We at The Data Monk hold the vision to make sure everyone in the IT industry has an equal stand to work in an open domain such as analytics. Analytics is one domain where there is no formal under-graduation degree and which is achievable to anyone and everyone in the World.

We are a team of 30+ mentors who have worked in various product-based companies in India and abroad, and we have come up with this idea to provide study materials directed to help you crack any analytics interview.

Every one of us has been interviewing for at least the last 6 to 8 years for different positions like Data Scientist, Data Analysts, Business Analysts, Product Analysts, Data Engineers, and other senior roles. We understand the gap between having good knowledge and converting an interview to a top product-based company.

Rest assured that if you follow our different mediums like our blog cum questions-answer portal www.TheDataMonk.com , our youtube channel - The Data Monk, and our e-books, then you will have a very strong candidature in whichever interview you participate in.

There are many blogs that provide free study materials or questions on different analytical tools and technologies, but we concentrate mostly on the questions which are asked in an interview. We have a set of 100+ books which are available both on Amazon and on The Data Monk e-shop page

We would recommend you to explore our website, youtube channel, and e-books to understand the type of questions covered in our articles. We went for the question-answer approach both on our website as well as our e-books just because we feel that the best way to go from beginner to advance level is by practicing a lot of questions on the topic.

We have launched a series of 50 e-books on our website on all the popular as well as niche topics. Our range of material ranges from SQL, Python, and Machine Learning algorithms to ANN, CNN, PCA, etc.

We are constantly working on our product and will keep on updating it. It is very necessary to go through all the questions present in this book.

Give a rating to the book on Amazon, do provide your feedback and if you want to help us grow then please subscribe to our Youtube channel.

A/B Testing

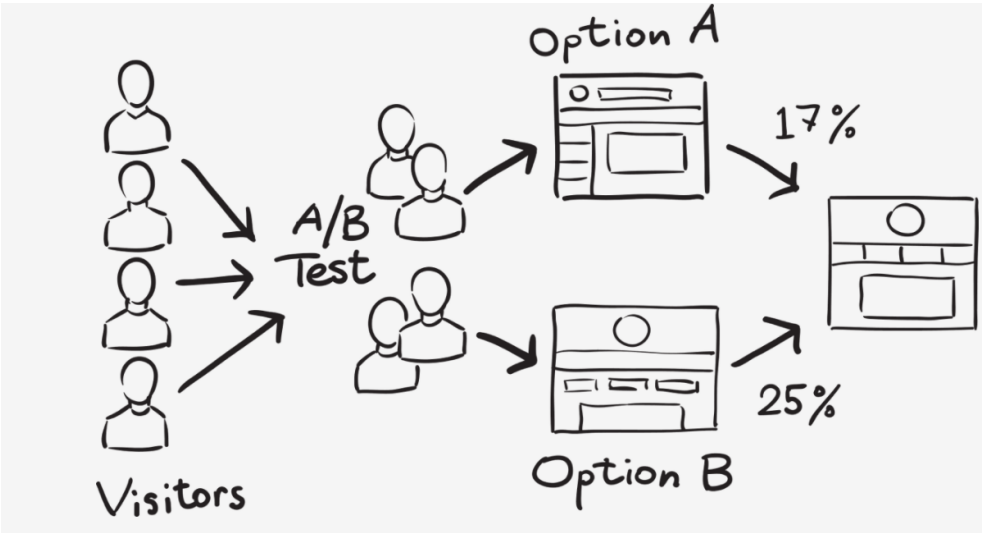
1) **What is A/B testing?**

   A/B testing is a randomized experiment where you have 2 versions of your webpage with some visible changes in it and statistically measuring which version could bring more impact to customers and drive business in an intended direction.

2) **Where exactly are we going to use A/B testing?**

   Imagine being a data scientist in a company and your company wants to know what it can do to engage with its audience in a better way. They thought they needed to make some changes in their webpage.

   Okay, this idea might work but how can we be sure about this?
   That's where you come in, as you know about a/b testing (let's assume that) you say we initially make the design public to few customers and compare how a set of customers are reacting to it and at the same time period how the other set of customers are reacting to old webpage.



Source

(All the visitors are split and shown two version and their conversion rates are measured and used as test size and control size for A/B testing)

We can definitely have certain factors to decide, for example: changing the search button from top to left in a new webpage might have made people click it more often.
You compare the 2 webpages engagement and do the A/B testing and submit the reports to the company which will make them decide if they should change the design of the webpage.

Now that we know what A/B testing is, let's know some math behind it and some concepts like p-values, hypothesis testing etc. Which helps us further understand A/B testing in a more intuitive way.

3) **What are some of the main metrics evaluated in A/B testing?**
   - Impression count
   - Click-through rate
   - Time period spent on particular page
   - Bounce rate on button's click-through link
   - Button hover time

     Do not consider too many metrics for an A/B test. It does not provide proper significant results.

