

**INFORME LABORATORIO 3:
ANÁLISIS ESTADÍSTICO**

DIEGO POLANCO BERRIOS, JUAN ROA CARVAJAL, DANY RUBIANO JIMÉNEZ

Profesores:

- Mónica Villanueva Ilufi
- Felipe Bello Robles

TABLA DE CONTENIDOS

ÍNDICE DE FIGURAS.....	v
ÍNDICE DE CUADROS	vi
CAPÍTULO 1. INTRODUCCIÓN.....	7
1.1 MOTIVACIÓN Y ANTECEDENTES	7
1.2 OBJETIVOS	7
1.3 ORGANIZACIÓN DEL DOCUMENTO	7
CAPÍTULO 2. MARCO TEÓRICO.....	9
2.1 TIPO DE POBLACIÓN	9
2.2 DEFINICIÓN DE TIPOS DE VARIABLES	9
2.3 ESTADÍSTICA UNIDIMENSIONAL	10
2.4 MEDIDAS DE CENTRALIZACIÓN	12
2.5 MEDIDAS DE DISPERSIÓN O MEDIDAS DE VARIABILIDAD	12
2.6 RELACIÓN FUNCIONAL ENTRE DOS VARIABLES	13
CAPÍTULO 3. RESULTADOS.....	15
3.1 DESTELLOS DE HADRÓN	17
3.1.1 Histogramas para las variables	17
3.1.2 Estadísticas para las variables	22
3.1.3 Moda para las variables	22
3.1.4 Correlación entre las variables	23
3.2 DESTELLOS GAMMA	23
3.2.1 Histogramas para las variables	23
3.2.2 Estadísticas para las variables	28
3.2.3 Moda para las variables	29
3.2.4 Correlación entre las variables	29
CAPÍTULO 4. ANÁLISIS	31
4.1 DESTELLOS DE HADRÓN	31

4.1.1	Histogramas para las variables	31
4.1.1.1	Histograma para fLength en los destellos de hadrones	31
4.1.1.2	Histograma para fWidth en los destellos de hadrones	31
4.1.1.3	Histograma para fSize en los destellos de hadrones	31
4.1.1.4	Histograma para fConc en los destellos de hadrones	31
4.1.1.5	Histograma para fConc1 en los destellos de hadrones	31
4.1.1.6	Histograma para fAsym en los destellos de hadrones	31
4.1.1.7	Histograma para fM3Long en los destellos de hadrones	31
4.1.1.8	Histograma para fM3Trans en los destellos de hadrones	32
4.1.1.9	Histograma para fAlpha en los destellos de hadrones	32
4.1.1.10	Histograma para fDist en los destellos de hadrones	32
4.1.2	Estadísticas para las variables	32
4.1.3	Moda para las variables	32
4.1.4	Correlación entre las variables	32
4.2	DESTELLOS GAMMA	33
4.2.1	Histogramas para las variables	33
4.2.1.11	Histograma para fLength en los destellos gamma	33
4.2.1.12	Histograma para fWidth en los destellos gamma	33
4.2.1.13	Histograma para fSize en los destellos gamma	33
4.2.1.14	Histograma para fConc en los destellos gamma	33
4.2.1.15	Histograma para fConc1 en los destellos gamma	33
4.2.1.16	Histograma para fAsym en los destellos gamma	33
4.2.1.17	Histograma para fM3Long en los destellos gamma	34
4.2.1.18	Histograma para fM3Trans en los destellos gamma	34
4.2.1.19	Histograma para fAlpha en los destellos gamma	34
4.2.1.20	Histograma para fDist en los destellos gamma	34
4.2.2	Estadísticas para las variables	34
4.2.3	Moda para las variables	34
4.2.4	Correlación entre las variables	35
CAPÍTULO 5. CONCLUSIONES		37
CAPÍTULO 6. BIBLIOGRAFÍA.....		39
CAPÍTULO 7. ANEXO: CÓDIGO SCRIPT		41

ÍNDICE DE FIGURAS

Figura 2-1: Histograma	11
Figura 2-2: Gráfico de barras	11
Figura 3-1: Diagrama de barras de las frecuencias absolutas.	16
Figura 3-2: Histograma para fLength en los destellos de hadrones.	17
Figura 3-3: Histograma para fWidth en los destellos de hadrones.	17
Figura 3-4: Histograma para fSize en los destellos de hadrones.	18
Figura 3-5: Histograma para fConc en los destellos de hadrones.	18
Figura 3-6: Histograma para fConc1 en los destellos de hadrones.	19
Figura 3-7: Histograma para fAsym en los destellos de hadrones.	19
Figura 3-8: Histograma para fM3Long en los destellos de hadrones.	20
Figura 3-9: Histograma para fM3Trans en los destellos de hadrones.	20
Figura 3-10: Histograma para fAlpha en los destellos de hadrones.	21
Figura 3-11: Histograma para fDist en los destellos de hadrones.	21
Figura 3-12: Estadísticas de los destellos de hadrón.	22
Figura 3-13: Correlación de Pearson para los hadrones.	23
Figura 3-14: Histograma para fLength en los destellos gamma.	23
Figura 3-15: Histograma para fWidth en los destellos gamma.	24
Figura 3-16: Histograma para fSize en los destellos gamma.	24
Figura 3-17: Histograma para fConc en los destellos gamma.	25
Figura 3-18: Histograma para fConc1 en los destellos gamma.	25
Figura 3-19: Histograma para fAsym en los destellos gamma.	26
Figura 3-20: Histograma para fM3Long en los destellos gamma.	26
Figura 3-21: Histograma para fM3Trans en los destellos gamma.	27
Figura 3-22: Histograma para fAlpha en los destellos gamma.	27
Figura 3-23: Histograma para fDist en los destellos gamma.	28
Figura 3-24: Estadísticas de los destellos gamma.	28
Figura 3-25: Correlación de Pearson para los destellos gamma.	29

ÍNDICE DE CUADROS

Tabla 3.1: Clasificación de variables del data set	15
Tabla 3.2: Tabla de frecuencias del data set	15
Tabla 3.3: Moda para la clase Hadrón	22
Tabla 3.4: Moda para la clase gamma	29

CAPÍTULO 1. INTRODUCCIÓN

1.1 MOTIVACIÓN Y ANTECEDENTES

Para realizar un estudio de datos óptimo es no se debe requerir solamente de los datos y aplicar mediciones estadísticas para establecer inferencias, sino que se debe conocer el contexto de los datos, es decir, el significado de los datos para poder realizar mediciones estadísticas adecuadas. Por ejemplo si se tiene un conjunto de datos de tipo binario no tiene ningún sentido obtener el promedio de estos datos ya que no tiene lógica alguna.

El obtener datos de forma adecuada mediante mediciones estadísticas puede permitir determinar tendencias de los datos y el significado de esta tendencia, se pueden inferir características destacables de entre los datos o hasta incluso obtener información inesperada.

El software R en conjunto con RStudio permite el uso de herramientas para realizar las mediciones estadísticas pertinentes a un conjunto de datos además de generar gráficos y tablas que facilitaran la inferencia de la información asociada al conjunto de datos. Se recalca nuevamente la obligación de comprender la naturaleza de los datos ya que no tiene sentido aplicar ciertos métodos de R a ciertos datos.

1.2 OBJETIVOS

El objetivo del presente documento es exponer al lector el análisis resultante de aplicar las distintas métricas estadísticas al respectivo conjunto de datos, es decir, el conjunto de datos de rayos Gamma y las partículas de hadron obtenidas por el telescopio MAGIC.

Para complementar la investigación también se expone al lector las características de las distintas métricas estadísticas utilizadas para el desarrollo de este documento para realizar el posterior análisis de datos.

1.3 ORGANIZACIÓN DEL DOCUMENTO

Para cumplir con los objetivos expuestos el presente documento distribuye su información de la siguiente forma: Primero se considero necesario un capítulo para que el lector tenga un conocimiento apropiado de las distintas métricas estadísticas utilizadas para el desarrollo del análisis, y de esta forma sea capaz de entender el porque se utilizan en ciertas variables y en otras no. Luego se cuenta con una sección en la que se explica el funcionamiento de los distintos comandos pertenecientes a R y que son utilizados para el desarrollo del análisis del respectivo conjunto de datos. Después se presentan los distintos resultados obtenidos mediante tablas, gráficos y explicando el sentido real de los resultados. Finalmente se presenta una conclusión en la que se comenta el si se cumplieron los objetivos y las distintas problemáticas asociadas al desarrollo del presente documento.

CAPÍTULO 2. MARCO TEÓRICO

Luego de haber definido nuestro estudio acerca de las partículas gama captadas por el telescopio Cherenkov en el informe 1, se continuó en el informe 2 con la capacitación del uso del software R. Con estos conocimientos adquiridos en ambos informes se proseguirá a realizar un análisis estadístico de nuestro data set con nuestro software R, es decir, diseñar, recolectar y analizar información recolectada por el telescopio Cherenkov. A continuación se definirán algunos conceptos relevantes en el marco de la estadística, para luego poder clasificar nuestras variables de análisis.

2.1 TIPO DE POBLACIÓN

La población es un conjunto de individuos o cosas a estudiar que podemos clasificarlas en población real, población hipotética, población estable, población aleatoria, entre otras. En este caso solo la dividiremos en 2 tipos, población finita y población infinita.

Población Finita: Nos referimos a población finita cuando el conjunto se encuentra compuesto por una cantidad limitada de elementos (José Luis García Cué, 1999).

Población Infinita: Es la que tiene un número extremadamente grande de componentes (José Luis García Cué, 1999). Es por esto que cuando tenemos que analizar una población infinita acotamos el conjunto de componentes, y ésta es llamada **muestra**.

2.2 DEFINICIÓN DE TIPOS DE VARIABLES

La variable es una medida en un experimento, representada generalmente por una (x) o por una (y) que puede tomar un valor de un conjunto de valores (José Luis García Cué, 1999).

Variable Cualitativa: Las variables cualitativas se refieren a características o cualidades que no pueden ser medidas con números. Podemos distinguir dos tipos:

Nominal: Una variable cualitativa nominal presenta modalidades no numéricas que no admiten un criterio de orden (Vitutor, 2014).

Ordinal: Una variable cualitativa ordinal presenta modalidades no numéricas, en las que existe un orden (Vitutor, 2014).

Variable Cuantitativa: Una variable cuantitativa es la que se expresa mediante un número, por tanto se pueden realizar operaciones aritméticas con ella. Podemos distinguir dos tipos:

Discreta: Una variable discreta es aquella que toma valores aislados, es decir no admite valores intermedios entre dos valores específicos (Vitutor, 2014).

