PLSC 504 - Fall 2017
Duration Dependence

October 3, 2017

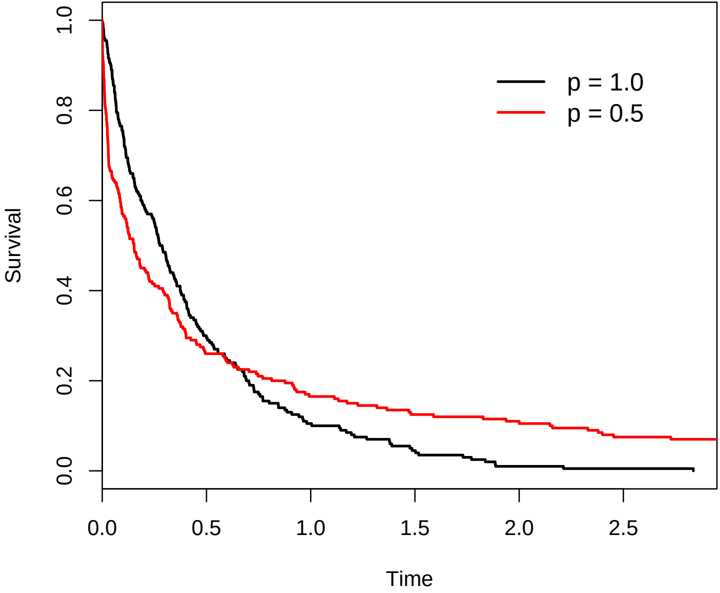
Stratification

- Allow different groups to have different baseline hazards
- Akin to different intercepts, but more flexible.
- *Assumes covariate effects are otherwise identical*
- Uses:
  - Unit/group heterogeneity
  - Nonproportional hazards
  - Simple models for duration dependence

Stratification, Simulated

```
> set.seed=7222009
> Z<-rnorm(200)
> X0<-rep(0,times=200)
> X1<-rep(1,times=200)
> T0<-rweibull(200,shape=1,scale=1/exp(2+0.5*Z))
> T1<-rweibull(200,shape=0.5,scale=1/exp(2+0.5*Z))
> C<-rep(1,times=400)
> X<-append(X0,X1)
> T<-append(T0,T1)
> data<-as.data.frame(cbind(T,C,X,rep(Z,times=2)))
> colnames(data)<-c("T","C","X","Z")
```

Stratified Weibull Hazards



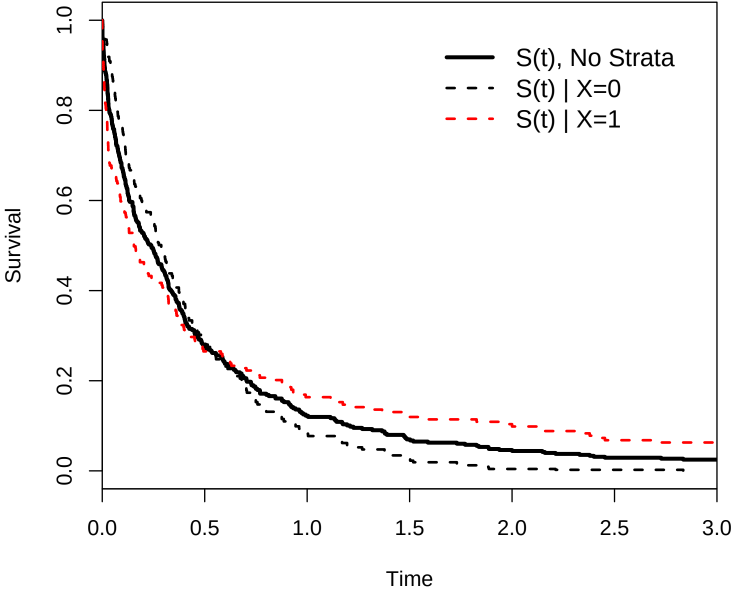
Stratification, Simulated

```
> cox<-coxph(S~Z+X,data=data)
> summary(cox)
Call:
coxph(formula = S ~ Z + X, data = data)

  n= 400, number of events= 400

      coef exp(coef) se(coef)      z Pr(>|z|)
Z  0.28286   1.32692  0.05133  5.510 3.58e-08 ***
X -0.22866   0.79560  0.10639 -2.149   0.0316 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

  exp(coef) exp(-coef) lower .95 upper .95
Z    1.3269     0.7536    1.1999    1.4674
X    0.7956     1.2569    0.6459    0.9801

Concordance= 0.571  (se = 0.017 )
Rsquare= 0.08   (max possible= 1 )
Likelihood ratio test= 33.25  on 2 df,   p=6.022e-08
Wald test            = 33.02  on 2 df,   p=6.749e-08
Score (logrank) test = 33.07  on 2 df,   p=6.601e-08
```

