Statistics 641, Fall 2014
Homework #3
Solutions

1. Suppose that we have 20 patients, 10 per treatment group, and we observe the following survival times:

A:	8+, 11+, 16+, 18+, 23, 24, 26, 28, 30, 31
B:	9, 12, 13, 14, 14, 16, 19+, 22+, 23+, 29+

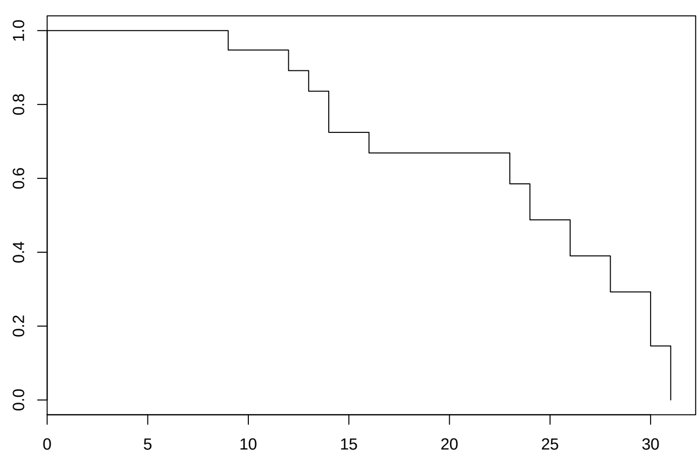
where the '+' indicates a censored observation.

For each of the following, please show your work (e.g., provide tables such as those on pages 5 and 8 of lecture 6, or page 7 of lecture 7). You may use software to check your work.

(a) Compute and plot the Kaplan-Meier estimate of survival, $\widehat{S}(t)$ for the two groups combined.

Letting n_j be the number at risk at each time, d_j the number of events, and $\widehat{S}(t_j)$ estimate of the survivor function.

t_j	n_j	d_j	$1 - d_j/n_j$	$\widehat{S}(t_j)$	$d_j/n_j(n_j - d_j)$	$\sum d_j/n_j(n_j - d_j)$	$\text{Var}(\widehat{S}(t))$	95%CI
0	20	–	–	1.00	0	0	–	
9	19	1	18/19	0.947	1/19(19-1) = 0.0029	0.0029	0.0026	(0.852, 1.000)
12	17	1	16/17	0.892	1/17(17-1) = 0.0037	0.0066	0.0052	(0.760, 1.000)
13	16	1	15/16	0.836	1/16(16-1) = 0.0042	0.0108	0.0075	(0.682, 1.000)
14	15	2	13/15	0.724	2/15(15-2) = 0.0103	0.0210	0.0110	(0.545, 0.963)
16	13	1	12/13	0.669	1/13(13-1) = 0.0064	0.0274	0.0123	(0.483, 0.925)
23	8	1	7/8	0.585	1/8(8-1) = 0.0179	0.0453	0.0155	(0.386, 0.888)
24	6	1	5/6	0.488	1/6(6-1) = 0.0333	0.0786	0.0187	(0.281, 0.845)
26	5	1	4/5	0.390	1/5(5-1) = 0.0500	0.1286	0.0196	(0.193, 0.788)
28	4	1	3/4	0.293	1/4(4-1) = 0.0833	0.2120	0.0181	(0.119, 0.721)
30	2	1	1/2	0.146	1/2(2-1) = 0.5000	0.7120	0.0152	(0.028, 0.765)
31	1	1	0/1	0.000	1/1(1-1) = ∞	∞	∞	(0, –)



(b) Compute the variance of $\widehat{S}(t)$ and a 95% confidence interval.

The variance of $\log(\widehat{S}(t))$ is $\sum_{j:t_j \le t} \frac{d_j}{n_j(n_j - d_j)}$. The table above shows the terms $d_j/n_j(n_j - d_j)$ and the cumulative sum. By the delta-method, $v(t) = \mathrm{Var}\big(\widehat{S}(t)\big) = \widehat{S}(t)^2 \mathrm{Var}\big(\log(\widehat{S}(t))\big)$, shown in the next column. The 95%-CI is computed on the $\log(S(t))$ scale and converted back:

$$\begin{aligned} \text{Upper limit} &= \widehat{S}(t) e^{1.96\sqrt{v(t)}} \\ \text{Lower limit} &= \widehat{S}(t) e^{-1.96\sqrt{v(t)}} \end{aligned}$$

(c) Assess equality of treatments using the log-rank test and the Gehan-Wilcoxon test.