Continua: Una variable continua es aquella que puede tomar valores comprendidos entre dos números (Vitutor, 2014).

2.3 ESTADÍSTICA UNIDIMENSIONAL

Al analizar características comunes a los elementos que componen la población de estudio y estudiamos un único carácter, entonces hablamos de estadística unidimensional.

Frecuencia absoluta: La frecuencia absoluta es el número de veces que aparece un determinado valor en un estudio estadístico. Se representa por f_i . La suma de las frecuencias absolutas es igual al número total de datos, que se representa por N (Vitutor, 2014)

$$\sum_{i=1}^n f_i = N$$

Frecuencias relativas: La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se puede expresar en tantos por ciento y se representa por n_i .

$$n_i = f_i/n$$

La suma de las frecuencias relativas es igual a 1 (Vitutor, 2014).

Frecuencia absoluta acumulada: La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado. Se representa por F_i (Vitutor, 2014).

Frecuencia relativas acumulada: La frecuencia relativa acumulada es el cociente entre la frecuencia acumulada de un determinado valor y el número total de datos. Se puede expresar en tantos por ciento. Este tipo de tablas de frecuencias se utiliza con variables discretas (Vitutor, 2014).

Histograma: Un histograma es una representación gráfica de una variable en forma de barras. Se utilizan para variables continuas o para variables discretas, con un gran número de datos, y que se han agrupado en clases. En el eje abscisas se construyen unos rectángulos que tienen por base la amplitud del intervalo, y por altura, la frecuencia absoluta de cada intervalo (Vitutor, 2014). La superficie de cada barra es proporcional a la frecuencia de los valores representados. Como podemos apreciarlo en el ejemplo de la Figura 2-1.

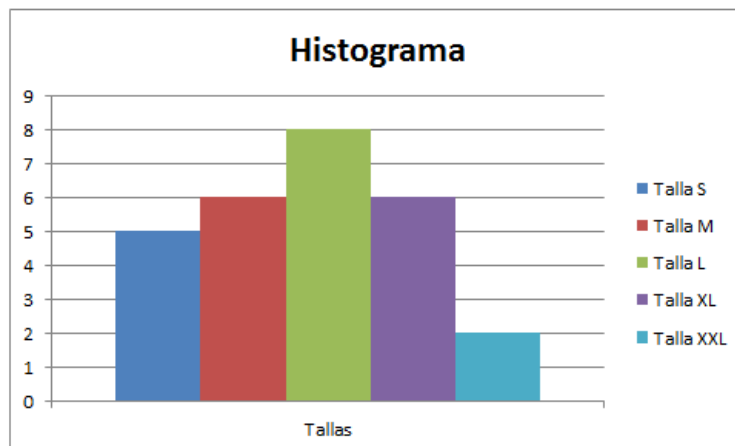
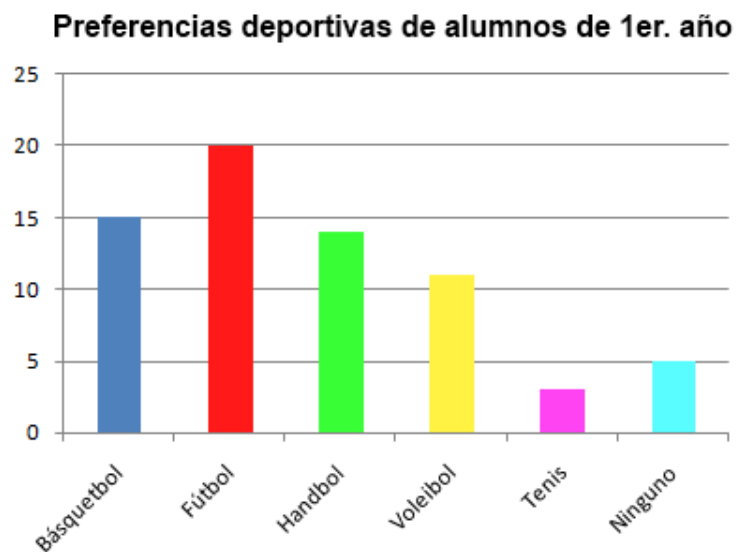


Figura 2-1: Histograma

Tabla de Frecuencias: La distribución de frecuencias o tabla de frecuencias es una ordenación en forma de tabla de los datos estadísticos, asignando a cada dato su frecuencia correspondiente.

Gráfico de barras separadas: Un diagrama de barras se utiliza para de presentar datos cualitativos o datos cuantitativos de tipo discreto. Se representan sobre unos ejes de coordenadas, en el eje de abscisas se colocan los valores de la variable, y sobre el eje de ordenadas las frecuencias absolutas o relativas o acumuladas. Los datos se representan mediante barras de una altura proporcional a la frecuencia (Vitutor, 2014). Como podemos apreciarlo en el ejemplo de la Figura 2-2.

*Figura 2-2: Gráfico de barras*

2.4 MEDIDAS DE CENTRALIZACIÓN

En situaciones se desea resumir la información expresada en las tablas en un solo número que nos diga algo, es decir, que nos represente el como están distribuidos los datos de una muestra o población.

Moda: Corresponde a la variable que más se repite en nuestra muestra o población, es decir, nuestra moda será la variable con mayor frecuencia absoluta en nuestra tabla de frecuencias.

Mediana: La mediana es un valor de la variable que deja por debajo de sí a la mitad de los datos, una vez que estos están ordenados de menor a mayor. La mediana se denota por Me y se calcula de manera sencilla.

Media: La media de una variable X se denota por \bar{x} , y se calcula como se indica a continuación. Dado un conjunto de datos numéricos x_1, x_2, \dots, x_n , la media no es más que la suma de todos los datos dividido por el número total de valores.

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

2.5 MEDIDAS DE DISPERSIÓN O MEDIDAS DE VARIABILIDAD

Nos indican por medio de un número que tan heterogénea o homogénea es la muestra o población, es decir, que tan parecidos son los valores obtenidos entre sí.

Dispersión respecto a Mediana: Es la diferencia entre el mayor y el menor valor que toma la variable. Es una medida poco precisa ya que sólo toma en consideración un par de observaciones y puede verse afectada por valores extremos.

Varianza: La varianza es la media aritmética del cuadrado de las desviaciones respecto a la media.

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

Coficiente de variación: Se define como cociente entre la desviación típica y la media:

$$CV = \frac{S}{\bar{x}}$$

Se interpreta como el número de veces que la media está contenida en la desviación típica. Suele darse su valor en tanto por ciento, multiplicando el resultado anterior por 100. De este modo se obtiene un porcentaje de la variabilidad.

2.6 RELACIÓN FUNCIONAL ENTRE DOS VARIABLES

Hablamos de relación funcional entre dos variables cuando mediante a una de las variables podemos conocer el valor de la otra variable. Es decir existe una variable dependiente y una variable independiente.

Razón de correlación: Es una medida de la relación entre la dispersión estadística entre categorías individuales y la dispersión entre la muestra o la población completa.

Kendall Spearman: Cuando los datos no se distribuyen según una normal bivariada o bien están medidos con una escala ordinal, una de las posibles medidas de asociación lineal es el coeficiente de correlación de Spearman

$$r_s = 1 - \frac{\sum_{i=1}^N (d_i^2)}{n(n^2 - 1)}$$

Donde d_i es la diferencia entre el rango del caso i en la variable X , y en la variable Y .

Coefficiente de correlación lineal (Pearson): El Coeficiente de Correlación de Pearson mide la relación lineal entre dos variables y se define como el cociente entre la covarianza y el producto de las desviaciones típicas de ambas variables.

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

CAPÍTULO 3. RESULTADOS

Recordando del informe 1, tenemos una muestra captada por nuestro telescopio de 19020 instancias, en donde 12332 son de clase gamma (g) y 6688 de clase hadrón (h). En cada una de esas instancias se pudieron observar 11 atributos o variables. Las cuales se pueden clasificar como:

Tabla 3.1: Clasificación de variables del data set

Variables	Tipo de variable
fLength	Continua
fWidth	Continua
fSize	Continua
fConc	Continua
fConc1	Continua
fAsym	Continua
fM3Long	Continua
fM3Trans	Continua
fAlpha	Continua
fDist	Continua
class	Nominal

Tal como se puede observar en la tabla 4.1, el ultimo elemento corresponde a la clase de destello que se presenta en la base de datos, estas pueden ser gamma (señal) denotada por 'g' y hadrón (fondo) denotado por 'h'.

A continuación se presentan las frecuencias tanto relativas como absolutas para cada una de las clases:

Tabla 3.2: Tabla de frecuencias del data set

Clases	Frecuencia Absoluta	Frecuencia Relativa
h: hadron	6688	0.648370136698212
g: gamma	12332	0.351629863301788

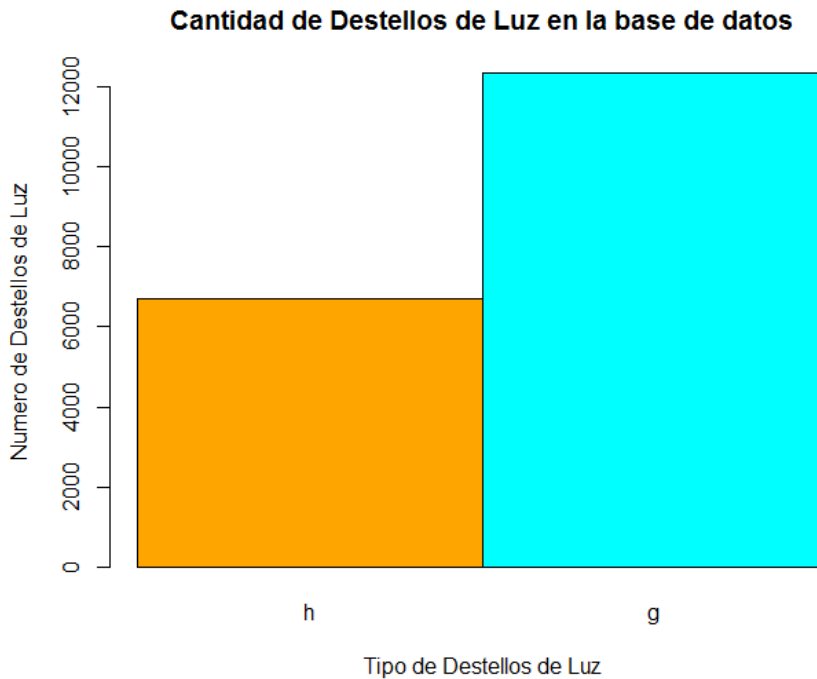


Figura 3-1: Diagrama de barras de las frecuencias absolutas.

