Hypothesis Testing starts with the formulation of these two hypotheses:

* **Null hypothesis (H₀)**: The status quo
* **Alternate hypothesis (H₁)**: The challenge to the status quo

**How to choose Null Hypothesis or Alternate hypothesis?**

The null hypothesis always has the following signs:

1. = equal to
2. ≤ maximum or at max
3. ≥ at least

The alternate hypothesis always has the following signs:

1. ≠ not equal to
2. > greater than
3. < less than

NOTE: Alternate Hypothesis is what researcher wants to prove.

**++++++++++++++++++++++++++++++++++++++++++++++++++++++**

**Situation 1**: Amazon claimed that its total valuation in December 2016 was **at least** $14 billion.

**Analysis: Now find here what is Null and alternate hypothesis?**

Here, the claim contains ≥ sign (i.e. the at least sign), so the null hypothesis is the original claim.

The hypothesis in this case can be formulated as:

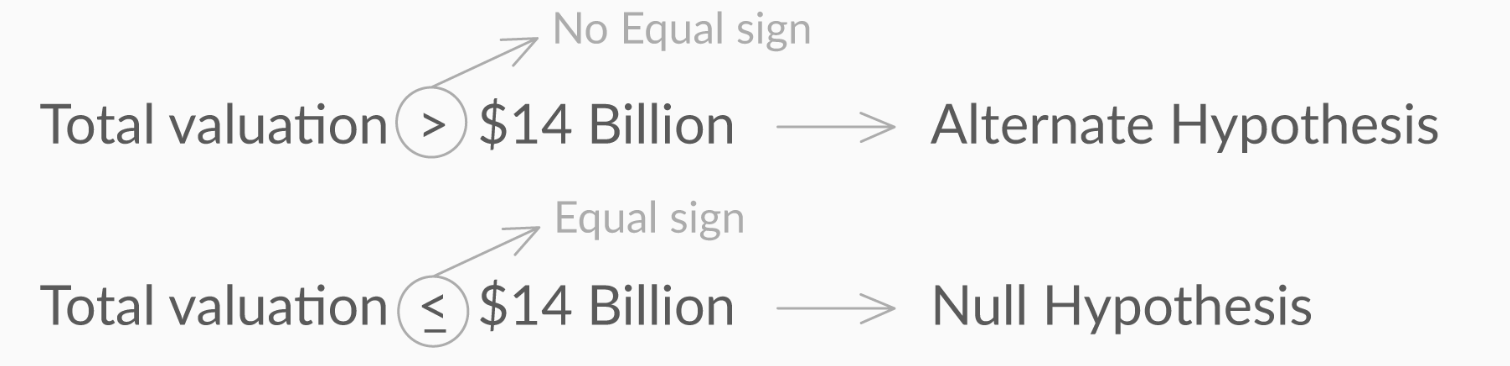


**Situation 2:** Amazon claimed that its total valuation in December 2016 was greater than $14 billion.

**Analysis: Now find here what is Null and alternate hypothesis?**

Here, **the claim contains > sign (i.e. the ‘more than’ sign)**, so the null hypothesis is the complement of the original claim.

The hypothesis in this case can be formulated as:



**Situation 3**: The average commute time for an Edureka employee to and from office is at least 35 minutes.

What will be the null and alternate hypotheses in this case if the average time is represented by μ?

Answer : H₀: μ ≥ 35 minutes and H₁: μ < 35 minutes

*The null hypothesis is always formulated by either = or ≤ or ≥ whereas the alternate hypothesis is formulated by ≠ or > or <. In this case, the average time taken was greater than or equal to 35 minutes. So, that becomes the null hypothesis. Less than 35 minutes becomes the alternate hypothesis.*

**How to make the decision to either reject or fail to reject the null hypothesis?**

**NOTE: we never accept NULL hypothesis instead we say “fail to reject the NULL hypothesis”**

**Ans**: Suppose your Friend Ravi said that his Average (population mean) archery score is 70 points.

**Analysis:**

So then Null Hypothesis H₀: μ = 70

Alternate Hypothesis: H₁: μ ≠ 70

To test the Hypothesis or to check if your friend is telling the truth or lie, let's say you take 5 samples and the sample mean score is 50.

So the question is how do we decide the critical value (above or less than that if a sample mean comes then we can say make decision to fail to reject or reject the null hypothesis).

**Lower Critical Value**: The Sample mean value below which if sample mean comes then we can Reject the Null Hypothesis.

**Upper Critical Value**: The Sample mean value above which if sample mean comes then we can Reject the Null Hypothesis.

**Acceptance Region**: Range of values of sample mean for which Null hypothesis can’t be rejected.

**NOTE**: if p < 5% then feature is significant, and that input feature would impact the output.

**Explanation**: In the class If you remember I had mentioned that if a input column(or feature) is having a strong correlation with Output column(remember Ice-cream sales and deaths by drowning example) then it doesn’t necessarily mean that I/P feature is impacting the output( or meaning feature is significant or not).

If an input feature impacts the output, then output then we say it is a significant feature and this decision we can make using the p-value of that feature after building the model.

Just remember if p-value of a feature is less than 0.05 or 5% then that feature is significant and if p-value is more than 0.05 then that feature is insignificant and therefore that feature must be dropped out from the input data set.

+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

**One Extra information Topic that is not present in the slides: see below.**

**p- value method for Hypothesis testing**

**Definition**: P-VALUE is the probability of failing to reject the null hypothesis.

OR

p-value as the probability that the null hypothesis will not be rejected( or failing to reject).

The higher the p-value, the higher is the probability of failing to reject a null hypothesis.

The lower the p-value, the higher is the probability of the null hypothesis being rejected.

**High p-value means accept Null hypothesis…… which is generally not good.**

**Low p-value means Reject the Null hypothesis…..which is generally what we want to do.**

**NOTE**: Alternate Hypothesis is what Data Scientist want to prove.