Teaching and Educational Methods

Enhancing Student Learning in Statistics and Econometrics Through Experiential Teaching Methods

Kedar Kulkarni[a]
[a]*Azim Premji University*

JEL Codes: A22, C90, I21
Keywords: Central Limit Theorem, experiential learning, pedagogy, statistical inference, undergraduate economics education

Abstract
Undergraduate students often struggle with the abstract and technical nature of statistical inference, especially in classrooms where prior mathematical exposure varies widely. This article evaluates the impact of experiential teaching techniques introduced in a second-year econometrics course at a liberal arts university in India. I designed two low-cost, intuitive interventions: (1) a classroom game using chocolates to demonstrate the Central Limit Theorem and (2) a video case study from professional cricket to explain hypothesis testing through the Decision Review System (DRS). These methods aimed to build conceptual clarity and bridge the gap between statistical theory and application. Using quiz performance data from three consecutive cohorts—two taught using traditional lectures and one using experiential methods—I estimate the effect of the intervention on student performance. Students in the experiential cohort scored 1.78 points higher on a 10-point quiz, representing a 35 percent improvement over the traditional cohort and a 0.64 standard deviation increase. The gains were particularly large for students with weaker quantitative backgrounds. Qualitative feedback further highlights strong student engagement and positive perceptions of the activities. Overall, the results suggest that simple, contextually grounded interventions can enhance students' understanding of statistical inference, especially when tailored to diverse learning needs.

1 Introduction

Undergraduate students often struggle to engage deeply with key concepts in statistical inference—such as the Central Limit Theorem, hypothesis testing, and the interpretation of p-values. These foundational ideas, typically introduced in introductory econometrics courses, can remain abstract when presented through traditional lecture-based instruction. While students may learn to execute formal procedures, they often lack a deeper intuition for uncertainty, sampling variation, and the logic underlying statistical inference. In response to these challenges, educators of statistics and econometrics have increasingly turned to active and experiential learning strategies aimed at improving comprehension and retention. Rooted in constructivist pedagogy, experiential learning emphasizes "learning by doing," encouraging students to develop conceptual understanding through hands-on engagement, interaction with real data, and iterative reflection (Kolb 2014). A growing body of evidence supports the effectiveness of such approaches in enhancing student performance (Prince 2004; Burch et al. 2019). In statistics classrooms, this has translated into the use of simulations, structured experiments, and student-generated datasets to help build applied reasoning and deepen understanding of inference (Wood 2005; Garfield and Ben-Zvi 2007; Sigal and Chalmers 2016; Park 2019).

In this article, I document the design and implementation of a targeted experiential learning module within an undergraduate econometrics course, developed to strengthen students' conceptual understanding of statistical inference. The instructional objective was to move students beyond the mechanical application of formulas and foster deeper intuition about sampling distributions, confidence intervals, *p*-values, and the logic of hypothesis testing. The intervention replaced a traditional lecture with a structured, hands-on classroom activity involving data simulation and collaborative interpretation. By encouraging students to read, write, reflect, and actively engage with abstract concepts, such approaches support the development of higher-order cognitive skills (Paul and Elder 2019). In addition, the article presents a formal evaluation of the intervention's effectiveness, using student performance on quizzes designed to assess conceptual understanding.

The experiential module was implemented at a private liberal arts institution in India, in an introductory econometrics course that is a core requirement for all undergraduate economics majors. The student cohort at the university is academically diverse, with many facing challenges related to mathematical preparation and language proficiency. The intervention replaced a traditional chalk-and-board session with an interactive, simulation-based classroom activity in which students generated data, worked collaboratively in small groups, and discussed their interpretations in real time. In contrast, the comparison group received standard board-based instruction covering the same conceptual material. The experiential format was designed to align with principles of active and interactive learning, encouraging students to construct understanding through doing, discussing, and reflecting. The intervention was evaluated using quiz scores designed to assess conceptual understanding of statistical inference. Results suggest that students who participated in the experiential activity performed better on inference-related questions than those who participated in the traditional format. By combining a detailed instructional account with formal evidence of learning gains, this article contributes to the growing literature on experiential learning pedagogy, active learning, and econometrics and statistics education, particularly in settings where students benefit from inclusive, hands-on approaches to learning technical material (Allgood et al. 2015; Angrist and Pischke 2017; Kassens 2019; Simons 2020).

2 Case Study

This section describes the institutional context, course structure, and student background relevant to the pedagogical intervention examined in this article.

2.1 Institutional and Course Background

Since January 2023, I have taught the Introduction to Econometrics course at Azim Premji University, a private liberal arts institution in India. The course is a core requirement for all economics majors enrolled in the BA (Honors) Economics program. It is typically offered in the second year and serves as the second course in the quantitative sequence, following Introduction to Statistics in the first year.

The course introduces students to the fundamentals of statistical inference, including estimation, hypothesis testing, and simple and multiple linear regression. The primary software used for statistical analysis is R, which is integrated throughout the course for demonstrations, assignments, and empirical exercises. Each cohort typically includes 70–75 students from diverse academic and linguistic backgrounds. Owing to the large enrollment, the course is conducted in two parallel sections, one of which I have taught continuously from 2023 onwards. Roughly half the students have studied advanced mathematics in high school, through international, national, or state-level curricula that include calculus-level content. The remaining students often come from commerce or humanities streams with limited formal mathematics training beyond Grade 10. Furthermore, several students also come from non-English-medium schooling backgrounds, having transitioned to English instruction only in their final years of schooling. This creates additional challenges when engaging with technical content presented in formal academic English.

Despite a well-structured curriculum, I encountered several recurring instructional challenges, particularly during my first time teaching the course. Students consistently reported difficulty in grasping foundational concepts such as confidence intervals, *p*-values, and the distribution of sampling statistics. Many expressed anxieties around mathematics and found it challenging to engage with abstract statistical reasoning. Even students who performed well on assessments often lacked clarity on the underlying logic of inferential procedures. Topics such as t-tests and regression assumptions were frequently cited as confusing, and the transition from theory to implementation in R programming was especially steep for many.

These issues were reflected in course feedback. One student noted that "it is a little difficult for the non-mathematics background students to follow… we had to face some problems in regard to understanding the technical aspects." Others emphasized the need for more intuitive and applied instruction, commenting "I wish we did more concepts in R, and it would help to focus on the intuition of each concept as well" and "more hands-on participation is required." Some also called for increased classroom interaction: "Instructors must increase class participation." These challenges and reflections motivated me to develop and implement a set of experiential and intuitive teaching strategies, described in the following sections.

