01

Use of Statistics in Educational Research

S.A.N. Danushka, Chinthaka Chandrakumara

Faculty of Education, University of Colombo

Abstract

The use of statistics is not restricted to a specific subject. Effective statistics are more useful for understanding and reading scenarios and epilogue. Hence, this article attempts to provide a brief but essential direction for educational researchers and students who are interested in working with statistics but still hesitant, as they are bound by discipline-related misunderstandings of utilizing statistics in their research studies. With two basic aims, the article provides simple description to comprehend what statistics could do for beginners in educational research and what they ought to know in statistics before it is used and benefits of using statistics in educational research have been discussed.

Keywords: Descriptive statistics, Inferential statistics, Measurement scales, Parametric and non-parametric statistics

Introduction

Education, as categorically labeled, is a mere social science which is seemingly dealing with many disciplines and the pace of dealing invariably increasing in the modern time due to many reasons. Educational research trends have also been widened to answer the concurrent demands and the related issues that the people are concerning about education. Further, we observe that education research is nonetheless qualitative by its nature rather than quantitative. Hence, many educationists are hesitant to use quantitative applications in educational research as they are controlled by so-called mindset, that is, the inability of using quantitative applications in educational research (Cohen, Manion & Morrison, 2018).

Three fundamental questions are addressed in this article:

(1) How can statistics be utilized in educational research?

(2) What types of statistical applications can be used? And

(3) Where can they be utilized in educational research?

Research is a common scientific way of in answering the discipline-related problems and issues, aiming at promoting and upgrading the subject knowledge, theory, and applications. Basically, three research approaches can be identified such as Qualitative, quantitative, and the combination of both, is called Mixed-method. In education, more qualitative researches are done due to its nature. Educational research issues and problems are human-centered and social-oriented, so the data is gathered via interviews, questionnaires, FGDs and observations. As the data is qualitative, the quality of the data is questionable, and further, it is difficult to quantify. Due to such inabilities, educational researchers are cautious to use statistics in research. Similarly, educational researchers are not confident in applying statistical techniques in research data analysis due to less exposure into numerical data handling (McLEAN & Ernest, 1998).

Once we take a look on some definitions provided, we could understand what statistics is, as defined by the *American Statistical Association (ASA),* "is the science of learning from data, and of measuring, controlling and communicating uncertainty." (Wild, Utts & Horton, 2017, p.2) and Statistics is a collection of methods for collecting, displaying, analyzing, and drawing conclusions from data (Shafer & Zhang, 2012, p.9). In the social sciences like education, statistics is essential because it uses exacting quantitative techniques to comprehend and address some of the biggest problems facing human society (Foster, L., Banton, J., & Diamond, 2014). At a time when the social sciences are changing due to the amount and kind of data available, statistics is particularly important. We understand that "words" could have different meanings and be differently understood, but "numbers" could convey the same massage for everyone and no hesitation over the numbered message. Hence, the use of statistics, undoubtedly, sheds light on educational research findings and eventually it leads the researcher to generalize his/her findings confidently and accurately (Felix & Felix, 2024).

Measurement scales

Use of statistics in research is to be done in a cautious manner, as it needs basic knowledge and experiences to decide the data scales (Measurement scales) and corresponding statistical techniques. First, we need to understand statistically different data scales as proposed in the table below.

Table 1: Comparison of Measurement Scales

Scale	Difference	Magnitude	Absolute Zero	Example
Nominal	Difference indicated in the form of label.	No equality in magnitude.	No absolute Zero	Use of number to place variables into categories.
Ordinal	Indication difference among individual.	No equality in Magnitude	No absolute Zero	Position in examination scores, Likert Scale, rating scale, etc
Interval	Indication difference among individual.	Equality in magnitude among individual.	No absolute Zero	Celsius scale, Fahrenheit Scale, IQ test
Ratio	Indication of difference among individual.	Equality in magnitude among individual.	Absolute Zero Counts	Test scores, Weighing balance, meter rule etc.

Source: Adapted from Isah, 2020, p.1 and Corder & Foreman, 2014, p.2

Table 1 shows that four measurements scales which is broadly labeled in statistics. It is useful to understand each as it is necessary for researchers to decide how to process and analyze data (Kothari, 2014, p.71; Corder & Foreman, 2014, p.1).

