3

ALGUNAS DISTRIBUCIONES DE PROBABILIDAD

3.1 INTRODUCCIÓN

La estadística inferencial tiene como problema general establecer las propiedades de un fenómeno aleatorio estudiando una parte del mismo. Para esto es necesario conocer la distribución de probabilidad de la variable aleatoria que se está estudiando. Esto puede ser complicado si no existe otra alternativa que deducir teóricamente la función de probabilidad. Afortunadamente, existen numerosos modelos de probabilidad muchos de los cuales, aunque hayan sido generados con otros fines, pueden ser usados para describir el comportamiento de la mayoría de las variables aleatorias que son estudiadas en las ciencias naturales.
Los modelos de distribuciones de probabilidad se clasifican de acuerdo a la naturaleza de la variable aleatoria en modelos probabilísticos discretos y en modelos probabilísticos continuos. En éste capítulo se estudiarán algunas distribuciones de probabilidad importantes para la aplicación de los métodos de inferencia estadística, como son la distribuciones Binomial, de Poisson y Normal. Otras distribuciones valiosas como las de T, de F y χ^2 se estudiaran en el momento que se requiera su uso. En este punto es necesario enfatizar que el esfuerzo que se haga en entender las propiedades de estos modelos permitirá, por una parte, comprender mejor el funcionamiento de los métodos de inferencia estadística y por otro lado, contar con más y mejores criterios para elegir el modelo apropiado en la aplicación de algún método estadístico.

3.2 MODELOS PROBABILÍSTICOS DISCRETOS

3.2.1. Modelo de Bernoulli.

Una gran cantidad de situaciones que se presentan en distintos campos de acción, tienen en común algunas cosas. Veamos algunos ejemplos:

a) Lanzar una moneda y determinar si sale cara en cada lanzamiento
b) Lanzar un dado y verificar cada vez si sale un número par.
c) Elegir aleatoriamente un individuo y determinar su sexo.
d) Determinar en una solución si el pH es básico.
e) Determinar si un individuo elegido al azar está enfermo.
f) Verificar para una sustancia su naturaleza orgánica.
g) Determinar si un elemento es metálico.
h) Verificar si un objeto o cosa es defectuoso.

Todos estos experimentos y otros similares reciben el nombre genérico de *Ensayos de Bernoulli*, por tener las características siguientes:

1) Cada vez que se repite el experimento se producen dos resultados mutuamente excluyentes: sale cara o sello, número par o impar, individuos machos o hembras, Ph básico o ácido, individuos enfermos o sanos, sustancias orgánicas o inorgánicas, metales o no metales, objetos buenos o defectuosos. Estos resultados de dos clases se identifican de forma general como éxito y fracaso.

2) Cada vez que se repite el experimento la probabilidad de ocurrencia del éxito p o del fracaso q no cambian. En los ejemplos citados se cumple esta condición: la probabilidad de encontrar una cara es la misma cada vez que se lanza la moneda, la probabilidad de obtener un número par no cambia en sucesivos lanzamientos del dado. Del experimento *c* en adelante, si el muestreo es con reposición y las condiciones del experimento no se alteran, la probabilidad de ocurrencia de los eventos estudiados se mantiene constante.

3) Los resultados son independientes. El hecho de que ocurra un fracaso o un éxito, no afecta la probabilidad de ocurrencia de un nuevo resultado al repetir el experimento.

El espacio muestral general para los distintos ensayos de Bernoulli está formado por dos resultados: el éxito (E) y el fracaso (F).

$$S = \{F; E\}$$

Conocido el espacio muestral, es posible definir una variable aleatoria que le asigne valores a estos dos resultados. Por ejemplo la variable $X = \textit{número de éxitos en un ensayo}$, genera el rango espacial siguiente:

$$Rx = \{0,1\}$$

Igualmente es posible encontrar una función que le asigne valores de probabilidad a éstos resultados. Para tal fin la probabilidad del éxito p y la probabilidad del fracaso $1-p$, se pueden representar de la forma siguiente:

$$P(X = 0) = p^0(1-p)^1$$
$$P(X = 1) = p^1(1-p)^0$$

De lo cual se deduce que la expresión matemática de la función de probabilidad para la distribución de Bernoulli es la siguiente

$$p_{(x)} = p^x(1-p)^{1-x}$$

El valor esperado y la varianza de esta distribución son: $E(x) = \mu = p$ y $Var(x) = \sigma^2 = pq$ respectivamente.

3.2.2. Modelo binomial.
Un experimento binomial consta de varios ensayos de Bernoulli, algunos ejemplos son los siguientes:

a) Lanzar una moneda n veces y determinar si sale cara en cada lanzamiento
b) Lanzar un dado n veces y verificar cada vez si sale un número par.
c) Determinar el sexo de cada uno de n individuos.
d) Medir en cada una de n soluciones si el pH es básico.
e) Determinar en cada uno de n individuos si están sanos o enfermos.
f) Verificar para cada una de n sustancias su naturaleza orgánica.
g) Determinar si cada uno de n elementos son metales.
h) Verificar si cada uno de n objetos son defectuosos.

El examen de los ejemplos anteriores deja ver que en cada repetición del experimento se mantienen las propiedades de los ensayos de Bernoulli y que la variable aleatoria que los caracteriza es el número de veces que ocurre el éxito (o el fracaso) en n repeticiones del experimento. La función de probabilidad para este tipo de variable, la vamos a deducir a través del ejemplo siguiente:

Ejemplo 3.1.
En una investigación de cierta parasitosis, pequeñas dosis de una vacuna experimental se inyectaron en ratones de laboratorio. Los resultados encontrados demostraron que 4 de cada 20 ratones mueren a causa de la vacuna. Si la misma dosis de la vacuna se aplica a cuatro ratones ¿cuál es la probabilidad de que mueran x individuos?

1) En primer lugar se verifica que se trata de un ensayo de Bernoulli.

   - Tiene dos resultados posibles: ratón muere (éxito); ratón sobrevive (fracaso). El significado del éxito en éste caso tiene una connotación simplemente estadística, porque para el objetivo de la investigación y de la ética científica no puede ser un éxito que muera un ratón.

   - La probabilidad del éxito $p = 1/5$ y del fracaso $q = 4/5$, no cambian en cada aplicación de la dosis.

   - El experimento se repitió cinco veces ($n = 4$).

2) Puesto que el número de repeticiones ($n=4$), es pequeño se puede determinar fácilmente los 16 resultados que conforman el espacio muestral. Si se representa con m el evento morir y con s el evento sobrevivir, los 16 resultados posible son los siguientes:

$$S = \begin{Bmatrix} ssss \\ sssm, ssms, smss, msss \\ ssmm, smsm, mssm, smms, msms, mmss \\ smmm, msmm, mmsm, mmms \end{Bmatrix}$$

La variable aleatoria X = *número de ratones muertos* genera el rango espacial siguiente:

$$Rx = \{0,\ 1,\ 2,\ 3,\ 4\}$$

3) Si $p =$ *probabilidad de morir* y $q =$ *probabilidad de sobrevivir*; la probabilidad con la cual ocurrirán los resultados del espacio muestral S, son las siguientes:

$P(X=0) = p(ssss)= qqqq = 1p^0q^4$

$P(X=1)=p(mmms)+p(mmsm)+p(msmm)+p(smmm)=4pqqq=4p^1q^3$

$P(X=2)= p(ssmm)+p(smsm)+p(smms)+p(msms)+p(mmss)+p(mssm)= 6ppqq = 6p^2q^2$

$P(X=3)= p(smmm)+p(msmm)+p(mmsm)+(mmms)= 4pppq= 4p^3q^1$

$P(X=4)= p(mmm)= 1pppp= 1p^4q^0$

4) Puesto que se está en busca de una fórmula general para calcular la probabilidad con la cual ocurre cada resultado de la variable aleatoria X = *número de ratones muertos*, una forma de hacerlo, es identificando los elementos comunes en las fórmulas usadas para el cálculo de cada probabilidad de X.
El primer hecho común que resalta es que los valores de p están elevados a una potencia que coincide con el valor x de la variable aleatoria. Igualmente, los valores de q están elevados a una potencia que es igual a $4-x$. De modo que se tiene como un término común, la expresión:

$$p^x q^{4-x}$$

Este último término se puede hacer mucho más general, si se sustituye el valor 4 por n, como valor genérico que representaría, en el ejemplo específico que se viene trabajando, un número cualquiera de ratones, o si se tratara de cualquier experimento binomial, n sería el número de repeticiones de ese experimento. El término queda entonces de la forma siguiente:

$$p^x q^{n-x}$$

5) Otro elemento común, que se observa en las probabilidades calculadas, es que cada término esta multiplicado por un coeficiente que representa el número de secuencias diferentes como pueden morir x ratones. Obviamente que no sería práctico que cada vez que se necesite calcular la probabilidad de que un evento ocurra x veces en n repeticiones se tengan que determinar las diferentes maneras como puede ocurrir un resultado particular. Si en el ejemplo se hubiesen vacunado 10 ratones en lugar de 4, el número de secuencias para cada valor de x se habría elevado enormemente, y sería muy complicado determinarlas todas. Afortunadamente los denominados métodos de enumeración ofrecen

una solución para calcular el número de secuencias diferentes que se pueden originar a partir de un conjunto de n elementos. El método que se usa en éste caso son las Permutaciones de *n* elementos diferentes, siendo *x* elementos de una clase (ratones que mueren) y *n – x* de otra clase (ratones que sobreviven), cuya fórmula de cálculo es la siguiente:

$$_nP_{x;n-x} = \frac{n!}{x!(n-x)!}$$

En consecuencia, la formulación general para calcular la probabilidad de tener x éxitos en n repeticiones es la siguiente:

$$P(X = x) = p_{(x)} = {}_nP_{(x;n-x)}\,p^x q^{n-x} = \frac{n!}{x!(n.x)!}\,p^x q^{n-x}$$

