Apuntes de Probabilidad y Estadística para Ingeniería y Administración

Ignacio Vélez Pareja
Decano
Facultad de Ingeniería Industrial
Politécnico Grancolombiano
Bogotá, Colombia
Octubre, 2002

Probabilidad y estadística: Conceptos básicos

Conceptos básicos de probabilidad

"La necesidad de jugar es tan apremiante y su práctica tan placentera, que supongo que debe ser pecado".
Heywood Broun (Citado por Thomas y Ronald Wonnacott)
"Debemos creer en la suerte. Porque, ¿de qué otra manera se explica el éxito de las personas que no nos gustan?"
Jean Cocteau (Citado por Thomas y Ronald Wonnacott)

Cuando se toman decisiones sobre resultados futuros que se conocen, la única razón para que se cometa un error es que exista un error en el análisis por parte del decisor. Esta situación se conoce como certidumbre completa.

Pero la realidad casi nunca es totalmente predecible. Por lo tanto, aunque el decisor haya hecho el análisis correcto, siempre hay factores que no puede controlar y que influyen para que los resultados sean imprevistos. Cuando prevalecen estas condiciones se dice que se trabaja "bajo incertidumbre" y, por lo tanto, el decisor se ve obligado a asumir riesgos. Por ejemplo, que los resultados de sus decisiones no sean favorables. Una forma de hacerlo es medir el riesgo asociado a cada predicción; riesgo que significa qué tantas posibilidades hay de que la decisión adoptada sea errónea. Con esa información el decisor tomará la "mejor" determinación y sólo queda esperar para saber si el resultado es o no favorable. Por más cuidado que se tenga en el análisis, siempre existirá la posibilidad de que el resultado sea desfavorable.

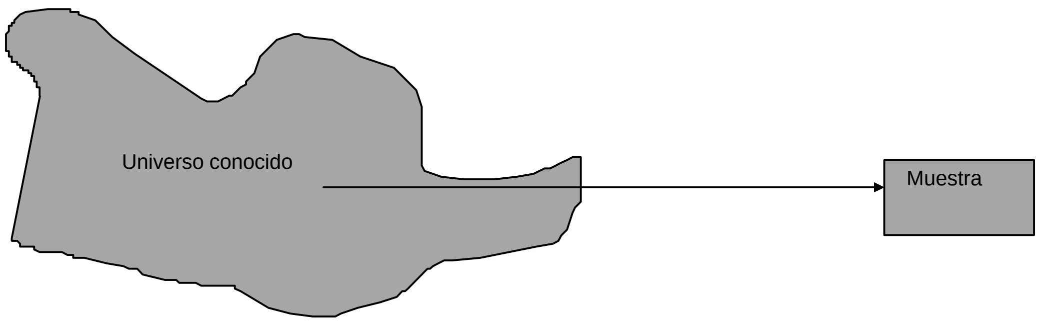
Hay que aceptar que este tipo de análisis es anticipado —"*a priori*"— y que se sabe que el resultado será único. Sin embargo, al hacer el análisis "*a priori*", hay que identificar el mayor número de resultados *posibles* y medir, para cada resultado, la probabilidad de que ocurra. Cuando se trabaja con decisiones bajo riesgo es necesario entonces introducir el concepto de probabilidad. Esta idea se utiliza en forma intuitiva y en el léxico corriente. Así, por ejemplo, se habla de la probabilidad de que llueva o de que un candidato gane cierta elección. Estas son probabilidades subjetivas que, ante escasez de información, son válidas.

Deducción e inducción

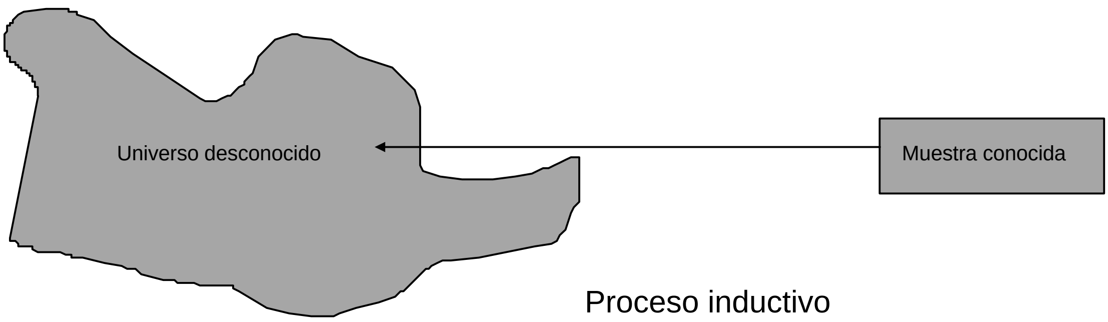
Al abordar el problema de la probabilidad y hacer análisis de tipo probabilístico, conviene distinguir entre un proceso deductivo y uno inductivo. Al estudiar la Teoría de la probabilidad y los rudimentos de la estadística, se está utilizando un proceso deductivo, esto es, que se parte de lo general para decir algo de lo particular. Por ejemplo, se conoce un universo de elementos y se desea saber cuál es el comportamiento de un grupo reducido de observaciones tomadas de ese universo.

En los estudios más avanzados de la Estadística se utiliza un método inductivo, esto es, que a partir de la información obtenida de unos pocos elementos de un universo —una muestra— se trata de encontrar las características de todo el universo. Una manera de recordar estas ideas es tener presente que la partícula *de* en *de*ducción indica hacia afuera y la partícula *in* que se encuentra en *in*ducción indica hacia adentro. Esto es que la deducción va del universo conocido (general) a la muestra desconocida (particular) —hacia afuera del universo— y la inducción va de la muestra conocida (particular) al universo desconocido (general) —hacia adentro del universo—.

Gráficamente:



Proceso deductivo



Proceso inductivo

Cuando se utiliza la inferencia estadística se está trabajando el concepto de inducción, o sea que de unas pocas observaciones se obtienen conclusiones sobre la totalidad del universo. El mejor ejemplo de este trabajo es el que hacen las firmas que realizan encuestas de opinión y las que miden el “rating” o la sintonía de los programas de televisión: A partir de 500 o 1.000 observaciones, hacen afirmaciones sobre el comportamiento de la totalidad de la población.

Probabilidad

Supóngase un experimento cualquiera, por ejemplo, el número dos en el lanzamiento de un dado. El conjunto de todos los resultados posibles se llama universo o espacio de la muestra, en este caso, los números de 1 a 6 en el lanzamiento del dado en cuestión.

Usualmente se utiliza el concepto de frecuencia para ilustrar el concepto de probabilidad. Supóngase que se estudian n resultados de un experimento, de los cuales m se consideran ocurrencias exitosas de un resultado deseado, E y $P(E)$ denota la probabilidad de ocurrencia de dicho resultado; la relación entre el número de resultados exitosos m y el número de resultados posibles n, es una medida aproximada de la probabilidad de ese resultado, es decir:

$$P(E)=\frac{m}{n} \tag{1}$$

Esto es rigurosamente cierto cuando n es muy grande. Más formalmente, se deberá escribir así:

$$P(E) = \lim_{n \to \infty} \frac{m}{n} \qquad (2)$$

Donde:

P(*E*): Probabilidad que el resultado *E* ocurra.

E: Resultado que interesa analizar.

M: Número de veces que ocurre *E*.

n: Número de veces que se ejecuta el experimento.

Por ejemplo, si se desea saber cuál es la probabilidad de ocurrencia de que aparezca el número 2 en la cara superior cuando se lanza un dado, se podrían hacer lanzamientos seguidos y anotar cuántas veces aparece cada número, en particular el 2. Si esto se repite varias veces, entonces la relación entre el número de veces que apareció el 2 y el número de lanzamientos será un estimativo de la probabilidad. Esta frecuencia relativa tiende a un número; en el caso de un dado que no esté cargado, esta frecuencia tiende a 1/6.

Una variable aleatoria está definida por una función que asigna un valor de dicha variable aleatoria a cada punto del universo. Por ejemplo, la variable aleatoria puede ser el valor que aparezca en la cara superior del dado, o el cuadrado de este valor, etc. En este ejemplo, *E=2*, *m* es el número de veces que aparece el número 2 y *n* es el número de lanzamientos.

Propiedades básicas de la probabilidad

A continuación se presentan algunas propiedades básicas de la probabilidad.

1) La probabilidad de un resultado del universo es una cantidad menor o igual que uno y mayor o igual que cero. Esto se explica porque la probabilidad está definida por la proporción entre un número de casos "exitosos" y el número total de casos. El número de casos "exitosos" es menor que el número total de casos.

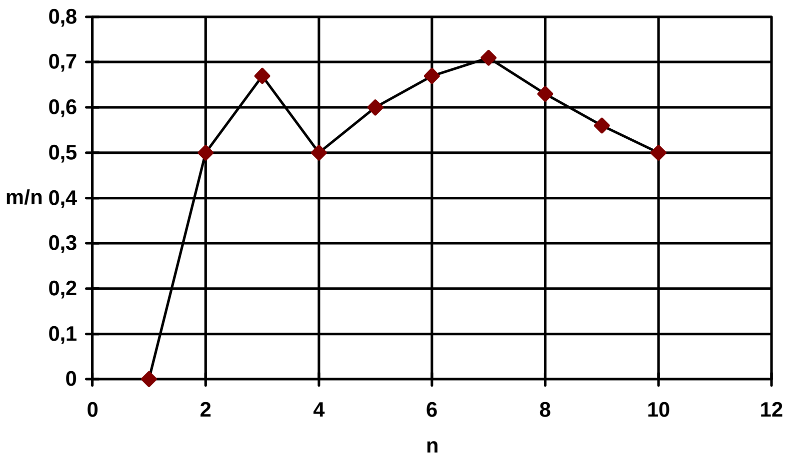
Ejercicio

Lanzar una moneda 50 veces. Construir y completar en la hoja de cálculo la siguiente tabla de ejemplo:

	A	**B**	**C**	**D**
1	Lanzamiento número (n)	Cara = 1 Sello = 0	Frecuencia acumulada de caras (m)	Frecuencia relativa de caras (m/n)
2	1	0	0	0,00
3	2	1	1	0,50
4	3	1	2	0,67
5	4	0	2	0,50
6	5	1	3	0,60
7	6	1	4	0,67
8	7	1	5	0,71
9	8	0	5	0,63
10	9	0	5	0,56
11	10	0	5	0,50

Construir una gráfica de los resultados con n en las abcisas y m/n en las ordenadas, como se ilustra a continuación.

FRECUENCIA RELATIVA ACUMULADA DEL LANZAMIENTO DE UNA MONEDA



2) La probabilidad de un resultado que no puede ocurrir, o sea que no pertenece al universo, es cero.

3) La probabilidad del universo es uno. Es decir, la probabilidad de que ocurra alguno de los resultados de todo el conjunto posible de ellos es $P(E_1+E_2+\ldots+E_m)$ y es igual a 1, donde $(E_1,E_2,\ldots,E_m)$, son todos los resultados posibles, mutuamente excluyentes y exhaustivos del universo.

Se dice que unos resultados son mutuamente excluyentes cuando la ocurrencia de cualquiera de ellos elimina la ocurrencia de cualquier otro.

Todos los resultados posibles m_i suman n, o sea:

$$m_1+m_2+m_3+\ldots m_k = n \tag{3}$$

Si esta ecuación se divide por n, entonces la suma de las frecuencias relativas es igual a 1.

$$m_1/n+m_2/n+m_3/n+\ldots m_k/n = n/n = 1 \tag{4}$$

Así pues en el límite:

$$P(E_1)+P(E_2)+P(E_3)+\ldots P(E_k) = 1 \tag{5}$$

4) Si E y F son resultados mutuamente excluyentes, o sea que sólo uno de ellos puede ocurrir, entonces la probabilidad de que ocurra E o F es $P(E+F) = P(E) + P(F)$. Nuevamente, en el lanzamiento de un dado de seis caras numeradas de 1 a 6, sólo un número aparecerá en la cara superior, por lo tanto, los resultados (E_2) y (E_6), o sea que aparezca 2 en un caso o que aparezca 6 en el otro, son resultados mutuamente excluyentes. La probabilidad de que ocurra E_2 o E_6 es de 1/6+1/6 o sea, 1/3.

5) Si *E* y *F* son resultados independientes, esto es, que la ocurrencia de uno no afecta la ocurrencia del otro, la probabilidad de que ocurran simultáneamente *P(EF),* es *P(E)* x *P(F).* Tomando como ejemplo el dado de seis caras, el hecho que en el primer lanzamiento del dado aparezca un 2, no influye para que en el segundo lanzamiento aparezca cierto número; los lanzamientos son resultados independientes. Entonces, la probabilidad de que en el primer lanzamiento aparezca un 2 y en el segundo aparezca un 6 será 1/6x1/6, o sea, 1/36.

NOTA

Obsérvese que cuando se trata de resultados mutuamente excluyentes y se desea saber la *probabilidad de que uno de los dos* ocurra, se expresa con frases ligadas por *o*; en el caso de resultados independientes y si se desea calcular la *probabilidad de que ambos* ocurran, las frases se ligan con *y*.

Estas propiedades son formales pero coinciden con las nociones intuitivas de probabilidad.

Eventos y sus probabilidades

En la realidad los hechos no son tan simples como en el ejemplo del dado. Ocurren combinaciones que complican un poco la situación. El cálculo de sus probabilidades es más complejo.

Eventos y sus probabilidades

La realidad es compleja y ocurren combinaciones de resultados; la combinación de varios resultados origina un evento. A través de un ejemplo se ilustrará esta idea.

Ejemplo

Supóngase que se desea analizar los resultados de una inversión $1.000 a tres años. El resultado de cada año es la ocurrencia de un ingreso por valor de $600 o $0. Los resultados posibles son:

$$
\begin{aligned}
(NNN) &= m_1\\
(NNS) &= m_2\\
(NSN) &= m_3\\
(NSS) &= m_4\\
(SNN) &= m_5\\
(SSN) &= m_6\\
(SNS) &= m_7\\
(SSS) &= m_8
\end{aligned}
$$

El orden de las letras se refiere año 1, 2 ó 3 y *S* indica si hay ingreso y *N* si no lo hay.

(NSS) significa un flujo de caja como este:

Año	Flujo de caja
0	-1.000
1	0
2	600
3	600

Y así para los demás casos. Las probabilidades de que el resultado sea cero son:

$$P(N)_1 = ,3$$
$$P(N)_2 = ,3$$
$$P(N)_3 = ,3$$

Se supone que los eventos son independientes entre sí. Esto significa que el resultado positivo de un año no influye en la probabilidad de que, en los años siguientes, el resultado sea también positivo. Esto en la realidad puede que no ocurra. Sin embargo, para efectos del análisis, se hará caso omiso de esta consideración.

Entonces, las probabilidades asociadas a cada resultado combinado son:

	A	**B**
1	Evento combinado	Probabilidad total
2	$P(N)_1$ =	30%
3	$P(N)_2$ =	30%
4	$P(N)_3$ =	30%
5	$(NNN) = m_1$	=B1*B2*B3 [0,027]
6	$(NNS) = m_2$	=B1*B2*(1-B3) [0,063]
7	$(NSN) = m_3$	=B1*(1-B2)*B3 [0,063]
8	$(NSS) = m_4$	=B1*(1-B2)*(1-B3) [0,147]
9	$(SNN) = m_5$	=(1-B1)*B2*B3 [0,063]
10	$(SSN) = m_6$	=(1-B1)*(1-B2)*B3 [0,147]
11	$(SNS) = m_7$	=(1-B1)*B2*(1-B3) [0,147]
12	$(SSS) = m_8$	=(1-B1)*(1-B2)*(1-B3)

			[0,343]

Estos resultados se denominarán puntos. Los eventos serán una combinación cualquiera de puntos. Así se puede pensar en el evento, “por lo menos un año con ingreso”, el cual incluiría los puntos m_2, m_3, m_4, m_5, m_6, m_7 y m_8, o en el evento “a lo sumo un año con ingreso cero”, el cual incluiría los puntos m_4, m_6, m_7 y m_8.

Si la probabilidad de que ocurra el ingreso es diferente a 70%, hay que introducir los valores adecuados en los cálculos.

La probabilidad de estos eventos será la suma de la probabilidad de los puntos. En el primer evento, la probabilidad será de:

$0{,}063+0{,}063+0{,}147+0{,}063+0{,}147+0{,}147+0{,}343 = 0{,}973$

en el segundo caso de:

$0{,}147+0{,}147+0{,}147+0{,}343 = 0{,}784$

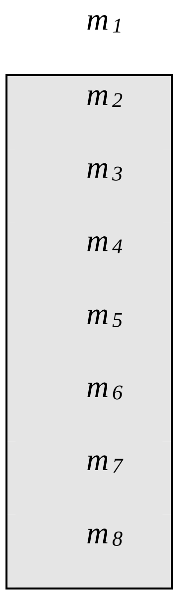
Diagramas de Venn

Los resultados y sus combinaciones en eventos se pueden visualizar en forma gráfica; estas gráficas se conocen como *diagramas de Venn* y se introducirán mediante la continuación del ejemplo anterior.

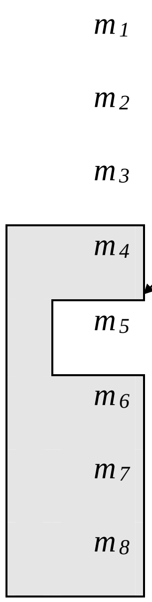
Ejemplo

Si se tienen los resultados *m*, entonces los eventos "por lo menos un año con ingreso", o "a lo sumo un año sin ingreso", se pueden ilustrar con una gráfica o *diagrama de Venn*.

Evento:
"por lo menos un año con ingreso"



Evento:
"máximo un año sin ingreso"



Combinación de eventos

Puede ser que se deseen evaluar combinaciones de eventos en términos probabilísticos. En ese caso se debe tener cuidado que no haya "duplicaciones" de puntos en los eventos combinados, puesto que al tener puntos "duplicados" las probabilidades se podrían doblar. Por ejemplo, que ocurrieran los eventos: "menos de dos años con ingreso" o "el primero sin ingreso seguido por dos años con ingresos".

Para facilitar el análisis se puede construir una tabla con algunos eventos posibles, así:

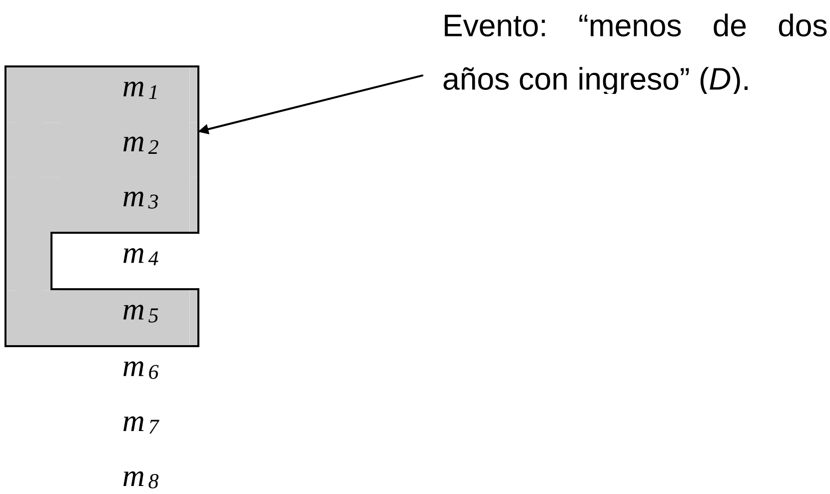
Evento	Descripción	Resultados que incluye	Probabilidad
A	Todos sin ingreso.	m_1	0,027
B	Todos con ingreso.	m_8	0,343
C	Primero sin ingreso, el resto con ingreso.	m_4	0,147
D	Menos de dos años con ingreso.	m_1, m_2, m_3, m_5	0,216
E	Un sólo año con ingreso.	m_2, m_3, m_5	0,189
F	Un sólo año sin ingreso.	m_6, m_7	0,294
G	Por lo menos dos años con ingreso.	m_4, m_6, m_7, m_8	0,784
H	Por lo menos un año con ingreso.	$m_2, m_3, m_4, m_5, m_6, m_7, m_8$	0,973

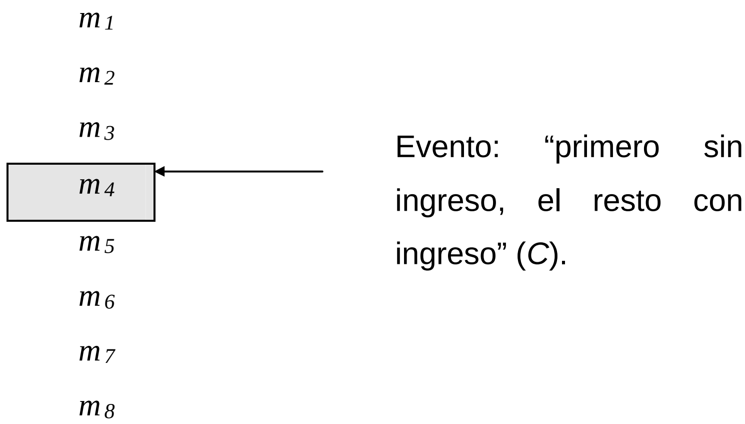
Completar esta tabla con diez eventos adicionales.

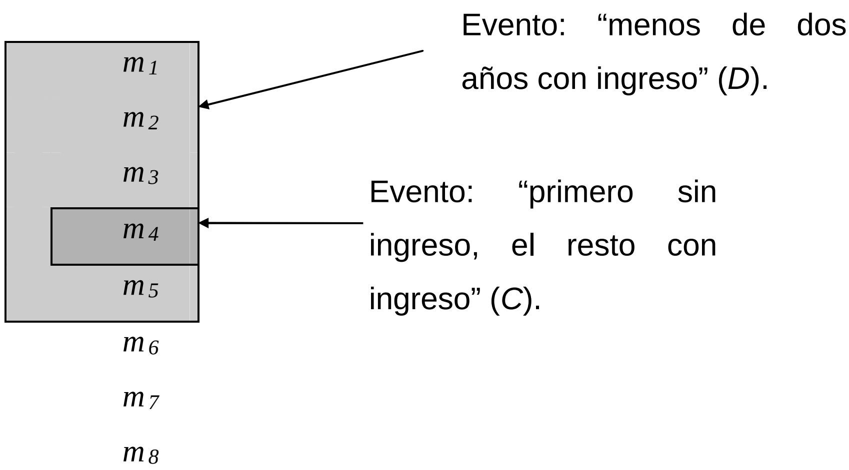
Los eventos mencionados "menos de dos años con ingreso" —evento *D*— y "el primero sin ingreso el resto con ingreso" —evento *C*— , contienen los puntos m_1, m_2, m_3, m_5 y m_4 respectivamente. Si se desea encontrar la probabilidad que de ocurra uno de los dos, entonces se dice que se halla la probabilidad de los eventos combinados *D* o *C*, en notación matemática, $P(D \cup C)$. Esta combinación de eventos contiene los puntos m_1, m_2, m_3, m_4 y m_5; su probabilidad es 0,363.

La notación $(D \cup C)$ quiere decir que el nuevo evento contiene los puntos que se encuentran en *D*, en *C* o en ambos. El nuevo evento se lee *D* “unión” *C*.

En diagrama de Venn:







Combinación de eventos *C* y *D* en $(D \cup C)$

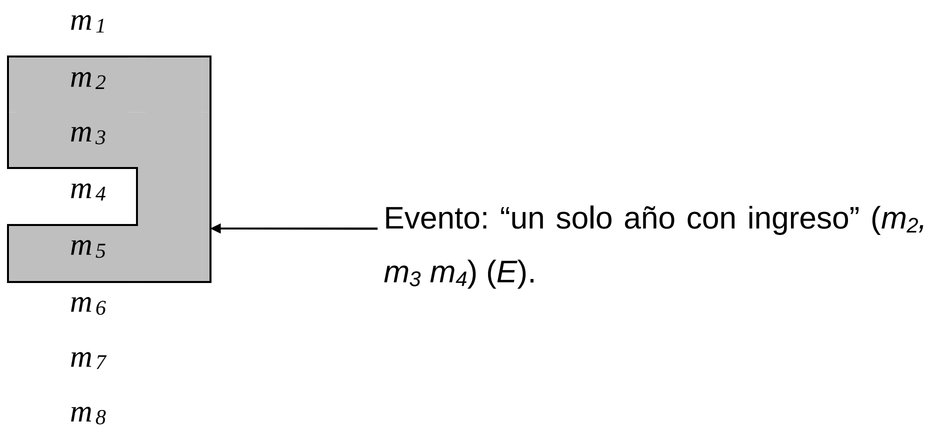
Si se estipulara que deben ocurrir ambos eventos, o sea que ocurriera el evento “menos de dos años con ingreso” —evento *D*— **y** además ocurriera el evento “el primero sin ingreso seguido por dos años con ingreso” —evento *C*—, se escribiría como $(D \cap C)$. La notación $(D \cap C)$ indica que el nuevo evento *D* y *C*, contiene sólo los elementos comunes a ambos. En este caso se ve claramente que no existen esos elementos comunes y se dice que es un conjunto *vacío.* Esto se puede verificar examinando los resultados m_i y el *diagrama de Venn*.

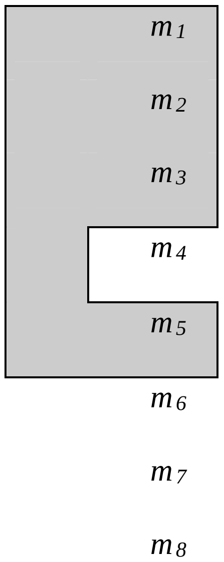
Por no existir ningún resultado posible común, su probabilidad es cero. El nuevo evento se lee *D* "intersección" *C*.

