ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

WSSPAS: An Interactive Web Application for Sample Size and Power Analysis with R Using Shiny*

WSSPAS: R Shiny Paketi Kullanılarak Örneklem Büyüklüğü ve Güç Analizi İçin İnteraktif Bir Web Uygulaması*

Ahmet Kadir ARSLAN,ᵃ
Şeyma YAŞAR,ᵃ
Cemil ÇOLAK,ᵃ
Saim YOLOĞLUᵃ

ᵃİnönü University Faculty of Medicine,
Department of Biostatistics and
Medical Informatics,
Malatya, TURKEY

Correspondence:
Şeyma YAŞAR
İnönü University Faculty of Medicine,
Department of Biostatistics and
Medical Informatics, Malatya,
TURKEY/TÜRKİYE
seyma.yasar@inonu.edu.tr

ABSTRACT Objective: The calculation of sample size and power analysis plays an important role in biomedical research. The most general definition of the calculation of the sample size and power analysis is to determine the minimum number of individuals that have the ability to represent the population during the planning phase of the study. Since the statistical methods for each research plan are different, the calculation of sample size and power analysis will be different. Therefore, it is difficult to calculate the sample size and power analysis manually for each clinical trial. The aim of this research is to develop a new user-friendly web-based tool that calculates sample size and power analysis for hypothesis testing, diagnostic tests, correlation and regression analysis using the open source software R Shiny package and guides the researchers with examples. **Material and Method:** This web tool will be updated upon the updated R software packages, including shiny, shinydashboard, pwr, powerAnalysis, powerMediation, MKmisc and rhandsontable. Scripts were written for calculations that could not be done by these packages. **Results:** Hypothetical samples were created to introduce menus in the web-based software developed for the calculation of sample size and power analysis, and screen images of the results of these samples were given. **Conclusion:** The designed interactive web application is freely accessible through http://biostatapps.inonu.edu.tr/WSSPAS. In the future studies, it is aimed to further strengthen the software by adding modules that can calculate sample size and power analysis for different multivariate statistical and machine learning methods.

Keywords: Sample size; power analysis; hypothesis test; diagnostic test; correlation, regression; web-based application

ÖZET Amaç: Biyomedikal araştırmalarda örneklem büyüklüğü ve güç analizinin hesaplanması önemli bir rol oynamaktadır. Örneklem büyüklüğünün ve güç analizinin hesaplanmasının en genel tanımı, çalışmanın planlama aşaması sırasında popülasyonu temsil etme kapasitesine sahip olan asgari kişi sayısını belirlemektir. Her bir araştırma planı için istatistiksel yöntemler farklı olduğundan, örneklem büyüklüğü ve güç analizinin hesaplanması farklı olacaktır. Bu nedenle, her klinik deneme için örneklem büyüklüğü ve güç analizini manuel olarak hesaplamak zordur. Bu araştırmanın amacı, açık kaynak kodlu yazılım R Shiny paketini kullanarak hipotez testi, tanı testleri, korelasyon ve regresyon analizi için örneklem büyüklüğü ve güç analizini hesaplayan ve araştırmacılara örneklerle rehberlik eden yeni bir kullanıcı dostu web tabanlı araç geliştirmektir. **Gereç ve Yöntemler:** Bu web tabanlı yazılım, shiny, shinydashboard, pwr, powerAnalysis, powerMediation, MKmisc, WebPower ve rhandsontable dahil olmak üzere güncellenmiş R yazılım paketleri üzerine geliştirilmiştir. Bu paketler tarafından yapılamayan hesaplamalar için komut dosyaları manuel olarak yazılmıştır. **Bulgular:** Örneklem büyüklüğü ve güç analizinin hesaplanması için geliştirilen web tabanlı yazılımda menüler tanıtmak için hipotetik örnekler oluşturulmuş ve bu örneklerin sonuçlarının ekran görüntüleri verilmiştir. **Sonuç:** Tasarlanan interaktif web uygulamasına http://biostatapps.inonu.edu.tr/WSSPAS aracılığıyla kolayca erişilebilir. Ayrıca ilerleyen çalışmalarda farklı çok değişkenli istatistiksel ve makine öğrenmesi yöntemleri için de örneklem büyüklüğü ve güç analizini hesaplayabilen modüllerin eklenmesiyle, yazılımın daha da güçlendirilmesi hedeflenmektedir.

Anahtar Kelimeler: Örneklem büyüklüğü; güç analizi; hipotez testi; tanı testi; korelasyon; regresyon; web-tabanlı uygulama

The calculation of sample size and power analysis is an important issue in biomedical studies. One of the frequently asked questions by researchers while planning a biomedical study is how many individuals should be involved in the study. The most basic step in planning is sample size and power analysis which is the method to determine the size of sample with the ability to represent.[1,2] There are many justifications for determining the sample size and power analysis. First, if the number of samples is too large for a data set which is not expected to yield statistically significant results, the study results are likely to be statistically significant. Secondly, if the sample size of a study for which significant results are expected is small, it is likely that the results of the study are statistically insignificant. In addition, sample size should not be ignored for economic reasons. A study with insufficient sample size may waste resources since it will not yield useful results. On the other hand, it uses more resources than the ones required for a study that has an excessive number of samples. At the same time, the size of the sample provides important evidence of the reliability and effectiveness of the study. Statistical power is the probability that this difference may be true if a statistically significant difference is found in the study conducted. While the small number of samples leads to the low power of the research, and a large number of samples leads to high power. On the other hand, the data obtained in the research that is prepared with different experimental setups are evaluated with appropriate statistical methods. Therefore, the calculation method of the sample size and the power analysis to be used at the beginning of the research can produce the possible differences depending on these statistical methods. Therefore, for each statistical test, there is a different sample size and power analysis formula.[3]

The American Psychological Association (APA) Publication Manual underlined that effect size should be reported together with p-values. Therefore, the effect size uncloses the magnitude of the treatment effect(s). According to this important result, reporting the effect size and p-value is necessary for improving the quality of the outputs of clinical trials. In this situation, the effect size may indicate the magnitude of the treatment effect and also compares the quantitative findings from the different types of research in meta-analyses.[4] For this reason, it is crucial to develop softwares/ applications that calculate the impact size for planning well-designed studies and improving the quality of the research.

The formulas used to calculate power and sample size for each clinical trial design are different and quite complex. It would be more appropriate for researchers to choose the correct computer programs and software developed for calculating the sample size in this respect. The aim of this study is to develop a new user-friendly web tool that allows the researchers to calculate the sample size and power analysis.

MATERIAL AND METHODS

SAMPLE SIZE AND POWER ANALYSIS

In the planning phase of a study, the population is first defined in which hypotheses are to be constructed. Then, the selected sample should have the ability to represent this identified population. One of the most important questions of the researcher at this stage of the research is to determine how many people should be selected as the sample in order to find a clinically and statistically significant results. In biomedical research, it is very important to calculate a sufficient sample size for the selected sample to represent the target population. There are many reasons for this. First, power analysis is carried out to prevent wasted resources and labor. Secondly, a planned study with a small sample may have low power to obtain significant results. On the other hand, the concept of power analysis is the possibility of detecting this difference in the statistical test used to compare the efficacy of two or more drugs or treatment methods. The calculated power (1-β) should be a minimum of 80% in clinical trials. For the sample size calculation, the

type of output variable, Type I Error (α), Type II Error (β), Effect Size, Allocation Ratio, The direction of Alternative Hypothesis (H_1) must be known.[5,6]

Type of Output Variable: The outcome variable used in a research can mostly be in two different types. One of these is the categorical variables expressed as a percentage of the incidence, side effect, or healing of a disease. The other one is the numerically expressed variable which is used to express the body functions obtained by means of a measuring instrument.

Type I Error (α): In a controlled trial in which the efficacy of any drug or method is investigated, Type I error is the probability that the researcher will inadvertently find a difference when there is not a real difference in terms of effectiveness. In clinical trials, this difference should be chosen as a maximum of 5%.

Type II Error (β): Type II error is the probability that the researcher will inadvertently find a "no difference" when there is a real difference in terms of effectiveness. In clinical trials, Type II error should be selected as a maximum of 20%. The 1- β value indicates the strength of the clinical work done, hence the lowest power value should be 80%.

