# Dance of Data in Pedagogical Research

*Peter Pšenák[1]*

## *Abstract*

This conference paper concentrates on the exploration of the interplay between statistics, probability distributions, t-tests, and the utilisation of the R language in pedagogical research. Statistics play a vital role in collecting, analysing, and interpreting data in educational research. The t-test, as a fundamental statistical tool, enables researchers to assess the significance of differences between groups, a common requirement in pedagogical studies. The integration of the R language, a powerful and flexible statistical computing tool, empowers researchers to conduct complex analyses efficiently. It acts as a bridge between statistical theory and practice, enabling the application of statistical methods to real-world educational data. This paper sheds light on the systematic approach to employing these elements, emphasising their practical importance in pedagogical research.

*Keywords:* Education, Research, R, Statistics, T-test

# 1 Introduction to Statistical Analysis in Educational Research

Statistical analysis is a critical component of educational research, playing a pivotal role in understanding, interpreting, and drawing conclusions from data collected within educational settings. Its importance lies in its ability to provide a systematic framework for researchers to analyse and interpret complex educational data, easing decision-making processes and shaping educational policies.

Statistical analysis is paramount in educational research due to its wide-ranging applications from evaluating teaching methods and student performance to examining socioeconomic influences on learning outcomes. Statistical analysis is a crucial tool for achieving the intended outcomes (Ary et al., 2018).

Educational research faces a significant challenge due to the overwhelming volume of published studies, increasing the complexity of conducting systematic reviews (Newman & Gough, 2020). Over recent years, the field has witnessed a surge in literature, coinciding

---

[1] Comenius University in Bratislava, Faculty of Management, Odbojárov 10, Bratislava.
*E-mail: peter.psenak@fm.uniba.sk*

notably with the disruptive impact of the COVID-19 pandemic on education worldwide (Abumalloh et al., 2021; Baker, 2020).

The pandemic prompted rapid and extensive changes in educational approaches, with a current trend leaning towards reinstating pre-pandemic educational norms. It acted as a catalyst for exploring innovative educational models like the flipped classroom (Szőköl et al., 2023) while emphasising the urgency of understanding students' psychological well-being (Ebrahim et al., 2022).

Statistical analysis in educational research involves using statistical tools to understand and interpret data collected in educational studies. Researchers use techniques like descriptive statistics to summarise data characteristics, inferential statistics to make predictions about larger populations, and experimental design to assess the effectiveness of educational interventions. This analysis helps identify trends and relationships between variables and assess the impact of teaching methods or policies. It also aids in survey analysis, tracking changes over time, and handling large educational datasets to derive evidence-based conclusions that inform educational practices and policies (Ary et al., 2018; Peck et al., 2020).

However, there are instances where educational research appears daunting, often resulting in a lack of enthusiasm among individuals. Quantitative research methodologies can seem challenging, with terms like hypothesis creation and testing evoking a sense of apprehension. This article aims to shift this perception by introducing accessible research methodologies and showing user-friendly approaches within educational research.

## 2  Fundamentals of Probability Distributions

Probability distributions form a foundational concept in educational research, serving as essential tools to understand and interpret data variability and uncertainty. In simplest terms, a probability distribution represents the likelihood of different outcomes in each scenario. In education, these distributions aid in modelling and analysing various educational phenomena, providing insights into student performance, learning outcomes, and the efficacy of teaching methodologies (Blitzstein & Hwang, 2019; Ross, 2020).

The relevance of probability distributions in educational research lies in their capacity to encapsulate uncertainty and variation within data. Researchers utilise these distributions to make informed decisions and predictions based on probabilities. For instance, in assessing student test scores, understanding the distribution of scores across a population helps educators identify the typical performance levels and calculate the probability of students achieving particular grades. Probability distributions also underpin inferential statistics, allowing researchers to conclude a larger population based on sample data (Jinming Zuo, 2021).

Several types of distributions commonly emerge in educational data analysis. The normal distribution shown in Figure 1 is prevalent in educational research due to its occurrence in naturally observed phenomena. This distribution is characterised by its unique properties,

such as the mean, median, and mode coinciding at the centre. This allows for a straightforward representation of variability and probabilities within a dataset. The binomial distribution is another significant type encountered in educational settings, mainly when dealing with discrete data or outcomes with two possibilities: pass or fail, success or failure in experiments, or responses to yes/no questions. Moreover, the Poisson distribution finds relevance in educational research when examining rare events or occurrences within a fixed interval, such as the frequency of student absences or rare academic achievements (Blitzstein & Hwang, 2019).
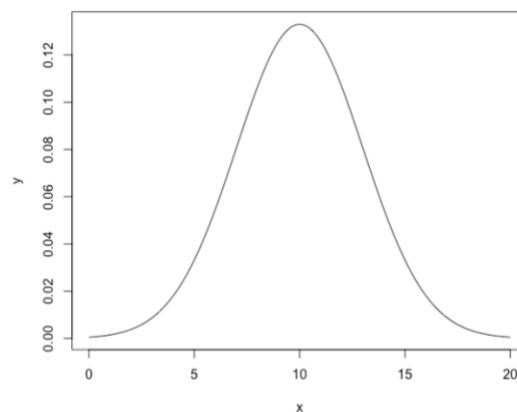


