2 DATA ANALYSIS

2.1 Descriptive statistics

2.1.1 cond. cond is a categorical variable. Therefore we calculate the frequency of each category. Frequency provides a count of the occurrences of each category in the data set and helps visualize the distribution of categorical variables.

cond

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	pm	27	49.1	49.1	49.1
	pw	28	50.9	50.9	100.0
	Total	55	100.0	100.0	

There are two categories: password memorised and password written. For the rest of the variables, we calculated the statistics based on each category.

2.1.2 successes. successes is a numerical variable that measures the number of successful logins over the study duration. The best way of showing central tendency is by getting the median which gets the average number of successful logins in a study duration. Calculating the range or standard deviation will show the variation of the data.

cond = pm

Statistics[a]

successes

N	Valid	27
	Missing	0
Mean		1.67
Std. Deviation		.620
Range		2

a. cond = pm

cond = pw

Statistics[a]

successes

N	Valid	28
	Missing	0
Mean		1.93
Std. Deviation		.466
Range		2

a. cond = pw

2.1.3 failures. failures is a numerical variable that measures the number of failed logins over the study duration. The best way of showing central tendency is by getting the median which gets the average number of failed logins in a study duration. Calculating the range or standard deviation will show the variation of the data.

cond = pm

Statistics[a]

failures

N	Valid	27
	Missing	0
Mean		1.26
Std. Deviation		1.483
Range		4

a. cond = pm

cond = pw

Statistics[a]

failures

N	Valid	28
	Missing	0
Mean		.61
Std. Deviation		2.315
Range		12

a. cond = pw

2.1.4 resets. resets is a numerical variable that measures the number of password resets over the study duration. The best way of showing central tendency is by getting the median which gets the average number of password resets in a study duration. Calculating the range or standard deviation will show the variation of the data.

cond = pw

Statistics[a]

resets

N	Valid	28
	Missing	0
Mean		.04
Std. Deviation		.189
Range		1

a. cond = pw

cond = pm

Statistics[a]

resets

N	Valid	27
	Missing	0
Mean		.30
Std. Deviation		.609
Range		2

a. cond = pm

2.1.5 slogintime. slogintime is a numerical variable that measures the average time taken to log in successfully. The best way of showing central tendency is by getting the mean which gets the average time it takes a user to log in successfully. Calculating the range or standard deviation will show the variation of the data.

cond = pw

Statistics[a]

slogintime

N	Valid	28
	Missing	0
Mean		9.8154
Std. Deviation		3.40427
Range		12.00

a. cond = pw

cond = pm

Statistics[a]

slogintime

N	Valid	27
	Missing	0
Mean		9.0370
Std. Deviation		6.88215
Range		32.50

a. cond = pm

2.1.6 maxmemtime. maxmemtime is a numerical variable that measures the longest duration over which it could be shown that a participant positively remembered their password. The best way of showing central tendency is by getting the mean which gets the average longest time that a participant positively remembered their password. Calculating the range or standard deviation will show the variation of the data.

cond = pm

Statistics[a]

maxmemtime

N	Valid	27
	Missing	0
Mean		169.0011
Std. Deviation		94.63317
Range		551.24

a. cond = pm

cond = pw

Statistics[a]

maxmemtime

N	Valid	28
	Missing	0
Mean		208.1464
Std. Deviation		95.81475
Range		569.03

a. cond = pw

2.1.7 totallastingsuccess. Although a number, the totallastingsuccess variable is a binary variable so it can be considered a categorical variable. Therefore we calculate the frequency of each category. Frequency provides a count of the occurrences of each category in the data set and helps visualize the distribution of categorical variables.

totalLastingSuccess[a]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	6	22.2	22.2	22.2
	1	21	77.8	77.8	100.0
	Total	27	100.0	100.0	

a. cond = pm

totalLastingSuccess[a]

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4	14.3	14.3	14.3
	1	24	85.7	85.7	100.0
	Total	28	100.0	100.0	

a. cond = pw

2.1.8 avgLeadingFailures. avgLeadingFailures is a numerical variable that measures the average number of times that the user had a failed password attempt before successfully logging in. The best way of showing central tendency is by getting the mean which gets the average time a user had a failed password attempt. Calculating the range or standard deviation will show the variation of the data.

cond = pm

Descriptive Statistics[a]