Como se puede apreciar en la Figura 3-1, existe una relación aproximadamente 2:1 entre las partículas gamma y los hadrones, lo que es considerable para la investigación puesto que la finalidad del telescopio es solo captar las partículas gamma.

Para cada destello, se presenta un conjunto de histogramas para cada una de las variables descritas. Además se presentan una serie de estadísticas diferentes como la media, la varianza, la desviación, los valores mínimos y máximos, junto con el coeficiente de variación. Así mismo también se obtiene el coeficiente de Pearson para cada una de las categorías descritas anteriormente.

A continuación se presentaran histogramas, tablas estadísticas, tablas de moda y tablas de correlación para las variables correspondientes a la clase hadrón y gamma, para luego analizar en el próximo capítulo como se distribuyen y comportan los datos obtenidos de cada variable.

3.1 DESTELLOS DE HADRÓN

3.1.1 Histogramas para las variables

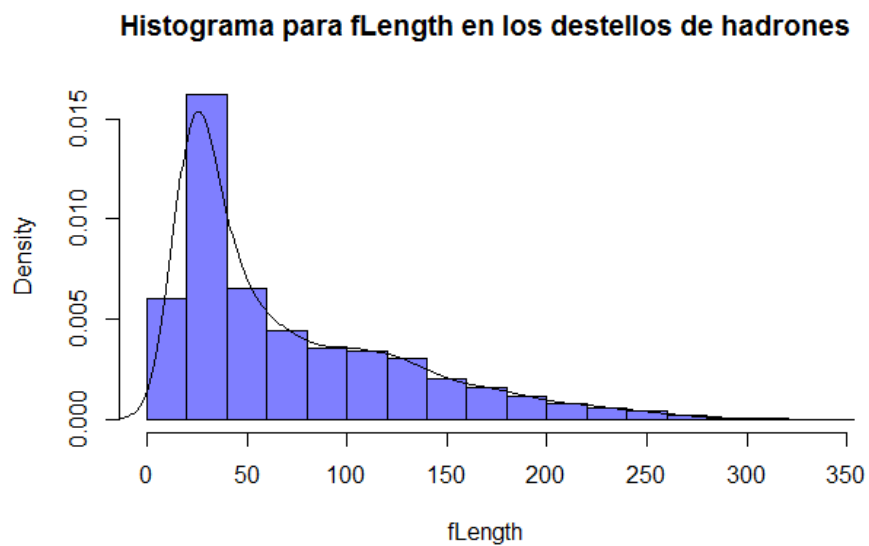


Figura 3-2: Histograma para fLength en los destellos de hadrones.

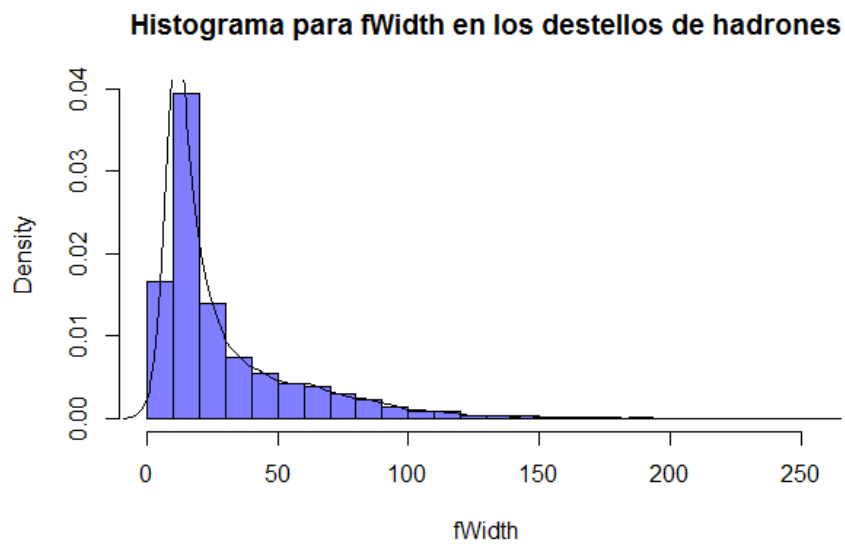


Figura 3-3: Histograma para fWidth en los destellos de hadrones.

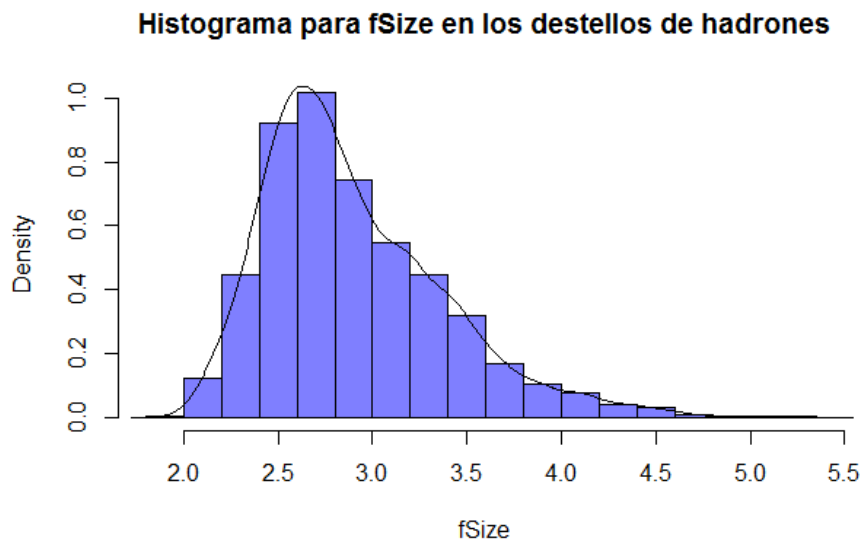


Figura 3-4: Histograma para fSize en los destellos de hadrones.

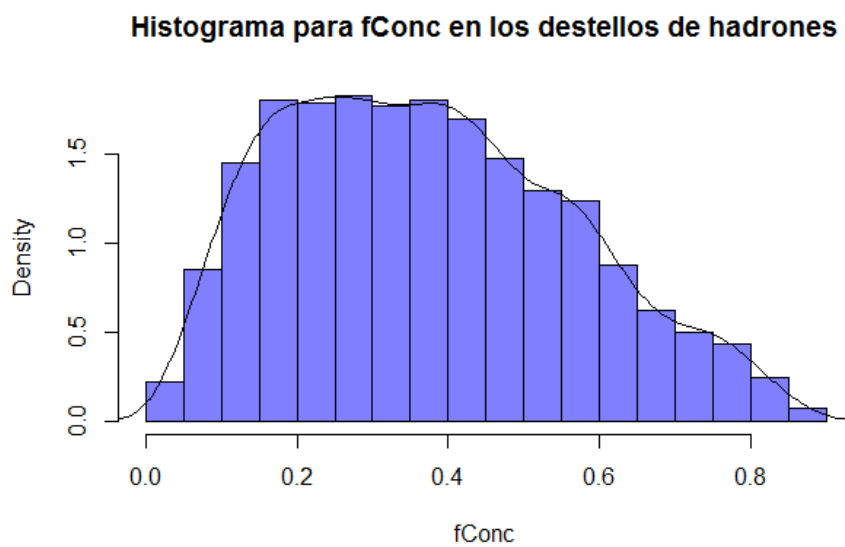
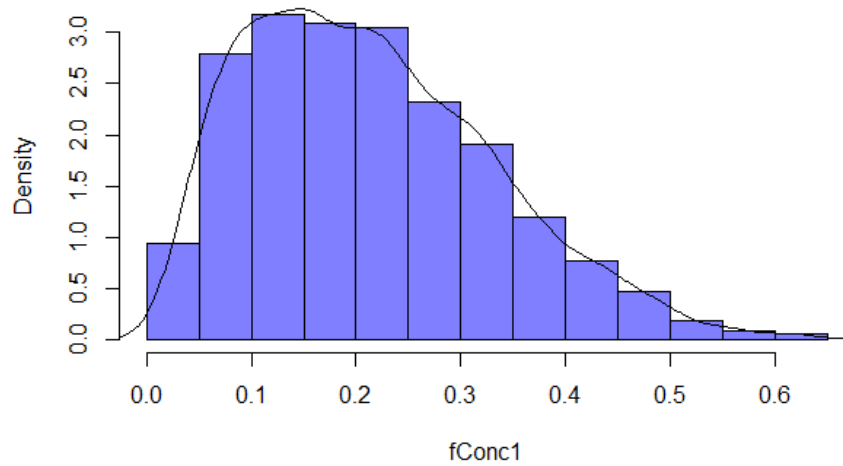
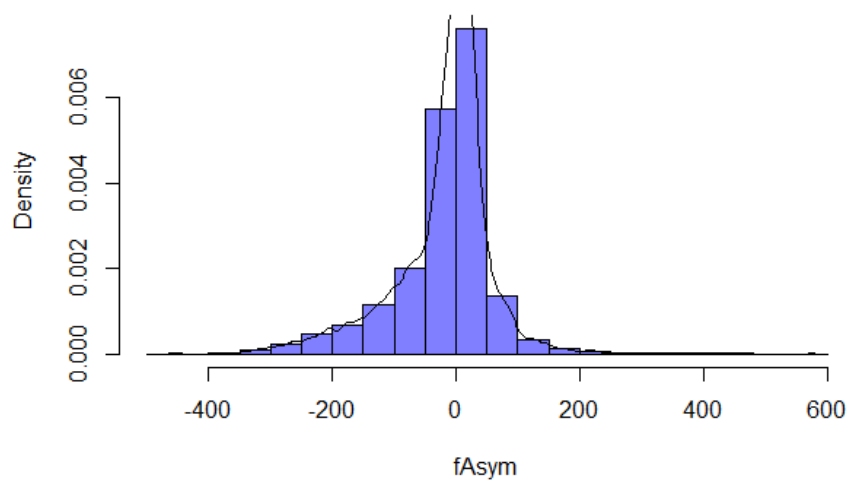


Figura 3-5: Histograma para fConc en los destellos de hadrones.

