R06 Hypothesis Testing

1. Introduction 2
2. The Process of Hypothesis Testing 2
   2.1 Stating the Hypotheses 2
   2.2 Two-Sided vs. One-Sided Hypotheses 3
   2.3 Selecting the Appropriate Hypotheses 3
3. Identify the Appropriate Test Statistic 4
   3.1 Test Statistics 4
   3.2 Identifying the Distribution of the Test Statistic 4
4. Specify the Level of Significance 5
5. State the Decision Rule 6
   5.1 Determining Critical Values 6
   5.2 Decision Rules and Confidence Intervals 7
   5.3 Collect the Data and Calculate the Test Statistic 7
6. Make a Decision 7
7. The Role of p-Values 8
8. Multiple Tests and Interpreting Significance 8
9. Tests Concerning a Single Mean 8
10. Test Concerning Differences Between Means with Independent Samples 10
11. Test Concerning Differences Between Means with Dependent Samples 11
12. Testing Concerning Tests of Variances (Chi-Square Test) 13
   12.1 Tests of a Single Variance 13
   12.2 Test Concerning the Equality of Two Variances (F-Test) 15
13. Parametric vs. Nonparametric Tests 16
14. Tests Concerning Correlation 16
   14.1 Parametric Test of a Correlation 17
   14.2 Tests Concerning Correlation: The Spearman Rank Correlation Coefficient 18
15. Test of Independence Using Contingency Table Data 18
Summary 19

This document should be read in conjunction with the corresponding reading in the 2022 Level I CFA® Program curriculum. Some of the graphs, charts, tables, examples, and figures are copyright 2021, CFA Institute. Reproduced and republished with permission from CFA Institute. All rights reserved.

Required disclaimer: CFA Institute does not endorse, promote, or warrant the accuracy or quality of the products or services offered by IFT. CFA Institute, CFA®, and Chartered Financial Analyst® are trademarks owned by CFA Institute.

Version 1.0

1. Introduction

Hypothesis testing is the process of making judgments about a larger group (a population) on the basis of observing a smaller group (a sample). The results of such a test then help us evaluate whether our hypothesis is true or false.

For example, let's say you are a researcher and you believe that the average return on all Asian stocks was greater than 2%. To test this belief, you can draw samples from a population of all Asian stocks and employ hypothesis testing procedures. The results of this test can tell you if your belief is statistically valid.

In this reading we will look at hypothesis tests concerning the mean, variance and correlation.

2. The Process of Hypothesis Testing

A hypothesis is defined as a statement about one or more populations. In order to test a hypothesis, we follow these steps:

1. State the hypothesis.
2. Identify the appropriate test statistic and its probability distribution.
3. Specify the significance level.
4. State the decision rule.
5. Collect data and calculate the test statistic.
6. Make a decision.

2.1 Stating the Hypotheses

For each hypothesis test, we always state two hypotheses: the null hypothesis (H_0) and the alternative hypothesis (H_a).

Null hypothesis (H_0): It is the hypothesis that the researcher wants to reject.

Alternative hypothesis (H_a)**:** It is the hypothesis that the researcher wants to prove. If the null hypothesis is rejected then the alternative hypothesis is considered valid.

Suppose you are a researcher and believe that the average return on all Asian stocks was greater than 2%. In this case, you are making a statement about the population mean (μ) of all Asian stocks.

For this example, the null and alternative hypotheses are:

H_0: $\mu \leq 2\%$

H_a: $\mu > 2\%$

(The value 2% is known as μ_0, the hypothesized value of the population mean.)

Instructor's Note:
An easy way to differentiate between the two hypotheses is to remember that the null

© IFT. All rights reserved

hypothesis always contains some form of the equal sign.

2.2 Two-Sided vs. One-Sided Hypotheses

The alternative hypothesis can be one-sided or two-sided depending on the proposition being tested. A one-sided test is also called a one-tailed test, and a two-sided test is also called a two-tailed test.

If we want to determine whether the estimated value of a population parameter is less than (or greater than) a hypothesized value we use a one-tailed test. However, if we want to determine whether the estimated value of a population parameter is different than a hypothesized value, we use a two tailed test.

Two-sided test: Suppose we want to test if the population mean is equal to 2%. The null and alternative hypothesis can be expressed as:

H_0: $\mu = 2\%$

H_a: $\mu \neq 2\%$

Since the alternative hypothesis contains a ≠ sign this is a two-sided test.

One-sided test (right side): Suppose we want to test if the population mean is greater than 2%. The null and alternative hypothesis can be expressed as:

H_0: $\mu \leq 2\%$

H_a: $\mu > 2\%$

Since the alternative hypothesis contains a > sign this is a one-sided test, and we are interested in the right side.

Instructor's Note
The sign in the alternative hypothesis points to the direction of the tail that we should use in our test. Since in our example the alternative hypothesis has a '>' sign it points to the right, therefore we are interested in the right tail.

One-sided test (left side): Suppose we want to test if the population mean is greater than 2%. The null and alternative hypothesis can be expressed as:

H_0: $\mu \geq 2\%$

H_a: $\mu < 2\%$

Since the alternative hypothesis contains a < sign this is a one-sided test, and we are interested in the left side.

