Logistic Regression

Practical in R - solutions

1. Getting started

- Log into www.ucl.ac.uk/moodle
- Download the file **logregression_data.dta** to your space on the workstation
- Start *RStudio*, and on "Environment" click on "Import Dataset" and choose "From Stata…". Just browse your dataset file from your folder and click "import".

2. Examine data

- names(logregression_data)

```
 [1] "sex"       "age"       "ethnici"   "diabetes"  "cholval"   "bmival"    "cigst1"
 [8] "adt30gp"   "omdiaval"  "omsysval"  "fatbanda"  "bmi30"     "sclass2"   "lowinc"
[15] "inactiv"   "smoker"    "overlim"   "cvddef1"   "ag16g10"   "agegr"     "_est_a"
[22] "_est_b"    "_est_x"
```

- summary(logregression_data)

```
      sex              age            ethnici         diabetes          cholval            bmival      
 Min.   :0.000   Min.   :16.00   Min.   :1.00   Min.   :0.00000   Min.   : 1.900   Min.   :14.06  
 1st Qu.:0.000   1st Qu.:34.00   1st Qu.:1.00   1st Qu.:0.00000   1st Qu.: 4.900   1st Qu.:23.44  
 Median :1.000   Median :47.00   Median :1.00   Median :0.00000   Median : 5.600   Median :26.35  
 Mean   :0.555   Mean   :48.21   Mean   :1.31   Mean   :0.04118   Mean   : 5.696   Mean   :26.96  
 3rd Qu.:1.000   3rd Qu.:62.00   3rd Qu.:1.00   3rd Qu.:0.00000   3rd Qu.: 6.400   3rd Qu.:29.69  
 Max.   :1.000   Max.   :99.00   Max.   :7.00   Max.   :1.00000   Max.   :13.900   Max.   :61.99  
                                 NA's   :35                       NA's   :7120     NA's   :1780   
     cigst1          adt30gp         omdiaval         omsysval         fatbanda         bmi30       
 Min.   :1.000   Min.   :1.000   Min.   : -8.00   Min.   : -8.00   Min.   :1.000   Min.   :0.0000  
 1st Qu.:1.000   1st Qu.:1.000   1st Qu.: -1.00   1st Qu.: -1.00   1st Qu.:1.000   1st Qu.:0.0000  
 Median :3.000   Median :2.000   Median : 64.00   Median :113.00   Median :1.000   Median :0.0000  
 Mean   :2.309   Mean   :1.914   Mean   : 45.21   Mean   : 79.31   Mean   :1.246   Mean   :0.2321  
 3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.: 76.50   3rd Qu.:131.00   3rd Qu.:1.000   3rd Qu.:0.0000  
 Max.   :4.000   Max.   :3.000   Max.   :151.50   Max.   :240.00   Max.   :3.000   Max.   :1.0000  
 NA's   :72      NA's   :45                                        NA's   :4166    NA's   :1780    
    sclass2          lowinc          inactiv          smoker          overlim          cvddef1       
 Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :1.000   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:0.0000  
 Median :0.0000   Median :0.0000   Median :0.000   Median :2.000   Median :0.0000   Median :0.0000  
 Mean   :0.4163   Mean   :0.1786   Mean   :0.375   Mean   :1.809   Mean   :0.3693   Mean   :0.1437  
 3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:0.0000  
 Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :3.000   Max.   :1.0000   Max.   :1.0000  
 NA's   :734      NA's   :2298     NA's   :45      NA's   :72      NA's   :221                      
    ag16g10          agegr           _est_a          _est_b      _est_x      
 Min.   :1.000   Min.   :0.0000   Min.   :0.0000   Min.   :1   Min.   :0.0000  
 1st Qu.:2.000   1st Qu.:0.0000   1st Qu.:1.0000   1st Qu.:1   1st Qu.:1.0000  
 Median :4.000   Median :0.0000   Median :1.0000   Median :1   Median :1.0000  
 Mean   :3.858   Mean   :0.4439   Mean   :0.9951   Mean   :1   Mean   :0.9951  
 3rd Qu.:5.000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1   3rd Qu.:1.0000  
 Max.   :7.000   Max.   :1.0000   Max.   :1.0000   Max.   :1   Max.   :1.0000  
```

- Our response variable will be diabetes
  - table(logregression_data$diabetes)

```
    0     1 
14225   611 
```

  - prop.table(table(logregression_data$diabetes))

```
         0          1 
0.95881639 0.04118361 
```

  - mean(logregression_data$diabetes)
  - sd(logregression_data$diabetes)

```
> mean(logregression_data$diabetes)
[1] 0.04118361
> sd(logregression_data$diabetes)
[1] 0.1987214
```

- summary(logregression_data$diabetes)

```
> summary(logregression_data$diabetes)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00000 0.00000 0.00000 0.04118 0.00000 1.00000
```

- prop.table(table(logregression_data$diabetes, logregression_data$sex),2)

```
           0         1
0 0.95213572 0.96417294
1 0.04786428 0.03582706
```

In case it is too hard to understand this table, you can add a label to the categories of variables diabetes and sex by typing:

— logregression_data$diabetes <- factor(logregression_data$diabetes, levels = c(0,1), labels = c("No", "Yes"))

— logregression_data$sex <- factor(logregression_data$sex, levels = c(0,1), labels = c("Men", "Women"))

```
            Men       Women
No  0.95213572 0.96417294
Yes 0.04786428 0.03582706
```

- prop.table(table(logregression_data$diabetes, logregression_data$ag16g10),2)

```
             1           2           3           4           5           6           7
No  0.993887531 0.993073593 0.980215203 0.970588235 0.938174274 0.901080159 0.907534247
Yes 0.006112469 0.006926407 0.019784797 0.029411765 0.061825726 0.098919841 0.092465753
```

QUESTION 1 - What are your preliminary ideas about the relationship between diabetes and sex and diabetes and ag16g10?

