Chapter 07

Hypothesis Testing: One-Sample Inference

Introduction

Hypothesis-testing framework specifies two hypotheses:
Null and alternative hypothesis

Hypothesis-testing provides an objective framework for making decisions using probabilities methods, rather than relying on subjective impressions.

It provides a uniform decision-making criterion that is consistent.

In a **one-sample problem**, hypotheses are specified about a single distribution.

In a **two-sample problem**, two different distributions are compared.

The null hypothesis, denoted by H_o, is the hypothesis that is to be tested. The alternative hypothesis, denoted by H_1 is the hypothesis that in some sense contradicts the null hypothesis.

$$H_o: \mu = \mu_o \qquad \text{vs.} \qquad H_1: \mu < \mu_o$$

Suppose the only possible decisions are whether H_o is true or H_1 is true. For ease of notation, all outcomes in a hypothesis testing situation generally refer to the null hypothesis.

If we decide H_o is true, then we say we accept H_o. If we decide H_1 is true, then we state that H_o is not true or, equivalently, that we reject H_o. Thus, four possible outcomes can occur:

Table 7.1 Four possible outcomes in hypothesis testing

Decision	Truth: H_0	Truth: H_1
Accept H_0	H_0 is true and H_0 is accepted	H_1 is true and H_0 is accepted
Reject H_0	H_0 is true and H_0 is rejected	H_1 is true and H_0 is rejected

- The **probability of a type I error** is the probability of rejecting the null hypothesis when H_0 is true.

It is denoted by α and is commonly referred to as the significance level of a test.

- The **probability of a type II error** is the probability of accepting the null hypothesis when H_1 is true.

This probability is a function of μ as well as other factors. It is usually denoted by β.

The power of a test is defined as

$1 - \beta = 1 -$ probability of a type II error $= Pr(\text{rejecting } H_0 | H_1 \text{ true})$

The aim in hypothesis testing is to use statistical tests that make α and β as small as possible, which means rejecting and accepting null hypothesis less often, resp.

One-Sample Test for the Mean of a Normal Distribution: One-Sided Alternatives

- **Acceptance region:** range of values of x for which H_o is accepted.
- **Rejection region:** range of values of x for which H_o is rejected.

If $\bar{x}$ is sufficiently smaller than μo, then H_o is rejected;
otherwise, H_o is accepted.
If H_o is true: the most likely values of $\bar{x}$ cluster around μ_o.
If H_1 is true: the most likely values of $\bar{x}$ cluster around μ_1.
Such a test based on sample mean has the highest power among all tests with a given type I error of α.

A **one-tailed test** is a test in which the values of the parameter being studied (in this case μ) under the alternative hypothesis are allowed to be either greater than or less than the values of the parameter under the null hypothesis (μ_o) but not both.

One-sample t Test for the Mean of a Normal Distribution with Unknown Variance (Alternative Mean < Null Mean)

To test the hypothesis H_o: $\mu = \mu_o$, σ unknown vs. H_1: $\mu < \mu_o$, σ unknown with a significance level of α, we compute

$$t = (\overline{x} - \mu_o)/(s/\sqrt{n})$$

If $t < t_{n-1,\alpha}$, then we reject H_o.

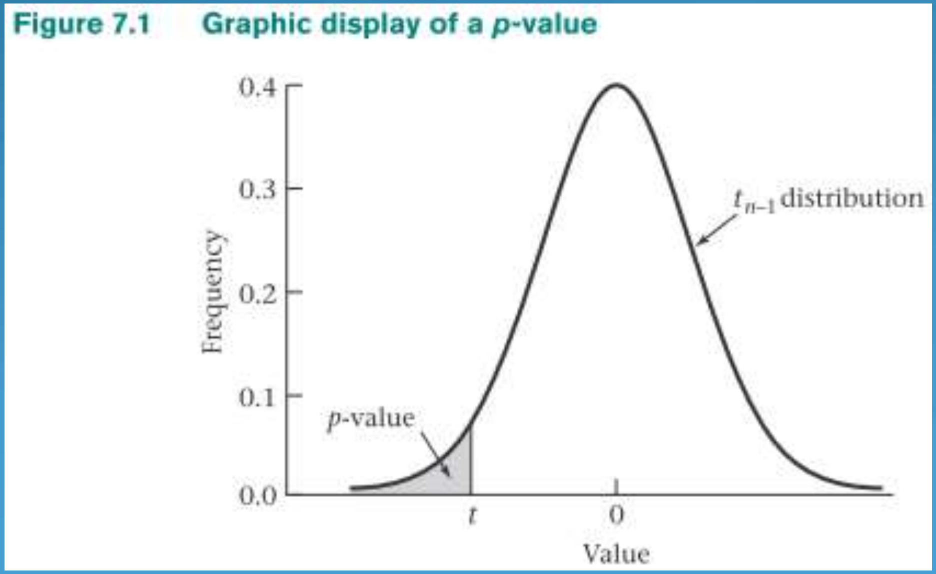
If $t \geq t_{n-1,\alpha}$, then we accept H_o.