Stratification, Simulated

```
> cox.strata<-coxph(S~Z+strata(X),data=data)
> summary(cox.strata)
Call:
coxph(formula = S ~ Z + strata(X), data = data)

  n= 400, number of events= 400

     coef exp(coef) se(coef)    z Pr(>|z|)
Z 0.32140   1.37906  0.05176 6.21  5.3e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

  exp(coef) exp(-coef) lower .95 upper .95
Z     1.379     0.7251     1.246     1.526

Concordance= 0.597  (se = 0.024 )
Rsquare= 0.092   (max possible= 1 )
Likelihood ratio test= 38.69  on 1 df,   p=4.955e-10
Wald test            = 38.56  on 1 df,   p=5.303e-10
Score (logrank) test = 38.62  on 1 df,   p=5.151e-10
```

Cox $\widehat{S(t)}$s



Stratified Weibull Model

```
> summary(survreg(S~Z+strata(X),data=data,dist="weibull"))

Call:
survreg(formula = S ~ Z + strata(X), data = data, dist = "weibull")
              Value Std. Error       z        p
(Intercept) -0.9976     0.0675 -14.781 1.93e-49
Z           -0.4140     0.0577  -7.178 7.06e-13
X=0          0.0152     0.0555   0.274 7.84e-01
X=1          0.6864     0.0543  12.650 1.11e-36

Scale:
 X=0  X=1   # Recall: scale = 1 / p
1.02 1.99

Weibull distribution
Loglik(model)= -7.8   Loglik(intercept only)= -31.4
Chisq= 47.36 on 1 degrees of freedom, p= 5.9e-12
Number of Newton-Raphson Iterations: 6
n= 400
```

Duration Dependence

1. *State Dependence*
   - E.g., Institutionalization / Degradation

Positive State Dependence $\longrightarrow$ Negative Duration Dependence

Negative State Dependence $\longrightarrow$ Positive Duration Dependence

Duration Dependence

2. *Unobserved / Unmodeled Heterogeneity*
    - $h(t|\mathbf{X}_i) \neq h(t|\mathbf{X}_j)$ for $\mathbf{X}_i = \mathbf{X}_j$
    - Adverse selection in the sample / data
    - Result: "Spurious" duration dependence

