Workshop 3: Planning tools

Before carrying out a time- and fund-consuming experiment, it is useful to get an idea of what to expect from the results. How big an effect are you expecting? What are the chances that you would detect it? What sample size would you need to have a reasonable chance of succeeding? How narrow a confidence interval around the estimated effect would you be happy with? In this workshop we will show how R can be used to address some of these questions.

Random sampling warm-up

To begin, get some practice sampling categorical and normally-distributed data from a population.

1. Sample 20 observations from a population having two groups of individuals, "infected" and "uninfected", in equal proportions. You could do this using the `sample` function:

   ```
   sample(c("Infected", "Uninfected"), 20, replace = TRUE)
   ```

   or using a call to `rbinom`

   ```
   rbinom(n=1, size=20, prob=0.5)
   ```

2. Repeat the previous step five times to convince yourself that the outcome varies from sample to sample.

3. Sample 18 individuals from a population having two groups of individuals, "mated" and "unmated", where the proportion mated in the population is 0.7. **Hint:** Modifying the `rbinom` should be straightforward. If you want to use the `sample` command, you can pass it a `prob` argument.

4. Repeat the previous step five times to convince yourself that the outcome varies from sample to sample.

5. Sample 30 observations from a normally-distributed population having mean 0 and standard deviation 2. Plot the results in a histogram.

   ```
   rnorm(30,0,2)
   ```

6. Repeat the following 5 times and calculate the mean each time: sample 30 observations from a normally-distributed population having mean 0 and standard deviation 2. Convince yourself that the sample mean is different each time.

For loops

In R (and many other programming languages), a `for` loop can be used dto repeat a block of code some number of times. The “`i`” in this loop is a counter that iterates over some specified set of values (e.g., `1:10`), each time executing the commands that are enclosed within the curly brackets `{}`

1. Start by writing a for loop that prints the numbers 1 to 10.

   ```
   for(i in 1:10) {
     print(i)
   }
   ```

2. Instead of printing out your results, you can concatenate them into a vector or a data frame.

   ```
   new_vec <- c()
   for(i in 1:10) {
     new_vec <- c(new_vec, i)
   }
   new_vec
   ```

3. Use a `for` loop to calculate and store the values of x^2 for $x = 1, \cdots, 10$.

4. Use a `for` loop to calculate and store the values of x^2 and x^3 for $x = 1, \cdots, 10$. **Hint:** You might need to create a more complex data-structure to store your results (e.g., a matrix). You can create an empty matrix with the following:

   ```
   matrix(NA, nrow=10, ncol=2)
   ```

   where `nrow` and `ncol` specify the number of rows and columns.

5. You can use for loops to iterate over more than just vectors of integers. E.g., try the following:

   ```
   myVec <- c('cat', 'dog', 'mouse')
   for(i in myVec) {
     print(i)
   }
   ```

Plan for precision

Consider an experiment to estimate mate preference of females of a species of jumping spiders. Each independent trial involves presenting a female spider with two tethered males. One of the males is from her own species, and the other is from its sister species. To avoid pseudoreplication, females are tested only once and males are replaced between tests. You want to estimate p, the proportion of female spiders that choose males of their own species. Before carrying out the experiment, it is useful to generate data under different scenarios to get a sense of the sample size you would need to estimate preference with sufficient precision.

Estimate weak or no preference

We'll start with the case of weak or no preference: Imagine that females choose males essentially randomly ($p = 0.5$), with half choosing the male from her own species and the other half picking the male of the other species. How much data would you need to demonstrate this (and convince your skeptical supervisory committee)? One idea is to collect data and use it to test the null hypothesis of no preference. If the null hypothesis is true, you should fail to reject it. However, this won't be very convincing to your committee. Failing to reject a null hypothesis is inconclusive by itself. Maybe your test won't have much power.

A second idea is to plan your sample size so as to obtain a narrow confidence interval (high precision) for the strength of preference. If, at the end of your experiment, you end up with an estimate of p close to 0.5 *and* your 95% confidence interval for p is relatively narrow, you'll be in a strong position to say that the true preference really is weak, even if you can't say it is exactly 0.5. What sample size is necessary to achieve a reasonably narrow confidence interval in this case? Investigate this question by simulating data.

