TEMA 8

REPASO DISTRIBUCIÓN NORMAL

Contenidos

1. Variables normales
2. La campana de Gauss
3. Probabilidades en la normal típica
4. Cálculos en distribuciones normales: tipificación
5. Intervalos característicos centrales

Tiempo estimado

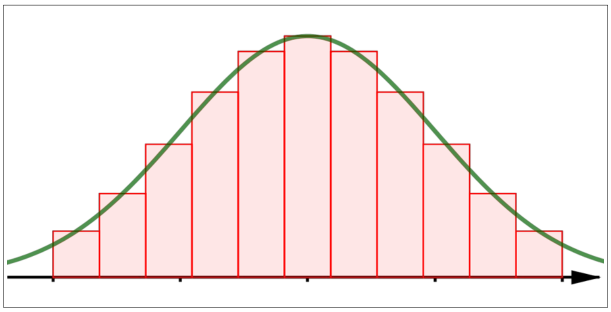
5 sesiones

Criterios de Evaluación

1. Conoce los aspectos básicos de las curvas normales e identifica sus parámetros básicos.
2. Sabe obtener probabilidades elementales usando la tabla de la normal típica.
3. Calcula probabilidades en problemas donde la variable es normal con media y desviación típica conocidas.

1.Variables normales

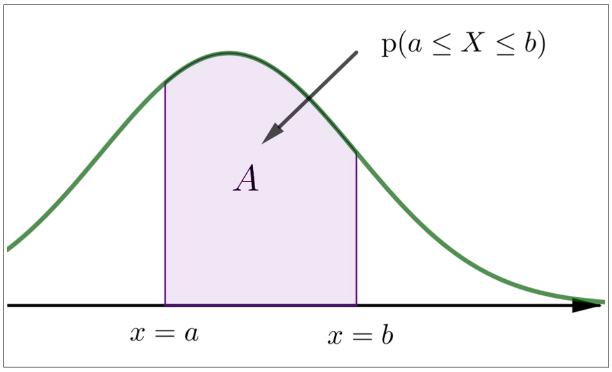
El matemático Gauss descubrió que una gran cantidad de variables estadísticas X tenían una características semejantes que podían estudiarse a partir de ciertas curvas exponenciales llamadas luego, en su honor, campanas de Gauss.



Dichas variables se denominan normales y se dice también que sus datos siguen una distribución normal.

La característica fundamental de las variables normales es la siguiente: la probabilidad de encontrar un individuo en el que la variable normal X tome un valor comprendido entre a y b viene dada por el área bajo la curva de Gauss en dicho intervalo:

Nota: una función cuya gráfica proporciona la probabilidad como antes se denomina **función de densidad**



Prácticamente toda variable que sea suma de una serie elevada de características sigue una distribución normal. Aunque una variable sea consecuencia de una serie de efectos y éstos no se distribuyan de forma normal, la concurrencia de todos ellos hace que la variable en cuestión sí se comporte como una distribución normal: aproximación de los valores a la media, simetría de éstos, escasos valores alejados de la media, etc.

Hay muchísimas variables que se distribuyen siguiendo el modelo normal:

a) Variables morfológicas de individuos como son la talla o el peso de las personas, la altura de los árboles de cierta especie, ...

b) Variables fisiológicas, como son la presión sanguínea, la frecuencia cardíaca, ...

c) Variables de tipo sociológico como son el consumo de ciertos productos por personas de un mismo colectivo.

d) Variables físicas, como la resistencia a la rotura de piezas.

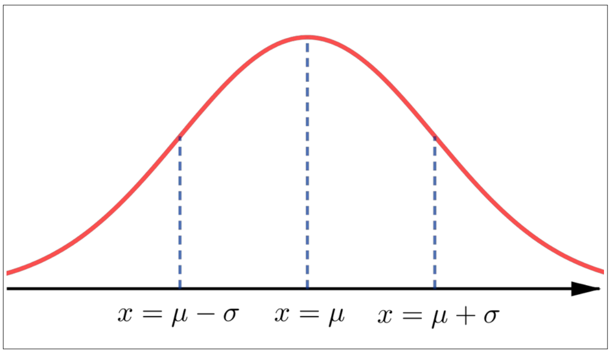
2.La campana de Gauss

Para cada media media μ y cada desviación típica σ hay una curva normal que se designa por $\mathcal{N}\,(\mu\,,\sigma)$.

