Probability and Statistics course

Introduction

Gabor Vigh

TTK Department of Probability Theory and Statistics
ELTE

November 23, 2025

Week 11

Hypothesis testing

A vital role of Statistics is in verifying statements, claims, conjectures, and in general - testing hypotheses. Based on a random sample, we can use Statistics to verify whether

1. a system has not been infected,
2. a hardware upgrade was efficient,
3. the average number of concurrent users increased by 2000 this year,
4. the average connection speed is 54 Mbps, as claimed by the internet service provider,

Testing statistical hypotheses has wide applications far beyond Computer Science. These methods are used to prove efficiency of a new medical treatment, safety of a new automobile brand, innocence of a defendant, and authorship of a document **Notation:**

- $H_0 =$ hypothesis (the null hypothesis)
- $H_A =$ alternative (the alternative hypothesis)

H_0 and H_A are simply two mutually exclusive statements. Each test results either in acceptance of H_0 or its rejection in favor of H_A.
A null hypothesis (H_0) is common belief. In order to overturn the common belief and to reject the hypothesis, we need **significant evidence**. Such evidence can only be provided by data. Only when such evidence is found, and when it strongly supports the alternative H_A, can the hypothesis H_0 be rejected in favor of H_A.
Based on a random sample, a statistician cannot tell whether the hypothesis is true or the alternative. We need to see the entire population to tell that. The purpose of each test is to determine whether the data provides sufficient evidence against H_0 in favor of H_A.

Hypothesis testing

Definition

Alternative of the type $H_A : \mu \neq \mu_0$ covering regions on both sides of the hypothesis ($H_0 : \mu = \mu_0$) is a **two-sided alternative**. Alternative $H_A : \mu < \mu_0$ covering the region to the left of H_0 is **one-sided, left-tail**. Alternative $H_A : \mu > \mu_0$ covering the region to the right of H_0 is **one-sided, right-tail**.

To verify that the average connection speed is 54 Mbps, we test the hypothesis $H_0 : \mu = 54$ against the two-sided alternative $H_A : \mu \neq 54$, where μ is the average speed of all connections. However, if we worry about a low

connection speed only, we can conduct a one-sided test of

$$H_0 : \mu = 54 \quad \text{vs} \quad H_A : \mu < 54.$$

In this case, we only measure the amount of evidence supporting the one-sided alternative $H_A : \mu < 54$. In the absence of such evidence, we gladly accept the null hypothesis.

Type I and II errors

When testing hypotheses, we realize that all we see is a random sample. Therefore, with all the best statistics skills, our decision to accept or to reject H_0 may still be wrong. Four situations are possible:

Result of the test	**Reject $\mathbf{H_0}$**	**Accept $\mathbf{H_0}$**
$\mathbf{H_0}$ is true	Type I error	correct
$\mathbf{H_0}$ is false	correct	Type II error

Each error occurs with a certain probability that we hope to keep small. A type I error is often considered more dangerous and undesired than a type II error. Making a type I error can be compared with sending a patient to a surgery when (s)he does not need one. For this reason, we shall design tests that bound the probability of type I error by a preassigned small number α. Under this condition, we may want to minimize the probability of type II error.

The **power of a statistical test** is the probability of correctly rejecting the null hypothesis (H_0) when the alternative hypothesis (H_a) is true. It represents the test's ability to detect an effect if one truly exists.

$$\text{Power} = P(\text{Reject } H_0 \mid H_a \text{ is true}) = 1 - \beta$$

Where:

- β is the probability of a **Type II Error** (failing to reject a false H_0).

Level α tests: general approach

1. Testing hypothesis is based on a **test statistic** T, a quantity computed from the data that has some known, tabulated distribution F_0 if the hypothesis H_0 is true.
2. **Acceptance region and rejection region**: Next, we consider the **null distribution** F_0. This is the distribution of the test statistic T when the hypothesis H_0 is true. If it has a density f_0, then the whole area under the density curve is 1. We can always find a portion of this area, α, which is called the **rejection region** (R), as shown below.

