Question 1 (Suhana)

Ahmad Firdaus Mohamed

2024-02-10

1. A dataset(data\_my.csv) contains these variables:  (SUHANA) (?data diabetes.dta)

-id : unique identification number

-mswho : status of metabolic syndrome WHO (yes/no) (change outcome to dmdx)

-age : age in years

-gender : gender of each respondent (male/female)

-hba1c : hba1c in percentage (%)

-msbp : mean BP in mmHg

-hdl : high density lipoprotein (mmol/l)

-ldl : low density lipoprotein (mmol/l)

-mogtt2h : 2 hours post prandial glucose (mmol/l)

## a) Analyze the relationship between variable ‘dmdx’ and these covariates, ‘age’,’gender’,’hba1c’ and   ‘msbp’. Name the model as ‘mod.metab’. Write the estimated regression parameters in a table based on ‘mod.metab’.

Interpret the parameters. (20 marks)

mod.metab <- glm(dmdx~ age + gender + hba1c + msbpr, family = binomial(link = 'logit'), data = Data1)  
summary(mod.metab)

##   
## Call:  
## glm(formula = dmdx ~ age + gender + hba1c + msbpr, family = binomial(link = "logit"),   
## data = Data1)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -12.003283 0.541290 -22.175 <2e-16 \*\*\*  
## age 0.051038 0.005797 8.804 <2e-16 \*\*\*  
## gendermale 0.120980 0.136531 0.886 0.376   
## hba1c 1.109134 0.050190 22.099 <2e-16 \*\*\*  
## msbpr 0.002134 0.003117 0.685 0.494   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2797.7 on 3819 degrees of freedom  
## Residual deviance: 1655.7 on 3815 degrees of freedom  
## AIC: 1665.7  
##   
## Number of Fisher Scoring iterations: 6

tidy(mod.metab, conf.int = T)

## # A tibble: 5 × 7  
## term estimate std.error statistic p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) -12.0 0.541 -22.2 5.94e-109 -13.1 -11.0   
## 2 age 0.0510 0.00580 8.80 1.32e- 18 0.0398 0.0625   
## 3 gendermale 0.121 0.137 0.886 3.76e- 1 -0.148 0.387   
## 4 hba1c 1.11 0.0502 22.1 3.25e-108 1.01 1.21   
## 5 msbpr 0.00213 0.00312 0.685 4.94e- 1 -0.00400 0.00823

Table.mod.metab <- tbl\_regression(mod.metab)  
Table.mod.metab

## Table printed with {flextable}, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

Table 1 Multiple Logistic regression without interaction (n=3820)

| **Characteristic** | **log(OR)**1 | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| reported age (years) | 0.05 | 0.04, 0.06 | <0.001 |
| gender |  |  |  |
| female | — | — |  |
| male | 0.12 | -0.15, 0.39 | 0.4 |
| mmol/l | 1.1 | 1.0, 1.2 | <0.001 |
| mean SBP for right arm (mmHG) | 0.00 | 0.00, 0.01 | 0.5 |
| 1OR = Odds Ratio, CI = Confidence Interval  Intercept = -12.0032  Hosmer-Lemeshow – not checked  AUROC – not checked | | | |

Logit (Y) = B0 + B1X1 + B2X2 + B3X3 + B4X4 + Error

Logit (dmdx) = -12.0032 + 0.051-3 (age) + 0.1209 (gendermale) + 1.1091 (hba1c + 0.0021(msbpr)

1. With every increase 1 year in age, the log odd for dmdx changes by 0.051 (95% CI = 0.0397, 0.0625), when adjusted for gender, hba1c and msbpr
2. Being male changes the log odd of dmdx by 0.1209 (95% CI = -0.1483, 0.3873), when adjusted for age, hba1c and msbpr
3. With every increase of 1mmol hba1c , the log odd for dmdx changes by 1.1091 (95% CI = 1.0133, 1.2101 ), when adjusted for gender, age and msbpr
4. With every increase of 1mmhg msbpr , the log odd for dmdx changes by 0.0021 (95% CI = -0.0040, 0.0082 ), when adjusted for gender, age and hba1c

## b) Add an interaction term between variable ‘age’ and ‘gender’ to the model ‘mod.metab’.

Run the analysis and name the model as ‘mod.metab.int’.

Write the estimated regression parameters in a table based on ‘mod.metab.int’.    (10 marks)

mod.metab.int <- glm(dmdx~ age + gender + hba1c + msbpr + age: gender, family = binomial(link = 'logit'), data = Data1)  
summary(mod.metab.int)

##   
## Call:  
## glm(formula = dmdx ~ age + gender + hba1c + msbpr + age:gender,   
## family = binomial(link = "logit"), data = Data1)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -11.868974 0.584790 -20.296 < 2e-16 \*\*\*  
## age 0.048543 0.007179 6.761 1.37e-11 \*\*\*  
## gendermale -0.244482 0.640133 -0.382 0.703   
## hba1c 1.109413 0.050236 22.084 < 2e-16 \*\*\*  
## msbpr 0.002163 0.003119 0.694 0.488   
## age:gendermale 0.006370 0.010885 0.585 0.558   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 2797.7 on 3819 degrees of freedom  
## Residual deviance: 1655.4 on 3814 degrees of freedom  
## AIC: 1667.4  
##   
## Number of Fisher Scoring iterations: 6

Table.mod.metab.int <- tbl\_regression(mod.metab.int)  
Table.mod.metab.int

## Table printed with {flextable}, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

Table 2 Multiple Logistic Regression with Interaction (n=3820)

| **Characteristic** | **log(OR)**1 | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| reported age (years) | 0.05 | 0.03, 0.06 | <0.001 |
| gender |  |  |  |
| female | — | — |  |
| male | -0.24 | -1.5, 1.0 | 0.7 |
| mmol/l | 1.1 | 1.0, 1.2 | <0.001 |
| mean SBP for right arm (mmHG) | 0.00 | 0.00, 0.01 | 0.5 |
| reported age (years) \* gender |  |  |  |
| reported age (years) \* male | 0.01 | -0.01, 0.03 | 0.6 |
| 1OR = Odds Ratio, CI = Confidence Interval  Intercept = -11.8689  Hosmer-Lemeshow – not checked  AUROC – not checked | | | |

## c) Compare model ‘mod.metab.int’ and ‘mod.metab’ using an appropriate test and interpret the result. (10 marks)

anova(mod.metab, mod.metab.int, test = 'Chisq')

## Analysis of Deviance Table  
##   
## Model 1: dmdx ~ age + gender + hba1c + msbpr  
## Model 2: dmdx ~ age + gender + hba1c + msbpr + age:gender  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1 3815 1655.7   
## 2 3814 1655.4 1 0.34377 0.5577

When the significant level set at 0.05, there is no difference between the 2 models