Survival Analysis: Left-Truncated Data

Introduction: The random variable of most interest in survival analysis is time-to-event. Often in biomedical studies, the event is death. It is because of this common application the field is termed *Survival Analysis*. Survival analysis methods can be applied to a wide range of data not just biomedical survival data. Other time-to-event data can include: time-to-relapse of a disease, length of stay in a hospital, duration of a strike, money paid by insurance company, time-to-employment (Klein), failure times of electrical components, etc. (Survival).

There is generally a relatively short window for the duration of any study where the event is observed to occur or not occur. When the duration of study is small, adjustments must be made to account for potential biases. Useful data will be excluded when data is censored but not accounted for, and biases can be introduced when data is truncated. Since censoring and truncation are often confused, a brief discussion on censoring with examples is helpful to more fully understand left-truncation.

There are three general types of censoring, right-censoring, left-censoring, and interval-censoring. The most common type of censoring encountered in survival analysis data is right censored (Survival). In our course, we adjusted our model for the herpes data to account for right censoring. It is called right censoring because the true unobserved event is to the right of the censoring time. Left-censoring occurs when we cannot observe the time when the event occurred. For obvious reasons if the event is death, the data can't be left-censored. A good example is discussed in an ASA paper on survival analysis, "e.g. [a] study of age at which African children learn a task. Some already knew (left-censored), some learned during a study (exact), some had not yet learned by end of study (right-censored)." Interval-censoring is also discussed in *Survival Analysis: Introduction* (Survival).

Truncation is due to sampling bias that only those individuals whose lifetimes lie within a certain interval can be observed. To further discuss left-truncation we will use the motivating example on the next page. A useful insight and directly applicable to the motivating example, "Age is often used as a covariate when it should be used as a left-truncation point. When age is used as a left-truncation point, it is unnecessary to use it as a covariate in the model" (Klein).

Motivating Example: Survival data on 26 psychiatric inpatients admitted to the University of Iowa hospitals during the years 1935–1948. This sample is part of a larger study of psychiatric inpatients discussed by Tsuang and Woolson (1977). Data for each patient consists of age at first admission to the hospital, sex, number of years of follow-up (years from admission to death or censoring) and patient status at the follow-up time. Ultimately we want to know if male psychiatric patients and female psychiatric patients are more likely to die than the general public. The data from the hospitals are displayed in the table below (Klein).

Obs	sex	age	time	death
1	2	51	1	1
2	2	58	1	1
3	2	55	2	1
4	2	28	22	1
5	1	21	30	0
6	1	19	28	1
7	2	25	32	1
8	2	48	11	1
9	2	47	14	1
10	2	25	36	0
11	2	31	31	0
12	1	24	33	0
13	1	25	33	0
14	2	30	37	0
15	2	33	35	0
16	1	36	25	1
17	1	30	31	0
18	1	41	22	1
19	2	43	26	1
20	2	45	24	1
21	2	35	35	0
22	1	29	34	0
23	1	35	30	0
24	1	32	35	1
25	2	36	40	1
26	1	32	39	0

The way we coded the variable death, death(0) indicates censoring time. It is important to adjust for left-truncation in this data because patients entered into the study (admitted to the hospital) have not died. This fact could bias the results, when results are compared to the general population of Iowa residents. We stratified by sex because many studies show women live longer than men so we want to compare by gender.

