Engaging Students in a Large Lecture: An Experiment using Sudoku Puzzles

Caroline Brophy
National University of Ireland Maynooth

Lukas Hahn
Ulm University

Journal of Statistics Education Volume 22, Number 1 (2014),
www.amstat.org/publications/jse/v22n1/brophy.pdf

Key Words: Descriptive statistics; Contingency tables; Logistic regression; ANOVA; Survival analysis; Active learning.

Abstract

In this paper, we describe an in-class experiment that is easy to implement with large groups of students. The experiment takes approximately 15-20 minutes to run and involves each student completing one of four types of Sudoku puzzles and recording the time it takes to completion. The resulting data set can be used as a teaching tool at an introductory level right through to an advanced level of statistics. Basic methods for describing and displaying data as well as the intricacies that arise with real data may be discussed in an introductory course. The range of more sophisticated analyses that can be taught with the data set include chi-squared tests for independence, ANOVA, t- and F-tests, logistic regression and survival analysis. We describe and provide the tools to implement the experiment and illustrate several potential teaching topics using a collected data set.

1. Introduction

Involving students with in-class activities is a pedagogical method intended to promote active learning. The extent to which this can help to develop statistical reasoning skills and improve grades is still under discussion; however, it is agreed that such teaching techniques can be useful to improve student attendance and engagement at lectures (Snee 1993, Pfaff and Weinberg

2009). Involving students in data collection should naturally pique their interest in the topic of the data from early on in the process (Cummiskey, Kuiper and Sturvdant 2012).

The data sets typically used in an introductory level statistics course are either taken from textbooks, or information such as height, shoe size and gender is recorded via an in-class survey. While it is easy to collect such personal information, it is mostly only useful for teaching graphical or tabular display methods. Even with an introductory level course it is preferable to introduce students to the world of how statistics can address questions and test hypotheses. Conducting an in-class experiment, where the students can identify research questions and suggest ways to address the questions, should enhance the learning opportunities for the students. Moreover, in our experience, students are more likely to be interested in finding results concerning an experiment they participated in rather than a presentation of data relating to their personal characteristics. However, in-class experiments can be challenging due to lecture time constraints and the logistics of dealing with large groups.

In this paper, we describe an experiment that can be conducted easily in around 15-20 minutes in large classes. The experiment involves each student completing one of four types of Sudoku puzzles and recording the time it takes to complete it. The types of Sudoku puzzles differ only in the characters required to be filled into each grid. The types are numbers, letters, Greek letters and random symbols. Thus the two main hypotheses that can be tested are: 1. Does the type of Sudoku puzzle affect ability to complete it correctly? 2. Does the type of Sudoku puzzle affect the time it takes to complete it?

2. Experiment Description

The factor manipulated in the experiment is the type of Sudoku. There are four levels of Sudoku type denoted Greek, letter, number and symbol, which are each the same puzzle but with different characters (Table 1, Figure 1). A correctly completed Sudoku puzzle will contain all six characters in each row, each column and each box of six cells (Appendix A). Sudoku puzzles are typically a 9 x 9 grid but here mini Sudoku puzzles, with a 6 x 6 grid, are used so that the puzzles can be completed in a short timeframe.

Table 1. The characters for each Sudoku type.

Sudoku type	Characters					
Greek	α	β	δ	ε	λ	μ
Letter	a	b	c	d	e	f
Number	1	2	3	4	5	6
Symbol	■	△	✔	✕	⊖	☺

To prepare for the experiment, the worksheets provided in Appendix B are printed. Each worksheet contains one of the four types of Sudoku puzzles, instructions on how to complete the puzzle, a space for recording puzzle completion time, and an additional question to find out if the student has played Sudoku before or not. It is recommended to print an equal number of each type of worksheet to have approximately equal replication at each factor level and to interleave the worksheets to facilitate a randomisation process in-class.

It is suggested to conduct the experiment during the first lecture of the semester and under exam-like conditions (i.e., silence and no conferring). The bundle of worksheets is passed around and students are told to take a worksheet and keep it face down. Since the types of Sudokus are interleaved in the bundle, no two students sitting beside each other should be attempting the same puzzle. A clock is placed on a screen where everyone can see (e.g., www.online-stopwatch.com/full-screen-stopwatch). Before students are allowed to commence, they are told to

1. read the instructions before commencing with the puzzle,
2. record the time on the clock when the puzzle is completed,
3. record their time even if they give up without completing the puzzle or know they have not completed the puzzle correctly,
4. remain silent when they have completed their puzzle to allow others who have received a more difficult puzzle sufficient time to complete theirs and
5. answer the question below the puzzle after their time to completion is recorded.

