Statistical Power

Objectives

Define Power and relate it to the Type II error.

Explain how the following factors contribute to power: sample size, effect size (difference between sample statistics and statistic formulated under the null), and significance level.

Compute power given a dataset and a problem.

Identify what can be done to increase power.

Estimate sample size required of a test (power analysis) for one sample mean or proportion case.

Hypothesis Testing (from yesterday)

1. Define null and alternative hypotheses.
2. Assume that the null hypothesis is true.
3. Collect evidence to disprove this assumption.
4. Outcome: we “reject the null hypothesis” or “fail to reject the null.”

Hypothesis Testing

1. State null hypothesis (H_0) and alternative hypothesis (H_A)

   H_0: $\mu = 100$

   H_A: $\mu \neq 100$

2. Choose significance level, α

   $\alpha = 0.05$ (usually)

3. Compute appropriate test statistic using collected data.

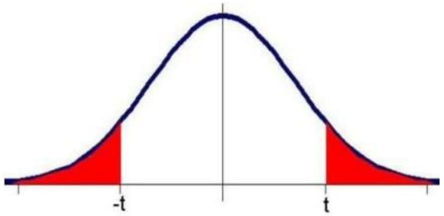
   t-statistic (with appropriate treatment of sample sizes/variance)

4. Compute p-value based on test statistic.

   Two-sided test of t-distribution

5. Reject or fail to reject null.

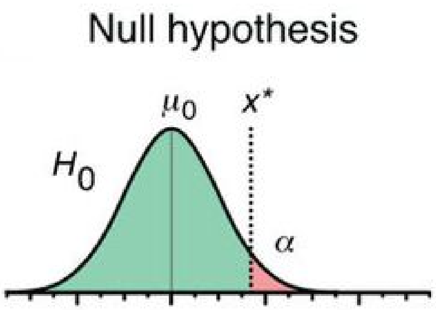
   P[reject H_0 | H_0 is true] = α



Hypothesis Testing

	H_0 is true	H_0 is false
Fail to reject H_0	Correct Decision $(1 - \alpha)$	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Decision $(1 - \beta)$

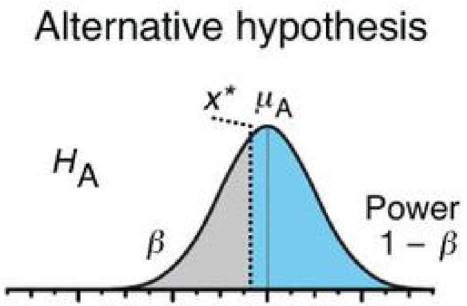
Hypothesis Testing



	H_0 is true	H_0 is false
Fail to reject H_0	Correct Decision (1 - α)	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Decision (1 - β)

Power Calculation

	H_0 is true	H_0 is false
Fail to reject H_0	Correct Decision $(1 - \alpha)$	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Decision $(1 - \beta)$



