CAPÍTULO 6
LA DISTRIBUCIÓN NORMAL

6.1 INTRODUCCIÓN

Frecuentemente se constata en variables aleatorias continuas, 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, dando lugar a histogramas cuya forma recuerda la de una campana.

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** (Otras denominaciones utilizadas son las de distribución de Laplace o distribución de Gauss, que hacen referencia a los apellidos de los dos famosos astrónomos y matemáticos que utilizaron dicha distribución para el estudio de los errores en sus observaciones).

En el presente capítulo se define la distribución normal y se exponen sus propiedades más importantes, haciéndose hincapié en el manejo de tablas y en la utilización de papel probabilístico para representar datos normales.

También se destina un apartado a introducir sucintamente el concepto de la distribución **Lognormal**.

Las posibilidades de aplicación de estas distribuciones a problemas reales se ilustran con los diferentes ejercicios que se proponen, cuya resolución no siempre es trivial.

6.2 DISTRIBUCIÓN NORMAL: CONCEPTO Y PROPIEDADES

Como sabemos toda variable aleatoria continua viene caracterizada por su función de densidad f(x), que indica la densidad de probabilidad asociada a cada valor posible x.

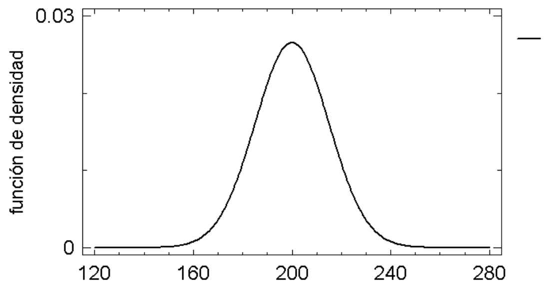
Como ya hemos indicado en la Introducción, la distribución de probabilidad continua más frecuentemente utilizada en las aplicaciones prácticas es la **distribución normal**, que se caracteriza porque su función de densidad viene dada por la expresión:

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

donde m = media de la distribución

σ = desviación típica de la distribución.

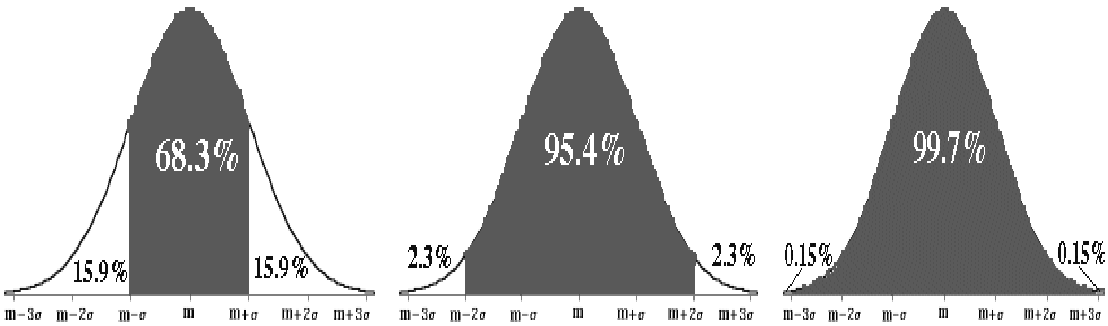
Dicha función de densidad f(x) tiene forma de una curva en campana, con una densidad máxima en m, que es la media y la mediana de la distribución. La densidad decrece de forma simétrica a ambos lados de m, de forma más o menos rápida en función del valor que tenga la desviación típica σ (puede comprobarse que σ coincide con la distancia desde la media m al punto de inflexión de la función de densidad). Los dos parámetros m y σ caracterizan por completo la distribución de una variable normal.



Toda distribución normal tiene coeficientes de asimetría y de curtosis nulos.

