UNIDAD 8 : DISTRIBUCIONES BINOMIAL Y NORMAL

1. VARIABLES ALEATORIAS

Una variable aleatoria es una función que asigna **a cada suceso del espacio muestral de un experimento un número real.**

Si el resultado del experimento es numérico porque contamos o medimos, los posibles valores de la variable coinciden con los resultados del experimento.

Si el resultado del experimento es cualitativo, hacemos corresponder a cada resultado un número siguiendo algún criterio, es decir definiendo una variable aleatoria.

En un mismo experimento aleatorio se pueden asignar múltiples variables aleatorias

Tipos de variables aleatorias:

1. Variable aleatoria discreta (x). Se le denomina variable porque puede tomar diferentes valores, aleatoria, porque el valor tomado es totalmente al azar y discreta porque solo puede tomar valores enteros y un número finito de ellos.

Ejemplos:
x→ Variable que nos define el número de burbujas por envase de vidrio que son generadas en un proceso dado.
x→0, 1, 2, 3, 4, 5, etc, etc. burbujas por envase
x→Variable que nos define el número de productos defectuosos en un lote de 25 productos.
x→0, 1, 2, 3,....,25 productos defectuosos en el lote

x→Variable que nos define el número de alumnos aprobados en la materia de probabilidad en un grupo de 40 alumnos.
x→0, 1, 2, 3, 4, 5,....,40 alumnos aprobados en probabilidad

Con los ejemplos anteriores nos damos cuenta claramente que los valores de la variable x siempre serán enteros, nunca fraccionarios.

2. Variable aleatoria continua (x). Se le denomina variable porque puede tomar diferentes valores, aleatoria, porque los valores que toma son totalmente al azar y continua porque puede tomar tanto valores enteros como fraccionarios y un número infinito de ellos.

x→Variable que nos define la longitud de un cable o circuito utilizado en un arnés de auto
x→20.5 cm, 20.1, 20.0, 19.8, 20,6, 20.0, 20.0

x→Variable que nos define la concentración en gramos de plata de algunas muestras de mineral
x→14.8gramos, 12.0, 10.0, 42.3, 15.0, 18.4, 19.0, 21.0, 20.8

Las variables descritas anteriormente nos generan una distribución de probabilidad, las que pueden ser.

1) **Distribución de probabilidad discreta.**
2) **Distribución de probabilidad continua.**

2. DISTRIBUCIÓN BINOMIAL o DE BERNOULLI

Es una de las distribuciones de probabilidad más utilizadas en la práctica estadística. Se emplea cuando el fenómeno de estudio queda determinado por dos sucesos complementarios: si/no; hombre/mujer; nacional/extranjero; trabajador en activo/parado;… En general, esas dos situaciones pueden considerarse resultados de un experimento aleatorio y a los sucesos contrarios, sin que indique valoración alguna, suelen llamárseles éxito y fracaso.

Características básicas de una distribución binomial :

- Cada prueba del experimento aleatorio presenta dos únicas opciones, que puede designarse como éxito (E) y fracaso (F).
- Se realizan n ensayos del experimento, independientes unos de otros e idénticos.
- La probabilidad de éxito es constante a lo largo de las n pruebas: $P(E) = p$.
- La probabilidad de fracaso también es constante: $P(F) = q = 1 - p$.

Veámoslo con un ejemplo:

Tiramos un dado 7 veces y contamos el número de cincos que obtenemos. ¿Cuál es la probabilidad de obtener tres cincos?.

Este es un típico ejemplo de distribución binomial, pues estamos repitiendo 7 veces el experimento de lanzar un dado. ¿Cuál es nuestro éxito?.

Evidentemente, sacar un 5, que es en lo que nos fijamos.

El fracaso, por tanto, será no sacar 5, sino sacar cualquier otro número.

Por tanto, Éxito = E = "sacar un 5" $\Longrightarrow p(E) = \dfrac{1}{6}$

Fracaso = F = "no sacar un 5" $\Longrightarrow p(F) = \dfrac{5}{6}$

Para calcular la probabilidad que nos piden, fijémonos en que nos dicen que sacamos 3 cincos y por lo tanto tenemos 3 éxitos y 4 fracasos, ¿de cuántas maneras pueden darse estas posibilidades?. Podríamos sacar 3 cincos en las 3 primeras tiradas y luego 4 tiradas sin sacar cinco, es decir: EEEFFFF Pero también podríamos sacar EFEFFFE, es decir que en realidad estamos calculando de cuántas maneras se pueden ordenar 4 fracasos y 3 éxitos. Recordando las técnicas combinatorias, este problema se reduce a calcular las permutaciones con elementos repetidos:

$$P_7^{3,4} = \frac{7!}{3! \cdot 4!} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35\text{formas}$$

Y por tanto, como $p(E) = \dfrac{1}{6}$ y tengo 3 éxitos y $p(F) = \dfrac{5}{6}$ y tengo 4 fracasos:

$$p(\text{tener 3 éxitos y 4 fracasos}) = 35 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = 0'0781$$

Formalizando lo obtenido, en una variable binomial con 7 repeticiones y con probabilidad de éxito $\frac{1}{6}$, la probabilidad de obtener 3 éxitos es 0'0781, y lo expresaríamos:

$$\text{Bin}\left(7; \frac{1}{6}\right), \text{entonces } p(X = 3) = 0'0781$$

Como repetir este proceso sería bastante penoso en la mayoría de los casos, lo mejor es recurrir a la siguiente fórmula que expresa la probabilidad de obtener cierto número de éxitos en una distribución binomial:

2.1 Cálculo de probabilidades de una distribución binomial

Si realizamos n veces un experimento en el que podemos obtener éxito, E, con probabilidad p y fracaso, F, con probabilidad q ($q = 1 - p$), diremos que estamos ante una distribución binomial de parámetros n y p, y lo representaremos por Bin(n;p). En este caso la probabilidad de obtener k éxitos viene dada por:

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

k = 0,1,2,...

