REPEATED EVENTS

Dr. Aric LaBarr

Institute for Advanced Analytics

MODELS FOR REPEATED EVENTS

Independence Model

Independence Model

- Easiest approach is modeling the recurrences as separate, independent events.
- Assumes that all recurrences are identical – the risk of the event is the same regardless of previous events.
- Only care about the overall effect, ignoring the order or type of recurrence.

Independence Model – Risk Set

- Each observation has time intervals of (start, 1st], (1st, 2nd], …, (kth,stop]

ID	start	stop	event	enum
5	0	6	1	1
5	6	10	0	2
13	0	3	1	1
13	3	9	1	2
13	9	21	1	3
13	21	23	0	4
16	0	26	0	1
41	0	35	1	1
41	35	51	0	2

Accounting for Dependence

- Easiest approach is modeling the recurrences as separate, **independent events**.
- But they aren't! Right?

- 2 Approaches:
  1. Time-Dependent Variables
  2. Robust Standard Errors

Independence Model – R

```
bladder.td <- coxph(Surv(start, stop, event == 1) ~ rx + number +
                    size + enum, data = bladder)

summary(bladder.td)
```

Independence Model – R

```
## Call:
## coxph(formula = Surv(start, stop, event == 1) ~ rx + number +
##     size + enum, data = bladder)
##
##   n= 178, number of events= 112
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## rx      -0.30125   0.73989  0.20440 -1.474  0.14052
## number   0.14193   1.15249  0.04949  2.868  0.00414 **
## size    -0.01586   0.98427  0.06926 -0.229  0.81892
## enum     0.53604   1.70922  0.10192  5.260 1.44e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Independence Model – R

```
##          exp(coef) exp(-coef) lower .95 upper .95
## rx          0.7399     1.3516    0.4957     1.104
## number      1.1525     0.8677    1.0459     1.270
## size        0.9843     1.0160    0.8593     1.127
## enum        1.7092     0.5851    1.3997     2.087
##
## Concordance= 0.673  (se = 0.03 )
## Likelihood ratio test= 43.89  on 4 df,   p=7e-09
## Wald test            = 45.65  on 4 df,   p=3e-09
## Score (logrank) test = 50.69  on 4 df,   p=3e-10
```

Accounting for Dependence

- Easiest approach is modeling the recurrences as separate, **independent events**.
- But they aren't! Right?

- 2 Approaches:
  1. Time-Dependent Variables
  2. Robust Standard Errors → Still possible correlation between observations that can not be explained away with time-dependent variables.

Independence Model – R

```
bladder.rse <- coxph(Surv(start, stop, event == 1) ~ rx + number +
                     size + enum + cluster(id), data = bladder)

summary(bladder.rse)
```

Independence Model – R

```
## Call:
## coxph(formula = Surv(start, stop, event == 1) ~ rx + number +
##     size + enum + cluster(id), data = bladder)
##
##   n= 178, number of events= 112
##
##            coef exp(coef) se(coef) robust se      z Pr(>|z|)
## rx     -0.30125   0.73989  0.20440   0.21277 -1.416  0.15682
## number  0.14193   1.15249  0.04949   0.05321  2.667  0.00764 **
## size   -0.01586   0.98427  0.06926   0.06175 -0.257  0.79734
## enum    0.53604   1.70922  0.10192   0.10516  5.097 3.45e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Independence Model – R

```
##        exp(coef) exp(-coef) lower .95 upper .95
## rx        0.7399     1.3516    0.4876     1.123
## number    1.1525     0.8677    1.0384     1.279
## size      0.9843     1.0160    0.8721     1.111
## enum      1.7092     0.5851    1.3909     2.100
##
## Concordance= 0.673  (se = 0.031 )
## Likelihood ratio test= 43.89  on 4 df,   p=7e-09
## Wald test            = 41.41  on 4 df,   p=2e-08
## Score (logrank) test = 50.69  on 4 df,   p=3e-10,   Robust = 21.88  p=2e-04
##
##   (Note: the likelihood ratio and score tests assume independence of
##      observations within a cluster, the Wald and robust score tests do not).
```

MODELS FOR REPEATED EVENTS

Conditional Model

Stratified Models

- Unlike the independence model, we can preserve the ordering of events if it's important.
- In the **conditional model**, we stratify on the number of events, so only those who have had a previous event are in the risk set for the next one.
  - Example: Not in the risk set for the 3rd event until you have had the 2nd event.
- Each recurrence is a separate stratum (imagine own model) with its **own baseline hazard** – no estimates/inferences on the number of recurrences.

