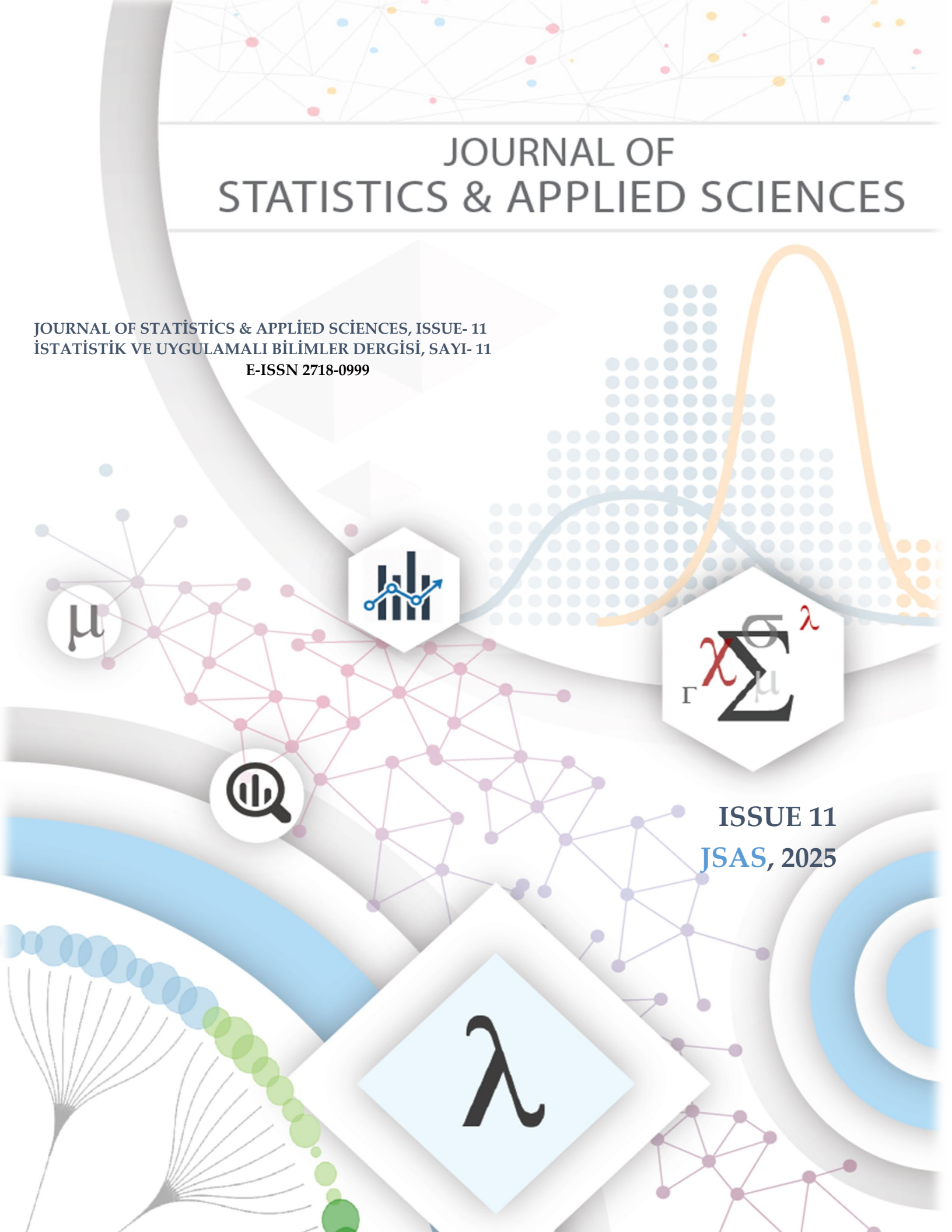
# JOURNAL OF STATISTICS & APPLIED SCIENCES

JOURNAL OF STATİSTİCS & APPLİED SCİENCES, ISSUE- 11
İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ, SAYI- 11
E-ISSN 2718-0999



ISSUE 11
JSAS, 2025

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

# -JOURNAL OF STATİSTİCS & APPLİED SCİENCE-
# -İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ-

**Aralık 2025**

**Issue: 11**

**Sayı: 11**



**Yayın Türü:** 6 Aylık, Uluslararası, Hakemli

**Publication Type:** 6 Monthly, International, Refereed

**e-ISSN:** 2718-0999

**Dizinleme Bilgileri/ Abstracting and Indexing Services**

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

# BİLİM VE DANIŞMA KURULU



# EDİTÖR KURULU



# YABANCI DİL EDİTÖRLERİ

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

# İÇİNDEKİLER (CONTENT)

## ARAŞTIRMA MAKALELERİ (RESEARCH ARTİCLES)



## Derleme Makale (Review Article)



## Editöre Mektup (Letter to the Editor)

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

# ÖNSÖZ

Sayın Okurlar,

İstatistik ve Uygulamalı Bilimler Dergisi'nin 11. sayısını siz değerli okuyucularımızla paylaşmaktan mutluluk duyuyoruz. Bu sayımızda üç özgün araştırma makalesi, bir derleme makalesi ve bir editöre mektup yer almaktadır. Sayı, teorik ve uygulamalı istatistik alanlarına disiplinlerarası katkılar sunan çalışmaları içermektedir.

İlk araştırma makalesinde Jose Moral De La Rubia, parametrik yedi sayılı özet temelli geliştirilmiş bir normallik testi önermektedir. İkinci çalışmada, Yalova iline ait meteorolojik veriler kullanılarak makine öğrenmesi yöntemleriyle rüzgar enerjisi üretimi tahmini yapılmıştır. Üçüncü araştırma, ilaç üretim süreçlerinde fire oranlarının izlenmesine yönelik istatistiksel süreç analizine dayanmaktadır. Sayının derleme makalesinde ise, Türkiye'deki proje çizelgeleme konulu lisansüstü tezler bibliyometrik açıdan incelenmiştir. Son olarak, editöre mektup niteliğindeki çalışmada, p-hacking uygulamalarının bilimsel araştırmalardaki güvenilirliği nasıl etkilediği tartışılmaktadır.

Bu sayıya katkı sunan tüm yazarlara ve değerlendirme sürecinde görev alan hakemlere teşekkür eder, okuyucularımıza faydalı ve ilham verici bir okuma deneyimi dileriz.

Saygılarımızla,

**İstatistik ve Uygulamalı Bilimler Dergisi Editörü**

Dr. Abdulkadir Keskin

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

# PREFACE

Dear Readers,

We are pleased to present the 11th issue of the *Journal of Statistics and Applied Sciences* to our valued readers. This issue features three original research articles, one review article, and one letter to the editor. The collection offers interdisciplinary contributions to both theoretical and applied aspects of statistics.

The first research article by Jose Moral De La Rubia introduces a refined normality test based on the parametric seven-number summary. The second article focuses on forecasting wind energy production using machine learning methods and meteorological data from Yalova, Türkiye. The third article presents a statistical process analysis of defect rates in pharmaceutical production. The review article provides a bibliometric analysis of graduate theses in Türkiye on project scheduling. Finally, the letter to the editor discusses the issue of p-hacking and its potential threat to the credibility of scientific research.

We sincerely thank all authors for their valuable contributions and all reviewers for their critical evaluations. We hope this issue will be both informative and inspiring to our readers.

Sincerely,

**Editor of the Journal of Statistics and Applied Sciences**

Dr. Abdulkadir Keskin

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

*Research Article*

Received: date:07.04.2025
Accepted: date:04.06.2025
Published: date:30.06.2025

# Refined Normality Test Based on the Parametric Seven-Number Summary

**Jose Moral de la Rubia**[1*]

[1]School of Psychology, Universidad Autónoma de Nuevo León, Monterrey Mexico; jose.morald@uanl.edu.mx
Orcid: 0000-0003-1856-1458[1]
*Correspondence: jose.morald@uanl.edu.mx

**Abstract:** In 2022, a normality test based on the parametric seven-number summary was proposed. Its test statistic is the sum of squared standardized quantiles and was initially approximated by a chi-square distribution with seven degrees of freedom, without accounting for correlation among quantiles. Objective: To improve the test by incorporating these correlations. Two alternatives were proposed: (1) estimating the sampling distribution of the Q-statistic via bootstrap ($Q_B$), and (2) using a quadratic form with a correlation matrix of quantiles under normality ($Q_T$). Methods: The three variants ($Q$, $Q_B$, $Q_T$) were compared with the Shapiro-Wilk W-test in terms of accuracy (hit ratio) and statistical power. A total of 372 random samples were generated across 31 sample sizes from twelve continuous distributions. Correct classifications were compared using Cochran's Q test, and power was assessed via repeated-measures ANOVA. Results: $Q_B$ was significantly the most accurate and showed the highest average power compared to $Q$ and $Q_T$. Its accuracy was equivalent to that of the Shapiro–Wilk W-test, although the latter outperformed all three Q variants in average power. Conclusions: $Q_B$ is a suitable inferential extension of the seven-number summary for testing normality.

**Keywords:** Testing normality, normal distribution, quadratic forms; bootstrap, generalized chi-square distribution, inference statistics.

## 1. Introduction

Bowley [1] proposed the five-number summary to describe the distribution of a variable based on five positional statistics: the minimum, the three quartiles, and the maximum. In 1910, he expanded this summary to seven numbers by incorporating the first and last deciles, allowing for the definition of two extreme deviation ranges. Very low values fall below the 10th percentile, while very high values exceed the 90th percentile. The central tendency, or shoulder area, extends across the interquartile range, spanning from the first to the third quartile. Low values lie between the 10th and 25th percentiles, while high values fall between the 75th and 90th percentiles [2]. Thus, the seven-number summary provides a structured description of the distribution and facilitates score interpretation using six class intervals: [$min$, $p_{10}$] very low scores, ($p_{10}$, $p_{25}$] low scores, ($p_{25}$, $p_{50}$] low-medium scores, ($p_{50}$, $p_{75}$] medium-high scores, ($p_{75}$, $p_{90}$] high scores, and ($p_{90}$, $max$] very high scores [3-4].

In recent years, Moral [5] proposed a method for testing whether a sample has been drawn from a normal distribution using seven quantiles that are equispaced at two-thirds from the central point of a standard normal distribution. This approach is an adaptation of the seven-number summary for assessing normality. Since this summary is defined with respect to the normal distribution, it is referred to as parametric. Previously, it had only been used for descriptive and indicative purposes in normality assessment. However, Moral [5] proposed its application as an omnibus statistical test. Notably, in a similar vein, Avdović and Jevremović [6] also suggested assessing normality by comparing empirical and theoretical quantiles, though their approach is based on the entire set of $n$ sample data rather than on seven specific quantiles.

**Citation:** J. M. de la Rubia, "Refined Normality Test Based on the Parametric Seven-Number Summary," *Journal of Statistics and Applied Sciences*, no. 11, pp. 1–39, 2025, doi: 10.52693/jsas.1670945.

In the proposal of the test based on the Parametric Seven-Number Summary (PSNS), a parametric approximation using the chi-square distribution with seven degrees of freedom was employed. This approximation assumes independence among the standardized quantiles under the assumption of normality, with their sum of squares serving as the test statistic (Q-statistic). Although this convenient approach simplifies the calculations and appears to perform well in practice [5], it does not hold up theoretically.

The first objective of this study is to propose two variants of the test to assess the sampling distribution of the test statistic. In the first variant, the sampling distribution of the Q-statistic—along with the calculation of the p-value, type II error, and statistical power—is determined using the bootstrap procedure. This variant is denoted as $Q_B$, to distinguish it from the original proposal ($Q$). An R script was developed for this purpose [7]. See Appendices A (Q-statistic), B (bootstrap version of PSNS Q test), and C (statistical power of bootstrap version of PSNS Q test).

Specifically, 10,000 resampled datasets of the same size as the original sample are generated by either (a) resampling with replacement from the observed data (nonparametric bootstrap) or (b) drawing 2,000 random samples from a normal distribution where the location parameter is estimated by the sample mean and the scale parameter is estimated by the sample standard deviation with Bessel's correction (parametric bootstrap). The first approach produces the empirical bootstrap distribution, while the second generates the normative bootstrap distribution. The latter is used both to calculate the p-value (the proportion of values in the normative bootstrap distribution that exceeds the test statistic value in the original sample) and to determine the critical value (the $0.95^{th}$ quantile of normative bootstrap distribution). This critical value is then used to compute statistical power via rejection rate in the empirical bootstrap distribution.

A sample size of 2,000 was chosen for the normative bootstrap distribution, as it is considered sufficiently large for the estimation of p-values and critical values [8]. However, for the calculation of statistical power, the number of bootstrap samples was increased to 10,000. This modification allowed for a single simulation to be used, thus avoiding the need to define a rejection threshold (e.g., 5% or 50% rejection across 1,000 simulations) for classifying each simulation as a rejection when computing the rejection rate [9].

The second variant introduces a new statistic based on a quadratic form, where the symmetric matrix corresponds to the expected correlations between the seven quantiles under the null hypothesis of normality [10]. This statistic is denoted as $Q_T$, with the subscript T referring to the theoretical values under the null hypothesis of a standard normal distribution used in its computation. The p-value and statistical power of $Q_T$ are obtained using the generalized chi-square distribution, implemented via the CompQuadForm package in R [11]. See Appendix D (quadratic form version of PSNS Q test). The original version of the Q Test can be found in Appendix E (chi-square approximation to PSNS Q test).

The second objective of this study is to determine which of the proposed methods ($Q$, $Q_B$, and $Q_T$) demonstrates the best performance. To this end, the three variants of the normality test based on the PSNS, along with the Shapiro-Wilk $W$ test [12], are compared in terms of accuracy (proportion of correct classifications or hit rate) and statistical power. The comparisons are conducted across 31 sample sizes (ranging from 50 to 1,500 in increments of 50, plus 2,000) and 12 distributions (one normal and 11 non-normal). The W test is included using Royston's standardization procedure [13], as it is regarded as the most powerful and accurate test for sample sizes ranging from 3 to 2,000 observations [14-15]. Specifically, the largest original sample size considered in this study was 2,000, which is unusually large for an empirical study in the social and health sciences. For larger samples sizes, the Shapiro-Francia W' test [16] using Royston's standardization procedure [17] could have been applied.

This is a methodological study, in which the use of simulated data is considered the most appropriate. This type of data allows us to determine whether the null hypothesis of normality should be retained or rejected—something that is not possible with real scale-based data, where the distribution is assumed and ultimately unknown. However, since the purpose of the study is to apply the normality test to real-world datasets, a third objective has been added: to demonstrate the practical effectiveness of the proposed test variants using R scripts applied to real-world data.

To achieve this, a personality measurement instrument was administered to a sample of 385 students who applied for admission to a psychology program at a public university in northeastern Mexico. The study assessed alexithymia, a trait characterized by difficulty in identifying and expressing feelings, as well as a cognitively externalized style disconnected from the internal experiential world of emotions, dreams, and fantasies. This personality trait is evaluated using the Toronto Alexithymia Scale (TAS-20) [18], which was validated in Mexico by Moral [19].

Alexithymia—measured on an interval scale such as the total TAS-20 score—is expected to follow a normal distribution, given that this personality trait is influenced by both environmental and biological factors [18]. It is hypothesized that TAS-20 total scores within the central range of the distribution reflect adaptive or functional expressions of the trait, whereas scores in the tails reflect maladaptive manifestations (right tail) and exceptional emotional competence and awareness (left tail). Specifically, individuals with alexithymic personality disorder are expected to score in the right tail of the distribution. Empirical evidence supports this model [20].

**2. The Summary of the Seven Numbers to Evaluate Normality**

In the test based on the PSNS, the test statistic, symbolized by $Q$, is defined as the sum of the squares of the quantiles corresponding to the orders -2, -4/3, -3/4, 0, 2/3, 4/3, and 2 of a standard normal distribution. Before squaring and summing the seven sample quantiles, they are standardized [5]. Initially, the sample mean and standard deviation are used for standardization. Subsequently, they are re-standardized using the expected mean and standard deviation under the null hypothesis of a standard normal distribution. Since these variables are correlated, the sampling distribution of this quadratic form follows a generalized chi-square distribution [21]. However, Moral [5] suggested a simplified approach for calculating the p-value by approximating it with a chi-square distribution with seven degrees of freedom.

It is worth noting that the 2nd, 8th, 25th, 50th, 75th, 92nd, and 98th percentiles of a standard normal distribution are approximately evenly spaced. When rounded to two decimal places, the second percentile corresponds to -2.05, the eighth to -1.34, the twenty-fifth to -0.67, the fiftieth to 0, the seventy-fifth to 0.67, the ninety-second to 1.34, and the ninety-eighth to 2.05. Within this sequence, the quantiles are separated by a distance of 0.67 along the Z-score axis, except at the two extremes, where a slight discrepancy of 0.04 is observed (-1.34 - (-2.05) = 0.71 = 0.67 + 0.04 and 2.05 - 1.34 = 0.71 = 0.67 + 0.04).

To achieve a more precise alignment with evenly spaced quantiles, the following orders must be used: 0.0228, 0.0912, 0.2525, 0.5, 0.7475, 0.9088, and 0.9772. These yield the quantiles -2, -4/3, -2/3, 0, 2/3, 4/3, and 2, which are evenly spaced at two-thirds intervals within the range [-2, 2], centered at 0. These seven quantiles closely approximate the percentiles used in the seven-number summary, namely the minimum (0th percentile), 10th percentile, 25th percentile, 50th percentile, 75th percentile, 90th percentile, and maximum (100th percentile) of a standard normal distribution. Therefore, they are considered a summary of seven key values for a normal distribution [5], also referred to as the PSNS [22-23].

It should be noted that, since the normal distribution has infinite support over the interval $(-\infty, +\infty)$, it does not inherently have a minimum or maximum. However, any finite sample x of size $n$ drawn from a normal distribution will have a limited range of scores, bounded by the sample minimum and maximum, $[min(x), max(x)]$. Moreover, approximately 95.44% of the distribution lies within the interval $[-2, 2]$. For this reason, in the PSNS [5], the minimum is replaced by the 0.0228 quantile and the maximum by the 0.9772 quantile.

In asymptotic sampling, the distribution of the quantiles of a distribution with finite moments follows a normal distribution [10]. If $X$ is a random variable with a density function $f_X(x)$ and finite moments, the sample quantile of order $p$, denoted as $q_x(p)$, follows a normal distribution. Its arithmetic mean or expected value, denoted as $\mu[q_x(p)]$, corresponds to the population quantile $Q_X(p)$. The variance is given by the ratio of the product of the quantile order and its complement (numerator) to the product of the sample size and the square of the density function evaluated at the quantile (denominator). See Equation 1. The standard deviation or standard error, denoted as $\sigma[q_x(p)]$, is the square root of this ratio (Equation 2).

$$x = \{x_i\}_{i=1}^{n} = \{x_1, x_2, \dots, x_n\} \subseteq X$$
$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$
$$q_x(p) = x_{(i)} \sim N\left(\mu[q_x(p)] = Q_X(p), \sigma^2[q_x(p)] = \frac{p \times (1-p)}{n \times f_X^2[Q_X(p)]}\right) \quad (1)$$

$$\sigma[q_x(p)] = \sqrt{\frac{p \times (1-p)}{n \times f_X^2[Q_X(p)]}} \quad (2)$$

Equation 2 can be used to obtain the asymptotic standard errors of the seven quantiles that make up the PSNS, based on the orders of the quantiles ($p$ = 0.02275, 0.09121, 0.25249, 0.5, 0.74751, 0.90879, and 0.97725) of a standard normal distribution [2, 5]. Table 1 presents these quantiles, specifically for the standard normal distribution N(0, 1).

**Table 1.** Value in the probit function, cumulative distribution function, density function, and asymptotic standard errors of the PSNS in a standard normal distribution

| $Q_Z(p) = z_p$ | $F_Z(z_p) = p$ | $f_Z(z_p)$ | $\sigma(z_p)$ |
|---|---|---|---|
| -2 | 0.02275 | 0.05399 | $2.76168/\sqrt{n}$ |
| -4/3 | 0.09121 | 0.16401 | $1.75544/\sqrt{n}$ |
| -2/3 | 0.25249 | 0.31945 | $1.35998/\sqrt{n}$ |
| 0 | 0.5 | 0.39894 | $1.25331/\sqrt{n}$ |
| 2/3 | 0.74751 | 0.31945 | $1.35998/\sqrt{n}$ |
| 4/3 | 0.90879 | 0.16401 | $1.75544/\sqrt{n}$ |
| 2 | 0.97725 | 0.05399 | $2.76168/\sqrt{n}$ |

Note. $Q_Z(p) = \Phi^{-1}(p) = z_p$ = probit function, quantile function, or inverse function of the cumulative distribution function of a variable following a standard normal distribution $N(0, 1)$, $p$ = order or cumulative probability associated with the quantile $z_p$ in the standard normal distribution, $f_Z(z_p) = \varphi(z_p)$ = density function of the standard normal distribution, $\sigma(z_p)$ = asymptotic standard error of the sample p-th quantile calculated from a random sample of size $n$ drawn from a standard normal distribution.

The standard errors in Table 1 ($\sigma[z_p]$) allow the definition of asymptotic confidence intervals within which sample quantiles are expected to fall when computed from random samples of size $n$ drawn from a standard normal distribution, with a probability of 1 - $\alpha$ [10]. See Equation 3.

$$Z \sim N(0, 1)$$
$$P\left(\hat{z}_p \in \left[z_p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p \times (1-p)}{n \times f_Z^2(z_p)}}, z_p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p \times (1-p)}{n \times f_Z^2(z_p)}}\right]\right) = 1 - \alpha \quad (3)$$

$z_p$ = population p-th quantile of the standard normal distribution.
$\hat{z}_p$= sample quantile or point estimate of the population p-th quantile of a variable X following a standard normal distribution.
$n$ = sample size.
$f_Z^2(z_p)$ = squared height or density of the $z_p$ in a standard normal distribution.
$z_{1-\alpha/2}$ = (1-$\alpha$/2)-th quantile of a standard normal distribution, where $\alpha$ is the significance level, typically 0.05 ($z_{0.975}$ = 1.96).

From this asymptotic confidence interval, it is possible to derive a normally distributed statistic with a mean of 0 and a variance of 1, which allows testing whether a standardized sample quantile (standardized using its sample mean and variance) corresponds to the expected quantile of a standard normal distribution [5]. See Equation 4.

$$Z = \frac{\hat{z}_p - z_p}{\sqrt{\frac{p \times (1-p)}{n \times f_Z^2(z_p)}}} \sim N(0,1) \tag{4}$$

**3. Testing the Expected Value Under Normality for a Number of the PSNS**

Based on the Z statistic in Equation 4, a statistical test can be formulated to assess whether the sample p-th quantile is equivalent to the expected value under a normal distribution, and this test can then be generalized to the seven values of the PSNS.

Equation 5 shows the statistical hypotheses of the two-tailed test.

$$\begin{aligned} H_0&: Z[Q_X(p)] = \frac{Q_X(p) - \mu_X}{\sigma_X} = Q_Z(p) = z_p, where\ Z \sim N(0,1) \\ H_1&: Z[Q_X(p)] = \frac{Q_X(p) - \mu_X}{\sigma_X} \neq Q_Z(p) = z_p, where\ Z \sim N(0,1) \end{aligned} \tag{5}$$

$Z[Q_X(p)]$ = population p-th quantile of variable X, standardized by population mean and variance.
$z_p$ = hypothetical value of the p-th quantile in a standard normal distribution N(0, 1).

Assumptions: A random sample of large size drawn from a normal distribution.

Test statistic: Equation 6 shows the double standardization (first by the sample mean and standard deviation and second by the mean and expected standard deviation under the null hypothesis of normality) experienced by the sample quantile $q_X(p)$.

$$Z\left[z\left(q_X(p)\right)\right] = \frac{z[q_X(p)] - z_p}{\sqrt{\frac{p(1-p)}{n f_Z^2(z_p)}}} = \frac{\frac{q_X(p) - \bar{x}}{s_{n-1}(x)} - z_p}{\sqrt{\frac{p(1-p)}{n f_Z^2(z_p)}}} \tag{6}$$

The theoretical quantile $z_p$ is obtained by evaluating the probit function or quantile function of a standard normal distribution in $p$ (Equation 7). In the context of the PSNS, $p$ takes the following values: $F_Z(-2) = 0.0228$, $F_Z(-2/3) = 0.0912$, $F_Z(-1/3) = 0.2525$, $F_Z(0) = 0.5$, $F_Z(1/3) = 0.7475$, $F_Z(2/3) = 0.9088$, and $F_Z(2) = 0.9772$, where $Z \sim N(0, 1)$ and $F_Z$ is its cumulative distribution function, usually expressed as $\Phi(z_p) = p$.

$$Q_Z(p) = \Phi^{-1}(p) = z_p, where\ X \sim N(0,1) \tag{7}$$

To obtain the standardized sample p-th quantile, denoted by $z[c_X(p)]$, we start by calculating this quantile. To do this, the $n$ sample data of X are ordered in ascending direction: $x_{(i)}$, $i = 1, 2, \cdots, n$. For more details refer to Equation 8.

$$\begin{aligned} x = \{x_i\}_{i=1}^n = \{x_1, x_2. \ldots, x_n\} \subseteq X \\ x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \end{aligned} \tag{8}$$

The position of the quantile $q_p$ among the $n$ sample observations is calculated using R's type-8 rule, which provides greater accuracy in estimating quantiles across different distributions compared to other linear interpolation methods [4, 24-25]. Refer to Equation 9 for more details. The type-6 rule is also recommended for general use with very different distributions [26], while the type-9 rule is specifically suggested for the normal distribution [27].

$$i = 1/3 + p \times (n + 1/3) = i + (i - \lfloor i \rfloor) \tag{9}$$

If the value of $i$ is an integer, the data is searched in that order; whereas if it is a number with decimal places, the quantile is obtained by linear interpolation (Equation 10).

$$\hat{Q}_X(p) = q(p) = x_{(\lfloor i \rfloor)} + (i - \lfloor i \rfloor)\left(x_{(\lfloor i \rfloor + 1)} - x_{(\lfloor i \rfloor)}\right) \tag{10}$$

The mean (Equation 11) and sample standard deviation (Equation 12) are further calculated.

$$\hat{\mu}_X = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \tag{11}$$

$$\hat{\sigma}_X = s_{n-1}(x) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{12}$$

Finally, the sample quantile is standardized, as shown in Equation 13.

$$z[q_x(p)] = \frac{\hat{Q}_X(p) - \hat{\mu}_X}{\hat{\sigma}_X} = \frac{q_x(p) - \bar{x}}{s_{n-1}(x)} \tag{13}$$

Sampling distribution: The random variable $Z[z(c_x(p))]$, proposed as a test statistic, follows a standard normal distribution under the null assumption or hypothesis that variable X follows a normal distribution (Equation 14).

$$Z\big[z\big(q_x(p)\big)\big] \sim N(0,1) \tag{14}$$

Decision based on probability value or critical level: Let $P(Z \leq |z|)$ the probability of obtaining a value less than or equal to the absolute value of the test statistic in a standard normal distribution. If 2-tailed $p$-value = $2 \times (1 - P(Z \leq |z|) \geq \alpha$, the null hypothesis is maintained in a two-tailed test with a $\alpha$ significance level. Conversely, if 2-tailed $p$-value = $2 \times (1 - P(Z \leq |z|) < \alpha$, it is rejected.

## 4. The Omnibus Test of Normality Based on the PSNS

### 4.1. Approach

If the seven standardized values $z[z(c_x(p))]$ are squared and added together, the resulting statistic, denoted by $Q$, follows a generalized chi-square distribution, since it is the sum of squares of seven dependent random variables with a standard normal distribution [28-29]. The chi-square distribution with seven degrees of freedom is a concrete case of this generalized chi-square distribution. This equality occurs when quantiles, considered as random variables, have very small or no correlations. Moral [5] proposed using the chi-square distribution with seven degrees of freedom as an approximation to the generalized chi-square distribution and, in this way, obtain an omnibus test of normality with simple and well-known calculations of the probability, power, and effect size value [30]. Analogous to the variance formula of a quantile (Equation 1), the formula of covariance between two quantiles is a quotient, the numerator of which is the product of the order of one quantile and the complement of the order of the other quantile, and its denominator is the product of the sample size and densities of both quantiles [10]. See Equation 15. This covariance converges to zero when the sample size ($n$) tends to infinity, which is favored by a greater distance between the two quantiles. Except in the uniform distribution, the more distant two quantiles are, the lower their correlation. Also, the variance and standard deviation of each quantile tend to zero as the sample size increases. However, the correlation or quotient between the covariance and the geometric mean of the variances (Equation 16) is independent of the sample size. The $n$ in the numerator of the covariance is simplified with the two square roots of $n$ of each standard deviation in the denominator. On the other hand, this correlation is always positive, since there is no term that could be negative in the formula. Precisely, the definition of quantile or order statistic implies a monotonic non-decreasing order [31].

$$\sigma\big(q_{p_1}, q_{p_2}\big) = \frac{p_1 \times (1 - p_2)}{n \times f_Z(z_{p_1}) \times f_Z(z_{p_2})}, where\ p_1 < p_2 \tag{15}$$

$$\rho\big(q_{p_1}, q_{p_2}\big) = \frac{\sigma\big(q_{p_1}, q_{p_2}\big)}{\sqrt{\sigma^2\big(q_{p_1}\big) \times \sigma^2\big(q_{p_2}\big)}} = \frac{\sigma\big(q_{p_1}, q_{p_2}\big)}{\sigma\big(q_{p_1}\big) \times \sigma\big(q_{p_2}\big)} \tag{16}$$

The correlations between the quantiles of the PSNS of a normal distribution vary from a minimum of 0.023 to a maximum of 0.581 with a mean of 0.275, 12 out of the 21 correlations are less than 0.3, with 5 out of these 12 correlations being less than 0.1. In turn, 5 out of the 21 correlations take values between 0.3 and 0.49 and 4 between 0.5 and 0.59. The four highest correlations are between the third and fourth numbers and between the fourth and fifth numbers, as well as between the third and fifth numbers with their values contiguous towards the corresponding tail (Table 2).

**Table 2.** Matrix of correlations between PSNS quantiles of a normal distribution

| $\rho(z_{[i,]}, z_{[,j]})$ | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] |
|---|---|---|---|---|---|---|---|
| [1,] | 1 | 0.4816 | 0.2625 | 0.1526 | 0.0887 | 0.0483 | 0.0233 |
| [2,] | 0.4816 | 1 | 0.5451 | 0.3168 | 0.1841 | 0.1004 | 0.0483 |
| [3,] | 0.2625 | 0.5451 | 1 | 0.5812 | 0.3378 | 0.1841 | 0.0887 |
| [4,] | 0.1526 | 0.3168 | 0.5812 | 1 | 0.5812 | 0.3168 | 0.1526 |
| [5,] | 0.0887 | 0.1841 | 0.3378 | 0.5812 | 1 | 0.5451 | 0.2625 |
| [6,] | 0.0483 | 0.1004 | 0.1841 | 0.3168 | 0.5451 | 1 | 0.4816 |
| [7,] | 0.0233 | 0.0483 | 0.0887 | 0.1526 | 0.2625 | 0.4816 | 1 |

Note. $\rho(z_{[i,]}, z_{[,j]})$ = correlation between quantiles of a standard normal distribution: 1 = $z(p = F_Z(-2))$, 2 = $z(p = F_Z(-2/3))$, 3 = $z(p = F_Z(-1/3))$, 4 = $z(p = F_Z(0))$, 5 = $z(p = F_Z(1/3))$, 6 = $z(p = F_Z(2/3))$, and 7 = $z(p = F_Z(2))$, where $F_Z$ is the cumulative distribution function of a standard normal distribution.

The formula shown in Equation 16, which requires Equations 2 and 15 for calculation, provides the population correlation, i.e., when the sample size tends to infinity. Appendix F contains an R script for computing this correlation matrix using a bootstrap procedure.

By extracting a large sample of 1,000 data points from a standard normal distribution using a single seed (original reproducible sample), generating 10,000 bootstrap samples, computing the seven quantiles in each bootstrap sample, calculating the 21 correlations between these quantiles, and averaging the 10,000 estimates for each correlation, the resulting correlation matrix closely resembles the one obtained using the formula above (Table 3). The absolute residuals, or differences, between the corresponding correlations (asymptotic vs. bootstrap) range from 0.0003 to 0.0237, with a mean of 0.0101. The Root Mean Square Error (RMSE) is 0.0090. Therefore, both matrices can be considered equivalent.

**Table 3.** Correlation matrix obtained by bootstrap (10,000 replications) from a random sample of 1000 data extracted from a standard normal distribution.

| $r(q_{x[i,]}, q_{[,j]})$ | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] |
|---|---|---|---|---|---|---|---|
| [1,] | 1 | 0.4677 | 0.2628 | 0.1465 | 0.0933 | 0.0323 | 0.0206 |
| [2,] | 0.4677 | 1 | 0.5301 | 0.2965 | 0.1751 | 0.0913 | 0.0581 |
| [3,] | 0.2628 | 0.5301 | 1 | 0.5575 | 0.3212 | 0.1758 | 0.0934 |
| [4,] | 0.1465 | 0.2965 | 0.5575 | 1 | 0.5647 | 0.3064 | 0.1507 |
| [5,] | 0.0933 | 0.1751 | 0.3212 | 0.5647 | 1 | 0.5374 | 0.2555 |
| [6,] | 0.0323 | 0.0913 | 0.1758 | 0.3064 | 0.5374 | 1 | 0.4732 |
| [7,] | 0.0206 | 0.0581 | 0.0934 | 0.1507 | 0.2555 | 0.4732 | 1 |

Note. $q_{x[i,]}$ = i-th quantile and $c_{[,j]}$ = j-th quantile of a sample x of size 1000 randomly drawn from a standard normal distribution using a fixed seed (123) to ensure reproducibility. Quantile order (*i* for the quantile per row and *j* for the quantile per column): 1 = Φ(-2), 2 = Φ(-4/3), 3 = Φ(-2/3), 4 = Φ(0), 5 = Φ(2/3), 6 = Φ(4/3), and 7 = Φ(2), where Φ is the cumulative distribution function of a standard normal distribution, and $r(q_{x[i,]}, q_{[,j]})$ = Pearson's product-moment correlation coefficient between the sample quantiles of order *i* and order *j* (computed from 10,000 bootstrap samples using R's type-9 rule).

It should be noted that the high correlations between the terms (random variables) in the sum of squares of the *Q* statistic cause the generalized chi-square distribution to deviate from the standard chi-square distribution due to pronounced leptokurtosis. This results in a greater concentration of values around the center and a shorter right tail. This effect occurs because, when variables are correlated, they tend to move together, reducing the overall dispersion of the sum of squares. Consequently, leptokurtosis increases (more values near the mean), and positive skewness decreases, with fewer extreme values appearing in the right tail.

One of the challenges of the generalized chi-square distribution is the absence of tables for its cumulative densities and probabilities, as it does not have a standard form, unlike the chi-square distribution. If this distribution is used as an approximation to compute the p-value, the significance level could be increased to compensate for the shortening of the right tail [5].

Another, much better option is to compute the p-value, statistical power, and effect size using the bootstrap procedure [32]. This empirical method allows for the generation of the generalized chi-square

distribution underlying the test statistic, which is a sum of squares of dependent standard normal variables, thereby preserving the original formulation. A third alternative is to express the statistic as a quadratic form–specifically, the product of the transposed vector of standardized quantiles, the correlation matrix of the quantiles, and the vector of standardized quantiles–and use R's CompQuadForm package [11]. In this approach, 21 duplicate summands are incorporated, corresponding to the cross-products of the seven quantiles weighted by their correlations.

**4.2. Statistical Hypotheses**

Equation 17 presents the statistical hypotheses of the $Q$ test. They are formulated as a two-tailed test and involve a random vector consisting of seven standardized normal variables (quantiles).

$$H_0: X \sim N(\mu_X, \sigma_X) \Rightarrow Z = \frac{X-\mu_X}{\sigma_X} \sim N(0,1) \Rightarrow \begin{pmatrix} z_{Q_Z(p=0.023)} \\ z_{Q_Z(p=0.091)} \\ z_{Q_Z(p=0.252)} \\ z_{Q_Z(p=0.5)} \\ z_{Q_Z(p=0.748)} \\ z_{Q_Z(p=0.977)} \end{pmatrix} = \begin{pmatrix} -2 \\ -4/3 \\ -2/3 \\ 0 \\ 2/3 \\ 4/3 \\ 2 \end{pmatrix}$$

$$H_1: X \nsim N(\mu_X, \sigma_X) \Rightarrow Z = \frac{X-\mu_X}{\sigma_X} \nsim N(0,1) \Rightarrow \begin{pmatrix} z_{Q_Z(p=0.023)} \\ z_{Q_Z(p=0.091)} \\ z_{Q_Z(p=0.252)} \\ z_{Q_Z(p=0.5)} \\ z_{Q_Z(p=0.748)} \\ z_{Q_Z(p=0.977)} \end{pmatrix} \neq \begin{pmatrix} -2 \\ -4/3 \\ -2/3 \\ 0 \\ 2/3 \\ 4/3 \\ 2 \end{pmatrix} \quad (17)$$

**4.3. Assumptions**

It is assumed that the sample is random and consists of $n$ scores or values from a continuous quantitative variable that follows a normal distribution. A large sample size is required.

Based on a simulation study, Moral [5] recommended a minimum sample size of 150 participants when approximating the p-value using the chi-square distribution with seven degrees of freedom (original formulation). This sample size ensures high accuracy (i.e., a proportion of correct decisions—retaining or rejecting the null hypothesis of normality—of at least 0.8), as well as high statistical power for detecting non-normality and a high complement of power (i.e., the probability of correctly retaining the null hypothesis of normality when the alternative is false) for the normal distribution, with both values ($\Phi$ when $H_0$ should be retained, or $1 - \Phi$ when $H_0$ should be rejected) being ≥ 0.8. However, the test achieved perfect accuracy (i.e., 1) in detecting normality with a sample size as small as 20, although with a power below 0.2. The test's limited power to reject the null hypothesis of normality was observed in the presence of symmetric mesokurtic distributions, such as the uniform and semicircular distributions, as well as symmetric distributions with slight leptokurtosis, such as the logistic distribution. This limitation can be attributed to the high correlations among the four central values of the PSNS (Table 2).

