Lecture 2: Survival model regression

Graeme L. Hickey

Department of Epidemiology and Population Health, Institute of Infection and Global Health, University of Liverpool, UK

27th November 2014

Spatial and Temporal Statistical Modelling for Population Health Sciences

Setup

For each subject $i = 1, \ldots, n$ we record m baseline measurements (can be continuous or categorial): $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})^T$

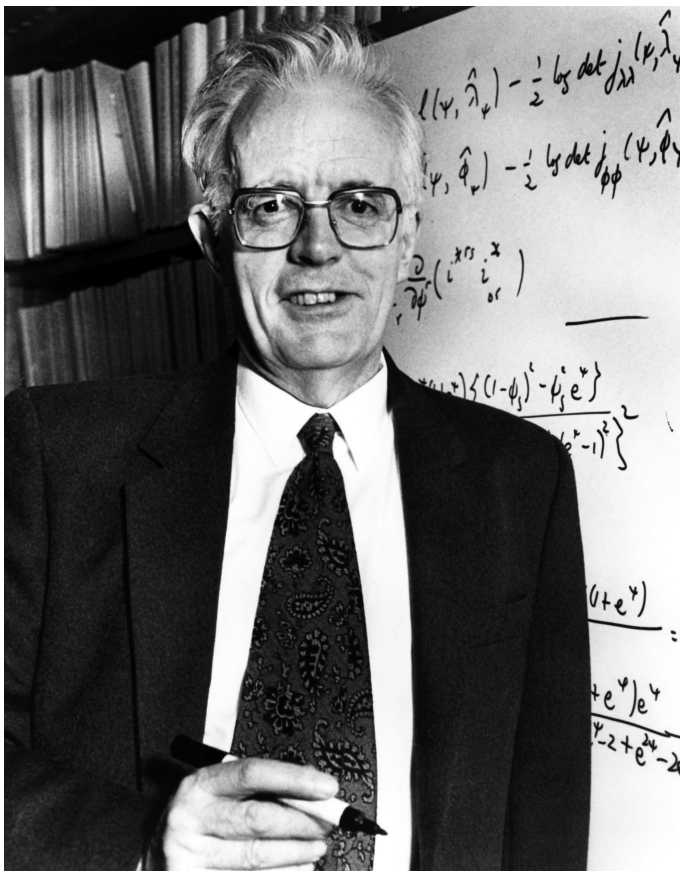
Goal

To quantify the effect of one or more explanatory variables on the failure time $\longrightarrow$ Regression

Approaches

1. Semi-parametric: Cox proportional hazards model
2. Parametric:
   - Proportional hazards models
   - Accelerated failure time models

Cox proportional hazards regression



Introduced by Sir David Cox in 1972

The Cox proportional hazards (PH) regression model is ubiquitous in the medical and epidemiology literature

Cited 38,000 times (Google Scholar)

The model

Definition

$$\begin{aligned} h(t, \mathbf{x}) &= \exp(\beta_0(t) + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m) \\ &= h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m) \end{aligned}$$

with $\beta_1, \beta_2, \ldots, \beta_m$ the model coefficients for estimation, and $h_0(t)$ the baseline hazard function (which might depend on t, but not the covariates)

We don't need to specify $h_0(t)$ — this is why we call it semi-parametric, and why it is so popular

Proportional hazards

Demonstration

Consider two subjects with measured covariates $\mathbf{x}_1 = (x_1^{(1)}, x_2, \ldots, x_m)$ and $\mathbf{x}_2 = (x_1^{(2)}, x_2, \ldots, x_m)$

$$\begin{aligned} \frac{h(t, \mathbf{x}_2)}{h(t, \mathbf{x}_1)} &= \frac{h_0(t)\exp(\beta_1 x_1^{(2)} + \beta_2 x_2 + \cdots + \beta_m x_m)}{h_0(t)\exp(\beta_1 x_1^{(1)} + \beta_2 x_2 + \cdots + \beta_m x_m)} \\ &= \exp\left[\beta_1 (x_1^{(2)} - x_1^{(1)})\right] \end{aligned}$$

