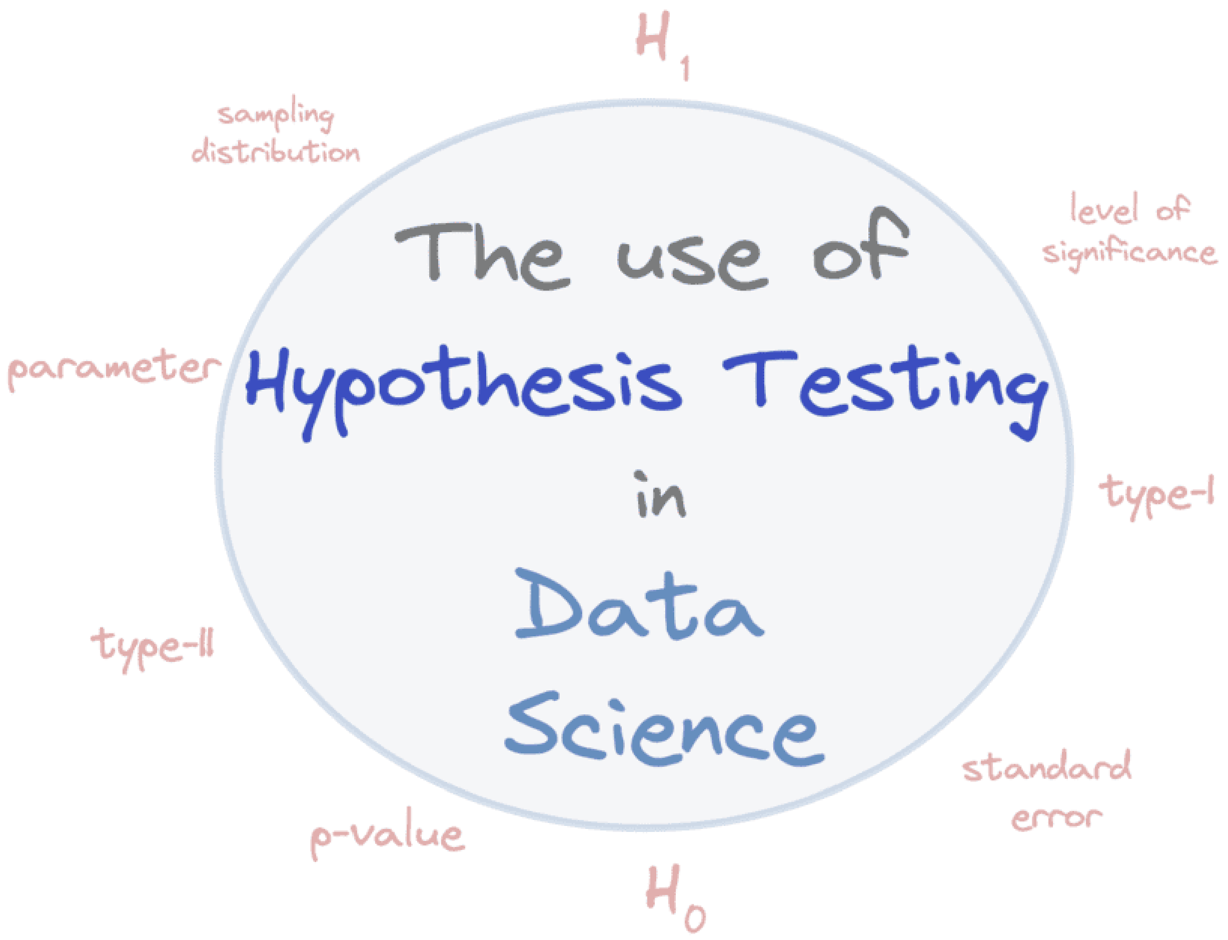
HYPOTHESIS TESTING

Komputasi Statistika

HYPOTHESIS TESTING

Defining a hypothesis allows you to collect data effectively and determine whether it provides enough evidence to support your hypothesis.



