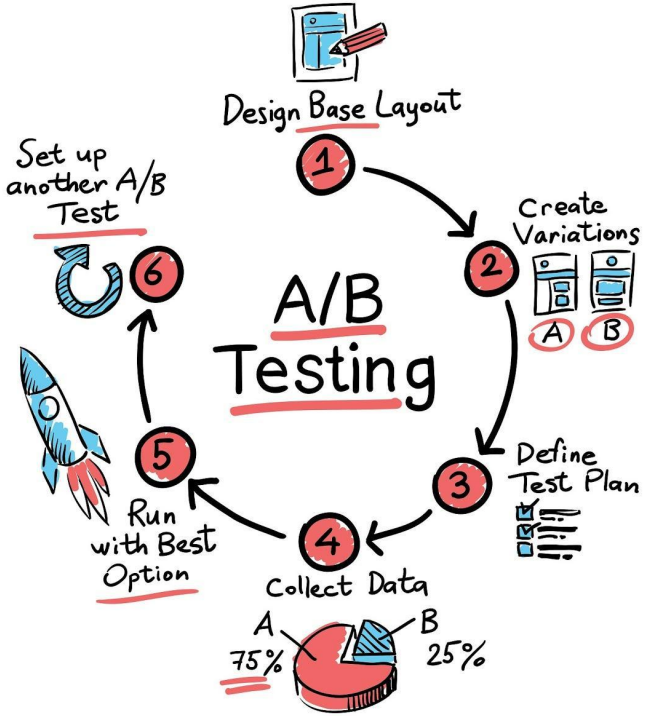
A/B testing

Maria Aguilera,
Kelsey Moon,
Michelle Joung

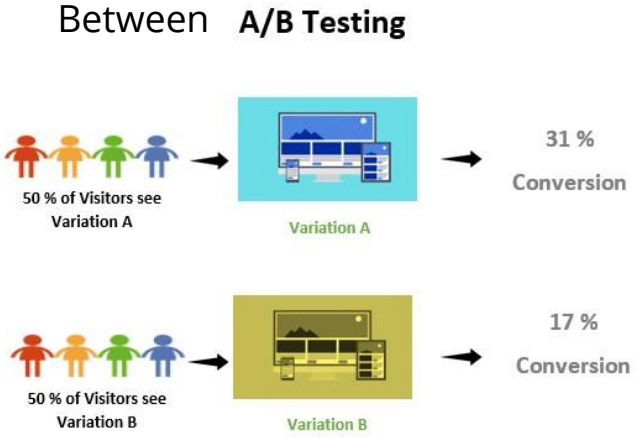
What is A/B testing

- Using experimental design and statistics to compare two or more variants of a design
- Usually deals with *conversion rates*



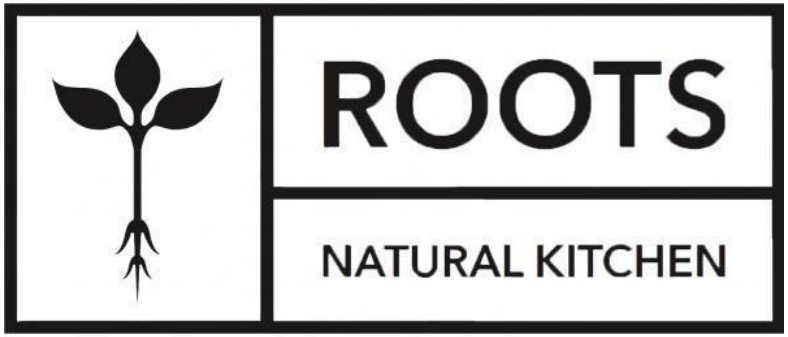
Utilization and Types of A/B Testing

- Types:
  - Within
    - Each participant sees both conditions
  - Between
    - Different groups see different conditions
- Uses:
  - Conversion Rate
  - Engagement
  - Drop off Rate
  - Time Spent
- Purpose: Maximize desired goal
- Assumptions



