CAPÍTULO 5

DISTRIBUCIONES DE PROBABILIDADES

En este capítulo se introducirá el concepto de **variable aleatoria**, cuya importancia radica en introducir modelos matemáticos en el cálculo de probabilidades. Luego, se considerarán las distribuciones de probabilidades de variables aleatorias discretas con su media y varianza respectiva. Existe un gran número de distribuciones discretas, pero en este texto sólo se discutirá en detalle la distribución binomial. Debido a que este texto no requiere un curso previo de Cálculo diferencial e integral, el estudio de las variables aleatorias continuas es omitido. Solamente se considera en el texto el estudio de la distribución Normal que es de crucial importancia para el proceso de Inferencia Estadística.

5.1 Variables Aleatorias

Una variable aleatoria es aquella que asume sus valores de acuerdo a los resultados de un experimento aleatorio. Usualmente se representa por las últimas letras del alfabeto: X, Y o Z.
Propiamente una variable aleatoria X es una función cuyo dominio es la colección de eventos del espacio muestral S y cuyo rango R_x, es un subconjunto de los números reales.

Algunos ejemplos de variables aleatorias son:

X: La suma que aparece al lanzar un par de dados.
Y: El número de caras que aparecen al lanzar una moneda tres veces.
Z: El número de errores que se encuentran en la página de un libro.

Ejemplo 5.1 De una caja que contiene 5 bolas numeradas del 1 al 5 se extraen 3 bolas una por una y sin reposición. Entonces **X**: El mayor de los tres números sacados, es una variable aleatoria.
Aqui el espacio muestral es:

S = {(1,2,3), (1,2,4), (1,2,5), (1,3,4), (1,3,5), (1,4,5), (2,3,4), (2,3,5), (2,4,5), (3,4,5)}

y la variable aleatoria X asume los valores: 3, 4 y 5. Por ejemplo, $X(2,3,4) = 4$.

El objetivo de la variable aleatoria es introducir notación matemática en el cálculo de probabilidades, la cual es mucho más simple y breve. Por ejemplo, en lugar de usar la frase "la probabilidad de que el mayor de los 3 números extraidos sea 4", se escribe simplemente como "P(X = 4)".
Por otro lado,

P(X = 4) = P(w están en S, tal que X(w) = 4)
= P({(1,2,4), (1,3,4), (2,3,4)}) = 3/10

Si el rango de valores R_x de la variable aleatoria X es finito o infinito enumerable entonces se dice que es una **variable aleatoria discreta**. Si su rango de valores R_x es infinito no enumerable entonces se dice que es una **variable aleatoria continua**.

5.1.1. Función de probabilidad de una variable aleatoria discreta

Si X es una variable aleatoria discreta con rango de valores R_x entonces, su función de probabilidad se define por:

$$p(x) = P[X = x], \text{ para todo } x \in R_x$$

y tiene las siguientes propiedades:

i) $p(x) > 0$ y
ii) $\Sigma\, p(x) = 1$.

Cuando R_x no contiene muchos valores es más conveniente expresar p(x) en una tabla de valores, la cual es llamada tabla de función de probabilidad.

Ejemplo 5.2 Hallar la función de probabilidad de la variable del ejemplo anterior

Solución:
Expresando p(x) en una tabla de valores se tiene que:

X	p(x)
3	1/10
4	3/10
5	6/10

Ejemplo 5.3. Se lanza una par de dados legales y distinguibles entre si. Hallar la función de probabilidad de **X**: la suma de los dos dados.

Solución:
Expresando p(x) en una tabla de valores y observando el espacio muestral del experimento se tiene que:

X	2	3	4	5	6	7	8	9	10	11	12
P(x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Ejemplo 5.4. De un lote que contiene 10 articulos, de los cuales 4 son dañados se extraen al azar y sin reposición 3. Se define la variable X: Número de artículos dañados que hay en la muestra. Hallar la función de probabilidad de X.

Solución: En este caso el rango de valores de X es $R_x = \{0, 1, 2, 3\}$ y en particular p(2) = Prob(sacar 2 dañados) = $\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}$, y en general p(x) = $\frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}$, para x = 0,1,2,3.

Calculando las combinaciones se obtiene la siguiente tabla de función de probabilidad:

X	p(x)
0	1/6
1	1/2
2	3/10
3	1/30

5.1.2. Función de distribución acumulativa

Sea X una variable aleatoria discreta con función de probabilidad p(x) y rango de valores R_x, entonces su función de distribución acumulativa se define por:

$$F(t) = P(X \leq t) = \sum_{x \leq t} p(x)$$

t es cualquier número real. En particular, si t es un valor que está en R_x , el cual consiste de enteros no negativos, entonces:

$$F(t) = p(0) + p(1) + p(2) + p(3) + \ldots + p(t)$$

Ejemplo 5.5. Hallar la función de distribución acumulativa para el Ejemplo anterior.

Solución:

X	p(x)	F(x)
0	1/6	1/6
1	½	4/6
2	3/10	29/30
3	1/30	1

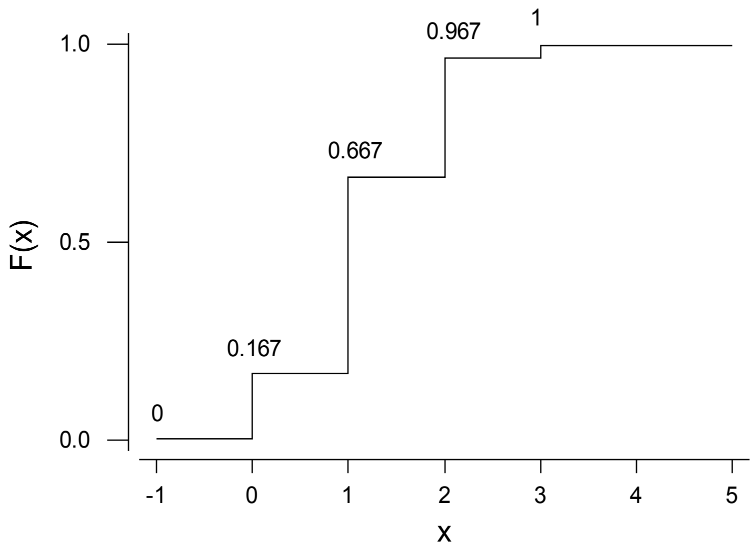
La gráfica de una función de distribución acumulativa es creciente y del tipo escalonado, con saltos en los puntos que están en el rango de valores y cuya magnitud es igual al valor de la función de probabilidad en dicho punto. Más formalmente tiene la siguiente propiedad:

Propiedad. *La relación entre la función de distribución de probabilidad y la función de distribución acumulativa está dada por:*

$$p(x) = F(x) - F(x-1)$$

para todo valor de x en el rango de valores de la variable aleatoria.

