A PROJECT REPORT
ON
"Factors Affecting Typing Speed"

MA4240 - Applied Statistics

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Team Members

Amulya Tallamraju	AI20BTECH11003
Vaibhav Chhabra	AI20BTECH11022
Anita Dash	MA20BTECH11001
Anjali	MA20BTECH11002
Ruthwika Boyapally	MA20BTECH11004
Kunal Nema	MA20BTECH11007
Sparsh Gupta	MA20BTECH11015
Tapishi Kaur	MA20BTECH11017

UNDER THE GUIDANCE OF
Prof. Sameen Naqvi

ABSTRACT

Typing is now one of the essential part of almost every job, since it lets you do your work very quickly and efficiently. This report aims to analyze the typing speed of students in IITH and tries to draw some conclusions based on the factors like gender, batch (degree program - UG, Masters or PhD) and branch, and checks if there is some relation between them and typing speed. We have used visualizations and tables in our report to present the data more effectively.

Introduction

The objective of our study is to ascertain factors affecting the typing speed of students. For logistical reasons, we are studying the population of only IIT Hyderabad.The factors that we considered for the purpose of this study are as follows:

- **Category 1:** Gender
- **Category 2:** Batch [1]
- **Category 3:** Branch

Through this report we strive to correlate these factors to the typing speed of the individuals. The two dependent variables we study to analyze the typing speed of an individual are

- WPM (Words per Minute)
- Accuracy during test (in percentage)

We perform an observational study to find how the above stated factors affect the typing speed of an individual in case they have any effect. We then summarise the the generated sample and analyze the sample statistics to estimate population parameters.

[1] We refer to the factor of the current program the student is enrolled into as its batch

Data Collection

Variables of Interest

The independent variables are :

1. **Gender:** Nominal Categorical variable
2. **Branch:** Nominal Categorical variable
3. **Batch:** Ordinal Categorical variable

For the purpose of analysis we will group branches into:

1. **Group 1**: Coding related engineering branches(CSE, MA, AI, ES)
2. **Group 2**: Bio-chemistry related engineering branches(CH, CE, MSME, BME, BTE)
3. **Group 3**: Physics related engineering branches(Physics, ME, EE, EP)
4. **Group 4**: Non-engineering branches(LA, Design, Others)

The dependant variables are :

1. **WPM:** Numerical Variable
2. **Accuracy:** Numerical Variable

Survey

A google form was circulated to collect data. To avoid measurement problem in responses, participants were urged to take a common typing test.[2]

The data frame generated had 157 entries with the following statistics:

Total distribution:

Male	Female	UG	PG	PHD	Group 1	Group 2	Group 3	Group 4
107	50	80	41	36	53	36	36	32

The survey we conducted is an example of convenience sampling hence the data frame generated would be biased. Any analysis of a biased sample will only represent the sample itself and not the entire population

[2]We made use of the website monkeytype

Sampling

To combat the inherent bias involved in convenience sampling, we use probabilistic sampling along with convenience sampling.

We stratify the data on the basis of value of one independent variable at a time and randomly sample fixed number of points from each group. We then analyze them to draw conclusions.

To determine the minimum number of samples needed to be drawn from each group, first the distribution of sample needs to be ascertained.

Sample Distribution

Histogram test

The histograms generated in python for the wpm and accuracy are as follows:

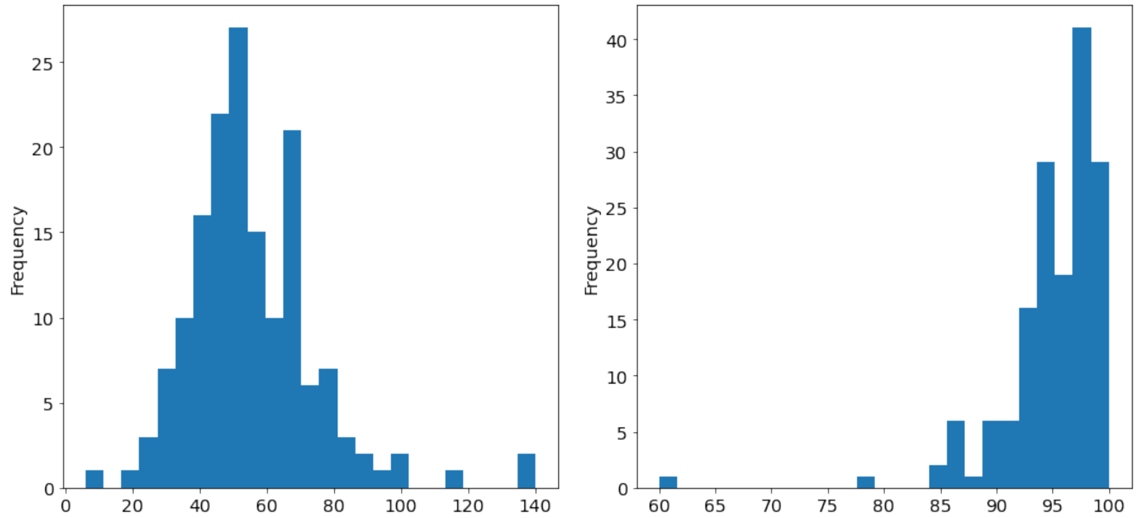


Figure 1: Histograms for WPM and Accuracy

The graph shows that the data for wpm is right skewed and the accuracy is left skewed. Hence it is likely that our sample is not normal.

