Paper ID #49437

Improving Statistics Education at Wright State University with Design Project Based Learning, Problem Solving, and Peer Review

Dr. Trevor Joseph Bihl, Wright State University
Dr. Terry Lynn Oroszi, Wright State University
Subhashini Ganapathy, Wright State University
Jeffrey B. Travers, Wright State University

Improving Statistics Education at Wright State University with Design Project Based Learning, Problem Solving, and Peer Review

Abstract

Widespread knowledge of statistical methods is essential in engineering and medical disciplines, primarily for experiment design, data analysis, and interpreting significant results. Often, this knowledge is expected to be gained through a service course in statistics whereby the focus is on breadth and covering the practical extent of methods commonly used in a field. However, students frequently enter statistics courses with negative attitudes, compounded by traditional teaching methods that focus on trivial examples and a general lack real-world relevance. At Wright State University, innovative approaches in statistical education were implemented across the Biomedical, Industrial, and Human Factors Engineering Department and Pharmacology and Toxicology Department. These innovations include design project-based learning, where students undertake semester-long projects to collect and analyze data, and problem-solving focused on real-world ambiguity and complexity. Additionally, innovation-based learning in the form of peer review—through blinded reviews of class projects and critical evaluations of published work—was integrated to foster critical thinking and exposure to real-world applications of statistics. These changes resulted in greater student engagement, practical learning outcomes, and multiple peer-reviewed publications from course projects.

Introduction

Statistics is the backbone of decision-making and problem-solving in virtually every scientific, engineering, and medical field. It informs everything from experimental design and clinical trials to quality control in manufacturing and predictive modeling in engineering systems. As the world becomes increasingly data-driven, the ability to understand and apply statistical methods has become a critical skill not only for specialists but for a wide range of professionals [1] [2]. In disciplines such as biomedical engineering, pharmacology, and industrial engineering, statistical literacy enables advancements in patient care, drug development, and operational efficiency [3]. Moreover, industries and organizations now rely heavily on statistical techniques to derive actionable insights from vast quantities of data, underscoring its pervasive role in both innovation and routine operations [4].

Despite the importance of statistics, student attitudes towards statistics education are often negative even before classes [5]. While unfortunate, this is a logical byproduct of traditional classes in statistics focusing on rote learning of methods and regurgitation through exams. Such approaches do not address the value of the material on appropriate problems and possibly the material learned it quickly lost since many users of statistics are infrequent users and not experts in the methods and software. Schau [6] further posited that these issues revolved around *affect* (feelings), *cognitive competence* (attitudes around knowledge/skills), *value* (perceived usefulness), and *difficulty* (student perception of course difficulty).

Herein the authors develop both a set of graduate level statistics courses and repeatable modules for these courses that focuses first on *value*, whereby the usefulness of the material is demonstrated through design project learning, and *cognitive competence*, whereby expertise in the use of methods is gained through progressive use of methods. The result is a plan to provide improvement in the feelings about the topic (*affect*) as the students begin with rote learning, move to more relevant problems, and receive peer and professor feedback. While *difficulty* is intrinsically difficult to address, since a course has stated learning objectives, through relevant examples and assignments this burden is ideally reduced as well.

Background

Despite its significance, statistics education across educational levels often faces challenges related to content delivery, pedagogy, and student attitudes. The course framework described herein aims to address these issues and thus first what is needed is a general understanding of statistics education, as well as pedagogical concepts embraced, and finally statistical concepts that are the focus of these courses.

Statistics Education

Statistics education can generally be divided into two groups: the first includes those who gain essential statistical literacy for their domain-specific applications, and the second who pursue advanced expertise in statistics to develop, refine, or innovate methods. Survey courses are often used for the first group and these emphasize foundational knowledge, such as probability and statistical inference, applications to specific domains like health or engineering, and practical skills with tools for data analysis. In contrast, the second group focuses on the deeper theoretical and methodological aspects of statistics, aiming to advance the field through novel statistical models, innovative computational techniques, and the novel frameworks for data interpretation.

Perceptions created in K-12 and undergraduate experiences frequently lay the groundwork for negative attitudes regarding statistics. This both introduces biases into those taking survey courses and reduces interest in specializing in statistics. If statistics is presented in K-12 settings, the focus is often on procedural fluency over conceptual understanding, while undergraduate courses can emphasize rote memorization rather than practical applications [7]. While the primary focus of this paper is on graduate-level education, where students encounter the discipline in depth, these foundational issues play a critical role in shaping statistical literacy.

Learning Elements

Graduate statistics courses at Wright State University address the dual need for technical expertise and transferable skills by integrating design project-based learning, problem-solving, and peer review. These methods enhance critical thinking, communication, and the ability to evaluate statistical methods, preparing students for real-world interdisciplinary challenges. Project-based learning (PBL) provides a practical framework to engage students with real-world problems, fostering active participation and deeper learning [8] [9]. Studies have shown that PBL enhances problem-solving skills and promotes critical thinking by requiring students to design and implement solutions to complex tasks [10].

Peer review is an invaluable method for fostering critical evaluation and collaborative learning. It encourages students to assess the quality of others' work while reflecting on their own [11] [12] [13]. This process helps students refine their analytical skills and improves their understanding of how to evaluate statistical results and experiment designs. Research demonstrates that peer review is a useful learning tool [14] and also strengthens students' writing and oral presentation skills [15]-[19], which are essential for engineers to communicate their findings effectively. Incorporating peer review in presentations further ensures students develop the ability to critique technical content in professional settings [20].