La fórmula $_nP_{x:n-x}$ con la cual se calcula el número de permutaciones diferentes que se pueden construir con *n* elementos, cuando *x* son de una clase y *n-x* de otra clase, es equivalente a la fórmula $_nC_x$ que se utiliza para calcula el número de combinaciones de *n* elementos diferentes tomados de *x* en *x*.

$$_nP_{(x;n-x)} = \frac{n!}{x!(n-x)!} = {}_nC_x$$

Por lo tanto, la función de probabilidad del modelo binomial puede ser expresado indistintamente en una de las dos formas siguientes:

a) $p_{(x)} = {}_nP_{(x;n-x)}\,p^x q^{n-x} = \dfrac{n!}{x!(n-x)!}\,p^x q^{n-x}$

b) $p_{(x)} = {}_nC_x\;p^x q^{n-x} = \dfrac{n!}{x!(n-x)!}\,p^x q^{n-x}$

Cualquiera de las dos fórmulas anteriores representa la función de probabilidad de la variable aleatoria *X = número de éxitos en n repeticiones de un experimento*. Su aplicación permitirá calcular la probabilidad de que un resultado ocurra *x* veces en n repeticiones. Para el caso específico del ejemplo con el cual se comenzó esta explicación, la aplicación de la función de probabilidad produce los resultados siguientes:

$$P(X = 0) = p_{(0)} = {}_4C_0\;p^0 q^4 = \frac{4!}{0!(4-0)!}(1/5)^0(4/5)^4 = (1)((1)(0{,}4096) = 0{,}4096$$

$$P(X = 1) = p_{(1)} = {}_4C_1\;p^1 q^3 = \frac{4!}{1!(4-1)!}(1/5)^1(4/5)^3 = (4)(0{,}2)(0{,}5120) = 0{,}4096$$

$$P(X=2)= p_{(2)}={}_4C_2\ p^2q^{4-2}=\frac{4!}{2!(4-2)!}\ (1/5)^2(4/5)^2= (6)(0,04)(0,64) = 0,1536$$

$$P(X=3)= p_{(3)}={}_4C_3\ p^3q^{4-3}=\frac{4!}{3!(4-3)!}\ (1/5)^3(4/5)^1= (4)(0,008)(0,8) = 0,0256$$

$$P(X=4)= p_{(4)}={}_4C_4\ p^4q^{4-4}=\frac{4!}{4!(4-4)!}\ (1/5)^4(4/5)^0= (1)(0,0016)(1) = 0,0016$$

3.2.2.1 Distribución de probabilidades del modelo binomial

El conjunto de pares ordenados $[\,x_i\,;\,p_{(\,x\,)}\,]$ genera una distribución binomial, nombre que se le da porque los sucesivos términos de la distribución de probabilidad son semejantes a los obtenidos con la expansión del Binomio de Newton $(\,p+q\,)^n$. Cuando una variable aleatoria se distribuye en forma binomial con parámetros *n* y *p* se puede representar mediante la expresión siguiente: $X : b(\,n;\,p\,)$. La forma de la distribución binomial cambia para cada combinación de valores diferente de *n* y/o *p* (Figura 3.1).

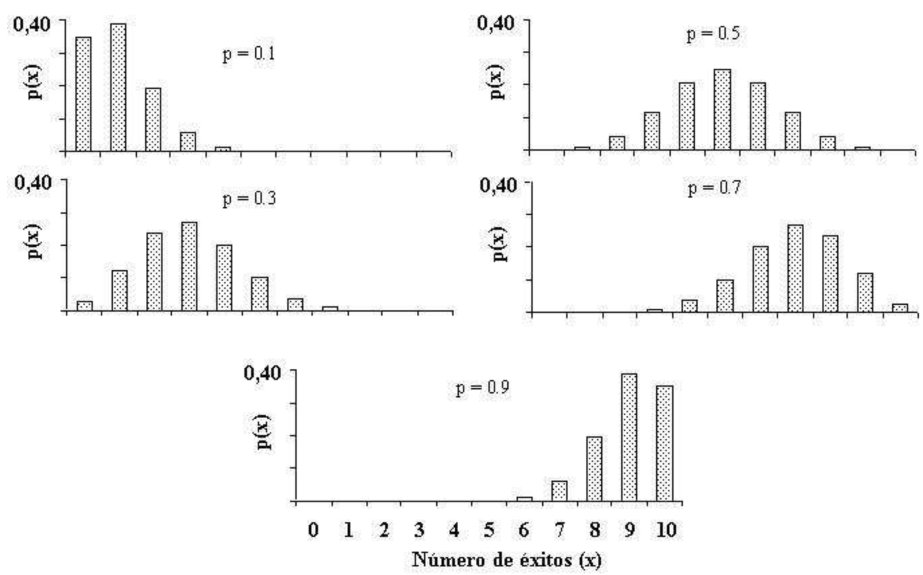


Figura 3.1. Distribuciones de probabilidad para una variable aleatoria que sigue el modelo binomial con *n = 10* y diferentes valores de *p*.

Ejemplo 3.2.

Para una pareja de individuos la probabilidad de tener un hijo con ojos azules es $1/4$ y la de tener hijos con ojos marrones es de $3/4$. Si tienen previsto tener cuatro hijos ¿Cuál será la distribución de probabilidades para la variable *X = número de hijos con ojos azules*?

Se trata de un ensayo binomial. Sólo hay dos posibles resultados, mutuamente excluyentes e independientes. La probabilidad de éxito (ojos azules) $p = 1/4$ y la probabilidad del fracaso $q = 3/4$ no cambiarán en los nacimientos sucesivos. Para obtener la distribución de probabilidades se debe usar la función de probabilidad del modelo binomial para calcular la probabilidad de ocurrencia de cada resultado del rango espacial: *Rx = {0, 1, 2, 3, 4}*,

$$P(X=0) = {}_4C_0\, p^0 q^{4-0} = \frac{4!}{0!(4-0)!}(1/4)^0(3/4)^4 = (1)(1)(0{,}3164) = 0{,}3164$$

$$P(X=1) = {}_4C_1\, p^1 q^{4-1} = \frac{4!}{1!(4-1)!}(1/4)^1(3/4)^3 = (4)(0{,}25)(0{,}4269) = 0{,}4269$$

$$P(X=2) = {}_4C_2\, p^2 q^{4-2} = \frac{4!}{2!(4-2)!}(1/4)^2(3/4)^2 = (6)(0{,}0625)(0{,}5625) = 0{,}2109$$

$$P(X=3) = {}_4C_3\, p^3 q^{4-3} = \frac{4!}{3!(4-3)!}(1/4)^3(3/4)^1 = (4)(0{,}0156)(0{,}75) = 0{,}0469$$

$$P(X=4) = {}_4C_4\, p^4 q^{4-4} = \frac{4!}{4!(4-4)!}(1/4)^4(3/4)^0 = (1)(0{,}0039)(1) = 0{,}0039$$

La distribución de probabilidades se muestra en la tabla siguiente y en la Figura 3.2.

X	$p_{(x)}$
0	0,3164
1	0,4269
2	0,2109
3	0,0469
4	0,0039

Figura 3.2. Distribución de probabilidades del número de hijos con ojos azules.

3.2.2.2 Función de probabilidad acumulada para el modelo binomial

La función acumulada $F_{(x)} = \sum_{i=1}^{n} p_{(x)}$, facilita el cálculo de probabilidades del tipo siguiente:

a) $P(X \leq x)$

b) $P(X \geq x)$

c) $P(x_1 \leq X \leq x_2)$

A fin de facilitar la aplicación de la distribución binomial, en la mayoría de los textos de estadística existen tablas con las probabilidades acumuladas. También muchos programas de utilidades para computadoras personales, como el Excel, proporcionan funciones que permitan calcular las probabilidades individuales o acumuladas. En la Tabla 3.1 se muestra una parte de una tabla de probabilidades acumuladas de la función binomial. Su uso es muy fácil, tiene tres entradas, el valor del parámetro *p* (probabilidad del éxito), el valor de *n* (número de repeticiones) y el valor de *x* (número de éxitos).

Tabla 3.1. Algunas probabilidades acumuladas de la función de probabilidad $p(x) = {}_nC_x\, p^x q^{n-x}$

$$F_{(x)} = P(X \leq x) = \sum^{Rx} {}_nC_x\, p^x q^{n-x}$$

		p										
n	X	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6
15	0	0,2059	0,0874	0,0352	0,0134	0,0047	0,0016	0,0005	0,0001	0	0	
15	1	0,549	0,3186	0,1671	0,0802	0,0353	0,0142	0,0052	0,0017	0,0005	0,0001	0
15	2	0,8159	0,6042	0,398	0,2361	0,1268	0,0617	0,0271	0,0107	0,0037	0,0011	0,0003
15	3	0,9444	0,8227	0,6482	0,4613	0,2969	0,1727	0,0905	0,0424	0,0176	0,0063	0,0019
15	4	0,9873	0,9383	0,8358	0,6865	0,5155	0,3519	0,2173	0,1204	0,0592	0,0255	0,0093
15	5	0,9978	0,9832	0,9389	0,8516	0,7216	0,5643	0,4032	0,2608	0,1509	0,0769	0,0338
15	6	0,9997	0,9964	0,9819	0,9434	0,8689	0,7548	0,6098	0,4522	0,3036	0,1818	0,095
15	7	10.000	0,9994	0,9958	0,9827	0,95	0,8868	0,7869	0,6535	0,5	0,3465	0,2131
15	8		0,9999	0,9992	0,9958	0,9848	0,9578	0,905	0,8182	0,6964	0,5478	0,3902
15	9		10.000	0,9999	0,9992	0,9963	0,9876	0,9662	0,9231	0,8491	0,7392	0,5968
15	10			10.000	0,9999	0,9993	0,9972	0,9907	0,9745	0,9408	0,8796	0,7827
15	11				10.000	0,9999	0,9995	0,9981	0,9937	0,9824	0,9576	0,9095
15	12					10.000	0,9999	0,9997	0,9989	0,9963	0,9893	0,9729
15	13						10.000	10.000	0,9999	0,9995	0,9983	0,9948
15	14								10.000	10.000	0,9999	0,9995
15	15										10.000	10.000

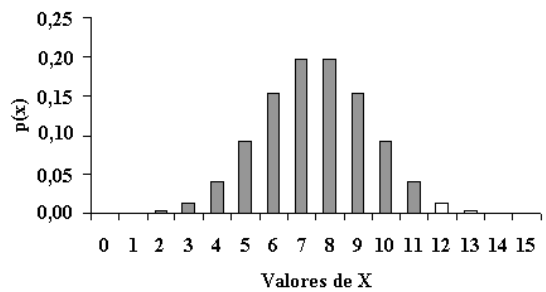
Ejemplo 3.3.

