Sample Size Calculation

Joyce

BITTIGER

来自硅谷的终身学习平台

Sample Size Questions

!!! Sample size calculation is an VERY important question in A/B testing !!!

- Type of questions
  - What is the minimum sample size?
  - What factors will impact your sample size?
  - How long are you going to run your experiment?
  - What's your roll-out plan?
  - What kind of issues may you anticipate in an A/B testing experiment?

Data Assumptions

- What distribution assumptions are you making to your data?

  i.i.d. Normal distribution, Central Limit Theorem

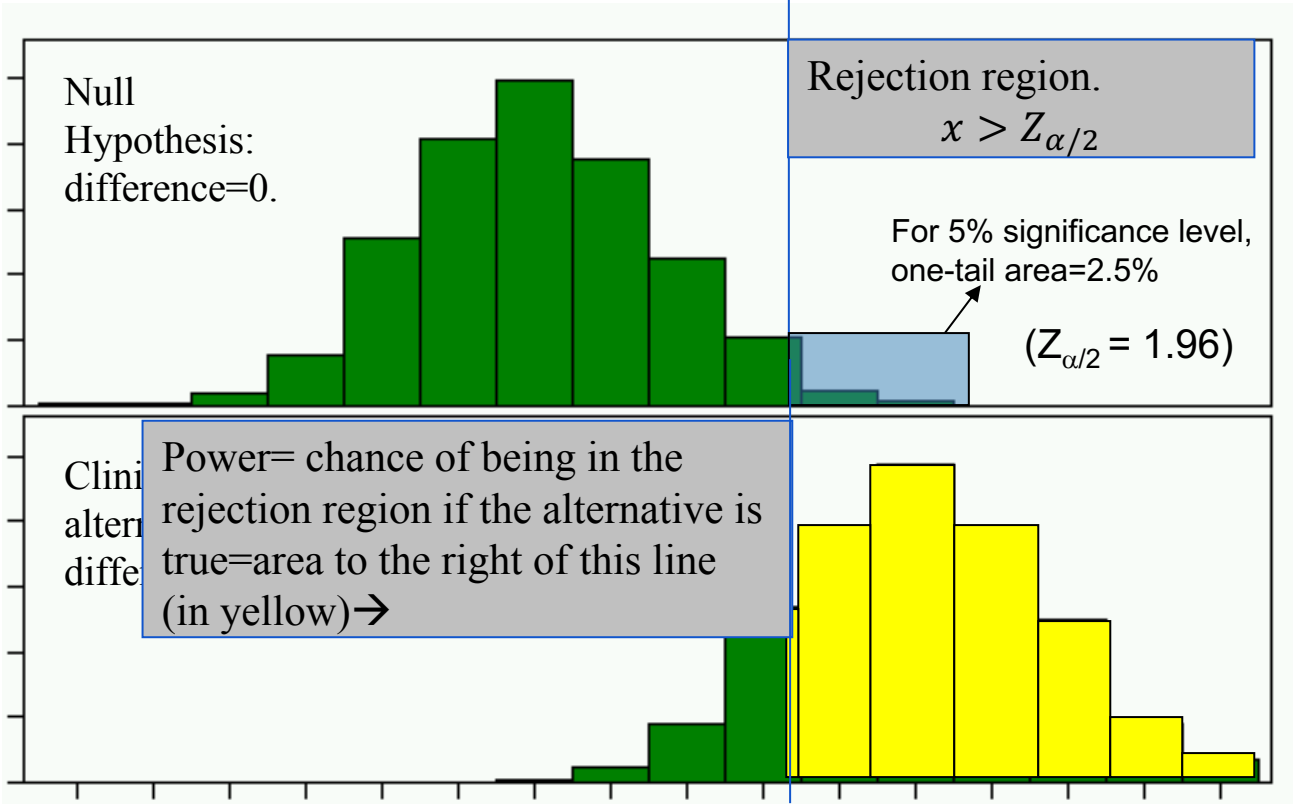
- What is the null hypothesis of your test?

  $diff = \mu_A - \mu_B = 0$

Power Analysis

$diff \sim N(0,1)$

$diff \sim N(3,1)$



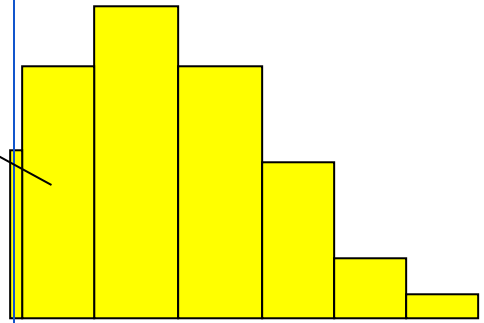
Power Analysis

Rejection region. Any value >= 6.5 (0+3.3*1.96)
Any value >= $Z_{\alpha/2}$

Power= chance of being in the rejection region if the alternative is true=area to the right of this line (in yellow)