Nominal scale is merely a method of allocating numerical symbols to occurrences for the purpose of tagging them. A common illustration of this is the numbering of basketball players to recognize them. These figures cannot be regarded as linked to a sequential scale. Their arrangement is insignificant; the figures are merely useful tags for the specific category of events and therefore, possess no numerical value. Nominal scales offer practical methods for

monitoring individuals, items, and occurrences. There is little one can accomplish with the given numbers. For instance, it is not practical to take an average of the numbers on the jerseys of a team of football players and produce a significant result. One cannot meaningfully compare the figures attributed to one group with the digits allocated to a different one. The tallying of individuals within each group is the sole feasible calculation. Consequently, we are limited to employing mode as the metric of central tendency. No commonly accepted metric for dispersion exists for nominal scales. The Chi-square test is the most frequently used method for assessing statistical significance that can be applied, and the contingency coefficient can be calculated using measures of correlation.

The nominal scale represents the most basic level of measurement. It shows no hierarchical or distance relationship and lacks a mathematical origin. A nominal scale merely categorizes differences between items. Consequently, nominal data is simply data that can be counted. This scale overlooks any insights we might have regarding varying levels of attitudes, skills, understandings, and so on. Despite these limitations, nominal scales remain highly practical and are frequently employed in surveys and other retrospective research when classifying data by significant sub-groups of the population.

The ordinal scale is the most basic form of ordered measurement. This scale arranges events sequentially, but it does not ensure that the distances between the values are uniform according to any specific criteria. Rank orders serve as examples of ordinal scales and are often utilized in research focused on qualitative aspects. A student's ranking in their graduating class exemplifies the use of an ordinal scale. It is important to exercise caution when interpreting scores derived from ordinal scales. For example, if Ranaya ranks 8^{th} in her class and Thisumi ranks 24^{th}, it cannot be stated that Ranaya's position is three times superior to Thisumi's. The assertion would be completely nonsensical. Ordinal scales solely allow for the arrangement of items in a hierarchy from highest to lowest. Ordinal measures lack absolute values, and the actual differences between neighboring ranks may not be uniform. The only conclusion is that one individual is positioned higher or lower on the scale compared to another, but more accurate comparisons cannot be established.

Therefore, utilizing an ordinal scale denotes a relationship of 'greater than' or 'less than' without allowing us to quantify the exact magnitude of the differences. The actual difference between ranks 6 and 7 could be more or less significant than the difference between ranks 7 and 8. Because the values in this scale only convey a ranking, the median is the most suitable measure

of central tendency. For measuring variability, percentile or quartile methods are employed. Correlations are limited to various rank-order techniques. Statistical significance measures are confined to non-parametric approaches.

When the ***interval scale*** is considered, the intervals are modified according to a predetermined rule that serves as the foundation for equalizing the units. The units are considered equal only to the extent that one agrees with the assumptions underlying this rule. Interval scales may have a zero arbitrary point, but it's not feasible to establish what could be deemed an absolute zero or a distinct origin for them. A key drawback of the interval scale is its absence of a true zero; it cannot measure the total lack of a trait or characteristic. The Fahrenheit scale serves as an example of an interval scale, illustrating the limitations and capabilities associated with it. The Fahrenheit scale represents an interval scale and illustrates the distinctions between what can and cannot be measured with it. It is possible to state that a temperature rise from 40° to 50° represents the same increase in temperature as a rise from 70° to 80°, but it is inaccurate to claim that 80° is twice as warm as 40°, since both values rely on the arbitrary setting of zero at the freezing point of water.

Interval scales offer a more robust form of measurement compared to ordinal scales because they include the principle of equal intervals. Consequently, more sophisticated statistical techniques can be applied to interval scales. The mean serves as a suitable measure of central tendency, whereas the standard deviation is the most commonly utilized measure of dispersion. Techniques for calculating product-moment correlation are fitting, and the commonly employed tests for statistical significance include the '*t*' test; a t-test is a statistical method that assesses whether there is a meaningful difference between the means of two groups. It serves as a type of hypothesis testing that is particularly beneficial for small sample sizes when the population's standard deviation is not known and the 'F' test; F-test is a statistical method utilized to assess the differences in variances among two or more groups and it is quite similar to the ANOVA test.

