Learning Module 8: Hypothesis Testing

LOS 8a: Explain hypothesis testing and its components, including statistical significance, Type I and Type II errors, and the power of a test

A hypothesis is an assumed statement about a population's characteristics, often considered an opinion or claim about an issue. To determine if a hypothesis is accurate, statistical tests are used. Hypothesis testing uses sample data to evaluate if a sample statistic reflects a population with the hypothesized value of the population parameter.

Below is an example of a hypothesis:

“The mean return of small-cap stock is higher than that of large-cap stock.”

Hypothesis testing involves collecting and examining a representative sample to verify the accuracy of a hypothesis. Hypothesis tests help analysts to answer questions such as:

- Is bond type A more profitable than type B?
- Does staff training lead to improved efficiency at the workplace?
- Are motor vehicle insurance claims consistent with a lognormal distribution?

Procedure Followed During Hypothesis Testing

Whenever a statistical test is being performed, the following procedure is generally considered ideal:

1. Statement of both the null and the alternative hypotheses.
2. Selection of the appropriate test statistic, i.e., what's being tested, e.g., the population mean, the difference between sample means, or variance.
3. Specification of the level of significance.
4. A clear statement of the decision rule to guide the choice of whether to reject or approve the null hypothesis.
5. Calculation of the sample statistic.

6. Arrival at a decision based on the sample results.

Step 1: Stating the Hypotheses

The Null vs. Alternative Hypothesis

The **null hypothesis**, denoted as H_0, signifies the existing knowledge regarding the population parameter under examination, essentially representing the "status quo." For example, when the U.S. Food and Drug Administration inspects a cooking oil manufacturing plant to confirm that the cholesterol content in 1 kg oil packages doesn't exceed 0.15%, they might create a hypothesis like:

H_0: Each 1 kg package has 0.15% cholesterol.

A test would then be carried out to confirm or reject the null hypothesis.

Typical statements of H_0 include:

$$
\begin{aligned}
H_0 &: \mu = \mu_0 \\
H_0 &: \mu \leq \mu_0 \\
H_0 &: \mu \geq \mu_0
\end{aligned}
$$

Where:

μ = True population mean.

μ_0 = Hypothesized population mean.

The **alternative hypothesis**, denoted as H_a, is a contradiction of the null hypothesis. Therefore, rejecting the H_0 makes H_a valid. We accept the alternative hypothesis when the “status quo” is discredited and found to be false.

Using our FDA example above, the alternative hypothesis would be:

H_a: Each 1 kg package does not have 0.15% cholesterol.

One-tailed vs. Two-tailed Hypothesis Testing

One-tailed Test

A one-tailed test (one-sided test) is a statistical test that considers a change in only one direction. In such a test, the alternative hypothesis either has a < (less than sign) or > (greater than sign), i.e., we consider either an increase or reduction, but not both.

A one-tailed test directs all the significance levels (α) to test statistical significance in one direction. In other words, we aim to test the possibility of a change in one direction and completely disregard the possibility of a change in the other direction.

If we have a 5% significance level, we shall allot 0.05 of the total area in one tail of the distribution of our test statistic.

Examples: Hypothesis Testing

Let us assume that we are using the standardized normal distribution to test the hypothesis that the population mean is equal to a given value X. Further, let us assume that we are using data from a sample drawn from the population of interest. Our null hypothesis can be expressed as:

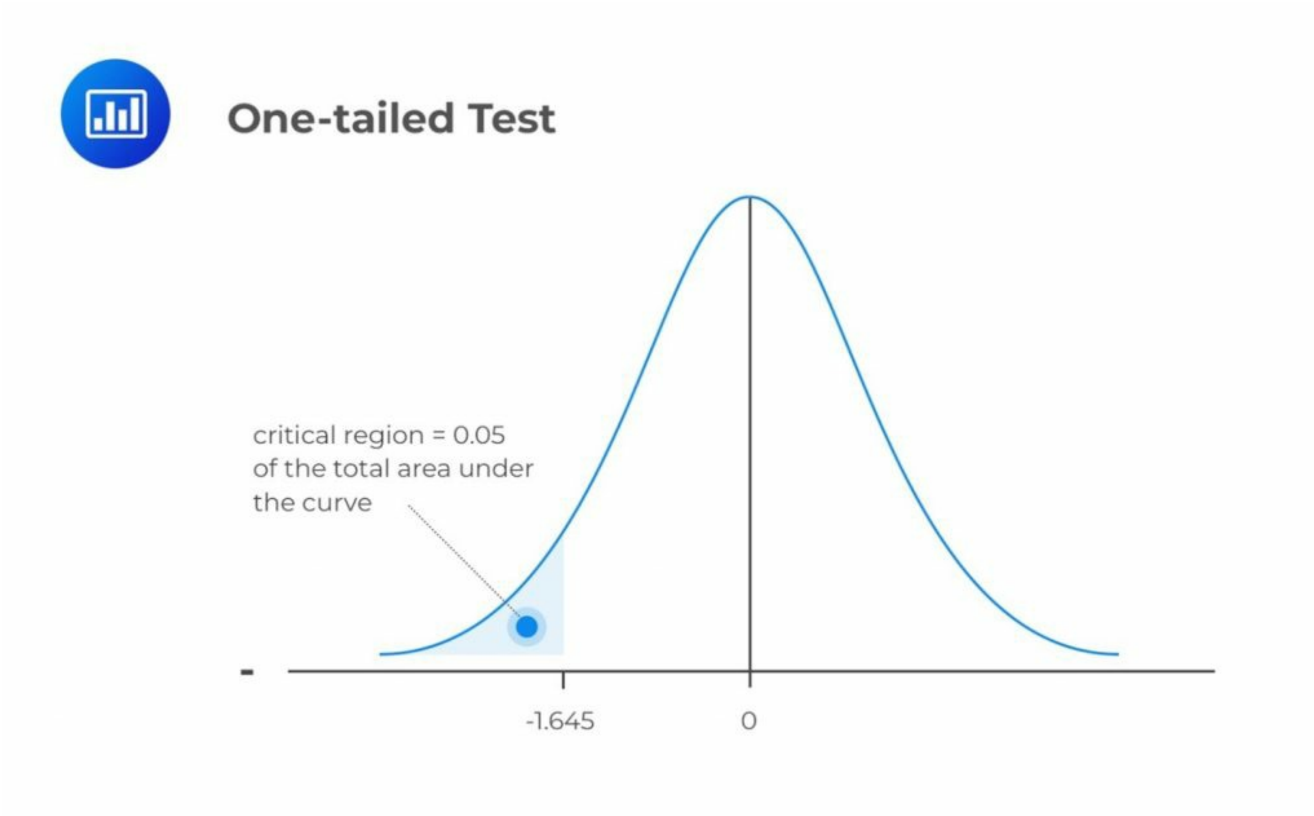
$$H_0 : \mu = X$$

If our test is one-tailed, the alternative hypothesis will test if the mean is either significantly greater than X or significantly less than X, but NOT both.

Case 1: At the 95% Confidence Level

$$H_a : \mu < X$$

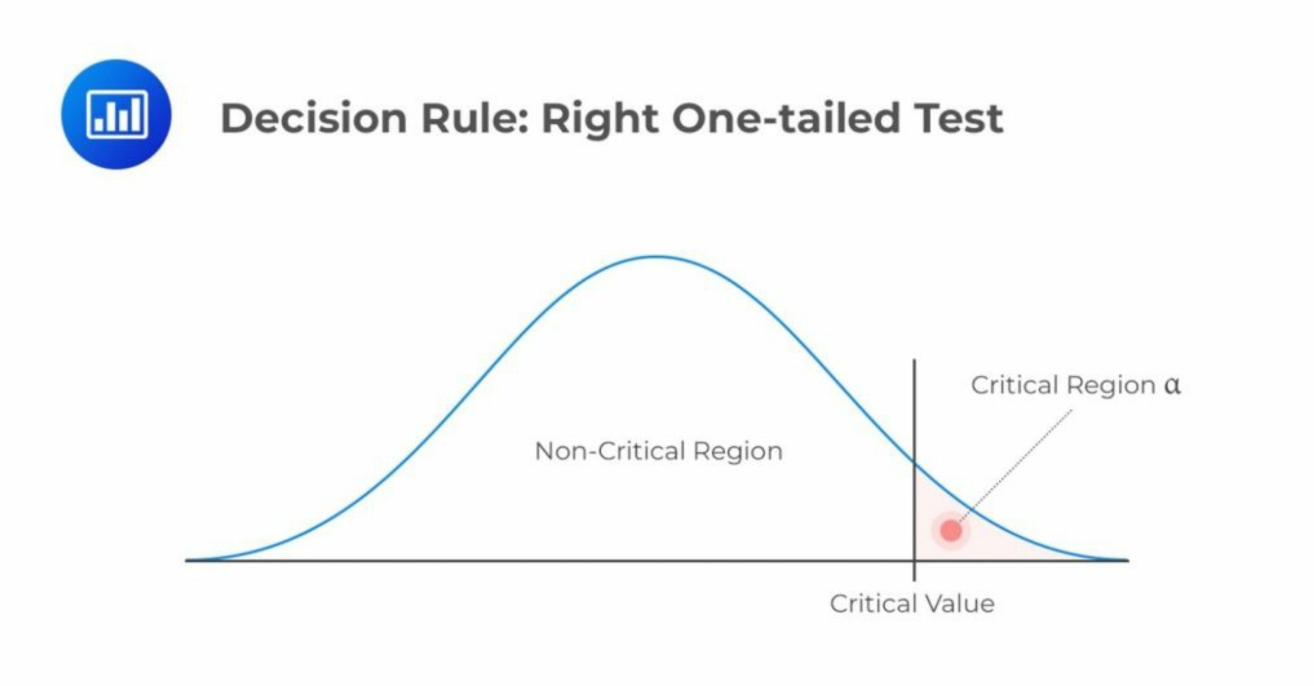
The mean is significantly less than X if the test statistic is in the bottom 5% of the probability distribution. This bottom area is known as the critical region (rejection region). We will reject the null hypothesis if the test statistic is less than -1.645.



