

# Biostatistics (MATH11230)

Vanda Inácio

In this supplement I reproduce the results in the slides about confounding and interaction/effect modification. This first block reproduces the confounding analysis results.

```
require(readxl)
data_wcgs <- read_excel("wcgsdata.xls")

smoking_binary <- ifelse(data_wcgs$Ncigs0 == 0, 0, 1)
data_wcgs$smoking_binary <- factor(smoking_binary, levels = c(0, 1),
                                   labels = c("no", "yes"))
res_1 <- glm(Chd69 ~ smoking_binary, family = "binomial",
            data = data_wcgs)
summary(res_1)

##
## Call:
## glm(formula = Chd69 ~ smoking_binary, family = "binomial", data = data_wcgs)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.7636     0.1042  -26.54 < 2e-16 ***
## smoking_binaryyes  0.6299     0.1337   4.71 2.47e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 1781.2  on 3153  degrees of freedom
## Residual deviance: 1758.4  on 3152  degrees of freedom
## AIC: 1762.4
##
## Number of Fisher Scoring iterations: 5
exp(res_1$coefficients[2])

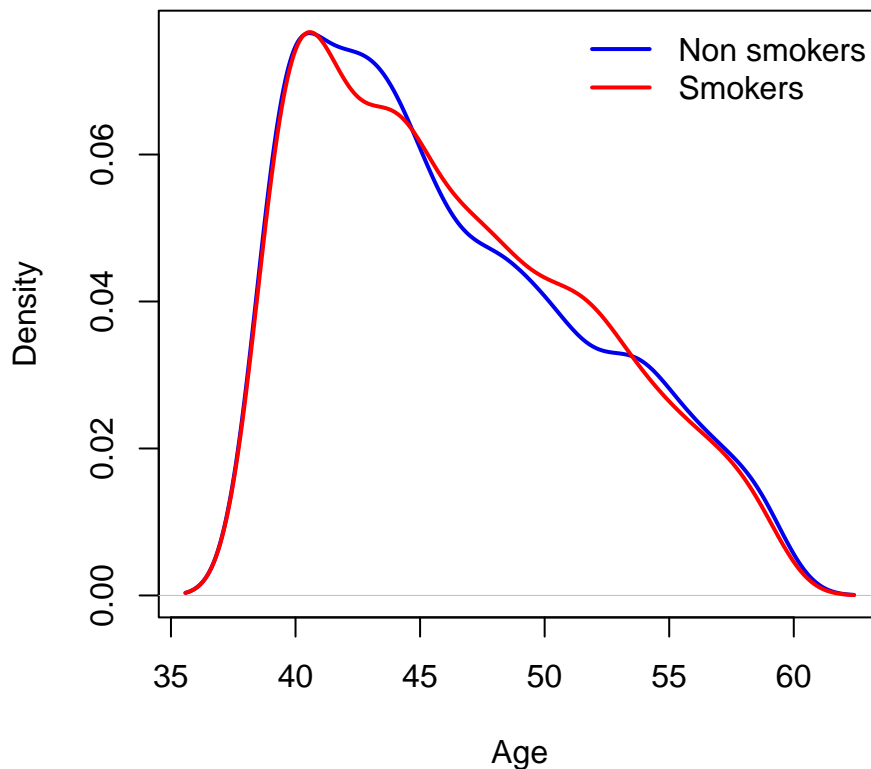
## smoking_binaryyes
##           1.877353
exp(confint.default(res_1)[2, ])

##    2.5 %    97.5 %
## 1.444517 2.439885
#association analysis investigations
Age_non_smokers <- data_wcgs$Age0[data_wcgs$smoking_binary == "no"]
Age_smokers <- data_wcgs$Age0[data_wcgs$smoking_binary == "yes"]
```