Q-Q plot test

The Q-Q plots in python for the accuracy and wpm are as follows:

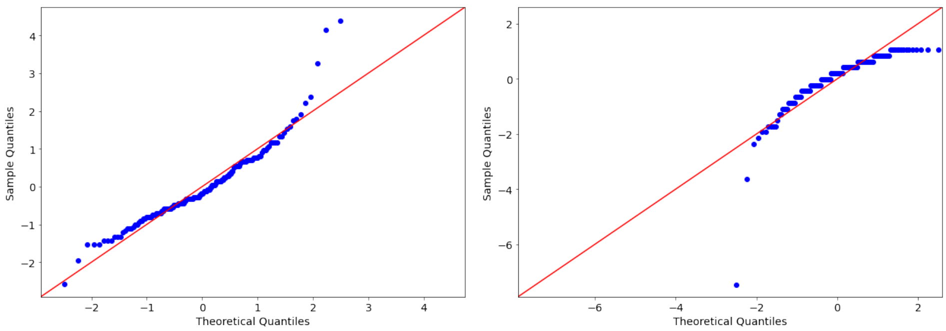


Figure 2: QQ plot for WPM and accuracy

Some of the plotted points fall far from straight line. It gives the impression that the sample is not normally distributed.

p-value test for normality

The p value test(more details) also concludes that our data is not normal.

```
[ ] from scipy.stats import shapiro
    stat, p = shapiro(sample_data['WPM'])
    alpha = 0.05
    if p > alpha:
        print('Sample is normal (fail to reject H0)')
    else:
        print('Sample is not normal (reject H0)')

    Sample is not normal (reject H0)
```

Figure 3: p-test for normality

Using Central Limit Theorem

The CLT states that in case of skewed data set, if the random sample size(n) is > 30, the sample distribution follows normal distribution.

Note:- The histogram and qq plot show that the distribution is not very heavily skewed hence CLT will hold for $n > 30$.

The following graphs show the mean for all possible random samples of size 30 from each group verifying the claim that CLT will hold.