The prevalence of being diagnosed with diabetes seems to be higher in men compared to women, and lower in younger groups compared to older groups.

3. LOGISTIC REGRESSION

- Use logistic regression to examine the association between diabetes and ag16g10 and interpret the OR

- model1 <- glm(diabetes ~ factor(ag16g10),
  data = logregression_data,
  family = binomial(link = "logit"))
- summary(model1)

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4564  -0.3573  -0.2443  -0.1179   3.1929

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -5.0913     0.3171 -16.055  < 2e-16 ***
factor(ag16g10)2   0.1258     0.4043   0.311 0.755645
factor(ag16g10)3   1.1884     0.3442   3.453 0.000555 ***
factor(ag16g10)4   1.5948     0.3395   4.697 2.64e-06 ***
factor(ag16g10)5   2.3717     0.3282   7.226 4.96e-13 ***
factor(ag16g10)6   2.8820     0.3270   8.813  < 2e-16 ***
factor(ag16g10)7   2.8074     0.3297   8.514  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5094.3  on 14835  degrees of freedom
Residual deviance: 4657.9  on 14829  degrees of freedom
AIC: 4671.9

Number of Fisher Scoring iterations: 7
```

- exp(cbind(odds=coef(model1), confint(model1)))

```
                            odds       2.5 %      97.5 %
(Intercept)       0.006150062 0.003076772  0.01081065
factor(ag16g10)2  1.134088807 0.520484258  2.59379869
factor(ag16g10)3  3.281940065 1.749495253  6.84515158
factor(ag16g10)4  4.927272059 2.656417869 10.20260977
factor(ag16g10)5 10.715345738 5.936763365 21.79686951
factor(ag16g10)6 17.850092216 9.918018583 36.24290596
factor(ag16g10)7 16.566790206 9.145429144 33.78174442
```

QUESTION 2 - How can you interpret the odds ratio of those aged 45-54? And of those aged 75+?

Compared to those in the youngest age group, those aged 45 to 54 are 4.9 (95% CI 2.5 to 9.6) times more likely to have diabetes; while those in the oldest age group are 16.6 (95% CI 8.7 to 31.6) times more likely to have diabetes. All ORs of age, except the first, are significantly different from 1, indicating that age group overall is statistically significant

- Use logistic regression to examine the association between **diabetes** and **sex** and interpret the OR
  - sex <- factor(sex)
  - sex <- relevel(sex, ref="0")
  - model2 <- glm(diabetes ~ sex,
    data = logregression_data,
    family = binomial(link = "logit"))
  - summary(model2)

```
Call:
glm(formula = diabetes ~ sex, family = binomial(link = "logit"),
    data = logregression_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.3132  -0.3132  -0.2701  -0.2701   2.5803

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.99034    0.05765 -51.870  < 2e-16 ***
sex         -0.30223    0.08270  -3.654 0.000258 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5094.3  on 14835  degrees of freedom
Residual deviance: 5081.0  on 14834  degrees of freedom
AIC: 5085

Number of Fisher Scoring iterations: 6
```

- exp(cbind(odds=coef(model2), confint(model2)))

```
                      odds      2.5 %     97.5 %
(Intercept) 0.05027044 0.04481117 0.05617781
sex         0.73916860 0.62842545 0.86918486
```

Women compared to men are less likely to have diabetes, the OR is 0.73 (95% CI 0.63 to 0.87).

4. MULTIVARIATE ANALYSES

Exercise 4.1

Run a logistic regression for **diabetes** (outcome variable) and **sex**, **ag16g10** (age groups), **ethnici** (ethnicity) as independent variables. Explain the results of the model and test whether ethnicity should be kept into the model or not.

- model3 <- glm(diabetes ~ sex + factor(ag16g10),
  data = subset(logregression_data, ethnici>0),
  family = binomial(link = "logit"))
- summary(model3)

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4953  -0.3883  -0.2251  -0.1287   3.2438

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -4.91862    0.31952 -15.394  < 2e-16 ***
sex              -0.33722    0.08417  -4.006 6.17e-05 ***
factor(ag16g10)2  0.12909    0.40437   0.319 0.749550
factor(ag16g10)3  1.19375    0.34423   3.468 0.000525 ***
factor(ag16g10)4  1.59261    0.33958   4.690 2.73e-06 ***
factor(ag16g10)5  2.37126    0.32825   7.224 5.05e-13 ***
factor(ag16g10)6  2.88211    0.32708   8.812  < 2e-16 ***
factor(ag16g10)7  2.83716    0.32989   8.600  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5091.4  on 14800  degrees of freedom
Residual deviance: 4638.8  on 14793  degrees of freedom
AIC: 4654.8

Number of Fisher Scoring iterations: 7
```

