La distribución Normal

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

1. Resumen de las ideas clave

En este artículo vamos a conocer las características básicas de la distribución Normal y sus posibles aplicaciones prácticas con la finalidad de elaborar una especie de catálogo al que acudir para determinar un modelo de probabilidad para describir el comportamiento de una variable continua.

2. Introducción

¿Qué conozco sobre variables aleatorias (V.A.) continuas y los tipos de distribuciones que éstas pueden seguir? ¿Cómo afectará a los análisis estadísticos el que los datos de una V.A. tengan una distribución normal?

Una variable aleatoria se define continua cuando el conjunto de valores que puede tomar es un infinito continuo, es decir, puede tomar cualquier valor en un intervalo. La distribución normal o distribución de Gauss es una de las distribuciones de probabilidad de variables continuas que aparece con más frecuencia en fenómenos reales, frente a otros tipos de distribuciones como las asimétricas o la exponencial

En este objeto de aprendizaje, conoceremos las características y propiedades de la distribución normal. Utilizamos ejemplos y ejercicios donde descubriremos los principales aspectos sobre el cálculo de probabilidades mediante la distribución normal y sus principales aplicaciones prácticas, para ayudar a la comprensión de las mismas. Finalmente, hacemos hincapié en los conceptos básicos de aprendizaje con respecto a la distribución normal y sus aplicaciones prácticas

3. Objetivos

- Identificar las propiedades de una distribución normal.
- Determinar cómo se tipifican las variables Normales.
- Buscar probabilidades en la tabla de la Normal Tipificada.
- Interpretar áreas bajo la curva normal de acuerdo al problema.

4. Definición y características de la distribución Normal

4.1. ¿Por qué es importante conocer la distribución normal?

La distribución normal fue desarrollada por Abraham de Moivre, (1667-1754). Posteriormente, Carl Friedrich Gauss (1777-1855) elaboró desarrollos más profundos y formuló la ecuación de la curva; de ahí que también se la conozca, más comúnmente, como la **"campana de Gauss"**. La distribución de una variable normal está completamente determinada por dos parámetros, su media y su desviación estándar, denotadas generalmente por μ y σ.

La distribución normal es la distribución de probabilidad más importante en estadística, debido a tres razones fundamentales (DeGroot, M.H., 1988):

- Desde un punto de vista matemático resulta conveniente suponer que la distribución de una población de donde se ha extraído una muestra aleatoria sigue una distribución normal, ya que entonces se pueden obtener las distribuciones de varias funciones importantes de las observaciones muestrales, que además resultan tener una forma sencilla.
- Desde un punto de vista científico, la distribución normal aproxima en muchas ocasiones los valores obtenidos para variables que se miden sin errores sistemáticos. Por ejemplo, se ha observado que muchos experimentos físicos frecuentemente tienen distribuciones que son aproximadamente normales, como estaturas o pesos de los individuos, beneficios medios de las empresas, la duración de un producto perecedero, el tiempo necesario para llevar a cabo un trabajo, etc.
- La última razón es la existencia del Teorema Central del Límite, establece que cuando se dispone de una muestra aleatoria grande, aunque presente una distribución no normal e incluso distribuciones típicas de variables aleatorias discretas, pueden tratarse como aproximadamente distribuciones normales.

Algunos ejemplos típicos de la distribución normal son:

~ Estatura de las personas.

~ Tª de una cámara frigorífica.

~ Dosis de un aditivo.

~ Precipitaciones anuales de un determinado país.

4.2. Definición y Características

Se dice que X sigue una distribución Normal, con media μ y desviación típica σ, con ($-\infty < \mu < +\infty$ y $\sigma > 0$), que se representa con la siguiente notación:

$$X \sim N(\mu, \sigma)$$

Su Función de Densidad viene definida por:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{(x.\mu)^2}{2\sigma^2}} \quad \text{con } (-\infty < \mu < +\infty)$$

Ecuación 1. Función de Densidad de la distribución Normal.

donde los símbolos *e* y π tienen valores aproximadamente 2.7183 y 3,1416.

La Función de Probabilidad permite calcular la probabilidad de que una variable aleatoria *X*, pertenezca a un intervalo [a,b]. Dicha probabilidad se obtiene como:

$$F(X) = \int_{b}^{a} f(X)dx$$

Ecuación 2.Función de Probabilidad distribución Normal.

