[2023] "Two truths and a lie" as a class-participation activity. American Statistician 77, 97--101.

"Two truths and a lie" as a class-participation activity*

Andrew Gelman†

20 Mar 2022

Abstract

We adapt the social game "Two truths and a lie" to a classroom setting to give an activity that introduces principles of statistical measurement, uncertainty, prediction, and calibration, while giving students an opportunity to meet each other. We discuss how this activity can be used in a range of different statistics courses.

1. Background

Class-participation activities are useful in many settings, but especially so at the beginning of the semester, where they can help build community and set up a norm of active student involvement. Here we describe an statistics-based activity that has the additional benefit of allowing students get to know each other on a social level.

There is a large literature on the benefits of active learning—classroom interactions that involve students doing things, talking with each other, and solving problems together; see, for example, Bligh (2000) regarding college teaching in general, and Cobb (1992), Magel (1998), Nolan and Speed (2000), and Rossman and Chance (2001) on statistics education more specifically. As Mazur and Watkins (2009) discuss in the context of physics instruction, active learning provides a "structured environment" that facilitates collaboration and awareness of learning strategies, and can work in small or large classrooms. It is recommended that each class-participation activity directly involves the students while being tied to a particular topic being covered in the course, and this has led to efforts such as Gnanadesikan et al. (1997) and Rossman and Chance (2008) to integrate active learning within introductory statistics courses. In the present article, we describe an activity that is designed to get students involved in class and with each other and which relates to several different areas of statistics. We designed and tested the activity in a class on applied regression for social scientists, but we think it should work in a wide variety of statistics courses.

2. The "Two truths and a lie" activity: Data gathering and analysis

This activity can be performed during the first week of class or later on during the semester if that seems to better fit with the sequence of topics in the course.

We start the activity by dividing students into groups of four—it's fine if some groups have three or five students in them—to play "two truths and a lie." We display the instructions in Figure 1 onto the screen and explain the procedure. In this game, one person makes three statements about him or herself; two of these statements should be true and one should be false. The other students in the group should then briefly confer and together guess which statement is the lie. They should jointly construct a numerical statement of their certainty about their guess, on a 0–10 scale, where 0 represents pure guessing and 10 corresponds to complete certainty. The true statement is then revealed so that the students know if they guessed correctly. Each group of students then rotates

*To appear in *The American Statistician*. We thank the students in our class for their participation and three anonymous reviewers for helpful comments.

†Department of Statistics and Department of Political Science, Columbia University, New York.

Within your group:

1. One person tells three personal statements, one of which is a lie.
2. Others discuss and guess which statement is the lie, and they jointly construct a numerical statement of their certainty in the guess (on a 0–10 scale).
3. The storyteller reveals which was the lie.
4. Enter the certainty number and the outcome (success or failure) and submit in the Google form.

Rotate through everyone in your group so that each person plays the storyteller role once.

Figure 1: *Instructions for the "two truths and a lie" activity, to project onto the screen for students.*

through, with each student playing the role of storyteller, so that when the activity is over, each group of four students has produced four certainty numbers, each corresponding to a success or failure. Figure 2 shows an example.

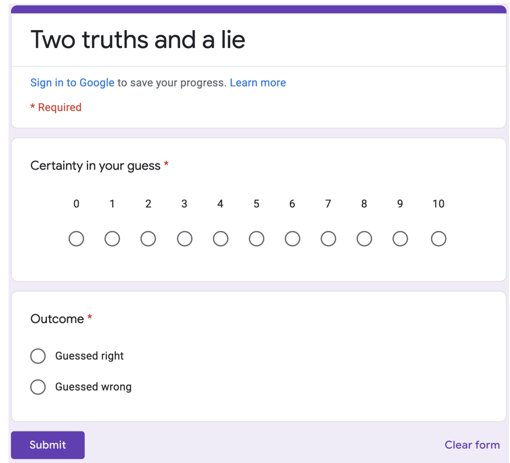
We then give students the url of a Google form where they can enter their data using their phone or laptop. The form is set up to take one response at a time, so each group should enter four responses corresponding to their four guesses. Alternatively we could set up a longer Google form allowing a group to enter all four responses together, but that would require additional data processing on the analysis end, so we go with this simpler approach that does not keep track of the clustering of the responses.

