Sample size and statistical power

Saengsit Kritsadee, ATM.Dr., Ph.D.

Going from population to sample

- Populations, parameters & taking a census
- Samples, statistics, and
- Getting a sample from a population
  - Random sampling process
    - Simple random selection of subjects from population
    - Stratified random sampling
    - Cluster/multistage sampling
  - Non-random sampling process
    - Convenience sampling
    - Snowball sampling

Hypothesis testing

- Begin with assumption of “no difference”, and when that is untenable, conclude a difference
  - Ho: $\mu = 100$ H_a: $\mu \neq 100$
- Determine standards of rareness, α
- Assuming Ho, what is μ & σ of sampling distribution?
- Assuming Ho, how rare is observed measure?
- Compare rareness of observed measure to α
  - If observation is rare, conclude Ho is false.
  - If observation is not rare, conclude Ho cannot be rejected.

Hypothesis testing & α

- The normal distribution extends to infinity in both directions.
- We choose our level at which results are "not normal"
- This level, α, expresses how "rare" something has to be to claim something is different
- But because the normal distribution extends to infinity, what we claim as different might not be…and we make a Type 1 error, Pr (Type 1 Error) = α

Hypothesis testing and errors

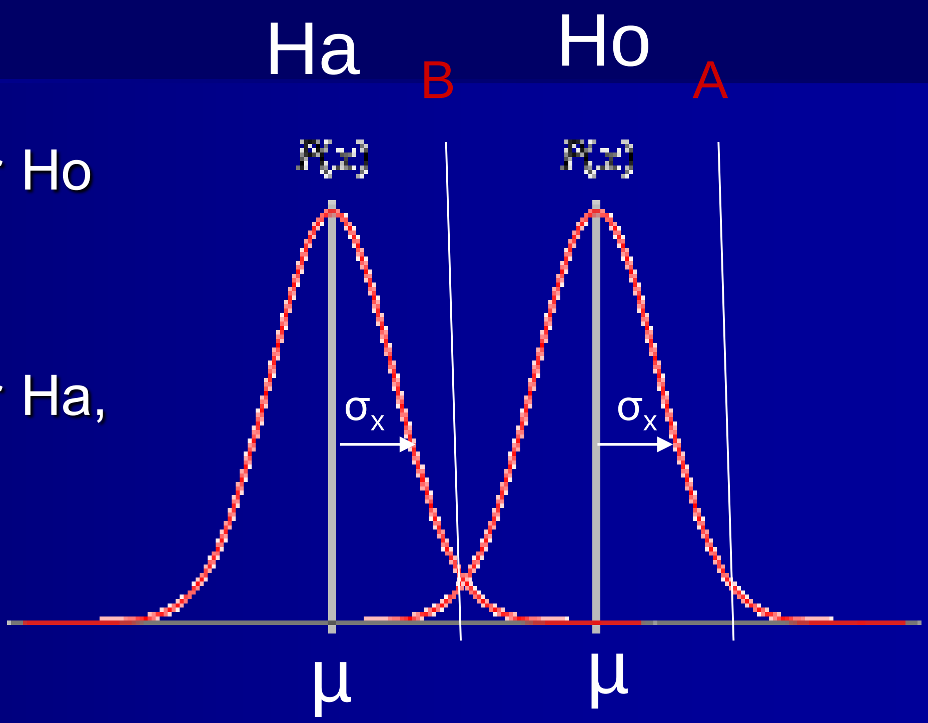
	Ho is True	Ho is False
You decide: Reject Ho	Type 1 Error Probability = α	Correct decision Probability = 1-β
You decide: Fail to reject Ho	Correct decision Probability = 1-α	Type II Error Probability = β

Hypothesis testing and errors

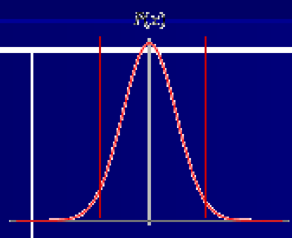
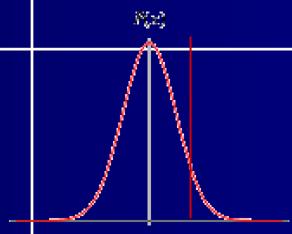
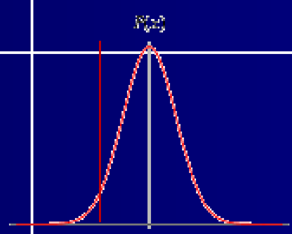
- You set alpha
- A type II error is when you fail to reject the null hypothesis, but you should have rejected it
- Both errors always exist when you test
- As you increase alpha you increase beta and vice versa.

