# Statistics: Interval Estimates, Hypothesis Testing

Probability Theory and Statistics, Fall 2023

By: Padmini Mukkamala

Last updated: December 4, 2023

# Contents

**Lecture 24** **2**
Interval estimates with known $\sigma$ . . . 2
Interval estimates with unknown $\sigma$ . . . 2
Interval estimates for $\sigma$ . . . 3
Difference of the mean of two populations . . . 4

**Lecture 25** **5**
Two-sided $z$-test . . . 5
One-sided $z$-test . . . 6
Two-sided $t$-test . . . 7
One-sided $t$-test . . . 7
Test for mean of two populations . . . 8

# Lecture 24

## Confidence Intervals

The estimates we have studied so far are also called **point estimates** because we give a single value as an estimate for the parameter or a function of it. The point estimate is our best guess for the value of the parameter. But it is useful sometimes to give instead a range of values in which the parameter will lie, along with the probability that it would be in this range. For example, instead of saying that the best value for a parameter is say, 1, it might be better to say that with 90% probability our parameter is in the interval $(0.5, 1.5)$. We will look at methods to do this here. In general, given an observation of the sample, we try to find an interval with $1-\alpha$ probability that the unknown parameter will be in that interval. The limits of this interval will both be functions of the observation, thereby, they are values of some statistic. We formalize all these notions below.

### Interval estimates with known $\sigma$

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be i.i.d. random variables, where the mean $\mu \in \mathbb{R}$ is unknown, but $\sigma^2 > 0$ is known. Given a $\alpha \in (0,1)$, a $100(1-\alpha)\%$ confidence interval for the mean of the normal distribution is given by:

$$\overline{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $\Phi(z_\alpha) = 1 - \alpha$.
Proof: We want to find a value $a$ such that,

$$P(-a \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq a) = 1 - \alpha$$

. But since $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ has $N(0,1)$ distribution, $a = \Phi^{-1}(1 - \frac{\alpha}{2})$. We denote this quantity with $z_{\alpha/2}$. And so,

$$P(-z_{\alpha/2} \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) = 1 - \alpha$$

But then the following inequality can be used to find an interval estimate for the population mean as follows:

$$-z_{\alpha/2} \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

would imply,

$$\overline{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Here, we call $100(1-\alpha)\%$ or $1-\alpha$ the confidence coefficient.

When the distributions of $X_i$'s is not normal, then also confidence intervals are made as above. We use Central Limit Theorem to note that $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$. Then the steps that follow are identical when the population standard deviation $\sigma$ is known. But we note here that in while with Normal distribution, the above probability was exact, in this case it is an approximation, and the larger the sample, the better is this approximation.

### Interval estimates with unknown $\sigma$

When the population standard deviation is unknown, then we can still find interval estimates, but would now require the $t$-distribution or Student's distribution. The reason for this is:

Let $n$ be a positive integer. A random variable $X$ is said to have **$t$-distribution with $n$ degrees of freedom**, or **Student's distribution with $n$ degrees of freedom**, if its density function is:

$$f_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \qquad x \in \mathbb{R},$$

where $\Gamma$ is again the Gamma function. We denote this as: $X \sim t(n)$.

The distribution takes its name from William Sealy Gosset, who wrote mathematical papers under the pseudonym "Student".

Let $X_1, \ldots, X_n$ be i.i.d. random variables with $N(\mu, \sigma^2)$ distribution. Let $\overline{X_n}$ denote the sample mean, and let $S_n^* = \sqrt{S_n^{*2}}$ denote the sample standard deviation. Then the random variable

$$\frac{(\overline{X_n} - \mu)}{S_n^*/\sqrt{n}}$$

has $t$-distribution with $n-1$ degrees of freedom.

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be i.i.d. random variables, where the mean $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown. Given a $\alpha \in (0,1)$, a $100(1-\alpha)\%$ confidence interval for the mean of the normal distribution is:

$$\overline{X}_n - t_{\alpha/2}\frac{S_n^*}{\sqrt{n}} \leq \mu \leq \overline{X}_n + t_{\alpha/2}\frac{S_n^*}{\sqrt{n}}$$

where $t_\alpha$ is the value for which the $t$-distribution with $n-1$ degrees of freedom has a tail probability of $\alpha$.
Proof: We again need a value $a$ such that,

$$P(-a \leq \frac{\overline{X}_n - \mu}{S_n^*/\sqrt{n}} \leq a) = 1 - \alpha$$

. But since $\frac{\overline{X}_n - \mu}{S_n^*/\sqrt{n}}$ has $t$ distribution with $n-1$ degrees of freedom, we can find the value of $a$ from the table, we will denote it with $t_{\alpha/2}$. And so,

$$P(-t_{\alpha/2} \leq \frac{\overline{X}_n - \mu}{S_n^*/\sqrt{n}} \leq t_{\alpha/2}) = 1 - \alpha$$

