Capítulo 8

Plan de muestreo

Objetivos

Al finalizar este capítulo el estudiante estará en condiciones de:

- **Tener** claro los conceptos de elemento, población, marco muestral y unidad de muestreo.
- **Diferenciar** una muestra de un censo.
- **Clasificar** las técnicas de muestreo, como técnicas probabilísticas y no probabilísticas.
- **Identificar y diferenciar** las técnicas de muestreo no probabilísticas de las probabilísticas.
- **Deducir** las fórmulas para determinar el tamaño de la muestra.
- **Estimar** el tamaño de la muestra por el procedimiento de muestreo aleatorio simple y el estratificado.

7. PLAN DE MUESTREO

En el capítulo anterior se explicó como diseñar el medio de recolección de la información, considerando, la encuesta, como el medio de mayor utilización en la investigación de mercados. Este capítulo se enfoca en el tema del plan de muestreo, en las técnicas de muestreo de mayor utilización, así como el cálculo del tamaño de la misma. Todo estudio de investigación de mercados requiere la selección de algún tipo de muestra. En la formulación y evaluación de proyectos, resulta indispensable, dado que en la mayor parte de ellos, se realiza para la producción y comercialización de un producto y/o servicio, es decir, que se coloca a prueba el producto, por tanto se requiere monitorear las ventas del área de influencia del mercado objetivo.

La alternativa al muestreo es hacer un censo. En un censo, se utiliza todos los elementos disponibles de una población definida, por ejemplo: si se requiere investigar la aceptación de un producto en el mercado, se tendría que llegar a todas las familias de la población objetivo.

El muestreo ofrece algunos beneficios importantes en comparación con la realización de un censo.

- Una muestra ahorra dinero.
- Una muestra ahorra tiempo.
- Una muestra puede ser más exacta.
- Una muestra es mejor, si el estudio conlleva la destrucción o contaminación del elemento muestreado.

Ahora recordemos algunos conceptos básicos de estadística que se requieren para adelantar el tema.

8.1. ELEMENTO

Es la unidad acerca de la cual se solicita información. Éste, suministra la base del análisis que se llevará a cabo. Los elementos más comunes del muestro en la investigación de mercados, son los individuos. En otros casos, los elementos podrían ser productos, almacenes, empresas, familias, etc. Los elementos de cualquier muestra específica dependerán de los objetivos del estudio.

8.2. POBLACIÓN

Una población o universo, como también se llama, es el conjunto de todos los elementos definidos antes de selección de la muestra.

8.3. UNIDAD DE MUESTREO

Es el elemento o los elementos disponibles para su selección en alguna etapa del proceso de muestreo.

8.3. MARCO MUESTRAL

Es una lista de todas las unidades de muestreo disponibles para su selección, en una etapa del proceso de muestreo. Uno de los pensamientos más creativos en un proyecto de investigación de mercados, puede relacionarse con la especificación de un marco muestral. Un marco puede ser una lista de alumnos, una lista de votantes inscritos, un directorio telefónico, una lista de empleados o incluso un mapa.

8.5. PROCESO DE MUESTREO

En la figura 8.1 se describe en forma general los pasos para seleccionar una muestra.

Figura 8.1. Proceso de muestreo.

PASO 1: Definir la población.

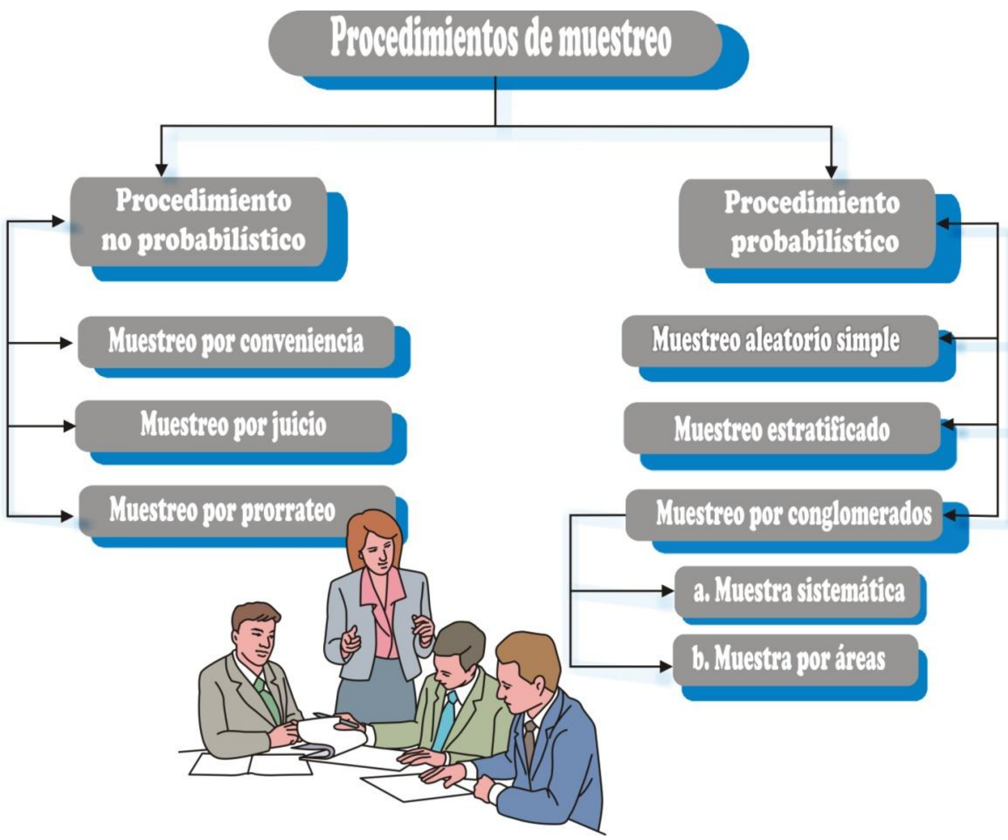
PASO 2: Identificar el marco muestral del cual se selecciona la muestra.

PASO 3: Decidir sobre el tamaño de la muestra.
PASO 4: Seleccionar un procedimiento específico mediante el cual se determina la muestra.
PASO 5: Seleccione físicamente la muestra con base en el procedimiento descrito en el paso 4.

8.6. PROCEDIMIENTOS DE MUESTREO

Existen muchos procedimientos mediante los cuales se puede seleccionar la muestra, pero fundamentalmente existen dos tipos de muestreo. El muestreo probabilístico y el muestreo no probabilístico.

Figura 8.2. Procedimiento de muestreo.



8.6.1. Procedimiento probabilístico: Cada elemento de la población tiene una posibilidad conocida de ser seleccionado para la muestra. Una muestra probabilística permite calcular el grado probable hasta el cual, el valor de la muestra, puede diferir del valor de la población de interés. Esta diferencia recibe el nombre de ERROR MUESTRAL.

8.6.2. Procedimiento no probabilístico. La selección de los elementos que hacen parte de la muestra se basa en el criterio y experiencia del investigador. No existe una posibilidad conocida de que se seleccione cualquier elemento particular de la población. Por tanto, no se puede calcular el error muestral.

Para adelantar este tema es preciso establecer la nomenclatura a utilizar y los conceptos de parámetros y estadísticos.

1.6.3. Parámetro. Un parámetro, es una descripción de una medida de la población bajo estudio. Ejemplos: Edad promedio de los estudiantes, Ingreso promedio.

1.6.4. Estadístico. Un estadístico, es una descripción resumida de una medida en la muestra seleccionada. Así la edad promedio de los estudiantes, será un estadístico, si se mide a través de una muestra.

8.7. NOMENCLATURA.

La nomenclatura que utilizaremos en este texto, se acoge a la mayoría de los textos de estadística y de econometría.

Tabla 8.1. Nomenclatura utilizada.

Parámetros - Estadísticos	Símbolos de población	Símbolos de la muestra
Media o promedio.	μ	X
Varianza.	σ^2	S^2
Proporción que responden “si”.	π	p
Proporción que responden “no”.	$(1-\pi)$	(1-p) o q

8.8. TEOREMA.

Dada una población, si extraemos todas las muestras posibles de un mismo tamaño, entonces las medias de la distribución de todas las medias muestrales posibles, será igual a la media poblacional.

$$\mu_x = \frac{\sum \overline{X}_j}{M}$$

$$\mu_x = \frac{\overline{X}_1 + \overline{X}_2 + \overline{X}_3 + ... + \overline{X}_M}{M} = \mu \qquad (1)$$

La varianza de todas las medias muestrales se simboliza por: $\sigma^2{}_{\bar{x}}$
El error estándar será simbolizada por: $\sigma_{\bar{x}}$

$$\sigma_x = \sqrt{\frac{\sum \left(\overline{X}_j - \mu\right)^2}{M}}$$

$$\sigma_x = \sqrt{\frac{\left(\overline{X}_1 - \mu\right)^2 + \left(\overline{X}_2 - \mu\right)^2 + ... + \left(\overline{X}_M - \mu\right)^2}{M}} \qquad (2)$$

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

La formula (2) se aplica para muestras grandes o sea n > 30 y se denomina: error estándar de la media.

8.9. TEOREMA DEL LÍMITE CENTRAL

Se cumple, cuando independientemente de la población origen, la distribución de las medias aleatorias, se aproxima a una distribución normal a medida que el tamaño de la muestra crece. Se podrá decir también, que si las muestras provienen de una población que no es normal, es de importancia tener en cuenta el tamaño de la muestra. Si el tamaño de la muestra es pequeño, la distribución obtenida de sus medias muestrales, tendrá un comportamiento similar al de la población de donde se extrajeron. Por el contrario, si el tamaño muestral es grande, el comportamiento de estas medias muestrales será igual al de una distribución normal independientemente de la población de donde fueron extraídas.

En su forma más simple el teorema indica que, si n variables aleatorias independientes tienen varianzas finitas, su suma, cuando se le expresa en medida estándar, tienden a estar normalmente distribuidas cuando n tiende a infinito.

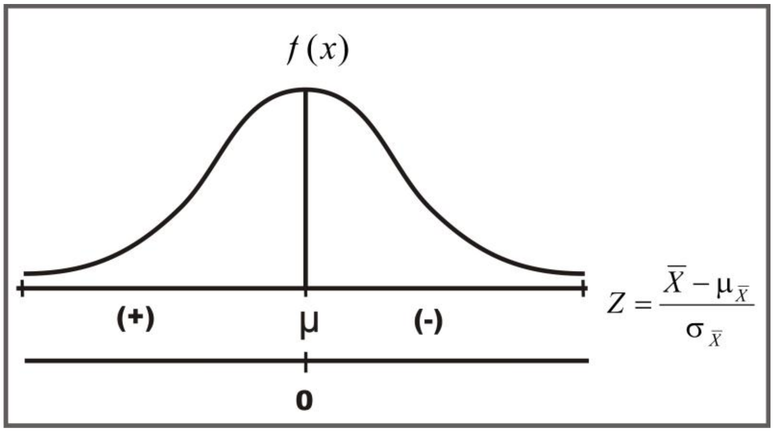
De acuerdo al teorema anterior, la varianza estadística para distribuciones de medias muestrales será:

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$$

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad (3)$$

Es decir, que se aproxima a una distribución normal de probabilidades.

Figura 8.3. Distribución normal.



