REAL-LIFE MODULE STATISTICS: A HAPPY HARVARD EXPERIMENT

Kari Lock and Xiao-Li Meng
Statistics Department, Harvard University, United States of America
lock@stat.harvard.edu

Five years ago, a discussion ensued over wine about how to make learning statistics a "happy" experience. This turned into many discussions over dinners and wine, and the formation of the "happy team": a team of faculty and grad students dedicated to creating the course "Real-Life Statistics: Your Chance for Happiness (or Misery)". The course is module based, featuring modules such as "Romance", "Wine and Chocolate", "Finance", "Medical", and more. We've taught this course at Harvard three times; twice as a second level course and once with no prerequisites. Here we discuss the team approach to creating a course, the module approach to teaching statistics, and the happiness (and misery) involved both for us and our students.

THE HAPPY TEAM

It all started over wine, with a conversation between a graduate student (Wee Lee Loh) and a faculty member (Xiao-Li Meng): "Since statistics is applicable to almost anything, why not teach a class applying statistics to wine?" Wine makes people happy, so surely students will enjoy a class on the statistics of wine! However, chocolate also makes people happy, as does falling in love, maintaining good health, and getting rich. Rather than teaching statistics as a series of techniques, why not teach it as a "life skill" for increasing your happiness? This idea blossomed into "Real-Life Statistics: Your Chance for Happiness (or Misery)".

An idea blossoming into a course is no trivial matter, and wouldn't have happened without the happy team. The team started with Wee Lee and Xiao-Li, who recruited a number of happy energetic Ph.D and Master's students. Excited by the opportunity to create this unique new course, all invitees enthusiastically joined. Other than Xiao-Li, the current team comprises six statistics Ph.D. candidates at Harvard (Simeng Han, Jonathan Hennessy, Bo Jiang, Victoria Liublinska, Kari Lock, Cassandra Wolos Pattanayak); six alumni (Paul Baines, Yves Chretien, Paul Edlefsen, Wee lee Loh, Reetu Kumra Mundi, Linjuan Qian) still help out in various ways. The purpose of the happy team is to create a course that conveys the power and wide applicability of statistics in a way students will find interesting and engaging, and of course to bring happiness to both our students and ourselves.

We have roughly monthly dinner meetings, each beginning with toasts in which each member shares what he/she is currently happy about. With everyone in good spirits, we then proceed to lively discussions about the class and pedagogy in general. For example, someone mentions that we need an engaging first-day activity to illustrate the general idea of statistical inference. We discuss many fun possibilities and decide to use chocolate and wine (it is the "happy course," after all!). Many chocolates will be made; a secret proportion will be filled with wine. Each student tastes a chocolate and reports "wine" or "no wine" via clickers. Students then predict (point estimate and interval) the true proportion of chocolates that contain wine, and by increasing the sample size—an activity well welcomed by students—we demonstrate how uncertainties in statistical estimators decrease accordingly. Aided by solid and liquefied food and many enthusiastic minds working together, these meetings generate creative new ideas for the course and happiness for the team members.

Outside of the dinner meetings, the team meets in smaller groups to work on all the minute details associated in creating a new course: finding data, creating lecture slides, writing problem sets, and creating group projects to assign to the students. Both the creative process and the hard work become much more manageable and enjoyable with a devoted team working together.

MODULE-BASED APPROACH

The course is built around five modules, based on real-life topics, with approximately two weeks per module. So far we have built six modules: Romance, Finance, Medical, Election, Legal, Wine and Chocolate. We plan to build more to better fit current interests; e.g., we added the Election module in association with the 2008 US Presidential Election. We've taught this course as

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

both a second level course (with a prerequisite of any previous statistics course), and as a general education course assuming no prior knowledge. The statistical content differs between the two versions, but the real-life topics remain the same.

The romance module begins by asking, "Where should you take someone on a first date?", covering the basics of questionnaire and survey designs, with an emphasis on data quality in statistical analysis. The module proceeds to online dating, discussing how to ask questions to generate successful matches, and interpreting the probability and odds of a successful match. Students see logistic regression (yes, before linear regression) for match prediction and variable selection with real speed-dating data. The speed-dating analysis assesses which variables are best at predicting initial attraction, something many students are curious about!

The finance module focuses on investing. Students learn about predicting stocks with topics such as expected return (expected value), volatility (variance), beta-coefficient (regression), not chasing hot stocks (regression towards the mean), and a bit of time series (for the second-level version). Students learn how correlation underlies the principle of portfolio diversification. Each module has a group project; in finance each group is given a virtual $10,000 to invest, and a custom-built course website tracks the value for the rest of the term. The team with the most money at the end wins a tour of the Harvard Management Company, and the team with the best interval prediction wins a dinner with the happy team. Students got extremely involved in this competition, checking and updating their portfolios daily! One student emailed us after the course: "I enjoyed the Finance module of Stat 105 and its two projects so much so that I have decided to study the maths behind finance in more depth."

The medical module was created primarily by Yves Chretien, pursuing an M.D. and statistics Ph.D. Clinical trials and hypothesis tests are illustrated with the Viagra trial, and observational studies with Fen-Phen. Students learn Simpson's Paradox by comparing two kidney stone treatments. The difference between association and causation is driven home with an article claiming that airports and casinos are safer than hospitals if you undergo cardiac arrest! For the medical group projects, students choose a published observational study, explain the results, report potential confounding factors, and design a hypothetical randomized experiment addressing the same issue, with a discussion on its pros and cons, including ethics issues.