Nota1:

Observar que las probabilidades de éxito y fracaso son complementarias, es decir, q = 1-p y p = 1-q, por lo que basta saber una de ellas para calcular la otra.

Nota 2:

La elección de éxito o fracaso es subjetiva y queda a elección de la persona que resuelve el problema,

2.2 Media y Desviación típica

Aunque no se demostará, en una distribución binomial Bin(n;p), el número esperado de éxitos o media, viene dado por $\bar{x} = n \cdot p$. (Recordemos que la media es una medidad de centralización).

La desviación típica, σ , que es una medida de dispersión y mide lo alejados que están los datos de la media, viene dada por $\sigma = \sqrt{n \cdot p \cdot q}$.

3. DISTRIBUCIÓN DE PROBABILIDAD CONTINUA. LA DISTRIBUCIÓN NORMAL

Una variable estadística se llama continua cuando puede tomar todos los valores de un intervalo. Así, por ejemplo, son variables estadísticas continuas, las estaturas y pesos de los individuos, los tiempos de espera de un autobús, el tamaño de una determinada variedad de manzanas, etc.

Para estas distribuciones, la probabilidad de un valor concreto es 0, pues el número de casos posibles es infinito. Por ejemplo, la probabilidad de que una persona mida exactamente 172,12345678910… cm es 0; altura tan improbable como que mida exactamente 172,000…cm.

En cambio, la probabilidad de que una persona mida entre 171,5 cm y 172,5 cm sí podrá calcularse.

Esto es, si X es la variable que mide la estatura de una persona, se tendrá:

$P(X = 172{,}12345...) = 0$; $P(X = 172{,}000...) = 0$.

En cambio, $P(171{,}5 < X < 172{,}5) = ?$, valor que dependerá de la población de estudio.

Al estudiar aspectos tan cotidianos como:

- Caracteres morfológicos de individuos (personas, animales, plantas) de una misma raza. como tallas, pesos, envergaduras, etc.
- Caracteres fisiológicos, como el efecto de una misma dosis de un fármaco, o de una misma cantidad de abono.
- Caracteres sociológicos, como el consumo de ciertos productos por individuos de un mismo grupo humano.
- Caracteres psicológicos, como el cociente intelectual, grado de adaptación a un medio.
- Caracteres físicos, como la resistencia a la rotura de ciertas piezas. . .

todos ellos tienen en común que se distribuyen "normalmente". ¿Qué quiere decir esta expresión?. Pués, por ejemplo, si hacemos una estadística para conocer la altura de 1400 mujeres y representamos los resultados en un diagrama de barras, obtenemos:

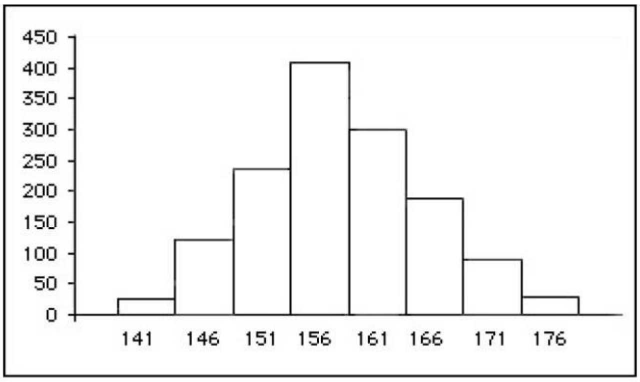


Figura 3.1: Distribución de estaturas de 1400 mujeres

Las gráficas de este tipo son muy corrientes: Hay pocos individuos en los extremos y un aumento paulatino hasta llegar a la parte central del recorrido, donde está la mayoría de ellos.

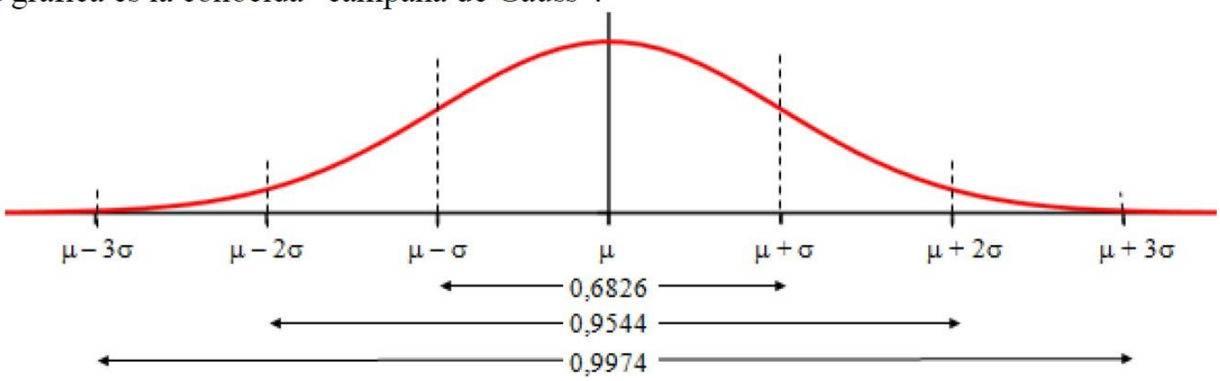
3.1 ¿Que és una distribución normal?