In other words, $h(t, \mathbf{x}_2) \propto h(t, \mathbf{x}_1)$

Hazard ratio

We call the proportionality constant for a unit increase in measured covariate x_1, which equals $\exp(\beta_1)$, the hazard ratio

Proportional hazards

Demonstration

Consider two subjects with measured covariates $\mathbf{x}_1 = (x_1^{(1)}, x_2, \ldots, x_m)$ and $\mathbf{x}_2 = (x_1^{(2)}, x_2, \ldots, x_m)$

$$\begin{aligned} \frac{h(t, \mathbf{x}_2)}{h(t, \mathbf{x}_1)} &= \frac{h_0(t)\exp(\beta_1 x_1^{(2)} + \beta_2 x_2 + \cdots + \beta_m x_m)}{h_0(t)\exp(\beta_1 x_1^{(1)} + \beta_2 x_2 + \cdots + \beta_m x_m)} \\ &= \exp\left[\beta_1 (x_1^{(2)} - x_1^{(1)})\right] \end{aligned}$$

In other words, $h(t, \mathbf{x}_2) \propto h(t, \mathbf{x}_1)$

Hazard ratio

We call the proportionality constant for a unit increase in measured covariate x_1, which equals $\exp(\beta_1)$, the hazard ratio

Hazard ratios

We can estimate multiple hazard ratios, as per ordinary regression, and interpret them as:

- $\beta_j = 1 \Rightarrow$ does not influence time to failure
- $\beta_j > 1 \Rightarrow$ risk of failure is $\exp(\beta_j)$ more likely per unit increase
- $\beta_j < 1 \Rightarrow$ risk of failure is $\exp(\beta_j)$ less likely per unit increase

Fitting Cox PH models

Could use standard maximum likelihood estimation techniques **if** we specified the baseline hazard function $h_0(t)$

Partial likelihood function

Let $R(t)$ be the set of all subjects alive and uncensored at time t (called the risk set). If a failure occurs at time $t^\star$

$$\begin{aligned} P(\text{subject } i \text{ dies}) &= \frac{h(t^\star, \mathbf{x}_i)}{\sum_{k \in R(t^\star)} h(t^\star, \mathbf{x}_k)} \\ &= \frac{\exp(\beta^T \mathbf{x}_i)}{\sum_{k \in R(t^\star)} \exp(\beta^T \mathbf{x}_k)} \end{aligned}$$

We multiply these probabilities over all patients who have an observed failure time to yield the partial likelihood function, which we maximise using standard numerical methods.

Fitting Cox PH models

Could use standard maximum likelihood estimation techniques **if** we specified the baseline hazard function $h_0(t)$

Partial likelihood function

Let $R(t)$ be the set of all subjects alive and uncensored at time t (called the risk set). If a failure occurs at time $t^\star$

$$\begin{aligned} P(\text{subject } i \text{ dies}) &= \frac{h(t^\star, \mathbf{x}_i)}{\sum_{k \in R(t^\star)} h(t^\star, \mathbf{x}_k)} \\ &= \frac{\exp(\beta^T \mathbf{x}_i)}{\sum_{k \in R(t^\star)} \exp(\beta^T \mathbf{x}_k)} \end{aligned}$$

We multiply these probabilities over all patients who have an observed failure time to yield the partial likelihood function, which we maximise using standard numerical methods.

