Tarea 1

Daniel Felipe Villa - 1005087556 dvilla@unal.edu.co

2022-03-22

Contents

Base de datos para el desarrollo del ejercicio:

CH	Y	CH	Y	CH	Y	CH	Y	CH	Y
1	0	21	1	41	0	61	0	81	0
2	0	22	1	42	0	62	0	82	0
3	0	23	0	43	1	63	1	83	0
4	0	24	0	44	1	64	0	84	0
5	0	25	0	45	1	65	0	85	0
6	0	26	0	46	1	66	1	86	0
7	0	27	1	47	0	67	0	87	1
8	0	28	0	48	1	68	1	88	0
9	0	29	0	49	0	69	0	89	0
10	0	30	1	50	0	70	0	90	0
11	1	31	0	51	0	71	1	91	0
12	0	32	0	52	0	72	1	92	0
13	0	33	0	53	0	73	0	93	0
14	0	34	0	54	0	74	0	94	0
15	0	35	0	55	0	75	1	95	0
16	0	36	0	56	0	76	1	96	0
17	1	37	0	57	0	77	1	97	0
18	1	38	0	58	1	78	0	98	0
19	0	39	0	59	0	79	0	99	0
20	1	40	0	60	0	80	0	100	0

Figure 1: Data_Set

Las variables de nuestro dataset son:

- `CH`: Hogares de un barrio en particular
- `Y`: ¿Tienen niños menores de **5** años? $(1 : SI \mid 0 : NO)$
- `N`: # total de Hogares encuestados
- `n`: tamaño de muestra seleccionada

1er Punto

Gracias a R utilizaremos el siguiente mecanismo para seleccionar numeros aleatorios del tamaño **N** de la muestra.

```
# Definimos una semilla para que nuestra muestra a seleccionar no cambie:
set.seed(1)

N <- 100
```

```r
# Seleccionamos nuestra MAS con:
n <- 20
muestra <- sort(sample(1:N,n))

# Hogares seleccionados Aleatoriamente
print(muestra)
```

```
##  [1]  1  7 14 21 34 37 39 43 51 54 59 68 73 74 79 82 83 85 87 97
```

2do punto

Aprox Binomial

para estimar la proporción necesitamos los siguientes datos:

- **a**: # de Exitos observados en la muestra (en este caso, hogares con niños menores de 5 años).

```r
# Creamos nuestro nuevo dataset a partir de "muestra":

df <- DataSets[muestra,]
print(df)
```

```
## # A tibble: 20 x 2
##       CH     Y
##    <dbl> <dbl>
##  1     1     0
##  2     7     0
##  3    14     0
##  4    21     1
##  5    34     0
##  6    37     0
##  7    39     0
##  8    43     1
##  9    51     0
## 10    54     0
## 11    59     0
## 12    68     1
## 13    73     0
## 14    74     0
## 15    79     0
## 16    82     0
## 17    83     0
## 18    85     0
## 19    87     1
## 20    97     0
```

```r
# Observamos los hogares que tienen niños menores de 5 años:

which(df[,2] == 1)
```

```
## [1]  4  8 12 19
```

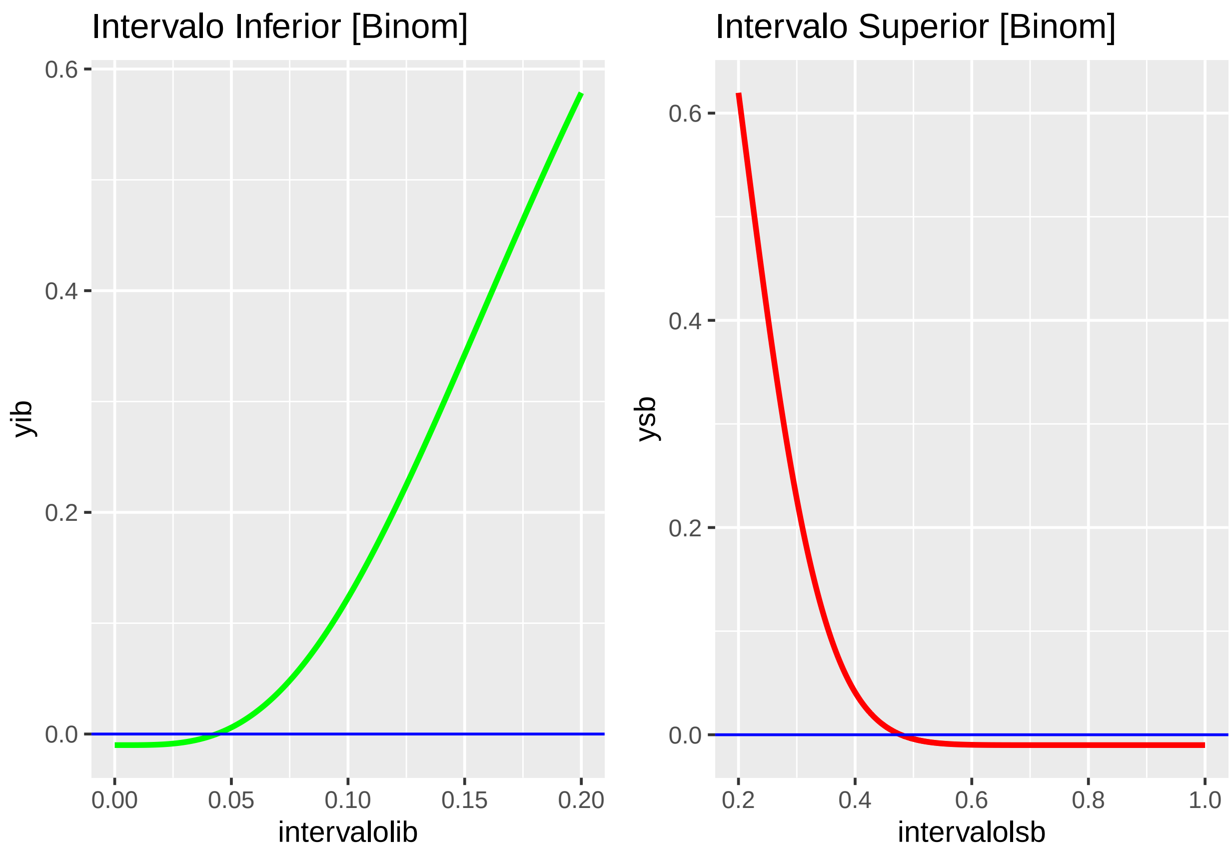
Ahora hacemos un conteo de cuantos hogares reflejan casos de *1*.

```
# Numero de exitos en nuestra muestra
a <- length(which(df[,2] == 1))
print(a)
```

```
## [1] 4
```

