A5 Y B5 ESTADÍSTICA

Función de probabilidad

La función de probabilidad de una variable que sigue una distribución binomial $B(n, p)$ es:

$$P(r\ éxitos) = P(X = r) = \binom{n}{r} \cdot p^r \cdot q^{n-r}$$

Una moneda esta trucada de manera que la probabilidad de salir cara es de 0,4 se lanza seis veces. Halla la probabilidad de que:

- Salgan exactamente cuatro caras
- Salgan seis cruces
- Salga, al menos, 4 caras
- Salgan, a lo sumo, dos caras
- Halla la media y la desviación.

Llamamos X a la variable que indica el numero de caras y observamos que sigue una distribución binomial $B(6{,}0'4)$, siendo $n = 6$ lanzamientos $p = 0{,}4$ la probabilidad de salir cara y $q = 1 - p = 1 - 0{,}4 = 0{,}6$ la probabilidad de cruz.

$$P(X = 4) = \binom{6}{4} \cdot 0{,}4^4 \cdot 0{,}6^2 = \frac{6!}{4! \cdot 2!} \cdot 0{,}0256 \cdot 0{,}36 = 0{,}13824$$

$$P(X = 0) = \binom{6}{0} \cdot 0{,}4^0 \cdot 0{,}6^6 = 1 \cdot 1 \cdot 0{,}046656 = 0{,}046656$$

$$P(X \geq 4) = P(X = 4) + P(X = 5) + P(X = 6)$$

$$= \binom{6}{4} \cdot 0{,}4^4 \cdot 0{,}6^2 + \binom{6}{5} \cdot 0{,}4^5 \cdot 0{,}6^1 + \binom{6}{6} \cdot 0{,}4^6 \cdot 0{,}6^0 = 0{,}1792$$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) =$$

$$\binom{6}{0} \cdot 0{,}4^0 \cdot 0{,}6^6 + \binom{6}{1} \cdot 0{,}4^1 \cdot 0{,}6^5 + \binom{6}{2} \cdot 0{,}4^2 \cdot 0{,}6^4 = 0{,}54432$$

Parámetros de una distribución binomial

$$Media\ o\ esperanza;\ \mu = n \cdot p$$

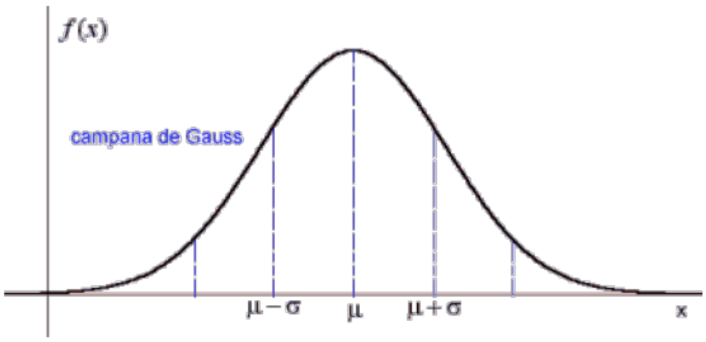
$$Varianza;\ \sigma^2 = n \cdot p \cdot q$$

$$Desviación\ típica;\ \sigma = \sqrt{n \cdot p \cdot q}$$

DISTRIBUACIÓN NORMAL

Una variable aleatoria continua X sigue una distribución normal de media μ y desviación típica σ , y se escribe, $N(\mu, \sigma)$ si se cumple:

- Toma valores en todo $\mathbb{R}$
- Su función de densidad es (Representar la campana de Gauss)



Algunas propiedades de esta grafica son:

- En el intervalo $(\mu - \sigma, \mu + \sigma)$ el área encerrada es $0{,}6826$.
- En el intervalo $(\mu - 2\sigma, \mu + 2\sigma)$ el área encerrada es $0{,}9544$.
- En el intervalo $(\mu - 3\sigma, \mu + 3\sigma)$ el área encerrada es $0{,}9973$.

La función depende de los valores de la media y de la desviación típica. El valor de la media desplaza la grafica a izquierda o derecha, mientras que el calor de la desviación, modifica la altura y anchura de la campana.

DISTRIBUCIÓN NORMAL ESTANDAR

Dada la complejidad de la función de densidad de la distribución normal, y teniendo en cuneta la dificultad que ello provoca en el calculo de probabilidades, es lógico pensar en encontrar los valores de $\mu \ \ y \, \sigma$ que faciliten al máximo la tarea.

