

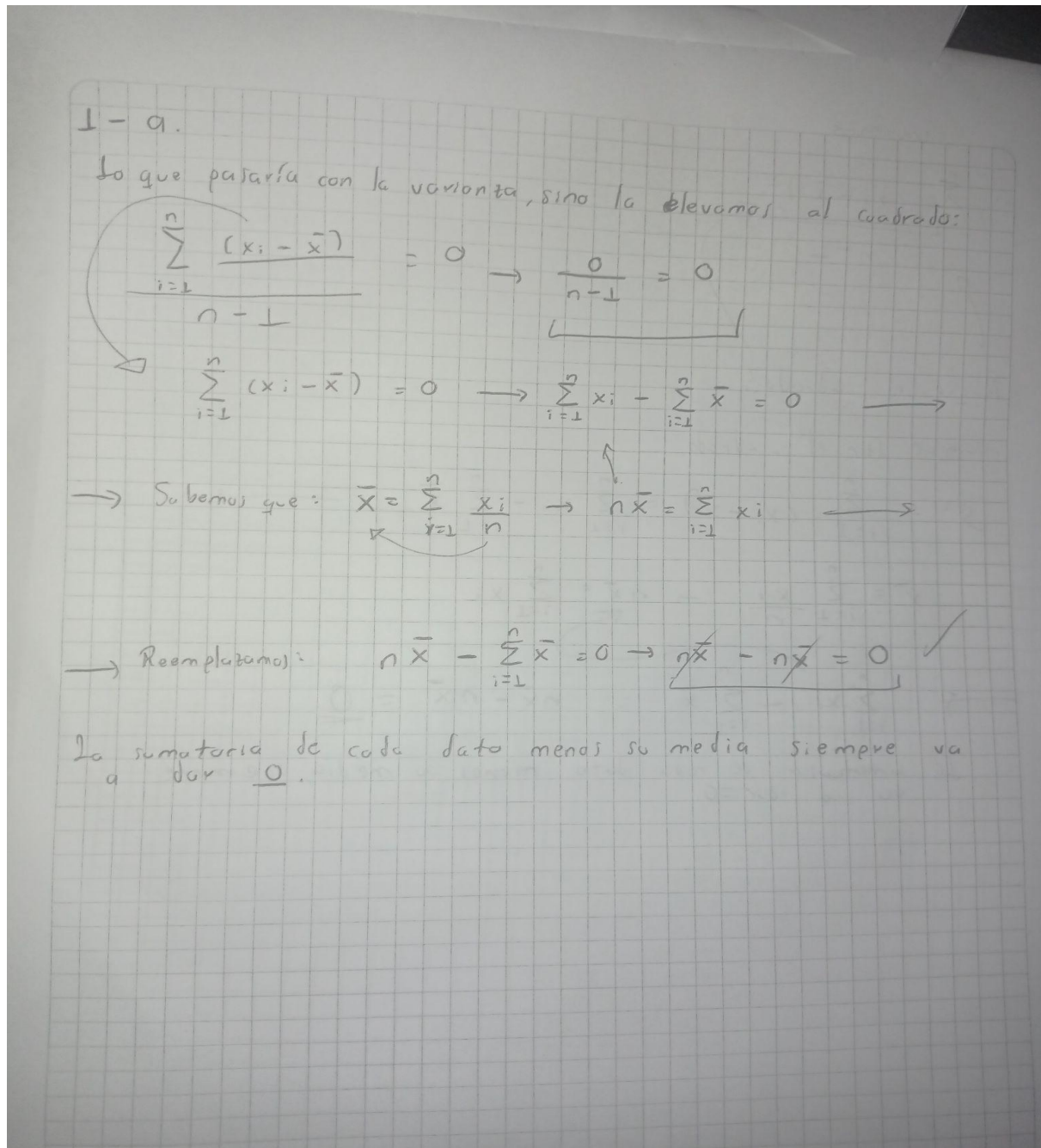
**Parcial Estadística**  
**Juan Sebastian Velasquez Acevedo**  
[juan.velasquez.acevedo@correounivalle.edu.co](mailto:juan.velasquez.acevedo@correounivalle.edu.co)  
Código 1744936-3743

Todo el contenido del parcial también está en el siguiente repositorio de Github:  
[https://github.com/Odzen/Probability\\_Statistics/tree/main/parcial](https://github.com/Odzen/Probability_Statistics/tree/main/parcial)

Parte 1 teórica (50%):

1.

A.



1.B

① = b

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}$$

↓  
→ ignora denominador

$$\rightarrow \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2x_i\bar{x} + \sum_{i=1}^n \bar{x}^2 \rightarrow$$

$$\rightarrow \text{Saco constantes} \rightarrow \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 \rightarrow$$

