Statistical tests (means and differences between groups)

venerdì 24 ottobre 2025 18:57

Z-test (μ unknown)

1. Purpose and Context

The **Z-test for a population mean** is used to determine whether the mean of a population is equal to a specified value when the **population standard deviation σ is known**. Because the test relies on the standard normal distribution, it is considered a *parametric* hypothesis test.

Typical applications include:

- industrial quality control
- validation of theoretical or historical mean values
- large-sample inference where σ is known or reliably estimated

2. Hypotheses

Given a random sample $X_1, X_2, \ldots, X_n$ drawn from a population with mean μ and known standard deviation σ, we test:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0 \quad \text{(two-tailed)}$$

or, for one-tailed alternatives:

$$H_1 : \mu > \mu_0 \quad \text{or} \quad H_1 : \mu < \mu_0$$

3. Assumptions

- The sample is drawn from a **normal population**, *or* the sample size is **large** ($n > 30$) so that the **Central Limit Theorem** applies.
- Observations are **independent and identically distributed (i.i.d.)**.
- The population standard deviation σ is **known**.

$\rightsquigarrow$ $X_1 \ldots X_n \overset{iid}{\sim} N(\mu, \sigma_0^2)$ where μ is unknown and σ_0 is known

$$\Rightarrow \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\,?\,,\,?\,)$$

$$?)\ \mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\cdot n\cdot\mu = \mu$$

$$?)\ Var(\bar{X}) = Var\left(\frac{1}{n}\sum X_i\right) = \left(\frac{1}{n}\right)^2 \sum Var(X_i) = \frac{1}{n^2}\cdot n\,\sigma_0^2 = \frac{\sigma_0^2}{n}$$

$$\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma_0^2}{n}\right)$$

$$\Rightarrow Z := \frac{\bar{X}-\mu}{\sqrt{\sigma_0^2/n}} = \frac{\bar{X}-\mu}{\sigma_0^2/\sqrt{n}} \sim N(0,1)$$

$\Rightarrow$ Under H_0, $Z_0 := \dfrac{\bar{X}-\mu_0}{\sigma_0^2/\sqrt{n}} \sim N(0,1)$

$\Rightarrow$ For a chosen significance level α :

- two-tailed test: reject H_0 if $|Z_0| > z_{\alpha/2}$
- right-tailed test: reject H_0 if $Z_0 > z_\alpha$
- left-tailed test: reject H_0 if $Z_0 < -z_\alpha$

example:

$\sigma = 10$
$n = 36$
$\mu_0 = 100$
$\bar{x} = 104$
$\alpha = 0{,}05 \Rightarrow z_{\alpha/2} = z_{0{,}025} \cong 1{,}96$

$$Z_0 = \frac{104 - 100}{10/\sqrt{36}} = \frac{4}{10/6} \simeq 2.40$$

$\Rightarrow$ Since $2.40 > 1{,}96$, we reject H_0

$\Rightarrow$ $\mu \neq 100$ $\rightarrow$ this is what we are going to suppose.

t-test (σ and μ unknown)

1. Purpose and Context

The **one-sample t-test** is used to test whether the mean of a population is equal to a specified value when the **population standard deviation is unknown.** In contrast to the Z-test, the t-test replaces the population standard deviation with the **sample standard deviation**, and the test statistic follows a **Student's t distribution** with $n - 1$ degrees of freedom.

Typical applications include:

- evaluating pre/post effects on a single group
- testing a theoretical or reference mean when data are limited
- small-sample inference in experimental and clinical research

2. Hypotheses

Given a sample $X_1, X_2, \ldots, X_n$ with sample mean $\bar{X}$ and sample standard deviation s, we test:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0 \quad \text{(two-tailed)}$$

or the one-tailed alternatives:

$$H_1 : \mu > \mu_0 \quad \text{or} \quad H_1 : \mu < \mu_0$$

3. Assumptions

- The population distribution is **normal** (this assumption is especially important for small n).
- Observations are **independent and identically distributed** (i.i.d.).
- The population standard deviation σ is **unknown.**

$\leadsto X_1 \dots X_n \sim N(\mu, \sigma^2)$ iid

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{and} \quad s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

$\Rightarrow$ We have already seen that $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

But what about s^2?

$$X_i \sim N(\mu, \sigma) \text{ and } \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \overset{?}{\Rightarrow} X_i - \overline{X} \sim N(?, ?)$$

