JUSTIFEST: A BLENDED LEARNING TOOL FOR STATISTICAL ANALYSIS PLANS

Frederik De Keersmaeker[1], Eva Vangenechten[1], Erik Fransen[2], and Els Goetghebeur[1]
[1]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
[2]Department of Biomedical Sciences, University of Antwerp, Belgium
frederik.dekeersmaeker@ugent.be

Statistics classes in higher education often focus on specific analyses, assuming a research question was previously formulated and an analysis plan for the available data was agreed on. These early stages of the investigative process are, however, critical steps where important choices are made and where statistical thinking must lead to a justified approach. We built a blended learning tool—Justifest—that stimulates students to practice planning skills, facing the consequences of various choices. Starting from a simple, small, but relevant case study, they come up with a plan to investigate whether university grades have changed during the coronavirus pandemic. We help to improve their initial analysis plan by asking critical questions at the conceptual level and revealing frequentist properties through plasmode simulation.

BACKGROUND

The Importance of Statistical Thinking

The importance of fostering statistical thinking in science and hence in statistics education at all levels has been well established (see, e.g., Cox & Efron, 2015; Pfannkuch, 2018; and references therein). Wild and Pfannkuch (1999) identify four dimensions of statistical thinking, the first of which is the investigative cycle described by the PPDAC (Problem, Plan, Data, Analysis, Conclusion) cycle. In their attempt to make the intuitive but vague notion of statistical thinking more concrete, they report that statisticians they interviewed "were particularly interested in giving prominence to the early stages of PPDAC" (p. 225).

The Lack of Attention for the Investigative Process in Higher Education

Even though any statistical inquiry first requires formulating a precise research question and a statistical analysis plan, most standard statistics textbooks skip over it to focus on the analysis and conclusion steps. In the seventeen chapters of *Introduction to the Practice of Statistics*, Moore, McCabe, and Craig (2021) discuss topics ranging from *t*-tests for comparing two means to logistic regression and non-parametric tests. However, there is no chapter that teaches readers how to deal with this overwhelming choice of techniques, i.e., how to weigh the advantages and disadvantages of several approaches in the likely situation where more than one is appropriate. Different methods are presented in a highly segmented way. Traditional approaches rarely use real data, and, consequently, avoid many questions that lead to an analysis of choice. Instead, *textbook examples* commonly satisfy all necessary assumptions for a particular analysis, or, at best, a series of examples is presented in which exactly one assumption of the approach is violated. It makes students ill prepared to tackle real-world problems independently. Students have no option but to learn these essential skills and attitudes by experience post-graduation.

Higher education students majoring in statistics or data science tend to have project work in specific courses and eventually present a thesis with more elaborate applied work. Even they often miss systematic training in how to apply what they learn in practical settings with various objectives and underlying data structures. They are nevertheless supposed to have mastered statistical thinking by the time they graduate. Adopting the mindset of problem solving under uncertainty as early as possible critically contributes to their success in more advanced statistics classes. In an introductory course on *Principles of Statistical Data Analysis* (*PrinStat*), we introduce master's students with different background training to the investigative process and encourage critical thinking, among other topics. In our experience, the traditional approach accomplishes the learning goals that are lower on Bloom's (1956) taxonomy: students remember and understand material on basic probability, confidence intervals, and hypothesis testing (*knowledge* and *comprehension*); they can use these techniques to solve straightforward problems (*application*). Weighing the potential of different methods and questioning assumptions (*analysis*) before engaging in a specific analysis that is appropriate and justified, is more challenging. It requires sufficiently broad knowledge of candidate methods as well as

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

practice. In general, students need to build confidence and skills to develop a plan from scratch, considering important and different options, selecting the most appropriate one, and defending their choice (*synthesis* and *evaluation*). This is all the more relevant because a pre-specified statistical analysis plan—with the option to modify it later—is seen as a critical component of reproducible research.

