

TEACHING STATISTICS TO REAL PEOPLE: ADVENTURES IN SOCIAL STOCHASTICS

Rachel M. Fewster

Department of Statistics

University of Auckland, New Zealand

r.fewster@auckland.ac.nz

Peer-supported learning, where students work in small groups and aim to learn with and from each other, is an attractive and increasingly popular model in tertiary education. Learning mathematical statistics has traditionally been a solitary activity, but there is much to be gained from social learning. I will describe several team activities that can be conducted in class sessions for a first undergraduate course in mathematical statistics. Primary aims are to foster versatile thinkers who can readily switch between language, diagrams, and mathematical notation, and to provide a context for creative and open-ended engagement with the subject matter.

BACKGROUND

Listen in to any conversation among contemporary tertiary educators, and you will soon hear phrases like ‘active learning’, ‘non-traditional delivery’, ‘flipping the classroom’, and ‘collaborative learning’. Indeed, if fame in the 21st century can be measured by having a page on Wikipedia, these phrases are famous. Bringing them to life in a large undergraduate statistics class, however, is another matter. In 2010 I began experimenting with non-traditional modes of delivery for STATS 210 (Statistical Theory), a first undergraduate course in mathematical statistics at the University of Auckland. STATS 210 takes about 100 students per semester, and covers basic probability manipulations, discrete and continuous random variables, the principles of hypothesis testing, and maximum likelihood estimation. It is designed as a second-year course, but is also taken by mathematically able first-years, and by math-averse third-years who have deferred the unwelcome theory course to the last possible moment. Consequently, the intake is a diverse mix of students who are variously motivated and demotivated by theory. The challenge for the instructor is to convey the fascination of the field to both groups, providing sufficient challenge to inspire the theory-driven students while simultaneously making the theory accessible, relevant, and inspiring to those who would rather be elsewhere.

Prior to my foray into non-traditional delivery, my philosophy for dealing with this diversity was based on the following quote by American actress Lisa Kirk:

“A gossip is one who talks to you about others; a bore is one who talks to you about himself; and a brilliant conversationalist is one who talks to you about yourself.” – Lisa Kirk

Considering that all my students are real people, whatever their attitudes to theory, it will be a win for statistics if the subject matter leads students to learn things they didn’t already know about themselves and their own lives. In a nutshell, a statistics instructor should be a ‘brilliant conversationalist’, with a mission to inform students about themselves. Applications of statistics are rife in popular psychology and ‘quirkology’ (Wiseman, 2007) and these examples seem to be successful in drawing a broad range of students to the subject. Having gained a strong conviction that the way to students’ hearts is through their biologically programmed social instincts, and as rumblings of non-traditional delivery became ever-louder in staff common rooms, I became keen to further exploit the humanness of my students for statistical gain.

My first attempt at non-traditional delivery was a full version of ‘flipping the classroom’, in which students were to read the course notes in their own time, and lecture time was spent in trying to guide the class to their own discovery and derivation of key ideas, such as the binomial probability formula and why it differs from the geometric, or where the formula for expectation comes from. This met with limited success and was progressively abandoned as the semester went on. One key problem was that discovery takes many times longer than delivery, so it was difficult to cover the syllabus. Additionally, students were not as punctilious about self-directed reading as the educational literature had led me to expect. Instead, I reverted to traditional lectures for delivery

– having discovered perhaps for the first time that lectures are in fact valuable and appreciated by students – but compressed them from three hours to two hours per week, and used the extra time for a collaborative team-learning model described below. The resulting delivery is two hours of lectures, one hour of traditional tutorial worksheets, and one to two hours of team-learning each week. The mix of styles is intentionally not too far from the traditional student comfort zone, but with a new leaning towards more active and collaborative learning.

