

M.Sc. in Applied Statistics and Analytics

Project Report
on
**Students' Performance in
Online Examination: The
Gender Perspective**

Submitted by,

Group: 6

Semester: 1

Abhik De	Anubhab Maity
Koyel Chakraborty	Kushaal Agarwal
Soumitro Mukherjee	Susnato Chakraborty

Supervised by,

Mr. Taranga Mukherjee,
Assistant Professor,
Department of Applied Science

Maulana Abul Kalam Azad University of Technology

Contents

1	Introduction	2
2	Database and Methodologies	2
2.1	Source of Data	2
2.2	Research Questions	3
2.3	Methodologies	3
3	Results and Discussion	13
4	Concluding Remarks	20
5	Conclusion	21
6	References	21

1 Introduction

When the recent pandemic was drastically transforming our socio-economic acumen, the academia faced a prominent paradigm shift. The initiation of virtual classrooms, whether a boon or a bane, is still a matter of debate for many. Moreover, online evaluation systems have received massive appreciation and criticism both at the same time. In this study, we intend to investigate how flexible the new framework is in granting decent grades. Moreover, our motivation is to examine the notions of women emancipation in terms of students' adaptability at the face of technological infusion. If such gender prerogatives are actually valid or not, can be deduced from the performance of students during the online examination.

2 Database and Methodologies

2.1 Source of Data

Primary data have been collected from the first semester students of M.Sc. in Applied Statistics and Analytics course conducted by the Maulana Abul Kalam Azad University of Technology, West Bengal. Our dataset consists of undergraduate scores (CGPAs) of 30 respondents from our class. The aforementioned undergraduate course had the duration of six consecutive semesters (2018-2021). During the course, all semester examinations except the

first and second one were conducted online. In this study, aggregate score (CGPA) of all the semesters taken together will be used. Based on this sample, we are interested to infer if the scores are significantly varied across gender.

2.2 Research Questions

- 1.** Is the overall grade more than 7.5 pointer?
- 2.** Did the female students perform better than their male peers in terms of scoring CGPA?
- 3.** Are the female examinees more consistent than the male aspirants appearing in the examination?

2.3 Methodologies

Exploratory Data Analysis

To check if the normality assumption is valid for the population under study, we have employed exploratory data analysis techniques on our random sample. These exploratory data analysis techniques include Histogram, Kernel Density Plot, Boxplot and Quantile-Quantile Plot.

Hypothesis Tests

Parametric hypothesis tests (One Sample t-Test, Fisher's t-Test, F-test) as well as their non-parametric counterparts (Sign Test, Wilcoxon Signed-Rank Test, Mann-Whitney U Test, Ansari-Bradley Test) have been used

in the study. To draw a final conclusion about the normality behaviour of the data, Kolmogorov-Smirnov Test has been applied prior to these tests.

1. Kolmogorov-Smirnov Test:

This test is used to check if a sample comes from a particular distribution. Under the null hypothesis, it tests the equality of the empirical distribution with that of the theoretical one. According to the assumption, the theoretical population must have an absolutely continuous distribution function. If $F_n(x)$ and $F(x)$ be the empirical and theoretical CDFs respectively, we define the test statistic as,

$$W = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \quad (1)$$

Note. Here, \mathbb{R} denotes the set of all real numbers. *Sup* is an abbreviation for Supremum.

When the null hypothesis is true, i.e., if the empirical distribution approaches the theoretical distribution, the test statistic approaches to zero.

2. Student's t-Test:

Originally devised by William Sealy Gosset in 1908, t-Test is used as an one-sample location test. One principle assumption is that, the sampling distribution is normal.

It is a mean-based test, i.e, it compares the sample mean with a hypothesized population mean value. The test statistic,

$$T = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \quad (2)$$

follows a t -distribution with $(n - 1)$ degrees of freedom.

Note. Here, \bar{x} represents sample mean, μ denotes the population mean, S is the sample standard deviation and n denotes the sample size.