4) **What is the p-value which measures the statistical significance for A/B testing?**

   p-value is a statistical measurement which tells us how common our selected value is to the whole data set or distribution and this value tells us if the given selected value came from the particular data set.

   p-value ranges from 0-1.
   The closer the p-value to 0 the more we are confident that the selected value is unique and has some statistical significance.
   We generally give a threshold 0.05 to decide the above. If the p-value of a certain point is <=0.05 then it means that the point is not that common to the dataset.
   If p-value is greater than 0.05 that means this point can be described by the distribution formed by the dataset or samples.

5) How should p-value be interpreted when performing A/B testing?

P in p-value means probability.
This is how we interpret p-value: suppose you rolled 10 dice and the outcome was number 5 in 9 dice and 1 in one die.

Now if some asks you, is that outcome rare or it can just be one of the random outcomes that often.
Your intuition says “yes, it is definitely not something random or happens many times, it is a rare event”.
But math does not rely on intuition. You need some quantity or number which says “yes, it is a rare event because the value is this or less than this or more than this something like that which deals in terms of number”.
p-value serves the purpose in this situation.
p-values Is meaningless without an initial hypothesis.
I will explain what this means.
Initially your intuition said it is rare. Did you question why you thought this way?

Let me answer this: You thought this way because you know that each number in dice has equal probability of occurrence and in spite of that you got a 5 in 9 dice so that is why you were amazed by the result.

Now answer this question.
Would you have been amazed with this result if you don’t know that all outcomes have equal chance of occurring. A big NO right?
We say something as rare or common based on some hypothesis. Which in this case is 1/6 chance of occurring any number on any dice.

This is what we mean saying that p-value does not make sense if you do not have a hypothesis attached to it.

In layman terms: if you want to say something is rare, you need to have an idea of what is common and what is rare. This idea or initial assumption which says that things are not different is called the null hypothesis. So p-value must be backed by a null hypothesis every time.

6) **What are the factors considered in calculating p-value while performing A/B testing or hypothesis testing?**

p-value comprises of 3 elements:

1) The probability of random chance would result in the observation.
2) The probability of other cases which is equally possible or has equal probability.
3) The cases for which probability is even rarer.

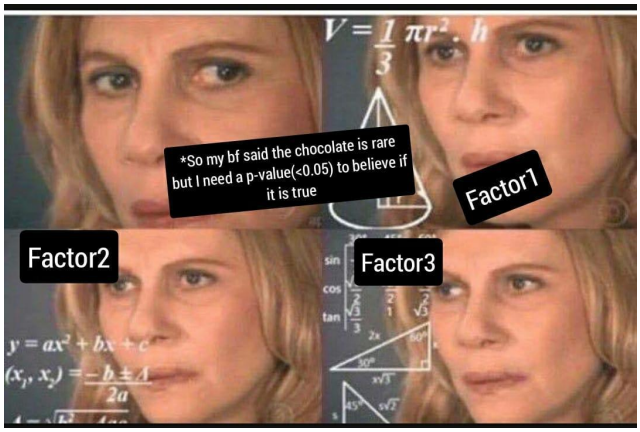
7) **Why exactly do we consider these 3 factors to determine p-value in A/B testing?**

Let us take an example and find out the reasons. And see what exactly we mean by p-value

Imagine that you gave a chocolate to your lover and claimed that it is a rare chocolate in order to impress her.

You said chocolate is rare.

But your girlfriend is a statistician. She can only believe it is different if the p-value is less than 0.05.



First, she assumes that this chocolate is not at all special (this is our null hypothesis which we discuss later).

Let me just ask you what factors she needs to consider in order to tell it is rare.

- She needs to know if there are any other chocolates that are rare like this, if there are many chocolates which are equally rare like this. she can come to the conclusion that, ok so there are many rarely available chocolates so what's special in it, this is one of them.

  BAZINGA!! She proved her point, it is not a special chocolate. We are not done yet. You are going to be screwed further.

- There are also other factors like are there any other chocolates which are even rarer. The more the number of chocolates that are rarer than the chocolate you gave the more you are screwed (screwed here tells that your point that this chocolate is special becomes less qualified)

- The last and common factor she considers is the general probability of the availability of chocolate.

So, what happened here is. The movement you claimed something she first assumed null hypothesis that nothing is different: this chocolate is like any other.
Now using the above 3 factors she added each factor result and she got a value greater than 0.05(p-value) for example 0.23. Now she says that her boyfriend is a liar.

8) Consider an example and calculate the p-value of it

Since we have understood the basic idea and concept behind p-value. We will take an example and find out the p-value.

Consider 3 coins tossed and find the p-value of 2 tails 1 head (H-head, T-tail) combination.

