Tarea 3 Probabilidad y Estadística

Integrantes: Pablo Moraga

Alex Parada

Fecha de entrega :10/12/2021

Profesor: Demian Schcolnik

Ayudante: Abel Sierra

Sección 1

# **Introducción**

En el presente informe se analiza la distribución que se atribuye a un set de datos otorgado, además de justificar el porque es la distribución que se elige, además de esto se encontrará la varianza, desviación típica y la esperanza de dicha distribución.

En la segunda parte de este trabajo se analizará un caso en donde se aplica la distribución normal, el cual hará el análisis previo que se describió para la primera parte del trabajo y las similitudes que pueden haber en casos en donde se aplica la distribución normal.

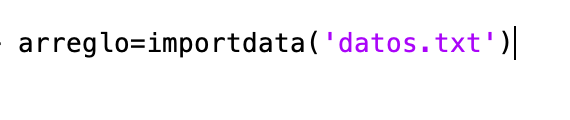
Cabe mencionar que el código que se creó para poder realizar esta tarea fue a través de Python y Matlab, el cual se describe en la sección siguiente.

# **Desarrollo del código**

# **Parte 1:**

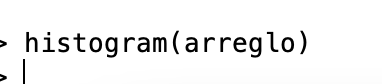
Para la primera parte de la actividad se utilizó el software MATLAB el cual nos proporciona los comandos necesarios para simplificar la realización de la tarea.

Para empezar se importan los datos provenientes del archivo con el comando importdata y se guardan en un arreglo denominado “arreglo”



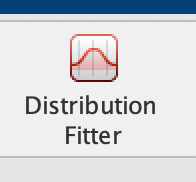
*Imagen 1:Importan el set de datos*

Luego se procede a realizar el gráfico correspondiente, el cual se realiza con el comando histogram proporcionado por matlab, además de esto para poder ver la distribución aplicada para el arreglo se instaló la extensión o app de MATLAB denominada “Statistics and Machine Learning Toolbox”, el cual nos proporciona la herramienta distribution filter que nos permitirá ajustar los datos a la distribución que nosotros queramos.



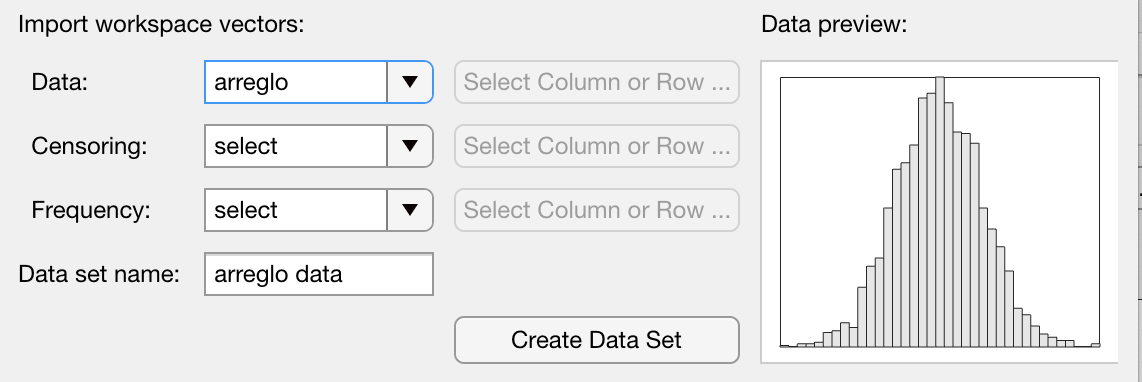
*Imagen 2:Uso de comando*

Para la herramienta usada anteriormente, al instalar la extensión en el apartado de apps de MATLAB se encontrará la siguiente opción.



*Imagen 3:Uso de la herramienta*

Al darle click se desplegará la siguiente ventana el cual debe elegir en data, el arreglo que proporcionará a aplicar la herramienta, luego de esto se presiona el botón create dataset y permitirá ver la visualización de la distribución.



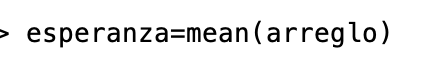
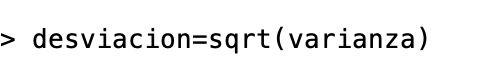
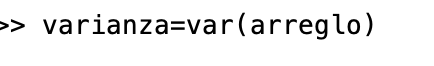
*Imagen 4:Uso de la herramienta*

Además para ajustar el gráfico que se realizó a una distribución normal directamente se realiza el siguiente comando.



