UNIDAD DIDÁCTICA 4

DISTRIBUCIONES DE PROBABILIDAD – PARTE 3

Objetivo: El objetivo de esta Unidad Didáctica es introducir de forma resumida los conceptos elementales sobre las distribuciones de probabilidad y la esperanza matemática indispensables para el resto de la asignatura, así como presentar sucintamente los modelos más importantes en la práctica para variables aleatorias discretas (Binomial y Poisson) y continuas (Uniforme, Exponencial y Normal).

De acuerdo con el enfoque general de la asignatura, el tratamiento de los conceptos expuestos se realiza a un nivel elemental e intuitivo, pero suficiente para que un ingeniero pueda entenderlos y aplicarlos a los problemas reales que puedan aparecer en su ejercicio profesional.

Contenido

Fuente: Romero, R.; Zúnica, L. R. *Métodos estadísticos en ingeniería*. I.S.B.N 84-9705-727-9

1. Principales distribuciones de probabilidad (continuación)

En esta tercera parte de la Unidad Didáctica se presenta un modelo de distribución de probabilidad para variables aleatorias continuas de especial interés por su importancia práctica. Esta distribución de probabilidad se denomina **distribución Normal**.

1.5. La distribución Normal

En este apartado se define la distribución normal y se exponen sus propiedades más importantes, haciendo hincapié en el manejo de tablas.

1.5.1. Concepto

En el mundo real aparecen con mucha frecuencia variables aleatorias continuas para las que se constata una pauta de variabilidad caracterizada por una acumulación de valores en el entorno de una zona central y unas frecuencias que decrecen de forma aproximadamente simétrica a medida que éstos se alejan de dicho valor central.

Este comportamiento probabilístico da lugar a histogramas cuya forma recuerda a la de una campana[1] (**Figura 9**). Dicha pauta de variabilidad, que como hemos indicado es muy corriente en datos reales, puede modelarse razonablemente asumiendo que la variable estudiada sigue en la población una distribución de probabilidad denominada **distribución Normal**.

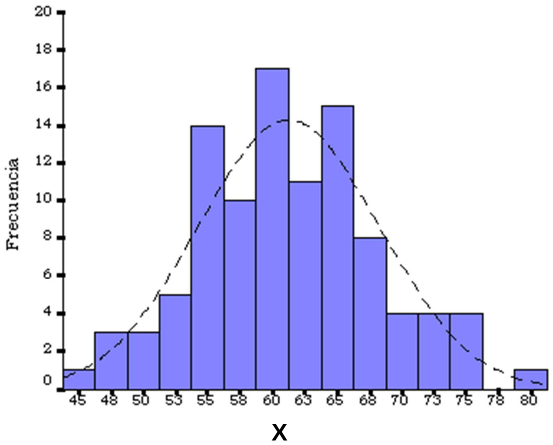


Figura 9

Para indicar que la variable aleatoria **X** sigue una distribución **Normal** de media **m** y desviación típica **σ** utilizamos la expresión:

$$\mathbf{X} \sim N(m,\sigma)$$

[1] Es por ello que a esta forma de la distribución también se le conoce como Campana de Gauss. También tiene otras denominaciones como distribución de Laplace o distribución de Gauss, que hacen referencia a los apellidos de dos famosos astrónomos y matemáticos que utilizaron esta distribución para el estudio de los errores en sus observaciones.

La distribución normal se caracteriza por su **función de densidad f(x)** que viene dada por la expresión:

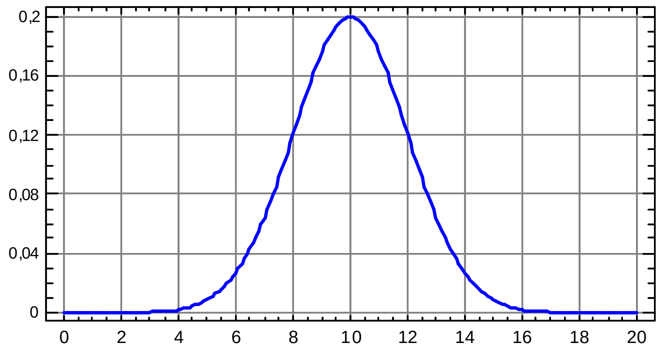
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad (-\infty < x < +\infty)$$

donde **m** = μ = media de la distribución

σ = desviación típica de la distribución

Dicha función de densidad f(x) (**Figura 9**) tiene forma, como ya se ha mencionado, de una curva en campana, con una densidad máxima en m, que es la media y la mediana de la distribución. La densidad decrece de forma simétrica a ambos lados de la media, de forma más o menos rápida en función del valor que tenga la desviación típica.

f(x)



media = mediana = moda = 10

Figura 9. f(x) para una variable aleatoria X~N(m=10, σ=2)

1.5.2. Propiedades

Dos propiedades importantes de la distribución normal son las siguientes:

1 - Si **X** se distribuye normalmente cualquier transformada lineal **Y** = a + b**X** se distribuye también normalmente.