Su ecuación es

$$y = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma}}$$

Se trata de una función continua, simétrica, cuyo máximo se alcanza para $x=\mu$ y que tiene dos puntos de inflexión situados a distancia σ de dicha media:



Una característica fundamental de todas ellas es que el área bajo cualquier curva normal es 1.

La campana de Gauss, también llamada curva normal, es la gráfica de una importante función que tiene muchas aplicaciones en diversas ramas de las matemáticas, la física, ingenierías, las ciencias sociales,...

Debe su nombre al insigne matemático Carl Friedrich Gauss (1777-1855)

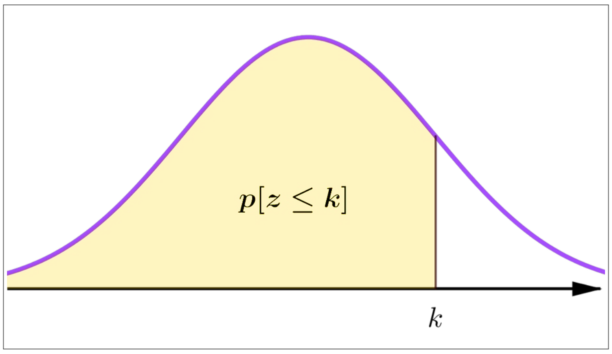
3.Probabilidades en la normal típica

Todas las curvas normales son esenciales idénticas, sólo se distinguen en un cambio de eje de simetría y un cambio de escala. Esto permite calcular cualquier área a partir de una de ellas.

Se ha tomado como base de cálculo a la que tiene $\mu=0$ y $\sigma=1$, denominándose normal típica y se reserva para ella la letra z:

$$Z \sim \mathcal{N}\,(0\,,1)$$

Vamos a calcular áreas (que son probabilidades) en la normal típica.



Su cálculo es complejo, pues el recinto está limitado por una curva. Hay programas que nos permiten calcular cualquier área bajo una curva normal cualquiera, pero para exámenes oficiales, sólo dispondremos de la tabla que nos proporciona las áreas limitadas por la curva $\mathcal{N}\,(0\,,1)$ desde $-\infty$ hasta un número $k>0$.

Es por ello que vamos a aprender, primero, a calcular cualquier tipo de área bajo la curva normal típica usando su tabla.

Llamaremos

$$\phi(k) = p\,(z \leq k)$$

☞ Ejemplo: $p\,(z \leq 2.35)$

Como $2.35 = 2.3 + 0.05$ la intersección de la fila 2.3 con la columna 0.05 nos da el resultado buscado:

$$p\,(z \leq 2.35) = \phi(2.35) = 0.9906$$

☞ Ejemplo: $p\,(z \geq 0.78)$

Teniendo en cuenta que el área encerrada por la curva es 1, tenemos:

$$p\,(z \geq 0.78) = 1 - p\,(z \leq 0.78) = 1 - \phi(0.78) = 1 - 0.7823 = 0.2177$$

☞ Ejemplo: $p\,(z \leq -2.18)$

Teniendo en cuenta la simetría de la curva:

$$p\,(z \leq -2.18) = p\,(z \geq 2.18) = 1 - \phi(2.18) = 1 - 0.9854 = 0.0146$$

☞ Ejemplo: $p\,(z \geq -1.5)$

De nuevo, teniendo en cuenta la simetría de la curva:

$$p\,(z \geq -1.5) = p\,(z \leq 1.5) = \phi(1.5) = 0.9332$$

Tenemos así estos tres casos:

Directo

$$p(z \geq -k) = p(z \leq k) = \phi(k)$$

Contrario

$$p(z \leq -k) = p(z \geq k) = 1 - \phi(k)$$

Diferencia

$$p(a \leq z \leq b) = p(z \leq b) - p(z \leq a)$$

☞ Ejemplo: $p\,(1.55 \leq z \leq 2.48)$

Basta expresar como diferencia de áreas:

$$p\,(1.55 \leq z \leq 2.48) = \phi(2.48) - \phi(1.55) = 0.9934 - 0.9394 = 0.054$$

También puede plantearse el problema inverso: obtener el valor de la variable conocida el área.