The structure of the team-learning sessions follows the Team-Based Learning model of Michaelsen et al. (2002; 2008), but implemented as a stand-alone structure without the corresponding changes to lecture style recommended by these authors. The Team-Based Learning model is becoming popular across the tertiary education spectrum, but it is still new in the mathematical sciences (Paterson & Sneddon, 2011) and a greater base of evidence and experience specific to these fields might legitimately be sought by instructors before adopting the model. The aim here is to describe a workable implementation for a moderately large, theoretically-oriented course, and encourage sharing of materials and ideas for team activities to reduce the burden on instructors and create more opportunities for studying the model's effectiveness.

STRUCTURE OF TEAM SESSIONS

Team sessions are scheduled in a weekly two-hour block, with the second hour only to ensure there is no time pressure; most teams finish within 50-75 minutes. Teams are composed of five or six students, with fixed membership for the duration of the course, and are chosen by the instructor. Experience suggests that teams comprising students with broadly similar ability are more successful in generating full participation than teams aimed at diverse ability. The 100-strong class is split between three flat-floored seminar rooms, with a tutor in each room. Shifting to small flat-floored rooms was a great improvement over initial attempts to use a tiered lecture theatre: being real people, students are highly influenced by the comfort and intimacy of their surroundings.

Following the model of Michaelsen et al. (2002; 2008), the weekly sessions alternate between multiple-choice quizzes and 'team tasks' or activities. The quiz sessions begin with students doing a quiz of 10-15 multiple choice questions individually under test conditions. Quiz questions are aimed at concept rather than calculation, and specifically aim to pose a different style of question from the other assessment methods of tutorials, assignments, and exams. There needs to be enough surprise in the questions to create ample opportunities for discussion within the teams. After about 20 minutes, the individual quiz answers are collected in and students join their teammates. They then redo the same quiz with their team, but are now provided with a scratch-card on which to scratch off the answers decided by the team. The correct answer is identified when obtained, so the team knows immediately if they have it right (4 marks). If they select a wrong answer, they must continue discussion, identify what went wrong and select an alternative answer for 2 marks. The process continues until they have only one choice left. This model is an effective way of delivering immediate feedback, and shifts the emphasis of the activity to understanding a suite of correct and incorrect arguments, rather than simply to complete the test and get a mark. Assessment for the quiz sessions operates on 50% individual mark and 50% team mark, with students almost invariably doing substantially better as a team than individually.

The team tasks that take place in alternate weeks involve open-ended exercises aimed to encourage creative thinking and exploration of the course concepts. Students work with their teams for the whole session, and assessment is based only on attendance. Each team produces a single set of answers or output, and when completed they must get a copy of sample answers from the tutor and discuss their own answers in the light of the sample answers. This eases the burden of marking on the instructor, especially as open-ended questions are time-consuming and hard to mark fairly, but it does put responsibility on the students for ensuring they have properly understood the activity. Tutors must be on high alert to intercept teams that try to leave before the feedback stage has been completed, and ensure everyone is happy with the concepts. The rest of this manuscript focuses on ideas and experiences relating to the design of these team activities.

TEAM ACTIVITIES

Task 1: Probability Concepts

This is the first activity in the course, and is a gentle introduction to help students to get to know each other as well as inducting them into the style of open-ended questions. The aim is to encourage versatility in thinking about simple probability concepts using language, diagrams, and mathematical notation, and in switching seamlessly between them. An example setting is below:

Mr Tambourine runs a musical coffee shop called Cafe Swan Cake, which attracts musical customers from all over the world. Define the sample space $\Omega = \{\text{customers}\}$, and events $I = \{\text{customer is Irish}\}$ and $B = \{\text{customer plays the Banjo}\}$. The information given is $P(B | I) = 0.4$, and $P(B \cap I) = 0.2$.

