MSCR 534: Final Exam

Anish Shah

May 1, 2020

Question 1

Answered in word document.

Question 2

```
B0 = -4.4195 # Intercept
B1 = 1.0221 # FPM exposure
B2 = 0.4210 # age <60
B3 = 0.5600 # age >= 60
B4 = 0.7712 # FPM with young age
B5 = 0.8211 # FPM with old age

# 65 year old with high FPM exposure
estimate = B0 + B1 + B3 + B5
cumInc = exp(estimate)

# 38 year old with low FPM exposure
estimate = B0 + B2
cumInc = exp(estimate)

# 42 with High FPM compared to 70 with low FPM
RR = exp(B1 + B2 + B4) / exp(B3)
```

Question 3

Answered in word document.

Question 4

Answered in word document.

Question 5

Table 3 filled out using this information.

```r
# Data
q5 <- read_sas("q5.sas7bdat") %>%
    clean_names()

df <- q5

# Risk in hip fracture
df$hip55 %<>% factor(levels = c(1,0), labels = c("Fx", "NoFx"))
df$livea %<>% factor(levels = c(1,0), labels = c("Alone", "Others"))

tbl <- table(df$livea, df$hip55)
epitbl <- epi.2by2(tbl)

# Print
print(epitbl)
```

```
##              Outcome +    Outcome -      Total        Inc risk *        Odds
## Exposed +          215          298        513              41.9       0.721
## Exposed -          355          842       1197              29.7       0.422
## Total              570         1140       1710              33.3       0.500
##
## Point estimates and 95% CIs:
## -------------------------------------------------------------------
## Inc risk ratio                               1.41 (1.24, 1.62)
## Odds ratio                                   1.71 (1.38, 2.12)
## Attrib risk *                                12.25 (7.26, 17.25)
## Attrib risk in population *                  3.68 (0.26, 7.09)
## Attrib fraction in exposed (%)               29.24 (19.08, 38.12)
## Attrib fraction in population (%)            11.03 (6.35, 15.47)
## -------------------------------------------------------------------
##  Test that OR = 1: chi2(1) = 24.261 Pr>chi2 = <0.001
##  Wald confidence limits
##  CI: confidence interval
##  * Outcomes per 100 population units
```

Verified odds ratio via comparison of groups.

```r
# Crude risk of hip fracture
compareGroups(hip55 ~ livea, data = df) %>%
    createTable(show.all = TRUE, show.ratio = TRUE) %>%
    export2md(format = "latex", caption = "Comparison of Hip Fx and Loneliness")
```

Comparison of Hip Fx and Loneliness

	[ALL] N=1710	Fx N=570	NoFx N=1140	OR	p.ratio	p.overall
livea:						<0.001
Alone	513 (30.0%)	215 (37.7%)	298 (26.1%)	Ref.	Ref.	
Others	1197 (70.0%)	355 (62.3%)	842 (73.9%)	1.71 [1.38;2.12]	<0.001	

Adjusted OR done through logistic regression.

```
m <- glm(hip55 ~ livea + alc + ager, family = binomial("logit"), data = q5)

# Make pretty
stargazer(m, header = FALSE, type = "latex",
                    apply.coef = exp,
                    p.auto = FALSE,
                    report = "vc*s", title = "Adjusted OR for Hip Fx",
                    ci.custom = list(exp(confint(m))))
```

Adjusted OR for Hip Fx

	Dependent variable:
	hip55
livea	1.380***
	(1.100, 1.740)
alc	0.866**
	(0.762, 0.983)
ager	3.290***
	(2.770, 3.930)
Constant	0.037***
	(0.023, 0.060)
Observations	1,710
Log Likelihood	−967.000
Akaike Inf. Crit.	1,942.000
Note:	*p<0.1; **p<0.05; ***p<0.01

Question 6

We will need to establish several relationships to study the mediation effects of depression on the relationship between living alone and hip fractures.