The value of t is called a test statistic because the test procedure is based on this statistic.

The value $t_{n-1,\alpha}$ is called a critical value because the outcome of the test depends on whether the test statistic $t < t_{n-1,\alpha}$ = critical value, whereby we reject H_o or $t \geq t_{n-1,\alpha}$, whereby we accept H_o.

The general approach in which we compute a test statistic and determine the outcome of a test by comparing the test statistic with a critical value determined by the type I error is called the **critical-value method** of hypothesis testing.

The ***p-value*** for any hypothesis test is the α level at which we would be indifferent between accepting or rejecting H_0 given the sample data at hand. That is, the *p*-value is at the α level at which the given value of the test statistic (such as *t*) is on the borderline between the acceptance and rejection regions.



Solving for p as a function of t we get $p = Pr(t_{n-1} \leq t)$

The ***p*-value** can also be thought of as the probability of obtaining a test statistic as extreme as or more extreme than the actual test statistic obtained, given that the null hypothesis is true.

The ***p*-value** indicates exactly how significant the results are without performing repeated significance tests at different α levels.

<u>Guidance for Judging the Significance of a p-Value</u>

- If $0.01 \le p < 0.05$, then the results are significant.
- If $0.001 \le p < 0.01$, then the results are highly significant.
- If $p < 0.001$, then the results are very highly significant.
- If $p > 0.05$, then the results are considered not statistically significant (denoted as NS).
- If $0.05 \le p < 0.1$, then a trend toward statistical significance is sometimes noted.

Determination of Statistical Significance for Results from Hypothesis Tests

Either of the following methods can be used to establish whether results from hypothesis tests are statistically significant:

(1) The test statistic t can be computed and compared with the critical value $t_{n-1,\alpha}$ at an α level of 0.05. If H_0: $\mu = \mu_0$ vs. H_1: $\mu < \mu_0$ is being tested and $t < t_{n-1,0.05}$, then H_0 is rejected and the results are declared *statistically significant* ($p < 0.05$). Otherwise, H_0 is accepted and the results are declared *not statistically significant* ($p \geq 0.05$) This approach is called the **critical-value method**.

(2) The exact *p*-value can be computed and, if $p < 0.05$, then H_0 is rejected and the results are declared *statistically significant*. Otherwise, if $p \geq 0.05$, then H_0 is accepted and the results are declared *not statistically significant*. This approach is called the ***p*-value method**.

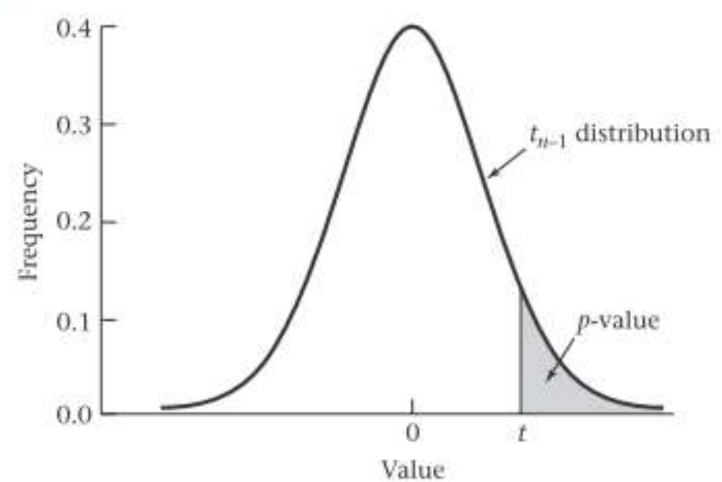
One-Sample t Test for the Mean of a Normal Distribution with Unknown Variance (Alternative Mean > Null Mean)