8.10. DISTRIBUCIÓN NORMAL DE PROBABILIDAD

Una distribución normal de probabilidad es una función de densidad de probabilidad definida como:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \times e^{\frac{-1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \qquad (4)$$

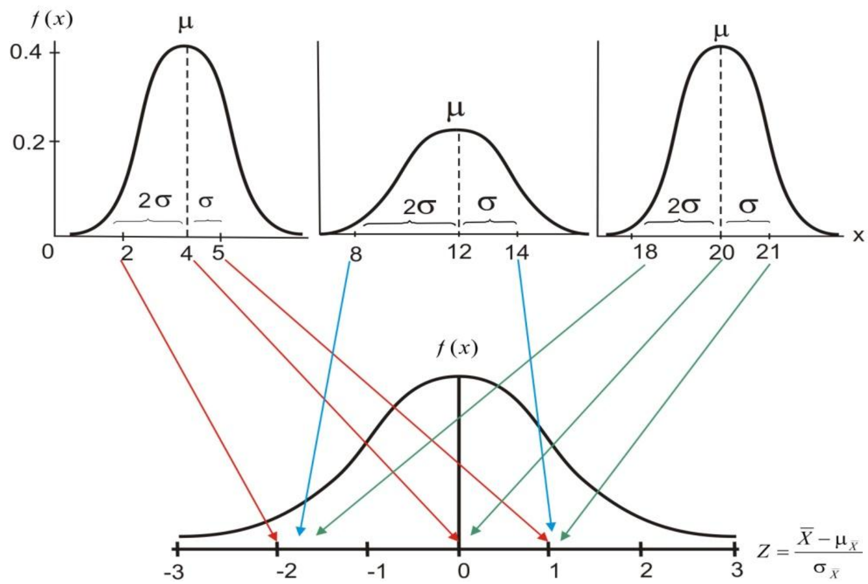
En donde $f(x)$ es una función del valor observado, x, mientras que μx, es la media y σ, la desviación estándar de la distribución de probabilidad de la variable aleatoria, X, π = 3.14159 y e = 2.71828. σ

La distribución normal de probabilidad tiene las siguientes características:

1. Un pico sobre la media, mediana y la moda de la variable aleatoria, todas iguales entre sí.
2. Perfectamente simétrica alrededor de este valor central en pico y, por lo tanto, llamada de “forma de campana”.

3. Caracterizada por las colas que se extienden indefinidamente en ambas direcciones desde el centro, y se aproximan pero nunca tocan el eje horizontal, lo que implica una probabilidad positiva de hallar valores de la variable aleatoria en cualquier punto ente menos infinito e infinito.

Figura 8.4. Construcción de la curva normal estándar[1].



Las tres curvas normales diferentes de la figura 8.4 se pueden convertir en una curva normal estándar, mostrada en el panel inferior, cada valor real, x, de la variable aleatoria normal se convierte en un valor estandarizado, z, si se calcula primeramente la desviación de x desde la media y luego se expresa en términos de unidades estándar de desviación. Este procedimiento es equivalente a hacer μ igual a cero y tratar la distancia de la media σ como igual a uno.

Por ejemplo: para el panel superior derecho tenemos los siguientes valores:

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{2-4}{1} = -2 \qquad (5)$$

Para el panel del centro se tiene:

[1] *Tomado de KIHLER. Heinz. Estadística para negocios y economía. Cecsa. México. 1996. P. 253*

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{8 - 12}{2} = -2 \qquad (6)$$

En aquellos casos de poblaciones finitas, es decir, cuando se da información sobre el tamaño poblacional y cuando el tamaño de la muestra es mayor del 5% de la población, se puede aplicar el factor de corrección, representado de diferentes maneras; cualquiera de estas formas podrá ser aplicada:

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{N-n}{N}} = \sqrt{\frac{N}{N} - \frac{n}{N}} = \sqrt{1 - \frac{n}{N}} = \sqrt{1-f} \quad donde \quad f = \frac{n}{N} \qquad (7)$$

En distribuciones de medias muestrales, la estandarización de Z, incluyendo el factor de corrección será:

$$Z = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)\sqrt{\frac{N-n}{N}}} \qquad (8)$$

8.11. TAMAÑO OPTIMO EN POBLACIONES FINITAS.

Para determinar la fórmula para el cálculo del tamaño de la muestra, se tomará el valor de $\bar{X} - \mu$ como el error máximo permitido, que denominaremos E.

Ahora haciendo algunas transformaciones algebraicas se obtiene la fórmula que se utilizará para el cálculo del tamaño de la muestra así:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}}}$$

$$Z = \frac{E}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}}}$$

$$E = \frac{Z\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

$$E^2 = \left(\frac{Z^2\sigma^2}{n}\right)\left(\frac{N-n}{N}\right)$$

$$E^2 = \left(\frac{Z^2\sigma^2}{n}\right)\left(\frac{N-n}{N}\right)$$

$$E^2 = \frac{\left(Z^2\sigma^2 N\right)-\left(Z^2\sigma^2 n\right)}{nN}$$

$$E^2 nN = \left(Z^2\sigma^2 N\right)-\left(Z^2\sigma^2 n\right)$$

$$E^2 nN + \left(Z^2\sigma^2 n\right) = \left(Z^2\sigma^2 N\right)$$

$$n\left(E^2 N + Z^{2^2}\sigma^2\right) = Z^2\sigma^2 N$$

$$n = \frac{Z^2 N\sigma^2}{NE^2 + Z^2\sigma^2} \qquad (9)$$

También:

$$.n = \frac{\sigma^2}{\left(\frac{E}{Z}\right)^2 + \frac{\sigma^2}{N}} \qquad (10)$$

Y la más utilizada $n = \dfrac{n_0}{1+\dfrac{n_0}{N}}$ *donde* $n_0 = \dfrac{Z^2\sigma^2}{E^2} = \left(\dfrac{Z\sigma}{E}\right)^2$ (11)

8.12. DISTRIBUCIÓN MUESTRAL DE UNA PROPORCIÓN.

En el análisis de una característica cualitativa o atributo, se emplea la proporción de éxito y no el número de éxitos como en la distribución binomial.

Se define la proporción de éxitos como:

$$P = \frac{Numero \quad de \quad casos \quad favorables \quad o \quad exitos}{Tatal \quad de \quad casos \quad posibles}$$

$$P = \frac{\sum a_j}{n}$$

SIMBOLOGÍA:

$A = \sum A_j = NP$ Total de elementos que presenta la característica en la población.

$P = \frac{A}{N} = \frac{\sum A_j}{N}$ Proporción de elementos que presentan la característica en la población.

$Q = \frac{N - A}{N} = 1 - P$ Proporción de elementos que no presentan la característica.

$$P + Q = 1 \qquad (12)$$

Varianza de la proporción en la población $\sigma_p^2 = PQ$

Desviación estándar $\sigma_P = \sqrt{PQ}$

$$\sigma_{\overline{P}} = \frac{\sigma_P}{\sqrt{n}} = \sqrt{\frac{PQ}{n}}$$ = Error estándar de la proporción. (13)

8.13. FÓRMULAS UTILIZADAS EN ATRIBUTOS:

$$n = \frac{Z^2 NPQ}{(N-1)E^2 + Z^2 PQ}. \qquad (14)$$

$$.n = \frac{PQ}{\left(\frac{E}{Z}\right)^2 + \frac{PQ}{N}} \qquad (15)$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad Siendo \quad n_0 = \frac{Z^2 PQ}{E^2} \qquad (16)$$

8.14. TAMAÑO DE LA MUESTRA

Para determinar n el tamaño de la muestra, es necesario identificar los siguientes componentes o elementos técnicos:

1. **La varianza** $(\sigma^2{}_X)$. En el caso de $(\sigma^2{}_X)$ = PQ, sucede algo similar, pero se tiene la costumbre de tomar P = 0,50 con lo cual se obtiene el máximo valor posible de n.

2. **El nivel de confianza.** Tiene relación directa con el tamaño de la muestra, por lo tanto, se dirá que a mayor nivel de confianza más grande debe ser el tamaño de la muestra. El nivel es fijado por el investigador, de acuerdo a su experiencia.

3. **Precisión de la estimación**: Corresponde al margen de error que el investigador fija de acuerdo al conocimiento que tenga acerca del parámetro que piensa estimar. Se le conoce como error de muestreo (E), siendo:

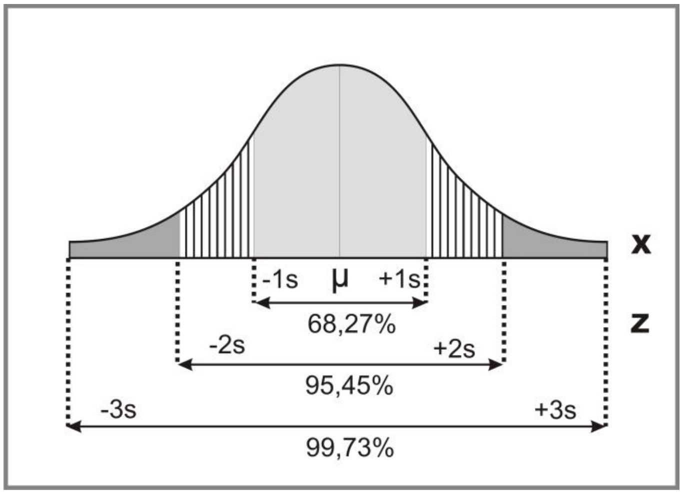
$$E = Z\frac{\sigma}{\sqrt{n}} \quad = Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}} \qquad (17)$$

4. **Estimación de intervalos de confianza para parámetros de población**. Sean μs y σs la media y la desviación (error típico) de la distribución de muestreo de un estadístico S. Entonces si la distribución de muestreo de S es aproximadamente normal (que como hemos visto es cierto para muchos estadísticos si el tamaño de la muestra n $\geq$ 30), podremos esperar hallar un estadístico real S que esté en los intervalos:

$$[\mu_s - \sigma_s \quad ; \quad \mu_s + \sigma_s] \; ; \; [\mu_s - 2\sigma_s \quad ; \quad \mu_s + 2\,\sigma_s] \; ; \; [\mu_s - 3\sigma_s \quad ; \quad \mu_s + 3\sigma_s]$$

Alrededor del 68.7%, 95.45% y 99.73% respectivamente.

Figura 8.5. Distribución normal estándar.



5. Cálculo de los coeficientes de confianza. $S \pm 1.96\sigma_S$ son los límites de confianza 95% para S. El porcentaje de confianza suele llamarse nivel de confianza. Los números 1.96, 2.58, etc., en los límites de confianza se llaman COEFICIENTES DE CONFIANZA O VALORES CRÍTICOS y se denota por Z. De los niveles de confianza podemos deducir los coeficientes de confianza y viceversa.

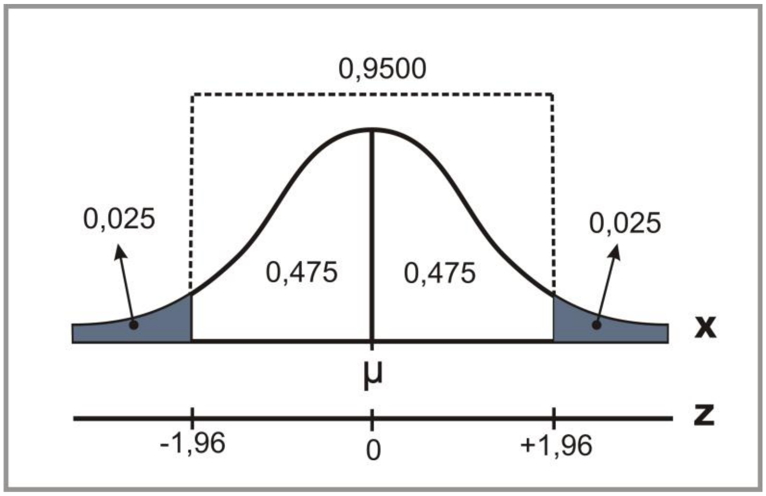
Ejemplo: Calcular el Coeficiente de confianza para un nivel de confianza del 95%, para ello, se debe seguir los siguientes pasos:

Primer paso: 1-0,95 = 0,05

Segundo paso: 0,05/2 = 0,025 (operación que se realiza por la tabla de distribución de frecuencia utilizada).

Tercer paso: 0,500-0,025 = 0,475 Se busca en la tabla de distribución normal el área 0,475 el valor de Z que le corresponde = 1.96

Figura 8.6. Distribución normal del cálculo del coeficiente de confianza.



En la tabla 8.2. Se presenta el cálculo de algunos niveles de confianza más utilizados.

Tabla 8.2. Coeficientes de confianza.

Nivel de confianza	**99.73%**	**99%**	**98%**	**96%**	**95.45%**	**95%**	**90%**	**80%**	**68.27%**	**50%**
Coeficiente de Z	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

8.15. MUESTREO ALEATORIO SIMPLE (M.A.S.)

Es el más elemental de los procedimientos estadísticos de muestreo, para facilitar la comprensión del tema a los estudiantes de formulación y evaluación de proyectos se harán los cálculos para variables discretas.

