Biostatistics 140.623
Third Term, 2017-2018

Laboratory Exercise 3
Answer Key

The times to “drug failure” (as determined by a treating psychiatrist) for 25 patients in a study comparing a new treatment for schizophrenia to a standard treatment as used in Self-Evaluation Problems Class 5 are:

Trt group	Times (weeks)
Standard	3, 5+, 6, 8, 8, 9, 13, 15+, 16, 16, 17, 18
New	4, 6, 9, 9, 10+, 11, 12, 13+, 14+, 16, 17, 18, 20

+ denotes a censored observation

The corresponding dataset can be found in the Stata file `trt.dta`

1. Check the listing below to confirm that the data set has been set up appropriately:

```
. list
          trt     weeks   failure        id
  1.        0         3         1         1
  2.        0         5         0         2
  3.        0         6         1         3
  4.        0         8         1         4
  5.        0         8         1         5
  6.        0         9         1         6
  7.        0        13         1         7
  8.        0        15         0         8
  9.        0        16         1         9
 10.        0        16         1        10
 11.        1        14         0        11
 12.        0        17         1        12
 13.        0        18         1        13
 14.        1         4         1        14
 15.        1         6         1        15
 16.        1         9         1        16
 17.        1         9         1        17
 18.        1        10         0        18
 19.        1        11         1        19
 20.        1        12         1        20
 21.        1        13         0        21
 22.        1        16         1        22
 23.        1        17         1        23
 24.        1        18         1        24
 25.        1        20         1        25
```


2. The next step is to define the data as "survival time data" using the "**stset**" command. Here the event is defined as drug failure.

```
. stset weeks, failure(failure==1) id(id)

                id:  id
     failure event:  failure == 1
obs. time interval:  (weeks[_n-1], weeks]
 exit on or before:  failure

------------------------------------------------------------------------------
       25  total obs.
        0  exclusions
------------------------------------------------------------------------------
       25  obs. remaining, representing
       25  subjects
       20  failures in single failure-per-subject data
      293  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        20
```

3. The **drug failure incidence rates** can be obtained in the new and standard treatment groups:

```
. stir trt

         failure _d:  failure == 1
   analysis time _t:  weeks
                 id:  id

note:  Exposed <-> trt==1 and Unexposed <-> trt==0

                 | trt                    |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
         Failure |        10          10  |         20
            Time |       159         134  |        293
-----------------+------------------------+------------
                 |                        |
  Incidence rate |  .0628931    .0746269  |   .0682594
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Inc. rate diff. |        -.0117338       |  -.0722224    .0487549
 Inc. rate ratio |         .8427673       |   .3148131    2.256122 (exact)
 Prev. frac. ex. |         .1572327       |  -1.256122    .6851869 (exact)
 Prev. frac. pop |         .0853242       |
                 +-------------------------------------------------
                    (midp)   Pr(k<=10) =                    0.3529 (exact)
                    (midp) 2*Pr(k<=10) =                    0.7058 (exact)
```

4. Next, we can obtain the **Kaplan-Meier estimates** of the survivor "drug-failure free" function for each treatment group: (Compare these to your solutions to **Self-Evaluation Problems Class 5**).

