VARIABLES ALEATORIAS

Concepto de variable aleatoria. Tipos

Una variable aleatoria (v.a.) es una función que asigna a cada resultado de un experimento aleatorio un número.

Ejemplos:

- En el experimento aleatorio que consiste en elegir un alumno al azar y preguntarle cuánto pesa, la función X que le hace corresponder a cada alumno su peso es una v.a.
Es decir, X = peso es una v.a.
- En el experimento aleatorio que consiste en lanzar una moneda 5 veces y anotar el nº de caras, la función X que le hace corresponder a cada resultado del experimento el nº de caras es una v.a.
Es decir, X = nº de caras es una v.a.

En las v.a., la media aritmética se representa por μ, la varianza por σ^2 y la desviación típica por σ.

Las v.a. se clasifican en:

Variables aleatorias continuas: Son aquellas que podrían tomar los infinitos valores de un intervalo.
Por ejemplo, si elegimos al azar un alumno y le preguntamos su estatura en cm entonces
X = "estatura" es una v.a. continua pues la estatura puede ser cualquier nº de un intervalo

Otros ejemplos de v.a. continuas son el peso de una persona, la longitud de un tornillo, el nivel de agua de un embalse, la temperatura, horas de duración de una pila, etc

Variables aleatorias discretas: Son aquellas que toman un número contable de valores, 0, 1, 2, 3, etc.
Por ejemplo, la v.a. X = número de personas que viven en una casa es una v.a. discreta

Otros ejemplos de v.a. discretas son suma de los puntos al tirar un dado dos veces, el número de hijos de un matrimonio, el número de asignaturas suspensas de un alumno, el número de libros vendidos en una librería, la edad de una persona, etc

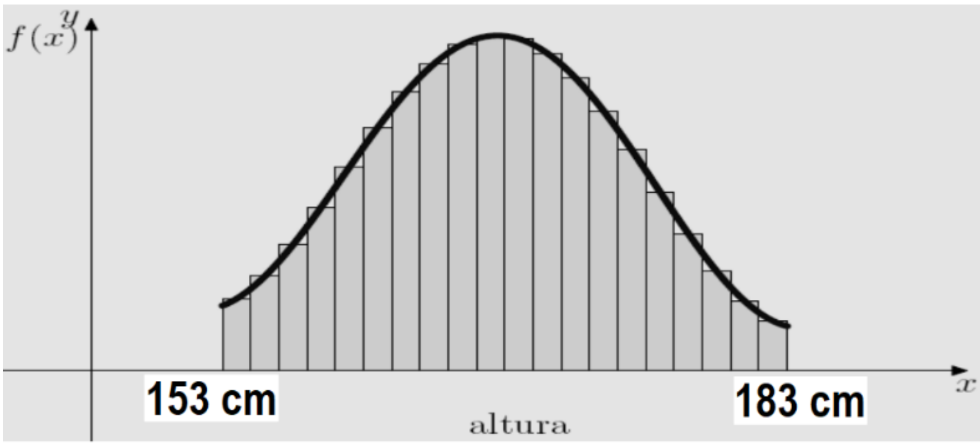
Los valores que toma una v.a. discreta X se suelen representar por $x_1, x_2, \ldots, x_n$.

VARIABLES ALEATORIAS CONTINUAS

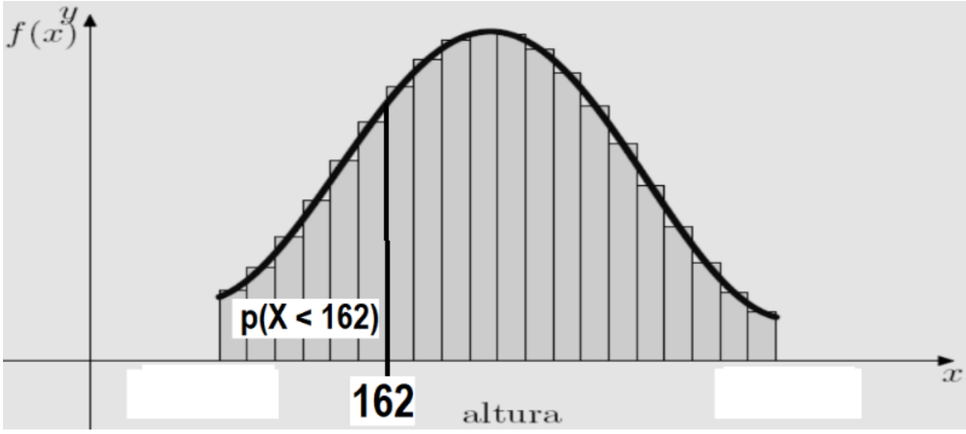
Concepto de variable aleatoria continua. Probabilidad.

Consideremos la v.a. continua X = "estatura, en cm, de los chicos de 17 años de una ciudad".
Supongamos que se sabe que dichas estaturas varían entre 183 cm, el más alto y 153 cm, el más bajo.

Si dibujamos el histograma de frecuencias el polígono de frecuencias se ajusta a una curva.
La función f(x) cuya gráfica es esa curva se llama *función de densidad de probabilidad* de X o simplemente curva de densidad.



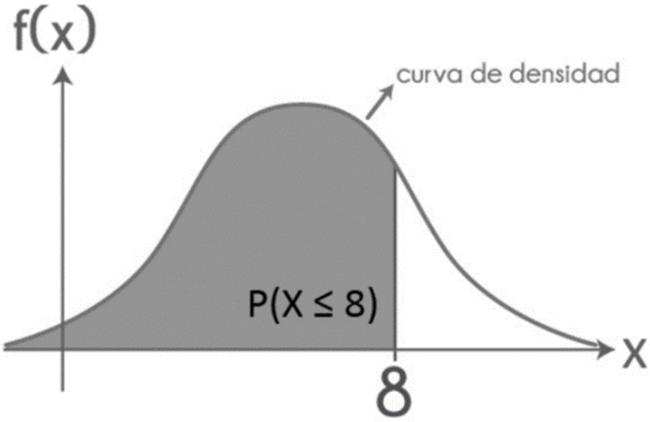
Usando el ejemplo anterior



se puede demostrar que, si se elige un chico al azar la probabilidad, por ejemplo, de que mida menos de 162 cm, $p(X < 162)$, corresponde al área entre la curva y el eje X, a la izquierda de 162.

De la misma forma $p(X > 162)$ sería el área entre la curva y el eje X a la derecha de 162.
Además, $p(X = 162)$ sería el área de segmento vertical que pasa por 162, o sea $p(X = 162) = 0$.
Según esto $p(X \leq 162) = p(X < 162) + p(X = 162) = p(X < 162) + 0 = p(X < 162)$.

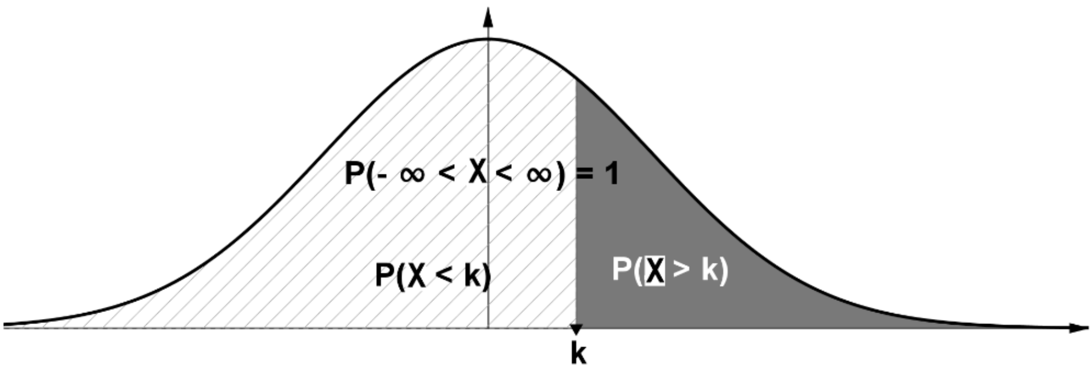
Otro ejemplo.



$p(X = 8) = 0$ $\qquad$ $p(X \leq 8) = p(X < 8) =$ área a la izquierda de 8

Observa que también se cumple que $p(-\infty < X < +\infty) = p(\text{del suceso seguro}) = 1$.
Es decir, el área entre la curva y el eje X siempre vale 1

En general, si X es una v.a. continua

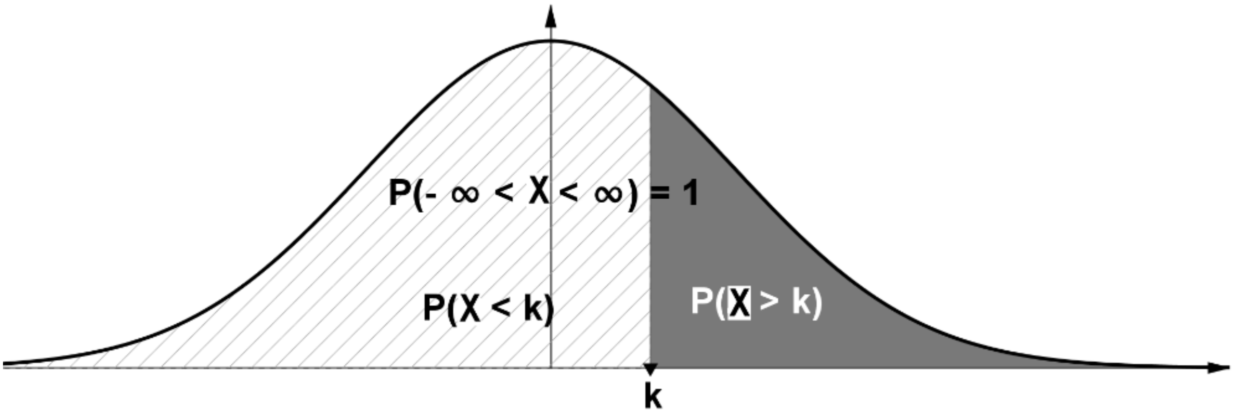


$p(X = k) = 0$ $\qquad$ $p(X < k) = p(X \leq k)$ $\qquad$ $p(-\infty < X < +\infty) = 1$

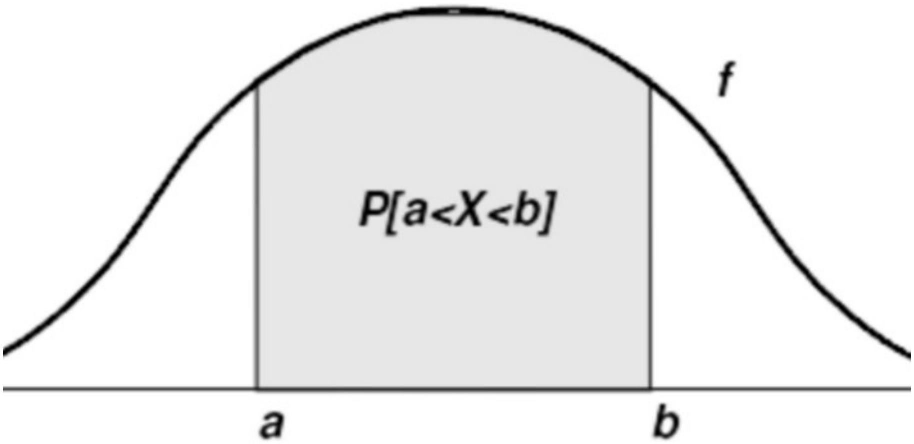
También se cumplen las siguientes propiedades:

1) $p(X > k) = 1 - p(X \leq k)$, por la propiedad de la probabilidad del suceso contrario.

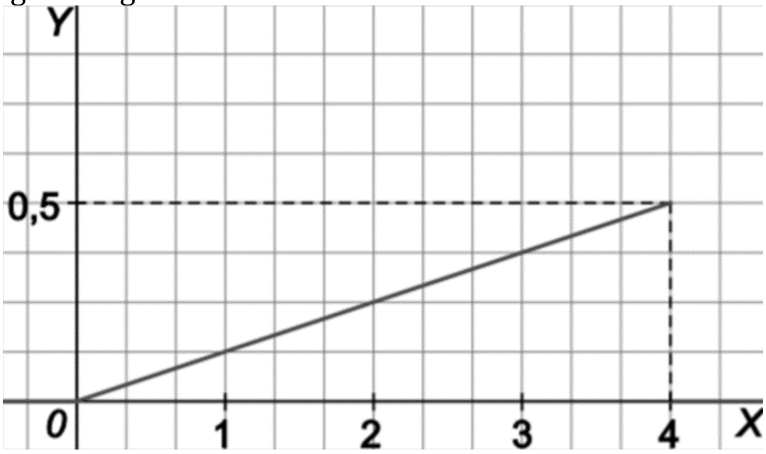
La interpretación gráfica es que el área a la derecha de k, igual al área total, que es 1, menos el área a la izquierda de k.



2) $p(a < X < b) = p(X < b) - p(X < a)$. La interpretación gráfica es que el área entre “a” y “b”, es el área a la izquierda de b menos el área a la izquierda de a.



Otro ejemplo:
Hallamos, a partir de la siguiente gráfica de la función de densidad de una v.a. continua X:



a) $p(X < 2)$
Es el área a la izquierda de 2, o sea el área del triángulo de base 2 y altura 0,25.

$$\text{Luego, } p(X < 2) = \frac{\text{base . altura}}{2} = \frac{2\ .\ 0,25}{2} = 0,25$$

b) $p(X > 2)$

$$p(X > 2 = = 1 - p(X \leq 2) = 1 - 0,25 = 0,75$$

c) $p(1 \leq X \leq 3)$
Es el área comprendida entre 1 y 3, o sea el área del trapecio de bases 0,125 y 0,375 y altura 2.

$$\text{Luego, } p(1 \leq X \leq 3) = \frac{(\text{base mayor} + \text{base menor})\ .\ \text{altura}}{2} = \frac{(0,375 + 0,125)\ .\ 2}{2} = 0,5$$

d) $p(X > 5)$ Es el área a la derecha de 5, o sea 0

e) $p(X > 3)$
Es el área a la derecha de 3, o sea el área del trapecio de bases 0,375 y 0,5 y altura 1.