Si se tiene una variable aleatoria que se distribuye binomialmente con n = 15 y p = 0,5, encuentre las probabilidades siguientes:

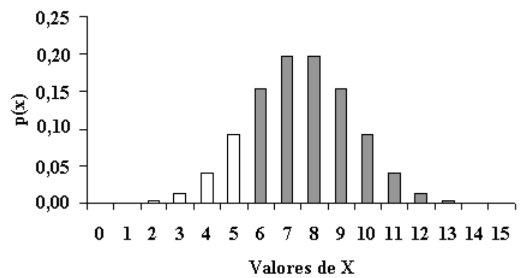
a) $P(X \leq 3)$; *b)* $P(X > 5)$; *c)* $P(X < 8)$; *d)* $P(X \geq 9)$; *e)* $P(X = 6)$; *f)* $P(10 < X \leq 14)$

Para encontrar estas probabilidades, primero ubicamos en la tabla los parámetros de la distribución $n = 15$ y $p = 0,5$. Luego encontramos la probabilidad acumulada para el valor de x correspondiente. Así tenemos que:

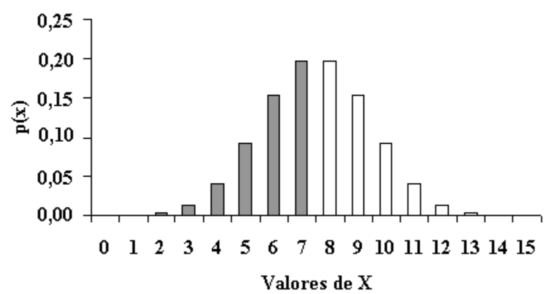
$a)\ P(X \le 11) = 0,9824$



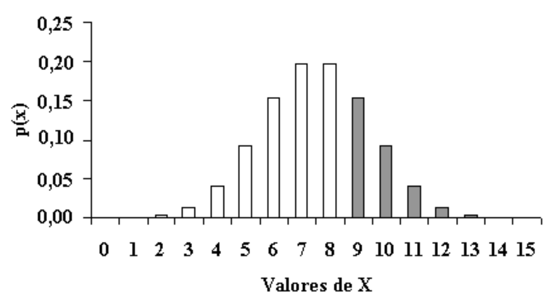
$b)\ P(X>5) = 1 - P(X \le 5) = 1 - 0,150 = 0,8491$



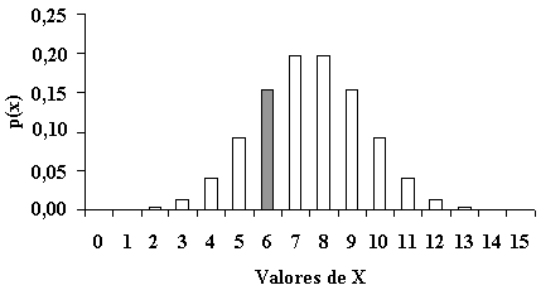
$c)\ P(X<8) = P(X \le 7) = 0,500$



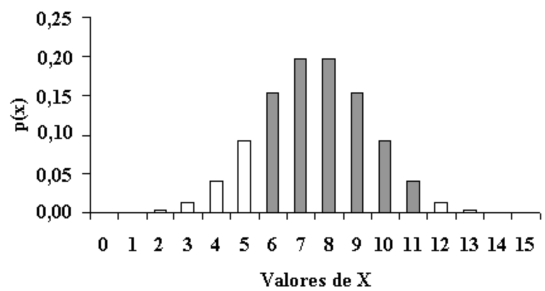
$d)\ P(X \ge 9) = 1 - P(X \le 8) = 1 - 0,6964 = 0,3036$



$$e)\ P(X=6) = P(X \le 6) - P(X \le 5) =$$
$$= 0,3036 - 0,1509 = 0,1527$$



f) $P(5<X \leq 11)=P(X \leq 11)-P(X \leq 5)=$
$= 0,9824 - 0,1509 = 0,8315$



Las áreas sombreadas de las figuras anteriores representan las probabilidades solicitadas.

3.2.2.3 Valor Esperado (Media) y Varianza de la Distribución Binomial

El valor esperado y la varianza de la distribución binomial son los siguientes:

$$E(X) = \mu = np \qquad Var(X) = \sigma^2 = npq$$

Ejemplo 3.4.
Calcule el número promedio y la varianza que se espera obtener del número de niños con fiebre cuando se vacunan 6 niños, si la probabilidad de que se enferme un niño es de 1/4.

$$\mu = np = 6(1/4) = 1{,}5 \text{ niños} \qquad \sigma^2 = npq = 6(1/4)(3/4) = 1{,}125$$

3.2.3. Modelo de Poisson.

Esta distribución fue introducida por el matemático francés S.D. Poisson en 1837. El modelo de Poisson a semejanza del binomial consta de varios ensayos de Bernoulli. La diferencia estriba en que el modelo binomial sirve para calcular la probabilidad de ocurrencia de un resultado particular en un número finito de repeticiones, mientras que con el modelo de Poisson se determina la probabilidad de ocurrencia de un determinado evento en el tiempo o el espacio y no en un número definido de repeticiones del experimento. En estos eventos que se producen aleatoriamente en el espacio o el tiempo, la frecuencia de ocurrencia de un evento es tan baja con relación a la frecuencia de no ocurrencia que se consideran como sucesos raros. Tratar de describir la distribución de una variable aleatoria de este tipo mediante el modelo binomial sería impráctico puesto que el número de ensayos tendría que ser extraordinariamente grande para que ocurriera el resultado esperado. Por ejemplo, para un biólogo que está colectando individuos de una especie de planta cuyos individuos están distribuidos aleatoriamente e independientemente en una sabana, es de suma importancia conocer la distribución de probabilidades de la variable número de plantas. Para obtener ésta distribución se podría usar el modelo binomial. Sólo se necesitaría considerar cada punto muestreado como una repetición del proceso, sin embargo esto implicaría trabajar con un número de repeticiones extremadamente grande, puesto que la presencia de una planta en un punto del área de búsqueda es un hecho muy poco frecuente con relación al número de puntos donde no se encuentra. Bajo el supuesto que se pudiera superar la dificultad del elevado número de repeticiones, se tendría otro problema, como es que la función binomial está caracterizada por un valor de n muy grande y un valor de p muy pequeño, lo que hace sumamente tedioso el cálculo de probabilidades por tener que usar factoriales de números muy grandes. Afortunadamente, situaciones como la planteada donde $n \to \infty$ y $p \to 0$, se pueden resolver usando el modelo probabilístico de Poisson.

Para deducir la función de probabilidad de Poisson seguiremos haciendo uso del mismo ejemplo de la planta en la sabana. Para tal efecto se hará uso de dos supuestos: el primero es que en esta sabana se delimitó una parcela de terreno que tiene un número promedio de plantas igual a λ; y el segundo es que el área de la parcela se corresponde con una unidad de superficie, de forma que λ representa el número promedio de plantas por unidad de superficie. El mayor interés es el de conocer la probabilidad con la cual la variable aleatoria asume los valores de su rango espacial, el cual es $Rx = \{0,1,2,3,........N\}$. Una manera de encontrar las probabilidades para cada resultado en Rx sería definir un número grande de parcelas del mismo tamaño, determinar su número de plantas y construir una distribución de frecuencias. Las frecuencias relativas se pueden usar para estimar la probabilidad de ocurrencia de la variable número de plantas. Otra forma sería dividir la parcela en *n* unidades del mismo tamaño lo suficientemente pequeñas para que en cada uno de ellas se produzca uno de dos resultados: presencia o ausencia de plantas (Figura 3.3).

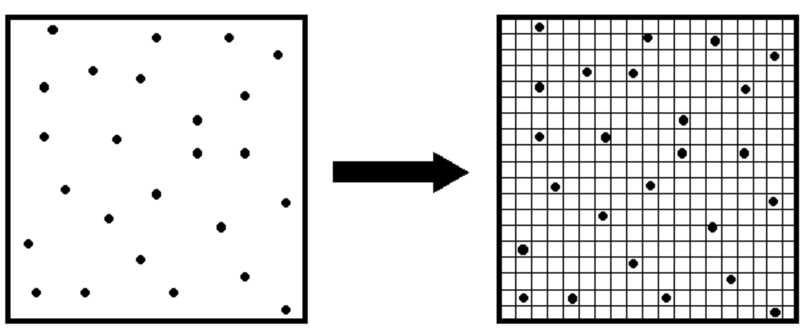


Figura 3.3. Parcela de terreno dividida en n subparcelas. Los puntos indican la ubicación hipotética de las plantas (ver texto)

Bajo estas nuevas condiciones el experimento presenta las características de un experimento binomial:

a) Se repite *n* veces.
b) Dos resultados posibles. Si la probabilidad de ocurrencia de una planta es muy pequeña con relación a la probabilidad de que no ocurra y cada subparcela es lo suficientemente pequeña, entonces la posibilidad de encontrar mas de una planta en una misma subparcela es infinitamente pequeña y para efectos prácticos se puede considerar que en cada subparcela existen dos resultados posibles: presencia o ausencia de la planta.

c) Para cada subparcela existe la misma posibilidad de tener una planta.

d) Los n ensayos son independientes. El hecho de que ocurra una planta en un punto, no afecta su ocurrencia en otro punto.

