What is hypothesis testing:

↳ A hypothesis is an assumption or statement about a population that we can test using sample data.

Example,

↳ A company claims the average salary of their employees is 500,000 Rs per year.

↳ We don't know the true average salary, so we want to test whether the claim is reasonable based on our sample data.

Null hypothesis (H_0):

↳ The statement of no effect or no difference.
↳ We assume it is true unless proven otherwise.

Alternative hypothesis (H_A):

↳ The statement that contradicts the null hypothesis.
↳ It is the outcome we might observe if the null hypothesis is false.

One-sided alternative hypothesis: $\mu < 50$ or $\mu > 50$

Two-sided alternative hypothesis: $\mu \neq 50$

It is important to note that hypotheses are always about the population and not about the sample.

Type I & Type II Errors:

↳ When testing hypothesis, all we see is a random sample

↳ Therefore, our decision to accept or reject H_0 may still be wrong, maybe due to sampling error.

	Result of the test	
	Reject H_0	**Accept H_0**
H_0 **is true**	*Type I error*	correct
H_0 **is false**	correct	*Type II error*

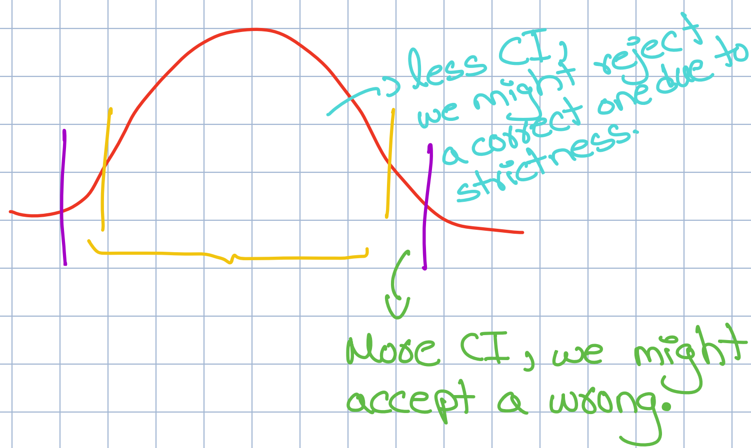
- This means that when we test a hypothesis, we are working with limited data — a single random sample taken from the population. This sample is just one of many possible samples we could have drawn, and it might not fully represent the true characteristics of the entire population.

- For example, imagine flipping a coin 10 times. If you get 7 heads and 3 tails in your sample, you might incorrectly conclude the coin is biased, even though the coin might actually be fair. This happens because the small sample doesn't perfectly represent the true behavior of the coin.

- In hypothesis testing, this randomness in the sample can lead us to make wrong decisions:
    - **Type I Error (False Positive):** Rejecting the null hypothesis H0 when it is actually true.
    - **Type II Error (False Negative):** Failing to reject H0 when it is actually false.

↳ Each error occurs with a certain probability that we hope to keep small. A good test results in a wrong decision only if observed data is somewhat extreme.

↳ Type I errors are more undesired, so we set a small α to minimize them.



Tradeoff between Type I & Type II.

Increasing the confidence level (CI) reduces the Type I error. Here's why:

- **What is Confidence Level and Type I Error?**
    - **Confidence Level (CI):** The percentage of times we expect the true population parameter to fall within our confidence interval if we repeated the sampling many times. For example, a 95% CI means we expect the true parameter to be in the interval 95 out of 100 times.
    - **Type I Error**: The probability of rejecting the null hypothesis H0 when it is actually true. This is typically denoted by alpha, which is also the significance level.

- **Relationship Between CI and Type I Error:**
    - The confidence level 1 - alpha and the Type I error rate alpha are directly linked. For instance:
    - A 95% CI corresponds to alpha = 0.05 (Type I error rate of 5%).
    - A 99% CI corresponds to alpha = 0.01 (Type I error rate of 1%).

    - When you increase the confidence level (e.g., from 95% to 99%), you are making the interval wider to ensure it captures the true population parameter more often. This reduces the risk of falsely rejecting H0, thereby decreasing the Type I error rate alpha.

