Actividad 1: Distribuciones de Probabilidad

Jacob Valdenegro Monzón A01640992
Roberto Ángel Rillo Calva A01642022

1.- Una barra de 12 pulg que está sujeta por ambos extremos se somete a una cantidad creciente de esfuerzo hasta que se rompe. Sea Y = la distancia del extremo izquierdo al punto donde ocurre la ruptura. Suponga que Y tiene la función de densidad de probabilidad:

$$f(y) = \{\left(\frac{1}{24}\right)y\left(1 - \frac{y}{12}\right),\ 0 \le y \le 12\ 0,\ \textit{De lo contrario}$$

Calcule lo siguiente:

a.- La función de distribución acumulativa de Y.
b.- P(Y ≤ 4), P(Y > 6) y P(4 ≤ Y ≤ 6)
c.- E(Y), E(Y^2) y Var(Y).
d.- La probabilidad de que el punto de ruptura ocurra a más de 2 pulg del punto de ruptura esperado.

a) La función de distribución acumulativa de Y = y**2*(0.0208333333333333 - 0.00115740740740741*y)

b) P(Y ≤ 4) = 0.259259259259259 2586
P(Y > 6) = 0.500000000000018
P(4 ≤ Y ≤ 6) = 0.2407407407407397

c) E(Y) = 6.00000000000000
E(Y^2) = 43.2000000000000
Var(Y) = 7.20000000000002

d) La probabilidad de que el punto de ruptura ocurra a más de 2 pulgadas del punto de ruptura esperado es 0.518518518518521

2.- Sea X la temperatura, en grados centígrados, a la cual ocurre una reacción química. Suponga que X tiene una función de densidad de probabilidad:

$$f(x) = \{\frac{1}{9}\quad 0\left(4 - x^2\right),\quad -1 \le x \le 2$$

a.- Corrobore que la función es una distribución válida.
b.- Determine la función de distribución acumulativa.
c.- E(Y), E(Y^2) y Var(Y).
d.- La probabilidad de que la temperatura sea menor a 0°C
e.- La probabilidad de que la temperatura sea entre 4°C y 6°C

a) La función es válida porque la integral de f(x) sobre su dominio es 1.00000000000000

b) La función de distribución acumulada es -0.037037037037037*x**3 + 0.444444444444444*x + 0.407407407407407

c) E(Y) = 0.250000000000000
E(Y^2) = 0.600000000000000
Var(Y) = 0.537500000000000

d) La probabilidad de que la temperatura sea menor a 0°C es 0.407407407407407

e) La probabilidad de que la temperatura sea entre 4°C y 6°C es 0 ya que el dominio de x está restringido a [-1, 2]

3.- El artículo "Computer Assisted Net Weight Control" (Quality Progress, 1983: 22-25) sugiere una distribución normal con media de 137.2 oz y una desviación estándar de 1.6 oz del contenido real de frascos de cierto tipo. El contenido declarado fue de 135 oz.

a.- ¿Cuál es la probabilidad de que un solo frasco contenga más que el contenido declarado?

b.- Suponiendo que la media permanece en 137.2, ¿a qué valor se tendría que cambiar la desviación estándar de modo que 95% de todos los frascos contengan más que el contenido declarado?

c.- Entre 10 frascos seleccionados al azar, ¿cuál es la probabilidad de que por lo menos ocho contengan más que el contenido declarado?

a) La probabilidad de que un solo frasco contenga más que el contenido declarado es 0.9154342776486631

b) La nueva desviación estándar necesaria es 1.3375050302058848

c) La probabilidad de que al menos 8 de 10 frascos contengan más que el contenido declarado es 0.9538238329543585

4.- El artículo "Characterization of Room Temperature Damping in Aluminum-Idium Alloys" (Metallurgical Trans., 1993: 1611-1619) sugiere que el tamaño de grano de matriz A1 (μm) de una aleación compuesta de 2% de indio podría ser modelado con una distribución normal con valor medio de 96 y desviación estándar de 14.

a.- ¿Cuál es la probabilidad de que el tamaño de grano exceda de 100?

b.- ¿Cuál es la probabilidad de que el tamaño de grano sea de 50 y 80?

c.- ¿Qué intervalo (a, b) incluye el 90% central de todos los tamaños de grano (de modo que 5% esté por debajo de a y 5% por encima de b)?

a) La probabilidad de que un solo frasco contenga más que el contenido declarado es 0.9154342776486631

