Biost 517 / Biost 514 Applied Biostatistics I / Biostatistics I

Timothy A. Thornton, Ph.D.
Associate Professor of Biostatistics
University of Washington

Lecture 21:

Power of a Statistical Test: Power and Sample Size Calculations

Hypothesis Testing: Recap

Before analyzing (or sometimes even before collecting) data:

- Specify null hypothesis H_0.
- Specify alternative hypothesis H_A.
- Specify the significance level α.

Data analysis:

- Compute test statistic T on the data.
- Compare observed value of T to the sampling distribution of T when the null hypothesis H_0 is true
- Use the alternative hypothesis to get the appropriate p-value P (one-sided or two-sided).
- Compare P to α and reject the null hypothesis if $P < \alpha$.

Type I and Type II errors

Truth about Population

Decision based on the data

	Null Hypothesis is True	Alternative Hypothesis is True
Do not reject the null hypothesis	Correct Decision	Type II Error
Reject the null hypothesis	Type I error	Correct Decision

Type I and Type II Errors

- P(type I error | H_0) = α
  - "alpha-level" of the test
  - "significance level" of the test
  - "size" of the test
  - α is 1 – Specificity of the test

- P(type II error | H_1) = β
  - P(correct decision | H_1) = 1– β.
  - 1– β is called the "power" of the test
  - Power is Sensitivity of the test
- Lower type 1 error and higher power lead to better PPV, NPV

Type II Errors: Distribution under H_A

- We have mainly focused on the distribution of the test statistic under the null hypothesis. Shouldn't we also consider the distribution under the alternative hypothesis?

	H_0 True	H_A True
Decide H_0	$1-\alpha$	β (type II error)
Decide H_A	α (type I error)	$1-\beta$

α = P(reject H_0 | H_0 true)

power ($1-\beta$) = P(reject H_0 | H_A true)

Study Design: Power and Sample Size

Power = P [reject H_0 | H_A true]

- Power calculations is an important component in study design and indicates the ability of the test procedure to reliably detect departures from the null hypothesis.
- In fact, in the acquisition of obtaining grants from the NIH one must show that the study is capable (has sufficient power) of detecting a meaningful difference, i.e., can discriminate between Hypotheses
- Sample size calculations ensure that the study (and the proposed test statistic) is sufficiently powered to detect departures from the null hypothesis.

Study Design: Power and Sample Size

Power = P [reject H_0 | H_A true]

- Power and sample size calculations require a model for the data under **both** the null and the alternative hypothesis.
- We will derive general formulas for computing power (given sample size) or sample size (given power) for widely used test statistics that we have previously covered

Power: One sample test

- We will first consider the 1-sample testing situation with hypotheses:

$$H_0 : \theta = \theta_0$$
$$H_a : \theta \neq \theta_0$$

- The scenario is enough to illustrate all of the important concepts. Details change when we consider other scenarios such as:
  - One-sided alternatives
  - 2-sample problems
  - Sample proportions
  - More complex scenarios

Power Derivation: One sample test

- For one sample testing of $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$, the test statistic has the form of

$$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{V/n}} \sim N\left(\theta - \theta_0, 1\right)$$

  where V is the variance in the population, e.g., σ^2.
- Under the null hypothesis, $Z \sim N(0, 1)$
- A two-sided test with significance level α would reject $H_0 : \theta = \theta_0$ if $|Z| > z_{1-\alpha/2}$
- Note that $P(|Z| > z_{1-\alpha/2} | H_0 \text{ is true}) = P(|Z| > z_{1-\alpha/2} | \theta_0) = \alpha$
- Power is the probability of rejecting H_0 for a given alternative value of θ, which we will refer to as θ_A:

$$Pow(\theta_A) = P(|Z| > z_{1-\alpha/2} | \theta_A)$$

Power Derivation: One sample test

- Power is the probability of rejecting H_0 for a given alternative value of θ, which we will refer to as θ_A:

$$Pow(\theta_A) = P(Z > z_{1-\alpha/2} | \theta_A) = 1 - \beta$$

- Power depends on θ_0, θ_A, V, and n.

