

Taller 4

Juan David Henao

22 de octubre de 2015

1. Test Exacto de Fisher:

La prueba de X^2 era la prueba utilizada para conocer la relación o la no relación entre datos de tipo cualitativo, pero esta prueba esta sujeta a varios supuestos par su correcta implementación. Uno de ellos es si las frecuencias esperadas de los valores en una tabla de contingencia son lo suficientemente grandes y ademas si los datos muestran algún tipo de dependencia, entonces el método X^2 no sera el mas apropiado.

Como una alternativa a esta prueba a aparecido el Test Exacto de Fisher, el cual permite analizar si dos variables dicotómicas, en una muestra demasiado pequeña, están asociadas, teniendo en cuenta ademas que no se cumplan los supuestos necesarios para aplicar la prueba de X^2 , esto incluye que, al menos en el 80 % de las celdas de la tabla de contingencia, los valores esperados sean igual o mayores que 5.

Si las variables que se están considerando dentro de la tabla e contingencia son dicotómicas, entonces se puede obtener una tabla 2x2, entonces el test exacto de Fisher evaluara las probabilidades para cada tabla 2x2 que se puedan crear teniendo en cuenta el total marginal de la tabla observada, esto si se parte del supuesto de la independencia de las dos variables que se están analizando.

Tabla de comparación para variables dicotómicas.

Característica 2	Característica 1		Total
	Presente	Ausente	
Presente	a	b	a+b
Ausente	c	d	c+d
Total	a+c	b+d	n

La probabilidad de observar las frecuencias a, b, c y d, como se menciona anteriormente, se obtienen bajo el supuesto de independencia y son dados por la distribución hipergeométrica:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} \quad (1)$$

Esta ecuación representa todas las posibles combinaciones filasxcolumnas de la tabla 2x2 y dicha probabilidad debe ser calculada para todas las tablas de 2x2 estimadas que se puedan originar con el mismo total marginal (n) que la tabla de 2x2 observada. Una vez calculadas las probabilidades, estas serán usadas para calcular el p-value de el test exacto de Fisher el cual indicara la probabilidad de obtener diferencias de los grupos iguales o mayores al observado, dado que el p-value sea pequeño ($p < 0.05$), se

asumirá entonces que las variables no son independientes.

Para estimar el p-value se pueden seguir dos métodos, el primero de ellos es sumando las probabilidades de las tablas estimadas cuyo valor de probabilidad asociado (ecuación 1), sea menor o igual al de la tabla observada, la segunda posibilidad es sumar todas las probabilidades que apoyen la hipótesis alternativa (no independencia) tanto como lo hagan los datos reales.

2. Test exacto de fisher usando GWASTools

Para este ejercicio se utilizaron 15 SNPs relacionados a la enfermedad de esclerosis múltiple obtenidos a partir de estudios de GWAS consultados en la base de datos GWAS Central, los identificadores de los estudios realizados son: HGVRs1735, HGVRs3140 y HGVRs643.

- Cargando la librería GWASTools.

```
> library("GWASTools")
```

- Cargando los datos de SNPs.

```
> SNPs<-read.table("SNPsTaller4.txt", header = T)
> SNPs
```

	FENO	C.T	G.T	A.G	A.C
1	rs7090512	1	0	0	0
2	rs669607	0	1	0	0
3	rs1738074	0	0	1	0
4	rs1077667	0	0	1	0
5	rs11154801	0	0	0	1
6	rs17066096	0	0	1	0
7	rs4271113	0	0	1	0
8	rs2920001	1	0	0	0
9	rs11026091	0	0	1	0
10	rs5978649	0	0	1	0
11	rs533259	1	0	0	0
12	rs9568281	0	0	1	0
13	rs2523393	1	0	0	0
14	rs1800693	0	0	1	0
15	rs2300747	0	0	1	0

- Creando el tipo de archivo *MatrixGenotypeReader*.

```
> genotype<-as.matrix(SNPs[2:16,2:5])
> SNPID <- 1:15
> chrom<-as.integer(c(10,3,6,19,6,6,23,18,11,23,1,13,6,12,1))
> pos<-as.integer(c(6110829,28071444,159465977,6668972,135739355,
+                  137452908,34892503,28789725,3259809,13615118,
+                  182549019,50185204,29705659,6440009,117104215))
> IDscan<-1:4
> InputData<-MatrixGenotypeReader(genotype=genotype,snpID=SNPID,
+                                 chromosome=chrom,position=pos,scanID=IDscan)
> InputData
```

An object of class `MatrixGenotypeReader`
with 4 scans and 15 snps

- Creando el tipo de dato `ScanAnnotationDataFrame`.

```
> sex <- c('F','F','F','F')
> scanID <- as.numeric(IDscan)
> frame <- data.frame(scanID,sex)
> scan <- ScanAnnotationDataFrame(frame)
> scan
```

An object of class `'ScanAnnotationDataFrame'`
scans: 1 2 3 4
varLabels: scanID sex
varMetadata: labelDescription

- Creando el tipo de dato `GenotypeData`.

```
> InputData2<-GenotypeData(InputData, snpAnnot = NULL, scanAnnot = scan)
> InputData2
```

An object of class `GenotypeData`
| data:
An object of class `MatrixGenotypeReader`
with 4 scans and 15 snps
| SNP Annotation:
NULL
| Scan Annotation:
An object of class `'ScanAnnotationDataFrame'`
scans: 1 2 3 4
varLabels: scanID sex
varMetadata: labelDescription

- Test exacto de fisher.

```
> fisher <- batchFisherTest(InputData2, batchVar = "scanID",
+                           chrom.include=23, sex.include = "F",return.by.snp=TRUE)
> fisher
```

```
$mean.or
  1    2    3    4
Inf Inf Inf Inf
```

```
$lambda
      1          2          3          4
2.000000e+00 1.601713e-16 1.601713e-16 1.601713e-16
```

```
$pval
  1 2 3 4
7  0.25 1 1 1
10 0.25 1 1 1
```

```
$oddsratio
      1 2 3 4
7  Inf 0 0 0
10 Inf 0 0 0
```

```
$allele.counts
      nA nB
7      1  7
10     1  7
```

```
$min.exp.freq
      1      2      3      4
7  0.25 0.25 0.25 0.25
10 0.25 0.25 0.25 0.25
```