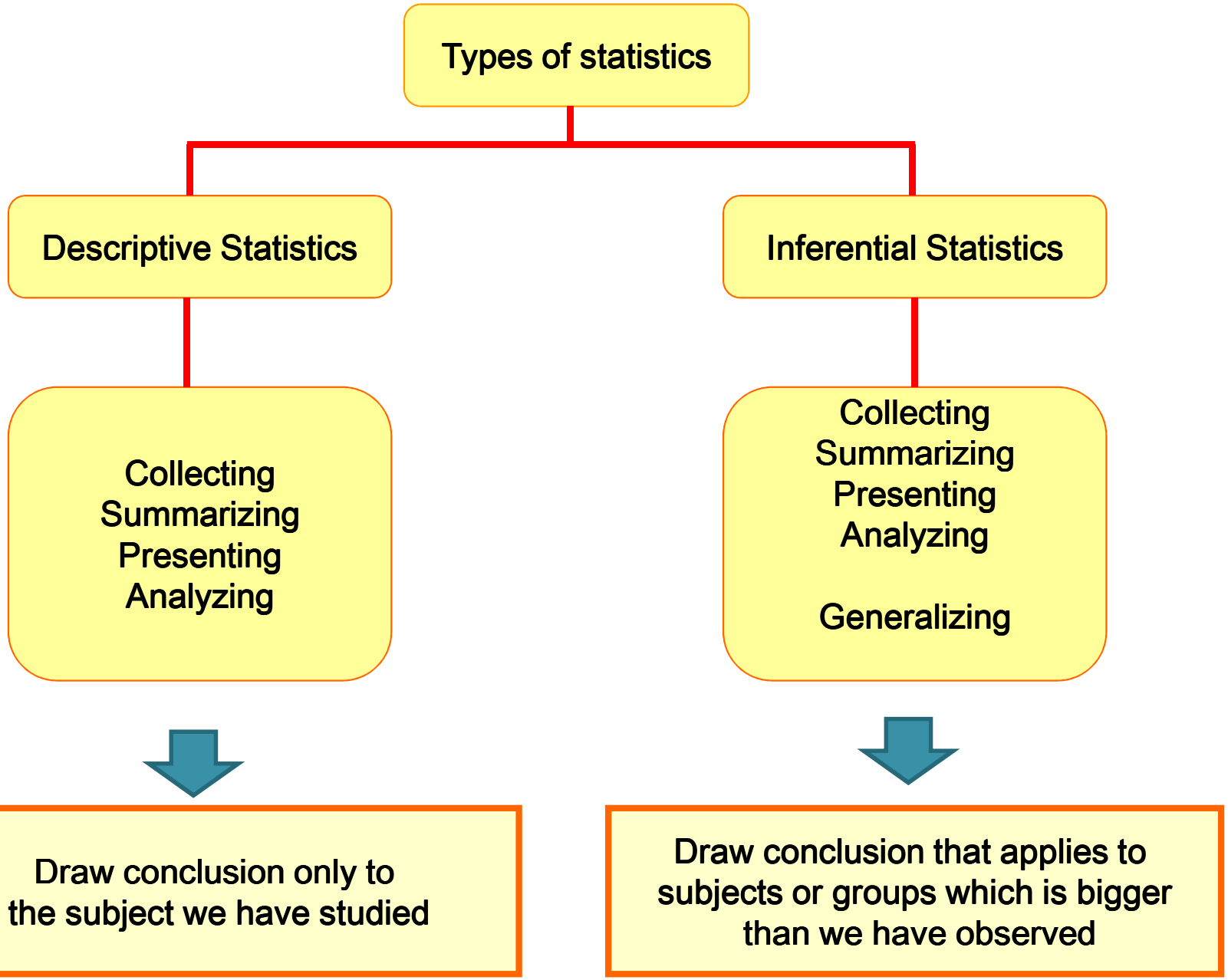
Inferential statistics

Wrap up

Pema Chophel


Types of statistics
Descriptive Statistics
Inferential Statistics
Collecting
Summarizing
Presenting
Analyzing
Collecting
Summarizing
Presenting
Analyzing
Generalizing
Draw conclusion only to the subject we have studied
Draw conclusion that applies to subjects or groups which is bigger than we have observed

