Hypothesis

In order to make decisions it is useful to make some assumptions about the population. Such assumptions, which may or may not be true, are known as hypothesis. These are the tentative, declarative statement about the relationship between two or more variables. There are two types of statistical hypotheses for each situation: the **null hypothesis** and the **alternative hypothesis**. Both of these hypotheses contain opposite view points.

Null Hypothesis (H_0)

- The **null hypothesis (H_0)** states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters (i.e. H_0 is a statement of a no relationship)- It explicitly says that the two groups we are studying are the same.

Alternative Hypothesis (H_1)

- The **alternative hypothesis (H_1)** states the existence of a difference between a parameter and a specific value, or states that there is a difference between two parameters- In other words it says that the two groups we are studying are different

Symbols used in hypothesis

H_0	H_1
equal (=)	not equal (≠) or greater than (>) or less than (<)
greater than or equal to (≥)	less than (<)
less than or equal to (≤)	more than (>)

Example 1

If we want to examine that on an average college student take less than five years to complete their education. The null and alternative hypotheses are:

$$H_0 : \mu \geq 5$$
$$H_1 : \mu < 5$$

Writing null hypothesis

Case 1: Suppose a cake baked through conventional method has an average life span of μ days and it is proposed to test a new process of baking cakes. So, we have two populations of cakes (one by conventional method and other by new process). Here hypothesis can be formed like:

(i) New method is better than conventional method.

(ii) New method is inferior to conventional method.

(iii) There is no difference between the two methods.

Since first two statements display a preferential mentality, they tend to be biased. As a result, adopting the hypothesis of no difference, i.e. a neutral or null attitude toward the outcome, is the safest course of action. Thus, if the average life of cakes baked using the new method is μ_0, the null hypothesis is:

$$H_0: \mu = \mu_0$$

Case 2 : Suppose a departmental store is planning to have its own android application (app) conditioned that new service will be introduced only if more than 60% of its customers use internet to shop. So here null hypothesis would be that % of customers using internet is less or equal to 60% and the alternative hypothesis will be its opposite.

H_0: Proportion of customers using internet for

shopping ≤ 60%

H_1: Proportion of customers using internet for

shopping > 60%

If the null hypothesis is rejected, then the alternative hypothesis H_1 will be accepted and as a result e-commerce shopping service will be introduced.

NOTE

- *After stating the hypothesis, the researcher designs the study. The researcher selects the correct statistical test, chooses an appropriate level of significance, and formulates a plan for conducting the study.*
- *If we discard the null hypothesis, then we can assume there is enough evidence to support the alternative hypothesis.*

Standard Error of Mean (σ_M)

When we take a sample from a population, we pick up one of many samples. Some of them will have the same mean whereas some will have very different means. Standard error of the mean (SEM) measures how much dispersion there is likely to be in a sample's mean compared to the population mean i.e it measures the standard deviation of sampling distribution about the mean.

$$\sigma_M = \frac{\sigma}{\sqrt{N}}$$

σ = Standard deviation of original

distribution

N = Sample size

- **Small SEM:** Having large number of observations and all of them being close to the sample mean (large N, small SD) gives us confidence that our estimation of the population means (i.e., that it equals the sample mean) is relatively accurate.
- **Large SEM:** Having small number of observations and they vary a lot (small N, large SD), then population estimation is likely to be quite inaccurate.

Degrees of freedom

The number of independent pieces of information on which an approximation is based is known as the degrees of freedom. You can also think of it as the number of values that are free to vary as you estimate parameters.

Example 1

Consider a classroom having seating capacity of 30 students. The first 29 students have a choice to sit but the 30th student can only sit on the one remaining seat. Therefore, the degrees of freedom are 29.

Example 2

For scheduling three hour-long tasks (read, eat and nap) between the hours of 5 p.m. and 8 p.m. we have two degrees of freedom as any two tasks can be scheduled at will, but after two of them have been set in time slots, the time slot for the third is decided by default.

Degrees of freedom is some or other way related with the size of the sample because higher the degrees of freedom generally mean larger sample sizes.