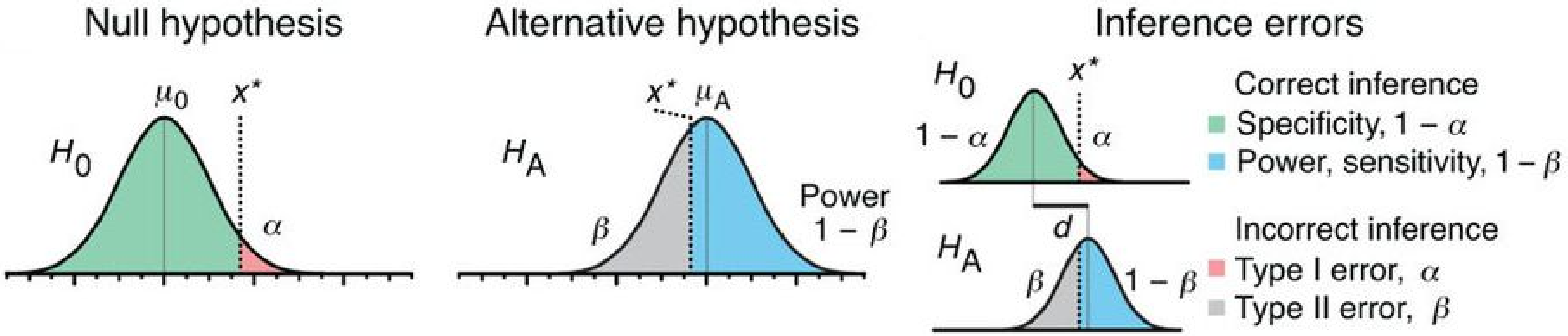
Power Calculation

1. Assume that the null hypothesis (H_0) is ***false***.
2. Given:
   a. Significance level, α
   b. Effect size, $\mu_A - \mu_0$
   c. Sample size, n
   d. Standard deviation, s

Calculate Power: P[reject H_0 | H_0 is false] = $1 - \beta$

Hyp. Testing & Power Calc.

	H_0 is true	H_0 is false
Fail to reject H_0	Correct Decision (1 - α)	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Decision (1 - β)



Power Calculation Example

1. Assume that the null hypothesis (H_0) is ***false***.
2. Given: (for AB-test of website)
   a. Significance level, $\alpha = 0.05$
   b. Effect size, $\mu_0 = 0.06$, $\mu_A = 0.07$
   c. Sample size, $n = 5000$
   d. Standard deviation, $s = 0.237$ (known since $s^2 = p(1-p)$ for proportions)

Calculate Power: $P[\text{reject } H_0 \mid H_0 \text{ is false}] = 1 - \beta$

Power Calculation

$\alpha = 0.05$, $\mu_0 = 0.06$, $\mu_A = 0.07$, $N = 5000$, $s = 0.237$

Calculate Power: P[reject H_0 | H_0 is false] = 1 - β

1. Calculate "critical value" for rejecting H_0 X^*:

$$Z_\alpha \leq Z = \frac{X - \mu_0}{s/\sqrt{n}} \; or \; \mu_0 + Z_\alpha \frac{s}{\sqrt{n}} = X^* \leq X$$

2. Now calculate the Z-score and p-value of the "critical value" under H_A

Power Calculation

$\alpha = 0.05$, $\mu_0 = 0.06$, $\mu_A = 0.07$, $N = 5000$, $s = 0.237$

Calculate Power: P[reject H_0 | H_0 is false] = 1 - β

1. Calculate "critical value" for rejecting H_0 X^*:

$$Z_\alpha \le Z = \frac{X - \mu_0}{s/\sqrt{n}} \;\; or \;\; \mu_0 + Z_\alpha \frac{s}{\sqrt{n}} = X^* \le X$$

(For this example, use one-sided test of proportions $\alpha=0.05 \Rightarrow Z_\alpha=1.645$)

$\boxed{X^* = 0.0655}$

2. Now calculate the Z-score and p-value of the "critical value" under H_A

Power Calculation

$\alpha = 0.05$, $\mu_0 = 0.06$, $\mu_A = 0.07$, $N = 5000$, $s = 0.237$

Calculate Power: P[reject H_0 | H_0 is false] = 1 - β

1. Calculate "critical value" for rejecting H_0 X^*:

$$Z_\alpha \le Z = \frac{X - \mu_0}{s/\sqrt{n}} \; or \; \mu_0 + Z_\alpha \frac{s}{\sqrt{n}} = X^* \le X$$