Effect Size: It is the statistical value that indicates the degree of deviation of the sample results from the predictions specified in the null hypothesis.[7,8] More simply, the effect size is the concept of how much a newly tried method makes a difference compared to the old one. Effect size is calculated by different formulas. For the statistical methods in which the difference between the two group averages is calculated, the Cohen's d formula is extensively preferred in calculating the effect size. The effect size ranges for the Student's t-test, which is one of the hypothesis tests, reported by Cohen are interpreted as with 0.20 for small effect size, 0.50 for medium effect size, and 0.80 for large effect size. Similarly, for the variance analysis reported by Cohen, the effect size ranges were equally interpreted as 0.01 or less for small effect size, between 0.01 to 0.14 for medium effect size, and 0.14 or larger for large effect size.[9,10] The value of the effect size for correlation analysis is equal to the correlation coefficient. Interpretation of effect sizes for correlation coefficients (r) according to Davis is interpreted as a negligible relationship between 0.01 and 0.09, a small relationship between 0.10 and 0.29, the medium relationship between 0.30 and 0.49, a strong relationship between 0.50 and 0.69. On the other hand, interpretation of effect sizes for correlation coefficients (r) according to Hinkle, Wiersma, and Jurs is interpreted as very small between 0.00 and 0.30, small between 0.30 and 0.50, medium between 0.50 and 0.70, high between 0.70 and 0.90 and very high between 0.90 and 1.00. According to Hopkins, interpretation of effect sizes for correlation coefficients (r) is interpreted as a negligible relationship between 0.00-0.10, the small relationship between 0.10-0.30, the medium relationship between 0.30-0.50, the high relationship between 0.50-0.70, the very high relationship between 0.70-0.90, perfect relationship between 0.90-1.00. The reported ranges of effect sizes for Cohen's correlation coefficient are interpreted as 0.01 or less for small effect size, 0.01 to 0.14 for medium effect size, and 0.14 or greater for large effect size.[11-13] Another simple measure for effect size calculations is the multiple regression coefficient R^2. All the basic statistical softwares automatically calculate the R^2 value. According to Cohen, the effect size (R^2) obtained can be interpreted as 0.0196 small, 0.1300 medium, and 0.2600 the great effect value.[8] For categorical measures, such as "yes" versus "no" or "sick" versus "not sick", two measures that can be used to assess effects are relative risks and odds ratios. Relative risk statistics are particularly useful in prospective clinical trials to assess the differences between treatments. The reported ranges of effect sizes for the relative risks are interpreted as 2 for the small effect size, 3 for the medium effect size, and 4 for the large effect size.[14] While the relative risk is an appropriate measure for prospective studies, such as randomized clinical trials or cohort studies, the odds ratio is suitable for case-control studies, usually when subjects with a given characteristic are compared to those without the characteristic. Similarly, the reported

ranges of effect sizes for odds ratio is interpreted as 1.5 for the small effect size, 2 for the medium effect size, and 4 for the large effect size.[15] As a rule of thumb, it is recommended that the effect size value be ≥ 0.5 for the comparison of two groups in clinical trials. Also, the effect size should be ≥ 0.45 for the comparison of three or more groups.[16]

Allocation Ratio: If the number of samples to be calculated is to be determined at the lowest possible level, it should be distributed according to 1:1 rule. That is, 1 patient should be taken in the placebo group versus 1 patient taken in the treatment or drug group. Sometimes this preference can be selected as 1:2 or 1:3. In this case, it is necessary to work with larger samples to have the same power as 1:1 ratio.[16]

The Direction of Alternative Hypothesis (H_1): The direction of the alternative hypothesis can be one-sided or two-sided. One-sided hypotheses test for difference in the direction of greatness or smallness, while two-sided hypotheses test for inequality (≠) without a sense of direction. The calculated power of two-sided hypothesis tests is lower than the power of one-sided hypothesis tests.[6]

Types of Power Analysis

Power analysis is applied in two different ways in scientific research.

1. A Priori Power Analysis: While the research is in the planning phase, it is the power analysis that is used to calculate the number of units of research carried out on the sample in order to reach the estimated research power based on the estimated parameters.[16]

2. Posteriori Power Analysis: It is the power analysis that is used to calculate the actual power of the decisions made in a research that is concluded according to the research plan.[16]

The Developed Web-Based Software

The web-based application was designed using Shiny version 1.0.5 package on the basis of the R programming language. The developed web-based software includes the calculation of sample size and power analysis for hypothesis tests, diagnostic tests, correlation and regression analyses. The main and sub menus of the web-based application are explained in detail below.

Examples

Hypothetical samples were created to introduce menus in the web-based software developed for the calculation of sample size and power analysis, and screen images of the results of these samples were given.

Hypothesis Test Menu

In this web-based tool, the "Hypothesis test" menu consists of two sub-sections, which are "Quantitative Variable" and "Qualitative Variable". In the "Quantitative Variable" sub-section, there are three different number of options, which are one sample(group), two sample(groups) and more than two sample (groups). In the "one sample (group)" sub-menu, there is a test type option, which is one sample t-test. Under the "two sample (groups)" sub-menu, there are two different types of test options, which are independent samples t-test and paired samples t-test. If groups are independent, "independent samples t-test" is selected. However, if the groups are dependent, "paired samples t-test" is selected. In the "more than two samples (groups)" sub-section, there are two different types of test options, which are independent samples one-way ANOVA (Analysis of Variance) and paired samples one-way ANOVA. Similarly, if groups are independent, "independent samples one-way ANOVA" is selected. However, if the groups are dependent, "paired samples one- way ANOVA" is selected. In the "Qualitative Variable" sub-section, there are three different types of test options, which are one-sample proportion test, two independent samples proportion test, and Chi-square test. For example, the view of sample size and power calculation menus for "One-way ANOVA (independent-repeated)" was given below (Figure 1).

Similarly, the a view of the module of the sample size and power analysis for the Chi-Square test from the hypothesis tests in developed web-based software shows below (Figure 2).

Diagnostic Tests Menu

This menu consists of three sub-sections, which are ROC (Receiver Operating Characteristic), sensitivity and specificity. In the "ROC Curve" sub-section, there is only one option which is "One-Sample ROC Curve Test". If the sample size needed to compare the area under the one ROC curve to the zero hypothesis value is desired to compute, the "One-Sample ROC Curve Test" option should be chosen. Screenshot of the ROC curve test is given below (Figure 3).
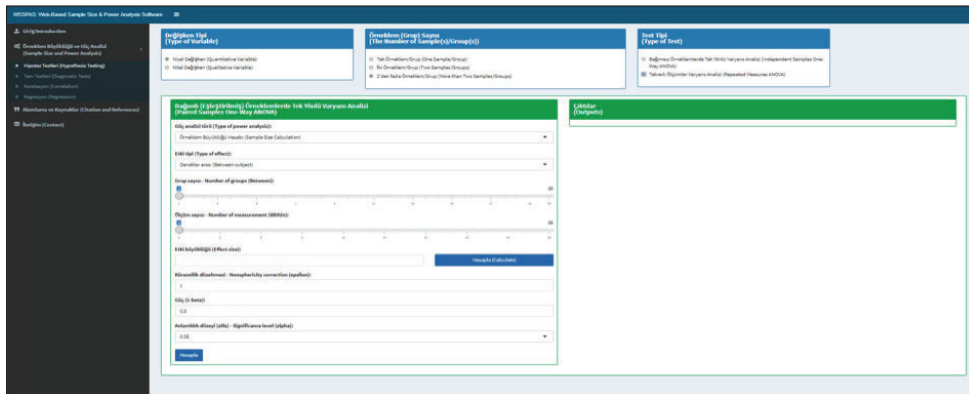

FIGURE 1: "Repeated Measures ANOVA and Independent Samples One-Way ANOVA" menu
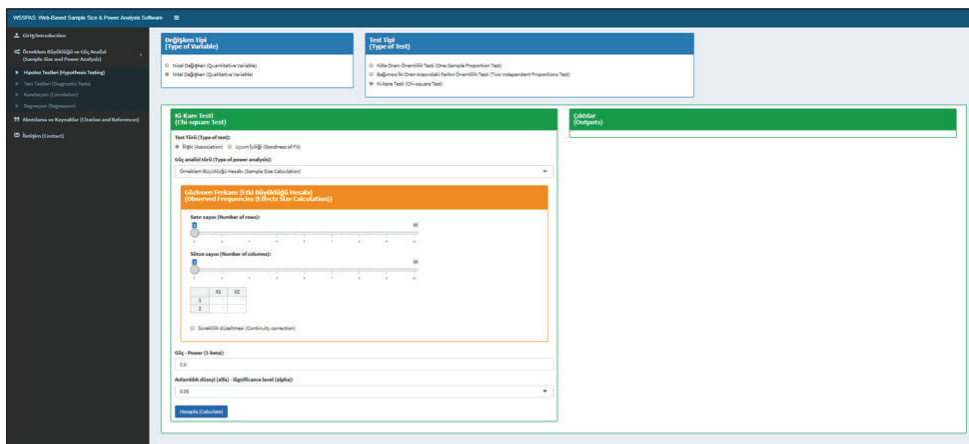
FIGURE 2: "Chi-Square Test" menu
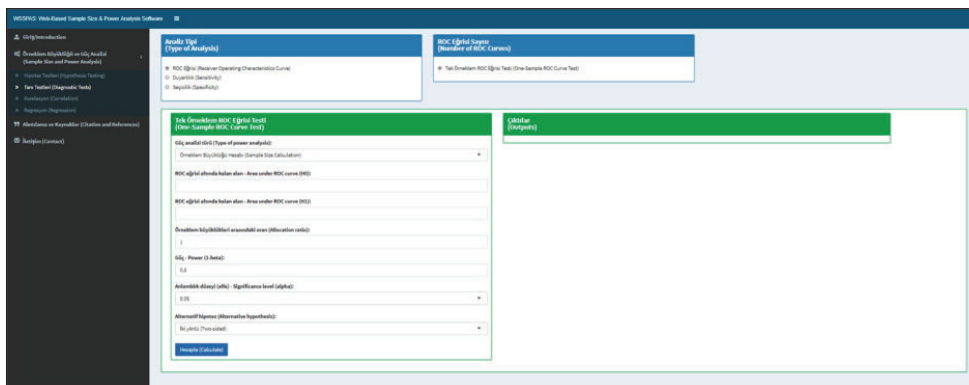
FIGURE 3: "One-Sample Roc Curve Test" menu

As for "The Sensitivity" and "The Specificity" sub-sections, there are two options called "One-Sample Sensitivity Test" and "One-Sample Specificity Test". If the sample size is calculated for one diagnostic test with sensitivity or specificity, the "One-Sample Sensitivity Test" or "One-Sample Specificity Test" sub-section is selected. The screenshots of "The Sensitivity" and "The Specificity" sub-sections are depicted below, respectively (Figure 4, Figure 5).