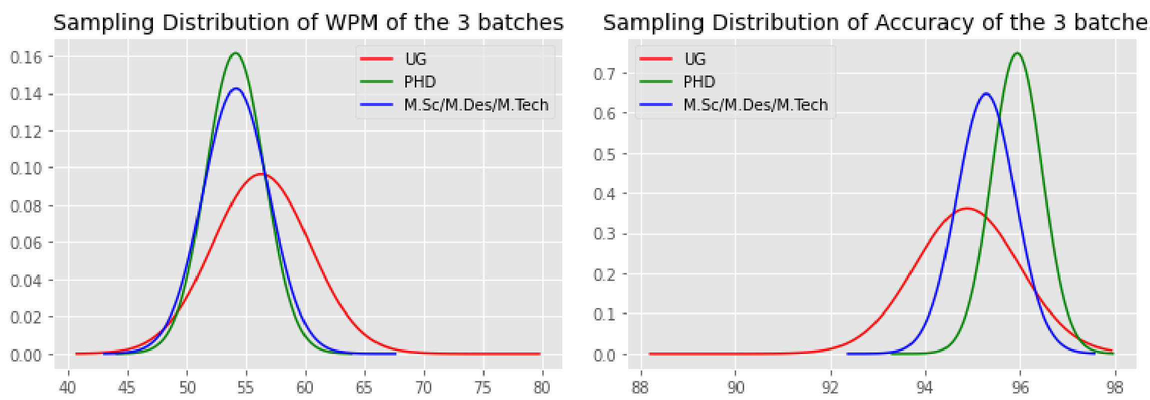


Figure 4: Distribution of sample mean divided by batch

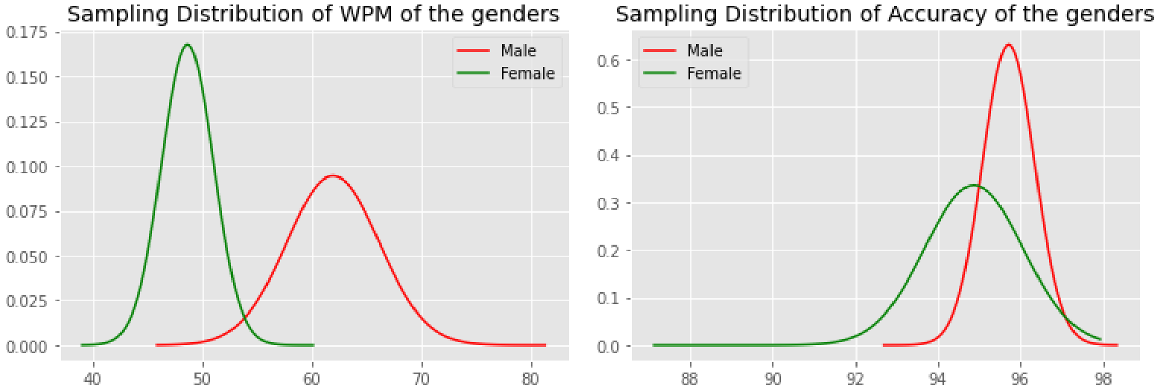


Figure 5: Distribution of sample mean divided by gender

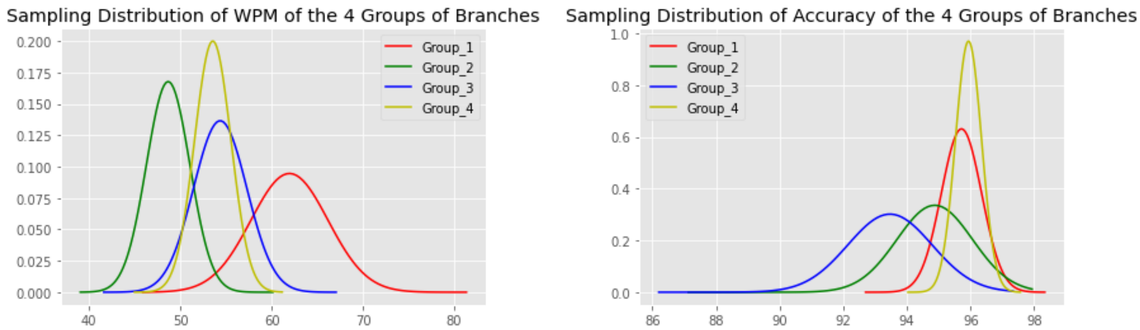


Figure 6: Distribution of sample mean divided by branch

Hence the samples selected for each category are as follows:-

Sample:

Male	Female	UG	PG	PHD	Group 1	Group 2	Group 3	Group 4
107	50	80	41	36	53	36	36	32
70	39	60	35	33	45	30	30	30

Summarizing and Visualizing Sample Data

The collected data has been summarized and visualized under **3** categories for which the values of mean and variances are calculated and graphs are made to help in visualise the data.
(NOTE: Since the data is being sampled randomly, below mentioned data is for one randomly generated sample.)

Gender-based Comparison

The data was divided into 2 strata, one representing the 'Males' and other for 'Females'. Consider the following statistics:

Sample size:

Male	70
Female	39

Mean:

	WPM	Accuracy
Male	61.98	94.86
Female	49.17	94.86

Variance:

	WPM	Accuracy
Male	539.23	43.90
Female	180.20	43.90

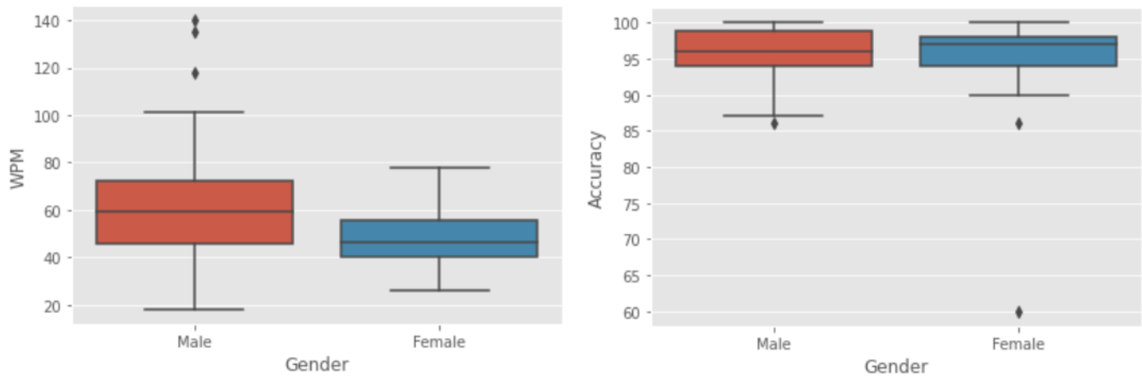


Figure 7: Side-by-Side Box plots for Gender-based Evaluation

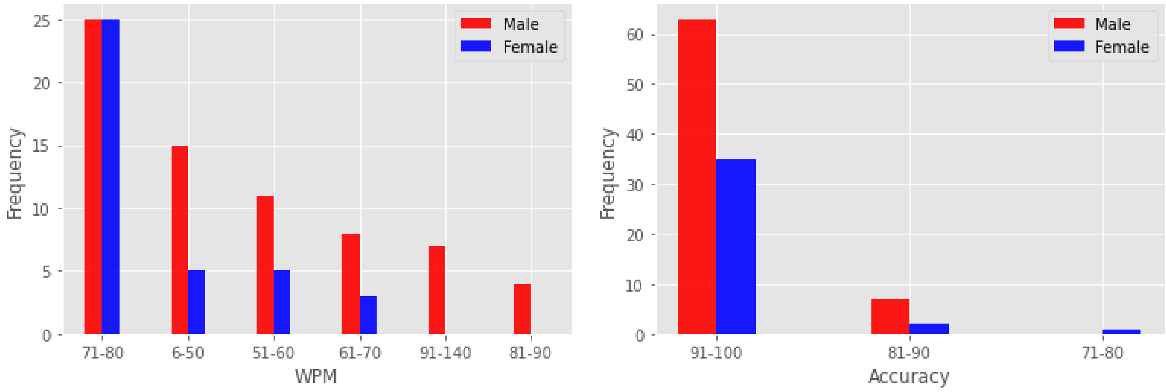


Figure 8: Side-by-Side Bar graph plots for Gender-based Evaluation

Batch-based Comparison

The data had been divided into 3 strata; one strata represents 'UG Students', second represents 'PhD Students' and third 'Masters Students'. Consider the following *statistics*:

Sample size:

UG	60
PHD	33
Masters	35

Mean:

	WPM	Accuracy
UG	56.31	94.88
PHD	54.12	95.93
Masters	54.17	95.28

Variance:

	WPM	Accuracy
UG	525.16	37.39
PHD	187.67	8.87
Masters	242.29	11.85

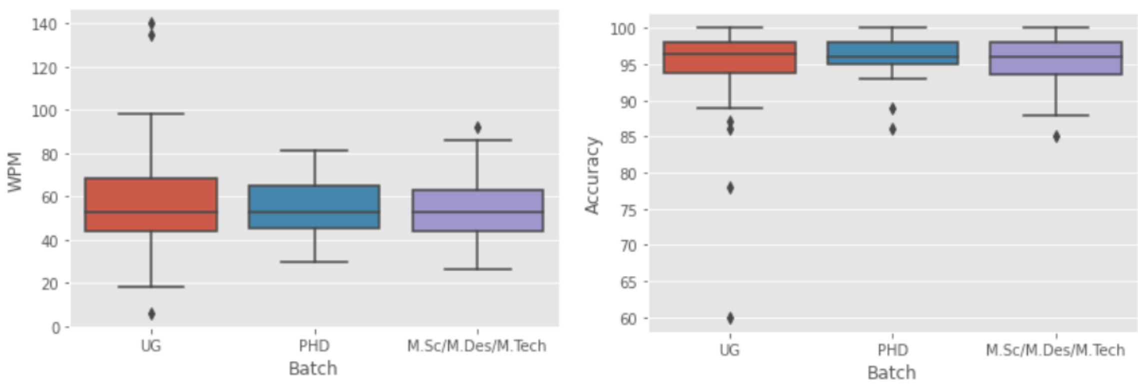
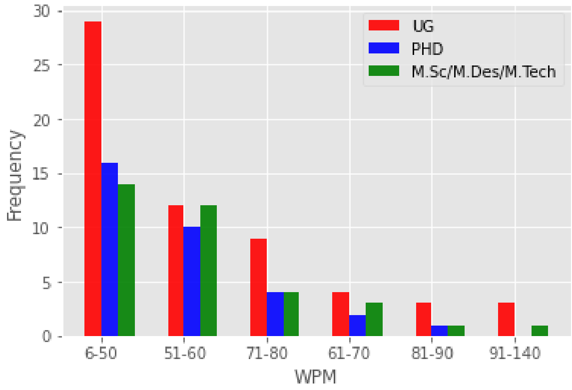


Figure 9: Side-by-Side Box plots for Batch-based Evaluation



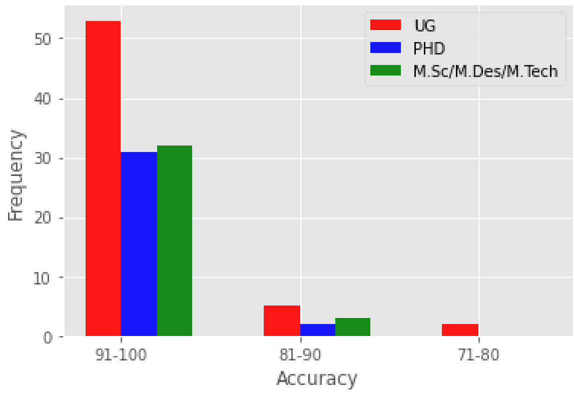


Figure 10: Side-by-Side Bar graph plots for Batch-based Evaluation

Branch-based Comparison

The data had been divided into 4 startas; each one representing a group of branches, following are the groups:

Group 1: CSE, MA, AI, ES
Group 2: CH, CE, MSME, BME, BTE
Group 3: Physics[3], ME, EE, EP
Group 4: LA[4], Des[5], Others

Consider the following *statistics*:

Sample size:

Group 1	45
Group 2	30
Group 3	30
Group 4	30

Mean:

	WPM	Accuracy
Group 1	66.29	95.22
Group 2	45.33	96.16
Group 3	54.37	93.47
Group 4	53.57	95.93

Variance:

	WPM	Accuracy
Group 1	650.66	11.90
Group 2	120.64	17.66
Group 3	262.52	53.91
Group 4	123.43	5.24

[3]For PhD students, pursuing for degree under department of Physics
[4]Liberal Arts
[5]Master/Doctorate in Design

