

ARBOLES DE DECISION ASOCIADOS AL CONTAGIO EN COLOMBIA AL 11 ABRIL DE 2020

Matemática, Administradora Pública y Magister Científico-Estadística

Analítica y Datos –Beatriz García Peña



1. Objetivo

Determinación de árboles de decisión para el contagio positivo, los fallecidos y los recuperados, según edad y grupos de edad, a nivel nacional, de Bogotá y para el resto de Colombia; utilizando la base de datos de contagio positivo del Instituto Nacional de Salud, la cual fue consultada en su sitio web oficial el 12 de abril de 2020 a las 18:44 p.m.

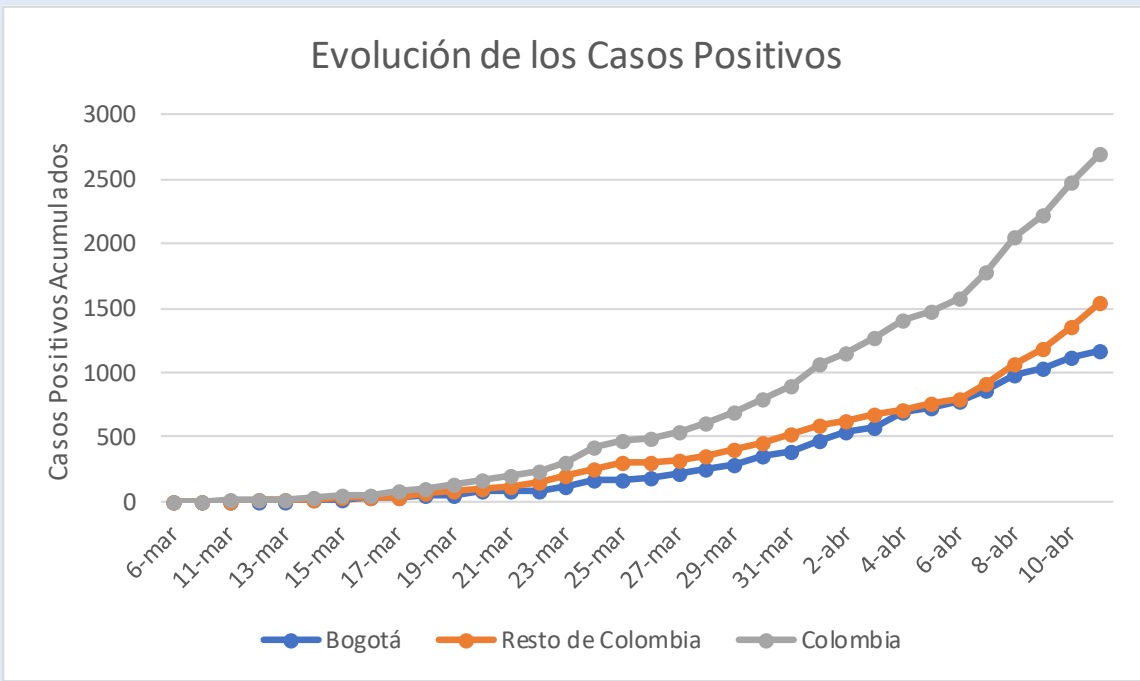
2. Palabras claves

Arboles de decisión, Prueba de Bondad de Ajuste, Prueba Ji-Cuadrado, Inferencia Estadística, Instituto Nacional de Salud, DANE, Base de datos de contagio positivo del Instituto Nacional de Salud, Beatriz García Peña, Analítica y Datos