Factor 1: random probability: 3/ 8 (TTH, THT, HTT) (3 coins total number of cases 8)

Factor2: cases which have equal probability
2 heads 1 tail can be obtained in 3 different ways (TTH, HTT, THT).
Probability= 3/8

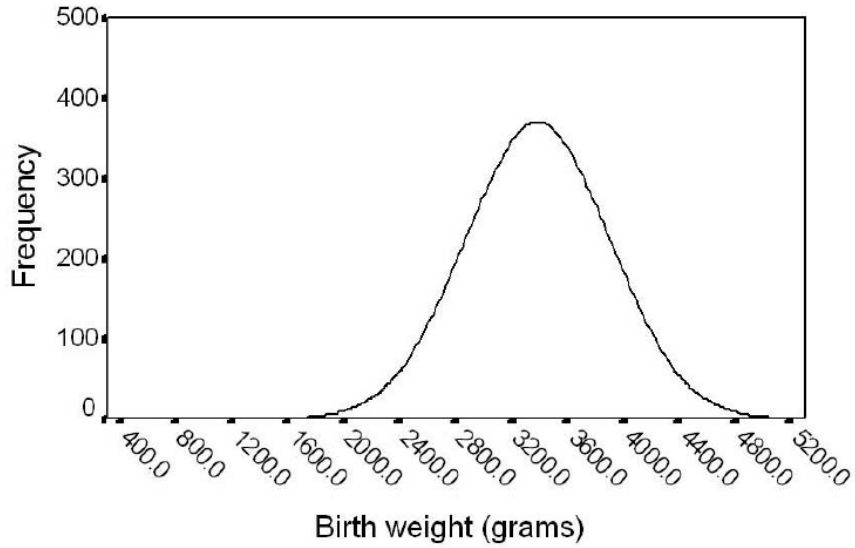
Factor3: cases that are even rarer than 1 head 2 tails: all heads or all tails.
TTT, HHH
Probability = 2/8

P-value = 3/8 +3/8 +2/8 =1 (>0.05) so 2 tails 1 head is not such a highly different outcome or rarer outcome.

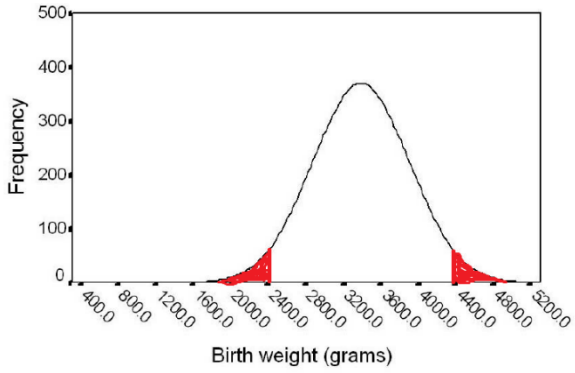
To see a less magnitude p-value you can consider other examples like all heads when 6 coins are tossed at a time. Solve this as an exercise.
You get 0.03125 as p-value.

9) **How do you calculate p-values for continuous values like weight of newborn babies?**

In the previous case, the number of cases are fixed but when we consider weights the values are continuous and you cannot count all and find p-values so we come up with density distribution.



This is a sample case to show how the distribution looks like.
In this case, Let's ask a question, what is the p-value of 2400 grams



We need to find the total area which has less density, check the total red area that will be the p-value of 2400 grams.

10) What is the confidence interval for A/B testing or hypothesis testing?

Confidence interval is the range in which a given percentage of test outcomes fall. Generally, we take 95% as a confidence interval. If value moves out of this range the hypothesis fails

Now let's see what are the important concepts in the A/B testing as we go further into the article.

11) What is hypothesis testing?

Hypothesis testing is a statistical method in which you were initially supposed to take a hypothesis. You collect samples to finally decide either to fail to reject the hypothesis or reject the hypothesis based on p-values or some other measures based on the dataset.

12) What is a null hypothesis?

Whenever we want to find out if 2 data samples are completely different or similar. We initially come up with a hypothesis that the two data samples from 2 different conditions are similar. This hypothesis which says both are similar is called the null hypothesis.

13) What is an alternate hypothesis?

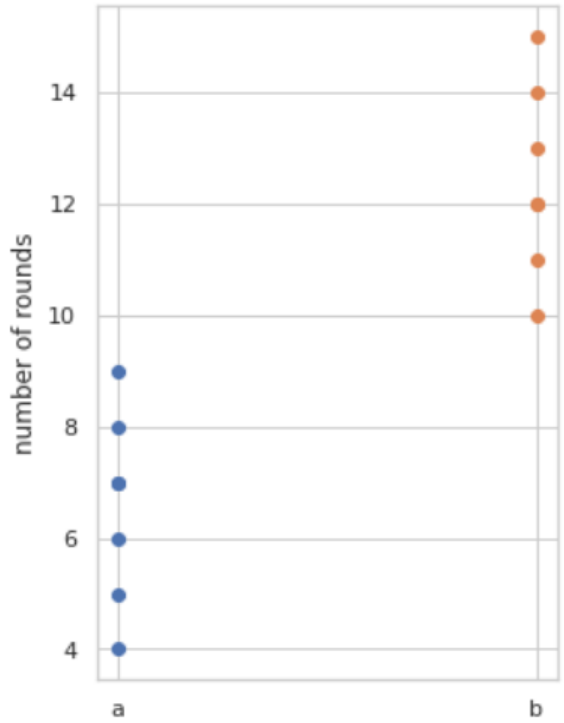
Null hypothesis says both are similar. The Alternative hypothesis is completely opposite; it says that 2 data samples are completely different and that these cannot be explained under a single graph.

14) With an example, show what is null and alternative hypothesis in A/B testing?

Imagine I have 2 lab rats (A and B). I am conducting an experiment which is to make them run in that circle to test their strength.

