Cox PH Models

Landi Luo

12/7/2018

Load packages:

```
if (!require("pacman"))
  install.packages("pacman", repos = "http://cran.us.r-project.org/")
p_load("tidyverse", "survival", "kableExtra", "survivalAnalysis")
```

Import data:

```
breast <- readRDS(file = "breast_final.rds")

# delete all survival time = 0
breast <- subset(breast, SRV_TIME_MON != 0)

# convert SEX to dummy: 0=male, 1=female
breast$SEX <- ifelse(breast$SEX == 1, 0, 1)
```

Cox Model: All Covariates

Using the Breslow method of handling ties, we fit a Cox proportional hazards model to the data including all 13 covariates: race, sex, stage, breast subtype, age dx, age, marital status, benign tumor count, malignant tumor count, primary site, pr status, er status, insurance status.

```
fit <- coxph(Surv(SRV_TIME_MON, delta) ~ factor(SEX) + factor(stage) + factor(RAC_RECY) +
             factor(BRST_SUB) + AGE_DX + Age + factor(MAR_STAT) + MALIGCOUNT +
             BENBORDCOUNT + factor(PRIMSITE) + factor(ERSTATUS) + factor(PRSTATUS) +
             factor(INSREC_PUB), data = breast, ties = "breslow")
summary(fit)
```

```
## Call:
## coxph(formula = Surv(SRV_TIME_MON, delta) ~ factor(SEX) + factor(stage) +
##     factor(RAC_RECY) + factor(BRST_SUB) + AGE_DX + Age + factor(MAR_STAT) +
##     MALIGCOUNT + BENBORDCOUNT + factor(PRIMSITE) + factor(ERSTATUS) +
##     factor(PRSTATUS) + factor(INSREC_PUB), data = breast, ties = "breslow")
##
##   n= 55333, number of events= 3033
##
##                           coef  exp(coef)   se(coef)      z Pr(>|z|)
## factor(SEX)1         -0.234365   0.791073   0.213390 -1.098 0.272077
## factor(stage)1        1.828641   6.225421   0.709709  2.577 0.009977 **
## factor(stage)2        3.106446  22.341492   0.708153  4.387 1.15e-05 ***
## factor(stage)3        4.240722  69.457958   0.708109  5.989 2.11e-09 ***
## factor(stage)4        5.860086 350.754397   0.707926  8.278  < 2e-16 ***
## factor(RAC_RECY)2     0.232315   1.261517   0.067114  3.462 0.000537 ***
## factor(RAC_RECY)3     0.081242   1.084634   0.197643  0.411 0.681031
## factor(RAC_RECY)4    -0.258741   0.772023   0.077398 -3.343 0.000829 ***
## factor(BRST_SUB)2    -0.969818   0.379152   0.152118 -6.375 1.82e-10 ***
## factor(BRST_SUB)3     0.174221   1.190319   0.064187  2.714 0.006642 **
## factor(BRST_SUB)4    -0.150397   0.860366   0.142171 -1.058 0.290118
```

```
## AGE_DX                 0.033083   1.033637   0.014080  2.350 0.018794 *
## Age                   -0.008105   0.991928   0.014056 -0.577 0.564208
## factor(MAR_STAT)2     -0.373226   0.688509   0.049662 -7.515 5.67e-14 ***
## factor(MAR_STAT)3     -0.140786   0.868675   0.146900 -0.958 0.337870
## factor(MAR_STAT)4     -0.154337   0.856983   0.064872 -2.379 0.017354 *
## factor(MAR_STAT)5      0.075775   1.078719   0.063998  1.184 0.236410
## factor(MAR_STAT)6     -0.121050   0.885989   0.336358 -0.360 0.718933
## MALIGCOUNT             0.253995   1.289166   0.055057  4.613 3.96e-06 ***
## BENBORDCOUNT           0.083414   1.086991   0.232919  0.358 0.720252
## factor(PRIMSITE)1     -0.244291   0.783260   0.255573 -0.956 0.339144
## factor(PRIMSITE)2     -0.164098   0.848658   0.254097 -0.646 0.518402
## factor(PRIMSITE)3     -0.043522   0.957412   0.259180 -0.168 0.866645
## factor(PRIMSITE)4     -0.276886   0.758141   0.247999 -1.116 0.264215
## factor(PRIMSITE)5     -0.306483   0.736031   0.256539 -1.195 0.232210
## factor(PRIMSITE)6     -0.218727   0.803541   0.336821 -0.649 0.516089
## factor(PRIMSITE)7     -0.095879   0.908574   0.248255 -0.386 0.699339
## factor(PRIMSITE)8      0.027267   1.027642   0.248121  0.110 0.912494
## factor(ERSTATUS)1      1.080860   2.947214   0.129059  8.375  < 2e-16 ***
## factor(PRSTATUS)1      0.653819   1.922871   0.053097 12.314  < 2e-16 ***
## factor(INSREC_PUB)1  -0.144237   0.865682   0.128615 -1.121 0.262091
## factor(INSREC_PUB)2  -0.513753   0.598246   0.126632 -4.057 4.97e-05 ***
## factor(INSREC_PUB)3  -0.385176   0.680331   0.132869 -2.899 0.003745 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                      exp(coef) exp(-coef) lower .95 upper .95
## factor(SEX)1            0.7911   1.264106    0.5207    1.2019
## factor(stage)1          6.2254   0.160632    1.5490   25.0192
## factor(stage)2         22.3415   0.044760    5.5761   89.5144
## factor(stage)3         69.4580   0.014397   17.3372  278.2695
## factor(stage)4        350.7544   0.002851   87.5821 1404.7232
## factor(RAC_RECY)2       1.2615   0.792697    1.1060    1.4389
## factor(RAC_RECY)3       1.0846   0.921970    0.7363    1.5978
## factor(RAC_RECY)4       0.7720   1.295298    0.6634    0.8985
## factor(BRST_SUB)2       0.3792   2.637465    0.2814    0.5109
## factor(BRST_SUB)3       1.1903   0.840111    1.0496    1.3499
## factor(BRST_SUB)4       0.8604   1.162296    0.6511    1.1368
## AGE_DX                  1.0336   0.967458    1.0055    1.0626
## Age                     0.9919   1.008138    0.9650    1.0196
## factor(MAR_STAT)2       0.6885   1.452413    0.6247    0.7589
## factor(MAR_STAT)3       0.8687   1.151178    0.6514    1.1585
## factor(MAR_STAT)4       0.8570   1.166884    0.7547    0.9732
## factor(MAR_STAT)5       1.0787   0.927025    0.9516    1.2229
## factor(MAR_STAT)6       0.8860   1.128682    0.4583    1.7129
## MALIGCOUNT              1.2892   0.775695    1.1573    1.4361
## BENBORDCOUNT            1.0870   0.919971    0.6886    1.7159
## factor(PRIMSITE)1       0.7833   1.276716    0.4746    1.2926
## factor(PRIMSITE)2       0.8487   1.178330    0.5158    1.3964
## factor(PRIMSITE)3       0.9574   1.044483    0.5761    1.5912
## factor(PRIMSITE)4       0.7581   1.319017    0.4663    1.2327
## factor(PRIMSITE)5       0.7360   1.358638    0.4452    1.2169
## factor(PRIMSITE)6       0.8035   1.244492    0.4152    1.5549
## factor(PRIMSITE)7       0.9086   1.100626    0.5585    1.4780
## factor(PRIMSITE)8       1.0276   0.973101    0.6319    1.6713
```

```
## factor(ERSTATUS)1      2.9472   0.339304    2.2885    3.7955
## factor(PRSTATUS)1      1.9229   0.520056    1.7328    2.1338
## factor(INSREC_PUB)1    0.8657   1.155158    0.6728    1.1139
## factor(INSREC_PUB)2    0.5982   1.671553    0.4668    0.7668
## factor(INSREC_PUB)3    0.6803   1.469873    0.5244    0.8827
##
## Concordance= 0.889  (se = 0.006 )
## Rsquare= 0.125   (max possible= 0.679 )
## Likelihood ratio test= 7415  on 33 df,   p=<2e-16
## Wald test            = 7475  on 33 df,   p=<2e-16
## Score (logrank) test = 15283  on 33 df,   p=<2e-16
```

