

AUTOMATED, RECEPTIVE AND INTERACTIVE: A CLASSROOM-BASED DATA GENERATION EXERCISE

LANGAN, Dean & WADE, Angie

Great Ormond Street Institute of Child Health, University College London, 30 Guilford Street,
London WC1N 1EH.

d.langan@ucl.ac.uk

It is easier to engage with statistics training when presented with examples from familiar subject areas. However, when teaching students of varying professional backgrounds, finding relatable examples can be especially challenging. Classroom-based data generation exercises offer a solution with students involved in the process from data collection through to choice and use of appropriate analyses. One such exercise that forms an integral part of an introductory statistics course is based on beermat (coaster) flipping, a popular pub game in the UK. We recently moved the data collection process online allowing students to enter data via smartphones. Furthermore, a web application has been developed using the shiny package in R. This application automizes data analysis and allows students to explore the results interactively and independently. The application comes to life with visual demonstrations of core concepts such as the central limit theorem and bootstrapping. This technology further engages students and the ensuing discussion comparing outputs and interpretation is a welcome addition to classroom interactivity. We present details of this exercise, focussing on use of the web application, example outputs, student feedback and guidance for best practice to maximise learning outcomes.

INTRODUCTION

Teaching statistics has transitioned from focus on probability theory and statistical inference as an abstract concept, to practical application and statistical reasoning in many applied disciplines (Bradsheet 1996, Smith 1998). Topics typically taught in today's classroom include the principles of good research design, data collection and generation of answers to research questions through data analysis. We have witnessed a shift towards cooperative learning; a technique giving students more ownership of their learning by allowing students to experience first-hand a quantitative research process (Garfield 1993). A natural synergy of this transition has been to engage students through group exercises that reflect a real quantitative research.

Smith (1998) states "students are more easily convinced of the power of statistical reasoning if they see it applied to questions that are interesting and real to them". In a homogenous classroom, a meaningful project is naturally one that relates to the group's primary research field. In a heterogenous group, such as those containing students from different programmes or short courses, hands-on data generation exercises are advocated. This type of activity gives students a sense of shared ownership over the data and provides motivation to learn universal statistical concepts applicable to many fields. Lee & Famoye (2006) state "data generated from students themselves tend to draw their attention and motivate interest more than a dataset disconnected from their everyday life".

In this paper, we present a collaborative exercise refined over a decade by the Centre for Applied Statistics Courses (www.ucl.ac.uk/stats-courses), who are based in University College London (UCL) (United Kingdom). This exercise is run within a stand-alone statistics course titled "Introduction to Statistics and Research Methods" that typically attracts heterogeneous students. They are asked to play a game in pairs that involves flipping a beermat (known as a coaster in the US and Australia); a game made popular in pubs across the UK. Instructions are provided in *Figure 1* on how to successfully flip a beermat. The main research question explored during this course is – is beermat flipping ability related to height? Previously, this exercise was performed using paper-based data collection forms, and analysis was performed by a statistician (Koutoumanou & Wade 2017). As of 2019, the data is entered by students in an online database and analysis are now automated, receptive and interactive through a new web application developed using the *shiny* package in R (Chang et al. 2021).

We provide further details of the development of this exercise split into two parts; data generation and statistical analysis. We present example output, summarise anonymous feedback

from students and reflect on its effectiveness measured in terms of learning outcomes. A teacher may wish to recreate the exercise in their own statistical course or simply gain inspiration for a similar exercise.

