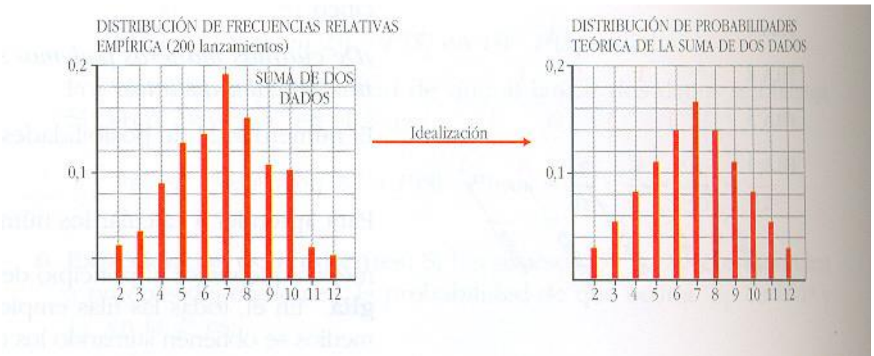
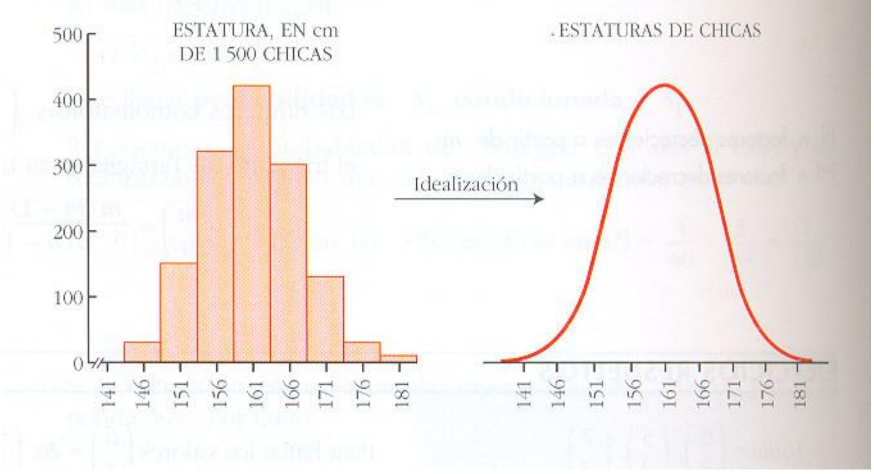
DISTRIBUCIONES DE PROBABILIDAD

Las distribuciones de probabilidad son idealizaciones de las distribuciones de frecuencias relativas. Estas son experimentales (lanzamiento de dos dados 180 veces), las distribuciones de probabilidad son teóricas.



Cuando la variable es continua, la distribución estadística viene dada mediante un histograma de frecuencias relativas y su idealización, que es la distribución de probabilidad, mediante una curva.



Variable aleatoria: Se llama variable aleatoria a toda aplicación del espacio muestral E en los números reales. $E \xrightarrow{\chi} \Re$.

Para definir una variable aleatoria se necesita:

a) Un experimento aleatorio con su espacio muestral.

b) Una regla o función que, a cada suceso elemental "e" del espacio muestral, le haga corresponder un número $\chi(e)$.

La variable aleatoria χ puede ser de dos tipos:

Discreta: solo toma un número finito de valores (o a lo sumo numerables)

Continua: puede tomar todos los valores de un determinado intervalo.

De los ejemplos siguientes, decir cuáles son variables aleatorias discretas y cuales continuas.

