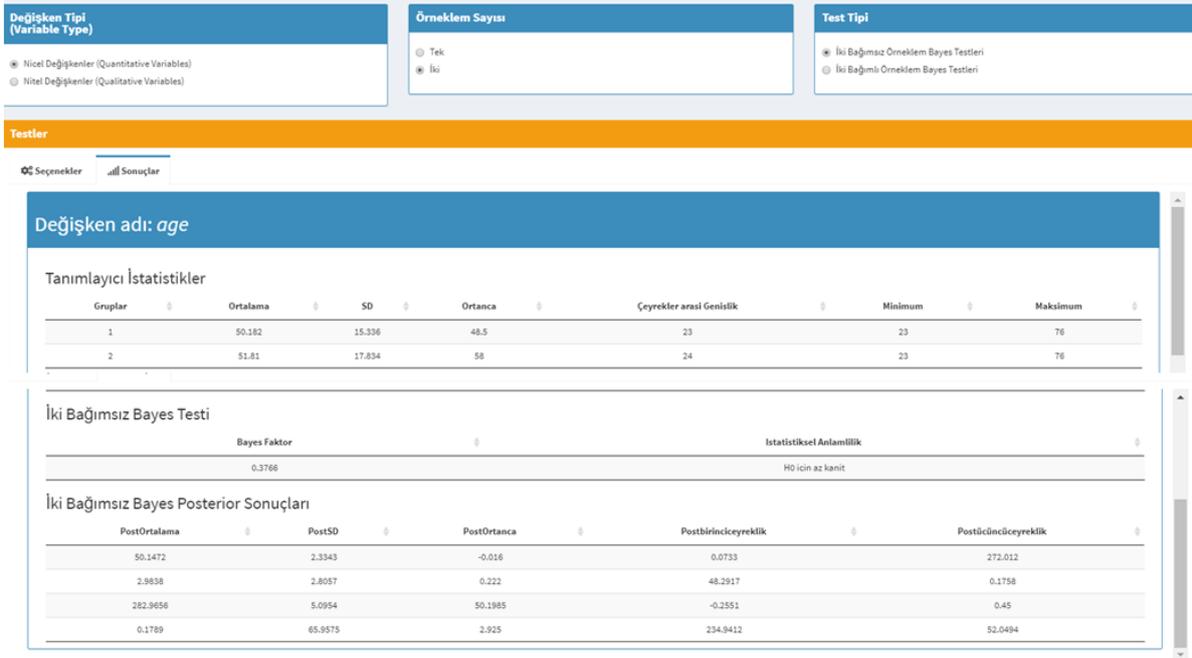


Figure 5. Two Independent Sample Bayes Test Results

Discussion

The foundations of the Bayesian approach emerged in the 18th century based on the "Bayesian Theorem". With the changes made since its occurrence until now, it has advanced considerably today and has been an approach used in almost all branches of science. The most important difference between the Bayesian approach and the classical approach is that the parameters are taken as a random variable for analysis in the Bayesian approach. Researchers supporting the classical approach did not accept the Bayesian approach since the prior distributions selected for the parameters were subjective (Press 1989, Demirhan 2004). Although it has sometimes been left behind by the advocates of the classical approach since its development, it has always been compared with the classical approach. In the working principle of classical statistical methods, unknown parameters are assumed to be fixed and probabilities related to these parameters are determined by using relative frequencies. According to these assumptions, since the parameters are fixed, probability inferences about them cannot be made. For this situation, Bayesian methods, which accept the probability as a "degrees of belief" (that is, the degree of belief in the accuracy of an event) and the trial parameter as a random variable, offer an alternative approach to the classical approach. In other words, parameters are accepted as random variables in the Bayesian approach (Bolstad 2007, Ibrahim et al. 2014).

For the studies, this developed web-based software has "One Sample Bayes Tests" for "One" sample size, "Two Independent Sample Bayes Tests" and "Two Dependent Sample Bayes Tests" and tabs for "Two" sample size. In the software, descriptive statistics of variables, outputs related to the test, comments about the outputs, and posterior results of the variable are presented to users. Some package programs such as IBM SPSS Statistics (Spss 2013), Minitab (Minitab 2000), and Stata (StataCorp 2007) and JASP (Team 2018), which are among the known package programs, carry out Bayesian analysis but there is no interpretation of the analysis output in these programs. Also, the software we developed offers free access.

