A Fully Detailed, Intuitive Example
How the p-value Is Calculated for Two Small Datasets

Overview

In this example, we manually compute and interpret the *p-value* step by step using two very small datasets. The goal is intuition and understanding rather than heavy mathematics.

Step 1: The Data (Very Small Samples)

We have two groups:

Group A

$$10,\ 12,\ 11$$

Group B

$$14,\ 15,\ 16$$

At first glance, Group B looks larger. However, statistics asks a deeper question:

> *Is this difference real, or could it just be random variation?*

Step 2: The Null Hypothesis

The null hypothesis (H_0) states:

> The two groups come from populations with the same mean.

We therefore assume *no real difference* and test how surprising our observed data is under this assumption.

Step 3: Compute the Sample Means

Group A mean

$$\bar{x}_A = \frac{10 + 12 + 11}{3} = 11$$

Group B mean

$$\bar{x}_B = \frac{14 + 15 + 16}{3} = 15$$

Difference of means

$$\bar{x}_B - \bar{x}_A = 4$$

Important: A difference of 4 alone means nothing until we compare it to the variability in the data.

Step 4: Compute the Sample Variance (Noise)

Group A Values: $10, 12, 11$
Mean: 11
Deviations:

$$-1,\ +1,\ 0$$

Squared deviations:

$$1,\ 1,\ 0$$

Sample variance:

$$s_A^2 = \frac{1+1+0}{3-1} = 1$$

Group B Values: $14, 15, 16$
Mean: 15
Deviations:

$$-1,\ 0,\ +1$$

Squared deviations:

$$1,\ 0,\ 1$$

Sample variance:

$$s_B^2 = \frac{1+0+1}{3-1} = 1$$

Observation Both groups have low variance, meaning there is little noise in the data.

Step 5: Compute the Standard Error

The standard error of the difference in means is:

$$SE = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

$$SE = \sqrt{\frac{1}{3} + \frac{1}{3}} = \sqrt{\frac{2}{3}} \approx 0.816$$

Interpretation If there were no real difference, we would expect mean differences of about 0.8 just by chance.

Step 6: Compute the t-statistic

$$t = \frac{\text{observed difference}}{\text{expected random difference}}$$

$$t = \frac{4}{0.816} \approx 4.9$$

Meaning The observed difference is almost five times larger than random noise.

Step 7: Degrees of Freedom

$$df = n_A + n_B - 2 = 3 + 3 - 2 = 4$$

Degrees of freedom determine which reference t-distribution is used.

Step 8: Convert t-statistic to p-value

We now ask:

> If the null hypothesis were true, how likely is it to observe a t-value this extreme?

Using a t-table for $df = 4$:

$\|t\|$	p-value
2.78	≈ 0.05
4.60	≈ 0.01

Our value:

$$t = 4.9 \ \Rightarrow \ p < 0.01$$

Step 9: Final Conclusion

Because:

$$p < 0.05$$

we conclude:

> The difference between the two groups is statistically significant and very unlikely to be due to random variation.

What the p-value Really Means

Correct meaning The probability of observing a difference this large or larger, assuming no real difference exists.

Wrong meanings

- Probability that the null hypothesis is true
- Size of the effect
- Probability that the result will repeat

One-Sentence Intuition

> *The p-value tells us whether the observed difference is too large to be explained by noise.*

Machine Learning Connection

Replace:

- Group A / Group B $\rightarrow$ Model A / Model B
- Numbers $\rightarrow$ accuracy, loss, F1-score

The same logic applies:

> Is the improvement real, or just randomness due to sampling?