

# Descriptive Statistics

1. **Mean:**  $\bar{x} = \frac{\sum x_i}{n} \rightarrow$  average
2. **Weighted Mean:**  $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} \rightarrow$  weighted avg
3. **Median:** Middle of ordered data
4. **Mode:** Most frequent value  
data.science.beginners
5. **Range:**  $R = \max(x) - \min(x) \rightarrow$  spread
6. **Variance (Population):**  $\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$
7. **Variance (Sample):**  $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
8. **Standard Deviation:**  $\sigma = \sqrt{\sigma^2}$
9. **Coefficient of Variation:**  $CV = \frac{\sigma}{\mu} \times 100$
10. **Quartiles:**  $Q_1, Q_2, Q_3$  (25%, 50%, 75%)

# Probability & Distributions

11. **Probability:**  $P(A) = \frac{\text{favorable outcomes}}{\text{total outcomes}}$

12. **Complement Rule:**  $P(A^c) = 1 - P(A)$

13. **Addition Rule:**  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

14. **Multiplication Rule:**  $P(A \cap B) = P(A)P(B|A)$

15. **Conditional Probability:**  $P(A|B) = \frac{P(A \cap B)}{P(B)}$

data.science.beginners

16. **Bayes' Theorem:**  $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

17. **Expected Value:**  $E[X] = \sum x_i P(x_i)$

18. **Binomial Distribution:**  $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

19. **Poisson Distribution:**  $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

20. **Normal PDF:**  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

# Interference Statistics

21. **Z-score:**  $Z = \frac{x - \mu}{\sigma}$

22. **t-score:**  $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$

23. **Confidence Interval (Mean, known  $\sigma$ ):**  
 $\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

24. **Confidence Interval (Mean, unknown  $\sigma$ ):**  
 $\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$

data.science.beginners

25. **Margin of Error:**  $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

26. **Standard Error:**  $SE = \frac{s}{\sqrt{n}}$

27. **Chi-Square Statistic:**  $\chi^2 = \sum \frac{(O - E)^2}{E}$

28. **F-Ratio:**  $F = \frac{s_1^2}{s_2^2}$

29. **ANOVA F-test:**  $F = \frac{MS_{between}}{MS_{within}}$

30. **Correlation (Pearson):**  
 $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$



# Regression & ML Stats

31. **Simple Linear Regression:**  $y = \beta_0 + \beta_1 x + \epsilon$

32. **Slope ( $\beta_1$ ):**  
$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

33. **Intercept ( $\beta_0$ ):**  $\beta_0 = \bar{y} - \beta_1 \bar{x}$

34. **R-squared:**  $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$

35. **Mean Squared Error (MSE):**  $MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$

36. **Root Mean Squared Error (RMSE):**  $RMSE = \sqrt{MSE}$

37. **Mean Absolute Error (MAE):**  $MAE = \frac{\sum |y_i - \hat{y}_i|}{n}$

38. **Log-Loss:**  
$$-\frac{1}{n} \sum [y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)]$$

39. **Odds:**  $Odds = \frac{p}{1-p}$

40. **Logit:**  $\log\left(\frac{p}{1-p}\right)$

# Advanced & Data Science Applications

41. **Entropy:**  $H(X) = - \sum p(x) \log p(x)$

42. **Gini Index:**  $G = 1 - \sum p_i^2$

43. **Information Gain:**  $IG = H(parent) - \sum \frac{n_i}{n} H(child_i)$

44. **KL Divergence:**  $D_{KL}(P||Q) = \sum P(x) \log \frac{P(x)}{Q(x)}$

45. **Covariance:**

$$\text{Cov}(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

46. **Standardization:**  $z = \frac{x - \bar{x}}{s}$

47. **Normalization (Min-Max):**

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

48. **Principal Component (PCA eigenvector):**

$$\text{Maximize } \frac{w^T S w}{w^T w}$$

49. **Mahalanobis Distance:**

$$D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

50. **Silhouette Score:**

$$s = \frac{b-a}{\max(a,b)}$$

(where  $a$  = intra-cluster distance,  $b$  = nearest-cluster distance)