Ratio scales feature an absolute or true zero point for measurement. The concept of 'absolute zero' is not as clearly defined as it was previously thought. We can imagine an absolute zero for length, just as we can for time. For instance, the zero mark on a centimeter ruler signifies the complete lack of length or height. However, achieving an absolute zero in temperature is theoretically impossible and remains solely a concept in the minds of scientists. If one item weights 10 grams and the other one weights 5 grams, it is correct to say that the first one is two

times as heavy as the second. A ratio scale indicates the actual quantities of variables. Examples include measurements of physical dimensions like weight, height, and distance. All statistical methods can be applied to ratio scales, and any operations that can be carried out with real numbers can also be carried out with ratio scale values. Both multiplication and division are applicable on this scale, unlike the other scales previously mentioned. The geometric and harmonic means can be utilized as measures of central tendency, and it is also possible to compute coefficients of variation.

It is noted that generally educational researchers' lack of understanding about the four measurement scales lead to inaccuracy and least-quality of data analysis in research and again it would be doubtful. Some reasons are visible like lack of knowledge about statistics, lack of practice, no appropriate course units introduced for novices to follow in respective courses in the universities and colleges and lack of supervision. Usually, in social science research, Ordinal or Nominal data are collected but it does not say that Interval or Ratio data are not applicable. Educational research studies, particularly in psychology and assessment, interval and ratio data can be collected, i.e. written test marks, Psychometric test marks. Selecting the type of analysis and statistical tool should then be decided depending upon the measurement scale as any statistical tool cannot be utilized for data come from different measurement scales.

Different types of statistics: Descriptive and Inferential Statistics

Statistics are generally classified into two categories: descriptive and inferential. Typically, these two types of statistics are utilized in scientific research. Quantitative research employs different functions for each of these statistics.

Descriptive statistics as a statistical approach, uses numerical and graphical or tabular methods to summarize data in such a way that characteristics of individuals or group are explained and differentiated from other individuals or group within a population. Examples of descriptive statistics include but not limited to:

- Frequency distribution (frequency polygon, pie chart, bar chart, histogram, Pictogram)
- Measure of Central Tendency (Mean, Median, Mode, Skewness, Kurtosis, normal distribution)
- Measure of Departure (range, mean deviation, variance, standard deviation)
- Measure of Partition (Quartile, Decile, Percentile)
- Measure of Relative Standing (T-score, Z-score, Percentile Rank, standard nines)

- Measure of Relationship (Pearson Product Moment Correlation, Spearman Rank Order Correlation, Phi Coefficient etc).

Inferential statistics uses data collected from a randomized sample to make decisions, estimates, or generalizations which can be counted on the general population. The essence of inferential statistics is to make conclusions that bind every member of a population by using a scientific tool known as hypothesis testing. Inferential Statistics is of two categories: *Parametric statistics and Non-Parametric statistics*. Each of these categories has its application based on certain assumptions.

Assumptions to Use Parametric Statistics

The following assumptions are used to employ parametric statistics;

- The variables should be normally distributed – i.e., if you have student test results, it is essential to check normality, which can be easily done in SPSS via a Normality Test. It means that data dispersion happens within a normal bell-shaped curve. Once you calculate both Skewness and Kurtosis of the test data, this assumption could be tested statistically.
- Equality of variance between the variables - Equality of variance refers to the uniformity in the spread or dispersion of data points around their respective averages among different groups or samples being analyzed. When you perform Levene's test for equality of variance in SPSS, this assumption can be precisely understood and decided the applicability of data to go for Student's test or the ANOVA test. In statistics, homoscedasticity is a prerequisite for checking the suitability of data for such tests. The variables must be measured using interval or ratio scale
- The representative sample should be randomly selected. Though this has been the argument in statistics if the data is normally distributed, there is no such obstacle to go for parametric measures as the data normality would be enough to go for parametric measures. Here, the sample size is significant, often exceeding 30–40, due to the Central Limit Theorem (Sedgwick, 2012; Vickers, 2005). It means that if your data's deviation from normality is significant, particularly with smaller sample sizes, it is much safer and more precise to utilize non-parametric tests depending upon the expected tasks of the analysis aligning to the data analysis plan. It reveals that before moving to data analysis, researcher must develop an analysis plan considering the research objectives.

