Actividad 1: Distribuciones de Probabilidad

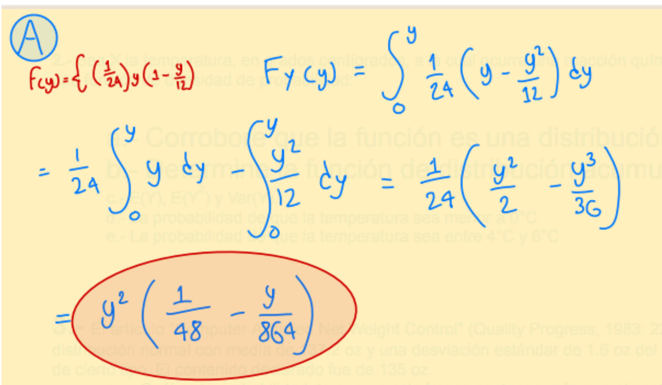
1.- Una barra de 12 pulg que está sujeta por ambos extremos se somete a una cantidad creciente de esfuerzo hasta que se rompe. Sea Y = la distancia del extremo izquierdo al punto donde ocurre la ruptura. Suponga que Y tiene la función de densidad de probabilidad:

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

Calcule lo siguiente:

a.- La función de distribución acumulativa de Y.

$y^2(\frac{1}{48}-\frac{y}{864})$

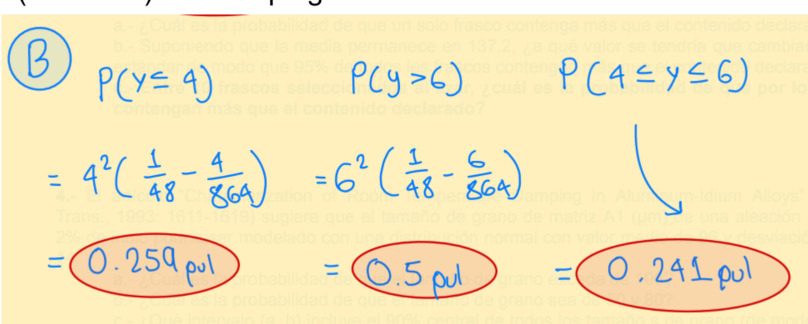


b.- P(Y ≤ 4), P(Y > 6) y P(4 ≤ Y ≤ 6)

P(Y ≤ 4) = 0.259 pulg

P(Y > 6) = 0.5 pulg

P(4 ≤ Y ≤ 6) = 0.241 pulg



c.- E(Y), E(Y^2) y Var(Y).

E(Y) = 43.3

E(Y^2) = 6

Var(Y) = 7.3

C

$$E(x) = \int_0^y y \cdot \left(\frac{1}{24} y \left(1 - \frac{y}{12}\right)\right) dy$$

$$= \frac{1}{24}\int_0^y y\left(y - \frac{y^2}{12}\right) dy = \frac{1}{24}\int_0^y \left(y^2 - \frac{y^3}{12}\right) d$$

$$= \frac{1}{24}\left(\frac{y^3}{3} - \frac{y^4}{48}\right) = y^3\left(\frac{1}{72} - \frac{y}{1152}\right)$$

$$E(y^2) = \frac{1}{24}\left(\frac{y^4}{4} - \frac{y^5}{60}\right) = y^4\left(\frac{1}{96} - \frac{y}{1440}\right)$$

$$Var(x) = E(x^2) - [E(x)]^2$$

$$Var(y) = 43.3 - 36 = 7.3 \text{ pul}$$

d.- La probabilidad de que el punto de ruptura ocurra a más de 2 pulg del punto de ruptura esperado.

45.6%

D

En python $\quad P(Y-6>2)$

$$Z = \frac{Y - E(y)}{\sigma}$$

$Y > 8$ 2 pulg Arriba

$Y < 4$ 2 pulg abajo

$$P(Y-6>2) = 2 = \frac{8-6}{2.683} + \frac{4-6}{2.683} = 45.60\ \%$$

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:

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

a.- Corrobore que la función es una distribución válida.

La función sí es válida, ya que al calcular la integral en el intervalo indicado el resultado es igual a 1.