Within A/B Test

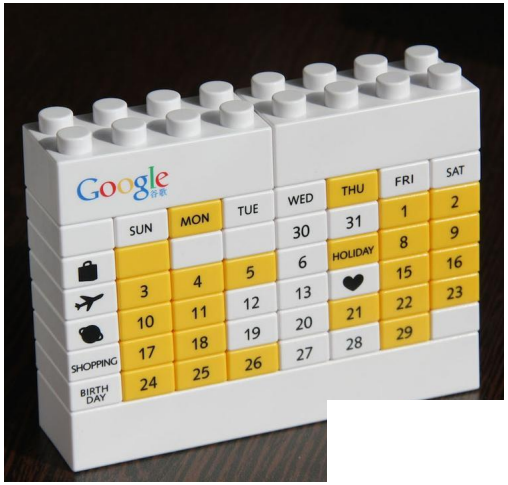
Old Version



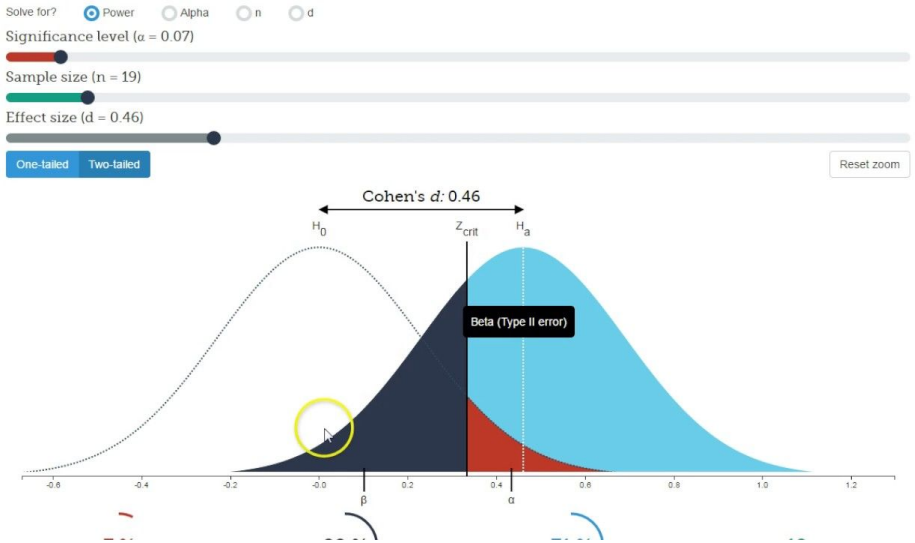
New Version

What now?

1. Time frame
2. Test and Control
3. Collect Data
4. Sample Size (Power Analysis)
5. Statistical Test
    a. Logistic Regression - for binary categorical DV
    b. T-test - continuous binary DV
    c. Linear Regression - continuous DV
    d. Chi-Squared



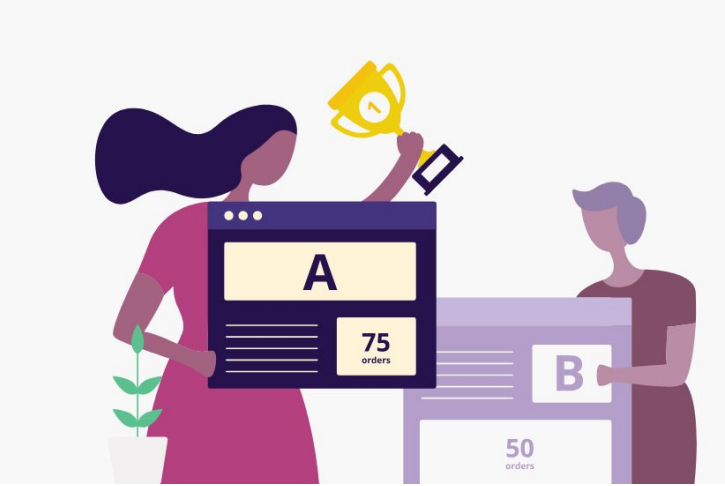
Power Analysis: Sample Size



- What is necessary:
  - Power (1 - Beta)
  - Significant level (alpha)
  - Effect Size
- Functions depend on statistical tests
  - SSizeLogisticBin()
  - pwr.t.test()
- Higher power, lower alpha, and smaller effect size will require **more** data points

What's Next?