8.15.1. Cálculo del tamaño de la muestra cuando no se conoce la varianza poblacional. Para diseñar la muestra es indispensable contar con un marco muestral, es decir, la lista, mapa u otra especificación de las unidades, que resulta de la información previamente disponible. En la práctica, es muy frecuente que no se conozca la varianza de la característica en la población (σ^2); en tales casos se debe recurrir a censos, a investigaciones similares realizadas con anterioridad o a investigaciones preliminares, como son las encuestas pilotos. Este último mecanismo, es el que más se utiliza para el cálculo del tamaño de la muestra, en caso de que no se conozca información sobre la población.

La tabla 8.3 presenta una población que utilizaremos para ilustrar el cálculo aleatorio simple. Esta población está compuesta por tiendas de la ciudad de Popayán, que fue suministrada por la Cámara de Comercio del Cauca. En la tabla, hay tres segmentos de información así:

1. **Columna 1.** El número de tiendas en su orden.
2. **Columna 2.** El nombre o razón social.
3. **Columna 3**. Dirección.
4. **Columna 4.** Activos en millones de pesos.
5. **Columna 5.** Ingresos en millones de pesos.
6. **Columna 6.** La propiedad del local.

El primer paso que se debe hacer es tomar una muestra piloto (Tabla 8.4).

Tabla 8.3. Tiendas de la Ciudad de Popayán, discriminadas por el valor en activos, ingresos y propiedad del local.

# de ord.	Direccion	Activos en millones	Ingresos en millones de pesos	Local propio
1	-CARRERA 7E # 17BIS-17	1.000	13.888	no
2	-TRANSV. 19 # 10-121	1.000	7.518	no
3	-CARRERA 1AE # 9A-41	1.050	9.600	si
4	-CALLE 5B # 18-31	1.080	13.311	si
5	-GALERIA SUR PUESTO # 12	1.100	11.910	si
6	-CALLE 12 # 28A-04	1.120	14.300	si
7	-CALLE 68 # 10-92	1.150	0.450	si
8	-CALLE 4 # 25-73	1.150	11.450	si
9	-CALLE 19 # 30-24	1.170	12.075	si
10	-CALLE 5 # 27A-12	1.175	2.450	no
11	-MANZANA 3 NO. 42A-11	1.200	1.200	no
12	-CARRERA 41 # 2-13	1.200	0.400	no
13	-MANZ. 25 #25-16 TOMAS CIPRIAN	1.270	11.270	si
14	-CALLE 63N NO. 7A-09	1.414	3.000	no
15	-CALLE 8 # 5-28	1.433	14.900	si
16	-CALLE 5 # 43-24	1.473	1.573	no
17	-CALLE 16A # 4-51	1.500	1.500	si
18	-CALLE 29 BLOQUE H CASA No 5	1.500	4.900	si
19	-CARRERA 41 # 4-11	1.547	2.160	no
20	-CARRERA 3 # 8-07	1.650	19.187	si
21	-CARRERA 9 No 7-99	1.650	13.500	no
22	-CALLE 4 # 36-11	1.690	5.800	no
23	-CALLE 12 # 4-93	1.710	1.030	si
24	-CALLE 7 # 19-114	1.900	11.900	no
25	-CALLE 7 # 21-62	2.140	0.945	si
26	-CARRERA 2A No.7A-40	2.185	28.652	si
27	-CALLE 12 13-03	2.200	6.938	si
28	CRA 12 66N-72 BELLOHORIZONTE	2.300	44.308	si
29	-CARRERA 9 # 7N-02	2.343	7.819	si
30	-CALLE 9 # 17-25	2.800	15.500	no
31	-CRA 3 # 9-84	3.300	16.720	si
32	-CARRERA 41 # 2-13	3.600	24.166	si
33	PASAJE CENTRO COMERCIAL LOCAL 46	3.750	21.676	si
34	-CALLE 9 # 17-55	4.000	10.800	si
35	-CALLE 56N # 10-110	4.115	9.295	si
36	CALLE 17 # 6E-19 B/ LOS SAUCES.	4.600	10.500	no
37	-CALLE 20N # 8-47 POPAYAN	5.000	80.000	si
38	-CARRERA 2 # 3-93	5.550	14.590	no
39	-CALLE 5B # 18- 17	6.500	13.005	no
40	-CALLE 70D # 7-15	6.920	31.650	si
41	-CALLE 5 No 18-50 POPAYAN	7.200	117.197	no
42	-CARRERA 6 # 43N-50	7.350	11.550	si
43	-CARRERA 11 # 12A-10	8.795	2.450	si
44	-CALLE 19 No 31-14	10.000	30.641	no
45	-CARRERA 9 # 63N-18	10.028	35.381	si
46	-CARRERA 6 # 12-51	10.500	521.794	si
47	-CALLE 11N # 9-44	14.384	75.990	no
48	-CARRERA 7 # 12 -106	18.000	58.535	no
49	-CARRERA 6A # 9N-92 B/BOLIVAR	23.178	111.957	si
50	-CARRERA 4 # 7-35	29.469	747.544	si

Fuente: Cámara de Comercio del Cauca

Tabla 8.4. Muestra Piloto, muestreo aleatorio simple (M.A.S.)

No. Aleatorio	Activos X_i	Local	$(Xi - X)^2$
48	10,028	si	25,198
33	5,550	si	0,294
24	2,200	no	7,886
3	1,414	si	12,918
18	1,100	no	15,274
35	5,000	si	0,000
39	4,115	si	0,798
41	1,175	no	14,693
44	1,500	si	12,307
49	18,000	no	168,787
SUMA	50,082		258,156
MEDIA	5,008		
VARIANZA MUESTRAL			28,684

8.15.2. Cálculo de la muestra para variables continúas.

$$\overline{X} = \frac{\sum X_i}{n} = \frac{10{,}028 + 5.550 + ... + 18{,}000}{10}$$

$$\overline{X} = \frac{50.082}{10} = 5.008$$

Se estima un error del 10% del promedio de los activos así:

$$E = 0{,}10(\overline{X}) = 0{,}10 \times 5.008 = 0.5008$$

$$Varianza \quad muestral \quad S^2 = \frac{\sum (X - \overline{X})^2}{n-1} = \frac{258.156}{9} = 28.684$$

Una vez estimados estos estadísticos, se procede a calcular el tamaño de la muestra. Se trabajará con un nivel de confianza del 90%, es decir un valor de z = 1,645.

$$n = \frac{Z^2 \times N \times S^2}{N \times E^2 + Z^2 \times S^2}$$

$$n = \frac{(1{,}645)^2 \times 50 \times 28..684}{50 \times (0{,}5008)^2 + (1.645)^2 \times 28.684} = 43{,}045 \approx 43$$

$$n = \frac{S^2}{\left(\frac{E}{Z}\right)^2 + \frac{S^2}{N}}$$

$$n = \frac{28.684}{\left(\frac{0.5008}{1.645}\right)^2 + \frac{28.684}{50}} = 43{,}045 \approx 43$$

$$n_0 = \frac{\frac{Z^2 S^2}{E^2}}{1 + \frac{\frac{Z^2 S^2}{E^2}}{N}}$$

$$n_0 = \frac{\frac{(1{,}645)^2 * 28.684}{(0.5008)^2}}{1 + \frac{\frac{(1{,}645)^2 * 28.684}{(0.5008)^2}}{50}} = 43{,}045 \approx 43$$

Ficha Técnica

1. Título	:Encuesta a tenderos
2. Responsables	:Estudiantes de administración de empresas de la Universidad del Cauca.
3. Cobertura	:Ciudad de Popayán.
4. Unidad de análisis	:Tiendas con activos mayores a $1.000.000.
5. Periodo	: Septiembre de 2003.
6. PLAN DE MUESTREO	
6.1. Población objetivo	: Tiendas ubicadas en la ciudad de Popayán
6.2. Marco	: Listado suministrado por la Cámara de Comercio del Cauca.
6.3. Unidades de muestreo	: Propietarios o administradores de las tiendas con patrimonio mayor o igual a $ 1.000.000.
6.4. Nivel de confianza	: 90%
6.5. Coeficiente de conf.	: 1,645
6.6. Varianza muestral	: 258,156
6.7. Error máximo	: 5%
6.8. Tipo de diseño	: Muestra probabilísticas.
6.9. Procedimiento	: Muestreo aleatorio simple (M.A.S.)
6.10. Variables utilizadas	: Promedio de ventas, tipo de productos vendidos.
6.11. Tipo de entrevista	: Entrevista directa con formulario en papel.
6.12. Tamaño poblacional	: 50

Tamaño de la muestra = 43

8.15.3. Cálculo del tamaño de la muestra para variables discretas.

$$p = \frac{\sum a_i}{n} = \frac{6}{10} = 0{,}6$$

En este caso, seis tiendas de las diez, poseen local propio correspondiendo al 60%, es decir, que el valor de p = 0.60 y q =0.4

$$pq = 0{,}6 \times 0{,}4 = 0{,}24$$

Entonces el tamaño de la muestra es:

$$n = \frac{Z^2 \times N \times p \times q}{N \times E^2 + Z^2 \times p \times q}$$

$$n = \frac{(1{,}645)^2 \times 50 \times 0{,}6 \times 0{,}40}{50 \times (0{,}10)^2 + (1.645)^2 \times 0{,}6 \times 0{,}40} = 28{,}25 \approx 28$$

Ficha Técnica

1. Título	:Encuesta a tenderos
2. Responsables	:Estudiantes de administración de empresas de la Universidad del Cauca.
3. Cobertura	:Ciudad de Popayán.
4. Unidad de análisis	:Tiendas con activos mayores a $1.000.000.
5. Periodo	: Septiembre de 2003.
6. PLAN DE MUESTREO	
6.1. Población objetivo	: Tiendas ubicadas en la ciudad de Popayán
6.2. Marco	: Listado suministrado por la Cámara de Comercio del Cauca.
6.3. Unidades de muestreo	: Propietarios o administradores de las tiendas con patrimonio mayor o igual a $ 1.000.000.
6.4. Nivel de confianza	: 90%
6.5. Coeficiente de conf.	: 1,645
6.6. Varianza muestral	: P=0,5 y Q=0,5
6.7. Error máximo	: 10%
6.8. Tipo de diseño	: Muestra probabilísticas.
6.9. Procedimiento	: Muestreo aleatorio simple (M.A.S.)
6.10. Variables utilizadas	: Promedio de ventas, tipo de productos vendidos.
6.11. Tipo de entrevista	: Entrevista directa con formulario en papel.
6.12. Tamaño poblacional	: 50

Tamaño de la muestra = 28

8.16. TAMAÑO ÓPTIMO DE LA MUESTRA CON VARIABLES DISCRETAS

Una forma de obtener el tamaño óptimo de la muestra asignado el valor de 0.5 a p y 0.5 a q. En la tabla 8.5 se realizaron los cálculos para diferentes valores de p y q. Como se puede observar el mayor valor se obtiene cuando estos parámetros toman estos valores.

Estadísticos de la ficha técnica:

N = 50
E= 2%
Z= 1,96

Tabla 8.5. Tamaño de la muestra, con diferentes valores de p.

p	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%
q	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%
n=	45	47	48	48	49	49	49	49	48,97	48,98	48,97	49	49	49	49	48	48	47	45

8.17. RELACIÓN DEL TAMAÑO DE LA MUESTRA CON EL NIVEL DE CONFIANZA Y EL ERROR.

En la tabla 8.6 se calculó el tamaño de la muestra dado los niveles de confianza del 96%, 95% y 80%, así para errores del 4%, 5% y el 20% sobre el valor de la media ($\overline{X} = 5$) y $S^2 = 28{,}684$

Tabla 8.6. Relación del tamaño de la muestra con el nivel de confianza y error.

