MSCR 534: Analysis Exercise 2

Anish Shah

April 28, 2020

Assignment Description

1. To design and create tables for presentation of epidemiologic data.
2. To perform Poission regression, regression with splines, and use competing risk analysis with Cox Hazards Regression.
3. To generate and interpret measures of association using SAS.

Question 1: Poisson Regression

A researcher aimed to compare the automobile mortality rates among drivers aged 18-70 between two states (state 1 and state 0) during the year 2003. The total number count of deaths (see variable *deaths*) due to automobile accidents within each of eight age groups are listed in the datelines. The age groups are 18-22, 23-35, 36-41, 42-47, 48-53, 54-58, 59-63, and 64-70. The average number of people in each age group are included in the variable *population* which was determined by the states Department of Motor Vehicles database on number of drivers licenses.

Table 1

Fill in Table 1. Use proc genmod to obtain the rate ratios and 95% confidence intervals.

```
# Dt
dt <- as.data.table(q1)

# Crude table
crude <- dt[, .(dead = sum(deaths), pop = sum(population)), by = State, with = TRUE]

# Crude rates
crude$dead / (crude$pop / 100000)
```

```
## [1]  9.61 26.60
```

```
# Summary table for crude rate ratios
epitab(crude$dead, crude$pop, method = "rateratio")
```

```
## $tab
##           Outcome
## Predictor  crude$dead crude$pop rateratio lower upper p.value
##   Exposed1        574   5973394      1.00    NA    NA      NA
##   Exposed2       1622   6098871      2.77  2.52  3.04       0
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "midp.exact"
```

```
# Adjusted rates
dt[, rates := deaths/(population/100000)]
dt$ln_pop <- log(dt$population)
dt[, agegp := factor(agegp)]
m <- glm(deaths ~ State + agegp + offset(ln_pop), family = poisson(link = "log"), data = dt)

# Present data
tidy(m, exponentiate = TRUE, conf.int = TRUE) %>%
    kable(format = "latex", booktabs = TRUE, caption = "Adj RR") %>%
    kable_styling(latex_options = "HOLD_position")
```

Adj RR

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.00	0.168	-58.007	0.000	0.000	0.00
State	2.80	0.049	20.987	0.000	2.545	3.08
agegp2	1.53	0.211	2.024	0.043	1.018	2.34
agegp3	1.56	0.181	2.442	0.015	1.105	2.25
agegp4	2.12	0.172	4.355	0.000	1.532	3.01
agegp5	2.01	0.171	4.091	0.000	1.461	2.86
agegp6	1.19	0.172	0.995	0.320	0.860	1.69
agegp7	1.89	0.171	3.713	0.000	1.370	2.69
agegp8	1.35	0.177	1.703	0.088	0.969	1.94

