



ORIGINAL

Entre bambalinas. Bondad de ajuste a una normal.

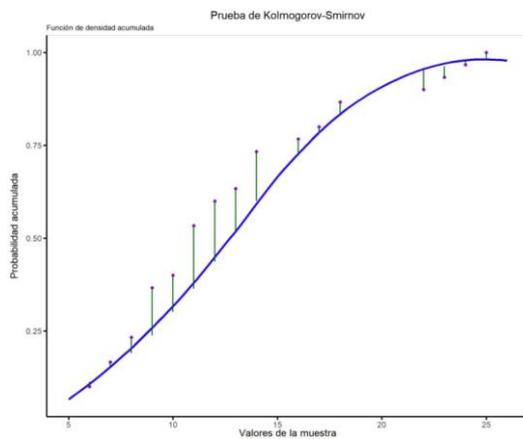
Molina Arias M

H.U. Getafe, España

Resumen

Las pruebas de Shapiro-Wilk y de Kolmogorov-Smirnov son los test de contraste más empleados para comprobar la bondad de ajuste a una distribución normal de nuestros datos. Se describe su realización paso a paso y su equivalencia en métodos gráficos como el gráfico de cuantiles teóricos y el gráfico de la función de densidad acumulada.

Introducción



Las pruebas de Shapiro-Wilk y de Kolmogorov-Smirnov son los test de contraste más empleados para comprobar la bondad de ajuste a una distribución normal de nuestros datos. Se describe su realización paso a paso y su equivalencia en métodos gráficos como el gráfico de cuantiles teóricos y el gráfico de la función de densidad acumulada.

Imaginad que os encontráis en el abarrotado vestíbulo de un teatro antiguo. Los candelabros de cristal tintinean suavemente mientras una multitud expectante murmura sobre el espectáculo que está a punto de comenzar, deseando descubrir los

secretos que yacen detrás del telón de terciopelo.

Dentro, los focos se encienden y el telón se levanta lentamente. En el escenario, un viejo mago se prepara para demostrar algunos de los actos más enigmáticos de su repertorio, con los que asombrará al entregado público, incapaz de comprender la magia que da lugar a la ilusión.

Sin embargo, todos sabemos que, detrás de la espectacularidad, seguramente se esconde la simplicidad del truco que la permite, que no llegamos ni a imaginar. Todo esto viene a cuento porque, aunque os parezca difícil de creer, algo parecido ocurre con las pruebas de bondad de ajuste a una normal que aplicamos a nuestros datos.

Al igual que en los actos de escapismo más complejos, donde las cadenas y las cerraduras juegan un papel esencial en el arte del ilusionismo, estas pruebas esconden bajo su complejidad matemática una simpleza y elegancia que puede ser visualizada y entendida a través de simples gráficos.

Así que no os dejéis abrumar por este enredo de cálculos y teorías. En esta entrada veremos cómo la complejidad de estas pruebas esconde una trama

simple y elegante. Seguid leyendo y descubriréis que la verdadera magia no está solo sobre el escenario, sino en el sutil arte de manejar los datos entre bambalinas, donde lo extraordinario se vuelve comprensible.

La importancia de ser normal

Son bien conocidas las características de la distribución normal y su papel central en la realización de nuestros contrastes de hipótesis. Esto es así porque las pruebas paramétricas de contraste de hipótesis suelen precisar que los datos se ajusten a una distribución normal.

Todos los métodos que tratan de verificar este supuesto tratan de analizar cuánto difiere la distribución de nuestros datos (los observados en nuestra muestra) respecto a lo que deberíamos esperar si los datos procediesen de una población en la que la variable siguiese una distribución normal con la misma media y desviación estándar que la observada en los datos de la muestra.

Para esta comprobación disponemos de tres estrategias posibles: los métodos basados en contrastes de hipótesis, los basados en representaciones gráficas y los llamados métodos analíticos. Hoy nos vamos a centrar en los primeros, las pruebas de bondad de ajuste a una normal, concretamente en las pruebas de Shapiro-Wilk y de Kolmogorov-Smirnov.