3. Fisher's t -Test:

This is a special version of the t -Test to compare the location parameters of two sample groups. One additional assumption is that, the observations are independent within and across groups. Besides, the two groups are required to be homoscedastic. It is a mean-based test, i.e, it compares the sample means of the two groups. The test statistic,

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}})} \quad (3)$$

follows a t -distribution with $(n_1 + n_2 - 2)$ degrees of freedom.

$$\text{where, } S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

Note. Here, \bar{x}_1, \bar{x}_2 represents sample means of two groups under study, μ_1, μ_2 denotes the population means of the respective groups, S_p is the pooled sample standard deviation and n_1, n_2 denotes the sample sizes.

In case, the two populations under study are not homoscedastic, the test can be performed applying Satterthwaite approximation (Satterthwaite, 1946). This test is also termed as **Welch Test**. The test statistic T' is given by,

$$T' = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (4)$$

which follows a t -distribution with v degrees of freedom.

$$\text{where, } v = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{(n_1-1)(\frac{S_1^2}{n_1})^2 + (n_2-1)(\frac{S_2^2}{n_2})^2}$$

Note. Here, \bar{x}_1, \bar{x}_2 represents sample means of two groups under study, μ_1, μ_2 denotes the population means of the respective groups, S_1, S_2 is the sample standard deviations of the two groups and n_1, n_2 denotes the sample sizes.

4. *F-Test:*

Initially developed by R. A. Fisher in the 1920's, the nomenclature was given by George W. Snedecor. It is a variance-based test, i.e, it compares the sample variances of two normal populations. The test statistic,

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \quad (5)$$

If the null hypothesis is $H_0 : \sigma_1 = \sigma_2$, then the above expression reduces to,

$$F = \frac{S_1^2}{S_2^2} \quad (6)$$

The test statistic follows a F -distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom.

Note. Here, S_1, S_2 represents sample standard deviations of two groups under study and σ_1, σ_2 denotes the population standard deviations of the respective groups. n_1, n_2 denotes the sample sizes.

5. *Sign Test:*

Initially developed by John Arbuthnot in 1710, the test is a non-parametric version of one sample t-Test. It is a median-based test, i.e, it compares the sample median with the hypothesized population median value, $\tilde{\mu}$. Ac-

According to the assumption, the population must have an absolutely continuous distribution function. For this, all sample observations exactly equal to the hypothesized median value needs to be excluded from the analysis. The test procedure comprises of calculating the differences, $d_i = x_i - \tilde{\mu}$ for each sample value x_i . Then, the indicator variable Z_i is defined as,

$$Z_i = \begin{cases} 1, & \text{if } d_i > 0 \\ 0, & \text{otherwise} \end{cases}$$

The test statistic is given by,

$$S_+ = \sum_{i=1}^n Z_i \quad (7)$$

Then probability of getting less than (or more than, depending on the alternative hypothesis) S_+ successes is calculated from a binomial distribution with parameters $n(\text{number of observations} \neq \tilde{\mu})$ and $p = 0.5$. This probability is used as the $p - \text{Value}$ of the test.

6. Wilcoxon Signed-Rank Test:

This test was devised by Frank Wilcoxon in 1945. The test is a non-parametric version of one sample t-Test and a modified version of the sign test. It is a median-based test, i.e, it compares the sample median with the hypothe-

sized population median value. According to the assumption, the population must have an absolutely continuous distribution function. The test procedure comprises of calculating $d_i = (x_i - \tilde{\mu})$ for each sample value x_i and subsequently ranking $|d_i|$'s. The test statistic,

for left-tailed test:

$$W_+ = \sum_i R_i \quad \forall i \ni d_i > 0 \quad (8)$$

for right-tailed test:

$$W_- = \sum_i R_i \quad \forall i \ni d_i < 0 \quad (9)$$

for both-tailed test:

$$W = \min\{W_+, W_-\} \quad (10)$$

Note. Here, R_i represents the i 'th rank. Prior to the test, all observations resulting in tie ranks and zero values of $|d_i|$ must be excluded from the analysis.

If the computed test statistic is less than the tabulated value of W_+ (for left-tailed test), W_- (for right-tailed test) or W (for both-tailed test) at the corresponding sample size n and level of significance α , the null hypothesis is rejected.

