DISTRIBUCIONES DE PROBABILIDAD. BINOMIAL Y NORMAL.

1.- Variables aleatorias.

Variable aleatoria: Se llama variable aleatoria a toda función que asocia a cada elemento del espacio muestral E un número real.

Se utilizan letras mayúsculas X, Y ... para designar variables aleatorias, y las respectivas minúsculas (x, y, ...) para designar valores concretos de las mismas.

Variable aleatoria discreta: Una variable discreta es aquella que solo puede tomar un número finito de valores entre dos valores cualesquiera de una característica.

Ejemplos: El número de hijos de una familia, la puntuación obtenida al lanzar un dado.

Variable aleatoria continua: Una variable continua es aquella que puede tomar un número infinito de valores entre dos valores cualesquiera de una característica.

Ejemplos: La altura de los alumnos de una clase, las horas de duración de una pila.

2. 2.1. Variables discretas. Función de probabilidad.

Se llama función de probabilidad de una variable aleatoria discreta X a la aplicación que asocia a cada valor de x_i de la variable su probabilidad p_i.

$\mathbf{0 \le p_i \le 1}$

$\mathbf{p_1 + p_2 + p_3 + \cdots + p_n = \Sigma\, p_i = 1}$

Ejemplo: Calcular la **distribución de probabilidad** de las puntuaciones obtenidas al lanzar un dado.

Todas las probabilidades son 1/6 y como son 6 posibles todas suman 1.

$P_1=1/6 \ldots P_2=1/6$

2.2. Función de distribución.

Sea X una variable aleatoria discreta cuyos valores suponemos ordenados de menor a mayor. Llamaremos **función de distribución de la variable X**, y escribiremos F(x) a la función:

F(x) = p(X ≤ x) La función de distribución asocia a cada valor de la variable aleatoria la probabilidad acumulada hasta ese valor.

2.3. Media, varianza, desviación típica.

Esperanza matemática o media

$$\mu = X_1 \cdot p_1 + X_2 \cdot p_2 + \cdots + X_i \cdot p_i = \sum_{i=1}^{n} x_i \cdot p_i$$

Varianza

$$\sigma^2 = \sum_{i=1}^{n} x_i^2 \cdot p_i - \mu^2$$

Desviación típica

$$\sigma = \sqrt{\sum_{i=1}^{n} x_i^2 \cdot p_i - \mu^2}$$

3. Distribución Binomial o de Bernoulli.

La variable aleatoria binomial es una variable discreta que representa el número de éxitos obtenidos relativos a esa variable.

Un experimento sigue el modelo de la **distribución binomial o de Bernoulli** si:

1. En cada prueba del experimento sólo son posibles **dos resultados**: el suceso A (**éxito**) y su contrario $\overline{A}$ (fracaso).

2. La **probabilidad del suceso A es constante**, es decir, que no varía de una prueba a otra. Se representa por **p**.

3. El **resultado** obtenido en cada prueba es **independiente** de los resultados obtenidos anteriormente.

La **distribución binomial** se suele representar por **B(n, p)**.

n es el número de pruebas de que consta el experimento.

p es la probabilidad de éxito.

La **probabilidad de** $\overline{A}$ **es 1− p**, la representamos por **q** y se denomina fracaso.

La **función de probabilidad de la distribución binomial**, también denominada **función de la distribución de Bernoulli**, es:

$$p(X = k) = \binom{n}{k} p^k \cdot q^{n-k}$$

n es el número de pruebas.

k es el número de éxitos.

p es la probabilidad de éxito.

q es la probabilidad de fracaso.

Y n sobre k es el número combinatorio:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Definimos Media, varianza y desviación típica de la variable binomial como:

$$\mu = n \cdot p \quad \sigma^2 = n \cdot p \cdot q \quad \sigma = \sqrt{n \cdot p \cdot q}$$

Ejercicio 1: La última novela de un autor ha tenido un gran éxito, hasta el punto de que el 80% de los lectores ya la han leido. Un grupo de 4 amigos son aficionados a la lectura:

a) ¿Cuál es la probabilidad de que el grupo hayan leido la novela 2 personas?
b) ¿Y cómo máximo 2?

