Cox regression models with multiple predictors

Shannon Pileggi

STAT 417

OUTLINE

Multiple predictors

Inference (β_i)

Inference ($\beta_i + \beta_j$)

Overall tests

Dummy variables

Cox regression with multiple predictors

When there are several predictors $X_1, X_2, \ldots, X_p$ that we believe are associated with hazard, then the CR model becomes:

Parameters are estimated by maximizing the partial (log) likelihood function and the fitted CR model is given by:

Time invariant predictors

We will assume that the values of each predictor are measured on each individual at the beginning of the study and remain fixed over time - these are called *time invariant predictors*.

For which of the following predictors is it reasonable to assume that they are *time invariant*? In a study of time to college graduation...

1. high school GPA
2. college GPA
3. gender
4. illegal drug use (yes/no)
5. weight

Proportionality assumption

The proportionality assumption of the Cox regression model implies that for two sets of values of predictors, $\{x_1, \ldots, x_p\}$ and $\{x_1^*, \ldots, x_p^*\}$, …

1. the hazard ratio remains constant over time
2. the difference between the hazards remains constant over time
3. the ratio of log hazards remains constant over time
4. the difference between the log hazards remains constant over time

Interpretation of β_j's and $e^{c\beta_j}$'s

β_j

$e^{c\beta_j}$

VALCG study with multiple predictors

Recall the lung cancer study, and consider the predictors:

- X_1 = Karnofsky score (quantitative)
- X_2 = Cancer treatment (0 = standard, 1 = test) (categorical)

1. Write the form of the CR model with the two explanatory variables.

2. Provide an interpretation for β_1.

VALCG study with multiple predictors, cont.

3. Provide an interpretation for e^{β_2}.

VALCG study with multiple predictors: hazard ratios

Set up the hazard ratios (in terms of β's) for:

1. Patients taking the test treatment to patients taking the standard treatment (fixing Karnofsky score).

2. A ten point increase in Karnofsky score, fixing the treatment.

VALCG study with multiple predictors: hazard ratios

Set up the hazard ratios (in terms of β's) for:

3. Patients taking the test treatment whose Karnofsky score is 10 points higher than patients taking the standard treatment.

VALCG study with multiple predictors: R output

R Code

```
CR_mod1 <- coxph(Surv(time, status) ~ karno + trt, data = veteran)
summary(CR_mod1)
```

R Code

R Output

```
            coef exp(coef)  se(coef)      z Pr(>|z|)
karno -0.033954  0.966616  0.005084 -6.679  2.4e-11 ***
trt    0.177322  1.194016  0.183149  0.968    0.333
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

      exp(coef) exp(-coef) lower .95 upper .95
karno    0.9666     1.0345    0.9570    0.9763
trt      1.1940     0.8375    0.8339    1.7096
```

R Output

VALCG study w/ mult. predictors: estimated hazard ratios

Compute and interpret estimated hazard ratios for:

1. Patients taking the test treatment to patients taking the standard treatment (fixing Karnofsky score).

2. A ten point increase in Karnofsky score, fixing the treatment.

VALCG study w/ mult. predictors: estimated hazard ratios

Compute and interpret estimated hazard ratios for:

3. Patients taking the test treatment whose Karnofsky score is 10 points higher than patients taking the standard treatment.

Multiple predictors

Inference (β_i)

Inference ($\beta_i + \beta_j$)

Overall tests

Dummy variables

Full R output

R Output

```
            coef exp(coef)  se(coef)      z Pr(>|z|)
karno -0.033954  0.966616  0.005084 -6.679  2.4e-11 ***
trt    0.177322  1.194016  0.183149  0.968    0.333
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

      exp(coef) exp(-coef) lower .95 upper .95
karno    0.9666     1.0345    0.9570    0.9763
trt      1.1940     0.8375    0.8339    1.7096

Concordance= 0.712  (se = 0.03 )
Rsquare= 0.269   (max possible= 0.999 )
Likelihood ratio test= 42.97  on 2 df,   p=4.676e-10
Wald test            = 44.66  on 2 df,   p=2.001e-10
Score (logrank) test = 46.78  on 2 df,   p=6.933e-11
```

R Output

Hypothesis test for β_j

Test whether the predictor X_j has a significant effect on hazard **when all other predictors are included in the model** with:

The Wald test statistic is:

VALCG study: Wald tests

Are treatment and Karnofsky score significant predictors of hazard?