7. Mann-Whitney U Test:

This test was proposed by Frank Wilcoxon in 1945. Later, Henry Mann and Donald Ransom Whitney (1947) performed a detailed analysis of the test. For this reason, the test is also called Mann-Whitney-Wilcoxon Test. It is a non-parametric version of the independent samples t-Test or Fisher's t-Test. Inferences are drawn about the location parameters of the two populations under study. According to the assumption, the populations must have an absolutely continuous distribution function. To perform the test, we define an indicator variable,

$$\Phi(X_i, Y_j) = \begin{cases} 1, & \text{if } X_i > Y_j & i = 1(1)n_1 \\ 0, & \text{otherwise} & j = 1(1)n_2 \end{cases}$$

The test statistic is given by,

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Phi(X_i, Y_j) \quad (11)$$

Note. Here, X, Y denotes the two populations under study. X_i denotes the i 'th observation in the random sample taken from population X . Y_j denotes the j 'th observation in the random sample taken from population Y .

When C is the tabulated critical value of U at the specified level of significance α and combined sample size $(n_1 + n_2)$, we reject null hypothesis if $U > C$ for right

tailed test. For left tailed test, we reject null hypothesis if $U < C$. For both tailed test, any of the two conditions needs to be satisfied at $\alpha/2$ level for acceptance of the alternative hypothesis.

8. Ansari-Bradley Test:

It is a non-parametric version of the Fisher's F Test. Inferences are drawn about the scale parameters (variance) of the two populations under study. Under the null hypothesis, it tests the equality of the scale parameters. According to the assumption, the populations must have an absolutely continuous distribution function. If X and Y be two populations under study, we define an indicator variable,

$$Z_i = \begin{cases} 1, & \text{if } i\text{'th combined sample} \\ & \text{is from population } X \\ 0, & \text{Otherwise} \end{cases}$$

The linear rank statistic is given by,

$$A_N = \sum_{i=1}^{n_1} \left| i - \frac{N+1}{2} \right| \cdot Z_i \quad (12)$$

Note. Here, N denotes the combined sample size ($n_1 + n_2$)

When C is the tabulated critical value of A_N at the specified level of significance α and combined sample size

$(n_1 + n_2)$, we reject null hypothesis if $A_N > C$ for right tailed test. For left tailed test, we reject null hypothesis if $A_N < C$. For both tailed test, any of the two conditions needs to be satisfied at $\alpha/2$ level for acceptance of the alternative hypothesis.

Table 1 outlines the application of parametric and non-parametric hypothesis tests in context of the present study.

*Table 1: Application of Hypothesis Tests w.r.t.
Research Questions*

Research Questions	Test Type	Parametric Tests	Non-Parametric Tests
1	One Sample Location	Student's t	Sign Wilcoxon Signed Rank
2	Two Sample Location	Fisher's t	Mann-Whitney U
3	Two Sample Variance	Fisher's F	Ansari-Bradley

3 Results and Discussion

Prior to the analysis, it is important to study the descriptive statistics. For this purpose, two summary tables of CGPA scores have been presented below.

Table 2: Basic Features of the Data

Group	Sample Size	Mean	Standard Deviation
Male	17	7.242	0.688
Female	13	7.589	0.367
Combined	30	7.393	0.589

Table 3: Five Number Summary

Group	Minimum	Q_1	Q_2	Q_3	Maximum
Male	5.300	7.000	7.400	7.520	8.500
Female	7.001	7.300	7.500	7.900	8.120
Combined	5.300	7.100	7.442	7.685	8.500

Note. Here, Q_1, Q_2, Q_3 denotes the first, second and third quartiles respectively.

Clearly the sample indicates that, on average, female students have performed better than their male peers. Additionally, the performance of the female students are comparatively less varied. But, to examine the significance of such claim at population level, we need to perform hypothesis tests. For this purpose, we need to know if

normality assumption is valid for the population under study.

From *Figure 1*, *Figure 2* and *Figure 3*, following interpretations can be deduced,

- 1.** The male students' sample as well as the combined sample contain outliers.
- 2.** Symmetricity is not largely compromised in any of the Histograms and KDE (Kernel Density Estimation) Plots. But, the Box Plot of male sample detects considerable difference between mean and median.
- 3.** In Quantile-Quantile Plots, no foolproof evidence of normality has been found.

