

# BST 210 HW7 Wenjie Gu

Wenjie Gu

```
library(knitr)
hook_output = knit_hooks$get('output')
knit_hooks$set(output = function(x, options) {
  # this hook is used only when the linewidth option is not NULL
  if (!is.null(n <- options$linewidth)) {
    x = knitr:::split_lines(x)
    # any lines wider than n should be wrapped
    if (any(nchar(x) > n)) x = strwrap(x, width = n)
    x = paste(x, collapse = '\n')
  }
  hook_output(x, options)
})
```

```
# load file
dat = read.table("mlowbwt_data_only.txt", head = T)
colnames(dat) = tolower(colnames(dat))
head(dat)
```

```
##   stratum obs age low lwt smoke ht ui ptd
## 1      1   1  16   1 130     0 0 0   0
## 2      1   2  16   0 112     0 0 0   0
## 3      1   3  16   0 135     1 0 0   0
## 4      1   4  16   0  95     0 0 0   0
## 5      2   1  17   1 130     1 0 1   1
## 6      2   2  17   0 103     0 0 0   0
```

## Problem 1

### 1(a)

```
library("survival")
mod1a = clogit(low ~ ptd + strata(stratum), data = dat)
summary(mod1a)
```

```
## Call:
## coxph(formula = Surv(rep(1, 116L), low) ~ ptd + strata(stratum),
##       data = dat, method = "exact")
##
##   n= 116, number of events= 29
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## ptd 1.9371      6.9389   0.5914 3.275  0.00105 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## ptd      6.939      0.1441      2.177      22.12
##
## Concordance= 0.649 (se = 0.055 )
```

```
## Likelihood ratio test= 12.5 on 1 df, p=4e-04
## Wald test = 10.73 on 1 df, p=0.001
## Score (logrank) test = 14.08 on 1 df, p=2e-04
```

```
exp(confint(mod1a))
```

```
##          2.5 %    97.5 %
## ptd 2.177144 22.11563
```

We estimated that on average the odds having low birth weight for mothers who have history of premature labor is 6.9389 times the odds of low birth weight among those without history of premature labor, with a 95% confidence interval of (2.177144, 22.11563).

### 1(b)

```
mod1b = clogit(low ~ ptd + age + strata(stratum), data = dat, method = "exact")
summary(mod1b)
```

```
## Call:
## coxph(formula = Surv(rep(1, 116L), low) ~ ptd + age + strata(stratum),
##       data = dat, method = "exact")
##
## n= 116, number of events= 29
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## ptd 1.9371    6.9389   0.5914 3.275 0.00105 **
## age      NA         NA   0.0000   NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## ptd    6.939    0.1441    2.177    22.12
## age      NA         NA      NA      NA
##
## Concordance= 0.649 (se = 0.055 )
## Likelihood ratio test= 12.5 on 1 df, p=4e-04
## Wald test = 10.73 on 1 df, p=0.001
## Score (logrank) test = 14.08 on 1 df, p=2e-04
```

The coefficient of age in this model is NA, meaning that we cannot estimate the coefficient of factor age. This makes sense because in this study cases and controls are matched on age and we already added strata() to account for this. If we are matching on a factor, we cannot estimate the beta for that factor and we can only say that we are controlling for that factor with CMLE.

### 1(c)

```
mod1c = clogit(low ~ ptd + strata(age), data = dat, method = "exact")
summary(mod1c)
```

```
## Call:
## coxph(formula = Surv(rep(1, 116L), low) ~ ptd + strata(age),
##       data = dat, method = "exact")
##
## n= 116, number of events= 29
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
```

```
## ptd 1.8061      6.0866    0.5394 3.349 0.000812 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## ptd      6.087      0.1643      2.115      17.52
##
## Concordance= 0.648 (se = 0.053 )
## Likelihood ratio test= 11.96 on 1 df,  p=5e-04
## Wald test              = 11.21 on 1 df,  p=8e-04
## Score (logrank) test = 13.81 on 1 df,  p=2e-04
```

```
summary(mod1a)
```

```
## Call:
## coxph(formula = Surv(rep(1, 116L), low) ~ ptd + strata(stratum),
##       data = dat, method = "exact")
##
##      n= 116, number of events= 29
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## ptd 1.9371      6.9389  0.5914 3.275  0.00105 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## ptd      6.939      0.1441      2.177      22.12
##
## Concordance= 0.649 (se = 0.055 )
## Likelihood ratio test= 12.5 on 1 df,  p=4e-04
## Wald test              = 10.73 on 1 df,  p=0.001
## Score (logrank) test = 14.08 on 1 df,  p=2e-04
```