Students learn by doing, so projects are important learning opportunities. To give students the opportunity to experience the investigative process from start to end, we offer a *Statistical Consulting* course in the second semester of our program. Final competences include:

1. Structure a client project in logically ordered identified components: the primary and secondary objectives and research question(s) in context, the database with its sampling design and possible limitations, simple and/or complex analysis methods that may answer the formalized question(s), priorities and constraints of the project (client), existing literature on the topic, the audience to report results to, the decisions that may be based on the statistical analysis results.
2. Recognize statistical issues in different phases, including research question formulation, study design, and sampling procedure.
3. Identify and motivate an appropriate data analysis technique to answer the client's research question in the available time.

Blended Learning Tools

We developed a blended learning tool called *Justifest*, short for Justified estimation, which we describe below. In our curriculum, we intend to use it in *Principles of Statistical Data Analysis,* and in *Statistical Consulting* at the start of the semester as a review of group comparisons and the investigative process. There exist other blended learning tools that responded to the need to develop statistical thinking in context, relying on simulated data. To mention a few:

1. Steiner and Mackay (2009) provided students with simulated data in a virtual environment and asked them to use the PPDAC cycle to reduce variation by implementing one of several variation reduction frameworks introduced in their course.
2. Darius, Portier, and Schrevens (2007) built a tool to allow students in experimental design courses to gain experience using simulations.
3. Bulmer and Haladyn (2011) and Baglin, Bedford, and Bulmer (2013) described The Island, a tool that gave students access to (epidemiological and demographic) data from a simulated population. Students could ask their own research question and design and run a study (including data collection) in this environment.

Tools 2 and 3 are used in advanced statistics classes in which the entire investigative process is the central object of study (namely experimental design and epidemiology). Ideally, statistics students are exposed to these experiences as early as possible. Justifest is therefore tailored to more elementary subject matter, i.e., group comparisons.

We can see several advantages that simulated data bring. If we expect students to learn how to plan larger studies addressing more complex questions, we see an additional need to start with simple—but real context—assignments, and work with real data.

Finally, all three tools are built to allow students to experiment in order to get a better grasp on specific concepts covered in the classes for which they are designed. Justifest targets the comparison of different analysis methods from an introductory class by running through an investigation and focusing on the strengths and weaknesses various options bring. It therefore offers the opportunity to simultaneously learn and apply what it aims to contribute, i.e., a critical attitude when developing a statistical analysis plan sparked by critical questions (see below).

Besides comparing across various blended learning tools, we see our tool as complementary to and supportive of traditional project work with a single evaluation at the end of term. Justifest allows students to experiment and can give ample feedback throughout the term if and when the student feels ready to apply a new approach without overburdening educational staff. It supports growth by inviting students to apply the lessons they learned, while the subject of learning is the investigative process itself. It stimulates this by asking students critical questions that foster statistical thinking, followed by a demonstration of the consequences of various choices (through simulation in an R shiny app) and a discussion of best practices, so that students can implement and learn (see below) statistics content.

A BLENDED LEARNING TOOL FOR GROUP COMPARISONS

After a short introduction to Justifest below, subsequent subsections zoom in on specific parts or issues of our tool.

Short Description of Justifest

Justifest is a learnr module (https://rstudio.github.io/learnr/) that will be available online and is structured like a consulting case. After an introduction where we explain the problem and an opportunity to acquire relevant background knowledge, students are asked to write down a relevant and precise research question and to come up with an initial analysis plan. The tool then asks them critical questions, their answers to which will invite them to adapt their analysis plan. Armed with their final analysis plan, they are given access to the data and asked to perform their analysis. Students then write down their conclusion and reflect on the lessons learned. Finally, we provide them with an overview of the results of different relevant approaches.

For some of these critical questions, students are invited to run plasmode simulations to discover frequentist properties of estimators or hypothesis tests, either by using an R Shiny app we provide, or by writing their own R code within the learnr module.

Learning Outcomes

Students should be able to distill a precise research question from the research objective and the context. They should subsequently be able to come up with a diverse list of possible approaches, to choose one of them, and to argue why their approach is the most appropriate one.

