Statistics 641, Spring 2018
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.

   (a) Below is the output from the Kaplan-Meier estimate of overall survival, ignoring treatment group. "By hand," find the values that are missing and indicated by dashes (e.g., "—"). Note that you can use `survfit` to check your final answers.

```
> summary(survfit(Surv(time,status)~1, data=data1, conf.type="log-log"))
. . .
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    9     19       1    0.947  0.0512       0.6812        0.992
   12     17       1    0.892  0.0724       0.6315        0.972
   13     16       1    0.836  0.0867       0.5727        0.944
   14     15       2    0.724  0.1050       0.4591        0.875
   16     --       -    -----  ------       ------        -----
   23     --       -    -----  ------       ------        -----
   24      6       1    0.488  0.1367       0.2136        0.716
  . . .
```

From `survfit`:

```
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
. . .
   14     15       2    0.724  0.1050       0.4591        0.875
   16     13       1    0.669  0.1108       0.4059        0.836
   23      8       1    0.585  0.1245       0.3113        0.782
   24      6       1    0.488  0.1367       0.2136        0.716
. . .
```

The number at risk at time 16 is 13, and there is one failure at time 16. Therefore survival at time 16 is

$$0.724 \times \left(1 - \frac{1}{13}\right) = .669$$

and $\widehat{\Lambda}(16) = -\log(0.669) = 0.40197$. From the table above, the variance of $\widehat{S}(14)$ is $0.1050^2 = 0.011025$, and therefore the variance of $\widehat{\Lambda}(14)$ is $0.011025/.724^2 = 0.021033$. Therefore

$$\text{Var}\left[\widehat{\Lambda}(16)\right] = 0.021033 + \frac{1}{13 \times 12} = 0.027443$$

and the standard error of $\widehat{S}(16)$ is $.669 \times \sqrt{0.027443} = 0.1108$. The confidence intervals are calculated on the $\log \Lambda(t)$ scale ('`conf.type="log-log"`'), so applying the delta-method,

$$\text{Var}\left[\log \Lambda(16)\right] = \frac{0.027443}{0.40197^2} = 0.1698$$

and a 95% CI is $\log 0.40197 \pm 1.96 \times \sqrt{.1698} = (-1.719, -0.1037)$. Transforming back to the $S(t)$ scale, we get $\exp(-\exp(-0.1037)) = 0.4059$ and $\exp(-\exp(-1.1719)) = 0.8359$. Similarly, survival at time 23 is

$$0.669 \times \left(1 - \frac{1}{8}\right) = .585,$$

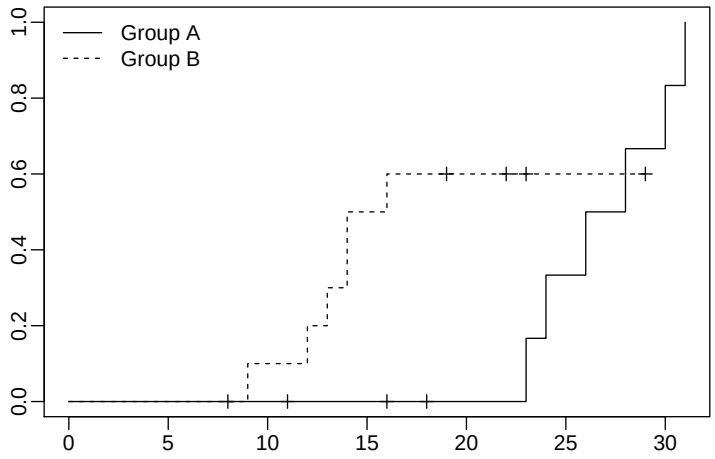
and $\widehat{\Lambda}(23) = -\log(0.585) = 0.53614$.

$$\text{Var}\left[\widehat{\Lambda}(23)\right] = 0.027443 + \frac{1}{8 \times 7} = 0.04530$$

and the standard error of $\widehat{S}(23)$ is $.585 \times \sqrt{0.04530} = 0.1245102$. The CI for $\log \Lambda(23)$ is $\log 0.53614 \pm 1.96 \times \sqrt{0.04530}/0.53614 = (-1.4014, 0.1547)$ and the CI for $\widehat{S}(23)$ is (0.3113,0.782).