I agree that one could use AGE as the matching variable. After this change, I found that the coefficient of ptd is different from the coefficient from 1b. I would prefer using AGE as matching variable rather than STRATUM. Because if we include subjects with the same age into the same matched set, the estimated coefficient will be the estimated log odds ratio of low birth weight between mothers with and without history of premature labor, holding mother's age constant rather than holding the "stratum group" constant.

1(d)

```
mod1d = clogit(low ~ ptd + age*ptd + strata(age), data = dat, method = "exact")
summary(mod1d)
```

```
## Call:
## coxph(formula = Surv(rep(1, 116L), low) ~ ptd + age * ptd + strata(age),
##       data = dat, method = "exact")
##
##      n= 116, number of events= 29
##
##      coef exp(coef) se(coef)      z Pr(>|z|)
## ptd      3.57700  35.76622  2.88673  1.239  0.215
## age              NA         NA  0.00000   NA     NA
## ptd:age -0.07523   0.92753  0.11936 -0.630  0.529
##
```

```
##          exp(coef) exp(-coef) lower .95 upper .95
## ptd      35.7662    0.02796    0.1248 10248.531
## age              NA          NA          NA      NA
## ptd:age    0.9275    1.07813    0.7341    1.172
##
## Concordance= 0.648 (se = 0.053 )
## Likelihood ratio test= 12.35 on 2 df,  p=0.002
## Wald test              = 11.36 on 2 df,  p=0.003
## Score (logrank) test = 14.4 on 2 df,  p=7e-04
```

Yes, we can assess if age is an effect modifier. After including the interaction between age and ptd into the conditional logistic regression model, we found that the coefficient of the interaction term is not significant ( $p > 0.05$ ). Therefore, age is not an effect modifier for the effect of ptd on low birth weight.

1(e)

```
mod1e = glm(low ~ ptd + age, family = binomial(), data = dat)
summary(mod1e)
```

```
##
## Call:
## glm(formula = low ~ ptd + age, family = binomial(), data = dat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4063  -0.6568  -0.6292  -0.2046   1.9202
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0631     1.2132  -0.876  0.38090
## ptd           1.9667     0.5546   3.546  0.00039 ***
## age          -0.0190     0.0528  -0.360  0.71891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 130.46 on 115 degrees of freedom
## Residual deviance: 117.40 on 113 degrees of freedom
## AIC: 123.4
##
## Number of Fisher Scoring iterations: 4
```

```
summary(mod1c)
```

```
## Call:
## coxph(formula = Surv(rep(1, 116L), low) ~ ptd + strata(age),
##       data = dat, method = "exact")
##
##      n= 116, number of events= 29
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## ptd 1.8061      6.0866   0.5394  3.349 0.000812 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##      exp(coef) exp(-coef) lower .95 upper .95
## ptd      6.087    0.1643    2.115    17.52
##
## Concordance= 0.648 (se = 0.053 )
## Likelihood ratio test= 11.96 on 1 df,  p=5e-04
## Wald test              = 11.21 on 1 df,  p=8e-04
## Score (logrank) test = 13.81 on 1 df,  p=2e-04
```

The coefficient of ptd is smaller in the conditional logistic regression model than the unconditional model. Also the standard error of the coefficient in the conditional logistic regression model is also smaller. The unconditional logistic regression model is not an appropriate approach in the analysis of a matched case-control study. The dataset was already collected in a matched design, we have to choose an analysis that accounts for that design => Conditional logistic regression model.

