International Journal of
Strategic Research in Education, Technology and Humanities
p-ISSN: 2465-731X | e-ISSN: 2467-818X
IJSRETH
October, 2024
Vol. 12, No. 2

Academic Research Analysis: Fundamentals, Quality, and Contextually Productive Application

Tombari James
Department of Industrial Technology
Federal College of Education (Technical), Omoku
PMB 11, Omoku, Rivers State, Nigeria.

Article DOI:
10.48028/iiprds/ijsreth.v12.i2.05

Keywords:
Survey analysis,
Data analysis,
Nonparametric tests,
Likert Scale datasets

Corresponding Author:
Tombari James

Abstract

To mitigate the challenges faced by early researchers in the analyses and presentation of their survey findings, this article explains and illustrates critically fundamental aspects of academic research data analyses. The contextual presentation, and illustration are an asset to the student-researcher or instructor who desires to quickly look up how to interpret: Likert Scale survey results, Difference tests, and correlation r-values. Also, this applied research clearly explains the concept of parametric and non-parametric tests using case studies to illustrate when non-parametric tests such as Mann-Whitney U test, Wilcoxon Signed Test, Kruskal-Wallis test, Chi square test, Spearman Rank correlation, and Puri and Sen test can be used. These are sequentially organized to promote constructive research manuscripts and contextually productive application by academics. A review of over 45 correlation analysis in published articles as presented in the findings revealed the need to call for standardization in the interpretation/reporting of educational surveyed data. Also, a statistically consistent approach to interpreting the mean of a Likert Scale test result was illustrated. The choice to maintain a standard/universal approach to the interpretation of test results is a choice to advance, standardize, and promote logical consistency, as well as quality in academic research analysis reporting.

Background of the Study

Academic research comprises a statement of specific objectives, research questions, and often a test of hypotheses concerning an identified problem. To achieve these, the researcher is expected to outline clear and achievable objectives (known as specific objectives); the pursuit of these objectives leads to research questions tailored to each specific objective; finally, the researcher might have an expected outcome or there might be a generally expected standard outcome/value against which the researcher would postulate and test the null hypotheses. Based on the findings to the research questions, and test of hypotheses, the researcher highlights descriptions and or make inferences then draw conclusions. Therefore, academic research requires a working knowledge of the *educational discipline*, some application of *psychology*, and indeed a working knowledge of *test statistics*. The combination of the above trio has often proved to be quite daunting to early researchers, perhaps because concise, contextual, and comprehensive guidelines are not easily available. Also, some experienced educators find it intimidating to interpret test statistic results. This paper attempts to simplify all that, to promote qualitative research, effective collaboration, and socio-economic usefulness of academic research publications.

Gasca, (2016) explained the need for quietness to promote creativity in today's ever buzzling world. Same thing applies to research. Important to point out is that the purpose of this applied research is to provide constructive approaches and engender progressive and advanced academic research reports, not to allot blames or condemnations; moreover, whatever we know today, we have learned from someone or something, and knowledge is progressive.

Aim & Objectives

1) Likert Scale: Data gathering, interpretation and decision making
2) Parametric and non-parametric tests
3) Case study: involving use of Mann-Whitney Utests, Wilcoxon Test, and Kruskal-Wallis
4) Difference test: p-values interpretation and hypothesis testing
5) Correlational test: R-values interpreation and hypthoses

Likert Scale: Data gathering and decision making

The Likert scale is a system of assigning weights to choices. A target audience can express the degree/level to which they accept/like/agree-with specific statements offered as a potential answer to specific question-items posed by a researcher.

As a researcher your research topic would typically have two or more objectives. To investigate these objectives, you would set research questions. Based on your copious review of literature and intuition, you would then provide tentative statements that could be answers to the research questions. Through the research instrument, your respondents can communicate their opinion about what potential answers they consider most valid using the Likert Scale that you have provided. When you retrieve the answered questionnaire, you would have a certain number of responses.

Each Likert scale is assigned an arbitral but logical and sequential weight. Typically, an odd number of options is recommended, this is to make the median response obvious. Thus, five scales (or options) can be chosen thus: Strongly Agree (having a weight of 5), Agree (having a weight of 4), Undecided (having a weight of 3), Disagree (having a weight of 2), and Strongly Disagree (having a weight of 1).