Figure 1. The four Sudoku puzzles. A) Greek, B) Letter, C) Number, D) Symbol.

A

β			μ		ε
		ε	λ		
λ	β	μ	α	ε	
	δ	α	β	λ	μ
		λ	ε		
α		β			λ

B

b			f		d
		d	e		
e	b	f	a	d	
	c	a	b	e	f
		e	d		
a		b			e

C

2			6		4
		4	5		
5	2	6	1	4	
	3	1	2	5	6
		5	4		
1		2			5

D

△			☺		×
		×	⊖		
⊖	△	☺	■	×	
	✔	■	△	⊖	☺
		⊖	×		
■		△			⊖

The lecturer should additionally emphasise that all students hand in their solutions anonymously and that the degree of difficulty varies between different puzzles. This should prevent disheartened students from entering a lower time than was needed when observing that other students finished more quickly than themselves. Students may commence when the clock is started. Towards the end of the experiment, the lecturer can begin collecting completed worksheets. The entire in-class process should take approximately 15-20 minutes in total for a group of size 200-300 and will vary for smaller / larger classes.

After the in-class experiment, the puzzles need to be corrected (each one takes a matter of seconds and is facilitated by the colour solutions in Appendix A) and recorded manually. The data variables recorded are the type of Sudoku played (Sudoku type: Greek, letter, number, symbol), whether or not the Sudoku was correct (correct: yes, no), the length of time it took to complete the Sudoku (time: in seconds) and whether or not the student has played Sudoku before (Sudoku experience: yes, no). While this data entry is tedious, the end result is a rich and valuable data set that can be used continually for the rest of the semester.

The research questions that can be addressed using the data set include:

1. Does Sudoku type affect ability to get the Sudoku correct?
2. Does Sudoku experience affect ability to get the Sudoku correct?
3. Does Sudoku type affect the time it takes to complete the Sudoku?
4. Does Sudoku experience affect the time it takes to complete the Sudoku?

A wide range of statistical methods can be used to address these questions, some of which are outlined in Section 3.

3. Teaching Applications Using the Data from the Experiment

The experiment was conducted in the first lecture of an introductory statistics class of 276 first year undergraduate students at the National University of Ireland Maynooth (NUIM) in February 2012. The data set is available here, along with a documentation file. In this section, we use this data to illustrate some of the statistical techniques that can be taught using the experiment. The possible analyses range from basic tabular and graphical descriptions (introductory level) to ANOVA and logistic regression models (intermediary level) to survival analysis methods for dealing with censored data (advanced level).

3.1 Descriptive Statistics

3.1.1 Types of Data

The data set allows for a rich in-class discussion on types of data. The categorical data variables are Sudoku type, correct and Sudoku experience, and time is a quantitative data variable. However, time is a censored quantitative data variable, since, in the case of an incorrect Sudoku, we only know that the time it would have taken the student to complete it correctly is longer than what was recorded. With an introductory or intermediary group this facilitates discussions on how complications with real data can arise. The analysis in an introductory level class will be on the subset of data with correct Sudokus only (Sections 3.3 and 3.4), although it is important to raise the students' awareness for the limitations in results and inference on any analysis of this subset of data. If censored data techniques are outside the scope of the course, at the end of the semester, the lecturer should stress again the potential bias effects of omitting such data and may refer to subsequent statistics courses which provide tools for more detailed analyses. With a more advanced group, the in-class discussion can lead more quickly to analyses that incorporate the right-censored structure of the data (Section 3.6).

The practical aspects of conducting experiments and analysing real data can also be discussed. For example, if a student chose not to do the experiment at all this would result in missing data or if a student recorded a lower time than they actually did then this data point would be invalid. These discussion points can be useful to introduce students to the idea of conducting experiments that balance quality with practicalities, without compromising the inferential ability of the study.

3.1.2 Categorical Data

The variables correct (whether or not the student got the Sudoku correct) and Sudoku type (Greek, letter, number, symbol) are examples of categorical data and can be displayed using various methods. Graphical techniques include a bar chart displaying the distribution of correct (Figure 2A) and a segmented bar chart displaying the conditional distribution of correct for each Sudoku type (Figure 2B). Tabular techniques include a frequency and relative frequency table showing the distribution of Sudoku type, which illustrates the experimental design aim to have approximately equal replication for each Sudoku type (Table 2). A range of other graphical and tabular methods for displaying single or joint categorical data variables can also be taught.

