BAREKENG: Journal of Mathematics and Its Applications
December 2024 Volume 18 Issue 4 Page 2517–2530
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss4pp2517-2530

RETROSPECTIVE ANALYSIS IN HYPOTHESIS TESTING TO EVALUATE INDONESIA'S GINI RATIO AFTER COVID-19 PANDEMIC

Karunia Eka Lestari [1*], Fitriani Agustina[2], Mokhammad Ridwan Yudhanegara[3], Edwin Setiawan Nugraha[4], Sisilia Sylviani[5]

[1,3]Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Singaperbangsa Karawang
Jln. H. S. Ronggowaluyo Teluk Jambe Timur, Karawang, 41361, Indonesia

[2]Department of Mathematics Education, Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia
Jln. Dr. Setiabudhi No 229, Bandung, 40154, Indonesia

[4]Department of Actuarial Science, Faculty of Business, President University
Jln. Ki Hajar Dewantara, Bekasi, 17550, Indonesia

[5]Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Padjajaran
Jln. Ir. Soekarno km 21, Sumedang, 45363, Indonesia

*Corresponding author's e-mail: * **karunia@staff.unsika.ac.id***

ABSTRACT

Article History:

Received: 18th, April 2024
Revised: 8th, June 2024
Accepted: 4th,August 2024
Published:14th, October 2024

Keywords:

Gini Ratio;
Hypothesis Testing;
Statistical Power;
Retrospective Analysis.

*The study highlighted three essential roles of retrospective analysis in hypothesis testing, particularly as a priori analysis, post hoc analysis, and sensitivity analysis. These approaches were applied to the Gini ratio data sourced from the National Socioeconomic Survey Indonesia 2023 to examine the income inequality level in Indonesia. The sample size, statistical power, and effect size for the one-sample t-test are evaluated by aid G*Power software. The test results show that for a sample size of 10, at the 95% confidence interval, there is not enough evidence to show that the Gini ratio in 2023 is smaller than 0.4. A retrospective analysis using G*power software reveals that for a sample size of 20 at the same confidence interval, there is enough evidence to suggest that the Gini ratio is statistically significant at less than 0.4 with a power of analysis of 90.8% and an effect size of 0.76. This study has important implications in hypothesis testing, especially in retrospective analysis, since understanding the effect of sample size and effect size makes it possible for academics or practitioners to optimize hypothesis testing and generate more accurate and reliable test results.*

How to cite this article:

K. E. Lestari., F. Agustina., M. R. Yudhanegara, E. S. Nugraha., and S. Sylviani "RETROSPECTIVE ANALYSIS IN HYPOTHESIS TESTING TO EVALUATE INDONESIA'S GINI RATIO AFTER COVID-19 PANDEMIC," *BAREKENG: J. Math. & App.*, vol. 18, iss. 4, pp. 2517-2530, December, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Hypothesis testing has a central role in decision-making from the inferential statistics framework. The hypothesis test is based on sample data that can provide information on whether an unknown population parameter differs sufficiently from its hypothesized value [1]. Furthermore, to evaluate the research hypothesis, the hypothesis is restated in terms of two statistical hypotheses, namely the null and the alternative hypothesis. The null hypothesis, denoted by H_0, is a statement regarding the hypothesized population parameter and is made to be rejected. The hypothesis contains a statement regarding no effect or no difference. The alternative hypothesis, denoted by H_1 or H_a, is a conjecture created in light of what is being investigated and found by considering what the researcher wants to accept that differs from the hypothesized population parameters [2]. Such a hypothesis usually reflects the researcher's underlying research hypothesis and specifies an alternative population condition that is "supported" or "affirmed" following the rejection of H_0.

The null hypothesis takes a leading role in statistical hypothesis testing, as it is the hypothesis that is assumed to be true and formally tested. It also determines the sampling distribution and forms the basis for the final decision to "reject" or "retain." Since research hypothesis statements generally predict the existence of an effect or difference concerning whatever is being studied, the null hypothesis will generally be the one the researcher expects to reject. However, suppose the sample data is not significantly different from what is stated in the null hypothesis, beyond reasonable doubt. In that case, this suggests that the null hypothesis cannot be ignored [3].

When the test results lead one to retain the null hypothesis, or H_0 is not rejected, as a consequence, the researcher does not have enough evidence to reject H_0 and thus cannot support the H_a claim. It implies that there is not enough evidence to state that the sample was drawn from a population that has parameters that are significantly different, smaller than, or larger than those specified in H_0. To make the conclusions of such test results correct, in this case, the researcher should decide whether the true population parameter is really not sufficiently different from the hypothesized value and not just that there is not a large enough sample to detect whether or not the parameter is sufficiently different from the hypothesized value [4]. For this reason, the idea of retrospective analysis in hypothesis testing comes in handy. This analysis provides information on the sample size required to detect minimal differences in population parameters. A sample that is not large enough to find a specific difference will indicate that the analytical results are less robust. In other words, the sample size is not large enough to find an effect of a specific size if such an effect exists.

Up to this point, the phenomenon in the population being statistically tested is considered not to exist (if H_0 is true) or existing (if H_0 is false). The nonexistence of the phenomenon implies some specific values for the population parameters. In short, when H_0 is false or rejected, the hypothesis is false at a certain level, referred to as the effect size. The effect size is a non-zero value of deviation from the nonexistence of the hypothesized phenomenon or effect in the population [5]. Thus, effect size measures the magnitude and practical significance of differences or relationships found in hypothesis testing. It allows us to determine the meaningfulness of the results and can be interpreted in various ways, depending on the statistical analysis used.