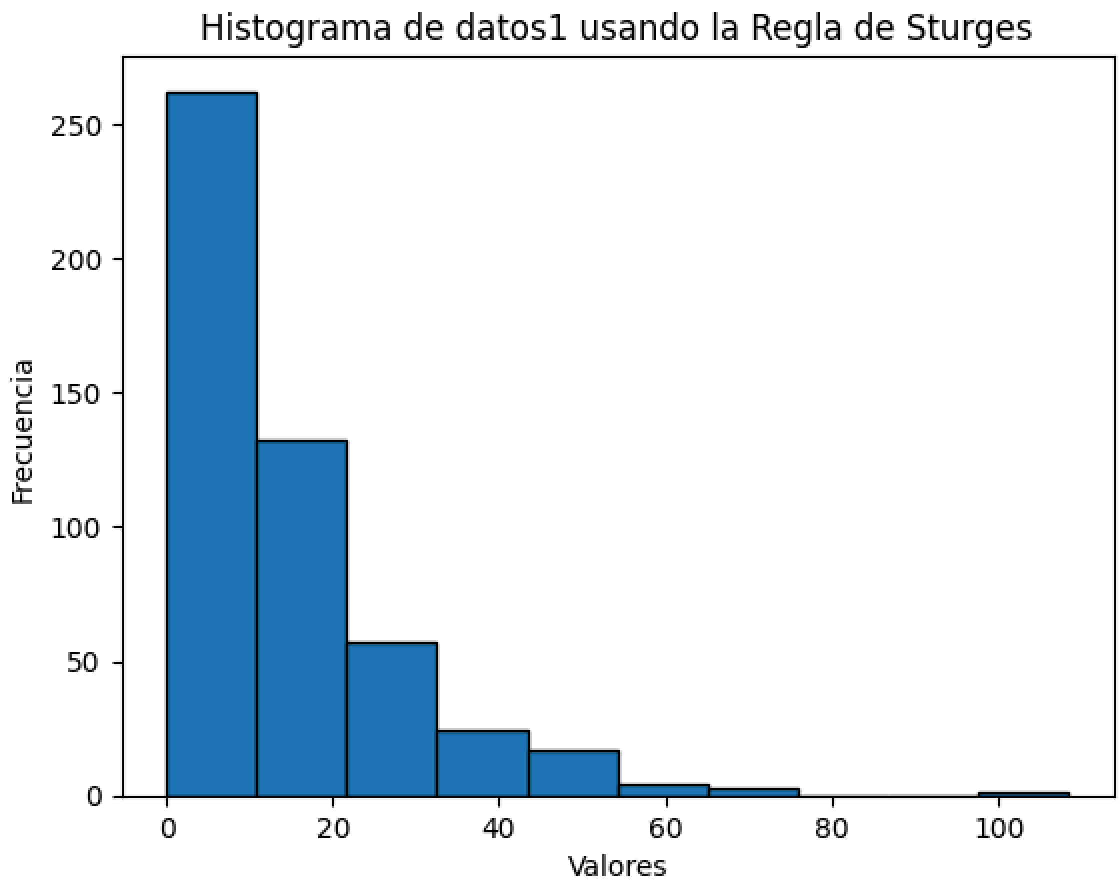
b) La nueva desviación estándar necesaria es 1.3375050302058848

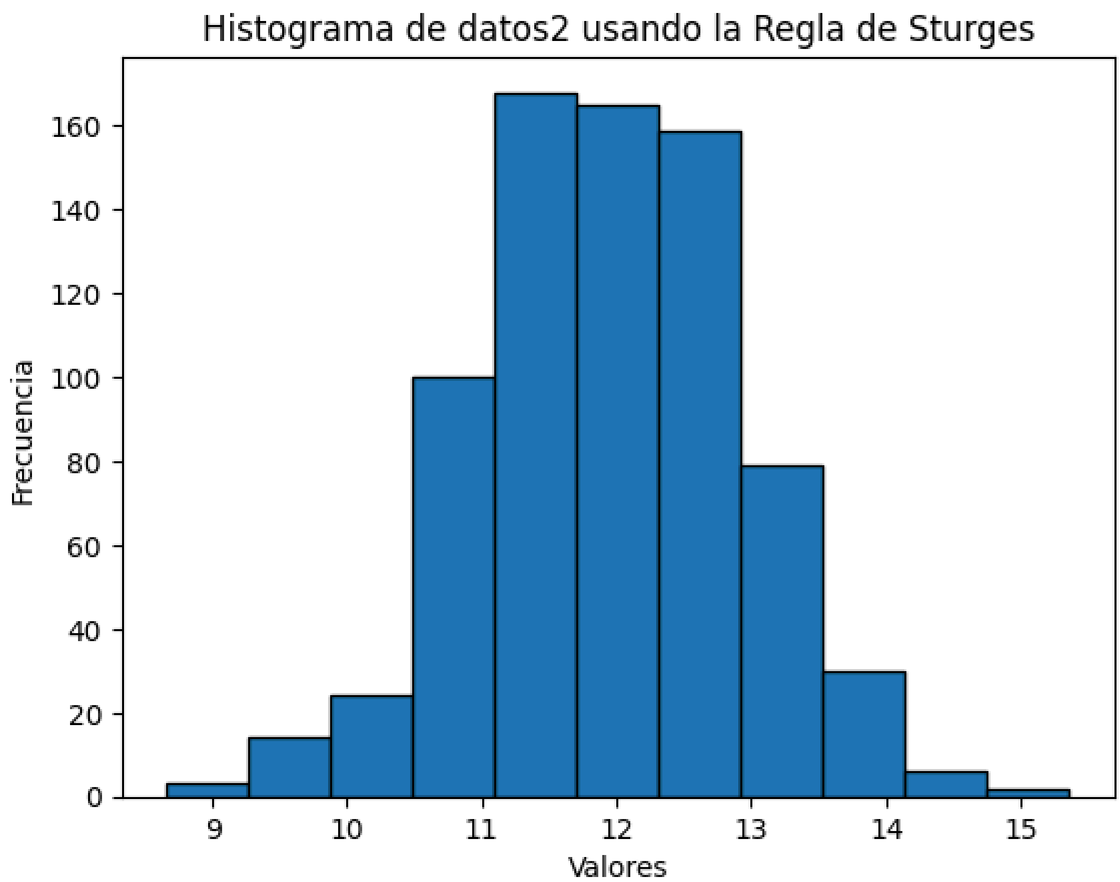
c) La probabilidad de que al menos 8 de 10 frascos contengan más que el contenido declarado es 0.9538238329543585

