Hypothesis Testing

By Dias Kulzhanov

Steps in Hypothesis Testing

- The steps in testing a hypothesis are as follows:

  1. Stating the hypotheses.
  2. Identifying the appropriate test statistic and its probability distribution.
  3. Specifying the significance level.
  4. Stating the decision rule.
  5. Collecting the data and calculating the test statistic.
  6. Making the statistical decision.
  7. Making the economic or investment decision.

1^{st} step: Stating the hypotheses

The first step in hypothesis testing is stating the hypotheses. We always state two hypotheses: the null hypothesis (or null), designated H_0, and the alternative hypothesis, designated H_a.

- **Definition of Null Hypothesis.** The null hypothesis is the hypothesis to be tested. For example, we could hypothesize that the population mean risk premium for Canadian equities is less than or equal to zero.

The null hypothesis is a proposition that is considered true unless the sample we use to conduct the hypothesis test gives convincing evidence that the null hypothesis is false. When such evidence is present, we are led to the alternative hypothesis.

- **Definition of Alternative Hypothesis.** The alternative hypothesis is the hypothesis accepted when the null hypothesis is rejected. Our alternative hypothesis is that the population mean risk premium for Canadian equities is greater than zero.

- **Formulations of Hypotheses.** We can formulate the null and alternative hypotheses in three different ways:

  1. H_0: $\theta = \theta_0$ versus H_a: $\theta \neq \theta_0$ (a "not equal to" alternative hypothesis)
  2. H_0: $\theta \leq \theta_0$ versus H_a: $\theta > \theta_0$ (a "greater than" alternative hypothesis)
  3. H_0: $\theta \geq \theta_0$ versus H_a: $\theta < \theta_0$ (a "less than" alternative hypothesis)

The first formulation is a **two-sided hypothesis test** (or **two-tailed hypothesis test**): We reject the null in favor of the alternative if the evidence indicates that the population parameter is either smaller or larger than θ_0. In contrast, Formulations 2 and 3 are each a **one-sided hypothesis test** (or **one-tailed hypothesis test**). For Formulations 2 and 3, we reject the null only if the evidence indicates that the population parameter is respectively greater than or less than θ_0. The alternative hypothesis has one side.

- When we have a "suspected" or "hoped for" condition for which we want to find supportive evidence, we frequently set up that condition as the alternative hypothesis and use a one-sided test. To emphasize a neutral attitude, however, the researcher may select a "not equal to" alternative hypothesis and conduct a two-sided test.

2nd step: Identifying the appropriate test statistic and its probability distribution

- **Definition of Test Statistic.** A test statistic is a quantity, calculated based on a sample, whose value is the basis for deciding whether or not to reject the null hypothesis.

The focal point of our statistical decision is the value of the test statistic. Frequently (in all the cases that we examine in this reading), the test statistic has the form

$$\text{Test statistic} = \frac{\text{Sample statistic} - \text{Value of the population parameter under } H_0}{\text{Standard error of the sample statistic}} \quad \textbf{(1)}$$

3rd: Specifying the significance level

- The **level of significance** reflects how much sample evidence we require to reject the null. Analogous to its counterpart in a court of law, the required standard of proof can change according to the nature of the hypotheses and the seriousness of the consequences of making a mistake. There are four possible outcomes when we test a null hypothesis:

	True Situation	
Decision	**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** in testing a hypothesis is denoted by the Greek letter alpha, α. This probability is also known as the **level of significance of the test.**

4^{th}: Stating the decision rule

- A **decision rule** consists of determining the rejection points (critical values) with which to compare the test statistic to decide whether to reject or not to reject the null hypothesis. When we reject the null hypothesis, the result is said to be **statistically significant**.

- **Definition of a Rejection Point (Critical Value) for the Test Statistic.** A rejection point (critical value) for a test statistic is a value with which the computed test statistic is compared to decide whether to reject or not reject the null hypothesis.