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \rightarrow \frac{1}{9}\int_{-1}^{2}(4-x^2)\,dx = \frac{1}{9}\left(\int_{-1}^{2} 4\,dx - \int_{-1}^{2} x^2\,dx\right)$$

$$\frac{1}{9}\left(\underbrace{4x\Big|_{-1}^{2}}_{12} - \underbrace{\frac{x^3}{3}\Big|_{-1}^{2}}_{3}\right) = \frac{1}{9}\left([12-3]\right) = 1$$

b.- Determine la función de distribución acumulativa.

$\frac{1}{9}(4x - \frac{x^3}{3} + \frac{11}{3})$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-1}^{x} \frac{1}{9}(4-x^2)\,dx = \frac{1}{9}\left(4x\Big|_{-1}^{x} - \frac{x^3}{3}\Big|_{-1}^{x}\right) = \frac{1}{9}\left(\left[4x-\frac{x^3}{3}\right]-\left[-4-\frac{-1}{3}\right]\right)$$

$$= \frac{1}{9}\left(4x-\frac{x^3}{3}+\frac{11}{3}\right)$$

c.- E(Y), E(Y^2) y Var(Y).

E(X) = 0.25
E(X^2) = 0.6
Var(X) = 0.5375

$$E(x) = \int_{-\infty}^{\infty} x\,f(x)\,dx = \int_{-1}^{2} x\cdot\frac{1}{9}(4-x^2)\,dx = \frac{1}{9}\left(\int_{-1}^{2} 4x\,dx - \int_{-1}^{2} x^3\,dx\right) = \frac{1}{9}\left(\underbrace{2x^2\Big|_{-1}^{2}}_{6} - \underbrace{\frac{x^4}{4}\Big|_{-1}^{2}}_{\frac{15}{4}}\right) = \frac{1}{4}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \frac{1}{9}\left(\int_{-1}^{2} 4x^2\,dx - \int_{-1}^{2} x^4\,dx\right) = \frac{1}{9}\left(\underbrace{\frac{4x^3}{3}\Big|_{-1}^{2}}_{12} - \underbrace{\frac{x^5}{5}\Big|_{-1}^{2}}_{\frac{33}{5}}\right) = \frac{3}{5}$$

$$Var(x) = E(x^2) - [E(x)]^2 = 0.5375$$

d.- La probabilidad de que la temperatura sea menor a 0°C

Es igual a 40.7%

$$P(X<0) = \int_{-1}^{0} f(x)\,dx = \frac{1}{9}\int_{-1}^{0}(4-x^2)\,dx = \frac{1}{9}\left(\underbrace{4x\Big|_{-1}^{0}}_{-4} - \underbrace{\frac{x^3}{3}\Big|_{-1}^{0}}_{-\frac{1}{3}}\right) = -\frac{11}{27} = 0.407$$

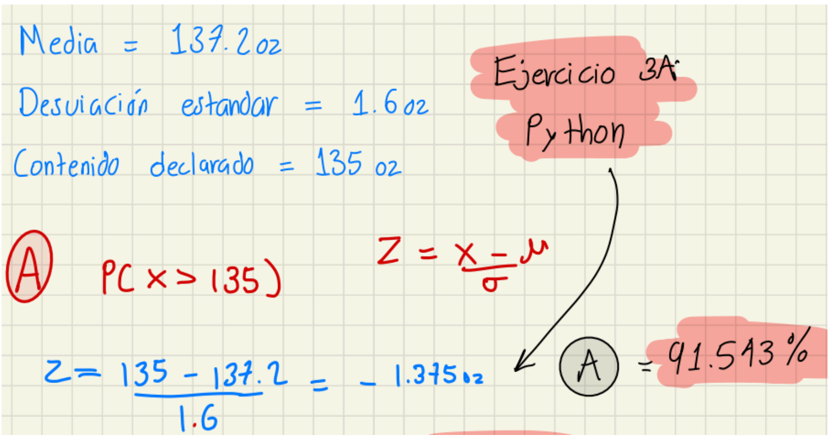
e.- La probabilidad de que la temperatura sea entre 4°C y 6°C

4°C y 6°C están fuera del intervalo por la que la probabilidad es igual a 0.

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.