Ejemplos

a) En el experimento lanzar dos veces un dado. Podemos definir las variables aleatorias:

- *A cada suceso le asociamos la suma de los valores obtenidos en cada dado. ¿Cuáles son los valores de la variable aleatoria?*

E={(1,1) (1,2)........(6,6)} la variable aleatoria χ = {2,3,4,5,6,7,8,9,10,11,12} en este caso la variable aleatoria es discreta

- *A cada suceso le asociamos la diferencia en valor absoluto de los valores obtenidos en cada dado.... ¿Cuáles son los valores de la variable aleatoria?.*

E={(1,1) (1,2)........(6,6)} la variable aleatoria χ = {0,1,2,3,4,5} en este caso la variable aleatoria es discreta

b) En el experimento de elegir una persona de una población, a cada persona le asociamos su altura. ¿Cuáles son los valores de la variable aleatoria?

χ= {[140, 210]} variable aleatoria continua

c) En el experimento de elegir una bombilla de las fabricadas por una determinada empresa, a cada bombilla elegida le asociamos el tiempo de vida. ¿Cuáles son los valores de la variable aleatoria?.

χ= {[0, 2000]} horas de vida, es una variable continua

VARIABLE ALEATORIA DISCRETA.

FUNCIÓN DE PROBABILIDAD

La función de probabilidad (o de masa) de una variable aleatoria discreta χ es la aplicación f que a cada valor de la variable aleatoria le hace corresponder su probabilidad.

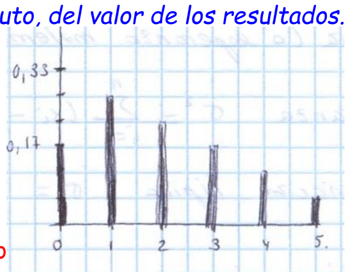
Se expresa de la forma $f(x) = P(\chi = x)$

Ejemplo:

1. Hallar la función de probabilidad de la variable aleatoria asociada al experimento de lanzar dos veces un dado y a cada suceso le asociamos la diferencia, en valor absoluto, del valor de los resultados.

x_i	0	1	2	3	4	5
P_i	6/36	10/36	8/36	6/36	4/36	2/36

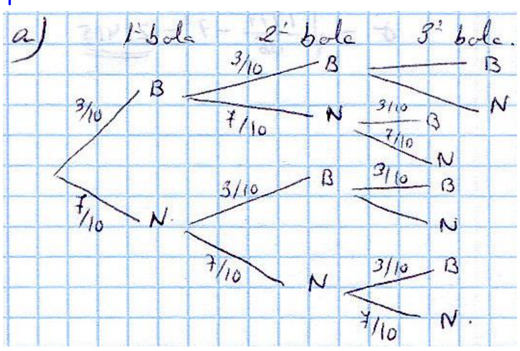
Hacer Gráfica:



Comparar variable estadística cuantitativa discreta con variable aleato
Frecuencia relativa con función de probabilidad
Diagrama de barras con gráfica de la función de probabilidad.

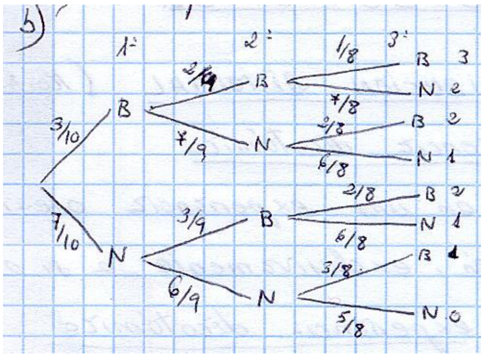
2.- En una urna hay 3 bolas blancas y 7 negras. Se extraen 3 bolas. Sea la variable aleatoria χ: número de bolas blancas extraídas. Halla su función de probabilidad. Hacer la representación gráfica. Hacer el ejercicio primero si la extracción se hace con reposición y luego sin reposición.

a) con reposición



χ= {0,1,2,3}

χ_i	0	1	2	3
p_i	0,34	0,44	0,19	0,08



χ = {0,1,2,3}

χ	0	1	2	3
p_i	0,29	0,53	0,18	0,008

3.- Número obtenido al lanzar un dado:

χ	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

4.– Número de caras al lanzar 2 monedas:

χ	0	1	2
p_i	1/4	2/4	1/4

FUNCIÓN DE DISTRIBUCIÓN.