R Output
```
          coef exp(coef)  se(coef)      z Pr(>|z|)
karno -0.033954  0.966616  0.005084 -6.679  2.4e-11 ***
trt    0.177322  1.194016  0.183149  0.968    0.333
```
R Output

VALCG study: CI for HR

Construct and interpret a 95% confidence interval for the population hazard ratio of patients on the test treatment to those on the standard treatment (for fixed Karnofsky score).

R Output
```
          coef exp(coef)  se(coef)      z Pr(>|z|)
karno -0.033954  0.966616  0.005084 -6.679  2.4e-11 ***
trt    0.177322  1.194016  0.183149  0.968    0.333
```
R Output

VALCG study: CI for HR

Interpret the interval associated with Karnofsky score.

R Output
```
      exp(coef) exp(-coef) lower .95 upper .95
karno    0.9666     1.0345    0.9570    0.9763
trt      1.1940     0.8375    0.8339    1.7096
```
R Output

Multiple predictors

Inference (β_i)

Inference ($\beta_i + \beta_j$)

Overall tests

Dummy variables

Inference for a linear combination of $\beta' s$

The general form of the CI:

$$\exp\left[c\hat{\beta}_j \pm z_{\alpha/2}|c|SE(\hat{\beta}_j)\right]$$

The form of the true hazard ratio for patients taking the test treatment whose Karnofsky score is 10 points higher than patients taking the standard treatment:

$$HR = \exp[10\beta_1 + \beta_2]$$

How could we extend the above expression to this linear combination?

1. Which parts are straightforward?
2. Which parts need care?

CI for a linear combination of $\beta' s$

The $100(1-\alpha)\%$ CI for the *HR* of the general form $e^{(a\beta_i + b\beta_j)}$ is given by:

Estimated variance-covariance matrix

R Code

```r
CR_mod1$var
```

R Code

R Output

```
              [,1]          [,2]
[1,]  2.584253e-05 -0.0001026675
[2,] -1.026675e-04  0.0335433798
```

R Output

Estimated variance-covariance matrix

- The estimated covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ is:
- The estimated variance of $\hat{\beta}_1$ is:
- The estimated variance of $\hat{\beta}_2$ is:
- The standard errors of $\hat{\beta}_1$ and $\hat{\beta}_2$ are:

VALCG study: CI for HR for linear combination of β's

Compute the confidence interval for the population hazard ratio for patients taking the test treatment whose Karnofsky score is 10 points higher than patients taking the standard treatment.

HR $= \exp[10\beta_1 + \beta_2]$

Multiple predictors

Inference (β_i)

Inference ($\beta_i + \beta_j$)

Overall tests

Dummy variables

Full R output

R Output

```
           coef exp(coef)  se(coef)      z Pr(>|z|)
karno -0.033954  0.966616  0.005084 -6.679  2.4e-11 ***
trt    0.177322  1.194016  0.183149  0.968    0.333
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

      exp(coef) exp(-coef) lower .95 upper .95
karno    0.9666     1.0345    0.9570    0.9763
trt      1.1940     0.8375    0.8339    1.7096

Concordance= 0.712  (se = 0.03 )
Rsquare= 0.269   (max possible= 0.999 )
Likelihood ratio test= 42.97  on 2 df,   p=4.676e-10
Wald test            = 44.66  on 2 df,   p=2.001e-10
Score (logrank) test = 46.78  on 2 df,   p=6.933e-11
```

R Output

Overall tests

What is different about the two highlighted lines?

Three tests

1. Partial Likelihood Ratio Test:
   $$\boxed{G_l = 2\left[l_p(\hat{\beta}) - l_p(0)\right]}$$
2. Wald Test:
   $$\boxed{G_W = \hat{\beta}^T \mathbf{I}(\hat{\beta})\hat{\beta}}$$
3. Score Test:
   $$\boxed{G_S = \mathbf{u}^T(\mathbf{0})[\mathbf{I}(\mathbf{0})]^{-1}\mathbf{u}(\mathbf{0})}$$

All three test statistics (G_l, G_W, G_S) follow a χ^2-distribution with p degrees of freedom.