En la siguiente Figura se muestra la función de distribución acumulativa para el ejemplo anterior.



Ejemplo 5.6. Una variable aleatoria X tiene función de distribución acumulativa dada por la siguiente tabla de valores:

X	F(x)
3	1/10
4	4/10
5	1

a) Hallar la probabilidad de que x sea menor o igual que 3.
b) Hallar la probabilidad de que x sea mayor o igual que 5.
c) Hallar la probabilidad de que x sea igual a 5.

Solución:

a) $P(X \leq 3) = F(3) = 1/10$.

b) $P(X \geq 5) = 1- P(X \leq 4) = 1-F(4) = 1-4/10 = 6/10$.

c) $p(4) = F(4) - F(3) = 4/110 = 1/10 = 3/10$.

5.1.3 Valor Esperado y Varianza de una Variable Aleatoria Discreta

Sea X una variable aleatoria discreta con función de probabilidad $p(x)$ y rango de valores R_x, entonces su Valor Esperado o Media se define como el número:

$$\mu = E(X) = \sum_x xp(x)$$

La suma es sobre todos los valores x que están en R_x.

Ejemplo 5.7. Hallar el valor esperado de la suma obtenida al lanzar un par de dados.

Solución.

X	2	3	4	5	6	7	8	9	10	11	12
p(x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
xp(x)	2/36	6/36	12/36	20/36	30/36	42/36	40/36	36/36	30/36	22/36	12/36

La suma de la fila $xp(x)$ es 252/36 = 7. O sea que el valor esperado es 7.

Ejemplo 5.8. Hallar el valor esperado del número de articulos dañados que hay en la muestra de tamaño 3 extraida de un lote que contiene 10 artículos de los cuales, 4 son dañados.

Solución:

x	p(x)	xp(x)
0	1/6	0
1	1/2	1/2
2	3/10	6/10
3	1/30	3/30

Sumando la última columna se obtiene que μ = 12/10 = 1.2 articulos dañados. O sea, se espera que en la muestra hayan 1.2 artículos dañados. No tiene mucho sentido la interpretación directa del número, pero equivale a decir que si se extraen 10 muestras independientes de tamaño 3, en promedio deben salir un total de 12 artículos dañados.

Ejemplo 5.9. Un juego consiste en acertar un número del 1 al 1000. A la persona que acierta el número se le da un premio de 500 dólares y a las dos personas que tienen el número que le antecede o precede se le dan 100 dólares. Si el boleto cuesta 1 dólar. ¿Cuál será la Ganancia Neta esperada de una persona que compra un boleto?

Solución:
La Ganancia Neta es igual a la ganancia por el premio recibido menos el costo del boleto. Sea G la ganancia por el premio recibido. Hallaremos primero la Ganancia Esperada:

G	P(G)	Gp(G)

500	1/1000	500/1000
100	2/1000	200/1000
0	997/1000	0

Luego, la ganancia esperada por boleto será 700/1000 = 0.70. Así que la Ganancia Neta esperada será 0.70 - 1.00 = -0.30. Lo que significa que una persona pierde 30 centavos por cada boleto que compra. O dicho de otra manera, la empresa que administra el juego gana 30 centavos por cada boleto que vende.

La **Varianza** de una variable aleatoria discreta x con función de probabilidad p(x) y media μ se define por:

$$\sigma^2 = \sum (x-\mu)^2 p(x),$$

Donde la suma es sobre todos los valores del rango de X.

Para calcular la varianza, es más conveniente construir una tabla de la siguiente manera:

X	$p(x)$	$xp(x)$	$(x-\mu)^2$	$(x-\mu)^2p(x)$

La varianza será la suma de la última columna.

Ejemplo 5.10. Hallar la varianza del número de artículos dañados del Ejemplo 5.8.

Solución:

x	p(x)	xp(x)	$(x-\mu)^2$	$(x-\mu)^2p(x)$
0	1/6	0	1.44	.24
1	½	.5	. 04	.02
2	3/10	.6	.64	.192
3	1/30	.1	3.64	.121

Luego la varianza será $\sigma^2 = 0.573$.

Otra forma alterna para calcular la varianza es

$$\sigma^2 = \sum x^2 p(x) - \mu^2$$

La raíz cuadrada positiva de la varianza es llamada la **desviación estándar** y es más conveniente porque está en la misma escala de valores de la variable.

5.2 La Distribución Binomial.

Un experimento es llamado de Bernoulli si satisface las siguientes características:

a) En cada repetición puede ocurrir sólo una de dos maneras, una de ellas es llamada *Exito* y la otra *Fracaso*.

b) La probabilidad de *Exito*, representada por p, debe permanecer constante cuando el experimento es repetido muchas veces.
c) Las repeticiones de los experimentos deben ser independientes entre sí.

Ejemplo 5.11. Los siguientes son experimentos de Bernoulli

a) Observar las veces que sale 6 al lanzar varias veces un dado, en este caso la probabilidad de éxito es 1/6.
b) Contar el número de pacientes que sobreviven a una operación de corazón abierto.
c) Contar el número de personas que se entrevistan por un empleo y a las que se le hace una oferta de empleo.

Una variable aleatoria X tiene una **distribución Binomial** con parámetros n y p si se define como el número de éxitos que ocurren cuando un experimento de Bernoulli se repite n veces en forma independiente.

Ejemplo 5.12. Las siguientes son variables aleatorias binomiales.

a) Número de veces que resulta suma 7 al lanzar un par de dados 10 veces es una variable binomial con parametros $p = 1/6$ y $n = 10$.
b) Número de preguntas bien contestadas en un examen de 10 preguntas de selección múltiple, donde cada una tiene 4 alternativas de las cuales una es la correcta. En este caso $n = 10$ y $p = ¼ = 0.25$.
c) Número de artículos dañados que hay en una muestra de tamaño 3 extraida CON REPOSICIÓN de un lote que contiene 10 artículos, de los cuales 4 son dañados. En este caso $n = 3$ y $p = 4/10$.

La función de probabilidad de una binomial es de la forma:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

para $x = 0, 1, \ldots, n$.

El valor de $p(x)$ para diversos valores de n y p aparece en tablas de todo texto básico de Estadística.
Se puede mostrar que el valor esperado de una Binomial es $\mu = np$ y que la varianza es $\sigma^2 = npq$. Las demostraciones de estas propiedades pueden ser encontradas en cualquier texto de Estadística Matemática.

En **MINITAB** se pueden calcular la función de probabilidad (***Probability***), la función de distribución acumalada (***Cumulative probability***) y los percentiles ***(Inverse cumulative probability***) de la distribución Binomial para cualquier valor de n y p. Para esto hay que seguir la secuencia **Calc Probability Distributions Binomial**.

