EXERCISE 4.2 (BMT)

```
/* To find the location of the file in SAS Studio: Find the file in the
folder you uploaded it to, under Server Files and Folders - Files
(Home). Right click and choose Properties. Copy the path from Location
(Ctrl/Command+C) and paste it in your code (Ctrl/Command+V), with
quotation marks around it (single or double).
An example of a path is used in the code below, replace that by your
own unique path for each exercise. */



data BMT; /* From section 1.3 */
infile '/home/username/Survival analysis/Book data/BMT.txt'
   firstobs=2;
input Group TD TDFS Death Relapse DFS TA A TC C TP P PatAge DonAge
PatSex DonSex PatCMV DonCMV Transpl FAB Hosp MTX;
run;

/* Variables used below:
TDFS: disease free survival time, i.e. time to relapse or death
DFS: event indicator, where 1 denotes relapse or death, and 0 denotes
an individual still alive and disease free.

You can if you prefer instead use TDFS together with the relapse
indicator Relapse, which would mean that you focus on time to relapse
and censor for deaths */



/* ASSUMPTIONS for the K-M estimator*/

/* Check censoring pattern */

/* Create format to be used in the plot below */

proc format;
value cens                    /* Creating a format named “cens” */
   1 = 'Not censored'
   0 = 'Censored'
   other = 'Unknown';
run;
```

```
/* Copy the code from SAS examples 4.1 and adjust to this example */

ods graphics / reset attrpriority=none;
/* Need to override the built-in attributes in proc sgplot */

proc sgplot data=BMT;
styleattrs          /* Without the statement above these changes would
                    not have any effect */
   datasymbols=(plus circle)
   datacolors=(black red);
scatter x=TDFS y=Group /group=DFS jitter; /* The blue "group" option
tells SAS which values that should get the different colors/symbols
we've assigned above. Not the same as our variable named "group". */
xaxis label= "Disease free survival (days from transplant)";
yaxis integer label='Risk group'; /* "Integer" ensures that no tick
marks are shown between the group values */
format DFS cens.;
label DFS="00"x; /* Remove the name "DFS" from the x axis legend */
run;

/* Censoring spread out over time without any particular patterns
(except for the right censoring, which is expected), and patterns not
that different across groups. Good. */

/* SAMPLE SIZE*/
/* To be investigated with proc lifetest below */



/* KAPLAN-MEIER SURVIVAL ESTIMATES */
/* Also provides a summary of the data. Remember to check for small
samples and/or heavy censoring. */

proc lifetest data=BMT;
time TDFS*DFS(0);   /*Time variable * event variable(censoring value)*/
strata Group;
run;
/* Output above shows n=83 events in total (24, 25, 34 in each group)*/
/* 39% censoring (37%, 54%, 24% in each group) */

/* Groups 1, 2, and 3 doesn't say much. Create a format for group */

proc format;
value grp           /* Creating a format named "grp" */
   1 = 'ALL'
   2 = 'AML low risk'
   3 = 'AML high risk'
   other = 'Unknown';
run;
```

```
proc lifetest data=BMT;
time TDFS*DFS(0);
strata Group;
format Group grp.;
run;



/* NELSON-AALEN CUMULATIVE HAZARD AND SURVIVAL */

proc lifetest data=BMT nelson method=breslow;
time TDFS*DFS(0);
strata Group;
format Group grp.;
run;
/* "nelson" gives the Nelson-Aalen cumulative hazards estimates */
/* "method=Breslow" gives the Breslow estimator, which is the
exponentiation of the negative Nelson-Aalen estimator of the cumulative
hazard function - i.e. the survival estimated by the Nelson-Aalen
method. */


proc lifetest data=BMT method=breslow plots=LOGSURV;
time TDFS*DFS(0);
strata Group;
format Group grp.;
run;
/* "plots=LOGSURV" constructs a plot of the negative log of the
estimated survival function, i.e., the cumulative hazards function. The
produced plot is however not a step function, so we'll have to
construct a better plot ourselves (see SAS examples 4.3). */


ods trace on;       /* ods = Output Delivery System */

proc lifetest data=BMT nelson method=breslow;
time TDFS*DFS(0);
strata Group;
format Group grp.;
run;

/* Check the log window to see which output can be produced from the
code above*/

ods trace off;      /* Switch off the ods trace */

proc lifetest data=BMT method=breslow nelson;