Considering that the one-sample t-test is the simplest test statistical tool and is widely applied in various fields, we will confine our attention to establishing (1) a minimum sample size, (2) a statistical power test, and (3) an effect size for the one-sample t-test. These calculations were performed by G*Power, which is free software to calculate statistical power. This method is not new, but we provide it in a different way to allow academics or practitioners to analyze the power of a test involving a one-sample t-test more simply and easily.

The focus of the research data in this paper is the Gini ratio. The Gini ratio is a major concern after the Corona Virus Disease 2019 (COVID-19) pandemic in Indonesia because Indonesia has been one of the countries affected by the COVID-19 pandemic since March 2020 until now. Fluctuations in cases have increased every day and have an impact on every sector, especially the economic sector [6]. The COVID-19 pandemic has weakened the economy on the production side, resulting in a decrease in national income. The decline in national income results in a decline in economic development [7]. So the evaluation of Indonesia's Gini Ratio after the COVID-19 Pandemic is very important.

2. RESEARCH METHODS

This section briefly provides conceptual theory and methods for data analysis, including a one-sample t-test, power analysis, sample size, and effect size.

2.1 The One-sample t-Test

In practice, one frequently has to estimate the population mean μ when the population standard deviation σ is unknown. Under this condition, one can conduct hypothesis tests related to one-sample mean using the one-sample t-test. The one-sample t-test is a parametric procedure for determining whether an unknown population mean is significantly different from a specific or hypothesized value.

Given random samples $X_1, X_2, \cdots, X_n$ that are mutually independent and follow the normal distribution with unknown μ and σ, then based on the Central Limit Theorem, the random variable T follows the t-distribution with degrees of freedom $\nu = n - 1$, that is

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \tag{1}$$

Here, $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is the sample distribution mean, and the standard deviation of the sample distribution is $S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$.

The t-distribution is also known as the Student-t distribution. The distribution is named after William Sealy Gosset who first published it in English in 1908 in the scientific journal Biometrika using the pseudonym "Student" because the company supervisor where Gosset worked preferred his staff to use pen names when publishing scientific papers related to company data to maintain company trade secrets [8]. The t-distribution resembles the standard normal distribution in that it is unimodal, symmetrically distributed, and has a bell-shaped curve centered around 0. However, unlike the z-distribution which has a variance of 1, the t-distribution has a variance equal to its independent degree which is more than 1. Therefore, the t-distribution is heavy-tailed compared to the z-distribution. This indicates that the t-distribution tends to have many outliers with extreme values. The thicker the tail, the more likely it is that there are one or more disproportionate values in the sample. The shape of the t-distribution depends on the degrees of freedom. The larger the degrees of freedom, the higher the curve and the thinner the tails. In other words, the larger the degrees of freedom, the closer the curve will be to the z-distribution.

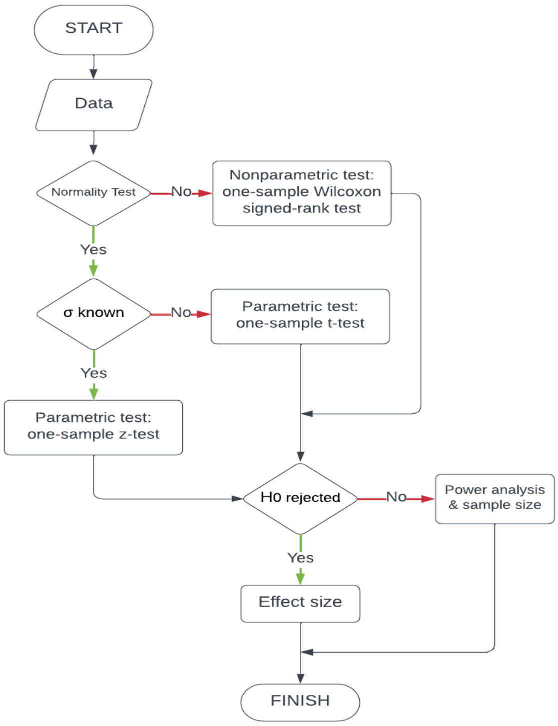


Figure 1. Flowchart of Statistical Test for one-Sample Mean with an Interval or Ratio Scale

The one-sample t-test is commonly employed on small sample sizes ($n < 30$), with the assumptions that:

1. The data is continuous with an interval or ratio scale.
2. The sample is randomly selected from the population of interest, such that each individual (observation) in the population has an equal chance of being selected as a sample.
3. Each observation is mutually independent.
4. The population standard deviation σ is unknown.
5. The data follows a normal distribution.

Based on **Equation (1)**, the t-test statistic is calculated by

$$t_{stat} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \tag{2}$$

Broadly defined, there are seven steps required for hypothesis testing using the one-sample t-test, as detailed in the next algorithm.

Algorithm 1. Statistical hypothesis testing steps.

Step 1: Formulate the statistical hypothesis.
Step 2: State the assumptions (if any) and check the conditions.
Step 3: Determine the critical value at the selected significance level α.
Step 4: Calculate the test statistic.
Step 5: Calculate the p-value.
Step 6: Determine the test criteria or critical region.
Step 7: Conclude.