We can test the relationship of living alone with depression.

```
# Data
df <- q5

# Tidy
df$livea %<>% factor(levels = c(0,1), labels = c("Others", "Alone"))
df$dep %<>% factor(levels = c(0,1), labels = c("Happy", "Sad"))
df$hip55 %<>% factor(levels = c(0,1), labels = c("NoFx", "Fx"))

compareGroups(dep ~ livea, data = df) %>%
    createTable(show.ratio = TRUE, show.all = TRUE) %>%
    export2md(format = "latex", caption = "Loneliness and Depression")
```

Loneliness and Depression

	[ALL] N=1710	Happy N=1487	Sad N=223	OR	p.ratio	p.overall
livea:						0.009
Others	1197 (70.0%)	1058 (71.1%)	139 (62.3%)	Ref.	Ref.	
Alone	513 (30.0%)	429 (28.9%)	84 (37.7%)	1.49 [1.11;2.00]	0.009	

We can then test the relationship of the potential mediator of depression on hip fractures.

```
# Data as above
compareGroups(hip55 ~ dep, data = df) %>%
    createTable(show.ratio = TRUE, show.all = TRUE) %>%
    export2md(format = "latex", caption = "Depression and Fx")
```

Depression and Fx

	[ALL] N=1710	NoFx N=1140	Fx N=570	OR	p.ratio	p.overall
dep:						0.045
Happy	1487 (87.0%)	1005 (88.2%)	482 (84.6%)	Ref.	Ref.	
Sad	223 (13.0%)	135 (11.8%)	88 (15.4%)	1.36 [1.01;1.81]	0.040	

Now, since depression is associated with both exposure and outcome, we can test it as a mediator using an adjusted regression model.

```
# Three models
m1 <- glm(hip55 ~ livea, family = binomial("logit"), data = q5)
m2 <- glm(hip55 ~ dep, family = binomial("logit"), data = q5)
m3 <- glm(hip55 ~ dep + livea, family = binomial("logit"), data = q5)

# Make pretty
stargazer(m1, m2, m3, header = FALSE, type = "latex",
                    apply.coef = exp,
                    p.auto = FALSE,
                    report = "vc*s", title = "Mediation Analysis of Loneliness, Sadness, and Fractures"
                    ci.custom = list(exp(confint(m1)), exp(confint(m2)), exp(confint(m3)))
                    )
```

To mesure effects, we need to look at measured relationships in the causal model.

- Living Alone -> Depression
- Living Alone + Depression -> Hip Fx

```
# Models
m <- glm(dep ~ livea, family = binomial("logit"), data = q5)
y <- glm(hip55 ~ livea + dep, family = binomial("logit"), data = q5)

# Mediation
res <- mediate(m, y, treat = "livea", mediator = "dep", boot = TRUE, sims = 500)
summary(res)
```

Mediation Analysis of Loneliness, Sadness, and Fractures

	Dependent variable:		
	hip55		
	(1)	(2)	(3)
livea	1.710***		1.690***
	(1.380, 2.120)		(1.360, 2.100)
dep		1.360**	1.300*
		(1.010, 1.810)	(0.969, 1.740)
Constant	0.422***	0.480***	0.408***
	(0.372, 0.477)	(0.430, 0.534)	(0.358, 0.464)
Observations	1,710	1,710	1,710
Log Likelihood	−1,077.000	−1,086.000	−1,075.000
Akaike Inf. Crit.	2,157.000	2,177.000	2,156.000
Note:			*p<0.1; **p<0.05; ***p<0.01

```
## 
## Causal Mediation Analysis 
## 
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
## 
##                           Estimate 95% CI Lower 95% CI Upper
## ACME (control)            0.003272    -0.000289         0.01
## ACME (treated)            0.003718    -0.000334         0.01
## ADE (control)             0.119373     0.066169         0.17
## ADE (treated)             0.119819     0.066559         0.17
## Total Effect              0.123091     0.069721         0.17
## Prop. Mediated (control)  0.026582    -0.002229         0.07
## Prop. Mediated (treated)  0.030204    -0.002574         0.08
## ACME (average)            0.003495    -0.000312         0.01
## ADE (average)             0.119596     0.066364         0.17
## Prop. Mediated (average)  0.028393    -0.002402         0.07
##                                     p-value    
## ACME (control)                          0.1 .  
## ACME (treated)                          0.1 .  
## ADE (control)            <0.0000000000000002 ***
## ADE (treated)            <0.0000000000000002 ***
## Total Effect             <0.0000000000000002 ***
## Prop. Mediated (control)                0.1 .  
## Prop. Mediated (treated)                0.1 .  
## ACME (average)                          0.1 .  
## ADE (average)            <0.0000000000000002 ***
## Prop. Mediated (average)                0.1 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Sample Size Used: 1710
```

```
##
##
## Simulations: 500
```