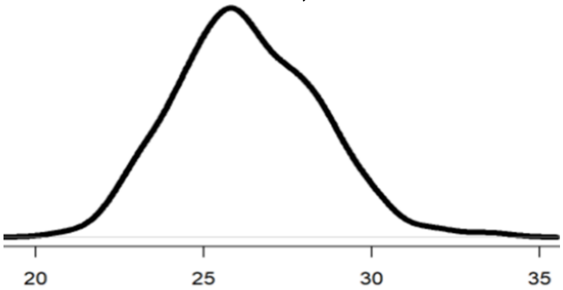
$$\text{Luego, } p(X > 3) = \frac{(\text{base mayor} + \text{base menor})\ .\ \text{altura}}{2} = \frac{(0,375 + 0,5)\ .\ 1}{2} = \frac{0,875}{2} = 0,4375$$

f) $p(X = 1)$ Es el área del segmento vertical que pasa por 1, o sea 0

g) $p(X \leq 0)$ Es el área a la izquierda de 0, o sea 0

Actividades resueltas

1) Esta es la curva de densidad de una v.a. continua X. Si se sabe que el área que hay a la izquierda de 30 vale 0,8 y el área que hay a la derecha de 25 es 0,72 determina:



a) $p(X > 30)$ **Resolución** $p(X > 30) = 1 - p(X \leq 30) = 1 - 0{,}8 = 0{,}2$

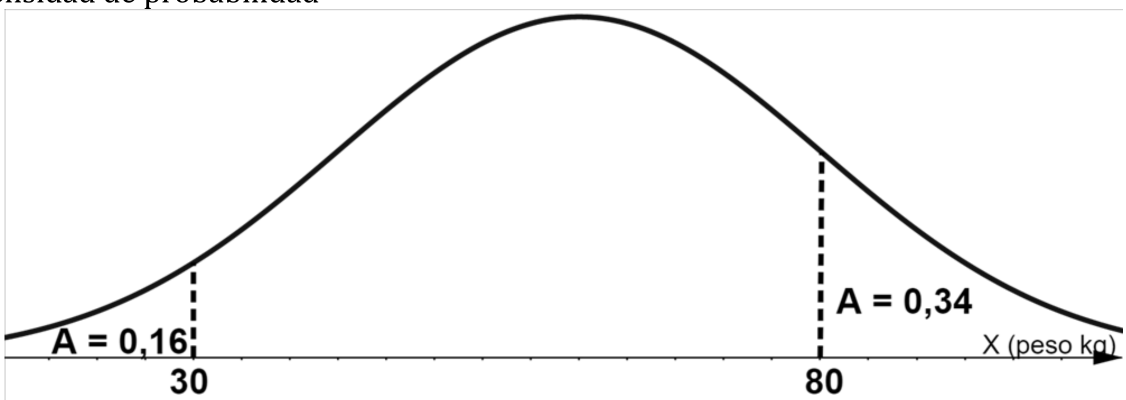
b) $p(X = 25)$ **Resolución** Es el área del segmento vertical que pasa por $x = 25$, o sea 0

c) $p(X < 25)$ **Resolución** $p(X < 25) = 1 - p(X \geq 25) = 1 - 0{,}72 = 0{,}28$

d) $p(25 < X < 30)$ **Resolución** $p(25 < X < 30) = p(X < 30) - p(X < 25) = 0{,}8 - 0{,}28 = 0{,}52$

e) Si además se sabe que $p(X > 20) = 0{,}93$, ¿cuál es el área que queda a la izquierda de 20?
Resolución Es $p(X < 20) = 1 - p(X \geq 20) = 1 - 0{,}93 = 0{,}07$

2) En un grupo de personas se está estudiando la variable aleatoria X = peso y se ha obtenido la siguiente curva de densidad de probabilidad



Se elige al azar una persona del grupo. Halla expresando el resultado en porcentaje:

a) La probabilidad de que el peso sea mayor de 80 kg
Resolución Es el área a la derecha de 80, o sea $0{,}34 = 34\%$

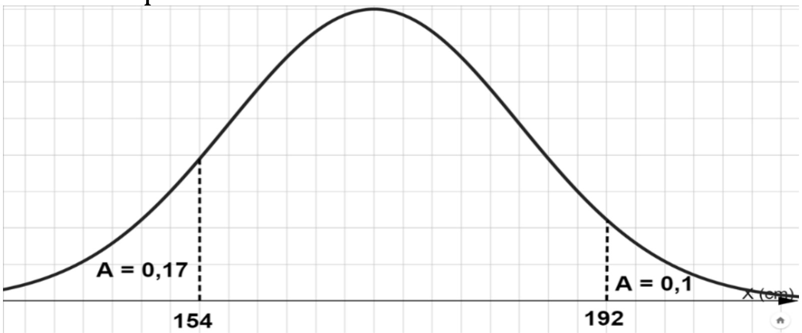
b) $p(-\infty < X < +\infty)$ **Resolución** Es el área entre la curva y el eje X, o sea $1 = 100\%$

c) La probabilidad de que el peso esté entre 30 kg y 80 kg
Resolución Es el área entre 30 y 80, o sea $1 - 0{,}16 - 0{,}34 = 0{,}5 = 50\%$

d) $p(X \leq 80)$ **Resolución** Es el área a la izquierda de 80, o sea $1 - 0{,}34 = 0{,}66 = 66\%$

e) $p(X = 30)$ **Resolución** Es el área del segmento vertical que pasa por 30, o sea $0 = 0\%$

3) En un grupo de personas se está estudiando la variable aleatoria X = estatura (en cm) y se ha obtenido la siguiente curva de densidad de probabilidad



Se elige al azar una persona del grupo. Halla expresando el resultado en porcentaje:

a) La probabilidad de que su estatura sea mayor de 192 cm

Resolución Es el área a la derecha de 192, o sea 0,1 = 10%

b) $p(-\infty < X < +\infty)$ **Resolución** Es el área entre la curva y el eje X, o sea 1 = 100%

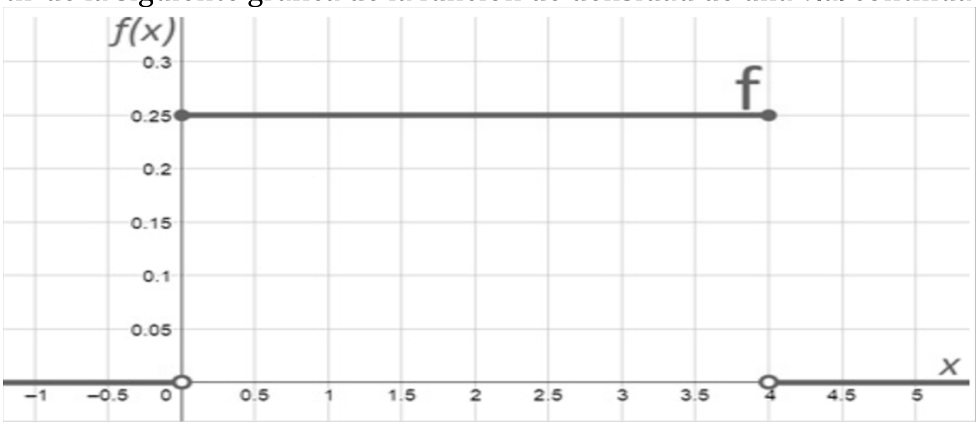
c) La probabilidad de que su estatura esté entre 154 cm y 192 cm

Resolución Es el área entre 154 y 192, o sea 1 – 0,17 – 0,1 = 0,73 = 73%

d) $p(X \geq 154)$ **Resolución** Es el área a la derecha de 154, o sea 1 – 0,17 = 0,83 = 83%

e) $p(X = 192)$ **Resolución** Es el área del segmento vertical que pasa por 192, o sea 0 = 0%

4) Calcular, a partir de la siguiente gráfica de la función de densidad de una v.a. continua X:



a) $p(X < 2)$

Resolución

Es el área a la izquierda de x = 2, o sea el área del rectángulo de base 2 y altura 0,25.

Luego, $p(X < 2) = 2 \cdot 0{,}25 = 0{,}5$

b) $p(1 \leq X \leq 3{,}5)$

Resolución Es el área comprendida entre x = 1, x = 3,5, o sea el área del rectángulo de base 2,5 y altura 0,25. Luego, $p(X < 2) = 2{,}5 \cdot 0{,}25 = 0{,}625$

c) $p(X > 4{,}5)$ **Resolución** Es el área a la derecha de x = 4,5, o sea 0

d) $p(X > 0{,}5)$

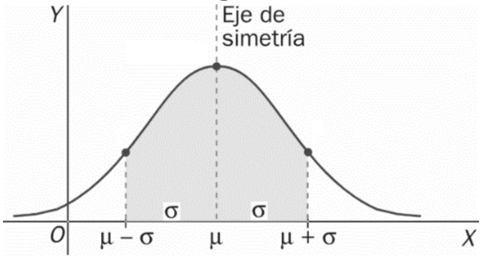
Resolución Es el área a la derecha de x = 0,5, o sea el área del rectángulo de base 3,5 y altura 0,25.

Luego, $p(X > 0{,}5) = 3{,}5 \cdot 0{,}25 = 0{,}875$

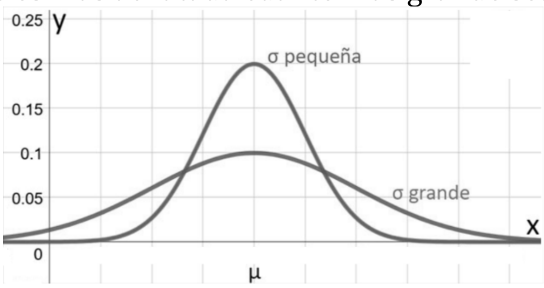
e) $p(X = 1{,}5)$ **Resolución** Es el área del segmento vertical que pasa por x = 1,5, o sea 0

Distribución normal, probabilidad

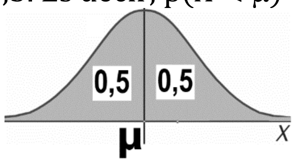
La mayoría de las v.a. continuas tienen una curva de densidad que tiene forma de campana, llamada **campana de Gauss**. La campana de Gauss tiene la siguiente forma:



Cuando para una v.a. X la curva de densidad tiene esta forma diremos que X tiene una **distribución normal** de media μ y desviación típica σ. Se escribe así: **$X \rightarrow N(\mu, \sigma)$**.
En la vida real nos podemos encontrar muy frecuentemente v.a. que tienen una distribución normal. Por ejemplo, altura de los individuos de una población, peso, temperatura ambiental de una ciudad, coeficiente intelectual, las calificaciones en un examen, etc.
La forma de la campana de Gauss es más achatada cuanto más grande sea la desviación típica

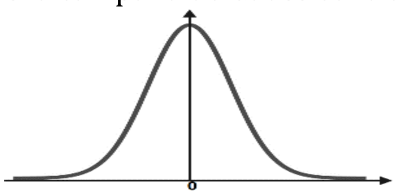


Como puedes observar la curva es simétrica respecto del eje de simetría, que es la recta vertical que pasa por μ. Luego, como el área entre la campana de Gauss y el eje X vale 1, el área a la izquierda de μ y a la derecha de μ son iguales y cada una vale 0,5. Es decir, $p(X < \mu) = p(X > \mu) = 0{,}5$



Distribución N(0,1), probabilidad

Si en una distribución normal **$X \rightarrow N(\mu, \sigma)$**, la media es $\mu = 0$ y la desviación típica es $\sigma = 1$, resulta entonces una distribución **N(0, 1)**.
Esta v.a. la representaremos por Z y se llama distribución normal tipificada, **$Z \rightarrow N(0, 1)$**.
La curva de densidad de la variable Z es la campana de Gauss centrada en 0



Probabilidades en la distribución N(0,1) usando una tabla

En la distribución normal $Z \longrightarrow N(0, 1)$, sabemos que $p(Z \leq k) = p(Z < k) =$ área a la izquierda de k entre la campana de Gauss y el eje horizontal.
Para el cálculo de dicha área se necesita una herramienta, que se llama integral que no es objeto de estudio en este curso. Sin embargo, existe una **tabla** donde se recogen todas esas áreas que corresponden a $p(Z < k)$, con 4 cifras decimales para k comprendido entre 0 y 3,49.

Para valores de k mayores que 3,49 tomaremos la probabilidad igual a 1.

La tabla de que nos indica $p(Z < k)$ para la variable $Z \rightarrow N(0, 1)$ es la siguiente:

k	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Uso de la tabla en sentido directo: usando la tabla se puede determinar $p(Z < k)$, con k entre 0 y 3,49.

Ejemplo:

Para hallar $p(Z < 1{,}24)$ buscamos en la 1ª columna las unidades,décimas, o sea 1,2 y en la 1ª fila las centésimas, o sea 0,04. La intersección de la fila con la columna nos da el valor 0,8925.
Por tanto, $p(Z < 1{,}24) = 0{,}8925$

Uso de la tabla en sentido inverso: también se puede usar la tabla en sentido inverso para hallar el valor de k.