Figure 2. A) The distribution of correct. B) The conditional distribution of correct for each level of Sudoku type.

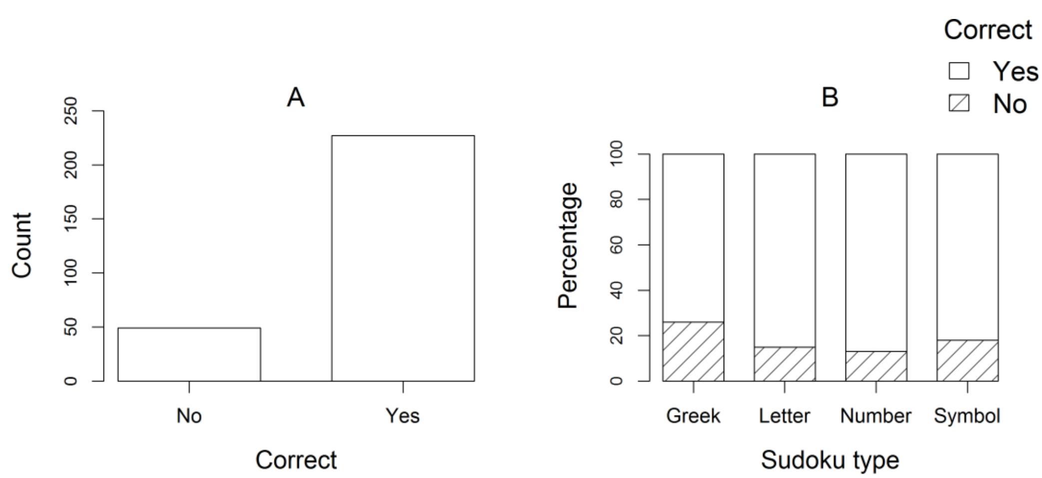


Table 2. The distribution of Sudoku type.

Sudoku type	Frequency	Relative frequency %
Greek	70	25.4
Letter	68	24.6
Number	70	25.4
Symbol	68	24.6
Total	276	100.0

3.1.3 Continuous data

Using the subset of data with correct Sudokus only, a histogram (Figure 3) and box plots (Figure 4) for the quantitative variable time can be demonstrated. Summary statistics can be illustrated for each Sudoku type and for all types combined (Table 3). The distribution of time (overall and within each Sudoku type) was skewed to the right (Figures 3 and 4, comparison of the mean and median in Table 3).

Figure 3. Histogram of time for correct Sudokus.

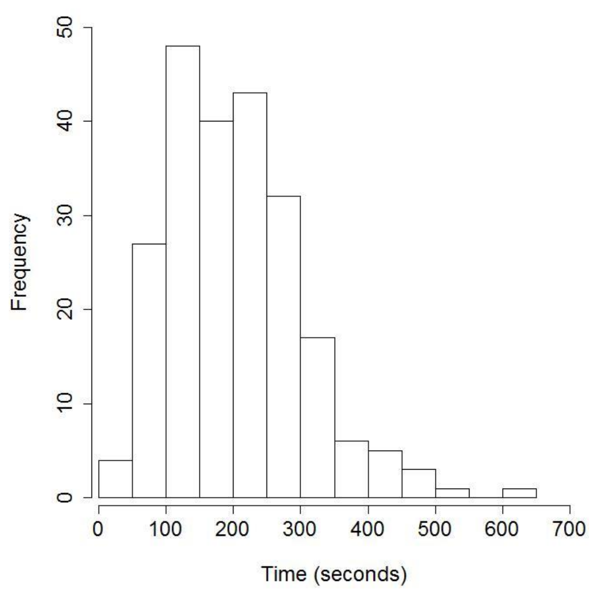


Figure 4. Box plots of time by Sudoku type for correct Sudokus.