La **Función de distribución** de una variable aleatoria discreta χ es la que a cada valor x de la variable le asigna la probabilidad de que ésta tome valores menores o iguales que x.

$$f(x) = P(\chi \leq x)$$

Ejemplo:
Hallar la función de distribución de los ejemplos anteriores.

PARÁMETROS DE UNA VARIABLE ALEATORIA DISCRETA.

a) **Media** (o esperanza matemática) $\mu = x_1 p_1 + x_2 p_2 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i$

b) **Varianza** $\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p_i = \sum_{i=1}^{n} x_i^2 \cdot p_i - \mu^2$

c) **Desviación típica**: raíz cuadrada de la varianza. (σ)

Ejemplo:

1) *Hallar la media y la desviación típica de los ejemplos anteriores.*
2) *Hallar la media y la desviación típica de la variable que se obtiene al sumar las puntuaciones de dos dados*

x	2	3	4	5	6	7	8	9	10	11	12
p	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$$\mu = \sum_{i=1}^{n} x_i \cdot p_i = \frac{252}{36} = 7 \quad \sigma = \sqrt{\sum_{i=1}^{n} x_i^2 \cdot p_i - \mu^2} = \sqrt{\frac{1974}{36} - 7^2} = 2{,}415 =$$

DISTRIBUCIÓN BINOMIAL
(Modelo de distribución de probabilidad discreta)

Experiencia dicotómica: Si en una experiencia aleatoria nos fijamos en un suceso A y prestamos atención, exclusivamente, a si ocurre A o su contrario A' se trata entonces de una experiencia dicotómica.

Al suceso A se le llama éxito, y su probabilidad p[A] = p, siendo la de su complementario p[A']= 1 – p = q

Ejemplos de experiencias dicotómicas:

1.– Lanzar una moneda A = cara; A' = cruz p[A]=1/2 = p ; p[A']=1/2 = q

2.– Lanzar una dado y ver si sale "5" ó no: A = {5} ; A' ={1,2,3,4,6} ; p = 1/6 ; q = 5/6

3.– Extraer una carta de una baraja y ver si es figura
A= figura = {As, Sota, Caballo, Rey} ; A'= no figura = {2,3,4,5,6,7}
p[A]=0,4 = p ; q = 0,6

4.– Tenemos un montón de tornillos fabricados por una máquina que por término medio, produce un 2% de defectuosos. Extraemos uno de ellos al azar : A= defectuoso ; A' = no defectuoso ; p = 0,02 q= 0,98

Tenemos una distribución binomial: cuando se repite "n" veces una misma experiencia dicotómica , nos preguntamos por el número x de éxitos. Siendo x una variable aleatoria discreta que puede tomar los valores 0,1,2,...n.

La experiencia se repite "n veces" en las mismas condiciones, la probabilidad de éxito no cambia. En este caso la distribución de probabilidad de la variable x se llama distribución binomial B(n,p) siendo n= nº de veces que se repite la experiencia y p la probabilidad de éxito.

Vemos si las siguientes distribuciones se ajustan a una Binomial:

1.– Lanzamos 10 monedas y nos preguntamos por el número de caras obtenidas.
Es una binomial B(10 , 0,5) χ = nº de caras = {0,1,2,3,4,5,6,7,8,9,10}
2.– Lanzamos 6 dados y nos preguntamos por el número de cincos
Es una binomial B(6 , 1/6) χ = {0,1,2,3,4,5,6}
3.– Extraemos 5 cartas de una baraja y nos preguntamos ¿Cuántas figuras habrá?
Si no hay reemplazo cambia la probabilidad de acierto, por tanto no se ajusta a una binomial
4.– Extraemos una carta de una baraja y la devolvemos al mazo. Barajamos y volvemos a extraer. Repetimos 5 veces la experiencia
Es una binomial B(5; 0,4) p(figura) = p = 0,4; χ = {0,1,2,3,4,5,}
5.– Nos preguntamos cuántos partidos ganará el Betis en sus próximos diez encuentros.