We download the data from the Google form as a csv file, read it into our statistical software, and announce that we will display the data (a scatterplot of the success/failure outcome vs. certainty score) along with a fitted curve showing probability of the guess being correct as a function of certainty score. If the class is sufficiently advanced, we explain that the fitted curve will be a logistic regression; otherwise we simply say we will fit a curve.

Before making the plot and displaying the data and fit, we we ask students in their groups to sketch what they think the scatterplot and fitted curve for the class will look like, and then we lead the class in discussion. Some possible prompts include: What do you think the range of certainty scores will look like: will there be any 0's or 10's? Will there be a positive relation between x and y: are guesses with higher certainty be more accurate, on average? How strong will the relation be between x and y: what will the curve look like? If students have seen logistic regression, we ask them to give approximate numerical values for the intercept and slope coefficients corresponding to their sketched curves.

After this discussion, we display the data and fitted curve and conduct a follow-up discussion of what has been learned. Figure 3 shows an example of real data from an applied regression class. In this case, there is essentially no relation between the certainty score and the outcome (coded as 1 for a successful guess and 0 for an error). In fact, the estimated logistic regression coefficient is negative: higher certainty scores correspond to slightly lower rates of accuracy! It's hard to see this in the plot of raw responses; in general, scatterplots are not so helpful for displaying discrete data. The standard error from the regression gives a sense of the uncertainty in the fit; in this case, the right panel of Figure 3 shows an estimated slope of -0.08 with standard error 0.15, indicating that

Certainty	Outcome
8	Success
4	Success
7	Failure
5	Success


Two truths and a lie

Sign in to Google to save your progress. Learn more

* Required

Certainty in your guess *

0 1 2 3 4 5 6 7 8 9 10

Outcome *

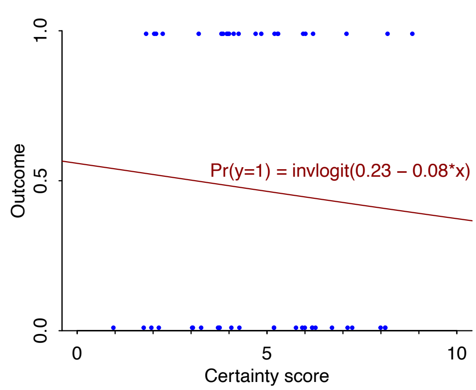
Guessed right

Guessed wrong

Submit

Clear form

Figure 2: *(a) Example of data produced by a group of four students playing the "two truths and a lie" game. Each number is a group consensus: here, the group of students B, C, D assigned a certainty of 8 to their guess of student A's lie, and they were correct; students A, C, D assigned a certainty of 4 to their guess of student B's lie, and they happened to be correct, and so on. (b) Students play the game in groups and then enter the data one at a time into a Google form, so that the instructor can analyze all the results together.*



Coefficient	Estimate (s.e.)
Intercept	0.23 (0.77)
Slope	−0.08 (0.15)

Figure 3: *(a) Scatterplot from the "two truths and a lie" activity performed in a class of 49 students, along with a curve showing a fitted model predicting correctness given the certainty scores (which have been jittered to avoid points overlapping on the graph); (b) Coefficient estimates and standard errors from the fitted logistic regression. The instructor can perform these steps in real time on the data that students entered in their Google forms.*

the sign of the underlying relationship is unclear from the data. For this class, the students' stated certainty scores do not form a useful predictor of their actual knowledge.

