Research Minute

Hypothesis Testing and the Story of *p*

Issue 14

Sandra Burge, Ph.D.

The Purpose of Research

The purpose of research is to understand general rules about the way things work. Most of the time we seek to understand ***the relationship between two phenomena***, like…

- Does Screening, Brief Interventions, Referral to Treatment (SBIRT - Phenom 1) reduce patients' alcohol consumption (Phenom 2) ?
- Do group visits with families (Phenom 1) improve children's BMI (Phenom 2) ?
- Does clinic-wide training (Phenom 1) improve Zoster immunization rates (Phenom 2)?
- In patients with diabetes, is BMI (Phenom 1) associated with A1c (Phenom 2) ?

Clinical observations may lead us to believe that SBIRT, group visits, training workshops, and BMI affect certain outcomes, but do they, really? Or are we seeing something that could have happened anyway (happened by chance), regardless of the presence or absence of Phenom 1?

Hypothesis Testing and the Story of *p*

To begin, we make a prediction, by writing a Research Hypothesis, $\mathbf{H_1}$:

H_1: SBIRT reduces alcohol consumption in heavy drinkers.

Next, we state a null hypothesis, $\mathbf{H_0}$. The null hypothesis should represent what we might expect by chance. If drinking outcomes are random and not due to our SBIRT efforts, the null hypothesis is…

H_0: SBIRT has no influence on alcohol consumption in these drinkers.

In our study, we draw a sample that (hopefully) is representative of a larger population. We are trying to determine if our findings are likely to be "true" for the population of heavy drinkers.

Statistical tests estimate the probability that the Null Hypothesis is true in the population, based on our findings in the sample. In other words, what is the probability that our original prediction is ***wrong***? That our findings in the sample are random and not true for the population?

I hate being wrong…

This is called a Type I error. We want the probability of that error to be very, very small. So we set our cutoff $\alpha = .05$. See the Table below.

Our statistical test will spit out a probability (p) of a Type I error, traditionally set at $p \leq \alpha$, or $p \leq .05$.

On the other hand, my sample may be particularly resistant to SBIRT interventions, even though SBIRT works for the rest of the population. My research hypothesis is not supported by my findings (even though it should be). That's a Type II Error. Traditionally β is set at .20 or .10.

"Reject the null" … means my findings are likely **not** random, and my research Hypothesis H_1 is supported.

Here's the crazy-making part of hypothesis-testing, the language of double- and triple-negatives.

"Fail to reject the null"… Are you kidding me? Triple negatives? Re-think this as "accept the null," or Hypothesis H_1 is **not supported** by the findings.

	Population: H_1 is true	**Population:** H_0 is true
Sample: Reject H_0	POWER! $1-\beta$	Type I error α
Sample: Fail to Reject H_0	Type II error β	$1 - \alpha$

p-value is all about minimizing Type I error, the probability that we are wrong .

burge.researchminutes@gmail.com

Art by Brian Hischier