PSYC 575: Week 13 Quiz Solution

Last updated on November 12, 2021

Part 1A

Consider that a research team is designing a study to evaluate the effect of a therapy on eating disorder (γ_{01}), and they will test against a null hypothesis ($H_0 : \gamma_{01} = 0$). Consider also that we are in an omniscient god view, and we know whether the null hypothesis is indeed true or false.

1. If the null hypothesis is, indeed, true (the therapy is ineffective),
   a. what is the probability that they will make a Type I error?
      (A) α
   b. what is the probability that they will make a Type II error?
      (C) 0

If we live in a world where the null is true, meaning that the therapy is ineffective, researchers can only either

1. correctly retain the null or
2. make a Type I error.

As the null is true, they can never make a Type II error.

2. If the null hypothesis is, indeed, false (the therapy is effective),
   a. what is the probability that they will make a Type I error?
      (C) 0
   b. what is the probability that they will make a Type II error?
      (B) β

If we live in a parallel world where the null is false, meaning that the therapy is effective, researchers can only either

1. correctly reject the null or
2. make a Type II erorr.

As the null is false, they can never make a Type I error.

Part 1B

Consider we collected some data about the effect of a therapy on eating disorder (γ_{01}), and we test against the null hypothesis ($H_0 : \gamma_{01} = 0$).

3. If we retain the null hypothesis,
   a. what is the probability that we made a Type I error?
      (A) 0
   b. what is the probability that we made a Type II error?
      (C) β

Because we have concluded that the null hypothesis is true,

1. if the null hypothesis is indeed true, then we made the correct decision and made no Type I error
2. if the null hypothesis is indeed false, then we made a Type II error.

4. If we reject the null hypothesis,
   a. what is the probability that we made a Type I error?
      (C) α
   b. what is the probability that we made a Type II error?
      (B) 0

Because we have concluded that the null hypothesis is false,

1. if the null hypothesis is indeed true, then we made a Type I error
2. if the null hypothesis is indeed false, then we made the correct decision and made no Type II error.

		Conclusion about null hypothesis from statistical test	
		Accept Null	Reject Null
Truth about null hypothesis in population	True	Correct	Type I error Observe difference when none exists
	False	Type II error Fail to observe difference when one exists	Correct

Figure from online.

Part 2A

When we design a study that evaluates the effect of a therapy on eating disorder, assuming $\alpha = .05$, if we

5. increase sample size,
    a. what happens to statistical power of detecting an effect (in this example, γ_{01})?
        - Power increases
    b. what happens to standard error (SE)?
        - SE decreases
    c. what happens to the width of 95% confidence interval (CI) for an effect estimate (in this example, $\hat{\gamma}_{10}$)?
        - CI width decreases
6. How does the width of 95% CI relate to
    a. standard error?
        - 95% CI $\approx \gamma_{01} \pm 2 * \text{SE}$
    b. precision?
        - The smaller the SE, the smaller the 95% CI width, the higher the precision.

> In summary, increasing sample
>
> 1. increases power
> 2. decreases standard error, hence decreases confidence interval width and increases precision.

Part 2B

Consider we are planning for a study that evaluates an educational program on students' self efficacy. Again, here we have an omniscient god view and know the true parameter values.

7. In which study we need a larger sample size? A study that examines
    (A) Program A, with a true effect size of .5, or
    (B) **Program B, with a true effect size of .2**
8. In which study we need a larger sample size? A study that has
    (A) **Classes as level-2 clusters, with a true ICC of .3**
    (B) Schools as level-2 clusters, with a true ICC of .1

(Hint: try to play around with the PowerUpR shiny app which we will discuss more in the lecture. Use their default values (main effect, statistical power, two-level CRT), change only the value of `rho2`, and see what happens to the resulting power)

> In summary, if the true effect size is large, we don't need many samples to detect such as effect. However, if the true ICC is large, we need more samples to detect an effect.

Why do we need more samples for a study with a larger ICC?

An intuitive explanation is that we get less information from groups of people that are more alike to one another. Putting it as an extreme, consider that everyone from the same group is identical (i.e., ICC = 1). This situation is like duplicating a data point from the same person over and over again. Although we have a lot of duplicated data points, we only know how the treatment works on this same person.

Another way of explaining this is with design factor. Recall from the week 3 lecture, design factor, Deff = $1 + (n - 1)\rho$, represents the variance inflation due to clustering, resulting in a wider confidence intervals than you would expect in data without clustering. Design effect can also be interpreted as the ratio of the actual sample size to the effective sample size (the sample size you would expect if there was no clustering). Substantial design effect (> 1.1; Lai & Kwok, 2015) indicates that the actual sample size is larger than the sample size you would expect if there was no clustering. In other words, there are not as many samples as you thought you have when people in the same cluster are more alike to each other.

In sum, the larger the true ICC, the more sample size you need to detect an effect.

How do we know the true effect size and ICC?

We can estimate the true effect size and ICC after running a study and collecting some data. However, we don't get such information before conducting a study, but the true values of effect size and ICC are required for sample size planning. A usual practice is to give researchers' best educated guess of these parameter values, based on literature or a pilot study. For example, if similar studies have found ICC in LAUSD schools to be .1 (picked arbitrarily) and we are going to run a study with students from LAUSD schools, then .1 is a reasonable guess of the true ICC in sample size planning. We can use the same way to make an educated guess of other parameter values.

My research is on incorporating researchers' uncertainty in sample size planning. While .1 is a reasonable guess of the true ICC in the above example, we may still feel uncertain about whether it is an accurate estimate of the true ICC due to sampling variability. Some studies may report ICC to be .1 but others may report .2 for students in LAUSD schools. What value of ICC should we use to compute the sample size? The hybrid classical-Bayesian (HCB) approach I propose for multilevel studies allows researchers to specify their degree of uncertainty in the true parameter value. This in my work (currently under review) has been found to improve accuracy in sample size planning. However, the current work only applies to two-level cluster randomized trials, and I will extend this approach to other types of multilevel designs in my career. I will discuss more about this in the lecture and please feel free to reach out to me if you would like to learn more :)

References

Lai, & Kwok, O. (2015). Examining the Rule of Thumb of Not Using Multilevel Modeling: The "Design Effect Smaller Than Two" Rule. *The Journal of Experimental Education, 83*(3), 423–438. https://doi.org/10.1080/00220973.2014.907229