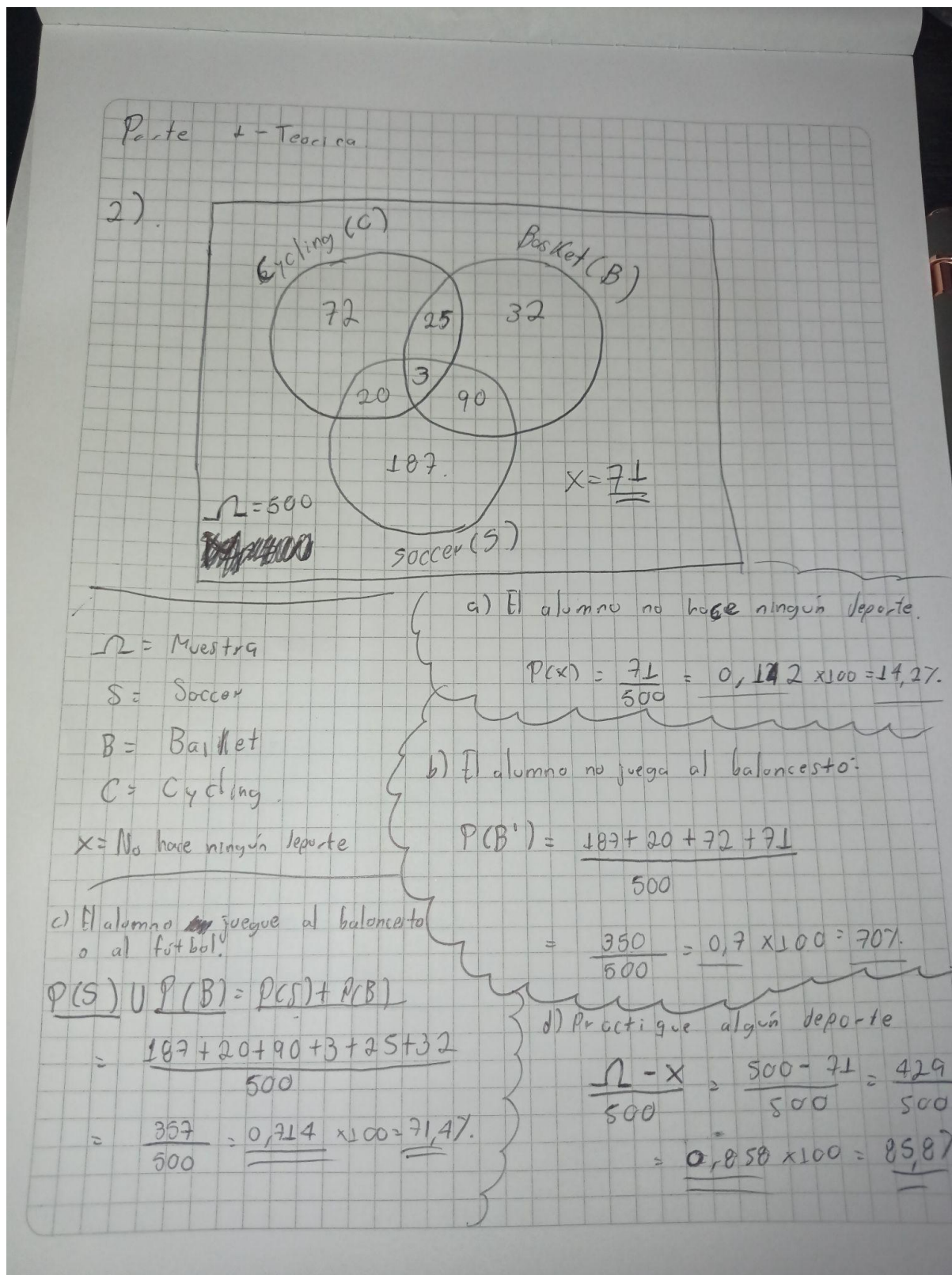
$$\rightarrow \text{Sabemos que: } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \rightarrow n\bar{x} = \sum_{i=1}^n x_i \rightarrow$$

$$\rightarrow \text{Reemplazamos: } \sum_{i=1}^n x_i^2 - 2\bar{x}^2 n + \sum_{i=1}^n \bar{x}^2 \rightarrow$$

$$\rightarrow \text{Resolvemos última sumatoria: } \sum_{i=1}^n x_i^2 - \underline{2\bar{x}^2 n} + \underline{n\bar{x}^2} \rightarrow$$

$$\rightarrow \text{Operamos términos semejantes: } \underline{\underline{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}$$

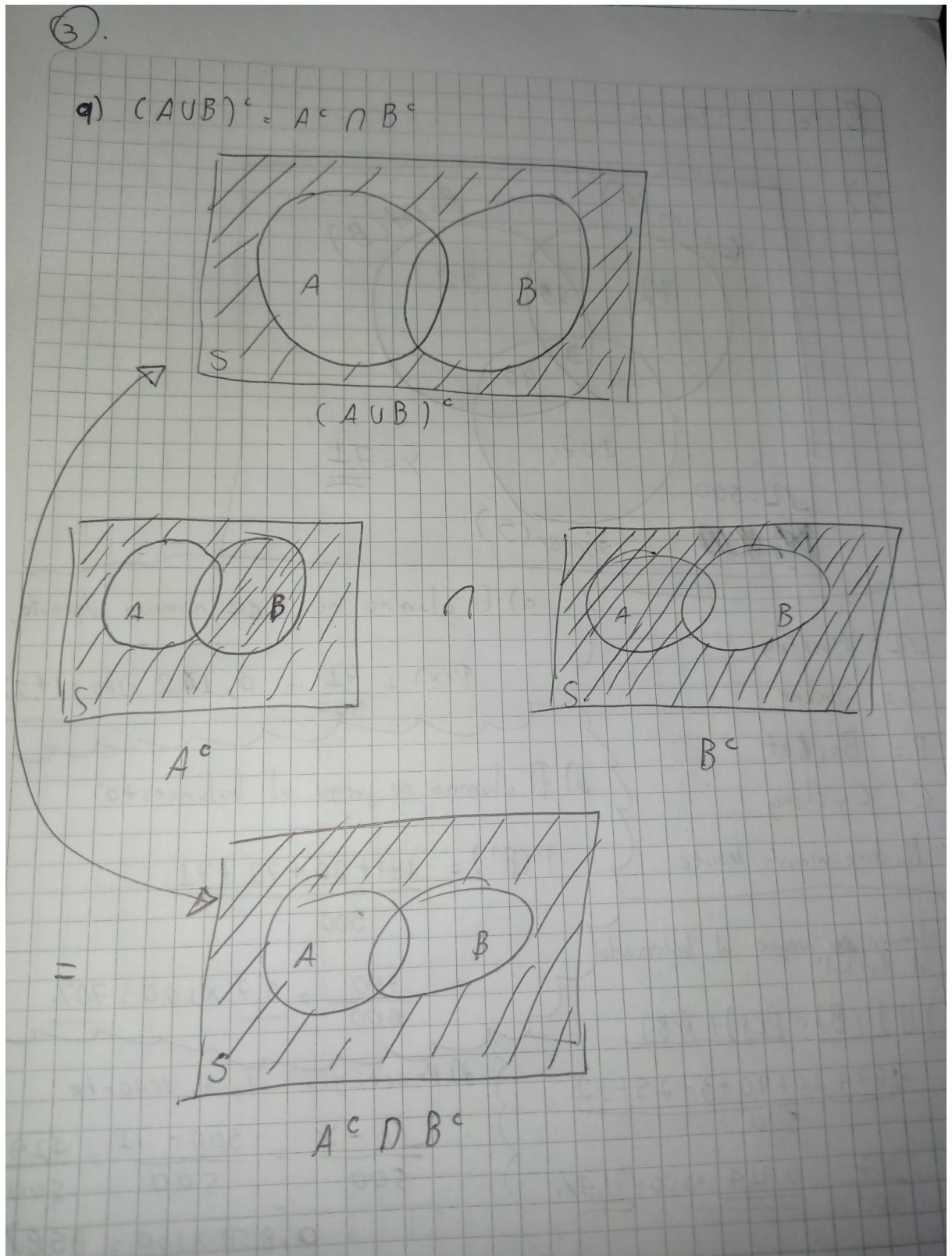
2.





