



## El papel probabilístico normal

<b>Apellidos, nombre</b>	Martínez Gómez, Mónica (momargo@eio.upv.es) Marí Benlloch, Manuel (mamaben@eio.upv.es)
<b>Departamento</b>	Estadística, Investigación Operativa Aplicadas y Calidad
<b>Centro</b>	Universidad Politécnica de Valencia



## 1. Resumen de las ideas clave

En este artículo vamos a comprender el funcionamiento y aplicaciones del papel probabilístico normal (ppn), con la finalidad de elaborar una especie de catálogo al que acudir por tratarse de una de las herramienta más utilizadas en el ámbito empresarial, tanto en el sector industrial como en el sector servicios, donde pueden aparecer problemas tanto en el proceso de fabricación como en la prestación de un servicio para ayudar a identificarlos y solventarlos.

## 2. Introducción

El empleo de papel probabilístico normal, propuesto por Daniel (1959), es una herramienta muy útil para distintos fines: comprobar la distribución que siguen los datos de una variable aleatoria continua, evaluar los efectos significativos en un diseño de experimentos mediante lo que se denomina Plot de Daniel, o bien en sistemas de control de calidad.

Con el fin de dominar esta herramienta, en primer lugar necesitamos partir de unos objetivos que pretendemos conseguir con la ayuda de este artículo; a continuación trabajaremos la definición y características del papel probabilístico normal, conociendo su formato gráfico específico, sus principales características y los pasos necesarios para la construcción del diagrama. Asimismo, se resuelven algunos ejemplos prácticos para facilitar su comprensión. Finalmente, destacaremos los conceptos básicos de aprendizaje con respecto al Papel probabilístico normal y sus aplicaciones prácticas.

## 3. Objetivos

- Construir la gráfica del papel probabilístico a partir de un conjunto de datos de la variable observada.
- Detectar el tipo de distribución que siguen las variables a partir de la representación en papel probabilístico.
- Identificar datos anómalos.
- Estimar media y desviación típica en el caso de variables normales.

## 4. Definición y características del Papel Probabilístico

La representación de los efectos en papel probabilístico normal (ppn) para analizar su significación estadística (método propuesto por Daniel en 1959) resulta una herramienta muy útil para analizar la significación estadística de distintos análisis: efectos que pueden considerarse significativos, tipo de

distribución que siguen las variables, presencia de datos anómalos, etc. Así es recogido en libros clásicos en el ámbito académico como el de Box, Hunter y Hunter (1978) o el de Montgomery (1997).

Dada la amplitud de este campo de estudio, en este artículo nos centraremos únicamente en la representación de los datos obtenidos de una variable aleatoria para comprobar el tipo de distribución que siguen: si es una distribución normal, o bien presenta asimetría, o es una mezcla de poblaciones, o bien, siendo normal presenta datos anómalos. Todo ello lo podremos detectar observando la forma que adquieren los puntos al representarlos en el papel probabilístico normal, como veremos a continuación.

## 4.2. Definición y Características

Básicamente consiste en una representación en el plano de un conjunto de datos, haciendo corresponder a la abscisa o eje x los valores de los datos de la variable estudiada, mientras que a la ordenada o eje y, le hace corresponder el porcentaje de valores en la muestra que son menores o iguales que el valor considerado, es decir, la frecuencia relativa acumulada.

El formato del papel probabilístico (imagen 3) es fijo y como se aprecia, la escala vertical de dicho papel está modificada al aplicar una pequeña corrección por continuidad, de manera que si la variable que estamos representando sigue una distribución normal, la curva típica de la función de distribución de las variables normales sufre una especie de “estiramiento” por la parte superior e inferior, transformándose en una recta. Es decir que cuando los datos que representamos proceden de una variable que sigue una distribución normal, los puntos correspondientes se sitúan aproximadamente en torno a una recta (Romero y Zúnica, 2000). En realidad, lo que hacemos al representar los valores es asumir que siguen una Normal con parámetros sin determinar, y a la vista de la disposición que adoptan los puntos en el ppn juzgamos correcta o no la suposición de normalidad. Una vez representados, observamos que algunos se alinean aproximadamente según una recta que pasa por el punto (0; 0,5), y estos son los que consideraremos que siguen una distribución normal con media cero. Los otros, los que se alejan de la recta por los extremos, serán considerados como datos anómalos.

Veamos a continuación como debemos proceder a la hora de representar los datos en papel probabilístico normal y a continuación veremos un ejemplo de los aspectos típicos de representaciones de las distintas situaciones con las que nos podemos encontrar.