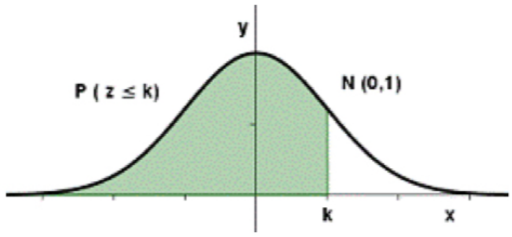
Se conoce como distribución normal estándar o tipificada a la distribución normal en la que $\mu = 0 \ y \ \sigma = 1$, es decir, $N(0,1)$. la variable aleatoria $N(0,1)$ se identifica con Z.

Lo bueno de esta distribución es que nos permite encontrar la solución y por tanto las probabilidades necesarias, sin tener que realizar cálculos integrales.

A continuación, aprenderemos a utilizar la tabla:

Tabla de distribución normal estándar N (0, 1)

Los valores de la tabla representan el área bajo la curva normal hasta un valor positivo de Z.

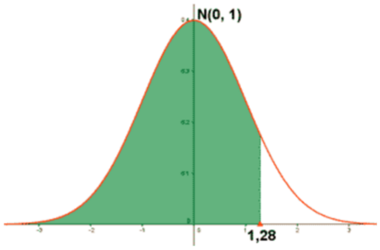


k	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7652
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8930
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9561	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9934	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9901	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9954	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0.9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

UTILIZACIÓN DE LAS TABALAS DE UN DISTRIBUCIÓN NORMAL $N(\mu, \sigma)$

Ejemplo 1:

Calcular P(Z ≤ 1,28)

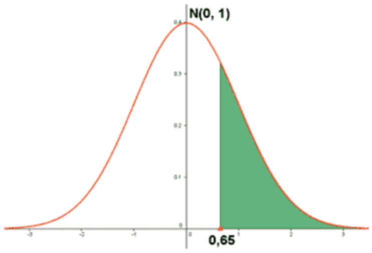


Esta probabilidad se puede encontrar directamente en la tabla. Buscamos en la tabla la intersección de la fila que comienza por 1,2 y la columna correspondiente a 0,08.
Y obtenemos P(Z ≤ 1,28) = 0,8997. Puede decirse que aproximadamente el 89,97% de los valores de la variable están distribuidos entre -∞ y 1,28.

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

Ejemplo 2:

Calcular P (Z ≥ 0,65)



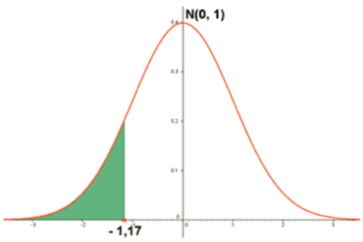
Esta probabilidad no se puede encontrar directamente en la tabla, tenemos que calcularla utilizando la probabilidad del suceso contrario.
P (Z ≥ 0,65) = 1 - P(Z ≤0,65)

Buscamos en la tabla la intersección de la fila que comienza por 0,6 y la columna correspondiente a 0,05.
Y obtenemos P(Z ≤ 0,65) = 0,7422.
P (Z ≥ 0,65) = 1 - P(Z ≤0,65) = 1 - 0,7422 = 0,2578
Puede decirse que aproximadamente el 25,78% de los valores de la variable están distribuidos entre 0,65 y +∞.

$$P(Z \leq -a) = P(Z > a) = 1 - P(Z \leq a)$$

Ejemplo 3:

Calcular P(Z ≤ -1,17)



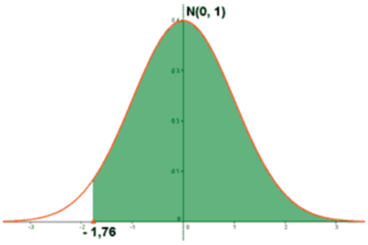
La tabla sólo ofrece probabilidades para valores positivos de la variable Z. Teniendo en cuenta la simetría de la función densidad, y que el área bajo toda la curva es una unidad, se obtiene:

P(Z ≤ -1,17) = P(Z ≥ 1,17) = 1 - P(Z ≤1,17) = 1-0,8790 = 0,121

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

Ejemplo 4:

Calcular P(Z ≥ -1,76)