*Imagen 5:Uso del comando*

Para calcular la varianza, desviación estándar y esperanza se utilizan los siguientes comandos los cuales nos permiten calcular dichos parámetros fácilmente.

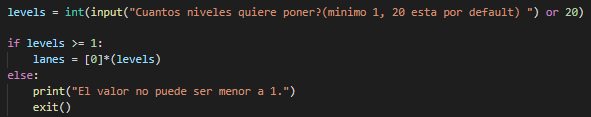


*Imagen 6:Usos de los comandos*

# **Parte 2:**

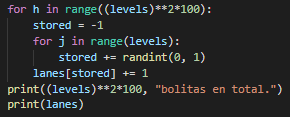
Para este caso se utilizó algo llamado la Máquina de Dalton, que consiste en unas pelotitas que se tiran y mientras van bajando tienden a ir hacia la derecha o hacia la izquierda y terminan cayendo en filas(niveles), las cuales al revisarlas tienden a una distribución normal.

Para simular la Máquina de Dalton se utilizó un arreglo en el cual se pregunta la cantidad de niveles que va a tener la simulación.



*Imagen 7:Creación de la máquina*

Entonces se hace una proporcion a esto de cuantas bolitas serían necesarias para que se note esta distribución, en este caso se usó la siguiente fórmula:



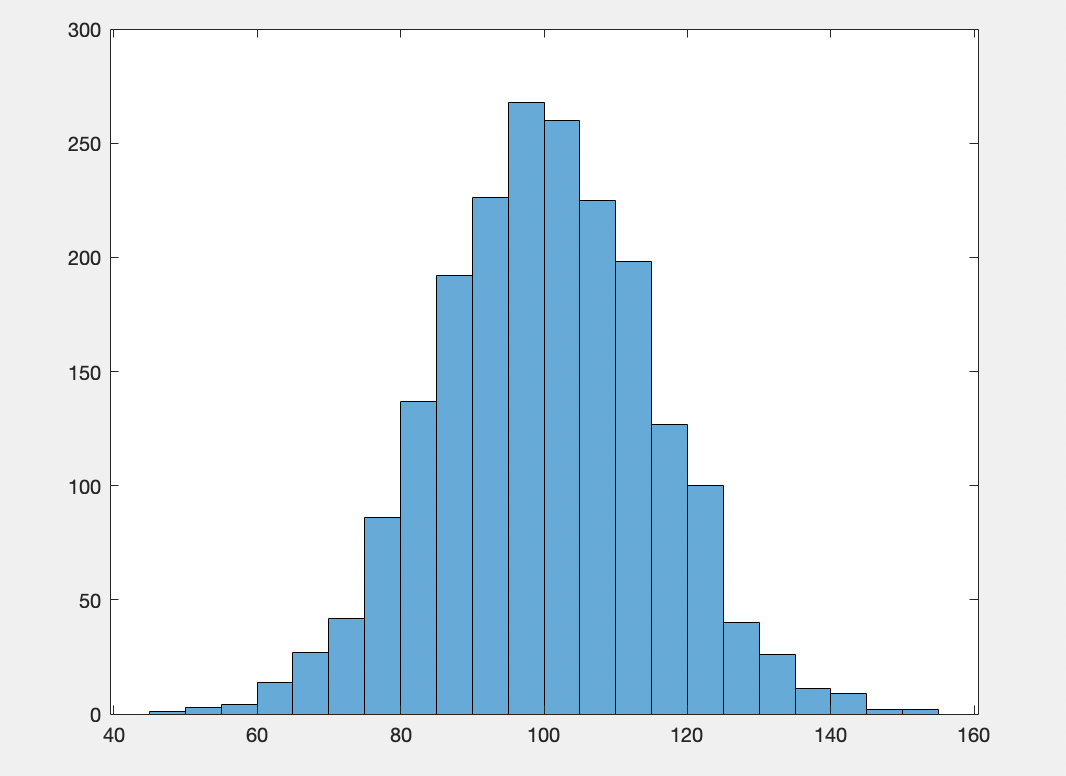
*Imagen 8:Creación de la máquina*

Finalmente cayendo en un nivel y guardándolo como se muestra en la anterior imagen, mostrando en pantalla cuántas bolitas se usaron e imprimiendo el arreglo por niveles de cuantas bolitas cayeron en qué nivel.