	N	Range	Mean	Std. Deviation
avgLeadingFailures	27	2.0	.500	.7206
Valid N (listwise)	27			

a. cond = pm

cond = pw

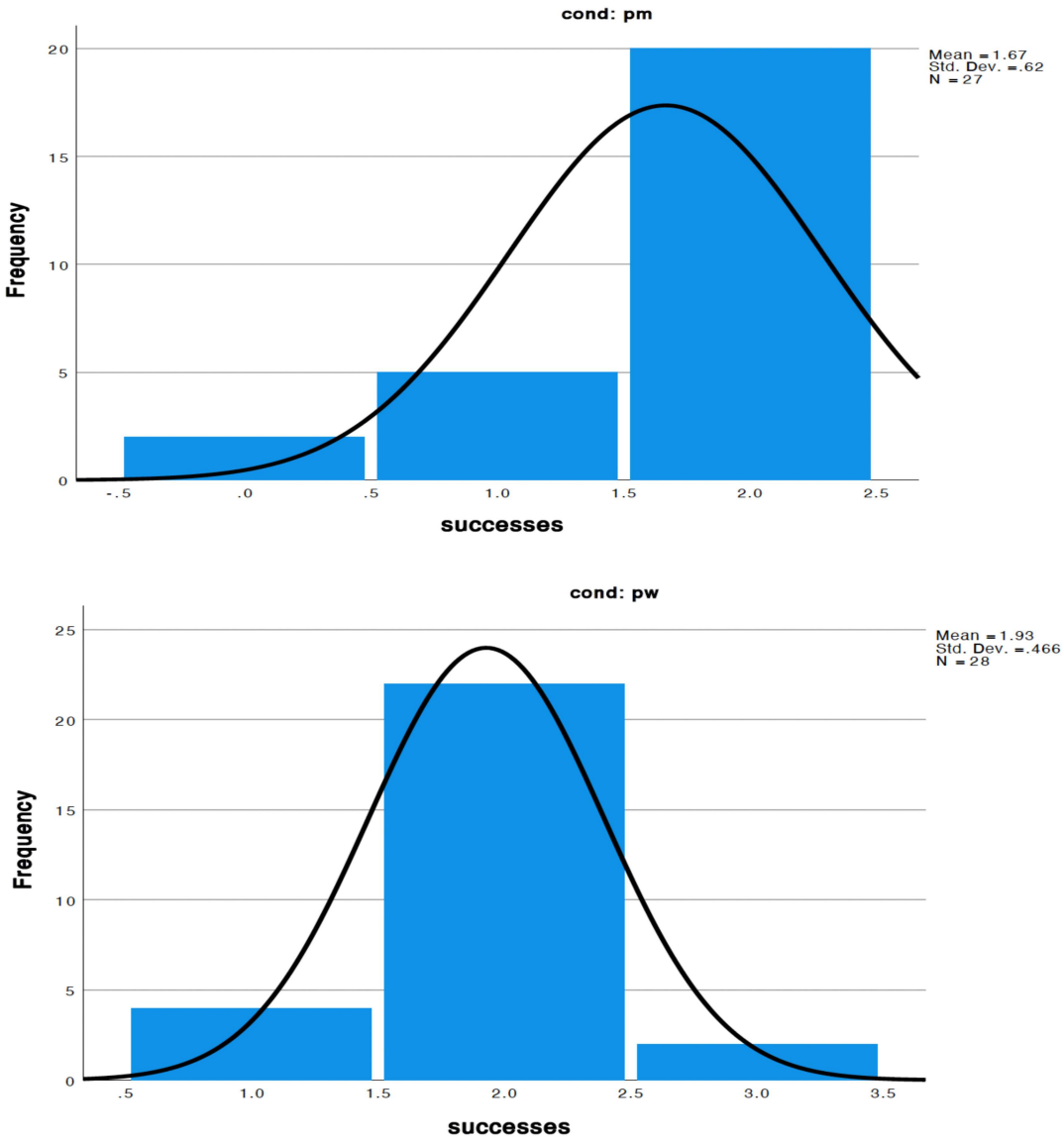
Descriptive Statistics[a]

	N	Range	Mean	Std. Deviation
avgLeadingFailures	28	6.0	.286	1.1420
Valid N (listwise)	28			