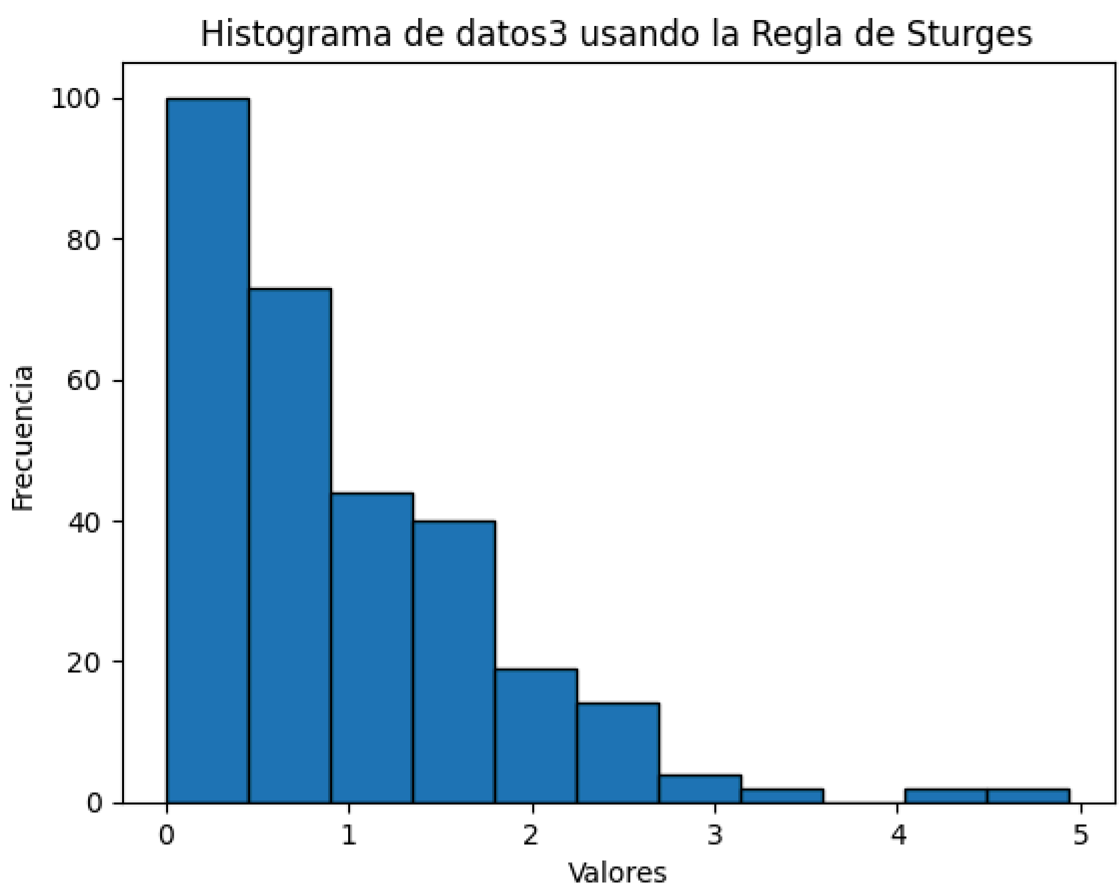
5.- Para los 3 conjuntos de datos que se proveen en el CSV:

a.- Construye e interpreta un histograma. Utiliza la regla de Sturges para calcular el número apropiado de clases.