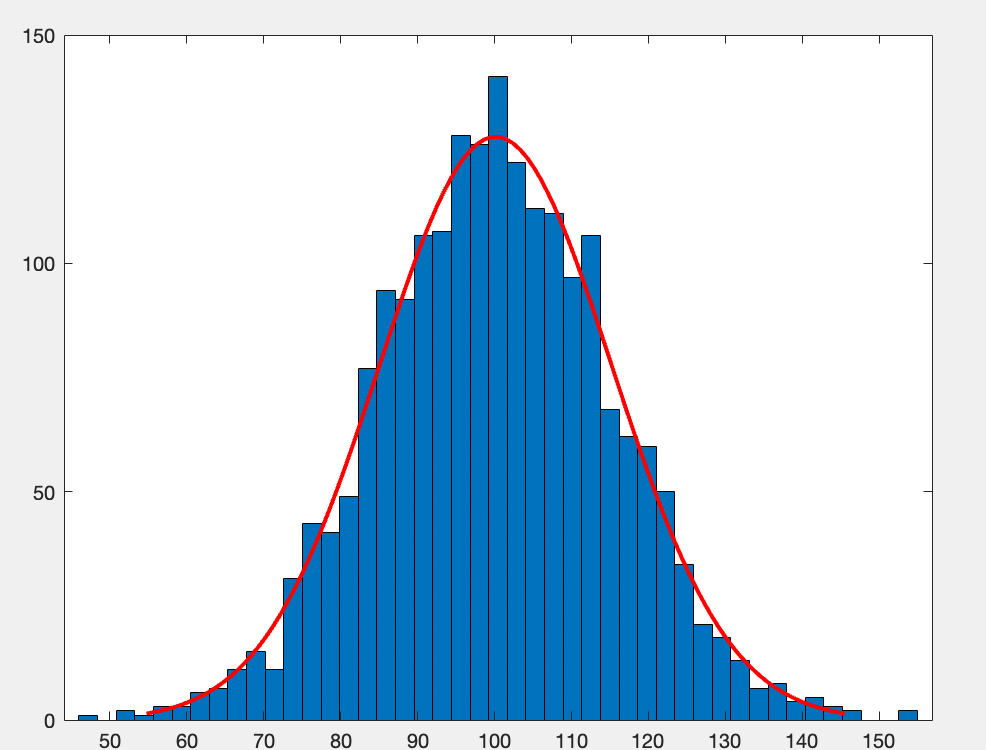
# **Resultados**

# **Parte 1:**

Primero que todo para calcular los parámetros que nos solicitan se debe analizar a qué distribución corresponde el set de datos proporcionados, por lo cual se realiza la gráfica que siguen los datos, la cual es la siguiente.



