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, 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.

```r
> summary(survfit(Surv(time,status)~1, data=data1, conf.type="log-log"))
```

<table>
<thead>
<tr>
<th>time</th>
<th>n.risk</th>
<th>n.event</th>
<th>survival</th>
<th>std.err</th>
<th>lower 95% CI</th>
<th>upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>19</td>
<td>1</td>
<td>0.947</td>
<td>0.0512</td>
<td>0.6812</td>
<td>0.992</td>
</tr>
<tr>
<td>12</td>
<td>17</td>
<td>1</td>
<td>0.892</td>
<td>0.0724</td>
<td>0.6315</td>
<td>0.972</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
<td>1</td>
<td>0.836</td>
<td>0.0867</td>
<td>0.5727</td>
<td>0.944</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>2</td>
<td>0.724</td>
<td>0.1050</td>
<td>0.4591</td>
<td>0.875</td>
</tr>
<tr>
<td>16</td>
<td>--</td>
<td>--</td>
<td>---</td>
<td>-------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>23</td>
<td>--</td>
<td>--</td>
<td>---</td>
<td>-------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>24</td>
<td>6</td>
<td>1</td>
<td>0.488</td>
<td>0.1367</td>
<td>0.2136</td>
<td>0.716</td>
</tr>
</tbody>
</table>

From `survfit`:

<table>
<thead>
<tr>
<th>time</th>
<th>n.risk</th>
<th>n.event</th>
<th>survival</th>
<th>std.err</th>
<th>lower 95% CI</th>
<th>upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>15</td>
<td>2</td>
<td>0.724</td>
<td>0.1050</td>
<td>0.4591</td>
<td>0.875</td>
</tr>
<tr>
<td>16</td>
<td>13</td>
<td>1</td>
<td>0.669</td>
<td>0.1108</td>
<td>0.4059</td>
<td>0.836</td>
</tr>
<tr>
<td>23</td>
<td>8</td>
<td>1</td>
<td>0.585</td>
<td>0.1245</td>
<td>0.3113</td>
<td>0.782</td>
</tr>
<tr>
<td>24</td>
<td>6</td>
<td>1</td>
<td>0.488</td>
<td>0.1367</td>
<td>0.2136</td>
<td>0.716</td>
</tr>
</tbody>
</table>

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

\[
\hat{S}(16) = 0.724 \times \left(1 - \frac{1}{13}\right) = 0.669
\]

and \(\hat{\Lambda}(16) = -\log(0.669) = 0.40197\). From the table above, the variance of \(\hat{S}(14)\) is \(0.1050^2 = 0.011025\), and therefore the variance of \(\hat{\Lambda}(14)\) is \(0.011025 / 0.724^2 = 0.021033\). Therefore

\[
\text{Var} \left(\hat{\Lambda}(16)\right) = 0.021033 + \frac{1}{13 \times 12} = 0.027443
\]
and the standard error of $\hat{S}(16)$ is $0.669 \times \sqrt{0.027443} = 0.1108$. 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 $\log 0.40197 \pm 1.96 \times \sqrt{0.1698} = (-1.719, -0.1037)$. 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 $\hat{\Lambda}(23) = -\log(0.585) = 0.53614$.

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

and the standard error of $\hat{S}(23)$ is $0.585 \times \sqrt{0.04530} = 0.1245102$. 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 $\hat{S}(23)$ is $(0.3113, 0.782)$.

(b) Plot cumulative mortality by treatment.

```r
> 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.
First get some summary counts via survfit:

```r
> summary(survfit(Surv(t,d)~z, data=data1),
+    time=sort(unique(data1$t[data1$d==1])))
```

### z=A

<table>
<thead>
<tr>
<th>time</th>
<th>n.risk</th>
<th>n.event</th>
<th>survival</th>
<th>std.err</th>
<th>lower 95% CI</th>
<th>upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>8</td>
<td>0</td>
<td>1.000</td>
<td>0.000</td>
<td>1.0000</td>
<td>1.000</td>
</tr>
<tr>
<td>14</td>
<td>8</td>
<td>0</td>
<td>1.000</td>
<td>0.000</td>
<td>1.0000</td>
<td>1.000</td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>0</td>
<td>1.000</td>
<td>0.000</td>
<td>1.0000</td>
<td>1.000</td>
</tr>
<tr>
<td>23</td>
<td>6</td>
<td>1</td>
<td>0.833</td>
<td>0.152</td>
<td>0.5827</td>
<td>1.000</td>
</tr>
<tr>
<td>24</td>
<td>5</td>
<td>1</td>
<td>0.667</td>
<td>0.192</td>
<td>0.3786</td>
<td>1.000</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>1</td>
<td>0.167</td>
<td>0.152</td>
<td>0.0278</td>
<td>0.997</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>1</td>
<td>0.000</td>
<td>NaN</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

### z=B

<table>
<thead>
<tr>
<th>time</th>
<th>n.risk</th>
<th>n.event</th>
<th>survival</th>
<th>std.err</th>
<th>lower 95% CI</th>
<th>upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>7</td>
<td>2</td>
<td>0.5</td>
<td>0.1581</td>
<td>0.269</td>
<td>0.929</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>1</td>
<td>0.4</td>
<td>0.1549</td>
<td>0.187</td>
<td>0.855</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>0</td>
<td>0.4</td>
<td>0.1549</td>
<td>0.187</td>
<td>0.855</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>0</td>
<td>0.4</td>
<td>0.1549</td>
<td>0.187</td>
<td>0.855</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>0</td>
<td>0.4</td>
<td>0.1549</td>
<td>0.187</td>
<td>0.855</td>
</tr>
</tbody>
</table>

log-rank $\sum \log-rank \sum GW\sum$

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[d_{j1}]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{Var}(d_{j1})$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$t_j$</th>
<th>$d_{j1}$</th>
<th>$n_{j1}$</th>
<th>$d_{j2}$</th>
<th>$n_{j2}$</th>
<th>$n_{j1} + n_{j2}$</th>
<th>$E[d_{j1}]$</th>
<th>$\text{Var}(d_{j1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>10</td>
<td>19</td>
<td>0.474</td>
<td>0.2493</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>17</td>
<td>0.471</td>
<td>0.2491</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>8</td>
<td>16</td>
<td>0.500</td>
<td>0.2500</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>13</td>
<td>0.615</td>
<td>0.2367</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0.833</td>
<td>0.1389</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0.800</td>
<td>0.1600</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0.750</td>
<td>0.1875</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The completed table is:

<table>
<thead>
<tr>
<th>(t_j)</th>
<th>(d_{j1})</th>
<th>(n_{j1})</th>
<th>(d_{j2})</th>
<th>(n_{j2})</th>
<th>(n_{j1} + n_{j2})</th>
<th>(E[d_{j1}])</th>
<th>(\text{Var}(d_{j1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>10</td>
<td>19</td>
<td>0.474</td>
<td>0.2493</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>9</td>
<td>17</td>
<td>0.471</td>
<td>0.2491</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>8</td>
<td>16</td>
<td>0.500</td>
<td>0.2500</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>15</td>
<td>1.067</td>
<td>0.4622</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>13</td>
<td>0.615</td>
<td>0.2367</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>0.750</td>
<td>0.1875</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0.833</td>
<td>0.1389</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0.800</td>
<td>0.1600</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0.750</td>
<td>0.1875</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| \(\log\text{-rank } \sum\) | 6 | 6.26 | 2.1212 | \(\text{GW } \sum\) | 26 | 70 | 394 |

For example, the \(E[d_{j1}]\) column for time 14 is \(2 \times 8/15 = 1.067\) and the \(\text{Var}(d_{j1})\) 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 \(\text{Var}(d_{j1})\), \(n_{j1} + n_{j2}\) 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 \texttt{survdiff}:

\begin{verbatim}
> 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
\end{verbatim}

This the same as the unweighted log-rank above. \texttt{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 \( t \leq 20 \), 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 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
- **miocati**: 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.

```r
> 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`).
\begin{verbatim}
> 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 \texttt{survdiff} output, Here's trick that uses the function \texttt{with} to do
the one-step estimator.

\begin{verbatim}
> 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
\end{verbatim}

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 the one-step from part (b).