Un mayor nivel de confianza menor tamaño de la muestra

Un mayor error menor tamaño de la muestra

N=50		NIVEL DE CONFIANZA		
		96%	95%	80%
		Z=2,05	Z=1,96	Z=1,28
ERROR	E=4%*5=0,2	49	49	48
	E=5%*5=0,25	48	48	47
	E=20%*5=1,0	35	33	23

Se puede concluir, que un mayor nivel de confianza, por ejemplo 96% y un error bajo en este caso del 4%, el tamaño de la muestra es de 49, que es casi el tamaño de la población; caso contrario, ocurre con un nivel de confianza bajo como 80% y un error alto como del 20%, el tamaño de la muestra es de 23, es decir que la muestra se reduce considerablemente, lo cual puede traer como consecuencia un estudio o investigación poco confiable.

8.18. MUESTREO ESTRATIFICADO - ASIGNACIÓN PROPORCIONAL

Este método permite determinar el tamaño óptimo de la muestra, así como los estimados puntuales y límites de confianza para el promedio, proporción, razón y proporciones en conglomerados. Con este método, los tamaños muestrales en cada estrato, se distribuyen en la misma proporción que las unidades en la población de cada uno de ellos; en otras palabras, el peso relativo dado por el número de unidades en cada estrato, en relación al total de elementos de la población, debe ser igual al obtenido en la muestra.

Antes de determinar el tamaño de la muestra, se elaboró la estratificación para la población de 50 tiendas. Se distribuyó la población de tiendas en dos estratos así:

Tabla 8.7. Estrato I: Tiendas con Activos menores e iguales a $2.000.000

# de ord.	Direccion	Activos en millones	Ingresos en millones de pesos	Local propio
1	-CARRERA 7E # 17BIS-17	1.000	13.888	no
2	-TRANSV. 19 # 10-121	1.000	7.518	no
3	-CARRERA 1AE # 9A-41	1.050	9.600	si
4	-CALLE 5B # 18-31	1.080	13.311	si
5	-GALERIA SUR PUESTO # 12	1.100	11.910	si
6	-CALLE 12 # 28A-04	1.120	14.300	si
7	-CALLE 68 # 10-92	1.150	0.450	si
8	-CALLE 4 # 25-73	1.150	11.450	si
9	-CALLE 19 # 30-24	1.170	12.075	si
10	-CALLE 5 # 27A-12	1.175	2.450	no
11	-MANZANA 3 NO. 42A-11	1.200	1.200	no
12	-CARRERA 41 # 2-13	1.200	0.400	no
13	-MANZ. 25 #25-16 TOMAS CIPRIAN	1.270	11.270	si
14	-CALLE 63N NO. 7A-09	1.414	3.000	no
15	-CALLE 8 # 5-28	1.433	14.900	si
16	-CALLE 5 # 43-24	1.473	1.573	no
17	-CALLE 16A # 4-51	1.500	1.500	si
18	-CALLE 29 BLOQUE H CASA No 5	1.500	4.900	si
19	-CARRERA 41 # 4-11	1.547	2.160	no
20	-CARRERA 3 # 8-07	1.650	19.187	si
21	-CARRERA 9 No 7-99	1.650	13.500	no
22	-CALLE 4 # 36-11	1.690	5.800	no
23	-CALLE 12 # 4-93	1.710	1.030	si
24	-CALLE 7 # 19-114	1.900	11.900	no

Fuente: Cámara de Comercio del Cauca

Tabla 8.8. Estrato II. Tiendas con activos de más de $2.000.0001

# de ord.	Direccion	Activos en millones	Ingresos en millones de pesos	Local propio
1	-CALLE 7 # 21-62	2.140	0.945	si
2	-CARRERA 2A No.7A-40	2.185	28.652	si
3	-CALLE 12 13-03	2.200	6.938	si
4	CRA 12 66N-72 BELLOHORIZONTE	2.300	44.308	si
5	-CARRERA 9 # 7N-02	2.343	7.819	si
6	-CALLE 9 # 17-25	2.800	15.500	no
7	-CRA 3 # 9-84	3.300	16.720	si
8	-CARRERA 41 # 2-13	3.600	24.166	si
9	PASAJE CENTRO COMERCIAL LOCAL 46	3.750	21.676	si
10	-CALLE 9 # 17-55	4.000	10.800	si
11	-CALLE 56N # 10-110	4.115	9.295	si
12	CALLE 17 # 6E-19 B/ LOS SAUCES.	4.600	10.500	no
13	-CALLE 20N # 8-47 POPAYAN	5.000	80.000	si
14	-CARRERA 2 # 3-93	5.550	14.590	no
15	-CALLE 5B # 18- 17	6.500	13.005	no
16	-CALLE 70D # 7-15	6.920	31.650	si
17	-CALLE 5 No 18-50 POPAYAN	7.200	117.197	no
18	-CARRERA 6 # 43N-50	7.350	11.550	si
19	-CARRERA 11 # 12A-10	8.795	2.450	si
20	-CALLE 19 No 31-14	10.000	30.641	no
21	-CARRERA 9 # 63N-18	10.028	35.381	si
22	-CARRERA 6 # 12-51	10.500	521.794	si
23	-CALLE 11N # 9-44	14.384	75.990	no
24	-CARRERA 7 # 12 -106	18.000	58.535	no
25	-CARRERA 6A # 9N-92 B/BOLIVAR	23.178	111.957	si
26	-CARRERA 4 # 7-35	29.469	747.544	si

Fuente: Cámara de Comercio del Cauca

Nomenclatura para calcular las proporciones del muestreo estratificado.

N	Total de unidades que constituyen la población objetivo.
N_h	Total de unidades que contiene cada estrato poblacional.
N_1, N_2, N_3 etc.	Serán los tamaños poblacionales en los estratos 1, 2, 3 etc.
$\Sigma N_h = N = N_1 + N_2 + N_3 + \ldots N_M$	
$\bar{Y}_h = \dfrac{\sum Y_{hj}}{N_h}$	Media aritmética poblacional para cada estrato.

$$\bar{Y}_{st} = \frac{\sum Y_h N_h}{N} = \sum \bar{Y}_h W_h$$

Media aritmética poblacional estratificada ponderada.

$$W_h = \frac{N_h}{N} \qquad W_1 = \frac{N_1}{N}$$

$$W_2 = \frac{N_2}{N} \qquad W_3 = \frac{N_3}{N}$$

$$\sum W_h = W_1 + W_2 + W_2 + ...W_M = 1$$

Proporción de elementos de cada estrato.

$$S^2{}_h = \frac{\sum Y^2{}_{hj} - N_h \bar{Y}^2{}_h}{N_h - 1}$$

Varianza poblacional en cada estrato.

n = Número de unidades que contiene la muestra total.

nh = Número de unidades que contiene la muestra en cada estrato muestral.

$$\sum n_h = n = n_1 + n_2 + n_3 + ...$$

$$\bar{y}_h = \frac{\sum \bar{y}_{hj}}{n_h} \qquad \bar{y}_1 = \frac{\sum \bar{y}_{1j}}{n_1}$$

$$\bar{y}_2 = \frac{\sum \bar{y}_{2j}}{n_2} \qquad \bar{y}_3 = \frac{\sum \bar{y}_{3j}}{n_3}$$

$\bar{y}_h = Media\ \ aritmética\ \ muestral\ \ para\ \ cada\ \ estrato.$

$$\bar{y}_{st} = \frac{\sum N_h \bar{y}_h}{N} \qquad \bar{y}_{st} = \sum \bar{y}_h W_h$$

Media aritmética muestral estratificada.

$$s^2{}_h = \frac{\sum y^2{}_{hj} - n_h \bar{y}^2{}_h}{n_h - 1}$$

Varianza muestral en cada estrato.

8.18.1. Cálculo del tamaño de la muestra. Se debe calcular la proporción de unidades en cada estrato y peso relativo.

Tabla 8.9. Cálculo de las proporciones de las variables discretas.

Estrato	Criterio de selección	No. Unid.	Proporción
Estrato I	Activos menores o iguales a $ 2.000.000	24	$W_1 = \frac{N_1}{N} = \frac{24}{50} = 0{,}48$
Estrato II	Activos mayores o guales a $ 2.000.0001.	26	$W_2 = \frac{N_2}{N} = \frac{26}{50} = 0{,}52$

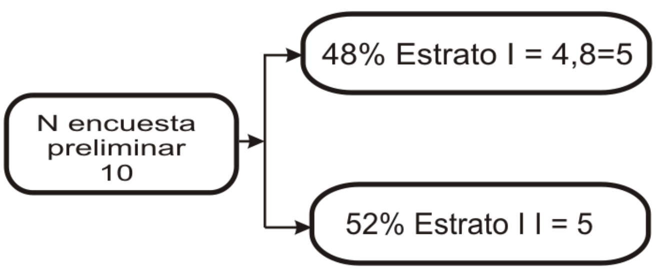
$$W_1 + W_2 = 0{,}48 + 0{,}52 = 1$$

$$\sum W_h = 1 ... ó ... 100\%$$

El tamaño de la muestra piloto es 10 encuestas.

En la asignación proporcional las 10 encuestas se distribuyen de la siguiente manera:

Figura 8.7. Tamaños proporcionales de la muestra según estratos



Determinados los tamaños muestrales para la encuesta preliminar, se procede a la selección de las unidades requeridas, que permitirán calcular las varianzas y el error. Haciendo uso de una tabla de números aleatorios para los dos estratos, se seleccionó de la siguiente forma:

Tabla 8.10. Muestra piloto Estrato I.

No. De orden	Números aleatorios	Local propio	Activos en miles de pesos	$(X - \overline{X})^2$
1	24	no	1,900	0,072
2	20	si	1,650	0,000
3	14	no	1,414	0,047
4	22	no	1,690	0,004
5	17	si	1,500	0,017
Totales			8,154	46,879
			Media	Varianza
Media y Varianza			1,631	11,720

Tabla 8.11. Muestra piloto Estrato II.

No. De orden	Números aleatorios	Local propio	Activos en miles de pesos	$(X - \overline{X})^2$
1	24	no	18,000	98,804
2	6	no	2,800	27,668
3	13	si	5,000	9,364
4	10	si	4,000	16,484
5	22	si	10,500	5,954
Totales			40,300	186,956
			Media	Varianza
Media y Varianza			8,060	46,739

La media ponderada es igual a:

$$\bar{x}_{st} = \sum W_h \bar{x}_h = 0,48(1,631) + 0,52(8.060) = 4.974$$

El error de muestreo corresponde a 0.05, $E = 0,05(\bar{x}) = 0,05(4.974) = 0,2487$

Se calcula de la varianza ponderada:

$$S^2{}_{st} = \sum W_h S^2{}_h$$

$$S^2{}_{st} = 0,48(11.72) + 0,52(0,46.739) = 29.93$$

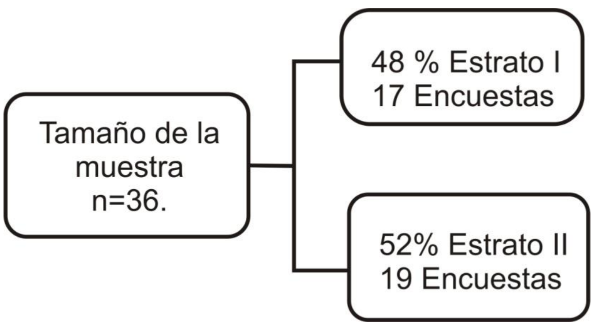
El tamaño de la muestra es igual a:

$$n = \frac{Z^2 N S^2}{N E^2 + Z^2 S^2}$$

$$n = \frac{1{,}645^2 \times 50 \times 29.93}{(50 \times 0{,}806^2) + (1{,}96^2 \times 29.93)} = 35.687 \cong 36$$

Una vez calculado el tamaño de la muestra se procede a distribuir el tamaño de la muestra en forma proporcional a los estratos así:

Figura 8.8. Tamaños proporcionales de la muestra según estrato.



Ahora veamos cómo sería el cálculo de n para variables discretas o atributos. Para ello consideramos como característica cualitativa la propiedad del local comercial, se establece un nivel confianza del 95% (Z = 1,96) y un error del 5%.

Tabla 8.12. Cálculo del parámetro p.