Ejemplo 5.13. Haciendo uso de **MINITAB**

a) Expresar en una tabla de valores la función de probabilidad y la función de distribución acumulada de la variable aleatoria X: Número de preguntas bien contestadas por un estudiante que responde al azar un examen tipo selección múltiple que consiste de 10 preguntas, cada una con 4 alternativas de las cuales sólo una es correcta.

b) Usar la tabla anterior para calcular la probabilidad de que el estudiante:
   - i) Tenga exactamente 3 preguntas buenas.
   - ii) Tenga 6 ó menos preguntas buenas.
   - iii) Tenga por lo menos 4 buenas.

Solución:

a) Primero hay que poner en una columna, llamada '*x*', todos los valores posibles de la variable. La ventana de diálogo para el cálculo de la probabilidad acumulada (similar es para calcular la probabilidad) y los resultados son como sigue:

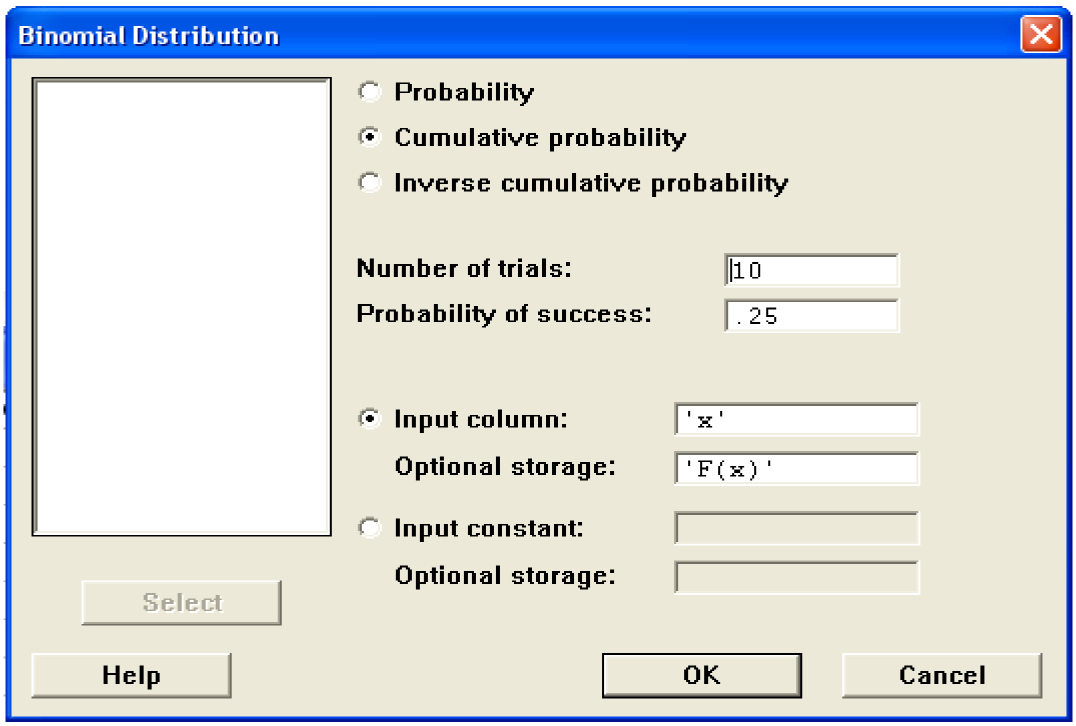


Figura 5.1. Ventana de diálogo para calcular probabilidades acumuladas de una distribución Binomial.

En la ventana **session** se presentarán los siguientes resultados:

```
Data Display
 Row     x       P(x)      F(x)

   1     0    0.056314  0.05631
   2     1    0.187712  0.24403
   3     2    0.281568  0.52559
   4     3    0.250282  0.77588
   5     4    0.145998  0.92187
   6     5    0.058399  0.98027
   7     6    0.016222  0.99649
   8     8    0.000386  0.99997
   9     9    0.000029  1.00000
  10    10    0.000001  1.00000
```

b) La probabilidad de tener 3 preguntas bien contestadas es P(3) = 0.2502, la probabilidad de tener 6 o menos preguntas bien contestadas es F(6) = 0.9964, la probabilidad de tener por lo menos 4 buenas es por complemento $P(X \geq 4) = 1 - P(X \leq 3) = 1 - F(3) = 1 - 0.77588 = 0.23412$.

También se puede hallar la probabilidad o la probabilidad acumulada para un número dado de éxitos. Para esto en ***Input constant*** se pone el número de éxitos.

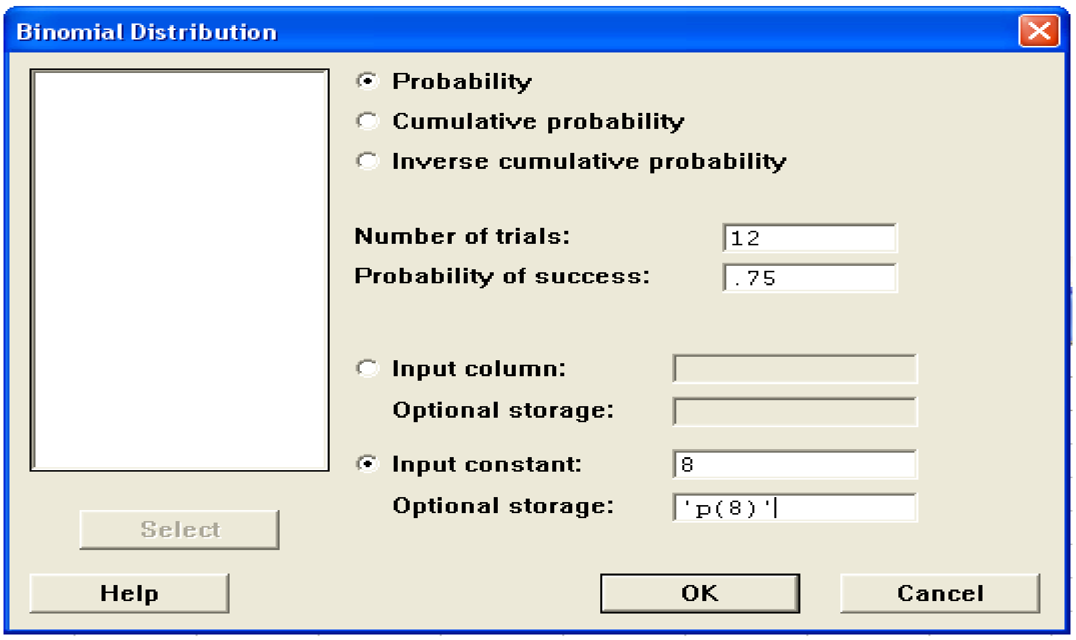


Figura 5.2. Ventana de diálogo para calcular probabilidades de una distribución Binomial.