(b) Plot cumulative mortality by treatment.

```
> plot(survfit(Surv(t,d)~1, data=D), fun="event")
# I don't like boxes around legends, so set bty="n"
> legend("topleft", bty="n", lty=1:2, c("Group A","Group B"))
```



(c) Below is table used for computing the (unweighted) log-rank and Gehan-Wilcoxon (GW) tests for equality of survival between treatments. Find the values that are missing and indicated by a dash (—). Calculate the log-rank and Gehan-Wilcoxon chi-square statistics. Note that you can use `survdiff` to check your final answers.

	t_j	d_{j1}	n_{j1}	d_{j2}	n_{j2}	$n_{j1}+n_{j2}$	$E[d_{j1}]$	$\mathrm{Var}(d_{j1})$
	9	0	9	1	10	19	0.474	0.2493
	12	0	8	1	9	17	0.471	0.2491
	13	0	8	1	8	16	0.500	0.2500
	14	—	—	—	—	—	—	—
	16	0	8	1	5	13	0.615	0.2367
	23	—	—	—	—	—	—	—
	24	1	5	0	1	6	0.833	0.1389
	26	1	4	0	1	5	0.800	0.1600
	28	1	3	0	1	4	0.750	0.1875
	30	—	—	—	—	—	—	—
	31	—	—	—	—	—	—	—
log-rank $\sum$		—					—	—
GW$\sum$		—					—	—

First get some summary counts via `survfit`:

```
> summary(survfit(Surv(t,d)~z, data=data1),
+    time=sort(unique(data1$t[data1$d==1])))
                z=A
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
. . .
   13      8       0    1.000   0.000       1.0000        1.000
   14      8       0    1.000   0.000       1.0000        1.000
   16      8       0    1.000   0.000       1.0000        1.000
   23      6       1    0.833   0.152       0.5827        1.000
   24      5       1    0.667   0.192       0.3786        1.000
. . .
   30      2       1    0.167   0.152       0.0278        0.997
   31      1       1    0.000     NaN           NA           NA

                z=B
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
. . .
   14      7       2      0.5  0.1581        0.269        0.929
   16      5       1      0.4  0.1549        0.187        0.855
   23      2       0      0.4  0.1549        0.187        0.855
. . .
   26      1       0      0.4  0.1549        0.187        0.855
   28      1       0      0.4  0.1549        0.187        0.855
```

The completed table is:

	t_j	d_{j1}	n_{j1}	d_{j2}	n_{j2}	$n_{j1}+n_{j2}$	$E[d_{j1}]$	$\mathrm{Var}(d_{j1})$
	9	0	9	1	10	19	0.474	0.2493
	12	0	8	1	9	17	0.471	0.2491
	13	0	8	1	8	16	0.500	0.2500
	14	0	8	2	7	15	1.067	0.4622
	16	0	8	1	5	13	0.615	0.2367
	23	1	6	0	2	8	0.750	0.1875
	24	1	5	0	1	6	0.833	0.1389
	26	1	4	0	1	5	0.800	0.1600
	28	1	3	0	1	4	0.750	0.1875
	30	1	2	0	0	2	1	0
	31	1	1	0	0	1	1	0
log-rank $\sum$		6					8.26	2.1212
GW $\sum$		26					70	394

For example, the $E[d_{j1}]$ column for time 14 is $2 \times 8/15 = 1.067$ and the $\mathrm{Var}(d_{j1})$ column is $8 \times 7 \times 2 \times 13/15^2/14 = .4622$. The entries "log-rank $\sum$" row are the sums of the corresponding columns. The entries in the "GW $\sum$" row are the sums of the columns above multiplied either by the $n_{j1}+n_{j2}$ column (first two) or this column squared (last column). Note that for $E[d_{j1}]$ and $\mathrm{Var}(d_{j1})$, $n_{j1}+n_{j2}$ appears in the denominators, so there is cancellation and the computation is actually slightly simpler.