Ejemplos:

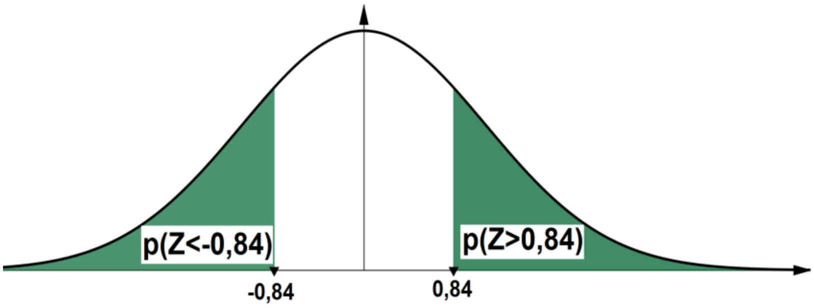
1) Hallar k sabiendo que $p(Z < k) = 0{,}7190$. Buscamos 0,7190 dentro de la tabla y vemos que le corresponde 0,5 (en la 1ª columna) y 0,08 (en la 1ª fila). Por tanto, $k = 0{,}58$

2) Hallar k sabiendo que $p(Z < k) = 0{,}6738$. Buscamos 0,6738 dentro de la tabla y vemos que no está pero 0,6738 está muy próximo a 0,6736 que le corresponde 0,4 (en la 1ª columna) y 0,05 (en la 1ª fila). Por tanto, $k = 0{,}45$

3) Hallar k sabiendo que $p(Z < k) = 0{,}9376$. En este caso 0,9376 no está en la tabla. Los valores más próximos son 0,9370, que corresponde a $k_1 = 1{,}53$, y 0,9382, que corresponde a $k_2 = 1{,}54$.
Como 0,9376 está prácticamente a la misma distancia de los valores encontrados, tomaremos como valor de k el punto medio de k_1 y k_2, es decir $k = (1{,}53+1{,}54)/2 = 1{,}535$

Uso de la simetría de la campana de Gauss para calcular probabilidades en la N(0, 1)

1) Cálculo de la probabilidad de que Z sea menor que un nº negativo. Por ejemplo, p(Z < – 0,84)



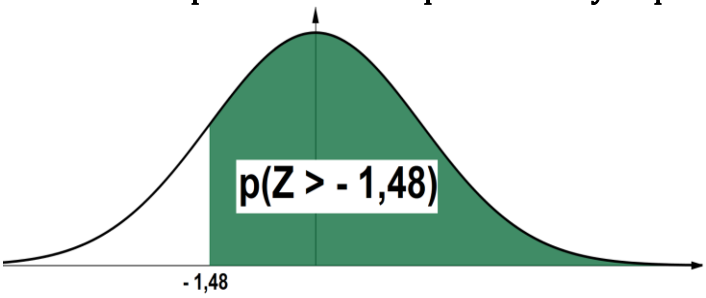
$$p(Z < -0,84) = \text{área a la izda de } -0,84 \xrightarrow{\text{por simetría}} \text{área a la dcha de } 0,84 = p(Z > 0,84)$$

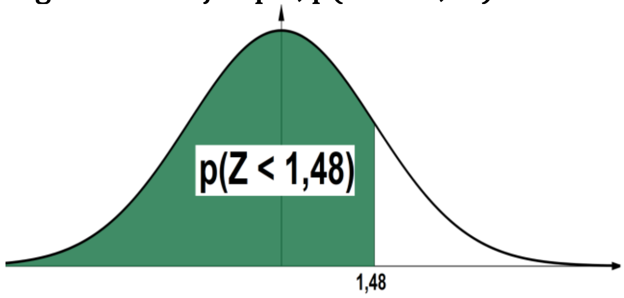
Por la propiedad de la probabilidad del suceso contrario

$$p(Z > 0,84) = 1 - \underbrace{p(Z \leq 0,84)}_{\text{viene en la tabla}} = 1 - 0,7995 = 0,2005$$

En general, $p(Z < -k) = p(Z > k) = 1 - p(Z \leq k)$

2) Cálculo de la probabilidad de que Z sea mayor que un nº negativo. Por ejemplo, p(Z > – 1,48)

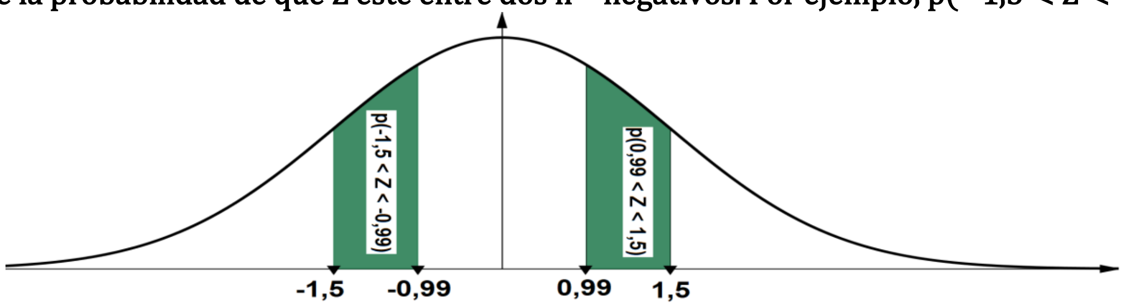




$$p(Z > -1,48) = \text{área a la dcha de } -1,48 \xrightarrow{\text{por simetría}} \text{área a la izda de } 1,48 = \underbrace{p(Z < 1,48)}_{\text{viene en la tabla}} = 0,9306$$

En general, $p(Z > -k) = p(Z < k)$

3) Cálculo de la probabilidad de que Z esté entre dos nºs negativos. Por ejemplo, p(– 1,5 < Z < – 0,99)



$$p(-1,5 < Z < -0,99) = \text{área entre} -1,5 \text{ y } -0,99 \xrightarrow{\text{por simetría}} \text{área entre } 0,99 \text{ y } 1,5$$

$$\text{Es decir, } p(0,99 < Z < 1,5) = \underbrace{p(Z < 1,5)}_{\text{en la tabla}} - \underbrace{p(Z < 0,99)}_{\text{en la tabla}} = 0,9332 - 0,8389 = 0,0943$$

En general, $p(-b < Z < -a) = p(a < Z < b) = p(Z < b) - p(Z < a)$

4) Cálculo de la probabilidad de que Z esté entre un nº negativo y otro positivo.
Por ejemplo, p(– 1 < Z < 2):

Podemos hallarla usando las propiedades ya vistas: $p(-1 < Z < 2) = p(Z < 2) - p(Z < -1) =$

$$= p(Z < 2) - p(Z > 1) = p(Z < 2) - [1 - p(Z < 1)] = \underbrace{p(Z < 2)}_{\text{en la tabla}} + \underbrace{p(Z < 1)}_{\text{en la tabla}} - 1 = 0,8413 + 0,9772 - 1 = 0,8185$$

En general, $p(-a < Z < b) = p(Z < a) + p(Z < b) - 1$

Actividades resueltas

1) Usando la tabla de la distribución normal N(0,1) y las propiedades de la probabilidad, calcula las siguientes probabilidades:

a) $p(Z < 2{,}78)$ **Resolución** $0{,}9973$

b) $p(Z < -1{,}35)$ **Resolución** $p(Z > 1{,}35) = 1 - p(Z < 1{,}35) = 1 - 0{,}9115 = 0{,}0885$

c) $p(Z > -1{,}27)$ **Resolución** $p(Z < 1{,}27) = 0{,}8980$

d) $p(0{,}7 < Z < 1{,}29)$ **Resolución** $p(Z < 1{,}29) - p(Z < 0{,}7) = 0{,}9015 - 0{,}7580 = 0{,}1435$

e) $p(-3{,}07 < Z < 2{,}77)$ **Resolución** $p(Z < 3{,}07) + p(Z < 2{,}77) - 1 = 0{,}9989 + 0{,}9972 - 1 = 0{,}9961$

f) $p(Z = 2{,}73)$ **Resolución** 0

g) $p(-2{,}25 < Z < -1{,}49)$ **Resolución** $p(Z < 2{,}25) - p(Z < 1{,}49) = 0{,}9878 - 0{,}9319 = 0{,}0559$

h) $p(-0{,}39 < Z < 1{,}1)$ **Resolución** $p(Z < 0{,}39) + p(Z < 1{,}1) - 1 = 0{,}6517 + 0{,}8643 - 1 = 0{,}516$

i) $p(Z > 1)$ **Resolución** $1 - p(Z < 1) = 1 - 0{,}8413 = 0{,}1587$

j) $p(Z < 2{,}17)$ **Resolución** $0{,}9846$

k) $p(Z < 1{,}46)$ **Resolución** $0{,}9279$

l) $p(Z > 0{,}83)$ **Resolución** $1 - p(Z \leq 0{,}83) = 1 - 0{,}7967 = 0{,}2033$

m) $p(1{,}45 < Z < 2{,}03)$ **Resolución** $p(Z < 2{,}03) - p(Z < 1{,}45) = 0{,}9788 - 0{,}9265 = 0{,}0523$

n) $p(Z > -2)$ **Resolución** $p(Z < 2) = 0{,}9772$

ñ) $p(-1{,}74 < Z < 2{,}6)$ **Resolución** $p(Z < 1{,}74) + p(Z < 2{,}6) - 1 = 0{,}9591 + 0{,}9953 - 1 = 0{,}9544$

o) $p(Z < -0{,}38)$ **Resolución** $p(Z > 0{,}38) = 1 - p(Z \leq 0{,}38) = 1 - 0{,}6480 = 0{,}352$

p) $p(Z > 1{,}53)$ **Resolución** $1 - p(Z \leq 1{,}53) = 1 - 0{,}9370 = 0{,}0{,}063$

q) $p(0{,}3 < Z < 2{,}44)$ **Resolución** $p(Z < 2{,}44) - p(Z < 0{,}3) = 0{,}9927 - 0{,}6179 = 0{,}3748$

r) $p(Z > -1{,}3)$ **Resolución** $p(Z < 1{,}3) = 0{,}9032$

s) $p(-1{,}06 < Z < 2{,}15)$ **Resolución** $p(Z < 1{,}06) + p(Z < 2{,}15) - 1 = 0{,}8554 + 0{,}9842 - 1 = 0{,}8396$

t) $p(Z < -1{,}29)$ **Resolución** $p(Z > 1{,}29) = 1 - p(Z \leq 1{,}29) = 1 - 0{,}9015 = 0{,}0985$

2) Usando la tabla de la distribución N(0, 1) halla el valor de k que cumple:

a) $p(Z < k) = 0{,}88$ **Resolución** $k = 1{,}175$

b) $p(Z \leq k) = 0{,}7823$ **Resolución** $k = 0{,}78$

c) $p(Z \leq k) = 0{,}8485$ **Resolución** $k = 1{,}03$

d) $p(Z < k) = 0{,}9612$ **Resolución** $k = 1{,}765$

e) $p(Z < k) = 0{,}7486$ **Resolución** $k = 0{,}67$

f) $p(Z < k) = 0{,}9981$ **Resolución** $k = 2{,}89$

g) $p(Z < k) = 0{,}9844$ **Resolución** $k = 2{,}155$

h) $p(Z < k) = 0{,}6718$ **Resolución** $k = 0{,}445$

i) $p(Z < k) = 0{,}7358$ **Resolución** $k = 0{,}63$

j) $p(Z < k) = 0{,}9417$ **Resolución** $k = 1{,}57$

Tipificación

Se puede demostrar que si una v.a. continua $X \to N(\mu, \sigma)$, entonces la variable $Z = \frac{X-\mu}{\sigma} \to N(0, 1)$.

Al proceso de restarle a X la media y luego dividir entre la desviación típica se le llama **tipificación** de la variable X.

La tipificación nos permite hallar probabilidades en una distribución $X \to N(\mu, \sigma)$ transformándola en una distribución $Z \to N(0, 1)$.

Actividades resueltas

1) Si $X \to N(5, 4)$, entonces $Z = \frac{X-5}{4} \to N(0, 1)$. Halla:

a) $p(X \leq 8)$ **Resolución** $p\left(\frac{X-5}{4} \leq \frac{8-5}{4}\right) = p(Z \leq 0{,}75) = 0{,}7734$

b) $p(1 < X < 7)$

Resolución

$$p\left(\frac{1-5}{4} < \frac{X-5}{4} < \frac{7-5}{4}\right) = p(-1 < Z < 0{,}5) = p(Z < 1) + p(Z < 0{,}5) - 1 = 0{,}8413 + 0{,}6915 - 1 = 0{,}5328$$

2) El peso de los toros de una determinada ganadería se distribuye normalmente con media 500 kg y desviación típica 45 kg. Si la ganadería tiene 2000 toros, calcula cuántos toros pesarán:

a) más de 540 kg

Resolución

$$X = peso \qquad X \to N(500, 45) \Rightarrow Z = \frac{X-500}{45} \to N(0, 1)$$

Hallemos $p(X > 540)$ que será el porcentaje de toros que pesan más de 540 kg:

$$p(X > 540) \xrightarrow[\text{dividimos entre 45}]{\text{restamos 500 y}} p\left(\frac{X-500}{45} > \frac{540-500}{45}\right) = p(Z > 0{,}89) = 1 - p(Z \leq 0{,}89) = 1 - 0{,}8133 = 0{,}1867 = 18{,}67\%$$

Luego, el nº de toros que pesan más de 540 kg es 18,67% de 2000 = 373,4 ~ 373 toros

b) menos de 480 kg

Resolución

Hallemos $p(X < 480)$ que será el porcentaje de toros que pesan menos de 480 kg:

$$p(X < 480) \xrightarrow[\text{dividimos entre 45}]{\text{restamos 500 y}} p\left(\frac{X-500}{45} < \frac{480-500}{45}\right) = p(Z < -0{,}45) = p(Z > 0{,}45) =$$

$$= 1 - p(Z \leq 0{,}45) = 1 - 0{,}6736 = 0{,}3264 = 32{,}64\%$$

Luego, el nº de toros que pesan menos de 480 kg es 32,64% de 2000 = 652,8 ~ 653 toros

c) entre 490 y 510 kg

Resolución

Hallemos p(490 < X < 510) que será el porcentaje de toros que pesan entre 490 kg y 510 kg:

$$p(490 < X < 510) \xrightarrow[\text{dividimos entre 45}]{\text{restamos 500 y}} p\left(\frac{490-500}{45} < \frac{X-500}{45} < \frac{510-500}{45}\right) = p(-0,22 < Z < 0,22) =$$

$$= p(Z < 0,22) + p(Z < 0,22) - 1 = 2\ .\ 0,5871 - 1 = 0,1742 = 17,42\%$$

Luego, el nº de toros que pesan entre 490 kg y 510 kg es 17,42% de 2000 = 348,4 ~ 348 toros

3) El tiempo, en minutos, que tiene que esperar un cliente para ser atendido en un Banco A sigue una distribución normal de media $\mu = 9$ minutos y desviación típica $\sigma = 1$ minuto, mientras que en otro Banco B sigue una distribución normal de media $\mu = 8,5$ minutos y varianza $\sigma^2 = 4$ min^2.