Se cumple que,

$$\int_{+\infty}^{-\infty} f(X)dx = 1$$

Ecuación 3. Propiedad de la Función de Probabilidad de la Distribución Normal

El histograma correspondiente a la Función de Densidad, tiene forma de una campana simétrica con una densidad o valor máximo en la media, μ y dicha densidad decrece de forma simétrica a ambos lados en función del valor de la desviación típica, σ. Gráficamente, como se puede apreciar en la figura 1, se observa que σ es la distancia desde la media μ al punto de inflexión de la campana (Romero y Zunica, 2000).

El parámetro de posición (es decir, donde caen más frecuentemente los valores), más adecuado en distribuciones normales es la media, μ, que como puede apreciarse en la figura de la función de densidad, se corresponde con la densidad máxima. En las variables que siguen una distribución normal la media toma un valor muy próximo o coincide con la mediana y con la moda, con lo ambos parámetros también serían buenos indicadores de posición.

El indicador de dispersión (mide la variabilidad de mis datos) más adecuado es la varianza, σ^2, aunque realmente el más utilizado es su raíz cuadrada o desviación típica, σ, al venir expresada en las mismas unidades que la variable medida y, por lo tanto, que la media.

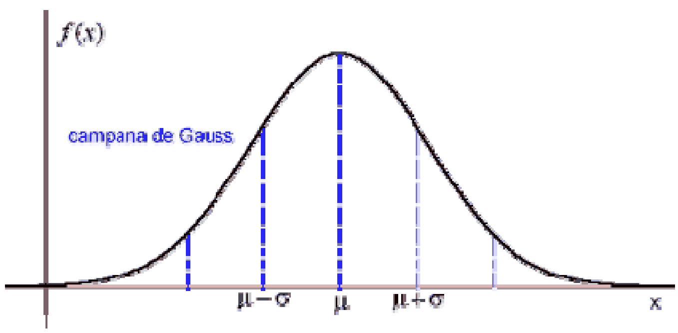

Imagen 1. La función de densidad de una distribución normal
Fuente: http://personal5.iddeo.es/ztt/Tem/t21_distribucion_normal.htm

Respecto a los parámetros de asimetría y curtosis, toda variable que presente una distribución normal tiene un coeficiente de asimetría nulo, como puede deducirse de la gráfica de la función de densidad, que decrece de forma simétrica a ambos lados. Así mismo presenta valores de curtosis de cero o tres en función de la fórmula utilizada en el cálculo de dicho coeficiente.

La distribución cumple tres propiedades básicas normal que constituyen la base de las técnicas utilizadas en control estadístico de procesos (Romero y Zúnica, 2000). Como puede observarse en la gráfica de la función de densidad, se comprueba que en toda distribución normal, en el intervalo:

1. $\mu \pm \sigma$ se encuentra el 68% de la distribución.
2. $\mu \pm 2\sigma$ se encuentra el 95,5% de la distribución.
3. $\mu \pm 3\sigma$ se encuentra el 99,7% de la distribución.

Es decir existe una probabilidad inferior al 5% de encontrar un valor de una variable aleatoria que siga una distribución normal que difiera de su media en más de dos desviaciones típicas y es prácticamente improbable, con una probabilidad inferior al 3 por mil, de encontrar un valor que difiera de su media en más de tres desviaciones típicas.

Un Teorema importante se desprende de las dos propiedades que es establecen a continuación:

1. Si $X_1, \ldots, X_n$ son variables aleatorias independientes, con media μ_i y varianza σ_i^2, y $b_1, \ldots b_k$ y a, son constantes distintas de cero, la transformada lineal $Y= a+(b_1X+ \ldots b_nX_n)$, se distribuye también normalmente con media $a+b_1\mu_1+\ldots+b_k\,\mu_k$ y varianza $b_1^2\sigma_1^2+\ldots+b_n^2\sigma_k^2$.
2. Si $X_1, \ldots, X_n$ son variables aleatorias independientes, con media μ_i y varianza σ_i^2, cualquier suma o resta de ellas, $Y= X_1+\ldots+X_n$, tiene también una distribución normal con media $\mu_1+\ldots+\mu_k$ y varianza $\sigma_1^2+\ldots+\sigma_k^2$.

Es decir, cualquier combinación lineal de variables normales independientes, sigue también una distribución normal.