1. Randomly sample $n = 10$ females from a population having equal numbers of "successes" (females who choose males of her own species) and "failures" (females who choose males of the other species). What was the proportion of successes in your sample?

2. Using the data from step 1, calculate an approximate 95% confidence interval for the population proportion of successes. Use the Agresti-Coull method in the `binom` package in R, which you will need to install if you haven't already done so.

```
install.packages('binom', dependencies = TRUE) # install once
    only
library(binom)                                 # load before
   using
```

   To obtain the 95% confidence interval, use the binom.confint function explained below. The argument `x` is the number of "successes" in your generated sample (number of females who chose males of her own species) and `n` is the sample size (number of females tested).

```
myCI <- binom.confint(x, n, method='ac') # calculate
   confidence interval
myCI                                     # shows the results
myCI$lower                               # the lower limit
myCI$upper                               # the upper limit
```

3. Repeat steps 1 and 2 five times and keep a record of the confidence intervals you obtained. What was the lowest value for the span of the confidence interval from the 5 samples?

4. You can speed up the effort if you create a for loop in R that automatically repeats steps 1 and 2 as many times as you decide.

5. Increase the sample size to `n=20` and run the loop from step 4 again. How much narrower are the confidence interval spans? Are the spans adequate?

6. By modifying the sample size and re-running the loop a bunch of times, find a sample size (ballpark, no need to be exact at this point) that usually produces a confidence interval having a span no greater than 0.2. This would be the span of a confidence interval that had, e.g., a lower limit of 0.4 and an upper limit of 0.6. Surely this would be convincing evidence that the mate preference really was weak.
   By this point you might wish to speed things up by saving the results of each iteration to a vector or data frame rather than print the results to the screen. This will make it possible to increase the number of iterations (say, to 100 times instead of just 10) for a given value of `n`.

7. Given the results of step 6, you would now have some design options before you. Is the sample size `n` that your simulation indicated was needed to generate a confidence interval of span 0.2 realistic? In other words, would an experiment with so many female spiders (and so many males) be feasible? If the answer is yes, great, get started on your experiment! If the answer is no, the sample size required is unrealistically large, then you have some decisions to make:

   - Forget all about doing the experiment. (Consider a thesis based on theory instead.)
   - Revise your concept of what represents a “narrow” confidence interval. Maybe a confidence interval for `p` spanning, say, 0.3 to 0.7 (a span of 0.4) would be good enough to allow you to conclude that the preference was “not strong”. This would not require as big a sample size as a narrower interval.

Plan for power

Assume that the preference p really is different from 0.5, and use null hypothesis significance testing to detect it. What strength of preference would we like to be able to detect in our experiment? To pick an extreme case, if the true proportion of females in the population choosing a male from her own species is 0.51 rather than 0.50, you would need an enormous sample size to detect it. But we don’t really care about such a small effect. Let’s start instead with the more realistic proportion $p = 0.7$. What sample size would be needed to detect it with reasonably high probability?

1. Sample 20 females from a population in which the true fraction of “successes” is 0.7

2. Apply the binomial test to your sample, to test the null hypothesis that the population proportion is 0.5. The binomial test calculates the exact 2-tailed probability of a result as extreme or more extreme as that observed if the null hypothesis is true. The method is implemented in R in the following command,

   ```
   z <- binom.test(x, n, p=0.5)
   ```

   where `x` is the observed number of successes in your sample from step 1, and `n` is the sample size. `z` here is an object that stores the result. To see the results of the test enter just `z` in the command line. If you just want to see the resulting P-value of the test, enter

```
    z$p.value
```

instead. Did you reject the null hypothesis?

3. Create a `for` loop to repeat steps 1 and 2 100 times. In what fraction of iterations was the null hypothesis rejected?
4. By modifying the sample size and re-running the loop multiple times, find a sample size (ballpark, no need to be exact at this point) that usually results in the null hypothesis being rejected. Compare your results to those from the confidence interval simulation above.
5. Is the sample size you determined feasible in an experiment? If the answer is yes, great! If the answer is no, because the sample size required is too large, then you have some decisions to make. You could decide not to run the experiment after all. Or, you could revise your aims. Perhaps your committee would be happy if you if you could detect a preference of 0.8 instead of 0.7.