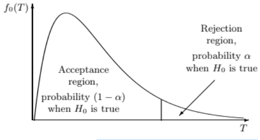


Figure: Acceptance and rejection regions

3. **Result and its interpretation** Accept the hypothesis H_0 if the test statistic T belongs to the acceptance region. Reject H_0 in favor of the alternative H_A if T belongs to the rejection region. Our acceptance and rejection regions guarantee that the significance level of our test is

$$\begin{aligned}\text{Significance level} &= P\{\text{Type I error}\}\\ &= P\{\text{Reject } H_0 \mid H_0 \text{ is true}\}\\ &= P\{T \in R \mid H_0\}\\ &= \alpha.\end{aligned}$$

Level α tests: general approach

If the test **rejects** the hypothesis, all we can state is that the data provides sufficient evidence against H_0 and in favor of H_A. It may either happen because H_0 is not true, or because our sample is too extreme. The latter, however, can only happen with probability α.

If the test **accepts** the hypothesis (i.e., fails to reject H_0), it only means that the evidence obtained from the data is not sufficient to reject it. In the absence of sufficient evidence, by default, we accept the null hypothesis.

Rejection regions and power

Our construction of the rejection region guaranteed the desired significance level α, as we proved in (9.13).
However, one can choose many regions that will also have probability α (see Figure 9.7). Among them, which one is the best choice?
To avoid Type II errors, we choose such a rejection region that will likely cover the test statistic T in case the alternative H_A is true. This **maximizes the power** of our test because we'll rarely accept H_0 in this case.

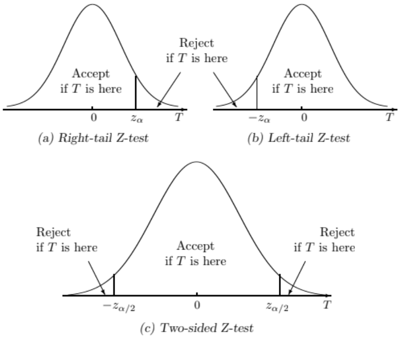


Figure: Acceptance and rejection regions for a Z-test

Standard Normal null distribution (Z-test)

a. A level α test with a **right-tail alternative** should
   - reject H_0 if $Z \geq z_\alpha$
   - accept H_0 if $Z < z_\alpha$

b. With a **left-tail alternative** (e.g., $H_A : \mu < \mu_0$), we should
   - reject H_0 if $Z \leq -z_\alpha$
   - accept H_0 if $Z > -z_\alpha$

c. With a **two-sided alternative** (e.g., $H_A : \mu \neq \mu_0$), we
   - reject H_0 if $|Z| \geq z_{\alpha/2}$
   - accept H_0 if $|Z| < z_{\alpha/2}$

Now consider testing a hypothesis about a population parameter θ. Suppose that its estimator $\hat{\theta}$ has Normal distribution, at least approximately, and we know $E(\hat{\theta})$ and $\text{Var}(\hat{\theta})$ if the hypothesis is true.
Then the Z test statistic is

$$Z = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{\text{Var}(\hat{\theta})}}$$

Z-tests for means and proportions

As we already know,

- sample means have **Normal distribution** when the distribution of data is Normal;
- sample means have **approximately Normal distribution** when they are computed from large samples (the distribution of data can be arbitrary, due to the Central Limit Theorem);
- sample proportions have **approximately Normal distribution** when they are computed from large samples;
- this extends to **differences** between means and between proportions.

Summary of Z tests

Null Hypothesis H_0	Parameter (θ), Estimator ($\hat{\theta}$)	If H_0 is true: $E(\hat{\theta})$	If H_0 is true: $\mathrm{Var}(\hat{\theta})$	Test Statistic Z
One-Sample Z-Tests (Sample Size n)				
$\mu = \mu_0$	μ, $\bar{X}$	μ_0	$\frac{\sigma^2}{n}$	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
$p = p_0$	p, $\hat{p}$	p_0	$\frac{p_0(1-p_0)}{n}$	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Two-Sample Z-Tests (Independent Samples n and m)				
$\mu_X - \mu_Y = D$	$\mu_X - \mu_Y$, $\bar{X} - \bar{Y}$	D	$\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$	$Z = \frac{(\bar{X}-\bar{Y})-D}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}$
$p_1 - p_2 = D$	$p_1 - p_2$, $\hat{p}_1 - \hat{p}_2$	D	$\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}$	$Z = \frac{(\hat{p}_1-\hat{p}_2)-D}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}}}$
$p_1 = p_2$	$p_1 - p_2$, $\hat{p}_1 - \hat{p}_2$	0	$p(1-p)\left(\frac{1}{n}+\frac{1}{m}\right)$ (where $p = p_1 = p_2$)	$Z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n}+\frac{1}{m}\right)}}$ (where $\hat{p} = \frac{n\hat{p}_1+m\hat{p}_2}{n+m}$)

