

Example 1

Estimation for exponential and Weibull distributions

```
#example.1.r -- Exponential and Weibull distributions (lectures 5 and 6)
#library(survival)
```

```
#data
```

```
t <- c(1.2,5.0,0.3,3.0,1.3,0.9,7.2,2.3,3.4,
2.7,2.8,1.6,1.1, 1.1,0.7,3.9, + 1.7,7.3,4.5,
7.5,1.2,0.9,0.6,0.2,2.1,2.1,5.0, 4.0,0.8,5.0,
0.5,1.8,3.6,0.1,7.9,4.2,0.1,3.4,0.4,3.6)
```

```
cc <- c(0,0,0,1,1,1,0,1,0,1,1,0,1,0,1,0,0,
0,1,0,0,0,0,0,1,0,1,0,0,1,0,0,0,0,1,1,0,1,0,1)
```

```
n <- length(t) #sample size
```

```
d <- sum(cc) #number of complete observations
```

```
#exponential estimates
```

```
mean <- sum(t)/d
```

```
l <- 1/mean
```

```
median <- mean*log(2)
```

```
round(cbind(n,d,l,mean,median),3)
```

```
      n  d    l  mean  median
[1,] 40 16 0.150 6.688   4.635
```

```
#Kaplan-Meier survival curve
```

```
fit <- survfit(Surv(t,cc)~1)
```

```
summary(fit)
```

```
Call: survfit(formula=Surv(t,cc)~1)
```

time	n.risk	n.event	survival	std.err	lower	upper
0.7	33	1	0.970	0.0298	0.913	1.000
0.9	31	1	0.938	0.0422	0.859	1.000
1.1	29	1	0.906	0.0517	0.810	1.000
1.3	25	1	0.870	0.0610	0.758	0.998
2.1	21	1	0.828	0.0708	0.701	0.979
2.3	19	1	0.785	0.0794	0.644	0.957
2.7	18	1	0.741	0.0861	0.590	0.931
2.8	17	1	0.698	0.0914	0.540	0.902
3.0	16	1	0.654	0.0955	0.491	0.871
3.4	15	1	0.610	0.0986	0.445	0.838
3.6	13	1	0.563	0.1016	0.396	0.802
4.2	9	1	0.501	0.1079	0.328	0.764
4.5	8	1	0.438	0.1111	0.267	0.720
5.0	7	2	0.313	0.1091	0.158	0.620
7.9	1	1	0.000	NA	NA	NA

```
plot(fit,xlab="time",ylab="survival probabilities", main="Example data")
```

```
#exponential model estimates
```

```
fit <- survreg(Surv(t,cc)~1, dist="exponential")
```

```
summary(fit)
```

```
Call: survreg(formula=Surv(t,cc)~1, dist="exponential")
```

	Value	Std. Error	z	p
(Intercept)	1.9	0.25	7.6	2.94e-14

```
Scale fixed at 1
```

```
Exponential distribution
```

```
Loglik(model)= -46.4 Loglik(intercept only)= -46.4
```

```
Number of Newton-Raphson Iterations: 5
```

```
n= 40
```

```
b <- fit$coefficients
L <- -2*fit$loglik[2]
mean <- exp(b)
l <- 1/mean
median <- log(2)/l
```

```
> round(cbind(n,d,l,mean,median,L),3)
      n d      l    mean  median      L
[1,] 40 16 0.150  6.688   4.635  92.808
```

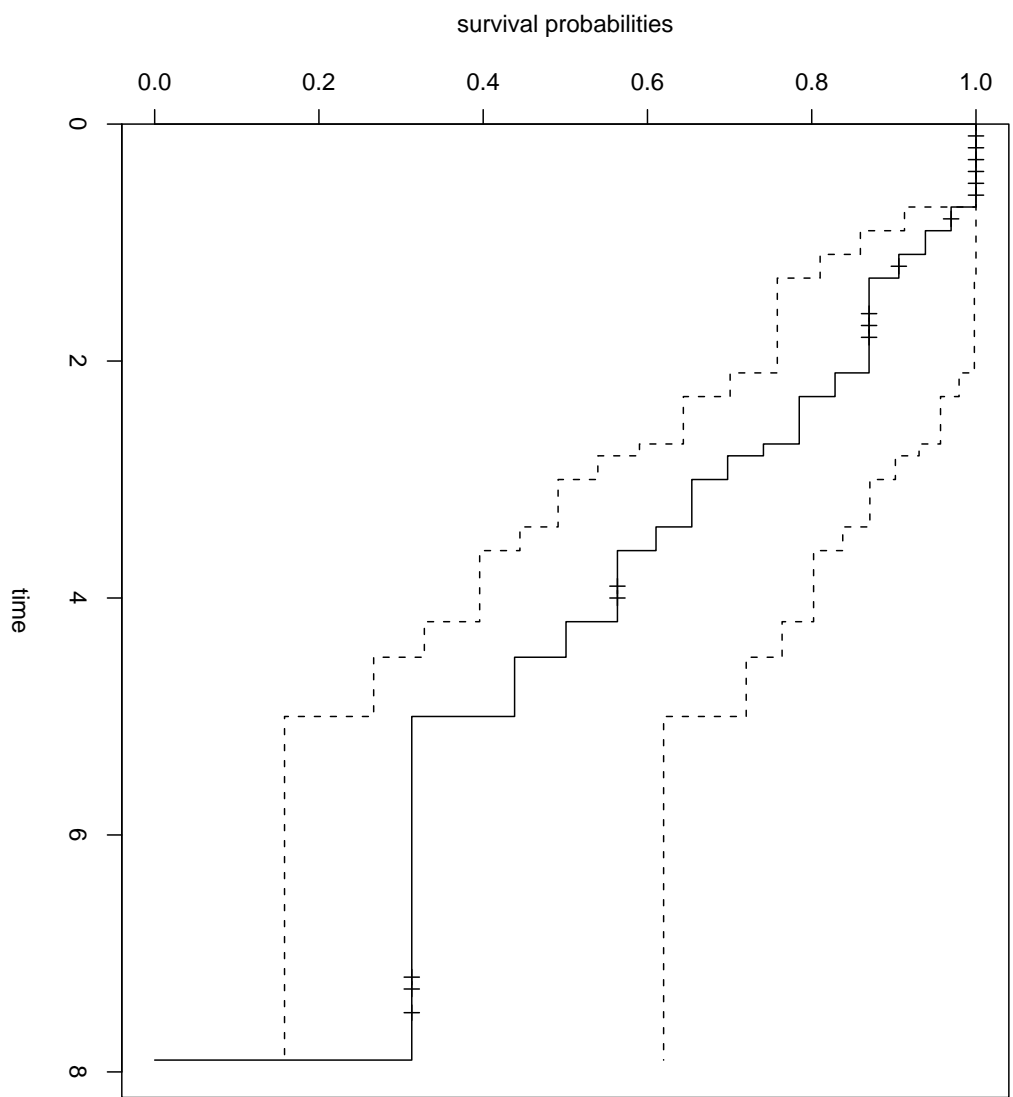
```
#Weibull model estimates
fit <- survreg(Surv(t,cc)~1, dist="weibull")
summary(fit)
```

```
Call: survreg(formula=Surv(t,cc)~1, dist="weibull")
              Value Std. Error      z      p
(Intercept)  1.714      0.152  11.30 1.25e-29
Log(scale)  -0.521      0.189  -2.76 5.78e-03
Scale= 0.594
```

```
Weibull distribution
Loglik(model)= -43.4  Loglik(intercept only)= -43.4
Number of Newton-Raphson Iterations: 8
n= 40
```

```
b <- fit$coefficients
L <- -2*fit$loglik[2]
s <- fit$scale
l <- exp(-b/s)
g <- 1/s
mean <- l^(-1/g)*gamma(1+1/g)
median <- (log(2)/l)^(1/g)
```

```
round(cbind(n,d,l,g,mean,median,L),3)
      n d      l      g    mean  median      L
(Intercept) 40 16 0.056 1.683 6.688   4.635  86.745
```



