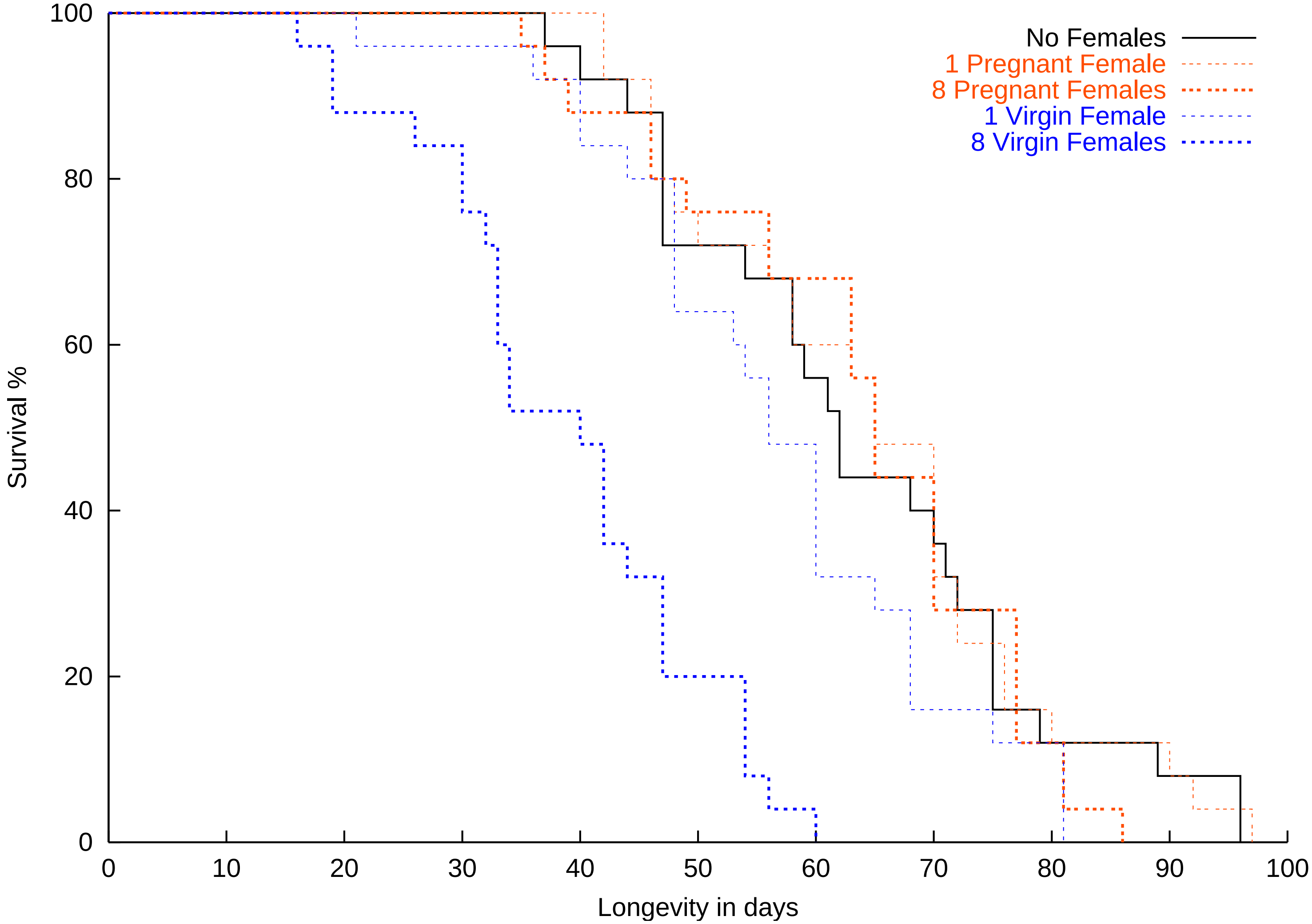
Survival Analysis

- Two sides to survival
  1. Survival to fixed point in time
  2. Length of time to failure

- Focus on length of time
  - Nonparametric
  - Parametric
    - Weibull

Cost of increased reproduction on longevity in male fruitflies

- Dependent variable: longevity of a male fruitfly in days.
- Treatments:
  - No females
  - One virgin female
  - Eight virgin females
  - One newly pregnant female
  - Eight newly pregnant females
- 25 fruitflies on each of the five treatments.


No Females
1 Pregnant Female
8 Pregnant Females
1 Virgin Female
8 Virgin Females
Survival %
100
80
60
40
20
0
0
10
20
30
40
50
60
70
80
90
100
Longevity in days

Survival/Longevity

- Time between two events
  - Stating point: Birth, Diagnosis, Start of Treatment
  - End point: Death, Recurrence of Disease, Cure

- Censoring
  - May only be able to narrow the time of an event to an interval
  - Disease started before diagnosis
  - Death occurred after the patient left the study

- Dependent variable: Failure time T_i
- Linear predictor: Risk Factor γ_i
- Males housed with eight virgin females have earlier failure times and therefore are at increased risk of dying

Survival function

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

- We will suppress the use of γ_i to keep the notation simpler
- It is also one minus the CDF $S(t) = 1 - F(t)$
- Decreasing function that goes from 1 to 0.