Possible team activities based on this setting include:

- Create two collections of sentences in natural language that express the information $P(B | I) = 0.4$ and $P(B \cap I) = 0.2$, using as many different sentence structures as possible. This aims to encourage students to expand their linguistic understanding of the notation, and think more deeply about the distinction between the two expressions.
- Draw Ω , B , and I on a Venn diagram that accurately depicts the information $P(B | I) = 0.4$ and $P(B \cap I) = 0.2$, using rectangular regions for events B and I and depicting probabilities by areas on the diagram. The diagram (Figure 1) emphasizes that $P(B | I)$ is the probability of encountering event B within the restricted sample space I . This way of understanding conditional probability links directly to the formula $P(B \cap I) / P(I)$, but is not well-conveyed by the traditional language ‘ B given I ’.
- Draw further Venn diagrams of Ω , B , and I , showing B and I variously as independent and non-independent events. The purpose here is to create a visual impression of what independence means (not easy!), and also to illustrate the limitations of Venn diagrams, motivating the need to develop a mathematical formulation of probability. Sample diagrams are in Figure 1.
- In one or two sentences, explain to a high school student why independence of B and I means that $P(B \cap I) = P(B)P(I)$.

Exercises like this show how very simple settings can be used to create considerable discussion and foster new insights. From observation, although not all students comprehended the visual representations of conditional probability and independence, the exercise did seem to lead to a widespread tendency to sketch diagrams as a first response to a verbal question. I logged this as a minor breakthrough, because students became inclined to sketch diagrams on their own volition instead of only when told to. This interest in visual thinking also seemed to transfer to later material in the course, as if a ‘doodling barrier’ had somehow been overcome by the group activity.

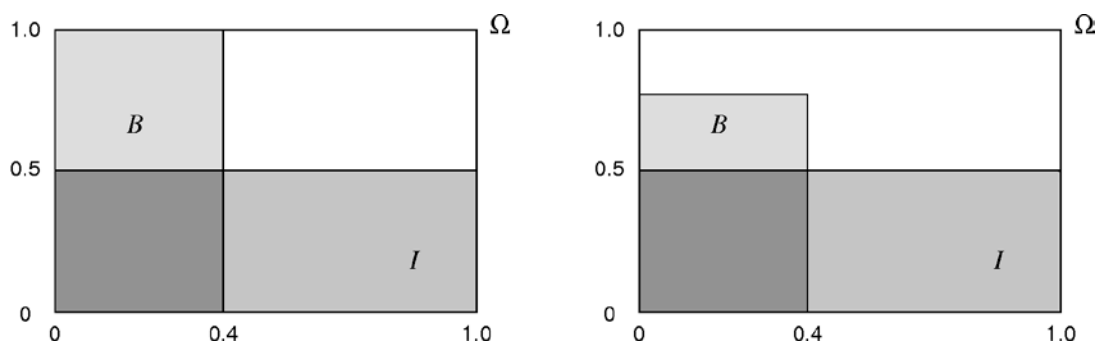


Figure 1. Events B and I within sample space Ω such that $P(B | I) = 0.4$ and $P(B \cap I) = 0.2$. On the left, B (vertical rectangle) occupies the same proportion within I (horizontal rectangle) as it occupies within the whole space Ω , so $P(B | I) = P(B)$ and the events are independent. Any other diagram, such as that on the right, shows B and I as non-independent events.

Task 2: Hypothesis Testing

Students are introduced to the idea of hypothesis testing in a simple binomial setting: toss a coin 10 times and observe 9 heads; can you continue to believe that the coin is fair? Soon after this they are asked to conduct their own hypothesis test as a team activity. All that needs to be decided by the instructor is a simple but engaging example that students can use to collect their own observations during the session.

My first attempt at this activity invited students to investigate the following question:

Is it true that dropped toast tends to land on the carpet butter-side down?

Teams were provided with ample quantities of toast (but with squeeze water bottles and marker pens instead of butter), and asked to investigate, keeping in mind the usual context in which this question might be of interest. I imagined this to evoke images of toast flying off the breakfast table, but to my surprise this was not evident to students, with one team even interpreting it as a requirement to stand on desks and drop toast from a precisely measured height of 2 metres. (I intend to decline any invitations to breakfast with these particular students!) Overall, this was my least successful attempt at designing a team activity, considering the general chaos of flying toast, the preparation and clearing up, and, significantly, the apparent lack of statistical learning that took place. My conclusion from this was that, although it is instructive to students to experience the data-collecting mechanism, if this phase is too distracting or energy-consuming it limits the scope for appreciating the real point of the exercise. To this day, I use the euphemism 'too much toast' to dismiss ideas for activities that over-emphasize distractions from the educational goal.