The p-value of a one-sample t-test is determined by taking the cumulative distribution function of the t-distribution, as follows: for $t > 0$

$$P(T \leq t) = P\left(T \leq \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}\right) = 1 - \frac{1}{2}\left(\frac{B[x(t); \mathrm{a}; \mathrm{b}]}{B[\mathrm{a}; \mathrm{b}]}\right), \tag{3}$$

with $x(t) = \frac{v}{t^2 + v} \in [0,1]$, $a = \frac{v}{2}$, and $b = \frac{1}{2}$. Here, $B[\mathrm{a}; \mathrm{b}]$ is a Beta function, that is

$$B[\mathrm{a}; \mathrm{b}] = \int_0^1 u^{a-1}(1-u)^{b-1} du\,, \tag{4}$$

and $B[x(t); \mathrm{a}; \mathrm{b}]$ is an incomplete Beta function [9], such that

$$B[x(t); \mathrm{a}; \mathrm{b}] = \int_0^{x(t)} u^{a-1}(1-u)^{b-1} du\,. \tag{5}$$

2.2 Statistical Power Analysis and Sample Size

Consider the t-test statistic in **Equation (2)**. Under the assumption H_0, the t statistic follows the Student's t-distribution with free degrees $v = n - 1$. Meanwhile, under the alternative hypothesis (H_a), the t statistic follows noncentral Student's t-distribution with free degrees $v = n - 1$ and noncentral parameters [10], that is

$$\lambda = \frac{\delta}{s/\sqrt{n}}, \tag{6}$$

where $\delta = \mu_\mathrm{a} - \mu_0$ denotes the difference between the true population mean μ_a and the hypothesized population mean μ_0. The statistical power test $(1 - \beta)$ for detecting specific differences δ at a given significance level α, is determined as follows:

$$1-\beta = \begin{cases} 1 - T\left(t_{(\alpha/2;\nu)};\nu,\lambda\right) + T\left(-t_{(\alpha/2;\nu)};\nu,\lambda\right) & \text{, if two} - \text{tailed} \\ 1 - T\left(t_{(\alpha;\nu)};\nu,\lambda\right) & \text{, if right} - \text{tailed} \\ T\left(-t_{(\alpha;\nu)};\nu,\lambda\right) & \text{, if left} - \text{tailed} \end{cases} \tag{7}$$

Here, $T(t;\nu,\lambda)$ is the cumulative distribution function of the non-centered Student's t-distribution with degrees of freedom $\nu = n-1$ and noncentral parameters λ at the value of $t = \{-t_{\alpha/2}, -t_{\alpha}, t_{\alpha}, t_{\alpha/2}\}$. Recall that $T \sim T(\nu,\lambda)$ is non-centered t-distribution, and its cumulative distribution function is

$$T(t;\nu,\lambda) = P(T \le t) = \begin{cases} \dfrac{1}{2}\displaystyle\sum_{j=0}^{\infty} \frac{1}{j!}\left(-\lambda\sqrt{2}\right)^{j} e^{\frac{-\lambda^2}{2}} \frac{\Gamma\left(\frac{j+1}{2}\right)}{\sqrt{\pi}} \frac{B\left[\frac{\nu}{\nu+t^2};\frac{\nu}{2};\frac{j+1}{2}\right]}{B\left[\frac{\nu}{2};\frac{j+1}{2}\right]} & , t \ge 0 \\ 1 - \dfrac{1}{2}\displaystyle\sum_{j=0}^{\infty} \frac{1}{j!}\left(-\lambda\sqrt{2}\right)^{j} e^{\frac{-\lambda^2}{2}} \frac{\Gamma\left(\frac{j+1}{2}\right)}{\sqrt{\pi}} \frac{B\left[\frac{\nu}{\nu+t^2};\frac{\nu}{2};\frac{j+1}{2}\right]}{B\left[\frac{\nu}{2};\frac{j+1}{2}\right]} & , t < 0 \end{cases} \tag{8}$$

with Γ is the Gamma function, such that

$$\Gamma(x) = \int_0^{\infty} u^{x-1} e^{-u}\, du \tag{9}$$

The minimum sample size n_β required to detect at least a difference δ from the hypothesized population mean μ_0 at a given $1-\beta$ power level can be determined by solving **Equation (7)** for a given value of β **[11]**. A power value of $1-\beta = 0.9$ is commonly considered adequate. A value of 0.9 indicates that one has a 90% chance of detecting a difference between the actual population mean μ_a and the target or hypothesized population mean μ_0 when such a difference exists **[11]**. Likewise, determining the minimum difference δ that can be attained at a given $1-\beta$ power level for each sample size n is obtained iteratively using these equations **[12]**. Since, the cumulative distribution function of the non-centered t-distribution in **Equation (8)** is relatively sophisticated. Therefore, in this study, it was calculated using the G*power software.

2.3 Effect size

Interpreting the effect size in hypothesis testing is essential for understanding the practical significance of the results obtained and providing a more comprehensive understanding of the research findings. When H_0 is rejected, the population effect size of the one-sample t-test is determined by the Cohen index below **[5]**

$$\delta_{ES} = \frac{|\mu_a - \mu_0|}{\sigma} \tag{10}$$

Lo and Chen **[13]** estimated δ_{ES} by

$$\hat{\delta}_{ES} = \frac{|\bar{x} - \mu_0|}{s}, \tag{11}$$

where μ_a is the true population mean, μ_0 is the hypothesized population mean, $\bar{x}$ is the sample mean, and s is the sample standard deviation. The magnitude of the effect size is interpreted according to the three criteria: (1) small effect size, if $0{,}2 < \hat{\delta}_{ES} \le 0{,}5$; (2) medium, if $0{,}5 < \hat{\delta}_{ES} \le 0{,}8$; and (3) high, if $\hat{\delta}_{ES} > 0{,}8$ **[5]**.

2.4 G*Power Software for Retrospective Analysis

G*Power is a free statistical power analysis software for various statistical tests such as exact test, z, t, F, χ^2, and related tests **[14]**. G*power provides power value, sample size, and effect size calculators including graphical options. The software and its manual book can be downloaded for free on the website https://www.psychologie.hhu.de. G*Power software offers five different types of statistical power analysis, consisting of a priori, compromise, criterion, post hoc, and sensitivity [15]. In terms of retrospective analysis, the current study will deal with three of the five analyses, specifically:

1. A priori, the minimum necessary sample size $\boldsymbol{n_\beta}$ is computed as a function of $\mathbf{1-\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\delta_{ES}}$ to evaluate the sample size that has been used in hypothesis testing.
2. Post hoc, statistical power level $\mathbf{1-\beta}$ is computed as a function of $\boldsymbol{\alpha}$, $\boldsymbol{\delta_{ES}}$, and $\boldsymbol{n_\beta}$ to reveal the statistical test capability in detecting the differences of the variable interest if such differences genuinely exist.
3. Sensitivity, population effect size $\boldsymbol{\delta_{ES}}$ is computed as a function of $\boldsymbol{\alpha}$, $\mathbf{1-\beta}$, and $\boldsymbol{n_\beta}$ to assess the practical significance of a hypothesis test.