Si se consideraran los eventos *D* y *E*, "menos de dos años con ingreso" (m_1, m_2, m_3, m_5) y "un sólo año con ingreso" (m_2, m_3, m_5), se tendría lo siguiente:

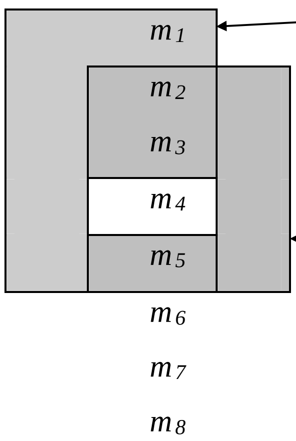
$$P(D \cup E) = P(m_1, m_2, m_3, m_5) = 0{,}027+0{,}063+0{,}063+0{,}063 = 0{,}216$$

En diagrama de Venn:





Evento: "menos de dos años con ingreso" (m_1, m_2, m_3, m_5) (D).

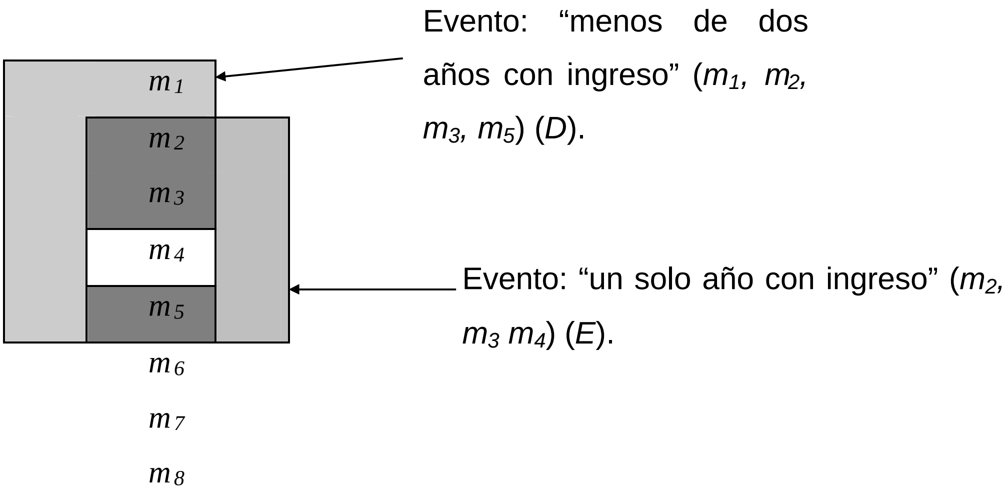


Evento: "menos de dos años con ingreso" (m_1, m_2, m_3, m_5) (D).

Evento: "un solo año con ingreso" (m_2, m_3 m_4) (E).

Combinación de eventos D o E en $(D \cup E)$

$$P(D \cap E) = P(m_2, m_3, m_5) = 0{,}063+0{,}063+0{,}063 = 0{,}189$$



Combinación de eventos D y E en $(D \cap E)$ (sombreado fuerte).

Obsérvese que $P(D \cup E) = P(m_1, m_2, m_3, m_5)$ no es igual a 0,500+0,375, o sea a la suma de las probabilidades de los dos eventos, D y E. Si se sumaran las dos probabilidades, se estaría contando dos veces la probabilidad asociada a los resultados comunes a los dos eventos. En forma matemática esta precaución se escribe así:

$$P(D \cup E) = P(D) + P(E) - P(D \cap E) \qquad (6)$$

Lo que en la práctica dice esta ecuación, es que se le elimina una vez la probabilidad asociada a los resultados que están en ambos eventos.

Cuando los eventos -*D* y *E*, por ejemplo- son mutuamente excluyentes, entonces sí se puede calcular la probabilidad de la unión de ellos como la suma:

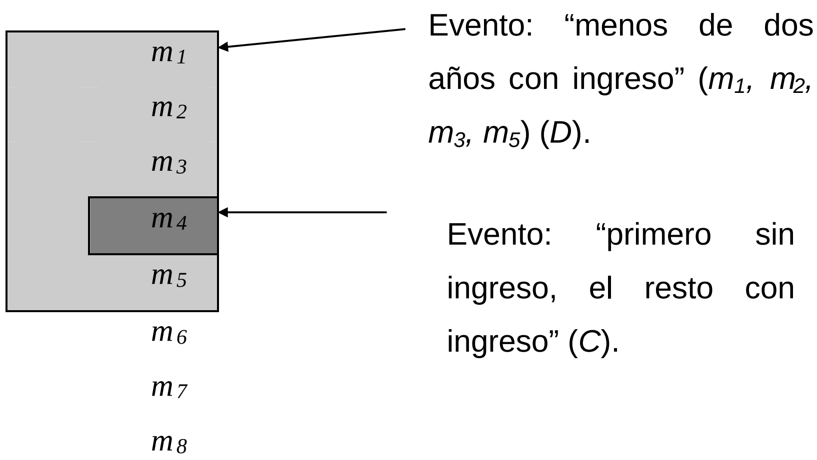
$$P(D \cup E) = P(D) + P(E) \tag{7}$$

Esto es fácil de deducir, ya que cuando dos eventos son excluyentes, la unión de ellos es cero, porque no existen resultados comunes a ambos.

Particiones y complementos

Cuando un grupo de eventos (colección de un cierto número de puntos o resultados posibles) es mutuamente excluyente, se dice que no existen resultados posibles comunes entre ellos, es decir que ningún resultado o punto pertenece a más de un evento. En el ejemplo de la inversión, se tiene que los eventos *C*, “menos de dos años con ingreso”, y *D* “primero sin ingreso y el resto con ingreso”, son mutuamente excluyentes; es imposible que esos dos eventos puedan darse al mismo tiempo.

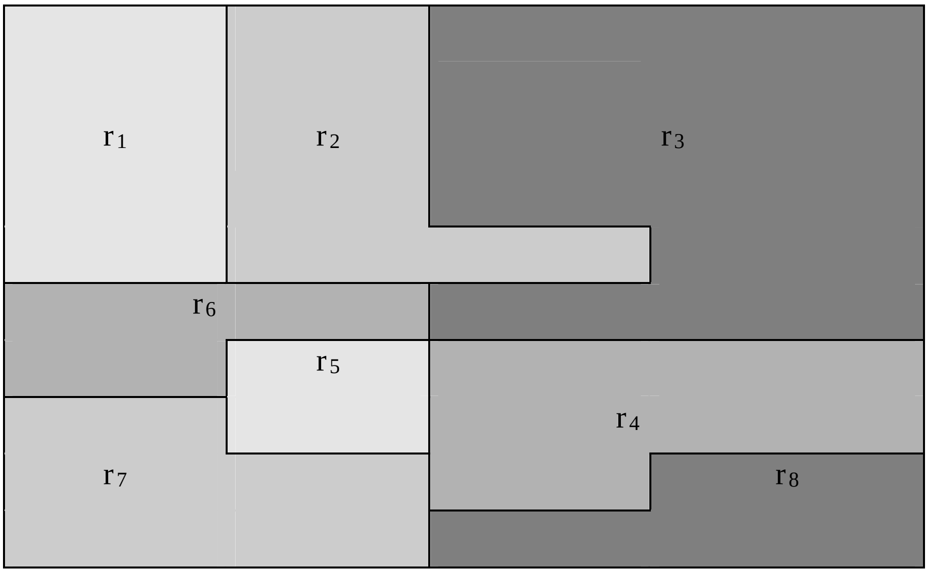
En diagrama de Venn:



Combinación de eventos mutuamente excluyentes, C y D

A partir de las posibilidades del ejemplo de la inversión a tres años, identificar tres eventos mutuamente excluyentes y construir los *diagramas de Venn* correspondientes, para verificar su resultado.

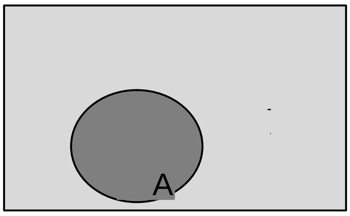
Cuando el conjunto de eventos mutuamente excluyentes es tal que cubre todos los resultados posibles del universo, entonces se les llama particiones. Esto significa que la unión de una colección de eventos mutuamente excluyentes y que conforman una partición es la totalidad del universo.



Partición de un universo

Por otro lado, cuando un evento contiene exactamente los puntos o resultados que no están en otro, y entre los dos se forma una partición, esto es, que entre los dos contienen todos los resultados posibles del universo, se dice que uno es el complemento del otro. Por ejemplo, el evento *A* “todos sin ingreso” y el evento *H* “por lo menos un año con ingreso”, son complementarios entre sí. Se dice que el evento *A* es el complemento del evento *H*. La notación es la siguiente: $\bar{A} = H$; lo cual quiere decir que *H* es el complemento de *A*.. Es fácil concluir lo siguiente:

$$P(A)+P(\overline{A})=1$$



Evento A y su complemento $(\overline{A})$

Probabilidad Condicional y Probabilidad Conjunta

Cuando se conoce que un evento ha ocurrido y se desea saber la probabilidad de que ocurra otro, "dado" que ocurrió el primero, se dice que se calcula la *probabilidad condicional*. O sea que se puede hablar de calcular la probabilidad condicional que ningún año tenga ingreso, dado que se sabe que en el primero no hubo.

Al observar la tabla de los ocho casos posibles, se concluye que al condicionar los eventos al hecho, que el primero no tenga ingreso, en la realidad se "recorta" el universo y por eso las probabilidades cambian. El nuevo universo, cuando se sabe que en el primer año no hubo ingresos queda así:

	A
13	Eventos condicionados a que en el año no hubo ingresos.
14	(*NNN*) = m_1
15	(*NNS*) = *m2*
16	(*NSN*) = *m3*
17	(*NSS*) = *m4*

Si alguien quisiera apostar al hecho de que en ningún año hubiera ingresos, tendría dos estimativos de la probabilidad diferentes: Antes de saber sobre el resultado de alguno de los años y después de tener alguna información sobre el primer año. Una persona razonable, por ejemplo, le asignará probabilidad cero a los resultados m_4, m_5, m_6 y m_7, después de saber que el primer año no tuvo ingresos, ya que esos resultados son imposibles, puesto que en ellos el primer año sí tiene ingresos; así las cosas, los eventos posibles son m_1, m_2, m_3 y m_4 y de estos cuatro eventos, sólo m_1 interesa.

	A	B
5	(*NNN*) = m_1	=B1*B2*B3 [0,027]
6	(*NNS*) = m_2	=B1*B2*(1-B3) [0,063]
7	(*NSN*) = m_3	=B1*(1-B2)*B3 [0,063]
8	(*NSS*) = m_4	=B1*(1-B2)*(1-B3) [0,147]
9	Total	0,30

Como los cuatro resultados tienen una probabilidad total de 0,30 y son los únicos posibles, se concluye que el evento m_1 tiene probabilidad 0,09, o sea 0,027/0,3. Se deduce fácilmente que la probabilidad ha sido "aumentada" y que este aumento debe ser proporcional en cada caso; esto es, que la probabilidad inicial de cada resultado debe ser dividida por la probabilidad total de los eventos que quedan al condicionar los eventos.

	A	**B**
13	Eventos condicionados a que el primero no tenga ingresos.	Probabilidad condicional que ninguno tenga ingreso, dado que el primero no lo tiene.
14	(*NNN*) = m_1	=B5/B18 [0,09]
15	(*NNS*) = m_2	0,21
16	(*NSN*) = m_3	0,21
17	(*NSS*) = m_4	0,49
18	Probabilidad que el primero no tenga ingresos.	=B5+B6+B7+B8 [0,300]

NOTA

La notación para la probabilidad condicional de un evento *A*, dado que ocurrió otro *B*, se indica así:

$$P(A/B)=\frac{P(A\cap B)}{P(B)} \tag{8}$$

Aquí se observa lo siguiente: El valor 0,027 de la celda *B5* es el valor de la probabilidad *P*("primer año cero"n "todos cero") y el denominador es la probabilidad *P*("primer año cero"), o sea que la probabilidad condicional es:

$$P(\text{"}todoscero\text{"}/\text{"}primerañocero\text{"}) = \frac{P(\text{"}primeraño\ cero\text{"} \cap \text{"}todos\ cero\text{"})}{P(\text{"}primer\ año\ cero\text{"})}$$

$$P(\text{"}todos\ cero\text{"}/\text{"}primer\ año\ cero\text{"}) = \frac{0{,}027}{0{,}3} = 0{,}09$$

La expresión de la probabilidad condicional puede ser escrita de otra forma -en “reversa”- y se conoce como la *regla multiplicativa*, así:

$$P(A \cap B) = P(B) \times P(A/B) \tag{9}$$

Análisis Bayesiano

La fortaleza de la teoría de la probabilidad es la de reforzar el buen juicio o criterio y la experiencia, nunca los reemplaza. De hecho, ningún método para tomar decisiones puede reemplazar ese criterio y experiencia. Los métodos para la toma de decisiones se enriquecen con la experiencia del decisor y lo ayudan a tomar mejores decisiones.

En los estudios de mercado se puede aplicar muy bien la idea del análisis bayesiano. Esto consiste en hacer inferencias sobre unas *causas*, a partir de los *efectos conocidos*; aquí se aplica la idea presentada arriba de la probabilidad condicional. Se calcula la *probabilidad a posteriori*, después de que se ha observado un efecto determinado. En otras palabras, el análisis bayesiano o estadística bayesiana se caracteriza por el hecho de *revisar* o *ajustar* una *probabilidad a priori* acerca de un determinado parámetro a una *probabilidad **a** posteriori* más confiable, basado en los resultados que ofrece la evidencia de una muestra o estudio adicional.

A través de dos ejemplos se ilustrará la idea del Análisis Bayesiano, para después introducir la presentación de tipo matemático y formal.

Ejemplo

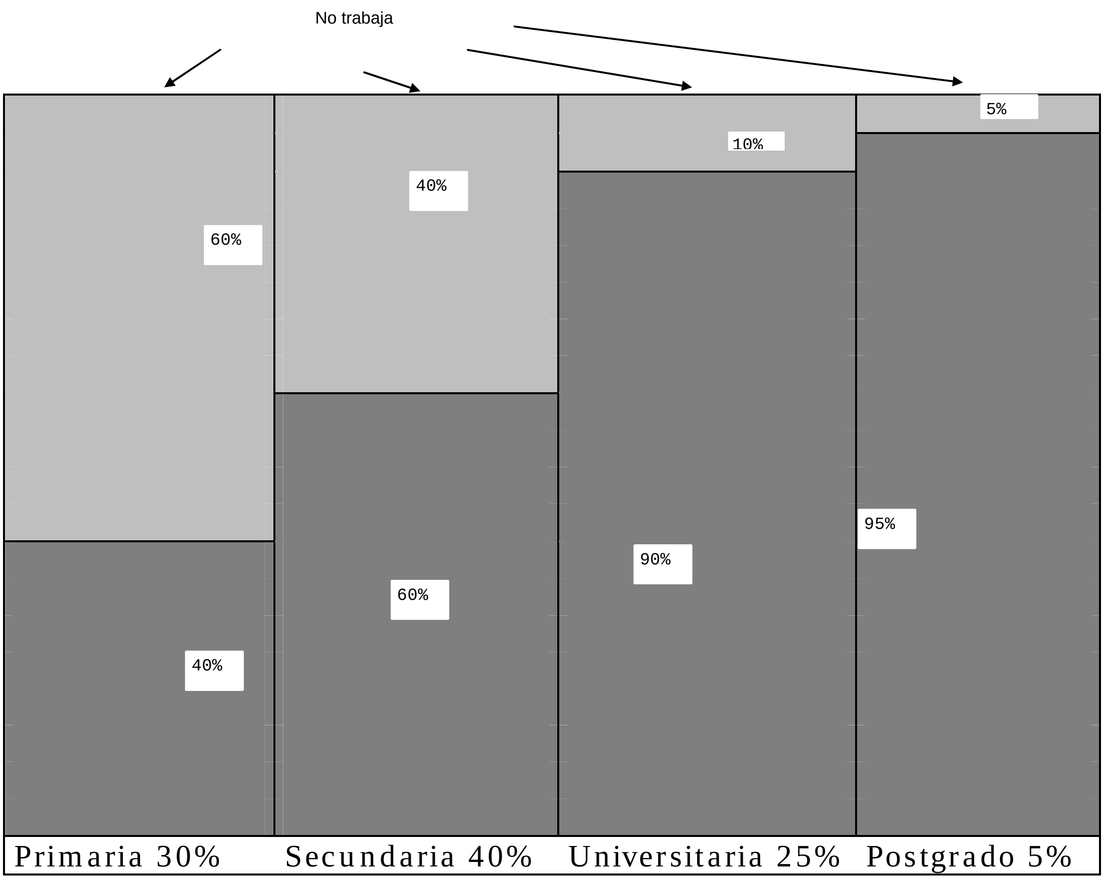
En un estudio de mercado se tienen los siguientes cálculos del mercado potencial para cierto producto:

Mercado objetivo: Amas de casa.

	A	**B**	**C**
1		Educación	Trabaja
2	Primaria	30%	40%
3	Secundaria	40%	60%
4	Universitara	25%	90%
5	Postgrado	5%	95%

Si se hiciera un muestreo de esta población y se escogiera una persona en forma aleatoria y se encuentra que está trabajando, ¿cuál es la probabilidad de que esta persona tuviera título de postgrado?

Una forma intuitiva de analizar esta situación es la de estimar qué porcentajes de esas amas de casa trabajan (Efecto (E)) en cada categoría de educación (Causa (C)). En forma gráfica se tiene:



Se podría esperar que de acuerdo con cada nivel de educación y desempleo, hubiera entonces el siguiente porcentaje total de gente trabajando:

	A	B	C	D
1		Educación	Trabaja	% Total trabajando
2	Primaria.	30%	40%	=B2*C2 [12%]
3	Secundaria.	40%	60%	=B3*C3 [24%]
4	Universitaria.	25%	90%	=B4*C4 [22,5%]
5	Postgrado.	5%	95%	=B5*C5 [4,75%]
6	Total.			=D2+D3+D4+D5 [63,25]
7	Probabilidad que entre las que estén trabajando, esa persona tenga título de Postgrado.			=D5/D6 [7,51%]

Ejemplo

Supóngase ahora que se tiene una inversión que tiene un beneficio neto presente (valor presente) de $7 millones si ocurre el evento *A1*, y de -$3 millones si ocurre el evento *A2*. Las probabilidades *a priori* son *P*(*A1*) = 0,3 y *P*(*A2*) = 0,7.

Esto significa que el valor esperado de la inversión es:

$$E(A) = \$7x0{,}3 - \$3x0{,}7 = 0 \text{ millones}$$

Si no se invierte el valor esperado también sería cero. Si se contrata un estudio adicional sobre este proyecto, los resultados sería que si el resultado del estudio fuera *B1*, entonces esto significaría que habría una ganancia *A1* de $7 millones, y si el resultado del estudio fuera *B2*, significaría que hay una pérdida *A2* de $3 millones. Si llegara a resultar *A1* —$7 millones— el estudio resultaría en *B1* con una probabilidad de 0,8, y si llegara a resultar *A2* - —$3 millones— el estudio resultaría en *B2* con probabilidad 0,6. En resumen se tiene:

Con esta información se puede mejorar el estimativo inicial que se tenía para la probabilidad de *A1* y de *A2*. El análisis sería el siguiente: Si se espera que *A1* ocurra el 30% de los casos y suponiendo que ello ocurriera, sólo el 65% de las veces, el estudio indicará que sí ocurrirá *B1*; así mismo, si se espera que *A2* ocurra el 70% de los casos, el estudio indicará que así ocurrirá -*B2*- el 60% de las veces. En términos de probabilidad condicional:

$P(A1) = 0,3 \quad P(A2) = 0,7$

$P(B1|A1)= 0,65$

$P(B2|A1) = 1- P(B1|A1) = 1-0,65 = 0,35$

$P(B2|A2) = 0,6$

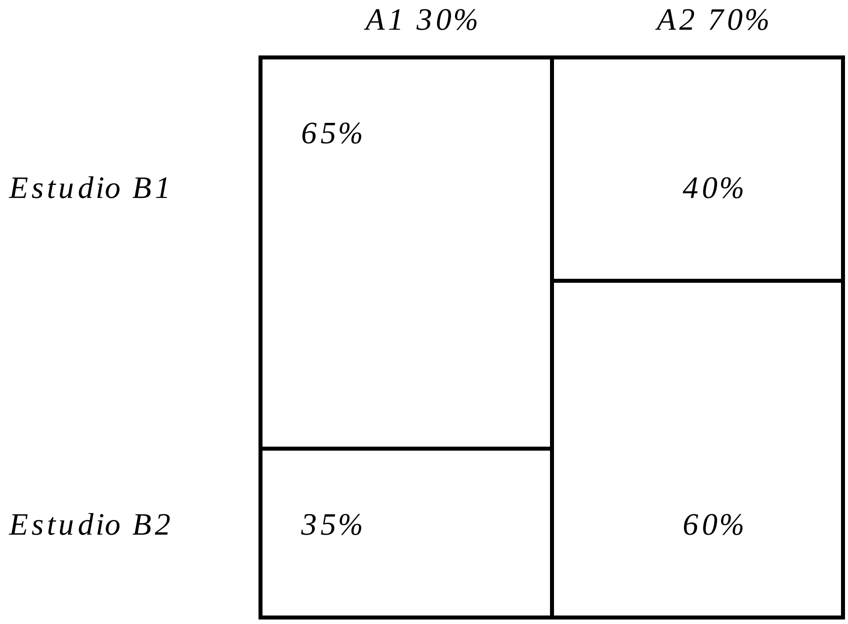
$P(B1|A2) = 1 - P(B2|A2) = 1 – 0,6 = 0,4$

En forma tabular se tiene:

	A	B	C
1		A1	A2
2	Estudio B1	65%	40%
3	Estudio B2	35%	60%
4	Por resultado, A1 o A2	30%	70%

Obsérvese que los porcentajes suman 100% en sentido vertical, esto es, que los porcentajes asignados a los resultados del estudio son las *probabilidades condicionales* de que sus resultados sean *B1* o *B2*, dado que la inversión producirá *A1* o *A2*. En cambio, los porcentajes asignados por resultado del estudio no son probabilidades condicionales, sino "**incondicionales**", esto es, que no dependen de si el resultado de la inversión es *A1* o *A2*. Si el estudio resulta en *B1*, ¿cuál es la probabilidad de que en la realidad ocurra *A1*?

Gráficamente se tiene:



	A	B	C
1		A1	A2
2	Estudio *B1*	65%	40%
3	Estudio *B2*	35%	60%
4	Por resultado, *A1* o *A2*	30%	70%
5	Proporción de veces en que si *A1* ocurre, el estudio diga que va a ocurrir *A1* (*B1*)	=B2*B4 [19,5%]	
6	Proporción de veces en que si *A2* ocurre, el estudio diga que va a ocurrir *A1* (*B1*)	=C2*C4 [28%]	
7	Proporción de veces en que el estudio acierta en su predicción. (Probabilidad de que el resultado sea *A1* dado que el resultado del estudio es B1)	=B5/(B5+B6) [41,05%]	

Obsérvese que antes de obtener la información adicional, la probabilidad que se le asignó al resultado *A1* era 40%, probabilidad a priori; después de contar con la información adicional que el estudio resultó en *B1*, esa probabilidad subió a 41,05%, probabilidad a posteriori o revisada. Ahora el valor esperado de esta inversión será:

$E(A)$ = \$7x0,4105-\$3x(1-0,4105) = 2,8735-1,7685 = 1,1050

Lo que estos ejemplos ilustran es conocido como el *Teorema de Bayes*.

En el primer ejemplo, se cuenta con la siguiente información:

$P(P)$ = 30%, $P(S)$ = 40%, $P(U)$= *25% y* $P(Pos)$ = 5% proporciones (probabilidades) por tipo de educación.

$P(T|P)$= 40%, proporción de amas de casa con primaria que trabaja y $P(NT|P)$=1-40% = 60%, proporción de amas de casa con primaria que no trabaja.

$P(T|S)$= 60%, proporción de amas de casa con secundaria que trabaja y $P(NT|S)$=1-60% = 40%, proporción de amas de casa con secundaria que no trabaja.

$P(T|U)$= 25%, proporción de amas de casa con universitaria que trabaja y $P(NT|U)$=1-25% = 75%, proporción de amas de casa con universitaria que no trabaja.

$P(T|Pos)$= 90%, proporción de amas de casa con postgrado que trabaja y $P(NT|P)$=1-90% = 10%, proporción de amas de casa con postgrado que no trabaja.

La probabilidad de que tenga título de postgrado y de que esté trabajando, en términos de probabilidades condicionales se calcula a continuación.

La probabilidad de que tenga primaria **y** trabaje:

$P(P \cap T) = P(P) \times P(T/P)$ =30%x40%= 12%

La probabilidad de que tenga secundaria **y** trabaje:

$P(S \cap T) = P(S) \times P(T/S)$ = 40%x60% = 24%

La probabilidad de que tenga universitaria **y** trabaje:

$P(U \cap T) = P(U) \times P(T/U)$ = 25%x90% = 22.5%

La probabilidad de que tenga postgrado **y** trabaje:

$P(Pos \cap T) = P(Pos) \times P(T/Pos)$ = 5%x95% = 4.75%

Si la persona seleccionada trabaja, podrá haber cursado primaria, secundaria, universidad o postgrado y estas posibilidades son excluyentes, por lo tanto, esas probabilidades se pueden sumar y se obtiene la probabilidad, o proporción de personas trabajando, así:

$$P(T) = P(P)xP(T/P) + P(S)xP(T/S) + P(U)xP(T/U) + P(Pos) \times P(T/Pos)$$

A esta probabilidad se le conoce como *probabilidad marginal*.