Fitting a Cox proportional hazards model,

\begin{verbatim}
> 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
\end{verbatim}

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.)
\end{verbatim}
(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:

```r
> 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)

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>-0.32409</td>
<td>0.72318</td>
<td>0.11266</td>
<td>-2.88</td>
<td>0.00402</td>
</tr>
<tr>
<td>milocatiAnterior</td>
<td>-0.29460</td>
<td>0.74483</td>
<td>0.17743</td>
<td>-1.66</td>
<td>0.09685</td>
</tr>
<tr>
<td>milocatiInferior</td>
<td>-0.55045</td>
<td>0.57669</td>
<td>0.18287</td>
<td>-3.01</td>
<td>0.00261</td>
</tr>
<tr>
<td>milocatinon-BHAT MI</td>
<td>-0.24898</td>
<td>0.77959</td>
<td>0.23162</td>
<td>-1.07</td>
<td>0.28238</td>
</tr>
<tr>
<td>milocatinontransmrl</td>
<td>-0.36988</td>
<td>0.69082</td>
<td>0.19063</td>
<td>-1.94</td>
<td>0.05234</td>
</tr>
<tr>
<td>smokerYes</td>
<td>0.48878</td>
<td>1.63032</td>
<td>0.12054</td>
<td>4.05</td>
<td>5.0e-05</td>
</tr>
<tr>
<td>anginaUnknown</td>
<td>-0.17891</td>
<td>0.87205</td>
<td>0.16056</td>
<td>-1.11</td>
<td>0.26879</td>
</tr>
<tr>
<td>anginaYes</td>
<td>0.36782</td>
<td>1.44458</td>
<td>0.11421</td>
<td>3.22</td>
<td>0.00128</td>
</tr>
<tr>
<td>sbp</td>
<td>0.00688</td>
<td>1.00690</td>
<td>0.00642</td>
<td>1.07</td>
<td>0.28404</td>
</tr>
<tr>
<td>age</td>
<td>0.05189</td>
<td>1.05326</td>
<td>0.00820</td>
<td>6.33</td>
<td>2.4e-10</td>
</tr>
<tr>
<td>dbp</td>
<td>0.00371</td>
<td>1.00372</td>
<td>0.00988</td>
<td>0.38</td>
<td>0.70708</td>
</tr>
<tr>
<td>chfUnknown</td>
<td>0.30347</td>
<td>1.35455</td>
<td>0.58381</td>
<td>0.52</td>
<td>0.60320</td>
</tr>
<tr>
<td>chfYes</td>
<td>0.81022</td>
<td>2.24839</td>
<td>0.14307</td>
<td>5.66</td>
<td>1.5e-08</td>
</tr>
<tr>
<td>sex</td>
<td>-0.13691</td>
<td>0.87205</td>
<td>0.16056</td>
<td>-0.85</td>
<td>0.39382</td>
</tr>
<tr>
<td>weightkg</td>
<td>-0.00436</td>
<td>0.99565</td>
<td>0.00453</td>
<td>-0.96</td>
<td>0.33573</td>
</tr>
<tr>
<td>heartrate</td>
<td>0.01986</td>
<td>1.02006</td>
<td>0.00555</td>
<td>3.57</td>
<td>0.00035</td>
</tr>
</tbody>
</table>

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:

```r
> 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 these variables as predictors, we can simply leave “unknown” as a separate category.

```r
> 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 ***
  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 ***
---
All variables reach at least $p < 0.1$.
> cox2
...

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>-0.31907</td>
<td>0.72682</td>
<td>0.11238</td>
<td>-2.84</td>
<td>0.00452</td>
</tr>
<tr>
<td>milocatiAnterior</td>
<td>-0.28288</td>
<td>0.75361</td>
<td>0.17721</td>
<td>-1.60</td>
<td>0.11043</td>
</tr>
<tr>
<td>milocatiInferior</td>
<td>-0.53751</td>
<td>0.58420</td>
<td>0.18241</td>
<td>-2.95</td>
<td>0.00321</td>
</tr>
<tr>
<td>milocatinon-BHAT MI</td>
<td>-0.24041</td>
<td>0.78630</td>
<td>0.23145</td>
<td>-1.04</td>
<td>0.29893</td>
</tr>
<tr>
<td>milocatinontransmrl</td>
<td>-0.35911</td>
<td>0.69830</td>
<td>0.18956</td>
<td>-1.89</td>
<td>0.05816</td>
</tr>
</tbody>
</table>
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
Matches the output from `anova`.

CHF can’t be a confounder because the study is randomized. (Although I didn’t explicitly tell you this!)

(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.

