

Se analiza una serie de datos que corresponden a los registros diarios de la temperatura promedio en la localidad de Apazulco del estado de Jalisco, México desde el año 1993 al 2014, luego de filtrar algunos intervalos de tiempo en los que no se tenía medición se redujo la serie a 7284 datos a los cuales se les realizará el análisis estadístico presentado en el siguiente informe.

En la Figura 1 se presenta la distribución de todos los datos de temperatura en el tiempo.

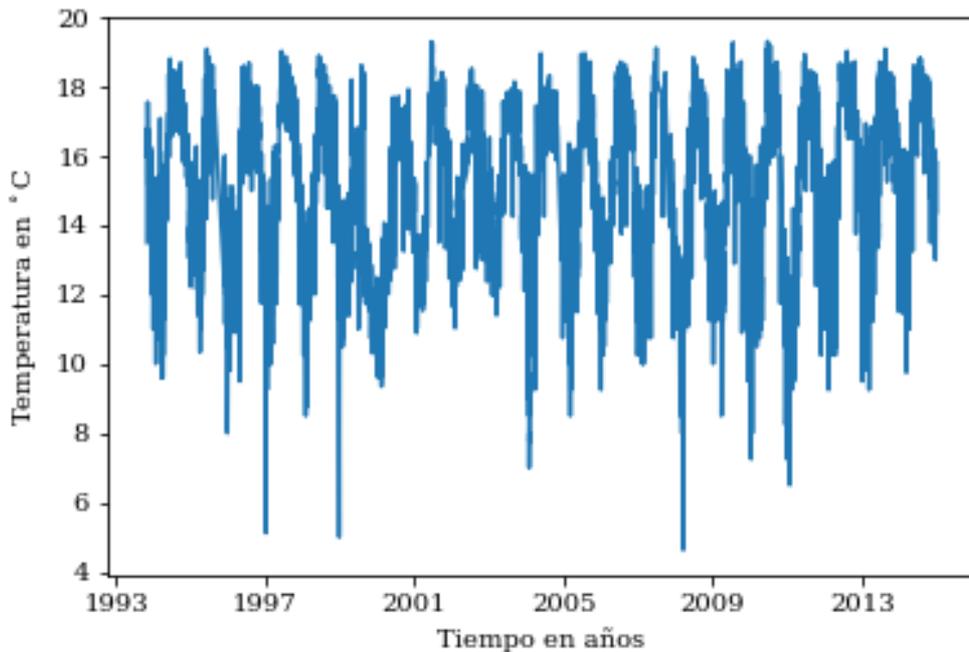


Figura 1 Cambio de temperatura en el tiempo.

Para la realización del histograma de probabilidades, teniendo en cuenta que las temperaturas varían entre 4.5°C y 19.5°C, se decidió dividir los datos en 30 intervalos los cuales representan una variación de 0.5°C entre ellos. Se observa un histograma bimodal sesgado a la izquierda, con un primer pico en 13.44°C y una probabilidad de 6.32% y un segundo pico en 17.6°C y una probabilidad de 12.05%