Case 2: Still at the 95% Confidence Level

$$H_{a1} : \mu > X$$

We would reject the null hypothesis only if the test statistic is greater than the upper 5% point of the distribution. In other words, we would reject H_0 if the test statistic is greater than 1.645.

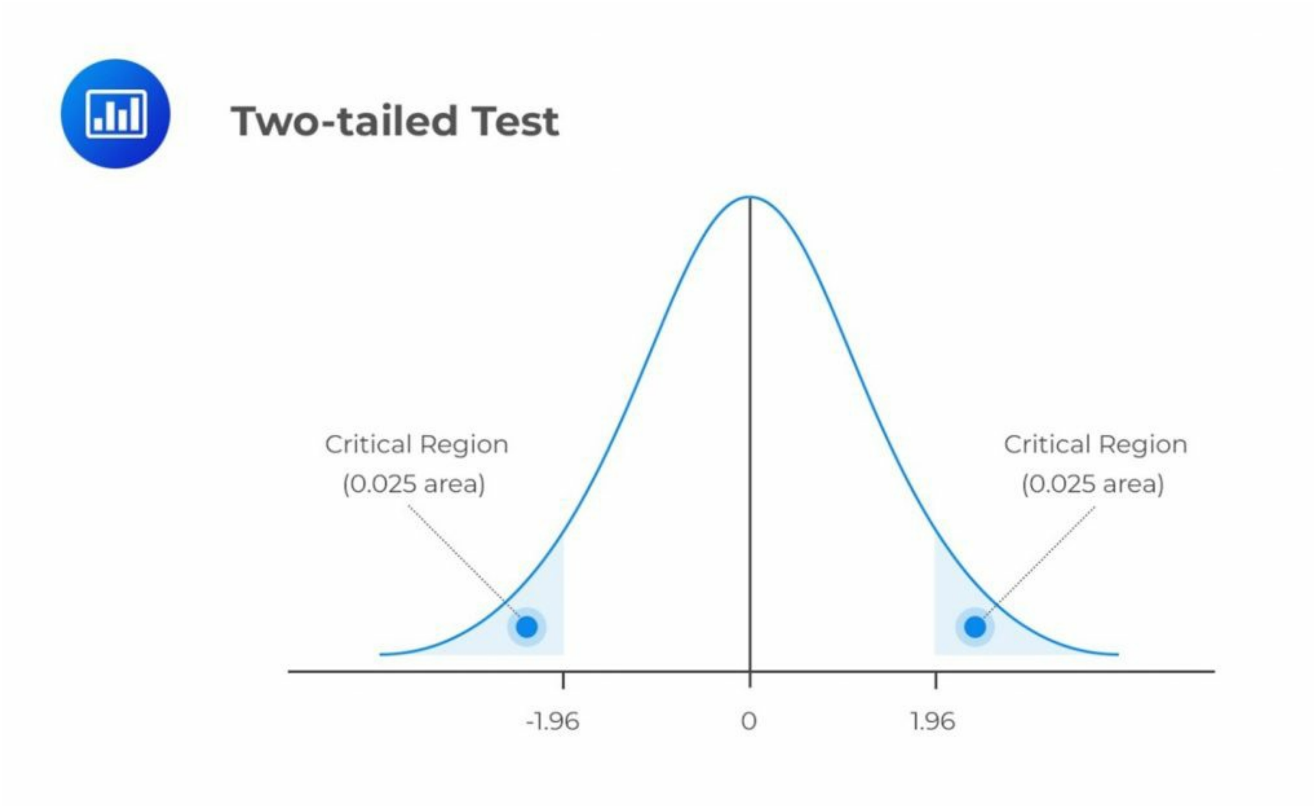


A Two-tailed Test

A two-tailed test considers the possibility of a change in either direction. It looks for a statistical relationship in both a distribution's positive and negative directions. Therefore, it allows half the value of α to test statistical significance in one direction and the other half to test the same in the opposite direction. A two-tailed test may have the following set of hypotheses:

$$H_0 : \mu = X$$
$$H_1 : \mu \neq X$$

Refer to our earlier example. If we were to carry out a two-tailed test, we would reject H_0 if the test statistic turned out to be less than the lower 2.5% point or greater than the upper 2.5% point of the normal distribution.



Step 2: Identify the Appropriate Test Statistic and Distribution

Test Statistic

A test statistic is a standardized value computed from sample information when testing hypotheses. It compares the given data with what an analyst would expect under a null hypothesis. As such, the null hypothesis is a major determinant of the decision to accept or reject H_0, the null hypothesis.

We use test statistic to gauge the degree of agreement between sample data and the null hypothesis. Analysts use the following formula when calculating the test statistic for most tests:

$$\text{Test statistic} = \frac{\text{Sample statistic} - \text{Hypothesized value}}{\text{Standard error}}$$

The test statistic is a random variable that varies with each sample. The table below provides an overview of commonly used test statistics, depending on the presumed data distribution:

Hypothesis Test	Test Statistic
Z-test	Z- statistic (Normal distribution)
Chi-Square Test	Chi-square statistic
t-test	t-statistics
ANOVA	F-statistic

We can subdivide the set of values that the test statistic can take into two regions: The non-rejection region, which is consistent with the H_0, and the rejection region (critical region), which is inconsistent with the H_0. If the test statistic has a value found within the critical region, we reject the H_0.

As is the case with any other statistic, the distribution of the test statistic must be completely specified under the H_0 when the H_0 is true.

The following is the list of test statistics and their distributions:

Test Subject	Test Statistic Formula	Test Statistic Distribution	Number of Degrees of Freedom
Single Mean	$t = \frac{\bar{X}-\mu_0}{\frac{s}{\sqrt{n}}}$	$t-$ distribution	$n-1$
Difference in Means	$t = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}}$	$t-$ distribution	n_1+n_2-1
Mean of Differences	$t = \frac{\bar{d}-\mu_{d0}}{s_{\bar{d}}}$	$t-$ distribution	$n-1$
Single Variance	$\chi^2 = \frac{s^2(n-1)}{\sigma_0^2}$	Chi-square Distribution	$n-1$
Difference in variances	$F = \frac{S_1^2}{S_2^2}$	F-distribution	n_1-1, n_2-1
Correlation	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	t-distribution	$n-2$
Independence (categorical data)	$\chi^2 = \sum_{i=1}^{m}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$	Chi-square Distribution	$(r-1)(c-1)$

Where:

μ_0, μ_{d_0}, and σ_0^2 denote hypothesized values of the mean, mean difference, and variance in that order.

$\bar{X}$, $\bar{b}$ s^2, s and r denote the sample mean of the differences, sample variance, sample standard deviation, and correlation, in that order.

O_{ij} and E_{ij} are observed and expected frequencies, respectively, with r indicating the number of rows and c indicating the number of columns in the contingency table.

Step 3: Specify the Level of Significance

The significance level represents the amount of sample proof needed to reject the null hypothesis. First, let us look at type I and type II errors.

Type I and Type II Errors

When using sample statistics to draw conclusions about an entire population, the sample might not accurately represent the population. This can result in statistical tests giving incorrect results, leading to either erroneous rejection or acceptance of the null hypothesis. This introduces the two errors discussed below.

Type I Error

Type I error occurs when we reject a true null hypothesis. For example, a type I error would manifest in the rejection of $H_0 = 0$ when it is zero.

