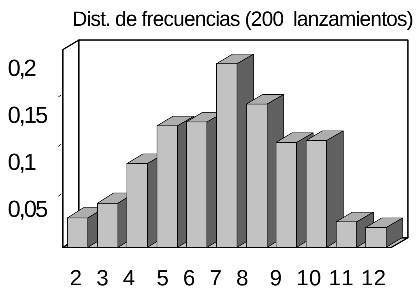
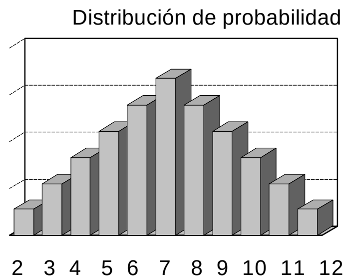
Distribuciones de probabilidad: Binomial y Normal

Recordarás que cuando se quiere estudiar un fenómeno, se recogen una serie de observaciones sobre los valores que presenta y sus frecuencias, confeccionando una tabla estadística, que nos permite conocer el comportamiento de los datos. Esta es una aproximación basada en datos observados en una muestra. Las distribuciones de probabilidad son modelos teóricos de como sería tal distribución, para la población completa. Construimos tablas de frecuencias usando datos reales observados, pero al construir distribuciones de probabilidad, usamos los posibles resultados y sus probables frecuencias.





Así por ejemplo, al contemplar el experimento "observar la suma obtenida al lanzar dos dados", podemos hacer una distribución de frecuencias, relativas resultante de realizar el experimento un nº grande de veces, o por el contrario, una distribución de probabilidad, en la que asignemos a cada resultado su probabilidad. En el primer caso utilizamos la frecuencia relativa de cada dato mientras en el segundo utilizamos la probabilidad de cada resultado.

De esta forma, al utilizar distribuciones de probabilidad, usamos un modelo teórico que correspondería a una distribución perfecta de frecuencias de una población, es decir, el que correspondería al fenómeno, si este se realizara un nº infinito de veces.

La utilidad de estos modelos teóricos perfectos es múltiple.

Si conocemos el modelo teórico al que se ha de adaptar un fenómeno, y variamos algún factor, podremos comparar los resultados obtenidos en un nuevo estudio de frecuencias y saber saber hasta que punto ese determinado factor influye en el fenómeno.

La inferencia estadística, que estudiarás el próximo curso, es importante para la toma de decisiones en muchos campos. Parte de los resultados obtenidos en pequeñas muestras, para inferir datos correspondientes a toda la población.Generalmente de las muestras no somos capaces de saber cual es la distribución de frecuencias de toda la población, pero sí que podremos si estas poblaciones se ajustan a modelos de probabilidad teóricos previamente estudiados.

Variables aleatorias

Al igual que para hacer estudios estadisticos necesitabamos variables que acogieran los valores observados (variables estadísticas), al utilizar distribuciones de probabilidad utilizaremos variables para describir los posibles resultados de un fenómeno. A estas variables se las denomina variables aleatorias. A efectos de notación utilizaremos la misma que la ya usada con variables estadísticas.

Ejemplos:

- *En un exámen tipo test,, de 10 preguntas, podemos definir X como la variable que mide el número de respuestas acertadas. X podrá tomar los valores 0,1,2,...,10*
- *Al lanzar un dado, si X designa el resultado obtenido, podrá tomar los valores 1,2,...,6*
- *Si X es el nº de piezas defectuosas encontradas en un lote de 100, podrá tomar los valores 0,1,2,....,100*
- *Si X designa la talla de los alumnos de la clase, podrá tomar valores comprendidos entre digamos 1,45 m y 2,10 m*

Al igual que con las variables estadísticas, distinguiremos dos tipos, las discretas y las contínuas.

Una variable es **discreta**, si toma sólo un número finito o contable de valores. A este tipo corresponden los tres primeros ejemplos anteriores. Si por el contrario, toma infinitos valores, que no presentan huecos diremos que es **contínua**. Las medidas tales como pesos, tallas, tiempos, temperaturas etc son de este segundo tipo.

Empezaremos por el primer grupo de variables, que es el más sencillo

Función de probabilidad

Imaginemos, que "Binter", ha estudiado durante algun tiempo, la puntualidad de sus tres vuelos diarios desde Las Palmas a Arrecife. Estos vuelos son independientes entre sí, y por tanto el retraso de uno de ellos no afecta a los demás. Ha denominado X al número de vuelos que llega a su hora al aeropuerto de Arrecife, y ha encontrado, que la probabilidad de que X sea igual a 0,1,2,3, viene dada por la siguiente tabla:

Valores (X)	*0*	*1*	*2*	*3*
Probabilidad (P(X))	*0,064*	*0,288*	*0,432*	*0,216*

Al realizar esta operación, ha asignado probabilidad a cada uno de los valores de la variable. A la función P(X), que asigna a cada valor de X, su probabilidad la denominaremos función de probabilidad. Puesto que cada valor es una probabilidad, todos ellos habrán de estar entre o y 1. Como recordarás, la probabilidad del suceso serguro es 1, y dado que 0,1,2,3 son todos los posibles valores de X, su suma ha de ser (salvo redondeo) igual a 1. Estas dos condiciones son necesarias y suficientes para que una asignación sea una función de probabilidad. En consecuencia:

Una función P(X) es una función de probabilidad si verifica: $\begin{cases} 0 \le P(X) \le 1 \\ \Sigma P(X) = 1 \end{cases}$

Ejemplos:

- *Si X representa el nº de hijos varones que puede tener una pareja con dos descendientes, la función de probabilidad asociada a X sería:*

X	0	1	2
P(X)	1/4	2/4	1/4

- *Revisando los resultados dados entre el Real Madrid y El Barcelona durante los últimos 40 años, se ha obtenido la tabla:*

X	0	1	2	3	4	5	6
P(X)	0,1	0,2	0,42	0,21	0,05	0,015	0,005

donde X es el nº de goles marcados por el Real Madrid.