**4.4. Test Statistics (Three Variants)**

4.4.1 Statistics $Q$ and $Q_B$

The test statistic is defined as the sum of squared standardized quantiles from the PSNS, using the expected values under the assumed distribution (standard normal), and excluding the influence of the sample mean ($\bar{x} = \sum_{i=1}^{n} x_i/n$) and standard deviation ($s_{n-1}(x) = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$). This new random variable is denoted as $Q$, in reference to the quadratic forms described in Cochran's theorem [28]. See Equation 18, where x is a random sample of $n$ data from a variable X with unknown distribution $F_X$ that is hypothesized to be normal and $Z$ is a variable defined over the domain (-∞, +∞) that follows a standard normal distribution. The R's implementation is provided in Appendix A.

$$x = \{x_i\}_{i=1}^{n} \sim F_n(x_i); \subseteq X \sim F_X(x_i); z_i \in Z \sim F_Z(z_i) \equiv N(0,1)$$

$$Q = \sum_{i=1}^{7} \left( \frac{\frac{q_x(p_i) - \bar{x}}{s_{n-1}(x)} - q_Z(p_i)}{\sigma\left(q_Z(p_i)\right)} \right)^2 = \sum_{i=1}^{7} \left( \frac{\frac{q_x(p_i) - \bar{x}}{s_{n-1}(x)} - z_{p_i}}{\sigma\left(z_{p_i}\right)} \right)^2 = \quad (18)$$

$$= \left( \frac{q_x(p_1 = 0.02275) - \bar{x}}{\frac{s_{n-1}(x)}{2.76168/\sqrt{n}}} + 2 \right)^2 + \left( \frac{q_x(p_2 = 0.09121) - \bar{x}}{\frac{s_{n-1}(x)}{1.75544/\sqrt{n}}} + \frac{4}{3} \right)^2$$
$$+ \left( \frac{q_x(p_3 = 0.25249) - \bar{x}}{\frac{s_{n-1}(x)}{1.35998/\sqrt{n}}} + \frac{2}{3} \right)^2 + \left( \frac{q_x(p_4 = 0.5) - \bar{x}}{\frac{s_{n-1}(x)}{1.25331/\sqrt{n}}} \right)^2$$
$$+ \left( z\frac{q_x(p_5 = 0.74751) - \bar{x}}{\frac{s_{n-1}(x)}{1.35998/\sqrt{n}}} - \frac{2}{3} \right)^2 + \left( \frac{q_x(p_6 = 0.90879) - \bar{x}}{\frac{s_{n-1}(x)}{1.75544/\sqrt{n}}} - \frac{4}{3} \right)^2$$
$$+ \left( \frac{q_x(p_7 = 0.97725) - \bar{x}}{\frac{s_{n-1}(x)}{2.76168/\sqrt{n}}} - 2 \right)^2$$

It should be noted that this sum of squares, derived from seven correlated Gaussian random variables with a mean of 0 and variance of 1, does not account for the standardized covariance matrix of the quadratic form. Strictly speaking, it is not a true quadratic form, as it does not assume that this matrix is an identity matrix. Therefore, the symbol $Q$ serves as a referential allusion to Cochran's theorem [28]. In the $Q_B$ variant, the distribution of the test statistic in the sampling process—used for computing the p-value (see Appendix B for the R implementation) and statistical power (see Appendix C for the R implementation)—is determined using the empirical resampling method with replacement (bootstrap), which provides a more accurate approximation than the original approach.

#### 4.4.2 $Q_T$ Statistic

Another option is to express the statistic as a quadratic form without ignoring the correlation matrix and use R's CompQuadForm package [11] to compute the probability value and power. This option is denoted $Q_T$ and is properly a quadratic form. The subscript T refers to the fact that the matrix of covariances is theoretically established from the null hypothesis of normal distribution.

A quadratic form is a linear combination involving a transposed vector (of order $1 \times k$), a symmetric matrix of order $k \times k$, and the original vector of order $k \times 1$. To obtain the quadratic form, the values $z[z(q_x(p)]$ are used as the vector, denoted by z (Equation 19). The symmetric matrix corresponds to the matrix of standardized covariances or correlations between the quantiles of the PSNS of a normal distribution. This matrix is denoted by R as the correlation matrix (Equation 20).

The quadratic form $z^T$ R z results in a second-degree polynomial with $k \times k$ terms, comprising the sum of the squared variables and the cross-products of the variables weighted by their corresponding correlations (Equation 21). If the vector and the matrix contain only numerical values, without variables, the quadratic form yields a scalar. See Appendix D for the calculation of the $Q_T$ statistic in R.

$$z^t = (z_1 \quad z_2 \quad z_3 \quad z_4 \quad z_5 \quad z_6 \quad z_7) \tag{19}$$

$$\mathrm{R} = \begin{pmatrix} 1 & 0.2819 & 0.2625 & 0.1526 & 0.0887 & 0.0826 & 0.0233 \\ 0.2819 & 1 & 0.9314 & 0.5413 & 0.3146 & 0.2930 & 0.0826 \\ 0.2625 & 0.9314 & 1 & 0.5812 & 0.3378 & 0.3146 & 0.0887 \\ 0.1526 & 0.5413 & 0.5812 & 1 & 0.5812 & 0.5413 & 0.1526 \\ 0.0887 & 0.3146 & 0.3378 & 0.5812 & 1 & 0.9314 & 0.2625 \\ 0.0826 & 0.2930 & 0.3146 & 0.5413 & 0.9314 & 1 & 0.2819 \\ 0.0233 & 0.0826 & 0.0887 & 0.1526 & 0.2625 & 0.2819 & 1 \end{pmatrix} \tag{20}$$

$$
\begin{aligned}
Q_T = z^t \times \mathrm{R} \times z = {} & z_1^2 + z_2^2 + z_3^2 + z_4^2 + z_5^2 + z_6^2 + z_7^2 + 2 \times 0.4816 \times z_1 \times z_2 \\
& + 2 \times 0.2625 \times z_1 \times z_3 + 2 \times 0.1526 \times z_1 \times z_4 + 2 \times 0.0887 \times z_1 \times z_5 \\
& + 2 \times 0.0483 \times z_1 \times z_6 + 2 \times 0.0233 \times z_1 \times z_7 + 2 \times 0.5451 \times z_2 \times z_3 \\
& + 2 \times 0.3168 \times z_2 \times z_4 + 2 \times 0.1841 \times z_2 \times z_5 + 2 \times 0.1004 \times z_2 \times z_6 \\
& + 2 \times 0.0483 \times z_2 \times z_7 + 2 \times 0.5812 \times z_3 \times z_4 + 2 \times 0.3378 \times z_3 \times z_5 \\
& + 2 \times 0.1841 \times z_3 \times z_6 + 2 \times 0.0887 \times z_3 \times z_7 + 2 \times 0.5812 \times z_4 \times z_5 \\
& + 2 \times 0.3168 \times z_4 \times z_6 + 2 \times 0.1526 \times z_4 \times z_7 + 2 \times 0.5451 \times z_5 \times z_6 \\
& + 2 \times 0.2625 \times z_5 \times z_7 + 2 \times 0.4816 \times z_6 \times z_7
\end{aligned} \tag{21}
$$

**4.5. Sampling Distribution**

The statistic or random variable $Q$ or $Q_B$ (sum of squares) follows a generalized chi-square distribution, $\chi^2_g(7, \Sigma)$, as does the quadratic form $Q_T$ (Equation 22). This distribution deviates from the chi-square distribution with seven degrees of freedom due to leptokurtosis and a shortened right tail, a consequence of the positive correlations among the seven random variables whose sum of squares defines it [21, 28-29].

The $Q$ statistic would follow a chi-square distribution if the seven variables were independent, and the $Q_T$ statistic would do so if the R matrix were an identity matrix. However, the seven quantiles are dependent because they originate from the same distribution, with an average correlation of 0.275.

If the chi-square distribution with seven degrees of freedom is used as an approximation, the significance level could be increased to 0.1 to compensate for the shortened right tail (see Appendix E for the calculation in R).

However, an approximation to the generalized chi-square distribution can be obtained via the bootstrap method. On the one hand, the empirical bootstrap distribution is generated from 10,000 replications of the original sample. On the other hand, the normative bootstrap distribution is obtained from 2,000 replications of a random sample drawn from a normal distribution whose location parameter is the arithmetic mean of the original sample and whose scale parameter is the standard deviation of the original sample, corrected using Bessel's correction. The 0.95 quantile (using R's type-8 method) of normative bootstrap distribution serves as the critical value. The bootstrap right-tailed p-value is computed, defined as the proportion of values (Q estimates) in the normative bootstrap distribution that exceeds the test statistic (from original sample). See Appendix B for the R code used in this calculation. This empirical approach provides a more accurate estimation of both the p-value (Appendix B) and the statistical power (see the R script in Appendix C) compared to the chi-square distribution with seven degrees of freedom (see the R script in Appendix E). This approach replaces the chi-square distribution with a normative bootstrap distribution that has a similar form, as it constitutes a generalized chi-square distribution. Consequently, the probability value is computed based on the right tail of the distribution.

$$
Q \sim \chi^2_g(7, \Sigma) \; and \; Q_T \sim \chi^2_g(7, \mathrm{R}) \tag{22}
$$

**4.6. Decision**

Let $P(\chi^2 g(7, \Sigma) \geq Q)$, be the probability of obtaining a value greater than or equal to the test statistic $Q$ in a generalized chi-square distribution with seven degrees of freedom. If $P(\chi^2_g(7, \Sigma) \geq Q) \geq \alpha$, the null hypothesis of normality is retained in a two-tailed test with a significance level of $\alpha$. Conversely, if $P(\chi^2_g(7, \Sigma) < \alpha$, the null hypothesis is rejected.

The same procedure applies to the quadratic form $Q_T$ and the initial approximation of the $Q$ statistic using the chi-square distribution with seven degrees of freedom [5].

**4.7. Statistical Power**

**4.7.1 With $Q_B$ Statistic**

Type II error is defined as the probability of retaining the null hypothesis given that the alternative hypothesis is true (Q > critical value). Therefore, it represents the error of failing to reject the null

hypothesis. It is denoted by β. Its complement, denoted by φ, represents the statistical power, that is the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true [33].
Using a resampling procedure with replacement from the original sample, 10,000 bootstrap samples are generated. For each bootstrap sample, the Q statistic is calculated, and the proportion of null hypothesis rejections is determined. The null hypothesis of normality is rejected if the Q statistic exceeds the critical value obtained from 2,000 samples generated under normality (normative bootstrap distribution). The rejection rate represents the bootstrap power, reflecting the test's ability to detect deviations from normality in the original sample. See Appendix E for the R implementation.

#### 4.7.2 With the $Q_T$ Statistic

The statistical power of the quadratic form $Q_T = z^T R z$ can be computed using Davies' AS 155 algorithm (1980) [34]. The initial evaluation of the generalized non-central chi-square distribution (which corresponds to the distribution of a quadratic form of correlated variables) is performed at a critical level associated with a specific alternative hypothesis. This level corresponds to the (1 - $\alpha$)th quantile of a chi-square distribution, where the degrees of freedom equal the number of eigenvalues of the correlation matrix. Since this is an iterative procedure, the critical value is adjusted at each step.
In Davies' model, the $\lambda$ parameter (degrees of freedom) is the eigenvalue vector of the correlation matrix R, while the δ parameter (non-centrality parameter) is defined as the element-wise product of the eigenvalues and the squared standardized quantiles. See Appendix D for the calculation in R.

#### 4.7.3 With the $Q$ Statistic

If we consider the approximation to the chi-square distribution with seven degrees of freedom of the original proposal, the Type II error (β) can be computed using the cumulative distribution function of a non-central chi-square distribution with seven degrees of freedom. Its non-centrality parameter is set to the value of the test statistic Q and is evaluated at the critical level corresponding to the true alternative hypothesis (see Equation 23). The statistical power (φ) is the complement of this cumulative probability (Equation 24). See Appendix E for the calculation in R.

$$\beta = \chi^2_{df=7,NCP=Q}\left(_{1-\alpha}\chi^2_{df}\right) \tag{23}$$

$$\phi = 1 - \beta = 1 - \chi^2_{df=7,NCP=Q}\left(_{0.90}\chi^2_7\right) \tag{24}$$

### 4.8. Effect Size

#### 4.8.1 With the $Q_B$ Statistic

The effect size of the deviation from normality for the $Q$ statistic can be calculated as a standardized distance, analogous to Cohen's $d$ statistic [35]. The mean of $Q$ in the normative bootstrap sampling distribution (obtained from 2,000 bootstrap samples of size $n$ drawn from a standard normal distribution) serves as the expected value of $Q$, while the sample standard deviation of $Q$ in that distribution represents the bootstrap standard error.
Using these two values, the standardization, denoted by $z$, is performed. This value can be interpreted in relation to the dispersion of a normal distribution: a $z$ value below 1.7 indicates a trivial effect size, below 2.5 a small effect, below 3 a medium effect, and above 3 a very large effect (Equation 25). See Appendix B for the calculation in R.

$$Z = \frac{\left|Q - \hat{Q}^N_{bootstrap}\right|}{se^N_{bootstrap}} = \begin{cases} < 1.7 & trivial \\ < 2.5 & small \\ < 3 & median \\ \geq 3 & large \end{cases} \tag{25}$$

#### 4.8.2 With the $Q_T$ Statistic

For the quadratic form $Q_T$, whose vector consists of correlated standard normal variables and whose symmetric matrix is the correlation matrix of these variables, the mathematical expectation of $Q_T$ is given by the trace of the correlation matrix R, i.e., $E[QT] = \text{tr}(R) = 7$. The variance of $Q_T$ is equal to twice the sum of the squared eigenvalues of R [36-37].
The seven eigenvalues of the correlation matrix of the quantiles from the PSNS of the normal distribution are: 3.281666, 1.449412, 0.8942277, 0.8293143, 0.4113026, 0.06717703, and 0.06690025.

Consequently, the standard deviation of $Q_T$ is the square root of twice the sum of the squares of these eigenvalues, resulting in $\sigma(Q_T) = 5.3918$.

By standardizing $Q_T$ using its mean and standard deviation, a measure of effect size in terms of standardized distance is obtained. This measure is analogous to Cohen's $d$ [35], though applied in the context of tests based on quadratic forms rather than mean differences. It can be interpreted relative to the dispersion of a normal distribution: a z-value below 1.7 indicates a trivial effect size, below 2.5 a small effect, below 3 a medium effect, and above 3 a very large effect (Equation 26). See Appendix D for the calculation in R.

$$Q_T = z^t \times \mathrm{R} \times z \sim \chi^2_g(7, \mathrm{R})$$

$$Z = \frac{|Q_T - E(Q_T)|}{\sigma(Q_T)} = \frac{|Q_T - tr(R)|}{\sqrt{2 \times \sum_{i=1}^{7} \lambda_i^2}} = \frac{|Q_T - 7|}{5.3918} = \begin{cases} < 1.7 & trivial \\ < 2.5 & small \\ < 3 & median \\ \geq 3 & large \end{cases} \quad (26)$$

#### 4.8.3 With $Q$ Statistic

If we consider the approximation to the chi-square distribution with seven degrees of freedom of the original proposal, the effect size of the deviation from normality in the sample distribution can be measured using Cramér's $V$ coefficient [38], which ranges from 0 to 1 [39]. Following Cohen [35], $V$ values below 0.1 indicate a trivial effect size, values between 0.1 and 0.29 indicate a small effect, between 0.3 and 0.49 a medium effect, between 0.5 and 0.69 a large effect, and values of 0.7 or higher indicate a very large effect (Equation 27). See Appendix E for the calculation in R.

$$V = \sqrt{\frac{Q}{df \times n}} = \sqrt{\frac{Q}{7 \times n}} = \begin{cases} < 0.1 & trivial \\ < 0.3 & small \\ < 0.5 & median \\ < 0.7 & large \\ \geq 0.7 & huge \end{cases} \quad (27)$$

## 5. Materials and Methods

### 5.1. Sample Generation

Samples were generated using the reverse transform procedure [40] in the R program for 31 sample sizes (ranging from 50 to 1500, with increments of 50 and 2000) and 12 continuous distributions: Laplace (mu = 50, sigma = 10), normal (mean = 50, sd = 10), Student's t-distribution with 4 degrees of freedom, linearly transformed to have a mean of 50 and a standard deviation of 10, chi-square with 5 degrees of freedom, arcsine or beta (shape1 = 0.5, shape2 = 0.5), uniform (min = 0, max = 100), triangular (min = 0, max = 3, mode = 2.99), semicircular Wigner or beta (shape1 = 1.5, shape2 = 1.5), linearly transformed to have a radius of 3, exponential (rate = 2), hyperbolic drying linearly transformed to have a mean of 50 and a standard deviation of 10, logistic (location = 50, scale = 10), and Rayleigh (sigma = 12). A seed was set for result reproduction (123, except for 12 with the normal distribution). You can visualize the samples used with the following script [41]. The result obtained from running the script in Appendix G with the first listed distribution—without a hash symbol, which is the Laplace distribution—is shown in Table 4. For presentation purposes, each one of 31 original samples (of different size) drawn from Laplace distribution in the Table 4 was reduced to the minimum, median and maximum values, with the other values being indicated by ellipses (…).

**Table 4.** Original sample drawn from a Laplace's distribution

| Sample size | Minimum | ... | Median | ... | Maximum |
|---|---|---|---|---|---|
| 50 | 9.5898829 | ... | 47.61486 | ... | 77.26236 |
| 100 | 20.4659908 | ... | 50.09903 | ... | 93.65911 |
| 150 | -6.1995893 | ... | 48.75828 | ... | 83.52246 |
| 200 | -14.2188404 | ... | 49.47437 | ... | 99.03407 |
| 250 | -2.2532879 | ... | 50.11989 | ... | 103.48888 |
| 300 | -1.2977199 | ... | 49.56351 | ... | 97.06442 |

| | | | | | |
|---|---|---|---|---|---|
| 350 | -14.1836689 | ... | 50.73552 | ... | 100.83601 |
| 400 | -40.8138988 | ... | 49.94166 | ... | 99.69524 |
| 450 | -5.7418691 | ... | 49.28067 | ... | 94.29004 |
| 500 | -9.2878793 | ... | 49.86201 | ... | 121.60666 |
| 550 | -26.2334393 | ... | 49.62307 | ... | 102.91311 |
| 600 | 0.7197058 | ... | 50.27604 | ... | 106.89202 |
| 650 | -8.4086652 | ... | 50.04799 | ... | 99.44288 |
| 700 | -35.1873205 | ... | 49.93920 | ... | 124.71132 |
| 750 | -4.3566404 | ... | 49.76694 | ... | 106.39675 |
| 800 | -27.0054577 | ... | 50.10379 | ... | 135.88705 |
| 850 | -6.2540337 | ... | 49.47890 | ... | 160.39084 |
| 900 | -23.8530300 | ... | 49.94832 | ... | 112.82285 |
| 950 | -21.2402997 | ... | 49.16473 | ... | 109.56630 |
| 1000 | -10.3551434 | ... | 50.22167 | ... | 121.36216 |
| 1050 | -10.9874128 | ... | 49.84465 | ... | 126.42730 |
| 1100 | -17.6727651 | ... | 49.13120 | ... | 112.89938 |
| 1150 | -19.3386696 | ... | 49.82527 | ... | 114.30884 |
| 1200 | -34.2327092 | ... | 50.41611 | ... | 112.19349 |
| 1250 | -11.8510114 | ... | 50.06876 | ... | 105.43839 |
| 1300 | 1.8400462 | ... | 49.71477 | ... | 119.94513 |
| 1350 | -24.1708992 | ... | 49.83466 | ... | 129.00185 |
| 1400 | -17.6022695 | ... | 50.03964 | ... | 108.28709 |
| 1450 | -34.8161317 | ... | 50.20337 | ... | 106.84666 |
| 1500 | -26.7905853 | ... | 49.88449 | ... | 111.58504 |
| 2000 | -27.9725620 | ... | 50.11953 | ... | 117.70299 |

**5.2. Statistical Analysis**

Two criteria were considered to evaluate the performance of the normality tests. One criterion was the proportion of hits, defined as the ratio between the number of correct or successful identifications and the total number of original samples drawn from each distribution (31 samples). A hit occurs when the null hypothesis of normality is retained at the 5% significance level for a sample drawn from a normal distribution with a mean of 50 and a standard deviation of 10. Conversely, a hit also occurs when the hypothesis is rejected at the 5% level for a sample drawn from a non-normal distribution. Another criterion was statistical power at the 5% significance level. For samples drawn from a non-normal distribution, power was defined as the probability of correctly rejecting the null hypothesis of normality, that is, detecting a deviation from normality when it truly exists. In the case of normal samples, its complement, the Type II error (i.e., the probability of failing to reject the null hypothesis when it is false), was assessed. A value close to 1 in this context would indicate a high probability of retaining the null hypothesis when the data are actually normal.

Since a single seed was used (123 for non-normal distributions and 12 for the normal distribution) to generate the 31-size samples across 12 distributions, the four normality tests are treated as a repeated-measures factor. Accordingly, hit rates were compared using Cochran's Q-test, with post-hoc comparisons performed using Dunn's test with Bonferroni correction [42]. The effect size in the omnibus test was measured using the eta-squared coefficient [43]. Confidence intervals for the hit rates were calculated using Wilson's procedure [44].

Statistical power means were compared using a repeated measures analysis of variance. The normality test served as the within-subjects factor with four levels. Distribution was included as a between-subjects factor with twelve independent groups, and sample size was treated as a covariate. The model included the main effect of the test as well as its interactions with sample size and distribution type.

The assumption of homogeneity of covariance matrices was assessed using Box's M test, and sphericity was evaluated using Mauchly's test. Due to a violation of the sphericity assumption, the multivariate approach based on Wilks' lambda was applied [41]. Effect size was estimated using partial eta squared ($\eta^2_p$) [43].

Normality in the distributions of mean differences was assessed using the Shapiro-Wilk W test [12-13]. In case of violation of this assumption, pairwise comparisons between the 4 levels of the repeated measures factor were performed using Student's t-test, with the p-value calculated using the bootstrap procedure (by drawing 1000 samples with replacement from the original sample), and confidence intervals were computed using the Bias-Corrected and Accelerated (BCa) percentile method. The interactions between the test and the distribution were examined by checking whether the 95% bootstrap BCa confidence intervals for the power means overlapped (not significant) or did not overlap (significant), in addition to analyzing crossovers in the power mean plot with lines separated by normality test (4 lines) [46]. The interactions between the test and the sample size were assessed using the power mean plot with lines separated by normality test, also evaluating the crossovers.

Statistical analyses were executed with R version 4.4 [41], Excel version 365 [47] and SPSS version 2021 [48].

## 6. Results

### 6.1. Comparison of Hit Ratio or Test Accuracy

When comparing the number of hits among the four normality tests using Cochran's *Q* test for the equality of binary proportions in related samples, the difference was significant ($Q_{[df = 3,\ N = 372]} = 158{,}980$, asymptotic two-tailed p-value < 0.001), with a large effect size calculated through the eta-squared coefficient ($\eta^2 = 0.142 > 0.14$). Comparisons performed using Dunn's post-hoc test revealed that five out of six differences were statistically significant, even after applying Bonferroni correction. The Shapiro-Wilk *W* test showed the highest hit rate (0.976, 95% Wilson-type *CI* [0.955, 0.987]), followed by the *Q* test based on bootstrap procedures (0.952, 95% Wilson-type *CI* [0.925, 0.969]), with their hit rates being statistically equivalent ($t = 1.137$, $p = 0.256$, $p_{BC} = 1$). Third place was for the Q test based on the chi-square distribution with seven degrees of freedom (0.849, 95% Wilson-type CI [0.810, 0.882]). The Q test based on a quadratic form using a generalized chi-square distribution for p-value calculation yielded the lowest hit rate (0.766, 95% Wilson-type *CI* [0.721, 0.806]). Refer to Table 5 and Figure 1 for further details. In the representation of the hit-and-miss proportion for each normality test in Figure 1, yellow is used for hits and red for misses.

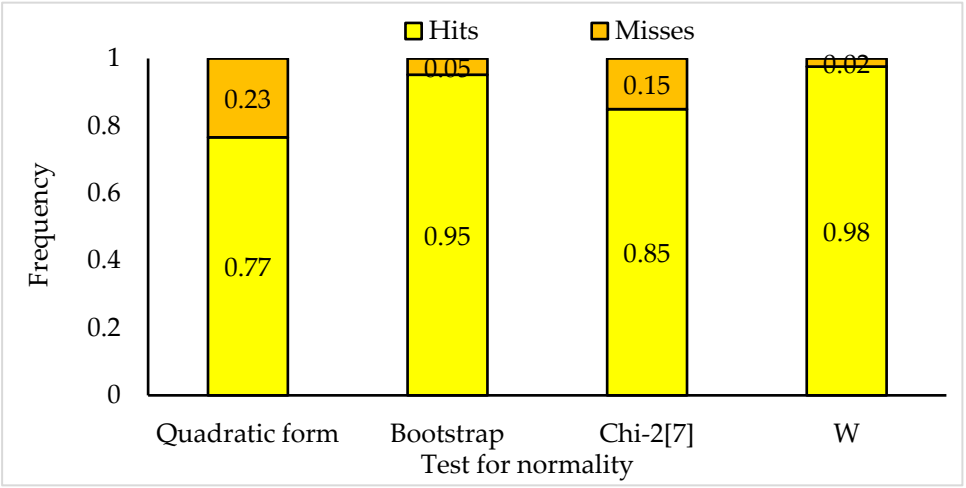



**Figure 1.** Stacked bar plot illustrates the hit-and-miss proportions for each of the four normality tests. These proportions reflect the retention of the null hypothesis of normality for samples drawn from a normal distribution, or its rejection for samples drawn from non-normal distributions, using a significance level of 0.05. Normality tests: Quadratic Form = *Q* test based on the quadratic form and using the generalized chi-square distribution to calculate the probability value; Bootstrap = *Q* test employing the bootstrap approach for this calculation; Chi-2[7] = *Q* test with approximate probability values calculated using the chi-square distribution

with seven degrees of freedom, assuming independence among random variables; and W = Shapiro-Wilk *W* test standardized using Royston's method [12-13].

**Table 5.** Pairwise comparisons of the proportions of hits using Dunn's test

| Test 1 - Test 2 | *MSD* | *se* | *t* | *p* | $p_c$ |
|---|---|---|---|---|---|
| Hits_Chi2 - Hits_Quadratic | 0.083 | 0.019 | 4.406 | < 0.001 | < 0.001 |
| Hits_Bootstrap - Hits_Quadratic | 0.188 | 0.019 | 9.949 | < 0.001 | < 0.001 |
| Hits_W - Hits_Quadratic | 0.210 | 0.019 | 11.086 | < 0.001 | < 0.001 |
| Hits_Bootstrap - Hits_Chi2 - | -0.105 | 0.019 | -5.543 | < 0.001 | < 0.001 |
| Hits_W - Hits_Chi2 | 0.126 | 0.019 | 6.680 | < 0.001 | < 0.001 |
| Hits_W - Hits_Bootstrap | 0.022 | 0.019 | 1.137 | 0.256 | 1 |

Note. *MSD* = Minimum Significant Difference. Each row tests the null hypothesis that the proportions of hits in Test 1 and Test 2 are equal. *se* = asymptotic standard error, *t* = test statistic value, *p* = asymptotic two-tailed probability value, and $p_c$ = adjusted probability value using Bonferroni correction at a significance level of 0.05. Hits corresponding to each normality test: Quadratic = *Q* test based on the quadratic form, using the generalized chi-square distribution to compute the probability value; Bootstrap = *Q* test employing the bootstrap approach for this computation; Chi2 = *Q* test using an approximate probability calculation based on the chi-square distribution with seven degrees of freedom (which assumes independence between random variables); and *W* = Shapiro-Wilk *W* test standardized using Royston's method.

When calculating the confidence intervals for the hit rates per distribution using Wilson's formula [44], the $Q_T$ test showed significantly lower proportions compared to the other three tests ($Q_B$, *Q*, and *W*) for the chi-square and uniform distributions. For the Rayleigh and logistic distributions, the hit rates of the $Q_T$ and *Q* tests were significantly lower than those of the *W* and $Q_B$ tests, whose confidence intervals overlapped. Across the full set of 12 distributions, the hit rates for the $Q_T$ and *Q* tests were significantly lower than those of the *W* and $Q_B$ tests, which were statistically equivalent (Table 6).

**Table 6.** Proportions of hits and their 95% Wilson-type confidence intervals for each normality test across different distributions

| **Distribution** | $Q_T$ | $Q_B$ | *Q* | *W* |
|---|---|---|---|---|
| Laplace | 0.871<br>[0.711, 0.949] | 0.935<br>[0.793, 0.982] | 0.903<br>[0.751, 0.967] | 0.968<br>[0.838, 0.994] |
| Normal | 1<br>[0.890, 1] | 1<br>[0.890, 1] | 1<br>[0.890, 1] | 1<br>[0.890, 1] |
| Student | 0.968<br>[0.838, 0.994] | 0.935<br>[0.793, 0.982] | 0.968<br>[0.838, 0.994] | 1<br>[0.890, 1] |
| Chi-2 | **0.484**<br>**[0.320, 0.652]** | 1<br>[0.890, 1] | 0.968<br>[0.838, 0.994] | 1<br>[0.890, 1] |
| Arcsine | 0.935<br>[0.793, 0.982] | 1<br>[0.890, 1] | 0.968<br>[0.838, 0.994] | 1<br>[0.890, 1] |
| Uniform | **0.774**<br>**[0.602, 0.886]** | 0.968<br>[0.838, 0.994] | 0.871<br>[0.711, 0.949] | 1<br>[0.890, 1] |
| Triangular | 0.839<br>[0.674, 0.929] | 0.968<br>[0.838, 0.994] | 0.871<br>[0.711, 0.949] | 1<br>[0.890, 1] |
| Semicircle | 0.839<br>[0.674, 0.929] | 0.903<br>[0.751, 0.967] | 0.871<br>[0.711, 0.949] | 1<br>[0.890, 1] |
| Exponential | 1<br>[0.890, 1] | 1<br>[0.890, 1] | 1<br>[0.890, 1] | 1<br>[0.890, 1] |
| HSD | 0.871<br>[0.711, 0.949] | 0.935<br>[0.793, 0.982] | 0.774<br>[0.602, 0.886] | 0.935<br>[0.793, 0.982] |
| Rayleigh | **0.645**<br>**[0.469, 0.789]** | 0.968<br>[0.838, 0.994] | **0.645**<br>**[0.469, 0.789]** | 0.968<br>[0.838, 0.994] |
| Logist | **0.355** | 0.839 | **0.516** | 0.871 |

| | | | | |
|---|---|---|---|---|
| | **[0.211, 0.531]** | [0.674, 0.929] | **[0.348, 0.680]** | [0.711, 0.949] |
| Total | **0.766** | 0.954 | **0.849** | 0.976 |
| | **[0.721, 0.806]** | [0.928, 0.971] | **[0.810, 0.882]** | [0.955, 0.987] |

Note. Normality tests: $Q_T$ = $Q$ test based on the quadratic form using the generalized chi-square distribution to calculate the probability value; B = $Q$ test employing the bootstrap approach for this calculation; $\chi^2[7]$ = $Q$ test with approximate probability values derived from the chi-square distribution with seven degrees of freedom, assuming independence among random variables; and $W$ = Shapiro-Wilk $W$ test standardized using Royston's method [12–13].

### 6.2. Comparison of Mean Statistical Power

A repeated-measures model was defined to compare the mean statistical power of the four normality tests. The type of distribution was included as a fixed-effects factor for independent measures, while sample size was treated as a covariate. The assumption of sphericity was not met, as indicated by the Mauchly test ($W$ = 0.900, $\chi^2[5]$ = 37.615, $p$ < 0.001), and the null hypothesis of equal covariance matrices was rejected by the Box test ($M$ of Box = 3005.680, $F$[90, 107675.456] =28.462, $p$ < 0.001). Consequently, the multivariate omnibus test of Wilks' lambda was used.

The effect of normality tests (an intragroup factor with four levels) was significant, with a large effect size. Additionally, the interaction between the normality test and sample size (a covariate with 31 values) was significant, showing a medium effect size, as was the interaction between the normality test and the type of distribution (a fixed-effect factor with 12 levels), which had a large effect size. The statistical power for these differences was unitary. For more details, refer to Table 7.

**Table 7.** Multivariate test: Wilks lambda

| **Effect** | **Λ** | ***F*** | $df_1$ | $df_2$ | ***p*** | $\eta_P^2$ | ***NCP*** | **ϕ** |
|---|---|---|---|---|---|---|---|---|
| Test | 0.710 | 48,700 | 3 | 357 | < 0.001 | 0.290 | 146.101 | 1 |
| Test*N | 0.840 | 22,679 | 3 | 357 | < 0.001 | 0.160 | 68.037 | 0.999 |
| Test*Dist | 0.407 | 11.375 | 33 | 1052.492 | < 0.001 | 0.259 | 367.578 | 1 |

Note. Intra-subject design: normality test (with four levels). Interaction of the test with N (sample size) and Distr. (type of distribution). The sample size includes 31 values (treated as a covariate), and the type of distribution encompasses 12 categories (a fixed-effects factor for independent groups). $\lambda$ = Wilks Lambda statistic, where the asterisk denotes its exact calculation; $df$ = degrees of freedom; $p$ = right-tailed probability in the Snedecor-Fisher F distribution; $\eta_P^2$ = partial eta-squared coefficient; $NCP$ = non-centrality parameter; and $\varphi$ = statistical power calculated at a significance level of 0.05.

The assumption of a normal distribution in the six power differences among the four normality tests is satisfied only with the logistic distribution in three out of six comparisons and with the normal distribution in two out of six comparisons. For the other distributions, as well as in the combined sample, the null hypothesis of normality is rejected using the Shapiro-Wilk test [12-13] at a significance level of 5% (Table 8).

**Table 8.** Testing normality for the difference in power between the four normality tests

| Distribution | Diff. | *n* | *W* | *z* | *p* | Diff. | *n* | *W* | *z* | *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| Laplace | $Q_T$-B | 31 | 0.651 | 5.036 | <0.001 | B-$\chi^2$ | 31 | 0.445 | 5.998 | <0.001 |
| Normal | | 31 | 0.919 | 2.010 | 0.022 | | 31 | 0.928 | 1.766 | 0.039 |
| Student | | 31 | 0.514 | 5.723 | <0.001 | | 31 | 0.472 | 5.894 | <0.001 |
| Chi-2 | | 31 | 0.789 | 3.994 | <0.001 | | 31 | 0.326 | 6.400 | <0.001 |
| Arcsine | | 31 | 0.269 | 6.568 | <0.001 | | 31 | 0.269 | 6.568 | <0.001 |
| Uniform | | 31 | 0.451 | 5.975 | <0.001 | | 31 | 0.533 | 5.640 | <0.001 |
| Triangular | | 31 | 0.616 | 5.234 | <0.001 | | 31 | 0.489 | 5.826 | <0.001 |
| Semicircle | | 31 | 0.54 | 5.609 | <0.001 | | 31 | 0.787 | 4.013 | <0.001 |
| Exponential | | 31 | 0.176 | 6.816 | <0.001 | | 31 | 0.192 | 6.776 | <0.001 |
| Hyperbolic | | 31 | 0.69 | 4.791 | <0.001 | | 31 | 0.643 | 5.083 | <0.001 |
| Rayleigh | | 31 | 0.522 | 5.688 | <0.001 | | 31 | 0.723 | 4.558 | <0.001 |
| Logistic | | 31 | 0.965 | 0.272 | 0.389 | | 31 | 0.945 | 1.208 | 0.116 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Total | | 372 | 0.683 | 10.444 | <0.001 | | 372 | 0.704 | 10.281 | <0.001 |
| Laplace | $Q_T$-$\chi^2$ | 31 | 0.676 | 4.882 | <0.001 | B-W | 31 | 0.4 | 6.159 | <0.001 |
| Normal | | 31 | 0.97 | -0.048 | 0.513 | | 31 | 0.945 | 1.208 | 0.116 |
| Student | | 31 | 0.334 | 6.375 | <0.001 | | 31 | 0.44 | 6.016 | <0.001 |
| Chi-2 | | 31 | 0.829 | 3.558 | <0.001 | | 31 | 0.263 | 6.585 | <0.001 |
| Arcsine | | 31 | 0.269 | 6.568 | <0.001 | | 31 | 0.202 | 6.750 | <0.001 |
| Uniform | | 31 | 0.566 | 5.488 | <0.001 | | 31 | 0.326 | 6.400 | <0.001 |
| Triangular | | 31 | 0.652 | 5.030 | <0.001 | | 31 | 0.414 | 6.110 | <0.001 |
| Semicircle | | 31 | 0.694 | 4.764 | <0.001 | | 31 | 0.375 | 6.244 | <0.001 |
| Exponential | | 31 | 0.282 | 6.531 | <0.001 | | 31 | 0.179 | 6.809 | <0.001 |
| Hyperbolic | | 31 | 0.65 | 5.042 | <0.001 | | 31 | 0.644 | 5.078 | <0.001 |
| Rayleigh | | 31 | 0.717 | 4.602 | <0.001 | | 31 | 0.54 | 5.609 | <0.001 |
| Logistic | | 31 | 0.875 | 2.909 | 0.002 | | 31 | 0.829 | 3.558 | <0.001 |
| Total | | 372 | 0.667 | 10.56 | <0.001 | | 372 | 0.549 | 11.280 | <0.001 |
| Laplace | $Q_T$-W | 31 | 0.639 | 5.106 | <0.001 | $\chi^2$-W | 31 | 0.435 | 6.035 | <0.001 |
| Normal | | 31 | 0.738 | 4.442 | <0.001 | | 31 | 0.755 | 4.303 | <0.001 |
| Student | | 31 | 0.356 | 6.306 | <0.001 | | 31 | 0.313 | 6.440 | <0.001 |
| Chi-2 | | 31 | 0.797 | 3.914 | <0.001 | | 31 | 0.332 | 6.382 | <0.001 |
| Arcsine | | 31 | 0.282 | 6.531 | <0.001 | | 31 | 0.276 | 6.548 | <0.001 |
| Uniform | | 31 | 0.456 | 5.956 | <0.001 | | 31 | 0.527 | 5.666 | <0.001 |
| Triangular | | 31 | 0.616 | 5.234 | <0.001 | | 31 | 0.487 | 5.835 | <0.001 |
| Semicircle | | 31 | 0.559 | 5.521 | <0.001 | | 31 | 0.721 | 4.573 | <0.001 |
| Exponential | | 31 | 0.331 | 6.385 | <0.001 | | 31 | 0.277 | 6.545 | <0.001 |
| Hyperbolic | | 31 | 0.694 | 4.764 | <0.001 | | 31 | 0.686 | 4.817 | <0.001 |
| Rayleigh | | 31 | 0.468 | 5.910 | <0.001 | | 31 | 0.702 | 4.709 | <0.001 |
| Logistic | | 31 | 0.927 | 1.795 | 0.035 | | 31 | 0.956 | 0.746 | 0.232 |
| Total | | 372 | 0.691 | 10.383 | <0.001 | | 372 | 0.651 | 10.672 | <0.001 |

Note. Diff = random variable of the power difference between tests of normality: $Q_T$ = $Q$ test from quadratic form using the generalized chi-square distribution for the calculation of the probability value, B = $Q$ test from its bootstrap approach and $\chi^2_{[7]}$ = $Q$ test from the chi-square approximation with seven degrees of freedom, and $W$ = Shapiro-Wilk test with Royston's standardization. Test to check the normality of power differences: $W$ = Shapiro-Wilk test statistic, $n$ = sample size, $z$ = standardized value of the logarithmic transformation of $W$ (Royston's standardization), and $p$ = probability to the right tail in a standard normal distribution.

When pairwise comparisons were conducted using the paired-samples Student's t-test with a bootstrap procedure for calculating p-values and obtaining BCa percentile confidence intervals (1,000 replications), the mean power of the W test was significantly higher than that of the other three tests. The statistical power of the bootstrap version of the Q test was significantly higher than that of the other two versions, which did not differ significantly from each other. See Table 9, which presents the group means, the mean differences with their corresponding 95% BCa confidence intervals, and the two-tailed bootstrap p-values.