- exp(cbind(odds=coef(model3), confint(model3)))

```
                        odds       2.5 %      97.5 %
(Intercept)       0.007309246 0.003642707  0.01292085
sexwomen          0.713751318 0.605061888  0.84171033
factor(ag16g10)2  1.137791434 0.522139856  2.60246123
factor(ag16g10)3  3.299445364 1.758639231  6.88220467
factor(ag16g10)4  4.916579209 2.650364854 10.18126629
factor(ag16g10)5 10.710899840 5.933621384 21.78955831
factor(ag16g10)6 17.851945616 9.917718677 36.25003954
factor(ag16g10)7 17.067231147 9.418337824 34.81100206
```

- model4<- glm(diabetes ~ sex + factor(ag16g10) + factor(ethnici),

  data = logregression_data,

  family = binomial(link = "logit"))

- summary(model4)

- exp(cbind(odds=coef(model4), confint(model4)))

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9401  -0.3717  -0.2097  -0.1115   3.3230

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)      -5.19749    0.32393 -16.045  < 2e-16 ***
sex              -0.31960    0.08456  -3.780 0.000157 ***
factor(ag16g10)2  0.11910    0.40530   0.294 0.768877
factor(ag16g10)3  1.28045    0.34542   3.707 0.000210 ***
factor(ag16g10)4  1.71074    0.34127   5.013 5.36e-07 ***
factor(ag16g10)5  2.55975    0.33081   7.738 1.01e-14 ***
factor(ag16g10)6  3.08948    0.33005   9.361  < 2e-16 ***
factor(ag16g10)7  3.07890    0.33329   9.238  < 2e-16 ***
factor(ethnici)2  0.75472    0.61091   1.235 0.216679
factor(ethnici)3  0.84404    0.38106   2.215 0.026761 *
factor(ethnici)4  0.87567    0.35939   2.437 0.014828 *
factor(ethnici)5  1.15835    0.24157   4.795 1.63e-06 ***
factor(ethnici)6  1.52031    0.23256   6.537 6.26e-11 ***
factor(ethnici)7  1.31179    0.34587   3.793 0.000149 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5091.4  on 14800  degrees of freedom
Residual deviance: 4573.7  on 14787  degrees of freedom
  (35 observations deleted due to missingness)
```

```
                        odds       2.5 %       97.5 %
(Intercept)       0.005530439  0.002736248  0.009875427
sexwomen          0.726441360  0.615369263  0.857332727
factor(ag16g10)2  1.126477252  0.515960621  2.580968751
factor(ag16g10)3  3.598256541  1.912829018  7.520860069
factor(ag16g10)4  5.533029475  2.971517057 11.490777058
factor(ag16g10)5 12.932644115  7.123300393 26.421878044
factor(ag16g10)6 21.965662142 12.121751940 44.824816965
factor(ag16g10)7 21.734403865 11.902924968 44.582502113
factor(ethnici)2  2.127025879  0.505774466  6.030207735
factor(ethnici)3  2.325744923  1.018224634  4.623594383
factor(ethnici)4  2.400484430  1.106946651  4.602838455
factor(ethnici)5  3.184684085  1.930987418  5.000983714
factor(ethnici)6  4.573626484  2.836757097  7.086132768
factor(ethnici)7  3.712819202  1.773825789  6.978257199
```

The direction of the association between age groups and diabetes, and sex and diabetes were the same when comparing the model with and without ethnicity. However, the odds ratios increased after including ethnicity into the model. Nevertheless, in model 3 we can see that apart from individuals aged 25-34 (OR 1.13, 95% CI 0.51 – 2.51), all other older age groups are more likely to be diagnosed with diabetes compared to those aged 16-24.

In model 4, females are less likely (OR 0.72, 95% CI 0.61 – 0.85) than males to be diagnosed with diabetes. Apart from individuals aged 25-34 (OR 1.12, 95% CI 0.51 – 2.49), all other older age groups are more likely to be diagnosed with diabetes compared to those aged 16-24. Additionally, apart from mixed ethnic group (OR 2.12, 95% 0.64 – 7.04), all other ethnic groups were more likely to be diagnosed with diabetes compared to Whites.

- install.packages("lmtest")
- library(lmtest)
- lrtest(model3, model4)

```
Model 1: diabetes ~ sex + factor(ag16g10)
Model 2: diabetes ~ sex + factor(ag16g10) + factor(ethnici)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   8 -2319.4
2  14 -2286.8  6 65.068  4.179e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The LRT ($p<0.0001$) provides strong evidence that there is an association between ethnicity and the odds of diabetes (after accounting for age and sex).

Exercise 4.2

Run a logistic regression for the association between obesity (**bmi30**) as outcome and social class (**sclass2**), smoking status (**smoker**), alcohol consumption (**overlim**), adjusted for **age** and **separately** for men and women.