Si un cliente tiene que hacer una gestión bancaria y sólo dispone de 10 minutos, ¿en qué Banco es más probable que le atiendan antes de los 10 minutos?

Resolución

$$X = tiempo\ en\ \text{A} \qquad \text{X} \to N(9,\ 1) \ \Rightarrow \ \text{Z} = \frac{X-9}{1} \to N(0,\ 1)$$

$$Y = tiempo\ en\ B \qquad Y \to N(8,5\ ;\sqrt{4}) = N(8,5\ ;2) \ \Rightarrow \ \text{Z} = \frac{Y-8,5}{2} \to N(0,\ 1)$$

$$\begin{cases} p(de\ que\ lo\ atiendan\ en\ A) = p(X < 10) = p\left(\frac{X-9}{1} < \frac{10-9}{1}\right) = p(Z < 1) = 0,8413 = 84,13\% \\ p(de\ que\ lo\ atiendan\ en\ B) = p(Y < 10) = p\left(\frac{Y-8,5}{2} < \frac{10-8,5}{2}\right) = p(Z < 0,75) = 0,7734 = 77,34\% \end{cases}$$

$Luego$, es más probable $que\ lo\ atiendan$ en $el\ Banco$ A

4) Una v.a X sigue una distribución normal N(8, 2), calcula:

a) $p(X \le 10)$

Resolución

$$Como\ \text{X} \to N(8,\ 2) \ \Rightarrow \ \text{Z} = \frac{X-8}{2} \to N(0,\ 1)$$

$$p(X \le 10) = p\left(\frac{X-8}{2} \le \frac{10-8}{2}\right) = p(Z \le 1) = 0,8413$$

b) $p(X \ge 5)$

Resolución

$$p(X \ge 5) = p\left(\frac{X-8}{2} \ge \frac{5-8}{2}\right) = p(Z \ge -1,5) = p(Z \le 1,5) = 0,9332$$

c) $p(X < 7)$

Resolución

$$p(X < 7) = p\left(\frac{X-8}{2} < \frac{7-8}{2}\right) = p(Z < -0,5) = p(Z > 0,5) = 1 - p(Z \le 0,5) = 1 - 0,6915 = 0,3085$$

d) $p(4 \le X < 9)$

Resolución

$$p(4 \le X < 9) = p\left(\frac{4-8}{2} \le \frac{X-8}{2} < \frac{9-8}{2}\right) = p(-2 \le Z < 0,5) = p(Z < 2) + p(Z < 0,5) - 1 = 0,9772 + 0,6915 - 1 = 0,6687$$

e) $p(7 < X \le 8)$

Resolución

$$p(7 < X \le 8) = p\left(\frac{7-8}{2} \le \frac{X-8}{2} < \frac{8-8}{2}\right) = p(-0,5 \le Z < 0) = p(0 < Z < 0,5) = p(Z < 0,5) + p(Z < 0) = 0,6915 - 0,5 = 0,1915$$

5) Los pesos de los habitantes de una ciudad se distribuyen normalmente, con media 70 kg y desviación típica 5 kg. Determina:

a) La probabilidad de que el peso de un individuo elegido al azar esté entre 60 y 65.

Resolución

$$X = peso \qquad X \to N(70,\ 5) \ \Rightarrow \ Z = \frac{X-70}{5} \to N(0,\ 1)$$

$$p(60 < X < 65) = p\left(\frac{60-70}{5} < \frac{X-70}{5} < \frac{65-70}{5}\right) = p(-2 < Z > -1) = p(1 < Z > 2) =$$

$$= p(Z < 2) - p(Z < 1) = 0,9772 - 0,8413 = 0,1359$$

b) El tanto por ciento de habitantes que pesa más de 68 kg?

Resolución

Para saber el % hallemos primero $p(X > 68)$ para saber el % de toros que pesan menos de 480 kg:

$$p(X > 68) = p\left(\frac{X-70}{5} > \frac{68-70}{5}\right) = p(Z > -0,4) = p(Z < 0,4) = 0,6554 = 65,54\%$$

c) El número de personas que pesa menos de 80,6 kg, sabiendo que la ciudad tiene 8000 habitantes

Resolución

Hallemos primero $p(X < 80,6)$ para saber el % de habitantes que pesan menos de 80,6 kg:

$$p(X < 80,6) = p\left(\frac{X-70}{5} < \frac{80,6-70}{5}\right) = p(Z < 2,12) = 0,983 = 98,3\%$$

Luego, el nº de habitantes es 98,3% de 8000 = 7864 habitantes

6) En un proceso fotográfico, el tiempo de revelado de las fotos puede considerarse una variable aleatoria normal para la cual μ = 12,9 minutos y σ = 2 minutos. Determina el porcentaje de probabilidad de que el tiempo de revelado …
a) tarde entre 16 y 16,5 minutos

Resolución

$$X = tiempo\ de\ revelado \qquad \mathrm{X} \to N(12{,}9\,;2) \Rightarrow \quad \mathrm{Z} = \frac{X-12{,}9}{2} \to N(0,\,1)$$

$$p(16 < X < 16{,}5) = p\left(\frac{16-12{,}9}{2} < \frac{X-12{,}9}{2} < \frac{16{,}5-12{,}9}{2}\right) = p(1{,}55 < Z > 1{,}8) =$$

$$= p(Z < 1{,}8) - p(Z < 1{,}55) = 0{,}9641 - 0{,}9394 = 0{,}0247 = 2{,}47\%$$

b) se lleve al menos 16,2 minutos

Resolución

$$\mathrm{p}(\mathrm{X} \geq 16{,}2) = \mathrm{p}\left(\frac{\mathrm{X}-12{,}9}{2} \geq \frac{16{,}2-12{,}9}{2}\right) = \mathrm{p}(\mathrm{Z} \geq 1{,}65) = 1 - \mathrm{p}(\mathrm{Z} < 1{,}65) = 1 - 0{,}9505 = 0{,}0495 = 4{,}95\%$$

c) se lleve a lo más 11,5 minutos

Resolución

$$p(X \leq 11{,}5) = p\left(\frac{X-12{,}9}{2} \leq \frac{11{,}5-12{,}9}{2}\right) = p(Z \leq -0{,}7) = p(Z \geq 0{,}7) =$$

$$= 1 - p(Z < 0{,}7) = 1 - 0{,}758 = 0{,}242 = 24{,}2\%$$

7) Se sabe que la distribución de las calificaciones de un curso de 1ª de Bachillerato sigue una distribución normal de media 6,2 puntos y desviación típica 0,8. Calcula el % de probabilidad de que la nota de un estudiante del curso elegido al azar esté entre 5 y 6

Resolución

$$\mathrm{X} = \text{nota del alumno}, \quad \mathrm{X} \to \mathrm{N}(6{,}2;\,0{,}8) \qquad \mathrm{Z} = \frac{\mathrm{X}-6{,}2}{0{,}8} \to \mathrm{N}(0,\,1)$$

$$\text{Nos piden } \mathrm{p}(5 < \mathrm{X} < 6) = \mathrm{p}\left(\frac{5-6{,}2}{0{,}8} < \frac{\mathrm{X}-6{,}2}{0{,}8} < \frac{6-6{,}2}{0{,}8}\right) = \mathrm{p}(-1{,}5 < \mathrm{Z} < -0{,}25) =$$

$$= \mathrm{p}(0{,}25 < \mathrm{Z} < 1{,}5) = \mathrm{p}(\mathrm{Z} < 1{,}5) - \mathrm{p}(\mathrm{Z} < 0{,}25) = 0{,}9332 - 0{,}5987 = 0{,}3345 = 33{,}45\%$$

8) Se sabe que el nivel de potasio presente en la sangre de una persona adulta se distribuye normalmente con media 3,8 y desviación típica 0,2. Se elige al azar una persona adulta.
Calcula el % de probabilidad de que su nivel de potasio esté entre 3,5 y 3,7

Resolución

$$\mathrm{X} = \text{nivel de potasio}, \quad \mathrm{X} \to \mathrm{N}(3{,}8;\,0{,}2) \qquad \mathrm{Z} = \frac{\mathrm{X}-3{,}8}{0{,}2} \to \mathrm{N}(0,\,1)$$

$$\text{Nos piden } \mathrm{p}(3{,}5 < \mathrm{X} < 3{,}7) = \mathrm{p}\left(\frac{3{,}5-3{,}8}{0{,}2} < \frac{\mathrm{X}-3{,}8}{0{,}2} < \frac{3{,}7-3{,}8}{0{,}2}\right) = \mathrm{p}(-1{,}5 < \mathrm{Z} < -0{,}5) =$$

$$= \mathrm{p}(0{,}5 < \mathrm{Z} < 1{,}5) = \mathrm{p}(\mathrm{Z} < 1{,}5) - \mathrm{p}(\mathrm{Z} < 0{,}5) = 0{,}9332 - 0{,}6915 = 0{,}2417 = 24{,}17\%$$

VARIABLES ALEATORIAS DISCRETAS

Concepto, probabilidad

Vamos a ver cómo se calculan probabilidades cuando tenemos una v.a. discreta, X.
Al conjunto de las probabilidades le llamaremos distribución de probabilidad.

Supongamos el experimento aleatorio de lanzar un dado dos veces.
Los resultados del experimento los podemos obtener mediante esta tabla:

	1	2	3	4	5	6
1	**11**	**12**	**13**	**14**	**15**	**16**
2	**21**	**22**	**23**	**24**	**25**	**26**
3	**31**	**32**	**33**	**34**	**35**	**36**
4	**41**	**42**	**43**	**44**	**45**	**46**
5	**51**	**52**	**53**	**54**	**55**	**56**
6	**61**	**62**	**63**	**64**	**65**	**66**

Sea la v.a. X = "nº de veces que sale el 6". Podemos observar que X es una v.a. discreta, porque toma valores aislados, $X = 0$, 1 y 2

La tabla de distribución de probabilidad de esta v.a. es:

x_i	$p_i = p(X = x_i)$
0	$p(X = 0) = p(\text{de que no salga el } 6) = \frac{25}{36} \sim 0{,}6944 = 69{,}44\%$
1	$p(X = 1) = p(\text{de que el 6 salga 1 vez}) = \frac{10}{36} \sim 0{,}2778 = 27{,}78\%$
2	$p(X = 2) = p(\text{de que el 6 salga 2 veces}) = \frac{1}{36} \sim 0{,}0278 = 2{,}78\%$
Total	$\sum p_i = 1 = 100\%$

Observa que la suma de las probabilidades es 1, $\sum p_i = 1$

Conociendo la distribución de probabilidad podemos calcular cualquier probabilidad.

Por ejemplo, si queremos calcular la probabilidad de que salga el 6 menos de 2 veces:

$$p(X < 2) = p(X = 0) + p(X = 1) = \frac{25}{36} + \frac{10}{36} = \frac{35}{36} \cong 0{,}9722 = 97{,}22\%$$

Otra forma de hallar la probabilidad es usando la del suceso contrario:

$$p(X < 2) = 1 - p(X = 2) = 1 - \frac{1}{36} = \frac{35}{36} \cong 0{,}9722 = 97{,}22\%$$

En general, se llama <u>distribución de probabilidad</u> de una v.a. discreta X que toma los valores $x_1, x_2, \ldots, x_n$ al conjunto de probabilidades: $p_1 = p(X = x_1)$, $p_2 = p(X = x_2)$, ... $p_n = p(X = x_n)$.

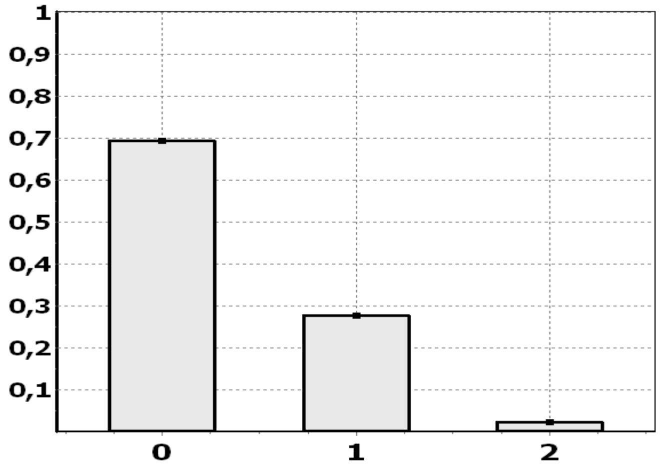
Se suelen representar mediante una <u>*tabla de distribución de probabilidades*</u> como la siguiente:

x_i	$p_i = p(X = x_i)$
x_1	p_1
x_2	p_2
....	....
x_n	p_n
Total	$\sum p_i = 1$

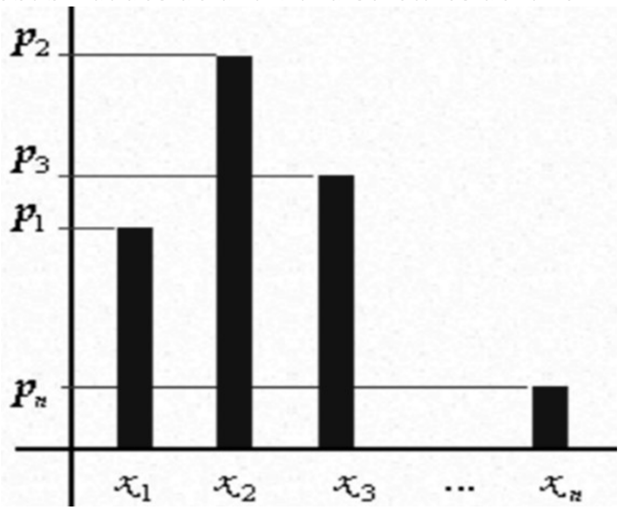
A partir de la tabla de distribución de probabilidad se puede dibujar un diagrama de barras, llamado <u>*gráfico de probabilidades*</u>, representando en el eje horizontal los valores x_i y en el eje vertical los valores p_i.