Ejercicio 2: En unas pruebas de alcoholemia se ha observado que el 5% de los conductores controlados dan positivo en la prueba y que el 10% de los conductores controlados no llevan puesto el cinturón de seguridad. También se ha observado que las dos infracciones son independientes. Un guardia de tráfico para cinco conductores al azar. Si tenemos en cuenta que el número de conductores es suficientemente importante como para estimar que la proporción de infractores no varía al hacer la selección

a) Determinar la probabilidad de que exactamente tres conductores hayan cometido alguna de las dos infracciones
b) Determine la probabilidad de que al menos uno de los conductores controlados haya cometido alguna de las dos infracciones

Ejercicio 3: Un laboratorio afirma que una droga causa efectos secundarios en una proporción de 3 de cada 100 pacientes. Para contrastar esta afirmación, otro laboratorio elige al azar a 5 pacientes a los que aplica la droga. ¿Cuál es la probabilidad de los siguientes sucesos?

a) Ningún paciente tenga efectos secundarios
b) Al menos dos tengan efectos secundarios
c) ¿Cuál es el número medio de pacientes que espera laboratorio que sufran efectos secundarios si elige 100 pacientes al azar?

4. Distribución Normal.

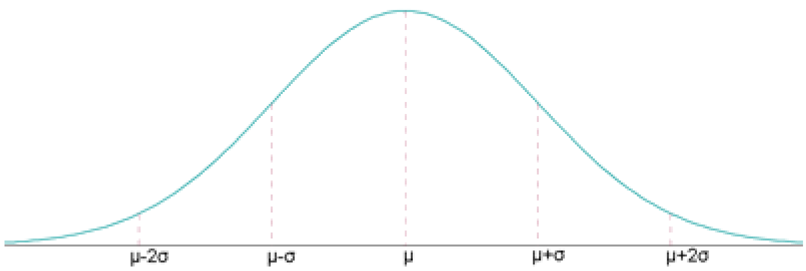
Una **variable aleatoria continua**, **X**, sigue una **distribución normal** de **media μ** y **desviación típica σ**, y se designa por **N(μ, σ)**, si se cumplen las siguientes condiciones:

1. La variable puede tomar cualquier valor: (-∞, +∞)

2. La **función de densidad**, es la expresión en términos de ecuación matemática de la **curva de Gauss**:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Curva de la distribución normal



- El campo de existencia es cualquier valor real, es decir, (-∞, +∞).
- Es simétrica respecto a la media μ.
- Tiene un máximo en la media μ.
- Crece hasta la media μ y decrece a partir de ella.
- En los puntos $\mu - \sigma$ y $\mu + \sigma$ presenta puntos de inflexión.
- El eje de abscisas es una asíntota de la curva.

El área del recinto determinado por la función y el eje de abscisas **es igual a la unidad**.

Al ser **simétrica** respecto al eje que pasa por **x = μ**, deja un **área igual a 0.5 a la izquierda y otra igual a 0.5 a la derecha**. x

Pero nosotros no vamos a trabajar con la distribución normal de media μ y desviación σ si no con la Distribución Normal Estándar, la conocida como Normal de media=0 y desviación=1, puesto que para esta distribución existe una tabla que nos determina las probabilidades deseadas. Tendremos entonces que **pasar** de una variable $\boldsymbol{X \epsilon N(\mu, \sigma)}$ con la que obtenemos $\boldsymbol{P(X \leq k)}$ a una variable $\boldsymbol{Z \epsilon N(0,1)}$ con la que obtendremos $\boldsymbol{P(Z \leq k)}$**.** Este proceso se denomina **TIPIFICAR** y se realizará de la siguiente forma:

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

Tabla de la curva normal (0, 1)

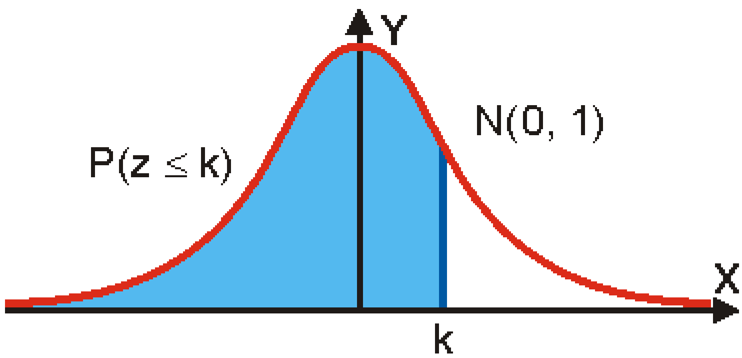
La **tabla** nos da las **probabilidades de P(z ≤ k)**, siendo **z** la variable tipificada.

Estas probabilidades nos dan la **función de distribución F(k)**.

