Monday Breakups

Facebook is a social networking website. One piece of data that members of Facebook often report is their relationship status: single, in a relationship, married, it's complicated, etc.

With the help of Lee Byron of Facebook, David McCandless—a London-based author, writer, and designer—examined changes in peoples' relationship status, in particular, breakups. A plot of the results showed that there were repeated peaks on Mondays. Based on this initial examination of data, McCandless speculated that breakups are reported at a higher frequency on Mondays. This is his **research hypothesis**.

To test this research hypothesis, McCandless collected a random sample of 50 breakups reported on Facebook within the last year. Of these sampled breakups, 13 occurred on a Monday.

In this activity, you will be exploring the following research question:

Is 13 (out of 50) breakups reported on Mondays consistent with the model where breakups are equally likely during the week? Or is it more consistent with a model where Mondays have a higher frequency of breakups?

Discuss the Following Questions

1. What does the observed data (13 out of 50 breakups on Monday) suggest about the answer to the research question? Explain.

'Equally Likely' Model

Suppose for the moment that the researchers' conjecture is wrong, and breakups *are not* reported on Monday more than any other day. In other words, breakups are reported uniformly throughout the week. This, is a statistical hypothesis. Namely,

> *Statistical Hypothesis:* Breakups are reported at the same frequency/ percentage on each day of the week.

This statistical hypothesis specifies an "equally likely" breakup reporting model for each day of the week. We can use TinkerPlots™ to create this model and generate random outcomes.

Statistical Hypothesis

A statistical hypothesis is a statement **specifying a model** that explains variation in a particular outcome.

Wait a minute…if McCandless believes that breakups are reported more frequently on Mondays, why wouldn't that be his statistical hypothesis? It could be, but remember, the statistical hypothesis has to specify a model that can be used to generate outcomes. One such model that fits his research hypothesis is that 30% of breakups are reported on Mondays. Another is that 50% of breakups are reported on Mondays. How many others are there?
In order to answer his research question, McCandless would have to examine the results from every single one of those models to see if 13 out of 50 breakups is consistent with one of those models.

#protip: Rather than examine each of those models, statisticians often use the 'equally likely' model. Mathematically, this model acts as a "lower bound" for all the models where Mondays has more reported breakups than other days. If 13 out of 50 breakups is an extreme (high) result given this model, we can rule out this model and suggest that the observed data are inconsistent with the model. This points toward a model in which the reported frequency of breakups is higher on Mondays, without having to know the exact percentage of breakups that occur on Monday!

2. Draw a picture of the sampler (model) that you will use to generate outcomes from the model specified in the statistical hypothesis. In the picture, be sure to (1) indicate the type of sampling device used (mixer, spinner, etc.); (2) label all the elements in your sampling device; (3) label the probability associated with each element; and (4) indicate the Repeat and Draw values you will use.

- Set up the model/sampler in TinkerPlots™.

Simulating the Data

- Carry out a single trial of the simulation in TinkerPlots™.
- Plot the outcomes from the trial.

3. Sketch a plot of the outcomes from this trial. Add all labels and statistics (counts, percentage, etc.) to your plot.

Remember what we are ultimately interested in is the number of breakups (out of 50) that are reported on Mondays.

- Collect the appropriate statistic
- Carry out 499 more trials (500 trials total) of the simulation in TinkerPlots™.

Evaluating the Hypothesized Model

- Plot the results from the simulation.

4. Sketch a plot of the results below.

5. Now, reconsider the observed data; the data that McCandless actually observed initially. He observed 13 breakups (out of 50) that were reported on Monday. Given the plot of results, is 13 a value that is consistent with the model's results or not? Explain.

6. What does this consistency/inconsistency suggest about the answer the research question? Explain.

Modeling the Monday Breakups Problem

MODEL

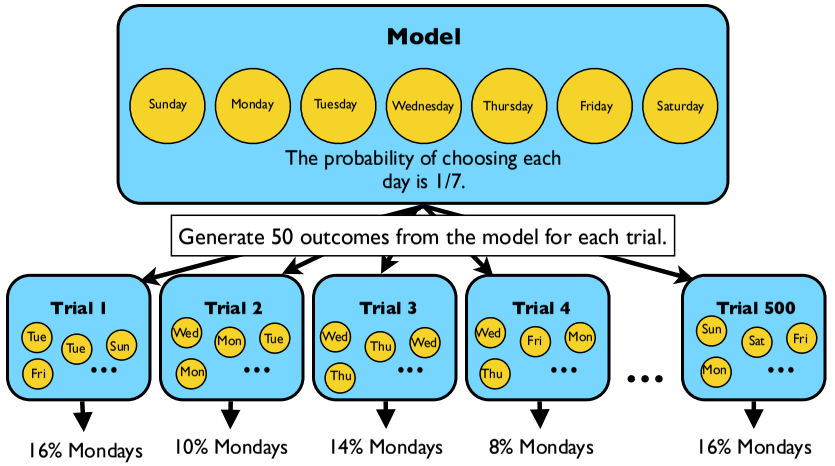
The model has seven outcomes–one for each day of the week. Each is equally likely.

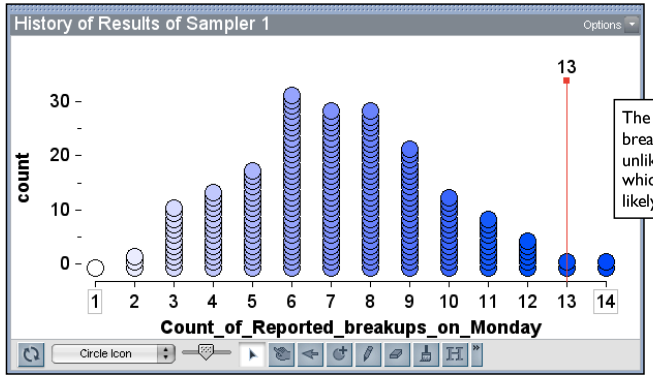
SIMULATE

A trial ends when 50 outcomes have been randomly generated from the model For each trial, compute the percentage of breakups "reported on Monday". Generate many trials.

EVALUATE

Compile all of the numerical summary measures into a single distribution, Answer the research question by evaluating whether the empirical evidence (observed result) tis consistent with the "equally likely" model.





The observed result (13 breakups on Monday) is unlikely given the model in which breakups are equally likely each day of the week.