Figure 1: Normal distribution example visualised in R

Understanding these distributions equips researchers and educators with the tools to model, analyse, and derive meaningful insights from educational data. They form the basis for informed decision-making, enabling a more profound comprehension of student performance trends, the effectiveness of interventions, and the evaluation of educational policies and practices.

# 3   Fundamentals of Probability Distributions

The creation of the student's t-test marks a significant milestone in the history of statistics, attributed to the work of William Sealy Gosset, a British statistician who worked for the Guinness Brewery in Dublin, Ireland, in the early 20th century. Gosset faced challenges in analysing small sample sizes typically encountered in Guinness's quality control and brewing processes. The existing statistical methods of the time were unsuitable for small sample sizes. In 1908, Gosset introduced what would later be known as the student's t-distribution and the t-test in a paper published under the pseudonym "Student". Due to the constraints of secrecy imposed by Guinness, Gosset used this pen name to protect the company's proprietary information (Brown A.,2005).

The t-test was born from Gosset's need for a robust method to make inferences from small sample sizes. He derived a distribution based on the normal distribution but with heavier tails, suitable for smaller samples, to account for increased uncertainty. This distribution allowed

for calculating a statistic, now known as the t-statistic, used to test hypotheses about population means when the sample size is small and the population standard deviation is unknown. Gosset's pioneering work on the t-distribution and the t-test revolutionised statistical inference for small sample sizes, laying the foundation for modern inferential statistics and shaping how researchers analyse and draw conclusions from limited data in numerous fields, including the realm of educational research.
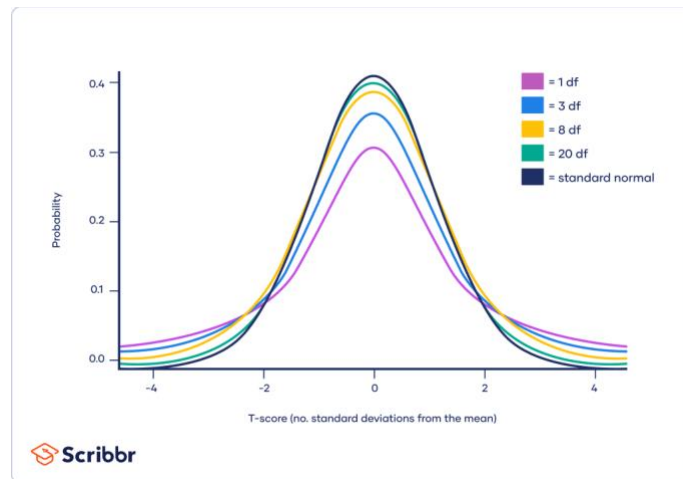


Figure 1: Comparison of standard normal and t-distribution (Bevans, 2020)

T-tests are fundamental statistical tools in pedagogical studies, playing a pivotal role in comparing groups and evaluating the effectiveness of educational interventions or methods. These tests assess whether observed differences between groups are statistically significant or simply due to random chance. Their significance lies in providing a structured framework for researchers to analyse and draw conclusions regarding the impact of teaching methods, interventions, or educational programs (J. et al., 2019; Rojewski et al., 2012).

In educational research, various types of t-tests are commonly employed. The independent samples t-test is prominent when comparing the means of two independent groups, such as assessing the performance of students taught through different teaching methodologies. Conversely, the paired samples t-test evaluates mean differences within the same group under different conditions or before and after interventions. This test helps study changes in student performance following an educational program. Both types of t-tests help researchers determine if observed differences in performance are statistically significant or if they could have occurred by chance (J. et al., 2010; Wright, 2006).

To illustrate, consider a study comparing the effectiveness of two teaching methods on student achievement. Researchers could use an independent samples t-test to analyse the test scores of students taught using Method A (for example, the flipped classroom method) versus Method B (classical education without flipped classroom). Similarly, in a scenario where an educational intervention aims to improve reading comprehension, a paired samples t-test could assess whether there is a significant difference in students' pre- and post-intervention

reading scores. These examples demonstrate how t-tests enable researchers to quantify and validate the impact of educational methods, providing empirical evidence to support educational decision-making.