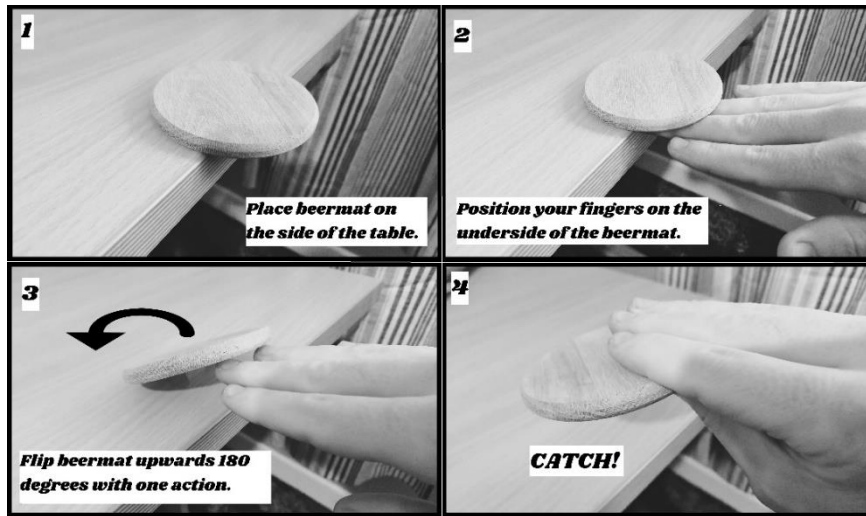


Figure 1: How to successfully flip a beer mat

A DATA GENERATION EXERCISE

The development of this data generation exercise falls broadly into three phases: (1) Up to 2018, when data collection was paper-based, and analysis was performed manually through the SPSS software package. (2) In 2019, the data collection and analysis were automated with the intention of running the exercise with students face-to-face in the classroom. (3) From 2020, the exercise was adapted for online teaching due to the COVID-19 pandemic. Given the popularity of online courses, this format is set to continue after students can return to the classroom. In this section we provide details of what the exercise entailed for the student and the teacher at each stage.

1. Manual (2010-2018)

Students are handed a paper form and beer mat and arranged randomly into pairs by calling out numbers in the classroom. The tallest person is asked to perform the task first. This is not the optimal study design and may introduce confounding bias (due to a learning effect), which the students are expected to recognise and thus provides a discussion point.

<p style="text-align: center;">Pair number <u>1</u></p> <p>Tallest of pair is person A Shortest is person B.</p> <p>Are you A or B? <u>A</u></p> <p>What is your height? <u>185cm</u></p> <p>Gender? <input checked="" type="radio"/> Male / Female</p> <p>Beer mat flips:</p> <p>Flip the mat once. Did you catch it? Yes <input checked="" type="radio"/> No</p> <p>Now perform 20 flips. How many catches from the 20 flips? <u>9</u></p> <p>See how many flips you can do in 1 minute <u>49</u></p> <p>How long does it take you (in seconds) to 10 catches? <u>29 sec</u></p> <p>Finally, do one more flip. Did you catch it? Yes <input checked="" type="radio"/> No</p>	<p>Is beer mat flipping ability related to height?</p> <p><small>You will be divided into pairs for this study. Each pair will be assigned a number, and will have a beer mat and some form of timing device. Each pair should follow the instructions on this form and fill out any results.</small></p> <p><small>The tallest person of the group should perform each task first.</small></p> <hr/> <p>Pair Number</p> <p>Your answer _____</p> <hr/> <p>Height of the tallest person (cm)</p> <p>Your answer _____</p> <hr/> <p>Height of the shortest person (cm)</p>
---	--

Figure 2: A paper data entry form collected from a student (left) compared against our new online form (right) (note: the questions have changed a little in the process)

The paper form is shown in *Figure 2 (left)*. Questions that students were asked on this paper form include (1) What is your height? (2) Gender? (3) Flip the beer mat once, did you catch it? (4) Now perform 20 flips, how many catches from 20 flips? (5) See how many flips you can do in 1 minute. (6) How long does it take you (in seconds) to 10 catches? (7) Finally, do one more flip, did you catch it? Questions 1-2 provide demographic details of the participants that can be used to explore relationships with subsequent measurements of beer mat flipping ability (questions 3 – 7). We gather information in a variety of formats to subsequently demonstrate to students how the type of data impacts approach to analysis. For example, students are asked to perform one flip at the beginning and end of the exercise, providing within-person paired data so that improvement/deterioration can be investigated (questions 3 and 7).