Cuidado: en estos dos ejemplos conocemos el área (que se encuentra en el interior de la tabla) y queremos hallar el valor de z (en los bordes de la tabla)

☞ Ejemplo: sabiendo que $p\,(z \leq k) = 0.9357$ obtengamos el valor de k

El área está en la intersección de la fila del 1.5 con la columna del 0.02, así:

$$k = 1.5 + 0.02 = 1.52$$

☞ Ejemplo: sabiendo que $p\,(z \geq k) = 0.0041$ obtengamos el valor de k.

Como el área dada es la que hay a la derecha de k, tenemos antes que restar de la unidad para obtener el área acumulada a la izquierda de ese valor desconocido, que ea la dada en la tabla:

$$p\,(z \geq k) = 0{,}0041 \rightarrow p\,(z \leq k) = 1 - 0{,}0041 = 0{,}9959 \rightarrow$$
$$\rightarrow \phi(k) = 0{,}9959 \rightarrow k = 2{,}64$$

4.Probabilidades en una normal: tipificación

Resumamos lo fundamental:

Consideremos una variable $\mathbf{X}$ en una población cuya media es μ y cuya desviación típica es σ. Diremos que la variable sigue una distribución normal si la probabilidad de que un individuo tome un valor comprendido entre a y b viene dada por el área bajo la correspondiente curva de Gauss en dicho intervalo.

En ese caso escribimos

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

Dicho de otro modo: las distribuciones normales son aquellas cuyas función de densidad es una campana de Gauss.

Para averiguar probabilidades en una distribución normal cualquiera, se reduce el cálculo manual al de la normal típica gracias al siguiente resultado, conocido como propiedad de tipificación:

Si $\mathbf{X} \sim \mathcal{N}(\mu, \sigma)$ entonces

$$p(X \leq x_0) = p\left(Z \leq \frac{x_0 - \mu}{\sigma}\right)$$

O dicho de otra forma:

$$\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma}$$

es una distribución $\mathcal{N}(0, 1)$.

☞ <u>Ejemplo</u>: La estatura de una población se distribuye normalmente con media 170 cm. y desviación típica 6 cm.

a) Calculemos la probabilidad de que elegido un individuo al azar tenga una estatura inferior a los 175 cm.

b) En un grupo de 40 individuos de ese colectivo, ¿cuántas personas con esa estatura esperamos encontrar?

Estamos estudiando la variable $\mathbf{X}$ = "estatura de un individuo". Se nos señala que sigue una distribución $\mathcal{N}(170, 6)$. Hemos de calcular la probabilidad $p(X \leq 175)$.

Ahora calculamos tipificando y recurriendo a las tablas:

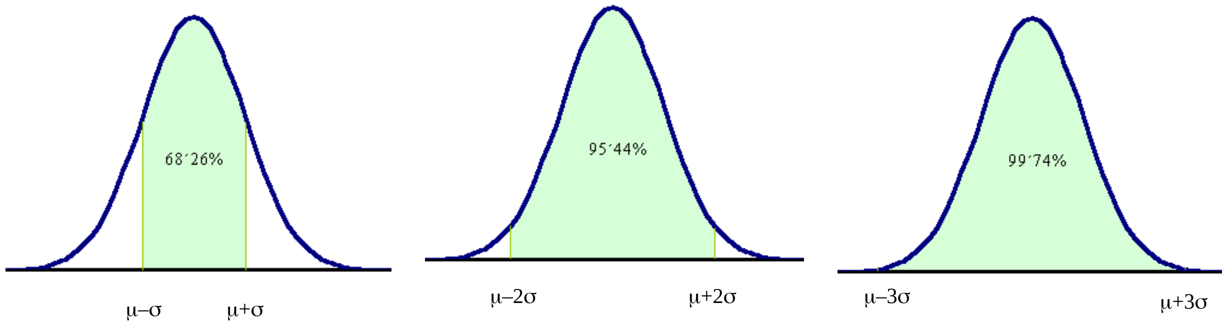
$$p(X \leq 175) = p\left(Z \leq \frac{175 - 170}{6}\right) = p(Z \leq 0.83) = 0.7967$$