***Note:** A higher degree of freedom means more power to reject a false null hypothesis and find a significant result.*

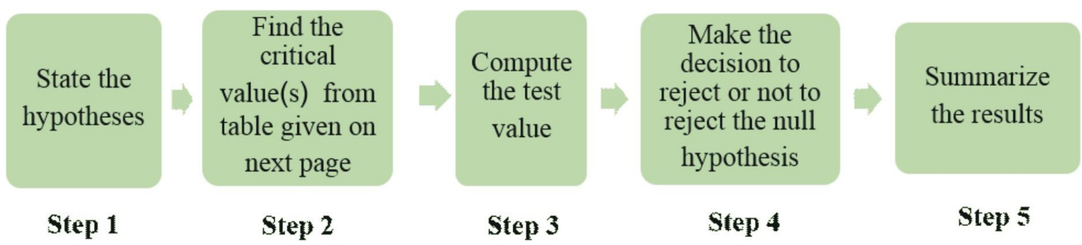
$$Df = N - 1$$

where: *Df = degrees of freedom and*

N = sample size

HYPOTHESIS TESTING PROCEDURE:

Use following procedure for testing the hypotheses by using the t test (traditional method):



PLEASE NOTE AND REFER TO THE NOTEBOOK WHERE THE STEP2 IS COMPUTED VIA SCIPY OR MATH LIBRARIES: https://github.com/Laxminarayen/Inceptez-batch-25-Classwork/blob/main/Python-Class16-Hypothesis2-Inferential3-Pandas/DifferentHypothesisTests.ipynb

Hypothesis Testing: Two Decision Approaches

When performing a hypothesis test, we have two main ways to decide whether to reject or fail to reject the null hypothesis (H_0):

1. Test Statistic (Critical Value) Method

Idea:
We compute a **test statistic** (Z, t, χ^2, F, etc.) from our sample and compare it to a **critical value** from the corresponding theoretical distribution.

Steps:

1. Choose a significance level **α** (usually 0.05).
2. Look up the **critical value** from the reference distribution (e.g., $Z_{0.05} = -1.645$ for left-tailed test).
3. Compute the test statistic.
4. **Decision rule**:
   - If the test statistic falls into the **rejection region** (beyond the critical value), reject H_0.
   - Otherwise, fail to reject H_0.

Example:
If $Z = -2.3$ and the critical value $Z\alpha = -1.645$ for $\alpha=0.05$ (left-tailed), since $-2.3 < -1.645$, we reject H_0.

P-value Method

Idea:
Instead of comparing to a fixed critical value, we compute the **probability of observing a test statistic as extreme or more extreme than the one observed, under H_0**.

Steps:

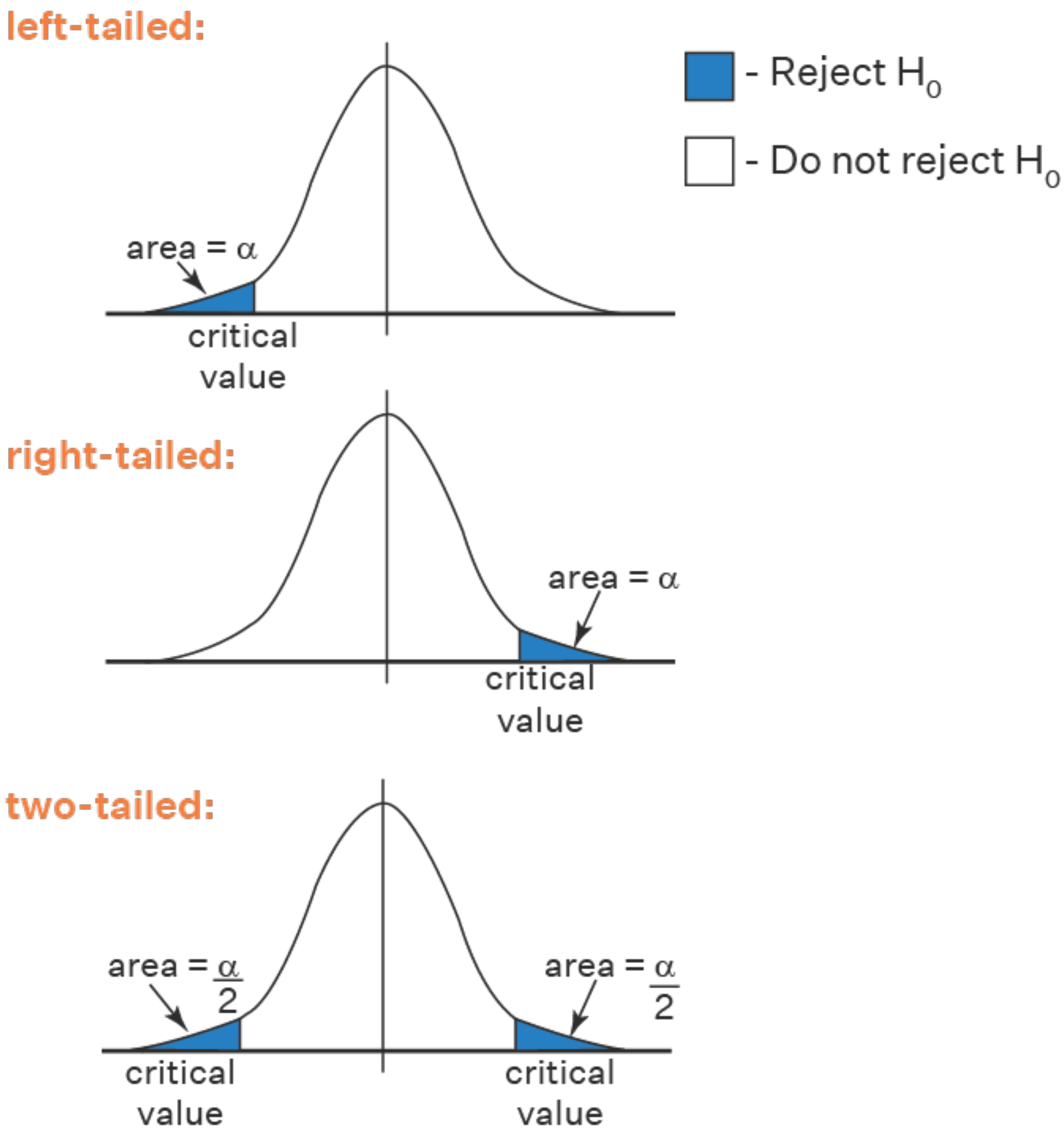
1. Compute the test statistic.
2. Find the **p-value**: the probability of getting a test statistic at least as extreme as observed, assuming H_0 is true.
   - Left-tailed: $p = P(T \leq t_{obs})$
   - Right-tailed: $p = P(T \geq t_{obs})$
   - Two-tailed: $p = 2\times P(T \geq |t_{obs}|)$
3. **Decision rule**:
   - If $\mathbf{p < \alpha}$, reject H_0.
   - If $\mathbf{p \geq \alpha}$, fail to reject H_0.

Example:
If $t = -2.3$ with $df=29$, then $p \approx 0.014$. Since $0.014 < 0.05$, reject H_0.

Key Insight:

- **Test Statistic Method** is older (tables, no computers).
- **P-value Method** is modern and more common (stat software computes p).
- In interviews:
  - If asked for decision-making: **p-value method** is preferred.
  - If asked for critical regions: mention **test statistic method**.

Example of Acceptance and Rejection Regions by the nature of tests:



One-Sample Z-Test (σ Known)

1. Purpose

The **Z-test** helps us check if a **sample mean** is "too far away" from a hypothesized **population mean** (μ_0).

- Every sample will have some random fluctuation.
- The key question is:
  > "Is the difference between our sample mean ({x}) and μ_0 small enough to be explained by chance, or is it so large that it suggests a real effect?"

So the Z-test is essentially asking:
- *"Is my sample mean just a noisy estimate of μ_0?"*
- *"Or is it statistically distant enough that I should doubt μ_0?"*

---

2. Hypotheses

- **Null hypothesis (H_0):** $\mu = \mu_0$
  > Assumes the population really has mean μ_0, and any observed difference is just random sampling error.

