APÉNDICE A

PRUEBA DE NORMALIDAD DE SHAPIRO-WILK

Cuando los datos resultan de un proceso de medición o conteo (variables cuantitativas), es necesario comprobar antes de cualquier análisis estadístico, si la variable aleatoria estudiada sigue el modelo normal de distribución de probabilidades. En el caso que los datos se ajustan a una distribución normal se les puede aplicar los métodos estadísticos denominados paramétricos. Así se denominan aquellos métodos cuya aplicación depende del cumplimiento de algunos supuestos sobre las propiedades de la población de datos. Estas propiedades se denominan parámetros, de allí el nombre de métodos paramétricos. Por ejemplo, el uso de muchos de los métodos de inferencia estadística más comunes (intervalos de confianza, pruebas de hipótesis, correlación, regresión y análisis de varianzas) requieren que las muestras de datos provengan de poblaciones de valores que se distribuyen normalmente.

La experiencia indica que para la mayoría de los datos biológicos una desviación moderada de la normalidad no es de importancia y los análisis antes mencionados pueden efectuarse sin mayores problemas. Sin embargo, hay ocasiones en las cuales no es posible obviar el incumplimiento de tal supuesto. En estos casos se tienen dos alternativas. Una vía es recurrir a la estadística no paramétrica y usar métodos equivalentes. Este tipo de estadística no requiere de suposiciones previas acerca de la distribución de los datos. Sin embargo, cuando se cumplen el supuesto de normalidad, aunque sea en forma aproximada, los métodos paramétricos son mucho más potentes que las pruebas no paramétricas, por lo que a menudo se recurre al uso de alguna función matemática que transforme los datos de tal forma que los nuevos valores cumplan con el supuesto requerido. Una vez transformados los datos se comprueba si los nuevos valores se distribuyen normalmente. De modo que es muy importante poder contar con un método para comprobar la normalidad de un conjunto de datos originales o transformados.

Entre los numerosos métodos usados para probar la normalidad de un conjunto de datos, destaca la prueba de Shapiro-Wilk por ser una de la más sencilla y potentes. La única condición es que el tamaño de la muestra debe ser igual o menor a 50.

Prueba de Shapiro-Wilk

a. Hipótesis.

H_0 : La variable aleatoria no tiene una distribución normal

H_1 : La variable aleatoria tiene una distribución normal

b. Estadístico de prueba.

$$W_c = \frac{b^2}{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2}$$

El término $b = \sum_{i=1}^{k} a_i \left[X_{(n-i+1)} - X_i \right]$, siendo a_i = el valor de un coeficiente que se encuentra tabulado para cada tamaño de muestra y la posición *i* de cada observación. El término $\left[X_{(n-i+1)} - X_i \right]$ = diferencias sucesivas que se obtienen al restar el primer valor al último valor, el segundo al penúltimo, el tercero al antepenúltimo y así hasta llegar a restar el último al primer valor. Por ejemplo si se tienen siete valores, la secuencia de diferencias es la siguiente:

Observación i	Valores X_i ordenados en orden ascendente	$\left[X_{(n-i+1)} - X_i \right]$
1	X_1	$X_7 - X_1$
2	X_2	$X_6 - X_2$
3	X_3	$X_5 - X_3$
4	X_4	$X_4 - X_4$
5	X_5	$X_3 - X_5$
6	X_6	$X_2 - X_6$
7	X_7	$X_1 - X_7$

c. Zona de aceptación para H_0:

La zona de aceptación para H_0 está formada por todos los valores del estadístico de prueba W_c menores al valor esperado o tabulado $W_{(1-\alpha;n)}$.

$$ZA = \left\{ W \,/\, W_{calculado} \leq W_{(1-\alpha;n)} \right\}$$

Ejemplo.

Una surtidora automática fue utilizada para llenar envases con 16 ml de un medicamento y mediante un muestreo aleatorio se seleccionaron 8 frascos y se les midió el volumen envasado, encontrándose los resultados siguientes:

16,0; 15,9; 15,97; 16,04; 16,05; 15,98; 15,96; 16,02.

Se quiere saber si la variable volumen servido se distribuye normalmente.

Procedimiento para la prueba de normalidad de Wilk-Shapiro

a. Hipótesis.

H_0: La variable aleatoria volumen servido no sigue una distribución normal

H_1: La variable aleatoria volumen servido sigue la distribución normal

b. Se calcula el estadístico de prueba:

En primer lugar se construye una tabla con las columnas siguientes:

Columna 1: Se enumeran todos los valores de la variable estudiada $(i = 1,2,3, \ldots . ,n)$

Columna 2: Se ordenan los valores de la variable en forma ascendente X_i.

Columna 3: Se ordenan los valores de la variable en forma descendente $\left[X_{(n-i+1)}\right]$

Columna 4: Se obtiene la diferencia $\left[X_{(n-i+1)} - X_i\right]$ = (Columna 3 – Columna 2)

Columna 5: Se obtienen los valores de a_i para n = 8 en la Tabla 10.

Columna 6: Se calcula el producto $a_i\left[X_{(n-i+1)} - X_i\right]$ = (Columna 5 x Columna 6)

Columna 7: Se calcula el término $\left(X_i - \bar{X}\right)^2$.

1	**2**	**3**	**4**	**5**	**6**	**7**
i	X_i	$X_{(n-i+1)}$	$X_{(n-i+1)} - X_i$	a_i	$a_i\left[X_{(n-i+1)} - X_i\right]$	$\left(X_i - \bar{X}\right)^2$
1	15,90	16,05	0,15	0,6052	0,090780	0,0081
2	15,96	16,04	0,08	0,3164	0,025312	0,0009
3	15,97	16,02	0,05	0,1743	0,008715	0,0004
4	15,98	16,00	0,02	0,0561	0,001122	0,0001
5	16,00	15,98	-0,02	0	0	0,0001
6	16,02	15,97	-0,05	0	0	0,0009
7	16,04	15,96	-0,08	0	0	0,0025
8	16,05	15,9	-0,15	0	0	0,0036
Total					0,0159	0,0166

Para obtener el valor del estadístico de prueba W_c se divide el cuadrado del total de la columna 5: $b^2 = \left(\sum_{i=1}^{n} a_i\left[X_{(n-i+1)} - X_i\right]\right)^2$ entre el total de la columna 6: $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$

$$W = \frac{b^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{(0,0159)^2}{0,0166} = 0,9553$$

c. Zona de aceptación para H_0.

En la Tabla 2 se obtiene el valor de $W_{(0.95;8)}$ y se define la zona de aceptación de H_0.

$$ZA = \{W / W_{calculado} \leq 0.818\}$$

Como el valor de $W_c = 0,9553$ es mayor al valor esperado $W_{(0,95;8)} = 0,818$ se rechaza H_0, por lo tanto se concluye que se tiene una confianza del 95% que la variable volumen servido se distribuye normalmente.