- Features:
  - Shape
  - Location:
    Decreasing risk will “stretch” out the survival function
    Effect is the same as scaling a χ^2 random variable by σ^2

Hazard function

- $f_T(t) \leftrightarrow F(t) \leftrightarrow S(t)$
- We define a hazard function and use this as our starting point.

  A hazard function is the short term risk of failure for an individual who hasn't failed at time t divided by the time interval

$$\lambda(t) = \lim_{\Delta \to 0} \frac{\Pr(T_i \leq t + \Delta | T_i > t)}{\Delta}.$$

- $\lambda(t) \leftrightarrow f_T(t) \leftrightarrow F(t) \leftrightarrow S(t)$

$$
\begin{aligned}
\lambda(t) &= \lim_{\Delta \to 0} \frac{S(t) - S(t+\Delta)}{S(t)\Delta} \\
&= \lim_{\Delta \to 0} \frac{F(t+\Delta) - F(t)}{S(t)\Delta} \\
&= \left[\lim_{\Delta \to 0} \frac{F(t+\Delta) - F(t)}{\Delta}\right] \frac{1}{S(t)} \\
&= \frac{f(t)}{S(t)} \\
\Leftrightarrow f(t) &= \lambda(t) S(t)
\end{aligned}
$$

$$S(t) = e^{-\Lambda(t)}$$

$$\Lambda(t) = -\ln[S(t)]$$

Using $S(t) = 1 - F(t)$

$$\frac{\partial \Lambda(t)}{\partial t} = \frac{f(t)}{S(T)}$$

$$= \lambda(t)$$

Using $S(0) = 1 \Leftrightarrow \Lambda(t) = 0$

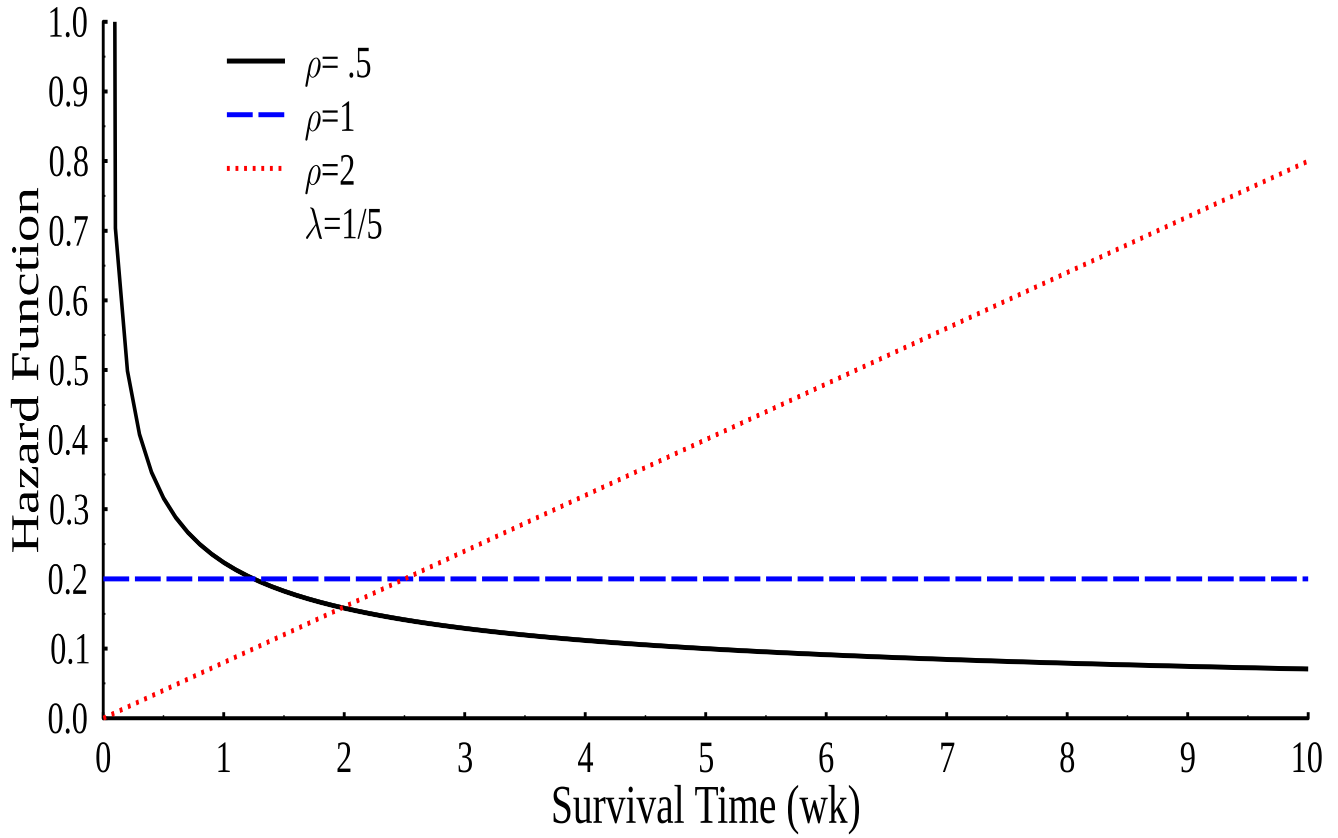
$$\Lambda(t) = \int_0^t \lambda(w) dw$$

- Hazard function are non-negative
- If the hazard function is constant,
  - then the probability of surviving another year is the same for a 20 year old and for a 100 year old
- If the hazard function is increasing,
  - then the probability of surviving another year is the greater for a 20 year old compared to a 100 year old