2.3 Selecting the Appropriate Hypotheses

The easiest approach is to specify the alternative hypothesis first and then specify the null. Using a '<' or '>' sign in the alternative hypothesis instead of a '≠' sign reflects that belief of the researcher more strongly. However, a researcher may sometimes select a two-sided test

© IFT. All rights reserved

to depict neutrality in his beliefs.

3. Identify the Appropriate Test Statistic

3.1 Test Statistics

A test statistic is calculated from sample data and is compared to a critical value to decide whether or not we can reject the null hypothesis. The test statistic that should be used depends on what we are testing. For example, the test statistic for the test of a population mean is calculated as:

$$\text{test statistic} = \frac{\text{sample statistic} - \text{value of the parameter under } H_0}{\text{standard error of the sample statistic}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Continuing with our Asian stocks example, suppose we want to test if the population mean is greater than a particular hypothesized value. We draw 36 observations and get a sample mean of 4. We are also told that the standard deviation of the population is 4. If the hypothesized value of the population mean is 2, the test statistic is calculated as:

$$\text{test statistic} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{4 - 2}{\frac{4}{\sqrt{36}}} = 3$$

However, if the hypothesized value of the population mean is 0, then the test statistic is calculated as:

$$\text{test statistic} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{4 - 0}{\frac{4}{\sqrt{36}}} = 6$$

3.2 Identifying the Distribution of the Test Statistic

Exhibit 4 from the curriculum shows which test-statistics should be used depending on what we want to test and their corresponding distributions.

Instructor's Note: You will understand this table better, after you finish reading the remaining sections.

What We Want to Test	Test Statistic	Probability Distribution of the Statistic	Degrees of Freedom
Test of a single mean	$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$	t-Distributed	$n - 1$
Test of the 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-Distributed	$n_1 + n_2 - 2$

© IFT. All rights reserved

Test of the mean of differences	$t = \frac{\bar{d} - \mu_{d0}}{s_{\bar{d}}}$	*t*-Distributed	$n - 1$
Test of a single variance	$\chi^2 = \frac{s^2(n-1)}{\sigma_0^2}$	Chi-square distributed	$n - 1$
Test of the difference in variances	$F = \frac{s_1^2}{s_2^2}$	*F*-distributed	$n_1 - 1, n_2 - 1$
Test of a correlation	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	*t*-Distributed	$n - 2$
Test of independence (categorical data)	$\chi^2 = \sum_{i=1}^{m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$	Chi-square distributed	$(r - 1)(c - 1)$

4. Specify the Level of Significance

In reaching a statistical decision, we can make two possible errors:

- Type I error: We may reject a true null hypothesis.
- Type II error: We may fail to reject a false null hypothesis.

The following table shows the possible outcomes of a test.

Decision	**True condition**	
	H_0 true	**H_0 false**
Do not reject H_0	Correct decision	Type II error
Reject H_0 (accept H_a)	Type I error	Correct decision

The probability of a Type I error is also known as the **level of significance** of a test and is denoted by 'α'. A related term, **confidence level** is calculated as (1 - α). For example, a level of significance of 5% for a test means that there is a 5% probability of rejecting a true null hypothesis and corresponds to the 95% confidence level.

Controlling the two types of errors involves a trade-off. If we decrease the probability of a Type I error by specifying a smaller significance level (for e.g., 1% instead of 5%), we increase the probability of a Type II error. The only way to reduce both types of error simultaneously is by increasing the sample size, n.

The probability of a Type II error is denoted by 'β'. The **power of test** is calculated as (1 - β). It represents the probability of correctly rejecting the null when it is false.

The different probabilities associated with the hypothesis testing decisions are presented in

© IFT. All rights reserved

the table below:

Decision	True condition	
	H_0 true	**H_0 false**
Do not reject H_0	Confidence level ($1 - \alpha$)	β
Reject H_0 (accept H_a)	Level of significance (α)	Power of the test ($1 - \beta$)

The most commonly used levels of significance are: 10%, 5% and 1%.

5. State the Decision Rule

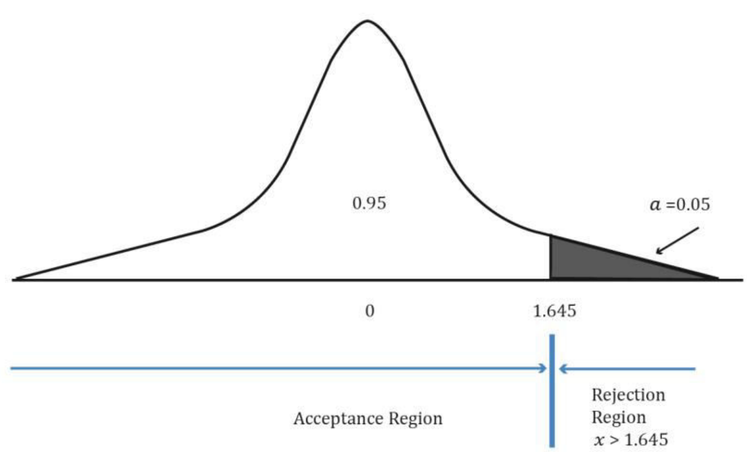
A decision rule involves determining the critical values based on the level of significance; and comparing the test statistic with the critical values to decide whether to reject or not reject the null hypothesis. When we reject the null hypothesis, the result is said to be statistically significant.