The following is the code used to obtain Kaplan-Meier curves, graphical representations of survival estimates. Use of the proc phreg procedure for left-truncated data is discussed in *Surviving Left Truncation Using PROC PHREG* (Foreman).

```
proc lifetest data=psych method=km plots=(survival(cl));
        strata sex;
        time time*death(0); /* time var is years from admission to death or censoring
        censoring indicated when death=0 */
        survival out=out1 conftype=log;
run;

proc phreg data=psych plots(cl overlay=stratum)=survival;
        strata sex;
        model time*death(0)=;
        output out=psych01 survival=S;
run;

proc print data=psych01;
run;

data temp; set psych;
        agetime=age+time;
run;

proc phreg data=temp plots(cl overlay=stratum)=survival;
        strata sex;
        model agetime*death(0)=/entry=age;
        output out=psych02 survival=S;
run;

proc sort data=psych02;
        by agetime;
run;
proc print data=psych02;
run;
```
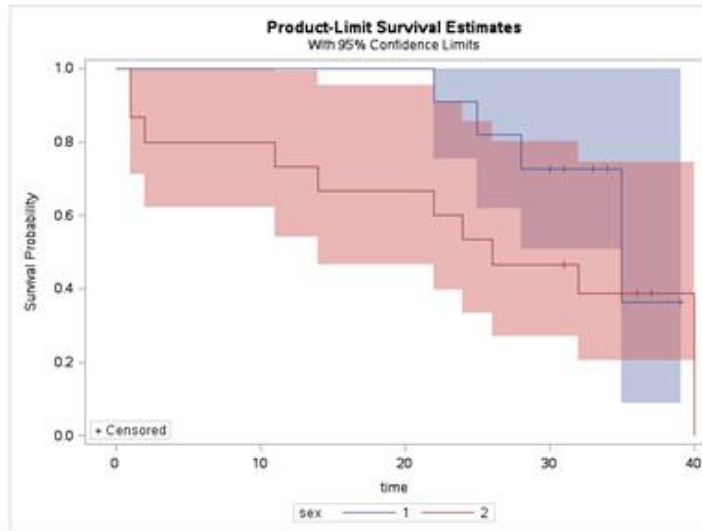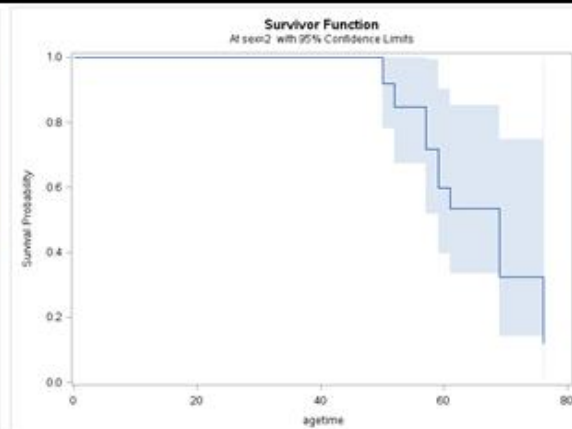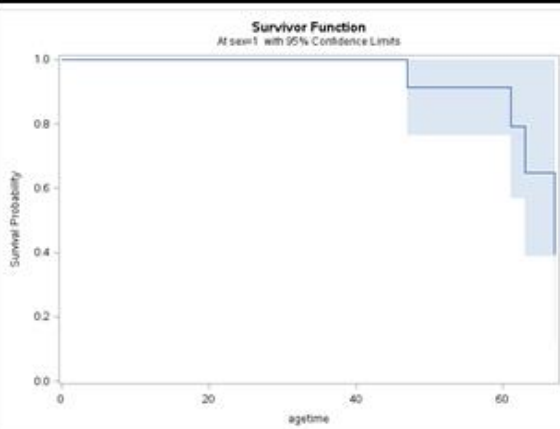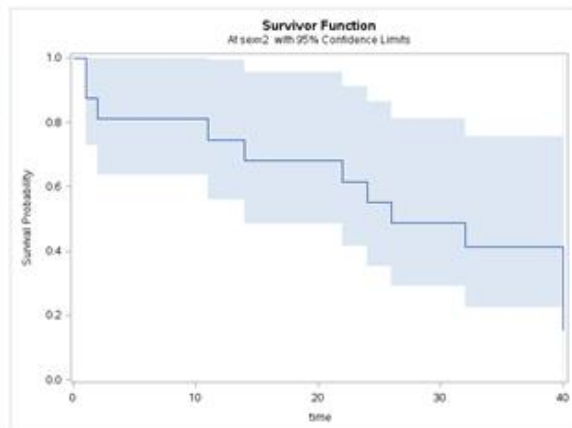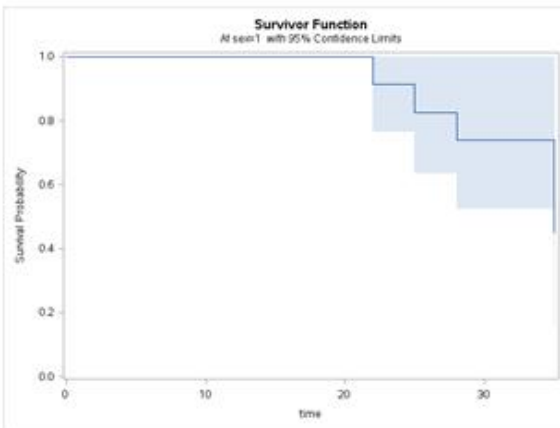
Product-Limit Survival Estimates
With 95% Confidence Limits

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	1.6074	1	0.2049
Wilcoxon	2.8418	1	0.0918
-2Log(LR)	2.5358	1	0.1113