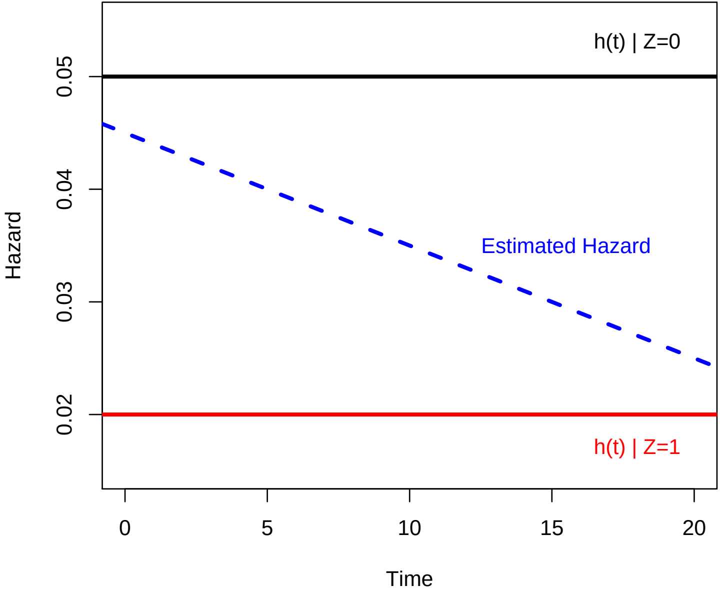
Suppose we have an unobserved Z, with

$$h_i(t|\mathbf{X}_i, Z_i = 0) = 0.05$$

and

$$h_i(t|\mathbf{X}_i, Z_i = 1) = 0.02.$$

Unobserved Heterogeneity Illustrated



Unobserved Heterogeneity: A Simulation

```
> set.seed(7222009)
> W<-rnorm(500)
> X<-rnorm(500)
> Z<-rnorm(500)
> T<-rexp(500,rate=(exp(0+0.5*W+0.5*X-0.6*Z))) # exponential hazard
> C<-rep(1,times=500)
> S<-Surv(T,C)
> summary(survreg(S~W,dist="weibull"))

Call:
survreg(formula = S ~ W, dist = "weibull")
              Value Std. Error      z        p
(Intercept) -0.0101     0.0629  -0.16 8.73e-01
W           -0.6339     0.0610 -10.40 2.47e-25
Log(scale)   0.2833     0.0333   8.52 1.62e-17

Scale= 1.33 # implies p = 1/Scale = 0.753

Weibull distribution
Loglik(model)= -568.1   Loglik(intercept only)= -615.3
Chisq= 94.47 on 1 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

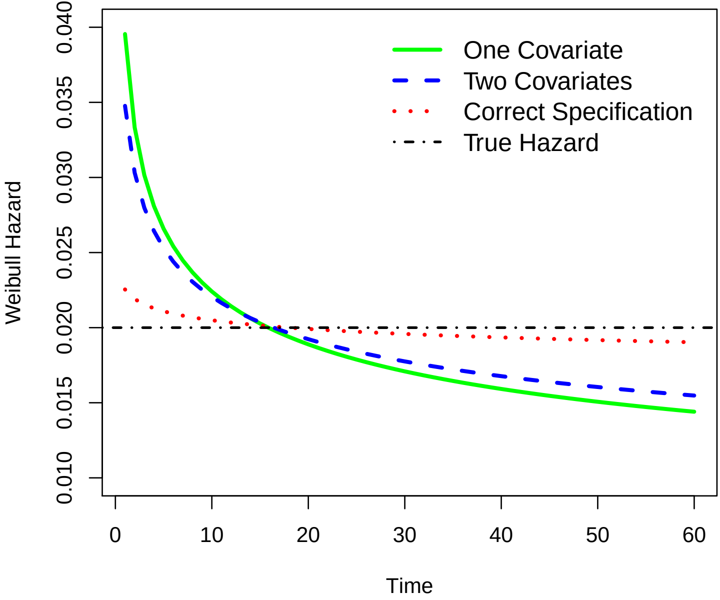
Unobserved Heterogeneity: A Simulation

```
> summary(survreg(S~W+X,dist="weibull"))

Call:
survreg(formula = S ~ W + X, dist = "weibull")
              Value Std. Error       z        p
(Intercept) -0.0511     0.0591  -0.865 3.87e-01
W           -0.5907     0.0581 -10.160 2.98e-24
X           -0.4750     0.0556  -8.549 1.24e-17
Log(scale)   0.2202     0.0329   6.689 2.24e-11

Scale= 1.25 # implies p = 1/Scale = 0.802

Weibull distribution
Loglik(model)= -534.5   Loglik(intercept only)= -615.3
Chisq= 161.6 on 2 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

