Practical Lecture 11: Using LDSC to estimate heritability and genetic correlations

Loic Yengo

2022-06-18

We provide summary statistics for two GWAS of Waist-to-Hip Ratio (WHR) in males and females conducted by the GIANT consortium.

Preparation

Create a directory where to run the practical and copy GWAS summary statistics there

```
mkdir prac11/
cd prac11
cp /data/module4/prac11/summary-statistics/GIANT_2015_WHR_*_EUR.txt .
```

Load LDSC software and required libraries (more details available on github: https://github.com/bulik/ldsc)

```
conda activate ldsc
```

Download LDSC files

```
git clone https://github.com/bulik/ldsc
```

Download LDSC scores calculated in European ancestries individuals from the 1000 Genomes Project

```
wget https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2
tar -xvf eur_w_ld_chr.tar.bz2
```

Format GWAS summary statistics for LDSC analysis

```
ldsc/munge_sumstats.py --sumstats GIANT_2015_WHR_FEMALES_EUR.txt \
  --merge-alleles eur_w_ld_chr/w_hm3.snplist --chunksize 1000000 \
  --out giant_whr_females
```

```
ldsc/munge_sumstats.py --sumstats GIANT_2015_WHR_MALES_EUR.txt \
  --merge-alleles eur_w_ld_chr/w_hm3.snplist --chunksize 1000000 \
  --out giant_whr_males
```

Part 1: Estimation of heritability using LDSC

Run the following commands

```
ldsc/ldsc.py --h2 giant_whr_females.sumstats.gz --ref-ld-chr eur_w_ld_chr/ \
  --w-ld-chr eur_w_ld_chr/ --out h2_giant_whr_females

ldsc/ldsc.py --h2 giant_whr_males.sumstats.gz --ref-ld-chr eur_w_ld_chr/ \
  --w-ld-chr eur_w_ld_chr/ --out h2_giant_whr_males
```

Question 1

  a) How many SNPs were used in each regression?

  b) What is the heritability of WHR in males and females?

  c) What is the LD score intercept? Is this expected? If not then what could explain this observation?

Part 2: Estimation of genetic correlation using LDSC

Run the following command

```
ldsc/ldsc.py --rg giant_whr_females.sumstats.gz,giant_whr_males.sumstats.gz \
  --ref-ld-chr eur_w_ld_chr/ --w-ld-chr eur_w_ld_chr/ --out rg_giant_whr
```

Question 2

  a) How many SNPs were used in each regression?

  b) What is the heritability of WHR in males and females? Is this different from your previous results? Rerun the LDSC commands from **Question 1** adding the following flag: `--two-step INFINITY`. What can you conclude?

  c) Interpret the bivariate LD score intercept.

  d) What can you conclude regarding the genetic architecture of WHR in males and females.

Part 3: Estimation of genetic correlation using LDSC

We now provide summary statistics for two GWAS of Waist-to-Hip Ratio (WHR) in males and females in the UK Biobank (UKB). The files are already formatted for LDSC.

Copy the new GWAS summary statistics

```
cp /data/module4/prac11/summary-statistics/ukb_whr_*.sumstats.gz .
```

Question 3

  a) Estimate the heritability of WHR in males and females from the UKB. How does it compare with your previous results. Is this expected?

  b) Estimate the genetic correlation between UKB males and UKB females for WHR. How does it compare with you previous analyses using data from GIANT?

  c) Interpret the bivariate LDSC intercept for the two WHR GWAS in the UK Biobank.

  d) Estimate the genetic correlation between UKB (fe)males and GIANT (fe)males for WHR.