No es binomial ya que la probabilidad de vencer en cada encuentro es muy distinta. Las diez experiencias no son idénticas

6.– Máquina que produce tornillos, un 3% son defectuosos. Se empaquetan en cajas de 100 tornillos. ¿Cuántos defectuosos habrá en cada caja? B(100, 0,03) ; χ = {0,1,2,3,.... ,100}

CALCULO DE PROBABILIDADES EN UNA BINOMIAL

Ejemplo: Extraer "figura" en una baraja de 40 cartas. Se realizan 5 extracciones con reemplazamiento, es decir, reponiendo la carta y barajando.
p[figura] = 0,4 = p[F]; p[no figura] = p[F'] = 0,6; B(5 ; 0,4)

Los posibles resultados serán por ejemplo:

F F' F F F'
F' F' F' F F'

Calculamos la probabilidad de obtener 3 figuras

p [F F' F F F'] = $0{,}4^3 \cdot 0{,}6^2$ esta sería la probabilidad en ese orden, pero hay 10 formas distintas en las que podemos obtener 3 figuras y 2 no figuras ($\binom{5}{3}$ = 10 formas)

La probabilidad de obtener 3 figuras será la probabilidad de obtener 3 aciertos:

p[x = 3] = $\binom{5}{3} \cdot (0{,}4)^3 \cdot (0{,}6)^2$

En general cuando tenemos una binomial B(n, p) ; p[A] = p y p[A'] = q = 1 – p ¿Cuál es la probabilidad de que ocurra " k " éxitos?

p[x = k] = $\binom{n}{k} \cdot p^k \cdot q^{n-k}$ **probabilidad de obtener k éxitos**

Parámetros de la distribución binomial:

media: $\mu = n \cdot p$

desviación típica: $\sigma = \sqrt{n \cdot p \cdot q}$

Ejercicios

1.– En una binomial B(8; 0,2) calcular p[x = 0] ; p[x $\neq$ 0]; p[x = 2] así como la media y la desviación típica.

p[x = 0] = $\binom{8}{0} \cdot (0{,}2)^0 \cdot (0{,}8)^8 = 0{,}167$

p[x $\neq$ 0] = 1 – p[x = 0] = 1 – 0,167 = 0,833

p[x = 2] = $\binom{8}{2} \cdot 0{,}2^2 \cdot 0{,}8^6$ = 28 · 0,04 · 0,262 = 0,29

$\mu = n \cdot p = 8 \cdot 0{,}2 = 1{,}6$

$\sigma = \sqrt{npq} = \sqrt{1{,}28} = 1{,}13$

2.– Una máquina produce tornillos, sabemos que un 2% son defectuosos. Se empaquetan cajas de 100 tornillos. Calcula la probabilidad de que en una caja:

a) No haya ninguno defectuoso
b) Haya alguno defectuoso
c) Haya exactamente 3 defectuosos
d) Parámetros

Es una binomial B(100; 0,02) p = 0,02 ; q = 0,98

a) p[ninguno defectuoso]= p[x = 0] = $\binom{100}{0} \cdot 0{,}02^0 \cdot 0{,}98^{100} = 0{,}98^{100} = 0{,}13$

b) p[alguno defectuoso]= p[x $\neq$ 0] = 1 – 0,13 = 0,97

c) p[x = 3] = $\binom{100}{3} \cdot 0{,}02^3 \cdot 0{,}98^{97} = 0{,}182$

d) $\mu = np = 100 \cdot 0{,}02 = 2$

$\sigma = \sqrt{npq} = \sqrt{100 \cdot 0{,}02 \cdot 0{,}98} = 1{,}4$

VARIABLE ALEATORIA CONTINUA.

Las distribuciones de probabilidad de variable continua son idealizaciones de las distribuciones estadísticas de variable continua.
Son variables continuas, estaturas, pesos, tiempo...