5.1 Determining Critical Values

One-tailed test:

Continuing with our Asian stocks example, suppose we want to test if the population mean is greater than 2%. Say we want to test our hypothesis at the 5% significance level. This is a one-tailed test and we are only interested in the right tail of the distribution. (If we were trying to assess whether the population mean is less than 2%, we would be interested in the left tail.)

The critical value is also known as the **rejection point** for the test statistic. Graphically, this point separates the acceptance and rejection regions for a set of values of the test statistic. This is shown below:



The region to the left of the test statistic is the 'acceptance region'. This represents the set of values for which we do not reject (accept) the null hypothesis. The region to the right of the

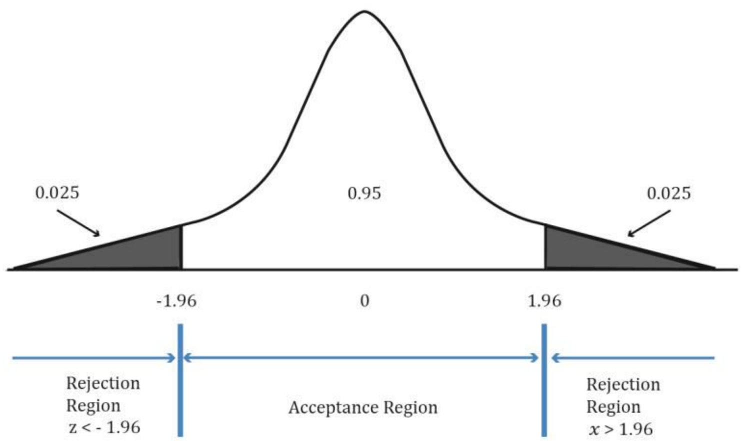
© IFT. All rights reserved

test statistic is known as the 'rejection region'.

Using the Z –table and 5% level of significance, the critical value = $Z_{0.05}$= 1.65

Two-tailed test:

In a two-tailed test, two critical values exist, one positive and one negative. For a two-sided test at the 5% level of significance, we split the level of significance equally between the left and right tail i.e. $\frac{0.05}{2}$ = 0.025 in each tail.
This corresponds to rejection points of +1.96 and -1.96. Therefore, we reject the null hypothesis if we find that the test statistic is less than -1.96 or greater than +1.96. We fail to reject the null hypothesis if -1.96 ≤ test statistic ≤ +1.96. Graphically, this can be shown as:



5.2 Decision Rules and Confidence Intervals

The above figure also illustrates the relationship between confidence intervals and hypothesis tests. A confidence interval specifies the range of values that may contain the hypothesized value of the population parameter. The 5% level of significance in the hypothesis tests corresponds to a 95% confidence interval. When the hypothesized value of the population parameter is outside the corresponding confidence interval, the null hypothesis is rejected. When the hypothesized value of the population parameter is inside the corresponding confidence interval, the null hypothesis is not rejected.

5.3 Collect the Data and Calculate the Test Statistic

In this step we first ensure that the sampling procedure does not include biases, such as sample selection or time bias. Then, we cleanse the data by removing inaccuracies and other measurement errors in the data. Once we are convinced that the sample data is unbiased and accurate, we use it to calculate the appropriate test statistic.

6. Make a Decision

A statistical decision simply consists of rejecting or not rejecting the null hypothesis. If the

© IFT. All rights reserved

test statistic lies in the rejection region, we will reject H_0. On the other hand, if the test statistic lies in the acceptance region, then we cannot reject H_0.

An economic decision takes into consideration all economic issues relevant to the decision, such as transaction costs, risk tolerance, and the impact on the existing portfolio. Sometimes a test may indicate a result that is statistically significant, but it may not be economically significant.

7. The Role of p-Values

The p-value is the smallest level of significance at which the null hypothesis can be rejected. It can be used in the hypothesis testing framework as an alternative to using rejection points.

- If the p-value is lower than our specified level of significance, we reject the null hypothesis.
- If the p-value is greater than our specified level of significance, we do not reject the null hypothesis.

For example, if the p-value of a test is 4%, then the hypothesis can be rejected at the 5% level of significance, but not at the 1% level of significance.

Relationship between test-statistic and p-value

A high test-statistic implies a low p-value.

A low test-statistic implies a high p-value.

8. Multiple Tests and Interpreting Significance

A Type I error represents a false positive result – rejecting the null when it is true. When multiple hypothesis tests are run on the same population, some tests will give false positive results. The expected portion of false positives is called the **false discovery rate** (FDR). For example, if you run 100 tests and use a 5% level of significance, you will get five false positives on average. This issue is called the **multiple testing problem**.

To overcome this issue, the **false discovery approach** is used to adjust the p-values when you run multiple tests. The researcher first ranks the p-values from the various tests from lowest to highest. He then makes the following comparison, starting with the lowest p-value (with k = 1), p(1):

$$p(1) \leq \alpha \frac{\text{Rank of i}}{\text{Number of tests}}$$

This comparison is repeated until we find the highest ranked p(k) for which this condition holds. For example, say we perform this check for k=1, k=2, k=3, and k=4; and we find that the condition holds true for k=4. Then we can say that the first four tests ranked on the basis of the lowest p-values are significant.

