> # load packages

> library(readxl)

> library(dplyr)

载入程辑包：‘dplyr’

The following objects are masked from ‘package:stats’:

filter, lag

The following objects are masked from ‘package:base’:

intersect, setdiff, setequal, union

> # import .xlsx data

> excel\_sheets("Desktop/regression/bloodtest.xls")

[1] "斑块成分和血压血脂" "Sheet1" "Sheet2"

> bloodtest <- read\_excel("Desktop/regression/bloodtest.xls", sheet = "斑块成分和血压血脂")

New names:

\* 用药 -> 用药...21

\* 用药 -> 用药...24

\* `血糖(4.16-6.44)` -> `血糖(4.16-6.44)...25`

\* `血糖(4.16-6.44)` -> `血糖(4.16-6.44)...34`

> View(bloodtest)

> # check classification of variable

> str(bloodtest)

tibble [300 × 55] (S3: tbl\_df/tbl/data.frame)

$ 检查日期 : chr [1:300] "2015-10-10" "2015-10-10" "2015-10-16" "2015-10-21" ...

$ 姓名 : chr [1:300] "尚君明" "张瑞萍" "刘晋军" "李春凤" ...

$ 年龄2 : num [1:300] 46 42 46 40 40 53 66 63 74 27 ...

$ 性别2为男性1为女性 : num [1:300] 2 1 2 2 2 2 2 2 1 1 ...

$ 年龄 : num [1:300] 46 42 46 40 40 53 66 63 74 27 ...

$ 身高m : chr [1:300] "1.72" "1.65" "1.68" "1.7" ...

$ 体重KG : num [1:300] 31.4 60.4 62.6 64.1 65.4 ...

$ 教育程度 : chr [1:300] "初" "高" "初" "高" ...

$ 婚姻 : chr [1:300] "已婚" "已婚" "已婚" "已婚" ...

$ 收入 : num [1:300] 5000 5000 5000 5000 5000 5000 5000 5000 5000 5000 ...

$ 吸烟 : chr [1:300] "不吸" NA NA NA ...

$ 饮酒 : chr [1:300] "偶尔" NA NA NA ...

$ 运动 : chr [1:300] "每天一次" "每天一次" "每天一次" "每天一次" ...

$ 饮食 : chr [1:300] "规律" "规律" "规律" "规律" ...

$ 睡眠 : chr [1:300] "好" "好" "好" "好" ...

$ BMI体重指数 : num [1:300] 10.6 22.2 22.2 22.2 22.1 ...

$ 电话 : chr [1:300] "0" "0" "0" "0" ...

$ 高血压 : num [1:300] 1 1 0 0 0 0 1 1 1 0 ...

$ 收缩压SBP : num [1:300] 165 160 126 128 136 138 150 149 158 148 ...

$ 舒张压 : num [1:300] 86 76 78 78 72 82 86 76 74 76 ...

$ 用药...21 : chr [1:300] "普萘洛尔" "倍他乐克" NA NA ...

$ 冠心病 : num [1:300] 0 0 0 0 0 0 0 0 0 0 ...

$ 糖尿病 : num [1:300] 0 0 0 0 0 0 1 0 0 0 ...

$ 用药...24 : chr [1:300] NA NA NA NA ...

$ 血糖(4.16-6.44)...25 : chr [1:300] NA NA NA NA ...

$ 高血脂 : num [1:300] 0 0 0 0 0 1 1 1 1 0 ...

$ 诊断意见 : chr [1:300] "双肺未见明显异常改变。 \n 左冠状动脉前降支近段可见非钙化斑块，管腔狭窄约50%左右。\n 头"| \_\_truncated\_\_ "CT冠状动脉造影未见明显异常。\n 右肺上叶钙化灶。 \n 肝右叶点状钙化灶。\n 头颈部动脉CTA未见明显异"| \_\_truncated\_\_ "左冠状动脉前降支近中段非钙化为主混合斑块，管腔狭窄约25%左右。\n 头颈部CTA未见明显异常。\n 请结合临床。\n" "冠状动脉CTA未见明显异常。\n 左侧颈总动脉、锁骨下动脉、椎动脉起源异常。\n 请结合临床。\n" ...