FUNCIÓN DE DENSIDAD

Cuando en la estadística descriptiva, hacíamos la representación gráfica de los datos de una variable estadística continua, utilizábamos el histograma. Recordamos que en el histograma la frecuencia relativa (o absoluta) de un determinado intervalo viene expresada por el área del rectángulo (NO POR SU ALTURA)

La función de densidad de una variable aleatoria continua, es una función no negativa que nos permite hallar la probabilidad de que la variable aleatoria tome valores en un determinado intervalo.

Determinar el tipo de distribución que sigue una determinada variable aleatoria continua es una tarea de investigación, así como determinar su función de densidad y sus parámetros.

La distribución más común entre las variables aleatorias continuas es la que se llama **normal**. (En un principio se pensaba que todas las variables aleatorias continuas se comportaban de la misma forma, tenían la misma distribución)

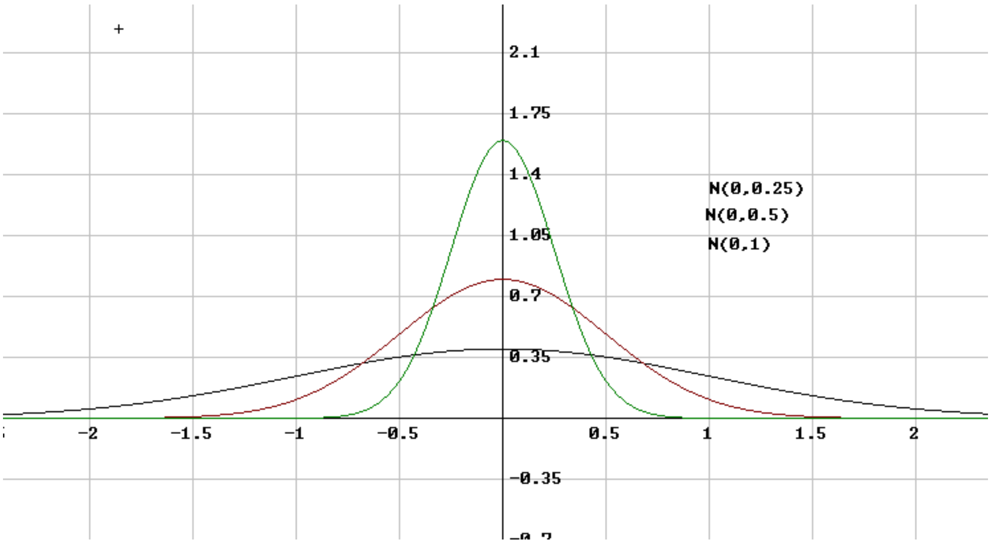
Para calcular probabilidades a partir de la función de densidad f(x) , hay que hallar las áreas bajo la curva (esto se hace por integrales), para la distribución normal estas áreas están tabuladas.

LA DISTRIBUCIÓN NORMAL

La función de densidad de una distribución normal de media μ y de desviación típica σ es:

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

La representación gráfica de estas funciones es de la forma:

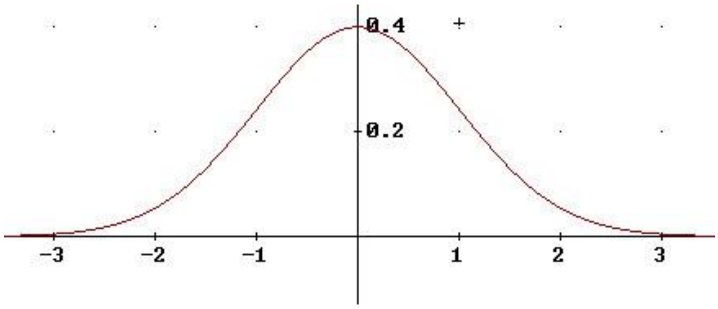


Y la probabilidad de que la variable aleatoria tome valores comprendidos entre a y b viene dado por el área comprendida entre la gráfica de la función f, el eje OX y las rectas x=a y x=b, o sea;

$$P(a \leq \chi \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Como esta integral no se puede calcular por los métodos habituales, se encuentran los valores tabulados.

