

# Engañando a Gauss

No cabe duda de que Gauss y su distribución con forma de campana son la base para la realización de gran parte de las pruebas de contraste de hipótesis e inferencia de datos en estadística. Por eso, a nadie le llama la atención que muchas pruebas solo puedan realizarse cuando la variable que se estudia sigue una [distribución normal](#).

Por ejemplo, si queremos comparar las medias de dos muestras, éstas tienen que ser independientes, seguir una distribución normal y tener una varianza similar (homocedasticidad). Lo mismo ocurre para muchas otras comparaciones, estudios de correlación, etc.

Cuando tenemos la mala suerte de que nuestra muestra no sigue una distribución normal debemos recurrir a las pruebas de contraste no paramétricas. Estas pruebas son igual de serias y rigurosas que las paramétricas, pero tienen el inconveniente de que son mucho más conservadoras, en el sentido de que cuesta más alcanzar el nivel de significación estadística necesario para poder rechazar la hipótesis nula. Podría darse el caso de que no obtengamos significación estadística con la prueba no paramétrica mientras que, si pudiésemos aplicarla, si podríamos obtenerla con la paramétrica.

Para evitar que pueda pasarnos esto, a alguien se le debió ocurrir que podemos transformar los datos de tal forma que los nuevos datos transformados sí sigan la distribución normal. Esto, que parece un truco sucio, es perfectamente lícito, siempre en cuanto tengamos en cuenta que luego tendremos que hacer la transformación inversa para interpretar correctamente los resultados.

Hay diversos métodos de transformación, aunque quizás el más utilizado es el de la [transformación logarítmica](#).

Pensemos un momento en los logaritmos decimales (base 10). En la escala logarítmica hay la misma distancia entre 1 y 10 que entre 10 y 100 y que entre 100 y 1000. ¿Qué quiere decir esto?. Pues que si transformamos cada variable en su logaritmo, los valores entre 1 y 10 se expandirán, mientras que los más altos se comprimirán. Por eso la transformación logarítmica es útil para transformar distribuciones con sesgo positivo (con cola más larga hacia la derecha): la parte izquierda se expandirá, mientras que la derecha se comprimirá, favoreciendo que la curva resultante se ajuste mejor a una normal. Solo comentar que, aunque hemos puesto el ejemplo con el logaritmo decimal, que se entiende mejor, en estadística suele hacerse esta transformación con logaritmos neperianos o naturales, que tienen por base el número e, que vale, aproximadamente, 2,7182812.

Esta transformación logarítmica solo vale para números mayores que cero, aunque si tenemos una distribución con valores negativos podríamos sumar una constante a cada valor para que fuese mayor que cero antes de calcular su logaritmo. Cuando la nueva curva se ajusta a la campana se dice que sigue una distribución lognormal.

En ocasiones, si la distribución está muy sesgada, puede hacerse la transformación recíproca ( $1/x$ ), más potente y que produce un efecto similar a la logarítmica. Otra tercera posibilidad, menos potente que la logarítmica, es transformar calculando la raíz cuadrada de cada valor.

Cuando el sesgo de la distribución es negativo (cola más larga hacia la izquierda) nos interesará lo contrario: comprimir la cola de la izquierda y extender la de la derecha. Si lo pensamos, esto puede hacerse elevando cada valor al cuadrado o al cubo. Los productos resultantes de los valores pequeños estarán menos alejados que los resultantes de valores grandes, con lo que la distribución se parecerá más a una normal.

Así que miramos nuestra distribución, hacemos la transformación que nos parezca más oportuna y comprobamos si ya es normal. En ese caso, hacemos la prueba paramétrica para obtener el nivel de significación. Finalmente, deshacemos la transformación para poder interpretar correctamente los resultados, aunque en este punto puede haber alguna dificultad.

Si hemos aplicado una transformación logarítmica y hemos obtenido una media, su antilogaritmo es la media geométrica y no la aritmética. Si se trata de una diferencia de medias, lo que obtenemos es la razón de medias geométricas.

Con los intervalos de confianza no hay mayor problema. Podemos transformar, obtenerlos y deshacer la transformación. La que se muestra intratable con todo este tipo de transformaciones y antitransformaciones es la desviación estándar, que no tiene sentido transformar en reverso porque sus unidades pierden totalmente el sentido de la realidad.

Las transformaciones  $1/x$  y raíz cuadrada permiten recuperar los valores de las medias y de los intervalos de confianza sin problemas, pero tampoco pueden hacer nada con las desviaciones estándar.

Comentar, para terminar, que hay otras dos situaciones en que puede ser útil transformar los datos. Una es cuando las varianzas de las muestras son diferentes (no hay homocedasticidad). En estos casos puede aplicarse la logarítmica (si la varianza aumenta en proporción con la media), el cuadrado (si aumenta en proporción con el cuadrado de la media) o la raíz cuadrada (si lo hace en proporción con la raíz cuadrada de la media).

La otra situación es cuando queremos forzar que la relación entre las dos variables sea lineal, como cuando queremos utilizar modelos de

regresión lineal. Claro que estos casos habría que tener algunas otras consideraciones de cómo afectan las transformaciones a los coeficientes de regresión. Pero esa es otra historia...

---