El gráfico de probabilidades de la v.a. X del ejemplo anterior es

x_i	p_i
0	0,6944
1	0,2778
2	0,0278



En general, el gráfico de de probabilidades de una v.a. discreta es de la forma



Parámetros de una distribución discreta

Los parámetros más utilizados en una v.a. discreta son:

Media aritmética o esperanza matemática: Se representa por μ y es la suma de todos los valores $x_i p_i$

$$\boxed{\mu = \sum x_i\, p_i}$$

Varianza: Se representa por σ^2 y es la suma de todos los valores $x_i^2 p_i$

$$\boxed{\sigma^2 = \Sigma x_i^2 p_i - \mu^2}$$

Desviación típica: Se representa por σ y es la raíz cuadrada de la varianza

$$\boxed{\sigma = \sqrt{\sigma^2}}$$

Calculemos los parámetros en el ejemplo expuesto anteriormente de lanzar un dado dos veces y la v.a. X = nº de veces que sale el 6:

x_i	p_i	$x_i\, p_i$	$x_i^2 p_i$
0	$\frac{25}{36}$	0	0
1	$\frac{10}{36}$	$\frac{10}{36}$	$\frac{10}{36}$
2	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{4}{36}$
Total	1	$\frac{12}{36} = \frac{1}{3}$	$\frac{14}{36} = \frac{7}{18}$

Media o esperanza matemática de X: $\mu = \sum x_i\, p_i = \frac{12}{36} = \frac{1}{3} \cong 0{,}3333$

Varianza de X: $\sigma^2 = \Sigma x_i^2 p_i - \mu^2 = \frac{7}{18} - \left(\frac{1}{3}\right)^2 = \frac{5}{18} \cong 0{,}2778 = \frac{7}{18} - \left(\frac{1}{3}\right)^2 = \frac{7}{18} - \frac{1}{9} = \frac{5}{18} \sim 0{,}278$

Desviación típica de X: $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{5}{18}} \cong 0{,}527$

En general, para calcular los parámetros de una v.a. discreta se puede completar una tabla como la siguiente:

x_i	p_i	$x_i p_i$	$x_i^2 p_i$
...	...	...	...
total	1	...	...

Actividades resueltas

1) Una compañía aérea dispone de un avión de 200 plazas. Sea X la variable aleatoria discreta que expresa el número de viajeros que va al aeropuerto para viajar en el avión. Nos dicen que X = 198, 199, 200, 201, 202, 203, 204, 205 y nos dan su tabla de distribución de probabilidad:

x_i	198	199	200	201	202	203	204	205
$p_i = p(X = x_i)$	0,05	0,09	0,15	0,20	0,23	0,17	0,09	0,02

a) Hallar la probabilidad de que todos viajeros que van al aeropuerto tengan plaza.

Resolución

Para que tengan plaza el nº de viajeros debe ser menor o igual a 200. Luego, nos piden

$$p(X \leq 200) = 0,05 + 0,09 + 0,15 = 0,29 = 29\%$$

b) Obtener la probabilidad de que se queden sin plaza al menos tres de los viajeros.

Resolución

Nos piden $p(X \geq 203) = 0,17 + 0,09 + 0,02 = 0,28 = 28\%$

c) Calcular la probabilidad de que el número viajeros que va al aeropuerto sea un número par.

Resolución

Nos piden $p(X = 198) + p(X = 200) + p(X = 202) + p(X = 204) = 0,05 + 0,15 + 0,23 + 0,09 = 52\%$

d) Calcular el número medio o esperado de viajeros que acude al aeropuerto y la desviación típica de X.

Resolución

Calculemos los parámetros μ y σ:

x_i	p_i	$x_i\,p_i$	$x_i^2\ p_i$
198	0,05	9,9	3528,36
199	0,09	17,91	3564,09
200	0,15	30	6000
201	0,2	40,2	8080,2
202	0,23	46,46	9384,92
203	0,17	34,51	7005,53
204	0,09	18,36	3745,44
205	0,02	4,1	840,5
Total	1	201,44	42149,04

Media o esperanza matemática de X: $\mu = \sum x_i p_i = 201,44 \simeq 201$ viajeros

Desviación típica de X: $\sigma = \sqrt{\sigma^2} = \sqrt{\sum x_i^2\, p_i - \mu^2} = \sqrt{42149,04 - 201,44^2} = \sqrt{1570,9664} \cong 39,6354$

2) Se lanza dos veces un dado tetraédrico con las caras numeradas del 1 al 4 y sea X la diferencia positiva de los números de las caras ocultas.

a) Elabora la tabla de distribución de probabilidad.

Resolución

Hallemos primero el espacio muestral, E = {11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 33, 34, 41, 42, 43, 44}

La tabla de la distribución de probabilidad es:

x_i	$p_i = p(X = x_i)$
0	$p(X = 0) = p(\text{de que la diferencia de los puntos sea } 0) = \frac{4}{16} = \frac{1}{4} = 0{,}25 = 25\%$
1	$p(X = 1) = p(\text{de que la diferencia de los puntos sea } 1) = \frac{6}{16} = \frac{3}{8} = 0{,}375 = 37{,}5\%$
2	$p(X = 2) = p(\text{de que la diferencia de los puntos sea } 2) = \frac{4}{16} = \frac{1}{4} = 0{,}25 = 25\%$
3	$p(X = 3) = p(\text{de que la diferencia de los puntos sea } 3) = \frac{2}{16} = \frac{1}{8} = 0{,}125 = 12{,}5\%$
Total	$\sum p_i = 1 = 100\%$

b) Calcula $p(X \geq 2)$

Resolución

Nos piden $p(X = 2) + p(X = 3) = 0{,}25 + 0{,}125 = 0{,}375 = 37{,}5\%$

c) Halla la probabilidad de que la diferencia entre las caras esté comprendida entre 0 y 3

Resolución

Nos piden $p(X = 1) + p(X = 2) = 0{,}25 + 0{,}375 = 0{,}625 = 62{,}5\%$

d) Calcula μ y σ

Resolución

Calculemos los parámetros μ y σ:

x_i	p_i	$x_i\,p_i$	$x_i^2\,p_i$
0	0,25	0	0
1	0,375	0,375	0,375
2	0,25	0,5	1
3	0,125	0,375	1,125
Total	1	1,25	2,5

Media o esperanza matemática de X: $\mu = \sum x_i p_i = 1{,}25$

Desviación típica de X: $\sigma = \sqrt{\sigma^2} = \sqrt{\sum x_i^2\,p_i - \mu^2} = \sqrt{2{,}5 - 1{,}25^2} = \sqrt{0{,}9375} \cong 0{,}9682$

3) Sea X la variable aleatoria que expresa el número de personas que habitan en una vivienda de un determinado barrio elegida al azar. Su distribución de probabilidad viene recogida en la siguiente tabla:

x_i	1	2	3	4	5	6	7
p_i	0,23	0,322	0,186	0,155	0,067	0,024	0,016

Elegimos una vivienda

a) Hallar la probabilidad de que el nº de personas que viven en un hogar sea menor que cuatro.

Resolución

Nos piden $p(X < 4) = p(X = 1) + p(X = 2) + p(X = 3) = 0{,}23 + 0{,}322 + 0{,}186 = 0{,}738 = 73{,}8\%$

b) Calcular la probabilidad de que al menos dos personas vivan en una vivienda.

Resolución

Nos piden $p(X \geq 2) = 1 - p(X < 2) = 1 - p(X = 1) = 1 - 0{,}23 = 0{,}77 = 77\%$

c) Determina la probabilidad de que el nº de personas que viven en una vivienda sea impar.

Resolución

Nos piden $p(X = 1) + p(X = 3) + p(X = 5) + p(X = 7) = 0{,}23 + 0{,}186 + 0{,}067 + 0{,}016 = 0{,}499 = 49{,}9\%$

d) Obtener el número medio de personas que habitan en una vivienda y la desviación típica de X.

Resolución

Calculemos los parámetros μ y σ:

x_i	p_i	$x_i\, p_i$	$x_i^2\ p_i$
1	0,23	0,23	0,23
2	0,322	0,644	1,288
3	0,186	0,558	1,674
4	0,155	0,62	2,48
5	0,067	0,335	1,675
6	0,024	0,144	0,864
7	0,016	0,112	0,784
Total	1	2,643	8,995

Media o esperanza matemática de X: $\mu = \sum x_i p_i = 2{,}643 \cong 3$ personas

Desviación típica de X: $\sigma = \sqrt{\sigma^2} = \sqrt{\sum x_i^2\, p_i - \mu^2} = \sqrt{8{,}995 - 2{,}643^2} = \sqrt{2{,}009551} \cong 1{,}4176$

Distribución binomial, probabilidad, parámetros

Empecemos con la siguiente situación:

Lanzamos un dado 15 veces y sea el suceso A ="salir el 6", la probabilidad de A es $p(A) = \frac{1}{6}$.

Tomemos la v.a. X = nº de veces que ocurre A es decir, nº de veces que sale el 6.

Fíjate que X = 0, 1, 2, 3, ..., 15

Decimos que la v.a. X tiene una distribución de probabilidad binomial de parámetros 15 y $\frac{1}{6}$

Se escribe así: $X \to B(15, \frac{1}{6})$

En general, si realizamos n veces el mismo experimento aleatorio y tomamos la v.a.
X = número de veces que ocurre un determinado suceso A" decimos que la variable aleatoria X tiene <u>*distribución de probabilidad binomial*</u> B(n, p), siendo p = p(A). Se representa así: $X \to B(n, p)$

X puede tomar los valores 0, 1, 2, 3,…., n.

Más ejemplos de distribuciones binomiales

1) En un parto la probabilidad de que nazca una niña es del 65%.
Si observamos 30 nacimientos y X = número de niñas. Entonces, $X \to B(30 ; 0{,}65)$

X = 0, 1, 2, 3, …, 30

2) La probabilidad de acertar al lanzar a canasta un determinado jugador de baloncesto es del 70%.
Si lanza a canasta 50 veces y X = número de aciertos. Entonces, $X \to B(50; 0{,}7)$

X = 0, 1, 2, 3, …, 50

3) Lanzamos una moneda 25 veces; X = número de cruces. Entonces, $X \to B(25; 0{,}5)$ porque la probabilidad del suceso A ="salir cruz" es $\frac{1}{2} = 0{,}5$. Observa que X = 0, 1, 2, 3, …, 25

<u>Distribución de probabilidad en una distribución B(n, p)</u>

Si X es una v.a. que tiene una distribución binomial B(n, p) entonces la distribución de probabilidad se puede obtener mediante la fórmula:

$$\boxed{p_k = p(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ donde } k \text{ toma } n+1 \text{ valores, } 0, 1, 2, \ldots, n}$$

Recuerda que $\binom{n}{k}$ es un número combinatorio que se lee "n sobre k" y se puede hallar con la calculadora científica mediante la secuencia n $\boxed{\text{nCr}}$ k $\boxed{=}$

Por ejemplo, $\binom{7}{3}$ se calcularía así: 7 $\boxed{\text{nCr}}$ 3 $\boxed{=}$. Nos da 35.

$\binom{n}{k}$ también se puede hallar con la fórmula $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$, donde n! = 1.2.3 ….n

Por ejemplo, Si $X \to B(5 ; 0{,}2)$, entonces n = 5, p = 0,2, X = 0, 1, 2, 3, 4, 5.

La probabilidad de que X tome el valor 3 es $p(X = 3) = \binom{5}{3} 0{,}2^3 0{,}8^2 = 10.0{,}008.0{,}64 = 0{,}0512 = 5{,}12\%$

<u>Parámetros de una distribución binomial</u>

Los parámetros o medidas más usadas en una distribución binomial son:

<u>*Media aritmética o esperanza matemática de X*</u>: $\boxed{\mu = np}$ <u>*Varianza de X*</u>: $\boxed{\sigma^2 = np(1-p)}$

<u>*Desviación típica de X*</u>: $\boxed{\sigma = \sqrt{\sigma^2} = \sqrt{np(1-p)}}$

Si $X \to B(5 ; 0{,}2)$, la media es $\mu = np = 5 \cdot 0{,}2 = 1$, la varianza es $\sigma^2 = np(1 - p) = 5 \cdot 0{,}2 \cdot 0{,}8 = 0{,}8$ y la desviación típica $\sigma = \sqrt{\sigma^2} = \sqrt{0{,}8} \cong 0{,}8944$

Actividades resueltas

1) Si $X \to B(5; 0{,}65)$, $X = 0, 1, 2, 3, 4, 5$.

Como $p(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $k=0, 1, 2,\ldots, n$ y en este caso, $n = 5$, $p = 0{,}65$,

$$p(X=k)=\binom{5}{k}0{,}65^k(1-0{,}65)^{5-k}=\binom{5}{k}0{,}65^k 0{,}35^{5-k},\ k=0, 1, 2, 3, 4, 5$$

Vamos a resolver algunas cuestiones:

$a)\ p(X=3)=\binom{5}{3}0{,}65^3.0{,}35^2=10\ .0{,}65^3.0{,}35^2=0{,}3364$

$b)\ p(X=0)=\binom{5}{0}0{,}65^0.0{,}35^5=1\ .\ 1\ .\ 0{,}35^5=0{,}005252$

$c)\ p(X\geq 0)=1,\ \textit{pues es el suceso seguro}$

d) $p(X \leq 4) = p(X = 0) + p(X = 1) + p(X = 2) + p(X = 3) + p(X = 4)$

Observa que es más rápido usar la probabilidad del suceso contrario:

$$p(X\leq 4)=1-p(X>4)=1-p(X=5)=1-[\binom{5}{5}0{,}65^5.0{,}35^0]=1-1\ .\ 0{,}65^5\ .\ 1=0{,}8840$$

e) $p(X \geq 2) = p(X = 2) + p(X = 3) + p(X = 4) + p(X = 5)$

Observa que también es más rápido usar la probabilidad del suceso contrario:

$$p(X\geq 2)=1-p(X<2)=1-[p(X=0)+\mathrm{p(X=1)}]=1-[0{,}005252+\binom{5}{1}0{,}65^1.0{,}35^4]=1-[0{,}005252+5\ .\ 0{,}65.0{,}35^4]=0{,}9460$$

f) $p(X \neq 1) = p(X = 2) + p(X = 3) + p(X = 4) + p(X = 5)$

Observa que también es más rápido usar la probabilidad del suceso contrario:

$$p(X\neq 1)=1-p(X=1)=1-\binom{5}{1}0{,}65^1.0{,}35^4=1-5\ .\ 0{,}65\ .\ 0{,}35^4=0{,}9512$$

$g)\ p(X<0)=0,\ \textit{pues es el suceso imposible}$

h) La media de X es $\mu = n.p = 5.0{,}65 = 3{,}25$.

