a p value is a probability that prvoides a measure of the evidence against the null hypothesis provided by the sample.  
small p values indicate more evidence against null hypothesis.   
  
P\_value: Assuming null hypothesis is true, what’s the probability of observing a as extreme or more extreme test statistics as the observed case  
  
A/A testing is the tactic of using A/B testing to test two identical versions of a page against each other. Typically, this is done to check that the tool being used to run the experiment is statistically fair. In an A/A test, the tool should report no difference in conversions between the control and variation, if the test is implemented correctly.

What is A/B Test?

A/B testing is general framework of hypothesis testing between two groups.

**A/B Testing** method covering the statistical hypothesis tests are used to reveal significance of the difference in a metric variable value between two groups such a Group A and Group B.

CLT - Interview Quiz

User CTR of your website is p

You sampled 1000 users from

your website. What is the

sampling distribution of the

sample mean?

Hypothesis Test - Interview Quiz

● Test a coin is fair. What test? What is null hypothesis?

● Test two group of users have same CTR (click through rate). What test? What is d.f. (degree of freedom)?

● Test two group of users on your website have the same mean spending. What test? What is d.f.?

DOE - A/A Test - Interview Quiz

Your colleague gave you two dataset and told you they aretest/control groups from an A/B test. What will you do to make sure the datasets are appropriate to use?

What will you do if you find users been assigned to both test and control group? Do you have any concern?

DOE - Metrics - Interview Quiz

What metrics are you going to evaluate the success of digital ads? How to convince clients to buy ads on our

website?

DOE – Sample Size - Interview Quiz

How long would you run your experiment?

What is your minimum sample size? What factors would you consider? How would these

factors impact your sample size?

What will you do to balance user experience and quick learning?

What is your roll-out plan?

What will you do if your experiment take too long to run?

● Can you run an experiment and keep reading until the result is

significant?

● What if your key metrics dropped by 5% on first day? What if

dropped by 20%?

● What is bootstrap? Boosting?

● The metrics of interest is 90th quantile of users’ spending. How to

estimate sample mean and variance?

RM – Interview Question

What if your result show positive impact on some metrics and

negative impact on some other metrics?

○ What if your result is neutral?

○ What if you result is statistically significant but the margin is very

small?

○