The responses retrieved from the target audience is called data. This data needs to be presented first in Tables to facilitate computations and reveal simple relationships or differences. This is achieved by creating a frequency table as illustrated in Table 1. This frequency (or contingency) Table is built by counting and recording the number of respondents that chose Strongly Agree (SA) for question-item 1. Repeat this step for Agree (A), Undecided (UD), Disagree (D), and Strongly Disagree (SD). Repeat the same for the remaining statements and research questions. Having performed this step, you would have a Table like Table 1.

Table 1: A frequency table of audience Likert scale responses

Questionnaire-item	Statement	SA	A	UD	D	SD	X	Decision
1	abc	2	5	1	1	1	3.60	
…	xyz	…	…	…	…	...	…	
n	jkl	0	1	2	3	4	2.00	

SA = Strongly Agree; A = Agree; UD = Undecided; D = Disagree; SD = Strongly Disagree

The mean (X) illustrated in Table 1. is obtained by:

1) Multiply the count of the option by the weight of the option e.g.,
   SA:
   2×5
2) Repeat step #1 for all other Likert Scale options e.g.,

A	UD	D	SD
5×4	1×3	1×2	1×1

3) Add up the value obtained from steps 1 and 2 e.g.,
   $10 + 20 + 3 + 2 + 1 = 36$
4) Divide the value obtained in step 3 by the total number of respondents e.g.,
   $\frac{36}{10} = \mathbf{3.60}$
   $\therefore \bar{x} = \mathbf{3.60}$

Likert Scale: Interpretation

Having created a contingency table (see Table 1.), the implication of a mean of 3.38, 3.42, 1.40, or 4.25, etc can be determined. To sustain a common standard, we need to submit to a statistical approach. The range of data needs to be determined. This refers to the difference between the highest and the lowest weight. Thus, the range for a 5-point Likert scale would be 4 (i.e., 5 – 1). Since the mean is often computed to 2 places of decimal, there is need to use an interpretation that accounts for intervals between our chosen weights. This interval is obtained by dividing the range by the number of weights. Thus, for this scenario, the interval would be 0.8 (i.e., 4 5). With this approach, a common decision (interpretation) would be arrived at by researchers irrespective of discipline, geographical location or institutional affiliation (see Table 2).

Table 2: Likert Scale Interpretation and Decision Making

Mean Score interval	Decision
4.21 – 5.00	Strongly Agree
3.41 – 4.20	Agree
2.61 – 3.40	Undecided
1.81 – 2.60	Disagree
1.00 – 1.80	Strongly Disagree

Table 2 illustrates how the Likert scale data scores should be interpreted. The *Mean Intervals* was achieved by:

1) Start from the least weight (in this case, 1.00)
2) Add the interval to the least weight (i.e., 1.00 + 0.80)
3) Begin the next level by adding 0.01 to the upper limit of the previous level (i.e., 1.80 + 0.01). The upper limit of this level is obtained by adding the interval to the upper limit of the previous level (i.e., 1.80 + 0.80).
4) Repeat step 3 until you arrive at the highest weight (Strongly Agree = 5.00).

By following the above steps, you would obtain a table similar to Table 2 with which you can interpret the average opinion of your respondents. Hence, the decision for statement #1 which had a X of 3.6 would be "Agree" (cf. Table 1. with Table 2). That is, "the respondents *agreed* with statement-abc" whereas, the respondents are "undecided" about the statement-xyz. Therefore, the Likert scale data which was gathered and presented in Table 1 is best interpreted as shown in Table 3.

Table 3.

Questionnaire-item	Statement	SA	A	UD	D	SD	X	Decision
1	abc	2	5	1	1	1	3.60	Agreed
…	xyz	1	4	2	1	3	3.20	Undecided
n	jkl	0	1	2	3	4	2.00	Disagree

Data gathered from the researcher's reputable respondents, analysed and presented in Table 3 can be interpreted thus:

The respondents *agreed* that abc is a Nigerian food. The respondents were undecided whether xyz is a Nigerian food. Lastly, the respondents disagree that jkl is a Nigerian food. This decision approach is indeed different from what has often been found in some literatures. It becomes obvious that using Table 2 to decide on the mean response is more appropriate. This approach is consistent with the views of (DATAtab, 2024; MATHStorya, 2023). Also, if the options/ratings are more than five or less, Table2 can be customized appropriately.

Note that, surveys are constructed and administered with the expectation that the respondents are custodians of knowledge on the subject matter for which the research questions pertain. Hence, the interpretation of these average findings could be generalized as the answers/solutions/ratings to the questions asked. For this reason, the researcher owes it as good practice to seek responses from individuals or organizations who are indeed practitioners or custodians of the knowledge being investigated.