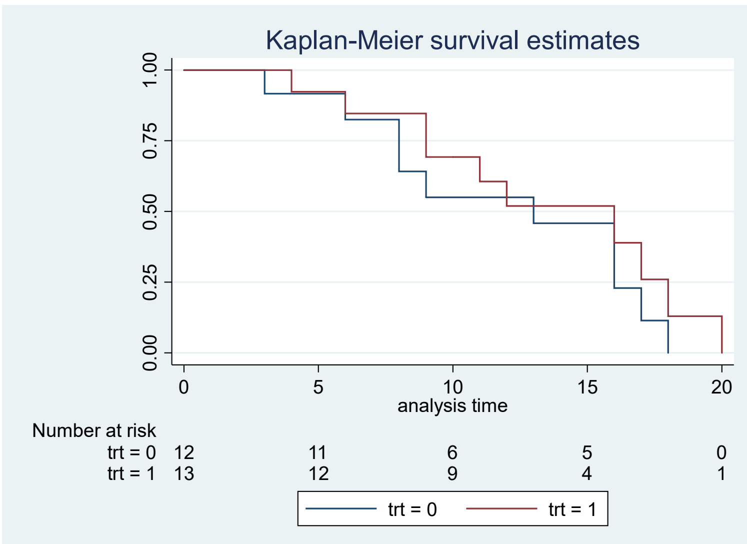
```
sts list if trt==0

         failure _d:  failure == 1
   analysis time _t:  weeks
                 id:  id

           Beg.          Net            Survivor      Std.
  Time    Total   Fail   Lost           Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
     3       12      1      0             0.9167    0.0798     0.5390    0.9878
     5       11      0      1             0.9167    0.0798     0.5390    0.9878
     6       10      1      0             0.8250    0.1128     0.4609    0.9533
     8        9      2      0             0.6417    0.1441     0.3022    0.8483
     9        7      1      0             0.5500    0.1499     0.2321    0.7829
    13        6      1      0             0.4583    0.1503     0.1689    0.7102
    15        5      0      1             0.4583    0.1503     0.1689    0.7102
    16        4      2      0             0.2292    0.1370     0.0382    0.5143
    17        2      1      0             0.1146    0.1061     0.0067    0.3917
    18        1      1      0             0.0000         .          .         .
-------------------------------------------------------------------------------
```

```
. sts list if trt==1

         failure _d:  failure == 1
   analysis time _t:  weeks
                 id:  id

           Beg.          Net            Survivor      Std.
  Time    Total   Fail   Lost           Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
     4       13      1      0             0.9231    0.0739     0.5664    0.9888
     6       12      1      0             0.8462    0.1001     0.5122    0.9591
     9       11      2      0             0.6923    0.1280     0.3734    0.8718
    10        9      0      1             0.6923    0.1280     0.3734    0.8718
    11        8      1      0             0.6058    0.1382     0.2943    0.8143
    12        7      1      0             0.5192    0.1430     0.2246    0.7500
    13        6      0      1             0.5192    0.1430     0.2246    0.7500
    14        5      0      1             0.5192    0.1430     0.2246    0.7500
    16        4      1      0             0.3894    0.1554     0.1152    0.6626
    17        3      1      0             0.2596    0.1482     0.0454    0.5553
    18        2      1      0             0.1298    0.1180     0.0076    0.4260
    20        1      1      0             0.0000         .          .         .
```


5. The "survival curves" by treatment group can be plotted along with a risk table:
   Based upon the plot of the Kaplan-Meier curves for each treatment group, which treatment, if any, should be preferred?

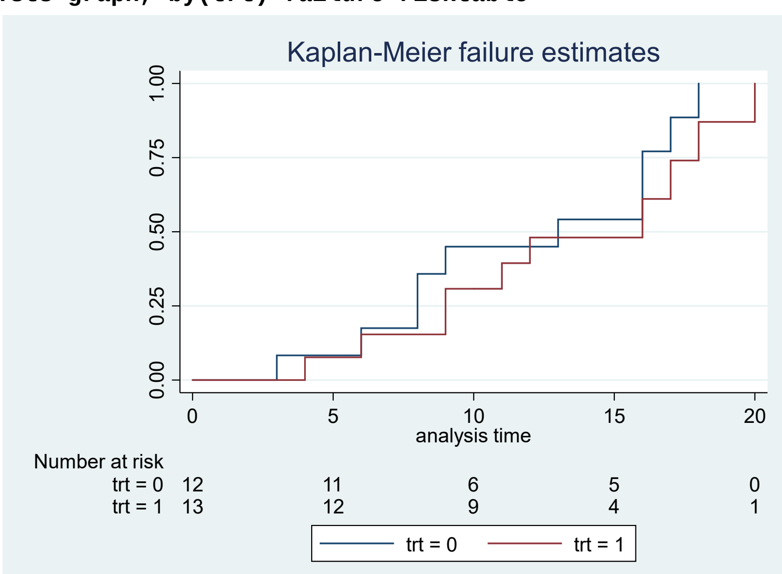
There is much overlap between the two curves and the sample sizes are very small. The curves do not appear to be distinguishable from each other.

```
.sts graph, by(trt) risktable
```