Número de clases de Datos1 según Sturges: 10
Número de clases de Datos2 según Sturges: 11
Número de clases de Datos3 según Sturges: 10

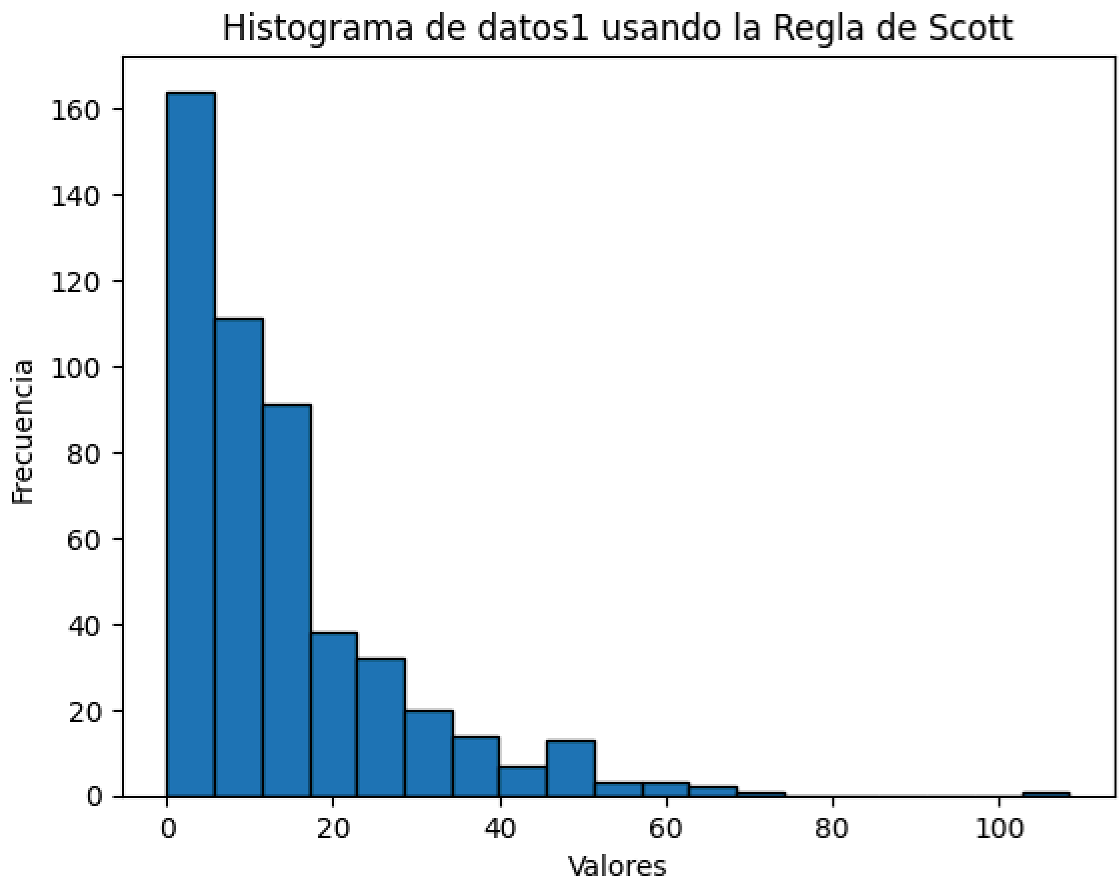


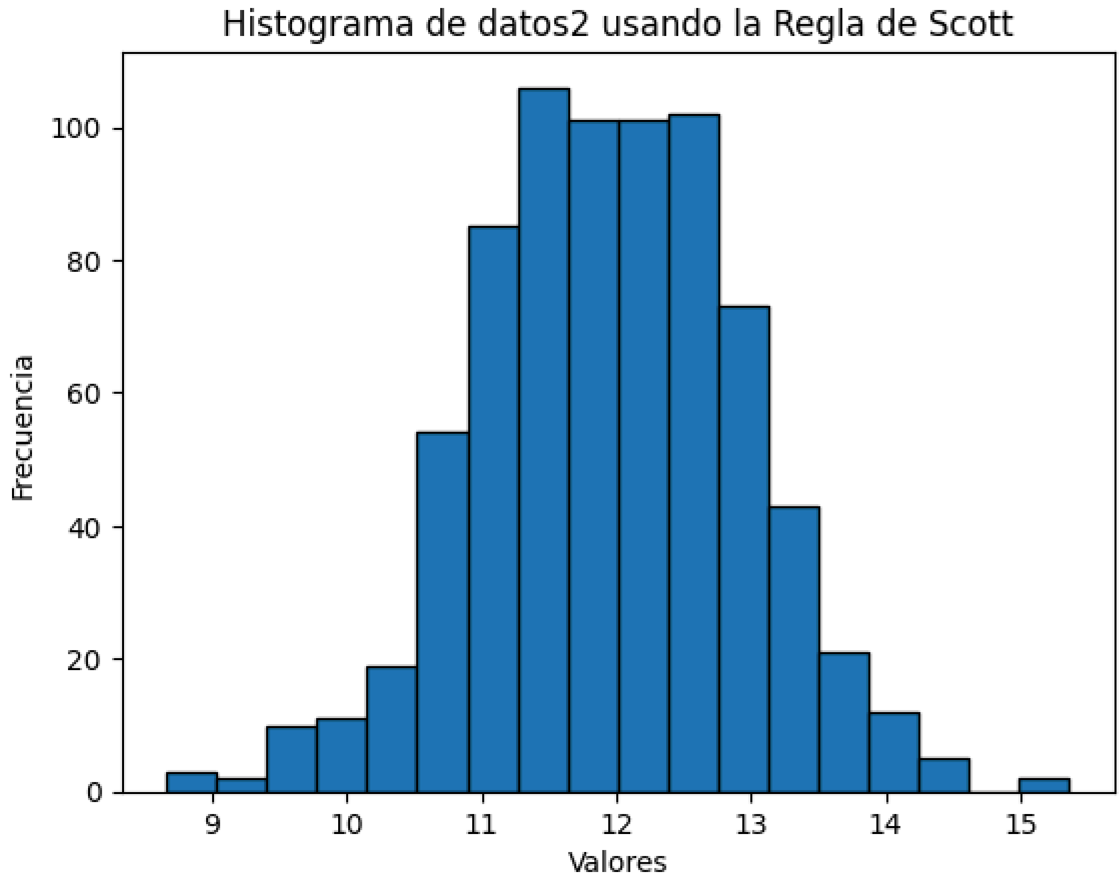

Histograma de datos2 usando la Regla de Sturges
Frecuencia