Case Study #1

Table 4 below illustrates a set of five (5) point ratings and their corresponding weights which can be applied to a questionnaire.

Table 4: An Example 5-point Weighting Scale

Rating	Weight
Very Good (VG)	5
Good (G)	4
Fair (F)	3
Poor (P)	2
Very Poor (VP)	1

Example Research Question: How do you rate your University/College on the following services?

Table5: A Sample University/College Service Questionnaire

University/College Services	VG	G	F	P	VP
Student-support					
Library					
Sports facilities					
Hostel Accommodation					
Lecture theatres					
Examination halls					
Environmental aesthetics					
Transportation services					
Sanitation and environment					
Electricity supply					
Water supply					
24hrs security					
Electrical Laboratory					
Electronics Laboratory					
Power Systems Laboratory					
Information Technology					

Parametric and Non-Parametric Tests

Nonparametric tests are performed on observed data which are neither interval or ratio. In other words, nonparametric tests are performed on categorical data, for example Likert Scale data, gender (male or female), groups (group1, group2, etc), age-grade, etc. On the other hand, parametric tests are performed on observed data which are interval or ratio, for example age (in months or years), body mass, distance, time (in seconds, minutes, hours, or years), weight (in grams, or kilogram), height (in metres or feet), electrical power (in watts or kilowatts), current capacity (mAh), water pH, etc.

Typically, nonparametric tests require the original observations to be transformed (Harwell, 1988). The most common of which is a rank transformation. Whichever test is chosen, it is desirable to mitigate type 1 errors. In research statistics, type 1 error refers to rejecting the null hypothesis in favour of the alternate whereas, the null hypothesis was valid. That is, stating that there exists a significant difference whereas, there was actually no difference. According to (Kasahara, 2020) a type 1 error implies rejecting the null-hypothesis whereas the null-hypothesis was actually true.

While some authors argue that larger sample size can mitigate type 1 errors, Harwell (1988) inform that the appropriate sample size depends on the robustness properties of the chosen statistical tests. For example, a sample size of less than 10 per group may be satisfactory (i.e., has a very small chance of causing a type 1 error) even when normality assumptions are violated for an ANOVA test, but a sample size of less than 30 per group for an ANCOVA test would most likely lead to type 1 error when normality assumptions are violated.

Nonparametric tests on the other hand, have the desirable feature of being able to control type 1 error rates for both normal and non-normal distributions, and equal and unequal sample sizes. Despite this benefit, most educational researchers are prone to choose parametric tests over nonparametric tests. This is because, over the years, most parametric tests have been said to be robust even when normality assumptions are violated by the given dataset. Nonetheless, when the assumption of normal distribution is violated, then the probability of a type 1 error is greatly increased, the test statistic robustness notwithstanding (Harwell, 1988). Therefore, it is desirable that an educational researcher always determine if the data he/she has gathered has a normal distribution or not. If the data has a normal distribution, then parametric test might be used. But if it has non-normal distribution then a non-parametric test might be used. Tests for normality include the Shapiro-Wilk test, Kolmogorov-Smirnov test, histogram, and box plot which can be achieved using IBM's® Statistical Package for the Social Sciences (SPSS).

Furthermore, Harwell (1988) inform that the nature of data alone is not a valid determinant of what type of test to use, rather the fit between the test and the data is a valid determinant. That is, a test should be performed after cross-examining the underlying test assumptions (is the test meant for parametric or for non-parametric data? What is the minimum sample size per group?).

Difference Test: P-values Interpretation

In academic survey, the null hypothesis is usually tested at an alpha-value of 0.05 (also called 5%). This alpha-value is a standard and generally accepted value. A difference test cross-examines the mean of two samples to determine if there is a significant difference. When the difference between the mean of more than two samples are examined at a time it is called an Analysis of Variance (ANOVA).

Generally, it is accepted that for the null hypothesis to be rejected, the p-value must be less than the alpha-value. That is; after stating the null hypothesis, the gathered data would be subjected to an appropriate difference test. If the p-value (i.e., probability value) from the test is less than the alpha-value, then the null hypothesis would be rejected, and the alternate hypothesis accepted. Note that the alternate hypothesis is the opposite statement of the null hypothesis.

Thus, if the test returned a p-value of 0.03, this is obviously less than 0.05. Therefore, the null hypothesis must be rejected and the alternate hypothesis accepted. Since the null hypothesis always posits that there is no significant difference, then a p-value that is less than the alpha-value implies that there is a significant difference.