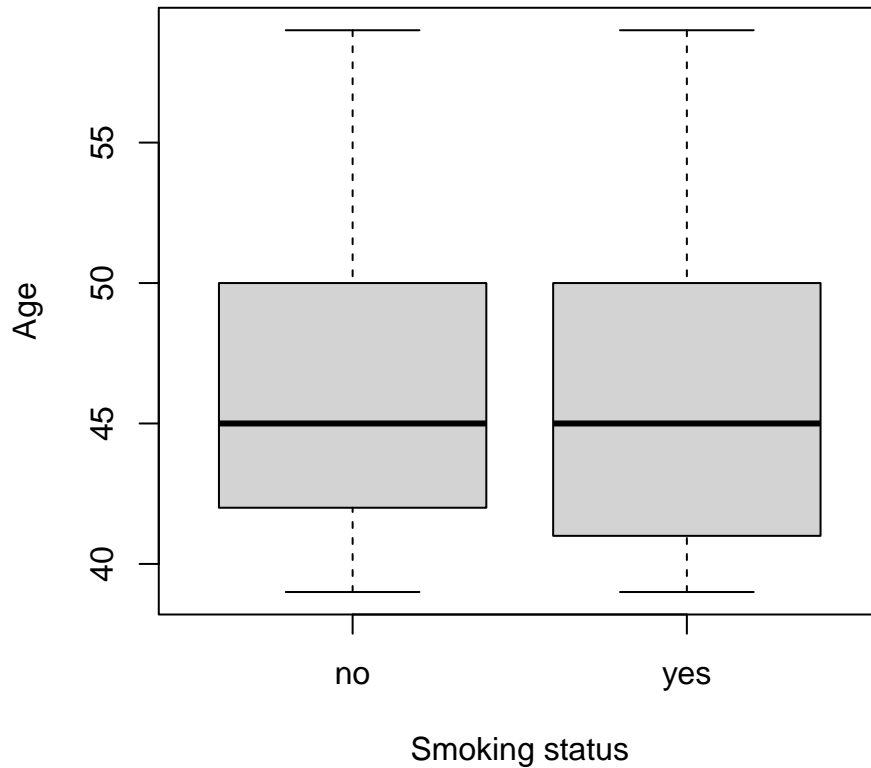
```
t.test(Age_non_smokers, Age_smokers, alternative = "two.sided", var.equal = FALSE)
```

```
##  
## Welch Two Sample t-test  
##  
## data: Age_non_smokers and Age_smokers  
## t = -0.26738, df = 3132.6, p-value = 0.7892  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.4385357 0.3332829  
## sample estimates:  
## mean of x mean of y  
## 46.25363 46.30626
```

```
plot(density(Age_non_smokers), col = "blue2", lwd = 2,  
     xlab = "Age", ylab = "Density", main = "")  
lines(density(Age_smokers), col = "red", lwd = 2)  
legend("topright", legend = c("Non smokers", "Smokers"),  
      lwd = c(2, 2), col = c("blue2", "red"), bty = "n")
```



```
boxplot(data_wcgs$Age0 ~ data_wcgs$smoking_binary,  
        ylab = "Age", xlab = "Smoking status")
```



```
res_association_exp_conf <- glm(smoking_binary ~ Age0, family = "binomial",
                               data = data_wcgs)
summary(res_association_exp_conf)

##
## Call:
## glm(formula = smoking_binary ~ Age0, family = "binomial", data = data_wcgs)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.175021  0.300841  -0.582   0.561
## Age0         0.001725  0.006454   0.267   0.789
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 4365.2  on 3153  degrees of freedom
## Residual deviance: 4365.2  on 3152  degrees of freedom
## AIC: 4369.2
##
## Number of Fisher Scoring iterations: 3
exp(res_association_exp_conf$coefficients[2])

##      Age0
## 1.001726
exp(confint.default(res_association_exp_conf)[2,])

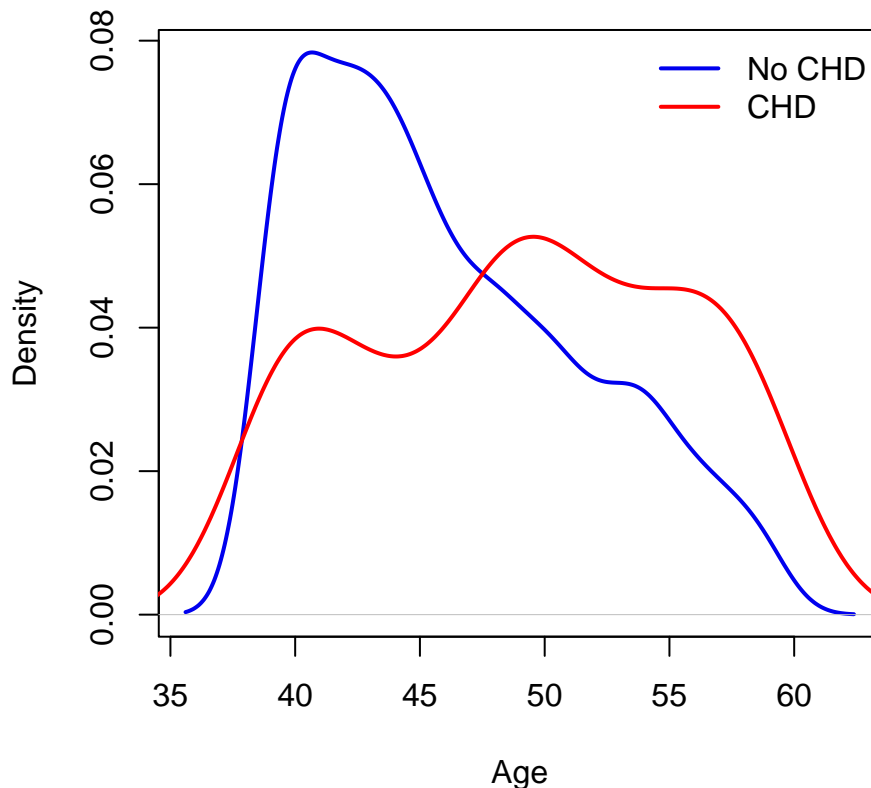
##      2.5 %      97.5 %
## 0.9891339 1.0144794
```

```