Type II Error

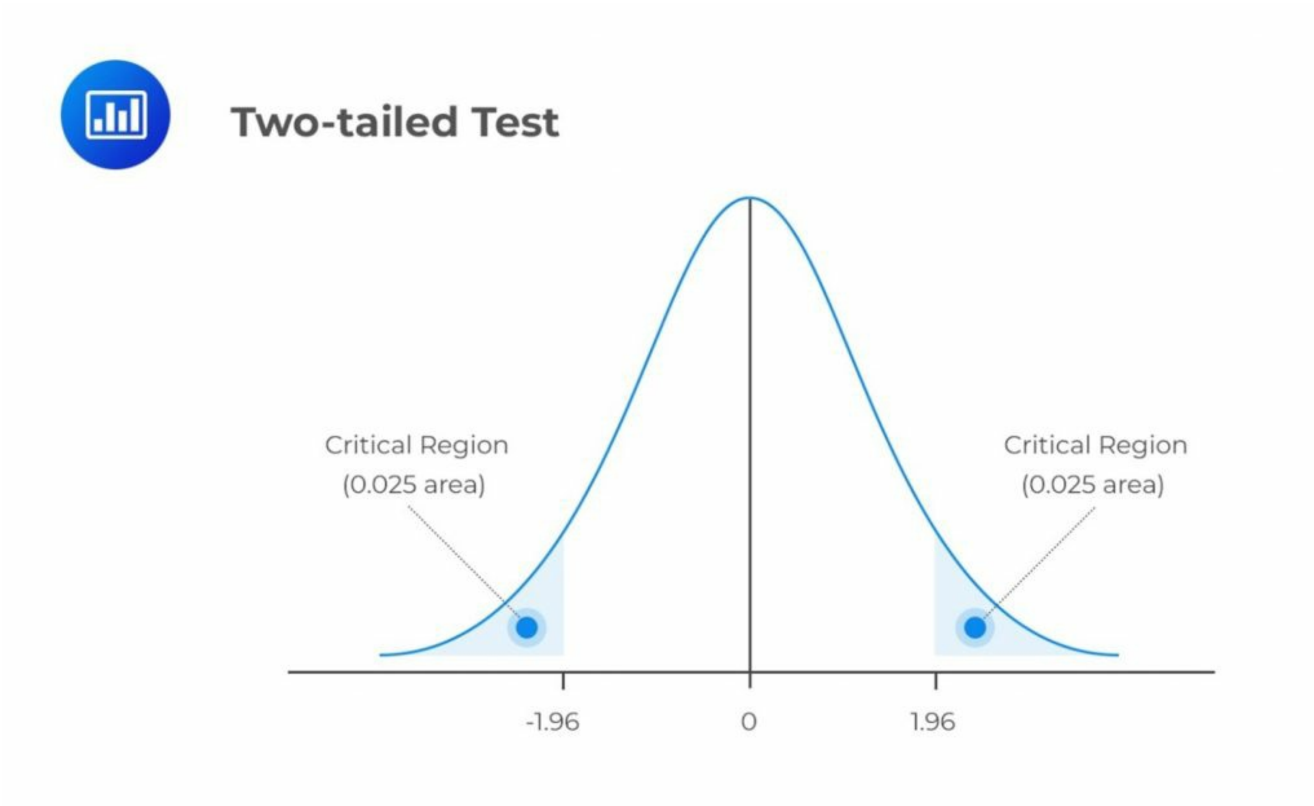
Type II error occurs when we fail to reject a false null hypothesis. In such a scenario, the evidence the test provides is insufficient and, as such, cannot justify the rejection of the null hypothesis when it is false.

Consider the following table:

Decision	True Null Hypothesis (H_0)	False Null Hypothesis (H_0)
Fail to reject the null hypothesis	Correct decision	Type II error
Reject null hypothesis	Type I error	Correct decision

The level of significance, denoted by α, represents the probability of making a type I error, i.e., rejecting the null hypothesis when it is true. The **confidence level** complements the significance level, $(1-\alpha)$.

We use α to determine critical values that subdivide a distribution into the rejection and the non-rejection regions. The figure below gives an example of the critical regions under a two-tailed normal distribution and 5% significance level:



Consequently, β, the direct opposite of α, is the probability of making a type II error within the bounds of statistical testing. The ideal but practically impossible statistical test would be one that **simultaneously** minimizes α and β.

The Power of a Test

The power of a test is the direct opposite of the significance level. The level of significance gives us the probability of rejecting the null hypothesis when it is, in fact, true. On the other hand, the power of a test gives us the probability of correctly discrediting and rejecting the null hypothesis when it is false. In other words, it gives the likelihood of rejecting H_0 when, indeed, it is false. Expressed mathematically,

$$\text{Power of a test} = 1 - \beta = 1 - P\,(\text{type II error})$$

In a scenario with multiple test results for the same purpose, the test with the highest power is considered the best.

Steps 4, 5, 6: State the Decision Rule, Calculate the Test Statistic, and Make a Decision

The decision rule is the procedure that analysts and researchers follow when deciding whether to reject or not reject a null hypothesis. We use the phrase "not to reject" because it's statistically incorrect to "accept" a null hypothesis. Instead, we can only gather enough evidence to support it.

Breaking Down the Decision Rule

The decision to reject or not reject a null hypothesis relies on the distribution of the test statistic. The decision rule compares the calculated test statistic to the critical value.

If we reject the null hypothesis, the test is considered statistically significant. If not, we fail to reject the null hypothesis, indicating insufficient evidence for rejection.

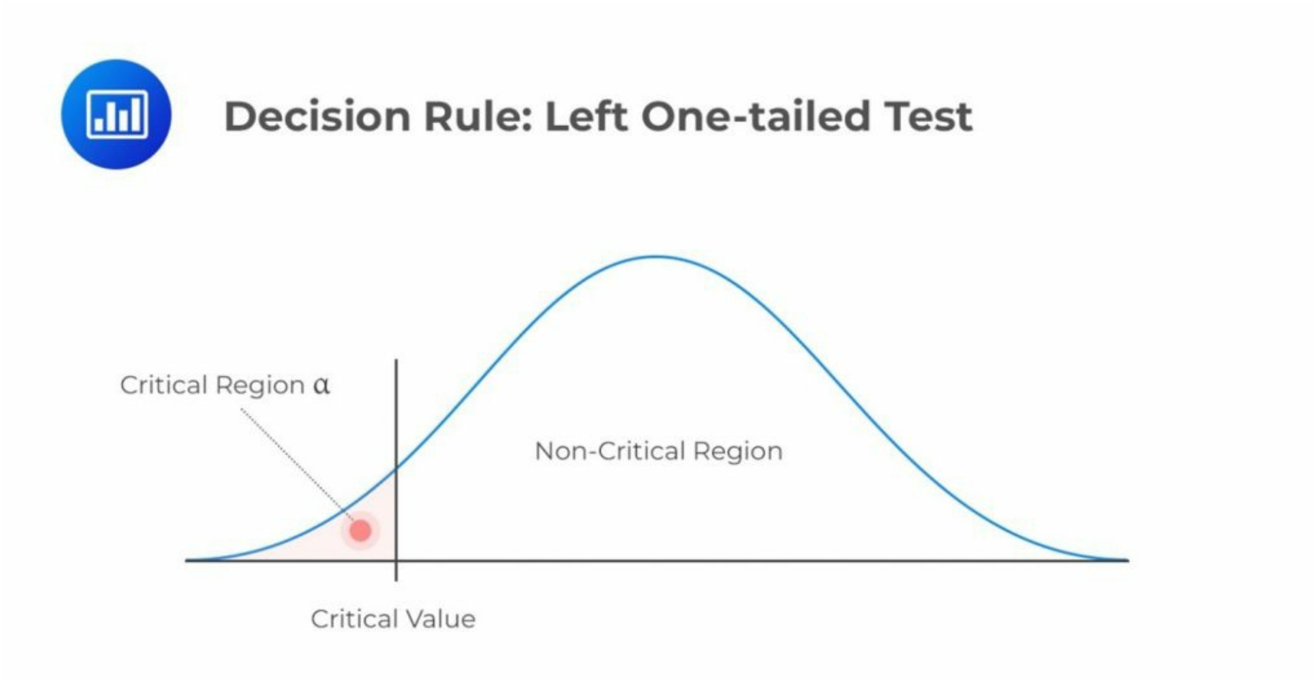
We use the test's significance level if a variable follows a normal distribution. This helps us find critical values corresponding to specific points on the standard normal distribution. These critical values guide the decision-making process for rejecting or not rejecting a null hypothesis.

Before deciding whether to reject or not reject a null hypothesis, it's crucial to determine whether the test should be one-tailed or two-tailed. This choice depends on the nature of the research question and the direction of the expected effect. Notably, the number of tails determines the value of α (significance level). The following is a summary of the decision rules under different scenarios.