Correlational Test: R-Values Interpretation

All correlational tests return a correlational strength and direction between each pair of the investigated variables. This value ranges from -1 to 1. And this is one of the critical purposes of a correlational analysis. To investigate the strength of correlation between the variables of interest and the direction in which these variables are related. With such a range comes the need for interpretation. The result of a correlation test has both a p-value (significance level of the analysed data) and an r-value (strength of the correlation and direction of the correlation between the variables under investigation).

The purpose of illustrating Table 6 below is to promote universality, standard, and a statistically consistent approach in the interpretation of r-values. In Table 6, seven (7) correlational strengths are considered (Very strong, Strong, above moderate, Moderate, Weak, Very Weak, and No Correlation). With an odd-number strengths of correlation, a median level becomes obvious to select.

Table6: Measured Strength of Relationship

****Absolute (r-value)**	**Correlation Strength**	**Direction**
0.81 – 1.00	Very Strong	Positive or Negative
0.61 – 0.80	Strong	Positive or Negative
0.41 – 0.60	above moderate	Positive or Negative
0.21 – 0.40	Moderate	Positive or Negative
0.14 – 0.20	Weak	Positive or Negative
0.08 – 0.13	Very weak	Positive or Negative
0.00 – 0.07	No	Positive or Negative

** A negative r-value means a negative correlation
A positive r-value means a positive correlation.

Based on Table 6, r-values of 0.06, 0.12, 0.19, 0.3, 0.5, 0.7, and 0.9 for example, would be interpreted as: No correlation, very weak correlation, weak correlation, moderate correlation, above moderate correlation, strong correlation, and Very strong correlation respectively. That is, the magnitude of the r-value indicates the correlation strength while the numeric sign indicates the direction of the correlation (which can be either positive or negative).

Case Study #2

Fifty-One (51) correlational academic survey tests related to Nigeria were extracted from Google Scholar and sampled. The indices of interest were to check if the following were presented or stated: r-values table, Measured strength of correlation. Typical findings are presented in Table7.

The survey revealed that while there was a difference between author's interpretation and the interpretation using a statistical approach, there were also some interpretations that had no difference. 100% of the sampled correlational tests interpreted their results without presenting (or citing) a Table (or Standard) of measured strength of relationship for the various levels of correlational strength that the author was willing to consider. Thus, those articles assigned arbitral correlational levels to r-values without maintaining a specific statistical approach. 100% of the sampled academic surveys presented their Table of test result r-values.

Table 7: Comment on 51 Sampled Correlational Studies

r-value	Author's interpretation	Statistical interpretation*	Difference
-0.77[a]	Strong negative correlation	Strong Negative Correlation	No
-0.16 [a]	Weak negative correlation	Weak negative correlation	No
-0.33 [a]	Weak negative correlation	Moderate Negative correlation	Yes
-0.47 [a]	Slightly weak negative correlation	moderate Negative correlation	Yes
-0.50 [a]	Strong correlation	Above moderate negative correlation	Yes
0.02 [a]	Extremely weak correlation	Positive No Correlation	Yes
0.41[b]	Significant relationship	above moderate Positive Correlation	Yes
0.24 [b]	Significant relationship	Moderate Positive Correlation	Yes
0.03[c]	Weak correlation	No Positive Correlation	Yes
0.18[c]	Weak correlation	Weak Positive Correlation	No
-0.05[c]	weak correlation	No Negative Correlation	Yes
0.24[c]	Weak correlation	Moderate Positive Correlation	Yes

*this is obtained using Table **Error! Reference source not found.**

a: (Adigun, 2020)

b: (Adeyemi, 2008)

c: (Ogedebe, Emmanuel, & Musa, 2012)

A cross-examination of these fifty-one (51) correlational tests revealed some arbitral interpretation of r-values. This was due to the absence of a standard table or reference table to interpret the strength of correlation; secondly, statistically inconsistent interpretation of p-values was observed. Some authors used p-values to decide the outcome of correlational analysis thus completely ignoring the relevance of r-values. However, the r-value is the major reason for a correlational analysis, since the aim of a correlational test is to reveal the strength and direction of the relationship.

Nonparametric Unpaired Sample Test

An unpaired sample test is performed on a dataset whose values are known or considered/expected to be free of a previous sample or event. That is, the responses were collected once without repetition. When there are two groups (e.g., control group and experimental group, male and female, lecturers and students, adolescents and adults, lecturers and industry professionals, etc) then the Mann-Whitney u test (also known as the Wilcoxon Rank Sum test) can be performed to determine if there exists a difference between the mean response of both groups. According to (Harwell, 1988) the Puri and Sen test can also be used for a grouped unpaired dataset.

