# LDA final project

*Kaitlin Maciejewski, Morgan de Ferrante, Bingnan Li, Volha Tryputsen*

## Contents

## Exploratory

### Smoking vs. Age, Sex

### (1): Smoking ~ age, sex

#### Is there a relationship between age and smoking status?

ANS: Yes, the proportion of smokers decreases with the age.

```
smoke_v_age = data %>%
  select(cursmoke, age) %>%
  ggplot(aes(x = age, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

**Does this relationship differ by sex?**

ANS: There is a higher proportion of smoker among men compared to women as both age ,but there is no interaction between age and sex.

```
smoke_age_sex = data %>%
  select(cursmoke, age, sex) %>%
  ggplot(aes(x = age, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

## (2) number of cigarettes ~ age, sex

**Is there a relationship between the number of cigarettes smoked per day and age?**

ANS: Yes, number of sigarets smoked per day stays constant for 30-50 years old and decreases with age after 50 years old.

**All**

```
n_c_age_all = data %>%
  select(cigpday, age) %>%
  ggplot(aes(x = age, y = cigpday)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

**Smokers only**

```
n_c_age_smoke = data %>%
  select(cigpday, age) %>%
  filter(cigpday > 0) %>%
  ggplot(aes(x = age, y = cigpday)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

**Does this relationship differ by sex?**

ANS: There is sex effect (men smoke higer number of sigarets per day than women across age), but there is no sex and age interaction.

**All**

```
n_c_age_s_all = data %>%
  select(cigpday, age, sex) %>%
  ggplot(aes(x = age, y = cigpday, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

**Smokers only**

```
n_c_age_s_smoke = data %>%
  select(cigpday, age, sex) %>%
  filter(cigpday > 0) %>%
  ggplot(aes(x = age, y = cigpday, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

## Smoking vs. health outcomes

### (1) The relationship between current smoking status and systolic blood pressure.

**smoking ~ sysbp**

ANS: Proportion of smokers decreases with increase of systolic blood presure

```
smoke_sbp = data %>%
  select(cursmoke, sysbp) %>%
  ggplot(aes(x = sysbp, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

ANS: slightly higher sysbp for non-smokers

```
smoke_sysbp_status = data %>%
  select(cursmoke, sysbp) %>%
  mutate(cursmoke = factor(cursmoke)) %>%
  ggplot(aes(y = sysbp, x = cursmoke)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```

**smoking ~ sysbp, sex**

ANS: Proportion of smokers decreases with increase of systolic blood presure; the proportion is higher for men (sex effect).

```
smoke_sysbp_sex = data %>%
  select(cursmoke, sysbp, sex) %>%
  ggplot(aes(x = sysbp, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

ANS: no differences in sysbp between male and female smokers and non-smokers

```
smoke_sysbp_sex_status = data %>%
  select(cursmoke, sex, sysbp) %>%
  mutate(cursmoke = factor(cursmoke),
         smoke_sex = interaction(cursmoke, sex)) %>%
  ggplot(aes(y = sysbp, x = smoke_sex)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```

## (2) The relationship between current smoking status and diastolic blood pressure.

**smoking ~ diabp**

ANS: Proportion of smokers decreases with increase of diastolic blood presure for BP=100 ad then proportion increases again (latter could be due to not enough data)

```
smoke_dbp = data %>%
  select(cursmoke, diabp) %>%
  ggplot(aes(x = diabp, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

ANS: no difference

```
smoke_dbp_box = data %>%
  select(cursmoke, diabp) %>%
  mutate(cursmoke = factor(cursmoke)) %>%
  ggplot(aes(y = diabp, x = cursmoke)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```

**smoking ~ diabp, sex**

ANS: Proportion of smokers decreases with increase of diastolic blood presure; the proprtions are higher for men (sex effect).

```
smoke_dbp_s = data %>%
  select(cursmoke, diabp, sex) %>%
  ggplot(aes(x = diabp, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

ANS: no difference

```
smoke_dbp_s_bp = data %>%
  select(cursmoke, sex, diabp) %>%
  mutate(cursmoke = factor(cursmoke),
         smoke_sex = interaction(cursmoke, sex)) %>%
  ggplot(aes(y = diabp, x = smoke_sex)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```

## (3) The relationship between current smoking status and serum total cholesterol.

**smoking ~ totchol**

ANS: Proportion of smokers slightly decreases with increase of total cholesterol values

```
smoke_tc = data %>%
  select(cursmoke, totchol) %>%
  ggplot(aes(x = totchol, y = cursmoke)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```

ANS: no difference