Examples

Example

The number of concurrent users for some internet service provider has always averaged $\mu_0 = 5000$ with a standard deviation of $\sigma = 800$. After an equipment upgrade, the average number of users at $n = 100$ randomly selected moments of time is $\bar{x} = 5200$. Does it indicate, at a 5% level of significance ($\alpha = 0.05$), that the mean number of concurrent users has increased? Assume that the standard deviation of the number of concurrent users has not changed.

Example

A quality inspector finds $x_A = 10$ defective parts in a sample of $n_A = 500$ parts received from manufacturer A. Out of $n_B = 400$ parts from manufacturer B, she finds $x_B = 12$ defective ones. A computer-making company uses these parts in their computers and claims that the quality of parts produced by A and B is the same. At the 5% level of significance ($\alpha = 0.05$), do we have enough evidence to disprove this claim?

Unknown σ: T-tests

The resulting T-statistic has the form

$$t = \frac{\hat{\theta} - E(\hat{\theta})}{s(\hat{\theta})} = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{d\,\text{Var}(\hat{\theta})}}$$

Example (From last week)

If an unauthorized person accesses a computer account with the correct username and password (stolen or cracked), can this intrusion be detected? Recently, a number of methods have been proposed to detect such unauthorized use. The time between keystrokes, the time a key is depressed, and the frequency of various keywords are measured and compared with those of the account owner. If there are significant differences, an intruder is detected. The following times between keystrokes (in seconds) were recorded when a user typed the username and password:

$$0.24, 0.22, 0.26, 0.34, 0.35, 0.32, 0.33, 0.29, 0.19, 0.36, 0.30, 0.15, 0.17, 0.28, 0.38, 0.40, 0.37, 0.27$$

As the first step in detecting an intrusion, let's construct a 99% confidence interval for the mean time between keystrokes, assuming a Normal distribution of these times.

Example

A long-time authorized user of the account makes $\mu_0 = 0.2$ seconds between keystrokes. One day, the data above are recorded as someone typed the correct username and password. At a 5% level of significance ($\alpha = 0.05$), is this an evidence of an unauthorized attempt?

Please check Baron's book for other examples to extend the last weeks examples.

T-tests

Hypothesis H_0	**Conditions**	**Test Statistic** t	**Degrees of Freedom** (ν)
One-Sample T-Test			
$\mu = \mu_0$	Sample size n; unknown σ	$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$	$n - 1$
Two-Sample T-Tests (Independent Samples)			
$\mu_X - \mu_Y = D$	Sample sizes n, m; unknown but equal standard deviations, $\sigma_X = \sigma_Y$ (Pooled Variance)	$t = \frac{\bar{X} - \bar{Y} - D}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$	$n + m - 2$
$\mu_X - \mu_Y = D$	Sample sizes n, m; unknown, unequal standard deviations, $\sigma_X \neq \sigma_Y$ (Non-Pooled Variance)	$t = \frac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$	Satterthwaite approximation, formula (9.12)

Connection between confidence intervals and two-sided tests

An interesting fact can be discovered if we look into our derivation of tests and confidence intervals. It turns out that we can conduct two-sided tests using nothing but the confidence intervals!
A level α Z-test of $H_0 : \theta = \theta_0$ vs $H_A : \theta \neq \theta_0$ accepts the null hypothesis if and only if a symmetric $(1 - \alpha)100\%$ confidence Z-interval for θ contains θ_0.