Visualizing α and β

- Pr (Type 1 Error) = Area under Ho to right of A and to left of B
- Pr (Type 2 Error) = Area under Ha, to right of B
- But we don't know Ha



Hypotheses: one & two tailed

Ho: $\mu = 8.0$ Ha: $\mu \neq 8.0$		Prob > $\lvert z \rvert$ Prob > $\lvert z \rvert$	Large (>0.05) Small (<0.05)
Ho: $\mu \leq 8.0$ Ha: $\mu \geq 8.0$		Prob > z Prob > z	Large (>0.05) Small (<0.05)
Ho: $\mu \geq 8.0$ Ha: $\mu \leq 8.0$		Prob < z Prob < z	Large (>0.05) Small (<0.05)

There are two types of error we can commit in hypothesis testing:

- Type I error: Reject H_0 when it is true.
- Type II error: Fail to reject H_0 when it is false.

We control the probability of committing a type I error by choosing a small value for α, e.g. $\alpha = 0.05$

Question: How can we control β, the probability of committing a type II error?

Answer: By controlling the sample size. As the sample size increases the probability of committing a type II error decreases.

Power = 1-β, It is the probability of rejecting H_0 when it is false, i.e. the probability of detecting a true alternative hypothesis.

Assume it is well established that participants in a 6 months smoking cessation program reduce their daily number of cigarettes by 10, on average. The program reduce their daily number of cigarettes by 10, on average. The program coordinators would like to know whether increasing the number of session per week will lead to a significantly greater average reduction in the number of cigarettes smoked per day.

They choose a random sample of 5 new participants and increase the number of sessions from 2 to 3 week. They find that the average reduction in the sample is 15 with a standard deviation of 10.

H_0: The increase is not significant

A one sample t-test yields a p-value greater than 0.1. We fail to reject H_0 and conclude that the increased reduction of cigarettes smoked per day is not significant.

This seems strange, since the increase appears to be important.

Question:

Was the sample size large enough to detect the increase as significant? Did we have enough power to detect the alternative hypothesis?

Answer:

Probably not.

Note:

Power depends on a many factors including the α-level and the magnitude of the effect we would like to detect. What magnitude is important depends on biological considerations or experience. It is not a statistical question.

Note:

The sample size is generally chosen such that the power is greater than 80%.

Why don’t we try to get more than 80% power?

Many studies have limited budgets and the sample size must be kept as small as possible. 80% is considered a reasonable compromise.

Problems with sample size/ power calculations:

- α and β level are completely arbitrary
- The magnitude of the effect we would like to detect is arbitrary
- To determine sample size or power we need an estimate of the standard deviation (if we are estimating a mean), or of the disease rate in the absence of exposure prevalence in the absence of disease) and the relative size of the compared groups (if we are estimating a RR or OR). These estimates are guess work or come from small pilot studies and are often inaccurate.
- The sample size is often predetermined by the availability of eligible study subjects rather than statistical formulas.

Note: Even power curves (plots of power vs. sample size) generally don't alleviate these problems.

solution

Avoid hypothesis testing when possible and use confidence intervals instead.

Even though the level of confidence is arbitrary, confidence intervals are preferable because they provide us with an estimate of the effect and a measure of the precision of the estimate. Even if the sample size is small, confidence intervals still provide us with a lot of information.

Reference

- ไพบูลย์ โล่ห์สุนทร ระบาดวิทยา ภาควิชาเวชศาสตร์ป้องกัน คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย 2540
- Annette Bachand, Introduction to Epidemiology: Colorado State University, Department of Environmental Health
- Leslie Gross Portney and Mary P. Watkins (2000). Foundations of Clinical Research: Applications to Practice. Prentice-Hall, Inc. New Jersey, USA