Age_no_CHD_no_smoke <- data_wcgs$Age0[data_wcgs$Chd69 == 0 & data_wcgs$smoking_binary == "no"]
Age_CHD_no_smoke <- data_wcgs$Age0[data_wcgs$Chd69 == 1 & data_wcgs$smoking_binary == "no"]

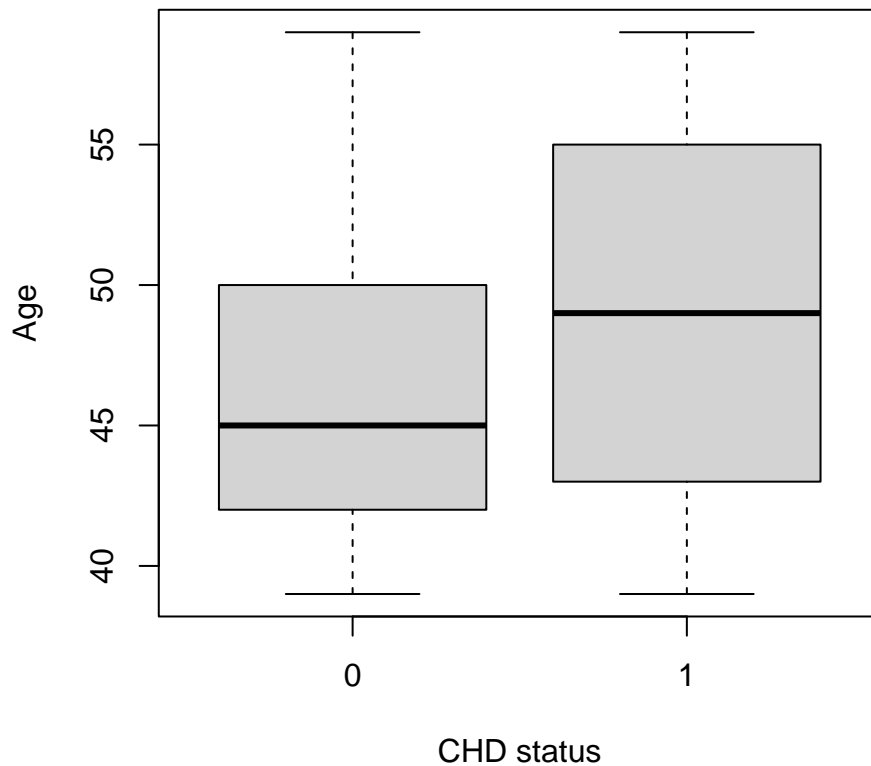
t.test(Age_no_CHD_no_smoke, Age_CHD_no_smoke, alternative = "two.sided", var.equal = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: Age_no_CHD_no_smoke and Age_CHD_no_smoke
## t = -4.6108, df = 106.25, p-value = 1.123e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -4.330029 -1.726047
## sample estimates:
## mean of x mean of y
## 46.07400 49.10204
```

```
plot(density(Age_no_CHD_no_smoke), col = "blue2", lwd = 2,
     xlab = "Age", ylab = "Density", main = "")
lines(density(Age_CHD_no_smoke), col = "red", lwd = 2)
legend("topright", legend = c("No CHD", "CHD"),
     lwd = c(2, 2), col = c("blue2", "red"), bty = "n")
```



```
boxplot(data_wcgs$Age0[data_wcgs$smoking_binary == "no"] ~
        data_wcgs$Chd69[data_wcgs$smoking_binary == "no"],
        ylab = "Age", xlab = "CHD status")
```



```
res_association_out_conf <- glm(data_wcgs$Chd69[which(data_wcgs$smoking_binary == "no")] ~
  data_wcgs$Age0[which(data_wcgs$smoking_binary == "no")],
  family = "binomial")

summary(res_association_out_conf)
```

```
##
## Call:
## glm(formula = data_wcgs$Chd69[which(data_wcgs$smoking_binary ==
##   "no")] ~ data_wcgs$Age0[which(data_wcgs$smoking_binary ==
##   "no")], family = "binomial")
##
## Coefficients:
##                                     Estimate Std. Error
## (Intercept)                        -7.07999    0.87867
## data_wcgs$Age0[which(data_wcgs$smoking_binary == "no")]  