© IFT. All rights reserved

9. Tests Concerning a Single Mean

One of the decisions we need to make in hypothesis testing is deciding which test statistic and which corresponding probability distribution to use. We use the following table to make this decision:

Sampling from		Small sample size (n<30)	Large sample size (n≥30)
Normal distribution	Variance known	z	z
	Variance unknown	t	t (or z)
Non –normal distribution	Variance known	NA	z
	Variance unknown	NA	t (or z)

If the population variance is known and our sample size is large, we can use the z-statistic and z-distribution to compute the critical value. However, if we do not know the population variance and we have a small sample size, then we have to use the t-statistic and t-distribution to compute the critical values.

Example

An analyst believes that the average return on all Asian stocks was *less* than 2%. The sample size is 36 observations with a sample mean of -3. The standard deviation of the population is 4. Will he reject the null hypothesis at the 5% level of significance?

Solution:

In this case, our null and alternative hypotheses are:

H_0: $\mu \geq 2$

H_a: $\mu < 2$

The standard error of the sample is: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{36}} = 0.67$

The test statistic is:

$$\text{test statistic} = \frac{\text{sample statistic} - \text{value of the parameter under } H_0}{\text{standard error of the sample statistic}} = \frac{-3-2}{0.67} = -7.5$$

The critical values corresponding to a 5% level of significance is -1.65.

When we consider the left tail of the distribution, our decision rule is as follows: Reject the null hypothesis if the test statistic is less than the critical value and vice versa. Since our calculated test statistic of -7.5 is less than the critical value of -1.65, we reject the null hypothesis.

Example

Fund Alpha has been in existence for 20 months and has achieved a mean monthly return of 2% with a sample standard deviation of 5%. The expected monthly return for a fund of this nature is 1.60%. Assuming monthly returns are normally distributed, are the actual results

© IFT. All rights reserved

consistent with an underlying population mean monthly return of 1.60%?

Solution:
The null and alternative hypotheses for this example will be:

H_0: μ = 1.60 versus H_a: μ ≠ 1.60

$$\text{test statistic} = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{2 - 1.60}{\frac{5}{\sqrt{20}}} = 0.36$$

Using this formula, we see that the value of the test statistic is 0.36.

The critical values at a 0.05 level of significance can be calculated from the t-distribution table. Since this is a two-tailed test, we should look at a 0.05/2 = 0.025 level of significance with df = n - 1 = 20 – 1 = 19. This gives us two values of -2.1 and +2.1.

Since our test statistic of 0.35 lies between -2.1 and +2.1, i.e., the acceptance region, we do not reject the null hypothesis.

10. Test Concerning Differences Between Means with Independent Samples

Instructor's Note:
Focus on the basics of this topic, the probability of being tested on the details is low.

In this section, we will learn how to calculate the difference between the means of two independent and normally distributed populations. We perform this test by drawing a sample from each group. If it is reasonable to believe that the samples are normally distributed and also independent of each other, we can proceed with the test. We may also assume that the population variances are equal or unequal. However, the curriculum focuses on tests under the assumption that the population variances are equal.

The test statistic is calculated as:

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

The term s_p^2 is known as the pooled estimator of the common variance. It is calculated by the following formula:

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

The number of degrees of freedom is $n_1 + n_2 - 2$.

Example
(*This is based on Example 9 from the curriculum.*)

An analyst wants to test if the returns for an index are different for two different time

© IFT. All rights reserved

periods. He gathers the following data:

	Period 1	**Period 2**
Mean	0.01775%	0.01134%
Standard deviation	0.31580%	0.38760%
Sample size	445 days	859 days

Note that these periods are of different lengths and the samples are independent; that is, there is no pairing of the days for the two periods.

Test whether there is a difference between the mean daily returns in Period 1 and in Period 2 using a 5% level of significance.

Solution:
The first step is to formulate the null and alternative hypotheses. Since we want to test whether the two means were equal or different, we define the hypotheses as:
H_0: $\mu_1 - \mu_2 = 0$
H_a: $\mu_1 - \mu_2 \neq 0$

We then calculate the test statistic:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(445 - 1)09973 + (859 - 1)15023}{445 + 859 - 2} = 0.1330$$

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{(\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2})^{1/2}} = \frac{(0.01775 - 0.01134) - 0}{(\frac{0.1330}{445} + \frac{0.1330}{859})^{1/2}} = 0.3099$$

For a 0.05 level of significance, we find the t-value for 0.05/2 = 0.025 using df = 445 + 859 - 2=1302. The critical t-values are ±1.962. Since our test statistic of 0.3099 lies in the acceptance region, we fail to reject the null hypothesis.

We conclude that there is insufficient evidence to indicate that the returns are different for the two time periods.

11. Test Concerning Differences Between Means with Dependent Samples

Instructor's Note:
Focus on the basics of this topic, the probability of being tested on the details is low.

In the previous section, in order to perform hypothesis tests on differences between means of two populations, we assumed that the samples were independent. What if the samples are not independent? For example, suppose you want to conduct tests on the mean monthly return on Toyota stock and mean monthly return on Honda stock. These two samples are believed to be dependent, as they are impacted by the same economic factors.

