Topics :- (1) Understanding Hypothesis Testing
(2) Null & Alternate hypothesis testing
(3) Making a Decision
(4) Critical Value Method

Hypothesis Testing

We are data scientist not data administrators so for us data alone is not interesting.

We are interested in the interpretation of the data.

So when we work to pose questions / hypothesis, we use statistical methods that output confidence or likelihood about it.

In general this type of methods falls under the category of statistical hypothesis testing or significance test.

We usually perform a hypothesis test to calculate some quantity/metric while making some assumption.

And that metric will inform us whether or not our assumptions were right.

Since we cannot test each & every one from the population for a given assumption, we create a sub sample.

There are some techniques for the sample size determination they typically use { Z-score, margin of error & standard deviation. }

For an binomial population :→ $\text{Sample size} = \frac{Z^2 \times P(1-P)}{e^2}$

TYPES OF HYPOTHESIS

Null Hypothesis

'H_0' we try to discredit the assumption.

It's a form of "And Reducto Absurdum" meaning of the assumption was correct something Absurd would happen.

So we under to validate our measurement/hypothesis we will try to discredit the Null hypothesis.

Alternate hypothesis

The alternative hypothesis denoted by H_1/H_a is the hypothesis that sample observation are not purely by chance or there's underlying cause that's not random in nature.

Can we accept the Null hypothesis ?

(Research) — A hypothesis test has 2 outcomes: Accept / Reject

(Statistician) — issue with accepting the null hypothesis. They say you can either Reject or fail to Reject the null hypothesis.

The difference b/w acceptance & failure to reject is that acc implies that the null hypothesis is true while the other implies that there is not sufficient evidence to support the alternate hypothesis over the null hypo.

Hypothesis Testing :-

Formal Procedure of hypothesis testing

(1) State the hypotheses in a mutually exclusive manner.
(2) Formulate an analysis plan → decide over the test to use
(3) Analyze the sample data → find the value for the test metric (like z-score, t-score, z-score)
(4) Interpret Results.
↳ If the value of your test score is unlikely based on the null hypothesis, Reject it.

Making a decision :-

Decision depends on the nature of your test and the value you get in the end.

To validate/invalidate a hypothesis you apply decision rules.

While making these decision we may make

Type-I error :- When the researcher reject the Null hypothesis even when its true.
Prob of committing a type-I error is called significance level or alpha α.

Type-II error :- This happens when we fail to reject the null hypothesis given its false.
It defined as Beta or β & the probability of not committing type II errors is called Power!

$\alpha \uparrow \rightarrow$ tend to reject the Null

Power of the test $\uparrow$ ⇒ acceptance of Null $\uparrow$.

Ideally we want significance level that minimizes the two types of error.

Like 5-sigma in physics refers to a significance level of 5×10^{-7} or in other words

The probability that we found the results by chance is 1 in 3.5 million independent repeats of experiment

Critical Values :→

So hypothesis test might not return a p-value/z-score etc. they might instead return a list of critical values and their associated significance level.

These are usually observed in Non-Parametric or distribution free statistical hypothesis test.

Works in a similar manner:
test stats < Critical value :→ fail to reject
test stats ≥ Critical value :→ Reject.

Confidence Interval :-

The 95% in a 95% confidence interval tells us that if we calculated a confidence interval from 100 different samples, about 95% of them would contain the true population.

This confidence is in the fact that the procedure of calculating these confidence intervals will only exclude (the population mean) 5% of the time.

Z-score :-

JEE Main vs RPET

$\left(\frac{x-\mu}{\sigma/\sqrt{n}}\right)$ ≈ $\left(\frac{x-\mu}{s.d}\right)$

Z-score helps us compare quantities that are not at the same level of measurement.

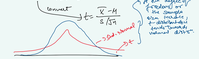
For Hypo. Testing, we need samples to be :-
- Random
- Normal
- Independence

We also need (Significance level) $\alpha = 0.05$

t-Distribution :→

$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ } we don't know σ

$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ } $\sigma \approx s$



A fn defined by the sample size (n-1 degrees of freedom) and "normal value" 95%

Two Sample mean test :-

Q. Driver of Honda & Car

Car: μ = 1.3 miles, σ = 0.5 miles } 22 rides

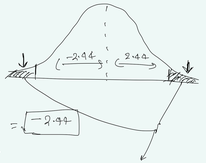
Honda: μ = 1.6 miles, σ = 0.3 miles } 24 rides

for α = 0.05, which unit better?