For a more formal conclusion about normality behaviour, *Kolmogorov-Smirnov Test* has been applied with respect to the following hypotheses,

Null hypothesis:

$$H_0 : P = P_N$$

Alternative hypothesis:

$$H_1 : P \neq P_N$$

Note. Here, P denotes the empirical distribution under study and P_0 denotes normal distribution. The findings of the test have been summarized in Table 4.

CGPA Scores

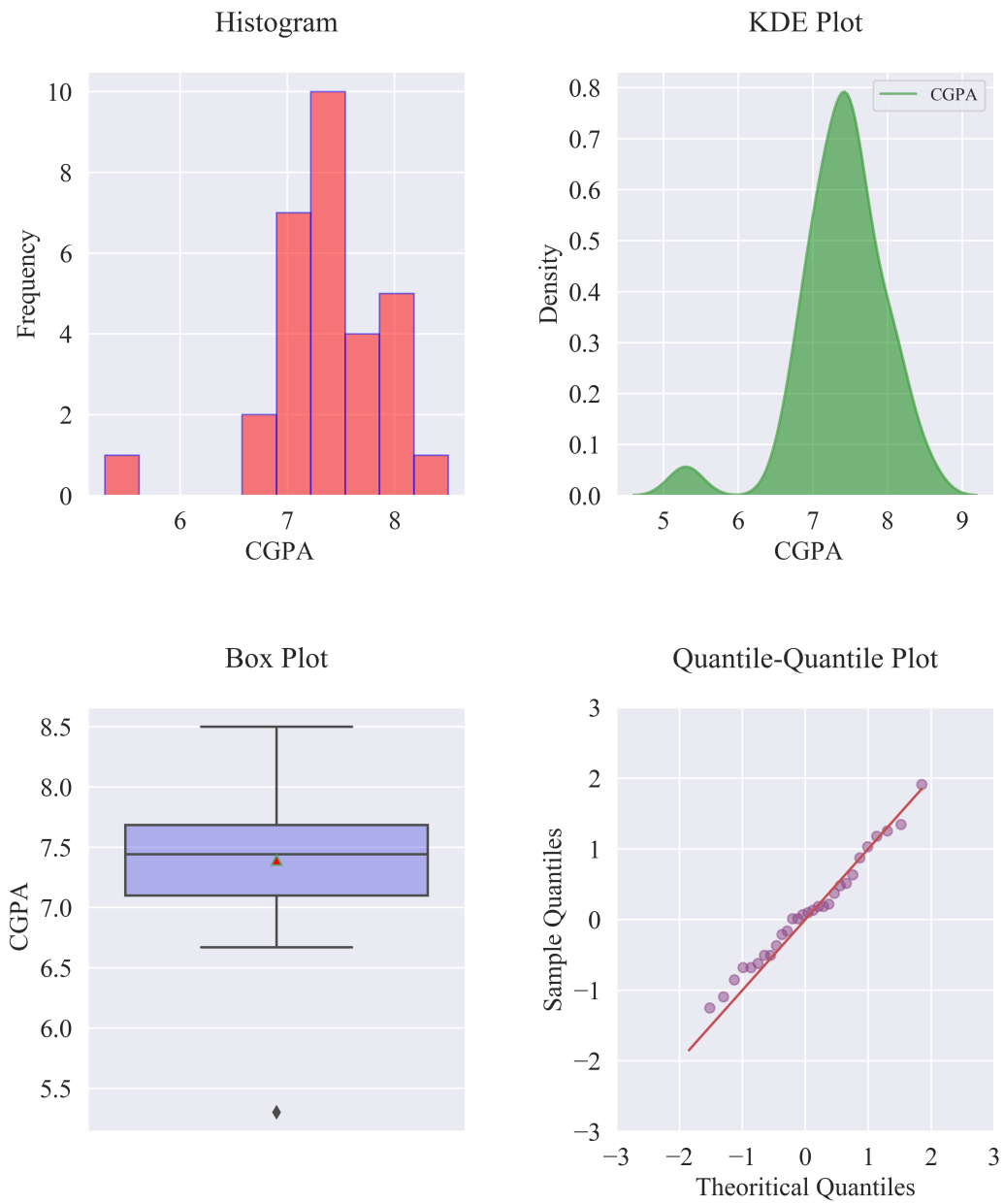


Figure 1: Exploratory Analysis of CGPA Scores

CGPA Scores (Genderwise)