Media y desviación típica

Dado que una distribución de probabilidad no es más que un modelo perfecto de la distribución de frecuencias de una población, podemos calcular su media y desviación típica de forma idéntica, cambiando cada f_i/N por la probabilidad, obteniendo:

$$\begin{cases} \mu = \Sigma XP(X) \\ \sigma = \sqrt{\left(\Sigma X^2 P(X)\right) - \mu^2} \end{cases}$$

donde la letra μ representa la media la distribución. Este valor es el valor medio que podemos esperar, si realizaramos este experimento un nº infinito de veces (por eso a veces se le denomina valor esperado o esperanza), pero no obtendremos así el valor que se espera que se repita más a menudo. De hecho el valor obtenido no tiene porqué tener sentido (por ejemplo, para el ejemplo inicial obtenemos 1,8 vuelos en su hora que no es un valor posible).Tanto la media como la desviación típica, tienen una interpretación análoga a la dada ya en el capítulo....... Puedes realizar su cálculo haciendo uso de la calculadora de forma similar a la explicada en dicho capítulo escribiendo probabilidades en lugar de frecuencias.

Ejemplos:

- *En el ejemplo de "Binter", es m = 1,8 y s = 0,8. Si los datos tienen forma normal, podremos asegurar que por ejemplo, en el intervalo* $(\mu-2\sigma,\mu+2\sigma)=(0,2\ ,\ 3,4)$ *que a efectos prácticos es el (1,3) , está el 95% de los datos. Esto significará que el 95% de los días habrá entre 1 y 3 vuelos no retrasados. Comparamos ahora con el resultado de la tabla, y vemos que la probabildad de esto es:*

  P(1 o 2 o 3) = P(1)+P(2)+P(3)=0,936 es decir un 93,6%, que es muy similar a lo predicho.

- *Análisis de juegos: Supongamos que por 100 pesetas tomamos parte en una rifa en la que hay 100 papeletas, y sólo una tiene premio de 1000 pesetas. Si llamamos X a las posibles ganancias obtenidas, tendremos la distribución de probabilidad:*

X=Posibles ganancias	-100	1.000
P(X)	99/100	1/100

La media de ganancias será m = -89, es decir, si repitieramos el juego un nº infinito de veces, perderíamos como media 89 pesteas por jugada. El juego es favorable al organizador. Esta es una forma de medir hasta qué punto es juesto el precio de un juego. Cuando la media de ganancias es mayor que cero, el juego sería favorable al jugador. En la práctica todos los juegos tienen una esperanza inferior a cero. (¿porqué habría de organizarse un juego sinó?)

- *Tablas de decisión: Un empresario tiene un taller de serigrafía en el que se estampan camisetas publicitarias. En éste momento opta a un contrato pero una de las máquinas necesarias se encuentra estropeada. Puede optar por repararla o cambiarla por una nueva más eficiente. Tras realizar un estudio de los posibles beneficios y pérdidas, la situación es la siguiente:*

	Obtiene el contrato	No obtiene el contrato
A: Repara la maquinaria	400.000 pts	-10.000 pts
B: Compra maquinaria nueva	700.000 pts	-80.000 pts

El conocimiento que del negocio tiene el empresario le permite asignar (de forma subjetiva) una probabilidad del 20% al suceso "obtener el contrato", con lo que si designamos X="posibles ganancias", tendremos para los opciones A y B:

X_A=Posibles ganancias	400.000	-10.000
P(X)	20/100	80/100

X_B=Posibles ganancias	700.000	-80.000
P(X)	20/100	80/100

Calculando las ganacias esperadas, tendremos: m_A*=72.000 ,* m_B*=76.000 . Dado que la opción A tiene una esperanza de ganancias menor que la B, el empresario elegirá la opción B.*

- *Toma de decisiones: La esperanza de vida de un paciente de cáncer, es de 6 años. Si se opera y la operación tiene éxito, su esperanza de vida es de 15 años, pero hay una probabilidad de 1/3 de quedarse igual que antes de la operación, y por otro lado una probabiliad p de morir en ella. ¿Qué valores debe tomar p para que resulte deseable realizar la operación?*

  Llamamos X, al nº de años que puede vivir elpaciente. X puede tomamr los valores 0 (si muere en la operación), 6 (si no resulta efectiva), o 15 (si todo sale bien). La probabilidad de X=0 es p, la de X=6 es 1/3 y la de X=15 será 1-(1/3+p). Para que la operación sea deseable,la media de la variable X, ha de ser como mínimo mayor que 6 que es la esperanza si no se realiza. Se tiene que:

  $\mu = 0\times p + 6\times 1/3 + 15\times(1-(1/3+p)) = 12 - 15p \geq 6$ *para lo cual ha de ser* $p \leq \frac{6}{15} = 0,4$

Nos concentraremos ahora en la distribución de variable discreta más útil, la distribución binomial:

Distribución binomial

Existen muchas situaciones en las que se presenta una experiencia binomial. Este tipo de experiencias se caracteriza por:

- Estar formada por un nº predeterminado n de experimentos iguales
- Cada uno de los experimentos es independiente de los restantes (la probabilidad del resultado de un experimento no dpende del resultado del resto)
- El resultado de cada experimento ha de admitir sólo dos categorías (a las que denominaremos éxito y fracaso)
- Las probabilidades de ambas posibilidades ha de ser constante en todos los experimentos. (p y q respectivamente)
- Designamos por X la variable que mide el nº de éxitos que se han producido en los n experimentos.