Ejemplo 5.14. La prueba ELISA es usada para detectar la presencia de anticuerpos al virus del SIDA. ELISA, detecta que hay anticuerpos presentes en el 97 por ciento de los casos de que la muestra de sangre está contaminada con el virus del SIDA. Suponga que entre las muchas muestras que pasan por un Banco de Sangre hay 12 que están contaminadas con SIDA.

a) ¿Cuál es la probabilidad de que ELISA detecte 9 de estos casos?
b) ¿Cuál es la probabilidad de que ELISA detecte por lo menos 2 de estos casos?
c) ¿Cuál es la probabilidad de que por lo menos 4 casos no sean detectados por ELISA?

Solución:
Sea **X**: número de casos detectados por ELISA en la muestra de 12.

X es una Binomial con $n = 12$ y $p = .97$

a) Es igual a $p(9)$. Haciendo uso de MINITAB con **input constant** igual a 9, se obtiene $p(9) = .0045$.
b) Es igual a $P(X \geq 2) = 1 - P(X \leq 1) = 1 - F(1) = 1 - .0000 = 1.000$

c) Si por lo menos 4 no son detectados, significa que A LO MÁS 8 son detectados, o sea P $(X \leq 8) = F(8) = 0.0003$.

También se puede resolver como P(Y ≥ 4), donde Y representa el número de casos No detectados por ELISA, o sea, es una binomial con $p = .03$. Por complemento P (Y ≥ 4) = 1-(P≤3) = 1-F(3) = 1-.9997 = .0003.

Ejemplo 5.15. El Departamento de Salud ha determinado que el 10% de los puertorriqueños son zurdos. Se elige al azar 9 estudiantes de una escuela en Puerto Rico. ¿Cuál es la probabilidad de que:

a) Exactamente 2 de ellos sean zurdos?
b) Exactamente 6 de ellos sean diestros?
c) Por lo menos 4 de ellos sean diestros?

Solución:
Sea X: número de zurdos en la muestra de 9 estudiantes. X es una binomial con $p = .10$ y $n = 9$.

a) $p(2) = .1722$
b) Si hay 6 diestros entonces 3 son zurdos. Luego, la probabilidad pedida es $p(3) = .0446$
c) Si hay por lo menos 4 derechos, significa que hay a lo más 5 zurdos. Luego, la probabilidad pedida es P (X ≤ 5) = F(5) = .9999. También puede ser resuelto cambiando la probabilidad de éxito a $p = .90$ y hallando P (X ≥ 4) = 1 – P (X ≤ 3) = 1 – F (3) = 1 - .0001 = .9999.

Por otro lado, dada una probabilidad, **MINITAB** produce los valores de la variable que tienen una probabilidad acumulada lo más cercano posible a dicha probabilidad, esto es posible si se selecciona **Inverse cumulative probability** en la ventana de diálogo.

5.3 La Distribución Normal

La distribución Normal, también llamada Distribución Gaussiana en honor a K. Gauss, es una del tipo continuo y es considerada la distribución más importante en Estadística por las numerosas aplicaciones que tiene. Su comportamiento es reflejado por la Curva Normal que es la gráfica de la siguiente ecuación

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

Donde la media μ y la desviación estándar σ son los parámetros de la distribución. En la Figura 5.3 se muestra una curva normal con media $\mu = 15$ y desviación estándar $\sigma = 3$.

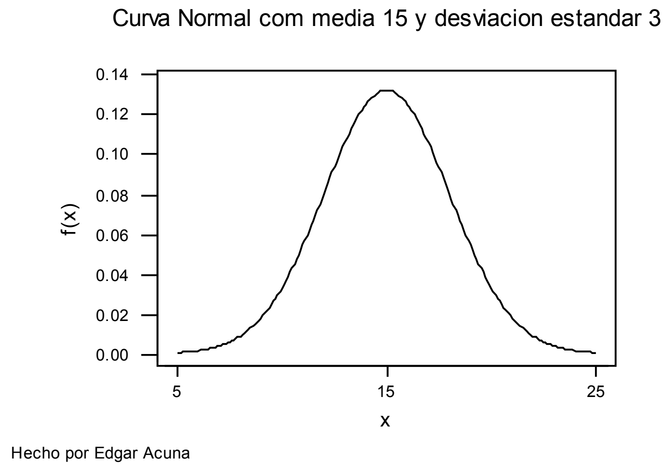


Figura 5.3. Gráfica de una curva normal con media 15 y desviación estándar 3.

Si una variable aleatoria X tiene una distribución Normal y queremos calcular la probabilidad de que X caiga entre dos valores a y b entonces, debemos hallar el área debajo de la curva entre a y b, esto se puede hacer por un proceso de Cálculo llamado Integración. Debido a que μ puede asumir cualquier valor real y que σ puede asumir cualquier valor real positivo habría que hacer un proceso de integración en cada caso, lo cual complicaría el proceso de calcular la probabilidad en lugar de simplificarlo. Afortunadamente se puede mostrar que cualquier normal puede ser transformada en una que tiene media 0 y desviación estandar 1 y la cual es llamada la **Distribución Normal Estándar** y se representa por Z. En el apéndice A de este texto se ha incluido una tabla que da el área debajo de la curva normal estándar a la izquierda de un valor de Z.

En **MINITAB** se pueden calcular la función de densidad (***Probability density***), la función de distribución acumalada (***Cumulative probability***) y los percentiles ***(Inverse cumulative probability***) de la distribución Normal para cualquier valor de la media μ y desviación estándar σ. No se requiere transformación a una normal estándar. Para esto hay que seguir la secuencia **Calc Probability Distributions Normal**.