For a one-tailed test, we indicate a rejection point using the symbol for the test statistic with a subscript indicating the specified probability of a Type I error, α; for example, z_α. For a two-tailed test, we indicate $z_{\alpha/2}$. To illustrate the use of rejection points, suppose we are using a z-test and have chosen a 0.05 level of significance.

5^{th}: Collecting the data and calculating the test statistic

- Collecting the data by sampling the population

6^{th}: Making the statistical decision

- To reject or not

7^{th}: Making the economic or investment decision

- The first six steps are the statistical steps. The final decision concerns our use of the statistical decision.
- The economic or investment decision takes into consideration not only the statistical decision but also all pertinent economic issues.

p-Value

- The **p-value** is the smallest level of significance at which the null hypothesis can be rejected. The smaller the p-value, the stronger the evidence against the null hypothesis and in favor of the alternative hypothesis. The p-value approach to hypothesis testing does not involve setting a significance level; rather it involves computing a p-value for the test statistic and allowing the consumer of the research to interpret its significance.

Tests Concerning a Single Mean

- For hypothesis tests concerning the population mean of a normally distributed population with unknown (known) variance, the theoretically correct test statistic is the t-statistic (z-statistic). In the unknown variance case, given large samples (generally, samples of 30 or more observations), the z-statistic may be used in place of the t-statistic because of the force of the central limit theorem.
- The t-distribution is a symmetrical distribution defined by a single parameter: degrees of freedom. Compared to the standard normal distribution, the t-distribution has fatter tails.

- **Test Statistic for Hypothesis Tests of the Population Mean (Practical Case—Population Variance Unknown).** If the population sampled has unknown variance and either of the conditions below holds:

  1. the sample is large, or
  2. the sample is small but the population sampled is normally distributed, or approximately normally distributed,

  then the test statistic for hypothesis tests concerning a single population mean, μ, is

$$t_{n-1} = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \tag{4}$$

  where

  t_{n-1} = t-statistic with $n - 1$ degrees of freedom (n is the sample size)

  $\overline{X}$ = the sample mean

  μ_0 = the hypothesized value of the population mean

  s = the sample standard deviation

The denominator of the t-statistic is an estimate of the sample mean standard error, $s_{\overline{X}} = s/\sqrt{n}$.[20]

The *z-Test Alternative*

- **The *z*-Test Alternative.**

1. If the population sampled is normally distributed with known variance σ^2, then the test statistic for a hypothesis test concerning a single population mean, μ, is

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \tag{5}$$

2. If the population sampled has unknown variance and the sample is large, in place of a *t*-test, an alternative test statistic (relying on the central limit theorem) is

$$z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \tag{6}$$

In the above equations,

σ = the known population standard deviation

s = the sample standard deviation

μ_0 = the hypothesized value of the population mean

- **Rejection Points for a z-Test.**

A Significance level of $\alpha = 0.10$.

1 H_0: $\theta = \theta_0$ versus H_a: $\theta \neq \theta_0$. The rejection points are $z_{0.05} = 1.645$ and $-z_{0.05} = -1.645$.

Reject the null hypothesis if $z > 1.645$ or if $z < -1.645$.

2 H_0: $\theta \leq \theta_0$ versus H_a: $\theta > \theta_0$. The rejection point is $z_{0.10} = 1.28$.

Reject the null hypothesis if $z > 1.28$.

3 H_0: $\theta \geq \theta_0$ versus H_a: $\theta < \theta_0$. The rejection point is $-z_{0.10} = -1.28$.

Reject the null hypothesis if $z < -1.28$.

B Significance level of $\alpha = 0.05$.

1 H_0: $\theta = \theta_0$ versus H_a: $\theta \neq \theta_0$. The rejection points are $z_{0.025} = 1.96$ and $-z_{0.025} = -1.96$.

Reject the null hypothesis if $z > 1.96$ or if $z < -1.96$.