time TDFS*DFS(0);
strata Group;
format Group grp.;
ods output BreslowEstimates=Estimates;  /*Name the output to be saved*/
run;
```

```
/* Plot Nelson-Aalen cumulative hazard rates*/

/* In order to plot the cumhaz, we need to replace the missing values
with previous existing value (for censored observations to show in the
plot) */
data figure; set estimates;
 retain impute_value .;
 if missing(cumhaz)
   then cumhaz = impute_value;
   else impute_value = cumhaz;
run;



ods graphics / attrpriority=none;
/* Need to override the built-in attributes in proc sgplot */

proc sgplot data=figure;
styleattrs datacontrastcolors=(orange purple blue)
      datalinepatterns=(solid shortdash mediumdash);
step x=TDFS y=cumhaz / group=Group;
title "Nelson-Aalen cumulative hazard";
xaxis label="Disease free survival (days from transplant)";
yaxis label="Cumulative hazard" grid;
format Group grp.;
label Group="00"x; /*Removes the word “Group” from the x-axis legend*/
run;

/* Transform the cumulative hazard to survival estimates */
data figure; set figure;
surv_nelson=exp(-cumhaz);
run;
/* Gives exactly the same values as the Breslow estimator */



/* POINTWISE CONFIDENCE INTERVALS */
/* The default log-transformation is okay to use */

proc lifetest data=BMT plots=survival(cl strata=panel);
time TDFS*DFS(0);
strata Group;
label TDFS='Disease free survival (days from transplant)';
format Group grp.;
run;
/* If you exclude “strata=panel” the three groups will be in the same
plot, with confidence intervals overlapping*/


/* If you want to make adjustments to the plot, e.g. titles, not
connecting the confidence intervals over time, etc, you have to store
the data and create a plot “manually” later on: */
```

```
proc lifetest data=BMT plots=none
outsurv=survCI;     /* Saves the results in data set survCI */
time TDFS*DFS(0);
strata Group;
format Group grp.;
run;



/* Plot survival curve with pointwise confidence intervals */

proc sort data=survCI; by Group TDFS; run;

proc sgplot data=survCI;
styleattrs datacontrastcolors=(orange purple blue)
   datalinepatterns=(solid);
step y=survival x=TDFS/Yerrorlower=SDF_LCL Yerrorupper=SDF_UCL
   group=Group markers markerattrs=(symbol=plus);
xaxis label ="Disease free survival (days from transplant)";
yaxis label='Estimated survival (probability of not relapsing)';
label Group="00"x; /* Remove the name "Group" from x-axis label */
format Group grp.;
run;
/*The censored observations at the end are not included in the graph */

/* Impute survival estimates for the censored times */
data survCI2; set survCI;
retain impute_value .;
if missing(survival)
/* Only S(t) for censored times, not the CI, to make
clear that the intervals are pointwise at event times */
   then survival = impute_value;
else impute_value = survival;
run;

/* Plot survival curve including censored observations at the end*/

proc sgplot data=survCI2;
styleattrs datacontrastcolors=(orange purple blue)
   datalinepatterns=(solid);
step y=survival x=TDFS/Yerrorlower=SDF_LCL Yerrorupper=SDF_UCL
   group=Group markers markerattrs=(symbol=plus);
title "Pointwise 95% log-transformed confidence intervals for Kaplan-
Meier survival";
xaxis label ="Disease free survival (days from transplant)";
yaxis label='Estimated survival (probability of not relapsing)';
format Group grp.;
label Group="00"x;
run;

/*Perhaps a bit cluttered to have all three groups in the same graph */
```

```
proc sgplot data=survCI2;
step y=survival x=TDFS/Yerrorlower=SDF_LCL Yerrorupper=SDF_UCL
   markers markerattrs=(symbol=plus);
title "Pointwise 95% log-transformed confidence intervals for Kaplan-
Meier survival";
xaxis label ="Disease free survival (days from transplant)";
yaxis label='Estimated survival (probability of not relapsing)';
format Group grp.;
by Group; /* Produces three separate graphs */
run;


/* EQUAL PROBABILITY BANDS */

/* Arcsine-square root transformed */

proc lifetest data=BMT plots=survival(cb=ep strata=panel)
   conftype=ASINSQRT;
time TDFS*DFS(0);
strata Group;
label TDFS='Disease free survival (days from transplant)';
format Group grp.;
run;


/* Again, if you want to make adjustments to titles, not connecting the
conficence intervals over time, etc: */