FORMULATE HYPOTHESIS

H_0	H_1 / Ha
Assumption, status quo	Rejection of H_0
Assumed to be "true", a given	Rejection of an assumption or the given
Negation of the research question	Research question to be "proven"
Always contains an equality (=, ≤, ≥)	Does not contain equality (≠, <, >)

$H_0: \mu =$ $\quad$ $H_0: \mu \leq$ $\quad$ $H_0: \mu \geq$

$H_1: \mu \neq$ $\quad$ $H_1: \mu >$ $\quad$ $H_1: \mu <$

The **decision** made for the hypothesis is **reject H_0** or **fail to reject H_0**, we don't say accept H_0

TYPE I AND TYPE II ERROR

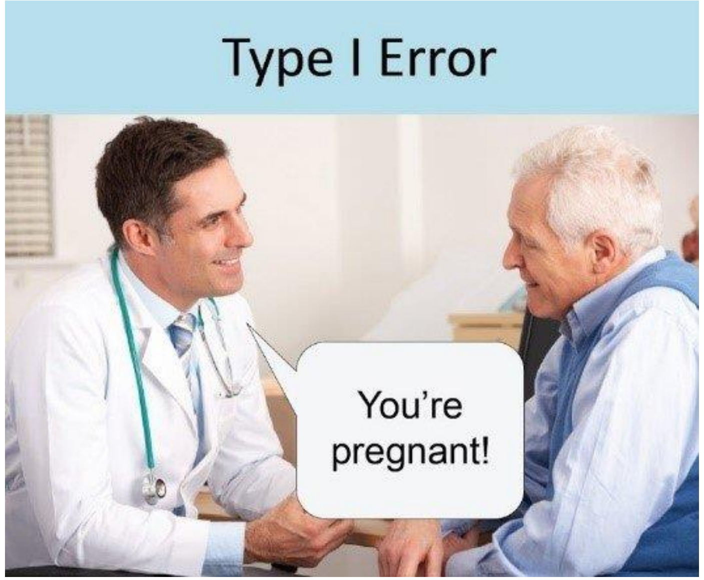
		Actual condition	
		H_0 TRUE	H_0 FALSE
Conclusion	Fail to reject H_0	Correct (True Positive)	**Type II error** (False Negative)
	Reject H_0	**Type I error** (False Positive)	Correct (True Negative)

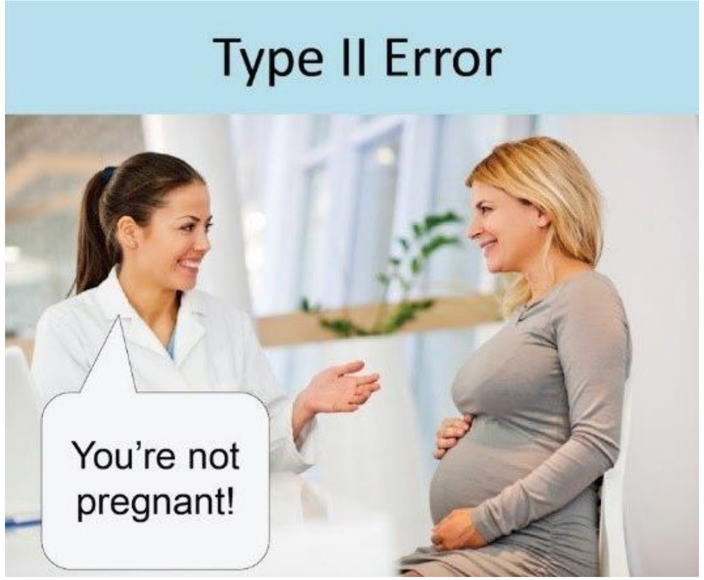
Probabilitas kesalahan tipe 1 (α)

