

HYPOTHESIS, NULL HYPOTHESIS, P Value

→ Hypothesis is an informed guess, with a reasonable chance of being right.

OR

A statement temporarily accepted as true.

NULL HYPOTHESIS →

It proposes that no statistical difference significance (the extent to which) a result deviates from expected value arising from random variation or error in sampling exists in a set of given observations.

NULL HYPOTHESIS attempts to show that no variation exists between variables or "simply that a single variable is no different than its mean"

NULL HYPOTHESIS is represented by " H_0 "

Eg → Earth is spherical in shape.

Now there is an alternative hypothesis associated with NULL Hypothesis. In above example alternative hypothesis would be that →

"Earth is not spherical in shape".

Another example can be →

→ There is no difference in education provided by government schools and private schools in DELHI.

This will be - the null hypothesis and the alternate hypothesis is There is a difference in education provided by govt. schools and pri. schools in Delhi.

Now, how to decide which hypothesis stands true, it'll be decided by 'P' Value. P value is probability and lies between 0 and 1.

P value measures compatibility of data with NULL hypothesis

- High P value → data likely with NULL hyp.
- Low P value → data unlikely with NULL.

Low P value suggests that the "sample" provides enough evidence to reject NULL hypothesis for entire "population".

Now how to decide how 'low' is enough to reject NULL hypothesis. This will be decided by a comparison b/w 'P' value and α , which is significance level.

Q What is α (Significance level)?

→ α is basically a fixed value of .01 or .05 i.e 1% or 5%.

and this is the threshold.

This is taken because there could be error in sample.

Now, this significance level is assumed before experiment either .05 or .01.

Q How to calculate P value?

→ It can be calculated from Z-table.

$$Z = \frac{\text{Value} - \text{Mean}}{\text{S.D.}}$$

I'll start explaining all the above mentioned concepts with LAYMAN examples.

First eg is based on "TRUE EVENTS"

→ An Indian mother thinks that her son is very obedient and is not NAUGHTY at all.

Now she received a complaint from the school that her son is NAUGHTY however the mother denies to accept it.

This is NULL HYPOTHESIS that the kid is "NOT NAUGHTY".

She again receives - the complaint 2nd time and she still doesn't believes - the school.

After few days she has received complaints 4 or 5 times, now she starts doubting herself and - the NULL hypothesis.

This point will be significance level (α).

→ Now out of 10 complaints received against her son, it came out that actually it was her son's mistake 8 times out of 10.

Now there is enough evidence to prove - the mother is wrong and to reject NULL hypothesis.

In this example no actual stats is used it is a very Layman idea of NULL Hypothesis, P-Value, significance level.

Example 2. I →

I built a machine to predict weather and I want funding for my start-up.

I went to the investors and I've to prove that the machine works fine however until proven the investors believe that the machine doesn't work.

$H_0 \rightarrow$ Null hypothesis will be that the machine doesn't work.

$H_a \rightarrow$ Machine does work.

Now I'll do 100 experiments and will decide from the probability if the machine works or not.

In this case $\alpha = .05$.

→

After the experiment is done for 100 times, it came out that it gave wrong predictions for 3 times.

$$P = \frac{\text{no. of wrong prediction}}{\text{Total no. of prediction}}$$

$$P = \frac{3}{100} = .03$$

Now compared to $\alpha = .05$ $P = .03$ is low and thus NULL has to go, which is machine doesn't work.

→ Thus it is concluded that
- the machine works just fine
and - the inventor is happy.

Example 3-

Now I'll take an example using Mean, Z, P, & A t-test NULL hypothesis' example.

→ Let's say I have a chocolate manufacturing company (RAHUL'S CHOCOLATE).

I'm putting 20 gram peanuts in 100 gram of chocolate bar.

So the null hypothesis is that each chocolate bar contains 20 gram of peanuts.

→ Now if someone comes up and say there is not enough peanuts in the chocolate, and it must be checked.

Now, here the stats really start.

I took a sample of 250 chocolates from 1000 chocolate population.

I melted the chocolates and calculated the weight of peanuts from chocolates.

The m.

Here I'll take 2 cases.

① The mean of chocolates is weight of grams. peanuts is 19 grams.

② The mean weight of peanuts in chocolates is 10 grams.

*** — $\alpha = .05$ significance value ***

Case 1 → The weight of peanuts is 19 gram. and $\sigma = 2$ grams

$$Z = \frac{19 - 20}{2} \quad (\frac{\text{Value} - \text{Mean}}{\sigma})$$

$$Z = \frac{1}{2} = -.5.$$

$$\text{for } Z = -.5 \quad P = .308538$$

Here. $P = .308538$ which is larger than $\alpha = .05$ \therefore there is no problem with the manufacturing of chocolates and we accept NULL HYPOTHESIS.

② The mean weight of peanuts in 10 grams in chocolate.

Again $\alpha = .05$, $\sigma = 2 \text{ gram}$,

$$Z = \frac{\text{Value} - \text{Mean}}{\sigma} = \frac{10 - 20}{2} = -5$$

$Z = -5$ and P value = .00001

Hence, $P = .00001$ which is very low as compared to $\alpha = .05$ therefore NULL has to go.

In simple terms there is enough evidence that the chocolates contains way too less of peanuts as the company claims.

Hence NULL HYPOTHESIS has to go.

BY:

RAHUL SHARMA
INSAID