Esta es la probabilidad de que esté trabajando. Pero de las que están trabajando sólo interesan las que tienen postgrado y, de acuerdo con la probabilidad condicional, se tiene:

$$P(Pos\,|T) = \frac{P(Pos \cap T)}{P(T)}$$

$$P(Pos\,|T) = \frac{P(Pos) \times P(T\,/\,Pos)}{P(P)xP(T\,/\,P) + P(S)xP(T\,/\,S) + P(U)xP(T\,/\,U) + P(Pos) \times P(T\,/\,Pos)}$$

En el ejemplo, usando directamente el *Teorema de Bayes*:

	A	**B**	**C**	**D**	**F**
8	Evento a analizar.	Probabilidades básicas.		Probabilidades calculadas	
9		Probabilidad del evento.	Probabilidad de que trabaje dado el evento.	Probabilidad conjunta *P*(*TnP*), *P*(*TnS*), *P*(*TnU*) *P*(*TnPos*)	Probabilidad del evento (*P, S, U, Pos*) dado que trabaja
10	Primaria	*=B2* [30%]	*=C2* [40%]	*=B10*C10* [12%]	*=E10/D14* [18,97%]
11	Secundaria	*=B3* [40%]	*=C3* [60%]	*=B11*C11* [24%]	*=D11/D14* [37,94%]
12	Universitaria	*=B4* [25%]	*=C4* [90%]	*=B12*C12* [22,5%]	*+D12/D14* [35,57%]
13	Postgrado	*=B5* [5%]	*=C5* [95%]	*=B13*C13* [4,75%]	*+D13/D14* [7,51%]
14	Total.			*=D10+D11+D12+D13* [63,25%]	

Obsérvese cómo las probabilidades o proporciones iniciales de los niveles educativos quedaron “revisadas”, después de que se obtuvo la información de la persona seleccionada trabaja. Nótese que las probabilidades de la columnas *B* y *C* son datos conocidos o estimados y no dependen de la teoría de la probabilidad; las probabilidades de las columnas *D* y *E* son calculadas, de acuerdo con la teoría de la probabilidad. El análisis bayesiano conduce a revisar las probabilidades asignadas tipo de educación si se sabe el resultado de la muestra.

En el segundo ejemplo, se cuenta con la siguiente información:

P(A1) = 30% y *P(A2)* = 70% proporciones (probabilidades) por resultados.

P(B1|A1)= 65%, proporción de veces que el estudio dice que va a resultar el evento *A1*, cuando en realidad va a ser *A1*.

P(B2|A1)=1-65% = 35% proporción de veces que el estudio se equivoca cuando en realidad el resultado va a ser *A1*.

P(B2|A2)= 60%, proporción de veces que el estudio dice que va a resultar el evento *A2*, cuando en realidad va a ser *A2*.

P(B1|A2)=1-60% = 40% proporción de veces que el estudio se equivoca cuando en realidad el resultado va a ser *A2*.

La probabilidad que la inversión resulte en *A1* si el estudio resulta en *B1*, en términos de probabilidades condicionales se calcula a continuación.

La probabilidad de que el estudio resulte en *B1* y ocurra *A1*:

$P(A1 \cap B1) = P(A1)xP(B1/A1)$ =30% 65%= 19,5%

La probabilidad de que el estudio resulte en *B2* y ocurra *A1*:

$P(A2 \cap B1) = P(A2)xP(B1/A2)$ = 70%x40% = 28%

Si el estudio indica *B1*, esto puede ocurrir siendo cierto que ocurra *A1* o siendo falso que ocurra *A1*, sino que ocurra *A2* y estas posibilidades son excluyentes, por lo tanto, esas probabilidades se pueden sumar y se obtiene la probabilidad, o proporción de veces que el estudio dice *B1*, así:

$P(B1) = P(A1)xP(B1/A1) + P(A2)xP(B1/A2)$

A esta probabilidad se le conoce como *probabilidad marginal*.

Esta es la probabilidad de que este estudio resulte en *B1*. Pero de los resultados *B1* sólo interesa el que dice la verdad, y de acuerdo con la probabilidad condicional, se tiene:

$$P(A1/B1) = \frac{P(B1 \cap A1)}{P(B1)}$$

$$P(A1/B1) = \frac{P(A1)xP(B1/A1)}{P(A1)xP(B1/A1) + P(A2)xP(B1/A2)}$$

En el ejemplo, usando directamente el *Teorema de Bayes*:

	A	B	C	D	F
15	Evento a analizar.	Probabilidades básicas.		Probabilidades calculadas	
16		Probabilidad del evento.	Probabilidad de que el estudio diga lo correcto dado el evento.	Probabilidad conjunta P(A1∪B1) y P(A2∪B1)	Probabilidad del evento (A1 o A2) dado que el estudio dice B1
17	A1	=B4 [30%]	=B2 [65%]	=B17*C17 [19,5%]	=D17/D19 [41,05%]
18	A2	=C4 [70%]	=C3 [60%]	=B18*(1-C18) [28%]	=D18/D19 [58,95%]
19	Total.			=D17+D18 [47,5%]	

Obsérvese cómo las probabilidades o proporciones iniciales de los resultados de la inversión quedaron "revisadas", después de que se obtuvo la información de que el estudio dice *B1*. Nótese que las probabilidades de la columnas *B* y *C* son datos conocidos o estimados y no dependen de la teoría de la probabilidad; las probabilidades de las columnas *D* y *E* son calculadas, de acuerdo con la teoría de la probabilidad. El análisis bayesiano conduce a revisar las probabilidades asignadas a los resultados de la inversión, dependiendo de los resultados del estudio.

Ahora el valor esperado de esta inversión será:

$E(A)$ = \$7x0,4105-\$3x(1-0,4105) = 2,8735-1,7685 = 1,1050

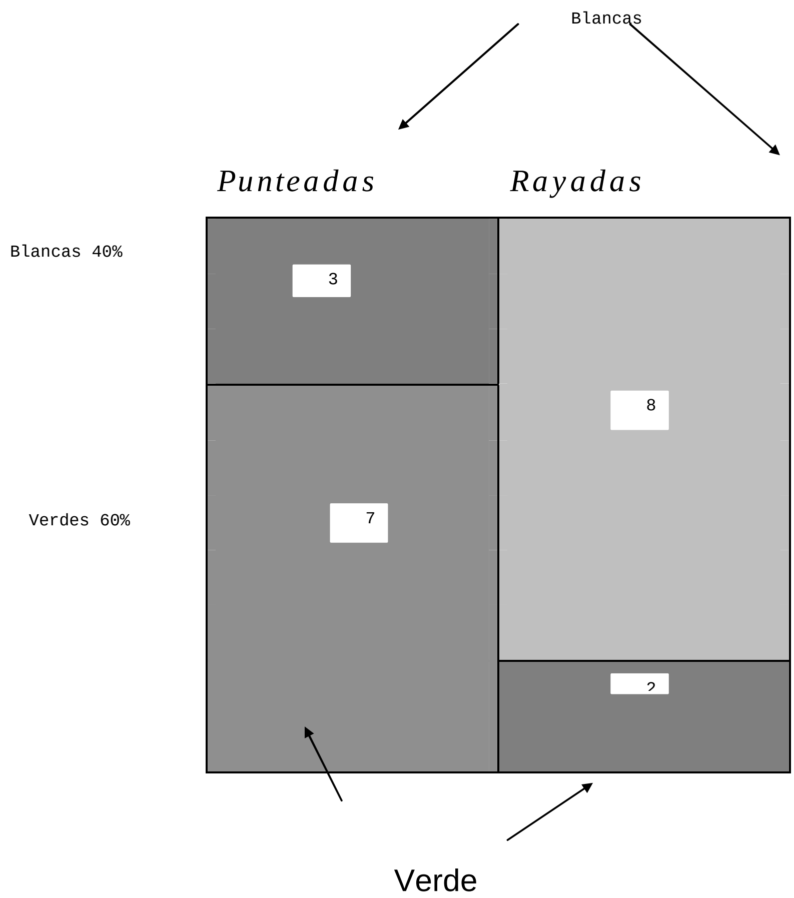
Ejemplo

Supóngase ahora que se tiene una urna con balotas blancas y verdes, y que, a su vez, estas tienen rayas y puntos. La distribución es la siguiente:

	A	**B**	**C**
1		Blancas	Verdes
2	Punteadas.	30%	70%
3	Rayadas.	80%	20%
4	% por color.	40%	60%

Obsérvese que los porcentajes suman 100% en sentido horizontal, esto es, que los porcentajes asignados a las balotas con puntos son las *probabilidades condicionales* que sean blancas o verdes, dado que la balota tenga puntos y lo mismo para las balotas con rayas. En cambio, los porcentajes asignados por color no son probabilidades condicionales, sino "*incondicionales*", esto es, que no dependen de si son o no balotas con puntos o con rayas. Si se extrae una balota de esta urna, ¿cuál es la probabilidad de que sea blanca, si la balota que ha salido es rayada?

Gráficamente se tiene:



	A	B	C
1		Blancas	Verdes
2	Puntos.	30%	70%
3	Rayas.	80%	20%
4	% por color.	40%	60%
5	Probabilidad que la rayada que ha salido sea blanca.	=B3*B4 [32%]	
6	Probabilidad que la rayada que ha salido sea verde.	=C3*C4 [12%]	
7	Probabilidad que la rayada que ha salido sea blanca.	=B5/(B5+B6) [72,73%]	

Obsérvese que antes de obtener la información adicional, la probabilidad (la proporción que había en la urna) que se le asignó a una balota blanca era de 40%; después de contar con la información adicional de que la balota seleccionada era rayada, esa probabilidad ascendió a 72,73%.

En el tercer ejemplo, las probabilidad de que sea rayada, en términos de probabilidades condicionales, es:

La probabilidad de que sea rayada y blanca:

$$P(B \cap R) = P(B) \times P(R / B)$$

La probabilidad de que sea rayada y verde:

$$P(V \cap R) = P(V) \times P(R / V)$$

La balota será rayada y blanca o rayada y verde; por lo tanto, esas probabilidades se pueden sumar y se obtiene la probabilidad de que sea rayada, así:

$$P(R) = P(B) \times P(R / B) + P(V) \times P(R / V)$$

A esta probabilidad se le conoce como *probabilidad marginal*.

Esta es la probabilidad de que sea rayada. De acuerdo con la probabilidad condicional, se tiene:

$$P(B|R) = \frac{\left(P(B \cap R)\right)}{P(R)}$$

$$P(B \mid R) = \frac{P(B \cap R)}{P(B) \times P(R \mid B) + P(V) \times P(R \mid V)}$$

$$P(B \mid R) = \frac{P(B) \times P(R / B)}{P(B) \times P(R \mid B) + P(V) \times P(R \mid V)}$$

En el ejemplo, usando directamente el *Teorema de Bayes*:

	A	B	C	D	E	F
8	Evento a analizar.	Probabilidades básicas.				
9		Probabilidad del evento.	Probabilidad de rayada dado el evento blanca.	Probabilidad de rayada dado el evento verde.	Probabilidad conjunta P(BnR) y P(VnR).	Probabilidad del evento dado que es rayada.
10	Blanca.	=B4 [40%]	=B3 [80%]		=B10*C10 [32%]	=E10/E12 [72,73%]
11	Verde.	=C4 [60%]		=C3 [20%]	=B11*D11 [12%]	=E11/E12 [27,27%]
12	Total.	=B10+B11 [100%]			=D10+D11 [44%]	

NOTA

Nótese que las probabilidades de la columnas *B*, *C* y *D* son datos conocidos o estimados y no dependen de la teoría de probabilidad; las probabilidades de las columnas *E* y *F* son calculadas, de acuerdo con la teoría de la probabilidad.

Independencia Estadística

Se dice que dos eventos son *estadísticamente independientes* cuando la ocurrencia de uno de ellos no afecta la probabilidad de ocurrencia del otro. Esto ocurre cuando se tienen ciertas combinaciones de probabilidades. Técnica y formalmente se dice que un evento *A* es estadísticamente independiente de otro evento *B* si y sólo si, se cumple la siguiente condición:

$$P(A \mid B)=P(A) \tag{10}$$

Esta expresión es muy importante tenerla presente. Cuando se dice de independencia estadística, se refiere en forma exclusiva a que se cumpla la anterior condición. No se hace referencia a ninguna otra clase de independencia, por ejemplo, independencia física al poder lanzar un dado o una moneda, o independencia lógica de poder hacerlo en “forma independiente”, etc.

Si esta condición se cumple, entonces cuando *dos eventos son estadísticamente independientes,* la relación conocida como *regla multiplicativa* queda así:

$$P(A \cap B) = P(B) \times P(A) \tag{11}$$

Se comete un error con frecuencia, al aplicar la regla multiplicativa de eventos estadísticamente independientes, a eventos que no lo son. Hay que ser cuidadoso en su uso. Sólo existe independencia estadística cuando la ocurrencia de un evento, no afecta la probabilidad de ocurrencia del otro.

Ejemplo

Si en el ejemplo anterior se tuvieran los siguientes datos:

	A	B	C
1		Blancas	Verdes
2	Puntos.	12%	18%
3	Rayas.	28%	42%
4	% por color.	40%	60%
5	Probabilidad que sea rayada.	=B3+C3 [70%]	
6	Probabilidad que sea rayada y blanca.	=B3 [28%]	
7	Probabilidad que la rayada que ha salido, sea blanca.	=B6/B5 [40%]	

Es decir, que la probabilidad de que sea blanca –"*a priori*"-, es 40%, y la probabilidad de que sea blanca, dado que salió rayada es también 40%, lo cual indica que con las probabilidades que se asignaron al ejemplo, en este último caso, hay independencia estadística entre el evento balotas rayada o punteada y balota verde o blanca.

Otra vez, no importa que conceptualmente "el ser balota rayada no tiene nada que ver con ser balota blanca", sino que lo importante es que las probabilidades guarden ciertas proporciones entre sí, para que se de la independencia estadística. Se puede pensar en si un evento aporta o no, más información de la que se tenía, para considerar la independencia estadística.

Conceptos básicos de estadística

"Hay gente que utiliza la Estadística como un borracho utiliza el poste de la luz; más para apoyarse que para iluminarse".
Andrew Lang (Citado por Thomas y Ronald Wonnacott)
"Las cifras no mienten, pero los mentirosos sí hacen cálculos con ellas"
General Charles H. Grosvenor (Citado por Thomas y Ronald Wonnacott)

La estadística es un método científico de análisis que se aplica a las ciencias sociales y naturales. Su principal utilización es la inferencia estadística, esto es, que a partir de la información obtenida de una muestra -reducido número de observaciones de un universo- se hacen "inferencias" sobre la población total. Este tema no es el objeto de este nivel introductorio y se tratará posteriormente.

Estadística Descriptiva

Lo primero que se debe hacer con la información obtenida de una muestra, es reducirla a unas cuantas cifras que condensen o concentren la información más importante. Estas cifras se conocen como las estadísticas de la muestra.

Obsérvese la diferencia entre Estadística, área del conocimiento que permite hacer inferencia sobre poblaciones, y la estadística de una muestra, que es una cifra que describe a esa muestra o al universo. También debe distinguirse entre la estadística de una muestra y el parámetro que describe a un universo. Lo que se calcula para una muestra son las estadísticas de la muestra que pueden servir para calcular o hacer una estimación de los parámetros del universo.

Supóngase que se ha obtenido una muestra de la audiencia de una cierta población y se indaga sobre hábitos de lectura. Si la muestra que se obtuvo es de 1.000 personas y 345 de ellas responden que no leen, entonces una forma de describir la muestra es diciendo que el 34,5% de ella no lee. Esta cifra puede ser utilizada para hacer una inferencia de la población en cuanto a los hábitos de lectura.

Ahora bien, los datos que se obtienen no pueden ser utilizados sin un previo análisis y sin reserva. Por lo general, cuando se toma una muestra se incurre en algún tipo de error estadístico, el cual tiene que ver con el tamaño de la muestra; intuitivamente es obvio que si se tiene un universo muy grande, a mayor información que se obtenga -mayor tamaño de la muestra- más cerca de la realidad van a estar las estadísticas de la muestra, comparadas con las estadísticas del universo. Los técnicos reconocen entonces un margen de error, y se dice que un dato tiene un margen de error. Por ejemplo, los datos de preferencia de votos en una campaña electoral se expresan como que el 65% de la gente votará por el candidato *A*, con un margen de error de ±5%. Esto es, que el verdadero valor se estima que está dentro del intervalo 60%-70% y esta afirmación tiene una determinada probabilidad de que sea cierta; se dice entonces que es un intervalo de *P*% de confianza (por ejemplo, 95% de confianza).

Distribuciones de probabilidad

Un universo tiene unas características estadísticas que, como se dijo arriba, pueden especificarse con unas cuantas cifras. Así mismo, los universos tienen dos características básicas: son discretos cuando los valores que toman las unidades que lo configuran toman valores finitos, por ejemplo, el número de años en que los ingresos son cero, 0, 1, 2, 3, etc. o son continuos lo cual significa que pueden tomar un infinito número de valores, como puede ser el espesor de una lámina de acero o, con mucho rigor, la edad del ser humano.

Histogramas y Tablas

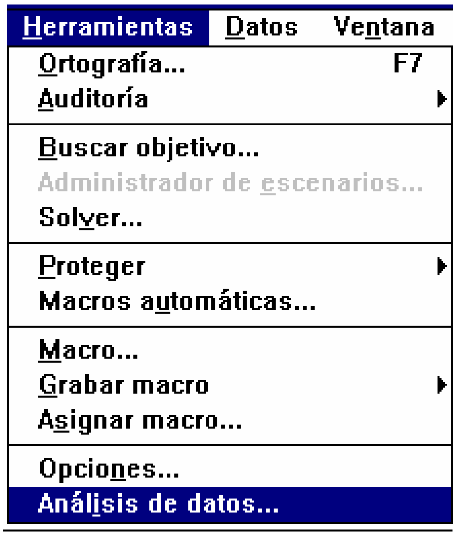
Una manera de visualizar la información de una muestra es tabularla o mostrar la gráfica de los valores obtenidos.

Caso discreto

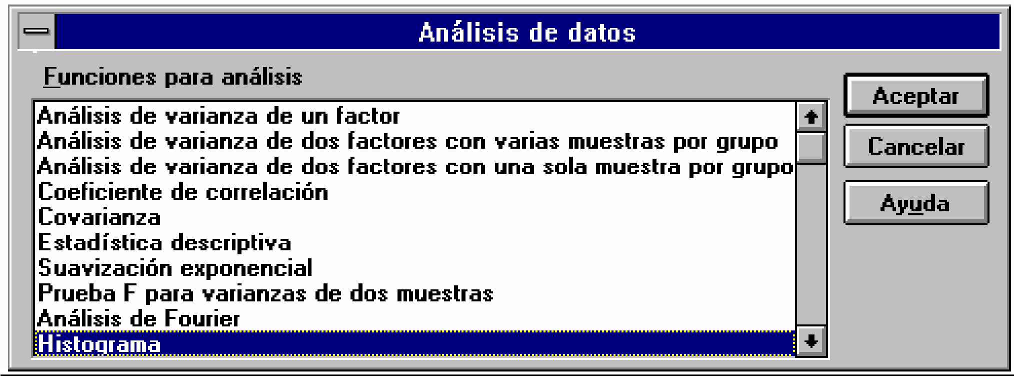
Suponga que se hace una muestra 6.400 viviendas de un país. (Esto puede hacerse con facilidad si se tiene acceso a los formularios de un censo de población o de una manera más compleja, construyendo una muestra aleatoria de las 6.400 viviendas, localizándolas y visitándolas para verificar los datos). La muestra indica que en las viviendas el número de habitaciones es de 1, 2, 3, 4, 5 ó 6.

En el archivo *Estadística.XLS* en la hoja *DISTRIVI* se encuentran los valores de una muestra de 6.400 viviendas en términos del número de habitaciones de cada vivienda. Estos datos están en el rango *A1:J640* y a ellos se hace referencia en la siguiente tabla. Se debe advertir que según los manuales de *Excel*, esta función no puede manejar más de 6.400 observaciones; sin embargo, el autor ha trabajado con 10.000 y se han obtenido resultados satisfactorios, excepto en la configuración de la fórmula. Para ilustrar el uso de la función se presenta un ejemplo:

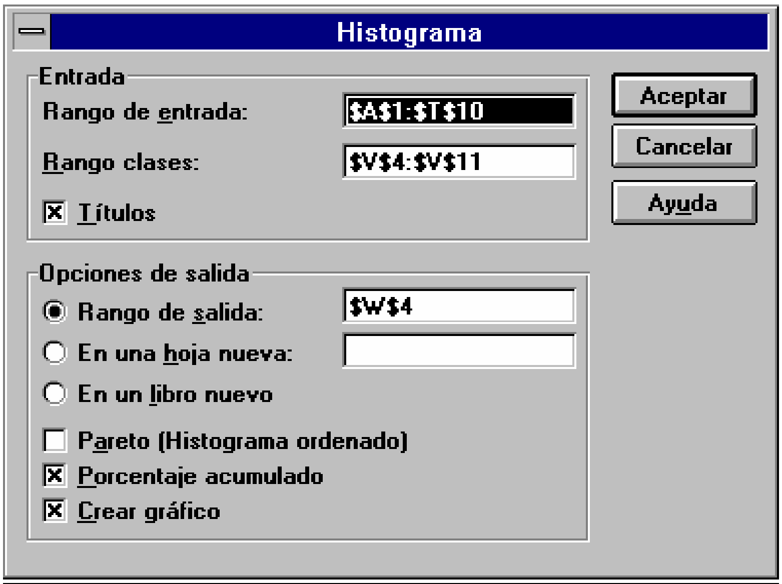
Si se tienen los siguientes datos y se desea calcular cuántas veces aparecen viviendas con 1,2,3,4,5 o 6 habitaciones, se debe usar la función *=FRECUENCIA(Datos;grupos)* o en la Barra de Herramientas de *Excel*, se oprime *Herramientas* y aparece el menú que se muestra a continuación. Allí se escoge *Análisis de Datos*. Esta opción se explicará en detalle**.**



Cuando se hace *click* con el *Mouse* en *Análisis de Datos*, aparece el siguiente cuadro de diálogo:



Allí se selecciona *Histograma* y aparece el cuadro de diálogo siguiente:



Se le debe indicar al programa en qué rango se hallan los datos, dónde insertar los grupos o rango de clases y cuál ha de ser el rango de salida. Además se debe indicar si se desea el porcentaje acumulado, además del histograma de frecuencias absolutas y si se desea hacer una gráfica. Hecho esto, se oprime el botón 

. También se puede pedir que construya una curva de *Pareto*, la cual consiste en ordenar los valores de mayor a menor frecuencia. Todo esto lo hace *Excel* en forma automática.

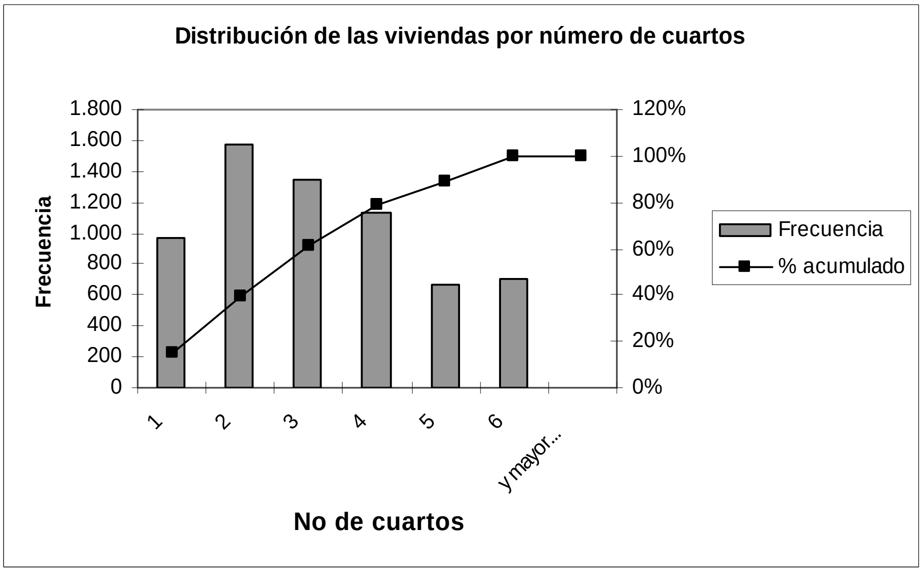
Con los datos del archivo *Estadística.XLS*, en la hoja *DISTRIVI* se introdujeron los rangos en el cuadro de diálogo anterior.