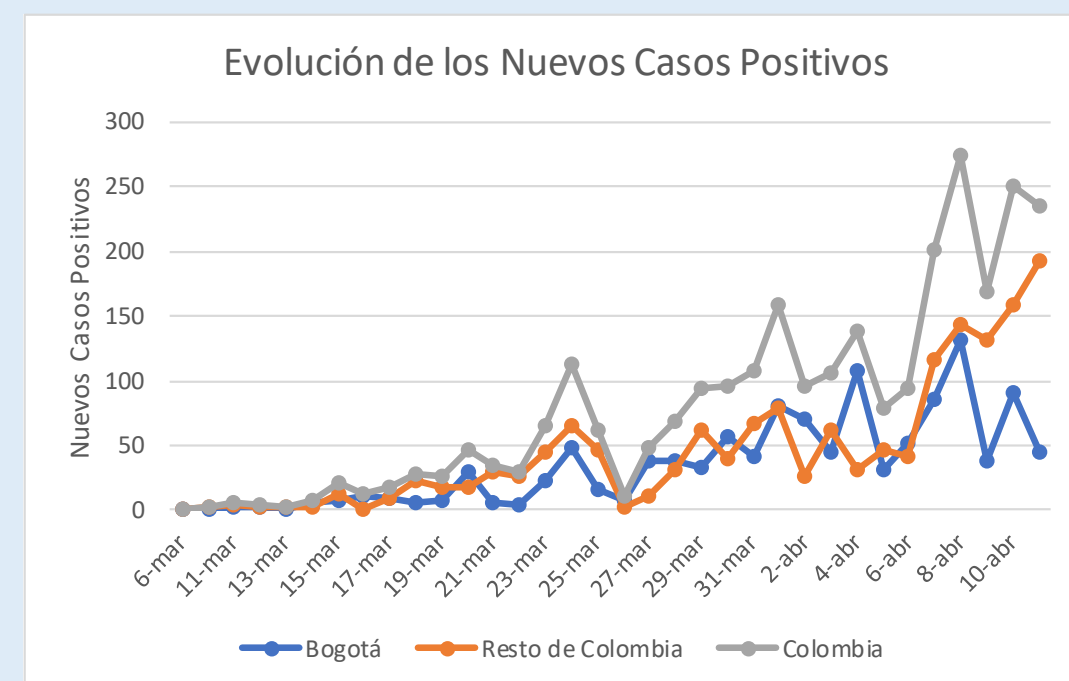
3. Metodología

- Acceso y validación de la base de datos
- Elaboración de tablas de frecuencias para Bogotá, el resto de Colombia y Colombia para las variables contagio positivo, fallecidos y recuperados por género y grupos de edad
- Cálculo y diagramación de árboles de decisión
- Elaboración de las respectivas pruebas de hipótesis, utilizando la distribución χ^2 , prueba de bondad de ajuste, para establecer si las frecuencias objeto de estudio son iguales o diferentes; utilizando niveles de significancia del 0.05, 0.02 y 0.01