Según se aprecia en la figura siguiente, si una variable aleatoria **X** se distribuye normalmente con media m y desviación típica σ se verifica:

$$P(m-\sigma < \mathbf{X} < m+\sigma) \approx 0.68$$
$$P(m-2\sigma < \mathbf{X} < m+2\sigma) \approx 0.95$$
$$P(m-3\sigma < \mathbf{X} < m+3\sigma) \approx 0.997$$



o sea que es bastante raro (probabilidad menor que el 5%) encontrar un valor de una variable normal que difiera de su media en más de dos desviaciones típicas, y es muy improbable (probabilidad menor que el 3 por mil) encontrar un valor que difiera de la media en más de tres desviaciones típicas.

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.

Nota técnica: el requisito de independencia exigido en la propiedad 2 es realmente una condición suficiente pero no necesaria. La condición necesaria y suficiente para que una combinación lineal de dos variables normales se distribuya normalmente es que dichas variables tengan una distribución conjunta normal bidimensional (concepto que no se aborda en este texto), siendo la independencia una condición suficiente para que esto se verifique

6.3 TABLA DE LA DISTRIBUCIÓN NORMAL

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

Autoevaluación: Calcular la probabilidad de que una variable N[0,1] sea mayor que 2. Comparar el resultado obtenido con lo indicado en la segunda de las figuras anteriores.

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).

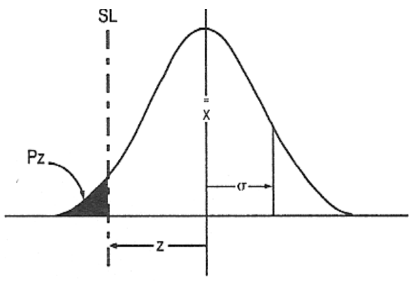
Autoevaluación: Calcular la probabilidad de que una variable N[0,1] tome valores comprendidos entre -1.5 y 2.5

¿Cómo se calculan probabilidades para las variables normales no tipificadas, que son las que se encuentran en la práctica? Basta para ello transformar la expresión probabilística de interés, en una equivalente relativa a una variable N[0,1], aprovechando la Propiedad 1 expuesta en el apartado anterior.

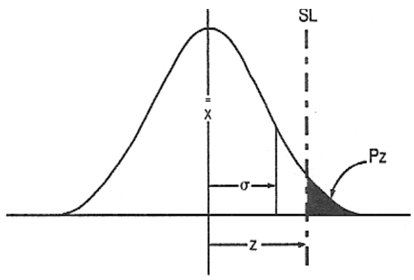
Así, siendo **X** una variable normal de media m y desviación típica σ, N[m,σ], se tiene:

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

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



OR



|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

Autoevaluación: 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?
Si se centrase correctamente el proceso de forma que su media resultara 180, por ejemplo utilizando el valor adecuado de la cantidad de poliol e isocianato, ¿en cuánto se reduciría el porcentaje de asientos defectuosos obtenidos?

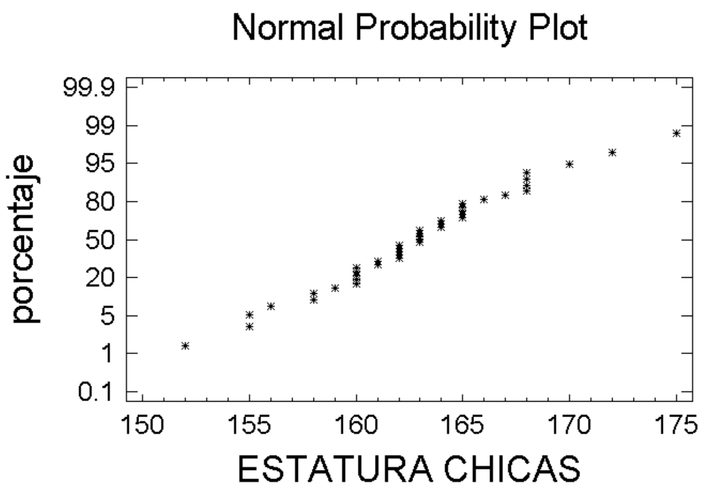
Autoevaluación: El peso de las naranjas de un determinado calibre fluctúa normalmente con media=150 grs. y σ=30 grs. Una bolsa se llena con 15 naranjas seleccionadas al azar.
a) ¿Cuál es la probabilidad de que la mayor naranja de la bolsa pese más de 200 grs.?
b) ¿Cuál es la probabilidad de que la menor naranja de la bolsa pese menos de 120 grs.?
c) ¿Cuál es la probabilidad de que el peso total de la bolsa sea inferior a 2 kgs?
(Ver respuesta en el Anejo al final del Tema)

6.4 PAPEL PROBABILÍSTICO

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.

