



Survey on Impact of Online Studies during Lockdown

Applied Statistics

MA4240

Project presentation

Advisor: Dr. Sameen Naqvi

J.Shashank (AI21MTECH11003)

Aman Ladkat (AI21MTECH14011)

Divya Gupta (AI21MTECH12002)

Akshad Shyam (AI22MTECH02006)

Sanyog Sharma (AI21MTECH12003)

Kaustubh Kartikey (AI21MTECH12006)

Ishita Bardhan (AI21MTECH12001)

Ujwal Sai (AI20MTECH01001)

Shailesh Agarwal (AI21MTECH11001)



Content

1. Introduction
2. Data Collection
3. Data Visualisation
4. Verifying Central Limit Theorem
5. Method of Moments and Maximum Likelihood Estimation of the parameters
6. Sampling Distribution of Sample mean when Population Variance is unknown
7. Hypothesis Testing method for Variables
8. Conclusion



Introduction

- Due to Covid-19 lockdown, in-person meets and classes have been replaced with google meets and Teams classes.
- The stress levels, motivation factors, and study environment affect the mental health of the students and in turn their course performance.
- It is necessary to examine the effect of course design and structure, quality of the instructor, and students' expectations on the satisfaction of students with online classes during the Covid-19 lockdown.
- This study aims to compare the perceptions of university students of IIT Hyderabad (both undergraduate and postgraduate) regarding the effectiveness of online courses.



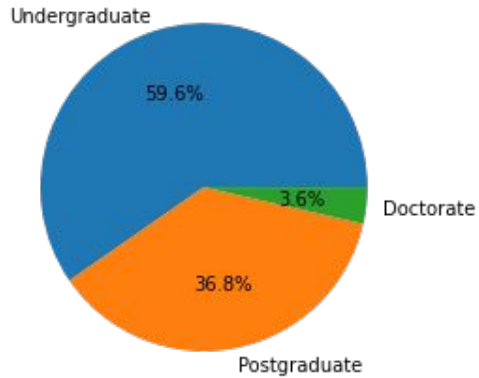
Data Collection

- Collected data from IITH community
- 223 responses
- Questionnaire had 16 columns consisting of numerical and categorical data
- Categorical columns: Gender, Degree, Internet connectivity issues, Area, Lecture modes, Online learning content quality, Class interaction with instructors, Personalised interaction with instructors, Level of interaction with your batchmates, Study environments, Online examination experience, Motivation, Work life balance, Stress levels
- Numerical columns: Study hours, Grades