Histograma para fConc1 en los destellos de hadrones*Figura 3-6: Histograma para fConc1 en los destellos de hadrones.***Histograma para fAsym en los destellos de hadrones***Figura 3-7: Histograma para fAsym en los destellos de hadrones.*

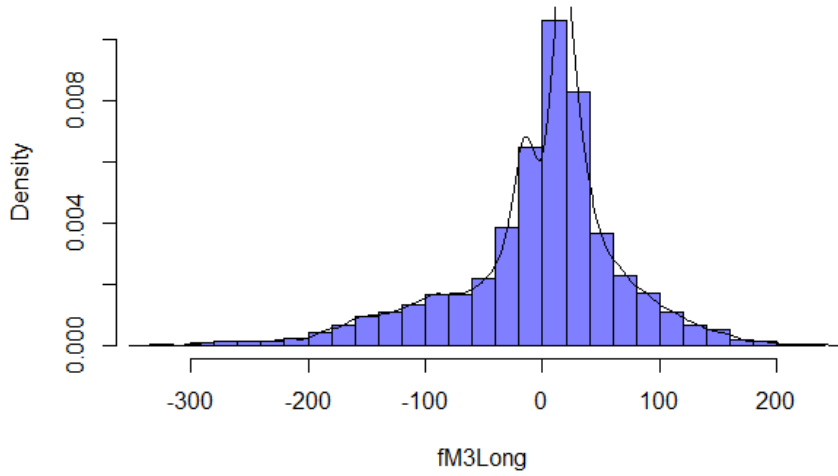
Histograma para fM3Long en los destellos de hadrones

Figura 3-8: Histograma para fM3Long en los destellos de hadrones.

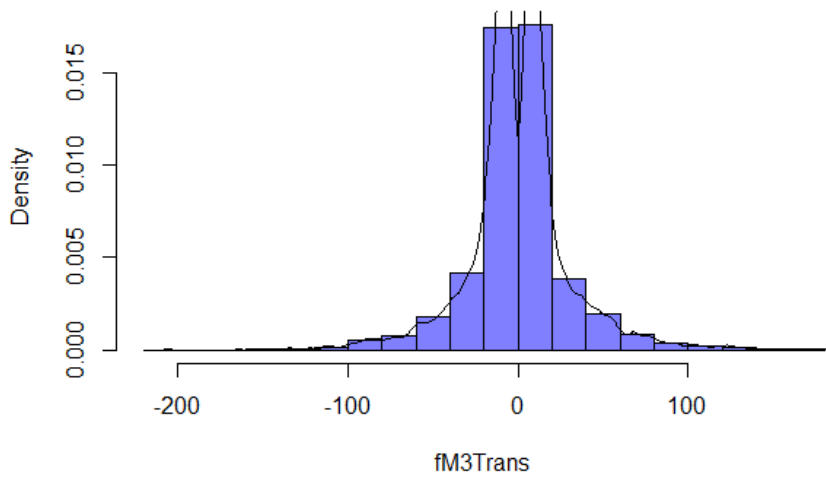
Histograma para fM3Trans en los destellos de hadrones

Figura 3-9: Histograma para fM3Trans en los destellos de hadrones.

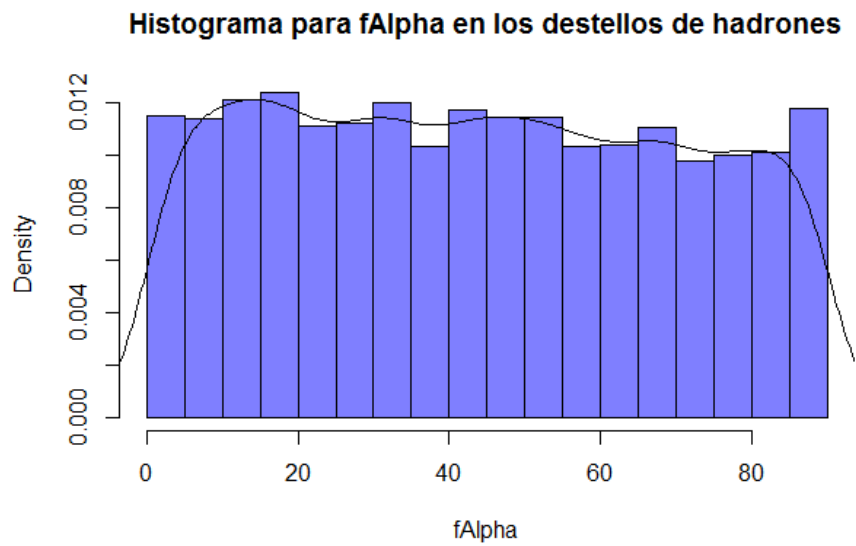


Figura 3-10: Histograma para fAlpha en los destellos de hadrones.

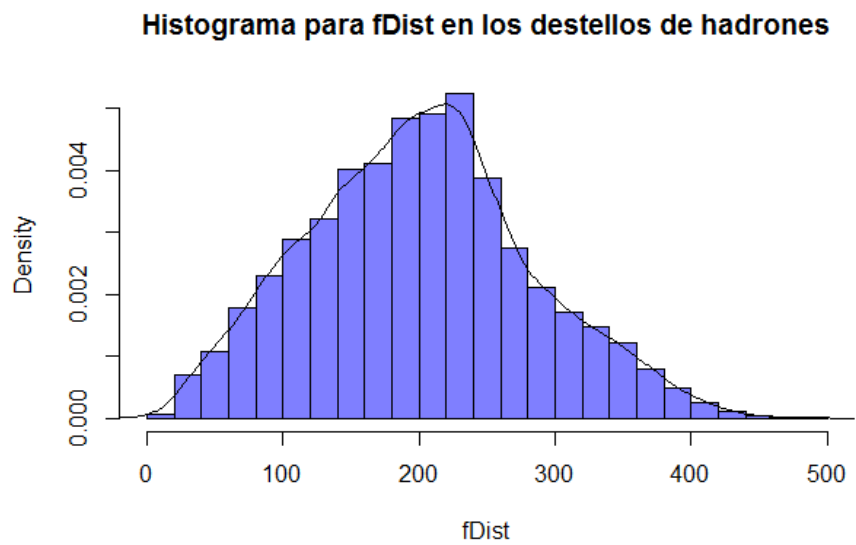


Figura 3-11: Histograma para fDist en los destellos de hadrones.

3.1.2 Estadísticas para las variables

	Media	Desviacion	Varianza	Minimo	Maximo	Coef. Var.
fLength	70.9435043	57.9527287	3.358519e+03	4.2835	334.1770	0.8168856
fWidth	28.7973727	27.1926226	7.394387e+02	0.0000	256.3820	0.9442744
fSize	2.9006088	0.4806117	2.309876e-01	1.9413	5.3233	0.1656934
fConc	0.3742169	0.1912700	3.658423e-02	0.0131	0.8930	0.5111208
fConc1	0.2139373	0.1180511	1.393607e-02	0.0003	0.6430	0.5518026
fAsym	-18.2871110	82.2978754	6.772940e+03	-457.9161	575.2407	-4.5003213
fM3Long	-2.8482976	70.6857971	4.996482e+03	-331.7800	238.3210	-24.8168583
fM3Trans	0.3582453	29.9181485	8.950956e+02	-205.8947	179.8510	83.5130163
fAlpha	43.9852332	25.9839346	6.751649e+02	0.0097	90.0000	0.5907422
fDist	200.4345174	81.8704735	6.702774e+03	1.2826	495.5610	0.4084649

Figura 3-12: Estadísticas de los destellos de hadrón.

3.1.3 Moda para las variables

Tabla 3.3: Moda para la clase Hadrón

Variables	Moda
fLength	NA
fWidth	0
fSize	2.6031
fConc	0.3696
fConc1	0.169 - 0.243
fAsym	0
fM3Long	0
fM3Trans	0
fAlpha	26.9 - 61.5 - 63.1
fDist	32.3 - 46.6 - 57.9 - 62.5 - 66.8 - . . .

3.1.4 Correlación entre las variables

	fLength	fWidth	fSize	fConc	fConc1	fAsym	fM3Long	fM3Trans	fAlpha	fDist
fLength	1.000000000	0.77388415	0.747319449	-0.688523876	-0.651878530	-0.423193551	-0.307349467	0.013006907	-0.009281786	0.32526695
fWidth	0.773884146	1.000000000	0.790355200	-0.672617563	-0.638750459	-0.304117537	-0.296955766	0.039518634	0.041616820	0.35499734
fSize	0.747319449	0.790355200	1.000000000	-0.839128348	-0.796836767	-0.305340042	-0.223553365	0.010991984	-0.004227923	0.39484159
fConc	-0.688523876	-0.67261756	-0.839128348	1.000000000	0.972725250	0.238805805	0.124154135	-0.007703533	-0.010714499	-0.26768378
fConc1	-0.651878530	-0.63875046	-0.796836767	0.972725250	1.000000000	0.220316949	0.115716462	-0.008234827	-0.015137427	-0.24517934
fAsym	-0.423193551	-0.30411754	-0.305340042	0.238805805	0.220316949	1.000000000	0.285738265	0.006849347	0.053334383	-0.22904934
fM3Long	-0.307349467	-0.29695577	-0.223553365	0.124154135	0.115716462	0.285738265	1.000000000	-0.026806057	-0.001489255	-0.13112105
fM3Trans	0.013006907	0.03951863	0.010991984	-0.007703533	-0.008234827	0.006849347	-0.026806057	1.000000000	0.010831730	0.01314991
fAlpha	-0.009281786	0.04161682	-0.004227923	-0.010714499	-0.015137427	0.053334383	-0.001489255	0.010831730	1.000000000	-0.09711218
fDist	0.325266948	0.35499734	0.394841585	-0.267683779	-0.245179342	-0.229049339	-0.131121045	0.013149912	-0.097112180	1.000000000

Figura 3-13: Correlación de Pearson para los hadrones.

3.2 DESTELLOS GAMMA

3.2.1 Histogramas para las variables

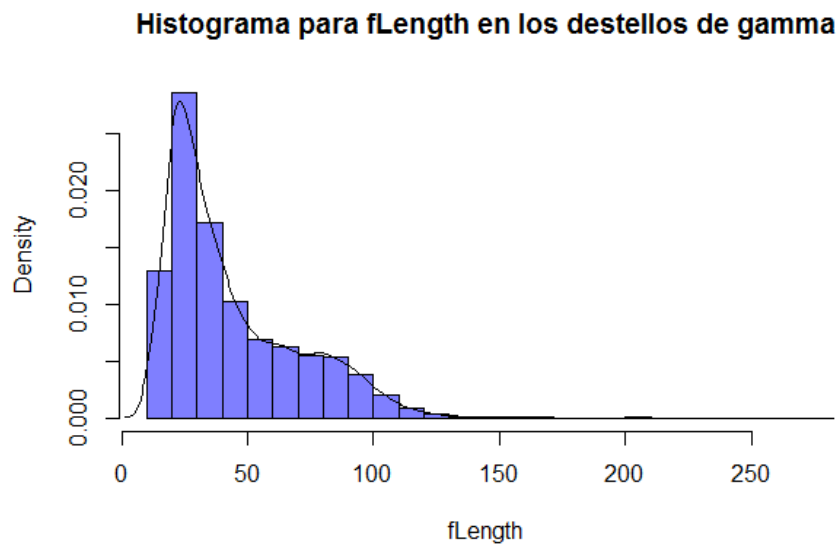


Figura 3-14: Histograma para fLength en los destellos gamma.

