Can Dolphins Communicate?

Adapted from an activity in "Introduction to Statistical Investigations", Tintle et al.; 2012, John Wiley and Sons

A famous study from the 1960s explored whether two dolphins (Doris and Buzz) could communicate abstract ideas. Researchers believed dolphins could communicate simple feelings like "Watch out!" or "I'm happy," but Dr. Jarvis Bastian wanted to explore whether they could also communicate in a more abstract way, much like humans do. To investigate this, Dr. Bastian spent many years training Doris and Buzz and exploring the limits of their communicative ability.

During a training period lasting many months, Dr. Bastian placed buttons underwater on each end of a large pool—two buttons for Doris and two buttons for Buzz. He then used an old automobile headlight as his signal. When he turned on the headlight and let it shine steadily, he intended for this signal to mean "push the button on the right." When he let the headlight blink on and off, this was meant as a signal to "push the button on the left." Every time the dolphins pushed the correct button, Dr. Bastian gave the dolphins a reward of some fish. Over time Doris and Buzz caught on and could earn their fish reward every time.

Then Dr. Bastian made things a bit harder. Now, Buzz had to push his button before Doris. If they didn't push the buttons in the correct order—no fish. After a bit more training, the dolphins caught on again and could get earn their fish reward every time. The dolphins were now ready to participate in the real study to examine whether they could communicate *with each other*.

Dr. Bastian placed a large canvas curtain in the middle of the pool. (See Figure 1.1.) Doris was on one side of the curtain and could see the headlight, whereas Buzz was on the other side of the curtain and could not see the headlight. Dr. Bastian turned on the headlight and let it shine steadily. He then watched to see what Doris would do. After looking at the light, Doris swam near the curtain and began to whistle loudly. Shortly after that, Buzz whistled back and then pressed the button on the right—he got it correct and so both dolphins got a fish. But this single attempt was not enough to convince Dr. Bastian that Doris had communicated with Buzz through her whistling. Dr. Bastian repeated the process several times, sometimes having the light blink (so Doris needed to let Buzz know to push the left button) and other times having it glow steadily (so Doris needed to let Buzz know to push the right button). He kept track of how often Buzz pushed the correct button.

Figure 1.1: Depending whether or not the light was blinking or shown steadily, Doris had to communicate to Buzz as to which button to push.
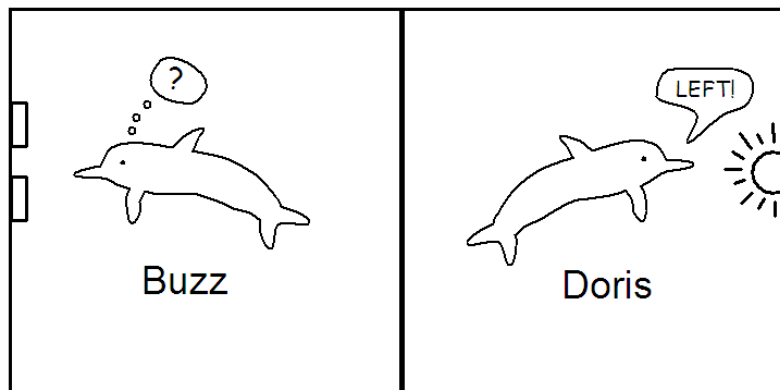
In one phase of the study, Dr. Bastian had Buzz attempt to push the correct button a total of 16 different times. In this phase of the study, Buzz pushed the correct button 15 out of 16 times. This means that the proportion of the time that Buzz was correct was 15/16 = 0.938.

Think about it: Based on these data, do you think Buzz somehow knew which button to push? Is 15 out of 16 correct pushes convincing to you? Or do you think that Buzz could have just been guessing? How might you justify your answer?

Buzz certainly pushed the correct button most of the time, so we might consider either:

- Buzz is doing something other than just guessing (his probability of a correct button push is more than 0.50).
- Buzz is just guessing (his probability of a correct button push is 0.50) and he got really lucky in these 16 attempts.

These are the two possible explanations to be evaluated. It's certainly *possible* that Buzz was just guessing and got lucky! But does this seem like a reasonable explanation to you? How would you argue against someone who thought this was the case?