We have for the unweighted log-rank:

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

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

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

Check using `survdiff`:

```
> survdiff(Surv(t,d)~z, data=data1)
Call:
survdiff(formula = Surv(t, d) ~ z, data = data1)

     N Observed Expected (O-E)^2/E (O-E)^2/V
z=A 10        6     8.26     0.618      2.41
z=B 10        6     3.74     1.365      2.41
 Chisq= 2.4  on 1 degrees of freedom, p= 0.121
```

This the same as the unweighted log-rank above. `survdiff` doesn't do the Gehan-Wilcoxon test, so we don't have any easy way to check this using `R`.

(d) Suppose that we terminate follow-up at time 20, so that all times beyond 20 are censored at time 20. Perform the log-rank test for the difference between groups censored at time 20. Does this provide insight into the difference between the two tests from part (c)?

First create new column in dataset (`d20`) that makes all the events beyond time 20 into non-events. In principle, the corresponding times should also be changed to 20, but after the time of the last event (now time 16), there is no contribution to the test, so the times don't matter. There are lots of ways to do this. Here's a simple one:

```
> data1$d20 <- data1$d*(data1$t<=20)
> survdiff(Surv(t,d20)~z, data=data1)
Call:
survdiff(formula = Surv(t, d20) ~ z, data = data1)

     N Observed Expected (O-E)^2/E (O-E)^2/V
z=A 10        0     3.13      3.13      6.75
z=B 10        6     2.87      3.40      6.75

 Chisq= 6.8  on 1 degrees of freedom, p= 0.00936
```

Equivalently, we could use just the first 5 rows of the table above. The total number of failures in column d_{j1} is zero, the sum of the expected values is 3.127, and the sum of the variances is 1.447. The test statistic is

$$\frac{(0-3.127)^2}{1.447} = 6.76$$

Note that for $t <= 20$, all the failures are in group B, whereas for $t > 20$ all the failures are in group A. Consequently, for the original log-rank test, cumulative sums of observed minus expected grow in absolute value (the sign depends on which cell we pick for the observed), until time 16, then begin to shrink starting at time 23 when the first failure occurs in group A. The unweighted log-rank statistic gives equal weight to all events so the early difference favoring group A is partially canceled by the later difference favoring group B. The Gehan-Wilcoxon weighted log-rank statistic gives more weight to the early differences relative to the later differences, so there is less cancellation, and the test statistic is larger. Censoring at time 20 effectively gives zero weight to the events favoring group B, and so there is no cancellation and the test statistic is larger yet.

The two unweighted log-rank tests, one over the whole follow-up interval, and one restricted to $t \leq 20$, have clear interpretations. On the other hand, while the GW-weighted test gives high weight to early events and low weight to later events, it does so in an ad-hoc manner that has no obvious interpretation.

---

2. The Beta-blocker Heart Attack Trial (BHAT) was a randomized trial conducted between 1978 and 1980 and assessed the effect of propranolol on mortality in subjects who had experienced at least one MI. Data from BHAT are available in dataset "bhat.csv".

   The variables in the dataset are:

`trt`	Treatment group (0=placebo/1=propranolol)
`day`	Follow-up time in days
`status`	censoring/failure indicator (1=dead, 0=censored)

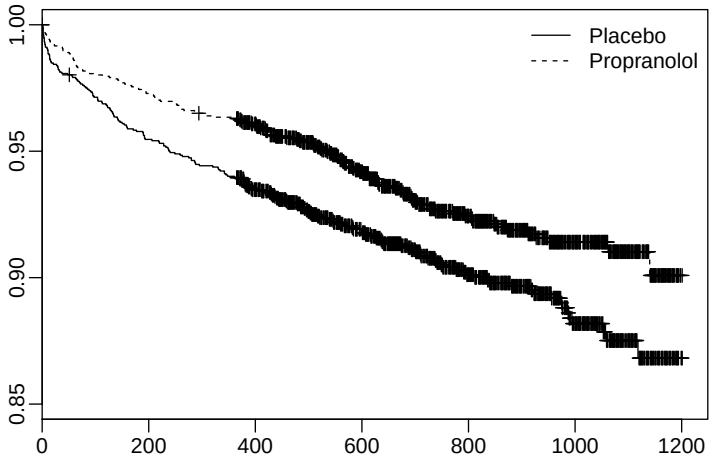
   Baseline (pre-treatment) variables:

`age`	Age at baseline in years
`sex`	Sex (1=Male, 2=Female)
`weightkg`	Baseline Weight in *kg*
`smoker`	Smoker
`sbp`	Systolic Blood Pressure
`dbp`	Diastolic Blood Pressure
`heartrate`	Heart Rate
`milocati`	Location of prior MI
`angina`	Suspected Angina Pectoris
`chf`	Suspected Chronic Heart Failure

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

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

```
> plot(survfit(Surv(day,status)~trt, data=bhat), lty=1:2, ylim=c(.85,1))
> legend("topright", bty="n", lty=1:2, c("Placebo","Propranolol"))
```



   (b) Compare treatment groups (unadjusted) using the log-rank test. Based on this test, find an estimate of the hazard ratio using the one-step estimator. Note that while `survdiff` doesn't print out the variance (Fisher information), it returns an object with components called `obs`, `exp`, and `var` (e.g., `survdiff(...)$var`).

```
>  survdiff(Surv(day,status)~trt, data=bhat)
Call:
survdiff(formula = Surv(day, status) ~ trt, data = bhat)
         N Observed Expected (O-E)^2/E (O-E)^2/V
trt=0 1921      188      162      4.25      8.43
trt=1 1916      138      164      4.18      8.43
Chisq= 8.4  on 1 degrees of freedom, p= 0.00369
```

There is "statistically significant" decrease in mortality in the propranolol group.

Rather than saving the `survdiff` output, Here's trick that uses the function `with` to do the one-step estimator.

```
>  with(survdiff(Surv(day,status)~trt, data=bhat), (obs-exp)[2]/var[1])
[1] -0.3216515   ##log HR
>  exp(with(survdiff(Surv(day,status)~trt, data=bhat), (obs-exp)[2]/var[1]))
[1] 0.7249508  ## HR
```

The one-step estimator suggests a 28% reduction in hazard with propranolol relative to placebo.

(c) Estimate the hazard ratio for treatment and test H_0 using the Wald and likelihood ratio tests. Compare to the one-step estimator from part (b).

Fitting a Cox proportional hazards model,

```
>  summary(coxph(Surv(day,status)~trt, data=bhat))
Call:
coxph(formula = Surv(day, status) ~ trt, data = bhat)
  n= 3837, number of events= 326
       coef exp(coef) se(coef)      z Pr(>|z|)
trt -0.3241    0.7232   0.1121 -2.891  0.00384 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    exp(coef) exp(-coef) lower .95 upper .95
trt    0.7232      1.383    0.5806    0.9009
. . .
Likelihood ratio test= 8.46  on 1 df,   p=0.003629
Wald test            = 8.36  on 1 df,   p=0.003841
Score (logrank) test = 8.43  on 1 df,   p=0.003689
```

The Wald statistic is $Z = -2.891$ (or the chi-square statistic is $8.36 = Z^2$, with 1 DF), and the likelihood ratio statistic is 8.46 (chi-square with 1 DF). The score test is the log-rank which yields a statistic of 8.43, same as in part (b).

The point estimate of the HR is 0.7232, quite close to the one-step estimate in part (b). (As expected, the MLE is slightly further from 1 than the one-step estimate.)

(d) Estimate the hazard ratio for treatment adjusted for important baseline variables and test H_0 using the Wald and likelihood ratio tests. Compare to the result from (b). Is CHF a confounder for the effect of treatment on the outcome?