ANOVA Table: All Covariates

We constructed an Analysis of Variance table to summarize estimates of the risk coefficients and the results of the one degree of freedom tests for each covariate in the model:

```
anova_table <- data.frame(summary(fit)$coefficients)
kable(anova_table, "latex", booktabs = TRUE,
      col.names = c("Coefficient", "Exp. Coeff.", "Std. Error", "Z-Score", "P-Value")) %>%
  kable_styling("striped")
```

	Coefficient	Exp. Coeff.	Std. Error	Z-Score	P-Value
factor(SEX)1	-0.2343650	0.7910730	0.2133903	-1.0982930	0.2720765
factor(stage)1	1.8286411	6.2254212	0.7097085	2.5766087	0.0099775
factor(stage)2	3.1064456	22.3414916	0.7081526	4.3866896	0.0000115
factor(stage)3	4.2407216	69.4579580	0.7081091	5.9887970	0.0000000
factor(stage)4	5.8600863	350.7543972	0.7079259	8.2778240	0.0000000
factor(RAC_RECY)2	0.2323147	1.2615166	0.0671137	3.4615099	0.0005372
factor(RAC_RECY)3	0.0812425	1.0846338	0.1976429	0.4110568	0.6810309
factor(RAC_RECY)4	-0.2587411	0.7720229	0.0773981	-3.3429909	0.0008288
factor(BRST_SUB)2	-0.9698181	0.3791520	0.1521181	-6.3754270	0.0000000
factor(BRST_SUB)3	0.1742212	1.1903188	0.0641868	2.7142821	0.0066420
factor(BRST_SUB)4	-0.1503972	0.8603662	0.1421707	-1.0578634	0.2901177
AGE_DX	0.0330834	1.0336367	0.0140805	2.3495887	0.0187942
Age	-0.0081050	0.9919278	0.0140565	-0.5766028	0.5642078
factor(MAR_STAT)2	-0.3732264	0.6885093	0.0496616	-7.5153933	0.0000000
factor(MAR_STAT)3	-0.1407861	0.8686751	0.1468997	-0.9583822	0.3378701
factor(MAR_STAT)4	-0.1543374	0.8569829	0.0648718	-2.3791128	0.0173544
factor(MAR_STAT)5	0.0757746	1.0787194	0.0639984	1.1840084	0.2364097
factor(MAR_STAT)6	-0.1210502	0.8859895	0.3363584	-0.3598847	0.7189334
MALIGCOUNT	0.2539954	1.2891659	0.0550571	4.6133054	0.0000040
BENBORDCOUNT	0.0834136	1.0869913	0.2329193	0.3581223	0.7202518
factor(PRIMSITE)1	-0.2442908	0.7832598	0.2555725	-0.9558572	0.3391444
factor(PRIMSITE)2	-0.1640985	0.8486585	0.2540967	-0.6458110	0.5184017
factor(PRIMSITE)3	-0.0435219	0.9574116	0.2591802	-0.1679214	0.8666451
factor(PRIMSITE)4	-0.2768864	0.7581406	0.2479985	-1.1164840	0.2642150
factor(PRIMSITE)5	-0.3064831	0.7360310	0.2565386	-1.1946861	0.2322097
factor(PRIMSITE)6	-0.2187273	0.8035408	0.3368214	-0.6493866	0.5160885
factor(PRIMSITE)7	-0.0958792	0.9085738	0.2482549	-0.3862127	0.6993392
factor(PRIMSITE)8	0.0272669	1.0276420	0.2481207	0.1098937	0.9124937
factor(ERSTATUS)1	1.0808603	2.9472139	0.1290589	8.3749408	0.0000000
factor(PRSTATUS)1	0.6538192	1.9228707	0.0530969	12.3137000	0.0000000
factor(INSREC_PUB)1	-0.1442372	0.8656823	0.1286153	-1.1214622	0.2620912
factor(INSREC_PUB)2	-0.5137528	0.5982462	0.1266319	-4.0570570	0.0000497
factor(INSREC_PUB)3	-0.3851760	0.6803309	0.1328691	-2.8989135	0.0037446