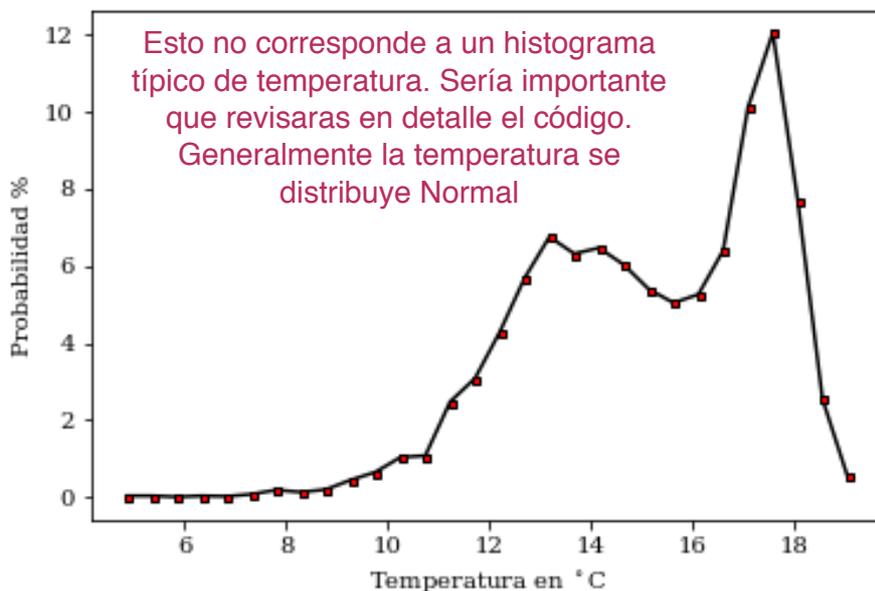


Figura 2 Histograma de probabilidades.

INDICES DE LOCALIZACIÓN

Se calcularon los índices de localización y algunos percentiles en la serie general de datos teniendo los siguientes resultados:

Media = 15.15°C P25 = 13.3°C

Mediana = 15.37°C P75 = 17.28°C

Moda = 17.6°C

La trimedia es una medida que reduce la sensibilidad de la medida a los valores extremos de la serie.

$$\text{Trimedia} = \frac{P25 + (2 \times P50) + P75}{4} = 15.33^\circ\text{C}$$

El valor de la trimedia es muy similar a el valor de la media aritmética, es decir, los valores extremos de la serie no influyen de manera considerable en el cálculo de la media.

Como se puede observar, el valor de la moda calculada concuerda con el valor pico del histograma de frecuencias y el valor del percentil 25 con el pico secundario.

En la Figura 3 se muestra la distribución de los percentiles notándose la mayor concentración de datos entre el percentil 25 y el 75 representada en el área sombreada de color rojo.

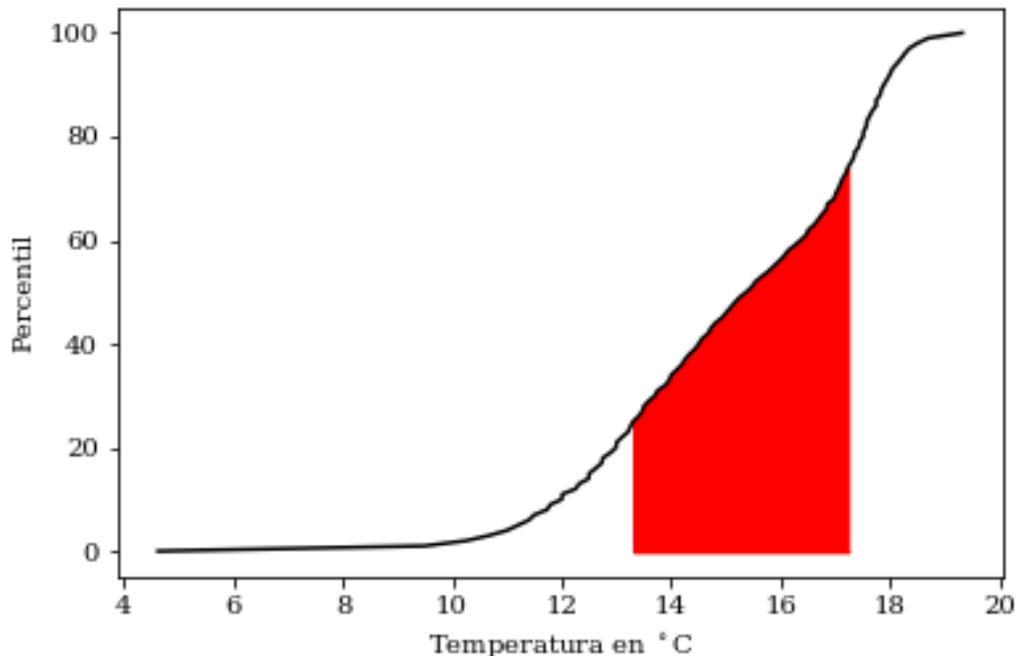


Figura 3 Probabilidad acumulada.

