# Survival analysis:
## Parametric regression

**Mark Stevenson**

Faculty of Veterinary and Agricultural Sciences

The University of Melbourne, Parkville Victoria 3010 Australia

mark.stevenson1@unimelb.edu.au

# Roadmap

- Background
- Exponential model
- Weibull model
- Accelerated failure time models
- Choosing a parametric model

## Background

- Semi-parametric models make no assumption about the distribution of failure times, but do make assumptions about how covariates change the survival experience

- Parametric models, on the other hand, make assumptions about the distribution of failure times and the relationship between covariates and survival experience
  - fully specify the distribution of the baseline hazard/survival function according to some (defined) probability distribution

## Background

- Semi-parametric models only use the order of failure times in an analysis
  - 1, 3, 4, 6, 89 and 1, 63, 65, 88, 89 are equivalent
- Parametric models uses exact times of failures
  - better for prediction, more efficient use of data
  - possible to generate predictions for out of sample values
  - computationally easier; easier to extend (e.g. random effects)

# Roadmap

- Background
- <u>Exponential model</u>
- Weibull model
- Accelerated failure time models
- Choosing a parametric model

# Exponential model

- Simplest type of parametric model in that it assumes that the baseline hazard is constant over time:

$$h(t) = h_0 \exp^{\beta_1 x_{1i} + \dots + \beta_m x_{mi}}$$
$$h_0 = \lambda$$

10,000 changing-color light globes at the Circo Massimo in Rome.

# Exponential model

- Possible to split the time at risk into short periods of constant (‘piece-wise exponential model’)

- Hazard allowed to vary among time intervals

# Roadmap

- Background
- Exponential model
- <u>Weibull model</u>
- Accelerated failure time models
- Choosing a parametric model

# Weibull model

- Assumed that the baseline hazard has a shape which gives rise to a Weibull distribution of survival times
- The baseline hazard is:

$$h(t) = h_0(t)\, \exp^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_m x_{mi}}$$
$$h_0 = \lambda p\, t^{\mathrm{p}-1}$$

Includes an intercept term $\beta_0$

$\lambda$ = scale parameter

$p$ = shape parameter

THE UNIVERSITY OF MELBOURNE

## Weibull model

- The shape parameter

  $p > 1$ hazard is increasing

  $p = 1$ hazard is constant

  $p < 1$ hazard is decreasing

# Roadmap

- Background
- Exponential model
- Weibull model
- <u>Accelerated failure time models</u>
- Choosing a parametric model

# Accelerated failure time models

- All parametric models are accelerated failure time models

- General form of an accelerated failure time model is:

  $\ln(t) = \beta X + \varepsilon$

  $\varepsilon = \ln(\tau)$, an error term on the $\ln(t)$ scale

  $\tau$ = an error term on the original time scale

- If $\varepsilon \sim N$ an AFT model is equivalent to ordinary linear regression on $\ln(t)$ scale except that the data are censored

## Accelerated failure time models

- $\tau$ is the distribution of survival times when $\beta X$ = 0, that is baseline survival (not hazard)

- We specify the distributional form of $\tau$ (e.g. gamma, log-normal, log-logistic, Weibull)

## Accelerated failure time models

- Predictors act additively on the $\ln(t)$ scale and multiplicatively on the original $(t)$ scale

- If the $\beta s$ from an AFT are positive, failure times are longer than average (time passes more slowly)

- If the $\beta s$ from an AFT are negative, failure times are shorter than average (time passes more quickly)

# Accelerated failure time models

- Accelerated failure time coefficients represent the expected change in $\ln(t)$ for a one unit change in the predictor

- This is useful, since we talk about the effect of covariates on survival time (rather than on hazard)

