Survival Analysis using MATVEC

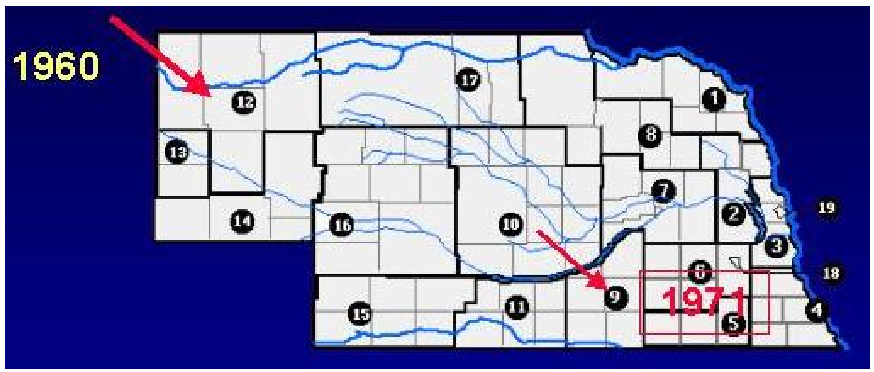
Stephen D. Kachman
Department of Biometry
Gonzalo Martínez
Animal Science Department
University of Nebraska Lincoln

Introduction

- Length of time between two events
  - Define:
    - Start point (birth, enter production, . . .)
      - Time $t = 0$
    - End point (death, sale, illness, . . .)
  - Length of productive life

Project

- Project 46-002 entitled "Effect of selection for weaning weight, yearling weight, and muscling in beef cattle" (Koch et al. 1974)
  - Initiated 1960 and data for cows born between 1964 and 1980
  - In 1971 the cows were moved from Fort Robinson, NE to Clay Center, NE



- Survival (`yol`): number of years between first calving and disposal

  1. Not pregnant at weaning
  2. Serious unsoundness
  3. Failure to raise a live calf for 2 consecutive years
  4. Old age

Censoring

- Incomplete record
  - End point hasn't occurred yet
  - Animal is removed from the herd before the end

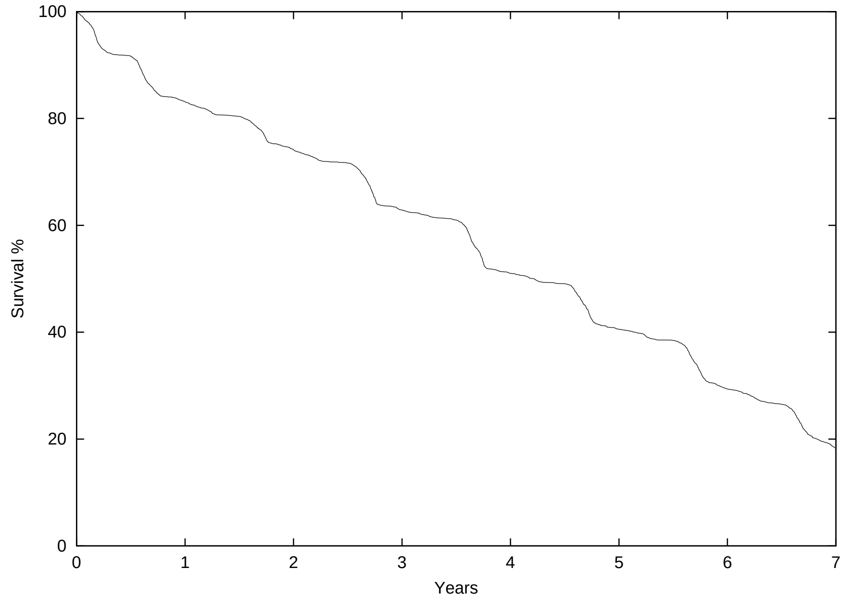
- A record was recorded as censored (`censor=0`)
  1. cow sold or culled for reasons not related to the experiment
  2. cow was still alive (7 years of age)

- Otherwise a record was recorded as uncensored (`censor=1`)

Survival analysis

- Packages
  - `SAS`: `Proc LIFEREG`
    - Fixed effects models
  - `Survival Kit` and `MATVEC`
    - Mixed models