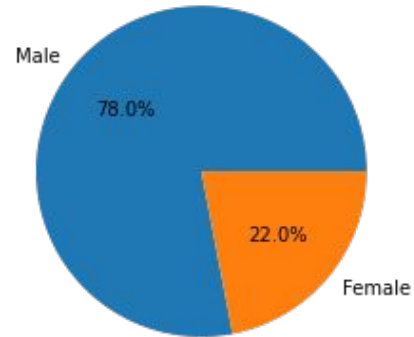


Data Visualization - Pie Charts

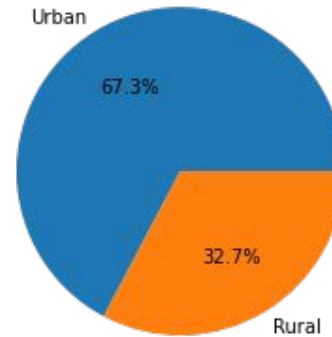
Degree



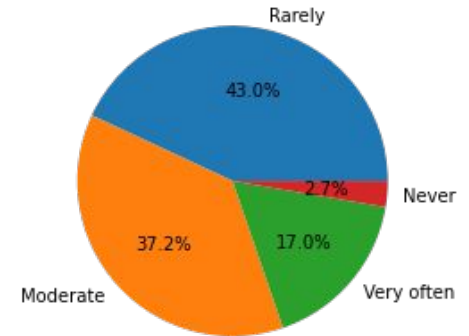
Gender



Area



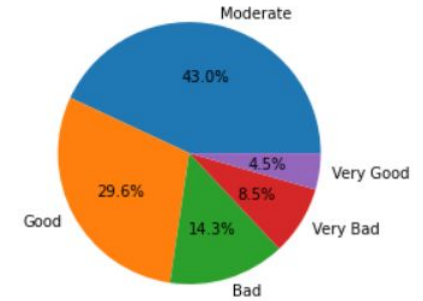
Internet connectivity issues



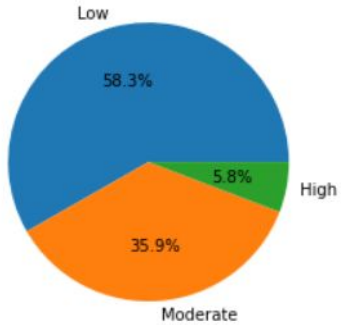


Data Visualization - Pie Charts

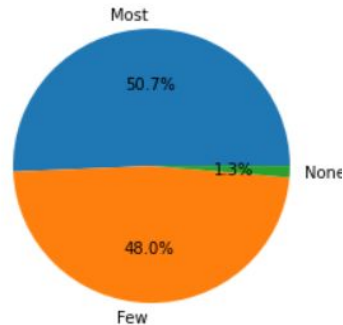
Online learning content quality



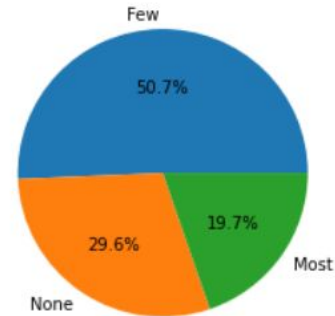
Level of interaction with your batchmates



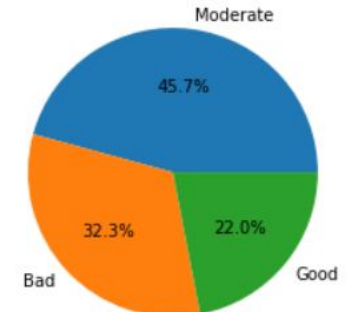
Class interaction with instructors



Personalized Interaction with instructors

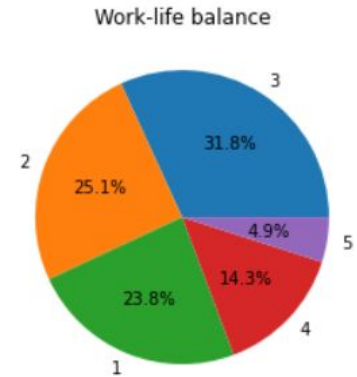
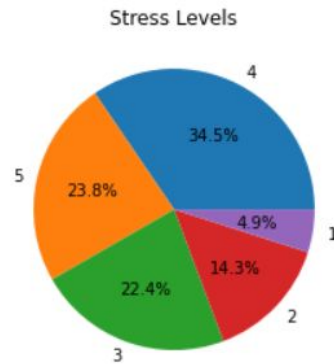
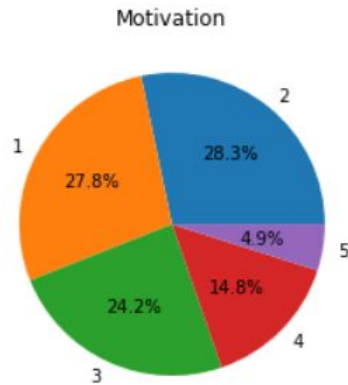
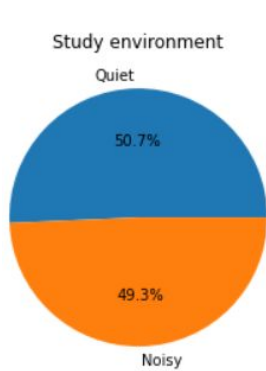