Even with small datasets, calculating t-tests manually has become outdated. Fortunately, the availability of free tools like R has simplified this process, allowing us to perform these calculations swiftly and accurately.

# 4   Fundamentals of Probability Distributions

The "R" language stands as a versatile and powerful tool in statistical analysis, renowned for its extensive range of statistical capabilities and open-source nature. It offers a comprehensive suite of functions and packages tailored explicitly for statistical computing, making it a preferred choice in educational research for conducting complex analyses efficiently (Barnwal, 2017; Bretz et al., 2016).

One of the critical benefits of R in educational research lies in its ability to handle diverse statistical methods and techniques. Its extensive collection of packages supports various statistical analyses, including regression modelling, hypothesis testing, factor analysis, and data visualisation. This flexibility allows educational researchers to perform intricate analyses without the constraints often found in other statistical software (Barnwal, 2017; Field et al., 2012).

In educational research, R finds practical applications in various scenarios. For instance, researchers might use R to conduct regression analyses to assess the relationship between student performance and multiple factors like teaching methods, socioeconomic status, or classroom environment. The language's visualisation capabilities also shine when illustrating trends or patterns in educational data, aiding in creating insightful graphs or charts for presentations and publications. Furthermore, R facilitates reproducibility in research by providing a transparent and documented workflow, enhancing the credibility and replicability of studies conducted in the field of education (Pšenák et al., 2018).

The accessibility of R, coupled with its vast statistical functionalities, empowers researchers in educational settings to conduct sophisticated analyses, explore intricate relationships within data, and derive nuanced insights. Its role in educational research extends beyond analysis; it fosters a culture of transparency, rigor, and innovation within the academic sphere. Getting started with R is incredibly straightforward. You can freely download it from the R website, and we highly recommend coupling it with R-Studio, also available for free. Once installed, a plethora of free tutorials on platforms like YouTube and other online resources await. These abundant educational materials offer a quick and easy way to learn R efficiently.

# 5 Systematic Approach to Applying Statistics in Pedagogical Research

Employing statistics systematically in pedagogical research involves a structured approach, integrating statistical tools like probability distributions, t-tests, and the R language methodically within research workflows. A crucial asset for any researcher is a structured workflow that guides the entire research process. We aim to offer a versatile workflow applicable to various educational research scenarios, simplifying the initiation phase for researchers.

A proposed step-by-step guide to the process would consist of the following steps:
1. Formulating Research Questions
2. Data Collection and Preparation
3. Application of Statistical Tools
4. Data Analysis and Interpretation
5. Validation and Reproducibility
6. Iterative Process and Refinement

Researchers can methodically navigate the complexities of pedagogical research by systematically integrating statistics, probability distributions, t-tests, and the R language into research workflows. Our proposed analysis method leverages the information provided in this publication to enhance research speed and relevance.

Consider a scenario in which we implement the above-written educational research steps:
*Research Question Formulation:* A team of educators aims to investigate the correlation between student engagement levels and academic performance across different learning environments. They precisely define their research objectives by framing questions that delve into specific aspects of student engagement, such as the impact of interactive learning versus traditional lecture-based teaching on academic outcomes.

*Data Collection and Preparation:* The team collects data from various schools, ensuring meticulous collection of student engagement metrics (such as participation in class discussions, completion of assignments, etc.) and corresponding academic scores. They prepare the dataset, ensuring accuracy and completeness by validating and organising the gathered information.

*Application of Statistical Tools:* Leveraging statistical software like R, the team conducts comprehensive analyses, applying suitable statistical methods to explore the relationship between student engagement metrics and academic performance, for example, with regression analysis.

*Data Analysis and Interpretation:* They derive meaningful insights through statistical analysis, identifying patterns and relationships within the data. The team interprets their findings.

*Validation and Reproducibility:* To ensure the credibility of their findings, the researchers diligently document their statistical procedures, codes used for analysis, and any data transformations undertaken. This documentation ensures the replicability and transparency of their research methodology.

*Iterative Process and Refinement:* Engaging in an iterative process, the team continually refines their analysis, considering alternative statistical approaches to strengthen the robustness of their findings. This iterative refinement cycle helps enhance the reliability and depth of their research outcomes.

# 6    Conclusion

In educational research, statistical analysis is an indispensable cornerstone, evident across data collection, analysis, and interpretation. This paper has journeyed through the multifaceted landscape of statistical methodologies and their pivotal significance within educational contexts.