```
smoke_tc_bp = data %>%
  select(cursmoke, totchol) %>%
  mutate(cursmoke = factor(cursmoke)) %>%
  ggplot(aes(y = totchol, x = cursmoke)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```
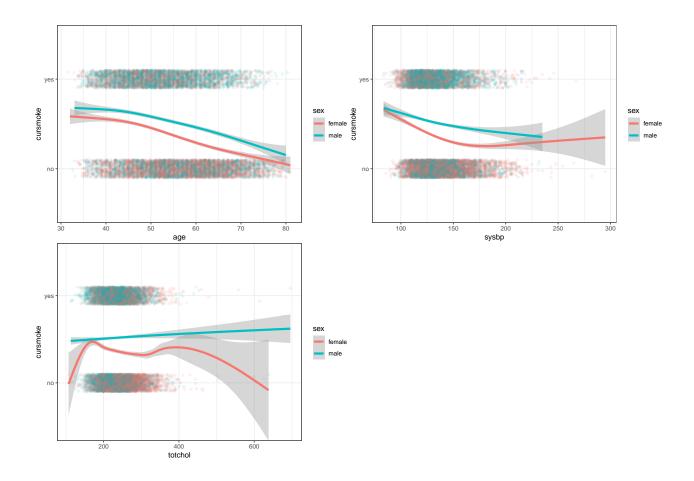
**smoking ~ totchol, sex [!!!]**

ANS: Proportion of smokers hasnonlinera relationship with total cholesterol for women; proprtions increases
with increase in total cholesterol for men (sex by totchol interaction effect).

```
smoke_tc_sex = data %>%
  select(cursmoke, totchol, sex) %>%
  ggplot(aes(x = totchol, y = cursmoke, group = sex, color = sex)) +
  geom_jitter(height = 0.1, alpha = 0.1) +
  geom_smooth(lwd = 1.5) +
  theme_bw()
```
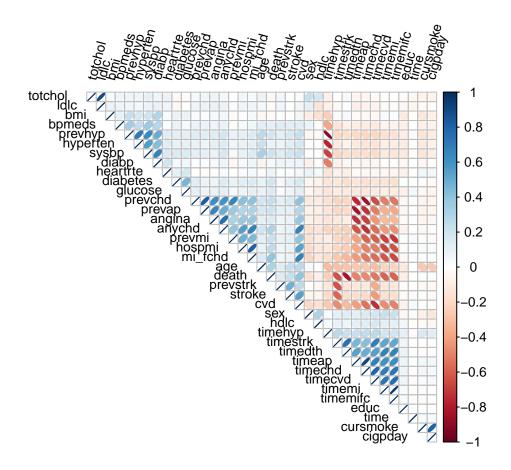
ANS: no difference

```
smoke_tc_sex_bp = data %>%
  select(cursmoke, sex, totchol) %>%
  mutate(cursmoke = factor(cursmoke),
         smoke_sex = interaction(cursmoke, sex)) %>%
  ggplot(aes(y = totchol, x = smoke_sex)) +
  geom_boxplot(outlier.colour = "white") +
  theme_bw()
```

```
smoke_age_sex
smoke_sysbp_sex
smoke_tc_sex
```

```
## Warning: Removed 409 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 409 rows containing missing values (geom_point).
```

## Cor plot

```r
library(dplyr)

corr <- data[,c(-1,-21)] %>% cor(., use = "complete.obs")

library(corrplot)

corrplot(corr, type = "upper", order = "hclust",
         tl.col = "black", tl.srt = 75,
         tl.offset = 1, tl.cex= .8, method = "ellipse")
```

## Missingness

```r
prop <- round(colSums(is.na(data))/dim(data)[1], 3)

knitr::kable(sort(prop, decreasing = TRUE)[1:9], col.names = "Proportion of NAs")
```

|         | Proportion of NAs |
|---------|-------------------|
| hdlc    | 0.740             |
| ldlc    | 0.740             |
| glucose | 0.124             |
| bpmeds  | 0.051             |
| totchol | 0.035             |
| educ    | 0.025             |
| cigpday | 0.007             |
| bmi     | 0.004             |
| heartrte| 0.001             |

```r
prob.data <- data %>%
  group_by(period) %>%
  summarise(sysbp_prob = sum(sysbp, na.rm = TRUE)/n())
prob.data

## # A tibble: 3 x 2
##    period sysbp_prob
```

```
##     <int>        <dbl>
## 1       1         133.
## 2       2         137.
## 3       3         140.
```

```
table(data$period)
```

```
##
##    1    2    3
## 4434 3930 3263
```