- **Alternative hypothesis (H_1):**

  - Left-tailed: $\mu < \mu_0$ (sample mean is significantly lower)
  - Right-tailed: $\mu > \mu_0$ (sample mean is significantly higher)
  - Two-tailed: $\mu \neq \mu_0$ (sample mean is significantly different, in either direction)

---

3. The Z Statistic

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$\bar{x}$ = sample mean
μ = population mean
σ =population standard deviation
n =sample size

Interpretation:
- Numerator = observed difference (how far sample mean is from μ_0).
- Denominator = expected fluctuation if H_0 is true.
- So Z = "How many SEMs away is my sample mean from μ_0?"

If Z is large (positive or negative), it means the sample mean is **too distant to be explained by chance alone**.

---

4. Decision Methods

A. Test Statistic (Critical Value) Method

- Compare Z to cutoff values from the **standard normal distribution**.
- For $\alpha = 0.05$:
  - Left-tailed → reject if $Z < -1.645$
  - Right-tailed → reject if $Z > +1.645$
  - Two-tailed → reject if $|Z| > 1.96$
- Critical region = "the zone so extreme we wouldn't expect it under H_0."

B. P-value Method

- Compute the exact probability of getting a test statistic as extreme as observed, under H_0.
- If this probability (p-value) < $\alpha \rightarrow$ Reject H_0.
- Benefit: gives not just a Yes/No, but a measure of "how strong" the evidence is.

---

5. Example: Hiring Bias in LLMs

- Global benchmark: $\mu_0 = 3.0$, $\sigma = 0.9$
- Sample: n = 400 Indian candidates, {x} = 2.92
- Question: Are Indian candidates being scored lower?

Step 1. Compute Z:

$$Z = \frac{2.92 - 3.0}{0.9/\sqrt{400}} = \frac{-0.08}{0.045} \approx -1.78$$

Step 2. Critical Value Method:
- At $\alpha=0.05$ (left-tailed), $Z_{crit} = -1.645$
- Since $-1.78 < -1.645 \rightarrow$ Reject H_0.

Step 3. P-value Method:
- $p = P(Z \leq -1.78) \approx 0.0375$
- Since $p < 0.05 \rightarrow$ Reject H_0.

Conclusion: There is statistically significant evidence that Indian candidates are being scored lower than the global benchmark.

---

6. Fine-Grain Insights

- The **Z-statistic is a signal-to-noise ratio**:
  - Numerator = signal (difference between sample and benchmark).
  - Denominator = noise (expected sampling fluctuation).
- If the difference is less than 1 SEM → not worth worrying.
- If it's 2–3 SEMs away → unlikely by chance → strong evidence against H_0.
- Larger sample size (n ↑) → SEM ↓ → test becomes more sensitive (small deviations become detectable).

---

7. Assumptions

- Population standard deviation (σ) is known.
- Observations are independent and random.
- Sample size is large ($n \geq 30$) or data are from a normal population.

---

8. Limitations

- In practice, σ is almost never known.
- For small n or unknown σ → use **t-test** instead.
- Statistically significant ≠ Practically significant. Always interpret effect size.

---

9. Common Interview Questions

- **When to use Z vs t?**
  - Z if σ known (rare), or sample size huge and σ well-estimated.
  - t if σ unknown (typical real-world case).
- **What does the p-value mean?**
  - Probability of getting results at least as extreme as observed, assuming H_0 is true.
- **Why divide by $\sigma/\sqrt{n}$?**
  - To scale difference into units of "expected sampling variability."
- **If sample size is huge, won't Z always reject H_0?**
  - Yes, even trivial differences become significant. That's why we check **effect size.**
- **Can Z-test apply to proportions?**
  - Yes, one-sample Z-test is also used for testing a proportion p against p_0 (via normal approximation).

---

10. Business Intuition

The Z-test is a **benchmark check**.
- If global data tell us σ exactly, then Z quantifies whether our group's mean is unusually far from the benchmark.
- It answers: *"Is my sample just normal noise, or do I need to rethink my assumption that my population matches the benchmark?"*

One-Sample t-Test (σ Unknown)

1. Purpose

The one-sample t-test evaluates whether the sample mean $\bar{x}$ is statistically different from a hypothesized population mean μ_0 when the **population standard deviation σ is unknown**.