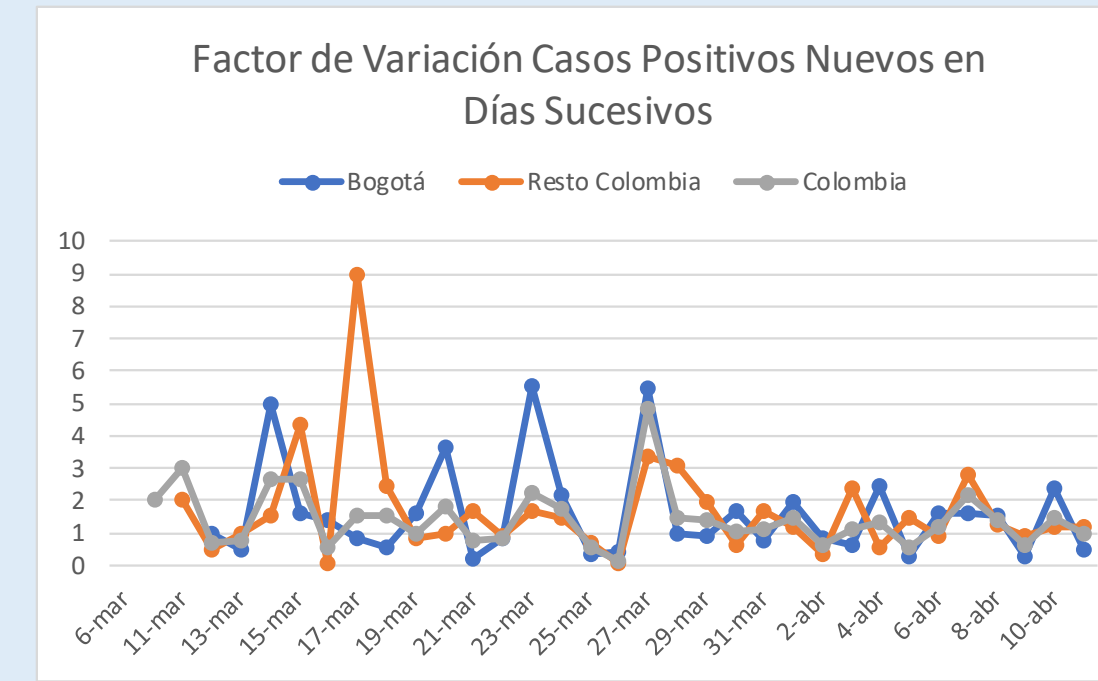
4. Evolución de los Casos Positivos



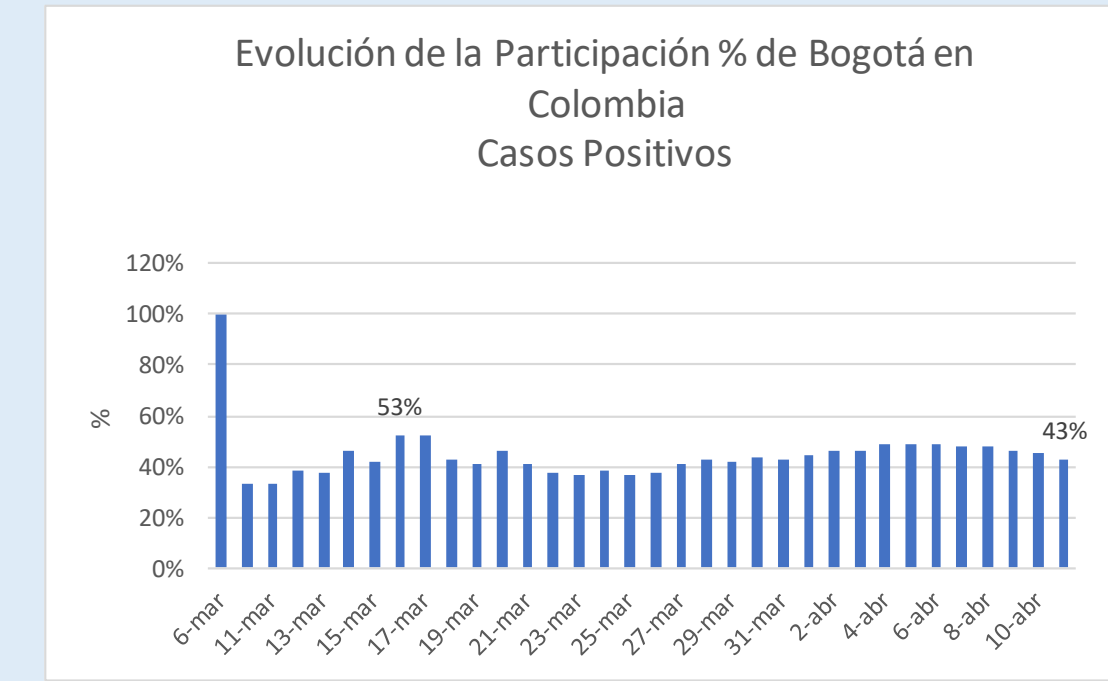
La evolución de los casos positivos de contagio acumulados en Colombia crece en promedio 50,5 cada día en el periodo comprendido entre el 6 de marzo y el 11 de abril de 2020. La contribución de Bogotá y el resto del país es muy similar en el periodo considerado solamente en los últimos 3 días se observa despejarse de manera significativa al resto de Colombia de la curva de Bogotá.



Los mayores nuevos casos positivos de contagio en Colombia entre el 6 de marzo y el 11 de abril de 2020 son superiores a 200 y máximos el 8 de abril con 274. En Bogotá y el resto del país los máximos corresponden a 131 y 192 ocurridos el 8 de abril y 11 de abril, respectivamente.