- **Key Tradeoff:**
  - While increasing the CI reduces Type I error, it can also increase the Type II error because the wider interval might make it harder to detect true differences or effects

The P-value:

- The p-value helps us decide if the results we see in our data are surprising or not, assuming the null hypothesis H0 is true. The null hypothesis is like the "default" assumption that there's no effect, no difference, or nothing special happening.

- **If the p-value is small:**
  - This means our data is unusual or surprising if H0 is true. For example, if you expected a coin to be fair H0, but you flipped it 10 times and got 9 heads, that's pretty surprising. A small p-value suggests the null hypothesis might be wrong, so we consider rejecting it.

- **If the p-value is large:**
  - This means our data looks pretty normal and doesn't contradict H0. For example, flipping a coin 10 times and getting 5 heads isn't surprising if the coin is fair. A large p-value means we don't have enough evidence to reject H0, so we stick with it.

- **In short:**
  - Small p-value = "This result is surprising if H0 is true. Maybe H0 is wrong."
  - Large p-value = "This result isn't surprising if H0 is true. H0 seems fine for now."

- The p-value tells us how likely it is to get your observed result by random chance if the H_0 is true. If the probability is less than alpha, we reject it because the probability of it happening by chance is very small, and it might be something significant.

Steps:

1. Define your hypothesis
2. Select α
3. Perform experiment
4. Choose a test & perform it
5. Find the p-value OR compare the test result to critical value.

6. Make a decision.

Example,

$\mu = 3$, std_err $= 0.827$

For $\overline{X} = 33$,

$$z = \frac{33-35}{0.827}$$

$$= -2.41$$

$Z_\alpha = -2.41$

$\alpha = 0.00798$

For $\overline{X} = 37$

$$Z = \frac{37-35}{0.827}$$

$$= 2.41$$

$Z\alpha = 2.41$

0.992

$$P(33 \leq X \leq 37) = P(X \geq 37) - P(X \leq 33)$$

$$= 0.992 - 0.00798$$

$$= 0.9844$$

Example,

$\mu = 27.8$ $N = 48$

stderr $= 0.7$ $\overline{x} = 29$

$$z = \frac{29 - 27.8}{0.7}$$

$$= 1.71$$

$Z_{1.71} = 0.95637$

$$P(x > 29) = 1 - 0.95637$$

$$= 0.0436$$

Chi-test:

↳ Z & t-tests are for comparing sample & population mean.

↳ But for categorical data, we use Chi-test.

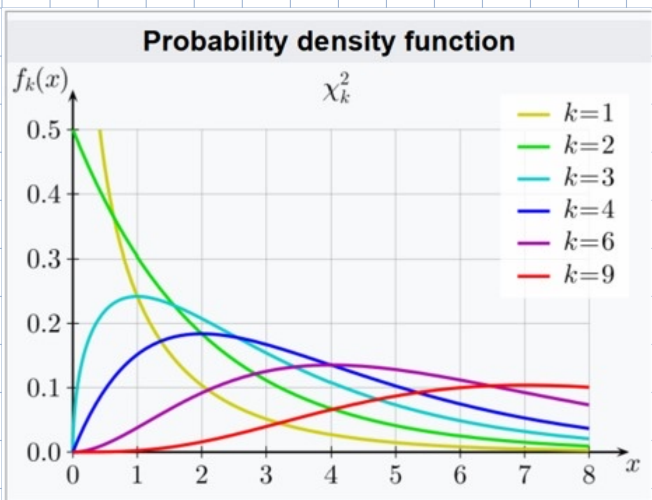
1. Whether gender is related to voting preference?

2. Whether smoking is related to lung cancer?

↳ If a random variable Z has the standard normal distribution, then Z^2 has the χ^2 distribution with one df.

$$\chi^2 = Z_1^2 + Z_2^2 + \dots + Z_k^2$$
$$= \sum_{i=1}^{k} Z_i^2$$

↳ k is degrees of freedom here.