(here they are in same ring but imagine them having their own separate ring for the experiment)

The experiment goes for a week for both A and B and we measure what the number of rounds both A and B can make individually before they are exhausted.



Consider this as the graph for the results of the experiment.

Now we initially have a null hypothesis. What would that be?

Once go check the previous question to have an idea.

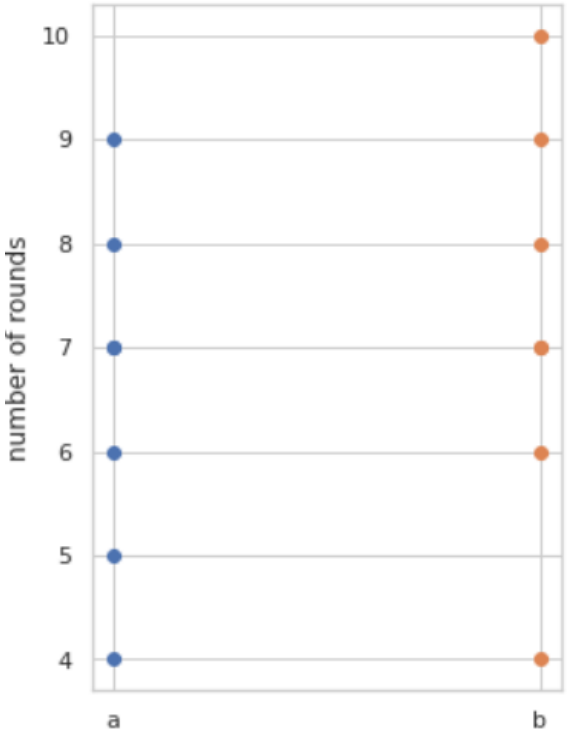
The **null hypothesis** for this example would be that.

Both rats are similar in making a number of rounds and it assumes that these differences in graphs are purely random.

The alternative hypothesis would be that both are different and cannot be because of randomness but some inherent property. To check if null hypothesis can be rejected or not rejected, we have tools like p-value.

In this example if we find p-value and we get the value as 0.03 which is less than 0.05 so we reject the null hypothesis.

In the other case, if we have data something like the below graph



You do not see all the 7 points for either a or b because some points are overlapping.

For this experiment it looks like both have similar capacities so you might get p-value as 0.4 which is less than 0.05 which fails to reject a null hypothesis.

We can conclude that both lab rats are similar.

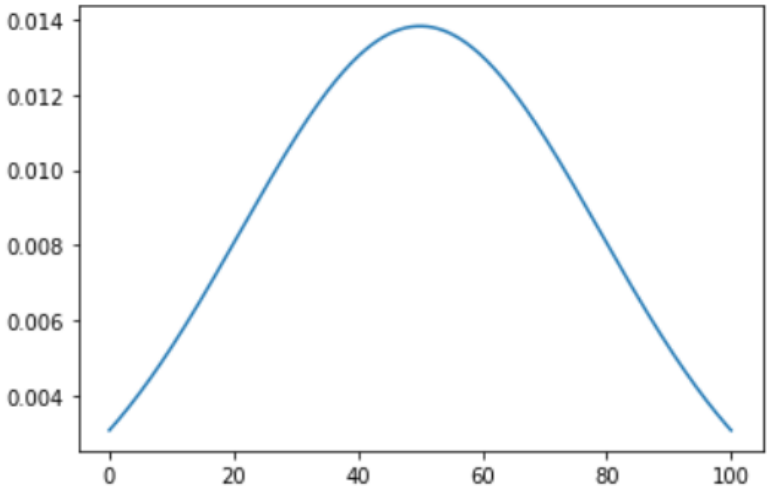
15) What is a type-1 error which occurs in hypothesis testing/ A/B testing?

Type-1 error occurs during hypothesis testing when we reject a null hypothesis even when it is true.
It is also represented by α

16) What is type-2 error which occurs during hypothesis testing?

This other type of error occurs when we fail to reject the null hypothesis even when it is false.
It is also represented by β.

17) How do we interpret typ1 and type 2 errors for A/B testing?



Consider this as density fraction v/s marks graph. We can see it being a normal distribution. It kind of mimics the true nature of most of the classes. Class dominated by average students and toppers and failed percentage of students being less in number.

Since the mean of this data is 50 marks.
So our null hypothesis would be

H_0: μ= 50
if I take some random sample out of class(like picking some 30 students out of 200 students) and I get the mean as 70. Then my alternate hypothesis would be $\mu \neq 50$.

Alternate hypothesis
H_1: μ≠50

Seeing this random sample, we may say that, yes, since we got a mean as70 we reject the null hypothesis (which says mean is 50).
But when we saw the whole sample we knew that the mean is 50 but we rejected the null hypothesis. This is known as type-1 error.

Type-2 error occurs when your random sample says the mean is 57 and as it is near to null hypothesis so you accept it. you consider it as true - mean by accepting but it is not true the value of true mean is different. In that case the error is called type 2 error.

β is called probability of type-2 error and 1- β is called power of the hypothesis test.

