

Practical Lecture 9: Heritability of binary traits

Loic Yengo

2022-06-18

Preparation

Create a directory where to run the practical

```
mkdir prac9/
```

Part 1

Load the practical R objects.

```
load("/data/module4/prac9/prac9.RData")
```

We provide an R object named PHENOMAT, which contains $N=10,000$ rows corresponding to N independent families from the population and 4 columns corresponding to fathers, mothers, sisters and brothers from these families. Each entry (Y_{ij}) corresponds to the measured trait of individual j ($j=1$ for fathers, 2 for mothers, 3 for sisters and 4 for brothers) in family $i=1, \dots, N$.

Question 1. Calculate the phenotypic correlation between siblings, as well as the regression of the average phenotype of siblings onto the average phenotypes of parents. What can you say about the heritability of that trait under the assumption that parents are unrelated (i.e. independent)? Are these results consistent? If not then what could explain this inconsistency? How can you verify your assumptions?

```
correlations_between_all_pairs <- cor(PHENOMAT)
midParent <- 0.5*(PHENOMAT[, "FATHER"] + PHENOMAT[, "MOTHER"])
midChild <- 0.5*(PHENOMAT[, "SISTER"] + PHENOMAT[, "BROTHER"])
```

Compare

```
lm(midChild~midParent)
```

and

```
cor(PHENOMAT[, "SISTER"], PHENOMAT[, "BROTHER"])
```

Question 2. We now assume that the trait in question defines the liability of a certain disease in the population. If the prevalence for that disease is $K=5\%$, then what can you say about the absolute risk (KR) of developing the disease among individuals with an affected brother. To answer that question, we would assume that all resemblance between relatives is due to additive genetic effects.

```

K=0.05
t=qnorm(1-K)
z=dnorm(t)
i=z/K
k=i*(i-t)
aR=0.5 # coefficient of co-ancestry (genetic relatedness) between siblings
tR=(t-aR*i*h2)/sqrt(1-aR*aR*h2*h2*k)
KR=1-pnorm(tR)

```

Question 3. Using the threshold (t) above which individuals are classified as diseased, estimate KR empirically for siblings pairs and for mothers-daughters pairs using the data provided in PHENOMAT (Tip: $KR = P(Y>t|Y_R>t) = P(Y>t \text{ and } Y_R>t)/P(Y_R>t) = P(Y>t \text{ and } Y_R>t)/K$). How does it compare with your results from **Question 2**?

Part 2

We provide a genomic relationship matrix (GRM) across 3000 individuals and a phenotype file with 1 quantitative trait (assumed to be an observed liability to disease) and 9 binary traits. These binary traits are obtained from discretizing the quantitative trait under a liability threshold model. The first binary trait corresponds to a prevalence $K=0.1$, the second to $K=0.2, \dots$, and the last one to a prevalence $K=0.9$.

Use GCTA to estimate the heritability of the observed liability

```

gcta --grm /data/module4/prac9/myData --pheno /data/module4/prac9/myData.phen \
--mphen 1 --reml --out observed_liability

```

Estimate the heritability on the observed (0/1) scale of the **first binary trait**

```

gcta --grm /data/module4/prac9/myData --pheno /data/module4/prac9/myData.phen \
--mphen 2 --reml --out observed_scale_K_0.1`

```

Re-estimate the heritability of the **first binary trait** but now on the liability scale using GCTA option `--prevalence`

```

gcta --grm /data/module4/prac9/myData --pheno /data/module4/prac9/myData.phen \
--mphen 2 --prevalence 0.1 --reml --out observed_liability_scale_K_0.1

```

Verify the formula $h_l^2 = h_o^2 \left(\frac{K(1-K)}{z^2(K)} \right)$, where $z=dnorm(qnorm(1-K))$.

Estimate the heritability on the observed (0/1) scale and the on the liability scale for all remaining binary traits. Plot these estimates against each other.