2.2 Making Statistical Inference Intuitive: A Pedagogical Intervention

In the third iteration of the course, offered in 2025, I introduced two experiential teaching techniques aimed at making statistical inference more engaging,[1] active, and accessible for students. The first intervention was an in-person classroom game designed to illustrate the Central Limit Theorem (CLT) and to highlight conditions under which its assumptions might be violated. This hands-on activity allowed students to visualize the behavior of sampling distributions and understand their importance in shaping inference. The second intervention involved a real-world video case study from the world of professional cricket to demonstrate the logic of hypothesis testing.[2] The example was selected to be both relatable and memorable while also providing a concrete structure through which key ideas—such as null and alternative hypotheses, test statistics, and *p*-values—could be explored in context.

2.2.1 Demonstrating the Central Limit Theorem Through Classroom Sampling

To make the Central Limit Theorem (CLT) more tangible, I designed a short classroom activity built around a simple sampling experiment using everyday objects—chocolates of varying weights. Students worked in pairs, drawing random samples of chocolates from a bag, weighing them using a digital scale, and calculating sample means across repeated trials. In the first version of the activity, the bag contained a large number of identical chocolates, approximating a population of independent and identically distributed (i.i.d.) observations. The resulting distribution of sample means closely resembled a bell curve and had an expected value near the known population mean. In the second version, I used a bag with a mix of different types of chocolates (e.g., toffees, bars) to violate the i.i.d. assumption. This led to a more dispersed and biased distribution of sample means, helping students see how violations of core assumptions can distort inference. The activity was completed within a 20-minute window and generated a lively discussion around sampling, distributional assumptions, and the role of the CLT in inference. A detailed description of the activity setup and classroom results are included in Appendix B.

[1] The term "inference" here refers to frequentist statistical inference, which forms the foundation of most undergraduate econometrics curricula. The distinction between Frequentist and Bayesian approaches is briefly introduced in the course to provide context, but the focus remains on the frequentist framework typically followed in standard textbooks.

[2] Cricket is a hugely popular sport in India, and its cultural familiarity made it an effective context for engaging students with abstract statistical concepts.

These activities reinforce the importance of the i.i.d. assumption underlying the Central Limit Theorem (CLT). In Activity 1, this assumption holds, resulting in a distribution of sample means that is centered closely around the true population mean (see Table A1 and Figure A1 in Appendix A). By contrast, Activity 2 violates the i.i.d. assumption due to the presence of heterogeneous chocolate weights, leading to a biased estimate of the mean (as illustrated in Figure A2 in Appendix A). Notably, the sampling distribution from Activity 1 resembles a bell-shaped curve, visually confirming the key insight of the CLT—even when the population distribution is not normal, the distribution of the sample mean will be approximately normal if the sample size is sufficiently large and the i.i.d. condition holds. To further demonstrate this concept, instructors can complement the classroom activity with a simulation exercise, where students draw repeated random samples from various non-normal distributions such as Poisson, binomial, or exponential. As students compute and plot the distribution of sample means for each case, they can observe that—regardless of the shape of the original population—the resulting sampling distribution converges to a normal one, reinforcing the robustness and power of the CLT in practice.[3]

Beyond illustrating the CLT, this activity provides a foundation for introducing the concept of descriptive statistics and estimation. (See Appendix C for more details.) Students can begin by computing basic summary statistics such as mean, standard deviation, range, and sample size-for their samples. This step not only helps them describe the properties of their data but also mirrors the way empirical research is typically reported. In journal articles, the first table often provides summary statistics, giving readers an immediate sense of the dataset's scope, variation, and potential issues. For students, replicating such tables builds both technical competence and an appreciation of how quantitative data is summarized and communicated in published research.

Once the summary statistics are computed, the instructor can introduce the idea of estimation, specifically, how we use a sample to infer something about a population. A helpful way to motivate this discussion is by returning to the activity itself: Each group collected a different sample, leading to variation in the sample means across the classroom. This variation is not error in the conventional sense but reflects the randomness inherent in the sampling process. It is this variation that motivates the need for interval estimation.

The instructor can then walk students through the construction and interpretation of a 95 percent confidence interval. This is an opportunity to emphasize that while each group has only one sample mean, the set of all possible sample means from repeated sampling would form a distribution. A 95 percent confidence interval provides a range of plausible values for the true population mean, derived from a sample, and reflects the uncertainty of sampling. The correct interpretation, often misunderstood, should be emphasized: If we were to repeat the sampling process many times, approximately 95 percent of the resulting confidence intervals would contain the true population mean.

Using the summary statistics already computed, each group can construct their own 95 percent confidence interval. The class can then discuss whether the actual population mean (as revealed by the instructor at the end of the activity) falls within each group's interval. This exercise reinforces the interpretation of confidence intervals and brings clarity to questions such as: What does the "95 percent" refer to? Why do some intervals not include the population mean? What affects the width of the interval?

Overall, the CLT activity helps students understand how sampling works, why assumptions like i.i.d. matter, and how sample means behave across repeated samples. It provides a practical introduction to key ideas such as summary statistics, estimation, and confidence intervals—forming the foundation for statistical inference used widely in economics and related fields.

[3] R code for the simulation is provided in Appendix D.

2.2.2 Understanding Hypothesis Testing Through Sports

Sports offer a natural and engaging context for introducing the logic of hypothesis testing. In games such as tennis and cricket, players are allowed to challenge official decisions using technology. These challenges closely parallel the fundamental structure of hypothesis testing: There is an initial decision (the null hypothesis), a challenge to that decision (the alternative hypothesis), and evidence provided through systems like Hawk-Eye in tennis or the Decision Review System (DRS) in cricket to determine whether the original decision should stand.

To make this abstract logic more relatable, I showed students a short video clip from a cricket match involving a leg before wicket (LBW) appeal that was reviewed using DRS.[4] In this example, the on-field umpire's initial call, say "not out," represents the null hypothesis. A review by the fielding team acts as the alternative hypothesis, asserting that the batter is actually out. The DRS provides data such as ball trajectory, point of impact, and projected path toward the stumps. A final decision is then made based on predefined rules. For instance, to overturn a "not out" call, more than 50 percent of the ball must be projected to hit the stumps. This requirement parallels the significance level (e.g., 5 percent) in statistical inference, which defines how strong the evidence must be to reject the null hypothesis.

The analogy also helps illustrate the concept of a *p*-value. If the ball is clearly projected to hit the stumps, the probability of observing such an outcome under the null hypothesis (that the batter is not out) is low. This constitutes strong evidence against the null—akin to a low *p*-value—and justifies overturning the original call. Conversely, if the projection is marginal, the probability under the null is higher, resulting in a higher *p*-value and insufficient evidence to reverse the decision. Through class discussion, I used this example to emphasize that the *p*-value reflects the probability of observing data as extreme as (or more extreme than) the actual data, assuming the null hypothesis is true—not the probability that the null hypothesis itself is true.