En un grupo de 40 individuos de ese colectivo, esperamos encontrar:

$$E = 40 \cdot p(X \leq 175) = 31.868 \approx 32 \text{ individuos}$$

5.Intervalos característicos centrales

En la distribución normal de las probabilidades tienen especial relevancia los siguientes intervalos característicos:



- En $(\mu - \sigma\,, \mu + \sigma)$ está el 68.26% de los individuos.
- En $(\mu - 2\sigma\,, \mu + 2\sigma)$ está el 95.44% de los individuos.
- En $(\mu - 3\sigma\,, \mu + 3\sigma)$ está el 99.74% de los individuos.

6.Ampliación: valores críticos

Son parámetros de gran utilidad que se aparecen en multitud de problemas que aprenderemos a resolver el curso próximo

Consideremos un número α, comprendido entre 0 y 1, que denominaremos nivel de significación.

Llamamos valor crítico unilateral asociado, a dicho nivel de significación, al número z_α tal que:

$$\mathrm{p}\,(z > z_\alpha) = \alpha$$

Llamamos valor crítico bilateral asociado, a dicho nivel de significación, al número $z_{\frac{\alpha}{2}}$ tal que:

$$\mathrm{p}\left(z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

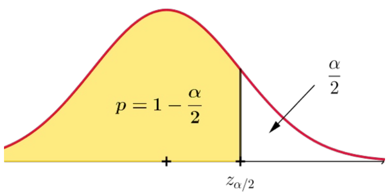
Suele decirse en estos casos que el nivel de confianza es (1 - α) ·100%

☞ Calculemos el valore crítico unilateral correspondiente a un nivel de significación $\alpha = 0.02$.

$$\mathrm{p}\,(z > z_\alpha) = 0.02 \to \mathrm{p}\,(z < z_\alpha) = 0.98 \to z_\alpha \approx 2.05$$

Significación: $\alpha = 0.02$ ➡ Valor crítico: $z_\alpha \approx 2.05$

Se dice también que el nivel de confianza es del 98%.

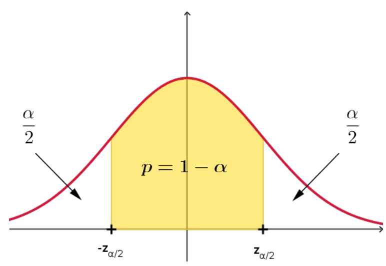


☞ Calculemos el valor crítico bilateral correspondiente a un nivel de significación $\alpha = 0.01$.

$$\mathrm{p}\left(z > z_{\alpha/2}\right) = 0.005 \to \mathrm{p}\left(z < z_{\alpha/2}\right) = 0.995 \to z_{\alpha/2} \approx 2.575$$

Significación: $\alpha = 0.01$ ➡ Valor crítico: $z_{\alpha/2} \approx 2.575$

Diríamos también que el nivel de confianza es del 99%.



Ejercicios

1. Calculemos:

a) $p\,(z \leq 1.47)$ b) $p\,(z \geq -2.24)$

c) $p\,(z \leq -0.8)$ d) $p\,(1.4 \leq z \leq 3.56)$

2. Halla el valor de k en los siguientes casos:

a) $p\,(z \leq k) = 0.9875$

b) $p\,(z \geq k) = 0.003$

3. Las precipitaciones anuales en una región son, en media, de 2000 mm, con una desviación típica de 300 mm. Calcula, suponiendo distribución normal, la probabilidad de que un año determinado la lluvia no supere los 1200 mm.