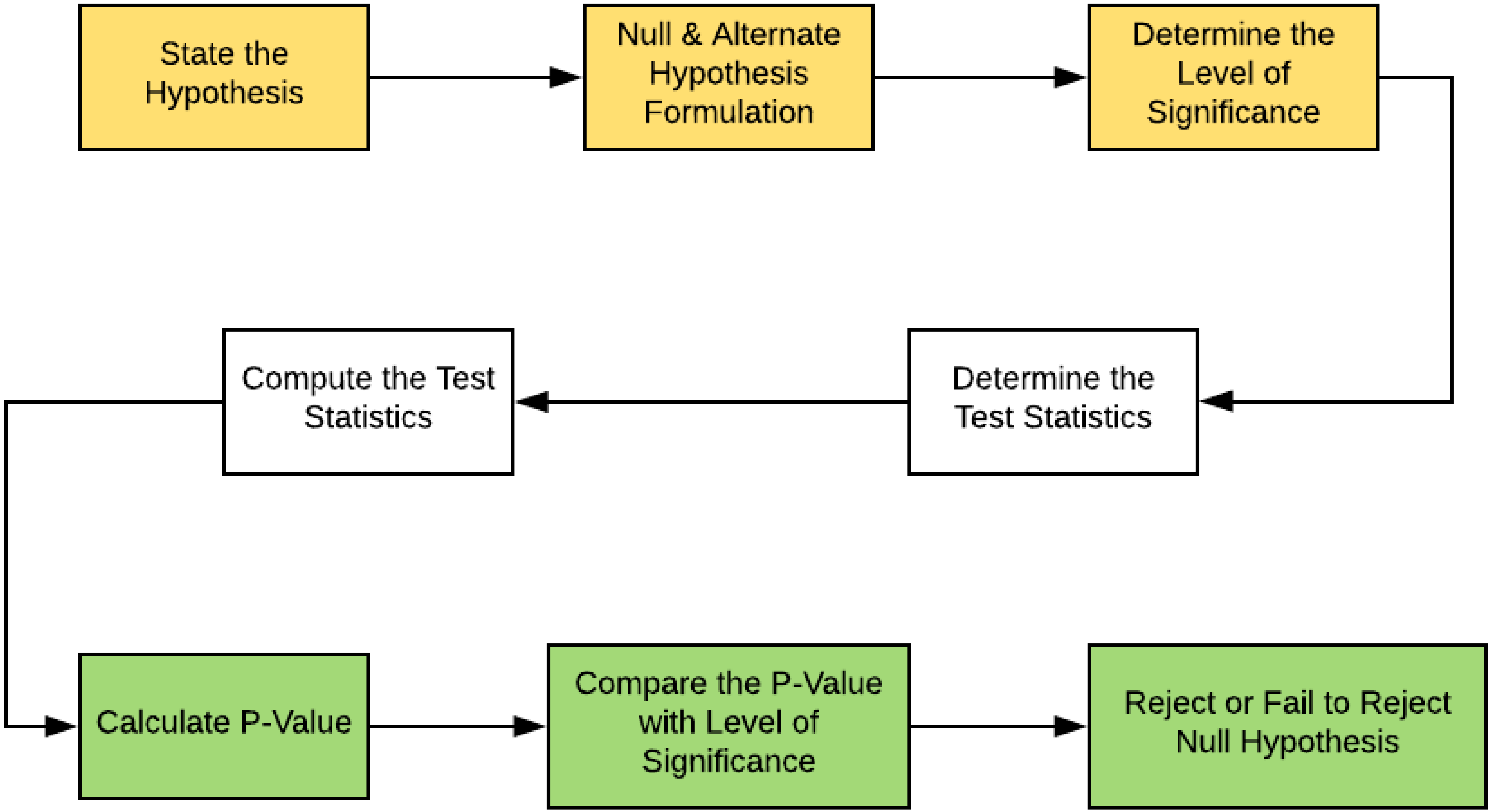
$$\alpha = P(menolak\ H_0 | H_0\ benar)$$

Probabilitas kesalahan tipe 2 (β)

$$\beta = P(menerima\ H_0 | H_0\ salah)$$







Hypothesis Testing Workflow

HYPOTHESIS TESTING POPULATION MEAN

Satu populasi

Large Sample Tests for μ

When the sample size is large, a **Z test** concerning μ is based on the normal test statistic

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

The rejection region is one- or two-sided depending on the alternative hypothesis. Specifically,

$$\begin{array}{lll} H_1: \mu > \mu_0 & \text{requires} & R: Z \geq z_\alpha \\ H_1: \mu < \mu_0 & & R: Z \leq -z_\alpha \\ H_1: \mu \neq \mu_0 & & R: |Z| \geq z_{\alpha/2} \end{array}$$