Cuando estemos en estas condiciones, diremos que la variable X sigue una distribución binomial de probabilidad, y la designaremos por B(n,p), donde n es el nº de pruebas de que consta el experimento y p la probabilidad de éxito-.

Ejemplos:

- *Si se investiga el nº de piezas defectuosas que existe en un lote de 6 , se está llevando a cabo una experiencia binomial de 6 experimentos (cada experimento es la observación de una pieza), cada uno de los cuales es independiente de los restantes, con dos posibilidades (éxito="ser defectuosa" fracaso="no ser def.") cuyas probabilidades (que supondremos p=0,1 , q=0,9) son constantes en todos los experimentos. La variable X="nº de piezas defectuosas" tiene una distribución B(6, 0,1)*
- *Cuando se estudia el sexo de los descendientes en las familias con diez hijos, se realiza un experiencia binomial de 10 experimentos. Si X="nº de varones", la variable X sigue la B(10, 0,5)*
- *Al contestar al azar un exámen tipo test, formado por 8 preguntas con 5 posibles repuestas cada una, si estudiamos el nº de aciertos tendremos una experiencia binomial de 8 experimentos independientes, con dos posibilidades (acierto o error) con probabilidades respectivas 1/5 y 4/5, que son constantes en cada experimento. Si X mide el nº de aciertos, seguirá la distribución B(8, 1/5)*

Si el nº de experimentos es n, los posibles valores de X son siempre 0,1,2,....,n. y podemos calcular la probabilidad de que obtengamos un nº k de éxitos usando la tabla .... Para ello, localizamos en la primera columna el valor de n, y dentro de este valor el de k. Sobre la horizontal localizamos el valor de p, y uniendo la vertical con la horizaontal, obtenemos la probabilidad buscada. Como puedes ver, la tabla sólo admite valores entre 0 y 0,5 para p. Cuando la probabilidad de p supere este último valor, habremos de planterar el problema a la inversa, es decir, llamando éxito al suceso contrario.

Ejemplos:

- *Si por ejemplo queremos calcular la probabilidad de acertar al azar 6 preguntas del exámen tipo test, habremos de denominar éxito al acierto, y para calcular la probabilidad hemos de ir a la tabla ..... con n=8 (nº de experimentos), P=1/5 (probabilidad de éxito), X=6 (nº de éxitos)., que en en este caso da como resultado P(6)=0,0011, es decir, que sólo en un 1,1% de los casos conseguiremos acertar 6 preguntas si respondemos al azar.*
- *Supongamos que un avión dispone de 3 turboreactores independientes, y que ha sido comprobado que cada uno de ellos tiene una probabilidad de fallar durante un vuelo regular de 0,05. Si consideramos X la variable que mide el nº de turboreactores que fallen durante un vuelo, utilizando la tabla para n=3, p=0,05, y X=0,1,2,3, obtendremos la distribución de probabilidad correspondiente a la variable X:*

X=Nº de turboreactores averiados	*0*	*1*	*2*	*3*
P(X)	*0,8574*	*0,1354*	*0,0071*	*0,0001*

De ella podemos deducir muchas probabilidades. Por ejemplo:

Prob. de que fallen las tres turbinas = p(3)=0,0001, que equivale al 0,1%

Probabilidad de haya 1 o 2 averiados = p(1 o 2)= p(1)+p(2) = 0,1425

Prob. de que como máximo haya una turbina estropeada = p(0 o 1) = p(0) + p(1)=0,9928

Prob. de que al menos una no falle = p(0 o 1 o 2)=p(0)+p(1)+p(2)=0,999 o bien un 99,9%. Podríamos también haberlo calculado como p(0ó1ó2) = 1 - p(3)=0,999.

- *Sabemos que la probabilidad de que un turista que ha visitado Gran Canaria, haya quedado satisfecho es 0,9. Si un grupo de 7 turistas abandona hoy la isla, ¿Cuál es la probabilidad de que exactamente 6 de ellos estén satisfechos?. Puesto que la probabilidad de estar satisfecho es superior a 0,5 planteamos como éxito el estar insatisfecho, es decir, calculamos la probabilidad de un insatisfecho y habrá que ir a la tabla con n=7, p=0,1 y X=1, obteniendose: p=0,3720*
- *Una cierta enfermedad tiene una tasa de mortalidad del 10%. Al ensayar un nuevo tratamiento en un grupo de 10 pacientes, 5 de ellos fallecieron. ¿hay evidencia suficiente para indicar que el tratamiento es insadecuado?*

  En este caso, calculamos la probabilidad de 5 fallecimientos para n=10, p=0,1, y obtenemos que la probabilidad de que esto ocurra es manor que un 0,5%. Ante esta situación, hay evidencia de que el tratamiento no es adecuado.