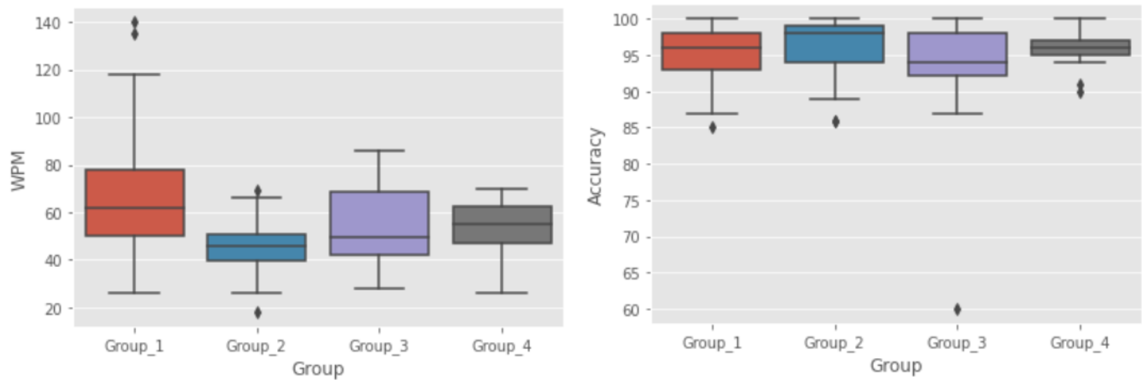


Figure 11: Side-by-Side Box plots for Group-based Evaluation

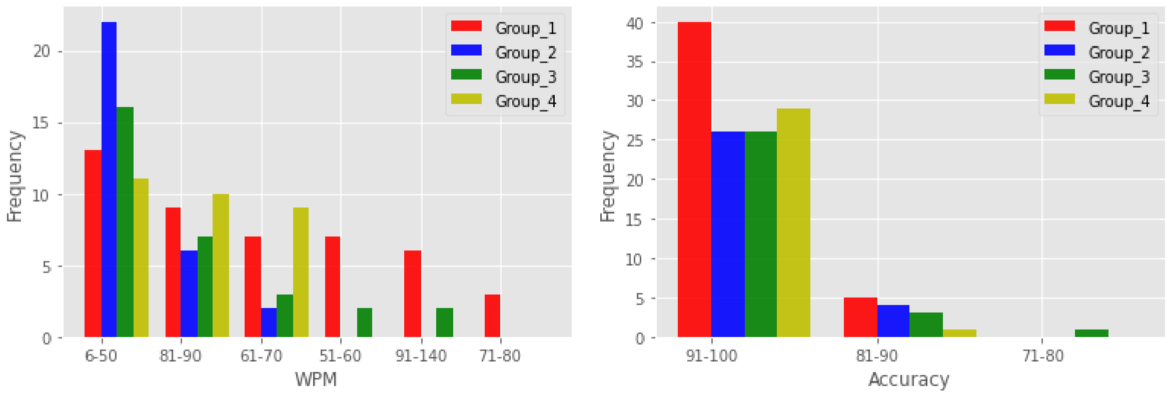


Figure 12: Side-by-Side Bar graph plots for Group-based Evaluation

Analyzing the Data

Confidence Interval Estimation

With the data of the sample, the quantities like population mean and variance can be estimated using the concept of *Confidence Interval*.

1) CI for estimating population mean

If $X_1, X_2, \ldots, X_n$ are normally distributed with unknown mean μ and variance σ^2, then a $(1-\alpha)100\%$ CI for the population mean μ is:

$$\left(\bar{X} - t_{\alpha/2,n-1}\left(\frac{S}{\sqrt{n}}\right),\ \bar{X} + t_{\alpha/2,n-1}\left(\frac{S}{\sqrt{n}}\right)\right)$$

(a) **Gender-based Estimation of mean:** Here we have two values of category i.e. Male and Female, so we can say with 95% confidence that population mean of both will lie within specified range as follows:

	WPM	Accuracy
Male	[56.45, 67.52]	[94.87, 96.56]
Female	[44.83, 53.53]	[92.72, 97.02]

(b) **Batch-based Estimation of mean:** Here we have three values of category i.e. UG, PHD and Masters, so we can say with 95% confidence that population mean of all variable will lie within specified range as follows:

	WPM	Accuracy
UG	[50.40, 62.23]	[93.30, 96.46]
PHD	[49.26, 58.97]	[94.88, 96.99]
Masters	[48.82, 59.52]	[94.10, 96.46]

(c) **Branch-based Estimation of mean:** Here we have four values of category i.e. Group 1, Group 2, Group 3, and Group 4, so we can say with 95% confidence that population mean of all variable will lie within specified range as follows:

	WPM	Accuracy
Group 1	[58.63, 73.95]	[94.19, 96.26]
Group 2	[41.23, 49.43]	[94.60, 97.74]
Group 3	[48.32, 60.42]	[90.72, 96.21]
Group 4	[49.42, 57.72]	[95.08, 96.79]

2) CI for estimating difference of population means

- **Two sampled pooled t-interval** $\left(\frac{1}{4} < \frac{S_X^2}{S_Y^2} < 4\right)$

