A/B Testing Review1

1. Main Idea:
   - Hypothesis testing: control and test group-> If there exists a significant difference under controlled condition.
   - H0: No difference; H1: under some treatment, there exists a significant difference.
   - Two tails or One tail test
2. Test Statistics:
   - ✓ T-test: normally distributed/unknown variance/independent small sample size
   - ✓ Z-score: independent large sample size/normally distributed/known variance we use Z-score method
   - ✓ Chi-Square test: in order to test whether the test group complies to theorical distribution (distribution of control group)
3. Compare Two Samples: pooled standard error

$$\hat{p}_{\text{pool}} = \frac{X_{\text{cont}} + X_{\text{exp}}}{N_{\text{cont}} + N_{exp}}$$

$$SE_{\text{pool}} = \sqrt{\hat{p}_{\text{pool}} * \left(1 - \hat{p}_{\text{pool}}\right) * \left(\frac{1}{N_{\text{cont}}} + \frac{1}{N_{\text{exp}}}\right)}$$

$$Practical\ significant\ boundary\ d = \hat{p}_{\text{exp}} - \hat{p}_{\text{cont}}$$

$$if\ \hat{d} > 1.96 * SE_{\text{pool}}\ or\ \hat{d} < -1.96 * SE_{\text{pool}} \rightarrow Reject\ H0.$$

4. Ensure Experiment Metrics: (from product or business respects)
   - Number of Cookies (page view)
   - Unique View (UV)
   - Number of user's id
   - Number of clicks
   - Click through probability=number of users who click/number of users who visit→measure the total impact
   - Retention
   - Net conversion
   - Gross conversion
   - DAU (Daily Activity User) /WAU/MAU
   - DNU
   - LTV (life time value) …
5. How to choose good metrics?
   1) Sensitivity and robustness→A/A test to measure, sanity check
   2) Variability
6. Multivariate Testing (MTV): we test several variables at the same time to get the best combined elements of product.
   - Bonferroni Correction:
     - Adjust significant level threshold.
     - If the original significant level (**α**)=0.05→After adjustment: **α***=0.05/numbers of multivariate tests
     - If the p-value we get at final< **α***, the experiment result is significant.
7. Sample Size:
   - ➢ Calculator: https://www.evanmiller.org/ab-testing/
   - ➢ Information we need:
     - ✓ Significant level (**α**): normally 0.05
     - ✓ Power (1-**β**): normally 0.8
     - ✓ Expected standard deviation
     - ✓ Minimum detectable size (difference between test and control group)
   - ➢ Formular:

   1) If the metrics are in form of **percentage or ratio**

$$n = \frac{\left(Z_{a/2} \cdot \sqrt{2 \cdot \frac{(p_1 + p_2)}{2} \cdot \left(1 - \frac{(p_1 + p_2)}{2}\right)} + Z_\beta \cdot \sqrt{p_1 \cdot (1 - p_1) + p_2 \cdot (1 - p_2)}\right)^2}{|p_1 - p_2|^2}$$

$p1: contol\ group's\ effect\ performance \quad p2: test\ group\ effect\ performance$

2) If the metrics are in form of count or difference

$$n = \frac{-\sigma^2}{\Delta^2}\left(Z_{\frac{\alpha}{2}} + Z_\beta\right)^2$$

8. Downside of A/B Testing:
   - Not good for testing new or big experience