Veamos primero su cara menos amable, la que nos muestra toda su complejidad matemática.

Prueba de Shapiro-Wilk

Esta es la prueba de bondad de ajuste a una normal que suele emplearse cuando la muestra es pequeña, digamos que menor de 30 elementos.

Como cualquier otra prueba de contraste paramétrica, la prueba de Shapiro-Wilk se basa en el cálculo de un estadístico, llamado W , que sigue una determinada distribución de probabilidad. Una vez calculado W , podemos obtener la probabilidad de obtener por azar un valor mayor o igual al observado bajo el supuesto de la hipótesis nula, que dice que los datos siguen una normal.

Esta probabilidad, efectivamente, es nuestra p . Si $p < 0,05$ (o el valor que elijamos para establecer la significación estadística), rechazaremos la hipótesis nula y asumiremos que los datos no se ajustan a una distribución normal. Por el contrario, si $p > 0,05$, nos quedaremos con la hipótesis nula y no podremos negar que los datos sean normales (ya sabéis, no rechazar la hipótesis nula no es sinónimo de demostrarla, especialmente en pruebas poco potentes como esta).

La dificultad de la prueba de Shapiro-Wilk es que el estadístico W no sigue una distribución estándar simple como la normal o la t de Student. En cambio, la distribución de W es única para cada tamaño de muestra n en función de las propiedades de los cuantiles de la distribución normal estándar y de nuestra muestra de datos.

Esta distribución, particular para cada n , se determina mediante simulaciones y cálculos matemáticos complejos, por lo que a nadie se le ocurre hacer los cálculos a mano, sino recurrir a tablas con valores de W precalculados o, mejor, a paquetes estadísticos que hacen todo esto en un abrir y cerrar de ojos, sin quejas ni esfuerzo.

Un ejemplo, paso a paso

Para entender algo mejor todo el proceso, vamos a utilizar un conjunto de

datos inventado que vamos a generar con el programa R.

Vamos a suponer que tenemos una muestra de 30 registros con los siguientes valores:

23, 9, 25, 11, 11, 8, 6, 9, 22, 12, 13, 18, 9, 14, 12, 11, 7, 18, 17, 7, 11, 10, 24, 14, 9, 6, 16, 14, 6 y 8.

Este conjunto de datos tiene una media de 12,66 y una desviación estándar de 5,5. Además, me vais a permitir una pequeña trampa, porque he diseñado los datos sabiendo que no siguen una distribución normal, sino una exponencial, lo que resultará más didáctico para nuestras explicaciones.

Ahora vamos a calcular el estadístico W y ver si es significativo. Aquí es donde comienza la parte más ardua de trabajo con este estadístico tan antipático, cuya fórmula es la siguiente:

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

Vamos a explicar los componentes de la fórmula anterior porque es fundamental no confundir los términos, ya que $x_{(i)}$ y x_i son bien diferentes.

Lo primero que tenemos que hacer es ordenar nuestros datos de menor a mayor. Una vez hecho, calculamos los valores absolutos de las diferencias entre el primero (el valor más bajo) y el último (el valor más alto), entre el segundo y el penúltimo, y así, con todos los datos, con lo que obtenemos un conjunto, con la mitad de elementos que los datos originales, con las diferencias absolutas de las parejas. Esto representa $x_{(i)}$, cada una de estas diferencias.

Por su parte, x_i representa a cada elemento de la serie de datos original.

Una vez explicado esto, vemos que el numerador es una suma ponderada de estas diferencias entre parejas de datos, cada una de las cuales se multiplicará por a_i , que son una serie de coeficientes que se obtienen de los cuantiles de la distribución normal estándar y el número de observaciones (son los que ponderan las diferencias de las parejas ordenadas).

El denominador es más sencillo: la suma de los cuadrados de las diferencias de cada valor con la media de la distribución. Calculemos ahora el valor de W. Para ello iremos calculamos los términos necesarios, según se muestra en la tabla adjunta.