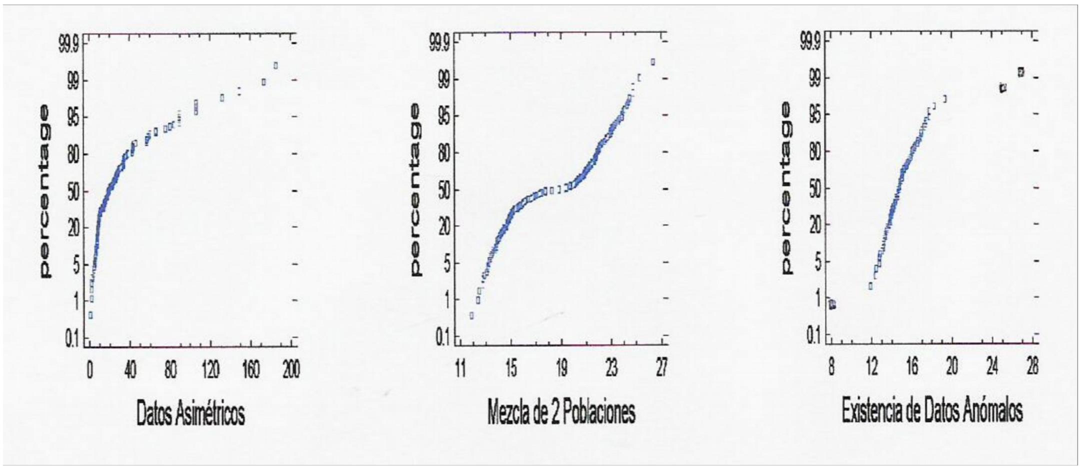
Una representación de un conjunto de datos en papel probabilístico normal hace corresponder a cada observación un punto en el plano. La abscisa del punto no es más que el valor observado, mientras que la ordenada corresponde al porcentaje de valores en la muestra que son menores o iguales que el considerado. (En general se aplica una corrección de continuidad, de forma que en una muestra de tamaño N a la observación i-ésima, una vez ordenadas de menor a mayor, le corresponde como ordenada 100(i-0.5)/N)

La escala vertical del papel probabilístico está modificada -tal como se aprecia en la siguiente figura que corresponde a las estaturas de las chicas en la encuesta curs8990.sf3- de forma que corresponde a los valores de la función de distribución de una normal tipificada. Así, por ejemplo, el intervalo entre los puntos correspondientes al 50% y al 80% es muy parecido al existente entre este último y el del 95%, dado que los valores correspondientes a dichas probabilidades acumuladas en una N[0,1] están aproximadamente equiespaciados (concretamente son 0, 0.84 y 1.65)



La idea básica para la utilización del papel probabilístico normal 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.

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 la siguiente figura vemos aspectos típicos de representaciones correspondientes a diferentes tipos de datos no normales.



Como se aprecia los datos de distribuciones asimétricas positivas presentan claramente una curvatura negativa cuando se representan en papel probabilístico normal. La mezcla de dos poblaciones con diferentes medias se detecta por la existencia de dos tramos de crecimiento rápido separados por uno de crecimiento lento. Por último los valores anormalmente altos o bajos se representan por puntos que se separan de la recta en la zona superior derecha e inferior izquierda respectivamente.