Estrato I	Estrato II
$P = \frac{2}{5} = 0{,}4$	$P = \frac{3}{5} = 0{,}6$
$Q = \frac{3}{5} = 0{,}6$	$Q = \frac{2}{5} = 0{,}4$
$P + Q = 1$	$P + Q = 1$

Calculamos PQ ponderado así:

$$PQ_{st} = \sum W_h P_h Q_h$$

$$PQ_{st} = 0{,}48(0{,}40)(0{,}60) + 0{,}52(0{,}60)(0{,}40) = 0{,}24$$

Se procede a calcular el tamaño de la muestra así:

$$n = \frac{Z^2 NPQ}{NE^2 + Z^2 PQ}$$

$$n = \frac{1{,}96^2 \times 50 \times 0{,}24}{50 \times 0{,}05^2 + 1{,}96^2 \times 0{,}24} = 44$$

Ficha Técnica

1. Título	:Encuesta a tenderos
2. Responsables	:Estudiantes de administración de empresas de la Universidad del Cauca.
3. Cobertura	:Ciudad de Popayán.
4. Unidad de análisis	:Tiendas con activos mayores a $1.000.000.
5. Periodo	: Septiembre de 2003.
6. PLAN DE MUESTREO	
6.1. Población objetivo	: Tiendas ubicadas en la ciudad de Popayán
6.2. Marco	: Listado suministrado por la Cámara de Comercio del Cauca.
6.3. Unidades de muestreo	: Propietarios o administradores de las tiendas con patrimonio mayor o igual a $ 1.000.000.
6.4. Nivel de confianza	: 95%
6.5. Coeficiente de conf.	: 1,96
6.6. Varianza muestral	: PQ PONDERADO=0,24
6.7. Error máximo	: 5%
6.8. Tipo de diseño	: Muestra probabilísticas.
6.9. Procedimiento	:Muestreo Estratificado -Asignación Proporcional
6.10. Variables utilizadas	: Promedio de ventas, tipo de productos vendidos.
6.11. Tipo de entrevista	: Entrevista directa con formulario en papel.
6.12. Tamaño poblacional	: 50

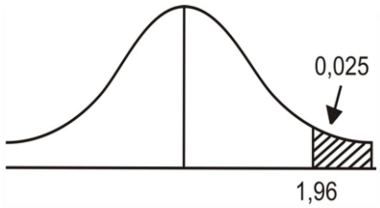
Tamaño de la muestra = 44

PREGUNTAS DE ANÁLISIS

1. Defina:

- Elemento.
- Población.
- Unidad de muestreo.
- Marco muestral.
- Estadístico.

2. Describa el proceso de diseño de la muestra.
3. ¿En qué se diferencia las técnicas de muestreo probabilístico de las técnicas de muestreo no probabilístico?
4. ¿Cuáles son las características distintivas del muestreo aleatorio simple?
5. ¿Describa el muestreo estratificado?
6. ¿Cómo se diferencia el nivel de confianza y el grado precisión?
7. ¿Cómo se afecta el tamaño de la muestra cuando se incremente el nivel de confianza?
8. ¿Cómo se afecta el tamaño de la muestra cuando se aumenta el error?



APÉNDICE.8. A.1.
DISTRIBUCIÓN NORMAL

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,10	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,20	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,30	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,40	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,50	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,60	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,70	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,80	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,90	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,00	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,10	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,20	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,30	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,40	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,50	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,60	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,70	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,80	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,90	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,00	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,10	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,20	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,30	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,40	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,50	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,60	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,70	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,80	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,90	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,00	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,10	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,20	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,30	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,40	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998

APÉNDICE 8. A.2.

TABLA PARA DETERMINAR EL TAMAÑO DE LA MUESTRA

Margen de confianza del	95%
z=	1,96
Hipótesis p=	50%

POBLACIÓN	TAMAÑO DE LA MUESTRA SEGÚN MARGENES DE ERROR					
N	**1%**	**2%**	**3%**	**4%**	**5%**	**10%**
500					217	81
1.000				375	278	88
1.500			624	429	306	90
2.000			696	462	322	92
2.500		1.225	748	484	333	92
3.000		1.334	787	500	341	93
3.500		1.424	818	512	346	93
4.000		1.500	842	522	350	94
4.500		1.566	863	530	354	94
5.000		1.622	879	536	357	94
5.500		1.671	894	541	359	94
6.000		1.715	906	546	361	95
6.500		1.753	917	550	363	95
10.000	4.899	1.936	964	566	370	95
15.000	5.855	2.070	996	577	375	95
20.000	6.488	2.144	1.013	583	377	96
25.000	6.939	2.191	1.023	586	378	96
30.000	7.275	2.223	1.030	588	379	96
35.000	7.536	2.247	1.036	590	380	96
40.000	7.745	2.265	1.039	591	381	96
60.000	8.279	2.309	1.048	594	382	96
65.000	8.368	2.315	1.050	595	382	96
70.000	8.445	2.321	1.051	595	382	96
100.000	8.762	2.345	1.056	597	383	96
200.000	9.164	2.373	1.061	598	383	96
300.000	9.306	2.382	1.063	599	384	96
900.000	9.503	2.395	1.066	600	384	96
1.000.000	9.513	2.395	1.066	600	384	96

APÉNDICE.8.A.3.

TABLA PARA DETERMINAR EL TAMAÑO DE LA MUESTRA

Margen de confianza del	99%
z=	2,58
Hipótesis p=	50%

POBLACIÓN	TAMAÑO DE LA MUESTRA SEGÚN MARGENES DE ERROR					
N	**1%**	**2%**	**3%**	**4%**	**5%**	**10%**
500					286	125
1.000				510	400	143
1.500			828	614	461	150
2.000			961	684	499	154
2.500		1.562	1.063	734	526	156
3.000		1.743	1.144	772	545	158
3.500		1.901	1.210	802	559	159
4.000		2.039	1.264	825	571	160
4.500		2.162	1.311	845	580	160
5.000		2.271	1.350	861	587	161
5.500		2.369	1.384	875	594	162
6.000		2.457	1.413	886	599	162
6.500		2.537	1.440	897	604	162
10.000	6.246	2.938	1.560	942	624	164
15.000	7.889	3.257	1.646	973	637	165
20.000	9.083	3.444	1.693	989	644	165
25.000	9.991	3.567	1.722	999	648	165
30.000	10.704	3.654	1.742	1.005	651	165
35.000	11.279	3.718	1.756	1.010	653	166
40.000	11.752	3.768	1.767	1.014	655	166
60.000	13.028	3.890	1.794	1.022	658	166
65.000	13.249	3.910	1.798	1.024	659	166
70.000	13.445	3.927	1.801	1.025	659	166
100.000	14.267	3.994	1.815	1.029	661	166
200.000	15.363	4.075	1.832	1.035	663	166
300.000	15.766	4.103	1.838	1.036	664	166
900.000	16.339	4.141	1.845	1.039	665	166
1.000.000	16.369	4.143	1.846	1.039	665	166

Capítulo 9

Métodos de proyección

Objetivos

Al finalizar este capítulo el estudiante estará en condiciones de:

- **Identificar y diferenciar** las técnicas de predicción.
- **Comprender el** análisis de regresión y correlación.
- **Identificar** la función de regresión lineal (FRL).
- **Explicar y comprender** el método de mínimos cuadrados ordinarios (MCO).
- **Diferenciar** la función de regresión muestral de la poblacional.
- **Interpretar** los parámetros de la función de regresión.
- **Interpretar y explicar** el coeficiente de determinación r^2.
- **Proyectar** las cifras mediante los parámetros de la Función de regresión muestral (FRM).
- **Conocer y utilizar** las herramientas de análisis de regresión de la hoja de cálculo EXCEL versiones 2003 y 2007.
- **Hacer uso de las funciones de TENDENCIA y PRONÓSTICO** de la hoja de calculo EXCEL.

9. MÉTODOS DE PROYECCIÓN

El pronóstico de la situación futura del mercado del producto, es uno de los aspectos de mayor importancia para todo proyecto. Debido a los continuos cambios que se presentan específicamente en variables de tipo económico, social, político, tecnológico y ecológico, los resultados proyectados, deben ser analizados con cierta reserva y tanto las cantidades demandadas y ofrecidas con los precios esperados de los años siguientes, deben tomarse como aproximaciones o datos de referencia para análisis y toma de decisiones y no como cálculos de absoluta certeza.

Antonio pulido, argumenta que:

> *"La predicción debe entenderse como un intento permanente de anticipación de un futuro incierto y sobre el que, además podemos incidir. Nada más lejano a un buen enfoque que la supuesta elaboración estática de predicciones justificadas como certeras. Defendemos la predicción que hace explícitos sus condicionantes, que marcan el camino de su propia revisión, que permite la elección entre hipotéticos futuros alternativos[2]".*

Existen variadas alternativas para determinar el comportamiento futuro de un producto, lo cual exige de los proyectistas, un análisis detenido de las implicaciones y exigencias de cada método, para seleccionar y aplicar correctamente el que se considere más adecuado a la situación objeto de estudio. De acuerdo con nuestra experiencia de más de 10 años dirigiendo y acompañando a los estudiantes en la elaboración de sus proyectos, consideramos oportuno mencionar algunos aspectos que nos condicionan a plantear algunas técnicas acorde con el medio, cantidad y calidad de información disponible.

Las siguientes son algunas de las limitaciones que se le presentan a los estudiantes en el momento de hacer sus proyectos:

1. Falta de base de datos regionales de tipo económico, pese a los esfuerzos que han venido adelantando las instituciones locales.
2. El problema de respuestas en blanco en las encuestas, que trae como consecuencia que no se refleje verdaderamente el comportamiento del mercado.
3. En la mayoría de los casos los métodos utilizados para la recolección de la información (trabajo de campo) no son bien aplicados, ya sea por negligencia de los estudiantes o por falta de colaboración de los encuestados.
4. Las cifras económicas generalmente están disponibles a niveles agregados. Por ejemplo, la mayor parte de los indicadores económicos como el PIB, ingreso disponible, consumos per-cápita están disponibles para la economía como un todo.
5. Hay información de carácter confidencial que las instituciones no están autorizadas para suministrar. Por ejemplo: La Administración de Impuestos y Aduanas Nacionales

[2] *PULIDO SAN ROMA, Antonio. Predicción Económica y empresarial. Madrid. Pirámides, S.A. 1989. Prólogo.*

(DIAN), no está autorizada para proporcionar información sobre declaraciones individuales de renta, solo puede dar algunos datos generales.

Debido a estos y otros múltiples problemas y limitaciones de información; presentamos en el punto siguiente las técnicas de predicción que se desarrollarán en este texto. De otra parte el estudiante encuentra una amplia bibliografía de técnicas de pronósticos, de estadística y econometrícas que explican sofisticados métodos como: Agregación de predicciones individuales, series de tiempo, AR (p), MA(q) y ARIMA modelos autoregresivos, Box-Jenkins y descomposición de series, entre otros.

9.1. TÉCNICAS DE PREDICCIÓN

Como se mencionó, presentamos esta clasificación porque nos hemos convencido que las técnicas más empleadas, dadas las limitaciones del medio, son habitualmente las más sencillas, y con un alto grado de satisfacción y confiabilidad. De otra parte, estas técnicas también se pueden combinar entre sí, como veremos más adelante.

1. **Predicciones en base a relaciones entre variables.** En esta técnica se clasifican los modelos econométricos, específicamente las funciones de Regresión, mediante el método de mínimos cuadrados ordinarios (MCO). Estos son métodos estadísticos que buscan establecer una ecuación, que permite estimar el valor desconocido de una variable, a partir del valor conocido de una o más variables, también se estudiará las formas funcionales del modelo como: Log lineal y semilogarítmicas.

2. **Predicción en base a información subjetiva.** Son datos suministrados por individuos aisladamente o como ponentes de un grupo con base a sus experiencias, opiniones y expectativas de futuro, entre los medios utilizados están: Encuestas de actitudes o sentimientos, encuestas de intención o expectativas, panel de expertos y técnicas exploratorias y normativas.

3. **Predicción con base en series de tiempo.** La característica clave de este enfoque es el estudio de un fenómeno en si mismo, a través de su evaluación temporal (Series temporales o históricas). También se analizará la técnica de alisado o suavización de series mediante promedios móviles ponderados.