2 - Si **X** e **Y** son dos variables **independientes** que se distribuyen normalmente, su suma **Z** = **X** + **Y** se distribuye también normalmente.

Las propiedades 1 y 2 implican que **cualquier combinación lineal de variables normales independientes se distribuye también de forma normal**.

1.5.3. Tabla

Una variable normal se dice que es **tipificada** si su **media es cero y** su **desviación típica es uno**. Nos referiremos a esta distribución utilizando la expresión:

Z ~ N(0,1)

La probabilidad de que una N(0,1) sea mayor que un valor dado z vendrá dada por la integral de su función de densidad desde z hasta 4

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

Los valores de esta integral deben obtenerse por métodos numéricos dado que no puede calcularse directamente al no existir la primitiva de f(x) y vienen recogidos en la tabla de la página 9 para valores de z entre 0 y 4.

Obviamente la tabla puede también utilizarse para calcular probabilidades del tipo P(**X**<a), sin más que hacer dicha probabilidad igual a 1 - P(**X**>a).

Dado, por otra parte, que una distribución N(0,1) es simétrica respecto al origen, P(N(0,1)>z) = P(N(0,1)<-z), por lo que la tabla da también directamente la probabilidad de que la variable sea menor que un valor negativo.

Adicionalmente la probabilidad de cualquier intervalo (a,b) puede calcularse sin más que restar P(**X**<b) - P(**X**<a), como se vio en el apartado 1.3. de esta Unidad Didáctica (parte 1).

Ejercicio 15: utilizando la tabla de probabilidades de la N(0,1) calcular:

a) La probabilidad de que una variable N(0,1) sea mayor que 2.

b) La probabilidad de que una variable N(0,1) sea mayor que -1.

c) La probabilidad de que una variable N(0,1) esté comprendida entre -1 y 2

1.5.4. Cálculo de probabilidades de una variable N(m,σ)

¿Cómo se calculan probabilidades para las variables normales no tipificadas, que son las que se encuentran en la práctica?. Para ello basta transformar la expresión probabilística de interés, en una equivalente relativa a una variable N(0,1), aprovechando la propiedad 1 expuesta en el apartado 1.5.2.

A esta transformación se le conoce como **tipificar** la variable aleatoria.

Así, siendo **X** una variable normal de media m y desviación típica σ, **N(m,σ)**, se tiene:

$$P(X > z) = P\left(\frac{x-m}{\sigma} > \frac{z-m}{\sigma}\right) = P\left(N(0,1) > \frac{z-m}{\sigma}\right)$$

probabilidad, esta última, que puede buscarse directamente en la tabla.

Ejercicio 16: La dureza de los asientos de poliuretano fabricados en una factoría fluctúa normalmente con media 185 newtons y desviación típica 12 newtons. ¿Qué porcentajes de los asientos fabricados cumplirán las especificaciones establecidas que son de 180 ± 20 newtons?

1.5.5. Aproximaciones normales

Un teorema de gran importancia en la Estadística Matemática postula que, **en condiciones muy generales respecto a las distribuciones de los sumandos, la suma de variables aleatorias independientes tiende a distribuirse normalmente a medida que aumenta el número de sumandos**. A este resultado se le conoce con el nombre de **Teorema Central del Límite**.

Este resultado teórico justifica, en cierto sentido, la frecuencia con la que se presentan en la realidad variables aleatorias cuya distribución se asemeja a la pauta de variabilidad teórica de una distribución normal. En efecto, muchas variables reales pueden considerarse como el resultado de la actuación de un conjunto de factores independientes. Así el rendimiento obtenido en una parcela cultivada depende de las características del suelo, de las condiciones microclimáticas, del grado de incidencia de diversas plagas, del potencial genético de las semillas concretas utilizadas, etcétera... Como consecuencia del teorema central del límite cabe esperar que dicho rendimiento, que en cierto sentido es la suma de una serie de factores independientes, se distribuya aproximadamente de forma normal.

Ejercicio 17: Continuando con los datos del Ejercicio 11, calcula la probabilidad de que el tiempo de acceso a 100 ficheros consecutivos supere los 31 segundos.

