¿Y si los datos no siguen una distribución normal?...

Bondad de ajuste a una normal.
Transformaciones.
Pruebas no paramétricas

Preparado por Luis M. Molinero (Alce Ingeniería)
CorreoE: **bioestadistica@alceingenieria.net**

[**Artículo en formato PDF**]

www.seh−lelha.org/stat1.htm

Julio 2003

Cuando se analizan datos medidos por una variable cuantitativa continua, las pruebas estadísticas de estimación y contraste frecuentemente empleadas se basan en suponer que se ha obtenido una muestra aleatoria de una distribución de probabilidad de tipo normal o de Gauss. Pero en muchas ocasiones esta suposición no resulta válida, y en otras la sospecha de que no sea adecuada no resulta fácil de comprobar, por tratarse de muestras pequeñas. En estos casos disponemos de dos posibles mecanismos: los datos se pueden **transformar** de tal manera que sigan una distribución normal, o bien se puede acudir a pruebas estadísticas que no se basan en ninguna suposición en cuanto a la distribución de probabilidad a partir de la que fueron obtenidos los datos, y por ello se denominan **pruebas no paramétricas** (*distribution free*), mientras que las pruebas que suponen una distribución de probabilidad determinada para los datos se denominan pruebas paramétricas.

Dentro de las pruebas paramétricas, las más habituales se basan en la **distribución de probabilidad normal**, y al estimar los parámetros del modelo se supone que los datos constituyen una muestra aleatoria de esa distribución, por lo que la elección del estimador y el cálculo de la precisión de la estimación, elementos básicos para construir intervalos de confianza y contrastar hipótesis, dependen del modelo probabilístico supuesto.

Cuando un procedimiento estadístico es poco sensible a alteraciones en el modelo probabilístico supuesto, es decir que los resultados obtenidos son aproximadamente válidos cuando éste varía, se dice que es un procedimiento **robusto**.

Las inferencias en cuanto a las medias son en general robustas, por lo que si el tamaño de muestra es grande, los intervalos de confianza y contrastes basados en la *t de Student* son aproximadamente válidos, con independencia de la verdadera distribución de probabilidad de los datos; pero si ésta distribución no es normal, los resultados de la estimación serán poco precisos.

Procedimientos para verificar el ajuste a una distribución de probabilidad

Existen diferentes pruebas para verificar el ajuste de nuestros datos a una distribución de probabilidad. Las dos más utilizadas son el contraste χ^2 **de Pearson**, y la **prueba de Kolmogorov−Smirnov**.

Contraste χ^2 de Pearson

La idea del contraste de Pearson es muy sencilla: se agrupan los datos en *k* clases ($k \geq 5$), como si fuéramos a construir un histograma, cubriendo todo el rango posible de valores, siendo deseable disponer, aproximadamente, del mismo número de datos en cada clase y al menos de tres datos en cada una.

Llamamos *Oi* al número de datos observado en la clase *i*. Mediante el modelo de probabilidad que se desea verificar se calcula la probabilidad *Pi* asignada a cada clase, y por lo tanto, para una muestra de *n* datos, la frecuencia esperada según ese modelo de probabilidad es *Ei=n.Pi*.

Se calcula entonces el siguiente índice de discrepancia entre las frecuencias observadas y las que era previsible encontrar si el modelo fuera el adecuado:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

que se distribuye aproximadamente como una χ^2 si el modelo es correcto.

Si el modelo se especifica de forma completa con las probabilidades *Pi*, conocidas antes de tomar los datos, el número de grados de libertad es $k-1$. Pero si se han estimado *r* parámetros del modelo a partir de los datos, entonces los grados de libertad son $k-r-1$.

