

Important Mathematical Formulas with Examples

1. Mean (Average)

Formula: $\text{Mean} = (\text{Sum of all values}) / (\text{Number of values})$

Examples:

- [10, 20, 30] $\rightarrow \text{Mean} = 60 / 3 = 20$
- [4, 6, 8, 10] $\rightarrow \text{Mean} = 28 / 4 = 7$
- [100, 200, 300, 400, 500] $\rightarrow \text{Mean} = 1500 / 5 = 300$

2. Median

If odd count: Middle value. If even: Average of two middle values.

Examples:

- [3, 5, 7] $\rightarrow \text{Median} = 5$
- [4, 6, 8, 10] $\rightarrow \text{Median} = (6+8)/2 = 7$
- [10, 20, 15] $\rightarrow \text{Sorted: [10, 15, 20]} \rightarrow \text{Median} = 15$

3. Mode

Most frequent number.

Examples:

- [2, 2, 3, 4, 5] $\rightarrow \text{Mode} = 2$
- [1, 1, 2, 3, 3, 3] $\rightarrow \text{Mode} = 3$
- [4, 5, 6, 7] $\rightarrow \text{No mode}$

4. Standard Deviation

Formula: $\sigma = \sqrt{[\sum (x_i - \text{mean})^2 / N]}$

Examples:

- [2, 4, 4, 4, 5, 5, 7, 9] $\rightarrow \text{Mean} = 5, \text{Variance} = 4 \rightarrow \text{SD} = 2$
- [10, 10, 10, 10] $\rightarrow \text{SD} = 0$
- [5, 10, 15] $\rightarrow \text{Mean} = 10, \text{SD} \approx 4.08$

5. Z-Score

Formula: $Z = (x - \text{mean}) / \text{standard deviation}$

Examples:

- $x = 20, \text{mean} = 10, \text{std} = 5 \rightarrow Z = 2$
- $x = 5, \text{mean} = 10, \text{std} = 5 \rightarrow Z = -1$
- $x = 10, \text{mean} = 10, \text{std} = 5 \rightarrow Z = 0$

6. Term Frequency (TF)

Formula: $\text{TF} = (\text{word count in doc}) / (\text{total words})$

Examples:

- 'apple apple orange' $\rightarrow \text{TF}(\text{apple}) = 2/3 \approx 0.66$
- $\text{TF}(\text{orange}) = 1/3 \approx 0.33$
- $\text{TF}(\text{banana}) = 0$

7. Inverse Document Frequency (IDF)

Formula: $\text{IDF} = \log(\text{Total docs} / \text{Docs with word})$

Examples:

- Word in 1 of 5 $\rightarrow \text{IDF} = \log(5/1) \approx 0.699$
- Word in all 5 $\rightarrow \text{IDF} = \log(5/5) = 0$
- Word in 2 of 5 $\rightarrow \text{IDF} \approx 0.397$

8. TF-IDF

Formula: $\text{TF-IDF} = \text{TF} \times \text{IDF}$

Examples:

- $\text{TF} = 0.5, \text{IDF} = 0.69 \rightarrow \text{TF-IDF} = 0.345$
- $\text{TF} = 0.2, \text{IDF} = 0.4 \rightarrow \text{TF-IDF} = 0.08$
- $\text{TF} = 0.1, \text{IDF} = 0.0 \rightarrow \text{TF-IDF} = 0$

9. Accuracy

Formula: $\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$

Examples:

- $TP=5, TN=3, FP=1, FN=1 \rightarrow \text{Accuracy} = 8/10 = 80\%$
- 45 correct out of 50 $\rightarrow \text{Accuracy} = 90\%$
- 10 correct out of 20 $\rightarrow \text{Accuracy} = 50\%$

10. Precision

Formula: $\text{Precision} = TP / (TP + FP)$

Examples:

- $TP=5, FP=1 \rightarrow \text{Precision} = 5/6 \approx 83.3\%$
- $TP=9, FP=1 \rightarrow \text{Precision} = 9/10 = 90\%$
- $TP=2, FP=3 \rightarrow \text{Precision} = 2/5 = 40\%$

11. Recall

Formula: $\text{Recall} = TP / (TP + FN)$

Examples:

- $TP=5, FN=1 \rightarrow \text{Recall} = 5/6 \approx 83.3\%$
- $TP=8, FN=2 \rightarrow \text{Recall} = 8/10 = 80\%$
- $TP=3, FN=1 \rightarrow \text{Recall} = 3/4 = 75\%$

12. F1 Score

Formula: $F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

Examples:

- $P=0.833, R=0.833 \rightarrow F1 \approx 83.3\%$
- $P=0.9, R=0.8 \rightarrow F1 = 2 \times 0.72 / 1.7 \approx 84.7\%$
- $P=0.6, R=0.75 \rightarrow F1 \approx 66.7\%$

13. Confusion Matrix Terms

TP: predicted YES, actually YES

TN: predicted NO, actually NO

FP: predicted YES, actually NO

FN: predicted NO, actually YES

Used to calculate Accuracy, Precision, Recall, F1.

14. Summary

- Use Mean, Median, Mode to describe data.
- Use SD, Z-Score to understand spread.
- Use TF, IDF, TF-IDF for text analysis.
- Use Accuracy, Precision, Recall, F1 to evaluate models.