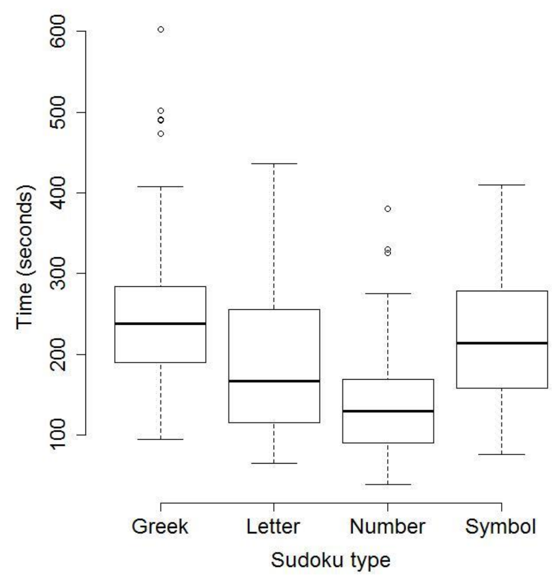


Table 3. Summary statistics for time (seconds) to completion of correct Sudokus for each Sudoku type and overall.

	Sudoku type				Types
	Greek	Letter	Number	Symbol	combined
Minimum	95.0	65.0	39.0	76.0	39.0
Lower quartile	190.5	116.0	90.0	159.0	125.0
Median	238.0	166.5	130.0	214.0	191.0
Mean	256.6	190.4	145.9	226.9	202.6
Upper quartile	281.5	252.0	169.0	277.5	263.5
Maximum	602.0	436.0	380.0	410.0	602.0
Interquartile range	91.0	136.0	79.0	118.5	138.5
Standard deviation	106.6	89.0	77.1	87.2	98.6

3.2 Chi-square test for independence between two categorical variables

A chi-square test for independence showed no evidence of a dependence between correct and Sudoku type (χ^2=4.62, df=3, p=0.2), despite the type Greek appearing to have a lower percentage of correct Sudokus than the other types (Table 4). There was however strong evidence of a dependence between correct and Sudoku experience (χ^2=43.48, df=1, p<0.001), with a much higher success rate for those with experience over those without (Table 4).

Table 4. Contingency table for the counts of correct by Sudoku type and by Sudoku experience.

	Sudoku type				Sudoku experience		
Correct	Greek	Letter	Number	Symbol	No	Yes	Total
No	18	10	9	12	25	24	49
Yes	52	58	61	56	25	202	227
Total	70	68	70	68	50	226	276
Percent Correct	74%	85%	87%	82%	50%	89%	82%

3.3 Two-sample t-test

Conditioning on the Sudoku being correct (i.e., using only the subset of data with correct Sudokus), the Sudoku types number and letter were classified as familiar and the remaining categories, Greek and symbol, as non-familiar. The logic for this classification is that students will be familiar with the numbers and letters but may not be as familiar with the Greek letters and the symbols; thus checking through and writing out the list of characters may take longer for the non-familiar group compared to the familiar group. A two sample t-test detected a significant difference in mean length of time to successful completion for the familiar versus the non-familiar groups (t=-6.05, df=225, p<0.001). With sample sizes of 119 and 108 observations respectively, approximate normality of the test statistic followed from the Central Limit Theorem. An F-test did not show evidence against the additional assumption of equal variances (F=0.78, df=118, 107, p=0.166). This F-test may however be misleading due to its strong sensitivity towards the skewed nature of the time data, but the Welch two-sample t-test, which allows for distinct variances, concluded the same as the standard t-test.

3.4 One-way ANOVA

Still conditioning on correct Sudokus only, there was evidence of an effect of Sudoku type on the length of time to Sudoku completion (Table 5). Tukey's adjustment for multiple comparisons showed that the type number differed significantly from all other Sudoku types and that the type Greek also differed from letter. Other comparisons, however, revealed no more significant results (Table 6). A normal probability plot (not shown) of the residuals from this one-way ANOVA model showed evidence of non-normality which was due to the skewed nature of the times within each group (Figure 4). Hence this data is useful for exploring model diagnostic tests and possible solutions. Two-way ANOVA, with factors Sudoku experience and Sudoku type, with and without their interaction, could also be illustrated with this data.

Table 5. One-way ANOVA table for the response time to completion of correct Sudokus for the factor Sudoku type.

	Df	SS	MS	F	P-value
Sudoku type	3	389466	129822	16.03	<0.001
Error	223	1805800	8098		
Total	226	2195266			

Table 6. Mean length of time to completion (seconds) for each Sudoku type.

	Sudoku type		
Number	Letter	Symbol	Greek
146	190	227	257

Means are computed using only correct Sudokus. Sudoku types that share a common bolded line do not differ significantly ($\alpha = 0.05$), tested using Tukey's adjustment for multiple comparisons.

