Gov 2001 Section 9: Duration Models

Konstantin Kashin[1]

April 5, 2013

[1]Thanks to Molly Roberts, Jen Pan, Brandon Stewart, Iain Osgood, and Patrick Lam for contributing to this material.

Outline

Replication Paper

- Re-replication was due today!
- Go through comments carefully and make the transition to thinking about final product!
- Problem Set 6 released today; due next Wednesday, April 10 at 7pm

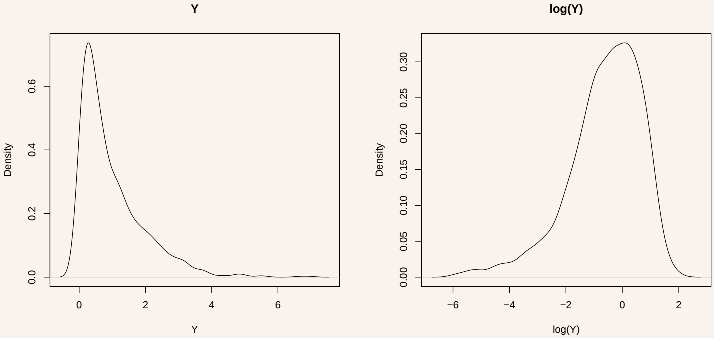
Outline

What are duration models used for?

- Survival models = duration models = event history models
- Dependent variable Y is the duration of time units spend in some state before experiencing an event (aka failure, death)
- Used in biostatistics and engineering: i.e. how long until a patient dies
- Models the relationship between duration and covariates (how does an increase in X affect the duration Y)
- In social science, used in questions such as how long a coalition government lasts, how long a war lasts, how long a regime stays in power, or how long until a legislator leaves office
- Observations should be measured in the same (temporal) units

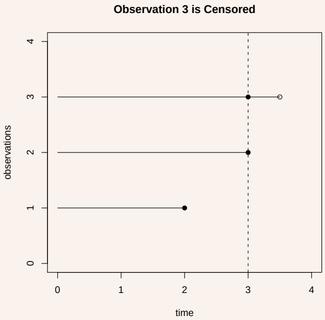
Why not use OLS?

1. OLS assumes Y is Normal but duration dependent variables are always positive (number of years, number of days. etc.)
   - Can possibly transform Y (log) to make it look Normal



Why not use OLS?

2. Duration models can handle censoring



Observation 3 is censored in that it has not experienced the event at the time we stop collecting data, so we don't know its true duration

Why not use OLS?

3. Duration models can handle time-varying covariates (TVCs)
   - If Y is duration of a regime, GDP may change during the duration of the regime
   - OLS cannot handle multiple values of GDP per observation
   - You can set up data in a special way with duration models such that you can accomodate TVCs
   - We won't cover this today but its the same principle as censoring

Dependent Variable

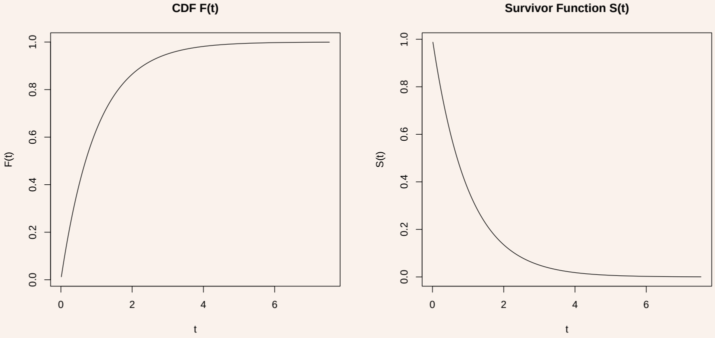
Let T denote a continuous positive random variable denoting the duration/survival times ($T = Y$)

T has a **probability density function** $f(t)$

Survivor Function

$f(t)$ has a corresponding CDF $F(t) = \int_0^t f(u)du = P(T \leq t)$, which is the probability of an event occurring by time t

Then the **survivor function** is $S(t) = 1 - F(t) = P(T > t)$: *probability of surviving (no event) until at least time t*