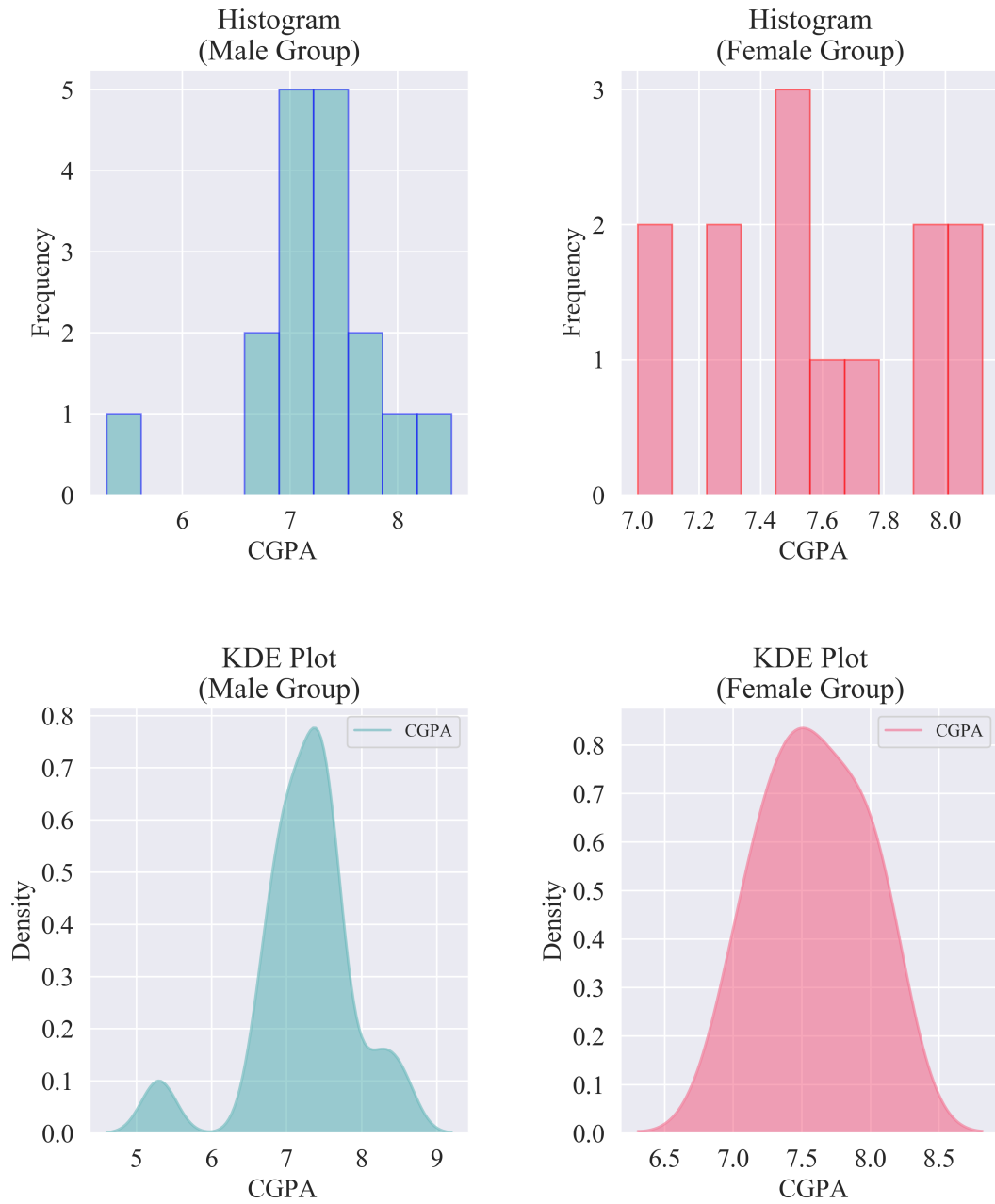


Figure 2: Histograms and KDE Plots of Male and Female Group

CGPA Scores (Genderwise)