**Table 9.** Comparison of mean statistical power using Student's t-test for two matched samples, with the BCa method applied for calculating the bootstrap confidence interval.

| Tets | $m_1$ | $m_2$ | $md$ | $p\_boot$ |
|---|---|---|---|---|
| $Q_T$-B | 0.902 | 0.935 | -0.033 | 0.001 |
| | [0.882, 0.922] | [0.920, 0.950][b] | [-0.049, -0.018][b] | |
| $Q_T$-$\chi^2_{[7]}$ | | 0.902 | 0 | 0.983 |
| | | [0.881, 0.920][b] | [-0.014, 0.014][b] | |
| $Q_T$-W | | 0.961 | -0.059 | 0.001 |
| | | [0.946, 0.975][b] | [-0.075, -0.046][b] | |
| B-$\chi^2_{[7]}$ | 0.935 | 0.902 | 0.034 | 0.001 |

| | | | | |
|---|---|---|---|---|
| | [0.920, 0.950]ᵇ | [0.881, 0.920]ᵇ | [0.020, 0.049]ᵇ | |
| B-W | | 0.961<br>[0.946, 0.975]ᵇ | -0.040<br>[-0.055, -0.026]ᵇ | 0.001 |
| $\chi^2$-W | 0.902<br>[0.881, 0.920]ᵇ | 0.961<br>[0.946, 0.975]ᵇ | -0.060<br>[-0.074, -0.046]ᵇ | 0.001 |

Note. Statistical tests of normality: $Q_T$ = *Q* test based on the quadratic form, using the generalized chi-square distribution to compute the p-value; B = *Q* test employing the bootstrap approach; $\chi^2[7]$ = *Q* test using the chi-square approximation with seven degrees of freedom; and *W* = Shapiro-Wilk test standardized using Royston's method. $m_1$ = mean power of the test serving as the minuend, $m_2$ = mean power of the test serving as the subtrahend, *md* = difference between the two means (mean difference) with a 95% bootstrap confidence interval computed using the BCa method, and p_boot = bootstrap probability for the null hypothesis of mean equality.

Figure 2 displays the average power values for the non-normal distributions and the average complement of power (beta error) for the normal distribution for the four normality tests. The curves are represented in different colors: the Shapiro–Wilk test is shown in yellow, the Q test (quadratic version) in blue, the bootstrap version of the Q test in red, and the version of the Q test that uses an approximate p-value based on the chi-square distribution with seven degrees of freedom in black. Statistical power increased with sample size. The W test showed the fastest improvement as sample size increased, exhibiting higher average power than the other tests. The bootstrap version of the Q test outperformed the other two Q test versions from the smallest sample size. At a sample size of 1,000, all four tests reached unit power.

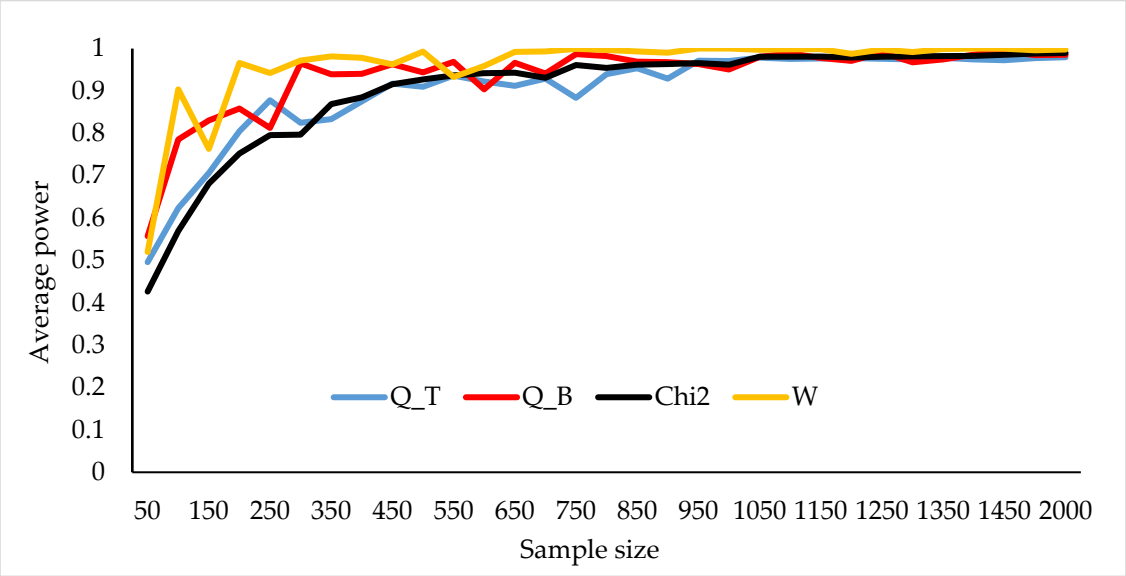



**Figure 2.** Diagram of the marginal means of power by normality test in relation to sample size. Tests: Q_T = *Q* test based on the quadratic form, using the generalized chi-square distribution to calculate the probability value; B = *Q* test employing the bootstrap approach; Chi2 = *Q* test with approximate probability values derived from the chi-square distribution with seven degrees of freedom (assuming independence among random variables); and W = Shapiro-Wilk test standardized using Royston's method [12-13].

The interaction effect between the distribution and the normality test was assessed using the mean diagram (Figure 3) and the overlap of the 95% bootstrap confidence intervals obtained via the BCa percentile method. If no overlap was observed, the difference was considered significant in a two-tailed test at a 5% significance level (Table 10).

**Table 10.** Mean power in the interaction between the distribution and the test, with 95% bootstrap confidence intervals obtained via the BCa method

| **Distr.** | $Q_T$ | **Bootstrap** | $\chi^2[7]$ | *W* |
|---|---|---|---|---|
| **Laplace** | 0.909 [0.834, 0.969]ᵇ | 0.950 [0.871, 0.998]ᵇ | 0.935 [0.863, 0.991]ᵇ | 0.967 [0.895, 1]ᵇ |
| **Normal** | 0.830 [0.812, 0.845]ᵇ | **0.724 [0.667, 0.776]ᵇ** | 0.853 [0.836, 0.870]ᵇ | 0.947 [0.910, 0.973]ᵇ |

| | | | | |
|---|---|---|---|---|
| **Student** | 0.979 [0.927, 1]ᵇ | 0.946 [0.891, 0.993]ᵇ | 0.985 [0.942, 1]ᵇ | 0.989 [0.963, 1]ᵇ |
| **Chi-2** | **0.806 [0.708, 0.890]ᵇ** | 0.997 [0.988, 1]ᵇ | 0.986 [0.963, 1]ᵇ | 0.995 [0.983, 1]ᵇ |
| **Arcsine** | 0.992 [0.976, 1]ᵇ | 0.999 [0.995, 1]ᵇ | 0.981 [0.943, 1]ᵇ | 0.991 [0.991, 1]ᵇ |
| **Uniforme** | 0.964 [0.915, 0.996]ᵇ | 0.982 [0.938, 1]ᵇ | 0.941 [0.874, 0.989]ᵇ | 0.955 [0.956, 1]ᵇ |
| **Triangular** | 0.873 [0.778, 0.956]ᵇ | 0.988 [0.956, 1]ᵇ | 0.944 [0.882, 0.993]ᵇ | 0.985 [0.959, 1]ᵇ |
| **Semicircular** | 0.908 [0.821, 0.974]ᵇ | 0.935 [0.857, 0.993]ᵇ | 0.830 [0.728, 0.916]ᵇ | 0.954 [0.882, 0.998]ᵇ |
| **Exponential** | 1 [0.999, 1]ᵇ | 1 [0.999, 1]ᵇ | 1 [0.999, 1]ᵇ | 1 [0.999, 1]ᵇ |
| **Secant** | 0.909 [0.830, 0.971]ᵇ | 0.924 [0.846, 0.982]ᵇ | 0.866 [0.768, 0.949]ᵇ | 0.934 [0.830, 0.996]ᵇ |
| **Rayleigh** | 0.959 [0.906, 0.992]ᵇ | 0.951 [0.889, 0.996]ᵇ | 0.795 [0.678, 0.898]ᵇ | 0.948 [0.859, 0.999]ᵇ |
| **Logistic** | 0.693 [0.611, 0.774]ᵇ | 0.830 [0.739, 0.909]ᵇ | 0.705 [0.608, 0.794]ᵇ | 0.836 [0.719, 0.927]ᵇ |

Note. Normality test: $Q_T$ = $Q$ test from the quadratic form using the generalized chi-square distribution for the calculation of the probability value, Bootstrap = $Q$ test from the bootstrap approach for this calculation, $\chi^2[7]$ = $Q$ test with the approximate probability calculation from the chi-square distribution with seven degrees of freedom (which implies assuming independence between random variables), and $W$ = Shapiro-Wilk $W$ test with Royston's standardization.

The mean diagram in Figure 3 shows that the highest average power is achieved by the Shapiro–Wilk W test across all distributions, except for the Rayleigh distribution, where the maximum occurs with the quadratic form of the Q test ($Q_T$). For the normal distribution, the average power of the bootstrap version of the Q test ($Q_B$) was significantly lower than that of the other three tests (Table 10). Additionally, for the chi-square distribution, the average power of the quadratic form of the Q test was significantly lower ($Q_T$) than that of the other three tests (Table 10). For the remaining distributions, no significant differences in average power were found among the four normality tests (Table 10).

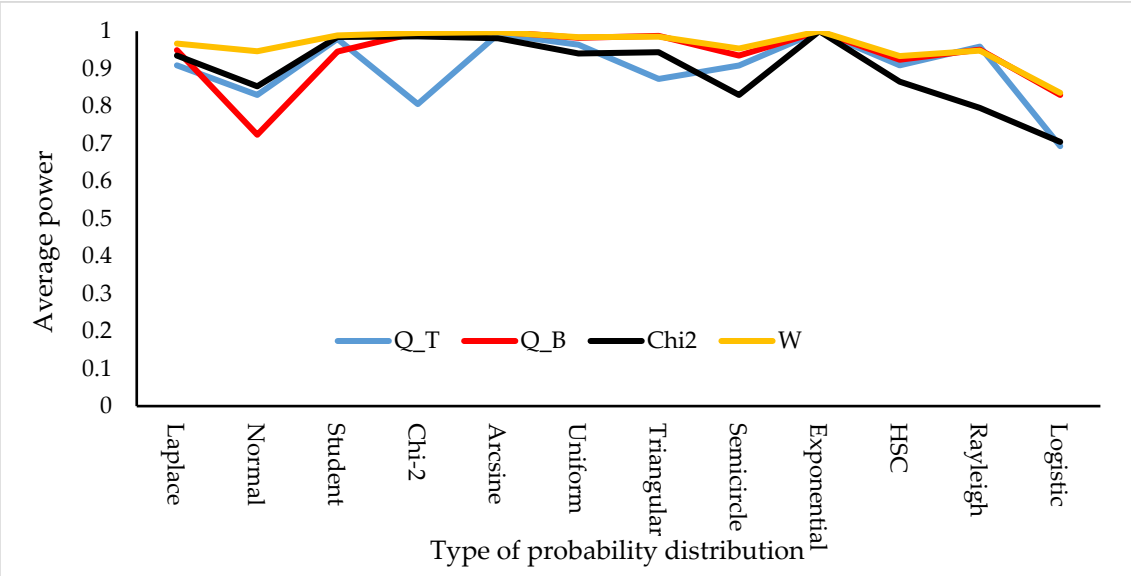



**Figure 3.** Diagram of the marginal mean power of the normality tests in relation to the type of distribution. Normality test: Q_T = Q test based on the quadratic form, using the generalized chi-square distribution to compute the probability value; B = Q test employing the bootstrap approach for this computation; Chi2 = Q test using an approximate probability calculation based on the chi-square distribution with seven degrees of freedom (which assumes independence between random variables); and W = Shapiro-Wilk W test standardized using Royston's method.

### 6.3. Application of R scripts to real-world data

Alexithymia was measured using the 20-item Toronto Alexithymia Scale (TAS-20) [18], adapted for the Mexican population by Moral [19]. The scale consists of 20 Likert-type items with six response options (0 to 5), yielding a total score ranging from 0 to 100. The sample consisted of 385 participants (85 men and 300 women) who took the entrance examination for a psychology program at a public university in northeastern Mexico. Below is the random vector of TAS-20 total scores.

x <- c(34, 14, 13, 17, 44, 4, 6, 33, 22, 28, 14, 23, 13, 12, 32, 12, 7, 48, 55, 45, 24, 13, 61, 23, 55, 35, 11, 13, 29, 30, 10, 29, 29, 26, 14, 41, 11, 17, 21, 26, 10, 38, 24, 20, 50, 26, 27, 27, 25, 17, 16, 27, 20, 28, 35, 28, 27, 40, 10, 12, 21, 19, 23, 40, 22, 4, 8, 26, 14, 35, 47, 24, 21, 31, 33, 23, 0, 33, 36, 31, 24, 25, 25, 4, 43, 21, 41, 26, 27, 21, 21, 32, 12, 22, 18, 30, 36, 26, 9, 38, 18, 35, 31, 41, 26, 15, 17, 31, 24, 44, 56, 16, 52, 30, 24, 25, 52, 44, 4, 19, 34, 22, 28, 31, 3, 12, 0, 34, 10, 44, 18, 17, 7, 48, 20, 17, 17, 42, 9, 22, 18, 58, 49, 32, 24, 16, 22, 9, 25, 28, 27, 46, 30, 10, 15, 29, 22, 18, 16, 13, 34, 13, 12, 14, 25, 33, 52, 5, 35, 20, 16, 28, 17, 20, 5, 25, 35, 9, 35, 45, 16, 26, 32, 27, 20, 27, 10, 23, 24, 20, 43, 28, 15, 21, 26, 32, 22, 51, 11, 27, 39, 21, 26, 18, 20, 19, 26, 14, 29, 25, 29, 23, 6, 19, 31, 5, 20, 33, 12, 32, 14, 24, 45, 39, 21, 20, 33, 18, 23, 15, 6, 25, 8, 54, 36, 16, 12, 28, 25, 17, 50, 7, 31, 28, 20, 20, 20, 27, 29, 32, 34, 41, 29, 31, 18, 4, 19, 30, 31, 5, 12, 35, 34, 17, 39, 14, 23, 18, 21, 22, 31, 36, 19, 50, 23, 10, 31, 36, 29, 21, 13, 28, 43, 17, 24, 43, 20, 17, 60, 39, 22, 30, 19, 37, 25, 21, 30, 9, 39, 21, 10, 17, 45, 33, 32, 33, 14, 31, 33, 16, 24, 19, 8, 4, 27, 21, 58, 38, 40, 14, 21, 14, 2, 29, 18, 42, 25, 28, 49, 37, 24, 34, 33, 30, 23, 10, 30, 46, 12, 38, 11, 32, 27, 37, 7, 25, 24, 12, 22, 10, 5, 24, 24, 20, 19, 16, 29, 28, 18, 19, 28, 17, 13, 47, 27, 15, 13, 25, 35, 6, 19, 45, 36, 17, 32, 25, 5, 30, 37, 41, 46, 15, 14, 13, 32)

The script in Appendix A yields the following results. First, it produces a density histogram of the empirical distribution with the empirical and normal density curves overlaid (Figure 4). The histogram was constructed following the Freedman-Diaconis rule [49], and the density was estimated using Epanechnikov's kernel [50] and the Sheather-Jones bandwidth [51]. The histogram profile (12 yellow-filled bins) and its corresponding density curve (dark blue line) resemble the normal curve (red line), although a slight positive (right-tailed) skewness is noticeable (Figure 4). Second the script provides the sample quantiles corresponding to the PSNS (Table 11), along with the statistics needed to standardize these quantiles (sample size, mean, and standard deviation), and a table showing the calculations used to calculate the Q-statistic, which is obtained by summing the values in the last column of this table and equals 13.7043 (Table 12).

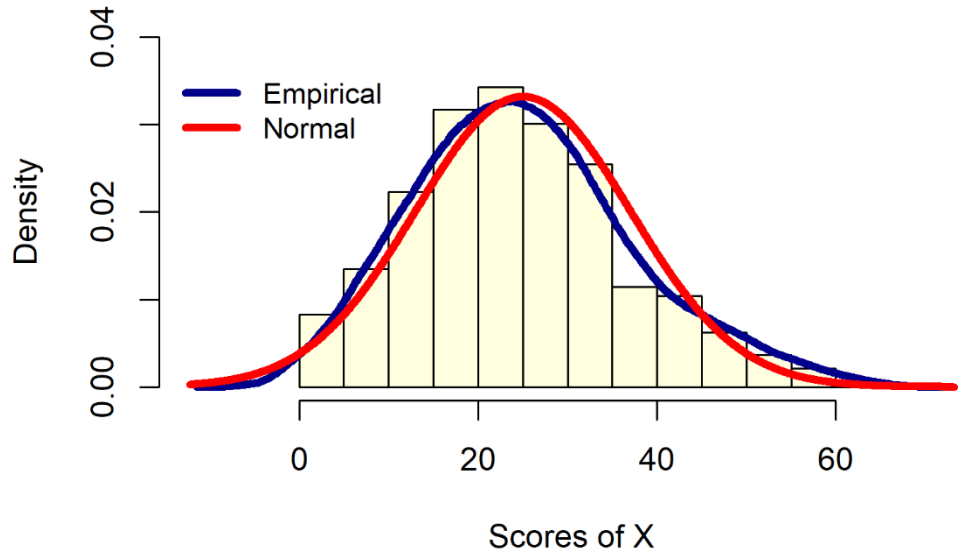



**Figure 4.** Density histogram of the empirical distribution (using the Freedman-Diaconis rule) with the empirical density curve (based on Epanechnikov's kernel and the Sheather–Jones bandwidth) and the normal density curve overlaid.

**Table 11.** Sample quantiles corresponding to Parametric Seven-Number Summary (PSNS)

| 2.275013% | 9.121122% | 25.24925% | 50% | 74.75075% | 90.87888% | 97.72499% |
|---|---|---|---|---|---|---|
| 4 | 10 | 17 | 24 | 32 | 43 | 52 |

Statistics needed to standardize sample quantiles corresponding to the PSNS:
Sample size: 385.
Sample mean: 25.01818.
Sample standard deviation: 12.01441.

**Table 12.** Testing normality using the test based on the Parametric Seven-Number Summary

| e_zp | p | d_zp | ee_zp | x_p | z_xp | z | z_sq |
|---|---|---|---|---|---|---|---|
| -2 | 0.0228 | 0.054 | 0.1407 | 4 | -1.7494 | 1.7804 | 3.1698 |
| -1.3333 | 0.0912 | 0.164 | 0.0895 | 10 | -1.25 | 0.9313 | 0.8673 |
| -0.6667 | 0.2525 | 0.3194 | 0.0693 | 17 | -0.6674 | -0.0103 | 1e-04 |
| 0 | 0.5 | 0.3989 | 0.0639 | 24 | -0.0847 | -1.3268 | 1.7603 |
| 0.6667 | 0.7475 | 0.3194 | 0.0693 | 32 | 0.5811 | -1.2342 | 1.5233 |
| 1.3333 | 0.9088 | 0.164 | 0.0895 | 43 | 1.4967 | 1.8259 | 3.3339 |
| 2 | 0.9772 | 0.054 | 0.1407 | 52 | 2.2458 | 1.7463 | 3.0495 |
| Sum | | | | | | | 13.7043 |

Note: e_zp = expected value for the p-th quantile in a N(0, 1) distribution, p = quantile order, d_zp = density of the p-th quantile in a N(0, 1) distribution, ee_zp = standard error of the p-th quantile for n normally distributed data points, x_p = p-th sample quantile, z_xp = standardized quantile (calculated using the sample mean and standard deviation), z = (z_xp - e_zp) / ee_zp = standardization of the z_xp statistic under the assumption of normality, z^2 = squared z_xp statistic.

Value of test statistic in original sample 13.70427.

The script in Appendix B applies bootstrap procedures to compute the critical value for a 5% significance level. This value is obtained from the normative bootstrap distribution (Figure 5) as its 0.95 quantile, using R's Type 8 quantile estimator (Q_critical = 6.43475). The script also calculates the p-value, defined as the proportion of values in the normative bootstrap distribution that are greater than the observed test statistic, based on the 2000 Q estimates that constitute the distribution (Figure 6). In addition, it computes the effect size as the standardized distance between the test statistic and its expected value (in the normative bootstrap distribution), using its standard deviation for scaling. The null hypothesis of normality is rejected (Q statistic= 13.70427 > Q_critical value = 6.43475; p-value = 0.0005 < $\alpha$ = 0.05), with a large effect size: Z = (13.70427 - 2.961112) / 1.902137 = 5.6479 > 3.

Bootstrap sampling distribution of the Q statistic generated from the original sample (Figure 5).

Number of extractions with replacement = 2000.

Bootstrap estimation of the Q = 17.59577.

Bootstrap standard error of the Q = 7.13478.

Bootstrap bias of the Q = 3.891502.

Normative bootstrap sampling distribution of the Q statistic (Figure 6).

This is generated from a normal distribution with mean 25.01818 and standard deviation 12.01441

Based on 2000 extractions of size 385.

Bootstrap expected value of Q under normal = 2.961112.

Bootstrap standard error of Q under normal = 1.902137.

Bootstrap critical value for the Q statistic with a significance level of 0.05 = 6.43475.

Bootstrap p-value = 0.0005.

The null hypothesis that the data follow a normal distribution is rejected with a significance level of 0.05 based on the bootstrap p-value.

Effect size as standardized distance: Z_Q = 5.647943.

Interpretation of effect size

< 1.7: Trivial

[1.7, 2.5): Small

[2.5, 3): Medium

$\geqslant$ 3: Large

Based on the Z statistic = 5.6479, the size effect of the deviation from the normal distribution model is classified as large.

The normative bootstrap distribution of the Q statistic (derived from normally distributed samples) is represented in Figure 5 by a histogram and a density curve, along with the chi-square distribution with 7 degrees of freedom (red curve). The vertical dotted lines indicate the test statistic (black) and the

bootstrap critical value (purple). The normative bootstrap distribution appears more peaked and has a longer right tail than the chi-square distribution with 7 degrees of freedom.

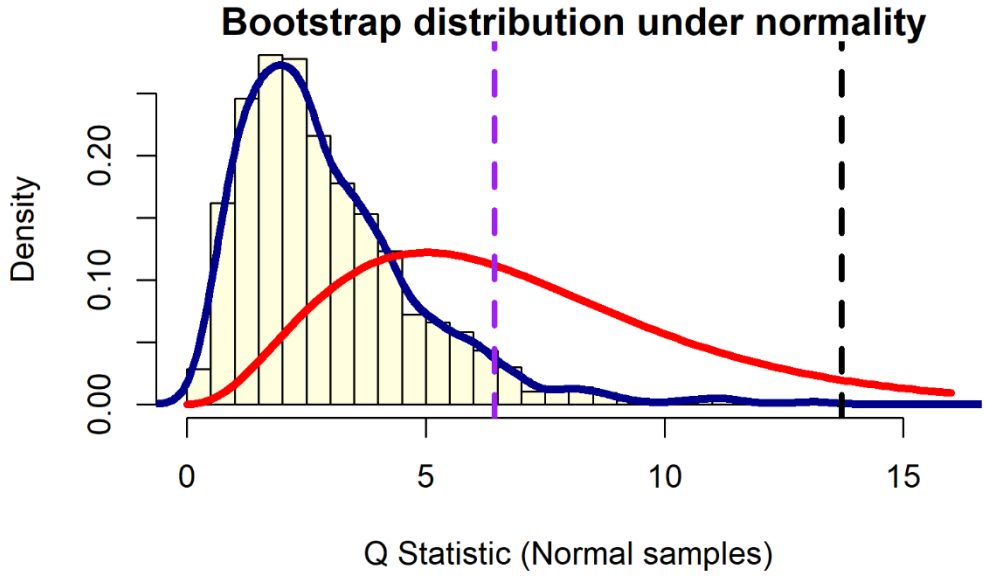



**Figure 5.** Histogram of normative bootstrap distribution (under null hypothesis of normality) of Q-Statistic with empirical and chi-2[7] density curves overlaid. The critical value (from the normative bootstrap distribution of the Q-statistic) is represented by a purple dotted line, while the test statistic value is shown with a black dotted line.

The empirical bootstrap distribution of the Q statistic (derived from the original sample) is shown in Figure 6 as a histogram and a density curve, together with the chi-square distribution with 7 degrees of freedom (red curve). The vertical dotted lines represent the test statistic (black) and the bootstrap critical value (purple). In contrast to what is observed in Figure 5, the chi-square distribution appears more peaked and exhibits a longer right tail than the empirical bootstrap distribution.

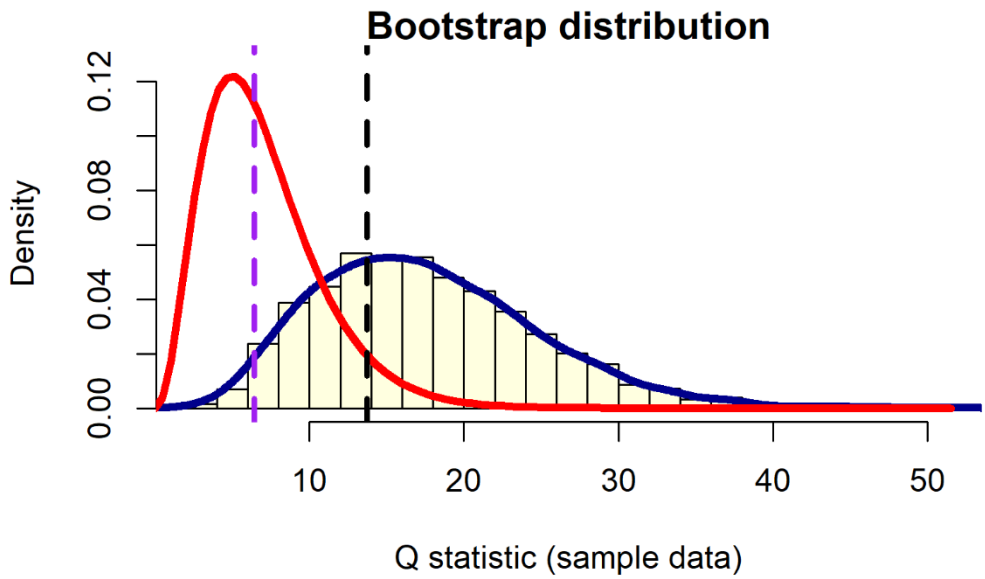



**Figure 6.** Histogram of bootstrap distribution (from original sample) of Q-Statistic with empirical and chi-2[7] density curves overlaid. The critical value (from the normative bootstrap distribution of the Q-statistic) is represented by a purple dotted line, while the test statistic value is shown with a black dotted line.

Appendix C script using 10,000 bootstrap samples (generated from original sample) computes statistical power or proportion of rejections across simulations. When you run the script, you get a very high power value ($\phi = 1 > 0.90$).

Number of bootstrap samples = 10000.

Number of data per bootstrap sample = 385.

Bootstrap power at a significance level of 0.05 = 1.
The script in Appendix D yields the quadratic form of the Q statistic. First, it presents the Gaussian vector with quantiles standardized under the null hypothesis of normality. Second, it displays the matrix of correlations between quantiles under the null hypothesis (Table 13). Third, it computes the scalar resulting from the quadratic form—namely, the row Gaussian vector multiplied by the square correlation matrix, and then by the column Gaussian vector—which gives the value of the test statistic (Q_T = 13.08073). The p-value is calculated using a generalized chi-square distribution and is greater than the significance level, so the null hypothesis of normality is retained (p-value = 0.9103499 > $\alpha$ = 0.05). Finally, the effect size is computed as the standardized distance between the test statistic and its expected value, which turns out to be trivial: Z_Q = (13.08073 − 7) / 4.829811 = 1.259 < 1.7.
Gaussian vector z:
z <- (1.780379, 0.9313009, -0.01029847, -1.326763, -1.23424, 1.825896, 1.746296)

**Table 13**. Correlation matrix of PSNS quantiles (under the assumption of normality)

| Quantile | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] |
|---|---|---|---|---|---|---|---|
| [1,] | 1 | 0.4816 | 0.2625 | 0.1526 | 0.0887 | 0.0483 | 0.0233 |
| [2,] | 0.4816 | 1 | 0.5451 | 0.3168 | 0.1841 | 0.1004 | 0.0483 |
| [3,] | 0.2625 | 0.5451 | 1 | 0.5812 | 0.3378 | 0.1841 | 0.0887 |
| [4,] | 0.1526 | 0.3168 | 0.5812 | 1 | 0.5812 | 0.3168 | 0.1526 |
| [5,] | 0.0887 | 0.1841 | 0.3378 | 0.5812 | 1 | 0.5451 | 0.2625 |
| [6,] | 0.0483 | 0.1004 | 0.1841 | 0.3168 | 0.5451 | 1 | 0.4816 |
| [7,] | 0.0233 | 0.0483 | 0.0887 | 0.1526 | 0.2625 | 0.4816 | 1 |

The Q statistic value in the quadratic form = 13.08073.
Eigenvalues: 2.747683, 1.535561, 0.9552459, 0.6382596, 0.4665808, 0.3647546, and 0.2919151.
Number of eigenvalues: 7.
Sum of eigenvalues: 7.
p-value for the Q_T statistic = 0.1019042.
Statistical power: 0.9103499.
Expected value of the Q_T statistic = 7.
Standard deviation of the Q_T statistic = 4.829811.
Effect size as standardized distance: Z_Q = 1.259.
Interpretation of effect size.
< 1.7: Trivial
[1.7, 2.5): Small
[2.5, 3): Medium
$\geqslant$ 3: Large
Based on Z statistic = 1.259, the effect size of the deviation from the normal distribution model is classified as trivial.
When you run the Appendix E script to calculate the p-value, statistical power, and effect size of the Q statistic using the chi-square approximation with 7 degrees of freedom, the result resembles the quadratic form of the Q statistic. This aligns with the hypothesized distribution of the TAS-20 total score but differs from the result obtained using the bootstrap version of the Q statistic, which does not support this hypothesis.
Test Statistic: Q = 13.70427, critical value with a significance level of 0.05 = 14.06714 , asymptotic p-value = 0.05669822.
The null hypothesis that the data follow a normal distribution is not rejected with a significance level of 0.05 by the Q test.
The statistical power to the right tail for the alternative hypothesis of non-normality for the Q test: $\phi$ = 0.2222.

If the null hypothesis of normality is maintained, the power must be less than 0.5; while, if rejected, it must be greater than 0.5. In the latter case, it is classified as good with a value of 0.8 and very good with a value of 0.9.
Effect size via Cramer's V-coefficient: V = 0.0713.
Interpretation of effect size based on Cohen (1988) [35]:
< 0.1: Trivial
[0.1, 0.3): Small
[0.3, 0.5): Medium
$\geqslant$ 0.5: Large
Based on the V statistic = 0.0713, the size of the effect of the deviation from the normal distribution model is classified as trivial.
The Shapiro–Wilk test, using Royston's procedure (Appendix H), confirms the result obtained with the bootstrap version of the Q statistic and the histogram of deviations from normality. In contrast, the quadratic form and chi-square versions of the Q statistic support the null hypothesis of normality. Consequently, this example validates the results of the analysis with simulated data, indicating that the bootstrap version is the most appropriate option for the PSNS Q test.
Shapiro-Wilk W-statistic: w = 0.9819425.
Logarithmic transformation of W-statistic: ln(1 - w) = -4.014193.
Expected value of ln(1 - w) = -5.583649.
Standard deviation of ln(1 - w) = 0.4210108.
Royston´s standardized W-statistic: z_w = 3.727829.
Right-tailed p-value = 9.656814e-05.
The null hypothesis that the data follow a normal distribution is rejected with a significance level of 0.05 by the Shapiro-Wilk W-test using Royston's procedure.
Statistical power: $\phi$ = 0.9813733.
It should be noted that both the Kolmogorov-Smirnov-Lilliefors test [52] (D = 0.051253, p-value = 0.01674) and the Anderson-Darling test [53] ($A^2$ = 1.4911, p-value = 0.000751) reject the null hypothesis of normality at the 5% significance level. The Anscombe-Glynn kurtosis test [54] indicates mesokurtosis (kurt = 2.99577, z = 0.16173, p-value = 0.8715), whereas the D'Agostino asymmetry test [55] reveals right-tail asymmetry (skew = 0.46965, z = 3.65637, p-value = 0.0002558). However, no outliers are detected (Grubbs test [56] for one outlier: G = 2.99489, U = 0.97658, p-value = 0.5018). See Appendix H for the implementation of these tests.

**7. Discussion**

The initial proposal of the *Q* test for normality, based on the PSNS, defined a sum of squares of seven standardized quantiles without interaction terms [5]. Although it does not assume independence between quantiles, their correlations are ignored to allow the use of the chi-square distribution with seven degrees of freedom as an approximate sampling distribution. This approximation greatly simplifies the calculation of the probability value, Type II error, statistical power, and effect size, as is also done in the $K^2$ test [55, 57-60]. The test requires a large random sample of a continuous variable.
Based on a simulation study and the approximate calculation of the probability value and statistical power using the chi-square distribution with seven degrees of freedom, Moral [5] suggested a minimum sample size of 150 participants for the *Q* test to achieve high accuracy (proportion of correct decisions in retaining or rejecting the null hypothesis of normality) and high power (for non-normal distributions) or complement of power (for the normal distribution). However, for the normal distribution, the test exhibited an accuracy of 1 in detecting or retaining the null hypothesis of normality with a sample size as small as 20, while maintaining a power below 0.2.
The lower power of the *Q* test in rejecting the null hypothesis for non-normal distributions was particularly evident in symmetric distributions with slight leptokurtosis (logistic) or platykurtosis (uniform, semicircular, and arcsine). This reduced power is explained by the high correlations observed among the four central numbers of the seven-number summary.

The present study does not ignore the correlation between quantiles but addresses it in two ways: by generating the sampling distribution of the $Q$ statistic using the bootstrap method ($Q_B$) or by introducing a new expression of the statistic that incorporates interaction terms between pairs of random variables, weighted by their corresponding correlations. This new approach takes the form of a quadratic expression, where the vector of standardized quantiles remains the same as in the previous proposal, but the symmetric square matrix is replaced by the correlation matrix of the quantiles under the null hypothesis of normality, instead of an identity matrix. Since the standardization of the quantiles and the correlation matrix is based on their expected values under the null hypothesis of normality, the subscript T (theoretical distribution) is added to $Q$ (quadratic form) to denote the test statistic.

A sample simulation was used as in the previous study. Thus, simulations were performed for 31 sample sizes across 12 distributions (one normal and 11 non-normal) using a fixed seed, resulting in 372 random samples, each of which was subjected to four normality tests. On the one hand, normality was assessed using three versions of the $Q$-test: one ignoring the correlations between quantiles, one generating the sampling distribution using a bootstrap procedure, and one specifying the quadratic form with the correlation matrix of the quantiles. On the other hand, normality was also tested using the Shapiro-Wilk $W$ test [12-13]. The effectiveness (hit rate) and power of the tests were evaluated, and both metrics were compared across the four normality tests using two repeated measures tests: Cochran's $Q$ test for proportion of correct classifications and Friedman's test for mean power.

As in the previous study [5], the $W$ test exhibited the highest accuracy and statistical power. Its power was significantly higher than those of the three versions of the $Q$ test and its hit ratio was significantly higher than those of the approach Q test based on chi-square distribution and generalized chi-square distribution, but not than bootstrap approach. The latter showed significant differences in accuracy (hit ratio) y statistical power than other two versions of Q test.

The $Q_B$ test demonstrated perfect accuracy, or hit ratio, for normal distribution that is a mesokurtic symmetrical distribution, as well as for distributions very distant from normality (exponential, arcsine, and chi-square), except for these last two distributions are results shared by the four normality tests. Its accuracy was very high for platykurtic symmetrical distributions (uniform, arcsine, and semicircular) with an average of 0.957, which was the weakness of the original version of Q test. It also was very high for leptokurtic symmetrical distribution (Laplace, Student, and hyperbolic secant) with an average of 0.935, even for a distribution with slight positive asymmetry ($\sqrt{\beta_1} \approx 0.631$) and slight pointing ($\beta_2 - 3 \approx 0.256$) as is the Rayleigh distribution with a value of 0.968. For the slightly leptokurtic symmetrical distribution included in the study, namely the logistic distribution, its accuracy was high of 0.806. In fact, its accuracy was equivalent to that of the $W$ test across all 12 distributions. Meanwhile, the $Q_T$ test showed lower accuracy than the $W$ test for the chi-square, uniform, logistic, and Rayleigh distributions, as well as lower accuracy than the original version of the $Q$ test for the logistic and Rayleigh distributions. Consequently, the bootstrap approach is the best choice among the three versions of the $Q$ test.

Is it worth questioning why the quadratic form is not the best option for the Q test—or, in fact, why it performs the worst? The quadratic form is the most theoretically grounded expression, clearly based on the assumption of normality, which serves as the null hypothesis of the distributional model. Accordingly, it is denoted with the subscript T (for theoretical) and has a specific implementation in R. The issue may lie in the second-degree polynomial involving squared standardized values and their cross-products, weighted by the corresponding correlation coefficients. This arithmetic structure makes the test overly conservative with respect to the null hypothesis, ultimately rendering it inadequate and lacking in statistical power. In contrast, the bootstrap procedure generates the sampling distribution in a way that incorporates the influence of quantile correlations, but without the excessive conservatism of the explicit quadratic form. It also avoids the theoretical limitation of assuming independence, as in the original test form, which proved to be both appropriate and powerful.

While the bootstrap approach is as accurate as the other two versions of the Q test in maintaining the null hypothesis under normally distributed data, it exhibits a lower type II error rate (i.e., higher

statistical power). Thus, it represents the least conservative formulation with respect to the null hypothesis of normality in the Q test.

Although the *W* test was a better option than the $Q_B$ test in its bootstrap approach, the latter remains an important alternative when using the seven-number summary to assess normality. Both tests showed similar accuracy (0.952, 95% Wilson-type *CI* [0.925, 0.969] vs. 0.976, 95% Wilson-type *CI* [0.955, 0.987]) and average power (0.935, 95% BCa *CI* [0.919, 0.950] vs. 0.961, 95% BCa *CI* [0.946, 0.975]). Notably, the *W* test does not clearly outperform the *Q* test with samples of 50, and a previous study had already reported this finding with smaller samples of 20 and 30 data points [5]. Similarly, Souza, Toebe, Mello, and Bittencourt [61] found that the Shapiro-Wilk test performed poorly with small samples.

The example with real data demonstrates that the bootstrap version of the Q test leads to the same conclusion as traditional normality tests—such as the Shapiro-Wilk test—namely, the rejection of the null hypothesis. In contrast, the other two versions of the Q test assume normality. Although this assumption is supported by previous studies conducted with psychology students [62], high school students [63], and both clinical [20] and general [64] populations, the current sample consists of students specifically aspiring to enter a psychology program during the selection process. Therefore, the deviation from normality, characterized by positive skewness, may be attributed to impression management (i.e., social desirability). Since alexithymia entails low emotional intelligence, it is not a desirable trait for candidates seeking to become psychologists. Consequently, participants provide lower scores on the alexithymia scale.

Regarding the limitations of the study, it should be noted that the simulation was restricted to a single sample per condition (sample size × distribution type) due to computational processing constraints. However, a seed was used to ensure the reproducibility of results and to achieve greater statistical power in comparisons among the four normality tests when applying repeated measures statistical tests (Cochran's *Q* and Friedman's test). Small sample sizes (20, 30, 40) were not included, as a previous study on the *Q* test indicated the necessity of larger samples, including *W* test [5]. Additionally, only 12 continuous distributions were considered, meaning the behavior of the *Q* test with discrete distributions, such as uniform, binomial, negative binomial, hypergeometric, negative hypergeometric, or Poisson, was not examined [65]. The comparison was limited to the Shapiro-Wilk test [12-13], which is currently regarded as the most powerful test for sample sizes ranging from 3 to 2000 [66]. However, other alternatives exist, such as the Shapiro-Francia test [16-17], which can be used for samples of up to 5,000 data points and has a statistical power similar to that of the Shapiro-Wilk test [15].