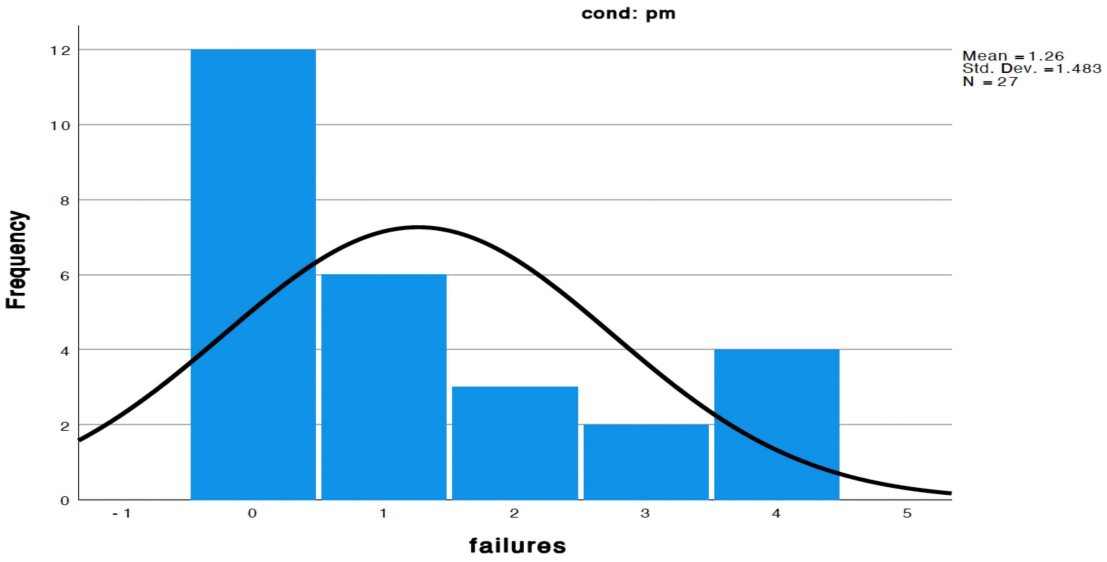
a. cond = pw

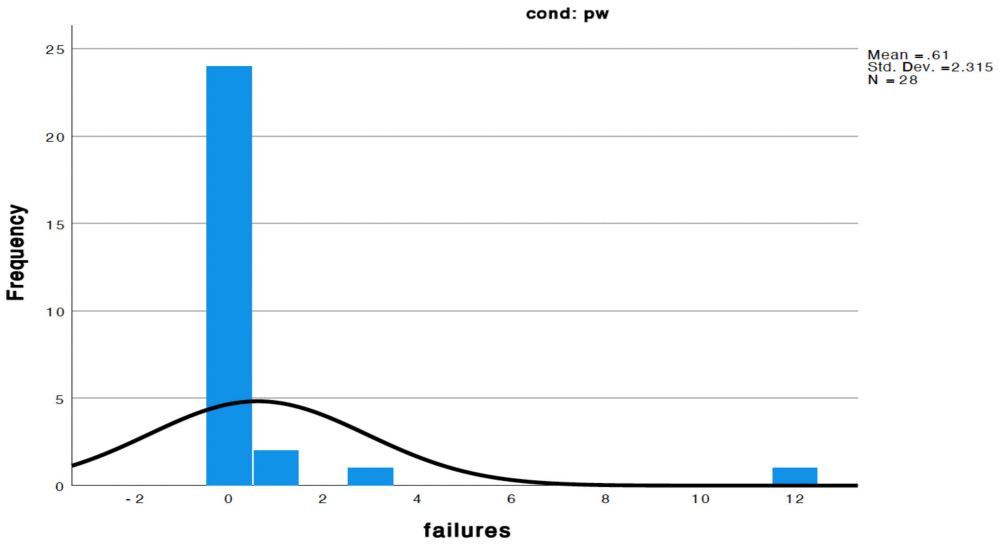
2.2 Data visualization

2.2.1 successes.