Lo que tenemos que aprender, es cómo hallar la probabilidad sabiendo manejar las tablas de la **distribución normal de media μ =0 y de desviación típica σ = 1**



En una distribución normal el reparto de probabilidades es :

$p[\mu - \sigma \leq x \leq \mu + \sigma] = 0{,}6826$, el 68,26% de los datos está en este intervalo

$p[\mu - 2\sigma \leq x \leq \mu + 2\sigma] = 0{,}9544$, el 95,44% de los datos está en este intervalo

$p[\mu - 3\sigma \leq x \leq \mu + 3\sigma] = 0{,}9974$, el 99,74 % de los datos está en este intervalo

Manejo de las tablas de Z=N(0,1):

- El área bajo la curva, es la probabilidad del suceso seguro y por lo tanto es 1.
- $P(Z \le 0) = 0,5$

❖ Calcular la probabilidad de los siguientes sucesos:

a) $P(Z \le 1,52)$, b) $P(Z \ge 0,65)$, c) $P(Z \le -1,35)$; d) $P(Z \ge -2,5)$; e) $P(0,74 \le Z \le 1,42)$
f) $P(-0,54 \le Z \le 2,1)$; g) $P(-1,2 \le Z \le -0,4)$, h) $P(-0,5 \le Z \le 0,5)$; i) $P(-1 \le Z \le 1)$;
j) $P(-2 \le Z \le 2)$; l) $P(-3 \le Z \le 3)$

❖ Calcular **a** en los siguientes casos:
a) $P(Z \le a) = 0,7517$; b) $P(Z \le a) = 0,0287$; c) $P(-a \le Z \le a) = 0,9$; d) $P(Z \ge a) = 0,05$;
e) $P(-a \le Z \le a) = 0,95$; f) $P(Z \ge a) = 0,025$; g) $P(-a \le Z \le a) = 0,99$

❖ Tipificación de la variable N(μ , σ).

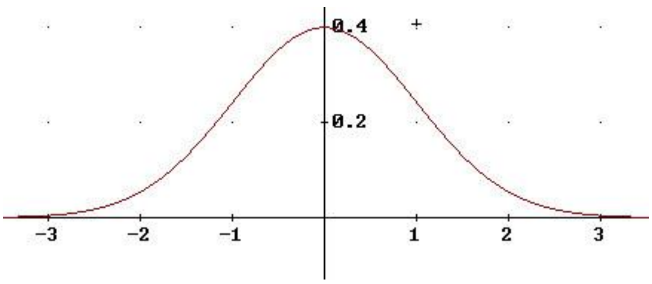
Cuando queremos calcular probabilidades en una distribución N(μ, σ) la relacionamos con una distribución N(0,1) de la que tenemos tablas tabuladas.
Si "x" es una variable que sigue N(μ,σ) para calcular la probabilidad p[h<x<k], debemos tipificar la variable pasando a la variable "z" correspondiente a la distribución N(0,1)

P[h < x < k] = $p\left[\frac{h-\mu}{\sigma} < z < \frac{k-\mu}{\sigma}\right]$ El cambio de la variable "x" a la variable "z" se llama tipificación de la variable. La variable tipificada sigue una distribución normal N(0, 1)

Ejemplo:

1) En una N(6,4) calcular las probabilidades:

a) $p[x \le 3] = p\left[z \le \frac{3-6}{4}\right] = p[z \le -0,75] = p[z \ge 0,75] = 1 - p[z \le 0,75] = 1 - 0,7734 = 0,2266$



Realiza siempre un dibujo para razonar el cálculo de la probabilidad.

