

```

# Survival Analysis in R
# Copyright 2013 by Ani Katchova

# install.packages("survival")
library(survival)

mydata<- read.csv("C:/Econometrics/Data/survival_unemployment.csv")
attach(mydata)

# Define variables
time <- spell
event <- event
X <- cbind(logwage, ui, age)
group <- ui

# Descriptive statistics
summary(time)
summary(event)
summary(X)
summary(group)

# Kaplan-Meier non-parametric analysis
kmsurvival <- survfit(Surv(time,event) ~ 1)
summary(kmsurvival)
plot(kmsurvival, xlab="Time", ylab="Survival Probability")

# Kaplan-Meier non-parametric analysis by group
kmsurvival1 <- survfit(Surv(time, event) ~ group)
summary(kmsurvival1)
plot(kmsurvival1, xlab="Time", ylab="Survival Probability")

# Nelson-Aalen non-parametric analysis
nasurvival <- survfit(coxph(Surv(time,event)~1), type="aalen")
summary(nasurvival)
plot(nasurvival, xlab="Time", ylab="Survival Probability")

# Cox proportional hazard model - coefficients and hazard rates
coxph <- coxph(Surv(time,event) ~ X, method="breslow")
summary(coxph)

# Exponential, Weibull, and log-logistic parametric model coefficients
# Opposite signs from Stata results, Weibull results differ; same as SAS
exponential <- survreg(Surv(time,event) ~ X, dist="exponential")
summary(exponential)

weibull <- survreg(Surv(time,event) ~ X, dist="weibull")
summary(weibull)

loglogistic <- survreg(Surv(time,event) ~ X, dist="loglogistic")
summary(loglogistic)

```

```

> # Survival Analysis in R
> # Copyright 2013 by Ani Katchova
>
> # install.packages("survival")
> library(survival)
Loading required package: splines
>
> mydata<- read.csv("C:/Econometrics/Data/survival_unemployment.csv")
> attach(mydata)
>
> # Define variables
> time <- spell
> event <- event
> X <- cbind(logwage, ui, age)
> group <- ui
>
> # Descriptive statistics
> summary(time)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  2.000   5.000   6.248   9.000  28.000
> summary(event)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  0.000   0.000   0.321   1.000   1.000
> summary(X)
      logwage          ui          age
Min.   :2.708  Min.   :0.0000  Min.   :20.00
1st Qu.:5.298  1st Qu.:0.0000  1st Qu.:27.00
Median :5.677  Median :1.0000  Median :34.00
Mean   :5.693  Mean   :0.5528  Mean   :35.44
3rd Qu.:6.052  3rd Qu.:1.0000  3rd Qu.:43.00
Max.   :7.600  Max.   :1.0000  Max.   :61.00
> summary(group)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000 0.0000  1.0000  0.5528  1.0000  1.0000
>
> # Kaplan-Meier non-parametric analysis
> kmsurvival <- survfit(Surv(time,event) ~ 1)
> summary(kmsurvival)
Call: survfit(formula = Surv(time, event) ~ 1)

```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	3343	294	0.912	0.00490	0.903	0.922
2	2803	178	0.854	0.00622	0.842	0.866
3	2321	119	0.810	0.00708	0.797	0.824
4	1897	56	0.786	0.00756	0.772	0.801
5	1676	104	0.738	0.00847	0.721	0.754
6	1339	32	0.720	0.00882	0.703	0.737
7	1196	85	0.669	0.00979	0.650	0.688
8	933	15	0.658	0.01001	0.639	0.678
9	848	33	0.632	0.01057	0.612	0.654
10	717	3	0.630	0.01064	0.609	0.651
11	659	26	0.605	0.01128	0.583	0.627
12	556	7	0.597	0.01150	0.575	0.620
13	509	25	0.568	0.01234	0.544	0.593
14	415	30	0.527	0.01353	0.501	0.554

```

15      311      19      0.495 0.01458      0.467      0.524
16      252      10      0.475 0.01527      0.446      0.506
17      201       8      0.456 0.01606      0.426      0.489
18      169       7      0.437 0.01691      0.405      0.472
19      149       4      0.426 0.01744      0.393      0.461
20      130       3      0.416 0.01794      0.382      0.452
21      109       4      0.400 0.01883      0.365      0.439
22       82       4      0.381 0.02029      0.343      0.423
26       48       2      0.365 0.02233      0.324      0.412
27       33       5      0.310 0.02964      0.257      0.374
> plot(kmsurvival, xlab="Time", ylab="Survival Probability")
>
> # Kaplan-Meier non-parametric analysis by group
> kmsurvival1 <- survfit(Surv(time, event) ~ group)
> summary(kmsurvival1)
Call: survfit(formula = Surv(time, event) ~ group)