Veamos algunos ejemplos:

- **Ejemplo 1.** *En una ciudad se estima que la temperatura máxima en el mes de junio si una distribución normal, con media 23° y desviación típica 5°. Calcular el número de días del mes en los que se espera alcanzar máximas entre 21° y 27°.*

  - X~ Temperatura Máxima en mes de junio~ N(*23, 5*)

$$P(21 < X < 27) = P(N(23,5) < 21) - P(23,5) < 27) =$$
$$P(N(0,1) < \frac{21-23}{5}) - P(N(0,1) < \frac{27-23}{5}) = [1 - P(N(0,1) > 0,4)] - [P(N(0,1) > 0,8)$$
$$= 0,7881 - (1 - 0.6554) = 0,4425$$

Luego el número de días sería: $0,44425 * 30 = 13$

- **Ejemplo 2.** *Las Empresas multinacionales del sector automoción presentan hasta enero del 2009 un volumen de facturación, que seguía una distribución normal con media 185 millones de euros y desviación típica 12 millones de euros. ¿Qué porcentaje de empresas facturará entre 160 y 200 millones?*

X~ volumen de facturación~ N(*185*, *12*)

$$P(160 < X < 200) = P(N(185{,}12) < 200) - P(185{,}12) < 160) =$$
$$P(N(0{,}1) < \frac{200-185}{12}) - P(N(0{,}1) < \frac{160-185}{12}) = [1 - P(N(0{,}1) > 1{,}25)] - [P(N(0{,}1) > 2{,}08)$$
$$= [1 - 0{,}1056] - 0{,}0188 = 0{,}8756$$

- **Ejemplo 3.** *El peso del las personas que utilizan un ascensor fluctúan normalmente con media 75 kg y desviación típica 11 kgs. Si la carga máxima del ascensor es de 500 kgs, ¿cuál es la probabilidad de que al subir 6 personas en el ascensor se sobrepase dicha carga? (Extraído de Romero y Zúnica, 2003).*

X~ Peso de una persona~ N(*75*, *11*)

Y~ Peso de seis personas=$X_1+X_2+X_3+X_4+X_5+X_6$~ N(*450*, *26,9*)

$$P(Y > 500) = P(N(0{,}1) > \frac{500-450}{26{,}9}) - P(N(0{,}1) > 1{,}86) = 0{,}0314$$

4.3. El proceso de Tipificación. La tabla de la distribución normal

La distribución normal con $\mu = 0$ y *σ=1*, *N(0,1)*, se llama *distribución normal tipificada o distribución normal estándar* y su función de distribución está tabulada para determinados valores, como se aprecia en la tabla 1 de la página siguiente, ya que, calcular la probabilidad de que una variable normal tome valores superiores a un *z* dado, equivale al cálculo de la integral de la función de densidad y esta integral no puede estimarse directamente para valores de *z*, entre 0 y 4, por no existir la primitiva de f(x):

$$P(N(0{,}1) > z) = \int_z^{\infty} f(x)dx = \int_z^{\infty} \frac{1}{\sqrt{2\pi}} * e^{-\frac{(x)^2}{2^2}} dx$$

siendo la representación gráfica de esta función la que se muestra en la imagen 2.

Para calcular probabilidades en el caso de variables normales, con media μ y desviación típica *σ*, *N(* μ, *σ)*, se debe transformar la variable a una variable normal tipificada, N(0,1) el siguientes proceso:

$$z = \frac{x-\mu}{\sigma}$$

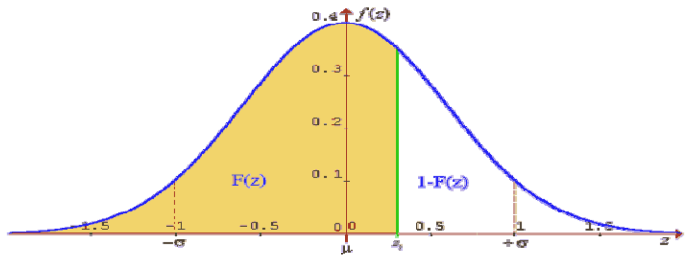

Imagen 2. *La función de densidad de una distribución normal tipificada N(0,1)*
Fuente: http://personal5.iddeo.es/ztt/Tem/t21_distribucion_normal.htm

Es decir, que es posible calcular el valor de la función de distribución de cualquier variable normal en cualquier punto si conocemos la distribución de la normal tipificada o estándar. Sólo tenemos que convertir el punto *a*, restándole la media y dividiendo por la desviación típica (Peña, 2001):

$$P(N(\mu,\sigma) > a) = P(N(0,1) > \frac{a-\mu}{\sigma}) = P(N(0,1) > z)$$

donde el z se obtiene de la tabla 1.