In the context of reviewing published articles, peer review offers students the opportunity to analyze and assess established work, enhancing their understanding of how research fits into broader disciplinary conversations [21]. Introducing students to technical articles through active learning and peer review enables them to engage deeply with professional literature, a process shown to enhance critical thinking and innovation-based learning [22]-[25].

Course Design

In preparing students (particularly engineers and scientists, in these courses) for the future there are increasing demands on educators to teach writing, oral communication, critical thinking and problem-solving skills in addition to the discipline content. For these aims, two graduate level courses (PTX 7000 – Biostatistics and IHE 6150– Engineering Statistics) were designed at Wright State University to cover the extent of statistics at a practical level. The primary goal of both courses was shared: that students are not expected to become experts in statistics but become familiar with the practical use of statistical methods in designing experiments, collecting data, and analyzing results through appropriate methods. The two courses further share similar philosophy and goals in content with the exception that some methods received different emphasis depending on domain. The two courses did differ on some content, textbooks, some lectures/examples, and relevant homework problems.

Constraints on Course Design

The courses are both 3 semester hour credits with 14 weeks of classes. The courses were designed for either fully remote (pre-recorded lectures) or in-class meeting twice weekly for two 90-minute periods. Expectations are that students are familiar with Microsoft Word, Excel, and PowerPoint; students are not expected to have experience with statistical or programming software (this will be gained during the course). Both courses relied on JMP (SAS, Cary, NC) as statistical software, with expertise gained over the semester. JMP was selected due to a combination of industry demand, simplicity (and lack of coding), and accuracy of the software.

Textbook selection took multiple iterations and experimentations with course notes revised per textbook. In IHE classes textbooks by both Montgomery and Runger [26], and Navidi [27] were used in different classes. In the end, Montgomery and Runger was ultimately retained due to its comparatively easy to understand presentation of material. Similarly in PTX classes, books by Rosner [28], van Belle et al. [29], and Daniels and Cross [30] were used in different offerings. Daniels and Cross was ultimately selected as it was found to be written sufficiently well for a broad audience who have maybe never taken a stats class before.

Course Goal and Outcomes

PTX7000 and IHE6150 are graduate-level courses at Wright State University designed to provide practical knowledge of statistics in the health professions and engineering disciplines, respectively. Both courses emphasize the application of statistical methods such as hypothesis testing, regression, design of experiments, and ANOVA. While PTX7000 focuses on health data analysis, including survival analysis and epidemiology, IHE6150 incorporates statistical process control and problem-solving for engineering applications. Shared goals include equipping students with the ability to design experiments, analyze data, and draw valid conclusions using industry-relevant tools.

Course Design

The courses were designed with not only the integration of key statistical learning elements, but also with design project-based learning by focusing on real statistical challenges, enabling students to apply theoretical knowledge to hands-on data analysis and experimental design, as supported by recent work [31] [32]. Further, Harirforoush [33] emphasized that design projects not only deepen technical understanding but also cultivate creativity and collaboration—skills crucial for students addressing interdisciplinary problems. Further important skills to learn in this course involved literature reviews and finding out what has been done before so as not to waste time duplicating previous work. Such background development is typically done through library research and by keeping up-to-date with current publications. One of the most difficult aspects of developing this skill, however, is learning to critically evaluate this previous work. It is of little value to know what has been done before if the engineer cannot evaluate the quality or importance of the work.

Design projects in these courses were structured to guide students through every stage of statistical analysis, from formulating hypotheses and designing experiments to analyzing and interpreting results. These activities ensure that students not only learn to apply statistical methods but also develop the ability to critically evaluate the validity and significance of their findings. By engaging with peer-reviewed work and participating in peer evaluations, students gain exposure to professional standards of quality and rigor in research. A key outcome of this approach is fostering the skill of reviewing and critiquing technical literature, helping students avoid the pitfalls of duplicating prior work and instead contributing to innovative solutions [34] [35].

For these objectives, the following elements were incorporated beyond the typical statistical learning elements:

- Problem Solving – Typical statistical courses do not focus students on solving problems beyond textbook problems which are often designed to provide reasonable results. Problem solving with questions and assignments that need critical thinking are incorporated into these courses
- Peer review – in addition to leveraging lectures and book materials, peer review is used to gain proficiency in data organization and interpretation and view statistical tools effectively for problem-solving.

- Design project based learning – where students must demonstrate learned knowledge and skills through defining problems, designing experiments, collecting and analyzing data, and describing results.

For these general objectives, the classes incorporate specific learning objectives, including:

- Peer Review – employed to develop critical evaluation and constructive feedback skills and enhance communication and collaboration.
- Analysis of real world data – used to gain proficiency in data organization and interpretation and view statistical tools effectively for problem-solving.
- Design of Experiments – employed to apply principles of experimental design and formulate hypotheses and plan valid experiments.
- Analysis of Results – statistical tools in the course are used to interpret outputs and assess reliability of results. Further, the students are expected to communicate these findings clearly to diverse audiences (who have their own, and often very different, projects).
- Critical Analysis of Published Results – tools in the class are used to evaluate quality and validity of published studies and identify limitations.

Course Structure

The courses were designed synergistically as depicted in Figure 1. In Figure 1, the approximate week of instruction is presented at the left followed by the lecture content (separated into biostatistics and engineering statistics options) and associated homework and their topic areas. Notably, this is designed for a 14 week semester with no lecture content in week 14; this is due to planned student project presentations during this week. To the right of this are literature review assignments and final project assignments. The scheduling of the assignments is not happenstance, deconfliction of assignments is taken into account to avoid two being due at the same time. High level goals of these assignment categories are presented in Table 1, which illustrate the procedural knowledge gained by homework, the project learning enabled by the final project, and the problem solving and peer review experiences found in the homework and literature review topics.