There are lots of ways of deciding which are "important" baseline variables, and I'll accept any reasonable set. Here's a model with all the baseline variables:

```
> coxph(Surv(day,status)~trt + milocati + smoker + angina + sbp +
 + age + dbp + chf + sex + weightkg + heartrate, data=bhat)
Call:
coxph(formula = Surv(day, status) ~ trt + milocati + smoker +
    angina + sbp + age + dbp + chf + sex + weightkg + heartrate,
    data = bhat)


                        coef exp(coef) se(coef)     z       p
trt                  -0.32409   0.72318  0.11266 -2.88 0.00402
milocatiAnterior     -0.29460   0.74483  0.17743 -1.66 0.09685
milocatiInferior     -0.55045   0.57669  0.18287 -3.01 0.00261
milocatinon-BHAT MI  -0.24898   0.77959  0.23162 -1.07 0.28238
milocatinontransmrl  -0.36988   0.69082  0.19063 -1.94 0.05234
smokerYes             0.48878   1.63032  0.12054  4.05 5.0e-05
anginaUnknown        -0.17891   0.83618  0.58536 -0.31 0.75987
anginaYes             0.36782   1.44458  0.11421  3.22 0.00128
sbp                   0.00688   1.00690  0.00642  1.07 0.28404
age                   0.05189   1.05326  0.00820  6.33 2.4e-10
dbp                   0.00371   1.00372  0.00988  0.38 0.70708
chfUnknown            0.30347   1.35455  0.58381  0.52 0.60320
chfYes                0.81022   2.24839  0.14307  5.66 1.5e-08
sex                  -0.13691   0.87205  0.16056 -0.85 0.39382
weightkg             -0.00436   0.99565  0.00453 -0.96 0.33573
heartrate             0.01986   1.02006  0.00555  3.57 0.00035

Likelihood ratio test=148  on 16 df, p=0
n= 3817, number of events= 325
   (20 observations deleted due to missingness)
```

Note that 1) 20 observations have missing values, and 2) inspection of the data shows that these are all for the variable `weightkg`, and 3) `weightkg` has a large p-value, so we can consider it "unimportant" and remove it from the model. Fit model without `weightkg`, and use the `anova` function so assess the "importance" (sequentially) of both continuous and categorical variables:

```
>  cox1 <- coxph(Surv(day,status)~trt + milocati + smoker + angina +
+ sbp + age + dbp + chf + sex + heartrate, data=bhat)
> anova(cox1)
Analysis of Deviance Table
 Cox model: response is Surv(day, status)
Terms added sequentially (first to last)
```

```
           loglik   Chisq Df Pr(>|Chi|)
NULL      -2616.0
trt       -2611.8  8.4606  1  0.0036293 **
milocati  -2603.6 16.3153  4  0.0026240 **
smoker    -2602.2  2.8517  1  0.0912793 .
angina    -2593.3 17.7469  2  0.0001401 ***
sbp       -2588.7  9.2602  1  0.0023418 **
age       -2565.1 47.0624  1  6.876e-12 ***
dbp       -2564.5  1.3098  1  0.2524374
chf       -2548.6 31.7618  2  1.268e-07 ***
sex       -2548.6  0.0865  1  0.7686677
heartrate -2542.2 12.7240  1  0.0003610 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we were being more systematic, we might want to consider each variable adjusted for all the others, but for simplicity, I'll keep everything with $p < 0.1$ in this ANOVA. Note also that from the previous model, `angina` and `chf` have an "unknown" category. Since we're not really interested in the these variables as predictors, we can simply leave "unknown" as a separate category.

```
>  cox2 <- coxph(Surv(day,status)~trt + milocati + smoker + angina +
+ sbp + age +  chf + heartrate, data=bhat)
> anova(cox2)
Analysis of Deviance Table
 Cox model: response is Surv(day, status)
Terms added sequentially (first to last)

           loglik   Chisq Df Pr(>|Chi|)
NULL      -2616.0
trt       -2611.8  8.4606  1  0.0036293 **
milocati  -2603.6 16.3153  4  0.0026240 **
smoker    -2602.2  2.8517  1  0.0912793 .
angina    -2593.3 17.7469  2  0.0001401 ***
sbp       -2588.7  9.2602  1  0.0023418 **
age       -2565.1 47.0624  1  6.876e-12 ***
chf       -2549.4 31.5125  2  1.436e-07 ***
heartrate -2542.5 13.7509  1  0.0002087 ***
```

All variables reach at least $p < 0.1$.