0.09077    0.01785
##                                     z value Pr(>|z|)
## (Intercept)                        -8.058 7.78e-16 ***
## data_wcgs$Age0[which(data_wcgs$smoking_binary == "no")]  5.086 3.66e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 743.72  on 1651  degrees of freedom
## Residual deviance: 717.99  on 1650  degrees of freedom
## AIC: 721.99
##
## Number of Fisher Scoring iterations: 6
```

```

exp(res_association_out_conf$coefficients[2])

## data_wcgs$Age0[which(data_wcgs$smoking_binary == "no")]
## 1.09502

exp(confint.default(res_association_out_conf))[2,]

## 2.5 % 97.5 %
## 1.057375 1.134004

#model adjusting for the potential confounding effect of age
res_2 <- glm(Chd69 ~ smoking_binary + Age0, family = "binomial",
             data = data_wcgs)
summary(res_2)

##
## Call:
## glm(formula = Chd69 ~ smoking_binary + Age0, family = "binomial",
## data = data_wcgs)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.32132 0.56144 -11.259 < 2e-16 ***
## smoking_binaryyes 0.63816 0.13472 4.737 2.17e-06 ***
## Age0 0.07518 0.01139 6.599 4.13e-11 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1781.2 on 3153 degrees of freedom
## Residual deviance: 1715.2 on 3151 degrees of freedom
## AIC: 1721.2
##
## Number of Fisher Scoring iterations: 5

exp(res_2$coefficients[2])

## smoking_binaryyes
## 1.892994

exp(confint.default(res_2)[2, ])

## 2.5 % 97.5 %
## 1.453712 2.465019

#rule of thumb
(as.numeric((exp(res_1$coefficients[2]) - exp(res_2$coefficients[2]))/exp(res_1$coefficients[2]))) * 100

## [1] -0.8331309

#information criteria
AIC(res_1); AIC(res_2)

## [1] 1762.381
## [1] 1721.234

