Significance Testing / Hypothesis Testing

Sample Statistics and Population Parameters

- Statistics' main objective is to make inferences about a population based on information contained in a sample.
- Populations are described by numerical summaries
  - A **parameter** describes a **population**.
    - Quantitative variables
      - population mean μ
      - Standard deviation, σ
    - Categorical variables
      - proportion, ρ
  - A **statistic** describes a **sample**.
    - sample mean x and
    - Standard deviation s.
    - proportion, p

College of Information Technology and Computer Science

Making statistical inferences

1. Estimation
   - "What is the value of the population parameter?"
   - Confidence interval for estimating a parameter
2. Significance hypothesis testing
   - "Does the population parameter satisfy a specified condition?"

College of Information Technology and Computer Science

Significance hypothesis testing

- Significance testing
  - uses probability to provide a way to quantify how plausible a parameter is while controlling the chance of an incorrect inference
  - method for using data to summarize the evidence about a hypothesis.
  - Determine whether there is enough evidence in favor of certain belief, or **hypothesis**, about a parameter

College of Information Technology and Computer Science

Significance hypothesis testing

- Questions about the parameter.
  1. Is there statistical evidence, from a random sample of potential customers, to support the hypothesis that more than 10% of the potential customers will purchase a new product when seen in Facebook?
  2. Is a new drug effective in curing a certain disease? A sample of patients is randomly selected. Half of them are given the drug while the other half are given a placebo. The conditions of the patients are then measured and compared.

College of Information Technology and Computer Science

The Five Steps of a Significance Test

Null: H_0
Alt: H_a / H_1

1. **Assumptions**
   First, specify the variable and parameter. The assumptions commonly pertain to the method of data production (randomization), the sample size, and the shape of the population distribution.
2. **Hypotheses**
   State the null hypothesis, H_0 (a single parameter value, usually no effect), and the alternative hypothesis, H_a (a set of alternative parameter values)
3. **Test statistic**
   The test statistic measures distance between the point estimate of the parameter and its null hypothesis value, usually by the number of standard errors between them.
4. **P-value**
   The P-value is the probability that the test statistic takes the observed value or a value more extreme if we presume H_0 is true. Smaller P-values represent stronger evidence against H_0.
5. **Conclusion**
   Report and interpret the P-value in the context of the study. Based on the P-value, make a decision about H_0 (either reject or do not reject H_0) if a decision is needed.

College of Information Technology and Computer Science

Formulation of Hypotheses

- A **hypothesis** is a statement about a population, usually claiming that a parameter takes a particular numerical value or falls in a certain range of values.
  - For a categorical variable the parameter is a proportion, ρ
  - For a quantitative variable the parameter is a mean, μ

College of Information Technology

Two types of Hypothesis

- **Null hypothesis H_0** : (=)
  - It is the hypothesis to be tested.
  - It is a statement that either a parameter takes on a particular value, or that there is no difference between parameters (i.e., there is no effect).
  - Examples
    - H_0: The probability ρ of guessing the correct hand is 50% (ρ = 50%).
    - H_0: The mean IQ of UC students is 110. (μ = 110)

College of Information Technology

Two types of Hypothesis

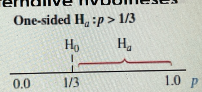
- **Alternative hypothesis H_A:**
  - It states that either the parameter falls within an alternate range of values, or that there is a difference between parameters (i.e., there is an effect).
  - It is a statement of what we believe is true if our sample data cause us to reject the null hypothesis.
  - Examples
    - H_A: The probability ρ of guessing the correct hand is larger than 50% (ρ > 50%).
    - H_A: The probability ρ of guessing the correct hand is less than 50% (ρ < 50%).
    - H_A: The mean IQ of UC students differs from 110. (μ ≠ 110).

College of Information Technology and Computer Science

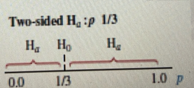
The Alternative Hypothesis

$>$ or $<$

- **Two Types of Hypotheses**
  - **One-sided alternative hypotheses**
    - H_a: $p > p_0$.



  - **Two-sided alternative hypotheses**
    - H_a: $p \neq p_0$.



College of Information Technology

H0 or Ha? For parts a and b, is the statement a null hypothesis, or an alternative hypothesis?