Beginning with lecture material and topics aligned to homework, as depicted in Figure 1, it is important to note that key concepts from earlier topics continue to be of importance in subsequent topics. For instance, distributional knowledge is needed to understand residuals and when to use nonparametric methods. Notably, some areas also overlap; for instance, equivalence exists between hypothesis tests and analysis of variance (ANOVA) for some cases, this is explored in Homework 3 (HW 3 in Figure 1). Similarly, residuals are of importance for both ANOVA and linear regression as they inform the user of the quality of a model and thus homework 3 and 5 overlap in this area. For this aim, homework is used to develop understanding of core concepts, consistent with [36].

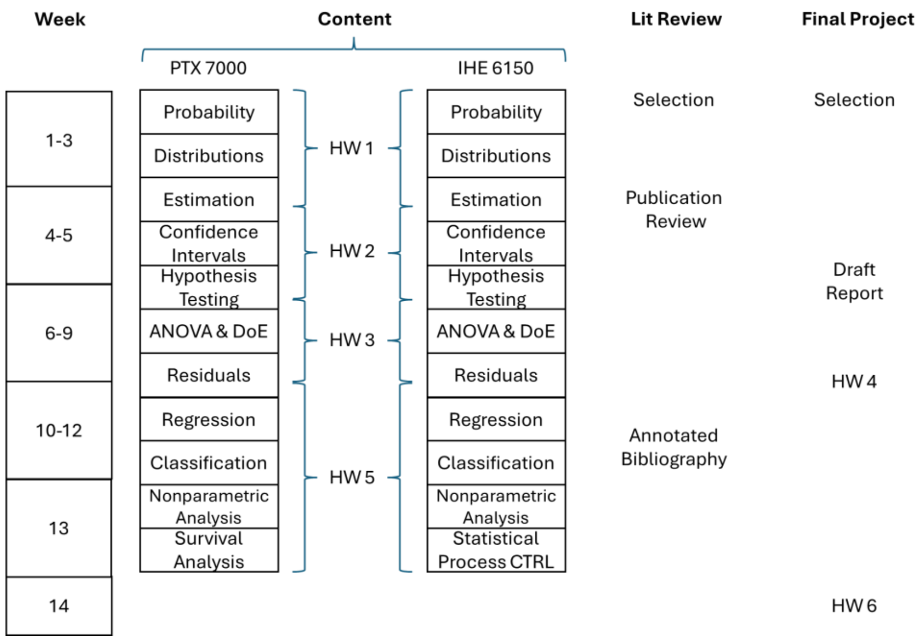


Figure 1. Course Topic Layout and Assignment Focus/Approximate Due Dates

Table 1. Pedagogical goals of gradable items

Modules	**Overview**	**Value (of course grade)**
Homework	Provide basic comprehensive understanding of statistics concepts	35%*
Statistical Paper Review	Perform peer review on selected published papers which employ statistical methods, evaluating papers on specific criteria	5%
Annotated Bibliography	Rigorously study literature in a specific statistical topic area	10%
Final Project Selection	Demonstrate exploration of possible research items with short paragraphs describing 5+ possible projects, statistical merit is not initial criteria, but developed through discussions and down selection to the most viable topic.	5%
Final Project Proposal	Develop a proposal which documents the literature on the single selected final project area, what has been done before, experimental plan, and possibly preliminary data. This is evaluated by the instructor and peers through blinded peer review	5%
Final Project Report	Final report including material from the *final project proposal*, peer and instructor feedback, and experimentation data and statistical analysis	20%
Final Project Presentation	Presentation of the results from *final project report*.	15%
Participation	Evaluate contributions students make through questions, email, dialogue and other useful interactions.	5%

*Any Phd Student will have homework worth 25% and have an in-class teaching expectation worth 10%

Problem Solving and Design Project Learning Modules

Problem solving and design project learning is primarily considered in both statistical courses in the form of two large assignments, the annotated bibliography and the final project. Both of these assignments consistent of an initial student-led instructor-approved selection of the topic. This was performed for quality control and to keep topics unique within a single course and avoid repetition across multiple semesters of courses.

Annotated Bibliography (Literature Depth)

A primary aspect of research is becoming familiar with prior works, thus this assignment serves two purposes: 1) to expose students to a wide variety of statistical publications and 2) facilitate general graduate skills in becoming familiar with research. The format was selected to focus on a broad understanding of a topic area with general summaries of each reviewed paper. However, a simple review is not sufficient and students are expected to include an introduction, background, discussion and synthesis of papers, conclusions, open questions/problems, bibliography of grouped papers and discussion on each paper. However, some freedom is given in format and the students are pointed to multiple examples, e.g. [37] [38].

Additionally, the primary focus of this report is on methods, not applications. For example, examining the literature on advances in regression (which includes variants such as total least squares) would be appropriate; but discussing papers which have boutique data applications but no discussion of the underlying methods would not. Students are not given direct numbers of references needed, as each domain is more or less mature, but are encouraged in dialogue on the appropriate scope for their topics. At an absolute minimum, one should expect 30 quality references and a solid reference selection scheme, such as the one presented in Figure 2.