We substitute σ with the sample standard deviation s, and account for the added variability using the **t-distribution**.

2. Hypotheses

- Null hypothesis: $H_0\!: \ \mu = \mu_0$
- Alternative hypothesis (choose one):
  - Left-tailed: $H_1\!: \ \mu < \mu_0$
  - Right-tailed: $H_1\!: \ \mu > \mu_0$
  - Two-tailed: $H_1\!: \ \mu \neq \mu_0$

3. Test Statistic

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

x̄ = obersved mean of the sample
μ = assumed mean
s = standard deviation
n = sample size

Distribution under H_0:

$$t \sim t_{n-1}$$

4. Decision Procedures

A) Critical-Value Method

1. Fix significance level α.
2. Find the critical value $t_{\alpha,df}$ from the t-distribution with $df = n - 1$.
3. Reject H_0 if the observed t lies in the rejection region:
   a. Left-tailed: $t \leq t_{\alpha,df}$
   b. Right-tailed: $t \geq t_{1-\alpha,df}$
   c. Two-tailed: $|t| \geq t_{1-\alpha/2,df}$

B) P-value Method

- Compute observed t.
- Compute p-value using t_{n-1} distribution:
  - Left-tailed: $p = P(T \leq t_{\text{obs}})$
  - Right-tailed: $p = P(T \geq t_{\text{obs}})$
  - Two-tailed: $p = 2P(T \geq |t_{\text{obs}}|)$

- Reject H_0 if $p < \alpha$.

5. Worked Example

Suppose we test whether the average productivity score of a new AI tool equals 75.

- Hypothesized mean: $\mu_0 = 75$
- Sample size: $n = 20$
- Sample mean: $\bar{x} = 72.5$
- Sample standard deviation: $s = 6$
- Significance: $\alpha = 0.05$, left-tailed test

Step 1. Compute test statistic

$$t = \frac{72.5 - 75}{6/\sqrt{20}} = \frac{-2.5}{1.3416} \approx -1.86$$

Step 2. Critical-Value Method
- Degrees of freedom: $df = 19$
- From t-table: $t_{0.05,19} \approx -1.729$
- Since $-1.86 < -1.729$, reject H_0.

Step 3. P-value Method
- $p = P(T_{19} \leq -1.86) \approx 0.039$
- Since $p < 0.05$, reject H_0.

Conclusion: The AI tool's mean productivity score is significantly lower than 75.

6. Insights

- The t-distribution has heavier tails than the normal, reflecting added uncertainty from estimating σ.
- As $n \to \infty$, t_{n-1} converges to $\mathcal{N}(0,1)$.
- With small n, critical values are larger, making it harder to reject H_0.

7. Assumptions

- Random, independent sample.
- Population distribution is normal (important when n is small).
- For large n, the test is robust to non-normality (CLT).

8. Limitations

- Sensitive to outliers in small samples.
- Cannot handle highly skewed data unless n is large.

9. Common Interview Questions

- When t vs Z?
  - Z if σ known (rare), or n huge.
  - t if σ unknown (most real-world cases).
- Why does t have heavier tails?
  - To reflect extra variability from using s instead of σ.
- Why $df = n - 1$?
  - One degree of freedom lost when estimating $\bar{x}$.
- What happens as n grows?
  - t-distribution approaches the normal.
- Is the t-test robust?

- Yes, for moderate-to-large n, but less so for tiny samples with skewness.

10. Business Intuition

The t-test is the “realistic” version of the Z-test.
- *“Given only my sample’s variability, do I have enough evidence to say my population mean differs from the benchmark?”*
- Small n → need strong evidence.
- Large n → easier to detect small differences.

One-Sample Chi-Square Test for Variance

1. Purpose

The chi-square (χ^2) test for variance evaluates whether the **population variance** σ^2 equals a hypothesized value σ_0^2.
It is used when we want to test **consistency of variability** against a benchmark.

Examples:
- Are test scores more variable in one country compared to the global benchmark?
- Is the manufacturing process variance within acceptable tolerance?