Dado que una **variable binomial** no es más que la suma de N variables de Bernuoilli independientes, cabe esperar que su distribución se vaya aproximando a la de una normal a medida que aumente N. En efecto si X es una variable Binomial (B(n,p)) y su varianza np(1-p) es moderadamente grande (valores mayores o iguales que 9 son suficientes para obtener aproximaciones razonables) la variable X tiende a distribuirse aproximadamente como una $N(m=np,\sigma^2=np(1-p))$, por tanto:

$$\text{Si}\ \ X \sim B(n,p)\ \ \text{y}\ \ (\sigma_x^2 = np(1-p)) \geq 9\ \Rightarrow \frac{X-np}{\sqrt{np(1-p)}} \approx N[0,1]$$

pudiendo utilizarse las tablas de la N[0,1] para calcular las probabilidades de X.

Ejercicio 18: Cuando se prueban las tarjetas de circuito que se usan en la fabricación de reproductores de discos compactos, el porcentaje de defectos a largo plazo es 5%. Suponiendo que se recibe un lote de 250 tarjetas, ¿cuál es la probabilidad aproximada de que al menos el 10% de las tarjetas en el lote sean defectuosas?

También una **variable de Poisson** puede aproximarse por una normal si su parámetro λ no es muy pequeño (valores del orden de 9 ó más son recomendables para obtener aproximaciones satisfactorias). Así si X sigue una distribución de Poisson de parámetro λ (Ps(λ)), la variable X tiende a distribuirse aproximadamente como una $N(m=\lambda,\sigma^2=\lambda)$, por tanto:

$$\text{Si}\ \ X \sim Ps(\lambda)\ \ \text{y}\ \ (\sigma_x^2 = \lambda) \geq 9\ \Rightarrow \frac{X-\lambda}{\sqrt{\lambda}} \approx N[0,1]$$

Ejercicio 19: Los códigos CRC utilizados en el envío de paquetes a través de la red son capaces de corregir como máximo 1000 errores en cada bloque de 200 paquetes. Se sabe que el número medio de errores en cada bloque enviado es de 400. Si tomamos al azar un bloque de 200 paquetes, ¿cuál es la probabilidad de que el número total de errores sea inferior a 380?

1.5.6. Papel probabilístico normal (PPN)

El papel probabilístico constituye una herramienta extremadamente práctica de análisis estadístico, utilizándose en el estudio de distintos tipos de distribuciones. En este apartado trataremos sólo del papel probabilístico para distribuciones normales.

Una representación de un conjunto de datos en papel probabilístico hace corresponder un punto a cada observación. La abcisa del punto no es más que el valor observado, mientras que la ordenada corresponde al porcentaje de valores en la muestra que son menores o iguales que el considerado. Los puntos se representan en un papel como el de la figura adjunta, que recoge los datos correspondientes a una variable aleatoria normal X.

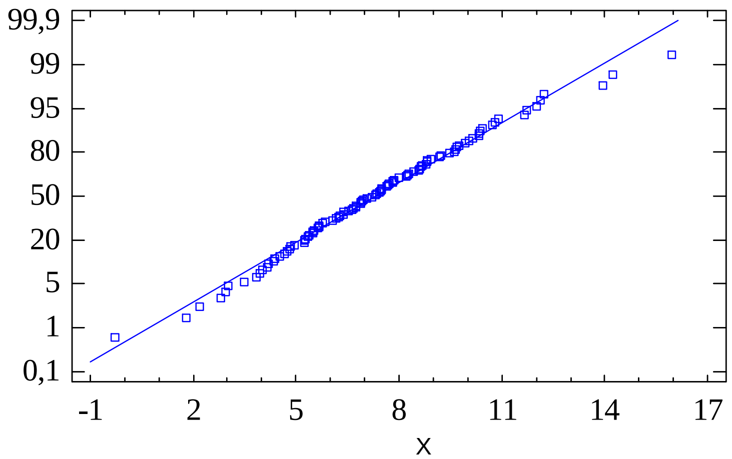


Figura 10. Representación en PPN de los datos de una v.a. X~N(m,σ)

Tal como se aprecia en la **Figura 10**, la escala vertical del papel probabilístico está modificada de forma que corresponde a las probabilidades acumuladas de una variable aleatoria normal tipificada.

La idea básica para la utilización del papel probabilístico es la siguiente: **cuando datos procedentes de una distribución normal se representan en este papel, los puntos correspondientes se sitúan aproximadamente a lo largo de una recta**.

Importante: El **PPN** no se trata de un test de inferencia estadística, es una herramienta descriptiva que permite ver si el **modelo normal** se ajusta lo suficientemente bien a la realidad de los datos observados (muestra) como para utilizarlo.