# Models

```
library(lme4)
library(dplyr)
```

**(1) Is there a relationship between age and smoking status? Does this relationship differ by sex?**

```
my.data <- read.csv("../final_data/frmgham2.csv")
library(gee)

model.q1 <- gee(CURSMOKE ~ AGE + as.factor(SEX) + as.factor(educ)
              + BMI + DIABETES + HEARTRTE + PREVCHD + PREVSTRK
              + PREVHYP + TIMEDTH,
              id = RANDID,
              data = my.data,
              family=binomial,
              corstr = "unstructured")
```

```
##      (Intercept)              AGE  as.factor(SEX)2 as.factor(educ)2
##     5.2810549757    -0.0570318456    -0.7246241605      0.0503157656
## as.factor(educ)3 as.factor(educ)4              BMI         DIABETES
##    -0.2263406725    -0.2104931713    -0.0940121564    -0.1114780579
##          HEARTRTE          PREVCHD          PREVSTRK          PREVHYP
##     0.0173911758    -0.0569530253    -0.2633153119    -0.2252617826
##          TIMEDTH
##    -0.0001021991
```

```
knitr::kable(summary(model.q1)$coefficients[,c(1,4,5)], digits = 3)
```

|                  | Estimate | Robust S.E. | Robust z |
|------------------|----------|-------------|----------|
| (Intercept)      | 5.149    | 0.278       | 18.543   |
| AGE              | -0.051   | 0.002       | -21.106  |
| as.factor(SEX)2  | -0.716   | 0.060       | -11.892  |
| as.factor(educ)2 | 0.083    | 0.071       | 1.168    |
| as.factor(educ)3 | -0.182   | 0.086       | -2.116   |
| as.factor(educ)4 | -0.212   | 0.097       | -2.185   |
| BMI              | -0.084   | 0.007       | -12.152  |
| DIABETES         | -0.060   | 0.102       | -0.595   |
| HEARTRTE         | 0.009    | 0.001       | 6.031    |

|  | Estimate | Robust S.E. | Robust z |
|---|---|---|---|
| PREVCHD | -0.259 | 0.090 | -2.872 |
| PREVSTRK | -0.189 | 0.176 | -1.073 |
| PREVHYP | -0.049 | 0.041 | -1.187 |
| TIMEDTH | 0.000 | 0.000 | -6.341 |

```
model.q1[["working.correlation"]]
```

```
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.7563191 0.5136670
## [2,] 0.7563191 1.0000000 0.5713635
## [3,] 0.5136670 0.5713635 1.0000000
```

```
QIC(model.q1)
```

```
##      QIC
## 13931.71
```

```
model.q1 <- gee(CURSMOKE ~ AGE + as.factor(SEX) + as.factor(educ)
                + BMI + DIABETES + HEARTRTE + PREVCHD + PREVSTRK
                + PREVHYP + TIMEDTH,
                id = RANDID,
                data = my.data,
                family=binomial,
                corstr = "exchangeable")
```

```
##      (Intercept)              AGE  as.factor(SEX)2 as.factor(educ)2
##     5.2810549757    -0.0570318456    -0.7246241605     0.0503157656
## as.factor(educ)3 as.factor(educ)4              BMI         DIABETES
##    -0.2263406725    -0.2104931713    -0.0940121564    -0.1114780579
##         HEARTRTE          PREVCHD         PREVSTRK          PREVHYP
##     0.0173911758    -0.0569530253    -0.2633153119    -0.2252617826
##          TIMEDTH
##    -0.0001021991
```

```
knitr::kable(summary(model.q1)$coefficients[,c(1,4,5)], digits = 3)
```

|  | Estimate | Robust S.E. | Robust z |
|---|---|---|---|
| (Intercept) | 5.185 | 0.278 | 18.645 |
| AGE | -0.051 | 0.002 | -21.542 |
| as.factor(SEX)2 | -0.741 | 0.061 | -12.229 |
| as.factor(educ)2 | 0.077 | 0.071 | 1.073 |
| as.factor(educ)3 | -0.178 | 0.087 | -2.061 |
| as.factor(educ)4 | -0.214 | 0.097 | -2.195 |
| BMI | -0.084 | 0.007 | -12.123 |
| DIABETES | -0.103 | 0.100 | -1.023 |
| HEARTRTE | 0.009 | 0.001 | 6.023 |
| PREVCHD | -0.276 | 0.088 | -3.156 |
| PREVSTRK | -0.166 | 0.176 | -0.943 |
| PREVHYP | -0.054 | 0.041 | -1.318 |
| TIMEDTH | 0.000 | 0.000 | -6.247 |

```
model.q1[["working.correlation"]]
```

```
##            [,1]      [,2]      [,3]
## [1,] 1.0000000 0.7094644 0.7094644
## [2,] 0.7094644 1.0000000 0.7094644
## [3,] 0.7094644 0.7094644 1.0000000
```

```
QIC(model.q1)
```

```
##      QIC
## 13931.65
```