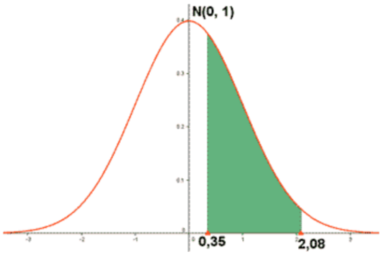
El área equivalente que podemos obtener directamente de la tabla es siguiendo el mismo procedimiento que el ejemplo 3.
Teniendo en cuenta la simetría de la función densidad, y que el área bajo toda la curva es una unidad, se obtiene:

P(Z ≥ -1,76) = P(Z ≤1,76) = 0,9608

$$P(a \le Z \le b) = P(Z \le b) - P(Z \le a)$$

Ejemplo 5:

Calcular P(0,35 ≤ Z ≤ 2,08)



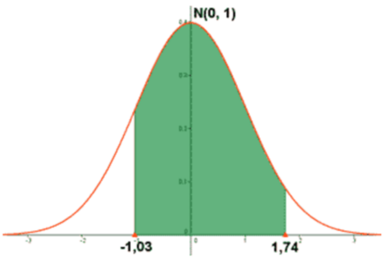
La probabilidad pedida se calcula restando el área mayor menos el área menor.

P(0,35 ≤ Z ≤ 2,08) = P(Z ≤ 2,08) - P(Z ≤ 0,35) =

= 0,9812 - 0,6368 = 0,3444

Ejemplo 6:

Calcular P(-1,03 ≤ Z ≤ 1,74)



Uno de los valores de la variable Z es positivo y otro negativo, por tanto, se resta el mayor al menor y a su vez, el negativo se pasa a positivo como en el ejemplo 4.

P(-1,03 ≤ Z ≤ 1,74) = P(Z ≤ 1,74) - P(Z ≤ -1,03) =

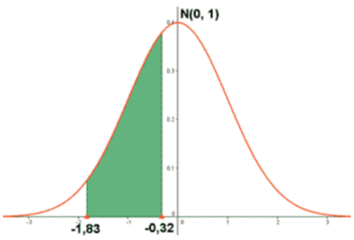
= P(Z ≤ 1,74) - [1 - P(Z ≤ 1,03)] = 0,9591 - (1 - 0,8485) =

= 0,8076

$$P(-a \le Z \le -b) = P(b \le Z \le a)$$

Ejemplo 7:

Calcular P(-1,83 ≤ Z ≤ -0,32)



Como consecuencia de la simetría de la función densidad: la probabilidad pedida es la misma que positiva. Luego:

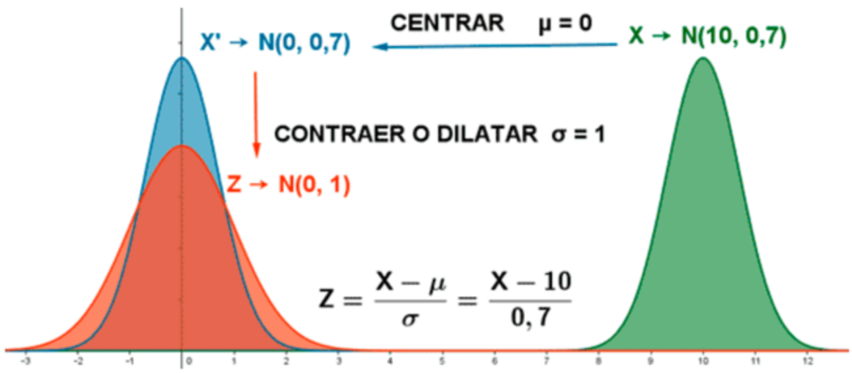
P(-1,83 ≤ Z ≤ -0,32) = P(0,32 ≤ Z ≤ 1,83) =

= P(Z ≤ 1,83) - P(Z ≤ 0,32) = 0,9664 - 0,6255 = 0,3409

Observaciones

1. En caso de que se quiera calcular $P(Z \leq k)$ siendo k un numero mayor que $3'59$, se puede decir que dicha probabilidad es igual a 1 aproximadamente, puesto que en la tabla se observa que $P(Z \leq 3'59) = 0'9999$ y para valores mayores tiene que aumentar por ser acumulativa, siendo uno el área total bajo la curva.
2. Observa que en todos los casos utilizamos indistintamente $>$, $<$, $\leq$, $\geq$ ya que en las distribuciones continuas $P(Z = a) = 0$. Es por eso que, al utilizar el suceso contrario, escribimos
   $P(Z \leq a) = 1 - P(Z > a)$. Sin embargo, debemos ser cuidadosos si la distribución es binomial, pues lo cierto seria $P(Z \leq a) = 1 - P(Z > a)$.
3. Es posible que la búsqueda se realice de manera indirecta, es decir, conocida
   $P(Z \leq a) = p$, hallar el valor de a.