Power here:

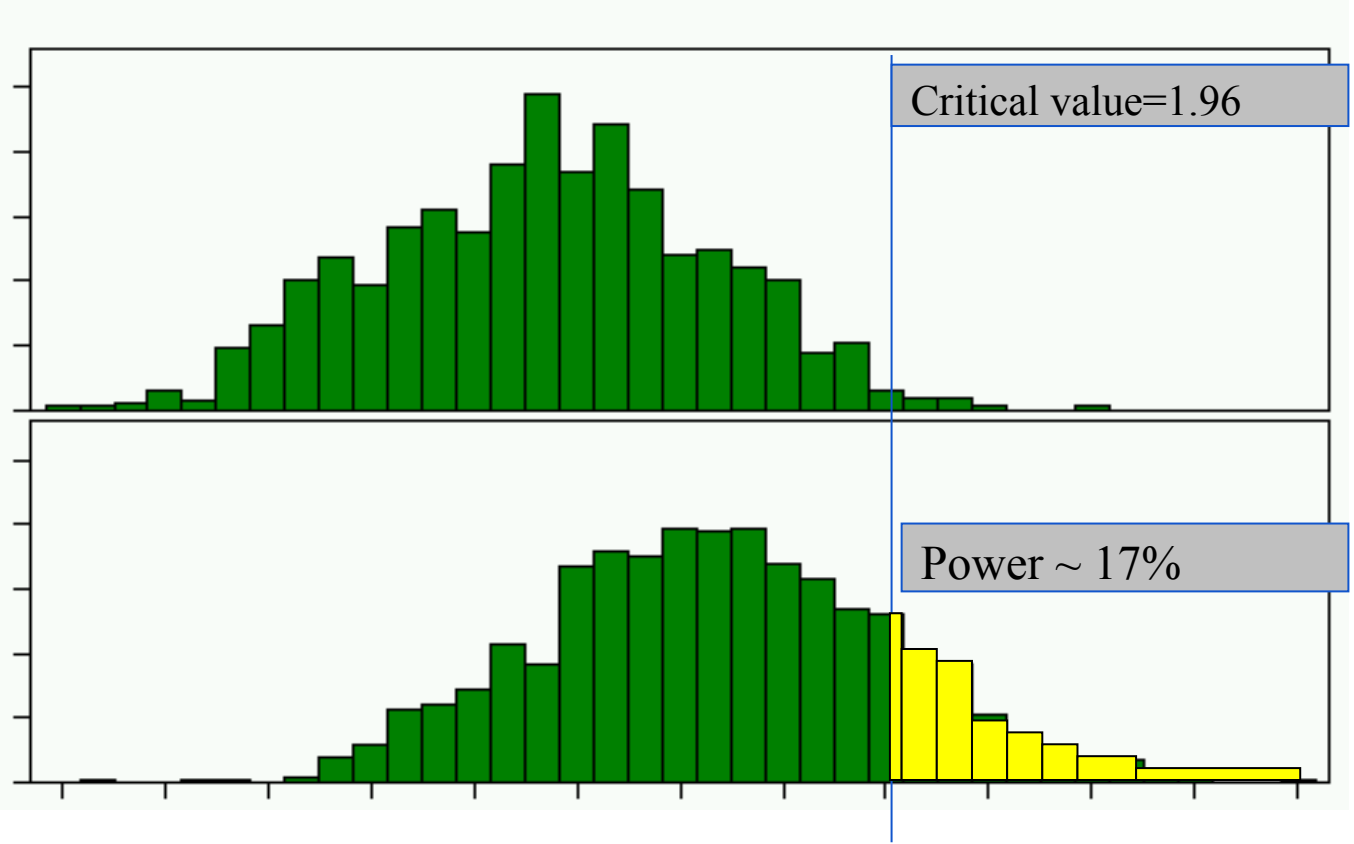
$$P(X > 1.96 \mid \mu = 3, \sigma = 1)$$
$$= P(Z > \frac{1.96 - 3}{1})$$
$$= 85\%$$