So how are we going to decide between these two possible explanations? One approach is to model the random process (repeated attempts to push the correct button) and then see whether the observed data is consistent with guessing. That is, is it the kind of result that we might expect to see just by chance if Buzz was just guessing?

Suppose "Buzz was simply guessing". What proportion of the time out of 16 tries would we expect him to be correct? To investigate, we can model guessing behavior using a coin flip to represent (or *simulate*) Buzz's choice *assuming he is just guessing* which button to push. To generate this artificial data, we can let "heads" represent the outcome that Buzz pushes the correct button and let "tails" be the outcome that Buzz pushes the incorrect button. This gives Buzz a 50% chance of pushing the correct button. This can be used to represent the "Buzz was just guessing" or the "random chance alone" explanation.

Now that we see how flipping a coin can simulate Buzz guessing, let's flip some coins to simulate Buzz's performance. Imagine that we get heads on the first flip. What does this mean? This would correspond to Buzz pushing the correct button! But, why did he push the correct button? In this guessing model, the only reason he pushed the correct button is because he happened to guess correctly—remember the coin is simulating what happens when Buzz is just guessing which button to push.

What if we keep flipping the coin? Each time we flip the coin we are simulating another attempt where Buzz guesses which button to push. Remember that heads represents Buzz guessing correctly and tails represents Buzz guessing incorrectly. How many times do we flip the coin? Sixteen, to match Buzz's 16 attempts in the actual study. After 16 tosses, Figure 1.2 shows our **simulated data**.

Figure 1.2: Results from tossing a coin 16 times

Here we got 11 heads and 5 tails (11 out of 16 or 0.6875 is the simulated statistic). This gives us an idea of what could have happened in the study if Buzz had been randomly guessing which button to push each time.

Will we get this same result every time we flip 16 coins? Let's flip our coin another 16 times and see what happens. When we did this we got 7 heads and 9 tails as shown in Figure 1.3 (7 out of 16 or 0.4375 is the simulated statistic).

**Figure 1.3:** Results from a second repetition of tossing a coin 16 times

So can we learn anything from these coin tosses when the results vary between the sets of 16 tosses?

Using and evaluating the coin flip guessing model

Because coin flipping is a random process, we know that we won't obtain the same number of heads with every set of 16 flips. But are some numbers of heads more likely than others? Does the *distribution* of the number of heads that result in 16 flips have a predictable long-run pattern? In particular, how much *variability* is there in our simulated statistics between repetitions (sets of 16 flips) just by random chance?

In order to investigate these questions, we need to continue to flip our coin to get many, many sets of 16 flips (or many repetitions of the 16 choices where we are modeling Buzz is simply guessing each time). We did this, and Figure 1.4 shows what we found when we graphed the number of heads from each set of 16 coin flips. Here, the process of flipping a coin 16 times was repeated 1,000 times—this number was chosen for convenience, but also appears to be large enough to give us a fairly accurate sense of the long-run behavior for the number of heads in 16 tosses.

Figure 1.4: A dotplot showing 1000 repetitions of flipping a coin 16 times and counting the number of heads
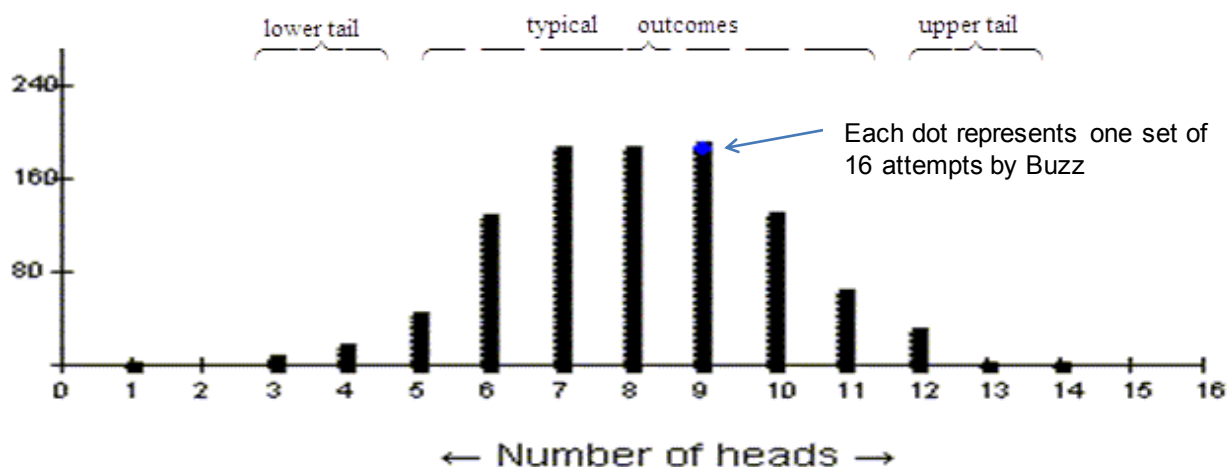
1. Build a worksheet to simulate this experiment and use it to obtain 1,000 repetitions.

