

This is an analysis note of the association analysis for Dr. Steven Kleeberger

Jianying Li

March 9th, 2018

Introduction

This exercise is to document how one can use R to perform logistic regression and further extract important estimates, as well as computing the odds ratios and construct the confident interval for the odds ratios.

Author: Jianying Li Purpose: to analyze the association of the genetic effect of a functional/coding SNP “rs2257167” on the disease phenotypes (severe vs. mild), and generate odds ratio and confident interval

Association analysis on disease severity against GoldenGate SNP panels with several potential covariates.

In this analysis, we tested three hypotheses each with different genetic assumptions (genetic mode): case I, II, and III. All the analyses will be done with each case respectively. Here, we are assuming that allele “A” is the reference and “B” is the variant (rs2257167). Here are those three cases as following:

Table 1: Genotype case I

Disease-level	AA	AB	BB
Severe	SNPCnt	SNPCnt	SNPCnt
Mild	SNPCnt	SNPCnt	SNPCnt

Table 2: Genotype case II

Disease-level	AA	AB/BB
Severe	SNPCnt	SNPCnt
Mild	SNPCnt	SNPCnt

Table 3: Genotype case III

Disease-level	AA/AB	BB
Severe	SNPCnt	SNPCnt
Mild	SNPCnt	SNPCnt

With the proposed three cases:

Case I: AA, AB, BB. With all three genotypes captured, we are testing the additive genetic effect.

Case II: AA, AB/BB. With collapsing any genotype with “B” allele, we are testing the dominate effect from the variant allele.

Case III: AA/AB, BB. With collapsing any genotype with “A” allele, we are testing the recessive effect from the variant allele.

In this study, we have recruited 748 infants with lung disease clinically diagnosed as “severe” or “mild”. The infants were tested for RSV infection when administered. This research focuses on the disease status as our major outcome with the infants tested for RSV positive, the question is whether such a coding SNP (rs2257167) is associated with the disease outcome. The ultimate question behind this research is when the kid is infected with RSV and further develops the lung disease, what is the effect their genetic background on their clinical symptom. In addition to the clinical diagnosis, we also collected other information, i.e. gender, social economical status etc. Those factors were each tested in the model as a covariate. Therefore, we are able to address the hidden impact from other covariates that could interact and affect their clinical symptom.

Accessing the minor allele frequency in this population

Getting the data

Allele frequency table

Table 4: Minor allele frequency – rs2257167

	Reference allele (G) frequency	Variant allele (C) frequency
ALL individuals	0.7313433	0.2686567
Individual with RSV positive	0.7319202	0.2680798

Logistic regression analysis on three cases without covariate

Case I

Call: glm(formula = m[, 2] ~ m[, 1], family = binomial(“logit”))

Deviance Residuals: Min 1Q Median 3Q Max
-1.4705 -1.2474 0.9103 0.9103 1.3250

Coefficients: Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.6669 0.1369 4.873 1.1e-06 * m[, 1] **-0.5037 0.1598 -3.153 0.00162** — Signif. codes: 0 ‘
0.001 ’ 0.01 ’ 0.05 ‘ 0.1 ‘ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 541.03 on 400 degrees of freedom

Residual deviance: 530.91 on 399 degrees of freedom AIC: 534.91

Number of Fisher Scoring iterations: 4

Call: glm(formula = m[, 2] ~ as.factor(m[, 1]), family = binomial("logit"))

Deviance Residuals: Min 1Q Median 3Q Max
-1.4549 -1.2881 0.9233 0.9233 1.4224

Coefficients: Estimate Std. Error z value Pr(>|z|)

(Intercept) 0.6321 0.1420 4.453 8.47e-06 * **as.factor(m[, 1])1 -0.3757 0.2178 -1.725 0.08456 .**
as.factor(m[, 1])2 -1.1917 0.3887 -3.066 0.00217 — Signif. codes: 0 ‘**0.001**’ ‘**0.01**’ ‘0.05’ ‘0.1’ ‘1’

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 541.03 on 400 degrees of freedom

Residual deviance: 530.16 on 398 degrees of freedom AIC: 536.16

Number of Fisher Scoring iterations: 4

Case II

Case III

Logistic regression analysis on three cases with covariate

Analysis with covariates: gender

Analysis with covariates: region

Analysis with covariates: breastfeeding

Analysis with covariates: social-status