Cox Model: Top 9 Significant Variables

Using variable selection methods (LASSO, SCAD, MCP), we decided the top 9 significant variables were:

- Stage
- ERSTATUS
- PRSTATUS
- MALIGCOUNT
- RAC_RECY
- PRIMSITE
- BRST_SUB
- MAR_STAT
- INSREC_PUB

We fit a Cox model with these covariates plus sex:

```r
fit2 <- coxph(Surv(SRV_TIME_MON, delta) ~ factor(SEX) + factor(stage) + factor(RAC_RECY) +
            factor(BRST_SUB) + factor(MAR_STAT) + MALIGCOUNT + factor(PRIMSITE) +
            factor(ERSTATUS) + factor(PRSTATUS) + factor(INSREC_PUB) ,
            data = breast, ties = "breslow" )
summary(fit2)
```

```
## Call:
## coxph(formula = Surv(SRV_TIME_MON, delta) ~ factor(SEX) + factor(stage) +
##     factor(RAC_RECY) + factor(BRST_SUB) + factor(MAR_STAT) +
##     MALIGCOUNT + factor(PRIMSITE) + factor(ERSTATUS) + factor(PRSTATUS) +
##     factor(INSREC_PUB), data = breast, ties = "breslow")
##
##   n= 55333, number of events= 3033
##
##                          coef  exp(coef)   se(coef)      z Pr(>|z|)
## factor(SEX)1        -0.349810   0.704822   0.213794 -1.636  0.10180
## factor(stage)1       1.856788   6.403136   0.709702  2.616  0.00889 **
## factor(stage)2       3.076275  21.677495   0.708156  4.344 1.40e-05 ***
## factor(stage)3       4.184014  65.628743   0.708103  5.909 3.45e-09 ***
## factor(stage)4       5.831913 341.010252   0.707915  8.238  < 2e-16 ***
## factor(RAC_RECY)2    0.191037   1.210504   0.066904  2.855  0.00430 **
## factor(RAC_RECY)3    0.067350   1.069670   0.197577  0.341  0.73319
## factor(RAC_RECY)4   -0.312395   0.731692   0.077236 -4.045 5.24e-05 ***
## factor(BRST_SUB)2   -0.965328   0.380858   0.152375 -6.335 2.37e-10 ***
## factor(BRST_SUB)3    0.251263   1.285648   0.063903  3.932 8.43e-05 ***
## factor(BRST_SUB)4   -0.103694   0.901502   0.142382 -0.728  0.46644
## factor(MAR_STAT)2   -0.315727   0.729258   0.049500 -6.378 1.79e-10 ***
## factor(MAR_STAT)3   -0.099548   0.905246   0.146908 -0.678  0.49801
## factor(MAR_STAT)4   -0.008995   0.991045   0.064307 -0.140  0.88876
## factor(MAR_STAT)5    0.515853   1.675066   0.057982  8.897  < 2e-16 ***
## factor(MAR_STAT)6   -0.126694   0.881003   0.336125 -0.377  0.70623
## MALIGCOUNT           0.288007   1.333767   0.054876  5.248 1.54e-07 ***
## factor(PRIMSITE)1   -0.245851   0.782039   0.255827 -0.961  0.33655
## factor(PRIMSITE)2   -0.244650   0.782979   0.254533 -0.961  0.33647
## factor(PRIMSITE)3   -0.120621   0.886370   0.259590 -0.465  0.64217
## factor(PRIMSITE)4   -0.363928   0.694941   0.248432 -1.465  0.14295
## factor(PRIMSITE)5   -0.370951   0.690078   0.256980 -1.443  0.14888
## factor(PRIMSITE)6   -0.175356   0.839158   0.336763 -0.521  0.60257
## factor(PRIMSITE)7   -0.185623   0.830586   0.248680 -0.746  0.45541
## factor(PRIMSITE)8   -0.044644   0.956338   0.248489 -0.180  0.85742
## factor(ERSTATUS)1    1.033302   2.810329   0.128667  8.031 9.68e-16 ***
## factor(PRSTATUS)1    0.678659   1.971233   0.053083 12.785  < 2e-16 ***
## factor(INSREC_PUB)1 -0.107959   0.897665   0.128455 -0.840  0.40066
## factor(INSREC_PUB)2 -0.357259   0.699591   0.126129 -2.832  0.00462 **
## factor(INSREC_PUB)3 -0.148713   0.861816   0.131929 -1.127  0.25965
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## factor(SEX)1           0.7048   1.418799    0.4636    1.0717
## factor(stage)1         6.4031   0.156173    1.5933   25.7331
## factor(stage)2        21.6775   0.046131    5.4104   86.8545
## factor(stage)3        65.6287   0.015237   16.3816  262.9256
## factor(stage)4       341.0103   0.002932   85.1508 1365.6711
```

```
## factor(RAC_RECY)2      1.2105   0.826102    1.0617    1.3801
## factor(RAC_RECY)3      1.0697   0.934868    0.7262    1.5755
## factor(RAC_RECY)4      0.7317   1.366695    0.6289    0.8513
## factor(BRST_SUB)2      0.3809   2.625650    0.2825    0.5134
## factor(BRST_SUB)3      1.2856   0.777818    1.1343    1.4572
## factor(BRST_SUB)4      0.9015   1.109260    0.6820    1.1917
## factor(MAR_STAT)2      0.7293   1.371256    0.6618    0.8036
## factor(MAR_STAT)3      0.9052   1.104672    0.6788    1.2073
## factor(MAR_STAT)4      0.9910   1.009035    0.8737    1.1242
## factor(MAR_STAT)5      1.6751   0.596991    1.4951    1.8767
## factor(MAR_STAT)6      0.8810   1.135070    0.4559    1.7025
## MALIGCOUNT             1.3338   0.749756    1.1978    1.4852
## factor(PRIMSITE)1      0.7820   1.278709    0.4737    1.2912
## factor(PRIMSITE)2      0.7830   1.277174    0.4754    1.2895
## factor(PRIMSITE)3      0.8864   1.128197    0.5329    1.4743
## factor(PRIMSITE)4      0.6949   1.438971    0.4271    1.1309
## factor(PRIMSITE)5      0.6901   1.449112    0.4170    1.1419
## factor(PRIMSITE)6      0.8392   1.191671    0.4337    1.6237
## factor(PRIMSITE)7      0.8306   1.203969    0.5102    1.3523
## factor(PRIMSITE)8      0.9563   1.045655    0.5876    1.5564
## factor(ERSTATUS)1      2.8103   0.355830    2.1839    3.6164
## factor(PRSTATUS)1      1.9712   0.507297    1.7765    2.1874
## factor(INSREC_PUB)1    0.8977   1.114002    0.6979    1.1547
## factor(INSREC_PUB)2    0.6996   1.429407    0.5464    0.8958
## factor(INSREC_PUB)3    0.8618   1.160340    0.6655    1.1161
##
## Concordance= 0.882  (se = 0.006 )
## Rsquare= 0.121   (max possible= 0.679 )
## Likelihood ratio test= 7165  on 30 df,   p=<2e-16
## Wald test            = 7211  on 30 df,   p=<2e-16
## Score (logrank) test = 15003  on 30 df,   p=<2e-16
```

```r
anova_table2 <- data.frame(summary(fit2)$coefficients)
kable(anova_table2, "latex", booktabs = TRUE,
      col.names = c("Coefficient", "Relative Risk", "Std. Error", "Z-Score", "P-Value")) %>%
  kable_styling(latex_options = "striped")
```