Valores
0
20
40
60
80
100
120
140
160
9
10
11
12
13
14
15

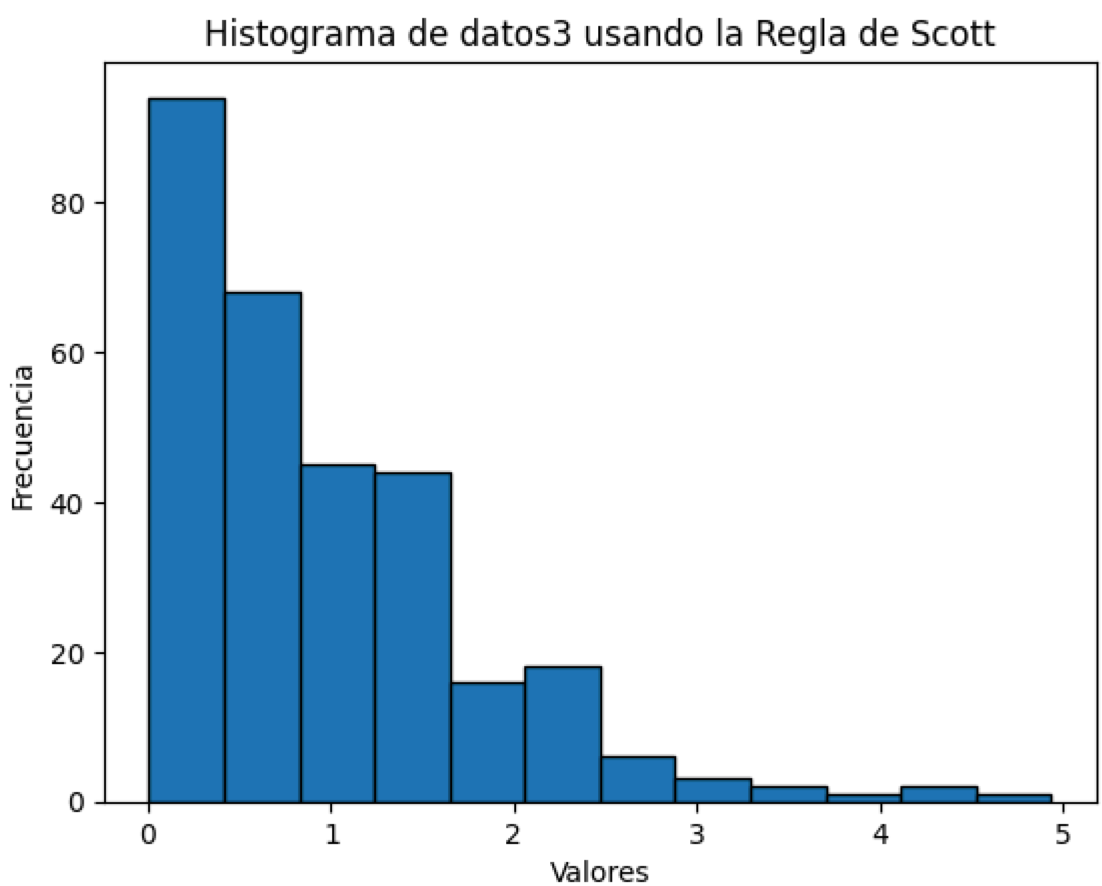

Histograma de datos3 usando la Regla de Sturges
Frecuencia
Valores
0
20
40
60
80
100
0
1
2
3
4
5

b.- Compara el número de clases con el obtenido con la regla de Scott.