Power Derivation: One sample test

$$
\begin{aligned}
P\left(Z > z_{1-\alpha/2} | \theta_A\right) &= P\left(Z - \frac{\theta_A - \theta_0}{\sqrt{V/n}} > z_{1-\alpha/2} - \frac{\theta_A - \theta_0}{\sqrt{V/n}} | \theta_A\right) \\
&= P\left(\frac{\hat{\theta} - \theta_0}{\sqrt{V/n}} - \frac{\theta_A - \theta_0}{\sqrt{V/n}} > z_{1-\alpha/2} - \frac{\theta_A - \theta_0}{\sqrt{V/n}} | \theta_A\right) \\
&= P\left(\frac{\hat{\theta} - \theta_A}{\sqrt{V/n}} > z_{1-\alpha/2} - \frac{\theta_A - \theta_0}{\sqrt{V/n}} | \theta_A\right) = 1 - \beta \\
&\Longrightarrow z_{1-\alpha/2} - \frac{\theta_A - \theta_0}{\sqrt{V/n}} = z_{1-\beta}
\end{aligned}
$$

Power and Sample Size Calculation: One sample test

- From the results on the previous page, we have the following:

$$\theta_0 - z_{1-\alpha/2}\sqrt{V/n} = \theta_A + z_{1-\beta}\sqrt{V/n}$$

- To get sample size for a given power level $1-\beta$, solve for n:

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 V}{\Delta^2}$$

where $\Delta = |\theta_A - \theta_0|$

- To get power for a given sample size first solve for $z_{1-\beta}$:

$$z_{1-\beta} = \Delta\sqrt{\frac{n}{V}} - z_{1-\alpha/2}$$

then calculate the power with $P\left(Z > z_{1-\beta}\right)$ where Z follows a $N(0,1)$ distribution.

Two-Sample Test: Power Calculations

- Power calculations for a two-sample test can be similarly obtained as this statistic also follows a standard normal distribution under the null hypothesis.

The **two-sample z-statistic** is

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

- To test the hypothesis $H_0 : \mu_1 = \mu_2$ (or equivalently $\mu_1 - \mu_2 = 0$) , we use

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \text{ under } H_0$$

General Power Calculations

- Typical values for power are:

Typical values for power:

$z_{.90}$ = 1.28 (90% power)

$z_{.80}$ = 0.84 (80% power)

- Power and sample sized calculations based on a *t* statistic can be similarly obtained. In this case, critical values used in the calculations are based on a *t* distribution with the appropriate degrees of freedom instead of a normal distribution.
- Power calculations comparing differences in proportions for two samples can be similarly be obtained, as this test is also based on a normal distrubtion

General Power Calculations

Steps for computing power/sample size

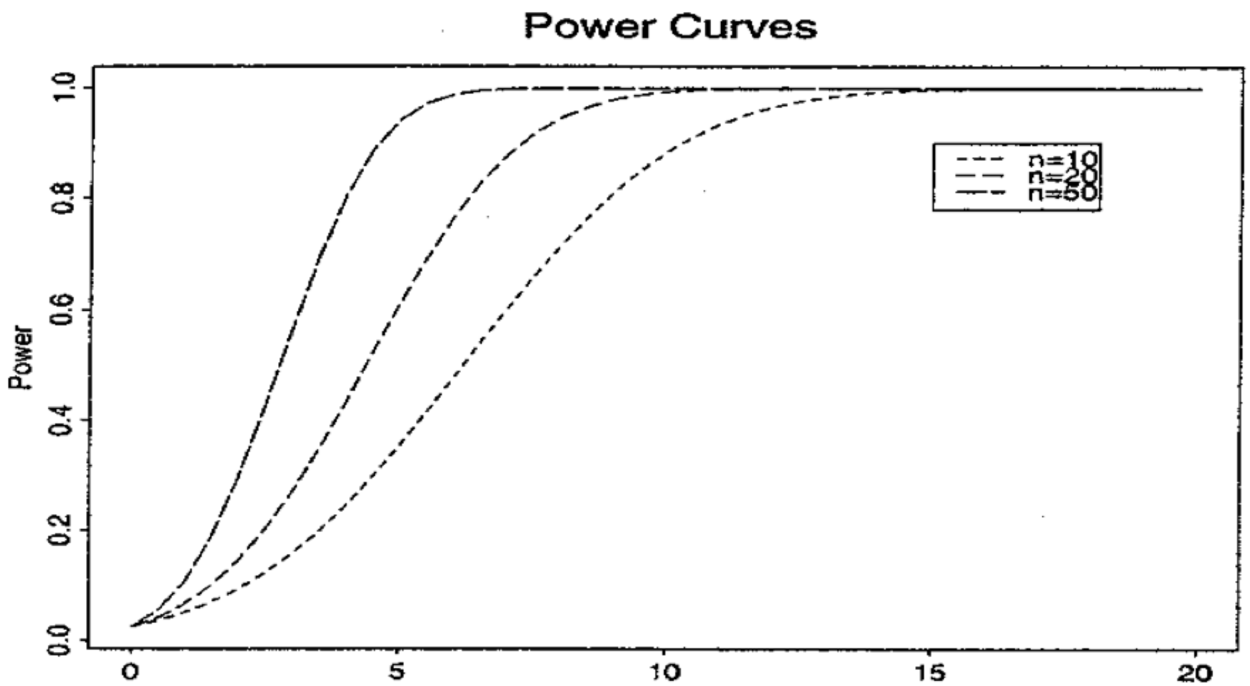
Assume that σ is known. Even if we don't know it, we'll need an estimate of it.