Online examination experience



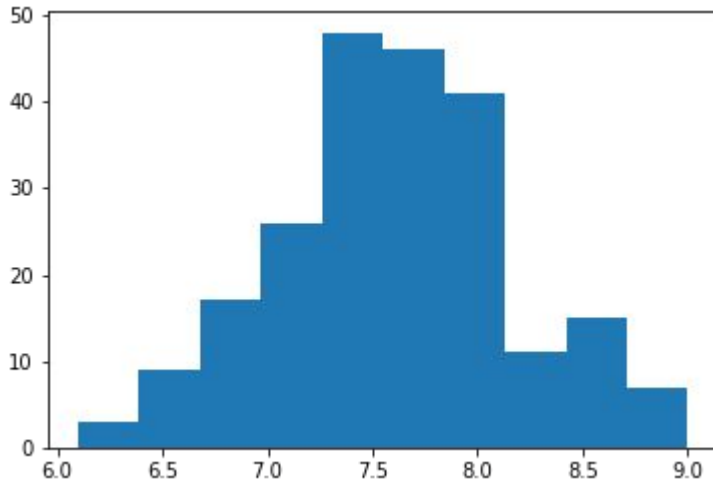


Data Visualization - Pie Charts

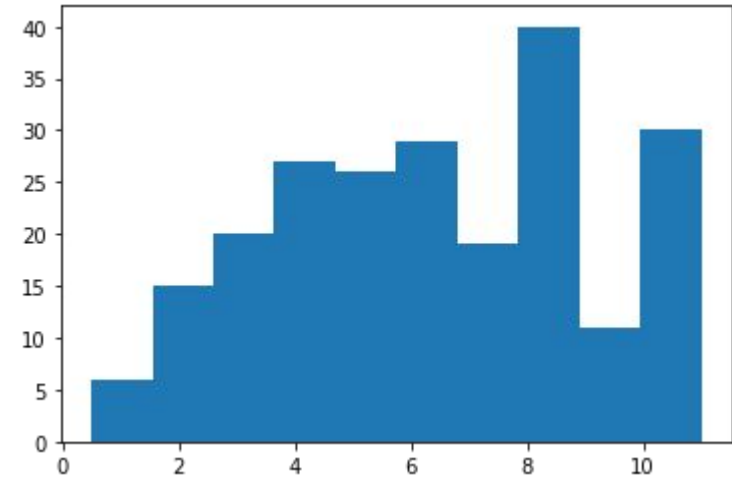




Data Visualization - Histograms



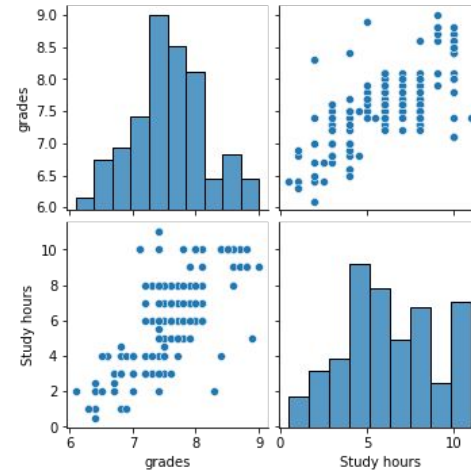
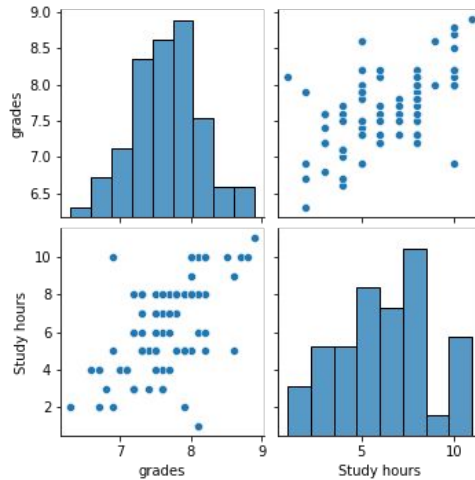
Histogram of Grades