## (2) Is there a relationship between the number of cigarettes smoked per day and age? Does this relationship differ by sex?

```
model.q2 <- lmer(CIGPDAY ~ AGE + as.factor(SEX) +
                 as.factor(educ) + (1|RANDID),
               data = my.data)
```

If we think cig per day as count data, it follows poisson distribution. Then we can fit GEE model as well:

```
model.q2_1 <- gee(CIGPDAY ~ AGE + as.factor(SEX) + as.factor(educ)
               + BMI + DIABETES + HEARTRTE + PREVCHD + PREVSTRK
               + PREVHYP + TIMEDTH,
               data = my.data,
               id = RANDID,
               family=poisson,
               corstr = "unstructured")
```

```
##      (Intercept)              AGE  as.factor(SEX)2 as.factor(educ)2
##     4.914149e+00    -3.395990e-02    -7.680017e-01     5.765596e-02
## as.factor(educ)3 as.factor(educ)4              BMI         DIABETES
##    -1.227240e-01    -8.833837e-02    -4.493994e-02    -1.608903e-01
##         HEARTRTE          PREVCHD         PREVSTRK          PREVHYP
##     1.281640e-02    -9.883049e-02    -1.115583e-01    -1.079832e-01
##          TIMEDTH
##    -5.152275e-05
```

```
knitr::kable(summary(model.q2_1)$coefficients[,c(1,4,5)], digits = 3)
```

|                  | Estimate | Robust S.E. | Robust z |
|------------------|----------|-------------|----------|
| (Intercept)      | 4.523    | 0.170       | 26.626   |
| AGE              | -0.024   | 0.001       | -15.826  |
| as.factor(SEX)2  | -0.795   | 0.039       | -20.223  |
| as.factor(educ)2 | 0.096    | 0.044       | 2.172    |
| as.factor(educ)3 | -0.072   | 0.057       | -1.245   |
| as.factor(educ)4 | -0.067   | 0.063       | -1.059   |
| BMI              | -0.039   | 0.005       | -8.032   |
| DIABETES         | -0.110   | 0.072       | -1.536   |
| HEARTRTE         | 0.007    | 0.001       | 8.453    |
| PREVCHD          | -0.241   | 0.066       | -3.635   |
| PREVSTRK         | -0.130   | 0.110       | -1.181   |
| PREVHYP          | -0.017   | 0.028       | -0.587   |
| TIMEDTH          | 0.000    | 0.000       | -5.505   |

```
model.q2_1[["working.correlation"]]
```

```
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.7425318 0.5006078
## [2,] 0.7425318 1.0000000 0.6615306
## [3,] 0.5006078 0.6615306 1.0000000
```

```
QIC(model.q2_1)
```

```
##       QIC
## -236402.4
```

```
model.q2_1 <- gee(CIGPDAY ~ AGE + as.factor(SEX) + as.factor(educ)
                  + BMI + DIABETES + HEARTRTE + PREVCHD + PREVSTRK
                  + PREVHYP + TIMEDTH,
                   data = my.data,
                   id = RANDID,
                   family=poisson,
                   corstr = "exchangeable")
```

```
##     (Intercept)               AGE  as.factor(SEX)2 as.factor(educ)2
##    4.914149e+00     -3.395990e-02     -7.680017e-01     5.765596e-02
## as.factor(educ)3 as.factor(educ)4               BMI         DIABETES
##   -1.227240e-01     -8.833837e-02     -4.493994e-02    -1.608903e-01
##        HEARTRTE          PREVCHD          PREVSTRK          PREVHYP
##    1.281640e-02     -9.883049e-02     -1.115583e-01    -1.079832e-01
##         TIMEDTH
##    -5.152275e-05
```

```
knitr::kable(summary(model.q2_1)$coefficients[,c(1,4,5)], digits = 3)
```

