**Lab: Genomic SEM**

Using Genomic SEM to:

-Estimate a common factor model

-Run a user specified model

-Run a multivariate GWAS

**Estimate a Common Factor Model**

**Step 1:** Munge the summary statistics (*munge*)

**Step 2:** Run multivariable ld-score regression (*ldsc*)

**Step 3:** Estimate the common factor model (*commonfactor*)

**Run a User Specified Model**

Follow same **Steps 1-2** for common factor model, but specify your own model for **Step 3** (*usermodel*)

**Multivariate GWAS in Genomic SEM**

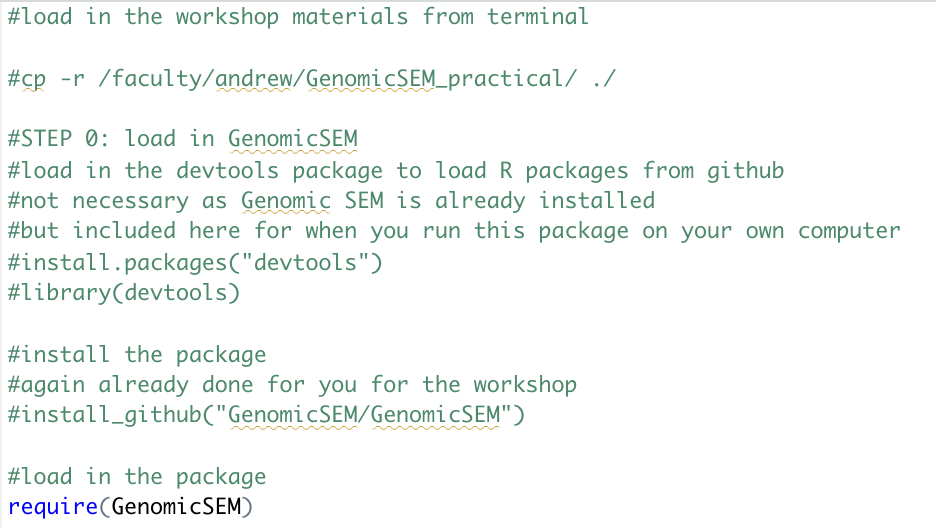
**Step 1:** Munge the summary statistics (*munge*)

**Step 2:** Run multivariable ld-score regression (*ldsc*)

**Step 3:** Prepare the summary statistics for GWAS (*sumstats*)

**Step 4:** Run the multivariate GWAS (*commonfactorGWAS* or *userGWAS*)

**Step 0:** Start up an R session, copy over workshop materials into your folder, and load in GenomicSEM

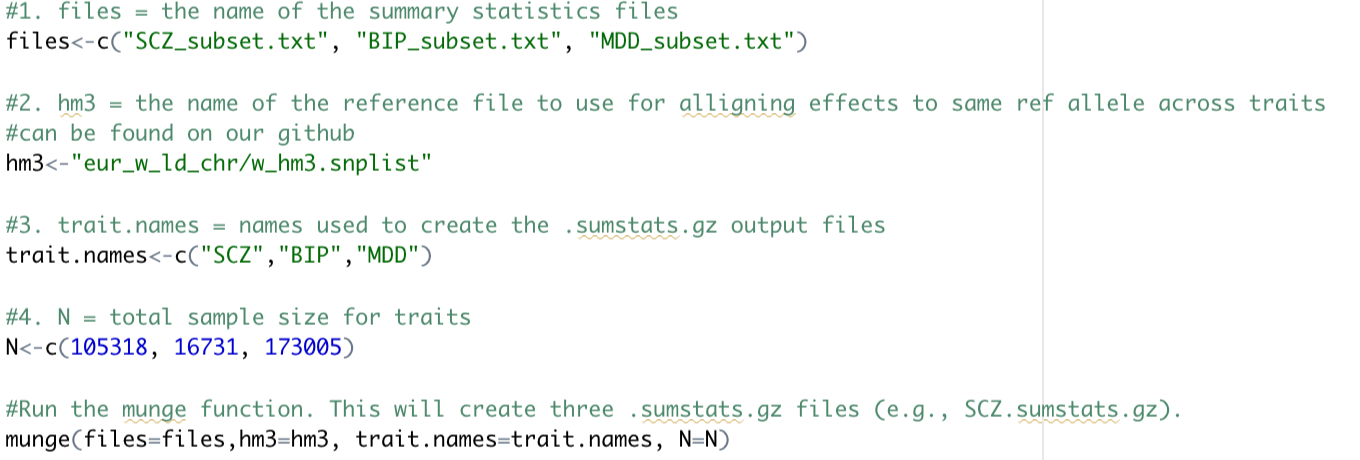


**Estimate a Common Factor Model**

**ACTIVITY**: Fit a common factor model for three psychiatric traits: Schizophrenia, Bipolar Disorder and Major Depressive Disorder