To allow users to explore the parameter space relevant to the power analysis being performed, an arbitrary parameter (α, $1-\beta$, δ_{ES}, or n_β) may be plotted as a function of the other parameters [14].

2.5 Gini Ratio

The Gini ratio measures the aggregate income inequality in a region [16]**.** This index is a summary statistic that measures how equitably resources are distributed within a population [17]. Statistics Indonesia (BPS) employs expenditure data as a proxy for income sourced from the National Socioeconomic Survey (Susenas) and adapted the World Bank's Gini ratio and distribution indicators to measure the inequality rate in Indonesia. The Gini ratio (GR) is calculated by the formula [18].

$$GR = 1 - \sum_{i=1}^{n} fp_i(Fc_i + Fc_{i-1}). \tag{12}$$

Here, fp_i is the frequency of population in the i-th expenditure class, Fc_i and Fc_{i-1} respectively are cumulative frequency of total expenditure in expenditure class i and $i-1$. The Gini ratio value lies in the interval [0,1]. A Gini ratio value closer to 1 indicates a higher level of inequality, and the reverse is true.

The Gini ratio is based on the Lorenz curve, which is a cumulative expenditure curve that compares the distribution of consumption expenditure values with a uniform distribution that represents the cumulative percentage of the population [19]. Thus, the basic idea of calculating the Gini ratio derives from gauging the area of a curve that describes the distribution of income for the entire income group, as illustrated in Figure 2. The horizontal line (red line) in Figure 2 represents the perfect equality. The magnitude of the inequality is depicted as the shaded area A. The Gini coefficient or Gini ratio is defined as the ratio between the shaded area A and the triangular area OPE. It implies that if income is perfectly equally distributed (uniformly distributed), then all points will lie on the diagonal line [20]. As a consequence, the shaded area will be zero since it is equal to the diagonal line. Thus, the ratio is equal to zero. Conversely, if only one entity receives all the income, then the shaded area will be equal to the triangular area, so the Gini ratio is one.

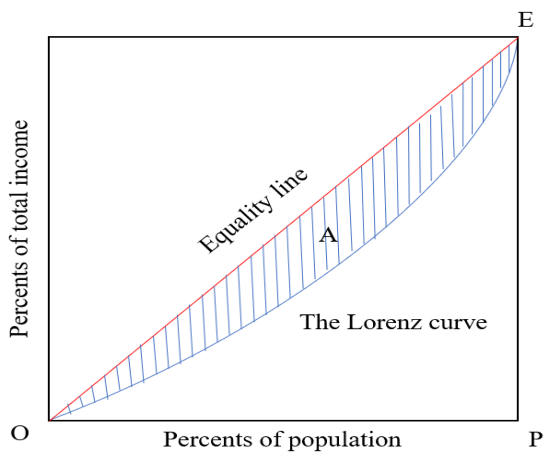


Figure 2. Lorenz Curve as a Graphical Representation of Income Inequality [21]

Furthermore, income inequality is interpreted according to the Gini ratio criteria [22] as follows

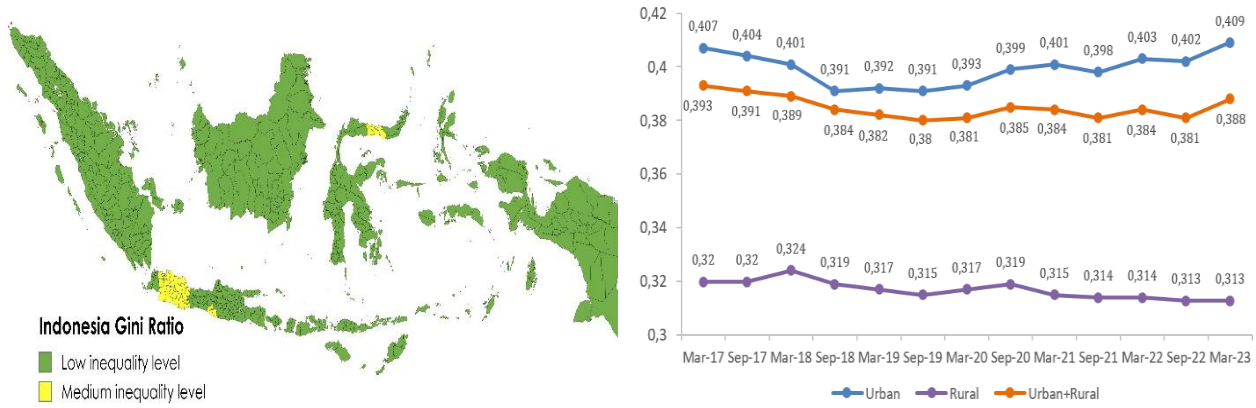
Table 1. Gini ratio criterion and interpretation

Gini ratio (GR)	Income distribution interpretation
$GR = 0$	Perfect equality
$0 < GR < 0.4$	Low inequality level
$0.4 \le GR < 0.5$	Medium inequality level
$0.5 \le GR < 1$	High inequality level
$GR = 1$	Perfect inequality (monopolized by one entity/company)

3. RESULTS AND DISCUSSION

3.1 Data description

Based on Statistics Indonesia data, the Gini Ratio has continued to decline nationwide from March 2017 to September 2019. It indicates that during this period, there was an improvement in the equitable distribution of population expenditure in Indonesia. However, the Gini ratio increased in March 2020 and September 2020 as a direct consequence of the COVID-19 pandemic. Figure 3 reveals that after 2020, the Gini ratio fluctuated until it reached 0.388 in March 2023.