Question 2: Splines

For Question 2, use the SAS dataset *cohort* from Lab 8. In Lab 8 we used four equal cut-points and a quadratic spline in a Cox Hazards regression using a SAS macro. For Question 2, redo this spline regression analysis by using the same SAS macro with 3 equal knots (tertiles). Use the data to complete Table 2 and answer Questions 2 below.

```
cohort <- read_sas("cohort.sas7bdat")

# Unadjusted
m <- coxph(Surv(timetooc, outc2) ~ baserbg2, data = cohort)

# Tertiles
cohort$glugrp <- cut2(cohort$baserbg2, g = 3) %>% factor()
cohort$agecat %<>% factor()
m3 <- coxph(Surv(timetooc, outc2) ~ glugrp, data = cohort)
m3adj <- coxph(Surv(timetooc, outc2) ~ glugrp + agecat, data = cohort)

# Splines
k <- quantile(cohort$baserbg2, p = c(0.25, 0.50, 0.75))
ms <- coxph(Surv(timetooc, outc2) ~ bs(baserbg2, knots = k, degree = 1), data = cohort)
coxph(Surv(timetooc, outc2) ~ bs(baserbg2, knots = k, degree = 1), data = cohort) %>% summary()
```

```
## Call:
## coxph(formula = Surv(timetooc, outc2) ~ bs(baserbg2, knots = k,
##     degree = 1), data = cohort)
##
##   n= 25872, number of events= 3986
##
##                                          coef exp(coef) se(coef)     z Pr(>|z|)
## bs(baserbg2, knots = k, degree = 1)1 -0.0982    0.9065   0.1206 -0.81     0.42
## bs(baserbg2, knots = k, degree = 1)2 -0.1275    0.8803   0.1016 -1.25     0.21
## bs(baserbg2, knots = k, degree = 1)3  0.1037    1.1093   0.1035  1.00     0.32
## bs(baserbg2, knots = k, degree = 1)4  0.2484    1.2819   0.3711  0.67     0.50
##
##                                      exp(coef) exp(-coef) lower .95 upper .95
## bs(baserbg2, knots = k, degree = 1)1     0.906      1.103     0.716      1.15
## bs(baserbg2, knots = k, degree = 1)2     0.880      1.136     0.721      1.07
## bs(baserbg2, knots = k, degree = 1)3     1.109      0.902     0.906      1.36
## bs(baserbg2, knots = k, degree = 1)4     1.282      0.780     0.619      2.65
##
## Concordance= 0.525  (se = 0.005 )
## Likelihood ratio test= 33.4  on 4 df,   p=0.000001
## Wald test            = 33.9  on 4 df,   p=0.0000008
## Score (logrank) test = 34  on 4 df,   p=0.0000008
```

```
tidy(ms, exponentiate = TRUE, conf.int = TRUE)
```

```
## # A tibble: 4 x 7
##   term                    estimate std.error statistic p.value conf.low conf.high
##   <chr>                      <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 bs(baserbg2, knots = ~     0.906     0.121    -0.814   0.416    0.716      1.15
```

```
## 2 bs(baserbg2, knots = ~    0.880     0.102    -1.25    0.210    0.721      1.07
## 3 bs(baserbg2, knots = ~    1.11      0.103     1.00    0.316    0.906      1.36
## 4 bs(baserbg2, knots = ~    1.28      0.371     0.669   0.503    0.619      2.65
```

Question 3: Competing Risk Analysis

A cohort study aimed to determine the hazard rate of incident of invasive pulmonary Aspergillosis (fungal infection) in patients with idiopathic pulmonary fibrosis. The primary exposure of interest was biologic sex and a competing risk of concern was all-cause mortality. Assume all participants were free of Aspergillosis at baseline and were at risk of disease.

```
q3 <- read_sas("analysisexercise2question3.sas7bdat") %>%
    as.data.table()
cols = c("sex", "censor1")
q3[, (cols) := lapply(.SD, as.factor), .SDcols = cols]
```

What is the crude risk ratio for incident Aspergillosis, comparing men to women?

```
# Crude risk
m <- glm(censor1 ~ sex, family=binomial("log"), q3[censor1 != 2])
tidy(m, exponentiate = TRUE, conf.int = TRUE) %>%
    kable(format = "latex", booktabs = TRUE, caption = "Crude RR") %>%
    kable_styling(latex_options = "HOLD_position")
```

Crude RR

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0.093	0.183	-12.95	0.000	0.063	0.13
sex1	1.322	0.211	1.32	0.187	0.887	2.04

```
# Median time
tmp <- median(q3$timetoc2)
```

Table 3A

```
# Turn death into potential event (dropping competing risks)
q3$censor <- ifelse(q3$censor1 == 1, 1, ifelse(q3$censor1 == 2, 1, 0))

# Crude HR
m <- coxph(Surv(timetoc2, censor) ~ sex, data = q3)
tidy(m, exponentiate = TRUE, conf.int = TRUE) %>%
    kable(format = "latex", booktabs = TRUE, caption = "Crude HR") %>%
    kable_styling(latex_options = "HOLD_position")
```

Crude HR

term	estimate	std.error	statistic	p.value	conf.low	conf.high
sex1	1.37	0.198	1.6	0.108	0.932	2.02

```r
# Adjusted
m <- coxph(Surv(timetoc2, censor) ~ sex + age + smk, data = q3)
tidy(m, exponentiate = TRUE, conf.int = TRUE)  %>%
    kable(format = "latex", booktabs = TRUE, caption = "Crude HR") %>%
    kable_styling(latex_options = "HOLD_position")
```