Hypotheses Tests for μ—Small Samples

To test $H_0: \mu = \mu_0$ concerning the mean of a **normal population,** the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

which has Student's t distribution with $n - 1$ degrees of freedom:

$$\begin{array}{ll} H_1: \mu > \mu_0 & R: T \geq t_\alpha \\ H_1: \mu < \mu_0 & R: T \leq -t_\alpha \\ H_1: \mu \neq \mu_0 & R: |T| \geq t_{\alpha/2} \end{array}$$

The test is called a **Student's *t* test** or simply a ***t* test.**

HYPOTHESIS TESTING POPULATION MEAN

Dua populasi independen

H_0: $\mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$ leading to two-tailed test
H_0: $\mu_1 - \mu_2 \geq (\mu_1 - \mu_2)_0$ leading to left-tailed test
H_0: $\mu_1 - \mu_2 \leq (\mu_1 - \mu_2)_0$ leading to right-tailed test

Cases in which the Test Statistic is Z

1. The **sample sizes** n_1 and n_2 are both **at least 30** and the population standard deviations $\boldsymbol{\sigma_1}$ **and** $\boldsymbol{\sigma_2}$ **are known.**
2. Both populations are **normally distributed** and the population standard deviations $\boldsymbol{\sigma_1}$ **and** $\boldsymbol{\sigma_2}$ **are known.**

The formula for the test statistic Z is

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

where $(\mu_1 - \mu_2)_0$ is the hypothesized value for the difference in the two population means.

HYPOTHESIS TESTING POPULATION MEAN

Dua populasi independen

Cases in which the Test Statistic is *t*

Both populations are **normally distributed;** population standard deviations σ_1 **and** σ_2 **are unknown, but** the sample standard deviations S_1 **and** S_2 **are known.**

The equations for the test statistic *t* depends on 2 subcases:

Subcase 1: σ_1 **and** σ_2 **are believed to be equal** (although unknown. We calculate *t* using the formula

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where S_p^2 is the pooled variance of the two samples, which serves as the estimate of the common population variance given by formula

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

The degrees of freedom for *t* are $(n_1 + n_2 - 2)$.

Cases in which the Test Statistic is *t*

Subcase 2: σ_1 **and** σ_2 **are believed to be unequal** (although unknown. We calculate *t* using the formula

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

The degrees of freedom for this *t* given by

$$df = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{(S_1^2/n_1)^2/(n_1 - 1) + (S_2^2/n_2)^2/(n_2 - 1)}$$

HYPOTHESIS TESTING POPULATION MEAN

Paired-observation

The test statistic for the pair-observation *t* test is

$$t = \frac{\bar{D} - \mu_{D\,0}}{s_D/\sqrt{n}}$$

Where :

$\bar{D}$ is the sample average difference between each pair of observation,

s_D is the sample standard deviation of there differences,

n is the number of pairs of observations, and

$\mu_{D\,0}$ is the population mean difference under the null hypothesis.

HYPOTHESIS TESTING POPULATION MEAN RESUME

H_0	Value of Test Statistic	H_1	Critical Region
$\mu = \mu_0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$; σ known	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$; $v = n - 1$, σ unknown	$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$; σ_1 and σ_2 known	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{1/n_1 + 1/n_2}}$; $v = n_1 + n_2 - 2$, $\sigma_1 = \sigma_2$ but unknown, $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu_1 - \mu_2 = d_0$	$t' = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$; $v = \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$, $\sigma_1 \neq \sigma_2$ and unknown	$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t' < -t_\alpha$ $t' > t_\alpha$ $t' < -t_{\alpha/2}$ or $t' > t_{\alpha/2}$
$\mu_D = d_0$ paired observations	$t = \dfrac{\bar{d} - d_0}{s_d/\sqrt{n}}$; $v = n - 1$	$\mu_D < d_0$ $\mu_D > d_0$ $\mu_D \neq d_0$	$t < -t_\alpha$ $t > t_\alpha$ $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

TESTING POPULATION PROPORTION

Satu populasi

Hypotheses about population proportions can be tested using the binomial distribution or normal approximation to calculate the p-value.