**8. Conclusions**

It is concluded that the bootstrap approach in the *Q* test, based on the parametric seven-number summary (PSNS), represents a substantial improvement over the original version, which relied on the chi-square distribution with seven degrees of freedom and did not account for correlations among the seven quantiles. The $Q_B$ variant significantly improves both the test's accuracy and statistical power. Furthermore, its performance is comparable to that of the Shapiro–Wilk W test, which is currently regarded as the most powerful normality test for samples of up to 2,000 observations.

**9. Suggestions**

Further investigation of the $Q_B$ test with discrete distributions that converge to normality is recommended, considering both parameterizations that deviate from normality and parameter values that ensure proximity to it. For instance, the binomial distribution B(10, 0.1) can be considered far from normal, warranting rejection of the null hypothesis of normality in this case. In contrast, the binomial distribution B(100, 0.5) has parameter values that closely approximate the normal distribution, supporting the retention of the null hypothesis.

Similarly, in the case of the Poisson distribution, a rate parameter of 3 results in a distribution that deviates significantly from normality, justifying rejection of the null hypothesis. Conversely, a rate parameter of 100 provides an acceptable approximation to normality, reinforcing the retention of the null hypothesis.

It is also suggested to include larger sample sizes (ranging from 1,000 to 10,000 in increments of 500) and to compare the $Q_B$ test with the Shapiro-France normality test [16-17] and the quantile-based test

by Avdović and Jevremović [6]. The expectation is that the $Q_T$ test will exhibit similar or slightly lower accuracy and power values. If additional tests such as the Kolmogorov-Smirnov test [67] or the G-test [68] are included, the $Q_B$ test is expected to demonstrate superior performance.

The application of the $Q_B$ test is recommended either as a standalone procedure—accompanied by graphical representations such as bar and box plots, histograms with overlaid density and normal curves, and normal quantile-quantile plots [69-70]—particularly when the seven-number summary is reported [4], or as a complementary test alongside the Shapiro–Wilk test [12-13]. The R scripts developed in this study are available for this purpose.

The Freedman–Diaconis rule is used to determine the bin width and number of bins for the histogram [49] representing the sample data. Epanechnikov's kernel function [50] and the Sheather–Jones bandwidth selector [51] are employed to estimate the density of the overlaid empirical curve. These options were selected due to their flexibility regarding the normality assumption and their closer alignment with empirical data [71]. Additionally, a normal curve is included as a theoretical reference in the plot, which can be saved in Tagged Image File Format (TIFF). This format, developed by Aldus Corporation, is known for its high quality and support for both lossless and lossy compression, making it suitable for various purposes, including archiving and printing [72].

**Funding:** This research received no external funding.

**Acknowledgments:** The author expresses his gratitude to the reviewers and editors for the suggestions and corrections received for the improvement of the manuscript.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

[1] A. L. Bowley, Elements of Statistics. P. S. King, London, 1901.

[2] A. L. Bowley, Elementary Manual of Statistics. P. S. King, London, 1910.

[3] P. Muthudoss, S. Kumar, E. Y. C. Ann, K. J. Young, R. L. R. Chi, R. Allada, B. Jayagopal, A. Dubala, I. B. Babla, S. Das, S. Mhetre, I. Saraf, and A. Paudel, "Topologically directed confocal Raman imaging (TD-CRI): Advanced Raman imaging towards compositional and micromeritic profiling of a commercial tablet components," J. Pharm. Biomed. Anal., vol 210, article 114581, February 2022. https://doi.org/10.1016/j.jpba.2022.114581

[4] J. W. Tukey, Exploratory Data Analysis. Addison and Wesley, Reading, MA, 1977.

[5] J. Moral, "Testing for normality from the parametric seven number summary," Open J. Stat., vol. 12, no. 1, pp. 118-154, February 2022. https://doi.org/10.4236/ojs.2022.121009

[6] A. Avdović, and V. Jevremović, "Quantile-zone based approach to normality testing," Mathematics, vol. 10, no. 11, article 1828, May 2022. https://doi.org/10.3390/math10111828

[7] G. Dikta, and M. Scheer, Bootstrap Methods: with Applications in R. Springer Nature, Cham, Switzerland, 2021. https://doi.org/10.1007/978-3-030-73480-0

[8] B. Efron, and B. Narasimhan, "The automatic construction of bootstrap confidence intervals," J Comput Graph Stat., vol. 29, no. 3, pp. 608-619, Mach 2020. https://doi.org/10.1080/10618600.2020.1714633

[9] G. Rousselet, C. R. Pernet, and R. R. Wilcox, "An introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations", Meta-Psychology, vol. 7, article 2058. https://doi.org/10.15626/MP.2019.2058

[10] A. Stuart, and J. K. Ord, Kendall's Advanced Theory of Statistics. Volume 1. Distribution Theory, 6th ed., Edward Arnold, London, 1994.

[11] P. Lafaye de Micheaux, Package CompQuadForm: Distribution function of quadratic forms in normal variables, 2025. https://cran.r-project.org/web/packages/CompQuadForm/CompQuadForm.pdf

[12] S. S. Shapiro, and M. B. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, vol. 52, no. 3-4, pp. 591-611, December 1965. https://doi.org/10.1093/biomet/52.3-4.591

[13] J. P. Royston, "Approximating the Shapiro-Wilk W-test for non-normality," Statistics and Computing, vol. 2, no. 3, pp. 117-119, September 1992. https://doi.org/10.1007/BF01891203

[14] J. Arnastauskaitė, T. Ruzgas, and M. Bražėnas, "An exhaustive power comparison of normality tests," Mathematics, vol. 9, no. 7, article 788, April 2021. https://doi.org/10.3390/math9070788

[15] N. Khatun, "Applications of normality test in statistical analysis," Open J. Stat., vol. 11, no. 1, pp. 113-122, January 2021. https://doi.org/10.4236/ojs.2021.111006

[16] S. S. Shapiro, and R. S. Francia, "An approximate analysis of variance test for normality," J. Am. Stat. Assoc., vol. 67, no 337, pp. 215-216, June 1972. https://doi.org/10.1080/01621459.1972.10481232

[17] J. P. Royston, "A toolkit of testing for non-normality in complete and censored samples," J. Roy. Stat. Soc. Ser. D-The Statistician, vol. 42, no. 1, pp. 37-43, March 1993. https://doi.org/10.2307/2348109

[18] R. M. Bagby, J. D. Parker, and G. J. Taylor, "The twenty-item Toronto Alexithymia Scale--I. Item selection and cross-validation of the factor structure," J. Psychosom. Res., vol. 38, no. 1, pp. 23-32, Janaury 1994. https://doi.org/10.1016/0022-3999(94)90005-1

[19] J. Moral, "Propiedades psicométricas de la Escala de Alexitimia de Toronto de 20 reactivos en México," Revista Electrónica de Psicología Clínica de Iztacala, vol. 11, no. 2, pp. 97-114, July 2008. Available at: https://www.iztacala.unam.mx/carreras/psicologia/psiclin/

[20] G. Pedersen, E. Normann‐Eide, I. U. M. Eikenæs, E. H. Kvarstein, and T. Wilberg, "Psychometric evaluation of the Norwegian Toronto Alexithymia Scale (TAS‐20) in a multisite clinical sample of patients with personality disorders and personality problems," J. Clin. Psychol., June 2022, vol. 78, no. 6, pp. 1118-1136, 2022. https://doi.org/10.1002/jclp.23270

[21] H. Zhang, J. Shen, and Z. Wu, "A fast and accurate approximation to the distributions of quadratic forms of Gaussian variables," J Comput. Graph. Stat., vol. 31, no. 1, pp. 304-311, January 2022. https://doi.org/10.1080/10618600.2021.2000423

[22] M. Estaki, L. Jiang, N. A. Bokulich, D. McDonald, A. González, T. Kosciolek, C. Martino, Q. Zhu, A. Birmingham, Y. Vázquez-Baeza, M. R. Dillon, E. Bolyen, J. Caporaso, G., and R. Knight, "QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data," Curr. Protoc. Bioinformatics, vol. 70, article e100, April 2020. https://doi.org/10.1002/cpbi.100

[23] J. M. O. Machado, Outlier detection in accounting, master thesis, University of Porto. Institutional repository of the University of Porto, 2018. https://sigarra.up.pt/fep/en/pub_geral.show_file?pi_doc_id=423163

[24] S. Cai, J. Zhou, and J. Pan, "Estimating the sample mean and standard deviation from order statistics and sample size in meta-analysis," Stat. Methods Med. Res., 30, vol. 12, pp. 2701-2719, October 2021. https://doi.org/10.1177/09622802211047348

[25] R. J. Hyndman, and Y. Fan, "Sample quantiles in statistical packages," Am. Stat., vol. 50, no. 4, pp. 361-365, November 1996. https://doi.org/10.1080/00031305.1996.10473566

[26] L. Makkonen, and M. Pajari, "Research article defining sample quantiles by the true rank probability," J. Probab. Stat., vol. 2014, article 326579, December 2014. http://dx.doi.org/10.1155/2014/326579

[27] D. M. Hawkins, "Quantile-quantile methodology-detailed results," arXiv, article 2303.03215, October 2023. https://doi.org/10.48550/arXiv.2303.03215

[28] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," Math. Proc. Camb. Philos. Soc., vol. 30, no. 2, pp. 178-191, April 1934. https://doi.org/10.1017/S0305004100016595

[29] P. McCullagh, "Gaussian distributions," in Ten Projects in Applied Statistics, Springer Series in Statistics. Springer, Cham, pp. 251-277. https://doi.org/10.1007/978-3-031-14275-8_15

[30] G. Casella, and R. Berger, Statistical Inference. CRC Press, Hoboken, NJ, 2024. https://doi.org/10.1201/9781003456285

[31] W. Y. Chen, G. W. Peters, R. H. Gerlach, and S. A. Sisson, "Dynamic quantile function models," Quantitative Finance, vol. 22, no. 9, pp. 1665-1691, April 2022. https://doi.org/10.1080/14697688.2022.2053193

[32] A. Canty, B. Ripley, and A. R. Brazzale, Package 'boot'. Bootstrap functions, 2024 https://cran.r-project.org/web/packages/boot/boot.pdf

[33] J. Cohen, "A power primer," in A. E. Kazdin, Ed., Methodological Issues and Strategies in Clinical Research, 4th ed. American Psychological Association, Washington, DC, 2016, pp. 279-284. https://doi.org/10.1037/14805-018

[34] R. B. Davies, "Algorithm AS 155: The distribution of a linear combination of chi-2 random variables." J. R. Stat. Soc. Ser. C Appl., vol. 29, no. 3, pp. 323-333, December 1980. https://doi.org/10.2307/2346911

[35] J. Cohen, Statistical Power Analysis for Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

[36] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification," Ann. Math. Stat., vol. 25, no. 3, pp. 484-498, September 1954. https://doi.org/10.1214/aoms/1177728717

[37] A. Das, "New methods to compute the generalized chi-square distribution," arXiv, article 2404.05062, February 2024. https://doi.org/10.48550/arXiv.2404.05062

[38] H. Cramer, Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ, 1946.

[39] S. J. Peterson, and S. Foley, "Clinician's guide to understanding effect size, alpha level, power, and sample size," Nutr. Clin. Pract., vol. 36, no. 3, pp. 598-605, June 2021. https://doi.org/10.1002/ncp.10674

[40] C. Baumgarten, and T. Patel, "Automatic random variate generation in Python," Proceedings of the 21st Python in Science Conference, July 11, 2022., pp. 46-51. https://doi.org/10.25080/majora-212e5952-007

[41] R Core Team (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

[42] N. Smeeton, N. Spencer, and P. Sprent, Applied Nonparametric Statistical Methods. CRC Press, Boca Raton, FL, 2025. https://doi.org/10.1201/9780429326172

[43] C. F. Fey, T. Hu, and A. Delios, "The measurement and communication of effect sizes in management research," Manag. Organ. Rev., vol. 19, no. 1, pp. 176-197, February 2023. https://doi.org/10.1017/mor.2022.2

[44] L. A. O. Orawo, "Confidence intervals for the binomial proportion: a comparison of four methods," Open J. Stat., vol. 11, no. 5, pp. 806-816, October 2021. https://doi.org/10.4236/ojs.2021.115047

[45] H. Pham, Springer Handbook of Engineering Statistics, 2023. https://doi.org/10.1007/978-1-4471-7503-2

[46] M. J. Blanca-Mena, R. Alarcón, J. Arnau, J. García-Castro and R. Bono, "How to proceed when normality and sphericity are violated in the repeated measures ANOVA," Ann. Psychol., vol. 40, no. 3, pp. 466-480, July 2024. https://doi.org/10.6018/analesps.594291

[47] Microsoft Corporation. Microsoft Excel (version 2021) [Software], Microsoft Corp., 2021.

[48] IBM Corporation, IBM SPSS Statistics (version 27) [Software], IBM Corp., 2020.

[49] D. Freedman, and P. Diaconis, "On the histogram as a density estimator: L2 theory," Probab. Theory Relat. Fields, vol. 57, no. 4, pp. 453-476, December 1981. https://doi.org/10.1007/BF01025868

[50] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," Theory Probab. Appl., vol. 14, no. 1, pp. 153-158, January 1969. https://doi.org/10.1137/1114019

[51] S. J. Sheather, and Jones, M. C. "A reliable data-based bandwidth selection method for kernel density estimation," J. R. Stat. Soc. Ser. B Stat. Method., vol. 53, no. 3, pp. 683-690, December 1991. https://doi.org/10.1111/j.2517-6161.1991.tb01857.x

[52] H. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," J. Am. Stat. Assoc., vol. 62, no. 318, pp. 399-402, June 1967. https://doi.org/10.1080/01621459.1967.10482916

[53] T. W. Anderson, and D. A. Darling, "Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes," Ann. Math. Stat., vol. 23, no. 2, pp. 193-212, June, 1952. https://doi.org/10.1214/aoms/1177729437

[54] F. J. Anscombe and W. J. Glynn, "Distribution of the kurtosis statistic b2 for normal samples," Biometrika, vol. 70, no. 1, pp. 227-234, April 1983 https://doi.org/10.1093/biomet/70.1.227R.

[55] R. B. D'Agostino, "Transformation to normality of the null distribution of g1," Biometrika, vol. 57, no. 3, pp. 679-681, December 1970. https://doi.org/10.1093/biomet/57.3.679

[56] F. E. Grubbs, "Sample criteria for testing outlying observations," Ann. Math. Statist. Vol. 21, no. 1, pp. 27-58, March, 1950. https://doi.org/10.1214/aoms/1177729885

[57] D'Agostino, and E. S. "Pearson, Tests for departure from normality: empirical results for the distributions of b2 and $\sqrt{b1}$," Biometrika, vol. 60, no. 3, pp. 613-622, 01 December 1973. https://doi.org/10.1093/biomet/60.3.613

[58] R. B. D'Agostino, A. Belanger, and R. B. Jr. D'Agostino, "A suggestion for using powerful and informative tests of normality," Am. Stat., vol. 44, no. 4, pp. 316-321, December 1990. https://doi.org/10.1080/00031305.1990.10475751

[59] C. M. Jarque, and A. Bera, "A test for normality of observations and regression residuals," Int. Stat. Rev., vol. 55, no. 2, pp. 163-172, August 1987. https://doi.org/10.2307/1403192

[60] C. Urzua, "On the correct use of omnibus tests for normality," Econ. Lett., vol. 53, no. 3, pp. 247-251, December 1996. https://doi.org/10.1016/S0165-1765(96)00923-8

[61] R. R. De Souza, M. Toebe, A. C. Mello, and K. C. Bittencourt, "Sample size and Shapiro-Wilk test: An analysis for soybean grain yield," Eur. J. Agron., vol. 142, no. 1, article 126666, January 2023. https://doi.org/10.1016/j.eja.2022.126666

[62] Y. Şener, and Y. Günaydın, "Attitudes Toward Dating Violence, Social Impact, and Alexithymia in University Students: A Structural Equation Modeling," Psychol Reports, 0(0), July, 2024. https://doi.org/10.1177/00332941241265618

[63] M. A. Terzioğlu, and A. Büber, "Alexithymia, internet addiction, and cyber-victimisation among high school students in Turkey: an exploratory study," Behav. Inform. Technol., vol. 44, no. 7, pp. 1350-1361, May 2024. https://doi.org/10.1080/0144929X.2024.2353273

[64] M. Miniati, C. Poidomani, C. Conversano, G. Orrù, R. Ciacchini, D. Marazziti, G. Perugi, A. Gemignani, and L. Palagini, "Alexithymia and interoceptive confusion in Covid-19 pandemic distress," Clinical Neuropsychiatry: Journal of Treatment Evaluation, vol. 20, no. 4, pp. 264-270, January 2023. https://doi.org/10.36131/cnfioritieditore20230405

[65] J. L. Devore, K. N. Berk, and M. A. Carlton, Modern Mathematical Statistics with Applications. Springer Nature, Berlin, 2021. https://doi.org/10.1007/978-3-030-55156-8

[66] S. Demir, "Comparison of normality tests in terms of sample sizes under different skewness and Kurtosis coefficients," IJATE, vol. 9, no. 2, pp. 397-409, June 2022. https://doi.org/10.21449/ijate.1101295

[67] J. Vrbik, "Deriving CDF of Kolmogorov-Smirnov test statistic," Appl. Math., vol. 11, pp. 227-246, March 2020. https://doi.org/10.4236/am.2020.113018

[68] S. S. Mangiafico, "G-test of goodness-of-fit," in An R Companion for the Handbook of Biological Statistics, Version 1.3.9, 2023. https://rcompanion.org/rcompanion/b_04.html

[69] H. Wickham, W. Chang, L. Henry, K. Takahashi, C. Wilke, K. Woo, H. Yutani, D. Dunnington, T. van den Brand, and P. L. Pedersen, ggplot2: Create elegant data visualisations using the grammar of graphics, 2024. https://doi.org/10.32614/CRAN.package.ggplot2

[70] S. Phuyal, M. Rashid, and J. Sarkar, GIplot: An R package for visualizing the summary statistics of a quantitative variable, 2021. https://cran.r-project.org/web/packages/GIplot/index.html

[71] L. Gonzales-Fuentes, L. Barbé, K., Barford, L. Lauwers, and L. Philips, "A qualitative study of probability density visualization techniques in measurements," Measurement, vol. 65, no. 1, pp. 94-111, April 2015. https://doi.org/10.1016/j.measurement.2014.12.022

[72] A. J. Qasim, and F. Q. A. Alyousuf, "History of image digital formats using in information technology," Qalaai Zanist Scientific Journal, vol. 6, no. 2, pp. 1098-1112, June 2021. https://doi.org/10.25212/lfu.qzj.6.2.41

## Appendices

### Appendix A. Q-statistic

```
# Shared script for the Q_B and Q tests.
# It serves as the introductory section of the scripts found in Appendices C and E.

# Define data vector.
x <- c()

# Sample representation using a histogram (based on the Freedman-Diaconis rule) with density curves (Epanechnikov's kernel and Sheather-Jones bandwidth) and normal curves overlaid.
# Remove the hash symbol from tiff() and dev.off to save the graphic as a TIFF file.
tiff("histogram.tiff", width = 1600, height = 900, units = "px", res = 300)
par(mar = c(4, 4, 1, 1) + 0.1) # Set the plot margins.

# Define a wider range for x to ensure complete visualization
x_range <- range(x)
x_buffer <- 0.2 * diff(x_range)
xlim_adjusted <- c(x_range[1] - x_buffer, x_range[2] + x_buffer)
# Compute density for ylim
density_x <- density(x, kernel = "epanechnikov", bw = "SJ")
y_normal <- dnorm(x, mean = mean(x), sd = sd(x))
# Histogram based on the Freedman-Diaconis rule.
hist(x, breaks = "FD", freq = FALSE, col = "lightyellow", border = "black",
```

```
main = "", xlab = "Scores of X", xlim = xlim_adjusted, ylab = "Density", ylim = c(0, max(density_x$y, y_normal) + 0.01))
# Overlay a density curve (Epanechnikov's kernel and Sheather-Jones bandwidth)
lines(density_x, col = "darkblue", lwd = 4)
# Overlay the expected density curve if data follow a normal distribution.
x_seq <- seq(xlim_adjusted[1], xlim_adjusted[2], length.out = 1000)
y_normal <- dnorm(x_seq, mean = mean(x), sd = sd(x))
lines(x_seq, y_normal, col = "red", lwd = 4)
# Add legend.
legend("topleft", legend = c("Empirical", "Normal"), col = c("darkblue", "red"),
lwd = 4, bty = "n", title = "", cex = 0.8)
dev.off()

# Calculation of sample quantiles at the orders corresponding to the parametric seven-number summary for a normal distribution, using the R type-8 rule.
p <- c(pnorm(-2), pnorm(-4/3), pnorm(-2/3), pnorm(0), pnorm(2/3), pnorm(4/3), pnorm(2))
q <- quantile(x, probs = p, type = 8)
cat("\nSample quantiles corresponding to Parametric Seven-Number Summary (PSNS)\n")
print(round(q, 4))

cat("\nStatistics needed to standardize sample quantiles corresponding to the PSNS:\n")
n <- length(x) # sample size
m <- mean(x) # sample mean
s <- sd(x) # sample standard deviation
cat("\nSample size:", n, ".\n")
cat("\nSample mean:", m, ".\n")
cat("\nSample standard deviation:", s, ".\n")

# Computation of the statistical value for the normality test, derived from the parametric seven-number summary.
e_zp <- qnorm(p) # Expected value under normal.
d_zp <- dnorm(qnorm(p)) # Expected density under normal.
ee_zp <- sqrt((p * (1 - p)) / (n * dnorm(qnorm(p))^2)) # Standard error under normal.
zx <- (q - m) / s # Standardized quantile.
z <- (zx - e_zp) / ee_zp # Standardization of zx under assumption of normality.
z_sq <- z^2 # squared of the z_q statistic.
Q_stat <- sum(z_sq) # test statistic for the normality Q test.

# Creating a table from a data frame with the previous calculations.
tabla <- data.frame(e_zp = round(e_zp, 4), p = round(p, 4), d_zp = round(d_zp, 4), ee_zp = round(ee_zp, 4), x_p = round(q, 4), z_xp = round(zx, 4), z = round(z, 4), z_sq = round(z_sq, 4))
fila_suma <- data.frame(e_zp = "Sum", p = "", d_zp = "", ee_zp = "", x_p = "", z_xp = "",
z = "", z_sq = format(Q_stat, digits = 6))
tabla <- rbind(tabla, fila_suma)
# Displaying the table.
cat("\nTable: Testing normality using the test based on the Parametric Seven-Number Summary\n")
print(tabla, row.names = FALSE)
cat("\nNote: e_zp = expected value for the p-th quantile in a N(0, 1) distribution,
p = quantile order,
d_zp = density of the p-th quantile in a N(0, 1) distribution,
ee_zp = standard error of the p-th quantile for n normally distributed data points,
```

```
x_p = p-th sample quantile,
z_xp = standardized quantile (calculated using the sample mean and standard deviation),
z = (z_xp - e_zp) / ee_zp = standardization of the z_xp statistic under the assumption of normality,
z^2 = squared z_xp statistic.\n")

# Display value of test statistic in original sample.
cat("\nValue of test statistic in original sample: Q-statistic =", Q_stat, ".\n")
```

**Appendix B. Bootstrap version of PSNS Q test**

```
# Script for the Q_B test: Bootstrap probability and effect size.
# This script is designed to be used in conjunction with the script from Appendix B, but it can also be executed independently.

# Function to compute the Q statistic from a sample vector using R's type-8 quantiles.
compute_q_statistic <- function(sample, type = 8) {
# Define the seven standard normal quantiles used in the PSNS:
p <- c(pnorm(-2), pnorm(-4/3), pnorm(-2/3), pnorm(0), pnorm(2/3), pnorm(4/3), pnorm(2))
# Obtain the sample quantiles at those probabilities
q <- quantile(sample, probs = p, type = type)
# Standardize the sample quantiles using the mean and standard deviation of the bootstrap sample.
zx <- (q - mean(sample)) / sd(sample)
# Compute the squared deviations between the standardized sample quantiles and the corresponding theoretical quantiles, scaled by its standard error (under null hypothesis of normality)
z_sq <- ((zx - qnorm(p)) / sqrt((p * (1 - p)) / (length(sample) * dnorm(qnorm(p))^2)))^2
sum(z_sq)
}

# Value of test statistic in original sample.
Q_stat <- compute_q_statistic(x)

# Bootstrap sampling distribution for the Q statistic in the normality test based on the Parametric Seven-Number Summary (PSNS).
set.seed(123) # Seed for reproducibility of results.
Q_bootstrap <- numeric(2000)
for (i in 1:2000) {
x_boot <- sample(x, length(x), replace = TRUE)
Q_bootstrap[i] <- compute_q_statistic(x_boot)
}
# print(round(Q_bootstrap, 4)) # Remove the hash symbol if you want to display the empirical bootstrap distribution.

# Bootstrap critical value for the null hypothesis of normality using R's type-9 quantiles in the function that computes the Q statistic.
set.seed(123) # Seed for reproducibility of results,
Q_null <- numeric(2000)
for (i in 1:2000) {
x_norm_boot <- rnorm(length(x), mean = mean(x), sd = sd(x))
Q_null[i] <- compute_q_statistic(x_norm_boot, type = 9)
}
# print(round(Q_null, 4)) # Remove the hash symbol if you want to display the normative bootstrap distribution.
```

```
# Critical value and bootstrap p-value
alpha <- 0.05
Q_critical <- quantile(Q_null, 1 - alpha, type = 8)

# Print results.
cat("\nValue of test statistic in original sample", Q_stat, ".\n")
cat("\nBootstrap sampling distribution of the Q statistic generated from the original sample\n")
cat("\nNumber of extractions with replacement =", length(Q_bootstrap), ".\n")
cat("\nBootstrap estimation of the Q =", mean(Q_bootstrap), ".\n")
cat("\nBootstrap standard error of the Q =", sd(Q_bootstrap), ".\n")
cat("\nBootstrap bias of the Q =", mean(Q_bootstrap) - Q_stat, ".\n")
cat("\nNormative bootstrap sampling distribution of the Q statistic\n")
cat("\nThis is generated from a normal distribution with mean", mean(x), "and standard deviation",
sd(x), "\n")
cat("\nBased on", length(Q_null), "extractions of size", length(x), ".\n")
cat("\nBootstrap expected value of Q under normal =", mean(Q_null), ".\n")
cat("\nBootstrap standard error of Q under normal =", sd(Q_null), ".\n")
p_bootstrap <- mean(Q_null > Q_stat)
cat("\nBootstrap critical value for the Q statistic with a significance level of", alpha, "=", Q_critical,
".\n")
cat("\nBootstrap p-value =", format(p_bootstrap, scientific = FALSE, digits = 5), ".\n")
if (p_bootstrap < alpha) {
cat("\nThe null hypothesis that the data follow a normal distribution is rejected \nwith a significance
level of", alpha, "based on the bootstrap p-value.\n")
} else {
cat("\nThe null hypothesis that the data follow a normal distribution is not rejected \nwith a
significance level of", alpha, "based on the bootstrap p-value.\n")
}

# Effect size.
Z_Q <- abs(Q_stat - mean(Q_null)) / sd(Q_null)
cat("\nEffect size as standardized distance: Z_Q =", Z_Q, ".\n")
cat("\nInterpretation of effect size\n")
cat("< 1.7: Trivial\n[1.7, 2.5): Small\n[2.5, 3): Medium\n ≥ 3: Large\n")
effect_size <- ifelse(Z_Q < 1.7, "trivial",
ifelse(Z_Q < 2.5, "small",
ifelse(Z_Q < 3, "medium", "large")))
cat("\nBased on the Z statistic = ", round(Z_Q, 4), ", the size effect of the deviation from the normal
distribution model is classified as", effect_size, ".\n")

# Histogram of the bootstrap sampling distribution of the Q statistic from the sample data,
# with an overlaid kernel density estimate and a chi-square density curve with seven degrees of
freedom.
# Remove the hash symbol from tiff() and dev.off to save the graphic as a Tagged Image File Format
(TIFF) file.
# tiff("Empiric_hist.tiff", width = 1600, height = 900, units = "px", res = 300)
par(mar = c(4, 4, 1, 1) + 0.1) # Set the plot margins
# Define x-axis limits
x_min <- min(Q_bootstrap)
x_max <- max(Q_bootstrap)
```

```
# Compute kernel density and chi-square density
dens_kernel <- density(Q_bootstrap)
dens_chisq <- dchisq(dens_kernel$x, df = 7)
# Determine max y value for ylim
y_max <- max(dens_kernel$y, dens_chisq)
# Plot histogram with adjusted ylim
hist(Q_bootstrap, breaks = "fd", freq = FALSE, col = "lightyellow", border = "black",
main = "Bootstrap distribution", xlab = "Q statistic (sample data)",
ylab = "Density", xlim = c(x_min, x_max), ylim = c(0, y_max * 1.05))
# Overlay kernel density estimate (Gaussian kernel)
lines(dens_kernel, col = "darkblue", lwd = 4)
# Overlay chi-square density curve with 7 degrees of freedom
curve(dchisq(x, df = 7), from = 0, to = x_max, col = "red", lwd = 4, add = TRUE)
# Add vertical line for Q_critical
abline(v = Q_critical, col = "purple", lwd = 3, lty = 2)
# Add vertical line for Q_statistic
abline(v = Q_stat, col = "black", lwd = 3, lty = 2)
# dev.off()

# Histogram of the bootstrap sampling distribution of the Q statistic under normality,
# with an overlaid kernel density estimate and a chi-square density curve with seven degrees of
freedom.
# Remove the hash symbol from tiff() and dev.off to save the graphic as a TIFF file.
# tiff("Normative_hist.tiff", width = 1600, height = 900, units = "px", res = 300)
par(mar = c(4, 4, 1, 1) + 0.1) # Set the plot margins.
# Density histogram using Freedman-Diaconis rule.
# Check if the upper limit of 16 for the ordinate axis is appropriate.
hist(Q_null, breaks = "fd", freq = FALSE, col = "lightyellow", border = "black",
main = " Bootstrap distribution under normality", xlab = "Q Statistic (Normal samples)", ylab =
"Density", xlim = c(0, 16))
# Overlay kernel density estimate (Gaussian kernel)
lines(density(Q_null), col = "darkblue", lwd = 4)
# Overlay chi-square density curve with 7 degrees of freedom
# Check if the upper limit of 16 for the curve is appropriate.
curve(dchisq(x, df = 7), from = 0, to = 16, col = "red", lwd = 4, add = TRUE)
# Add vertical line for Q_critical
abline(v = Q_critical, col = "purple", lwd = 3, lty = 2)
# Add vertical line for Q_statistic
abline(v = Q_stat, col = "black", lwd = 3, lty = 2)
# dev.off()
```

**Appendix C. Statistical power of bootstrap version of PSNS Q test**

```
# Script for computing the bootstrap power of the normality test using the QB statistic.

x <- c() # Define data vector.
Q_critical <- 6.43475 # Bootstrap critical value. It is taken from the result of the previous script.
alpha <- 0.05 # Level of significance or complement of the critical value order (Q_critical).

# Bootstrap power calculation.
set.seed(456) # Seed for reproducibility of results.
# Create a vector to store results of 10000 simulated tests
```

```
for (i in 1:10000) {
x_boot <- sample(x, length(x), replace = TRUE)
p <- c(pnorm(-2), pnorm(-4/3), pnorm(-2/3), pnorm(0), pnorm(2/3), pnorm(4/3), pnorm(2))
q <- quantile(x_boot, probs = p, type = 8)
zx <- (q - mean(x_boot)) / sd(x_boot)
z_sq <- ((zx - qnorm(p)) / sqrt((p * (1 - p)) / (length(x_boot) * dnorm(qnorm(p))^2)))^2
Q_sim <- sum(z_sq)
}
# Compute empirical power as the proportion of rejections across simulations.
power <- mean(Q_sim > Q_critical)

# Display results
cat("\nNumber of bootstrap samples =", length(Q_sim), ".\n")
cat("\nNumber of data per bootstrap sample =", length(x_boot), ".\n")
cat("\nBootstrap power at a significance level of", alpha, "=", format(power, scientific = FALSE, digits =
5), ".\n")
```

**Appendix D. Quadratic form version of PSNS Q test**

```
# Script for calculating the Q_T statistic (quadratic form), its probability value, statistical power, and
effect size.

# Replace the example with your own data vector.
x <- c(-0.23, -1.39, 0.38, 0.52, -0.49, 0.28, -0.04, 0.11, 1.03, -0.33, -0.33, 0.06, 0.16, 0.29, -0.16, -1.06, 0.54,
0.88, -1.64, -0.31)

# Gaussian vector z.
p <- c(pnorm(-2), pnorm(-4/3), pnorm(-2/3), pnorm(0), pnorm(2/3), pnorm(4/3), pnorm(2))
q <- quantile(x, probs = p, type = 8) # Parametric seven-number summary
n <- length(x) # Sample size
m <- mean(x) # Sample Mean
s <- sd(x) # Standard Sample Deviation
e_zp <- qnorm(p) # Expected value under normal
d_zp <- dnorm(qnorm(p)) # Expected density under normal
ee_zp <- sqrt((p * (1 - p)) / (n * dnorm(qnorm(p))^2)) # Standard error under normal
zx <- (q - m) / s # Standardized quantile
z <- (zx - e_zp) / ee_zp # Standardization of zx statistic under assumption of normality
cat("\nGaussian vector z:\n")
cat(z, "\n")

# Matrix of correlations R under the assumption of normality.
p <- pnorm(c(-2, -4/3, -2/3, 0, 2/3, 4/3, 2)) # Cumulative probabilities corresponding to PSNS z-scores
d <- dnorm(c(-2, -4/3, -2/3, 0, 2/3, 4/3, 2)) # Density values corresponding to PSNS z-scores
n <- 1000 # Number of simulations
# Initialize a square matrix of zeros to store the theoretical covariance values.
cov_matrix <- matrix(0, nrow = length(p), ncol = length(p))
# Compute the diagonal elements (variances) of the covariance matrix.
# These are based on the variance of the sample quantiles under normality.
for (i in 1:length(p)) {
cov_matrix[i, i] <- (p[i] * (1 - p[i])) / (n * d[i]^2)}
# Compute the off-diagonal elements (covariances) of the covariance matrix.
# These reflect the expected covariances between different quantiles under normality.
```

```
for (i in 1:(length(p) - 1)) {
for (j in (i + 1):length(p)) {
cov_matrix[i, j] <- (p[i] * (1 - p[j])) / (n * d[i] * d[j])
cov_matrix[j, i] <- cov_matrix[i, j]}}
# Convert the covariance matrix to a correlation matrix.
# This is done by standardizing each covariance by the square roots of the corresponding variances.
corr_matrix <- cov_matrix / sqrt(outer(diag(cov_matrix), diag(cov_matrix)))
cat("\nCorrelation matrix under the assumption of normality\n")
print(round(corr_matrix, 4))

# Quadratic form.
Q_T <- t(z) %*% corr_matrix %*% z
cat("\nThe Q statistic value in the quadratic form =", Q_T, ".\n")

# Probability value with the generalized chi-square distribution (Davies method).
library(CompQuadForm)
eigenvalues <- eigen(corr_matrix, symmetric = TRUE, only.values = TRUE)$values
p_value <- davies(Q_T, lambda = eigenvalues)$Qq
cat("\nEigenvalues:", eigenvalues, ".\n")
cat("\nNumber of eigenvalues:", length(eigenvalues), "\n")
cat("\nSum of eigenvalues:", sum(eigenvalues), ".\n")
cat("\np-value for the Q_T statistic =", p_value, ".\n")
# Exact calculation of power using the generalized chi-square distribution.
alpha <- 0.05
Q_alpha <- qchisq(1 - alpha, df = length(eigenvalues)) # Initial approach
delta <- eigenvalues * z^2
power <- davies(Q_alpha, lambda = eigenvalues, delta = delta)$Qq
cat("\nStatistical power:", power, ".\n")

# Effect size.
ev_Q <- sum(diag(corr_matrix)) # Expected value of the Q_T statistic.
sd_Q <- sqrt(2 * sum(eigenvalues ^2)) # Standard deviation of the Q_T statistic.
Z_Q <- abs(Q_T - ev_Q) / sd_Q # Standarized value of the Q_T statistic.
cat("\nExpected value of the Q_T statistic =", ev_Q, ".\n")
cat("\nStandard deviation of the Q_T statistic =", sd_Q, ".\n")
cat("\nEffect size as standardized distance: Z_Q =", Z_Q, ".\n")
cat("\nInterpretation of effect size.\n")
cat("< 1.7: Trivial\n[1.7, 2.5): Small\n[2.5, 3): Medium\n ≥ 3: Large\n")
effect_size <- ifelse(Z_Q < 1.7, "trivial",
ifelse(Z_Q < 2.5, "small",
ifelse(Z_Q < 3, "medium", "large")))
cat("\nBased on Z statistic =", round(Z_Q, 4), ", the effect size of the deviation from the normal
distribution model is classified as", effect_size, ".\n")
```

**Appendix E. Chi-square approximation to PSNS Q test**

```
# Script for calculating the probability value, statistical power, and effect size based on the chi-square
approximation with 7 degrees of freedom.
# To be used following script in Appendix B.

cat("\nCalculation of probability value, statistical power, and effect size using the chi-square
approximation with 7 degrees of freedom.\n")
```

```
# Calculation of the probability value from the chi-square approximation with 7 degrees of freedom.
alpha <- 0.05
p_asint <- pchisq(Q_stat, df = 7, lower.tail = FALSE)
cat("\nTest Statistic: Q =", Q_stat, ",", "critical value with a significance level of", alpha,"=",
qchisq(alpha, df = 7, lower.tail = FALSE), ",", "asymptotic p-value =", p_asint, ".\n")
if (p_asint < alpha) {
cat("\nThe null hypothesis that the data follow a normal distribution is rejected \nwith a significance
level of", alpha, "by the Q test.\n")
} else {cat("\nThe null hypothesis that the data follow a normal distribution is not rejected \nwith a
significance level of", alpha, "by the Q test.\n")}

# Statistical power of the Q test using the chi-square distribution approximation with 7 degrees of
freedom.
power <- 1 - pchisq(qchisq(alpha, df = 7, lower.tail = FALSE), df = 7, ncp = Q_stat, lower.tail = FALSE,
log.p = FALSE)
cat("\nThe statistical power to the right tail for the alternative hypothesis of non-normality for the Q
test: ϕ =", round(power, 4), ".\n")
cat("\nIf the null hypothesis of normality is maintained, the power must be less than 0.5; \nwhile, if
rejected, it must be greater than 0.5. In the latter case, it is classified as \ngood with a value of 0.8 and
very good with a value of 0.9.\n")

# Effect size calculation using the chi-square distribution approximation with 7 degrees of freedom.
v <- sqrt(Q_stat / (n * 7))
cat("\nEffect size via Cramer's V-coefficient: V =", round(v, 4), ".\n")

# Interpretation based on Cohen (1988) [29].
cat("\nInterpretation of effect size based on Cohen (1988):\n")
cat("< 0.1: Trivial\n[0.1, 0.3): Small\n[0.3, 0.5): Medium\n≥ 0.5: Large\n")
effect_size <- ifelse(v < 0.1, "trivial",
ifelse(v < 0.3, "small",
ifelse(v < 0.5, "medium", "large")))
cat("\nBased on the V statistic =", round(v, 4), ", the size of the effect of the deviation from the normal
distribution model is classified as", effect_size, ".\n")
```