$ 心脏评分 : num [1:300] 2 1 1 1 1 1 1 2 1 1 ...

$ 头颈评分 : num [1:300] 1 1 1 1 1 1 1 1 1 1 ...

$ 非稳定非钙化 : num [1:300] 1 0 1 0 0 1 0 1 1 0 ...

$ 稳定钙化 : num [1:300] 0 0 0 0 0 0 0 0 0 0 ...

$ 狭窄 : num [1:300] 0 0 2 0 0 0 0 2 1 0 ...

$ 支架 : num [1:300] 0 0 0 0 0 0 1 0 0 0 ...

$ 血糖(4.16-6.44)...34 : chr [1:300] NA NA NA NA ...

$ 甘油三脂TG（0.45-1.81） : num [1:300] NA NA NA NA 2.88 2.6 3.94 4.2 4.73 0.68 ...

$ 总胆固醇TCH（2.9-5.68） : num [1:300] NA NA NA NA 5.54 6.12 5.9 7.21 6.75 3.56 ...

$ 高密度脂蛋白胆固醇（0.9-2.19） : num [1:300] NA NA NA NA 1.24 1.04 1.26 1.2 1.21 1.22 ...

$ 低密度脂蛋白胆固醇LDL（0.00-3.36）: num [1:300] NA NA NA NA 3.44 3.9 1.77 1.38 2.45 2.68 ...

$ 尿素（1.7-8.3） : num [1:300] NA NA NA NA 1.82 2.36 3.32 3.46 4.21 4.76 ...

$ 肌酐（40-106） : num [1:300] NA NA NA NA 68 67 45 58 48 47 ...

$ CK-MB(0.02-0.13)ug/L : num [1:300] NA NA NA NA 0.75 0.88 0.83 0.9 0.94 0.98 ...

$ LDH : num [1:300] NA NA NA NA NA NA NA NA NA NA ...

$ 糖化血红蛋白c : num [1:300] NA NA NA NA 4.9 5.2 NA 5.5 5.8 NA ...

$ HbA1a : num [1:300] NA NA NA NA 1.3 1.2 NA 1.1 1 NA ...

$ HbA1b : num [1:300] NA NA NA NA 0.9 0.9 NA 0.9 0.9 NA ...

$ Cysc（0-1.35） : num [1:300] NA NA NA NA NA NA NA NA NA NA ...

$ ALT : logi [1:300] NA NA NA NA NA NA ...

$ AST : logi [1:300] NA NA NA NA NA NA ...

$ B2微球蛋白 : num [1:300] NA NA NA NA NA NA NA NA NA NA ...

$ 氧化低密度脂蛋白 : logi [1:300] NA NA NA NA NA NA ...

$ 人脂蛋白 : logi [1:300] NA NA NA NA NA NA ...

$ 白蛋白 : logi [1:300] NA NA NA NA NA NA ...

$ 球蛋白 : logi [1:300] NA NA NA NA NA NA ...

$ 影像号 : logi [1:300] NA NA NA NA NA NA ...

$ 出生日期 : logi [1:300] NA NA NA NA NA NA ...

> # select variable

> bloodtest <- bloodtest %>% select(c("血糖(4.16-6.44)...34","甘油三脂TG（0.45-1.81）","总胆固醇TCH（2.9-5.68）","高密度脂蛋白胆固醇（0.9-2.19）","低密度脂蛋白胆固醇LDL（0.00-3.36）","收缩压SBP","舒张压"))

> # remove NA value

> bloodtest <- na.omit(bloodtest)

> # change column names

> colnames(bloodtest) <- c("FG","TG","TC","HDL-C","LDL-C", "SBP","DBP")

> # check classification of variable

> str(bloodtest)

'data.frame': 262 obs. of 7 variables:

$ FG : num 5.89 5.78 7.82 4.38 4.38 5.69 6.72 6.02 6.8 4.5 ...