Una vez que hemos aprendido a trabajar con la tabla, nos planteamos como calcular probabilidades si la variable no es $N(0,1)$, cosa prácticamente segura en un contexto real. Como seria tarea imposible tabular todas las posibles distribuciones normales, y para evitar el calculo integral, se ha buscado una estrategia que permite transformar una variable X de distribución $N(\mu, \sigma)$ en otra variable Z de distribución $N(0,1)$. Dicha estrategia se conoce como tipificación de la variable.



TIPIFICACIÓN DE LA VARIABLE

Tipificar una variable X que sigue una distribución normal $N(\mu,\sigma)$, es ajustarla a otra variable Z normal $N(0,1)$. Para ello, es necesario trasladar su centro desde $\mu\ a\ 0$ y reducir o ampliar su forma acampanada desde $\sigma\ a\ 1$.

Ambas cosas se consiguen sin mas que hacer el cambio de variable $Z=\frac{X-\mu}{\sigma}$

De esta forma, si necesitamos calcula $P(Z\leq a)$ solo tenemos que realizar dicho cambio de variable de la siguiente forma:

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

Con lo que con esta situación estaríamos en condiciones de hacer uso de la tabla.

Las tallas de los alumnos de 17 $años$ de un centro escolar siguen una distribución $N(165,10)$.

- Eligiendo un alumno al azar, ¿Cuál es la probabilidad de que mida menos de 170 centiemtros? ¿Y entre155 y 170cm?
- ¿Qué porcentaje de alumnos mide mas de 182 cm?
- ¿A partir de que talla se encuentra el 10 % de los alumnos mas altos?

Sea X la variable normal $N(165,10)$. Se pide:

- $P(X<170)=P(X\leq 170)=P\left(\frac{X-165}{10}\leq\frac{170-165}{10}\right)=P(Z\leq 0{,}5)=0{,}6915$

$$P(155\leq X\leq 170)=P\left(\frac{155-165}{10}\leq\frac{X-165}{10}\leq\frac{170-165}{10}\right)=P(-1\leq Z\leq 0{,}5)=$$

$$P(Z\leq 0{,}5)-P(Z\leq -1)=0{,}6915-P(Z\geq 1)=0{,}6915-[1-P(Z\leq 1)]=$$

$$0{,}6915-1+0{,}8413=0{,}5328$$

- $P(X\geq 182)=P\left(Z\geq\frac{182-165}{10}\right)=P(Z\geq 1{,}7)=1-P(Z\leq 1{,}7)=1-0{,}9554=0{,}0446$

Lo que significa que el 4,46% de los estudiantes miden mas de 182 cm

- $P(X\geq a)=P\left(\frac{X-165}{10}\geq\frac{a-165}{10}\right)=0{,}1\ \rightarrow\ P\left(Z\geq\frac{a-165}{10}\right)=0{,}1\ \rightarrow$

$$1-P\left(Z\leq\frac{a-165}{10}\right)=0{,}1\ \rightarrow\ 1-0{,}1=P\left(Z\leq\frac{a-165}{10}\right)\rightarrow P\left(Z\leq\frac{a-165}{10}\right)=0{,}9$$

y buscamos en la tabla el valor mas próximo al obtenido $\frac{a-165}{10}=1{,}28\ \rightarrow a=177{,}8$

Se deduce que el 10% de los alumnos mas altos miden mas de 177,8 cm.

APROXIMACIÓN DE LA BINOMIAL POR LA NORMAL

Se demostró que una variable X que sigue una distribución binomial $B(n,p)$ de media $\mu = n \cdot p\ y\ desviacion\ \sigma = \sqrt{npq}$, puede aproximarse por una variable X' normal $N(\mu,\sigma) = N(np,\sqrt{npq})$, y que esta aproximación es buena siempre que $n \cdot p \geq 5\ y\ n \cdot q \geq 5$. El siguiente paso que tenemos que topar es tipificar la variable para poder trabajar con la tabla, tal y como lo hemos explicado anteriormente.