3.5 Logistic regression

Logistic regression models (Agresti 2002) are appropriate when the dependent variable is dichotomous. In our case, we examine which variables are useful predictors of the response Sudoku correct. We can use the fitted model to jointly estimate the effects of Sudoku type and Sudoku experience on the odds and the probability of completing Sudoku correctly. A range of logistic regression models were fitted and likelihood ratio tests were used to compare models. Sudoku experience was the only significant predictor of successfully completing Sudoku (Table 7). The odds of successfully completing the Sudoku correctly were 8.4 times higher for those who had played Sudoku before relative to those who had not played before (95% CI: (4.2, 16.9)). The choice of final model (Model 4) agrees with the chi-square tests for independence in Section 3.2 and additionally shows no evidence of an interaction between correct and Sudoku type on success probability (Table 7).

Table 7. Comparisons of the fitted logistic regression models.

Model #	Terms	-2LL	Comparison	Term tested	LRT	df	P-value
1	ST SE ST*SE	213.5					
2	ST SE	217.5	M1 v M2	ST*SE	4.0	3	0.262
3	ST	253.7	M2 v M3	SE	36.1	1	<0.001
4	SE	222.3	M2 v M4	ST	4.8	3	0.189
5	Intercept	258.1	M4 v M5	SE	35.8	1	<0.001

-2 Log Likelihood (-2LL) values are shown for each logistic regression model and comparisons are made using likelihood ratio tests. ST = Sudoku type, SE = Sudoku experience, LRT = likelihood ratio test statistic and df = degrees of freedom.

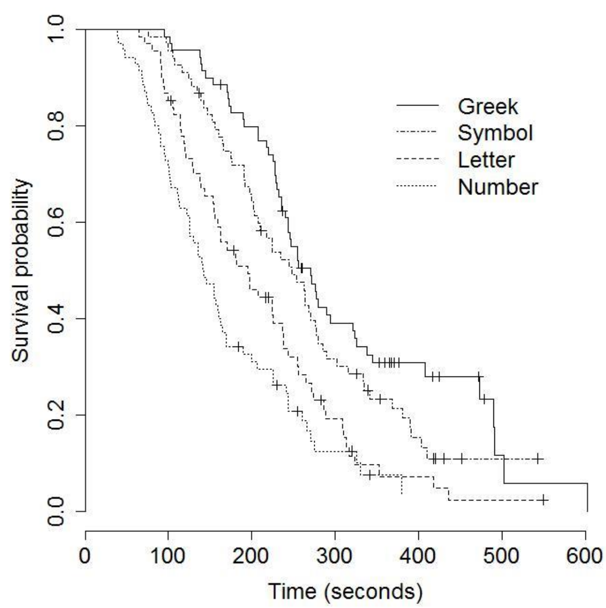
3.6 Survival Analysis for Censored Data

In the analyses of Sections 3.3 and 3.4, the recorded times for incorrect Sudokus were ignored. Hence, statistical inference from these results is only valid conditioning on correct puzzle takers and cannot be generalised to all puzzle takers. Due to relatively large time values for many incorrect Sudokus, the mean and median times to completion for each Sudoku type (Tables 3 and 6) are likely to be underestimated. Time entries on worksheets with incorrect solutions contain information on the quantity of interest since they give lower bounds for true completion times. The continuous variable time is therefore referred to as right-censored. Survival analysis techniques (Kalbfleisch and Prentice 2002) can use both censored and non-censored data to test for differences in true completion times among the Sudoku types. Putting survival analysis terminology in the context of our experiment gives the following: ‘survival’ indicates the event that correct completion of the puzzle has not happened, ‘death’ indicates the event that a person has correctly completed the puzzle and a right-censored data point corresponds to the time of a person who has incorrectly completed the puzzle. The non-negative random variable T denotes the time to death, i.e. correct completion.

The Kaplan-Meier estimator (Kaplan and Meier 1958) estimates the survival function, $S(t) = P(T \geq t)$ for time t. Using this method, Figure 5 shows the estimated survival probabilities (or probabilities of correct completion having not yet occurred) (y-axis) versus time (x-axis) for each Sudoku type. Censored data points are indicated by a plus symbol and the probability of completing the Sudoku puzzle correctly by time t is estimated by $1-\widehat{S(t)}$ (where ^ indicates estimated). Conditioning on each Sudoku type, the Kaplan-Meier method estimated that the median times to correct completion (Figure 5, x-values at y=0.5) were 142 seconds for the type number, 195 for letter, 248 for symbol and 270 for Greek. As expected, these unbiased median estimates are all higher than the biased medians presented for the subset of correct Sudokus only (Table 3). The log-rank test (Mantel 1966, Peto and Peto 1972) is a standard method for group comparisons in survival analysis. It showed significant differences among the survival functions ($p<0.001$).