The following investigation proved more successful:

Can you cheat with dice by baking them?

The idea is that baking dice (for 20 minutes at 120°C to be precise) causes the plastic to melt slightly and sink to the bottom, loading the dice in favor of the number showing at the top. For fan-assisted ovens, it is worth ensuring all dice are baked with the same orientation in case there is a greater impact of the fan action. Having baked the dice myself (ensure adequate ventilation if doing this at home) the students only need to roll the dice enough times to collect their data and conduct the hypothesis test. No significant effects have been found in my classes to date, despite promises on the internet of endless riches to be gained by treating dice in this manner!

For further challenge, red dice can be treated to fall in favor of 6 and white dice treated to fall against 6. Teams are asked to design a way of investigating whether a red die is *more likely than a white die* to display a 6, still using only a one-sample binomial test. They can do this by setting up a contest between the two dice, rolling them together until they have observed n occasions on which one die displays a 6 and the other does not. They then test whether there is a significant imbalance of red 6s among these n trials. Most teams need some hints to arrive at this design, but they enjoy the twist on the binomial theme. They then experience the wastefulness of discarding most of their die-rolls, which is a good way to motivate the need for more advanced testing designs such as a two-sample test.

Task 3: Movie Making

I have long been interested in getting students to create their own dialogues as a way of personalizing and relating to the course material. Putting this into practice has been problematic, not least because students are reluctant to explain concepts on record in case they say something perceived as stupid. New web-based tools provide a superb opportunity to gain all the benefits of this exercise with none of the disadvantages. Teams can create animated movies or cartoon strips online, and importantly, if anything less than sensible is said, it is not said by the students themselves but by their cartoon characters.

The best tool I found for animated movie-making was called Xtranormal, which sadly went offline in late 2013. Free, web-based, and very easy to use, the team just needed to select characters and type in their dialogue, which was then animated and spoken by the characters. Figure 2 shows

a screen capture and script from an Xtranormal movie that was created in one hour – starting with no prior knowledge of the software – by a team with the following brief:

Make a movie on a topic connected with sample size.