	Coefficient	Relative Risk	Std. Error	Z-Score	P-Value
factor(SEX)1	-0.3498104	0.7048217	0.2137943	-1.6362007	0.1017976
factor(stage)1	1.8567879	6.4031362	0.7097021	2.6162919	0.0088891
factor(stage)2	3.0762746	21.6774947	0.7081560	4.3440637	0.0000140
factor(stage)3	4.1840138	65.6287428	0.7081035	5.9087604	0.0000000
factor(stage)4	5.8319125	341.0102520	0.7079154	8.2381491	0.0000000
factor(RAC_RECY)2	0.1910369	1.2105041	0.0669042	2.8553800	0.0042985
factor(RAC_RECY)3	0.0673499	1.0696697	0.1975770	0.3408793	0.7331945
factor(RAC_RECY)4	-0.3123954	0.7316922	0.0772355	-4.0447106	0.0000524
factor(BRST_SUB)2	-0.9653283	0.3808581	0.1523746	-6.3352323	0.0000000
factor(BRST_SUB)3	0.2512626	1.2856476	0.0639033	3.9319198	0.0000843
factor(BRST_SUB)4	-0.1036935	0.9015015	0.1423817	-0.7282787	0.4664430
factor(MAR_STAT)2	-0.3157273	0.7292583	0.0494999	-6.3783367	0.0000000
factor(MAR_STAT)3	-0.0995482	0.9052464	0.1469079	-0.6776229	0.4980108
factor(MAR_STAT)4	-0.0089949	0.9910455	0.0643074	-0.1398729	0.8887604
factor(MAR_STAT)5	0.5158527	1.6750663	0.0579821	8.8967507	0.0000000
factor(MAR_STAT)6	-0.1266944	0.8810029	0.3361255	-0.3769258	0.7062288
MALIGCOUNT	0.2880071	1.3337668	0.0548764	5.2482854	0.0000002
factor(PRIMSITE)1	-0.2458512	0.7820386	0.2558266	-0.9610074	0.3365485
factor(PRIMSITE)2	-0.2446495	0.7829789	0.2545332	-0.9611693	0.3364671
factor(PRIMSITE)3	-0.1206211	0.8863697	0.2595897	-0.4646607	0.6421744
factor(PRIMSITE)4	-0.3639285	0.6949409	0.2484321	-1.4649012	0.1429479
factor(PRIMSITE)5	-0.3709508	0.6900779	0.2569803	-1.4434992	0.1488799
factor(PRIMSITE)6	-0.1753562	0.8391580	0.3367633	-0.5207106	0.6025684
factor(PRIMSITE)7	-0.1856234	0.8305863	0.2486804	-0.7464339	0.4554054
factor(PRIMSITE)8	-0.0446436	0.9563382	0.2484893	-0.1796602	0.8574194
factor(ERSTATUS)1	1.0333016	2.8103290	0.1286672	8.0308106	0.0000000
factor(PRSTATUS)1	0.6786592	1.9712329	0.0530830	12.7848714	0.0000000
factor(INSREC_PUB)1	-0.1079585	0.8976648	0.1284547	-0.8404406	0.4006614
factor(INSREC_PUB)2	-0.3572593	0.6995910	0.1261293	-2.8324858	0.0046188
factor(INSREC_PUB)3	-0.1487134	0.8618161	0.1319291	-1.1272218	0.2596487