Por lo tanto, es posible utilizar la función de probabilidad del modelo binomial para el cálculo de probabilidades.

X = número de plantas en n repeticiones

$$\downarrow$$

$$Rx = \{0, 1, 2, 3, \ldots\ldots\ldots n\}$$

$$\downarrow$$

$$P(X = x) = p(x) = {}_nC_x\, p^x q^{n-x}$$

Para aplicar la función de probabilidades falta conocer el valor de *p*. Este se puede deducir a partir de λ que es el número promedio de plantas por parcela o por unidad de superficie (véase segundo supuesto). Puesto que la parcela se dividió en n subparcelas, la probabilidad de que ocurra una planta en cada una de las n subparcelas de tamaño 1/n será $p = \lambda/n$ y la probabilidad de que no ocurra será $q = 1-(\lambda/n)$, de modo que la función $p_{(x)} = {}_nC_x p^x q^{n-x}$ queda de la manera siguiente:

$$p(x) = {}_nC_x\, [\lambda/n]^x\, [1\text{-}\lambda/n]^{n-x}$$

Sin embargo esta función sólo es una aproximación, pues toma en cuenta *n* subparcelas. Como la superficie es una variable continua, el área de la parcela se puede dividir en infinitas subparcelas, de modo que cuando *n* tiende a infinito la función de probabilidad binomial se aproxima al valor siguiente:

$$\underset{n \to \infty}{Limite}\ {}_nC_x\, [\lambda/n]^x\ [(1-\lambda)/n]^{n-x} = \frac{e^{-\lambda}\ \lambda^x}{x\,!}$$

Donde λ = número de ocurrencia del evento de interés en una unidad de espacio (o tiempo). Para cualquier otro valor de espacio (o tiempo) la función de probabilidad será:

$$p(x) = {}_nC_x\, [\lambda a/n]^x\ [(1-\lambda a)/n]^{n-x} = \frac{e^{-\lambda a}\ \lambda a^x}{x\,!}$$

Donde *a* es un factor de proporcionalidad que permite calcular el número de ocurrencias del éxito en un tiempo o espacio dado diferente a la unidad.

Sí se hace $\lambda a = \mu$ la función de probabilidades para el modelo de Poisson queda de la manera siguiente:

$$p_{(x)} = \frac{e^{-\mu}\ \mu^{x}}{x!}$$

Donde μ = es el número promedio de ocurrencias en un espacio o tiempo dado y x = número de veces que ocurre el éxito en ese mismo espacio o tiempo.

3.2.3.1 Distribución de probabilidades del modelo de Poisson

La distribución de probabilidades de Poisson esta formada por los pares ordenados de valores [x_i; $p(x_i)$] y la misma está caracterizada por un sólo parámetro: el promedio μ. En forma similar a la distribución binomial, la distribución de Poisson es una familia de curvas, cuya forma depende del valor de μ (Figura 3.4)

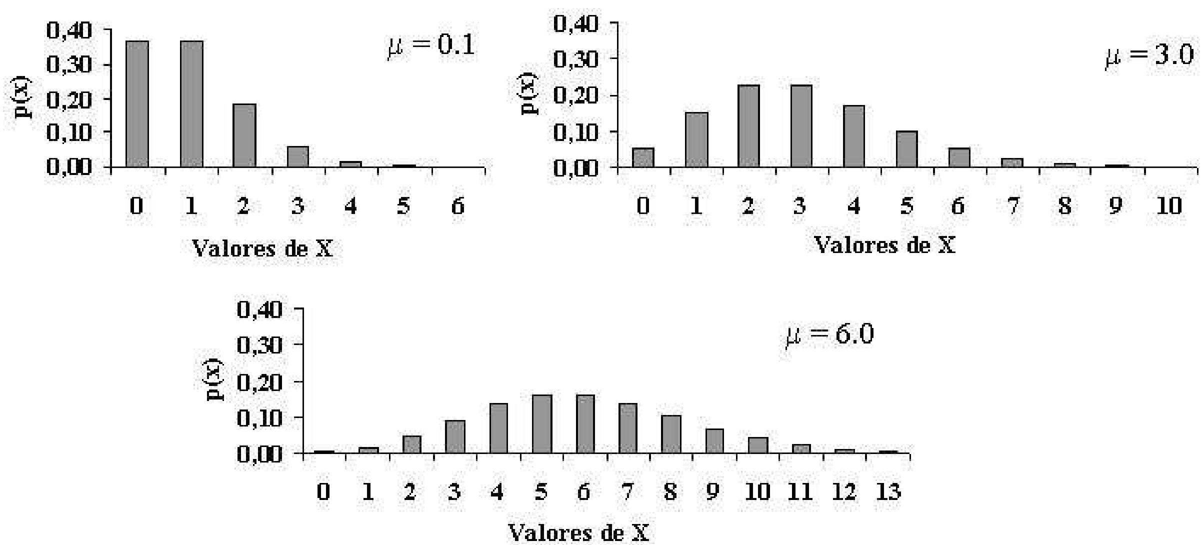


Figura 3.4. Distribución de probabilidades para el modelo de Poisson

Ejemplo 3.5.

Supóngase que el número de partículas radiactivas emitidas por cierto material durante una hora tiene una distribución de Poisson cuyo promedio es de 0,8 partículas/hora. ¿Cuál es la probabilidad de que en cinco horas se emitan mas de 3 y menos de 7 partículas?.

Para encontrar la probabilidad solicitada $P(3 < X < 7)$ será necesario calcular las probabilidades individuales $p_{(4)}$, $p_{(5)}$ y $p_{(6)}$, de forma que,

$$P(3 < X < 7) = P(4 \leq X \leq 6) = p_{(4)} + p_{(5)} + p_{(6)}$$

Para el cálculo de cada probabilidad se usa la función de Poisson: $p_{(x)} = e^{-\mu}\ \mu^{x} / x!$.

Si $\lambda = 0{,}8\ emisiones/hora$, el número promedio esperado para cinco horas será $\mu = \lambda t = (0{,}8)(5) = 4{,}0\ emisiones$. Las probabilidades requeridas son:

$$p(4) = e^{-4}4^4/4! = (0{,}0183)(256)/24 = 4{,}688/24 = 0{,}1954$$

$$p(5) = e^{-4}4^5/5! = (0{,}0183)(1024)/120 = 18{,}74/120 = 0{,}1562$$

$$p(6) = e^{-4}4^6/6! = (0{,}0183)(4096)/720 = 74{,}96/720 = 0{,}1041$$

Al sumar estos valores se obtiene la probabilidad total buscada,

$$P(3 < X < 7) = P(4 \leq X \leq 6) = 0{,}1954 + 0{,}1562 + 0{,}1041 = 0{,}4557$$

Ejemplo 3.6.
Se ha encontrado que el número de fallas en un transistor de un aparato de medición es una variable aleatoria que se distribuye como Poisson. El promedio de fallas es de 0,1 cada 10 horas. ¿Cuál será la probabilidad de que el transistor falle en dos horas?

La probabilidad de que el transistor falle será la probabilidad de que la variable aleatoria *X* = *número de fallas en dos horas* tome un valor mayor o igual a 1.

Si $\lambda = 0{,}1\ fallas/10\ horas = 0{,}01\ fallas/hora$, entonces $\mu = \lambda t = (0{,}01)(2) = 0{,}002\ fallas$

La probabilidad requerida es:

$$P(X \geq 1) = 1 - P(X \leq 0) = 1 - p_{(0)} = 1 - \frac{e^{\mu} - \mu^0}{0!} = 1 - \frac{e^{0{,}002} - 0{,}002^0}{0!} = 1 - 0{,}9892 = 0{,}0198$$

3.2.3.2 Función de probabilidad acumulada para el modelo de Poisson
En la Tabla 3.2 se muestran una parte de la Tabla de probabilidades acumuladas de la función de probabilidad de Poisson. Las entradas para ésta tabla son el parámetro μ y el número *x* de éxitos.

Tabla 3.2: Algunas probabilidades acumuladas de la Distribución de Poisson

$$F(x) = P(X \leq x) = \sum^{Rx} \frac{e^{-\mu}\mu^x}{x!}$$

	μ									
X	1,8	1,9	2	2,1	2,2	2,3	2,4	2,5	2,6	2,7
0	0,1653	0,1496	0,1353	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672
1	0,4628	0,4337	0,406	0,3796	0,3546	0,3309	0,3084	0,2873	0,2674	0,2487
2	0,7306	0,7037	0,6767	0,6496	0,6227	0,596	0,5697	0,5438	0,5184	0,4936
3	0,8913	0,8747	0,8571	0,8386	0,8194	0,7993	0,7787	0,7576	0,736	0,7141
4	0,9636	0,9559	0,9473	0,9379	0,9275	0,9162	0,9041	0,8912	0,8774	0,8629
5	0,9896	0,9868	0,9834	0,9796	0,9751	0,97	0,9643	0,958	0,951	0,9433
6	0,9974	0,9966	0,9955	0,9941	0,9925	0,9906	0,9884	0,9858	0,9828	0,9794
7	0,9994	0,9992	0,9989	0,9985	0,998	0,9974	0,9967	0,9958	0,9947	0,9934
8	0,9999	0,9998	0,9998	0,9997	0,9995	0,9994	0,9991	0,9989	0,9985	0,9981
9	10.000	10.000	10.000	0,9999	0,9999	0,9999	0,9998	0,9997	0,9996	0,9995
10				10.000	10.000	10.000	10.000	0,9999	0,9999	0,9999

Ejemplo 3.7.
Supóngase que el número de impulsos que recibe una central telefónica es una variable que se distribuye como Poisson El promedio es de 120 impulsos recibidos por hora. La central tiene una capacidad máxima de 4 impulsos/minuto. ¿Cuál es la probabilidad de que en un minuto determinado la central se congestione?