INDICES DE LOCALIZACIÓN MOVILES

Luego de analizar los índices de localización para toda la serie, se calcularon nuevamente estos índices, pero en esta ocasión considerando una rango o ventana móvil de longitud = 7 años, esto con el fin de identificar si los índices de localización son estacionarios o varían en el tiempo o si varían, identificar si lo hacen con alguna tendencia.

Como primer paso se analizó la distribución de los histogramas en el tiempo (ver Figura 4), donde se pudo notar un constante color amarillo en la probabilidad cerca al 12%, esta probabilidad corresponde al pico del histograma de la serie original, por lo que podemos afirmar que los histogramas son estacionarios.

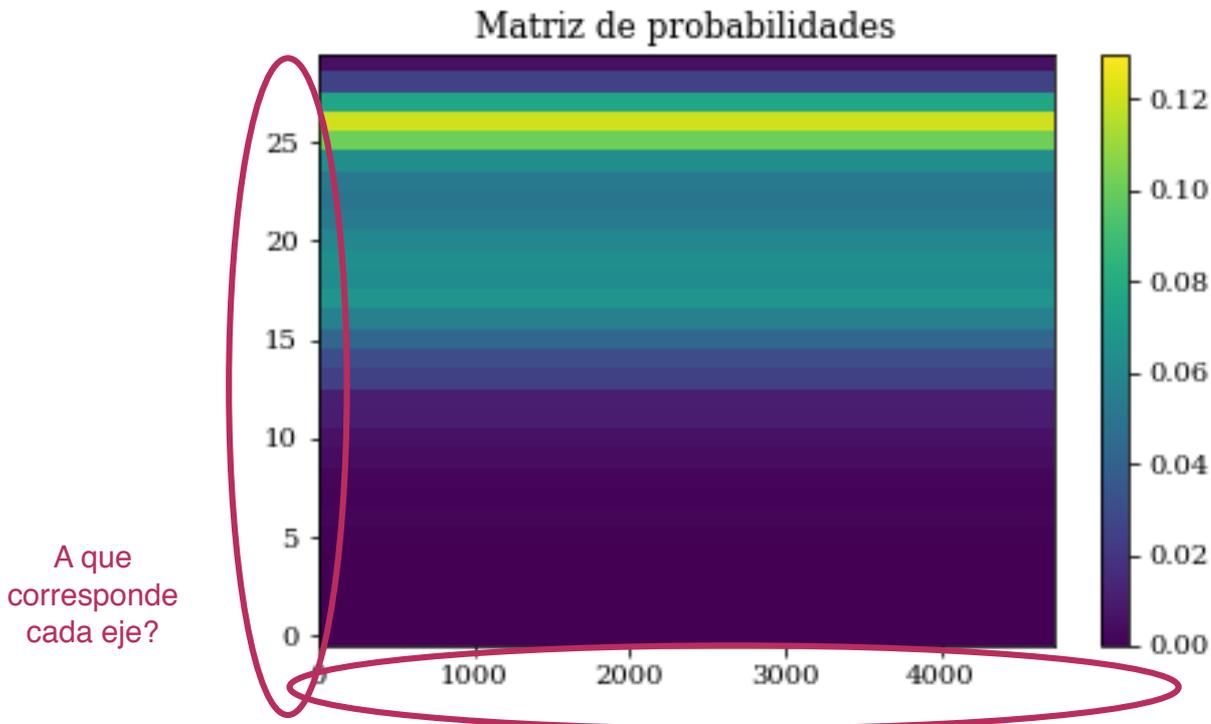


Figura 4 Matriz de probabilidades

Sabiendo que las probabilidades se mantienen constantes en el tiempo podemos analizar los índices de localización.

Usando la misma ventana de tiempo se analizó la media y la mediana obteniéndose los siguientes resultados:

Valores asociados a la ventana móvil			
	Mínimo	Máximo	Promedio
Media_s	14.84	15.38	15.07
Mediana_s	14.8	15.75x	15.22

Adicionalmente se presenta la **Figura 5** donde se puede apreciar de manera gráfica el comportamiento de estos dos índices largo de la ventana.