Extract HR and 95% CI's

```
fit2summary <- cox_as_data_frame(summary(fit2))

relrisk <- fit2summary[, c(1, 4:6, 10)]

kable(relrisk, "latex", booktabs = TRUE,
      col.names = c("Covariate", "Relative Risk", "Lower 95% CI", "Upper 95% CI", "P-Value")) %>%
  kable_styling(latex_options = "striped")
```

Covariate	Relative Risk	Lower 95% CI	Upper 95% CI	P-Value
factor:(SEX)1	0.7048217	0.4635506	1.0716706	0.1017976
factor:(stage)1	6.4031362	1.5932840	25.7331098	0.0088891
factor:(stage)2	21.6774947	5.4103535	86.8545430	0.0000140
factor:(stage)3	65.6287428	16.3815621	262.9255904	0.0000000
factor:(stage)4	341.0102520	85.1508065	1365.6710583	0.0000000
factor:(RAC_RECY)2	1.2105041	1.0617379	1.3801148	0.0042985
factor:(RAC_RECY)3	1.0696697	0.7262261	1.5755330	0.7331945
factor:(RAC_RECY)4	0.7316922	0.6289055	0.8512780	0.0000524
factor:(BRST_SUB)2	0.3808581	0.2825282	0.5134105	0.0000000
factor:(BRST_SUB)3	1.2856476	1.1342986	1.4571912	0.0000843
factor:(BRST_SUB)4	0.9015015	0.6819789	1.1916864	0.4664430
factor:(MAR_STAT)2	0.7292583	0.6618307	0.8035553	0.0000000
factor:(MAR_STAT)3	0.9052464	0.6787636	1.2072995	0.4980108
factor:(MAR_STAT)4	0.9910455	0.8736853	1.1241704	0.8887604
factor:(MAR_STAT)5	1.6750663	1.4951250	1.8766639	0.0000000
factor:(MAR_STAT)6	0.8810029	0.4558971	1.7025030	0.7062288
MALIGCOUNT	1.3337668	1.1977578	1.4852201	0.0000002
factor:(PRIMSITE)1	0.7820386	0.4736617	1.2911840	0.3365485
factor:(PRIMSITE)2	0.7829789	0.4754348	1.2894638	0.3364671
factor:(PRIMSITE)3	0.8863697	0.5329074	1.4742734	0.6421744
factor:(PRIMSITE)4	0.6949409	0.4270533	1.1308725	0.1429479
factor:(PRIMSITE)5	0.6900779	0.4170192	1.1419317	0.1488799
factor:(PRIMSITE)6	0.8391580	0.4337009	1.6236679	0.6025684
factor:(PRIMSITE)7	0.8305863	0.5101615	1.3522652	0.4554054
factor:(PRIMSITE)8	0.9563382	0.5876206	1.5564172	0.8574194
factor:(ERSTATUS)1	2.8103290	2.1839137	3.6164198	0.0000000
factor:(PRSTATUS)1	1.9712329	1.7764526	2.1873701	0.0000000
factor:(INSREC_PUB)1	0.8976648	0.6978682	1.1546624	0.4006614
factor:(INSREC_PUB)2	0.6995910	0.5463649	0.8957889	0.0046188
factor:(INSREC_PUB)3	0.8618161	0.6654514	1.1161251	0.2596487

Table with ANOVA estimates, HR and 95% CIs

```
table <- cbind(relrisk[, 1], anova_table2[, c(1,3)], relrisk[, -1], row.names = NULL)
kable(table, "latex", booktabs = TRUE, col.names = c("Covariate", "Coefficient", "Std. Error",
      "Relative Risk", "Lower 95% CI", "Upper 95% CI", "P-Value")) %>%
  kable_styling(latex_options = "striped")
```