En particular la distribución binomial B(n,p), tiene como media y desviación típica:

$$\begin{cases} \mu = np \\ \sigma = \sqrt{npq} \end{cases}$$

- *Imaginemos que un exámen tipo test consta de 100 preguntas todas ellas con 5 opciones de las que sólo una es correcta. Si una persona contestara al azar, estaría frente a una experiencia binomial de n=100 experimentos, con probabilidad de éxito p=1/5 (consideramos éxito a acertar la respuesta). El nº medio de éxitos y la desviación típica para alqguien que conteste al azar resultará ser:*

  $$\mu = np = 100 \times \frac{1}{5} = 20 \quad \textit{respuestas correctas} \quad \sigma = \sqrt{npq} = \sqrt{100 \times \frac{1}{5} \times \frac{4}{5}} = 4$$

  Usando la regla para distribuciones con forma normal, un 95% de las personas que han respondido al azar, habrán contestado bien entre 12 y 28 respuestas, y puesto que un 99,8% de las repuestas de estos últimos se encuentran entre 8 y 32 respuestas acertadas, la probabilidad de que alguien que no estudió saque más de 32 respuestas acertadas es realmente muy pequeña.

Ajuste a una binomial

En algunas situaciones, se dispone de un conjunto de datos de los que se sospecha que puedan estar distribuidos binomialmente, y se quiere confirmar esta intuición.

Por ejemplo, supongamos que un importador de juguetes recibe mensualmente 100 cajas de un determinada modelo, y cada una de las cajas contiene 4 juguetes. Durante el primer mes, para poder calcular los costes reales de los juguetes, lleva a cabo un estudio en el que obtiene:

Nº de juguetes defectuosos por caja	*0*	*1*	*2*	*3*	*4*
Nº de cajas	*65*	*30*	*4*	*1*	*0*

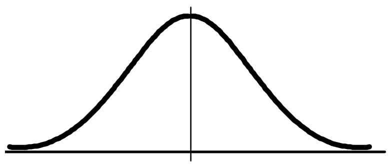
El empresario calcula en primer lugar el nº medio de juguetes defectuosos por caja, obteniendo media=0,41. Si la distribución de juguetes defectuosos por caja fuera una B(n=4,p), la media sería m=np=0,41 , luego despejando, p=0,102. Para ver si se ajusta o no a la B(4 , 0,1), calculamos para esta última la probabilidad de 0,1,2,3,4 "éxitos" (juguetes defectuosos), y obtenemos:

X="éxitos"	*0*	*1*	*2*	*3*	*4*
p(X)	*0,656*	*0,292*	*0,049*	*0,004*	*0*

y por tanto, en 100 cajas, las "teóricas" apariciones de los valores 0,1,2,3,4 , serían: 65,6 , 29,2 , 4,9 , 0,4 y 0, que son muy similares a las encontradas en la muestra.

A partir de aquí, el empresario puede asumir que la probabilidad de que un juguete sea defectuoso es del 10%, podrá hacer previsiones para las 5000 cajas de que va a constar su próximo pedido.

Distribución Normal

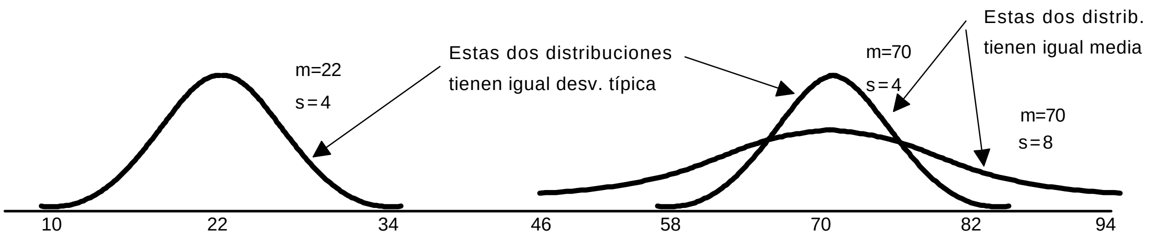


Abordaremos ahora el estudio de la más importante de las distribuciones de tipo contínuo, la distribución normal, cuya función de probabilidad tiene una representación gráfica (denominada para variables contínuas, curva de densidad) que se caracteriza por la forma de campana invertida (Fig. 1). Hemos hablado con anterioridad de lo importantes que son y de la frecuencia con que aparecen en estadística este tipo de distribuciones. Corresponde a fenómenos en los que existen unos pocos datos en los extremos, y estos aumentan paulatinamente hasta la parte central donde está la mayoría de ellos. Se presenta en fenómenos tan dispares, como la longitud de una pieza fabricada por una máquina o el efecto que una misma dosis de un fármaco produce en diferentes individuos. Son de este tipo, la mayoría de los caracteres morfológicos de las poblaciones (tallas, pesos,...), sociológicos (consumo de productos, valoración de un mismo fenómeno,...), psicológicos (coeficiente de inteligencia, grado de adaptación al medio,...) físicos (resistencia a la rotura de una pieza, duración de una pila,.....) y en general todas aquellas características que se obtengan como suma de muchos factores.

Ante este tipo de fenómenos, existe un modelo matemático que nos permite el tratamiento de todos ellos, la curva o campana de Gauss. Veremos ahora las características más importantes de este modelo, y como usarlo para calcular probabilidades.

Diremos que una variable aleatoria tiene una **distribución normal** si su curva de densidad es simétrica , con forma de campana invertida (fig. 1).