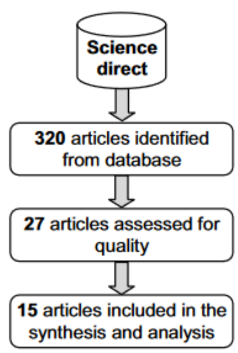


Figure 2. Example literature review selection process provided to students, from [39]

Final Project

A comprehensive analysis project constitutes an important element of the course. The purpose of this project is to collect, analyze, and provide insight into data using the tools from this class. This project includes discussing ideas with the professor, submitting a proposal, performing research, writing a final report and developing a final report presentation. The final project is divided into four parts: 1) project selection, 2) project proposal, 3) project report, and 4) project presentation. The project serves many purposes, in replacing a final exam the project demonstrates mastery of material through the appropriate use of statistical methods and interpreting their results. Beyond this, the students must further communicate these findings clearly to a diverse audience (who have their own, and often very different, projects).

Project selection involves the students choosing a topic; these are available first-come, first-served; however, the courses assume that the students will develop their own topics (with instructor supervision, not direction) and the students are not provided with a list of ideas. Students are encouraged to look to laboratory experiments in literature, or even science fair project ideas[1]. Given the possibility that many concepts are not practical, the students are expected to submit 5 candidate problem ideas and the feasibility of each (1 paragraph, 5 sentences (max) per problem) for the student and instructor to collaboratively select the move viable project. In general, the weighting is first on collecting data, second on aggregating data from multiple sources (or using unpublished laboratory data from a professor colleague), and third on using already available data (such as from Kaggle); the reasoning is that the first provides valuable experimental design and hypothesis generation experiences that are not found otherwise.

Project proposals are midterm checkpoints where both the professor and other students can provide feedback. At this point in the course, results are not expected, students should have an introduction and abstract to the problem, a basic literature review on similar studies, and an experimentation plan. The focus of this assignment is encouraging problem solving through developing an experimentation plan, formulation of hypotheses, and data collection expectations.

The final project report and presentation are used to evaluate comprehensive understanding of course material through 1) the usage of appropriate statistical methods to analyze the student collected data and 2) the appropriate inferences and conclusions drawn from the analysis. These are presented to the class as a mini-conference the last week of the semester, with each student having 10 minutes to present the problem, background, and results. Anecdotally, many students have expressed concerns with having only 10 minutes; which reinforces the need for brevity since many funding and workspace decisions are made based on short “elevator speeches” [40].

Report and Presentation Grading Approach

All reports are graded using a 100-point rubric (Table 2) based on scope and understanding, organization, coherence, application of methods, and conclusions. The same rubric is used for double-blinded peer reviews of student reports.

Table 2. Overview of report gradable items.

Subject	Weight
Original Work / Late / approved topic / approved format	Binary
Breadth and Depth (including scope, understanding of data, use of references, etc.)	30%
Organization (including format, readability, etc.)	10%
Coherence (including writing; e.g. is the paper understandable? do I understand what you did? do I think you understand what you did?)	20%
Application, Appropriate Use, Understanding of Methods	30%
Conclusions	10%

[1] https://www.lumiere-education.com/post/15-journals-to-publish-your-research-in-high-school

The final presentation is evaluated using the rubric in Table 3. These are given with less specificity to allow for some subjective interpretation and naturalness in presentation quality. Each of the 10 items is graded equally and the total is graded out of 100. While the instructor's evaluation is the most meaningful one for the grade, each student is also provided with an evaluation sheet for single-blinded presentation evaluations. These are also sometimes helpful to the instructor in reviewing before final grades.

Table 3. Overview of presentation gradable items.

<table>
<tr><td rowspan="5">Research and Slide quality metrics</td><td>Organization</td><td rowspan="5">Presentation Quality Metrics</td><td>Clarity</td></tr>
<tr><td>Scope</td><td>Interaction</td></tr>
<tr><td>Understandable</td><td>Pace</td></tr>
<tr><td>Application/Insights</td><td>Timing</td></tr>
<tr><td>Examples/Graphics</td><td>Style</td></tr>
</table>

Peer Review Modules

Peer review modules are provided in three different assignments: the publication review, homework 4 (proposal draft review), and homework 6 (final presentation evaluation). Each peer review module serves a different purpose and different approach.

Statistical Publication Review

The publication review serves the opposite purpose to the Annotated Bibliography. While the Annotated Bibliography demonstrates a high-level understanding of a large number of papers within a topic area, the publication review serves as a deep and forensic dive into a smaller set of papers. Many published papers using statistical methods (ranging from simple regressions to complex neural networks) are often not repeatable due to insufficient details being presented [41] [42]. That this is common in published papers indicates that peer review doesn't find all of these issues. The focus of this assignment is to rigorously evaluate instructor selected publications against specific algorithmic and data repeatability questions. These questions are found in Table 4.

Homework 4 – Proposal Draft Peer Review

Homework 4 introduces students to the critical skill of peer review through a double-blinded evaluation of proposal drafts. Each student assesses two anonymized reports based on defined criteria, such as clarity, coherence, and application of statistical methods; per the same criteria the instructors use in Table 2. This assignment serves multiple purposes: students gain insights into the evaluation process, improve their ability to critique technical work, and receive diverse feedback on their own drafts. The peer review process fosters collaborative learning and helps students refine their proposals, ultimately enhancing the quality of their final projects.

Homework 6 – Final Presentation Evaluation

Homework 6 extends peer review into the realm of presentation skills, requiring students to evaluate the final project presentations of their peers. Students accomplish this by providing single-blinded feedback per Table 3, which is then shared with the students at the end of the course. By providing structured feedback on elements like clarity, pacing, and use of visuals, students learn to assess technical communication critically. This assignment not only engages students in active participation during presentations but also encourages them to reflect on their own presentation styles. The structured evaluation criteria ensure that feedback is constructive and actionable, further contributing to the improvement of their communication skills.