Número de clases de datos1 según la Regla de Scott: 19
Número de clases de datos2 según la Regla de Scott: 18
Número de clases de datos3 según la Regla de Scott: 12



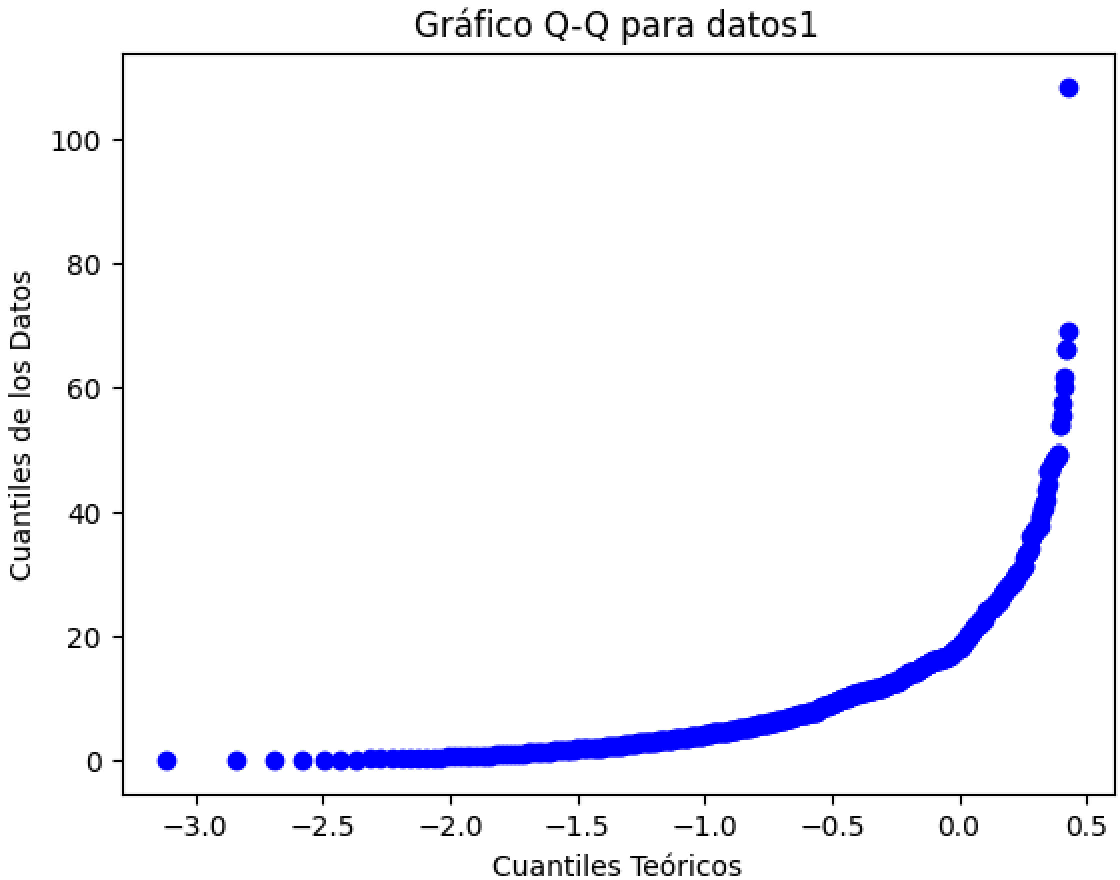

Histograma de datos2 usando la Regla de Scott
Frecuencia
0
20
40
60
80
100
9
10
11
12
13
14
15
Valores


Histograma de datos3 usando la Regla de Scott
Frecuencia
0
20
40
60
80
0
1
2
3
4
5
Valores

Interpretación: El histograma 1 y 3 muestran una distribución asimétrica a la derecha, no obstante, en el histograma 2 podemos observar una distribución más normalizada debido a su forma de “campana”. Al igual que en la regla de Scott pero con una mayor cantidad de clases en cuanto a los diferentes datos analizados

c.- Construye e interpreta un gráfico Q-Q para comprobar si los datos provienen de una distribución normal. Estima los parámetros utilizando la regresión de un gráfico probabilístico.