Los resultados con los datos anteriores son los siguientes:

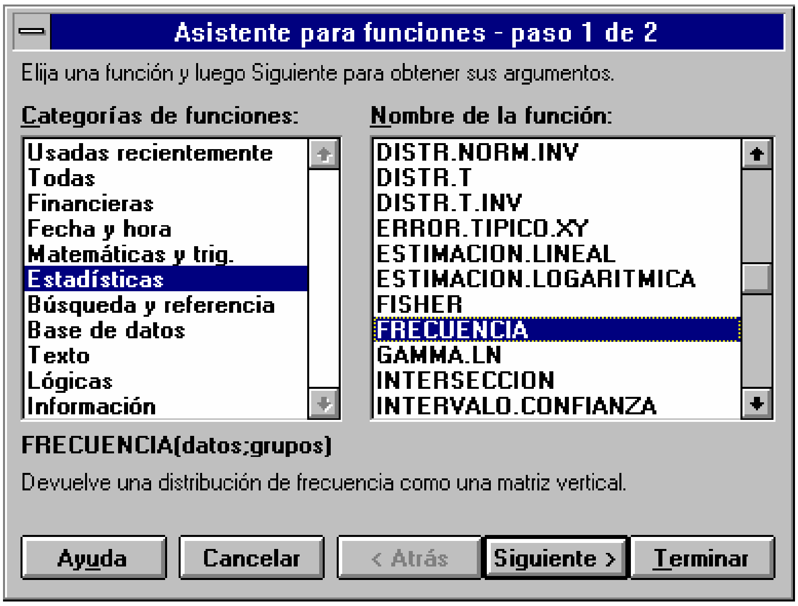
En forma tabular:

No de cuartos	*Frecuencia*	*% acumulado*
1	970	15,16%
2	1.575	39,77%
3	1.349	60,84%
4	1.136	78,59%
5	665	88,98%
6	705	100,00%
y mayor...	0	100,00%

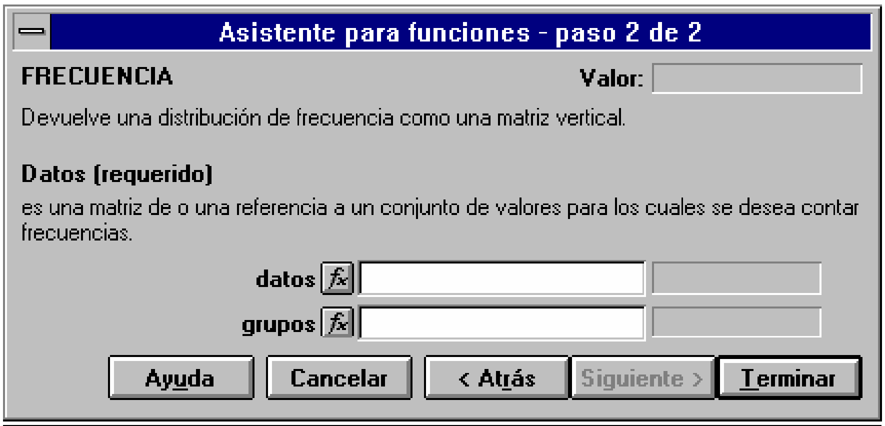
En forma gráfica:



Cuando se usa la función ***frecuencia***, el procedimiento es más complicado y menos espectacular. En la barra de herramientas se encuentra el botón del "*Asistente de Funciones*",  al oprimirlo aparece el siguiente cuadro de diálogo:



En este cuadro ya se señaló las funciones "*Estadísticas*" y dentro de ellas se escogió "*Frecuencia*". Cuando ya ha sido seleccionada la función que interesa, se oprime el botón Siguiente > y aparece este cuadro de diálogo:



En la casilla 

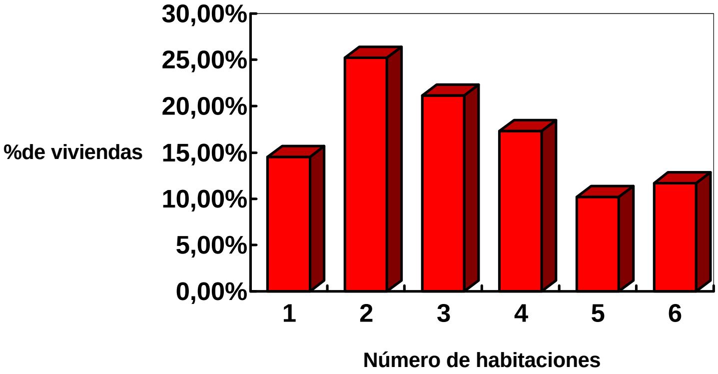
se introduce el rango donde se encuentran los datos; en el ejemplo, *A1:T10*. En la casilla 

se introduce el rango de los grupos en que se desea clasificar los datos; en el ejemplo, *M3:M8*. Al oprimir el botón 

aparece el resultado en la celda *O3*; a partir de esa celda se debe señalar el rango *O3:O8*, inmediatamente, estando sobre la celda *O3*, se oprime *F2* y simultáneamente las teclas *CTRL+MAYUSCULAS+ENTRAR (en Windows)* y *COMANDO+ENTRAR (en Macintosh).* Así se convierte la función (fórmula) *FRECUENCIA*, en una matriz y se obtiene el valor de la frecuencia de ocurrencia para cada grupo.

	M	**N**	**O**
1	Número de habitaciones.	Frecuencia absoluta acumulada.	Frecuencia relativa.
2			
3	1	{=FRECUENCIA(A1:J640;M3:M8)} [937]	=N3/$N$9 [14,64%]
4	2	{=FRECUENCIA(A1:J640;M3:M8)} [1,603]	=N4/$N$9 [25,05%]
5	3	{=FRECUENCIA(A1:J640;M3:M8)} [1,363]	=N5/$N$9 [21,30%]
6	4	{=FRECUENCIA(A1:J640;M3:M8)} [1,109]	=N6/$N$9 [17,33%]
7	5	{=FRECUENCIA(A1:J640;M3:M8)} [650]	=N7/$N$9 [10,16%]
8	6	{=FRECUENCIA(A1:J640;M3:M8)} [738]	=N8/$N$9 [11,53%]
9	TOTAL	=SUMA(N3:N8) [6,400]	=SUMA(O3:O8) [100,00%]

Los valores obtenidos se pueden mostrar en una gráfica que se llama histograma de frecuencia.



En el archivo *Estadística.XLS* en la hoja *DISTRIVI* construir la tabla y gráfica anteriores.

Caso continuo

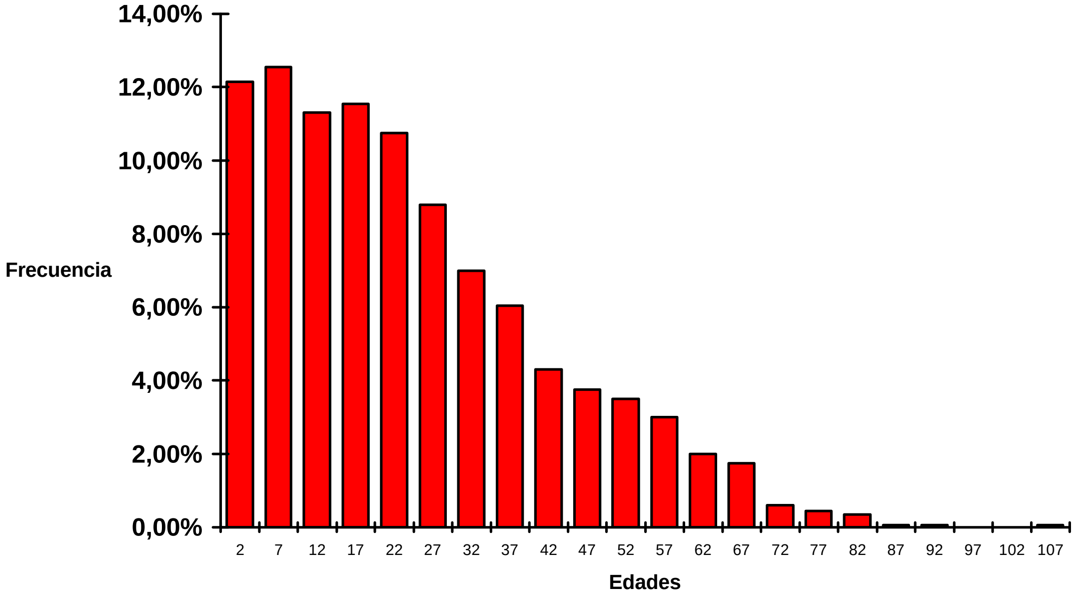
Nuevamente, si se toman los datos de un censo de población y se obtiene una muestra de 2.000 personas, las edades se clasifican en intervalos y no en valores puntuales. Estrictamente, la edad de una persona se comporta como una variable continua, a pesar de que en la práctica la gente “redondea” su edad en números enteros y casi nunca, o nunca, se dice que alguien tiene 22 años, 3 meses, 27 días, 4 horas, etc. También en la práctica nadie tiene la misma y exacta edad de otra persona. Por consiguiente carece de sentido hablar de valores concretos, antes bien, se habla de rangos de edad. Más aun, en el caso de las edades, se acostumbra a definir los rangos con extremos enteros, por ejemplo, se habla del grupo de edad de 0-4 años o de 5-9 años. La muestra indica que las personas tienen edades entre 0 y 100 años (en la práctica, aunque se puede exceder esa cifra). La muestra se puede clasificar de acuerdo con los grupos de edad quinquenales (cinco años) y su tabulación se puede presentar así:

Grupos de Edad (años)	Frecuencia absoluta	Frecuencia relativa
0-4	243	12,15%
5-9	251	24,70%
10-14	226	36,00%
15-19	231	47,55%
20-24	215	58,30%
25-29	176	67,10%
30-34	140	74,10%
35-39	121	80,15%
40-44	86	84,45%
45-49	75	88,20%
50-54	70	91,70%
55-59	60	94,70%
60-64	40	96,70%
65-69	35	98,45%
70-74	12	99,05%
75-79	9	99,50%
80-84	7	99,85%
85-89	1	99,90%
90-94	1	99,95%
95-99	0	99,95%
100-104	0	99,95%
105-109	1	100,00%
Total	0	100,00%

Siempre surge la pregunta de: ¿cuántos intervalos se deben construir? La respuesta es que esto depende de los datos que se deseen analizar y no deben ser, ni muchos, ni pocos. Se puede considerar que entre 5 y 15 intervalos sería razonable. En cuanto al punto medio de cada intervalo, es preferible considerar un número entero.

Si se analizan las edades de la muestra y de la población de manera estricta, se tendría un patrón de muchos valores continuos con una concentración en los primeros 40 años, que se iría reduciendo a medida que se aumenta la edad.

El histograma de frecuencias de las edades sería así, considerando el valor central de cada intervalo como el valor del mismo:



En el archivo *Estadística.XLS* y en la hoja *DISTRIED* construir la tabla y gráfica anteriores. Mostrar también la gráfica de la frecuencia relativa.

Estadísticas de una distribución

Arriba se mencionó que la distribución de un universo se podía representar por las estadísticas de la muestra o del universo. Las estadísticas más comunes son aquellas que muestran la tendencia central o valor alrededor del cual se agrupan los elementos del universo y el grado de dispersión. Estas dos ideas se ilustrarán con el caso discreto de las habitaciones de las viviendas de la muestra seleccionada.

Tendencia central de la distribución

Con esta estadística se trata de examinar hacia qué valor se concentran los valores de la distribución. Las estadísticas más conocidas que miden la tendencia central son: La *moda*, la *mediana* y la *media* o *valor esperado.*

La moda

La moda se define como el valor más frecuente. Esto es, aquel valor que tiene mayor frecuencia. Debido a que los datos pueden agruparse de manera arbitraria —en el caso de la distribución continua— la moda no es la mejor medida de tendencia central. También puede suceder que haya dos “modas” iguales, en ese caso se dice que la distribución es bimodal y se presenta una ambigüedad. La forma más fácil de determinar la moda es utilizando el histograma de frecuencias.

Por medio del histograma de frecuencias de los ejemplos anteriores, identificar la moda. *Excel* tiene la fórmula para ello, *=MODA(Datos)* pero está restringida a un número reducido de observaciones; para 400 observaciones calcula el valor, para 10.000 arroja error.

La mediana

La mediana es aquel valor que divide la distribución en partes iguales, o sea que el número de observaciones por encima de la mediana es igual al número de observaciones por debajo de ella. Se conoce también como el valor medio o percentile 50.

Con los datos de los ejemplos anteriores debe usted identificar la mediana. Excel tiene la fórmula para ello, *=MEDIANA(Datos).*

La media o valor esperado

El valor esperado o media indica la tendencia central de los datos. Esto significa que es el valor alrededor del cual tienden a agruparse los datos de una distribución. En el caso de una variable aleatoria discreta, se calcula multiplicando cada valor posible por su probabilidad y sumando sus resultados. En el caso de una variable aleatoria continua, se debe recurrir al concepto de integral que se estudia en el cálculo integral. Generalmente se expresa por medio de la letra griega μ (parámetro) para el universo y por la notación $E(\ )$ o $\overline{X}$ (estadística) para una muestra.

Con los datos de los ejemplos anteriores calcular la media. En Excel la fórmula es *=PROMEDIO(Datos).*

Medidas de la dispersión de la distribución

Las estadísticas que describen a una muestra o universo muestran qué tan dispersas están las observaciones o los elementos del universo. Las más comunes son la varianza, la desviación estándar (es la raíz cuadrada de la varianza) y el rango. Intuitivamente se puede pensar en medir las diferencias entre cada observación y el valor central, por ejemplo, el valor esperado o media. Eso va a producir valores negativos y positivos y al sumarse entre sí deben cancelarse y producir el valor cero. Se puede obviar este inconveniente trabajando con el valor absoluto, en *Excel*, *=DESVPROM(Datos)* o con los cuadrados de las diferencias en *Excel*, *=DESVIA2(Datos)*. Cuando se desea medir las *variaciones entre dos o más variables*, entre sí, entonces se habla de la *covarianza*.

Varianza

Una medida de la dispersión de unos datos es la varianza. Esta se calcula así:

$$s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n} \qquad (12)$$

O sea que es el promedio del cuadrado de las diferencias de cada dato con el promedio. Esta expresión se aplica para la distribución y la muestra; cuando se refiere a la población o universo, se utiliza la letra griega sigma s^2(parámetro) y s^2 (estadística), cuando se trata de la muestra. Sin embargo, cuando se trata de estimar la varianza de un universo o distribución a partir de la varianza de una muestra de tamaño n, entonces la fórmula debe modificarse así:

$$s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} \qquad (13)$$

Con los datos de los ejemplos anteriores calcular la varianza de los datos obtenidos y de la distribución o universo de donde procedieron esos datos. En *Excel* la fórmula es *=VAR(Datos)* cuando se trata de medir la varianza de la muestra, y *=VARP(Datos),* cuando se trata, a partir de la muestra, calcular la varianza del universo.

Desviación estándar

La desviación estándar (s) es la raíz cuadrada de la varianza. Se puede demostrar que si X_1, X_2, X_3....X_n son variables aleatorias independientes con media m_i y desviación estándar s_i, entonces la suma de esas variables tendrán una distribución normal con media m_i y desviación estándar nxs_i^2.

Con los datos de los ejemplos anteriores, calcular la desviación estándar de los datos obtenidos y de la distribución o universo de donde procedieron esos datos. En *Excel* la fórmula es *=DESVEST(Datos),* cuando se trata de medir la desviación estándar de la muestra (con esta función se hace una estimación del parámetrto del universo), y *=DESVESTP(Datos),* cuando se trata de calcular la desviación estándar del universo, a partir de la totalidad de los datos de ese universo. (DESVEST parte de la hipótesis de que los argumentos representan la muestra de una población. Si sus datos representan la población total, utilice DESVESTP para calcular la desviación estándar).

Rango

Otra manera de estimar la dispersión de unos datos es medir su rango. Esta es la diferencia entre el valor máximo y el valor mínimo.

Con los datos de los ejemplos anteriores calcular el rango de los datos obtenidos. En *Excel* la fórmula para el valor máximo es *=MAX(Datos)* y para el valor mínimo es *=MIN(Datos).*

Covarianza

La covarianza indica en qué medida dos variables se mueven al unísono. Si se observa el comportamiento de la rentabilidad de las acciones en la Bolsa, se encontrará que algunas de ellas aumentan al mismo tiempo y otras disminuyen mientras las otras aumentan. El cálculo de la covarianza relaciona las diferencias entre las variables y sus medias, unas con otras, así:

$$s_{ij}^{2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(X_i - \overline{X}_i\right)\left(X_j - \overline{X}_j\right)}{n} \qquad (14)$$

o

$$s_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}\left(X_i - E(X_{i)}\right)\left(X_j - E(X_j)\right) \qquad (15)$$

También se puede expresar como:

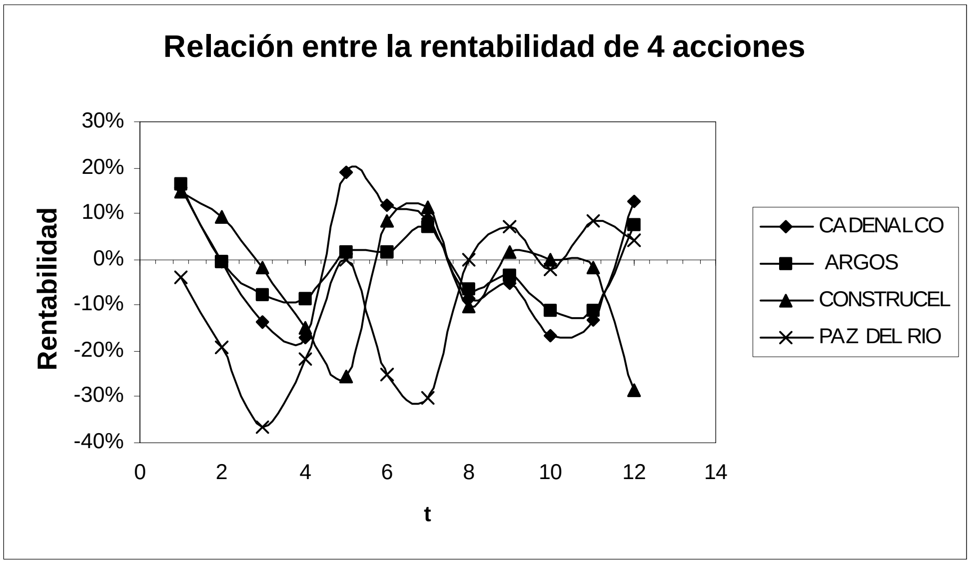
$$s_{ij} = r_{ij} s_i s_j \qquad (16)$$

El resultado del cálculo de la covarianza es una tabla como la siguiente:

$$\begin{matrix} s_{11} & s_{21} & \dots & \dots & s_{n1} \\ s_{12} & s_{22} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & s_{nn-1} \\ s_{1n} & \dots & \dots & s_{n-1n} & s_{nn} \end{matrix} \tag{17}$$

Hay que observar que los datos de la diagonal son las varianzas de cada variable. Las demás son las covarianzas y son simétricas. Por ejemplo, si se desea saber cómo varían cuatro acciones de la Bolsa de Bogotá, se tiene:

Mes	CADENALCO	ARGOS	CONSTRUCEL	PAZ DEL RIO
1	15,75%	16,63%	14,87%	-3,91%
2	-0,47%	-0,36%	9,42%	-19,33%
3	-13,65%	-7,77%	-1,68%	-36,61%
4	-17,00%	-8,77%	-14,81%	-21,58%
5	18,87%	1,51%	-25,58%	0,04%
6	11,78%	1,50%	8,57%	-25,22%
7	9,00%	6,90%	11,42%	-30,23%
8	-8,61%	-6,54%	-10,32%	0,00%
9	-5,31%	-3,57%	1,71%	7,22%
10	-16,73%	-11,06%	0,00%	-2,25%
11	-13,21%	-11,33%	-1,88%	8,52%
12	12,65%	7,72%	-28,44%	4,22%



La covarianza se puede calcular en *Excel* con la opción de menú *Herramientas* y allí se selecciona *Análisis de datos*. En el cuadro de diálogo que aparece, se escoge *Covarianza* y se indica el rango donde están los datos para los cuales se desea calcular la covarianza. La matriz de covarianza de las rentabilidades de las cuatro acciones es:

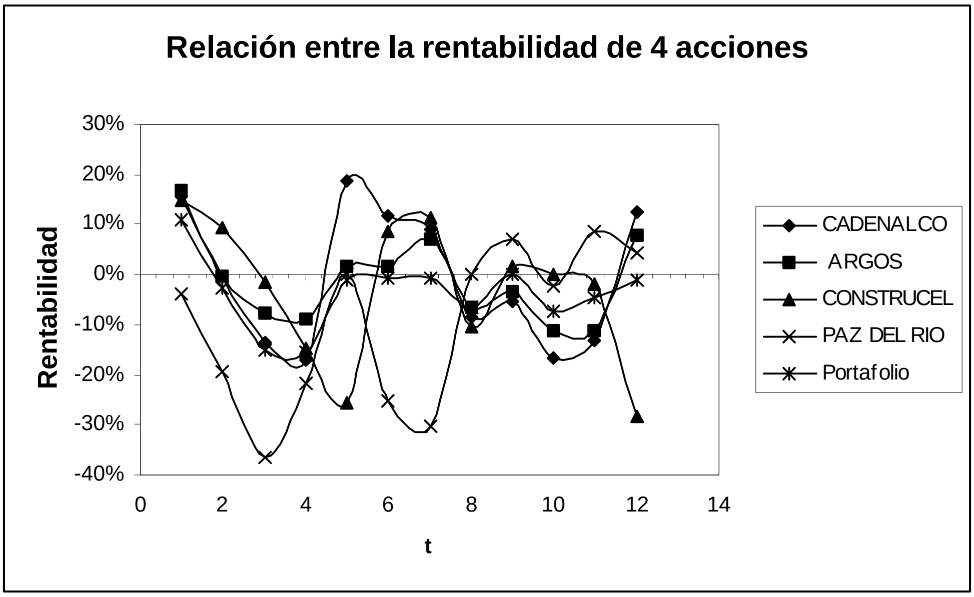
	CADENALCO	*ARGOS*	*CONSTRUCEL*	*PAZ DEL RIO*
CADENALCO	0,0182927	0,01022173	-0,00046223	0,00106384
ARGOS	0,01022173	0,00738926	0,00233332	-0,00041683
CONSTRUCEL	-0,00046223	0,00233332	0,01999673	-0,00041683
PAZ DEL RIO	0,00106384	-0,00041683	-0,00827034	0,02463414

La covarianza de *Argos* y *Argos* es su varianza; en forma similar, la covarianza entre Cadenalco y Cadenalco es su varianza. La covarianza de

Construcel y Construcel es su varianza; en forma similar, la covarianza entre Paz del Río y Paz del Río es su varianza.

En estas gráficas se observa que mientras dos de las acciones tienden a seguir el mismo comportamiento (suben o bajan ambas) las otras dos se comportan de manera contraria (mientras una sube, la otra baja). La medida del grado de coincidencia en el comportamiento está dada por la covarianza entre ellas.

Si se mezclan las cuatro acciones en partes iguales (si se construye un portafolio de las cuatro acciones por partes iguales), gráficamente se tiene lo siguiente:



Y los resultados de la media y la desviación estándar son los siguientes:

	CADENALCO	ARGOS	CONSTRUCEL	PAZ DEL RIO	PORTAFOLIO
PROMEDIO	-0,58%	-1,26%	-3,06%	-9,93%	-3,71%
DESVEST	12,95%	8,23%	13,54%	15,03%	6,74%

Obsérvese cómo se redujo la desviación estándar (la variabilidad del portafolio), debido a la combinación de cuatro variables (rentabilidad) y a la inclusión de acciones con covarianzas negativas.

Correlación

El dato que proporciona la covarianza no es fácil de interpretar cuando se expresa en las unidades de las variables analizadas; se necesita un indicador que sea independiente de las unidades de las variables analizadas; para evitar este problema, se puede encontrar la correlación o índice de correlación entre dos variables, escalándolo o "normalizándolo" con las desviaciones estándar, así:

$$r_{ij} = \frac{s_{ij}}{s_i s_j} \qquad (18)$$

Esta expresión se conoce como coeficiente de y está entre -1 y 1. Sirve para medir el grado de asociación entre dos variables u observaciones. En el ejemplo de las acciones se tiene, para Construcel y Paz del Río, desviaciones estándar:

	CADENALCO	*ARGOS*	*CONSTRUCEL*	*PAZ DEL RIO*
CADENALCO	0,0182927	0,01022173	-0,00046223	0,00106384
ARGOS	0,01022173	0,00738926	0,00233332	-0,00041683
CONSTRUCEL	-0,00046223	0,00233332	0,01999673	-0,00827034
PAZ DEL RIO	0,00106384	-0,00041683	-0,00827034	0,02463414

	CONSTRUCEL	PAZ DEL RIO
Varianza	0,01999673	0,02463414
Desviación estándar	0,14140981	0,15695266

$$r(c,p) = \frac{-0{,}00827034}{0{,}14140981 \times 0.15695266} = -0{,}37262751$$

Este valor indica que están correlacionadas negativamente; o sea, cuando la rentabilidad de una acción aumenta, la rentabilidad de la otra tiende a bajar.

Variable aleatoria

Una variable aleatoria es el valor que se le asigna a un determinado evento. Por ejemplo, en el ejemplo de la inversión a tres años, se le puede asignar un valor monetario a cada evento y un *Valor Presente Neto* (VPN) a la inversión; el VPN es una variable aleatoria.

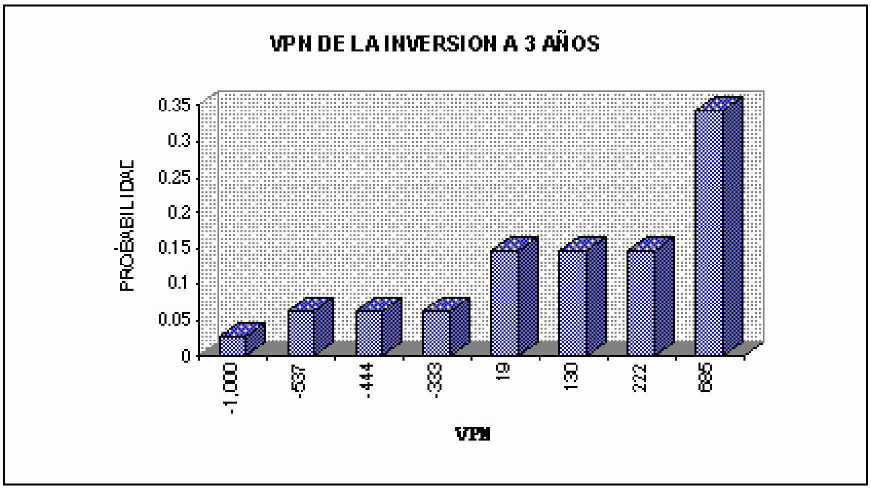
Si se retoma el ejemplo de la inversión a tres años, se habían previsto ciertos resultados y si se supone que cada año sin ingreso se le asocia el valor cero y cada año con ingreso se le asocia el valor $600, al año sin ingreso se le asocia una probabilidad de 30% y se tiene una tasa de descuento de 20% anual, entonces se tiene lo siguiente:

	A	B	C
1	$P(N)_1$=	30%	Inversión
2	$P(N)_2$=	30%	$1,000
3	$P(N)_3$=	30%	Tasa de descuento anual.
4	Valor de un año con ingreso.	$600	20%
5	Valor de un año sin ingreso.	$0	
6	Evento combinado.	Probabilidad total.	Variable aleatoria (VPN).
7	(NNN) = m_1	=B1*B2*B3 [0,027]	-1,000
8	(NNS) = m_2	=B1*B2*(1-B3) [0,063]	=B4/(1+C4) ^3-C2 [-537,04]
9	(NSN) = m_3	=B1*(1-B2)*B3 [0,063]	=B4/(1+C4) ^2-C2 [-444,44]
10	(NSS) = m_4	=B1*(1-B2)*(1-B3) [0,147]	=B4/(1+C4) ^2+B4/(1+C4) ^3-C2 [18,52]
11	(SNN) = m_5	=(1-B1)*B2*B3 [0,063]	-333,33
12	(SSN) = m_6	=(1-B1)*(1-B2)*B3 [0,147]	222,22
13	(SNS) = m_7	=(1-B1)*B2*(1-B3) [0,147]	129,63
14	(SSS) = m_8	=(1-B1)*(1-B2)*(1-B3) [0,343]	685,19

La distribución de la variable aleatoria será:

	A	B	C
13	Variable aleatoria (VPN)	Probabilidad	
14	-1.000,00	2,7%	
15	-537,04	6,3%	
16	-444,44	6,3%	
17	-333,33	6,3%	
18	18,52	14,7%	
19	129,63	14,7%	
20	222,22	14,7%	
21	685,19	34,3%	
22	Total	100,0%	

Con los datos de la tabla anterior construir el histograma que aparece a continuación.