```r
> 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`:

```r
> cox.zph(cox2)

<table>
<thead>
<tr>
<th></th>
<th>rho</th>
<th>chisq</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0.06256</td>
<td>1.26897</td>
<td>0.25996</td>
</tr>
<tr>
<td>milocatiAnterior</td>
<td>0.05238</td>
<td>0.92149</td>
<td>0.33708</td>
</tr>
<tr>
<td>milocatiInferior</td>
<td>0.04403</td>
<td>0.63500</td>
<td>0.42553</td>
</tr>
<tr>
<td>milocatinon-BHAT MI</td>
<td>0.03456</td>
<td>0.39795</td>
<td>0.52815</td>
</tr>
</tbody>
</table>
```
Maybe there is non-proportionality for milocatinontransmrl, so stratify by this variable.

```r
> cox6 <- coxph(Surv(day,status)~trt + strata(milocatinontransmrl) + smoker +
+ angina + sbp + age + chf + heartrate, data=bhat)
> cox.zph(cox6)
```

<table>
<thead>
<tr>
<th></th>
<th>rho</th>
<th>chisq</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>0.06612</td>
<td>1.417639</td>
<td>0.2338</td>
</tr>
<tr>
<td>smokerYes</td>
<td>-0.05981</td>
<td>1.149683</td>
<td>0.2836</td>
</tr>
<tr>
<td>anginaUnknown</td>
<td>0.06737</td>
<td>1.512130</td>
<td>0.2188</td>
</tr>
<tr>
<td>anginaYes</td>
<td>-0.00789</td>
<td>0.020484</td>
<td>0.8862</td>
</tr>
<tr>
<td>sbp</td>
<td>0.02687</td>
<td>0.257855</td>
<td>0.6116</td>
</tr>
<tr>
<td>age</td>
<td>-0.00109</td>
<td>0.000426</td>
<td>0.9835</td>
</tr>
<tr>
<td>chfUnknown</td>
<td>-0.08444</td>
<td>2.348590</td>
<td>0.1254</td>
</tr>
<tr>
<td>chfYes</td>
<td>-0.03538</td>
<td>0.408394</td>
<td>0.5228</td>
</tr>
<tr>
<td>heartrate</td>
<td>-0.12801</td>
<td>5.377292</td>
<td>0.0204</td>
</tr>
<tr>
<td>GLOBAL</td>
<td>NA</td>
<td>11.649394</td>
<td>0.2338</td>
</tr>
</tbody>
</table>

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.

```r
> cox6
```

<table>
<thead>
<tr>
<th></th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>-0.31331</td>
<td>0.73102</td>
<td>0.11238</td>
<td>-2.79</td>
<td>0.00530</td>
</tr>
</tbody>
</table>

Coefficient for treatment is largely unaffected by baseline model.

3. Suppose that the dataset "data3.csv" represents the results from a randomized trial and contains the following variables

<table>
<thead>
<tr>
<th>t</th>
<th>Follow-up time</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>Event indicator (1=dead, 0=censored)</td>
</tr>
<tr>
<td>z</td>
<td>Treatment group</td>
</tr>
<tr>
<td>w</td>
<td>Binary baseline variable</td>
</tr>
</tbody>
</table>

\( H_0 \) is the null hypothesis that there is no difference in survival by treatment.
(a) Plot cumulative mortality by treatment, test $H_0$ using the log-rank test, and estimate the hazard ratio for treatment, ignoring $w$. What does this indicate regarding the effect of treatment on survival.

```
> plot(survfit(Surv(t,d)~z, data=data3), lty=1:2, fun="event")
> legend("topleft", bty="n", lty=1:2, c("z=0","z=1"))
> survdiff(Surv(t,d)~z, data=data3)
Call:
survdiff(formula = Surv(t, d) ~ z, data = data3)

             N Observed Expected (O-E)^2/E (O-E)^2/V
z=0   826     392     353     4.25     8.01
z=1   841     362     401     3.75     8.01

Chisq= 8 on 1 degrees of freedom, p= 0.00466
```

```
> coxph(Surv(t,d)~z, data=data3)

 coef   exp(coef) se(coef)    z     p
z -0.2060  0.8139  0.0729 -2.82 0.0047

Likelihood ratio test=7.99 on 1 df, p=0.00471
n= 1667, number of events= 754
```

Survival is better in the $z=1$ group relative to the $z=0$ group, with $p < 0.005$. Aside from a possible type I error, the randomization ensures that this association is causal. Estimated hazard ratio is 0.814.