**Appendix F. Correlation matrix**

```
# Population correlation matrix of PSNS quantiles in a normal distribution
set.seed(123) # Seed for reproducibility of results.
n <- 1000 # Original sample size and each bootstrap sample.
B <- 10000 # Number of bootstrap samples.
p <- pnorm(c(-2, -4/3, -2/3, 0, 2/3, 4/3, 2)) # Quantile orders.
# Function for extracting sample quantiles by R type-9 rule.
get_quantiles <- function(sample) {quantile(sample, probs = p, type = 9)}
# Generating bootstrap distributions.
original_sample <- rnorm(n, 0, 1)
bootstrap_samples <- replicate(B, sample(original_sample, n, replace = TRUE))
# Calculation of quantiles for each bootstrap sample.
quantiles_matrix <- apply(bootstrap_samples, 2, get_quantiles)
# Calculation and printing of the correlation matrix.
correlation_matrix <- cor(t(quantiles_matrix))
print(correlation_matrix)
```

**Appendix G. Generation of original samples**

```
# Generation of original samples of different sizes, drawn from various distributions

# Required libraries.
library(VGAM)
library(EnvStats)
library(extraDistr)

# Define an empty dataframe to store the results.
results <- data.frame(Sample_Size = integer(), Min = numeric(), Gap1 = character(), Mdn = numeric(),
Gap2 = character(), Max = numeric())

# Define sample sizes.
n_values <- c(seq(50, 1500, by = 50), 2000) # De 50 a 1500 en pasos de 50, más 2000

# Set a seed for reproducibility.
set.seed(123)
# set.sed (12) # For normal distribution
# Iterate over sample sizes
for (sample_size in n_values) {

# Generate data from a distribution.
# To select a distribution, remove the hash symbol (#) from the corresponding sample generator. In this
run, the hash was removed from the generator for the Laplace distribution.
x <- rlaplace(sample_size, mu = 50, sigma = 10) # Laplace's distribution.
# x <- rnorm(sample_size, mean = 50, sd = 10) # Normal distribution.
# x <- rt(sample_size, df = 4) * 10 + 50 # Student's t-distribution with 4 degrees of freedom.
# x <- rchisq(sample_size, df = 5) # Chi-square distribution with 4 degrees of freedom.
# x <- rbeta(sample_size, shape1 = 0.5, shape2 = 0.5) # Arcsine distribution.
# x <- runif(sample_size, min = 0, max = 100) # Continuous uniform distribution.
# x <- rtri(sample_size, min = 0, max = 3, mode = 2.99) # Triangular distribution.
# x <- 3 * (2 * rbeta(sample_size, shape1 = 1.5, shape2 = 1.5) - 1) # Wigner's semicircle distribution.
# x <- rexp(sample_size, rate = 2) # Exponential distribution.
# x <- 50 + 10 * 2/pi * log(tan(pi/2 * runif(sample_size, min = 0, max = 1))) # Hyperbolic secant distribution.
# x <- rlogis(sample_size, location = 50, scale = 10) # Logistic distribution.
# x <- rrayleigh(sample_size, sigma = 12) # Rayleigh's distribution.

# Calculate the minimum, median and maximum.
results <- rbind(results, data.frame(Sample_Size = sample_size, Min = min(x), Gap1 = "...", Mdn =
median(x), Gap2 = "...", Max = max(x)))
}
# Print the final table.
# Remove the hash symbol from the corresponding distribution.
cat("\nOriginal sample drawn from a Laplace's distribution\n")
# cat("\nOriginal sample drawn from a normal distribution\n")
# cat("\nOriginal sample drawn from a Student's t-distribution with 4 degrees of freedom\n")
# cat("\nOriginal sample drawn from a chi-square distribution with 4 degrees of freedom\n")
# cat("\nOriginal sample drawn from an arcsine distribution\n")
# cat("\nOriginal sample drawn from a continuous uniform distribution\n")
# cat("\nOriginal sample drawn from a triangular distribution\n")
```

```
# cat("\nOriginal sample drawn from a Wigner's semicircle distribution\n")
# cat("\nOriginal sample drawn from an exponential distribution\n")
# cat("\nOriginal sample drawn from a hyperbolic secant distribution\n")
# cat("\nOriginal sample drawn from a logistic distribution\n")
# cat("\nOriginal sample drawn from a Rayleigh's distribution\n")
colnames(results) <- c("Sample size", "Minimum", "...", "Median", "...", "Maximum")
print(results, row.names = FALSE)
```

**Appendix H. Shapiro-Wilk W test and other tests of normality**

# Shapiro-Wilk W test using Royston's procedure for samples of 12 to 2000 data points and statistical power estimation.

```
# Define a vector of data points
x <- c()

# Load required library
library(nortest)

# Define significance level
alpha <- 0.05

# Compute statistics
sw <- shapiro.test(x)
w <- sw$statistic
p_value <- sw$p.value
sample_size <- length(x)
m <- 0.0038915 * log(sample_size)^3 - 0.083751 * log(sample_size)^2 - 0.31082 * log(sample_size) - 1.5861
sd <- exp(0.0030302 * log(sample_size)^2 - 0.082676 * log(sample_size) - 0.4803)
z_w <- (log(1 - w) - m) / sd
zc <- qnorm(0.95, mean = m, sd = sd)
power <- 1 - pnorm((zc - log(1 - w)) / sd)

# Display results
cat("Shapiro–Wilk W statistic: w =", w, "\n")
cat("Logarithmic transformation of W statistic: ln(1 - w) =", log(1 - w), "\n")
cat("Expected value of ln(1 - w):", m, "\n")
cat("Standard deviation of ln(1 - w):", sd, "\n")
cat("Royston's standardized W statistic: z_w =", z_w, "\n")
cat("Right-tailed p-value:", p_value, "\n")

if (p_value < alpha) {
cat("\nThe null hypothesis that the data follow a normal distribution is rejected\nat the", alpha, "significance level by the Shapiro–Wilk W test using Royston's procedure.\n")
} else {
cat("\nThe null hypothesis that the data follow a normal distribution is not rejected\nat the", alpha, "significance level by the Shapiro–Wilk W test using Royston's procedure.\n")
}

cat("\nStatistical power: ϕ =", power, "\n")

# Lilliefors (Kolmogorov-Smirnov) normality test
```

```
lillie.test(x)
# Anderson-Darling normality test
ad.test(x)
# Outliers, symmetry, and kurtosis tests.
library(outliers)
grubbs.test(x, type = 10)
library(moments)
anscombe.test(x)
agostino.test(x)
```

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

*Research Article*

Received: date:06.04.2025
Accepted: date:21.05.2025
Published: date:30.06.2025

# Machine Learning-Based Wind Energy Forecasting Using Weather Parameters: The Example of Yalova

**Abdulkadir Atalan[1,] *, İlker Çam[2], Lütfi Alper Gündoğdu[3], Harun Kahyalık[4] and Yasemin Ayaz Atalan[5]**

[1]Department of Industrial Engineering, Çanakkale Onsekiz Mart University, Çanakkale Türkiye; abdulkadiratalan@gmail.com
[2]Department of Energy Department; Çanakkale Onsekiz Mart University, Çanakkale Türkiye; ilkerarda15@gmail.com
[3]Department of Energy Department; Çanakkale Onsekiz Mart University, Çanakkale Türkiye; lutfialpergundogdu@gmail.com
[4]Department of Energy Department; Çanakkale Onsekiz Mart University, Çanakkale Türkiye; hrnkhylk@gmail.com
[5]Department of Energy Department; Çanakkale Onsekiz Mart University, Çanakkale Türkiye; yayazatalan@gmail.com
Orcid: 0000-0003-0924-3685[1] Orcid: 0009-0007-5313-5776[2] Orcid: 0000-0002-2002-0952[3] Orcid: 0009-0008-9636-8008[4] Orcid: 0000-0001-7767-0342[5]
*Correspondence: abdulkadiratalan@gmail.com, yayazatalan@gmail.com

**Abstract:** In this study, various machine learning algorithms were evaluated for estimating wind energy production using hourly meteorological data of Yalova province in 2018. The input parameters were input parameters of weather parameters such as temperature, relative humidity, air pressure, wind direction, and wind speed. In the analysis performed on a total of 50530 data points, methods such as Gradient Boosting (GB), Random Forests (RF), k-nearest neighbor (*k*NN), and Stochastic gradient descent (GBD) were compared. Model performances were evaluated according to Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), MAPE, and $R^2$ criteria. According to the results, the best-performing algorithm was RF with an MSE value of 0.039, RMSE value of 0.197, MAE value of 0.081, MAPE value of 0.377, and $R^2$ score of 0.961. On the other hand, the SGD model showed the lowest performance with an MSE value of 0.175, RMSE value of 0.418, MAE value of 0.303, MAPE value of 0.581, and $R^2$ score of 0.822. These findings show that machine learning models, supported by selecting the correct weather parameters, can provide high accuracy in estimating wind energy production and contribute to energy management policies in this direction.
**Keywords:** Wind energy, weather parameters, machine learning, forecasting

## 1. Introduction

Today, energy holds much greater global significance than in the past. A significant portion of the current energy demand is met through fossil fuels [1]. The various environmental, economic, and health-related negative impacts of fossil fuels make it necessary to seek new and sustainable solutions in the energy sector [2]. In this regard, renewable energy sources, developed as alternatives to depletable energy resources, are expected to play a much more significant role in future energy production [3]. Renewable energy sources, such as wind, solar, geothermal, biomass, hydroelectric, and wave energy, stand out due to their sustainable and environmentally friendly nature [4]. Enabling countries to produce energy within their borders, reducing dependence on foreign sources, and ensuring long-term economic benefits make these energy sources strategically valuable. Furthermore, the inexhaustible nature of renewable energy sources and their ability to minimize carbon emissions and reduce environmental impacts strengthen their critical role in the global energy transition process.

Wind energy has gained particular prominence among renewable resources due to its clean and inexhaustible nature [5]. As a renewable and clean energy source, wind energy has emerged as an alternative for energy production [6]. Its environmental benefits, increased electricity production, and

**Citation:** A. Atalan, İ. Çam, L. A. Gündoğdu, H. Kahyalık, and Y. A. Atalan, "Machine Learning-Based Wind Energy Forecasting Using Weather Parameters: The Example of Yalova," *Journal of Statistics and Applied Sciences*, no. 11, pp. 40–49, 2025, doi: 10.52693/jsas.1670486.

reduced fossil-based energy production have made wind energy increasingly important [7]. The amount of energy produced in wind farms is related to weather parameters [8]. For efficient use of wind energy, changes in wind speed need to be observed, and production potential must be accurately predicted [9].

Electricity production from wind energy is highly variable because it directly depends on atmospheric conditions. Variations in wind speed, changes in direction, and other meteorological factors are the main factors determining the daily and seasonal fluctuations in energy production. Machine learning techniques are frequently used to model these variations, and hybrid models are developed to improve accuracy [10]. Atmospheric parameters such as temperature, pressure changes, humidity, and air density directly affect the efficiency of wind turbines. Changes in air density can cause the same wind speed to carry different amounts of energy. These variations necessitate reliable forecasting methods to ensure the stable operation of energy plants. In this context, machine learning algorithms are used to forecast the power produced by wind turbines, and the performance of these methods is evaluated using various statistical criteria [11]. Ensuring grid security, making energy production planning more predictable, and enhancing economic sustainability are critical aspects where the accuracy of wind energy forecasts plays a vital role.

Wind energy generation occurs when wind turbines convert kinetic energy into electrical energy [12]. The movement of the wind strikes the blades of the turbine, causing them to rotate. This rotational movement is converted into mechanical energy through the rotor and transmitted to the turbine shaft [13]. The mechanical energy transferred from the turbine shaft to the electric generator is converted into electrical energy and is ready for use [14]. The electricity is increased to the appropriate voltage level with the help of transformers and transmitted to the electricity grid after adjusting the frequency [15]. Thus, wind energy is vital in meeting sustainable energy needs, reaching various applications from households to industrial facilities [16].

Wind energy fluctuates depending on meteorological variables [8]. Traditional statistical forecasting methods often fail to model the complex relationships between these variables accurately and may not provide sufficiently accurate predictions [17]. These limitations become more apparent in complex systems like energy markets with dynamic and multidimensional datasets. Traditional approaches typically rely on linear assumptions and fixed parameters, which cannot fully reflect the nonlinear interactions and time-dependent changes between variables [18].

Traditional forecasting methods include physical modeling techniques, statistical analysis methods, and time series models; however, their limitations in accurately modeling the complex variables in the atmosphere have increased the need for new and more advanced approaches [19], [20]. Today, artificial intelligence (AI) and machine learning techniques are increasingly used for more precise evaluation of meteorological data [21]. Deep learning algorithms, big data analytics, and hybrid forecasting models improve the predictability of wind energy production by modeling atmospheric variables more accurately [22]. In this regard, the detailed analysis of weather parameters and their integration into forecasting models significantly improves operational efficiency in the renewable energy sector. Machine learning algorithms, a subset of AI, can process large energy datasets and make more accurate predictions [23]. Therefore, data-driven models in wind power forecasting help make energy production processes more efficient and reliable [24].

Machine learning algorithms enable predicting future events by analyzing large datasets [25]. Used across many sectors, these algorithms are particularly significant for more efficient management of renewable resources in the energy sector [26]. In wind power forecasting, machine learning methods assist in identifying the relationships between variables affecting energy production by processing large datasets [27]. The most commonly used methods for wind power forecasting include Random Forest (RF), *k*-nearest Neighbors (*k*NN), gradient-boosted trees (GBT), and Artificial Neural Networks (ANN) [28].

Studies on machine learning methods in wind power forecasting aim to determine the most suitable methods by comparing the forecasting performances of different algorithms [29]. In wind power forecasting, the use of Genetic Algorithm (GA) has achieved a prediction accuracy of 98% by analyzing

SCADA data and NASA meteorological data collected [30]. In a similar study, Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Decision Trees (DT) algorithms were used for short-term wind power forecasting, with the DT algorithm providing a prediction accuracy of $R^2$=0.92 [31]. In a previous study, Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) models were compared, and the highest accuracy for wind power forecasting was achieved with the LSTM model [32]. In another study, the effect of the placement of wind turbines within the plant on energy production was analyzed, and plant data were evaluated using Random Forest (RF) and Gradient Boosting Descent (GBD) models, and it was found that optimal turbine placement could increase energy production [33].

This study consists of four main sections. The first section provides a literature review on the subject and methodology of the study. The second section includes information about the methods used. The third section discusses the analysis and forecasting results obtained from the proposed methods. Finally, the last section presents the general structure of the study and its contributions to the literature.

## 2. Materials and Methods

This study aims to estimate wind energy production using hourly meteorological data from 2018 in Yalova province. The data set used includes weather parameters. These meteorological variables were included in the model as independent variables (input), while the amount of wind energy production was used as the dependent variable (output). In the pre-processing process of the data, missing and outlier values were detected and corrected with appropriate statistical techniques. Then, all variables were normalized and prepared for the modeling phase. This preparation process was implemented to increase the accuracy of the models and strengthen the learning processes. The variables considered in the study are provided in **Table 1.**

**Table 1.** The features of the dependent and independent variables

| Variable | Types | Unit | Data Type | Data Source |
|---|---|---|---|---|
| Energy Output | Dependent | kW | Continuous | kaggle |
| Wind Speed | Independent | m/sec | Continuous | Open-Meteo |
| Wind Direction | Independent | (°) | Continuous | Open-Meteo |
| Temperature | Independent | $^0$C | Continuous | Open-Meteo |
| Humidity | Independent | RH | Continuous | Open-Meteo |
| Dew Point 2m | Independent | °C | Continuous | Open-Meteo |
| Pressure MSL | Independent | hPa | Continuous | Open-Meteo |
| Cloud Cover | Independent | % | Continuous | Open-Meteo |
| Wind Gusts | Independent | m/s | Continuous | Open-Meteo |
| Precipitation | Independent | mm | Continuous | Open-Meteo |

The primary data source used in this study is the Wind Turbine SCADA Dataset, obtained from the Kaggle platform [34]. This dataset includes operational data on wind turbines. In addition, Open API ) where other datasets are provided, is a weather data source known for its reliability and widespread use [35]. However, there may be deviations in the data supplied by Open-Meteo API. This is because API includes model predictions and meteorological measurements; sometimes, deviations from real-time values may be observed. The following code was used to retrieve data from the Open-Meteo API Archive for this study:

```
import requests
import pandas as pd
import numpy as np

def open_meteo_to_full_climate_data_excel(lat, lon, start_date, end_date, output_file):
    """
    Retrieves climate data from the Open-Meteo API for the specified location (Yalova) and date range,
    performs interpolation at 10-minute intervals, and saves the data to an Excel file.
    This function fetches the following climate parameters:
```

```
    - temperature_2m, relative_humidity_2m,
dew_point_2m, apparent_temperature,
      pressure_msl, cloud_cover, wind_speed_10m,
wind_direction_10m, wind_gusts_10m,
precipitation.
    """

    base_url = "https://archive-api.open-
meteo.com/v1/archive"
    params = {
        "latitude": lat,
        "longitude": lon,
        "start_date": start_date,
        "end_date": end_date,
        "temperature_unit": "celsius",
        "wind_speed_unit": "kmh",
        "timezone": "Europe/Istanbul",
        "hourly": ",".join([
            "temperature_2m",              #
Temperature at 2m height
            "relative_humidity_2m",      #
Relative humidity at 2m height
            "dew_point_2m",                #
Dew point at 2m height
            "apparent_temperature",      #
Apparent temperature
            "pressure_msl",               # Mean
sea level pressure
            "cloud_cover",                # Cloud
cover
            "wind_speed_10m",             #
Wind speed at 10m height
            "wind_direction_10m",        # Wind
direction at 10m height
            "wind_gusts_10m",             #
Wind gusts at 10m height
            "precipitation"              #
Precipitation amount
        ]),
    }

    try:
        # Fetch data from the Open-Meteo API
        response = requests.get(base_url,
params=params)
        response.raise_for_status()
        data = response.json()

        if "hourly" in data:
            hourly_df =
pd.DataFrame(data["hourly"])
            hourly_df["time"] =
pd.to_datetime(hourly_df["time"])

            # Interpolate the data to 10-minute
intervals
            hourly_df.set_index("time",
inplace=True)

            # Resample hourly data to 10-minute
intervals with linear interpolation
            ten_minute_df =
hourly_df.resample('10min').interpolate(method='li
near')

            # Save the data to an Excel file
            ten_minute_df.to_excel(output_file,
index=True)
            print(f"Data has been successfully
written to {output_file}.")
        else:
            print("Hourly data not found.")

    except requests.exceptions.RequestException as
e:
        print(f"Error during API request: {e}")
    except Exception as e:
        print(f"An unexpected error occurred: {e}")

# Yalova coordinates
yalova_lat = 40.6559
yalova_lon = 29.2715

# Define the date range for 2018
start_date = "2018-01-01"
end_date = "2018-12-31"
output_file =
"yalova_2018_full_climate_data_10min.xlsx"

# Call the function to fetch and save the data
open_meteo_to_full_climate_data_excel(yalova_lat,
yalova_lon, start_date, end_date, output_file)
```

In this study, ML algorithms, Random Forest (RF), Gradient Boosting (GB), *k*-nearest neighbor (*k*NN), and Stochastic Gradient Descent (SGD), were used to obtain the amount of energy produced from estimated wind energy. These algorithms were run in the Orange 3.34 Data Mining program to get

performance and prediction data. The workflow diagram of the methodology of the study is given in **Figure 1**.

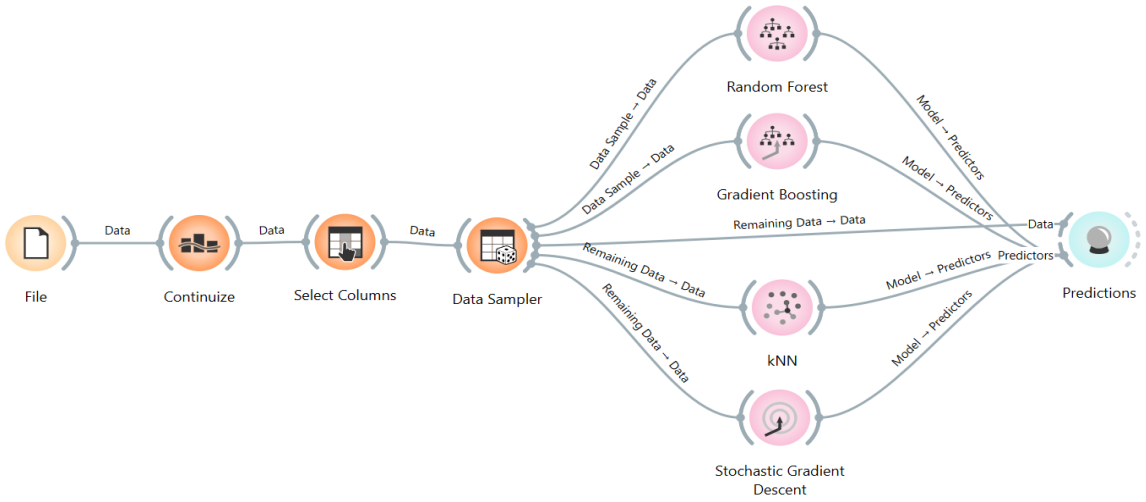



**Figure 1.** The flowchart of the method

Random Forest (RF) is an ensemble learning algorithm combining multiple decision trees. Each decision tree is trained based on a different subset of the dataset, and the results obtained are combined to increase the prediction accuracy [36]. The basic operating principle of the RF algorithm is based on training each decision tree independently by selecting random samples from the dataset and determining random subsets of features [37]. In this process, each decision tree produces its prediction. At the same time, the final result is determined by two different methods: majority voting in classification problems, and the final decision is formed by averaging the predictions in regression problems [38].

The Gradient Boosting (GB) algorithm is an ensemble learning method that focuses on obtaining more accurate predictions by minimizing error rates in the learning process [39]. The GB model aims to create a strong prediction model by sequentially training weak learners (usually decision trees) and performing optimizations to reduce the error rates of previous models at each stage [40]. Each new model improves the learning process by taking into account the errors made by the previous model more and increases the prediction accuracy.

k-nearest neighbor (k-NN) is an unsupervised learning algorithm that estimates a data point's class or value by looking at its nearest neighbors [41]. This algorithm is generally used in classification and regression problems. To determine the class or value of a data point, k-NN examines the labels or values of the nearest neighbors in the specified k number and usually estimates according to the class or average of these neighbors [42]. The user determines the k value, and it is generally necessary to select the most appropriate value using cross-validation. Although k-NN is widely preferred with its simple structure and understandability, the computational cost can be high in large data sets because distance calculations must be made over the entire data set for each estimate [43].

Stochastic gradient descent (SGD) is a widely used optimization algorithm in machine learning [44]. This method updates parameters using only the gradient of one data point at each iteration, to minimize the model's error function (usually the loss function) [45]. While the traditional gradient descent algorithm updates parameters by computing the gradient over the entire training data, SGD calculates only one randomly selected data point at a time. This provides a faster and lighter optimization process but can sometimes experience more fluctuations due to drift and noise. SGD is often preferred in large datasets and deep learning models because each iteration is faster and puts less pressure on the local memory [46]. However, it is important to choose the right learning rate and carefully monitor the parameter updates, as incorrect settings can negatively affect the success of the model.

To compare the performances of the ML algorithms used in the study, the root mean squared error (RMSE), the mean squared error (MSE), the mean absolute error (MAE), and the mean absolute

percentage error (MAPE) criteria, which express the error coefficients, as well as the $R^2$ values were calculated [47, 48]. The following formulas were used to calculate the data belonging to the performance criteria.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_j - \hat{y}_i\right| \quad (3)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (5)$$

The formulas presented are common error metrics used in regression analysis to evaluate the performance of predictive models. In these equations, $y_i$ represents the actual (observed) values, $\hat{y}_i$ represents the predicted values, $\bar{y}$ denotes the mean of the actual values, and $n$ is the total number of observations.

**3. Results**

In this study, wind energy data obtained by considering the effects of environmental factors belonging to Yalova province were estimated with ML algorithms. Descriptive statistics of dependent and independent variables constituting the data set of the study are given in **Table 1**.

**Table 1.** Descriptive statistics of variables

| Variable | N | Mean | StDev | Variance | Min | Max | Skew | Kurt |
|---|---|---|---|---|---|---|---|---|
| Wind Speed (m/s) | 50530 | 75,580 | 42,272,0 | 178,689,0 | 0.00000 | 252,060 | 0.62 | 0.06 |
| Wind Direction (°) | 50530 | 123.69 | 93.4400 | 8731.730 | 0.00000 | 360.000 | 0.70 | -0.75 |
| Energy Produced (MW) | 50530 | 13,077 | 13,125,0 | 17,225,00 | -0.00247 | 36,1870 | 0.60 | -1.16 |
| Temperature 2m (°C) | 50530 | 16,551 | 7,143,00 | 51,016,00 | -0.20000 | 33,6000 | 0.01 | -0.91 |
| Relative Humidity 2m (%) | 50530 | 75,736 | 13,699,0 | 187,657,0 | 23,0000 | 100,000 | -0.69 | -0.13 |
| Dew Point 2m (°C) | 50530 | 11,921 | 6,204,00 | 38,484,00 | -5,80000 | 23,6000 | -0.32 | -0.78 |
| Pressure MSL (hPa) | 50530 | 1014.4 | 6.40000 | 41.00000 | 990.300 | 1033.80 | 0.20 | -0.02 |
| Cloud Cover (%) | 50530 | 51,094 | 39,042,0 | 1,524,277 | 0.00000 | 100,000 | -0.01 | -1.64 |
| Wind Gusts 10m (m/s) | 50530 | 22,851 | 12,055,0 | 145,316,0 | 1,80000 | 79,6000 | 0.71 | 0.03 |
| Precipitation (mm) | 50530 | 0.08851 | 0.33285 | 0.110790 | 0.00000 | 710,000 | 7.47 | 84.84 |

Descriptive statistics for meteorological variables presented in **Table 1** reveal the environmental factors affecting wind energy production in Yalova province in detail. Wind speed and direction variables constitute the essential inputs of wind energy production (MW). The average wind speed is 75.58 m/s and varies between a minimum of zero and a maximum of 252.06 m/s. The high standard deviation (42.27) and variance (178.689) reveal that wind speed varies greatly. The average wind direction is 123.69° and is observed to differ slightly from the normal distribution with a kurtosis value of -0.75. The MW variable reflects the fluctuation in energy production and varies between -0.002 and 36.18 MW. In

addition, meteorological variables such as temperature (Temperature 2m), relative humidity (Relative Humidity 2m), dew point (Dew Point 2m), and atmospheric pressure (Pressure MSL) can also be related to wind energy production. While the average temperature is 16.55°C, the dew point is calculated as 11.92°C. The pressure variable shows a relatively low variability, averaging around 1014.4 hPa. The cloud cover variable varies between 0 and 100 and has a wide distribution due to the high standard deviation (39.04). Wind gusts (10m) have an average of 22.85 m/s and can reach a maximum value of 79.6 m/s. The precipitation amount variable varies between 0 and 710 mm, while the high kurtosis value (84.84) shows that the data distribution contains sharp and extreme values. These data provide an essential basis for analyzing how wind energy production is affected by atmospheric conditions.

Linear regression analysis was performed to measure the statistical effect and strength of the independent variables on the dependent variable. In this study, linear regression analysis was applied to measure the statistical impact and strength of the effect of various independent variables (Temperature, Humidity, Wind Speed, etc.) on a dependent variable. The ANOVA table (**Table 2**) of the linear regression analysis is presented to evaluate how much of the variance each independent variable explains on the dependent variable and whether this explanation is statistically significant. ANOVA is a method used to assess the mean differences between groups statistically. In **Table 2**, the degrees of freedom (DF), corrected sum of squares (Adj SS), corrected mean squares (Adj MS), F value, and P value of each independent variable are presented. While P values show the statistical significance of the effect of independent variables on the dependent variable, F values and Adj SS values show the effect size. **Figure 2** shows the power of influence of independent variables on the dependent variable. The variable with the most impact on the dependent variable has been determined as Wind Speed (m/s).

**Table 2.** ANOVA analysis

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Temperature at 2m (°C) | 1 | 13.4000 | 13.4000 | 47.3900 | 0.001 |
| Relative Humidity at 2m (%) | 1 | 8.30000 | 8.30000 | 29.3000 | 0.000 |
| Dew Point at 2m (°C) | 1 | 12.6000 | 12.6000 | 44.5800 | 0.001 |
| Mean Sea Level Pressure (hPa) | 1 | 62.6000 | 62.6000 | 221.310 | 0.001 |
| Cloud Cover (%) | 1 | 1.80000 | 1.80000 | 6.46000 | 0.011 |
| Wind Gusts at 10m (m/s) | 1 | 94.6000 | 94.6000 | 334.400 | 0.002 |
| Precipitation (mm) | 1 | 13.2000 | 13.2000 | 46.6500 | 0.003 |
| Wind Speed (m/s) | 1 | 72157.9 | 72157.9 | 255115 | 0.002 |
| Wind Direction (°) | 1 | 4.90000 | 4.90000 | 17.4500 | 0.001 |

ANOVA analysis statistically reveals the effect of various meteorological variables on a specific dependent variable. All analyzed variables were statistically significant, as all p-values were less than 0.05. In particular, variables such as temperature (p=0.001), relative humidity (p=0.000), dew point (p=0.001), sea level pressure (p=0.001), and wind direction (p=0.001) all have high F-values and strongly reveal their effects with very low p-values. For example, sea level pressure stands out with its adjusted mean square value of 62.6 and F-value of 221.31. At the same time, wind direction has a relatively lower effect (F=17.45) but still has a statistically significant effect. This situation reveals that many components related to weather conditions directly and significantly affect the dependent variable in the system.

As a result of the analysis, the wind speed variable stands out with a corrected mean square value of 72,157.9 and an extraordinarily high F-value of 255,115, making it the most effective factor on the dependent variable. This is followed by wind intensity with a corrected mean square value of 94.6 and an F-value of 334.4. In addition, factors such as precipitation amount (F=46.65, p=0.003) and cloudiness rate (F=6.46, p=0.011) also have significant effects despite lower F-values. This shows that the system is affected by thermodynamic and atmospheric variables and that complex interactions between meteorological parameters should be considered. In conclusion, this ANOVA table powerfully reveals the effect of meteorological variables at individual and comparative levels and provides important clues about which factors should be prioritized in modeling.

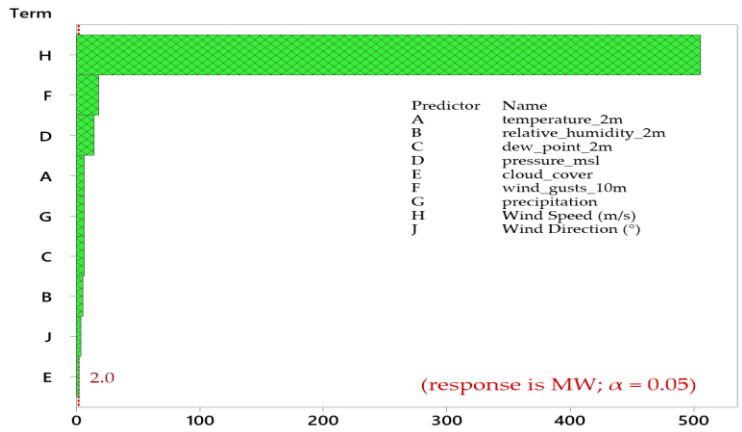



**Figure 2.** Pareto chart of standardized effects of independent variables

The Pareto chart presented in **Figure 2** shows the standardized effects of the independent variables on the dependent variable (MW). According to the chart, the Wind Speed (m/s) (H) variable has the most potent effect on the dependent variable, while the impact of the Cloud Cover (E) variable is relatively low. This contradicts the claim in the text that the "Cloud Cover" variable is the most effective. The chart also shows that the variables Precipitation (F), Sea Level Pressure (D), Temperature at 2m (A), and Dew Point at 2m (G) have significant effects on the dependent variable. The effects of the variables Relative Humidity (B) at 2m and Wind Direction (°) (J) are at the lowest level. The expression $\alpha = 0.05$ in the chart indicates the level of statistical significance. This is a threshold value used to evaluate whether the length of the bars is statistically significant. If the bar length exceeds this threshold value, the effect of the relevant independent variable is statistically significant. This analysis shows that the variables with the most substantial impact on the dependent variable are Wind Speed and Rainfall. The performance measurement criteria of ML algorithms, RMSE, MSE, MAE, MAPE, and $R^2$ data, are given in **Table 3.**

**Table 3.** Data on performance measurement criteria of ML algorithms

| Model | MSE | RMSE | MAE | MAPE | $R^2$ |
|---|---|---|---|---|---|
| RF | 0.039 | 0.197 | 0.081 | 0.377 | 0.961 |
| GB | 0.071 | 0.266 | 0.117 | 0.421 | 0.928 |
| *k*NN | 0.039 | 0.198 | 0.111 | 0.689 | 0.960 |
| SGD | 0.175 | 0.418 | 0.303 | 0.581 | 0.822 |

**Table 3** presents the data for RMSE, MSE, MAE, MAPE, and $R^2$ metrics used to evaluate the performance of machine learning algorithms. According to the results in the table, the Random Forests (RF) algorithm shows the best performance with the lowest MSE (0.039), RMSE (0.197), and MAE (0.081) values. This indicates that the RF algorithm makes fewer prediction errors and produces results closer to the actual values. In addition, the $R^2$ value of RF is relatively high (0.961), indicating that the model can explain a large portion of the variance in the data. Although the *k*NN algorithm has similar MSE (0.039) and RMSE (0.198) values to RF, its MAE value is higher than RF (0.111). The Gradient Boosting (GB) algorithm has higher MSE (0.071), RMSE (0.266), and MAE (0.117) values than RF and K-NN, indicating that GB performs worse than the other two algorithms.

SGD algorithm has the highest MSE (0.175), RMSE (0.418), and MAE (0.303) values compared to the other algorithms, indicating that SGD performs the worst. In addition, SGD's $R^2$ value is the lowest (0.822) compared to the other algorithms, indicating that the model cannot explain the variance in the data. The results in **Table 3** reveal that RF and K-NN algorithms perform better than the other algorithms for this dataset. However, which algorithm is most suitable may vary depending on the specific requirements and priorities of the application. For example, when the MAPE value is taken into account, it is seen that RF (0.377) performs better than GB (0.421). This means that the percentage of errors for RF is lower than that of GB.

## 4. Conclusion

This study aims to increase predictability in renewable energy production by investigating the effectiveness of machine learning (ML) algorithms in wind energy forecasting in Yalova province. Comparisons between different ML models revealed that Random Forest (RF) and Gradient Boosting algorithms provide high accuracy rates. The effect of meteorological variables (e.g., temperature, humidity, pressure, and wind speed) used in the study on wind power production was analyzed in detail, and it was seen that wind speed was the most dominant parameter in forecast performance. It was shown that even in regions with moderate wind potential, such as Yalova, satisfactory forecast results can be achieved with appropriate modeling and data analysis. These findings emphasize the importance of environmental variables and data-driven approaches in energy production planning. As a result, this study constitutes an example that will contribute to renewable energy management at the local level and shows the applicability of similar methodologies in different geographical regions. The results prove that wind energy forecasts should be addressed from a multi-dimensional perspective, not only with technical data but also by considering environmental and statistical components. In future studies, it may be possible to increase the prediction accuracy by using longer data sets and integrating more advanced ML and deep learning methods. In addition, integrating these models in energy supply-demand management with real-time systems will play a strategic role in implementing sustainable energy policies.

**Author Contributions:** The conceptualization of the study was carried out by A.A. and Y.A.A. Data analysis and methodology development were conducted by L.A.G. and H.K. The original draft was prepared by A.A. and Y.A.A. Review, editing, and final revisions were performed by İ.Ç. and Y.A.A. All authors have approved the final version of the manuscript and share equal responsibility for the content.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## References

[1] N. Abas, A. Kalair, and N. Khan, "Review of fossil fuels and future energy technologies," *Futures*, vol. 69, pp. 31–49, 2015.

[2] P. Wilkinson, K. R. Smith, M. Joffe, and A. Haines, "A global perspective on energy: health effects and injustices," *Lancet*, vol. 370, no. 9591, pp. 965–978, 2007.

[3] O. Ellabban, H. Abu-Rub, and F. Blaabjerg, "Renewable energy resources: Current status, future prospects and their enabling technology," *Renew. Sustain. energy Rev.*, vol. 39, pp. 748–764, 2014.

[4] A. Rahman, O. Farrok, and M. M. Haque, "Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic," *Renew. Sustain. energy Rev.*, vol. 161,
p. 112279, 2022.

[5] A. Atalan and Y. A. Atalan, "Nonlinear Optimization Models of Box-Behnken Experimental Design: Turbine Simulation for Wind Power Plant," *3rd International Conference on Engineering and Applied Natural Sciences*, 2023.

[6] A. D. Şahin, "Progress and recent trends in wind energy," *Prog. energy Combust. Sci.*, vol. 30, no. 5, pp. 501–543, 2004.

[7] E. Toklu, "Overview of potential and utilization of renewable energy sources in Turkey," *Renew. Energy*, vol. 50, pp. 456–463, 2013, doi: 10.1016/j.renene.2012.06.035.

[8] F. Cassola and M. Burlando, "Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output," *Appl. Energy*, vol. 99, pp. 154–166, 2012.

[9] I. Sanchez, "Short-term prediction of wind energy production," *Int. J. Forecast.*, vol. 22, no. 1, pp. 43–56, 2006.

[10] S.-Q. Dotse, I. Larbi, A. M. Limantol, and L. C. De Silva, "A review of the application of hybrid machine learning models to improve rainfall prediction," *Model. Earth Syst. Environ.*, vol. 10, no. 1, pp. 19–44, 2024.

[11] A. Alkesaiberi, F. Harrou, and Y. Sun, "Efficient wind power prediction using machine learning methods: A comparative study," *Energies*, vol. 15, no. 7, p. 2327, 2022.

[12] D. Wang, X. Gao, K. Meng, J. Qiu, L. L. Lai, and S. Gao, "Utilisation of kinetic energy from wind turbine for grid connections: a review paper," *IET Renew. Power Gener.*, vol. 12, no. 6, pp. 615–624, 2018.

[13] J. W. Zhou, W. Zhang, X. Jiang, and E. D. Zhai, "Investigation on dynamics of rotating wind turbine blade using transferred differential transformation method," *Renew. Energy*, vol. 188, pp. 96–113, 2022.

[14] C.-T. Chen, R. A. Islam, and S. Priya, "Electric energy generator," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 53, no. 3, pp. 656–661, 2006.

[15] E. Serban, M. Ordonez, and C. Pondiche, "Voltage and frequency grid support strategies beyond standards," *IEEE Trans. power Electron.*, vol. 32, no. 1, pp. 298–309, 2016.

[16] J. B. Welch and A. Venkateswaran, "The dual sustainability of wind energy," *Renew. Sustain. Energy Rev.*, vol. 13, no. 5, pp. 1121–1126, 2009.