Example data

Example 2

Estimation for exponential and Weibull two-sample models

```
#example.2.r -- Two sample problem (lecture 7)
#library(survival)

#data
t <- c(1.2,5.0,0.3,3.0,1.3,0.9,7.2,2.3,3.4,
2.7,2.8,1.6,1.1, 1.1,0.7,3.9, + 1.7,7.3,4.5,
7.5,1.2,0.9,0.6,0.2,2.1,2.1,5.0, 4.0,0.8,5.0,
0.5,1.8,3.6,0.1,7.9,4.2,0.1,3.4,0.4,3.6)

cc <- c(0,0,0,1,1,1,0,1,0,1,1,0,1,0,1,0,0,
0,1,0,0,0,0,0,1,0,1,0,0,1,0,0,0,0,1,1,0,1,0,1)

sex <- c(1,1,0,1,0,1,0,0,1,0,1,0,0,1,1,0,0,0,1,
0,1,0,1,1,0,0,1,0,1,0,0,1,0,1,0,0,1,1,0,1)

n <- length(t) #sample size
d <- sum(cc) #number of complete observations

#Kaplan-Meier survival curve
fit <- survfit(Surv(t,cc)~sex)
summary(fit)

Call: survfit(formula=Surv(t,cc)~ sex)
```

sex=0

time	n.risk	n.event	survival	std.err	lower	upper
1.1	17	1	0.941	0.0571	0.836	1.000
1.3	16	1	0.882	0.0781	0.742	1.000
2.1	13	1	0.814	0.0972	0.645	1.000
2.3	11	1	0.740	0.1131	0.549	0.999
2.7	10	1	0.666	0.1237	0.463	0.959
4.2	6	1	0.555	0.1446	0.333	0.925
5.0	5	1	0.444	0.1525	0.227	0.871
7.9	1	1	0.000	NA	NA	NA

sex=1

time	n.risk	n.event	survival	std.err	lower	upper
0.7	15	1	0.933	0.0644	0.815	1.000
0.9	13	1	0.862	0.0911	0.700	1.000
2.8	8	1	0.754	0.1284	0.534	1.000
3.0	7	1	0.646	0.1485	0.412	1.000
3.4	6	1	0.538	0.1581	0.303	0.957
3.6	4	1	0.404	0.1663	0.180	0.905
4.5	3	1	0.269	0.1561	0.086	0.839
5.0	2	1	0.135	0.1231	0.022	0.808

```
plot(fi t,xlab="time",ylab="survival probabilities",lty=1:2,main="Two-sample data")
text(3,0.4,"Male",cex=1.5)
text(6,0.55,"Female",cex=1.5)
```

```
#exponential estimates
#females
t0 <- t[sex==0]
c0 <- cc[sex==0]
n0 <- length(t0)
d0 <- sum(c0)
mean0 <- sum(t0)/d0
l0 <- 1/mean0
median0 <- mean0*log(2)
round(cbind(n0,d0,l0,mean0,median0),3)
      n0 d0    l0  mean0  median0
[1,] 21  8 0.118  8.450   5.8571
```

```
#males
t1 <- t[sex==1]
c1 <- cc[sex==1]
n1 <- length(t1)
d1 <- sum(c1)
mean1 <- sum(t1)/d1
l1 <- 1/mean1
median1 <- mean1*log(2)
round(cbind(n1,d1,l1,mean1,median1),3)
      n1 d1    l1  mean1  median1
[1,] 19  8 0.203   4.93    3.414

#logrank test
survdif(Surv(t,cc)~sex)

Call: survdif(formula=Surv(t,cc)~sex)

      N Observed Expected (O-E)^2/E (O-E)^2/V
sex=0 21         8    10.21    0.478    1.42
sex=1 19         8     5.79    0.844    1.42

Chisq= 1.4  on 1 degrees of freedom, p= 0.233

#Cox model score test is identical when no tied values occur
fit <- coxph(Surv(t,cc)~sex)
summary(fit)

Call: coxph(formula = Surv(t, cc) ~ sex)
      n= 40

      coef exp(coef) se(coef)      z      p
sex 0.619      1.86    0.522 1.18 0.24

      exp(coef) exp(-coef) lower .95 upper .95
sex      1.86      0.539    0.667    5.17

Rsquare= 0.034 (max possible= 0.873 )
Likelihood ratio test= 1.4  on 1 df,  p=0.236
```

```
Wald test          = 1.4  on 1 df,    p=0.236
Score (logrank) test = 1.45 on 1 df,    p=0.229
```

```
#exponential model estimates
```

```
fit <- survreg(Surv(t,cc)~ sex, dist="exponential")
summary(fit)
```

```
Call: survreg(formula=Surv(t,cc)~sex, dist="exponential")
              Value Std. Error      z      p
(Intercept)  2.13      0.354  6.04 1.58e-09
sex          -0.54      0.500 -1.08 2.80e-01
Scale fixed at 1
```

```
Exponential distribution
```

```
Loglik(model)= -45.8  Loglik(intercept only)= -46.4
  Chisq= 1.15 on 1 degrees of freedom, p= 0.28
Number of Newton-Raphson Iterations: 5
```

```
b <- fit$coefficients
L <- -2*fit$loglik[2]
mean0 <- exp(b[1])
ratio <- exp(b[2])
mean1 <- mean0*ratio
l0 <- 1/mean0
l1 <- l0/ratio
mean0 <- log(2)/l0
mean1 <- log(2)/l1
```

```
round(cbind(l0,l1,mean0,mean1,median0,median1,ratio,L),3)
              l0    l1 mean0 mean1 median0 median1 ratio    L
(Intercept) 0.118 0.203 5.857 3.414  5.857  3.414  0.583 91.656
```

```
#Weibull model estimates
```

```
fit <- survreg(Surv(t,cc)~ sex, dist="weibull")
summary(fit)
```

```
Call: survreg(formula=Surv(t,c)~sex, dist="weibull")
```

	Value	Std. Error	z	p
(Intercept)	1.891	0.196	9.64	5.30e-22
sex	-0.477	0.273	-1.75	8.08e-02
Log(scale)	-0.605	0.194	-3.12	1.83e-03

Scale= 0.546

Weibull distribution

Loglik(model)= -42 Loglik(intercept only)= -43.4

Chisq= 2.75 on 1 degrees of freedom, p= 0.097

Number of Newton-Raphson Iterations: 7

n= 40

```
s <- fi t$scale
```

```
b <- -fi t$coefficients/s
```

```
L <- -2*fi t$loglik[2]
```

```
l0 <- exp(b[1])
```

```
l1 <- exp(b[1]+b[2])
```

```
g <- 1/s
```

```
mean0 <- l0^ (-1/g)*gamma(1+1/g)
```

```
mean1 <- l1^ (-1/g)*gamma(1+1/g)
```

```
median0 <- (log(2)/l0)^ (1/g)
```

```
median1 <- (log(2)/l1)^ (1/g)
```

```
round(cbind(n,d,l0,l1,g,mean0,mean1,median0,median1,L),3)
```

	n	d	l0	l1	g	mean0	mean1	median0	median1	L
(Intercept)	40	16	0.0313	0.075	1.832	5.886	3.653	5.423	3.366	83.993

```
#comparison of survival functions
```

```
fi t <- survfi t(Surv(t,cc)~ sex)
```

```
summary(fi t)
```

```
Call: survfit(formula=Surv(t,cc)~sex)
```