	t_j	d_{j1}	n_{j1}	d_{j2}	n_{j2}	$n_{j1}+n_{j2}$ $(=w_j)$	$E[d_{j1}]$	$\mathrm{Var}(d_{j1})$
	9	0	9	1	10	19	$1 \times 9/19$	$1 \times 9 \times 10 \times 18/19^2 \times 18$
	12	0	8	1	9	17	$1 \times 8/17$	$1 \times 8 \times 9 \times 16/17^2 \times 16$
	13	0	8	1	8	16	$1 \times 8/16$	$1 \times 8 \times 8 \times 15/16^2 \times 15$
	14	0	8	2	7	14	$2 \times 8/15$	$2 \times 8 \times 7 \times 13/15^2 \times 14$
	16	0	8	1	5	13	$1 \times 8/13$	$1 \times 8 \times 5 \times 12/13^2 \times 12$
	23	1	6	0	2	8	$1 \times 6/8$	$1 \times 6 \times 2 \times 7/8^2 \times 7$
	24	1	5	0	1	6	$1 \times 5/6$	$1 \times 5 \times 1 \times 5/6^2 \times 5$
	26	1	4	0	1	5	$1 \times 4/5$	$1 \times 4 \times 1 \times 4/5^2 \times 4$
	28	1	3	0	1	4	$1 \times 3/4$	$1 \times 3 \times 1 \times 3/4^2 \times 3$
	30	1	2	0	0	2	$1 \times 2/2$	$1 \times 2 \times 0 \times 1/2^2 \times 1$
	31	1	1	0	0	1	$1 \times 1/1$	$1 \times 1 \times 0 \times 0/1^2 \times 0$
$\sum(\cdot)$		6					8.26	2.12
$\sum w_j(\cdot)$		26					70	394

We have for the unweighted log-rank:

$$\frac{(6 - 8.26)^2}{2.12} = 2.41$$

and for the Gehan-Wilcoxon-weighted log-rank:

$$\frac{(26 - 70)^2}{394} = 4.91$$

2. For the following use the data from the file `data3.csv`

The variables in the dataset are:

`trt` Treatment group (0/1)

`days`	Follow-up time in days
`status`	censoring/failure indicator (1=failure, 0=censored)
`sex`	Sex (1=Male, 2=Female)
`age`	Age at baseline in years

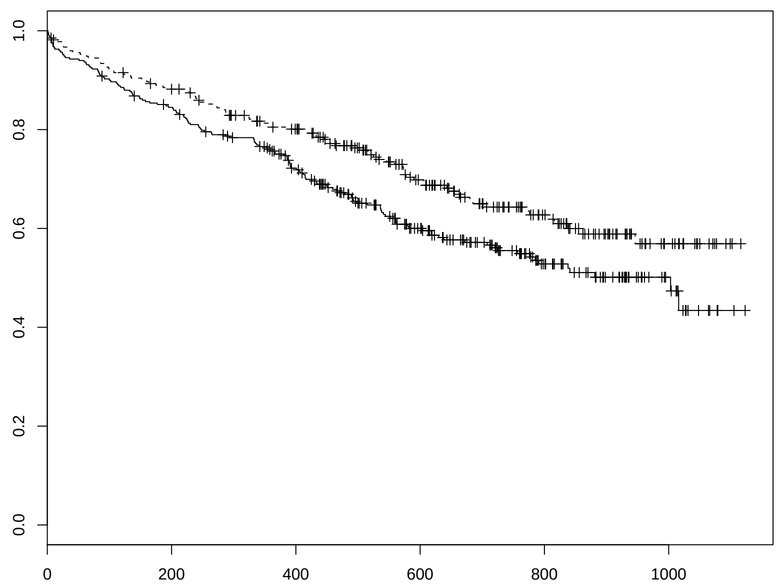
Let H_0 be the null hypothesis that there is no difference in survival by treatment.

First, read data into R:

```
> D <- read.csv("hw1.csv")
```

(a) Plot the Kaplan-Meier estimates of event-free survival by treatment group.

```
> plot(survfit(Surv(days,status)~trt, data=D), lty=1:2)
```



(b) Test H_0 using the Wald, score and likelihood ratio tests.

Fitting a Cox proportional hazards model,

```
> summary(coxph(Surv(days,status)~trt, data=D))
Call:
coxph(formula = Surv(days, status) ~ trt, data = D)

  n= 622

       coef exp(coef) se(coef)     z Pr(>|z|)
trt -0.3231    0.7239   0.1335 -2.42   0.0155 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
    exp(coef) exp(-coef) lower .95 upper .95
trt    0.7239      1.381    0.5573    0.9404

Rsquare= 0.01   (max possible= 0.99 )
Likelihood ratio test= 5.98  on 1 df,   p=0.01445
Wald test            = 5.86  on 1 df,   p=0.01552
Score (logrank) test = 5.91  on 1 df,   p=0.01507
```

the Wald statistic is $Z = -2.42$ (or the chi-square statistic is $5.86{=}Z^2$, with 1 DF), and the likelihood ratio statistic is 5.98 (chi-square with 1 DF). The score test is the log-rank which yields a statistic of 5.91. All are close together and provide modest evidence of a difference between treatment groups.

(c) Provide a summary and 95% confidence interval for the observed treatment difference.