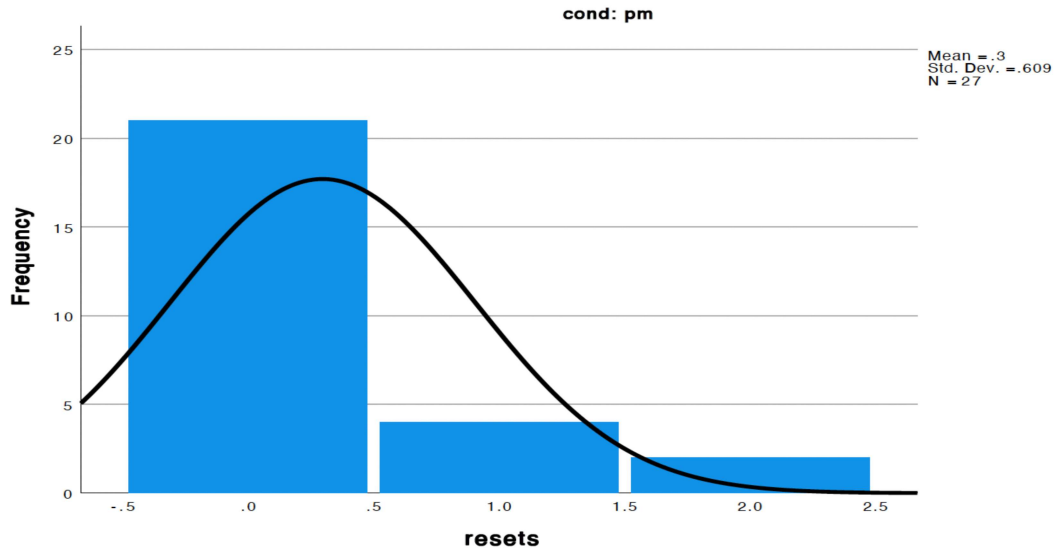


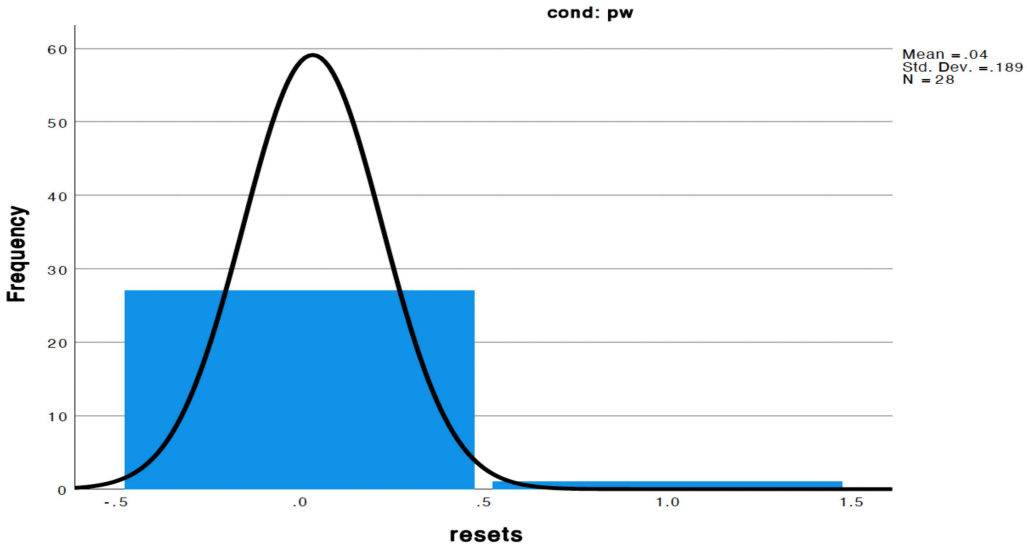
2.2.2 *failures.*



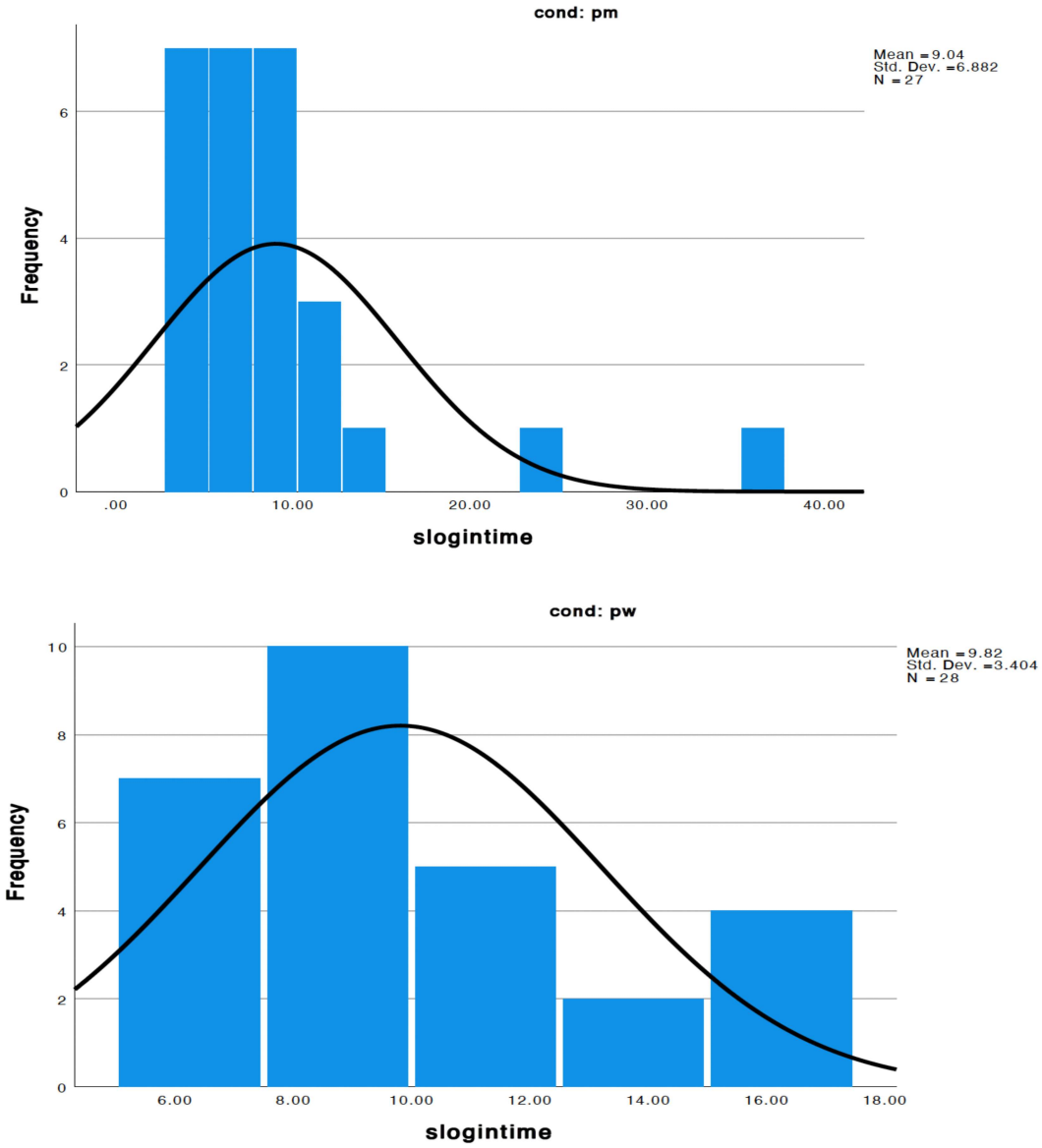


2.2.3 resets.