We know that a linear combination of normal variables is still normal

$\Rightarrow s^2$ is a χ^2 but we need to discover the degree of freedom

PROOF

$$X_i - \overline{X} = X_i - \frac{1}{n} \sum_{i=1}^{n} X_i = X_i - \frac{1}{n} X_i - \frac{1}{n} X_j - \frac{1}{n} X_t \dots = \left(1 - \frac{1}{n}\right) X_i - \frac{1}{n} \sum_{j \neq i} X_j$$

$$\text{*)}\ \mathbb{E}(X_i - \overline{X}) = \mathbb{E}\left[\left(1 - \frac{1}{n}\right) X_i - \frac{1}{n} \sum_{j \neq i} X_j\right] =$$

because of linearity of $\mathbb{E}(\cdot)$

and of independence of X_i and X_j

$$= \left(1 - \frac{1}{n}\right) \mathbb{E}(X_i) - \frac{1}{n} \sum_{j \neq i} \mathbb{E}(X_j) =$$

$$= \left(1 - \frac{1}{n}\right) \mu - \frac{1}{n} \cdot (n-1) \mu = \mu \left[\left(1 - \frac{1}{n}\right) - \frac{n-1}{n}\right] = \mu \left[\frac{n-1}{n} - \frac{n-1}{n}\right] = 0$$

$$\text{*)}\ Var(X_i - \overline{X}) = Var\left(X_i - \frac{1}{n} \sum X_i\right) = Var\left(\left(1 - \frac{1}{n}\right) X_i - \frac{1}{n} \sum_{j \neq i} X_j\right) =$$

$$= \left(1 - \frac{1}{n}\right)^2 \sigma^2 + \frac{1}{n^2} (n-1) \sigma^2 = \sigma^2 \left[\left(\frac{n-1}{n}\right)^2 + \frac{(n-1)}{n^2}\right] =$$

$$= \sigma^2 \left[\frac{(n-1)^2 + (n-1)}{n^2}\right] = \sigma^2 \left[\frac{n^2 + 1 - 2n + n - 1}{n^2}\right] = \sigma^2 \frac{n^2 - n}{n^2} =$$

$$= \sigma^2 \frac{\cancel{n}(n-1)}{\cancel{n^2}} = \sigma^2 \frac{n-1}{n}$$

$$\Rightarrow \frac{S^2(n-1)}{\sigma^2} \sim \chi^2(n-1)$$

$\Rightarrow$ If we define $T_0 := \dfrac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$

PROOF

$$\bar{X}-\mu_0 \sim N\left(0, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim N(0,1)$$

$$\Rightarrow \frac{(\bar{X}-\mu_0)\sqrt{n}}{\sigma} \cdot \sqrt{\frac{(n-1)}{\frac{(n-1)S^2}{\sigma^2}}} \sim t(n-1)$$

(It's the definition of t-student : see pdf "Probability distributions")

$$= \frac{(\bar{X}-\mu_0)\sqrt{n}}{\sigma} \cdot \sqrt{\frac{\sigma^2}{S^2}} = \frac{\bar{X}-\mu_0}{S} \cdot \sqrt{n} \quad \square$$

$\Rightarrow$ For a chosen significance level α :

- Two-tailed test: reject H_0 if $|T| > t_{\alpha/2, n-1}$
- right-tailed test : reject H_0 if $T > t_{\alpha, n-1}$
- left-tailed test: reject H_0 if $T < -t_{\alpha, n-1}$

example

$n=16 \Rightarrow n-1=15 \quad t_{0.025,15} \cong 2.131$

$\bar{X}=52$

$s=8$

$\mu_0=50$

$\alpha=0.05$

$\Rightarrow T_0 = \frac{52-50}{8} \cdot 4 = \frac{2 \cdot 4}{8} = 1$

$\Rightarrow$ Since $1.00 < 2.13$, we fail to reject H_0

$\Rightarrow$ This sample doesn't provide sufficient evidence that the mean differs from 50.

But the p-value for $t=1.00$ with 15 df is ~ 0.33, which is > 0.05, therefor the result is not statistically significant at the 5% level.