Covariate	Coefficient	Std. Error	Relative Risk	Lower 95% CI	Upper 95% CI	P-Value
factor:(SEX)1	-0.3498104	0.2137943	0.7048217	0.4635506	1.0716706	0.1017976
factor:(stage)1	1.8567879	0.7097021	6.4031362	1.5932840	25.7331098	0.0088891
factor:(stage)2	3.0762746	0.7081560	21.6774947	5.4103535	86.8545430	0.0000140
factor:(stage)3	4.1840138	0.7081035	65.6287428	16.3815621	262.9255904	0.0000000
factor:(stage)4	5.8319125	0.7079154	341.0102520	85.1508065	1365.6710583	0.0000000
factor:(RAC_RECY)2	0.1910369	0.0669042	1.2105041	1.0617379	1.3801148	0.0042985
factor:(RAC_RECY)3	0.0673499	0.1975770	1.0696697	0.7262261	1.5755330	0.7331945
factor:(RAC_RECY)4	-0.3123954	0.0772355	0.7316922	0.6289055	0.8512780	0.0000524
factor:(BRST_SUB)2	-0.9653283	0.1523746	0.3808581	0.2825282	0.5134105	0.0000000
factor:(BRST_SUB)3	0.2512626	0.0639033	1.2856476	1.1342986	1.4571912	0.0000843
factor:(BRST_SUB)4	-0.1036935	0.1423817	0.9015015	0.6819789	1.1916864	0.4664430
factor:(MAR_STAT)2	-0.3157273	0.0494999	0.7292583	0.6618307	0.8035553	0.0000000
factor:(MAR_STAT)3	-0.0995482	0.1469079	0.9052464	0.6787636	1.2072995	0.4980108
factor:(MAR_STAT)4	-0.0089949	0.0643074	0.9910455	0.8736853	1.1241704	0.8887604
factor:(MAR_STAT)5	0.5158527	0.0579821	1.6750663	1.4951250	1.8766639	0.0000000
factor:(MAR_STAT)6	-0.1266944	0.3361255	0.8810029	0.4558971	1.7025030	0.7062288
MALIGCOUNT	0.2880071	0.0548764	1.3337668	1.1977578	1.4852201	0.0000002
factor:(PRIMSITE)1	-0.2458512	0.2558266	0.7820386	0.4736617	1.2911840	0.3365485
factor:(PRIMSITE)2	-0.2446495	0.2545332	0.7829789	0.4754348	1.2894638	0.3364671
factor:(PRIMSITE)3	-0.1206211	0.2595897	0.8863697	0.5329074	1.4742734	0.6421744
factor:(PRIMSITE)4	-0.3639285	0.2484321	0.6949409	0.4270533	1.1308725	0.1429479
factor:(PRIMSITE)5	-0.3709508	0.2569803	0.6900779	0.4170192	1.1419317	0.1488799
factor:(PRIMSITE)6	-0.1753562	0.3367633	0.8391580	0.4337009	1.6236679	0.6025684
factor:(PRIMSITE)7	-0.1856234	0.2486804	0.8305863	0.5101615	1.3522652	0.4554054
factor:(PRIMSITE)8	-0.0446436	0.2484893	0.9563382	0.5876206	1.5564172	0.8574194
factor:(ERSTATUS)1	1.0333016	0.1286672	2.8103290	2.1839137	3.6164198	0.0000000
factor:(PRSTATUS)1	0.6786592	0.0530830	1.9712329	1.7764526	2.1873701	0.0000000
factor:(INSREC_PUB)1	-0.1079585	0.1284547	0.8976648	0.6978682	1.1546624	0.4006614
factor:(INSREC_PUB)2	-0.3572593	0.1261293	0.6995910	0.5463649	0.8957889	0.0046188
factor:(INSREC_PUB)3	-0.1487134	0.1319291	0.8618161	0.6654514	1.1161251	0.2596487

Test PH Assumption for Sex

To test the proportional hazards assumption for Sex (a fixed-time covariate), we can create a time-dependent covariate $Z_2(t)$, defined as $Z_2(t) = Z_1 \times g(t)$, where g(t) is a known function of the time t. In most applications, we take $g(t) = \ln(t)$. A test of $H_0 : \beta_2 = 0$ is a test of the proportional hazards assumption.

```
# convert dataset into a counting process-like dataset
cut.points <- unique(breast$SRV_TIME_MON[breast$delta == 1])
breast2 <- survSplit(data = breast, cut = cut.points, end = "SRV_TIME_MON", start = "t0",
                     event = "delta")
head(breast2)
```

```
##   stage RAC_RECY SEX BRST_SUB AGE_DX Age MAR_STAT MALIGCOUNT BENBORDCOUNT
## 1     1        1   1        3     45  49        2          1            0
## 2     1        1   1        3     45  49        2          1            0
## 3     1        1   1        3     45  49        2          1            0
## 4     1        1   1        3     45  49        2          1            0
## 5     1        1   1        3     45  49        2          1            0
## 6     1        1   1        3     45  49        2          1            0
```

```
##   PRIMSITE ERSTATUS PRSTATUS INSREC_PUB t0 SRV_TIME_MON delta
## 1        4        0        0          2  0            1     0
## 2        4        0        0          2  1            2     0
## 3        4        0        0          2  2            3     0
## 4        4        0        0          2  3            4     0
## 5        4        0        0          2  4            5     0
## 6        4        0        0          2  5            6     0
```

```r
# create time-dependent covariate
breast2$tdc_sex <- breast2$SEX * log(breast2$SRV_TIME_MON)

coxph(Surv(t0, SRV_TIME_MON, delta) ~ SEX + tdc_sex, data = breast2, ties = "breslow")
```

```
## Call:
## coxph(formula = Surv(t0, SRV_TIME_MON, delta) ~ SEX + tdc_sex,
##     data = breast2, ties = "breslow")
##
##           coef exp(coef) se(coef)     z      p
## SEX      2.777    16.070    1.165  2.38 0.0171
## tdc_sex -0.976     0.377    0.342 -2.85 0.0044
##
## Likelihood ratio test=12.61  on 2 df, p=0.002
## n= 1742258, number of events= 3033
```

Using $g(t) = \ln(t)$, the Wald p-value for the test of $H_0 : \beta_2 = 0$ is 0.0044, which is significant at $\alpha = 0.05$. Thus, there is evidence that `SEX` covariate has nonproportional hazards. Therefore, we should stratify on Sex.