2 H_0: $\theta \leq \theta_0$ versus H_a: $\theta > \theta_0$. The rejection point is $z_{0.05} = 1.645$.

Reject the null hypothesis if $z > 1.645$.

3 H_0: $\theta \geq \theta_0$ versus H_a: $\theta < \theta_0$. The rejection point is $-z_{0.05} = -1.645$.

Reject the null hypothesis if $z < -1.645$.

C Significance level of $\alpha = 0.01$.

1 H_0: $\theta = \theta_0$ versus H_a: $\theta \neq \theta_0$. The rejection points are $z_{0.005} = 2.575$ and $-z_{0.005} = -2.575$.

Reject the null hypothesis if $z > 2.575$ or if $z < -2.575$.

2 H_0: $\theta \leq \theta_0$ versus H_a: $\theta > \theta_0$. The rejection point is $z_{0.01} = 2.33$.

Reject the null hypothesis if $z > 2.33$.

3 H_0: $\theta \geq \theta_0$ versus H_a: $\theta < \theta_0$. The rejection point is $-z_{0.01} = -2.33$.

Reject the null hypothesis if $z < -2.33$.

Tests Concerning Differences between Means

- When we want to test whether the observed **difference between two means** is statistically significant, we must first decide whether the samples are **independent or dependent** (related). If the samples are independent, we conduct tests concerning **differences between means**. If the samples are dependent, we conduct tests of **mean differences (paired comparisons tests)**.
- When we conduct a test of the difference between two population means from normally distributed populations with **unknown variances, if we can assume the variances are equal**, we use a **t-test** based on pooling the observations of the two samples to estimate the common (but unknown) variance. This test is based on an assumption of independent samples.

Thus we usually formulate the following hypotheses:

1 H_0: $\mu_1 - \mu_2 = 0$ versus H_a: $\mu_1 - \mu_2 \neq 0$ (the alternative is that $\mu_1 \neq \mu_2$)

2 H_0: $\mu_1 - \mu_2 \leq 0$ versus H_a: $\mu_1 - \mu_2 > 0$ (the alternative is that $\mu_1 > \mu_2$)

3 H_0: $\mu_1 - \mu_2 \geq 0$ versus H_a: $\mu_1 - \mu_2 < 0$ (the alternative is that $\mu_1 < \mu_2$)

- **Test Statistic for a Test of the Difference between Two Population Means (Normally Distributed Populations, Population Variances Unknown but Assumed Equal).** When we can assume that the two populations are normally distributed and that the unknown population variances are equal, a t-test based on independent random samples is given by

$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{\left(\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}\right)^{1/2}} \tag{7}$$

where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ is a pooled estimator of the common variance.

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

- **Test Statistic for a Test of the Difference between Two Population Means (Normally Distributed Populations, Unequal and Unknown Population Variances).** When we can assume that the two populations are normally distributed but do not know the population variances and cannot assume that they are equal, an approximate t-test based on independent random samples is given by

$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^{1/2}} \tag{8}$$

where we use tables of the t-distribution using “modified” degrees of freedom computed with the formula

$$\mathrm{df} = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2}} \tag{9}$$

Tests Concerning Mean Differences

- In tests concerning two means based on two samples that **are not independent**, we often can arrange the data in paired observations and conduct a test of mean differences **(a paired comparisons test**). When the samples are from normally distributed populations with unknown variances, the appropriate test statistic is a t-statistic. The denominator of the t-statistic, the standard error of the mean differences, takes account of correlation between the samples.