Challenges in Using Real Data

Gaining familiarity with the investigative process requires a significant investment of time and effort on the part of students and educators because it unavoidably includes seeing a project through from start to finish. To increase the likelihood of learning, it helps that students have ownership of their projects and are personally affected by or interested in the subject. For this reason, we have chosen to let students explore the difference in grade distributions before and during the coronavirus pandemic for a specific course. The possible effect of the pandemic on grades impacts students directly and has been widely debated in educational research. On the one hand, there is consensus that it has negatively impacted learning outcomes (see, e.g., Farnell et al., 2021). On the other hand, because educators considered the pandemic as a mitigating factor when assigning grades, it may have led to inflated grades (see, e.g., Karadag, 2021). Universities are sitting on a trove of data to assess this question. We give students access to grades from a specific (but anonymized) course over three years (one pre-pandemic year and two pandemic years).

The first steps of the investigative cycle, *problem* and *plan*, require expert input. Ideally, a study about the impact of the coronavirus pandemic on grades in higher education would at the very least involve a specialist who knows how the grades are typically distributed and what the magnitude is of a meaningful difference on a specific scale. Students who are not domain experts, however, will typically need to consult published resources or explore historical data to attain some level of familiarity and expertise with the anticipated data distribution. In strict terms, the practice of examining the data to be analyzed before writing a protocol is considered *data snooping*, the dangers of which are exactly what we want to educate students on. To reconcile these conflicting goals, we first offer the students data from two years: one pre-pandemic year to act as a benchmark and one pandemic year. They are encouraged to explore these data before developing their analysis plan. Afterwards, we offer them data from a second pandemic year which they are to compare in a well-chosen meaningful way with the benchmark.

Structure of Justifest

Justifest includes the following sections.

- An introduction offering extended context, a description of the data source and type of available data, and the research objective.
- A background knowledge section where students can explore benchmark data and data from one pandemic year. They are asked critical questions to stimulate more in-depth thinking about the setting, how the data are sampled and distributed, and to encourage them to come up with and

compute relevant summary statistics, e.g., mean, median, pass-fail rate, 95% percentile, etc. In a next iteration of Justifest (with a different problem, context, and data), we will consider other ways to provide the necessary background knowledge: prior publications, an existing data source that contains pertinent available information, other useful resources, or the student's own experience and plans.

- A research question section where they use the domain knowledge from the previous step to translate the research objective into a concrete research question.
- Several sections covering the analysis plan, augmented with plasmode simulations revealing a method's statistical properties in the current data setting (Morris et al., 2019), resulting in a final plan that is well thought through. We discuss this in more detail below.
- An analysis section where they perform their analysis using previously unaccessed data.
- A conclusion section where they interpret the results of their analysis and form a conclusion.
- A section where they reflect on lessons learned and propose an improved study design and analysis plan for the next study on the same topic.
- A feedback section where they get an overview of results from different possible approaches to answer the question they have asked, together with pros and cons.

Guiding Students Towards an Analysis Plan by Having Them Answer Critical Questions

In the analysis plan section, students are asked to come up with an initial plan, which they will gradually adapt and improve. Changes to their plan fall into one of two categories. On the one hand, they may change their approach after learning about a method they had not previously considered, or after more carefully considering the advantages, disadvantages, underlying assumptions, etc. of their method or an alternative (e.g., using a permutation test instead of a *t*-test). On the other hand, they may state more clearly an aspect relevant to their approach that they had previously left unspecified (e.g., specifying whether they are targeting evidence of equivalence or superiority).

Students are prompted to examine and revise their approach by answering critical questions that we ask them in the tool, e.g., how will you determine the distribution of the test statistic under the null hypothesis? Every question is split into three parts. The first part offers background to make sure students understand the relevant concepts, e.g., an explanation of the difference between superiority, non-inferiority, and equivalence testing. The second part is the actual question. Finally, the third part is a *discussion* about the analysis approach—a term we use instead of *answer* because there is not necessarily a single clear-cut right or wrong answer in this context.