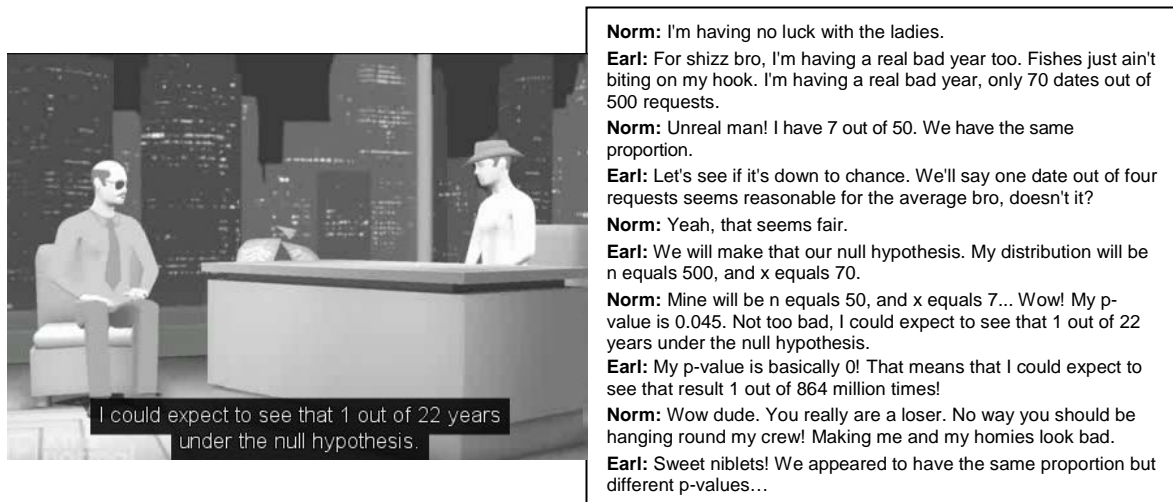


Figure 2. Screen capture and script of the 2010 Falcon team's Xtranormal movie: written by Mark Burgess, Wei Guang Chen, Adele Havinga, Georgia Miskell, Jessica Simons, Han Kil Suh, and Taskeen Yaqoob. The movie can be watched at www.stat.auckland.ac.nz/~fewster/falcon.html. The characters use one-tailed p -values where two-tailed would be preferable; however, they set up their null hypothesis beautifully and the script alternates brilliantly between colloquial and technical language.

The movie in Figure 2 addresses the effect of sample size on the p -value, a topic that I have grappled with for years with respect to improving students' understanding. The movie-making exercise, and this movie in particular, seems to have created a real breakthrough in understanding this topic – not just for the team who wrote it, but also for all the other students who have viewed and enjoyed it. Since this movie was created, my 100-strong classes almost unanimously give flawless answers to exam questions on the topic (and they include their own diagrams, which I trace back to the legacy of Task 1). Indeed, students who evidently do not have a clear understanding of what a p -value is, nonetheless expertly address the matter of how it changes with sample size.

Although Xtranormal is no longer available, there are several alternatives, and new options appear frequently. Some possibilities (not yet road-tested) are: GoAnimate, which does the same as Xtranormal but requires a subscription; Plotagon, which is similar to Xtranormal and free but needs to be installed; PowToon, which makes free animated videos without the text-to-speech facility but with a number of other interesting features; and Pixton, which is free web software for making comic strips. Websites for all these tools, and more, are readily found by Google searches.

A challenge associated with the movie activity is how to exemplify the possibilities without stifling creativity. I have run this activity twice: the first time without any previous student exemplars, and the second time using the exemplar in Figure 2. On the first occasion, there were a few outstanding movies, but also several humdrum results. On the second occasion, all of the movies were good, but in the statistical sense they were clones of the exemplar. In future I plan to run the activity with a different topic or a different medium (e.g. comic strips rather than movies), in the hope that exemplars from a previous topic or medium might be successful at inspiring without interfering with the creative juices.

Task 4: Spies and Agents – Maximum Likelihood Estimation of a Mixture Proportion

Students in STATS 210 are introduced to the method of maximum likelihood through one-parameter examples such as estimating the binomial parameter p in a situation with known n . A problem with this approach is that, although the method of maximum likelihood seems sensible, it also seems like a lot of work to obtain perfectly obvious answers. This activity is aimed at illustrating maximum likelihood in a case where the estimate is not so obvious, but the distributional form is still simple. The idea is to create a context for estimating a mixture proportion s , such that the binomial p parameter is $\alpha s + \beta(1-s)$ for known α and β .

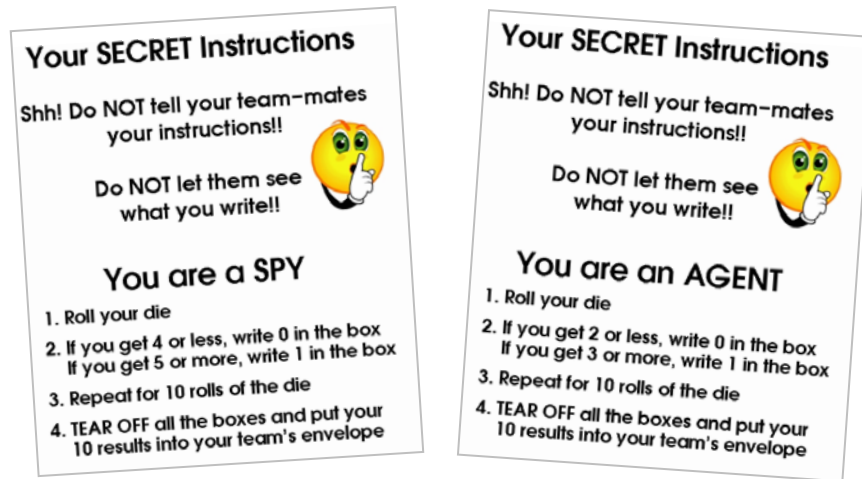


Figure 3. Secret instructions distributed to each team member. The team's data of 0s and 1s results from an unknown mix of Spies and Agents.