  If $X_1, X_2, ..., X_n \sim N(\mu_1, \sigma^2)$ and $Y_1, Y_2, ..., Y_m \sim N(\mu_2, \sigma^2)$ are independent random samples, then a $(1-\alpha)100\%$ CI for the difference in the population means, $\mu_1 - \mu_2$ is:

$$\left((\bar{X}-\bar{Y}) - t_{\alpha/2,n+m-2} S_P \sqrt{\frac{1}{n}+\frac{1}{m}},\ (\bar{X}-\bar{Y}) + t_{\alpha/2,n+m-2} S_P \sqrt{\frac{1}{n}+\frac{1}{m}}\right)$$

  where

$$S_P^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

- **Welch's t-interval** $\left(\frac{S_X^2}{S_Y^2} < \frac{1}{4} \textbf{ or } 4 < \frac{S_X^2}{S_Y^2}\right)$

  If $X_1, X_2, ..., X_n \sim N(\mu_1, \sigma_X^2)$ and $Y_1, Y_2, ..., Y_m \sim N(\mu_2, \sigma_Y^2)$ are independent random samples, then a $(1-\alpha)100\%$ CI for the difference in the population means, $\mu_1 - \mu_2$ is:

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

  where

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

(a) **Estimation difference: Gender-based** With 95% confidence we can say that difference of mean in male and female will lie within:

	WPM	Accuracy
Male-Female	[5.85, 19.77]	[-1.44, 3.14]

(b) **Estimation difference: Batch-based** With 95% confidence we can say that difference of mean in all possible batches will lie within range as follows:

	WPM	Accuracy
UG-PHD	[-6.48, 10.87]	[-0.82, 2.93]
PHD-Masters	[-7.06, 7.16]	[-0.90, 2.12]
UG-Masters	[-8.62, 8.72]	[-1.58, 2.89]

(c) **Estimation difference: Branch-based** With 95% confidence we can say that difference of mean in all possible batches will lie within range as follows:

	WPM	Accuracy
Group 1-2	[12.39, 29.54]	[-0.83, 2.71]
Group 1-3	[1.46, 22.39]	[-0.76, 4.27]
Group 1-4	[4.11, 21.33]	[-0.72, 2.14]
Group 2-3	[1.88, 16.19]	[-0.39, 5.79]
Group 2-4	[2.52, 13.94]	[-1.52, 1.98]
Group 3-4	[-6.38, 7.98]	[-6.38, 7.98]

3) CI for estimating population variance

If $X_1, X_2, ..., X_n$ are normally distributed and $a = \chi^2_{1-\alpha/2,n-1}$, $b = \chi^2_{\alpha/2,n-1}$, then a $(1-\alpha)100\%$ CI for the population variance σ^2 is:

$$\left(\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right)$$

(a) **Gender-based Estimation of variance:** Here we have two values of category i.e. Male and Female, so we can say with 95% confidence that population variance of both will lie within specified range as follows:

	WPM	Accuracy
Male	[653.82, 1626.51]	[14.82, 36.85]
Female	[120.34, 299.29]	[29.27, 72.93]

(b) **Batch-based Estimation of variance:** Here we have three values of category i.e. UG, PHD and Masters, so we can say with 95% confidence that population variance of all variable will lie within specified range as follows:

	WPM	Accuracy
UG	[377.33, 781.23]	[26.87, 55.63]
PHD	[121.37, 328.33]	[5.73, 15.52]
Masters	[158.66, 416.28]	[7.75, 20.35]

(c) **Branch-based Estimation of variance:** Here we have four values of category i.e. Group 1, Group 2, Group 3, and Group 4, so we can say with 95% confidence that population variance of all variable will lie within specified range as follows:

	WPM	Accuracy
Group 1	[445.93,1038.25]	[8.16, 18.99]
Group 2	[76.52, 218.03]	[11.20, 31.92]
Group 3	[166.50, 474.41]	[34.19, 97.43]
Group 4	[78.28, 223.05]	[3.32, 9.46]

4) CI for estimating ratio of population variance

If $X_1, X_2, ..., X_n \sim N(\mu_X, \sigma_X^2)$ and $Y_1, Y_2, ..., Y_m \sim N(\mu_Y, \sigma_Y^2)$ are independent samples and $c = F_{\alpha/2}(m-1, n-1)$, $d = F_{1-\alpha/2}(m-1, n-1)$,

$$\left(c\frac{S_X^2}{S_Y^2},\ d\frac{S_X^2}{S_Y^2}\right)$$

(a) **Estimation ratio: Gender-based** With 95% confidence we can say that ratio of variance in male and female will lie within:

	WPM	Accuracy
Male-Female	[1.66,5.15]	[0.15,0.48]

(b) **Estimation ratio: Batch-based** With 95% confidence we can say that ratio of variance in all possible batches will lie within range as follows:

	WPM	Accuracy
UG-PHD	[1.47, 5.04]	[2.21, 7.60]
PHD-Masters	[0.38, 1.55]	[0.37, 1.50]
UG-Masters	[0.26, 0.87]	[0.18, 0.60]

(c) **Estimation ratio: Branch-based** With 95% confidence we can say that ratio of variance in all possible batches will lie within range as follows:

	WPM	Accuracy
Group 1-2	[3.13,9.31]	[0.39,1.163]
Group 1-3	[1.44,4.28]	[0.13,0.38]
Group 1-4	[3.06,9.095]	[1.32,3.92]
Group 2-3	[0.24,1.73]	[0.17,1.23]
Group 2-4	[0.52,3.68]	[1.79,12.69]
Group 3-4	[0.24,3.63]	[0.05,0.75]

Hypothesis Testing

Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution.