Histogram of Study Hours



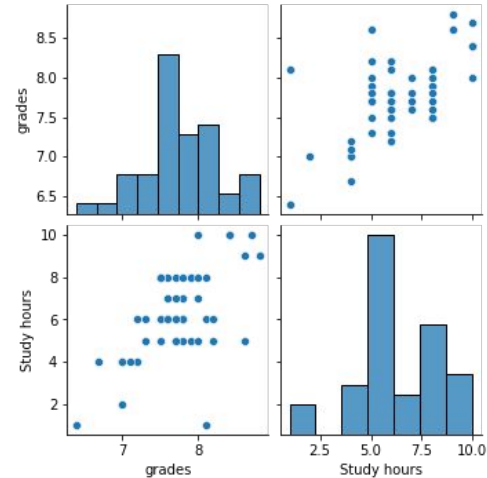
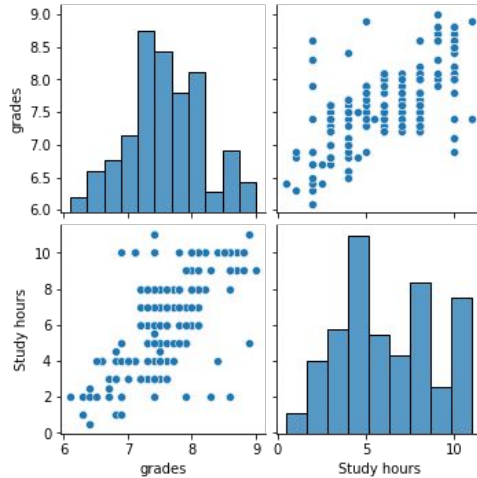
Data Visualization - Pairplots



Comparison of numerical attributes for postgraduates and undergraduates



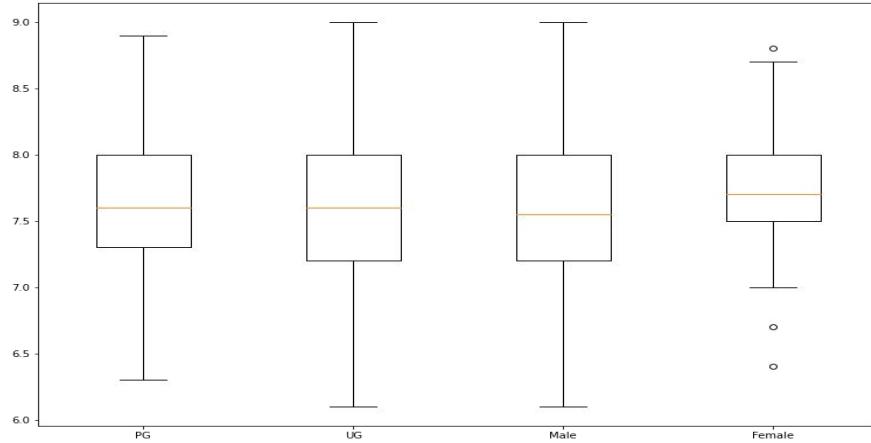
Data Visualization - Pairplots



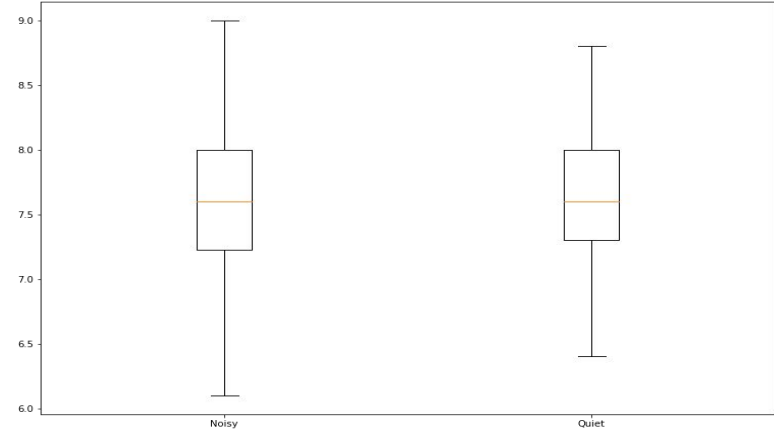
Comparison of numerical attributes for male and female responses

A slight positive correlation between study hours and grades can be observed. However, some outliers can also be seen in each of the cases.