Weibull

- Weibull the log hazard function is linear in log time

$$\begin{aligned} \ln(\lambda(t;\boldsymbol{\gamma}_i)) &= [\ln(\rho) + (\rho)\ln(\lambda)] + (\rho - 1)\ln(t) \\ \lambda(t;\boldsymbol{\gamma}_i) &= \rho\lambda^{\rho}t^{\rho-1} \\ S(t;\boldsymbol{\gamma}_i) &= e^{-(\lambda t)^{\rho}}. \end{aligned}$$


Hazard Function
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
ρ= .5
ρ=1
ρ=2
λ=1/5
0
1
2
3
4
5
6
7
8
9
10
Survival Time (wk)

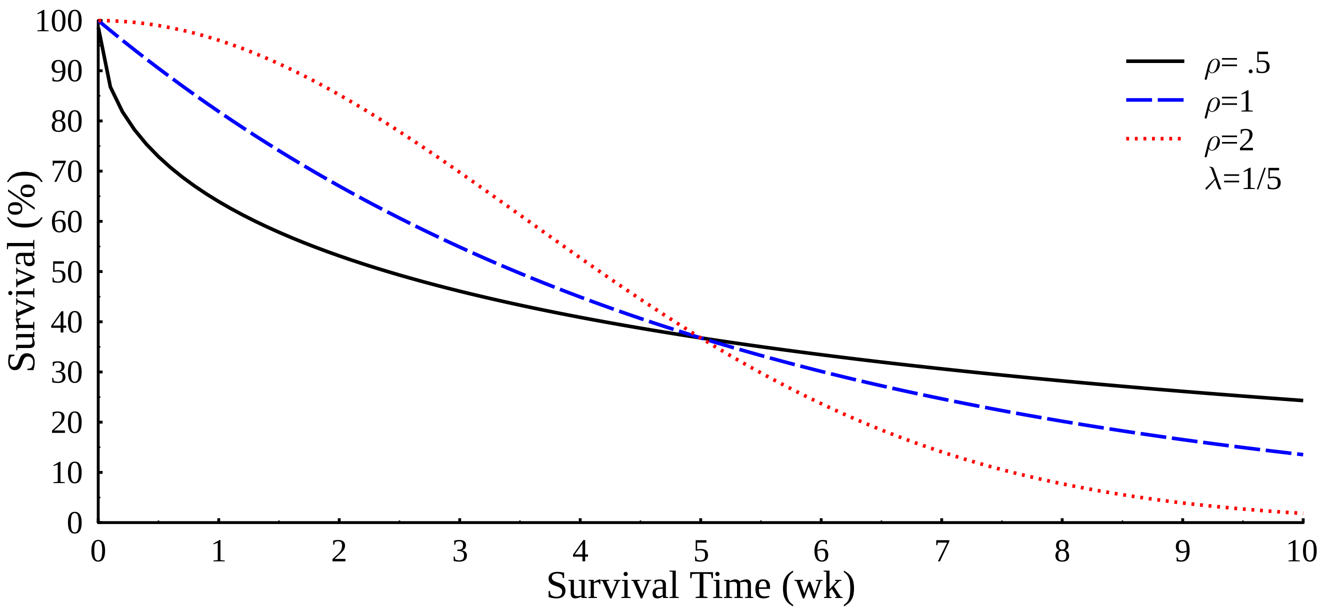
Rate Parameter

- Rate parameter ρ determines whether the hazard function increases or decreases over time.

- $\rho > 1$
  - hazard function increases in time
  - hazard function is zero when time is zero.

- $\rho < 1$
  - hazard function decreases in times
  - hazard function becomes infinite when time approaches zero

Weibull survival function



- $\rho > 1$

  Survival function starts out flat and then steepens

- $\rho < 1$

  Survival curve starts out steep and flattens out

- $1/\lambda$

  Approximately equal to the time when 2/3 of the individual have failed

Proportional Hazard Models

$$\lambda(t; \boldsymbol{\gamma}_i) = \rho\, t^{\rho-1} e^{-\gamma_i}$$

- Proportional hazard function scales a baseline hazard function $\lambda_0(t)$ by a scaling factor $e^{-\gamma}$.
- Not appropriate when effects change over time.

$$\begin{aligned}
\lambda(t; \boldsymbol{\gamma}_i) &= \lambda_0(t) e^{-\gamma} = \rho\, t^{\rho-1} e^{-\gamma} \\
\Lambda(t; \boldsymbol{\gamma}_i) &= \Lambda_0(t) e^{-\gamma} = t^{\rho} e^{-\gamma} \\
S(t; \boldsymbol{\gamma}_i) &= \exp(-\Lambda_0(t) e^{-\gamma}) = \exp(-t^{\rho} e^{-\gamma}).
\end{aligned}$$