Hazard Rate

The **hazard rate** (or hazard function) $h(t)$ is roughly *the probability of an event at time t given survival up to time t*

$$\begin{aligned} h(t) &= P(t \leq T < t+\tau | T \geq t) \\ &= P(\text{event at } t | \text{survival up to } t) \\ &= \frac{P(\text{survival up to } t | \text{event at } t) P(\text{event at } t)}{P(\text{survival up to } t)} \\ &= \frac{P(\text{event at } t)}{P(\text{survival up to } t)} \\ &= \frac{f(t)}{S(t)} \end{aligned}$$

Relating the Density, Surival, and Hazard Functions

$$\underbrace{f(t)}_{\text{density function}} = \underbrace{h(t)}_{\text{hazard function}} \underbrace{S(t)}_{\text{survival function}}$$

Modeling with Covariates

We can model the mean of the duration times as a function of covariates via a link function $g(\cdot)$

$$g(E[T_i]) = X_i\beta$$

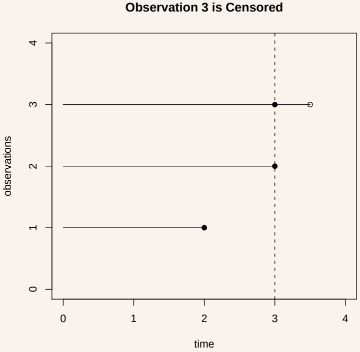
and estimate β via maximum likelihood.

Estimating Parametric Survival Models

1. Make an assumption that T_i follows a specific distribution $f(t)$ (stochastic component).
2. Model the hazard rate with covariates (systematic component).
3. Estimate via ML.
4. Interpret quantities of interest (hazard ratios, expected survival times).

What's Special About Survival Models?

Censoring:



Observation 3 is censored in that it has not experienced the event at the time we observe data, so we don't know its true duration.

Censoring

Observations that are censored give us information about how long they survive.

For censored observations, we know that they survived at least up to some observed time t^c and that their true duration is some $t \geq t^c$.

For each observation, create a censoring indicator c_i such that

$$c_i = \begin{cases} 1 & \text{if not censored} \\ 0 & \text{if censored} \end{cases}$$

Censoring

Our likelihood is function is then

$$
\begin{aligned}
L &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [P(T_i \geq t_i^c)]^{1-c_i} \\
&= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [1 - F(t_i)]^{1-c_i}
\end{aligned}
$$

We can incorporate the information from the censored observations into the likelihood function.

$$
\begin{aligned}
L &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [P(T_i^* \geq t_i^c)]^{1-c_i} \\
&= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [1 - F(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i}
\end{aligned}
$$

So uncensored observations contribute to the density function and censored observations contribute to the survivor function in the likelihood.

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government
- Independent variables:
  - investiture (`invest`)
  - fractionalization (`fract`)
  - polarization (`polar`)
  - numerical status (`numst2`)
  - crisis duration (`crisis`)
- Censoring occurs because of constitutionally mandated election period: governments end when this period ends (`ciep12`)

Outline

The Poisson Process

- Popular example of stochastic process (particularly a Levy process)
- Principles of Poisson process:
  - **Independent increments**: number of events occuring in two disjoint intervals is independent
  - **Stationary increments**: probability distribution of number of occurences depends only on the time length of interval (because of common rate)
- Homogenous Poisson process is governed by a common rate
- Events occur at rate λ (expected occurences per unit of time)
- N_τ = number of arrivals in time period of length τ
  - $N_\tau \sim \text{Poisson}(\lambda\tau)$