Correlation Menu

In this web-based software, the "Correlation" menu contains " One Sample Correlation Test" sub-section. The one sample correlation test calculates the size of samples required to obtain the H_1 correlation coefficient to be obtained from the sample with a certain power when the correlation coefficient is equal to a value such as H_0 (5). The view of the Correlation menu is given below (Figure 6).
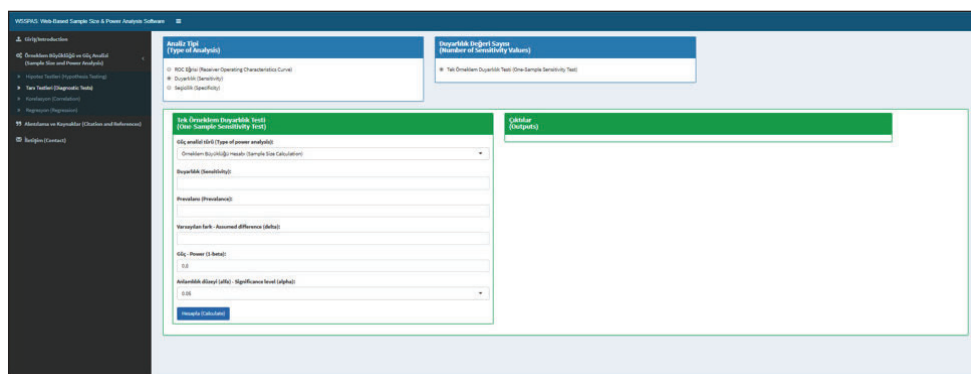
FIGURE 4: "One-Sample Sensitivity Test" menu
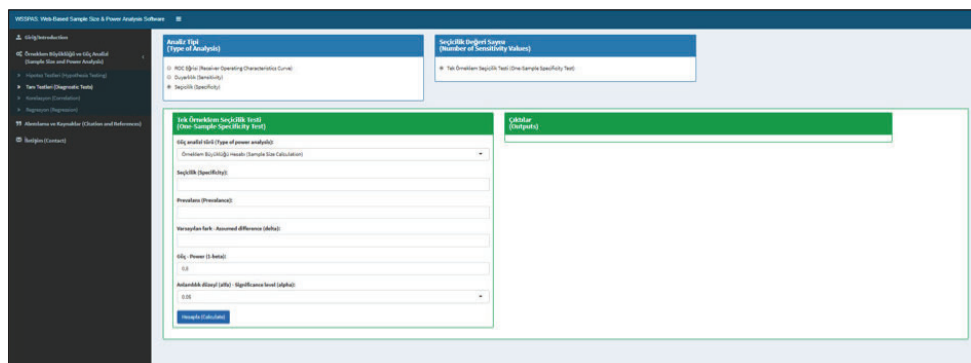
FIGURE 5: "One-Sample Specificity Test" menu
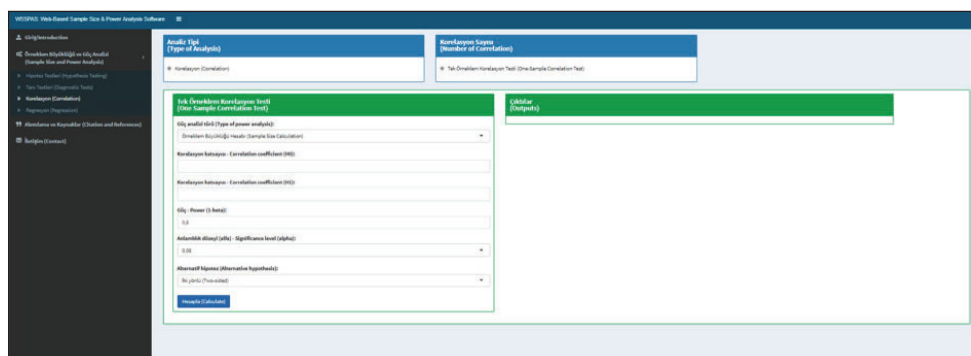
FIGURE 6: "Correlation" menu

Regression Menu

This menu in the developed web-based software consists of two sub-sections, which are "Linear Regression Analysis" and "Logistic Regression Analysis". In the study of sample size calculation and power analysis, if the dependent variable is quantitative, the menu "Linear Regression Analysis" is selected. However, if the dependent variable is qualitative, the "Logistic Regression Analysis" menu is selected. If the type of the dependent variable is quantitative and the number of independent variable is one, the "Simple Linear Regression" sub-section is used. Similarly, the "Simple Logistic Regression" sub-section is used if the dependent variable is of type qualitative and the number of independent variable is one. A screenshot of the regression menu is given below (Figure 7).

ACCESS TO THE DEVELOPED INTERACTIVE WEB APPLICATION

The developed interactive web application can be accessed free at http://biostatapps.inonu.edu.tr/WSS-PAS and this web tool will be updated upon the updated R software packages, including shiny, shiny-dashboard, pwr, powerAnalysis, powerMediation, MKmisc and rhandsontable.[17-23] Scripts were written for calculations that could not be done by these packages.

▌ RESULTS

A researcher would like to determine whether a diet program will reduce the concentration of C-Reactive Protein (CRP), a determinant of heart disease risk. For this purpose, the researcher is measuring the CRP concentrations at the end of the first month, the second month and the fourth month in a group of patients. It is assumed that the mean of the CRP concentration measurements is 4.8 with a standard deviation of 0.96 in the first month; 5.2 with a standard deviation of 1.04 in the second month; 4.1 with a standard deviation of 0.82 in the fourth month. What should the sample size (n) be when considering $\alpha = 0.05$ and $\beta = 0.20$ (1-β = 0.80 power)?

When the sampling size calculation stages were selected as Type of power analysis: Sample Size Calculation; Type of effect: Wihin-subject; Number of groups: 1; Number of measurements: 3; Effect size: Means: 4.8, 5.2, 4.1; Within-group standart deviations: 0.96, 1.04, 0.82; The calculated effect size: 1.01; Correlation among measurements: 0.50; Nonsphericity correction:1; Power (1-beta): 0.80 and Significance Level (alpha): 0.05, it was calculated that minimum 11 patients should be taken to the study. A screenshot of this calculation is given below (Figure 8).
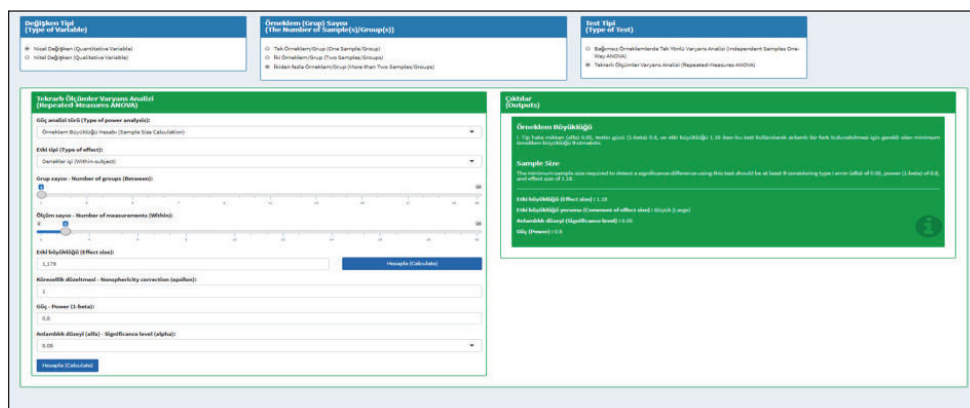