Survival Function



$$S(t; \boldsymbol{\eta}_i) = \Pr(T_i \geq t)$$

- Probability that an individual with a given risk factor η_i will survive till time t
- Length of time, implies that Survival is 100% at time 0
- Decreasing function
  - Nobody "unfails"
- Features
  - Shape of the survival function
  - Location "Stretching of the time scale"

Kaplan-Meyer

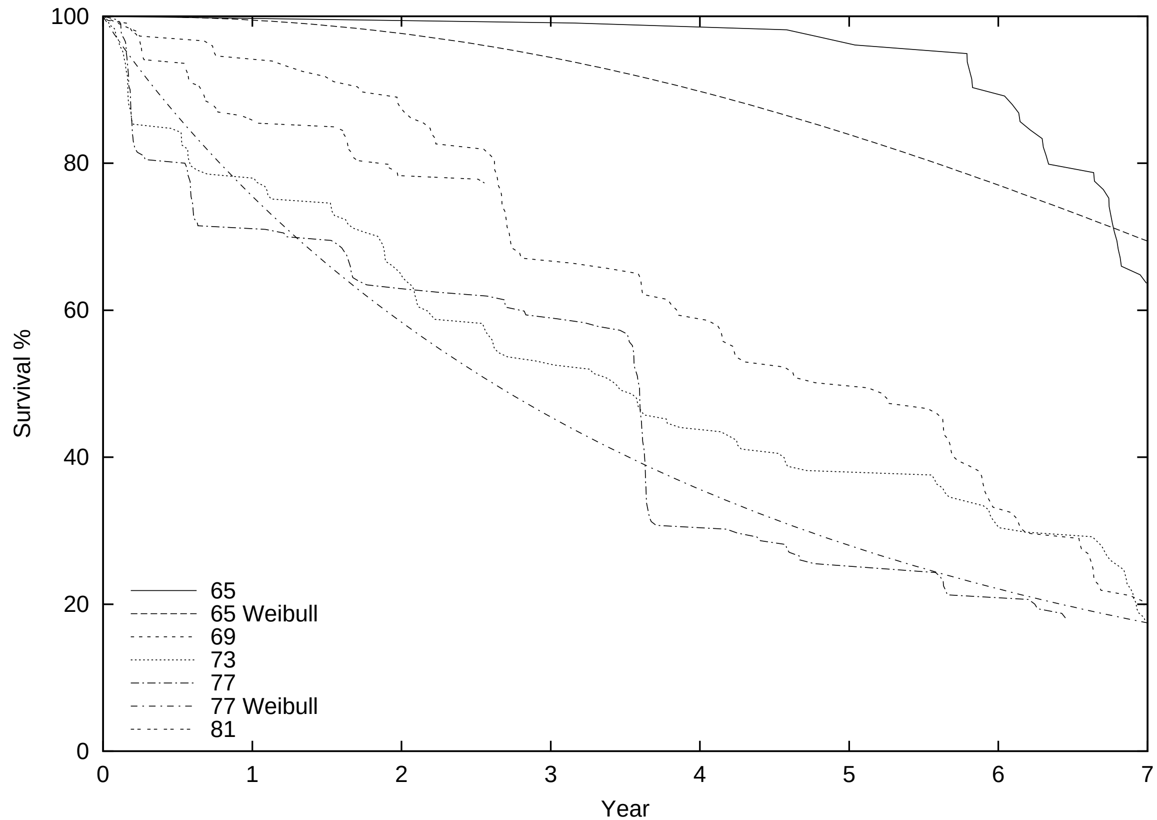
```
D=Data();
D.input("sire_yob.dat","anim sire dam line yob lod censor yol");
K=D.kaplan("yol","censor");
K.save("kap_sur.dat","out");
```

```
        0         1       inf
        0 0.999680306905 182.441705768
 0.00547945 0.999360613811 364.882708284
. . .
  6.98082 0.18524381674 181.262497358
   6.9863 0.184006107497 182.072599915
  6.99178 0.183593537749 182.072599915
```

- Assumed that there was a single survival function

- Could be that the survival function varied by year of birth

```
for(yob=64;yob<=82;yob++) {
D=Data();
datfile=string("YOB-",yob,".dat")
D.input(datfile,"anim sire dam line yob lod oldcensor
        yol grp $ censor");
K=D.kaplan("yol","censor");
kapfile=string("Kap-",yob,".dat")
K.save(kapfile,"out");
}
```