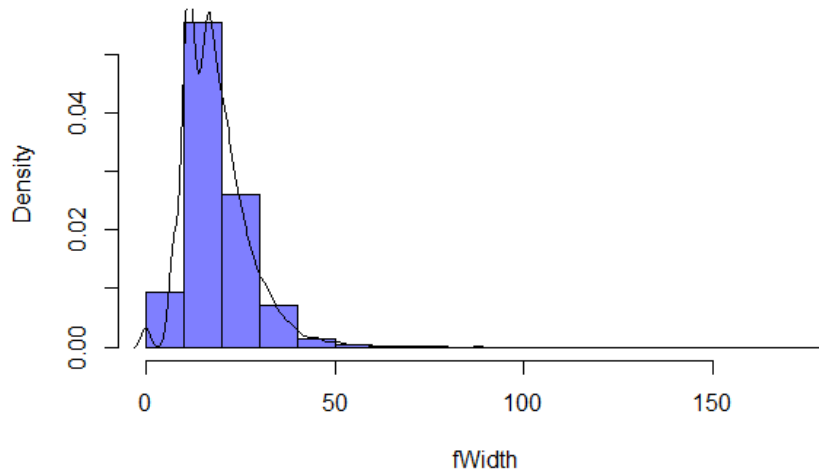
Histograma para fWidth en los destellos de gamma

Figura 3-15: Histograma para fWidth en los destellos gamma.

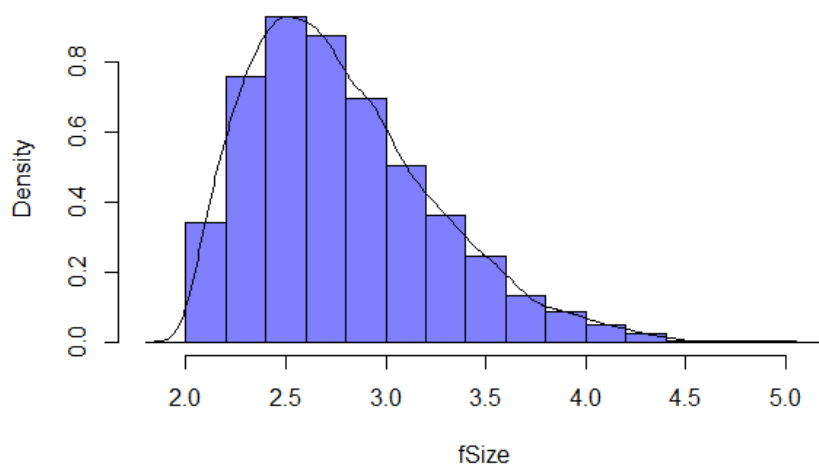
Histograma para fSize en los destellos de gamma

Figura 3-16: Histograma para fSize en los destellos gamma.

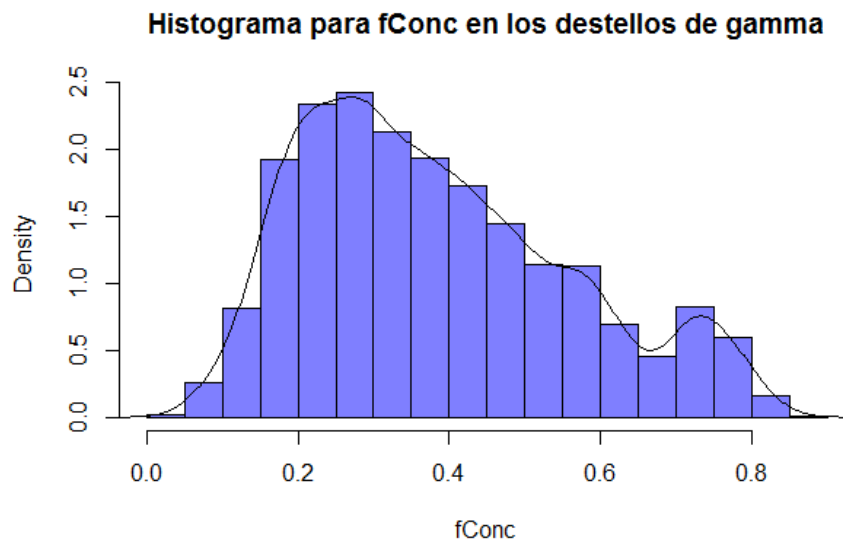


Figura 3-17: Histograma para fConc en los destellos gamma.

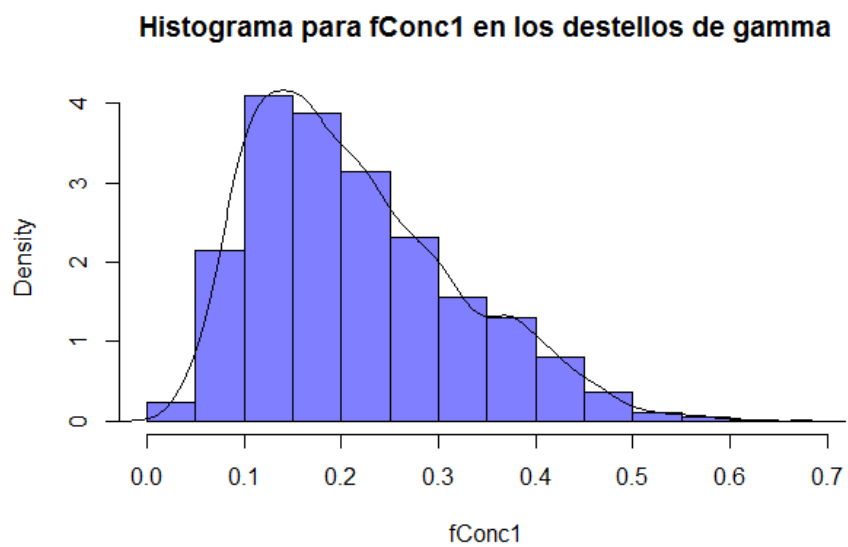


Figura 3-18: Histograma para fConc1 en los destellos gamma.

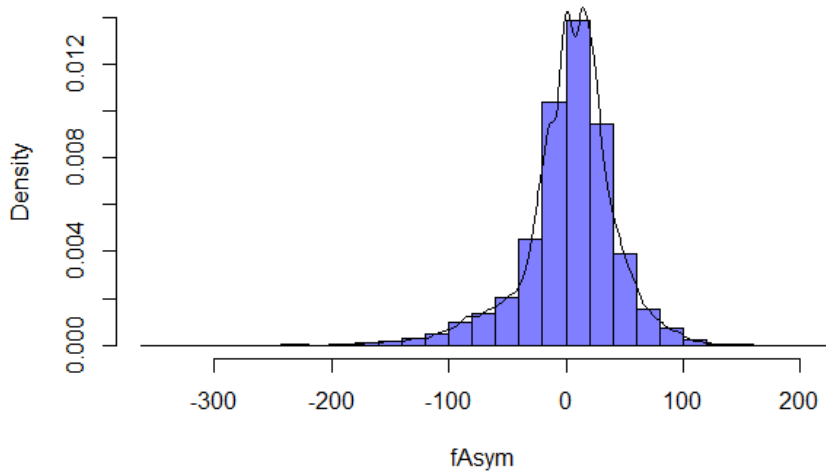
Histograma para fAsym en los destellos de gamma

Figura 3-19: Histograma para fAsym en los destellos gamma.

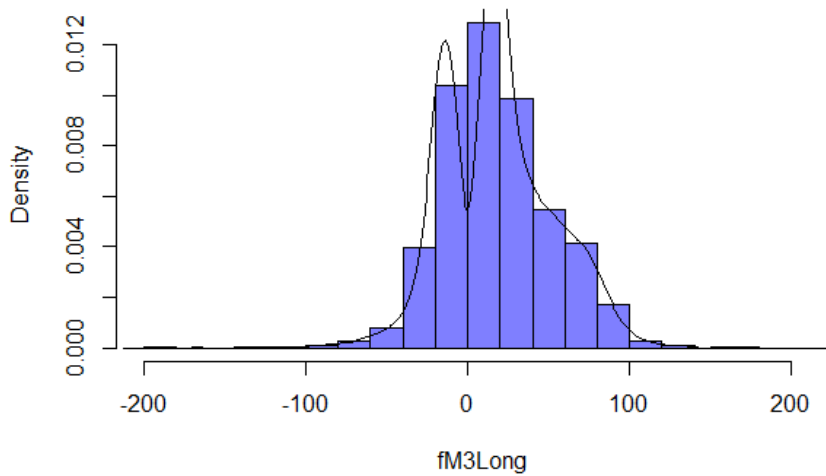
Histograma para fM3Long en los destellos de gamma

Figura 3-20: Histograma para fM3Long en los destellos gamma.

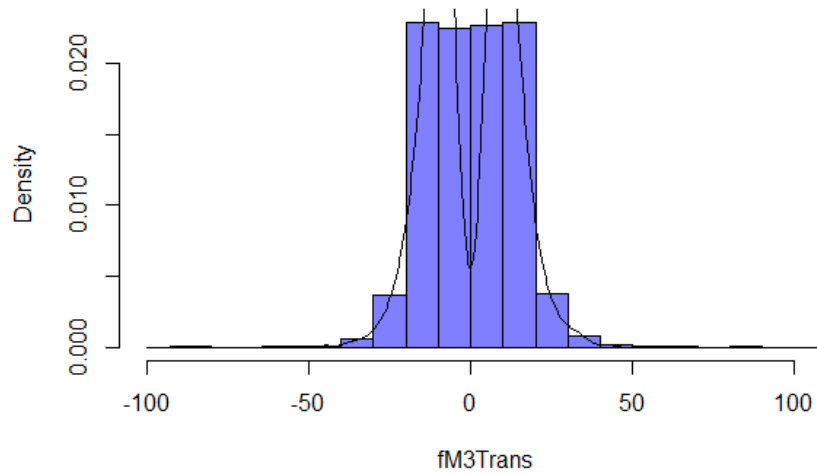
Histograma para fM3Trans en los destellos de gamma

Figura 3-21: Histograma para fM3Trans en los destellos gamma.