- Consider a Weibull accelerated failure time model fitted to the `addict` data …

```
library(survival); setwd("D:\\TEMP");
dat <- read.table("addict.csv", header = TRUE, sep = ",");
dat$clinic <- factor(dat$clinic, levels = c(1,2), labels = c("1","2"))
dat$prison <- factor(dat$prison, levels = c(0,1), labels = c("0","1"))
```

```
library(rms)
addict.wei <- psm(Surv(stop, status) ~ clinic + prison + dose, dist =
"weibull", data = dat)

Obs       Events Model L.R.         d.f.            P          R2
238          150       60.73           3            0        0.23

              Value Std. Error     z        p
(Intercept)  4.7915    0.27825 17.22 1.87e-66
clinic=2     0.7198    0.15951  4.51 6.39e-06
prison=1    -0.2232    0.12245 -1.82 6.84e-02
dose         0.0247    0.00464  5.31 1.08e-07
Log(scale)  -0.3006    0.06759 -4.45 8.70e-06

Scale= 0.74
```

| Variable | Subjects | Failed | Coefficient (SE) | P | Survival (95% CI) |
|---|---|---|---|---|---|
| Intercept | 238 | 150 | 4.7915 (0.2782) | < 0.01 | |
| Clinic: | | | | < 0.01[a] | |
| Clinic 1 | 163 | 122 | - | | 1.00 |
| Clinic 2 | 74 | 28 | 0.7198 (0.1595) | | 2.05 (1.50 -- 2.81) [b] |
| Prison: | | | | 0.07 | |
| Absent | 127 | 81 | - | | 1.00 |
| Present | 111 | 69 | -0.2232 (0.1224) | | 0.80 (0.63 -- 1.02) |
| Dose | 238 | 150 | 0.0247 (0.0046) | < 0.01 | 1.02 (1.01 -- 1.03) |

[a] Significance of the two clinic variables in the model.
[b] Interpretation: after adjusting for the effect of methadone dose and prison status retention time for patients from Clinic 2 was 2.05 times that of patients from Clinic 1 (95% CI 1.50 -- 2.81).

| Variable | Subjects | Failed | Coefficient (SE) | P | Hazard ratio (95%) |
|---|---|---|---|---|---|
| Clinic: | | | | < 0.01[a] | |
| Clinic 1 | 163 | 122 | - | | 1.00 |
| Clinic 2 | 74 | 28 | -1.0091 (0.2147) | | 0.36 (0.24 - 0.55) [b] |
| Prison: | | | | 0.06 | |
| Absent | 127 | 81 | - | | 1.00 |
| Present | 111 | 69 | 0.3146 (0.1672) | | 1.37 (0.98 - 1.90) |
| Dose | 238 | 150 | - 0.0352 (0.0064) | < 0.01 | 0.96 (0.95 - 0.98) |

[a] Significance of the two clinic variables in the model.
[b] Interpretation: compared with the reference category (patients from Clinic 1), after adjusting for the effect of methadone dose and prison status, patients from Clinic 2 had 0.36 (95% CI 0.24 - 0.55) times the daily hazard of relapse.

# Accelerated failure time models

- What is the effect of Clinic 2 on retention time?

  $\ln(t) = 4.7915 + (0.7198 \times 1)$
  $\ln(t) = 5.5113$
  $t = \exp^{5.5113}$
  $t = 247$ days

  - patients from Clinic 2 survived, on average, $\exp^{5.5113} = 247$ days
  - patients from Clinic 1 survived, on average, $\exp^{4.7915} = 120$ days

## Accelerated failure time models

- Exponential and Weibull models can be parameterised as either proportional hazards or accelerated failure time models

- Other parametric models (e.g. the log-normal, the log-logistic, and gamma) can only be expressed as accelerated failure time

# Roadmap

- Background
- Exponential model
- Weibull model
- Accelerated failure time models
- <u>Choosing a parametric model</u>

# Choosing a parametric model

- Choice depends on
  - biological knowledge of outcome, knowledge of expected shape of baseline hazard
  - nested models: use likelihood ratio test
  - non-nested models: use AIC

- Hybrid = piecewise exponential?

# Roadmap

- Background
- Exponential model
- Weibull model
- Accelerated failure time models
- Choosing a parametric model