This activity also provided an intuitive framework for explaining Type I and Type II errors. A Type I error occurs when the umpire's original decision is correct (the batter is not out) but is incorrectly overturned, analogous to rejecting a true null hypothesis. A Type II error occurs when the original decision is incorrect (the batter is out), but insufficient evidence leads to the decision being upheld, analogous to failing to reject a false null hypothesis.

These scenarios are intuitive for students because they link statistical logic to concrete outcomes: A Type I error results in a wrongful dismissal, while a Type II error allows a batter to continue unfairly. By encouraging students to map each element of the DRS process—initial call, challenge, evidence, and final decision—onto the formal framework of hypothesis testing, this analogy provided a meaningful and memorable way to engage with otherwise abstract concepts such as significance levels, *p*-values, and decision errors in statistical inference.

3 An Evaluation of the Interventions

The experiential teaching methods described above were introduced during the third iteration of the Introduction to Econometrics course, taught in early 2025. The course is a core requirement offered to all economics majors during the winter semester (January–April) in the second year of their undergraduate program. In contrast, the first two iterations of the course, offered in 2023 and 2024, followed a traditional chalk-and-board lecture format with minimal use of interactive or applied classroom activities. This shift in instructional design creates a natural experiment to evaluate the effectiveness of experiential pedagogy.

To assess the impact of this intervention, I use student performance on a short in-class quiz administered uniformly across all three cohorts. The full sample consists of 95 students: 60 students from the two traditional cohorts (31 in 2023 and 29 in 2024) and 35 students from the experiential

[4] See here for a simple explanation on LBW. The video clip can be accessed here.

cohort taught in 2025. All three cohorts were taught by the same instructor using an identical syllabus, assessment structure, and learning resources. The only systematic difference across cohorts was the instructional method adopted for the experiential cohort.

I chose the quiz scores as the primary outcome measure rather than written examination scores because the latter primarily emphasize numerical procedures and practice-driven problem solving. In contrast, the quiz was explicitly designed to test conceptual understanding, making it more sensitive to differences in instructional method.

Table 1 presents summary statistics for key variables, disaggregated by teaching cohort. While the cohorts are broadly similar in terms of gender composition, GPA in the first year, and age, quiz scores are notably higher in the experiential learning group.

Table 1. Summary statistics by teaching method

	Mean	SD	Min.	Max.	*N*
Traditional chalk & board cohort					
Score on quiz (out of 10)	5.10	2.82	0	10	60
GPA in first year	8.00	0.93	5.40	9.60	60
Age	18.21	0.45	17	19	60
Female	0.58	0.49	0	1	60
MathSupport	0.73	0.44	0	1	60
Experiential learning cohort					
Score on quiz (out of 10)	6.94	2.28	2	10	35
GPA in first year	7.80	0.99	5.00	9.50	35
Age	18.22	0.42	18	19	35
Female	0.71	0.45	0	1	35
MathSupport	0.57	0.50	0	1	35
All cohorts					
Score on quiz (out of 10)	5.78	2.77	0	10	95
GPA in first year	7.93	0.95	5.00	9.60	95
Age	18.22	0.44	17	19	95
Female	0.63	0.48	0	1	95
MathSupport	0.67	0.47	0	1	95

Notes: This table presents summary statistics of key variables by teaching cohort. GPA is measured on a 10-point scale. *Female* is a binary variable equal to 1 if the student identifies as a female. *MathSupport* is a binary variable equal to 1 if the student was enrolled in a summer support course on quantitative methods.

To formally estimate the relationship between teaching method and quiz performance, I use the following regression specification:

(1) $$Q_i = \alpha + \beta EL_i + X_i'\gamma + \lambda t + \varepsilon_i ,$$

where Q_i is the quiz score (out of 10) of student i and EL_i is a binary indicator equal to 1 if student i belonged to the experiential learning cohort (2025), and 0 otherwise. The vector X_i includes student-level controls—first-year GPA, gender, and age—to account for observable differences in academic ability and demographics. The error term ε_i captures idiosyncratic determinants of quiz performance. The term

t denotes a linear cohort time trend, capturing smooth changes in academic performance that would have occurred absent the intervention, such as increasing familiarity with educational technology or shifts in institutional standards.

The coefficient β therefore captures the average difference in quiz scores between students taught using experiential learning methods and those taught using traditional methods, conditional on observable student characteristics and underlying time trends. Under the assumption that, absent the introduction of experiential learning, quiz performance would have continued to evolve linearly across cohorts, β can be interpreted as the causal effect of the teaching method on quiz outcomes.

Table 2. Effect of experiential learning method on quiz performance

	(1) Baseline OLS	(2) OLS w/Controls	(3) Time Trend	(4) Robust SE
EL	1.843***	1.778***	1.780***	1.780***
	(0.561)	(0.514)	(0.608)	(0.623)
Female		0.846*	0.847	0.847*
		(0.505)	(0.516)	(0.489)
GPA in first year		1.043***	1.044***	1.044***
		(0.298)	(0.305)	(0.275)
Age		0.177	0.177	0.177
		(0.559)	(0.563)	(0.586)
MathSupport		-1.046*	-1.045	-1.045
		(0.615)	(0.642)	(0.686)
Constant	5.100***	-6.198	-6.203	-6.203
	(0.341)	(10.607)	(10.715)	(10.930)
R^2	0.104	0.355	0.355	0.355
No. of obs.	95	95	95	95

Notes: The table presents least squares estimates of the pedagogical intervention on quiz performance. Dependent variable is quiz scores (out of 10). EL is a binary indicator equal to 1 if the student was in the experiential learning cohort, and 0 otherwise. Column (1) presents baseline results; Column (2) adds controls for observed student characteristics (GPA, Gender, *MathSupport* and Age) where *MathSupport* takes the value of 1 if the student was enrolled in a summer quantitative methods course and 0 otherwise; column (3) includes linear time trend. Column (4) reports OLS estimates in column (3) with standard errors corrected for heteroskedasticity. Standard errors are in parentheses. Asterisks indicate significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

3.1 Main Results

Table 2 reports the estimated impact of the experiential learning intervention on student quiz performance using ordinary least squares (OLS), based on the regression model specified in Equation (1). In the baseline specification (column 1), students exposed to experiential teaching scored, on average, 1.84 points higher than those in the traditional lecture-based cohort. Controlling for student characteristics—first-year GPA, gender, age, and math support status—in column 2 slightly reduces the estimated effect to 1.78 points. This estimate remains stable with the inclusion of year fixed effects in column 3, and column 4 replicates the same specification with heteroskedasticity-robust standard errors.

Across all specifications, the coefficient on *EL* remains statistically significant at the 1 percent level. Given a baseline average quiz score of 5.1 in the traditional cohort, the fully controlled estimate of 1.78 points represents 0.64 standard deviation (SD) increase in performance. These findings suggest that the experiential learning intervention significantly enhanced students' conceptual grasp of inferential statistics, improving their ability to reason about estimation and hypothesis testing.

