

# Homework 4

Denis Ostroushko

## Problem 1

(a)

Proportional Hazards Model:  $h(t|Z) = h_0(t) \exp(Z' \hat{\beta})$ ,

where  $Z' \hat{\beta}$  is a linear combination of predictors. In our case,  $\hat{\beta}$  is a vector of length one, a single model estimate that governs estimated differences between groups.

So, the full expression for the model is:  $h(t|Z) = h_0(t) \exp(\hat{\beta} * I(\text{Group } 2))$ , thus *Group 1* is a reference level.

(b)

We create a dummy variable to conduct regression analysis, set indicator to 0 when an individual is in an 'Untreated' group, and 1 when in 'Radiated' group

To write down the partial likelihood for the model, we need to specify the hazard function for the survival times. For Cox proportional hazards model we assume that the hazard function for each group is proportional to the hazard function for the other group, with a constant hazard ratio over the entire observed timeline.

Generally, partial likelihood for the Cox model is given by:

$$L(\beta) = \prod_{i=1}^n \frac{\exp(\beta Z_i)}{\sum_{j \in R(t_i)} \exp(\beta Z_j)}$$

where:

1.  $N$  is the sample size
2.  $\beta$  is the log hazard ratio for the radiated group compared to the untreated group

3.  $X_i$  is the value of the dummy variable for the  $i$ th individual (i.e., 0 for untreated, 1 for radiated)
4.  $T_i$  is the observed survival time for the  $i$ th individual,
5.  $\delta_i$  is an indicator variable that takes the value 1 if the survival time is uncensored

For this case with one dummy predictor the likelihood function has a simpler form

$$L = \prod_{i=1}^{12} \frac{\exp(\beta Z_i)}{\sum_{j \in R(t_i)} \exp(\beta Z_j)}$$

(c)

The data patterns and order of events and corresponding groups remains the same, so the likelihood function will remain almost the same.

(d)

The baseline hazard rate for the two models will be different due to different time values that go into the hazard equation. However, variance and betas should be the same due to similar pattern and ordering of events and corresponding groups.

## Problem 2

(a)

Summary of fitted Cox PH model is given below. We will use the test statistics and corresponding p-values to conduct the tests we are required.

```
tongue$descr <-
  factor(
    ifelse(tongue$type == 1, "Aneuploid", "Diplod"),
    levels = c("Diplod", "Aneuploid")
  )

fit <- coxph(Surv(time, delta) ~ descr, method="breslow", data=tongue)
summary(fit)
```

Call:

```
coxph(formula = Surv(time, delta) ~ descr, data = tongue, method = "breslow")
```

```
n= 80, number of events= 53
```

	coef	exp(coef)	se(coef)	z	Pr(> z )
descrAneuploid	-0.4610	0.6307	0.2805	-1.643	0.1

	exp(coef)	exp(-coef)	lower .95	upper .95
descrAneuploid	0.6307	1.586	0.3639	1.093

Concordance= 0.564 (se = 0.036 )

Likelihood ratio test= 2.61 on 1 df, p=0.1

Wald test = 2.7 on 1 df, p=0.1

Score (logrank) test = 2.75 on 1 df, p=0.1

### Score test

1. Null hypothesis: no difference in survival rates between groups
2. Alternative hypothesis: some difference in survival rates between groups
3. Test statistic: 2.75; P-value: 0.1
4. Conclusion: we do not reject null hypothesis due to lack of overwhelming statistical evidence. Although, these results might be suggestive of the fact that there is some difference between the groups

### (b)

95% confidence interval is given below, we exponentiate model estimate

```
exp(confint(fit))
```

	2.5 %	97.5 %
descrAneuploid	0.3639301	1.092955

**(c)**

Keep model with Breslow ties handling. Grab likelihood ratio test statistic and p-value.

1. Null hypothesis: no difference in survival rates between groups
2. Alternative hypothesis: some difference in survival rates between groups
3. Test statistic: 2.61; P-value = 0.1
4. Conclusion: we do not reject null hypothesis due to lack of overwhelming statistical evidence. Although, these results might be suggestive of the fact that there is some difference between the groups

**(d)**

Keep model with Breslow ties handling. Grab Wald test statistic and p-value.

1. Null hypothesis: no difference in survival rates between groups
2. Alternative hypothesis: some difference in survival rates between groups
3. Test statistic: 2.7; P-value = 0.1
4. Conclusion: we do not reject null hypothesis due to lack of overwhelming statistical evidence. Although, these results might be suggestive of the fact that there is some difference between the groups

**(e)**

Breslow

coef	exp(coef)	se(coef)
-0.4609544	0.6306814	0.2805353

coef	exp(coef)	se(coef)
-0.4663742	0.6272725	0.2804476

Estimates are almost identical. I speculate that the small number of ties is the reason for a minor discrepancy.