FIGURE 7: "Simple Linear Regression" menu

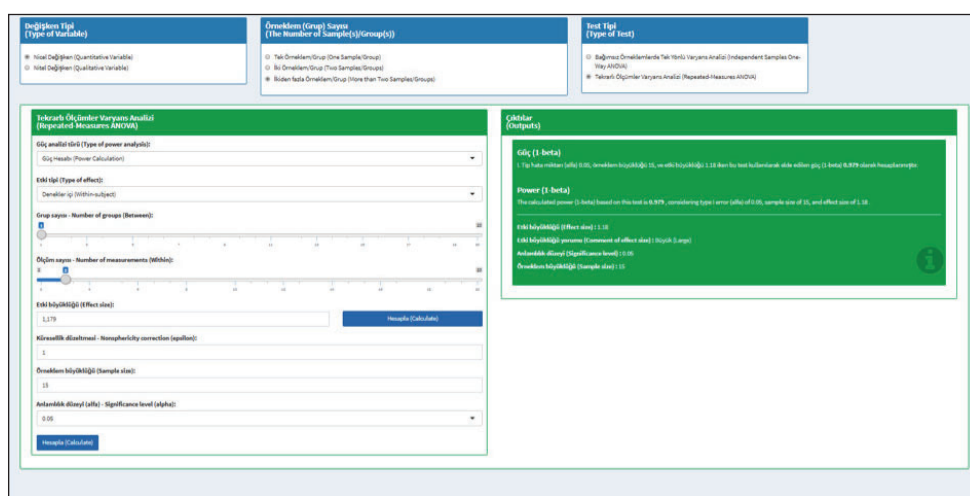**FIGURE 8:** Input and output for the sample size of "Repeated-Measures ANOVA"

FIGURE 9: Input and output for calculating the power of "Repeated-Measures ANOVA"

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 15 individuals? When the same values for this calculation are entered in the corresponding software and the sample size is taken 15 individuals, the power of the performed research is found to be 98%. A screenshot of this result is presented below (Figure 9).

A HYPOTHETICAL APPLICATION FOR INDEPENDENT SAMPLES ONE-WAY ANOVA

A pharmaceutical company wants to test three different formulations of a pain medication for a migraine. For this application, it is assumed that the first group has a pain mean of 9, the second group has a pain mean of 7, the third group has a pain mean of 6, the standard deviation of the first group is 1.8, the standard deviation of the second group is 1.4 and the standard deviation of the third group is 1.2, what should the sample size (n) be, when it assumed that the sample size in each group is equal and α=0.05 and β=0.20 (1-β=0.80 power)?

As a result of the calculation of the sample size, it is necessary to take at least 6 individuals in each group. Screenshot for this calculation is shown below (Figure 10).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 8 individuals in each group? When the same values for this calculation are entered in the corresponding software and the sample size is taken 8 individuals for each group, the power of the performed research is found to be 94%. A screenshot of this result is presented below (Figure 11).

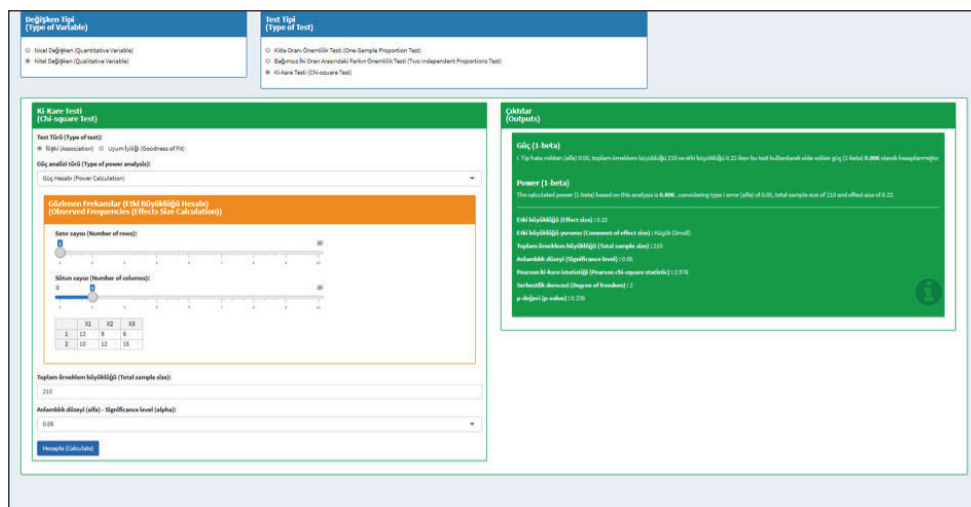**FIGURE 10:** Input and output for the sample size of "Independent Samples One-Way ANOVA"

FIGURE 11: Input and output for calculating the power of "Independent Samples One-Way ANOVA"

A HYPOTHETICAL APPLICATION FOR CHI-SQUARE TEST

A dietitian wants to examine the relationship between the results of three different exercise studies in a group of obese women and men. The pilot study consisted of 12 obese women participants and 10 obese men participants in the 1st exercise program, 9 obese women participants and 12 obese men participants in the 2nd exercise program, 6 obese women participants and 15 obese men participants in the 3rd exercise program. When it is assumed that $\alpha = 0.05$ and $\beta = 0.20$ (1-β = 0.80 power), what should the sample size (n) be?

As a result of the calculation of the sample size, it is found that at least 208 obese individulas should be taken into the study. A screenshot of this calculation is given below (Figure 12).
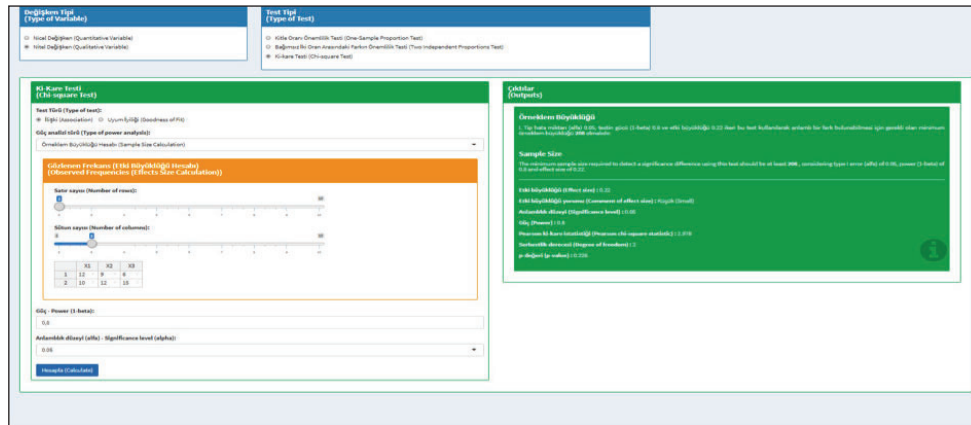
FIGURE 12: Input and output for the sample size of "Chi- Square Test"

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 210 patients with obese? When the same values for this calculation are entered in the corresponding software and the sample size is taken 210 patients with obese, the power of the performed research is found to be 81%. A screenshot of this result is presented below (Figure 13).

A HYPOTHETICAL APPLICATION FOR TWO PAIRED SAMPLES T-TEST

In a study, it is desirable to test how much the mean arterial blood pressure (MAP) of a new drug is reduced in 12 minutes in patients with hypertension. Before the experiment, the MAP mean of the hypertensive patient group (N = 200) is known 110 mmHg and the MAP average (μs) is known 90 mmHg. When the standard deviation of the differences is 30 mmHg, α = 0.05 and 1-β = 0.80, what should the sample size (n) be?

As a result of the calculation of the sample size, it is found that at least 20 patients with hypertension should be taken into the study. A screenshot of this calculation is given below (Figure 14).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 30 patients with hypertension? When the same values for this calculation are entered in the corresponding software and the sample size is taken 30 patients with hypertension, the power of the performed research is found to be 94%. A screenshot of this result is presented below (Figure 15).

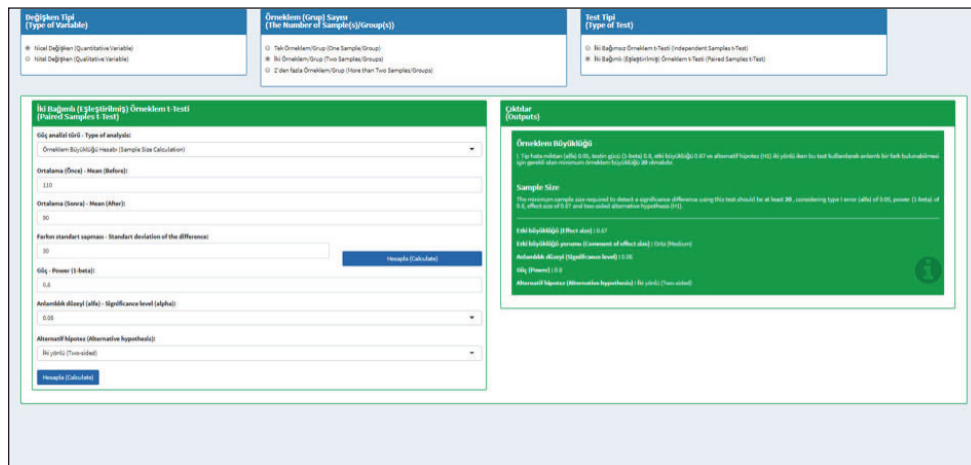**FIGURE 13:** Input and output for calculating the power of "Chi- Square Test"