Assumptions to Use Non-Parametric Statistics

The following assumptions are used to employ non-parametric statistics;

- The variables may not necessarily be normally distributed
- No equality of variance between the variables
    - The variables were measured using nominal or ordinal scale (Corder & Foreman, 2014, p.2). These assumptions could be understood when you refer to the brief description postulated related to the aforementioned assumptions under parametric statistics.

The following table shows how different data types come via diverse measurement scales could lead to corresponding statistical data analysis tools. And, it would be beneficial for educational researchers, particularly for beginners, to select statistical data analysis tools with respect to the data at hand.

Table 2: Conditions to Apply Statistics in Research

What are you looking for?	What type of scale is used?	Parametric Test	Non-Parametric Test
Relationship between two independent continuous variables	Interval/Ratio scale for both variables	Pearson Product-Moment Correlation (PPMC)	
Relationship between two independent continuous variables	Ordinal Scale for both variables		Spearman Rank Order Correlation
Relationship between two independent categorical variables	Nominal scale for both variables (Dichotomies)		Phi Correlation Coefficient
Difference between two independent continuous variables	Ratio scale for both variables	Independent t-test	
Difference between two independent continuous variables	Ordinal scale for both variables		Mann Whitney's U Test
Difference between two dependent continuous variables	Ratio scale for both variables	Dependent t-test	

Difference between two dependent continuous variables	Ordinal scale for both variables		Wilcoxon Test/Sign Test

Source: Adapted from Isah, 2020, p.3 and and Corder & Foreman, 2014, pp. 3-4

Use of Statistics by Educationists: Purposes

It is well-known that statistics provides certain benefits for students, researchers, and academics in education. Some situations won't be able to understand with the absence of numerical interpretations, where verbal explanations/statements would not make sense or a failure to provide to do so. i.e. If you claimed that *"Sadun is good in English"*. Sadun is good in English due to some reasons as he is regular in the class, well obedient and the statement is totally subjective because he is good for English teacher but not for other teachers' subjects. But such questions could not raise if the Sadun's English monthly test marks have been given, then the matter is understandable. This shows that numerical explanations are necessarily important for some situations to get them understood. And also the combination of both "words" and "numbers" would ensure proficient clarifications over the educational concerns either in research or in general affairs. Some noticeable points are given below, aligning to some personal research experiences.

1. **Utilizing Statistics to Influence Educational Policies**:

   Decisions Based on Data one of the most emotional moments when we confronted a policy choice that would have implications for thousands of students. When we were debating whether to adopt a more traditional, rigid curriculum for teaching math or to embrace a more contemporary, flexible approach. It is understood that while personal narratives were impactful, they couldn't be the only foundation for decisions affecting so many lives. That's when we decided to utilize statistics. We conducted a pilot study and collected data from several colleges that had used both methodologies. We examined student performance, levels of engagement, and even dropout statistics. The findings were remarkable. Although the traditional method was effective for some, the flexible curriculum resulted in greater engagement and higher test scores among a wider array of students, particularly those who found mathematics challenging. Equipped with this data, we were able to move forward with confidence, enacting a policy rooted not just in theory, but in real-world outcomes, thereby enhancing the educational experience for numerous students. Statistics not only guided our decision-making process but also ensured that we made the correct choice.

2. **Assessing learning outcomes: Power of numbers**

   Teacher expressed. *"I'm putting in so much effort to teach these students, but their test performance is poor. Where am I going wrong?"* Rather than making assumptions or assigning blame, we looked at the numbers. Patterns were identified by evaluating the data, which included not only test scores but also attendance, completion of homework, and levels of participation. Those students who regularly attended morning tutoring sessions had much better results than those who didn't, as it became clear. Additionally, we discovered that students with difficulties in reading comprehension were also having trouble with math word problems, which affected their overall performance. Equipped with this insight, the teacher could modify her teaching strategies, integrating more reading techniques into her math lessons and motivating students to join tutoring sessions. Within a few months, we witnessed a notable rise in test scores, not just in math, but in other subjects too. This served as a valuable lesson for everyone: data doesn't only highlight issues; it also uncovers solutions.