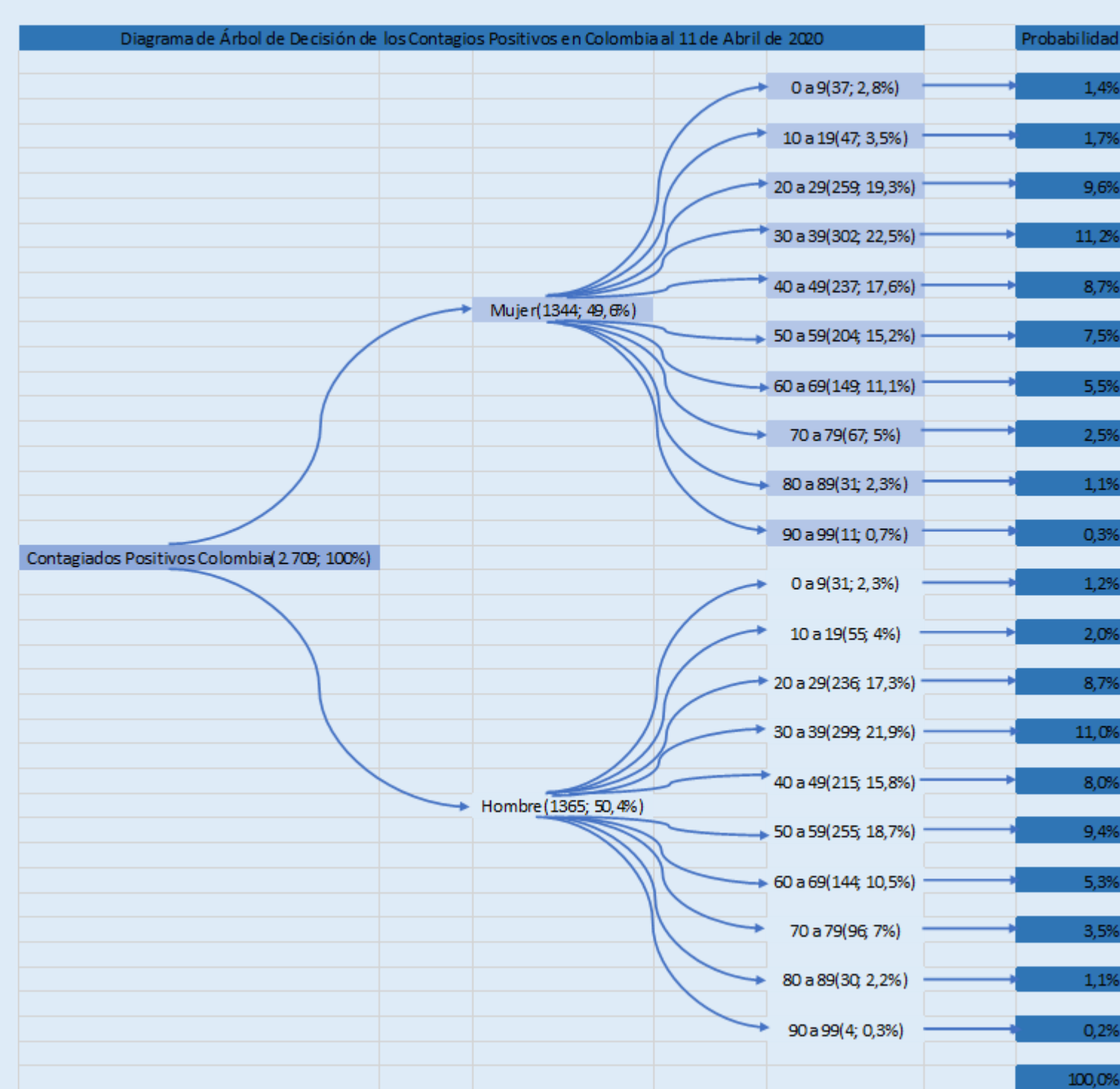


El factor promedio de variación de casos positivos nuevos en días sucesivos para Colombia es de 1,42 con una desviación estándar de 0,91, mientras que para Bogotá es de 1,59 con una desviación estándar de 1,46 y para el resto de Colombia es de 1,69 con una desviación estándar de 1,65. Es significativamente inferior el factor promedio de Bogotá y con menor variabilidad en relación con el del resto de Colombia.