- model5<- glm(bmi30 ~ sclass2 + factor(smoker) + overlim + age,
  data = subset(logregression_data, sex==0),
  family = binomial(link = "logit"))
- summary(model5)

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0271  -0.7719  -0.6565  -0.5220   2.0947

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.825470   0.117333 -15.558  < 2e-16 ***
sclass2          0.215174   0.064462   3.338 0.000844 ***
factor(smoker)2  0.275890   0.074852   3.686 0.000228 ***
factor(smoker)3 -0.450278   0.090466  -4.977 6.45e-07 ***
overlim          0.151200   0.065595   2.305 0.021163 *
age              0.009099   0.002024   4.496 6.91e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6212.1  on 5721  degrees of freedom
Residual deviance: 6082.2  on 5716  degrees of freedom
  (880 observations deleted due to missingness)
AIC: 6094.2

Number of Fisher Scoring iterations: 4
```

- Ask for odds ratio: exp(cbind(odds=coef(model5), confint(model5)))

```
                      odds     2.5 %    97.5 %
(Intercept)     0.1611419 0.1278456 0.2025200
sclass2         1.2400774 1.0929880 1.4072530
factor(smoker)2 1.3177033 1.1380607 1.5262152
factor(smoker)3 0.6374507 0.5332631 0.7603373
overlim         1.1632291 1.0229289 1.3229124
age             1.0091409 1.0051497 1.0131564
```

Men exceeding the recommended daily units of alcohol consumption are more likely (OR 1.16 95%CI:1.02; 1.32 p<0.05) to be obese that those who don't exceed the recommended daily units of alcohol (when we adjust for age, social class and smoking status). Current smokers are less likely (OR 0.63 95%CI: 0.53; 0.76 p<0.001) to be obese that non-smokers, while ex-smokers are 1.3 times (95%CI: 1.14; 1.52 p<0.001) more likely to be obese that non-smokers, adjusting for age, social class and alcohol consumption.

- model6<- glm(bmi30 ~ sclass2 + factor(smoker) + overlim + age,
  data = subset(logregression_data, sex==1),
  family = binomial(link = "logit"))
- summary(model6)

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1110  -0.7551  -0.6562  -0.5283   2.0537

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.836113   0.104153 -17.629  < 2e-16 ***
sclass2          0.443479   0.060677   7.309 2.70e-13 ***
factor(smoker)2  0.295976   0.068858   4.298 1.72e-05 ***
factor(smoker)3 -0.059408   0.076904  -0.773 0.439818
overlim         -0.255518   0.068533  -3.728 0.000193 ***
age              0.010089   0.001758   5.739 9.53e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7235.1  on 6630  degrees of freedom
Residual deviance: 7068.0  on 6625  degrees of freedom
  (1603 observations deleted due to missingness)
AIC: 7080

Number of Fisher Scoring iterations: 4
```

- exp(cbind(odds=coef(model6), confint(model6)))

```
                          odds     2.5 %    97.5 %
(Intercept)          0.1594359 0.1298345 0.1953105
sclass2              1.5581188 1.3832083 1.7546817
factor(smoker)2      1.3444378 1.1744853 1.5384722
factor(smoker)3      0.9423218 0.8099333 1.0949705
overlim              0.7745152 0.6767372 0.8853500
age                  1.0101404 1.0066698 1.0136323
```

Women exceeding the recommended daily units of alcohol consumption are less likely (OR 0.77 95%CI: 0.68; 0.89 p<0.001) to be obese that those who don't exceed the recommended daily units of alcohol (when we adjust for age, social class and smoking status). Current smokers are not significantly less likely to be obese that non-smokers as indicated by the p-value greater than 0.05, while ex-smokers are more 1.34 times more (95%CI:1.17; 1.53 p<0.005) likely to be obese that non-smokers, adjusting for age, social class and alcohol consumption.

The main difference between the two models is that women exceeding recommended daily units of alcohol are less likely to be obese than those who don't while for men it is the opposite! Also among women, the OR for obesity is not statistically significant in current smoker vs non-smoker, meaning that the two groups of smokers do not differ significantly in their risk of being obese.

5. INTERACTION TERMS

Exercise 5.1

- Fit a logistic regression model for the effects of **sex**, **agegr** (indicates whether age is above or below 50) and their interaction on the odds of CVD
  - model7 <- glm(cvddef1 ~ sex + agegr + sex:agegr, data = logregression_data, family = binomial(link = "logit"))
  - summary(model7)

```
Call:
glm(formula = cvddef1 ~ sex + agegr + sex:agegr, family = binomial(link = "logi
t"),
    data = logregression_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7905  -0.7036  -0.3855  -0.3493   2.3779

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.76626    0.06983 -39.617  < 2e-16 ***
sex          0.20390    0.09039   2.256   0.0241 *
agegr        1.76317    0.08137  21.669  < 2e-16 ***
sex:agegr   -0.47084    0.10728  -4.389 1.14e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 12214  on 14835  degrees of freedom
Residual deviance: 11261  on 14832  degrees of freedom
AIC: 11269

Number of Fisher Scoring iterations: 5
```

- exp(cbind(odds=coef(model7), confint(model7)))

```
                odds      2.5 %     97.5 %
(Intercept) 0.06289671 0.05469722 0.07192762
sex         1.22617924 1.02795216 1.46531456
agegr       5.83090475 4.98041156 6.85254935
sex:agegr   0.62447759 0.50570032 0.77014887
```

- Write down the following odds ratios from the output:

  1) The odds ratio for the effect of sex (women versus men) at the baseline value of age (≤50): **1.23 (1.02 to 1.46)**

  2) The odds ratio for the effect of age at the baseline value of sex (men): **5.83 (4.97 to 6.84)**

  3) The interaction term between sex and age: **0.62 (0.51 to 0.77)**

  4) The estimated odds ratio for women vs men among those not at the baseline of age (aged 51+): **1.22*0.62=0.756**

  5) The estimated odds ratio for the effect of age (51+ vs ≤50) among women: **5.83*0.62=3.61**

Summarise your results: The odds ratio of having CVD among younger women compared to younger men is 1.23 (95% CI 1.03; 1.46), the OR of older women compared to older men is 0.756 (95% CI 0.68; 0.86), so at older ages women are less likely to have CVD than men. However, we find that the odds of having CVD are 5.8 times higher among older men compared to younger men (95% CI 4.97; 6.83), whereas the odds of having CVD are 3.6 times higher in older women compared to younger women (3.17; 4.17).

