

# 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 relacion o la no relacion entre datos de tipo cualitativo, pero esta prueba esta sujeta a varios supuestos par su correcta implementacion. 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 mestran algun tipo de dependencia, entonces el metodo  $X^2$  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 dicotomicas, en una muestra demasiado pequeña, estan asociadas, teniendo en cuenta aemas 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 estan considerando dentro de la tabla e contingencia son dicotomicas, 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 estan analizando.

Tabla de comparacion para variables dicotomicas.

Característica 2	Característica 1		
	Presente	Ausente	Total
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 distribucion hipergeometrica:

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

Esta ecuacion represent 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 ls probabilidades, estas seran usadas para calcular el p-value de el test exacto de Fisher el cual indcara la probabilidad de obtener diferencias de los grupos iguales o mayores al observado, dado que el p-value sea pequeño ( $p < 0.05$ ), se asumira

entonces que las variables no son independientes.

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

## 2. Test exacto de fisher usando GWASTools

```
> library("GWASTools")
> SNPs<-read.table("SNPsTaller4.txt", header = T)
> 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(6110829,28071444,159465977,6668972,135739355,137452908,34892503,2878972))
> IDscan<-1:4
> InputData<-MatrixGenotypeReader(genotype=genotype,snpID=SNPID,chromosome=chrom,position=pos)
> sex <- c('F','F','F','F')
> scanID <- as.numeric(IDscan)
> frame <- data.frame(scanID,sex)
> scan <- ScanAnnotationDataFrame(frame)
> InputData2<-GenotypeData(InputData, snpAnnot = NULL, scanAnnot = scan)
> 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

>
```