1) $Ha: p > 100$

1. In Canada, the proportion of adults who favor legalized gambling is 0.50. $H_0: p = 50\%$
2. The proportion of all Canadian college students who are regular smokers is less than 0.24, the value it was 10 years ago. $Ha: p < 0.24$
3. The mean IQ of all students at UC is larger than 100.
4. The mean working hours for the working population differs from 40 hours. $Ha: p \neq 40$

- Introducing notation for a parameter, state the hypotheses in parts a and b in terms of the parameter values.

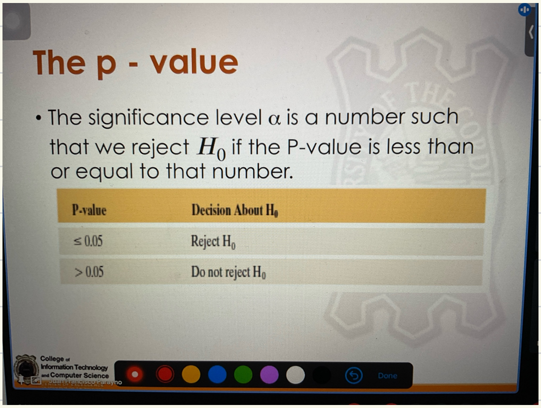
College of Information Technology

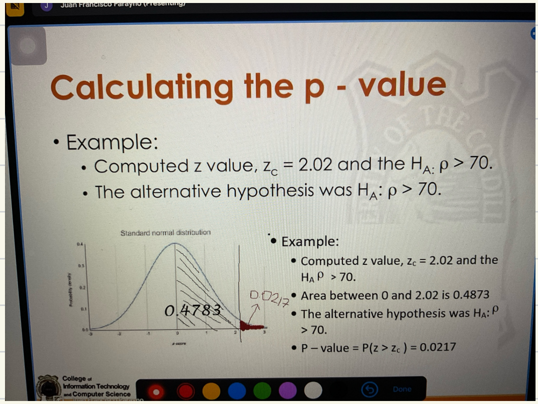
The p - value

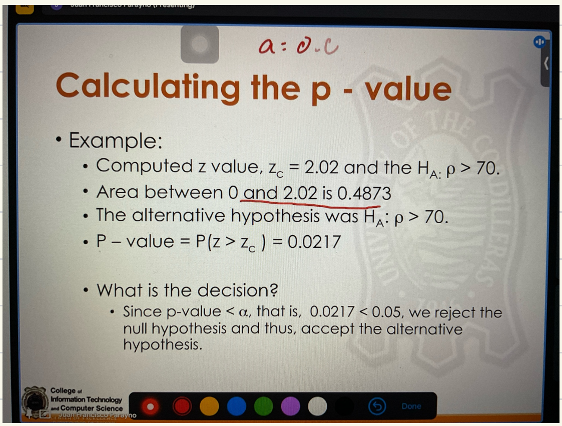
- The significance level (α) is the probability of rejecting a null hypothesis.
  - In practice, the most common significance level is 0.05.

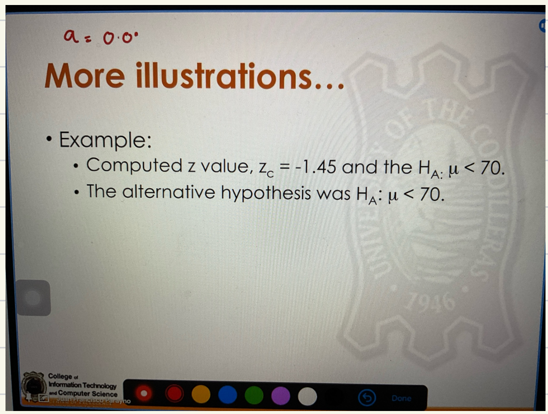
		Condition of Null Hypothesis	
		True	False
Possible Action	Fail to reject H_0	Correct action	Type II error
	Reject H_0	Type I error	Correct action

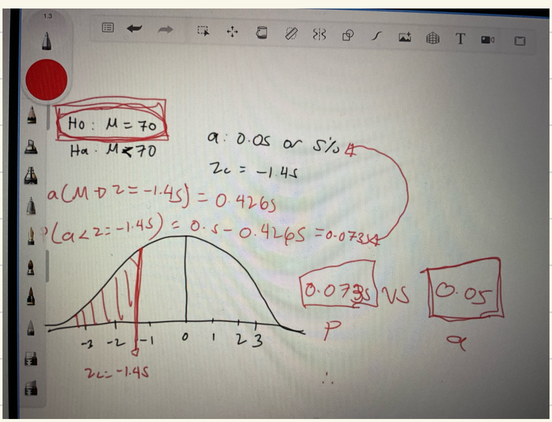
College of Information Technology and Computer Science

