

Workshop: Mendelian Randomization

Paquetes usados en el Workshop

MendelianRandomization

Paquete desarrollado para llevar a cabo varios análisis de aleatorización mendeliana sobre datos genéticos en R. El paquete utiliza varios métodos para evaluar si un factor de riesgo (o exposición) tiene un efecto causal sobre un evento.

metafor

Proporciona funciones para llevar a cabo metaanálisis en R. Incluye funciones para ajustar los modelos de efectos fijos y aleatorios. Permite la inclusión de variables moderadoras en los modelos.

TwoSampleMR

Paquete enfocado en el análisis de aleatorización mendeliana de dos muestras (2SMR). Este es un método para estimar el efecto causal de una exposición sobre un resultado utilizando sólo estadísticas resumidas de estudios de asociación de genoma completo (GWAS).

MRPRESSO

Mendelian Randomization Pleiotropy RESidual Sum and Outlier, es un método que permite evaluar la pleiotropía horizontal en la aleatorización mendeliana multi instrumento utilizando estadísticas de asociación resumidas de todo el genoma.

Bases de datos

Para este workshop se utiliza la base de datos del estudio: “A Genome-wide Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic Community Health Study/Study of Latinos”, que es un GWAS sobre el metaboloma humano del cual obtenemos los datos de exposición de ácido úrico en población latina que contiene 1,169 observaciones. El artículo se puede acceder por medio de este **enlace**.

La base de datos que se utiliza para obtener el evento de enfermedad arterial coronaria (CAD) se obtiene del artículo: “A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease”. Que tiene 933 variantes de CAD. Se puede acceder a este por medio de este **enlace**

Limpieza y unión de las bases de datos

Parte del proceso de análisis compete a la limpieza y de las bases de datos y homogeneización de la información. Después de descargadas las bases de datos y manejo de la información (Ver 01 - MR Analysis.R, seccion 1, 2 y 3) debemos obtener una base de datos con 1,023 obsservaciones con la siguiente estructura:

```
head(exp_out)
```

```
##          cpaid beta.outcome se.outcome effect_allele.outcome
## 1 10:1012315:A:G    -0.013016  0.0109204                A
## 2 10:1012372:A:G    -0.013041  0.0109193                A
## 3 10:1013535:A:C    -0.012598  0.0109019                A
## 4 10:1014603:A:G    -0.013341  0.0109008                A
## 5 10:1014603:A:G    -0.013341  0.0109008                A
## 6 10:1019138:A:T    -0.012900  0.0109106                T
##  other_allele.outcome eaf.outcome pval.outcome      SNP beta.exposure
## 1                   G    0.725704    0.2332993 rs10795302    0.1325959
## 2                   G    0.725832    0.2323564 rs10795303    0.1324959
## 3                   C    0.725339    0.2478531 rs6560863    0.1307491
## 4                   G    0.725719    0.2210069 rs1904671    0.1318013
## 5                   G    0.725719    0.2210069 rs1904671    0.1315361
## 6                   A    0.723641    0.2370712 rs10795306    0.1289839
##  se.exposure effect_allele.exposure other_allele.exposure eaf.exposure
## 1 0.02808845                G                A    0.211479
## 2 0.02808579                G                A    0.211491
## 3 0.02804180                C                A    0.211538
## 4 0.02804180                G                A    0.211861
## 5 0.02804489                G                A    0.211825
## 6 0.02804754                A                T    0.212326
##  pval.exposure
## 1 2.35084e-06
## 2 2.38735e-06
## 3 3.12169e-06
## 4 2.59926e-06
## 5 2.72924e-06
## 6 4.25009e-06
```

SNPs independientes

Procedemos a encontrar todos los SNPs independientes por medio del proceso de Linkage Disequilibrium (LD) Clumplig. Este proceso informa las asociaciones genéticas más significativas en una región en términos de un número menor de “grupos” de SNP genéticamente vinculados. Esto puede ayudar a evaluar cuántos loci independientes están asociados con un rasgo determinado.

```
all <- TwoSampleMR::clump_data(exp_out,          # base de datos
                               clump_r2 = 0.05,
                               pop = "AMR"       # Population reference panel - America
                               )
```

```
## pval.exposure and pval.outcome columns present. Using pval.exposure for clumping.
```

```
## API: public: http://gwas-api.mrcieu.ac.uk/
```

```
## Please look at vignettes for options on running this locally if you need to run many instances of this command.
```

```
## Clumping I3X4M1, 1023 variants, using AMR population reference
```

```
## Server code: 502; Server is possibly experiencing traffic, trying again...
```

```
## Retry succeeded!
```

```
## Removing 1000 of 1023 variants due to LD with other variants or absence from LD reference panel
```

