STATISTICS ULTIMATE CHEAT SHEET

1. Descriptive Statistics

Measures of Central Tendency:

- **Sample Mean:** $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- **Population Mean:** $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$
- **Median:** Middle value of ordered dataset.
- **Mode:** Most frequently occurring value.

Measures of Dispersion:

- **Sample Variance:** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- **Population Variance:** $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$
- **Standard Deviation:** $s = \sqrt{s^2}$, $\sigma = \sqrt{\sigma^2}$
- **Interquartile Range:** $\text{IQR} = Q_3 - Q_1$
- **Coeff. of Variation:** $\text{CV} = \frac{s}{\bar{x}} \times 100\%$

2. Probability Theory

Axioms and Basic Rules:

- $0 \leq P(A) \leq 1, \quad P(\Omega) = 1$
- **Complement:** $P(A^c) = 1 - P(A)$
- **Addition Rule:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Conditional Prob.:** $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- **Independence:** A, B independent $\iff$ $P(A \cap B) = P(A)P(B)$

Bayes' Theorem:

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

3. Random Variables

Expectation and Variance:

- **Discrete:** $E[X] = \sum xP(X = x)$
- **Continuous:** $E[X] = \int_{-\infty}^{\infty} xf(x)dx$
- **Variance Property:** $\text{Var}(X) = E[X^2] - (E[X])^2$
- **Linear Transformations:**
  - $E[aX + b] = aE[X] + b$
  - $\text{Var}(aX + b) = a^2\text{Var}(X)$
  - $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$

4. Key Distributions

Discrete Distributions:

- **Binomial** $X \sim \text{Bin}(n, p)$: P(X=k) = n$\left(_{kp^k(1-p)^{n-k}}\right)$

  $E[X] = np, \quad \text{Var}(X) = np(1-p)$

- **Poisson** $X \sim \text{Pois}(\lambda)$: P(X=k) = $\lambda^k e^{-\lambda}\overline{k!}$

  $E[X] = \lambda, \quad \text{Var}(X) = \lambda$

Continuous Distributions:

- **Normal** $X \sim \mathcal{N}(\mu, \sigma^2)$: f(x) = 1$\frac{}{\sigma\sqrt{2\pi}\exp(-\frac{(x-\mu)^2}{2\sigma^2})}$

  Standardization: $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$

- **Exponential** $X \sim \text{Exp}(\lambda)$: f(x) = $\lambda e^{-\lambda x} \quad (x \geq 0)$

  $E[X] = 1/\lambda, \quad \text{Var}(X) = 1/\lambda^2$

5. Sampling Distributions

Central Limit Theorem (CLT): For i.i.d. $X_1, X_2, \ldots, X_n$ with mean μ and variance σ^2, as $n \to \infty$:

$$\bar{X} \approx \mathcal{N}(\mu, \frac{\sigma^2}{n}) \quad \Longrightarrow \quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Standard Error:

$$\text{SE}(\bar{X}) = \frac{s}{\sqrt{n}}$$

6. Hypothesis Testing

General Framework:

1. State H_0 (Null) and H_1 (Alternative).
2. Choose Significance Level α (e.g., 0.05).
3. Compute Test Statistic (Z, t, χ^2, F).
4. Determine p-value or Critical Region.
5. Reject H_0 if $p \leq \alpha$ or $|\text{Test Stat}| \geq$ Critical Value.

Error Types:

	H_0 **True**	H_0 **False**
Reject H_0	Type I Error (α)	Correct ($1-\beta$)
Fail to Reject	Correct ($1-\alpha$)	Type II Error (β)

7. Summary of Key Tests

1-Sample Z-Test (σ known):

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

1-Sample t-Test (σ unknown):

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \quad df = n - 1$$

2-Sample t-Test (Independent):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Chi-Square Goodness-of-Fit Test:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}, \quad df = k - 1$$

8. Regression & Correlation

Pearson Correlation Coefficient:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{\text{Cov}(X, Y)}{s_X s_Y}$$

Simple Linear Regression:

$$Y = \beta_0 + \beta_1 X + \varepsilon]$$

Slope: $\hat{\beta}_1 = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = r\frac{s_Y}{s_X}$

Intercept: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

Coeff. of Determination (R^2): Proportion of variance explained by model ($R^2 = r^2$).

9. Confidence Intervals

CI for Mean (σ known):

$$\bar{x} \pm Z_{\alpha/2}(\frac{\sigma}{\sqrt{n}})$$

CI for Mean (σ unknown):

$$\bar{x} \pm t_{\alpha/2, n-1}(\frac{s}{\sqrt{n}})$$

CI for Proportion:

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$