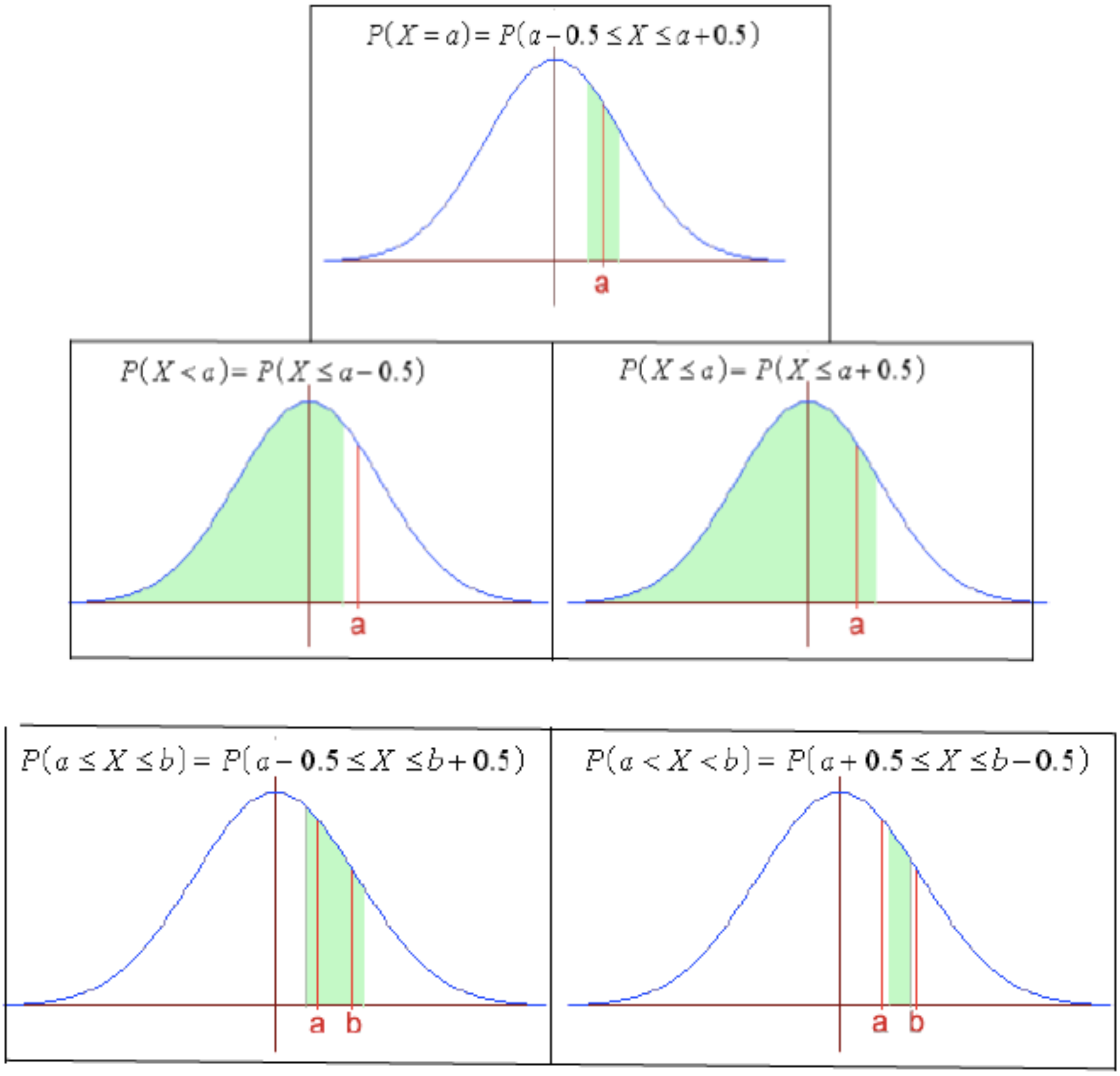
De esta manera, podremos resolver problemas relacionados con distribuciones binomiales donde el numero de repeticiones es elevado.

$$\begin{tikzcd} X\ binomial\ B(n,p) \arrow[r, "\text{Normaliza}"] & X'\ normal\ N(np,\sqrt{npq}\ ) \arrow[d, "\text{Tipificar}"] \\ & Z = \dfrac{X-np}{\sqrt{npq}}\ \ normal\ N(0,1) \end{tikzcd}$$

Es necesario, sin embargo, realizar algunas correcciones para salvar las diferencias entre discretas y continuas, dado que en las primeras existe la probabilidad de que la variable tome un valor concreto, y en las segundas, este siempre es cero. Por esta razón, en la distribución binomial no es lo mismo

$P(X < a)\ que\ P(X \leq a)$.