9.2. PREDICCIONES EN BASE A LA RELACIÓN ENTRE VARIABLES

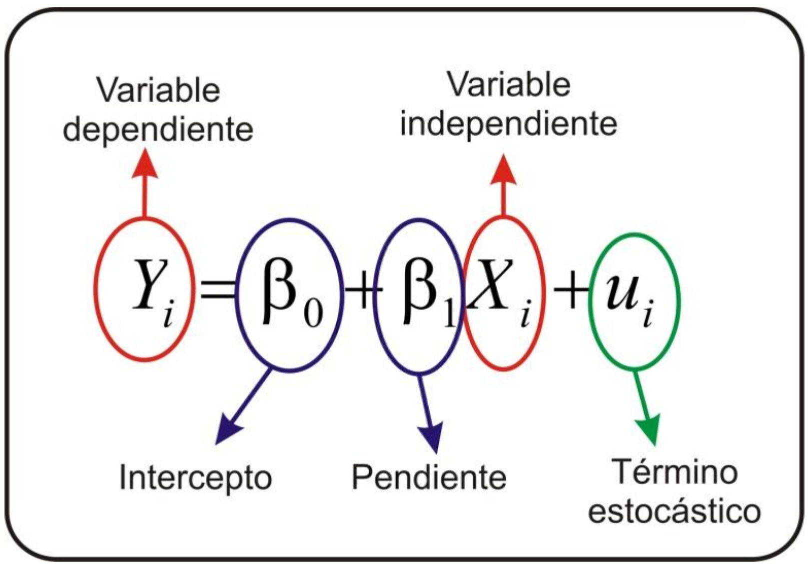
La dependencia entre el comportamiento de diferentes fenómenos posibilita la introducción de predicciones condicionadas con diversos grados de formalización y un análisis de causalidad. La técnica que se incluye en esta clasificación es el Análisis de Regresión y Correlación, mediante el método de Mínimos Cuadrados Ordinarios (MCO).

9.2.1. Análisis de regresión y correlación[3]. El análisis de regresión trata del estudio de la dependencia de la variable dependiente, en una o más variables; las variables explicativas, con el objetivo de estimar y/o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las últimas.

El análisis de regresión da lugar a una ecuación matemática que permite describir la relación existente entre dos variables. Es decir, obtener una línea "ideal" conocida como línea de regresión que nos describa la relación o dependencia entre dos variables.

Esta línea o función matemática, en caso de una sola variable independiente o explicativa (X_i) puede ser expresada, a través de una ecuación presentada en la figura 9.1.

Figura 9.1. Función de regresión lineal con una variable independiente.



.

Este modelo planteado con una sola variable independiente, frecuentemente es inadecuado para explicar comportamientos del mercado, por ejemplo: al plantear una función de ingreso - consumo donde se toma como variable dependiente (Y_i) el consumo, y la variable independiente (X_i) el ingreso (*seteris peribus*), es decir, que las demás variables no se tienen en cuenta.

$$Consumo(Y_i) = \beta_1 + \beta_2 Ingreso(X_i) + u_i \qquad (9.1.)$$

---

[3] *Los conceptos visto en esta técnica "Predicción en base a la relación entre variables", son una adaptación a los contenidos en los capítulos uno, dos y tres de GUJARATI. Damodar N. Econometría. 2003. McGrawHill. México.*

Pero en la práctica las cosas no son tan simples, dado que, además del ingreso, existen otras variables que probablemente afectan el gasto de consumo, como el tamaño del grupo familiar.

$$Consumo(Y_i) = \beta_1 + \beta_2 Ingreso(X_i) + \beta_3 Grupofamiliar + u_i \qquad (9.2)$$

En este sentido este modelo se extiende a otro tipo de variables, que denominaremos función de regresión múltiple

9.2.2. Método de mínimos cuadrados ordinarios (MCO). El método de mínimos cuadrados ordinarios se atribuye a Carl Friedrich Gauss, un matemático alemán. Bajo ciertos supuestos, este método de mínimos cuadrados tiene algunas propiedades estadísticas muy atractivas, que lo han convertido en uno de los más eficaces y populares del análisis de regresión. Para entenderlo, se explicará el principio de los mínimos cuadrados.

La Función de regresión poblacional de dos variables

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (9.3.)$$

Esta función no es observable directamente, ya sea porque no se dispone de la información, porque no hay entes especializados en la recolección de la misma o por el costo que implica realizar un censo. Entonces, esta función debe ser estimada por una función muestral, dado que la mayoría de las investigaciones, por lo general, trabajan con muestras tomadas de la población, en otras palabras, trabajan con la **función de regresión muestral (FRM).** La función de regresión muestral se denotara con un gorro en los parámetros, por ejemplo $\hat{\beta}_1$.

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + u_i \\ Y_i &= \hat{Y}_i + u_i \end{aligned} \qquad (9.4.)$$

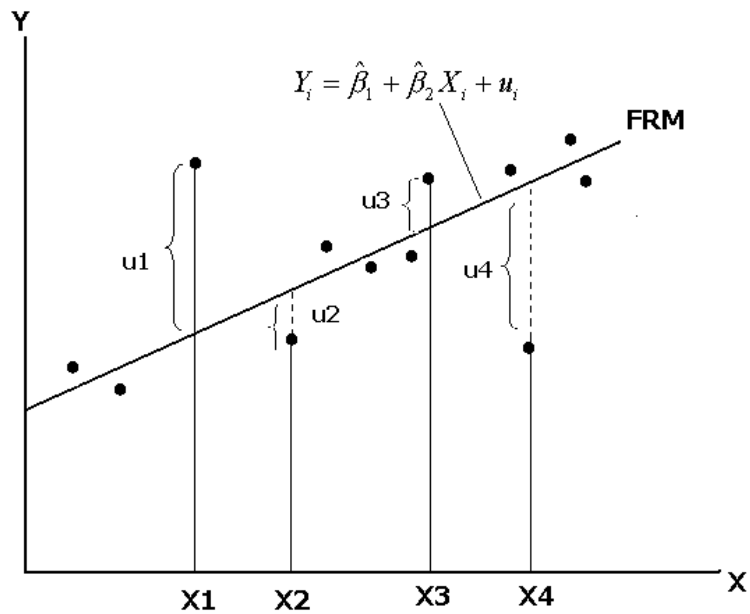
Ahora, dados n pares de observaciones de Y y X, se estará interesado en determinar la FRM de tal manera que esté lo más cerca posible a Y observado. Con este fin, se puede adoptar el siguiente criterio: seleccionar la FRM de tal manera que la suma de los residuos sea la menor posible. Este criterio, aunque es intuitivamente atractivo, no es muy bueno, como puede verse en el diagrama de dispersión que aparece en la Figura 9.2

Si se adopta el criterio de minimizar $\sum u_i$, la Figura 9.2 muestra que los residuos u_2 y u_3 al igual que los residuos u_1 y u_4 reciben el mismo peso en la suma ($u_1 + u_2 + u_3 + u_4$), aunque los dos primeros están mucho más cerca de la FRM que los dos últimos. Es decir, que a todos los residuales se les da la misma importancia sin importar qué tan cerca o qué tan lejos estén las observaciones individuales de la FRM. De tal forma, que es muy posible que la suma algebraica de los u_i sea pequeña (aun cero). Se puede evitar este problema si se adopta el criterio de mínimos cuadrados, el cual establece que la FRM pueda determinarse en forma que la suma de los u_i sea lo más pequeña posible.

$$\sum \hat{u}_i^{\;2} = \sum (Y_i - \hat{Y}_i)^2$$
$$\sum \hat{u}_i^{\;2} = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2 \quad (9.5.)$$

Donde u_{i2} son los residuos elevados al cuadrado. Al elevar al cuadrado u_i este método da más peso a los residuos tales como u_1 y u_4 en la figura No. 9.2 que a los residuos u_2 y u_3.

Figura 9.2 Función de regresión muestral.



El principio de mínimos cuadrados escoge β_1 y β_2 de tal manera que para una muestra dada o conjunto de datos, $\sum u_i^{\;2}$ es la más pequeña posible. En otras palabras, una muestra dada, proporciona valores estimados únicos de β_1 y de β_2 que producen el valor más pequeño o reducido posible de $\sum u_i^{\;2}$.

Haciendo uso del cálculo y derivando la ecuación 9.1 con respecto β_1 y β_2 y mediante algunas manipulaciones algebraicas se obtienen las ecuaciones normales:

Primera ecuación $\sum Y = n\hat{\beta}_i + \hat{\beta}_2 \sum X_i$ (9.6.)

Segunda ecuación $\sum X_i Y_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^{\,2}$ (9.7.)

Resolviendo las ecuaciones normales simultáneamente, se obtiene:

Método de mínimos cuadrados ordinarios (MCO).
(9.8.)

$$\hat{\beta}_1 = \frac{\sum X_i^{\,2} \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^{\,2} - \left(\sum X_i\right)^2} \tag{9.9.}$$

Los estimadores obtenidos previamente se conocen como estimadores de mínimos cuadrados.

9.2.3. Coeficiente de determinación r^2. Es considerada la medida que determina la bondad del ajuste, es decir, que tan bien se ajusta la línea de regresión a los datos, es claro que si todas las observaciones caen en la línea de regresión, se obtendrá un ajuste perfecto, pero raramente se presenta este caso. Generalmente hay algunas u_i positivas y algunas u_i negativas, se espera que estos residuos alrededor de la muestra sean lo más pequeños posibles. El coeficiente de determinación r^2 (caso de dos variables) o R^2 (regresión múltiple) es una medida resumen que nos dice qué tan bien se ajusta la línea de regresión muestral a los datos, en otras palabras **el coeficiente de determinación (muestral) r^2 mide la proporción o el porcentaje de la variabilidad en Y explicada por el modelo de regresión.**

9.2.4. Coeficiente de correlación r. Es un parámetro estrechamente relacionado con r^2 pero conceptualmente muy diferente de éste, es el coeficiente de correlación, el cual es una medida de asociación entre dos variables. Puede ser calculado a partir de:

$$r = \pm\sqrt{r^2} \tag{9.10}$$

o a partir de su definición:

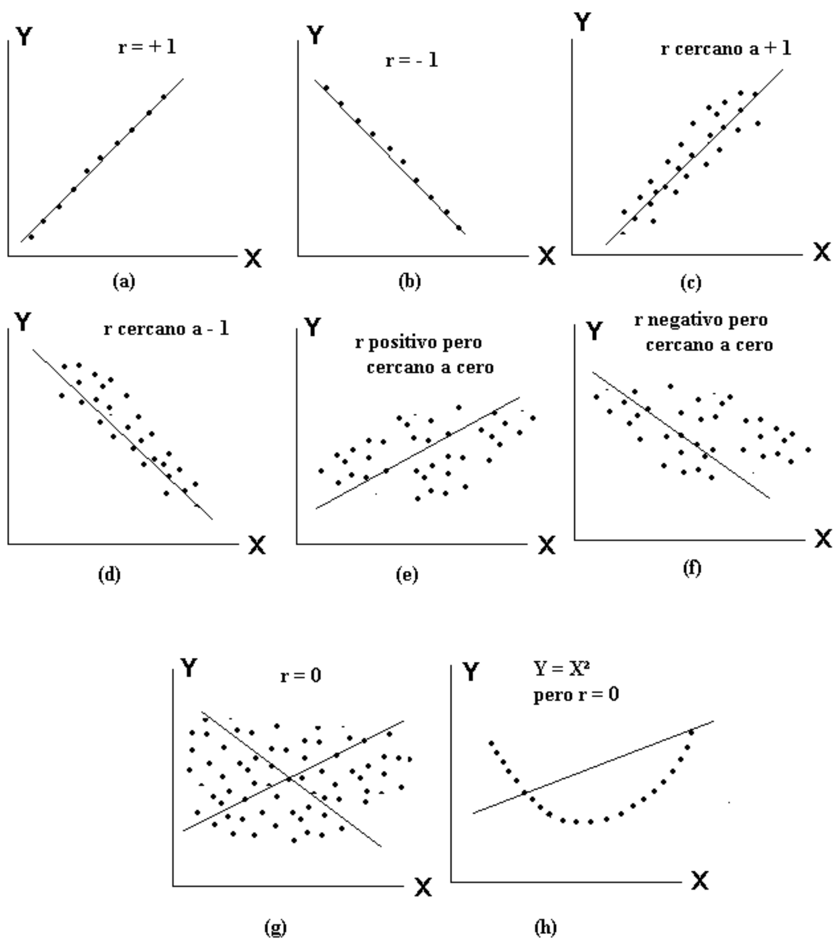
$$r = \frac{\sum x_i y_i}{\sqrt{(\sum x_i^{\,2})(\sum y_i^{\,2})}} \tag{9.11}$$

$$r = \frac{n\sum X_i Y_i - \left(\sum X_i\right)\left(\sum Y_i\right)}{\sqrt{..\left[\; n\sum X_i^{\;2} - (\sum X_i)^2 \;\right] \times \left[\; n\sum Y_i^{\;2} - (\sum Y_i)^2 \;\right]}} \qquad (9.12)$$

En el contexto de la regresión, r^2, es una medida con más significado que r. El grado de correlación lo podemos clasificar tentativamente, evitando un tanto la rigidez de sus límites

a) Correlación perfecta, cuando	(r = 1 o –1)
b) Correlación excelente, cuando	(-1 < r < -0,90)
c) Correlación aceptable, cuando	(-0,90 < r < -0,80)
d) Correlación regular, cuando	(-0,80 < r < -0,60)
e) Correlación mínima, cuando	(-0,60 < r < -0,30)
f) No hay correlación, cuando	(-0,30 < r < 0)

Figura 9.3. Coeficiente de determinación r^2.