233

FIGURE 14: Input and output for the sample size of "Two Paired Samples t- Test"
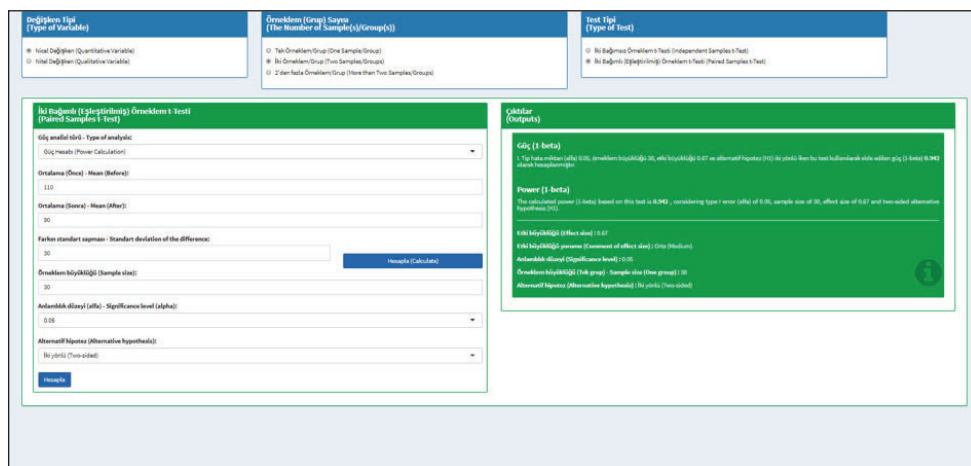

FIGURE 15: Input and output for the sample size of "Two Paired Samples t- Test"

A HYPOTHETICAL APPLICATION FOR TWO INDEPENDENT SAMPLES T-TEST

A clinical dietician would like to compare the efficacy of two different dietary applications (D_1 and D_2) for diabetic patients. After 8 weeks of treatment, each patient's fasting blood glucose level will be measured. Dietitian predicts the mean fasting blood glucose level for D_1 is 95 mg/dl, and the mean fasting blood glucose level for D_2 is 90 mg/dl. It also assumes that the standard deviation of the fasting blood glucose level distribution for D_1 is 15 mg/dl and the standard deviation of the fasting blood glucose level distribution for D_2 is 12 mg/dl. The dietician would like to accept the number of samples needed in each group as equal size groups. When the values of $\alpha = 0.05$ and $\beta = 0.20$ (1-β = 0.80 power) are taken into consideration, what should the sample size (n) be? When the values of $\alpha = 0.05$ and $\beta = 0.20$ (1-β = 0.80 power) are taken into consideration, what should the sample size (n) be?

As a result of sample size calculation; at least 117 diabetic patients in each group should be taken into the study. Screenshot for this calculation is shown below (Figure 16).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 120 patients with hypertension in each group? When the same values for this calculation are entered in the corresponding software and the sample size is taken 120 patients with hypertension in each group, the power of the performed research is found to be 81%. A screenshot of this result is presented below (Figure 17).
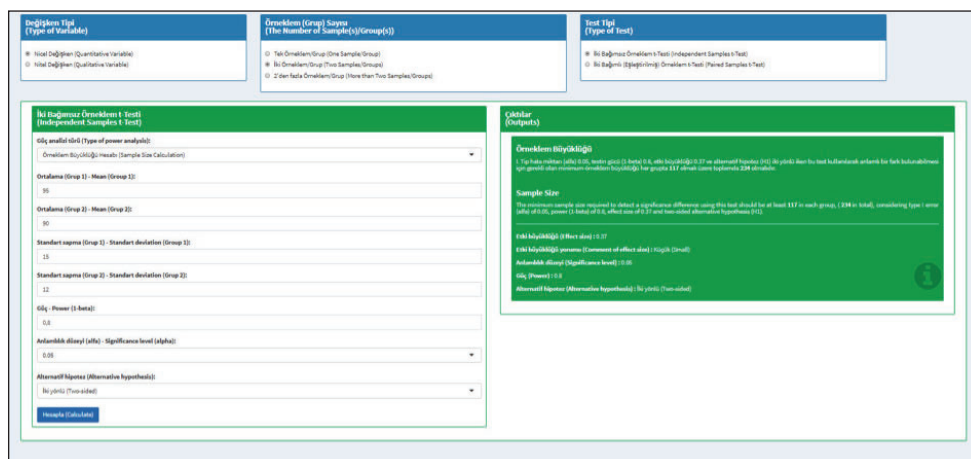
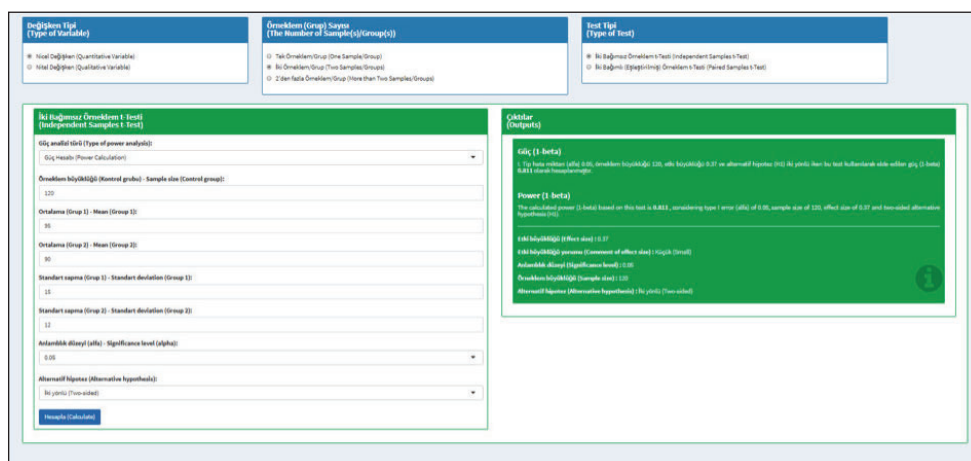**FIGURE 16:** Input and output for the sample size of "Two Independent Samples t- Test"

FIGURE 17: Input and output for calculating the power of "Two Independent Sample t- Test"

A HYPOTHETICAL APPLICATION FOR TWO INDEPENDENT PROPORTIONS TEST

In a clinical trial to compare rates of urinary tract infections, it is assumed that the ratio for the first group is 0.73 and the ratio for the second group is 0.61. When it is assumed that the sample size in both groups is equal and $\alpha = 0.05$ and $\beta = 0.20$ ($1-\beta = 0.80$ power), what should the sample size (n) be?

As a result of the calculation of the sample size, at least 238 individuals in each group have to be taken into the study. A screenshot of this calculation is given below (Figure 18).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 250 individuals in each group? When the same values for this calculation are entered in the corresponding software and the sample size is taken 250 individuals in each group, the power of the performed research is found to be 82%. A screenshot of this result is presented below (Figure 19).
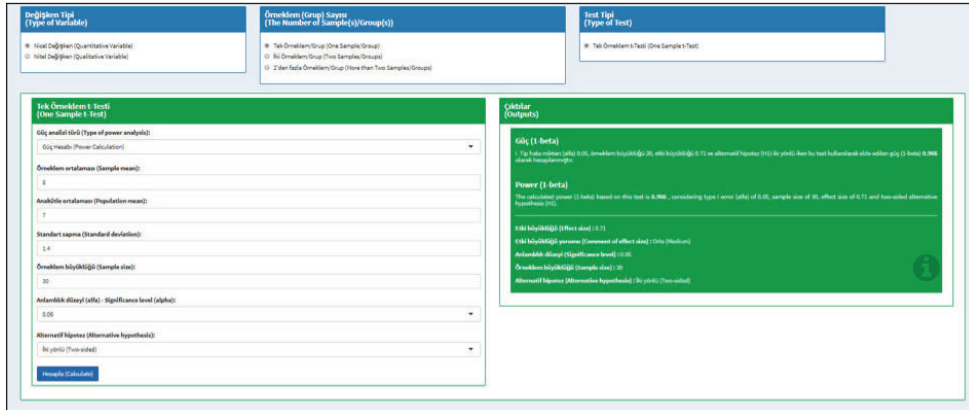
A HYPOTHETICAL APPLICATION FOR ONE-SAMPLE T-TEST

In the screening studies, it was determined that in patients with coronary bypass surgery, the mean length of stay after surgery was 7 days and the standard deviation was 1.4 days. What should the sample size (n) be for the values of $\alpha = 0.05$, $\beta = 0.20$ in a research that is assumed to be 8 days of mean length of stay and is planned to be conducted in Inonu University Faculty of Medicine Department of Cardiovascular Surgery?

As a result of the calculation of the sample size, at least 18 patients had to be involved in the study. A screenshot of this calculation is given below (Figure 20).