Debido a las propiedades que se especifican a continuación, es posible calcular también la P(*N(μ, σ)*< a), la P(*N(μ, σ)*> -a) y la P(*a*< *N(μ, σ)*< b):

1. La curva normal es asintótica al eje de abscisas. Por ello, cualquier valor entre - ∞ y +∞ es teóricamente posible. El área total bajo la curva es, por tanto, igual a 1.
2. Es simétrica con respecto a su media. Según esto, para este tipo de variables existe una probabilidad de un 50% de observar un dato mayor que la media, y un 50% de observar un dato menor.
3. En consecuencia, y a partir de estas propiedades, gráficamente, podemos observar que podemos estimar:
    a. P(*N(μ, σ)*< a) = 1-P(*N(μ, σ)*> a), por ser ambas probabilidades complementarias y el área total bajo la curva es igual a 1

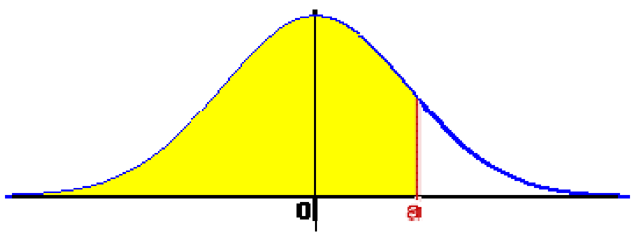

Imagen 3. *Área de probabilidad menor que un valor "a"*
Fuente: http://personal5.iddeo.es/ztt/Tem/t21_distribucion_normal.htm

    b. P(*N(μ, σ)*> -a) = P(*N(μ, σ)*< a) = 1-P(*N(μ, σ)*> a), por simetría respecto a su media.

c. P(a< $N(\mu, \sigma)$< b) = P($N(\mu, \sigma)$< b)- P($N(\mu, \sigma)$< a) = [1-P($N(\mu, \sigma)$> b)]-[1-P($N(\mu, \sigma)$> a)]

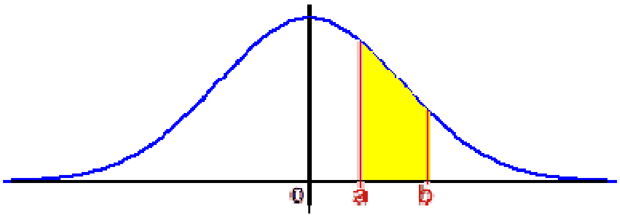

Imagen 4. *Área de probabilidad de un intervalo*
Fuente: http://personal5.iddeo.es/ztt/Tem/t21_distribucion_normal.htm

Tabla 1. *Probabilidades de que una normal tipificada tome valores mayores que z*

DEPARTAMENT D'ESTADÍSTICA
I INVESTIGACIÓ OPERATIVA
APLICADES I QUALITAT

DISTRIBUCIÓN
Normal Tipificada

|z|	0	1	2	3	4	5	6	7	8	9
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3	0.0014	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0010	0.0010	0.0010
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.6	0.0047	0.0045	0.0044	0.0043	0.0042	0.0041	0.0039	0.0038	0.0037	0.0036
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.4	0.0082	0.0080	0.0078	0.0076	0.0073	0.0071	0.0070	0.0068	0.0066	0.0064
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.2	0.0139	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0376	0.0367
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.2	0.1151	0.1131	0.1112	0.1094	0.1075	0.1057	0.1038	0.1020	0.1003	0.0985
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.0	0.1587	0.1563	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2207	0.2177	0.2148
0.6	0.2743	0.2709	0.2676	0.2644	0.2611	0.2579	0.2546	0.2514	0.2483	0.2451
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.2	0.4207	0.4168	0.4129	0.4091	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4841	0.4801	0.4761	0.4721	0.4681	0.4641

Veamos algunos ejemplos:

- **Ejemplo 1**: *Calcular la probabilidad de que una variable X con distribución normal del media 4 y desviación típica 2 sea menor que 5.*

$$P(N(4,2)<5)=1-P(N(4,2)>5=1-P(N(0,1)>\frac{5-4}{2})=$$
$$1-P(N(0,1)>0,5)=0,3085$$

- **Ejemplo 2**: *Calcular la probabilidad de que una variable X con distribución normal del media 4 y desviación típica 2 sea mayor que -5.*

$$P(N(4,2)>-5)=P(N(4,2)<-5)=P(N(0,1)>\frac{5-4}{2})=$$
$$P(N(0,1)>0,5)=1-0,3085=0,6915$$