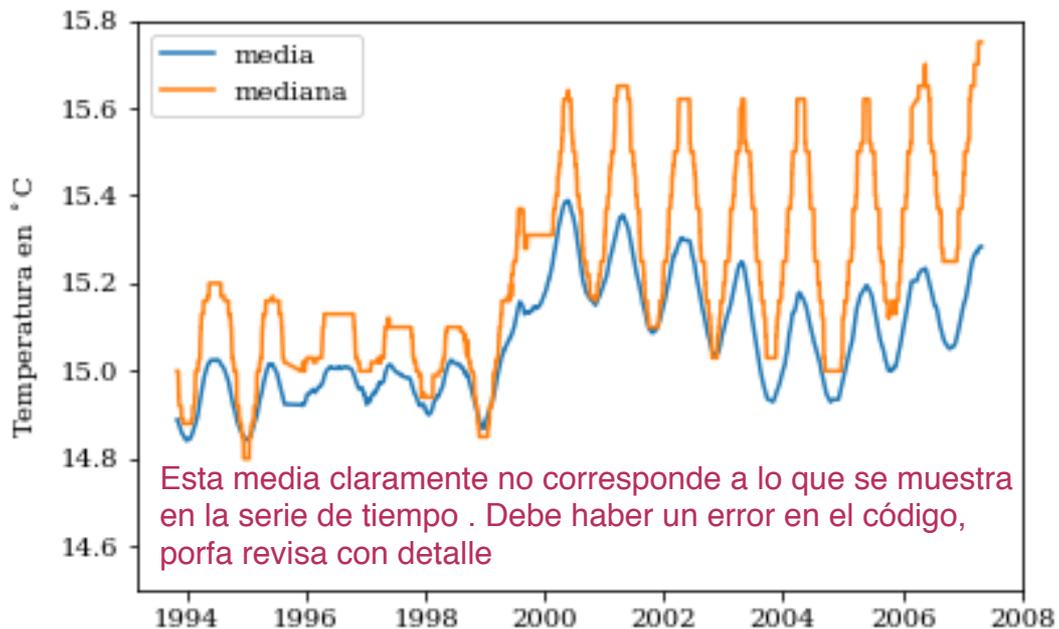


Figura 5 Índices de localización móviles.

Nótese que los índices no son estacionarios, varían en el tiempo, más adelante se buscara identificar si existe alguna tendencia en la manera como estos valores cambian, inicialmente se observa una tendencia positiva ya a pesar de mostrar un comportamiento periódico la mediana y la media aumentan en el tiempo

INDICES DE DISPERSIÓN

De igual manera se calcularon los índices de dispersión con el fin de analizar la variabilidad de los datos de la serie, se encontraron los siguientes resultados:

Rango = Máximo valor de la serie – Mínimo valor de la serie

$$\text{Rango} = 19.31 - 4.63 = 14.68^{\circ}\text{C}$$

Sin embargo, este valor no es muy representativo, si observamos la **Figura 3**, es claro que el valor de 4.63°C es un valor aislado con una probabilidad muy cercana a 0.

Desviación estándar (S)

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

$$S = 2.34^{\circ}\text{C}$$

Este valor nos indica que a pesar de la gran variabilidad que se observa en la **Figura 1** en realidad los datos se están alejando en promedio 2.34°C de la media.

Cabe resaltar también, que el valor de la desviación estándar es apenas mayor a la diferencia entre el percentil 25 y el percentil 50 = 2.07°C y entre el percentil 50 y el percentil 75 que es = 1.91°C

Rango intercuartil (IQR) : Se define como la diferencia entre el valor del tercer y primer cuartil

$$q_{75} - q_{25} = 3.98^{\circ}\text{C}$$

Indice de asimetría de Yule Kendall

Está definido por la expresión:

$$Y_{YK} = \frac{(P75 - p50) - (p50 - p25)}{IQR} = -0.04$$

El valor negativo este índice me indica una distribución sesgada a la izquierda como se había observado visualmente en la Figura 2.

Que datos estén sesgados a la izquierda me indica que la temperatura en esta localidad es un dato que se mantiene constante entre 15 y 17°C y que los valores extremos inferiores son poco probables, algo coherente si tenemos en cuenta la cercanía del lugar con el trópico.

