

Module 4: Regression Methods: Concepts and Applications

Lab 4: Logistic Regression and GLMs

The goal of this lab is to answer the following scientific questions using the cholesterol dataset.

- Is hypertension associated with rs174548?
- Is hypertension associated with triglycerides?
- Is hypertension associated with rs174548 after adjusting for triglyceride levels?

The cholesterol data set is available for download from the module Github repository and contains the following variables:

ID: Subject ID

sex: Sex: 0 = male, 1 = female

age: Age in years

chol: Serum total cholesterol, mg/dl

BMI: Body-mass index, kg/m²

TG: Serum triglycerides, mg/dl

APOE: Apolipoprotein E genotype, with six genotypes coded 1-6: 1 = e₂/e₂, 2 = e₂/e₃, 3 = e₂/e₄, 4 = e₃/e₃, 5 = e₃/e₄, 6 = e₄/e₄

rs174548: Candidate SNP 1 genotype, chromosome 11, physical position 61,327,924. Coded as the number of minor alleles: 0 = C/C, 1 = C/G, 2 = G/G.

rs4775401: Candidate SNP 2 genotype, chromosome 15, physical position 59,476,915. Coded as the number of minor alleles: 0 = C/C, 1 = C/T, 2 = T/T.

HTN: diagnosed hypertension: 0 = no, 1 = yes

chd: diagnosis of coronary heart disease: 0 = no, 1 = yes

You can download the data file and read it into R as follows:

```
cholesterol = read.csv("https://raw.githubusercontent.com/rhubb/SISG2018/master/data/SISG-D
ata-cholesterol.csv", header=T)
```

Install R packages

- For this lab you will need the *gee* and *lmtree* packages.
- If you have not already, install the packages first. You will then need to load the libraries each time you execute your R script.

```
install.packages("gee")
install.packages("lmtree")
library(gee)
library(lmtree)
```

Exercises

1. Is there an association between rs174548 and hypertension? Analyze this relationship using descriptive statistics as well as a logistic regression analysis.
2. Use logistic regression to investigate the association between triglycerides and hypertension. Interpret the results of this model. Make sure that you can interpret the model coefficients and hypothesis testing.
3. Analyze the association between hypertension and rs174548 adjusted for triglycerides using logistic regression. What does this model tell you about the association between rs174548 and hypertension? What role does triglycerides play in this analysis?
4. Use a GLM to estimate the relative risk of hypertension for patients with different rs174548 genotypes, adjusting for triglycerides. Make sure you can interpret the coefficients. How do these results compare to the results of the logistic regression analysis?
5. Use a GLM to estimate the risk difference for hypertension according to rs174548 genotypes, adjusting for triglycerides. Make sure you can interpret the coefficients. How do these results compare to the results of the logistic regression and relative risk regression analyses?

Once your group has completed the lab exercises, please submit your R script file to the class Github repository:

<https://github.com/rhubb/SISG2018/tree/master/submit> (<https://github.com/rhubb/SISG2018/tree/master/submit>)

Sign in using the class username and password. Then click upload files to save your R script file to the repository.