[17] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns

and ways forward," *PLoS One*, vol. 13, no. 3, p. e0194889, 2018.
[18] P. Reichert and J. Mieleitner, "Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters," *Water Resour. Res.*, vol. 45, no. 10, 2009.
[19] Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, and A. Baklanov, "Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects," *Atmos. Environ.*, vol. 60, pp. 656–676, 2012.
[20] A. Atalan, "The ChatGPT application on quality management: a comprehensive review," *J. Manag. Anal.*, pp. 1–31, May 2025, doi: 10.1080/23270012.2025.2484225.
[21] S. Dewitte, J. P. Cornelis, R. Müller, and A. Munteanu, "Artificial intelligence revolutionises weather forecast, climate monitoring and decadal prediction," *Remote Sens.*, vol. 13, no. 16, p. 3209, 2021.
[22] J. Devaraj, R. Madurai Elavarasan, G. M. Shafiullah, T. Jamal, and I. Khan, "A holistic review on energy forecasting using big data and deep learning models," *Int. J. energy Res.*, vol. 45, no. 9, pp. 13489–13530, 2021.
[23] Y. A. Atalan and A. Atalan, "Integration of the Machine Learning Algorithms and I-MR Statistical Process Control for Solar Energy," *Sustain.*, vol. 15, no. 18, Sep. 2023, doi: 10.3390/su151813782.
[24] E. T. Renani, M. F. M. Elias, and N. A. Rahim, "Using data-driven approach for wind power prediction: A comparative study," *Energy Convers. Manag.*, vol. 118, pp. 193–203, 2016.
[25] A. Atalan and C. Ç. Dönmez, "Dynamic Price Application to Prevent Financial Losses to Hospitals Based on Machine Learning Algorithms," *Healthcare*, vol. 12, no. 13, p. 1272, Jun. 2024, doi: 10.3390/healthcare12131272.
[26] H. Szczepaniuk and E. K. Szczepaniuk, "Applications of artificial intelligence algorithms in the energy sector," *Energies*, vol. 16, no. 1, p. 347, 2022.
[27] H. Demolli, A. S. Dokuz, A. Ecemis, and M. Gokcek, "Wind power forecasting based on daily wind speed data using machine learning algorithms," *Energy Convers. Manag.*, vol. 198, p. 111823, 2019, doi: https://doi.org/10.1016/j.enconman.2019.111823.
[28] P. Piotrowski, D. Baczyński, M. Kopyt, and T. Gulczyński, "Advanced ensemble methods using machine learning and deep learning for one-day-ahead forecasts of electric energy production in wind farms," *Energies*, vol. 15, no. 4, p. 1252, 2022.
[29] L. Wang, X. Zhou, X. Zhu, Z. Dong, and W. Guo, "Estimation of biomass in wheat using random forest regression algorithm and remote sensing data," *Crop J.*, vol. 4, no. 3, pp. 212–219, Jun. 2016, doi: 10.1016/j.cj.2016.01.008.
[30] P. Moshtaghi, N. Hajialigol, and B. Rafiei, "A Comprehensive Review of Artificial Intelligence Applications in Wind Energy Power Generation," *Available SSRN 5061006*, 2024.
[31] K. Yazıcı, "Makine öğrenmesi yöntemleri kullanılarak kısa dönem rüzgar gücü tahmini." Sakarya Üniversitesi, 2021.
[32] Y. Qin *et al.*, "Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal," *Appl. Energy*, vol. 236, pp. 262–272, 2019.
[33] U. Singh, M. Rizwan, M. Alaraj, and I. Alsaidan, "A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments," *Energies*, vol. 14, no. 16, p. 5196, 2021.
[34] B. Erisen, "Wind Turbine Scada Dataset 2018 Scada Data of a Wind Turbine in Turkey," *Kaggle*, 2018. https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset
[35] Open-meteo, "Accurate Weather Forecasts for Any Location," 2025. https://archive-api.open-meteo.com/v1/archive
[36] A. Atalan, "Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms," *Agribusiness*, vol. 39, no. 1, pp. 214–241, Jan. 2023, doi: 10.1002/agr.21773.
[37] D. Thakur and S. Biswas, "Permutation importance based modified guided regularized random forest in human activity recognition with smartphone," *Eng. Appl. Artif. Intell.*, vol. 129, p. 107681, 2024, doi: https://doi.org/10.1016/j.engappai.2023.107681.
[38] H. İnaç, Y. E. Ayözen, A. Atalan, and C. Ç. Dönmez, "Estimation of Postal Service Delivery Time and Energy Cost with E-Scooter by Machine Learning Algorithms," *Appl. Sci.*, vol. 12, no. 23, p. 12266, Nov. 2022, doi: 10.3390/app122312266.
[39] W. Khan, S. Walker, and W. Zeiler, "Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach," *Energy*, vol. 240, p. 122812, 2022, doi: https://doi.org/10.1016/j.energy.2021.122812.
[40] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, R. Tibshirani, and J. Friedman, "Boosting and additive trees," *Elem. Stat. Learn. data mining, inference, Predict.*, pp. 337–387, 2009.
[41] G. McKenzie and D. Romm, "Measuring urban regional similarity through mobility signatures," *Comput. Environ. Urban Syst.*, vol. 89, p. 101684, Sep. 2021, doi: 10.1016/j.compenvurbsys.2021.101684.
[42] Z. Liu, Q. Pan, and J. Dezert, "A new belief-based K-nearest neighbor classification method," *Pattern Recognit.*, vol. 46, no. 3, pp. 834–844, 2013.
[43] J. Huang *et al.*, "Cross-validation based K nearest neighbor imputation for software quality datasets: an empirical study," *J. Syst. Softw.*, vol. 132, pp. 226–252, 2017.
[44] Y. Tian, Y. Zhang, and H. Zhang, "Recent advances in stochastic gradient descent in deep learning," *Mathematics*, vol. 11, no. 3, p. 682, 2023.
[45] J. Yang and G. Yang, "Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer," *Algorithms*, vol. 11, no. 3, p. 28, 2018.
[46] S. U. Stich, J.-B. Cordonnier, and M. Jaggi, "Sparsified SGD with memory," *Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.
[47] A. Keskin, "Türkiye enerji piyasasında piyasa takas fiyatı tahmini: Makine öğrenimi yöntemlerinin karşılaştırılması," *Üçüncü Sektör Sosyal Ekonomi Dergisi*, vol. 60, no. 2, pp. 1707–1719, 2025, doi: 10.63556/tisej.2025.1509.
[48] Y. Ayaz Atalan and A. Atalan, "Testing the Wind Energy Data Based on Environmental Factors Predicted by Machine Learning with Analysis of Variance," *Appl. Sci.*, vol. 15, no. 1, p. 241, Dec. 2024, doi: 10.3390/app15010241.

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

*Research Article*

Received: date:28.04.2025
Accepted: date:05.06.2025
Published: date:30.06.2025

# İlaç Üretim Sürecinde Fire Oranlarinin İzlenmesi: İstatistiksel Süreç Analizi

**Metin Berk Çetin [1], Beyza Özkan [2], Yavuz Özdemir [3], Mustafa Yıldırım [4,*]**

[1]Department of Industrial Engineering, Istanbul Health and Technology University, Istanbul, Türkiye; c.metinberk@hotmail.com
[2]Department of Industrial Engineering, Istanbul Health and Technology University, Istanbul, Türkiye; beyzaaozkan2003@gmail.com
[3]Department of Industrial Engineering, Istanbul Health and Technology University, Istanbul, Türkiye; yavuz.ozdemir@istun.edu.tr
[4]Department of Industrial Engineering, Istanbul Health and Technology University, Istanbul, Türkiye; mustafa.yildirim@istun.edu.tr
Orcid: 0009-0006-5935-4091[1], Orcid: 0009-0000-8709-4596[2] , Orcid: 0000-0001-6821- 9867[3], Orcid: 0000-0001-5709-4421[4]
*Correspondence: mustafa.yildirim@istun.edu.tr

**Öz:** Bu çalışmada, bir ilaç üretim hattındaki süreç kararlılığı ve fire oranlarının analizi hedeflenmiştir. İlk olarak, p diyagramı kullanılarak süreç kararlılığı değerlendirilmiş ve sürecin kontrol dışı olduğu belirlenmiştir. Daha sonra, farklı vardiyalardaki fire oranlarının incelenmesi amacıyla ANOVA analizi yapılmıştır. Analiz sonuçları, özellikle 4. vardiyalarda fire oranlarının istatistiksel olarak anlamlı şekilde daha yüksek olduğunu göstermiştir. Bu durumun nedenlerini anlamak için Kök Neden Analizi uygulanmış ve 4. vardiyalardaki yüksek fire oranlarının makinelerde biriken toz sızdırmalarından kaynaklandığı tespit edilmiştir. Bu sorun, makine temizliğinin yetersiz sıklıkta yapılmasıyla ilişkilendirilmiştir. Çalışma, süreç verimliliğini artırmaya yönelik çözüm önerileri geliştirerek, kalite kontrol ve süreç yönetimi alanlarında katkı sağlamaktadır.
**Anahtar Kelimeler:** süreç kararlılığı, p diyagramı, ANOVA, kök neden analizi, süreç yönetimi

# Monitoring Defect Rates in the Pharmaceutical Production Process: A Statistical Process Analysis

**Abstract:** This study aims to analyze process stability and defect rates in a pharmaceutical production line. First, process stability was evaluated using a p-chart, which indicated that the process was out of control. Then, an ANOVA analysis was conducted to examine defect rates across different shifts. The results showed that defect rates were statistically significantly higher, particularly in the 4th shifts. To understand the root causes of this issue, a Root Cause Analysis was performed, revealing that the high defect rates in the 4th shifts were due to dust accumulation in the machines. This problem was linked to the insufficient frequency of machine cleaning. The study contributes to quality control and process management by providing solutions to improve process efficiency.
**Keywords:** process stability, p-chart; ANOVA, root cause analysis, process management

## 1. Giriş

Günümüzün yoğun rekabet ortamında işletmelerin sektörde tutunabilmesi ve rekabet avantajı elde edebilmesi için maliyetleri en düşük seviyede tutarak, yüksek kaliteli ürünleri müşterilere zamanında ulaştırması gerekmektedir [1]. Firmalar arası rekabetin artmasıyla birlikte işletmeler için verimlilik, kalite ve maliyet yönetimi hayati bir öneme sahip hale gelmiştir. Üretim süreçlerinde ortaya çıkan kayıpların azaltılması, işletmelerin sürdürülebilirliği ve kârlılığı açısından kritik bir faktördür [2]. Çoğu organizasyon, müşteriye birimden birime her zaman aynı olan veya müşteri beklentilerini karşılayan seviyelerde kalite özelliklerine sahip ürünler sunmayı zor (ve pahalı) bulur. Bunun başlıca nedeni değişkenliktir. Her üründe belirli bir miktarda değişkenlik bulunur; dolayısıyla hiçbir iki ürün tamamen aynı değildir [3]. Fire oranlarının yüksek olması, yalnızca işletmelerin maliyetlerini artırmakla kalmaz, aynı zamanda süreçlerin etkinliğini ve ürün kalitesini doğrudan etkileyerek müşteri memnuniyetini azaltır. Bu durum, özellikle yoğun rekabetin olduğu sektörlerde daha büyük bir risk oluşturur. Kalite ve verimlilik dengesini sağlayabilmek, işletmelerin uzun vadede sürdürülebilir bir başarı yakalayabilmesi

**Citation :** M. B. Çetin, B. Özkan, Y. Özdemir, and M. Yıldırım, "İlaç üretim sürecinde fire oranlarının izlenmesi: İstatistiksel süreç analizi," Journal of Statistics and Applied Sciences, no. 11, pp. 50–58, 2025, doi: 10.52693/jsas.1681851.

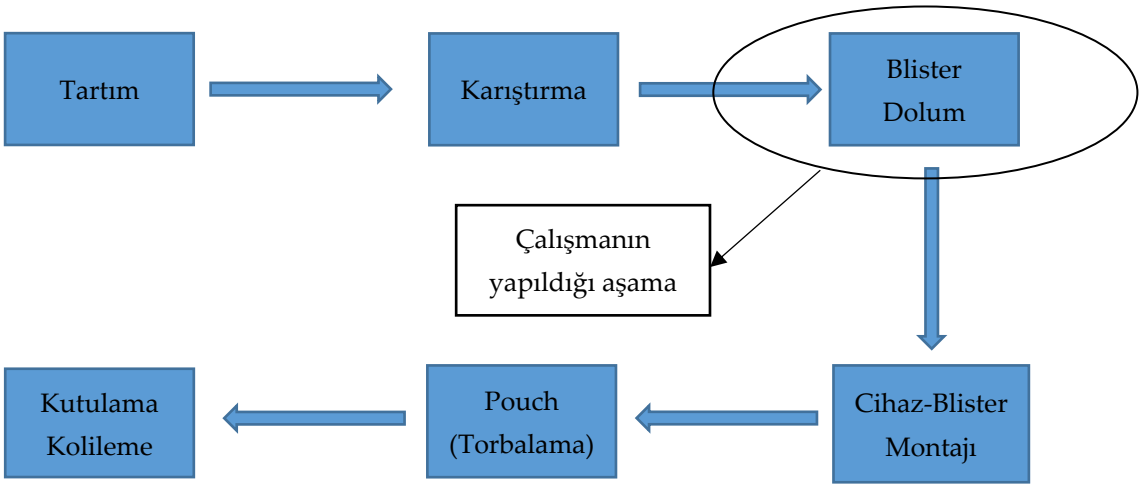
için kritik bir öneme sahiptir. İstatistiksel Proses Kontrol (İPK) ve ilgili analiz yöntemleri, bu hedefe ulaşmada işletmelere güçlü araçlar sunarak süreçlerin potansiyelini tam olarak kullanmalarına olanak tanır. İstatistiksel Proses Kontrolü (İPK), bir prosesin izlenmesi ve kontrolü için istatistiksel yöntemlerin uygulanmasıdır ve bu prosesin tam potansiyeliyle çalışmasını ve uyumlu bir ürün üretmesini sağlar. İstatistiksel Proses Kontrolü (İPK), süreç kabiliyetini iyileştirmek için kontrol grafikleri, bireysel ve alt grup ölçümleri ve nitelik verileri kullanarak süreçleri yönetmek ve izlemek için kapsamlı bir yaklaşımdır [4]. İPK kapsamında bir proses, mümkün olan en az atıkla mümkün olduğunca fazla uyumlu ürün üretmek için öngörülebilir şekilde davranır [5]. Kontrol grafikleri üretim süreçlerinin kararlılığını ölçmek ve süreçteki değişkenlikleri izlemek için önemli bir araçtır. Bu grafikler, bir sürecin istatistiksel kontrol altında olup olmadığını belirlemek için kullanılır. Kontrol sınırları dahilinde çalışan bir süreç, genellikle öngörülebilir sonuçlar üretir ve bu durum süreç kararlılığının göstergesidir. Ancak kontrol sınırlarının dışında meydana gelen dalgalanmalar, süreçteki potansiyel bir problemin işareti olabilir ve bu tür sorunların hızlı bir şekilde ele alınması gerekir. Bir süreç kontrolden çıktığında, mühendislerin atanabilir varyasyon nedenlerini belirleyip bunları ortadan kaldırmaya çalışabilmesi için bir uyarı mekanizması devreye girer. Ancak, süreçteki kontrol dışı durumların oluşmasını önlemek için proaktif bir yaklaşım benimsemek daha etkili bir yöntemdir. Bu sayede, süreç önleyici bir şekilde optimize edilerek daha az uygunsuz ürün üretilmesi sağlanabilir [6]. Modern üretim sistemlerinde, üretim süreçlerinin verimliliğini artırmak ve hataları en aza indirmek önemli bir hedef haline gelmiştir. Bu çerçevede, kalite kontrol teknikleri ve istatistiksel analiz yöntemleri, süreç iyileştirme çalışmalarında yaygın olarak kullanılmaktadır. Özellikle Süreç Kontrol Diyagramları ve ANOVA gibi yöntemler, üretim hatalarının kök nedenlerini belirlemek ve süreçleri geliştirmek için etkili araçlar sunmaktadır.

Literatürde, bu tekniklerin yüksek hassasiyet gerektiren alanlarda başarılı bir şekilde uygulandığına dair birçok çalışma mevcuttur. Ayrıca, belirli üretim süreçlerindeki verimlilik kayıplarını inceleyen vaka çalışmaları, hem teorik bilgi birikimine hem de sektörel uygulamalara önemli katkılar sağlamaktadır. Aichouni vd.'nin 2012 yılında yayınladıkları makale, süreç iyileştirme konusundaki temel kavramların bir incelemesini yapmaya ve bu kavramların inşaat firmalarının iyileştirme projelerinde uygulanmasının faydalarını göstermeye adanmıştır. Analiz, beton üreticilerinin üretim süreçlerini etkili bir şekilde iyileştirebileceklerini, paradan ve malzemeden tasarruf edebileceklerini ve süreçlerini sürdürülebilir hale getirebileceklerini açıkça göstermektedir [7]. Yıldırım vd. 2013 yılında yaptıkları bu çalışmada, elektronik sektöründe yangın ve gaz algılama sistemlerinin üretimini yapan bir işletmede İstatistiksel Proses Kontrol tekniklerinin uygulaması yapılmıştır. Prosesin kontrol altında olma durumunun belirlenmesi için kalite kontrol kısmından alınan veriler ile kontrol diyagramları oluşturulmuştur. Süreçte geliştirme çalışmaları daha uyumlu, az toleransa sahip alıcı malzeme kullanılması, devrenin bu şekilde tasarlanması ile yapılabilir olduğu sonucuna ulaşılmıştır. Duman yoğunluğu miktarının ölçüldüğü, rüzgar tünelinin bakımlarının aksamadan yapılması, cihazların kalibrasyon ve doğrulamalarının düzenli takip edilmesinin faydalı olacağı belirtilmiştir [8]. Abtew vd.'nin 2018'de yaptıkları araştırmanın temel amacı, hazır giyim endüstrisinin dikim katındaki kalite özellikleri için uygun İstatistiksel Proses Kontrol (İPK) tekniklerini uygulamaktır. Meydana gelen özel nedenli değişimler ve proses istikrarı için düzeltici eylem planları geliştirilmiştir. Proje, İPK kavramını ve uygulama prosedürünü anlamak için farklı kalite ekibi üyeleri için teorik ve iş başında eğitim şemalarını içermektedir. Uygulamadan sonra, dikim bölümünde önemli iyileştirmeler elde edilmiştir. İPK araçlarının uygulanmasından önce ve sonra yapılan dört aylık analiz, reddetme oranının %9,141'den %6,4'e düştüğünü göstermiştir [9]. Simatupang vd. 2024 yılında yaptıkları çalışmada Anova analizi kullanılarak palmiye çekirdeği yağı üretimi yapan bir işletmede yağ kalitesini etkileyen varyansı belirlemeyi amaçlamışlardır. Çalışmanın sonuçları, serbest yağ asitleri, nem içeriği ve safsızlık içeriği için varyans analizi Fcount değerlerinin Ftable değerlerinden daha küçük olduğunu, bu da bu dört veri grubunun ölçüm sonuçlarının veya ortalama değerlerinin istatistiksel olarak anlamlı farklılıklara sahip olmadığını göstermiştir [10]. Flavius A. Ardelean 2017 yılında gerçekleştirdiği çalışmasında Altı Sigma Tasarımı ANOVA (Varyans Analizi) yöntemini kullanarak, belirli bir karakteristik boyuta sahip üç vardiyada üretilen bir parçanın analizini sunmayı amaçlamıştır. Çalışmanın sonucunda gözlemlenen varyansa göre süreci iyileştirme adına çalışmalar yapılması gerektiği vurgulanmıştır [11]. Paese vd. 2001 yılında yaptıkları çalışmada Rio Grande do Sul'da

bulunan bir meta/mekanik tesisinde çelik çubukların laminasyon sürecinde gözlemlenen değişkenlik kaynaklarını belirlemek için kullanılan Varyans Analizi'nin (ANOVA) bir uygulamasını sunmaktadır. Belirlenen değişkenliğin ana kaynağının çelik türü olduğu saptanmıştır. Değişkenlik kaynaklarının incelenmesi, geleneksel kararlılık ve kapasite çalışmalarıyla birlikte açıklanan İstatistiksel Süreç Kontrolünün kurulması için bir temel görevi görmüştür [12].

Modern üretim sistemlerinde fire oranlarının düşürülmesine yönelik çalışmalar, yalnızca sorunları tespit etmekle kalmayıp, aynı zamanda bu sorunların kök nedenlerini anlamaya yönelik yöntemlerin geliştirilmesiyle desteklenmektedir. Bu bağlamda, istatistiksel süreç kontrolü, kalite yönetimi ve veri analitiği gibi yöntemler, üretim süreçlerindeki problemleri anlamak ve iyileştirme fırsatlarını belirlemek için kritik araçlar haline gelmiştir. Özellikle istatistiksel yöntemler, süreçlerin kararlılığını ölçmek, problemleri önceliklendirmek ve süreç iyileştirme kararlarını daha bilinçli bir şekilde alabilmek için büyük bir avantaj sunar. Vardiyalı çalışma düzenine sahip üretim tesislerinde, farklı vardiyaların performansı da üretim süreçlerinin genel verimliliği üzerinde önemli bir etkiye sahiptir. Özellikle gece vardiyalarında yorgunluk, dikkat eksikliği ve motivasyon düşüklüğü gibi faktörler, kötü ürün oranlarının artmasına neden olabilir. Vardiyalar arasındaki bu performans farklılıklarının doğru bir şekilde analiz edilmesi, üretim süreçlerinde adaletli bir iş yükü dağılımı yapılmasını ve süreç iyileştirme çalışmalarının daha verimli hale getirilmesini sağlar. Varyans analizi (ANOVA), iki farklı grup ortalamaları arasında fark olup olmadığını varyans kullanarak araştıran istatistiksel bir yaklaşımdır. İki grubun karşılaştırılmasında kullanılan farklı yöntemler olsa da (Z veya T testi) en yaygın olanı F testi yani Varyans Analizi (ANOVA, Analysis of Variance) dir. İkiden fazla grubun önem seviyesi değerlendirilmek istediğinde aralarındaki farklılık sadece Varyans Analizi ile yapılabilmektedir [13]. Fire oranlarını azaltmaya yönelik çalışmalar, yalnızca mevcut sorunları çözmekle sınırlı kalmamalı, aynı zamanda uzun vadeli bir kalite kültürü oluşturmayı hedeflemelidir. Bu bağlamda, sürekli iyileştirme anlayışı, işletmelerin üretim süreçlerinde daha yüksek bir kalite standardına ulaşmalarını sağlarken, aynı zamanda maliyetlerin kontrol altına alınmasına da olanak tanır. Sürekli iyileştirme yaklaşımı, hem teknik iyileştirmeleri hem de çalışanların süreçlere daha aktif bir şekilde katılmasını teşvik eden bir liderlik anlayışını içerir.

Bu çalışmada, bir ilaç üretim hattındaki süreç kararlılığı ve fire oranları incelenerek üretim verimliliğini artırmaya yönelik iyileştirme fırsatları belirlenmesi amaçlanmıştır. Üretim süreci; Tartım, Karıştırma, Blister Dolum, Cihaz-Blister Montajı, Pouch (Torbalama) ve Kutulama/Kolileme aşamalarından oluşmaktadır. Şema 1'de görüldüğü üzere çalışma, özellikle Blister Dolum aşamasına odaklanarak gerçekleştirilmiştir. Yapılan saha gözlemleri ve üretim mühendislerinin geri bildirimleri doğrultusunda, üretim sürecinde en fazla fire oranının blisterleme aşamasında gerçekleştiği bildirilmiştir. Bu aşamada meydana gelen teknik aksaklıklar ve ürün yerleşim hataları, hatalı ürün sayısında ciddi artışlara neden olmaktadır. Dolayısıyla, blisterleme süreci üretim verimliliği açısından kritik bir nokta olarak öne çıkmakta ve iyileştirme çalışmaları için öncelikli müdahale alanı olarak değerlendirilmektedir. İlk olarak, süreç istikrarını değerlendirmek için p diyagramı kullanılmış ve sürecin kontrol dışında olduğu tespit edilmiştir. Bu durumun olası nedenlerini anlamak adına, farklı vardiyalardaki fire oranları arasında istatistiksel olarak anlamlı bir fark olup olmadığını belirlemek için ANOVA analizi gerçekleştirilmiştir. Analiz sonuçları, özellikle 4. vardiyalarda fire oranlarının belirgin şekilde yüksek olduğunu göstermiştir. Bunun nedenlerini daha ayrıntılı incelemek amacıyla Kök Neden Analizi uygulanmış ve temel sorunun, makinelerde biriken toz sızdırmalarından kaynaklandığı belirlenmiştir. Bu sorunun, makine temizliğinin belirlenen sıklıkta yapılmamasından kaynaklandığı anlaşılmıştır. Elde edilen bulgular doğrultusunda, üretim sürecinin verimliliğini artırmak ve kalite kontrol süreçlerini iyileştirmek için çeşitli öneriler geliştirilmiştir. Çalışma, ilaç üretim süreçlerinde kalite güvencesini artırmaya ve üretim kayıplarını azaltmaya yönelik bilimsel ve pratik bir katkı sağlamayı hedeflemektedir.




**Şema 1.** Çalışmanın yapıldığı aşamayı gösteren akış şeması

## 2. Materyal ve Yöntemler

Bu çalışmada, bir ilaç üretim hattındaki süreç kararlılığını ve fire oranlarını analiz etmek için nicel yöntemler kullanılmıştır. İlk olarak, süreç kararlılığını değerlendirmek amacıyla 30 partiden elde edilen fire oranları kullanılarak p diyagramı oluşturulmuş ve sürecin kontrol dışı olduğu tespit edilmiştir. Daha sonra, vardiyalar arasındaki fire oranı farklılıklarını incelemek ve bu farklılıkların istatistiksel olarak anlamlı olup olmadığını belirlemek için ANOVA analizi gerçekleştirilmiştir. Elde edilen sonuçlar doğrultusunda, 4. vardiyalarda fire oranlarının diğer vardiyalara kıyasla daha yüksek olduğu belirlenmiş ve bu durumun kök nedenleri Kök Neden Analizi yöntemiyle analiz edilmiştir. Bu kapsamda, mevcut ara temizlik uygulamaları; yalnızca ürün parametrelerinde ciddi sapmalar meydana geldiğinde, operatörler tarafından ve sistematik olmayan bir şekilde yapılmaktadır. Temizlik sıklığı üretim planlamasını aksatmamak amacıyla sabit tutulmakta, ancak bu yaklaşım filtrelerdeki toz birikimini önlemeye yetmemektedir. Bu metodolojik yaklaşım, üretim hattındaki temel problemleri tespit etmek ve çözüm önerileri geliştirmek için bütüncül bir bakış açısı sağlamıştır.

Kontrol şeması, bir süreci zaman içinde ölçmek, izlemek ve kontrol etmek için kullanılan, kalite kontrolde yaygın olarak kullanılan ve güçlü bir grafik araçtır [14]. P diyagramı, özellikle üretim süreçlerinde hatalı ürün oranlarının kontrol edilmesi ve süreç kararlılığının değerlendirilmesi için önemli bir araçtır. Bu diyagram, belirli bir zaman diliminde veya parti bazında oluşan hatalı ürün oranlarını analiz ederek süreçteki sapmaların ve kontrol dışı noktaların tespit edilmesine olanak tanır. Süreç performansının görsel olarak izlenmesine yardımcı olan p diyagramı, aynı zamanda sistematik hataların veya özel nedenlere bağlı sapmaların belirlenmesini kolaylaştırır. Bir süreç, çıktılardaki değişimlerin rastgele varyasyonlardan kaynaklandığı durumlarda "kontrol altında" olarak tanımlanır. Eğer çıktı deseni, rastgele nedenlerden beklenen dağılımı izlemiyorsa, bu süreç "kontrolden çıkmış" sayılır ve bu durumda muhtemel bir neden belirlenebilir[15]. Üretim süreçlerinde, kalite standartlarını sürekli olarak korumak ve hatalı ürünleri minimuma indirmek büyük önem taşır. P diyagramı, sürecin kontrol altında olup olmadığını göstermekle kalmaz, aynı zamanda potansiyel problemleri erken aşamada tespit ederek süreç iyileştirme çalışmalarına rehberlik eder. Bu yönüyle, kalite kontrol ve süreç yönetimi alanlarında güvenilir ve sıklıkla kullanılan yöntemlerdendir. Bu çalışmada süreç kararlılığını değerlendirmek amacıyla p diyagramı kullanılmıştır. P diyagramı, süreçte üretilen ürünlerin hata oranlarını inceleyerek sürecin kontrol altında olup olmadığını değerlendirmek amacıyla kullanılmıştır. Analiz kapsamında, 30 partiye ait fire oranları hesaplanmış ve her parti için hatalı ürün oranı, toplam ürün sayısına bölünerek elde edilmiştir. Daha sonra, bu oranlar kullanılarak kontrol limitleri hesaplanmıştır. Kontrol limitlerinin belirlenmesinde aşağıdaki formül uygulanmıştır:

- **Orta Hata (p):**

p = (Hatalı Ürün Sayısı) / (Toplam Ürün Sayısı)

- **Üst Kontrol Limiti (UCL):**

$UCL = \bar{p} + 3 * \sqrt{[\bar{p} * (1 - \bar{p}) / n]}$

- **Alt Kontrol Limiti (LCL):**

$LCL = \bar{p} - 3 * \sqrt{[\bar{p} * (1 - \bar{p}) / n]}$

Tek yönlü ANOVA, üretim süreçlerinde farklı koşulların (örneğin, farklı vardiyalar, makineler veya üretim yöntemleri) ürün kalitesi veya verimliliği üzerindeki etkilerini incelemek için oldukça önemli bir araçtır. Üretim süreçlerinde, birden fazla faktörün etkileşimi ve bu faktörlerin üretim çıktıları üzerindeki etkisi genellikle karmaşık olabilir. Tek yönlü ANOVA, bu faktörlerin her birini tek başına değerlendirerek, farklı gruplar (örneğin, farklı vardiyalarda üretilen ürünler) arasındaki ortalama farkların anlamlı olup olmadığını belirler. Bu sayede, üretim hattında herhangi bir faktörün (örneğin, vardiya değişiklikleri veya makineler arasındaki farklar) ürün kalitesi veya verimliliği üzerinde anlamlı bir etki yaratıp yaratmadığı anlaşılabilir. Üretim hattındaki vardiyalar arasında gözlemlenen kalite farklılıklarını değerlendirmek için tek yönlü ANOVA analizi uygulanmıştır. ANOVA, grup içi ve grup arası varyansların oranı olan F istatistiğini kullanır. Analizin temel ilgi alanı grup ortalamalarının farklılıklarına odaklanır; ancak ANOVA, varyans farkına odaklanır [13]. Bu yöntem, dört vardiyadaki kötü ürün oranlarının istatistiksel olarak anlamlı bir farklılık gösterip göstermediğini belirlemek amacıyla kullanılmıştır. İlk adımda, her vardiya için kötü ürün oranlarının ortalamaları hesaplanmıştır. Sonraki adımda, grup içi varyans (grup içindeki bireysel ölçümler arasındaki fark) ve grup dışı varyans (grup ortalamaları arasındaki fark) hesaplanarak F-değeri bulunmuştur. F-değeri, grup içi ve grup dışı varyans oranı olarak tanımlanır. Bu değerin yüksek olması, gruplar arasında anlamlı bir fark olduğuna işaret eder. Ardından, p-değeri hesaplanarak, bu değerin istatistiksel olarak anlamlı olup olmadığı test edilmiştir. Eğer p-değeri belirlenen anlamlılık seviyesinden (genellikle 0.05) küçükse, null hipotez reddedilerek vardiyalar arasında anlamlı bir fark olduğu sonucuna varılmıştır.

Uygulama Adımları:

- Her vardiyaya ait kötü ürün oranları hesaplanmıştır.
- Vardiyalar arasında kötü ürün oranlarındaki farklılıklar tek yönlü ANOVA testi ile analiz edilmiştir.
- ANOVA sonucunda elde edilen p-değeri yorumlanarak vardiyalar arasında anlamlı bir fark olup olmadığı değerlendirilmiştir.

Hipotezler:

- $H_0$ (null hipotez): Tüm vardiyaların ortalama kötü ürün oranları birbirine eşittir.
- $H_1$ (alternatif hipotez): En az bir vardiyanın ortalama kötü ürün oranı diğerlerinden farklıdır.

Analiz sonucunda, özellikle 4. vardiyada kötü ürün oranlarının diğer vardiyalara göre anlamlı bir şekilde daha yüksek olduğu tespit edilmiştir ($p < 0.05$). Bu durum, dördüncü vardiyalarda üretim performansını düşüren belirli faktörlerin varlığına işaret etmektedir.

Kök Neden Analizi, bir problemin yüzeysel belirtilerine değil, ardındaki gerçek sebeplere odaklanarak çözüm üretmeye yönelik güçlü bir araçtır. Bu analiz yöntemi, özellikle karmaşık süreçlerde tekrarlayan sorunları anlamak ve bunların temel nedenlerini ortaya çıkarmak için kullanılır. Geleneksel yöntemler çoğunlukla sorunun geçici çözümleriyle sınırlı kalırken, kök neden analizi, problemin kökenine inerek daha kalıcı ve etkili çözümler geliştirmeyi sağlar. Üretim hattındaki fire oranları veya verimlilik sorunları gibi durumlarda, kök neden analizi kullanılarak yalnızca dışsal faktörler değil, süreçlerin, ekipmanların, insan hatalarının ve organizasyonel zayıflıkların da etkileri göz önünde bulundurulur. Bu nedenle, kök neden analizi, problemlerin tekrarlanmasını önlemek, kaynakları daha verimli kullanmak ve süreçlerin sürdürülebilirliğini sağlamak için kritik bir yöntemdir. Bu çalışmada, 4. vardiyalarda gözlemlenen yüksek fire oranlarının temel nedenlerini analiz etmek amacıyla 5 Neden yöntemi kullanılmıştır. Bu metodoloji, sorunun kök nedenlerine ulaşmayı ve çözüm önerileri geliştirmeyi hedeflemiştir.

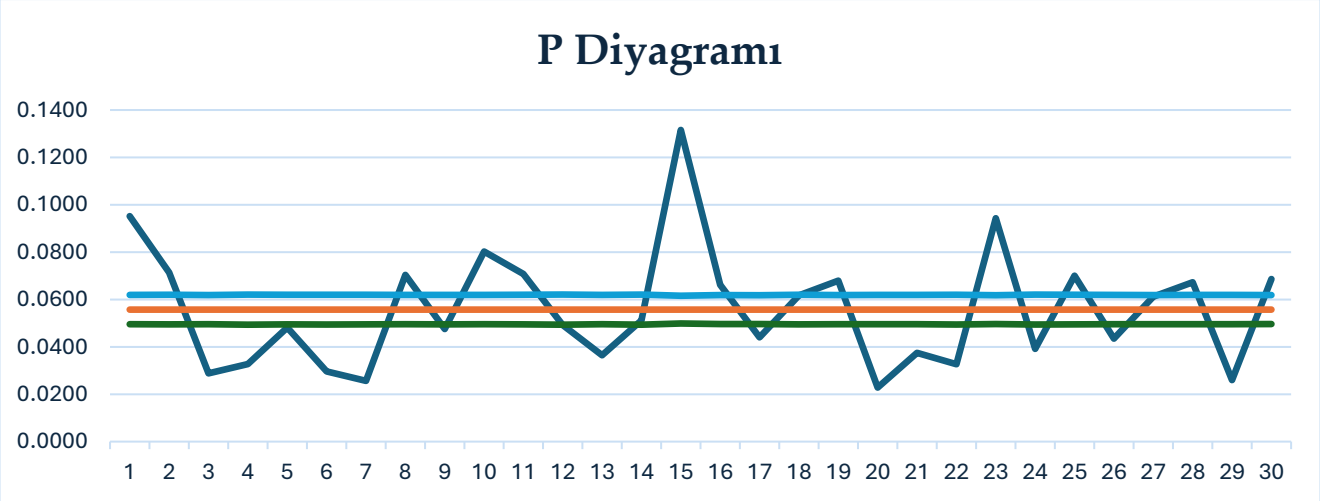
## 3. Sonuçlar

### 3.1. P Diyagramı

Üretim süreci, 24 saatlik zaman dilimi içerisinde 3 vardiya esasına göre yürütülmektedir. Her bir parti ürün, ardışık 4 vardiyada tamamlanacak şekilde üretilmektedir. Bu çalışmada, her vardiyada elde edilen toplam ürün sayısı ile hatalı ürün sayısı kullanılarak ilgili vardiya için p değeri (hatalı ürün oranı) hesaplanmış ve bu değerler p diyagramında görselleştirilmiştir.

30 Parti (120 vardiya) için oluşturulan p diyagramı, süreçteki fire oranlarının belirli bir düzende dağılmadığını ve sürecin kontrol dışı olduğunu açıkça göstermektedir. Bu sonuç, üretim sürecinde

istenmeyen dalgalanmaların ve hataların varlığını işaret etmektedir. Ancak, p diyagramı sadece sürecin kontrol dışında olduğunu göstermekte olup, bu durumun sebeplerine dair daha detaylı bir inceleme yapılması gerekmektedir. Bu bağlamda, ANOVA analizi kullanılarak, sürecin kontrol dışı olmasının vardiya bazında farklılıklar gösterip göstermediği daha derinlemesine analiz edilecektir. Böylece, sürecin kontrol dışı olmasına neden olan faktörlerin belirlenmesi ve bunların düzeltilmesine yönelik daha somut adımlar atılabilecektir.




**Figür 1.** P Diyagramı

### 3.2 Vardiya Bazlı Anova Analizi

Figür 1'de görüldüğü üzere, çok sayıda kontrol dışı nokta bulunmaktadır. Bu, sürecin belirli aralıklarla istenmeyen dalgalanmalara ve hatalara neden olduğunu göstermektedir. Daha detaylı inceleme bir sonraki adımda ANOVA analizi ile gerçekleştirilmiştir.

**Tablo 1.** Anova Sonuçları

| Anova: Tek Etken | | | |
|---|---|---|---|
| **Gruplar** | **Vardiya Sayısı** | **Ortalama Kötü Ürün Oranı** | **Varyans** |
| VARDİYA 1 | 15 | 0,0413 | 0,0008 |
| VARDİYA 2 | 15 | 0,0414 | 0,0004 |
| VARDİYA 3 | 15 | 0,0465 | 0,0006 |
| VARDİYA 4 | 15 | 0,0932 | 0,0025 |
| *F* | *P-değeri* | | *F ölçütü* |
| 9,0098 | 0,0000580 | | 2,7694 |

Bu analizde, 4 farklı vardiya arasındaki ortalama kötü ürün oranlarının istatistiksel olarak anlamlı bir fark gösterip göstermediğini incelemek amacıyla tek yönlü (tek etkenli) ANOVA testi uygulanmıştır.
Tablo 1'de en yüksek ortalama kötü ürün oranı 4. vardiyada görülmektedir. Bu da, ANOVA sonuçlarının desteklediği şekilde, 4. vardiyanın diğer vardiyalara göre daha yüksek kötü ürün oranına sahip olduğunu göstermektedir.

- $H_0$ (null hipotez): Tüm vardiyaların ortalama kötü ürün oranları birbirine eşittir.
- $H_1$ (alternatif hipotez): En az bir vardiyanın ortalama kötü ürün oranı diğerlerinden farklıdır.

ANOVA sonuçlarına göre:

- Hesaplanan F değeri: 9,0098
- Kritik F değeri (F ölçütü): 2,7694
- P-değeri: 0,0000580

P-değeri, %5 anlamlılık düzeyinden ($\alpha = 0,05$) çok daha küçük olduğundan, $H_0$ reddedilmiştir. Bu durum, vardiyalar arasında ortalama kötü ürün oranları açısından istatistiksel olarak anlamlı bir fark olduğunu göstermektedir.