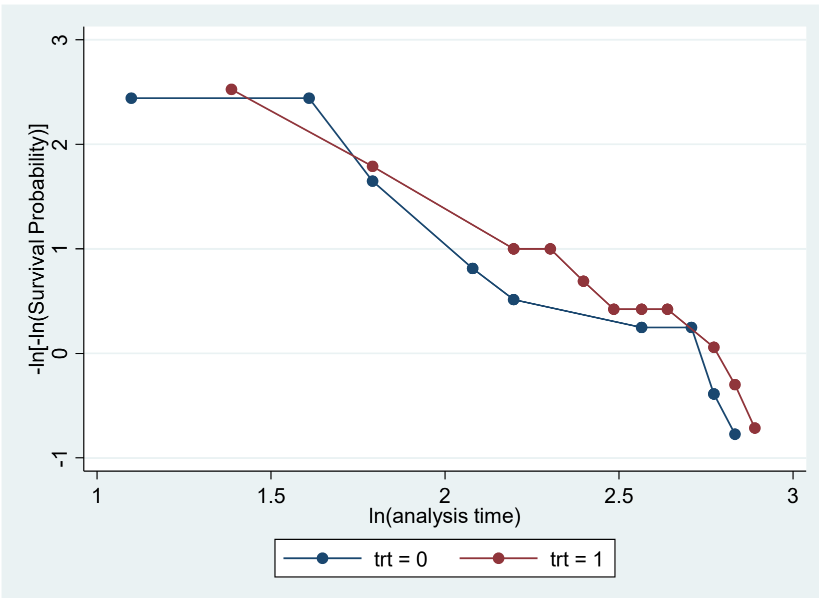
6. The cumulative hazard of drug failure by treatment group can be plotted using the "failure" option:

```
.sts graph, by(trt) failure risktable
```



7. What do you observe in the **complementary log log (CLL) plot** below?

```
.stphplot, by(trt)
```



We are hoping to observe a straight line relationship between the CLL function and log t. The sample sizes are very small, which makes the assessment difficult. However, a straight line relationship appears to hold for certain values of t. However, the curves cross during the early time period.

Also, the lines appear parallel for the two treatment groups; this would support the proportional hazards assumption that the ratio of hazards (comparing the two treatments) is constant (same) regardless of time t.

8. The l**og-rank statistic** can be calculated to test the null hypothesis of no difference in overall survival by treatment.

```
.sts test trt

         failure _d:  failure == 1
   analysis time _t:  weeks
                 id:  id
```



```
Log-rank test for equality of survivor functions

      |   Events         Events
trt   |  observed       expected
------+-------------------------
0     |        10           8.27
1     |        10          11.73
------+-------------------------
Total |        20          20.00

            chi2(1) =       0.76
            Pr>chi2 =     0.3831
```

9. Set up the log-rank statistic to test whether overall drug failure differs between the two treatments. **The log-rank test statistic can be computed by hand from the 2x2 tables based on each observed event time. In lab, you could split into groups, each group calculating the pieces needed from one 2x2 table.**

where

$$\chi^2_{LR} = \frac{\left[\sum_j (a_j - E(a_j))\right]^2}{\sum_j \mathrm{Var}(a_j)} \text{ where } E(a_j) = d_j \cdot \frac{n_{ja}}{n_j} \text{ and } \mathrm{Var}(a_j) = \frac{d_j(n_j - d_j)n_{ja}n_{jb}}{n_j^2(n_j - 1)}$$

	Event	No Event	Total
Standard Trt	a_j		n_{ja}
New Trt	c_j		n_{jb}
Total	d_j		n_j

Observed Event Times:
Time=j = 3 weeks:

	Event	No Event	Total
Standard Trt	1	11	12
New Trt	0	13	13
Total	1	24	25

a_j =1 $E(a_j)$= (1)(12)/25 a_j-$E(a_j)$ = 13/25 $Var(a_j)$=(1)(24)(12)(13)/[25^2(24)] = 0.2496

Time=4 weeks:

	Event	No Event	Total
Standard Trt	0	11	11
New Trt	1	12	13
Total	1	23	24

a_j =0 E(a_j)= (1)(11)/24 a_j-E(a_j) = -11/24 Var(a_j)=(1)(23)(11)(13)/[24^2(23)] = 0.2483