In such situations, we conduct a t-test that is based on data arranged in **paired observations.** Paired observations are observations that are dependent because they have

© IFT. All rights reserved

something in common.

We will now discuss the process for conducting such a t-test.

Example:

Suppose that we gather data regarding the mean monthly returns on stocks of Toyota and Honda for the last 20 months, as shown in the table below:

Month	**Mean return of Toyota stock**	**Mean monthly return of Honda stock**	**Difference in mean monthly returns (d_i)**
1	0.5%	0.4%	0.1%
2	0.7%	1.0%	-0.3%
3	0.3%	0.7%	-0.4%
…	…	…	…
20	0.9%	0.6%	0.3%
Average	0.750%	0.600%	**0.075%**

Here is a simplified process for conducting the hypothesis test:

Step 1: Define the null and alternate hypotheses

We believe that the mean difference is not 0. Hence the null and alternate hypotheses are:

$H_0: \mu_d = \mu_{d0}$ versus $H_a: \mu_d \neq \mu_{d0}$

μ_d stands for the population mean difference and μ_{d0} stands for the hypothesized value for the population mean difference.

Step 2: Calculate the test-statistic

Determine the sample mean difference using:

$$\bar{d} = \frac{1}{n}\sum_{i=0}^{n} d_i$$

For the data given, the sample mean difference is 0.075%.

Calculate the sample standard deviation. The process for calculating the sample standard deviation has been discussed in an earlier reading. The simplest method is to plug the numbers (0.1, -0.3, -0.4…0.3) into a financial calculator. The entire data set has not been provided. We'll take it as a given that the sample standard deviation is 0.150%.

Use this formula to calculate the standard error of the mean difference:

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

For our data this is 0.150 / $\sqrt{20}$ = 0.03354.

We now have the required data to calculate the test statistic using a t-test. This is calculated using the following formula using n - 1 degrees of freedom:

© IFT. All rights reserved

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

For our data, the test statistic is $\frac{0.075 - 0}{0.03354} = 2.24$.

Step 3: Determine the critical value based on the level of significance

We will use a 5% level of significance. Since this is a two-tailed test we have a probability of 2.5% (0.025) in each tail. This critical value is determined from a t-table using a one-tailed probability of 0.025 and df = 20 – 1 = 19. This value is 2.093.

Step 4: Compare the test statistic with the critical value and make a decision

In our case, the test statistic (2.23) is greater than the critical value (2.093). Hence we will reject the null hypothesis.

Conclusion: The data seems to indicate that the mean difference is not 0.

The hypothesis test presented above is based on the belief that the population mean difference is not equal to 0. If μ_{d0} is the hypothesized value for the population mean difference, then we can formulate the following hypotheses:

1. If we believe the population mean difference is greater than 0:
   $H_0: \mu_d \leq \mu_{d0}$ versus $H_a: \mu_d > \mu_{d0}$
2. If we believe the population mean difference is less than 0:
   $H_0: \mu_d \geq \mu_{d0}$ versus $H_a: \mu_d < \mu_{d0}$
3. If we believe the population mean difference is not 0:
   $H_0: \mu_d = \mu_{d0}$ versus $H_a: \mu_d \neq \mu_{d0}$

12. Testing Concerning Tests of Variances (Chi-Square Test)

Instructor's Note:
Focus on the basics of this topic, the probability of being tested on the details is low.

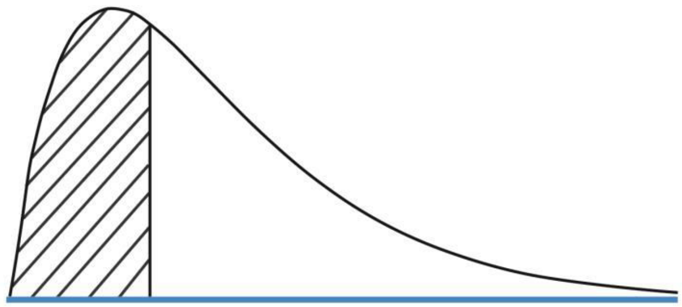
12.1 Tests of a Single Variance

In tests concerning the variance of a single normally distributed population, we use the chi-square test statistic, denoted by χ^2.

Properties of the chi-square distribution

The chi-square distribution is asymmetrical and like the t-distribution, is a family of distributions. This means that a different distribution exists for each possible value of degrees of freedom, n - 1. Since the variance is a squared term, the minimum value can only be 0. Hence, the chi-square distribution is bounded below by 0. The graph below shows the shape of a chi-square distribution:

© IFT. All rights reserved



There are three hypotheses that can be formulated (σ^2 represents the true population variance and σ_0^2 represents the hypothesized variance):

1. $H_0: \sigma^2 = \sigma_0^2$ versus $H_a: \sigma^2 \neq \sigma_0^2$. This is used when we believe the population variance is not equal to 0, or it is different from the hypothesized variance. It is a two-tailed test.
2. $H_0: \sigma^2 \geq \sigma_0^2$ versus $H_a: \sigma^2 < \sigma_0^2$. This is used when we believe the population variance is less than the hypothesized variance. It is a one-tailed test.
3. $H_0: \sigma^2 \leq \sigma_0^2$ versus $H_a: \sigma^2 > \sigma_0^2$. This is used when we believe the population variance is greater than the hypothesized variance. It is a one-tailed test.