Example

Recall the dialysis data from earlier, with method = 1 (APD), = 0 (CAPD)

$$h(t) = h_0(t) \exp\{\beta_1 \text{method} + \beta_2 \text{age}\}$$

	HR	95% CI	Wald's Z	P
method	0.37	(0.22, 0.62)	-3.75	< 0.001
age	1.06	(1.04, 1.08)	6.76	< 0.001

Interpretation: an increase in age of 1-year increases the hazard of instantaneous death by 6%, and APD is strongly associated with longer survival after adjustment for age

Example

Recall the dialysis data from earlier, with method = 1 (APD), = 0 (CAPD)

$$h(t) = h_0(t) \exp\{\beta_1 \text{method} + \beta_2 \text{age}\}$$

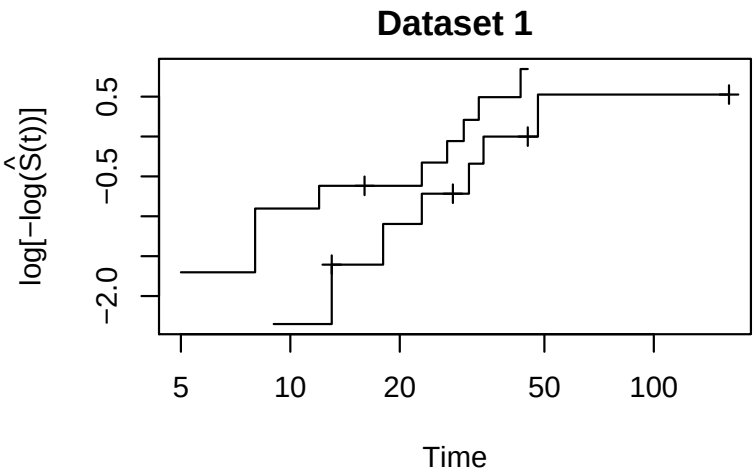
	HR	95% CI	Wald's Z	P
method	0.37	(0.22, 0.62)	-3.75	< 0.001
age	1.06	(1.04, 1.08)	6.76	< 0.001

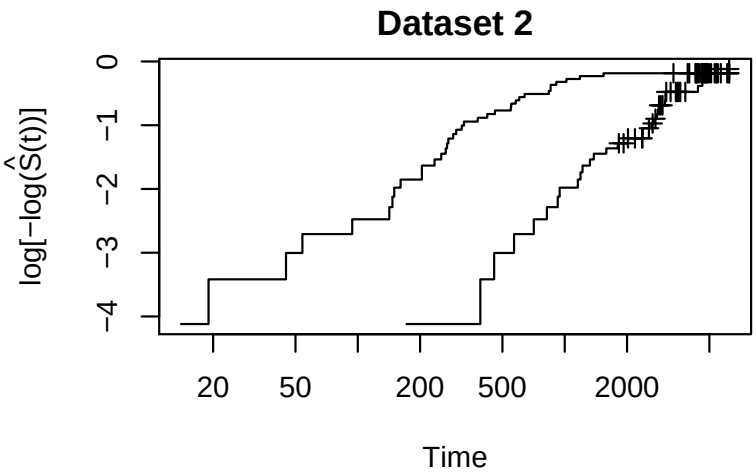
Interpretation: an increase in age of 1-year increases the hazard of instantaneous death by 6%, and APD is strongly associated with longer survival after adjustment for age

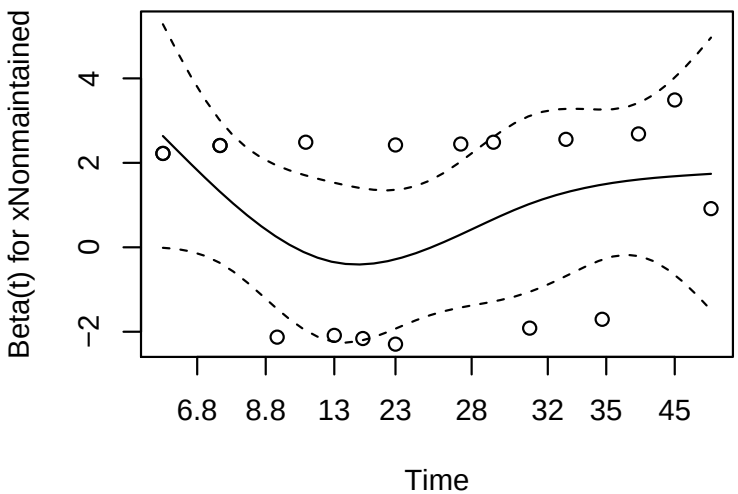
Is the PH assumption met?