Las variables aleatorias se representarán por X y los valores específicos por x. En el ejemplo X es el *Valor Presente Neto* (VPN) de la inversión a tres años. Se dice entonces que esta variable aleatoria X toma los siguientes valores:

X_1=	-1.000,00
X_2=	-537,04
X_3=	-444,44
X_4=	-333,,33
X_5=	18,,52
X_6=	129,63
X_7=	222,22
X_8=	685,19

En general, se escribiría $X=x$. La probabilidad se escribiría $P(X=222,22)$, $P(X=685,19)$, etc.

Asociada a toda variable aleatoria existe una función de distribución acumulada. Si se define el evento: variable aleatoria X menor o igual que b, como $E = (X< b)$, entonces $P(E)$ es la probabilidad de este evento y se denomina probabilidad acumulada de b, $F(b)$. La función acumulada de probabilidad es una función numérica definida para todos los valores posibles de b y tiene las siguientes propiedades:

- $F(b)$ es una función no decreciente de b. Esto es, que a medida que b aumenta, $F(b)$ aumenta o permanece igual y nunca disminuye.
- La variable aleatoria puede tomar valores entre menos infinito ($-\infty$) y más infinito ($+\infty$).

Por lo tanto:

- $F(-\infty) = 0$ y $F(\infty) = 1$ (19)

De acuerdo con estas definiciones y propiedades, se puede establecer el valor de la probabilidad que una variable aleatoria se encuentre entre dos valores dados como:

$P(a < X < b) = F(b) - F(a)$ (20)

Esto significa que cuando se tiene una variable continua, la probabilidad de un valor preciso es cero.

La distribución de probabilidad discreta

Una distribución de probabilidad es discreta cuando para cada valor existe una probabilidad de ocurrencia. Ejemplos de fenómenos que se clasifican como discretos, son el valor del lanzamiento de un dado, de una moneda, el número de hijos de una pareja, etc.

En términos generales, la distribución acumulada de esta clase de leyes probabilísticas está dada por:

$$F(b) = P(x \leq b) = \sum_{i=-\infty}^{b} P(x = x_i) \quad \textit{para todas las } x_i \leq b \qquad (21)$$

Esto significa que la probabilidad de que ocurra un valor menor o igual a b, es igual a la suma de todas las probabilidades de los valores de X menores que b. Por ejemplo, la probabilidad de que el resultado del lanzamiento de un dado de seis caras sea menor o igual que 3 es igual a la suma de las probabilidades de que el valor del lanzamiento sea 1, 2 ó 3.

Por otro lado, la media y la varianza se calculan así:

Media de la población: $m = \sum_x xp(x)$ (22)

Varianza de la población: $s^2 = \sum_x (x-m)^2 p(x)$ (23)

La Distribución Binomial

Existen muchas leyes de probabilidad discretas; la más común es la binomial. Esta distribución se utiliza en situaciones con un número fijo de pruebas o ensayos, cuando los resultados de un ensayo son sólo éxito o fracaso, cuando los ensayos son independientes y cuando la probabilidad de éxito es constante durante todo el experimento. Por ejemplo, se puede calcular la probabilidad que dos de los próximos tres bebés que nazcan de un pareja sean hombres. A continuación se muestran algunos fenómenos regidos por la distribución binomial:

Fenómeno	**Éxito**	**Fracaso**	***p* (probabilidad de éxito)**	***n* (casos totales)**	***r* (éxitos)**
Lanzar una moneda.	Cara.	Sello.	0,50	n lanzamientos.	Número de caras.
Nacimientos en una familia.	Niño.	Niña.	0,50	Tamaño de la familia.	Número de niños en la familia.
Lanzamiento de 3 dados.	8 puntos.	Cualquier otro resultado.	21/216	*n* lanzamientos.	Número de 8's.
Resultado de inversiones.	Ingreso.	No ingreso.	Asignada según el caso.	Períodos futuros en que pueden ocurrir los ingresos.	Períodos con ingresos.
Nacimientos de terneros.	Ternera.	Ternero.	0,50	Número de partos.	Número de terneras.
Escogencia de un votante en una encuesta de opinión política.	Liberal.	Otros partidos.	Proporción de Liberales en la población.	Tamaño de la muestra.	Número de liberales en la muestra.
Lanzamiento de un dado.	1 punto.	Cualquier otro valor.	1/36	n lanzamientos.	Número de 1's.
Aprobación de leyes en el Congreso.	Aprobada.	Rechazada.	Proporción de congresistas a favor de la ley.	Número de congresistas.	Aprobación de la ley.

La distribución binomial típica es el lanzamiento de una moneda: S = número de caras en n lanzamientos.

Se dice que son *n* lanzamientos independientes y para cada lanzamiento hay un éxito (cara) o un fracaso (sello) y las probabilidades son *p* para cara y *(1-p)* para sello.

Si se tiene 1 dado, la distribución binomial que rige este fenómeno se puede deducir examinando el caso mencionado de obtener cierto número de *1's* en *n* lanzamientos. Si se estipula que sean tres *1's* en 6 lanzamientos, los casos posibles en que aparecen tres *1's* son:

AAAFFF	AFAFAF	FAAAFF	FAFFAA
AAFAFF	AFAFFA	FAAFAF	FFAAAF
AAFFAF	AFFAAF	FAAFFA	FFAAFA
AAFFFA	AFFAFA	FAFAAF	FFAFAA
AFAAFF	AFFFAA	FAFAFA	FFFAAA

El resultado acertado (1) se indicará con la letra *A* y no acertado (diferente a 1) se indicará con la letra *F*.

La probabilidad de cualquier evento, por ejemplo *AFFAFA*, es:

$px(1-p)x(1-p)xpx(1-p)xp$

Esta probabilidad es igual para cualquiera de los eventos señalados, pues sólo cambia el orden. Como estos eventos son excluyentes, entonces la probabilidad de obtener tres *1's* en seis lanzamientos es la suma de cada una de esas probabilidades. O sea:

$20x(1/6)^3(1-1/6)^3 = 20x(1/216)x(125/216) = 2{,}500/46{,}656 = 0{,}05358$

En general para calcular el número de *r* casos exitosos de *n* intentos, en cualquier orden, se tiene:

$$C_r^n = \frac{n!}{r!(n-r)!} \tag{24}$$

n se define como *n* factorial y es igual a *1x2x3x4...x(n-1)xn* y el caso especial de 0 es igual a 1.

En *Excel =DISTR.BINOM(num_éxitos (r);intentos (n);prob de éxito (p);acumulado).* En el argumento acumulado, “falso”=valor de la densidad de probabilidad “verdadero”= valor acumulado de la probabilidad.

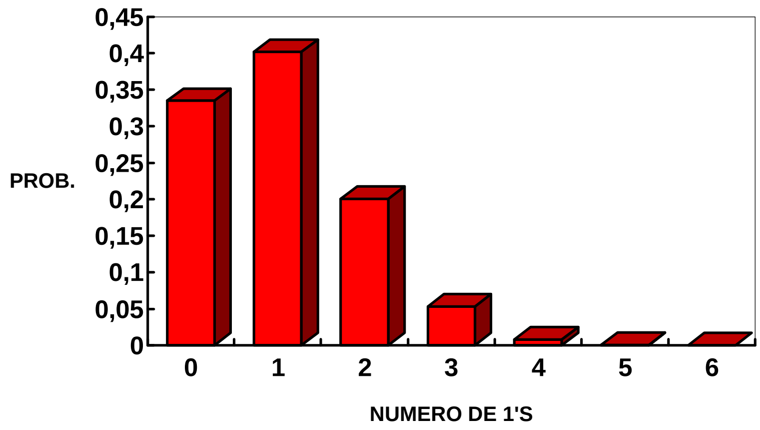
En el ejemplo, obtener tres unos en seis lanzamientos del dado:

	A	**B**
19	*p*=	=1/6 [0,1667]
20	*n*=	6
21	*r*=	3
22	*P*=	***=DISTR.BINOM(B21;B20;B19 “FALSO”)*** [0,05358368]

Si se analizan todos los casos posibles, se tiene:

	A	B	C	D	E
22	n	r	p	ACUMULADO = “VERDADERO”	ACUMULADO = “FALSO”
23	6	6	0.1667	***=DISTR.BINOM(A23;B23;B23; “VERDADERO”)*** 1	***=DISTR.BINOM(A23;B23;B23 “FALSO”)*** [2,1433E-05]
24	6	5	0.1667	**=DISTR.BINOM(A24;B24;B24; “VERDADERO”)** [0,99997857]	**=DISTR.BINOM(A24;B24;B24 “FALSO”)** [0,000643]
25	6	4	0.1667	[0,99933556]	0,00803755
26	6	3	0.1667	0,99129801	0,05358368
27	6	2	0.1667	0,93771433	0,20093879
28	6	1	0.1667	0,73677555	0,40187757
29	6	0	0.1667	0,33489798	0,33489798
30				TOTAL	1

DISTRIBUCION BINOMIAL (NUMERO DE 1'S OBTENIDOS EN 6 LANZAMIENTOS DE UN DADO)



Construir en la hoja de cálculo la tabla y la gráfica anteriores.

La distribución binomial tiene los siguientes parámetros:

Media = p

Varianza = pq = p(1-p)

La distribución de probabilidad continua

Cuando la variable que se está analizando puede tomar cualquier valor entre -∞y ∞de una manera "continua", esto es, que se admite cualquier valor, entero o no dentro de esos límites, entonces se dice que es una variable continua. A diferencia de la distribución discreta, donde cada valor tiene asociada una probabilidad, en este caso cada valor tiene asociado un valor que se llama función de densidad de probabilidad. Esta función de densidad de probabilidad no es un histograma de frecuencia, sino de una curva. La probabilidad se le asigna a un rango de valores y se mide en términos de la proporción del área bajo la curva entre esos dos valores y el área total.

La distribución acumulada es en este caso:

$$F(b) = P(x \le b) = \int_{-\infty}^{b} P(x)dx \qquad (25)$$

La media y la varianza serán:

Media de la población, $m = \int_{-\infty}^{+\infty} xp(x)dx$

(26)

Varianza de la población, $s^2 = \int_{-\infty}^{+\infty} (x-m)^2 p(x)dx$ (27)

La Distribución de Probabilidad Normal

También existen muchas leyes de probabilidad continuas; la más conocida y frecuente es la que se conoce como distribución normal o de Gauss. Esta ley de probabilidad rige muchos fenómenos de la naturaleza.

Esta distribución de probabilidad (función de densidad de probabilidad) se expresa así:

$$P(x) = \frac{e^{-\left(\frac{1}{2}\right)\left(\frac{x-m}{s}\right)^2}}{\sqrt{2ps}} \qquad (28)$$

Donde:

x = Variable aleatoria con distribución normal.

e = Base de los logaritmos naturales, 2.71828183.

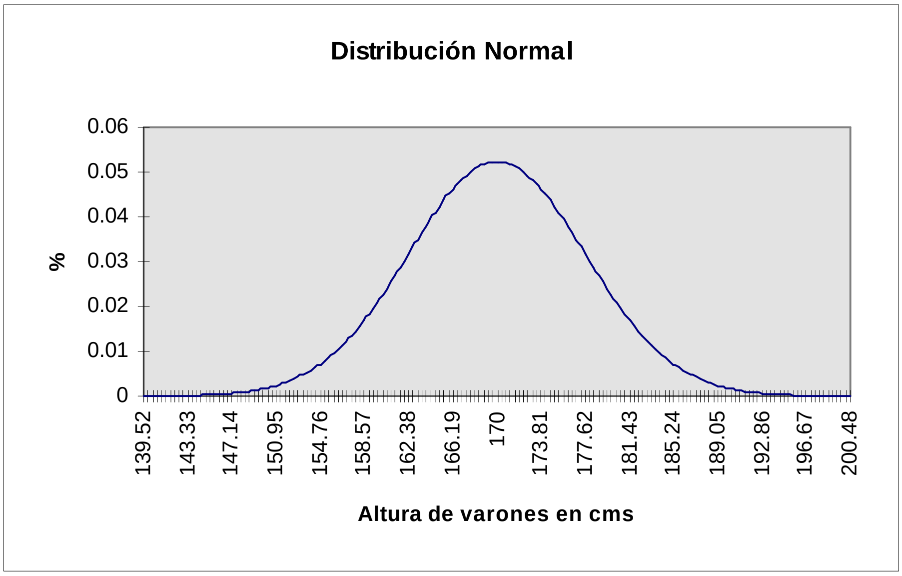
p = Número pi 3.141516....

m = Media de la distribución.

s = Desviación estándar de la distribución.

Esta distribución tiene unas características que la hacen muy especial:

- La moda, la media y la mediana son iguales.
- Es simétrica alrededor de la media.
- La curva tiene dos puntos de inflexión en la media ± una desviación estándar.
- Es asintótica en cero alejándose de media.
- El área bajo la curva es igual a 1.



Hay un caso especial que consiste en "estandarizar" la distribución normal; esto consiste en cambiar el origen de la distribución y suponer que la media es cero y que la desviación estándar es 1. Para lograr esto se hace la siguiente transformación:

$$Z = \frac{x - m}{s} \tag{29}$$

Z es la variable normal estandarizada.

Esto es muy utilizado cuando no se usan hojas de cálculo. Con la normal estandarizada es como se calculan las tablas de la distribución normal acumulada.

Cuando se trabaja con distribuciones continuas se manejan áreas, ya que se supone que cualquier valor puede ocurrir y que entonces los intervalos son infinitesimales. Sin embargo, para hablar de probabilidad se debe referir a un intervalo y, como se estudió en las características de una variable aleatoria, la probabilidad entre $-\infty$ y ∞ es igual a 1. Como se dijo también arriba, cuando se trata de distribuciones continuas, la probabilidad debe calcularse como el área bajo la curva entre dos valores. El área bajo la curva entre $-\infty$ y ∞ es 1. Todo esto significa que la probabilidad de un valor exacto, por ejemplo la probabilidad de que una persona tenga 32 años, 7 meses, 4 días, 3 horas, 3 minutos 22 segundos (inclusive se puede llegar a expresar esto de manera infinetisimal) es cero, puesto que el área entre dos valores iguales (esto es, el mismo valor) es cero.

Muchos fenómenos de la naturaleza pueden ser descritos por la distribución normal o de *Gauss*. Inclusive, algunos fenómenos que no siguen esta ley de probabilidad pueden ser analizados suponiendo que siguen esa distribución, y los resultados, en términos prácticos son bastantes aceptables.

La distribución normal tiene ciertas características que sirven para hacer más fácil su manejo; una de ellas es la simetría —ya mencionada— y que el área debajo de la curva es proporcional, independientemente del fenómeno que se esté analizando y de los valores de sus parámetros (valor esperado o media y desviación estándar) a la probabilidad de los valores que la limitan a cada lado.

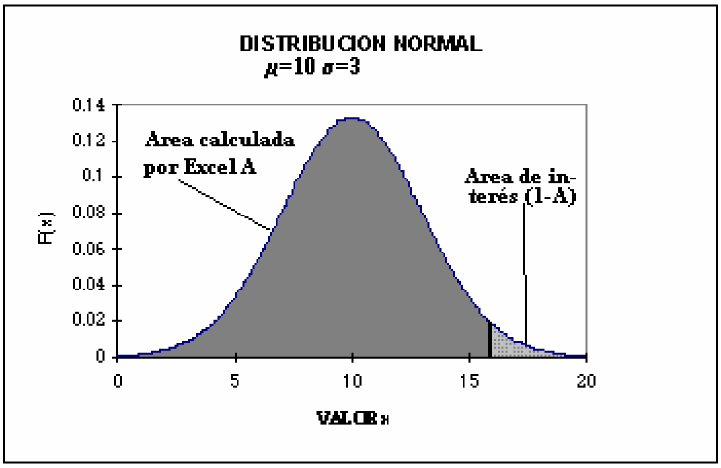
Esto significa que si se tiene información sobre el área bajo la curva normal con parámetros $m = 0$ y $s = 1$ esta información se puede utilizar para otra distribución normal con parámetros diferentes, si se hacen ciertas transformaciones. Ahora bien, esta transformación era válida cuando era necesario recurrir a tablas por la dificultad del cálculo. Hoy, las hojas electrónicas como *Excel*, por ejemplo, traen funciones que no sólo manejan la distribución estandarizada con parámetros $m = 0$ y $s = 1$ -la función es *=DISTR.NORM.ESTAND(z)*- sino que tiene funciones que manejan directamente el valor de la probabilidad deseada, incluyendo los parámetros de la distribución que se estudia. Para este caso, también se utiliza el *Asistente de Funciones* y se aplica la función *=DISTR.NORM(Valor que interesa x;media;desv.estándar;acum).* Si se escribe *Acumulado* en *acum*, entonces arroja el valor acumulado entre -∞? y el valor que interesa, *x*; si no se escribe *Acumulado*, arroja el valor de la densidad de probabilidad, o sea el valor de la función $P(x) = \frac{e^{-\left(\frac{1}{2}\right)\left(\frac{x-m}{s}\right)^2}}{\sqrt{2p}s}$.

Ejemplo

Si se tiene una distribución normal con $m = 10$ y $s = 3$ y se desea calcular la probabilidad que la variable en cuestión tome un valor mayor que 16.

Ahora bien, como se desea saber la probabilidad de que la variable sea mayor que 16 y lo que arroja la función es la probabilidad de que sea menor que 16, entonces se debe restar esta última de 1.

	A	B
1	Valor x.	16
2	Media.	10
3	Desviación estándar.	3
4	Probabilidad de que *x* sea menor que 16.	*=DISTR.NORM(B1;B2;B3;VERDADERO)* [97,72%]
5	Probabilidad de que *x* sea mayor que 16.	=1-B4 [2,28%]



Si se estandariza, esto es que la media se convierte en 0 y la desviación estándar en 1:

$$z = \frac{16-10}{3} = 2$$

Esto es, se supone que la media se traslada a 0 y se calcula cuántas veces está s = 3 en el intervalo entre 10 y 16. Esta transformación equivale a trabajar con una distribución normal con parámetros m = 0 y s = 1, y se debe calcular la probabilidad de que z sea mayor que 2. Si la probabilidad de que sea menor que 2 —o 16 en el problema original— entonces la probabilidad que sea mayor que 2 -o sea, mayor que 16 en la variable original- será 1 – 0,9773 = 0,0227 o sea 2,27%.

	A	B
1	Valor *x*.	16
2	Media.	10
3	Desviación estándar.	3
4	Probabilidad de que *x* sea menor que 16.	*=DISTR.NORM(B1;B2;B3;VERDADERO)* [97,72%]
5	Probabilidad de que *x* sea mayor que 16.	=1-B4 [2,28%]
6	z	=(B1-B2)/B3 [2]
7	Probabilidad de que *z* sea menor que 2.	*=DISTR.NORM.ESTAND(V17)* [97,72%]
8	Probabilidad de que *z* sea mayor que 2.	=1-B8 [2,28%]

Estadística no paramétrica

Hasta ahora nos hemos ocupado de distribuciones que nos permiten "medir" ciertas estadísticas tales como media, varianza, etc. Existen muchas situaciones en la realidad que requieren más que todo verificar si ciertos resultados son independientes o si siguen determinada ley de probabilidad (distribución de probabilidad). También es necesario encontrar herramientas que nos permitan evaluar resultados basados en ordinalidad, más que en valores absolutos. De estos temas se ocupa esta sección. La estadística no paramétrica se utiliza cuando no se cumplen ciertas condiciones rigurosas, por ejemplo, en lo que se conoce como estadística paramétrica se puede requerir que las muestras a estudiar sean independientes y que provengan de distribuciones normales con varianzas iguales. Las pruebas de hipótesis no paramétricas son muy populares debido a dos razones: La primera, es que requieren suposiciones menos restrictivas que las paramétricas y con frecuencia los cálculos son cortos y simples. La segunda, es el hecho de que las pruebas no paramétricas son las más adecuadas cuando se trata de analizar información de muestras que sólo pueden ser ordenadas.

La distribución c^2 (chi cuadrado o ji cuadrado)

La distribución c^2 es muy útil para hacer pruebas de hipótesis, en particular pruebas de bondad de ajuste. Algunas de ellas son las pruebas de normalidad (verificar si una variable tiene distribución normal) y tablas de contingencia que permiten evaluar la independencia de los resultados.

Esta distribución tiene una formulación tan poco amigable como la distribución normal:

$$Y = \frac{e^{-c^2/2}\left(c^2\right)^{(n-2)/2}}{2^{n/2}\left(\frac{n-2}{2}\right)!} \qquad (30)$$

Donde e es la base de los logaritmos naturales (2,7148...), c^2 es la variable *chi* cuadrado y n son los grados de libertad.

La prueba c^2 compara las frecuencias obtenidas de la variable con la frecuencia esperada de esa variable. Si las dos frecuencias son “parecidas” no se rechaza la hipótesis de que la frecuencia observada procede de una distribución dada. La forma general para estas pruebas es la siguiente:

$$\sum_{i=1}^{k} \frac{(F_i - f_i)^2}{f_i} \tag{31}$$

Donde F_i es la frecuencia observada, f_i es la frecuencia teórica o esperada y k es el número de observaciones. En cada caso hay que definir el número de grados de libertad.

Pruebas de normalidad

Muchas pruebas estadísticas están basadas en el supuesto de que la variable que se estudia tenga una distribución normal. Aunque existen varias formas de hacer la prueba de normalidad, sólo se va a presentar la que utiliza la distribución *chi cuadrado.*

En este caso, la frecuencia esperada es la que se obtiene de calcular el valor de la distribución normal entre ciertos valores. Los grados de libertad son *k-3*. Esto se debe a que se tienen que cumplir tres condiciones en el proceso de ajuste:

$$\begin{aligned} \sum f_i &= \sum F_i \\ \overline{X'} &= \sum \frac{F_i X_i}{n} \\ s'^2 &= \sum \frac{F_i x_i^2}{n} \end{aligned} \tag{32}$$

Al tener que satisfacer esas tres condiciones se pierden tres grados de libertad por tanto, los grados de libertad para la prueba serán *k-3*.

Ejemplo

Con una distribución normal cuya media sea 10 y la desviación estándar fuera de 3, se han recolectado 250 observaciones de cierta variable y han sido clasificadas en los rangos que indica la siguiente tabla:

entre	y	Frecuencia absoluta observada	Frecuencia relativa acumulada teórica f_i	Frecuencia relativa teórica	Frecuencia absoluta observada F_i
	-2		0,00%		
-2	1	1	0,13%	0,13%	0,34
1	4	6	2,28%	2,14%	5,35
4	7	33	15,87%	13,59%	33,98
7	10	110	50,00%	34,13%	85,34
10	13	75	84,13%	34,13%	85,34
13	16	15	97,72%	13,59%	33,98
16	19	10	99,87%	2,14%	5,35

Para las pruebas de ajuste *chi cuadrado* se recomienda que por lo menos en cada intervalo o valor, haya por lo menos 5 observaciones. Por tanto, se fusionan los dos primeros intervalos.

Entre	y	Frecuencia absoluta observada	Frecuencia absoluta observada Fi	*Fi--fi*	$\frac{(F_i - f_i)^2}{f_i}$
-2	4	7	5,68751551	1,31248449	0,30287663
4	7	33	33,9762994	-0,975	0,02797906
7	10	110	85,336185	24,663815	7,12832159
10	13	75	85,3361851	-10,3361851	1,25195101
13	16	15	33,9762994	-18,9762994	10,5985627
16	19	10	5,3500237	4,6499763	4,04152968
19	22				
		Prueba chi cuadrado inversa (5%, 5)	11,0704826	Resultado $\sum \frac{(F_i - f_i)^2}{f_i}$	23,3512207

Si los parámetros se estiman a partir de los datos de la muestra, entonces se pierden 3 grados de libertad. Si los parámetros no se estiman, sino que se conocen desde el universo, entonces se pierde un grado de libertad. En este caso los parámetros se conocen, por tanto, el número de grados de libertad será 5 (6-1). La Frecuencia relativa acumulada teórica se calcula con la función de Excel *=DISTR.NORM(valor de X (valor superior del rango); media;desviación estándar;VERDADERO). VERDADERO* indica que se está calculando el valor acumulado entre menos infinito y el valor superior del rango (*X*).

La estadística $\frac{(F_i - f_i)^2}{f_i}$ se debe calcular con base *en las observaciones*, no en la frecuencia relativa. El valor máximo permisible del total es de 11,0704826 y se calcula con la función *=PRUEBA.CHI.INV(probabilidad %; grados de libertad).* En este ejemplo se utilizó un nivel de 5% y 5 grados de libertad. Como el resultado 23,3512207 es mucho menor que el permitido, 11,0704826, entonces *se rechaza* la hipótesis de que las observaciones provienen de una distribución normal.