- **Ejemplo 3**: *Calcular la probabilidad de que una variable X con distribución normal del media 4 y desviación típica 2 sea menor que -5.*

$$P(N(4,2)<-5)=P(N(4,2)>5)=P(N(0,1)>\frac{5-4}{2})=$$
$$P(N(0,1)>0,5)=0,3085$$

- **Ejemplo 4**: *Calcular la probabilidad de que una variable X con distribución normal del media 4 y desviación típica 2 este en l intervalo 3 y 5.*

$$P(N(2<(4,2)<5)=P(N(4,2)<5)-P(N(4,2)<2)=$$
$$[1-P(N(0,1)>\frac{5-4}{2})]-P(N(0,1)<\frac{2-4}{2}=$$
$$[1-P(N(0,1)>0,5)]-(P(N(0,1)>1)=1-0,3085-0,1587=0,5328$$

5. Cierre

Una variable, X que sigue una distribución Normal, con media μ y desviación típica σ , tiene una función de distribución característica, con la típica forma de campana de Gauss, con una densidad o valor máximo en la media, μ y dicha densidad decrece de forma simétrica a ambos lados en función del valor de la desviación típica, σ. Esta función de densidad cumple tres condiciones básicas que constituyen la base de las técnicas utilizadas en control estadístico de procesos: **$\mu \pm \sigma$ se encuentra el 68% de la distribución; $\mu \pm 2\sigma$ se encuentra el 95,5% de la distribución; $\mu \pm 3\sigma$ se encuentra el 99,7% de la distribución.**

Dos propiedades importantes a recordar para el cálculo de probabilidades, se desprenden de la forma típica de campana de Gauss de la función de densidad:

1. La curva normal es asintótica al eje de abscisas. Por ello, cualquier valor entre - ∞ y +∞ es teóricamente posible. El área total bajo la curva es, por tanto, igual a 1.
2. Es simétrica con respecto a su media. Según esto, para este tipo de variables existe una probabilidad de un 50% de observar un dato mayor que la media, y un 50% de observar un dato menor

Se cumple la condición de que cualquier combinación lineal de variables normales independientes, sigue también una distribución normal.

Para efectuar el cálculo de probabilidades de variables normales, con media μ y desviación típica *σ, N(μ, σ),* se debe transformar la variable a una variable normal tipificada, N(0,1), cuya función de distribución está tabulada para determinados valores y se obtiene directamente de tablas.

6. Bibliografía

6.1. Libros:

[1] DeGroot, M.H. (1988). *Probabilidad y Estadística.* (2ª Ed.). Addison-Wesley Iberoamericana. ISBN 0-201-64405-3.

[2] Martín Pliego, F.J. (2004). *Introducción a la Estadística Económica y Empresarial*. (Ed.) Thomson. Madrid.

[3] Mendenhall, W.; Reinmuth, J.E. (1978). *Estadística para administración y economía*. (Ed.) Grupo Editorial Iberoamericana. ISBN 968-7270-13-6.

[4] Montiel, A.M.; Rius, F.; Barón F.J. (1997*). Elementos básicos de Estadística Económica y Empresarial*. (2ª Ed.) Prentice Hall, Madrid.

[5] Peña, D. (2001). *Fundamentos de Estadística*. (Ed.) Alianza Editorial, S.A. Madrid. ISBN: 84-206-8696-4.

[6] Romero, R y Zúnica, L.R. (1993). Estadística (Proyecto de Innovación Educativa). SPUPV-93.637.

[7] Romero, R y Zúnica, L.R. (2000). Introducción a la Estadística. (Ed.). SPUPV- 2000.4071.

6.2. Referencias de fuentes electrónicas:

[8] 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/usuaris/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

[12] http://es.geocities.com/pilar_zutabe/EJERCICIOS/1BACHILLERHUMANISTICO/Ejerciciosdistribucionnormal.htm

[13] http://www.digeo.cl/asignaturas/mat/Ejercicios-Distribucion-Normal.pdf

[14] http://www.fisterra.com/mbe/investiga/distr_normal/distr_normal.asp

[15] www1.uprh.edu/.../La%20distribucion%20normal/Modulo%20Sobre%20La%20Distribucion%20Normal%

[16] https://polimedia.upv.es/visor/?id=e7dd2019-e8f4-a44c-8935-aa3e3da14449