In the discussion that followed in our class, students conjectured why their guesses were so bad (23 our 49 correct, not much better than the 1/3 success rate that would come from random guessing) and why their certainty judgments were not predictive of their accuracy. In many ways, the class discussion before seeing the data was better than the post-data followup, which illustrates a general point that concepts can sometimes be clearer in theory, with real data providing a useful check on speculation.

For any particular class, the interpretation of the "two truths and a lie" experiment will depend on the data that come in, and you should be prepared for anything. In the class discussion before the scatterplot and fitted model are revealed, it is natural for students to expect the certainty score to be a strong predictor of empirical accuracy. If this occurs, great; if not, this is an excellent opportunity to discuss the challenges of measurement and the value of statistical evaluation of a measurement protocol.

3. Adapting the activity to courses at different levels

The "Two truths and a lie" game should be fun for any group of students, but the relevant statistical lesson will depend on the level of the course being taught. This activity connects to several important topics, including measurement, uncertainty, prediction, calibration, and logistic regression. Because of its social aspect, it makes sense to do this during the first or second week of the semester, but it is also always important to explicitly connect classroom activities to the material being covered that week, as well as to the course as whole. We give some details about how this could go for a few different courses at different levels.

3.1. Introductory statistics

For an introductory course, the focus can be on probability and uncertainty. Before the activity begins, ask students to speculate on how accurate their guesses will be? On average, will they be able to guess the lie every time? 90% of the time? 50%? More than 33%, we hope, right? This can be an opportunity to introduce the concept of a null hypothesis: from pure random guessing, the number of correct guesses would have the binomial distribution with probability 1/3. Depending on when during the semester this activity is done, we can follow up with an estimate of average success probability with standard error, a confidence interval, and a hypothesis test. At the same time it is important to keep the larger perspective of the sampling distribution, so when presenting these results we should engage the students in discussion of how these numbers would all change if the data were different: with the given n, how many successful guesses would be needed for the null hypothesis to be rejected at the 5% level or the 95% interval for the success rate to exclude 1/3, and so forth.

We can connect this to other problems of probability estimation such as weather forecasting, election forecasting, and so on. We can display the fitted curve without going into detail on logistic regression, just giving this as an example of an advanced statistical method. For an example during the first weeks of an introductory class, the lesson here is not any particular technique, but rather the way that statistical analysis can be used to learn information from subjective certainty statements. Statistical modeling is playing an important role as a bridge between the qualitative and quantitative worlds.

3.2. Bayesian statistics

For a course on Bayesian statistics, the activity can be used to demonstrate the principle of calibration. In this activity, the certainty judgments represent prior information but not prior *distributions*, and the step of fitting a model to predict accuracy of guesses y given certainty judgments x can be seen as a data-based construction of a prior distribution. For example, suppose the model is $\Pr(y = 1) = \text{logit}^{-1}(-0.6 + 0.3x)$; this corresponds to a probability of correct guess ranging from 0.35 when $x = 0$ to 0.92 when $x = 10$, and for a new guess with certainty judgment x, the value $\text{logit}^{-1}(-0.6 + 0.3x)$ can be taken as the prior probability that this guess is correct. The point here is that priors for real problems can be calibrated based on the accuracy of past guesses; this is, for example, how point spreads for sporting events can be translated into betting odds (Stern, 1997).

These points can be placed in the context of a class discussion via a series of prompts. As always, it is best to start the discussion *before* the data have been revealed. To start, students can consider in pairs what is their prior probability that a particular guess is correct: at what odds would they be willing to bet that they actually caught the lie? The next question is how this prior probability varies with x. From this they can see the certainty judgments as a device for constructing an empirical prior distribution. We can then ask how large a dataset might be needed for this prior to be useful in practice. It is easiest to get a sense of this using simulation, starting with some assumption about the function $\text{E}(y|x)$, trying out a sample size, simulating data, and seeing what the plot of $\text{E}(y)$ vs. x looks like. The connection of all this to the class-participation activity is that, by giving the certainty statements and guesses themselves, students should get a picture of the challenges of constructing empirically-based priors.

