

Multiple linear regression for GWAS

$$y_1 = \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} + \cdots + \beta_p x_{1,p} + \epsilon$$

$$y_2 = \beta_0 + \beta_1 x_{2,1} + \beta_2 x_{2,2} + \cdots + \beta_p x_{2,p} + \epsilon$$

...

$$y_n = \beta_0 + \beta_1 x_{n,1} + \beta_2 x_{n,2} + \cdots + \beta_p x_{n,p} + \epsilon$$

y_n = phenotype of person n

β_p = slope for SNP number p

$x_{n,p}$ = 0, 1 or 2 for SNP number p of person n

ϵ = environmental factors (covariants)