



Gráficos y parámetros de posición, dispersión y forma de estadística descriptiva

Ejemplos de aplicación

Apellidos, nombre	Prats Montalbán, José Manuel (joprmon@eio.upv.es)
Departamento	Estadística, Investigación Operativa Aplicadas y Calidad
Centro	ETS Ingenieros Industriales



1 Resumen de las ideas clave

El presente artículo docente pretende mostrar al alumno de Ingeniería Química cómo utilizar de manera conjunta los parámetros de posición, dispersión y forma junto con los gráficos comúnmente utilizados dentro de cualquier análisis descriptivo a través de diferentes ejemplos de datos reales. Para ello, se toma como base el análisis de variables continuas.

2 Introducción

En la actualidad, las cantidades ingentes de datos relacionados con diferentes variables de calidad (e.g. octanaje final de un lote de gasolina, pureza y distribución de una mezcla farmacéutica, tono de una pieza cerámica, porcentaje de impurezas a la salida de un proceso de reciclaje de PET, densidad final de un polímero, etc.), así como variables propias generadas durante el propio proceso de producción (e.g. temperatura, presión, tipo y cantidad de materia prima utilizada, proveedor, tensiones e intensidades, etc.) hacen necesaria de la utilización de gráficos y parámetros que ayuden a tener una idea de qué tipo de distribución siguen los mismos; así como detectar posibles anomalías en los datos.

Ante esta situación, el ingeniero se puede plantear qué tipo de parámetros y/o gráficos utilizar con el fin de sacar el mayor rendimiento posible a los datos recogidos. El presente artículo docente presenta ejemplos en los que la aplicación conjunta de todos ellos (siempre que sea posible) genera un valor añadido a la hora de ganar conocimiento acerca de las variables (de tipo continuo) analizadas.

Los ejemplos de aplicación se presentan sobre el formato de *Matlab*.

3 Objetivos

El valor añadido del presente objeto de aprendizaje se encuentra en la aplicación conjunta de los gráficos y parámetros descriptivos a cualquier tipo de variable continua, así como en su explicación correspondiente. De esta manera, se plantean como objetivos los siguientes:

- Hacer reflexionar al alumno acerca de la necesidad de analizar los datos antes de cualquier decisión y/o utilización posterior de los mismos (e.g. para construir un modelo de modelización, un gráfico de control estadístico de calidad, etc.)
- Utilizar cada uno de los parámetros y gráficos de manera conjunta, determinando las posibles distribuciones que pueden seguir; así como la posible existencia de datos anómalos, censurados, asimetrías, etc.

4 Ejemplos de aplicación

4.1 Estatura de los alumnos

El primer ejemplo presenta el análisis de una serie de 219 datos correspondientes a la altura de unos jóvenes. En primer lugar, los valores estandarizados de asimetría (0.08) y curtosis (-0.33) se encuentran dentro del intervalo [-2,2].

Asimismo, el diagrama Box & Whisker (Figura 1 a) presenta un bigote algo mayor por la derecha, si bien la parte izquierda de la caja es más ancha que la derecha. Dado que la mayor variabilidad de esta muestra es en un caso por la derecha y en el otro por la izquierda, ambos criterios (caja y principalmente bigotes) se contradicen a la hora de establecer una cierta asimetría. Todo ello, unido a la ausencia de datos anómalos, no hace sospechar nada desde el punto de vista de este gráfico.

Por su parte, el Histograma (Figura 1 b), el cual hemos realizado dado que el tamaño de muestra es mayor a 40-50 datos, presenta una distribución de los mismos que se asemeja a una campana.

Finalmente, el Papel Probabilístico Normal (Figura 1 c) presenta los datos de la muestra alineados aproximadamente a lo largo de una recta.

Por todo ello, parece que los datos se pueden aproximar a una distribución Normal.