|                  | Estimate | Robust S.E. | Robust z |
|------------------|----------|-------------|----------|
| (Intercept)      | 4.403    | 0.171       | 25.730   |
| AGE              | -0.021   | 0.001       | -14.173  |
| as.factor(SEX)2  | -0.833   | 0.039       | -21.366  |
| as.factor(educ)2 | 0.098    | 0.044       | 2.243    |
| as.factor(educ)3 | -0.070   | 0.057       | -1.230   |
| as.factor(educ)4 | -0.063   | 0.063       | -1.011   |
| BMI              | -0.039   | 0.005       | -8.108   |
| DIABETES         | -0.121   | 0.070       | -1.734   |
| HEARTRTE         | 0.007    | 0.001       | 8.366    |
| PREVCHD          | -0.268   | 0.066       | -4.041   |
| PREVSTRK         | -0.144   | 0.106       | -1.356   |
| PREVHYP          | -0.020   | 0.028       | -0.709   |
| TIMEDTH          | 0.000    | 0.000       | -4.943   |

```
model.q2_1[["working.correlation"]]
```

```
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.7347634 0.7347634
## [2,] 0.7347634 1.0000000 0.7347634
## [3,] 0.7347634 0.7347634 1.0000000
```

```
QIC(model.q2_1)
```

```
##        QIC
## -236402.8
```

Using mixed effect model using cig per day instead of smoking status:

```r
#saturated model
model.saturated <- lmer(CIGPDAY ~ as.factor(SEX) + AGE
                        + BPMEDS + as.factor(educ)
                        + TOTCHOL + BMI + GLUCOSE + DIABETES + HEARTRTE + PREVAP
                        + PREVCHD + PREVMI + PREVSTRK +STROKE+ PREVHYP + (1|RANDID),
                        na.action = 'na.omit',
                        data = my.data)
summary(model.saturated)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## CIGPDAY ~ as.factor(SEX) + AGE + BPMEDS + as.factor(educ) + TOTCHOL +
##     BMI + GLUCOSE + DIABETES + HEARTRTE + PREVAP + PREVCHD +
##     PREVMI + PREVSTRK + STROKE + PREVHYP + (1 | RANDID)
##    Data: my.data
##
## REML criterion at convergence: 67692.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4444 -0.3150 -0.1078  0.2067  6.2548
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  RANDID   (Intercept) 95.78    9.787
##  Residual             35.77    5.981
## Number of obs: 9310, groups:  RANDID, 4213
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)      23.237579   1.393589  16.675
## as.factor(SEX)2  -7.062484   0.339110 -20.827
## AGE              -0.178024   0.013735 -12.962
## BPMEDS            0.147068   0.340662   0.432
## as.factor(educ)2  0.602137   0.400578   1.503
## as.factor(educ)3 -0.984441   0.482907  -2.039
## as.factor(educ)4 -1.015103   0.548951  -1.849
## TOTCHOL           0.011407   0.002557   4.461
## BMI              -0.313257   0.035227  -8.893
## GLUCOSE          -0.010167   0.004203  -2.419
## DIABETES         -0.254321   0.570383  -0.446
## HEARTRTE          0.070622   0.008232   8.579
## PREVAP           -3.058426   0.985947  -3.102
## PREVCHD           0.948740   1.050262   0.903
## PREVMI           -2.594434   0.894896  -2.899
## PREVSTRK         -1.029740   0.917968  -1.122
## STROKE            0.943808   0.586941   1.608
## PREVHYP          -0.205972   0.240530  -0.856
```

```r
#using variables that selected
model.mixed2 <- lmer(CIGPDAY~ AGE + as.factor(SEX) + SYSBP
```

```
                        + DIABP + TOTCHOL + as.factor(educ)
                        + (1|RANDID),
                        data = my.data,
                        na.action = "na.omit")
summary(model.mixed2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## CIGPDAY ~ AGE + as.factor(SEX) + SYSBP + DIABP + TOTCHOL + as.factor(educ) +
##     (1 | RANDID)
##    Data: my.data
##
## REML criterion at convergence: 78607.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.6343 -0.3025 -0.1015  0.1642  6.6280
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  RANDID   (Intercept) 97.30    9.864
##  Residual             36.52    6.043
## Number of obs: 10868, groups:  RANDID, 4306
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)       21.750845   1.049612  20.723
## AGE               -0.199583   0.012148 -16.429
## as.factor(SEX)2   -6.448164   0.332812 -19.375
## SYSBP             -0.001364   0.006670  -0.205
## DIABP             -0.018833   0.011514  -1.636
## TOTCHOL            0.011959   0.002373   5.039
## as.factor(educ)2   1.055542   0.392720   2.688
## as.factor(educ)3  -0.625323   0.474072  -1.319
## as.factor(educ)4  -0.827192   0.539360  -1.534
##
## Correlation of Fixed Effects:
##             (Intr) AGE    a.(SEX SYSBP  DIABP  TOTCHO as.()2 as.()3
## AGE         -0.490
## as.fc(SEX)2 -0.144  0.019
## SYSBP        0.023 -0.460 -0.064
## DIABP       -0.452  0.314  0.076 -0.679
## TOTCHOL     -0.368 -0.095 -0.094  0.047 -0.154
## as.fctr(d)2 -0.225  0.110 -0.055  0.008  0.001  0.002
## as.fctr(d)3 -0.161  0.051 -0.092  0.028 -0.004 -0.007  0.351
## as.fctr(d)4 -0.175  0.044  0.074  0.032 -0.005 -0.002  0.301  0.244
```

