Power and Sample Size

BIOS 6611

CU Anschutz

Week 4

Review of Important Definitions

Detectable Difference, Power, and Sample Size

Often in planning studies we are asked *"How many subjects do I need for the study?"* Or, *"If I have 30 subjects, do I have enough power to see a significant result?"*

These questions are not as readily answered as they might seem. The answers depend on several factors, which we will now examine. Finding rejection regions and p-values requires calculating probabilities assuming H_0: $\mu = \mu_0$ is true. To find power we need to consider the flip side: H_0 is not true, and some specific alternative hypothesis H_1 is true.

Generic Example: A collaborating investigator says, "I want to do a study where I compare my new method to the standard method. How many subjects do I need? I think I can get 50, is that enough?"

Important Probabilities and Definitions

Recall, based on the data we make a decision to reject H_0 or to not reject H_0, and we quantify the evidence against H_0 in the form of a p-value.

	Reality	
What we decide	**H_0 True**	**H_0 False/H_1 True**
Fail to reject H_0	*Correct* Probability of correct decision $=$ $1-\alpha =$ level of confidence	*Type II Error* P(Type II Error) $= \beta$
Reject H_0	*Type I Error* P(Type I Error) $= \alpha$ (**Level of significance**)	*Correct* Probability of correct decision $=$ $1-\beta =$ **Power**

Type I Error, Type II Error, and Power

	$\mathbf{H_0}$ **True**	$\mathbf{H_0}$ **False**
Fail to reject $\mathbf{H_0}$	$1-\alpha$ Level of confidence	β P(Type II Error)
Reject $\mathbf{H_0}$	α Level of significance P(Type I Error)	$1-\beta$ Power: Probability of finding difference if it exists

P(Type I Error) $= \alpha =$ Probability of rejecting the null hypothesis when it is true; P(Reject $H_0 | H_0$ is true)

P(Type II Error) $= \beta =$ Probability of failing to reject the null hypothesis when it is false; P(Fail to Reject $H_0 | H_0$ is false)

Power $= 1 - \beta =$ Probability of rejecting the null hypothesis when it is false; P(Reject $H_0 | H_0$ is false)

Approach: We want both α and β to be small. Since β increases as α decreases, this is not a well-defined problem.

Power Calculation Derivation for a Two-Sided, One-Sample Z-test

Our Motivating Context

For our derivations we will assume the population standard deviation, σ, is known. Therefore we can work with the standard normal distribution (i.e., $Z \sim N(0, 1)$) to derive closed-form equations for a one-sample two-sided Z-test.

In a different lecture we will see examples where we assume σ is unknown as estimated as s. To reflect this uncertainty, we would use the t-distribution with $df = n - 1$, which leads to an iterative process to estimate power, sample size, etc.

The 5 Important Quantities

When designing a study we generally have 5 important quantities to consider:

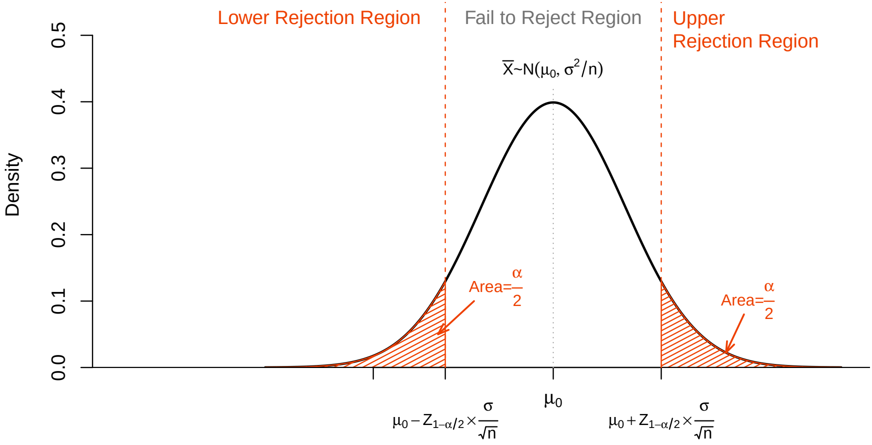
1. Level of significance of a test $= \alpha$
2. Power of a test $= 1 - \beta$
3. Sample size $= n$
4. Effect size (e.g., detectable difference in means $|\mu_0 - \mu_1|$)
5. Estimate of variability: σ^2

We usually set α (typically at 0.05) and assume a known σ^2. Then, by fixed two of the other quantities, we can solve for the last remaining one.