**Step 1:** Munge the three provided summary statistics using the *munge* function

*\*Note that all code provided for Genomic SEM below is written in R.*

Now open the .log file produced by munge to answer questions below:

Checking Understanding Questions (to discuss as group):

1. How many SNPs are left for MDD?
2. How many SNPs are pruned for imputation quality (INFO score) for SCZ?
3. How many rows are removed from BIP due to low minor allele frequency (MAF)?

**Step 2:** Run multivariable LD-Score Regression. We specifically use the *ldsc* function from Genomic SEM as this accounts for sample overlap

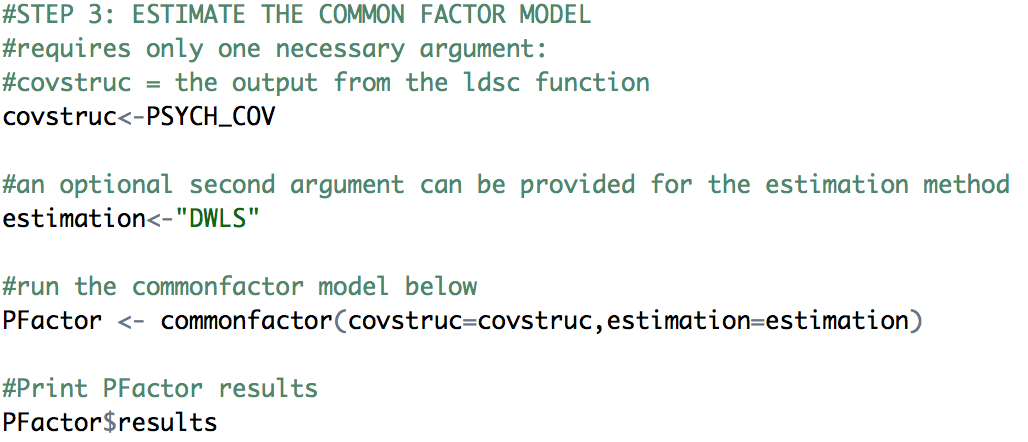


Answers to these questions will be in the results Rstudio prints to the screen:

Checking Understanding Questions (to discuss as group):

1. What is the liability scale h2 for BIP?
2. What is the observed scale h2 for SCZ?
3. What is the genetic correlation between SCZ and MDD?

**Step 3:** Run the model using the *commonfactor* function



Checking Understanding Questions (to discuss as group):

1. What is the unstandardized factor loading for SCZ?
2. What is the standardized residual variance of MDD?

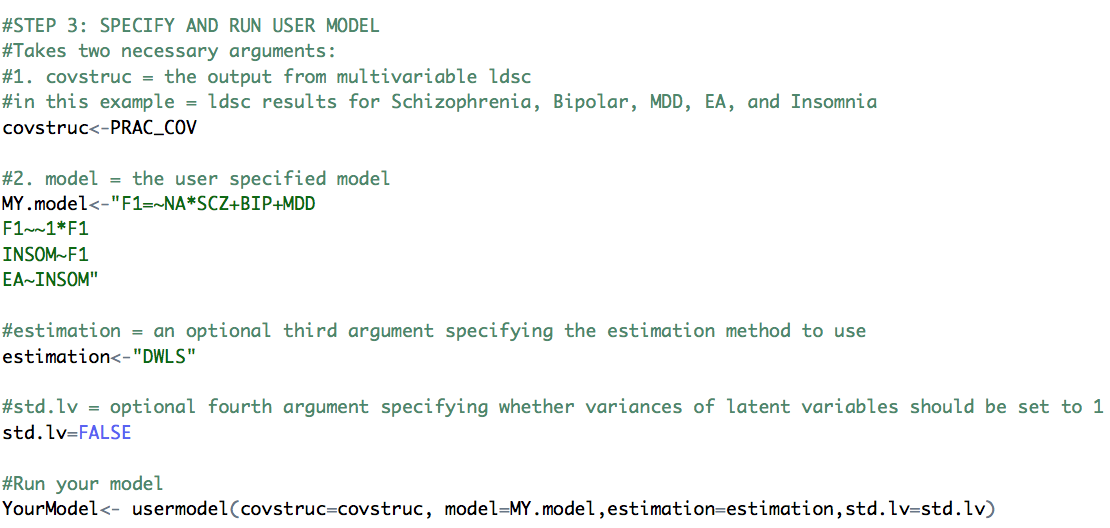
Pause here; we will discuss next steps as a group.