Cada distribución normal tiene dos parámetros que son los que la determinan: su media m y su desviación típica s . Por ello, se suele denotar por N(μ,σ). En la figura 2, vemos diferentes distribuciones normales:

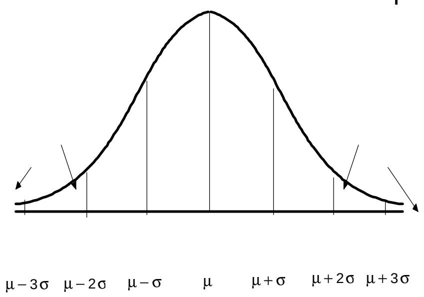


Como puedes observar, la media indica el eje de simetría de la distribución, mientras la desviación típica es la que determina el recorrido de la misma.

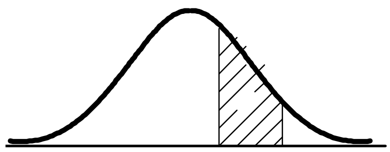
Ya dijimos que para distribuciones de tipo discreto, la suma de todos los valores de la probabilidad debía ser 1. Para el caso de las distribuciones de tipo contínuo esta condición se transforma en que el área total bajo la curva ha de ser 1. La clave de este tipo de distribuciones está en que existe una correspondencia entre área y probabilidad, de forma que la probabilidad de que la variable esté entre dos valores a y b es exactamente el área marcada en la figura 3. Hemos de aclarar que puesto que en una distribución contínua existen infinitos valores la probabilidad de cada uno de ellos ha de ser necesariamente nula o lo que es lo mismo, el área bajo un determinado punto es nula. (veremos más adelante como solventar este problema),

La distribución normal estandar

Nosotros haremos uso siempre de la llamada normal estandar N(0,1), que tiene media 0 y desviación típica 1 y al que denotaremos con la letra Z. Esta es la distribución representada en la figura lateral. Como puedes ver, si tratamos con la N(0,1), los intervalos centrados en la media y de radios iguales a una d.t., dos d.t.,.... son los intervalos (-1,1) , (-2,2), (-3,3). Podrás observar que lo deducido del teorema de Chebyshev, nos indica en porcentaje la cantidad de probabilidad que existe entre los valores representados.



Pero no sólo disponemos de estos valores, sino que además disponemos para esta distribución de una tabla, confeccionada a partir del cálculo del área que existe desde el extremo inferior hasta cualquier valor positivo de la variable Z. Con esta tabla y utilizando las técnicas que describiremos más adelante, estaremos en condiciones de calcular probabilidades. En particular veremos como calcular las probabilidades:



$P(a<Z<b)$ = probabilidad de que la variable Z se encuentre entre los valores a y b (o lo que es lo mismo, probabilidad de que un valor de Z tenga puntuación típica comprendida entre a y b)

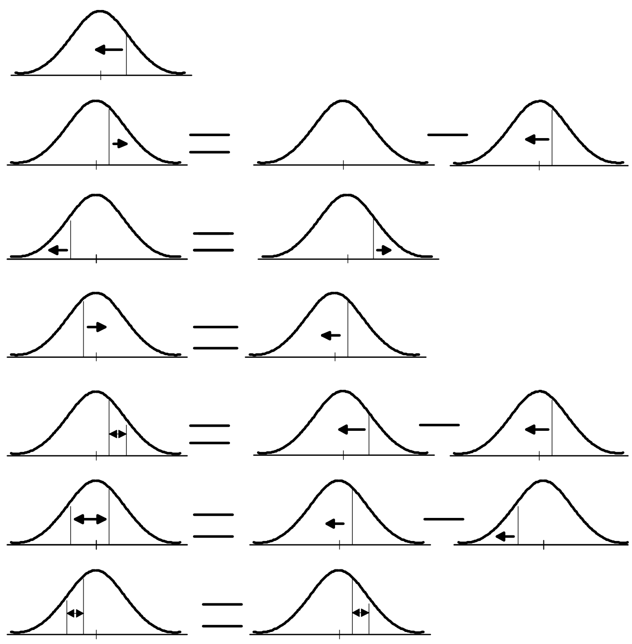
$P(Z>a)$ = probabilidad de que la variable Z sea mayor que el valor a (probabilidad de que un valor tenga p.t. mayor que a)

$P(Z<a)$=probabilidad de que la variable Z sea menor que el valor a. (probabilidad de que un valor tenga p.t. menor que a). Estas son precisamente las probabilidades que se pueden calcular en la tabla, para a positivo.

Por lo explicado anteriormente sobre la probabilidad de un valor concreto, no tendrá para nosotros sentido utilizar $\leq o \geq$ en las expresiones anteriores.