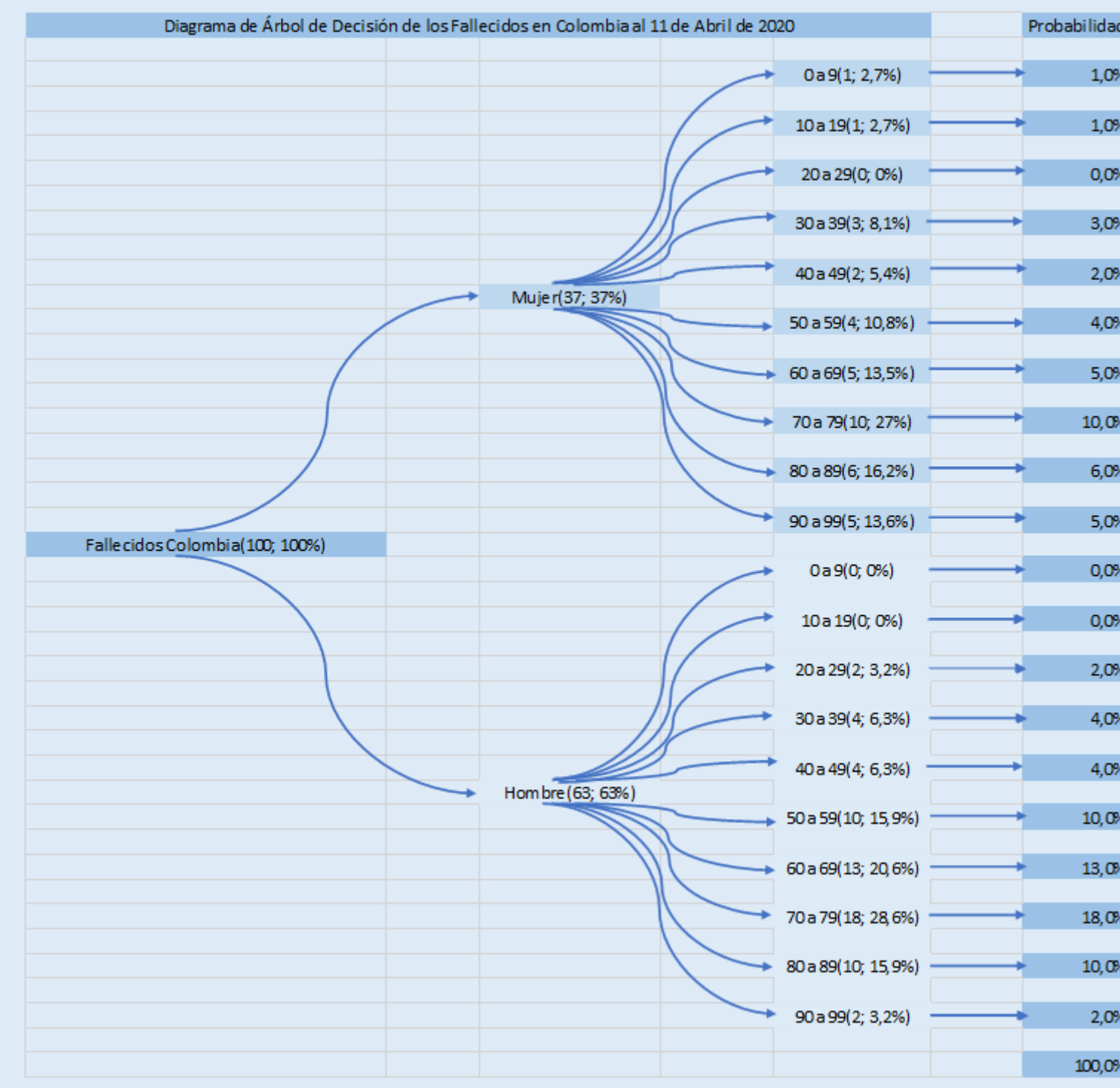


El % de contagios positivos correspondiente a Bogotá corresponde en promedio al 44% de los contagios positivos de Colombia en el periodo comprendido entre el 6 de marzo y el 11 de abril de 2020 y es el mayor % correspondiente a las distintas ciudades y departamentos del país.