After drawing a random sample from a normally distributed population, we calculate the test statistic using the following formula using n - 1 degrees of freedom:

$$\chi^2 = \frac{(n-1)(s^2)}{\sigma_0^2}$$

where:
n = sample size
s = sample variance

We then determine the critical values using the level of significance and degrees of freedom. The chi-square distribution table is used to calculate the critical value.

Example
Consider Fund Alpha which we discussed in an earlier example. This fund has been in existence for 20 months. During this period the standard deviation of monthly returns was 5%. You want to test a claim by the fund manager that the standard deviation of monthly returns is less than 6%.

Solution:
The null and alternate hypotheses are: H_0: $\sigma^2 \geq 36$ versus H_a: $\sigma^2 < 36$

Note that the standard deviation is 6%. Since we are dealing with population variance, we will square this number to arrive at a variance of 36%.

We then calculate the value of the chi-square test statistic:

$\chi^2 = (n - 1) s^2 / \sigma_0^2 = 19 \times 25/36 = 13.19$

Next, we determine the rejection point based on df = 19 and significance = 0.05. Using the chi-square table, we find that this number is 10.117.

Since the test statistic (13.19) is higher than the rejection point (10.117) we cannot reject H_0. In other words, the sample standard deviation is not small enough to validate the fund manager's claim that population standard deviation is less than 6%.

12.2 Test Concerning the Equality of Two Variances (F-Test)

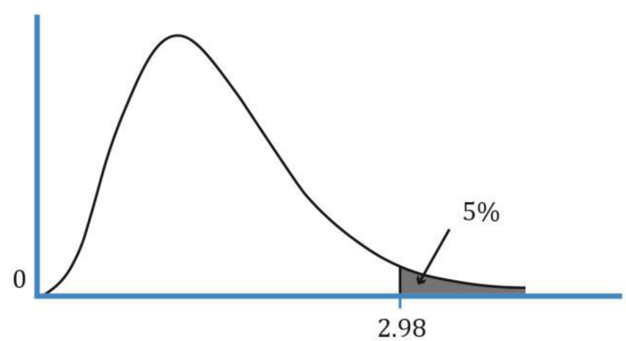
In order to test the equality or inequality of two variances, we use an F-test which is the ratio of sample variances.

The assumptions for a F-test to be valid are:

- The samples must be independent.
- The populations from which the samples are taken are normally distributed.

Properties of the F-distribution

The F-distribution, like the chi-square distribution, is a family of asymmetrical distributions bounded from below by 0. Each F-distribution is defined by two values of degrees of freedom, called the numerator and denominator degrees of freedom. As shown in the figure below, the F-distribution is skewed to the right and is truncated at zero on the left hand side.



The rejection region is always in the right-side tail of the distribution.

When working with F-tests, there are three hypotheses that can be formulated:

1. $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_a: \sigma_1^2 \neq \sigma_2^2$. This is used when we believe the two population variances are not equal.
2. $H_0: \sigma_1^2 \leq \sigma_2^2$ versus $H_a: \sigma_1^2 > \sigma_2^2$. This is used when we believe the variance of the first population is greater than the variance of the second population.
3. $H_0: \sigma_1^2 \geq \sigma_2^2$ versus $H_a: \sigma_1^2 < \sigma_2^2$. This is used when we believe the variance of the first population is less than the variance of the second population.

The term σ_1^2 represents the population variance of the first population and σ_2^2 represents

© IFT. All rights reserved

the population variance of the second population.

The formula for the test statistic of the F-test is:

$$F = \frac{s_1^2}{s_2^2}$$

where:
s_1^2 = the sample variance of the first population with n observations
s_2^2 = the sample variance of the second population with n observations

A convention is to put the larger sample variance in the numerator and the smaller sample variance in the denominator.

$df_1 = n_1 - 1$ numerator degrees of freedom
$df_2 = n_2 - 1$ denominator degrees of freedom

The test statistic is then compared with the critical values found using the two degrees of freedom and the F-tables.

Finally, a decision is made whether to reject or not to reject the null hypothesis.

Example

You are investigating whether the population variance of the Indian equity market changed after the deregulation of 1991. You collect 120 months of data before and after deregulation. Variance of returns before deregulation was 13. Variance of returns after deregulation was 18. Check your hypothesis at a confidence level of 99%.

Solution:

Null and alternate hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$ versus H_A: $\sigma_1^2 \neq \sigma_2^2$

F-statistic: $\frac{18}{13} = 1.4$

df = 119 for the numerator and denominator

α = 0.01 which means 0.005 in each tail. From the F-table: critical value = 1.6

Since the F-stat is less than the critical value, do not reject the null hypothesis.

13. Parametric vs. Nonparametric Tests

The hypothesis-testing procedures we have discussed so far have two characteristics in common:

- They are concerned with parameters, such as the mean and variance.
- Their validity depends on a set of assumptions.

Any procedure which has either of the two characteristics is known as a **parametric test**.

Nonparametric tests are not concerned with a parameter and/or make few assumptions about the population from which the sample are drawn. We use nonparametric procedures in three situations:

© IFT. All rights reserved

- Data does not meet distributional assumptions.
- Data has outliers
- Data are given in ranks. (Example: relative size of the company and use of derivatives.)
- The hypothesis does not concern a parameter. (Example: Is a sample random or not?)