Una variable con distribución normal queda totalmente definida por su media μ y por su desviación típica σ. Se denota como $N(\mu, \sigma)$.

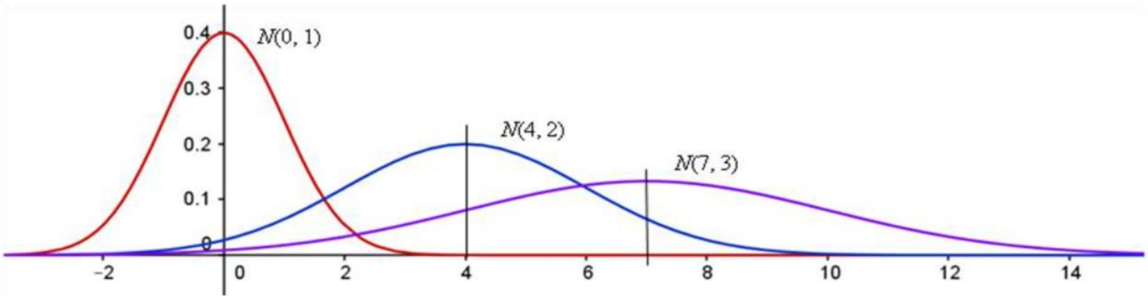
→ La expresión analítica de la función de densidad de la distribución normal es

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

Su gráfica es la conocida "campana de Gauss".



→ La variación de la media y de la desviación típica originan cambios en la curva, desplazándose a izquierda o derecha o haciéndose más esbelta o más baja, como puede verse en la figura.



Recuérdese que la desviación típica es una medida de la dispersión de los elementos de una población. Una desviación típica más grande significa que los datos son más heterogéneos; por eso las curvas normales con mayor desviación típica son más planas, que es un indicador de que los datos pueden estar más alejados de la media. (Cuando la igualdad entre los datos es grande, la desviación típica es pequeña; y al revés).

3.2 Distribución normal estándar

De entre todas las curvas normales $N(\overline{x};\sigma)$, la más sencilla, usada y conocida es aquella que tiene por media 0 y por desviación típica 1, N(0, 1).

Esta normal estándar se suele representar por Z.

La gráfica de esta curva se denomina campana de Gauss y se puede observar en la figura:

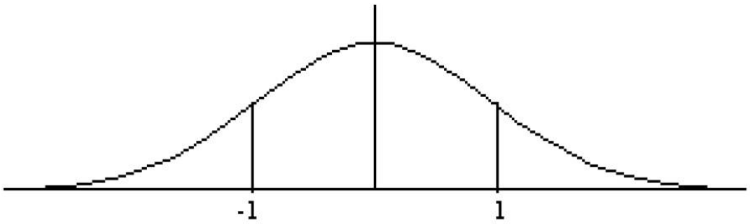


Figura 3.3: Distribución normal $N(0;1)$. El máximo está en $(0, \frac{1}{\sqrt{2\cdot\pi}})$

Su función de densidad será:

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

Definición :

Para un valor cualquiera k, definimos la probabilidad de que la distribución Z, N(0;1) , sea menor o igual que k como:

$p(Z \leq k)$= "Área encerrada bajo la curva normal N(0,1) desde $-\infty$ hasta k"

(es decir la parte rayada de la figura siguiente).

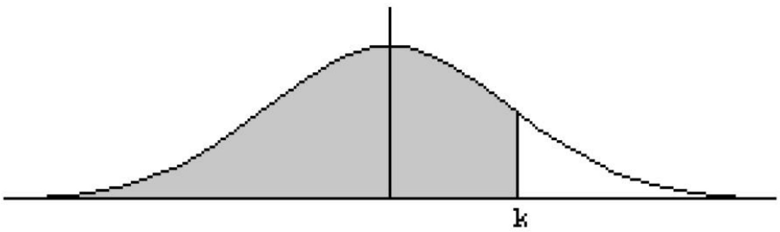
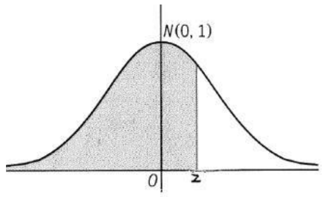


Figura 3.4: Área encerrada por la curva normal desde $-\infty$ hasta k

Ahora bien, ¿cómo calcular dicha área?. Fácil: Dichas áreas o probabilidades se encuentran tabuladas.

DISTRIBUCIÓN NORMAL $N(0, 1)$



$\Phi(z) = P(Z \leq z)$

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad -\infty < z < +\infty$$

x	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	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
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	0,8830
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	0,9564	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	0,9834	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	0,9904	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	0,9955	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	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

3.3 Uso de las tablas de la normal estándar

La normal N(0;1) se encuentra tabulada, para valores a partir de 0 y hasta 3'99. Si por ejemplo queremos calcular $p(Z \leq 2'78)$, hemos de realizar los pasos:

1. Buscar la parte entera y las décimas en la primera columna (en este caso 2'7).
2. Buscar las centésimas en la primera fila (en este caso 8).
3. En el punto común a la fila y la columna que hemos encontrado, tenemos la probabilidad buscada, en este caso 0'9973.