Example

Election example from last, we computed a 95% confidence interval for the difference of proportions supporting a candidate in towns A and B: $[-0.14, 0.16]$. This interval contains 0, therefore, the test of

$$H_0 : p_1 = p_2 \quad \text{vs} \quad H_A : p_1 \neq p_2$$

accepts the null hypothesis at the 5% level. Apparently, there is no evidence of unequal support of this candidate in the two towns.

A level α T-test of $H_0 : \theta = \theta_0$ vs $H_A : \theta \neq \theta_0$ accepts the null hypothesis if and only if a symmetric $(1 - \alpha)100\%$ confidence T-interval for θ contains θ_0.

P-value

Suppose that the result of our test is crucially important. For example, the choice of a business strategy for the next ten years depends on it. In this case, can we rely so heavily on the choice of α? And if we rejected the true hypothesis just because we chose $\alpha = 0.05$ instead of $\alpha = 0.04$, then how do we explain to the chief executive officer that the situation was marginal? The statistical term for “too close to call” is the **p-value**.

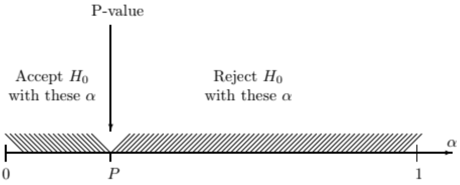


Figure: P-value

Definition

The **P-value** is the lowest significance level α that forces rejection of the null hypothesis (H_0). P-value is also the highest significance level α that forces acceptance (failure to reject) of the null hypothesis (H_0).

P-value

Decision Rule for Hypothesis Testing using the *p*-value:

- If $\alpha < P$, accept H_0.
- If $\alpha > P$, reject H_0.

Practical Rule of Thumb:

- If $P < 0.01$, reject H_0 (very strong evidence against H_0).
- If $P > 0.1$, accept H_0 (weak or no evidence against H_0).

Computing P-values

Also, at this border our observed Z-statistic coincides with the critical value z_α,

$$Z_{\text{obs}} = z_\alpha,$$

and thus,

$$P = \alpha = P\{Z \geq z_\alpha\} = P\{Z \geq Z_{\text{obs}}\}.$$

In this formula, Z is any Standard Normal random variable, and Z_{obs} is our observed test statistic, which is a concrete number, computed from data. First, we compute Z_{obs},

$$P\{Z \geq Z_{\text{obs}}\} = 1 - \Phi(Z_{\text{obs}}).$$

We are deciding between the null hypothesis H_0 and the alternative H_A. Observed is a test statistic Z_{obs}. If H_0 were true, how likely would it be to observe such a statistic? In other words, are the observed data consistent with H_0?

A **high** P-**value** tells us that this or even a more extreme value of Z_{obs} is quite possible under H_0, and therefore, we see no contradiction with H_0. The null hypothesis is **not rejected**.
Conversely, a **low** P-**value** signals that such an extreme test statistic is unlikely if H_0 is true. However, we really observed it. Then, our data are **not consistent** with the hypothesis, and we should **reject** H_0.

Calculating P-values

Table: P-value Computation for Z-Tests

Hypothesis $\mathbf{H_0}$	Alternative $\mathbf{H_A}$	P-value (Probability)	P-value (Using Φ)
$\theta = \theta_0$	right-tail $\theta > \theta_0$	$P\{Z \geq Z_{obs}\}$	$1 - \Phi(Z_{obs})$
	left-tail $\theta < \theta_0$	$P\{Z \leq Z_{obs}\}$	$\Phi(Z_{obs})$
	two-sided $\theta \neq \theta_0$	$P\{\lvert Z\rvert \geq \lvert Z_{obs}\rvert\}$	$2(1 - \Phi(\lvert Z_{obs}\rvert))$

Table: P-value Computation for T-Tests

Hypothesis $\mathbf{H_0}$	Alternative $\mathbf{H_A}$	P-value (Probability)	P-value (Using CDF)
$\theta = \theta_0$	right-tail $\theta > \theta_0$	$P\{t \geq t_{obs}\}$	$1 - F_\nu(t_{obs})$
	left-tail $\theta < \theta_0$	$P\{t \leq t_{obs}\}$	$F_\nu(t_{obs})$
	two-sided $\theta \neq \theta_0$	$P\{\lvert t\rvert \geq \lvert t_{obs}\rvert\}$	$2(1 - F_\nu(\lvert t_{obs}\rvert))$