Figure 3 shows instructions used for a classroom game of 'Spies and Agents' that creates such a scenario. Each team member is given a set of secret instructions that defines them as a spy or an agent. Spies – being subject to distractions – tend to be less successful at completing their missions than agents. Each team member rolls a die ten times, to generate outcomes failure or success for ten missions. The success probability is $1/3$ for spies and $2/3$ for agents. Each student writes the results of ten missions into a strip of ten boxes provided, then tears off each box and submits the data to a common envelope. When the data are emptied out of the envelope, they constitute a barrage of 0s and 1s with no knowledge of who contributed which items. At this point the team learns about the two sets of instructions, but not about the identities of team members. As a team they calculate that the overall probability of a successful mission is $(2-s)/3$, where s is the proportion of spies in the team, and formulate the binomial likelihood. Having obtained their maximum likelihood estimate of s , they mutually confess their status and compare their estimate of s to the corresponding true value.

Initial experience of this activity suggested that students appreciated the opportunity to see maximum likelihood uncover something that was not blindingly obvious, especially with the conspiratorial game-like connotations of uncovering illicit knowledge about team members. However, a much greater (and unexpected) benefit seemed to emerge on the final exam, where the class was set a statistically identical problem but in a very different context involving apple blight. The exam question demanded considerably deeper understanding, including derivation of estimators and estimator variances and comparison of different estimation methods. About a third of the class demonstrated a deep and mature understanding, contrasted with the roughly 10% that I expected from previous experience. This suggests that the team activity might have acted in an enabling fashion, perhaps transforming the mixture example from an abstract problem to a real-world experience, and providing the seed for further development in private study. Although this result is anecdotal and may be an outcome of wishful thinking on the part of the instructor, it hints of subtle effects of social learning that would be worthwhile investigating in a more scientific way.

Task 5: Poisson Regression

Poisson regression offers another opportunity to find a one-parameter maximum likelihood estimate with a simple likelihood structure, but with an estimator that is not obvious and differs from that of ordinary least-squares regression. Data points satisfy $Y_x \sim \text{Poisson}(\beta x)$ for $x = 1, 2, \dots$, where β is the parameter to be estimated; thus β corresponds to the slope of the best fit line through the points (x, Y_x) . The difficulty comes in finding a Poisson data-generating mechanism that can be easily used in class, without involving ‘too much toast’ (see Task 2). I settled on using a Poisson point process, where each student is given their own individual sheet with the point process ready-printed, also showing randomly-chosen centers of three non-overlapping circles of different sizes. Generating a PDF of 100 different point processes for printing is easily done in software such as R. The circles themselves are printed onto a transparency that is passed around the team, so all the student needs to do to collect their individual data is to center the transparency correctly for each circle and count the number of points that lie within each circle. The number of points inside circle x is Y_x for $x = 1, 2, 3$. The students know that the circle sizes have been chosen such that $Y_x \sim \text{Poisson}(\beta x)$ for some β , and their job is to estimate β . It later emerges that the true value of β is $\beta = \pi = 3.14$. Thus the exercise provides a way of estimating the mysterious value π , noting that knowledge of the true value of π is never needed to obtain the estimate.