Survivor Function
At sex=1 with 95% Confidence Limits

Survivor Function
At sex=2 with 95% Confidence Limits

Survivor Function
At sex=1 with 95% Confidence Limits

Survivor Function
At sex=2 with 95% Confidence Limits

A combination of survival curves over strata (male, female) with accounting for left-truncation can be obtained in R with the following code (Diaz).

```
install.packages('survival')
install.packages('KMsurv')

library(survival)
library(KMsurv)


data(psych); attach(psych)
my.surv.object <- Surv(age, age+time, death)
my.surv.object
survfit(my.surv.object~sex)
my.fit <- survfit(my.surv.object~sex)

plot(my.fit, main="Kaplan-Meier estimate with 95% confidence bounds",
    xlab="time", ylab="survival function")
```
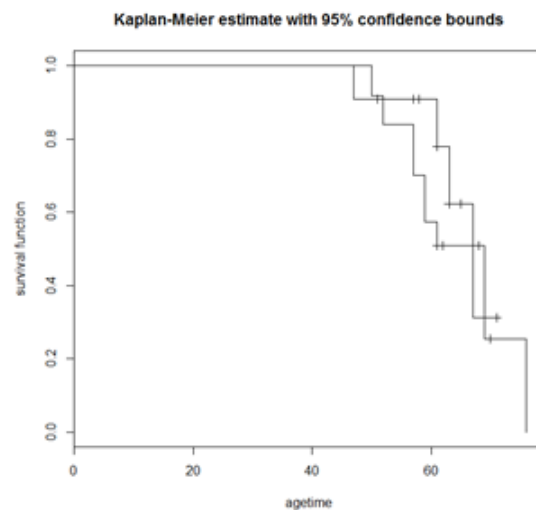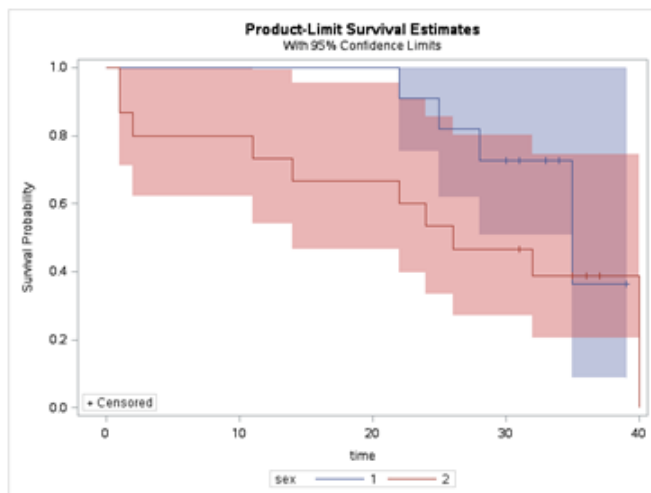
References:

Surviving Left Truncation Using PROC PHREG
Aimee J. Foreman, Ginny P. Lai, Dave P. Miller,
ICON Clinical Research, San Francisco, CA
http://www.wuss.org/proceedings08/08WUSS%20Proceedings/papers/anl/anl03.pdf

Survival Analysis in R
David Diez
http://stat.ethz.ch/education/semesters/ss2011/seminar/diez.pdf

The Basic of Survival Analysis
www.stat.nus.edu.sg%2F~stachenz%2FST3242Notes1.pdf&ei=sFqeUrm9EInTqgHEwoDQDg&usg=AFQjC
NG9xBEUwNzBPhit2rxcdpehosjx4w&sig2=6fdDO0e2ey4kGSUzUVobrg

Survival Analysis: Introduction – American Statistical Association
www.amstat.org%2Fchapters%2Fnortheasternillinois%2Fpastevents%2Fpresentations%2Fsummer05_Ib
rahim_J.pdf&ei=TlqeUsrSF82grgH2q4AQ&usg=AFQjCNEdeHLLNgy6lSIyCSXdiurzHexzpg&sig2=RA5v5VTuL
YRObayH1n8QhQ

Survival Analysis: Techniques for Censored and Truncated Data
John P. Klein and Melvin L. Moeschberger
http://f3.tiera.ru/2/M_Mathematics/MV_Probability/MVsa_Statistics%20and%20applications/Klein%20
J.,%20Moeschberger%20M.%20Survival%20analysis..%20Techniques%20for%20censored%20and%20tru
ncated%20data%20(2ed.,%20Springer,%202003)(ISBN%20038795399X)(542s)_MVsa_.pdf