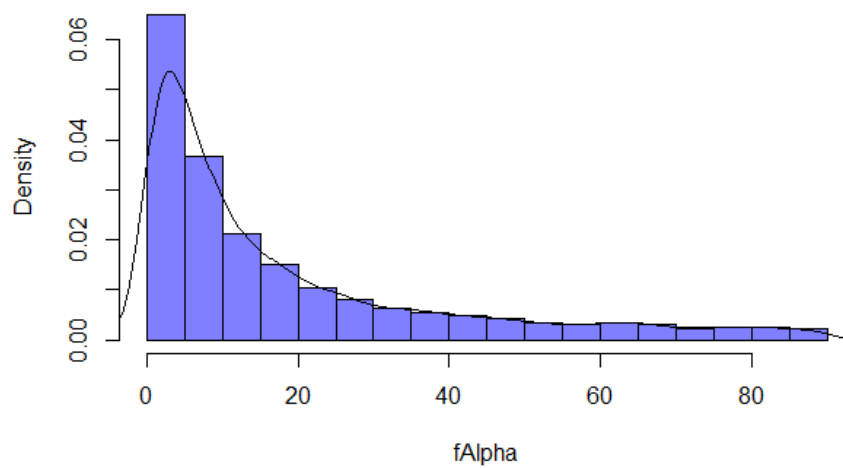
Histograma para fAlpha en los destellos de gamma

Figura 3-22: Histograma para fAlpha en los destellos gamma.

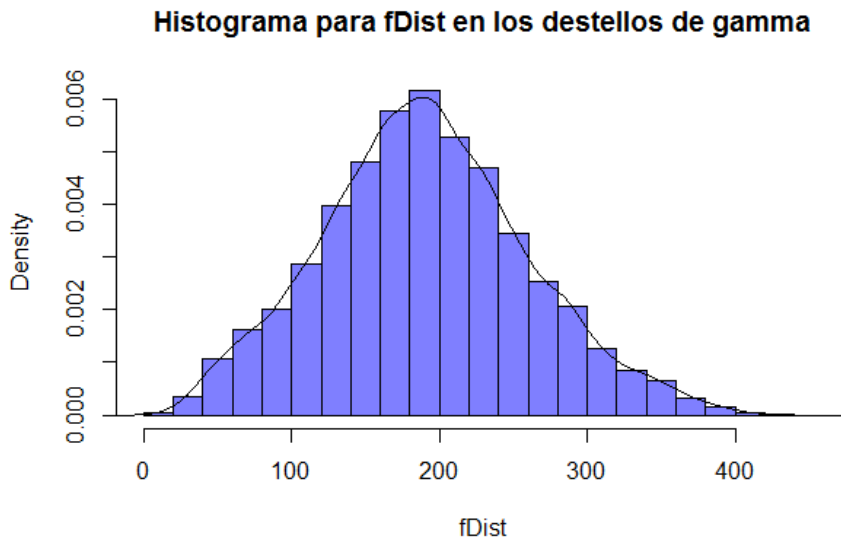


Figura 3-23: Histograma para fDist en los destellos gamma.

3.2.2 Estadísticas para las variables

	Media	Desviacion	Varianza	Minimo	Maximo	Coef. Var.
fLength	70.9435043	26.1734342	6.850487e+02	12.1929	272.0630	0.5995581
fWidth	28.7973727	9.0286662	8.151681e+01	0.0000	176.3350	0.4856028
fSize	2.9006088	0.4630827	2.144456e-01	1.9978	5.0100	0.1663359
fConc	0.3742169	0.1779791	3.167654e-02	0.0139	0.8889	0.4639211
fConc1	0.2139373	0.1062007	1.127859e-02	0.0073	0.6752	0.4938475
fAsym	-18.2871110	39.6326288	1.570745e+03	-349.7570	219.8960	12.2449420
fM3Long	-2.8482976	33.9427796	1.152112e+03	-198.8670	215.8940	1.9058899
fM3Trans	0.3582453	13.5514233	1.836411e+02	-91.3485	101.3930	70.9971565
fAlpha	43.9852332	21.4825854	4.615015e+02	0.0000	90.0000	1.1436452
fDist	200.4345174	70.3023841	4.942425e+03	5.7456	450.4020	0.3695657

Figura 3-24: Estadísticas de los destellos gamma.

3.2.3 Moda para las variables

Tabla 3.4: Moda para la clase gamma

Variables	Moda
fLength	24.382
fWidth	0
fSize	2.1508
fConc	0.6
fConc1	0.194
fAsym	0
fM3Long	0
fM3Trans	0
fAlpha	2e-04
fDist	117 186 210 216 246

3.2.4 Correlación entre las variables

	fLength	fWidth	fSize	fConc	fConc1	fAsym	fM3Long	fM3Trans	fAlpha	fDist
fLength	1.00000000	0.695645604	0.76611863	-0.71020137	-0.67997682	-0.120236696	0.542494040	0.012365077	-0.443254341	0.602484254
fWidth	0.69564560	1.000000000	0.85515611	-0.76133892	-0.73142715	-0.002603284	0.422040086	0.044029211	-0.296895911	0.389202911
fSize	0.76611863	0.855156110	1.000000000	-0.86343788	-0.82538765	0.029023346	0.540957933	0.022636604	-0.458026829	0.459632536
fConc	-0.71020137	-0.761338915	-0.86343788	1.000000000	0.97933050	-0.033394763	-0.452183185	-0.017358046	0.482306093	-0.368562686
fConc1	-0.67997682	-0.731427149	-0.82538765	0.97933050	1.000000000	-0.037630972	-0.433914104	-0.016044296	0.463504592	-0.347069275
fAsym	-0.12023670	-0.002603284	0.02902335	-0.03339476	-0.03763097	1.000000000	0.162438350	-0.005629217	-0.004521284	-0.178247494
fM3Long	0.54249404	0.422040086	0.54095793	-0.45218318	-0.43391410	0.162438350	1.000000000	0.008227403	-0.271443903	0.294720759
fM3Trans	0.01236508	0.044029211	0.02263660	-0.01735805	-0.01604430	-0.005629217	0.008227403	1.000000000	-0.007360395	0.009828714
fAlpha	-0.44325434	-0.296895911	-0.45802683	0.48230609	0.46350459	-0.004521284	-0.271443903	-0.007360395	1.000000000	-0.425098133
fDist	0.60248425	0.389202911	0.45963254	-0.36856269	-0.34706927	-0.178247494	0.294720759	0.009828714	-0.425098133	1.000000000

Figura 3-25: Correlación de Pearson para los destellos gamma.

CAPÍTULO 4. ANÁLISIS

4.1 DESTELLOS DE HADRÓN

4.1.1 Histogramas para las variables

Estos histogramas sirven para obtener una "primera vista" general, o panorama, de la distribución de la muestra de los destellos de hadrones, respecto a cada una de las variables presentes. De esta manera ofrece una visión en grupo permitiendo observar una preferencia, o tendencia, por parte de la muestra por ubicarse hacia una determinada región de valores dentro del espectro de valores presentados para cada variable.

4.1.1.1 Histograma para *fLength* en los destellos de hadrones

Como se puede observar a través de este histograma, el largo en mm de los destellos de hadrones varía entre 0 y 300, teniendo una clara tendencia en los valores cercanos a 50 mm.

4.1.1.2 Histograma para *fWidth* en los destellos de hadrones

Se aprecia que el ancho en mm de los destellos de hadrones varía entre 0 y 150 aproximadamente, notando que la mayoría de los datos se concentran en valores cercanos a los 20 mm. Situación similar a la vista anteriormente en el histograma de *fLength*.

4.1.1.3 Histograma para *fSize* en los destellos de hadrones

En este caso si bien podemos notar la existencia de una mayor densidad entre 2.5 y 3.0 números de fotones por pixel, se puede decir que existe una densidad de datos mejor distribuida que las vistas en los casos anteriores.

4.1.1.4 Histograma para *fConc* en los destellos de hadrones

Aquí podemos notar que los datos de *fConc* de los hadrones van desde 0 a 0.9 una densidad mucho mas pareja, esto es esperable pues *fConc* corresponde a un promedio de los 2 valores superiores al dato seleccionado de *fSize*.

4.1.1.5 Histograma para *fConc1* en los destellos de hadrones

Aquí podemos ver que los datos van desde el 0 al 0.6, un rango inferior al del *fConc*. Esto se debe a que *fConc1* es calculado con el promedio de solo un valor superior al seleccionado de *fSize*. Por esto es que la densidad mayor llega hasta 3.0, el doble de la densidad mayor anterior.

4.1.1.6 Histograma para *fAsym* en los destellos de hadrones

Se puede apreciar en este caso que los valores van desde -400 a 200 mm, esto se debe al punto de referencia que se toma al analizar el *fAsym*, pues *fAsym* es calculado con la distancia del píxel más alto al centro, a lo largo del eje mayor. Entonces la mayoría se encuentra por sobre el 0 del eje y el resto por lo general se encuentra por debajo del 0 en el eje.

4.1.1.7 Histograma para *fM3Long* en los destellos de hadrones

En este caso tenemos una distribución que va desde el -200 al 200, con la mayoría de datos en el centro. Por lo tanto la medida de centralización media puede ser bastante representativa para este caso.

4.1.1.8 Histograma para fM3Trans en los destellos de hadrones

Al igual que el caso anterior tenemos una distribución centrada alrededor del 0, que va entre los valores -100 a 100. Por lo que una medida de centralización como la media o promedio puede ser representativa también. La similitud de la distribución se puede deber a que el calculo tanto de fM3Trans como el de fM3Long son los mismos pero en distintos ejes.

4.1.1.9 Histograma para fAlpha en los destellos de hadrones

En este caso tenemos una densidad a primera vista homogénea que va desde los 0 grados hasta aproximadamente 90 grados desde el mayor eje hasta el origen. Lo que nos dice que la elipse que capta el telescopio no es siempre en una misma posición y no presenta además una posición considerablemente mayor a las demás.

4.1.1.10 Histograma para fDist en los destellos de hadrones

Aquí podemos ver que desde el origen hasta el centro de la elipse la distancias que predominan son la de 220 - 240 mm aproximame. Con una densidad de datos distribuida principalmente alrededor de esos valores.