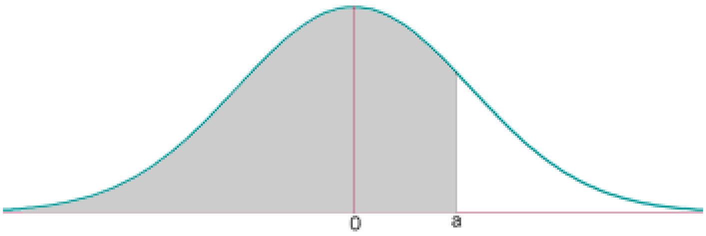
F(k) = P(z ≤ k)

Búsqueda en la tabla de valor de k: **Unidades y décimas** en la columna de la izquierda. **Céntesimas** en la fila de arriba.

Asociada a la distribución normal tipificada existe la denominada Tabla de la normal que nos indica la probabilidad de que una variable tipificada sea menor o mayor que un dato cualquiera.
$\boldsymbol{F(a) = P(Z \leq k)}$ siendo $\boldsymbol{Z \in N(0,1)}$



P(Z ≤ a)



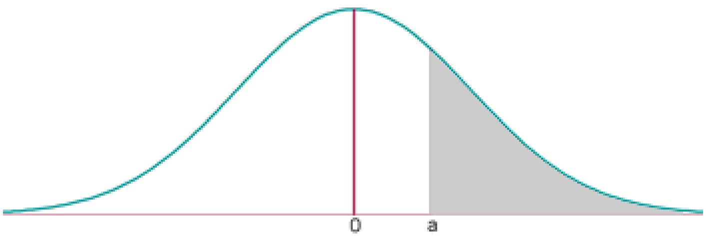
P(Z ≤ 1.47) = 0.9292

Pero no siempre la probabilidad que nos mandan es menor e igual que, por los que no se corresponderá con la probabilidad proporcionada por la tabla. Para calcular las demás probabilidades debemos ponerlas en función de probabilidades como la anterior que ya conocemos. Veamos los diferentes casos teóricos que posteriormente aparecen como ejemplos:

- $P(Z \geq a) = 1 - P(Z \leq a)$
- $P(Z \leq -a) = P(Z \geq a) = 1 - P(Z \leq a)$
- $P(Z \geq -a) = P(Z \leq a)$
- $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$
- $P(-a \leq Z \leq b) = P(Z \leq b) - P(Z \leq -a)$
- $P(-a \leq Z \leq -b) = P(Z \leq -b) - P(Z \leq -a)$

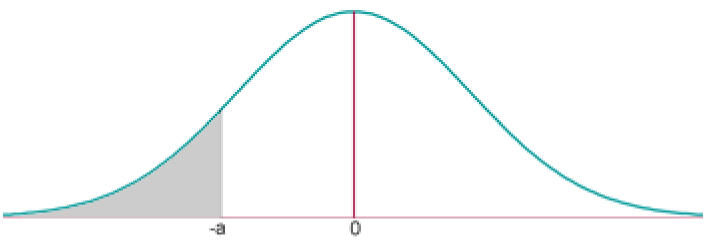
De todas formas, lo mejor para ver cada uno de los casos es hacer **el dibujo en la campana de Gauss y asociarlo con lo que nos proporciona la tabla.**

P(Z > a) = 1 - P(Z ≤ a)



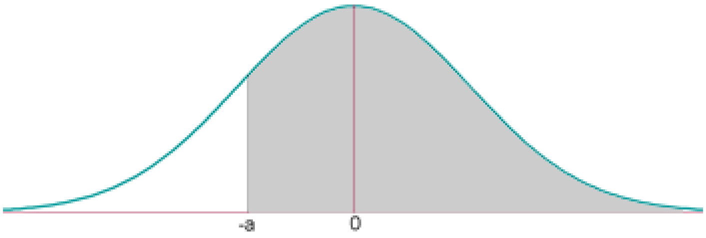
P(Z > 1.47) = 1 – P(Z ≤ 1.47) = 1 – 0.9292 = 0.0708

P(Z ≤ –a) = 1 – P(Z ≤ a)