Por tanto $p(Z \leq 2'78) = 0'9973$.

Si queremos calcular una probabilidad de un valor mayor que 3'99, basta fijarse en que las probabilidades correpondientes a valores tales como 3'62 y mayores ya valen 0'9999 (prácticamente 1). Por eso, para estos valores mayores que 3'99, diremos que la probabilidad es aproximadamente 1. Así:

$$p(Z \leq 5'62) \approx 1$$

aunque no aparezca en la tabla.

Por otra parte, fijémonos en que en este tipo de distribuciones no tiene sentido plantearse probabilidades del tipo p(Z=k), ya que siempre valen 0, al no encerrar ningún área. Por tanto, si nos pidiesen p(Z=3'2), basta decir que p(Z=3'2)=0.

Así, al pasar al complementario, si tenemos $Z \geq k$, su complementario será $Z < k$, pero como incluir k no influye en la probabilidad,al calcular probabilidades podemos escribir:

$$p(Z \geq k) = 1 - p(Z < k) = 1 - p(Z \leq k)$$

Sólo se puede hacer esto en distribuciones continuas, en el caso de la binomial esto no se puede hacer y hay que ser cuidadosos con el paso al complementario.

Cálculo de otras probabilidades:

1. Si k es positivo y queremos calcular $p(Z \geq k)$, es decir el área rayada:

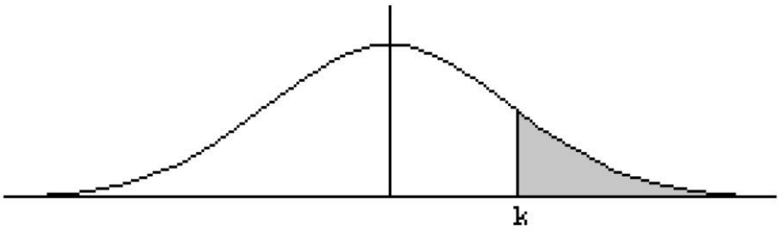


Figura 3.5: $p(Z \geq k)$. Basta pasar al complementario

basta pasar al complementario, es decir: $p(Z \geq k) = 1 - p(Z \leq k)$ y esta última probabilidad ya se encuentra tabulada.

Ejercicio: Calcular $p(Z \geq 0'3)$ y $p(Z \geq 2'07)$.

2. Si k es positivo y queremos calcular $p(Z \leq -k)$, es decir el área: por simetría, $p(Z \leq -k) =$

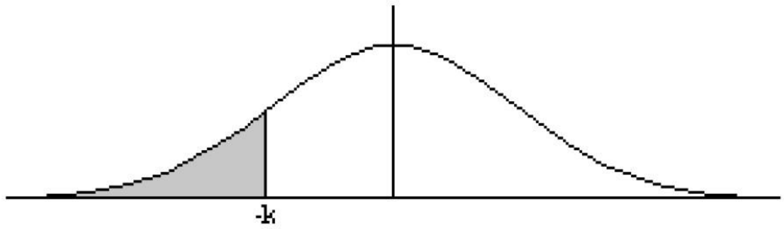


Figura 3.6: $p(Z \leq -k)$.Las probabilidades de valores negativos no están tabuladas

$p(Z \geq k)$ y ésta se calcula como en el caso anterior. Se puede observar la igualdad de áreas en la figura:

Figura 3.7: $p(Z \leq -k) = p(Z \geq k)$. La simetría permite reducir este caso al anterior

Ejercicio: Calcular $p(Z \leq -0'78)$ y $p(Z \leq -3'2)$.

3. Si k es positivo y queremos calcular $p(Z \geq -k)$, es decir el área rayada:

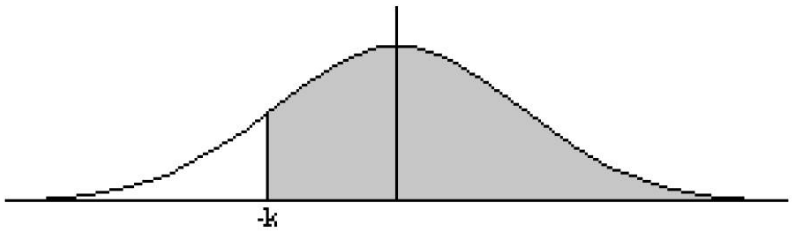


Figura 3.8: $p(Z \geq -k)$

entonces, por simetría $p(Z \geq -k) = p(Z \leq k)$:

Figura 3.9: $p(Z \geq -k) = p(Z \leq k)$.La simetría permite reducir este caso al que ya está tabulado

Ejercicio: Calcular $p(Z \geq -0'96)$ y $p(Z \geq -1'01)$.

4. Probabilidades comprendidas entre dos valores,$p(k_1 \leq Z \leq k_2)$,es decir el área rayada:

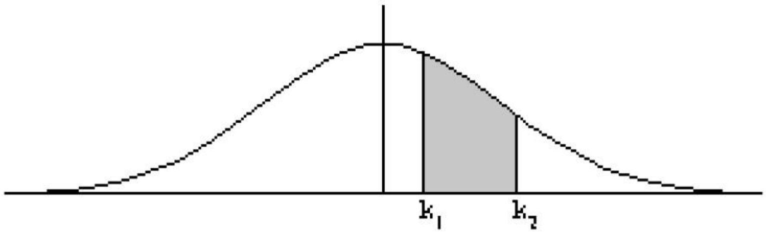


Figura 3.10: $p(k_1 \leq Z \leq k_2)$. Probabilidad comprendida entre dos valores

se calcula restando las áreas:

Figura 3.11: $p(Z \leq k_2)$ en la primera imagen.$p(Z \leq k_1)$ en la segunda. Al restar obtenemos el área pedida.