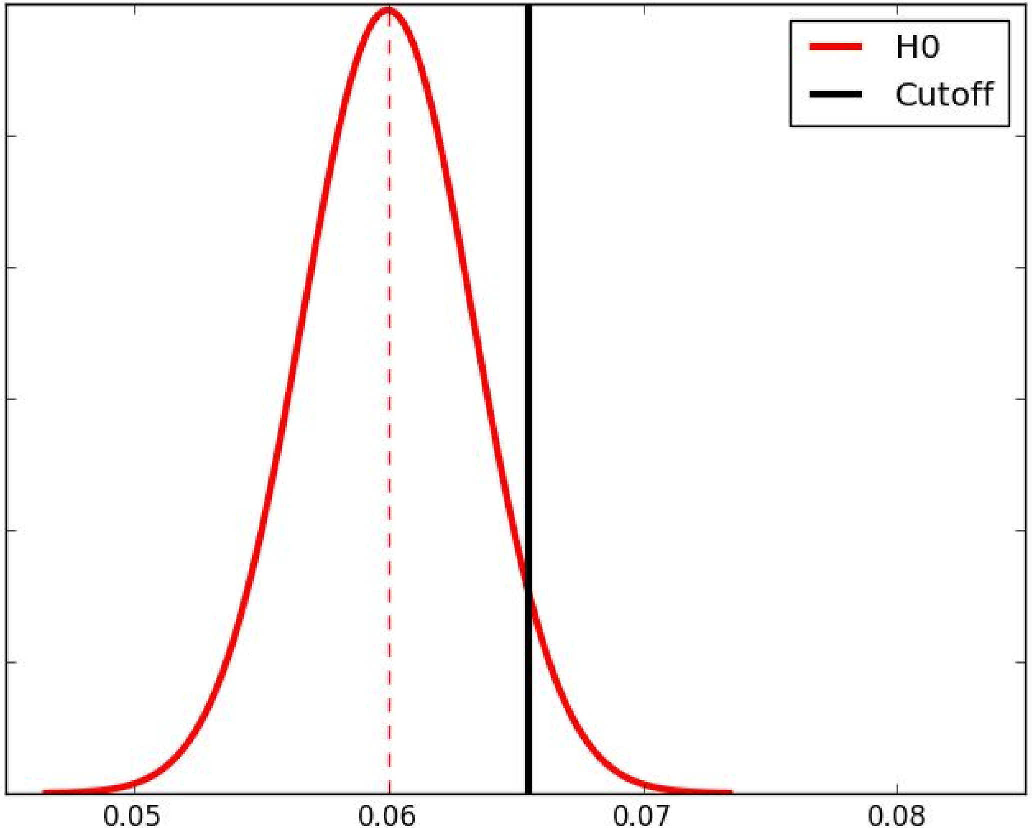
(For this example, use one-sided test of proportions α=0.05 ⇨ Z_α=1.645)