1. Choose α and 1-sided / 2-sided
2. Identify $\boldsymbol{\theta}_0$ (usually obvious)
3. Identify $\boldsymbol{\theta}_A$ (many possible values; how to choose?)
4. Compute n (given power) or power (given n)

Note that 1 & 2 determine what the rejection region will be…

Power Curves

- It is often useful to display a Power Curve with different values of $|\theta_A - \theta_0|$ for different sample sizes considered.



$|\theta_A - \theta_0|$

Power Calculations in R

- The **power.t.test()** function in R can be used to compute the power of the one- or two- sample t test, or determine parameters to obtain a target power.

- Suppose we are testing blood pressure medication among hypertensive men. We want to design a study that can detect a clinically relevant decline in blood pressure due to the treatment. Suppose 5 mmHg is considered clinically relevant and I enroll 50 participants in the study. Suppose the standard deviation of changes is 10. What is the power for an α= 0.05, 2-sided test?

Example 1: Power Calculations in R

- Suppose we are testing blood pressure medication among hypertensive men. We want to design a study that can detect a clinically relevant decline in blood pressure due to the treatment. Suppose 5 mmHg is considered clinically relevant and I enroll 50 participants in the study. Suppose the standard deviation of changes is 10. What is the power for an α= 0.05, 2-sided test?

```
power.t.test(n=50,delta=5,sd=10,sig.level=0.05,type="one.sample",alternative="two.sided")
```

```
     One-sample t test power calculation

              n = 50
          delta = 5
             sd = 10
      sig.level = 0.05
          power = 0.9338976
    alternative = two.sided
```

Example 2: Sample Size Calculation in R

- Same setting as in example 1. How many patients are required to obtain 80% power using a 2-sided $\alpha = 0.05$ test

```
power.t.test(power=.80,delta=5,sd=10,sig.level=0.05,type="one.sample",alternative="two.sided")
```

```
     One-sample t test power calculation

              n = 33.3672
          delta = 5
             sd = 10
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

- So we should recruit n=34 patients into our study

Factors that influence Power

$$z_{1-\beta} = \Delta\sqrt{\frac{n}{V}} - z_{1-\alpha/2}$$

- Power is expected to increase as:
  - Sample size increases.
  - Distance between $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_0$ ($\boldsymbol{\Delta}$) increases.
  - Variance gets smaller.
  - Significance level α gets larger (easier to reject)

Factors that influence Sample Size

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 V}{\Delta^2}$$

- The expected sample size increases as:
  - Power increases
  - Distance between $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_0$ ($\boldsymbol{\Delta}$) decreases
  - Variance increases
  - Significance level α decreases

Summary

- Power is an important component in study design.
- Sample size calculations ensure that the study is capable of detecting departures from the null hypothesis.
- Power and Sample size require more than a test. A model for sampling distribution under both the null and the alternative hypotheses are also required.