## Problem 2

### 2(a)

```
library("foreign")
dat2 = read.table("fromlungcancer.txt", sep = "", header = FALSE)
colnames(dat2) = c("smokedur", "cigpday", "cases", "manyears")
View(dat2)

mod2a = glm(cases~cigpday, family = poisson(), offset = log(manyears), data = dat2)
mod2a.1 = glm(cases~cigpday + smokedur, family = poisson(), offset = log(manyears), data = dat2)
# Assess confounding
summary(mod2a)

##
## Call:
## glm(formula = cases ~ cigpday, family = poisson(), data = dat2,
##      offset = log(manyears))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.8684  -1.6542  -0.5111   1.8415   4.8937
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -8.159268   0.175063  -46.61  <2e-16 ***
## cigpday      0.070359   0.006468   10.88  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 445.10  on 62  degrees of freedom
## Residual deviance: 324.71  on 61  degrees of freedom
## AIC: 448.55
##
## Number of Fisher Scoring iterations: 6
summary(mod2a.1)

##
```

```
## Call:
## glm(formula = cases ~ cigpday + smokedur, family = poisson(),
##      data = dat2, offset = log(manyyears))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2706  -1.1369  -0.4592   0.6675   2.4863
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -12.215635   0.377854  -32.33  <2e-16 ***
## cigpday      0.066712   0.006323   10.55  <2e-16 ***
## smokedur     0.111457   0.007508   14.85  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 445.099  on 62  degrees of freedom
## Residual deviance:  82.243  on 60  degrees of freedom
## AIC: 208.09
##
## Number of Fisher Scoring iterations: 5
```

Smoking duration is a potential confounder for the effect of cigarettes smoking per day on lung cancer incidence: 1) Smoking duration is associated with cigarettes smoking per day in the absence of lung cancer. 2) Smoking duration is associated with lung cancer. 3) Smoking duration is not a downstream effect of cigarettes smoking per day.

Now, we can assess if smoking duration is a confounder using our 10% rule: Since the coefficient for cigpday changes from 0.070359 to 0.066712 after including smokedur in the poisson regression model (decrease by 5.1%), which is less than 10%, smokedur is not a confounder of the effect of cigpday on log lung cancer incidence.

```
# Assess effect modification
```

```
mod2a.2 = glm(cases ~ cigpday + smokedur * cigpday, family = poisson(), offset = log(manyyears), data = dat2)
summary(mod2a.2)
```

```
##
## Call:
## glm(formula = cases ~ cigpday + smokedur * cigpday, family = poisson(),
##      data = dat2, offset = log(manyyears))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3352  -1.1180  -0.4311   0.6454   2.4962
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.247e+01  7.386e-01 -16.887  < 2e-16 ***
## cigpday       7.784e-02  2.796e-02   2.784  0.00537 **
## smokedur     1.172e-01  1.598e-02   7.335  2.21e-13 ***
## cigpday:smokedur -2.516e-04  6.152e-04  -0.409  0.68256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 445.099 on 62 degrees of freedom
## Residual deviance: 82.076 on 59 degrees of freedom
## AIC: 209.92
##
## Number of Fisher Scoring iterations: 5
```

Since the interaction between cigpday and smokedur is not significant according to Wald test, we can say that cigpday does not modify the effect of smokedur on log lung cancer incidence.

```
summary(mod2a)
```

```
##
## Call:
## glm(formula = cases ~ cigpday, family = poisson(), data = dat2,
## offset = log(manyyears))
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -3.8684 -1.6542 -0.5111 1.8415 4.8937
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -8.159268 0.175063 -46.61 <2e-16 ***
## cigpday 0.070359 0.006468 10.88 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 445.10 on 62 degrees of freedom
## Residual deviance: 324.71 on 61 degrees of freedom
## AIC: 448.55
##
## Number of Fisher Scoring iterations: 6
```

```
confint(mod2a)
```

```
## Waiting for profiling to be done...
##
## 2.5 % 97.5 %
## (Intercept) -8.51368127 -7.82700207
## cigpday 0.05773514 0.08310548
```

```
cat(sprintf("The incidence rate ratio of lung cancer comparing a population of individuals with a fixed
```

```
## The incidence rate ratio of lung cancer comparing a population of
## individuals with a fixed average cigarette smoking per day to a
## population of individuals with a 1-unit lower average cigarette
## smoking per day is estimated to be 1.073, with a 95% confidence
## interval of (1.059, 1.087).")
```