b) **p[x ≥ 12] = p[z≥ $\frac{12-6}{4}$] = p [z ≥ 1,5] = 1- p[z ≤ 1,5] = 1 - 0,9332 = 0,0668**

c) **p[5≤x≤8] = p[**$\frac{5-6}{4} \le z \le \frac{8-6}{4}] = p[-0,25 \le z \le 0,5] = p[z \le 0,5] - p[z \le -0,25] = p[z \le 0,5] - p[z \ge 0,25] = p[z \le 0,5] - (1 - p[z \le 0,25]) = 0,6915 - (1 - 0,5987) = 0,2902$

APROXIMACIÓN DE UNA BINOMIAL POR UNA NORMAL

Sabiendo que el 3% de las personas son daltónicas. Tomamos al azar 7 personas y nos preguntamos cuál es la probabilidad de que dos de ellas sean daltónicas, entonces se trata de una distribución binomial con n = 7 y p = 0,03. Es decir x es B(7, 0,03)

P(x=2) = $\binom{7}{2}(0{,}03)^2 \cdot 0{,}97^5 = 0{,}01623$

Parámetros de la distribución: $\mu = np = 7 \cdot 0{,}03 = 0{,}21$; $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{7 \cdot 0{,}03 \cdot 0{,}97} = 0{,}4513$

Para aproximar una binomial a una normal n·$p > 5$ y $n \cdot q > 5$, siendo mayores de 3, la aproximación es buena y si es mayor de 5, la aproximación es casi perfecta.

Como la binomial es discreta y la normal es continua para calcular p(x = k) con una normal se hace: p(x= k) = p(k-0,5<x'<k+0,5);

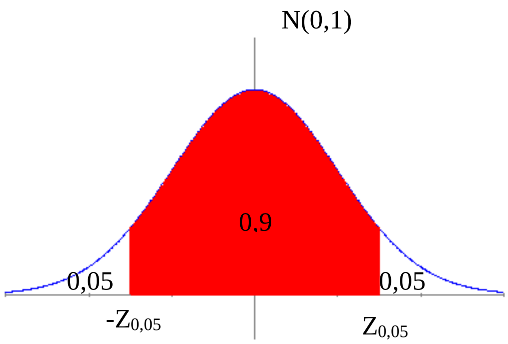
Si consideramos 200 personas, el número de daltónicas, x, es una binomial B(200, 0,03) cuyos parámetros son $\mu = 200 \cdot 0{,}03 = 6$ $\sigma = \sqrt{200 \cdot 0{,}03 \cdot 0{,}97} = 2{,}41$ Se aproxima a una normal con los mismos parámetros: N(6, 2,41)

Calculamos la p(x = 3) = p(3-0,5< x' < 3+ 0,5) = p (2,5<x'<3,5) se pasa a una normal N(0,1) y se calcula la probabilidad.

INTERVALOS CARACTERÍSTICOS

Si la variable x sigue una distribución de media μ, se llama intervalo característico correspondiente a una probabilidad "p" a un intervalo centrado en la media (μ-K, μ+K) tal que la probabilidad de que x pertenezca a dicho intervalo es p:

p[(μ-K<x<μ+K] = p



Hallemos el intervalo característico de una distribución N(0,1) correspondiente a una probabilidad p = 0,9

Si dentro hay un área de 0,9, fuera queda 0,1. Puesto que el intervalo es simétrico, el área de cada una de las dos colas es 0,05

p[z>K] = 0,05 ⇒ p[z>k] = 1- p[z≤k] = 0,05 ⇒ p[z≤k] = 0,95

Buscando en las tablas el valor de la variable correspondiente a esa probabilidad (0.95) encontramos los valores de probabilidad 0,9495 y 0,9505 correspondientes a un valor de k = 1,64 y k= 1,65 respectivamente ⇒ el valor buscado de k = la media aritmética de los dos = 1,645⇒ el intervalo buscado será (-1,645; 1,645) es decir p[(-1,645<z<1,645] = 0,9 ⇒ El intervalo simétrico respecto a 0 (media de esta normal) con un área de 0,9 es el indicado.