Uso de la tabla de valores

Como ya hemos comentado, disponemos de una tabla de la N(0,1) , en la que aparecen las probabilidades de la forma p(Z<a), siendo a positivo. Haciendo uso de la simetría de la distribución normal, y de que el área total es 1, podemos calcular todos los casos que se presenten. Ahora mediante ejemplos verás como solucionar cada caso:



a) p(Z<1,23) = 0,8907

b) p(Z>0,84)=1-p(Z<0,84)=1-0,7996=0,2004

c) p(Z<-1,23)=p(Z>1,23)=1-p(Z<1,23)=1-0,8907=0,1193

d) p(Z>-0,84)=p(Z<0,84)=0,7996

e) p(0,84<Z<1,23)=p(Z<1,23)-p(z<0,84)=0,0911

f) p(-1,23<Z<0,84) = p(Z<0,84)-p(Z<-1,23) = p(Z<0,84)-[1-p(Z<1,23) =0,6803

g) p(-1,23<Z<-0,84)=p(0,84<Z<1,23)=0,0911

Tipificación de una variable

Aunque la mayoría de las distribuciones no tiene media 0 y desviación típica 1, existe una forma de transformarlas para conseguirlo. Este proceso se denomina tipificación de la variable, y consiste en sustituir la variable original X, por una nueva Z que sigue la distribución estandar, haciendo uso de la expresión:

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

O lo que es lo mismo, transformamos los valores de la variable X por su correspondiente puntuación típica.

Imaginemos que una máquina encargada de rellenar con 10 ml una botella de perfume, sigue una distribución N(10,0´1). Para calcular que la probabilidad de que un fasco de perfume tenga un contenido entre 9,85 y 10 ml, haríamos:

$$p(9,85 < X < 10) = p(\frac{9,85-10}{0,1} < Z < \frac{10-10}{0,1}) = p(-1,5 < Z < 0)$$

De esta forma, sea cual sea la media y d.t. de una distribución normal, podremos hacer el cálculo de probabilidades utilizando la tabla correspondiente a la estandar.

Ejemplos:

- *La duración media de un televisor de una determinada marca es de 8 años con una desviación típica de medio año y se distribuye normalmente. Si quisieramos calcular la proporción de ellos que durarán más de 9 años, tendríamos:*

$$p(X>9) = p(Z > \frac{9-8}{0,5}) = p(Z>2) = 1-p(Z<2) = 1-0,9772 = 0,0228 \sim 22,8\%$$

- *La cosecha de tomates de una empresa agrícola canaria se sabe por experiencia que se distribuye con respecto al diámetro en cm. de los tomates siguiendo la N(4, 0,5). Este año, le informan de que sólo podrá vender en el mercado exterior tomates con diámetro superior a 4,5 cm. Si se prevee una producción total de 50.000 kg., podrá prever cuantos kilos podrá destinar a exportación, calculando* $p(X>4,5)=p(Z>1)=0,1587 \sim 15,87\%$ *, es decir 0,1587x5000=7935 kg.*
- *Si el tiempo (en meses) X que tarda en aparecer el primer diente en un bebé sigue la N(7,5 , 2), podremos calcular:*

  p(6<X<9)=p(-0,75<Z<0,75)=p(Z<0,75)-[1-p(Z<0,75)]=0,5468

  es decir el 54,68% de los niños echa los dientes entre los 6 y los 9 meses
- *Supongamos con los datos del problema anterior, que se hiciera la siguiente clasificación:*

  Dentición prematura :el 10% que menos tarda en echar los dientes

  Dentición tardía el 20% que más tarda en echar los dientes

  Dentición normal : el resto

  Para saber cuales son los valores correspondientes a cada categoría, deberemos calcular:

  Dentición prematura: el valor de k de forma que p(Z<k)=0,1. Un niño tendrá d. prematura si tarda entre 0 y k meses en echar los dientes.

  D. tardía: el valor de k de forma que p(Z>k)=0,2. Se tendrá dentición tardía cuando se tarden más de k meses en echar los dientes.

  D. normal será la que se eche entre los valores calculados anteriormente.
- *En un exámen de matemáticas las notas se distribuyen según N(6 , 2,5). Si el profesor decide poner sobresaliente al 20% con mayor nota, para saber a partir de que nota ha de poner sobresaliente, ha de calcular para que valor k es p(X>k)=0,2, es decir:* $p\left(Z > \frac{k-6}{2,5}\right)=0,20 \Rightarrow 1-p\left(z < \frac{k-6}{2,5}\right)=0,20 \Rightarrow p\left(z < \frac{k-6}{2,5}\right)=0,8 \Rightarrow \frac{k-6}{2,5} \cong 0,84 \Rightarrow k \cong 8,1$

Ajuste a una normal

Para averiguar si unos datos se ajustan a una distribución normal, se calculan la media y desviación típica de dichos datos. Si con la normal de igual media y desviación típica, las frecuencias teóricas de los valores se ajustan a las reales, podremos concluir que dichos datos siguen una distribución normal.

Imaginemos por ejemplo que tras analizar el peso de una muestra de 200 paquetes de azúcar, hemos obtenido los siguientes datos:

Pesos (en kg)	*Frecuencia*
2,30 - 2,35	*4*
2,35 -2,40	*16*
2,40- 2,45	*34*
2,45- 2,50	*45*
2,50 - 2,55	*49*
2,55-2,60	*38*
2,60-2,65	*14*

Calculamos la media y desv. típica de esta muestra y obtenemos media=2,497, d.t.=0,0725. Comparando las frecuencias reales con las teóricas para N(2,497 , 0,0725) , tendremos:

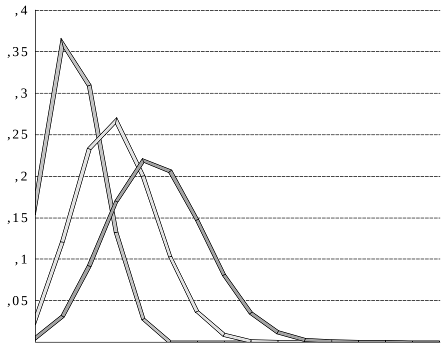
Pesos (en kg)	*Frecuencia*	*Probabilidades*	*Frecuencias teóricas*
2,30 - 2,35	*4*	*p(2,3<X<2,35)=0,0183*	*3,66*
2,35 -2,40	*16*	*p(2,35<X<2,4)=0,0701*	*14,02*
2,40- 2,45	*34*	*p(2,4<X<2,45)=0,1693*	*33,86*
2,45- 2,50	*45*	*p(2,45<X<2,5)=0,2549*	*50,98*
2,50 - 2,55	*49*	*p(2,5<X<2,55)=0,2513*	*50,26*
2,55-2,60	*38*	*p(2,55<X<2,6)=0,1549*	*30,98*
2,60-2,65	*14*	*p(2,6<X<2,65)=0,0604*	*12,08*

y podemos considerar aceptables las diferencias.

Aproximación de la binomial mediante la normal

Recordarás que en un epígrafe anterior hemos hablado de la distribución binomial, y de como hacer uso de la tabla de valores correspondientes. Sin embargo, recuerda que dicha tabla está confeccionada para pequeños valores de n. ¿qué ocurriría si tuvieramos que hacer el cálculo para valores de n superiores? Veremos ahora que bajo determinadas condiciones, la distribución binomial es muy semejante a la normal, y que podremos hallar probabilidades utilizando ésta última.

Para justificar lo dicho anteriormente, observa lo que ocurre con la distribución binomial con p=0,3, cuando los valores de n van creciendo. Como puedes ver, la distribución binomial tiende a tener la forma de la normal, para valores crecientes de n.



<u>En la práctica, cuando np y nq son ambos mayores que 3, la aproximación es bastante buena, y si superan a 5, es prácticamente perfecta En tales casos, $B(n,p) \sim N(np, \sqrt{npq})$ es decir, la binomial de n expermientos con probabilidad de éxito p es equivalente a una normal que tenga igual media y desviación típica que ella</u>.

Veremos un ejemplo que ilustre como utilizar esta propiedad para calcular probabilidades. Imaginemos que en Canarias sabemos que el porcentaje de fracaso escolar en BUP es del 40%, y que sobre una población de 1000 estudiantes, queremos calcular la probabilidad de que no se superen los 380 fracasos. Sabemos que X="nº de fracasos " sigue una B(1000, 0,4). Para calcular lo pedido habríamos de tener una tabla que contuviera el valor n=1000 (te imaginas que tabla !!), además habría que calcular $p(X \leq 380)$ que es la suma de las 381 probabilidades p(0),p(1),...,p(379),p(380). Para utilizar una distribución normal que se asemeje mucho a la B(1000, 0,4), primero hemos de comprobar que np y nq superan ambas el 5.

np=1000x0,4=400>5 ; nq=1000x0,6=600>5.

En consecuencia, la B(1000, 0,4) es equivalente a la N(400, 15,49). Tipificando esta variable normal tendremos:

$$p(X \leq 380) = p(Z \leq \tfrac{380-400}{15,49}) = p(Z \leq -1,25) = 0,1056 \sim 10,56\%$$

y por tanto sólo el 10,46% de los centros de 1000 estudiantes tendrán teóricamente menos de 380 fracasos.

Si quisieramos calcular la probabilidad de que fuera 400 el número de fracasos, deberíamos realizar una corrección para que puediera hacerse el cálculo, que consistiría en calcular la probabilidad de un intervalo de centro en 400 y radio media unidad , es decir: $p(X=400) = p(399,5 < X < 400,5) = p(-0,03 < Z < 0,03) = 0,024 \sim 2,4\%$

Ejemplo:

- *La probabilidad de que una copa de cristal se rompa durante su transporte es del 1%. Si se hace un transporte de 1000 copas, podremos calcular el nº esperado de roturas como m=np=0,01x1000=10, con desv. típica: s=3,146. Puesto que np y nq superan el 5, podremos utilizar la N(10, 3,146) para calcular la probabilidad de que haya más de 15 roturas* $p(X > 15) = p(Z > 1,59) = 0,0559 \sim 5,6\%$

Ejercicios

1.- En 1992 el 20% de las películas de cine proyectadas en España eran de nacionalidad española. Si tomaramos al azar una muestra de 5 películas proyectadas en alguna sala, calcula la probabilidad de que:

a) haya 1 película española

b) haya 3 o más películas españolas.

c) haya 2 películas extranjeras.

2.- Un proveedor de bolígrafos publicitarios afirma que el 95% de ellos no tiene ningún defecto. Tú seleccionas 10 al azar y ves que sólo 4 de ellos funcionan bien. Calcula la probabilidad de que hayan 4 o menos bolígrafos que funcionen suponiendo cierto lo que dice el proveedor. Basandote en este resultado dí que opinas de la afirmación del proveedor.

3.- La opinión que tiene la población española sobre la terapia de grupo es favorable en un 40%, y desfavorable en el resto. Se elígen 8 personas al azar. Hallar:

a) Probabilidad de que sólo 3 tengan opinión favorable.

b) Prob. de que más de 3 tengan opinión favorable.

4.- Según el I.N.E. la población activa española en 1992 era de 15.201.000 personas, de los que 2.789.000 estaban parados. Se escogen 10 personas al azar de la población activa. Hallar:

a) Probabilidad de que 5 de ellos estén parados.

b) Probabilidad de que todos trabajen

c) Cuál es el nº medio de parados en muestras de 10 personas.