proc lifetest data=BMT plots=none confband=EP conftype=ASINSQRT
   bandmintime=100 bandmaxtime=600  /* tL and tU */
   outsurv=survEP stderr;  /* Saves the results (incl std) in dataset
   survEP */
time TDFS*DFS(0);
strata Group;
run;

/* The survEP dataset will still contain the full set of data, which
means that you have to constraint the graph too (see below)*/

/* Plot survival curve with EP confidence bands */

proc sgplot data=survEP;
where 100<=TDFS<=600;
step x=TDFS y=survival / lineattrs=(color=black pattern=1);
step x=TDFS y= EP_LCL / lineattrs=(color=red pattern=3);
step x=TDFS y= EP_UCL / lineattrs=(color=red pattern=3);
title "95% Arcsine-square root transformed equal probability bands";
xaxis label ="Disease free survival (days from transplant)";
yaxis label="Estimated Survival Function";
by Group;
format Group grp.;
run;
```

```
/* HALL-WELLNER CONFIDENCE BANDS */

/* Arcsine-square root transformed */

proc lifetest data=BMT plots=survival(cb=hw strata=panel)
    conftype=ASINSQRT;
time TDFS*DFS(0);
strata Group;
label TDFS='Disease free survival (days from transplant)';
format Group grp.;
run;
/* Again, if you want to make adjustments to titles, not connecting the
conficence intervals over time, etc: */

proc lifetest data=BMT plots=none
   bandmintime=100 bandmaxtime=600  /* tL and tU */
   confband=HW conftype=asinsqrt outsurv=survHW stderr;
time TDFS*DFS(0);
strata Group;
run;

/* Plot survival curve with HW confidence bands */
proc sgplot data=survHW;
where 100<=TDFS<=600;
step x=TDFS y=survival / lineattrs=(color=black pattern=1);
step x=TDFS y= HW_LCL / lineattrs=(color=red pattern=3);
step x=TDFS y= HW_UCL / lineattrs=(color=red pattern=3);
title "95% Arcsine-square root transformed Hall-Wellner confidence
bands";
xaxis label ="Disease free survival (days from transplant)";
yaxis label="Estimated Survival Function";
by Group;
format Group grp.;
run;
/* If we calculate the HW bands for all times, it will show a bit odd
appearance at early times. */



/* ALL THREE IN THE SAME GRAPH */

data figure_all;
merge survEP(keep=group TDFS _Censor_ survival SDF_LCL SDF_UCL EP_LCL
EP_UCL) survHW(keep=group TDFS _Censor_ survival HW_LCL HW_UCL);
by Group TDFS;
run;
```

```
/* Plot survival curve with pointwise CI:s and confidence bands */

proc sgplot data=figure_all;
where 100<=TDFS<=600;
step x=TDFS y=survival / lineattrs=(color=black pattern=1);
step x=TDFS y= SDF_LCL / lineattrs=(color=red pattern=2);
step x=TDFS y= SDF_UCL / lineattrs=(color=red pattern=2);
step x=TDFS y= EP_LCL / lineattrs=(color=blue pattern=3);
step x=TDFS y= EP_UCL / lineattrs=(color=blue pattern=3);
step x=TDFS y= HW_LCL / lineattrs=(color=green pattern=4);
step x=TDFS y= HW_UCL / lineattrs=(color=green pattern=4);
title "95% Arcsine-square root transformed Hall-Wellner confidence
bands";
xaxis label ="Disease free survival (days from transplant)";
yaxis label="Estimated Survival Function";
by Group;
format Group grp.;
run;

/* Not exactly the same as Figure 4.7 in the book, since the pointwise
intervals are log-transformed, and the confidence bands are arcsine-
square root transformed */


/* MEAN SURVIVAL TIME */

proc lifetest data=BMT plots=none;
   time TDFS*DFS(0);
   strata Group;
run;

/* "Note: The mean survival time and its standard error were
underestimated because the largest observation was censored and the
estimation was restricted to the largest event time." */


/* Efron's tail correction */

proc lifetest data=BMT plots=none
timelim=observed; /* Uses the largest observed time */
time TDFS*DFS(0);
strata Group;
ods output Means=mean_surv;
/*Store mean and stderr for each group in dataset mean_surv */
run;

/* Note: The mean survival time and its standard error may have been
underestimated because the estimation was restricted to the largest
observation, which was censored. */

data mean_surv; set mean_surv;
LCL=mean-1.96*stderr;
UCL=mean+1.96*stderr;
run;
```

```
/* MEDIAN SURVIVAL TIME */

proc lifetest data=BMT plots=none;
time TDFS*DFS(0);
strata Group;
run;



****************** CLEAN SAS WORK DATASETS ******************;
proc datasets lib=work nolist memtype=data kill;
run; quit;
/*==================== End of Programme =======================*/
```