100
80
60
40
20
0
Survival %
0
1
2
3
4
5
6
7
Year
65
65 Weibull
69
73
77
77 Weibull
81

- It appears that for
  - Fort Robinson: the culling pressure was initially low and then increased
  - Clay Center: The culling pressure was initially high and then leveled off

Hazard Function

Short term risk of failure for animal alive at time t

$$\lambda(t; \boldsymbol{\eta}_i)$$

Over <u>short</u> periods of time, the probability that an animal fails is approximately equal to the hazard rate times the period of time.

- The higher the hazard rate the shorter the time period that this approximation is reasonable.
- Dramatic shifts will also make the approximate relationship poorer.

Weibull

- Hazard function is nonnegative
- Look at the log of the hazard function $\ln(\lambda(t, \boldsymbol{\eta}_i))$
- Parameterize so the Survival function looks nice

$$\ln(\lambda(t; \boldsymbol{\eta}_i)) = [\ln(\rho) + \eta_i] + (\rho - 1)\ln(t)$$
$$S(t; \boldsymbol{\eta}_i) = e^{-t^{\rho}\exp(\eta_i)}$$

- Watch out for time zero!

Weibull hazard function

- $\rho > 1$
  - Increasing Hazard function
  - $\lambda(0) \to 0$

- $\rho < 1$
  - Decreasing Hazard function
  - $\lambda(0) \to \infty$

Model

$$t_i \sim \mathrm{Weib}(\eta_i, \rho_i)$$

- Risk factor (η_i)
  - Year of Birth
  - Sire $\mathrm{N}(\mathbf{0}, \boldsymbol{A}\sigma_s^2)$
- Rate parameter (ρ_i)
  - Birth Location

MATVEC

```
P=Pedigree();
P.input("sire.ped","individual father mother");
M=Model();
M.equation("yol=  yob sire,censor= grp");
M.param(0,"yol","censor",.5/365.25);
M.variance("sire",P,1,0.00,0.00,1);
M.link("weibull",0);
Init=Vector(1,1.3)
M.init(Init);
M.fitdata(D);
```

Analysis

- Analysis by Penalized Quasi-likelihood

```
nest=M.glim(10).nrow();
vce=M.vce_aireml(10,1.e-6);
M.info("anisur.infout","out");
M.save("anisur.out","out");
```

```
Original  Residual Log Likelihood:-2509.20208087

 Iteration 1.0 Res Log Like -2440.55481456 Change 68.647266308
 Iteration 2.0 Res Log Like -2439.96041245 Change 0.594402110166
 Iteration 3.0 Res Log Like -2439.94952014 Change 0.0108923095286
 Iteration 4.0 Res Log Like -2439.94921835 Change 0.000301791919355
 Iteration 5.0 Res Log Like -2439.94921139 Change 6.95912103765e-06
 Iteration 6.0 Res Log Like -2439.94921151 Change -1.21479388326e-07

 Iteration 6 Converged
 Final Estimates
0.0534112407297
       0
       1
```

Info File

```
             some extra information in the model
 --------------------------------------------------------
AI REML converged
maximum log restricted likelihood = -2439.94921151
 variance for sire =
 0.0534112407297        0
        0        1
 residual variance =
        1

 MME dimension   : 928
 non-zeros in MME: 3010
```

Rate parameter test

- Estimate the rate parameters

```
M.contrast("grp",identity(2,2).kron([0 1]));
```

- Are the rate parameters the same for Fort Robinson and Clay Center?

```
M.contrast("grp",[0 1 0,-1]);
```

```
           RESULTS FROM CONTRAST(S)
-----------------------------------------------------------
 Contrast MME_addr    K_coef   Raw_data_code
 ---------------------------------------------
  1         926             1   censor:grp:FR
         estimated value (K'b-M) = 2.18090847582 +- 0.0881395475837
            Prob(|t| > 24.743812915) = 0 (p_value)

  2         928             1   censor:grp:CC
         estimated value (K'b-M) = 0.9389580339 +- 0.0219260283712
            Prob(|t| > 42.8238994314) = 0 (p_value)
-----------------------------------------------------------
  joint hypothesis test H: K'b = M
  Prob(F > 1223.19861203) = 0 (p_value)
```

```
 2000 (Error degrees of freedom)

           RESULTS FROM CONTRAST(S)
-----------------------------------------------------------
 Contrast MME_addr    K_coef   Raw_data_code
 ---------------------------------------------
  1        926            1   censor:grp:FR
  1        928           -1   censor:grp:CC
         estimated value (K'b-M) = 1.24195044192 +- 0.09082838111
            Prob(|t| > 13.6735943847) = 0 (p_value)
            2000 (Error degrees of freedom)
-----------------------------------------------------------
```

