PracticeSSMLR

Ahmad Firdaus Mohamed

2024-02-10

# Modelling Linear Regression

Using the same dataset (data\_my.csv), analyze the relationship between variable ‘hdl’ and these covariates, ‘ldl’,’mogtt2h’,’gender’. Name the model as Model A.  
a) Write the regression model and regression estimate. (10 marks)

1. Interpret the variances between regression model and regression estimates (10 marks)
2. Add on interaction between ‘mogtt2h’ and ‘gender’ to the model and name the model  as Model B.

 Compare model A and model B using appropriate test and interpret the findings. (10 marks)

1. Predict the mean hdl level in a population with ldl = 3.5, mogtt2h = 7.0, gender = male according the Model A and Model B. (10 marks)

# Prepare Environment

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ lubridate 1.9.3 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.0  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

## Read data

Using the same dataset (data\_my.csv), analyze the relationship between variable ‘hdl’ and these covariates, ‘ldl’,’mogtt2h’,’gender’. Name the model as Model A.  
a) Write the regression model and regression estimate. (10 marks)

Model.A <- lm(hdl~ldl + mogtt2h + gender, data = Data1)  
summary(Model.A)

##   
## Call:  
## lm(formula = hdl ~ ldl + mogtt2h + gender, data = Data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.23360 -0.22664 -0.03627 0.17738 2.94062   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.117167 0.020464 54.593 < 2e-16 \*\*\*  
## ldl 0.096612 0.005117 18.882 < 2e-16 \*\*\*  
## mogtt2h -0.005839 0.001527 -3.823 0.000134 \*\*\*  
## gendermale -0.185443 0.011663 -15.900 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3372 on 3725 degrees of freedom  
## (611 observations deleted due to missingness)  
## Multiple R-squared: 0.1386, Adjusted R-squared: 0.138   
## F-statistic: 199.9 on 3 and 3725 DF, p-value: < 2.2e-16

tidy(Model.A, conf.int = T)

## # A tibble: 4 × 7  
## term estimate std.error statistic p.value conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 1.12 0.0205 54.6 0 1.08 1.16   
## 2 ldl 0.0966 0.00512 18.9 5.16e-76 0.0866 0.107   
## 3 mogtt2h -0.00584 0.00153 -3.82 1.34e- 4 -0.00883 -0.00284  
## 4 gendermale -0.185 0.0117 -15.9 3.97e-55 -0.208 -0.163

Regression Model Y= B0 + B1X1 + B2X2 + B3X + E

Regression Estimate

hdl = 1.1171 + 0.0966(ldl) - 0.0058(mogtt2h)- 0.18544(gendermale)

1. Interpret the variances between regression model and regression estimates (10 marks)
2. with every increase of 1mmol ldl, hdl changes by 0.0966 (95% CI = 0.0865. 0.1066) , when mogtt2h and gender are adjusted
3. with every increase of 1 mmol moggt2h, hdl changes by - 0.0058 (95% CI = -0.0088, -0.0028) when ldl and gender are adjusted
4. being male, the hdl changes by -0.1854 (95% CI = -0.2083, -0.1625) compared to female when ldl and mogtt2h are adjusted
5. Add on interaction between ‘mogtt2h’ and ‘gender’ to the model and name the model  as Model B.

 Compare model A and model B using appropriate test and interpret the findings. (10 marks)

Model.B <- lm(hdl~ldl + mogtt2h + gender + mogtt2h:gender, data = Data1)  
summary(Model.B)

##   
## Call:  
## lm(formula = hdl ~ ldl + mogtt2h + gender + mogtt2h:gender, data = Data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.23246 -0.22538 -0.03637 0.17771 2.95019   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.128000 0.021901 51.504 < 2e-16 \*\*\*  
## ldl 0.096877 0.005120 18.923 < 2e-16 \*\*\*  
## mogtt2h -0.007383 0.001890 -3.907 9.53e-05 \*\*\*  
## gendermale -0.216319 0.025132 -8.607 < 2e-16 \*\*\*  
## mogtt2h:gendermale 0.004340 0.003130 1.387 0.166   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3371 on 3724 degrees of freedom  
## (611 observations deleted due to missingness)  
## Multiple R-squared: 0.1391, Adjusted R-squared: 0.1382   
## F-statistic: 150.4 on 4 and 3724 DF, p-value: < 2.2e-16

anova (Model.A, Model.B)

