Here we will test how each method fairs under each set of conditions. Since we are estimating the heritability it's useful to note that heritability from the lens of a linear model is the partial R^2 value associated with genetic variation (R^2_G). Since in all cases the R^2 values will share the same denominator ($\hat{V}ar(Y)$), we only need to consider the numerator, the expected sum of squares.

1 H_0: AdjHE + site effect

$$Y = \begin{bmatrix} X_g \\ X_c \\ X_s \\ I \end{bmatrix} \begin{bmatrix} \beta_g & \beta_c & \beta_s & \epsilon \end{bmatrix}$$

$$Q_c Y = \begin{bmatrix} X_g \\ X_s \\ I \end{bmatrix} \begin{bmatrix} \beta_g & \beta_s & \epsilon \end{bmatrix}, E\beta_g = E\beta_s = E\epsilon = 0$$

$$EQ_c Y = 0 \qquad\qquad\qquad\text{(First Moment)}$$

$$\forall i \in \{1..M\} \forall j \in \{1..N_s\}, Var(\beta_{gi}, \beta_{sj}, \epsilon) = \begin{bmatrix} \sigma_g^2 & 0 & 0 \\ 0 & \sigma_s^2 & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix}$$

$$EYY' = Var(Y) + EYEY' = EYY'$$
$$\text{(Second Moment)}$$

$$YY' = A\sigma_G^2 + S\sigma_S^2 + I\sigma_E^2$$
$$\text{(Regression Equation)}$$

$$X_c \perp X_s, X_g$$

2 AdjHE Estimator

From previous paper the expectation of the AdjHE estimator is

$$E\hat{h}^2 = h^2 \frac{tr(A-I)Q(A^* - I - \sigma_S^2 S - X_c\beta_c\beta_c'X_c')Q}{tr(A-I)Q(A-I)Q}$$

Where Q is a projection matrix defined for the variance effecting terms (i.e. not fixed effect terms). The most basic to account for population substructure is the project ion matrix defined by the pc loading vectors

$$Q_{P_k} = PC_k(PC_k'PC_k)^{-1}PC_k'$$
$$Q_P = I - \sum_k Q_{P_k}$$

But in addition, could include other covariates

3 Covariates

The mean can just be fit with OLS with an appeal to M-estimation theory since it's a finite covariance, it will be consistent.

3.1 SWD

$$Q_1 = I - X_s(X_s'X_s)^{-1}X_s'$$

$$Q_2 = I - X_c(X_c'X_c)^{-1}X_c''$$

$$E\hat{h}^2 = h^2 \frac{tr(A-I)Q_2Q_1(A^* - I - \sigma_S^2 S - X_c\beta_c\beta_c'X_c')Q_1Q_2}{tr(A-I)Q(A-I)Q}$$

$$= h^2 \frac{tr(A-I)Q_2Q_1(A^* - I - X_c\beta_c\beta_c'X_c')Q_1Q_2}{tr(A-I)Q_2Q_1(A-I)Q_1Q_2}$$

$$= h^2 \frac{tr(A-I)Q_2Q_1(A^* - I)Q_1Q_2}{tr(A-I)Q_2Q_1(A-I)Q_1Q_2} \qquad (X_c \perp X_s)$$

Therefore unbiased when $X_g \perp X_S$

3.2 Adjusted residual harmonization

$$Q = I - X_{sc}(X_{sc}'X_{sc})^{-1}X_{sc}'$$

$$E\hat{h}^2 = h^2 \frac{tr(A-I)Q(A^* - I - \sigma_S^2 S - X_c\beta_c\beta_c'X_c')Q}{tr(A-I)Q(A-I)Q}$$

$$= h^2 \frac{tr(A-I)Q(A^* - I)Q}{tr(A-I)Q(A-I)Q}$$

$$= h^2 \frac{tr(A-I)Q(A-I)Q}{tr(A-I)Q(A-I)Q}$$

Therefore unbiased

3.3 Normalize

Mean center

$$\hat{\beta}_S = (X_sX_S')^{-1}X_s'Y$$

$$R = y - X_S\hat{\beta}_S$$

Variances

$$SS_i = (y - X_S\hat\beta_S)' \begin{bmatrix} I_{n_i} & 0 \\ 0 & 0 \end{bmatrix} (y - X_S\hat\beta_S)$$