2. Construct a column chart similar to Figure 1.4. Comment on the differences and similarities between the dotplot and the column chart.

Let's think carefully about what the graph in Figure 1.4 shows. For this graph, one *observational unit* (one dot) represents one set of 16 coin tosses and the *variable*, what we measure about each set of 16 coin tosses, is the number of heads. This is a quantitative variable. We see that the resulting number of heads follows a clear pattern: 7, 8, and 9 heads happened quite a lot, 10 was pretty common also (though less so than 8), 6 happened some of the time, 1 happened once. But we never got 15 heads in our 16 tosses! We might consider any outcome between about 5 and 11 heads to be typical, but getting fewer than 5 heads or more than 11 heads happened rarely enough to be considered a bit unusual. We refer to these results as being out in the tails of the distribution.

Think about it: How does the above analysis help us address the strength of evidence for our research conjecture that Buzz was doing something other than just guessing?

So what does this have to do with the dolphin communication study? Well, we said that we would flip a coin to simulate what could happen if Buzz was really just guessing each time he pushed the button in 16 attempts. We saw that getting results like 15 heads out of 16 never happened in our 1000 repetitions—showing that 15 is a very unusual outcome, far in the tail of the distribution of the simulated statistics, if Buzz is simply guessing. In short, even though we expect some variability in the results for different sets of 16 tosses, the pattern shown in this distribution indicates that an outcome of 15 heads is outside the typical chance variability we would expect to see when Buzz is simply guessing.

In the actual study, Buzz really did push the correct button 15 times out of 16, an outcome that would rarely occur if Buzz was just guessing. So, our coin flip guessing model tells us that we have very strong evidence that Buzz was not just guessing to make his choices. Therefore, we don't believe the "by chance alone" explanation is a good one for Buzz. We instead believe the study result provides strong enough evidence to be statistically significant, not a result that happened by chance alone.

Another Doris and Buzz study

One goal of statistical significance is to rule out random chance as a *plausible* explanation for what we have observed. We still need to worry about how well the study was conducted. For example, are we absolutely sure Buzz couldn't see the headlight around the curtain? Are we sure there was no pattern to which headlight setting was displayed that he might have detected? And of course we haven't completely ruled out random chance, he may have had an incredibly lucky day. But the chance of him being that lucky is so small, that we conclude that other explanations are more plausible or credible.

One option that Dr. Bastian pursued was to re-do the study except now he replaced the curtain with a wooden barrier between the two sides of the tank, to ensure a more complete separation between the dolphins to see whether that would diminish the effectiveness of their communication.

In this case, Buzz pushed the correct button only 16 out of 28 times. The variable is the same (whether or not Buzz pushed the correct button), but the number of observational units has changed to 28 (the number of attempts).

Think about it: For this phase of the study, do you think that Doris could tell Buzz which button to push, even under these conditions? Or is it believable that Buzz could have just been guessing?

3. What proportion of the time did Buzz choose correctly in this second study?

4. Is it plausible (believable) that Buzz was simply guessing in this set of attempts? Carry out a simulation similar to the one described for the first study and create a column chart that shows the distribution of sample proportions under the guessing model for this experiment.

5. Based on the simulated results, are you convinced that Buzz is not just guessing? Explain your reasoning