Besides, it is planned to add "More Than Two Independent Sample Bayes Tests" and "More Than Two Dependent Sample Bayes Tests" tabs for the "More than Two" sample size. Codes that cannot be found in the R programming language for more than two samples codes in the Python programming language will be interoperable by adding a Python session to the R session with the Reticulate package. As a result, the developed software is a user-friendly web-based software that allows users to perform Bayesian tests that can be used as an alternative to single and two-sample tests in the classical approach.

Ethics Committee Approval: Ethics committee approval is not required in this study.

Peer-review: Externally peer-reviewed.

Author Contributions: Author Contributions: Concept - C.C.; Design- I.B.C., S.Y.; Supervision- C.C.; Materials- I.B.C., S.Y.; Data Collection and/or Processing- I.B.C.; Analysis and/or Interpretation- I.B.C., S.Y., Z.T.; Literature Review- I.B.C., Z.T.; Writing- I.B.C.; Critical Review- C.C.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study hasn't received no financial support.

References

- Akar, M. & S. Gundogdu. Use of Bayes theory in seafood. *Journal of Fisheries Sciences*. com, 2014; 8: 8-16.
- Altuntas, M. Bayesian Approaches in Statistical Model Selection and Bayes Factor. Master Thesis. TC Sinop University Institute of Science, Sinop, 2011
- Ashby, D. & A. F. Smith Evidence-based medicine as Bayesian decision-making. *Statistics in medicine*, 2000; 19: 3291-3305.
- Berg, B. A. Introduction to Markov chain Monte Carlo simulations and their statistical analysis. *Markov Chain Monte Carlo Lect Notes Ser Inst Math Sci Natl Univ Singap*, 2005; 7: 1-52.
- Berger, J. The case for objective Bayesian analysis. *Bayesian analysis*, 2006; 1: 385-402.
- Bolstad, W. Introduction to Bayesian Statistics Second Edition, A John Wiley & Sons. Inc., New York, 2007
- Cengiz, M., Y. Terzi & B. Yuksel A Bayesian approach in detecting vascular occlusion. *Journal of Experimental and Clinical Medicine*, 2009; 21.
- Cengiz, M. A., E. Terzi, T. Senel & N. Murat A bayesian approach to parameter estimation in logistic regression. *Afyon Kocatepe University Journal of Science and Engineering Sciences*, 2012; 12: 15-22.
- Chang, W., J. Cheng, J. Allaire, Y. Xie & J. McPherson shiny: Web Application Framework for R; 2016. R package version 0.13. 2. URL: <http://CRAN.R-project.org/package=shiny>.
- Congdon, P. Bayesian statistical modelling. John Wiley & Sons, 2007
- Demirci, M. Bayes Theorem and its applications in business administration. *The Journal of Academic Social Science Studies*, 2016; 43: 439-462.
- Demirhan, H. Bayesian estimation of parameters and expected cell frequencies in logarithmic linear models. Master of Science Thesis. Hacettepe University, 2004
- Gelfand, A. E., S. E. Hills, A. Racine-Poon & A. F. Smith Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 1990; 85: 972-985.
- Gilks, W., S. Richardson & D. Spiegelhalter. Markov chain Monte Carlo in practice. 1995. Chapman and Hall/CRC.
- Ibrahim, J. G., M. H. Chen & D. Sinha Bayesian Survival Analysis. Wiley StatsRef: Statistics Reference Online, 2014
- Jeffreys, H. The theory of probability. OUP Oxford, 1998
- Kass, R. E. & L. Wasserman Formal rules for selecting prior distributions: A review and annotated bibliography. *Journal of the American Statistical Association*, 1996; 435: 1343-1370.
- Minitab, I. MINITAB statistical software. Minitab Release, 13, 0., 2000
- Morey, R. D., J. N. Rouder, T. Jamil & M. R. D. Morey Package 'bayesfactor'; 2015. URL <http://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf> (accessed 1006 15).
- Press, S. J. Bayesian statistics: principles, models, and applications. John Wiley & Sons Inc., 1989.
- Spss, I. IBM SPSS statistics 22. New York: IBM Corp., 2013
- StataCorp, L. Stata data analysis and statistical Software. Special Edition Release, 2007; 10: 733.
- Team, J. JASP (Version 0.9)., 2018; Computer software. <https://jasp-stats.org>.
- Wong, M., K. Lam & E. Lo Bayesian analysis of clustered interval-censored data. *Journal of dental research*, 2005; 84: 817-821.
- Yin, G. & J. G. Ibrahim. Bayesian transformation hazard models. In *Optimality*, 2006; 170-182. Institute of Mathematical Statistics.