La central comenzará a fallar cuando el número de impulsos sea superior a 4 por minuto, de modo que la probabilidad solicitada es $P(X > 4)$.

Si $\lambda = 120 \text{ impulsos/hora} = 120 \text{ impulsos/60 minutos} = 2 \text{ impulsos/minuto}$, entonces se tiene que $\mu = \lambda t = (2 \text{ impulsos / minuto})(1 \text{ minuto}) = 2 \text{ impulsos}$

La probabilidad solicitada es $P(X > 4) = 1 - P(X \leq 4) = 1 - F_{(4)}$. Para encontrar el valor de F(4) se localiza el valor de $\mu = 2$ en la primera fila de la Tabla 3.2, luego en la primera columna se ubica el valor de $x = 4$. El valor de la intersección *0,9473* representa las probabilidades acumuladas de la distribución de Poisson para $x = 4$ cuando $\mu = 2$, por tanto

$$P(X > 4)= 1- P(X \leq 4)= 1 - F(4)= 1- 0,9473= 0,0527$$

Ejemplo 3.8.
Encontrar la probabilidad solicitada en el Ejemplo 3.5 utilizando la función acumulativa

La probabilidad solicitada es:

$$P(3 < X < 7) = P(4 \leq X \leq 6) = P(X \leq 6) - P(X \leq 3) = F_{(6)} - F_{(3)}$$

En una Tabla de probabilidades acumuladas para el modelo de Poisson se encuentran los valores de F(6) y F(3) para $\mu = 4$ y se resuelve.

$$P(3 < X < 7) = P(4 \leq X \leq 6) = P(X \leq 6) - P(X \leq 3) = F_{(6)} - F_{(3)} = 0,8893 - 0,4335 = 0,4558$$

Ejemplo 3.9.

El número de accidentes severos que ocurren anualmente en los laboratorios de química según la Federación Internacional de Laboratorios de Química es de 2,5 casos por cada 1000 profesionales del área. ¿Cuál es la probabilidad de que en una población de 80 personas que trabajan en los laboratorios de un Departamento de Química, se produzca algún accidente en el año?

$$\lambda = 2{,}5 \text{ casos/1000 personas} = 2{,}5 \times 10^{-3} \text{ casos/persona}$$

$$\mu = \lambda t = (2{,}5 \times 10^{-3} \text{ casos/persona})(80 \text{ personas}) = 0{,}2 \text{ casos}$$

$$P(X > 0) = 1 - P(X \leq 0) = 1 - F_{(0)} = 1 - 0{,}8187 = 0{,}1813$$

3.2.3.3 Relación entre los modelos Binomial y Poisson

La deducción de la función de probabilidad del modelo de Poisson se hizo a partir de la función de probabilidad del modelo binomial. Con éste propósito en un experimento binomial se aumentó infinitamente el número de repeticiones n, y la probabilidad de ocurrencia del éxito se disminuyó proporcionalmente a éste aumento, $p = \lambda/n$. De modo que en cualquier ensayo de Bernoulli donde n sea muy grande y p muy pequeño, se puede utilizar la función de Poisson para calcular las probabilidades de ocurrencia del éxito, sabiendo que $\mu = np$.

Ejemplo 3.10.

En una fábrica, se sabe que en el proceso de envasado de alimentos uno de cada 1000 envases producidos no cumple con las normas de calidad del producto. ¿Cuál es la probabilidad de que en una muestra aleatoria de 8000 envases, siete o menos no cumplan con las normas de calidad?

La variable ***X*** = *número de envases sin calidad* se distribuye binomialmente: a) en cada elección existe dos resultados posibles, b) la probabilidad de encontrar un envase que no cumpla con la garantía no cambia, c) los ensayos son independientes y d) el experimento se repitió n veces.

$$n = 8000 \quad \text{y} \quad p = 1/1000 = 0{,}001$$

La función acumulativa no esta tabulada para valores de n tan grandes. El cálculo mediante la función de probabilidad se dificulta por que hay que trabajar con factoriales muy grandes. Sin embargo si consideramos que n tiende a infinito y p tiende a cero, podemos aproximar la Binomial a Poisson, de modo que: $\mu = np = (8000)(0{,}001) = 8 \text{ envases}$

La probabilidad deseada las buscamos en la tabla de probabilidades acumuladas del modelo de Poisson para $\mu = 8$.

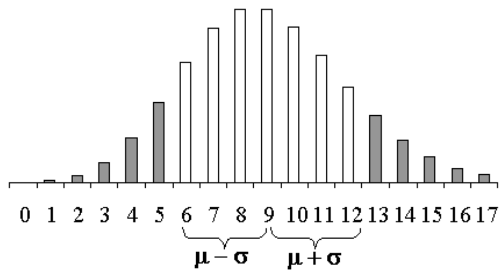
$$P(X \leq 7) = F_{(7)} = 0{,}3134$$

3.2.3.4 Valor Esperado y Varianza de la Distribución de Poisson

El valor esperado $\left[E(X) = \mu\right]$ y la varianza $\left[Var(X) = \sigma^2\right]$ de la distribución de Poisson son iguales: $\mu = \sigma^2$.

Ejemplo 3.11.
Sea X una variable que se distribuye según el modelo de Poisson, sabiendo que $\mu = 9$ calcule la probabilidad que tiene la variable aleatoria de ser mayor o menor a la media en más de una desviación estándar.

La probabilidad solicitada es $P[\, X < (\mu - \sigma) \text{ ó } X > (\mu + \sigma)]$



Como se sabe que en el modelo de Poisson $\mu = \sigma^2$, se deduce que $\sigma^2 = \mu = 9$. Por lo tanto el valor de la desviación estándar es $\sigma = \sqrt{9} = 3$, de modo que:

$$P[X < (\mu-\sigma) \text{ ó } X > (\mu+\sigma)] = P(X < \mu-\sigma) + P(X > \mu+\sigma) = P(X<9-3) + P(X > 9+3) = P(X < 6) + P(X > 12) =$$

$$= P(X \leq 5) + [1 - P(X \leq 12)] = P(X \leq 5) + 1 - P(X \leq 11) = 0{,}1157 + 1 - 0{,}8758 = 0{,}2399$$

3.3 MODELOS PROBABILÍSTICOS CONTINUOS

3.3.1. Modelo Normal.

La función de probabilidad de la distribución normal sirve de modelo para una gran cantidad de variables continuas naturales, tales como la temperatura, la humedad, la precipitación, la altura, el peso, la concentración, el coeficiente de inteligencia, los errores instrumentales, etc. Igualmente la distribución de muchos estadísticos tienden hacia la distribución normal, por lo cual ésta distribución adquiere una gran importancia en el análisis de datos mediante la inferencia estadística. La fórmula de la función de probabilidad del modelo normal fue introducida por Abraham De Moivre en las primeras décadas del siglo XVIII, como una derivación de la distribución binomial cuando n es muy grande. Luego fue usada por Pierre Laplace y Carl Gauss con diferentes propósitos. Erróneamente, se le ha atribuido a éste último el descubrimiento de la función de probabilidad normal, por lo que su distribución muchas veces se le denomina Campana de Gauss, para honrar a su supuesto creador y por la forma de la distribución que semeja a una campana. El nombre de distribución normal le fue dado por Karl Pearson, otro insigne matemático. Una variable aleatoria **X** se encuentra distribuida normalmente si su función de probabilidad es la siguiente:

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

Esta función esta caracterizada por dos parámetros: la media (μ) y la desviación estándar (σ). El valor de μ define la posición de la distribución y el valor de σ define la forma de la distribución (Figura 3.5). Para cada par de valores distintos de μ y σ existe una distribución normal, por lo tanto hay un número infinito de distribuciones normales diferentes.

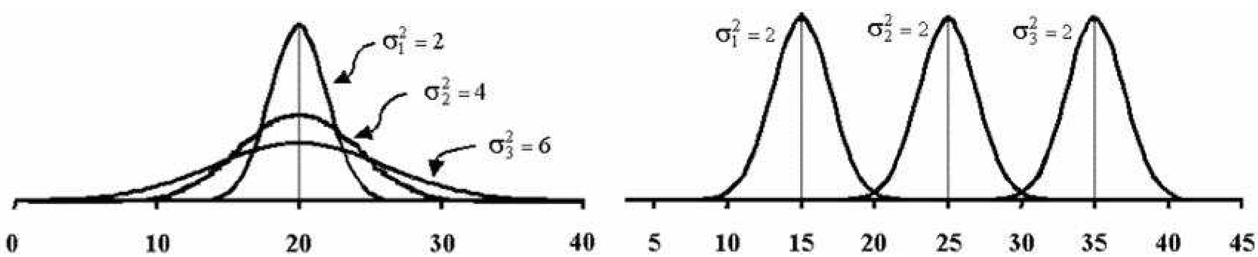


Figura 3.5. Distribuciones normales con la misma media ($\mu = 20$) y diferentes varianzas (arriba); Distribuciones normales con diferentes medias y la misma varianza (abajo)

La distribución normal es simétrica, con un valor máximo para $x = \mu$ y presenta dos puntos de inflexión para $x = \pm\sigma$ (Figura 3.6)

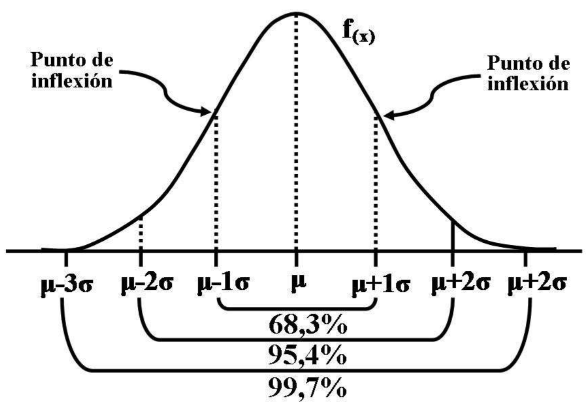


Figura 3.6. Distribución normal y arreas contenidas por los intervalos definidos a una, dos y tres desviaciones estándar alrededor de μ.