Stratify on Sex

Fix Cox model with the top 9 covariates, stratified on `SEX`.

```r
fit3 <- coxph(Surv(SRV_TIME_MON, delta) ~ strata(SEX) + factor(stage) + factor(RAC_RECY) +
             factor(BRST_SUB) + factor(MAR_STAT) + MALIGCOUNT + factor(PRIMSITE) +
             factor(ERSTATUS) + factor(PRSTATUS) + factor(INSREC_PUB) ,
             data = breast, ties = "breslow" )
summary(fit3)
```

```
## Call:
## coxph(formula = Surv(SRV_TIME_MON, delta) ~ strata(SEX) + factor(stage) +
##     factor(RAC_RECY) + factor(BRST_SUB) + factor(MAR_STAT) +
##     MALIGCOUNT + factor(PRIMSITE) + factor(ERSTATUS) + factor(PRSTATUS) +
##     factor(INSREC_PUB), data = breast, ties = "breslow")
##
##   n= 55333, number of events= 3033
##
##                         coef  exp(coef)   se(coef)      z Pr(>|z|)
## factor(stage)1      1.857621   6.408470   0.709702  2.617  0.00886 **
## factor(stage)2      3.077618  21.706631   0.708156  4.346 1.39e-05 ***
## factor(stage)3      4.184598  65.667066   0.708103  5.910 3.43e-09 ***
## factor(stage)4      5.832492 341.208011   0.707915  8.239  < 2e-16 ***
## factor(RAC_RECY)2   0.189521   1.208670   0.066937  2.831  0.00464 **
## factor(RAC_RECY)3   0.067450   1.069777   0.197574  0.341  0.73281
## factor(RAC_RECY)4  -0.311933   0.732030   0.077238 -4.039 5.38e-05 ***
## factor(BRST_SUB)2  -0.965198   0.380908   0.152381 -6.334 2.39e-10 ***
## factor(BRST_SUB)3   0.251216   1.285588   0.063901  3.931 8.45e-05 ***
```

```
## factor(BRST_SUB)4    -0.104480   0.900793   0.142392 -0.734  0.46310
## factor(MAR_STAT)2    -0.314578   0.730097   0.049510 -6.354 2.10e-10 ***
## factor(MAR_STAT)3    -0.099928   0.904903   0.146909 -0.680  0.49638
## factor(MAR_STAT)4    -0.008725   0.991312   0.064331 -0.136  0.89211
## factor(MAR_STAT)5     0.516334   1.675872   0.057990  8.904  < 2e-16 ***
## factor(MAR_STAT)6    -0.125466   0.882086   0.336123 -0.373  0.70895
## MALIGCOUNT            0.288792   1.334814   0.054880  5.262 1.42e-07 ***
## factor(PRIMSITE)1    -0.221547   0.801278   0.256714 -0.863  0.38813
## factor(PRIMSITE)2    -0.218838   0.803452   0.255639 -0.856  0.39197
## factor(PRIMSITE)3    -0.093926   0.910350   0.260677 -0.360  0.71861
## factor(PRIMSITE)4    -0.337854   0.713299   0.249553 -1.354  0.17579
## factor(PRIMSITE)5    -0.345290   0.708015   0.258044 -1.338  0.18086
## factor(PRIMSITE)6    -0.149278   0.861330   0.337596 -0.442  0.65836
## factor(PRIMSITE)7    -0.158763   0.853198   0.249813 -0.636  0.52508
## factor(PRIMSITE)8    -0.019407   0.980780   0.249559 -0.078  0.93801
## factor(ERSTATUS)1     1.033526   2.810961   0.128666  8.033 9.54e-16 ***
## factor(PRSTATUS)1     0.679725   1.973336   0.053092 12.803  < 2e-16 ***
## factor(INSREC_PUB)1  -0.107544   0.898037   0.128449 -0.837  0.40245
## factor(INSREC_PUB)2  -0.357413   0.699484   0.126123 -2.834  0.00460 **
## factor(INSREC_PUB)3  -0.149441   0.861189   0.131925 -1.133  0.25731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                     exp(coef) exp(-coef) lower .95 upper .95
## factor(stage)1         6.4085   0.156043    1.5946   25.7545
## factor(stage)2        21.7066   0.046069    5.4176   86.9713
## factor(stage)3        65.6671   0.015228   16.3911  263.0789
## factor(stage)4       341.2080   0.002931   85.2002 1366.4628
## factor(RAC_RECY)2      1.2087   0.827355    1.0601    1.3781
## factor(RAC_RECY)3      1.0698   0.934774    0.7263    1.5757
## factor(RAC_RECY)4      0.7320   1.366064    0.6292    0.8517
## factor(BRST_SUB)2      0.3809   2.625306    0.2826    0.5135
## factor(BRST_SUB)3      1.2856   0.777854    1.1343    1.4571
## factor(BRST_SUB)4      0.9008   1.110133    0.6814    1.1908
## factor(MAR_STAT)2      0.7301   1.369681    0.6626    0.8045
## factor(MAR_STAT)3      0.9049   1.105091    0.6785    1.2068
## factor(MAR_STAT)4      0.9913   1.008764    0.8739    1.1245
## factor(MAR_STAT)5      1.6759   0.596704    1.4958    1.8776
## factor(MAR_STAT)6      0.8821   1.133676    0.4565    1.7046
## MALIGCOUNT             1.3348   0.749168    1.1987    1.4864
## factor(PRIMSITE)1      0.8013   1.248006    0.4845    1.3253
## factor(PRIMSITE)2      0.8035   1.244630    0.4868    1.3261
## factor(PRIMSITE)3      0.9104   1.098478    0.5462    1.5174
## factor(PRIMSITE)4      0.7133   1.401936    0.4374    1.1633
## factor(PRIMSITE)5      0.7080   1.412399    0.4270    1.1741
## factor(PRIMSITE)6      0.8613   1.160996    0.4444    1.6693
## factor(PRIMSITE)7      0.8532   1.172060    0.5229    1.3922
## factor(PRIMSITE)8      0.9808   1.019597    0.6014    1.5995
## factor(ERSTATUS)1      2.8110   0.355750    2.1844    3.6172
## factor(PRSTATUS)1      1.9733   0.506756    1.7783    2.1897
## factor(INSREC_PUB)1    0.8980   1.113539    0.6982    1.1551
## factor(INSREC_PUB)2    0.6995   1.429626    0.5463    0.8956
## factor(INSREC_PUB)3    0.8612   1.161185    0.6650    1.1153
##
```

```
## Concordance= 0.883  (se = 0.006 )
## Rsquare= 0.121   (max possible= 0.677 )
## Likelihood ratio test= 7163  on 29 df,   p=<2e-16
## Wald test            = 7208  on 29 df,   p=<2e-16
## Score (logrank) test = 14991  on 29 df,   p=<2e-16
```