The Poisson Process

- Exponential distribution measures inter-arrival times in a Poisson process
- T = time to wait until next event in a Poisson process with rate λ
- $T \sim \text{Expo}(\lambda)$
- **Memorylessness property**: how much you have waited already is irrelevant

$$P(T > t + 5 | T > t) = P(t > 5)$$

The Two Parametrization of the Exponential Model

- $\lambda_i > 0$ is **rate** parameter

$$T_i \sim \text{Exponential}(\lambda_i)$$

$$f(t_i) = \lambda_i e^{-\lambda_i t_i}$$

$$E(T_i) = \mu_i = \frac{1}{\lambda_i}$$

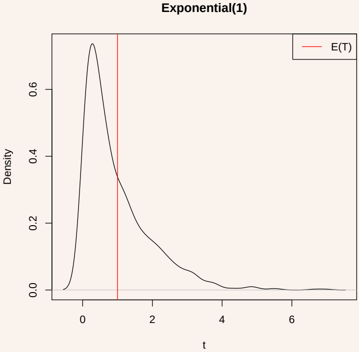
- $\theta_i > 0$ is **scale** parameter ($\theta_i = \frac{1}{\lambda_i}$)

$$T_i \sim \text{Exponential}(\theta_i)$$

$$f(t_i) = \frac{1}{\theta_i} e^{-\frac{t_i}{\theta_i}}$$

$$E(T_i) = \mu_i = \theta_i$$

The Exponential Model



Link Functions

- $\lambda_i > 0$ is **rate** parameter $\longrightarrow$ inverse link

$$E(T_i) = \frac{1}{\lambda_i} = \frac{1}{\exp(x_i\beta)}$$

Positive β implies that expected duration time decreases as x increases.

- $\theta_i > 0$ is **scale** parameter ($\theta_i = \frac{1}{\lambda_i}$) $\longrightarrow$ identity link

$$E(T_i) = \theta_i = \exp(x_i\beta)$$

$$\log(\theta_i) = x_i\beta$$

Positive β implies that expected duration time increases as x increases.

Hazard Function for Rate Parametrization

For $T_i \sim \text{Exponential}(\lambda_i)$:

$$f(t) = \lambda_i e^{-\lambda_i t}$$

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - \left(1 - e^{-\lambda t}\right) \\ &= e^{-\lambda_i t} \end{aligned}$$

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{\lambda_i e^{-\lambda_i t}}{e^{-\lambda_i t}} \\ &= \lambda_i \end{aligned}$$

Hazard Function for Scale Parametrization

For $T_i \sim \text{Exponential}(\theta_i)$:

$$
\begin{aligned}
f(t) &= \frac{1}{\theta_i}\exp[-\frac{t}{\theta_i}] \\
S(t) &= 1 - F(t) \\
&= 1 - (1 - \exp[-\frac{t}{\theta_i}]) \\
&= \exp[-\frac{t}{\theta_i}] \\
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{\frac{1}{\theta_i}\exp[-\frac{t}{\theta_i}]}{\exp[-\frac{t}{\theta_i}]} \\
&= \frac{1}{\theta_i}
\end{aligned}
$$

Let's work with the scale parametrization

- Note that $h(t) = \frac{1}{\theta_i}$, which does not depend on t!
  - The exponential model thus assume a **flat hazard**
  - Every unit / individual has their own hazard rate, but it does not change over time
  - Connected to **memorylessness property** of the exponential distribution

Modeling $h(t)$ with covariates:

$$h(t) = \frac{1}{\theta_i} = \exp[-x_i\beta]$$

Positive β implies that hazard decreases and average survival time increases as x increases.

Estimation via ML:

$$
\begin{aligned}
L &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [1 - F(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\frac{1}{\theta_i} e^{-\frac{t_i}{\theta_i}}\right]^{c_i} \; \left[e^{-\frac{t_i}{\theta_i}}\right]^{1-c_i} \\
\ln L &= \sum_{i=1}^{n} c_i \left(\ln \frac{1}{\theta_i} - \frac{t_i}{\theta_i}\right) + (1 - c_i)\left(-\frac{t_i}{\theta_i}\right) \\
&= \sum_{i=1}^{n} c_i \left(\ln e^{-x_i\beta} - e^{-x_i\beta} t_i\right) + (1 - c_i)(-e^{-x_i\beta} t_i) \\
&= \sum_{i=1}^{n} c_i \left(\ln e^{-x_i\beta} - e^{-x_i\beta} t_i + e^{-x_i\beta} t_i\right) - e^{-x_i\beta} t_i \\
&= \sum_{i=1}^{n} c_i \left(-x_i\beta\right) - e^{-x_i\beta} t_i
\end{aligned}
$$