Özellikle 4. vardiya, ortalama kötü ürün oranı açısından diğer vardiyalara kıyasla belirgin şekilde daha yüksek bir değere (0,0932) sahiptir. Bu bulgu, 4. vardiyanın kötü ürün oranının diğer vardiyalara kıyasla istatistiksel olarak anlamlı derecede daha yüksek olduğunu ve bu farkın ANOVA testi ile doğrulandığını ortaya koymaktadır. Bu durum, üretim sürecindeki sorunların özellikle 4. vardiyada yoğunlaştığını ve bu vardiyadaki operasyonel farklılıkların süreç performansını olumsuz etkileyebileceğini göstermektedir.

Bir sonraki adım olarak, bu anlamlı farkların kök nedenlerini belirlemek için kök neden analizi yapılacaktır. Bu analiz, makinelerdeki performans düşüşlerinin, özellikle 4. vardiya ile ilişkilendirilen sorunların nedenlerini keşfetmeye ve çözüm önerileri geliştirmeye yardımcı olacaktır. Kök neden analizi, 4. vardiyadaki makine verimliliğini etkileyebilecek faktörleri derinlemesine inceleyecek ve sürecin iyileştirilmesine yönelik somut adımlar atılmasını sağlayacaktır.

### 3.3 Kök Neden Analizi

Bu çalışmada, 4. vardiyalarda gözlemlenen yüksek fire oranlarının nedenlerini belirlemek amacıyla kök neden analizine başvurulmuştur. İlk olarak, 4. vardiyalardaki fire oranlarının diğer vardiyalara göre daha fazla olduğu gözlemiyle başlanmıştır. Bunun ardından, makinelerdeki toz sızdırmalarının artmasının, ürün kalitesini düşüren temel faktör olduğu tespit edilmiştir. Toz sızdırmalarının artmasının nedeninin, mevcut ara temizlik uygulamaları; yalnızca ürün parametrelerinde ciddi sapmalar meydana geldiğinde, operatörler tarafından ve sistematik olmayan bir şekilde yapılmakta olduğu gözlemlenmiştir. Temizlik sıklığı üretim planlamasını aksatmamak amacıyla sabit tutulmakta, ancak bu yaklaşım filtrelerdeki toz birikimini önlemeye yetmemektedir. 5 Neden analizi sonucunda, mevcut temizlik prosedürünün yeterli sıklıkta uygulanmaması ve yoğun üretim planlarının temizlik için yeterli duruş süreleri tanımaması gibi faktörlerin, makinelerdeki toz birikimini ve dolayısıyla 4. vardiyadaki fire oranlarını artırdığı anlaşılmıştır. Yoğun üretim programı nedeniyle temizlik sıklığının artırılmasının engellenmesi, makinelerdeki verimliliği ve kaliteyi olumsuz yönde etkilemiştir. Bu analiz, makinelerdeki temizlik sıklığının arttırılması gerektiğini ortaya koymuş ve üretim kapasitesini etkileyebilecek bu durumu göz önünde bulundurarak çözüm önerileri geliştirilmiştir. Çalışmada bu temel nedenlerin tespit edilmesi, üretim sürecindeki verimliliği artırmak ve kaliteyi iyileştirmek adına temizlik prosedürlerinin gözden geçirilmesi gerektiğini göstermektedir. Bu bağlamda, önerilen çözüm yolları, makinelerdeki toz birikimini ve sızdırmaları azaltarak, 4. vardiyadaki fire oranlarını iyileştirmeyi hedeflemektedir.

Fire oranlarının diğer vardiyalara kıyasla anlamlı şekilde yüksek çıkması, üretim sürecinde belirli bir noktada süreklilik gösteren bir probleme işaret etmektedir. Bu kapsamda uygulanan "5 Neden Analizi", sorunun temel kaynağını sistematik olarak açığa çıkarmıştır. Analiz sonucunda, 4. vardiyalarda artan fire oranlarının temel nedeni olarak makinelerdeki toz sızdırmaları belirlenmiştir. Sızdırmaların yoğun olarak bu vardiyada ortaya çıkmasının, makinelerde uygulanan ara temizlik prosedürlerinin yetersizliğinden kaynaklandığı tespit edilmiştir. Mevcut ara temizlik uygulamaları; yalnızca ürün parametrelerinde ciddi sapmalar meydana geldiğinde, operatörler tarafından ve sistematik olmayan bir şekilde yapılmaktadır. Temizlik sıklığı üretim planlamasını aksatmamak amacıyla sabit tutulmakta, ancak bu yaklaşım filtrelerdeki toz birikimini önlemeye yetmemektedir. Ayrıca, yoğun üretim planları nedeniyle ara temizlik sıklığını artırmaya yönelik sistematik bir düzenleme yapılmamıştır. Bu durum, üretim sürekliliği önceliği nedeniyle, temizlik işlemlerinin gerekli sıklıkta yapılmasının önüne geçmektedir. Bu bağlamda gerçekleştirilen kök neden analizi sonucunda, üretim hattında ürün parametrelerinde sapmaların meydana geldiği durumlarda operatörler tarafından gerçekleştirilen ara temizlik uygulamalarının, ürün kalitesi üzerindeki belirleyici etkisi bir kez daha ortaya konmuştur. Her ne kadar bu temizlik faaliyetleri, üretim planlamasını aksatmayacak şekilde ve kısa süreli müdahaleler olarak gerçekleştiriliyor olsa da, mevcut temizlik periyodu makinelerdeki toz birikimini tam anlamıyla engelleyememektedir. Özellikle filtreleme sisteminde zamanla oluşan toz yükü, belirli vardiyalarda (özellikle 4. vardiyada) makinelerin sızdırma yapmasına ve bu durumun da fire oranlarının artmasına neden olduğu tespit edilmiştir. Bu doğrultuda yapılan iyileştirme çalışması kapsamında, ara temizlik işlemlerinin daha sistematik, standartlaştırılmış ve izlenebilir hale getirilmesi önerilmektedir. Özellikle kalite kaybının yüksek olduğu kritik vardiyalarda, filtreleme sistemlerine yönelik kısa süreli fakat etkili temizlik müdahalelerinin düzenli aralıklarla uygulanması ve her bir müdahalenin kayıt altına alınarak

takip edilmesi gerekmektedir. Bununla birlikte, mevcut sabit temizlik periyotlarının yetersiz kaldığı göz önünde bulundurularak, temizlik sıklığının üretim hacmi, ürün tipi, makinenin ortalama çalıştırılma süresi ve kirlenme eğilimi gibi dinamik değişkenlere bağlı olarak esnek şekilde belirlenmesi gerektiği sonucuna varılmıştır. Bu yaklaşım sayesinde, üretim süresi boyunca makinelerde oluşabilecek kirlenmeler önceden kontrol altına alınabilecek; böylece hem makine performansı korunacak hem de fire oranlarında gözle görülür bir azalma sağlanacaktır.

**Tablo 2.** 5 Neden Analizi

| ***Sorun*** | |
|---|---|
| 4. vardiyalarda fire oranlarının diğer vardiyalara kıyasla istatistiksel olarak fazla olması. | |
| ***Nedenler*** | ***Cevaplar*** |
| 4. vardiyalarda fire oranları neden daha yüksek? | Çünkü makinelerde toz sızdırmaları bu vardiyada daha fazla artıyor ve ürünlerin kalite standartlarını düşürüyor. |
| Makinelerde neden 4. vardiyada toz sızdırmaları daha fazla oluyor? | Çünkü makinelerin ara temizliği ürün değerlerinin sapmalarının artması halinde operatörler tarafından sistematik olmayan şekilde yapılıyor ve bu süre içerisinde filtreler aşırı derecede toz biriktiriyor. |
| Temizlik neden bu şekilde yapılıyor? | Çünkü mevcut ara temizlik prosedürü bu şekilde belirlenmiş ve daha sık temizlik yapılmasının gerekli olduğu düşünülmemiş. |
| Temizlik prosedürü neden daha sık yapılacak şekilde düzenlenmemiş? | Çünkü yoğun üretim planları nedeniyle ara temizlik için sistematik bir çalışma düzeni oluşturulmamış. |
| Yoğun üretim planları neden temizlik sıklığının artırılmasına engel oluyor? | Çünkü ara temizlik duruşlarının üretim kapasitesini düşüreceği varsayılıyor ve bu durum ara temizlik sıklığının artırılmasını öncelikli bir çözüm olarak görmeyi zorlaştırıyor. |
| ***KÖK NEDEN*** | Mevcut ara temizlik prosedürü, makinelerdeki filtrelerde biriken tozun zamanla artmasına ve 4. vardiyada sızdırmaların ürün kalitesini düşürmesine neden oluyor. |

## 4. Tartışma

Bu çalışmada, bir ilaç üretim hattındaki süreç kararlılığı ve fire oranları analiz edilerek, üretim verimliliğini etkileyen temel faktörler belirlenmiştir. Elde edilen bulgular, p diyagramı ile sürecin kontrol dışında olduğunu gösterirken, ANOVA analizi özellikle 4. vardiyalarda fire oranlarının istatistiksel olarak anlamlı şekilde daha yüksek olduğunu ortaya koymuştur. Kök Neden Analizi sonucunda, bu durumun makinelerde biriken toz sızdırmalarından kaynaklandığı ve temizlik sıklığının yetersiz olmasının süreci olumsuz etkilediği tespit edilmiştir. Önceki çalışmalar, üretim süreçlerinde düzenli bakım ve temizliğin süreç kararlılığı ve kalite üzerindeki kritik rolünü vurgulamaktadır. Bu bağlamda, çalışmamız mevcut literatür ile uyumlu olup, temizliğin yetersiz sıklıkta yapılmasının üretim hatlarında fire oranlarını artırabileceğini somut verilerle desteklemektedir. Ancak, bu çalışmada yalnızca belirli bir üretim hattı ve belirli bir makine tipi ele alınmıştır. Farklı üretim süreçleri ve makine tipleri üzerindeki etkileri değerlendirmek için daha geniş kapsamlı araştırmalara ihtiyaç duyulmaktadır. Bulgular, ilaç üretim sektöründe kalite kontrol süreçlerinin yalnızca son ürün değerlendirmesiyle sınırlı kalmaması gerektiğini ve süreç içerisindeki kritik noktaların da sürekli izlenmesi gerektiğini göstermektedir. Özellikle vardiya bazlı analizlerin, süreç içindeki değişkenlikleri anlamada önemli bir yöntem olduğu görülmüştür. Gelecekteki araştırmalar, temizlik sıklığının artırılmasının üretim verimliliği üzerindeki doğrudan etkilerini inceleyerek, optimum temizlik aralığını belirlemeye yönelik deneysel çalışmalar yapabilir. Ayrıca, farklı üretim hatlarında benzer analizler gerçekleştirilerek, süreç kararlılığını etkileyen diğer faktörler de incelenebilir. Üretim hattında otomasyon ve sensör destekli veri toplama sistemlerinin entegrasyonu ile süreç içi kontrollerin daha etkin hale getirilmesi de gelecekteki çalışmalar için önemli bir araştırma alanı olabilir.

**Yazar Katkıları:** Çalışmanın literatür taraması, araştırma sürecinin tasarımı ve uygulanması, veri toplama, analiz yöntemlerinin seçimi ve uygulanması ile elde edilen bulguların yorumlanması aşamalarında M.B.Ç. ve B.Ö. etkin sorumluluk üstlenmiştir. Çalışmanın yürütülmesi, planlanması ve makalenin yazım süreci bu yazarlar tarafından gerçekleştirilmiştir. Y.Ö. ve M.Y., çalışmanın bütününe yönelik bilimsel, teknik ve yöntemsel rehberlik sağlamış; özellikle kavramsallaştırma, metodolojik çerçevenin oluşturulması, analizlerin doğrulanması ve biçimsel bütünlüğün sağlanması süreçlerinde M.B.Ç. ve B.Ö. ile yakın iş birliği içerisinde yönlendirici katkılarda bulunmuştur. Tüm yazarlar makalenin son hâlini onaylamış ve içeriğinden ortak sorumluluk üstlenmiştir.

**Finansman:** Bu araştırma kapsamında herhangi bir dış finansman alınmamıştır.

**Çıkar çatışmaları:** Yazarlar çıkar çatışması beyan etmemektedir.

## Kaynaklar

[1] A. Sağbaş, D. Hasan, O. Çapraz ve N. Karakurt, "Yalın üretime geçiş sürecinde seri üretim hattında üretim sistemi optimizasyonu," *Verimlilik Dergisi*, no. 2, pp. 7–27, 2018.

[2] B. Dulkadir, "İşletme Yönetiminde Firenin Azaltılarak Verimliliğin Artırılması: İplik Üretim Tesislerinde Bir Uygulama," *Akademik Yaklaşımlar Dergisi*, vol. 7, no. 2, pp. 19–28, 2016.

[3] D. C. Montgomery, *Introduction to Statistical Quality Control*, 2nd ed. John Wiley & Sons, 1991.

[4] L. C. Alwan, *Statistical Process Analysis*, 2000.

[5] I. Madanhire and C. Mbohwa, "Application of statistical process control (SPC) in manufacturing industry in a developing country," *Procedia CIRP*, vol. 40, pp. 580–583, 2016, doi: 10.1016/j.procir.2016.01.129.

[6] M. Aichouni, "On the use of the basic quality tools for the improvement of the construction industry: A case study of a ready mixed concrete production process," *International Journal of Civil & Environmental Engineering*, vol. 12, no. 1, pp. 31–38, 2012.

[7] H. Yıldırım and E. Karaca, "Üretim sürecinde istatistiksel proses kontrol (İPK) uygulamaları ve elektronik sektöründe bir inceleme," *Öneri Dergisi*, vol. 10, no. 39, pp. 77–87, 2013, doi: 10.14783/od.v10i39.1012000309.

[8] M. Abtew, S. Kropi, Y. Hong, and L. Pu, "Implementation of Statistical Process Control (SPC) in the sewing section of garment industry for quality improvement," *Autex Research Journal*, vol. 18, no. 2, pp. 160–172, 2018, doi: 10.1515/aut-2017-0034.

[9] W. R. Simatupang, N. Sutrisno, H. Haniza, N. Siregar, B. S. Kesuma, and R. Zulvatricia, "Application of SQC and ANOVA in quality control of palm kernel oil," *Journal of Industrial and Manufacture Engineering*, 2024.

[10] F. A. Ardelean, "Case study using analysis of variance to determine groups' variations," 2017.

[11] C. Paese, C. S. Caten, and J. L. Ribeiro, "Aplicação da análise de variância na implantação do CEP," *Production Journal*, vol. 11, pp. 17–26, 2001.

[12] B. Çayır Ervural, "Varyans Analizi (ANOVA) ve Kovaryans Analizi (ANCOVA) İle Deney Tasarımı: Bir Gıda İşletmesinin Tedarik Süresine Etki Eden Faktörlerin Belirlenmesi," *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, vol. 7, no. 2, pp. 923–941, 2020, doi: 10.35193/bseufbd.719341.

[13] Y. Y. Tesfay, "Process control charts," in *Developing Structured Procedural and Methodological Engineering Designs*, Springer, Cham, Switzerland, 2021.

[14] [H. S. Oberoi, M. Parmar, H. Kaur, ve R. Mehra, "Statistical Process Control: A Quality Control Technique for Confirmation to Ability of Process," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 6, pp. 666–672, 2016.

[15] T. Kim, "Kavramsal figürleri kullanarak tek yönlü ANOVA'yı anlamak," *Kore Anesteziyoloji Dergisi*, vol. 70, no. 1, pp. 22–26, 2017, doi: 10.4097/kjae.2017.70.1.22

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

*Rewiev Article*

Received: date:04.12.2024
Accepted: date:21.05.2025
Published: date:30.06.2025

# Türkiye'de Proje Çizelgeleme Konulu Lisansüstü Tezlerin Bibliyometrik Analizi

**Ecem Şevval Pınarcı[1], Emel Güven[2] ve Tamer Eren[3*]**

[1]Department of Industrial Engineering, Kırıkkale University, Kırıkkale Türkiye; ecemsevvalpinarci@gmail.com
[2]Department of Industrial Engineering, Kirikkale University, Kırıkkale Türkiye; emel-gvn@hotmail.com
[3*]Department of Industrial Engineering, Kirikkale University, Kırıkkale Türkiye; tamereren@gmail.com
Orcid: 0009-0004-6784-0925[1] Orcid: 0000-0001-6106-9720[2] Orcid: 0000-0001-5282-3138[3]
*Correspondence: tamereren@gmail.com

**Öz:** Proje çizelgelemesi, bir projeyi oluşturan görevlerin başlangıç ve bitiş zamanlarını, bağımlılık ilişkilerini ve kaynak kullanımını dikkate alarak zaman ekseninde planlanması sürecidir. Bu süreçte, her bir faaliyetin ne zaman başlayacağı ne kadar süreceği ve hangi kaynaklarla gerçekleştirileceği belirlenir. Çalışmanın amacı, Yüksek Öğretim Kurulu Başkanlığı Ulusal Tez Merkezi Elektronik Arşivi (YÖKTEZ) veri tabanında bulunan "Proje Çizelgeleme "alanında yapılan çalışmaları bibliyometrik analiz yöntemiyle incelemektir. Bu doğrultuda, YÖKTEZ üzerinden 1991-2024 yılları aralığında "Proje Çizelgeleme "taraması yapılıp toplamda 66 lisansüstü tez çalışması üzerinden analiz gerçekleştirilmiştir. Analiz kapsamında "Proje Çizelgeleme "alanında en çok çalışmanın "2022" yılında yapıldığı ve yapılan çalışmaların büyük çoğunluğunun yüksek lisans tezi olduğu görülmüştür. YÖKTEZ veri tabanında yapılan bibliyometrik analiz, Türkiye'de proje çizelgeleme alanındaki akademik çalışmaları sistematik bir şekilde incelemeyi ve bu çalışmaların hangi konulara odaklandığını, hangi yöntemlerin kullanıldığını ve hangi sonuçların elde edildiğini ortaya koymayı amaçlamaktadır. Bu tür bir analiz, proje çizelgeleme konusundaki bilgi boşluklarını belirlemeye, yeni araştırma alanları önermeye ve bu alandaki bilgi birikimini artırmaya yardımcı olacaktır.

**Anahtar Kelimeler:** Proje Çizelgeleme, Bibliyometrik Analiz, Lisansüstü Tez

# Bibliometric Analysis of Graduate Theses on Project Scheduling in Türkiye

**Abstract:** Project scheduling is the process of planning the tasks that make up a project on a timeline by taking into account their start and end times, dependency relationships, and resource usage. In this process, the start time, duration, and required resources for each activity are determined. The aim of the study is to examine the works conducted in the field of "Project Scheduling" in the YÖKTEZ (Council of Higher Education National Thesis Center) database using bibliometric analysis. Accordingly, a search for "Project Scheduling" was conducted in YÖKTEZ for the years 1991-2024, and an analysis was performed based on a total of 66 graduate theses. The analysis revealed that 2022 witnessed the highest number of studies in the field of 'Project Scheduling,' with a significant proportion being master's theses. The bibliometric analysis conducted in the YÖKTEZ database aims to systematically review academic works in the field of project scheduling in Turkey and to reveal the topics these studies focus on, the methods used, and the results obtained. This analysis aims to uncover existing knowledge gaps in the field of project scheduling, propose potential avenues for future research, and enrich the academic discourse surrounding this topic.

**Keywords:** Project Scheduling, Bibliometric Analysis, Graduate Theses

## 1. Giriş

Çizelgeleme proje yönetimi, üretim ve ekip atamaları gibi alanlarda, kaynakların ve işlemlerin en uygun şekilde düzenlenmesi ve zamanlanması sürecidir. "Proje" kavramı, belirlenmiş bir amaca ulaşmak için, başlangıcı ve

**Citation** : E. Ş. Pınarcı, E. Güven, and T. Eren, "Türkiye'de proje çizelgeleme konulu lisansüstü tezlerin bibliyometrik analizi," *Journal of Statistics and Applied Sciences*, no. 11, pp. 59–69, 2025, doi: 10.52693/jsas.1596530.

bitişi belli olan bir çalışmayı ifade eder. Bu amaç doğrultusunda kullanılan en etkili araçlardan biri de proje çizelgelemedir. Proje çizelgeleme, izlenebilirliği ve kontrol altına alınabilirliği olduğundan çok yardımcı olmaktadır [1].

Çizelgeleme problemi birçok alana ayrılmaktadır. Hemşire çizelgeleme, personel çizelgeleme, makine ve ekipman çizelgeleme, üretim çizelgeleme gibi birçok alanda kullanılmaktadır. En çok kullanılan alanlardan biri de proje çizelgelemedir. Proje çizelgeleme firma veya kurumlar için büyük bir önem taşımaktadır. Fırsatları değerlendirmek, mevcut problemleri çözmek veya değişen koşullara uyum sağlamak amacıyla projeler üretirler. Bu süreçte, karşılaşılan riskler göz önünde bulundurularak, en yüksek kârın elde edilmesi hedeflenir ve buna göre çizelgeleme oluşturulur [2].

Projelerin en kısa sürede, minimum maliyetle ve yüksek kalite standartlarında tamamlanmasını sağlamak için, proje yönetimi yaklaşımı geliştirilmiştir. Bu yaklaşım, insan gücü, ekipman, makine, bilgi birikimi, beceriler ve diğer kaynakların etkin ve uyumlu şekilde kullanımı üzerine odaklanır. Proje yönetimi, işletmelerde projelerin başarı oranını artırmanın yanı sıra, kaynak kullanımında verimliliği sağlamak ve kurumsal iletişimi güçlendirmek için önemli bir araçtır [3]. Proje yönetimi, planlanan hedefe doğru ilerlenmesinde çizelgeleme ile ele alınan bir süreçtir. Proje çizelgeleme, kısıtların belirlenip optimum zamanda veya maliyette projelerin tamamlanmasını hedefleyen bir problemdir [4].

Proje çizelgeleme, firmalar ve kurumlar için iş süreçlerinin düzenli ve verimli bir şekilde yürütülmesini sağlayan önemli bir araçtır. Bu süreçler, iş akışındaki tüm faaliyetlerin ayrıntılı şekilde analiz edilmesini ve görselleştirilmesini mümkün kılarak, proje yönetimindeki karmaşıklığın azaltılmasına katkıda bulunur. Ayrıca, projelerin bütçe kısıtları içinde kalmasına yardımcı olurken, kaynakların etkin kullanımını desteklemektedir. Aynı zamanda, proje ilerleyişinin izlenmesini kolaylaştırır ve zaman yönetiminde şeffaflık sağlar. Tüm bu unsurlar, projenin başarılı bir şekilde tamamlanma olasılığını artırırken, işletmelerin stratejik hedeflerine ulaşmasına da olanak tanımaktadır [5]. Projeler, belirli bir başlangıç ve bitiş tarihine sahip olup, bu süreçte kaynakların verimli bir biçimde yönetilerek hedeflenen sonuçların elde edilmesini amaçlayan çalışmalardır. Rekabetin yoğun olduğu iş dünyasında, projelerin hızlı bir şekilde tamamlanması, firmaların pazarda öne çıkabilmesi için kritik bir avantaj yaratır. Bu sebeple, projelerin zamanında ve belirlenen hedeflere uygun olarak sonlandırılması büyük önem taşımaktadır. Bu bağlamda, proje çizelgeleme, kaynakların etkin bir şekilde yönetilmesine ve projelerin başarıyla sonuçlanmasına olanak tanıyan temel bir araç olarak karşımıza çıkmaktadır [6].

Proje çizelgeleme probleminin birçok farklı türü bulunmaktadır. Bunlar; Tek modlu klasik kaynak kısıtlı proje çizelgeleme (KKPÇ) problemleri, Minimum ve maksimum zaman gecikmeli KKPÇ problemleri, Çok modlu KKPÇ (ÇM-KKPÇ) problemleri, minimum ve maksimum zaman gecikmeli çok modlu KKPÇ problemleri, minimum ve maksimum zaman gecikmeli kaynak yatırım problemleridir [7]. Kaynak kısıtlarının dikkate alındığı proje çizelgeleme problemleri, faaliyetlerin en uygun şekilde sıralanarak hem kaynakların etkin kullanımını hem de proje süresinin minimize edilmesini hedeflemektedir. Bu tür modeller, kaynakların sınırlı olduğu durumlarda daha gerçekçi bir yaklaşım sunmaktadır [8]. Bu çerçevede genetik algoritma gibi sezgisel yöntemler, büyük ve karmaşık projelerde çözüm süresini kısaltarak etkili sonuçlar elde etme potansiyeline sahiptir. Sezgisel yöntemlerin kullanımı, klasik optimizasyon yöntemleriyle çözülemeyen büyük ölçekli problemlerde alternatif bir çözüm yolu sunmaktadır [9].

Paksoy ve Uzun [9], genetik algoritma tabanlı bir yaklaşım geliştirerek kaynak kısıtlı proje çizelgeleme problemlerinin çözümüne odaklanmışlardır. Çalışma, faaliyetlerin çizelgelenmesinde kaynak sınırlamalarının dikkate alınarak optimal bir çözüm elde edilmesini hedeflemiştir. Bettemir ve Çakmak [10], küçük ölçekli kaynak kısıtlı proje çizelgeleme problemlerinin çözümünde, tüm arama uzayını tarayan bir yaklaşımın optimal çözüm elde etmedeki etkinliğini araştırma konusu yapmışlardır. Soysal ve diğerleri [11], kaynak kısıtlı çok modlu çoklu proje çizelgeleme problemlerini belirsizlik altında ele alarak, bu tür problemlerin çözümüne yönelik bir model geliştirilmiştir. Kolaylıoğlu [12], inşaat sektöründe proje yönetimi süreçlerini ele alarak, proje yöneticisinin rol ve sorumluluklarının proje başarısına olan etkilerini incelemiştir. Çınar [13], kaynak kısıtlı bilgi teknolojisi projelerinin çizelgelenmesinde meta-sezgisel algoritmaların kullanımını incelemiş ve bu yöntemlerin etkinliğini örnek problemler üzerinde test etmiştir. Özdemir ve Karacabey [14], kaynak kısıtlı proje çizelgeleme problemlerinin çözümünde genetik algoritma yöntemlerinin performansını karşılaştırarak, bu yöntemlerin avantajlarını ve eksiklerini detaylı bir şekilde analizi yapılmıştır. Durucasu ve diğerleri [1], inşaat

projelerinde belirsizlikleri dikkate alarak bulanık CPM yöntemi kullanılmış ve bu yaklaşımın proje çizelgelemedeki etkinliğini değerlendirilmiştir. Joushani [15], değişken yoğunluklu kaynak kısıtlı proje çizelgeleme problemleri için bir matematiksel model ve genetik algoritma yaklaşımı önererek, bu yöntemlerin proje planlamadaki etkisini değerlendirmiştir. Taş vd. [16], analitik hiyerarşi prosesi ve hedef programlama yöntemlerini birleştirerek monoray projelerinin seçiminde kullanılabilecek bir karma model geliştirilmiş ve bu modelin karar süreçlerindeki etkisini analiz etmişlerdir. Hamurcu ve Eren [17], kent içi ulaşım projelerinin değerlendirilmesi için bulanık analitik hiyerarşi prosesi tabanlı çok kriterli optimizasyon ve uzlaşık çözüm yöntemini kullanarak alternatiflerin sıralanması ve en uygun projenin seçilmesi amacıyla bir karar destek modeli geliştirmişlerdir.

Proje çizelgeleme konusunda yazılmış tezler ve makalelerdeki değişimler ile bu alandaki araştırma eğilimleri üzerine yapılan incelemeler, proje çizelgeleme çalışmalarının zamanla nasıl evrildiğinin anlaşılmasına yardımcı olmaktadır. Bu tür analizler, proje çizelgeleme araştırmalarının hangi dönemlerde ve hangi bağlamlarda yoğunlaştığını ortaya koyarak, gelecekte yapılacak çalışmalar için stratejik bir yol haritası sunmaktadır. Bu çalışma, proje çizelgeleme literatürüne bibliyometrik analiz yöntemiyle katkıda bulunarak hem akademisyenler hem de bu alanda çalışan araştırmacılar ve sektör uzmanları için yol gösterici ve destekleyici bir kaynak niteliği taşımaktadır. Çalışmanın temel amacı, proje çizelgeleme literatürünü sistematik olarak inceleyerek, mevcut yaklaşımları, kullanılan çözüm tekniklerini ve uygulama alanlarını bibliyometrik analiz yöntemiyle ortaya koymaktır. Bu kapsamda, ilgili literatürde yer alan tezler ve akademik makaleler üzerinden alanın tarihsel gelişimi, eğilimleri ve öne çıkan konuları değerlendirilmiştir. Yöntem olarak bibliyometrik analiz tercih edilmiş olup, veriler belirli bir zaman aralığında taranmış ve analiz edilmiştir. Çalışma, giriş bölümünün ardından sırasıyla literatür incelemesi, analiz yöntemi ve bulgular, tartışma ve sonuç bölümleriyle yapılandırılmıştır.

## 2. Materyal Ve Yöntem

Bu kısımda, araştırmanın hedefleri ve kapsamı yanı sıra çalışma sürecinde dikkate alınan varsayımlar ve kısıtlar ele alınmıştır. Ayrıca, araştırma yöntemi ve kullanılan yaklaşımlar da detaylı bir şekilde açıklanmıştır.

### 2.1. Araştırmanın Amacı ve Kapsamı

Bu araştırmanın merkezindeki konusu, proje çizelgeleme üzerine yapılan lisansüstü tezlerin incelenmesi ve bu tezlerin sonuçlarının sentezlenerek yorumlanmasıdır. Bu bağlamda, gerçekleştirilen bibliyometrik analizde aşağıda yer alan soruların cevaplandırılması amaçlanmıştır.

1. Proje çizelgeleme alanında yapılan ve YÖKTEZ'de taranan tez çalışmalarının yıllara göre dağılımı nasıldır?
2. Proje çizelgeleme konulu lisansüstü tezlerin araştırma türlerine göre dağılımı nedir?
3. Proje çizelgeleme konulu lisansüstü tezlerin danışman unvanlarına göre dağılımı nedir?
4. Proje çizelgeleme konulu lisansüstü tezlerin ana bilim dallarına göre dağılımı nedir?
5. Proje çizelgeleme yazınına en fazla katkı sağlayan üniversiteler hangileridir?
6. Proje çizelgeleme konulu lisansüstü tezlerin kullanılan yöntemlere göre dağılımları nedir?
7. Proje çizelgeleme konulu lisansüstü tezlerin konularına göre dağılımları nedir?
8. Proje çizelgeleme konulu lisansüstü tezlerde kullanılan anahtar kelimeler nedir?

### 2.2. Varsayımlar ve Kısıtlar

Çalışmadaki ilk varsayım, lisansüstü tez çalışmalarında bilimsel yöntemlerin kullanıldığıdır. Diğer bir varsayım ise, Türkiye'deki üniversitelerde 1991-2024 yılları arasında proje çizelgeleme konusunda yapılan lisansüstü tezlerin büyük çoğunluğunun YÖKTEZ veri tabanında yer aldığıdır. Ancak, YÖKTEZ veri tabanının doğruluğu ve eksiksizliği kesin olarak garanti edilemediğinden, bu durum araştırmanın sınırlılıklarından biri olarak değerlendirilmektedir. Araştırmanın 1991 yılından başlamasının nedeni, YÖKTEZ'de bu konu ile ilgili ilk lisansüstü tezin bu yıla ait olmasıdır. Veriler, 22.10.2024 tarihine kadar olan çalışmaları kapsamaktadır ve bu tarih aralığındaki YÖKTEZ'deki lisansüstü tez çalışmalarıyla sınırlıdır.

### 2.3. Bibliyometrik Analiz

Bibliyometri, ülkeler, kurumlar, yazarlar, makaleler veya tezlerin arasındaki araştırma alanlarının katkılarını ve üretkenliğini analiz etmek ve niceliksel olarak değerlendirebilmek için matematiksel ve istatiksel yöntemler kullanan disiplinler arası bir bilim dalıdır. Bibliyometrik analizler; makalelerin hem nicel hem de nitel değerlendirmelerinin yapılmasına olanak sağlayan, araştırmacıların yolunu aydınlatan, güçlü istatistiksel bir yöntem olarak kullanılmaktadır [18]. Bu analiz türü, yayın sıklığı, yazarları, kurumları, alıntı sayıları ve kullanılan anahtar kelimeler gibi çeşitli göstergeleri kullanarak yayınları inceleyerek belirli bir bilimsel disiplin

veya konudaki yayınların farklı özelliklerini değerlendirir. Bibliyometrik analiz, bilimsel alanlardaki eğilimlerin izlenmesi, araştırma alanlarının ve ilgili konuların tanımlanması, bu alanlardaki yeniliklerin takip edilmesi ve araştırma stratejilerinin belirlenmesi amacı ile kullanılmaktadır [19].

Bibliyometrik analiz, yaygın olarak bilimsel yayınları içeren bibliyografik veri tabanlarından elde edilen veriler üzerinde gerçekleştirilir. Bu analiz türü, yayınlanan makale sayısı, yazar ve kurum dağılımı, atıf sayıları ve atıf ilişkileri gibi farklı göstergeler aracılığıyla bilimsel literatürün genel bir analizini sağlar [20]. Literatürde farklı alanlarda yapılan çok sayıda bibliyometrik analiz bulunmaktadır. Pınarcı ve diğerleri [18], Türkiye'de ekip çizelgeleme konulu lisansüstü tezlerin bibliyometrik analizi yöntemini kullanarak incelemişlerdir. Merigo ve Yang [21], operasyonel araştırma ve yönetim bilimi alanındaki literatürün bibliyometrik analizini yaparak, bu alandaki temel yayınları, yazarları ve araştırma eğilimlerini değerlendirmişlerdir. Aksungur ve diğerleri [22], insansız hava araçları (İHA) konulu lisansüstü tezlerin bibliyometrik analizini yapmışlardır. Öztürk ve Kurutkan [23], kalite yönetimi alanını bibliyometrik analiz yöntemiyle inceleyerek, bu alandaki bilimsel eğilimleri ve araştırma yoğunluklarını değerlendirmiştir. Sanlı ve diğerleri [24], siber güvenlik alanındaki akademik çalışmaları bibliyometrik analiz yöntemleriyle incelemişlerdir. Gaferoğlu ve diğerleri [25], tsunami konulu lisansüstü tezlerin bibliyometrik yöntemlerle değerlendirmişlerdir. Birinci [26], Turkish Journal of Chemistry dergisinin bibliyometrik analizini yaparak, dergideki makalelerin bilimsel gelişimini ve alandaki eğilimleri incelemiştir. Beşel [27], Türkiye'de maliye alanında gerçekleştirilen lisansüstü tezleri bibliyometrik yöntemlerle incelemiştir. Öztürk ve diğerleri [28], sağlık turizmi konulu lisansüstü tezlerin bibliyometrik analizi ele alınmıştır. Zeren ve Kaya [29], dijital pazarlama alanındaki ulusal yazını bibliyometrik analiz ile inceleyerek, bu alandaki araştırma eğilimlerini ve bilimsel gelişimleri ortaya koymuşlardır. Duran ve Çelikkaya [30], Türkiye'de lojistik üzerine yapılmış lisansüstü tezlerin bibliyometrik analizini gerçekleştirmişlerdir Küpçüoğlu ve diğerleri [31], blok zincir teknolojisi üzerine yazılan yüksek lisans ve doktora tezlerinin bibliyometrik analizini gerçekleştirmişlerdir. Budd [32], yükseköğretim literatüründe yayınların dağılımını ve eğilimlerini analiz ederek bu alandaki ana temaları ortaya koymuştur.

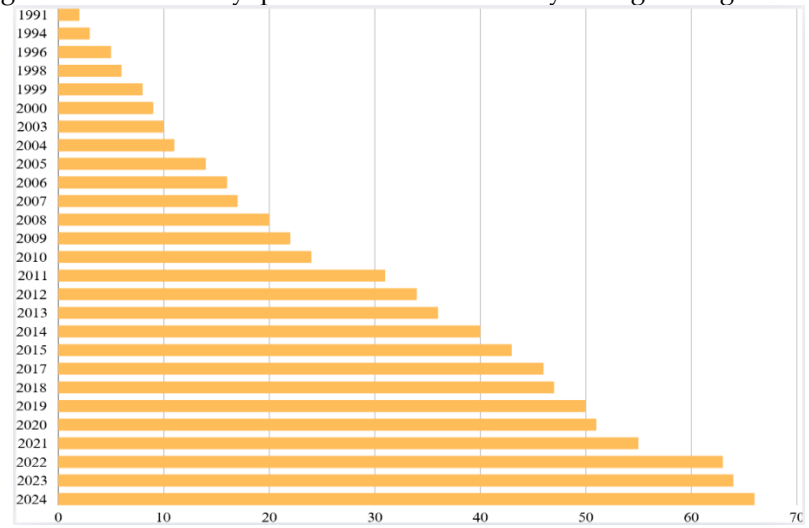
Bu çalışmada, yükseköğretim araştırmalarında hangi konuların ön plana çıktığı ve bu alandaki bilgi üretim süreçleri değerlendirilmiştir. Literatür taramasında "proje çizelgeleme" alanında bibliyometrik analiz çalışmasına rastlanamadığı için bu konudaki çalışmaların çok sınırlı olduğunu söyleyebilmek mümkündür. Çalışmanın bu yönüyle literatüre katkı sağlayacağı düşünülmektedir. Bu kapsamda Türkiye'deki üniversiteler bünyesinde yapılan lisansüstü tezler, YÖKTEZ Merkezi Elektronik Arşivinde "proje çizelgeleme" anahtar sözcüğü yardımı ile araştırılmış ve elde edilen bulgular figürler kullanılarak görselleştirilip yorumlanmıştır.

## 3. Sonuçlar

Bu bölümde proje çizelgeleme anahtar kelimesiyle yapılan tarama sonucunda, ulaşılan lisansüstü tezler ayrıntılı bir biçimde gözden geçirilip ayrıştırılmış ve araştırmanın soruları çerçevesinde değerlendirmeye alınmıştır.

### 3.1. Lisansüstü Tezlerin Yıllara Göre Dağılımları

Şekil 1'de proje çizelgeleme konusunda yapılan lisansüstü tezlerin yıllara göre dağılımları verilmiştir.