Ejemplo 5.16. En este ejemplo en la columna llamada Z se han puesto 15 valores y se quiere hallar el área a la derecha de dichos valores. Las áreas serán guardadas en una columna llamada ***Area***. Por otro lado en la columna *alpha* se han puesto 11 valores de área y se desea hallar los valores de z correspondientes, estos son llamados **percentiles.** La ventana de diálogo y los resultados son como sigue:

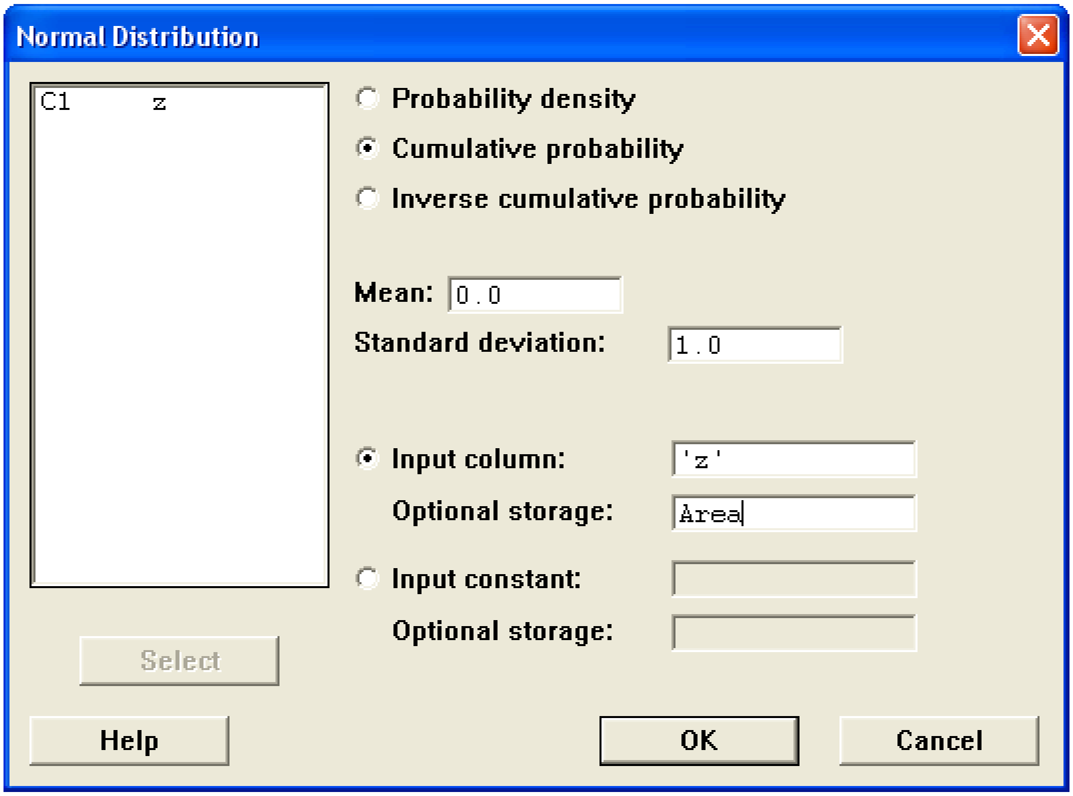


Figura 5.4. Ventana de diálogo para calcular areas debajo de una curva normal.

Data Display

Row	z	Area	alpha	z(alpha)
1	-3.00	0.001350	0.010	-2.32635
2	-2.57	0.005085	0.050	-1.64485
3	-2.23	0.012874	0.150	-1.03643
4	-2.00	0.022750	0.250	-0.67449
5	-1.64	0.050503	0.300	-0.52440
6	-1.00	0.158655	0.500	0.00000
7	-0.73	0.232695	0.800	0.84162
8	0.00	0.500000	0.900	1.28155
9	0.63	0.735653	0.950	1.64485
10	1.96	0.975002	0.975	1.95996
11	2.33	0.990097	0.995	2.57583
12	2.54	0.994457		
13	2.97	0.998511		
14	3.33	0.999566		
15	3.67	0.999879		

Para hallar los percentiles se elige ***Inverse cumulative probability*** y se escribe *alpha* en ***input column*** y *z(alpha)* en ***Optional storage***

El percentil del 90 por ciento será 1.28155 y el percentil del 25 por ciento será -.67449.

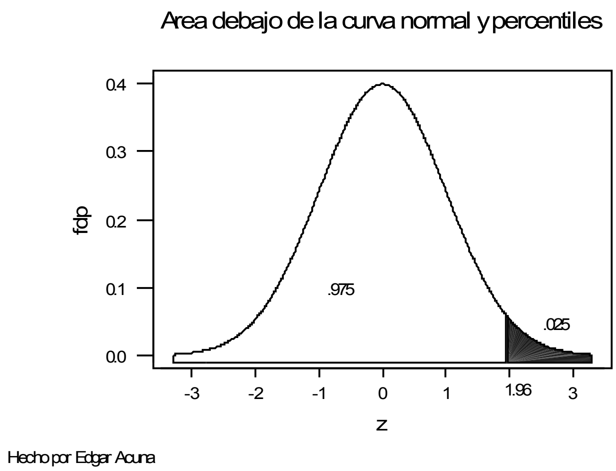


Figura 5.5. Areas debajo de una curva normal y percentil del 97.5%

En la gráfica se representa que el percentil del 97.5% es 1.96 y que el área que queda en el extremo derecho más alla de 1.96 es del 2.5%.

Estandarización de una Normal
Dada una variable aleatoria X distribuida Normalmente con media μ y desviación estándar σ entonces puede ser convertida a una normal estándar mediante el proceso de estandarización, definido por $Z = (X - \mu)/\sigma$*, donde X es* $N(\mu, \sigma^2)$.
Además si X_p y Z_p representen sus respectivos percentiles entonces:

$$X_p = \mu + \sigma Z_p$$

Fórmulas para calcular área debajo de la curva normal
En las siguientes fórmulas, F representa la distribución acumulada de la Normal, es decir el área acumulada a la izquierda del valor dado
a) P (X < a) = F(a) b) P (a < X < b) = F(b) - F(a) c) P (X > b) = 1 - F(b)

Ejemplo 5.17. Si X es una población Normal con media μ = 70 y σ = 10. Hallar las siguientes probabilidades:

a) P (X < 60)
b) P (X > 95)

c) $P(50 < X < 80)$

Solución:
Usando **MINITAB** con **mean** = 70 y **standard deviation** = 10, se tiene que:

a) $P(X < 60) = F(60) = .1587$

b) $P(X > 95) = 1 - F(95) = 1 - .9938 = .0062$

c) $P(50 < X < 80) = F(80) - F(50) = .8413 - .0228 = .8185$

Ejemplo 5.18. El Nivel de potasio presente en la sangre de una persona adulta se distribuye normalmente con media 3.8 y desviación estandar 0.2. Se elige al azar una persona:

a) ¿Cuál es la probabilidad de que el nivel de potasio de la persona sea mayor que 4.1?
b) Si el nivel de potasio es menor que 3.4 se dice que la persona sufre de hipocalcemia. ¿Cuál es la probabilidad de que una persona padezca de ésta enfermedad?
c) ¿Cuál es la probabilidad de que el nivel de potasio sea mayor que 3.25 pero menor que 3.75?
d) A las personas con el 15% más bajo de nivel de potasio se las someterá a una dieta para subirle el nivel. ¿Cuál debe ser el nivel de potasio requerido como máximo para ser sometido a la dieta?
e) A las personas con el 10% más alto de nivel de potasio se las someterá a una dieta para bajarles el nivel. ¿Cuál debe ser el nivel de potasio requerido como minimo para ser sometido a la dieta?