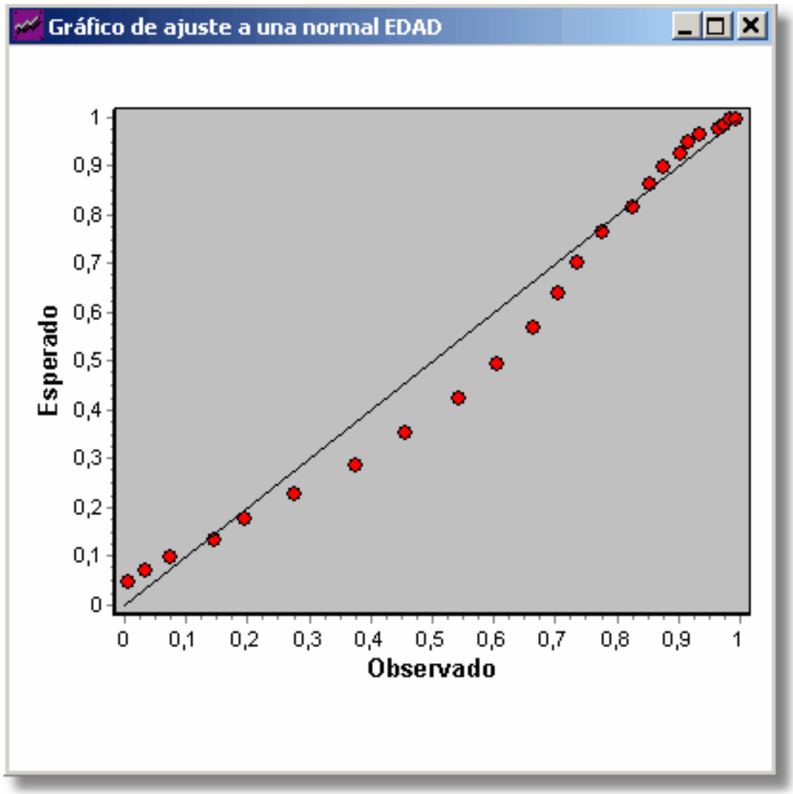
- **Prueba de Kolmogorov−Smirnov**

Este contraste, que es válido únicamente para variables continuas, compara la función de distribución (probabilidad acumulada) teórica con la observada, y calcula un valor de discrepancia, representado habitualmente como *D*, que corresponde a la discrepancia máxima en valor absoluto entre la distribución observada y la distribución teórica, proporcionando asimismo un valor de probabilidad *P*, que corresponde, si estamos verificando un ajuste a la distribución normal, a la probabilidad de obtener una distribución que discrepe tanto como la observada si verdaderamente se hubiera obtenido una muestra aleatoria, de tamaño *n*, de una distribución normal. Si esa probabilidad es grande no habrá por tanto razones estadísticas para suponer que nuestros datos no proceden de una distribución, mientras que si es muy pequeña, no será aceptable suponer ese modelo probabilístico para los datos.

- **Prueba de Shapiro−Wilks**

Aunque esta prueba es menos conocida es la que **se recomienda** para contrastar el ajuste de nuestros datos a una distribución normal, sobre todo cuando la muestra es pequeña ($n<30$).

Mide el ajuste de la muestra a una recta, al dibujarla en papel probabilístico normal. Este tipo de representación también lo proporcionan algunos programas de estadística, de tal manera que nos permite además apreciar el ajuste o desajuste de forma visual:



En escala probabilística normal se representa en el eje horizontal, para cada valor observado en nuestros datos, la función de distribución o probabilidad acumulada observada, y en el eje vertical la prevista por el modelo de distribución normal. Si el ajuste es bueno, los puntos se deben distribuir aproximadamente según una recta a 45º. En la imagen vemos que en este ejemplo existe cierta discrepancia.

En cualquier caso siempre es adecuado efectuar una representación gráfica de tipo histograma de los datos, y comparar el valor de la media y la mediana, así como evaluar el coeficiente de asimetría y apuntamiento, además de llevar a cabo una representación en escala probabilística de la distribución de probabilidad esperada versus observada, como la de la figura.