Se quita la parte correspondiente a $Z \leq k_1$,$p(Z \leq k_2) - p(Z \leq k_1)$.

Ejercicio: Calcular $p(-0'96 \leq Z \leq 1'49)$ y $p(-1'32 \leq Z \leq -0'57)$.

Ejercicio: Calcular p(Z=2), $p(Z \leq 2)$, $p(Z \geq 2)$, $p(Z \leq -2)$, $p(Z \geq -2)$, $p(-2 \leq Z \leq 2)$, $p(0'81 \leq Z \leq 1'33)$.

Cálculo del valor de z a partir de la probabilidad asociada

Hasta ahora nos han dado la distribución normal N(0;1) y nos pedían $p(Z \leq k)$ siendo k un cierto número, y nos pedían calcular dicha probabilidad.

Ahora bien, otra pregunta puede ser: Dado que en una normal N(0;1) sabemos que $p(Z \leq k) = 0'9573$, ¿quién es k?.

La resolución es bien sencilla. Basta buscar 0'9573 *dentro de la tabla* de la distribución normal, y lo encontramos en el cruce de la fila 1'7 con la columna 2, y por lo tanto k debe ser 1'72.

En caso de que el valor a buscar no aparezca directamente dentro de la tabla de la distribución normal, pueden ocurrir dos posibilidades:

a) Si el valor se encuentra entre dos valores de la tabla y a la misma distancia (aproximadamente) de cada uno de ellos, por ejemplo: $p(Z \leq k) = 0'7982$. En este caso el valor buscado será la media entre los valores extremos.

Si buscamos en la tabla este valor no aparece directamente, sino que se encuentra entre los valores 0'7967 (que corresponde a 0'83) y 0'7996 (que corresponde a 0'84). Por tanto el valor de k será:

$$k = \frac{0'83 + 0'84}{2} = 0'835$$

b) Si el valor está entre dos valores, pero muy cercano a uno de ellos, directamente tomamos este valor, por ejemplo: $p(Z \leq k) = 0'7970$. El valor más cercano es 0'9767 (que corresponde a 0'83) y como el valor buscado está muy cerca de él, entonces directamente k=0'83.

b) Si la pregunta es: ¿cuánto debe valer Z para $P(-z_2 < Z < z_2) = 0,80$? Se procede así:
Como

$$P(-z_2 < Z < z_2) = P(Z < z_2) - P(Z < -z_2) = P(Z < z_2) - [1 - P(Z < z_2)] \Rightarrow$$
$$\Rightarrow P(-z_2 < Z < z_2) = 2 \cdot P(Z < z_2) - 1.$$

Así: $P(-z_2 < Z < z_2) = 0,80 \Rightarrow 2 \cdot P(Z < z_2) - 1 = 0,80 \Rightarrow P(Z < z_2) = 0,90 \Rightarrow z_2 = 1,28$.

c) Un caso que se presenta con frecuencia es encontrar el intervalo $(-z, z)$ que contiene el 95% de los datos de la variable estadística. Esto es, hallar el valor z tal que $P(-z < Z < z) = 0,95$.
Por el ejemplo anterior:

$$P(-z < Z < z) = 0,95 \Rightarrow 2 \cdot P(Z < z) - 1 = 0,95 \Rightarrow P(Z < z) = 0,9750 \Rightarrow z = 1,96.$$

3.4 Tipificación

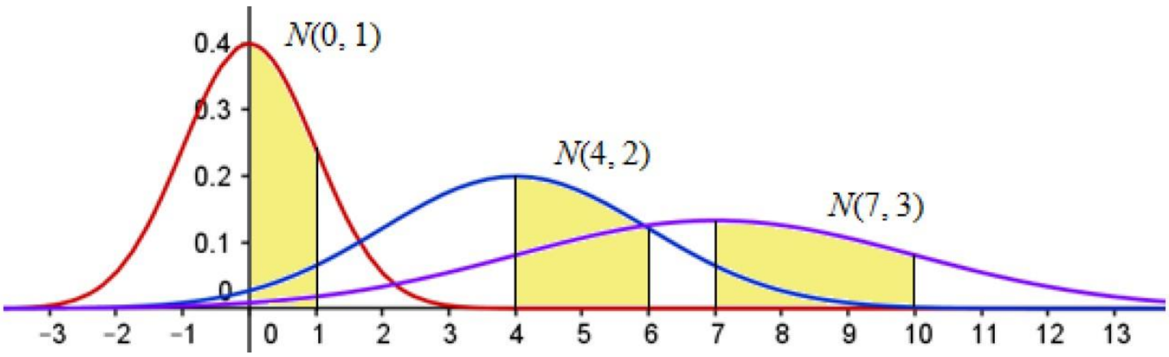
Se puede demostrar que:

Si X ∈N (μ ,σ) , entonces $Z = \frac{X - \mu}{\sigma}$ **∈N (0 , 1)**

Este proceso se llama **tipificación**

Además **P (a < X < b) = P** $\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$ **,**

- El proceso de tipificación se puede explicar gráficamente con ayuda de la siguiente figura. En ella, los recintos coloreados tienen la misma área.