To determine if more than two groups differ in their responses to a question-item, the Kruskal-Wallis test can be performed. According to (Stat59, 2021) these groups can include a control group, course group, and a Course Plus Simulation group. The Chi square test offers a special use case beyond what the typical t-test can handle. With a Chi square test, a researcher can investigate if two or more groups differ in more than one preference or achievements. But researchers should note that to satisfy the requirement of the chi square test, the data should have been gathered such that once a respondent chooses one preference, he/she cannot choose the other.

A research question requiring the use of Chi Square can be: *Do men and women differ in their preference of car brands?* Here, the independent variables would be gender (male and female) and the dependent variable would be *car brand* preferences (e.g., Innoson, Toyota, Honda, Hyundai, Ford, Benz, Kia, Volkswagen). A simple survey instrument to collect data for a chi square test is presented in Table 8. The analyses of which would require the presentation of a contingency table and the calculation of the Chi-square value by combining the contingency table with an expected frequency table. Thus, a contingency table like Table 9 would need to be developed with a corresponding expected frequency table (see Table 10) assuming there are 92 respondents.

Table 8: A simple Chi Square data research instrument

	Ins.	Tyt.	Hnd.	Kia
Which of these cars do you prefer				

Table 9: Contingency Table (Observation)

Gender	Innoson	Toyota	Kia	Honda	Total
Male	7.00	16.00	15.00	11.00	49
Female	6.00	13.00	16.00	8.00	43
Total	**13.00**	**29.00**	**31.00**	**19.00**	**92**

Table 10: Expected Frequency

Gender	Innoson	Toyota	Kia	Honda	Total
Male	6.92	15.45	16.51	10.12	49.00
Female	6.08	13.55	14.49	8.88	43.00
Total	13.00	29.00	31.00	19.00	92.00

Nonparametric Paired Sample Test

A paired sample test is performed on a dataset whose values are known or considered/expected to vary due to a mediating action. That is, responses to the same question items would be collected more than once from the same sample/respondents, hence we have a pair of tests to evaluate. That is, *does the second response differ significantly from the first?* Usually, the second response is collected after administering an intervention on the response group. This is why the first response is called pre-test and the second response is called post-test. Examples of paired sample tests include Wilcoxon Signed-Rank test.

Nonparametric Correlational test, Regression, and Difference Test

To analyse the influence of two or more variables on another variable, we use a regression test. It can also be used to predict a variable based on one or more variables. Examples of regression tests include Logistic regression (see DATAtab, 2024 for more details), and Puri and Sen test (see Harwell, 1988 for more details). The Spearman rank correlation test can be used to test the correlation between two nonparametric pairs (Morse, 2022).

Nonparametric tests equivalent to unpaired samples t-test, paired samples t-test, test of correlation, and test of regression have been in existence for decades, but their widespread adoption have been constrained by the unavailability of software programs capable of such calculations, and the low intensity of study of these tests by software related doctoral programs (Harwell, 1988). This conclusion still holds true today. The later reason for the unpopular utilization of nonparametric tests can be reasoned to have persisted because most software related doctoral programs have typically been fundamentally science-based programs. And science largely investigates numbers (parametric values) not feelings, emotions, and beliefs – a select choice of psychology and sociology which yield mostly categorical variables. Nonetheless, the central limit theorem has often been cited over the years to justify the choice of parametric tests on sample data that are substantively nonparametric (Harwell, 1988; Stat59, 2021). With these extensive applied research, academic researchers are welcome to embrace the challenge and apply nonparametric tests on their nonparametric datasets confidently.

Case Study #3: Attitude of men and Women towards Government Financed Childcare

Far back as 1988, Harwell recommended and demonstrated the use of the Puri and Sen test to evaluate and analyse nonparametric sample data. This article extracts the data used by Harwell and applies the Mann-Whitney UTest on the same data. The analytical result and interpretation are compared with that obtained by Harwell.

Table 11: Attitude towards government financed child-care.

Men	Women
30	22
33	11
35	14
36	12
23	24

Source: (Harwell, 1988)

Null Hypothesis: Men and women do not differ in their attitudes towards government-financed childcare.