To test the hypothesis $H_o: \mu = \mu_o$ vs. $H_1: \mu > \mu_o$
with a significance level of α, the best test is based on t, where
$t = (\bar{x} - \mu_o)/(s/\sqrt{n})$

- If $t > t_{n-1,1-\alpha}$ then H_o is rejected
- If $t \leq t_{n-1,1-\alpha}$ then H_o is accepted

The p-value for this test is given by $p = Pr(t_{n-1} > t)$

Figure 7.2 p-value for the one-sample t test when the alternative mean (μ_1) > null mean (μ_0)



One-Sample Test for the Mean of a Normal Distribution: Two-Sided Alternatives

A **two-tailed test** is a test in which the values of the parameter being studied (in this case μ) under the alternative hypothesis are allowed to be either greater than or less than the values of the parameter under the null hypothesis (μ_0).

A reasonable decision rule to test for alternatives on either side of the null mean is to reject H_0 if it is either too small or too large. Another way of stating the rule is that H_0 will be rejected if t is either $< c_1$ or $> c_2$ for some constants c_1, c_2, and H_0 will be accepted if $c_1 \le t \le c_2$.

$$Pr(\text{reject } H_0 | H_0 \text{ true}) = Pr(t < c_1 \text{ or } t > c_2 | H_0 \text{ true})$$
$$= Pr(t < c_1 | H_0 \text{ true}) + Pr(t > c_2 | H_0 \text{ true}) = \alpha$$

$$Pr(t < c_1 | H_0 \text{ true}) = Pr(t > c_2 | H_0 \text{ true}) = \alpha/2$$

One-sample t Test for the Mean of a Normal Distribution with Unknown Variance (Two-Sided Alternative)

To test the hypothesis $H_o: \mu = \mu_o$ vs. $H_1: \mu \neq \mu_o$ with a significance level of α, the best test is based on $t = (\bar{x} - \mu_o)/ (s/\sqrt{n})$.

If $|t| > t_{n-1,1-\alpha/2}$ then H_o is rejected.
If $|t| > t_{n-1,1-\alpha/2}$ then H_o is accepted.

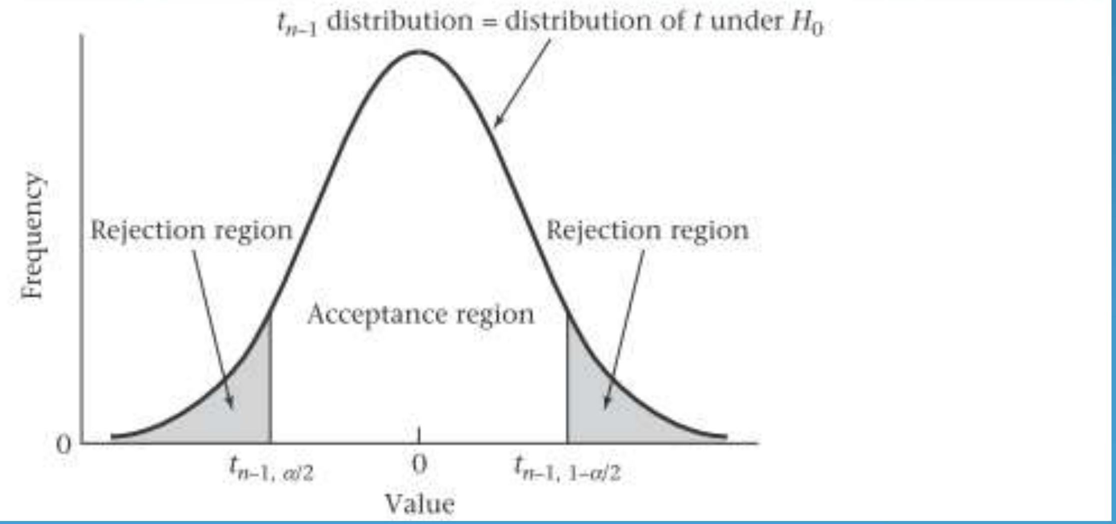


Figure 7.3 One-sample t test for the mean of a normal distribution (two-sided alternative)