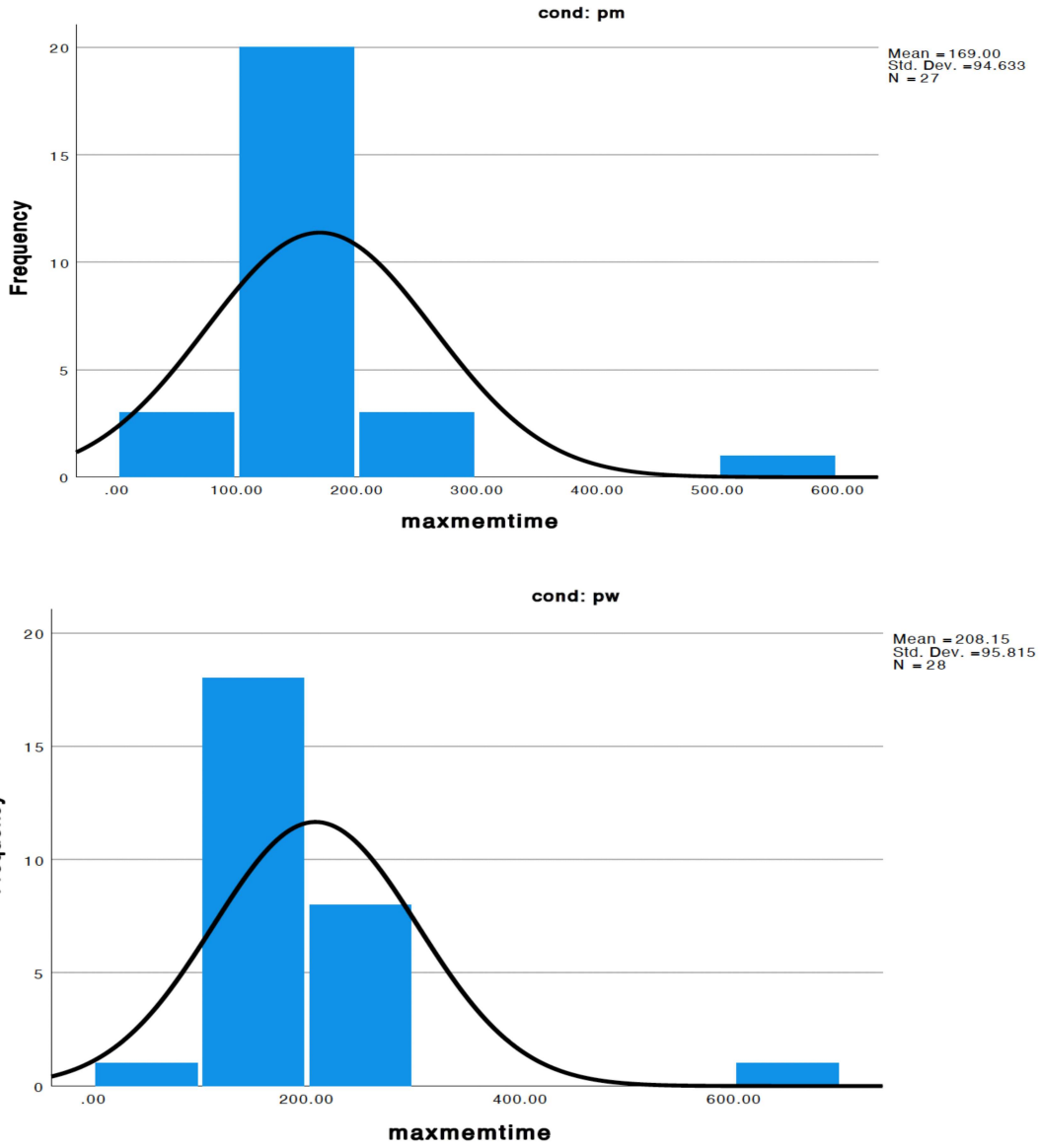

cond: pm
Mean = .3
Std. Dev. = .609
N = 27
Frequency
25
20
15
10
5
0
-.5
.0
.5
1.0
1.5
2.0
2.5
resets


cond: pw
Mean = .04
Std. Dev. = .189
N = 28
Frequency
60
50
40
30
20
10
0
-.5
.0
.5
1.0
1.5
resets

