Two estimators of survival

$$S_{\mathrm{acturial}}(t) = \prod_{t_i : t_i \le t} \left(1 - \frac{d_i}{l_i - w_i/2}\right)$$

where i is an index for the intervals before t, d_i are the deaths and w_i are the censored values.

$$S_{\mathrm{KM}}(t) = \prod_{t_i : t_i \le t} \left(1 - \frac{d_i}{l_i}\right)$$

where i is now an index over the distinct event times.

Calculation of the number at risk

Note that we calculate the number at risk using:

$$l_i = l_{i-1} - d_{i-1} - w_{i-1}$$

In words, the new number at risk is the equal to the old number at risk less the number of events less the number censored.

Log-exponential relationships

log (in Stata: ln) and exp are inverse functions for each other

$$\log(\exp(x)) = \exp(\log(x)) = x$$

Log-exponential relationships

log (in Stata: ln) and exp are inverse functions for each other

$$\log(\exp(x)) = \exp(\log(x)) = x$$

log takes values $(0, \infty)$; $\log(0) = -\infty$; $\log(1) = 0$; $\log(\infty) = \infty$
exp takes values $(-\infty, \infty)$; $\exp(-\infty) = 0$; $\exp(0) = 1$; $\exp(\infty) = \infty$

Log-exponential relationships

log (in Stata: ln) and exp are inverse functions for each other

$$\log(\exp(x)) = \exp(\log(x)) = x$$

log takes values $(0, \infty)$; $\log(0) = -\infty$; $\log(1) = 0$; $\log(\infty) = \infty$
exp takes values $(-\infty, \infty)$; $\exp(-\infty) = 0$; $\exp(0) = 1$; $\exp(\infty) = \infty$
log of a product is the sum of the log values

$$\log(xy) = \log(x) + \log(y)$$

Log-exponential relationships

log (in Stata: ln) and exp are inverse functions for each other

$$\log(\exp(x)) = \exp(\log(x)) = x$$

log takes values $(0, \infty)$; $\log(0) = -\infty$; $\log(1) = 0$; $\log(\infty) = \infty$
exp takes values $(-\infty, \infty)$; $\exp(-\infty) = 0$; $\exp(0) = 1$; $\exp(\infty) = \infty$
log of a product is the sum of the log values

$$\log(xy) = \log(x) + \log(y)$$

A product of exponentials is the exponential of the sum

$$\exp(x)\exp(y) = \exp(x + y)$$

Log-exponential relationships

log (in Stata: ln) and exp are inverse functions for each other

$$\log(\exp(x)) = \exp(\log(x)) = x$$

log takes values $(0, \infty)$; $\log(0) = -\infty$; $\log(1) = 0$; $\log(\infty) = \infty$
exp takes values $(-\infty, \infty)$; $\exp(-\infty) = 0$; $\exp(0) = 1$; $\exp(\infty) = \infty$
log of a product is the sum of the log values

$$\log(xy) = \log(x) + \log(y)$$

A product of exponentials is the exponential of the sum

$$\exp(x)\exp(y) = \exp(x + y)$$

log of a power is the power term times log of the base

$$\log(x^y) = y\log(x)$$

Log-exponential relationships

log (in Stata: ln) and exp are inverse functions for each other

$$\log(\exp(x)) = \exp(\log(x)) = x$$

log takes values $(0, \infty)$; $\log(0) = -\infty$; $\log(1) = 0$; $\log(\infty) = \infty$
exp takes values $(-\infty, \infty)$; $\exp(-\infty) = 0$; $\exp(0) = 1$; $\exp(\infty) = \infty$
log of a product is the sum of the log values

$$\log(xy) = \log(x) + \log(y)$$

A product of exponentials is the exponential of the sum

$$\exp(x)\exp(y) = \exp(x + y)$$

log of a power is the power term times log of the base

$$\log(x^y) = y\log(x)$$

exponential raised to a power is the exponential of the product

$$\exp(x)^y = \exp(xy)$$

Risks and rates in an interval

For an interval i of width Δ_i,

$$\begin{aligned} \text{risk}_{\text{act}} &= \frac{d_i}{l_i - c_i/2} \\ \text{risk}_{\text{KM}} &= \frac{d_i}{l_i} \\ \text{rate} &= \frac{d_i}{\Delta_i(l_i - d_i/2 - c_i/2)} \end{aligned}$$