**Estimate a User Specified Model**

**ACTIVITY**: Fit a user specified model for three psychiatric traits (Schizophrenia, Bipolar Disorder and Major Depressive Disorder), and two external phenotypes (Educational Attainment and Insomnia)

**Steps 1 and 2:** Already done for you, but as with the common factor model, involved munging the summary statistics (**Step 1)** and then running multivariable ld-score regression (**Step 2**)

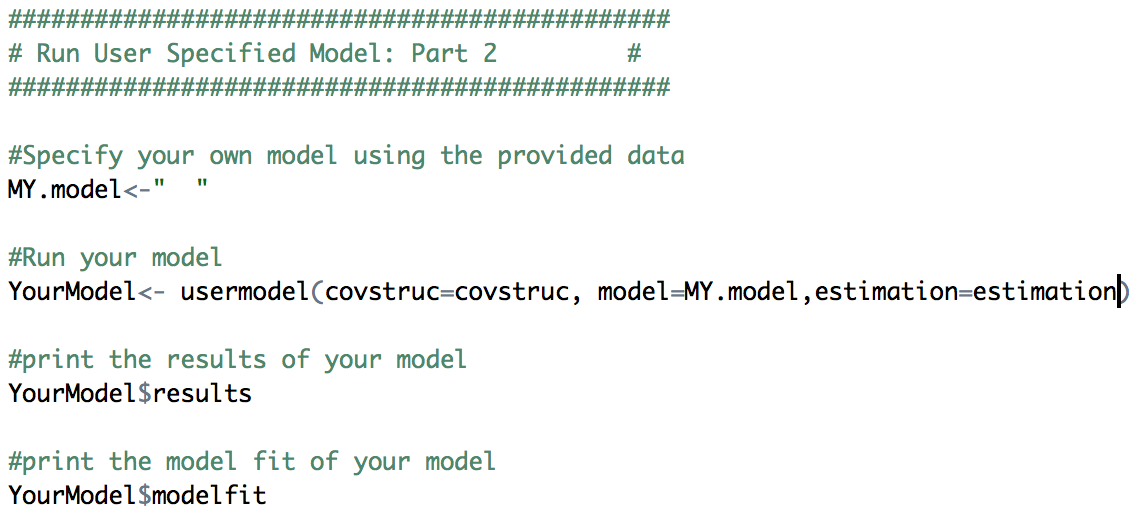
**Step 3a:** Specify and run a user-specified model that is provided for you



**ACTIVITY**: Run your own user-specified model using the same dataset

**Steps 1 and 2:** As before, already done for you.

**Step 3:** Specify your own model! Pre-register the model by it down on paper beforehand



Checking Understanding Questions (to discuss as group):

1. How would you verbally explain your model?
2. What is the CFI and AIC for your model?
3. Would you describe your model as fitting the data well?

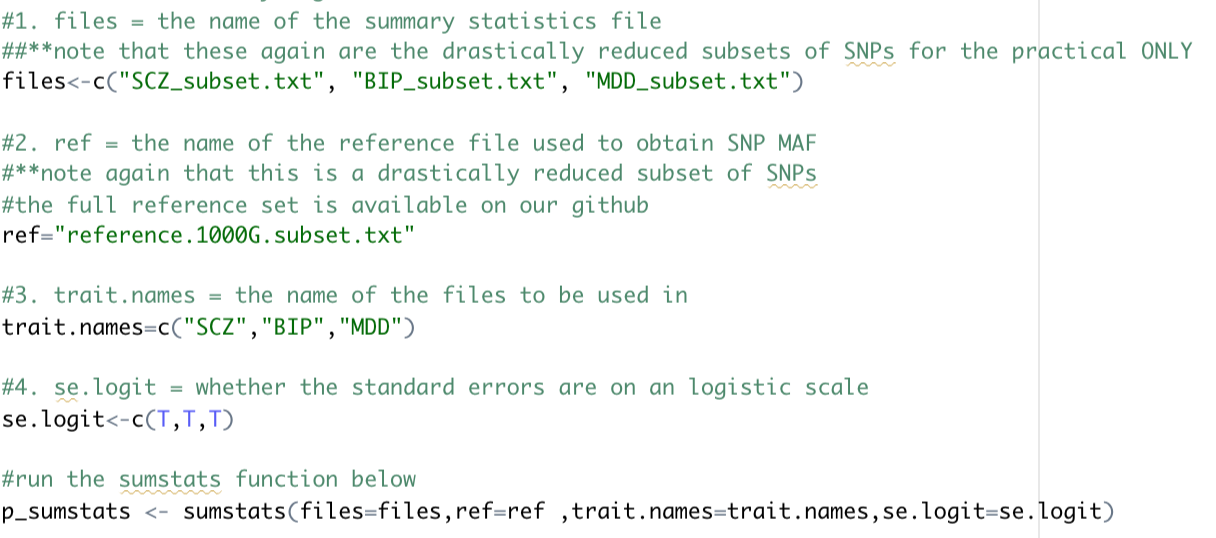
Pause here; we will discuss next steps as a group.