Datos ordenados	$x_{(i)}$	a_i	$x_{(i)} a_i$	$(\text{datos} - \text{media})^2$	Estadístico W	
6	25	19	0,4254	8,0826	44,4444	152,1111
6	24	18	0,2944	5,2992	44,4444	128,4444
6	23	17	0,2487	4,2279	44,4444	106,7778
7	22	15	0,2148	3,2220	32,1111	87,1111
7	18	11	0,1870	2,0570	32,1111	28,4444
8	18	10	0,1630	1,6300	21,7778	28,4444
8	17	9	0,1415	1,2735	21,7778	18,7778
9	16	7	0,1219	0,8533	13,4444	11,1111
9	14	5	0,1036	0,5180	13,4444	1,7778
9	14	5	0,0862	0,4310	13,4444	1,7778
9	14	5	0,0697	0,3485	13,4444	1,7778
10	13	3	0,0537	0,1611	7,1111	0,1111
11	12	1	0,0381	0,0381	2,7778	0,4444
11	12	1	0,0227	0,0227	2,7778	0,4444
11	11	0	0,0076	0,0000	2,7778	2,7778
Suma			28,1649	880,6667		

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{28,1649^2}{880,6667} = 0,901$$

Paso 1. Ordenar los datos y calcular las diferencias.

La primera columna de la tabla muestra los datos ordenados. He dividido la columna en dos y los datos van de menor a mayor en sentido descendente en la primera mitad y en sentido ascendente en la segunda. Lo hago así para facilitarme el cálculo de las diferencias, que están en la segunda columna.

Paso 2. Obtener los coeficientes.

El cálculo de los coeficientes tiene cierta complejidad, ya que involucra la matriz de covarianza de los cuantiles de la orden normal y su inversa, según una

fórmula todavía más antipática que la de W y que no os voy a mostrar aquí.

Por eso, lo habitual es recurrir a una tabla de valores precalculados o a un programa de estadística que sea capaz de obtenerlos y mostrarlos. Yo voy a tomar los datos de una de las [tablas que hay por Internet](#) y que os adjunto para que la veáis bien.

j	21	22	23	24	25	26	27	28	29	30
1	0'4643	0'4590	0'4542	0'4493	0'4450	0'4407	0'4366	0'4328	0'4291	0'4254
2	0'3185	0'3156	0'3126	0'3098	0'3069	0'3043	0'3018	0'2992	0'2968	0'2944
3	0'2578	0'2571	0'2563	0'2554	0'2543	0'2533	0'2522	0'2510	0'2499	0'2487
4	0'2119	0'2131	0'2139	0'2145	0'2148	0'2151	0'2152	0'2151	0'2150	0'2148
5	0'1736	0'1764	0'1787	0'1807	0'1822	0'1836	0'1848	0'1857	0'1864	0'1870
6	0'1399	0'1443	0'1480	0'1512	0'1539	0'1563	0'1584	0'1601	0'1616	0'1630
7	0'1092	0'1115	0'1201	0'1245	0'1283	0'1316	0'1346	0'1372	0'1395	0'1415
8	0'0804	0'0878	0'0941	0'0997	0'1046	0'1089	0'1128	0'1162	0'1192	0'1219
9	0'0530	0'0618	0'0696	0'0764	0'0823	0'0876	0'0923	0'0965	0'1002	0'1036
10	0'0263	0'0368	0'0459	0'0539	0'0610	0'0672	0'0728	0'0778	0'0822	0'0862
11	0'0000	0'0122	0'0228	0'0321	0'0403	0'0476	0'0540	0'0598	0'0650	0'0697
12	-	-	0'0000	0'0107	0'0200	0'0284	0'0358	0'0424	0'0483	0'0537
13	-	-	-	-	0'0000	0'0094	0'0178	0'0253	0'0320	0'0381
14	-	-	-	-	-	-	0'0000	0'0084	0'0159	0'0227
15	-	-	-	-	-	-	-	-	0'0000	0'0076

Selecciono los 15 coeficientes que da la tabla para cuando la muestra es de 30 elementos (15 diferencias de parejas) y lo paso a la tercera columna.