Son 4 hogares dentro de la muestra que presentan la propiedad de interés.

Ya teniendo todo listo podemos proceder para hallar el I.C.




```
## [1] "#  pp   Error comparado  con alfa/2 = 0.01"

##         [,1]          [,2]
##  [1,] 0.4779  8.832299e-05
##  [2,] 0.4780  6.473857e-05
##  [3,] 0.4781  4.120214e-05
##  [4,] 0.4782  1.771361e-05
##  [5,] 0.4783 -5.727088e-06
##  [6,] 0.4784 -2.912004e-05
##  [7,] 0.4785 -5.246534e-05
##  [8,] 0.4786 -7.576304e-05
##  [9,] 0.4787 -9.901325e-05
```

```
## [1] "#  pp   Error comparado  con alfa/2 = 0.01"
```

```
##        [,1]          [,2]
## [1,] 0.2435 -9.050622e-05
## [2,] 0.2436 -1.199806e-05
## [3,] 0.2437  6.692002e-05
```

Escogiendo el menor (de la columna 2) de cada uno de las 2 matrices mostradas anteriormente tenemos que, nuestro **Intervalo de Confianza** por medio de la aproximación binomial es:

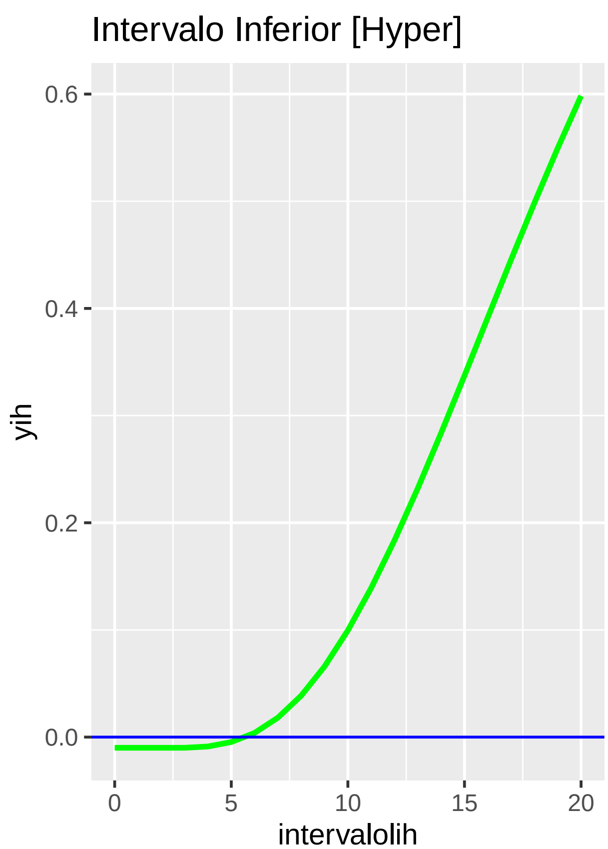
```
## [1] "( 0.2436 , 0.4783 )"
```

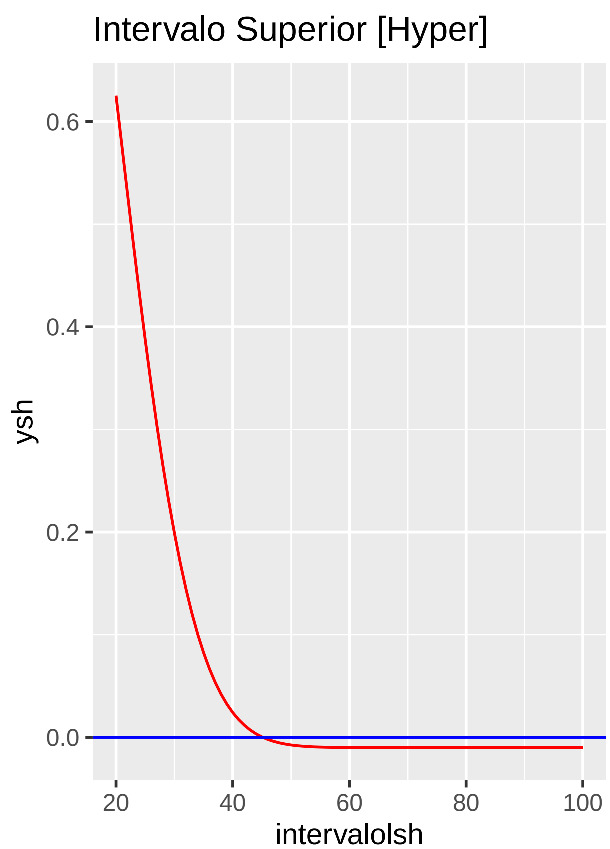
Aprox Hyper

```
# Agregamos unas nuevas funciones que dependan de la distribución Hypergeometrica:
Ls.conf.hiper.prop <- function(A,a,N,n,alfa){
  phyper(a,A,N-A,n)-alfa*0.5
}

Li.conf.hiper.prop <- function(A,a,N,n,alf){
  1-phyper(a-1,A,N-A,n)-alf/2
}
```

ya habiendo definido de antemano nuestros valores `a`, `n`,`N` y $\alpha = 0.02$ procedemos a replicar el codigo de "manera similar" al anterior:





```
## [1] "#  pp   Error comparado  con alfa/2 = 0.01"
```

```
##        [,1]          [,2]
##  [1,]   43  0.0071235023
##  [2,]   44  0.0033811968
##  [3,]   45  0.0003673043
##  [4,]   46 -0.0020384635
##  [5,]   47 -0.0039414386
##  [6,]   48 -0.0054327565
##  [7,]   49 -0.0065903529
##  [8,]   50 -0.0074801133
##  [9,]   51 -0.0081571049
## [10,]   52 -0.0086668328
## [11,]   53 -0.0090464739
## [12,]   54 -0.0093260533
## [13,]   55 -0.0095295385
## [14,]   56 -0.0096758350
## [15,]   57 -0.0097796749
## [16,]   58 -0.0098523946
## [17,]   59 -0.0099026059
## [18,]   60 -0.0099367630
## [19,]   