Table 4. Statistical Publication Review Peer Review Questions

You will assess these qualities on the following scale:
1 = none, 2=insufficient, 3= medium,4 = sufficient 5 = excellent, 6 = N/A
Descriptions for these attributes are as follows (*full credit for those who evaluate all with number rating and describe reasonings*):
Algorithm Details – do the authors name the machine learning method(s) used? Do they cite a quality paper for these method(s)? Do they discuss algorithmic settings?
Example 1:"Linear discriminant analysis" has no algorithmic settings and means a specific function
Example 2: "discriminant analysis" is unclear (i.e. there are many discriminant variants such as linear and quadratic)
Example 3: Artificial neural networks have many settings (number of nodes, number of layers, types of nodes, training methods, architecture variant). All of these must be specified for repeatability
Data Details – do the authors describe the source of the data or the collection means? Do they cite a source? Do they describe all data variables?
Performance Result Sufficiency – Do the authors present accuracy values? Do they compare results to different methods? Are error bars or confidence intervals used to discuss results (or do they include standard error values)?
Training/Testing Method – Do the authors divide their data into a training set and a testing set? Do they describe how the data was divided?
Graphical Excellence – Do the authors use effective graphics in their paper to present results? Are the figures readable? Are colors in the figure easy to distinguish?
Overall Statistical Method Quality Score – How do you personally combine the 5 assessments you performed and grade this paper on the 1 to 6 scale?
Statistical Methods Used in Paper: Additionally, list all stated (or suspected) machine learning methods used in the paper

Problem Solving Modules and Critical Analysis of Published Results

The following critical thinking problems are designed to help students develop a deeper understanding of data analysis, statistical reasoning, and the importance of selecting appropriate methods for interpreting results. These exercises highlight real-world issues such as software precision errors, outlier interpretation, confidence interval analysis, and comparing categorical data. Each problem encourages students to critically evaluate published results, challenge assumptions, and apply correct statistical methods when analyzing data. Through these examples, students not only improve their statistical skills but also learn to question the validity of results, fostering a mindset of evidence-based reasoning and careful data interpretation.

Precision and Software (Homework 1)

The first question involves analyzing two files of data, each with 500 entries of 1.1, 500 of 1.3, and 1 of 1.2. File 1 uses 1 significant digit, while File 2 extends to 16 digits (e.g., 1.1000000000000000). They calculate the mean by a) estimating it, b) using Excel's AVERAGE function, and c) using statistical software (JMP), then interpret the results. Excel often gives inconsistent results due to known statistical assumptions) [43] [44]. This exercise highlights the importance of understanding tools before trusting them and creates teachable moments about precision and software differences.

A	B	C	D
1.2	Mean	1.2	
1.1			
1.1			
1.1			
1.1			
1.1			

A	B	C	D
1.2000000000000000	Mean	1.1999999999999900	
1.1000000000000000			
1.3000000000000000			
1.1000000000000000			
1.3000000000000000			

Figure 3. Example results from MS Excel's Average. Left: 1 significant digit and correct answer; right: 16 significant digits and incorrect results

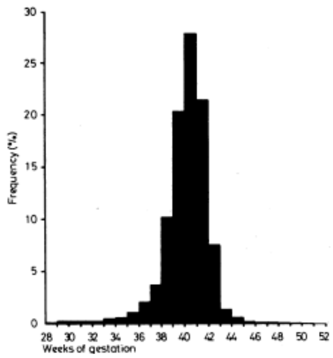
Outliers and the British Divorce Case (Homework 1)

Another critical thinking question is often included in homework 1, adapted from the example of [45] [46]. The general problem is presented in left panel of Figure 4. Here, the concern is on outliers and evaluating data and assertions of others: a husband wants a divorce because his wife gave birth 50 weeks after they last saw each other, considerably far from the typical 37-42 human gestation period. In this problem the students must interpret the data, decisions, and draw inferences with meaningful conclusions. Many students disagree with the judge's verdict due to the extreme outlier this pregnancy would have been (possibly record setting); however, some students defer to the judge's opinion and this provides a further opportunity to explain outliers with a concrete example.

Confidence Intervals for Comparison of Published Results (Homework 3)

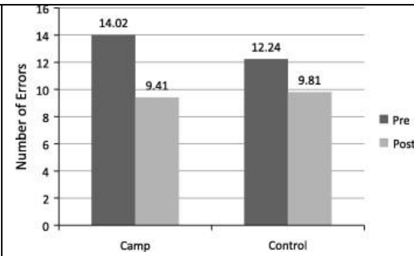
A critical thinking question in homework focuses on evaluating the results of a publication without having access to the original data. Here, Figure 4 middle, some results from [47] are explored; this paper presented pre- and post-test scores for 2 groups (camp and control) and stated that the camp group improved significantly over the control group. However, the results were presented without a confidence interval and the students are asked to critically analyze the histogram data using confidence intervals to assess whether the claimed treatment effect is valid. This question encourages deeper evaluation of experimental design, statistical rigor, and the reliability of the conclusions/publications.

Since the students do not have access to the original data, appropriate statistical methods are needed. Students usually correctly opt for proportional confidence intervals and determine that the results in the paper are possibly invalid. Generally valid answers include: A) Contrary to the paper, there are no apparent differences between groups since all confidence intervals overlap, or B) There is no difference between groups using proportional confidence intervals, but there could be a difference since these intervals are likely wider than those from original data.