Use LRT to test whether the covariate effects are different between the 2 strata

```
breastSEX0 <- breast[breast$SEX == 0, ]
breastSEX1 <- breast[breast$SEX == 1, ]
fit0 <- coxph(Surv(SRV_TIME_MON, delta) ~ factor(stage) + factor(RAC_RECY) +
              factor(BRST_SUB) + factor(MAR_STAT) + MALIGCOUNT + factor(PRIMSITE) +
              factor(ERSTATUS) + factor(PRSTATUS) + factor(INSREC_PUB),
              data = breastSEX0, ties = "breslow")
```

```
## Warning in fitter(X, Y, strats, offset, init, control, weights = weights, :
## Ran out of iterations and did not converge
```

```
fit1 <- coxph(Surv(SRV_TIME_MON, delta) ~ factor(stage) + factor(RAC_RECY) +
              factor(BRST_SUB) + factor(MAR_STAT) + MALIGCOUNT + factor(PRIMSITE) +
              factor(ERSTATUS) + factor(PRSTATUS) + factor(INSREC_PUB),
              data = breastSEX1, ties = "breslow")
X2 <- -2*(fit2$loglik[2] - (fit0$loglik[2] + fit1$loglik[2])); X2
```

```
## [1] 174.0883
```

```
1 - pchisq(X2, 9) #9 degrees of freedom for each covariate
```

```
## [1] 0
```

The p-value is <<0.0001, so the assumption of using a stratified model is not met; the covariate effects are not the same between the two strata. So a stratified model is not appropriate.

```
# check ph assumption for all covariates?
test.ph <- cox.zph(fit2)
test.ph$table
```

```
##                               rho        chisq            p
## factor(SEX)1        -0.0388974061 4.684311e+00 3.043928e-02
## factor(stage)1       0.0223478133 1.512495e+00 2.187588e-01
## factor(stage)2       0.0206794504 1.295129e+00 2.551048e-01
## factor(stage)3       0.0194346335 1.143794e+00 2.848520e-01
## factor(stage)4       0.0127558319 4.931393e-01 4.825303e-01
## factor(RAC_RECY)2    0.0255096012 2.010226e+00 1.562420e-01
## factor(RAC_RECY)3    0.0004195743 5.317587e-04 9.816025e-01
## factor(RAC_RECY)4   -0.0193026541 1.124223e+00 2.890110e-01
## factor(BRST_SUB)2    0.0354526999 3.710423e+00 5.407369e-02
## factor(BRST_SUB)3   -0.0229581177 1.694065e+00 1.930660e-01
## factor(BRST_SUB)4    0.0194033685 1.112461e+00 2.915476e-01
## factor(MAR_STAT)2    0.0090206466 2.451301e-01 6.205252e-01
## factor(MAR_STAT)3   -0.0013505632 5.548827e-03 9.406201e-01
## factor(MAR_STAT)4   -0.0120024011 4.366946e-01 5.087222e-01
## factor(MAR_STAT)5   -0.0569397144 9.666059e+00 1.877036e-03
## factor(MAR_STAT)6    0.0037008708 4.145386e-02 8.386644e-01
## MALIGCOUNT           0.0429480155 5.875959e+00 1.534901e-02
## factor(PRIMSITE)1   -0.0058074981 1.051136e-01 7.457773e-01
## factor(PRIMSITE)2   -0.0099267242 3.064839e-01 5.798459e-01
## factor(PRIMSITE)3    0.0024895494 1.927140e-02 8.895912e-01
```

```
## factor(PRIMSITE)4   -0.0079056624 1.947678e-01 6.589777e-01
## factor(PRIMSITE)5    0.0027872395 2.419965e-02 8.763780e-01
## factor(PRIMSITE)6   -0.0174081386 9.365014e-01 3.331792e-01
## factor(PRIMSITE)7   -0.0108012139 3.635855e-01 5.465216e-01
## factor(PRIMSITE)8   -0.0169931415 8.985629e-01 3.431673e-01
## factor(ERSTATUS)1   -0.0559378159 9.826392e+00 1.720255e-03
## factor(PRSTATUS)1   -0.0544337799 9.306847e+00 2.282990e-03
## factor(INSREC_PUB)1  0.0252318027 1.977785e+00 1.596239e-01
## factor(INSREC_PUB)2  0.0200394778 1.258547e+00 2.619263e-01
## factor(INSREC_PUB)3  0.0136865502 5.848580e-01 4.444144e-01
## GLOBAL                         NA 1.863941e+02 1.682704e-24
```