3.

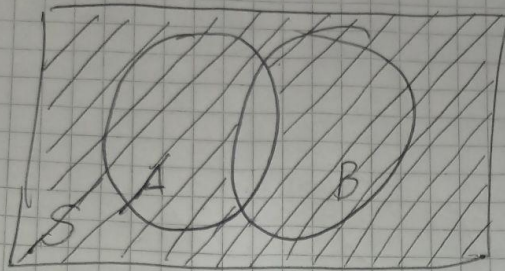
A.



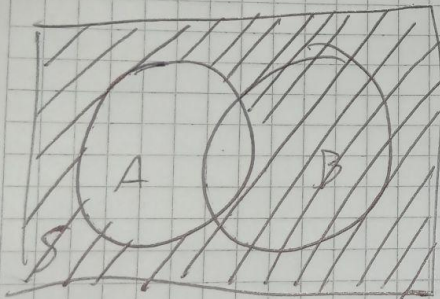
3.B.

③

$$6. (A \cap B)^c = A^c \cup B^c$$

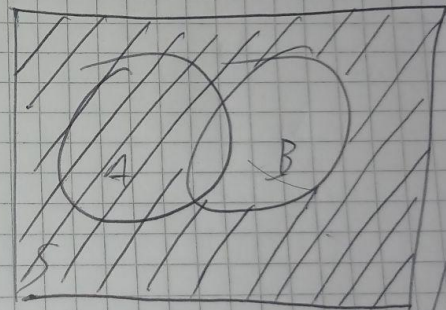


$(A \cap B)^c$



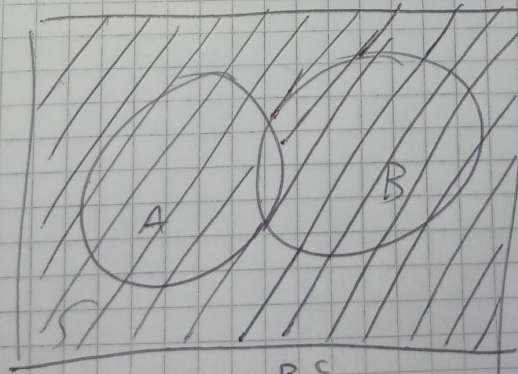
$A^c$

$\cup$



$B^c$

$=$



$A^c \cup B^c$



4.

$$4) A) P(E_1) = P(E_2) = 0,15$$

$$P(E_3) = 0,4$$

$$P(E_4) = 2P(E_5)$$

$$P(E_4) = ? = 0,2$$

$$P(E_5) = ? = 0,1$$

$$P(E_1) + P(E_2) + P(E_3) + P(E_4) + P(E_5) = 1$$

$$0,15 + 0,15 + 0,4 + 2P(E_5) + P(E_5) = 1$$

$$0,3 + 0,4 + 3P(E_5) = 1$$

$$0,7 + 3P(E_5) = 1$$

$$P(E_5) = \frac{1 - 0,7}{3} = \underline{\underline{0,1}} \times 100 = \underline{\underline{10\%}}$$

$$P(E_4) = 2P(E_5) = 2(0,1) = \underline{\underline{0,2}} \times 100 = \underline{\underline{20\%}}$$

$$B) P(E_1) = 3P(E_2) = 0,3$$

$$P(E_3) = P(E_4) = P(E_5)$$

$$3P(E_2) = 0,3$$

$$P(E_1) = 3(0,1)$$

$$P(E_1) + P(E_2) + P(E_3) + P(E_4) + P(E_5) = 1$$

$$P(E_2) = \frac{0,3}{3}$$

$$P(E_1) = 0,3$$

$$P(E_1) + P(E_2) + P(E_3) + P(E_4) + P(E_5) = 1$$

$$0,3 + 0,1 + 3P(E_3) = 1$$

$$\underline{\underline{P(E_2) = 0,1}}$$

$$P(E_3) = \frac{1 - 0,4}{3}$$

$$R// P(E_3) = P(E_4) = P(E_5) = \underline{\underline{0,2}}$$

$$P(E_3) = \frac{0,6}{3} = \underline{\underline{0,2}} \times 100 = \underline{\underline{20\%}}$$

5.

⑤.

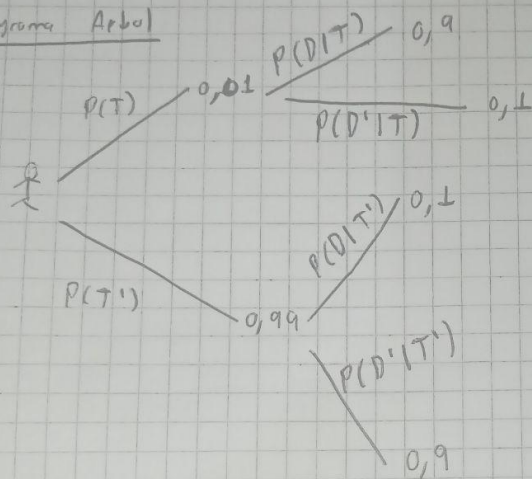
$P(T)$  = Tiene la enfermedad

$P(T')$  = No tiene la enfermedad.

$P(D)$  = Prueba detecta que la tiene

$P(D')$  = Prueba detecta que no la tiene.

Diagrama Arbol



Probabilidad de que tenga en realidad la enfermedad:

$$P(T|D) = \frac{P(T \cap D)}{P(D)}$$

$$= \frac{P(T) P(D|T)}{P(D)}$$

$$P(T) = 0,01$$

$$P(D|T) = 0,9$$

$$P(D) = (0,01 \times 0,9) + (0,99 \times 0,1)$$

$$\rightarrow P(T|D) = \frac{(0,01)(0,9)}{(0,01 \times 0,9) + (0,99 \times 0,1)} = \frac{0,009}{0,009 + 0,099} = \frac{0,009}{0,108}$$

$$\rightarrow \frac{0,0833}{1} \times 100 \approx 8,3\%$$

Solo el 8,3% de las personas que están enfermas en realidad recibieron un diagnostico correcto. Es por esto que, la prueba diagnostica no es confiable, la probabilidad es muy baja para llegar a un resultado verdadero.

Sorprende bastante este resultado, ya que por culpa de un mal diagnostico pueden haber muchas personas enfermas que no lleguen a recibir un tratamiento, por culpa de este examen incorrecto.