La función de probabilidad *f(x)* tiende a cero a medida que *x* tiende a $\pm\infty$, por lo que las dos colas de la distribución se aproximan asintóticamente a cero. Todas estas propiedades le confieren a la distribución normal una forma característica semejante a una campana. Cuando una variable aleatoria sigue la distribución normal se indica $X : N(\mu;\sigma)$. Por tratarse de un modelo para variables continuas, la probabilidad de que la variable se encuentre en un intervalo se obtiene integrando la función *f(x)* entre los límites del intervalo.

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f_{(x)}dx$$

Igualmente se puede calcular la probabilidad utilizando la función acumulativa $\phi_{(x)}$.

$$P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = \int_{-\infty}^{x_2} f_{(x)}dx - \int_{-\infty}^{x_1} f_{(x)}dx = \phi_{(2)} - \phi_{(1)}$$

En el caso de las distribuciones discretas los valores de la función acumulativa están tabulados para diferentes valores de los parámetros que caracterizan estas distribuciones. Esto no es posible en el caso de la distribución normal porque al ser aplicable a variables continuas existen infinitos valores de μ y σ. Afortunadamente ésta situación se resolvió tabulando las probabilidades acumuladas para una única distribución, con valores de μ y σ específicos, y mediante el procedimiento de tipificación transformar cualquier variable normal, en esa variable estándar o patrón. La variable que se seleccionó como estándar es aquella cuya función tiene como parámetros $\mu = 0$ y $\sigma = 1$, por lo cual se le denominó variable normal estándar, unitaria o tipificada, identificándose con la letra Z para diferenciarla de las otras variables cuya distribución de probabilidad tiene $\mu \neq 0$ y $\sigma \neq 1$. La función de probabilidad de la variable Z es la siguiente:

$$f_{(z)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}[z]^2}$$

La probabilidad de encontrar un valor de **Z** en un intervalo dado, se obtiene calculando el área que se encuentra entre la curva y el intervalo definido en el eje de coordenadas. Pero en lugar de integrar $f_{(z)}$ entre los límites del intervalo, ésta área se puede calcular utilizando la tabla de la función acumulada $\phi_{(z)}$, que proporciona los valores de integración entre -∞ y un dado valor de Z **(**Figura 3.7). En la Tabla 3.3 se muestra una parte de la tabla de $\phi_{(z)}$

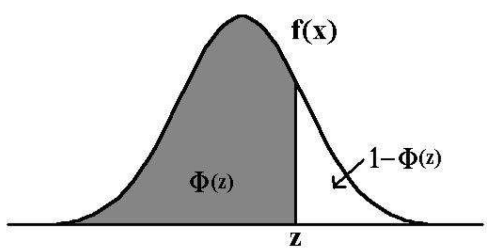


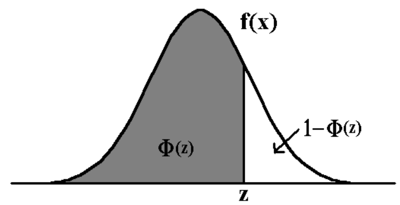
Figura 3.7. Función acumulativa Φ(z)

La primera columna de la Tabla 3.3 son los valores de Z con una apreciación en décimas y la primera fila son las centésimas de Z. Las cantidades dentro de la tabla son las probabilidades acumuladas desde - ∞ hasta un valor dado de Z. Por ejemplo, a la izquierda de z = 1,25 la probabilidad acumulada es 0,8944.

Tabla 3.3: Algunas Probabilidades acumuladas Φ(z) de la Distribución Normal Estandarizada.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406

$$\phi_{(z)} = \int_{-\infty}^{z} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}[z]^2}$$



A continuación se desarrollarán algunos ejemplos sobre el uso de la función acumulada $\phi_{(z)}$.

Ejemplo 3.12.
Si se elige aleatoriamente un valor de la variable **Z**, calcule las probabilidades siguientes:

a) $P(Z < 1,2)$

b) $P(-0,2 \le Z \le 1,35)$

c) $P(Z \ge 0,75)$

d) $P(-0,45 \le Z \le -0,13)$

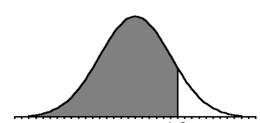
e) $P(0,12 \le Z \le 1,14)$

f) $P(-0,45 \ge Z \ge 0,45)$

Respuestas

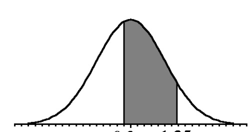
a) $P(Z < 1,2) = P(Z \le 1,2) = \phi_{(1,2)} = 0,8849$

Las áreas sombreadas en la figura de la derecha representan las áreas o probabilidades solicitadas



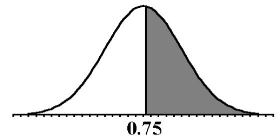
b) $P(-0,2 \le Z \le 1,35) = P(Z \le 1,35) - P(Z \le -0.2) =$

$= \phi_{(1,35)} - \phi_{(-0,2)} = 0,9115 - 0,4207 = 0,4908$



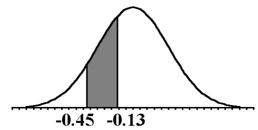
$c)\ P(Z \geq 0{,}75) = 1 - P(Z \leq 0{,}75) = 1 - \phi_{(0{,}75)} =$

$= 1 - 0{,}7734 = 0{,}2260$



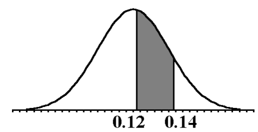
$d)\ P(-0{,}45 \leq Z \leq -0{,}13) = P(Z \leq -0{,}13) - P(Z \leq -0{,}45) =$

$= \phi_{(-0{,}13)} - \phi_{(-0{,}45)} = 0{,}4483 - 0{,}3264 = 0{,}1219$



$e)\ P(0{,}12 \leq Z \leq 1{,}14) = P(Z \leq 1{,}14) - P(Z \leq 0{,}12) =$

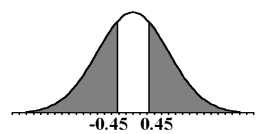
$= \phi_{(1{,}14)} - \phi_{(0{,}12)} = 0{,}8729 - 0{,}5478 = 0{,}3251$



$f)\ P(-0{,}45 \geq Z \geq 0{,}45) = P(Z \leq -0{,}45) + P(Z \geq 0{,}45) =$

$= P(Z \leq -0{,}45) + [1 - P(Z \leq 0{,}45] = \phi_{(-0{,}45)} + 1 - \phi_{(0{,}45)} =$

$= 0{,}3264 + 1 - 0{,}6736 = 0{,}3264 + 0{,}3264 = 0{,}6528$



3.3.3.1 Transformación de una variable X en la variable Z.

Al transformar la función $f_{(x)}$ en la función $f_{(z)}$, lo que realmente se hizo fue sustituir el término $\frac{x-\mu}{\sigma}$ por la variable z

$$f_{(x)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \xrightarrow[\sigma = 1]{Z = \frac{x-\mu}{\sigma}} f_{(z)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

De manera que cualquier variable **X** que se distribuye normalmente con $\mu \neq 0$ y/o $\sigma \neq 1$, se puede convertir en la variable **Z**, restando a todo valor de X su media μ y dividiendo esta diferencia entre su desviación estándar σ.

Al dividir la distancia $x - \mu$ entre σ, se está determinando el número de desviaciones estándar que caben en esa distancia. Por lo tanto los valores de Z expresan la distancia de X respecto a su media μ en términos de desviación estándar. Por ejemplo si un valor de una variable X al transformarse produce un valor de $z = 1{,}5$, este último indica que el valor de X está a $1{,}5\sigma$ a la derecha de μ. Igualmente un valor de $z = -2$, indica que el valor de X está a 2σ a la izquierda de μ. (Figura 3.7)

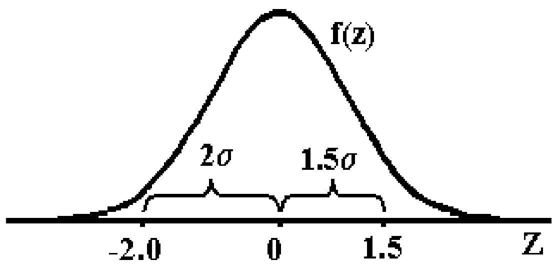


Figura 3.7. Equivalencia entre valores de Z y el número de desviaciones estándar

Otro aspecto importante a considerar en la transformación de una variable X en la variable Z, es que para toda distribución normal, independientemente de cuales sean sus valores de μ y σ, el área bajo la curva es la misma para todo intervalo cuyos límites estén ubicados a una misma distancia de μ en términos del número de desviaciones estándar. Por ejemplo, en toda distribución normal, el área sobre el intervalo $\mu \pm 1\sigma$ representa el de 68,3% del área total; si el intervalo tiene como límites $\mu \pm 2\sigma$ lo cubre un área igual al 95,4% y sobre el intervalo $\mu \pm 3\sigma$ el área representa un 99,7% del total.
Estas dos propiedades que acabamos de ver son las que permiten calcular las probabilidades de una distribución normal, cualquiera sean los valores de μ y σ, a partir de las probabilidades acumuladas para la distribución normal estandarizada o tipificada.

Ejemplo 3.13. Sea $X : N(20;4)$ y se quiere conocer la probabilidad de que la variable tenga un valor menor a 16.