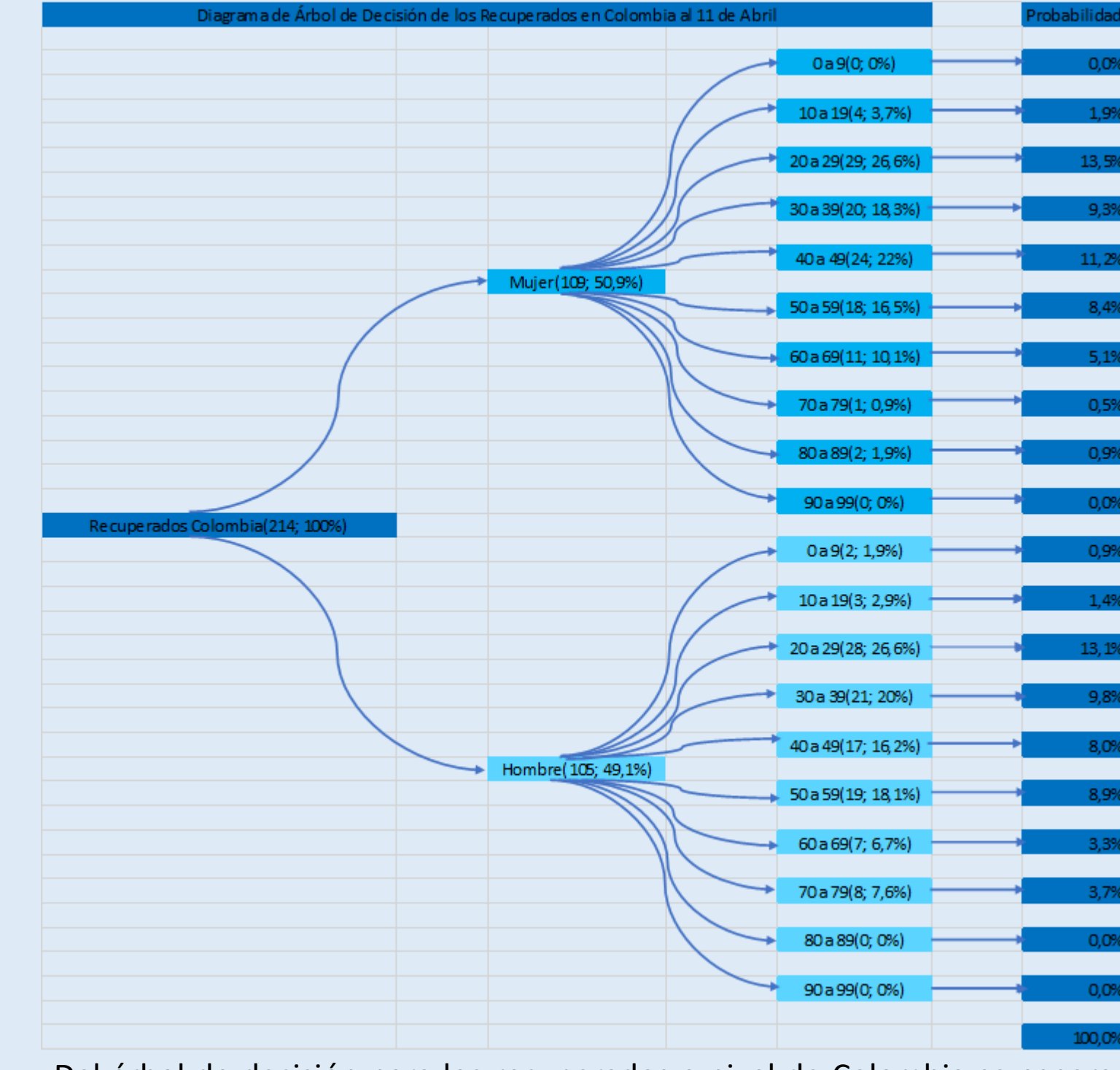
5. Probabilidad de Contagio, Fatalidad y Recuperación en Colombia



El árbol de decisión para el contagio positivo colombiano muestra las probabilidades para la combinación de género y grupos de edad y dado que se realizó con 2.709 casos disponibles al 11 de Abril, se espera que las probabilidades sean estables. La mayor probabilidad de contagio positivo en Colombia es para mujeres de 30 a 39 años, correspondiente al 11,2%.