3.2 Heterogeneity Analysis

To explore whether the experiential learning intervention had differential effects based on students' prior quantitative preparedness, I conducted a heterogeneity analysis using an interaction between the treatment indicator, *EL*, and a variable capturing math background. At the end of the first academic year, all economics majors take a diagnostic test to assess readiness for quantitatively demanding coursework. Students identified as weak in mathematics are required to enroll in a compulsory summer course on quantitative methods, which introduces foundational topics such as set theory, functions and relations, linear algebra, and basic calculus. I use enrolment in this support course as a proxy for quantitative weakness, coded as the binary variable *MathSupport*, and estimate the following modified model from equation (1):

(2) $$Q_i = \alpha + \beta_1 EL_i + \beta_2 EL_i * MathSupport_i + X_i' \gamma + \lambda t + \varepsilon_i \, .$$

Table 3 reports the least squares estimates based on the regression specification in Equation (2). Across all specifications, the coefficient on *MathSupport* is negative and statistically significant, indicating that, on average, students who enrolled in the summer quantitative methods course performed worse on the quiz than those who did not take the support course. This finding is consistent with the interpretation that students who opted into the support course entered with weaker quantitative preparation.

Moreover, the coefficient on the interaction term *EL × MathSupport* in column (4) is 1.52 and statistically significant at the 5 percent level. This implies that, among students in the experiential learning cohort, those who had been enrolled in the math support course scored, on average, 1.52 points higher on the quiz than their peers in the same cohort who were not enrolled in the support course. This corresponds to an effect size of approximately 0.55 standard deviations. In other words, within the experiential learning group, students with weaker quantitative backgrounds appear to have gained more from the intervention than their better-prepared counterparts.

The cumulative effect of experiential learning for students who also received math support is given by the sum of the coefficients on EL and EL × *MathSupport* (i.e., $\beta_1 + \beta_2$). In column (4), this cumulative effect is 2.567, implying that a student in the experiential cohort with prior math support scored, on average, about 2.5 points higher on the quiz than an otherwise comparable student in the traditional cohort without math support (approximately 0.92 SD increase). Although the baseline *EL* effect is only weakly statistically significant, the sizable cumulative effect suggests that the intervention was particularly beneficial for students with weaker quantitative preparation.

3.3 Qualitative Feedback on Experiential Methods

In addition to performance-based evidence, student feedback collected at multiple points in the semester reinforces the effectiveness of the experiential teaching approach. For the 2025 cohort, immediately after the mid-semester examination, an anonymous survey was conducted with a 100 percent response rate from students in the experiential learning cohort (see Figure 1). When asked to rate how effective the new pedagogical methods (e.g., the classroom sampling game and the DRS cricket video) were in helping them understand estimation and hypothesis testing, 57 percent of students rated the methods as "Very Effective" and 34 percent as "Somewhat Effective," with very few students rating them negatively.

Table 3. Heterogeneous effects of experiential learning on quiz performance

	(1) Baseline OLS	(2) OLS w/ Controls	(3) Time Trend	(4) Robust SE
EL	0.608	0.839	1.047	1.047*
	(0.877)	(0.830)	(1.176)	(0.141)
MathSupport	-2.761***	-1.650**	-1.712**	-1.712*
	(0.712)	(0.743)	(0.786)	(0.473)
EL * *MathSupport*	1.378	1.474	1.520	1.520**
	(1.096)	(1.026)	(1.048)	(0.239)
Female		0.854*	0.832	0.832
		(0.502)	(0.513)	(0.384)
Age		0.113	0.110	0.110
		(0.558)	(0.561)	(0.166)
GPA in first year		1.048***	1.035***	1.035**
		(0.296)	(0.303)	(0.240)
Constant	7.125***	-4.645	-4.182	-4.341
	(0.610)	(10.600)	(10.815)	(5.296)
Cumulative Effect - EL	1.986***	2.313***	2.567**	2.567***
	(0.065)	(0.063)	(1.196)	(0.329)
R^2	0.250	0.370	0.370	0.370
No. of obs.	95	95	95	95

Notes: The table presents least squares estimates of the pedagogical intervention on quiz performance. Dependent variable is quiz scores (out of 10). *EL* is a binary indicator equal to 1 if the student was in the experiential learning cohort, and 0 otherwise. Column (1) presents baseline results; column (2) adds controls for observed student characteristics (GPA, gender, and age); column (3) includes a linear time trend. Column (4) reports OLS estimates in column (3) with standard errors corrected for heteroskedasticity. Standard errors in parentheses. Asterisks indicate significance levels: $^{*}p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$.

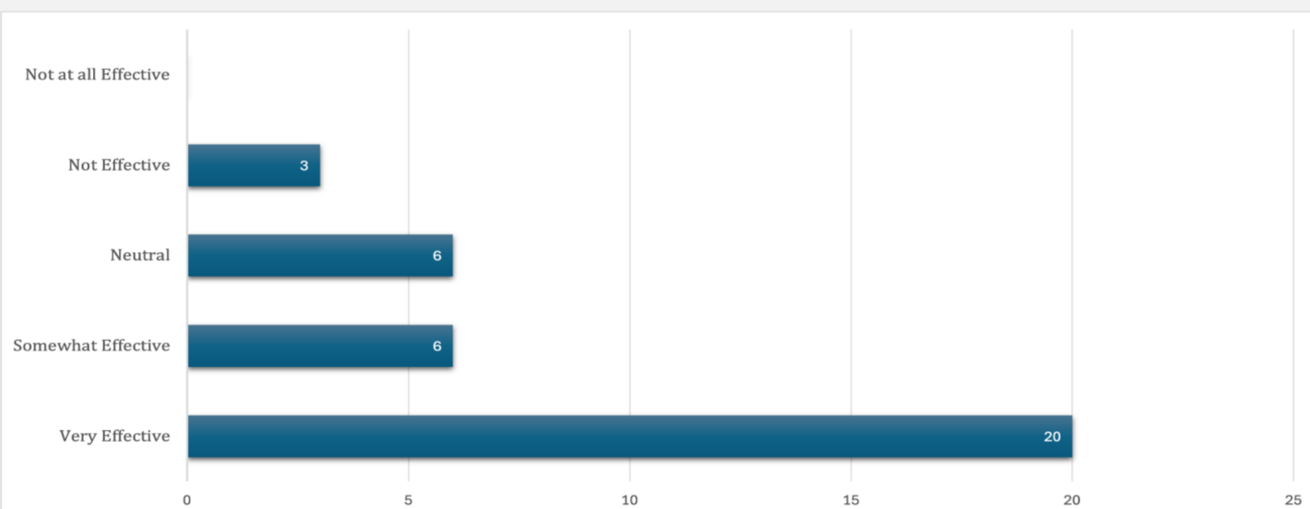

Figure 1. Student perceptions of the effectiveness of experiential learning methods

Note: The figure reports anonymous responses from a midsemester survey (N = 35). Students rated the effectiveness of experiential activities in understanding concepts related to estimation and hypothesis testing.