---

2. Hypotheses

- Null hypothesis:
  [H_0: $\sigma^2 = \sigma_0^2$]
- Alternative hypothesis (choose one):
  - Left-tailed: H_1: $\sigma^2 < \sigma_0^2$
  - Right-tailed: H_1: $\sigma^2 > \sigma_0^2$
  - Two-tailed: H_1: $\sigma^2 \neq \sigma_0^2$

---

3. Test Statistic

If $X_1, \ldots, X_n$ is a random sample from a normal distribution, then

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

where
- s^2 = sample variance
- n = sample size
- $df = n - 1$ degrees of freedom

This comes from the result that $(n-1)s^2/\sigma^2$ follows a chi-square distribution when sampling from a normal population.

---

4. Decision Procedures

A) Critical-Value Method

- Choose significance level α.
- Find chi-square critical value(s) from χ^2_{n-1} distribution.
- Decision rules:
  - Left-tailed: reject H_0 if $\chi^2 \le \chi^2_{\alpha,n-1}$
  - Right-tailed: reject H_0 if $\chi^2 \ge \chi^2_{1-\alpha,n-1}$
  - Two-tailed: reject H_0 if χ^2 lies outside interval

$$\chi^2 \notin \left[\chi^2_{\alpha/2,n-1},\ \chi^2_{1-\alpha/2,n-1}\right].$$

B) P-value Method

- Left-tailed: $p = P(\chi^2 \le \chi^2_{\text{obs}})$
- Right-tailed: $p = P(\chi^2 \ge \chi^2_{\text{obs}})$
- Two-tailed: $p = 2 \times \min\{P(\chi^2 \le \chi^2_{\text{obs}}),\ P(\chi^2 \ge \chi^2_{\text{obs}})\}$
- Reject H_0 if $p < \alpha$.

---

5. Worked Example

Suppose a manufacturer claims the process variance is $\sigma_0^2 = 16$ (i.e., $\sigma_0 = 4$).
We sample $n = 25$ items and compute sample variance $s^2 = 25$.
We test $H_0: \sigma^2 = 16$ against $H_1: \sigma^2 > 16$ at $\alpha = 0.05$.

Step 1. Compute statistic

$$\chi^2 = \frac{(25-1)\cdot 25}{16} = \frac{600}{16} = 37.5$$

Step 2. Critical value method
- Degrees of freedom $df = 24$
- Critical value $\chi^2_{0.95,24} \approx 36.42$
- Since $37.5 > 36.42$, reject H_0.

Step 3. P-value method
- $p = P(\chi^2_{24} \ge 37.5) \approx 0.034$
- Since $p < 0.05$, reject H_0.

Conclusion: The process variance is significantly larger than claimed.

---

6. Insights

- This test is **sensitive to normality**. If data are not normal, results may be misleading.
- The chi-square distribution is **skewed right**, especially for small n.
- As n grows, it becomes more symmetric.

---

7. Assumptions

- Random independent sample.
- Population distribution is normal.
- Hypothesized variance σ_0^2 is specified in advance.

---

8. Limitations

- Not robust to deviations from normality.
- Sensitive to outliers, since variance inflates with extreme values.
- Only tests variance, not mean.

---

9. Common Interview Questions

- What does the chi-square test for variance check?
  - Whether the population variance differs from a specified benchmark.
- Why $df = n - 1$?
  - Because sample variance uses $\bar{x}$, costing one degree of freedom.
- Is this test robust?
  - No, it requires normality assumption.
- What if I want to compare two variances?
  - Use an F-test.

---

10. Business Intuition

This test addresses **stability of processes**:
- *"Is my process variability within the promised range?"*
- Useful in **quality control, risk analysis, and benchmarking consistency**.

F-Test for Comparing Two Population Variances

1. Purpose

The **F-test** evaluates whether two populations have the same variance.
It is commonly used in:
- Testing equality of variability between two groups.
- As a building block for **ANOVA**.

It directly extends the chi-square variance test:
- Recall that for one sample,

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

- If we have two independent samples, each with its own chi-square distribution, their **ratio** follows an **F distribution**.

---

2. Hypotheses

Let σ_1^2 and σ_2^2 be the population variances of groups 1 and 2.