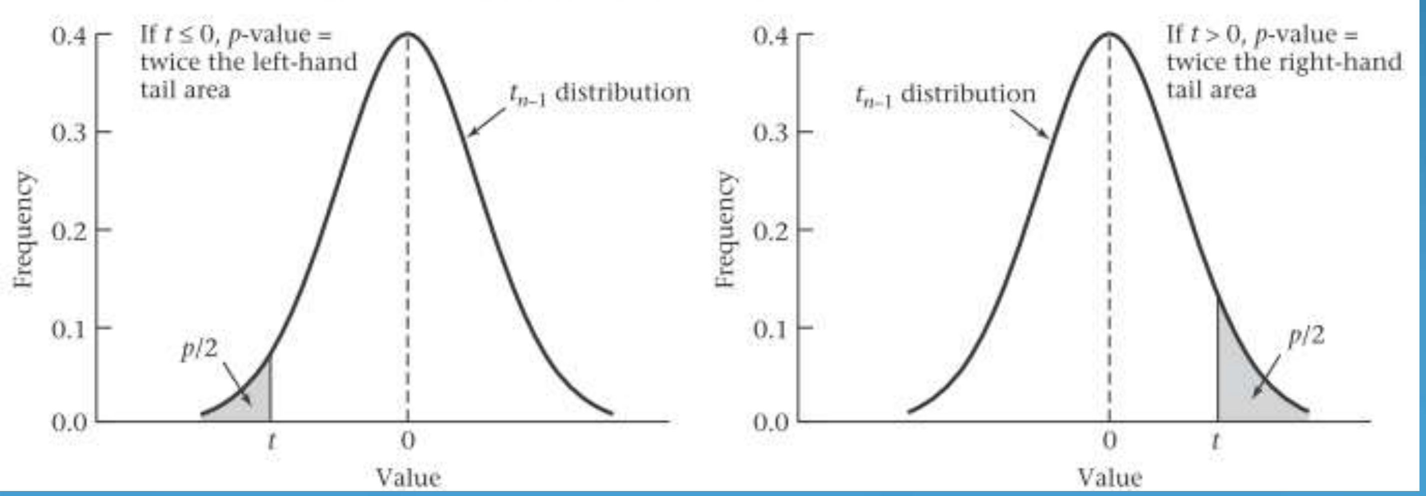
p-Value for the One-Sample t Test for the Mean of a Normal Distribution (Two-Sided Alternative)

Let $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$

$$p = \begin{cases} 2 \times Pr(t_{n-1} \le t), \text{ if } t \le 0 \\ 2 \times [1 - Pr(t_{n-1} \le t)], \text{ if } t > 0 \end{cases}$$

The ***p*-value** is the probability under the null hypothesis of obtaining a test statistic as extreme as or more extreme than the observed test statistic, where, because a two-sided alternative hypothesis is being used, extremeness is measured by the **absolute value** of the test statistic.

Figure 7.4 Illustration of the *p*-value for a one-sample *t* test for the mean of a normal distribution (two-sided alternative)



When is a one-sided test more appropriate than a two-sided test?

Generally, the sample mean falls in the expected direction from μ_o and it is easier to reject H_o using a one-sided test than using a two-sided test.

A two-sided test can be more conservative because it is not necessary to guess the appropriate side of the null hypothesis for the alternative hypothesis.

In some cases, only alternatives on one side of the null mean are of interest or are possible, and a one-sided test is better than a two-sided test because it has more power (since it is easier to reject H_o based on a finite sample if H_1 is actually true).

The decision whether to use a one-sided or two-sided test must be made before the data analysis (or before data collection) begins so as not to bias conclusions based on results of hypothesis testing.

Do not change from a two-sided to a one-sided test after looking at the data.