Test statistics :

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

The rejection region is right-sided, left-sided, or two-sided according to

$H_1 : p > p_0, \quad Z \geq Z_\alpha$

$H_1 : p < p_0, \quad Z \leq -Z_\alpha$

$H_1 : p \neq p_0, \quad |Z| \geq Z_{\alpha/2}$

TESTING POPULATION PROPORTION

Dua populasi

Situation I: $H_0: p_1 - p_2 = 0$
$H_1: p_1 - p_2 \neq 0$

Situation II: $H_0: p_1 - p_2 \geq 0$
$H_1: p_1 - p_2 < 0$

Situation III: $H_0: p_1 - p_2 \leq 0$
$H_1: p_1 - p_2 > 0$

Situation A: $H_0: p_1 - p_2 = D$
$H_1: p_1 - p_2 \neq D$

Situation B: $H_0: p_1 - p_2 \leq D$
$H_1: p_1 - p_2 > D$

Situation C: $H_0: p_1 - p_2 \leq D$
$H_1: p_1 - p_2 > D$

The test statistic for the difference between two population proportions when **the null hypothesis difference is zero** is

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

$\hat{p}_1 = x_1/n_1$ is the sample proportion in sample 1

$\hat{p}_2 = x_2/n_2$ is the sample proportion in sample 2

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

The test statistic for the difference between two population proportions when **the null hypothesis difference is some number *D*, other than zero,** is

$$z = \frac{\hat{p}_1 - \hat{p}_2 - D}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

TESTING POPULATION VARIANCE

Satu populasi

To **test hypotheses about** σ, the test statistic is

$$\frac{(n-1)S^2}{\sigma_0^2}$$

Given a level of significance α,

$$\left.\begin{array}{r} \text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma < \sigma_0 \end{array}\right\} \text{ if } \quad \frac{(n-1)S^2}{\sigma_0^2} \le \chi^2_{1-\alpha}$$

$$\left.\begin{array}{r} \text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma > \sigma_0 \end{array}\right\} \text{ if } \quad \frac{(n-1)S^2}{\sigma_0^2} \ge \chi^2_{\alpha}$$

$$\left.\begin{array}{r} \text{Reject } H_0: \sigma = \sigma_0 \\ \text{in favor of } H_1: \sigma \ne \sigma_0 \end{array}\right\} \begin{array}{l} \text{if} \quad \dfrac{(n-1)S^2}{\sigma_0^2} \le \chi^2_{1-\alpha/2} \\ \text{or} \quad \dfrac{(n-1)S^2}{\sigma_0^2} \ge \chi^2_{\alpha/2} \end{array}$$

TESTING POPULATION VARIANCE

Dua populasi

(a two-tailed test)	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$
(a left-tailed test)	$H_0: \sigma_1^2 \geq \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$
(a right-tailed test)	$H_0: \sigma_1^2 \leq \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$

The test statistic for the equality of variances of two normally distributed populations is

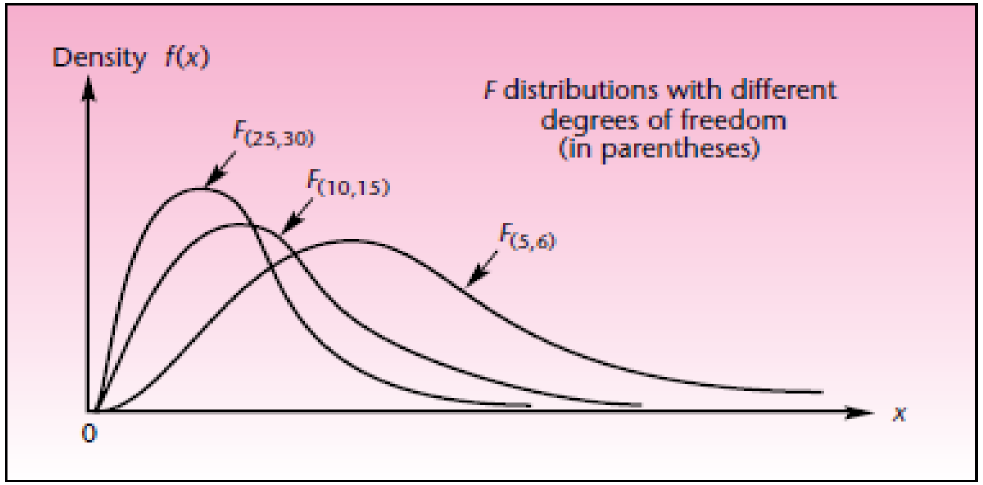
$$F_{(n_1-1,\, n_2-1)} = \frac{S_1^2}{S_2^2}$$