But then the following inequality can be used to find an interval estimate for the population mean as follows:

$$-t_{\alpha/2} \leq \frac{\overline{X}_n - \mu}{S_n^*/\sqrt{n}} \leq t_{\alpha/2}$$

Or,

$$\overline{X}_n - t_{\alpha/2}\frac{S_n^*}{\sqrt{n}} \leq \mu \leq \overline{X}_n + t_{\alpha/2}\frac{S_n^*}{\sqrt{n}}$$

Here also $100(1-\alpha)\%$ or $1-\alpha$ is called the confidence coefficient.

## Interval estimates for $\sigma$

In many industrial quality control operations, it is important that the standard deviation of the process be limited. For example, in computer chip manufacturing, even a micrometer of difference can lead to faulty equipment. In such cases, we want to sample some manufactured chips and use the sample variance to determine a good estimate for the population variance or standard deviation. We will use the following result without proving it.

Let $X_1, \ldots, X_n$ be i.i.d. random variables with $N(\mu, \sigma^2)$ distribution. Let $S_n^* = \sqrt{S_n^{*2}}$ denote the sample standard deviation. Then the random variable

$$\frac{(n-1)S_n^{*2}}{\sigma^2}$$

has $\chi^2$-distribution with $n-1$ degrees of freedom.

Let $X_1,\ldots,X_n \sim N(\mu,\sigma^2)$ be i.i.d. random variables, where the mean $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown. Given a $\alpha \in (0,1)$, a $100(1-\alpha)\%$ confidence interval for the standard deviation of the normal distribution is given by:

$$S_n^* \sqrt{\frac{n-1}{\chi^2_{\alpha/2}}} \leq \sigma \leq S_n^* \sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}}$$

where $\chi^2_\alpha$ is the value for which the $\chi^2$ distribution with $n-1$ degrees of freedom has the tail probability of $\alpha$.
Proof: Since $\frac{(n-1)S_n^{*2}}{\sigma^2}$ has $\chi^2$ distribution, we know that

$$P(\chi^2_{1-\alpha/2} \leq \frac{(n-1)S_n^{*2}}{\sigma^2} \leq \chi^2_{\alpha/2}) = 1-\alpha$$

so we want,

$$\chi^2_{1-\alpha/2} \leq \frac{(n-1)S_n^{*2}}{\sigma^2} \leq \chi^2_{\alpha/2}$$

we get the required result after rearranging the terms.

### Difference of the mean of two populations

Let $X_1,\ldots,X_n \sim N(\mu_1,\sigma_1^2)$ be i.i.d. random variables and $Y_1, Y_2, ..., Y_n \sim N(\mu_2, \sigma_2^2)$ be another set of i.i.d. random variables. Assume that the means $\mu_1, \mu_2 \in \mathbb{R}$ are unknown, but $\sigma_1^2, \sigma_2^2 > 0$ are known. Given $\alpha \in (0,1)$, a $100(1-\alpha)\%$ confidence interval for the difference $\mu_1 - \mu_2$ is given by

$$[\overline{X}_{n_1} - \overline{Y}_{n_2} - z_{\alpha/2}\sigma^*, \overline{X}_{n_1} - \overline{Y}_{n_2} + z_{\alpha/2}\sigma^*]$$

where

$$\sigma^* = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Let $X_1,\ldots,X_n \sim N(\mu_1,\sigma_1^2)$ be i.i.d. random variables and $Y_1, Y_2, ..., Y_n \sim N(\mu_2, \sigma_2^2)$ be another set of i.i.d. random variables. Assume that the means $\mu_1, \mu_2 \in \mathbb{R}$ and the variances $\sigma_1^2, \sigma_2^2 > 0$ are unknown. Given $\alpha \in (0,1)$, a $100(1-\alpha)\%$ confidence interval for the difference $\mu_1 - \mu_2$ is given by

$$[\overline{X}_{n_1} - \overline{Y}_{n_2} - t_{\alpha/2}\frac{S_P}{\sqrt{n^*}}, \overline{X}_{n_1} - \overline{Y}_{n_2} + t_{\alpha/2}\frac{S_P}{\sqrt{n^*}}]$$

where

$$S_P = \sqrt{\frac{S_X^{*2}(n_1-1) + S_Y^{*2}(n_2-1)}{n_1+n_2-2}}$$

$$n^* = \frac{n_1 n_2}{n_1+n_2}$$

and the $t$-distribution is considered with $n_1+n_2-2$ degrees of freedom.