In addition to the non-parametric Kaplan-Meier approach, survival functions for each Sudoku type can be estimated via the semi-parametric Cox's proportional hazards model (Cox 1972). This regression method measures the impact of Sudoku type on the hazard function, $\lambda(t; x) = \lim_{h \to 0+} P(t \leq T < t + h \mid T \geq t, x)/h$ for time t and covariates x, and where hazard functions for different x settings are assumed to be proportional to each other. Significant differences ($p<0.05$) in hazard functions were found between each pair of Sudoku types except for Greek and symbol. The estimated survival functions using this method (not shown) gave similar trends to the estimated Kaplan-Meier survival functions in Figure 5. A plot of log(-log($\widehat{S(t)}$) survival function estimates)) versus time t (not shown) for each Sudoku type showed a somewhat constant difference between the curves suggesting that the proportionality assumption was reliable. Full parametric regression models could also be investigated for these data, as could further goodness of fit tests for Cox's proportional hazards model.

Figure 5. Kaplan-Meier curves for estimated survival probability versus time for each Sudoku type.



4. Discussion

This paper describes an in-class experiment and illustrates many statistical analyses on the resulting data, ranging from introductory to advanced techniques. The lecturer only needs a short period of time to prepare for and perform the experiment. Ideally, it is run in the first lecture of the term and the data are revisited throughout the semester, making different examples from textbooks or dry demographic analyses obsolete. The range of topics taught using the experimental data can be modified depending on the scope of the course. Since the students are automatically part of the study, it is the intention that their participation will have positive impacts on engagement, learning ability and class environment.

The Sudoku-based activity shows similarity to the in-class experiment using online versions of the puzzle Tangram developed by Cummiskey et al. (2012). Similar to the Tangram study, with the Sudoku study, students partake in the experiment and in discussions on possible research questions, although in our study students do not have any input to the design of the experiment. The Tangram study is suited to a group that already has some statistical knowledge and can form the basis of laboratory practical classes, while our Sudoku study is more suited to the first lecture of an elementary statistics class.

Our Sudoku experiment has some limitations. For an introductory or intermediate group of students, the interpretation of some of the results may become problematic when survival analysis is not part of the curriculum. The censored structure of the variable time could be misunderstood such that students interpret results from Sections 3.3 and 3.4, which condition on having correctly completed the puzzle, to hold for all puzzle takers (correct and incorrect). Additionally, students could falsely conclude that it was acceptable to omit problematic or complex data. We recommend that this topic should be discussed intensively throughout the course and that the bias in the ANOVA analysis due to omitting the censored data should be explained in full detail.

Another limitation to the experiment arises when the number of participating students is smaller than was the case in this paper. Fundamental analyses, as in Sections 3.1 to 3.3, may not suffer from smaller sample sizes, but ANOVA, logistic regression, and survival analysis may be less robust. We recommend running the experiment in courses of class size of somewhat similar magnitude to ours (illustrations in this paper are from a class of 276 students). In case of a substantially lower number of students, it is also possible to drop the advanced techniques and use only the more basic analyses. Alternatively, a simpler design which compared only two types of Sudoku puzzles (e.g., only numbers and Greek letters) could be employed with a smaller group. A discussion on sample size issues could also naturally lead to the introduction of the topic of power calculations and whether or not the designed study had sufficient power to address the hypotheses being tested.

In addition to the advantage of generating a rich dataset, involving the students in the data collection process should positively affect the students’ interest in statistics (Snee 1993, Pfaff and Weinberg 2009). However, we had no scientific tool to test this effect on the participants. Although many students commented on their end of course evaluation that they enjoyed the fact that data collected on them was used in class, we did not quantify how well the experiment engaged students in participating or how it influenced learning ability. In particular, we did not have any possibility to test for whether or not the increase in positive attitude improved understanding or end of year grades. The opportunity to test these hypotheses may arise in the future. For example, if a particularly large class was split into two parallel classes, the experiment could be conducted with one group only. Using the second lecture as control, differences in learning environment and grades could be assessed. However, for now, we can only use the information provided by students on their end of year evaluation to judge the success of using the experiment. In this paper, we only illustrate the data from a single implementation of the experiment, but the experiment has been conducted with several different classes at the National University of Ireland Maynooth. The feedback has been consistently positive, indicating that the intention to further engage students in the course has been at least somewhat achieved with the inclusion of the experiment.