3. **Creating Effective Curricula: A Guide to Success**

   When we were assigned to develop a new curriculum, we were determined not to create something ineffective for our students. Therefore, rather than starting from the ground up or depending solely on theoretical frameworks, we engaged in statistical analysis of existing curricula. We examined years of student performance data to identify which aspects of prior curricula yielded the best educational results and which segments frequently caused confusion or disinterest. We conducted surveys on students and teachers to gain insights into their experiences, preferences, and challenges. Using this collected data, we developed a curriculum that was specifically designed for our students' needs. We understood which teaching methods resonated with various learner types and identified areas that required greater emphasis. Our choices were informed by evidence, not mere speculation.

4. **Enhancing fairness in education: Data as an approach to equity**

   One of the most touching and motivating aspects of using statistics in education is its capacity to highlight disparities that might otherwise remain hidden. I remember examining school data from various neighborhoods, anticipating the typical differences in performance. However, what was discovered exceeded mere variations in test scores;

it uncovered an unambiguous inequality in available resources. Schools situated in lower-income areas faced larger class sizes, a scarcity of experienced teachers, and outdated supplies. The students attending these schools possessed the same potential as those in wealthier neighborhoods, yet the statistical analysis indicated they were being disadvantaged through no fault of their own. This data served as the basis for a movement towards equity within the educational zone. We campaigned for increased resources for underfunded schools, smaller class sizes, and enhanced support for educators. Statistics illuminated the unfair treatment of students, which was not our intention.

5. **Improving Educational Research**

   Statistics play a crucial role in educational research. They enable researchers to gather and analyze data, test hypotheses, and reach dependable conclusions. By using statistical analysis, researchers can assess the effectiveness of various teaching methods, evaluate the effects of educational policies, and examine the elements affecting student success. Without the application of statistics, educational research would be lacking of the accuracy and precision needed to guide practice and policy.

Conclusion

It is observed that undergraduate and graduate students are usually asked to conduct research in higher academic scenarios as a part of their respective degrees. The issue is not just about conducting research; how do they ensure the quality of research, whether in qualitative or quantitative terms? Social science researchers often choose qualitative research approaches because they lack the necessary quantitative data handling and analyzing skills, even with computer software like SPSS and NVivo. Sometimes, they feel that quantitative tools are not applicable in social science research due to their nature. It is advisable to note that if the students are engaged in scientific research, quality maintenance deals with the overall research plan, or even it could be understood as research methodology, like design, sampling, developing a data collection instrument, and data analysis.

Predominantly, researchers tend to apply simple statistical techniques in research and dissertations, such as percentages, histograms, and polygons, without utilizing suitable statistical techniques to estimate accurate figures about data dispersion. Hence, the accuracy of

their respective research findings is questionable and non-generalizable. Further, there are possibilities to convert qualitative data into numerical data via appropriate coding methods so that social science researchers, particularly educational researchers, could go for effective and proficient data analysis if they are capable enough to identify the applicable statistical techniques that can be used while having a prior understanding of measurement scales and corresponding statistical techniques in data analysis. This article would be helpful for novices who have such a willingness in educational research.

References

Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education (8th ed.). Routledge.

Corder, G. W., & Foreman, D. I. (2014). Nonparametric statistics: A step-by-step approach. John Wiley & Sons.

Felix, C. S., & Felix, C. M. (2024). Literature review: Basic statistics in educational research. Cognizance Journal of Multidisciplinary Studies, 4(3), 206–215.

Foster, L., Banton, J., & Diamond, I. (2014). Beginning statistics: An introduction for social scientists. Sage Publications.

Ji, X. (2023). Analysis of the application of statistics in education. Lecture Notes in Education Psychology and Public Media. https://doi.org/10.54254/2753-7048/11/20230707

Kothari, C. R. (2004). Research methodology: Methods and techniques (2nd ed.). New Age International.

McLean, J. E., & Ernest, J. M. (1998). The role of statistical significance testing in educational research. Research in the Schools, 5(2), 15–22.

Sedgwick, P. (2012). Parametric v non-parametric statistical tests. BMJ, 344, e1753. https://doi.org/10.1136/bmj.e1753

Shafer, D. S., & Zhang, Z. (2012). Beginning statistics (2nd ed.). Phylis-Barnidge Publisher.

Vickers, A. J. (2005). Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. BMC Medical Research Methodology, 5(1), 35. https://doi.org/10.1186/1471-2288-5-35

Wild, C. J., Utts, J. M., & Horton, N. J. (2017). What is statistics? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), International handbook of research in statistics education (pp. 5–36). Springer. https://doi.org/10.1007/978-3-319-66195-7_1