• **Ejemplo:**

Representar en papel probabilístico normal los 10 datos obtenidos de la variable aleatoria,  $X \sim$  Millones de facturación de las Empresas del Mueble de la Comunidad Valenciana en el 2008.

Datos (Millones de facturación): 45, 80, 70, 55, 30, 26, 58, 62, 75, 40

Pasos:

1. En primer lugar procedemos a ordenar los datos de menor a mayor ya que representaremos frecuencia relativa acumuladas, y en consecuencia estaremos considerando la posición relativa del dato y no su propio valor. Este primer paso es similar a la representación de diagrama Box Whisker.

Datos ordenados de mayor a menor: 26, 30, 40, 45, 55, 58, 62, 70, 75, 80

2. En segundo lugar, para cada valor del dato, que vendrá representado en el eje x, debemos estimar su correspondiente valor de ordenada, mediante la expresión:

$$\frac{i - 0,5}{n} * 100$$

*Ecuación 1. Fórmula para el cálculo de la posición "i"*

donde "i" es la posición del dato y n es el número de datos de la muestra..

*Tabla 1. Valores y posición ("i") de los mismos para la representación*

Valor o Dato	$\frac{i - 0,5}{n} * 100$
26	$\frac{1 - 0,5}{10} * 100 = 5\%$
30	15%
40	25%
45	35%
55	45%
58	55%
62	65%
70	75%
75	85%
80	95%

3. A continuación procedemos a representar los valores obtenidos en la hoja del papel probabilístico.
4. El siguiente paso es interpretar los resultados en función de la forma obtenida al representar de los puntos. En este caso, los puntos correspondientes se sitúan aproximadamente a lo largo de una recta, con lo que podemos afirmar que la distribución es normal y la variable  $X \sim$

Millones de facturación de las Empresas del Mueble de la Comunidad Valenciana en el 2008, sigue una distribución  $N(\mu;\sigma)$  y el siguiente paso será proceder a la estimación de sus parámetros  $\mu$  y  $\sigma$ .

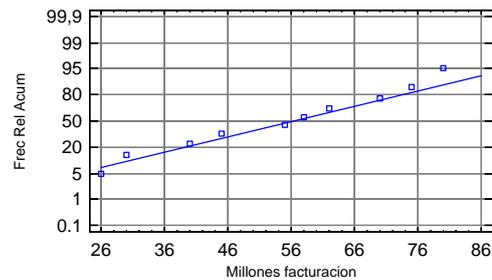


Imagen 1. Representación de los datos en papel probabilístico normal  
Fuente: Elaboración propia

5. Este paso **sólo será necesario** en caso de haber obtenido la conclusión de que mis datos siguen una distribución normal ya que si mi distribución no es normal, ni la media es un buen indicador de la posición ni la varianza o desviación típica es buen indicador de dispersión.
- **Cálculo de la media:** si la variable es normal media y mediana coinciden aproximadamente. Sabiendo que la mediana es el valor que deja el 50% de los datos por debajo, entramos por dicho valor del eje y, cortamos con la recta y leemos el valor del eje x correspondiente. En este caso,  $\mu$  sería aproximadamente 56 millones.

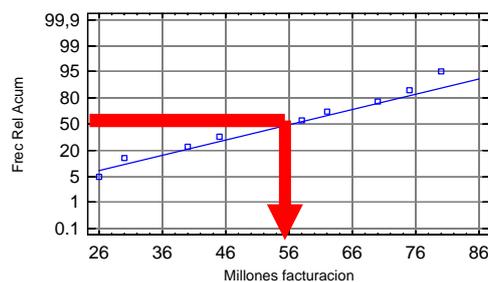


Imagen 2. Estimación del valor de la media mediante papel probabilístico normal  
Fuente: Elaboración propia

- **Cálculo de la desviación típica:** para la estimación de  $\sigma$ , utilizaremos la segunda propiedad que caracteriza a la distribución normal: "entre  $\mu \pm 2\sigma$  se encuentra el 95,5% de la distribución".



En consecuencia, si entramos por el valor del 95% del eje y procedemos como en el caso anterior, cortando con la recta y leyendo el valor del eje x correspondiente, podemos estimar  $\sigma$  como:

$$\sigma = \frac{\text{Valor del 95\%} - \text{Valor inicial}}{4} \cong \frac{\text{Valor Mediana} - \text{Valor inicial}}{2}$$