18) Show the 2 types of errors in A/B testing in confusion matrix

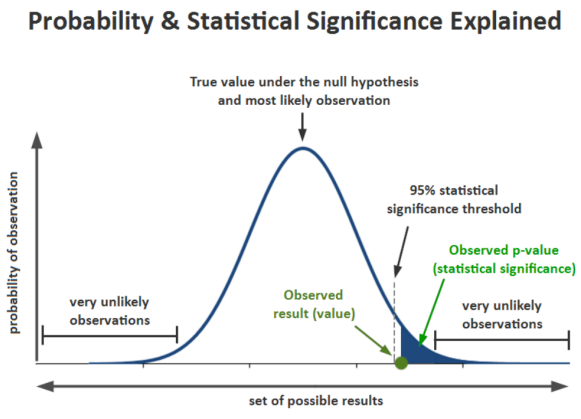
Null hypothesis(H_0)[true nature]

sample case	**Valid/True**	**Invalid/False**
Do not reject	**OK**	**Type-2 error**
reject	**Type-1 error**	**OK**

19) When can we say that the two data samples are so different that this difference is statistically significant in A/B testing?

When we compare 2 datasets we see one data being on the higher side and may directly assume that this is always true or we can also assume that this difference is completely random.

In statistics we need definite answers we just cannot say okay this looks so different and it is not because of randomness or it is nearly similar. We need certain thresholds to decide if the changes are due to randomness or if they are really different.



source

If we observe the above diagram, we see the threshold which determines the hypothesis. If the p-value is extremely away from the centre and threshold then these are unlikely observations.

Let's look at some common misconceptions people will generally have with respect to A/B testing and hypothesis testing.

20) Does low statistical significance or high p-value always mean that there is no improvement while performing A/B testing ?

Let's take an example to illustrate the situation.
Consider we have 2 coins tossing and calculate p-value for TT (2 tails which is like least possible outcome among the possible outcomes) even though it has least outcome among others it still gets p-value greater than 0.05 we get p-value as 0.5. In fact any outcome cannot produce a p-value less than 0.5 so we need to check these conditions before we finally conclude if there is statistical significance for a particular data or not.

21) In the previous question we saw that no matter what case you consider you never get a p-value less than 0.05. If something like that happens when you run A/B tests on 2 versions of webpages How do you solve it for your company?

There is a term called "statistical power" or to understand in this context we can say sensitivity.
This increase or decrease in statistical power changes the sensitivity or significance level.

To understand this concept, consider this example.

Suppose you got 10 and 15 in A and B versions respectively so you come to a decision by finding p-value that both are similar but we can clearly see that there is a difference of 33% . Why did we get this discrepancy?

In order to reliably measure the uncertainty attached to a claim of no improvement in views, you need to look at the statistical power of the test.

Statistical power measures the **sensitivity** of the test to a certain true effect, that is: how likely you would be to detect a real discrepancy of some minimum magnitude at a desired statistical significance level.

Solution:

- Now to work with the same statistical power we need to increase our sample size, that is to consider more views for both A and B to get reliable results.
- Or if you are not able to increase the data set make changes in the statistical power that increase the sensitivity: make the test identify even more little changes.

22) Does small p-value indicate that effective size or difference between version A and B is large in A/B testing?

The answer to this is a big NOOO.

The sample size also matters a lot. If the sample size is too large then the test assumes that you took good care of randomness and so small change can also make p-value less than 0.05.

Similarly, if the data set is too small the test thinks that there might be more randomness so your both versions need to have large differences in order to make p-value less than 0.05.

23) With an example, prove the above answer that small p-value does not indicate large difference between 2 versions in A/B testing.

There are 2 companies A and B:

Over the years they have conducted interviews to thousands of people.

	Got the job	Did not get the job
Company A	73	125
Company B	59	131

In the above case we took small data size
Company A had 37% acceptance
Whereas company B had 29% acceptance. There is a difference of 8% percent but still we get a p-value of 0.24(>0.05).

Consider another case with large data.

	Got the job	Did not get the Job
Company A	5005	9868
Company B	4800	9000

We have a considerably large dataset here.

In this case A has a 34% acceptance rate.
Whereas B had a 35% acceptance rate. The difference is only 1% still you get a p-value of 0.04 which says both company's hiring is different.

So the size of the data set also plays a key role in measuring statistical significance.

24) Address on of the most common mistakes that happen in using statistical significance in A/B testing

Lack of fixed sample size:
One of the common practises which backfires later is using the same statistical significance level for evaluating weekly, monthly and annual data.
As we have seen in the previous question the same significance definitely did not reflect the reality when the sample size is less in the first case of the previous question.

One way to fix this is to fix your sample size in advance and stick to just a single observation at the end of a test. It might be inflexible and inefficient, but you will know what is warranted by the data at hand.

The other approach could be trying a good sequential testing method. For this we can use an A/B agile testing approach but this topic is out of scope for this article. If you are interested, dig deep in it. I would suggest this resource.

25) What could be another commonly made mistake when performing A/B testing?

Generally, at industry scale for running A/B tests. We need a sample size of at least 25000 visitors to reach a significant sample count(source from this article). Most companies perform A/B testing for small sample size which is clearly not a true indication or representation of true population or behaviour, hence, results end up being unreliable.