Tablas de contingencia

Las tablas de contingencia muestran asociaciones entre clasificaciones. La forma más simple de tablas de contingencia es la llamada tablas 2x2. Todas las tablas de contingencia se pueden construir con una opción en el menú *Datos de Excel*, bajo *Asistente de tablas dinámicas*.

Tablas 2x2

La siguiente tabla clasifica a una población de adultos entre filiación política y edad, así:

Edad	Partido A	Partido B	Total
Menor de 30 años	77	323	400
Mayor de 30 años	177	223	400
Total	254	546	800

Lo que se pretende con el análisis de las tablas de contingencia 2x2 es verificar si una clasificación es independiente de la otra. ¿Tiene algún efecto la edad de la población en la escogencia de la filiación política? Esto quiere decir que se trata de una prueba de independencia.

Cuando un elemento se tabula como en la tabla anterior si las frecuencias de cada fila son proporcionales a las de las otras filas o si las frecuencias de cada columna son proporcionales a las de las otras columnas, entonces las dos clasificaciones son independientes una de otra.

Para calcular la frecuencia esperada de cada celda se toma el total de cada columna y se divide en la misma proporción en que están divididos los grandes totales de las filas. Así, la celda *Partido A* – menor de 30 años, tendrá como frecuencia esperada 254x400/800 = 127, y así las demás. Entonces las frecuencias esperadas para cada celda serán:

Edad	*Partido A*	*Partido B*	*Total*
Menor de 30 años	127	273	400
Mayor de 30 años	127	273	400
Total	254	546	800

El cálculo de $\sum_{i=1}^{k} \frac{(F_i - f_i)}{f_i}$ será:

$(77\text{-}127)^2/127 + (177\text{-}127)^2/127 \quad (323\text{-}273)^2/273 + (223\text{-}273)^2/273 = 57{,}7$

Para determinar los grados de libertad se debe tener en cuenta cuáles son las restricciones que se imponen en la tabla 2x2. Estas son:

$f_{11} + f_{12}$ = Total de la fila 1

$f_{21} + f_{22}$ = Total de la fila 2

$f_{11} + f_{21}$ = Total de la columna 1

$f_{12} + f_{22}$ = Total de la columna 2

Como una de ellas se puede obtener de las otras, entonces los grados de libertad que se pierden son 3 y los grados de libertad para la distribución valen 1 (2-1)(2-1).

El máximo valor permitido con un nivel de 5% es de 3,84, por tanto se rechaza la hipótesis de independencia y se dice que la edad sí es determinante de la filiación política.

Tablas de contingencia rxc

El análisis de las tablas de contingencia se puede generalizar para cualquier número de grupos de clasificación en los dos sentidos. En ese caso se dice que son tablas de contingencia rxc y los grados de libertad serán (r-1)x(c-1). El procedimiento es similar al presentado.

Pruebas de signo y orden

Estas son pruebas típicas no paramétricas. Miran más a las relaciones entre los valores que los valores mismos. Por ejemplo, lo importante no es si los valores de unas muestras son 2 y 5, sino que el valor de la muestra dos es más alto que el de la uno.

Prueba de signos

Cuando se trata de medir la diferencia en la media de dos poblaciones en condiciones paramétricas, se requiere que las dos muestras sean independientes y que provengan de universos normales con igual varianza. Si alguna de estas dos condiciones no se cumple hay que usar una prueba no paramétrica llamada del signo. Esto es, se comparan los valores y se determina cuál es el mayor y con base en el número de signos positivos (si fuera mayor) o negativos (si fuera menor o viceversa) se analiza la información.

Ejemplo

Se registra el promedio 25 estudiantes antes y después de tomar un taller de métodos de estudio. Se trata de estudiar si ese taller aumenta o no el promedio.

Efecto de un taller de métodos de estudio sobre el rendimiento

Sujeto	Antes	después	Signo del cambio		
1	3,7	3,8	+	1	1
2	3,7	3,5	-	0	1
3	3,55	3,6	+	1	1
4	3,4	3,35	-	0	1
5	3,7	3,65	-	0	1
6	3,5	3,55	+	1	1
7	3,6	3,65	+	1	1
8	3,7	3,8	+	1	1
9	3,45	3,5	+	1	1
10	3,4	3,5	+	1	1
11	3,55	3,6	+	1	1
12	3,45	3,6	+	1	1
13	3,65	3,45	-	0	1
14	3,35	3,4	+	1	1
15	3,55	3,65	+	1	1
16	3,6	3,65	+	1	1
17	3,6	3,5	-	0	1
18	3,6	3,6	No cambia	0	0
19	3,45	3,6	+	1	1
20	3,65	3,45	-	0	1
21	3,35	3,4	+	1	1
22	3,4	3,5	+	1	1
23	3,45	3,45	No cambia	0	0
24	3,6	3,75	+	1	1
25	3,7	3,65	-	0	1
				16	23

Si el taller no tuviera efecto sobre el promedio se esperaría que la mitad de las veces el promedio aumentara y la otra mitad disminuyera. Las muestras con igual valor de promedio se desechan.

Para una muestra de 23, se espera que la media de cambios de signo sea de:

Media: pxn 11,5.

La desviación estándar sería: $(np(1\text{-}p))1/2$ 2,39791576

Si la hipótesis es cierta, entonces el promedio debería ser 11,5. La pregunta ahora es: ¿Cuál es la probabilidad de que el número de signos + sea de 16 o mayor?

Usamos la normal con media 11,5 y desviación estándar 2,39791576; probabilidad de que sea mayor 3,03%.

Esto significa que la probabilidad de que esto ocurra por azar es de 3,03%.

Si aceptamos un error de 5% el valor obtenido está dentro de ese margen de error y no rechazaríamos la hipótesis de que tomar un curso de métodos de estudio influye en el rendimiento.

La binomial se puede aproximar a la normal cuando n>=10.

Ejemplo

Un panel debe hacer una prueba de sabor de dos productos. Le asigna 1 al sabor que prefiere y 1 al otro.

Jurado	Sabor uno	Sabor dos	Signo del cambio		
A	1	0	+	1	1
B	1	0	+	1	1
C	0	1	-	0	1
D	1	0	+	1	1
E	1	0	+	1	1
F	1	0	+	1	1
G	0	1	-	0	1
H	1	0	+	1	1
I	1	0	+	1	1
J	0	1	-	0	1
K	1	0	+	1	1
L	0	1	-	0	1
				8	12

Si no hubiera preferencia sobre los sabores se esperaría que la mitad de las veces el panel *prefiriera* el uno y la otra mitad el dos. Las muestras con igual valor de calificación se desechan.

Para una muestra de 12 se espera que la media de cambios de signo sea: media *pxn* = 6; la desviación estándar es desviación estándar = (*np*(*1-p*))1/2 = 1,73205081.

Si la hipótesis es cierta, entonces el promedio signos positivos debería ser 6.

La pregunta ahora es: ¿Cuál es la probabilidad de que el número de signos + sea de 8 o mayor?

Usamos la normal con media 6 y desviación estándar 1,7320508.

Probabilidad de que sea mayor: 12,41%.

Esto significa que la probabilidad de que esto ocurra por azar es de 12,41%.

Si aceptamos un error de 5% el valor obtenido no está dentro de ese margen de error y rechazaríamos la hipótesis de que no hay preferencia entre los jurados del panel por ningún sabor.

La binomial se puede aproximar a la normal cuando $n>=10$

Prueba U de Mann-Whitney

Cuando las muestras son independientes y tienen varianzas diferentes se puede utilizar esta prueba.

Ejemplo

Se aplica un cierto examen a estudiantes de la jornada diurna y ese mismo examen a estudiantes de la jornada nocturna. Se quiere probar si el rendimiento de los dos grupos es diferente o no.

Nota	Jornada	Nota	Jornada
3,50	Diurno	3,60	Nocturno
3,40	Diurno	3,35	Nocturno
3,65	Diurno	3,70	Nocturno
4,05	Diurno	3,25	Nocturno
3,30	Diurno	3,15	Nocturno
2,80	Diurno	3,85	Nocturno
3,10	Diurno	3,55	Nocturno
3,75	Diurno	3,00	Nocturno
4,15	Diurno	3,80	Nocturno
2,40	Diurno	3,05	Nocturno
		3,20	Nocturno

Esta prueba exige los siguientes pasos:

Paso 1

Asignar un orden a toda la muestra combinada (diurno y nocturno). Así, 1 a la más baja nota, 2 a la siguiente, etc. En el ejemplo, 1 a 2,4, 2 a 2,8, 3 a 3,0, etc.

Nota	Jornada	Orden
2,40	Diurno	1
2,80	Diurno	2
3,00	Nocturno	3
3,05	Nocturno	4
3,10	Diurno	5
3,15	Nocturno	6
3,20	Nocturno	7
3,25	Nocturno	8
3,30	Diurno	9
3,35	Nocturno	10
3,40	Diurno	11
3,50	Diurno	12
3,55	Nocturno	13
3,60	Nocturno	14
3,65	Diurno	15
3,70	Nocturno	16
3,75	Diurno	17
3,80	Nocturno	18
3,85	Nocturno	19
4,05	Diurno	20
4,15	Diurno	21

Paso 2

Se suman todos los rangos de cada grupo:

Nota	Jornada	Orden	Nota	Jornada	Orden
2,40	Diurno	1	3,00	Nocturno	3
2,80	Diurno	2	3,05	Nocturno	4
3,10	Diurno	5	3,15	Nocturno	6
3,30	Diurno	9	3,20	Nocturno	7
3,40	Diurno	11	3,25	Nocturno	8
3,50	Diurno	12	3,35	Nocturno	10
3,65	Diurno	15	3,55	Nocturno	13
3,75	Diurno	17	3,60	Nocturno	14
4,05	Diurno	20	3,70	Nocturno	16
4,15	Diurno	21	3,80	Nocturno	18
			3,85	Nocturno	19
	R1	113		R2	118

$$U = n_1 n_2 + \left(\frac{n_1(n_1+1)}{2} \right) - R_1 \qquad (33)$$

o:

$$\boxed{U = n_1 n_2 + \left(\frac{n_2(n2+1)}{2} \right) - R2} \qquad (34)$$

Con la primera fórmula, *n1*= 10, *n2* = 11, *R1* = 113

U= 52

Paso 4

Se determina la media y la desviación estándar de *U*

Media:

$$\boxed{E(U) = \frac{n_1 n_2}{2}} \qquad (35)$$

Media: 55.

Desviación estándar:

$$\boxed{s_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \qquad (36)$$

Desviación estándar:14,2009389

Si *n1* y *n2* son mayores que 8, entonces se puede aproximar a la normal. La probabilidad de que el valor de *U*>=52, sea producto del azar.

Probabilidad *U*>= 52   58,37%.

Si se acepta un nivel de error de 5%, entonces la prueba indicaría que no hay evidencia de que el rendimiento es el mismo.

Observaciones:

1) Si hay empates, los valores iguales reciben el rango promedio de sus rangos empatados (por ej. Si los valores 5º y 6º están empatados ambos reciben entonces el rango 5,5) y la que sigue recibe el rango siguiente (por ej. 7º)

2) *U* tiene distribución aproximadamente normal sólo cuando *n1* y *n2* son >= 8. Si esto no se cumple la normal no sirve.

Prueba H de Kruskal-Wallis

Es una generalización de la *Prueba U de Mann-Whitney*. Se utiliza para examinar la hipótesis nula de que varias muestras independientes pertenecen a poblaciones idénticas. Se asignan los ordenamientos a cada observación teniendo en cuenta todas las muestras. Al menor 1, al siguiente 2, etc. Con esos datos se calcula la estadística *H*:

$$H = \frac{12}{n(n+1)}\left(\frac{R_1^2}{n_1}+\frac{R_2^2}{n_2}+\cdots+\frac{R_k^2}{n_k}\right)-3(n+1) \qquad (37)$$

Donde k es el número de muestras, R_k es la suma de los rangos de la k-ésima muestra, y n es el número total de observaciones ($n_1+n_2+...+n_k$)

Si se supone que la hipótesis nula es verdadera y que cada muestra consiste de por lo menos 5 observaciones, *H* tiene una distribución que se puede aproximar a la *Chi-cuadrado* con (k-1) grados de libertad.

Ejemplo

Se quiere probar una metodología educativa con tres grupos de estudiantes y se mide su mejora porcentual en el promedio acumulado:

Aumento %	Metodología	Aumento %	Metodología	Aumento %	Metodología
21,65%	A	24,79%	B	12,54%	C
18,19%	A	31,39%	B	14,74%	C
16,62%	A	34,22%	B	16,62%	C
27,62%	A	25,42%	B	17,88%	C
23,85%	A	29,19%	B	12,85%	C
20,08%	A	30,76%	B	10,34%	C
		34,85%	B	21,65%	C
				19,14%	C
				9,08%	C

Aumento %	Met.	Orden	Aumento %	Met.	Orden	Aumento %	Met	Orden
16,62%	A	6,5	24,79%	B	15	9,08%	C	
18,19%	A	9	25,42%	B	16	10,34%	C	
20,08%	A	11	29,19%	B	18	12,54%	C	
21,65%	A	12,5	30,76%	B	19	12,85%	C	
23,85%	A	14	31,39%	B	20	14,74%	C	
27,62%	A	17	34,22%	B	21	16,62%	C	6
			34,85%	B	22	17,88%	C	
						19,14%	C	
						21,65%	C	12
	R1	70		*R2*	131		*R3*	
n1	6		*n3*	7		*n3*	9	
n=	22							

H = 15,6327875

Valor de *chi cuadrado* a 1% = 9,21035104

Como H es mayor entonces se rechaza que las tres metodologías no tienen igual efectividad.

Correlación de orden

El coeficiente de correlación, como ya se sabe, va a permitir medir la asociación entre dos variables, en este caso, de orden de una variable. En este caso se utilizará el coeficiente de correlación de orden o rango de Spearman. Por ejemplo, para medir calificaciones de evaluación de un grupo de personas o productos, a veces puede interesar si los evaluadores han sido coherentes en sus evaluaciones del mismo grupo de sujetos. Por ejemplo, la evaluación que hace un director de departamento comparado con la evaluación del decano.

Ejemplo

Tanto el decano como el director del departamento, han evaluado a los profesores y su resultado ha sido ordenado de acuerdo con las calificaciones de cada uno, así:

	Evaluación de	
Profesor	Decano	Director
A	7	8
B	5	6
C	4	5
D	3	4
E	2	1
F	6	3
G	8	10
H	10	9
I	9	7
J	1	2

El coeficiente r_s de *Spearman* está definido como:

$$r_s = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)} \tag{38}$$

Donde *n* es el número de observaciones pareadas y *d* es la diferencia entre cada par de rangos.

Profesor	Decano	Director	d	d^2
A	7	8	1	1
B	5	6	1	1
C	4	5	1	1
D	3	4	1	1
E	2	1	-1	1
F	6	3	-3	9
G	8	10	2	4
H	10	9	-1	1
I	9	7	-2	4
J	1	2	1	1
10			suma	24
r_s	0,85			

Ejemplo

Se quiere saber si existe alguna relación entre la prueba del ICFES y el rendimiento de los estudiantes.

Estudiante	Promedio ponderado	Puntaje ICFES
1	3,20	253
2	3,96	313
3	3,02	239
4	4,13	327
5	2,69	213
6	3,89	308
7	4,34	344
8	4,66	350
9	4,78	330
10	4,80	321
11	4,05	322
12	3,66	321
13	4,25	318
14	3,51	278
15	3,61	286
16	3,57	261
17	4,40	320
18	3,08	293
19	4,60	322
20	3,15	222
21	4,10	313
22	4,42	338
23	4,75	352
24	4,80	322
25	4,60	339
26	4,16	295

Estudiante	Promedio ponderado	Puntaje ICFES	Orden de promedio	Orden de puntaje	d	d^2
24	4,80	322	1	9	8	64
10	4,80	321	1	11,5	10,5	110,25
9	4,78	330	3	6	3	9
23	4,75	352	4	1	-3	9
8	4,66	350	5	2	-3	9
19	4,60	322	6,5	9	2,5	6,25
25	4,60	339	6,5	4	-2,5	6,25
22	4,42	338	8	5	-3	9
17	4,40	320	9	13	4	16
7	4,34	344	10	3	-7	49
13	4,25	318	11	14	3	9
26	4,16	295	12	18	6	36
4	4,13	327	13	7	-6	36
21	4,10	313	14	15,5	1,5	2,25
11	4,05	322	15	9	-6	36
2	3,96	313	16	15,5	-0,5	0,25
6	3,89	308	17	17	0	0
12	3,66	321	18	11,5	-6,5	42,25
15	3,61	286	19	20	1	1
16	3,57	261	20	22	2	4
14	3,51	278	21	21	0	0
1	3,20	253	22	23	1	1
20	3,15	222	23	25	2	4
18	3,08	293	24	19	-5	25
3	3,02	239	25	24	-1	1
5	2,69	213	26	26	0	0

N = 26 Suma total = 485,50.

r_s = 0,83.

Significancia estadística de r_s

Hipótesis: No existe relación entre los ordenamientos. Si n>=25, entonces se podría suponer distribución normal. La media de *rs* es 0 y desviación estándar es:

$$s_{r_s} = \frac{1}{\sqrt{n-1}} \qquad (39)$$

= 0,19611614

Probabilidad de que r_s sea mayor en valor absoluto que 0,83 = 2,11379*E*-05.

Esto significa que el coeficiente r_s es significativo al 1%.

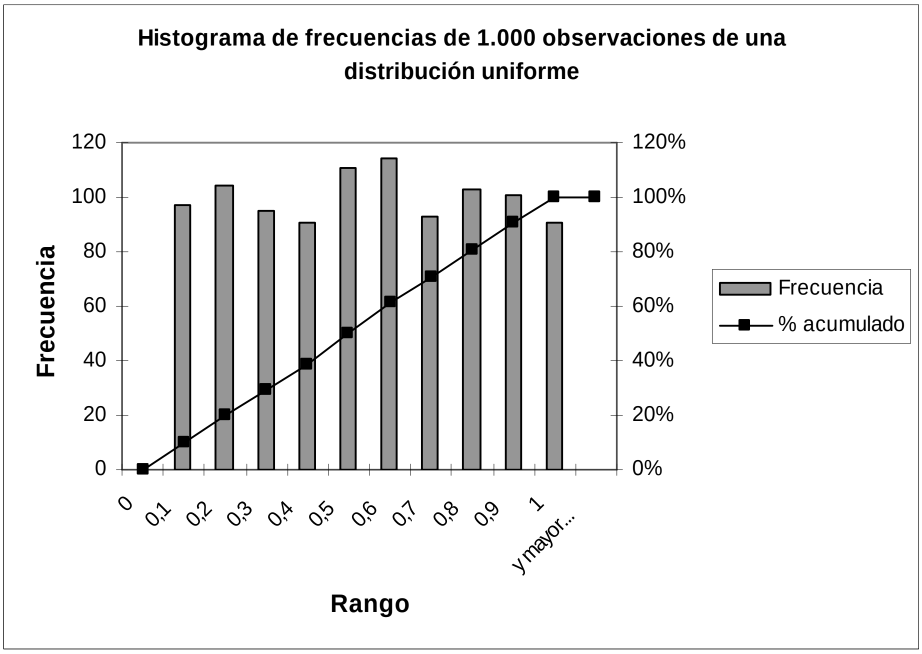
Muestreo aleatorio

En el archivo *Estadística.XLS* hay una hoja que se llama *aleatorio*. Allí aparecen 1.000 números aleatorios, entre 0 y 1. Un número aleatorio tiene como característica que tiene igual probabilidad de "salir" que cualquier otro. Un ejemplo simple es una bolsa con 10 bolas numeradas del 0 al 9. Esos son 10 dígitos que si sacamos al azar cualquier bola, cualquiera de los números tiene igual probabilidad de salir. En este caso, la probabilidad es de 1/10. Si se reemplaza la bola y siempre sacamos bolas de la bolsa con las 10 numeradas como se indicó, estamos ante un mecanismo de generación, números aleatorios. Esta distribución se conoce como distribución uniforme.

Cuando se hace el histograma de frecuencias de los 1.000 se encuentra lo siguiente:

Rangos	*Frecuencia absoluta*	*% acumulado*
0	0	,00%
0,1	97	9,70%
0,2	104	20,10%
0,3	95	29,60%
0,4	91	38,70%
0,5	111	49,80%
0,6	114	61,20%
0,7	93	70,50%
0,8	103	80,80%
0,9	101	90,90%
1	91	100,00%
y mayor...	0	100,00%

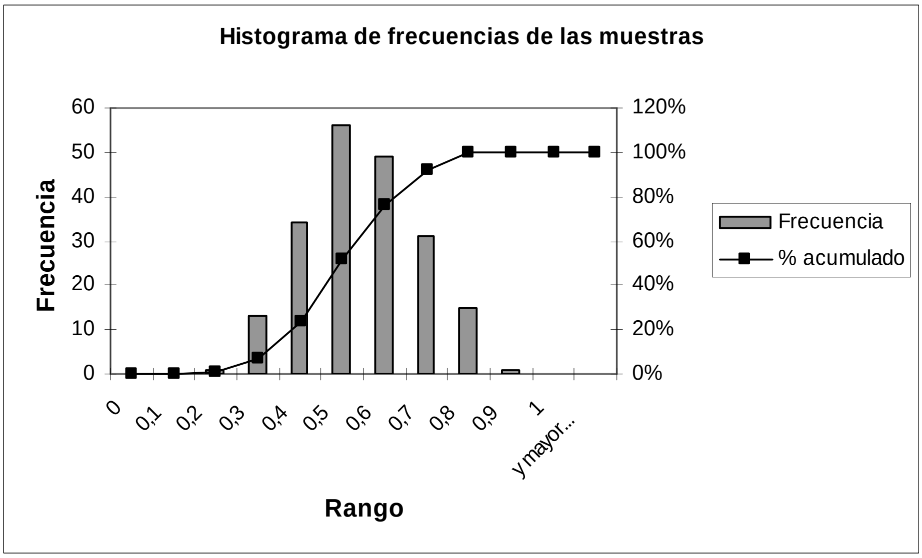
Y su gráfica es:



En esta gráfica se observa que la frecuencia es cercana a 100, por tratarse de una distribución uniforme. Es fácil intuir que la media de esta distribución es 0,5. Si se dibuja el histograma de frecuencias de los promedios de las muestras, se encuentra lo siguiente:

Rangos	*Frecuencia absoluta*	*% acumulado*
0	0	,00%
0,1	0	,00%
0,2	1	,50%
0,3	13	7,00%
0,4	34	24,00%
0,5	56	52,00%
0,6	49	76,50%
0,7	31	92,00%
0,8	15	99,50%
0,9	1	100,00%
1	0	100,00%
y mayor...	0	100,00%

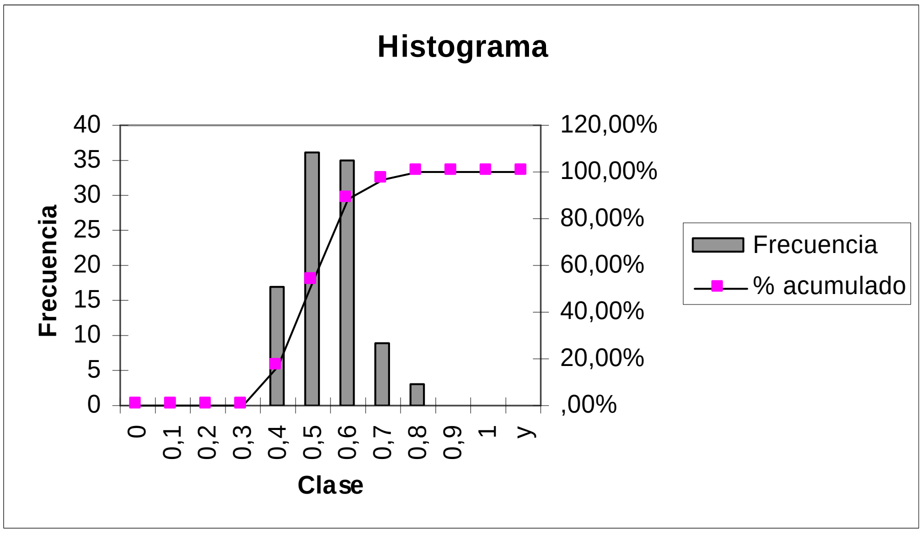
Y en la gráfica:



Si la muestra es de tamaño 10, los resultados son los siguientes:

Clase	Frecuencia	% acumulado
0	0	,00%
0,1	0	,00%
0,2	0	,00%
0,3	0	,00%
0,4	17	17,00%
0,5	36	53,00%
0,6	35	88,00%
0,7	9	97,00%
0,8	3	100,00%
0,9	0	100,00%
1	0	100,00%
y mayor...	0	100,00%

En gráfica:



Observe que ahora el histograma es más estrecho. ¿Qué ha sucedido? Pues que al tomar muestras de una distribución uniforme, y al calcular, además, el promedio de cada muestra, éste tiende a comportarse como una distribución normal.

Esta es una propiedad muy importante que se debe tener en cuenta. Se puede observar además que el rango de variación es mucho menor para los promedios de las muestras que para los datos originales. Esto se puede apreciar en la siguiente tabla:

	Muestras		Valores originales	
Rangos	**Frecuencia absoluta**	*% acumulado*	*Frecuencia absoluta*	*% acumulado*
0	0	,00%	0	,00%
0,1	0	,00%	97	9,70%
0,2	1	,50%	104	20,10%
0,3	13	7,00%	95	29,60%
0,4	34	24,00%	91	38,70%
0,5	56	52,00%	111	49,80%
0,6	49	76,50%	114	61,20%
0,7	31	92,00%	93	70,50%
0,8	15	99,50%	103	80,80%
0,9	1	100,00%	101	90,90%
1	0	100,00%	91	100,00%
y mayor...	0	100,00%	0	100,00%

Observado esto, se puede plantear una ley estadística que dice que la distribución de probabilidad del promedio de las muestras tomadas de una distribución cualquiera, tiende a ser normal con media igual a la media de la distribución original, y la desviación estándar igual a la desviación estándar de la distribución original dividida por la raíz cuadrada del tamaño de la muestra.