2. Automated (2019)

Most stages of the exercise were automated in 2019, now requiring little time on the part of the statistician. Students can enter data directly into an online spreadsheet through a google form (as shown in *Figure 2, right*). As a side note, we now ask students whether they were born in the UK rather than gender, since this information is less sensitive.

3. Online (from 2020)

In 2020, the core introductory course moved online due to the COVID-19 pandemic, and so did the beer mat flipping exercise. Given much of the exercise had been automated in the previous year, this meant coincidentally that minimal changes were required to adapt to teaching and performing this exercise online. First, we recognised that not everyone will have access to a beer mat in the home, so instead recommend using some form of credit or membership card instead. Arranging students into pairs can no longer be done physically, so instead we make use of the breakout rooms facility in Zoom (zoom.us). This arrangement runs much more smoothly than the equivalent task in the classroom. We found that such interactive exercises that allow students to connect is essential in a virtual environment to avoid disengagement with a lecture-style delivery.

STATISTICAL ANALYSIS

The development of the analysis phase of this exercise falls broadly into two time periods: (1) Before automation up to 2018, when the statistician performed analysis manually and students were provided with static printed results, and (2) After automation from 2019, when analysis is performed through a web application. Little has changed since providing online courses, so we describe the exercise within this setting in phase (2) alongside the equivalent classroom-based exercise. In this section, we provide details of what analysis entails.

1. Manual (2010-2019)

The paper forms are completed by students and entered manually into a spreadsheet by a statistician outside the classroom. SPSS syntax was created so that results can be generated automatically, and a report of the results is produced containing only graphs and tables. The report is printed for the students in time for the next class, where results can be discussed either as a whole class or in smaller groups. This exercise is either for revision in the final afternoon of the course or run in stages to give students a regular break from the core materials.

Students are presented with a series of questions that prompt them to relate the exercise to the course materials and relate readily to the type of questions the student might face in a real research project. These prompts touch on ideas of sampling bias, sample size, the limitations of pairing, appropriate analysis methods and whether the study could have been improved in some way. More specifically, there are questions relating to study design and those that require use of the printed output, such as:

- Could the order of performing the flips, tallest always first, have affected the results?
- How could the outcomes be displayed to show the relationship between ability to flip beer mats and height?
- How could you assess whether the ability to catch a single toss improves with practice?

- How would you analyse the within pair differences to see whether there was a consistent tendency for taller individuals to have a better (or worse) ability to flip beer mats?

2. Automated (from 2019)

As of 2019, the data (collected electronically) can be immediately incorporated into the analysis and results can be viewed in real-time through a web application developed using the shiny package in R (Chang et al. 2021). The application reads in the data from an online spreadsheet that automatically collects all entries from the data entry form. Screenshots are provided from this application in *Figures 3 and 4*. The application is available through the link tinyurl.com/OZCOTS-app and source code available via tinyurl.com/OZCOTS-code. Similar questions are proposed to the students in the classroom as those before the exercise was automated (see above). The key difference being that students can now access the application, make their own selections for choice of analysis and see the results appear in real time.

Rank	pair	height	uk	flip1	flip20	flipmin	flip10sec	flip2	tallest
1	1.00	177.00	No	No	11.00	38.00	25.00	No	No
2	1.00	182.00	Yes	Yes	9.00	37.00	20.00	Yes	Yes
3	2.00	165.00	No	No	4.00	4.00	120.00	No	No
4	2.00	173.00	No	No	13.00	9.00	70.00	Yes	Yes
5	3.00	162.00	No	No	15.00	22.00	29.00	Yes	No
6	3.00	183.00	Yes	Yes	14.00	17.00	15.00	Yes	Yes
7	4.00	160.00	No	Yes	14.00	32.00	22.00	Yes	No
8	4.00	175.00	Yes	Yes	15.00	32.00	21.00	No	Yes
9	5.00	157.00	Yes	No	16.00	25.00	31.00	Yes	No
10	5.00	180.00	Yes	Yes	14.00	17.00	26.00	Yes	Yes
11	6.00	160.00	Yes	Yes	17.00	28.00	22.00	Yes	No
12	6.00	162.00	Yes	Yes	12.00	11.00	57.00	No	Yes