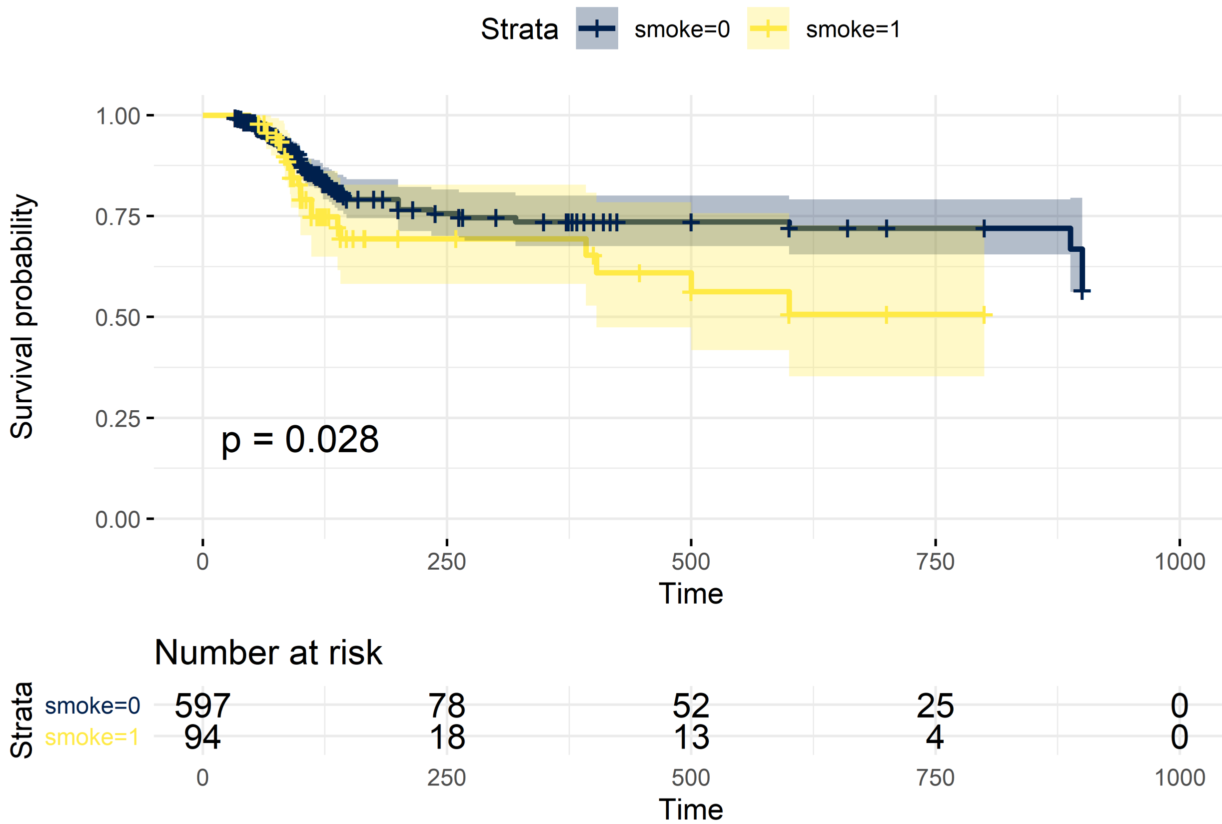
This supports the idea the depression only accounts for a very small part (indirect effect) of the total effect of living alone on hip fractures.