```

```
BIC(res_1); BIC(res_2)
```

```
## [1] 1774.494
```

```
## [1] 1739.404
```

```
#Likelihood ratio test
```

```
dif <- res_1$deviance - res_2$deviance
```

```
dif
```

```
## [1] 43.14663
```

```
pchisq(dif, df = 2-1, lower = FALSE)
```

```
## [1] 5.07874e-11
```

```
#alternative and easier way
```

```
anova(res_1, res_2, test = "Chisq")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: Chd69 ~ smoking_binary
```

```
## Model 2: Chd69 ~ smoking_binary + Age0
```

```
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
```

```
## 1      3152      1758.4
```

```
## 2      3151      1715.2  1   43.147 5.079e-11 ***
```

```
## ---
```

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

We now reproduce the results for the interaction (between smoking status and age) case.

```
#Fitting the model with the interaction term
```

```
res_interaction <- glm(Chd69 ~ smoking_binary*Age0,
```

```
                        family = "binomial", data = data_wcgs)
```

```
summary(res_interaction)
```

```
##
```

```
## Call:
```

```
## glm(formula = Chd69 ~ smoking_binary * Age0, family = "binomial",
```

```
##   data = data_wcgs)
```

```
##
```

```
## Coefficients:
```

```
##               Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)    -7.07999    0.87867  -8.058 7.78e-16 ***
```

```
## smoking_binaryyes    1.91472    1.13209   1.691  0.0908 .
```

```
## Age0            0.09077    0.01785   5.086 3.66e-07 ***
```

```
## smoking_binaryyes:Age0 -0.02639    0.02319  -1.138  0.2553
```

```
## ---
```

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

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
```

```
##   Null deviance: 1781.2  on 3153  degrees of freedom
```

```
## Residual deviance: 1713.9  on 3150  degrees of freedom
```

```
## AIC: 1721.9
```

```
##
```

```
## Number of Fisher Scoring iterations: 6
```

```

#Fitting the model without the interaction term
res <- glm(Chd69 ~ smoking_binary + Age0,
           family = "binomial", data = data_wcgs)
summary(res)

##
## Call:
## glm(formula = Chd69 ~ smoking_binary + Age0, family = "binomial",
##      data = data_wcgs)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -6.32132    0.56144 -11.259  < 2e-16 ***
## smoking_binaryyes  0.63816    0.13472   4.737 2.17e-06 ***
## Age0           0.07518    0.01139   6.599 4.13e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1781.2  on 3153  degrees of freedom
## Residual deviance: 1715.2  on 3151  degrees of freedom
## AIC: 1721.2
##
## Number of Fisher Scoring iterations: 5
anova(res, res_interaction, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: Chd69 ~ smoking_binary + Age0
## Model 2: Chd69 ~ smoking_binary * Age0
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       3151      1715.2
## 2       3150      1713.9  1   1.2962  0.2549

dif_deviance <- res$deviance - res_interaction$deviance
pchisq(dif_deviance, df = 1, lower = FALSE)

## [1] 0.2549

#Computing the variance for the estimated log odds ratio associated with smoking for
#an individual aged 50 years old
est_odds_ratio <- as.numeric(exp(res_interaction$coefficients[2] + 50*res_interaction$coefficients[4]))
est_odds_ratio

## [1] 1.813858

cov_matrix <- vcov(res_interaction)
cov_matrix

##              (Intercept) smoking_binaryyes           Age0
## (Intercept)    0.77206343    -0.77206343 -0.0155705387
## smoking_binaryyes -0.77206343     1.28162683  0.0155705387
## Age0           -0.01557054     0.01557054  0.0003185725
## smoking_binaryyes:Age0  0.01557054    -0.02606741 -0.0003185725
##              smoking_binaryyes:Age0

```



```
## (Intercept)          0.0155705387
## smoking_binaryyes    -0.0260674138
## Age0                 -0.0003185725
## smoking_binaryyes:Age0 0.0005378728

est_var <- cov_matrix[2,2] + (50^2)*cov_matrix[4,4] + 2*50*cov_matrix[2,4]
exp(log(est_odds_ratio) - 1.96*sqrt(est_var))

## [1] 1.378896

exp(log(est_odds_ratio) + 1.96*sqrt(est_var))

## [1] 2.386025
```