Exponential Regression in Zelig

```
z.out <- zelig(Surv(duration, event = ciep12)
~ invest + fract + polar + numst2 + crisis,
 data = coalition, model = "exp")
z.out
```

```
Call:
"survreg"(formula = formula, dist = "exponential", robust = robust,
    data = data)
               Value Std. Error     z        p
(Intercept)  4.82672   0.615377  7.84 4.38e-15
invest      -0.50476   0.136144 -3.71 2.09e-04
fract       -0.00225   0.000881 -2.55 1.06e-02
polar       -0.02880   0.005919 -4.86 1.14e-06
numst2       0.46132   0.128172  3.60 3.19e-04
crisis       0.00559   0.002106  2.65 7.98e-03

Scale fixed at 1

Exponential distribution
Loglik(model)= -1046.3   Loglik(intercept only)= -1100.7
Chisq= 108.86 on 5 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 4
n= 314
```

How do we get quantities of interest?

Variable of interest: majority versus minority governments (`numst2`), with all other variables set at mean or mode.

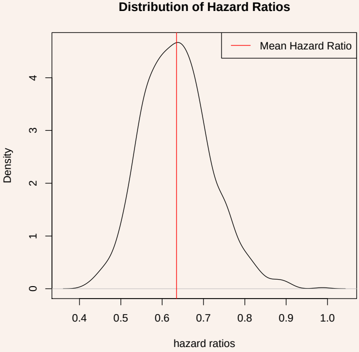
We could calculate:

- Find the hazard ratio of majority to minority governments
- Expected survival time for majority and minority governments
- Predicted survival times for majority and minority governments
- First differences in expected survival times between majority and minority governments

Hazard Ratios:

$$
\begin{aligned}
\text{HR} &= \frac{h(t|x_{\text{maj}})}{h(t|x_{\text{min}})} \\
&= \frac{e^{-x_{\text{maj}}\beta}}{e^{-x_{\text{min}}\beta}} \\
&= \frac{e^{-\beta_0} e^{-x_1\beta_1} e^{-x_2\beta_2} e^{-x_3\beta_3} e^{-x_{\text{maj}}\beta_4} e^{-x_5\beta_5}}{e^{-\beta_0} e^{-x_1\beta_1} e^{-x_2\beta_2} e^{-x_3\beta_3} e^{-x_{\text{min}}\beta_4} e^{-x_5\beta_5}} \\
&= \frac{e^{-x_{\text{maj}}\beta_4}}{e^{-x_{\text{min}}\beta_4}} \\
&= e^{-\beta_4}
\end{aligned}
$$

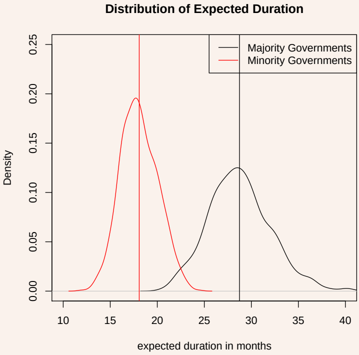
Hazard ratio greater than 1 implies that majority governments fall faster (shorter survival time) than minority governments.



Majority governments survive longer than minority governments.

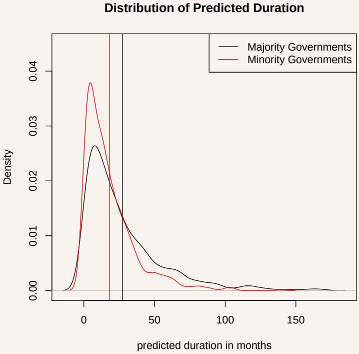
Expected (average) Survival Time:

$$
\begin{aligned}
E(T|x_i) &= \theta_i \\
&= \exp[x_i\beta]
\end{aligned}
$$



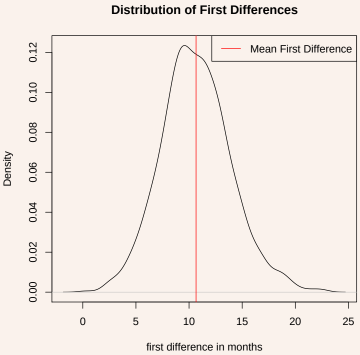
Predicted Survival Time:

Draw predicted values from the exponential distribution.