Solución:
Sea X: Nivel de potasio, X es normal con media 3.8 y desviación estándar 0.2

a) $P(X > 4.1) = 1 - F(4.1) = 1 - .9332 = .0668$.
b) $P(X < 3.4) = F(3.4) = .0228$.
c) $P(3.25 < X < 3.75) = F(3.75) - F(3.25) = .4013 - .0030 = .3983$.
d) Es equivalente a hallar el percentil del 15%. Usando **Inverse cumulative probability** en **MINITAB** se obtiene que 3.5927 debe ser el nivel de potasio requerido.
e) Es equivalente a hallar el percentil del (100-10)% = 90%. Usando **Inverse cumulative probability** en **MINITAB,** se obtiene que 4.0563 debe ser el nivel de potasio requerido.

Ejemplo 5.19. El tiempo que le toma a los estudiantes en ir de su casa a la Universidad se distribuye normalmente con media 20 minutos y desviación estándar 5.

a) ¿Cuál es la probabilidad de que a un estudiante le tome más de 18 minutos en llegar a la universidad?
b) ¿Cuál es la probabilidad de que un estudiante llegue a la universidad en menos de 30 minutos?
c) ¿A qué hora debe salir el estudiante de su casa si se desea que llegue tarde a su clase de la 8:00 a.m. solamente un 5 por ciento de las veces?

Solución

Sea la variable aleatoria X: El tiempo que le toma al estudiante en llegar de su casa a la Universidad, X es normal con media 20 y desviación estándar 5.

a) P (X > 18) = 1 – F (18) = 1 - .3446 = 6554.

b) P (X < 30) = .9772.

c) Equivale a hallar el percentil del 95%, y después restarle el tiempo hallado a las 8:00 am. Usando **Inverse cumulative probability** se obtiene que el percentil del 95 % es 28.2243. Luego el estudiante debe salir alrededor de 8.00 am.-28 minutos=7.32 am.

Ejemplo 5.20. Si la variable aleatoria X se distribuye normalmente con media μ y desviación estándar σ. Entonces hallar el valor *k* tal que

$$P\ (\ |X - \mu| < k\sigma) = .95$$

Solución:

Puesto que $|X-\mu|<k\sigma$ es equivalente a $\frac{|X-\mu|}{\sigma} < k$, por la fórmula de estándarización se obtiene que P(|Z| <k)=.95. Desdoblando el valor absoluto se obtiene que P(-k< Z <k)=.95. Por simetría de la distribución Normal el área que queda a la derecha del valor k es igual a 0.05/2 = 0.025. Es decir, $k = Z_{.975}$, Usando **MINITAB** o la tabla normal estándar del apéndice se obtiene k = 1.96.

5.4 Cotejando si hay Normalidad

Cuando se trata de sacar conclusiones acerca de la población usando los datos de la muestra, se asume generalmente que la los datos de la población se distribuyen de forma normal. Como no se conocen todos los elementos de la población, se deben usar los datos de la muestra para verificar si efectivamente la población es Normal. Existen varias pruebas estadisticas para verificar Normalidad.

En **MINITAB**, primero se elige la opción ***Basic Statistics*** de **Stat** y luego ***Normality Test*** del submenú que aparece.

En este texto nosotros sólo discutiremos la forma básica de detectar normalidad, la cual es a través del *plot de Normalida*. El *plot de Normalidad* consiste de un diagrama de puntos donde en el eje vertical se considera los escores normales y en el eje horizontal los valores de la variable. Si los puntos caen cerca de una línea, entonces se dice que hay **Normalidad**. En **MINITAB** este plot es obtenido siguiendo la secuencia **Graph Probability Plot**. En la ventana que aparece elegir la opcion ***Single*** como se muestra en la Figura 5.6

Ejemplo 5.21. Usar un plot de Normalidad para verificar si la siguiente muestra proviene de una población Normal

3.1 .9 2.8 4.3 .6 1.4 5.8 9.9 6.3 10.4 0 11.5

La ventana de diálogo se completará como se muestra en la Figura 5.7. En la opción **_Distribution.._** elegir normal y entrar los valores de la media y de la desviacion estandar correspondientes. Si estos valores no son entrados manualmente, MINITAB los estimará utilizando los datos.

MINITAB produce el plot que aparece en la Figura 5.8. En el eje horizontal aparecen los escores normales y en el eje vertical las probabilidades acumuladas de dichos escores.

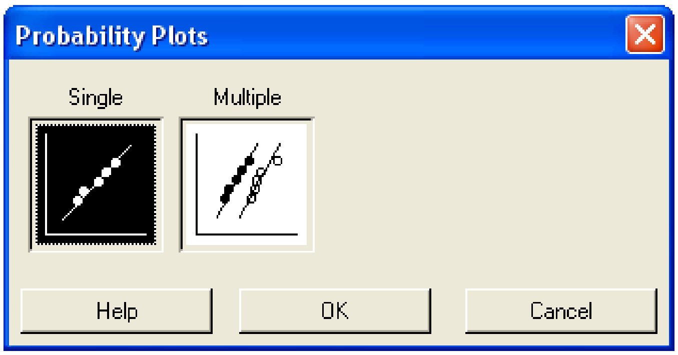


Figura 5.6. Ventana de dialogo de Probability Plots.

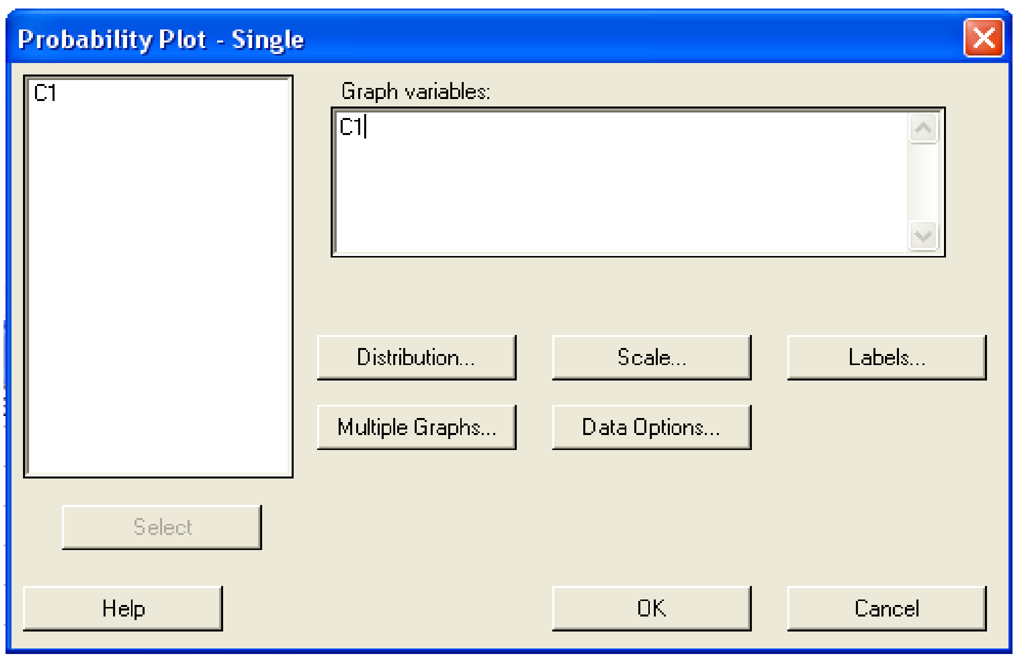


Figura 5.7 Ventana de diálogo de **Probability Plot - Single** para hacer un plot de Normalidad.