1. Goodness of fit

To see if an observed frequency distribution fits a specific expected distribution. Also known as goodness of fit. We can do it using C statistic.

$$C = \sum_{i=1}^{m} \frac{[f_o(\epsilon_i) - f_t(\epsilon_i)]^2}{f_t(\epsilon_i)}$$

We use one tail test for goodness of fit as we are looking for large deviations.

$C \le \chi^2_\alpha \rightarrow$ fail to reject H_0

$C > \chi^2_\alpha \rightarrow$ reject H_0

Think of it as a test of how "different" the observed data is from what the model predicts:

- **A small C says:** "The differences are small enough to consider random; the model is likely correct."
- **A large C says:** "The differences are too large to be random; the model likely does not represent the data."

Example,

Suppose that after losing a large amount of money, an unlucky gambler questions whether the game was fair and the die was really unbiased.

The last 90 tosses of this die gave the following results

Number of dots on the die	1	2	3	4	5	6
Number of times it occurred	20	15	12	17	9	17

H_0: Die is fair $P = 1/6$

H_A: Die is NOT fair $P \neq 1/6$

If fair,

$$\frac{90}{6} = 15$$

$$C = \frac{(20-15)^2 + (15-15)^2 + (12-15)^2 + (17-15)^2 + (9-15)^2 + (17-15)^2}{15}$$

$$= 5.2$$

$$\chi_{0.05,5} = 11.07$$

$$5.2 \leq 11.07$$

$\hookrightarrow$ We fail to reject H_0.

2. Independence Analysis:

To test whether two categorical variables are independent or if there is some association between them.

↳ Degrees of freedom for the r by c table:

$$df = (r-1)\times(c-1) \quad , \; r>1 \;\&\; c>1$$

Example,

Given the two way table, test whether the column and row are independent.

Observed:

	Boy	Girl	Total
Grades	117	130	247
Popular	50	91	141
Sports	60	30	90
Total	227	251	478

$$f_e = \frac{\text{Row total} \times \text{column total}}{\text{Total}}$$

Expected:

	Boys	Girls
Grades	$\frac{247\times 227}{478} = 117.29$	$\frac{247\times 251}{478} = 129.7$
popular	$\frac{141\times 227}{478} = 66.96$	$\frac{141\times 251}{478} = 74.03$
Sports	$\frac{90\times 227}{478} = 42.74$	$\frac{90\times 251}{478} = 47.25$

For boys,

$$C_b = \frac{(117-117.29)^2}{117.29} + \frac{(50-66.96)^2}{66.96} + \frac{(60-42.74)^2}{42.74}$$

$$= 11.14$$

For girls,

$$C_g = \frac{(130-129.7)^2}{129.7} + \frac{(91-74.03)^2}{74.03} + \frac{(30-47.25)^2}{47.25}$$

$$= 10.19$$

$$C = 10.9 + 11.14$$

$$= 21.33$$

$$df = (3-1) \times (2-1)$$
$$= 2$$

$$\chi^2_{0.05,2} = 5.99$$

$$21.33 > 5.99$$

↳ We reject H_0

3. Variance Estimator:

↳ For large samples we can,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

But for small samples, we use χ^2 distribution.

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{x_i - x}{\sigma}\right)^2$$

↳ For confidence interval;

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

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

Example,

Let we have a measurement device giving the data (n=6 measurements) as found below. Construct 90% confidence interval for the standard deviation.

2.5, 7.4, 8.0, 4.5, 7.4, and 9.2.

$$\bar{x} = \frac{2.5 + 7.4 + 8 + 4.5 + 7.4 + 9.2}{6}$$

$$= 6.5$$

$$s^2 = \frac{1}{6-1} \times \left[(2.5 - 6.5)^2 + \cdots + (9.2 - 6.5)^2 \right]$$

$$= 6.232$$

$$\chi^2_{0.05, 5} = 11.07 \qquad \chi^2_{0.95, 5} = 1.15$$

$$\frac{(6-1) \times 6.232}{11.07} \leq \sigma^2 \leq \frac{(6-1) \times 6.232}{1.15}$$

$$2.81 \leq \sigma^2 \leq 27.09$$

$$1.67 \leq \sigma \leq 5.2$$