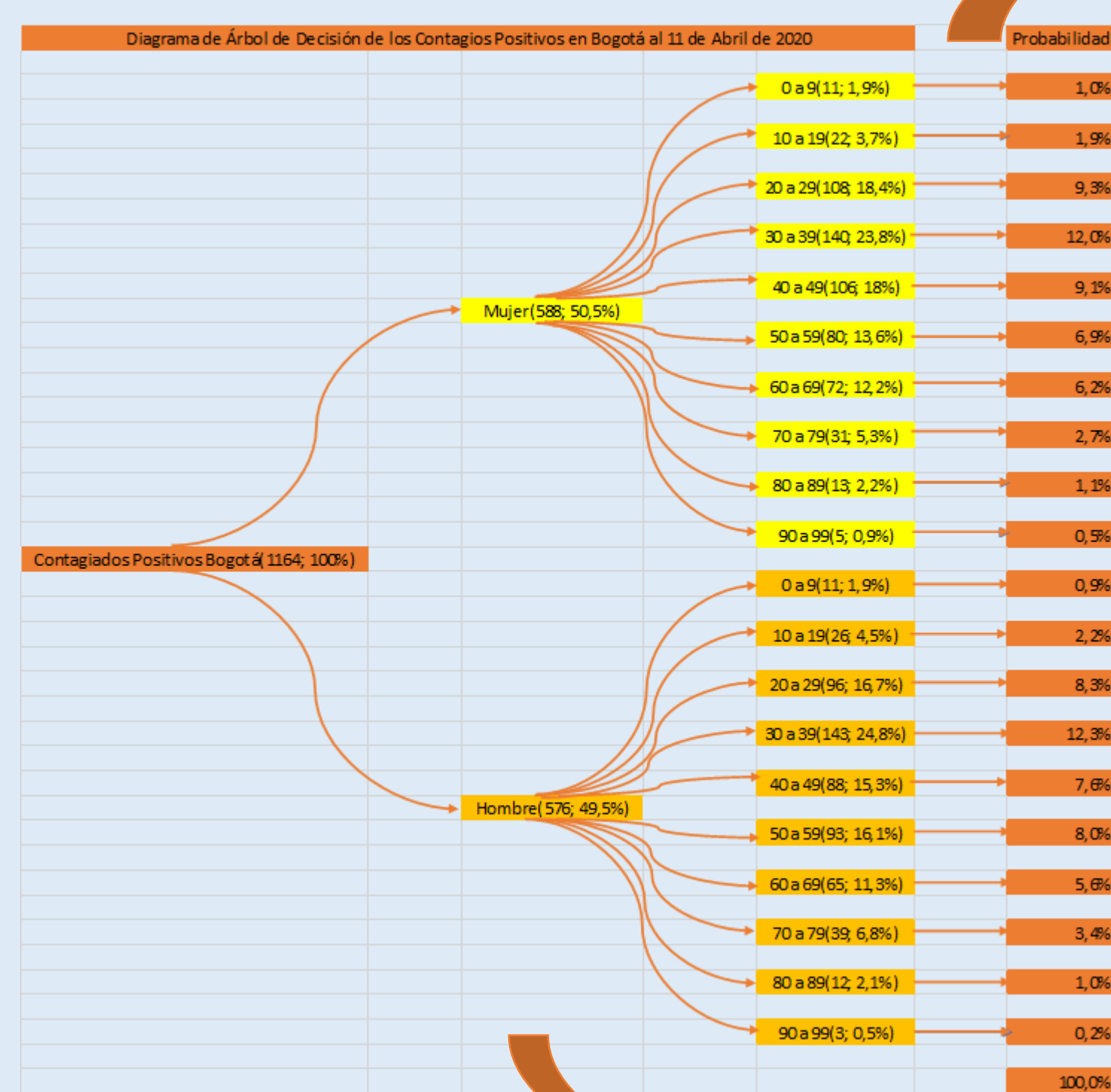


El árbol de decisión de los fallecidos por el contagio positivo a nivel de Colombia puede mostrar ligeras variaciones ya que se elaboró con los 100 casos disponibles al 11 de Abril. La mayor probabilidad de fallecer es para los hombres cuya edad corresponde al grupo de 70 a 79 años.



Del árbol de decisión para los recuperados a nivel de Colombia se espera que sus probabilidades sean relativamente estables ya que se elaboró con los casos disponibles al 11 de Abril y estos ascienden a 214 casos. La mayor probabilidad de recuperación en Colombia es para las mujeres del grupo de 20 a 29 años y esta asciende a 13,5%.

6. Probabilidad de Contagio Positivo para Bogotá y para el Resto de Colombia



La distribución del contagio según género y grupos de edad para Bogotá es significativamente diferente de la distribución de la población en Bogotá según género y grupos de edad en el 2020 con una probabilidad de error del 1% y $\chi^2 = 306,9^{***}$ con 19 grados de libertad.