Interpretación: *Los puntos caen cerca de la linea y todos caen dentro de las bandas de confianza, luego se puede concluir que la población de donde proviene la muestra es Normal.*

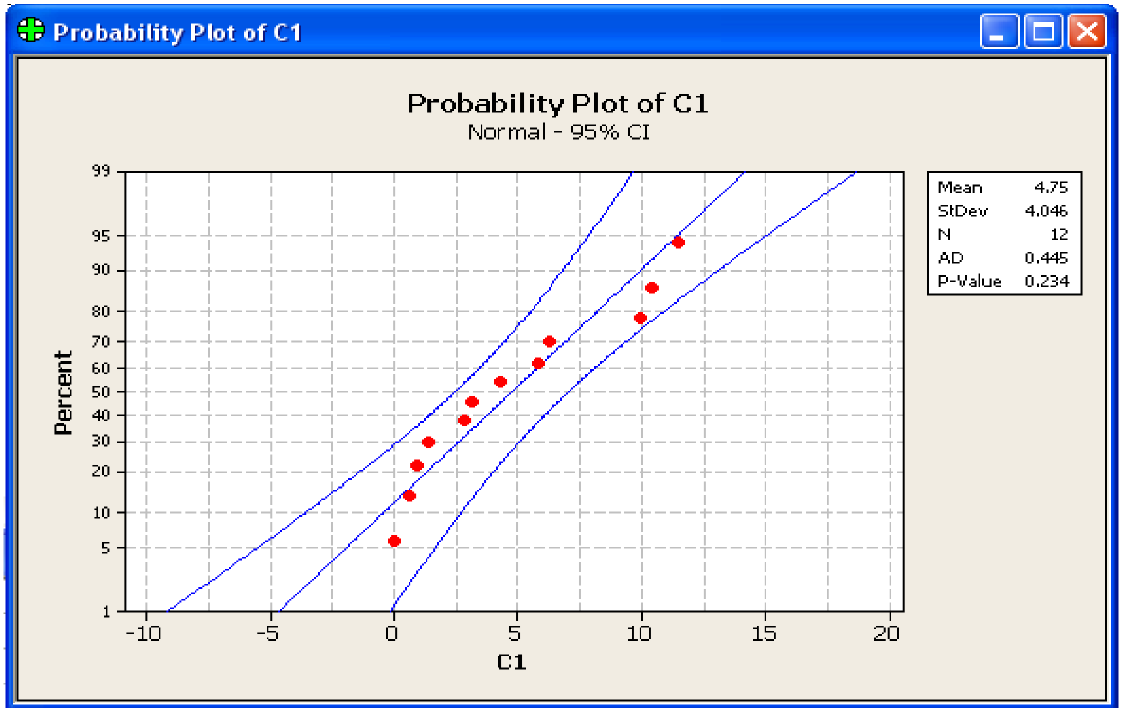


Figura 5.8. Plot de Normalidad para los datos del Ejemplo 5.21.

5.5 Simulando datos de una distribución conocida

Muchas veces se hace dificil conseguir datos reales para corroborar un método estadístico, una manera de resolver dicho problema es hacer que la computadora produzca mediante simulación dichos datos.

MINITAB tiene una lista grande de distribuciones conocidas, que pueden ser simuladas, esta lista se puede ver seleccionando **Random Data** en el menú **Calc**.

Ejemplo 5.22. Supongamos que deseamos simular 30 notas de una población normal que tiene media 70 y desviación estándar 10. La ventana de diálogo correpondiente será como sigue:

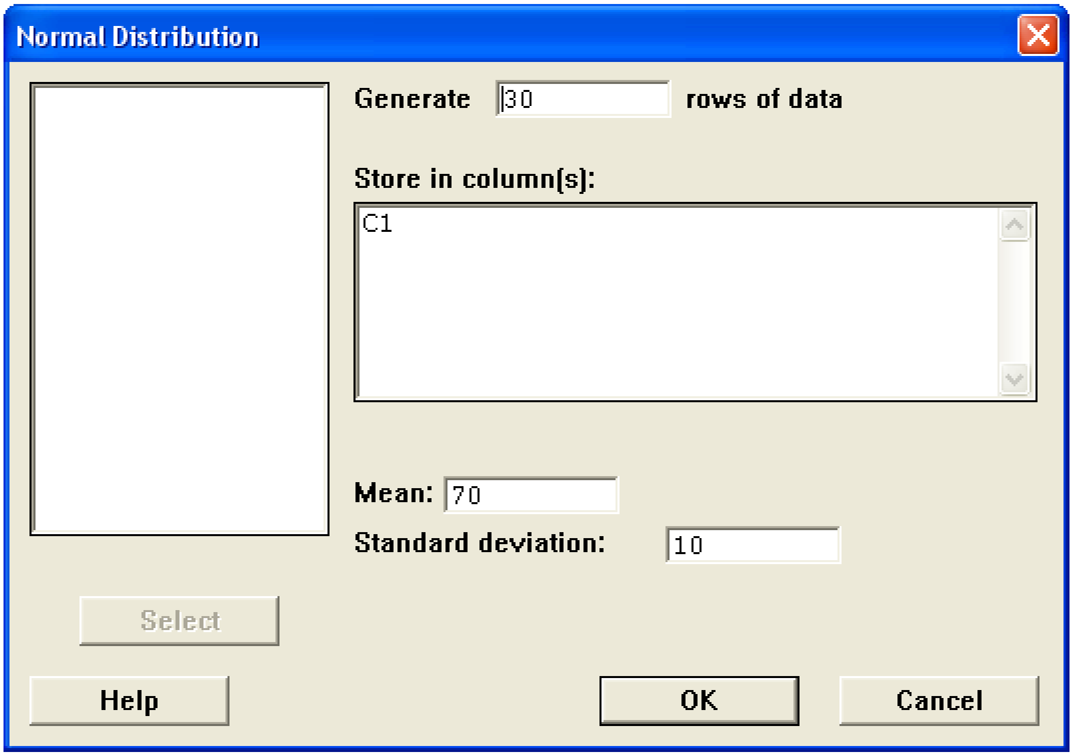


Figura 5.9. Ventana de diálogo para generar al azar una muestra de una población Normal.