Case Study 3 (H_{03}) can be tested using Mann-Whitney's U test (see Table11,12, and13). The results (see Table 12 and 13) indicates that men had significantly different attitude to government financed childcare than women, z = -2.402, p = 0.016. Therefore, men differ significantly in their attitude towards government financed childcare than women. These results, interpretation and conclusion corresponds withHarwell (1988) use of the Puri and Sen test.

Table 12: Case Study 3 Ranks

	Gender	N	Mean Rank	Sum of Ranks
GFCC	male	5	7.80	39.00
	Female	5	3.20	16.00
	Total	10		

Computed using IBM SPSS ver. 23

Table 13: Case Study 3 Test Statistics[a]

	Rank_GFCC
Mann-Whitney U	1.000
Wilcoxon W	16.000
Z	-2.402
Asymp. Sig. (2-tailed)	.016
Exact Sig. [2*(1-tailed Sig.)]	.016[b]

a. Grouping Variable: Gender

b. Not corrected for ties.

Computed using IBM SPSS ver. 23

From Table 12 and 13, it can be seen that to perform the Mann-Whitney U test, the raw values of men and women's attitude had to be ranked. It was on these ranked values that the computations were performed. A similar test can be performed on the question *do* *lecturers and students differ in their rating of University/College services*? (see Table5) Here, the independent variable would be status (i.e., student or Lecturer) while the dependent variable would be one of the University/College services (Electricity supply, Water supply, Transportation services, Technology Laboratories, Electronics Laboratory, etc) rating. However, if the researcher desired to investigate whether students and lecturers differed on all the University/College Services, then an ANOVA would be appropriate.

Conclusion

The Mann-Whitney U test, Wilcoxon Signed Rank test, Kruskal Wallis test, Chi Square test, Spearman Rank test, and Logistic Regression can be applied to the analysis of a nonparametric sample data. The Mann-Whitney U test yielded the same interpretations and conclusions on the data analysed by Harwell (1988) using the Puri and Sen test. It is good practice for an academic researcher to present the reference table or statement by which he/she interprets Likert Scale survey results, and correlational r-values. There is a statistically consistent approach to interpreting the mean of a Likert Scale test result. A choice to maintain a standard/universal approach to the interpretation of test results is a choice to advance, standardize, and promote logical consistency, as well as quality in academic research analysis reporting.

Reference

Adeyemi, T. O. (2008). Teachers ' teaching experience and students ' learning outcomes in secondary schools in Ondo State, Nigeria. *Educational Research and Review*, *3*(6), 204–212. Retrieved from http://www.academicjournals.org/ERR

Adigun, O. W. (2020). The Factors determining voter turnouts in presidential elections in Nigeria : Multivariate Correlation analysis of the 2019 Presidential election, *Open Political Science*, *3*, 11–33. Retrieved from https://doi.org/10.1515/openps-2020-0002

DATAtab. (2024). *Statistics made easy* (5th ed., M. Jesussek & H. Volk-Jesussek, eds.). Graz: DATAtab e.U. Retrieved from https://datatab.net

Gasca, P. (2016). Si*lence: It's one simple thing that will spark your creativity,* Retrieved from LinkedIn website: https://www.linkedin.com/pulse/silence-its-one-simple-thing-spark-your-creativity-peter-gasca

Harwell, M. R. (1988). Choosing between parametric and nonparametric tests, *Journal of Counselling and Development*, *67*, 34–37.

Kasahara, H. (2020). *Type 1 error, Type II error, and power of test example*, Ontario: The University of British Columbia. Retrieved from https://hkasahara.sites.olt.ubc.ca/files/2020/11/notes_power.pdf

MATHStorya. (2023). *How to interpret the Likert Scale ||5-pint Likert Scale*. YouTube. Retrieved from https://youtu.be/tZyPYpdblnU?si=jGi714FoJ9WzKzlk

Morse, D. (2022). *Re: How do I correlate two sets of Likert scale questions*? Retrieved June 16, 2024, from https:/www.researchgate.net/post/How_do_I_correlate_two_sets_of_Likert_scale_questions/61e4b3b2ced4cc56c6359f7d/citations/download

Ogedebe, P. M., Emmanuel, J. A., & Musa, Y. (2012). A survey on facebook and academic performance in Nigeria universities, *International Journal of Engineering Research and Applications*, *2*(4), 788–797. Retrieved from www.ijera.com

Stat59. (2021). N*on-Parametric analysis: an example using likert dat*a, Retrieved November 1, 2023, from https://youtu.be/31GY7my9xCw?si=vk_ZL9Aj2co3efOG