La desviación típica es $\sigma=\sqrt{np(1-p)}=\sqrt{5.0{,}65.(1-0{,}65)}=\sqrt{5.0{,}65.0{,}35}\cong 1{,}0665$

2) Lanzamos una moneda 10 veces. Sea X = nº de cruces y A = salir cruz.
Como la moneda se lanza 10 veces, n = 10 y como p(A) = 0,5 , p = 0,5. Luego, $X \to B(10; 0,5)$. Observa que X = 0, 1, 2, 3, …, 10

$$p(X=k)=\binom{10}{k}0,5^k(1-0,5)^{10-k}=\binom{10}{k}0,5^k 0,5^{10-k}=\binom{10}{k}0,5^{10}\cong 9,77.10^{-4}\binom{10}{k},\ k=0,1,2,3,\ldots,10$$

Vamos a resolver algunas cuestiones:
a) Porcentaje de probabilidad de que salgan 7 cruces:

$$p(X=7)=9,77.10^{-4}\binom{10}{7}=9,77.10^{-4}.\,120=0,11724\cong 11,72\%$$

b) Porcentaje de probabilidad de que salga más de una cruz:
$p(X > 1) = p(X = 2) + p(X = 3) + \ldots + p(X = 10)$

Observa que es más rápido usar la probabilidad del suceso contrario:

$$p(X>1)=1-p(X\le 1)=1-p(X=0)-p(X=1)=1-9,77.10^{-4}\binom{10}{0}-9,77.10^{-4}\binom{10}{1}=$$

$$=1-9,77.10^{-4}.\,1-9,77.10^{-4}.\,10=0,989253\cong 98,93\%$$

3) El 42% de los alumnos de un cierto instituto cursan estudios universitarios al terminar el Bachillerato. En un grupo de ocho alumnos elegidos al azar, halla el % de probabilidad de que estudien carrera: a) Al menos uno b) Más de 6 c) Menos de 2 d) Mas de 2 y menos de 5

Calcula la media y la desviación típica de esta distribución

Resolución

Sea X = nº de alumnos que estudian carrera y A = estudiar carrera. Como el grupo tiene 8 alumnos, n = 8 y como p(A) = 42% = 0,42, p = 0,42. Luego, $X \to B(8; 0,42)$

$$p(X=k)=\binom{8}{k}0,42^k(1-0,42)^{10-k}=\binom{8}{k}0,42^k 0,58^{8-k},\ k=0,1,2,3,\ldots,8$$

a) $p(X\ge 1)=1-p(X<1)=1-p(X=0)=1-\binom{8}{0}0,42^0.0,58^8=1-0,012806\ldots\cong 0,9872=98,72\%$

b) $p(X>6)=p(X=7)+p(X=8)=\binom{8}{7}0,42^7.0,58^1+\binom{8}{8}0,42^8.0,58^0=$

$=0,010697\ldots+0,0009682\ldots\cong 0,0117=1,17\%$

c) $p(X<2)=p(X=0)+p(X=1)=\binom{8}{0}0,42^0.0,58^8+\binom{8}{1}0,42^1.0,58^7=$

$=0,012806\ldots+0,074188\ldots\cong 0,087=8,7\%$

d) $p(2<X<5)=p(X=3)+p(X=4)=\binom{8}{3}0,42^3.0,58^5+\binom{8}{4}0,42^4.0,58^4=$

$=0,272317\ldots+0,246494\ldots\cong 0,5188=51,88\%$

La media de X es $\mu = n.p = 8 \,.\, 0,42 = 3,36$.

La desviación típica es $\sigma=\sqrt{np(1-p)}=\sqrt{8.0,42.(1-0,42)}=\sqrt{8.0,42.0,58}\cong 1,396$

4) Si X una variable aleatoria que sigue una distribución B(6 ; 0,3). Calcula:
a) $p(X = 2)$ b) $p(X \geq 1)$ c) $p(X \neq 4)$ d) $p(X < 5)$ e) La media y la desviación típica de X

Resolución

Como n = 6, p = 0,3, $p(X=k)=\binom{6}{k}0,3^k(1-0,3)^{6-k}=\binom{6}{k}0,3^k0,7^{6-k}$, $k=0,\ 1,\ 2,3,\ 4,\ 5,\ 6$

a) $p(X=2)=\binom{6}{2}0,3^2 0,7^4=15\ .\ 0,09\ .\ 0,2401=0,324135$

b) $p(X\geq 1)=1-p(X<1)=1-p(X=0)=1-\binom{6}{0}0,3^0 0,7^6=1-1\ .\ 1\ .\ 0,117649=0,882351$

c) $p(X\neq 4)=1-p(X=4)=1-\binom{6}{4}0,3^4 0,7^2=1-0,059535=0,940465$

d) $p(X<5)=1-p(X\geq 5)=1-\left[p(X=5)+p(X=6)\right]=1-\left[\binom{6}{5}0,3^5 0,7^1+\binom{6}{6}0,3^6 0,7^0\right]=$

$=1-(0,010206+0,000729)=0,989065$

e) La media de X es $\mu = n.p = 6.0,3 = 1,8$.
La desviación típica es $\sigma=\sqrt{np(1-p)}=\sqrt{6.0,3.(1-0,3)}=\sqrt{6.0,3.0,7}\cong 1,1225$

5) Lanzamos un dado 7 veces. Hallar la probabilidad, en %, de que salga un nº par:
a) 5 veces b) Alguna una vez

Resolución

Sea X = veces que sale un nº par y A = salir nº par.

Como el dado se lanza 7 veces, n = 7 y como p(A) = 0,5, p = 0,5. Luego, $X \to B(7;\ 0,5)$

$$p(X=k)=\binom{7}{k}0,5^k(1-0,5)^{7-k}=\binom{7}{k}0,5^k0,5^{7-k}=\binom{7}{k}0,5^7\cong 7,8125.10^{-3}\binom{7}{k},\ k=0,\ 1,\ 2,3,\ ...,7$$

a) $p(X=5)=7,8125.10^{-3}\binom{7}{5}=7,8125.10^{-3}.\ 21=0,1640625\cong 16,41\%$

b) $p(X>0)=1-p(X=0)=1-7,8125.10^{-3}\binom{7}{0}=1-7,8125.10^{-3}.\ 1=0,9921875\cong 99,22\%$

6) Una encuesta revela que el 20% de la población es favorable a un político y el resto es desfavorable.
Se eligen 6 personas al azar. Usando la distribución binomial, calcula la probabilidad
a) de que las personas favorables las personas favorables al político sean más de 2 pero menos de 5

Resolución

X = nº de personas favorables al político, A = "la persona es favorable al político"

Como se pregunta a 6 personas, n = 6 y como p(A) = 20% = 0,2, p = 0,2. Luego, $X \to B(6;\ 0{,}2)$

$$p(X=k)=\binom{6}{k}0{,}2^k(1-0{,}2)^{6-k}=\binom{6}{k}0{,}2^k0{,}8^{6-k},\ k=0,\ 1,\ 2,3,\ \ldots,6$$

$$\text{a) } p(2<X<5)=p(X=3)+p(X=4)=\binom{6}{3}0{,}2^30{,}8^3+\binom{6}{4}0{,}2^40{,}8^2=0{,}08192+0{,}01536=0{,}09728\cong 9{,}73\%$$

b) de que haya al menos 2 personas favorables a dicho partido.

Resolución

Nos piden $p(X \geq 2) = 1 - p(X < 2) = 1 - p(X = 0) - p(X = 1) =$

$$=1-\binom{6}{0}0{,}2^00{,}8^6-\binom{6}{1}0{,}2^10{,}8^5=1-1.1.0{,}262144-6.0{,}2.0{,}32768=0{,}3446696\cong 34{,}47\%$$

7) Una empresa vendedora de automóviles recibe el pago al contado del 20% de sus ventas.
En un determinado día ha vendido 5 automóviles. Usando la distribución binomial, halla la probabilidad de que al menos 2 automóviles se hayan vendido al contado

Resolución

$$X = n^{o}\ de\ ventas\ al\ contado\ ;\quad X \to B(5\ ;\ 0{,}2) \qquad p(X=k)=\binom{5}{k}0{,}2^k.0{,}8^{5-k}$$

$$p(X\geq 2)=1-p(X<2)=1-[p(X=0)+p(X=1)]=1-[\binom{5}{0}0{,}2^0.0{,}8^5+\binom{5}{1}0{,}2^1.0{,}8^4]=1-[\ 1\ .1.0{,}8^5+5\ .\ 0{,}2.0{,}8^4]=0{,}26272$$

8) La probabilidad de que un estudiante obtenga el título de arquitecto es del 30%.
Se eligen 5 estudiantes al azar. Usando la distribución binomial, calcula el porcentaje de probabilidad de que haya al menos 2 que obtengan el título de arquitecto.

Resolución

X = nº de estudiantes que obtienen el título de arquitecto, A = "el estudiante obtiene el título"

Como se eligen 5 estudiantes, n = 5 y como p(A) = 30% = 0,3, p = 0,3. Luego, $X \to B(5;\ 0{,}3)$

$$p(X=k)=\binom{5}{k}0{,}3^k(1-0{,}3)^{5-k}=\binom{5}{k}0{,}3^k0{,}7^{5-k},\ k=0,\ 1,\ 2,3,4,5$$

Nos piden $p(X \geq 2) = 1 - p(X < 2) = 1 - p(X = 0) - p(X = 1) =$

$$=1-\binom{5}{0}0{,}3^00{,}7^5-\binom{5}{1}0{,}3^10{,}7^4=1-1.1.0{,}16807-5.0{,}3.0{,}2401=0{,}47178\cong 47{,}18\%$$

Teorema de De Moivre de la binomial.

Antes de enunciar el teorema de De Moivre consideremos la siguiente situación:

Lanzamos un dado 100 veces y queremos calcular, por ejemplo, la probabilidad de obtener más de 23 veces el 6.

$$X = \text{nº de veces que sale el 6},\ X \to B(100, 1/6).$$

Como $p(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $k=0, 1, 2, \ldots, n$ y en este caso, n = 100, p = 1/6,

$$p(X=k)=\binom{100}{k}\left(\frac{1}{6}\right)^k\left(1-\frac{1}{6}\right)^{100-k}=\binom{100}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{100-k},\ k=0, 1, 2, 3, \ldots, 100$$

$p(X > 23) = p(X = 24) + p(X = 25) + p(X = 26) + \ldots + p(X = 100)$. Como ves, bastante complicado pues hay que hacer muchas operaciones: hay que calcular 77 probabilidades sustituyendo en la fórmula

Si aplicáramos la probabilidad del suceso contrario:
$p(X > 23) = 1 - p(X \leq 23) = 1 - p(X = 0) - p(X = 1) - p(X = 2) - \ldots - p(X = 23)$.
También bastante complicado pues hay que hacer muchas operaciones: hay que calcular 24 probabilidades sustituyendo en la fórmula

El matemático francés Abraham De Moivre nacido en el siglo XVII descubrió una forma de resolver este problema y enunció el **teorema de De Moivre** que dice:

"Sea una v.a. X que sigue una distribución binomial B(n , p) que cumple:

$$n \geq 30,\ np \geq 5 \quad y \quad n(1-p) \geq 5.$$

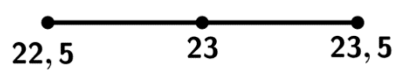
Entonces, la v.a. X se puede sustituir por otra v.a. $X' \to N(\mu, \sigma)$, siendo μ la media de X, $\mu = np$ y σ la desviación típica, $\sigma=\sqrt{np(1-p)}$. Es decir, $X' \to N\left(np, \sqrt{np(1-p)}\right)$ "

De esta forma, podemos calcular probabilidades de una B(n, p) de forma más sencilla y rápida.