Power Analysis

$diff \sim N(0,1)$

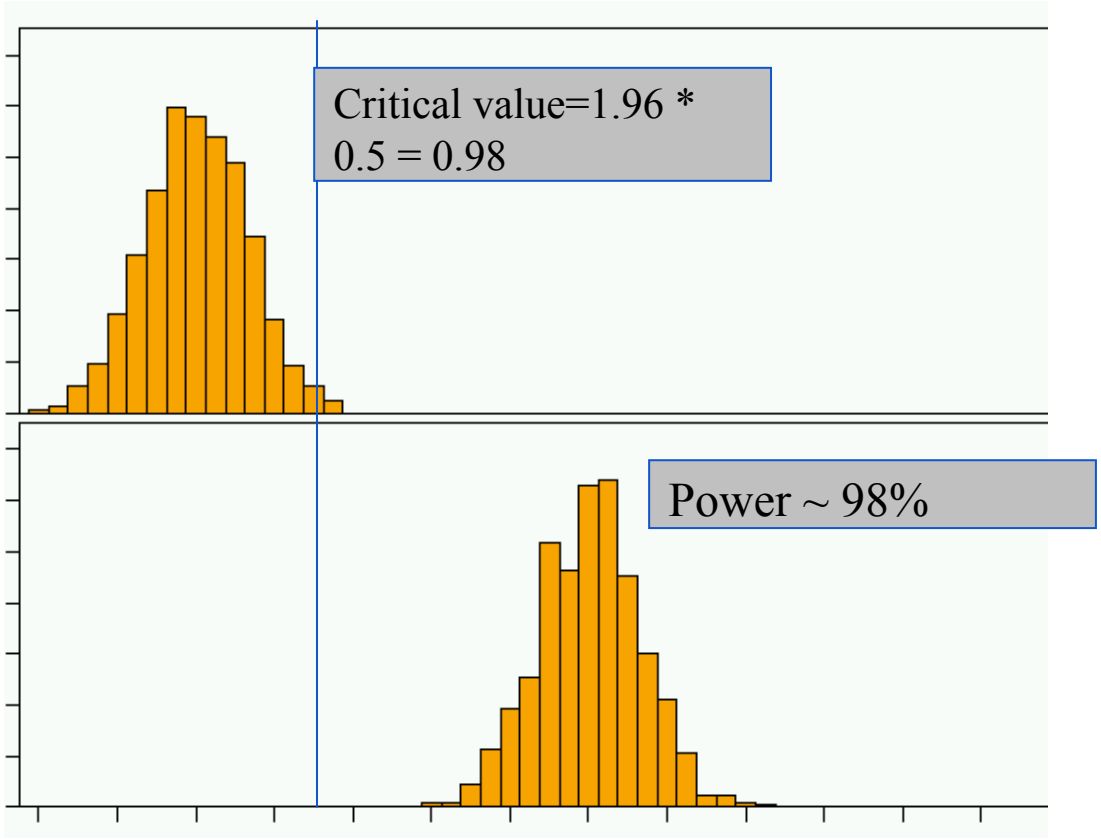
$diff \sim N(1,1)$



Power Analysis

$diff \sim N(0, 0.5)$

$diff \sim N(2, 0.5)$



Factors Impact Power

How is the power change if the following factors increase?

1. Size of the effect ↑
2. Variance of distribution ↓
3. Significance level desired α ↓

Sample Size Calculation

You are designing an experiment to evaluate two versions of website.

Assume the population has mean μ and variance σ^2 . You randomly selected n (n>30) samples from test/control group respectively.

What is the sample distribution?

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

More samples, less variance, higher power

Sample Size Calculation

Represents the desired power (typically .84 for 80% power).

Sample size in each group (assumes equal sized groups)

$$n = \frac{2\sigma^2 (Z_\beta + Z_{\alpha/2})^2}{\text{difference}^2}$$

Standard deviation of the outcome variable

Effect Size (the difference in means)

Represents the desired level of statistical significance (typically 1.96 for 95%).

Sample Size Calculation

For given β (power), α(significance level), σ (standard diviation of data)

$$Z_\beta = \frac{critical\ value - diff}{standard\ error(diff)} = \frac{z_{1-\alpha/2} * SE(diff) - diff}{SE(diff)}$$

$$= -z_{\alpha/2} - \frac{\text{diff}}{\text{SE(diff)}} = -z_{\alpha/2} - \frac{\text{diff}}{\sqrt{2\sigma^2/n}}$$

$$\therefore \text{n} = \frac{2\sigma^2\left(Z_\beta + Z_{\alpha/2}\right)^2}{diff^2}$$

$$\text{SE(diff)} = \sqrt{Var(diff)} = \sqrt{\text{Var}(\overline{X_a} - \overline{X_b})}$$

$$= \sqrt{\text{Var}(\overline{X_a}) + Var(\overline{X_b})} \text{ as } \text{X}_\text{a} \text{ and } \text{X}_\text{b} \text{ are independent}$$

$= \sqrt{2\sigma^2/n}$ n is sample size of one group, assuming two groups have equal sample size

If not equal variance, $\text{SE(diff)} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$

Questions

- What is the minimum sample size?

  Calculate use the formula

- What factors will impact your sample size?

  Power, significance, effect size, variance

- How long are you going to run your experiment?

  Calculate minimum sample size, daily volume, also need to consider seasonality

- What's your roll-out plan?

  Roll out gradually, 5% -> 10% -> 50%

- What kind of issues may you anticipate in an A/B testing experiment?

  Not enough samples, imbalanced samples