```

```

              group=0
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1   1495     266   0.822 0.00989   0.803   0.842
  2   1038     116   0.730 0.01191   0.707   0.754
  3    717      55   0.674 0.01317   0.649   0.701
  4    501      20   0.647 0.01396   0.620   0.675
  5    423      36   0.592 0.01550   0.563   0.623
  6    305       8   0.577 0.01603   0.546   0.609
  7    265      28   0.516 0.01801   0.482   0.552
  8    191       4   0.505 0.01842   0.470   0.542
  9    176       5   0.491 0.01898   0.455   0.529
 10    151       1   0.487 0.01913   0.451   0.526
 11    141       6   0.467 0.02010   0.429   0.508
 12    116       1   0.463 0.02033   0.424   0.504
 13    111       5   0.442 0.02144   0.402   0.486
 14     91       9   0.398 0.02376   0.354   0.447
 15     73       3   0.382 0.02459   0.336   0.433
 16     61       3   0.363 0.02566   0.316   0.417
 17     45       4   0.331 0.02799   0.280   0.390
 18     39       3   0.305 0.02944   0.253   0.369
 19     35       1   0.297 0.02986   0.243   0.361
 26     15       1   0.277 0.03379   0.218   0.352
 27     11       1   0.252 0.03897   0.186   0.341

```

```

              group=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1   1848      28   0.985 0.00284   0.979   0.990
  2   1765      62   0.950 0.00511   0.940   0.960
  3   1604      64   0.912 0.00676   0.899   0.926
  4   1396      36   0.889 0.00764   0.874   0.904
  5   1253      68   0.841 0.00919   0.823   0.859
  6   1034      24   0.821 0.00980   0.802   0.841
  7    931      57   0.771 0.01124   0.749   0.793
  8    742      11   0.759 0.01159   0.737   0.782
  9    672      28   0.728 0.01255   0.704   0.753
 10    566       2   0.725 0.01264   0.701   0.750
 11    518      20   0.697 0.01362   0.671   0.724
 12    440       6   0.688 0.01397   0.661   0.716

```

13	398	20	0.653	0.01526	0.624	0.684
14	324	21	0.611	0.01683	0.579	0.645
15	238	16	0.570	0.01857	0.534	0.607
16	191	7	0.549	0.01949	0.512	0.588
17	156	4	0.535	0.02022	0.497	0.576
18	130	4	0.518	0.02121	0.478	0.562
19	114	3	0.505	0.02207	0.463	0.550
20	99	3	0.489	0.02310	0.446	0.537
21	81	4	0.465	0.02492	0.419	0.517
22	61	4	0.435	0.02756	0.384	0.492
26	33	1	0.422	0.02970	0.367	0.484
27	22	4	0.345	0.04233	0.271	0.439

```
> plot(kmsurvival1, xlab="Time", ylab="Survival Probability")
>
> # Nelson-Aalen non-parametric analysis
> nasurvival <- survfit(coxph(Surv(time,event)~1), type="aalen")
> summary(nasurvival)
Call: survfit(formula = coxph(Surv(time, event) ~ 1), type = "aalen")
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	3343	294	0.916	0.00470	0.907	0.925
2	2803	178	0.859	0.00601	0.848	0.871
3	2321	119	0.817	0.00688	0.803	0.830
4	1897	56	0.793	0.00738	0.778	0.807
5	1676	104	0.745	0.00828	0.729	0.761
6	1339	32	0.727	0.00865	0.711	0.745
7	1196	85	0.678	0.00960	0.659	0.697
8	933	15	0.667	0.00985	0.648	0.686
9	848	33	0.641	0.01042	0.621	0.662
10	717	3	0.639	0.01049	0.618	0.660
11	659	26	0.614	0.01115	0.592	0.636
12	556	7	0.606	0.01138	0.584	0.629
13	509	25	0.577	0.01223	0.554	0.602
14	415	30	0.537	0.01340	0.511	0.564
15	311	19	0.505	0.01446	0.478	0.534
16	252	10	0.485	0.01517	0.457	0.516
17	201	8	0.467	0.01599	0.436	0.499
18	169	7	0.448	0.01687	0.416	0.482
19	149	4	0.436	0.01743	0.403	0.471
20	130	3	0.426	0.01795	0.392	0.462
21	109	4	0.410	0.01887	0.375	0.449
22	82	4	0.391	0.02035	0.353	0.433
26	48	2	0.375	0.02243	0.333	0.422
27	33	5	0.322	0.02912	0.270	0.385