(a) (b)

Figure 3. The data acquired from the National Socioeconomic Survey, (a) Indonesia's Gini ratio Distribution Map and (b) Gini ratio trends from March 2017 to March 2023

The government has taken extensive strategies to revitalize the economy due to the COVID-19 pandemic, including ensuring the inflation rate, food price stability, and sustainable social welfare distributions [23]. These strategies had a positive economic impact on the poverty rate in March 2023. Statistics Indonesia data reported that the poverty rate in Indonesia in March 2023 decreased by 0.21% compared to September 2022, which was 9.36%. In particular, Statistics Indonesia claims that the expenditure of the population in March 2023 was at a low inequality level [24].

3.2 Hypothesis Testing

The rest of this study focuses on hypothesis testing to verify Statistics Indonesia's claim. For this purpose, data on the Gini ratio in March 2023 is taken from the Statistics Indonesia official website, available at https://www.bps.go.id. The website provides Gini ratio data for March 2023, organized by province (34 provinces) and region (urban, rural, urban-rural). Furthermore, a random sample of 10 out of 34 provinces from urban-rural regions was selected, as reported in Table 2.

Table 2. Gini ratio from a Random Sample of 10 Provinces in Indonesia

Province	Gini ratio (GR)
East Java	0.387
Gorontalo	0.417
North Sulawesi	0.370
Papua	0.386
South Sulawesi	0.377
Southeast Sulawesi	0.371
Special Capital Region of Jakarta	0.431
Special Region of Yogyakarta	0.449
West Java	0.425
West Nusa Tenggara	0.375

Data source: Statistics Indonesia website (period March 2023)

Here are the hypothesis testing steps according to **Algorithm 1**,

Step 1: Formulate the statistical hypothesis.
Research hypothesis: The average Gini ratio in March 2023 is at a low inequality level ($GR < 0.4$).
The statistical hypothesis using the left-tailed test:
$H_0: \mu = 0.4$ vs $H_a: \mu < 0.4$.

Step 2: State the assumptions and check the conditions.

Assumptions: (1) The Gini ratio data is continuous with a ratio scale; (2) a random sample of 10 out of 34 provinces from urban-rural regions was selected; (3) each province is mutually independent; (4) the population standard deviation σ of Gini ratio is unknown; and (5) The normality test results using the Anderson-Darling test obtained a p-value = 0.086, which is greater than the significance level $\alpha = 0.05$. **Figure 4** suggests that at the 95% confidence interval, the Gini ratio data follows a normal distribution. Figure 5 provides a graphical plot of the normality of the residuals along with the Anderson-Darling test statistic and the corresponding p-value presented in the box on the top right-hand side of the plot.

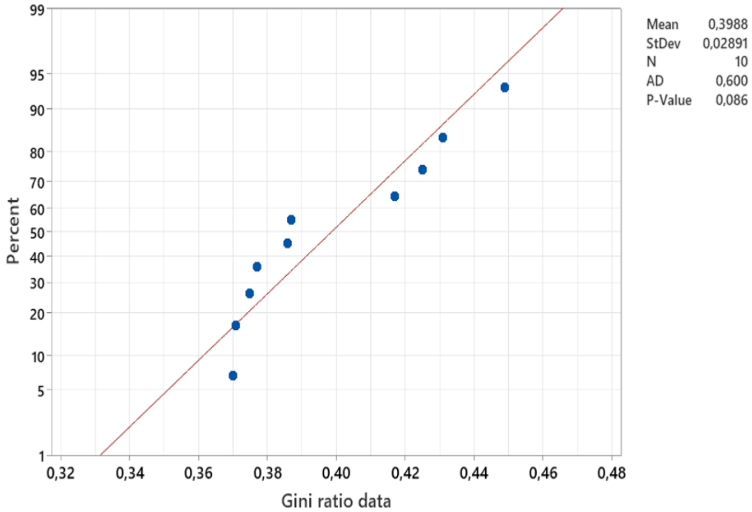


Figure 4. Probability plot of Gini Ratio Data and Normality Test Results using the Anderson-Darling test

Step 3: Determine the critical value at the selected α.
Since hypothesis testing involves a left-tailed test, at the significance level $\alpha = 0.05$, the critical value is determined by finding the t_{crit} that fills the 0.05 area to the left, and leaving the area to the right is 0.95 with degrees of freedom $\nu = n - 1 = 10 - 1 = 9$, such that $t_{crit} = -t_{(\alpha;\nu)} = -t_{(0.05;9)} = -1.883$.

Step 4: Calculate the test statistic.
According to the data in **Table 2**, it was found that the mean $\bar{x} = 0.399$ and standard deviation $s = 0.029$. Hence, the test statistic by using **Equation (2)** is

$$t_{stat} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.399 - 0.40}{0.0029/\sqrt{10}} = -0.131.$$

Step 5: Calculate the p-value.
Based on Equations (3), (4), (5), and (8), the p-value for the left-tailed test of the one-sample t-test is as follows.

$$p - value = P(T \leq t) = \frac{1}{2}\left(\frac{B[x(t); \mathrm{a}; \mathrm{b}]}{B[\mathrm{a}; \mathrm{b}]}\right) = \frac{1}{2}\left(\frac{B[0.998; 4.5; 0.5]}{B[4.5; 0.5]}\right) = \frac{1}{2}\left(\frac{0{,}898}{1}\right) = 0.449.$$

Step 6: Determine the test criteria or critical region.
The test statistic $t_{stat} = -0.131$ is greater than the critical value $t_{crit} = -1.883$, such that the t_{stat} does not fall into the H_0 rejection area. The area from t_{stat} to the left, representing the p-value with an area of 0.449, is larger than the area from t_{crit} to the left, which represents the significance level α that has an area of 0.05. It leads us not to reject H_0 **[25]**.