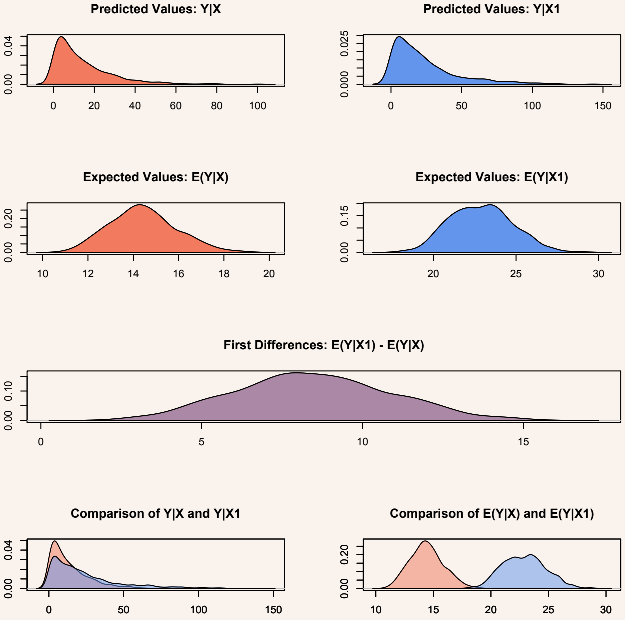
First Differences:

$$E(T|x_{\text{maj}}) - E(T|x_{\text{min}})$$



Quantities of Interest in Zelig

```
x.min <- setx(z.out,numst2=0)
x.maj <- setx(z.out,numst2=1)
s.out <- sim(z.out, x=x.min,x1=x.maj)
summary(s.out)
plot(s.out)
```


Predicted Values: Y|X
Predicted Values: Y|X1
Expected Values: E(Y|X)
Expected Values: E(Y|X1)
First Differences: E(Y|X1) - E(Y|X)
Comparison of Y|X and Y|X1
Comparison of E(Y|X) and E(Y|X1)

The exponential model is nice and simple, but the assumption of a flat hazard may be too restrictive.

What if we want to loosen that restriction by assuming a monotonic hazard?

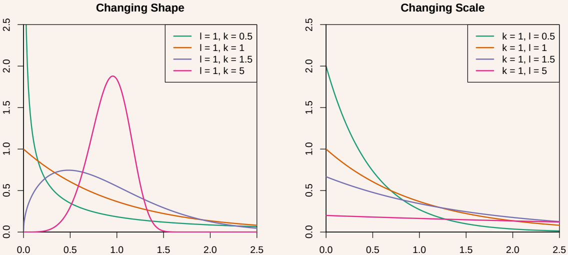
We can use the Weibull model.

Outline

The Weibull Model

$$\begin{aligned} T_i &\sim \text{Weibull}(\lambda_i, \alpha) \\ E(T_i) &= \lambda_i \Gamma\left(1 + \frac{1}{\alpha}\right) \end{aligned}$$

$\lambda_i > 0$ is the scale parameter and $\alpha > 0$ is the shape parameter.