This was further supported by the end-of-semester feedback (see Figure 2), in which students rated both the course and the instructor highly. The experiential learning cohort (2025) reported an average course feedback score of 4.56 and an instructor rating of 4.69 on a 5-point scale, the highest among the past 3 years. These positive ratings suggest that the shift toward active, intuitive instruction not only improved learning outcomes but also enhanced the classroom experience from the students' perspective.

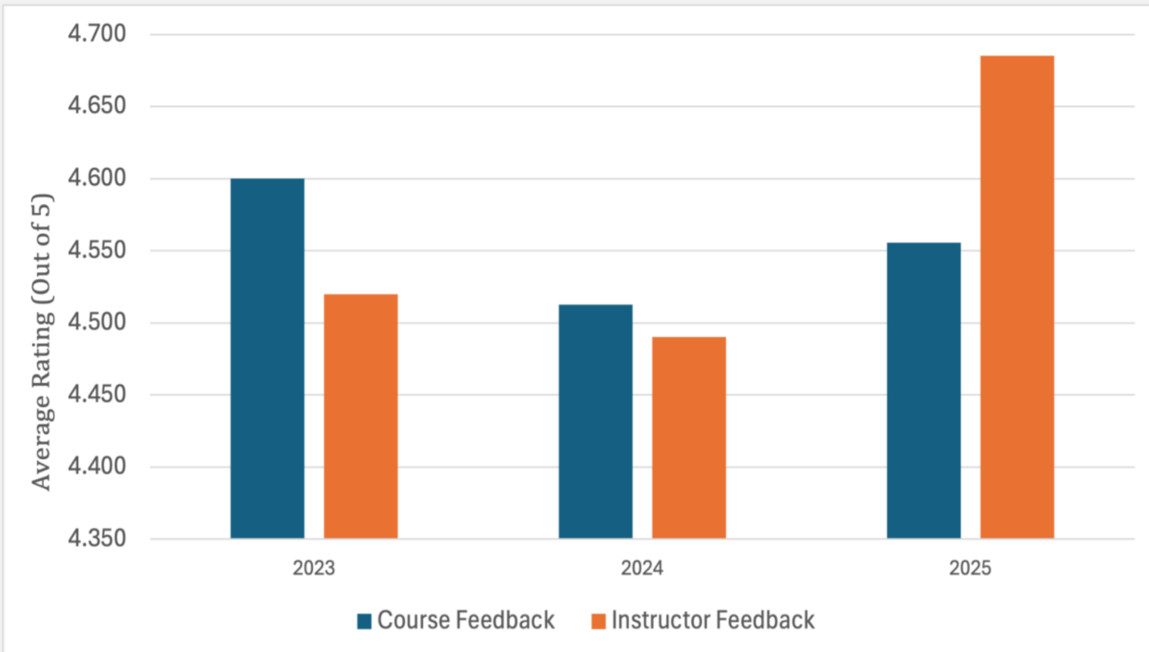

Figure 2. End-of-semester student feedback

Note: The figure reports end-of-semester student feedback scores on course and instructor effectiveness across cohorts. Sample sizes are N = 24 for 2023, N = 17 for 2024, and N = 14 for 2025.

Finally, in open-ended responses to the end-of-semester feedback form, several students highlighted the value of experiential and intuitive teaching methods in supporting their learning. Many pointed to the classroom activities and hands-on experiments as particularly effective, with one student noting, "The activities done in class and the instructor's notes accompanying the concepts were very helpful," and another reflecting that "Hands-on learning was great." The chocolate sampling experiment designed to illustrate the Central Limit Theorem (CLT) stood out in multiple responses; one student remarked, "I liked the part of the course where we actually got to know the CLT using the chocolates experiment." Others appreciated the integration of R simulations and the use of intuitive examples, which helped bridge theory and application. As one student observed, "All the different examples and classroom activities we did in class proved to be helpful in learning the material," while another emphasized that "Intuitive thinking is encouraged by the instructor via discussions and classroom games." Notably, one student commented that the concepts learned in this class helped them better understand material in other courses in the major, underscoring the broader relevance of developing statistical intuition through active learning.

4 Conclusion

This article reflects on a pedagogical intervention implemented in an undergraduate econometrics course at a liberal arts university in India. Motivated by persistent instructional challenges—particularly students' difficulty in grasping core inferential concepts such as confidence intervals, *p*-values, and sampling distributions—I introduced a set of experiential teaching methods designed to make statistical inference more intuitive, interactive, and accessible. These included a hands-on classroom activity to demonstrate the Central Limit Theorem using everyday objects and a real-world video case from cricket to explain the logic of hypothesis testing.

Quantitative evidence from a natural experiment setting—where the same instructor taught the course using traditional methods in two years and experiential methods in the third—indicates that the intervention led to a substantial improvement in students' conceptual understanding. Students in the experiential cohort scored, on average, 1.78 points higher (out of 10) on a reasoning-based quiz, corresponding to a 0.64 standard deviation increase and a gain of over 35 percent relative to the traditional cohorts. Importantly, heterogeneity analysis suggests that the gains were even larger among students with weaker quantitative backgrounds, indicating that the intervention helped narrow performance gaps.

These findings are reinforced by qualitative feedback collected during and at the end of the semester. Students overwhelmingly rated the new methods as effective and emphasized the value of intuitive thinking, hands-on activities, and relatable analogies in helping them understand abstract statistical ideas. The positive feedback suggests that experiential approaches not only enhance learning but also improve student engagement and confidence, particularly for those with limited prior exposure to formal mathematics.

In sum, the results support the case for rethinking how statistical inference is taught in undergraduate economics. While rigorous theory and formal methods remain essential, introducing active and intuitive learning strategies can play a powerful role in bridging the gap between abstract concepts and real-world understanding—especially in diverse classrooms where students arrive with varying levels of preparation. The interventions described here are low-cost, easily replicable, and adaptable to a wide range of teaching contexts, making them a promising addition to the instructional toolkit for quantitative economics education.

About the Authors: Kedar Kulkarni (kedar.kulkarni@apu.edu.in) is an Assistant Professor, School of Arts and Sciences, Azim Premji University, India.