Figure 3: Data and leader board tab of the web application

Figure 3 presents a screenshot of the ‘data and leader board’ tab of the application. First, students can select data belonging to any cohort of students from the left-hand side of the application (label 1). The most recent date is selected as a default since this date most likely belongs to the student’s cohort, although any combination of dates can also be selected too to increase the sample size. Data collected on the selected dates then becomes the analysis dataset for all other tabs in the application. On this same tab, a leader board is presented (label 2), which can be sorted in order of any of the variables (label 3). The student might choose to order by some variable relating to beer mat flipping ability to see how they fared in relation to their peers.

Figure 4: Linear regression tab of the web application

Figure 4 presents a screenshot of the application from the final ‘regression’ tab. We provide this as a typical example how students can interact and create results with other results tabs being set up in a similar way. The students are able to select only a numeric variable as the outcome (since only linear regression is taught during the course) (label 1) and one or more independent variables that can take any form (binary or numeric) (label 2). In the screenshot, ‘number of successful flips in one minute’ is chosen as the outcome and ‘height’ and ‘born in the UK’ as the two independent variables. Results are shown as they would appear in the R console (label 3) and also equivalent results pasted into an equation format (label 4). One of the strengths of these shiny apps is the ability to merge results into the text, providing students with a fuller explanation than you would typically see in the output for a statistical software package.

FEEDBACK

We present feedback that was collected anonymously via an online questionnaire after an iteration of the course in 2019 within a classroom. 28 of 35 students taking part in the course responded. Students were asked to rate the data generation exercise, the ease of data entry and the web application showing the results on scales of 1 (poor) to 5 (excellent). There were 23 (82%) respondents that rated the data generation aspect of the exercise a four or above; the equivalent numbers in relation to the data entry form and web application were 23 (82%) and 22 (79%) respectively (Figure 5).

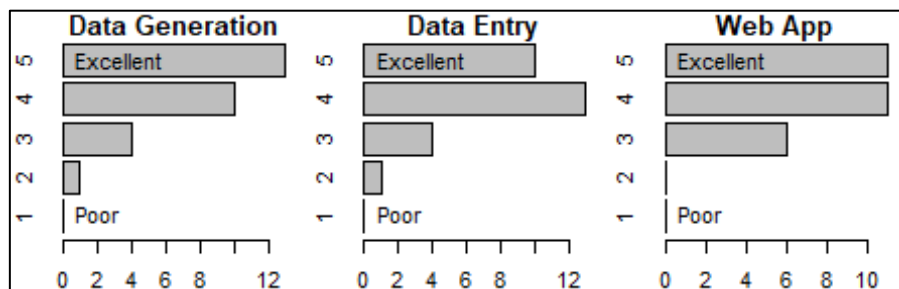


Figure 5: Feedback on data generation, data entry and the web application components of the exercise on a scale of 1 (poor) to 5 (excellent). Frequencies (x-axis) and score (y-axis) presented.

We received 15 free text comments, of which 8 (53%) were solely positive, 6 (40%) neutral and one (7%) negative. Of the positive comments, one student stated, “it was a great ice breaker, and being able to apply what we learned to this data (via [the] app) helped me to understand how to apply the concepts to real data”. Another student commented “[I] thought this was really useful and I was really impressed by how well integrated it was throughout the course”. The six neutral comments included four students that suggested more time allocated for the exercise, one student felt a simple written example without participation would have personally fulfilled the same purpose, and a final student stated, “I would like to see this being run on SPSS or Stata”. Our students have very different preferences for statistical software and varying computer programming skills, but perhaps the application could provide access to the data for download and include the R code used to generate results.

