A/B Testing

Case: Facebook status background color change

class starts @ 9:03 PM

Drug / Medicine : Fever

(200)

(random)

G1 (100) A — Control (Paracetamol) or placebo

G2 (100) B — Treatment (New Med)

#days to recover

A/B Test

↓

randomized Control-Treatment Test

[Phased trails]

T-Test ✓

Case:

Facebook - "status" → PM

colors background

status msgs

↳ Test on a reasonable sample size

1. What do we hope to achieve with this feature. → more engagement

2. has this been tried earlier?

3. external proof that this works?

4. Subset or all users?

Metrics

1

NSM: % age of "engaged" METRIC users ← Measure

liked, Comment, Save

logged-in + Spent 2 min in a week

Alt: Total mins spent on FB

② supporting metric (s):

↳ DAV

→ ↑ increase in status messages

⑤ Guard rail metric:

↳ that should change in a negative manner 'coz of new feature

- avg time spent per user per week
- % age of users consuming rich media
  - Images / Videos

Experimentation:

① $H_0: p_A = p_B$

$H_a: p_A \neq p_B$ (or)

Two-sided

METRIC = percentage of "engaged" users (p)

$p_A < p_B$ ✓

one-sided

② Test: t-Test (?) → X

χ^2-test → X

Proportions test → z, t ✓

- p = (41 + 351) / (195 + 605) = 0.49.

Set this number aside for a moment.

$\hat{p} = \frac{n_1 + n_2}{d_1 + d_2}$

Step 3: Insert the numbers from Step 1 and Step 2 into the test statistic formula:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\sim Z(0,1)$

#successin G_1

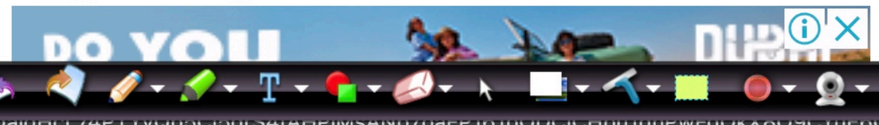
$\hat{p}_1 = \frac{n_1}{d_1} = \frac{\#\text{engaged users}}{\text{Total \# users}}$

$\hat{p}_2 = \frac{n_2}{d_2}$

Solving the formula, we get:

Z = 8.99

We need to find out if the z-score falls into the "rejection region."



Step 4: Find the z-score associat

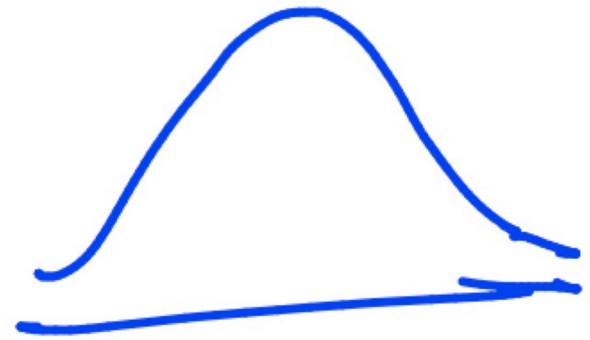
- p = (41 + 351) / (195 + 605) = 0.49.

Set this number aside for a moment.

Step 3: Insert the numbers from Step 1 and Step 2 into the test statistic formula:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\sim Z(0,1)$



Solving the formula, we get:
Z = 8.99

We need to find out if the z-score falls into the "rejection region."

- p = (41 + 351) / (195 + 605) = 0.49.

Set this number aside for a moment.

Step 3: Insert the numbers from Step 1 and Step 2 into the test statistic formula:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Solving the formula, we get:

Z = 8.99

We need to find out if the z-score falls into the "rejection region."

Step 4: Find the z-score associat

- p = (41 + 351) / (195 + 605) = 0.49.

Set this number aside for a moment.