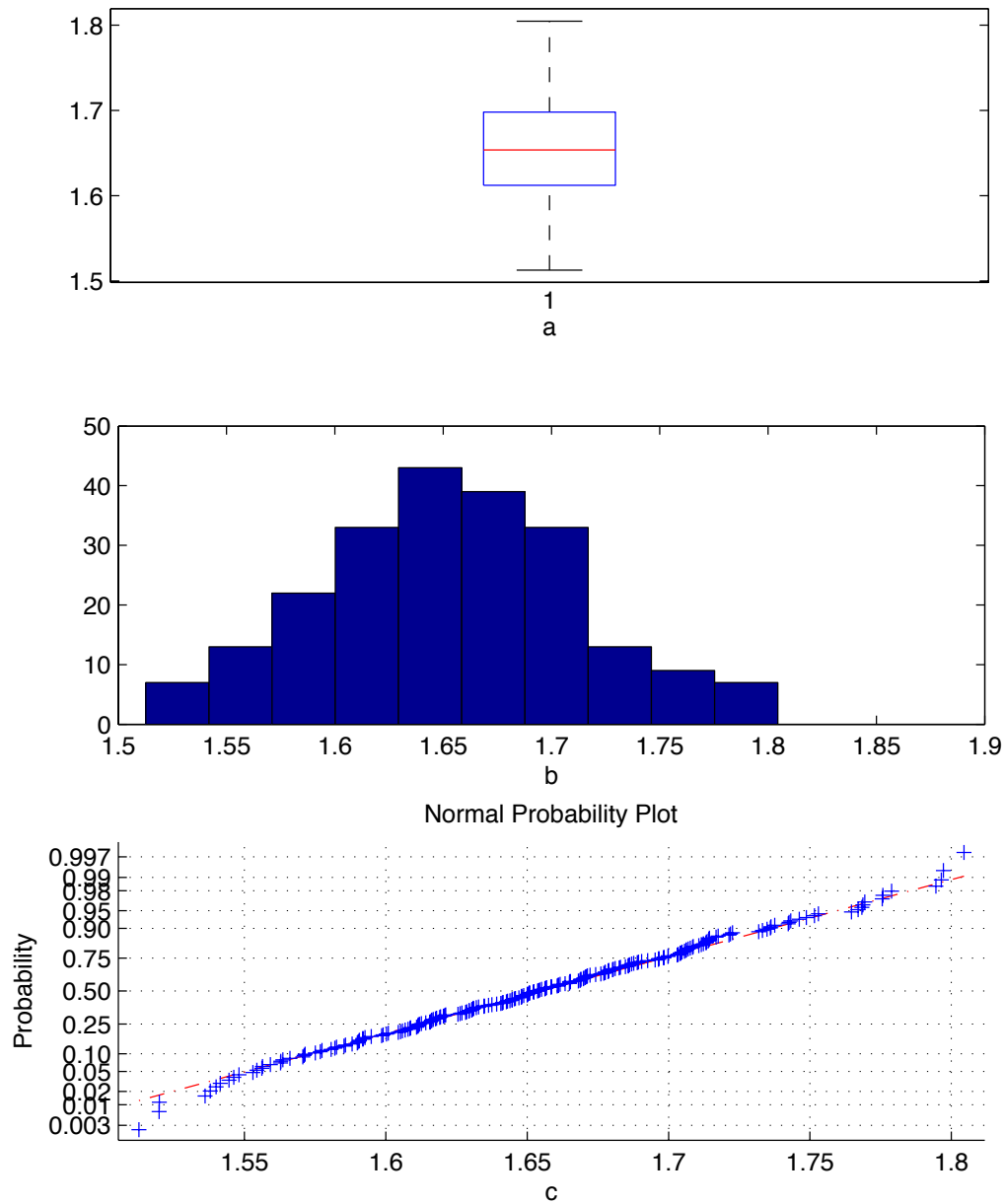


Figura 1. Diagrama Caja-Bigote, Histograma y Papel Probabilístico Normal para la variable ESTATURA

4.2 Caracterización energética del consumo de un horno

El segundo ejemplo presenta datos relativos a una empresa metalúrgica que quiere realizar una caracterización energética del consumo de un horno. Dicha empresa se dedica a la transformación de metales, desde el proceso de fundición, homogenizado, corte y laminación para obtener productos terminados de estos metales.