Risks and rates in an interval

For an interval i of width Δ_i,

$$\mathrm{risk}_{\mathrm{act}} = \frac{d_i}{l_i - c_i/2}$$

$$\mathrm{risk}_{\mathrm{KM}} = \frac{d_i}{l_i}$$

$$\mathrm{rate} = \frac{d_i}{\Delta_i(l_i - d_i/2 - c_i/2)}$$

The risk can be calculated from the rate:

$$\begin{aligned}\mathrm{risk} = 1 - S(t) = 1 - \exp(-\Lambda(t)) &= 1 - \exp(-\mathrm{rate}\Delta_i) \\ &= 1 - \exp\left(-\frac{d_i}{l_i - d_i/2 - c_i/2}\right) \\ &\approx \frac{d_i}{l_i - d_i/2 - c_i/2}\end{aligned}$$

where $\Lambda(t) = \int_0^t \lambda(s)ds$ (that is, the area under the hazard curve), which we call the cumulative hazard.

Interesting questions

- When should one use the actuarial estimator (or a discrete time formulation) compared with a continuous time formulation?

Interesting questions

- When should one use the actuarial estimator (or a discrete time formulation) compared with a continuous time formulation? Good question! The actuarial estimator is preferred for many ties and Cox regression is most common for few ties. A related question is: when should we use a discrete time survival regression model?

Interesting questions

- When should one use the actuarial estimator (or a discrete time formulation) compared with a continuous time formulation? Good question! The actuarial estimator is preferred for many ties and Cox regression is most common for few ties. A related question is: when should we use a discrete time survival regression model?
- Are we going to consider recurrent events?

Interesting questions

- When should one use the actuarial estimator (or a discrete time formulation) compared with a continuous time formulation? Good question! The actuarial estimator is preferred for many ties and Cox regression is most common for few ties. A related question is: when should we use a discrete time survival regression model?
- Are we going to consider recurrent events? No. These require special methods that take account of whether a subsequent event affects the hazard of a subsequent event, and whether we should model for a frailty/random effect.

Interesting questions

- When should one use the actuarial estimator (or a discrete time formulation) compared with a continuous time formulation? Good question! The actuarial estimator is preferred for many ties and Cox regression is most common for few ties. A related question is: when should we use a discrete time survival regression model?
- Are we going to consider recurrent events? No. These require special methods that take account of whether a subsequent event affects the hazard of a subsequent event, and whether we should model for a frailty/random effect.
- Are ties a problem in Cox regression?

Interesting questions

- When should one use the actuarial estimator (or a discrete time formulation) compared with a continuous time formulation? Good question! The actuarial estimator is preferred for many ties and Cox regression is most common for few ties. A related question is: when should we use a discrete time survival regression model?
- Are we going to consider recurrent events? No. These require special methods that take account of whether a subsequent event affects the hazard of a subsequent event, and whether we should model for a frailty/random effect.
- Are ties a problem in Cox regression? Yes. We will cover this on Friday.