Step 7: Conclude.
At the 95% confidence interval, the population average Gini ratio in March 2023 is not statistically significantly different from 0.4. In other words, there is not enough evidence to justify Statistics Indonesia's claim that the average Gini ratio is at a low inequality level.

3.3 Retrospective Analysis

Since H_0 was not rejected, using **Equations (7)** and **Equation (8)**, we can establish the magnitude of the statistical power test performed for a sample size of $n = 10$ with a specific difference, let $\delta = -0.02$ and $\delta_{ES} = 0.69$. By running the G*Power software derived power values are provided in **Figure 5**.

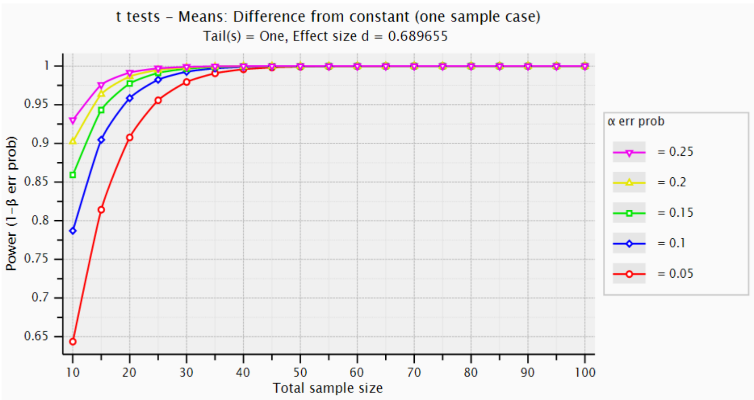


Figure 5. Power Plot as a Function of Sample Size with a Specific Effect Size

The results suggest that for a 95% confidence interval, we only have a 64.4% chance of detecting a difference $\delta = -0.02$ in the Gini ratio with a sample size of $n = 10$, once that difference exists. This probability is depicted as an area $1 - \beta$ in **Figure 6**. If a hypothesis test has weak power, researchers may fail to detect a difference and inaccurately conclude that there is no difference. It typically occurs when the sample size or the difference is smaller. Consequently, the statistical test has less power to detect such a difference.

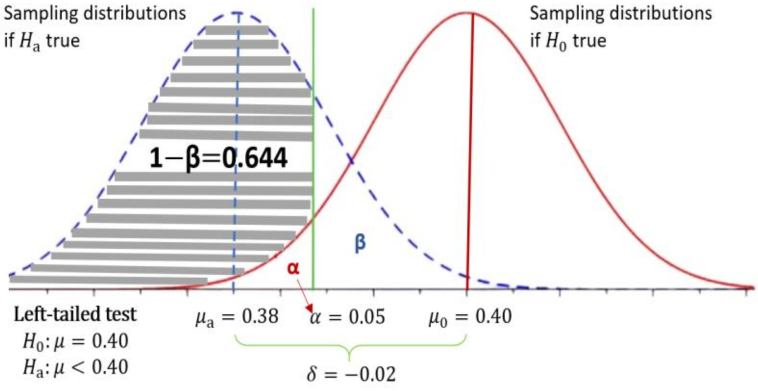


Figure 6. Power Test Area for a Left-Tailed Test of the Indonesia Gini Ratio Population Mean

Due to the statistical power test yield not providing a strong justification, it may be reasonable to consider increasing the sample size such that it is possible to detect differences with greater probability. Further action is taken by performing a retrospective analysis to investigate whether the sample is large enough to detect differences if they exist [1]. Suppose we want to detect a minimum difference of at least 0.02 in the average Gini ratio in March 2023 with a power of 0.90. Starting from Equation (7) and employing G*Power software, the result was iteratively derived, as presented in Figure 7.

A priori analysis results indicate that a minimum sample size of $n_\beta = 20$ is required to detect whether the population mean of the Gini ratio in March 2023 is less than the hypothesized mean ($\mu_0 = 0.40$) with the target power is 0.90 and a minimum difference of 0.02 such that the effect size is $\delta_{ES} = 0.69$. Due to the selected sample size of $n = 10$ being smaller than the necessary sample size of $n_\beta = 20$, the non-rejection of H_0 in the previous hypothesis testing may be caused by the sample size that is not adequate. In addition, G*Power always rounds non-integer sample sizes up to yield a consistent integer value with a power level that is not lower than the pre-specified one. Therefore, in most cases, the actual power is slightly larger than the pre-specified power in the a priori power analysis [14]. It could be verified from the actual power value in the test of 0.97 which was larger than the pre-specified power value of 0.90. In particular, if the researcher increases the sample size, the statistical power test will improve as well.

