GENETIC-CORRELATIONS-IN-LDSC

Michel Nivard

GENETIC CORRELATION IN LD SCORE REGRESSION

This lecture:

- the LDSC genetic correlation estimator
- Making sure the LD score you use, are appropriate for the cohort/GWAS you have,
- The Practical!

GET THE DATA FOR THE PRACTICAL

Lets start grabbing the data, its a bit more data then some of the other practicals.

Code is in the forum ("ISG Workshop logisitcs"), and in the shared files

ISGW Website International Statistical Genetics Workshop Computing Environment Often called "The Boulder Twin Workshop," the ISGW has been teaching behavioral genetics and other statistical genetics topics in Boulder, Colorado and 1987. This is the entry point for the computing environment used for the course. **Resources for registered participants** Rstudio Password change Shared file area Access Rstudio in the Access a command line in Click "Log In" and then Upload or download posters change your Workshop or presentations Workshop computing the Workshop computing environment. Or use your environment. password own SSH client and connect workshop.colorado.edu.

- 1 system("mkdir \$HOME/michel")
- 2 system("cp -R /faculty/michel/2024/practical* \$HOME/michel")
- 3 setwd("~/michel/practical")

GENETIC CORRELATION IN LD SCORE REGRESSION

Univariate LDSC:

Bivariate LDSC:

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., ... & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature genetics*, 47(11), 1236-1241.

LDSC AS A REGRESSION

Bivariate LDSC:

It's just a regression:

DATA

GWAS sumstats for traits 1 and 2 for SNP j:

Audience Q: "What is this Z?"

The LD score, a per SNP measure of the size of the part of the genome tagged by SNP j:



WORKS FOR GWAS THAT DO NOT OVERLAP!

BUT!

Its a regression where the "outcome" (GWAS results) and the "predictor" (LD score) come form different sources...

That's really weird... and it puts a little bit of responsibility on the user to ensure the variables are from a common population...

A LOOK AT VARIATION IN THE LD MAP



Salehi Nowbandegani, P., Wohns, A. W., Ballard, J. L., Lander, E. S., Bloemendal, A., Neale, B. M., & O'Connor, L. J. (2023). Extremely sparse models of linkage disequilibrium in ancestrally diverse association studies. *Nature Genetics*, *55*(9), 1494-1502.

SO WE WANT TO MATCH ANCESTRY IN THE LDSCORE TO THE GWAS SAMPLE

Use ancestry specific LD scores, for example:

- GNOMAD: https://gnomad.broadinstitute.org/downloads#v2-linkagedisequilibrium
- UKBiobank Pan ancestry: https://pan-dev.ukbb.broadinstitute.org/downloads

What if I still don't really think their appropriate for my sample?

- Make your won LD scores: https://github.com/bulik/ldsc/wiki/LD-Score-Estimation-Tutorial
- (You need sequencing data!)

CAN WE USE LDSC IN ADMIXED SAMPLES?

HOW CAN WE USE COV-LDSC IN ADMIXED SAMPLES?

- Use "covariate adjusted" LD scores in admixed samples.
- cov-LDSC software: https://github.com/immunogenomics/cov-ldsc

WHAT IF YOU WANT TO COMPUTE THE RG BETWEEN GWASS IN DIFFERENT ANCESTRIES?

- There is a tool called popcorn that models the differences in LD and AF and rg at the same time
- https://github.com/brielin/Popcorn

WHAT IF YOU WANT TO COMPUTE THE RG BETWEEN GWASS IN DIFFERENT ANCESTRIES?

- BUT it obviously also pulls in all cultural, cohort, measurement differences between your GWASs!
- Solutions: compute rg between traits, within an admixed person correlating the genetic effects betwene their two ancestries.
- Acessible explanation: https://www.nature.com/articles/s41588-023-01325-x
- Original paper: https://www.nature.com/articles/s41588-023-01338-6
- Software: https://github.com/KangchengHou/admix-kit

POPCORN VS ADMIX-KIT



LET'S LOOK AT THE PRACTICAL FOR TODAY!

- we have GWAS data and LD scores of MDD and BMI for people of European and east-Asian ancestry
- What is YOUR estimand? (what do YOU want to know?)
- We estimate the rg's using LDSC (estimator)
- We see if we can learn about your estimand using our estimates?