Linear γ

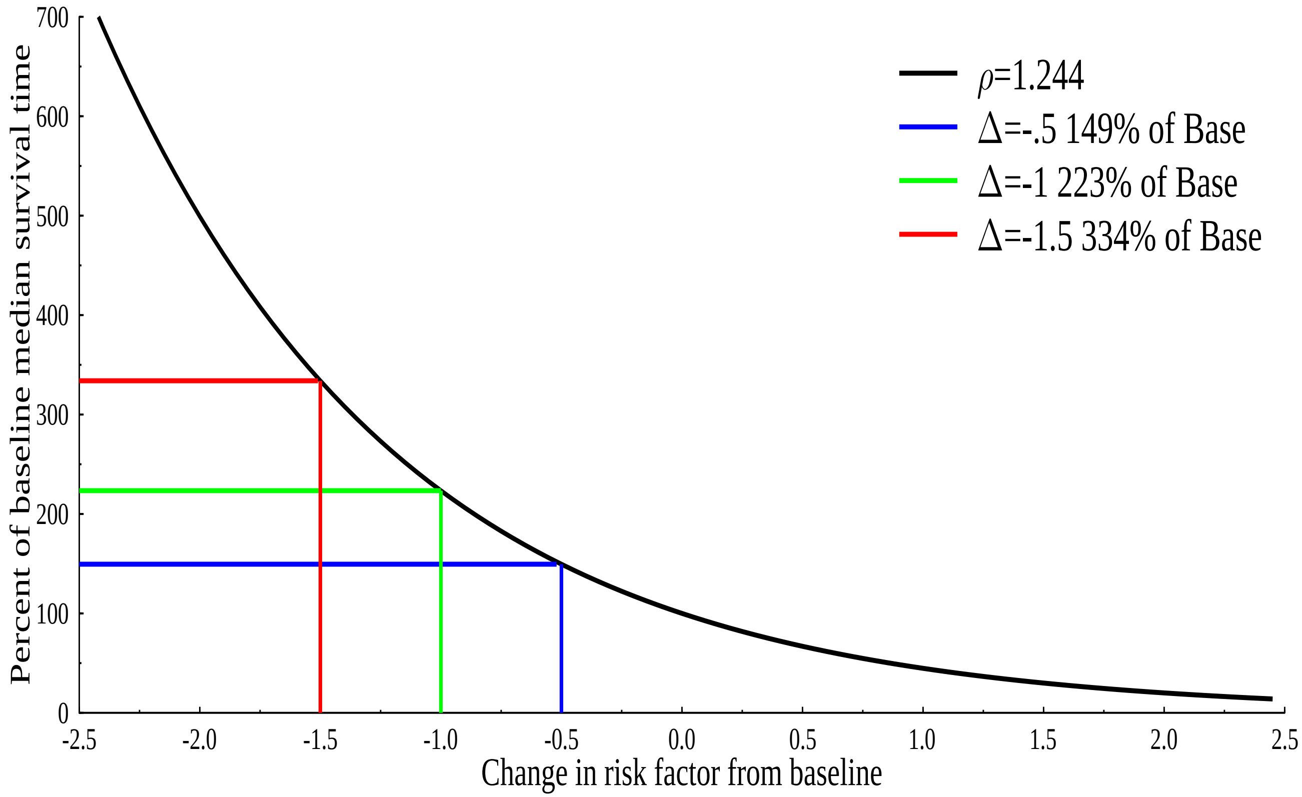
$$\boldsymbol{\gamma} = \boldsymbol{X\beta}$$

- If the rate parameter is known then we can transform the distribution into an exponential distribution.
- Otherwise we will need to find a "new" $\boldsymbol{H}'\boldsymbol{R}^{-1}\boldsymbol{H}$ and $\boldsymbol{H}'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{\mu} + \boldsymbol{H\eta})$.
- Larger γ_i are associated with longer median survival times.

Median Survival Time m_γ

$$m_\gamma = [-\ln(.5)]^{1/\rho} e^{\gamma/\rho}$$
$$m_{\gamma+\Delta} = m_\gamma e^{\Delta/\rho}$$

If the risk factor for one group is Δ more than the risk factor for another group, then their median survival time will be $e^{\Delta/\rho}$ times of the other groups median survival time.



ρ=1.244
Δ=-.5 149% of Base
Δ=-1 223% of Base
Δ=-1.5 334% of Base
Percent of baseline median survival time
Change in risk factor from baseline
0
100
200
300
400
500
600
700
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5

Estimating equations

Starting with the likelihood:

$$
\begin{aligned}
\ell(\boldsymbol{\gamma}, \rho) &= \sum_i \ell_i \\
\ell_i &= \ln(\lambda_0(t_i)) - \gamma_i - \Lambda_0(t_i) \exp(-\gamma_i) \\
\lambda_0(t_i) &= \rho\, t^{\rho-1} \\
\Lambda_0(t_i) &= t_i^{\rho} \\
\ell_i &= \ln(\rho) + (\rho - 1) \ln(t_i) - \gamma_i - {t_i}^{\rho} \exp(-\gamma_i).
\end{aligned}
$$

Next we need the score function:

$$\begin{aligned}\frac{\partial \ell_i}{\partial \gamma_i} &= -1 + {t_i}^\rho \exp(-\gamma_i) \\ \frac{\partial \ell_i}{\partial \rho} &= \frac{1}{\rho} + \ln(t_i)[1 - t_i^\rho \exp(-\gamma_i)].\end{aligned}$$

Followed by the Hessian:

$$
\begin{aligned}
\frac{\partial^2 \ell_i}{\partial \gamma_i \partial \gamma_i} &= -t_i^{\rho} \exp(-\gamma_i) \\
\frac{\partial^2 \ell_i}{\partial \gamma_i \partial \rho} &= -t_i^{\rho} \exp(-\gamma_i) \ln(t_i) \\
\frac{\partial^2 \ell_i}{\partial \rho \partial \rho} &= -\frac{1}{\rho^2} - t_i^{\rho} \exp(-\gamma_i) \ln(t_i)^2
\end{aligned}
$$