*Imagen 9:Histograma*

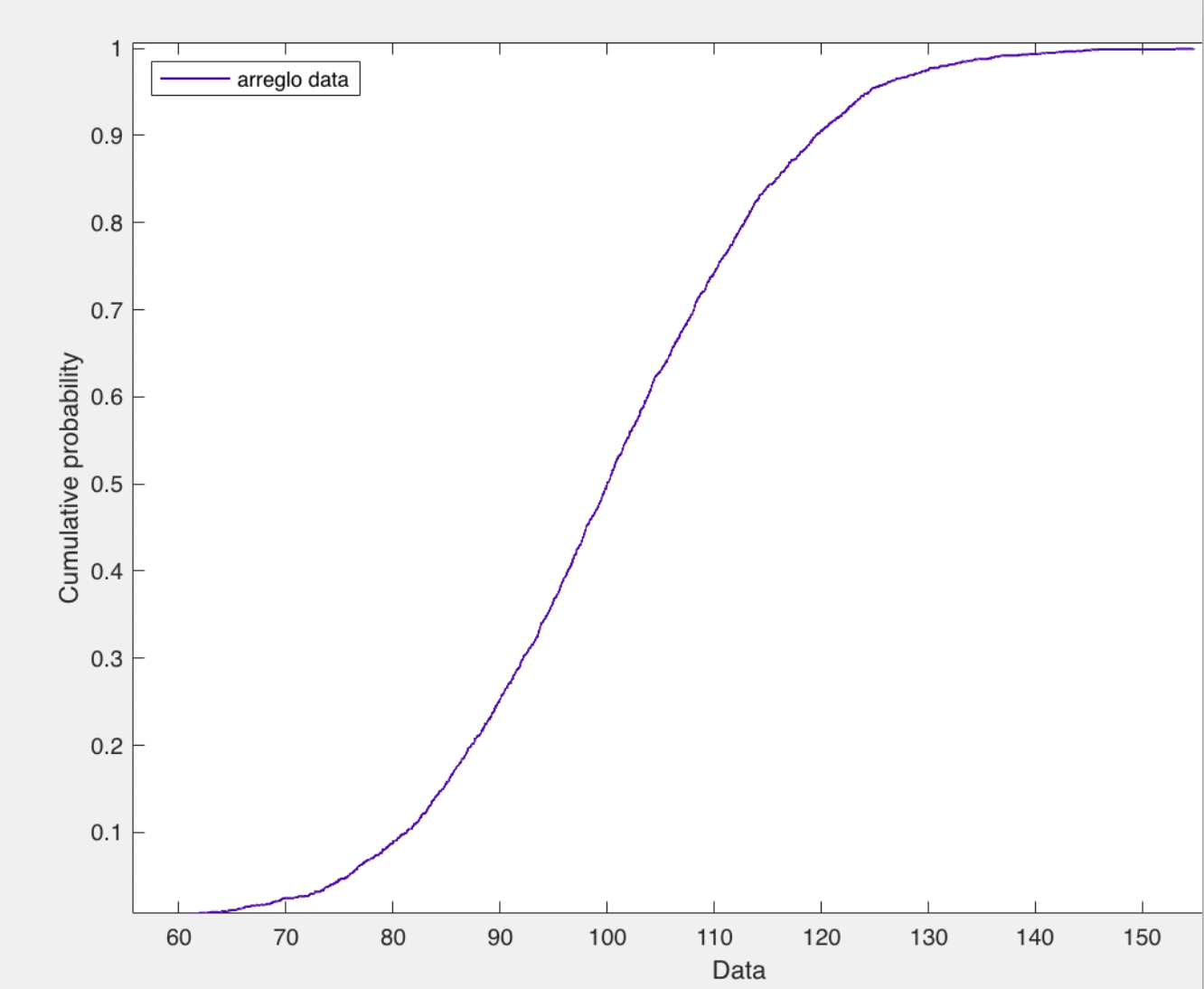


*Imagen 10:Gráfico distribución normal*

Como se muestra en el gráfico anterior, la gráfica de los datos se puede atribuir a una distribución normal por las siguientes razones.

1. Si bien el gráfico se atribuye a un histograma, se observa que la gráfica va tomando forma de campana de Gauss la cual es característica de una distribución normal, esto se observa más detalladamente en la figura que le acontece.
2. La media de una distribución normal es idéntica a la moda y mediana de dicha distribución, si se calcula estos parámetros se puede observar que los resultados son casi idénticos a la media calculada para este set de datos (los datos se muestran en la siguiente sección),sin embargo se puede atribuir que estos valores se encuentran en el centro.
3. Como se puede observar en la imagen 10,el gráfico tiende a ser simétrico, el cual es una característica fundamental en la distribución normal.
4. Por último se observa que el máximo que alcanza a llegar en el gráfico es cercano al valor de la media, lo cual es otra característica de la distribución normal la cual su media toma el valor de su máximo.

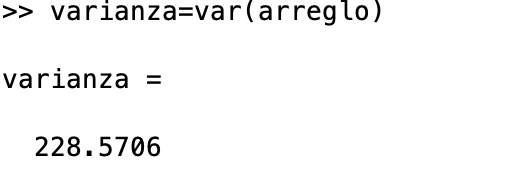
Antes de continuar, se adjunta la gráfica acumulada por porcentaje creada a partir de la extensión mencionada en el apartado “desarrollo de código”



*Imagen 11:Distribución acumulada*

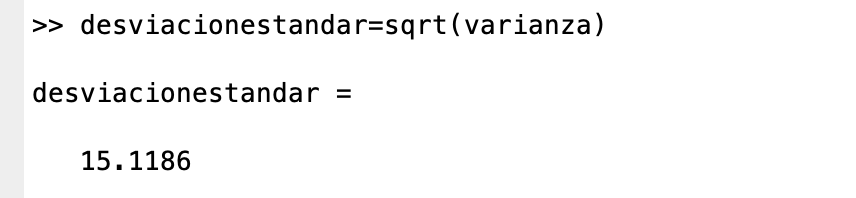
A continuación se muestran los resultados obtenidos al calcular los parámetros varianza,desviación estándar y esperanza.