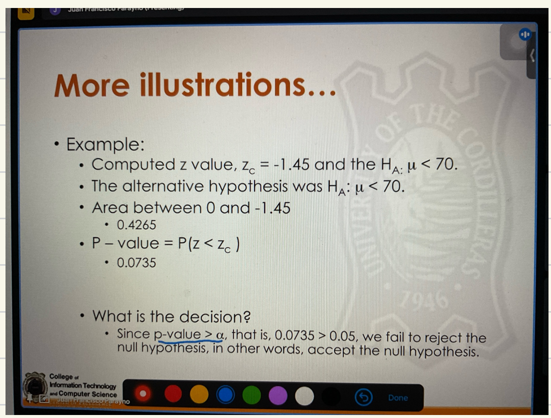

The p - value
• The significance level α is a number such that we reject H_0 if the P-value is less than or equal to that number.
P-value | Decision About H_0
≤ 0.05 | Reject H_0
> 0.05 | Do not reject H_0
College of Information Technology and Computer Science
Done


Calculating the p - value
• Example:
• Computed z value, z_c = 2.02 and the $H_{A:}$ ρ > 70.
• The alternative hypothesis was H_A: ρ > 70.
Standard normal distribution
0.4783
0.0217
• Example:
• Computed z value, z_c = 2.02 and the H_A ρ > 70.
• Area between 0 and 2.02 is 0.4873
• The alternative hypothesis was H_A: ρ > 70.
• P – value = P(z > z_c) = 0.0217
College of Information Technology and Computer Science
Done


a: 0.05
Calculating the p - value
• Example:
• Computed z value, z_c = 2.02 and the $H_{A:}$ ρ > 70.
• Area between 0 and 2.02 is 0.4873
• The alternative hypothesis was H_A: ρ > 70.
• P – value = P(z > z_c) = 0.0217
• What is the decision?
• Since p-value < α, that is, 0.0217 < 0.05, we reject the null hypothesis and thus, accept the alternative hypothesis.
College of Information Technology and Computer Science
Done


a = 0.05
More illustrations…
• Example:
• Computed z value, z_c = -1.45 and the $H_{A:}$ μ < 70.
• The alternative hypothesis was H_A: μ < 70.
College of Information Technology and Computer Science
Done

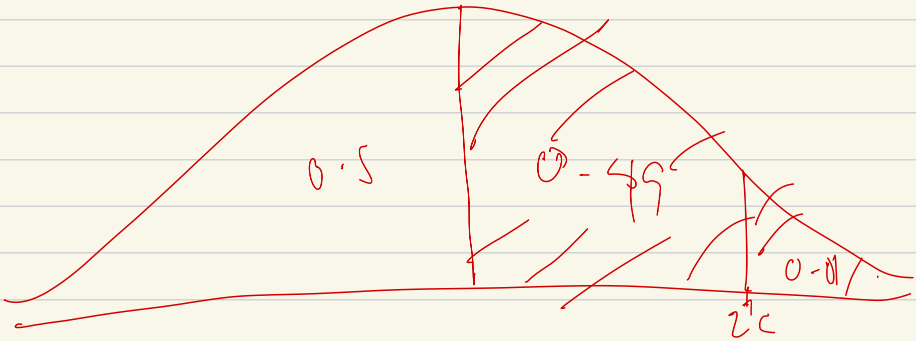

Ho : μ = 70
Ha · μ < 70
a : 0.05 or 5%
zc = -1.45
a(μ + z = -1.45) = 0.4265
P(a < z = -1.45) = 0.5 - 0.4265 = 0.0735
0.0735 vs 0.05
P
a
-3 -2 -1 0 1 2 3
zc = -1.45


More illustrations…
• Example:
• Computed z value, z_c = -1.45 and the $H_{A:}$ μ < 70.
• The alternative hypothesis was H_A: μ < 70.
• Area between 0 and -1.45
• 0.4265
• P – value = P(z < z_c)
• 0.0735
• What is the decision?
• Since p-value > α, that is, 0.0735 > 0.05, we fail to reject the null hypothesis, in other words, accept the null hypothesis.
College of Information Technology and Computer Science
Done

Right-tailed (>)

$\alpha = 0.01$ / 1%

$c = 0.99$ or 99%



$a(\mu - 1 > z_c) = 0.4$

$\to 0.3997$

$\to 1.28$

Questions: CITCS

① What is the level of awareness of the CITCS students about their carbon footprint