Score and Hessian for $\boldsymbol{\beta}$:

$$\begin{aligned}
\boldsymbol{\eta} &= \boldsymbol{X}\boldsymbol{\beta} \\
\frac{\partial \ell}{\partial \boldsymbol{\beta}'} &= \frac{\partial \ell}{\partial \boldsymbol{\eta}'}\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}'} \\
\frac{\partial \ell}{\partial \boldsymbol{\beta}'} &= \frac{\partial \ell}{\partial \boldsymbol{\eta}'}\boldsymbol{X} \\
\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} &= \frac{\partial \left(\boldsymbol{X}'\frac{\partial \ell}{\partial \boldsymbol{\eta}}\right)}{\partial \boldsymbol{\eta}'}\frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}'} \\
&= \boldsymbol{X}'\left[\frac{\partial^2 \ell}{\partial \boldsymbol{\eta}\partial \boldsymbol{\eta}'}\right]\boldsymbol{X}
\end{aligned}$$

Newton-Raphson equations:

$$
\begin{aligned}
(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{y}* \\
\boldsymbol{W} &= \begin{pmatrix} \mathrm{Diag}\left(-\frac{\partial^2 \ell_i}{\partial\gamma_i \partial\gamma_i}\right) & \mathrm{Diag}\left(-\frac{\partial^2 \ell_i}{\partial\gamma_i \partial\rho}\right) \\ \mathrm{Diag}\left(-\frac{\partial^2 \ell_i}{\partial\gamma_i \partial\rho}\right) & \mathrm{Diag}\left(-\frac{\partial^2 \ell_i}{\partial\rho \partial\rho}\right) \end{pmatrix} \\
\boldsymbol{y}* &= \begin{pmatrix} \{_C \frac{\partial \ell_i}{\partial \gamma_i}\} \\ \{_C \frac{\partial \ell_i}{\partial \rho}\} \end{pmatrix} + \boldsymbol{W}\boldsymbol{\eta} \\
\boldsymbol{\eta} &= \begin{pmatrix} \{_C \gamma_i\} \\ \mathbf{1}_N \rho \end{pmatrix}
\end{aligned}
$$

where $\boldsymbol{W}$ plays the same role as $\boldsymbol{H}'\boldsymbol{R}^{-1}\boldsymbol{H}$ and $\boldsymbol{y}^*$ plays the role of $\boldsymbol{H}'\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{\mu} + \boldsymbol{H}\boldsymbol{\eta})$.

Transformation to Exponential

- If $t \sim Weibull(\lambda, \rho)$, then $u = t^{\rho} \sim \mathrm{Exp}(\lambda)$.

- Requires that we know ρ.
  - Start with an initial estimate of ρ
  - Transform the data and calculate the log likelihood of the unadjusted data
  - Repeat