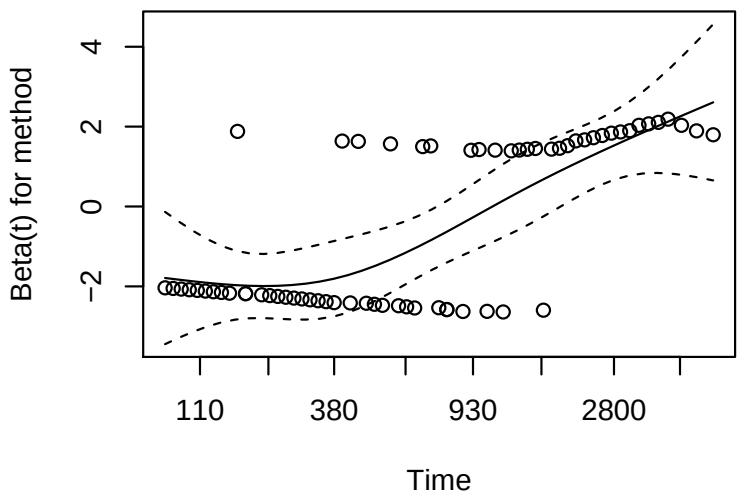
1. Simple check: if Kaplan-Meier curves cross the PH violated (oneway argument: non-crossing curves does not ensure PH satisfied)
2. Plot $\log\left[-\log(\hat{S}(t))\right]$ against $\log(t)$ and check they are parallel
3. Plot smoothed Schoenfeld residuals against predictor and check slope $= 0$
4. Grambsch-Therneau test[1]

[1] See Practical 2 sheet


Dataset 1
log[−log(Ŝ(t))]
−2.0
−0.5
0.5
5
10
20
50
100
Time


Dataset 2
log[−log(Ŝ(t))]
−4
−3
−2
−1
0
20
50
200
500
2000
Time


Beta(t) for xNonmaintained
−2
0
2
4
6.8
8.8
13
23
28
32
35
45
Time


Beta(t) for method
−2
0
2
4
110
380
930
2800
Time

How to overcome violations?

- Stratified baseline hazards

$$h(t, \mathbf{x}_i, \mathcal{S} = s_i) = h_0^{s_i}(t) \exp(\beta^T \mathbf{x}_i)$$

Example: $\mathcal{S} = \{\text{male}, \text{female}\}$

- Extended (piecewise) Cox model — induces a time-dependent covariate
- Parametric model that does not depend on PH assumption

Parametric models

Majority of models applied in practice can be framed as either:

1. A proportional hazards model
2. An accelerated failure time model

Proportional hazards models

PH model

$$h(t, \mathbf{x}) = h_0(t) \exp(\beta^T \mathbf{x})$$

where

- $h_0(t)$ is the baseline hazard function defined by the parametric family independent of the covariates
- $\exp(\beta_k)$ denotes the hazard ratio for a unit increase in variable x_k

Weibull model

Recall from this morning that $h(t) = \lambda p t^{p-1}$

We can re-parameterize λ as

$$\lambda = \exp(\beta_0 + \beta^T \mathbf{x}_i) = \lambda_0 \exp(\beta^T \mathbf{x}_i)$$

If $\mathbf{x}_1 = 0$ and $\mathbf{x}_2 = 1$ (i.e. single covariate), then

$$\text{HR} = \frac{h_2(t, \mathbf{x}_2)}{h_1(t, \mathbf{x}_1)} = \frac{\lambda_0 \exp(\beta_1) p t^{p-1}}{\lambda_0 \exp(0) p t^{p-1}} = \exp(\beta_1)$$

clearly indicating that the PH assumption is satisfied (so long as p is **independent** of $\mathbf{x}$)