2(b)

```
mod2b = glm(cases ~ cigpday + I(cigpday^2) + smokedur + I(smokedur^2), family = poisson(), offset = log
summary(mod2b)
```

```
##
## Call:
## glm(formula = cases ~ cigpday + I(cigpday^2) + smokedur + I(smokedur^2),
##      family = poisson(), data = dat2, offset = log(manyyears))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1705  -0.9117  -0.4424   0.6165   1.8939
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.607e+01  1.358e+00 -11.836  < 2e-16 ***
## cigpday       1.561e-01  2.799e-02   5.576  2.46e-08 ***
## I(cigpday^2)  -1.938e-03  5.565e-04  -3.483  0.000496 ***
## smokedur      2.667e-01  6.520e-02   4.090  4.32e-05 ***
## I(smokedur^2) -1.855e-03  7.735e-04  -2.398  0.016503 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 445.099  on 62  degrees of freedom
## Residual deviance:  61.369  on 58  degrees of freedom
## AIC: 191.21
##
## Number of Fisher Scoring iterations: 5
```

```
anova(mod2a, mod2b, test= "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: cases ~ cigpday
## Model 2: cases ~ cigpday + I(cigpday^2) + smokedur + I(smokedur^2)
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1         61      324.71
## 2         58       61.37  3   263.34 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This model including linear and quadratic terms of both cigpday and smokedur has an AIC score of 191.21, which is lower than the AIC score of the model in 2a) (AIC = 448.55). Conducting an anova test on the two models also gives a p-value < 0.05, indicating that including linear and quadratic terms of both cigpday and smokedur is necessary.

```
vcov(mod2b)
```

```
##              (Intercept)      cigpday I(cigpday^2)      smokedur
## (Intercept)  1.8440980737 -7.191181e-03  1.380791e-04 -8.452314e-02
## cigpday      -0.0071911809  7.836882e-04 -1.498570e-05 -9.286770e-05
## I(cigpday^2)  0.0001380791 -1.498570e-05  3.096828e-07  8.711042e-07
## smokedur     -0.0845231418 -9.286770e-05  8.711042e-07  4.250815e-03
## I(smokedur^2) 0.0009712369  1.298306e-06 -1.331609e-08 -5.000045e-05
##              I(smokedur^2)
## (Intercept)  9.712369e-04
## cigpday      1.298306e-06
## I(cigpday^2) -1.331609e-08
```



```
## smokedur      -5.000045e-05
## I(smokedur^2)  5.982694e-07

pointest1 = 1.561e-01 * 20 + (-1.938e-03) * 20^2
std1 = sqrt(20^2 * 7.836882e-04 + 400^2 * 3.096828e-07 + 2*20*400 * -1.498570e-05)
lower1 = pointest1 - 1.96* std1
upper1 = pointest1 + 1.96* std1

pointest2 = 1.561e-01 * 20 + (-1.938e-03) * (40^2 - 20^2)
std2 = sqrt(20^2 * 7.836882e-04 + 1200^2 * 3.096828e-07 + 2*20*1200 * -1.498570e-05 )
lower2 = pointest2 - 1.96*std2
upper2 = pointest2 + 1.96*std2

cat(sprintf("The estimated IRR for the effect of 20 vs. 0 cigarettes per day is %.3f, with a 95%% confi

## The estimated IRR for the effect of 20 vs. 0 cigarettes
per day is 10.452, with a 95% confidence interval of
(5.252, 20.799).

cat(sprintf("The estimated IRR for the effect of 20 vs. 0 cigarettes per day is %.3f, with a 95%% confi

## The estimated IRR for the effect of 20 vs. 0 cigarettes
per day is 2.218, with a 95% confidence interval of (1.498,
3.284).
```

## 2(c)

```
mod2c = glm(cases ~ as.factor(cigpday) + as.factor(smokedur), family = poisson(), offset = log(manyyears))
summary(mod2c)