Data Visualization - Box Plots

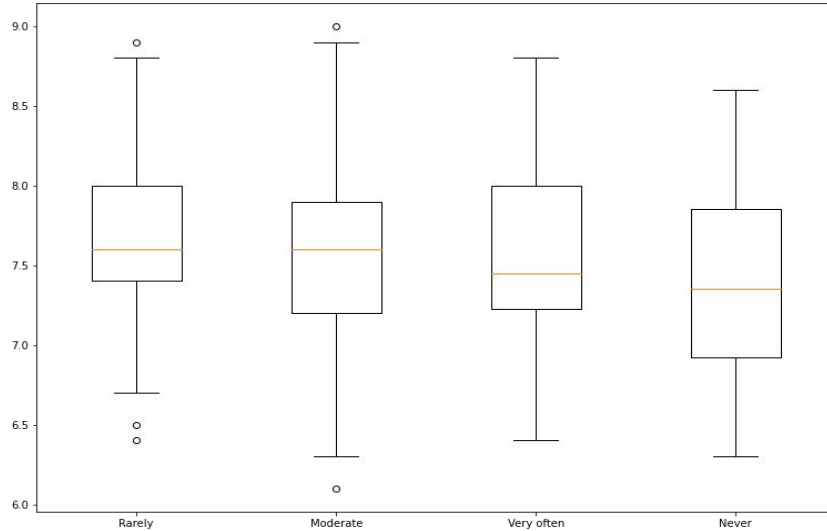


Box plot comparing grades based on degree and gender

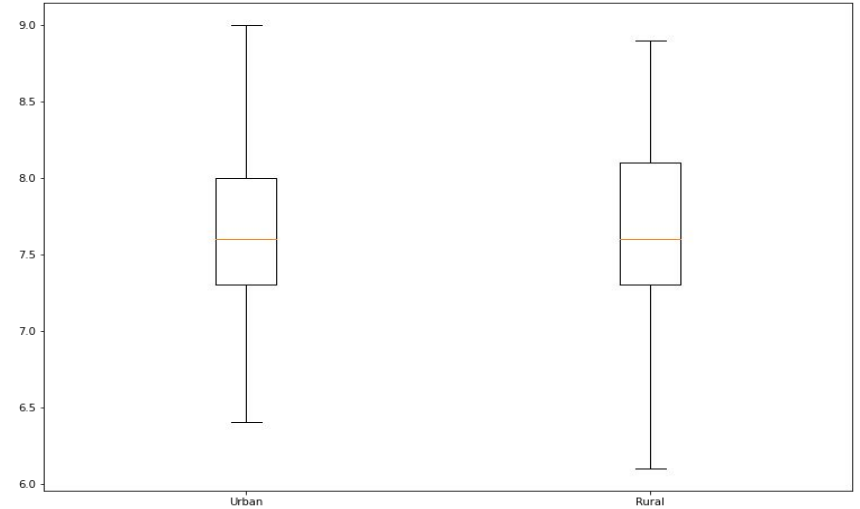


Box plot comparing grades based on study environment

Data Visualization - Box Plots



Box plots comparing grades for various Internet connectivity issues

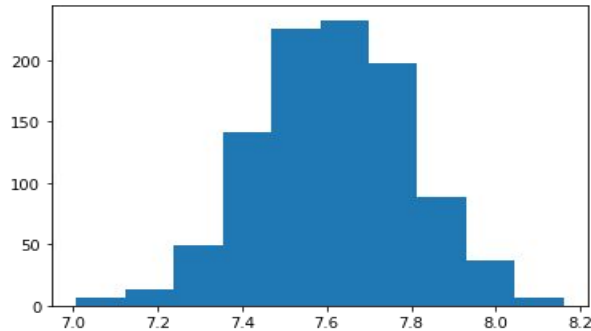


Box plot comparing grades based on area

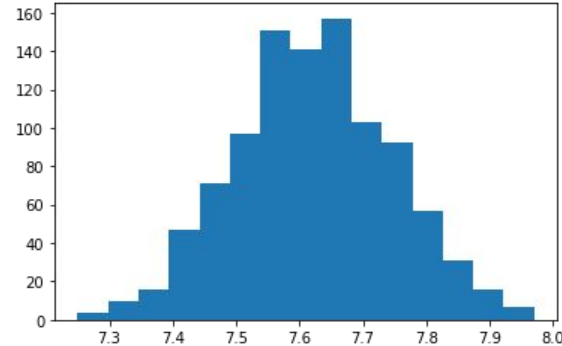


Verifying the Central Limit Theorem

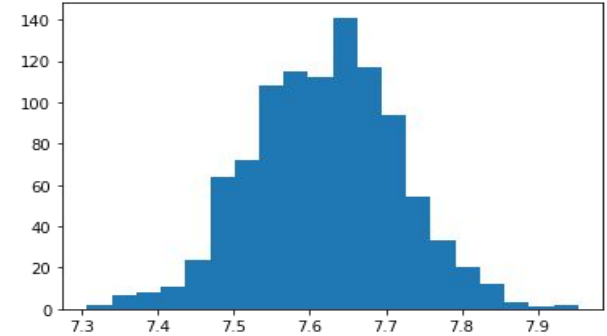
Grades column is considered for verifying the Central Limit Theorem (CLT). We take 1000 batches from the data and find the mean by taking sample size equal to 10, 20 and 30.



$n=10$



$n=20$



$n=30$



Method of Moments and Maximum Likelihood Estimation

- MME: we compare the corresponding moments of the population with the sample and estimate the mean and variance of the data.
- MLE: we multiply the pdf of the random sample and take log on it. We have considered the data to be of Normal distribution as $n > 30$. We then find the maximum value of the log-likelihood function and estimate the parameters, mean and variance of the data.

Method of Moments Estimate of Mean: 7.620535714285714

Method of Moments Estimate of Variance: 0.3387747130102028

Maximum Likelihood Estimate of Mean: 7.6400000000000006

Maximum Likelihood Estimate of Variance : 0.33563587443946213



Confidence Interval Estimation for Population Mean

Estimating the average grade of the student community of IIT-H using the average grade of the collected sample of students as the point estimate. The margin error was calculated using the sample variance and sample size. Since the grade values were normally distributed, the below formula was used for the calculation of lower and upper bounds:

$$\bar{x} \pm t_{\alpha/2, n-1} \left(\frac{s}{\sqrt{n}} \right)$$

```