The Weibull Model

$$f(t_i) \quad = \quad \left(\frac{\alpha}{\lambda_i^\alpha}\right) t_i^{\alpha-1} \exp\left[-\left(\frac{t_i}{\lambda_i}\right)^\alpha\right]$$

Model λ_i with covariates:

$$\lambda_i = \exp(x_i\beta)$$

$$log(\lambda_i) \quad = x_i\beta$$

Positive β implies that expected duration time increases as x increases.

$$f(t_i) = \left(\frac{\alpha}{\lambda_i^\alpha}\right) t_i^{\alpha-1} \exp\left[-\left(\frac{t_i}{\lambda_i}\right)^\alpha\right]$$

$$\begin{aligned} S(t_i) &= 1 - F(t_i) \\ &= 1 - \left(1 - e^{-(t_i/\lambda_i)^\alpha}\right) \\ &= e^{-(t_i)/\lambda_i)^\alpha} \end{aligned}$$

$$\begin{aligned} h(t_i) &= \frac{f(t_i)}{S(t_i)} \\ &= \frac{\left(\frac{\alpha}{\lambda_i^\alpha}\right) t_i^{\alpha-1} \exp\left[-\left(\frac{t_i}{\lambda_i}\right)^\alpha\right]}{e^{-(t_i/\lambda_i)^\alpha}} \\ &= \left(\frac{\alpha}{\lambda_i}\right)\left(\frac{t_i}{\lambda_i}\right)^{\alpha-1} \\ &= \left(\frac{\alpha}{\lambda_i^\alpha}\right) t_i^{\alpha-1} \end{aligned}$$

Modeling $h(t_i)$ with covariates:

$$h(t_i) \quad = \quad \left(\frac{\alpha}{\lambda_i^{\alpha}}\right) t_i^{\alpha-1}$$

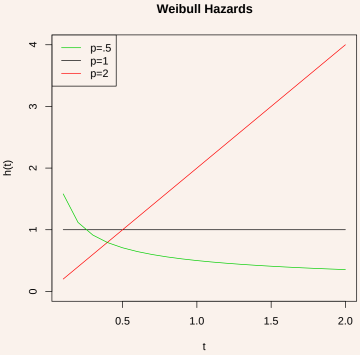
We want to constrain the $h(t)$ to be positive.

$$\lambda_i \quad = \quad \exp[x_i \beta]$$

Positive β implies that hazard decreases and average survival time increases as x increases. This is the same interpretation as the scale parametrization of the exponential model, but different than the rate parametrization!

$h(t_i)$ is modeled with both λ_i and α and is a function of t_i. Thus, the Weibull model assumes a **monotonic hazard**.

- If $\alpha = 1$, $h(t_i)$ is flat and the model is the exponential model.
- If $\alpha > 1$, $h(t_i)$ is monotonically increasing.
- If $\alpha < 1$, $h(t_i)$ is monotonically decreasing.



Deriving the Log-Likelihood for the Weibull Model

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \left[f(t_i)\right]^{c_i} \left[1 - F(t_i)\right]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{\alpha}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{\alpha-1} e^{-(t/\lambda)^{\alpha}}\right]^{c_i} \left[e^{-(t/\lambda)^{\alpha}}\right]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{\alpha}{\lambda^{\alpha}}\right) t^{\alpha-1} e^{-(t/\lambda)^{\alpha}}\right]^{c_i} \left[e^{-(t/\lambda)^{\alpha}}\right]^{1-c_i} \\
\ln L &= \sum_{i=1}^{n} c_i \left[\ln \alpha - \alpha \ln \lambda + (\alpha - 1) \ln t_i\right] - \left(\frac{t_i}{\lambda}\right)^{\alpha} \\
&= \sum_{i=1}^{n} c_i \left[\ln \alpha - \alpha\, x_i\beta + (\alpha - 1) \ln t_i\right] - \left(\frac{t_i}{e^{x_i\beta}}\right)^{\alpha}
\end{aligned}
$$

```
Call:
survreg(formula = Surv(duration, event = ciep12) ~ invest + fract +
    polar + numst2 + crisis, data = coalition, dist = "weibull")
               Value Std. Error     z        p
(Intercept)  4.75007   0.530723  8.95 3.55e-19
invest      -0.47160   0.116433 -4.05 5.11e-05
fract       -0.00212   0.000759 -2.79 5.26e-03
polar       -0.02792   0.005056 -5.52 3.33e-08
numst2       0.42746   0.110252  3.88 1.06e-04
crisis       0.00538   0.001829  2.94 3.28e-03
Log(scale)  -0.15644   0.049708 -3.15 1.65e-03

Scale= 0.855

Weibull distribution
Loglik(model)= -1041.7   Loglik(intercept only)= -1100.6
Chisq= 117.84 on 5 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 314
```

The shape parameter α for the Weibull distribution is the inverse of the scale parameter given by `survreg()`.