## Parte 2 práctica (50%):

### Código en R, Link Github:

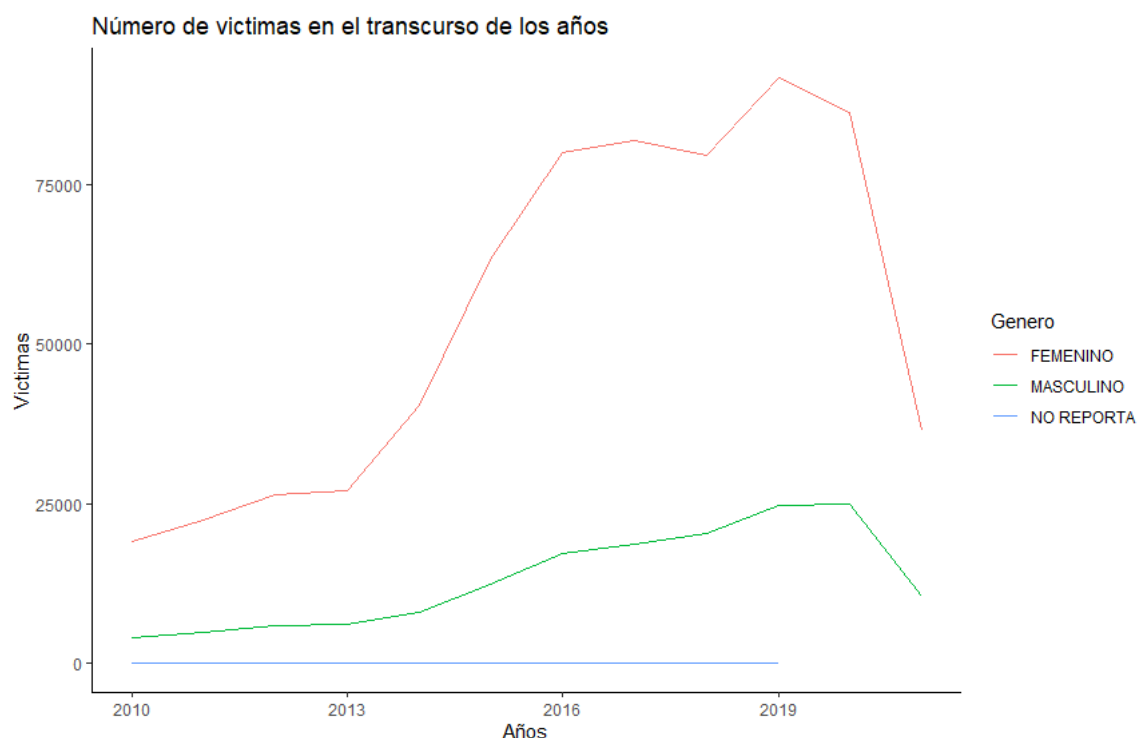
[https://github.com/Odzen/Probability\\_Statistics/tree/main/parcial/part2](https://github.com/Odzen/Probability_Statistics/tree/main/parcial/part2)

Para el análisis de los datos presentados en el archivo se usó el software R, se leyó el archivo .csv y luego se usó la librería tidyverse para ciertas manipulaciones y formatos de los datos con dataframes que me permitieron adecuar los datos para los diferentes tipos de gráficos que planté, con el fin de darle un mayor entendimiento y facilitar el análisis y lectura.

Lo primero que hicimos en el código, fue formatear la fecha y crear una nueva columna que me mostrase solo el año. Luego de esto transformé los años, que eran strings, a números. Esto fue necesario para poder usarlo en diferentes gráficos, sobre todo en la serie de tiempo. Al parecer la variable para el eje x debería ser numérica, para que geom\_line sepa cómo conectar los puntos para dibujar la línea.

### Gráficas:

#### 1. Serie de tiempo



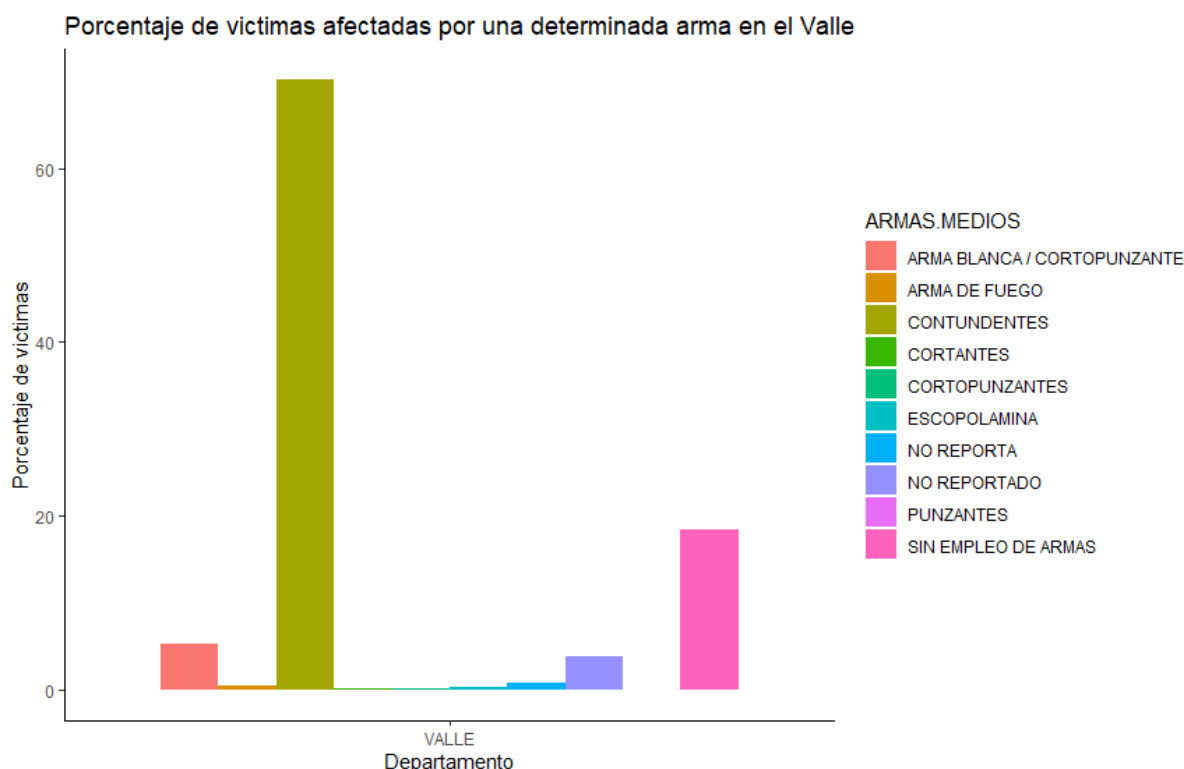
**Análisis:** En este caso, se decidió construir un gráfico de tipo serie de tiempo. En donde me mostrase el número de víctimas de violencia doméstica a través de los



años, también marcando los casos por género, ya sea masculino, femenino o no reportado. Podemos observar que:

- El género femenino siempre ha sido más violentado en el transcurso de los años estudiados. Las víctimas masculinas siempre son menos, pero también tienen un número significativo. Y los casos en que el género no es reportado, son tan poco, que pueden ser despreciables para este estudio, pues comparado con el resto de víctimas, se aproxima a 0.
- Los años más violentos fueron el 2019 y 2020.
- Vemos que con el tiempo, la violencia doméstica ha incrementado con creces en Colombia, y a pesar que en el último año haya bajado un poco, es preocupante la situación en nuestro país con respecto a este tema.

## 2. Gráfico de barras:

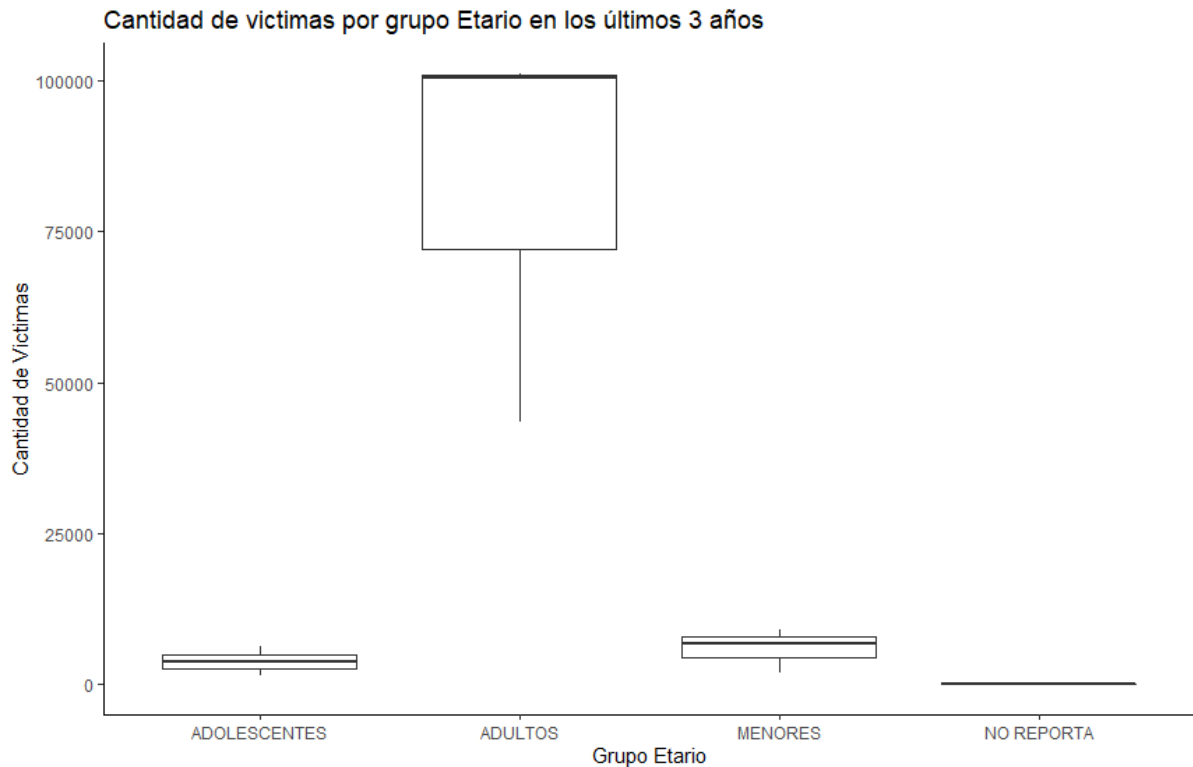


**Análisis:** Esta vez usamos un gráfico de barras, para mostrar el porcentaje de víctimas afectadas con alguna determinada arma, en el departamento del Valle del Cauca. En el gráfico nos podemos dar cuenta de:

- La mayoría de víctimas, por mucho, en el Valle, fueron atacadas con armas contundentes.
- Muchas otras víctimas, fueron atacadas sin armas, dando a entender que para que haya violencia doméstica, no necesariamente tiene que haber un arma física de por medio.
- Luego podemos observar, que en la violencia doméstica en el Valle, también hay un porcentaje en el que los agresores usan armas blancas.

- d. En comparación con los casos anteriores, hay muy poco porcentaje de víctimas que son agredidas con escopolamina u otras armas, y otras cuantos casos en donde no reportan el arma que usó el agresor.