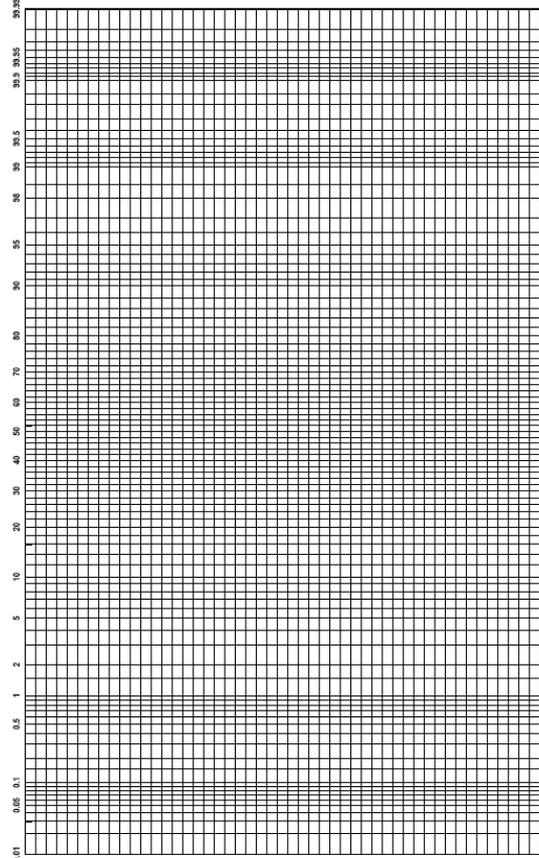
*Ecuación2. Fórmula para el cálculo de la desviación típica*

ya que por la propiedades de la campana de Gauss, sabemos que dos desviaciones típica quedan por debajo de la media y dos por encima. Es decir, en este caso:

$$\sigma = \frac{26 - 88}{4} \cong \frac{26 - 56}{2} \cong 15$$

*Ecuación3. Fórmula para el cálculo de la desviación típica*

Papel Probabilístico Normal



*Imagen 3. Papel probabilístico Normal*

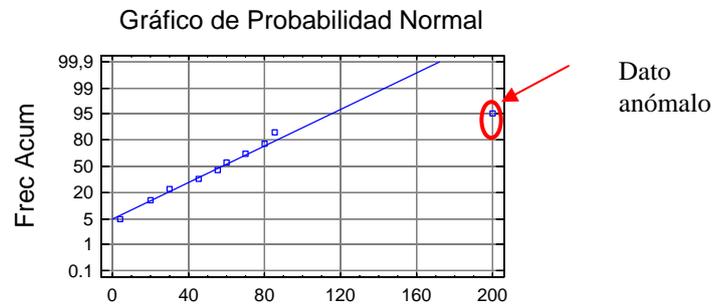
**Fuente:** <http://taylor.us.es/pl/personal/docencia/mdpp/2005-06/repositorio/ppn.pdf>

### 4.3. Interpretación de los posibles resultados obtenidos en el papel probabilístico

Los distintos aspectos que pueden aparecernos al representar los datos en papel probabilístico normal y su interpretación son los que se muestran a continuación:

a. *Presencia de dato anómalo.*

Los datos se sitúan aproximadamente en torno a una recta, pero hay algunos se alejan de ella, bien en la zona superior derecha, bien en la zona inferior izquierda.



*Imagen 4. Papel probabilístico Normal de datos normales con datos anómalos*  
**Fuente:** Elaboración propia

b. *Asimetría Positiva.*

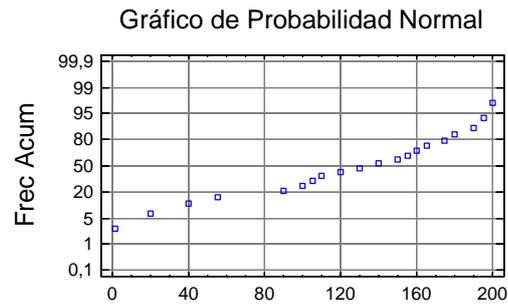
Las distribuciones asimétricas positivas presentan un papel probabilístico con una figura de puntos que forma una especie de curvatura hacia abajo, como se muestra en la figura siguiente:



*Imagen 5. Papel probabilístico Normal de datos asimétricos positivos*  
**Fuente:** Elaboración propia

c. *Asimetría Negativa.*

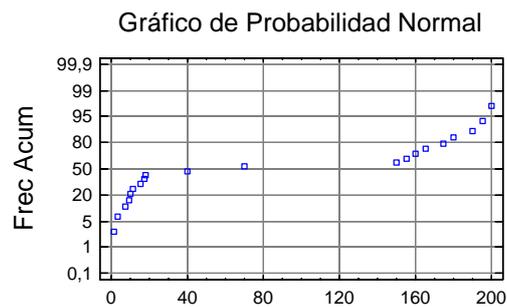
Las distribuciones asimétricas negativas presentan un papel probabilístico con una figura de puntos que forma una especie de curvatura hacia arriba, como se muestra en la figura siguiente:



*Imagen 6. Papel probabilístico Normal de datos asimétricos negativos*  
**Fuente:** Elaboración propia

d. *Mezcla de dos poblaciones.*

La mezcla de dos poblaciones con media diferentes, presenta una forma típica don dos zonas de crecimiento rápido separado por un tramo de crecimiento lento (Romero y Zúnica, 2000).



