**实验报告**

|  |  |  |
| --- | --- | --- |
| **课程名称**： **应用回归分析实验** | **学期： 2023年秋季学期** | **成绩**： |
| **指导教师**： 陈丹 | **学生姓名**： **枫叶** | **学生学号**： |
| **实验名称**： **logistic回归** | | |
| **实验编号**：8 | **实验日期**： | **实验学时**： |
| **学院： 数学与统计学院** | **专业： 统计学** | **年级**： **2021级** |

**一、实验目的**

掌握logistic回归

**二、使用环境**

R4.3.1

**三、实验内容**

**糖尿病概率预测**

1. **算法介绍及结果**

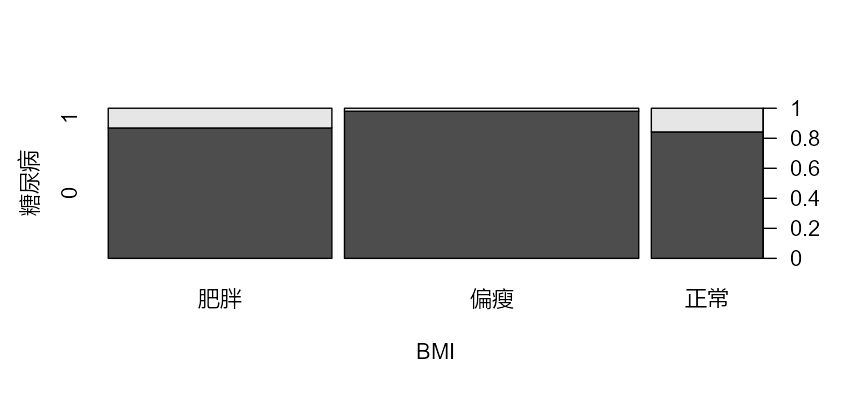


图 1

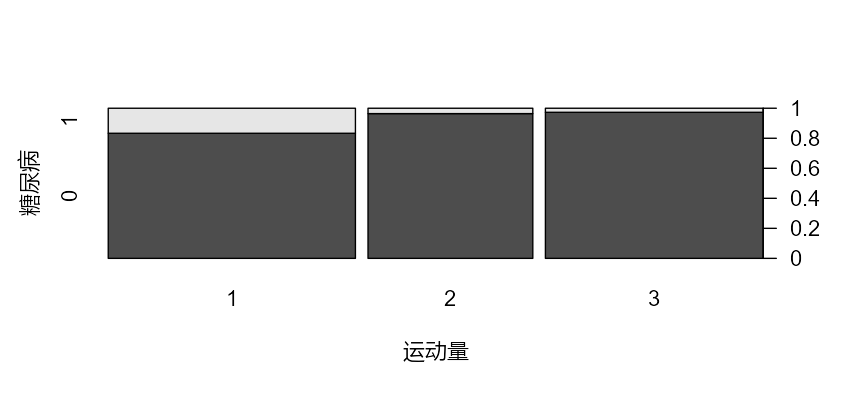


图 2

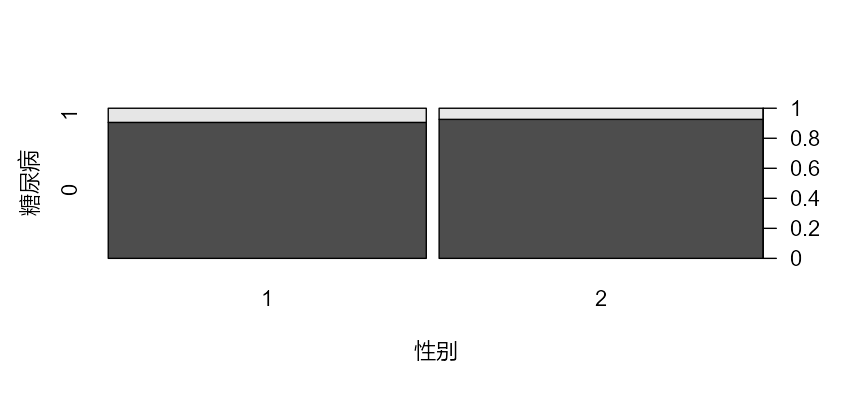


图 3

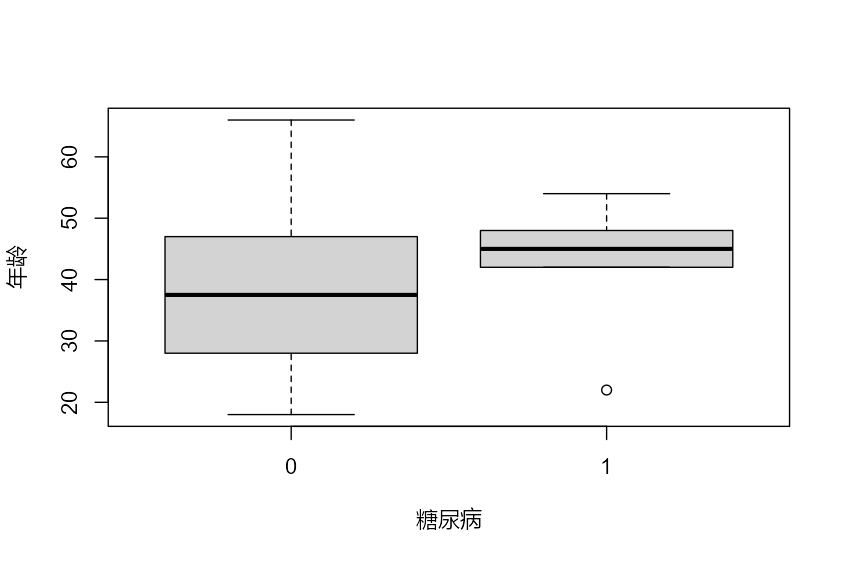


图 4

从棘状图与箱线图可以看出，BMI偏瘦以上的、运动量少的群体中有更多糖尿病患者，男女患糖尿病的占比差异较不明显，患糖尿病的群体年龄中位数要高于不患糖尿病的

表1给出了全部变量的回归结果（1）和逐步回归结果（2），可以看到经变量筛选后只有运动量的类别3显著，且系数为负，这可能表明需要达到一定的运动量才能起到防止患糖尿病的作用。模型（2）的LRT检验p值为0.04708，方程显著。

表 1

|  | （1） | （2） |
| --- | --- | --- |
| (Intercept) | -2.885 | -1.609\*\*\* |
|  | (1.758) | (0.414) |
| 年龄 | 0.036 |  |
|  | (0.036) |  |
| 运动量2 | -1.845 | -1.686 |
|  | (1.195) | (1.099) |
| 运动量3 | -1.997+ | -1.974+ |
|  | (1.183) | (1.095) |
| 性别2 | 0.550 |  |
|  | (0.837) |  |
| BMI偏瘦 | -1.703 |  |
|  | (1.179) |  |
| BMI正常 | -0.044 |  |
|  | (0.840) |  |
| Num.Obs. | 107 | 107 |
| AIC | 64.6 | 61.7 |
| BIC | 83.4 | 69.7 |
| Log.Lik. | -25.324 | -27.835 |
| F | 1.194 | 2.458 |