Autoevaluación: construirse un impreso aproximado para papel probabilístico normal, poniendo equiespaciados en la escala vertical los valores 0.1 1 5 20 50 80 95 99 y 99.9. Representar en el mismo los siguientes datos de una muestra de 8 observaciones: 246 301 186 200 182 385 292 215. ¿Es razonable asumir que esta muestra procede de una población normal?

Autoevaluación: Obtener una representación en papel probabilístico normal de los datos de PESO en los chicos. ¿Se distribuye el peso de los chicos de forma aproximadamente normal? ¿Cómo podría estimarse aproximadamente la estatura media de los chicos a partir de la representación anterior? ¿Cómo podría estimarse aproximadamente la desviación típica de la distribución a partir de la representación considerada?

Autoevaluación: Obtener a partir de los datos de la encuesta un diagrama en papel probabilístico normal de la variable TIEMPO (tiempo en minutos que se tarda en venir a la Universidad). ¿Se distribuye la variable TIEMPO de forma aproximadamente normal o, por el contrario, su distribución es muy asimétrica? Obtener una representación en papel normal del logaritmo de la variable TIEMPO. ¿Se distribuye el logaritmo de TIEMPO de forma aproximadamente normal?

6.5 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**.

Nota técnica: la condición necesaria y suficiente para que se verifique dicho teorema es complicada. Intuitivamente lo que exige es que ninguno de los sumandos predomine claramente sobre el resto. Una sencilla condición suficiente al respecto es que todas las variables que se suman tengan la misma distribución (Teorema de Linderbeg-Levy)

*Autoevaluación: Representar la función de probabilidad de la variable aleatoria **X** : {número de puntos al lanzar un dado}. Constatar que tiene una forma rectangular, muy diferente a una campana. Representar ahora la función de probabilidad de la variable aleatoria: {suma de puntos al lanzar 2 dados} Constatar que ahora las probabilidades empiezan a acumularse de forma simétrica en torno a un valor central, como sucede en una variable normal.*
Una sugerente visualización de la convergencia a la normalidad de la suma de variables uniformes [0,1] independientes, puede encontrarse en la dirección de Internet http://www.univie.ac.at/future.media/moe/galerie/wstat1/wstat1.html

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.

Autoevaluación: Calcular aproximadamente la probabilidad de sacar más de 80 puntos al lanzar 20 dados (Ver respuesta en el Anejo al final del Tema).

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 (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 tipificada Y = $\frac{X-Np}{\sqrt{Np(1-p)}}$ tiende a distribuirse aproximadamente como una N[0,1], pudiendo utilizarse las tablas de ésta para calcular las probabilidades de la primera.

Nota técnica: al aproximar una variable Binomial por una Normal (o, en general, siempre que se aproxime una variable discreta con valores enteros por una continua) es aconsejable, para mejorar la calidad de la aproximación, realizar una corrección de continuidad. Esta corrección consiste en asimilar cada valor entero x de la variable discreta, al intervalo [x-0.5 x+0.5] para la variable continua

Autoevaluación: Suponiendo que la probabilidad de elegir un dígito impar es la misma que la de un dígito par, calcular aproximadamente la probabilidad de que de 131 dígitos elegidos al azar más de 87 sean impares. (Ver respuesta en el Anejo al final del Tema)

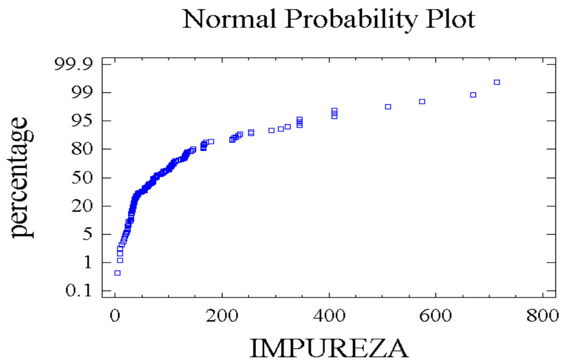
También una variable de Poisson puede aproximarse por una normal si su parámetro λ no es muy pequeño (valores del orden de 9 ó más son recomendables para obtener aproximaciones satisfactorias). Así si X sigue una distribución de Poisson de parámetro λ, la variable tipificada Y = $\frac{X-\lambda}{\sqrt{\lambda}}$ sigue aproximadamente una distribución N[0,1], si λ es suficientemente grande.