NOTA : Si queremos calcular un valor en una distribución normal cualquiera que cumple una determinada probabilidad, tendremos que tipificar previamente:

Por ejemplo, si X sigue una normal N(6;3) y $p(X \leq k) = 0'9082$, calcula k.
Tipificando:

$$p\left(\frac{X-6}{3} \leq \frac{k-6}{3}\right) = 0'9082 \longrightarrow p\left(Z \leq \frac{k-6}{3}\right) = 0'9082$$

Y buscando en la tabla,

$$\frac{k-6}{3} = 1'33 \Rightarrow k - 6 = 3'99 \Rightarrow k = 9'99$$

EJEMPLO

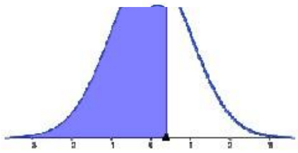
Veamos en el siguiente ejemplo, el manejo de la tipificación y de la tabla normal. Supongamos que una variable X, que describe la nota media de 2º de Bachillerato del alumnado de un instituto sigue una distribución normal de media 6.4 años con desviación típica 2. Abreviadamente $X \to N(6.4, 2)$. Elegido un alumno al azar, calculemos las probabilidades siguientes:

a) Tenga una nota media de un 7.

Evidentemente $P(X=7)=0$ ya que se trata de un valor puntual y la distribución normal es continua.

b) Tenga una nota media de menos de 7.2.

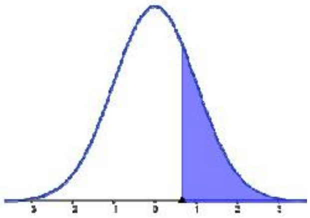
$$P(X \le 7.2) \overset{Tipificando}{=} P\left(\frac{X-6.4}{2} \le \frac{7.2-6.4}{2}\right) = P(Z \le 0.4) \overset{Tabla}{=} 0.6554$$



c) Tenga más de un 7.75 de nota media.

$$P(X \ge 7.75) \overset{Tipificando}{=} P\left(\frac{X-6.4}{2} \ge \frac{7.75-6.4}{2}\right) =$$

$$P(Z \ge 0.675) == 1 - P(Z \le 0.675) \overset{Tabla}{=} 1 - 0.7517 = 0.2483$$

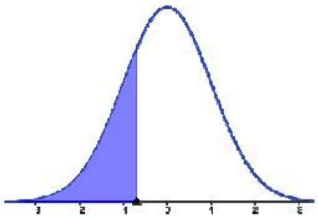


d) Sabiendo que hay 76 alumnos en total, ¿Cuantos/as alumnos/as habrá con menos de un 5 de nota media?

$$P(X \le 5) \overset{Tipificando}{=} P\left(\frac{X-6.4}{2} \le \frac{5-6.4}{2}\right) = P(Z \le -0.7) \overset{Simetría}{=}$$

$$= P(Z \ge 0{,}7) = 1 - P(Z \le 0.7) \overset{Tabla}{=} = 1 - 0.7580 = 0.242.$$

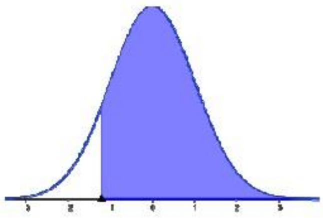
Por lo tanto, habrá $0.242 \cdot 76 = 18.392 \simeq 18$ alumnos/as



e) Tenga más de un 4 de nota media.

$$P(X \ge 4) \overset{Tipificando}{=} P\left(\frac{X-14.5}{2} \ge \frac{4-6.4}{2}\right) = P(Z \ge -1.2) \overset{Simetría}{=}$$

$$= P(Z \le 1.2) \overset{Tabla}{=} 0.8849$$

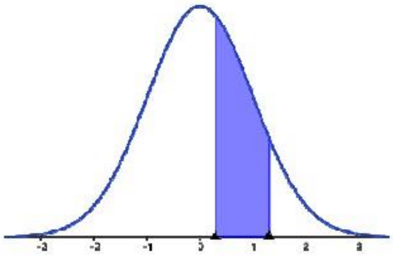


f) Tenga un notable (entre 7 y 9)

$$P(7 \le X \le 9) = P\left(\frac{7-6.4}{2} \le \frac{X-6.4}{2} \le \frac{9-6.4}{2}\right) =$$

$$P(0.3 \le Z \le 1.3) = P(Z \le 1.3) - P(Z \le 0.3) \overset{Tabla}{=}$$

$$= 0.9032 - 0.6179 = 0.2853$$

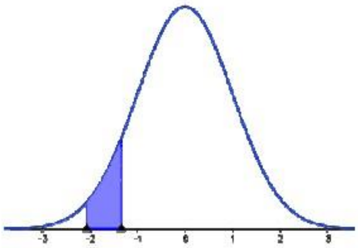


g) Tenga entre un 2.38 y un 3.72.