Delving deeper, the exploration of t-tests explains their profound relevance in comparing groups within educational research scenarios. Some theoretical examples showcased their application. Integrating the R language helps improve the research experience, empowering researchers to conduct sophisticated analyses efficiently. Its capability as a statistical computing tool resonated across practical examples, underscoring its role in enhancing analytical prowess and expediting complex analyses within educational research.

Lastly, the systematic approach delineated herein provides a structured methodology for researchers to employ statistics effectively in educational research. This methodological framework emphasises open-source tools, offering a guide for navigating the intricacies of educational research workflows.

In conclusion, this paper encompassed various statistical probabilities, explored the historical context of the t-test, delved into the application of R, and introduced a framework tailored to educational research. Using freely available software tools was proposed to augment research efficiency and transparency.

## *References*

Abumalloh, R. A., Asadi, S., Nilashi, M., Minaei-Bidgoli, B., Nayer, F. K., Samad, S., Mohd, S., & Ibrahim, O. (2021). The impact of coronavirus pandemic (COVID-19) on education: The role of virtual and remote laboratories in education. *Technology in Society,* 67, 101728. https://doi.org/10.1016/j.techsoc.2021.101728

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to Research in Education*. Cengage Learning.

Baker, K. (2020, May 18). Covid-19 is changing education for the better. https://www.ft.com/content/51496fde-98e7-11ea-871b-edeb99a20c6e

Barnwal, M. (2017, May 3). Time series and forecasting using R. http://manishbarnwal.com/blog/2017/05/03/time_series_and_forecasting_using_R/

Bevans, R. (2020, August 28). T-distribution: What it is and how to use it. Scribbr. https://www.scribbr.com/statistics/t-distribution/

Blitzstein, J. K., & Hwang, J. (2019). *Introduction to Probability*, Second Edition. CRC Press.

Bretz, F., Hothorn, T., & Westfall, P. (2016). *Multiple Comparisons Using R*. CRC Press.

Brown, A., The strange origins of the Student's t-test. (2008). The Physiological Society. Cit 03. December 2023, z https://www.physoc.org/magazine-articles/the-strange-origins-of-the-students-t-test/, https://doi.org/10.36866/pn.71.13

de Winter, J. C. F. (2019). The Student's t-test was used with extremely small sample sizes. Practical Assessment, Research, and Evaluation, 18(1). https://doi.org/10.7275/e4r6-dj05

de Winter, J., & Dodou, D. (2010). Five-Point Likert Items: T-Test Versus Mann–Whitney–Wilcoxon. Practical Assessment, Research and Evaluation, 15.

Ebrahim, A. H., Dhahi, A., Husain, M. A., & Jahrami, H. (2022). The Psychological Well-Being of University Students amidst COVID-19 Pandemic. Sultan Qaboos University Medical Journal, 22(2), 179–197. https://doi.org/10.18295/squmj.6.2021.081

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics using R. SAGE Publications. https://thuvienso.hoasen.edu.vn/handle/123456789/12146

Jinming Zuo. (2021). Standard normal distribution test of college students' test score distribution. International Conference on Economics. 2021 5th International Conference on Economics, Management Engineering and Education Technology (ICEMEET 2021), Canada.

Newman, M., & Gough, D. (2020). Systematic Reviews in Educational Research: Methodology, Perspectives and Application. V O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, & K. Buntins (Ed.), Systematic Reviews in Educational Research: Methodology, Perspectives and Application (s. 3–22). Springer Fachmedien. https://doi.org/10.1007/978-3-658-27602-7_1

Peck, R., Short, T., & Olsen, C. (2020). Introduction to statistics and data analysis. Cengage Learning. https://thuvienso.hoasen.edu.vn/handle/123456789/12547

Pšenák, P., Kováč, U., Káčer, J., & Skýpalová, M. (2018). Impact of University degree on the labour market in Slovakia. Economic and Social Development: 36th International Scientific Conference on Economic and Social Development - "Building Resilient Society".

Rojewski, J., Lee, I. H., & Gemici, S. (2012). Use of t-test and ANOVA in Career-Technical Education Research. Career and Technical Education Research, *37*(3), 263–275. https://doi.org/10.5328/cter37.3.263

Ross, S. M. (2020). Introduction to Probability and Statistics for Engineers and Scientists. Academic Press.

Szőköl, I., Pšenáková, I., & Kováč, O. (2023). The Influence of Innovative Teaching Methods on Student Motivation. R&E-SOURCE, 196–203. https://doi.org/10.53349/resource.2023.is1.a1203

Wright, D. B. (2006). Comparing groups in a before–after design: When t-test and ANCOVA produce different results. British Journal of Educational Psychology, 76(3), 663–675. https://doi.org/10.1348/000709905X52210