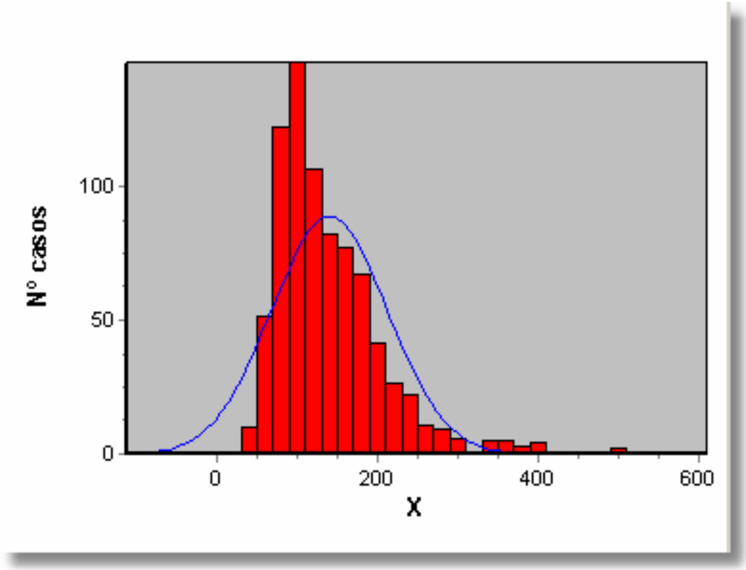
Posibles soluciones cuando se rechaza la hipótesis de normalidad

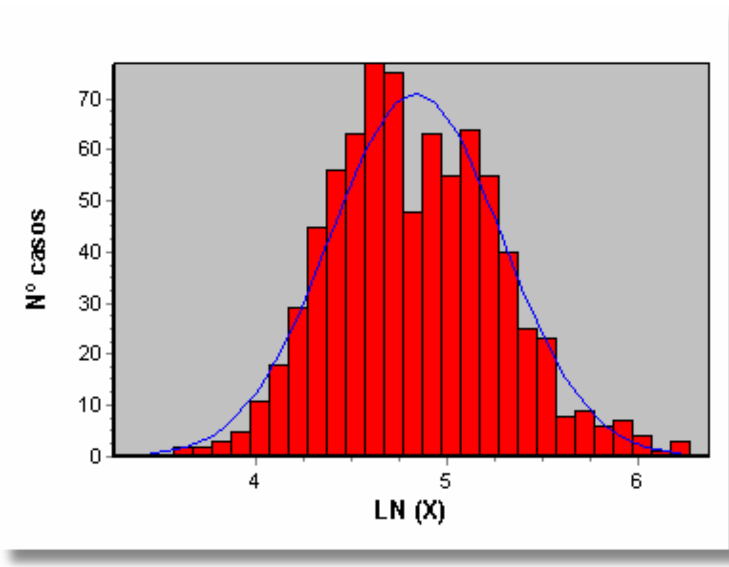
Si rechazamos o dudamos de la normalidad de nuestros datos, existen varias soluciones posibles:

- Si la distribución es más apuntada que la normal (mayor parte de los valores agrupados en torno de la media y colas más largas en los extremos), se debe investigar la presencia de heterogeneidad en los datos y de posibles valores atípicos o errores en los datos. La solución puede ser emplear pruebas no paramétricas.
- Si la distribución es unimodal y asimétrica, la solución más simple y efectiva suele ser utilizar una transformación para convertir los datos en normales.
- Cuando la distribución no es unimodal hay que investigar la presencia de heterogeneidad, ya que en estos casos la utilización de transformaciones no es adecuada y los métodos no paramétricos pueden también no serlo.
- Una alternativa muy interesante a los métodos paramétricos y a las pruebas no paramétricas clásicas, la constituye la metodología de **estimación autosuficiente** ya esbozada en otro artículo de esta serie.