La probabilidad que nos interesa es $P(X \leq 16)$. El área que representa esta probabilidad se encuentra a la izquierda del valor 16 (Figura 3.8). Para poder determinar el valor de ésta área mediante la tabla de probabilidades acumuladas de la distribución normal estándar, se debe convertir el valor de x en su respectivo valor z, lo cual se hace mediante la operación siguiente:

$$z = \frac{x-\mu}{\sigma} = \frac{16-20}{4} = -1$$

Ahora se busca el área que se encuentra a la izquierda de $z = -1$ en la tabla de probabilidades acumuladas para la variable Z, y se toma dicha área como la probabilidad conque la variable aleatoria X asume un valor igual o menor a 16 (Figura 3.8). En consecuencia se tiene,

$$P(X \leq 16) = P\left(Z \leq \frac{16-20}{4}\right) = P(Z \leq -1) = \phi_{(-1)} = 0,1587$$

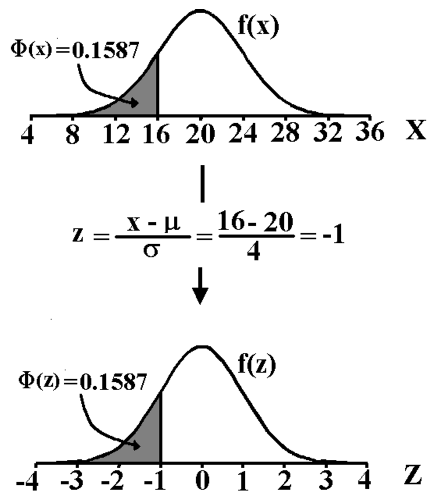
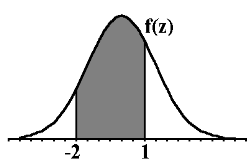


Figura 3.8.

Ejemplo 3.14.
El nivel de colesterol en la sangre en los individuos adultos de 25 a 30 años de cierta población es una variable que se distribuye normalmente con una media de 190 mg/dl y una desviación de 20 mg/dl, encuentre la probabilidad que un individuo elegido aleatoriamente tenga valores de colesterol a) entre 150 y 210 mg/dl; b) mayor a 230 o menor a 180

$$a)\ P(150 \le X \le 210) = P\left(\frac{150-190}{20} \le Z \le \frac{210-190}{20}\right) =$$

$$= P(Z \le 1) - P(Z \le -2) = 0{,}8413 - 0.02258 = 0{,}8185$$

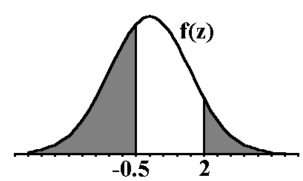


$$b)\ P(180 \ge X \ge 230) = P(X \le 180) + P(X \ge 230) =$$

$$= P\left(Z \le \frac{180\text{-}190}{20}\right) + P\left(Z \ge \frac{230-190}{20}\right) =$$

$$= P(Z \le \text{-}0{,}5) + P(Z \ge 2) = P(Z \le \text{-}0{,}5) + [1\text{-}P(Z \le 2)] =$$

$$= 0{,}3085 + 1\text{-}0{,}9772 = 0{,}3313$$



3.4 EJERCICIOS

1. En una experiencia de laboratorio se inocularon 15 conejos con un suero que contiene un germen que produce cierta enfermedad. Se sabe que el 40% de todos los conejos inoculados con éste suero contraerán dicha enfermedad (reacción positiva). ¿Responda?:

a. ¿Cuál es la variable aleatoria en éste experimento?
b. ¿Cuál es la distribución probabilística que sigue la variable aleatoria?
c. ¿Cuál es el espacio muestral de la variable aleatoria?
d. Calcule la probabilidad de que se produzca una reacción positiva:
  d1. En 12 conejos.
  d2. En menos de tres conejos.
  d3. En al menos tres conejos.
  d4. A lo sumo en tres conejos.

2. Se espera que el 90% de las semillas de una nueva variedad de maíz germine. Para verificar tal presunción se siembran 20 semillas en recipientes separados y se mantienen bajo las mismas condiciones ambientales ¿Calcule la probabilidad de que:
a. Germinen 15 semillas y 5 no germinen?
b. Germinen 13 o menos semillas?
c. Germinen entre 15 y 19 semillas (ambas inclusive)
d. Germinen más semillas que las que no germinan?

3. En un estudio sobre granivoría (depredación de semillas) se encuentra que el 60% de las hormigas que regresan al nido lleva cargada una semilla. Calcule la probabilidad que de las próximas 15 hormigas que regresan al hormiguero:
a. Más de 12 estén cargadas.
b. Entre 2 y 14 estén cargadas (ambos inclusive).
c. Se encuentren entre ellas 10 sin carga alguna.

4. Sea X una variable aleatoria discreta que sigue la distribución binomial con p = 0,5. Suponiendo que n = 5, encuentre la probabilidad de que la variable tome un valor en el intervalo.
a. $\mu - \sigma < X < \mu + \sigma$
b. $x < \mu - \sigma \quad ó \quad x > \mu + \sigma$
c. $x < \mu - \sigma \quad y \quad x > \mu + \sigma$

5. En un poblado el porcentaje de individuos cuya sangre tiene el factor Rh(+) es de 65%. Si se toma una muestra de sangre a 6 individuos en el mismo poblado, calcule:
a. Distribución de probabilidad de la variable número de individuos Rh(+).
b. Probabilidad de que se encuentren menos de dos, o más de cuatro individuos Rh(+).
c. Si se han encontrado más de tres individuos Rh(+) ¿cuál es la probabilidad de que se encuentren entre dos y cinco individuos Rh(+)? (ambos valores inclusive).

6. El 40% de las flores de un jardín no tienen néctar al final del día. ¿Cuál es la probabilidad de que al final de un día cualquiera después de examinar 15 flores al azar, el número de flores vacías difiera de la media poblacional en más de 2 individuos?

7. Se sabe que el 10% de los huevos producidos por una especie de tortuga son infértiles. Un biólogo necesita obtener crías de la tortuga y recolectó 20 huevos de la misma especie y formó dos grupos de 10 huevos que fueron colocados en una incubadora ¿cuál es la

probabilidad de que:

a. En los dos grupos nazcan menos de 8 crías
b. En el primer grupo nazcan menos de 8 crías y en el segundo grupo más de 8 crías
c. En el primer grupo nazcan 5 o menos crías ó más de 8 crías en el segundo grupo

8. Se censó el número de individuos adultos de las dos únicas especies de conejos que habitan en una pequeña isla, encontrándose los resultados siguientes:

	Especie A	Especie B
Hembras	70	80
Machos	30	20
Total	100	100

Si se capturan al azar 20 conejos de la especie A ¿cuál es la probabilidad de que:

a. Sean 10 machos y 10 hembras.
b. Sean 5 ó más hembras.
c. Sean menos de 4 ó más de 6 hembras.
d. Sean meno de 7 hembras si se sabe que se han seleccionado entre más de 2 y menos de 10 hembras.
e. Si se capturan 20 conejos de la especie A y 15 de la especie B. ¿cuál es la probabilidad de que sean 8 ó más machos de la especie A y 8 ó menos machos de la especie B?

9. Se sabe que en una zona donde persiste si se capturan 3 mosquitos del género *Anopheles* la probabilidad que los tres estén infectados con parásitos maláricos es 1/27. Conociendo esto determine:

a. ¿Cuál es la probabilidad de que uno este infectado?
b. Si se capturan 4 mosquitos de ésta especie ¿cuál es la probabilidad de que al menos uno esté infectado?
c. Si se capturan 6 mosquitos ¿cuántos se espera que estén infectados y cuál sería su desviación típica?

10. En un estudio sobre la regulación de la tensión arterial las ratas a las cuales se le inyecta una droga les aumenta la presión sanguínea. Después de haber realizado esta experiencia muchas veces se sabe que antes que el experimento haya concluido muere el 30% de las ratas por causa de la droga. Se realiza un nuevo experimento con un lote de 5 ratas, X representa el número de ratas sobrevivientes, calcule:

a. La probabilidad para cada uno de los valores que puede tomar dicha variable aleatoria.
b. La probabilidad de que sobrevivan 3 ó más ratas.
c. La probabilidad de que mueran entre 1 y 4 ratas (inclusive).

11. Un Biólogo esta interesado en los venados de la especie *Rabitus blancus*. El investigador después que capturaba un venado lo identificaba y liberaba inmediatamente. Mediante este procedimiento logró examinar 1000 venados en una región con una superficie de

20.000 Ha., y encontró que solamente el 0,01% de los venados era de la especie anterior. El investigador quiere saber:

a. ¿Cuál es la variable aleatoria de este experimento?
b. ¿Cuáles son las características del modelo probabilístico que sigue la variable aleatoria en este experimento?
c. ¿Cuál es el rango espacial de la experiencia?
d. Probabilidad de encontrar mas de dos venados en un área 5000 Ha.:
e. Probabilidad de que el número de venados sea menor o mayor a $\mu \pm \sigma$.

12. La población de valores de una variable aleatoria que sigue el modelo binomial presenta la distribución siguiente:

x	**f**
0	1
1	2
2	3
3	4
4	5
5	4
6	3
7	2
8	1

a. ¿Cuál será la probabilidad de que la variable esté incluida en el intervalo $\mu \pm \sigma$?
b. ¿Cuál será la frecuencia que predice el modelo binomial para x = 5?

13. En cierta población la incidencia de gripe es de 80 por cada 200.000 personas. Si en esa población se examinan 5000 personas, ¿cuál es la probabilidad de que se encuentre:

a. ¿Más de una persona con gripe?
b. ¿Menos de 4 ó más de 5 persona con gripe?

14. Un distribuidor vende semillas de tulipán rojo en paquetes de 1000 Y sabe, por experiencias anteriores, que el 1% de un gran número de ellas no será de la variedad que se desea. ¿Cuál es la probabilidad de que un paquete dado contenga más del 1% de semillas de otra clase?

15. Los registros de pluviometría en una zona caracterizada por la presencia de extensas selvas nubladas indican que el promedio de precipitaciones en 24 horas es de 36 mm. Calcule la probabilidad de que para una hora de cualquier día en particular, la precipitación sea de:

a. Mayor a 4 mm.
b. Igual o menor a 2 mm.
c. Exactamente 3 mm o más de 6 mm.