4.1.2 Estadísticas para las variables

A través de estos cálculos estadísticos, se puede observar que para el caso de las variables fLength, fWidth, fAsym, fM3Long, fM3Trans, fAlpha y fDist, estas presentan una desviación alta, evidenciando una heterogeneidad de los datos, de esta forma, se puede observar un volatilidad en los datos, expresado en la gran diferencia existente entre los valores máximos y mínimos. En el caso de fSize, fConc, y fConc1, se presenta según su desviación que los datos son más homogéneos. Con el cálculo del coeficiente de variación (C.V), se observan que fLength, fWidth, fSize, fConc, fConc1, fAlpha y fDist, presentan valores menores a 1, lo que evidencia que presentan distribuciones con baja varianza. Se dan casos especiales en el cálculo del C.V calculado para la variable fM3Trans, fM3Long y fAsym, donde para la primera, debido a que la media aritmética es cercana a 0 y la desviación es alta, se obtiene un valor muy grande, lo que no necesariamente implica una dispersión de datos, ya que se pierde bajo estas condiciones el cálculo del C.V. Para las otras dos variables, si se compara con lo valores típicos del C.V, los cuales están entre 0 y 1, estos se enmarcan fuera de lo típico,(Wikipedia, s.f.).

4.1.3 Moda para las variables

Se puede apreciar que existen varias variables que presentan distribuciones multimodales, siendo el caso de las variable fConc1 con 2 modas existentes, fAlpha con 3 modas y un caso particular que se presenta en la variable fDist, donde según los cálculos con las funciones de R, se presenta 141 modas. Así mismo las variables fSize y fConc presentan una distribución unimodal.

4.1.4 Correlación entre las variables

Se puede observar que en el caso de la variable fLength, esta tiene una relación con fWidth y fSize, con esta última, su relación no es claramente apreciable, pues su valor es cercano a 0.75. Para fWidth se tiene que existe una relación cercana con fSize, teniendo un valor cercano a 0.79 fsize presenta una relación con fConc y fConc1, con valores cercanos a -1, por lo tanto se evidencia que existe una relación inversa con respecto a

estas variables. f_{Conc} presenta una relación fuerte con f_{Conc1} , con valores cercanos a 0.93, esto evidencia la similitud entre estas variables, ya que f_{Conc} corresponde a la suma de los contenidos de todos los píxeles en la imagen y f_{Conc1} que representa la proporción de la mayor suma de dos píxeles sobre f_{Size} . Esto explica así mismo la relación inversa que f_{Size} presenta con ambas variables debido a la naturaleza de estas. Para f_{Asym} no se puede inferir que exista una relación con otras variables, debido a que los coeficientes son cercanos a 0. Así mismo se observa esto mismo con las variables f_{M3Long} , f_{M3Trans} , f_{Alpha} y f_{Dist} .

4.2 DESTELLOS GAMMA

4.2.1 Histogramas para las variables

Al igual que la sección de los histogramas para las variables de tipo hadrón en esta sección se muestran los histogramas respectivos para obtener una "primera vista" general, de la distribución de la muestra de los destellos de señales gamma, respecto a cada una de las variables presentes. De esta manera ofrece es posible observar una tendencia, por parte de la muestra por ubicarse hacia una determinada región de valores dentro del espectro de valores presentados para cada variable.

4.2.1.11 Histograma para f_{Length} en los destellos gamma

Como se puede observar a través de este histograma, el largo en mm de los destellos de señales gamma varía entre 0 y 250, teniendo una clara tendencia en los valores entre 0 a 50 mm.

4.2.1.12 Histograma para f_{Width} en los destellos gamma

Se aprecia que el ancho en mm de los destellos de señales gamma varía entre 0 y 150 aproximadamente, notando que la mayoría de los datos se concentran en valores cercanos a los 20 mm. Situación similar a la vista anteriormente en el histograma de f_{Length} . De hecho es bastante similar a lo ocurrido con la información análoga en los destellos de hadrón.

4.2.1.13 Histograma para f_{Size} en los destellos gamma

En este caso se puede observar la existencia de una mayor densidad de 2.5 números de fotones por pixel.

4.2.1.14 Histograma para f_{Conc} en los destellos gamma

Aquí podemos notar que los datos de f_{Conc} de las señales gamma van desde 0 a 0.8 con una densidad mas pareja, esto es esperable pues f_{Conc} corresponde a un promedio de los 2 valores superiores al dato seleccionado de f_{Size} .

4.2.1.15 Histograma para f_{Conc1} en los destellos gamma

Aquí podemos ver que los datos van desde el 0 al 0.7, un rango inferior al del f_{Conc} . Esto se debe a que f_{Conc1} es calculado con el promedio de solo un valor superior al seleccionado de f_{Size} . Por esto es que la densidad mayor llega hasta 4, un valor mayor de la densidad mayor anterior.

4.2.1.16 Histograma para f_{Asym} en los destellos gamma

Se puede apreciar en este caso que los valores van desde -300 a 200 mm, esto se debe al punto de referencia que se toma al analizar el f_{Asym} , pues f_{Asym} es calculado con la distancia del píxel más alto al

centro, a lo largo del eje mayor. Por ende, la mayoría se encuentra por sobre el 0 del eje y el resto por lo general se encuentra por debajo del 0 en el eje.

4.2.1.17 Histograma para fM3Long en los destellos gamma

En este caso tenemos una distribución que va desde el -200 al 200, con la mayoría de datos en el centro. Por lo tanto la medida de centralización media puede ser bastante representativa para este caso. Se debe hacer notar que es una situación similar a lo ocurrido en este mismo histograma para el caso de los destellos de hadrón.

4.2.1.18 Histograma para fM3Trans en los destellos gamma

Al igual que el caso anterior tenemos una distribución centrada alrededor del 0, que va entre los valores -100 a 100. Por lo que una medida de centralización como la media o promedio puede ser representativa también. La similitud de la distribución se puede deber a que el calculo tanto de fM3Trans como el de fM3Long son los mismos pero en distintos ejes. Al igual que en caso anteriores se produce una situación análoga para el caso de los destellos de hadrón.

4.2.1.19 Histograma para fAlpha en los destellos gamma

En este caso tenemos una densidad a primera vista que se diferencia totalmente de su análogo en el caso de los destellos de hadrón. Los valores van desde los 0 grados hasta aproximadamente 90 grados desde el mayor eje hasta el origen.

4.2.1.20 Histograma para fDist en los destellos gamma

Aquí podemos ver que desde el origen hasta el centro de la elipse la distancia que predomina son las cercanas a 200. Con una densidad de datos distribuida principalmente alrededor de esos valores.

4.2.2 Estadísticas para las variables

Se puede observar que para el caso de las variables fLength, fWidth, fAsym, fM3Long, fM3Trans, fAlpha y fDist, así como lo presentado en los estadísticos calculados para los destellos de hadrones, estas presentan una desviación alta, evidenciando una heterogeneidad de los datos, de esta forma, se puede observar una volatilidad en los datos, expresado en la gran diferencia existente entre los valores máximos y mínimos. En el caso de fSize, fConc, y fConc1, se presenta según su desviación que los datos son más homogéneos. Con el cálculo del coeficiente de variación (C.V), se observan que fLength, fWidth, fSize, fConc, fConc1 y fDist, presentan valores menores a 1, lo que evidencia que presentan distribuciones con baja varianza. En el caso de las variables fM3Long y fAlpha, con valores mayores que 1, estas presentan una distribución con una alta varianza. Se dan casos especiales con los C.V calculados para las variables fM3Trans y fAsym, donde el primero de ellos respectivamente debido a que la media aritmética es cercana a 0 y la desviación es alta, se obtiene un valor muy grande, lo que no necesariamente implica una dispersión de datos, ya que se pierde bajo estas condiciones el cálculo del C.V, (Wikipedia, s.f.), y algo parecido resulta con el C.V de fAsym.

4.2.3 Moda para las variables

Se puede apreciar que existe una variable que presenta una distribución multimodal, siendo esta fDist

con 5 modas presentes. Existe así, una diferencia con respecto a lo presentado en los destellos de hadrones, donde para esta variable se presentaron 141 modas. En el caso de las otras variables estas presentan una distribución unimodal.

4.2.4 Correlación entre las variables

Se puede apreciar por la correlación calculada, que en el caso de la variable `fLength`, esta presenta una cierta relación con `fSize`, con valores cercanos a 0.77. Si se compara las relaciones de `fLength` con lo presentado en los destellos de hadrones, esta vez no se puede evidenciar que exista una relación con `fWidth`, ya que su valor es cercano a 0.695. Para `fWidth` se puede afirmar que existe una relación con `fSize` y una cierta relación inversa con `fConc` con un valor cercano a -0.76, así mismo no existe evidencia suficiente para decir que existe una relación inversa con `fConc1`, por el valor del coeficiente. En el caso de `fsize` presenta una relación inversa con `fConc` y `fConc1`, con valores cercanos a -1. `fConc` presenta una fuerte relación con `fConc1`, por lo explicado en las relaciones de los destellos de hadrones. En el caso de `fAsym`, `fM3Long`, `fM3Trans`, `fAlpha` y `fDist`, no se evidencia una posible relación con otras variables, reflejando una similitud con las inferencias obtenidas a partir de la correlación de las variables presentadas para los destellos de los hadrones.

CAPÍTULO 5. CONCLUSIONES

Como se ha mencionado antes (y como de ha aplicado en este presente informe), para obtener información adecuada del respectivo conjunto de datos es necesario conocer el origen de la información, es decir, el origen de las variables estudiadas. Específicamente en este caso las variables estudiadas son en su mayoría de tipo continua(específicamente 10 de 11 tipos), mientras que solo una variable (tipo de clase hadrón o gamma) era una variable nominal. Las variables continuas en este caso son longitudes en milímetros de distintas características de los datos, por lo tanto corresponden a variables cuantitativas.

Luego de realizado el presente informe (más en específico la realización del capítulo análisis) se ha logrado completar con éxito el objetivo principal presentado en la respectiva introducción, es decir, el lector en este punto ya debe ser capaz de entender el porqué se aplicaron ciertas medidas estadísticas dependiendo del tipo de variable con la cual se trabaja. Esto ultimo ya que el lector se ha interiorizado con las distintas métricas utilizadas y que han sido descritas en el marco teórico del presente documento. Además se han descrito los distintos tipos de variables utilizadas en el estudio y es debido a ello que el lector ya es consciente de la aplicación de cierta medida estadística a cierto tipo de variable.

El uso de la herramienta R nos permitió utilizar gráficas como histogramas, en la que pudimos evidenciar las distribuciones de las distintas variables. El uso de estos histogramas ayudó para hacer la comparación entre variables, lo que nos permitió percatarnos que generalmente las variables poseen una distribución similar. Finalmente se debe hacer notar que el estudio estadístico de la muestra permite a los interesados obtener información del conjunto de datos que a simple vista es prácticamente imposible obtener. El análisis de datos realizado es fundamental para poder estudiar tendencias de datos, encontrar anomalías, entre otras características interesantes del conjunto de datos estudiado. Estar al tanto de estas características del conjunto de datos encausa a un dominio del tema mas completo, lo que conlleva a interpretaciones acertadas con respecto al tema de interés.