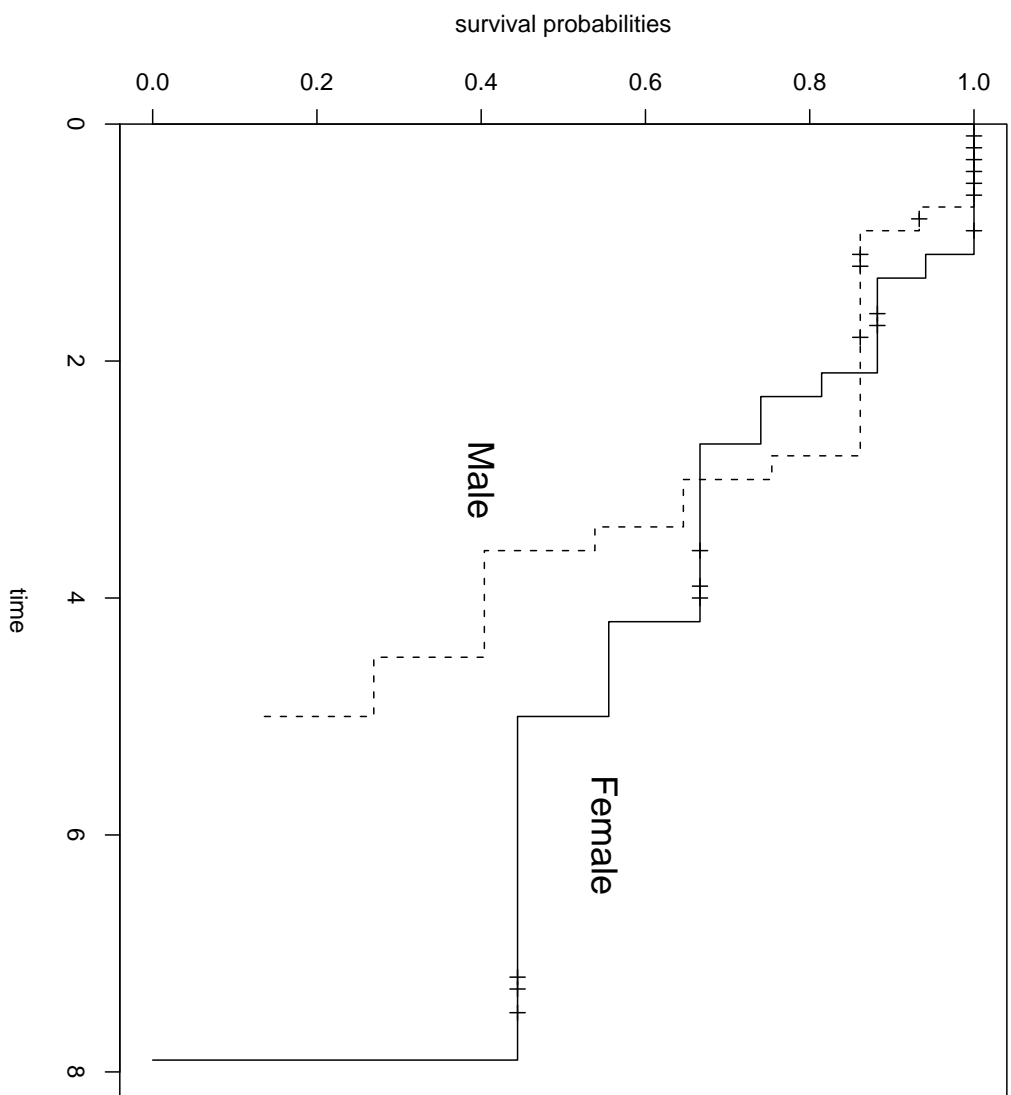
```
sex=0
```

time	n.risk	n.event	survival	std.err	lower	upper
1.1	17	1	0.941	0.0571	0.836	1.000
1.3	16	1	0.882	0.0781	0.742	1.000
2.1	13	1	0.814	0.0972	0.645	1.000
2.3	11	1	0.740	0.1131	0.549	0.999
2.7	10	1	0.666	0.1237	0.463	0.959
4.2	6	1	0.555	0.1446	0.333	0.925
5.0	5	1	0.444	0.1525	0.227	0.871
7.9	1	1	0.000	NA	NA	NA

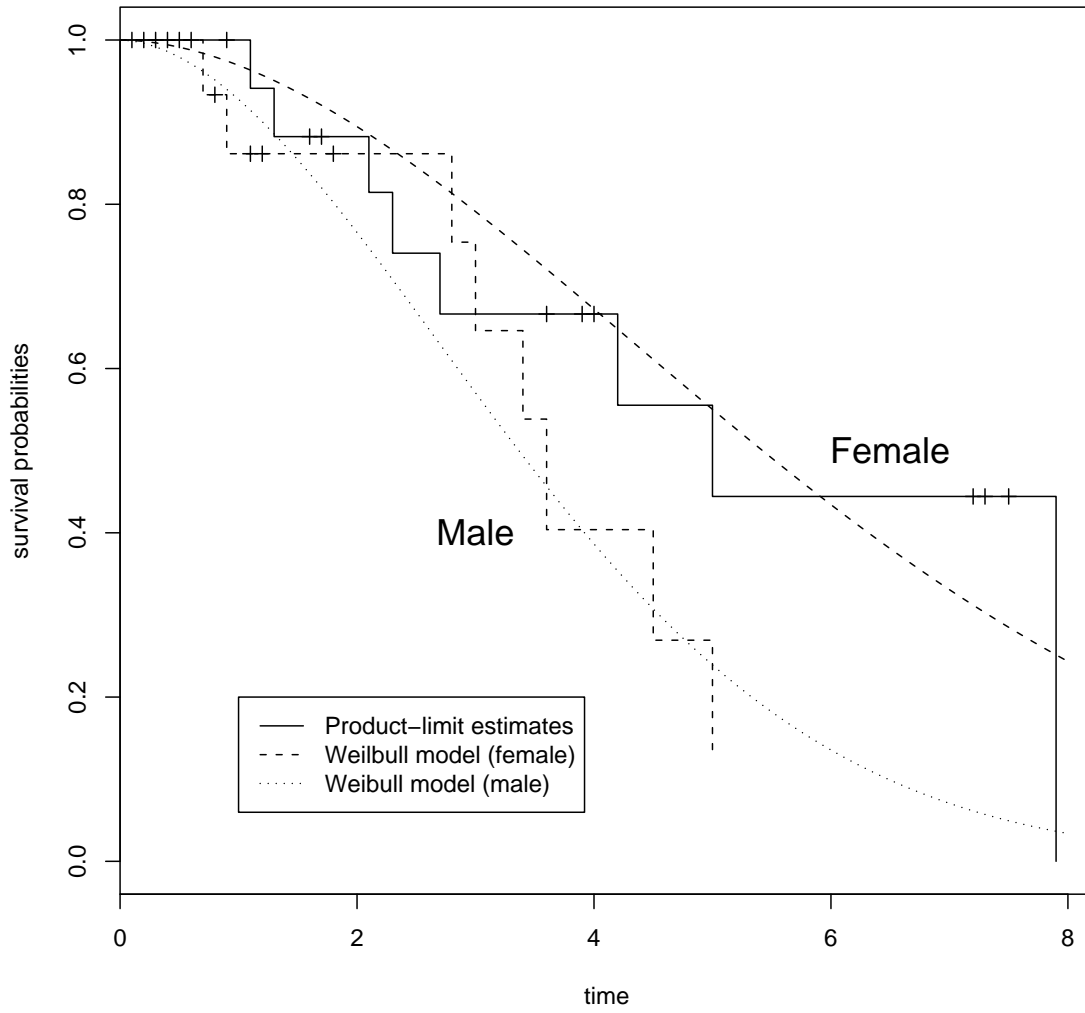
```
sex=1
```

time	n.risk	n.event	survival	std.err	lower	upper
0.7	15	1	0.933	0.0644	0.815	1.000
0.9	13	1	0.862	0.0911	0.700	1.000
2.8	8	1	0.754	0.1284	0.539	1.000
3.0	7	1	0.646	0.1485	0.413	1.000
3.4	6	1	0.538	0.1581	0.303	0.957
3.6	4	1	0.404	0.1663	0.180	0.905
4.5	3	1	0.269	0.1561	0.086	0.839
5.0	2	1	0.135	0.1231	0.022	0.808

```
#plot  
plot(fi t,xlab="time",lty=1:2,ylab="survival probabilities",main="Two-sample data")  
T <- seq(0,8,0.1)  
S0 <- exp(-10*T^ g)  
lines(T,S0,lty=2)  
S1 <- exp(-11*T^ g)  
lines(T,S1,lty=3)  
text(3,0.4,"Male",cex=1.5)  
text(6.5,0.5,"Female",cex=1.5)  
label <- c("Product-limit estimates","Weibull model (female)","Weibull model (male)")  
legend(1,0.2,label,lty=1:3)
```



Two-sample data -- by sex



Example 3

Estimation for general proportional hazards models

```
#example.3.r -- General proportional hazards model (lectures 8 and 9)
#library(survival)

#data
t <- c(1.2,5.0,0.3,3.0,1.3,0.9,7.2,2.3,3.4,
2.7,2.8,1.6,1.1, 1.1,0.7,3.9, + 1.7,7.3,4.5,
7.5,1.2,0.9,0.6,0.2,2.1,2.1,5.0, 4.0,0.8,5.0,
0.5,1.8,3.6,0.1,7.9,4.2,0.1,3.4,0.4,3.6)

cc <- c(0,0,0,1,1,1,0,1,0,1,1,0,1,0,1,0,0,
0,1,0,0,0,0,0,1,0,1,0,0,1,0,0,0,0,1,1,0,1,0,1)

sex <- c(1,1,0,1,0,1,0,0,1,0,1,0,0,1,1,0,0,0,1,
0,1,0,1,1,0,0,1,0,1,0,0,1,0,1,0,0,1,1,0,1)

age <- c(40,48,33,42,41,46,53,57,55,48,33,59,44,
52,51,40,33,37,41,41,46,47,46,41,44,53,37,38,
38,47,54,44,42,51,35,43,53,48,37,52)

n <- length(t)   #sample size
d <- sum(cc)     #number of complete observations

#Kaplan-Meier survival functions
fi t <- survfi t(Surv(t,cc)~ sex)
summary(fi t)
```

```
Call: survfit(formula = Surv(t, cc) ~ sex)
```

```
sex=0
```