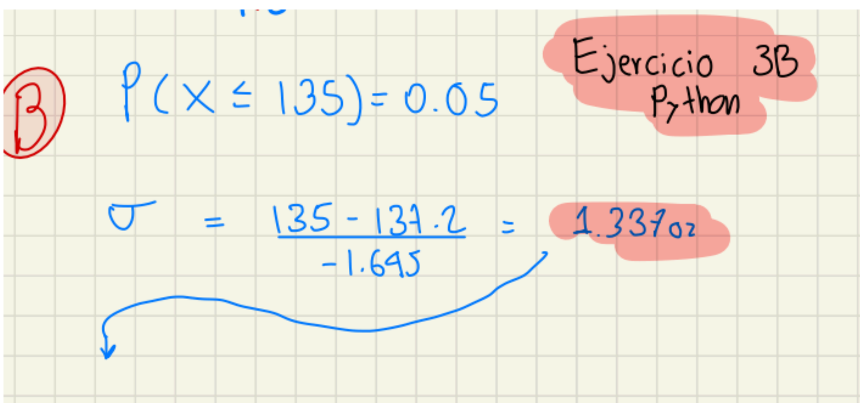
a.- ¿Cuál es la probabilidad de que un solo frasco contenga más que el contenido declarado?

91.5%



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?

1..337 oz



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

La probabilidad de que al menos 8 de los 10 frascos contengan más que el contenido declarado es: 95.22%

C

Dado que buscamos al menos 8, tenemos que sumar las probabilidades de que 8, 9, 10 contengan mas que el contenido declarado.

Distribución Binomial

Ejercicio 3C
Python

$n = 10$

$p = 0.914$

Biom.sf(7, n, p)

↓

Calcula la probabilidad de que haya mas de 7 exitos.

95.22%

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?

38.75%

$\mu = 96$ A

Ejercicio 4A
Python

38.75%

$\sigma = 14$

$X = 100 \rightarrow z = \frac{X - \mu}{\sigma} = 0.28571$

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

12.6%

B

$x = 50$

$x = 80$

Ejercicio 4B
Python

12.60%

$$z_{50} = \frac{50-96}{14} = -3.2857$$

$$z_{80} = \frac{80-96}{14} = -1.14285$$

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

El intervalo que incluye el 90% central de los tamaños de grano es: (72.97, 119.03)

Norm.cdf: Te da la probabilidad de que una variable aleatoria este por debajo de (X)

C

$$x = z \cdot \sigma + \mu$$

Ejercicio 4C
Python

Necesitamos encontrar los puntos de corte en cuales el 5% quede en la parte inferior y 5% en la superior.

Utilizando el percentil 5 y 95

(72.97 µm, 119.03 µm)

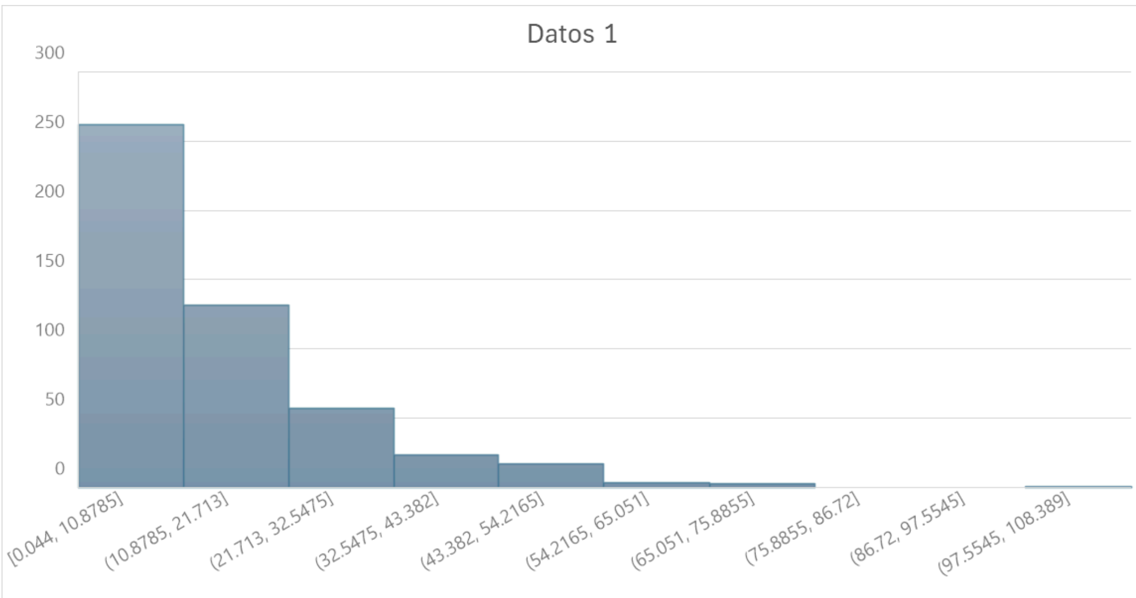
Esto significa que solo el 5% de los granos sera menor a 72.97 µm y el otro 5% sera mayor a 119.03 µm.