$$P(2.28 \le X \le 3.72) = P\left(\frac{2.28-6.4}{2} \le \frac{X-6.4}{2} \le \frac{3.72-6.4}{2}\right) =$$

$$P(-2.06 \le Z \le -1.34) \overset{Simetría}{=} P(2.06 \le Z \le 1.34) =$$

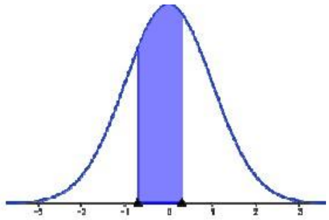
$$P(Z \le 2.06) - P(Z \le 1.34) \overset{Tabla}{=} 0.9803 - 0.9099 = 0.0704$$



DPTO. MATEMATICAS

IES SANTOMÉ

h) Tenga un suficiente o un bien (entre un 5 y un 7)



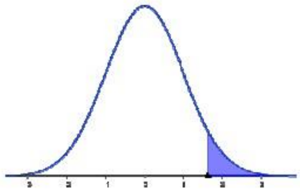
$$P(5 \le X \le 7) = P\left(\frac{5-6.4}{2} \le \frac{X-6.4}{2} \le \frac{7-6.4}{2}\right) =$$

$$P(-0.7 \le Z \le 0.3) = P(Z \le 0.3) - P(Z \le -0.7) =$$

$$\overset{Simetría}{=} P(Z \le 0.3) - P(Z \ge 0.7) = P(Z \le 0.3) - (1 - P(Z \le 0.7)) =$$

$$\overset{Tabla}{=} 0.6179 - 1 + 0.7580 = 0.3759$$

i) Se sabe que sólo un 5% del alumnado puede obtener una matrícula de honor. Halla la nota media que debe alcanzar un alumno/a para que se le conceda. Buscamos una nota k que cumpla:

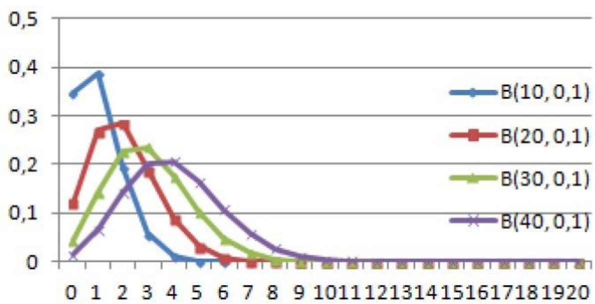


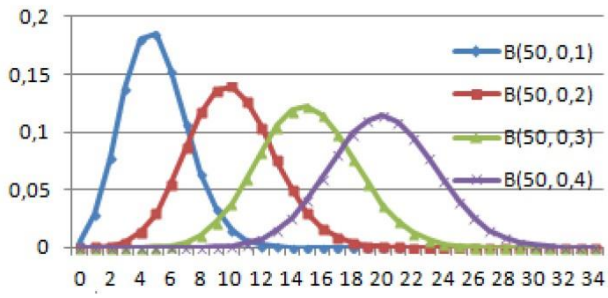
$$P(X \ge k) = 0.05 \rightarrow P(X \le k) = 0.95 \overset{Tipificando}{\rightarrow} P\left(\frac{X-6.4}{2} \le \frac{k-6.4}{2}\right) = 0.95 \rightarrow$$

$$\rightarrow P\left(Z \le \frac{k-6.4}{2}\right) = 0.95 \overset{Tabla}{\rightarrow} \frac{k-6.4}{2} = 1.645 \rightarrow k = 2 \cdot 1.645 + 6.4 = 9.69$$

4. APROXIMACIÓN DE LA DISTRIBUCIÓN BINOMIAL MEDIANTE UNA NORMAL

Al estudiar la distribución binomial, $B(n, p)$, se ha indicado que para valores grandes de n el cálculo de probabilidades se hace engorroso; además, los valores de probabilidad pueden resultar muy pequeños. Así, por ejemplo, para la binomial $B(50, 0{,}12)$, las probabilidades de r = 6 y de r = 12 son: $P(X=6) = 0{,}171185935$ y $P(X=12) = 0{,}0084088$. Más complicada resulta, por ejemplo, calcular la probabilidad de que $X > 10$ o de que $4 < X < 12$.
En las siguientes figuras puede observarse la evolución de las poligonales de frecuencias de la distribución binomial $B(n, 0{,}1)$ cuando n va creciendo, y de la distribución binomial $B(50, p)$ cuando p se acerca a 0,5.





Como ves, cuando n aumenta la poligonal se parece más a una campana de Gauss; y es mucho más evidente cuando n es grande y p se acerca más a 0,5. En general, se admite que la aproximación es buena cuando $n \ge 25$ y el producto $np \ge 5$.

Así, para las poligonales de la figura de la izquierda la única que puede admitirse como aceptable es la $B(40, 0{,}1)$, aunque $np = 4$. En cambio, las binomiales representadas en la figura de la derecha se aproximan bastante bien a campanas de Gauss.

En la práctica , aceptamos la aproximación de la binomial por la normal si :

$$n \geq 30 \ , \ np \geq 5 \ , \ nq \geq 5$$

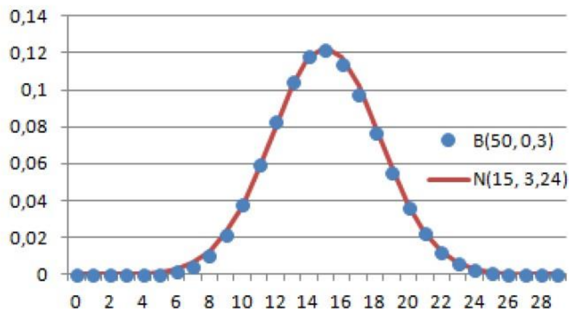
→ <u>Cuando el ajuste sea posible</u>, la distribución normal que mejor se aproxima a la $B(n, p)$ es la que tiene por media y desviación típica la de la distribución binomial. Como la media y desviación típica de la variable $X \approx B(n, p)$ son $\mu = np$ y $\sigma = \sqrt{npq}$, el ajuste se hace por la variable $X' \approx N(np, \sqrt{npq})$, se tipifica haciendo $Z = \dfrac{X'-\mu}{\sigma} = \dfrac{X'-np}{\sqrt{npq}}$.