FIGURE 18: Input and output for the sample size of "Two Independent Proportions t- Test"

FIGURE 19: Input and output for calculating the power of "Two Independent Proportions t- Test"

FIGURE 20: Input and output for the sample size of "One-Sample t- Test"

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 30 individuals? When the same values for this calculation are entered in the corresponding software and the sample size is taken 30, the power of the performed research is found to be 97%. A screenshot of this result is presented below (Figure 21).

A HYPOTHETICAL APPLICATION FOR ONE-SAMPLE PROPORTION TEST

The prevalence of obesity in studies conducted throughout Turkey was determined to be about 0.20. What should the sample size (n) be for the values of $\alpha = 0.05$, $\beta = 0.20$ ($1-\beta = 0.80$ power) in a study that is planned

to be conducted in the Department of Internal Medicine and assumed that the rate of obesity in individuals is 0.25?

As a result of the calculation of the sample size, at least 546 patients had to be taken into the study. A screenshot of this calculation is given below (Figure 22).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 600 individuals? When the same values for this calculation are entered in the corresponding software and the sample size is taken 600, the power of the performed research is found to be 84%. A screenshot of this result is presented below (Figure 23).

FIGURE 21: Input and output for calculating the power of "One-Sample t- Test"
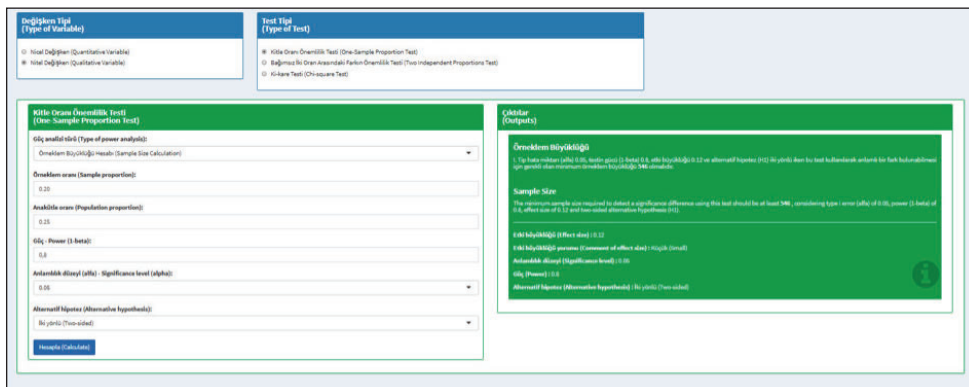
FIGURE 22: Input and output for calculating the power of "One-Sample t- Test"
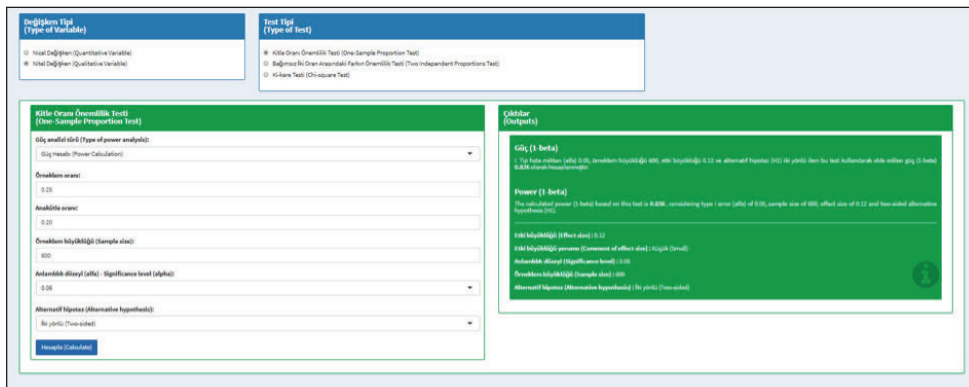
FIGURE 23: Input and output for the sample size of "One-Sample Proportion Test"

A HYPOTHETICAL APPLICATION FOR ONE ROC CURVE

In a study conducted in a radiology clinic, the diagnostic accuracy of Magnetic Resonance (MR) results, one of the imaging techniques, in prostate cancer diagnosis, is being investigated. When the 0.875 value of the area under the expected curve in the study is assumed to be significant from 0.5 null hypothesis test value and $\alpha = 0.05$ and $\beta = 0.20$ (1-β = 0.80 power) are taken, what should the sample size (n) be?

As a result of the calculation of the sample size, it is found that at least 8 individuals should be taken into the study. A screenshot of this calculation is given below (Figure 24).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 10 individuals? When the same values for this calculation are entered in the corresponding software and the sample size is taken 10, the power of the performed research is found to be 94%. A screenshot of this result is presented below (Figure 25).

FIGURE 24: Input and output for the sample size of "One-Sample ROC Curve Test"

FIGURE 25: Input and output for calculating the power of "One-Sample ROC Curve Test"

A HYPOTHETICAL APPLICATION FOR ONE-SAMPLE SENSITIVITY TEST AND ONE-SAMPLE SPECIFICITY TEST

A study aims to determine the sensitivity of a newly developed instrument to Obstructive Sleep Apnea (OSA) having a prevalence of 0.20 in patients referred to the Chest Diseases clinic. It is supposed that the reference test has sensitivity 85% and the specificity 80%. Assuming that this new test has 95% sensitivity and the 95% specificity when it is accepted that α = 0.05 and β = 0.20 (1-β = 0.80 power), what should the sample size be (n)?

As a result of the calculation of the sample size for sensitivity and specificity, it is found that minimum 340 individuals (272 in case group and 68 in control group) and minimum 47 individuals (10 in case group and 37 in control group) should be taken to the research. The screenshots of the calculations are given below (Figures 26, 27).

What is the power for sensitivity and specificity when we assume that the research is performed on 75 individuals in each group considering the same values above? When the same values for the calculations are entered and the sample size 75 in each group is taken, the calculated power values are 83% for the sensitivity and 99% for the specificity. The screenshots of this results are presented below (Figure 28, 29).

FIGURE 26: Input and output for the sample size of "One-Sample Sensitivity Test"

FIGURE 27: Input and output for the sample size of "One-Sample Specificity Test"
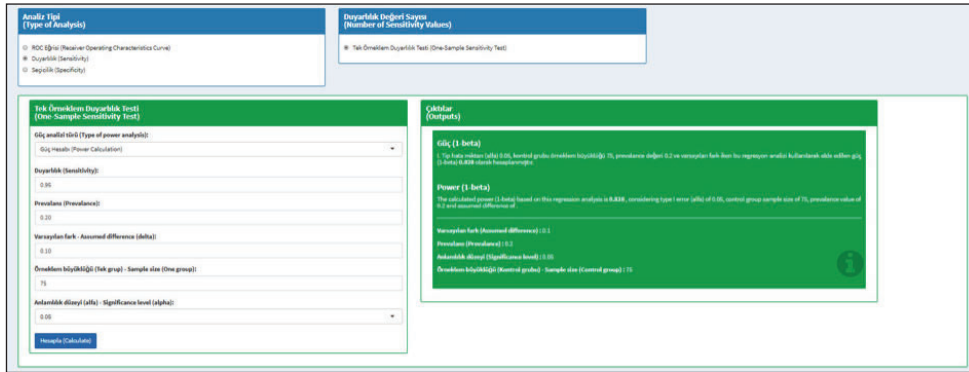
FIGURE 28: Input and output for calculating the power of "One-Sample Sensitivity Test"
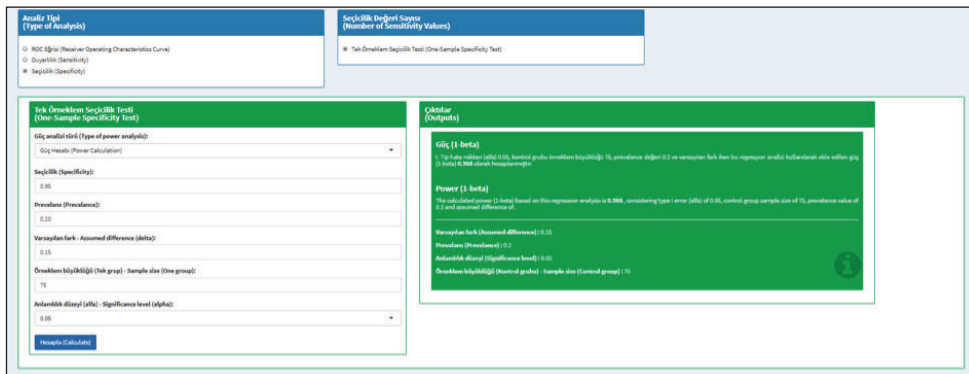

FIGURE 29: Input and output for calculating the power of "One-Sample Specificity Test"

A HYPOTHETICAL APPLICATION FOR ONE SAMPLE CORRELATION TEST

In a study conducted across Turkey, a relationship between the number of cigarettes smoked per day and the Systolic Blood Pressure has been found at 0.50 level. A study is planned to be conducted among the patients who are admitted to Family Medicine Smoking Cessation policlinic. In this study, the relationship between the two variables is expected to be 0.76. If we assume that $\alpha = 0.05$ and $\beta = 0.20$ (1-β = 0.80 power) in these conditions, what should the sample size (n) be?