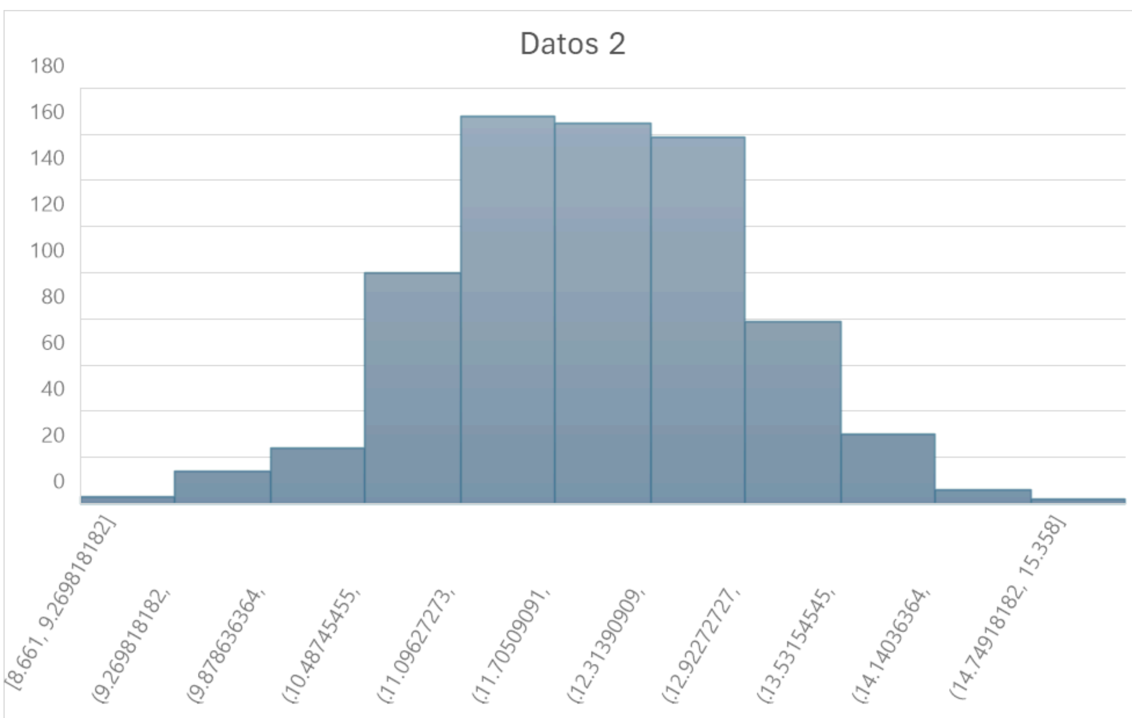
5.- Para los 3 conjuntos de datos que se proveen en el CSV:

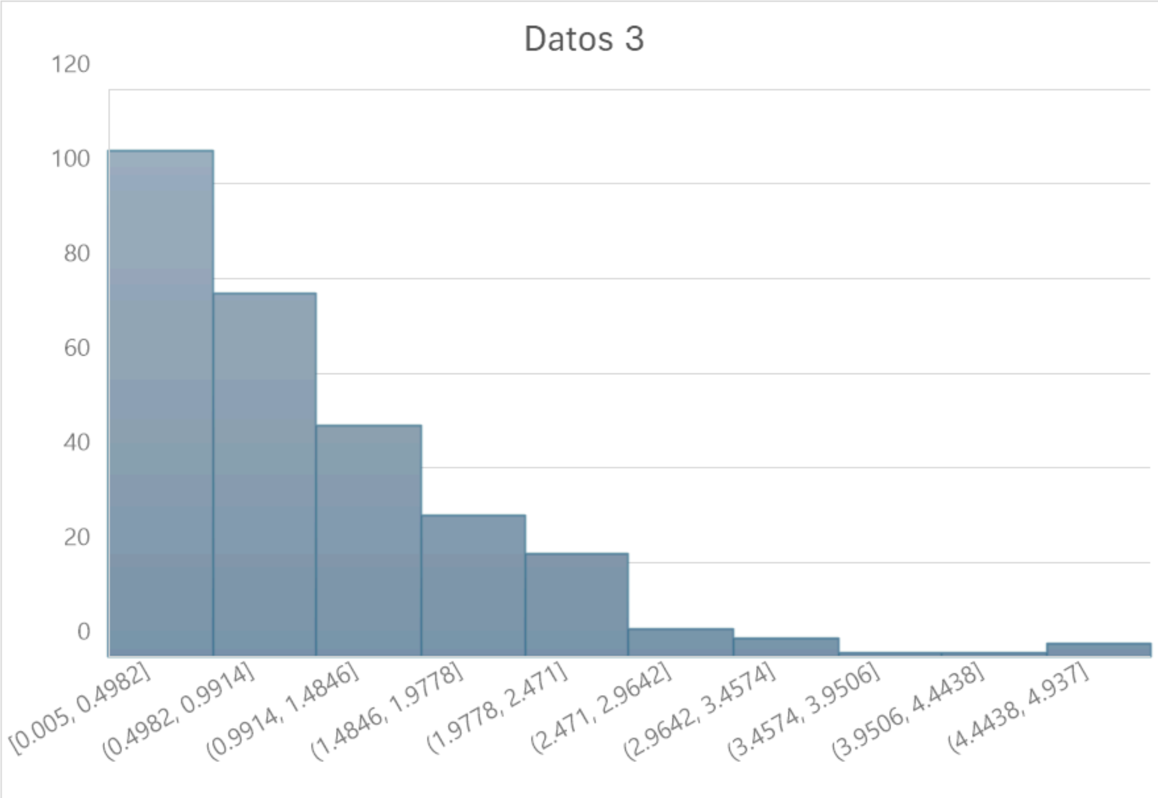
a.- Construye e interpreta un histograma. Utiliza la regla de Sturges para calcular el número apropiado de clases.

Para Datos1 k=10
Para Datos2 k=11
Para Datos3 k=10

Los histogramas muestran que en Datos1 y Datos3 la mayoría de los valores están agrupados en números bajos, por lo que se ve una clara acumulación de datos hacia la izquierda. Por otro lado, Datos2 tiene una forma más equilibrada, parecida a la de una campana. Se podría decir que la regla de Sturges funciona bien para Datos2 debido a su simetría, pero no parece funcionar bien para Datos1 y Datos3.