(b) Assess the proportional hazards assumption for the model in part (a).

```
> cox.zph(coxph(Surv(t,d)~z, data=data3))

 rho    chisq     p
z 0.163 19.9 0.00000834
```
cox.zph strongly suggests that proportionality does not hold. The figure suggests that the ratio $\Lambda_1(t)/\Lambda_0(t)$ is decreasing over time.

(c) Repeat parts (a) and (b) separately for subgroups $w=0$ and $w=1$. Note that the subset argument allows you to restrict the analysis to a subset of observations.

```R
# for w==0:
> survdiff(Surv(t,d)~z, data=data3, subset=w==0)
   N  Observed  Expected   (O-E)^2/E (O-E)^2/V
z=0 487  56  55.60    0.00333   0.00645
z=1 501  59  59.40    0.00311   0.00645
Chisq= 0 on 1 degrees of freedom, p= 0.936
> coxph(Surv(t,d)~z, data=data3, subset=w==0)
   coef exp(coef) se(coef)  z  p
z -0.015  0.985   0.187 -0.08 0.94 Likelihood ratio test=0.01 on 1 df, p=0.936
> cox.zph(coxph(Surv(t,d)~z, data=data3, subset=w==0))
   rho  chisq  p
z -0.0284 0.0927 0.761
> plot(survfit(Surv(t,d)~z, data=data3, subset=w==0), lty=1:2,
   + fun="event")
> plot(survfit(Surv(t,d)~z, data=data3, subset=w==0), lty=1:2,
   + fun="cloglog")
```

$w=0$
> survdiff(Surv(t,d)~z, data=data3, subset=w==1)
> N Observed Expected (O-E)^2/E (O-E)^2/V
> z=0 339 336 236 42.7 71.5
> z=1 340 303 403 24.9 71.5
> Chisq= 71.5 on 1 degrees of freedom, p= 0
> coxph(Surv(t,d)~z, data=data3, subset=w==1)
> coef exp(coef) se(coef) z p
> z -0.6837 0.5047 0.0823 -8.31 <2e-16
> cox.zph(coxph(Surv(t,d)~z, data=data3, subset=w==1))
> rho chisq p
> z -0.0351 0.804 0.37
> plot(survfit(Surv(t,d)~z, data=data3, subset=w==1), lty=1:2,
+ fun="event")
> plot(survfit(Surv(t,d)~z, data=data3, subset=w==1), lty=1:2,
+ fun="cloglog")

For the stratum \(w=0\), there is no evidence of a treatment difference at all and proportionality appears to hold, although with a hazard ratio of one. For the stratum \(w=1\), there is strong evidence of a treatment difference, and proportionality appears to hold, with a hazard ratio of about 0.50. We also note that mortality is dramatically higher in the group \(w=1\). E.g., for \(w=0\), cumulative mortality at time 60 is about 12%, whereas for \(w=0\), cumulative mortality at time 60 is over 80%.

(d) Using the log-rank tests from part (c), calculate the log-rank chi-square statistic, stratified by \(w\), and one-step estimate of the hazard ratio for \(z\), adjusted for \(w\). Compare to results using survdiff with a term `strata(w)`. Compare to the result from part (a).

```r
## Extract obs-exp, and var (fisher information) from survdiff
> uv0 <- with(survdiff(Surv(t,d)~z, data=data3, subset=w==0),
+ c(obs[2]-exp[2], var[1]))
> uv1 <- with(survdiff(Surv(t,d)~z, data=data3, subset=w==1),
+ c(obs[2]-exp[2], var[1]))
## stratified chisquare statistic:
> (uv0+uv1)[1]^2/(uv0+uv1)[2]
[1] 59.86843
> survdiff(Surv(t,d)~z + strata(w), data=data3)
```
Call:
survdiff(formula = Surv(t, d) ~ z + strata(w), data = data3)

<table>
<thead>
<tr>
<th>Make</th>
<th>Observed</th>
<th>Expected</th>
<th>(O-E)^2/E</th>
<th>(O-E)^2/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>z=0</td>
<td>826</td>
<td>392</td>
<td>291</td>
<td>34.8</td>
</tr>
<tr>
<td>z=1</td>
<td>841</td>
<td>362</td>
<td>463</td>
<td>21.9</td>
</tr>
</tbody>
</table>