<u>Nota</u>: dado que se dispone de un ábaco que permite fácilmente calcular probabilidades de variables de Poisson de parámetro $\lambda \leq 30$, la aproximación normal se utilizará para valores de λ mayores que 30.

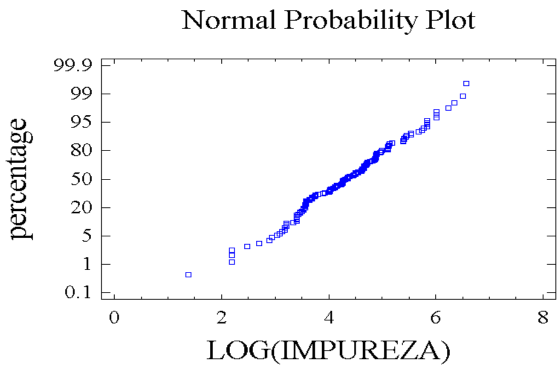
Autoevaluación: A una centralita telefónica llegan en promedio 2 llamadas por minuto. Calcular aproximadamente la probabilidad de que en una hora se reciban más de 150 llamadas. (Ver respuesta en el Anejo al final del Tema)

6.6 LA DISTRIBUCIÓN LOGNORMAL

En la siguiente figura se recoge la representación en papel probabilístico normal del contenido en impurezas (en partes por millón) en 139 lotes de un proceso químico. Se aprecia claramente una fuerte asimetría positiva que hace que los datos no se ajusten al modelo normal



En estos casos de fuerte asimetría positiva, frecuentemente el logaritmo de los datos se ajusta bien a una distribución normal, tal como se constata en la siguiente figura en la que se recoge la representación en papel probabilístico normal del logaritmo (neperiano) del contenido en impurezas

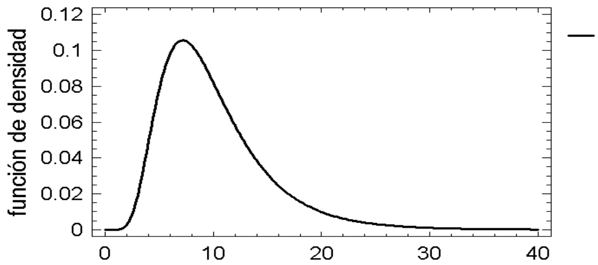


Una variable **X** sigue una distribución logarítmico normal (LogNormal) de parámetros μ y σ^2, si la variable **Y = log(X)** (logaritmo neperiano de **X**) sigue una distribución Normal(μ,σ^2)

X ~ LogNormal(μ,σ^2) <===> **Y**=log(**X**) ~ Normal(μ,σ^2)

<u>Nota</u>: el símbolo ~ es equivalente a la expresión "se distribuye como"

Toda distribución logarítmico normal es no negativa y presenta asimetría positiva, tal como se refleja en la siguiente figura, constituyendo frecuentemente un modelo adecuado para variables asimétricas positivas (como, por ejemplo, el tiempo hasta el fallo de equipos o piezas, el contenido en impurezas resultante de un proceso, el descentrado en operaciones de mecanizado, etcétera...)