16. En un estudio sobre la ecología de un insecto que es plaga de un cultivo se sabe que las ninfas de insecto se distribuye al azar en las hojas de la planta, siguiendo el modelo de Poisson. A fin de determinar la distribución de frecuencia del número de insectos presentes se tomó una muestra de 96 hojas contándose en cada una de ellas el número de ninfas presentes. Los resultados obtenidos fueron los siguientes:

N° insectos/hoja	0	1	2	3	4	5	6	7	8	> 9
Frecuencia (f)	7	13	22	20	14	10	6	3	1	0

a. Calcule la probabilidad *a posteriori* de cada valor de X.
b. Calcule el valor de probabilidad esperado según el modelo de Poisson para cada valor de X.

17. En cierta ciudad se sabe que una persona por cada 10.000 habitantes muere de cáncer. Calcule, ¿cuál es la probabilidad de que en las próximas 38000 muertes que ocurran
a. Exactamente 7 personas mueran de cáncer.
b. Entre 5 Y 10 personas (ambas inclusive) mueran de cáncer.
c. Más de 6 personas mueran de cáncer.
d. Mueran menos de 7 personas de cáncer si se saben que han muerto por esta causa entre 3 y 12 personas.

18. En un estudio de una muestra de sangre se observa al microscopio la distribución de los hematíes en los cuadritos de una retícula siendo la superficie de cada cuadrito igual a 20 milimicras cuadradas $(m\mu^2)$. Si el número promedio de hematíes es de 5 células por cuadrito. Calcule la probabilidad de encontrar en 100 $m\mu^2$,
a. Exactamente 25 hematíes.
b. Entre 18 y 30 hematíes.
c. Menos de 20 ó más de 30 hematíes.

19. Se sabe que de las papas almacenadas en un Silo cada 30 días se pudre un promedio igual al 15%. Si el silo contiene 2000 papas, calcule la probabilidad que en un día cualquiera:
a. Se pudran 6 ó mas papas.
b. Se pudran 4 ó 7 papas.

20. Para que el silo de la pregunta anterior se considere que está funcionando correctamente, el número de papas podridas cada día no debe exceder a μ. Si lo excede el sistema de conservación está fallando. ¿Cuál es la probabilidad de que el sistema falle?

21. En un estudio sobre una población de murciélagos se determinó que por término medio mueren 2 individuos por día. Calcule la probabilidad de que:
a. En una semana mueran 10 o menos individuos.
b. En tres días mueran exactamente 5 individuos
c. Mañana mueran como máximo 4 individuos.
d. Entre el sábado y el martes, ambos inclusive, mueran al menos seis individuos.

22. El número promedio de glóbulos rojos en un volumen determinado de suero sanguíneo es de 9 células/ml para personas normales. En base al planteamiento anterior responda lo siguiente:
    a. ¿Cuál es la variable aleatoria (X) en éste experimento?
    b. ¿Que tipo de modelo probabilística sigue X?
    c. ¿Diga las razones por las cuales escogió el modelo probabilístico anterior?
    d. ¿Cual es el rango espacial de la experiencia?
    e. Determine la probabilidad de que el número de glóbulos rojos para una persona cualquiera sea mayor o menor a dos desviaciones estándar del promedio.

23. Sea X:N(0;1), determine el valor de "a" cuando:
    a. $P(0 \leq X \leq a) = 0{,}3531$
    b. $P(a \leq X \leq 0) = 0{,}2673$
    c. $P(-a \leq X \leq a) = 0{,}8132$
    d. $P(|X| \leq a) = 0{,}7850$

24. Sea X una variable que se distribuye normalmente con media igual a 0 y desviación igual a 1. Explique sin cálculos si la $P(0 \leq X \leq 2)$ es igual, mayor o menor a $P(-1 \leq X \leq 1)$.

25. Sea X:N(5;2). Represente gráficamente las probabilidades siguientes:
    a. $P(-s \leq X \leq +s)$
    b. $P(-2s \leq X \leq 2s)$
    c. $P(X > \mu + s)$
    d. $P(X \leq \mu - s)$

26. Sea X una variable aleatoria que se distribuye normalmente con media igual 20 y desviación igual a 5. Determine la probabilidad de que un valor elegido en forma aleatoria:
    a. Sea mayor a 22
    b. Se encuentre entre 19 y 23
    c. Sea menor 18,2 o mayor a 21,8.

27. La altura de ciertas plantas se distribuye normalmente con media igual a 31,9 cm y desviación igual a 0,45 cm. Calcule la probabilidad que una planta elegida al azar:
    a. Sea menor a 31 cm.
    b. Este comprendida entre 31,5 y 32,5 cm.
    c. Sea mayor a 32,8 cm.

28. La concentración de plomo en la sangre de cierto grupo de individuos expuestos a este elemento se distribuye normalmente con $\mu = 0{,}25$ y $\sigma = 0{,}11$. Una concentración superior o igual a 0,6 ppm se considera extremadamente alta ¿Cuál es la probabilidad de que un individuo seleccionado aleatoriamente este incluido en ésta categoría?

29. Se sabe que la duración en días de un medicamento se distribuye normalmente con una vida media útil de 1860 días y una desviación de 90 días. Si se lanza al mercado un lote de éste medicamento con la misma fecha de expedición. Qué tanto por ciento del medicamento se espera que dure
    a. Más de 2000 días.
    b. Menos de 1800 días.
    c. Entre 1800 y 1900 días

30. Si la concentración de colesterol en la sangre de los 4578 adultos que viven en cierta región se distribuye normalmente con $\mu = 217,6$ mg/100 ml y $s = 32,32$ mg/100 ml ¿Cuántos individuos tendrán un valor de colesterol superior a los 220 mg/100 ml?

31. Un productor sabe que el peso de los frutos que produce se distribuye normalmente con media igual a 14,5 g y desviación igual a 0,5 g. En la última cosecha recolectó 10.000 frutos y los quiere repartir en seis categorías diferentes: P1, P2, P3, P4, P5 y P6, donde cada categoría tiene una amplitud de 0,5 g. Es decir que en P1 se encuentran los frutos que pesan entre 13,0 y 13,5 g, en P2 los que pesan entre 13,5 y 14 g. Y así hasta P6 donde están incluidos los que pesan entre 15,5 y 16 g. ¿Cuántos frutos le corresponden a cada categoría?

32. Si la cantidad de radiación que puede ser absorbida por un individuo antes que le sobrevenga la muerte es una variable que se distribuye normalmente con una media igual a 500 roentgen y una desviación de 150 roentgen ¿Sobre cual nivel de radiación sobrevivirá solamente el 5% de los expuestos?

33. El contenido promedio de ácido acetilsalicílico es de 0,5 g/tableta de aspirina. Si la concentración de este producto se distribuye normalmente con una desviación de 0,05 g. Determine:
    a. Que % de las tabletas producidas contendrá más de 0,575 g. del ácido.
    b. Cuantas tabletas tendrán un contenido del ácido que varíe entre 0,45 y 0,55 g/tableta en un lote de 5750 pastillas.
    c. Cuantas tabletas tendrán un contenido superior a 0,58 g/tableta en las próximas 1000 tabletas que se produzcan.
    d. Para cual contenido de ácido se obtiene el 25% de pastillas con el valor más bajo de ácido.

34. Al clasificar cierto tipo de semillas cuyo peso se distribuye normalmente se encontró que un 20% son pequeñas, 55% son medianas, 15% son grandes y 10% son extra-grandes. Si el peso medio de las semillas es de 4,83 g y su desviación es de 1,2 g. ¿Cuáles son los pesos límites de las semillas medianas?

35. En cierta población humana, la anchura del cráneo (X) se distribuye normalmente. Sí, entre los individuos de esa población hay 58% de dolicocéfalos ($X \leq 75$); 38 % de mesocéfalos ($75 \leq X \leq 80$) y 4% de braquicéfalos ($X \geq 80$). Determine el valor de la media y la desviación de la anchura del cráneo.

36. Suponiendo que la duración de cierto tipo de instrumentos electrónicos es una variable $X:N(\mu;\sigma^2)$ y se tiene que elegir entre los instrumentos electrónicos del tipo A y B, los cuales tienen una duración cuyas distribuciones son Xa:N(44;36) y Xb:N(45;9) respectivamente. ¿Cuál debe preferirse para usarlo durante un período mínimo de 45 horas? ¿Cuál debe preferirse para usarlo durante un período mínimo de 48 horas?

37. Una bureta automática esta ajustada para servir un promedio de 200 ml de una solución de HCl dentro de recipientes especiales. Si la cantidad de ácido servido es una variable aleatoria que se distribuye normalmente con una desviación estándar de 15 ml, responda
    a. ¿Que fracción de los recipientes contendrá mas de 224 ml de ácido?
    b. ¿Cual es la probabilidad de que un tubo de ensayo contenga entre 191 y 209 ml?
    c. ¿Cuantos recipientes se derramaran si los siguientes 1000 recipientes a llenar tienen una capacidad de 230 ml?
    d. ¿Por debajo de que valor se obtiene el 25% de recipientes con el menor volumen de ácido?

38. La concentración de glóbulos blancos en cierta población humana se distribuye normalmente con una media igual a 190 células/μl y una desviación estándar de 48 céluas/μl, se quiere conocer lo siguiente:
    a. ¿Cuál es la probabilidad de que al elegir aleatoriamente una persona de esa población la concentración de glóbulos blancos sea igual a 190 células/μl?
    b. ¿Cuál es la concentración máxima de glóbulos blancos que puede tener el 39% de las personas con valores más bajos de esta variable?
    c. Para un estudio se han seleccionado aleatoriamente 9 personas de esa población.
        c1. ¿Cuál es la probabilidad de que en total posean más de 1725 células/μl?
        c2. ¿Que número esperado de personas tienen una concentración mayor a 200 células/μl?