Left One-tailed Test

H_a : Parameter $< X$

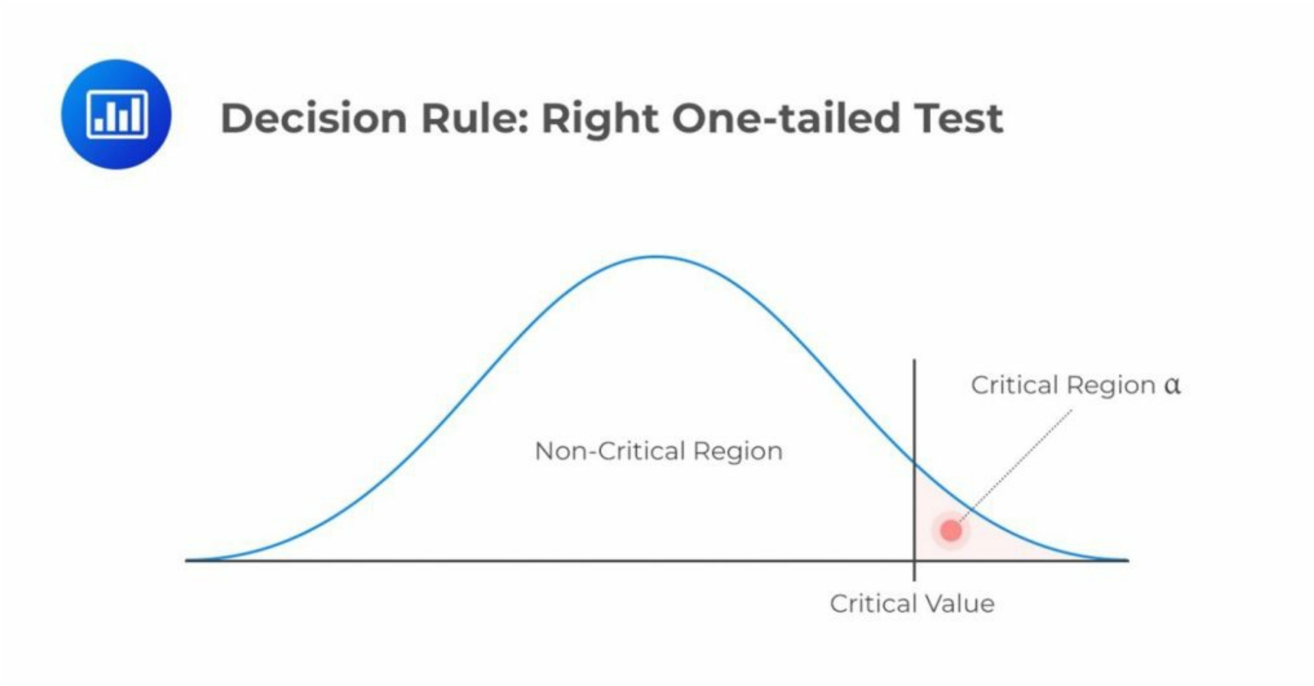
Decision rule: Reject H_0 if the test statistic is less than the critical value. Otherwise, **do not reject** H_0.



Right One-tailed Test

H_a : Parameter $> X$

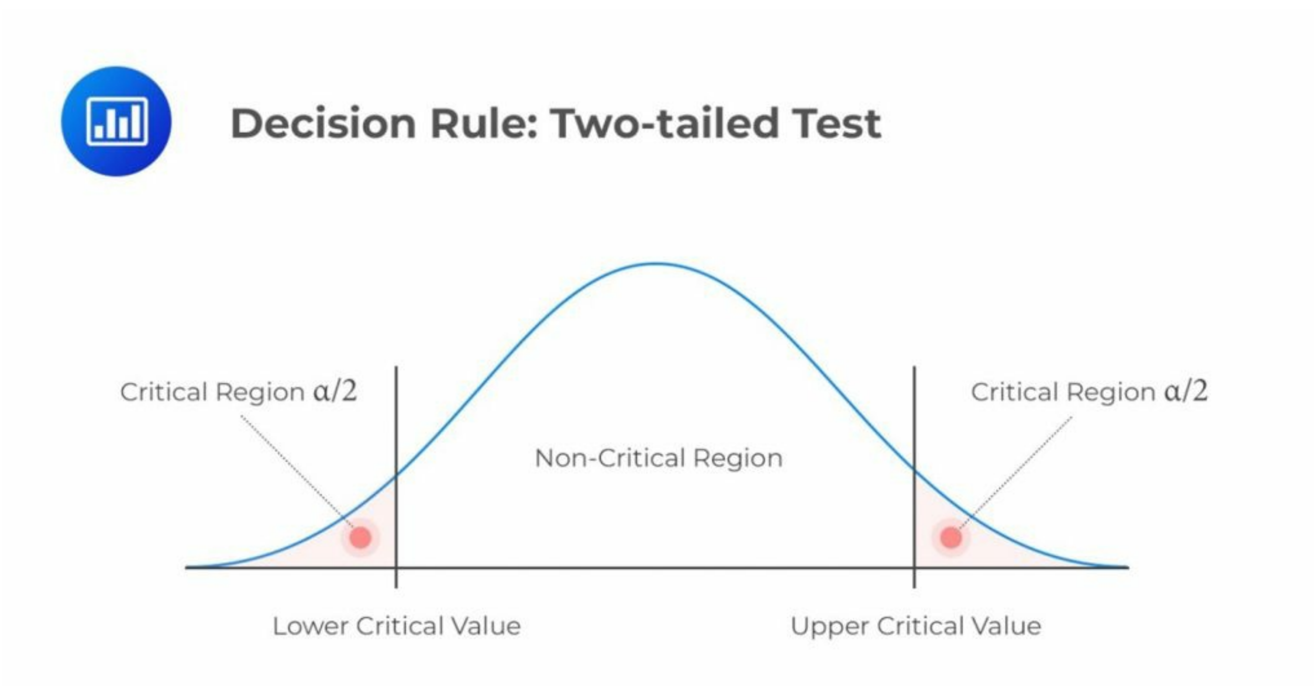
Decision rule: Reject H_0 if the test statistic is greater than the critical value. Otherwise, **do not reject** H_0.



Two-tailed Test

H_a : Parameter $\neq$ X (not equal to X)

Decision rule: Reject H_0 if the test statistic is greater than the upper critical value or less than the lower critical value.



The p-Value in Hypothesis Testing

The p-value is the lowest level of significance at which we can reject a null hypothesis. The probability of coming up with a test statistic would justify our rejection of a null hypothesis, assuming that the null hypothesis is indeed true.

Breaking Down the p-value

When carrying out a statistical test with a fixed value of the significance level (?), we merely compare the observed test statistic with some critical value. For example, we might "reject an H_0 using a 5% test" or "reject an H_0 at a 1% significance level." The problem with this 'classical' approach is that it does not give us details about the **strength of the evidence** against the null hypothesis.

Determination of the p-value gives statisticians a more informative approach to hypothesis testing. The p-value is the lowest level at which we can reject an H_0. This means that the strength of the evidence against an H_0 increases as the p-value becomes smaller.

In one-tailed tests, the p-value is the probability below the calculated test statistic for left-tailed tests or above the test statistic for right-tailed tests. For two-tailed tests, we find the probability below the negative test statistic and add it to the probability above the positive test statistic. This combines both tails for the p-value calculation.

Example: p-value

θ represents the probability of obtaining a head when a coin is tossed. Assume we tossed a coin 200 times, and the head came up in 85 out of the 200 trials. Test the following hypothesis at a 5% level of significance.

$H_0: \ \theta = 0.5$

$H_1: \ \theta < 0.5$

Solution

First, note that repeatedly tossing a coin follows a binomial distribution.

Our p-value will be given by $P(X < 85)$, where X follows a binomial (200,0.5), assuming the H_0 is true.

$$= P[Z < \frac{(85–100)}{\sqrt{50}}]$$
$$= P(Z < -2.12) = 1 - 0.9834 = 0.01660$$

(We have applied the Central Limit Theorem by taking the binomial distribution as approximately normal.)

Since the probability is less than 0.05, the H0 is extremely unlikely, and we have strong evidence against an H_0 that favors H_1. Therefore, clearly expressing this result, we could say:

“There is very strong evidence against the hypothesis that the coin is fair. We, therefore, conclude that the coin is biased against heads.”

Remember, failure to reject a H_0 does not mean it is true. It means there is insufficient evidence to justify the rejection of the H_0 given a certain level of significance.

Question

A CFA candidate conducts a statistical test about the mean value of a random variable X.

$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

She obtains a test statistic of 2.2. Given a 5% significance level, determine the p-value.