Two-sample indipendent t-test

1. Purpose and Context

The two-sample independent t-test (pooled t-test) compares the means of two independent groups to assess whether they come from populations with the same mean, under the **assumption that the two populations have equal variances** ($\sigma_1^2 = \sigma_2^2$). It is appropriate when samples are independent, the data are (approximately) normal, and the homogeneity of variance assumption is plausible.

Typical applications:

- comparing treatment vs control group outcomes
- comparing means across two independent demographic groups
- laboratory experiments with two independent samples where variance homogeneity is plausible

2. Hypotheses

Let μ_1 and μ_2 be the two population means. The null and alternative hypotheses are:

$$H_0 : \mu_1 = \mu_2 \quad \text{(equivalently } \mu_1 - \mu_2 = 0)$$

$$H_1 : \mu_1 \neq \mu_2 \quad \text{(two-tailed)}$$

(or one-tailed alternatives $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$ when appropriate).

3. Assumptions

- Observations within each sample are i.i.d. and independent across samples.
- Each population is approximately normal (important for small samples; CLT helps when samples are large).
- **Equal population variances:** $\sigma_1^2 = \sigma_2^2 = \sigma^2$. This is the critical assumption that justifies pooling the variance estimate.
- Samples are random and measured on an interval/ratio scale (or approximately so).

If the equal-variance assumption is doubtful, use Welch's t-test (unequal variances) rather than pooling.

$X_1 \dots X_n \sim N(\mu_1, \sigma^2) \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$

$Y_1 \dots Y_m \sim N(\mu_2, \sigma^2)$

$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}) \qquad S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})$

$$\Rightarrow S_p^2 := \frac{S_x^2(n-1) + S_y^2(m-1)}{n+m-2}$$

<u>oss</u> $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$

<u>PROOF</u>

$$*)\ \mathbb{E}(\bar{X} - \bar{Y}) \underset{\substack{\text{for independence} \\ \text{between } X_i \text{ and } Y_j \\ \forall i = 1:n \text{ and } \forall j = 1:m}}{=} \mathbb{E}(\bar{X}) - \mathbb{E}(\bar{Y}) = \mu_1 - \mu_2$$

$$*)\ Var(\bar{X} - \bar{Y}) = Var\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}) - \frac{1}{m}\sum_{i=1}^{m}(Y_i - \bar{Y})\right) =$$

$$= \frac{1}{n^2} \cdot n\, Var(X_i) - \frac{1}{m^2} \cdot m \cdot Var(Y_i) = \frac{1}{n}\sigma^2 + \frac{1}{m}\sigma^2 = \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right) \qquad \square$$

$\Rightarrow$ Under H_0 we have $\Delta_0 := \mu_1 - \mu_2 \underset{H_0}{=} 0$

$\Rightarrow$ Let's define $T := \dfrac{\bar{X} - \bar{Y} - \Delta_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$

$\Rightarrow T_0 = \dfrac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2)$ (the proof is the same of the previous one)

$\Rightarrow$ For a chosen significance level α:

- Two-tailed test: reject H_0 if $|T| > t_{\alpha/2, \nu}$
- right-tailed test: reject H_0 if $T > t_{\alpha, \nu}$
- left-tailed test: reject H_0 if $T < -t_{\alpha, \nu}$

where $\nu = n + m - 2$

example

Group 1: $n = 20$, $\overline{X} = 105$, $S_X^2 = 100$ $\alpha = 0.05$

Group 2: $m = 18$, $\overline{Y} = 100$, $S_Y^2 = 144$

$$\Rightarrow S_p^2 = \frac{(20-1)\cdot 100 + (18-1)\cdot 144}{20+18-2} \cong 120{,}78 \quad \Rightarrow S_p = \sqrt{S_p^2} \cong 10.99$$

$$\Rightarrow T_0 = \frac{105-100}{10.99\sqrt{\frac{1}{20}+\frac{1}{18}}} \cong 1.40 \qquad \nu = 20+18-2 = 36$$

$$\Rightarrow t_{0{,}025,\,36} \cong 2{,}028$$

$\Rightarrow$ since $1.40 < 2.028$, we fail to reject H_0

But the p-value for $t_{36} = 1.40$ is $\sim 0{,}17$, therefor the observed difference is not statistically significant at the 5% level.

Welch's test

1. Purpose and Context

Welch's t-test compares the means of two independent samples when **the two population variances cannot be assumed equal**. It is the recommended two-sample test when heteroscedasticity (unequal variances) is suspected because it adjusts the standard error and uses an approximate degrees-of-freedom (Welch–Satterthwaite) to control Type I error better than the pooled t.