*Imagen 7. Papel probabilístico Normal de mezcla de poblaciones*  
**Fuente:** Elaboración propia



## 5. Cierre

En este objeto de aprendizaje hemos tratado la construcción y utilidad del papel probabilístico normal.

Hemos visto que es una herramienta muy útil para:

- ✓ comprobar la distribución que siguen los datos de una variable aleatoria continua,
- ✓ evaluar los efectos significativos en un diseño de experimentos,
- ✓ detectar datos anómalos en sistemas de control de calidad,
- ✓ estimar los parámetros de ciertas distribuciones.

### ¿Y en qué consiste el ppn?

Consiste en una representación en el plano de un conjunto de datos, haciendo corresponder a la abscisa (eje x) los valores de los datos de la variable estudiada, mientras que a la ordenada (eje y), le hace corresponder la frecuencia relativa acumulada con una ligera modificación de escala de manera que ordenados los datos de menor a mayor, a la observación “i-esima”, la hace corresponder como ordenada el valor  $((i-0,5)/N)*100$ .

## 6. Bibliografía

### 6.1. Libros:

- [1] Box, E. P., Hunter, W. y Hunter, J. (1993). Estadística para Investigadores, Introducción al Diseño de Experimentos, Análisis de Datos y Construcción de Modelos. Editorial Reverté
- [2] DeGroot, M.H. (1988). *Probabilidad y Estadística*. (2ª Ed.). Addison-Wesley Iberoamericana. ISBN 0-201-64405-3.
- [3] Martín Pliego, F.J. (2004). *Introducción a la Estadística Económica y Empresarial*. (Ed.) Thomson. Madrid.
- [4] Mendenhall, W.; Reinmuth, J.E. (1978). *Estadística para administración y economía*. (Ed.) Grupo Editorial Iberoamericana. ISBN 968-7270-13-6.
- [5] Montgomery, D. C. (1997). *Design and Analysis of Experiments*. John Wiley
- [6] Montiel, A.M.; Rius, F.; Barón F.J. (1997). *Elementos básicos de Estadística Económica y Empresarial*. (2ª Ed.) Prentice Hall, Madrid.
- [7] Peña, D. (2001). *Fundamentos de Estadística*. (Ed.) Alianza Editorial, S.A. Madrid. ISBN: 84-206-8696-4.
- [8] ROMERO, R y ZÚNICA, L.R. (2000). *Introducción a la Estadística*. (Ed.) SPUPV 4071.
- [9] Uña Juárez, I; Tomero, V; San Martín, J. (2003). *Lecciones de cálculo de probabilidades*. (Ed.) Thomson Editores Spain. ISBN 84-9732-193-6.



## 6.2. Referencias de fuentes electrónicas:

- [8] [http://personal5.iddeo.es/ztt/Tem/t21\\_distribucion\\_normal.htm](http://personal5.iddeo.es/ztt/Tem/t21_distribucion_normal.htm) (Consultado 17/10/08).
- [9] [http://74.125.39.104/search?q=cache:nIXKW90gaC0J:www.udl.es/usuarios/seio2003/treballs/06\\_2\\_1.pdf+PAPEL+PROBABIL%C3%8DSTICO&hl=es&ct=clnk&cd=7&gl=es](http://74.125.39.104/search?q=cache:nIXKW90gaC0J:www.udl.es/usuarios/seio2003/treballs/06_2_1.pdf+PAPEL+PROBABIL%C3%8DSTICO&hl=es&ct=clnk&cd=7&gl=es)
- [10] <http://sauce.pntic.mec.es/~jpeo0002/Archivos/PDF/XT03.pdf>
- [11] [http://www.vitutor.com/pro/5/a\\_g.html](http://www.vitutor.com/pro/5/a_g.html)
- [12] [http://es.geocities.com/pilar\\_zutabe/EJERCICIOS/1BACHILLERHUMANISTICO/Ejerciciosdistribucion\\_normal.htm](http://es.geocities.com/pilar_zutabe/EJERCICIOS/1BACHILLERHUMANISTICO/Ejerciciosdistribucion_normal.htm)
- [13] [http://www.eticayempresa.com/congreso/06\\_2\\_1.pdf](http://www.eticayempresa.com/congreso/06_2_1.pdf)