Al pasar de la v.a. discreta X a la variable continua X´ hay que hacer algunas modificaciones en las probabilidades. Estas modificaciones se llaman "correcciones de Yates":

Veamos los casos que se nos pueden presentar a través de algunos ejemplos.
Sea la v.a. $X \to B(100, 1/6)$, $X = 0, 1, 2, 3, \ldots, 100$:

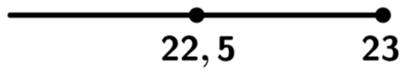
1) $p(X = 23)$. Tomamos un intervalo de la recta que contenga sólo al número 23. Se toma (22,5 ; 23,5)



Entonces, $p(X = 23) = p(22{,}5 < X' < 23{,}5)$

$$\text{En general, } p(X = k) = p(k - 0{,}5 < X' < k + 0{,}5)$$

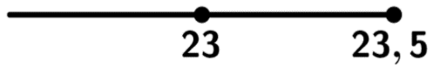
2) $p(X < 23)$. Como los valores de X, menores que 23 son 0, 1, 2, ….,22 tomamos un intervalo de la recta que contenga exactamente esos números. Se toma $(-\infty\ ;\ 22{,}5)$



Entonces, $p(X < 23) = p(X' < 22{,}5)$

En general, $p(X < k) = p(X' < k - 0{,}5)$

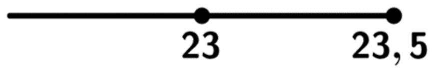
3) $p(X \leq 23)$. Como los valores de X, menores o iguales que 23 son 0, 1, 2, ….,22, 23 tomamos un intervalo de la recta que contenga exactamente esos números. Se toma $(-\infty\ ;\ 23{,}5)$



Entonces, $p(X \leq 23) = p(X' < 23{,}5)$

En general, $p(X \leq k) = p(X' < k + 0{,}5)$

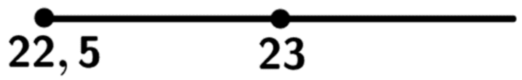
4) $p(X > 23)$. Como los valores de X, mayores que 23 son 24, 25,….,100 tomamos un intervalo de la recta que contenga exactamente esos números. Se toma $(23{,}5\ ;\ +\infty)$



Entonces, $p(X > 23) = p(X' > 23{,}5)$

En general, $p(X > k) = p(X' > k + 0{,}5)$

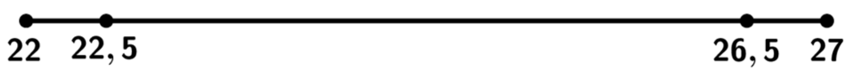
5) $p(X \geq 23)$. Como los valores de X, mayores o iguales que 23 son 23, 24,….,100 tomamos un intervalo de la recta que contenga exactamente esos números. Se toma $(22{,}5\ ;\ +\infty)$



Entonces, $p(X \geq 23) = p(X' > 22{,}5)$

En general, $p(X \geq k) = p(X' > k - 0{,}5)$

6) $p(22 < X < 27)$. Como los valores de X, mayores que 22 y menores que 27 son 23, 24, 25 y 26 tomamos un intervalo de la recta que contenga exactamente esos números. Se toma $(22{,}5\ ;\ 26{,}5)$



Entonces, $p(22 < X < 27) = p(22{,}5 < X' < 26{,}5)$

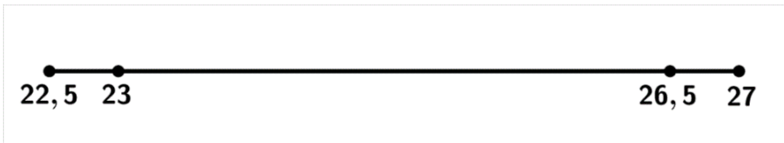
En general, $p(a < X < b) = p(a + 0{,}5 < X' < b - 0{,}5)$

7) $p(23 \le X \le 26)$. Como los valores de X, mayores o iguales que 23 y menores o iguales que 26 son 23, 24, 25 y 26 tomamos un intervalo de la recta que contenga exactamente esos números.
Se toma (22,5 ; 26,5)

Entonces, $p(23 \le X \le 26) = p(22{,}5 < X' < 26{,}5)$.

En general, $p(a \le X \le b) = p(a - 0{,}5 < X' < b + 0{,}5)$

8) $p(23 \le X < 27)$. Como los valores de X, mayores o iguales que 23 y menores que 27 son 23, 24, 25 y 26 tomamos un intervalo de la recta que contenga exactamente esos números.
Se toma (22,5 ; 26,5)



Entonces, $p(23 \le X < 27) = p(22{,}5 < X' < 26{,}5)$

En general, $p(a \le X < b) = p(a - 0{,}5 < X' < b - 0{,}5)$

9) $p(22 < X \le 26)$. Como los valores de X, mayores que 22 y menores o iguales que 26 son 23, 24, 25 y 26 tomamos un intervalo de la recta que contenga exactamente esos números.
Se toma (22,5 ; 26,5)

Entonces, $p(22 < X \le 26) = p(22{,}5 < X' < 26{,}5)$

En general, $p(a < X \le b) = p(a + 0{,}5 < X' < b + 0{,}5)$

Como ejemplo, vamos a resolver usando el teorema de De Moivre el problema antes expuesto de tirar un dado 100 veces:
$X \rightarrow B(100,1/6)$ y queremos hallar $p(X > 23)$. En este caso, $n = 100$, $p = 1/6$

Veamos si se puede aplicar el teorema de De Moivre:

$$np = 100.\frac{1}{6} = \frac{100}{6} \cong 16{,}7 \; ; \; n(1-p) = 100.\left(1-\frac{1}{6}\right) = 100.\frac{5}{6} = \frac{500}{6} \cong 83{,}3 .$$

Se puede aplicar el teorema porque $n \ge 30$, $np \ge 5$ y $n(1 - p) \ge 5$

Aproximamos $X \rightarrow B(100,1/6)$ por $X' \rightarrow N(\mu, \sigma)$,

siendo $\mu = np = 16{,}7$, $\sigma = \sqrt{n.p(1-p)} = \sqrt{100.\frac{1}{6}\left(1-\frac{1}{6}\right)} = \sqrt{100.\frac{1}{6}\frac{5}{6}} = \sqrt{\frac{500}{36}} \cong 3{,}7 .$

Como $X' \rightarrow N(16{,}7 \; ; \; 3{,}7)$, tipificando, $Z = \dfrac{X'-16{,}7}{3{,}7} \rightarrow N(0,\, 1)$

$p(X > 23) = p(X' > 23{,}5)$. Tipificamos:

$$p\left(\frac{X'-16{,}7}{3{,}7} > \frac{23{,}5-16{,}7}{3{,}7}\right) = p(Z > 1{,}84) = 1 - p(Z \le 1{,}84) = 1 - 0{,}9671 = 0{,}0329 = 3{,}29\%$$

Actividades resueltas

1) El 2% de los tornillos fabricados por una máquina presentan defectos.
Si tenemos un lote de 300 tornillos, ¿cuál es la probabilidad de que haya más de 10 defectuosos?

Resolución

$X = n^o$ *de tornillos defectuosos* ; $X \to B(300\ ;\ 0{,}02)$; $n = 300,\ p = 0{,}02 \to np = 300\ .\ 0{,}02 = 6 \geq 5\ ,\ n(1-p) = 300\ .\ 0{,}98 = 294 \geq 5$

Por tanto, se puede aplicar el teorema de De moivre. Sustituimos X *por otra v.a.* $X' \to N(np\ ,\ \sqrt{np(1-p)}\ ) = N(6\ ;\ 2{,}42)$

$$p(X > 10) = p(X' > 10{,}5) = p(\frac{X'-6}{2{,}42} > \frac{10{,}5-6}{2{,}42}) = p(Z > 1{,}86) = 1 - p(Z \leq 1{,}86) = 1 - 0{,}9686 = 0{,}0314 = 3{,}14\%$$

2) En una ciudad una de cada tres familias posee teléfono. Si se eligen al azar 90 familias, calcula la probabilidad de que entre ellas haya por lo menos 30 tengan teléfono.

Resolución

$X = n^o$ *de familias con teléfono* ; $X \to B(90\ ;\ \frac{1}{3}) = B(90\ ;0{,}33)$; $n = 90,\ p = 0{,}33 \to np = 90\ .\ 0{,}33 = 29{,}7 \geq 5\ ,\ n(1-p) = 90\ .\ 0{,}67 = 60{,}3 \geq 5$

Por tanto, se puede aplicar el teorema de De moivre. Sustituimos X *por otra v.a.* $X' \to N(np\ ,\ \sqrt{np(1-p)}\ ) = N(29{,}7\ ;\ 4{,}46)$

$$p(X \geq 30) = p(X' > 29{,}5) = p(\frac{X'-29{,}7}{4{,}46} > \frac{29{,}5-29{,}7}{4{,}46}) = p(Z > -0{,}04) = p(Z < 0{,}04) = 0{,}5160 = 51{,}6\%$$

3) El 5% de los libros prestados en una biblioteca de un centro escolar son técnicos.
Si se toman los últimos 500 préstamos, calcula la probabilidad de que se hayan prestado más de 25 y menos de 30 libros técnicos.

Resolución

$X = n^o$ *de libros técnicos* ; $X \to B(500\ ;\ 0{,}05)$; $n = 500,\ p = 0{,}05 \to np = 500\ .\ 0{,}05 = 25 \geq 5\ ,\ n(1-p) = 500\ .\ 0{,}95 = 475 \geq 5$

Por tanto, se puede aplicar el teorema de De moivre. Sustituimos X *por otra v.a.* $X' \to N(np\ ,\ \sqrt{np(1-p)}\ ) = N(25\ ;\ 4{,}87)$

$$p(25 < X < 30) = p(25{,}5 < X' < 29{,}5) = p(\frac{25{,}5-25}{4{,}87} < \frac{X'-25}{4{,}87} < \frac{29{,}5-25}{4{,}87}) = p(0{,}1 < Z < 0{,}92) =$$

$$= p(Z < 0{,}92) - p(Z < 0{,}1) = 0{,}8212 - 0{,}5398 = 0{,}2814 = 28{,}14\%$$

4) Después de realizar varios sondeos sobre una población se ha conseguido averiguar que únicamente el 15% de la misma es favorable a los tratamientos de psicoterapia.
Elegimos 50 personas al azar. Halla, usando el teorema de De Moivre, el % de probabilidad de que haya más de 10 personas favorables a los tratamientos de psicoterapia.

Resolución

X = nº de personas favorables, $X \to B(300; 0{,}15)$; $n = 50$, $p = 0{,}15$, $np = 7{,}5 \geq 5$, $n(1 - p) = 42{,}5 \geq 5$

Por tanto, se puede aplicar el teorema de De Moivre.

Sustituimos X por X´→ $X \to N\left(np, \sqrt{np(1-p)}\ \right) = N(7{,}5; 6{,}2)$

Nos piden $p(X > 10) = p(\frac{X-7{,}5}{6{,}2} > \frac{10-7{,}5}{6{,}2}) = p(Z > 0{,}4) = 1 - p(Z \leq 0{,}4) = 1 - 0{,}6554 = 0{,}3446 = 34{,}46\%$

TEOREMA CENTRAL DEL LÍMITE

Muestreo aleatorio simple

Consiste en tomar al azar unos pocos elementos de la población.
Por ejemplo, si la población es P = { a, b, c, d } y tomamos muestras de 2 elementos con reemplazamiento obtendríamos el conjunto de todas las muestras:

M = { a-a, a-b, a-c, a-d, b-a, b-b, b-c, b-d, c-a, c-b, c-c, c-d, d-a, d-b, d-c, d-d }

Teorema central del límite para la distribución de la media muestral

Sea X una v.a. continua con media μ y desviación típica σ.

Consideramos todas las muestras con reemplazamiento de tamaño n de dicha población:

M_1 , M_2 , M_3 , …., etc

Sea $\overline{x_1}$ la media de la muestra M_1 , $\overline{x_2}$ la media de M_2 , $\overline{x_3}$ la media de M_3, … , etc

La variable aleatoria $\overline{X}$ = media de las muestras de tamaño n: $\overline{X} = \overline{x_1}$, $\overline{x_2}$, $\overline{x_3}$, … , etc

tiene una distribución con media también μ y varianza $\frac{\sigma^2}{n}$ (su desviación típica es $\frac{\sigma}{\sqrt{n}}$)

Si además $n \geq 30$ ó $X \to \mathbf{N}(\mu,\sigma)$, entonces $\overline{X} \to N(\mu , \frac{\sigma}{\sqrt{n}})$

Si μ es desconocida, $\overline{X} \to N(\overline{x} , \frac{\sigma}{\sqrt{n}})$, siendo $\overline{x}$ la media de una de las muestras

Ejemplo:

Consideremos una población formada por 4 estudiantes (N = 4)

Sea la v.a. continua X = notas de un examen = 8, 9, 5, 6.

Puedes comprobar que la media es $\mu = 7$ y la desviación típica es $\sigma = \sqrt{\frac{5}{2}}$

Si ahora consideramos todas las muestras con reemplazamiento de tamaño dos (n = 2) y las medias de estas muestras, obtenemos una nueva variable $\overline{X}$, que viene expresada en la siguiente tabla:

Muestras	8-8	8-9	8-5	8-6	9-8	9-9	9-5	9-6	5-8	5-9	5-5	5-6	6-8	6-9	6-5	6-6
Variable media = $\overline{X}$	8	8.5	6.5	7	8.5	9	7	7.5	6.5	7	5	5.5	7	7.5	5.5	6

La media de $\overline{X}$ es también 7 y su desviación típica es $\frac{\sigma}{\sqrt{n}} = \frac{\sqrt{5/2}}{\sqrt{2}} = \sqrt{\frac{5}{4}} = \frac{\sqrt{5}}{2}$

Teorema central del límite para la distribución de la proporción muestral

Sea X una v.a. discreta que tiene una distribución binomial B(n,p).