Median Survival Time

$$m_{.5}(\eta_i, \rho_i) = [-\ln(.5)]^{1/\rho_i} e^{-\eta_i/\rho_i}$$

```
"Fort Robinson"
Krho=[zeros(1,nest-3) 1];
rho=M.estimate(Krho);
k=(-log(.5))^(1/rho);
est=k*(M.estimate(identity(7,7).kron([1 0]))/(-rho)).exp();
stderr=k*est.diag()*(M.covmat(identity(7,7).kron([1 0])).diag(0).sqrt())/rho;
[est; stderr].t()
"Clay Center"
```

Fort Robinson			Clay Center		
64	12.442216813	1.42430935298	71	2.75543995995	0.204315497852
65	9.38051934074	0.910194812466	72	3.29744639986	0.244672389254
66	7.28420006216	0.61200081431	73	3.04704022065	0.227866286068
67	6.61889595216	0.542290579784	74	2.23033787166	0.151723080441
68	5.79398184037	0.464671633113	75	3.39259584975	0.255167514455
69	4.71691133429	0.346856701048	76	1.45803918349	0.0974536922244
70	4.79263365665	0.351204028866	77	2.616410182	0.184691638011
			78	5.40491339814	0.45196617648
			79	6.26233020749	0.590162057172
			80	5.51588632867	0.505238454286
			81	7.42208354415	0.851485170547

Heritability

- Log scale

$$h_l^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \frac{\pi^2}{6}}$$

```
h2=4*sigma2s/(sigma2s+pi*pi/6);
se=(h2/sigma2s-h2*h2/(sigma2s*4))*sesigma2s;
string("\nh2 Log Scale=",h2," +- ",se,"\n")

h2 Log Scale=0.125796 +- 0.042036
```

- Observed Scale

$$h_o^2 = \exp\left(\frac{2\gamma}{\rho_i}\right) h_l^2$$

```
h2o=h2*exp(2*.5772/rho);
h2ose=se*exp(2*.5772/rho);
string("\nh2 Observed scale=",h2o," +- ",h2ose,"\n")

h2 Observed scale=0.430139 +- 0.143735
```

Expected Progeny Differences

- Percentage change in median survival time

$$EPD = 100\left(e^{-S_i/\rho} - 1\right)$$

```
nsires=nest/2-21;
Ksire=[zeros(nsires,19) identity(nsires,nsires)].kron([1 0]);
bv=M.estimate(Ksire);
bvse=M.covmat(Ksire).diag(0);
EPD=100*(-bv/rho).exp();
EPDse=EPD.diag()*bvse/rho;
EPD=EPD-100;
```

```
j=19*2+1;
for(i=1;i<=10;i++){
string(M.label(j)," ",EPD(i)," +- ",EPDse(i))
j+=2;
}
        yol:sire:450117 2.58581 +- 5.79113
        yol:sire:470018 0.258434 +- 5.67911
        yol:sire:470182 11.4184 +- 5.5594
        yol:sire:474160 -1.93891 +- 5.51125
        yol:sire:481657 -0.577784 +- 5.44309
        yol:sire:494424 6.96224 +- 5.4505
        yol:sire:566187 0.236756 +- 5.47451
        yol:sire:310982 1.15697 +- 5.74872
        yol:sire:420122 0.258434 +- 5.67911
```

Summary

- The analysis took about 13 seconds
  - About 1,000 equations

- Hazard functions can change
  - Between and within environments
  - Time dependent covariates
    Can be modeled using a censor code of -1 for left censored record
  - Rate parameter does not need to be a constant

- Definition of "Survival"