De igual manera de analizaron los índices de dispersión en la misma ventana móvil que se analizó para los índices de localización, obteniéndose los siguientes resultados:



Figura 6 Desviación Estandar(movil)

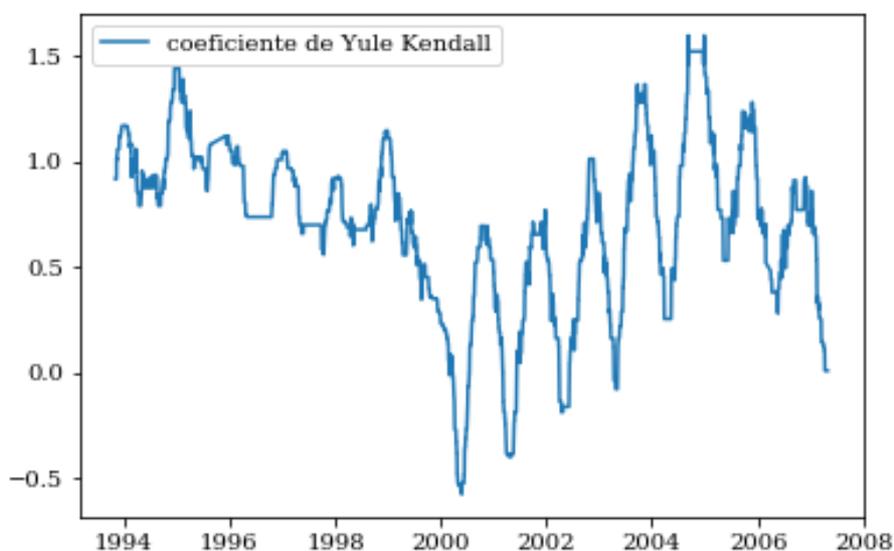


Figura 7 Indice de Yule Kendall móvil.

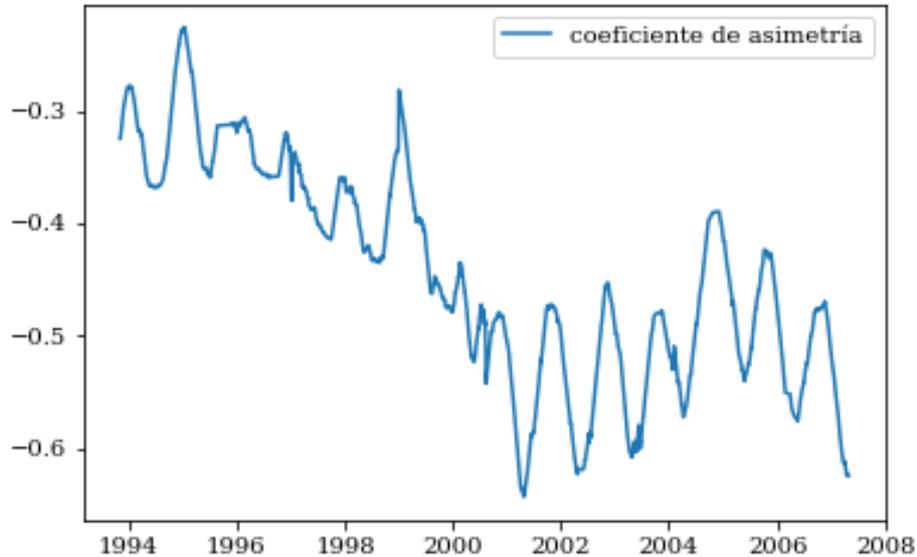


Figura 8 Coeficiente de asimetría móvil.

Observando la **Figura 6**, la **Figura 7** y la **Figura 8** Se puede afirmar que los índices de dispersión no son estacionarios.

Tendencia en la serie/Percentiles

Luego de analizar diferentes índices de localización, dispersión y simetría y observar que estos no son estacionarios en el tiempo, buscaremos determinar si existe o no tendencia en la serie de datos y sus percentiles.