A. 1.39%.

B. 2.78.

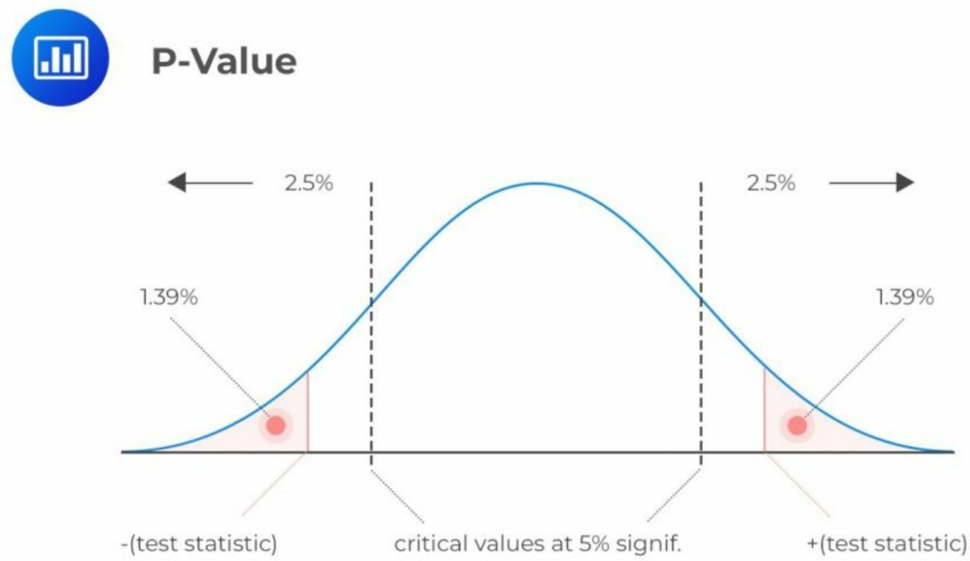
C. 2.78%.

Solution

The correct answer is C.

$$P - \text{value} = P(Z > 2.2) = 1 - P(Z < 2.2) = 1.39\% \times 2 = 2.78\%$$

(We have multiplied by two since this is a two-tailed test.)



Interpretation: The p-value (2.78%) is less than the significance level (5%). Therefore, we have sufficient evidence to reject the H_0. In fact, the evidence is so strong that we would also reject the H_0 at significance levels of 4% and 3%. However, at significance levels of 2% or 1%, we would not reject the H0 since the p-value surpasses these values.

LOS 8b: Construct hypothesis tests and determine their statistical significance, the associated Type I and Type II errors, and the power of the test given a significance level

Hypothesis Test Concerning Single Mean

The z-test is the ideal hypothesis test when the sampling distribution of the sample is normally distributed or when the standard deviation is known.

The z-statistic is the test statistic used in hypothesis testing.

Testing $H_0 : \mu = \mu_0$ Using the z-test

Given a random sample of size n from a normally distributed population with mean μ, variance σ^2, and a sample mean $\bar{X}$, we can compute the z-statistic as follows:

$$z - \text{statistic} = \frac{(\bar{X} - \mu_0)}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

Where:

$\bar{X}$ is the sample mean.

$?_0$ is the hypothesized mean of the population.

σ is the standard deviation of the population.

n is the sample size.

Once computed, the z-statistic is compared to the critical value that corresponds to the level of significance of the test. For example, if the significance level is 5%, the z-statistic is screened against the upper or lower 95% point of the normal distribution (±1.96). The decision rule is to reject the H0 if the z-statistic falls within the critical or rejection region.

Example: z-test

Academics carried out a study on 50 former United States presidents and found an average IQ of 135. You are required to carry out a 5% statistical test to determine whether the average IQ of presidents is greater than 130. (IQs are distributed normally, and previous studies indicate that $\sigma = 25$.)

Solution

Step 1: State the hypothesis:

$H_0 : \mu \leq 130$

$H_1 : \mu > 130$

Step 2: Identify the appropriate t-statistic:

$$z - \text{statistic} = \frac{(\bar{X} - \mu_0)}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

Assuming the H_0 is true, $\frac{(\bar{X}-130)}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0, 1)$

Step 3: Specify the level of significance:

This is a right-tailed test. Therefore, we compare our test statistic to the upper 95% point of the standard normal distribution (1.645).

Step 4: State the decision rule:

Reject the null hypothesis if the z-statistic is greater than 1.645.

Step 5: Calculate the test statistic:

The z-statistic is $\frac{(135 – 130)}{\left(\frac{25}{\sqrt{50}}\right)} = 1.414$

Step 6: Make a decision:

Since 1.414 is less than 1.645, we **do not have** sufficient evidence to reject the H_0. As such, it would be **reasonable** to conclude that the average IQ of U.S. presidents is not more than 130.

The t-test

The t-test is based on the t-distribution. The test is appropriate for testing the value of a population mean when:

- σ is unknown.
- The sample size is large (n ? 30), and if $n < 30$, the distribution must be normal or approximately normal.

Testing $H_0: \mu = \mu_0$ Using the t-test

We compute a t-statistic with $n-1$ degrees of freedom as follows:

$$t_{n-1} = \frac{(\bar{X} - \mu_0)}{\left(\frac{s}{\sqrt{n}}\right)}$$

Where:

$\bar{X}$ is the sample mean.

μ_0 is the hypothesized mean of the population.

s is the standard deviation of the sample.

n is the sample size.

Example: t-Test

Financial analysts in a certain equatorial country are interested in evaluating the potential impact of rainfall on agricultural investments. They have gathered data on the annual rate of rainfall (cm) over the last 10 years, as shown below:

{ 25, 26, 25, 27, 28, 29, 28, 27, 26, 25 }

Previously, the recorded average rainfall was 23 cm. The analysts want to find out if there's been an increase in the average rainfall rate, which could impact agricultural investment. Conduct a statistical test at a 5% significance level to investigate this.

Solution

Follow the steps outlined above.

Step 1: State the hypothesis:

As always, you should begin by stating the hypothesis:

$H_0 : \mu \le 23$

$H_1 : \mu > 23$

Step 2: Identify the appropriate t-statistic:

If we assume that the annual rainfall quantities are distributed normally and recorded independently, then:

$$\frac{(\bar{X} - \mu_0)}{\left(\frac{s}{\sqrt{n}}\right)} \sim t_{n-1}$$

Please, confirm that $\bar{X} = 26.6$ and $s = 1.43$

Step 3: Specify the level of significance:

$$\alpha = 5\% \, (\text{right} - \text{tailed})$$

Step 4: State the decision rule:

Reject the null hypothesis if the t-statistic is greater than $t_{0.05,9} = 1.833$

Step 5: Calculate the test statistic:

Therefore, our t-statistic $= \frac{(26.6 - 23)}{\left(\frac{1.43}{\sqrt{10}}\right)} = 7.96$

Step 6: Make a decision:

Our test statistic (7.96) is greater than the upper 95% point of the $t_{0.05,9}$ distribution (1.833).

df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.31[illegible]752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.91[illegible]986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.35[illegible]363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.13[illegible]847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.01[illegible]048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.94[illegible]180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.89[illegible]579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.85[illegible]548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869

Therefore, we have **sufficient evidence** to reject the H0. As such, it is **reasonable** to conclude that the average annual rainfall has increased from its former long-term average of 23.

Question

What is the value of t in the example above if the significance level is reduced from 5% to 0.5%, and does this change the decision rule?

A. 2.02; it does not change the decision rule.

B. 3.25; it does not change the decision rule.

C. 3.25; it changes the decision rule.

Solution

The correct answer is B.

A quick glance at the $t_{0.005,9}$ distribution when $\alpha = 0.5\%$ gives a value of 3.25.

df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869

However, the evidence against the H_0 is overwhelming since our test statistic (7.96) is still greater than 3.25. As such, the conclusion would remain unchanged.

Hypothesis Test Concerning the Equality of the Population Means

Analysts are often interested in establishing whether there exists a significant difference between the means of two different populations. For instance, they might want to know whether the average returns for two subsidiaries of a given company exhibit a **significant** variance. Such a test may then be used to make decisions regarding resource allocation or the reward of the directors. Before embarking on such an exercise, it is paramount to ensure that the samples taken are independent and sourced from normally distributed populations. It can either be assumed that the population variances are equal or unequal. In this reading, we will assume that the population variances are equal.