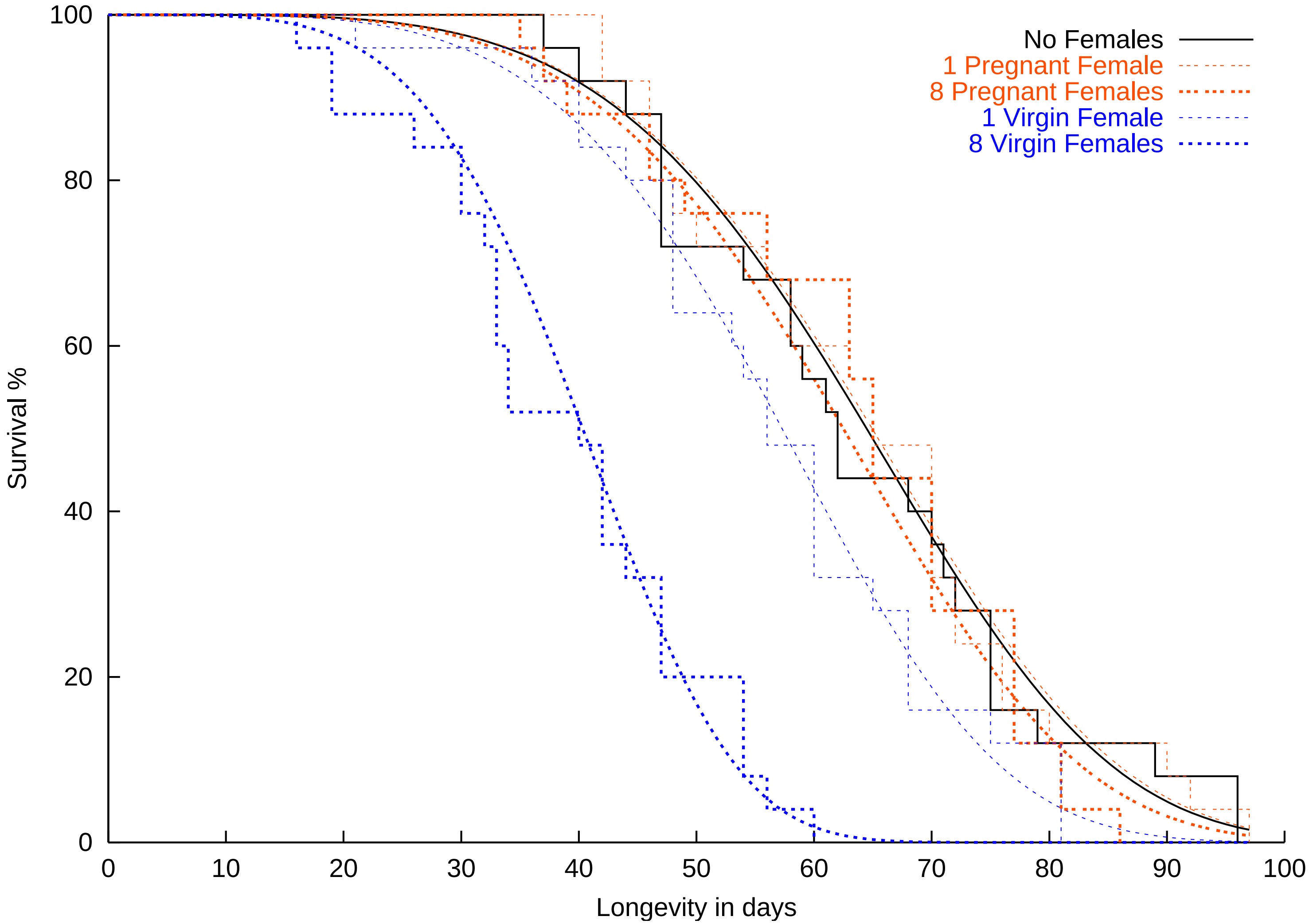
- But first we will assume that ρ is known

```
Survival Analysis
  Id * !I
  Partners * !I
  Type * !I
  Longevity !/100
  Thorax
  Sleep
  Rho !=4.4011
  N != 125 # Number of records
  Trans !=Longevity !^Rho
  RhoAdj !=Rho !^0 !-1 !*N # Log Like = RhoAdj - Deviance/2
fruitfly.dat
Trans !GAMMA !LOGARITHM  ~ mu Partners.Type
0 0 0
predict Partners Type !TDIFF
```

```
Using       125 records of      125 read

 Model term                      Size #miss #zero   MinNon0     Mean      MaxNon0
  1 Id                             25      0      0       1     13.0000         25
  2 Partners                        3      0      0       1      2.0000          3
  3 Type                            3      0      0       1      2.0000          3
  4 Longevity                              0      0 0.1600      0.5744     0.9700
  5 Thorax                                 0      0 0.6400      0.8210     0.9400
  6 Sleep                                  0      0  1.000       23.46      83.00
  7 Rho                                    0      0  4.401       4.401      4.401
  8 N                                      0      0  125.0       125.0      125.0
  9 Trans                    Variate       0      0 0.3142E-03 0.1515     0.8745
 10 RhoAdj                                 0      0  60.23       60.23      60.23
Warning: If RhoAdj is fitted as a covariate, it should be centred first.
 11 mu                              1
 12 Partners.Type                   9  2 Partners   :    3   3 Type           :    3
Forming      10 equations:  10 dense.
```

The 60.23 along with the deviance will be used to calculate the the log likelihood at the $\rho = 4.4011$.

```
Notice: Algebraic ANOVA Denominator DF calculation is not available
        Numerical derivatives will be used.
Distribution and link: Gamma; Log  Mu=exp(XB)=PHI/GAMMA
                                         V=Mu^2/PHI/N=PHI/N/Gamma^2
Warning: The LogL value is unsuitable for comparing GLM models
Notice:      5 singularities detected in design matrix.
  1 LogL=-49.7281     S2= 0.73689        120 df    1.000
        ....
  5 LogL=-67.9807     S2= 0.99889        120 df    1.000
  6 LogL=-67.9807     S2= 0.99889        120 df    1.000
Final parameter values                          1.0000
Deviance from GLM fit             120      143.46←Dev_Exp
Variance heterogenity factor [Deviance/DF]     1.20

Source                Model  terms     Gamma     Component    Comp/SE   % C
Variance                125    120   1.00000      0.998893       7.75   0 U

Analysis of Variance              NumDF      DenDF     F_inc             Prob
 11 mu                                1      120.0    553.84             <.001
 12 Partners.Type                     4      120.0     19.43             <.001
Notice: The DenDF values are calculated ignoring fixed/boundary/singular
            variance parameters using numerical derivatives.
Warning: This Analysis of Variance based on the working variable is not
         equivalent to the Analysis of Deviance.  Standard errors are scaled
         by the variance of the working variable, not the residual deviance.
```



100
80
60
40
20
0
Survival %
0
10
20
30
40
50
60
70
80
90
100
Longevity in days
No Females
1 Pregnant Female
8 Pregnant Females
1 Virgin Female
8 Virgin Females

- Now to find ρ
- Need the correct log likelihood
- Need to try several ρ values

Adjustment Factor

Weibull log likelihood

$$
\begin{aligned}
\ell_i(\gamma, \rho, t_i) &= \ln(\rho) + (\rho - 1)\ln(t_i) - \gamma_i - {t_i}^{\rho}\exp(-\gamma_i) \\
&= [\ln(\rho) + (\rho - 1)\ln(t_i)] - \ln(\lambda_i) - \frac{{t_i}^{\rho}}{\lambda_i}
\end{aligned}
$$

Exponential

$$
\begin{aligned}
\ell_i^*(\lambda; u_i = t_i^{\rho}) &= -\ln(\lambda_i) - \frac{u_i}{\lambda_i} \\
&= -\ln(\lambda_i) - \frac{{t_i}^{\rho}}{\lambda_i}
\end{aligned}
$$