Test for $\mu_1 - \mu_2$: Independent Samples, Unequal Variances

The sample means (μ_1) and (μ_2) are independent and the variance is assumed to be unequal. So if we are performing **Left-tailed test:**

$$\text{Null Hypothesis } (H_0)\text{: } \mu_1 \geq \mu_2$$
$$\text{Alternative Hypothesis } (H_a)\text{: } \mu_1 < \mu_2$$

The **Test Statistic (TS)** is:

$$t' = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Now,

$$df = \frac{(n_1 - 1)(n_2 - 1)}{(1-c)^2(n_1 - 1) + c^2(n_2 - 1)} \quad \text{where} \quad c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

and df represents degree of freedom.

Using p-value approach, we find the p-value for the test statistic (TS) by using t-distribution table and if

$$\text{p-value} \leq \alpha \text{ : Reject } H_0$$
$$\text{p-value} > \alpha \text{ : Fail to reject } H_0$$

1) **Gender-based:**

- **Test for difference in mean of male and female**
  Here, (WPM)
  H_0: $\mu_1 \geq \mu_2$ and H_a: $\mu_1 < \mu_2$

  For female: n_1 = 39, $\bar{x}_1$ = 49.18, s_1^2 = 180.20
  For male: n_2 = 70, $\bar{x}_2$ = 61.96, s_2^2 = 539.23
  For above values,

$$t' = -3.640, \quad c = 0.375, \quad df = 106.82$$

$$\text{p-value} = 0.000211$$

Since, p-value = 0.000211 $<$ (0.05 = α)
We reject Null Hypothesis (H_0)
Conclusion: The average typing speed (WPM) of male is greater is than average typing speed (WPM) of female.

2) **Batch-based:**

- **Test for difference in mean of UG(μ_1) and Masters(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$0.31 > 0.05$	$0.63 > 0.05$
Conclusion	Fail to reject H_0	Fail to reject H_0

- **Test for difference in mean of UG(μ_1) and PHD(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$0.28 > 0.05$	$0.86 > 0.05$
Conclusion	Fail to reject H_0	Fail to reject H_0

- **Test for difference in mean of Masters(μ_1) and PHD(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$0.49 > 0.05$	$0.79 > 0.05$
Conclusion	Fail to reject H_0	Fail to reject H_0

3) **Branch-based:**

- **Test for difference in mean of Group-1(μ_1) and Group-2(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$3.76 > 0.05$	$0.84 > 0.05$
Conclusion	Fail to reject H_0	Fail to reject H_0

- **Test for difference in mean of Group-1(μ_1) and Group-3(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$0.007 > 0.05$	$0.11 > 0.05$
Conclusion	reject H_0	Fail to reject H_0

- **Test for difference in mean of Group-1(μ_1) and Group-4(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$0.0022 > 0.05$	$0.85 > 0.05$
Conclusion	reject H_0	Fail to reject H_0

- **Test for difference in mean of Group-2(μ_1) and Group-3(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$0.007 > 0.05$	$0.95 > 0.05$
Conclusion	Fail to reject H_0	Fail to reject H_0

- **Test for difference in mean of Group-2(μ_1) and Group-4(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$0.0027 > 0.05$	$0.6 > 0.05$
Conclusion	reject H_0	Fail to reject H_0

- **Test for difference in mean of Group-3(μ_1) and Group-4(μ_2)**

	WPM	Accuracy
Null Hypothesis (H_0)	$\mu_1 \geq \mu_2$	$\mu_1 \geq \mu_2$
Alternative Hypothesis (H_a)	$\mu_1 < \mu_2$	$\mu_1 < \mu_2$
p-value	$0.41 > 0.05$	$0.95 > 0.05$
Conclusion	Fail to reject H_0	Fail to reject H_0

Test for σ_1^2 and σ_2^2

For comparing variances σ_1^2, σ_2^2, we are performing the **Two tailed test**:

$$\text{Null Hypothesis } (H_0)\text{: } \sigma_1^2 = \sigma_2^2$$
$$\text{Alternative Hypothesis } (H_a)\text{: } \sigma_1^2 \neq \sigma_2^2$$

The **Test Statistic (TS)** is:

$$F = \frac{s_1^2}{s_2^2}$$

For a significance level α, with $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$, if

$$F \leq F_{1-\alpha/2,df_1,df_2} \text{ or } F \geq F_{\alpha/2,df_1,df_2}\text{: Reject } H_0$$
$$F_{1-\alpha/2,df_1,df_2} < F < F_{\alpha/2,df_1,df_2} \text{ : Fail to reject } H_0$$

1. **Gender-based:**

   - **Test for comparing variance of male and female**
     (H_0): $\sigma_1^2 = \sigma_2^2$
     (H_a): $\sigma_1^2 \neq \sigma_2^2$
     Here, (WPM)
     For female: n_1 = 39, $\bar{x}_1$ = 49.18, s_1^2 = 180.20
     For male: n_2 = 70, $\bar{x}_2$ = 61.96, s_2^2 = 539.23
     For above values,

     $$F = 0.34, \quad df_1 = 38, \quad df_2 = 69$$

     For $\alpha = 0.05$,

     $$F_{0.975,38,69} = 0.56, \quad F_{0.025,38,69} = 1.94$$

     Since $F < F_{0.975,38,69} = 0.56$, we reject H_0.
     Conclusion: The null hypothesis H_0 is rejected. The variances of typing speed of females are not equal.