3.3. Generalized linear models or machine learning

For a class on generalized linear models or machine learning, you can use this as an introduction to logistic regression, showing the details of fitting and graphing the model, interpreting the coefficient estimates and standard errors, and using the prediction from the model to make probabilistic forecasts for new cases. Here the activity ties directly into the material taught in the class, and after the model has been fit, graphed, and explained, there is a sequence of logical followups. Students can discuss the range of predicted probabilities: will they always fall between 0 and 1? (Yes.) Will they always fall between 1/3 and 1? (Not necessarily.) How many measurements would be necessary for the slope of the curve to be estimated with some desired level of accuracy? (We can approximately figure this out using the rule that the standard error scales like $1/\sqrt{n}$ and can also check by simulating fake data.)

Depending on the course material, this activity can be followed up in different ways. For example, a simulation can be performed to assess the statistical power of the study given sample size n, the distribution of observed certainty scores x in the observed data, and assumed values of the intercept a and slope b of the logistic regression; in R:

```
n_loop <- 1000
slope_est <- rep(NA, n_loop)
slope_se <- rep(NA, n_loop)
for (i in 1:n_loop){
  x_sim <- sample(x, n, replace=TRUE)
  y_sim <- rbinom(n, 1, invlogit(a + b*x_sim))
  sim <- data.frame(x_sim, y_sim)
  fit <- glm(y_sim ~ x_sim, family=binomial(link="logit"), data=sim)
```

```
  slope_est[i] <- coef(fit)["x_sim"]
  slope_se[i] <- se.coef(fit)["x_sim"]
}
power <- mean(abs(slope_est)/slope_se > 2)
```

Here we have computed the power following the conventional rule that the estimated slope is statistically significant if it is more than two standard errors from zero. There is no need to perform this particular calculation; we are just illustrating how the data collected in this activity can be used as a starting point for relevant lessons.

3.4. Psychometrics and multilevel modeling

Another direction is to turn this into a lesson on reliability and validity of measurement. What is meant by that certainty score? How useful would we expect the certainty score to be in making a probabilistic forecast? This sort of calibration problem arises in many areas of science and policy. For example, consider a hiring setting where interviewers give numerical ratings for the candidates, and then later when there is data on job performance, the ratings can be retrospectively calibrated. One direction is to set up some comparison points, for example by asking respondents to give certainty scores for other outcomes such as weather or sporting events. There is a large literature on the difficulty of assessing accuracy of subjective guesses; see for example Nisbett and Wilson (1977) and Vredeveldt and Sauer (2014).

For a class on psychometrics or multilevel modeling, this discussion of measurement can serve as an entry point to the design and analysis of repeated measures data. What if the confidence of the guess is highly predictive of accuracy at the individual level, but with an effect that disappears when aggregated across guessers? Students can discuss in pairs how such a pattern can arise, if less accurate guessers tend also to be overconfident. To learn this pattern we would need to gather multiple measurements on each guesser, for example by having students make guesses and certainty statements individually rather than via consultation, and then the resulting data could be fit using a multilevel model with intercept and perhaps slope that vary by guesser.

4. Discussion

A key part of any class-participation activity is how things go after the data have been collected and analyzed.

We do not want to just dump the data and analysis (as in Figure 3) onto the screen and stop there. It is fun when an activity has a twist, but it is not a magic trick; the point is not to amaze students but to bring them closer to the material being taught. We want the activity not to mystify but to de-mystify. So it is important to follow up the activity with explicit discussion, both of its connection to the material being taught in the class and its relevance to real-world applications of statistics in areas such as education, business, politics, or health, depending on the interests of the students.