Alignment with Learning Theories

Students must play an active role when they learn about the investigative process. Classical approaches from *behavioral learning theory*, e.g., rote memorization or learning by observing a teacher solve problems, fall woefully short of what is required. Statistical thinking requires a collection of skills and attitudes on top of a body of knowledge. Therefore, it cannot be distilled into a list of rules or guidelines that can easily be transferred to students.

By using Justifest, students discover their own thought process and make new associations building on their prior knowledge. Awareness of and a critical attitude towards their own internal mental process is exactly what they are supposed to learn. The tool gives them the opportunity to tailor their learning path to their individual needs, i.e., to build on their particular set of skills and knowledge. The teaching philosophy behind our tool therefore aligns with *cognitive learning theory*, especially with *constructivist learning theory* (see, e.g., Pritchard, 2017).

Moreover, Justifest fits well within the framework of two more modern learning theories: transformative learning and experiential learning. *Transformative learning theory* was spearheaded by Mezirow (1978a, 1978b) in the 1970s as a better way to describe how adults learn. More so than children, adults rely on their large arsenal of past experiences and their current understanding, which they transform into new insights through critical reflection and review. Teachers offer opportunities for learning when they question assumptions and encourage discussion so students can discover new points of view. The structure of our tool, i.e., having students write down a statistical analysis plan using the knowledge and skills they already have, and subsequently inviting them to continuously reevaluate it by probing them with critical questions, matches the transformative learning process of higher education students.

Just like cooking or riding a bike, developing statistical analysis plans is a skill you learn by doing, not by merely retaining facts. This is the philosophy of *experiential learning*: learning, experiencing, and applying all occur simultaneously. Kolb and Fry (1974) described the experiential learning process as a four-step cycle in which learners have a concrete experience, reflect on it, form new abstract concepts, and test or apply those new concepts. In our *Statistical Consulting* class, Justifest can stimulate the first iteration of this cycle for our students who can apply the lessons they learned from the tool in their next consulting projects.

VALIDATION

Justifest can be used in any course where students have the prerequisite background knowledge on confidence intervals and hypothesis testing (both parametric and non-parametric methods for group comparison). We intend to deploy it in *Principles of Statistical Data Analysis* (*PrinStat*), giving students the opportunity to review and integrate the material learned in the course, or in the beginning of the semester in *Statistical Consulting* as a review of prerequisites on group comparisons and as an introduction to the investigative process. By surveying students and instructors, we will evaluate the efficacy and user-friendliness of the tool and determine in what context and how it is best used.

To evaluate the impact of exploiting the tool on student grades, we could experiment with and without Justifest and then engage in the same exercise that we present to the students, namely compare the grade distributions with or without introduction of the tool. If we want to draw valid causal inference, we run into the problem that the effect of the tool on the grades is heavily confounded by several factors unless we involve randomization. If the epidemiological situation in the coming academic year does not require severe measures as it did the year before, then any increase in grades that we attribute to the introduction use of Justifest may very well be the result of the absence of burdensome coronavirus restrictions.

To avoid disadvantaging students, we must be creative when organizing a randomized controlled trial, where only one group of students is offered extra assistance in the form of the tool. In a *randomized encouragement trial*, everyone in the course has access, but a random subgroup receives additional encouragement. Beyond an intention-to-treat analysis, this design allows for a randomization-based effect measure of the actual use of the tool, e.g., through structural mean models or a principal stratum approach where the Complier Average Causal Effect (CACE) stands to estimate the average effect of the tool in the subgroup of students who are assigned to use it and actually do use it as suggested (compliers) (see, e.g., Ertefaie et al., 2018; Jin & Rubin, 2009; Ruggeri et al., 2013).

This investigation itself can form the basis of a next edition of Justifest. We look forward to sharing our experience with and getting some feedback—or indeed beta testing—in the near future.

