Heritability Analysis

The estimation of heritability using LD Score regression (LDSC) is a method introduced by Bulik-Sullivan et al. in 2015. It is a way to estimate heritability attributed to all SNPs for a given trait as well as the genetic correlation between traits, by utilizing the linkage disequilibrium (LD) structure of the SNPs.

Traditional methods to estimate heritability usually rely on having data from twins or families, but LDSC is particularly useful because it can be applied to the results from genome-wide association studies (GWAS) without needing raw genotypic data.

Here is a brief step-by-step overview of how LDSC works:

1. Calculate LD Scores:
    - For each SNP, an LD score is calculated. This is essentially a sum of the squared correlations (r^2 values) between that SNP and every other SNP within a certain window around it.
    - The idea is that SNPs with higher LD scores are in higher LD with many other SNPs.

2. Quantify SNP-heritability:
    - For each SNP, you'll have a chi-squared statistic from the GWAS, which is a measure of the association between that SNP and the trait in question.
    - If you plot the LD score of each SNP against its chi-squared statistic, the slope of the line of best fit represents the SNP heritability (scaled by a factor).

3. Regression:
    - A simple linear regression is done where the chi-squared statistics are the dependent variable, and the LD scores are the independent variable.
    - The intercept of this regression provides an estimate of the confounding bias, which can arise due to issues like population stratification.
    - The slope of this regression, as mentioned earlier, is an estimate of the heritability.

4. Heritability Estimation:
    - The SNP heritability can be estimated from the slope after accounting for sample size.

5. Genetic Correlation Between Traits:
    - If you have GWAS results for two different traits, you can calculate the genetic correlation using LD scores.
    - This is done by taking the product of the Z-scores (standard normal deviates) for each SNP for both traits and regressing this against the LD scores.
    - The slope from this regression provides an estimate of the genetic correlation.

LDSC is a valuable tool because it accounts for the effect of LD on GWAS results and can be used to provide insights into the genetic architecture of complex traits and diseases. One thing to note is that while LDSC provides estimates for SNP-based heritability, it might not capture the entire genetic variance, especially if there are contributions from rare variants or structural variants that are not well tagged by common SNPs.

Genetic Correlation

Estimating genetic correlation between two traits using LD Score Regression (LDSC) involves leveraging the genetic signals (often in the form of GWAS summary statistics) from both traits to understand the degree to which their genetic architectures overlap.

Here's a step-by-step breakdown:

1. Input Data:
    - GWAS summary statistics for two traits you're interested in. This data includes, for each SNP:
        i. SNP identifier
        ii. Effect size (often a beta coefficient)
        iii. Standard error of the effect size
        iv. P-value (or Z-score)

2. Standardize GWAS Results:
    - Convert the effect sizes from each GWAS to Z-scores by dividing the effect size by its standard error.

3. Compute Cross-products of Z-scores:
    - For each SNP, compute the product of the Z-scores from the two GWAS. This product serves as the cross-trait LD score statistic.

4. LD Score Regression:
    - You'll regress the cross-trait LD score statistic for each SNP (from step 3) onto the LD scores of that SNP.
    - The slope of this regression is proportional to the genetic correlation between the two traits.

5. Estimate Genetic Correlation:
    - The genetic correlation (rg) between the two traits is estimated as:

$$r_g = \frac{slope\ of\ regression}{\sqrt{Heritability\ trait\ 1 \times Heritability\ Trait\ 2}}$$

- The heritability's of the traits can be estimated using LDSC on each trait separately, as described in the previous explanation.

6. Interpreting the Genetic Correlation:
    - A rg close to 1 suggests the two traits share a lot of genetic factors in the same direction.
    - A rg close to -1 indicates that the shared genetic factors have opposite effects on the two traits.
    - A rg close to 0 means there's little genetic overlap between the two traits.

<u>Observed Scale vs. Liability Scale</u>

1. Observed Scale: This refers to the scale on which the phenotype is directly observed. For a binary trait, this is the simple proportion of individuals who have the trait (e.g., disease prevalence in the sample).

2. Liability Scale: This is a hypothetical underlying continuous scale that determines the risk of an individual manifesting the binary phenotype. When this risk surpasses a certain threshold, the individual manifests the disease (or the binary trait). The distribution of this risk across a population is usually assumed to be normal (Gaussian). The actual manifestation of the disease happens when an individual's liability exceeds a certain threshold.

    a. The beauty of the liability scale is that it lets us talk about the genetics of binary (yes/no) traits in a continuous way. It's a bit like converting a switch (on/off) into a sliding scale (from low to high). This makes it easier to analyze and compare with other traits.

<u>Why is Liability Scale Important?</u>

1. Comparability: Transforming binary traits to the liability scale makes their heritabilities more comparable to those of continuous traits. Additionally, when estimating genetic correlations between a binary trait and a continuous trait using methods like LDSC, it's crucial to have both traits on a similar scale.

2. Accounting for Population Prevalence: When calculating heritability on the liability scale, the known or estimated prevalence of the trait in the general population (not just the sample) is used. This makes the estimates more generalizable and reflective of the underlying genetic architecture, rather than being solely based on the specific sample's characteristics.

3. Statistical Power: Estimating heritability on the liability scale can lead to increased statistical power. Especially in cases where the binary trait has a low prevalence, the observed-scale heritability can be much lower than the liability-scale heritability.