Si la representación de los datos difiere claramente de una línea recta ello es una prueba de que la población muestreada no se distribuye normalmente.

En las siguientes figuras vemos aspectos típicos de representaciones correspondientes a datos no normales.

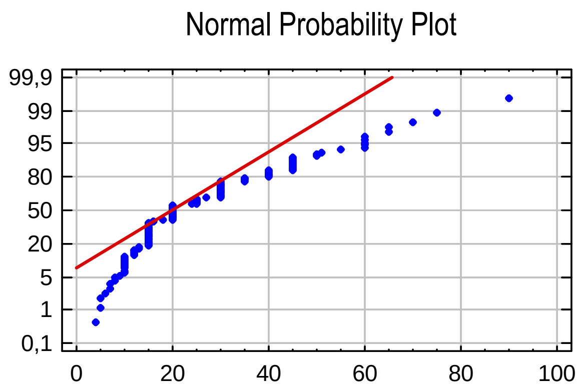


Figura 11. Representación en PPN de datos procedentes de una distribución asimétrica positiva

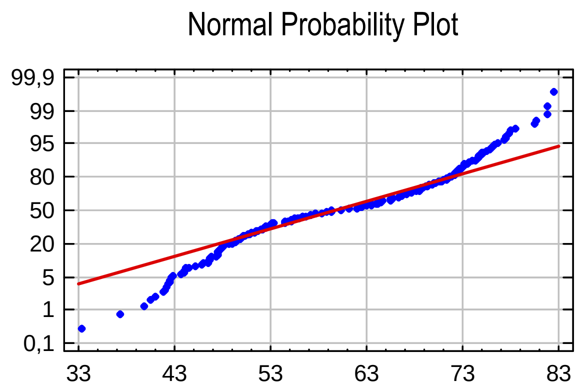


Figura 12. Representación en PPN de datos procedentes de dos poblaciones con distinta **m**

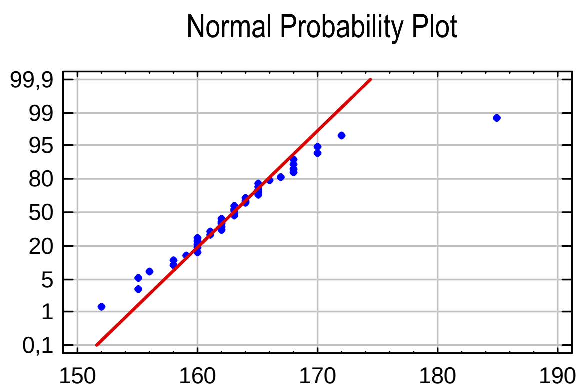


Figura 13. Representación en PPN de datos con un dato anómalo

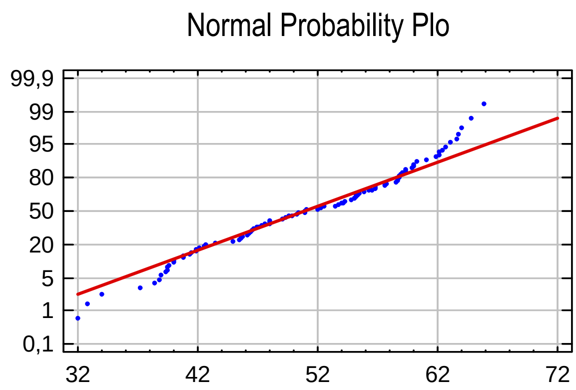


Figura 14. Representación en PPN de datos procedentes de una distribución normal

Ejercicio 20: En un estudio sobre la resistencia a la torsión de un tipo de tornillo fabricado por una cierta factoría se ha tomado una muestra de 10 tornillos de una partida. La resistencia observada (medida en Nw) de los tornillos de la muestra ha sido: 14,1 13,1 12,5 15,6 14,6 15 17 13,5 16,3 17