Details:

- $l_p(\beta)$ is the log partial likelihood function
- $\mathbf{I}(\beta)$ is the observed information matrix
- $\mathbf{0} = (0, 0, \ldots, 0)^T$ is a p-vector of 0's
- $\mathbf{u}^T(\mathbf{0})$ is the vector of partial derivatives of the log partial likelihood function (also called a vector of scores) evaluated at $\beta = \mathbf{0}$

VALCG study: interpret the results

R Output

```
Likelihood ratio test= 42.97  on 2 df,   p=4.676e-10
Wald test            = 44.66  on 2 df,   p=2.001e-10
Score (logrank) test = 46.78  on 2 df,   p=6.933e-11
```

R Output

Partial likelihood ratio test

$l_p(\hat{\beta}) =$ log partial likelihood function evaluated at the parameter estimates (measures of goodness-of-fit of the CR model to the data when the predictors $X_1, X_2, \ldots, X_p$ are included)

$l_p(0) =$ log partial likelihood function evaluated at 0 values (measures of the fit of the *null model*, i.e. a CR model with no predictors and consisting of only the baseline hazard function)

- If $l_p(\hat{\beta})$ is "much larger" than $l_p(0)$, then:
- The partial likelihood ratio test statistic compares $l_p(\hat{\beta})$ to $l_p(0)$:

Log partial likelihoods in R

R Code

```
CR_mod1$loglik
```

R Code

R Output

```
[1] -505.4491 -483.9657
```

R Output

$l_p(0) =$ left value (-505.4491)

$l_p(\hat{\beta}) =$ right value (-483.9657)

Verify that the partial likelihood ratio statistic is $G_l = 42.97$.

p-values from χ^2 distribution in R

R Code

```
pchisq(q = 42.97, df = 2, lower.tail = F)
```

R Code

R Output

```
4.668561e-10
```

R Output

Multiple predictors

Inference (β_i)

Inference ($\beta_i + \beta_j$)

Overall tests

Dummy variables

Categorical predictors with > 2 levels

- To include a categorical predictor with k levels (i.e. k different possible values) into the CR model, a set of k **dummy variables**, $D_1, D_2, \ldots, D_k$, must be created to "represent" the different values that X can take.

Categorical predictors with > 2 levels

Write out the dummy variables required for a CR model for the VALCG lung caner study with the categorical predictor X = *cancer cell type* (**small cell**, **squamous**, **large cell**, and **adenocarcinoma**).

Using dummy variables in the model

- Use $k-1$ dummy variables in your model:
- For our lung cancer model, this is:
- The k^{th} dummy variable omitted corresponds to the **reference cell**. The results for all other groups are compared relative to the k^{th} value.

Dummy variables in R

- R automatically creates dummy variables for you! If your categorical variable is:
  - *character*: you do not need to use `as.factor()` (but it won't hurt anything).
  - *numeric*: you **must** use `as.factor()` to create the dummy variables.
- R also automatically assigns the **reference** group for you:
  - *character*: first alphabetical value is the reference group
  - *numeric*: first numeric value is the reference group

  You can change the reference group in R.

R Code

```
CR_mod2 <- coxph(Surv(time, status) ~ celltype, data = veteran)
summary(CR_mod2)
```

R Code

VALCG study - CR model with `celltype`

Recall that `celltype` takes on values of:
`adeno`, `large`, `smallcell`, and `squamous`.

R Output

```
                    coef exp(coef) se(coef)      z Pr(>|z|)
celltypelarge     -0.9176    0.3995   0.2880 -3.186  0.00144 **
celltypesmallcell -0.1465    0.8638   0.2493 -0.587  0.55687
celltypesquamous  -1.1477    0.3174   0.2929 -3.919  8.9e-05 ***
```

R Output

What is the reference group?

1. `adeno`
2. `large`
3. `smallcell`
4. `squamous`

VALCG study - CR model with `celltype`

Write out the estimated Cox regression model.