$H_0: \mu_A = \mu_B$
$H_a: \mu_A \neq \mu_B$

⇒ Assuming the null being true:

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} = \frac{1.3-1.6}{\sqrt{\frac{0.25}{22} + \frac{0.09}{24}}} = -2.94$$



Prob (shaded region)
$= P(|t| \geq 2.44) \approx 0.024$

∴ p-value < α ⇒ Reject H_0
H_a

Two sample Proportion test :→

Q. Researchers suspect that use PC is becoming more common with students.

Data:	Year	No. of Students	No. of students with PC
	2000	600	132
	2015	600	228

H_0 :- $P_{2015} = P_{2000}$

H_a :- $P_{2015} > P_{2000}$

Step 1 :- Calculate Z-score :-

$$z = \frac{\hat{p}_{2015} - \hat{p}_{2000}}{\sigma_{\hat{p}_{2015} - \hat{p}_{2000}}} = \frac{\frac{228}{600} - \frac{132}{600}}{\sqrt{\frac{0.22 \times 0.78}{600} + \frac{0.22 \times 0.78}{600}}}$$

$\left\{ \hat{p} = \frac{228+132}{600+600} = \frac{360}{1200} = \frac{3}{10} \right\}$

$z = 1.61$

P-value $(z \geq 1.61) = 1 - 0.9463 = 0.0537$ (2 dec)

∴ P-value > α ⇒ we fail to reject for H_0

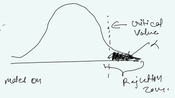
	Null Hypothesis	Alternative Hypothesis
Null Hypothesis (H_0)	No error ($1-\alpha$)	Type 2 error
Alternative Hypothesis (H_a)	Type 1 error (α)	No error

A/B testing :- To compare 2 different products. It consists of a randomized experiment with two variant A & B.

eg :-

Assumed Distribution	Example Case	Std Test	Alt. Test
Gaussian / normal	Avg $ / user	Z-test / t-test	Student t-test
Binomial	Click-through rate	Fisher's exact test	Barnard's test
Poisson	Text^n / week	E-test	C-test
Multinomial	No. of ads per product	chi-sq test	Multi-Variate Z-test / Tukey Simplify
Unknown			

Directional Hypothesis :-



eg :- Females will score significantly higher than males on IQ scores.

Non-Directional Hypothesis

eg :- Females will score significantly different than males on IQ test.

Z-test :-

{ → we know the population variance or
→ we don't know pop. var. but n ≥ 30 (large enough)

Z-score $= \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ (sample mean, population mean, population standard deviation, sample size)

But when { ① We know about the population variance
② Sample size small. (n < 30) }

$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ (s: Sample standard deviation)

One Sample Z test

Q. Rents score higher than 600 on IQ exam

Z-score $= \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Hypothesis: $H_0: \mu \leq 600$, $H_a: \mu > 600$

⇒ n = 36

P(X ≥ 36) = 0.0668 (one tail)
Critical value = 1.645

Z-score > critical val.
p-value < 0.05

Two Sample Z test

Q. Calculate scores higher than boys by 10 marks.

Z-Score $= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$H_0: \mu_1 - \mu_2 \leq 10$ ✓
$H_1: \mu_1 - \mu_2 > 10$ ✗

P-val.

Z-score < critical value
P-val > 0.05

Chi-Square Table (great for Categorical)

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$

Goodness of fit / Test of independence

$\chi^2 = \sum \left\{ \frac{(\text{observed} - \text{expected})}{\text{expected}} \right\}$??

Expected Value $= \frac{(\text{Row})_{total} * (\text{Column})_{total}}{(\text{Total Values})}$

Eg: A game has come out with following distribution of players.

(Goodness of fit)

	character selection			
	Healer.	Tank.	Assassin	Fighter.
Expected :-	15%	20%	20%	45%
Observed :-	25	35	50	70
for 200 players :-	30	40	40	90

df = 4-1 ⇒ 3

$\frac{(O_{bs_1} - \exp_1)^2}{\exp_1} + \frac{(O_{bs_2} - \exp_2)^2}{\exp_2}$

⇒ $\frac{25}{30} + \frac{35}{40} + \frac{50}{40} + \frac{70}{90}$

→ 0.83 + 0.625 + 2.5 + 0.8 = 3.75

Hence there is no statistically significant evidence that the game distribution was unfair.