The collected data set could be made richer by asking additional student information questions on the worksheets. In the worksheets we provide, only Sudoku experience, i.e., whether or not the student had ever played Sudoku before, was enquired. Other variables such as gender, academic measurements or general interest in logic puzzles could also be recorded. This would expand the scope of the hypotheses that could be tested with the study and allow for additional in-class discussions such as confounding of or correlation among explanatory variables used in modelling.

Since it is crucial for the data quality that all students take the time they need, a silent work environment during the experiment is essential. In our setup, the students who had already finished their Sudoku had to wait for the others to complete theirs. These students may be impatient and start talking or disturbing the others. We suggest providing an article, for example on a statistics topic or the history of Sudokus, as reading material for those who finished early; the article could be located on-screen beside the clock.

The in-class experiment allows the opportunity for students to see for themselves some of the practical issues that can arise when carrying out research studies. For example, in this experiment, any individual student may not have read the instructions before commencing, may not have bothered recording their time if they knew their Sudoku was incorrect or may have (accidentally or on purpose) recorded a different time than what they actually completed the puzzle in. In the history of running the experiment at NUIM, in the lecture following the experiment, students have been asked if they were honest in the recording of their time and no student has so far admitted to lying on their time. Of course it is impossible to know if this is true. If a student did not record their time the missing value could be removed from the data set. Drawbacks of such complete-case analyses should be pointed out and more sophisticated techniques such as inverse probability weighting or data imputation techniques (Little and Rubin 2002) could be introduced. This can easily lead into a discussion of how it is necessary, when designing any experiment, to consider the ability to address the research questions along with potential practical, time or financial constraints and how to deal with peculiarities that may arise.

With a class size of 276 students, the in-class experiment took 15-20 minutes. We believe that time consumption will not change significantly for groups between 200 and 300 students. Larger groups would need more time due to the spreading and collecting of the additional worksheets, although the duration of puzzle completion part of the process should not change. With a significantly larger group, it might be useful to have a class assistant to help or to ask some students in the class to assist. We recommend conducting the experiment in the first lecture, since the experiment should be fun for the students and thereby giving a nice introduction to the course. Also, 20 minutes of the first lecture, where often organisational issues dominate, should be a reasonable investment considering the rich analyses that can be carried out afterwards. After the class, the data are recorded manually which is time-consuming and for our sample size took approximately two hours. An online version of the puzzle, as in Cummiskey et al. (2012), would be helpful because the correction of the Sudokus and data recording would be done automatically. However, implementing such an electronic tool was not the intention of this paper. Moreover, the easiness of the experiment in paper format avoids additional time-consuming organisation of computer labs, or laptops, and instructions for students on how to use the software.

Overall, the experiment has proven through experience to be easily run with large groups, can illustrate real testable hypotheses and can facilitate the teaching of a wide range of statistical analyses. The study can form the basis for many discussions on types of data and the complexities that can arise with real data sets. It can also provide the opportunity to introduce new statistics students to the benefits of having statistical tools to address real world questions.

Appendix A
Solutions to the four Sudoku puzzles

β	λ	δ	μ	α	ε
μ	α	ε	λ	δ	β
λ	β	μ	α	ε	δ
ε	δ	α	β	λ	μ
δ	μ	λ	ε	β	α
α	ε	β	δ	μ	λ

b	e	c	f	a	d
f	a	d	e	c	b
e	b	f	a	d	c
d	c	a	b	e	f
c	f	e	d	b	a
a	d	b	c	f	e

2	5	3	6	1	4
6	1	4	5	3	2
5	2	6	1	4	3
4	3	1	2	5	6
3	6	5	4	2	1
1	4	2	3	6	5

△	⊖	✓	☺	■	✕
☺	■	✕	⊖	✓	△
⊖	△	☺	■	✕	✓
✕	✓	■	△	⊖	☺
✓	☺	⊖	✕	△	■
■	✕	△	✓	☺	⊖

Appendix B
Experiment worksheets

B.1 Worksheet 1: Sudoku Experiment

Instructions:

The Greek letters α, β, δ, ε, λ, μ must each appear once in each box, once in each row and once in each column. Use logic (i.e. do not guess) to determine what goes in each empty space.