Adjusted log likelihood

$$\ell_{Weib} = \ell_{Exp} + \sum_{i=1}^{n} [\ln(\rho) + (\rho - 1)\ln(t_i)]$$

However, ASREML reports:

```
Warning:  The LogL value is unsuitable for comparing GLM
models
```

So we will use the deviance instead

Deviance

$$
\begin{aligned}
Deviance &= [\ell(\mu = y) - \ell(\mu = \widehat{\mu})] \\
Dev_{\mathsf{Exp}} &= 2\sum_i \left[-\ln(t_i^\rho) - 1 + \ln(\lambda_i) + \frac{{t_i}^\rho}{\lambda_i}\right] \\
\frac{-Dev_{\mathsf{Exp}}}{2} &= \ell_{Exp} + \sum_i \left[\rho \ln(t_i) + 1\right] \\
\ell_{Weib} &= N\left[\ln(\rho) - 1\right] - \frac{Dev_{\mathsf{Exp}}}{2}
\end{aligned}
$$

Doing the search

- Under Linux
- Use a PERL script to run ASREML several times

```
#!/usr/bin/perl

die "Usage: $ARGV[0] <basename> <rho_start> <rho stop> <rho step>$#ARGV\n"
  if $#ARGV < 3;
```

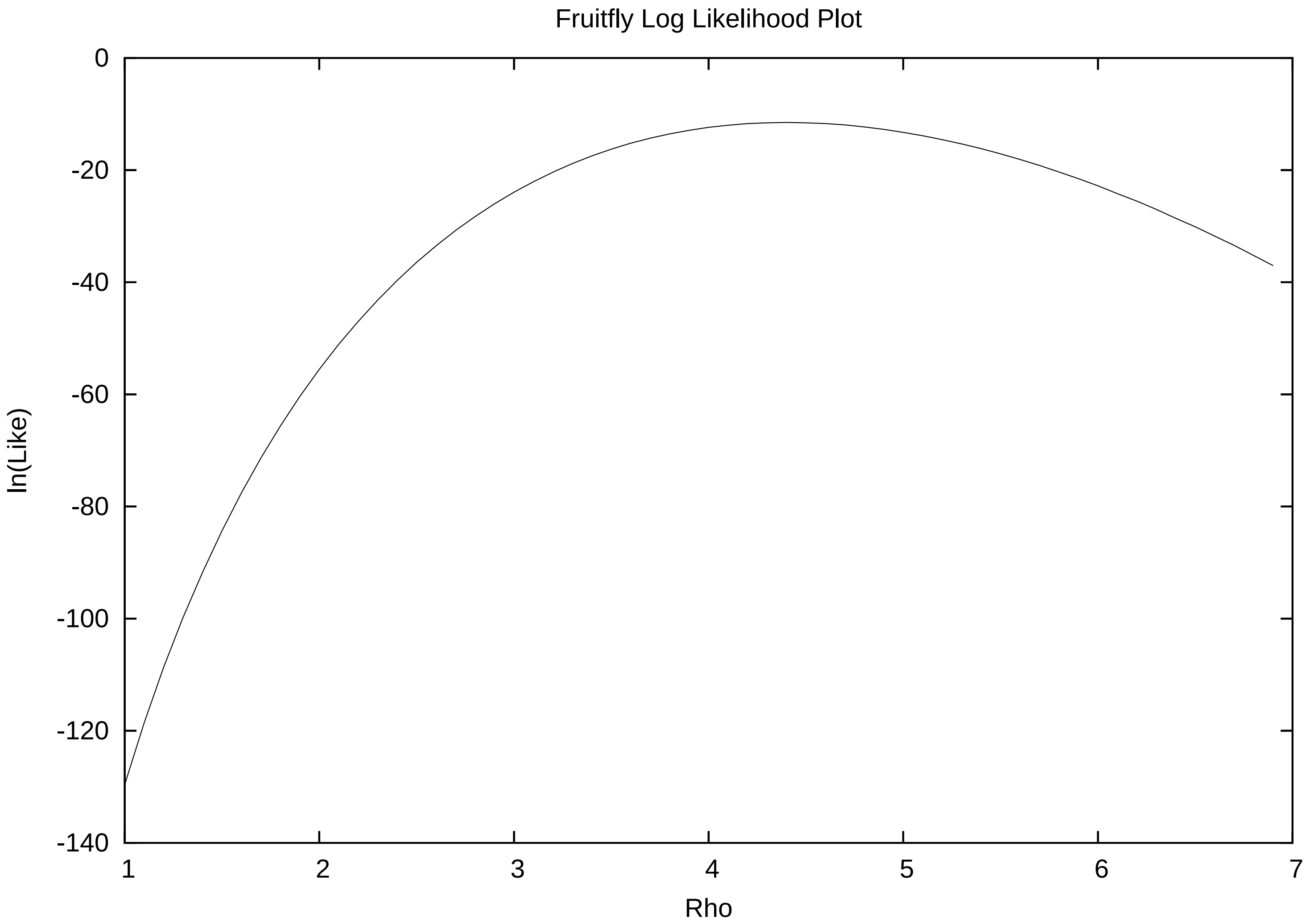
```
$basename=$ARGV[0];
$start=$ARGV[1];
$stop=$ARGV[2];
$rhostep=$ARGV[3];
$script=$basename . ".as";
$script2=$basename . "2.as";
$likedata=$basename . "_like.dat";

open LIKEDATA, ">$likedata" or die "Can't open $likedata\n";
print LIKEDATA  "Rho LogLike RhoAdj Deviance\n";
```

```
for($rho=$start;$rho<=$stop;$rho+=$rhostep){
  open SCRIPT , "< $script" or die "Can't open $script\n";
  open OUTSCRIPT,"> $script2" or die "Can't open $script2\n";

  while(<SCRIPT>){
    s/RHOVAR/$rho/;
    print  OUTSCRIPT $_;
  }

  system("ASReml -n $script2 > log");
```

```
  open RESULTS, "<$script2" . "r" or die "Can't open $script2" . "r/n";
  while(<RESULTS>){
    chomp();
    @parts=split/ +/;
    $rhoadj=$parts[6] if $parts[2]=~/RhoAdj/;
    $dev=$parts[6] if $_=~/Deviance from GLM fit/;
  }

  $loglike=$rhoadj-$dev/2.;
  if($rho==$start | $loglike> $maxlike){
    $maxlike=$loglike ;
    $maxrho=$rho;
  }
  print LIKEDATA "$rho $loglike $rhoadj $dev\n";

}
print "MLE rho=$maxrho with logLike=$maxlike\n";
```

```
$ perl calclike.pl fruit 1 7 .1
MLE rho=4.4 with logLike=-11.5
$ perl calclike.pl fruit 4.3 4.5 .01
MLE rho=4.42 with logLike=-11.495
$ perl calclike.pl fruit 4.41 4.43 .001
MLE rho=4.414 with logLike=-11.495
```