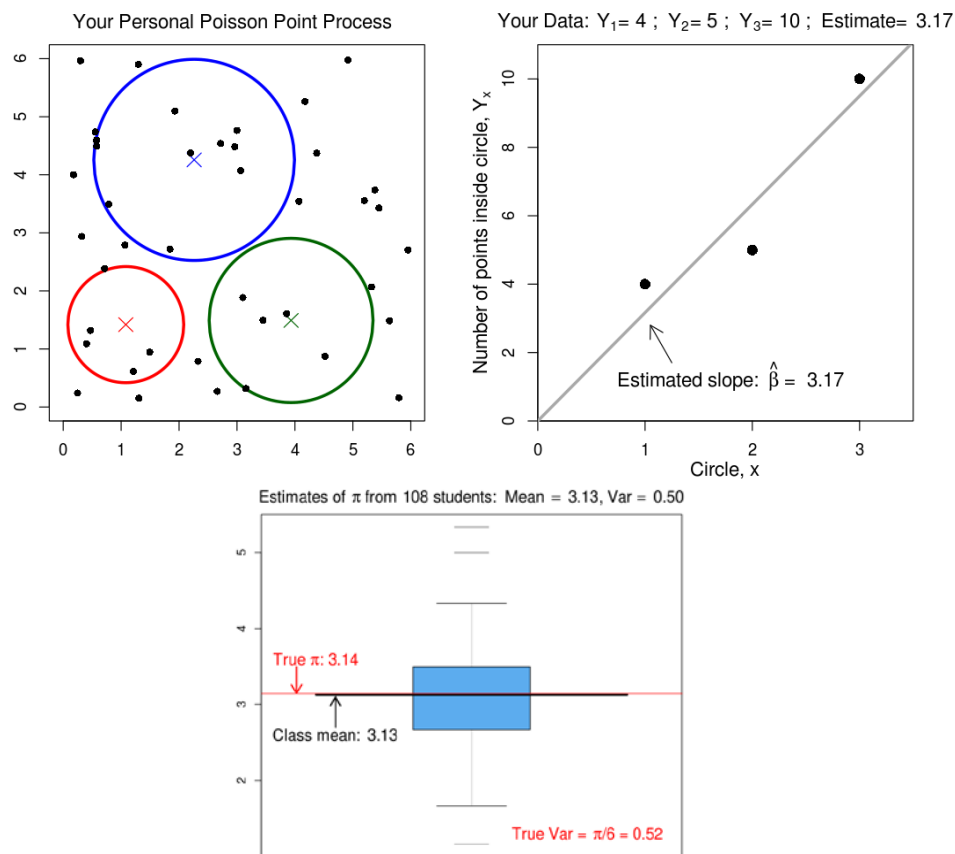


Figure 4. Poisson regression activity. The top panels show the individual data and regression estimate for one student. The Personal Point Process is provided on a printed sheet, which also shows the centers of the three circles. The circles are provided on a transparency that the student places over the printed sheet. Students place the circles correctly and count the numbers of points inside circle 1 (small), circle 2 (medium), and circle 3 (large) to gain their data Y_1, Y_2 , and Y_3 . The circles and point process are designed such that $Y_x \sim \text{Poisson}(\beta x)$ where β is the parameter to be estimated: in fact the true value of β is $\pi=3.14$. The estimate of π from this student’s data is 3.17 (top right). Results from the class’s 108 estimates of π are shown in the bottom panel. The class mean of 3.13 and variance of 0.50 are very close to the estimator mean and variance of 3.14 and 0.52 respectively. No knowledge of the value of π is needed to create the estimate.

Having collected individual data, the team works through the problem together to find the maximum likelihood estimator of β , which is $(Y_1+Y_2+Y_3)/6$. They each substitute their own values of Y_1 , Y_2 , and Y_3 , to get a collection of estimates from the teammates. They then derive the estimator mean to be β , and the estimator variance to be $\beta/6$. After it is revealed that the true value of β is $\pi = 3.14$, they can assess the sample mean and variance of their team's estimates against the estimator mean of π and variance of $\pi/6 = 0.52$. Figure 4 shows the set-up of the activity, and the collection of individual estimates from a class of 108 students.