time	n.risk	n.event	survival	std.err	lower	upper
1.1	17	1	0.941	0.0571	0.836	1.000
1.3	16	1	0.882	0.0781	0.742	1.000
2.1	13	1	0.814	0.0972	0.645	1.000
2.3	11	1	0.740	0.1131	0.549	0.999
2.7	10	1	0.666	0.1237	0.463	0.959
4.2	6	1	0.555	0.1446	0.333	0.925
5.0	5	1	0.444	0.1525	0.227	0.871
7.9	1	1	0.000	NA	NA	NA

```
sex=1
```

time	n.risk	n.event	survival	std.err	lower	upper
0.7	15	1	0.933	0.0644	0.815	1.000
0.9	13	1	0.862	0.0911	0.700	1.000
2.8	8	1	0.754	0.1284	0.540	1.000
3.0	7	1	0.646	0.1485	0.412	1.000
3.4	6	1	0.538	0.1581	0.303	0.957
3.6	4	1	0.404	0.1663	0.180	0.905
4.5	3	1	0.269	0.1561	0.086	0.839
5.0	2	1	0.135	0.1231	0.022	0.808

```
plot(fi t,xlab="time",ylab="survival probabilities",lty=1:2,main="Two-sample data")
```

```
text(6,0.65,"Female",cex=1.5)
```

```
text(3,4,"Male",cex=1.5)
```

```
#age 45 and older
```

```
fi t <- survfi t(Surv(t[age>=45],cc[age>=45])~ sex[age>=45])
```

```
#age less than 45
```

```
fi t <- survfi t(Surv(t[age<45],cc[age<45])~ sex[age<45])
```

```
summary(fi t)
```

```
Call: survfit(formula=Surv(t[age<45],cc[age<45])~sex[age<45])
```

```
sex[age < 45]=0
time n.risk n.event survival std.err lower upper
1.1 11 1 0.909 0.0867 0.754 1.000
1.3 10 1 0.818 0.1163 0.619 1.000
2.1 8 1 0.716 0.1397 0.488 1.000
4.2 4 1 0.537 0.1871 0.271 1.000
7.9 1 1 0.000 NA NA NA
```

```
sex[age < 45]=1
time n.risk n.event survival std.err lower upper
2.8 4 1 0.75 0.217 0.426 1.000
3.0 3 1 0.50 0.250 0.188 1.000
4.5 2 1 0.25 0.217 0.046 1.000
5.0 1 1 0.00 NA NA NA
```

```
plot(fi t,xlab="time",ylab="survival probabilities",lty=1:2,main="Two-sample data")
```

```
text(6,0.9,"age less than 45",cex=1.5)
```

```
text(6,0.65,"Female",cex=1.5)
```

```
text(3.5,.4,"Male",cex=1.5)
```

```
#age 45 and older
```

```
fi t <- survfit(Surv(t[age>=45],cc[age>=45])~ sex[age>=45])
```

```
summary(fi t)
```

```
Call: survfit(formula=Surv(t[age>=45],cc[age>=45])~sex[age>=45])
```

```
sex[age >= 45]=0
time n.risk n.event survival std.err lower upper
2.3 4 1 0.75 0.217 0.426 1.000
2.7 3 1 0.50 0.250 0.188 1.000
5.0 2 1 0.25 0.217 0.046 1.000
```

```
sex[age >= 45]=1
time n.risk n.event survival std.err lower upper
0.7 8 1 0.875 0.117 0.673 1.000
0.9 7 1 0.750 0.153 0.503 1.000
3.4 4 1 0.562 0.199 0.281 1.000
3.6 2 1 0.281 0.222 0.060 1.000
```

```
plot(fi t,xlab="time",ylab="survival probabilities",lty=1:2,main="Two-sample data")
text(4,0.9,"age 45 and older",cex=1.5)
text(2.0,0.5,"Female",cex=1.5)
text(4.0,0.65,"Male",cex=1.5)
```

```
#Weibull model
#interaction with age?
fi t <- survreg(Surv(t,cc)~ (sex+age)^ 2, dist="weibull")
summary(fi t)
```

```
Call:
survreg(formula=Surv(t,cc)~(sex+age)^2, dist="weibull")

              Value Std. Error      z      p
(Intercept)  3.1374      1.1986  2.618 0.00886
sex          -2.0919      1.7867 -1.171 0.24167
age          -0.0287      0.0263 -1.092 0.27488
sex:age       0.0370      0.0399  0.928 0.35336
Log(scale)  -0.6306      0.1928 -3.270 0.00108
Scale= 0.532
```

```
Weibull distribution
Loglik(model)= -41.4   Loglik(intercept only)= -43.4
  Chisq= 3.96 on 3 degrees of freedom, p= 0.27
Number of Newton-Raphson Iterations: 7
n= 40
```

```
L <- -2*fi t$loglik[2]
fi t <- survreg(Surv(t,cc)~ sex+age, dist="weibull")
summary(fi t)
```

```
Call:
survreg(formula = Surv(t, cc) ~ sex + age, dist = "weibull")

              Value Std. Error      z      p
(Intercept)  2.4085      0.9016  2.671 0.00755
sex          -0.4514      0.2722 -1.658 0.09728
age          -0.0122      0.0203 -0.598 0.54998
Log(scale)  -0.6202      0.1947 -3.185 0.00145
Scale= 0.538
```

```
Weibull distribution
Loglik(model)= -41.8   Loglik(intercept only)= -43.4
  Chisq= 3.1 on 2 degrees of freedom, p= 0.21
Number of Newton-Raphson Iterations: 7
n= 40
```

```
L0 <- -2*fit$loglik[2]
X2 <- L0-L   #likelihood ratio statistic
pvalue <- 1-pchisq(X2,1)
round(cbind(X2,pvalue),3)
```

```
      X2 pvalue
[1,] 0.864 0.353
```

```
#additive model -- parameter estimates
fit <- survreg(Surv(t,cc)~ sex+age, dist="weibull")
summary(fit)
```

```
Call:
survreg(formula=Surv(t,cc)~sex+age, dist="weibull")

              Value Std. Error      z      p
(Intercept)  2.4085     0.9016  2.671 0.0076
sex          -0.4514     0.2722 -1.658 0.0973
age          -0.0122     0.0203 -0.598 0.5500
Log(scale)  -0.6202     0.1947 -3.185 0.0015
Scale= 0.538
```

```
Weibull distribution
Loglik(model)= -41.8   Loglik(intercept only)= -43.4
  Chisq= 3.1 on 2 degrees of freedom, p= 0.21
Number of Newton-Raphson Iterations: 7
n= 40
```

```
s <- fit$scale
b <- -fit$coefficients/s
g <- 1/s
l0 <- exp(b[1])
l1 <- exp(b[1]+b[2])
```

```
median0 <- (-log(0.5)/10)^(1/g)
median1 <- (-log(0.5)/11)^(1/g)
mean0 <- 10^(-1/g)*gamma(1+1/g)
mean1 <- 11^(-1/g)*gamma(1+1/g)

round(cbind(n,d,10,11,g,mean0,mean1,median0,median1),3)
      n  d   10   11    g mean0 mean1 median0 median1
(Intercept) 40 16 0.011 0.026 1.859 9.873 6.286   9.129   5.812
```

#impact of age? and confounding influence of age?

```
fit <- survreg(Surv(t,cc)~sex, dist="weibull")
```

```
summary(fit)
```

Call:

```
survreg(formula=Surv(t,cc)~sex, dist="weibull")
```

	Value	Std. Error	z	p
(Intercept)	1.891	0.196	9.64	5.30e-22
sex	-0.477	0.273	-1.75	8.08e-02
Log(scale)	-0.605	0.194	-3.12	1.83e-03