26) Is it sufficient to test the A/B test once?

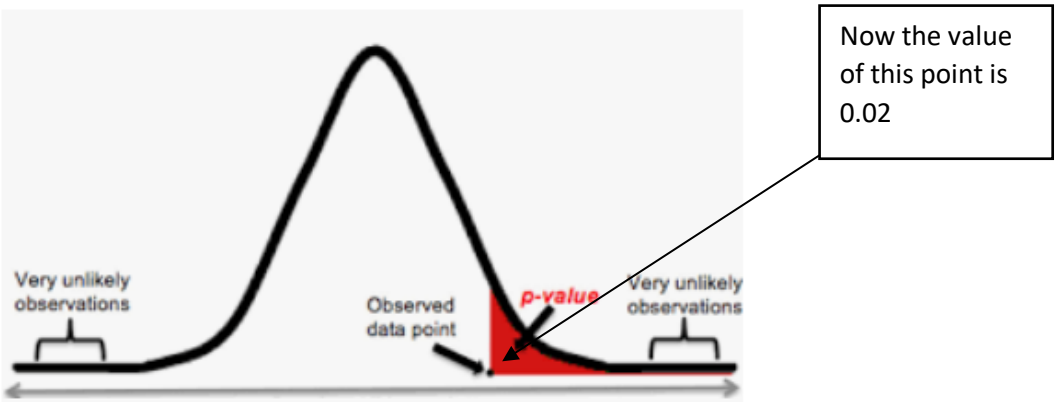
Most of the companies do this they test only once. But it is not an ideal practise.

As the meme explains, we need to perform testing every few months so that we are aware of the interests of new visitors and can measure primacy effects.

27) What is the main trade-off involved in choosing a particular statistical significance level?

People in industry often face this situation where they need to set a significance level depending on the problem. They need to weigh the factors before choosing a significance level (threshold like 0.05 when measured in p-value).

The main trade off is when we increase the statistical significance to a higher level that is making the p-value threshold smaller. Like for example shifting it from 0.05 to 0.02.



Now what happens is we reduced the threshold so high that now the difference between view count of 2 versions of a web page should be so high in order for the A/B test to say that there is an improvement.

In order for that 2 to happen at a relatively low view count difference we need to collect lots of data or conduct tests for a longer time.
For your clear understanding consider the 2nd case in question number 21.
In that case a small difference managed to reject the null hypothesis that 2 companies are similar in hiring percentages.

The trade here is accuracy with size. Since we increased the threshold now, we manage to fail the null hypothesis that truly indicates the difference between 2 versions when compared with the older threshold.

28) Explain how A/B testing works on big websites with millions of users.

When we have big websites with millions of users. The slightest improvements can result in many thousands or even millions of dollars of worth of revenue or profits. This makes them use high-sensitive tests and power quickly ramps up to the user requirement levels where even the most visited sites on earth need weeks or months to run a proper test.

29) explain how the A/B test works on a small website.

Small companies often aim big when they use A/B testing. If the view count is small then we can accept high uncertainty and fail to reject the null hypothesis. Or test so fast that low sensitivity is not such a great issue.

30) Suggest one good practise for A/B testing.

However big the sample size is, always do the test for at least 1 whole week. Because Mondays are not the same as Thursday or Friday, to get a general idea, collect the samples for the whole week even if you are satisfied for 3 days.

31) Implement A/B testing using python.

Step 1: import the libraries

```
import pandas as pd
import numpy as np
import seaborn as sns
import scipy.stats as ss
```

Step2: present the dataset

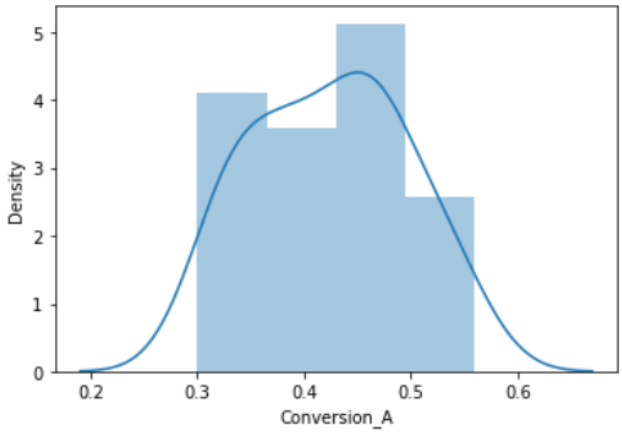
```
df.head(10)
```

	Day	Conversion_A	Conversion_B
0	1	0.40	0.48
1	2	0.34	0.50
2	3	0.46	0.46
3	4	0.48	0.54
4	5	0.44	0.48
5	6	0.36	0.44
6	7	0.38	0.46
7	8	0.56	0.50
8	9	0.32	0.44
9	10	0.40	0.54