Veamos con se compensan estas diferencias:



Se observa que el hecho de sumar o restar $0'5$ al valor de a, se produce dependiendo de si dicho valor a debe estar o no incluido en el calculo de la probabilidad.

DISTRIBUCIÖN MUESTRAL DE MEDIAS

TEOREMA CENTRAL DEL LIMITE

Dada una población de media μ y desviación σ, consideramos todas las muestras aleatorias de tamaño n y la variable aleatoria $\bar{X}$ que a cada muestra le hace corresponder su media, es decir, $\bar{X} = \{\overline{x_1}, \overline{x_2}, \dots, \overline{x_k}\}$ entonces se cumple:

- la media de las medias de las muestras es la misma que la de la población.
- La desviación típica de las medias es $\frac{\sigma}{\sqrt{n}}$
- La variable $\bar{X}$ que sigue una distribución normal siempre que la distribución inicial lo sea o, si no es así, siempre que el tamaño de las muestras sea igual o superior a treinta. $(n \geq 30)$.

Por tanto, si una población tiene media μ y desviación típica σ, y consideramos todas las muestras aleatorias de tamaño n (n cualquiera si la población es ``normal´´ y $n \geq 30$ si no lo es), entonces las medias de esas muestras siguen aproximadamente una distribución normal de media μ y desviación $\frac{\sigma}{\sqrt{n}}$, es decir $\bar{X}$ sigue una distribución normal $N(\mu, \frac{\sigma}{\sqrt{n}})$.

Se deduce que cuento mayor sea el tamaño de las muestras mejor es la aproximación y mas pequeña es la desviación.

DISTRIBUCIONES MUESTRALES DE PROPORCIONES

Supongamos que conocemos que el $40\ \%$ de los alumnos de bachillerato del centro realiza alguna actividad solidaria en su tiempo libre. Si eligiéramos un alumno al azar, la probabilidad de que realizase actividades solidarias (éxito) Seria $P = 0{,}4$ y estaríamos ante una prueba binomial.

Seleccionamos una muestra de 50 alumnos y nos preguntamos cual será la proporción (p_r) de alumnos solidarios en dicha muestra.

Dado que hay $n = 50$ alumnos, seguirá una distribución binomial $B(50, 0´4)$.

Teniendo en cuenta que:

$$np = 20 \geq 5$$

$$nq = 30 \geq 5$$

podremos aproximar la variable por una distribución normal

$$N(np, \sqrt{npq}), es\ decir\ N(20{,}3´46)$$

Como la proporción de alumnos solidarios en la muestra es $p_r = \frac{X}{50}$, la variable p_r seguirá una distribución $N\left(\frac{np}{n}, \frac{\sqrt{npq}}{n}\right) = N\left(p, \sqrt{\frac{pq}{n}}\right)$ en este caso $N\left(\frac{20}{50}, \frac{3{,}46}{50}\right) = N(0'4{,}0'069)$.

Podemos generalizar entonces:

Si en una población aparece una característica binomial determinada (el individuo la tiene o no) con una proporción p, entonces la proporción de individuos con dicha característica (p_r) en las muestras de tamaño $n, siendo\ n \geq 30$, sigue una distribución $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

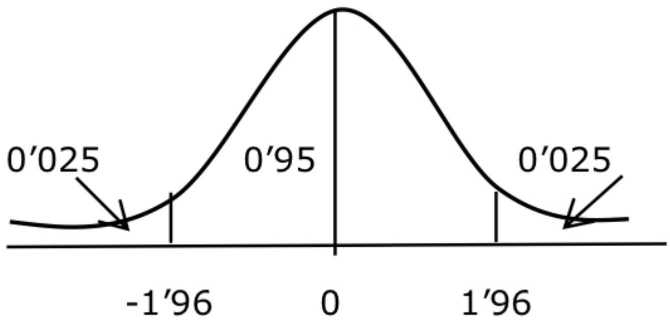
A esta distribución se le llama distribución muestral de proporciones.

INTERVALOS DE CONFIANZA

Dada una variable $X\ con\ media\ \mu$, llamamos intervalo característico correspondiente a una probabilidad p, al intervalo centrado en μ $(\mu - k, \mu + k)$ tal que la probabilidad de p es la probabilidad de que X pertenezca a ese intervalo.