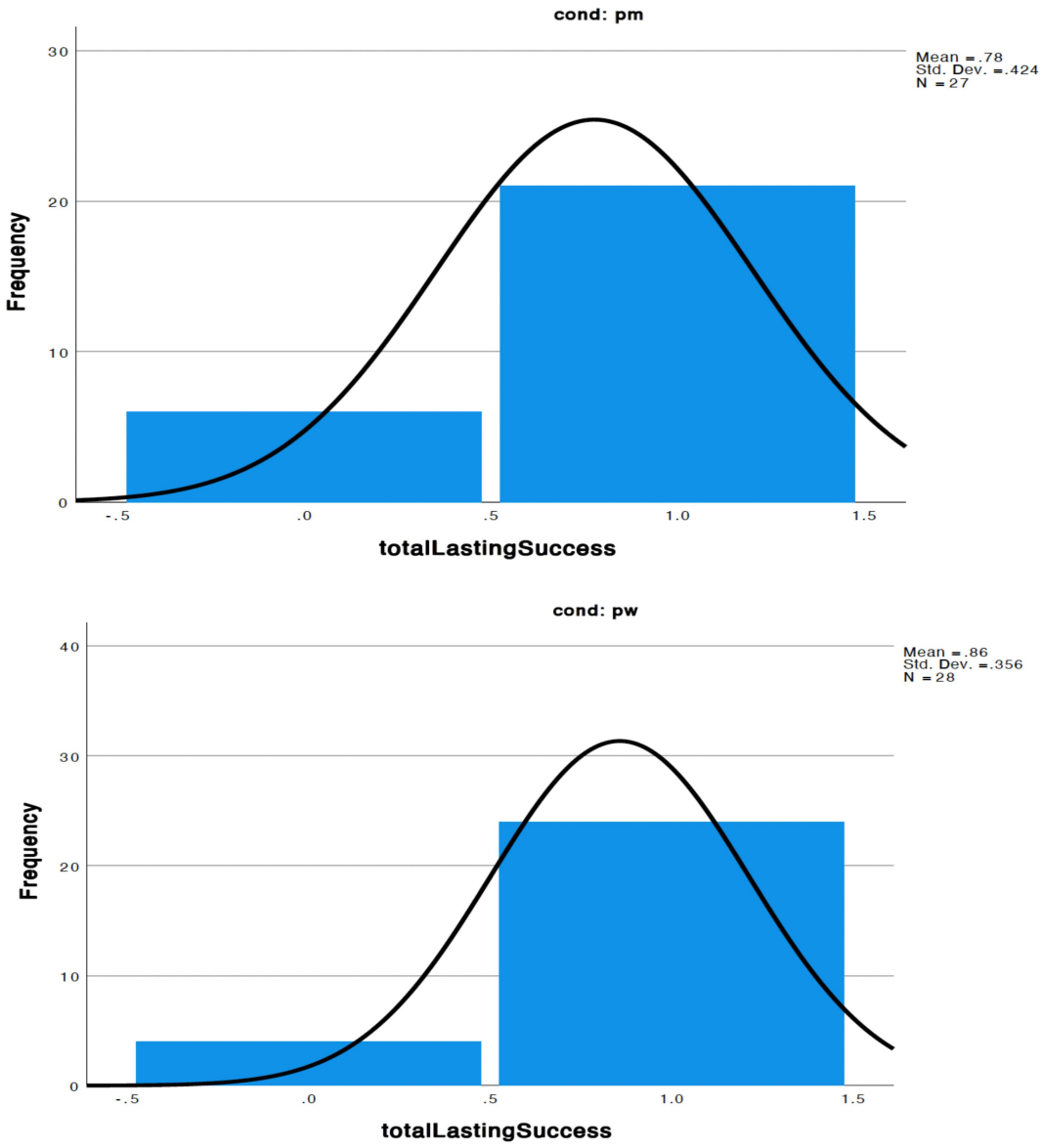
2.2.4 *slogintime.*



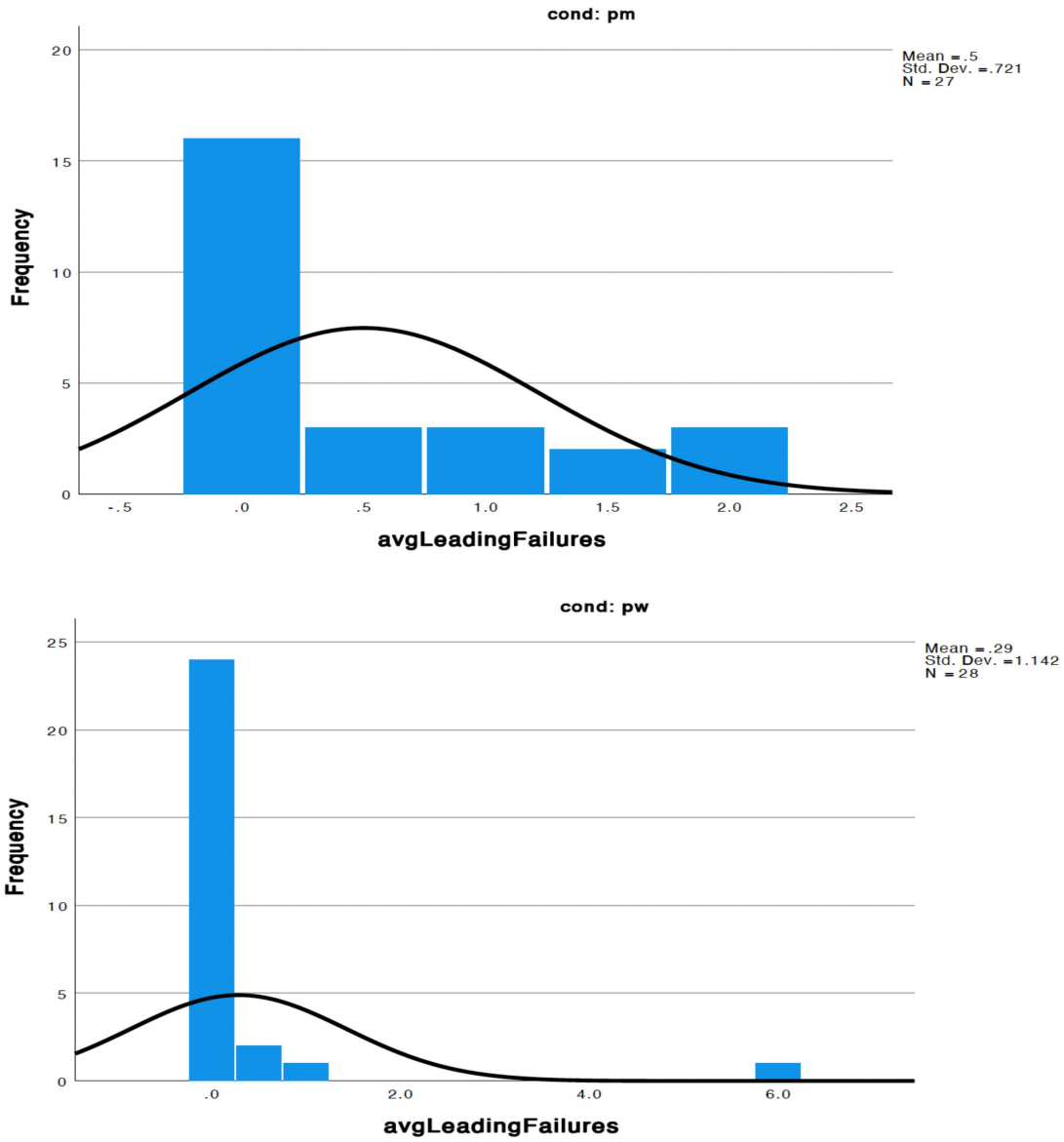
2.2.5 *maxmemtime.*



2.2.6 *totallastingsuccess.*



2.2.7 *avgleadingfailures.*



2.3 Inferential statistics

2.3.1 *Variable: Successes.* Hypothesis:

H0: There is no difference in the average number of successes between participants who memorized their password and those who wrote it down.

H1:There is a significant difference in the average number of successes between participants who memorized their password and those who wrote it down.

Statistical Test: Mann-Whitney U test

Justification: This test is appropriate for comparing means of two independent groups in which normality is not assumed. Success is a discrete variable and we are comparing the average number of successes.

Result: Using SPSS, p-value(p) = 0.123 U = 446 Since the p - value is greater than 0.05, we fail to reject the null hypothesis Therefore, there is no difference in the average number of successes between participants who memorized their password and those who wrote it down.

2.3.2 *Variable: Failures.* Hypothesis:

H0: There is no difference in the average number of failures between participants who memorized their password and those who wrote it down.

H1:There is a significant difference in the average number of failures between participants who memorized their password and those who wrote it down.

Statistical Test: Mann-Whitney U test

Justification: This test is appropriate for comparing means of two independent groups in which normality is not assumed. Failure is a discrete variable and we are comparing the average number of failures.

Result: Using SPSS, p-value(p) = 0.002 U = 223.000 Since the p - value is lesser than 0.05, we reject the null hypothesis Therefore, there is a significant difference in the average number of failures between participants who memorized their password and those who wrote it down.

2.3.3 *Variable: Resets.* Hypothesis:

H0: There is no difference in the average number of resets between participants who memorized their password and those who wrote it down.

H1:There is a significant difference in the average number of resets between participants who memorized their password and those who wrote it down.

Statistical Test: Mann-Whitney U test Justification: This test is appropriate for comparing means of two independent groups in which normality is not assumed. The distribution is skewed. Resets is a discrete variable and we are comparing the average number of resets.

Result: Using SPSS, p-value(p) = 0.037 U = 306.500 Since the p - value is lesser than 0.05, we reject the null hypothesis Therefore, there is a significant difference in the average number of resets between participants who memorized their password and those who wrote it down.