One-Sample z Test for the Mean of a Normal Distribution with Known Variance (Two-Sided Alternative)

To test the hypothesis H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$ with a significance level of α, where the underlying standard deviation σ is known, the best test is based on $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$

If $z < z_{\alpha/2}$ or $z > z_{1-\alpha/2}$ then Ho is rejected.

If $z_{\alpha/2} \leq z \leq z_{1-\alpha/2}$ then Ho is accepted.

To compute a two-sided p-value, we have $p = 2\Phi(z)$ if $z \leq 0$

$= 2[1 - \Phi(z)]$ if $z > 0$

One-Sample z Test for the Mean of a Normal Distribution with Known Variance (One-Sided Alternative)

($\mu_1 < \mu_0$) To test the hypothesis H_0: $\mu = \mu_0$ vs. H_1: $\mu < \mu_0$ with a significance level of α, where the underlying standard deviation σ is known, the best test is based on $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$

If $z < z_\alpha$ then H_0 is rejected; if $z \geq z_\alpha$, then H_0 is accepted. The *p*-value is given by $p = \Phi(z)$.

($\mu_1 > \mu_0$) To test the hypothesis H_0: $\mu = \mu_0$ vs. H_1: $\mu > \mu_0$ with a significance level of α, where the underlying standard deviation σ is known, the best test is based on $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$

If $z > z_{1-\alpha}$ then H_0 is rejected; if $z \leq z_{1-\alpha}$ then H_0 is accepted. The *p*-value is given by $p = 1 - \Phi(z)$.

Sample-Size Estimation Based on CI Width

Suppose we wish to estimate the mean of a normal distribution with sample variance s^2 and require that the two-sided 100% × (1 - α) CI for μ be no wider than L. The number of subjects needed is approximately

$$n = 4z_{1-\alpha/2}^2 s^2 / L^2$$

The Relationship Between Hypothesis Testing and Confidence Intervals (Two-Sided Case)

Suppose we are testing H_o: $\mu = \mu_o$ vs. H_1: $\mu \neq \mu_o$ is rejected with a two-sided level α test if and only I the two-sided 100% × (1 - α) CI for μ does not contain μ_o. H_o is accepted with a two-sided level α test if and only if the two-sided 100% × (1 - α) CI for μ does contain μ_o.

The two-sided 100% × (1 - α) CI for μ contains all values μ_o such that we accept H_o using a two-sided test with significance level α, where the hypotheses are H_o: $\mu = \mu_o$ vs. H_1: $\mu \neq \mu_o$. Conversely, the 100% × (1 - α) CI does not contain any value μo for which we can reject H_o using a two-sided test with significance level with significance level α, where H_o: $\mu = \mu_o$ vs. H_1: $\mu \neq \mu_o$.

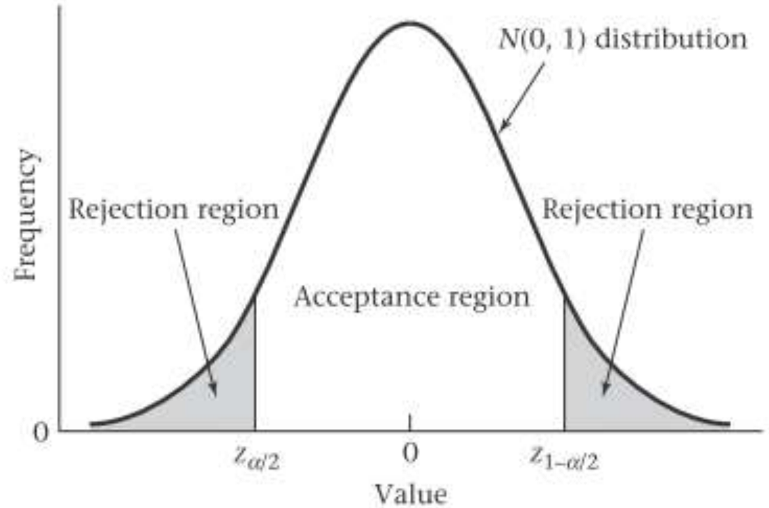
One-sample Inference for the Binomial Distribution: Normal-Theory Methods

Let the test statistic $z = (\hat{p} - p_0)/\sqrt{p_0 q_0 / n}$

If $z < z_{\alpha/2}$ or $z > z_{1-\alpha/2}$, then H_o is rejected.