The hazard ratio estimate from the Cox-model above is HR = 0.7239 (0.5573, 0.9404)

(d) Test H_0 adjusted for age and sex using the Wald and likelihood ratio tests.

First fit the Cox-model including age and sex effects:

```
> summary(coxph(Surv(days,status)~trt + age + sex, data=D))
Call:
coxph(formula = Surv(days, status) ~ trt + age + sex, data = D)

  n= 622

         coef exp(coef)  se(coef)      z Pr(>|z|)
trt -0.337584  0.713492  0.133585 -2.527  0.01150 *
age  0.021732  1.021970  0.007538  2.883  0.00394 **
sex -0.082668  0.920657  0.173959 -0.475  0.63463
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    exp(coef) exp(-coef) lower .95 upper .95
trt    0.7135     1.4016    0.5491     0.927
age    1.0220     0.9785    1.0070     1.037
sex    0.9207     1.0862    0.6547     1.295

Rsquare= 0.023   (max possible= 0.99 )
Likelihood ratio test= 14.54  on 3 df,   p=0.002257
Wald test            = 14.23  on 3 df,   p=0.002613
Score (logrank) test = 14.31  on 3 df,   p=0.002512
```

The Z statistic from the Wald test is $Z = -2.527$. To derive the likelihood ratio test we can simply take the difference between the likelihood ratio statistics for the models with and without treatment, or we can use the anova function:

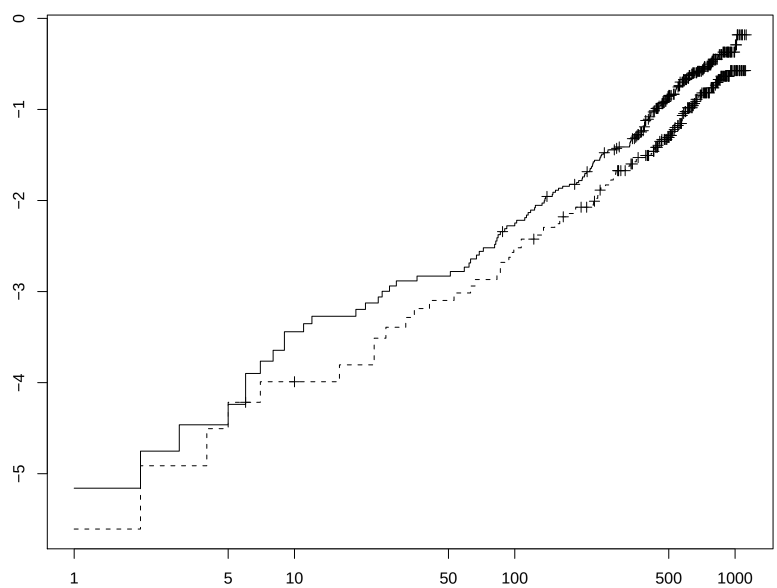
```
> anova(coxph(Surv(days,status)~ age + sex, data=D),
+    coxph(Surv(days,status)~trt + age + sex, data=D), test="Chis")
Analysis of Deviance Table
 Cox model: response is  Surv(days, status)
```

```
 Model 1: ~ age + sex
 Model 2: ~ trt + age + sex
   loglik  Chisq Df P(>|Chi|)
1 -1426.1
2 -1422.8 6.5277  1   0.01062 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The adjusted likelihood-ratio statistic is 6.5277.

(e) Assess whether the proportional hazards assumption for the treatment difference is reasonable.

```
> plot(survfit(Surv(days,status)~trt, data=D), lty=1:2, fun="cloglog")
```



These curves remain roughly the same distance apart for the portion where they are most stable—there is no evidence from the plot that the PH assumption does not hold. Using the `cox.zph` function for the unadjusted and adjusted models, we have

```
> cox.zph(coxph(Surv(days,status)~ trt, data=D))
       rho chisq     p
trt 0.0279 0.185 0.667
> cox.zph(coxph(Surv(days,status)~ trt + age + sex, data=D))
          rho   chisq     p
trt    0.0296 0.20868 0.648
age    0.0024 0.00132 0.971
sex    0.0154 0.05721 0.811
GLOBAL     NA 0.27726 0.964
```

Neither suggest that there is evidence the PH assumption fails.

In SAS, the PH test can be run as follows:

```
proc import datafile="hw1.csv" dbms=csv out=hw1 ;

proc phreg  data = hw1;
   model days*status(0) = trt  daystrt ;
   daystrt = trt*log(days);
```

```
             Analysis of Maximum Likelihood Estimates

              Parameter   Standard                           Hazard
Variable DF    Estimate      Error Chi-Square Pr > ChiSq      Ratio

trt       1    -0.43342    0.51810     0.6998     0.4028      0.648
daystrt   1     0.02098    0.09502     0.0488     0.8252      1.021
```

The Wald chi-square statistic is 0.0488 (p=.82), so, again, there is no evidence that the PH assumption fails. (Note that the estimate for treatment in the above is not meaningful because of the presence of the time-treatment interaction term.)