## (3) Totchol and cursmoke

```
#############################################
# Models After Removing Non Significant Terms #
#############################################
```

```
totchol_fit <- gee(totchol ~ cursmoke + age + factor(sex) + bmi +
                   diabetes + heartrte  + prevhyp ,
                id = randid,
                family = "gaussian",
                corstr = "unstructured",
                na.action = "na.omit")
```

```
##  (Intercept)      cursmoke           age factor(sex)2          bmi
##  162.1032803     2.3672102     0.6268036   12.3710689    0.7814908
##     diabetes      heartrte       prevhyp
##   -5.8194286     0.1891382     5.2800067
```

```
knitr::kable(summary(totchol_fit)$coefficients[,c(1,4,5)], digits = 3)
```

|              | Estimate | Robust S.E. | Robust z |
|--------------|----------|-------------|----------|
| (Intercept)  | 149.257  | 5.279       | 28.274   |
| cursmoke     | 3.430    | 1.009       | 3.399    |
| age          | 0.648    | 0.053       | 12.309   |
| factor(sex)2 | 13.474   | 1.219       | 11.050   |
| bmi          | 1.404    | 0.144       | 9.748    |
| diabetes     | -6.628   | 2.638       | -2.513   |
| heartrte     | 0.117    | 0.032       | 3.626    |
| prevhyp      | 3.451    | 0.903       | 3.820    |

```
knitr::kable(round(2 * pnorm(abs(coef(summary(totchol_fit))[,5]), lower.tail = FALSE), 3))
```

|              | x     |
|--------------|-------|
| (Intercept)  | 0.000 |
| cursmoke     | 0.001 |
| age          | 0.000 |
| factor(sex)2 | 0.000 |
| bmi          | 0.000 |
| diabetes     | 0.012 |
| heartrte     | 0.000 |
| prevhyp      | 0.000 |

```
QIC(totchol_fit)
```

```
##       QIC
## 84588.89
```

```
totchol_fit2 <- gee(totchol ~ cursmoke + age + factor(sex) + bmi +
                   diabetes + heartrte  + prevhyp ,
                id = randid,
                family = "gaussian",
                corstr = "exchangeable",
                na.action = "na.omit")
```

```
##  (Intercept)      cursmoke           age factor(sex)2          bmi
##  162.1032803     2.3672102     0.6268036   12.3710689    0.7814908
##     diabetes      heartrte       prevhyp
##   -5.8194286     0.1891382     5.2800067
```

```r
knitr::kable(summary(totchol_fit2)$coefficients[,c(1,4,5)], digits = 3)
```

|             | Estimate | Robust S.E. | Robust z |
|-------------|---------:|------------:|---------:|
| (Intercept) | 157.548  | 5.162       | 30.518   |
| cursmoke    | 3.824    | 1.035       | 3.696    |
| age         | 0.402    | 0.051       | 7.824    |
| factor(sex)2| 13.291   | 1.222       | 10.874   |
| bmi         | 1.656    | 0.143       | 11.578   |
| diabetes    | -7.648   | 2.322       | -3.293   |
| heartrte    | 0.118    | 0.033       | 3.607    |
| prevhyp     | 2.423    | 0.894       | 2.710    |

```r
knitr::kable(round(2 * pnorm(abs(coef(summary(totchol_fit2))[,5]), lower.tail = FALSE), 3))
```

|             | x     |
|-------------|------:|
| (Intercept) | 0.000 |
| cursmoke    | 0.000 |
| age         | 0.000 |
| factor(sex)2| 0.000 |
| bmi         | 0.000 |
| diabetes    | 0.001 |
| heartrte    | 0.000 |
| prevhyp     | 0.007 |

```r
QIC(totchol_fit2)
```

```
##      QIC
## 84588.27
```