Step3: visualize the data for both versions

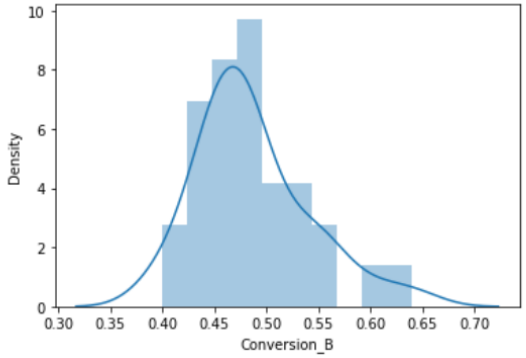
```
[11] sns.distplot(df.Conversion_A)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distr
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7fa42037
```




```
[12] sns.distplot(df.Conversion_B)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributio
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7fa41dc15c10>
```



Step 4: perform the test:

t_stat, p_value= ss.ttest_ind(df.Conversion_B,df.Conversion_A)
t_stat , p_value

```
t_stat, p_value= ss.ttest_ind(df.Conversion_B,df.Conversion_A)
t_stat , p_value
```

```
(3.7873679309192756, 0.0003637960128287791)
```

We get the p-value less than 0.05 thus rejecting the null hypothesis that both versions are similar.

32) When should you use and not use A/B testing?

This is something every analyst should be aware of: when to use A/B testing and when not to use A/B testing.

A/B testing is good and best for incremental changes: small changes like font, colour, position of data in the webpages, page load times etc.
We need some good large number of visitors in our planned stipulated time to perform the testing. If these conditions are satisfied, we are good to go with A/B testing.

A/B testing is not good for big changes like if a company introduces its new products and it puts it in its e-commerce website and see how customers respond to it. A/B test is not going to work in this case. In these cases, there may be novelty effects that drive higher than normal engagement or emotional responses that cause users to resist a change initially.

33) Is it good to consider as many metrics as possible while performing A/B tests?

This is also one of the serious mistakes that needs to be considered when you perform A/B testing.
Too many metrics leads to **"spurious correlation".** When companies prepare the test design layout it should select a few metrics and stick to it throughout the experiment. Using too many metrics leads to random fluctuations. We can interpret it like this: if there are too many metrics you tend to always look at outlier things instead of understanding what exactly data is telling you.

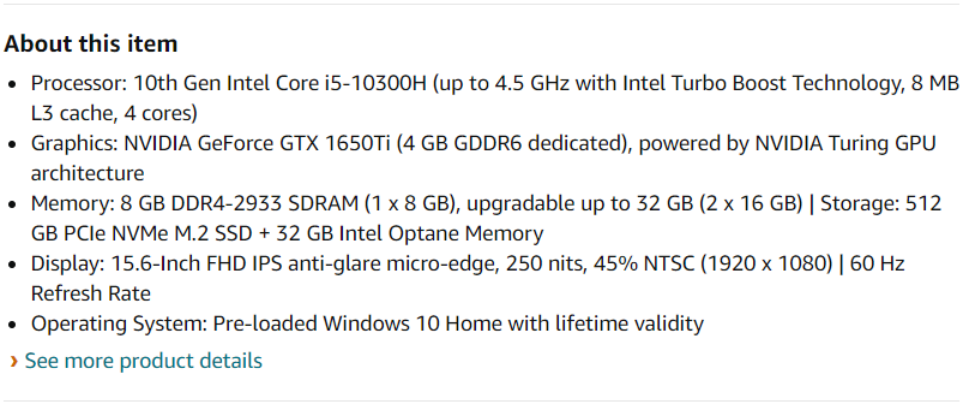
34) What kind of changes or aspects are majorly tested in A/B testing?

Changing 2 to 3 words in the last line of the front page does not exactly make a good test case for A/B testing and neither big change is useful either, so what kind of changes can be tested here.

1) **CTA (call to action)**: It is a prompt on the website that sends a specific URL. for example: these can be like LOG in, SIGN up, ADD ADDRESS etc.

2) **Images, audio and videos:** Images like below can increase CTA and these changes can also be measured by A/B testing. Make sure you are not inculcating too many changes while performing A/B testing.

3) **Captions and headlines:** understanding the mental state of the user and quoting right captions can bring a lot of conversion rate.
4) **Subject Lines:** making customers read your emails are the most important things. Shifting company emails from spam to the main folder brings a lot of conversion rate and with great subject lines this can be possible.
5) **Diving deep into content:** giving a little brief intro of the content or straight away giving all the info in the front page are 2 different options available to see where your customers land. This can be tested by A/B. You can have 1 to 5 product information on the front page or 10/15 depending on how customers respond in A/B test.
6) **Product description:**
   How should it be represented? Bullet points? Or paragraphs? A/B is your solution.