4. Se sabe que la talla media de la población en edad escolar es de 165 cm con desviación típica de 12 cm. Un Centro tiene 1400 alumnos matriculados.

a) ¿Cuántos alumnos miden más de 155 cm?

b) ¿Qué porcentaje de alumnos mide entre 150 y 178 cm?

c) Determina la probabilidad de que un cierto alumno mida entre 170 y 185 cm.

5. El peso de 600 alumnos se distribuye según una ley $\mathcal{N}\,(67\,,5)$. Calcula cuántos de ellos pesan:

a) Más de 80 kilos.

b) Menos de 50 kilos.

c) Entre 50 y 80 kilos.

6. La presión sanguínea de ciertos enfermos sigue una ley normal de media 90 mm Hg y de desviación típica 12 mm Hg. Halla la probabilidad de que elegido un paciente al azar:

a) Su presión sea mayor de 115 mm Hg.

b) Su presión esté entre 80 y 110 mm Hg.

7. Varios test de inteligencia dieron una puntuación que sigue una ley normal con media 100 y desviación típica 15.

a) Determina el porcentaje de población que obtendría un coeficiente entre 95 y 110.

b) ¿Qué intervalo centrado en 100 contiene al 50% de la población?

c) En una población de 2500 individuos, ¿cuántos esperamos que tengan un C.I. superior a 125?

8. El peso de los adultos de una población numerosa se distribuye normalmente con media 65 kg y desviación típica 3 kg. Se eligen dos individuos al azar. Justifica qué es más probable:

a) Que cada uno de los individuos tenga un peso comprendido entre 63,5 kg y 66,5 kg.

b) Que uno de ellos tenga un peso comprendido entre 62 y 68 kg y que el otro tenga un peso no comprendido entre 62 y 68 kg.

9. El tiempo necesario para que una ambulancia llegue a su destino se distribuye normalmente con una media de 17 minutos y desviación típica de 3 minutos.

a) Calcula la probabilidad de que el tiempo de llegada esté comprendido entre 13 y 21 minutos.

b) ¿Para qué valor de t la probabilidad de que la ambulancia emplee más de t minutos en llegar es el 5%?

10. Para aprobar unas oposiciones se necesita obtener 100 puntos, o más, en una prueba. Por experiencias anteriores, se sabe que la distribución de los puntos obtenidos por los opositores es una normal de media 110 puntos y desviación típica 15.

a) ¿Cuál es la probabilidad de que un opositor apruebe?

b) Si sabemos que hay 1000 opositores y sólo 300 plazas, ¿cuántos puntos se deberá exigir para ajustar el número de plazas al número de opositores aprobados?

11. Sea $\mathbf{X}$ una variable de media 0 que se distribuye normalmente. Calcula su varianza sabiendo que es $p(X \geq 2) = 0.1587$.

12. El tiempo de vida de un tipo de foco se distribuye siguiendo una distribución normal con desviación típica 4 horas.. Se comprueba que de 1500 focos, sólo 1400 lucen más de 30 horas. ¿Cuál es su media?

13. Las calificaciones de los estudiantes de un curso siguen una distribución normal. Si las puntuaciones tipificadas de dos de ellos fueron 0,8 y –0,4 y sus notas reales fueron 88 y 64 puntos, ¿cuál es la media y la desviación típica de las puntuaciones del examen?

Cuestiones

1. Si **X** es una distribución normal, ¿cuál es la probabilidad $p(X = k)$? ¿Cómo se calcula la probabilidad $p(a < X < b)$? Señala si esta probabilidad es igual o no a $p(a \leq X \leq b)$.

2. En la distribución normal tipificada, ¿cuáles son los intervalos característicos?

3. En la distribución $\mathcal{N}(0, 1)$ tenemos que es k un número positivo tal que $\phi(k) = p$. Expresa en función de p las probabilidades siguientes:

a) $p(Z \leq k)$

b) $p(Z \geq k)$

c) $p(Z \leq -k)$

d) $p(Z \geq -k)$

e) $p(0 \leq Z \leq k)$

f) $p(-k \leq Z \leq k)$

4. En la distribución $\mathcal{N}(0, 1)$ tenemos $0 < a < b$ con $\phi(a) = p$ y $\phi(b) = q$. Expresa en función de p y de q las probabilidades siguientes:

a) $p(a \leq X \leq b)$

b) $p(-a \leq X \leq b)$

c) $p(-b \leq X \leq a)$

d) $p(-a \leq X \leq b)$

e) $p(-b \leq X \leq -a)$

5. En la distribución $\mathcal{N}(20, 3)$, señala cuál es el intervalo característico $(\mu - \sigma, \mu + \sigma)$. Comprueba que en él se encuentra el 68,26% de las observaciones.

