

Matemáticas Computacionales

Practica 2

Cynthia Ivanna Cruz Quiñones

Matricula: 1854499

Grupo: 002

A 7 de Marzo del 2021

Índice

1. Introducción	3
2. Base de Datos HairEyeColor	3
2.1. Descripción del conjunto de datos	3
2.2. Estadística descriptiva de una variable	4
2.3. Estadística descriptiva de dos variable	6
3. Conclusión	8
4. Referencias	9

1. Introducción

En esta practica se estudiara una base de datos con estadística descriptiva en R, se analizaran los tipos datos que nos proporciona la misma y sus atributos, para asi mostrar gráficos y tablas del procesamiento derivado o implicado de la investigación.

2. Base de Datos HairEyeColor

La base de datos HairEyeColor[1] que contiene el color del pelo, color de ojos y sexo de 592 estudiantes de la University of Delaware, por Snee en el año de 1974. Después, en el año de 1992, fue realizada la separacion por sexo por Friendly, para una mejor intervencion o desglosamiento de la información.

Esta base de datos es ilustrativa para varias técnicas de análisis de contingencia de tablas, generalmente por metodos gráficos, como los diagramas de mosaico, diagramas de asociación, modelo de log-lineal, chi-squared tes, entre otros.

2.1. Descripción del conjunto de datos

El dataset "HairEyeColor" cuenta con 32 combinaciones entre tipos de color de cabello y ojos entre el género masculino y femenino.

Cuenta con las siguientes variables:

- **Hair:** Black, Brown, Red, Blond.
- **Eye:** Brown, Blue, Hazel, Green.
- **Sex:** Male, Female.

Estas son clasificadas como variables las cuales se atribuyen a la frecuencia en la cual sucede tal coincidencia entre el color de ojos y cabello.

En la figura1 se observa la estructura mosaico, la cual es la mas generalizada para el uso de este tipo de base de datos; en esta se representan los datos prolíficamente[3].

Diagrama de mosaico de la tabla bidimensional de frecuencia de colores de cabello y ojos

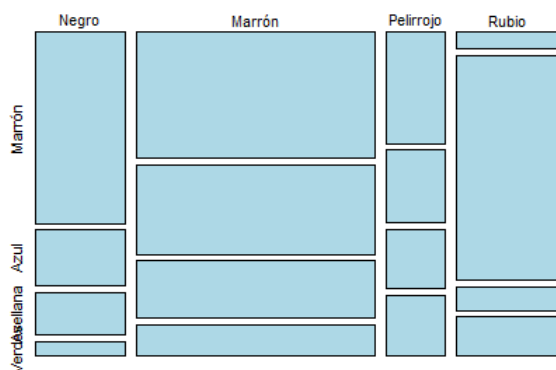


Figura 1: Representación grafica de mosaico

2.2. Estadística descriptiva de una variable

En las siguientes gráficas se muestra la variación que hay entre el color de cabello 2a, color de ojos 2b y el género mayor volube en cuanto a distincion del color de ojos y cabello2c, ya que estas son una variación.

Esté tipo de gráficas visualiza la distribución de datos en un intervalo o período de tiempo continuo. Este gráfico es una variación de un Histograma que usa el suavizado de cerner para trazar valores, permitiendo distribuciones más suaves al suavizar el ruido. Los picos de un gráfico de densidad ayudan a mostrar dónde los valores se concentran en el intervalo.

