Cox Proportional Hazards Model

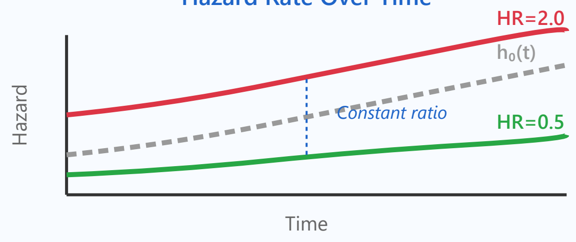
Gold standard for survival analysis in clinical research

Model Formula

$$h(t|X) = h_0(t) \cdot \exp(\beta_1 X_1 + \ldots + \beta_p X_p)$$

Hazard ratio = $\exp(\beta)$

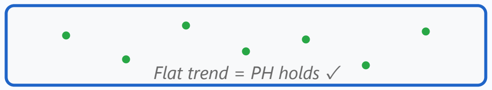
Hazard Rate Over Time



Key Assumptions

- **Proportional hazards:** HR constant over time
- **Linear relationship:** with log-hazard
- **Independent censoring:** uninformative

Testing PH Assumption



Time-varying Covariates

Handle changing predictors

Stratification

When PH assumption violated

Penalized Cox

High-dimensional data

🔍 Principles and Detailed Explanation of Cox Regression

📊 Semi-parametric Approach

The Cox model is a **semi-parametric** method that does not assume a specific form for the baseline hazard function $h_0(t)$.

⚖️ Partial Likelihood Estimation

The Cox model estimates β using **partial likelihood**.

At each event time, it compares "who has a higher probability of experiencing the event?"

Advantage: It can estimate the effect of covariates without needing to know the exact distribution of hazard over time.

All Forms of Baseline Hazard Allowed

$$L(\beta) = \prod_i \left[\exp(\beta X_i) / \sum_{j \in R_i} \exp(\beta X_j) \right]$$

R_i = risk set at time i

🔢 Cox Model Fitting Process

STEP 1: Data Preparation

- Verify survival time, event indicator, and covariates
- Identify censored observations (event=0)

STEP 2: Construct Risk Sets

- At each event time, create a set of all subjects who have not yet experienced the event
- Example: Event at t=5 → All subjects surviving up to time 5 form the risk set

STEP 3: Calculate Partial Likelihood

- Compute the relative hazard of the subject who actually experienced the event at each time point
- Baseline hazard $h_0(t)$ cancels out in calculations (only ratios matter)

STEP 4: Maximize and Estimate β

- Use Newton-Raphson algorithm to maximize partial likelihood
- Estimate regression coefficients β and standard errors

STEP 5: Assumption Testing and Model Diagnostics

- Test proportional hazards assumption (Schoenfeld residuals test)
- Check for influential observations (dfbeta, score residuals)

💡 Hazard Ratio (HR) Interpretation Guide

HR = exp(β) represents the change in hazard when a covariate increases by one unit.

HR = 2.0 2x increase in risk (100% ↑) e.g., Smoker vs Non-smoker	**HR = 1.0** No effect (0% change) Null hypothesis state	**HR = 0.5** 50% reduction in risk (halved ↓) e.g., Treatment effect

📋 Real Research Example: Cancer Patient Survival Analysis

Study Setting: 200 cancer patients, 5-year follow-up
Variables: Age, tumor size, treatment type
Results:
• Age: $\beta=0.03$, HR=1.03 (3% increase in risk per 1-year increase)
• Tumor size: $\beta=0.25$, HR=1.28 (28% increase in risk per 1cm increase)
• New treatment: $\beta=-0.51$, HR=0.60 (40% reduction in mortality risk vs. standard, $p<0.001$)

Interpretation: The new treatment significantly improves survival even after adjusting for age and tumor size.

✅ When to Use Cox Model

- Survival analysis with censored data
- When the exact distribution of hazard over time is unknown
- To quantify the effect of covariates on risk
- When the proportional hazards assumption is valid (or can be appropriately transformed)

⚠️ Cautions

- Interpretation requires caution when proportional hazards assumption is violated
- Estimation becomes unstable with small sample size or few events
- Time-dependent covariates cannot be handled by standard Cox model
- Competing risks require specialized models