The setting: A British divorce case was heard in court. The husband was seeking a divorce from his wife based on his wife giving birth to a baby 50 weeks after the husband left for military service in WWII with no contact with his wife. The husband had no physical contact with his wife during these 50 weeks, and therefore adultery was suspected (adultery was a legal requirement divorce in Britain at the time-1949). While the average human gestation period is 40 weeks (typical range: 37-42 weeks), the judge ruled that since a 349 day pregnancy was scientifically possible, although highly improbable, that the divorce would not be granted.

A) Evaluate and comment on the reasoning for the judge's verdict. Use appropriate statistical methods to justify your conclusion.
B) If you were an expert witness in this case, how and where would you draw the line?



Consider this histogram from a published study [47]. Here two groups of children were studied with the camp group, deprived of electronics, attending a weekend retreat, while the control group did not. Both groups took the DANVA2 test before and after. DANVA2 presents 48 face photos (happy, sad, angry, fearful) for 2 seconds each, and participants identify the emotion. Scoring is based on errors out of 48. Uhls et al. found significant improvement in the camp group, indicating the treatment's effectiveness.

A) Use confidence intervals of a proportion to critically examine their results. Considering a significance level of 5%, evaluate the conclusions of Uhls et al. Assume that the mean number of error is out of 48 total.
B) Do you believe that there is difference between the pre and post test results?

Consider the following results presented in a journal [47]. Assume a 5% significance whenever needed. This study examines non-verbal cue recognition using a test.

Table: Ethnic Makeup of Camp and Control Groups

	Camp	Control
Sample Size	51	54
White	26 (51%)	11 (20%)
Hispanic	9 (18%)	9 (17%)
Asian	9 (17%)	19 (35%)
African American	1 (2%)	1 (2%)
Other/mixed	6 (12%)	14 (26%)

A) Uhls et al. [47] used a t-test (unstated which type, but assumed to paired or pooled) to determine the groups were different with the results of "was a difference in ethnic composition ($t_{(105)} = -2.95$, $P < 0.01$)."
B) Is this test an appropriate method to compare these groups?
C) Did [47] perform the t-test appropriately?
D) Do you agree with their approach and result?
E) Using the data table, verify if these groups are similar or not

Figure 4. Example critical thinking questions. Left: Outliers in the British divorce case, middle: Uhls and confidence intervals, right: Uhls and incorrect statistical tests

Comparing Categorical Groups (Homework 3)

A further critical thinking question was developed, again considering the result of [47], but on comparing two groups based on ethnic categories. This is presented in the right side of Figure 4. Here, a table of ethnicities for the two groups considered in [47] is presented to the students with multiple questions. According to [47], a t-test was conducted and it was determined that there was a significant difference in ethnic composition between the groups.

The students are asked to critically analyze whether a t-test is appropriate for comparing categorical data, to evaluate the validity of the reported results, and to redo/verify the analysis. Since the data consists of proportions, students are encouraged to apply alternative statistical methods, such as chi-square or proportion tests, to verify if the groups are truly different. Most students correctly identify that a t-test (as used in [47]) is unsuitable for categorical data and then redo the analysis using an appropriate test, such as chi-squared. A few students even notice that the t-test has an incorrect number of degrees of freedom.

Impact and Results

Ideally, an evaluation would consider results and impacts using the developed method and an alternative teaching method through possibly a pre- and post-test of desired learning goals. An alternative would be a pre- and post-test within the course itself with comparison to other published results from statistical courses. However, it is intrinsically hard to perform such studies due to scale, timespans, and other practicalities. Thus, alternative impact measures will be used of a) grade distribution before/after changes, b) tangible course outcomes, and c) intangible course outcomes.

Table 5 presents the grade distribution from before changes were implemented (Spring 2015 for PTX and Fall 2017 for IHE) from a total of 324 students since Spring 2014. Following curriculum changes, the grade distribution improved, with the percentage of A grades increasing slightly from 55.5% to 56.6%, while B grades decreased from 33.6% to 30.7%. Notably, D grades were significantly reduced from 4.2% to 0.5%, suggesting an improvement in student comprehension and retention. A chi-square test yielded a p-value of 0.08, indicating that the change is statistically significant at the 10% level (but not at the 5% level), with the results overall suggesting a positive trend in student outcomes.

Table 5. Grade distributions before and after changes by percentage (number of students)

	Before Implementation	**After Implementation**
A	55.5% (66)	56.6% (116)
B	33.6% (40)	30.7% (63)
C	5.0% (6)	9.7% (20)
D	4.2% (5)	0.5% (1)
F	1.7% (2)	2.4% (5)

Table 6. Example Publications from Projects in IHE 6150 and PTX 7000

Publication	Project Name / Publication
[48]	*Epidemiological Study of Risk Factors for COVID-19 in Ohio*
[49]	*Annotated bibliography on Reinforcement learning and its clinical applications in medicine*
[50]	*Testing of Streptococcal Pharyngitis in Children Less than Three Years Old*
[51]	*Physician variation in quantifying actinic keratosis skin photodamage*
[52]	*Trends in Advanced Tabletop Digital DJ Media Players from 2000-2023*
[53]	*Analytics for non-invasive skin imaging and actinic damage*

Another tangible outcome are publications resulting from course projects. While this is not the primary goal of the final project, novel research and publications are possible. Additionally, the benefit of this metric is that it independently confirms the value of content demonstrated in the course through peer review. Example publications and associated projects are presented in Table 6 as an example of what is possible. Intangible course outcomes include statements from PTX students that the new course was the “most valuable course they took” in their program.