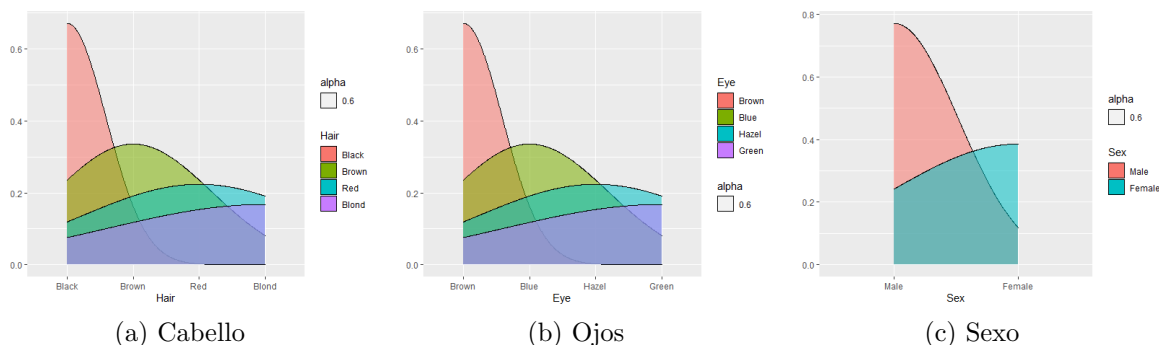


Figura 2: Density plot de cabello, ojos y sexo

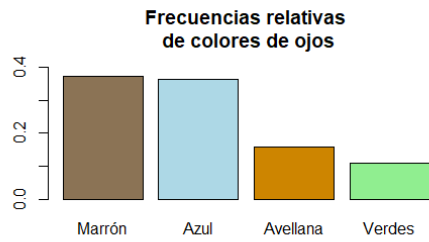
Su objetivo es extraer una tabla bidimensional de frecuencias absolutas de las variables Eye y hair, sin distinguir según el sexo. La manera más fácil de obtener esta tabla es sumando las subtabla de frecuencias para hombres y mujeres, y aplicando `as.table()` al resultado para transformarlo en una tabla por si no lo es.

Hair	Eye	Sex	Freq
Black:8	Brown:8	Male:16	Min.:2.00
Brown:8	Blue :8	Female:16	1st Qu.: 7.00
Red:8	Hazel:8		Median :10.00
Blond:8	Green:8		Mean:18.50
			3rd Qu.:29.25
			Max.:66.00

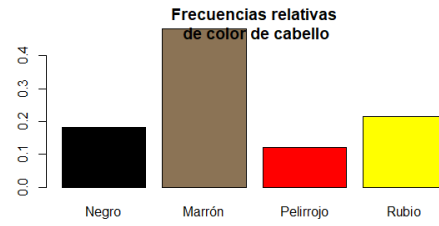
Cuadro 1: Frecuencia mediana, minima y maxima

En la tabla se puede observar que la minima es de 2.00, una mediana de 10.00 y una maxima de 66.00, con el primer lugar de 7.00, segundo de 18.50 y tercero de 29.25.

De esto se desglosa sus respectivas gráficas de frecuencia relativa entre la cantidad de color de ojos 3a y el color de cabello 3b.



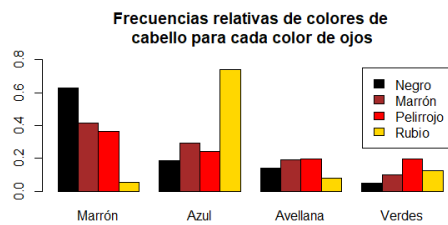
(a) Ojos



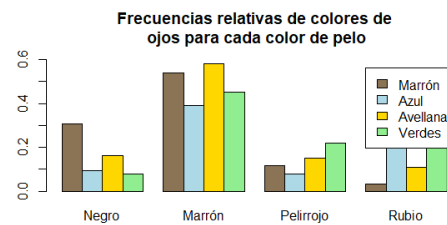
(b) Cabello

Figura 3: Frecuencia de color de cabello y de ojos

En comparacion a las gráficas anteriores, se muestran la correlatividad entre el color de cabello para cada color de ojos 4a y las frecuencias relativas de colores de ojos para cada color de cabello 4b.



(a) Ojos



(b) Cabello

Figura 4: Frecuencia correlativa entre el color de cabello y de ojos

2.3. Estadística descriptiva de dos variable

Boxplot de correlación [4]

El diagrama de caja es un gráfico utilizado para representar una variable cuantitativa (variable numérica). El gráfico es una herramienta que permite visualizar, a través de los cuartiles, cómo es la distribución, su grado de asimetría, los valores extremos, la posición de la mediana, etc.