Ejercicio

Se toma los datos de un proyecto de comercialización de carne de res en canal, realizado por estudiantes de Agrozootecnia de la Facultad de ciencias Agropecuarias de la Universidad del cauca. A continuación se presenta la información disponible.

Tabla 9.1. Series estadísticas de las variables utilizadas en el modelo.

Años	**Res** Consumo per cápita de carne de res **(ton/año)**	**Precio** Precio al detal de la carne de res **($/kg)**	**Ave** Precio al detal de la carne de aves **($/kg)**	**Ingreso** Ingreso personal disponible per cápita **($/Hab)**	**Ten** Variable de tendencia
1986	560.986	462	393	27.468	1
1987	549.194	579	464	28.740	2
1988	610.223	729	545	29.964	3
1989	672.461	910	637	33.396	4
1990	726.742	1.140	741	31.836	5
1991	651.834	1.786	1.072	31.284	6
1992	568.010	2.641	1.056	29.424	7
1993	592.160	3.063	1.225	30.288	8
1994	637.450	3.665	1.466	32.208	9
1995	696.181	4.336	2.074	32.364	10
1996	727.857	4.737	2.422	33.096	11
1997	760.590	5.175	2.791	31.920	12
1998	581.705	6.500	3.324	32.892	13

Fuente: Departamento Administrativo Nacional de Estadística – DANE
Centro de estudios gastronómicos –CEGA.

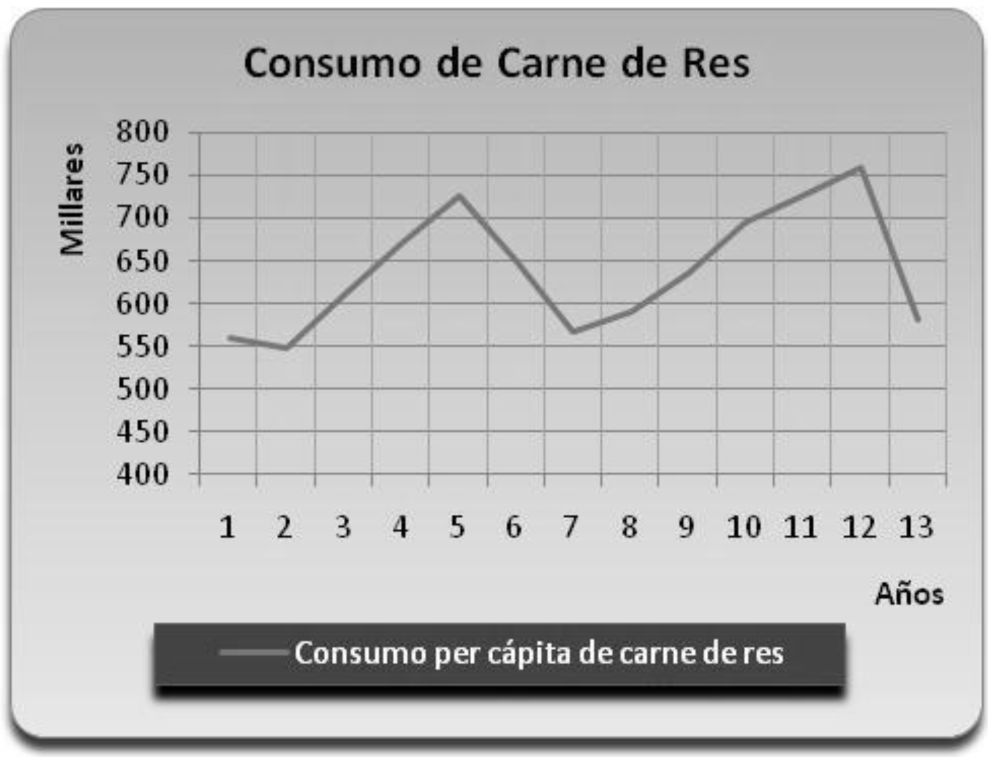
9.3. PROCEDIMIENTO PARA LA CONSTRUCCIÓN DEL MODELO

Lo primero que el estudiante debe realizar es un diagrama de dispersión, que le permite en forma visual, ver la tendencia de los valores y la relación que hay entre las variables seleccionadas. En este caso se seleccionaron las variables precio al detal de la carne de res ($/kg) y la variable de tendencia. La variable de tendencia se construyó tomando como uno el primer año, dos el segundo año y así sucesivamente.

Nota	No se analiza la variable consumo per cápita dado que existe dos ciclos en los datos, uno del año 2 al 7 y el otro del año 7 al 13, como se puede

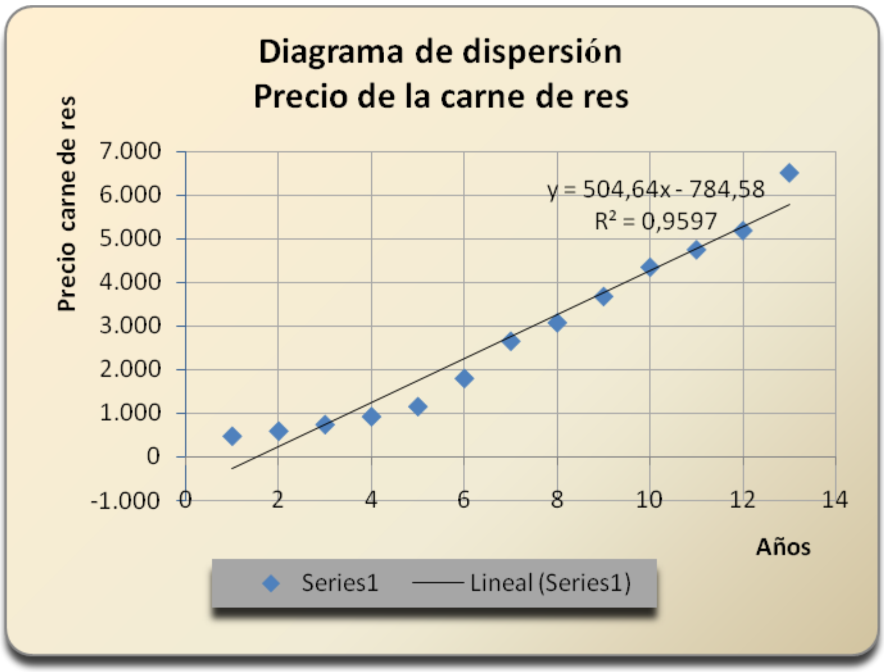
	observar en la figura 9.5, en este sentido se dividió la tabla en los dos grupos identificados y se calculó las regresiones, y mediante algunas técnicas econométricas se estableció la tendencia y pronostico del consumo de carne de res. Estos procedimientos están fuera del alcance de este texto, por tanto se explica el tema con la variable precio.

Figura 9.4. Diagrama de dispersión de la variable consumo de carne de res.



Fuente: Tabla 9.1.

Figura 9.5. Diagrama de dispersión de la variable precio al detal de la carne de res ($/Kg).



Fuente: Tabla 9.1.

Como se observa la relación entre las dos variables es de tipo lineal y creciente, es decir, que la función presenta una pendiente positiva.

Tabla 9.2. Datos estadísticos de las variables utilizadas en el modelo.

Ten (X_i)	Precio (Y_i)	X_iY_i	X^2	Y^2
1	462	462	1	213.444
2	579	1158	4	335.241
3	726	2178	9	527.076
4	910	3640	16	828.100
5	1.140	5700	25	1.299.600
6	1.786	10716	36	3.189.796
7	2.641	18487	49	6.974.881
8	3.063	24504	64	9.381.969
9	3.665	32985	81	13.432.225
10	4.336	43360	100	18.800.896

Ten (X_i)		Precio (Y_i)	X_iY_i	X^2	Y^2
11		4.737	52107	121	22.439.169
12		5.175	62100	144	26.780.625
13		6.500	84500	169	42.250.000
SUMAS	91	35.720	341.897	819	146.453.022
MEDIAS	7,00	2.747,69			

Fuente: Tabla 9.1.

9.3.1. Calculo de los paramentos $\hat{\beta}_2$ y $\hat{\beta}_1$

Se reemplaza los datos en la fórmula (9.8) y (9.9) y que ahora se reproducen a:

$$\hat{\beta}_2 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} = \frac{13(341.906) - (91)(35.720)}{13(819) - (91)^2} = 504{,}71 \qquad (9.13)$$

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} = \frac{(819)(35.720) - (91)(341.897)}{13(819) - (91)^2} = -785{,}27 \qquad (9.14)$$

Una fórmula alternativa y más sencilla es la siguiente:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} = 2.747{,}69 - (504{,}71x7) = -785{,}27$$

La ecuación encontrada es:

$$Y = \beta_1 + \beta_2 X_i$$

$$Y = -785{,}27 + 504{,}71X_i + u_i \qquad (9.15)$$

9.3.2. Cálculo del coeficiente de correlación r.

$$r = \frac{n\sum X_iY_i - \left(\sum X_i\right)\left(\sum Y_i\right)}{\sqrt{..\left[\; n\sum X_i^2 - (\sum X_i)^2 \;\right] \times \left[\; n\sum Y_i^2 - (\sum Y_i)^2 \;\right]}}$$

(9.16)

$$r = \frac{(13\times 341.897) - (91\times 35.720)}{\sqrt{\left(13\times 819 - (91)^2\right) \times \left(13\times 146.453.023 - (109.982)^2\right)}} = 0{,}979667$$

9.3.3. Cálculo del coeficiente de determinación r^2.

$$r^2 = 0{,}979667^2 = 0{,}9597 \qquad (9.17)$$

9.3.4. Interpretación del coeficiente de pendiente β_2. β_2 = 504,71, mide la pendiente de la recta, indica que dentro del período de estudio (1986-1998), anualmente se incrementó en promedio, el precio por kilogramo de la carne de res en $504,71.

9.3.5. Interpretación del coeficiente de determinación r^2. El coeficiente de determinación r^2 = 0,9597 significa que cerca del 96% de la variaciones del precio están explicadas por la variable de tendencia, puesto que r^2 puede llegar a ser máximo 1, el r^2 observado, sugiere que la línea de regresión muestral se ajusta muy bien a los datos. El coeficiente de correlación r =0,9796 indica que las dos variables, precio de la carne de res y la variable de tendencia, tienen una alta correlación positiva.

9.3.6. Cálculo de las proyecciones. Uno de los objetivos del proyecto, es estimar la proyección de los precios de la carne de res para los años 1999, 2000, 2001, 2002 y 2003, en términos de la variable tendencia corresponde a 14, 15, 16, 17 y 18. Para calcular los precios proyectados se reemplaza en la función (9.13) el valor de Xi como se muestra en la tabla 9.3.