ACKNOWLEDGMENT

We are grateful for several initiatives in the context of continuing education that support the development of blended learning tools. We mention the doctoral schools and FLAMES (Flanders' training network in methodology and statistics) that support the work of Frederik De Keersmaeker and of a group of young colleagues across five Flemish Universities. We further mention the *Voorsprongfonds* that supports lifelong learning in science and the work of Eva Vangenechten.

REFERENCES

Baglin, J., Bedford, A., & Bulmer, M. (2013). Students' experiences and perceptions of using a virtual environment for project-based assessment in an online introductory statistics course. *Technology Innovations in Statistics Education, 7*(2). https://doi.org/10.5070/T572014740

Bloom, B. S. (1956). *Taxonomy of educational objectives. Handbook: The cognitive domain*. David McKay.

Bulmer, M., & Haladyn, J. K. (2011). Life on an island: A simulated population to support student projects in statistics. *Technology Innovations in Statistics Education, 5*(1). https://doi.org/10.5070/T551000187

Cox, D. R., & Efron, B. (2017). Statistical thinking for 21st century scientists. *Science Advances*, *3*(6), Article e1700768. https://doi.org/10.1126/sciadv.1700768
Darius, P. L., Portier, K. M., & Schrevens, E. (2007). Virtual experiments and their use in teaching experimental design. *International Statistical Review/Revue Internationale de Statistique, 75*(3), 281–294. https://doi.org/10.1111/j.1751-5823.2007.00028.x
Ertefaie, A., Hsu, J. Y., Page, L. C., & Small, D. S. (2018). Discovering treatment effect heterogeneity through post-treatment variables with application to the effect of class size on mathematics scores. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 67*(4), 917–938. https://doi.org/10.1111/rssc.12265
Farnell, T., Skledar Matijević, A., & Šćukanec Schmidt, N. (2021). The impact of COVID-19 on higher education: A review of emerging evidence [NESET Report]. *Publications Office of the European Union*. https://data.europa.eu/doi/10.2766/069216
Jin, H., & Rubin, D. B. (2009). Public schools versus private schools: Causal inference with partial compliance. *Journal of Educational and Behavioral Statistics*, *34*(1), 24–45.
Karadag, E. (2021). Effect of COVID-19 pandemic on grade inflation in higher education in Turkey. *PLOS ONE*, *16*(8), Article e0256688. https://doi.org/10.1371/journal.pone.0256688
Kolb, D., & Fry, R. (1974). Toward an applied theory of experiential learning. In C. Cooper (Ed.), *Theories of group process*. John Wiley.
Mezirow, J. (1978a). Perspective transformation. *Adult Education, 28*(2), 100–110. https://doi.org/10.1177/074171367802800202
Mezirow, J. (with Marsick, V.). (1978b). *Education for perspective transformation. Women's re-entry programs in community colleges*. Teachers College, Columbia University.
Moore, D. S., McCabe, G. P, & Craig, B. (2021). *Introduction to the practice of statistics* (10th ed.). MacMillan Learning.
Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*, 2074–2102. https://doi.org/10.1002/sim.8086
Pfannkuch, M. (2018). Reimagining curriculum approaches. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 387–413). Springer Cham. https://doi.org/10.1007/978-3-319-66195-7_12
Pritchard, A. (2017). *Ways of learning: Learning theories for the classroom* (4th ed.). Routledge. https://doi.org/10.4324/9781315460611
Ruggeri, K., Farrington, C., & Brayne, C. (2013). A global model for effective use and evaluation of e-learning in health. *Telemedicine Journal and e-Health: The Official Journal of the American Telemedicine Association*, *19*(4), 312–321. https://doi.org/10.1089/tmj.2012.0175
Steiner, S., & Mackay, R. (2009). Teaching variation reduction using a virtual manufacturing environment. *The American Statistician, 63*, 361–365. https://doi.org/10.1198/tast.2009.08042
Wild, C. J., & Pfannkuch, M.. (1999). Statistical thinking in emperical enquiry. *International Statistical Review, 67*(3), 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x