## Problem 3

(a)

columns `coef` and `se(coef)` provide beta estimates for Z1 and Z2

```
fit1 <- coxph(Surv(time, death) ~Z1 + Z2, data=bnct, method = "breslow")  
  
summary(fit1)$coefficient
```

	coef	exp(coef)	se(coef)	z	Pr(> z )
Z1	-1.811971	0.1633320	0.5597120	-3.237327	1.206552e-03
Z2	-3.557371	0.0285137	0.7582538	-4.691530	2.711695e-06

(b)

there is no intercept in a cox ph model, so to compare radiated to untreated we need to use estimate and variance for radiated group. First, grab a model estimate

```
      Z1  
-1.811971
```

Then grab variance of the estimate to get a standard error

```
[1] 0.3132775
```

Create a confidence interval and exponentiate the end points to get a confidence interval for the comparison

```
exp(  
  c(  
    coef(fit1)[1] - 1.96 * sqrt(vcov(fit1)[1,1]),  
    coef(fit1)[1] + 1.96 * sqrt(vcov(fit1)[1,1])  
  )  
)
```

	Z1	Z1
	0.0545299	0.4892239

(c)

To conduct this test we compare the two Cox PH models: with and without two radiation treatment variables. Test output is given below:

#### Analysis of Deviance Table

Cox model: response is Surv(time, death)

Model 1: ~ 1

Model 2: ~ Z1 + Z2

	loglik	Chisq	Df	Pr(> Chi )
1	-73.016			
2	-59.331	27.371	2	1.139e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### LRT

1. Null hypothesis:  $\beta_{Z_1} = \beta_{Z_2} = 0$
2. Not all coefficients are equal to 0
3. Test statistic: 27.371; p-value < 0.001
4. Conclusion: there is strong evidence that not all beta estimates are equal to 0 and some radiation treatment has effect on survival rates

#### Wald Test

Code for the test is given below:

```
library(aod)
wald_test <- wald.test(b = coef(fit1), Sigma = vcov(fit1), Terms = 2:length(coef(fit1)))

# View the results
wald_test
```

Wald test:

-----

Chi-squared test:

X2 = 22.0, df = 1, P(> X2) = 2.7e-06

1. Null hypothesis:  $\beta_{Z_1} = \beta_{Z_2} = 0$

2. Not all coefficients are equal to 0
3. Test statistic: 22.0; p-value < 0.001
4. Conclusion: there is strong evidence that not all beta estimates are equal to 0 and some radiation treatment has effect on survival rates

**(d)**

First, obtain the difference between coefficients:

```
diff = coef(fit1)[1] - coef(fit1)[2]
diff
```

```

      Z1
1.7454
```

Now obtain pooled variance,  $Var(Diff) = Var(\beta_1) + Var(\beta_2) - 2Cov(\beta_1, \beta_2)$

```
pooled_var = vcov(fit1)[1,1] + vcov(fit1)[2,2] - 2*vcov(fit1)[1,2]
pooled_var
```

```
[1] 0.4002733
```

Conduct Z-test

```
(diff - 0) / sqrt(pooled_var)
```

```

      Z1
2.758778
```

1. Null hypothesis:  $\beta_1 = \beta_2 = 0$
2. Alternative hypothesis:  $\beta_1 \neq \beta_2$
3. Test statistic: 2.7587776; P-value: 0.0058018
4. Conclusion: we have strong evidence to believe that the two coefficients are different. Therefore, effect of two treatments on the survival rates are not the same

**(e)**

Using previous part, obtain a confidence interval for the difference using quantiles of normal distribution. Then exponentiate the endpoints to get a confidence interval for the relative risk.

```
      Z1      Z1
1.657588 19.795135
```

**(e) again**

same as part (c)

```
Analysis of Deviance Table
Cox model: response is Surv(time, death)
Model 1: ~ 1
Model 2: ~ Z1 + Z2
      loglik  Chisq Df Pr(>|Chi|)
1 -73.016
2 -59.331 27.371  2  1.139e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Same conclusion follows

**(g)**

same as part (c)

```
Wald test:
-----
```

```
Chi-squared test:
X2 = 22.0, df = 1, P(> X2) = 2.7e-06
```

Same conclusion follows