- At which point we are beyond the precision of our results file.



Fruitfly Log Likelihood Plot
0
-20
-40
-60
-80
-100
-120
-140
ln(Like)
1
2
3
4
5
6
7
Rho

```
Rho LogLike RhoAdj Deviance
1.000 -129.585 -125.0 9.17
1.100 -118.61 -113.1 11.02
...
4.000 -12.38 48.29 121.34
4.100 -11.995 51.37 126.73
4.200 -11.71 54.39 132.20
4.300 -11.55 57.33 137.76
4.400 -11.5 60.20 143.40
4.500 -11.55 63.01 149.12
4.600 -11.695 65.76 154.91
4.700 -11.945 68.45 160.79
4.800 -12.29 71.08 166.74
...
6.800 -35.225 114.6 299.65
6.900 -37.05 116.4 306.90
```

$\leftarrow 60.20 - 143.40/2 = -11.50$

Censoring

- Occurs when T_i is known to fall within a interval.
  - Right Censored: $c_i < T_i$ or $T_i \in (c_i, \infty)$
  - Interval Censored: $l_i < T_i < u_i$ or $T_i \in (l_i, u_i)$
  - Left Censored: $T_i < c_i$ or $T_i \in (-\infty, c_i)$

- We will assume that censoring time occurs independently of failure time.

  For example with right censoring: $f_{CT}(c, t; \gamma) = f_C(c) f_T(t; \gamma)$

  Each individual will have a censoring time and a failure time. However, only one of which will be observed.

For individual i we generate a $C_i = c$ and $T_i = t$.

Next we observe a censored observation with a value of c if $t > c$ and an uncensored observation with a value of t otherwise.

An observation is then coded as an observed time value and censor code (e.g. 0=censored and 1=uncensored) pair.

If an observation is uncensored, then we proceed as before.

If an observation is right censored at c, then we know $T_i > c$. The probability that $T_i > c$ is $S(c; \gamma_i, \rho)$.

Likelihood:

- Uncensored: $(t, 1)$

$$
\begin{aligned}
L_i &= \lambda(t; \gamma_i, \rho) e^{-\Lambda(t; \gamma_i, \rho)} \\
\ell_i &= \ln(\lambda(t; \gamma_i, \rho)) - \Lambda(t; \gamma_i, \rho)
\end{aligned}
$$

- Censored: $(t, 0)$

$$
\begin{aligned}
L_i &= \qquad\qquad\quad e^{-\Lambda(t; \gamma_i, \rho)} \\
\ell_i &= \qquad\qquad\qquad\quad - \Lambda(t; \gamma_i, \rho)
\end{aligned}
$$

- Combined: (t, c)

$$\ell_i = c \times \ln(\lambda(t;\gamma_i,\rho)) - \Lambda(t;\gamma_i,\rho)$$

$$\frac{\partial \ell_i}{\partial \gamma_i} = c \times \frac{\partial \ln(\lambda(t;\gamma_i,\rho))}{\partial \gamma_i} - \frac{\partial \Lambda(t;\gamma_i,\rho)}{\partial \gamma_i}$$

$$\frac{\partial^2 \ell_i}{\partial \gamma_i \partial \gamma_i} = c \times \frac{\partial^2 \ln(\lambda(t;\gamma_i,\rho))}{\partial \gamma_i \partial \gamma_i} - \frac{\partial^2 \Lambda(t;\gamma_i,\rho)}{\partial \gamma_i \partial \gamma_i}$$

What happens if all the observations within a group are right censored?

$$\frac{\partial \ell_i}{\partial \gamma_i} = t^\rho e^{\gamma_i}$$
$$> 0$$

The effect of that group will go to ∞.