Step 3: Insert the numbers from Step 1 and Step 2 into the test statistic formula:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$H_0: p_1 = p_2$

$\hookrightarrow p_1 - p_2 = 0$

$p_1 - p_2 = \delta$

Solving the formula, we get:

Z = 8.99

We need to find out if the z-score falls into the "rejection region."

Step 4: Find the z-score associat

- p = (41 + 351) / (195 + 605) = 0.49.

Set this number aside for a moment.

$N(\mu, \sigma)$

Step 3: Insert the numbers from Step 1 and Step 2 into the test statistic formula:

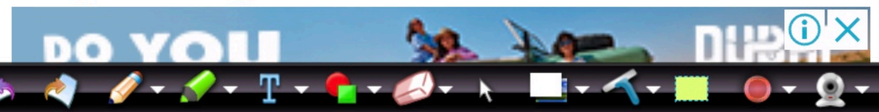
$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\sim Z(0,1)$

Solving the formula, we get:

Z = 8.99

We need to find out if the z-score falls into the "rejection region."



Step 4: Find the z-score associat

100K

50K

no change

A

50K

new feature

B

P_A

P_B

p-value = P(observing | H0) < 5%

100K (?)
(all - users)
or Geo - specific

A

$\hat{p}_A \approx 0.45$

B

$\hat{p}_B \approx$

Minimum detectable effect (MDE)

Biz impact

$p_B = 0.46$: 1% (?)

0.451

0.1%

Sample size

0.1%? or 1%

Guess $10B profit

← 2B users →

45% or 0.45

0.45%

(0.1%)

0.1% ↑

⇒ 10% → 1B → 10M $ additional money

{ 0.1% →

sample-size → 100k users

50k

50k

Multi-day →

10k

10k

10k

10k

10k

10k

10k

10k

10k

10k

e.g:

10,000 → 100 random users

IDs 1 to 10,000

1 2 - 10K

1/10K

np.random() → Unif r.v ~ U(0,1)

0 to 1

(?)

PARTIAL CODE

```
cnt = 0
while (cnt < 100)
{
    0 to 1        tmp = np.random()
    if tmp ≤ 1/10,000
        .
        .
        .
    cnt++
```

Recap:

1. Clarify objective & questions
2. Metrics: NSM, Supporting : Guardrails
3. Expt: Ho, Ha, test; α: 5%; MDE → Sample size