- **Null hypothesis:**
  [H_0: $\sigma_1^2 = \sigma_2^2$]
- **Alternative hypothesis:**
  - Two-tailed: H_1: $\sigma_1^2 \neq \sigma_2^2$
  - One-tailed: H_1: $\sigma_1^2 > \sigma_2^2$ or H_1: $\sigma_1^2 < \sigma_2^2$

---

3. Test Statistic

For samples of sizes n_1 and n_2, sample variances s_1^2 and s_2^2:

$$F = \frac{s_1^2}{s_2^2}$$

where we usually arrange so that $s_1^2 \geq s_2^2$, making $F \geq 1$ (for a two-tailed test).

with degrees of freedom
- $df_1 = n_1 - 1$ (numerator)
- $df_2 = n_2 - 1$ (denominator)

Derivation:
- $(n_1 - 1)\, s_1^2/\sigma_1^2 \sim \chi^2_{n_1 - 1}$
- $(n_2 - 1)\, s_2^2/\sigma_2^2 \sim \chi^2_{n_2 - 1}$
- If $\sigma_1^2 = \sigma_2^2$, then their ratio:

$$F = \frac{\big((n_1 - 1)s_1^2/\sigma^2\big)/(n_1 - 1)}{\big((n_2 - 1)s_2^2/\sigma^2\big)/(n_2 - 1)} \sim F_{n_1 - 1,\, n_2 - 1}.$$

---

4. Decision Procedures

A) Critical-Value Method

- Fix α.
- Obtain critical value(s) F_{α, df_1, df_2} from F-distribution.
- Decision:
  - Right-tailed: reject H_0 if $F \geq F_{1-\alpha, df_1, df_2}$
  - Left-tailed: reject H_0 if $F \leq F_{\alpha, df_1, df_2}$
  - Two-tailed: reject H_0 if F is either too large or too small:

$$F \notin \left[F_{\alpha/2, df_1, df_2},\ F_{1-\alpha/2, df_1, df_2}\right].$$

B) P-value Method

- Compute p from F-distribution with (df_1, df_2).
- Reject H_0 if $p < \alpha$.

---

5. Worked Example

Two machines produce bolts. We test if their variances differ.

- Sample 1: $n_1 = 16, s_1^2 = 0.012$
- Sample 2: $n_2 = 21, s_2^2 = 0.008$
- Hypotheses: $H_0: \sigma_1^2 = \sigma_2^2, H_1: \sigma_1^2 \neq \sigma_2^2$

Step 1. Compute statistic

$$F = \frac{0.012}{0.008} = 1.5$$

Step 2. Critical-value method
- $df_1 = 15, df_2 = 20$
- For $\alpha = 0.05$ two-tailed:
- $F_{0.975,15,20} \approx 3.01$
- $F_{0.025,15,20} \approx 1/3.01 = 0.332$
- Since 1.5 is within $[0.332, 3.01]$, fail to reject H_0.

Step 3. P-value method
- $p = P\left(F_{15,20} \geq 1.5\right) \approx 0.24$ (two-tailed ≈ 0.48)
- $p > 0.05$, fail to reject H_0.

Conclusion: No significant evidence of variance difference between the machines.

---

6. Insights

- The F-distribution is skewed right, especially for small df.
- As df grow, the F-distribution approaches 1 (becomes more symmetric).
- F is always nonnegative, since it is a ratio of variances.

---

7. Assumptions

- Both samples are random and independent.
- Populations are normally distributed.
- Variances tested are for continuous data.

---

8. Limitations

- Highly sensitive to non-normality.
- Sensitive to outliers (variance inflates).
- Alternative robust tests: Levene's test, Bartlett's test, Brown-Forsythe.

---

9. Common Interview Questions

- How is F-test derived from chi-square?
  - Ratio of two scaled chi-square variables forms an F distribution.
- Why two degrees of freedom?
  - One from each sample variance estimation.
- When to use F-test?
  - To compare variances, or as part of ANOVA.
- Is it robust?
  - Not very; assumes normality.

---

10. Business Intuition

The F-test checks **consistency between two processes**:
- *"Is one process more variable than the other?"*
- Used in **quality control**, **finance (risk comparison)**, and as the **foundation of ANOVA**.