✓ [10] lower_bound = sample_mean - error_margin  
0s      upper_bound = sample_mean + error_margin  
      print("Confidence interval for population mean: [", lower_bound, upper_bound, "]")
```

```
Confidence interval for population mean: [ 7.57127395690354 7.660116177625606 ]
```



Confidence Interval Estimation for difference of Means between two categories

We also estimate the confidence interval of the difference of grades between undergraduate and postgraduate students. Since both populations are normally distributed with unknown and unequal variances, we use the **Welch's t-interval**.

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2, r} \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}},$$

$$r = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right)^2}{\frac{(S_X^2/n)^2}{n-1} + \frac{(S_Y^2/m)^2}{m-1}}.$$

```
✓ [39] lower_bound = point_estimate - error_margin  
0s      upper_bound = point_estimate + error_margin  
      print("Welch's t-interval for difference of means of UG and PG grades: [", lower_bound, upper_bound, "])
```

Welch's t-interval for difference of means of UG and PG grades: [-0.21457873535796915 0.09255416172876488]



Hypothesis Testing

We are using the chi-square hypothesis test of independence to prove whether the variables are dependent or independent.

Below is the explanation on how this test is used:

- State the Null and Alternative Hypotheses
- Compute the degrees of freedom
 - In this case, we use of contingency table for categorical variables (most of the features or variables are categorical),
 - degree of freedom = $(\text{no_of_rows} - 1)(\text{no_of_columns} - 1)$
- Identify the critical value of test statistic and state the decision rule
 - We will find the χ^2_{crit} (df, 1- α) where $\alpha = 0.05$
 - In these case, if the $\chi^2_{\text{obt}} > \chi^2_{\text{crit}}$, then Null Hypothesis is rejected.