**Multivariate GWAS in Genomic SEM**

**ACTIVITY**: Estimate Multivariate GWAS for common factor using the three psychiatric traits (Schizophrenia, Bipolar Disorder, Major Depressive Disorder)

**Steps 1 and 2:** You already did them in the first example!

**Step 3:** Prepare summary statistics for multivariate GWAS using the *sumstats* function

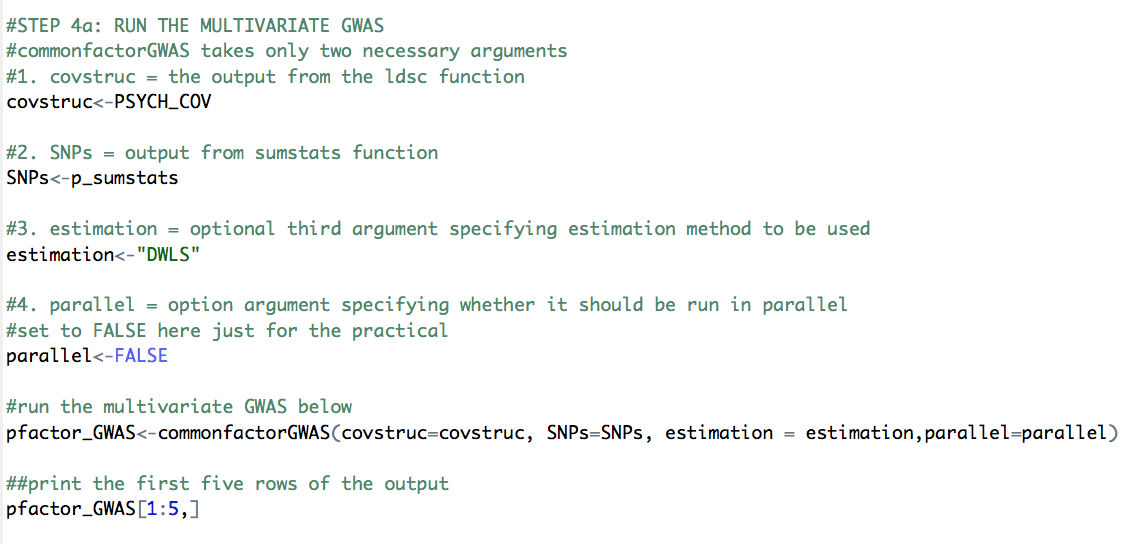


You can to open the .log file created by *sumstats* to answer some of these questions or examine what R is printing to screen when running *sumstats*.

Checking Understanding Questions (to discuss as group):

1. What columns is *sumstats* using for MDD to compute total sample size?
2. How many total SNPs are left across all three traits?
3. What is being interpreted as the “effect” column for SCZ?

