

Final Project

Invariant Metrics

- # of cookies
- # of clicks
- CTR

Evaluation Metrics

- Gross conversion
 - Control > Exp
- Retention
 - control < Exp
- Net conversion
 - control ≈ Exp
- # of user_ids (not one)
 - control < Exp

S.D. of Evaluation Metrics

Gross conversion

$$\begin{aligned} \text{# of user-ids who checkout} / \text{# of cookies who click start free} \\ = 660 / 3200 \\ = 0.2063 \\ SD = \sqrt{\frac{P(1-P)}{n}} \\ = 0.0057 \end{aligned}$$

$$\begin{aligned} \text{Retention} \\ \text{# of user-ids remain enrolled past 14-day boundary} / \text{# of user-ids who check out} \\ = 0.53 \end{aligned}$$

$$\begin{aligned} \text{Net conversion} \\ \text{# of user-ids remain enrolled} / \text{# of unique cookies who click,} \\ = 660 * 0.53 / 3200 \\ = 0.1093125 \end{aligned}$$

Sample Size

$$\alpha = 0.05 \quad \beta = 0.2$$

Gross conversion

$$0.2063$$

$$\begin{aligned} \rightarrow \text{# of cookies who click start} &\rightarrow \text{# of page views} \\ \rightarrow 26177 &\rightarrow 26177 \times (40000/3200) \\ = 32723 \times 2 \end{aligned}$$

Retention

$$0.53$$

$$\begin{aligned} \rightarrow \text{# of user-ids who check out} &\rightarrow \text{# of page views} \\ \rightarrow 39115 &\rightarrow 39115 \times (4000/12) \times 2 \quad \text{too large, not practical} \\ \rightarrow \text{# of unique cookies who click} &\rightarrow \text{# of page views} \\ \rightarrow 27413 &\rightarrow 27413 \times (40000/3200) \times 2 \end{aligned}$$

Sanity Check

$$\begin{aligned} \text{# of cookies} \\ P_{con} = 345543 / (345543 + 344160) \\ = 0.5006 \end{aligned}$$

$$H_0: P_{con} = 0.5 \quad H_a: P_{con} \neq 0.5$$

$$\begin{aligned} SE = \sqrt{\frac{0.5 \times 0.5}{345543 + 344160}} \\ = 0.0000 \end{aligned}$$

$$\begin{aligned} \text{# of clicks} \\ P_{con} = 28378 / (28378 + 28326) \\ = 0.5025 \end{aligned}$$

$$\begin{aligned} SE = \sqrt{\frac{0.5 \times 0.5}{(28378 + 28326)}} \\ = 0.0021 \end{aligned}$$

Both test I could assume, if the test was set up correctly, the probability of being in control vs. Exp is 0.5, and when sample size is large enough, the sample mean (here is p) distribution is approx. Normal with $\mu = P$, $b = \sqrt{P(1-P)/N}$

\rightarrow CI G C $M - 1.96\sigma$, $M + 1.96\sigma$
 if what you observed is within CI, then you are 95% confidence your experiment was set up correctly

• CTP

$$CTP_{con} = 0.08210$$

$$CTP_{con} = 0.08218$$

if equal

$$CTP_{pool} = \frac{28378 + 28325}{345543 + 344660}$$

$$= 0.08215$$

$$SE_{pool} = \sqrt{\frac{0.08215 \times (1 - 0.08215)}{345543 + 344660}}$$

$$= 0.00033$$

This portion is calculated wrong.

This confused with the previous calculations.

Actually, one should use SE_{pool}

$$= \sqrt{\hat{p}_1 * (1 - \hat{p}_1) * \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

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

Effect Size Tests

Gross conversion

$$\frac{\# \text{ user-ids}}{\# \text{ user-ids who checkout}} / \frac{\# \text{ cookies}}{\# \text{ cookies who click start free}}$$

Net Conversion

$$\frac{\# \text{ user-ids remain enrolled}}{\# \text{ unique user-ids who click}}$$

$$\text{let } \hat{d} = GC_{cont} - GC_{exp}$$

$$= 0.218875 - 0.198320$$

$$= 0.02$$

$$GC_{pool} = 0.208607$$

$$SE_{pool} = \sqrt{0.208607 * (1 - 0.208607) * \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

$$= 0.004372$$

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

$$\text{let } \hat{d} = NC_{cont} - NC_{exp}$$

$$= 0.004874$$

$$NC_{pool} = 0.115127$$

$$SE_{pool} = 0.003434$$

$$ME = 0.003434 * 1.96$$

$$= 0.006731$$

$$CI = (-0.0019, 0.0116)$$

$$ME = 0.004372 * 1.96 = 0.008568$$

$$CI = (0.2 \sim 0.2171)$$