### 3. Gráfico de cajas:

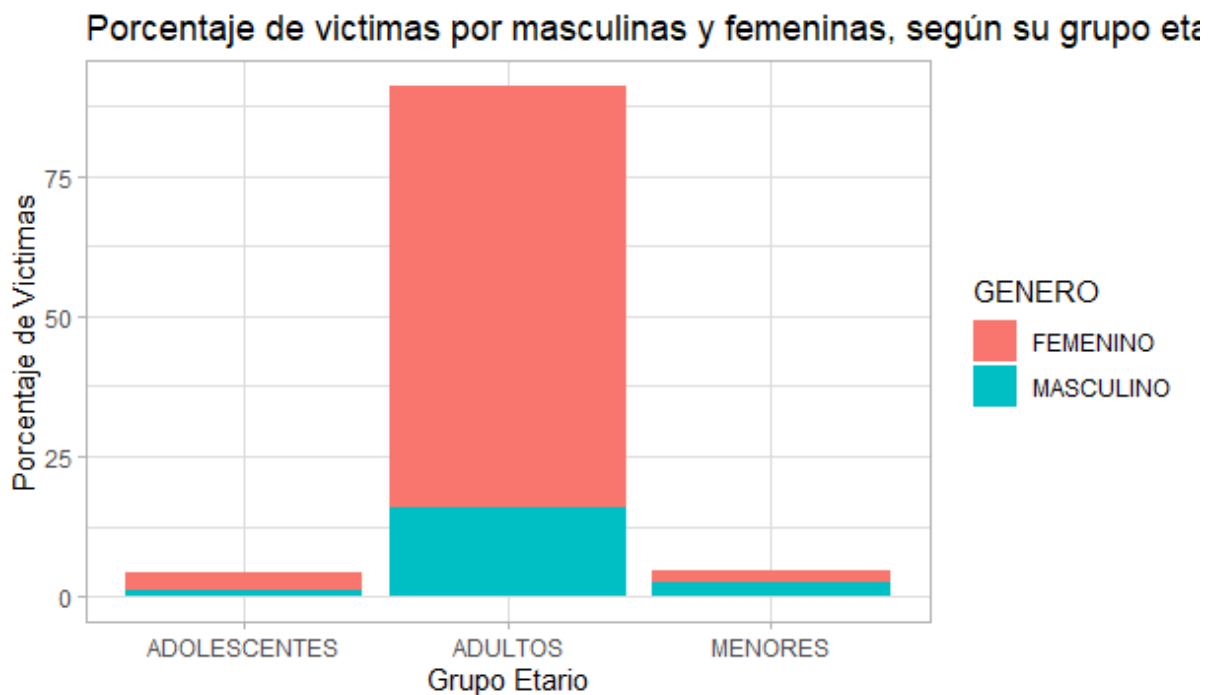


**Análisis:** En este caso, a través de un gráfico de cajas, se muestra la cantidad de víctimas de violencia doméstica, por grupo etario en los últimos 3 años (2019, 2020 y 2021) en el país. El gráfico nos muestra que:

- La mayoría de víctimas son adultos, su caja también me muestra que hay cierta varianza a través de los últimos años, puesto que su mínimo está muy lejos de su máximo, además de que su mediana está muy arriba en la caja lo que significa que el histograma asociado a esta caja presenta una asimetría a la derecha.
- De los adultos, el segundo grupo etario más afectado en los últimos años fue el de los menores de edad. A pesar de que el número de víctimas no es tanto como el de los adultos, es preocupante ver que muchos menores sufren de violencia doméstica en Colombia. En este caso, la caja muestra una varianza un poco menor entre los cuantiles. Y su mediana es más simétrica que en la caja de adultos.
- Luego vienen los adolescentes maltratados en este caso podemos ver que la cantidad de adolescentes se asemeja un poco a la de menores, aunque en este caso la gráfica nos muestra que tiene un comportamiento casi perfectamente simétrico, no ha variado mucho a través de los últimos años.

- d. Y los casos donde no se reporta el grupo etario, son casi despreciables y no tienen gran varianza, casi nula.

#### 4. Barras Apiladas:



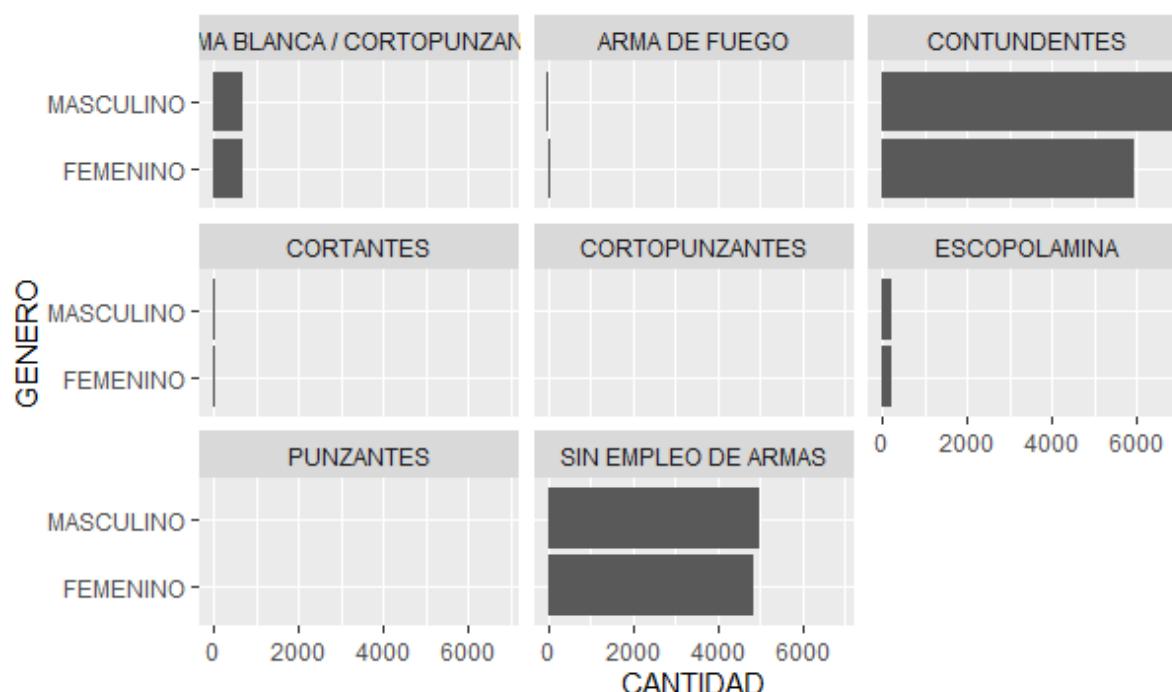
**Análisis:** Realizamos un gráfico de barras apiladas con referencias cruzadas, para mostrar el porcentaje de víctimas afectadas según su grupo etario, y su género. Ignorando los casos no reportados puesto que no alteran demasiado los resultados y para el análisis pueden ser despreciables. Según el gráfico, nos podemos dar cuenta de:

- Como ya vimos anteriormente, los adultos es el grupo etario más violentado, pero entre ellos, el mayor porcentaje es de género femenino, aproximadamente un 80%. Aunque, por el volumen de víctimas totales, el porcentaje restante, también representa una gran cantidad de hombres adultos que sufren violencia doméstica.
- Luego, en los menores la proporción entre niñas y niños que sufren violencia doméstica en Colombia es casi la misma.
- En los adolescentes, es más el porcentaje de víctimas mujeres que de víctimas hombres.
- Como podemos ver en los tres grupos etarios. La mayoría de los casos, el género femenino es el más afectado. Pero esto no le quita peso, al hecho que en ningún caso, la cantidad de víctimas masculinas es despreciable. La violencia doméstica se presenta en Colombia para los dos géneros.



- e. Al ver que los adultos son por mucho el grupo más afectado, podría haber alguna conexión entre estos adultos y sus hijos. Ya que en una misma casa, podrían haber varios signos de violencia doméstica, entre la pareja, sus hijos, etc. Esto podría afectar psicológicamente a estos niños, un problema social bastante grande en Colombia.