CONCLUSIONS

All teaching innovations are likely to create a mix of good and not-so-good experiences. Overall, my experience with the teamwork model has been rewarding and eye-opening, albeit demanding of time and energy. Oddly, the less structure there is in an activity, such as movie-making or toast-throwing, the more preparation it seems to demand from the instructor. My initial aims of attempting a scientific study of teamwork effectiveness were abandoned due to the time involved merely to introduce the model. Consequently, opinions expressed here about specific triumphs and anti-triumphs are only anecdotal. A key motivation of this article is to provide a source of ideas that may help other instructors to start implementing team learning, in the hope that these anecdotal theories may eventually be placed on a more evidential footing.

The greatest triumphs that I attribute to the team model have been in expanding students' ways of thinking about and relating to the subject material. These include changed behavior in visual thinking observed after Task 1; the embracing of the subject matter into the students' own language and culture through movie-making in Task 3; and the enabling effect for private study conjectured to have resulted from the team game in Task 4. The new slant on the subject exposed some misunderstandings that I had not previously realized: for example, in movie-making, a number of teams created dialogues involving hypothesis tests without first formulating a statistically-meaningful hypothesis.

I would not say that teamwork is a magic bullet for all statistics teaching: indeed, the experience has also rekindled my enthusiasm for traditional modes of delivery. Even if students value lectures for no better reason than inertia or lack of imagination, the bare fact that they do value lectures will engender a negative response if they are suddenly withdrawn. In my opinion, the winning combination comprises a mix of styles and activities, including both traditional and non-traditional elements. My experiences agree with the study of Gokhale (1995), who found that collaborative learning made a significant difference to critical thinking, but not to performance on drill-and-practice exercises pertaining to factual knowledge and comprehension.

Feedback on the team structure from student questionnaires has been very positive. On open-ended questions, more than twice as many students cited teamwork as helpful (29/60) than the next most commonly favored activity of worksheet tutorials (14/60), which in turn was slightly ahead of lectures and assignments. Only 3/60 responses suggested that teamwork was not helpful. A number of students commented that they liked either the multi-choice quizzes or the team activities, but not the other; however, preferences were roughly equally split between the two styles. Some aspects need careful handling to avoid making students anxious: for example, multi-choice quizzes need to be challenging enough to provide meat for discussion, but this means that students score lower on their individual quizzes than they are accustomed to. In STATS 210, quiz performance trends at about 10% below final exam performance. Some academically-oriented students who are accustomed to high marks do not perform so well on creative and open-ended tasks, which can sow seeds of discontent; this is offset by other students who unexpectedly come into their element in a team setting. More technical tasks such as Task 5 can result in one or two students doing all the mathematical work, while the other students get left behind, and more thought about how to address this issue may be needed. The entire teamwork component in STATS 210 contributes only 7% of the final mark, with the aim that students will recognize that it is very much intended as a formative rather than a summative framework.

ACKNOWLEDGMENTS

I am indebted to Judy Paterson, who introduced me to team-based learning and acted as mentor and support team throughout, and to Chieh-Hsi Wu and Stephanie Budgett for help and insights in the classroom.

REFERENCES

- Gokhale, A. A. (1995). Collaborative Learning Enhances Critical Thinking. *Journal of Technology Education* 7, 22-30.
- Michaelsen, L., Fink, D., & Knight, A. (2002). *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*. Sterling, Virginia: Stylus Publishing.
- Michaelsen, L. K., Sweet, M., & Parmelee, D. X. (2008). *Team-Based Learning: Small Group Learning's Next Big Step. New Directions for Teaching and Learning*. San Francisco: Jossey-Bass.
- Paterson, J., & Sneddon, J. (2011). Conversations about Curriculum Change: Mathematical Thinking and Team-Based Learning in a Discrete Mathematics Course. *International Journal of Mathematical Education in Science and Technology*, 42, 879-889.
- Wiseman, R. (2007) *Quirkology: How We Discover The Big Truths In Small Things*. New York: Basic Books.