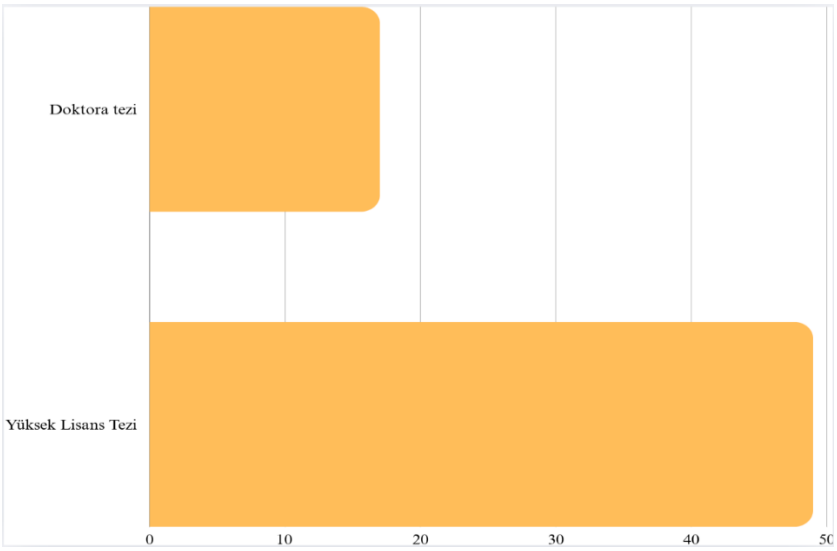



**Şekil 1.** Lisansüstü Tezlerin Yıllara Göre Dağılımları

1991-2024 yılları arasında araştırma kapsamında yapılan tez çalışmalarının yıllara göre dağılım grafiği detaylı olarak incelendiğinde, proje çizelgeleme konusunda en fazla tez çalışmasının 2022 yılında yapıldığı ve bu yılın toplam tezlerin %12'sini (8 adet) oluşturduğu gözlemlenmiştir. Bu grup içerisinde, 2011 yılına ait çalışmaların sayısı 7 tez olup, bu rakam toplam tezlerin yaklaşık %11'ine denk gelmektedir. Yıllara göre dağılım incelendiğinde, 1991–1994 ile 2000–2003 yılları arasında üçer yıl boyunca hiçbir tez çalışmasına rastlanmamıştır. Buna karşın, 2011 ve 2022 yıllarında belirgin bir artış olduğu saptanmıştır. 2011 yılı, proje yönetimi alanında teknolojik gelişmelerin hız kazanması ve büyük ölçekli projelere olan ilginin artmasıyla öne çıkarken; 2022 yılındaki artışın, COVID-19 pandemisi sonrası dijitalleşme, uzaktan çalışma uygulamaları ve proje yönetiminde yeni yaklaşımlara duyulan ihtiyacın bir yansıması olduğu düşünülmektedir. Sonuç olarak proje çizelgeleme konusundaki çalışmalarının 1991 yılından itibaren belirli aralıklarla sürdüğü tespit edilmiştir.

### 3.2. Lisansüstü Tezlerin Araştırma Türlerine Göre Dağılımları

Şekil 2'de proje çizelgeleme konusunda yapılan lisansüstü tezlerin araştırma türlerine göre dağılımları verilmiştir.

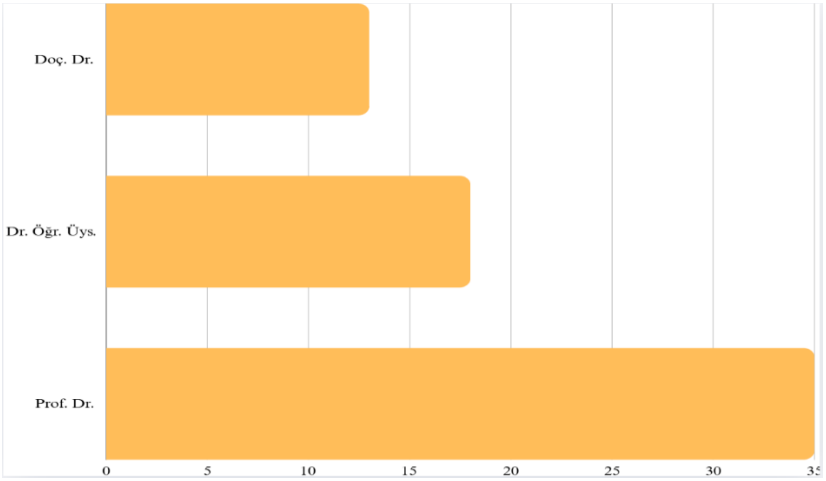



**Şekil 2.** Lisansüstü Tezlerin Araştırma Türlerine Göre Dağılımları

1991-2024 yılları arasında gözlemlenen ve incelenen 66 tez çalışmasının %74'ünü (49 adet) yüksek lisans tezleri, %26'sını (17 adet) ise doktora tezleri oluşturmaktadır. Şekil 2 incelendiğinde, proje çizelgeleme konusunda en fazla yüksek lisans tezinin hazırlandığı görülmektedir.

### 3.3. Lisansüstü Tezlerin Danışman Unvanlarına Göre Dağılımları

Şekil 3'te proje çizelgeleme konusunda yapılan lisansüstü tezlerin danışman unvanlarına göre dağılımları verilmiştir.

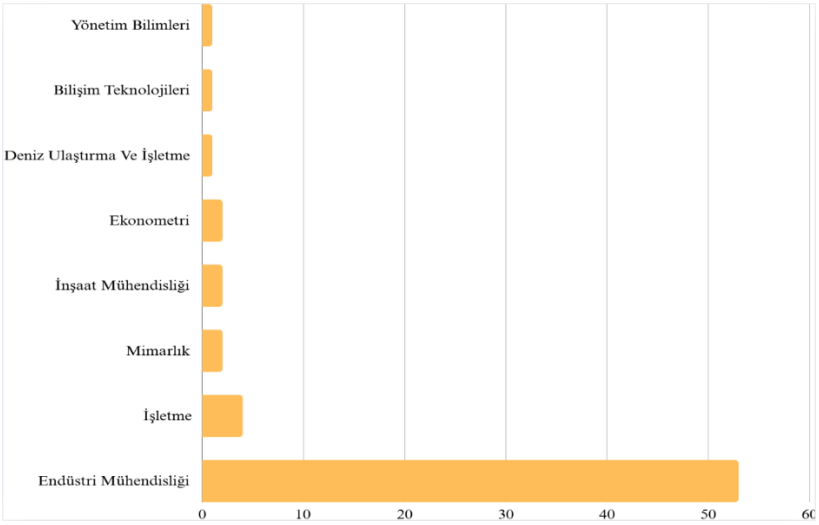



**Şekil 3.** Lisansüstü Tezlerin Danışman Unvanlarına Göre Dağılımları

Proje çizelgeleme konusunda yapılan 66 tezde danışmanların %53'ü (35 adet) Profesör Doktor, %27'ı (18 adet) Doktor Öğretim Üyesi (Yrd. Doç. Dr.) ve %20'u (13 adet) Doçent Doktor unvanına sahiptir.

### 3.4. Lisansüstü Tezlerin Ana Bilim Dallarına Göre Dağılımları

Şekil 4'te proje çizelgeleme konusunda yapılan lisansüstü tezlerin ana bilim dallarına göre dağılımları verilmiştir.

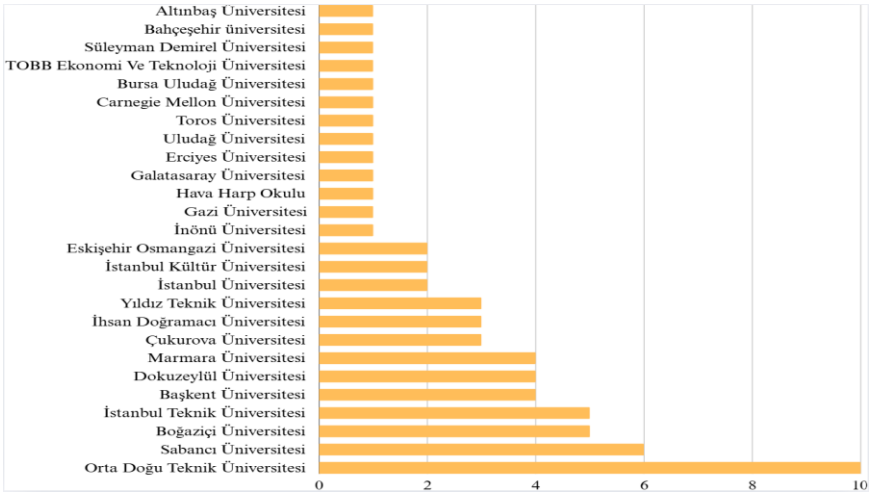



**Şekil 4.** Lisansüstü Tezlerin Ana Bilim Dallarına Göre Dağılımları

Şekil 4'e bakıldığında en çok Endüstri Mühendisliği Ana Bilim Dalı'nda %80'i (53 adet) tez çalışmasının yapıldığı görülmüştür. Çıkarılan sonuç ise, bu disiplin sınırlı kaynaklarla süreçlerin en verimli şekilde yönetilmesini amaçlamaktadır. Endüstri mühendisliği, insan, malzeme, bilgi, donanım ve enerjiden oluşan operasyonel sistemlerin tasarımı, analizi ve optimizasyonu üzerine yoğunlaşır ve proje çizelgeleme, bu hedeflere doğrudan katkı sağlamaktadır [33]. Proje çizelgeleme, iş gücü, zaman ve ekipman gibi kaynakların etkin kullanımını sağlayarak, projelerin maliyet ve zaman açısından optimize edilmesine olanak tanımaktadır [3]. Ayrıca, bu problemlerin çözümü için kullanılan matematiksel modelleme ve algoritmalar, endüstri mühendisliği eğitiminde sıklıkla öğretilen tekniklerle paralellik göstermektedir. Bu nedenle, bu alanda yapılan tezler ve araştırmalar, sadece teorik katkılar sunmakla kalmaz, aynı zamanda gerçek dünyadaki uygulamalara da rehberlik edecektir. Bunun dışında proje çizelgeleme konusunun; İşletme Ana Bilim Dalı %6'sı (4 adet), Mimarlık %3'ü (2 adet), İnşaat Mühendisliği Ana Bilim Dalı %3'ü (2 adet), Ekonometri Ana Bilim Dalı %3'ü (2 adet), Deniz Ulaştırma Ve İşletme Mühendisliği Ana Bilim Dalı %2'i (1 adet), Bilişim Teknolojileri Ana Bilim Dalı %2'si (1 adet) ve Yönetim Bilimleri Ana Bilim Dalı %2'si (1 adet) yakından ilgili olduğu söylenebilir.

### 3.5. Lisansüstü Tezlerin Üniversiteye Göre Dağılımları

Şekil 5'te proje çizelgeleme konusunda yapılan lisansüstü tezlerin üniversiteye göre dağılımları verilmiştir.

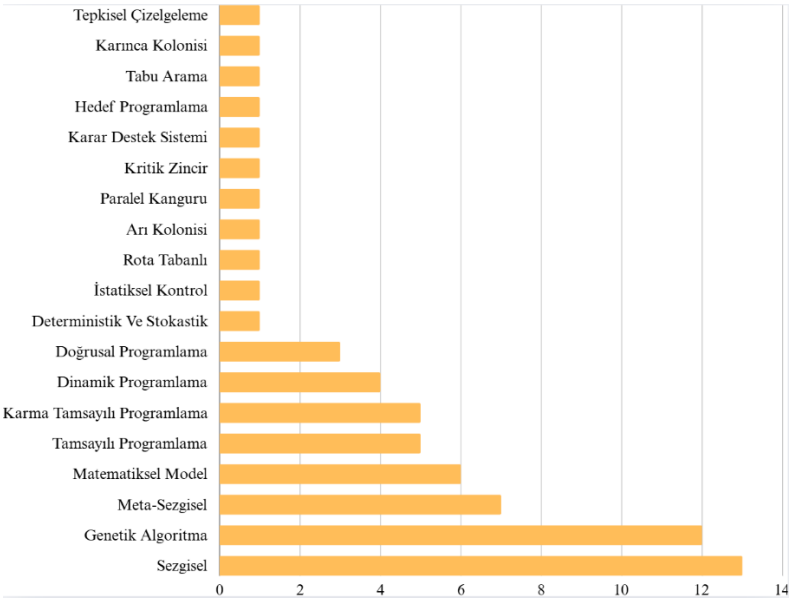



**Şekil 5.** Lisansüstü Tezlerin Üniversiteye Göre Dağılımları

Şekil 5'te üniversitelerin dağılımı göz önünde bulundurulduğunda, proje çizelgeleme konusunda yapılan tezlerin %15'inin (10 adet) Orta Doğu Teknik Üniversitesi'ne ait olduğu görülmektedir. Bu üniversiteyi, %9 (6 adet) ile Sabancı Üniversitesi, %7 (5 adet) ile Boğaziçi Üniversitesi ve yine %7 (5 adet) ile İstanbul Teknik Üniversitesi takip etmektedir. Ayrıca, %6'sı (4 adet) Başkent Üniversitesi, %6'sı (4 adet) Dokuz Eylül Üniversitesi ve %6'sı (4 adet) Marmara Üniversitesi'ne aittir. Bunları ise %4 (3 adet) ile Çukurova Üniversitesi, %4 (3 adet) ile İhsan Doğramacı Bilkent Üniversitesi, %4 (3 adet) ile Yıldız Teknik Üniversitesi, %3 (2 adet) ile İstanbul Üniversitesi, %3 (2 adet) ile İstanbul Kültür Üniversitesi ve %3 (2 adet) ile Eskişehir Osmangazi Üniversitesi izlemektedir. Diğer üniversitelerde ise yalnızca birer lisansüstü çalışmaya rastlanmıştır.

### 3.6. Lisansüstü Tezlerin Kullanılan Yöntemlere Göre Dağılımları

Şekil 6'da proje çizelgeleme konusunda yapılan lisansüstü tezlerin kullanılan yöntemlere göre dağılımları verilmiştir.




**Şekil 6.** Lisansüstü Tezlerin Kullanılan Yöntemlere Göre Dağılımları

Şekil 6'da yöntemlerin dağılımına bakıldığında, lisansüstü tezlerin %19'unun (13 adet) sezgisel yöntem kullandığı görülmektedir. Bunu %18'i (12 adet) genetik algoritma, %10'u (7 adet) meta-sezgisel yöntem, %9'u (6 adet) matematiksel model, %7'si (5 adet) tamsayılı programlama ve %7'si (5 adet) karma tamsayılı programlama takip etmektedir. Diğer yöntemlerden %6'sı (4 adet) dinamik programlama, %4'ü (3 adet) doğrusal programlama şeklinde dağılım göstermektedir. Geriye kalan yöntemlerin ise her birine yalnızca bir çalışmada yer verildiği saptanmıştır.

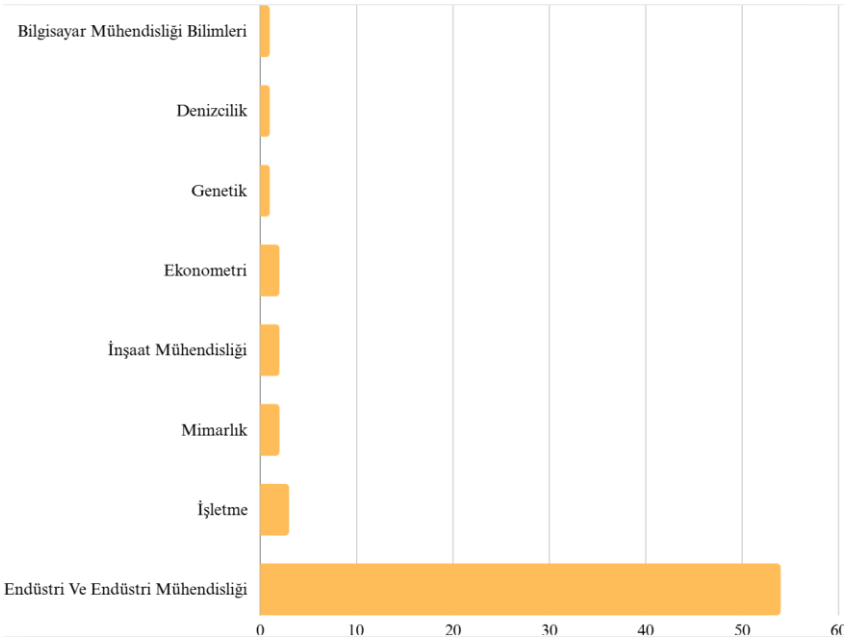
Bu yöntemlerin kullanım nedenlerini ve etkilerini daha iyi anlayabilmek amacıyla literatürde yer alan bazı çalışmalara yer verilmiştir.

Satiç [34], çok kaynak kısıtlı projelerin çizelgelenmesinde sezgisel yöntemlerin uygulanabilirliğini incelemiş ve bu yöntemlerin, özellikle kaynak sınırlamaları altındaki projelerde etkili bir çözüm sağlayabileceğini ortaya koymuştur. Akçay [35], jet eğitim uçaklarının fabrika seviyesi bakımlarında proje çizelgeleme problemi için sezgisel yöntemleri kullanmış ve kaynak kısıtları altında süreçleri optimize etmiştir.Balkaya [36], kaynak kısıtlı proje çizelgeleme problemlerinin optimizasyonunda genetik algoritma yaklaşımını uygulayarak, bu yöntemin performansını test etmiş ve etkili sonuçlar elde etmiştir. Taşıyıcı [37], tek modelli montaj hattı dengeleme problemlerinin çözümünde çok amaçlı genetik algoritmaların etkinliğini ortaya koymuştur. Çalışmada, montaj hatlarında görevlerin dengelemesi sırasında maliyetin minimize edilmesi ve iş yükünün dengelenmesi gibi birden fazla hedefin eş zamanlı olarak optimize edilebileceği gösterilmiştir. Sarı[38], metasezgisel algoritmaların proje çizelgeleme problemlerinin çözümündeki etkinliğini araştırmış ve bu yöntemlerin, kaynak kısıtlarını göz önünde bulunduran çizelgeleme optimizasyonunda önemli faydalar sağladığını ortaya koymuştur. Çınar [39], kaynak kısıtlı bilgi teknolojisi projelerinin çizelgelenmesi problemine yönelik meta-sezgisel algoritmalar kullanarak bir çözüm yaklaşımı sunmuştur. Çalışma, sınırlı kaynaklar altında projelerin

etkin bir şekilde planlanmasını sağlamayı amaçlamış ve bu süreçte meta-sezgisel yöntemlerin gücünden yararlanmıştır. Genel olarak nicel gözlemler kullanılmıştır.

### 3.7. Lisansüstü Tezlerin Kullanılan Konulara Göre Dağılımları

Şekil 7'de proje çizelgeleme konusunda yapılan lisansüstü tezlerin kullanılan konulara göre dağılımları verilmiştir.




**Şekil 7.** Lisansüstü Tezlerin Kullanılan Konulara Göre Dağılımları

Şekil 7'de konulara bakıldığında lisansüstü tezlerin %81'i (54 adet) Endüstri ve Endüstri Mühendisliği olduğu saptanmıştır. Bilim dallarına bakıldığında da Endüstri Mühendisliğinin çoğunlukta olduğu görülmüştür ve bağlantılı oldukları anlaşılmıştır. Takip eden konularda İşletme %4'ü (3 adet), Mimarlık %3'ü (2 adet), İnşaat Mühendisliği %3'ü (2 adet), Ekonometri %3'ü (2 adet), Genetik %2'si (1 adet), Denizcilik %2'si (1 adet) ve Bilgisayar Mühendisliği Bilimleri %2'si (1 adet) olduğu görülmektedir.

### 3.8. Lisansüstü Tezlerin Anahtar Kelime Bulutu

Şekil 8'de proje çizelgeleme konusunda yapılan lisansüstü tezlerin anahtar kelime bulutu verilmiştir.


**Şekil 8.** Lisansüstü Tezlerin Anahtar Kelime Bulutu

İlk olarak tez başlıkları, tez özetleri anahtar kelimeler detaylı olarak incelenip proje çizelgeleme konuları ile ilişkilendirilen konular analiz edilmiştir. Daha sonra, belirlenen konuların sayısına göre bir anahtar kelime bulutu oluşturulmuştur. Bu kelime bulutuna göre, proje çizelgeleme üzerine yapılan lisansüstü tezlerin proje

çizelgeleme, kaynak kısıtlı proje çizelgeleme, proje yönetimi, proje planlama, çoklu proje çizelgeleme, optimizasyon, matematiksel model, bulanık mantık, genetik algoritma, sezgisel yöntem gibi kavramları yoğun bir biçimde araştırıldığı belirlenmiştir. Araştırılan diğer kavramlar, daha detaylı bir şekilde Şekil 8'de sunulmaktadır.

**4. Tartışma ve Sonuç**

Proje çizelgeleme, bir projenin tüm aşamalarını planlayarak iş süreçlerini düzenli ve verimli bir şekilde yürütmeyi amaçlamaktadır. Faaliyetlerin başlangıç ve bitiş zamanları belirlenmesi, kaynakların doğru tahsis edilmesine ve işlerin belirlenen takvime uygun olarak tamamlanmasına olanak tanımaktadır. Bu süreç, projenin maliyetlerini kontrol altında tutarak verimliliği artırır ve gecikmeleri önler. Aynı zamanda, görevlerin ve sürelerin net bir şekilde tanımlanması, proje sürecinde belirsizlikleri azaltır ve iş süreçlerini daha şeffaf hale getirmektedir. Proje çizelgeleme, tüm süreçlerin zamanında ve uyumlu bir şekilde ilerlemesini sağlayarak hedeflere ulaşmayı kolaylaştırır.

Sonuç olarak, proje çizelgeleme, kaynakların en verimli şekilde kullanılmasını sağlar, görevlerin zamanında ve düzenli bir şekilde tamamlanmasına yardımcı olmaktadır. Faaliyetlerin doğru sırayla ve uygun zaman dilimlerinde gerçekleştirilmesi, projenin başarıyla tamamlanmasını desteklemektedir. Zaman yönetimi açısından kritik olan bu süreç, potansiyel gecikmeleri en aza indirerek proje hedeflerine ulaşılmasını sağlamaktadır. Proje çizelgeleme, her aşamanın planlı bir şekilde yönetilmesini sağladığı için projelerin bütçe ve zaman açısından daha kontrollü bir şekilde ilerlemesine imkân tanımaktadır. Bu nedenle, proje yönetiminde önemli bir rol oynamaktadır.

Çalışmanın amacı, proje çizelgelemeyle ilgili yapılan lisansüstü tezler ile ilgili bibliyometrik yöntemler kullanarak bir inceleme yapmaktır. Bu doğrultuda Ulusal Tez Merkezindeki lisansüstü tezler incelenip proje çizelgeleme ile ilgili birçok çalışmanın olduğu sonucu gözlemlenmiştir. Bu çalışmanın amaçları; bir kavramın ne sıklıkla ele alındığını, hangi kavramlara çevrildiğini ve hangi alanlarda daha sık kullandığını belirlemektir.

Bu araştırmada bibliyometrik analiz yöntemi ile Ulusal Tez Merkezinde bulunan proje çizelgeleme konusunda yazılan 66 lisansüstü tez incelenmiştir. Tezlerin 49'u yüksek lisans tezleri oluşturmaktadır. Lisansüstü tez çalışmalarında danışmanların unvanları incelendiğinde 35'inin profesör olduğu gözlemlenmiştir. Bir diğer açıdan proje çizelgeleme konusu üzerine yapılan lisansüstü tezlerin en fazla Orta Doğu Teknik Üniversitesi bünyesinde olduğu gözlemlenmiştir. Sabancı Üniversitesi, Boğaziçi Üniversitesi, İstanbul Teknik Üniversitesi, Başkent Üniversitesi, Dokuzeylül Üniversitesi ve Marmara Üniversitesi takip etmektedir. Ayrıca proje çizelgeleme konusunda yazılan tezlerin bilim dallarına bakıldığında en çok Endüstri Mühendisliği Ana Bilim Dalı'nın çalışma yaptığı gözlemlenmiştir. Bunu İşletme Ana Bilim Dalı, Mimarlık, İnşaat Mühendisliği Ana Bilim Dalı ve Ekonometri Ana Bilim Dalı takip etmektedir. Proje çizelgeleme konusunda yazılan tezlerin konulara bakıldığında lisansüstü tezlerin en çok çalışma Endüstri ve Endüstri Mühendisliği olduğu saptanmıştır. Bilim dallarına bakıldığında da Endüstri Mühendisliğinin çoğunlukta olduğu görülmüştü ve bağlantılı oldukları saptanmıştır.

Proje çizelgelemenin önemi, modern iş dünyasında giderek artmaktadır. Proje yönetimi, zaman ve kaynak verimliliği gibi konularda başarının temel unsurlarından biri olarak kabul edilmektedir. Bu nedenle, akademik araştırmaların yanı sıra, sektör profesyonellerinin de proje çizelgeleme konusunda bilgi ve becerilerini geliştirmeleri büyük önem taşımaktadır. Bu durum, sadece akademik alanda değil, aynı zamanda iş dünyasında da proje çizelgeleme uygulamalarının yaygınlaşmasını ve gelişmesini sağlayacaktır.

Proje çizelgeleme alanındaki araştırmaların artması, yeni yöntem ve teknolojilerin geliştirilmesine katkı sağlayacaktır. Bu bağlamda, üniversiteler, araştırma kurumları ve endüstri iş birliği yaparak proje çizelgeleme konusundaki bilgi birikimini ve uygulama alanlarını genişletebilir. Bu tür iş birlikleri hem akademik dünyada hem de iş dünyasında proje çizelgeleme uygulamalarının daha etkin ve verimli hale gelmesini sağlayacaktır.

Bu çalışmada Türkiye'de 1991-2024 yılları arasında proje çizelgeleme konusunda yapılan lisansüstü tezler incelenmiştir. Elde edilen bulgulara göre, proje çizelgeleme çalışmaları özellikle mühendislik ve işletme alanlarında yoğunlaşmıştır; ancak sağlık, tarım veya bilişim gibi alanlarda sınırlı sayıda çalışma olduğu görülmüştür. Bu durum, proje çizelgelemenin disiplinler arası bir yaklaşımla ele alınması gerektiğini ve bazı alanlarda daha fazla araştırmaya ihtiyaç duyulduğunu göstermektedir

Çalışmanın sonuçları Türkiye'de bu konuda yapılan tezlerin sayıca artmakla birlikte hâlâ gelişmiş ülkelerdeki literatür yoğunluğunun gerisinde olduğunu ortaya koymaktadır. Bu durum, Türkiye'deki üniversitelerin proje

çizelgeleme konusuna yeterince odaklanıp odaklanmadığına dair tartışmalara da zemin hazırlamaktadır. Ayrıca, proje çizelgeleme yöntemlerinin özellikle kaynak planlaması, zaman yönetimi, maliyet kontrolü ve risk analizi gibi alanlarda stratejik bir araç olarak kullanılabileceği vurgulanmalıdır.
Sonuç olarak, bu çalışma yalnızca tez sayılarını ve eğilimleri betimlemekle kalmamakta, aynı zamanda proje çizelgeleme konusundaki araştırma eksikliklerine dikkat çekerek literatüre yön verici önerilerde bulunmaktadır. Bu bağlamda, elde edilen bulgular proje çizelgeleme araştırmalarının hangi alanlarda yoğunlaştığını, hangi alanlarda eksiklikler olduğunu ve bu eksikliklerin nasıl giderilebileceğine yönelik yol gösterici niteliktedir.

**Yazar Katkıları:** E. Ş. P., E. G. ve T. E. çalışmanın tüm aşamalarına eşit katkı sağlamışlardır. Literatür taraması, veri toplama, analiz, yorumlama ve yazım süreçlerinde tüm yazarlar ortak sorumluluk üstlenmiştir.

**Finansman:** Bu araştırma dışarıdan fon almadı.

**Çıkar çatışmaları:** Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

## Kaynaklar

[1] Durucasu, H., İcan, Ö., Karamaşa, Ç., Yeşilaydın, G., & Gülcan, B. (2015). Bulanık CPM yöntemiyle proje çizelgeleme: İnşaat sektöründe bir uygulama.
[2] Ercan, P., & Tanrıöver, Ö. Ö. (2020). İş paketi sürelerinin belirli olduğu kaynak kısıtlı proje planlama problemi için öncelik kuralı. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, *35*(3), 1537-1550.
[3] Aksoy, A., Akansel, M., Atalay, C., Çamlıbel, A. M., Yaşar, D., Keseroğlu, D., & Vanlıoğlu, S. (2019). Proje Yönetiminde Zaman ve Maliyet Odaklı Bütünleşik Planlama Yaklaşımı ve Bir Uygulama. *Journal of Entrepreneurship and Innovation Management*, *8*(1), 1-20.
[4] Atlı, Ö., & Aydın, S. (2021). Çok Modlu Kaynak Kısıtlı Proje Çizelgeleme Problemlerinin Belirsizlik Ortamında Modellenmesi. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, *21*(3), 586-605.
[5] Özköse, H., & Gencer, C. (2019). Proje Planlama ve Çizelgelemede Genetik Algoritma Tabanlı Bir Yöntem ile Kritik Yolun-Proje Tamamlanma Zamanının Tespiti ve Zaman-Maliyet Analizi. *Bartın Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *10*(20), 278-300.
[6] Çorumlu, V., Atalay, K. D., & Dinler, E. (2024). Proje çizelgelemede bulanık doğrusal programlama ile yeni bir yöntem önerisi: Yazılım projesinde uygulama. *Journal of Turkish Operations Management*, *8*(1), 20-38.
[7] Özdamar, L., & Ulusoy, G. (1995). A survey on the resource-constrained project scheduling problem. *IIE transactions*, *27*(5), 574-586.
[8] Ulusoy, G. (2002). Proje planlamada kaynak kısıtlı çizelgeleme.
[9] Paksoy, Ö. G. D. S., & Uzun, A. (2008). Genetik algoritma ile kaynak kısıtlı proje çizelgeleme. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *17*(2), 345-362.
[10] Bettemir, Ö. H., & Çakmak, D. (2021). Tüm arama uzayını tarayarak küçük ölçekli kaynak kısıtlı proje çizelgeleme problemlerinin çözümü.
[11] Soysal, S., Dengiz, B., & Atalay, K. (2021). Belirsizlik altında kaynak kısıtlı çok modlu çoklu proje çizelgeleme. *Journal of Turkish Operations Management*, *5*(1), 598-614.
[12] Kolaylioğlu, Ö. (2006). *İnşaat sektöründe proje yönetimi ve proje yöneticisi* (Master's thesis, Dokuz Eylul Universitesi (Turkey)).
[13] Çinar, G. (2011). *Meta-sezgisel algoritmalar kullanılarak kaynak kısıtlı bilgi teknolojisi projelerinin çizelgelenmesi* (Master's thesis, Fen Bilimleri Enstitüsü).
[14] Özdemir, G. Ö., & Karacabey, A. A. (2006). Kısıtlı kaynaklarla proje çizelgelemesi problemlerinde kullanılan genetik algoritma yöntemleri ve bunların karşılaştırılması.
[15] Joushani, M. (2022). *Değişken yoğunluklu kaynak kısıtlı proje çizelgeleme için matematiksel modelleme ve genetik algoritma yaklaşımı* (Master's thesis, Bursa Uludağ Üniversitesi).
[16] Taş, M., Özlemiş, Ş. N., Hamurcu, M., & Eren, T. (2017). Analitik hiyerarşi prosesi ve hedef programlama karma modeli kullanılarak monoray projelerinin seçimi. *Harran Üniversitesi Mühendislik Dergisi*, *2*(2), 24-34.
[17] Hamurcu, M., & Eren, T. (2018). Kent içi ulaşım için bulanık ahp tabanlı vıkor yöntemi ile proje seçimi. *Engineering Sciences*, *13*(3), 217-228.
[18] Pınarcı, E. Ş., Vuruşkan, C. T., Güven, E., & Eren, T. (2024). Türkiye'de Ekip Çizelgeleme Konulu Lisansüstü Tezlerin Bibliyometrik Analizi. *Harran Üniversitesi Mühendislik Dergisi*, *9*(2), 118-130.
[19] Lawani, S. M. (1981). Bibliyometri: Kuramsal temelleri, yöntemleri ve uygulamaları. *Libri*, *31* (Cehresband), 294-315.

[20] Genç, G., & Sarı, M. E-sağlık alanındaki bilimsel yayınların bibliyometrik analiz yöntemi ile incelenmesi. *Sosyal Araştırmalar ve Yönetim Dergisi*, (1), 58-72.
[21] Merigó, J. M., & Yang, J. B. (2017). A bibliometric analysis of operations research and management science. *Omega*, *73*, 37-48.
[22] Aksungur, B. N., Sever, H., & Güven, E. & Eren, T. (2024). İnsansız Hava Araçları Konulu Lisansüstü Tezlerin Bibliyometrik Analizi. *Masyarakat, Kebudayaan & Politik*, *37*(1).
[23] Öztürk, N., & Kurutkan, M. N. (2020). Kalite yönetiminin bibliyometrik analiz yöntemi ile incelenmesi. *Journal of Innovative Healthcare Practices*, *1*(1), 1-13.
[24] Sanlı, Y. B., Baltacı, F., Güven, E., & Eren, T. (2024). Siber Güvenlik Çalışmaları Üzerine Bibliyometrik Analiz. *Bilişim Teknolojileri Dergisi*, *17*(3), 223-229..
[25] Gaferoğlu, İ., Kaya, S., Kalemler, Y. B., Güven, E., & Eren, T. (2024). Tsunami Konulu Lisansüstü Tezlerin Bibliyometrik Analizi. Urban 21 Journal, 2(2), 153-164.
[26] Birinci, H. G. (2008). Turkish Journal of Chemistry'nin bibliyometrik analizi. *Bilgi Dünyası*, *9*(2), 348-369.
[27] Beşel, F. (2017). Türkiye'de maliye alanında yapılmış lisansüstü tezlerin bibliyometrik analizi (2003-2017). International Journal of Public Finance, 2(1), 27-62.
[28] Öztürk, S., Taş, B. S., Kaplan, D., Keskin, S., Çoruk, E. N., Güven, E., & Eren, T. (2024). Sağlık turizmi konulu lisansüstü tezlerin bibliyometrik analizi. Sağlıkta Performans ve Kalite Dergisi, 21(3), 184-204.
[29] Zeren, D., & Kaya, N. (2020). Dijital pazarlama: Ulusal yazının bibliyometrik analizi. *Çağ Üniversitesi Sosyal Bilimler Dergisi*, *17*(1), 35-52.
[30] Duran, G., & Çelikkaya, S. (2019). Türkiye'de Lojistik Üzerine Yapilmiş Lisansüstü Tezlerin Bibliyometrik Analizi. GÜ İslahiye İibf Uluslararası E-Dergi, 3(3), 152-167.
[31] Küpçüoğlu, E., Alakaş, H. M., & Eren, T. (2024). Blok Zincir Üzerine Yazılan Yüksek Lisans ve Doktora Tezlerinin Bibliyometrik Analizi. Bilişim Teknolojileri Dergisi, 17(4), 281-293.
[32] Budd, J. M. (1988). A bibliometric analysis of higher education literature. *Research in Higher Education*, *28*, 180-190.
[33] Ercan, S., Metin, B. C., & Düzdar, İ. (2005). Endüstri mühendisliğine güncel bir bakış. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, *4*(7), 1-18.
[34] Satiç, U. (2014). Çok kaynak kısıtlı projelerin sezgisel yöntemlerle çizelgelenmesi (Master's thesis, Fen Bilimleri Enstitüsü).
[35] Akçay, A. (2017). Jet eğitim uçaklarının fabrika seviyesi bakımlarında proje çizelgeleme (Master's thesis, Fen Bilimleri Enstitüsü).
[36] Balkaya, A. H. (2011). Kaynak kısıtlı proje çizelgeleme problemlerinin genetik algoritma yaklaşımıyla optimizasyonu (Master's thesis, Dokuz Eylul Universitesi (Turkey)).
[37] Taşıyıcı, G. (2004). Tek modelli montaj hattı dengeleme problemlerinin çözülmesi için çok amaçlı genetik algoritma tasarımı.
[38] Sarı, T. (2008). Metasezgisel Yöntemlerle Proje Çizelgeleme Optimizasyonu (Doctoral dissertation, Marmara Universitesi (Turkey)).
[39] Çinar, G. (2011). Meta-sezgisel algoritmalar kullanılarak kaynak kısıtlı bilgi teknolojisi projelerinin çizelgelenmesi (Master's thesis, Fen Bilimleri Enstitüsü).

Journal of Statistics & Applied Sciences, Issue - 11
ISSN 2718-0999

*Letter to the Editor*

Received: date:20.03.2025
Accepted: date:17.06.2025
Published: date:30.06.2025

# P-Hacking in Scientific Research: Is the Reliability of Scientific Results at Risk?

**Sadi Elasan** [1*]

[1]Department of Medical Sciences, Van Yuzuncu Yil University, Van, Türkiye; sadielasan@yyu.edu.tr
Orcid: 0000-0002-3149-6462
*Correspondence: sadielasan@yyu.edu.tr

The increasing prevalence of statistical manipulations, known as p-hacking, poses a significant challenge to the reliability of scientific findings. This issue arises when researchers repeatedly adjust data analyses until a p-value falls below 0.05, often due to publication pressure or insufficient statistical training. Such practices compromise the validity of research outcomes, particularly in fields like medicine and biomedical sciences, where false conclusions can have serious implications.

P-hacking commonly involves selective data reporting, data dredging, and multiple hypothesis testing without proper corrections, leading to inflated false positive rates [1]. For instance, when 20 independent hypothesis tests are conducted at a 5% significance level, the probability of obtaining at least one false positive result rises to approximately 64% (1-0.95^20=0.64). If 100 hypotheses are tested, this probability escalates to 99% (1-0.95^100=0.99). Small sample sizes further exacerbate this issue, with false positive rates reaching up to 50% in studies with n<30. These inflated error rates mislead researchers and clinicians, potentially affecting medical decision-making [1-2].

To mitigate p-hacking, a multifaceted approach is necessary. First, strengthening statistical training is crucial. Researchers should move beyond reliance on p-values and incorporate alternative metrics such as effect sizes, confidence intervals, and Bayesian methods. For example, a study with p<0.05 but a negligible effect size (e.g., Cohen's d=0.2) may lack real-world significance [3-5].

Second, open science practices, including preregistration and data sharing, should be widely adopted. Journals implementing open data policies, such as Psychological Science and PLOS Biology, have reported a reduction in questionable p-values, with Psychological Science observing a 40% decline in studies reporting p-values just below 0.05 (1-2-3-4).

Third, proper statistical corrections, such as the Bonferroni and Benjamini-Hochberg methods, should be routinely applied. When 20 hypotheses are tested, the Bonferroni correction adjusts the significance threshold to 0.05/20=0.0025, reducing false positive risks. Additionally, prioritizing effect size metrics like Hedges' g and reporting confidence intervals can improve result interpretation.

Finally, addressing publication bias is essential. Journals should encourage the publication of null results and studies with robust methodological designs, regardless of significance. Moreover, training programs on ethical statistical practices should be expanded, as researchers with formal statistical education are less likely to engage in p-hacking.

In conclusion, tackling p-hacking requires improved statistical training, open science initiatives, and methodological rigor. By revising editorial policies and promoting best practices, journals can play a pivotal role in preserving the integrity of scientific research.

**Citation :** S. Elasan, "P-hacking in scientific research: Is the reliability of scientific results at risk?" *Journal of Statistics and Applied Sciences*, no. 11, pp. 70–71, 2025. [Editor's note]

## References


[1] J. P. Simmons, L. D. Nelson, and U. Simonsohn, "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant," Psychol. Sci., vol. 22, no. 11, pp. 1359–1366, 2011, doi: 10.1177/0956797611417632.

[2] M. L. Head, L. Holman, R. Lanfear, A. T. Kahn, and M. D. Jennions, "The extent and consequences of P-hacking in science," PLoS Biol., vol. 13, no. 3, p. e1002106, 2015, doi: 10.1371/journal.pbio.1002106.

[3] D. J. Benjamin et al., "Redefine statistical significance," Nat. Hum. Behav. vol. 2, pp. 6–10, 2018, doi: 10.1038/s41562-017-0189-z.

[4] J. M. Wicherts, M. Bakker, and D. Molenaar, "Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results," PLoS ONE, vol. 6, no. 11, p. e26828, 2011, doi: 10.1371/journal.pone.0026828.

[5] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor, "The preregistration revolution," Proc. Natl. Acad. Sci. U.S.A., vol. 115, no. 11, pp. 2600–2606, 2018, doi: 10.1073/pnas.1708274114.

30.06.2025

Journal of Statistics & Applied Sciences, Issue – 11
ISSN 2718-0999