Hypothesis Testing contd

- Compute the obtained value for the test statistic
 - From the contingency table, we get obtained frequency matrix f_o and we need to find obtained frequency matrix f_e which can be obtained by below formula
$$f_e(i) = (rm(i) * clm(i))/N$$
$$\chi^2_{obt} = \sum (f_o(i) - f_e(i))/f_e(i)$$
 - where, χ^2_{obt} = obtained value
 - $f_o(i)$ = obtained frequency for cell i
 - $f_e(i)$ = expected frequency for cell i
 - $rm(i)$ = row marginal/sum of row of cell i
 - $clm(i)$ = column marginal/sum of column of cell i
 - N = total sample size
- Make a decision about Null and state the substantive conclusion based on the decision rule



Hypothesis Testing contd

Internet Connectivity and Area of living

- Null Hypothesis : Area of Living does not affect Internet Connectivity issues during Online learning.
- Alternative Hypothesis: Area of Living affects Internet Connectivity issues during Online learning.

$$df = 3 \quad \chi^2_{\text{crit}} = 7.81472790 \quad \chi^2_{\text{obt}} = 40.19730779884739$$

Here $\chi^2_{\text{obt}} > \chi^2_{\text{crit}}$, we **reject** the null hypothesis.

	Internet Connectivity Issue			
Area	Moderate	Never	Rarely	Very Often
Rural	27	2	16	28
Urban	56	4	80	10



Hypothesis Testing contd

Degree and Work-life Balance

- Null Hypothesis : Degree pursued by student is not related to Work-life Balance.
- Alternative Hypothesis : Degree pursued by student is related to Work-life Balance.

$$df = 8$$

$$\chi^2_{\text{crit}} = 15.50731306$$

$$\chi^2_{\text{obt}} = 8.695489284389$$

$$\text{Here } \chi^2_{\text{obt}} < \chi^2_{\text{crit}},$$

	Work life Balance				
Degree	1	2	3	4	5
Doctorate	0	2	4	2	0
Postgraduate	18	19	24	14	7
Undergraduate	35	35	43	16	4

We cannot reject the null hypothesis



Hypothesis Testing contd

Internet Connectivity Issues

- S_1 = samples with Internet Connectivity Issues as 'Very often' and S_2 = samples with internet Connectivity issues not as 'Very often'
- **Null hypothesis** : $\mu_2 - \mu_1 \leq 0$
- **Alternative Hypothesis** : $\mu_2 - \mu_1 > 0$
- Test statistic value: 1.699
- Critical value (alpha = 0.05) : 1.65177
- Since, test statistic > critical value, we **reject** the null hypothesis



Hypothesis Testing contd

Study Environment

- S_1 = samples with Study Environment as 'Noisy' and S_2 = samples with Study Environment as 'Quiet'
- **Null hypothesis** : $\mu_2 - \mu_1 \leq 0$
- **Alternative Hypothesis** : $\mu_2 - \mu_1 > 0$
- Test statistic value: 5.002099
- Critical value (alpha = 0.05) : 1.65177
- Since, test statistic > critical value, we **reject** the null hypothesis



Hypothesis Testing contd

Stress Levels

- S_1 = samples with stress Levels greater than 3 and S_2 = samples with stress Levels less than or equal to 3
- **Null hypothesis** : $\mu_2 - \mu_1 \leq 0$
- **Alternative Hypothesis** : $\mu_2 - \mu_1 > 0$
- Test statistic value: -10.575068
- Critical value (alpha = 0.05) : 1.65177
- Since, test statistic < critical value, we **fail to reject** null hypothesis.



Conclusion

- The most influential factors in determining the quality of students' learning experiences would appear to be the resources.
- Other factors of often greater importance are those associated with the environment of the learning setting.
- The level of interaction with the instructor and batch mates, study environment and stress levels play significant roles in students' perceptions of the quality of their learning experience.



Thank you!