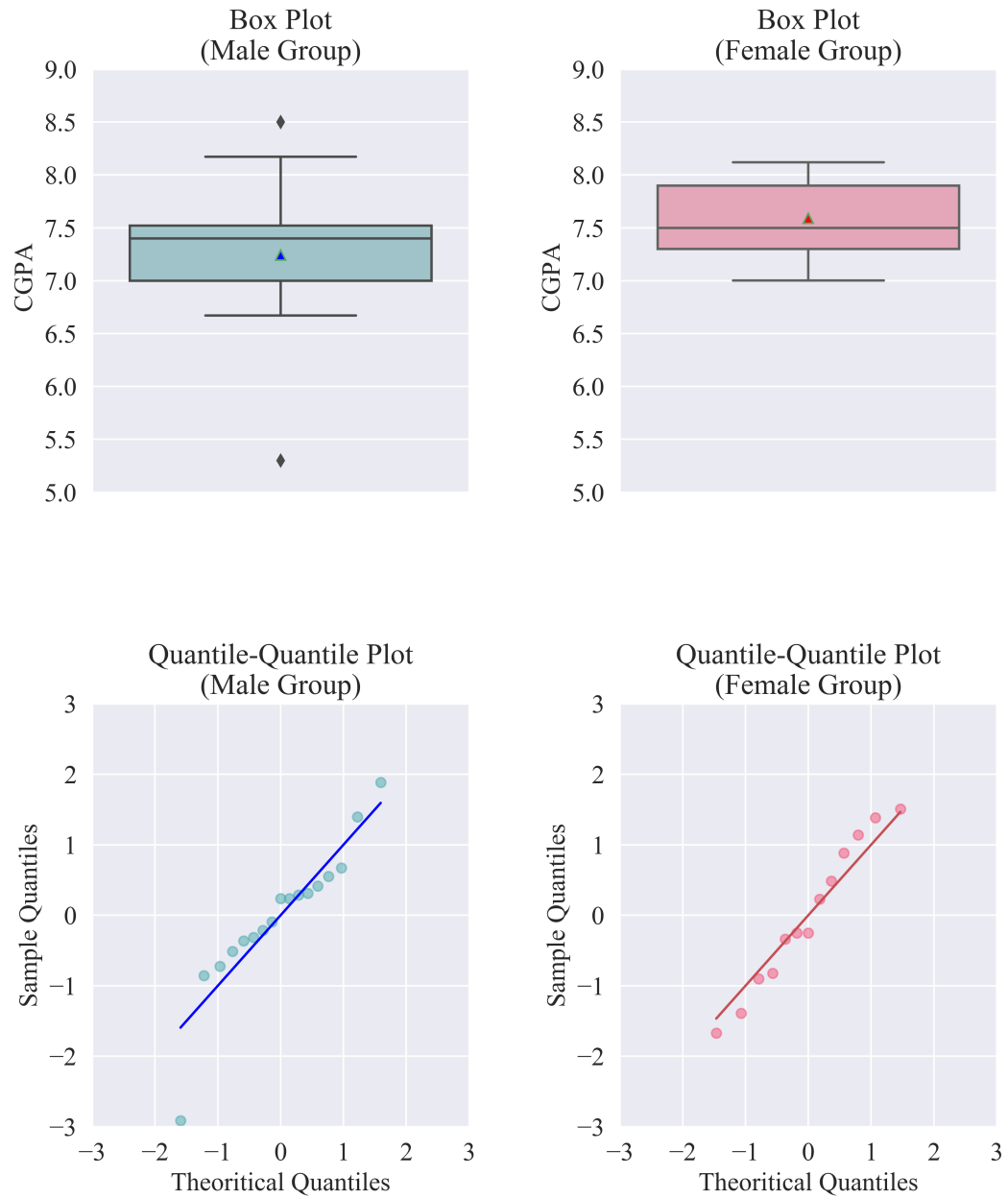


Figure 3: Box Plots and Q-Q Plots of Male and Female Group

Table 4: Results of Kolmogorov-Smirnov Test

Sample	Test Statistic	p-Value
Male	0.15115	0.8341
Female	0.12001	0.9868
Combined	0.11137	0.8548

Table 4 clearly depicts that, null hypotheses can not be rejected for any of the populations under study. Thus, no significant evidence of non-normality has been found. For this reason, we can safely incorporate parametric tests in our study. But, as a buffer against probable *Type II error* of Kolmogorov-Smirnov Test, non-parametric tests have also been conducted. *Table 5* contains the set of hypotheses used in the study.