Paso 3. Calcular el estadístico W.

Primero calculamos los componentes que nos faltan de la fórmula de W, que son la suma de los productos de los coeficientes por las diferencias de las parejas, que podéis ver en las dos columnas.

Finalmente, en la columna de la derecha sustituimos en la fórmula los valores calculados para el numerador y el denominador y obtenemos W, que tiene un valor de 0,901.

Paso 4. Resolver el contraste de hipótesis.

Tenemos que decidir si este valor es estadísticamente significativo o si podemos haberlo obtenido por azar. Si lo hacemos a mano, lo más sencillo es buscar una tabla con los valores críticos de W para cada tamaño muestral y nivel de significación. Podéis ver una en la [misma página en la que encontramos los coeficientes](#).

Para un valor de $n = 30$ y $p = 0,05$, el valor crítico de $W = 0,927$. Al ser nuestro valor de W menor que el valor crítico, podemos rechazar la hipótesis nula y concluir que los datos no siguen una distribución normal. El valor de p será menor de 0,05, aunque con estos métodos manuales es mejor que nos conformemos sin el valor exacto, ya que resulta complejo calcularlo.

Prueba de Kolmogorov-Smirnov

Vamos ahora con la prueba de Kolmogorov-Smirnov, casi tan antipática y compleja como la anterior.

Una ventaja de esta prueba es que permite comprobar la bondad de ajuste a cualquier distribución de probabilidad, no solo a la normal. Sin embargo, tiene dos inconvenientes que contrarrestan esta ventaja.

Primero, necesita conocer la media y la desviación estándar de la población de la que procede la muestra, datos que solemos desconocer. Para intentar solventar esto, se aplica la corrección de Lilliefors para comparar la bondad de ajuste con la distribución normal, tabulando el estadístico con estimaciones muestrales de la media y la desviación estándar.

Pero esto no funciona muy bien, lo que causa su segundo inconveniente: es una prueba poco potente, por lo que tiene una gran tendencia a no poder rechazar la hipótesis nula que, al igual que en la prueba de Shapiro-Wilk, asume que los datos siguen una distribución normal.

Vamos a ver un ejemplo práctico de cómo se hace el contraste con esta prueba, utilizando los mismos datos que en el ejemplo anterior, como podéis ver en la tabla que iremos rellenando con los datos que calculemos.

Datos ordenados		OCDF		CDF		OCDF - CDF	
6	11	0,10	0,53	0,11	0,38	0,01	0,15
6	12	0,10	0,60	0,11	0,45	0,01	0,15
6	12	0,10	0,60	0,11	0,45	0,01	0,15
7	13	0,17	0,63	0,15	0,52	0,01	0,11
7	14	0,17	0,73	0,15	0,60	0,01	0,14
8	14	0,23	0,73	0,20	0,60	0,03	0,14
8	14	0,23	0,73	0,20	0,60	0,03	0,14
9	16	0,37	0,77	0,25	0,73	0,11	0,04
9	17	0,37	0,80	0,25	0,78	0,11	0,02
9	18	0,37	0,87	0,25	0,83	0,11	0,03
9	18	0,37	0,87	0,25	0,83	0,11	0,03
10	22	0,40	0,90	0,31	0,95	0,09	0,05
11	23	0,53	0,93	0,38	0,97	0,15	0,04
11	24	0,53	0,97	0,38	0,98	0,15	0,01
11	25	0,53	1,00	0,38	0,99	0,15	0,01

Otro ejemplo, paso a paso

Esta prueba calcula un estadístico llamado D , según la fórmula que os muestro a continuación:

$$D = \max_{1 \leq i \leq n} |F_n(x_i) - F_0(x_i)|$$

Veamos qué significan los componentes de la fórmula. x_i representa cada uno de los valores de la muestra, ordenados de menor a mayor. $F_n(x_i)$ es un estimador de la probabilidad de observar, en nuestra muestra, valores menores o iguales a x_i . Por último, $F_0(x_i)$ es la probabilidad de observar valores menores o iguales de x_i cuando la hipótesis nula es cierta lo que, en nuestro ejemplo, se cumplirá cuando los datos se distribuyan de forma normal.