Weibull model

Recall from this morning that $h(t) = \lambda p t^{p-1}$

We can re-parameterize λ as

$$\lambda = \exp(\beta_0 + \beta^T \mathbf{x}_i) = \lambda_0 \exp(\beta^T \mathbf{x}_i)$$

If $\mathbf{x}_1 = 0$ and $\mathbf{x}_2 = 1$ (i.e. single covariate), then

$$\text{HR} = \frac{h_2(t, \mathbf{x}_2)}{h_1(t, \mathbf{x}_1)} = \frac{\lambda_0 \exp(\beta_1) p t^{p-1}}{\lambda_0 \exp(0) p t^{p-1}} = \exp(\beta_1)$$

clearly indicating that the PH assumption is satisfied (so long as p is **independent** of $\mathbf{x}$)

Weibull model

Recall from this morning that $h(t) = \lambda p t^{p-1}$

We can re-parameterize λ as

$$\lambda = \exp(\beta_0 + \beta^T \mathbf{x}_i) = \lambda_0 \exp(\beta^T \mathbf{x}_i)$$

If $\mathbf{x}_1 = 0$ and $\mathbf{x}_2 = 1$ (i.e. single covariate), then

$$\text{HR} = \frac{h_2(t, \mathbf{x}_2)}{h_1(t, \mathbf{x}_1)} = \frac{\lambda_0 \exp(\beta_1) p t^{p-1}}{\lambda_0 \exp(0) p t^{p-1}} = \exp(\beta_1)$$

clearly indicating that the PH assumption is satisfied (so long as p is **independent** of $\mathbf{x}$)

Accelerated failure time models

AFT model

$$S(t) = S_0(\gamma t)$$

where

- $S_0(t)$ is the baseline survival function
- $\gamma = \exp(-\alpha^T \mathbf{x})$ is the acceleration factor, which can depend on covariates $\mathbf{x}$ but not on time t

Acceleration factor

Properties

- If $\gamma > 1$ time to event accelerates relative to the baseline
- If $\gamma = 1$ time to event is 'normal' relative to the baseline
- If $\gamma < 1$ time to event decelerates relative to the baseline

Example

If we measured time to death of humans and dogs, and recorded species as single covariate:

$$S_{\text{dog}}(t) = S_{\text{human}}(7t)$$

In words: time accelerates for dogs by a factor of 7 relative to humans

Acceleration factor

Properties

- If $\gamma > 1$ time to event accelerates relative to the baseline
- If $\gamma = 1$ time to event is 'normal' relative to the baseline
- If $\gamma < 1$ time to event decelerates relative to the baseline

Example

If we measured time to death of humans and dogs, and recorded species as single covariate:

$$S_{\text{dog}}(t) = S_{\text{human}}(7t)$$

In words: time accelerates for dogs by a factor of 7 relative to humans

Accelerated failure time models

Log-linear form of the AFT model

$$\log(T_i) = \mu + \alpha^T \mathbf{x}_i + \sigma W_i$$

where

- T_i and $\mathbf{x}_i$ are defined as per before for subject i
- $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ are the parameters for estimation
- μ is a constant intercept term that we also estimate
- σ is a scale coefficient
- W_i is a random error term that has a certain (specified) distribution

Weibull AFT model

If we let:

- $\sigma = 1/p$
- $\lambda = \exp(-\mu/\sigma)$
- $W_i \sim$ standard extreme value distribution

then it can shown that T_i follows a Weibull distribution (as defined in lecture 1) and therefore it satisfies the AFT property

Weibull AFT model

The Weibull regression model is the **only** one that can be written as both PH and AFT models

Weibull PH model

$$\log(\lambda) = \beta_0 + \beta^T \mathbf{x}$$

Weibull AFT model

$$\log(\lambda) = -p\left(\alpha_0 + \alpha^T \mathbf{x}\right)$$

Correspondence between hazard ratios and acceleration factors: $\beta_j = -p\alpha_j$, but note the different interpretations