Calculating P-values

Example

A manager evaluates effectiveness of a major hardware upgrade by running a certain process 50 times before the upgrade and 50 times after it. Based on these data, the average running time is 8.5 minutes before the upgrade, 7.2 minutes after it. Historically, the standard deviation has been 1.8 minutes, and presumably it has not changed. Construct a 90% confidence interval showing how much the mean running time reduced due to the hardware upgrade.

Example

Refer to example from last week At the 10% level of significance, we know that the hardware upgrade was successful. Was it marginally successful or very highly successful?

Inference about variances

We'll derive confidence intervals and tests for the population variance:

a variance is a scale and not a location parameter,

b the distribution of its estimator, the sample variance, is not symmetric.

Variance often needs to be estimated or tested for quality control, in order to assess stability and accuracy, evaluate various risks, and also, for tests and confidence intervals for the population means when variance is unknown. The sample variance formula is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

The summands $(X_i - \bar{X})^2$ are not quite independent, as the Central Limit Theorem on p. 93 requires, because they all depend on $\bar{X}$. Nevertheless, the distribution of s^2 is approximately Normal, under mild conditions, when the sample is large.
For small to moderate samples, the distribution of s^2 is not Normal at all. It is not even symmetric. Indeed, why should it be symmetric if s^2 is always non-negative!

Chi-square distribution

When observations $X_1, \ldots, X_n$ are independent and Normal with $\mathrm{Var}(X_i) = \sigma^2$, the distribution of

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

is Chi-square with $(n-1)$ degrees of freedom.

The probability density function (PDF) for a Chi-square random variable X with ν degrees of freedom is:

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, \quad x > 0$$

where Γ is the Gamma function.

The expected value (mean) and variance are:

$$E(X) = \nu$$

$$\mathrm{Var}(X) = 2\nu$$

Confidence interval for the population variance

Let us construct a $(1-\alpha)100\%$ confidence interval for the population variance σ^2, based on a sample of size n. As always, we start with the estimator, the sample variance s^2.

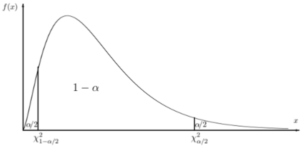


Figure: Critical values of the Chi-square distribution

A rescaled sample variance $\frac{(n-1)s^2}{\sigma^2}$ has χ^2 density

$$P\left(\chi^2_{1-\alpha/2} \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi^2_{\alpha/2}\right) = 1-\alpha.$$

$$P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right) = 1-\alpha.$$

Confidence interval for the population variance and std

Confidence interval for the variance

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right]$$

Confidence interval for the standard deviation

$$\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}\right]$$

Example

There is a sample containing n = 6 measurements, 2.5, 7.4, 8.0, 4.5, 7.4, and 9.2. Give confidence interval for the std deviation!

Testing variance

Table: Summary of Chi-square Test for Population Variance

Null $\mathbf{H_0}$	**Alternative** $\mathbf{H_A}$	**Test Statistic**	**Rejection Region**	P**-value Computation**
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2_{\text{obs}} = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2_{\text{obs}} \geq \chi^2_{\alpha}$	$P\left\{\chi^2 \geq \chi^2_{\text{obs}}\right\}$
	$\sigma^2 < \sigma_0^2$		$\chi^2_{\text{obs}} \leq \chi^2_{1-\alpha}$	$P\left\{\chi^2 \leq \chi^2_{\text{obs}}\right\}$
	$\sigma^2 \neq \sigma_0^2$		$\chi^2_{\text{obs}} \geq \chi^2_{\alpha/2}$ or $\chi^2_{\text{obs}} \leq \chi^2_{1-\alpha/2}$	$2\min\left(P\left\{\chi^2 \geq \chi^2_{\text{obs}}\right\}, P\left\{\chi^2 \leq \chi^2_{\text{obs}}\right\}\right)$

Example

There is a sample containing $n = 6$ measurements, 2.5, 7.4, 8.0, 4.5, 7.4, and 9.2. Calculate P-value!