- Figure out which model performed better
  - Statistical Tests: P- value, standard error, estimates, test statistic
- Update to the winning model
  - If the new model won, compare it to the old one
- Follow up experiments
  - Take things one step at a time
  - Small changes



Bag


ROOTS
NATURAL KITCHEN


ROOTS
NATURAL KITCHEN
HOURS
MONDAY - FRIDAY
10:30 AM - 9:00 PM
SATURDAY & SUNDAY
11:00 AM - 9:00 PM

Door

Meat

Vegetarian

Benefits

- Improve content
  - Better understanding of customer base
- Higher page engagement
- Can test multiple elements
- May increase revenue
- Increase conversion rates
- Decrease acquisition rates

Common Issues

- "Flicker effect"
- Small sample size can lead to inaccurate results
- Requires a minimum of 2 - 3 weeks
  - Or at least 5,000 ‘conversions’
- Needs regular updating
- False positives
- Confounding Variables
  - Internal
  - External
- Side Effect
  - Unintended consequence of a change

A/A testing

- Test one version of a page with two different groups

Before the A/B test:

- Helps find minimum sample size
- Baseline conversion rate

A/A Testing

After the A/B Test:

- Ensures it was efficient
    - No difference vs difference in conversion rates

A/B/N or Multivariate Testing

- Compare a control condition to “N” number of different versions of a web page
- Test to see which specific feature generates the best conversion rates

Cons of A/B/N testing

- May require lot of time
- Poor efficiency, more complicated statistics
- Each variant tested increases the sample size requirements