CONCLUSIONS

Data generation exercises such as that demonstrated in this paper give students an opportunity to take ownership over their learning. These exercises give students shared interest, which is particularly challenging in a diverse classroom and facilitates understanding of steps involved in typical research processes. Our exercise encourages aspects of statistical thinking as defined by Wild (1999), closely mimicking a typical interrogative research cycle, where students interpret results and critically appraise them in light of the limitations of the data collection process.

Related classroom exercises that involve students generating their own data can also be found. Zeleke & Lee (2010) suggest research questions such as “is hand size a good predictor of height” and “how many raisins in a 0.5 oz. raisin box”. Zetterqvist (1997) gives similar examples

of exercises within a chemistry class, such as “determine the concentration of copper in a piece of impregnated wood”. Other exercises focus on theoretical understanding. For example, demonstrating the central limit theorem by having students pick random numbers in small groups (Zacharopoulou 2006). These exercises typically involve manual data collection and analysis, setting a clear distinction from the beermat exercise demonstrated in this paper. For example, Zeleke & Lee (2010) ask volunteer students to collect the paper data entry forms and create appropriate plots, but stated a limitation of their exercises were lack of time and recommend online resources for data collection to speed up the process. Applications developed through the shiny package, much like our application, are commonplace in statistical education as demonstrated in a review by Doi (2016). However, these are generally built to explain concepts such as the central limit theorem, rather than incorporate student-generated data.

Automating data collection and analysis saves time for lecturers and students, providing an opportunity for data collection and analysis to take place within the same class. These stages of the research process don't typically run as smoothly in real life and this point should be emphasised in class. One of the themes coming from feedback was that students like to see how analysis might be carried out in a particular software package; this can be challenging in a classroom filled with diverse student preferences, but highlights how providing the data, code and examples for different packages could further enhance the exercise. This would be a simple process and illustrates the value of student feedback to improve classroom activities and make them more inclusive. The feedback collected for this exercise meant that a full evaluation of the learning outcomes was not possible, but will be a valuable avenue for future research. The online environment has led to the rapid development of this useful tool, the benefits of which will also be apparent with face-to-face courses when these return.

REFERENCES

- Bradstreet, T. E. (1996). Teaching introductory statistics courses so that non-statisticians experience statistical reasoning. *The American Statistician*, 50(1), 69-78.
- Chang, W., Cheng, J., Allaire J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, D., & Borges, B. (2021). shiny: Web Application Framework for R. R package version 1.6.0. <https://CRAN.R-project.org/package=shiny>
- Doi, J., Potter, G., Wong, J., Alcaraz, I., and Chi, P. (2016), Web Application Teaching Tools for Statistics Using R and Shiny. *Technology Innovations in Statistics Education*, 9(1), 1–33.
- Fawcett, L. (2018). Using interactive shiny applications to facilitate research-informed learning and teaching. *Journal of Statistics Education*, 26(1), 2-16.
- Garfield, J. (1993). Teaching Statistics Using Small-Group Cooperative Learning. *Journal of Statistics Education*, 1(1).
- Koutoumanou, E. & Wade, A. (2017). Students Generating and Using Their Own Data in a 5-day Basic Statistics Course. *United States Conference on Teaching Statistics (USCOTS, May, 2017)*.
- Lee, C., & Famoye, F. (2006). Teaching statistics using a real time online database created by students. *Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS7, July, 2006)*.
- Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education*, 6(3).
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International statistical review*, 67(3), 223-248.
- Zacharopoulou, H. (2006). Two learning activities for a large introductory statistics class. *Journal of Statistics Education*, 14(1).
- Zeleke, A., & Lee, C. (2010). Teaching introductory statistics using student generated data in a large class. *Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010)*.
- Zetterqvist, L. (1997). Statistics for chemistry students: how to make a statistics course useful by focusing on applications. *Journal of Statistics Education*, 5(1).