Con los datos disponibles se tiene:

Desviación estándar total de observaciones: 0,28472219.

Desviación estándar de la muestra: 0,13323466.

Promedio del total de observaciones: 0,4977773.

Promedio de la muestra: 0,4977773.

Los valores teóricos de la media y de la desviación estándar de la distribución uniforme son:

$$Media\ m = \frac{Valor\ máximo + valor\ mínimo}{2} = \frac{1-0}{2} = 0{,}5$$

$$Desviación\ estándar\ s\ = \frac{Valor\ máximo - valor\ mínimo}{\sqrt{12}} = \frac{1}{\sqrt{12}} = 0{,}28867513$$

Los valores no coinciden exactamente porque los de la tabla fueron calculados de una muestra de 1.000 observaciones y no es el universo total.

$$Desviación\ estándar\ del\ promedio\ de\ la\ muestra\ s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0{,}28867513}{\sqrt{5}} = 0{,}12909944$$

Cuando se toma una muestra, se espera entonces que los resultados obtenidos por la muestra fluctúen alrededor de su media y que cerca del 100% de los valores se encuentren entre $\overline{X} \pm 3s_{\bar{x}}$. Esta variación hacia arriba o hacia abajo de la media es el error permitido, y éste puede ser definido a voluntad. De manera que se puede construir una tabla que muestre el porcentaje de las muestras que caen entre dos valores:

Entre	% observaciones
$\overline{X} \pm s_{\overline{X}}$	68,27
$\overline{X} \pm 2s_{\overline{X}}$	95,45
$\overline{X} \pm 3s_{\overline{X}}$	99,73

Esto indica la confiabilidad del muestreo. De modo que es posible definir la confiabilidad del muestreo (95%, 90%, etc.), el error (e) que se acepta y, conociendo (o calculando aproximadamente) la desviación estándar de la población, se puede definir el tamaño de la muestra.

Si se denomina el número de desviaciones estándar por arriba y por debajo de la media como Z, entonces se puede expresar el error como:

$$\begin{aligned} &error\ e = Zs_{\overline{X}} = Z\frac{s}{\sqrt{n}} \\ &n = \frac{Z^2 s^2}{e^2} \end{aligned} \qquad (40)$$

Este valor de n es válido cuando se tiene una población infinita, de modo que no hay agotamiento del universo. Cuando se trata de una población finita, se incurre en agotamiento del universo y la probabilidad de la primera muestra es diferente de todas las subsiguientes.

En el caso de un universo finito de tamaño N, es necesario hacer un ajuste, así:

$$Tamaño\ de\ la\ muestra\ n_m = \frac{n}{\left(1 + \frac{n}{N}\right)} \qquad (41)$$

Hay un caso de especial interés y es la distribución binomial que rige experimentos tales como encuestas de opinión o de mercado, donde el resultado a medir es un porcentaje. En ese caso, se sabe ya que los parámetros de la distribución son *p* y *pq*. Entonces el cálculo del tamaño de la muestra será:

$$error\ e = Zs_{\bar{p}} = Z\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$
$$n = \frac{Z^2 p(1-p)}{e^2} \tag{42}$$

En este caso el error *e* es un porcentaje en relación con el porcentaje que se desea medir. Como no siempre se puede calcular el valor de *p*, se adopta una posición conservadora para *p*. Con *p*=0,5 se garantiza la máxima varianza, por tanto, el cálculo de n resulta el mayor posible, dados unos niveles de error y de confiabilidad.

Por ejemplo, si se desea hacer una encuesta sobre un universo finito de 400 elementos, se desea una confiabilidad de 95% y se acepta un error en el cálculo de la respuesta a una determinada pregunta de ±2%. Entonces esto debe interpretarse así: el valor de *Z* debe ser tal que el área bajo la curva normal sea 95%, pues cobija valores por encima y por debajo de la media. Esto quiere decir que se debe encontrar un *Z* tal que el área desde menos infinito y *Z* sea de 97,5%. Esto se encuentra con la función de *Excel*

=*DISTR.NORM.ESTAND.INV(0,975)*. Y arroja como resultado 1,96. ¡De aquí sale el famoso 1,96!.

Por otro lado, decir que se acepta un error de 2% significa que si la respuesta a la pregunta indica 10%, el verdadero valor está entre 8% y 12%, con una probabilidad de 95%. Con estos datos, el tamaño de la muestra, si el universo fuera infinito, sería de 2.401. Para un universo finito de 400 elementos, el tamaño sería de 343. En la siguiente tabla se puede observar que a medida que el decisor está dispuesto a tolerar un mayor error, entonces la muestra se reduce, tanto para el universo infinito como para el finito (400 elementos en este ejemplo):

Error	Muestra (n) universo infinito	Muestra (n_m) universo finito N=400
1%	9.603,61861	384,005788
2%	*2.400,90465*	*342,875599*
3%	1.067,06873	290,938989
4%	600,226163	240,036178
5%	384,144744	195,956039

Un análisis similar puede hacerse fijando el error y variando la confiabilidad. Se deja esto como ejercicio al lector. ¿Qué se espera del tamaño de la muestra si se desea más confiabilidad? ¿menos confiabilidad?

En últimas todo termina siendo un problema de costos: Cuánto se está dispuesto a pagar (por hacer una encuesta con mayor o menor cobertura) por reducir el error o aumentar la confiabilidad.

Métodos de pronóstico

...Aureliano saltó once páginas para
no perder el tiempo en hechos demasiado
conocidos, y empezó a descifrar
el instante que estaba viviendo,
descifrándolo a medida que lo vivía,
profetizándose a sí mismo en el acto
de descifrar la última página
de los pergaminos,...
GABRIEL GARCÍA MÁRQUEZ, CIEN AÑOS DE SOLEDAD

La Humanidad ha tratado siempre de predecir el futuro. Basta recordar todos los intentos de las tribus primitivas de controlar -prediciendo- los fenómenos naturales o el oráculo de Delfos en Grecia. Así mismo, los decisores se enfrentan día a día con la necesidad de tomar decisiones hoy con consecuencias futuras; desearían, sin duda, tener, como Aureliano Buendía, los pergaminos de Melquíades que les describieran, hoja por hoja, día por día, las consecuencias de sus decisiones del momento. Sin embargo, esto no ha sido posible hasta ahora, pero hay quienes ofrecen bolas de cristal y muchos otros métodos mientras aparecen los pergaminos.

Las técnicas de pronóstico son una herramienta necesaria para la planeación macro y microeconómica. Para el caso del gerente su quehacer básico es la toma de decisiones con consecuencias futuras y por lo tanto debe elaborar estimativos de lo que sucederá en el futuro. Por otro lado, debe prever escenarios que le permitan anticiparse a las posibles eventualidades que le indicarán la conveniencia o inconveniencia de una alternativa. En particular para analizar decisiones de inversión es necesario hacer estimativos de muy diversas variables: precios, tasas de interés, volúmenes de venta o de producción, etc., por lo tanto, es necesario que el analista conozca, por lo menos la existencia de ciertas técnicas que le ayuden en esta tarea.

Para elaborar pronósticos se pueden encontrar dos grandes clases de modelos: causales y de series de tiempo. Los primeros tratan de encontrar las relaciones de causalidad entre diferentes variables, de manera que conociendo o prediciendo alguna o algunas de ellas, se pueda encontrar el valor de otra. En el segundo caso no interesa encontrar esas relaciones, sino que se requiere solamente encontrar los posibles valores que asumirá una determinada variable. En todos los casos siempre se hace uso de la información histórica, ya sea para predecir el comportamiento futuro o para suponer que el comportamiento histórico se mantendrá hacia el futuro y sobre esta base hacer los estimativos. Aquí se estudiarán algunos métodos de pronóstico de series de tiempo. No se pretende ser exhaustivo sobre el tema porque el alcance de este texto no lo considera y porque además, existen textos especializados sobre pronósticos (ver bibliografía al final del capítulo).

Se debe tener presente que no existe ningún método de pronóstico infalible; lo que hacen estos procedimientos es estimar un valor posible, pero siempre sujeto a errores. Si el fenómeno que se va a pronosticar fuera determinístico, solo bastaría utilizar la ley matemática que lo rige y predecir con exactitud el resultado; este sería el caso de fenómenos físicos, como por ejemplo la caída libre de un cuerpo. En el proceso de toma de decisiones se involucra el comportamiento humano, por ejemplo, a través de las decisiones de los individuos a quienes está dirigida un determinado producto o servicio; las decisiones del mercado están compuestas por muchísimas decisiones individuales, imposibles de predecir con exactitud.

La mayoría de los datos incluyen combinaciones de estas tendencias y se deben generar procedimientos para separarlos. Existen otras clases de pronósticos denominados cualitativos o de pronóstico tecnológico, tales como el Método Delphi. Este método busca, a través de múltiples rondas o iteraciones donde se comparte la información, encontrar consenso sobre valores o escenarios posibles.

Se hace énfasis en que no hay un método de pronóstico perfecto, aunque se podría construir un modelo que ajuste perfectamente los datos que se tienen de un fenómeno; sin embargo, esto no es recomendable puesto que el elemento aleatorio o de error siempre estará presente y será impredecible y es mejor identificar los patrones predecibles y asumir el error que se presente que tratar de introducir en el modelo el elemento error que, se repite, es completamente impredecible e inevitable. En otras palabras, cualquier estimativo implica un cierto grado de error inevitable.

Existen muchos métodos de pronóstico y en esta nota no se hará una revisión exhaustiva de ellos. Además, para calificar la bondad de cada uno de ellos se debe acudir al método de los mínimos cuadrados, esto es, se considera el mejor método aquel que minimiza la suma de los cuadrados de los errores (diferencias entre el valor estimado y el observado).

Métodos de Suavización

Dentro de los métodos de suavización se pueden considerar tres categorías: a) Promedios móviles, b) suavización exponencial y c) otros.

Promedios móviles.

Esta técnica consiste en tomar un grupo de valores observados, calcularle el promedio y utilizarlo como pronóstico para el siguiente período. Sólo sirve para pronosticar un sólo período: el siguiente. Se debe especificar el número de observaciones que se tomarán; se llama móvil porque siempre se toman las N últimas observaciones para hacer el pronóstico.

Se pueden considerar promedios móviles simples y promedios móviles lineales. En el primer caso se toman los N últimos datos y se calcula el promedio; en el segundo caso se construyen además promedios de los promedios y con ellos se establece una ecuación lineal que permite elaborar el pronóstico.

Para el caso de los promedios móviles simples, algebráicamente se representa así:

$$F_{t+1} = (X_t + X_{t-1} + \ldots + X_{t-N+1})/N$$

$$F_{t+1} \frac{\sum_{i=t-N+1}^{t} X_t}{N}$$

Este método puede utilizarse cuando se sabe que los datos son estacionarios. La ventaja sobre el promedio total es que permite ajustar el valor de N para que responda al comportamiento de los datos.

Ejemplo:

Compras realizadas por los clientes que entran a una tienda. Se va a utilizar el promedio móvil con N= 3.

AÑO	COMPRAS	PRONOSTICO
1970	10	
1971	11	10.00
1972	15	10.50
1973	11	12.00
1974	15	12.33
1975	11	13.67
1976	9	12.33
1977	14	11.67
1978	11	11.33
1979	16	11.33
1980	12	13.67
1981	14	13.00
1982		14.00

Suavización exponencial

Existen muchos métodos de suavización exponencial: simple, de tasa de respuesta de adaptación, método de Brown de un solo parámetro, método de Holt de dos parámetros, método cuadrático de Brown, etc. Aquí se considerarán un método de suavización: suavización exponencial simple.

Suavización exponencial simple.

Este método consiste en asignar un peso a la última información (dato) disponible y al último pronóstico, el cual, a su vez, contiene la información pasada, así:

$$F_{t+1} = ?\, X_t + (1-a)F_t$$

Para F_2, se tiene:

$$F_2 = F_1$$

Otra forma de expresar el pronóstico es:

$$F_{t+1} = F_t + a\ e$$

donde e es el error incurrido en el último pronóstico.

Ejemplo: Datos de demanda de un producto Se utilizará suavización exponencial simple con a = .3

CLIENTE	DESPACHOS	PRONOSTICO
1970	628	
1971	424	628.00
1972	613	566.80
1973	620	580.66
1974	974	592.46
1975	550	706.92
1976	487	659.85
1977	408	607.99
1978	691	547.99
1979	872	590.90
1980	738	675.23
1981	767	694.06
1982		715.94

Los métodos hasta aquí presentados son muy adecuados para pronosticar el siguiente período; no se recomiendan para hacer predicciones a largo plazo.

Otros métodos de suavización

Solo se mencionarán otros métodos de suavización existentes y no menos importantes: método de control de adaptación de Chow, método de suavización de tres parámetros de Box y Jenkins, método multiplicativo de Winter y el sistema de monitoreo de Trigg.

Métodos de Tendencia

Uno de los métodos más conocidos, pero también de los más mal utilizados es la regresión lineal. En cualquier curso de Presupuesto es tema obligado. Sin embargo, como se mencionó, se tiende a utilizar este procedimiento. En cualquier caso en que se utilice un modelo, es necesario validarlo: esto es, verificar si los supuestos del modelo coinciden con la realidad. Y esto no es lo que hace la mayoría de los usuarios. La regresión lineal implica por lo menos, distribución normal de los errores de la variable dependiente, que no están correlacionados y para utilizarlo con validez estadística, además debe contarse con un tamaño de muestra n de por lo menos 30 datos históricos. ¡¡¡Cuántos cursos de finanzas y de presupuestos en particular no se hacen invitando a los estudiantes a utilizar la regresión lineal con 3 ó 5 datos!!! Otro supuesto obvio es que la tendencia observada de los datos puede ser descrita por una recta. Sin embargo, este supuesto se puede obviar haciendo las substituciones necesarias, por ejemplo, si se considera que una variable tiene un comportamiento exponencial (no lineal), estos datos podrían "linealizarse" calculando el logaritmo de los datos y proyectar el logaritmo. Después se halla el antilogaritmo y esa sería la proyección.

La idea de la regresión lineal es hallar una recta que cumpla con un requisito básico común para muchos métodos de pronóstico: la suma de los cuadrados de la diferencia entre el valor estimado y el observado es mínima. Por eso se llama también método de mínimos cuadrados.

En general, se trata de encontrar (en el caso de la regresión lineal), una recta que cumpla esa condición y que se expresa así:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n + e$$

Donde

Y = variable dependiente

X_j = variable independiente

e = error

a = intercepción con el eje de las abcisas (y)

b_j = coeficiente de cada variable X_j

El caso particular de una variable independiente la "fórmula" será:

$$Y = a + bX + e$$

Excel presenta varias alternativas para calcular proyecciones de variables que tienen un comportamiento lineal.

Métodos de Descomposición

Un método de pronóstico es el de descomposición, para analizar series de tiempo. Un paso importante en el proceso de determinar el método de series de tiempo adecuado es considerar los diferentes patrones que se encuentran en los datos. Se pueden identificar cuatro patrones típicos: horizontal o estacionaria, estacional, cíclico y de tendencia.

1. Se presenta un patrón horizontal o estacionario (H) cuando los datos fluctúan alrededor de un valor promedio constante. Las ventas que no aumentan ni disminuyen con el tiempo, es un ejemplo de este tipo de comportamiento.
2. Se presenta un patrón estacional (E) cuando los datos están afectados por factores que se repiten con cierta frecuencia (trimestral, mensual o en determinadas fechas, por ejemplo, Navidad, Semana Santa, etc.).
3. Un patrón cíclico (C) se presenta debido a efectos económicos de largo plazo y generalmente asociados con el ciclo económico. La construcción de vivienda puede ser un ejemplo de este tipo.

4. Existe un patrón de tendencia (T) cuando existe un aumento o disminución secular de los datos. Las ventas de la mayoría de las firmas presentan este comportamiento.

Los métodos de descomposición suponen que los datos contienen patrones estacionales, cíclicos y de tendencia; una función que representa esta relación puede ser la siguiente:

dato = patrón + error.

= f(tendencia, estacionalidad, ciclo) + error.

$X_t = f(T_t, E_t, C_t, Er_t)$

donde

X_t es el dato al período t.

T_t es el componente de tendencia en el período t.

E_t es el componente o índice de estacionalidad del período t.

C_t es el componente cíclico del período t.

y Er_t es el error del período t.

El procedimiento general para aislar los diversos componentes es el siguiente y se aplica a los diferentes métodos de descomposición.

1) Con los datos disponibles calcule el promedio con un N igual a la longitud de la estacionalidad (12 meses, 6 meses, 4 trimestres, o 7 días,

por ejemplo). Con esto se elimina la estacionalidad y el error, por lo tanto en el promedio móvil se encuentra sólo la tendencia y el ciclo.

2) Separe el resultado de 1) -el promedio móvil- de los datos. Lo que queda es la estacionalidad y el error.

3) Aísle los factores estacionales promediándolos para cada período que constituyen el período completo de estacionalidad (cada mes, semestre o trimestre, por ejemplo).

4) Identifique la forma de la tendencia con los resultados de 1) (lineal, exponencial, etc.) y calcule su valor para cada uno de los períodos para los cuales se tienen datos.

5) Separe el resultado de 4) de los resultados de 1) para obtener el factor cíclico.

6) Separe la estacionalidad, la tendencia y el ciclo de los datos para obtener el error.

Este método es útil cuando se considera que existe una tendencia y estacionalidad. La estacionalidad se puede identificar en los datos si se observan ciertos "picos" o "baches" en los datos con regularidad; por ejemplo, si encuentra que el consumo de gaseosa es siempre mayor en los días sábados y domingos y menor en los días jueves, se podría sospechar que existe una estacionalidad asociada a esos días de la semana. Por otro lado, se puede llegar a la conclusión acerca de la existencia de la estacionalidad deduciéndola a partir del comportamiento del negocio; por ejemplo, antes de examinar cualquier dato, se podría pensar que la venta de juguetes o de calendarios y agendas van a presentar picos en los tres últimos meses del año. Obsérvese que se habla de estacionalidad cuando los períodos de análisis son menores de un año. Por ejemplo, semestres, trimestres o meses en relación con un año; quincenas, décadas o semanas en relación con mes; días de la semana con relación a la misma. Esto es, si los datos son anuales, por ejemplo, no tiene sentido pensar en la existencia de un patrón estacional.

Uno de los modelos de descomposición más utilizados es el multiplicativo, o sea,

$X_t = T_t \text{ x } E_t \text{ x } C_t \text{ x } Er_t$

Al aplicar los seis pasos propuestos se tiene:

1) y 2) Calcule el promedio móvil y aísle los factores estacionales:

$M_t = T_t \text{ x } C_t$

$$\frac{X_t}{M_t} = \frac{T_t \times E_t \times C_t \times Er_t}{T_t \times C_t}$$

La expresión anterior aísla la estacionalidad y el error.

3) El siguiente paso es eliminar el error de los valores obtenidos con la última expresión. Los modelos clásicos de descomposición utilizan el enfoque del promedio medial. Para calcular el promedio medial se toman todos los datos de promedio móvil para cada período (mes, trimestre, etc.) y se eliminan los valores extremos, con los datos restantes se calcula el

promedio. Los datos obtenidos para cada período se ajustan al 100% multiplicando el promedio medial por 100 x número de períodos/suma de todos los promedios mediales

4) y 5) Los pasos finales es el de calcular la tendencia y separarla del ciclo. Se identifica el patrón de la tendencia y se calcula el valor de ella para cada uno de los períodos para los cuales se tienen datos. En este modelo se elimina así:

$$\frac{M_t}{T_t} = \frac{T_t \times C_t}{f(a,b,c...t)} = C_t$$

donde a,b,c... son las constantes de la regresión y t es el período correspondiente.

En el caso de una regresión lineal se tendría:

$$\frac{M_t}{T_t} = \frac{T_t C_t}{a + b(t)} = C_t$$

6) Con estos factores, estacionalidad, tendencia y ciclo, se puede estimar el error.

$$Er_t = \frac{X_t}{T_t \times E_t \times C_t}$$

Ejemplo:

Se va a estudiar el comportamiento del Indice de Ventas a Precios Corrientes del Comercio Detallista Alimentos y Bebidas 1975-1983 y se va a hacer una proyección utilizando el método de descomposición. (Los datos están tabulados más adelante). Se va a fraccionar la información de 1975 a 1982 para hacer una proyección de los siguientes doce meses (de 1983), para evaluar la bondad del método.

Indice de ventas a precios corrientes del comercio detallista[1]
Alimentos y bebidas 1975-1982

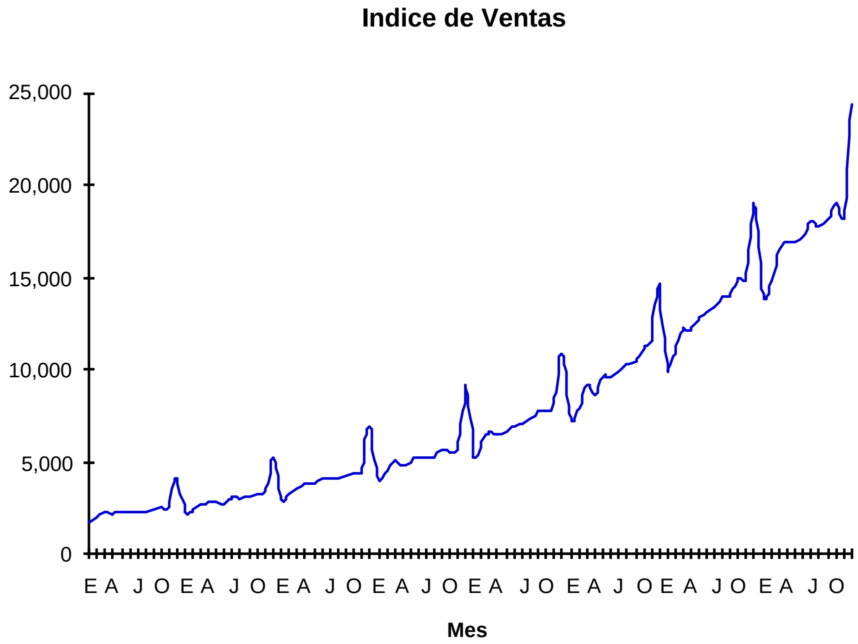
Dato real	Mes	Indice (1)	Dato real	Mes	Indice (1)	Dato real	Mes	Indice (1)	Dato real	Mes	Indice (1)
1	Ene 75	1,738	25	Ene 77	2,939	49	Ene 79	5,236	73	Ene 81	10,058
2	Feb	1,983	26	Feb	3,320	50	Feb	6,037	74	Feb	10,997
3	Mar	2,240	27	Mar	3,623	51	Mar	6,608	75	Mar	12,133
4	Abr	2,138	28	Abr	3,765	52	Abr	6,426	76	Abr	12,182
5	May	2,330	29	May	3,812	53	May	6,614	77	May	12,916
6	Jun	2,333	30	Jun	4,157	54	Jun	6,986	78	Jun	13,032
7	Jul	2,320	31	Jul	4,100	55	Jul	7,017	79	Jul	13,379
8	Ago	2,364	32	Ago	4,135	56	Ago	7,322	80	Ago	13,955
9	Sep	2,463	33	Sep	4,220	57	Sep	7,658	81	Sep	14,185
10	Oct	2,579	34	Oct	4,313	58	Oct	7,661	82	Oct	14,984
11	Nov	2,545	35	Nov	4,440	59	Nov	8,070	83	Nov	14,825
12	Dic	4,091	36	Dic	6,909	60	Dic	10,877	84	Dic	18,964
13	Ene 76	2,224	37	Ene 78	4,009	61	Ene 80	7,242	85	Ene 82	13,959
14	Feb	2,459	38	Feb	4,577	62	Feb	7,955	86	Feb	14,873
15	Mar	2,657	39	Mar	5,039	63	Mar	9,149	87	Mar	16,568
16	Abr	2,858	40	Abr	4,717	64	Abr	8,670	88	Abr	16,972
17	May	2,832	41	May	4,969	65	May	9,561	89	May	17,023
18	Jun	2,782	42	Jun	5,283	66	Jun	9,636	90	Jun	17,237
19	Jul	3,025	43	Jul	5,209	67	Jul	9,790	91	Jul	18,034
20	Ago	3,002	44	Ago	5,250	68	Ago	10,182	92	Ago	17,789
21	Sep	3,151	45	Sep	5,597	69	Sep	10,441	93	Sep	18,168
22	Oct	3,283	46	Oct	5,449	70	Oct	11,184	94	Oct	19,031
23	Nov	3,355	47	Nov	5,704	71	Nov	11,562	95	Nov	18,234
24	Dic	5,296	48	Dic	9,154	72	Dic	14,628	96	Dic	24,379

[1] Fuente DANE

Índice de ventas a precios corrientes del comercio detallista
Alimentos y bebidas 1983 (para comparación)

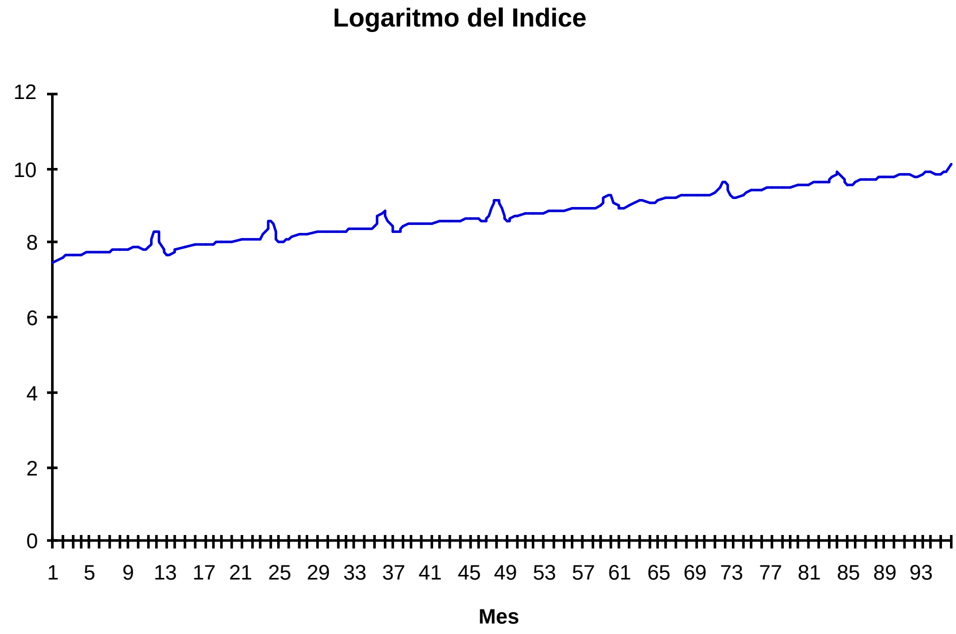
Dato real	Mes	Indice (1)
97	Ene 83	17,176
98	Feb	18,327
99	Mar	21,819
100	Abr	19,627
101	May	20,969
102	Jun	22,046
103	Jul	22,068
104	Ago	21,949
105	Sep	22,630
106	Oct	23,609
107	Nov	22,573
108	Dic	29,583

Lo primero que se requiere es observar el patrón de los datos. Para esto se grafican los datos en el tiempo, así:



Lo primero que se observa en esta gráfica es que hay una tendencia (crecimiento) y una estacionalidad (picos y valles). Otra información que se deduce de la gráfica es que la tendencia no parece lineal; por lo tanto, se puede explorar la posibilidad de linealizar los datos a través de una transformación logarítmica.