Conclusions

The integration of design project-based learning, problem-solving, and peer review into graduate-level statistics education at Wright State University has provided a robust framework for addressing traditional challenges in the field. Through these methods, students not only gained technical proficiency in statistical tools and methods but also developed critical thinking, collaboration, and communication skills. Evaluation of these approaches showed promising outcomes, with increased student engagement and improved understanding of course material, as evidenced by the quality of final projects and positive peer feedback during blinded reviews. Peer review, in particular, emerged as a valuable tool, fostering constructive critique and reflection that enhanced both individual and collective learning.

The implementation of real-world, discipline-specific projects further demonstrated the relevance of statistical methods in diverse fields such as health sciences and engineering. These projects encouraged students to apply their knowledge to meaningful problems, including experimental design and data analysis, resulting in practical, actionable insights. While further quantitative assessment is needed, anecdotal evidence suggests that these methods have significantly improved student confidence and interest in statistical applications.

This approach offers a scalable and replicable model for other institutions, emphasizing the importance of combining technical training with problem-solving and evaluative skills. Future work will explore expanded evaluation metrics, including student outcomes in subsequent courses or professional contexts, to further validate and refine the impact of these innovations on statistics education.

Bibliography

[1] A. McAfee, E. Brynjolfsson, T. Davenport, D. Patil and D. Barton, "Big data: the management revolution," *Harvard business review,* vol. 90, no. 10, pp. 60-68, 2012.

[2] R. Butler, "On the failure of the widespread use of statistics.," *Amstat News,* p. 84, 1998.

[3] G. Van Belle, Statistical rules of thumb, John Wiley & Sons., 2011.

[4] D. Montgomery and G. Runger, Applied statistics and probability for engineers, John wiley & sons., 2010.

[5] C. Schau, J. Stevens, T. Dauphinee and A. Vecchio, "The development and validation of the survey of antitudes toward statistics," *Educational and psychological measurement,* vol. 55, no. 5, pp. 868-875, 1995.

[6] C. Schau, "Students’ attitudes: The “other” important outcome in statistics education," *Proceedings of the joint statistical meetings,* pp. 3673-3681, 2003.

[7] P. Petocz and A. Reid, "Relationship between students' experience of learning statistics and teaching statistics," *Statistics Education Research Journal,* vol. 2, no. 1, pp. 39-53, 2003.

[8] S. Bell, "Project-Based Learning for the 21st Century: Skills for the Future," *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas,* vol. 83, no. 2, p. 39–43, 2010.

[9] A. Shekar, "Project-Based Learning in Engineering Design Education: Sharing Best Practices," *121st ASEE Annual Conference and Exposition,* 2014.

[10] M. Frank, I. Lavy and D. Elata, "Implementing the Project-Based Learning Approach in an Academic Engineering Course," *International Journal of Technology and Design Education,* vol. 13, p. 273–88, 2003.

[11] S. Hamilton, "Learning Through Peer Review Of Work," *ASEE Annual Conference,* 2005.

[12] J. Sharp and M. Dyrud, "Two Perspectives On Peer Review," *ASEE Annual Conference & Exposition*, 2009.

[13] L. Benson, R. A. Bates, K. Jensen, G. Lichtenstein, K. Watts, E. Ko and B. Albayati, "Building Research Skills through Being a Peer Reviewer," *ASEE Virtual Annual Conference Content Access,* 2021.

[14] S. Hamilton, L. Brunell, G. Tamm and O. Arnas, "Peer Review In Engineering Courses As A Learning Tool," *Annual Conference & Exposition,* 2006.

[15] D. Ludlow, "Using Critical Evaluation And Peer Review Writing Assignments In A Chemical Process Safety Course," *ASEE Annual Conference,* 2001.

[16] A. Rose, "Using The Peer Review Process To Implement Writing Assignments In An Engineering Technology Course," *ASEE Annual Conference,* 2001.

[17] T. Larkin, "Peer Review From A Student Perspective," *ASEE Annual Conference & Exposition,* 2009.

[18] S. I. Ringleb, O. M. Ayala and J. Kidd, "Implementing Peer-Review Activities for Engineering Writing Assignments," *ASEE Annual Conference & Exposition,* 2017.

[19] N. Smith, "Guided Peer Review of Technical Writing for Large Laboratory Course," *ASEE Annual Conference & Exposition,* 2019.

[20] J. Lax and A. W. Fentiman, "Improving Graduate Student Oral Presentations Through Peer Review," *ASEE Annual Conference & Exposition,* 2016.

[21] A. Clobes and W. H. Guilford, "Improved retention and recall with a peer reviewed writing assignment," *ASEE Annual Conference & Exposition*, 2013.

[22] P. Hill, "Introducing Technical Articles to Graduate Students Through Active Learning," *Southeast Regional ASEE meeting,* 2005.

[23] P. Hill and D. Swalm, "Using Technical Articles to Teach Entering Graduate Students the Role of Journal Articles in Research," *American Institute of Chemical Engineers,* 2005.

[24] R. Striker, M. Pearson, E. M. Swartz, E. A. Vazquez, L. Singelman and N. S. S., "Anonymous Online Peer Review for Innovation-Based Learning," *ASEE Virtual Annual Conference Content Access,* 2021.

[25] V. Svihla, C. Hubka and E. Chi, "Peer review and reflection in engineering labs: Writing to learn and learning to write," *ASEE 125th annual conference and exhibition,* 2018.

[26] D. Montgomery and G. Runger, Applied statistics and probability for engineers., John wiley & sons., 2010.

[27] W. Navidi, Statistics for engineers and scientists, New York: McGraw-Hill, 2006.

[28] B. Rosner, Fundamentals of Biostatistics (The 7th edition)., Boston, MA: Brooks/Cole., 2011.