SNP	id.exposure
rs11934731	I3X4Ml
rs76879454	I3X4Ml
rs2633842	I3X4Ml
rs72738425	I3X4Ml
rs17740450	I3X4Ml
rs76233293	I3X4Ml
rs10006210	I3X4Ml
rs74460821	I3X4Ml
rs11786665	I3X4Ml
rs2278122	I3X4Ml

Esta es la base de datos final con 23 SNPs con la que vamos a estar trabajando.

Estadística F

Los estadísticos F describen el nivel estadísticamente esperado de heterocigosidad en una población; más concretamente, el grado esperado de reducción de la heterocigosidad en comparación con la expectativa de Hardy-Weinberg.

```
all$f1 <-(all$beta.exposure * all$beta.exposure) / (all$se.exposure * all$se.exposure)

all$f1[1:10] %>%
  knitr::kable(caption = "Muestra aleatoria de la estadística F de 10 SNPs",
               col.names = "F-stat")
```

Table 2: Muestra aleatoria de la estadística F de 10 SNPs

F-stat
22.28460
22.75549
22.23903
23.56047
23.17756
21.10180
22.85019
21.49357
21.51265
24.23331

```
mean(all$f1)
```

```
## [1] 31.05569
```

Alineación de los SNPs con mismo alelo de efecto para la exposición y el evento

Alineamos los SNPs con el mismo alelo de efecto para la exposición y el evento al cambiar el símbolo de la beta de la exposición si los efectos de los alelos no son similares.

```
all$effect_allele.outcome <- as.factor(all$effect_allele.outcome)
all$effect_allele.exposure <- as.factor(all$effect_allele.exposure)

lev2 <- unique(c(levels(all$effect_allele.outcome), levels(all$effect_allele.exposure)))

all$effect_allele.outcome <- factor(all$effect_allele.outcome, levels=lev2)
all$effect_allele.exposure <- factor(all$effect_allele.exposure, levels=lev2)

all$effect_allele.exposure <- gsub(" ", "", all$effect_allele.exposure, fixed = TRUE)

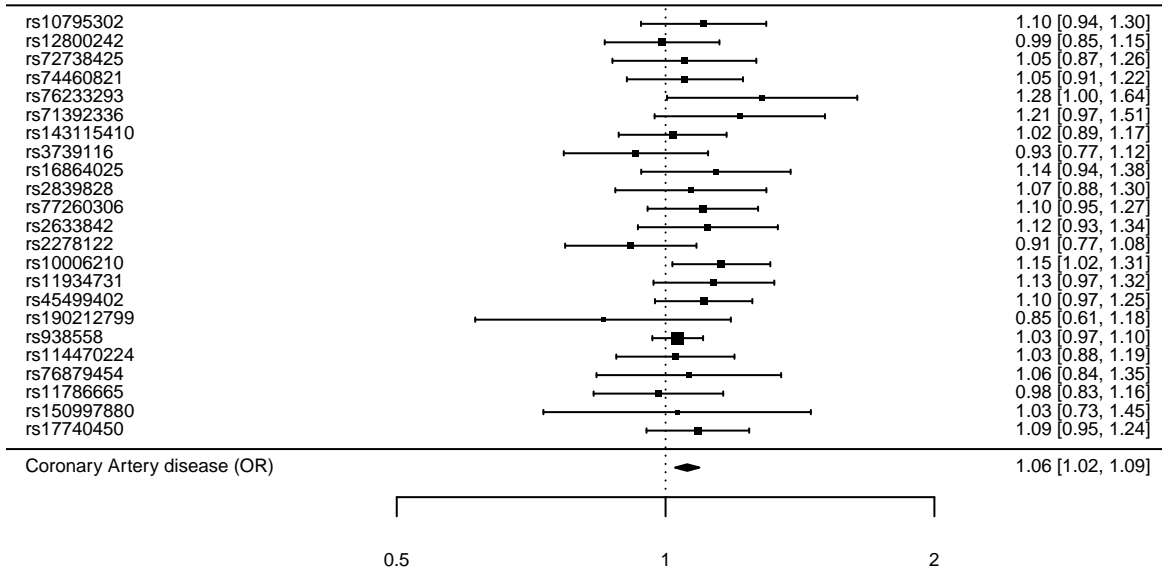
all$beta.exposure[all$effect_allele.exposure != all$effect_allele.outcome] <-
  all$beta.exposure[all$effect_allele.exposure != all$effect_allele.outcome] * -1
```