14. Tests Concerning Correlation

The strength of linear relationship between two variables is assessed through correlation coefficient. The significance of a correlation coefficient is tested by using hypothesis tests concerning correlation.

There are two hypotheses that can be formulated (ρ represents the population correlation coefficient):

- $H_0: \rho = 0$
- $H_a: \rho \neq 0$

This test is used when we believe the population correlation is not equal to 0, or it is different from the hypothesized correlation. It is a two-tailed test.

14.1 Parametric Test of a Correlation

As long as the two variables are distributed normally, we can use sample correlation, r for our hypothesis testing. The formula for the t-test is

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where: n – 2 = degrees of freedom if H_0 is true.

The magnitude of r needed to reject the null hypothesis H_0: ρ= 0 decreases as sample size n increases due to the following:

i. As n increases, the number of degrees of freedom increases and the absolute value of the critical value t_c decreases.

ii. As n increases, the absolute value of the numerator increases, leading to larger-magnitude t-values.

In other words, as n increases, the probability of Type-II error decreases, all else equal.

Example

The sample correlation between the oil prices and monthly returns of energy stocks in a Country A is 0.7986 for the period from January 2014 through December 2018. Can we reject a null hypothesis that the underlying or population correlation equals 0 at the 0.05 level of significance?

© IFT. All rights reserved

Solution:

H_0: $\rho = 0$ → true correlation in the population is 0.

H_a: $\rho \neq 0$ → correlation in the population is different from 0.

From January 2014 through December 2018, there are 60 months, so n = 60. We use the following statistic to test the above.

$$t = \frac{0.7986\sqrt{60-2}}{\sqrt{1-0.7986^2}} = \frac{6.0820}{0.6019} = 10.1052$$

At the 0.05 significance level, the critical level for this test statistic is 2.00 (n = 60, degrees of freedom = 58). When the test statistic is either larger than 2.00 or smaller than 2.00, we can reject the hypothesis that the correlation in the population is 0. The test statistic is 10.1052, so we can reject the null hypothesis.

14.2 Tests Concerning Correlation: The Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient is equivalent to the usual correlation coefficient but is calculated on the *ranks* of two variables within their respective samples.

15. Test of Independence Using Contingency Table Data

A chi-square distributed test statistic is used to test for independence of two categorical variables. This nonparametric test compares actual frequencies with those expected on the basis of independence.

The test statistic is calculated as:

$$\chi^2 = \sum_{i=1}^{m} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}},$$

where: $E_{ij} = \frac{(\text{Total row } i) \times (\text{Total column } j)}{\text{Overall total}}.$

This test statistic has degrees of freedom of (r − 1)(c − 2), where r is the number of categories for the first variable and c is the number of categories of the second variable.

© IFT. All rights reserved

Summary

LO.a: Define a hypothesis, describe the steps of hypothesis testing, and describe and interpret the choice of the null and alternative hypotheses.

A hypothesis is a statement about the value of a population parameter developed for the purpose of testing a theory.

In order to test a hypothesis, we follow these steps:

1. State the hypothesis.
2. Identify the appropriate test statistic and its probability distribution.
3. Specify the significance level.
4. State the decision rule.
5. Collect data and calculate the test statistic.
6. Make a decision.

The null hypothesis (H_0) is the hypothesis that the researcher wants to reject. It should always include some form of the 'equal to' condition.

The alternative hypothesis (H_a) is the hypothesis that the researcher wants to prove. If the null hypothesis is rejected then the alternative hypothesis is considered valid.

LO.b: Compare and contrast one-tailed and two-tailed tests of hypotheses.

In one-tailed tests, we are assessing if the value of a population parameter is greater than or less than a hypothesized value.

In two-tailed tests, we are assessing if the value of a population parameter is different from a hypothesized value.

LO.c: Explain a test statistic, Type I and Type II errors, a significance level, how significance levels are used in hypothesis testing, and the power of a test.

A test statistic is a quantity, calculated on the basis of a sample, and is used to decide whether to reject or not to reject the null hypothesis. The formula for computing the test statistic is:

$$\text{test statistic} = \frac{\text{sample statistic} - \text{value of the parameter under } H_0}{\text{standard error of the sample statistic}}$$

In reaching a statistical decision, we can make two possible errors: We may reject a true null hypothesis (a Type I error), or we may fail to reject a false null hypothesis (a Type II error).

The level of significance of a test is the probability of a Type I error. As α gets smaller the critical value gets larger and it becomes more difficult to reject the null hypothesis.

The power of a test is the probability of correctly rejecting the null (rejecting the null when it is false). It is expressed as:

Power of a test = 1 – P (Type II error)

© IFT. All rights reserved

LO.d: Explain a decision rule and the relation between confidence intervals and hypothesis tests, and determine whether a statistically significant result is also economically meaningful.

A decision rule consists of comparing the computed test statistic to the critical values (rejection points) based on the level of significance to decide whether to reject or not to reject the null hypothesis.

A confidence interval gives us the range of values within which a population parameter is expected to lie. Confidence intervals and hypothesis tests are linked through critical values. The null hypothesis will be rejected only if the test statistic lies outside the confidence interval.