Assume that μ_1 is the mean of the first population while μ_2 is the mean of the second population. In testing the equality of two population means, we wish to determine if they are equal or not. As such, the hypotheses can be any of the following:

I. **Two-sided**:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ vs. } H_a ? \mu_1 - \mu_2 \neq 0,$$

This can be stated as follows:

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a ? \mu_1 \neq \mu_2$$

II. **One-sided (right)**:

$$H_0 : \mu_1 - \mu_2 \leq 0 \text{ vs. } H_a ? \mu_1 - \mu_2 > 0,$$

This can be stated as follows:

$$H_0 : \mu_1 \leq \mu_2 \text{ vs. } H_a ? \mu_1 > \mu_2$$

III. **One-sided (Left)**:

$$H_0 : \mu_1 - \mu_2 \geq 0 \text{ vs. } H_a ? \mu_1 - \mu_2 < 0,$$

This can be stated as follows:

$$H_0 : \mu_1 \geq \mu_2 \text{ vs. } H_a?\mu_1 < \mu_2$$

However, note that tests such as $H_0 : \mu_1 - \mu_2 = 3$ vs. $H_a?\mu_1 - \mu_2 \neq 3$ are valid. The process is similar in both cases.

Hypothesis Test Concerning Differences between Means With Independent Samples

When testing for the difference between two population means, we assume that the two populations are distributed normally. Further, we assume that they have equal and unknown variances. We always make use of the student's t-distribution where the test statistic is given by:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p\left(\sqrt{\frac{1}{n1} + \frac{1}{n2}}\right)}$$

Where s_p^2 is the pooled estimator of the common variance and is given by:

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}$$

Also, the variables are defined as follows:

$\bar{X}_1$ = Mean of the first sample.

$\bar{X}_2$ = Mean of the second sample.

s_1^2 = variance of the first sample.

s_2^2 = variance of the second sample.

n_1 = Sample size of the first sample.

n_2 = Sample size of the second sample.

Example 1: Hypothesis Test Concerning the Equality of the Population Means

Nutritionists want to establish whether obese patients on a new special diet have a lower weight than the control group. After six weeks, the average weight of 10 patients (group A) on the special diet is 75kg, while that of 10 more patients of the control group (B) is 72kg. Carry out a 5% test to determine if the patients on the special diet have a lower weight.

Additional information: $\sum A^2 = 59520$ and $\sum B^2 = 56430$

Solution

As is the norm, start by stating the hypothesis:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ Vs } H_a ? \mu_1 - \mu_2 \neq 0,$$

We assume that the two samples have equal variances, are independent, and are normally distributed. Then, under H_0,

$$\frac{\bar{B} - \bar{A}}{S\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}$$

Note that the sample variance is given by:

$$s^2 = \frac{\sum X^2 - n\bar{X}^2}{n-1}$$

So,

$$S_A^2 = \frac{\{59520 - (10 * 75^2)\}}{9} = 363.33$$
$$S_B^2 = \frac{\{56430 - (10 * 72^2)\}}{9} = 510$$

Therefore,

$$S_2^p = \frac{(9 \times 363.33 + 9 \times 510)}{(10 + 10 - 2)} = 436.665$$

And

$$\text{Test statistic} = \frac{(75 - 72)}{\{\sqrt{439.665} \times \sqrt{(\frac{1}{10} + \frac{1}{10})}\}} = 0.3210$$

Our test statistic (0.3210) is less than the upper 5% point (1.734) of the t-distribution with 18 degrees of freedom.

df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.31[illegible]752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.91[illegible]986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.35[illegible]363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.13[illegible]847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.01[illegible]048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.94[illegible]180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.89[illegible]579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.85[illegible]548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.83[illegible]113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.81[illegible]461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.79[illegible]885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.78[illegible]288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.77[illegible]933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.76[illegible]310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.75[illegible]050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.74[illegible]884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.73[illegible]607	2.10982	2.56693	2.89823	3.9651
18	[illegible]	[illegible]	[illegible]	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495

Therefore, we **do not have sufficient** evidence to reject the H_0 at 5% significance. As such, it is **reasonable** to conclude that the special diet has the same effect on body weight as the placebo.

Note to candidates: You could choose to work with the p-value and determine $P(t_{18} > 0.937)$ and then establish whether this probability is less than 0.05. Working out the problem this way would lead to the same conclusion as above.

Example 2: **Hypothesis Test Concerning the Equality of the Population Means**

Suppose we replace '>' with '≠' in H_1 in the example above, would the decision rule change?

Replacing '>' with '≠' in H_1 would change the test from a one-tailed one to a two-tailed test. We would compute the test statistic just as demonstrated above. However, we would have to divide the level of significance by two and compare the test statistic to both the lower and upper 2.5% points of the t18-distribution (±2.101).

df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495

Since our test statistic lies within these limits (non-rejection region), the decision rule would remain unchanged.

Hypothesis Test Concerning Differences between Means With Dependent Samples

There are some challenges when testing the difference between two population means using independent samples. Variability within each sample, caused by factors unrelated to the research, can obscure the real difference of interest. Random variation within a sample might be so substantial that it obscures the actual difference caused by the specific phenomenon the analyst is studying.

When we want to test the differences between means with dependent samples, we use the paired comparison test (test of the mean of the differences).

Paired Comparisons Test

Assume that we have observations for the random variables X_A and X_B and that the samples are dependent.

Organize the observations in pairs and denote the differences between the two paired observations by d_i. That is $d_i = x_{Ai} - x_{Bi}$ where x_{Ai} and x_{Bi} are the ith pair of observations $1, 2, \ldots, n$.

Also, let μ_d be the population mean difference and μ_{d0} be the hypothesized value for the population mean difference. At this point, we can state the hypotheses:

- **Two-sided**: $H_0 : \mu_d = \mu_{d0}$ versus $H_a : \mu_d \neq \mu_{d0}$
- **One-sided (right side)**: $H_0 : \mu_d \leq \mu_{d0}$ versus $H_a : \mu_d > \mu_{d0}$
- **One-sided (left side)**: $H_0 : \mu_d \geq \mu_{d0}$ versus $H_a : \mu_d < \mu_{d0}$

Practically, $\mu_{d0} = 0$.

We are considering normally distributed populations with unknown population variances. As such, we will use the t-distributed statistic given by:

$$t = \frac{\bar{d} - \mu_{d0}}{s_{\bar{d}}}$$

Where:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$s_{\bar{d}} = s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \text{standard error of the mean differences}$$

$$s_d = \text{standard deviation of the differences}$$

Note that the degree of freedom is $n - 1$ where n is the number of the paired observations.

Example 2: **Paired Comparison Test**

An analyst aims to compare the performance of the BCD High Growth Index and the BCD Investment Grade Index. They collect data for both indexes over 2,050 days and calculate the means and standard deviations, as shown in the table below:

	BCD High Growth Index (%)	BCD Investment Grade Index (%)	Difference (%)
Mean return	0.0183	0.0161	−0.0022
Standard deviation	0.2789	0.3298	0.3321

Using a 5% significance level, determine whether the mean of the differences is different from zero.

Solution

Step 1: State the hypothesis:

$$H_0 : \mu_{d0} = 0 \text{ vs } H_a : \mu_{d0} \neq 0$$

Step 2: Identify the appropriate t-statistic:

We use the t-statistic, calculated by:

$$t = \frac{\bar{d} - \mu_{d0}}{s_{\bar{d}}}$$

Step 3: Specify the level of significance:

$$\alpha = 5\% \text{ (two-tailed test)}$$

Step 4: State the decision rule:

The degrees of freedom amount to $n - 1 = 2{,}050 - 1 = 2{,}049$; thus, the critical values are ± 1.960. Therefore, we will reject the null hypothesis if the calculated t-statistic is less than -1.96 or greater than 1.96.

Step 5: Calculate the test statistic:

Note that from the table $\bar{d} = -0.0022$ and $s_d = 0.3321$ so that the t-statistic is given:

$$t = \frac{\bar{d} - \mu_{d0}}{s_{\bar{d}}} = \frac{-0.0022 - 0}{\frac{0.3321}{\sqrt{2050}}} = -0.30$$

Step 6: Make a Decision:

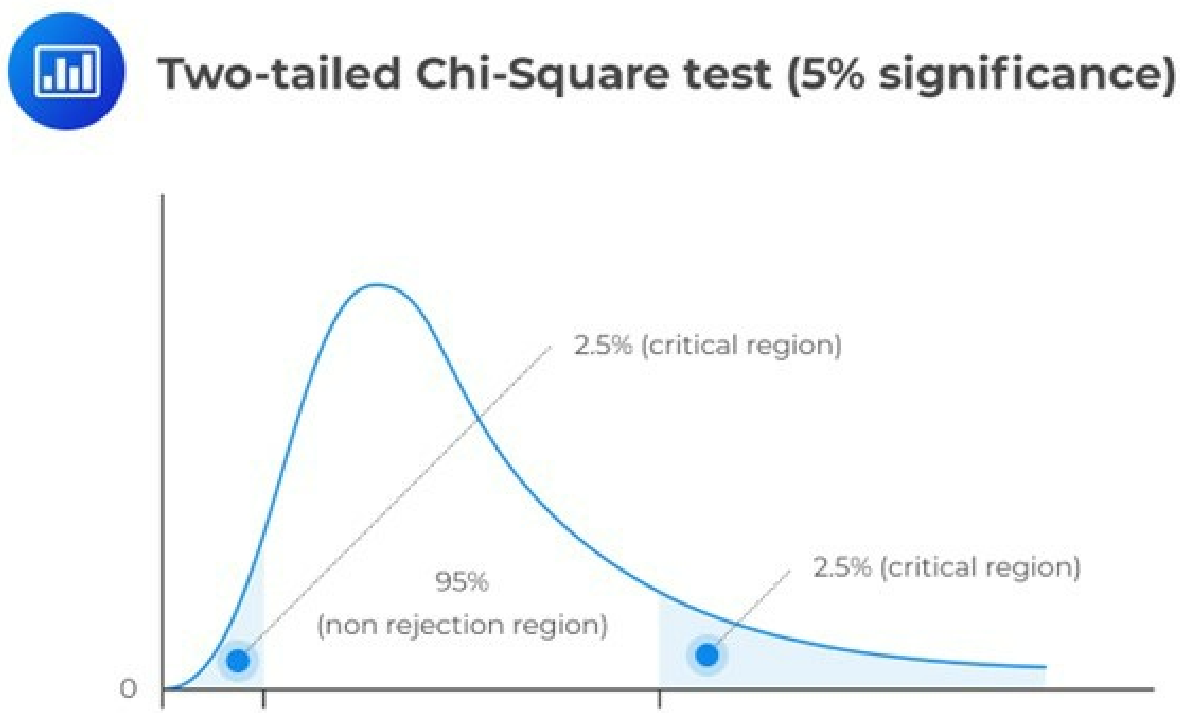
-0.30 falls within the bounds of the critical values of ±1.960. As such, there is insufficient evidence to show that the mean of the differences in returns differs from zero.

ν \ α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	$\boxed{1.960}$	2.326	2.576	3.090	3.291

Test Variance and Equality of Variance

Testing of a Single Variance

A chi-square test helps determine if a hypothesized variance value matches the true population variance. Unlike other distributions in the CFA® Program, the chi-square distribution is asymmetrical. Yet, as degrees of freedom increase, it approaches a more normal distribution.



As a natural consequence, the chi-square distribution has no negative values and is bounded by zero. A chi-square statistic with $(n-1)$ degrees of freedom is computed as:

$$\chi^2_{n-1} = \frac{(n-1)\,S^2}{\sigma_0^2}$$

Where:

n = Sample size.

S^2 = Sample variance.

σ_0^2 = Hypothesized population variance.

Example: Chi-square Test

For the 15-year period between 1995 and 2010, ABC's monthly return had a standard deviation of 5%. John Matthew, CFA, wishes to establish whether the standard deviation witnessed during that period still adequately describes the long-term standard deviation of the company's return. To achieve this end, he collects data on the monthly returns recorded between January 1, 2015, and December 31, 2016, and computes a monthly standard deviation of 4%.

Carry out a 5% test to determine if the standard deviation computed in the latter period differs from the 15-year value.

Solution

As is the norm, start by writing down the hypothesis.

$H_0 : \sigma_0^2 = 0.0025$

$H_1 : \sigma^2 \neq 0.0025$

Since the latter period has 24 months, n = 24, the test statistic is:

$$\chi^2_{24-1} = \frac{(24-1)\,0.0016}{0.0025} = 14.72$$

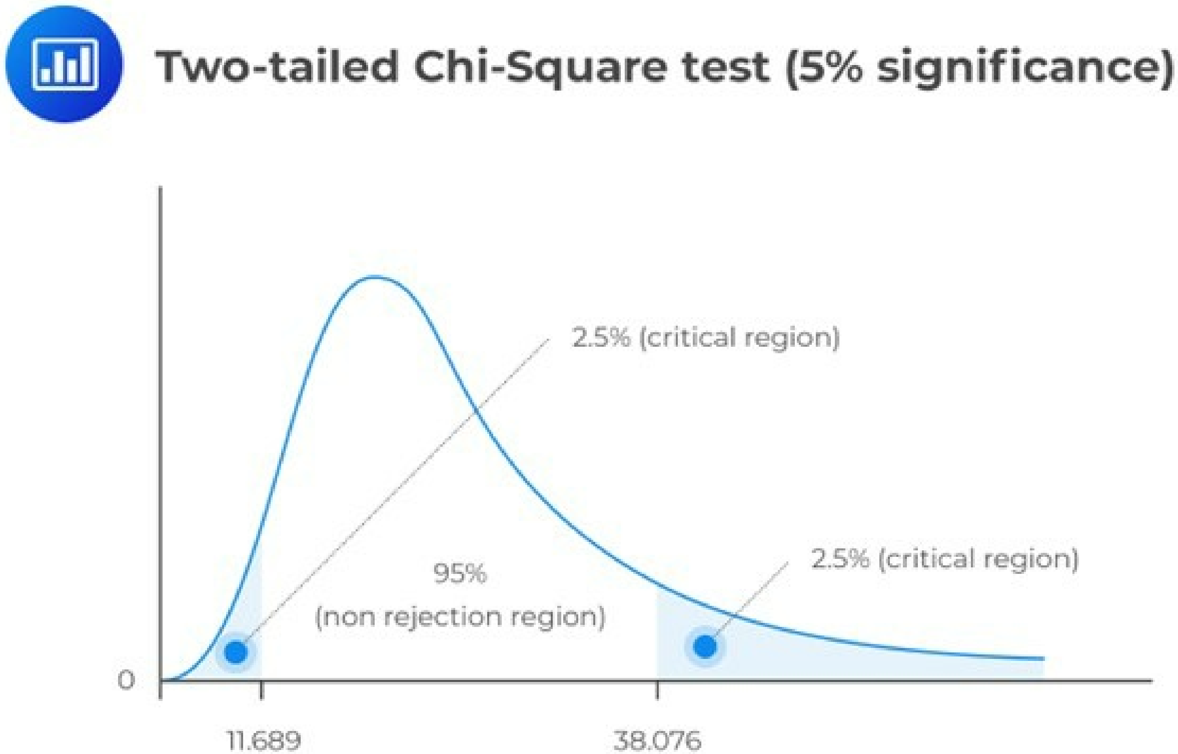
This is a two-tailed test. As such, we have to divide the significance level by two and screen our test statistic against the lower and upper 2.5% points of χ^2_{23}.

Consulting the chi-square table, the test statistic (14.72) lies between the lower (11.689) and the upper (38.076) 2.5% points of the chi-square distribution.

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.0[illegible]01	0.004	0.016	2.706	3.841	5.[illegible]24	6.635
2	0.020	0.0[illegible]51	0.103	0.211	4.605	5.991	7.[illegible]78	9.210
3	0.115	0.[illegible]16	0.352	0.584	6.251	7.815	9.[illegible]48	11.345
4	0.297	0.[illegible]84	0.711	1.064	7.779	9.488	11.[illegible]43	13.277
5	0.554	0.[illegible]31	1.145	1.610	9.236	11.071	12.[illegible]33	15.086
6	0.872	1.[illegible]37	1.635	2.204	10.645	12.592	14.[illegible]49	16.812
7	1.239	1.[illegible]90	2.167	2.833	12.017	14.067	16.[illegible]13	18.475
8	1.646	2.[illegible]80	2.733	3.490	13.362	15.507	17.[illegible]35	20.090
9	2.088	2.[illegible]00	3.325	4.168	14.684	16.919	19.[illegible]23	21.666
10	2.558	3.[illegible]47	3.940	4.865	15.987	18.307	20.[illegible]83	23.209
11	3.053	3.[illegible]16	4.575	5.578	17.275	19.675	21.[illegible]20	24.725
12	3.571	4.[illegible]04	5.226	6.304	18.549	21.026	23.[illegible]37	26.217
13	4.107	5.[illegible]09	5.892	7.042	19.812	22.362	24.[illegible]36	27.688
14	4.660	5.[illegible]29	6.571	7.790	21.064	23.685	26.[illegible]19	29.141
15	5.229	6.[illegible]62	7.261	8.547	22.307	24.996	27.[illegible]88	30.578
16	5.812	6.[illegible]08	7.962	9.312	23.542	26.296	28.[illegible]45	32.000
17	6.408	7.[illegible]64	8.672	10.085	24.769	27.587	30.[illegible]91	33.409
18	7.015	8.[illegible]31	9.390	10.865	25.989	28.869	31.[illegible]26	34.805
19	7.633	8.[illegible]07	10.117	11.651	27.204	30.144	32.[illegible]52	36.191
20	8.260	9.[illegible]91	10.851	12.443	28.412	31.410	34.[illegible]70	37.566
21	8.897	10.[illegible]83	11.591	13.240	29.615	32.671	35.[illegible]79	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	[illegible]	$\boxed{11.689}$	13.091	14.848	32.007	35.172	$\boxed{38.076}$	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892

Note that you will be given a simplified critical value table in the exam situation.