## 5. Gráfico de barras cruzadas:

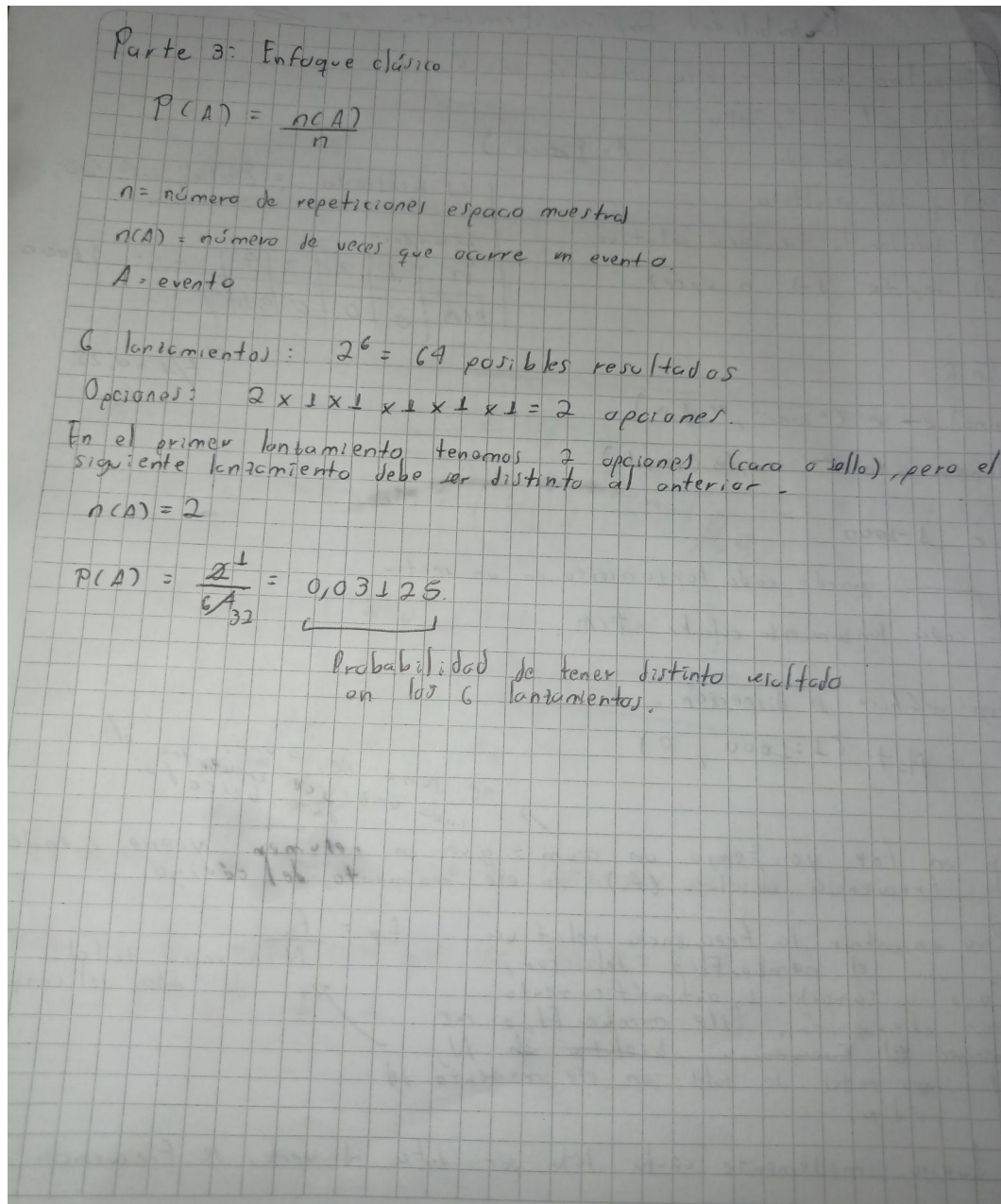


**Análisis:** Está gráfica representa la cantidad de niños y niñas que fueron atacados por una determinada arma. En este caso podemos observar que:

- a. Las armas más usadas son las contundentes, lo que preocupa, puesto que es inaudito que una cantidad tan grande de menores esté sufriendo de violencia doméstica y además usando armas.
- b. Tanto niños como niñas, casi en igual cantidad, son atacados sin armas en sus casas en Colombia.
- c. La tercera arma más común que es usada en Colombia para practicar violencia doméstica, son las armas cortopunzantes y la cuarta es la escopolamina.
- d. El resto de armas las usan poco, es casi despreciable la cantidad de niños víctimas en estos casos.
- e. Aquí en este rango de edad, grupo etario de menores, se rompe la tendencia un poco. Puesto que en otros grupos etarios, el género más damnificado usualmente siempre es el femenino más que el masculino. Y en esta última gráfica, podemos ver que en todos los casos, la violencia hacía los menores de edad en sus casas es más marcada en el género masculino. Más a los niños que a las niñas, aunque no hay tampoco una diferencia muy notoria.

### Parte 3

#### Enfoque Clásico



En R:

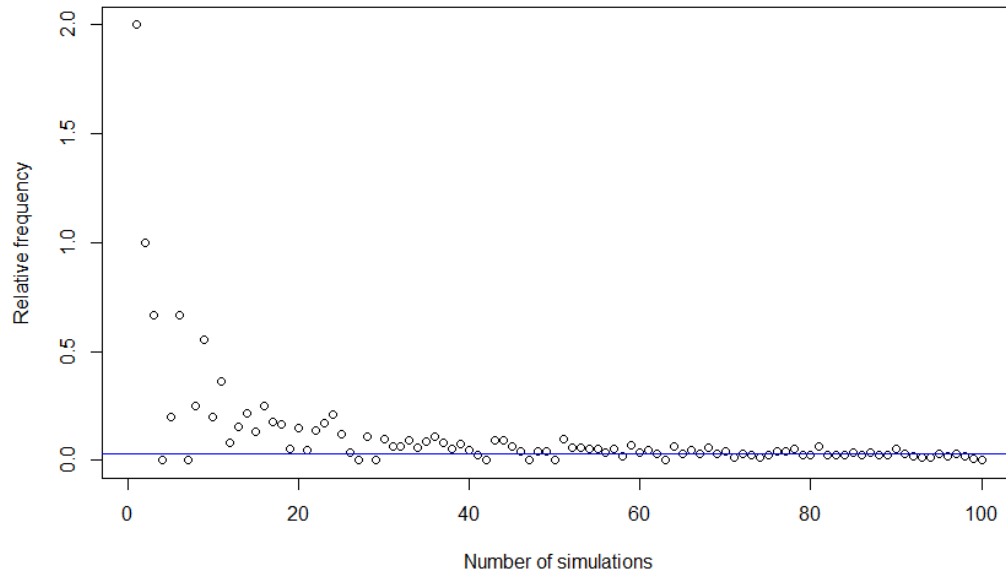
Código en GitHub:

[https://github.com/Odzen/Probability\\_Statistics/tree/main/parcial/part3](https://github.com/Odzen/Probability_Statistics/tree/main/parcial/part3)

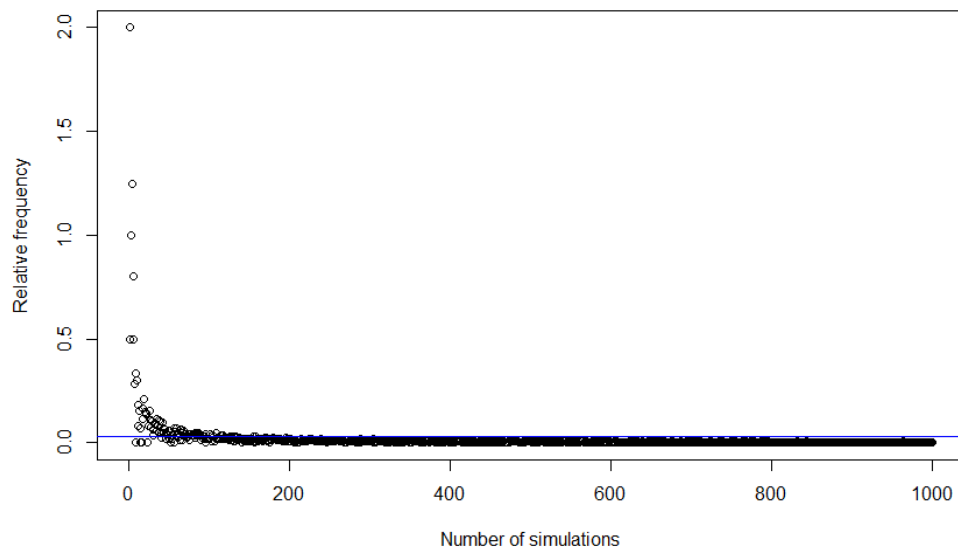
Para este ejercicio no usé la función `rle()` como se aconsejaba en las notas del parcial, sino que me enfoqué un poco más en seguir el funcionamiento similar a un código que ya había usado anteriormente para comprobar la probabilidad de que al lanzar una moneda de cara o

sello. Usando ciclos For. Esto hace un poco ineficiente mi algoritmo, pero teóricamente está fundamentado y funciona para los objetivos de este ejercicio práctico

### Gráfico 100 simulaciones

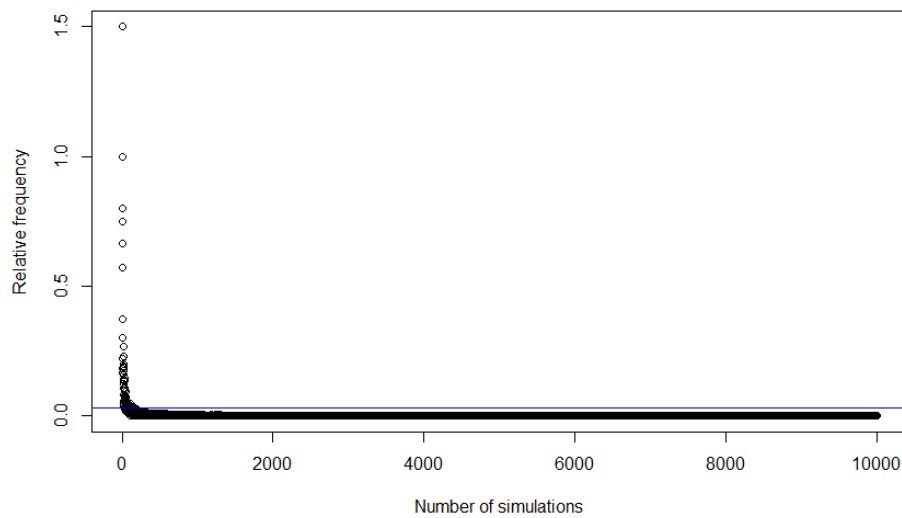


### Gráfico 1.000 simulaciones



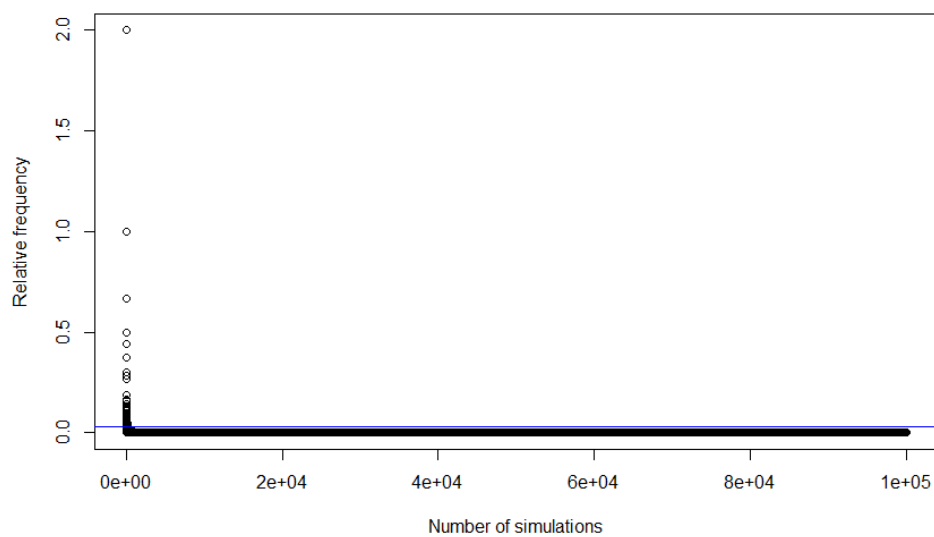


### Gráfico 10.000 simulaciones



### Gráfico 100.000 simulaciones

En este caso, por la complejidad de mi algoritmo y el número de ciclos que hace. Se demora un poco en ejecutar.



**Conclusión:** El valor teórico calculado con R y visualizado con los gráficos coincide con el valor dado por el enfoque clásico = 0.03125.