The scale parameter given by `survreg()` is NOT the same as the scale parameter in the Weibull distribution, which should be $\lambda_i = e^{x_i\beta}$.

What's the Difference?

Assumptions about the hazard rate.

The **hazard rate** (or hazard function) $h(t)$ is roughly *the probability of an event at time t given survival up to time t*.

$$\begin{aligned} h(t) &= P(\text{event at } t | \text{survival up to } t) \\ &= \frac{f(t)}{1 - F(t)} \end{aligned}$$

$$h(t|x) = h_0(t) exp(x_i \beta)$$

$h_0(t)$ is the baseline hazard. This describes how hazard changes over time. The Exponential model assumes a **flat hazard**, Weibull assumes **monotonic hazard**.

Hazard Ratios

One quantity of interest is the hazard ratio:

$$HR = \frac{h(t|x=1)}{h(t|x=0)}$$

Proportional hazards assumption: hazard ratio does not depend t.
Covariates are multiplicatively related to hazard.

Other Parametric Models

- Gompertz model: monotonic hazard
- Log-logistic or log-normal model: nonmonotonic hazard
- Generalized gamma model: nests the exponential, Weibull, log-normal, and gamma models with an extra parameter

But what if we don't want to make an assumption about the shape of the hazard?

Outline

The Cox Proportional Hazards Model

- Often described as a semi-parametric model.
- Makes no assumptions about the shape of the hazard or the distribution of T_i.
- Takes advantage of the proportional hazards assumption.

Pros:

- Makes no restrictive assumption about the shape of the hazard.
- A better choice if you want the effects of the covariates and the nature of the time dependence is unimportant.

Cons:

- Only quantities of interest are hazard ratios.
- Can be subject to overfitting
- Shape of hazard is unknown (although there are semi-parametric ways to derive the hazard and survivor functions)

```
Call:
coxph(formula = Surv(duration, event = ciep12) ~ invest + fract +
    polar + numst2 + crisis, data = coalition)

  n= 314

            coef  exp(coef)   se(coef)      z Pr(>|z|)
invest  0.5518753  1.7365064  0.1368178  4.034 5.49e-05 ***
fract   0.0023978  1.0024006  0.0008976  2.671 0.007557 **
polar   0.0327720  1.0333149  0.0061055  5.368 7.98e-08 ***
numst2 -0.4852138  0.6155655  0.1294150 -3.749 0.000177 ***
crisis -0.0061808  0.9938383  0.0021566 -2.866 0.004158 **
---
Signif. codes:  0 *** 0.001 **
0.01 * 0.05 . 0.1 1
```

```
       exp(coef) exp(-coef) lower .95 upper .95
invest    1.7365     0.5759    1.3281    2.2706
fract     1.0024     0.9976    1.0006    1.0042
polar     1.0333     0.9678    1.0210    1.0458
numst2    0.6156     1.6245    0.4777    0.7933
crisis    0.9938     1.0062    0.9896    0.9980

Rsquare= 0.299   (max possible= 1 )
Likelihood ratio test= 111.5  on 5 df,   p=0
Wald test            = 109.5  on 5 df,   p=0
Score (logrank) test = 117.1  on 5 df,   p=0
```

References

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling.* Cambridge University Press.

King, Gary, James E. Alt, Nancy E. Burns, and Michael Laver. 1990. "A Unified Model of Cabinet Dissolution in Parliamentary Democracies." *American Journal of Political Science* 34(3): 846-971

Long, S. J. (1997) *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: SAGE Publications, Inc.

McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models*, Second Edition. Boca Raton: Chapman and Hall/CRC

Lam, Patrick. Survival Model Notes. http://www.people.fas.harvard.edu/~plam/teaching.html