En este apartado se muestran tres Boxplot en relación a los datos Sexo 5a, cabello 5b y ojos 5c

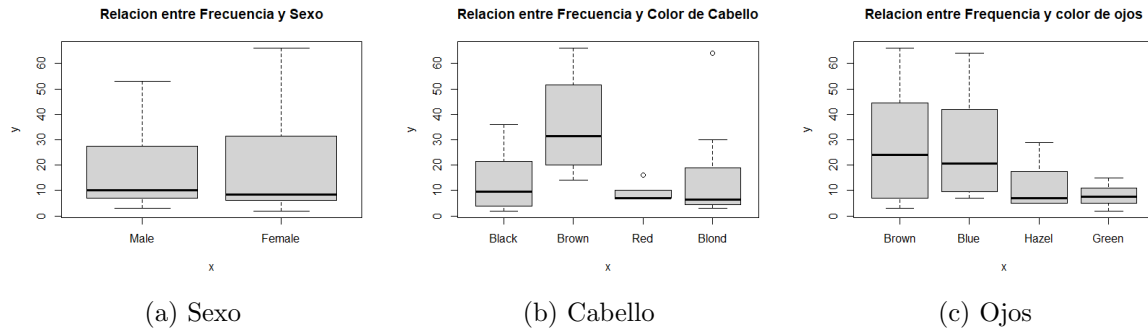


Figura 5: Box de correlación entre frecuencia y las variables

De estos tres diagramas de caja, se muestra que las mujeres tienen un mayor índice en comparación a los hombres en cuanto al color del cabello y ojos, mientras que el color de cabello con más índice es el café al igual que el color de los ojos.

A continuación se muestran otros dos diagramas que filtran esta información por género, en el cual el cabello cuenta con mayor índice de mujeres, las cuales tienen el cabello castaño, mientras que los hombres destacan con el color de ojos café y azul. Lo cual nos indica que el sexo masculino tiene un mayor índice de variación de iris 6b y las mujeres cuentan con mayor índice de cabello castaño 6a.

