PSYC 5301 - Week 1

Thomas J. Faulkenberry, Ph.D.

Jan 16, 2018

An example

Suppose you have a treatment that you suspect may alter performance on a certain task. The two groups were significantly different, $t(18) = 2.7$, $p = 0.01$. Decide whether each of the following statements is true or false:

1. You have disproved the null hypothesis

2. You have found the probability of the null hypothesis being true

3. You have proved your experimental hypothesis

4. You can deduce the probability of the experimental hypothesis being true

5. If you decide to reject the null hypothesis, you know the probability that you are making the wrong decision

6. You have a reliable experimental finding in the sense that if, hypothetically, the experiment were repeated a

large number of times, you would obtain a significant result on 99% of the replications.

Some definitions

1. What is a p-value?

   - p-values tell you how *surprising* the *data* is, assuming there is *no effect*.
   - Benjamini (2016): "In some sense it offers a first line of defense against being fooled by randomness, separating signal from noise"
   - from sample statistics (M, SD, n), we calculate a *test statistic* and compare against a distribution (e.g., z, t, F)
     – $p < 0.05$ –> data is surprising
     – $p > 0.05$ –> data is *not* surprising
   - p-value is the probability of getting the observed (or more extreme) data, *assuming the null hypothesis is true*
     – Note: a p-value is the probability of the *data*, not the probability of a *theory*
     – $p = P(D|H) \neq P(H|D)$

2. Decisions

action / truth	H0 false (effect)	H0 true (no effect)
reject H0	correct decision	**Type 1 error**
"accept" H0	**Type II error**	correct decision

- more definitions:
  - α = probability of finding significant result when H0 is true (Type I error rate)
  - β = probability of finding nonsignificant result when H0 is false (Type II error rate)
  - $1-\beta$ = probability of finding signficant result when H0 is false (statistical power)

Philosophical underpinnings

The goal of research is to find the **one truth**. . . however, the **paths are many**. Let's see how an ancient Hindu text can actually serve as a metaphor for how we do science.

Three paths to enlightenment (Bhagavad Gita, 500 BCE):

1. Karma yoga - the path of *action*

2. Jnana yoga - path of *knowledge*

3. Bhakti yoga - path of *devotion*

These map nicely onto Royall's (1997) three questions one should ask regarding data:

1. What should I do?

2. What's the relative evidence?

3. What should I believe?

Paths for research:

1. **Path of action**: search for rules to govern our *behavior* such that, in the long run, we will not be wrong too often
   - $p < \alpha$: reject H_0. *Act* as if data is not noise
   - $p > \alpha$: remain in doubt. *Act* as if data is just noise
   - A rule to govern our *behavior* in the *long run*. It tells us *nothing* about the *current test*.

2. **Path of knowledge**: compare the likelihood of different hypotheses, given the data.
   - suppose you flip a coin 10 times: you get 6 heads and 4 tails. Is the coin biased (unfair)?
   - Two hypotheses:
     - H_1: the coin is biased (the true proportion of heads/tails is 0.6
     - H_2: the coin is fair (true proportion of heads/tails is 0.5

- Question: given the data, how much more likely is H_1 than H_2

figures/coinFlip.png

-

3. **Path of belief**: do I really *believe* this coin will come up heads 60% of the time?

   - No. . . I have *prior* beliefs.
   - One "experiment" with 6 heads does not *change* my prior beliefs

These paths form the basis of three dominant statistical paradigms in the psychological literature:

1. Neyman-Pearson (the most common)

2. Likelihood

3. Bayesian

1. Neyman-Pearson method

   Historically, our method of hypothesis testing (using p-values) is an amalgamation of two (quite different) ideas from a couple of early 20th century statisticians:

   - Jerzy Neyman: p-value tells you what *action* to perform. If $p < \alpha$, then we reject null hypothesis
     - When we *act* as if there is an effect when $p < 0.05$, in the *long run* we won't be wrong more than 5% of the time
   - Ronald Fisher: p-value measures evidence. . . the smaller the p-value, the greater the evidence (this is actually incorrect)

- Note: when I teach undergraduate statistics, I teach *only* the Neyman method.
  - define H_0
  - set α (usually 0.05) and find the critical test statistic
  - if test statistic exceeds critical, we we reject H_0 (action)
- However, most psychological literature (and many courses) implicitly tack on the incorrect Fisher ideas.
  - Example: I got $p = 0.03$ for "Effect 1" and $p = 0.003$ for "Effect 2"..which has "more evidence"?
  - Answer: neither, but Fisher thought Effect 2 would have more evidence
  - this understanding is implicit everywhere in psychology, but it is wrong!
- Goal of Neyman-Pearson method: error control
  - don't make a fool out of yourself in the long run