61 -0.0099596360
## [20,]   62 -0.0099746994
## [21,]   63 -0.0099844456
## [22,]   64 -0.0099906335
## [23,]   65 -0.0099944837
## [24,]   66 -0.0099968281
## [25,]   67 -0.0099982228
## [26,]   68 -0.0099990318
## [27,]   69 -0.0099994884
## [28,]   70 -0.0099997386
## [29,]   71 -0.0099998713
## [30,]   72 -0.0099999392
## [31,]   73 -0.0099999725
## [32,]   74 -0.0099999882
## [33,]   75 -0.0099999952
## [34,]   76 -0.0099999982
## [35,]   77 -0.0099999994
## [36,]   78 -0.0099999998
## [37,]   79 -0.0099999999
## [38,]   80 -0.0100000000
## [39,]   81 -0.0100000000
## [40,]   82 -0.0100000000
## [41,]   83 -0.0100000000
## [42,]   84 -0.0100000000
```

```
## [1] "#  pp   Error comparado  con alfa/2 = 0.01"
```

```
##      [,1]         [,2]
## [1,]   24 -0.008764417
## [2,]   25 -0.004645806
## [3,]   26  0.003916569
```

Escogiendo el menor (de la columna 2) de cada uno de las 2 matrices mostradas anteriormente tenemos que, nuestro **Intervalo de Confianza** por medio de la aproximación hypergeometrica es:

```
## [1] "( 0.26 , 0.45 )"
```

Punto 3

Utilizando la aproximación binomial:

```
## [1] "## Así un intervalo de confianza está dado por:"
```

```
## ( 25 , 48 )
```

Utilizando la aproximación hypergeometrica.

```
## [1] "## Así un intervalo de confianza está dado por:"
```

```
## ( 26 , 45 )
```

Punto 4

Para concluir se toma como criterio la longitud del intervalo, con esto podemos observar su presición por tanto el intervalo más preciso es:

- para el caso de la **binomial**: Restamos

```
## [1] "#Prop"
```

```
## [1] 0.2347
```

```
## [1] "# Total"
```

```
## [1] 23
```

- para el caso de la **hipergeometrica**:

```
## [1] "#Prop"
```

```
## [1] 0.19
```

```
## [1] "# Total"
```

```
## [1] 19
```

Viendo los anteriores resultados el caso de la Aprox. hipergeometrica tiene menor longitud en contra parte con el de la aprox. binomial por tanto el mejor intervalo para estimar el parametro de la proporción y el total sera el de la aprox. hipergeometrica.