Interesting questions

- When should one use the actuarial estimator (or a discrete time formulation) compared with a continuous time formulation? Good question! The actuarial estimator is preferred for many ties and Cox regression is most common for few ties. A related question is: when should we use a discrete time survival regression model?
- Are we going to consider recurrent events? No. These require special methods that take account of whether a subsequent event affects the hazard of a subsequent event, and whether we should model for a frailty/random effect.
- Are ties a problem in Cox regression? Yes. We will cover this on Friday.
- What is the effect of informative censoring on a survival analysis? [Not currently answered]

Main effects model

$$\log(\lambda) = \beta_0 + \beta_1 \mathrm{I}(\text{hieng} = 1) + \beta_2 \mathrm{I}(\text{job} = 2) + \beta_3 \mathrm{I}(\text{job} = 3)$$

Main effects model

$$\log(\lambda) = \beta_0 + \beta_1 \mathrm{I}(\mathrm{hieng} = 1) + \beta_2 \mathrm{I}(\mathrm{job} = 2) + \beta_3 \mathrm{I}(\mathrm{job} = 3)$$

$$\begin{aligned}
\lambda_{\mathrm{driver,low}} &= \exp(\beta_0 + \beta_1 \cdot 0 + \beta_2 \cdot 0 + \beta_3 \cdot 0) = \exp(\beta_0) \\
\lambda_{\mathrm{driver,high}} &= \exp(\beta_0 + \beta_1) = \exp(\beta_0)\exp(\beta_1) \\
\lambda_{\mathrm{cond,low}} &= \exp(\beta_0 + \beta_2) = \exp(\beta_0)\exp(\beta_2) \\
\lambda_{\mathrm{cond,high}} &= \exp(\beta_0 + \beta_1 + \beta_2) = \exp(\beta_0)\exp(\beta_1)\exp(\beta_2) \\
\lambda_{\mathrm{bank,low}} &= \exp(\beta_0 + \beta_3) = \exp(\beta_0)\exp(\beta_3) \\
\lambda_{\mathrm{bank,high}} &= \exp(\beta_0 + \beta_1 + \beta_3) = \exp(\beta_0)\exp(\beta_1)\exp(\beta_3)
\end{aligned}$$

Main effects model: Rate ratios

$$
\begin{aligned}
\mathrm{RR}_{\text{driver,high vs low}} &= \frac{\exp(\beta_0)\exp(\beta_1)}{\exp(\beta_0)} = \exp(\beta_1) \\
\mathrm{RR}_{\text{cond,high vs low}} &= \frac{\exp(\beta_0)\exp(\beta_1)\exp(\beta_2)}{\exp(\beta_0)\exp(\beta_2)} = \exp(\beta_1) \\
\mathrm{RR}_{\text{bank,high vs low}} &= \frac{\exp(\beta_0)\exp(\beta_1)\exp(\beta_3)}{\exp(\beta_0)\exp(\beta_3)} = \exp(\beta_1) \\
\mathrm{RR}_{\text{cond vs bank,low}} &= \frac{\exp(\beta_0)\exp(\beta_2)}{\exp(\beta_0)} = \exp(\beta_2) \\
\mathrm{RR}_{\text{cond vs bank,high}} &= \frac{\exp(\beta_0)\exp(\beta_1)\exp(\beta_2)}{\exp(\beta_0)\exp(\beta_1)} = \exp(\beta_2)\mathrm{RR}_{\text{bank,high vs driver,low}} =
\end{aligned}
$$

Interaction models

$$\log(\lambda) = \beta_0 + \beta_1 \mathrm{I}(\text{hieng} = 1) + \beta_2 \mathrm{I}(\text{job} = 2) + \beta_3 \mathrm{I}(\text{job} = 3) + \beta_4 \mathrm{I}(\text{hieng} = 1 \;\&\; \text{job} = 2) + \beta_5 \mathrm{I}(\text{hieng} = 1 \;\&\; \text{job} = 3)$$

Interaction models

$$\log(\lambda) = \beta_0 + \beta_1 \mathrm{I}(\text{hieng} = 1) + \beta_2 \mathrm{I}(\text{job} = 2) + \beta_3 \mathrm{I}(\text{job} = 3) + \\ \beta_4 \mathrm{I}(\text{hieng} = 1 \ \& \ \text{job} = 2) + \beta_5 \mathrm{I}(\text{hieng} = 1 \ \& \ \text{job} = 3)$$

or. . .