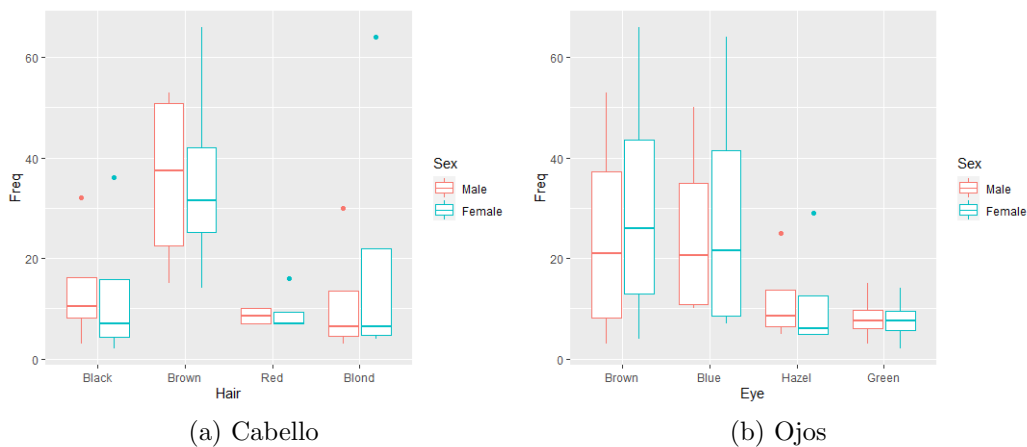


Figura 6: Boxplot de correlación entre género y color de cabello y ojos

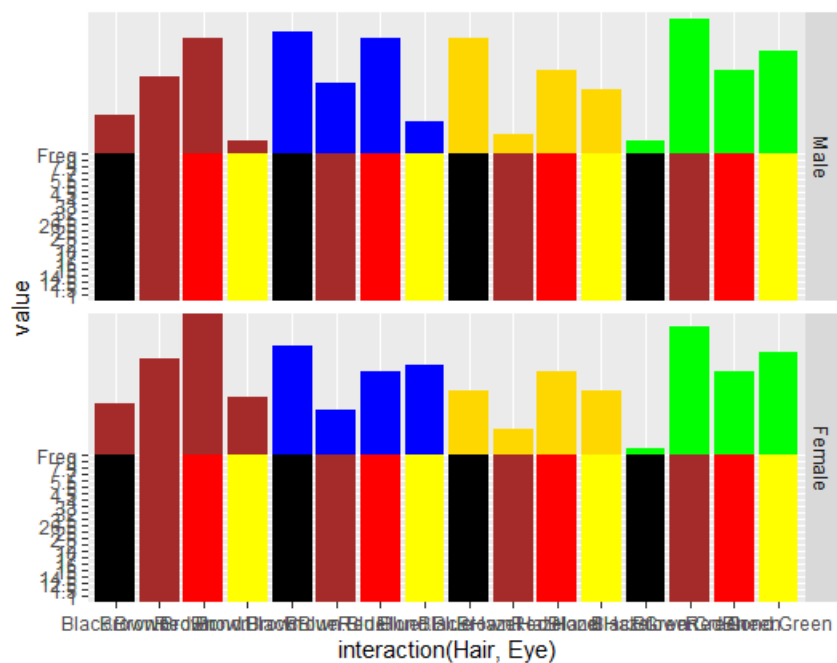


Figura 7: Escala de variación

Escala [fig7] de variación entre género y frecuencia en el que se presentan los distintos colores de cabello junto con el color de ojos [2].

A continuación se desglosan dos tablas de las cuales se presenta su respectiva representación gráfica escalar; dichas tablas se dividen en 2, por el género masculino 2 y género femenino 3.

Hair	Eyes Brown	Eyes Blue	Eyes Hazel	Eyes Green
Black	32	11	10	3
Brown	53	50	25	15
Red	10	10	7	7
Blond	3	30	5	8

Cuadro 2: Tabla relativas entre cabello y ojos de género masculino

Hair	Eyes Brown	Eyes Blue	Eyes Hazel	Eyes Green
Black	36	9	5	2
Brown	66	34	29	14
Red	16	7	7	7
Blond	4	64	5	8

Cuadro 3: Tabla relativas entre cabello y ojos de género femenino

3. Conclusión

La base de datos de HairEyesColor, se usa como referencia a las mediciones o frecuencia en que existe la relacion entre el color de cabello y ojos de las personas, llevando un listado de cada persona que fue medido y relacionado a este tipo.

Los datos proporcionados por la base de datos fue de muha ayuda ya que, al momento de buscar y realizar el filtramiento de la informacion para su representación gráfica, se tenia un mejor manejo de la informacion y su categorización por sexo.

Se mostraba un mejor manejo de datos, pero a la vez que esta era mas eficiente en su distribución, se tiene un menor rango para realizar un análisis mas profundo de diagramas.

Es interesante la investigación de dataset en R, al igual que sus implicaciones en estadística, es manejable para numerosas cantidades de información de las cuales se puede tener un mayor indice de error y ser frustantes.

Se tuvo que realizar un exhaustivo procesamiento de códigos, ya que en si este tipo de base de datos se maneja mas por gráficas de mosaico y de indice, lo cual ocasionó que navegara en distintas fuentes y coementarios de usuarios en base a dudas o optimisación para la creación de escalas didacticas.[5]

4. Referencias

Referencias

- [1] R core R-core@R project.org. Hair And Eye Color Of Statistics Students. <https://www.rdocumentation.org/packages/datasets/versions/3.6.2/topics/HairEyeColor>, 2021.
- [2] <https://stackoverflow.com/users/2994949/rawr>. HairEyeColor bar chart in R. <https://stackoverflow.com/questions/26394927/haireyecolor-bar-chart-in-r>, 2014.
- [3] JClave. HairEyeColor bar chart in R. <https://www.rpubs.com/JoanClaverol/485799>, 2019.
- [4] Danny Lwetable. RPubs, HairEyeColor. <https://rpubs.com/Dannylwe/178349>, 2016.
- [5] Cynthia Ivanna Cruz Quinones. Repositorio de Github. <https://github.com/CynthiaCruzqn/MatematicasComputacionales.git>, 2021.