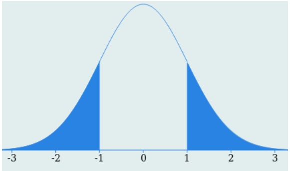
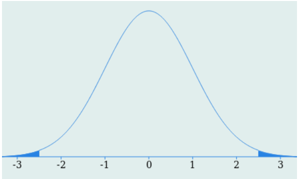
Hypothesis testing

- Hypothesis tests are techniques for making rational decisions in the context of incomplete information or uncertainty.
- It gives methods to decide whether claims about the behavioural effect(s) of one (or more) independent variable(s) can be believed.
- A hypothesis or claim is a precise statement about some population parameter(s) - like mean, median, proportion, difference of means, difference of proportions etc.
- Examples:
  - Average height of the IITK population is 155cm.
  - The amount of caffeine in a cup of coffee (150ml) will reduce the time students sleep in a class by at least 10 mins.
- Typical hypotheses: $\mu = \mu_0$, $\mu_1 = \mu_2$, $\mu_1 = \mu_2 = \mu_3$, $\pi = 0.4$, $\pi_1 - \pi_2 = 0$, $\rho_1 - \rho_2 = 0$ etc.

Logic of hypothesis testing

- How can we say that the independent variable(s) had an effect and it was not a matter of chance (or other factors)?
- If the probability that it is by chance is extremely small then one would tend to believe that the independent variable had an effect.





- What is extremely small is subjective. Typically, it is 5% or less or 1% or less. In behavioural studies the convention is 5%.

Null, alternate hypotheses I

- Every hypothesis testing problem has two competing hypotheses - the **null** hypothesis and the **alternate** hypothesis.
- The **null hypothesis** is the claim that there is no effect or difference. This is usually written H_0.
- The **alternate hypothesis** H_a (some write H_1) is the negation of the null hypothesis and is the hypothesis the experimenter wishes to establish.
- H_a is viable or is accepted if H_0 can be shown to be not viable or rejected.
- The way this is done is to show that the sample statistic value makes H_0 extremely unlikely.

Null, alternate hypotheses II

- Note that we avoid saying that H_a is true and H_0 is false. All that can be said is that H_a is highly probable and H_0 is extremely improbable.

Null and alternate hypotheses

- H_0 is usually a statement that some population parameter, θ has a specific value.
- The H_a is the opposite/ different from the null hypothesis. It can have four different forms. If H_0 is $\theta = \theta_0$:
  - H_a $\theta \neq \theta$. Two sided alternative. Non-directional.
  - H_a $\theta < \theta_0$. One sided alternative or left alternative.
  - H_a $\theta > \theta_0$. One sided alternative or right alternative.
  - H_a $\theta = \theta_1$., $\theta_1 \neq \theta_0$. This is needed for calculating the **power** of the test.

P value, significance level, statistical significance

- Assuming H_0 is true the probability that the value of the statistic is as extreme or more extreme than the value actually observed is called the **P value** of the test. Note that P is not the probability of H_0 itself.
- The **significance level** (symbolized by α) is a pre-chosen value/level such that if $P \leq \alpha$ we will reject H_0.
- If $P \leq \alpha$ the results (or H_a) is **statistically significant** at level α. Note the close resemblance of the P-value to the CI.

Hypothesis testing works backwards I

- To show that an effect exists (is highly probable) hypothesis testing works backwards:
  a) First it assumes that the effect does not exist - null hypothesis.
  b) Before data collection the rules for deciding whether the data is consistent with the null hypothesis are frozen - fix α.
  c) If the data are not consistent with the null hypothesis ($P \leq \alpha$) then our assumption that the effect does not exist (H_0) is rejected and therefore an effect (H_a) most probably exists.
- Not finding an effect is different from there is no effect.
- Also, finding an effect is different from the effect is important/ useful/ practical.

One-sided, two-sided tests

- Most often tests are two-sided. For example $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$. For $\alpha = 0.5$ the critical/ rejection region are two tails covering probability mass of 0.025 respectively assuming a z-test and the critical value is 1.96.
- If the same is one sided where the test is one-sided i.e. $\mu_1 - \mu_2 < 0$ or $\mu_1 - \mu_2 > 0$ then the probability mass of 0.05 is only on one side and the critical value becomes 1.635 - so easier to reject the H_0.
- Many statisticians are wary of one sided tests since the effect in some sense is already known.

Statistical and practical significance

- A result may be statistically significant but practically unimportant or vice versa. An example:

  Two diets are being tested for lowering cholesterol (mg/dl). $|\bar{X}_1 - \bar{X}_2| \geq 10$ have health effects. The table below shows some possible results. A -ve difference between *means* favours diet 1, +ve diet 2.

No.	$\bar{X}_1 - \bar{X}_2$	$\sigma_{d_{sd}}$	t	P	$P < 0.05$?	95%CI	Pract. imp.
1	2	0.5	4	< 0.0001	Y	(1,3)	?
2	30	5	6	< 0.0001	Y	(20,40)	?
3	30	14	2.1	0.032	Y	(2,58)	?
4	1	1	1	0.317	N	(-1,3)	?
5	2	30	0.1	0.947	N	(-58,62)	?
6	30	16	1.9	0.061	N	(-2,62)	?

Example contd.

No.	$\bar{X}_1 - \bar{X}_2$	$\sigma_{d_{sd}}$	t	P	$P < 0.05$?	95%CI	Pract. imp.
1	2	0.5	4	< 0.0001	Y	(1,3)	N
2	30	5	6	< 0.0001	Y	(20,40)	Y
3	30	14	2.1	0.032	Y	(2,58)	?
4	1	1	1	0.317	N	(-1,3)	N
5	2	30	0.1	0.947	N	(-58,62)	?
6	30	16	1.9	0.061	N	(-2,62)	?

Type I, Type II errors

- Two kinds of errors can arise a) A true null hypothesis is rejected and b) A false null hypothesis is not rejected. The table summarizes this:

	Actual state of H_0.	
Stat. inference	H_0 True	H_0 False
Reject H_0 - False	Type I error	Correct
Do not reject H_0 - True	Correct	Type II error

- α the significance level gives the probability of type I error.
- Hyp. testing is focused on reducing α. Possibly, because of widespread use in the bio-medical domain.
- The type II error is symbolized by β.