About this item

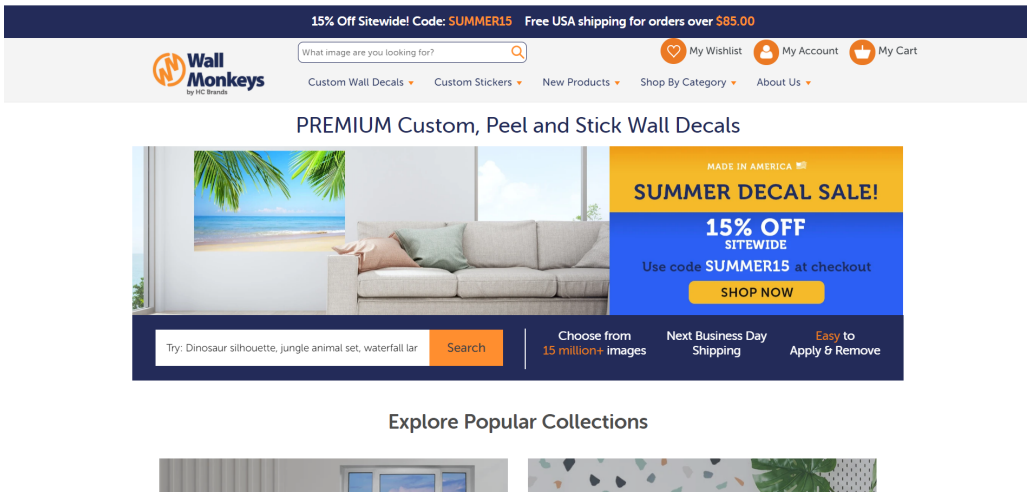
- Processor: 10th Gen Intel Core i5-10300H (up to 4.5 GHz with Intel Turbo Boost Technology, 8 MB L3 cache, 4 cores)
- Graphics: NVIDIA GeForce GTX 1650Ti (4 GB GDDR6 dedicated), powered by NVIDIA Turing GPU architecture
- Memory: 8 GB DDR4-2933 SDRAM (1 x 8 GB), upgradable up to 32 GB (2 x 16 GB) | Storage: 512 GB PCIe NVMe M.2 SSD + 32 GB Intel Optane Memory
- Display: 15.6-Inch FHD IPS anti-glare micro-edge, 250 nits, 45% NTSC (1920 x 1080) | 60 Hz Refresh Rate
- Operating System: Pre-loaded Windows 10 Home with lifetime validity

› See more product details

See how organized the points are, so from now on when you subconsciously feel attracted to a website just appreciate what kind of A/B testing has been done to make it look that way.

7) **Media mentions**
8) **Landing pages**

35) Give an example of real time A/B testing which brought huge profits to a company.



Wall Monkeys

If you are previously unaware of this company. This company sells myriad wall decals for home or for professional infrastructures.

This company took the help of another company named Crazy egg. Which has various CRO Tools for conducting various tests on website data.
It ran an A/B test for our WallMonkeys company.

36) What is the goal of the A/B test for this case Study?

WallMonkeys wanted to optimize its homepage for clicks and conversions. It started with its original homepage, which featured a stock-style image with a headline overlay.

After conducting behaviour tests with the initial designs. The company exchanged the stock-image style with an alternative that would show visitors the opportunities they could enjoy with Wall Monkey Products.

Now with these changes they conducted an A/B test. The results were good; they got an increment of 27% more that of the control group.

Satisfied with the results but it did not make them stop the test. Wall Monkey continued the test.
This time the second change was the business replaced its slider with a prominent search bar. The intention was that customers would be more drawn to searching for items in which they were specifically interested.

Now get ready for some big BAMMMM!

The second A/B testing example resulted in a conversion rate increase of 550 percent. This is definitely an incredible result for any range company.

37) What can we draw from this Case Study with respect to the A/B test?

In the first A/B test wall monkey got a pretty good result. If they have compromised or satisfied with the result. They would not have achieved this incredible result. This shows that the A/B test has a lot of potential if you could bring the right changes.

38) Companies perform A/B testing with its new version. Everything seems good. We got some improvement but when we completely launched it. The new version is not as encouraging as it was when we performed A/B testing. What happened here and how can we alter tests so that we do not fall into false hopes this way?

To understand this situation, we need to be familiar with 2 effects namely:

1) Primacy effect
2) Novelty effect

Primacy effect: it can also be called change aversion. As the name suggests, in this case people are reluctant to change. They are pretty much satisfied with the old version.

Novelty effect: This is exactly opposite to the primacy effect. The other set of customers may like the new addition or change and contribute towards high conversion rate in the test.

So how do we deal with the situation?

One obvious way and people's favourite solution is just ignore it. This actually happens, companies just move on with making another change and performing a new A/B test at some point in future afterwards.

One viable solution is, we could run tests only on first-time users because the novelty effect and primacy effect obviously doesn't affect such users.

But the company is at the middle of the test and it wants to check these effects then you could do these things:

1) Compare new users in the control group to those in the treatment group to evaluate novelty effect.
2) Compare first time users' results with existing users' results in the treatment group to get an actual estimate of the impact of the novelty or primacy effect.

39) What if your company tells you to find the best design for a search bar out of 4 different options available? How would you perform that? What kind of changes would you bring to A/B testing?

When we have multiple versions to test we can do A/B testing but with some changes in it. One commonly used method in A/B is to use Bonferroni correlation. It divides the significant level by 0.05 with the total number of variant groups, in this case it is 4: 0.05/4 =0.0125.
We set the p-value threshold to reject the null hypothesis of 0.0125.

40) What are some of the reasons for A/B tests not able to provide desired insights or outcomes?

Some of the main possible reasons are as follows:

1) Considering too many metrics.
2) Did not select proper metrics for the desired change.

3) Not having proper sample size.
4) Not able to interpret a statistically insignificant p-value.