A/B/N example


ROOTS
NATURAL KITCHEN

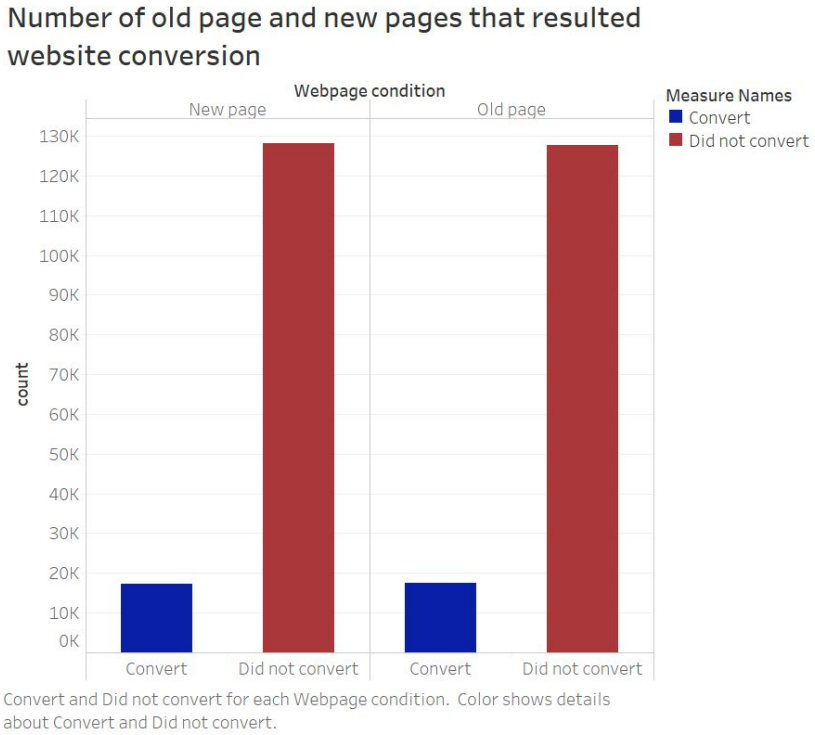

ROOTS
NATURAL KITCHEN


ROOTS
NATURAL KITCHEN


ROOTS
NATURAL KITCHEN

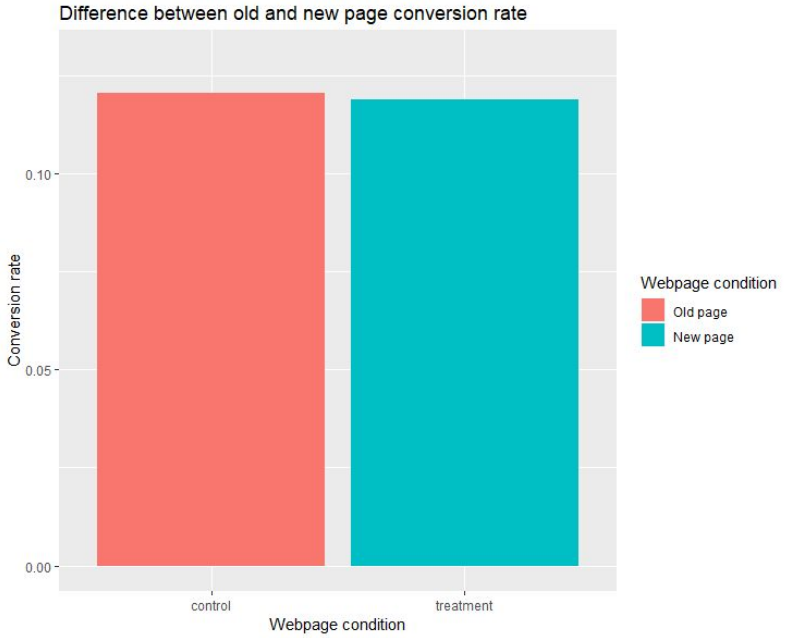
Data Set

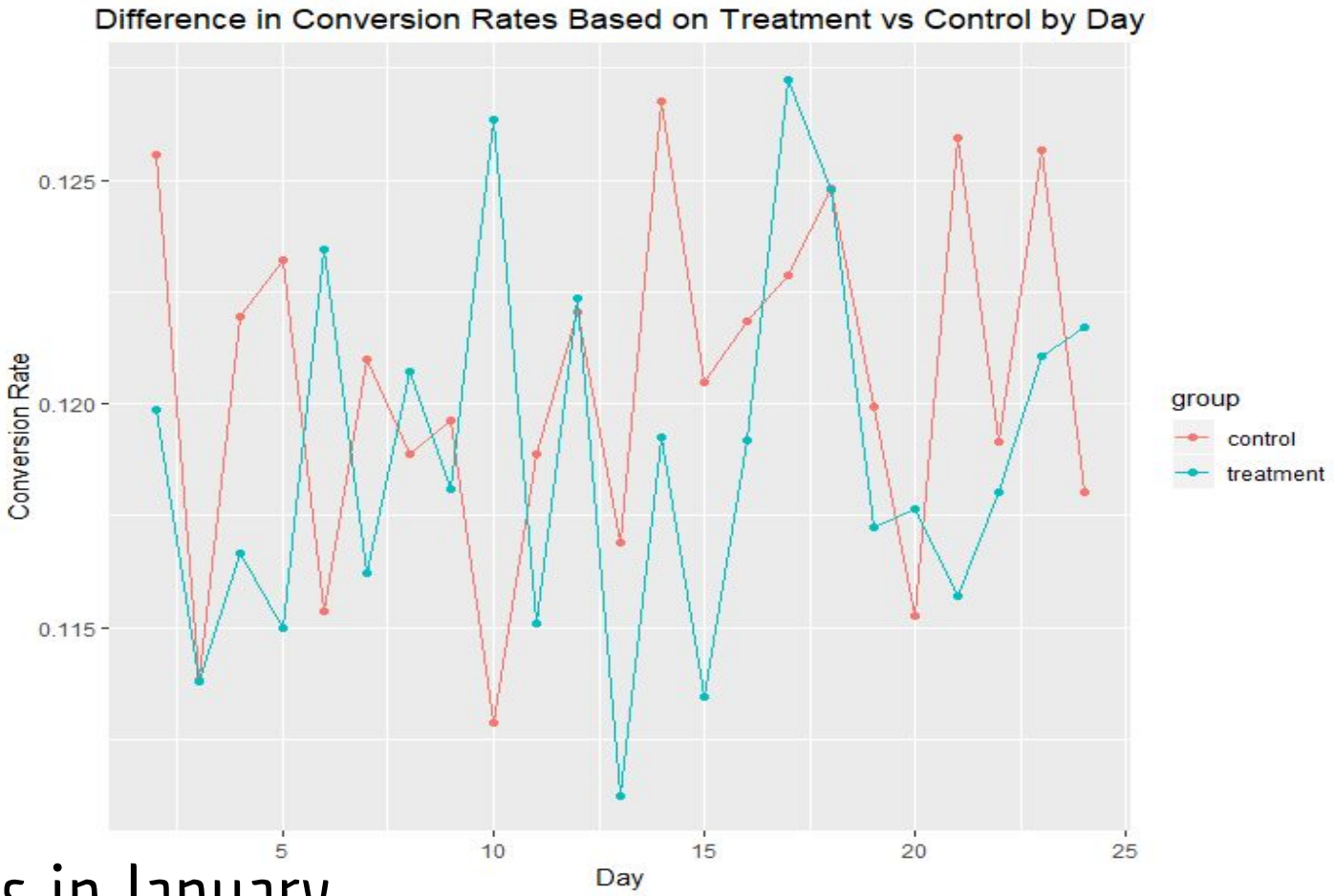
- Found on Kaggle
- 294,478 observations
- Five variables:
  - user_id, timestamp, group, landing_page, and converted
- Data collection period:
  - January 2nd, 2017 to January 24th, 2017
- Purpose : To determine whether the new website design results in a higher conversion rate than the old website design.



