PRUEBA DE SHAPIRO- WILK

La prueba de Shapiro-Wilk es una prueba estadística utilizada para determinar si una muestra de datos sigue una distribución normal. Fue desarrollada por Samuel Shapiro y Martin Wilk en 1965 y se basa en la covarianza entre los valores ordenados de la muestra y los valores esperados bajo la hipótesis de normalidad.

El procedimiento de la prueba de Shapiro-Wilk se puede resumir en los siguientes pasos:

1. Hipótesis nula (H0): La hipótesis nula establece que la muestra de datos proviene de una distribución normal.
2. Hipótesis alternativa (H1): La hipótesis alternativa afirma que la muestra de datos no sigue una distribución normal.
3. Ordenar los datos: Los datos se ordenan de menor a mayor.
4. Cálculo de los coeficientes de covarianza: Se calculan los coeficientes de covarianza que estiman la covarianza entre los valores ordenados y los valores esperados bajo la hipótesis de normalidad.
5. Cálculo del estadístico de prueba: Se calcula el estadístico de prueba W, que se basa en los coeficientes de covarianza. Cuanto más cercano esté el valor de W a 1, mayor evidencia de que los datos siguen una distribución normal.
6. Valor crítico: Se compara el valor de W con un valor crítico, obtenido de tablas de referencia o mediante software estadístico. Si el valor de W es menor que el valor crítico, se rechaza la hipótesis nula y se concluye que los datos no siguen una distribución normal.

Es importante tener en cuenta que la prueba de Shapiro-Wilk es sensible al tamaño de la muestra. Para muestras pequeñas (generalmente $n \leq 50$), la prueba tiene más poder para detectar desviaciones de la normalidad. Además, la prueba puede ser afectada por la presencia de valores atípicos en los datos, lo que puede influir en la interpretación de los resultados.

La prueba de Shapiro-Wilk es ampliamente utilizada en la práctica estadística como una herramienta inicial para evaluar la normalidad de los datos. Sin embargo, es importante recordar que ninguna prueba de normalidad puede demostrar que los datos provienen de una distribución normal exacta. Además, la interpretación adecuada de los resultados debe considerar el contexto y los supuestos del análisis estadístico realizado.

En resumen, la prueba de Shapiro-Wilk es una prueba estadística utilizada para evaluar si una muestra de datos sigue una distribución normal. Proporciona evidencia a favor o en contra de la hipótesis nula de normalidad basada en la covarianza entre los valores ordenados y los valores esperados bajo la hipótesis de normalidad. Sin embargo, se debe tener en cuenta el tamaño de la muestra y la presencia de valores atípicos al interpretar los resultados.

El estadístico de la prueba es Prueba de Shapiro-Wilk:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

donde
$x_{(i)}$ (con el subíndice i entre paréntesis) es el número que ocupa la i-énesima posición en la muestra (con la muestra ordenada de menor a mayor);

$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$ es la muestra muestral;

las variables $a_1, \ldots, a_n = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$

donde

$m = ((m_1, \ldots, m_n)^T$

Interpretación de la prueba de Shapiro-Wilk
La hipótesis nula se rechazará si W es demasiado pequeño. El valor de W puede oscilar entre 0 y 1.

Siendo la hipótesis nula que la población está distribuida normalmente, si el p-valor es menor a alfa (nivel de significancia) entonces la hipótesis nula es rechazada (se concluye que los datos no vienen de una distribución normal). Si el p-valor es mayor a alfa, se concluye que no se puede rechazar dicha hipótesis.

La normalidad se verifica confrontando dos estimadores alternativos de la varianza σ^2:

- un estimador no paramétrico al numerador, y
- un estimador paramétrico (varianza muestral), al denominador.

Referencias:

Box, G. E., Hunter, W. G., & Hunter, J. S. (2005). Estadística para investigadores: Diseño, innovación y descubrimiento (2a ed.). Wiley.

Cochran, W. G. (1977). Técnicas de muestreo (3a ed.). Ediciones Díaz de Santos.

Devore, J. L., & Peck, R. (2018). Estadística para ingenieros (9a ed.). Cengage Learning.
Johnson, R. A., & Wichern, D. W. (2007). Applied multivariate statistical analysis (6th ed.). Pearson.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introducción al análisis de regresión lineal (7a ed.). Cengage Learning.

Montgomery, D. C., Runger, G. C., & Hubele, N. F. (2019). Estadística aplicada y probabilidad para ingenieros (7a ed.). Cengage Learning.

Rice, J. A. (2006). Mathematical statistics and data analysis (3rd ed.). Cengage Learning.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2018). Probabilidad y estadística para ingeniería y ciencias (9a ed.). Pearson.