Como ejemplo, supongamos que la variable es $Z\ N(0,1)$. Queremos hallar el intervalo simétrico respecto a la media tal que la probabilidad de que Z este dentro es 95 %, es decir, $p = 0`95$.

Por simetría, si $0´95$ es el área dentro del intervalo, el $0´05$ restante se reparte idénticamente entre las dos colas : $0´025\ cada\ una$.



$P(-k < Z < k) = 0´95 \rightarrow P(Z < k) = 0´975$ y buscando en las tablas el valor mas próximo obtendremos $k = 1´96$.

Luego $P(-1´96 < Z < 1´96) = 0´95$, es decir, hay un 95 % de probabilidad de que Z tome valores en ese intervalo.

Realizando el mismo proceso obtendríamos para el 90 % $k = 1´645\ y\ para$ 99 % $k = 2´575$.

Podemos afirmas que $P\left(-z\alpha/_2 < Z < z\alpha/_2\right) = 1 - \alpha$

Los valores críticos mas utilizados son:

$1-\alpha$	$0´9$	$0´95$	$0´99$
$z\alpha/_2$	$1´645$	$1{,}96$	$2´575$

INTERVALOS DE CONFIANZA PARA LA MEDIA DE LA POBLACIÓN

Si deseamos estimar la media μ de una población de la que conocemos la desviación típica σ, extraemos una muestra de tamaño n de la que se obtiene una media muestral $\bar{x}$.

Si la población de una partida sigue una distribución normal o bien $n \geq 30$, entonces el intervalo de confianza de μ con un nivel de confianza de $(1-\alpha)\ \%$ es:

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\ ,\quad \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Si en algunos de los casos se desconociera la desviación de la población podríamos realizar el intervalo con la desviación de la muestra siempre y cuando la muestra sea lo suficientemente grande como para eludir el error que cometeremos.

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\ ,\quad \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

Para definir el error admisible tendremos en cuenta el intervalo de confianza. Lógicamente, el error máximo será el radio del intervalo:

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

Determinar el tamaño de la muestra:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

INTERVALOS DE CONFIANZA PARA LA PROPORCIÓN

Se tata ahora de encontrar un intervalo en el que afirmar, con un nivel de confianza determinada, que se encuentra la proporción total de la población.

Sabemos que las proporciones de las muestras (siempre que $n \geq 30$) siguen una distribución $N\left(p, \sqrt{\frac{pq}{n}}\right)$ y siguiendo un procedimiento totalmente análogo al de las medias muéstrales, podemos afirmar que el intervalo de confianza para la proporción de individuos de una población que cumplen una determinada característica:

$$\left(P_r - z\alpha/_2 \cdot \sqrt{\frac{pq}{n}} \ , \ P_r + z\alpha/_2 \cdot \sqrt{\frac{pq}{n}}\right)$$

El error máximo cometido será igualmente el radio del intervalo:

$$E = z\alpha/_2 \cdot \sqrt{\frac{pq}{n}}$$

Determinar el tamaño de la muestra:

$$n = \frac{p \cdot q}{\left(\frac{E}{z\alpha/_2}\right)^2}$$

ALGO FUNDAMENTAL:

$$< \boldsymbol{significa} \rightarrow \begin{cases} \boldsymbol{menor\ que} \\ \boldsymbol{menos\ que} \end{cases} \qquad > \boldsymbol{significa} \rightarrow \begin{cases} \boldsymbol{mayor\ que} \\ \boldsymbol{mas\ que} \end{cases}$$

$$\leq \boldsymbol{significa} \rightarrow \begin{cases} \boldsymbol{menor\ o\ igual\ que} \\ \boldsymbol{como\ mucho} \\ \boldsymbol{como\ maximo} \\ \boldsymbol{a\ lo\ sumo} \end{cases} \qquad \geq \boldsymbol{significa} \rightarrow \begin{cases} \boldsymbol{mayor\ o\ igual\ que} \\ \boldsymbol{como\ poco} \\ \boldsymbol{como\ minimo} \\ \boldsymbol{al\ menos, por\ lo\ menos} \end{cases}$$

$$\boldsymbol{NIVEL\ SIGNIFICACION\% = 100\% - NIVEL\ DE\ CONFIANZA\%}$$

$$\boldsymbol{LONGITUD = 2 \cdot ERROR}$$