##
## Call:
## glm(formula = cases ~ as.factor(cigpday) + as.factor(smokedur),
##      family = poisson(), data = dat2, offset = log(manyyears))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8329  -0.8560  -0.3808   0.4241   2.1762
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -12.5784     1.1475 -10.961  < 2e-16 ***
## as.factor(cigpday)5.2     1.2200     0.7073   1.725 0.084547 .
## as.factor(cigpday)11.2    2.0991     0.6363   3.299 0.000971 ***
## as.factor(cigpday)15.9    2.3089     0.6327   3.649 0.000263 ***
## as.factor(cigpday)20.4    2.9009     0.5956   4.870 1.11e-06 ***
## as.factor(cigpday)27.4    3.1162     0.5947   5.240 1.61e-07 ***
## as.factor(cigpday)40.8    3.6059     0.6048   5.962 2.49e-09 ***
## as.factor(smokedur)22.5    0.9469     1.1548   0.820 0.412202
## as.factor(smokedur)27.5    1.7016     1.0805   1.575 0.115284
## as.factor(smokedur)32.5    3.2029     1.0204   3.139 0.001695 **
## as.factor(smokedur)37.5    3.2423     1.0242   3.166 0.001547 **
## as.factor(smokedur)42.5    4.2088     1.0137   4.152 3.30e-05 ***
## as.factor(smokedur)47.5    4.4476     1.0171   4.373 1.23e-05 ***
## as.factor(smokedur)52.5    4.9048     1.0201   4.808 1.52e-06 ***
## as.factor(smokedur)57.5    5.4134     1.0239   5.287 1.24e-07 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 445.099  on 62  degrees of freedom
## Residual deviance:  51.471  on 48  degrees of freedom
## AIC: 201.31
##
## Number of Fisher Scoring iterations: 6
```

```
vcov(mod2c)
```

```
##              (Intercept) as.factor(cigpday)5.2
## (Intercept)          1.3167853          -0.327313901
## as.factor(cigpday)5.2  -0.3273139           0.500295074
## as.factor(cigpday)11.2 -0.3287141           0.333429768
## as.factor(cigpday)15.9 -0.3318027           0.333134070
## as.factor(cigpday)20.4 -0.3293749           0.333093843
## as.factor(cigpday)27.4 -0.3267085           0.333044220
## as.factor(cigpday)40.8 -0.3248529           0.333074938
## as.factor(smokedur)22.5 -0.9964739          -0.002671511
## as.factor(smokedur)27.5 -0.9950380          -0.003607696
## as.factor(smokedur)32.5 -0.9940707          -0.004575735
## as.factor(smokedur)37.5 -0.9939881          -0.004790859
## as.factor(smokedur)42.5 -0.9935631          -0.005839608
## as.factor(smokedur)47.5 -0.9936543          -0.006297780
## as.factor(smokedur)52.5 -0.9957556          -0.006335382
## as.factor(smokedur)57.5 -0.9986858          -0.008321486
##              as.factor(cigpday)11.2 as.factor(cigpday)15.9
## (Intercept)          -0.328714062          -0.3318026890
## as.factor(cigpday)5.2   0.333429768           0.3331340697
## as.factor(cigpday)11.2  0.404881358           0.3333860176
## as.factor(cigpday)15.9  0.333386018           0.4003402066
## as.factor(cigpday)20.4  0.333423838           0.3337683174
## as.factor(cigpday)27.4  0.333473103           0.3339247929
## as.factor(cigpday)40.8  0.333509597           0.3339407827
## as.factor(smokedur)22.5 -0.002344217          -0.0022739312
## as.factor(smokedur)27.5 -0.003143169          -0.0024481355
## as.factor(smokedur)32.5 -0.004043162          -0.0027398625
## as.factor(smokedur)37.5 -0.004294252          -0.0024180218
## as.factor(smokedur)42.5 -0.005193574          -0.0029806167
## as.factor(smokedur)47.5 -0.005742194          -0.0029265341
## as.factor(smokedur)52.5 -0.005200470          -0.0008688943
## as.factor(smokedur)57.5 -0.004227906           0.0013161827
##              as.factor(cigpday)20.4 as.factor(cigpday)27.4
## (Intercept)          -0.3293748556          -0.326708460
## as.factor(cigpday)5.2   0.3330938426           0.333044220
## as.factor(cigpday)11.2  0.3334238380           0.333473103
## as.factor(cigpday)15.9  0.3337683174           0.333924793
## as.factor(cigpday)20.4  0.3547427890           0.334121220
## as.factor(cigpday)27.4  0.3341212204           0.353671634
## as.factor(cigpday)40.8  0.3341526757           0.334492127
## as.factor(smokedur)22.5 -0.0036373134          -0.004566856
## as.factor(smokedur)27.5 -0.0048416734          -0.007066308
```