Crude HR

term	estimate	std.error	statistic	p.value	conf.low	conf.high
sex1	1.512	0.199	2.08	0.038	1.023	2.235
age	0.827	0.096	-1.97	0.049	0.685	0.999
smk	1.140	0.028	4.74	0.000	1.080	1.203

Table 3B

```r
### Cause specific for aspy
q3$censor <- ifelse(q3$censor1 == 1, 1, 0)

# Crude HR
m <- coxph(Surv(timetoc2, censor) ~ sex, data = q3)
tidy(m, exponentiate = TRUE, conf.int = TRUE) %>%
    kable(format = "latex", booktabs = TRUE, caption = "Crude Cause = Asp") %>%
    kable_styling(latex_options = "HOLD_position")
```

Crude Cause = Asp

term	estimate	std.error	statistic	p.value	conf.low	conf.high
sex1	1.23	0.226	0.92	0.357	0.79	1.92

```r
# Adjusted
m <- coxph(Surv(timetoc2, censor) ~ sex + age + smk, data = q3)
tidy(m, exponentiate = TRUE, conf.int = TRUE) %>%
    kable(format = "latex", booktabs = TRUE, caption = "Adj Cause = Asp") %>%
    kable_styling(latex_options = "HOLD_position")
```

Adj Cause = Asp

term	estimate	std.error	statistic	p.value	conf.low	conf.high
sex1	1.416	0.227	1.53	0.126	0.907	2.211
age	0.578	0.115	-4.76	0.000	0.462	0.725
smk	1.194	0.030	5.84	0.000	1.125	1.267

```r
### Cause specific for dead
q3$censor <- ifelse(q3$censor1 == 2, 1, 0)

# Crude HR
```

```r
m <- coxph(Surv(timetoc2, censor) ~ sex, data = q3)
tidy(m, exponentiate = TRUE, conf.int = TRUE) %>%
    kable(format = "latex", booktabs = TRUE, caption = "Crude Cause = Dead") %>%
    kable_styling(latex_options = "HOLD_position")
```

Crude Cause = Dead

term	estimate	std.error	statistic	p.value	conf.low	conf.high
sex1	1.89	0.414	1.54	0.123	0.841	4.27

```r
# Adjusted
m <- coxph(Surv(timetoc2, censor) ~ sex + age + smk, data = q3)
tidy(m, exponentiate = TRUE, conf.int = TRUE) %>%
    kable(format = "latex", booktabs = TRUE, caption = "Adj Cause = Dead") %>%
    kable_styling(latex_options = "HOLD_position")
```

Adj Cause = Dead

term	estimate	std.error	statistic	p.value	conf.low	conf.high
sex1	1.701	0.421	1.260	0.208	0.745	3.88
age	2.201	0.202	3.897	0.000	1.480	3.27
smk	0.994	0.064	-0.099	0.921	0.876	1.13

```r
### Subdistribution for asp

# Crude
m <- FGR(Hist(timetoc2, censor1, cens.code = 0) ~ sex, cause = 1, data = q3)

# Adjustd
m <- FGR(Hist(timetoc2, censor1, cens.code = 0) ~ sex + age + smk, cause = 1, data = q3)

### Subdistribution for asp

# Crude
m <- FGR(Hist(timetoc2, censor1, cens.code = 0) ~ sex, cause = 2, data = q3)

# Adjustd
m <- FGR(Hist(timetoc2, censor1, cens.code = 0) ~ sex + age + smk, cause = 2, data = q3)
```

For the subdistribution model, with outcome of death, here are the summary of failure outcomes.

```r
x <- addmargins(table(q3$censor1))
names(x) <- c("Censored", "Competing Event", "Event of Interest", "Total")
x %>%
    kable(caption = "Summary: Failure Outcomes", format = "latex") %>%
    kable_styling(latex_options = "HOLD_position")
```

Summary: Failure Outcomes

Var1	Freq
Censored	826
Competing Event	106
Event of Interest	44
Total	976