Los datos 1 y 3 no provienen de una distribución normal ya que no siguen una línea recta cercana a la diagonal y los puntos se desvían en las colas. los datos 2 provienen de una distribución normal



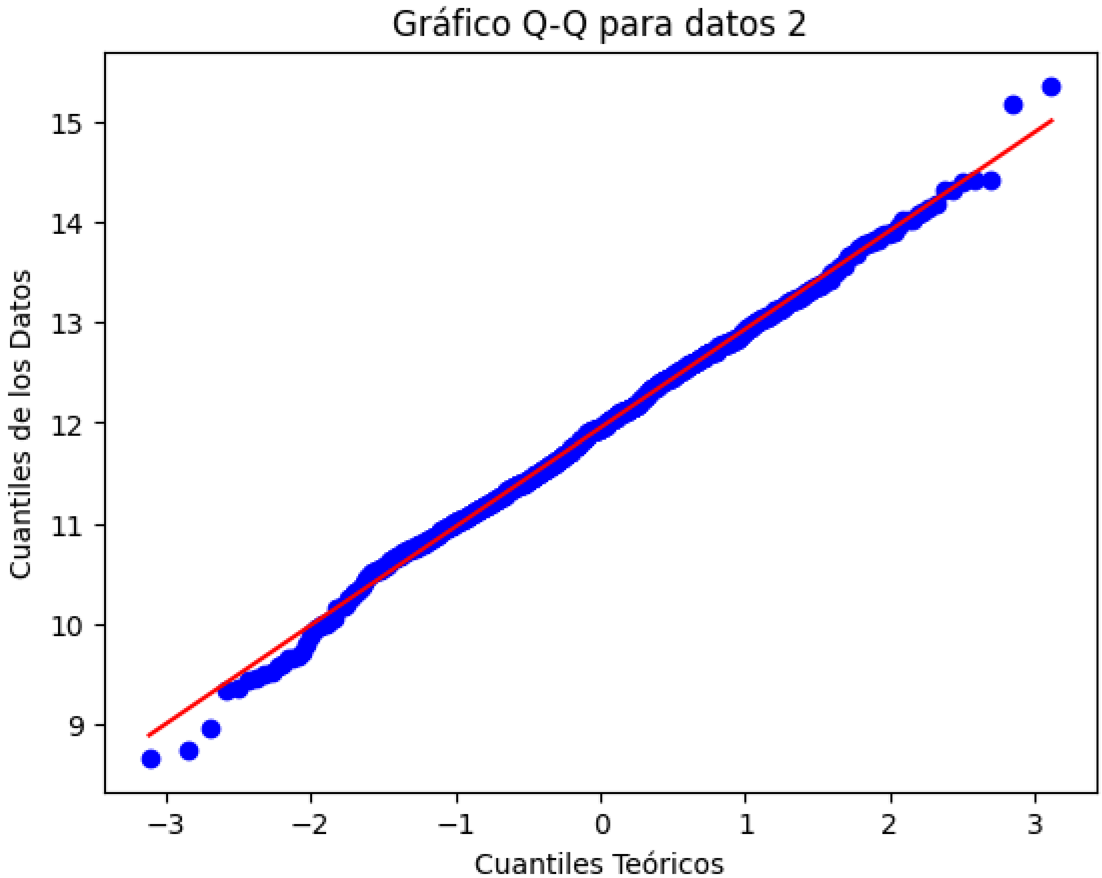
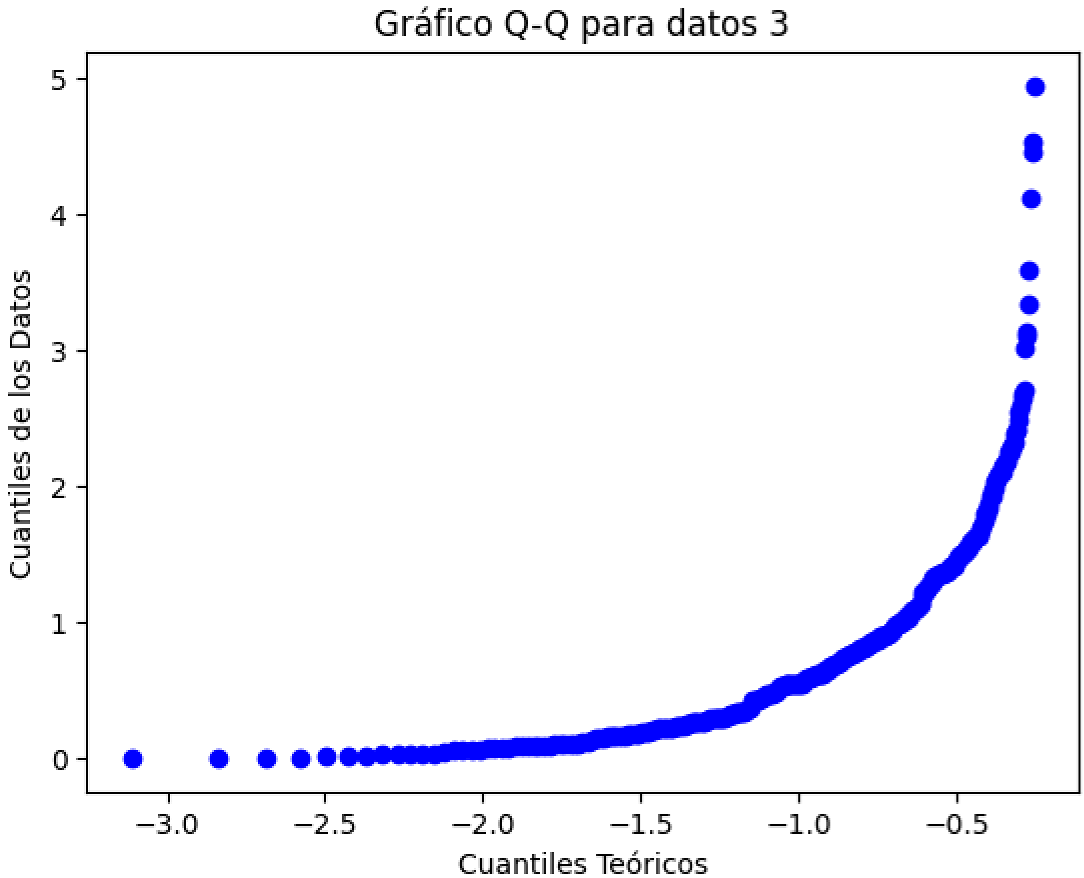