```

## as.factor(smokedur)32.5      -0.0055228778      -0.008693571
## as.factor(smokedur)37.5      -0.0054722841      -0.009016441
## as.factor(smokedur)42.5      -0.0056788150      -0.009435728
## as.factor(smokedur)47.5      -0.0057449251      -0.008825450
## as.factor(smokedur)52.5      -0.0036151997      -0.005817593
## as.factor(smokedur)57.5      -0.0001268521      -0.001727122
##                               as.factor(cigpday)40.8 as.factor(smokedur)22.5
## (Intercept)                  -0.324852948      -0.996473950
## as.factor(cigpday)5.2         0.333074938      -0.002671511
## as.factor(cigpday)11.2        0.333509597      -0.002344217
## as.factor(cigpday)15.9        0.333940783      -0.002273931
## as.factor(cigpday)20.4        0.334152676      -0.003637313
## as.factor(cigpday)27.4        0.334492127      -0.004566856
## as.factor(cigpday)40.8        0.365806495      -0.005230410
## as.factor(smokedur)22.5      -0.005230410      1.333508356
## as.factor(smokedur)27.5      -0.008471219      1.000354094
## as.factor(smokedur)32.5      -0.010476839      1.000452705
## as.factor(smokedur)37.5      -0.010940428      1.000484790
## as.factor(smokedur)42.5      -0.011404135      1.000476110
## as.factor(smokedur)47.5      -0.011002090      1.000424080
## as.factor(smokedur)52.5      -0.007820748      1.000274820
## as.factor(smokedur)57.5      -0.003490186      1.000055663
##                               as.factor(smokedur)27.5 as.factor(smokedur)32.5
## (Intercept)                  -0.995038027      -0.994070740
## as.factor(cigpday)5.2         -0.003607696      -0.004575735
## as.factor(cigpday)11.2        -0.003143169      -0.004043162
## as.factor(cigpday)15.9        -0.002448136      -0.002739862
## as.factor(cigpday)20.4        -0.004841673      -0.005522878
## as.factor(cigpday)27.4        -0.007066308      -0.008693571
## as.factor(cigpday)40.8        -0.008471219      -0.010476839
## as.factor(smokedur)22.5       1.000354094      1.000452705
## as.factor(smokedur)27.5       1.167374000      1.000906236
## as.factor(smokedur)32.5       1.000906236      1.041165949
## as.factor(smokedur)37.5       1.000974925      1.001257829
## as.factor(smokedur)42.5       1.000963414      1.001249577
## as.factor(smokedur)47.5       1.000861921      1.001120856
## as.factor(smokedur)52.5       1.000589273      1.000784666
## as.factor(smokedur)57.5       1.000198212      1.000312038
##                               as.factor(smokedur)37.5 as.factor(smokedur)42.5
## (Intercept)                  -0.993988106      -0.993563147
## as.factor(cigpday)5.2         -0.004790859      -0.005839608
## as.factor(cigpday)11.2        -0.004294252      -0.005193574
## as.factor(cigpday)15.9        -0.002418022      -0.002980617
## as.factor(cigpday)20.4        -0.005472284      -0.005678815
## as.factor(cigpday)27.4        -0.009016441      -0.009435728
## as.factor(cigpday)40.8        -0.010940428      -0.011404135
## as.factor(smokedur)22.5       1.000484790      1.000476110
## as.factor(smokedur)27.5       1.000974925      1.000963414
## as.factor(smokedur)32.5       1.001257829      1.001249577
## as.factor(smokedur)37.5       1.048978832      1.001354103
## as.factor(smokedur)42.5       1.001354103      1.027672779
## as.factor(smokedur)47.5       1.001217374      1.001224649
## as.factor(smokedur)52.5       1.000865489      1.000887732
## as.factor(smokedur)57.5       1.000373549      1.000429294

```