As a result of the calculation of the sample size, it is found that at least 43 individuals should be taken to the study. A screenshot of this calculation is given below (Figure 30).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 45 individuals? When the same values for this calculation are entered in the corresponding software and the sample size is taken 45, the power of the performed research is found to be 83%. A screenshot of this result is presented below (Figure 31).
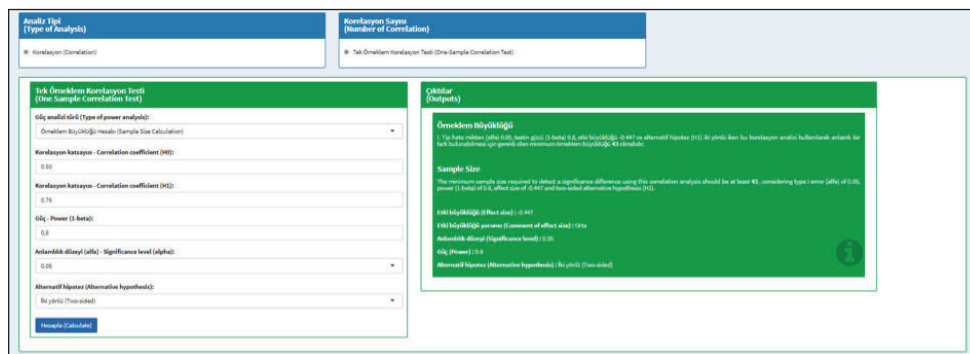

FIGURE 30: Input and output for the sample size of "One-Sample Correlation Test"

FIGURE 31: Input and output for calculating the power of "One-Sample Correlation Test"

A HYPOTHETICAL APPLICATION FOR SIMPLE LINEAR REGRESSION ANALYSIS

In a medical center, a study is planned to show that there is a linear relationship between body weight and systolic blood pressure among patients who applied to the internal medicine policlinic. It is assuming that the standard deviation value for systolic blood pressure (dependent variable) is 1.2, the standard deviation value for body weight (independent) is 0.3, and the regression coefficient is 3.2. When it is assumed that $\alpha = 0.05$ and $\beta = 0.20$ (1-β = 0.80 power), what should the sample size (n) be?

As a result of the calculation of the sample size, it is found that at least 7 individuals should be taken for the study. A screenshot of this calculation is given below (Figure 32).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 12 individuals? When the same values for this calculation are entered in the corresponding software and the sample size is taken 12 individuals, the power of the performed research is found to be 98%. A screenshot of this result is presented below (Figure 33).

FIGURE 32: Input and output for calculating the power of "One-Sample Correlation Test"

FIGURE 33: Input and output for calculating the power of "Simple Linear Regression Analysis"

A HYPOTHETICAL APPLICATION FOR SIMPLE LOGISTIC REGRESSION ANALYSIS WITH A CONTINUOUS INDEPENDENT VARIABLE

After viewing videotapes containing violent sequences, a study is planned to investigate the relationship between post-traumatic stress disorder and heart rate. It is assumed that the heart rate is normally distributed. The incidence rate is thought to be 7% (0.07) among soldiers. When researchhers are assuming that odds ratio=2, $\alpha = 0.05$ and $\beta = 0.10$ (1-β = 0.90 power) with a two-sided test, what should the sample size (n) be?

As a result of the calculation of the sample size, it is found that at least 336 individuals should be taken for the study. A screenshot of this calculation is given below (Figure 34).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 350 individuals? When the same values for this calculation are entered in the corresponding software and the sample size is taken 350 individuals, the power of the performed research is found to be 91%. A screenshot of this result is presented below (Figure 35).

FIGURE 34: Input and output for the sample size of "Simple Logistic Regression Analysis with a Continuous Independent Variable"
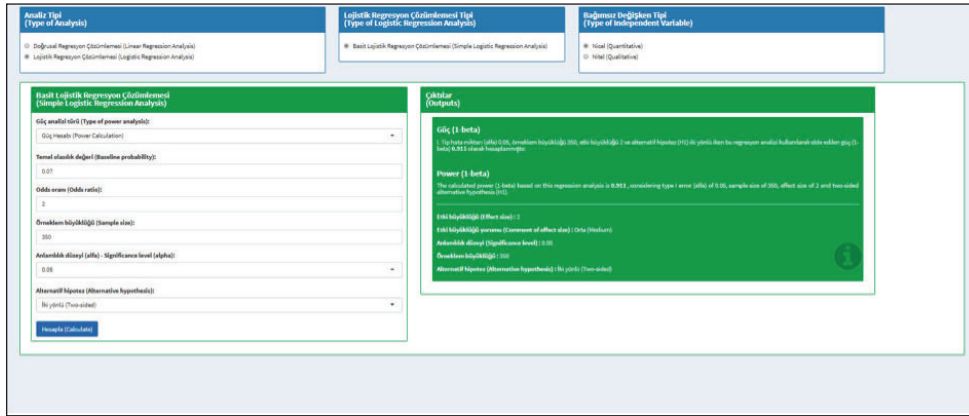
FIGURE 35: Input and output for calculating the power of "Simple Logistic Regression Analysis with a Continuous Independent Variable"

A HYPOTHETICAL APPLICATION FOR SIMPLE LOGISTıC REGRESSION ANALYSIS WITH A BINARY INDEPENDENT VARIABLE

A study will be carried out to investigate the relationship between post-traumatic stress disorder and gender. The incidence rate is thought to be 7% (0.07) among males. When researchers are assuming that sample proportion of 50% (0.50), an alternative probability of 10% (0.10), $\alpha = 0.05$ and $\beta = 0.10$ (1-β = 0.90 power) with a two-sided test, what should the sample size (n) be?

As a result of the calculation of the sample size, it is found that at least 3628 individuals should be taken for the study. A screenshot of this calculation is given below (Figure 36).

Considering the same values on the above example, what is the achieved power when it is assumed that the research is conducted with 3800 individuals? When the same values for this calculation are entered in the corresponding software and the sample size is taken 3800 individuals, the power of the performed research is found to be 91%. A screenshot of this result is presented below (Figure 37).
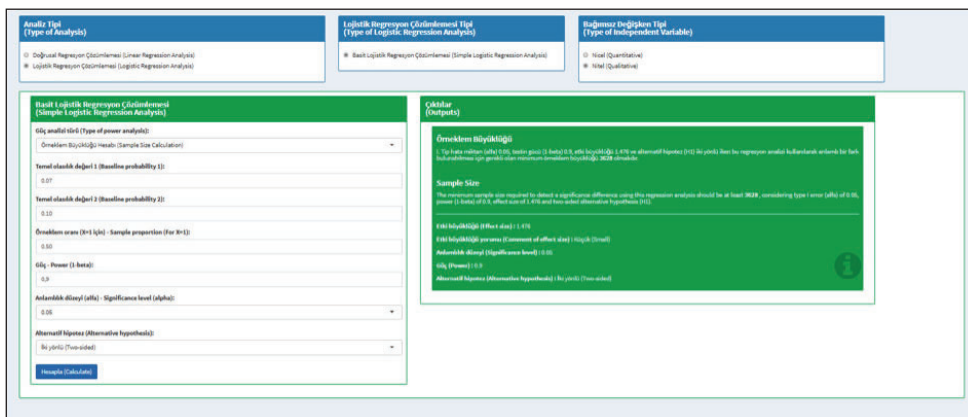

FIGURE 36: Input and output for the sample size of "Simple Logistic Regression Analysis with a Binary Independent Variable"

FIGURE 37: Input and output for calculating the power of "Simple Logistic Regression Analysis with a Binary Independent Variable"

DISCUSSION

Calculation of sample size and power analysis is an important issue to consider when planning biomedical research. While the adequate sample size has a crucial importance in reaching the correct results, the more or less than of required sample size may be disadvantageous for some reasons (time, labor, costs etc.).[24] The sample size should be calculated with reference to previous research or pilot work in the planning stage of the study.[25] In this study, a user-friendly web-based software which calculates the sample size required for hypothesis tests, ROC curve, diagnostic tests, correlation, simple linear regression and simple logistic regression analysis was developed and explained with various examples.

As the sample size and power analysis are calculated with different formulas for each research design and statistical hypothesis test, it takes a long time to process manually and errors may occur in the calculations. In order to remove these difficulties, many softwares (PASS[26], Minitab[27], XLSTAT-Power [28], G- Power[29], SPSS Sample Power[30], MedCalc[31], Statistica[32], GraphPad Prism[32], Stata[33], etc.) have been developed that perform sample size and power analysis. Most of these software programs (PASS, Minitab, XLSTAT- Power, SPSS Sample Power, MedCalc, Statistica, Stata, GraphPad Prism, etc.) are commercial, or dependent on an operating system (Windows, Mac). The proposed web-based application that is free and does not depend on any operating system has different and superior properties from other softwares. Thus, sample size and power analysis can be easily calculated online via the developed web-based software.