Evidently, we have insufficient evidence to reject the H_0. It is, therefore, reasonable to conclude that the latter standard deviation value is close enough to the 15-year value.

Test Concerning the Equality of the Variances

To test the equality concerning the variances, we use the F-test. Assume that we have 2 independent random samples of sizes n1 and n2 from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.

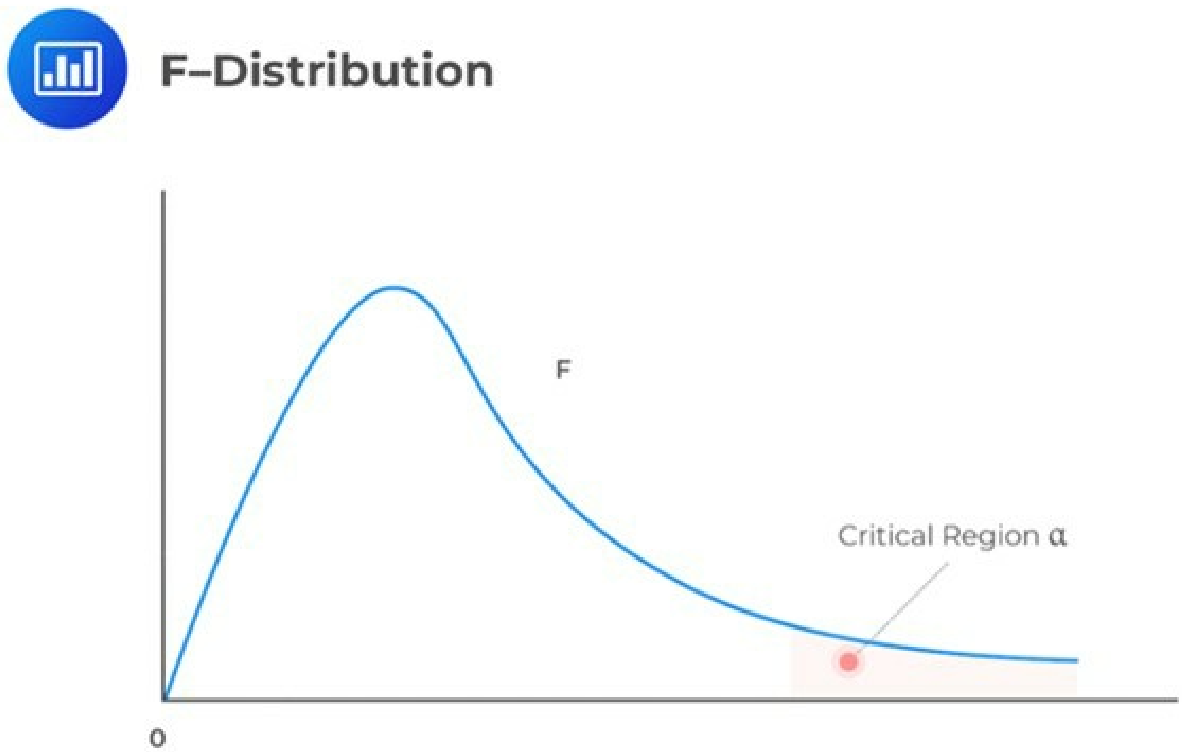
Also, let us consider a scenario where we have the sample variances as S_1^2 and S_2^2. The basic situation usually involves the following hypothesis:

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_a : \sigma_1^2 \neq \sigma_2^2$

The test statistic is $\frac{S_1^2}{S_2^2} \sim F_{n1-1,n2-1}$ under H_0.

The decision rule is to reject the null hypothesis if the test statistic falls within the critical region of the F-distribution.



Example: F-test

An analyst is studying whether the population variance of returns on a commodity index changed after the introduction of new trading guidelines. The first 320 weeks elapsed before the guidelines were introduced, and the second 320 weeks came after the introduction. The analyst gathers the data in the table below for 320 weeks of returns both before and after the change in guidelines.

	Mean Weekly Return (%)	Variance of Returns
Before guidelines change	0.180	3.520
After guidelines change	0.090	2.967

Do the variances of returns differ before and after the guideline change? Employ a 5 percent significance level.

Solution

Step 1: State the hypothesis:

$$H_0 : \sigma^2_{\text{Before}} = \sigma^2_{\text{After}} \text{ vs } H_a : \sigma^2_{\text{Before}} \neq \sigma^2_{\text{After}}$$

Step 2: Identify the appropriate t-statistic:

$$t = \frac{s^2_{\text{Before}}}{s^2_{\text{After}}}$$

Step 3: Specify the level of significance:

$$\alpha = 5\% \text{ (two-tailed)}$$

Step 4: State the decision rule:

$$\begin{aligned} \text{Left side} &= 0.803 \\ \text{Right side} &= 1.246 \end{aligned}$$

Reject the null if the calculated t-statistic is less than 0.803 and reject the null if the calculated t-statistic is greater than 1.246.

Step 5: Calculate the test statistic:

$$t = \frac{s^2_{\text{Before}}}{s^2_{\text{After}}} = \frac{3.520}{2.967} = 1.1864$$

Step 6: Make a decision:

Fail to reject the null hypothesis because 1.1864 falls within the bounds of the critical values of [0.80,1.246]. There is not sufficient evidence to indicate that the weekly variances of returns are different in the periods before and after the guidelines change.

LOS 8c: Compare and contrast parametric and nonparametric tests, and describe situations where each is the more appropriate type of test

Parametric Tests

Parametric tests are statistical tests in which we make assumptions regarding population distribution. Such tests involve the estimation of the key parameters of a distribution. For example, we may wish to estimate the mean or compare population proportions.

When conducting statistical tests, the choice of distribution directly influences how the test statistic is calculated. The tests we've discussed are considered parametric tests. For example, assuming a parameter follows a normal distribution leads to the computation of the z-statistic.

During parametric testing, approximating normal distributions for non-normal data may be required. This approximation is valid due to the central limit theorem, which states that as sample sizes increase, non-normal distributions tend to become more normal.

Parametric tests are generally considered to be stronger than nonparametric ones.

Nonparametric Tests

Nonparametric tests, sometimes known as distribution-free tests, don't assume anything about the parameter's distribution being studied. Researchers turn to nonparametric testing when they have concerns about factors other than the parameter's distribution.

The following table gives the alternative nonparametric tests for the parametric tests.

	Parametric Test	Non-parametric Test
Test concerning single mean	t-distributed test	Wilcoxon signed-rank test
Tests concerning differences between means	z-distributed test	Mann–Whitney U test (Wilcoxon rank sum test)
Test concerning mean differences (pair comparison test)	t-distributed test	Wilcoxon signed-rank test Sign test

Situations Where Nonparametric Tests are Appropriate

The data do not meet distributional assumptions:

This happens when the distributional assumptions of the parametric tests are not met. For instance, we may find parametric tests such as t-test are inappropriate because the sample size is small and may be drawn from non-normally distributed. As such, a nonparametric test is appropriate.

When there are outliers:

Outliers can affect the parametric statistics. On the other hand, outliers do not affect parametric tests.

Consider a situation where we want to establish the center of a rather skewed distribution, such as that of the income of the residents of a given city. While the majority of the residents could be categorized as the middle class, the presence of just a few billionaires in a sample can greatly increase the mean income. Such a mean, therefore, may not provide a very reliable or realistic measure of income.

Instead, it may be more appropriate to use the median. Compared to the mean, the median can better represent the center of the income distribution. This is due to the fact that 50% of the residents will be above the median and the remaining 50% below it.

In summary, “outliers” affect the mean when dealing with skewed data. The median, on the other hand, sticks closer to the center of the distribution.

When data is given in the form of ranks or uses an ordinal scale:

Although nonparametric tests are usually easier to conduct than parametric ones, they do not have as much statistical power. Nonetheless, they provide an efficient tool for analyzing ordinal, ranked, or very skewed data.

When the hypotheses do not concern a parameter:

We often use nonparametric tests, such as the runs test, when our goal is to determine if a sample from a population isn't random. Since randomness isn't a parameter, nonparametric tests are the right choice in such cases.

Nonparametric inference:

Nonparametric methods make our statistical analysis broader. They work with limited assumptions and can be used for ordered data. Plus, they handle questions that aren't tied to specific parameters.

Nonparametric tests are commonly used alongside parametric tests. They help analysts understand how sensitive the statistical results are to the assumptions of parametric tests. But when the conditions for a parametric test are met, we usually choose it over nonparametric tests. We prefer parametric tests because they often have more statistical power, which means they are better at detecting when the null hypothesis is incorrect.