$ TG : num 2.88 2.6 3.94 4.2 4.73 0.68 0.72 1.12 1.21 1.32 ...

$ TC : num 5.54 6.12 5.9 7.21 6.75 3.56 4.15 5.89 5.89 5.38 ...

$ HDL-C: num 1.24 1.04 1.26 1.2 1.21 1.22 1.23 1.24 1.92 1.26 ...

$ LDL-C: num 3.44 3.9 1.77 1.38 2.45 2.68 2.19 2.45 2.67 2.85 ...

$ SBP : num 136 138 150 149 158 148 136 136 148 135 ...

$ DBP : num 72 82 86 76 74 76 78 78 86 82 ...

- attr(\*, "na.action")= 'omit' Named int [1:38] 1 2 3 4 36 37 38 39 40 41 ...

..- attr(\*, "names")= chr [1:38] "1" "2" "3" "4" ...

> # switch FG to numeric variable

> bloodtest$FG <- as.numeric(bloodtest$FG)

> # check classification again

> str(bloodtest)

'data.frame': 262 obs. of 7 variables:

$ FG : num 5.89 5.78 7.82 4.38 4.38 5.69 6.72 6.02 6.8 4.5 ...

$ TG : num 2.88 2.6 3.94 4.2 4.73 0.68 0.72 1.12 1.21 1.32 ...

$ TC : num 5.54 6.12 5.9 7.21 6.75 3.56 4.15 5.89 5.89 5.38 ...

$ HDL-C: num 1.24 1.04 1.26 1.2 1.21 1.22 1.23 1.24 1.92 1.26 ...

$ LDL-C: num 3.44 3.9 1.77 1.38 2.45 2.68 2.19 2.45 2.67 2.85 ...

$ SBP : num 136 138 150 149 158 148 136 136 148 135 ...

$ DBP : num 72 82 86 76 74 76 78 78 86 82 ...

- attr(\*, "na.action")= 'omit' Named int [1:38] 1 2 3 4 36 37 38 39 40 41 ...

..- attr(\*, "names")= chr [1:38] "1" "2" "3" "4" ...

> # build model

> model\_lm <- lm(FG~., data = bloodtest)

> summary(model\_lm)

Call:

lm(formula = FG ~ ., data = bloodtest)

Residuals:

Min 1Q Median 3Q Max

-2.1493 -0.8115 -0.1042 0.6003 3.3438

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.739111 0.852417 3.213 0.00148 \*\*

TG -0.154495 0.069815 -2.213 0.02779 \*

TC 0.008878 0.037991 0.234 0.81542

`HDL-C` 0.191392 0.090579 2.113 0.03558 \*

`LDL-C` 0.169340 0.102243 1.656 0.09891 .

SBP 0.025106 0.004731 5.306 2.44e-07 \*\*\*

DBP -0.010273 0.009816 -1.047 0.29632

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.094 on 254 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.1451, Adjusted R-squared: 0.1249

F-statistic: 7.187 on 6 and 254 DF, p-value: 4.277e-07

> tidy(model\_lm)

# A tibble: 7 × 5

term estimate std.error statistic p.value

<chr> <dbl> <dbl> <dbl> <dbl>

1 (Intercept) 2.74 0.852 3.21 0.00148

2 TG -0.154 0.0698 -2.21 0.0278

3 TC 0.00888 0.0380 0.234 0.815

4 `HDL-C` 0.191 0.0906 2.11 0.0356

5 `LDL-C` 0.169 0.102 1.66 0.0989

6 SBP 0.0251 0.00473 5.31 0.000000244

7 DBP -0.0103 0.00982 -1.05 0.296

> confint(model\_lm)

2.5 % 97.5 %

(Intercept) 1.06040540 4.417816270

TG -0.29198529 -0.017004078

TC -0.06593983 0.083695290

`HDL-C` 0.01300976 0.369773252

`LDL-C` -0.03201227 0.370692301

SBP 0.01578834 0.034423945

DBP -0.02960409 0.009058777