Si μ es la media de la distribución del log(X), la media de X tiene una expresión algo complicada, pero la **mediana** de X (que en casos de una distribución asimétrica es un indicador de posición más adecuado que la mediana) viene dada simplemente por e^{μ}
si μ = media de log(**X**) y log(**X**) es Normal ➔ mediana(**X**) = e^{μ}

Autoevaluación: justificar la afirmación anterior

Autoevaluación: La superficie X de las explotaciones hortofrutícolas en una región sigue una distribución LogNormal. Se sabe que la mediana de X es 15 hanegadas y que sólo el 1% de las explotaciones son mayores de 50 hanegadas. ¿Qué porcentaje de las explotaciones serán menores de 5 hanegadas? (Ver respuesta en el Anejo al final del Tema)

6.A AUTOEVALUACIONES RESUELTAS Y EJERCICIOS

6.A.1 Respuesta a algunas Autoevaluaciones

Autoevaluación: El peso de las naranjas de un determinado calibre fluctúa normalmente con media=150 grs. y σ=30 grs. Una bolsa se llena con 15 naranjas seleccionadas al azar.
a) ¿Cuál es la probabilidad de que la mayor naranja de la bolsa pese más de 200 grs?
b) ¿Cuál es la probabilidad de que la menor naranja de la bolsa pese menos de 120 grs?
c) ¿Cuál es la probabilidad de que el peso total de la bolsa sea inferior a 2 kgs?

a) La probabilidad de que una naranja seleccionada al azar pese más de 200 grs. será

$$P(x>200) = P\left(\frac{x-150}{30} > \frac{200-150}{30}\right) = P(N(0,1) > 1.67) = 0.0475$$

P(mayor > 200) = 1 – P(mayor<200) = 1 – P(todas pesen < 200) = 1 – $(1\text{-}0.0475)^{15}$ = 0.518

b) La probabilidad de que una naranja seleccionada al azar pese menos de 120 grs. será

$$P(X<120) = P\left(\frac{X-150}{30} < \frac{120-150}{30}\right) = P(N(0,1) < -1) = 0.1587$$

P(menor < 120) = 1 – P(menor>120) = 1 – P(todas pesen > 200) = 1 – $(1\text{-}0.1587)^{15}$ = 0.925

c) Siendo X_i el peso de la naranja i-ésima (i=1...15) e Y el peso neto total de la bolsa:

$Y = X_1 + ... + X_{15} \Rightarrow Y \sim$ Normal con $m_Y = 150 +...+ 150 = 2250$ y $\sigma^2_Y = 30^2 +...+ 30^2 = 13500$
(la segunda propiedad se cumple por ser los pesos X_i independientes)

$$P(Y < 200) = P\left(\frac{Y-2250}{\sqrt{13500}} < \frac{2000-2250}{\sqrt{13500}}\right) = P(N(0,1) < -2.15) = 0.0158$$

Autoevaluación: Calcular aproximadamente la probabilidad de sacar más de 80 puntos al lanzar 20 dados

.
Siendo X_i el número de puntos al lanzar el dado i, se tiene:

$E(X_i) = 1(1/6) + 2(1/6) + ... + 6(1/6) = 3.5$
$\sigma^2(X_i) = (1-3.5)^2(1/6) + (2-3.5)^2(1/6) +...+(6-3.5)^2(1/6) = 2.917$

Total puntos $Y = X_1 + ... + X_{20} \Rightarrow Y \approx$ Normal $m_Y = 20x3.5 = 70$ y $\sigma_Y = \sqrt{20 \times 2.917} = 7.64$

$$P(Y>80) \approx P(Z\sim N(70,7.64)>80.5) = P\left(\frac{Z-70}{7.64} > \frac{80.5-70}{7.64}\right) = P(N(0,1) > 1.37) = 0.0853$$

Autoevaluación: Suponiendo que la probabilidad de elegir un dígito impar es la misma que la de un dígito par, calcular aproximadamente la probabilidad de que de 131 dígitos elegidos al azar más de 87 sean impares.

X: número de dígitos impares al elegir 131, suponiendo p(impar) = p(par) = 0.5 será un binomial con N=131 y p=0.5. Por ser Np(1-p) = 131x0.5x(1-0.5) = 32.75 >> 9, se podrá aproximar por una normal con su misma media (131x0.5=65.5) y desviación típica $\sqrt{32.75}$ =5.72

$$P(X>87) \approx P(Z\sim N(65.5,72)>87.5) = P\left(\frac{Z-65.5}{5.72} > \frac{87.5-65.5}{5.72}\right) = P(N(0,1) > 3.85) = 0.00006$$

Autoevaluación: A una centralita telefónica llegan en promedio 2 llamadas por minuto. Calcular aproximadamente la probabilidad de que en una hora se reciban más de 150 llamadas.