# Lecture 25

## Hypothesis Testing

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an i.i.d. sample of $n$ elements such that the distribution of the sample elements is unknown. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be an observation of $\mathbf{X} = (X_1, \ldots, X_n)$.

1. $H_0$, or the **null hypothesis** is a statement whose validity depends only on the joint distribution of sample $X_1, \ldots, X_n$, for which $\{H_0 \text{ is true }\}$ is an event. $H_1$, is called the **alternate hypothesis** is the negation of $H_0$: $H_1 = \neg H_0$.

2. The values of $\boldsymbol{X}$ for which we accept the claim that $H_0$ is true are said to be **acceptance region**. It complement, or the values of $\boldsymbol{X}$ for which we reject $H_0$ is said to be the **critical region**.

3. The statistic on the basis of which we define the critical region is called that **test statistic**.

4. We define **Type I error** as the error where we reject $H_0$, but $H_0$ is true. And similarly, we define **Type II error** as the error where we accept $H_0$, but $H_0$ was not true.

5. The **significance level** of the test is a number $0 < \alpha < 1$, and denotes the maximum probability of a Type I error, that is $P(\text{ reject } H_0 | H_0 \text{ is true }) = \alpha$.

6. The tail probability of the test statistic is called the **p-value**. If the p-value is less than significance, then we reject the null hypothesis.

We will look at the different types of $H_0$ and how we test to accept or reject them.

### Two-sided $z$-test

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be i.i.d. random variables, where the mean $\mu \in \mathbb{R}$ is unknown, but $\sigma^2 > 0$ is known. We are given $\alpha \in (0, 1)$, where $\alpha$ is the significance level. Further let the null hypothesis be $H_0 : \mu = \mu_0$ where $\mu_0 \in \mathbb{R}$ is a constant, and so the alternate hypothesis is $H_1 : \mu \neq \mu_0$. Then,
(a) Method of Acceptance region:
$H_0$ is accepted if the sample mean $\overline{X}_n$ is in the range:

$$(\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

where $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$).
(b) Method of Test-Statistic:
We will call $z = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$ as the test statistic. Then, $H_0$ is accepted if it is in the range:

$$(-z_{\alpha/2}, z_{\alpha/2})$$

(c) Method of $p$-value:
Let $p = 1 - \Phi(|\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}|)$. Then we accept the null hypothesis if $p > \alpha/2$ while we reject if $p \leq \alpha/2$. Here $p$ is called the $p$-value of the sample.
We note here that when carrying out the hypothesis test, we will evaluate each of the expressions above for the given observation $\overline{x}_n$.
Proof (only for (a)): Since $\alpha$ is the significance level, it is the probability of Type I error, that is, the null hypothesis is true, but we reject it. If the null hypothesis is true, then $\mu = \mu_0$, so the distribution of all the $X_i$'s is $N(\mu_0, \sigma^2)$.

Given the sample mean, let us assume that we reject the null hypothesis if it is $a$ standard deviations away. Then the probability of Type I error is $P(\overline{X}_n$ is in the critical region $|H_0$ is true) is

$$P(|\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}| \geq a) = \alpha$$

But then,

$$P(-a < \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} < a) = 1 - \alpha$$

But since $\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$ has $N(0,1)$ distribution, $a = \Phi^{-1}(1 - \frac{\alpha}{2})$. We denote this quantity with $z_{\alpha/2}$ And so,

$$P(-z_{\alpha/2} < \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$$

So we will accept the null hypothesis if

$$-z_{\alpha/2} < \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}$$

which is true if $\overline{X}_n$ is in the range:

$$(\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

### One-sided $z$-test

Below we see the case when the null hypothesis is an inequality. First we see the left-sided test.