Multi-day test

Pitfalls / Problems

(1) Primacy →

✓

- Wk1 → C 45% T 45%
- Wk2 → 45%
- Wk3 → 45%
- Wk4 → 45%

C T

↓

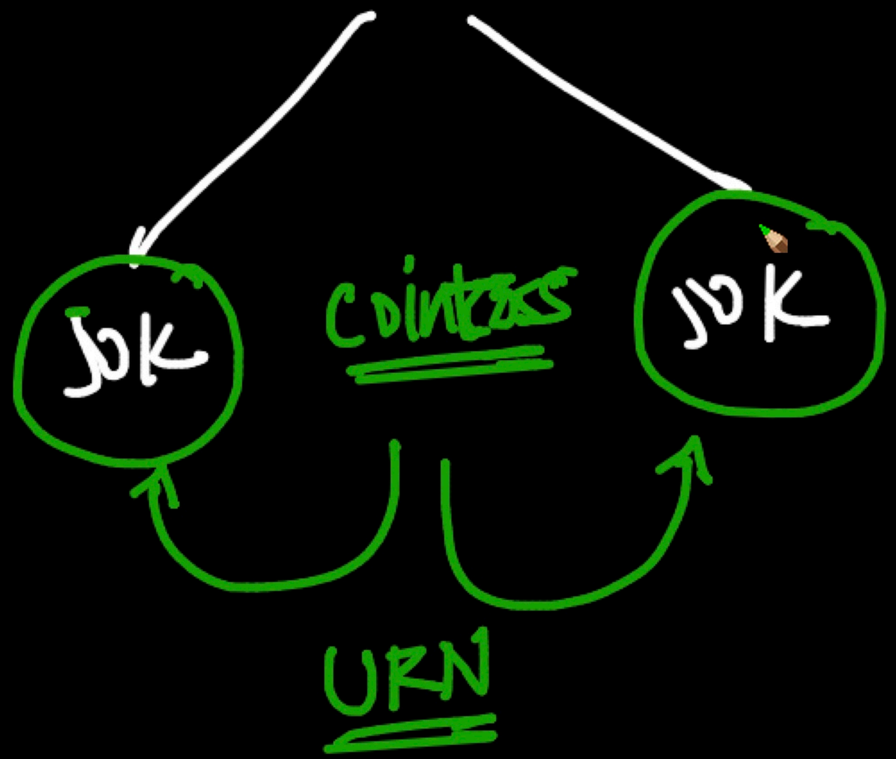
do not want to use the new feature

↓

50%

50.1%

look users



(2) Novelty:

	C	T
Wk 1	45%	65%
Wk 2	45%	68%
Wk 3	45%	52%
⋮		

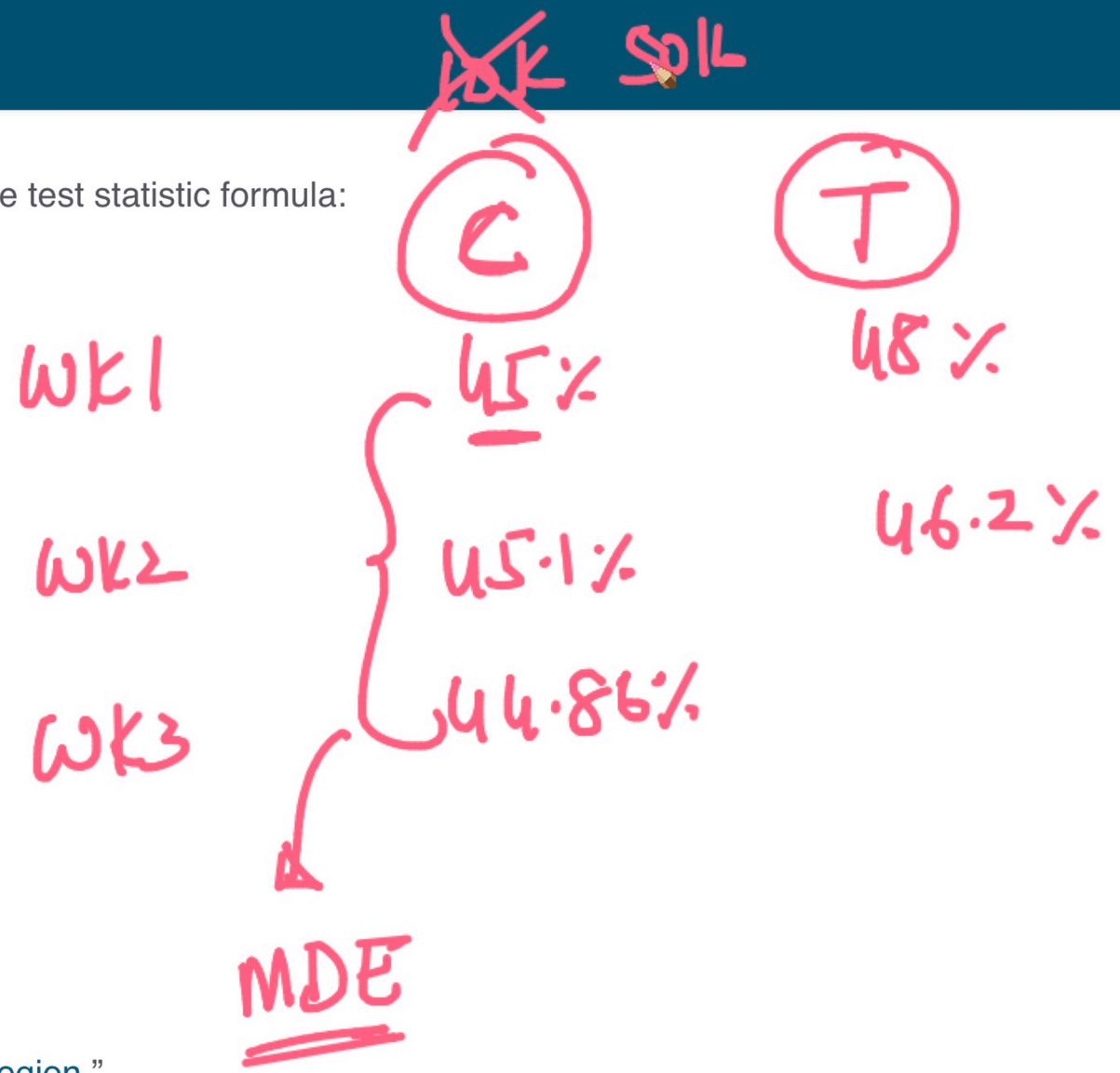
51.2%

51%

Statistics How To

Step 3: Insert the numbers from Step 1 and Step 2 into the test statistic formula:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$



Solving the formula, we get:

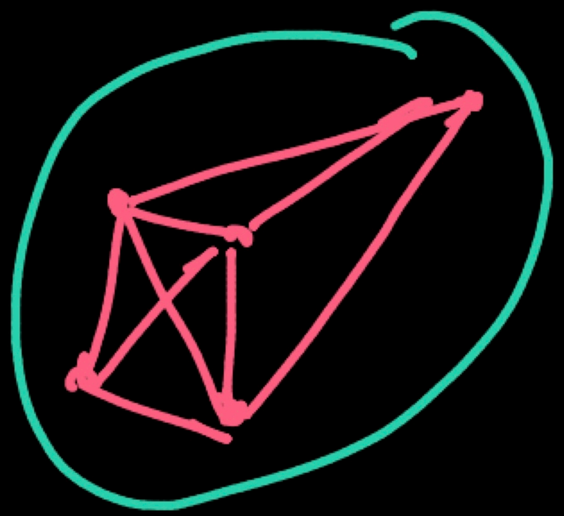
Z = 8.99

We need to find out if the z-score falls into the "rejection region."

Step 4: Find the z-score associated with α/2. I'll use the following table of known values:

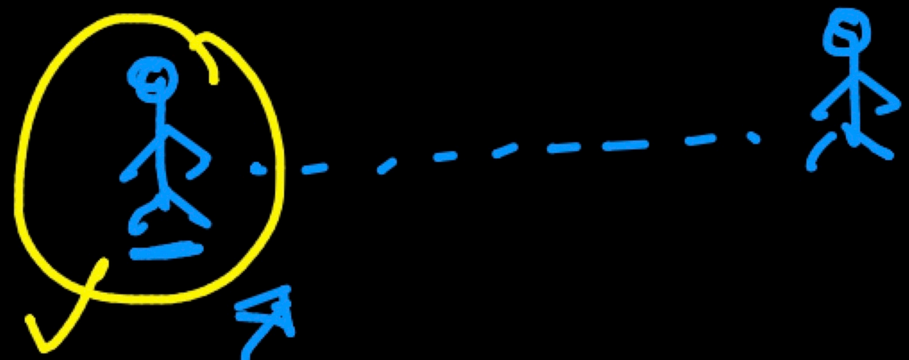
network effect

C



SPILLOVER

T



RD

T

45%

46%

44%

SPILLOVER

50%

50%

44.6%

Outcome bias:

DS|ML

Caution

look ✓

Launch to 1B+ users

↳ 0.1% or 1% ↑ mean to biz ✓

→ Cost of launching this feature

↳ hardware ✓

API/....

→ long-term impact (4-6 wks)

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:57

Z Test: Definition & Two Propor | S.6 Test of Proportion | STAT O | pivot sql - Google Search

Search Google or type a URL

Update

29 / 29