Acknowledgments: The author gratefully acknowledges Anand Shrivastava and Kade Finnoff for their valuable comments during the initial conceptualization and implementation of the methods. The author also thanks the undergraduate economics students at Azim Premji University for their constructive feedback, which helped refine the design and presentation of the activity. The author further thanks the editor and two anonymous reviewers for their constructive and insightful feedback, which substantially improved the quality of this manuscript. This article does not present any conflicts of interest or financial support for the research conducted. The study has been reviewed and approved by the Azim Premji University Institutional Review Board (APU-SAS-2025-07) and found to be in compliance with relevant policies and regulations concerning research involving human subjects. The findings, interpretations, and conclusions expressed in this article are solely those of the author and do not necessarily reflect the views of Azim Premji University.

AI Disclosure: Generative AI tools (ChatGPT 5.1) were used to support language editing, grammar correction, and stylistic refinement of the manuscript. AI was also used to assist with cleaning and formatting regression tables for presentation purposes. All substantive intellectual content, data analysis, interpretation of results, and final editorial decisions were performed by the author. The author reviewed, verified, and takes full responsibility for all content in the manuscript.

Appendix A: Results from the Classroom Activity

Table A1. Comparison of population and sample means from classroom activities		
Activity	**Population Mean (grams)**	**Sample Mean (grams)**
Activity 1 (i.i.d.)	3.75	3.82
Activity 2 (non-i.i.d.)	4.38	5.91

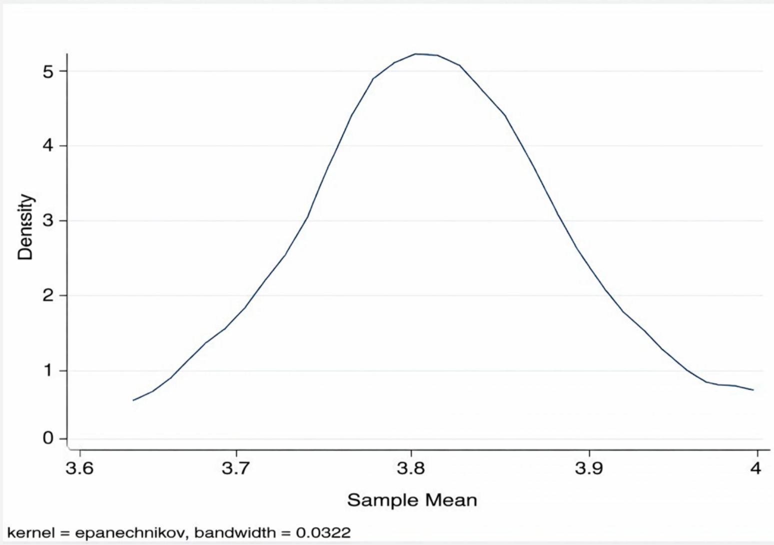

Figure A1. Distribution of sample means (in grams) under i.i.d. assumption (Activity 1)

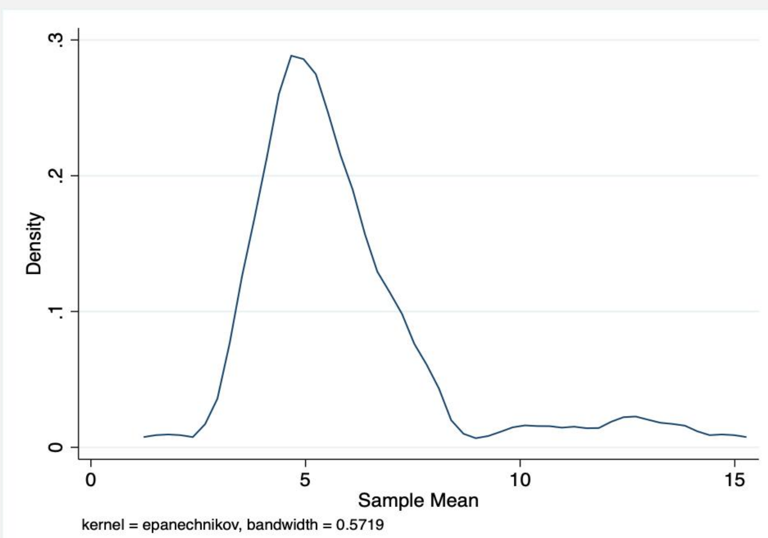

Figure A2. Distribution of sample means (in grams) under violation of i.i.d. assumption (Activity 2)

Appendix B: Implementation Guide for Experiential Activities

B1 Activity 1: Sampling from an Identical Distribution

Aim: To Demonstrate the Central Limit Theorem under ideal i.i.d. conditions
Materials: A single-type pack of chocolates, a bag, a serving bowl, digital scale, record sheets/spreadsheet, pens/pencils
Duration: 25 minutes

B1.1 Setup (~ 4 min)

Students are seated at their own desks throughout the activity. Each group consists of two students, and prior to the start of the activity, every group is assigned a unique Group ID (e.g., G1, G2, …) for ease of data recording and coordination. All required materials for each group (chocolates, serving bowl, and access to a digital scale) are either distributed in advance or rotated systematically by the instructor/teaching assistant so that each group can perform the sampling and weighing directly at their desks.

For large classes (more than 50 students), where it may not be feasible for every group to participate or where classroom management becomes challenging, a subset of the class actively conducts the experiment while the remaining students engage with pre-assigned discussion questions related to the expected outcomes and theoretical implications. These students are asked to reflect on the questions during the activity and later contribute to the collective discussion. This approach ensures that all students are actively engaged, either through hands-on experimentation or through guided analytical participation.

B1.2 Explanation of the Activity and Demonstration by the Instructor (~4 min)

At the beginning of the activity, the instructor briefly reminds students of the statement and intuition of the Central Limit Theorem (CLT). The aim of the exercise is then clearly stated as follows: "Today we will generate our own data to observe the CLT in action under ideal i.i.d. conditions." To activate prior thinking, the class is asked to predict the average weight of a single chocolate in the pack. A few guesses are invited from students, but the actual value is not revealed at this stage.

Next, the instructor explains the sampling procedure by demonstrating one round of data collection. A handful of chocolates is randomly drawn from the bag into the serving bowl ***without looking into the bag***, so that the idea of random sampling is maintained. The bowl is then placed on the digital scale to record the total weight. The number of chocolates in the bowl is counted, and the known weight of the empty bowl is subtracted to obtain the net weight. The average weight per chocolate is then computed as the net weight divided by the number of chocolates. Finally, the instructor shows students how and where these values should be entered into the record sheet for systematic data collection.

B1.3 Implementation of the Activity by the Students (~7 min)

Each group is instructed to complete five independent sampling rounds. Within each group, one student acts as the sampler and randomly draws a handful of chocolates from the bag into the serving bowl without inspecting individual pieces, in order to approximate random sampling. The second student acts as the recorder and notes the total weight displayed on the digital scale and the number of chocolates in the sample and then enters these values into the record sheet. For every round, each group records the activity number (e.g., Activity 1), their assigned Group ID, the round number (1–5), the total number of chocolates drawn, and the combined weight of the bowl and chocolates. This process is repeated until all groups have completed five rounds of sampling ($N = 5$). To save class time and facilitate later analysis, a shared spreadsheet (see the sample record sheet in Table B1) may be created and circulated to students before class, preferably through the course learning management system (e.g., Canvas, Blackboard, or Moodle). Students then enter their values directly into this spreadsheet, which also allows for faster aggregation and visualization.