## (4) Sysbp and cursmoke

```r
sysbp_fit <- gee(sysbp ~ cursmoke + age + factor(sex)  + bmi +
                  diabetes + heartrte + prevchd + prevstrk + death,
               id = randid,
               family = "gaussian",
               corstr = "unstructured",
               na.action = "na.omit")
```

```
## (Intercept)       cursmoke            age factor(sex)2          bmi
##   40.1191453     -1.1246140      0.7236371    2.9078810    1.3070103
##     diabetes       heartrte        prevchd     prevstrk        death
##    4.8530466      0.2464253      2.4405511    8.6116353    7.0492511
```

```r
knitr::kable(summary(sysbp_fit)$coefficients[,c(1,4,5)], digits = 3)
```

|             | Estimate | Robust S.E. | Robust z |
|-------------|---------:|------------:|---------:|
| (Intercept) | 42.702   | 2.163       | 19.740   |
| cursmoke    | -0.702   | 0.433       | -1.619   |
| age         | 0.741    | 0.022       | 33.618   |

|  | Estimate | Robust S.E. | Robust z |
|---|---|---|---|
| factor(sex)2 | 3.272 | 0.542 | 6.042 |
| bmi | 1.373 | 0.062 | 22.301 |
| diabetes | 3.664 | 1.127 | 3.250 |
| heartrte | 0.175 | 0.016 | 11.231 |
| prevchd | 1.670 | 0.876 | 1.906 |
| prevstrk | 4.894 | 2.004 | 2.443 |
| death | 7.650 | 0.639 | 11.967 |

```r
knitr::kable(round(2 * pnorm(abs(coef(summary(sysbp_fit))[,5]), lower.tail = FALSE), 3))
```

|  | x |
|---|---|
| (Intercept) | 0.000 |
| cursmoke | 0.105 |
| age | 0.000 |
| factor(sex)2 | 0.000 |
| bmi | 0.000 |
| diabetes | 0.001 |
| heartrte | 0.000 |
| prevchd | 0.057 |
| prevstrk | 0.015 |
| death | 0.000 |

```r
QIC(sysbp_fit)
```

```
##      QIC
## 68837.26
```

```r
sysbp_fit2 <- gee(sysbp ~ cursmoke + age + factor(sex)  + bmi +
                  diabetes + heartrte + prevchd + prevstrk + death,
                id = randid,
                family = "gaussian",
                corstr = "exchangeable",
                na.action = "na.omit")
```

```
##  (Intercept)      cursmoke          age factor(sex)2          bmi
##   40.1191453    -1.1246140    0.7236371    2.9078810    1.3070103
##     diabetes      heartrte      prevchd     prevstrk        death
##    4.8530466     0.2464253    2.4405511    8.6116353    7.0492511
```

```r
knitr::kable(summary(sysbp_fit2)$coefficients[,c(1,4,5)], digits = 3)
```

|  | Estimate | Robust S.E. | Robust z |
|---|---|---|---|
| (Intercept) | 43.000 | 2.176 | 19.758 |
| cursmoke | -0.772 | 0.431 | -1.789 |
| age | 0.741 | 0.021 | 34.477 |
| factor(sex)2 | 3.321 | 0.545 | 6.093 |
| bmi | 1.370 | 0.063 | 21.783 |
| diabetes | 3.173 | 1.114 | 2.848 |
| heartrte | 0.172 | 0.016 | 10.967 |
| prevchd | 1.553 | 0.862 | 1.800 |
| prevstrk | 4.201 | 2.094 | 2.006 |

|         | Estimate | Robust S.E. | Robust z |
|---------|----------|-------------|----------|
| death   | 7.794    | 0.644       | 12.109   |

```r
knitr::kable(round(2 * pnorm(abs(coef(summary(sysbp_fit2))[,5]), lower.tail = FALSE), 3))
```

|               | x     |
|---------------|-------|
| (Intercept)   | 0.000 |
| cursmoke      | 0.074 |
| age           | 0.000 |
| factor(sex)2  | 0.000 |
| bmi           | 0.000 |
| diabetes      | 0.004 |
| heartrte      | 0.000 |
| prevchd       | 0.072 |
| prevstrk      | 0.045 |
| death         | 0.000 |

```r
QIC(sysbp_fit2)
```

```
##      QIC
## 68837.55
```