```
##               as.factor(smokedur)47.5 as.factor(smokedur)52.5
## (Intercept)                -0.993654319      -0.9957555854
## as.factor(cigpday)5.2        -0.006297780      -0.0063353821
## as.factor(cigpday)11.2       -0.005742194      -0.0052004702
## as.factor(cigpday)15.9       -0.002926534      -0.0008688943
## as.factor(cigpday)20.4       -0.005744925      -0.0036151997
## as.factor(cigpday)27.4       -0.008825450      -0.0058175927
## as.factor(cigpday)40.8       -0.011002090      -0.0078207482
## as.factor(smokedur)22.5        1.000424080        1.0002748203
## as.factor(smokedur)27.5        1.000861921        1.0005892728
## as.factor(smokedur)32.5        1.001120856        1.0007846662
## as.factor(smokedur)37.5        1.001217374        1.0008654889
## as.factor(smokedur)42.5        1.001224649        1.0008877318
## as.factor(smokedur)47.5        1.034449542        1.0008307848
## as.factor(smokedur)52.5        1.000830785        1.0406869361
## as.factor(smokedur)57.5        1.000441261        1.0005076245
##               as.factor(smokedur)57.5
## (Intercept)                -0.9986858401
## as.factor(cigpday)5.2        -0.0083214860
## as.factor(cigpday)11.2       -0.0042279062
## as.factor(cigpday)15.9        0.0013161827
## as.factor(cigpday)20.4       -0.0001268521
## as.factor(cigpday)27.4       -0.0017271215
## as.factor(cigpday)40.8       -0.0034901864
## as.factor(smokedur)22.5        1.0000556634
## as.factor(smokedur)27.5        1.0001982120
## as.factor(smokedur)32.5        1.0003120382
## as.factor(smokedur)37.5        1.0003735490
## as.factor(smokedur)42.5        1.0004292944
## as.factor(smokedur)47.5        1.0004412606
## as.factor(smokedur)52.5        1.0005076245
## as.factor(smokedur)57.5        1.0483021062
```

```
pointest3 = 2.9009
lower3 = 2.9009 -1.96* 0.5956
upper3 = 2.9009 +1.96* 0.5956
```

```
pointest4 = 3.6059 - 2.9009
std4 = sqrt(0.365806495 + 0.3547427890 - 2*0.3341526757)
lower4 = pointest4 - 1.96*std4
upper4 = pointest4 + 1.96*std4
```

```
cat(sprintf("The estimated IRR for the effect of 20.4 vs. 0 cigarettes per day is %.3f, with a 95%% confidence interval of (%.3f, %.3f).", 18.191, 5.661, 58.456))
```

```
## The estimated IRR for the effect of 20.4 vs. 0
cigarettes per day is 18.191, with a 95% confidence
interval of (5.661, 58.456).
```

```
cat(sprintf("The estimated IRR for the effect of 40.8 vs. 20.4 cigarettes per day is %.3f, with a 95%% confidence interval of (%.3f, %.3f).", 2.024, 1.293, 3.168))
```

```
## The estimated IRR for the effect of 40.8 vs. 20.4
cigarettes per day is 2.024, with a 95% confidence interval
of (1.293, 3.168).
```

```
summary(mod2b)$aic
```

```
## [1] 191.2109
```

```
summary(mod2c)$aic
```

```
## [1] 201.3131
```

I would prefer the model in 2b using continuous linear and quadratic terms of both cigpday and smokedur better. Because the model in 2b gives an AIC score of 191.2109, which is lower than the AIC score of 201.3131 given by the model in 2c using cigpday and smokedur as categorical variables. Also the confidence interval of model 2c is wider than the confidence interval of model 2b, indicating more uncertainty in the estimated coefficient of cigpday.