B1.4 Discussion and Interpretation of Results (~10 min)

While the instructor generates the distribution of sample means using the values entered by the students (typically over 2–3 minutes), the class is asked to collectively reflect on the following guiding questions:

B1.4.1 What Is a Sample Mean and How Is It Calculated? Is It a Random Variable? Why? How Many Sample Means Have We Generated?
Students are first asked to recall that a sample mean is the average value computed from a single random sample and is therefore itself a *random variable*, since its value depends on the particular random draw of chocolates in each round. In the context of this activity, the average weight per chocolate calculated in each round by each group represents one realization of this random variable. Since every group performs five rounds of sampling, the total number of observed sample means is equal to five times the number of participating groups. This makes it clear that the class has collectively generated multiple realizations of the same random variable, thereby constructing an empirical sampling distribution.

B1.4.2 If All the Sample Means (Random Variables) Are Plotted as a Histogram or a Continuous Density Function, What Should It Ideally Look Like? Why?

Students are asked to predict the shape of the resulting distribution before it is displayed. They are guided to recall that each chocolate weight is itself a random variable drawn from the same population and that repeated samples are approximately independent. Under these ideal i.i.d. conditions, the Central Limit Theorem states that the distribution of the sample mean—despite being formed from underlying random variables of possibly unknown shape or form—converges to an approximately normal distribution as the sample size increases. Accordingly, the histogram or density of the sample means is expected to exhibit a bell-shaped, approximately normal curve.

B1.4.3 What Should the Mean of the Distribution Resemble?

Students are then asked to consider the expected value of the random variable, the sample mean. They are encouraged to reason that this expected value should be close to the true population mean, that is, the actual average weight of one chocolate in the pack. Once the empirical distribution is displayed, the instructor reveals the true average weight of a chocolate (obtained from the full pack) and asks students to verify whether it matches the observed mean of the sample means. This comparison highlights both the unbiasedness of the sample mean as an estimator and the probabilistic nature of convergence implied by the Central Limit Theorem under ideal i.i.d. assumptions.

B2 Activity 2: Sampling from a Mixed Distribution

Aim: To explore the consequences of violating the i.i.d. assumption
Materials: A mixed-pack of chocolates, a bag, a serving bowl, digital scale, record sheets/spreadsheet, pens/pencils.
Duration: 15 minutes

B2.1 Setup (~1 min)

Students remain seated at their desks throughout the activity. Each group consists of two students and is assigned a unique Group ID prior to the start of the activity. Materials are either distributed in advance or rotated among groups by the instructor/teaching assistant so that each group can carry out the sampling and weighing at their own desks. To ensure wider participation, students who did not actively perform Activity 1 are encouraged to take part in Activity 2. Any remaining nonparticipants are provided with guiding questions in advance and asked to reflect on the expected outcomes while observing the activity, so that all students remain engaged.

B2.2 Explanation of the Activity and Demonstration by the Instructor (~1 min)

The instructor briefly recalls the key idea of the Central Limit Theorem and explains that, unlike Activity 1, the chocolates are now drawn from a *mixed population*. The sampling and recording procedure is the same as in Activity 1 and is demonstrated quickly using one round: A handful of mixed chocolates is drawn into the bowl, weighed, counted, and the net and average weights are computed. Students are reminded to enter their values in the same record sheet or shared spreadsheet used previously.

B2.3 Implementation of the Activity by the Students (~7 min)

Each participating group completes five sampling rounds following the same procedure as in Activity 1. For each round, groups randomly draw a handful of mixed chocolates, record the total number of chocolates and the combined weight of the bowl and contents, and enter these values under Activity 2 in the record sheet or shared spreadsheet. Data entry may again be done directly in a precirculated spreadsheet via the course learning management system to enable quicker aggregation and visualization.

B2.4 Discussion and Interpretation of Results (~6 min)

While the instructor generates the distribution of sample means using the values entered by the students, the class is asked to collectively reflect on the following questions:

B2.4.1 What Should the Distribution of the Sample Means Look Like in This Case? Will It Still Be Normal? if Yes, Why? if No, Why Not?

Students are asked to predict the shape of the sampling distribution before it is displayed. They are reminded that, unlike Activity 1, the underlying population now consists of a mixture of different types of chocolates with different weights. As a result, the individual observations are no longer i.i.d. Students are encouraged to reason whether the CLT should still apply and under what conditions approximate normality might still emerge. When the empirical distribution is displayed, they observe that the shape may deviate from the bell-shaped curve seen earlier, for example showing skewness or increased variability, depending on the composition of the mixed pack.

B2.4.2 What Are the Consequences of Violating the i.i.d. Assumption, and Why Is This Assumption Central to Inferential Statistics?

The discussion then focuses on the implications of violating independence and identical distribution. Students are guided to recognize that many standard results in inferential statistics such as the Central Limit Theorem, properties of estimators, confidence intervals, and hypothesis tests critically rely on the i.i.d. assumption. When this assumption is violated, the sampling distribution of estimators can change in shape, variance, and bias, making usual inferential procedures unreliable or misleading. Through this activity, students directly observe that departure from i.i.d. conditions leads to greater uncertainty and less predictable sampling behavior, thereby highlighting why justifying the i.i.d. assumption is fundamental in statistical inference.

Table B1. Example of a record sheet for CLT activity

Activity No.	Group ID	Round No.	Total No. of Chocolates Drawn	Weight of the Bowl and Chocolates (in grams)	Weight of the bowl (in grams)	Net Weight of the Chocolates drawn (in grams)	Average Weight per Chocolate (in grams)

Appendix C: Point and Interval Estimators

After completing Activity 1 and visualizing the distribution of sample means under ideal i.i.d. conditions, the instructor uses the collected data to introduce students to descriptive statistics, estimation, and confidence intervals.

C1 Construction of Summary Statistics

Using their own Activity 1 data (with sample size $N = 5$ for each group), students are first asked to compute basic summary statistics for their sample, including the sample mean, sample standard deviation, minimum, and maximum. The instructor highlights that this mirrors the structure of a typical summary statistics table reported in economics and other empirical research. This step reinforces both computational skills and the role of descriptive statistics in empirical analysis.

C2 Introduction to Estimation: Point vs. Interval Estimators

Once summary statistics are computed, the instructor introduces the idea of statistical inference, specifically the problem of estimating the unknown population mean using sample data. Students are first introduced to *point estimation*, where the sample mean serves as an estimator of the population mean. Its limitation is emphasized: different samples yield different point estimates due to sampling variability. This naturally motivates the need for *interval estimation*, which explicitly accounts for uncertainty.