Exercise 5.2

- Fit a logistic regression model for CVD with an interaction term between sex and physical activity (**adt30gp**). Check the variables first, then interpret the results of each of the odds ratios obtained. Finally do a likelihood ratio test to see if there is an effect modification and based on the result run the appropriate model.
  - model8 <- glm(cvddef1 ~ factor(sex) + factor(adt30gp) + factor(sex):factor(adt30gp), data = logregression_data, family = binomial(link = "logit"))
  - summary(model8)

```
Call:
glm(formula = cvddef1 ~ factor(sex) + factor(adt30gp) + factor(sex):factor(adt30gp),
    family = binomial(link = "logit"), data = logregression_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7631  -0.6728  -0.4534  -0.4081   2.2481

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -1.08493    0.04871 -22.275  < 2e-16 ***
factor(sex)Women                 -0.28560    0.06512  -4.385 1.16e-05 ***
factor(adt30gp)2                 -0.87729    0.08306 -10.563  < 2e-16 ***
factor(adt30gp)3                 -1.35884    0.09099 -14.934  < 2e-16 ***
factor(sex)Women:factor(adt30gp)2  0.02451    0.11244   0.218  0.82741
factor(sex)Women:factor(adt30gp)3  0.38933    0.12839   3.032  0.00243 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 12193  on 14790  degrees of freedom
Residual deviance: 11718  on 14785  degrees of freedom
  (45 observations deleted due to missingness)
AIC: 11730

Number of Fisher Scoring iterations: 5
```

- exp(cbind(odds=coef(model8), confint(model8)))

```
                                      odds      2.5 %    97.5 %
(Intercept)                      0.3379269 0.3069228 0.371503
factor(sex)Women                 0.7515657 0.6615634 0.853995
factor(adt30gp)2                 0.4159082 0.3529914 0.488887
factor(adt30gp)3                 0.2569592 0.2145267 0.306521
factor(sex)Women:factor(adt30gp)2 1.0248163 0.8221975 1.277711
factor(sex)Women:factor(adt30gp)3 1.4759987 1.1472930 1.898215
```

The odds ratio for **factor(sex)women** is the effect of sex (women versus men) among those with a low physical activity level.

The odds ratio for **factor(adt30gp)2** is the odds ratio for the effect of medium physical activity (versus low physical activity) at the baseline value of sex (men).

The odds ratio for **factor(adt30gp)3** is the odds ratio for the effect of high physical activity (versus low physical activity) at the baseline value of sex (men).

The odds ratio for **factor(sex)women: factor(adt30gp)2** comes into play when comparing the cvd risk for a female with medium physical activity with a male with low physical activity, and needs to be multiplied by the odds ratios for **factor(sex)women** (female vs male) and **factor(adt30gp)2** (medium vs low physical activity) to calculate the odds ratio of cvd risk for a female with medium physical activity with a male with low physical activity.

The odds ratio for **factor(sex)women: factor(adt30gp)3** comes into play when comparing the cvd risk for a female with high physical activity with a male with low physical activity, and needs to be multiplied by the odds ratios for **factor(sex)women** (female vs male) and **factor(adt30gp)3** (high vs low physical activity) to calculate the odds ratio of cvd risk for a female with high physical activity with a male with low physical activity.

- model9 <- glm(cvddef1 ~ factor(sex) + factor(adt30gp), data = logregression_data, family = binomial(link = "logit"))
- summary(model9)

```
Call:
glm(formula = cvddef1 ~ factor(sex) + factor(adt30gp), family = binomial(link = "logit"),
    data = logregression_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7484  -0.6832  -0.4588  -0.3952   2.2754

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.12956    0.04235 -26.673  < 2e-16 ***
factor(sex)Women -0.20655    0.04805  -4.298 1.72e-05 ***
factor(adt30gp)2 -0.86233    0.05595 -15.411  < 2e-16 ***
factor(adt30gp)3 -1.17447    0.06446 -18.219  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 12193  on 14790  degrees of freedom
Residual deviance: 11728  on 14787  degrees of freedom
  (45 observations deleted due to missingness)
AIC: 11736

Number of Fisher Scoring iterations: 5
```

- exp(cbind(odds=coef(model9), confint(model9)))

```
                        odds      2.5 %    97.5 %
(Intercept)       0.3231760 0.2973004 0.3509915
factor(sex)Women  0.8133896 0.7403004 0.8937567
factor(adt30gp)2  0.4221784 0.3781024 0.4708504
factor(adt30gp)3  0.3089818 0.2719954 0.3502131
```