| RMSE | 0.26 | 0.27 |
| + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001 | | |

模型（2）的混淆矩阵如下表所示，ROC曲线如图5所示，模型具有一定预测能力，但性能较差

表 2

| 0 | 1 |
| --- | --- |
| 63 | 35 |
| 2 | 7 |

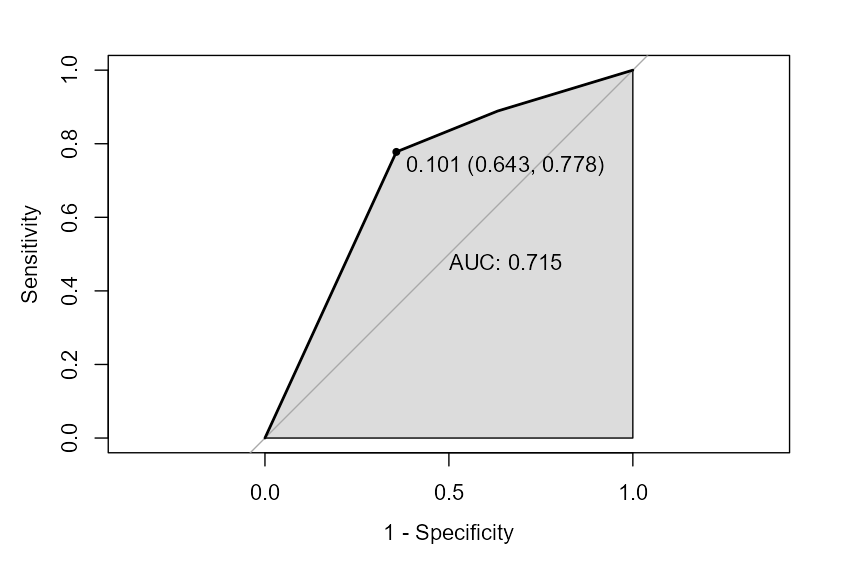


图 5

代码

|  |
| --- |
| library(readxl)  library(modelsummary)  library(flextable)  library(dplyr)  library(pROC)  library(vcd)  data <- read\_xlsx("D:/预删除文件夹/大三上/应用回归分析/实验8数据.xlsx") %>%  mutate(across(-"年龄",as.factor))  #描述统计  spine(糖尿病~BMI,data)  spine(糖尿病~运动量,data)  spine(糖尿病~性别,data)  boxplot(年龄~糖尿病,data)  #logistic回归  model0 <- glm(data = Boston,糖尿病~.,family = binomial(link = "logit"))  model1 <- step(model0)  modelsummary(list(model0,model1),output = "flextable",stars = T) %>%  align(align = "center",part = "all") %>%  save\_as\_docx(path = "D:/预删除文件夹/大三上/应用回归分析/临时.docx")  anova(model1,test = "LRT")  #模型评估  y\_prob <- model1$fitted.values  yhat <- 1\*(y\_prob>9/107)  as.data.frame.matrix(table(data$糖尿病,yhat)) %>%  flextable() %>%  align(align = "center",part = "all") %>%  save\_as\_docx(path = "D:/预删除文件夹/大三上/应用回归分析/临时.docx")  myroc <- roc(data$糖尿病,y\_prob,algorithm=2,)  plot(myroc,print.auc = T,auc.polygon=TRUE,print.thres=TRUE,legacy.axes=T) |

**六、参考文献**

**七、教师评语**