```
> plot(nasurvival, xlab="Time", ylab="Survival Probability")
>
>
> # Cox proportional hazard model - coefficients and hazard rates
> coxph <- coxph(Surv(time,event) ~ X, method="breslow")
> summary(coxph)
Call:
coxph(formula = Surv(time, event) ~ X, method = "breslow")
```

n= 3343, number of events= 1073

	coef	exp(coef)	se(coef)	z	Pr(> z)	
Xlogwage	0.461553	1.586535	0.057190	8.070	6.66e-16	***
Xui	-0.979578	0.375470	0.063979	-15.311	< 2e-16	***
Xage	-0.010850	0.989209	0.003132	-3.465	0.000531	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Xlogwage	1.5865	0.6303	1.4183	1.7747
Xui	0.3755	2.6633	0.3312	0.4256
Xage	0.9892	1.0109	0.9832	0.9953

Concordance= 0.693 (se = 0.011)
Rsquare= 0.081 (max possible= 0.992)
Likelihood ratio test= 281.5 on 3 df, p=0
Wald test = 286.3 on 3 df, p=0
Score (logrank) test = 300 on 3 df, p=0

```
>
>
> # Exponential, Weibull, and log-logistic parametric model coefficients
> # Opposite signs from Stata results, Weibull results differ; same as SAS
> exponential <- survreg(Surv(time,event) ~ X, dist="exponential")
> summary(exponential)
```

Call:
survreg(formula = Surv(time, event) ~ X, dist = "exponential")

	Value	Std. Error	z	p
(Intercept)	4.6426	0.30841	15.05	3.27e-51
Xlogwage	-0.4810	0.05678	-8.47	2.43e-17
Xui	1.0775	0.06269	17.19	3.29e-66
Xage	0.0126	0.00312	4.05	5.18e-05

Scale fixed at 1

Exponential distribution
Loglik(model)= -4083.8 Loglik(intercept only)= -4258.4
Chisq= 349.26 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 3343

```
>
> weibull <- survreg(Surv(time,event) ~ X, dist="weibull")
> summary(weibull)
```

Call:
survreg(formula = Surv(time, event) ~ X, dist = "weibull")

	Value	Std. Error	z	p
(Intercept)	4.4784	0.29145	15.37	2.78e-53
Xlogwage	-0.4567	0.05343	-8.55	1.26e-17
Xui	1.0352	0.06014	17.21	2.15e-66
Xage	0.0125	0.00292	4.28	1.90e-05
Log(scale)	-0.0695	0.02328	-2.99	2.81e-03

Scale= 0.933

Weibull distribution

Loglik(model)= -4079.5 Loglik(intercept only)= -4258.2

Chisq= 357.57 on 3 degrees of freedom, p= 0

Number of Newton-Raphson Iterations: 5

n= 3343

>

> loglogistic <- survreg(Surv(time,event) ~ X, dist="loglogistic")

> summary(loglogistic)

Call:

survreg(formula = Surv(time, event) ~ X, dist = "loglogistic")

	Value	Std. Error	z	p
(Intercept)	4.0408	0.31258	12.93	3.16e-38
Xlogwage	-0.4622	0.05653	-8.18	2.94e-16
Xui	1.2099	0.05939	20.37	3.07e-92
Xage	0.0106	0.00291	3.64	2.77e-04
Log(scale)	-0.3063	0.02437	-12.57	3.20e-36

Scale= 0.736

Log logistic distribution

Loglik(model)= -4014.1 Loglik(intercept only)= -4232

Chisq= 435.7 on 3 degrees of freedom, p= 0

Number of Newton-Raphson Iterations: 4

n= 3343