One of the important outputs of web-based software developed in this research is that it can easily calculate sample size and power analysis for repeated measures one way ANOVA. The sample size and power analysis for this method can also be calculated by means of many commercial (PASS, XLSTAT-Power, etc.) and limited number of free (G-Power, etc.) softwares. However, this softwares depends on the operating systems (Windows, Mac and so on) to be installed on personal computers. Quite a few web-based programs (GLIMMPSE, WebPower, and so fort) are also available for calculating sample size and power analysis of repeated measures designs. The developed web-based software is both free and independent of the operating systems, making a significant contribution to the area of interest in the calculation of sample size and power analysis.

The statistical significance of a result is the probability that the observed difference between the two variables is due to chance.[14] Clinical significance, on the other hand, is an indicator of how effective research findings can be on patients.[34] Statistically significant research results are not always expected

to be clinically significant. Similarly, it is not able to be said that the results of a statistically insignificant study are not clinically meaningless. In this context, the effect size, known as a sign of clinical significance, is defined as a measure giving the minimum amount of change that we want to be able to accurately determine in the end result.[35]

On the other hand, in a clinical study, significance (the probability that an observed outcome of an experiment or trial is not due to chance alone), direction (positive or negative), and magnitude (absolute or relative) are critical factors in the interpretation of treatment effects. In order to emphasize the importance of the results of the clinical trial, these factors need to be presented during the results section. However, in most cases, clinicians evaluate with only p, a statistical significance measure, the results obtained from the clinical study. But that is not enough. Because in a clinical study, p-value gives information about the efficacy of treatment between experimental and control groups. In clinical studies, effect size statistics which are a measure of the size of the effect must be reported in addition to the presence of a treatment effect. The effect size estimation provides an interpretable value on the direction and magnitude of a treatment effect and allows the results to be compared with the results of other studies using comparable measures.[36]

On the other hand, when the effect size is exactly zero with a sufficiently large sample, a statistical test will almost always make a statistically significant difference. For this reason, reporting only the significant p-value for a clinical trial is not enough for clinicians to fully understand the results. Unlike significance tests, the effect size is independent of the sample size (the number of observations), while the statistical significance (p-value) is both dependent on sample size and effect size. For this reason, p values are considered confusing due to their dependence on sample size. Sometimes a statistically significant result may only mean that a large sample size is used. Therefore, the results of studies limited by significance to p-value are not reliable. Therefore, when the research results are reported, the effect size should be reported in the Results section together with the p-value.[14]

To the best of our knowledge, existing software that calculates sample size and power analysis does not have a screen that warns about the choice of effect size. The module offers suggestions on effect size selection for clinicians. It is thought that web-based software developed with this aspect will make an important contribution to this field.

One of the other features of this software is also to give outputs of sample size and power analysis with the scientific expressions and allow researchers copy and paste the outputs to the related subsections of their scientific studies.

As a result, the base software developed using R Shiny can easily calculate sample size and power analysis for Hypothesis tests, Diagnostic Tests, Correlation and Regression analyzes in clinical trials. In the future studies, it is aimed to further strengthen the software by adding modules that can calculate sample size and power analysis for different multivariate statistical and machine learning methods.

Source of Finance

Conflict of Interest

Authorship Contributions

Idea/Concept: *Ahmet Kadir Arslan, Şeyma Yaşar, Cemil Çolak, Saim YOLOĞLU*; **Design:** *Ahmet Kadir Arslan, Şeyma Yaşar, Cemil Çolak*; **Control/Supervision:** *Ahmet Kadir Arslan, Şeyma Yaşar, Cemil Çolak, Saim YOLOĞLU*; **Data Collection And/Or Processing:** *Ahmet Kadir Arslan, Şeyma Yaşar*; **Analysis And/Or Interpretation:** *Ahmet Kadir Arslan, Şeyma Yaşar*; **Literature Review:** *Ahmet Kadir Arslan, Şeyma Yaşar*; **Writing The Article:** *Ahmet Kadir Arslan, Şeyma Yaşar, Cemil Çolak*; **Critical Review:** *Saim YOLOĞLU, Cemil Çolak*

REFERENCES

1. Dupont WD, Plummer WD Jr. Power and sample size calculations. A review and computer program. Control Clin Trials 1990;11(2):116-28.
2. Zodpey SP. Sample size and power analysis in medical research. Indian J Dermatol Venereol Leprol 2004;70(2):123-8.
3. Şenocak M. Klinik Biyoistatistik. 1. Baskı. Ankara: Nobel Tıp Kitabevleri Ltd Şti; 2009. p.216.
4. Vacha-Haase T, Nilsson JE, Reetz DR, Lance TS, Thompson B. Reporting practices and APA editorial policies regarding statistical significance and effect size. Theory & Psychology 2000;10(3):413-25.
5. Wassertheil-Smoller S, Smoller J. Biostatistics and Epidemiology: a Primer for Health and Biomedical Professionals: 4th ed. New York: Springer; 2015. p.260.
6. Özdamar K. Modern Bilimsel Araştırma Yöntemleri. 1. Baskı. Eskişehir: Kaan Kitabevi;2003.p.270.
7. Cohen J. The earth is round (p<.05). American Psychologist 1994;49:997.
8. Kirby A, Gebski V, Keech AC. Determining the sample size in a clinical trial. Med J Aust 2002;177(5):256-7.
9. Vacha-Haase T, Thompson B. How to estimate and interpret various effect sizes. J Couns Psychol 2004;51(4):473-81.
10. Yıldırım HH, Yıldırım S. [On hypothesis testing, confidence interval, effect size and noncentral probability distributions]. Elementary Education Online 2011;10(3):1112-3.
11. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988. p.474.
12. Davis JA. Elementary Survey Analysis. 1st ed. New Jersey, Prentice-Hall; 1971. p.195.
13. Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 1st ed. Chicago: Rand McNally College Publishing; 1979. p.489.
14. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012;4(3):279-82.
15. Chen H, Cohen P, Chen S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. Commun Stat Simul Comput 2010;39(4):860-4.
16. Süt N. [Sample size determination and power analysis in clinical trials]. RAED Dergisi 2011;3(1-2):29-33.
17. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Shiny: web application framework for R. R package version 0.13. R Foundation for Statistical Computing; 2016.
18. Chang W. shinydashboard: Create Dash boards with 'Shiny'. R package version 0.5. 3. R Foundation for Statistical Computing; 2016.
19. Champely S, Ekstrom C, Dalgaard P, Gill J, Weibelzahl S, Anandkumar A, et al. pwr package. R Foundation for Statistical Computing; 2018.
20. Fan FY. Package 'powerAnalysis'. R Foundation for Statistical Computing; 2017.
21. Qiu W. Package 'powerMediation'. R Foundation for Statistical Computing; 2018.
22. Kohl M. Package 'MKmisc'. R Foundation for Statistical Computing; 2018.
23. Owen J, Allen J, Xie Y, Martoglio E, Ger I, Marcin W, et al. Package 'rhandsontable'. R Foundation for Statistical Computing; 2018.
24. Gravetter FJ, Wallnau LB. Introduction to the t statistic. Essentials of Statistics for the Behavioral Sciences. 8th ed. USA: Cengage Learning; 2014. p.249-313.
25. Kumar GS. Importance of sample size in clinical trials. Int J Clin Exp Physiol 2014;1(1):10-2.
26. Hintze J. NCSS and PASS software. Kaysville, UT: Number Cruncher Statistical Systems; 2004.
27. Minitab. MINITAB statistical software. Minitab Release; 2000.
28. Addinsoft. XLSTAT statistical and data analysis solution. Boston, USA 2018.
29. Faul F, Erdfelder E, Buchner A, Lang A. G* Power Version 3.1. 7 [computer software]. Kiel, Germany: Universität Kiel; 2009.
30. Borenstein M, Hedges L, Rothstein H, Cohen J, Schoenfeld D. IBM SPSS SamplePower Release 3.0.1. 2010.
31. Schoonjans F, Zalata A, Depuydt CE, Comhaire FH. MedCalc: a new computer program for medical statistics. Comput Methods Programs Biomed 1995;48(3):257-62.
32. StatSoft. STATISTICA (data analysis software system), version 6. Tulsa, USA: 2001. p.150:
33. GraphPad Prism Version 6.00 for Windows La Jolla California USA: GraphPad Software; 2016.
34. Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. Int J Sports Phys Ther 2014;9(5):726-36.
35. Alpar R. Spor, Sağlık ve Eğitim Bilimlerinden Örneklerle Uygulamalı İstatistik ve Geçerlik-Güvenirlik. 4. Baskı. Ankara: Detay Yayıncılık; 2010. p.668.
36. McGough JJ, Faraone SV. Estimating the size of treatment effects: moving beyond p values. Psychiatry (Edgmont) 2009;6(10):21-9.