- lrtest(model8, model9)

```
Likelihood ratio test

Model 1: cvddef1 ~ factor(sex) + factor(adt30gp) + factor(sex):factor(adt30gp)
Model 2: cvddef1 ~ factor(sex) + factor(adt30gp)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   6 -5858.9
2   4 -5863.8 -2 9.6442    0.00805 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is strong evidence of an interaction between sex and levels of physical activity, and therefore the appropriate model is the one with interaction terms. Equally valid would be to choose to stratify the model by sex, or physical activity. If you stratify by physical activity groups you will see that the difference in the odds of having CVD between men and women is no longer significant in those who are highly active.

- model10 <- glm(cvddef1 ~ factor(sex), data = subset(logregression_data, adt30gp==1), family = binomial(link = "logit"))
- summary(model10)
- exp(cbind(odds=coef(model10), confint(model10)))

```
Call:
glm(formula = cvddef1 ~ factor(sex), family = binomial(link = "logit"),
    data = subset(logregression_data, adt30gp == 1))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7631  -0.7631  -0.6728  -0.6728   1.7871

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.08493    0.04871 -22.275  < 2e-16 ***
factor(sex)Women -0.28560    0.06512  -4.385 1.16e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5881.9  on 5545  degrees of freedom
Residual deviance: 5862.8  on 5544  degrees of freedom
AIC: 5866.8

Number of Fisher Scoring iterations: 4
```

```
                         odds      2.5 %    97.5 %
(Intercept)       0.3379269 0.3069228 0.371503
factor(sex)Women 0.7515657 0.6615634 0.853995
```

- model11 <- glm(cvddef1 ~ factor(sex), data = subset(logregression_data, adt30gp==2), family = binomial(link = "logit"))
- summary(model11)
- exp(cbind(odds=coef(model11), confint(model11)))

```
Call:
glm(formula = cvddef1 ~ factor(sex), family = binomial(link = "logit"),
    data = subset(logregression_data, adt30gp == 2))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5129  -0.5129  -0.4534  -0.4534   2.1569

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.96222    0.06728 -29.167  < 2e-16 ***
factor(sex)Women -0.26108    0.09165  -2.849  0.00439 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3408.5  on 4972  degrees of freedom
Residual deviance: 3400.5  on 4971  degrees of freedom
AIC: 3404.5

Number of Fisher Scoring iterations: 4
```

```
                         odds      2.5 %     97.5 %
(Intercept)       0.1405466 0.1229081 0.1600146
factor(sex)Women 0.7702168 0.6436779 0.9221041
```

- model12 <- glm(cvddef1 ~ factor(sex), data = subset(logregression_data, adt30gp==3), family = binomial(link = "logit"))
- summary(model12)
- exp(cbind(odds=coef(model12), confint(model12)))

```
Call:
glm(formula = cvddef1 ~ factor(sex), family = binomial(link = "logit"),
    data = subset(logregression_data, adt30gp == 3))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4289  -0.4289  -0.4081  -0.4081   2.2481

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -2.44376    0.07686 -31.797   <2e-16 ***
factor(sex)Women  0.10374    0.11065   0.938    0.348
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2455.5  on 4271  degrees of freedom
Residual deviance: 2454.6  on 4270  degrees of freedom
AIC: 2458.6

Number of Fisher Scoring iterations: 5
```

	odds	2.5 %	97.5 %
(Intercept)	0.08683341	0.07444694	0.1006379
factor(sex)Women	1.10930994	0.89270533	1.3779234

OPTIONAL EXERCISE

- In the model for exercise 1 we want to test whether there is an effect modification between sex and alcohol and sex and smoking – i.e. whether sex modifies the effect of alcohol, and the effect of smoking. To do that we include in the model 2 interaction terms as follows:
    - model13 <- glm(bmi30 ~ age + sclass2 + factor(sex) + overlim + factor(smoker) + factor(sex):overlim + factor(sex):factor(smoker), data = logregression_data, family = binomial(link = "logit"))
    - summary(model13)
    - exp(cbind(odds=coef(model13), confint(model13)))

```
Call:
glm(formula = bmi30 ~ age + sclass2 + factor(sex) + overlim +
    factor(smoker) + factor(sex):overlim + factor(sex):factor(smoker),
    family = binomial(link = "logit"), data = logregression_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0791  -0.7621  -0.6589  -0.5196   2.1373

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -1.908825   0.089368 -21.359  < 2e-16 ***
age                             0.009705   0.001327   7.316 2.56e-13 ***
sclass2                         0.336399   0.044332   7.588 3.24e-14 ***
factor(sex)Women                0.132022   0.075373   1.752 0.079846 .
overlim                         0.159129   0.064738   2.458 0.013970 *
factor(smoker)2                 0.257287   0.072695   3.539 0.000401 ***
factor(smoker)3                -0.481228   0.089751  -5.362 8.24e-08 ***
factor(sex)Women:overlim       -0.428181   0.091990  -4.655 3.25e-06 ***
factor(sex)Women:factor(smoker)2  0.044974   0.098626   0.456 0.648389
factor(sex)Women:factor(smoker)3  0.437625   0.117031   3.739 0.000184 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13447  on 12352  degrees of freedom
Residual deviance: 13157  on 12343  degrees of freedom
  (2483 observations deleted due to missingness)
AIC: 13177