$\boxed{X^* = 0.0655}$

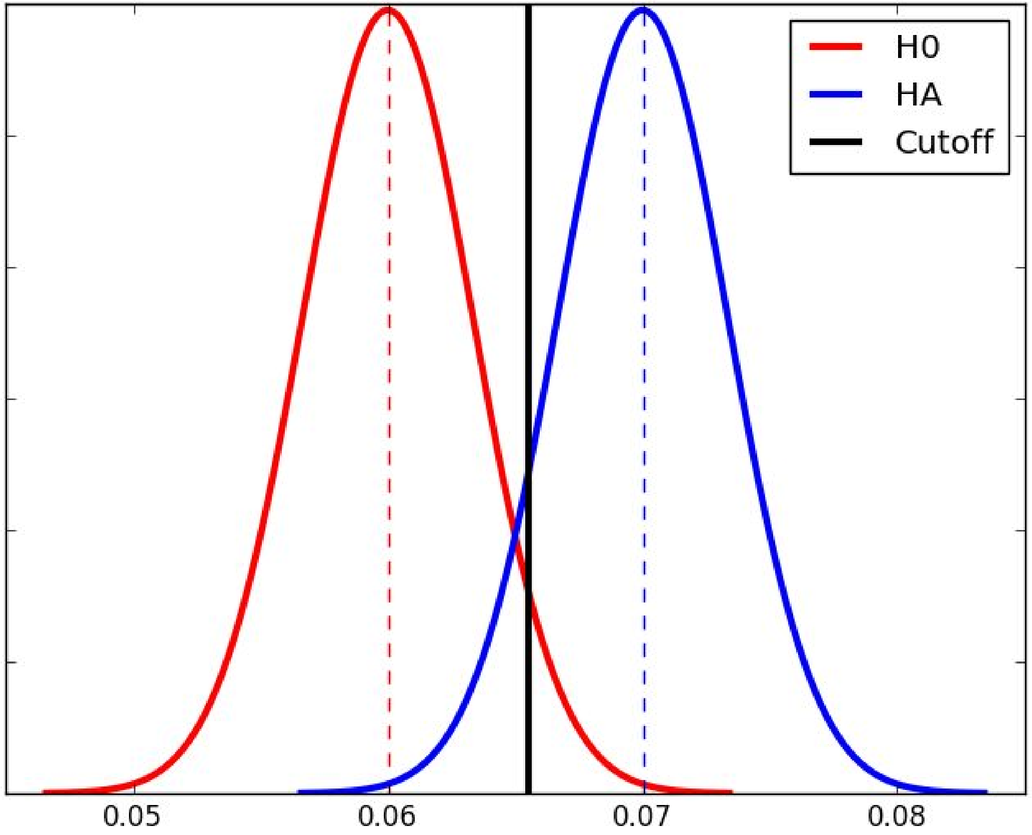
2. Now calculate the p-value of the "critical value" under H_A
P[μ_A >= X^*] = 0.91

Power of this test is 91%

Power Calc
Step 1.



Power Calc
Step 2.



Relating Power and Significance Level

First, we reject H_0 when:

$$Z_\alpha \le Z = \frac{X - \mu_0}{s/\sqrt{n}} \quad or \quad \mu_0 + Z_\alpha \frac{s}{\sqrt{n}} = X^* \le X$$

Then, we find the corresponding cut-off of this value under H_A is:

$$X^* = \mu_1 + Z_{1-\beta} \frac{s}{\sqrt{n}} = \mu_1 - Z_\beta \frac{s}{\sqrt{n}}$$

$$\boxed{Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}}$$

Define 4 variables, solve for the remaining 1.

- $$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

The equation above links the following variables:

- α (type I error; significance level)
- β (type II error; $\pi = 1 - \beta$, the statistical power)
- $\mu_1 - \mu_0$ (effect size)
- s (standard deviation)
- n (sample size)

Some Experimental Design Questions

After choosing significance level and power, what effect size can I distinguish with a sample of N subjects?

After choosing significance level and power, how many subjects do I need to observe to be able to identify a particular effect size?

Conceptual Questions

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If I decide that I need a more stringent significance level cutoff (e.g. α = 0.05 to 0.01) what happens to the power of the experiment (assuming everything else stays constant)?

What is the primary means available to increase the power of an experiment?

Conceptual Questions

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If I decide that I need a more stringent significance level cutoff (e.g. $\alpha = 0.05$ to 0.01) what happens to the power of the experiment (assuming everything else stays constant)? **-- Power is reduced**

What is the primary means available to increase the power of an experiment?

Conceptual Questions

$$Z_\alpha + Z_\beta = \frac{\mu_1 - \mu_0}{s/\sqrt{n}}$$

If I decide that I need a more stringent significance level cutoff (e.g. $\alpha = 0.05$ to 0.01) what happens to the power of the experiment (assuming everything else stays constant)? **-- Power is reduced**

What is the primary means available to increase the power of an experiment?

-- Increase the sample size

Objectives

Define Power and relate it to the Type II error.

Explain how the following factors contribute to power: sample size, effect size (difference between sample statistics and statistic formulated under the null), and significance level.

Compute power given a dataset and a problem.

Identify what can be done to increase power.

Estimate sample size required of a test (power analysis) for one sample mean or proportion case.