In the legal module, we focus on several high-profile court cases (e.g., Sally Clark) that hinged on the fundamentals of probabilistic reasoning and teach students how to avoid the "Prosecutor's Fallacy." This module, taught only in the second-level version of the course, introduces Bayesian reasoning with prior and posterior odds, and shows how to incorporate new information to update probabilities.

The election module was added in the second year by popular demand from the students. It takes students through the design and analysis of a pre-election poll: identifying sources of bias, the principles of survey sampling, and the crucial role of uncertainty in evaluating polls. To analyze poll results, students learn about the Binomial distribution, the Normal approximation, confidence intervals, and margin of error. For their project, students conduct a poll on a topic of choice, analyze the results, and discuss potential sources of bias.

The final module, Wine and Chocolate, begins with a wine tasting for students of age. The tasting illustrates principles of experimental design, and the resulting data are analyzed in class using ANOVA. Besides demonstrating a factorial experiment and providing data, the tasting serves as a great way to mingle informally with the students. For the project, each group is given (real) money for chocolate, and designs and analyzes a chocolate tasting experiment; students loved this, especially those who were too young to taste wine!

PEDAGOGICAL INNOVATIONS

We attribute the success of this course not only to the dedication of the happy team and the emphasis on real-life topics, but also to several innovative features of the course.

One feature that contributed immensely to student involvement is the Personal Response System, i.e., clickers. Participation, based mostly on clicker participation, makes up 15% of each student's grade. This greatly encouraged students to come to class and remain actively engaged during class. Clicker questions were designed to help students think intuitively about a concept before it was taught, to review concepts learned in previous modules, and to assess whether

students were getting it. Clicker questions spurred vibrant classroom discussion, as clickers force each student to form his or her own opinion, which they are then usually willing to share and debate. Classroom participation is fantastic, with attendance always above 90%, and often 100%, despite the fact that all lecture videos and slides are posted on the course webpage immediately after each lecture.

To help emphasize the idea of "real-life statistics," the final lecture of each module was by a guest speaker, an expert in the module topic. For example, we had social psychologist David Kenny for Romance module, vice president of the Harvard Management Company Stephen Blyth for Finance, brothers Yves Chretien and medical researcher Jean-Paul Chretien, M.D., Ph.D. for medical, and Harvard law professor James Greiner for legal. Grand finale guest lecturers included New York Times columnist David Leonhardt speaking about the use and misuse of statistics, and the Netflix prize winner Robert Bell about their team work that led to the prize (before it was announced). All these lectures were extremely well received by students in and outside the course (as we advertise them generally), and showcase how statistics is interwoven into every fabric of our life.

Each module had a problem set to be done individually and a group project. The problem sets ensured that the students learned the needed statistical techniques. The projects enabled students to apply the concepts in a realistic setting. Each group chose one project to present to the class, either as a live presentation or as a mini-movie. Students put a great deal of effort into the projects, and the entire teaching team has had a blast enjoying the presentations and movies!

The prerequisite version of the class had no in-class exams, and only a small final exam worth 5%, distributed on the first day of class and due on the last day. This put the emphasis on learning the material, rather than learning for an exam. (The general education version of the class had an in-class final exam, as required by the General Education office.)

Xiao-Li, the course professor, holds a weekly open "happy hour" with food and drink. Students are encouraged to come and talk with him about anything besides homework, making for great informal intellectual discussion outside the classroom. Happy team members serve as course teaching fellows, offering office hours and additional weekly sections to smaller groups of students.

CHALLENGES & CONCLUSIONS

One obvious challenge in creating a new course is the amount of preparatory work required. This becomes greatly facilitated as a team effort. Our course was offered for the first time as a general education course last spring, so in the prior fall semester we offered a graduate level workshop course on statistical fallacies and paradoxes, with an emphasis on pedagogy and designing materials for a general education course. Students explored concepts such as Simpson's Paradox and regression towards the mean, with a final project to tie these into the existing module topics and produce slides making them understandable to a general audience. Besides the happy team, both graduate students and faculty members joined this workshop for its rich pedagogical discussions and a deepened understanding of fundamental statistical concepts.

The module approach made the class structure less flexible than is sometimes the case. Having a certain amount of material that needs to be covered in two weeks allows less spontaneity and flexibility during lectures. The module approach also led to statistical concepts not being covered in the customary order (i.e., logistic regression before linear regression). Whereas the modules could have been carefully designed to avoid this, we did not find it to be a detriment, but rather a freedom not present in typical courses. It demonstrated well that the course focus is on real-life applications, not finding examples for the sake of illustrating statistical techniques, as typical in many statistical courses.

Too many students leave statistics courses with the impression that statistics is dry or boring. With a subject as applicable as statistics, this should never be the case. By creating a course that applies statistics to topics relevant to the students' lives, we excite and engage students and open their eyes to the power of statistics. By bringing together a happy team working toward this common goal, we have made the creation of this course an enjoyable and rewarding experience, bringing happiness (and hopefully not misery!) to our students.

And finally, for your happy viewing, check out the course trailer at http://www.stat.harvard.edu/happy.

ACKNOWLEDGEMENTS

We thank our entire Happy Team for enriching our pedagogical life, Harvard's Provost's Office, Graduate Dean's Office, and Undergraduate Dean's Office for financial and moral support, and all our students for their full participation in this happy experiment! We also thank Yves Chretien for careful proofreading.