Typical applications:

- comparing group means with different variability (e.g., different instruments, subpopulations)
- observational studies where homogeneity of variance is unlikely
- routine practice when you don't want to rely on the equal-variance assumption

2. Hypotheses

Let μ_1 and μ_2 be the two population means. Common formulations:

$$H_0 : \mu_1 = \mu_2 \quad (\text{or } \mu_1 - \mu_2 = 0)$$

$$H_1 : \mu_1 \neq \mu_2 \quad (\text{two-tailed})$$

(or one-tailed alternatives $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$ when appropriate).

Welch's test **does not** assume $\sigma_1^2 = \sigma_2^2$.

3. Assumptions

- Two samples are independent.
- Observations within each sample are independent and identically distributed.
- Each population is approximately normal (for small sample sizes). With moderate/large samples the test is robust by the CLT.
- Variances may be unequal (that is the point of Welch's test).

The statistic is:

$$T := \frac{\overline{X} - \overline{Y} - \Delta_0}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$

Under H_0, $T \sim t(\nu)$ where ν is given by the following Welch-Satterthwaite formula:

$$\nu = \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{(S_x^2/n)^2}{n-1} + \frac{(S_y^2/m)^2}{m-1}}$$

It is quite similar to the previous test, but we don't have the pooled variance and the degree of freedom of T is more complicated to be obtained

Paired t-test

1. Purpose and Context

The paired t-test compares **two related measurements on the same subjects** (or matched/blocked units). It tests whether the **mean of their differences** is significantly different from zero.

Typical scenarios:

- *Before vs. After* measurements on the same individuals
- Matched pairs (e.g., twins, or cases matched by age/sex)
- Repeated measures on the same experimental unit

The key idea: instead of comparing two independent samples, we reduce the problem to a **one-sample t-test on the differences.**

2. Hypotheses

Let $D_i = X_{i,\text{after}} - X_{i,\text{before}}$ be the paired differences. Then the hypotheses are:

$$H_0 : \mu_D = 0$$

$$H_1 : \mu_D \neq 0 \quad \text{(two-tailed)}$$

(or one-tailed versions $\mu_D > 0$ or $\mu_D < 0$ if direction is specified).

3. Assumptions

- The pairs are dependent (same subject or matched design).
- The differences D_i are i.i.d.
- The distribution of the differences is approximately normal (especially important if n is small).
- The pairs themselves are independent across subjects.

Note: **we do not assume equal variances**, because there is only one set of differences.

n = number of differences $\quad \bar{D} = \frac{1}{n} \sum_{i=1}^{n} \Delta_i \quad s_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\Delta_i - \bar{D})^2$

$\rightsquigarrow$ the statistic is

$$T := \frac{\bar{D} - \Delta_0}{s_D / \sqrt{n}} \qquad \text{Under } H_0,\ T_0 \sim t(n-1)$$

The decision rule is as usual

example

Suppose we measure blood pressure of $n = 8$ patients before and after treatment

$$D = \{-2, 1, 0, -1, -3, 2, -2, -1\}$$

$$\Rightarrow \bar{D} = \frac{1}{8}(-2+1+0-1-3+2-2-1) = -0.75$$

$$s_D^2 = \frac{1}{8-1} \sum_{i=1}^{8} (\Delta_i + 0.75)^2 \cong 2.79 \Rightarrow s_D = \sqrt{s_D^2} \cong 1.67$$

$$\Rightarrow T = \frac{-0.75 - 0}{1.67/\sqrt{8}} \cong -1.271 \qquad t_{0,025}(7) = x$$

if $|T| > x$, we reject H_0

Chi-square test for variance

1. Purpose and Context

The **chi-square test for population variance** is used to determine whether the **variance (or standard deviation) of a normally distributed population** equals a specified value.

It is typically used when you want to check **consistency or stability of variability**, for instance:

- To test if a manufacturing process has changed in variability,
- To verify whether the observed sample variation meets quality control specifications,
- To compare empirical variance with a theoretical model's variance.

The test directly concerns the **population variance** σ^2, not the mean.

2. Hypotheses

Let s^2 be the sample variance from a sample of size n.