β			μ		ε
		ε	λ		
λ	β	μ	α	ε	
	δ	α	β	λ	μ
		λ	ε		
α		β			λ

Time to completion: Minutes:_________ Seconds:_________

When you have completed the puzzle, please answer the following question:

Have you ever played Sudoku before today? Yes ☐ No ☐

B.2 Worksheet 2: Sudoku Experiment

Instructions:

The letters a, b, c, d, e, f must each appear once in each box, once in each row and once in each column. Use logic (i.e. do not guess) to determine what goes in each empty space.

b			f		d
		d	e		
e	b	f	a	d	
	c	a	b	e	f
		e	d		
a		b			e

Time to completion: Minutes:_________ Seconds:_________

When you have completed the puzzle, please answer the following question:

Have you ever played Sudoku before today? Yes ☐ No ☐

B.3 Worksheet 3: Sudoku Experiment

Instructions:

The numbers 1, 2, 3, 4, 5, 6 must each appear once in each box, once in each row and once in each column. Use logic (i.e. do not guess) to determine what goes in each empty space.

2			6		4
		4	5		
5	2	6	1	4	
	3	1	2	5	6
		5	4		
1		2			5

Time to completion: Minutes:_________ Seconds:_________

When you have completed the puzzle, please answer the following question:

Have you ever played Sudoku before today? Yes ☐ No ☐

B.4 Worksheet 4: Sudoku Experiment

Instructions:

The symbols ■, △, ✔, ✕, ⊖, ☺ must each appear once in each box, once in each row and once in each column. Use logic (i.e. do not guess) to determine what goes in each empty space.

△			☺		✕
		✕	⊖		
⊖	△	☺	■	✕	
	✔	■	△	⊖	☺
		⊖	✕		
■		△			⊖

Time to completion: Minutes:_________ Seconds:_________

When you have completed the puzzle, please answer the following question:

Have you ever played Sudoku before today? Yes ☐ No ☐

Acknowledgements

This work has partly been performed while Lukas Hahn visited the National University of Ireland Maynooth. The warm hospitality of the Department of Mathematics and Statistics at NUIM is gratefully acknowledged. Lukas Hahn was supported by a Talanx Stiftung (Talanx foundation) scholarship.

References

Agresti, A. (2002), *Categorical Data Analysis* (Second ed.), New Jersey: John Wiley & Sons, Inc.

Cox, D. R. (1972), "Regression Models and Life-Tables," Journal of the Royal Statistical Society. Series B (Methodological), 34, 187-220. (www.jstor.org/stable/2985181)

Cummiskey, K., Kuiper, S., and Sturvdant, R. (2012), "Using Classroom Data to Teach Students About Data Cleaning and Testing Assumptions," Frontiers in Psychology, 3, Article 354. (www.ncbi.nlm.nih.gov/pmc/articles/PMC3457080/)

Kalbfleisch, J. D., and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data* (Second ed.), New Jersey: John Wiley & Sons, Inc.

Kaplan, E. L., and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, 53, 457-481. (http://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501452)

Little, R. J. A., and Rubin, D. B. (2002), *Statistical Analysis with Missing Data* (Second ed.), New Jersey: John Wiley & Sons. Inc.

Mantel, N. (1966), "Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration," Cancer Chemother Rep, 50, 163-170. (http://ci.nii.ac.jp/naid/10004996110/en/)

Peto, R., and Peto, J. (1972), "Asymptotically Efficient Rank Invariant Test Procedures," Journal of the Royal Statistical Society. Series A (General), 135, 185-207. (http://www.jstor.org/stable/2344317)

Pfaff, T. J., and Weinberg, A. (2009), "Do Hands on Activities Increase Student Understanding?: A Case Study," Journal of Statistics Education, [Online], 17(3). (www.amstat.org/publications/jse/v17n3/pfaff.html)

Snee, R. D. (1993), "What's Missing in Statistical Education?," The American Statistician, 47, 149-154. (www.tandfonline.com/doi/abs/10.1080/00031305.1993.10475964)

Caroline Brophy
Department of Mathematics and Statistics
National University of Ireland Maynooth
Maynooth
Co Kildare
Ireland.
caroline.brophy@nuim.ie
+353 1 7083914

Lukas Hahn
Faculty of Mathematics and Economics
Ulm University
D-89069 Ulm
Germany.
lukas-1.hahn@uni-ulm.de