## Analysis of Variance Table  
##   
## Model 1: hdl ~ ldl + mogtt2h + gender  
## Model 2: hdl ~ ldl + mogtt2h + gender + mogtt2h:gender  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 3725 423.50   
## 2 3724 423.28 1 0.21864 1.9235 0.1656

There is no significant difference between Model.A and Model.B when the significant value is set at 0.05

1. Predict the mean hdl level in a population with ldl = 3.5, mogtt2h = 7.0, gender = male according the Model A and Model B. (10 marks)

Expand Grind Method

new\_data <- expand.grid (ldl=3.5, mogtt2h=7.0, gender = "male")

For model A

Predict.Model.A <- predict(Model.A, newdata = new\_data)  
summary(Predict.Model.A)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.229 1.229 1.229 1.229 1.229 1.229

using equation

hdl = 1.1171 + 0.0966(ldl) - 0.0058(mogtt2h)- 0.18544(gendermale)

1.1171 + (0.0966\*3.5) - (0.0058\*7) - 0.18544

## [1] 1.22916

For Model B

Expand Grid Method

Predict.Model.B <- predict(Model.B, newdata = new\_data)  
Predict.Model.B

## 1   
## 1.229449

using equation

summary(Model.B)

##   
## Call:  
## lm(formula = hdl ~ ldl + mogtt2h + gender + mogtt2h:gender, data = Data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.23246 -0.22538 -0.03637 0.17771 2.95019   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.128000 0.021901 51.504 < 2e-16 \*\*\*  
## ldl 0.096877 0.005120 18.923 < 2e-16 \*\*\*  
## mogtt2h -0.007383 0.001890 -3.907 9.53e-05 \*\*\*  
## gendermale -0.216319 0.025132 -8.607 < 2e-16 \*\*\*  
## mogtt2h:gendermale 0.004340 0.003130 1.387 0.166   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3371 on 3724 degrees of freedom  
## (611 observations deleted due to missingness)  
## Multiple R-squared: 0.1391, Adjusted R-squared: 0.1382   
## F-statistic: 150.4 on 4 and 3724 DF, p-value: < 2.2e-16

ldl = 1.128 + 0.0968 (ldl) - 0.0073(mogtt2h) - 0.2163 (gendermale) + 0.0043 (mogtt2h:gendermale)

1.128 + (0.0968\*3.5) - (0.0073\*7) - 0.2163 + (0.0043\*7)

## [1] 1.2295

TableModelA <- tbl\_regression(Model.A)  
TableModelA

## 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.

| **Characteristic** | **Beta** | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| ldl | 0.10 | 0.09, 0.11 | <0.001 |
| mogtt2h | -0.01 | -0.01, 0.00 | <0.001 |
| gender |  |  |  |
| female | — | — |  |
| male | -0.19 | -0.21, -0.16 | <0.001 |
| 1CI = Confidence Interval | | | |

TableModelB <- tbl\_regression(Model.B)  
TableModelB

## 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.

| **Characteristic** | **Beta** | **95% CI**1 | **p-value** |
| --- | --- | --- | --- |
| ldl | 0.10 | 0.09, 0.11 | <0.001 |
| mogtt2h | -0.01 | -0.01, 0.00 | <0.001 |
| gender |  |  |  |
| female | — | — |  |
| male | -0.22 | -0.27, -0.17 | <0.001 |
| mogtt2h \* gender |  |  |  |
| mogtt2h \* male | 0.00 | 0.00, 0.01 | 0.2 |
| 1CI = Confidence Interval | | | |