

In XPASS, the heritability and coheritability are estimated using the LD information from the entire chromosomes. This is different from LDSC, which only utilizes LD from local SNPs. While XPASS usually yields smaller standard error for estimated heritability's, it requires the population structures in the reference pannel to be properly corrected by including the covariates. Otherwise, the estimated heritability and coheritability can be biased by the population structures. Therefore, while the covariates files are optional in the software, we **strongly suggest** users to include them when using XPASS.

In the BMI example shown in the manual, we have included 5 and 20 PCs derived from the East Asian and European reference genotypes to correct for population stratification, respectively. Here we demonstrate that not including the PCs as covariates can bias the heritability and coheritability estimates.

- Prepare input files

```
# library(devtools)
# install_github("https://github.com/YangLabHKUST/XPASS")
library(XPASS)
library(data.table)
library(RhpcBLASctl)
blas_set_num_threads(30)

# reference genotypes for EAS (prefix of plink file bim/bed/fam)
ref_EAS <- "1000G.EAS.QC.hm3.ind"

# covariates of EAS reference genotypes
cov_EAS <- "1000G.EAS.QC.hm3.ind.pc5.txt"

# reference genotypes for EUR (prefix of plink file bim/bed/fam)
ref_EUR <- "1000G.EUR.QC.hm3.ind"

# covariates of EUR reference genotypes
cov_EUR <- "1000G.EUR.QC.hm3.ind.pc20.txt"

# genotype file of test data (plink prefix).
# Note: for demonstration, we assume that the genotypes of prediction target
# are used as the reference panel of target population.
# In practice, one can also use genotypes from other sources as reference
# panel.
BMI_test <- "1000G.EAS.QC.hm3.ind"

# sumstats of height
BMI_bbj_male <- "BMI_bbj_male_3M_format_3MImp.txt" # target
BMI_ukb <- "BMI_ukb_sumstat_format_all.txt" # auxiliary

BMI_bbj_female <- "BMI_bbj_female_3M_format_3MImp.txt" # external validation
```

- Run XPASS for estimating parameters by **including PCs as covariates**

```
# Run XPASS by including PCs as covariates
```

```
fit_bbj <-XPASS(file_z1 = BMI_bbj_male,file_z2 = BMI_ukb,file_ref1 = ref_EAS,
               file_ref2 = ref_EUR,
               file_cov1 = cov_EAS,file_cov2 = cov_EUR,
               file_predGeno = BMI_test,
               compPRS=F,
               pop = "EAS",sd_method="LD_block",compPosMean = F,
               file_out = "BMI_bbj_ukb_ref_TGP")
```

Summary statistics file 1: BMI_bbj_male_3M_format_3MImp.txt

Summary statistics file 2: BMI_ukb_sumstat_format_all.txt

Reference file 1: 1000G.EAS.QC.hm3.ind

Reference file 2: 1000G.EUR.QC.hm3.ind

Covariates file 1: 1000G.EAS.QC.hm3.ind.pc5.txt

Covariates file 2: 1000G.EUR.QC.hm3.ind.pc20.txt

Reading data from summary statisitcs...

3506148 and 3777871 SNPs found in summary statistics files 1 and 2.

Reading SNP info from reference panels...

1209411 and 1313833 SNPs found in reference panel 1 and 2.

746454 SNPs are matched in all files.

0 SNPs are removed because of ambiguity; 746454 SNPs remained.

Calculating kinship matrix from the both reference panels...

127749 SNPs in the second reference panel are aligned for alleles according to the first.

14337 SNPs have different minor alleles in population 1, z-scores are corrected according to reference panel.

14332 SNPs have different minor alleles in population 2, z-scores are corrected according to reference panel.

Assigning SNPs to LD Blocks...

Calculate PVE...

	h1	h2	h12	rho
[1,]	0.165790395	0.247603545	0.129399058	0.63866483
[2,]	0.007631346	0.008542654	0.006856057	0.02002658

Done.

- Run XPASS for estimating parameters while **not including PCs**

```
> fit_bbj <-XPASS(file_z1 = BMI_bbj_male,file_z2 = BMI_ukb,file_ref1 = ref_EAS,
                 file_ref2 = ref_EUR,
                 # file_cov1 = cov_EAS,file_cov2 = cov_EUR,
                 file_predGeno = BMI_test,
                 compPRS=F,
```

```

        pop = "EAS",sd_method="LD_block",compPosMean = F,
        file_out = "BMI_bbj_ukb_ref_TGP_nocvt")
Writing to log file: BMI_bbj_ukb_ref_TGP_nocvt.log
Summary statistics file 1: BMI_bbj_male_3M_format_3MImp.txt
Summary statistics file 2: BMI_ukb_sumstat_format_all.txt
Reference file 1: 1000G.EAS.QC.hm3.ind
Reference file 2: 1000G.EUR.QC.hm3.ind
Test genotype file : 1000G.EAS.QC.hm3.ind
Reading data from summary statistics...
3506148 and 3777871 SNPs found in summary statistics files 1 and 2.
Reading SNP info from reference panels...
1209411 and 1313833 SNPs found in reference panel 1 and 2.
746454 SNPs are matched in all files.
0 SNPs are removed because of ambiguity; 746454 SNPs remained.
Calculating kinship matrix from the both reference panels...
127749 SNPs in the second reference panel are aligned for alleles according to
the first.
14337 SNPs have different minor alleles in population 1, z-scores are corrected
according to reference panel.
14332 SNPs have different minor alleles in population 2, z-scores are corrected
according to reference panel.
Assigning SNPs to LD Blocks...
Calculate PVE...
          h1          h2          h12          rho
[1,] 0.072180463 0.15389606 0.124566236 1.18188918
[2,] 0.003322473 0.00530962 0.006599995 0.03706045
Done.

```

By comparing the results obtained from the above two approaches, we can observe that when the PCs are not included as covariates in the model, heritabilities from both populations are underestimated and the genetic correlation estimate exceeds 1, suggesting the estimates are biased by the uncorrected population structure. Therefore, including the PCs as covariates is very important for parameter estimation in XPASS and we **strongly suggest** users to include them when using XPASS.