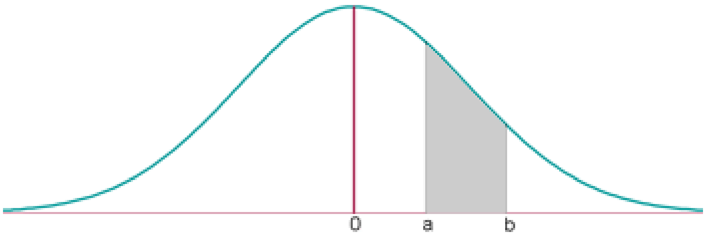
$P(Z \leq -1.47) = 1 - P(Z \leq 1.47) = 1 - 0.9292 = 0.0708$

$P(Z > -a) = P(Z \leq a)$



$p(Z > -1.47) = p(Z \leq 1.47) = 0.9292$

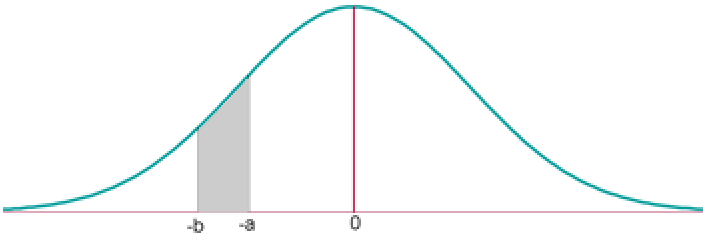
$P(a < Z \leq b) = P(Z \leq b) - P(Z \leq a)$



$P(0.45 < Z \leq 1.47) = P(Z \leq 1.47) - P(Z \leq 0.45) =$

$= 0.9292 - 0.6736 = 0.2556$

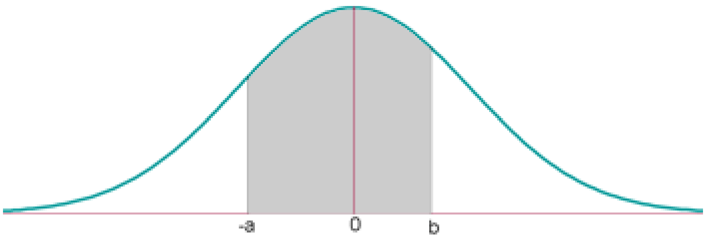
$P(-b < Z \leq -a) = P(a < Z \leq b)$



$P(-1.47 < Z \leq -0.45) = P(0.45 < Z \leq 1.47) =$

$= P(Z \leq 1.47) - P(Z \leq 0.45) = 0.9292 - 0.6736 = 0.2556$

$P(-a < Z \leq b) = P(Z \leq b) - [1 - P(Z \leq a)]$



$P(-1.47 < Z \leq 0.45) = P(Z \leq 0.45) - [1 - P(Z \leq 1.47)] =$

= 0.6736 − (1 − 0.9292) = 0.6028

Ejemplo1: La temperatura durante septiembre está distribuida normalmente con media de 18,7ºC y desviación 5ºC. Calcula la probabilidad de que la temperatura durante septiembre estea por debajo de los 21ºC.

Ejemplo2: La media de los pesos de 5000 estudiantes de un instituto es de 70kg y la desviación típica 3kg. Suponiendo que los pesos se distribuyen normalmente, calcula cuántos estudiantes pesan menos de 60Kg

Nota: no podemos calcular el número de estudiantes pero si la probabilidad de que un estudiante pese menos de 60 y finalmente aplicar esa probabilidad a los 5000 estudiantes.

APROXIMACIÓN DE LA DISTRIBUCIÓN BINOMIAL A UNA DISTRIBUCIÓN NORMAL

Cuando "n" es grande, la distribución Binomial resulta laboriosa y complicada, por lo que se demostró que cuando se dan ciertas circunstancias una distribución binomial se puede aproximar a una distribución normal de media $\mu = n \cdot p$ y desviación típica $\sigma = \sqrt{n \cdot p \cdot q}$

Normalmente se efectúa la aproximación con n mayor que 30 y p próximo a 0,5 pero esta aproximación será mejor cuanto mayor sea **n** y para valores de **p** y **q** no muy próximos a cero. Se considera que si **n·p** > 5 y **n·q** >5 la aproximación es bastante buena. Por lo tanto podemos usar esta condición.

$$\boldsymbol{X \in B(n,p) \to X \in N(n \cdot p, \sqrt{n \cdot p \cdot q})}$$

Sin embargo, nos encontramos que estamos pasando de una variable discreta en la que la probabilidad de un punto tiene un valor concreto a una variable continua en la que esto no ocurre, por lo tanto es necesario utilizar el conocido **factor de corrección** en algunos casos:

Sea $X \in B(n,p)$ e $Y \in N(n \cdot p; \sqrt{n \cdot p \cdot q})$ una aproximación de X, entonces:

- $p(X = a) = p(a - 0{,}5 \le Y \le a + 0{,}5)$
- $p(a \le X \le b) = p(a - 0{,}5 \le Y \le b + 0{,}5)$
- $p(a \le X < b) = p(a - 0{,}5 \le Y \le b - 0{,}5)$
- $p(a < X \le b) = p(a + 0{,}5 \le Y \le b + 0{,}5)$
- $p(a < X < b) = p(a + 0{,}5 \le Y \le b - 0{,}5)$
- los demás casos se corrigen de la misma forma.