Para ello, se dispone de un total de 27 datos relativos al consumo durante los últimos 27 meses. El primer análisis indica que los valores estandarizados de asimetría y curtosis son de 1.28 y de 1 respectivamente, lo cual está dentro de los intervalos $[-2, 2]$ en ambos casos.

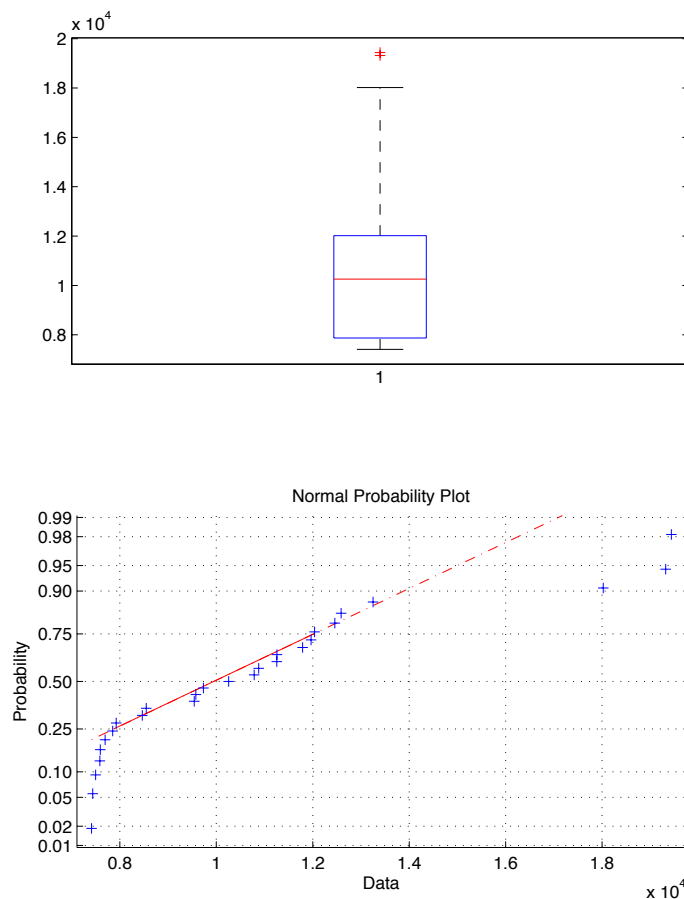


Figura 2. Diagrama Caja-Bigote y Papel probabilístico Normal para la variable CONSUMO.

Por otro lado, el diagrama Caja-Bigote presenta dos puntos que parecen anómalos, con un bigote derecho mucho más largo que el izquierdo (Figura 2, arriba), lo cual hace pensar en una asimetría positiva, a pesar de que el lado izquierdo de la Caja es algo más ancho que el de la derecha.

Sin embargo, es la inspección del Papel Probabilístico Normal (Figura 2, bajo) el que proporciona en este caso información clara, detectando la presencia de tres datos anómalos en la zona derecha del mismo (no dos como indicaba el diagrama Caja-Bigote). La diferencia entre ambos gráficos radica en que el valor 1.8×10^4 entra dentro del intervalo $C3 + 1.5 \times (C3 - C1)$, por lo que asume que pertenece a la distribución, cuando realmente no es así.

Tras eliminar los tres datos anómalos, la muestra presenta unos valores de asimetría y curtosis estandarizados de 0.17 y -1.41 respectivamente, indicando una posible distribución planicúrtica (lo cual también se aprecia en el diagrama Caja-Bigote por la mayor anchura de la Caja en relación con los Bigotes, Figura 3 arriba), si bien dentro de los límites $[-2, 2]$, por lo que no tenemos evidencia estadística suficiente de que esto sea así.