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be i.i.d. random variables, where the mean $\mu \in \mathbb{R}$ is unknown, but $\sigma^2 > 0$ is known. We are given $\alpha \in (0,1)$, where $\alpha$ is the significance level. Further let the null hypothesis be $H_0 : \mu \leq \mu_0$ where $\mu_0 \in \mathbb{R}$ is a constant, and so the alternate hypothesis is $\mu > \mu_0$. Then,
(a) Method of Acceptance region:
$H_0$ is accepted if the sample mean $\overline{X}_n$ is in the range:

$$\overline{X}_n < \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

where $\Phi(z_\alpha) = 1 - \frac{\alpha}{)}$.
(b) Method of Test-Statistic:
We will call $z = \frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$ as the test statistic. Then, $H_0$ is accepted if $z$ is in the range:

$$(-\infty, z_\alpha)$$

(c) Method of $p$-value:
Let $p = (1 - \Phi(\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}))$. Then we accept the null hypothesis if $p > \alpha$ while we reject if $p \leq \alpha$.
We note here that when carrying out the hypothesis test, we will evaluate each of the expressions above for the given observation $\overline{x}_n$.

Proof: Since $\alpha$ is the significance level, it is the probability of Type I error, that is, the null hypothesis is true, but we reject it. If the null hypothesis is true, then $\mu \leq \mu_0$. But our key observation here is that the Type I error is maximum if $\mu = \mu_0$. From this point, our analysis is very similar to what we did before. For maximum Type I error, the distribution of the $X_i$'s is $N(\mu_0, \sigma^2)$.

Given the sample mean, let us assume that we reject the null hypothesis if it is $a$ standard deviations away. Then the probability of Type I error is $P(\overline{X}_n$ is in the critical region $|H_0$ is true) is

$$P(\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} \geq a) = \alpha$$

or

$$P(\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} < a) = 1 - \alpha$$

But since $\frac{\overline{X}_n-\mu_0}{\sigma/\sqrt{n}}$ has $N(0,1)$ distribution, $a = \Phi^{-1}(1-\alpha)$. We denote this quantity with $z_\alpha$ So we will accept the null hypothesis if

$$\frac{\overline{X}_n-\mu_0}{\sigma/\sqrt{n}} < z_\alpha$$

which is true if:

$$\overline{X}_n < \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

And the following is the right-sided test.

Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ be i.i.d. random variables, where the mean $\mu \in \mathbb{R}$ is unknown, but $\sigma^2 > 0$ is known. We are given $\alpha \in (0,1)$, where $\alpha$ is the significance level. Further let the null hypothesis be $H_0 : \mu \geq \mu_0$ where $\mu_0 \in \mathbb{R}$ is a constant, and so the alternate hypothesis is $\mu < \mu_0$. Then,
(a) Method of Acceptance region:
$H_0$ is accepted if the sample mean $\overline{X}_n$ is in the range:

$$\overline{X}_n > \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$$

where $\Phi(z_\alpha) = 1 - \frac{\alpha}{)}$ (b) Method of Test-Statistic:
Let $z = \frac{\overline{X}_n-\mu_0}{\sigma/\sqrt{n}}$. Then, $H_0$ is accepted if it is in the range:

$$(-z_\alpha, \infty)$$

(c) Method of $p$-value:
Let $p = \Phi(\frac{\overline{X}_n-\mu_0}{\sigma/\sqrt{n}})$. Then we accept the null hypothesis if $p > \alpha$ while we reject if $p \leq \alpha$.
We note here that when carrying out the hypothesis test, we will evaluate each of the expressions above for the given observation $\overline{x}_n$.

Proof: left as an exercise!

### Two-sided $t$-test

When we don't know the standard deviation $\sigma$, then as before with confidence intervals, we will use the $t$-distribution instead of Normal distribution.

Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ be i.i.d. random variables, where both the mean $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown. We are given $\alpha \in (0,1)$, where $\alpha$ is the significance level. Further let the null hypothesis be $H_0 : \mu = \mu_0$ where $\mu_0 \in \mathbb{R}$ is a constant, and so the alternate hypothesis is $H_1 : \mu \neq \mu_0$. Then,
(a) Method of Acceptance region:
$H_0$ is accepted if the sample mean $\overline{X}_n$ is in the range:

$$(\mu_0 - t_{\alpha/2}\frac{S_n^*}{\sqrt{n}}, \mu_0 + t_{\alpha/2}\frac{S_n^*}{\sqrt{n}})$$

where the probability of being at most $t_{\alpha/2}$ for the $t$ distribution with $n-1$ degrees of freedom is $1-\frac{\alpha}{2}$).
(b) Method of Test-Statistic:
We will call $t = \frac{\overline{X}_n-\mu_0}{S_n^*/\sqrt{n}}$ as the test statistic. Then, $H_0$ is accepted if $t$ is in the range:

$$(-t_{\alpha/2}, t_{\alpha/2})$$

### One-sided $t$-test

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ be i.i.d. random variables, where both the mean $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are unknown. We are given $\alpha \in (0, 1)$, where $\alpha$ is the significance level. Further let the null hypothesis be $H_0 : \mu \leq \mu_0$ where $\mu_0 \in \mathbb{R}$ is a constant, and so the alternate hypothesis is $H_1 : \mu > \mu_0$. Then,
(a) Method of Acceptance region:
$H_0$ is accepted if the sample mean $\overline{X}_n$ is in the range:

$$\overline{X}_n < \mu_0 + t_{\alpha/2} \frac{S_n^*}{\sqrt{n}}$$

where the probability of being at most $t_\alpha$ for the $t$ distribution with $n-1$ degrees of freedom is $1-\alpha$.
(b) Method of Test-Statistic:
We will call $t = \frac{\overline{X}_n - \mu_0}{S_n^*/\sqrt{n}}$ as the test statistic. Then, $H_0$ is accepted if $t$ is in the range:

$$(-\infty, t_\alpha)$$

The right-sided counterpart of this is left as an exercise to the reader.

### Test for mean of two populations

First we consider the case when the standard deviations are known.

Let $X_1, \ldots, X_n \sim N(\mu_1, \sigma_1^2)$ be i.i.d. random variables and $Y_1, Y_2, ..., Y_n \sim N(\mu_2, \sigma_2^2)$ be another set of i.i.d. random variables. Assume that the means $\mu_1, \mu_2 \in \mathbb{R}$ are unknown, but $\sigma_1^2, \sigma_2^2 > 0$ are known. We are given $\alpha \in (0, 1)$, where $\alpha$ is the significance level. Further let the null hypothesis be $H_0 : \mu_1 - \mu_2 = D_0$ and so the alternate hypothesis is $H_1 : \mu_1 - \mu_2 \neq D_0$. Then,
(a) Method of Acceptance region:
$H_0$ is accepted if $\overline{X}_{n_1} - \overline{Y}_{n_2}$ is in the range:

$$(D_0 - z_{\alpha/2}\sigma^*, D_0 + z_{\alpha/2}\sigma^*)$$

(b) Method of Test-Statistic:
We will call $z = \frac{\overline{X}_{n_1} - \overline{Y}_{n_2} - D_0}{\sigma^*}$ as the test statistic. Then, $H_0$ is accepted if $z$ is in the range:

$$(-z_{\alpha/2}, z_{\alpha/2})$$

where

$$\sigma^* = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Now we consider the case when the standard deviations are not known.

Let $X_1, \ldots, X_n \sim N(\mu_1, \sigma_1^2)$ be i.i.d. random variables and $Y_1, Y_2, ..., Y_n \sim N(\mu_2, \sigma_2^2)$ be another set of i.i.d. random variables. Assume that the means $\mu_1, \mu_2 \in \mathbb{R}$ and the variances $\sigma_1^2, \sigma_2^2 > 0$ are unknown. We are given $\alpha \in (0, 1)$, and let $H_0 : \mu_1 - \mu_2 = D_0$ and so the alternate hypothesis is $H_1 : \mu_1 - \mu_2 \neq D_0$. Then,
(a) Method of Acceptance region:
$H_0$ is accepted if $\overline{X}_n - \overline{Y}_n$ is in the range:

$$(D_0 - t_{\alpha/2} \frac{S_P}{\sqrt{n^*}}, D_0 + t_{\alpha/2} \frac{S_P}{\sqrt{n^*}})$$

and the $t$-distribution is considered with $n_1 + n_2 - 2$ degrees of freedom.
(b) Method of Test-Statistic:

We will call $t = \frac{\overline{X}_n - \overline{Y}_n - D_0}{S_P/\sqrt{n^*}}$ as the test statistic. Then, $H_0$ is accepted if $t$ is in the range:

$$(-t_{\alpha/2}, t_{\alpha/2})$$

where

$$S_P = \sqrt{\frac{S_X^{*2}(n_1 - 1) + S_Y^{*2}(n_2 - 1)}{n_1 + n_2 - 2}}$$

$$n^* = \frac{n_1 n_2}{n_1 + n_2}$$

and the $t$-distribution is considered with $n_1 + n_2 - 2$ degrees of freedom.

# Wish you Good Luck in all future endeavors!