False positive statistics

Day 3 concepts covered:

1. Plot experimental data and outcomes of simulations
2. Compute summary statistics from a dataset
3. Use computer simulations to build intuitions about random variables
4. Simulate random distributions
5. Use computer simulations to inform experimental design
6. Articulate and encode the Null Hypothesis
7. Generate a simulated distribution under the Null Hypothesis
8. Compare simulated results to data and obtain a p value

My postdoc and I have a new audio-visual theory of perception that predicts that neurons in primary visual cortex should shift their preferred orientations in a clockwise direction when a monkey is listening to Led Zeppelin vs. Hoyt Axton. We collect data from 10 neurons under the two different auditory conditions and find that there is indeed a difference, which, though not statistically significant ($p=0.12$), is trending in the right direction. We decide that we need to collect more data, so we record from two new neurons each day until we obtain significance ($p<0.05$) or reach 50 neurons, in which case we will quit and declare the experiment a failure. If we find a significant difference, what false positive rate (i.e. the probability that we have rejected H_0 when H_0 was true) should we report in our paper?

Step 1:

Decide on the procedure you will use for simulating the experiment and note your steps in pseudo-code. See Hints 1 to 3 for help.

Step 2:

Write your implementation. What false-positive rate did you find?

Step 3:

In a paper called “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant” (Psychological Science 22(11): 1359, 2011) Simmons and colleagues discuss common mistakes in data collection and analysis that result in a false positive rate much higher than the 5% usually reported. We have provided an annotated version of the paper in this folder, under the file name simmons-nelson-simonsohn-false_positive_statistics-psycholsci2011.pdf

One common mistake is the absence of a “termination rule” for data collection. This occurs if a researcher does not decide in advance to end data collection after a fixed number of samples (or at a specific time), but continues collecting data until the results are significant.

What do the authors show in figures 1 and 2 of the paper? How do your results compare to theirs?

Extra Step 4:

We provide two files that have been used to reproduce the results of the paper: SampleSizeDF.m (reproducing figure 2) and in file dfSim.m (reproducing figure 1). See if you can read the code and understand what it does. Add comments to the code explaining what each block or line is doing

Hints

Hint 1: If you have trouble getting started, think through the experimental procedure. What is the Null Hypothesis. What do we measure initially? What steps are repeated, and when do we stop? What output do we need to collect?

Hint 2: Do not worry about the exact nature of the statistical test that we performed—use any old 2-sample test on H0 data. Similarly, do not worry about what sort of distribution the data should be coming from, as long as your choice of distribution is consistent within the simulation. Does the nature of the test or of the underlying distribution matter?

Hint 3: Do not get hung up on the value of our first test ($p = 0.12$). What we want to know is the true false positive rate of the *procedure* we followed: 10 observations for each of two conditions, test, 2 more observations for each conditions, re-test, repeat until $p < 0.05$ or $n = 50$ observations.

Revisions:
RTB created files SampleSizeDF.m and dfSim.m, 12/21/2012
MIS wrote step-by-step guide, 5/12/2014
MIS merged with RTB's document (including monkey study) 5/20/2014