Number of Fisher Scoring iterations: 4
```

```
                                      odds     2.5 %    97.5 %
(Intercept)                       0.1482545 0.1243193 0.1764792
age                               1.0097519 1.0071321 1.0123830
sclass2                           1.3998981 1.2833706 1.5269604
factor(sex)Women                  1.1411335 0.9847738 1.3233433
overlim                           1.1724892 1.0327624 1.3311483
factor(smoker)2                   1.2934169 1.1218512 1.4918017
factor(smoker)3                   0.6180242 0.5177318 0.7361272
factor(sex)Women:overlim          0.6516935 0.5439764 0.7801890
factor(sex)Women:factor(smoker)2 1.0460004 0.8619942 1.2688794
factor(sex)Women:factor(smoker)3 1.5490235 1.2319425 1.9492341
```

- **Is the interaction term between sex and alcohol significant? What do you conclude?**

The p-value for the interaction term between sex and alcohol **factor(sex):overlim** is <0.001 therefore we can conclude that after adjusting for age, social class, smoking and sex (and the interaction) there is a significant effect modification. We know, however, that the best way of checking this is using the likelihood ratio test, so we compare the results with a model with does not have interaction between sex and alcohol.

You don't need to interpret the results of this model, you should focus on the LRT for the interaction and decide whether or not your model should be stratified by gender

- model14 <- glm(bmi30 ~ age + sclass2 + factor(sex) + overlim + factor(smoker) + factor(sex):factor(smoker), data = logregression_data, family = binomial(link = "logit"))
- summary(model14)
- exp(cbind(odds=coef(model14), confint(model14)))

```
Call:
glm(formula = bmi30 ~ age + sclass2 + factor(sex) + overlim +
    factor(smoker) + factor(sex):factor(smoker), family = binomial(link = "logit"),
    data = logregression_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0579  -0.7595  -0.6612  -0.5149   2.1105

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -1.832714   0.087405 -20.968  < 2e-16 ***
age                              0.009958   0.001325   7.515 5.69e-14 ***
sclass2                          0.342864   0.044301   7.739 9.98e-15 ***
factor(sex)Women                -0.008685   0.068700  -0.126 0.899402
overlim                         -0.049117   0.046914  -1.047 0.295116
factor(smoker)2                  0.252506   0.072733   3.472 0.000517 ***
factor(smoker)3                 -0.450128   0.089499  -5.029 4.92e-07 ***
factor(sex)Women:factor(smoker)2  0.032173   0.098507   0.327 0.743964
factor(sex)Women:factor(smoker)3  0.364429   0.115899   3.144 0.001664 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13447  on 12352  degrees of freedom
Residual deviance: 13179  on 12344  degrees of freedom
  (2483 observations deleted due to missingness)
AIC: 13197

Number of Fisher Scoring iterations: 4
```

```
                                      odds     2.5 %    97.5 %
(Intercept)                       0.1599788 0.1346767 0.1897166
age                               1.0100077 1.0073902 1.0126368
sclass2                           1.4089770 1.2917729 1.5367707
factor(sex)Women                  0.9913528 0.8666869 1.1345885
overlim                           0.9520699 0.8683195 1.0436432
factor(smoker)2                   1.2872478 1.1164144 1.4847951
factor(smoker)3                   0.6375463 0.5343541 0.7590112
factor(sex)Women:factor(smoker)2 1.0326963 0.8512270 1.2524425
factor(sex)Women:factor(smoker)3 1.4396915 1.1475203 1.8076188
```

- lrtest(model13,model14)

```
Model 1: bmi30 ~ age + sclass2 + factor(sex) + overlim + factor(smoker) +
    factor(sex):overlim + factor(sex):factor(smoker)
Model 2: bmi30 ~ age + sclass2 + factor(sex) + overlim + factor(smoker) +
    factor(sex):factor(smoker)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  10 -6578.6
2   9 -6589.5 -1 21.845  2.956e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value for likelihood ratio statistic

The likelihood ratio confirms what we found with the p-value for the coefficient, i.e. sex modifies the effect of drinking on BMI after accounting for social class and age.

- **Is the interaction term between sex and smoking significant? How would you assess the overall significance of the interaction?**

  The p-value for the interaction term between sex and ex-smoker is not significant, while that of the interaction term between current smokers and sex is significant, after adjusting for all other variables. The way to test whether there should be an interaction term between sex and smoking is to do a likelihood ratio test, comparing models with, and without, an interaction between sex and smoking:

  Run the LR test to check for the overall significance of the interaction term between sex and smoking status. Discuss the result.

  - model15 <- glm(bmi30 ~ age + sclass2 + factor(sex) + overlim + factor(smoker) + factor(sex):overlim, data = logregression_data, family = binomial(link = "logit"))
  - summary(model15)
  - exp(cbind(odds=coef(model15), confint(model15)))

```
Call:
glm(formula = bmi30 ~ age + sclass2 + factor(sex) + overlim +
    factor(smoker) + factor(sex):overlim, family = binomial(link = "logit"),
    data = logregression_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0894  -0.7590  -0.6592  -0.5272   2.0833

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -1.961451   0.086077 -22.787  < 2e-16 ***
age                       0.009721   0.001321   7.360 1.83e-13 ***
sclass2                   0.330622   0.044344   7.456 8.93e-14 ***
factor(sex)Women          0.224591   0.056407   3.982 6.84e-05 ***
overlim                   0.135068   0.064277   2.101   0.0356 *
factor(smoker)2           0.291709   0.050352   5.793 6.90e-09 ***
factor(smoker)3          -0.233581   0.058522  -3.991 6.57e-05 ***
factor(sex)Women:overlim -0.378771   0.091058  -4.160 3.19e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13447  on 12352  degrees of freedom
Residual deviance: 13173  on 12345  degrees of freedom
  (2483 observations deleted due to missingness)
AIC: 13189

Number of Fisher Scoring iterations: 4
```

```
                              odds     2.5 %    97.5 %
(Intercept)               0.1406542 0.1187241 0.1663752
age                       1.0097682 1.0071598 1.0123879
sclass2                   1.3918336 1.2759502 1.5182056
factor(sex)Women          1.2518104 1.1210873 1.3985494
overlim                   1.1446146 1.0090968 1.2982938
factor(smoker)2           1.3387140 1.2128904 1.4775684
factor(smoker)3           0.7916933 0.7056007 0.8875689
factor(sex)Women:overlim 0.6847024 0.5725786 0.8182152
```