Transformaciones para conseguir datos normales

La utilización de transformaciones para lograr que los datos se ajusten a una distribución normal es en muchas ocasiones la solución más natural, ya que existen gran cantidad de parámetros biológicos que tienen una distribución asimétrica como la de la figura de la izquierda, y que se convierten en aproximadamente simétricas al transformarlas mediante el **logaritmo**.





Tenemos problemas con la transformación logarítmica ln(x) si la variable puede tomar el valor 0, por lo que en esos casos, o incluso si existen valores muy pequeños, será adecuado emplear la transformación ln(x+1). Cuando la desviación típica de los datos es proporcional a la media o cuando el efecto de los factores es multiplicativo, en lugar de aditivo, está indicado el uso de la transformación logarítmica.

Otra transformación posible es $\sqrt{x}$, que es aplicable cuando las varianzas son proporcionales a la media, lo que ocurre a menudo cuando los datos provienen de una distribución de Poisson (recuentos).

Otra transformación habitualmente empleada es 1/x, que también precisa que sumemos una cantidad a cada valor si existen ceros.

Estas tres transformaciones comprimen los valores altos de los datos y expanden los bajos, en sentido creciente en el siguiente orden: $\sqrt{x}$ (la que menos), ln x, 1/x.

Si la concentración de datos está, a diferencia de la figura anterior, en el lado de la derecha y la cola en la izquierda, se puede utilizar la transformación x^2, que comprime la escala para valores pequeños y la expande para valores altos.

Cuando los datos son proporciones o porcentajes de una distribución binomial, las diferencias con una distribución normal son más acusadas para valores pequeños o grandes de las proporciones, utilizándose entonces transformaciones basadas en $arcsen\sqrt{p}$.

En todos los casos para los cálculos estadísticos basados en la teoría normal, se utilizarán los valores transformados, pero después para la presentación de los resultados se efectuará la transformación inversa para presentarlos en su escala de medida natural.

Más abajo se proporcionan algunos enlaces sobre el tema de las transformaciones, de fácil lectura.

Pruebas no paramétricas

Se denominan pruebas no paramétricas aquellas que no presuponen una distribución de probabilidad para los datos, por ello se conocen también como de distribución libre (*distribution free*). En la mayor parte de ellas los resultados estadísticos se derivan únicamente a partir de procedimientos de ordenación y recuento, por lo que su base lógica es de fácil comprensión. Cuando trabajamos con muestras pequeñas ($n \leq 10$) en las que se desconoce si es válido suponer la normalidad de los datos, conviene utilizar pruebas no paramétricas, al menos para corroborar los resultados obtenidos a partir de la utilización de la teoría basada en la normal.

En estos casos se emplea como parámetro de centralización la **mediana**, que es aquel punto para el que el valor de X está el 50% de las veces por debajo y el 50% por encima.

Vamos a comentar la filosofía de alguna de las pruebas no paramétricas y en los enlaces se puede aumentar esta información.

- **Prueba de Wilcoxon de los rangos con signo**

Esta prueba nos permite comparar nuestros datos con una mediana teórica (por ejemplo un valor publicado en un artículo).

Llamemos M_0 a la mediana frente a la que vamos a contrastar nuestros datos, y sea $X1, X2 .. Xn$ los valores observados. Se calcula las diferencias $X1-M_0$, $X2-M_0$, ..., $Xn-M_0$. Si la hipótesis nula fuera cierta estas diferencias se distribuirían de forma simétrica en torno a cero.