2. **Batch-based:**

   - **Test for comparing variance of UG and PHD**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$0.35 \leq 0.55$ or $0.35 \geq 1.90$	$0.23 \leq 0.55$ or $0.23 \geq 1.90$
Conclusion	H_0 is rejected	H_0 is rejected

- **Test for comparing variance of Masters and PHD**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$0.50 \leq 0.77 \leq 2.01$	$0.74 \leq 0.50$ or $0.74 \geq 2.01$
Conclusion	Fail to reject H_0	H_0 is rejected

- **Test for comparing variance of Masters and UG**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$2.16 \leq 0.55$ or $2.16 \geq 1.87$	$3.15 \leq 0.55$ or $3.15 \geq 1.87$
Conclusion	H_0 is rejected	H_0 is rejected

3. **Branch-based:**

- **Test for comparing variance of Group-1 and Group-2**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$0.18 \leq 0.5$ or $0.18 \geq 2.0$	$0.52 \leq 1.48 \leq 2.0$
Conclusion	H_0 is rejected	Fail to reject H_0

- **Test for comparing variance of Group-1 and Group-3**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$0.4 \leq 0.52$ or $0.4 \geq 2.009$	$4.5 \leq 0.5$ or $4.5 \geq 2.009$
Conclusion	H_0 is rejected	H_0 is rejected

- **Test for comparing variance of Group-1 and Group-4**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$0.18 \leq 0.52$ or $0.18 \geq 2.009$	$0.43 \leq 0.5$ or $0.43 \geq 2.00$
Conclusion	H_0 is rejected	H_0 is rejected

- **Test for comparing variance of Group-2 and Group-3**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$0.45 \leq 0.47$ or $0.45 \geq 2.10$	$0.32 \leq 0.47$ or $0.32 \geq 2.10$
Conclusion	H_0 is rejected	H_0 is rejected

- **Test for comparing variance of Group-2 and Group-4**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$0.4 \leq 0.90 \leq 2.10$	$3.3 \leq 0.4$ or $3.33 \geq 2.10$
Conclusion	Fail to reject H_0	H_0 is rejected

- **Test for comparing variance of Group-3 and Group-4**

	WPM	Accuracy
H_0	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 = \sigma_2^2$
H_a	$\sigma_1^2 \neq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
F-value	$0.47 \leq 0.47$ or $0.47 \geq 2.10$	$0.09 \leq 0.47$ or $0.09 \geq 2.10$
Conclusion	H_0 is rejected	H_0 is rejected

Conclusion

We draw the following conclusions:

- Since we performed an observational study, we can only ascertain correlation between the dependent and independent variables.
- In terms of words per minute, males have been found to perform better than their female counterparts. The spread[6] of data is not similar in both cases.
- In terms of wpm, Group 1(CSE, MA, AI, ES) performed the best .Group 2 (CH, CE, MSME, BME,BTE) had the least typing speed. No concrete conclusion can be drawn about the order between Group 3(Physics, ME, EE, EP) and Group 4(LA, Design, Others). Spread of all groups is different but group 1 and group 2 can not be conclusively compared.
- No clear correlation was found between the typing speed of an individual and the course they are enrolled in.UG's have different spread than Masters and PhD's. Masters and PhD's can not be conclusively compared.
- No such conclusion can be drawn about accuracy in any of the above stated cases.

Google Sheet Link with all the data related to the study

[6] By spread, it means the comparison of variances of the groups

Contributions

Amulya Tallamraju:

- Data Collection
- Sampling Distribution for overall population
- Hypothesis Testing for mean, variance based on Gender, Branch and Batch
- Helped out with making the presentation
- Made the Google Form for collecting data

Vaibhav Chhabra:

- Data Collection
- Data Analysis based on batch(graduation-level)
- Made excel sheet for the batch data
- Made the Report on Latex

Anita Dash

- Data Collection
- Sampling Distribution
- Data Analysis Based on Gender
- Plotting the necessary Graphs (side by side box plots, bar plots etc.)
- Made the Google Sheet with all the Data related to the study
- Made the presentation on Latex Beamer

Anjali:

- Data Collection
- Data Analysis based on branch
- Made the Google Sheet with all the Data related to the study

Ruthwika Boyapally:

- Floated Google form via e-mail
- Data Collection
- Data Analysis based on Branch
- Made excel sheet based on branch data
- Made the presentation on Latex Beamer

Kunal Nema:

- Made and floated the Google form for survey
- Data Collection
- Summarizing and Analysis of Data based on Batch
- Made excel sheet for data on Batch
- Made Report on Latex

Sparsh Gupta:

- Data Collection
- Data Analysis based on Batch
- Helped in presentation making
- Made Report on LaTex.

Tapishi Kaur:

- Data Collection
- Sampling Distribution
- Data Analysis on basis of Branches
- Made Report on Latex