Letting A represent "after" and B "before," suppose we have observations for the random variables X_A and X_B and that the samples are dependent. We arrange the observations in pairs. Let d_i denote the difference between two paired observations. We can use the notation $d_i = x_{Ai} - x_{Bi}$, where x_{Ai} and x_{Bi} are the ith pair of observations, $i = 1, 2, \ldots, n$ on the two variables. Let μ_d stand for the population mean difference. We can formulate the following hypotheses, where μ_{d0} is a hypothesized value for the population mean difference:

1. $H_0: \mu_d = \mu_{d0}$ versus $H_a: \mu_d \neq \mu_{d0}$
2. $H_0: \mu_d \leq \mu_{d0}$ versus $H_a: \mu_d > \mu_{d0}$
3. $H_0: \mu_d \geq \mu_{d0}$ versus $H_a: \mu_d < \mu_{d0}$

As usual, we are concerned with the case of normally distributed populations with unknown population variances, and we will formulate a t-test. To calculate the t-statistic, we first need to find the sample mean difference:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad (10)$$

where n is the number of pairs of observations. The sample variance, denoted by s_d^2, is

$$s_d^2 = \frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n-1} \qquad (11)$$

Taking the square root of this quantity, we have the sample standard deviation, s_d, which then allows us to calculate the standard error of the mean difference as follows:[24]

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} \tag{12}$$

- **Test Statistic for a Test of Mean Differences (Normally Distributed Populations, Unknown Population Variances).** When we have data consisting of paired observations from samples generated by normally distributed populations with unknown variances, a t-test is based on

$$t = \frac{\bar{d} - \mu_{d0}}{s_{\bar{d}}} \tag{13}$$

with $n - 1$ degrees of freedom, where n is the number of paired observations, $\bar{d}$ is the sample mean difference (as given by Equation 10), and $s_{\bar{d}}$ is the standard error of $\bar{d}$ (as given by Equation 12).

Tests Concerning a Single Variance

- In tests concerning the variance of a single, normally distributed population, the test statistic is **chi-square (χ2) with n − 1 degrees of freedom**, where n is sample size.

1 $H_0: \sigma^2 = \sigma_0^2$ versus $H_a: \sigma^2 \neq \sigma_0^2$ (a "not equal to" alternative hypothesis)

2 $H_0: \sigma^2 \leq \sigma_0^2$ versus $H_a: \sigma^2 > \sigma_0^2$ (a "greater than" alternative hypothesis)

3 $H_0: \sigma^2 \geq \sigma_0^2$ versus $H_a: \sigma^2 < \sigma_0^2$ (a "less than" alternative hypothesis)

■ **Test Statistic for Tests Concerning the Value of a Population Variance (Normal Population).** If we have n independent observations from a normally distributed population, the appropriate test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \tag{14}$$

with $n - 1$ degrees of freedom. In the numerator of the expression is the sample variance, calculated as

$$s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} \tag{15}$$

In contrast to the t-test, for example, the chi-square test is sensitive to violations of its assumptions. If the sample is not actually random or if it does not come from a normally distributed population, inferences based on a chi-square test are likely to be faulty.

- **Rejection Points for Hypothesis Tests on the Population Variance.**

1. "Not equal to" H_a: Reject the null hypothesis if the test statistic is greater than the upper $\alpha/2$ point (denoted $\chi^2_{\alpha/2}$) or less than the lower $\alpha/2$ point (denoted $\chi^2_{1-\alpha/2}$) of the chi-square distribution with df = $n - 1$.[25]
2. "Greater than" H_a: Reject the null hypothesis if the test statistic is greater than the upper α point of the chi-square distribution with df = $n - 1$.
3. "Less than" H_a: Reject the null hypothesis if the test statistic is less than the lower α point of the chi-square distribution with df = $n - 1$.

Tests Concerning the Equality (Inequality) of Two Variances

- For tests concerning differences between the variances of two normally distributed populations based on two random, independent samples, the appropriate test statistic **is based on an F-test (the ratio of the sample variances**).