We can also consider what lessons students might take away from this activity. "Statistics is fun": that's a good memory. "I got fooled by Jason's lie: he's not really adopted": that's fine too, as it serves the goal of students getting to know each other. "You can use logistic regression to convert a certainty score into a predicted probability": that's good because it's a vivification of a general mathematical lesson. "The estimated slope was smaller than the standard error so we couldn't distinguish it from zero": that's not a bad lesson either. Think about what memories you want to create, and keep the discussion focused. For example, the details of the truths and lies are fun, and there could be a temptation to share some of the most successful lies with the class—but for a class on statistics or research methods, those sorts of details could be counterproductive, eliciting memories that would distract from the statistical lessons. We want the activity to be vivid and memorable but for the right reasons.

In our experience we have seen three sorts of positive outcomes associated with this sort of activity, especially when performed near the beginning of the semester. The first is that students get used to the idea that attendance is active, not passive, and we hope the alertness required to perform these activities translates into better participation throughout the class period. The second is that people typically find data more interesting and relatable when they can see themselves in the scatterplot. The third valuable outcome is that the "Two truths and a lie" activity is a social icebreaker. That said, we do not have direct empirical evidence of the effectiveness of this activity on student learning. It is our hope that in laying out this activity—not just the general concepts but also the details of implementation, including instructions, Google form, sample data and analysis, and post-analysis discussion points—we have lowered the barrier of difficulty so that instructors in a wide range of statistics courses can try it out in their own classes, at minimal cost in classroom time and with the potential to get students more involved in their learning of statistics.

That said, we have not offered any formal evaluation of this activity. As is typically the case in education, it is easier to develop a new idea than it is to quantitatively evaluate its effects in the classroom (Chance et al., 2008, Gelman and Loken, 2012). We can still learn from experience, but such learning tends to be qualitative, from observing student reactions and discussions. The biggest risk or opportunity cost we see in introducing a new class-participation activity is that time spent in the activity could be spent working on lecturing or problem solving. For this reason, it is important that the activity be closely tied to the course material (as discussed in Section 3) and that it be performed efficiently, with instructions and Google form prepared ahead of time and with code all set up to analyze the data when they come in. You can also use the present article as a template for designing and implementing your own class-participation activities.

References

Bligh, D. A. (2000). *What's the Point in Discussion?* Exeter, England: Intellect.

Chance, B., Garfield, J., Medina, E., and Ben-Zvi, D. (2008). Assessment in statistics education. In *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*, ed. J. Garfield and D. Ben-Zvi. New York: Springer.

Cobb, G. (1992). Teaching statistics. In *Heeding the Call for Change: Suggestions for Curricular Action*, ed. L. A. Steen, 3–34. Mathematical Association of America.

Gelman, A., and Loken, E. (2012). Statisticians: When we teach, we don't practice what we preach. *Chance* **25** (1), 47–48.

Gnanadesikan, M., Scheaffer, R. L., Watkins, A. E., and Witmer, J. A. (1997). An activity-based statistics course. *Journal of Statistics Education* **5** (2).

Magel, R. C. (1998). Using cooperative learning in a large introductory statistics class. *Journal of Statistics Education* **6** (3).

Mazur, E., and Watkins, J. (2009). Just-in-time teaching and peer instruction. In *Just in Time Teaching: Across the Disciplines, Across the Academy*, ed. S. Simkins and M. H. Maier, 39–62. Sterling, Va.: Stylus Publishing.

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review* **84**, 231–259.

Nolan, D., and Speed, T. (2000). *Stat Labs: Mathematical Statistics Through Applications.* New York: Springer-Verlag.

Rossman, A., and Chance, B. (2001). Teaching contemporary statistics through active learning. `http://www.rossmanchance.com/pbs/pbs.html`

Rossman, A. J., and Chance, B. L. (2008). *Workshop Statistics: Discovery with Data*, third edition. New York: Key College Publishing.

Stern, H. S. (1997). How accurately can sports outcomes be predicted? *Chance* **10** (4), 19–23.

Vredeveldt, A., and Sauer, J. D. (2015). Effects of eye-closure on confidence-accuracy relations in eyewitness testimony. *Journal of Applied Research in Memory and Cognition* **4**, 51–58.