EJERCICIOS BINOMIAL Y NORMAL:

1.- Sabemos que el 25% de los alumnos de 2º bach salen los sábados. Si elegimos 5 alumnos al azar, ¿cuál es la probabilidad de que salgan 3 este sábado? Y si elegimos los 70 alumnos que tiene el centro, calcula la probabilidad de que salgan entre 15 y 20 alumnos.

2.- La cantidad de café depositada en cada bolsa por una máquina envasadora automática sigue una distribución normal de media 1040gramos y desviación típica 50 gramos.

a. Calcula el tanto por ciento de paquetes que contienen más de 1 Kg.

b. Calcula la probabilidad de que un paquete tenga un peso comprendido entre 950 gramos y 1050 gramos.

3.- Supongamos que la probabilidad de tener una unidad defectuosa en una línea de ensamblaje es de 0.05. Si el conjunto de unidades terminadas constituye un conjunto de ensayos independientes:

a) ¿cuál es la probabilidad de que entre diez unidades dos se encuentren defectuosas?

b) ¿y de que a lo sumo dos se encuentren defectuosas? c) ¿cuál es la probabilidad de que por lo menos una se encuentre defectuosa? d) En estos casos, ¿cuál es el número esperado de unidades defectuosas? e) Si en vez de 10 unidades hablamos de 200, ¿cuál es la probabilidad de que encontrar al menos 9 defectuosas?

4.- Se supone que los resultados de un examen siguen una distribución normal con media 78 y desviación 36. Se pide:

a) ¿Cuál es la probabilidad de que una persona que se presenta al examen obtenga una calificación superior a 72? b) Sabiendo que la calificación de un estudiante es mayor que 72, ¿cuál es la prioridad de que su calificación sea, de hecho, superior a 84?

5.- En un examen tipo test de 200 preguntas de elección múltiple, cada pregunta tiene una respueseta correcta y una incorrecta. Se aprueba si se contesta a más de 110 respuestas correctas. Suponiendo que se contesta al azar, calcula la probabilidad de aprobar el examen.

6.- Una distribución normal tiene una media de 80 y una desviación de 14. Determina el valor por encima del cual se presentará el 80% de las observaciones.

7.- Las puntuaciones de un examen se distribuyen normalmente con media 15 puntos. La puntuación A ha sido superada por un 23% de los alumnos. La puntuación B está situada a 5 puntos diferenciales por debajo de la media. Entre B y la media se encuentra el 30% de los alumnos. Calcular : a) La desviación típica de las notas. b) Las puntuaciones directas de A y B. c) El porcentaje de alumnos entre A y B.

8.- El 65% de los alumnos de un cierto instituto cursan estudios universitarios al terminar el Bachillerato. En un grupo de ocho alumnos elegidos al azar, halla la probabilidad de que estudien una carrera: a)Alguno de ellos. b) Más de seis. c) Calcula la media y la desviación típica.

9.- En una urna hay 3 bolas rojas, 2 blancas y 5 verdes. Sacamos una bola, anotamos su color y la devolvemos a la urna. Si repetimos la experiencia 50 veces, ¿cuál es la probabilidad de sacar roja en más de 20 ocasiones?

10.- En una familia con seis hijos, ¿cuál de estas dos opciones es más probable?

- Que haya tantas chicas como chicos. - Que haya más chicas que chicos.

11.- La distribución del número de atunes captados por los barcos pesqueros que salen a faenar en una cierta zona es una normal de media 110. Se sabe que, tomando un barco al azar, la probabilidad de que capture más de 125 atunes es 0,1587.

a) Calcular la desviación típica de la distribución. b) Se considera que la campaña ha sido buena si se capturan más de 100 atunes, ¿qué porcentaje estimado de barcos no tendrán una buena campaña?

c) ¿Cuántos atunes debe capturar un barco para estar en el percentil 90?