Siendo X_i las llamadas en el minuto i $\Rightarrow$ Y: número de llamadas en una hora = $X_1 +...+ X_{60}$. Suponiendo que las X_i son independientes (criticar esta hipótesis) Y $\sim$ Poisson(λ=2+...+2=120). Como λ >> 9 Y se puede aproximar por una normal con su misma media (120) y desviación típica $\sqrt{120}$ = 10.95
$P(Y>150) \approx P(Z\sim N(120,10.95)>150.5) =$

$$P\left(\frac{Z-120}{10.95} > \frac{150.5-120}{10.95}\right) = P(N(0,1) > 2.79) = 0.0026$$

Autoevaluación: La superficie X de las explotaciones hortofrutícolas en una región sigue una distribución LogNormal. Se sabe que la mediana de X es 15 hanegadas y que sólo el 1% de las explotaciones son mayores de 50 hanegadas. ¿Qué porcentaje de las explotaciones serán menores de 5 hanegadas?

X: superficie explotaciones $\sim$ LogNormal $\Rightarrow$ Y = log(X) $\sim$ Normal(m_Y ,σ_Y)

P(X>15) = 0.5 $\Rightarrow$ P(Y > log(15)=2.708) = 0.5 $\Rightarrow$ m_Y = 2.71

P(X>50) = 0.01 $\Rightarrow$ P(Y > log(50)=3.912) = 0.01 $\Rightarrow$ $\frac{3.912-2.708}{\sigma_Y}$ habrá de ser = 2.33
(mirando en la tabla de la Normal tipificada el valor que corresponde a una probabilidad 0.01)

$\sigma_Y = \frac{3.912-2.708}{2.33} = 0.517$

P(X<5) = P(Y < log(5)=1.609) = $P\left(N(0,1) < \frac{1.609-2.708}{0.517} = -2.13\right)$ = 0.0166

(el 1.66% de las explotaciones serán inferiores a 5 hanegadas)

6.A.2 Ejercicios resueltos

El peso de las naranjas de cierto calibre se distribuye normalmente con media 150 grs y desviación típica 25 grs. Estas naranjas se expiden en bolsas de malla que se llenan seleccionando al azar las naranjas
¿Cuál es el número mínimo de naranjas a meter en cada bolsa si se desea tener una probabilidad mayor que 0.99 de que el peso total sea superior a 2 kgs

Y: peso total = X_1 +...+ X_N $\sim$ Normal con m_Y = 150N y $\sigma_Y = \sqrt{25^2 N} = 25\sqrt{N}$

Para que P(Y>2000) = $P\left(N(0,1) > \frac{2000-150N}{25\sqrt{N}}\right)$ sea > 0.99 $\Rightarrow$ $\frac{2000-150N}{25\sqrt{N}}$ ha de ser < -2.33

El menor valor de N que verifica $\frac{2000-150N}{25\sqrt{N}}$ < -2.33 es <u>N = 15</u>

*Una partida defectuosa de tornillos tiene una resistencia a la torsión **X** que fluctúa normalmente con media 25 nwt y σ = 4 nwt. Los tornillos se utilizan para cerrar las cajas que contiene ciertos módulos electrónicos y se atornillan mediante una atornilladora que produce un par de apriete **Y** que fluctúa normalmente con una media de 12 nwt y una σ = 3 nwt.*
¿Qué porcentaje de los tornillos se partirán al apretarse?