Gráfico Q-Q para datos 2
Cuantiles de los Datos
Cuantiles Teóricos
15
14
13
12
11
10
9
−3
−2
−1
0
1
2
3



Observación: Para este ejemplo se han llenado los campos null de los datos1 y datos3 como objeto de prueba, sin embargo, los resultados son los mismos en cuestión de normalidad distributiva

d.- Utilizando Minitab o algún otro software, ¿a qué distribución es más probable que pertenezca cada conjunto de datos y cuáles serían sus respectivos parámetros?

Los datos analizados dan a entender los datos 1 y 3 pertenecen a una distribución asimétrica" ya que no tienen forma de "campana" y tienen forma descendente hacia la derecha, sin embargo los datos 2 tienen una distribución normal o geométrica y sus respectivos parámetros de los datos 2 serían:

Media estimada: 11.949312
Desviación estándar estimada: 0.9830894759821239

No obstante, para los demás datos tendría los resultados de nan ya que no son los parámetros esperados para ese tipo de distribuciones, así que usamos minitab para la solución de de este inciso y este indica que la distribución 1 y 3 es exponencial o gama

Distribution Identification for Datos1

Distribution	AD	P	LRT P
Normal	23.861	<0.005	
Box-Cox Transformation	0.202	0.878	
Lognormal	5.489	<0.005	
3-Parameter Lognormal	1.098	*	0.000
Exponential	0.588	0.387	
2-Parameter Exponential	0.552	>0.250	0.284
Weibull	0.291	>0.250	
3-Parameter Weibull	0.303	>0.500	0.133
Smallest Extreme Value	53.690	<0.010	
Largest Extreme Value	8.461	<0.010	
Gamma	0.252	>0.250	
3-Parameter Gamma	0.251	*	0.735
Logistic	15.578	<0.005	
Loglogistic	2.923	<0.005	
3-Parameter Loglogistic	1.961	*	0.006
Johnson Transformation	0.278	0.649	

Distribution Identification for Datos2

Distribution	AD	P	LRT P
Normal	0.384	0.393	
Box-Cox Transformation	0.384	0.393	
Lognormal	0.986	0.013	
3-Parameter Lognormal	0.382	*	0.000
Exponential	291.447	<0.003	
2-Parameter Exponential	169.727	<0.010	0.000
Weibull	5.340	<0.010	
3-Parameter Weibull	1.012	<0.005	0.000
Smallest Extreme Value	9.321	<0.010	
Largest Extreme Value	11.033	<0.010	
Gamma	0.668	0.085	
3-Parameter Gamma	0.641	*	0.158
Logistic	0.961	0.007	
Loglogistic	1.162	<0.005	
3-Parameter Loglogistic	1.508	*	0.035

Distribution Identification for Datos3

Goodness of Fit Test

Distribution	AD	P	LRT P
Normal	8.664	<0.005	
Box-Cox Transformation	0.667	0.081	
Lognormal	5.440	<0.005	
3-Parameter Lognormal	2.050	*	0.000
Exponential	1.188	0.072	
2-Parameter Exponential	1.154	0.078	0.285
Weibull	0.575	0.151	
3-Parameter Weibull	0.663	0.089	0.254
Smallest Extreme Value	22.691	<0.010	
Largest Extreme Value	3.367	<0.010	
Gamma	0.703	0.082	
3-Parameter Gamma	0.695	*	1.000
Logistic	6.155	<0.005	
Loglogistic	3.780	<0.005	
3-Parameter Loglogistic	2.789	*	0.015