- lrtest(model13,model15)

```
Likelihood ratio test

Model 1: bmi30 ~ age + sclass2 + factor(sex) + overlim + factor(smoker) +
    factor(sex):overlim + factor(sex):factor(smoker)
Model 2: bmi30 ~ age + sclass2 + factor(sex) + overlim + factor(smoker) +
    factor(sex):overlim
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  10 -6578.6
2   8 -6586.3 -2 15.477  0.0004358 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The likelihood ratio test confirms that there is a significant interaction between sex and smoking, i.e. sex modifies the effect of smoking on BMI after accounting for social class and age. Therefore one might conclude (as sex modifies both the effect of drinking, and the effect of smoking) that it would be appropriate to stratify the model by sex

- model20 <- glm(bmi30 ~ age + sclass2 + overlim + factor(smoker), data = subset(logregression_data, sex==0), family = binomial(link = "logit"))
- summary(model20)

- exp(cbind(odds=coef(model20), confint(model20)))

```
Call:
glm(formula = bmi30 ~ age + sclass2 + overlim + factor(smoker),
    family = binomial(link = "logit"), data = subset(logregression_data,
        sex == 0))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0271  -0.7719  -0.6565  -0.5220   2.0947

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.825470   0.117333 -15.558  < 2e-16 ***
age              0.009099   0.002024   4.496 6.91e-06 ***
sclass2          0.215174   0.064462   3.338 0.000844 ***
overlim          0.151200   0.065595   2.305 0.021163 *
factor(smoker)2  0.275890   0.074852   3.686 0.000228 ***
factor(smoker)3 -0.450278   0.090466  -4.977 6.45e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6212.1  on 5721  degrees of freedom
Residual deviance: 6082.2  on 5716  degrees of freedom
  (880 observations deleted due to missingness)
AIC: 6094.2

Number of Fisher Scoring iterations: 4
```

	odds	2.5 %	97.5 %
(Intercept)	0.1611419	0.1278456	0.2025200
age	1.0091409	1.0051497	1.0131564
sclass2	1.2400774	1.0929880	1.4072530
overlim	1.1632291	1.0229289	1.3229124
factor(smoker)2	1.3177033	1.1380607	1.5262152
factor(smoker)3	0.6374507	0.5332631	0.7603373

- model21 <- glm(bmi30 ~ age + sclass2 + overlim + factor(smoker), data = subset(logregression_data, sex==0), family = binomial(link = "logit"))
- summary(model21)
- exp(cbind(odds=coef(model21), confint(model21)))

```
Call:
glm(formula = bmi30 ~ age + sclass2 + overlim + factor(smoker),
    family = binomial(link = "logit"), data = subset(logregression_data,
        sex == 1))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1110  -0.7551  -0.6562  -0.5283   2.0537

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.836113   0.104153 -17.629  < 2e-16 ***
age              0.010089   0.001758   5.739 9.53e-09 ***
sclass2          0.443479   0.060677   7.309 2.70e-13 ***
overlim         -0.255518   0.068533  -3.728 0.000193 ***
factor(smoker)2  0.295976   0.068858   4.298 1.72e-05 ***
factor(smoker)3 -0.059408   0.076904  -0.773 0.439818
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7235.1  on 6630  degrees of freedom
Residual deviance: 7068.0  on 6625  degrees of freedom
  (1603 observations deleted due to missingness)
AIC: 7080

Number of Fisher Scoring iterations: 4
```

	odds	2.5 %	97.5 %
(Intercept)	0.1594359	0.1298345	0.1953105
age	1.0101404	1.0066698	1.0136323
sclass2	1.5581188	1.3832083	1.7546817
overlim	0.7745152	0.6767372	0.8853500
factor(smoker)2	1.3444378	1.1744853	1.5384722
factor(smoker)3	0.9423218	0.8099333	1.0949705

After adjusting for age, social class and smoking, men who exceed the recommended alcohol limit are 1.16 times more likely to be obese compared to those who don't exceed the recommended alcohol limit 95%CI 1.02; 1.32 $p<0.05$. For women, we find the inverse association for alcohol (OR 0.77, 95%CI 0.68; 0.88 $p<0.05$).

After adjusting for all the other variables men who are current smokers are less likely to be obese than those who never smoked (OR 0.64 95%CI: 0.53; 0.76, $p<0.05$), for women the difference between current smokers and those who never smoked is not-statistically significant (OR 0.94 95%CI: 0.81; 1.09, $p=0.440$).