Question 7

This creates a Kaplan Meier curve for assessing lung cancer incidence by smoking strata.

```
q7 <- read_sas("q7.sas7bdat") %>%
    clean_names()

# Create a KM graph (exclude death during f/u)
df <- subset(q7, outcome != 1)
df$outcome[df$outcome == 2] <- 1
df$smoke %<>% factor()
fit <- survfit(Surv(survival_t, outcome) ~ smoke, data = df)
ggsurvplot(
    fit,
    data = df,
    risk.table = TRUE,
    pval = TRUE,
    conf.int = TRUE,
    ggtheme = theme_minimal(),
    palette = viridis_pal(option = "E")(2)
)
```



Now we need to calculate Cox regression to create a hazard rate ratio, as well the subdistribution models.

```
# Data excluding those who died from analysis, as above
df <- subset(q7, outcome != 1)
df$outcome[df$outcome == 2] <- 1
df$smoke %<>% factor()
df$agecat %<>% factor()
df$sex %<>% factor()

# Cox regression
m1 <- coxph(Surv(survival_t, outcome) ~ smoke + agecat + sex, data = df)
tidy(m1, exponentiate = TRUE, conf.int = TRUE) %>%
    kable(format = "latex",
                booktabs = TRUE,
                caption = "Adjusted HR for Lung Cancer") %>%
    kable_styling(latex_options = "HOLD_position")
```

Adjusted HR for Lung Cancer

term	estimate	std.error	statistic	p.value	conf.low	conf.high
smoke1	1.74	0.234	2.36	0.018	1.098	2.75
agecat1	1.59	0.253	1.84	0.066	0.969	2.61
sex1	1.30	0.225	1.18	0.238	0.839	2.03

```r
# Subdistribution for lung cancer
df <- q7
df$smoke %<>% factor()
df$agecat %<>% factor()
df$sex %<>% factor()

m2 <- FGR(Hist(survival_t, outcome, cens.code = 0) ~ smoke + agecat + sex,
                    cause = 2,
                    data = df)
print(m2)
```

```
##
## Right-censored response of a competing.risks model
##
## No.Observations: 729
##
## Pattern:
##
## Cause     event right.censored
##   1          38              0
##   2         110              0
##   unknown     0            581
##
##
## Fine-Gray model: analysis of cause 2
##
## Competing Risks Regression
##
## Call:
## FGR(formula = Hist(survival_t, outcome, cens.code = 0) ~ smoke +
##     agecat + sex, data = df, cause = 2)
##
##          coef exp(coef) se(coef)    z p-value
## smoke1  0.344      1.41    0.227 1.52    0.13
## agecat1 0.295      1.34    0.251 1.18    0.24
## sex1    0.210      1.23    0.224 0.94    0.35
##
##         exp(coef) exp(-coef)  2.5% 97.5%
## smoke1       1.41      0.709 0.904  2.20
## agecat1      1.34      0.744 0.821  2.20
## sex1         1.23      0.810 0.796  1.91
##
## Num. cases = 729
## Pseudo Log-likelihood = -648
## Pseudo likelihood ratio test = 4.12  on 3 df,
##
## Convergence: TRUE
```

```r
# Subdistribution for death
m3 <- FGR(Hist(survival_t, outcome, cens.code = 0) ~ smoke + agecat + sex,
                    cause = 1,
                    data = df)
print(m3)
```

```
##
## Right-censored response of a competing.risks model
##
## No.Observations: 729
##
## Pattern:
##
## Cause     event right.censored
##   1          38              0
##   2         110              0
##   unknown     0            581
##
##
## Fine-Gray model: analysis of cause 1
##
## Competing Risks Regression
##
## Call:
## FGR(formula = Hist(survival_t, outcome, cens.code = 0) ~ smoke +
##     agecat + sex, data = df, cause = 1)
##
##          coef exp(coef) se(coef)    z p-value
## smoke1  1.051      2.86    0.331 3.18  0.0015
## agecat1 1.095      2.99    0.355 3.08  0.0021
## sex1    0.538      1.71    0.424 1.27  0.2000
##
##         exp(coef) exp(-coef)  2.5% 97.5%
## smoke1       2.86      0.350 1.495  5.47
## agecat1      2.99      0.335 1.490  5.99
## sex1         1.71      0.584 0.746  3.93
##
## Num. cases = 729
## Pseudo Log-likelihood = -203
## Pseudo likelihood ratio test = 16.1  on 3 df,
##
## Convergence: TRUE
```