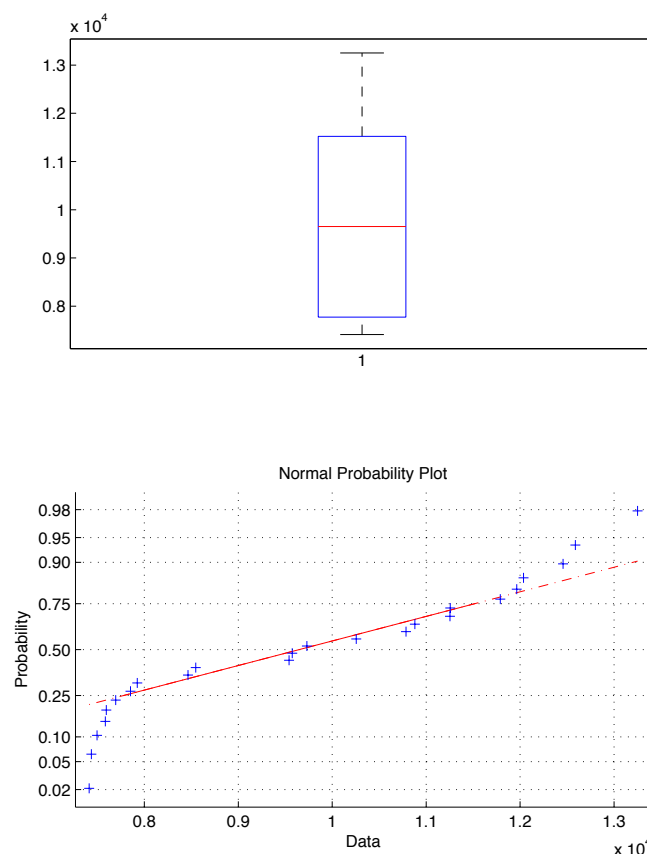


Figura 3. Diagrama Caja-Bigote y Papel probabilístico Normal para la variable CONSUMO.

Por último, el Papel Probabilístico Normal (Fig. 3, bajo) presenta en sus extremos ciertas curvaturas que también hacen pensar en cierta distribución planicúrtica, si bien pudiendo aproximar los datos de manera razonable a una distribución Normal.

4.3 No ajuste normal

El último ejemplo analiza una serie de datos en los que la distribución no se puede asimilar a una normal. En este caso, los datos provienen de los valores de intensidad a dos determinadas longitudes de onda, correspondientes a la captación de imágenes de alimentos. En este caso, la asimetría estandarizada presenta un valor de 0.09, y el de curtosis estandarizada de -1.75, ambos dentro del intervalo $[-2, 2]$.

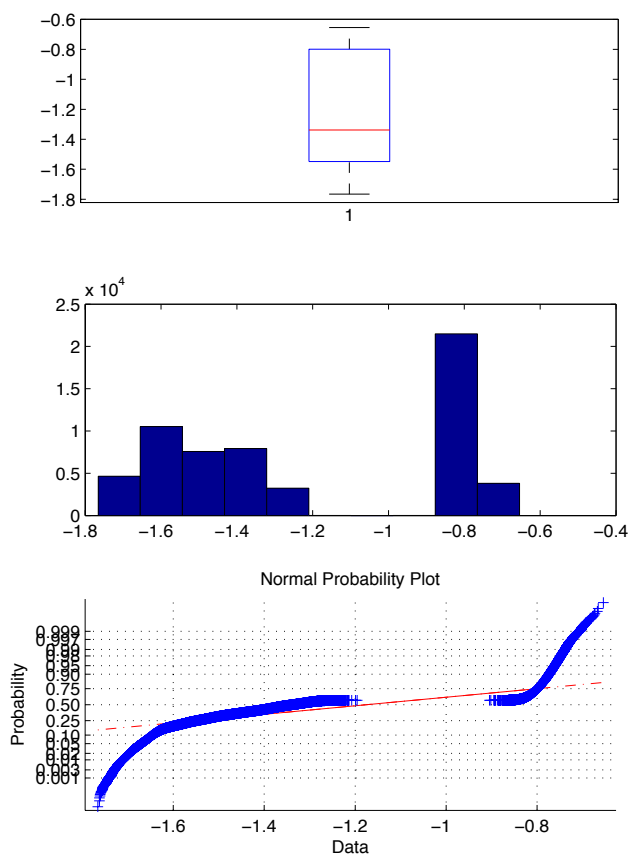


Figura 4. Diagrama Caja-Bigote, Histograma y Papel Probabilístico Normal para la variable INTENSIDAD

Sin embargo, el Histograma y el Papel Probabilístico Normal (Figura 4, medio y bajo), muestran claramente que nos encontramos frente a un caso de mezcla de poblaciones, posiblemente debido a las diferentes intensidades de dos fenómenos o tipo de objetos presentes en la imagen. En este caso, sería necesario analizar ambas poblaciones por separado.