Unobserved Heterogeneity: A Simulation

```
> summary(survreg(S~W+X+Z,dist="weibull"))

Call:
survreg(formula = S ~ W + X + Z, dist = "weibull")
              Value Std. Error      z        p
(Intercept) -0.0777     0.0494  -1.57 1.16e-01
W           -0.5665     0.0468 -12.11 9.17e-34
X           -0.5041     0.0473 -10.66 1.58e-26
Z            0.5923     0.0446  13.29 2.73e-40
Log(scale)   0.0423     0.0345   1.22 2.21e-01

Scale= 1.04 # implies p = 1/Scale = 0.959

Weibull distribution
Loglik(model)= -464.3   Loglik(intercept only)= -615.3
Chisq= 302.01 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

Unobserved Heterogeneity: A Simulation



Duration Dependence: What To Do?

(At least) Three Options:

1. Model Specification
2. Unit-Level Effects
3. Model the Duration Dependence

Modeling Duration Dependence

Weibull with:

$$p = \exp(\mathbf{Z}_i\gamma)$$

Gives:

$$h_i(t) = \exp(\mathbf{X}_i\beta)\exp(\mathbf{Z}_i\gamma)[\exp(\mathbf{X}_i\beta)t]^{[\exp(\mathbf{Z}_i\gamma)]-1}$$

and (more usefully):

$$S(t) = \exp(-\exp(\mathbf{X}_i\beta)t)^{\exp(\mathbf{Z}_i\gamma)}$$

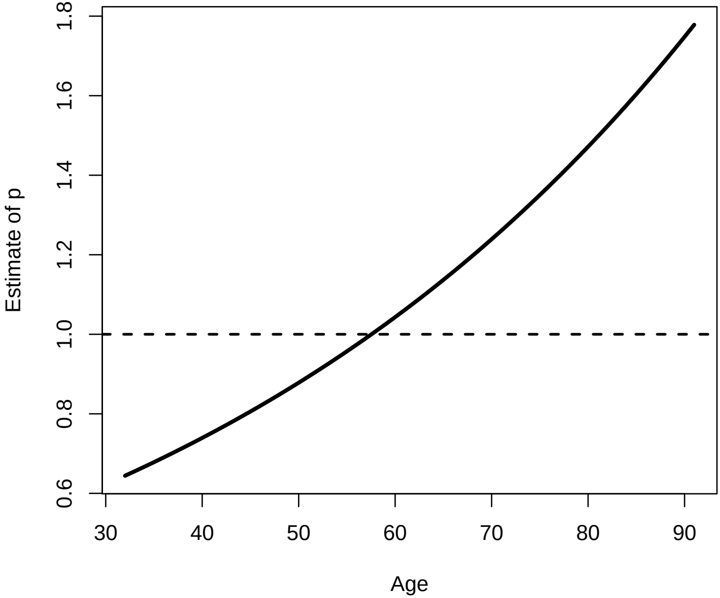
Example: SCOTUS Departures

```
> library(flexsurv)
> ct.weib<-flexsurvreg(scotus.S~age+pension+pagree,
                    data=scotus,dist="weibull")
> ct.weib

Estimates:
        data mean  est        L95%       U95%       exp(est)
shape          NA     0.999      0.637      1.570         NA
scale          NA   942.000     13.700  64800.000         NA
age        62.100    -0.041     -0.102      0.019      0.959
pension     0.199    -1.310     -2.360     -0.265      0.269
pagree      0.616    -0.113     -0.673      0.447      0.893
        L95%       U95%
shape          NA        NA
scale          NA        NA
age         0.903     1.020
pension     0.095     0.767
pagree      0.510     1.560

N = 1765,  Events: 51,  Censored: 1714
Total time at risk: 1765
Log-likelihood = -209, df = 5
AIC = 429
```

Example: SCOTUS Departures

```
> ct.weib.DD<-flexsurvreg(scotus.S~age+pension+pagree+shape(age),
                         data=scotus,dist="weibull")
> ct.weib.DD

Estimates:
           data mean  est         L95%        U95%
shape             NA     0.3710      0.1260      1.0900
scale             NA   491.0000     16.7000  14500.0000
age          62.1000    -0.0307     -0.0779      0.0164
pension       0.1990    -1.0900     -1.9700     -0.2190
pagree        0.6160    -0.0328     -0.4840      0.4180
shape(age)   62.1000     0.0172     -0.0011      0.0356
           exp(est)   L95%        U95%
shape             NA          NA          NA
scale             NA          NA          NA
age           0.9700      0.9250      1.0200
pension       0.3350      0.1400      0.8030
pagree        0.9680      0.6160      1.5200
shape(age)    1.0200      0.9990      1.0400

N = 1765,  Events: 51,  Censored: 1714
Total time at risk: 1765
Log-likelihood = -208, df = 6
AIC = 427
```

$\hat{p}$ by Age



$\widehat{h(t)}$s by Age and Model