De esta forma, D será igual a la mayor diferencia absoluta observada entre la frecuencia acumulada y la frecuencia teórica obtenida de la distribución de probabilidad que se especifica como hipótesis nula.

Ya solo queda hacer los cálculos.

Paso 1. Ordenar los datos.

Ya lo tenemos hecho del ejemplo anterior. Lo podéis ver en la primera columna de la tabla, aunque ahora todos siguen orden descendente de arriba a abajo.

Paso 2. Calcular la función de distribución acumulada empírica.

Esta es la frecuencia acumulada observada en nuestros datos (OCDF). Se muestra en la segunda columna.

Paso 3. Calcular la función acumulada teórica.

Es la frecuencia acumulada teórica (CDF) si los datos siguen una normal.

Como es lógico, este paso y el anterior pueden hacerse a mano, pero yo me he ayudado del programa R.

Paso 4. Calcular las diferencias absolutas.

Restamos las diferencias entre las frecuencias acumuladas y nos quedamos con el valor absoluto. Vemos que el valor máximo es de 0,15. Este es el valor del estadístico D .

Paso 5. Resolver el contraste.

El cálculo del valor crítico para D es complejo, así que volvemos a recurrir a los datos tabulados. Podéis ver un ejemplo en la [tabla 8 de este enlace](#). El valor de D para $\alpha = 0,05$ y $n = 30$ es de 0,24.

En este caso, rechazaremos la hipótesis nula si nuestro D es mayor que el valor crítico. Como nuestro valor de $D = 0,15$ es menor que el valor crítico, no podemos rechazar la hipótesis nula y asumimos que los datos siguen una distribución normal.

¡Un momento! ¿No os dije que había hecho una pequeña trampa y había generado una serie de datos que sigue una distribución exponencial? Acabáis de ver en directo el mayor inconveniente de estas pruebas de contraste: su baja potencia y la

“facilidad” que tienen para asumir la normalidad de los datos.

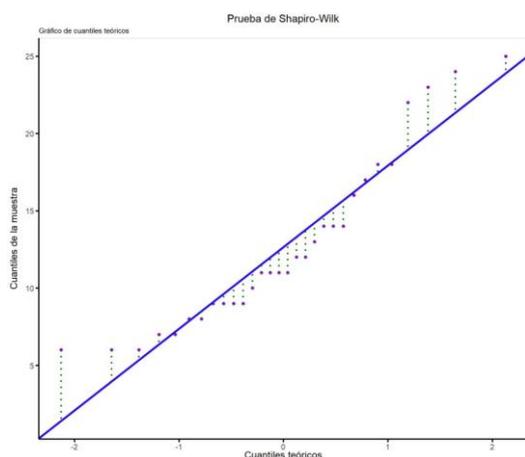
Por eso, siempre que no podamos rechazar la hipótesis nula, se aconseja complementar el contraste mediante algún medio gráfico como el histograma o el gráfico de cuantiles teóricos.

Entre bambalinas

Los pocos que hayáis seguido leyendo hasta aquí os preguntareis a cuento de qué venía la historia del principio sobre los magos, la espectacularidad de sus trucos y la simplicidad que suele haber tras ellos.

Pues resulta que estas dos pruebas, lo que hacen entre bambalinas es mucho más sencillo: recurren a métodos gráficos para hacer su magia.