5.- El director de un hotel del sur grancanario está estudiando los porcentajes de ocupación durante el invierno, para planificar futuras temporadas. El hotel tiene 100 habitaciones, y la tasa de ocupación durante el invierno es del 80%. Halla la media y desviación típica de ocupación durante el invierno. ¿Sería raro que en dicha temporada hubiera sólo 70 habitaciones ocupadas durante el invierno?. (Usa el intervalo $\bar{x}-3\sigma$ *y* $\bar{x}+3\sigma$ *como límites para los valores raros)*

6.- Se ha hecho un seguimiento a 200 personas elegidas al azar y se ha anotado el nº de veces que contraen la gripe al cabo de un año. Los resultadios obtenidos son:

Nº de resfriados	0	1	2	3	4
Nº de personas	18	97	80	3	2

Ajusta estos datos a una distribución binomial. Compara los datos reales con los teóricos.

7.- Un tipo de piezas requiere de 3 soldaduras. Se ha hecho un control de calidad a 600 de esas piezas y se ha obtenido:

Nº de soldaduras defectuosas	0	1	2	3
Nº de piezas	311	227	59	3

¿Se ajustan estos datos a una distribución binomial?

8.- Una fábrica de conservas vegetales ha comprado a un agricultor toda su cosecha de guisantes. Los diámetros de los guisantes comprados se distribuyen según una normal de media 1 cm. y d.t. 0,1 cm. Para enlatarlos, se clasifican según su diámetro en dos tipos. Clase A: más de 1,1 cm de diámetro ; Clase B: entre 0,85 cm y 1,1 cm, y el resto de los guisantes se rechazan. Calcula:

a) Porcentaje de guisantes de cada tipo

b) Qué porcentaje de los guisantes es rechazado.

9.- Un banco tiene cuatro ventanillas para atención alos clientes. Tras estudiar el tiempo de espera de estos, se sabe que el tiempo medio de espera es de 3,8 min. con una d.t. de 1,5 min. y que se distribuyen normalmente. Por experiencia se sabe que los clientes se irritan si el tiempo excede de 5 min.

a) ¿Qué porcentaje de los clientes estarán irritados con el actual sistema?

b) Añadiendo una ventanilla se reduciría el tiempo medio de espera a 3 min. con d.t. 1,5 min. También sabemos que se pierde tiempo porque todas las ventanillas comparten el mismo monitor. Si se pusiera un monitor por ventanilla se reduciría el tiempo medio a 3,7 min con una desviación típica de 0,8 min. Calcula el porcentaje de insatisfechos con cada uno de los sistemas, y decide cuál de las dos soluciones es más adecuada.

10.- Las calificaciones en una oposición en el sector bancario se deistribuyen según N(6´3 , 1´8). En esta oposición hay 1000 aspirantes y tan sólo 300 plazas. ¿Qué puntuación mínima hará falta para obtener plaza?.

11.- Un fabricante de ruedas para cohes sabe que la duración en km. de sus ruedas se distribuye normalmente con media 56.000 km. y desviación típica 8.000 km. El fabricante quiere dar una garantía de forma que sólo el 3% de ellas sea reemplazada por fallo antes del nº garantizado de km. ¿Por cuántos kilómetros deberán ser garantizadas las ruedas?

12.- Una empresa realiza una oposición para cubrir vacantes, que consta de dos pruebas, la 1ª de mecanografía, y la 2ª una entrevista personal. Si sabe que los resultados de la 1ª *prueba siguen la N(6,3) y la segunda la N(7,2) y que ambas pruebas son independientes, calcula:*

a) Probabilidad de que un opositor obtenga más de 9 en la 1ª prueba

b) Probabilidad de que obtenga más de 8 en la 2ª prueba.

c) La empresa sólo aceptará a aquellos que obtengan más de 9 en la 1ª y más de 8 en la 2ª. ¿Qué porcentaje de los opositores obtendrá la plaza?

13.-En una ganadería se sabe que el peso de los toros se distribuye según la normal de media 420 kg. y d.t. 43 kg. La ganadería tiene 2300 toros. Calcula:

a) ¿Cuántos pesarán más de 450 kg.

b) ¿Cuántos pesarán menos de 400 kg?

c) ¿Cuántos pesarán entre 410 y 430 kg?

14.- Se ha pasado un test a 60 individuos obteniendose las siguientes puntuaciones:

10	10	11	12	12	14	14	14	15	15	15	16	16	16	16	16	17	17	17	17
17	17	19	19	19	19	19	19	19	19	19	20	20	20	20	20	20	20	20	20
21	21	21	21	21	22	22	22	22	24	24	24	25	25	26	26	28	28	29	30

Estudiar si estos datos siguen una distribución normal.

15.- En una estación marítima de transbordadores se sabe que el 65% de los vehículos que se trasladan son turismos, y el resto motocicletas. Si este verano han pasado 1500 vehículos, calcula la probabilidad de que:

a) el nº de turismos esté entre 1100 y 1300

b) sea menor que 1375

16.- Según un estudio del ministerio de Hacienda, de las declaraciones que son sujetas a revisión un 70% acaba teniendo que pagar algún tipo de recargo. Si este año la delegación de Las Palmas a mandado a revisar 3000 declaraciones, ¿cuál es la probabilidad de que al menos 2000 de ellas tengan recargo?