Conditional Model – Risk Set

- Risk set for 1st event:

ID	start	stop	event	enum
5	0	6	1	1
13	0	3	1	1
16	0	26	0	1
41	0	35	1	1

- Risk set for 2nd event:

ID	start	stop	event	enum
5	6	10	0	2
13	3	9	1	2
41	35	51	0	2

Conditional Model – R

```
bladder.con <- coxph(Surv(start, stop, event == 1) ~ rx + number +
                       size + strata(enum), data = bladder)

summary(bladder.con)
```

Conditional Model – R

```
## coxph(formula = Surv(start, stop, event == 1) ~ rx + number +
##     size + strata(enum), data = bladder)
##
##   n= 178, number of events= 112
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## rx     -0.333489  0.716420  0.216168 -1.543   0.1229
## number  0.119617  1.127065  0.053338  2.243   0.0249 *
## size   -0.008495  0.991541  0.072762 -0.117   0.9071
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## rx        0.7164     1.3958    0.4690     1.094
## number    1.1271     0.8873    1.0152     1.251
## size      0.9915     1.0085    0.8598     1.144
##
## Concordance= 0.616  (se = 0.038 )
## Likelihood ratio test= 6.51  on 3 df,   p=0.09
## Wald test            = 6.85  on 3 df,   p=0.08
## Score (logrank) test = 6.91  on 3 df,   p=0.07
```

Conditional Model – R

```
bladder.con2 <- coxph(Surv(start, stop, event == 1) ~
                      rx:strata(enum) + number:strata(enum) +
                      size:strata(enum), data = bladder)

summary(bladder.con2)
```

Conditional Model – R

```
## coxph(formula = Surv(start, stop, event == 1) ~ rx:strata(enum) +
##     number:strata(enum) + size:strata(enum), data = bladder)
##
##   n= 178, number of events= 112
##
##                                coef exp(coef)  se(coef)      z Pr(>|z|)
## rx:strata(enum)enum=1     -0.525984  0.590973  0.315826 -1.665   0.0958 .
## rx:strata(enum)enum=2     -0.503837  0.604208  0.406167 -1.240   0.2148
## rx:strata(enum)enum=3      0.140657  1.151029  0.673063  0.209   0.8345
## rx:strata(enum)enum=4      0.050331  1.051619  0.791710  0.064   0.9493
## strata(enum)enum=1:number  0.238180  1.268937  0.075885  3.139   0.0017 **
## strata(enum)enum=2:number -0.024641  0.975660  0.089873 -0.274   0.7840
## strata(enum)enum=3:number  0.049661  1.050915  0.185323  0.268   0.7887
## strata(enum)enum=4:number  0.204277  1.226637  0.242040  0.844   0.3987
## strata(enum)enum=1:size    0.069613  1.072094  0.101559  0.685   0.4931
## strata(enum)enum=2:size   -0.160716  0.851534  0.122467 -1.312   0.1894
## strata(enum)enum=3:size    0.168099  1.183053  0.269040  0.625   0.5321
## strata(enum)enum=4:size    0.009095  1.009137  0.338928  0.027   0.9786
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gap Time

- Notice that in the conditional model, each event's **start** time is determined by the previous event's **stop** time!
- An alternative time scale is the **gap time**, where we instead choose to model the time *since last event*.
- In gap-time models, time is reset to 0 after each event, so the time until the prior event has no bearing on the current event's risk set.

Gap Time – Risk Set

- Risk set for 1st event:

ID	start	stop	event	enum
5	0	6	1	1
13	0	3	1	1
16	0	26	0	1
41	0	35	1	1

- Risk set for 2nd event:

ID	start	stop	event	enum
5	0	4	0	2
13	0	6	1	2
41	0	16	0	2

Gap Time – R

```
bladder.gap <- coxph(Surv(time = (stop - start), event == 1) ~ rx +
                       number + size + strata(enum), data = bladder)

summary(bladder.gap)
```

Gap Time – R

```
## coxph(formula = Surv(time = (stop - start), event == 1) ~ rx +
##     number + size + strata(enum), data = bladder)
##
##   n= 178, number of events= 112
##
##             coef exp(coef)  se(coef)      z Pr(>|z|)
## rx     -0.279005  0.756536  0.207348 -1.346  0.17844
## number  0.158046  1.171220  0.051942  3.043  0.00234 **
## size    0.007415  1.007443  0.070023  0.106  0.91567
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## rx        0.7565     1.3218    0.5039     1.136
## number    1.1712     0.8538    1.0579     1.297
## size      1.0074     0.9926    0.8782     1.156
##
## Concordance= 0.596  (se = 0.035 )
## Likelihood ratio test= 9.33  on 3 df,   p=0.03
## Wald test            = 10.11  on 3 df,   p=0.02
## Score (logrank) test = 10.27  on 3 df,   p=0.02
```