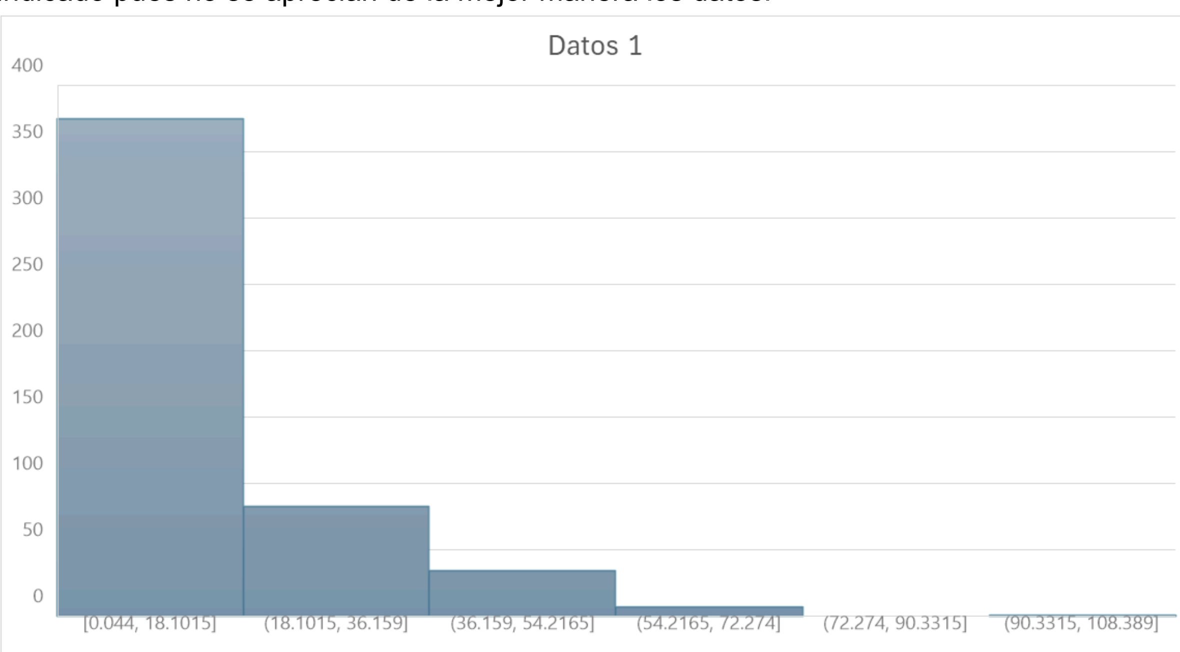






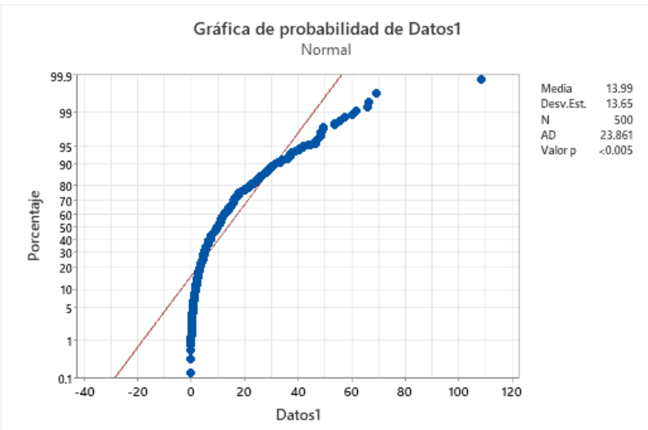
b.- Compara el número de clases con el obtenido con la regla de Scott.
Para Datos1 k= 6.003382065
Para Datos2 k= 0.377047342
Para Datos3 k= 0.441236289

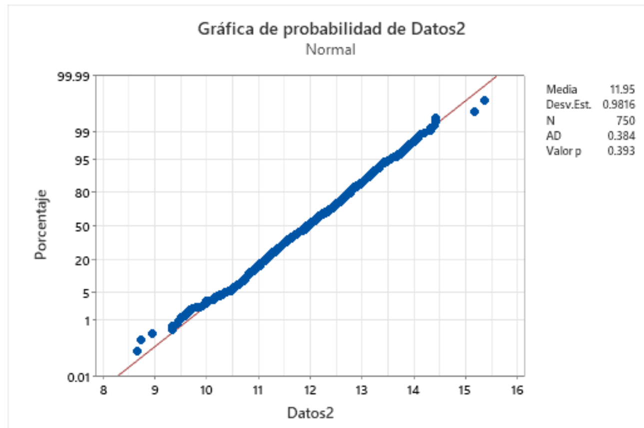
La regla de Scott no parece funcionar muy bien con el conjunto de datos, esto se podría deber al hecho de que para Datos2 y Datos3 la desviación estándar es muy pequeña lo que reduce demasiado el número de clases, por otro lado la desviación estándar en Datos 1 tiene un número un poco más grande lo cual no afecta tanto el número de clases pero aún así parece que ese número no es el indicado pues no se aprecian de la mejor manera los datos.