CAPÍTULO 6. BIBLIOGRAFÍA

José Luis García Cué. (1999, Agosto). Conceptos de estadística. Recuperado desde <http://colposfesz.galeon.com/est501/suma/sumahtml/conceptos/estadistica.htm>

Vitutor. (2014). Variable estadística. Recuperado desde http://www.vitutor.com/estadistica/descriptiva/a_2.html

Wikipedia. (s.f.). Coeficiente de variación. Recuperado desde https://es.wikipedia.org/wiki/Coeficiente_de_variaci%C3%83%C2%B3n

CAPÍTULO 7. ANEXO: CÓDIGO SCRIPT

```
1
2 library(psych)
3 library(help = "datasets")
4 library(modeest)
5 #histBy(typeh,var="fM3Trans",main="Histograma para fM3Trans en los destellos de hadrones")
6 setwd("C:/Users/Dany/Documents/Inferencia Estadística")
7 data<-read.csv(header=FALSE,col.names=c("fLength","fwidth","fSize","fConc","fConc1","fAsym",
8     "fM3Long","fM3Trans","fAlpha","fDist","class"),"magic04.data")
9
10 ~ #####
11 #Funciones usadas en el laboratorio 2
12 ~ #####
13
14
15 primerosdiez<-head(data,n=10)
16
17 fLength<-data$fLength
18 fwidth<-data$fwidth
19 fSize<-data$fSize
20 fConc<-data$fConc
21 fConc1<-data$fConc1
22 fAsym<-data$fAsym
23 fM3Long<-data$fM3Long
24 fM3Trans<-data$fM3Trans
25 fAlpha<-data$fAlpha
26 fDist<-data$fDist
27 fclass<-data$class
28
29
30 atributo<-c("fLength","fwidth","fSize","fConc","fConc1","fAsym","fM3Long",
31     "fM3Trans","fAlpha","fDist")
32 minimo<-c(min(fLength),min(fwidth),min(fSize),min(fConc),min(fConc1),min(fAsym),
33     min(fM3Long),min(fM3Trans),min(fAlpha),min(fDist))
34 maximo<-c(max(fLength),max(fwidth),max(fSize),max(fConc),max(fConc1),max(fAsym),
35     max(fM3Long),max(fM3Trans),max(fAlpha),max(fDist))
36
37 table<-data.frame(atributo,minimo,maximo)
38 rm(fLength,fwidth,fSize,fConc,fConc1,fAsym,fM3Long,fM3Trans,fAlpha,fDist,atributo,
39     minimo,maximo)
40
41 ~ #####
42
43 #Funciones para obtener los totales de cada clase
44 temp<-which(data$class=="g")
45 typeg<-data[temp,]
46 temp<-which(data$class=="h")
47 typeh<-data[temp,]
48
49 #funciones para obtener las frecuencias de cada clase
50 fg<-nrow(typeg)
51 fh<-nrow(typeh)
52 total<-nrow(data)
53
54 relfg<-fg/total
55 relfh<-fh/total
```

```

56
57 typeNames<-c("hadron","gamma")
58
59 absFreq<-c(fh,fg)
60 relFreq<-c(re1fh,relfg)
61 freqTable<-matrix(c(absFreq,relFreq),nrow=2,ncol=2)
62
63 dimnames(freqTable) = list(c("Hadron","Gamma"),c("Frecuencia absoluta","Frecuencia relativa"))
64
65 #Grafico de barras con la frecuencia absoluta
66 barplot(c(fh,fg),space=0,names.arg=typeNames,main="Cantidad de Destellos de Luz en la base de datos",
67        xlab="Tipo de Destellos de Luz",ylab="Numero de Destellos de Luz", col=c("orange","cyan"))
68
69
70 #funciones creadas
71
72 #Funcion para calcular el coef de variacion
73 CoefVar<-function(x){
74   cv<-0
75   mean<-0
76   sd<-0
77   {
78     mean<-sapply(x,mean,na.rm=TRUE)
79     sd<-sapply(x,sd,na.rm=TRUE)
80     cv<-sd/mean
81   }
82   return(cv)
83 }
84
85 #Función para obtener la correlacion de Pearson
86 PearsonCorrelation<-function(data){
87   correlationFactor<-0
88   {
89     correlationFactor<-cor(data,use="complete.obs",method="pearson")
90   }
91   return(correlationFactor)
92 }
93
94 #Función que elimina la columna de clase
95 DeleteIDtype<-function(data){
96   {
97     data<-data[, -11]
98   }
99   return(data)
100 }
101
102 #CATEGORÍA de Hadrones
103 #Estadísticas de la primera categoria hadrones
104 typeh<-DeleteIDtype(typeh)
105 meanh<-sapply(typeh,mean,na.rm=TRUE)
106 sdh<-sapply(typeh,sd,na.rm=TRUE)
107 varh<-sapply(typeh,var,na.rm=TRUE)
108 minh<-sapply(typeh,min,na.rm=TRUE)
109 maxh<-sapply(typeh,max,na.rm=TRUE)
110 coefvarh<-CoefVar(typeh)

```

```

111
112 statsh<-matrix(c(meanh,sdh,varh,minh,maxh,coefvarh),nrow=10,ncol=6)
113 dimnames(statsh) = list(c("fLength","fwidth","fSize","fConc","fConc1","fAsym","fM3Long","fM3Trans",
114 "fAlpha","fDist"),c("Media","Desviacion","Varianza","Minimo","Maximo","Coef. Var."))
115
116 #Calculo de las modas
117 trendh_fLenght<-mfv(typeh$fLenght)
118 trendh_fwidth<-mfv(typeh$fwidth)
119 trendh_fSize<-mfv(typeh$fSize)
120 trendh_fConc<-mfv(typeh$fConc)
121 trendh_fConc1<-mfv(typeh$fConc1)
122 trendh_fAsym<-mfv(typeh$fAsym)
123 trendh_fM3Long<-mfv(typeh$fM3Long)
124 trendh_fM3Trans<-mfv(typeh$fM3Trans)
125 trendh_fAlpha<-mfv(typeh$fAlpha)
126 trendh_fDist<-mfv(typeh$fDist)
127
128 trendsh<-c(trendh_fLenght,trendh_fwidth,trendh_fSize,trendh_fConc,trendh_fConc1,trendh_fAsym,
129 trendh_fM3Long,trendh_fM3Trans,trendh_fAlpha,trendh_fDist)
130
131 #Histogramas para las variables continuas de la categoria hadron
132 histBy(typeh,var="fLength",main="Histograma para fLength en los destellos de hadrones")
133 histBy(typeh,var="fwidth",main="Histograma para fwidth en los destellos de hadrones")
134 histBy(typeh,var="fSize",main="Histograma para fSize en los destellos de hadrones")
135 histBy(typeh,var="fConc",main="Histograma para fConc en los destellos de hadrones")
136 histBy(typeh,var="fConc1",main="Histograma para fConc1 en los destellos de hadrones")
137 histBy(typeh,var="fAsym",main="Histograma para fAsym en los destellos de hadrones")
138 histBy(typeh,var="fM3Long",main="Histograma para fM3Long en los destellos de hadrones")
139 histBy(typeh,var="fM3Trans",main="Histograma para fM3Trans en los destellos de hadrones")
140 histBy(typeh,var="fAlpha",main="Histograma para fAlpha en los destellos de hadrones")
141 histBy(typeh,var="fDist",main="Histograma para fDist en los destellos de hadrones")
142
143 #calculo de la correlacion entre las variables
144 corh<-PearsonCorrelation(typeh)
145
146
147 #CATEGORIA de Gamma
148 #Estadísticas de la primera categoria gamma
149 typeg<-deleteIDtype(typeg)
150 meang<-sapply(typeg,mean,na.rm=TRUE)
151 sdg<-sapply(typeg,sd,na.rm=TRUE)
152 varg<-sapply(typeg,var,na.rm=TRUE)
153 ming<-sapply(typeg,min,na.rm=TRUE)
154 maxg<-sapply(typeg,max,na.rm=TRUE)
155 coefvarg<-CoefVar(typeg)
156
157 statsg<-matrix(c(meanh,sdg,varg,ming,maxg,coefVarg),nrow=10,ncol=6)
158 dimnames(statsg) = list(c("fLength","fwidth","fSize","fConc","fConc1","fAsym","fM3Long","fM3Trans",
159 "fAlpha","fDist"),c("Media","Desviacion","Varianza","Minimo","Maximo","Coef. Var."))
160
161 #Calculo de las modas
162 trendg_fLenght<-mfv(typeg$fLenght)
163 trendg_fwidth<-mfv(typeg$fwidth)
164 trendg_fSize<-mfv(typeg$fSize)
165 trendg_fConc<-mfv(typeg$fConc)

```

```
166 trendg_fConc1<-mfv(typeg$fConc1)
167 trendg_fAsym<-mfv(typeg$fAsym)
168 trendg_fm3Long<-mfv(typeg$fm3Long)
169 trendg_fm3Trans<-mfv(typeg$fm3Trans)
170 trendg_fAlpha<-mfv(typeg$fAlpha)
171 trendg_fDist<-mfv(typeg$fDist)
172
173 trendsh<-c(trendg_fLength,trendg_fwidth,trendg_fSize,trendg_fConc,trendg_fConc1,trendg_fAsym,
174           trendg_fm3Long,trendg_fm3Trans,trendg_fAlpha,trendg_fDist)
175
176 #Histogramas para las variables continuas de la categoria hadron
177 histBy(typeg,var="fLength",main="Histograma para fLength en los destellos de gamma")
178 histBy(typeg,var="fwidth",main="Histograma para fwidth en los destellos de gamma")
179 histBy(typeg,var="fSize",main="Histograma para fSize en los destellos de gamma")
180 histBy(typeg,var="fConc",main="Histograma para fConc en los destellos de gamma")
181 histBy(typeg,var="fConc1",main="Histograma para fConc1 en los destellos de gamma")
182 histBy(typeg,var="fAsym",main="Histograma para fAsym en los destellos de gamma")
183 histBy(typeg,var="fm3Long",main="Histograma para fm3Long en los destellos de gamma")
184 histBy(typeg,var="fm3Trans",main="Histograma para fm3Trans en los destellos de gamma")
185 histBy(typeg,var="fAlpha",main="Histograma para fAlpha en los destellos de gamma")
186 histBy(typeg,var="fDist",main="Histograma para fDist en los destellos de gamma")
187
188 #calculo de la correlacion entre las variables
189 corg<-PearsonCorrelation(typeg)
190
```