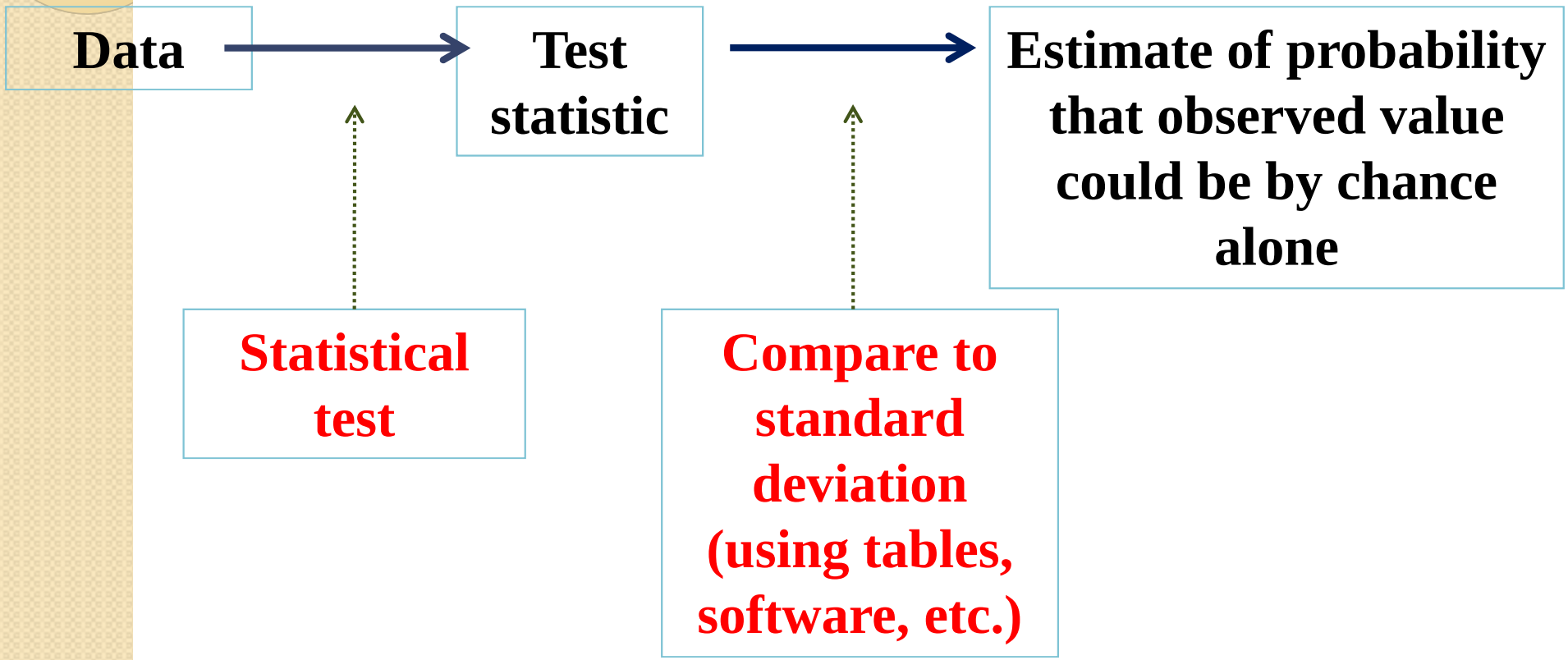
Inferential statistical methods

1. Hypothesis Testing

2. Parameter Estimation

1. **Hypothesis Testing:**
   - To ask whether an effect is present or not among different groups.

Concept in hypothesis testing



Steps to consider for a statistical test of a hypothesis

1. State the problem
2. Formulate hypotheses: one- or two-tailed
3. Choose α
4. Determine the test statistic
5. Calculate the test statistic
6. Decision and conclusion

1. State the problem

- One-sample problem
  - Hypotheses are specified about a single distribution (population)
- Two-sample problem
  - Two different distributions (populaitons) are compared

2. Formulate hypotheses

A. Null hypothesis (H_o):

- This is usually a statement of no difference

B. Alternative hypothesis (H_1 or H_a):

- This is the statement we will accept if we reject the null hypothesis

Possible scenarios

	H_0 is True	H_0 is False
Accept H_0	**$1 - \alpha$ (Correct decision)**	β **(Type II error)**
Reject H_0	α **(Type I error)**	**$1 - \beta$** = power **(Correct decision)**

- Type I error is the probability of rejecting H_0 when H_0 is true
- Type II error is the probability of accepting H_0 when H_0 is false

3. Choosing α level

- Level of significance
  - 1% (0.01), 5% (0.05), 10% (0.10)

4. Determine the test statistic

Need understanding probability distribution and asumption

- Analysis of continuous outcome
- Analysis of categorical outcome

5. Calculate the test statistic

- Most test statistics are of the form

1. One sample

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \qquad z = \frac{\overline{X} - \mu_o}{\sigma/\sqrt{n}} \qquad z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o q_o}{n}}}$$

2. Two sample

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad \boldsymbol{X}^2 = \sum_{i,j} \frac{\left(\boldsymbol{O}_{ij} - \boldsymbol{E}_{ij}\right)^2}{\boldsymbol{E}_{ij}}$$

6. Decision and conclusion

Depends on:-

1. Critical value
2. p-value

1. Critical value:- The critical value is a factor used to compute the margin of error

2. p value

- This is the probability of getting a value of the test statistic this extreme or more extreme, given H_0 is true
- Suggested terminology (conventional)
  - $p \leq 0.001$ Very highly statistically significant
  - $0.001 \leq p < 0.01$ Highly statistically significant
  - $0.01 \leq p < 0.05$ Statistically significant
  - $p \geq 0.05$ Not statistically significant

p value property

- Small difference and small sample size
  - Very large p value
- Small difference and large sample size
  - Small or large p value
- Large difference and small sample size
  - Small or large p value
- Large difference and large sample size
  - Very small p value