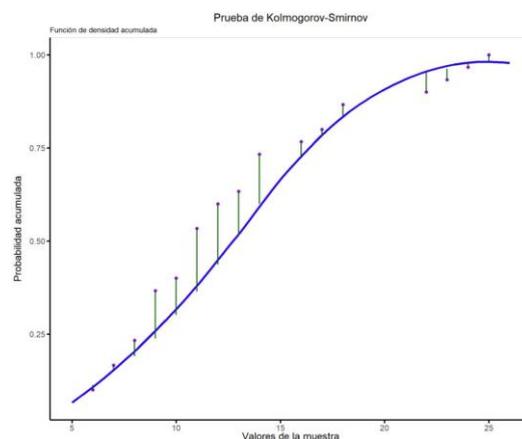
Detrás del telón, la prueba de Shapiro-Wilk imagina un gráfico de cuantiles teóricos y compara las distancias entre los datos reales y la recta que marca dónde deberían estar los datos si se distribuyesen según una normal. Podéis verlo en el primer gráfico.



Es lo mismo que hacemos nosotros para complementar el contraste numérico cuando no se rechaza la hipótesis nula, pero con la diferencia de que nosotros lo hacemos de forma subjetiva y el método de contraste hace un análisis

cuantitativo (aunque poco potente, como hemos visto).

La prueba de Kolmogorov-Smirnov hace algo parecido, pero recurre al gráfico de la función de probabilidad o densidad acumulada de los datos teóricos si estos siguiesen una normal. Aquí, compara las distancias entre los datos reales y la línea de probabilidad acumulada, tal como veis en el segundo gráfico.



Y esto es lo que se esconde detrás de todo el galimatías de números que hemos visto en los dos ejemplos anteriores. Me pregunto si detrás de las actuaciones de David Copperfield se esconderán trucos de esta belleza. A diferencia de hacer levitar un avión o desaparecer a alguien, aquí todo lo que “desaparece” son nuestras dudas sobre la distribución de los datos (siempre con un margen de error, claro), ¡y no se necesita magia, solo un poco de estadística!

Nos vamos

Y con esto creo que ha llegado la hora de airearnos un poco las cabezas, que deben estar echando humo.

Hemos visto cómo realizar de forma manual dos pruebas de bondad de ajuste a una normal, la prueba de Shapiro-Wilk y la prueba de Kolmogorov-Smirnov. Sobra decir que a nadie se le tiene que ocurrir hacer estas pruebas a

mano. El ejemplo está hecho para comprender su funcionamiento, pero siempre emplearemos un programa estadístico para llevarlas a cabo.

Yo he usado R incluso para hacer los cálculos manuales. Los amantes de los deportes de riesgo podéis descargar el *script* en este [enlace](#) y repetir el experimento por vuestra cuenta.

Para acabar, podemos decir que la prueba de Shapiro-Wilk puede ser la más idónea cuando los tamaños muestrales son pequeños (<30), mientras que preferiremos la de Kolmogorov-Smirnov cuando las muestras sean más grandes (a pesar de ser la menos potente de las dos), siempre aplicando la corrección de Lilliefors si desconocemos la media y la desviación estándar en la población.

Por último, no debemos olvidar complementar el contraste con otro método cuando no podamos rechazar la

hipótesis de normalidad, ya que, como hemos visto, son pruebas poco potentes. Lo más sencillo es recurrir a métodos gráficos, aunque hay otros métodos analíticos más sofisticados para aquellos que gusten de complicarse la vida. Pero esa es otra historia...

Bibliografía

- Romero Saldaña M. Pruebas de bondad de ajuste a una distribución normal. Rev Enferm Trabajo.2016;6:105-14. ([HTML](#))
- Classical tests. En: Crawley MJ, ed. The R book, 2ª ed. John Wiley & Sons, Ltd. West Sussex, UK, 2016: 344-87. ([HTML](#))

Correspondencia al autor

Manuel Molina Arias
mma1961@gmail.com
Servicio de Gastroenterología Hospital Infantil
Universitario La Paz, de Madrid.

Aceptado para el blog en julio de 2024