6. Sin hacer uso de las tablas, calcula qué porcentaje de observaciones en la distribución normal $\mathcal{N}(30, 5)$ están fuera del intervalo $(20, 40)$.

7. Sea $\mathbf{X} \sim \mathcal{N}(\mu, \sigma)$. Si designamos por z_α al valor crítico unilateral asociado a α, deduce de la fórmula de tipificación

$$\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma}$$

que es

$$p(X \leq \mu + z_\alpha \cdot \sigma) = 1 - \alpha$$

Autoevaluación

1. Calcula:

a) $p\left(Z \geq 2.28\right)$

b) $p\left(Z \leq -1\right)$

c) $p\left(Z \geq -3.25\right)$

2. Obtén el valor de k sabiendo que:

a) $p\left(Z \leq k\right) = 0.99886.$

b) $p\left(Z > k\right) = 0.0039.$

3. Se eligió una muestra de 1000 personas de una determinada población y resultó que su peso medio era de 70 Kg., con una desviación típica de 10 Kg.

Sabiendo que el peso es una variable que se distribuye normalmente, hállese:

a) El número de personas de la población que pesa menos de 60 Kg.

b) ¿Cuántas personas pesarán más de 100 Kg.?

c) ¿Qué porcentaje tiene un peso comprendido entre los 60 y los 90 Kg.?

d) Se eligen dos individuos al azar. Halla la probabilidad de que uno pese menos de 60 Kg. y de que el otro pese más de 100 Kg.

Autoevaluación

1. Calcula:

a) $p\,(Z \geq 2.28)$

b) $p\,(Z \leq -1)$

c) $p\,(Z \geq -3.25)$

2.

a) Basta buscar en la tabla directamente:

$$p\,(Z \leq k) = 0.99886 \;\rightarrow\; k = 3.06$$

b) Buscamos en la tabla a la inversa pero primero hallamos el área acumulada:

$$p\,(Z \leq k) = 1 - 0.0039 = 0.9961 \rightarrow k = 2.66$$

3. La variable X = "peso de un individuo" sigue una distribución $\mathcal{N}\,(70\,,10)$

a) Hemos de hallar primero $p\,(X < 60)$:

$$p\,(X < 60) = p\left(Z < \frac{60-70}{10}\right) = p\,(Z < -1)$$
$$= 1 - 0{,}8413 = 0{,}1587$$

Luego esperamos:

$$E = N \cdot p = 1000 \cdot p\,(X < 60) \approx 159 \text{ personas}$$

b) Calculemos primero $p\,(X > 100)$:

$$p\,(X > 100) = p\left(Z > \frac{100-70}{10}\right) = p\,(Z > 3)$$
$$= 1 - 0,99865 = 0{,}00135$$

Luego esperamos:

$$E = N \cdot p = 1000 \cdot p\,(X > 100) \approx 13 \text{ personas}$$

c) Calcularemos $p\,(60 \leq X \leq 90)$

$$p = p\left(\frac{60-70}{10} \leq Z \leq \frac{90-70}{10}\right)$$
$$= p\,(-1 \leq Z \leq 2) = 0{,}9772 - (1 - 0{,}8413)$$
$$= 0{,}8185$$

En porcentajes: el 81,85%.

d) Como el número de individuos es grande, ambos son sucesos independientes. Así la probabilidad pedida es:

$$p = p\,([X_1 < 60] \cap [X_2 > 100])$$
$$= p\,(X < 60) \cdot p\,(X > 100) = 0{,}1587 \cdot 0{,}0135$$
$$\approx 0{,}00021$$