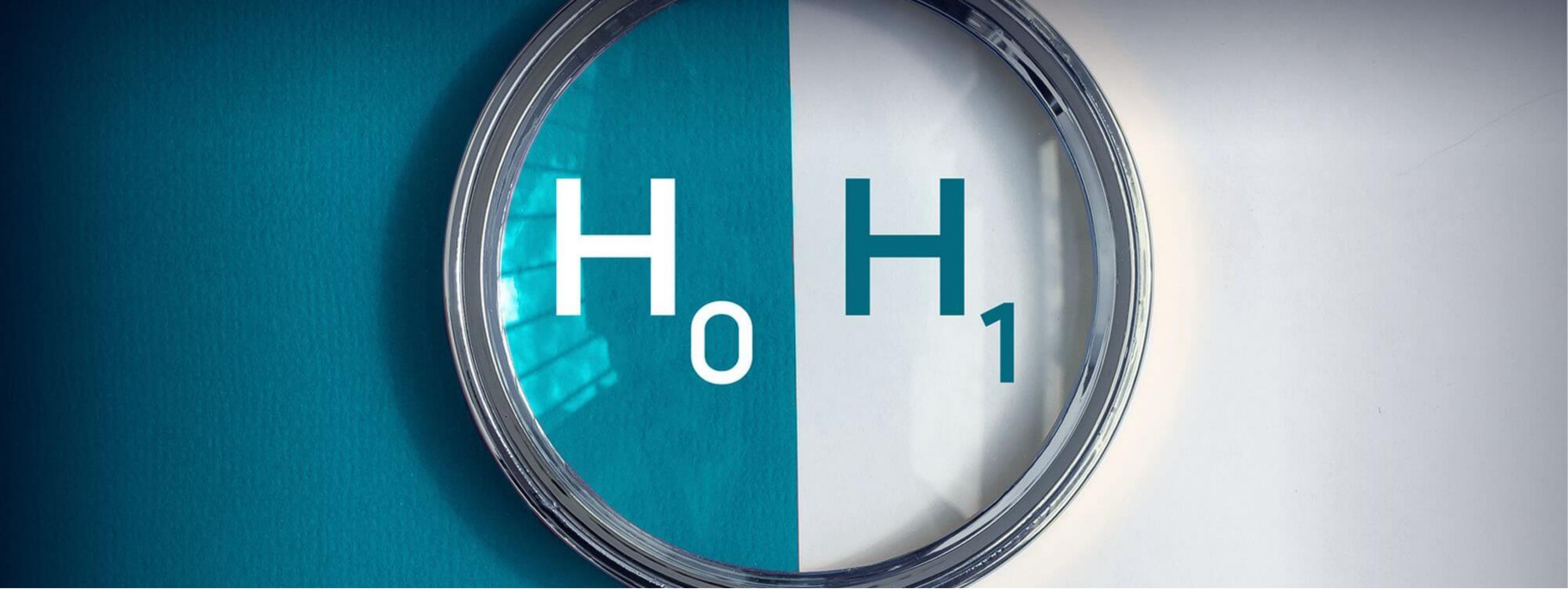
$$\frac{S_1^2}{S_2^2} = \frac{\chi_1^2 \sigma_1^2 (n_1 - 1)}{\chi_2^2 \sigma_2^2 (n_2 - 1)}$$

An F random variable with v_1 and v_2 degrees of freedom is

$$F_{(v_1,\, v_2)} = \frac{\chi_1^2 / v_1}{\chi_2^2 / v_2}$$

where χ_1^2 is chi-square random variable with v_1 degrees of freedom and χ_2^2 is chi-square random variable with v_2 degrees of freedom.





TERIMA KASIH

Prodi S1 Teknologi Sains Data