```
> cox2
...
                        coef exp(coef) se(coef)     z       p
trt                  -0.31907   0.72682  0.11238 -2.84 0.00452
milocatiAnterior     -0.28288   0.75361  0.17721 -1.60 0.11043
milocatiInferior     -0.53751   0.58420  0.18241 -2.95 0.00321
milocatinon-BHAT MI  -0.24041   0.78630  0.23145 -1.04 0.29893
milocatinontransmrl  -0.35911   0.69830  0.18956 -1.89 0.05816
```

```
smokerYes            0.50130   1.65086  0.11930  4.20 2.6e-05
anginaUnknown       -0.16242   0.85008  0.58607 -0.28 0.78168
anginaYes            0.36218   1.43645  0.11381  3.18 0.00146
sbp                  0.00787   1.00790  0.00469  1.68 0.09308
age                  0.05277   1.05419  0.00783  6.74 1.6e-11
chfUnknown           0.27542   1.31709  0.58415  0.47 0.63729
chfYes               0.79701   2.21890  0.14273  5.58 2.4e-08
heartrate            0.02053   1.02074  0.00542  3.79 0.00015
Likelihood ratio test=147  on 13 df, p=0
n= 3837, number of events= 326
```

The MLE for the log-hazard ratio is -.31907, and HR = 0.7268, very similar to part (b). The Wald $Z = -2.84$, again very close the unadjusted test. To perform the likelihood ratio test, we need to compare the full model (`cox2`) to a model with the same set of baseline variables, but without treatment. We can do this directly, or we can fit a model in which `trt` is listed *last*, and use the `anova` function.

```
>  cox3 <- coxph(Surv(day,status)~milocati + smoker + angina + sbp + age +
+  chf + heartrate, data=bhat)
> cox3
. . .
Likelihood ratio test=139  on 12 df, p=0
> cox4 <- coxph(Surv(day,status)~milocati + smoker + angina + sbp + age +
+ chf + heartrate + trt, data=bhat)
> anova(cox4)
Analysis of Deviance Table
 Cox model: response is Surv(day, status)
Terms added sequentially (first to last)

           loglik   Chisq Df Pr(>|Chi|)
NULL      -2616.0
milocati  -2607.8 16.3696  4  0.0025613 **
smoker    -2606.4  2.7117  1  0.0996120 .
angina    -2597.5 17.8132  2  0.0001355 ***
sbp       -2593.2  8.6912  1  0.0031974 **
age       -2569.3 47.7479  1  4.847e-12 ***
chf       -2553.2 32.1512  2  1.043e-07 ***
heartrate -2546.6 13.3161  1  0.0002631 ***
trt       -2542.5  8.1594  1  0.0042838 **
>
```

From `anova(cox4)` the likelihood ratio test statistic is 8.1594 with 1 df. Comparing `cox3` to `cox2`, the difference in likelihood ratio statistics as given in the output above is $147 - 139 = 8$, but this is rounded off to integers. We can extract the log-likelihood ratios from the fitted models. The first entry is the log-likelihood for the null model (no predictors), while the second entry is the log-likelihood for the fitted model.