Scale= 0.546

Weibull distribution

```
Loglik(model)= -42   Loglik(intercept only)= -43.4
```

```
Chisq= 2.75 on 1 degrees of freedom, p= 0.097
```

```
Number of Newton-Raphson Iterations: 7
```

```
n= 40
```

```
s <- fit$scale
```

```
B <- -fit$coefficients/s
```

```
L1 <- -2*fit$loglik[2]
```

```
X2 <- L1-L0 #likelihood ratio statistic
```

```
pvalue <- 1-pchisq(X2,1)
```

```
round(cbind(X2,pvalue),3)
```

	x2	pvalue
[1,]	0.347	0.556

```
#confounding
bias <- B[2]-b[2]
age0 <- mean(age[sex==0])
age1 <- mean(age[sex==1])
B0 <- B[2] # age removed from the model
b0 <- b[2] # age included in the model
round(cbind(age0,age1,B0,b0,bias),3)
      age0 age1 B0 b0 bias
sex 44.095 45.474 0.874 0.839 0.034
```

```
#baseline survival and hazard functions
```

```
par(mfrow=c(1,2))
par(pty="s")
T <- seq(0,10)
S <- exp(-10*T^ g)
plot(T,S,type="l")
title("Survival function")
H <- 10*g*T^ (g-1)
plot(T,H,type="l")
title("Hazard function")
```

```
#cox proportional hazard model
```

```
f <- coxph(Surv(t,cc)~ sex+age)
summary(f)
Call:
coxph(formula=Surv(t,cc)~sex+age)
```

```
n= 40
```

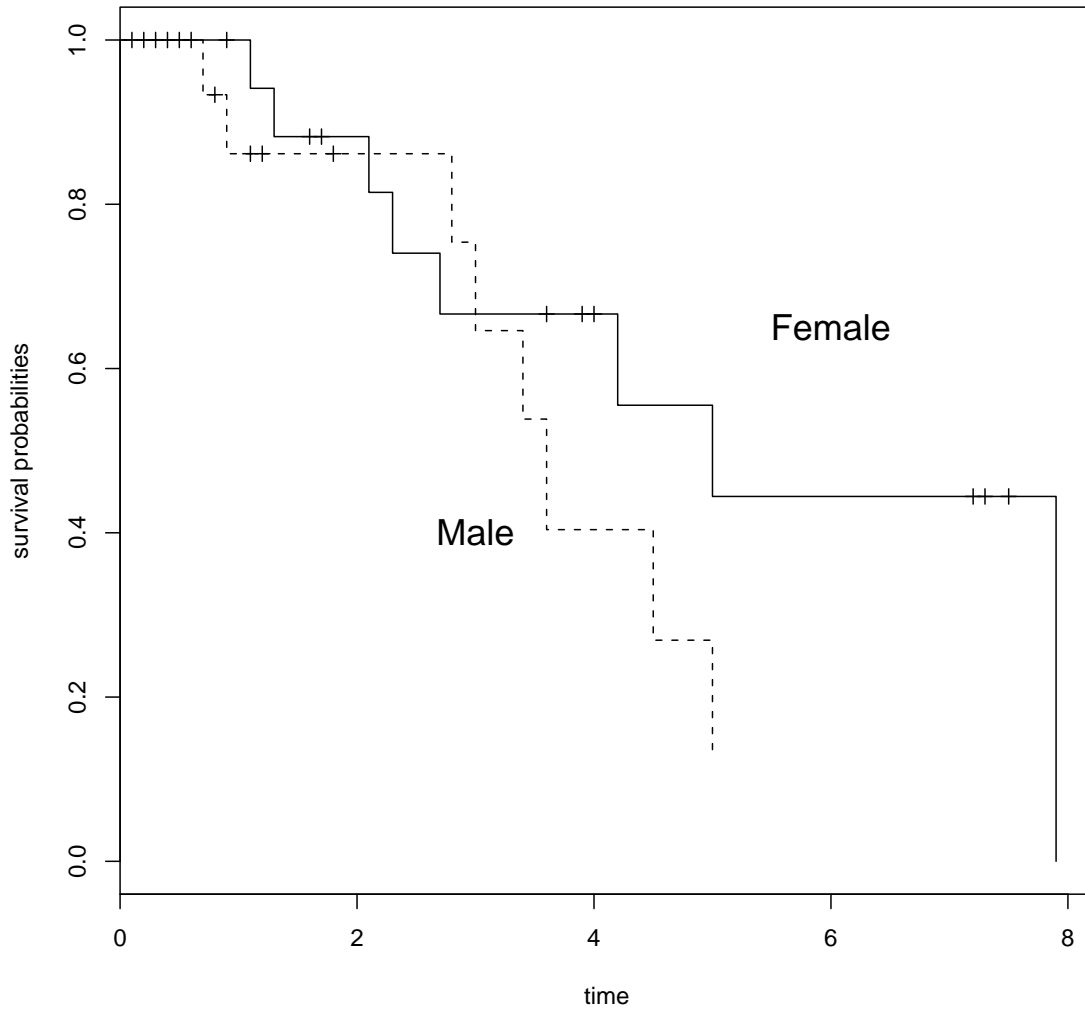
	coef	exp(coef)	se(coef)	z	p
sex	0.5834	1.79	0.5251	1.111	0.27
age	0.0288	1.03	0.0403	0.714	0.48

	exp(coef)	exp(-coef)	lower .95	upper .95
sex	1.79	0.558	0.640	5.02
age	1.03	0.972	0.951	1.11

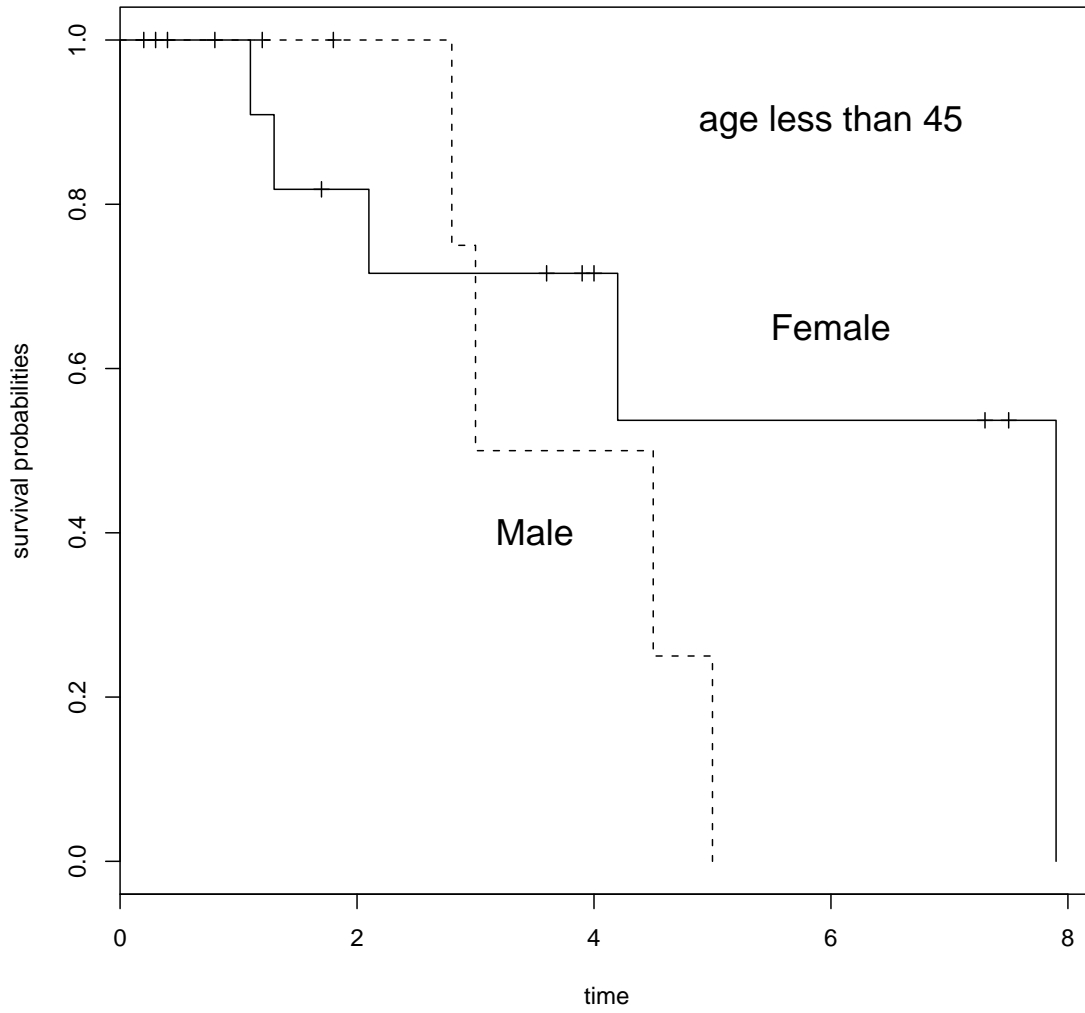
```
Rsquare= 0.047    (max possible= 0.873 )  
Likelihood ratio test= 1.91  on 2 df,    p=0.385  
Wald test          = 1.90  on 2 df,    p=0.387  
Score (logrank) test = 1.95  on 2 df,    p=0.377
```

```
par(mfrow=c(1,1))  
plot(survfit(f),xlab="time",ylab="survival probability")  
title("Baseline survival function -- Cox estimated model")  
#plots  
f0 <- survfit(f)  
T <- f0$time  
S <- f0$surv  
plot(T,S,type="s",xlab="time",ylab="survival probability")  
S1 <- S ^ exp(b0)    #b0 from additive model  
lines(T,S1,type="s",lty=2)  
title("Survival functions -- Cox estimated model")  
text(2.5,0.4,"Males",cex=1.5)  
text(6,0.4,"Females",cex=1.5)
```