Esto es, se va a trabajar no con los datos, sino con su logaritmo natural. Al graficar los datos así transformados, se obtiene lo siguiente:



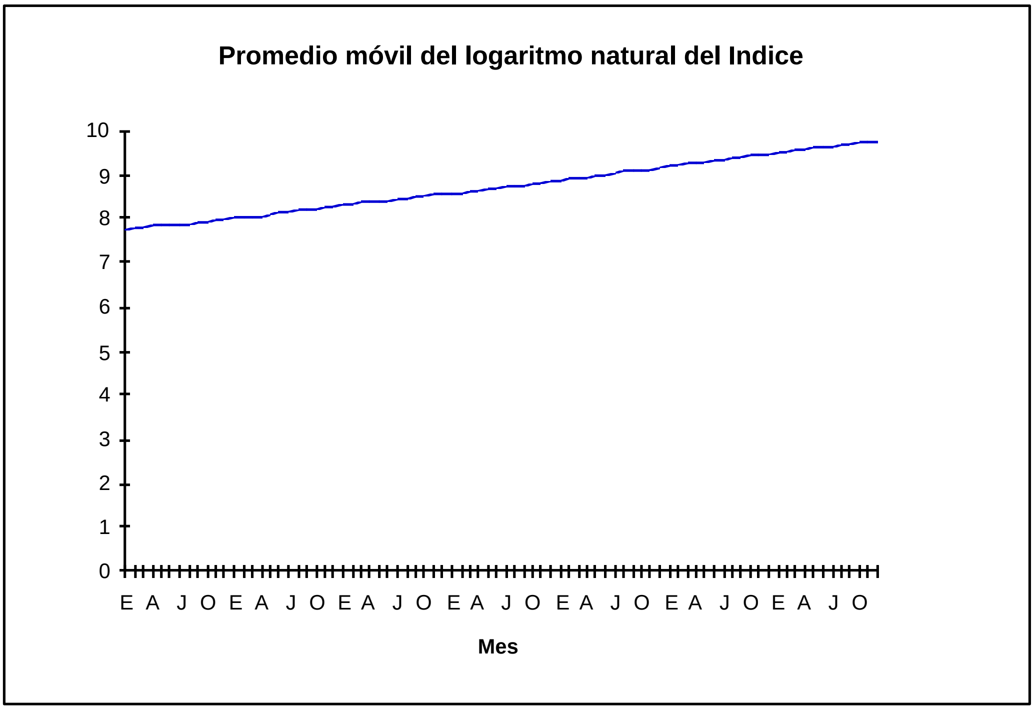
La transformación "linealizó" los datos, lo cual puede facilitar su tratamiento numérico. Sin embargo, la estacionalidad, ni el error se han perdido. Este resultado se hubiera podido visualizar utilizando la opción de gráfica semilogarítmica que ofrece Excel; como se trata de hacer cálculos, se hizo explícito el tratamiento logarítmico de los datos y después de eso se graficó el logaritmo natural (ln) de los datos originales. Se calcula el promedio móvil de 12 meses -para incluir todo el ciclo de estacionalidad- y se obtienen los siguientes resultados:

Índice de ventas a precios corrientes del comercio detallista
Alimentos y bebidas 1975-1983

Mes	ln del Indice (2)	Promedio móvil CxT (3)	ln del Indice (2)	Promedio móvil CxT (3)	ln del Indice (2)	Promedio móvil CxT (3)	ln del Indice (2)	Promedio móvil CxT (3)
	75-76		77-78		79-80		81-82	
Ene	7.460	#N/A	7.986	8.009	8.563	8.577	9.216	9.195
Feb	7.592	#N/A	8.108	8.032	8.706	8.599	9.305	9.222
Mar	7.714	#N/A	8.195	8.057	8.796	8.622	9.404	9.249
Abr	7.668	#N/A	8.234	8.083	8.768	8.645	9.408	9.273
May	7.754	#N/A	8.246	8.106	8.797	8.670	9.466	9.301
Jun	7.755	#N/A	8.333	8.130	8.852	8.694	9.475	9.326
Jul	7.749	#N/A	8.319	8.164	8.856	8.717	9.501	9.351
Ago	7.768	#N/A	8.327	8.189	8.899	8.742	9.544	9.377
Sep	7.809	#N/A	8.348	8.216	8.944	8.770	9.560	9.403
Oct	7.855	#N/A	8.369	8.240	8.944	8.796	9.615	9.429
Nov	7.842	#N/A	8.398	8.263	8.996	8.825	9.604	9.453
Dic	8.317	#N/A	8.841	8.286	9.294	8.853	9.850	9.474
Ene	7.707	7.774	8.296	8.309	8.888	8.868	9.544	9.496
Feb	7.808	7.794	8.429	8.334	8.982	8.895	9.607	9.523
Mar	7.885	7.812	8.525	8.361	9.121	8.918	9.715	9.548
Abr	7.958	7.826	8.459	8.389	9.068	8.945	9.739	9.574
May	7.949	7.851	8.511	8.407	9.165	8.970	9.742	9.602
Jun	7.931	7.867	8.572	8.430	9.173	9.001	9.755	9.625
Jul	8.015	7.881	8.558	8.450	9.189	9.027	9.800	9.648
Ago	8.007	7.904	8.566	8.469	9.228	9.055	9.786	9.673
Sep	8.055	7.923	8.630	8.489	9.253	9.083	9.807	9.693
Oct	8.097	7.944	8.603	8.513	9.322	9.109	9.854	9.714
Nov	8.118	7.964	8.649	8.532	9.355	9.140	9.811	9.734
Dic	8.575	7.987	9.122	8.553	9.591	9.170	10.101	9.751

El promedio móvil contiene, entonces, la tendencia (T) y el ciclo (C). Al dividir el dato original por el promedio móvil, el resultado entonces contendrá la estacionalidad (Est) y el error (Err). Como el ciclo es un ”movimiento” de largo plazo y de alguna manera la tendencia T, calculada como una regresión lineal, es un promedio, se puede suponer que la tendencia (T) calculada, elimina el ciclo C.

Al examinar el comportamiento del promedio móvil, se observa una casi perfecta linealidad de los datos.



Por lo tanto, se le puede proyectar por medio de una regresión lineal. Excel tiene varios modos de trabajar la proyección lineal. Aquí se ha escogido la función **=PRONOSTICO(valor de x;matriz y;matriz x)** que es muy sencilla.

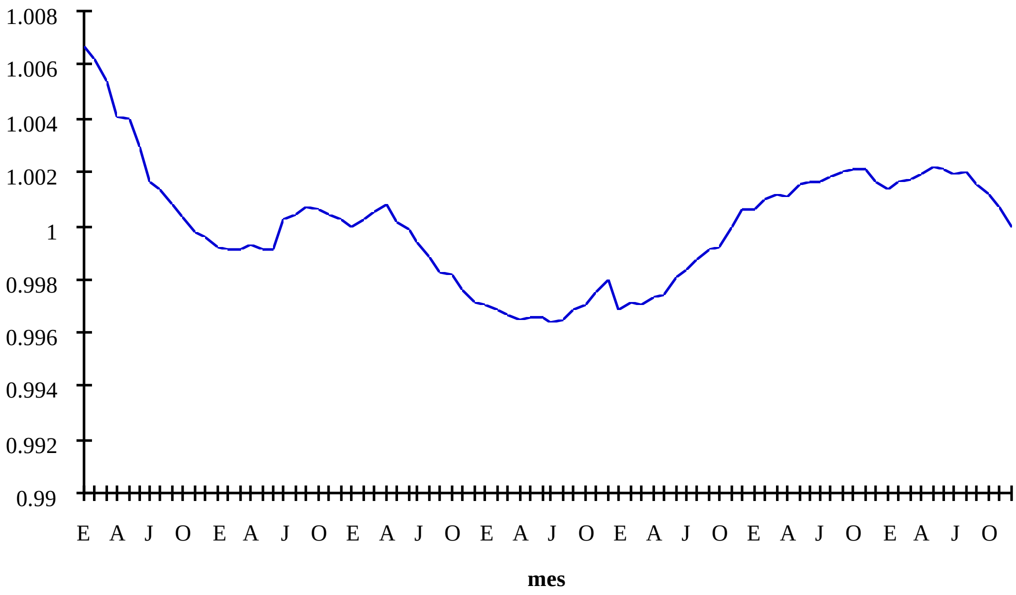
Indice de ventas a precios corrientes del comercio detallista alimentos y bebidas 1975-1983

EstxErr (4) =(2)/(3)	Tendencia T calculada (5)	EstxErr (4) =(2)/(3)	Tendencia T calculada (5)	EstxErr (4) =(2)/(3)	Tendencia T calculada (5)	EstxErr (4) =(2)/(3)	Tendencia T calculada (5)
76	76	77-78	77-78	79-80	79-80	81-82	81-82
		0.997	8.015	0.998	8.602	1.002	9.189
		1.009	8.039	1.012	8.626	1.009	9.213
		1.017	8.064	1.020	8.651	1.017	9.238
		1.019	8.088	1.014	8.675	1.015	9.262
		1.017	8.113	1.015	8.700	1.018	9.287
		1.025	8.137	1.018	8.724	1.016	9.311
		1.019	8.162	1.016	8.749	1.016	9.336
		1.017	8.186	1.018	8.773	1.018	9.360
		1.016	8.211	1.020	8.798	1.017	9.385
		1.016	8.235	1.017	8.822	1.020	9.409
		1.016	8.260	1.019	8.847	1.016	9.434
		1.067	8.284	1.050	8.871	1.040	9.458
0.991	7.721	0.999	8.308	1.002	8.895	1.005	9.483
1.002	7.746	1.011	8.333	1.010	8.920	1.009	9.507
1.009	7.770	1.020	8.357	1.023	8.944	1.017	9.531
1.017	7.795	1.008	8.382	1.014	8.969	1.017	9.556
1.013	7.819	1.012	8.406	1.022	8.993	1.015	9.580
1.008	7.844	1.017	8.431	1.019	9.018	1.014	9.605
1.017	7.868	1.013	8.455	1.018	9.042	1.016	9.629
1.013	7.893	1.011	8.480	1.019	9.067	1.012	9.654
1.017	7.917	1.017	8.504	1.019	9.091	1.012	9.678
1.019	7.942	1.011	8.529	1.023	9.116	1.014	9.703
1.019	7.966	1.014	8.553	1.024	9.140	1.008	9.727
1.074	7.990	1.066	8.577	1.046	9.165	1.036	9.752

Índice de ventas a precios corrientes del comercio detallista
Alimentos y bebidas 1975-1983

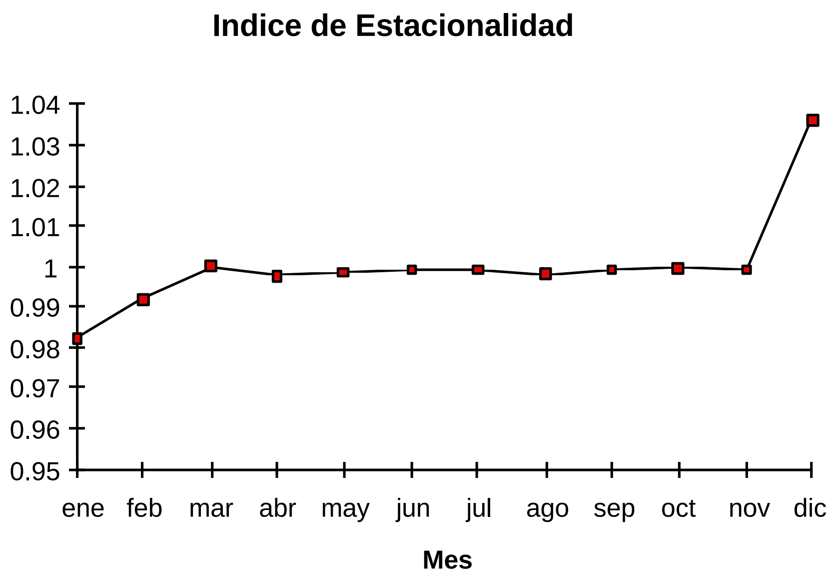
Mes	Ciclo C (6) =(3)/(5)	Ciclo C (6) =(3)/(5)	Ciclo C (6) =(3)/(5)	Ciclo C (6) =(3)/(5)	Ciclo C (6) =(3)/(5)	Ciclo C (6) =(3)/(5)	Ciclo C (6) =(3)/(5)
	1976	1977	1978	1979	1980	1981	1982
Ene	1.007	0.999	1.000	0.997	0.997	1.001	1.001
Feb	1.006	0.999	1.000	0.997	0.997	1.001	1.002
Mar	1.005	0.999	1.000	0.997	0.997	1.001	1.002
Abr	1.004	0.999	1.001	0.996	0.997	1.001	1.002
May	1.004	0.999	1.000	0.997	0.997	1.002	1.002
Jun	1.003	0.999	1.000	0.997	0.998	1.002	1.002
Jul	1.002	1.000	0.999	0.996	0.998	1.002	1.002
Ago	1.001	1.000	0.999	0.996	0.999	1.002	1.002
Sep	1.001	1.001	0.998	0.997	0.999	1.002	1.002
Oct	1.000	1.001	0.998	0.997	0.999	1.002	1.001
Nov	1.000	1.000	0.998	0.998	1.000	1.002	1.001
Dic	1.000	1.000	0.997	0.998	1.001	1.002	1.000

Ciclo de largo plazo



Indice de estacionalidad

								promedio	ajuste a 12 en total
	76	77	78	79	80	81	82		
E	0.9914	0.9972	0.9985	0.9984	1.0022	1.0023	1.0051	0.9993	0.9822
F	1.0017	1.0094	1.0113	1.0124	1.0097	1.0090	1.0089	1.0089	0.9916
M	1.0093	1.0171	1.0196	1.0202	1.0228	1.0167	1.0175	1.0176	1.0002
A	1.0168	1.0187	1.0084	1.0143	1.0137	1.0146	1.0173	1.0148	0.9974
M	1.0125	1.0173	1.0123	1.0146	1.0218	1.0178	1.0146	1.0158	0.9984
J	1.0082	1.0249	1.0169	1.0181	1.0192	1.0160	1.0135	1.0167	0.9992
J	1.0169	1.0190	1.0129	1.0159	1.0179	1.0161	1.0157	1.0163	0.9989
A	1.0131	1.0168	1.0114	1.0179	1.0191	1.0177	1.0117	1.0154	0.9980
S	1.0167	1.0160	1.0166	1.0198	1.0188	1.0166	1.0118	1.0166	0.9992
O	1.0192	1.0157	1.0106	1.0168	1.0235	1.0197	1.0144	1.0171	0.9997
N	1.0194	1.0164	1.0137	1.0194	1.0236	1.0159	1.0079	1.0166	0.9992
D	1.0736	1.0669	1.0665	1.0498	1.0459	1.0397	1.0359	1.0540	1.0360
							SUMA	12.2093	12.0000



Para preparar un pronóstico se multiplica el valor de la tendencia calculada por el índice de estacionalidad y por el factor cíclico que se estime. Para estimar el factor cíclico se debe tener un cierto conocimiento del devenir de la economía y no es calculable en forma directa como puede ser la tendencia o la estacionalidad; esta estimación del ciclo se basa en la información disponible sobre la economía, la observación del ciclo y en algún grado es un estimativo de tipo subjetivo. Una posibilidad es examinar la tendencia que muestre la gráfica, como aparece a continuación. En el caso del ejemplo, si se desea pronosticar el año 1983 se tiene:

1983	Dato No	Tendencia	Ciclo	Estacionalidad	Proyección del logaritmo TxCxEst (ln)	Proyección del dato o índice (antilogaritmo)
E	97	9.7760	1.0000	0.9822	9.6021	14,796.11
F	98	9.8005	1.0001	0.9916	9.7197	16,641.62
M	99	9.8249	1.0002	1.0002	9.8286	18,557.58
A	100	9.8494	1.0003	0.9974	9.8268	18,524.29
M	101	9.8739	1.0004	0.9984	9.8622	19,191.01
J	102	9.8983	1.0005	0.9992	9.8956	19,842.59
J	103	9.9228	1.0006	0.9989	9.9176	20,284.25
A	104	9.9472	1.0007	0.9980	9.9338	20,615.41
S	105	9.9717	1.0007	0.9992	9.9710	21,395.81
O	106	9.9962	1.0008	0.9997	10.0013	22,055.93
N	107	10.0206	1.0009	0.9992	10.0217	22,508.78
D	108	10.0451	1.0010	1.0360	10.4169	33,420.56

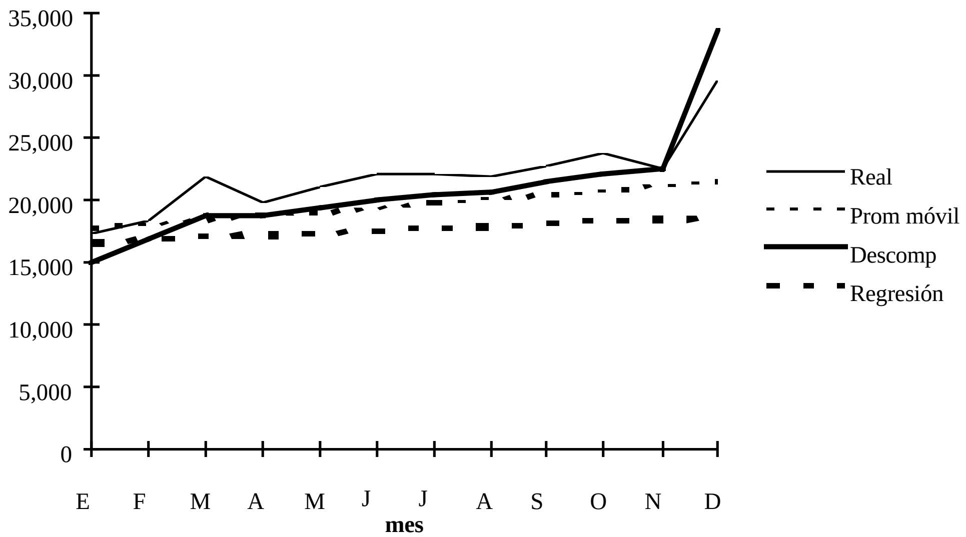
(El último dato que se utilizó para calcular la tendencia fue el número 96 (diciembre de 1982); por lo tanto, el mes de enero de 1983 corresponde al 97.

Los promedios móviles que se utilizan en este método pueden ser centrados o no. Si el número de términos en el promedio móvil es impar la colocación del promedio móvil no tiene problema pues se sitúa en (N+1)/2.

Si es par, se tendría que colocar medio período rezagado o medio período adelantado.

Se puede comparar gráficamente el resultado de algunas proyecciones, incluida ésta que se acaba de calcular.

Indice real vs varias proyecciones



Una forma analítica de evaluar qué tan buena es una proyección es calcular la suma de los cuadrados de los errores. Entre varios métodos, se considerará mejor aquel que presente menor suma de los cuadrados de los errores. Si se consideran los métodos de regresión lineal para proyectar los siguientes doce meses, o el promedio móvil, para proyectar los mismos doce meses, pero uno a la vez, o sea, que al finalizar enero se proyecta febrero, y así sucesivamente, se obtiene lo siguiente:

		Real	Prom móvil	Reg lineal	Descomp	Prom móvil (Cuadrado de los errores)	Reg lineal (Cuadrado de los errores)	Descomp (Cuadrado de los errores)
97	Ene	17,176	17,689	16,628	14,796	263,084	300,807	5,663,877
98	Feb	18,327	17,957	16,807	16,642	136,900	2,310,689	2,840,509
99	Mar	21,819	18,245	16,986	18,558	12,774,667	23,355,292	10,636,882
100	Abr	19,627	18,682	17,166	18,524	892,238	6,058,331	1,215,977
101	May	20,969	18,904	17,345	19,191	4,265,602	13,133,406	3,161,261
102	Jun	22,046	19,233	17,524	19,843	7,915,782	20,445,233	4,855,031
103	Jul	22,068	19,633	17,704	20,284	5,928,008	19,046,913	3,181,753
104	Ago	21,949	19,969	17,883	20,615	3,918,750	16,531,651	1,778,459
105	Sep	22,630	20,316	18,062	21,396	5,354,210	20,862,510	1,523,223
106	Oct	23,609	20,688	18,242	22,056	8,532,728	28,806,686	2,412,033
107	Nov	22,573	21,069	18,421	22,509	2,260,763	17,237,629	4,124
108	Dic	29,583	21,431	18,601	33,421	66,455,104	120,614,402	14,726,872
					Suma de cuadrados	118,697,835	288,703,547	52,000,000

Con estos datos se concluye que el mejor método de pronóstico en este caso particular es el de descomposición por tener menor suma de cuadrado de los errores. Debe observarse que este es un análisis *a posteriori* suponiendo que se está en diciembre de 1982 y se hacen las proyecciones (con excepción del promedio móvil que debe hacerse mes a mes); después se espera a diciembre de 1983 “para ver” qué tal resultaron las proyecciones. De modo que el análisis de los mínimos cuadrados sólo podría hacerse en diciembre de 1983.

El uso de los métodos de pronósticos depara sorpresas y deben explorarse muy bien las cifras. Lo que se ha hecho en este ejemplo es relativamente sencillo, porque se conocen unas cifras (1983) contra las cuales comparar la bondad del pronóstico; sin embargo, el problema reside en que en la realidad obviamente no se sabe qué tan bueno va a resultar el pronóstico hacia el futuro. Se sugiere al lector que trabaje los datos de los últimos cuatro años y deberá encontrar que el pronóstico que se obtiene es mejor que con todos los datos. Esto puede ocurrir porque al considerar la totalidad de los datos el modelo tiene en cuenta tendencias o hechos que ya no afectan a la situación actual; en otras palabras, replica un patrón que ya no es válido.

Referencias

Drake, Alwin W. *Fundamentals of Applied Probability Theory*, McGraw-Hill Book Co., 1967.

Makridakis, S., S.C. Wheelwright, *Forecasting. Methods and Applications*, John Wiley, 1978. Existe tercera edición, 1998).

Wonnacott, Thomas H., Ronald J. Wonnacot, *Introductory Statistics for Business and Economics*, 2ª ed., John Wiley, 1977.

Zuwaylif, Fadil H., *Estadística general aplicada*, Addison-Wesley Iberoamericana, 1987.

EJERCICIOS

1. Suponga que una empresa de encuestas encuentra la siguiente información acerca de si se favorece o no la legalización del consumo de droga. La muestra se ha dividido según la siguiente tabla:

	A favor	En contra
Menor de 30 años.	26%	21%
Mayor de 30 años.	19%	34%

Si se escoge un individuo en forma aleatoria, ¿cuál es la probabilidad que sea

¿A favor de la legalización?

¿A favor de la legalización, si es menor de 30 años?

¿A favor de la legalización, si es mayor de 30 años?

2. Una universidad matricula 10.000 estudiantes, de los cuales 4.950 son mujeres y 5.050 son hombres. Si la Facultad de Ingeniería tiene 20% de las mujeres y 35% de los hombres, si se escoge a un estudiante de manera aleatoria, ¿cuál es la probabilidad que sea mujer? ¿que sea hombre?

3. Simule el lanzamiento de un dado de seis caras utilizando la fórmula **=ALEATORIO** de Excel. Esta fórmula arroja números menores que 1, cuya característica es que los dígitos que lo conforman provienen de un universo en el cual los dígitos de 0 a 9 tienen igual probabilidad de aparecer. Para simular el lanzamiento de un dado de seis caras con valores de 1 a 6, descarte los dígitos 7, 8, 9 y 0.

Construya la gráfica en cada caso, calcule la media, la varianza y la desviación estándar de la muestra y del universo, para los siguientes casos:

- Para 25 casos.
- Para 100 casos.
- Para 500 casos.
- Para diez millones de casos (haga un estimativo basado en la información de los casos anteriores).

5. Para una perinola de siete caras, donde a cada cara se le asigna un número de 1 a 7, calcule en forma analítica, esto es, sin hacer lanzamientos, la media y la varianza del valor que aparece en la cara sobre la cual reposa la perinola al haber sido lanzada. ¿Cuál es la probabilidad de obtener el valor 5 en la base? ¿Cuál es la probabilidad de obtener un valor menor que 6?