Conversion rate

- Rate = (Those who clicked) / (Total visitors)
- Difference between the old and new page:
  - 0.16%.
- Conversion rate control:
  - 12.04%
- Conversion rate for new page:
  - 11.88%





Conversion rates in January

Statistical Analysis - Chi-Square

- Used to test if two categorical variables are related in population.
  - Large sample size
  - Independent observations

H0: There would not be a significant difference of conversion rate between old page and new page

HA: There would be a significant difference of conversion rate between old page and new page

- P value of chi square test = 0.19 (>0.05)
- Null hypothesis failed to reject.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Confidence Interval

- The 95% confidence interval : -0.0007144889 and 0.0039880781
- Confirms that there’s no actual difference in conversion rate between old website and new website.

```
        2-sample test for equality of proportions with continuity correction

data:  t2
X-squared = 1.8568, df = 1, p-value = 0.173
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.0007144889  0.0039880781
sample estimates:
   prop 1    prop 2
0.1204776 0.1188408
```

Recommendations/Future Research

- Test more specific variables
  - Exact word clicked for conversion
  - Amount of time spent on page
- Longer data collection period that accounts for seasonal variation
- Perform other A/B tests
- Test demographics of visitors

Real World Example

Version 1

Version 2

Conclusion

- Our Data:
  - Based on the statistical tests, the new site did not change the conversion rates
  - Maintain the old website design
- General:
  - Know the type of A/B test you want to perform prior to starting
  - Collect enough data
  - Start small!
  - Continue to make improvements

Questions?

Citations

Brebion, Anthony. “How to Avoid Flickering (Flash of Original Content) in A/B Tests.” *AB Tasty*, 9 Oct. 2018, www.abtasty.com/blog/ab-testing-flicker-effect/.

Cabot, Jordi. “A/B Testing for Dummies.” *Nelio Software*, 9 Apr. 2018, neliosoftware.com/blog/ab-testing-dummies/.

Chapman, Cameron. “A Beginner's Guide To AB Testing: An Introduction.” *Neil Patel*, 15 Feb. 2019, neilpatel.com/blog/ab-testing-introduction/.

Deshpande, Bala. “2 Key Assumptions to Be Aware of before Applying the Chi-Square Test.” *Analytics Made Accessible: for Small and Medium Business and Beyond*, Soliton Technologies, www.simafore.com/blog/bid/56480/2-key-assumptions-to-be-aware-of-before-applying-the-chi-square-test.

Parmar, Divya. “Introduction to A/B Testing.” *Data Science Central*, DSC, 23 May 2017, www.datasciencecentral.com/profiles/blogs/introduction-to-a-b-testing.

Piccinini, Page, director. *Assumptions and Types of A/B Testing*. *R*, *2018* campus.datacamp.com/courses/ab-testing-in-r/chapter-3-experimental-design-in-ab-testing?ex=5.

Shukairy, Ayat. “Momentum Behavior: Are Visitors in Love with Your Old Site Design?” *Invespcro*, Invesp, 2017, www.invespcro.com/blog/momentum-behavior-are-visitors-in-love-with-your-old-site-design/

Zhang, Luyuan. “A/B Testing.” *Kaggle*, Kaggle Inc, 25 June 2018, www.kaggle.com/zhangluyuan/ab-testing.