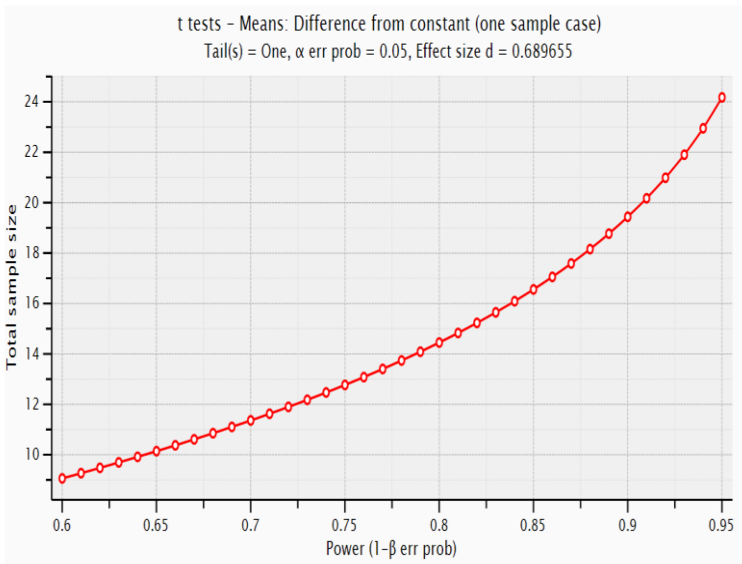


Figure 7. Sample Size Plot as a Function of Power with a Specific Effect Size

A priori analysis leads us to select at least a sample size $n_\beta = 20$ and repeat hypothesis testing. A resample of size ten was selected and added such that Table 3 exhibits a random sample of Gini ratios of 20 of the 34 provinces from urban-rural regions.

Table 3. Gini ratio from a Random Sample of 20 Provinces in Indonesia

Province	Gini ratio	Province	Gini ratio
East Java	0.387	Bali	0.362
Gorontalo	0.417	Banten	0.368
North Sulawesi	0.370	Bengkulu	0.333
Papua	0.386	Central Java	0.369
South Sulawesi	0.377	East Nusa Tenggara	0.325
Southeast Sulawesi	0.371	Jambi	0.343
Special Capital Region of Jakarta	0.431	Riau Islands	0.340
Special Region of Yogyakarta	0.449	South Sumatra	0.338
West Java	0.425	West Papua	0.370
West Nusa Tenggara	0.375	West Sulawesi	0.351

Data source: Statistics Indonesia website (period March 2023)

By applying **Algorithm 1**, similarly, the test statistic $t_{stat} = -3.378$ is less than the critical value $t_{crit} = -1.729$, then t_{stat} lies in the critical area of rejection H_0. Additionally, the area from t_{stat} to the left, representing the p-value with an area of 0.002, is smaller than the area from t_{crit} to the left, representing the significance level α with an area of 0.05. Thus, there is enough evidence to reject H_0. In conclusion, at the 95% confidence interval, Statistics Indonesia's claim that the average Gini ratio in March 2023 is at a low level of inequality is supported by statistically significant evidence.

In recognition of the fact that H_0 is rejected, implying a significant difference from a hypothesized population mean, the remaining step of the retrospective analysis is to execute a sensitivity analysis. This analysis aims to ensure that the observed differences are not solely due to random chance or sampling error but have real implications. Additionally, it allows us to assess whether the differences are meaningful regarding their practical implications or merely statistically significant [26]. Hence, the effect size defined by **Equation (11)** was calculated using G*Power software to gather the results in **Figure 8**.

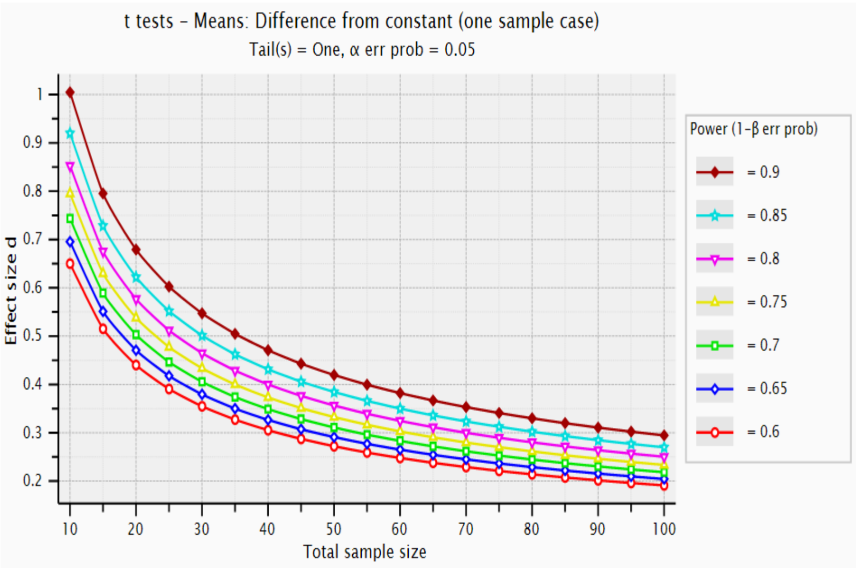


Figure 8. Effect Size Plot as a Function of Sample Size with a Specific Power

The value of $\delta_{ES} = 0.68$ indicates that the hypothesis testing has a medium effect size. It implies that the effect of the difference in the Gini ratio is not only statistically significant but adequately significant in practice. Moreover, an oversized sample size decreases the effect size due to its impracticality and potentially wasted time or unnecessary sampling costs.

4. CONCLUSIONS

This study highlights three essential roles of retrospective analysis in hypothesis testing, specifically as a priori analysis, post hoc analysis, and sensitivity analysis. The analysis was applied to Gini ratio data to verify Statistics Indonesia's claim that the expenditure of the population in March 2023 was at a low inequality level. The analysis was conducted with the aid of G*Power software. The result reveals that for a sample size of 20 at the 95% confidence interval, there is enough evidence to suggest that the Gini ratio is statistically significant at less than 0.4, with a power of analysis of 90.8% and an effect size of 0.68. As a result, retrospective analysis using G*Power software can be helpful for academics or practitioners in planning and evaluating hypothesis testing, especially concerning power, sample size, and effect size.

ACKNOWLEDGMENT

The authors would like to thank LPPM Universitas Singaperbangsa Karawang for the research funding through the HIPDA 2023 Scheme, also Mr. Dian Pebriana, Mr. Doni Ramdan, and Mrs. Julia Permata as the research assistant team for their technical contributions.