# **Varianza:**



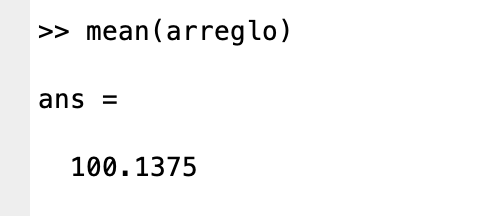
*Imagen 12:Varianza*

# **Desviación estándar:**



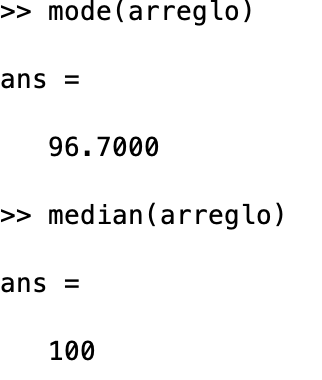
*Imagen 13:Desviación estándar*

# **Esperanza:**



*Imagen 14:Esperanza*

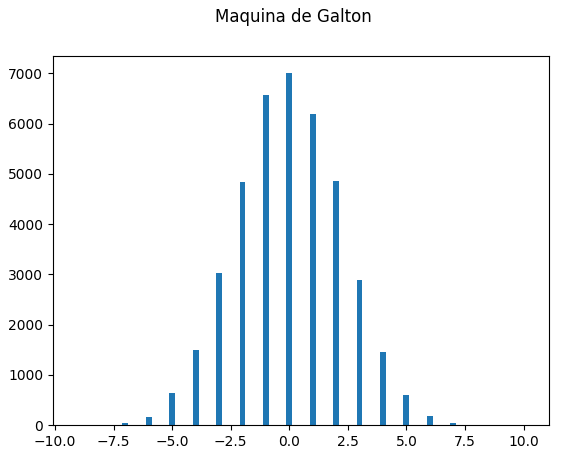
Además de los datos anteriores, se adjuntan la moda y mediana de los datos del gráfico para corroborar lo dicho del apartado anterior.



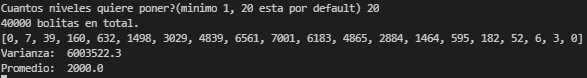
*Imagen 16:Moda y Mediana*

# **Parte 2:**

El resultado de la Máquina de Galton usando como referencia 20 niveles, es el siguiente:



*Imagen 17: Gráfico 2*



*Imagen 18: Set de datos*

En este resultado la varianza y la esperanza fueron los siguientes:

**Varianza: 6.003.522,3**

**Desviación estándar: 2513**

**Promedio: 2000.0**

Este histograma como se puede apreciar se asemeja bastante a la campana de Gauss

Ya realizado el análisis correspondiente para cada caso, se procede a comparar ambos gráficos obtenidos.

A primera vista lo que más resalta en los gráficos es la distribución de los datos, los cuales se muestra que el valor máximo(moda) y la mediana calzan con el valor de la media o son casi iguales, esto se atribuye a que la distribución que se imparte es una distribución normal en ambos casos.

Otra comparación es la varianza la cual con respecto al valor de los datos en cada uno de los casos es muy alta y esto es porque en una distribución normal los datos están muy dispersos y mientras más dispersos estén los datos la varianza es exponencial entonces es un valor muy alto.

# **Conclusión**

En este informe se evidenció la distribución normal para diferentes set de datos, por una parte un set otorgado y por otro lado un set proveniente de una máquina de galton simulada, cabe mencionar que esta máquina se puede encontrar en la vida real.

Bajo esto se evidencio que la distribución normal establece diferentes condiciones que son comunes a diferentes circunstancias que se pueden encontrar en la vida cotidiana, estas condiciones son características que se encuentran en una distribución normal las cuales son: una forma de campana de Gauss, una varianza alta y las medidas estadísticas se encuentran en el centro de esta distribución.

Además al tener una varianza alta como se mencionó anteriormente, se establece que la distribución normal posee una dispersión alta de datos por lo cual la media aritmética no es muy representativa.