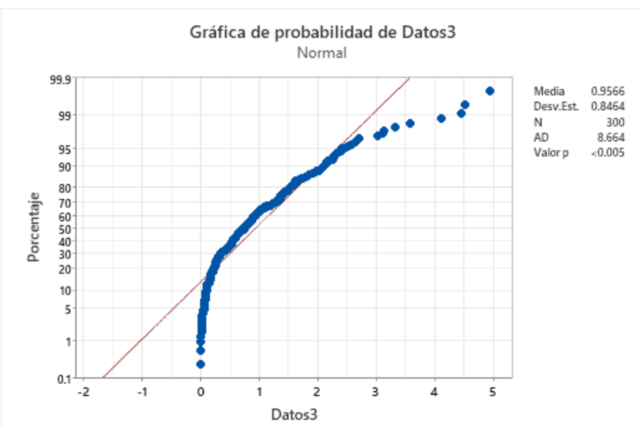


c.- Construye e interpreta un gráfico Q-Q para comprobar si los datos provienen de una distribución normal. Estima los parámetros utilizando la regresión de un gráfico probabilístico.

En un gráfico Q-Q, si los puntos siguen una línea recta cerca de la diagonal, sugiere que los cuantiles de los datos observados coinciden bien con los cuantiles esperados de una distribución normal. Como se puede observar en los gráficos Q-Q de abajo, tanto los puntos de Datos1 y Datos3 no siguen una línea recta pues los puntos se desvían a los extremos, pero en la gráfica de Datos2 podemos ver que los puntos siguen una línea recta recta a lo largo de toda la recta, por lo que podemos asumir que la gráfica de Datos2 SÍ pertenece o por lo menos se asemeja a una distribución normal.







Asimismo con el p-valor nos podemos dar cuenta si se trata de una distribución normal. Si el p-valor es p>0.005 no se rechaza la hipótesis nula, por lo que es muy probable que sea una distribución normal. Si el p-valor es p p<0.005 se rechaza la hipotesi nula lo que indica que probablemente NO sea una distribución normal.

Minitab & Excel

	Dato 1	Dato 2	Dato 3
P-Valor =	<0.05	0.393	<0.005
¿Distribución Normal?	✗	✓	✗

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?

Gracias al software de minitab podemos concluir que es más probable que Datos1 pertenezca a la Distribución Box-Cox, Datos2 a la Distribución Normal o Box-Cox y por último Datos3 a Gamma de 3 parámetros.

Distribución	AD	P	LRT P
Normal	23.861	<0.005	
Transformación Box-Cox	0.202	0.878	
Lognormal	5.489	<0.005	
Lognormal de 3 parámetros	1.098	*	0.000
Exponencial	0.588	0.387	
Exponencial de 2 parámetros	0.552	>0.250	0.284
Weibull	0.291	>0.250	
Weibull de 3 parámetros	0.303	>0.500	0.133
Valor extremo más pequeño	53.690	<0.010	
Valor extremo por máximos	8.461	<0.010	
Gamma	0.252	>0.250	
Gamma de 3 parámetros	0.251	*	0.735
Logística	15.578	<0.005	
Loglogística	2.923	<0.005	
Loglogística de 3 parámetros	1.961	*	0.006
Transformación de Johnson	0.278	0.649	

Distribución	AD	P	LRT P
Normal	0.384	0.393	
Transformación Box-Cox	0.384	0.393	
Lognormal	0.986	0.013	
Lognormal de 3 parámetros	0.382	*	0.000
Exponencial	291.447	<0.003	
Exponencial de 2 parámetros	169.727	<0.010	0.000
Weibull	5.340	<0.010	
Weibull de 3 parámetros	1.012	<0.005	0.000
Valor extremo más pequeño	9.321	<0.010	
Valor extremo por máximos	11.033	<0.010	
Gamma	0.668	0.085	
Gamma de 3 parámetros	0.600	*	0.157
Logística	0.961	0.007	
Loglogística	1.162	<0.005	
Loglogística de 3 parámetros	1.537	*	0.037

Distribución	AD	P	LRT P
Normal	8.664	<0.005	
Transformación Box-Cox	0.667	0.081	
Lognormal	5.440	<0.005	
Lognormal de 3 parámetros	2.050	*	0.000
Exponencial	1.188	0.072	
Exponencial de 2 parámetros	1.154	0.078	0.285
Weibull	0.575	0.151	
Weibull de 3 parámetros	0.663	0.089	0.254
Valor extremo más pequeño	22.691	<0.010	
Valor extremo por máximos	3.367	<0.010	
Gamma	0.703	0.082	
Gamma de 3 parámetros	0.695	*	1.000
Logística	6.155	<0.005	
Loglogística	3.780	<0.005	
Loglogística de 3 parámetros	2.789	*	0.015