Time=6 weeks:

	Event	No Event	Total
Standard Trt	1	9	10
New Trt	1	11	12
Total	2	20	22

a_j =1 E(a_j)= (2)(10)/22 a_j-E(a_j) = 2/22 Var(a_j)=(2)(20)(10)(12)/[22^2(21)] = 0.4723

Time=8 weeks:

	Event	No Event	Total
Standard Trt	2	7	9
New Trt	0	11	11
Total	2	18	20

a_j =2 E(a_j)= (2)(9)/20 a_j-E(a_j) = 22/20 Var(a_j)=(2)(18)(9)(11)/[20^2(19)] = 0.4689

Time=9 weeks:

	Event	No Event	Total
Standard Trt	1	6	7
New Trt	2	9	11
Total	3	15	18

a_j =1 E(a_j)= (3)(7)/18 a_j-E(a_j) = 3/18 Var(a_j)=(3)(15)(7)(11)/[18^2(17)] = 0.6291

Time=11 weeks:

	Event	No Event	Total
Standard Trt	0	6	6
New Trt	1	7	8
Total	1	13	14

a_j =0 E(a_j)= (1)(6)/14 a_j-E(a_j) = -6/14 Var(a_j)=(1)(13)(6)(8)/[14^2(13)] = 0.2449

Time=12 weeks:

	Event	No Event	Total
Standard Trt	0	6	6
New Trt	1	6	7
Total	1	12	13

a_j =0 $E(a_j)$= (1)(6)/13 a_j-$E(a_j)$ = -6/13 $Var(a_j)$=(1)(12)(6)(7)/[13^2(12)] = 0.2485

Time=13 weeks:

	Event	No Event	Total
Standard Trt	1	5	6
New Trt	0	6	6
Total	1	11	12

a_j =1 $E(a_j)$= (1)(6)/12 a_j-$E(a_j)$ = 6/12 $Var(a_j)$=(1)(11)(6)(6)/[12^2(11)] = 0.2500

Time=16 weeks:

	Event	No Event	Total
Standard Trt	2	2	4
New Trt	1	3	4
Total	3	5	8

a_j =2 $E(a_j)$= (3)(4)/8 a_j-$E(a_j)$ = 4/8 $Var(a_j)$=(3)(5)(4)(4)/[8^2(7)] = 0.5357

Time=17 weeks:

	Event	No Event	Total
Standard Trt	1	1	2
New Trt	1	2	3
Total	2	3	5

a_j =1 $E(a_j)$= (2)(2)/5 a_j-$E(a_j)$ = 1/5 $Var(a_j)$=(2)(3)(2)(3)/[5^2(4)] = 0.3600

Time=18 weeks:

	Event	No Event	Total
Standard Trt	1	0	1
New Trt	1	1	2
Total	2	1	3

a_j =1 $E(a_j)$= (2)(1)/3 a_j-$E(a_j)$ = 1/3 $Var(a_j)$=(2)(1)(1)(2)/[3^2(2)] = 0.2222

Time=20 weeks:

	Event	No Event	Total
Standard Trt	0	0	0
New Trt	1	0	1
Total	1	0	1

a_j =0 $E(a_j)$= (1)(0)/1=0 a_j-$E(a_j)$ = 0 $Var(a_j)$=(1)(0)(0)(1)/[1^2(0)] = undefined

Thus, we have:

$$\chi^2_{LR} = \frac{\left[\sum_j (a_j - E(a_j))\right]^2}{\sum_j \mathrm{Var}(a_j)}$$

$$= \frac{\left[(13/25)-(11/24)+(2/22)+(22/20)-(3/18)-(6/14)-(6/13)+(6/12)+(4/8)+(1/5)+(1/3)\right]^2}{(.2496+.2483+.4723+.4689+.6291+.2449+.2485+.2500+.5357+.3600+.2222)} = 0.7609$$

where p > 0.05 and we would fail to reject the null hypothesis of equal survival in both groups

(e.g. Ho: overall survival in the std treatment = overall survival in the new treatment).

Compare your calculation to that obtained by `Stata` above in step 8.