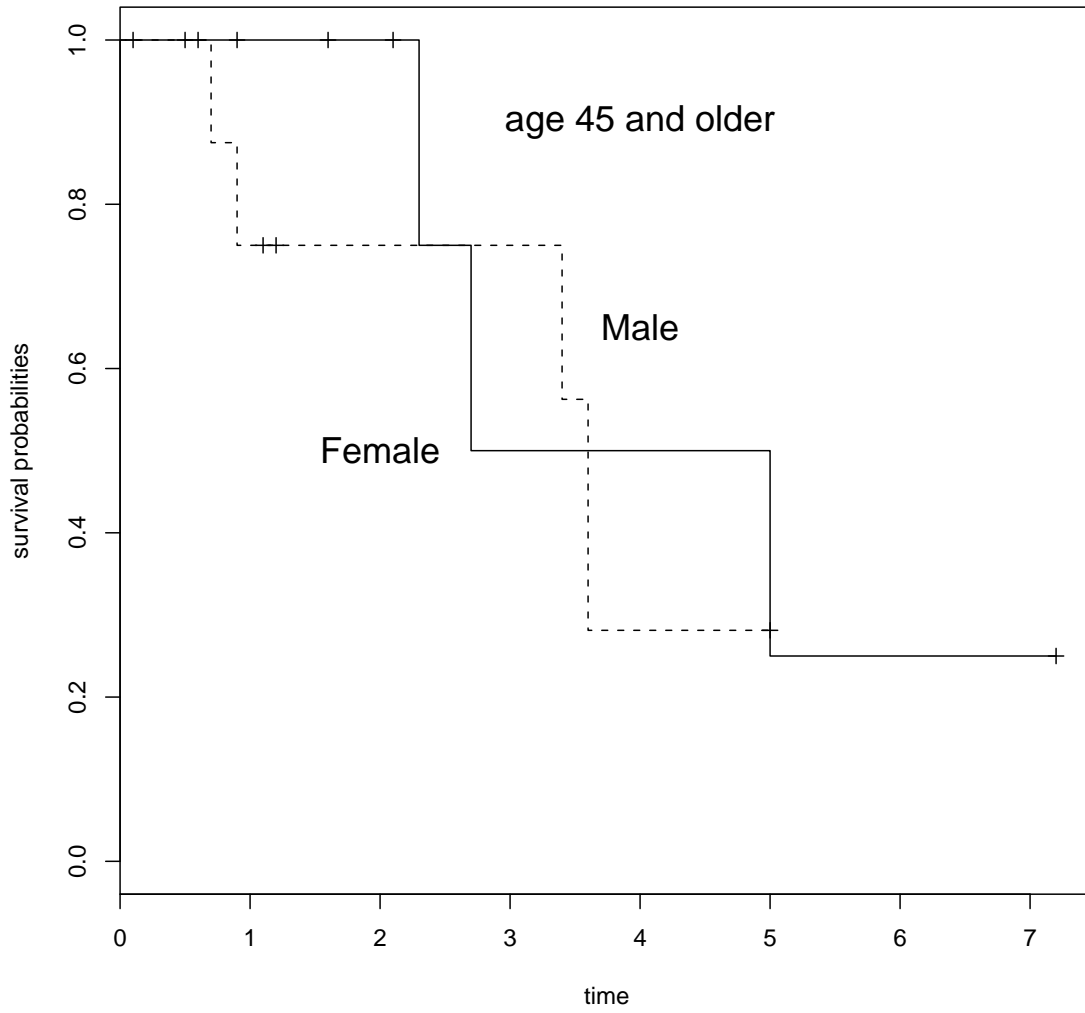
Two-sample data by sex

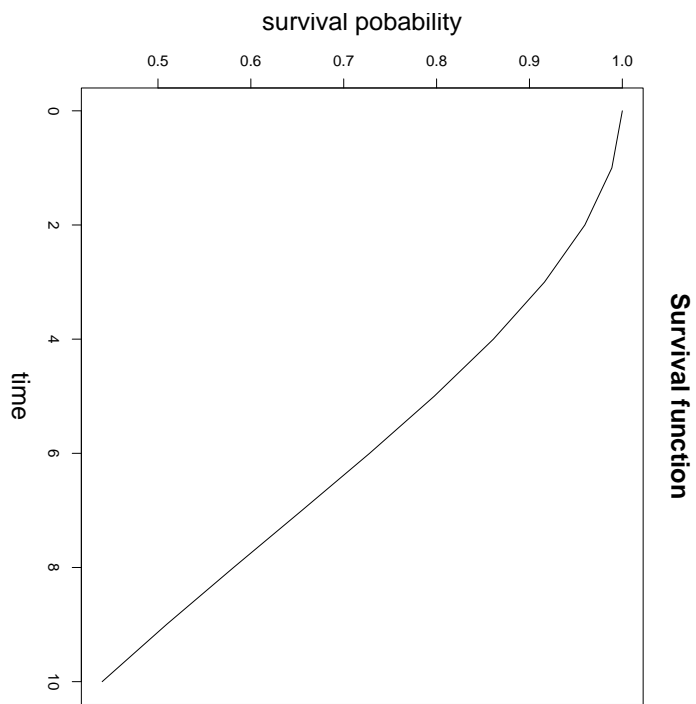
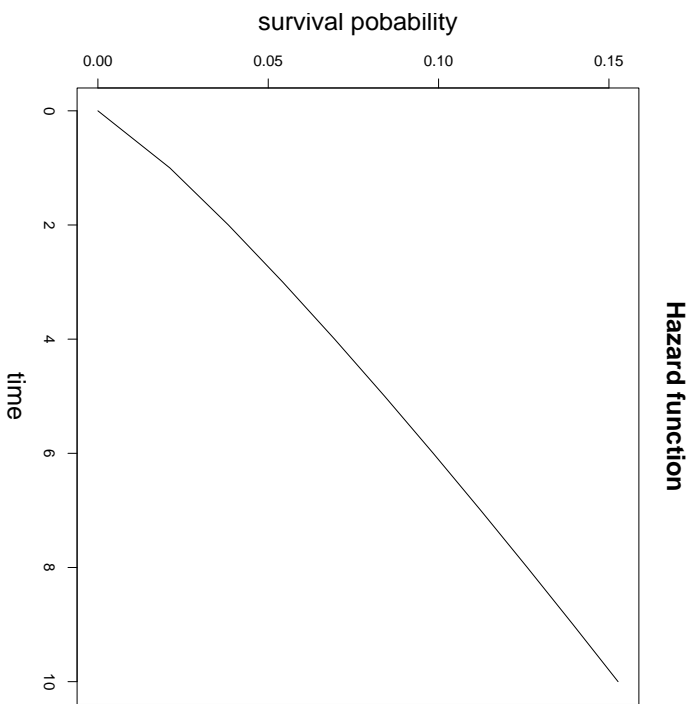


Two-sample data by sex (age less than 45)

