

# Statistics on further studies of students in IIT H

## MA4240 Applied Statistics

Shreyas Wankhede, Pradeep Mundlik, Akshitha Kola, Dhatri Reddy,  
Mouktika Cherukupalli, Avinash Malothu, Sumeeth Kumar, Sathvika  
Marri

Indian Institute of Technology Hyderabad

April 28, 2023

# Outline

- 1 Introduction
- 2 Data Visualization
- 3 Hypothesis Testing

# Introduction

This project is based on further studies of students studying at IITH. We have used statistics to deduce few conclusions from the given data, assuming that the data is random sample population.

We used sampling, more specifically volunteer sampling for collection of data through mail from students at IITH but only a few of them volunteered to respond. The data was collected from 114 students, and it is diverse with data from different years of UG.

## Variables of interest

- ① Which department?
- ② Gender?
- ③ Annual family income
- ④ CGPA?
- ⑤ Interested in further studies (if yes)
  - After how many years of work experience?
  - Which degree?( Masters/ PhD/ MBA)
  - Location?
  - In which department(Same as bachelors or different?)
- ⑥ Interested in further studies(if no)
  - Are you interested in civils?
  - Are you interested in software development?

# Data visualization

# Hypothesis Testing

## Case 1: Comparing CGPA of students who are willing to pursue higher studies with students who don't want to pursue

We assume our null hypothesis to be that average CGPA of students willing to go for higher studies is greater than those who don't want to. Let  $\alpha = 0.05$ .

$\bar{x}_1$  = sample mean of CGPA of people willing to go for higher studies

$\bar{x}_2$  = sample mean of CGPA of people who don't want to go for higher studies

$s_1^2$  = sample standard deviation of CGPA of people willing for higher studies

$s_2^2$  = sample standard deviation of CGPA of people who don't want

For Hypothesis Testing we make the following statements:

$$H_0 = \mu_1 - \mu_2 \geq 0$$

$$H_a = \mu_1 - \mu_2 < 0$$

## Case 1 continued

Information:

$$\bar{x}_1 = 8.4447 \quad (1)$$

$$\bar{x}_2 = 8.3919 \quad (2)$$

$$s_1^2 = 0.7694 \quad (3)$$

$$s_2^2 = 0.5672 \quad (4)$$

$$n_1 = 88 \quad (5)$$

$$n_2 = 26 \quad (6)$$

since  $\frac{s_1^2}{s_2^2} < 4$ , we can assume the population variances would be equal.

Degrees of freedom,

$$dof = n_1 + n_2 - 2 = 112$$

# Hypothesis testing

## Case 1 continued

The pooled variance will be:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 0.5317 \quad (7)$$

The test statistic is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{n_2 + n_1}{n_1 n_2}}} = 0.3244 \quad (8)$$

Using the rejection region approach, we reject  $H_0$  if  $t_{0.05,112} \geq -t$  where  $t_{0.05,112} = -1.6586$ . we have enough statistical evidence to reject null hypothesis since observed  $t$  is lesser than 1.6586



## Case 1 continued

Information:

$$\bar{x}_1 = 8.4447 \quad (9)$$

$$\bar{x}_2 = 8.3919 \quad (10)$$

$$s_1^2 = 0.7694 \quad (11)$$

$$s_2^2 = 0.5672 \quad (12)$$

$$n_1 = 88 \quad (13)$$

$$n_2 = 26 \quad (14)$$

since  $\frac{s_1^2}{s_2^2} < 4$ , we can assume the population variances would be equal.

Degrees of freedom,

$$dof = n_1 + n_2 - 2 = 112$$

# Hypothesis testing

## Case 1 continued

The pooled variance will be:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 0.5317 \quad (15)$$

The test statistic is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{s_p \sqrt{\frac{n_2 + n_1}{n_1 n_2}}} = 0.3244 \quad (16)$$

Using the rejection region approach, we reject  $H_0$  if  $t_{0.05,112} \geq -t$  where  $t_{0.05,112} = -1.6586$ . we have enough statistical evidence to reject null hypothesis since observed  $t$  is lesser than 1.6586

# Hypothesis Testing

**Case 2:** Hypothesized testing if students want an average work experience of more than 1 year before going for further studies

Let us assume the null hypothesis as the average work experience of students willing to go for higher studies is less than 1 year. Let  $\alpha = 0.05$ . The hypotheses are:

$$H_0 = \mu \leq \mu_0$$

$$H_a = \mu > \mu_0$$

where  $\mu_0 = 1$

$\bar{X}$  = Average years of work experience before going for further studies

$S^2$  = Sample standard deviation of number of years of work experience of students willing to go for higher studies

$n$  = Number of students planning to go for further studies

# Hypothesis testing

## Case 2 continued

Information:

$$\bar{X} = 1.33 \quad (17)$$

$$S^2 = 1.74 \quad (18)$$

$$n = 87 \quad (19)$$

$$df = n - 1 = 86 \quad (20)$$

The test static is given by:

$$t^* = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{1.33 - 1}{\frac{\sqrt{1.74}}{\sqrt{87}}} = 2.34 \quad (21)$$

Using the rejection region approach, we fail reject  $H_0$  if  $t_{0.05,86} \geq t$  where  $t_{0.05,86} = 1.6628$ . Since,  $t^*$  does not fall within the region, we fail to reject  $H_0$ .

# Hypothesis testing

**Case 4:** Hypothesized proportion testing if there is enough evidence that the proportions of people opting for masters, MBA, Phd are not all equal

Sample data :

Masters	MBA	PhD	Total
50	24	13	87

Let  $P_{Ms}$ ,  $P_{MBA}$ ,  $P_{PhD}$  denote proportions of students willing to pursue Masters, MBA, PhD for higher studies

$$H_0 : P_{Ms} = P_{MBA} = P_{PhD} = \frac{1}{3} \quad H_a : \text{atleast one } P \neq \frac{1}{3} \quad \alpha = 0.05$$

Also,

$$E = \frac{1}{3} \times 87 = 29 \quad (22)$$

and,

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad (23)$$

## Case 4 continued

$$\begin{aligned}\chi^2 &= \frac{(50 - 29)^2}{29} + \frac{(24 - 29)^2}{29} + \frac{(13 - 29)^2}{29} \\ &= 15.2 + 0.862 + 8.827 \\ &= 24.889\end{aligned}\tag{24}$$

At  $df = 3 - 1 = 2$ ,  $p \text{ value} = 0.0001$

$p \text{ value} < \alpha = 0.05$

Hence there is enough evidence that population proportions are not all equal.