2.3.4 *Variable: slogintime.* Hypothesis:

H0: There is no difference in the average time taken to log in successfully between participants who memorized their password and those who wrote it down. H1:There is a significant difference in average time taken to log in successfully between participants who memorized their password and those who wrote it down. Statistical Test: Mann-Whitney U test

Justification: The Mann-Whitney U test is appropriate for comparing distributions of two independent groups when the dependent variable is continuous but not normally distributed. The graph of slogintime is left-skewed.
Result: Using SPSS, p-value(p) = 0.32 U = 505.000 Since the p - value is greater than 0.05, we fail to reject the null hypothesis Therefore, there is no significant difference in average time taken to log in successfully between participants who memorized their password and those who wrote it down.

2.3.5 *Variable: maxmemtime.* Hypothesis:

H0: There is no difference in the longest duration over which it could be shown that a participant positively remembered their password between participants who memorized their password and those who wrote it down.
H1:There is a significant difference in average time taken to log in successfully between participants who memorized their password and those who wrote it down.
Statistical Test: Independent t-test
Justification: Slogintime is a continuous variable, and the t-test is appropriate for comparing means of normally distributed data between two independent groups. The graph of maxmemtime shows a normal distribution so a t-test is valid.
Result: Using SPSS, p-value(p) = 0.133 t = 1.524 Since the p - value is greater than 0.05, we fail reject the null hypothesis Therefore, there is no difference in the longest duration over which it could be shown that a participant positively remembered their password between participants who memorized their password and those who wrote it down.

2.3.6 *Variable: totalLastingSuccess.* Hypothesis:

H0: There is no association between password condition (memorized vs. written) and the likelihood of using the same password for the duration of the study.
H1:There is a significant association between password condition (memorized vs. written) and the likelihood of using the same password for the duration of the study.
Statistical Test: Fisher's test
Justification: Fisher's test is appropriate for comparing the distribution of a binary variable between two groups. Totallastingsuccess is binary/categorical so it's a valid test.
Result: Using SPSS, p-value(p) = 0.503 Since the p - value is greater than 0.05, we fail reject the null hypothesis Therefore, there is no association between password condition (memorized vs. written) and the likelihood of using the same password for the duration of the study.

2.3.7 *Variable: avgLeadingFailures.* Hypothesis:

H0: There is no difference in average number of times that the user had a failed password attempt before successfully logging in between participants who memorized their password and those who wrote it down.
H1:There is a significant difference in average number of times that the user had a failed password attempt before successfully logging in between participants who memorized their password and those who wrote it down.
Statistical Test: Mann-Whitney U test Justification: This test is appropriate for comparing means of two independent groups in which normality is not assumed. The distribution is skewed so this test is valid for this situation. Result: Using SPSS, p-value(p) = 0.026 U = 274.500 Since the p - value is lesser than 0.05, we reject the null hypothesis Therefore, there is a significant difference in average number of times that the user had a failed password attempt before successfully logging in between participants who memorized their password and those who wrote it down.

2.4 Discussion

The research study aimed to investigate how password recording, specifically the act of writing it down, influences the memorability and usability of text passwords. The study involved two conditions: one where participants were instructed to write down their passwords and another where they were instructed not to. The data analysis integrated descriptive statistics, graphical representations, and inferential statistics to provide insights into potential differences between the groups.

Participants who wrote down their passwords had fewer failures and resets, suggesting better usability in this aspect. While the impact on memorability is not the most, the study suggests that allowing users to write down their passwords might improve usability by reducing the number of failures and resets, and in turn increasing the number of successes. Perhaps a wider number of participants and more open-ended questions could establish a clearer impact. Participants who wrote down passwords may have had a more tangible reference for successful login, leading to fewer failures and resets. The act of writing down passwords eased recognition process.

Graphical representations also backed up the pre-mentioned observations. The bar charts visually depict the distribution of successes, failures, resets, slogintime maxmemtime, totalLastingSuccess and avgLeadingFailures across the two conditions.

Inferential statistics, employing Mann-Whitney U tests, Fisher's test and independent t-tests corroborated the findings. The Mann-Whitney U tests indicated significant differences in failures and resets, with participants that memorised encountering more problems. The Fisher's tests and independent t-test demonstrated no noticeable association between password recording and the likelihood of using the same password throughout the study.

The differences in usability metrics suggest that allowing users to write down passwords has a meaningful impact on reducing the number of failures and password resets during login attempts. The act of recording passwords may serve as a practical aid, providing users with a tangible reference that enhances the accuracy of their login attempts. This finding is particularly relevant in real-world scenarios where the usability of password systems directly influences user experience.

Because the memorability-related variables don't have a very noticeable difference, (eg. maxmemtime and totallastingsuccess), we can assume that the act of writing down passwords does not significantly affect participants' ability to remember their passwords over the short term.

In conclusion, the study's implications highlight the relationship between password recording and user experience. While the impact is not much, the fewer failures and resets from the password written suggest that allowing users to write down passwords should be the preferred method.