We test:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 \neq \sigma_0^2 \quad \text{(two-tailed)}$$

One-tailed versions are also possible:

$$H_1 : \sigma^2 > \sigma_0^2 \quad \text{or} \quad H_1 : \sigma^2 < \sigma_0^2$$

3. Assumptions

1. The data come from a **normally distributed population.**
2. Observations are **independent.**
3. The hypothesized variance σ_0^2 is a **fixed, known value.**

Note that normality is **essential**; the test statistic's distribution depends critically on it.

$X_1 \ldots X_n \sim N(\mu_0, \sigma^2)$ we have seen at the beginning that

$$\chi^2 := \frac{s^2 (n-1)}{\sigma^2} \sim \chi^2(n-1)$$

$\rightsquigarrow$ The statistic is χ^2

$\Rightarrow$ The decision rule will be (at α-level):

- Two-tailed test: reject H_0 if $\chi^2 < \chi^2_{\alpha/2}(n-1)$ or $\chi^2 > \chi^2_{1-\alpha/2}(n-1)$
- Right-tailed test: reject H_0 if $\chi^2 > \chi^2_{1-\alpha}(n-1)$
- Left-tailed test: reject H_0 if $\chi^2 < \chi^2_{\alpha}(n-1)$

example

A company claims that the standard deviation of weights of packaged goods is $\sigma_0 = 0.5$ kg.
A random sample of $n = 10$ packages has a sample standard deviation of $s = 0.8$ kg.
We test whether the true variance differs from the claimed value at the $\alpha = 0.05$ level.

$\Rightarrow \quad \sigma_0^2 = 0.5^2 = 0.25 \qquad s^2 = 0.8^2 = 0.64$

$\Rightarrow \quad \chi^2(10-1) = \chi^2(9) := \dfrac{9 \cdot 0.64}{0.25} = 23.04$

while $\chi^2_{1-\alpha/2}(9) = 19.023$ $\chi^2_{\alpha/2}(9) = 2.700$

But since $\chi^2 = 23.04 > 19.023$, we reject H_0

⇒ There is significance evidence that the population variance differs from 0.25; the process is more variable than claimed

F-test for equality of two variances

1. Purpose and Context

The **F-test for equality of variances** is used to compare the **variability of two populations** that are assumed to be **normally distributed and independent**.
It tests whether the two population variances are equal:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

This test is foundational for:

- Checking the assumption of **homogeneity of variances** before applying a two-sample t-test (pooled version),
- Serving as the theoretical basis for **ANOVA**,
- Assessing consistency between two measurement systems or production processes.

2. Hypotheses

Let s_1^2 and s_2^2 be the sample variances of two independent samples of sizes n_1 and n_2.

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad \text{(two-tailed)}$$

or, if directionality is known:

$$H_1 : \sigma_1^2 > \sigma_2^2 \quad \text{or} \quad H_1 : \sigma_1^2 < \sigma_2^2$$

3. Assumptions

1. The two populations are **independent**.
2. Each population is **normally distributed**.
3. The samples are **randomly drawn** and **independent**.

Violation of normality can greatly distort the F-test — it is **not robust** to non-normal data.

$X_1 \dots X_n \sim N(\mu, \sigma_1)$ $s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})$ where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

$Y_1 \dots Y_m \sim N(\mu, \sigma_2)$ $s_y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})$ where $\bar{Y} = \frac{1}{m} \sum_{i=1}^{m} Y_i$

$\Rightarrow F := \frac{s_x^2}{s_y^2} \sim F_{n-1,m-1}$ (see the definition of Fisher distribution in the pdf "Probability distributions")

⇒ The decision rule (at α-level) will be:

- Two-tailed test: reject H_0 if $F < F_{\alpha/2}(n-1, m-1)$ or if

$$F > F_{1-\alpha/2}(n-1, m-1)$$

- Right-tailed test: reject H_0 if $F > F_{1-\alpha}(n-1, m-1)$

- Left-tailed test: reject H_0 if $F < F_{\alpha}(n-1, m-1)$

example

Suppose we have two machines producing metal rods, and we wish to test whether they produce with the same variance in length.