C3 Construction and Interpretation of Confidence Intervals

The instructor then introduces the most common confidence levels used in economics and the social sciences—90%, 95%, and 99%—and explains their interpretation in terms of repeated sampling. Each group is asked to construct such confidence intervals for the population mean using their sample data. The instructor emphasizes that, since the population standard deviation is unknown, the sample standard deviation is used as its best available estimator.[5] Here, emphasis should be on the correct

[5] For example, see Eisenhauer (2025).

interpretation of confidence intervals: "If the sampling process were repeated many times, approximately 95% of the resulting 95% confidence intervals would contain the true population mean."

To make this interpretation concrete, the instructor asks groups to indicate whether their 95% confidence interval contains the true population mean (previously revealed from Activity 1). The class observes that a small fraction of intervals—approximately 5%—do not contain the true population mean, providing a direct empirical validation of the confidence level.

Appendix D: Sample R Code for Simulating CLT

The below R code illustrates the Central Limit Theorem.[6] The underlying population distribution in the simulation can be altered from uniform to Poisson, binomial, or normal by modifying only the data-generating step. When this is done, the resulting distribution of sample means remains approximately normal in all cases, provided the sample size is sufficiently large.

```
# *********** CLT Simulation: Change Only the First Distribution ***********
rm(list=ls())
set.seed(1234)
n <- 30      # Sample size
N <- 1000    # Number of repeated samples

# ----- UNDERLYING POPULATION DISTRIBUTION -----

# 1. Uniform Distribution (original case)
data <- runif(n, min = 0, max = 15)

# 2. Poisson Distribution (uncomment to use)
# lambda = 5 is the average rate
# data <- rpois(n, lambda = 5)

# 3. Binomial Distribution (uncomment to use)
# size = 10 trials, prob = 0.4 success probability
# data <- rbinom(n, size = 10, prob = 0.4)

# 4. Normal Distribution (uncomment to use)
# mean = 10, sd = 2
# data <- rnorm(n, mean = 10, sd = 2)
# ------------------------------------------------------------------
# Histogram of original population sample
hist(data, main = "Histogram of Original Data", xlab = "Values")

# Sample means
sample_means <- c()

for (i in 1:N){
  sample_means[i] <- mean(sample(data, n, replace = TRUE))
}

# Mean and SD of sample means
x_bar <- mean(sample_means)
```

[6] The R Script and corresponding Stata Do File are provided in the supplementary materials.

```
stdev <- sd(sample_means)

print(x_bar)
print(stdev)

# Histogram of sample means
hist(sample_means, main = "Distribution of Sample Means",
     xlab = "Sample Means")
```

Appendix E: Sample Quiz Questions

The quiz used to assess conceptual understanding consisted of five short-answer reasoning questions. Each question was worth 2 points, for a total of 10 points. Students were asked to state whether each of the following statements was true or false and to provide a brief justification for their response. The quiz focused on fundamental ideas in estimation and hypothesis testing and was administered uniformly across all cohorts.

State whether each statement is true or false and provide a brief justification.

(1) Suppose you take two random samples of sizes, $N = 36$ and $N = 64$ to estimate the population mean. The confidence interval for the population mean will be wider for sample with $N = 36$.
(2) The Central Limit Theorem forms the basis for inferential statistics.
(3) A high p-value is evidence that the null hypothesis is true.
(4) Suppose two economists, Amit and Anand, collect data on wages in Indian agriculture to test the effect of the MNREGA policy on labor wages (consider the null hypothesis is of no effect). Amit concludes from his analysis that wages after the introduction of MNREGA are higher than wages before, when in fact, wages have remained the same. Anand, on the other hand, concludes that wages have not changed, even though they increased. The analysis conducted by both Amit and Anand is incorrect. Specifically, Amit has committed a Type I error, while Anand has committed a Type II error.
(5) (a) Increasing the sample size will decrease the margin of error.
(b) The larger the margin of error, the less confident one is about the estimates.

References

Allgood, S., W.B. Walstad, and J.J. Siegfried. 2015. "Research on Teaching Economics to Undergraduates." *Journal of Economic Literature* 53(2):285–325. https://doi.org/10.1257/jel.53.2.285

Angrist, J.D., and J.S. Pischke. 2017. "Undergraduate Econometrics Instruction: Through Our Classes, Darkly." *Journal of Economic Perspectives* 31(2):125–144. https://doi.org/10.1257/jep.31.2.125

Burch, G.F., R. Giambatista, J.H. Batchelor, J.J. Burch, J.D. Hoover, and N.A. Heller. 2019. "A Meta-Analysis of the Relationship Between Experiential Learning and Learning Outcomes." *Decision Sciences Journal of Innovative Education* 17(3):239–273. https://doi.org/10.1111/dsji.12188

Eisenhauer, J.G. 2025. "Impossible Inferences: A Note on Inconsistency in Business Statistics." *Applied Economics Teaching Resources* 7(4):1–6 . https://doi.org/10.71162/aetr.224304

Garfield, J., and D. Ben-Zvi. 2007. "How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics." *International Statistical Review* 75(3):372–396. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Kassens, A.L. 2019. "Theory vs. Practice: Teaching Undergraduate Econometrics." *Journal of Economic Education* 50(4):367–370. https://doi.org/10.1080/00220485.2019.1654958

Kolb, D.A. 2014. *Experiential Learning: Experience as the Source of Learning and Development*. FT Press. https://doi.org/10.1002/job.4030080408.

Park, R. 2019. "Practical Teaching Strategies for Hypothesis Testing." *American Statistician* 73(3):282–287. https://doi.org/10.1080/00031305.2018.1424034

Paul, R., and L. Elder. 2019. *The Miniature Guide to Critical Thinking Concepts and Tools*, 8th ed. Rowman & Littlefield.

Prince, M. 2004. "Does Active Learning Work? A Review of the Research." *Journal of Engineering Education* 93(3):223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Sigal, M.J., and R.P. Chalmers. 2016. "Play It Again: Teaching Statistics with Monte Carlo Simulation." *Journal of Statistics Education* 24(3):136–156. https://doi.org/10.1080/10691898.2016.1246953

Simons, A.M. 2020. "Making Business Statistics Come Alive: Incorporating Field Trial Data from a Cookstove Study into the Classroom." *Applied Economics Teaching Resources* 2(3):76–86. https://doi.org/10.22004/ag.econ.303918

Wood, M. 2005. "The Role of Simulation Approaches in Statistics." *Journal of Statistics Education* 13(3). https://doi.org/10.1080/10691898.2005.11910562

DOI: https://doi.org/10.71162/aetr.626500