1. $H_0\colon \sigma_1^2 = \sigma_2^2$ versus $H_a\colon \sigma_1^2 \neq \sigma_2^2$
2. $H_0\colon \sigma_1^2 \leq \sigma_2^2$ versus $H_a\colon \sigma_1^2 > \sigma_2^2$
3. $H_0\colon \sigma_1^2 \geq \sigma_2^2$ versus $H_a\colon \sigma_1^2 < \sigma_2^2$

- **Test Statistic for Tests Concerning Differences between the Variances of Two Populations (Normally Distributed Populations).** Suppose we have two samples, the first with n_1 observations and sample variance s_1^2, the second with n_2 observations and sample variance s_2^2. The samples are random, independent of each other, and generated by normally distributed populations. A test concerning differences between the variances of the two populations is based on the ratio of sample variances

$$F = \frac{s_1^2}{s_2^2} \qquad (16)$$

with $\text{df}_1 = n_1 - 1$ numerator degrees of freedom and $\text{df}_2 = n_2 - 1$ denominator degrees of freedom. Note that df_1 and df_2 are the divisors used in calculating s_1^2 and s_2^2, respectively.

A convention, or usual practice, is to use the larger of the two ratios s_1^2/s_2^2 or s_2^2/s_1^2 as the actual test statistic. When we follow this convention, the value of the test statistic is always greater than or equal to 1; tables of critical values of F then need include only values greater than or equal to 1. Under this convention, the rejection point for any formulation of hypotheses is a single value in the right-hand side of the relevant F-distribution. Note that the labeling of populations as "1" or "2" is arbitrary in any case.

- **Rejection Points for Hypothesis Tests on the Relative Values of Two Population Variances.** Follow the convention of using the larger of the two ratios s_1^2 / s_2^2 and s_2^2 / s_1^2 and consider two cases:

  1. A "not equal to" alternative hypothesis: Reject the null hypothesis at the α significance level if the test statistic is greater than the upper α/2 point of the *F*-distribution with the specified numerator and denominator degrees of freedom.

  2. A "greater than" or "less than" alternative hypothesis: Reject the null hypothesis at the α significance level if the test statistic is greater than the upper α point of the *F*-distribution with the specified number of numerator and denominator degrees of freedom.

NONPARAMETRIC INFERENCE

- A parametric test is a hypothesis test concerning a parameter or a hypothesis test based on specific distributional assumptions. In contrast, a nonparametric test either is not concerned with a parameter or makes minimal assumptions about the population from which the sample comes.
- A nonparametric test is primarily used in three situations: **when data do not meet distributional assumptions, when data are given in ranks, or when the hypothesis we are addressing does not concern a parameter**.

	Parametric	Nonparametric
Tests concerning a single mean	*t*-test *z*-test	Wilcoxon signed-rank test
Tests concerning differences between means	*t*-test Approximate *t*-test	Mann–Whitney U test
Tests concerning mean differences (paired comparisons tests)	*t*-test	Wilcoxon signed-rank test Sign test

Tests Concerning Correlation: The Spearman Rank Correlation Coefficient

- The Spearman rank correlation coefficient is essentially equivalent to the usual correlation coefficient **calculated on the ranks of the two variables** (say X and Y) within their respective samples. Thus it is a number between −1 and +1, where −1 (+1) denotes a perfect inverse (positive) straight-line relationship between the variables and 0 represents the absence of any straight-line relationship (no correlation). The calculation of rS requires the following steps:

1 Rank the observations on X from largest to smallest. Assign the number 1 to the observation with the largest value, the number 2 to the observation with second-largest value, and so on. In case of ties, we assign to each tied observation the average of the ranks that they jointly occupy. For example, if the third- and fourth-largest values are tied, we assign both observations the rank of 3.5 (the average of 3 and 4). Perform the same procedure for the observations on Y.

2 Calculate the difference, d_i, between the ranks of each pair of observations on X and Y.

3 Then, with n the sample size, the Spearman rank correlation is given by[33]

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)} \tag{17}$$

For large samples (say $n > 30$), we can conduct a t-test using

$$t = \frac{(n-2)^{1/2} r_s}{\left(1 - r_s^2\right)^{1/2}} \tag{18}$$

based on $n - 2$ degrees of freedom.