```
> cox2$loglik
[1] -2615.986 -2542.506
> cox3$loglik
```

```
[1] -2615.986 -2546.586
> -2*(cox3$loglik[2] - cox2$loglik[2])
[1] 8.159439
```

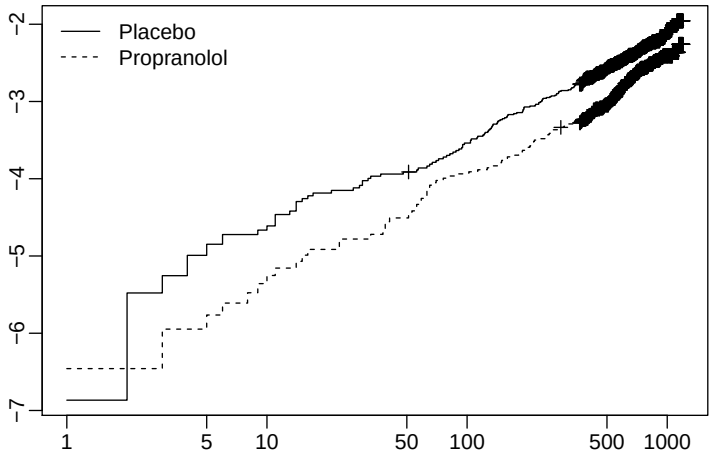
Matches the output from `anova`.

CHF can't be a confounder because the study is randomized.

(e) Assess whether the proportional hazards assumption for the model in part (d) is reasonable.

Graphically, we could look at unadjusted log-hazard versus log-time, for each treatment group, but this may not necessarily detect departures from proportionality in the *adjusted* model. One trick is to fit a model in which treatment is a stratification variable, so separate baseline hazards will be used for each group. The `basehaz` function extracts estimates of the baseline hazard for each treatment group.

```
> cox5 <- coxph(Surv(day,status)~strata(trt) + milocati + smoker + angina +
+ sbp + age +  chf + heartrate, data=bhat)
> blhaz <- basehaz(cox5)
## plot hazards separately for each treatment:  'with' lets us easily extract one
## treatment group at a time.
> with(subset(blhaz,strata=="trt=0"), plot(time,hazard, type="s",log="xy",
+ xlim=c(1,1200), ylim=c(.0008,.12)))
> with(subset(blhaz,strata=="trt=1"), lines(time,hazard, lty=2, col=2, type="s"))
```



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.

We can also use `cox.zph` with the model `cox2`:

```
> cox.zph(cox2)
                        rho    chisq       p
trt                 0.06256  1.26897 0.25996
milocatiAnterior    0.05238  0.92149 0.33708
milocatiInferior    0.04403  0.63500 0.42553
milocatinon-BHAT MI 0.03456  0.39795 0.52815
milocatinontransmrl 0.14926  7.29057 0.00693
```

```
smokerYes           -0.06073  1.18884 0.27556
anginaUnknown        0.06532  1.42162 0.23314
anginaYes           -0.00557  0.01027 0.91930
sbp                  0.02603  0.24343 0.62174
age                 -0.00403  0.00585 0.93903
chfUnknown          -0.08259  2.24240 0.13427
chfYes              -0.03548  0.41386 0.52002
heartrate           -0.12531  5.18353 0.02280
GLOBAL                    NA 21.82745 0.05809
```

Maybe there is non-proportionality for `milocati`, so stratify by this variable

```
>  cox6 <- coxph(Surv(day,status)~trt + strata(milocati) + smoker +
+ angina + sbp + age +  chf + heartrate, data=bhat)
> cox.zph(cox6)
                    rho     chisq      p
trt             0.06612  1.417639 0.2338
smokerYes      -0.05981  1.149683 0.2836
anginaUnknown   0.06737  1.512130 0.2188
anginaYes      -0.00789  0.020484 0.8862
sbp             0.02687  0.257855 0.6116
age            -0.00109  0.000426 0.9835
chfUnknown     -0.08444  2.348590 0.1254
chfYes         -0.03538  0.408394 0.5228
heartrate      -0.12801  5.377292 0.0204
GLOBAL               NA 11.649394 0.2338
```

The GLOBAL test suggests there is no remaining non-proportionality. Ignore $p = 0.02$ for `heartrate`. Non-proportionality with respect to baseline variables doesn't really matter anyway,

```
> cox6
                  coef exp(coef) se(coef)     z       p
trt            -0.31331   0.73102  0.11238 -2.79 0.00530
. . .
```

Coefficient for treatment is largely unaffected by baseline model.