Los datos aparecen con 4 decimales, pero si se elige la opción ***Format column*** del menú **Editor**, se puede definir que el número de decimales sean cero para que los datos salgan enteros, que es lo más común para notas. Los datos generados aparecen en la ventana **session** como sigue:

```
Data Display

C1
        80        80        77        75        54        69        53
        79        81        64        73        64        69        84
        60        95        71        63        58        65        79
        69        64        87        75        95        58        68
        63        81
```

EJERCICIOS

1. En una caja hay 5 fichas numeradas del 3 al 7. Se extraen al azar 3 de ellas a la vez. Hallar la función de probabilidad y el valor esperado de la variable aleatoria X: El menor de los números extraidos. (Por ejemplo si se extrajo la muestra 4, 3 y 6 entonces X=3).

2. De acuerdo a datos del gobierno, 30% de las mujeres que trabajan nunca han estado casadas, se elige al azar una muestra de 11 mujeres trabajadoras. ¿Cuál es la probabilidad de que:
   a) Exactamente 2 de ellas nunca hayan estado casadas?
   b) A lo más 3 de ellas nunca hayan estado casadas?
   c) Por lo menos 7 de ellas hayan estado casadas?

3. Un criminólogo afirma que el 80% de los condenados por "lavado de dinero" no vuelven a cometer un acto criminal por lo menos durante los primeros cinco años de ser liberados. Se elige al azar una muestra de 8 criminales que han sido liberados despues de estar encarcelados por "lavado" de dinero. ¿Cuál es la probabilidad de que:
   a) Ninguno de ellos comete crimen alguno por lo menos durante los cinco primeros años?
   b) Por lo menos 2 de ellos no cometan algún crimen por lo menos durante los cinco primeros años?
   c) No más de 3 de ellos cometan algún crimen por lo menos durante los primeros cinco años?

4. En un estudio clínico se determinó que 1 de cada 5 personas sufren de enfermedades mentales. Se seleccionaron al azar 30 personas:
   a) ¿Cuál es la probabilidad de que 7 de estas personas sufran de enfermedades mentales?
   b) ¿Cuál es la probabilidad de que al menos 8 de estas personas no sufran de enfermedades mentales?
   c) ¿Cuál es la probabilidad de que a lo más 6 sufran de enfermedades mentales?

5. Se ha encontrado que el 16% de los articulos producidos por una maquinaria tienen defectos. Un inspector de control de calidad selecciona 30 articulos aleatoriamente encuentre la probabilidad de que:
   a) 6 de los articulos seleccionados sean defectuosos .
   b) a lo más 10 de éstos articulos sean defectuosos.
   c) Al menos 15 de ellos no sean defectuosos.
   d) Al menos 6 de ellos pero, no más de 18 sean defectuosos.

6. Se estima que el 30% de los accidentes automovilisticos se debe a que el conductor está ebrio.
   a) Calcular en promedio cuántos accidentes se deberán al hecho de que el conductor esté ebrio en los siguientes 82 accidentes reportados.
   b) Calcular la desviación estandar del número medio de accidentes en los siguientes 82 accidentes reportados.

7. Una empresa tiene dos plantas de producción: A y B. En A se produce un 40% de la producción total y en B un 60%. Se sabe además que un 2% de la produccion de A y un 7% de la producción de B son defectuosas. Se elige al azar 12 articulos producidos por la empresa. ¿Cuál es la probabilidad de que:
   a) Solamente 3 salgan defectuosos?
   b) A lo mas 2 salgan defectuosos?
   c) Por lo menos 9 salgan buenos?

8. En el estudio Framingham acerca de factores que afectan las enfermedades cardíacas se hizo un seguimiento por un período de 16 años a una gran cantidad de hombres sanos. Se encontró que inicialmente la distribución de los niveles de colesterol de los hombres era Normal con media $\mu = 224$ y con desviación estándar $\sigma = 48$
   a) Una persona con un colesterol menor de 200 es considerada como una con bajo riesgo de tener complicaciones cardíacas. ¿Qué porcentaje de hombres tendrán bajo riesgo?
   b) Si el colesterol de la persona es mayor de 250 entonces tendrá problemas cardiacos en el futuro. ¿Qué porcentaje de hombres tendrán problemas cardiacos?
   c) Los hombres que tienen el 5% más alto de colesterol serán sometidos a una dieta, para bajarle su colesterol y evitar que tenga problemas cardiacos en el futuro. ¿Cuál será el nivel de colesterol máximo permitido para NO someterse a la dieta?

9. Un profesor considera que el tiempo que los estudiantes necesitan para terminar el examen se distribuye normalmente con media $\mu = 60$ minutos y desviación estándar $\sigma =$ 10 minutos.
   a) ¿Cuál es la probabilidad de que un estudiante demore más de una hora y 15 minutos en terminar el examen?
   b) ¿Cuál es la probabilidad de que un estudiante demore más de 45 minutos pero menos de 85 minutos en terminar el examen?
   c) Se elige al azar 8 estudiantes que cogieron el examen, ¿Cuál es la probabilidad que exactamente 5 de ellos tarden más de 40.4 minutos pero menos de 79.6 minutos en terminar el examen?

10. El contenido de las botella de jugo de naranja llenadas por una máquina automática tiene una distribución aproximadamente normal con media 63.9 onzas y desviación estándar de 0.25. Encontrar la probabilidad de que:
   a) Una botella contenga menos de 64 onzas de jugo de naranja.
   b) Una botella contenga al menos 63.75 onzas de jugo de naranja.

11. Un análisis realizado al contenido de grasa en jamones determina que en cada corte de 5 onzas de jamón se tiene en promedio 12.34 gramos de grasa si se asume que la cantidad de grasa tiene distribución normal con desviación estándar de 0.8 gramos.
   a) ¿Qué porcentaje de cortes de jamón de 5 onzas tiene un contenido de grasa entre 10.2 gramos y 12.5 gramos.
   b) ¿Qué porcentaje de cortes de jamón de 5 onzas tienen más de 14 gramos de grasa

12. Se sabe que X es una variable aleatoria con distribución normal y con media 72. Hallar la desviación estándar si en un 10% de las veces X tiene un valor mayor a 89.

13. Se estima que un conductor conduce un promedio de 12,400 millas al año, con una desviación estándar de 3800 millas. Calcular la probabilidad de que en el próximo año el conductor conduzca:

    a) Más 12,100 millas pero menos que 13,200 millas
    b) Más de 15,000 millas.