Some statistical methods will have additional quantities to consider.

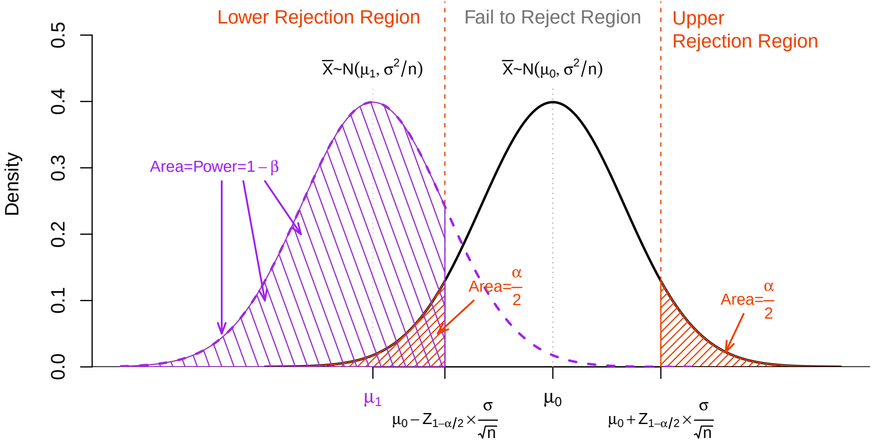
Derivation of Power for a Two-Sided One-Sample Z-test

Assume we want to test H_0: $\mu = \mu_0$ vs. H_1: $\mu = \mu_1$.



Derivation of Power for a Two-Sided One-Sample Z-test

Assume we want to test H_0: $\mu = \mu_0$ vs. H_1: $\mu = \mu_1$.



Check out https://rpsychologist.com/d3/nhst/ for an interactive version!

Derivation of Power for a Two-Sided, One-Sample Z-test

The power, P(Reject $H_0 | H_0$ False), to detect the difference $|\mu_0 - \mu_1|$ is

$$
\begin{aligned}
\text{Power} =& P(\text{Reject } H_0 | \mu = \mu_1) \\
=& P\left(\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ OR } \bar{X} > \mu_0 + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_1 \right) \\
=& P\left(\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_1 \right) + P\left(\bar{X} > \mu_0 + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_1 \right) \\
=& P\left(\bar{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_1 \right) + \approx 0 \\
=& P\left(\frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} < \frac{(\mu_0 - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) - \mu_1}{\frac{\sigma}{\sqrt{n}}} \middle| \mu = \mu_1 \right) \\
=& P\left(Z < \frac{\mu_0 - \mu_1}{\frac{\sigma}{\sqrt{n}}} - Z_{1-\frac{\alpha}{2}} \right) \\
=& \Phi(Z < Z_{1-\beta}) \\
=& 1 - \beta
\end{aligned}
$$

Two-Sided One-Sample Z-Test Calculations

Solve for Power

Find the *power* of a test with n subjects, a given α, and known σ to detect a difference in means of $|\mu_0 - \mu_1|$:

$$Z_{1-\beta} = \frac{|\mu_0 - \mu_1|}{se(\bar{X})} - Z_{1-\frac{\alpha}{2}} = \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} - Z_{1-\frac{\alpha}{2}} \implies$$

$$1 - \beta = \Phi\left[\frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} - Z_{1-\frac{\alpha}{2}}\right]$$

On the next slide we will see we can rearrange these equations to solve for other quantities of interest, such as n or the detectable difference.