$$\log(\lambda) = \beta_1 \mathrm{I}(\text{job} = 1) + \beta_2 \mathrm{I}(\text{job} = 2) + \beta_3 \mathrm{I}(\text{job} = 3) + \\ \beta_4 \mathrm{I}(\text{hieng} = 1 \ \& \ \text{job} = 1) + \beta_5 \mathrm{I}(\text{hieng} = 1 \ \& \ \text{job} = 2) + \\ \beta_6 \mathrm{I}(\text{hieng} = 1 \ \& \ \text{job} = 3)$$

Regression equation

$$\log(\lambda) = \beta_0 + \beta_1 \mathrm{I}(\mathrm{fup} = 1) + \beta_2 \mathrm{I}(\mathrm{fup} = 2) + \\ \beta_3 \mathrm{I}(\mathrm{fup} = 3) + \ldots + \beta_9 \mathrm{I}(\mathrm{fup} = 9) + \beta_{10} \mathrm{I}(\mathsf{year8594} = 1)$$

Regression equation

$$\log(\lambda) = \beta_0 + \beta_1 \mathrm{I}(\mathrm{fup} = 1) + \beta_2 \mathrm{I}(\mathrm{fup} = 2) + \\ \beta_3 \mathrm{I}(\mathrm{fup} = 3) + \ldots + \beta_9 \mathrm{I}(\mathrm{fup} = 9) + \beta_{10} \mathrm{I}(\text{year8594} = 1)$$

The rate for the period 75–84 for year *k* after diagnosis is

$$\lambda(\text{year8594} = 0, \mathrm{fup} = k) = \exp(\beta_0 + \beta_k)$$

Regression equation

$$\log(\lambda) = \beta_0 + \beta_1 \mathrm{I}(\mathrm{fup} = 1) + \beta_2 \mathrm{I}(\mathrm{fup} = 2) + \beta_3 \mathrm{I}(\mathrm{fup} = 3) + \ldots + \beta_9 \mathrm{I}(\mathrm{fup} = 9) + \beta_{10} \mathrm{I}(\text{year8594} = 1)$$

The rate for the period 75–84 for year k after diagnosis is

$$\lambda(\text{year8594} = 0, \mathrm{fup} = k) = \exp(\beta_0 + \beta_k)$$

The rate for the period 85–94 for year k after diagnosis is

$$\lambda(\text{year8594} = 1, \mathrm{fup} = k) = \exp(\beta_0 + \beta_k + \beta_{10})$$

Regression equation

$$\log(\lambda) = \beta_0 + \beta_1 \mathrm{I}(\mathrm{fup} = 1) + \beta_2 \mathrm{I}(\mathrm{fup} = 2) + \beta_3 \mathrm{I}(\mathrm{fup} = 3) + \ldots + \beta_9 \mathrm{I}(\mathrm{fup} = 9) + \beta_{10} \mathrm{I}(\mathsf{year8594} = 1)$$

The rate for the period 75–84 for year k after diagnosis is

$$\lambda(\mathsf{year8594} = 0, \mathrm{fup} = k) = \exp(\beta_0 + \beta_k)$$

The rate for the period 85–94 for year k after diagnosis is

$$\lambda(\mathsf{year8594} = 1, \mathrm{fup} = k) = \exp(\beta_0 + \beta_k + \beta_{10})$$

The rate ratio comparing the period 85–94 with 75–84 is for year k after diagnosis is:

$$\frac{\lambda(\mathsf{year8594} = 1, \mathrm{fup} = k)}{\lambda(\mathsf{year8594} = 0, \mathrm{fup} = k)} = \frac{\exp(\beta_0 + \beta_k + \beta_{10})}{\exp(\beta_0 + \beta_k)} = \exp(\beta_{10})$$