Chisq= 59.9 on 1 degrees of freedom, p= 1.01e-14
## One-step estimator of hazard ratio:
> exp((uv0+uv1)[1]/(uv0+uv1)[2])
[1] 0.5518251
## check against stratified Cox-model:
> coxph(Surv(t,d)~z + strata(w), data=data3)

<table>
<thead>
<tr>
<th>Make</th>
<th>coef</th>
<th>exp(coef)</th>
<th>se(coef)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>-0.5769</td>
<td>0.5616</td>
<td>0.0755</td>
<td>-7.64</td>
<td>2.1e-14</td>
</tr>
</tbody>
</table>

The stratified chi-square statistic from the separate calls to `survdiff` matches the call with stratification. The one-step estimator differs slightly from the MLE from the Cox-model.

(e) The argument `rho` in `survdiff` is the exponent of $S(t)$ in a weighted log-rank test, so that `rho=1` corresponds to the Peto-Peto weighted log-rank test. Compare the result using `rho=1` to the result from part (a). How does the conclusion change? What explains the difference?

> survdiff(Surv(t,d)~z, rho=1, data=data3)

<table>
<thead>
<tr>
<th>Make</th>
<th>Observed</th>
<th>Expected</th>
<th>(O-E)^2/E</th>
<th>(O-E)^2/V</th>
</tr>
</thead>
<tbody>
<tr>
<td>z=0</td>
<td>826</td>
<td>312</td>
<td>274</td>
<td>5.32</td>
</tr>
<tr>
<td>z=1</td>
<td>841</td>
<td>270</td>
<td>308</td>
<td>4.73</td>
</tr>
</tbody>
</table>

Chisq= 12.6 on 1 degrees of freedom, p= 0.000377
The chi-square statistic from part (a) is 8.01, and from the Peto-Peto weighted log-rank is 12.6. Because the hazard ratio from (a) appears to be decreasing, the weighted log-rank downweights the events from later times where the difference is smaller, and upweights earlier events where the difference is larger. Hence, the chi-square statistic for the weighted test is larger than the unweighted.

(f) To illustrate the problem that can arise using increasing weights, in the extreme case events before some fixed time receive weight zero, and events on or after this time receive weight one. We can accomplish this easily by restricting the analysis to subjects still at risk at the time in question. Perform the (unweighted) log-rank test and estimate the hazard ratio for subjects still at risk at time $t = 20$. How does this differ from the analysis in part (a)?
> survdiff(Surv(t,d)~z, data=data3, subset=t>=20)
    N Observed Expected (O-E)^2/E (O-E)^2/V
    z=0 522 88 104 2.43 4.47
    z=1 619 140 124 2.04 4.47
    Chisq= 4.5 on 1 degrees of freedom, p= 0.0345

The chi-square statistic is 4.47, whereas in part (a) it is 8.01. More important, though is that the sign changes. In part (a), in group z=1, the observed minus expected is 362-401=-39, fewer deaths than expected, whereas conditional on survival to time 20, in group z=1, the observed minus expected is 140-124=16, more deaths than expected. The conditional analysis shows that beyond time 20, conditional survival is better in group w=0.

(g) For each treatment group, find the number and percent of subjects with w=1. Repeat for subjects still at risk at time $t = 20$. How does this explain the difference between parts (a) and (f)? Repeat the analyses from part (f) stratified by w and compare to parts (a) and (f). Is w a confounder for the observed association between z and survival?

> tapply(data3$w, data3$z, sum)
     0   1
   339 340
> tapply(data3$w, data3$z, mean)*100
     0   1
61.04116 40.42806
## after time 20
> with(subset(data3, t>=20) , tapply(w, z, sum))
     0   1
    58 140
> with(subset(data3, t>=20) , tapply(w, z, mean))*100
     0   1
11.11111 22.61712

First, the proportion of subjects with w=1 at baseline is about 40% in each treatment group. Conditional on survival to time 20, the proportions change to about 11% and 23% in groups w=0 and w=1 respectively. Because subjects with w=1 are at dramatically higher risk, at time 20 the average risk in group z=0 is higher than in group z=1.