Solve for n or Detectable Difference

Find the *number of subjects* needed to detect a difference of $|\mu_0 - \mu_1|$ with α type I error, $1 - \beta$ power, and known σ:

$$n = \frac{\sigma^2 \left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}} \right)^2}{(\mu_0 - \mu_1)^2}$$

Find the *difference in means* that can be detected with a given n, α type I error, $1 - \beta$ power, and known σ:

$$|\mu_0 - \mu_1| = (Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}) \times \frac{\sigma}{\sqrt{n}}$$

Example - Sample Size

A new calcium channel blocker is to be tested for treatment of unstable angina, with an outcome of change in heart rate after 48 hours. A researcher wants to know the sample size needed to detect a difference of 5 bpm with σ=10 bpm, $\alpha = 0.05$, and 80% power ($1 - \beta = 0.8$).

$$
\begin{aligned}
n &= \frac{\sigma^2 \left(Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}\right)^2}{(\mu_0 - \mu_1)^2} \\
&= \frac{10^2 (Z_{0.8} + Z_{0.975})^2}{5^2} \\
&= \frac{10^2 (0.84 + 1.96)^2}{25} \\
&= 31.36
\end{aligned}
$$

The investigator will need 32 participants. (Note, in order to achieve *at least* the desired power we **must** round up, even if the estimate was $n = 31.01$. Otherwise we would be slightly **underpowered**.)

Example - Power

Suppose the same investigator expects they can only enroll 20 participants. Assuming the same difference of 5 bpm, $\sigma = 10$ bpm, and $\alpha = 0.05$, what is their power in this scenario?

$$\begin{aligned} Z_{1-\beta} &= \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} - Z_{1-\frac{\alpha}{2}} \\ &= \frac{5}{\frac{10}{\sqrt{20}}} - Z_{0.975} \\ &= 2.236 - 1.96 \\ &= 0.276 \end{aligned}$$

Next we use the CDF of the standard normal distribution to calculate the power at $Z_{1-\beta} = 0.276$:

$\Phi(0.276) = 0.608726 \approx 0.61 = 1 - \beta =$ Power

A sample size of 20 provides approximately 61% power to detect a difference of at least 5 bpm, given that it exists.

Example - Detectable Difference

The investigator is concerned about their low power on the previous slide. They ask what is their detectable difference if $n = 20$, $\sigma = 10$ bpm, $\alpha = 0.05$, and they want 80% power?

$$
\begin{aligned}
|\mu_0 - \mu_1| &= (Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}) \times \frac{\sigma}{\sqrt{n}} \\
&= (Z_{0.8} + Z_{0.975}) \times \frac{10}{\sqrt{20}} \\
&= (0.84 + 1.96)(2.236) \\
&= 6.26
\end{aligned}
$$

With a sample size of 20, there is 80% power to detect a difference of at least 6.26 bpm, given that it exists.

Additional Considerations

One-Sided Tests

The power calculation for a one-sided test looks very similar to our previous formula, but instead of $\frac{\alpha}{2}$ (i.e., half the α in each tail), we put it all in the one tail.

For H_1: $\mu_1 > \mu_0$:

$$Z_{1-\beta} = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} + Z_{\alpha} = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - Z_{1-\alpha} \implies 1-\beta = \Phi\left[\frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - Z_{1-\alpha}\right]$$

For H_1: $\mu_1 < \mu_0$:

$$Z_{1-\beta} = \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + Z_{\alpha} = \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - Z_{1-\alpha} \implies 1-\beta = \Phi\left[\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - Z_{1-\alpha}\right]$$

What a Power Calculation Does (and Doesn't) Do

- Power calculations are **not** guarantees of a successful outcome for future research!
- Power/sample size help us to determine if a certain research study is feasible based on the assumed scenario.
- The power calculation serves as an educated guess. Even if we knew the underlying population distribution, we are still taking a *sample*, and therefore will have variability around our estimates.
- Power calculations are generally required for most funded research studies (and should be carried out for most unfunded studies as well).
- In practice we often calculate power over a range of (realistic) scenarios.