Comparison of two variances. F-distribution

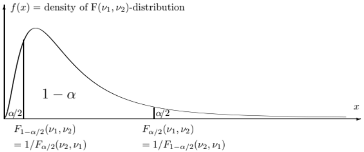
To compare variances or standard deviations, two independent samples $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_m)$ are collected, one from each population. Unlike population means or proportions, variances are scale factors, and they are compared through their ratio $\theta = \frac{\sigma_X^2}{\sigma_Y^2}$. A natural estimator for the ratio of population variances $\theta = \sigma_X^2/\sigma_Y^2$ is the ratio of sample variances $\hat{\theta} = \frac{s_X^2}{s_Y^2} = \frac{\sum(X_i - \bar{X})^2/(n-1)}{\sum(Y_i - \bar{Y})^2/(m-1)}$. The distribution of this statistic is simply called **F-distribution** with $(n-1)$ and $(m-1)$ degrees of freedom.

Distribution of the ratio of sample variances

For independent samples $X_1, \dots, X_n$ from Normal(μ_X, σ_X) and $Y_1, \dots, Y_m$ from Normal(μ_Y, σ_Y), the standardized ratio of variances

$$F = \frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} = \frac{\sum(X_i - \bar{X})^2/\sigma_X^2/(n-1)}{\sum(Y_i - \bar{Y})^2/\sigma_Y^2/(m-1)}$$

has F-distribution with $(n-1)$ and $(m-1)$ degrees of freedom. If F has $F(\nu_1, \nu_2)$ distribution, then the distribution of $\frac{1}{F}$ is $F(\nu_2, \nu_1)$.



Confidence interval for the ratio of population variances

Start with the estimator, $\hat{\theta} = s_X^2 / s_Y^2$. Standardizing it to

$$F = \frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} = \frac{s_X^2/s_Y^2}{\sigma_X^2/\sigma_Y^2} = \frac{\hat{\theta}}{\theta},$$

we get an F-variable with $(n-1)$ and $(m-1)$ degrees of freedom. Therefore,

$$P\left(F_{1-\alpha/2}(n-1, m-1) \le \frac{\hat{\theta}}{\theta} \le F_{\alpha/2}(n-1, m-1)\right) = 1-\alpha,$$

Solving the double inequality for the unknown parameter θ, we get

$$P\left(\frac{\hat{\theta}}{F_{\alpha/2}(n-1, m-1)} \le \theta \le \frac{\hat{\theta}}{F_{1-\alpha/2}(n-1, m-1)}\right) = 1-\alpha.$$

Therefore, the $(1-\alpha)100\%$ confidence interval for the ratio of variances $\theta = \frac{\sigma_X^2}{\sigma_Y^2}$ is:

$$\left[\frac{\hat{\theta}}{F_{\alpha/2}(n-1, m-1)}, \frac{\hat{\theta}}{F_{1-\alpha/2}(n-1, m-1)}\right] = \left[\frac{s_X^2/s_Y^2}{F_{\alpha/2}(n-1, m-1)}, \frac{s_X^2/s_Y^2}{F_{1-\alpha/2}(n-1, m-1)}\right]$$

The critical values of $F(\nu_1, \nu_2)$ and $F(\nu_2, \nu_1)$ distributions are related as follows:

$$F_{1-\alpha}(\nu_1, \nu_2) = \frac{1}{F_\alpha(\nu_2, \nu_1)}$$

The confidence interval is

$$\left[\frac{s_X^2}{s_Y^2 F_{\alpha/2}(n-1, m-1)}, \quad \frac{s_X^2 F_{\alpha/2}(m-1, n-1)}{s_Y^2}\right]$$