$n = 10 \qquad S_1^2 = 0.025$

$m = 12 \qquad S_2^2 = 0.010$

we test $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_0^2 \neq \sigma_1^2$ at $\alpha = 0.05$

$$\rightsquigarrow F = \frac{S_1^2}{S_2^2} = \frac{0.025}{0.01} = 2.5$$

$$\rightsquigarrow F_{0.025}(10-1, 12-1) = F_{0.025}(9, 11) = 0.248 \quad \text{and}$$

$$F_{1-0.025}(9, 11) = 3.59$$

$\Rightarrow$ Since $F = 2.5$ falls between 0.248 and 3.59, we fail to reject H_0.

$\Rightarrow$ There is no significance evidence that the variances differ; both machines appear to have similar variability.

One-way ANOVA

1. Purpose and Context

The **One-Way Analysis of Variance (ANOVA)** is used to test whether **three or more independent samples** come from populations with **equal means.**

It generalizes the two-sample t-test:
instead of comparing means of two groups, it evaluates whether **any difference exists among k group means**, using a ratio of between-group to within-group variability.

2. Hypotheses

Suppose we have k independent groups, each with sample size n_i, mean $\bar{X}_i$, and common population variance σ^2.

The hypotheses are:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_1 : \text{At least one } \mu_i \text{ differs}$$

3. Assumptions

1. **Independence**: observations are independent both within and between groups.
2. **Normality**: each group is normally distributed.
3. **Homogeneity of variances**: all populations share a common variance ($\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$).

ANOVA is **robust** to moderate violations of normality, but **sensitive** to large differences in group variances (especially with unequal sample sizes).

The key idea is to partition the total variation in the data in:

1) Between group variation: variation due to differences among group means

2) Within group variation: variation due to random nois within each group

$\Rightarrow$ $SST := SSB + SSW$ where:

$$\begin{cases} X_1^1 \dots X_{n_1}^1 & : 1° \text{ group} \\ X_1^2 \dots X_{n_2}^2 & : 2° \text{ group} \\ \vdots & \\ X_1^k \dots X_{n_k}^k & : k° \text{ group} \end{cases}$$

1) $SSB := \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2$

2) $SSW := \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_j^i - \bar{X}_i)^2$

$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j^i$ and

$\bar{X} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_j^i$

where $N = \sum_{i=1}^{k} n_i$

$\Rightarrow \quad MS_B = \dfrac{SSB}{k-1}$ and $MS_W = \dfrac{SSW}{N-k}$

$\rightsquigarrow$ the statistic is $F := \dfrac{MS_B}{MS_W} \sim F_{k-1,N-k}$

$\Rightarrow$ The decision rule (at α-level) will be:

we reject H_0 if $F > F_{1-\alpha}(k-1, N-k)$

example

Suppose three fertilizers A, B, C are tested on crops, and we record yields:

fertilizer	observations
A	20, 21, 19, 22
B	18, 17, 20, 19
C	23, 22, 24, 21

we want to test if the fertilizers differ in mean yield at $\alpha = 0.05$ level.

$\Rightarrow \quad \bar{X}_A = \dfrac{20+21+19+22}{4} = 20.5 \qquad \bar{X}_B = \ldots = 18.5 \qquad \bar{X}_C = \ldots = 22.5$

$$\bar{X} = \frac{20.5 + 18.5 + 22.5}{3} = 20.5$$

$\Rightarrow \quad SSB = 4 \cdot \left[(20.5 - 20.5)^2 + (18.5 - 20.5)^2 + (22.5 - 20.5)^2 \right] = 32$

$\Rightarrow SSW = SSW_A + SSW_B + SSW_C$ where:

$$SSW_A = (20-20.5)^2 + (21-20.5)^2 + (19-20.5)^2 + (22-20.5)^2 = 5.0$$

$$SSW_B = (18-18.5)^2 + (17-18.5)^2 + (20-18.5)^2 + (19-18.5)^2 = 5.0$$

$$SSW_C = (23-22.5)^2 + (22-22.5)^2 + (24-22.5)^2 + (21-22.5)^2 = 5.0$$

$\rightsquigarrow \quad SSW = 5 + 5 + 5 = 15$

$SSW = 5 + 5 + 5 = 15$

$\Rightarrow \quad MS_B = \dfrac{32}{3-1} = 16$ and $MS_W = \dfrac{15}{12-3} = 1.667$

$\Rightarrow \quad F = 16/1.667 = 9.6$ and $F_{1-0.05}(2,9) = 4.26$

$\Rightarrow$ Since $9.6 > 4.26$, we reject H_0. There is statistically significant difference among the fertiliter means.