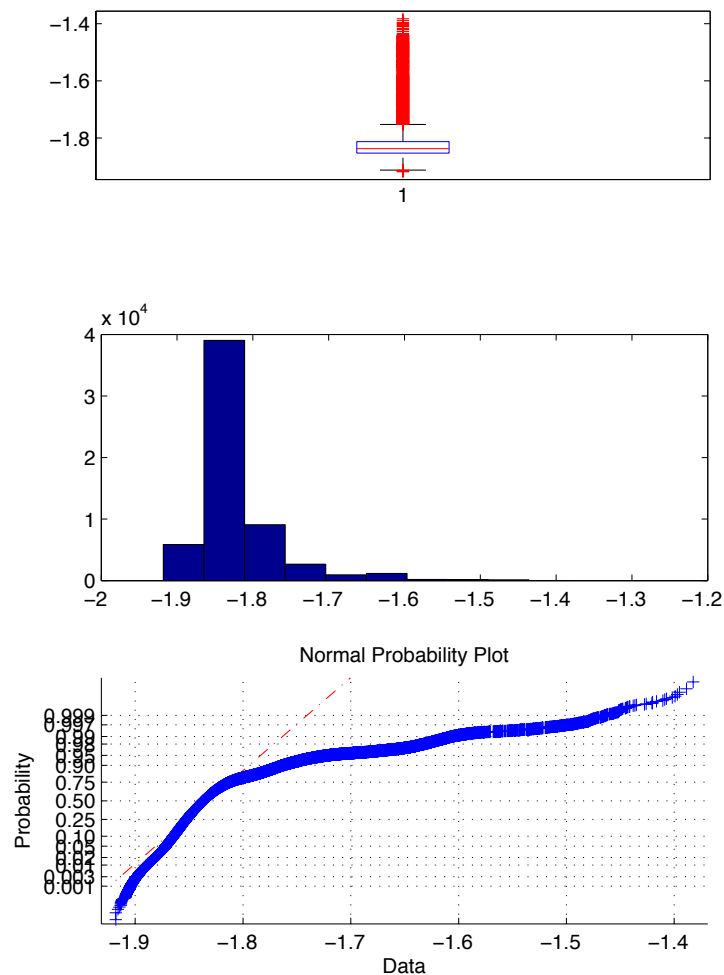


Figura 5. Diagrama Caja-Bigote, Histograma y Papel Probabilístico Normal para la variable INTENSIDAD



Para la siguiente longitud de onda, los valores de asimetría estandarizada y de kurtosis son 2.68 y de 9.91, respectivamente. En este caso, la Figura 5 revela la fuerte asimetría positiva existente, por lo que no es posible ajustar e una Normal. Es este caso, se podría intentar algún tipo de transformación (e.g. logarítmica) si bien en este caso no ha dado resultado, por lo que no se presentan.

5 Cierre

El presente objeto de aprendizaje ha presentado varios ejemplos de la utilización de los parámetros y gráficos que ayudan a caracterizar una distribución. Los aspectos prácticos se encuentran en el artículo docente "Gráficos y parámetros de posición, dispersión y forma de estadística descriptiva. Aspectos prácticos"; mientras que las cuestiones teóricas relativos a los mismos se encuentran en libro de texto tales como los recomendados en la Bibliografía.

6 Bibliografía

- [1] Peña, D. (2001). *Fundamentos de Estadística*. (Ed.) Alianza Editorial, S.A. Madrid. ISBN: 84-206-8696-4.
- [2] Romero, R y Zúnica, L.R. (1993). *Estadística (Proyecto de Innovación Educativa)*. SPUPV-93.637.