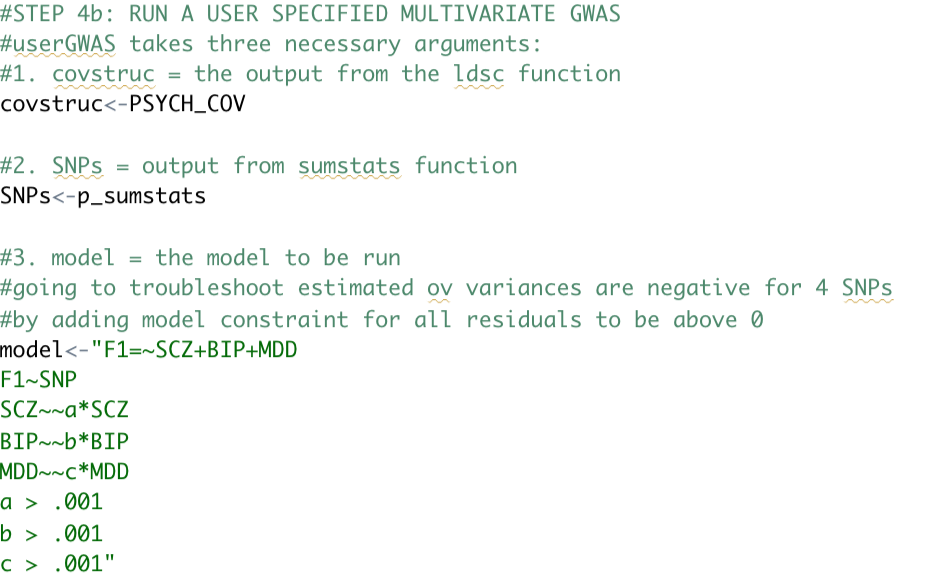
**Step 4a:** Run the multivariate GWAS using the *commonfactorGWAS* function

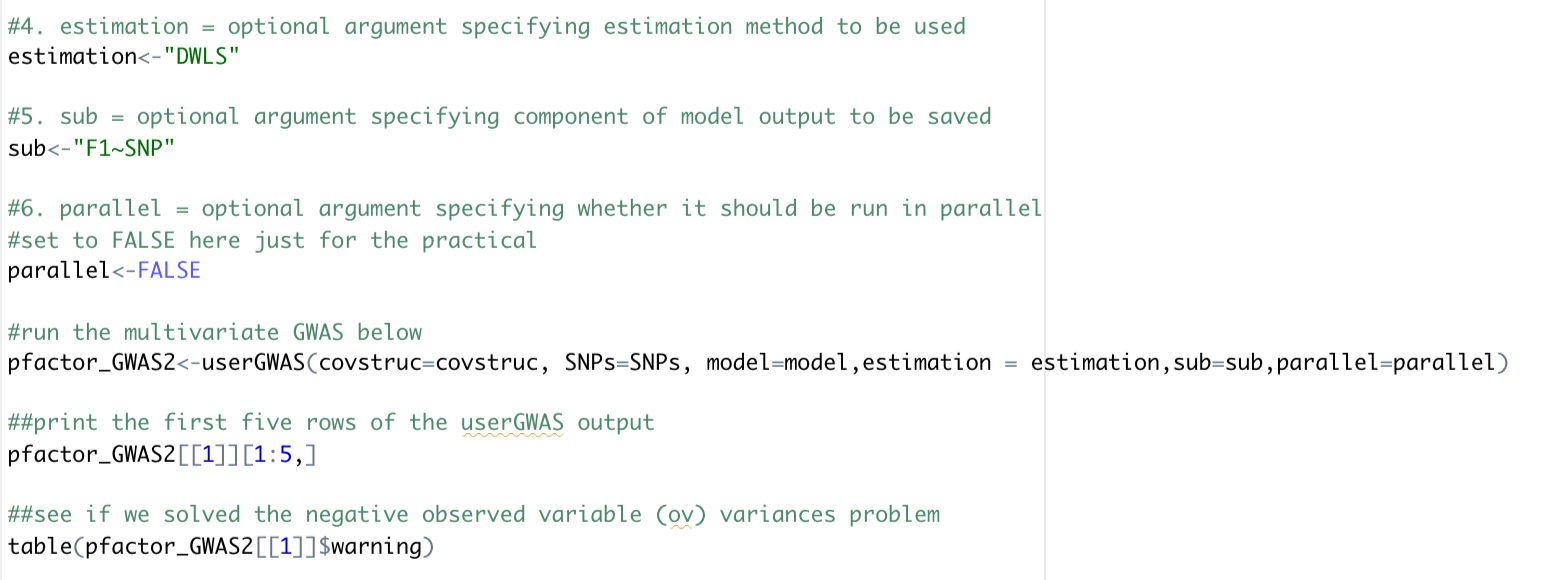


Checking Understanding Questions (to discuss as group):

1. How many warnings are there?
2. What is the p-value for the SNP effect on the factor for rs100053?

**Step 4b:** Run the same common factor model (with constraints on the residual variances) using the *userGWAS* function

****

****

**Feel free to move on to anthropometric traits example at the end of the R code if you have time.**