$$(X_S'X_S)^{-1} \qquad\qquad\qquad\qquad \text{(Site counts)}$$

$$\hat{V}ar(R_i) = SS_i/n_i$$

Variance standardizer

$$Q_V = diag(\hat{V}ar(R_i) \otimes J_{n_i})$$

$$E\hat{h}^2 = E\frac{trQ_VQ_S(YY-I)Q_SQ_V'}{trQ_VQ_S(A-I)Q_SQ_V'}$$

Not clear how to proceed

3.4 Combat

$$Z = Q_cY/\hat\sigma^2 \qquad\qquad\qquad \text{(Standardize)}$$

$$= \begin{bmatrix} X_g & X_s & I \end{bmatrix} \begin{bmatrix} \beta_g \\ \beta_s \\ \epsilon \end{bmatrix} /\hat\sigma^2$$

Site mean

$$\hat\beta_{si}^* = \frac{n_i\bar\tau_i^2\hat\beta_{si} + \delta_i^2\bar\gamma_i}{n_i\bar\tau_i^2 + \delta_i^2}$$

Site Variance

$$\hat\delta_i^2 = \frac{\bar\theta_i + 1/2\sum(Z - \hat\beta_i^*)^2}{\frac{n_i}{2} + \bar\lambda_i - 1}$$

$$T(Y) = \frac{\hat\sigma_s^2}{\hat\delta_s^2}(Y - \hat\beta_s) + X_c\hat\beta_c, \hat\beta_c \text{ is OLS for each coefficient}$$

$\sigma_g^2 = 0$ it is set to zero since we standardize it

$$T(Y) = \frac{1}{\hat\delta_s^2}(Y - \hat\beta_s) + X_c\hat\beta_c$$

3

3.5   Combat Verbatim

Gene specific (g) site mean(i) centered at overall site mean

$$\gamma_{ig} \sim N(\gamma_i, \tau_i^2), \quad \delta_{ig}^2 \sim IG(\lambda_i, \theta_i)$$

$$\hat{\gamma}_{ig} = \frac{1}{n_i} \sum_j Z_{ijg} = J'_{n_i} Z_{ig}/n_i$$

$$\hat{\gamma} = diag(\frac{1}{n_i}) J'_{n_i} Z_{ig}$$

$$\hat{\gamma}_{ig} = \frac{n_i \tau_i^{-2} \hat{\gamma}_{ig} + \delta_{ig}^2 \bar{\gamma}_i}{n_i \tau_i^{-2} + \delta_{ig}^2}$$

For one phenotype $\bar{\gamma}_i = \hat{\gamma}_{ig}, \tau_i^{-2} = 0$
Meanin batch mean is just the sample mean. for the batch variance

$$\hat{\delta}_i^2 = \frac{1}{n_i - 1} \sum (Y_{ij} - \hat{\gamma}_i)^2$$

$$\delta_i^* = \frac{\bar{\theta}_i + 1/2 \sum_j (Y_{ijg} - \gamma_i^*)^2}{\frac{n_i}{n} + \bar{\lambda}_i - 1}$$

$$\bar{\lambda}_i = \frac{\bar{V}_i + 2\bar{S}_i^2}{\bar{S}_i^2}$$

for single phenotype

$$\bar{V}_i = \hat{\delta}_i^2, \bar{S}_i = 0$$

Meaning

$$\bar{\lambda}_i = \frac{\bar{V}_i + 2\bar{S}_i^2}{\bar{S}_i^2}$$

???????
???????
? Combat not really comparable since it borrows data across multiple phenotypes???
???????

$$\begin{bmatrix} trA^2 - \sum s_j^2 & trA - \sum s_j \\ trA - \sum s_j & N - k \end{bmatrix} \begin{bmatrix} \sigma_g^2 \\ \sigma_e^2 \end{bmatrix} = \begin{bmatrix} y'Ay - \sum t_j s_j \\ y'y - \sum t_j \end{bmatrix}$$