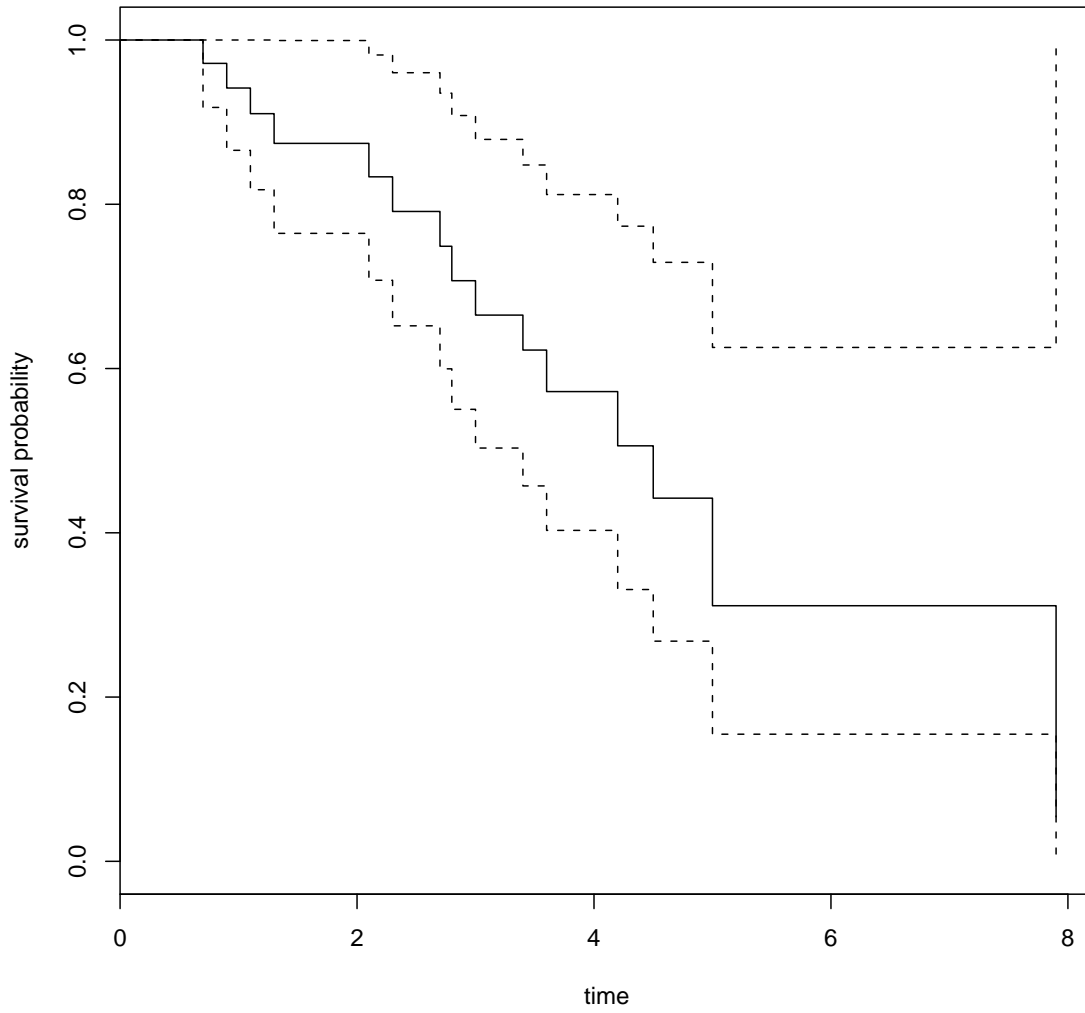


Two-sample data by sex (age 45 and older)

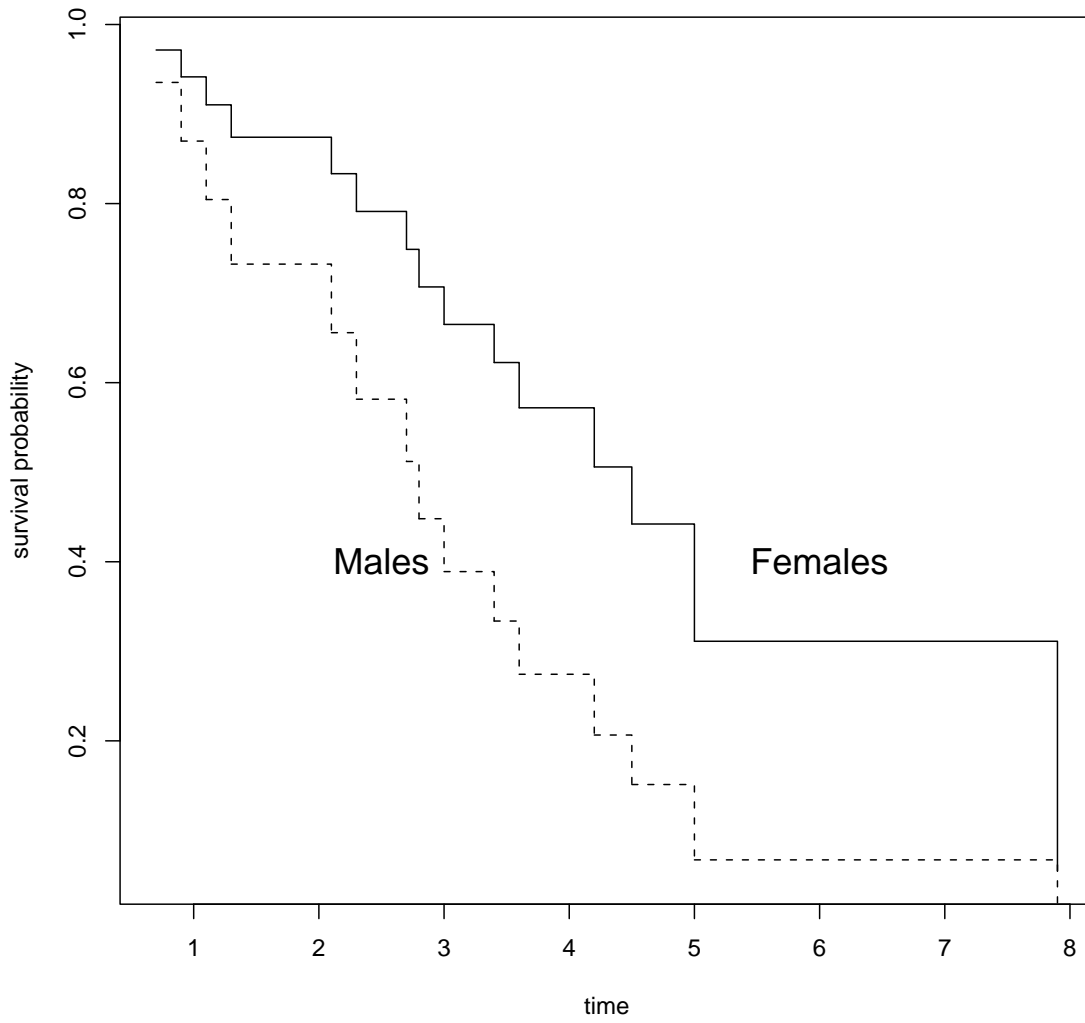




Baseline survival function -- Cox estimated model



Survival function -- Cox estimated model



Example 4

Residual analysis: general proportional hazards models

```
#example.4.r -- Residual values from the Cox approach (Lecture 9)
#library(survival)

#data
t <- c(1.2,5.0,0.3,3.0,1.3,0.9,7.2,2.3,3.4,
2.7,2.8,1.6,1.1, 1.1,0.7,3.9, + 1.7,7.3,4.5,
7.5,1.2,0.9,0.6,0.2,2.1,2.1,5.0, 4.0,0.8,5.0,
0.5,1.8,3.6,0.1,7.9,4.2,0.1,3.4,0.4,3.6)

cc <- c(0,0,0,1,1,1,0,1,0,1,1,0,1,0,1,0,0,
0,1,0,0,0,0,0,1,0,1,0,0,1,0,0,0,0,1,1,0,1,0,1)

sex <- c(1,1,0,1,0,1,0,0,1,0,1,0,0,1,1,0,0,0,1,
0,1,0,1,1,0,0,1,0,1,0,0,1,0,1,0,0,1,1,0,1)

age <- c(40,48,33,42,41,46,53,57,55,48,33,59,44,
52,51,40,33,37,41,41,46,47,46,41,44,53,37,38,
38,47,54,44,42,51,35,43,53,48,37,52)

n <- length(t)   #sample size
d <- sum(cc)     #number of complete observations

#Cox proportional hazard model
f <- coxph(Surv(t,cc)~ sex+age,method='breslow')
summary(f)
```

```
Call:
coxph(formula = Surv(t, cc) ~ sex + age, method = "breslow")
```

```
n= 40
```

	coef	exp(coef)	se(coef)	z	p
sex	0.5748	1.78	0.5241	1.097	0.27
age	0.0295	1.03	0.0403	0.733	0.46

	exp(coef)	exp(-coef)	lower .95	upper .95
sex	1.78	0.563	0.636	4.96
age	1.03	0.971	0.952	1.11

```
Rsquare= 0.046 (max possible= 0.874 )
Likelihood ratio test= 1.89 on 2 df, p=0.390
Wald test = 1.87 on 2 df, p=0.392
Score (logrank) test = 1.93 on 2 df, p=0.381
```

```
r <- f$residual
rr <- cc-r #recovering the residual values from the modified values
fit <- survfit(Surv(rr,cc)~ 1)
S <- fit$surv
T <- fit$time

#plot residual values
#postscript("ex.10.ps",width=8,height=8,horizontal=F)
plot(T,S,xlim=range(T),ylim=c(0,1),xlab="time",ylab="residual values")
title("Residual plot: residual values (exponential ?)")
t0 <- seq(0,max(T),0.05)
S0 <- exp(-t0)
lines(t0,S0)

#transformed
#postscript("ex.11.ps",width=8,height=8,horizontal=F)
plot(log(T),log(-log(S)),xlab="log(time)",ylab="log(-log[residual values])")
abline(0,1)
title("Residual plot: transformed residual values (straight-line?)")
```