Weibull AFT model

The Weibull regression model is the **only** one that can be written as both PH and AFT models

Weibull PH model

$$\log(\lambda) = \beta_0 + \beta^T \mathbf{x}$$

Weibull AFT model

$$\log(\lambda) = -p\left(\alpha_0 + \alpha^T \mathbf{x}\right)$$

Correspondence between hazard ratios and acceleration factors: $\beta_j = -p\alpha_j$, but note the different interpretations

Example

Back to the dialysis data from earlier...

$$log(T) = \alpha_0 + \alpha_1 \text{method} + \alpha_2 \text{age} + \frac{1}{p} W$$

	$\hat{\alpha}$	95% CI	Wald's Z	P
intercept	12.11	(10.74, 13.48)	17.32	< 0.001
method	1.32	(0.66, 1.98)	3.93	< 0.001
age	-0.08	$(-0.10, -0.05)$	-6.73	< 0.001

Interpretation: the survival time for APD (method $= 1$) patients is increased by a factor of $e^{1.32} = 3.74$ compared to the CAPD group (method $= 0$)

Example

Back to the dialysis data from earlier...

$$log(T) = \alpha_0 + \alpha_1 \text{method} + \alpha_2 \text{age} + \frac{1}{p} W$$

	$\hat{\alpha}$	95% CI	Wald's Z	P
intercept	12.11	(10.74, 13.48)	17.32	< 0.001
method	1.32	(0.66, 1.98)	3.93	< 0.001
age	-0.08	$(-0.10, -0.05)$	-6.73	< 0.001

Interpretation: the survival time for APD (method $= 1$) patients is increased by a factor of $e^{1.32} = 3.74$ compared to the CAPD group (method $= 0$)

Frailty

Just another name for random effects:

- $\lambda_i(t) = \lambda_0(t)\exp(\mathbf{x}_i\beta + U_i)$
- $U_i \sim N(0, \tau^2)$

Interpretation: subjects have different hazards, even after adjusting for available explanatory variables

Rationale: over-dispersion; clustering (e.g. patients within hospitals for a large multi-centre trial)

General regression modelling principles

In addition to satisfying the core model assumptions (e.g. PH in Cox regression), general regression model development thoughts should include

- **Variable selection**
  Example: Pre-screening; stepwise regression; EPV ≥ 10?
- **Interaction terms**
- **Functional form**
  Example: Martingale residual plots for functional form and linearity assessment; splines; categorisation/dichotomisation
- **Correct modelling distribution**
  Example: Likelihood ratio test (nested); AIC (in general)
- **Missing data**
  Example: Multiple imputation; case-deletion

Situations that require more complex methods

- **Ties**
  Survival models assume continuous time, so ties are not possible. However if we round to nearest day etc. they are inevitable. In Cox PH model: use exact method, Efron or Breslow approximations.
- **Time-dependent variables**
  Some variables can change over time, e.g. blood pressure. Utilising the longitudinal data can improve the model. It might even be sensible to jointly model the survival and longitudinal data.

Suggested reading

- Therneau TM, Grambsch PM (2001). Modeling Survival Data: Extending the Cox Model. New York: *Springer-Verlag*.
  ✎ Useful text book with R code
- Cox DR (1972). Regression models and life tables, *J R Stat Soc B*, 34:187-220.
  ✎ The seminal Cox PH model manuscript
- Grambsch PM, Therneau TM (1994), Proportional hazards tests and diagnostics based on weighted residuals. Biometrika, 81, 515-26.
  ✎ Methods for assessment of PH assumption using Schoenfeld residuals
- Therneau TM, et al. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1), 147-160.
  ✎ Residual based method for estimating the functional form of a continuous covariate
- Henderson R, et al. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465-80.
  ✎ Formulates a class of models that jointly models survival and longitudinal data