La probabilidad se designa como 1 - α

El valor crítico que nos da el extremo del intervalo se denomina: $Z\frac{\alpha}{2}$

INTERVALOS CARACTERÍSTICOS EN DISTRIBUCIONES NORMALES CUALESQUIERA

Sea x una variable que sigue una distribución normal N(μ,σ) y queremos encontrar un intervalo centrado en la media (μ+ k, μ- k) tal que la probabilidad de que la variable x tome un valor en ese intervalo es "p" = 1- α p[μ-k≤x≤μ+k] = p = 1-α

Si x sigue una N(μ,σ), entonces la variable $Z = \frac{x-\mu}{\sigma}$ sigue una N(0,1)

El intervalo característico de $Z = \frac{x-\mu}{\sigma}$ correspondiente a p = 1-α ⇒

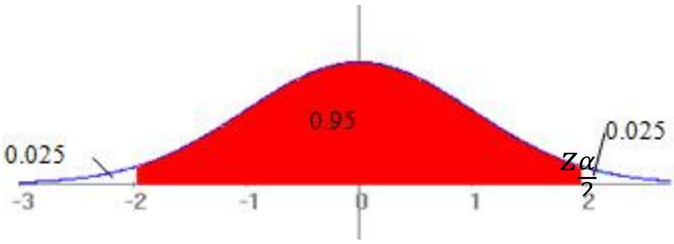
$-z_{\frac{\alpha}{2}} \cdot \sigma < x - \mu < z_{\frac{\alpha}{2}} \cdot \sigma \Rightarrow \mu - z_{\frac{\alpha}{2}} \cdot \sigma < x < \mu + z_{\frac{\alpha}{2}} \cdot \sigma \Rightarrow$ El intervalo característico buscado será: $\left(\mu - z_{\frac{\alpha}{2}} \cdot \sigma,\ \mu + z_{\frac{\alpha}{2}} \cdot \sigma\right)$

Ejemplo:

1.- En una distribución normal N(66, 8) obtener los intervalos característicos (intervalos centrados en la media) para los siguientes casos:

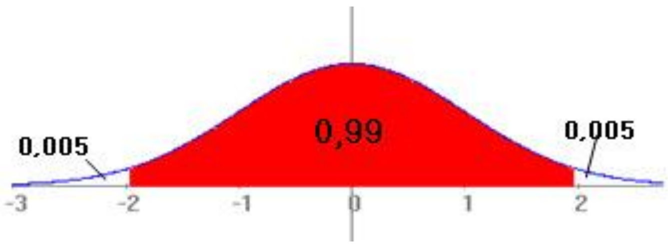
a) p[μ-k≤x≤μ+k] = 0,95
b) p[μ-k≤x≤μ+k] = 0,99

a) 1-α = 0,95⇒ α = 0,05 ⇒$\frac{\alpha}{2} = 0,025$ ⇒ p[z ≤$Z_{\frac{\alpha}{2}}$] = 0,025 + 0,95 = 0,975



buscando el valor de la variable z en las tablas correspondiente a este valor de probabilidad, encontramos $Z_{\frac{\alpha}{2}}$ = 1,96 ⇒ el intervalo buscado será: (μ- 1,96·σ, μ+1,96·σ) = (66-1,96·8 , 66+1,96·8) = (50,32; 81,68)

b) 1- α = 0,99 ⇒ α = 0,01 ⇒ $\frac{\alpha}{2} = 0,005$⇒ p[z ≤$Z_{\frac{\alpha}{2}}$] = 0,005 + 0,99 = 0,995



buscando el valor de la variable z en las tablas correspondiente a este valor de probabilidad, encontramos $Z_{\frac{\alpha}{2}}$ = 2,575, que es la media aritmética de los dos valores que aparecen por defecto y exceso, ya que la probabilidad buscada no está directamente en la tabla.
el intervalo buscado será:
(μ- 2,575·σ, μ+2,575·σ) = (66 - 2,575·8; 66 + 2,575·8) = (45,4; 86,6)