## (5) Diabp and cursmoke

```r
diabp_fit <- gee(diabp ~ cursmoke + factor(sex) + factor(educ) + bmi +
                   diabetes + heartrte  + prevstrk +death,
                 id = randid,
                 family = "gaussian",
                 corstr = "unstructured",
                 na.action = "na.omit")
```

```
##   (Intercept)      cursmoke  factor(sex)2 factor(educ)2 factor(educ)3
##    47.3480563    -1.0968017    -0.8594048     0.7880155     0.6921165
## factor(educ)4           bmi      diabetes      heartrte      prevstrk
##     0.5395935     0.9078065    -1.6162433     0.1550842     4.3067909
##         death
##     2.8662538
```

```r
knitr::kable(summary(diabp_fit)$coefficients[,c(1,4,5)], digits = 3)
```

|               | Estimate | Robust S.E. | Robust z |
|---------------|----------|-------------|----------|
| (Intercept)   | 48.560   | 1.142       | 42.511   |
| cursmoke      | -0.607   | 0.243       | -2.501   |
| factor(sex)2  | -0.591   | 0.295       | -1.999   |
| factor(educ)2 | 0.864    | 0.344       | 2.513    |
| factor(educ)3 | 0.769    | 0.400       | 1.924    |
| factor(educ)4 | 0.470    | 0.450       | 1.043    |
| bmi           | 0.971    | 0.035       | 28.015   |
| diabetes      | -1.731   | 0.598       | -2.895   |
| heartrte      | 0.112    | 0.009       | 12.846   |

|  | Estimate | Robust S.E. | Robust z |
|---|---|---|---|
| prevstrk | 1.842 | 1.207 | 1.526 |
| death | 3.219 | 0.336 | 9.582 |

```
knitr::kable(round(2 * pnorm(abs(coef(summary(diabp_fit))[,5]), lower.tail = FALSE), 3))
```

|  | x |
|---|---|
| (Intercept) | 0.000 |
| cursmoke | 0.012 |
| factor(sex)2 | 0.046 |
| factor(educ)2 | 0.012 |
| factor(educ)3 | 0.054 |
| factor(educ)4 | 0.297 |
| bmi | 0.000 |
| diabetes | 0.004 |
| heartrte | 0.000 |
| prevstrk | 0.127 |
| death | 0.000 |

```
QIC(diabp_fit)
```

```
##      QIC
## 53507.67
```

```
diabp_fit2 <- gee(diabp ~ cursmoke + factor(sex) + factor(educ) + bmi +
                  diabetes + heartrte  + prevstrk +death,
               id = randid,
               family = "gaussian",
               corstr = "exchangeable",
               na.action = "na.omit")
```

```
##   (Intercept)       cursmoke  factor(sex)2 factor(educ)2 factor(educ)3
##    47.3480563     -1.0968017    -0.8594048     0.7880155     0.6921165
## factor(educ)4           bmi      diabetes      heartrte      prevstrk
##     0.5395935      0.9078065    -1.6162433     0.1550842     4.3067909
##         death
##     2.8662538
```

```
knitr::kable(summary(diabp_fit2)$coefficients[,c(1,4,5)], digits = 3)
```

|  | Estimate | Robust S.E. | Robust z |
|---|---|---|---|
| (Intercept) | 48.526 | 1.156 | 41.980 |
| cursmoke | -0.525 | 0.242 | -2.166 |
| factor(sex)2 | -0.518 | 0.298 | -1.738 |
| factor(educ)2 | 0.857 | 0.346 | 2.476 |
| factor(educ)3 | 0.776 | 0.403 | 1.927 |
| factor(educ)4 | 0.531 | 0.453 | 1.171 |
| bmi | 0.991 | 0.035 | 28.079 |
| diabetes | -2.072 | 0.582 | -3.564 |
| heartrte | 0.106 | 0.009 | 12.102 |
| prevstrk | 1.049 | 1.215 | 0.863 |
| death | 3.317 | 0.338 | 9.800 |

```
knitr::kable(round(2 * pnorm(abs(coef(summary(diabp_fit2))[,5]), lower.tail = FALSE), 3))
```

|              | x     |
|--------------|-------|
| (Intercept)  | 0.000 |
| cursmoke     | 0.030 |
| factor(sex)2 | 0.082 |
| factor(educ)2| 0.013 |
| factor(educ)3| 0.054 |
| factor(educ)4| 0.241 |
| bmi          | 0.000 |
| diabetes     | 0.000 |
| heartrte     | 0.000 |
| prevstrk     | 0.388 |
| death        | 0.000 |

```
QIC(diabp_fit2)
```

```
##    QIC
## 53508
```