The statistical decision consists of rejecting or not rejecting the null hypothesis. The economic decision takes into consideration all economic issues relevant to the decision.

LO.e: Explain and interpret the p-value as it relates to hypothesis testing.

The p-value is the smallest level of significance at which the null hypothesis can be rejected. It can be used in the hypothesis testing framework as an alternative to using rejection points.

- If the p-value is lower than our specified level of significance, we reject the null hypothesis.
- If the p-value is greater than our specified level of significance, we do not reject the null hypothesis.

LO.f: Describe how to interpret the significance of a test in the context of multiple tests.

The false discovery approach is used to adjust the p-values when you run multiple tests. The researcher first ranks the p-values from the various tests from lowest to highest. He then makes the following comparison, starting with the lowest p-value (with k = 1), p(1):

$$p(1) \leq \alpha \frac{\text{Rank of i}}{\text{Number of tests}}$$

This comparison is repeated until we find the highest ranked p(k) for which this condition holds. If, say, k is 4, then the first four tests (ranked on the basis of the lowest p-values) are said to be significant.

LO.g: Identify the appropriate test statistic and interpret the results for a hypothesis test concerning the population mean of both large and small samples when the population is normally or approximately normally distributed and the variance is (1) known or (2) unknown.

We use the following table to decide which test statistic and which corresponding probability distribution to use for hypothesis testing.

© IFT. All rights reserved

Sampling from		Small sample size (n<30)	Large sample size (n≥30)
Normal distribution	Variance known	z	z
	Variance unknown	t	t (or z)
Non –normal distribution	Variance known	NA	z
	Variance unknown	NA	t (or z)

LO.h: Identify the appropriate test statistic and interpret the results for a hypothesis test concerning the equality of the population means of two at least approximately normally distributed populations based on independent random samples with equal assumed variances.

When we can assume that the two populations are normally distributed and that the unknown population variances are equal, the t-test based on independent random samples is given by:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{(\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2})^{1/2}}$$

The number of degrees of freedom is $n_1 + n_2 - 2$. The term s_p^2 is known as the pooled estimator of the common variance. A pooled estimate is an estimate drawn from the combination of two different samples. It is calculated as:

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

LO.i: Identify the appropriate test statistic and interpret the results for a hypothesis test concerning the mean difference of two normally distributed populations.

In cases where we have a test concerning the mean difference of two normally distributed populations that are dependent, we conduct a t-test that is based on data arranged in paired observations.

The hypothesis is formed on the difference between means of two populations e.g. H_0: $\mu_d = \mu_{d0}$ versus H_a: $\mu_d \neq \mu_{d0}$

In order to arrive at the test statistic, we first determine the sample mean difference using:

$$\bar{d} = \frac{1}{n}\sum_{i=0}^{n} d_i$$

And the standard error of the mean difference is computed as follows:

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$$

Once we have these two values, we can calculate the test statistic using a t-test. This is calculated using the following formula using n - 1 degrees of freedom:

© IFT. All rights reserved

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

The value of calculated test statistic is compared with the t-distribution values in the usual manner to arrive at a decision on our hypothesis.

LO.j: Identify the appropriate test statistic and interpret the results for a hypothesis test concerning (1) the variance of a normally distributed population and (2) the equality of the variances of two normally distributed populations based on two independent random samples.

In tests concerning the variance of a single normally distributed population, we use the chi-square test statistic, denoted by χ^2. After drawing a random sample from a normally distributed population, we calculate the test statistic using the following formula using n - 1 degrees of freedom:

$$\chi^2 = \frac{(n-1)(s^2)}{\sigma_0^2}$$

We then determine the critical values using the level of significance and degrees of freedom. The chi-square distribution table is used to calculate the critical value.

In order to test the equality or inequality of two variances, we use an F-test. The critical value is computed as:

$$F = \frac{s_1^2}{s_2^2}$$

The test statistic is then compared with the critical values found using the two degrees of freedom and the F-tables. Finally, a decision is made whether to reject or not to reject the null hypothesis.

LO.k: Compare and contrast parametric and nonparametric tests, and describe situations where each is the more appropriate type of test.

A parametric test is a hypothesis test concerning a parameter or a hypothesis test based on specific distributional assumptions. In contrast, a nonparametric test is either not concerned with a parameter or makes minimal assumptions about the population from which the sample is drawn.

A nonparametric test is primarily used in three situations: when data do not meet distributional assumptions, when data is given in ranks, or when the hypothesis we are addressing does not concern a parameter.

LO.l: Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

© IFT. All rights reserved

Parametric test: The significance of a correlation coefficient is tested by using hypothesis tests concerning correlation. The formula for the t-test is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where: n – 2 = degrees of freedom

Non-parametric test: The Spearman rank correlation coefficient is equivalent to the usual correlation coefficient but is calculated on the *ranks* of two variables within their respective samples.

LO.m: Explain tests of independence based on contingency table data.

A chi-square distributed test statistic is used to test for independence of two categorical variables. This nonparametric test compares actual frequencies with those expected on the basis of independence.

This test statistic has degrees of freedom of (r − 1)(c – 2), where r is the number of categories for the first variable and c is the number of categories of the second variable.

© IFT. All rights reserved