Audacity example:

creates online finance courses ⇒ user engagement.

User flow: customer funnel

- homepage visits
- Exploring sites
- Create account
- Complete.

Experiment

Hypothesis: changing "button" color will increase how many students explore Audacity courses.

Metric Choice

✗ • Total # of courses completed. (too long)

✓ • $\frac{\text{\# of clicks}}{\text{\# of pageviews}}$ = CTR, can > 1. ⇒ measure the usability of a button

✓ • $\frac{\text{\# of unique visitors who click}}{\text{\# of unique visitors to pages}}$ = Click Through probability.

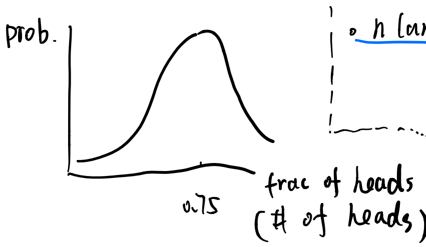
⇒ how often users go to 2nd page

✓ better for funnel

Data Distribution

Binomial Distribution.

- (¢) $p = \frac{3}{4}$. suc / fail



- n large → normal

$p = \mu$

$\sqrt{\frac{p(1-p)}{n}}$ = std.

- $N = 20$. $X = 16$ $\hat{p} = \frac{16}{20} = \frac{4}{5}$

When can use binomial?

① 2 types of outcomes
② Independent events
③ Identical distribution. p for all

✗ click a search page (not independent)

Confidence Interval

95% CI: if we theoretically repeated the experiment we would expect the interval we construct around sample mean to cover the true value in the population 95% of time

Calculate:

$\hat{p} = \frac{X}{N} = 0.1$

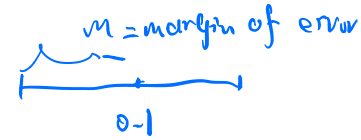
✓ To use normal: Check $N \cdot \hat{p} > 5$ & $N(1-\hat{p}) > 5$

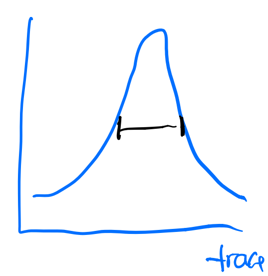
$m = Z * SE$

$m = Z * \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$

$[\hat{p} \pm m]$

$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$





Statistical Significance

Hypothesis Testing

$P(\hat{p}_{exp} - \hat{p}_{cont} \mid H_0)$: rej null Hyp

p-value

⇒ practical : 1%~2% change in CTR is significant.

Compare two samples

$$\hat{P}_{pool} = \frac{X_{cont}\ X_{exp}}{N_{cont} + N_{exp}}, \quad SE_{pool} = \sqrt{\hat{P}_{pool} * (1-\hat{P}_{pool}) \times \left(\frac{1}{N_{exp}} + \frac{1}{N_{cot}}\right)}$$

$$\hat{d} = \hat{P}_{exp} - \hat{P}_{cot}$$

H_0: $d=0$ $\quad \hat{d} \sim N(0, SE_{pool})$

If $\hat{d} > 1.96 \times SE_{poo}$ / $< -1.96 \times SE_{pool}$ ⇒ reject null.

⇒ larger exp. ⇒ more page views.

effect Size vs Power Trade-off.

CI ↑

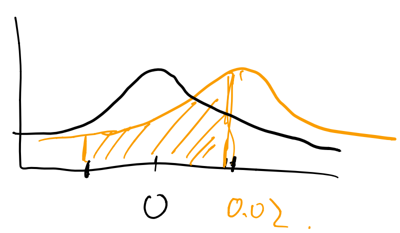
make sure that we have enough power to conclude with high prob that the result is infact statistical significant.

Design

Sample size, how many page views?

$\alpha = P(\text{rej} \mid \text{null true})$

$\beta = P(\text{fail to rej} \mid \text{null false})$.



Sample
- small: α low, β high
- large: α same, β lower

effect size:
- how large the difference that you want to detect from the testing.
- smaller effect size is harder to detect.
- you want when the H_a is true, you can rej H_0 correctly so that the diff is detected.

practical significance boundry

Minimum Detectable ~~Effect~~

(d_{min})

(dmin) ..... is a way of quantifying the size of difference between two groups.

power: - the probability of making a correct decision when H_0 is false.
- the probability that a test of significance will pickup on an effect that is present.

Analyze

$N_{cont} = 10{,}072$ $\quad$ $N_{exp} = 9886$ $\quad$ [practicee significance level] $d_{min} = 0.02$

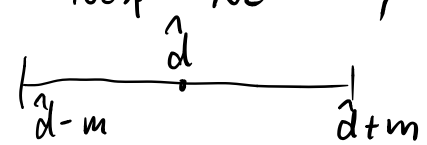
$X_{cont} = 974$ $\quad$ $X_{exp} = 1242.$ $\quad$ Confidence level = 95%.

Sol:

$$\hat{p}_{pool} = \frac{974+1242}{10072+9886} = 0.111$$

$$SE_{pool} = \sqrt{0.111(1-0.111)\left(\frac{1}{10072}+\frac{1}{9886}\right)} = 0.00445$$

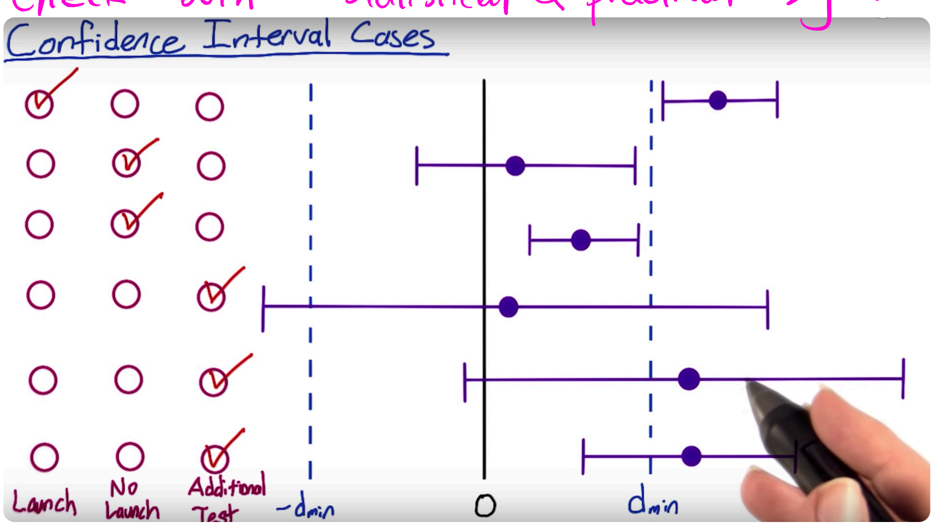
$$\hat{d} = \frac{X_{exp}}{N_{exp}} - \frac{X_c}{N_c} = 0.0289 \qquad m = SE_{pool} \times 1.96 = 0.0087$$



0.0202 > 0.02

☑ launch new version.

Check both statistical & practical sig level.



} more power