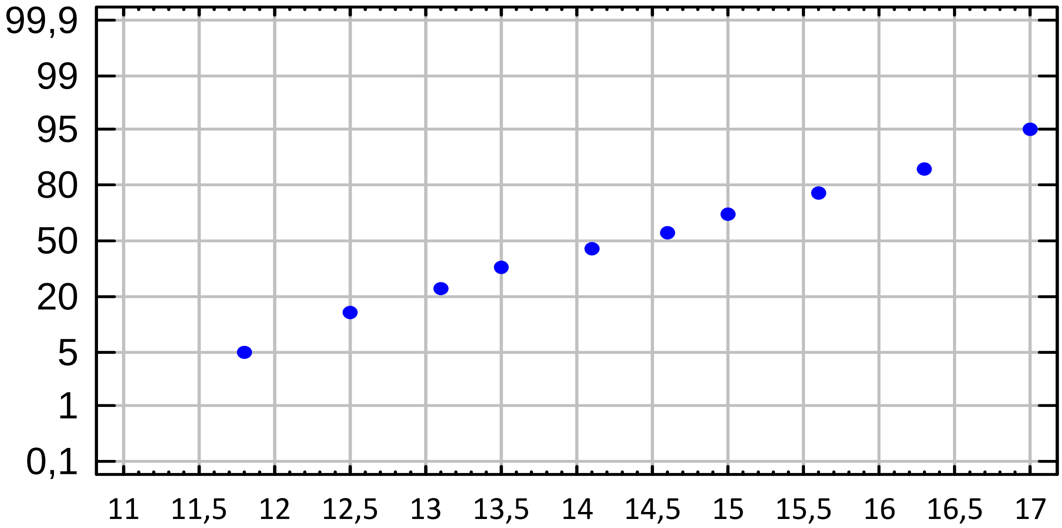
Estos datos se han representado en PPN (**Figura 15**).

A la vista del gráfico, se pide:

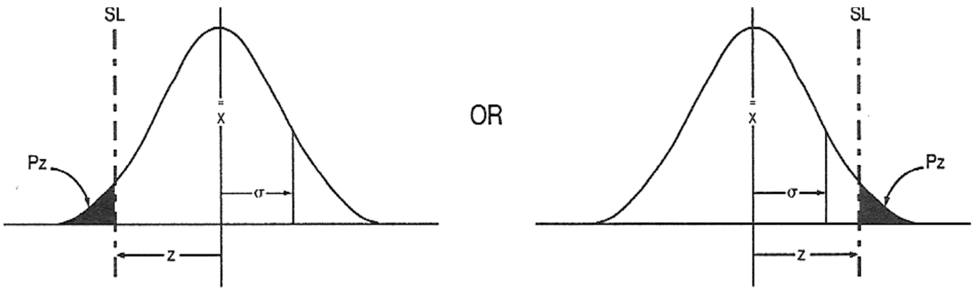
a) ¿Es admisible la hipótesis de que los datos proceden de una distribución normal?. Justificar.

b) Calcular, aproximadamente, sobre el PPN la media y desviación típica de la distribución.

Figura 15. Representación en PPN de la resistencia a la torsión de una muestra de 10 tornillos



Distribución Normal



|z|	x.x0	x.x1	x.x2	x.x3	x.x4	x.x5	x.x6	x.x7	x.x8	x.x9
4.0	.00003									
3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Ejercicios resueltos

Apartado 5.A.1 del Capítulo 5 del libro de R. Romero y L.R. Zúnica "Métodos Estadísticos en Ingeniería" SPUPV 637

Apartado 6.A.1 del Capítulo 6 del libro de R. Romero y L.R. Zúnica "Métodos Estadísticos en Ingeniería" SPUPV 637

Apartados 7.A.1 y 7.A.2 del Capítulo 7 del libro de R. Romero y L.R. Zúnica "Métodos Estadísticos en Ingeniería" SPUPV 637

Ver boletín correspondiente en PoliformaT (EST GII: Recursos / 04 | Ejercicios)

Para saber más

- Descartes. Instituto de Tecnologías Educativas (ITE) del Ministerio de Educación: http://recursostic.educacion.es/descartes/web/materiales_didacticos/distribuciones_probabilidad/dis_normal.htm
- VESTAC. Internet Scout Project. http://www.kuleuven.ac.be/ucs/java/gent/Ap4a.html
- Distribución Normal y propiedades. (*Balasbramanian Narasimhan; Stanford University, Department of Statistics*): http://www-stat.stanford.edu/~naras/jsm/NormalDensity/NormalDensity.html
- El tablero de Galton (Galton's board) es un experimento para ilustrar el comportamiento *casi gaussiano* de una distribución binomial con $p = 1/2$: http://goo.gl/g78V
- The Gauss distribution: suma de n variables aleatorias uniformemente distribuidas (*Franz Embacher and Petra Oberhuemer;Maths Online*): http://www.univie.ac.at/future.media/moe/galerie/wstat1/wstat1.html
- Distribución Binomial y aproximación normal. (David Lane, Rice University, Texas (USA)): http://www.ruf.rice.edu/~lane/stat_sim/normal_approx/index.html
- Otras distribuciones discretas que puede ser interesante conocer: http://goo.gl/Ft6U

Fuentes

Métodos Estadísticos en Ingeniería (Romero Villafranca, Rafael)