[29] G. Van Belle, L. Fisher, P. Heagerty and T. Lumley, Biostatistics: A Methodology for the Health Sciences., John Wiley & Sons Inc., 2004.

[30] W. Daniel and C. Cross, Biostatistics: a foundation for analysis in the health sciences, Wiley, 2018.

[31] B. Ansaf, "Teaching an undergraduate manufacturing course using a design-based teaching approach," *126th ASEE Annual conference and exposition,* 2019.

[32] L. Wei, Z. Wei and L. Chenhua, "The Study of the Effectiveness of Design-Based Engineering Learning: The Mediating Role of Cognitive Engagement and the Moderating Role of Modes of Engagement," *Frontiers in Psychology,* vol. 14, 2023.

[33] R. Harirforoush, "The benefits of design-project learning approach in an engineering course.," *2024 ASEE North Central Section Conference,* pp. 1-8, 2024.

[34] S. Turner, M. Pérez-Quiñones and S. Edwards, "Peer review in CS2: Conceptual learning and high-level thinking," *ACM Transactions on Computing Education (TOCE),* vol. 18, no. 3, pp. 1-37, 2018.

[35] S. Turner and M. Pérez-Quiñones, "Exploring peer review in the computer science classroom.," *arXiv preprint arXiv:0907.3456.,* 2009.

[36] P. Wankat, "The role of homework," *2001 American Society for Engineering Education Annual Conference & Exposition,* pp. 1-7, 2001.

[37] G. Di Battista, P. Eades, R. Tamassia and I. Tollis, "Algorithms for drawing graphs: an annotated bibliography," *Computational Geometry,* vol. 4, no. 5, pp. 235-282, 1994.

[38] R. Clemen, "Combining forecasts: A review and annotated bibliography," *International journal of forecasting,* vol. 5, no. 4, pp. 559-583, 1989.

[39] M. Feki, I. Boughzala and S. Wamba, "Big data analytics-enabled supply chain transformation: A literature review," *2016 49th Hawaii International Conference on System Sciences (HICSS),* pp. 1123-1132, 2016.

[40] A. Cox and L. Marris, "Introducing elevator speeches into the curriculum," *Journal of Education for Library and Information Science,* pp. 133-141, 2011.

[41] G. Zhang, "Avoiding pitfalls in neural network research.," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* vol. 37, no. 1, pp. 3-16, 2006.

[42] T. Bihl, J. Schoenbeck, D. Steeneck and J. Jordan, "Easy and efficient hyperparameter optimization to address some artificial intelligence “ilities”.," *Hawaii International Conference on System Scienes,* 2020.

[43] A. Yalta, "The accuracy of statistical distributions in Microsoft® Excel 2007.," *Computational Statistics & Data Analysis,* vol. 52, no. 10, pp. 4579-4586, 2008.

[44] B. McCullough and D. Heiser, "On the accuracy of statistical procedures in Microsoft Excel 2007.," *Computational Statistics & Data Analysis,* vol. 52, no. 10, pp. 4570-4578, 2008.

[45] V. Barnett, "The study of outliers: purpose and model," *Appied Statistics,* vol. 27, pp. 242-50, 1978.

[46] D. Altman, "Statistics and ethics in medical research. Misuse of statistics is unethical.," *British Medical Journal,* vol. 281, no. 6249, pp. 1182-1184, 1980.

[47] Y. M. M. Uhls, J. Morris, D. Garcia, G. Small, E. Zgourou and P. Greenfield, "Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues," *Computers in Human Behavior,* vol. 39, pp. 387-392, 2014.

[48] Y. Park, J. Travers and T. J. Bihl, "Risk Factors for COVID-19 in Ohio: A Retrospective Study at the County Level," *Marshall Journal of Medicine,* submitted: June. 2024.

[49] M. Gilbert, T. Frommeyer, R. Fursmith, G. Brttain and T. J. Bihl, "Reinforcement learning and its clinical applications in medicine," *Frontiers in Digital Health,* in revision: 2024.

[50] S. Bayless, T. J. Bihl, C. Rohan, J. Travers and E. Whitney, "Inappropriate Testing of Streptococcal Pharyngitis in Children Less than Three Years Old: Application of Statistical Process Control," *Clinical Pediatrics,* vol. 62, no. 4, pp. 309-315, 2023.

[51] B. Schmeusser, C. Borchers, J. Travers, S. Borchers, J. Trevino, M. Rubin, H. Donnelly, K. Kellawan, L. Carpenter, S. Bahl, C. Rohan, E. Mueenich, S. Guenthner, H. Hahn, A. Rkein, M. Darst, N. Mousdicas, E. Cates, U. Sunar and T. Bihl, "Inter-and Intra-physician variation in quantifying actinic keratosis skin photodamage," *Journal of Clinical and Investigative Dermatology,* vol. 8, no. 2, 2020.

[52] B. Goulart, T. J. Bihl and J. Travers, "Trends in Advanced Tabletop Digital DJ Media Players from 2000-2023," *IEEE NAECON,* 2024.

[53] J. B. Travers, C. Poon, T. J. Bihl, C. Borchers, D. J. Rohrbach, S. Borchers, J. Trevino, M. Rubin, H. Donnelly, K. Kellawan, L. Carpenter, S. Bahl, C. Rohan, E. Muennich, S. Guenthner, H. Hahn, A. Rkein, M. Darst, N. Mousdicas and Cates, "Quantifying skin photodamage with spatial frequency domain imaging: Statistical results," *Biomedical Optics Express,* vol. 10, no. 9, pp. 4676-83, 2019.