Para el análisis de la tendencia se utilizó la prueba de tendencias de Mann Kendall con la siguiente hipótesis

Hipotesis inicial H_0 = Los datos son aleatorios

La prueba de tendencias de Mann Kendall se basa en el cálculo del estadístico S que se define como:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(X_j - X_i)$$

Donde se cumple:

$$\text{Si } (X_j - X_i) > 0 \quad \text{sgn}(X_j - X_i) = 1$$

$$\text{Si } (X_j - X_i) = 0 \quad \text{sgn}(X_j - X_i) = 0$$

$$\text{Si } (X_j - X_i) < 0 \quad \text{sgn}(X_j - X_i) = -1$$

Siendo X_j y X_i datos consecutivos de la serie en estudio

Teniendo el valor de S se calcula el estadístico Z definido como:

$$\text{Si } S > 0; Z_{MK} = \frac{S - 1}{\sqrt{V(s)}}$$

$$\text{Si } S < 0; Z_{MK} = \frac{S + 1}{\sqrt{V(s)}}$$

$$Si S = 0 \quad Z_{Mk} = 0$$

$$V(S) = \frac{n(n-1)(2n+5) - \sum_t t(t-1)(2t+5)}{18}$$

El Z encontrado se debe comparar con el Z de una distribución normal (ver **Figura 9**) y así poder dar respuesta a las diferentes hipótesis.

La hipótesis puede ser rechazada si:

$$|Z_{Mk}| \geq Z_{1-\alpha/2}$$

Siendo alfa el valor de la probabilidad de incertidumbre con la que se va a rechazar la hipótesis, para este ejercicio se considerara una incertidumbre de 1% con el cual se obtuvieron los siguientes resultados:

$$|Z_{Mk}| \geq Z_{1-\alpha/2}$$

$$|6.34| \geq Z_{1-0.01/2}$$

$$|6.34| \geq Z_{0.995}$$

$$|6.34| \geq 2.58(\text{ver Figura 9})$$

Es claro que la serie no tiene tendencia, obtuviste este resultado dado los errores que hay en el cálculo de la media móvil. Nuevamente te pido que revises por favor

La hipótesis se rechaza con una certeza del 99.5%, lo que significa que la serie de datos no es aleatoria, es decir, tiene tendencia.

Teniendo en cuenta el análisis anterior, se continúa evaluando la tendencia de la serie ahora con el fin de conocer si esta es negativa o positiva.

Para que la tendencia sea negativa se debe cumplir:

$$Z_{Mk} \leq -(Z_{1-\alpha})$$

$$6.34 \leq -Z_{1-0.01}$$

$$6.34 \leq -Z_{0.99}$$

$$6.34 \leq -2.33(\text{ver Figura 9})$$

No se cumple, es decir, hay un 99% de certeza en rechazar la hipótesis de que la serie tiene una tendencia negativa.

Para tendencia positiva se debe cumplir:

$$Z_{Mk} \geq (Z_{1-\alpha})$$

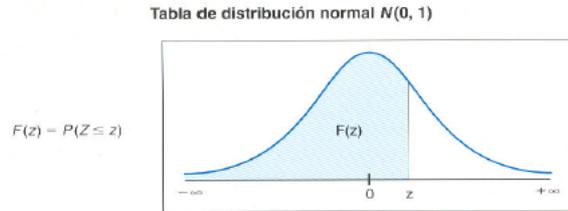
$$6.34 \geq Z_{1-0.01}$$

$$6.34 \geq Z_{0.99}$$

$$6.34 \geq 2.33(\text{ver Figura 9})$$

Se cumple la igualdad por lo que se puede decir, con un 99% de certeza que la serie tiene una tendencia positiva.

Se le aplico esta prueba al diferentes vectores como los percentiles, la media móvil, la mediana móvil y se obtuvieron $Z_{Mk} = 14.8, 38.05$ y 43.46 respectivamente, lo cual confirma la tendencia positiva de la serie.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Figura 9 Tabla de distribución normal (0,1)

Tu tarea carece de análisis físico de la serie. No concluyes mucho en relación a la naturaleza de los datos, en ese sentido es complejo tener pistas acerca de tus resultados. Recuerda que esta materia es de análisis de datos y no solo de programación.

3.4