PHP2510

HYPOTHESIS TESTING WEEK 7

CLINICAL TRIAL DESIGN CHALLENGE

This activity is designed to help you explore the relationships between core components of hypothesis testing and statistical power (sample size, effect size, significance level). You will need R for looking up cutoff values. Reading the content in CHIHARA Chapter 8.4 first will help you succeed. You may work in groups if you prefer.

DIRECTIONS

You work for a biotech company that has just developed a new drug to lower anxiety, as measured by the Generalized Anxiety Disorder 7-item (GAD-7) scale, a self-reported survey that provides a total score from 0 ("not at all" bothered across all 7 anxiety-related problems) to 21 (bothered "nearly every day" across all 7 problems). They are planning a Phase III clinical trial and have come to you for help designing the trial.

Based on their Phase II data, you know the following:

- The placebo group shows an anxiety reduction of roughly 1 pt
- The standard deviation of GAD-7 response changes in the population is 10 pts
- The company considers the drug successful if it lowers anxiety by 4+ pts more than the placebo group

EXERCISES

0. Set-up

Write down your basic set-up, specifying the:

- Null hypothesis
- Alternative hypothesis (is this one-sided or two-sided?)
- Test statistic
- Test statistic distribution under the null hypothesis
- Test statistic distribution under the alternative hypothesis

1. Baseline Calculation

Assume the following industry standards:

- Equal group sizes
- 5% significance level (= type 1 error rate); this is the same as saying $p < 0.05$ is "significant"
- 80% power (= 1 - type 2 error rate)

What is the required sample size per group? To calculate this:

1. First, write down the standardized test statistic under the null (that you will use to make decisions).
2. Then, use the significance level to determine the critical value for rejecting the null hypothesis. Use *qnorm* in R.
3. Then, set up a probability statement: the likelihood of rejecting your (standardized) test statistic (per 1) under the alternative hypothesis. What must this equal given the industry standards?
4. What is the distribution of your test statistic under the alternative? Use this fact to rewrite the probability statement in (2) to something you can work with.
5. Now solve for n using *qnorm* again.

Now that you've calculated this by hand, check your work with the built-in R function power.t.test. For the rest of this handout, you may use this function directly.

2. Drug Effectiveness

- What happens to the required sample sizes if the effect size is actually 1 pt less? 2 pts less?

3. Statistical Certainty

- What happens to the required sample size if we want more power to detect the original effect size? Calculate this for power = 90% and power = 95%.
- What happens to the required sample size if the review board requires lower type 1 error rates? Calculate this for a 2.5% and 1% significance level.

4. Budgets

Your boss tells you that the company only has the budget to test the drug on 50 people per group.

- What is the minimum effect size you can reliably detect, still with 80% power and alpha = 0.05?
- Find another choice for power and alpha with this budget.

5. Conclusions & Reflections

- Discuss the pros and cons of the following possible recommendations:
    - Option A: Go ahead with the study as is (What's the risk?)
    - Option B: Relax the significance level (What's the risk?)
    - Option C: Secure more funding to increase the sample size (What's the risk?)
- Comment on the relative scale of the change in sample size required compared to the experimental design changes (i.e. does a 5% change in power lead to a 5% change in sample size%?)
- Comment on the impact of performing one-sided vs two-sided tests to the sample size calculations
- Someone suggests using a different test statistic (perhaps the difference in group medians, rather than group means). How would you determine whether such a test statistic was preferred?