Table 5: Declaration of Hypotheses w.r.t Reasearch Questions

Research Questions	Null Hypothesis	Alternative Hypothesis
1	$H_0 : \mu_0 = 7.5$	$H_1 : \mu_0 > 7.5$
2	$H_0 : \mu_F = \mu_M$	$H_1 : \mu_F > \mu_M$
3	$H_0 : \sigma_F = \sigma_M$	$H_1 : \sigma_F < \sigma_M$

Note. Here, μ_0 represents the center of the population (CGPAs scored by the students), μ_M, μ_F denote centers of the male and female populations under study. σ_M, σ_F are the respective standard deviations.

Based on our methodological framework, hypothesis tests have been performed. The results regarding the same have been summarized in *Table 6*.

Table 6: Results of Parametric and Non-Parametric Tests

Test	H_1	$p - Value$
Student's t	$H_1 : \mu_0 > 7.5$	0.8369
Sign Test	$H_1 : \mu_0 > 7.5$	0.9075
Wilcoxon Signed Rank	$H_1 : \mu_0 > 7.5$	0.6925
Fisher's t (Equal Variance)	$H_1 : \mu_F > \mu_M$	0.0557
Fisher's t (Unequal Variance)	$H_1 : \mu_F > \mu_M$	0.0439
Mann-Whitney U	$H_1 : \mu_F > \mu_M$	0.0434
Fisher's F	$H_1 : \sigma_F < \sigma_M$	0.0163
Ansari-Bradley	$H_1 : \sigma_F < \sigma_M$	0.2699

Note. Fisher's t Test has been done twice, one with equal variance and another with unequal variance assumption. H_1 denotes the alternative hypotheses respective to the tests.

In this study, level of significance (α) has been fixed at 0.05. From *Table 6*, it is clear that,

- 1.** Alternative hypothesis $H_1 : \mu_0 > 7.5$ can not be accepted for any of the tests under study.
- 2.** Alternative hypothesis $H_1 : \mu_F > \mu_M$ can be accepted for Fisher's t Test with unequal variance as-

sumption and also for Mann-Whitney U Test. But, Fisher's t Test with equal variance assumption fails to reject H_0 .

- 3.** Alternative hypothesis $H_1 : \sigma_F < \sigma_M$ can be accepted for Fisher's F Test. But, it's non-parametric alternative, i.e., Ansari-Bradley Test could not reject H_0 . Since, the parametric version is more powerful and as our data have not shown any significant non-normal behaviour, we accept the verdict of F test in this case.

4 Concluding Remarks

Thus, we reach at the following final remarks,

- 1.** In this study, we could not find sufficient evidence of average CGPA being more than 7.5 pointer.
- 2.** On average, female students performed even better than their male peers.
- 3.** Female examinees are found to be even more consistent in their performance than male candidates of the examination.

5 Conclusion

Thus, the findings of the study suggest that, emancipation of women has erased all concepts of feminine inferiority in terms of technological advancements. Rather, women have effortlessly habituated themselves with the new evaluation regime, so much so that they could supersede men. The reason can be attributed to their inherent capability of adaptation and naturalization. The study culminates on a positive note that, technology could not become a barrier against this exquisite quality of women.

6 References

Books

1. Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.
2. Hogg, R. V., McKean, J., & Craig, A. T. (2005). *Introduction to mathematical statistics*. Pearson Education.
3. Rohatgi, V. K., & Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.
4. Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (1993). *Probability and statistics for engineers and scientists* (Vol. 5). New York: Macmillan.

Web Links

1. <https://www.rdocumentation.org/>
2. <https://www.statsmodels.org/>
3. <https://docs.scipy.org/doc/scipy/>

Submitted on: January 13, 2022