REFERENCES

[1] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability & Statistics for Engineers & Scientists*. Boston: Pearson, 2016.
[2] K. Andrew and R. Eckersley, *Statistics for biomedical engineers and scientists: How to visualize and analyze data*. London: Academic Press, 2019.
[3] G. V Glass and K. D. Hopkins, *Statistical Methods in Education and Psycology*, 3rd ed. Boston: Pearson Education, 1996.
[4] S. A. Lesik, *Applied statistical inference with MINITAB®*. Boca Raton: CRC Press, 2019.
[5] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. New York: Lauwrence Erlbaum Associates, 2013. doi: 10.4324/9780203771587.
[6] N. R. Chayyani, *Ketimpangan Pendapatan dan Pemulihan Ekonomi Nasional*. Jakarta: The Indonesian Institute Center of Public Policy Research, 2021. [Online]. Available: https://www.theindonesianinstitute.com/wp-content/uploads/2021/11/Ketimpangan-Pendapatan-dan-PEN-Nuri.pdf
[7] K. Mdingi and S. Ho, "Literature review on income inequality and economic growth," *MethodsX*, vol. 8, no. May, p. 101402, 2021, doi: 10.1016/j.mex.2021.101402.
[8] M. C. Wendl, "Pseudonymous fame," *Science (80-.).*, vol. 351, no. 6280, p. 1406, 2016.
[9] A. Li and H. Qin, "Some transformation properties of the incomplete beta function and its partial derivatives," *IAENG Int. J. Appl. Math.*, vol. 49, no. 1, pp. 1–14, 2019.
[10] D. A. Fitts, "Expected and empirical coverages of different methods for generating noncentral t confidence intervals for a standardized mean difference," *Behav. Res. Methods*, vol. 53, no. 6, pp. 2412–2429, 2021, doi: 10.3758/s13428-021-01550-4.
[11] D. A. Harrison and A. R. Brady, "Sample size and power calculations using the noncentral t-distribution," *Stata J. Promot. Commun. Stat. Stata*, vol. 4, no. 2, pp. 142–153, 2004, doi: 10.1177/1536867x0400400205.
[12] P. M. B. Cahusac and S. E. Mansour, "Estimating sample sizes for evidential t tests," *Res. Math.*, vol. 9, no. 1, pp. 1–12, 2022, doi: 10.1080/27684830.2022.2089373.
[13] W. H. Lo and S. H. Chen, "The analytical estimator for sparse data," *IAENG Int. J. Appl. Math.*, vol. 39, no. 1, pp. 1–8, 2009.
[14] F. Faul, E. Erdfelder, A. Buchner, and A. G. Lang, "G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.," *Behav. Res. Methods*, vol. 39, no. 2, pp. 175–191, 2007.
[15] H. Kang, "Sample size determination and power analysis using the G*Power software," *J. Educ. Eval. Health Prof.*, vol. 18, no. 17, pp. 1–12, 2021, doi: 10.3352/JEEHP.2021.18.17.
[16] T. Sitthiyot and K. Holasut, "A simple method for measuring inequality," *Palgrave Commun.*, vol. 6, no. 112, pp. 1–9, 2020, doi: 10.1057/s41599-020-0484-6.
[17] F. A. Farris, "The gini index and measures of inequality," *Am. Math. Mon.*, vol. 117, no. 10, pp. 851–864, 2010, doi: 10.4169/000298910X523344.
[18] B. D. Jakarta, *Profil Kemiskinan Provinsi DKI Jakarta Tahun 2020*. BPS Provinsi DKI Jakarta, 2020.
[19] Y. Liu and J. L. Gastwirth, "On the capacity of the Gini index to represent income distributions," *Metron*, vol. 78, no. 1, pp. 61–69, 2020, doi: 10.1007/s40300-020-00164-8.
[20] I. Drudi and G. Tassinari, "The Turn of the Screw . Changes in income distribution in Italy (2002-2010)," *Stat. Appl.*, vol. 12, no. 2, pp. 123–137, 2014.
[21] M. O. Lorenz, "Methods of measuring the concentration of wealth," *Am. Stat. Assoc.*, vol. 9, no. 70, pp. 209–219, 1905.
[22] P. K. Sen, "The gini coefficient and poverty indexes: Some reconciliations," *J. Am. Stat. Assoc.*, vol. 81, no. 396, pp. 1050–1057, 1986, doi: 10.1080/01621459.1986.10478372.
[23] A. Halimatussadiah, A. A. Widyasanti, A. Damayanti, K. Verico, R. M. Qibthiyyah, R. Kurniawan, J. F. Rezki, F. Rahardi, N. K. Sholihah, and S. Budiantoro, *Thinking Ahead: Indonesia' s Agenda on Sustainable Recovery from COVID -19 Pandemic*.

Jakarta: Institute for Economic and Social Research Faculty of Economics and Business, Universitas Indonesia (LPEM FEB UI) and Ministry of National Development Planning/ National Development Planning Agency (BAPPENAS), 2020, p. 125.

[24] BPS Statistics Indonesia, *Tingkat Ketimpangan Pengeluaran Penduduk Indonesia Maret 2023*. BPS Statistics Indonesia, 2023.

[25] H. H. S. Kyaw, S. S. Wint, N. Funabiki, and W. C. Kao, “A code completion problem in java programming learning assistant system,” *IAENG Int. J. Comput. Sci.*, vol. 47, no. 3, pp. 350–359, 2020.

[26] J. C. F. de Winter, “Using the student’s t-test with extremely small sample sizes,” *Pract. Assessment, Res. Eval.*, vol. 18, no. 10, pp. 1–12, 2013.