R Output


```
                    coef exp(coef) se(coef)      z Pr(>|z|)
celltypelarge     -0.9176    0.3995   0.2880 -3.186  0.00144 **
celltypesmallcell -0.1465    0.8638   0.2493 -0.587  0.55687
celltypesquamous  -1.1477    0.3174   0.2929 -3.919  8.9e-05 ***
```


R Output

VALCG study - CR model with `celltype`

What is the interpretation of $\hat{\beta}_1 = -0.9176$? The ___1___ of death for patients with ___2___ lung cancer is estimated to be 0.92 ___3___ than the ___4___ of death for patients with ___5___.

1. hazard, log hazard, hazard ratio
2. adeno, large, smallcell, squamous
3. lower, higher, times
4. hazard, log hazard, hazard ratio
5. adeno, large, smallcell, squamous

VALCG study - CR model with `celltype`

R Output

```
                      coef exp(coef) se(coef)      z Pr(>|z|)
celltypelarge      -0.9176    0.3995   0.2880 -3.186  0.00144 **
celltypesmallcell -0.1465    0.8638   0.2493 -0.587  0.55687
celltypesquamous  -1.1477    0.3174   0.2929 -3.919  8.9e-05 ***
```

R Output

Identify two sets of two groups of cancers that appear to have a similar effect on hazard. Classify these sets as "better off" or "worse off".

Set 1: adeno large smallcell squamous

Set 2: adeno large smallcell squamous

VALCG study - CR model with `celltype`

R Output

```
                  exp(coef) exp(-coef) lower .95 upper .95
celltypelarge        0.3995      2.503    0.2272    0.7025
celltypesmallcell    0.8638      1.158    0.5299    1.4080
celltypesquamous     0.3174      3.151    0.1788    0.5634
```

R Output

What is the interpretation of 0.3995?

VALCG study - CR model with `celltype`

R Output

```
                  exp(coef) exp(-coef) lower .95 upper .95
celltypelarge        0.3995      2.503    0.2272    0.7025
celltypesmallcell    0.8638      1.158    0.5299    1.4080
celltypesquamous     0.3174      3.151    0.1788    0.5634
```

R Output

What is the interpretation of the interval 0.5299 - 1.4080?

VALCG study - CR model with `celltype`

R Output

```
                    coef exp(coef) se(coef)      z Pr(>|z|)
celltypelarge     -0.9176    0.3995   0.2880 -3.186  0.00144 **
celltypesmallcell -0.1465    0.8638   0.2493 -0.587  0.55687
celltypesquamous  -1.1477    0.3174   0.2929 -3.919  8.9e-05 ***
```

R Output

Estimate how many times higher (or percentage points lower) the hazard rate is for patients with small cell cancer than patients with large cell cancer.

Multiple predictors | Inference (β_i) | Inference ($\beta_i + \beta_j$) | Overall tests | Dummy variables

VALCG study - CR model with `celltype`

R Output

```
                    coef exp(coef) se(coef)      z Pr(>|z|)
celltypelarge     -0.9176    0.3995   0.2880 -3.186  0.00144 **
celltypesmallcell -0.1465    0.8638   0.2493 -0.587  0.55687
celltypesquamous  -1.1477    0.3174   0.2929 -3.919  8.9e-05 ***
```

R Output

What is the general approach to construct a CI for the population hazard ratio for patients with small cell cancer relative to patients large cell cancer?

Multiple predictors | Inference (β_i) | Inference ($\beta_i + \beta_j$) | Overall tests | Dummy variables

VALCG study - CR model with `celltype`

R Output

```
                    coef exp(coef) se(coef)      z Pr(>|z|)
celltypelarge     -0.9176    0.3995   0.2880 -3.186  0.00144 **
celltypesmallcell -0.1465    0.8638   0.2493 -0.587  0.55687
celltypesquamous  -1.1477    0.3174   0.2929 -3.919  8.9e-05 ***
...
Likelihood ratio test= 24.85  on 3 df,   p=1.661e-05
Wald test            = 24.09  on 3 df,   p=2.387e-05
Score (logrank) test = 25.51  on 3 df,   p=1.208e-05
```

R Output

Is type of cancer associated with hazard of death?