P(se parta el tornillo) = P(Y > X) = P((Y-X) > 0)
Z = (Y-X) será Normal, por ser combinación lineal de dos variables normales independientes (justificar la independencia), con $m_Z = m_Y - m_X = 12 - 25 = -13$ y $\sigma_Y = \sqrt{4^2+3^2} = 5$

P(parta tornillo) = P(Z>0) = $P\left(N(0,1) > \frac{0-(-13)}{5} = 2.6\right)$ = 0.0047
(se partirán el 4.7 por mil de los tornillos)

6.A.3 Ejercicios adicionales

El tiempo de lectura de un dato en una unidad de disco se distribuye exponencialmente con media 2 milisegundos. En un determinado proceso deben leerse consecutivamente 400 datos situados aleatoriamente en el disco. Calcular la probabilidad de que el tiempo total de lectura de los datos sea mayor que 1 segundo.

El peso neto de los botes de un determinado tipo de conserva fluctúa normalmente siendo los dos cuartiles de la distribución iguales a 480 y 530 gramos respectivamente. Calcular la media y la desviación típica de dicha distribución

El diámetro Φ de las naranjas de una partida que llega a un almacén se distribuye normalmente. En un calibrado previo se constata que un 10% de las naranjas tienen $\Phi<40$ mm y que una 20% tienen $\Phi>70$ mm. Calcular el % de naranjas cuyo Φ estará comprendido entre 50 y 60 mm.

La carga de seguridad del ascensor de un hotel es 1000 kgs. El peso de las personas que suben al mismo fluctúa normalmente con media 65 kgs y desviación típica 10 kgs. ¿Cuál es el número máximo de personas que pueden subir a la vez, si se desea que la probabilidad de rebasar la carga de seguridad sea inferior a 0.001?

En promedio el 20% de los asientos fabricados en una factoría deben repararse para solucionar problemas ocasionados por arrugas en la tela del forro. Si un día se fabrican 400 asientos ¿cuál es la probabilidad de que haya que reparar más de 100 asientos? (Aproximar la binomial por una normal, haciendo la correspondiente corrección de continuidad)
El diámetro Φ_a de ciertos agujeros que se realizan en unas planchas fluctúa normalmente con media 20 mm y desviación típica 1 mm. En dicho agujeros se han de fijar unos casquillos de caucho fabricados por un proveedor, cuyo diámetro Φ_c fluctúa normalmente con media 20.5 mm y desviación típica 1 mm. Se sabe que si $\Phi_a - \Phi_c$ es mayor que 1 mm el casquillo se cae, mientras que si $\Phi_c - \Phi_a$ es mayor que 1.5 mm el casquillo no entra.
Calcular el % de veces que se presentarán problemas al montar los casquillos.
¿Cuál debería ser el valor medio de Φ_a para minimizar dicho %, suponiendo que no es posible modificar Φ_c?

Para venir en coche a la universidad un estudiante puede elegir entre dos caminos alternativos. El tiempo que tarda por el camino A fluctúa aleatoriamente de unos días a otros, distribuyéndose como una variable normal de media=25' y σ=7', mientras que el tiempo que tarda por el camino B se distribuye normalmente con media=30' y σ=3'.
a) ¿Qué camino debe elegir si desea salir lo más tarde posible de su casa pero teniendo la garantía de llegar a tiempo al menos el 95% de los días? (Razonar la respuesta calculando con cuantos minutos de antelación debería salir según el camino que escoja)
b) Si las clases comienzan a las 9 de la mañana calcular el camino que se debe escoger en función de la hora a la que se salga de casa, con el fin de maximizar la probabilidad de llegar a tiempo

*El deterioro de una pieza a lo largo del tiempo viene provocado por un agrietado de la misma. El tamaño **X** (en micras) de dicha grieta aumenta con el tiempo **T** de funcionamiento (en miles de horas) según la expresión $\boldsymbol{X = \alpha T}$, donde $\boldsymbol{\alpha}$ fluctúa aleatoriamente en la población de las piezas siguiendo una distribución LogNormal de parámetros $\mu = 0.8$ y $\sigma^2 = 0.02$.*
Hallar la vida mediana de las piezas, sabiendo que el fallo se produce cuando el tamaño de la grieta alcanza las 2.5 micras
¿Qué porcentaje de las piezas durarán más de 1000 horas?