Con esto, cuando sea preciso calcular probabilidades de una variable X binomial $B(n, p)$ puede hacerse recurriendo a una variable normal X' asociada. Por tanto:

$$P(X < k) = P(X' < k) = P\left(Z < \frac{k-np}{\sqrt{npq}}\right).$$

En la figura adjunta se han dibujado la $B(50, 0{,}3)$ y la $N\left(50 \cdot 0{,}3, \ \sqrt{50 \cdot 0{,}3 \cdot 0{,}7}\right) = N(15, \ 3{,}24)$.

Las probabilidades binomiales se representan por puntos; la normal es la curva continua. Es evidente la gran coincidencia.



<u>6.1 Corrección por continuidad (correcciones de Yates)</u>

La distribución binomial es de variable discreta y, por tanto, tiene sentido calcular probabilidades puntuales. Cuando se aproxima una binomial por una normal, al ser esta última de variable continua no tiene sentido calcular probabilidades puntuales, pues son todas nulas, tal y como ya se ha visto en un apartado anterior.

La aproximación de una variable de distribución binomial X por una normal Y, genera un error que se corrige modificando el intervalo cuya probabilidad se quiere calcular. Así pues:

- $P(X = a) = P(a - 0{,}5 \leqslant Y \leqslant a + 0{,}5)$.
- $P(X \leqslant a) = P(Y \leqslant a + 0{,}5)$ (para que contenga al punto a).
- $P(X < a) = P(Y \leqslant a - 0{,}5)$ (para que no contenga al punto a).
- $P(X > a) = P(Y \leqslant a + 0{,}5)$ (para que no contenga al punto a).
- $P(X \geqslant a) = P(Y \geqslant a - 0{,}5)$ (para que contenga al punto a).
- $P(a \leqslant X \leqslant b) = P(a - 0{,}5 \leqslant Y \leqslant b + 0{,}5)$.

Ejemplo: Se sabe que la probabilidad de que un piloto de Fórmula 1 sufra un reventón en un circuito es de 0.04. Si en una carrera participan 200 conductores, calcula:

a) La probabilidad de que se registren entre 12 y 18 reventones.

Si consideramos como éxito el que sufra un reventón, es evidente que la variable $X \to Bi(200, 0.04)$. Evidentemente, calcular esta probabilidad con la función de distribución o la de probabilidad de la binomial es una tarea bastante larga. Vemos si cumple las hipótesis para aproximarla por una normal: $n = 200 \geq 30$; $np = 200 \cdot 0.04 = 8 \geq 5$ y $nq = 200 \cdot 0.96 = 192 \geq 5$. Así pues, podemos aproximar la variable X por la normal $X' \to Bi\left(200 \cdot 0.04, \sqrt{200 \cdot 0.04 \cdot 0.96}\right) \simeq Bi(8, 2.77)$. Así pues, la probabilidad pedida será, aproximadamente:

$$P(12 \leq X \leq 18) \overset{Yates}{\simeq} P(11.5 \leq X' \leq 18.5) = P\left(\frac{11.5-8}{2.77} \leq \frac{X'-8}{2.77} \leq \frac{18.5-8}{2.77}\right) = P(1.26 \leq Z \leq 3.79) =$$

$$P(Z \leq 3.79) - P(Z \leq 1.26) = 0.99992 - 0.8962 = 0.10372$$

Si lo calculamos con Geogebra, el valor utilizando la binomial sale 0.107, que es bastante aproximado.

b) La probabilidad de que se registren exactamente 5 reventones.

$$P(X = 5) \overset{Yates}{\simeq} P(4.5 \leq X' \leq 5.5) \overset{Tipificando}{=} P(-1.26 \leq Z \leq -0.90) = P(1.26 \leq Z \leq 0.90) = P(Z \leq 1.26) -$$

$$-P(Z \leq 0.90) = 0.8962 - 0.8159 = 0.0803$$

Aunque este cálculo sí se podría hacer fácilmente con la calculadora:

$$P(X = 5) = \binom{200}{5} \cdot 0.04^5 \cdot 0.96^{195} \simeq 0.0906.$$

"La Real Academia define la palabra imposible como algo que no tiene ni facultad ni medios para llegar a ser o suceder, y define improbable como algo inverosímil que no se funda en una razón prudente. Puesto a escoger, a mí me gusta más la improbabilidad que la imposibilidad, como a todo el mundo supongo. La improbabilidad duele menos y deja un resquicio a la esperanza, a la ética. Que David ganara a Goliat era improbable pero sucedió. Un afroamericano habitando la Casa Blanca era improbable, pero sucedió. El amor, las relaciones, los sentimientos, no se fundan en una razón prudente, por eso no me gusta hablar de amores y retos imposibles, sino de amores y retos improbables."