Wald test & Forest plot

Corremos el test de Wald para efectos fijos y generamos un forest plot Igual que con la Aleatorización Mendeliana de varianza inversa ponderada (IVW) de efectos fijos.

```
dmres
```

```
##
## Fixed-Effects Model (k = 23)
##
## I2 (total heterogeneity / total variability): 0.00%
## H2 (total variability / sampling variability): 0.80
##
## Test for Heterogeneity:
## Q(df = 22) = 17.5035, p-val = 0.7350
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## 0.0558 0.0160 3.4886 0.0005 0.0245 0.0872 ***
##
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Estimaciones MR

Obtenemos las estimaciones para la CAD usando el paquete de aleatorización mendeliana con el modelo de efectos fijos por medio del método de varianza inversa.

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 23
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW      0.056      0.016 0.024, 0.087    0.000
## -----
## Residual standard error = 0.892
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption.
## Residual standard error is set to 1 in calculation of confidence interval when its estimate is less than 1.
## Heterogeneity test statistic (Cochran's Q) = 17.5035 on 22 degrees of freedom, (p-value = 0.7350). I2 = 0.0%.
## F statistic = 31.1.
```

Obtenemos las estimaciones para enfermedad cardíaca isquémica usando el paquete de aleatorización mendeliana con el modelo de efectos aleatorios.

```
##
```

```
## Inverse-variance weighted method
## (variants uncorrelated, random-effect model)
##
## Number of Variants : 23
##
## -----
## Method Estimate Std Error 95% CI      p-value
## IVW      0.056      0.016 0.024, 0.087    0.000
## -----
## Residual standard error = 0.892
## Residual standard error is set to 1 in calculation of confidence interval when its estimate is less than 1.
## Heterogeneity test statistic (Cochran's Q) = 17.5035 on 22 degrees of freedom, (p-value = 0.7350). I2 = 0.0%.
## F statistic = 31.1.
```

Obtenemos la mediana ponderada y la estimación MR-Egger

```
##
## Weighted median method
##
## Number of Variants : 23
## -----
## Method Estimate Std Error 95% CI      p-value
## Weighted median method    0.049      0.023 0.003, 0.094    0.037
## -----
##
## MR-Egger method
## (variants uncorrelated, random-effect model)
##
## Number of Variants = 23
##
## -----
## Method Estimate Std Error 95% CI      p-value
## MR-Egger      0.062      0.040 -0.015, 0.140    0.116
## (intercept)   -0.001      0.008 -0.017, 0.014    0.855
## -----
## Residual Standard Error : 0.912
## Residual standard error is set to 1 in calculation of confidence interval when its estimate is less than 1.
## Heterogeneity test statistic = 17.4701 on 21 degrees of freedom, (p-value = 0.6822)
## I2GX statistic: 82.4%
```

Mendelian Randomization Pleiotropy

Con nuestros datos corremos la prueba Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-Presso), este método ayuda a identificar SNPs que componen la CAD revisando cuál de estos son outlayers.

```
## Warning in mr_presso(BetaOutcome = "beta.outcome", BetaExposure =
## "beta.exposure", : No outlier were identified, therefore the results for the
## outlier-corrected MR are set to NA

## $'Main MR results'
## Exposure      MR Analysis Causal Estimate      Sd T-stat
## 1 beta.exposure      Raw      0.05581901 0.01427209 3.91106
```

```

## 2 beta.exposure Outlier-corrected          NA          NA          NA
##      P-value
## 1 0.0007490214
## 2          NA
##
## $'MR-PRESSO results'
## $'MR-PRESSO results'$'Global Test'
## $'MR-PRESSO results'$'Global Test'$RSSobs
## [1] 19.04688
##
## $'MR-PRESSO results'$'Global Test'$Pvalue
## [1] 0.748

##
## Random-Effects Model (k = 23; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0014)
## tau (square root of estimated tau^2 value):      0
## I^2 (total heterogeneity / total variability):   0.00%
## H^2 (total variability / sampling variability):   1.00
##
## Test for Heterogeneity:
## Q(df = 22) = 17.5035, p-val = 0.7350
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## 0.0558 0.0160 3.4886 0.0005 0.0245 0.0872 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