If $z_{\alpha/2} \leq z \leq z_{1-\alpha/2}$, then H_o is accepted. This test should only be used if $np_o q_o \geq 5$.

Figure 7.12 Acceptance and rejection regions for the one-sample binomial test—normal-theory method (two-sided alternative)

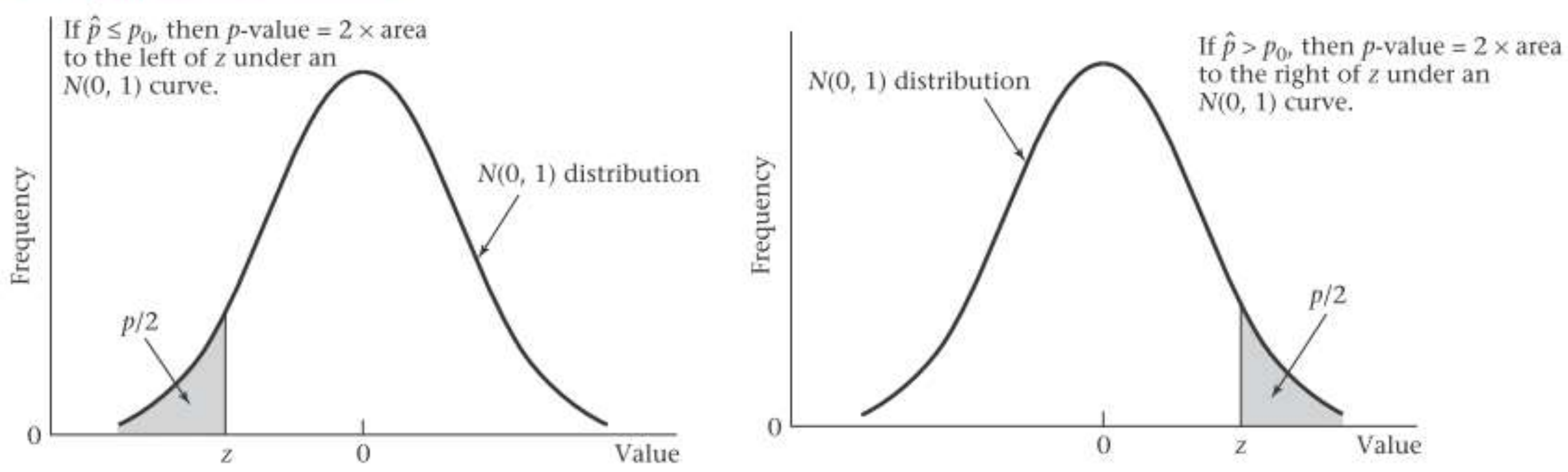


Computation of the p-value for the one-Sample Binomial Test: Normal Theory Method (Two-Sided Alternative)

Let the test statistic $z = (\hat{p} - p_0)/\sqrt{p_0 q_0 / n}$

If $\hat{p} \le p_0$, then *p*-value = 2 × Φ(z) = twice the area to the left of z under an N(0,1) curve. If $\hat{p} > p_0$, then *p*-value = 2 × [1- Φ(z)] = twice the area to the right of z under an N(0,1) curve.

Illustration of the *p*-value for a one-sample binomial test— normal-theory method (two-sided alternative)



Summary

In this chapter, we introduced

1. Specification of the null (H_0) and alternative (H_1) hypotheses;
2. type I error (α), type II error (β), and the power ($1-\beta$) of a hypothesis test; the p-value of a hypothesis test and the distinction between on-sided and two-sided tests;
3. methods for estimating appropriate sample size as determined by the prespecified null and alternative hypotheses and type I and type II errors.

   These concepts were applied to many one-sample hypothesis- testing cases. Each of the hypothesis tests was shown to be conducted in one of two ways

1. Specifying critical values to determine the acceptance and rejection regions (critical-value method) based on a specified type I error α.
2. Computing p-values (p-value method)

The End