BMT_Public_Value
Friday, 9 July, 2021 9:47 PM

4) statistical decision:

In hypothesis testing, the p-value and critical value are used to make decision whether to reject the null hypothesis

Probability (P)-Value:-

If is the probability that the test statistics equals the observed value or a value even more extreme, assuming that the null hypothesis is true.

It helps determine the strength of the evidence against the null hypothesis.

$$\boxed{\text{p-value} = P(\text{test statistics takes observed value or beyond in tails} \mid H_0 \text{ true})}$$

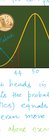
Example: Suppose that we want to test whether a coin is fair or, we have:

H_0: the coin is fair ($P(H) = P(T) = 0.5$)

H_1: the coin is not fair

For that let we have 100 coin flips.

If we draw the prob distribution of the number of heads:



Let we observe 60 heads in our 100 coin flips. Then, the p-value is the probability of getting a 60 heads (test statistics) equals the observed value (60) or a value even more extreme.

The p-value in the above example is 0.045. This p-value is calculated using z-score (see later) how to calculate it.

The above p-value means that if the coin is fair (H=T=50), the prob of getting 60 heads (or a 60 heads more extreme) is 0.045.

Let us now assume that we get 60 heads in 100 coin flips.

p-value now is 0.052.

The value 0.052 of p-value indicates that if the coin is fair, the prob of getting 60 heads (or a 60 heads more extreme) is 0.052.

We can start thinking that the null hypothesis may not be true because we are getting 60 heads out of 100 coin flips.

p-value (60 heads in 100 coin flips) = 0.0001 (almost zero)

Thus, smaller the p-value stronger is the evidence against the null hypothesis.

To put it in other words, larger is the difference between sample statistic and hypothesized population parameter, smaller is the p-value and stronger is the evidence against the null hypothesis.

The obvious question now is what should be the p-value to reject a null hypothesis?

The answer is the significance level (α).

Significance level (α):- The significance level or is the threshold for rejecting the null hypothesis when it is actually true.

Commonly used value of α is 0.05, meaning there is 5% chance of rejecting a null hypothesis when it is actually true. If α set by the researchers.

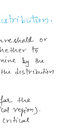
Interpretation:

1. If p ≤ α (significance level), we reject the null hypothesis, suggesting that the observed data is statistically significant.
2. If p > α, we fail to reject the null hypothesis, meaning that there is not enough evidence to support the alternative hypothesis.

How to map z-score to probability & vice-versa:-

It is calculated using z-score table. A part of this table is shown below. Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...										
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441



For example, let us find the value of z for cumulative prob of 0.9418. From the table this corresponds to $z = 1.3 + 0.04 = 1.34$. Then to get a cumulative prob of 0.9418, we have to move 1.34 σ away from the mean:

$$\int_{-\infty}^{1.34} f_N(x)\,dx = 0.9418$$

Where $f_N(x)$ is the normal distribution.

Critical value:- A critical value is a threshold or cut-off point that helps in deciding whether to reject the null hypothesis. It is determined by the chosen significance level (α) and the distribution of the test statistics.

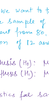
The critical value marks the boundary for the region of rejection (also called the critical region). If the test statistics falls beyond this critical value, the null hypothesis is rejected.

Critical value for a two-tailed test:-

Critical value for a two-tailed (two-sided) test for a given α is the z-score at the value $\alpha/2$.

For example, for $\alpha = 0.05$ the critical value is the z-score for $\alpha/2 = 0.025$, which is ±1.96. As show in fig.

The z-values in the red portion are the region of rejection for null hypothesis



(a) Critical value left tail

(b) Critical value right tail

Critical value for a one-sided test:-

The critical value for a one-sided (one-tailed) test for a given α is the z-score at the value α. For the above example with $\alpha = 0.05$, the critical value for one-tailed test is the z-score at $\alpha = 0.05$ which is 1.64 for right tail and -1.64 for the left tail.

Example:

Scenario: We want to test if the average score on test for a sample of 15 students is significantly different from 80, using the population standard deviation of 15 and sample mean of 85.

Solution:

Null hypothesis (H_0): $\mu = 80$

Alternative hypothesis (H_1): $\mu \neq 80$ (two-tailed test)

We can find z-statistics for sample as

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{85 - 80}{15/\sqrt{15}}$$

$$\boxed{z = 1.61}$$

We never need to find p-value from the z-score from the z-score table. $z = 1.61$ corresponds to the cumulative prob of $P = 0.9463$.

The area under the upper tail beyond the z-value $1 - P = 1 - 0.9463 = 0.0537$

Since it is two-tailed test, the total p-value

$$\text{p-value} = 2 \times 0.0537 = 0.1074$$

The critical value with $\alpha = 0.05$ for two tail test is 0.025, which leads to critical z-value of ±1.96 (using the z-table).

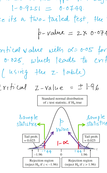
Critical z-value = ±1.96

We see that p-value (0.1) > α/2 (0.025). So we fail to reject the null hypothesis.

In other words, there is not enough evidence to conclude that the average score is significantly different from 80.

or

The Critical z-value (1.96) is higher than 1.61, measuring that the z-statistic does not fall in the rejection region. Thus, there is not enough evidence to reject the null hypothesis.



Example:

Scenario: A company claims that 60% of its customers are satisfied with their service. A survey of 200 customers shows that 150 are satisfied. We want to test if the proportion of satisfied customers is different from the claimed 60% at $\alpha = 0.05$.

Solution:

Null hypothesis (H_0): $p = 0.6$

Alternative hypothesis (H_1): $p \neq 0.6$

The sample size $n = 200$

Recall that the z-statistics for proportion is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \qquad \hat{p} = \frac{150}{200} = 0.75$$

$$p_0 = 0.60$$

$$n = 200$$

$$z = \frac{0.75 - 0.60}{\sqrt{\frac{0.6(1-0.6)}{200}}} = \frac{0.15}{0.03464} \approx 1.44$$

From the z-table, the cumulative prob for z = 1.44 is approximately 0.9251.

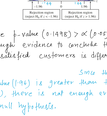
The area in the upper tail beyond this z-value is $1 - 0.9251 = 0.0749$.

Since it is two-tailed test, the total p-value

$$\text{p-value} = 2 \times 0.0749 = 0.1498$$

The critical value with $\alpha/2 = 0.05$ for two tail test is 0.025, which leads to critical z-value ±1.96 (using the z-table).

Critical z-value = ±1.96



Since p-value (0.1498) > α (0.05), there is not enough evidence to conclude that the proportion of satisfied customers is different 60%.

or

Since the critical z-value (1.96) is greater than the z-statistic (1.44), there is not enough evidence to reject the null hypothesis.