Para efectuar esta prueba se calculan las diferencias en valor absoluto $|Xi-M_0|$ y se ordenan de menor a mayor, asignándoles su rango (número de orden). Si hubiera dos o más diferencias con igual valor (empates), se les asigna el rango medio (es decir que si tenemos un empate en las posiciones 2 y 3 se les asigna el valor 2.5 a ambas). Ahora calculamos $R+$ la suma de todos los rangos de las diferencias positivas, aquellas en las que Xi es mayor que M_0 y $R-$ la suma de todos los rangos correspondientes a las diferencias negativas. Si la hipótesis nula es cierta ambos estadísticos deberán ser parecidos, mientras que si nuestros datos tienen a ser más altos que la mediana M_0, se reflejará en un valor mayor de $R+$, y al contrario si son más bajos. Se trata de contrastar si la menor de las sumas de rangos es excesivamente pequeña para ser atribuida al azar, o, lo que es equivalente, si la mayor de las dos sumas de rangos es excesivamente grande.

- **Prueba de Wilcoxon para contrastar datos pareados**

El mismo razonamiento lo podemos aplicar cuando tenemos una muestra de parejas de valores, por ejemplo antes y después del tratamiento, que podemos denominar $(X1,Y1), (X2,Y2), \ldots ,(Xn,Yn)$. De la misma forma, ahora calcularemos las diferencias $X1-Y1, X2-Y2, \ldots , Xn-Yn$ y las ordenaremos en valor absoluto, asignándoles el rango correspondiente. Calculamos $R+$ la suma de rangos positivos (cuando Xi es mayor que Yi), y la suma de rangos negativos $R-$. Ahora la hipótesis nula es que esas diferencias proceden de una distribución simétrica en torno a cero y si fuera cierta los valores de $R+$ y $R-$ serán parecidos.

- **Prueba de Mann−Whitney para muestras independientes**

Si tenemos dos series de valores de una variable continua obtenidas en dos muestras independientes: $X1, X2, \ldots , Xn, Y1, Y2, \ldots , Ym$, procederemos a ordenar conjuntamente todos los valores en sentido creciente, asignándoles su rango, corrigiendo con el rango medio los empates. Calculamos luego la suma de rangos para las observaciones de la primera muestra Sx, y la suma de rangos de la segunda muestra Sy. Si los valores de la población de la que se extrajo la muestra aleatoria de X se localizan por debajo de los valores de Y, entonces la muestra de X tendrá probablemente rangos más bajos, lo que se reflejará en un valor menor de Sx del teóricamente probable. Si la menor de las sumas de rangos es excesivamente baja, muy improbable en el caso de que fuera cierta la hipótesis nula, ésta será rechazada.

Existen más pruebas no paramétricas de entre las que a continuación mencionamos las más habituales, remitiendo al lector interesado a cualquier libro básico de bioestadística:

- Prueba de Kruskal−Wallis para comparar K muestras
- Prueba de Friedman para comparar K muestras pareadas (bloques)
- Coeficiente de correlación de Spearman para rangos
- Prueba de rachas de Wald−Wolfowitz

Enlaces de interés

- The normal distribution Douglas G Altman & J Martin Bland BMJ 1995;310:298 (4 Febrero)
- Transforming data J Martin Bland & Douglas G Altman BMJ 1996;312:770 (23 March)
- Transformations, means, and confidence intervals J Martin Bland & Douglas G Altman BMJ 1996;312:1079 (27 April)
- The use of transformation when comparing two means J Martin Bland & Douglas G Altman BMJ 1996;312:1153 (4 May)
- Algunas direcciones que permiten efectuar cálculos de pruebas no paramétricas:
  - Prueba de los signos
  - Contraste para la mediana de dos muestras independientes
  - Prueba de Wilcoxon para datos pareados
  - Prueba de Mann−Whitney para comparar dos muestras independientes
    Otro enlace para esta prueba
  - Prueba de rachas de Wald−Wolfowitz
  - Prueba de Kruskall−Wallis
  - Prueba de Friedman
  - Coeficiente de correlación de Spearman

Bibliografía seleccionada

- J.S. Milton, J.O. Tsokos. **Estadística para biología y ciencias de la salud**. Madrid: Interamericana−McGraw Hill; 1989