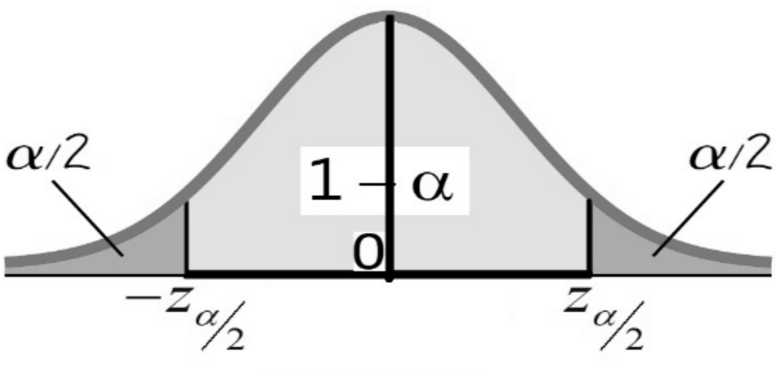
Para tamaños grandes de n, n > 30, la distribución binomial se aproxima a una distribución Normal.
Se puede considerar que $X \to N(np, \sqrt{np(1-p)})$

La variable aleatoria $p = \dfrac{X}{n}$ = proporción muestral, tiene una distribución normal de media p y desviación típica $\sqrt{\dfrac{p(1-p)}{n}}$.
$$\boxed{p \to N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)}$$

INTERVALOS DE CONFIANZA

Intervalo de confianza para la media poblacional

Sea X es una v.a. tal que $X \to N(\mu, \sigma)$ de la que desconocemos la media poblacional μ.
Un intervalo de confianza para μ es un intervalo I en el que $p(\mu \in I) = \mathbf{n_c}$.
El valor n_c se llama nivel de confianza y $\mathbf{n_c = 1-\alpha}$ (A α se le llama "nivel de significación")



$$z_{\alpha/2} \text{ cumple } p(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha = n_c \;\Rightarrow\; p(Z < z_{\alpha/2}) = 1-\alpha+\frac{\alpha}{2} = \frac{2-\alpha}{2} = \frac{1+\overbrace{1-\alpha}^{es\ n_c}}{2}$$

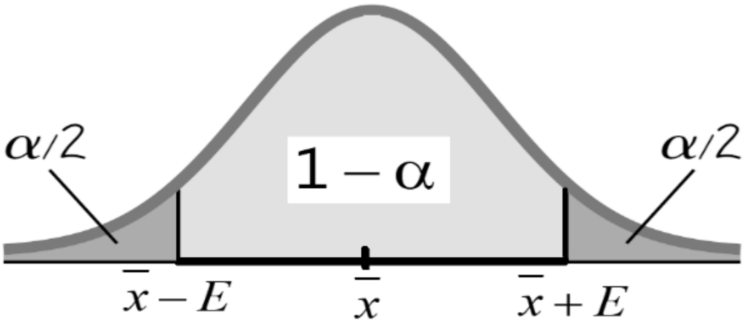
$$\text{Luego, } \boxed{p(Z < z_{\alpha/2}) = \frac{1+n_c}{2}}$$

Para estimar cuánto vale μ, tomamos una muestra de tamaño n cuya media es $\bar{x}$

Se puede demostrar que el intervalo de confianza para la media es

$$\boxed{I = (\bar{x} - E\ ,\ \bar{x} + E)} \quad \text{siendo} \quad \boxed{E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

E se llama *<u>error de estimación o error máximo admisible</u>* y es el mayor error que podemos cometer cuando tomamos $\bar{x}$ como aproximación de μ

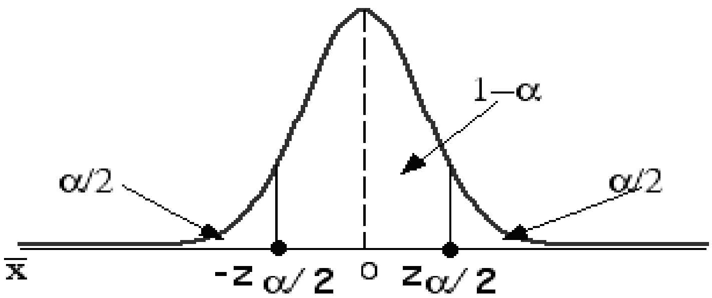


$n_c = p(\bar{x} - E < \mu < \bar{x} + E)$

$$\bar{x} - E < \mu < \bar{x} + E \xleftrightarrow{.(-1)} -\bar{x} + E > -\mu > -\bar{x} - E \Leftrightarrow -\bar{x} - E < -\mu < -\bar{x} + E$$

Según el teorema central del límite, $\bar{x} \to N(\mu, \frac{\sigma}{\sqrt{n}})$, luego $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \to N(0,1)$

Tipificando (sumamos $\bar{x}$ *y dividimos entre* $\sigma/\sqrt{n}$*)*: $\frac{-E}{\sigma/\sqrt{n}} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{E}{\sigma/\sqrt{n}} \Leftrightarrow \frac{-E}{\sigma/\sqrt{n}} < Z < \frac{E}{\sigma/\sqrt{n}}$



n_c = 1 – α (α se llama nivel de significación)

$$Como\ \ n_c = p(-\bar{x} - E < \mu < \bar{x} + E) = p\left(\frac{-E}{\sigma/\sqrt{n}} < Z < \frac{E}{\sigma/\sqrt{n}}\right) = p(-z_{\alpha/2} < Z < z_{\alpha/2})$$

$$z_{\alpha/2} = \frac{E}{\sigma/\sqrt{n}},\ despejando\ \boxed{E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$$

$$Por\ otra\ parte,\ z_{\alpha/2}\ cumple\ n_c = p(-z_{\alpha/2} < Z < z_{\alpha/2}) = 2\phi(z_{\alpha/2}) - 1_c \to \boxed{\phi(z_{\alpha/2}) = \frac{1 + n_c}{2}}$$

Intervalo de confianza para la proporción poblacional

Sea X es una v.a. tal que $X \to N(\mu, \sigma)$ de la que desconocemos la proporción, p, de individuos de la población que tiene una cierta característica. Un intervalo de confianza para p es un intervalo I en el que la probabilidad de que p pertenezca a I es n_c. El valor n_c se llama nivel de confianza y $\mathbf{n_c = 1-\alpha}$ (A α se le llama "nivel de significación")

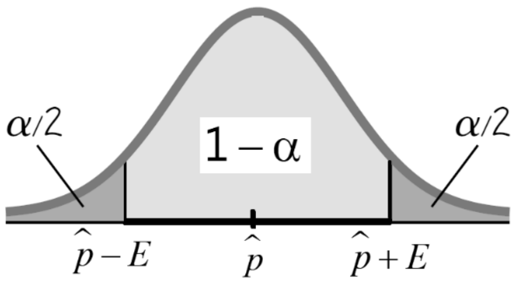
Razonando de forma análoga que en los intervalos para la media: $\boxed{p(Z < z_{\alpha/2}) = \frac{1+n_c}{2}}$

Para estimar cuánto vale p, tomamos una muestra de tamaño n cuya proporción de individuos que cumple dicha característica es $\bar{\mathbf{p}}$

Se puede demostrar que el intervalo de confianza para la proporción es

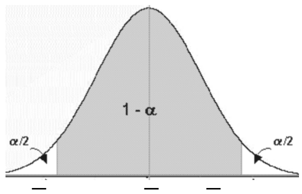
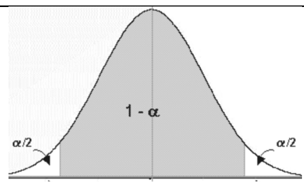
$$\boxed{I = (\bar{\mathbf{p}} - E\ ,\ \bar{\mathbf{p}} + E)} \quad \text{siendo} \quad \boxed{E = z_{\alpha/2} \cdot \sqrt{\frac{\bar{\mathbf{p}}\,(1-\bar{\mathbf{p}})}{n}}}$$

E se llama *error de estimación o error máximo admisible* y es el mayor error que podemos cometer cuando tomamos p como aproximación de p



$n_c = p(p - E < p < p + E)$

Razonando de manera similar al caso anterior, $\boxed{E = z_{\alpha/2} \cdot \sqrt{\frac{p\ (1-p)}{n}}}$ $\boxed{\phi(z_{\alpha/2}) = \frac{1+n_c}{2}}$

Parámetro a estimar	Intervalo de confianza	Error máximo admisible	Interpretación gráfica
Media	$I = (\bar{x} - E\ ,\ \bar{x} + E)$ El punto medio del intervalo es $\bar{x}$ y su amplitud es 2E	$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	 $\bar{x} - E$ $\bar{x}$ $\bar{x} + E$
Proporción	$I = (p - E\ ,\ p + E)$ El punto medio del intervalo es p y la amplitud es 2E	$E = z_{\alpha/2} \cdot \sqrt{\frac{p\ (1-p)}{n}}$	 $p - E$ p $p + E$

Propiedades de los intervalos de confianza

- El punto medio del intervalo de confianza para la media es $\overline{x}$ y para la proporción es p

$$punto\ medio = \frac{extremo\ \sup erior + extremo\ \inf erior}{2} \begin{cases} para\ la\ media:\ punto\ medio = \dfrac{\left(\overline{x}+E\right)+\left(\overline{x}-E\right)}{2} = \overline{x} \\ para\ la\ proporción:\ punto\ medio = \dfrac{\left(\overline{\mathbf{p}}+E\right)+\left(\overline{\mathbf{p}}-E\right)}{2} = \overline{\mathbf{p}} \end{cases}$$

- La amplitud del intervalo de confianza, es A = 2E y, por tanto, el error es $E = \frac{A}{2}$

$$A = amplitud = extremo\ \sup erior - extremo\ \inf erior \begin{cases} para\ la\ media:\ A = \left(\overline{x}+E\right)-\left(\overline{x}-E\right) = 2E \\ para\ la\ proporción:\ A = \left(\overline{\mathbf{p}}+E\right)-\left(\overline{\mathbf{p}}-E\right) = 2E \end{cases}$$

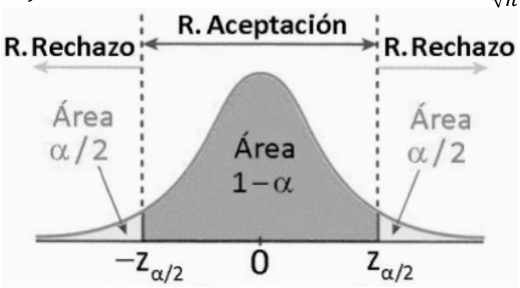
- Cuanto mayor es el tamaño de la muestra, n, menor es el error E y por tanto más "estrecho" es el intervalo de confianza

- Cuanto mayor es el nivel de confianza, n_c, mayor es el error E y por tanto más amplio es el intervalo de confianza

ANEXO: CONTRASTES DE HIPÓTESIS

Contraste de hipótesis bilateral para la media con desviación típica conocida

$$\begin{cases} H_0: \mu = \mu_0 \ \ (hipótesis\ nula) \\ H_1: \mu \neq \mu_0 \ \ (hipótesis\ alternativa) \end{cases}$$ Estadístico de contraste $z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$



Para un nivel de significación α:

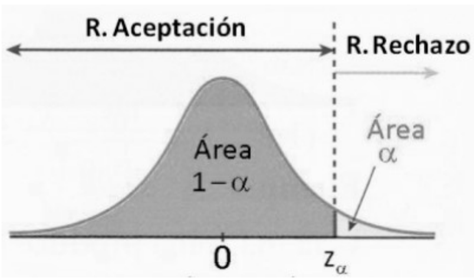
Región de aceptación, $R_A = [-z_{\alpha/2},\ z_{\alpha/2}]$ Región crítica o de rechazo, $R_c = (-\infty, -z_{\alpha/2}) \cup (z_{\alpha/2}, +\infty)$

Si el estadístico z ∈ R_A se acepta la hipótesis nula y en caso contrario se rechaza

Contrastes de hipótesis unilateral para la media con desviación típica conocida

Tipo I

$$\begin{cases} H_0: \mu \leq \mu_0 \ \ (hipótesis\ nula) \\ H_1: \mu > \mu_0 \ \ (hipótesis\ alternativa) \end{cases}$$ Estadístico de contraste $z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$



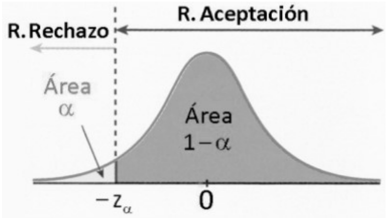
Para un nivel de significación α:

Región de aceptación, $R_A = (-\infty,\ z_\alpha]$ Región crítica o de rechazo, $R_c = (z_\alpha, +\infty)$

Si el estadístico z ∈ R_A se acepta la hipótesis nula y en caso contrario se rechaza

Tipo II

$\begin{cases} H_0: \mu \geq \mu_0 \ \ (hipótesis\ nula) \\ H_1: \mu < \mu_0 \ \ (hipótesis\ alternativa) \end{cases}$ Estadístico de contraste $z = \dfrac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$



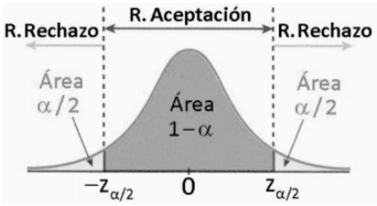
Para un nivel de significación α:

Región de aceptación, $R_A = [-z_\alpha, +\infty)$ Región crítica o de rechazo, $R_c = (-\infty, -z_\alpha)$

Si el estadístico z ∈ R_A se acepta la hipótesis nula y en caso contrario se rechaza

Contraste de hipótesis bilateral para la proporción

$\begin{cases} H_0: p = p_0 \ \ (hipótesis\ nula) \\ H_1: p \neq p_0 \ \ (hipótesis\ alternativa) \end{cases}$ Estadístico de contraste $z = \dfrac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$



Para un nivel de significación α:

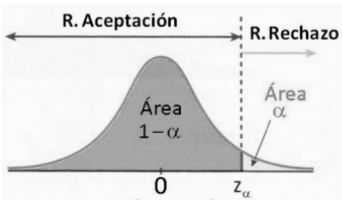
Región de aceptación, $R_A = [-z_{\alpha/2},\ z_{\alpha/2}]$ Región crítica o de rechazo, $R_c = (-\infty, -z_{\alpha/2}) \cup\ (z_{\alpha/2}, +\infty)$

Si el estadístico z ∈ R_A se acepta la hipótesis nula y en caso contrario se rechaza

Contrastes de hipótesis unilateral para la proporción

Tipo I

$\begin{cases} H_0: p \leq p_0 \ \ (hipótesis\ nula) \\ H_1: p > p_0 \ \ (hipótesis\ alternativa) \end{cases}$ Estadístico de contraste $z = \dfrac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$



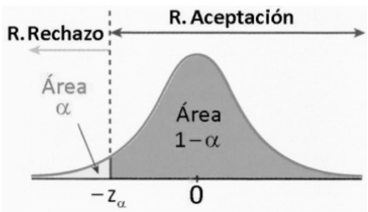
Para un nivel de significación α:

Región de aceptación, $R_A = (-\infty,\ z_\alpha]$ Región crítica o de rechazo, $R_c = (z_\alpha, +\infty)$

Si el estadístico z ∈ R_A se acepta la hipótesis nula y en caso contrario se rechaza

Tipo II

$\begin{cases} H_0: p \geq p_0 \ \ (hipótesis\ nula) \\ H_1: p < p_0 \ \ (hipótesis\ alternativa) \end{cases}$ Estadístico de contraste $z = \dfrac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$



Para un nivel de significación α:

Región de aceptación, $R_A = [-z_\alpha, +\infty)$ Región crítica o de rechazo, $R_c = (-\infty, -z_\alpha)$

Si el estadístico z ∈ R_A se acepta la hipótesis nula y en caso contrario se rechaza