F-tests comparing two variances

Table: Summary of F-Test for Comparing Two Population Variances

Null $H_0: \frac{\sigma_X^2}{\sigma_Y^2} = \theta_0$	Alternative H_A	Rejection Region	P-value
θ_0	$\frac{\sigma_X^2}{\sigma_Y^2} > \theta_0$	$F_{obs} \geq F_{\alpha}(n-1, m-1)$	$P\{F \geq F_{obs}\}$
	$\frac{\sigma_X^2}{\sigma_Y^2} < \theta_0$	$F_{obs} \leq F_{1-\alpha}(n-1, m-1)$	$P\{F \leq F_{obs}\}$
	$\frac{\sigma_X^2}{\sigma_Y^2} \neq \theta_0$	$F_{obs} \geq F_{\alpha/2}(n-1, m-1)$ or $F_{obs} \leq F_{1-\alpha/2}(n-1, m-1)$	$2\min\left(P\{F \geq F_{obs}\}, P\{F \leq F_{obs}\}\right)$

Test Statistic: $F_{obs} = \frac{s_X^2}{s_Y^2}/\theta_0$, where F has $F(n-1, m-1)$ distribution.

Example

A data channel has the average speed of 180 Megabytes per second. A hardware upgrade is supposed to improve stability of the data transfer while maintaining the same average speed. Stable data transfer rate implies low standard deviation.

- **Before Upgrade (Population 1, or X):** Sample size $n_1 = 27$ Sample standard deviation $s_1 = 22$ Mbps
- **After Upgrade (Population 2, or Y):** Sample size $n_2 = 16$ Sample standard deviation $s_2 = 14$ Mbps

We are asked to construct a **90% confidence interval** for the relative change in the standard deviation ($\frac{\sigma_1}{\sigma_2}$) (assume Normal distribution of the speed).

Can we infer that the channel became twice as stable as it was, if the increase of stability is measured by the proportional reduction of standard deviation?

Examples

1. We want to investigate whether the daily mean temperature in Budapest on October 18th was **below** 15°**C**. The daily mean temperatures from the past 4 years were as follows: $14.8, 12.2, 16.8, 11.1$ °C Assume that the data originates from a Normal distribution.
   a Write down the **null and alternative hypotheses**.
   b Assume the population standard deviation is known: $\sigma = 2$. Test the hypothesis using a **significance level of** $\alpha = 0.05$.
      - Specify the **critical region** and the *p*-**value**.
      - What is the decision?

   c Test the hypothesis **without using the prior information about the standard deviation** $(\sigma = 2)$.
   d What hypotheses should be formulated if we want to investigate whether the daily mean temperature in Budapest on October 18th was **different from** 15°**C**? Test this hypothesis using the given data.

   Given Critical Values: $(z_{0.05} = -1.645, \quad \Phi(1.275) \approx 0.899, \quad t_{3;0.05} = -2.353, \quad z_{0.975} = 1.96)$

Example

2 The two samples below relate to the defect rates (in per mille) observed in two different factory units. Can we state that factory unit "A" performed better? (We can assume that the samples are normally distributed and independent.)

Unit A	11.9	12.1	12.8	12.2	12.5	11.9	12.5	11.8	12.4	12.9
Unit B	12.1	12.0	12.9	12.2	12.7	12.6	12.6	12.8	12.0	13.1

Given Critical Values:

$$(F_{9,9;0.975} \approx 4.026, \quad t_{18;0.05} \approx -1.734)$$

Example

3 Two servers were compared. The average running time for 30 executions on the first server was 6.7 seconds, while, independently, the average running time for 20 executions on the second server was 7.2 seconds. Investigate whether there is a significant difference between the speeds of the two servers, assuming the standard deviation of the running times was 0.5 seconds on both machines. ($z_{0.975} = 1.96$)

4 The two samples below contain concentration data for an air pollutant found in the atmosphere at 10 busy intersections. The first row contains the figures for November 15th, and the second row contains the figures for November 29th. Has the air pollution level significantly changed?

Date	**Concentration Data**									
November 15th (X)	20.9	17.1	15.8	18.8	20.1	15.6	14.8	24.1	18.9	12.5
November 29th (Y)	21.4	16.7	16.4	19.2	19.9	16.6	15.0	24.0	19.2	13.2

Given Critical Value:

$$(t_{9;0.975} = 2.262)$$