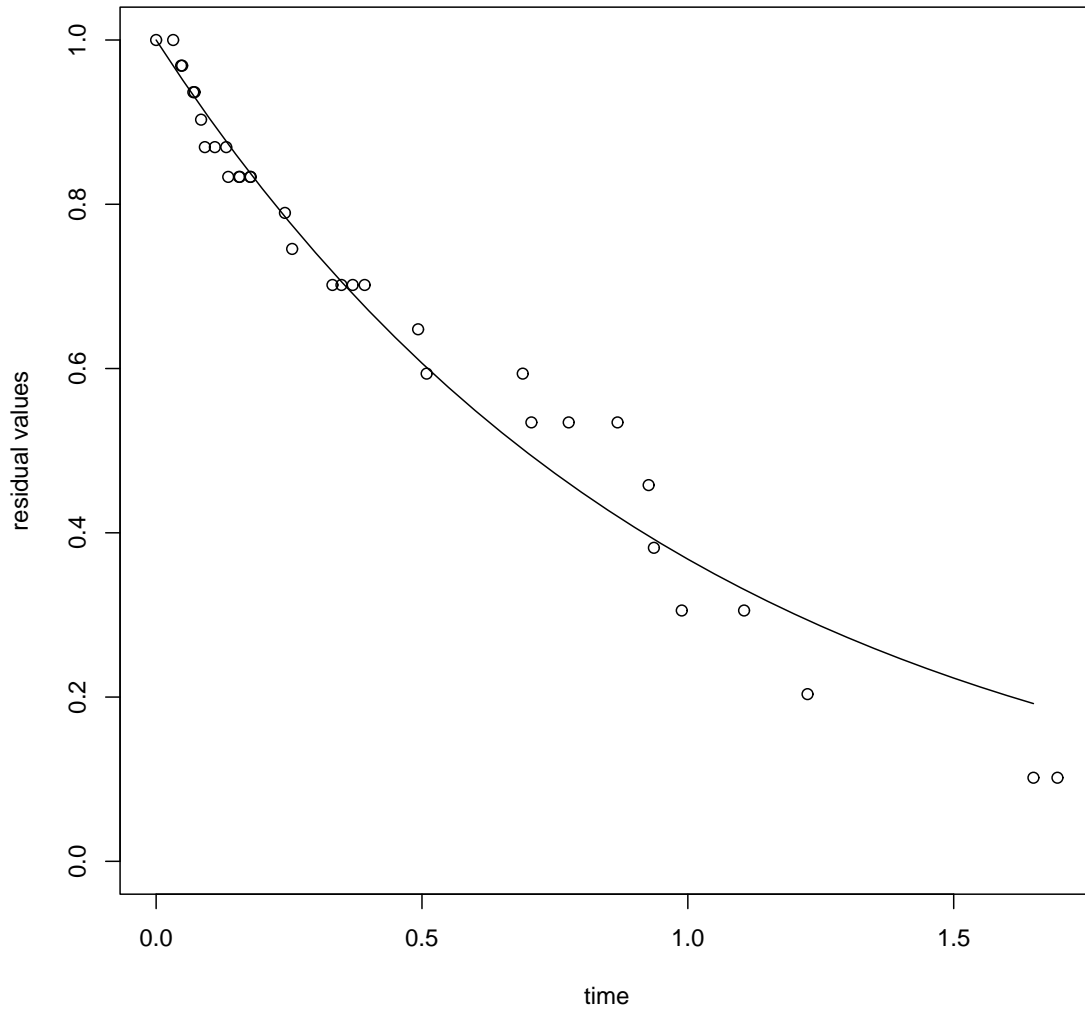
```
#binomial test of distribution censored values: males versus females
n0 <- length(cc[sex==0])
n <- length(cc)
binom.test(n0,n,p=0.5)
```

Exact binomial test

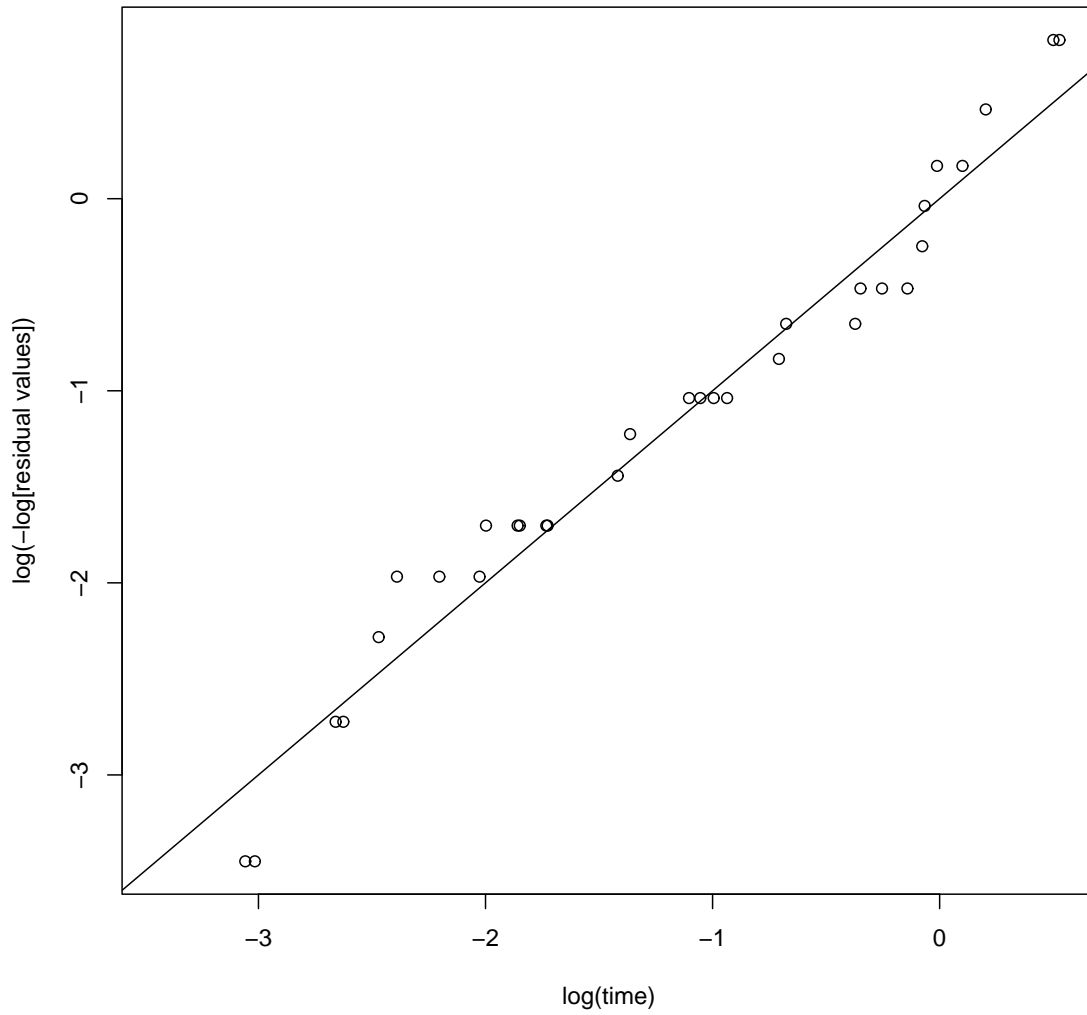
```
data: n0 and n
number of successes = 21, number of trials = 40, p-value = 0.8746
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.3612801 0.6848803
sample estimates:
probability of success
                0.525
```

```
#transformed modified Cox-Snell residuals by age
m <- sign(r)*sqrt(-2*(r+cc*log(cc-r)))
#postscript("ex.12.ps",width=8,height=8,horizontal=F)
plot(age,m,xlab="age",ylab="modified residual values")
abline(h=0)
title("Modified residual values by age -- example data")
```

Residual plot: residual values (exponential ?)



Residual plot: transformed residual values (straight-line?)



Modified residual values by age -- example data