El árbol de decisión de contagio de Bogotá es significativamente diferente del árbol de contagio del resto de Colombia con una probabilidad de error del 1% y $\chi^2 = 37,5^{***}$ con 19 grados de libertad.

Población Bogotá 2020		
Clase	Frecuencia	%
Mujer, 0 a 9	477.952	6,2%
Mujer, 10 a 19	514.910	6,6%
Mujer, 20 a 29	732.813	9,5%
Mujer, 30 a 39	656.159	8,5%
Mujer, 40 a 49	542.429	7,0%
Mujer, 50 a 59	499.193	6,4%
Mujer, 60 a 69	347.918	4,5%
Mujer, 70 a 79	177.725	2,3%
Mujer, 80 a 89	67.142	0,9%
Mujer, 90 a 99*	19.876	0,3%
Hombre, 0 a 9	498.158	6,4%
Hombre, 10 a 19	526.669	6,8%
Hombre, 20 a 29	728.023	9,4%
Hombre, 30 a 39	630.708	8,1%
Hombre, 40 a 49	474.619	6,1%
Hombre, 50 a 59	404.113	5,2%
Hombre, 60 a 69	265.255	3,4%
Hombre, 70 a 79	125.203	1,6%
Hombre, 80 a 89	43.585	0,6%
Hombre, 90 a 99*	11.505	0,1%
Total	7.743.955	100,0%

Fuente: Elaboración Propia
* Incluye la población desde los 90 años

Población Resto Colombia 2020		
Clase	Frecuencia	%
Mujer, 0 a 9	3.367.097	7,9%
Mujer, 10 a 19	3.464.696	8,1%
Mujer, 20 a 29	3.537.452	8,3%
Mujer, 30 a 39	3.160.431	7,4%
Mujer, 40 a 49	2.670.955	6,3%
Mujer, 50 a 59	2.410.954	5,7%
Mujer, 60 a 69	1.711.709	4,0%
Mujer, 70 a 79	928.826	2,2%
Mujer, 80 a 89	372.859	0,9%
Mujer, 90 a 99*	116.446	0,3%
Hombre, 0 a 9	3.520.618	8,3%
Hombre, 10 a 19	3.606.052	8,5%
Hombre, 20 a 29	3.553.568	8,3%
Hombre, 30 a 39	3.023.383	7,1%
Hombre, 40 a 49	2.442.201	5,7%
Hombre, 50 a 59	2.120.630	5,0%
Hombre, 60 a 69	1.470.440	3,4%
Hombre, 70 a 79	772.073	1,8%
Hombre, 80 a 89	293.927	0,7%
Hombre, 90 a 99*	84.152	0,2%
Total	42.628.469	100,0%

Fuente: Elaboración Propia
* Incluye la población desde los 90 años



La distribución de contagio según género y grupos de edad para el resto de Colombia es significativamente diferente de la distribución de la población para el resto de Colombia según género y grupos de edad en el 2020 con una probabilidad de error del 1% y $\chi^2 = 563,8^{***}$ con 19 grados de libertad.

7. Conclusiones y Recomendaciones

- Los árboles de decisión permiten detectar grupos vulnerables según género y grupos de edad para contagios, fallecidos y recuperados. Lo cual aporta información valiosa para los tomadores de decisiones a nivel de la política pública.
- El contagio positivo en Bogotá es significativamente diferente del respectivo en el resto de Colombia, es decir, que el abordaje del problema a estos niveles debe ser diferencial por género y grupos de edad.
- La distribución % del contagio positivo en Bogotá y en el resto de Colombia es diferente de la distribución % de la población Bogotá y del resto de Colombia según género y edad respectivamente.

8. Fuentes de Información

- Instituto Nacional de Salud, Base de datos sobre casos positivos, <https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx>
- DANE, Censo 2018, Estimaciones Proyecciones 1985-2020, <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018>
- EsMateria.com; Periodismo Científico; Nuevas Tecnologías de vacunas, ¿el camino hacia el fin de la COVID-19?, 8 abr 2020
- ElTiempo.com, Lo que dice el primer genoma del coronavirus colombiano, Nicolás Bustamante Hernández, 13 abr 2020