Tabla 9.3. Proyección de los precios de la carne de res.

Años Proyectados	**Datos que toma la variable Xi**	**Estimaciones**
1999	14	$Y = -785{,}27 + 504{,}71X_i + u_i$ $Y = -785{,}27 + (504{,}71 \times 14) + u_i = 6.281$
2000	15	$Y = -785{,}27 + (504{,}71 \times 15) + u_i = 6.785$
2001	16	$Y = -785{,}27 + (504{,}71 \times 16) + u_i = 7.290$
2002	17	$Y = -785{,}27 + (504{,}71 \times 17) + u_i = 7.794$
2003	18	$Y = -785{,}27 + (504{,}71 \times 18) + u_i = 8.299$

Fuente: Tabla 9.1.

Figura 9.6. Función de regresión lineal estimada y pronosticada de la variable precio de la carne de res.

Fuente: Tablas 9.1 y 9.3

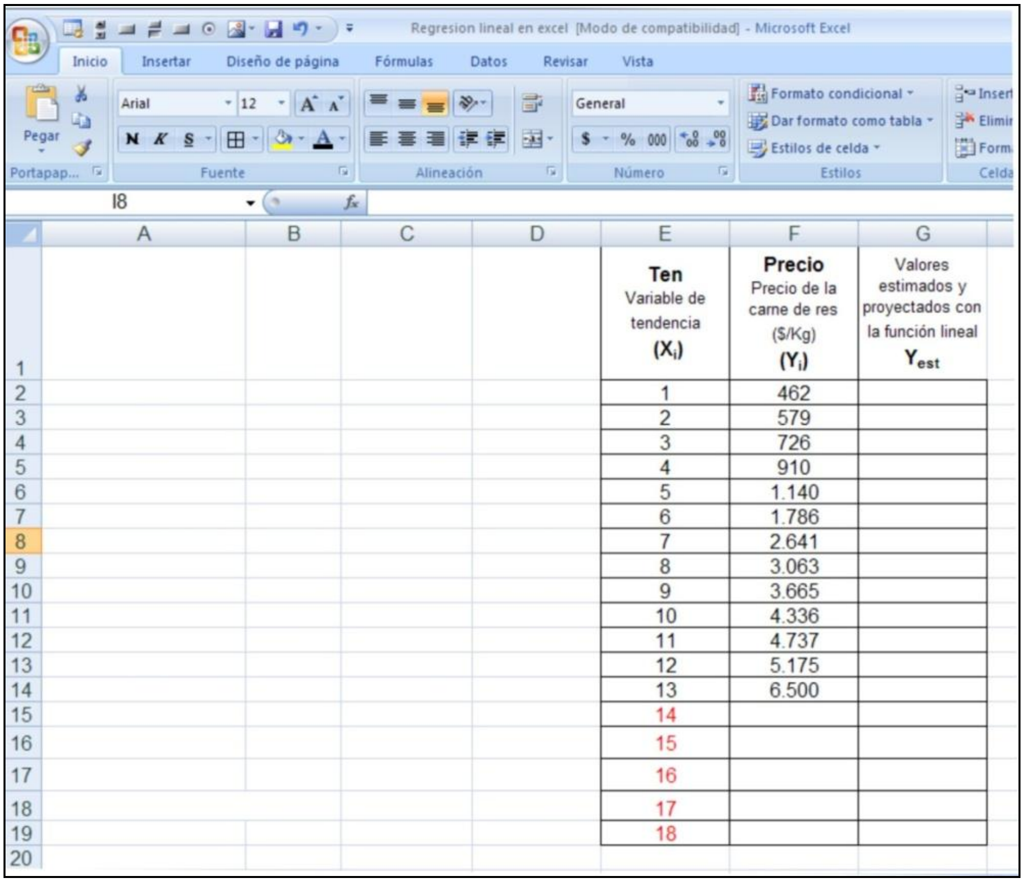
9.4. TÉCNICAS DE REGRESIÓN CON HOJA DE CALCULO EXCEL 2007

A continuación se da las instrucciones para realizar la proyección en la hoja de cálculo Excel, mediante regresión simple. Si realizó una instalación completa de Excel, las herramientas para análisis estarán disponibles cada vez que inicie el programa. Iniciamos con las funciones de la hoja de cálculo, de Office 2007.

Para trabajar en la hoja de cálculo siga los siguientes pasos:

- Digite los datos como aparecen en la Figura 9.7.

Figura 9.7. Digitación de la información en la hoja de cálculo.



	E	F	G
1	**Ten** Variable de tendencia **(X_i)**	**Precio** Precio de la carne de res ($/Kg) **($Y_i$)**	Valores estimados y proyectados con la función lineal **Y_{est}**
2	1	462	
3	2	579	
4	3	726	
5	4	910	
6	5	1.140	
7	6	1.786	
8	7	2.641	
9	8	3.063	
10	9	3.665	
11	10	4.336	
12	11	4.737	
13	12	5.175	
14	13	6.500	
15	14		
16	15		
17	16		
18	17		
19	18		
20			

Fuente: Tabla 9.1.

El estudiante se podrá dar cuenta que la versión de Excel 2007, ha estructurado mejor los menús, agregando herramientas más versátiles que permiten realizar trabajos con mejor calidad.
Utilizaremos la función ESTIMACION.LINEAL para calcular la pendiente de la ecuación y el intercepto. Esta función toma la forma

9.4.1. Calculo de la pendiente

=INDICE(ESTIMACION.LINEAL(Valores conocidos de Y; Valores conocidos de X);1)

9.4.2. Calculo del intercepto

=INDICE(ESTIMACION.LINEAL(Valores conocidos de Y; Valores conocidos de X);2)

9.4.3. Calculo del coeficiente de determinación r^2

=COEFICIENTE.R2(Valores conocidos de Y; Valores conocidos de X)

- Digite las fórmulas como se muestra en la figura 9.8.

Figura 9.8. Digitación de fórmulas de estimación lineal.

	A	B	C	D	E	F	G
1					**Ten** Variable de tendencia **(X_i)**	**Precio** Precio de la carne de res ($/Kg) **($Y_i$)**	Valores estimados y proyectados con la función lineal **Y_{est}**
2					1	462	
3					2	579	
4					3	726	
5					4	910	
6					5	1.140	
7					6	1.786	
8					7	2.641	
9					8	3.063	
10					9	3.665	
11					10	4.336	
12					11	4.737	
13					12	5.175	
14					13	6.500	
15					14		
16	Pendiente β_2=		=INDICE(ESTIMACION.LINEAL(F2:F14;E2:E14);1)		15		
17	Entercepto β_1 =		=INDICE(ESTIMACION.LINEAL(F2:F14;E2:E14);2)		16		
18	Coefiente de determinación r^2=		=COEFICIENTE.R2(F2:F14;E2:E14)		17		
19					18		

Fuente. Tabla 9.1.

- Los resultados de su computadora deben ser iguales a los de la figura 9.9.

Figura 9.9. Resultados de los estimadores de regresión lineal.

	A	B	C	D	E	F	G	H	I
1					**Ten** Variable de tendencia **(X_i)**	**Precio** Precio de la carne de res ($/Kg) **($Y_i$)**	Valores estimados y proyectados con la función lineal **Y_{est}**		
2					1	462			
3					2	579			
4					3	726			
5					4	910			
6					5	1.140			
7					6	1.786			
8					7	2.641			
9					8	3.063			
10					9	3.665			
11					10	4.336			
12					11	4.737			
13					12	5.175			
14					13	6.500			
15					14				
16	Pendiente β_2=		504,7087912		15				
17	Entercepto β_1 =		-785,269231		16				
18	Coefiente de determinación r^2=		0,959747457		17				
19					18				
20									

Hoja1 Hoja2 Hoja3 Hoja4 Hoja5

Listo 100%

Fuente: Tabla 9.1.

Realice las estimaciones y las proyecciones en la columna g, como se muestra en la figura 9.10.

Figura 9.10. Estimación y proyecciones de la función de regresión lineal.

	A	B	C	D	E	F	G	H	I
1					**Ten** Variable de tendencia **(X_i)**	**Precio** Precio de la carne de res ($/Kg) **($Y_i$)**	Valores estimados y proyectados con la función lineal **Y_{est}**		
2					1	462	=$C$17+($C$16*E2)		
3					2	579	=$C$17+($C$16*E3)		
4					3	726	=$C$17+($C$16*E4)		
5					4	910	=$C$17+($C$16*E5)		
6					5	1.140	=$C$17+($C$16*E6)		
7	Los datos a partir de la celda				6	1.786	=$C$17+($C$16*E7)		
8	E15, corresponde a los valores				7	2.641	=$C$17+($C$16*E8)		
9	proyectados con la función de				8	3.063	=$C$17+($C$16*E9)		
10	regresión lineal				9	3.665	=$C$17+($C$16*E10)		
11					10	4.336	=$C$17+($C$16*E11)		
12					11	4.737	=$C$17+($C$16*E12)		
13					12	5.175	=$C$17+($C$16*E13)		
14					13	6.500	=$C$17+($C$16*E14)		
15					14		=$C$17+($C$16*E15)		
16	Pendiente β_2=		504,7087912		15		=$C$17+($C$16*E16)		
17	Intercepto β_1 =		-785,269231		16		=$C$17+($C$16*E17)		
18	Coefiente de determinación r^2=		0,959747457		17		=$C$17+($C$16*E18)		
19					18		=$C$17+($C$16*E19)		
20									

Fuente: Figura 9.9.

Ubíquese en la celda G2 y digite la siguiente fórmula:

= $C$17+($C$16*E2)

Recuerde que se debe hacer buen uso de la potencia de la hoja de cálculo, digite una sola fórmula, las demás las puede copiar, de esta forma ahorra tiempo y agiliza el trabajo.

Con respecto al signo $ ubicado en la filas y columnas, hacen referencia a celdas absolutas, como se pude dar cuenta la fórmula es igual, la única celda relativa que experimenta cambios es la columna E con sus respectivas filas.

Los resultados de la estimación y predicción se muestran en la grafica 9.11.

Figura 9.11. Resultados de estimación y proyección de la función de regresión lineal.

	A	B	C	D	E	F	G	H	I
1					**Ten** Variable de tendencia **(X_i)**	**Precio** Precio de la carne de res ($/Kg) **($Y_i$)**	Valores estimados y proyectados con la función lineal **Y_{est}**		
2					1	462	-281		
3					2	579	224		
4					3	726	729		
5					4	910	1.234		
6					5	1.140	1.738		
7					6	1.786	2.243		
8					7	2.641	2.748		
9					8	3.063	3.252		
10					9	3.665	3.757		
11					10	4.336	4.262		
12					11	4.737	4.767		
13					12	5.175	5.271		
14					13	6.500	5.776		
15					14		6.281		
16	Pendiente β_2=		504,7087912		15		6.785		
17	Entercepto β_1 =		-785,269231		16		7.290		
18	Coefiente de determinación r^2=		0,959747457		17		7.795		
19					18		8.299		
20									

Hoja1 Hoja2 Hoja3 Hoja4 Hoja5

Listo 100%

Fuente: Figura 9.10.

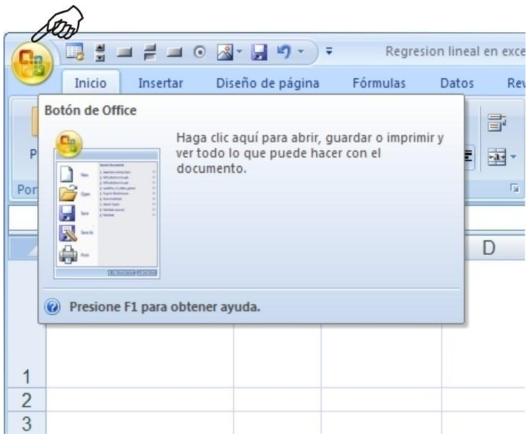
Compare los valores de la columna G con los obtenidos en la tabla 9.3.

9.5. HERRAMIENTA DE ANÁLISIS DE DATOS

La herramienta de análisis de datos, son instrucciones más potentes para el análisis de regresión. Para agregar esta herramienta a la hoja de cálculo haga los siguientes pasos:

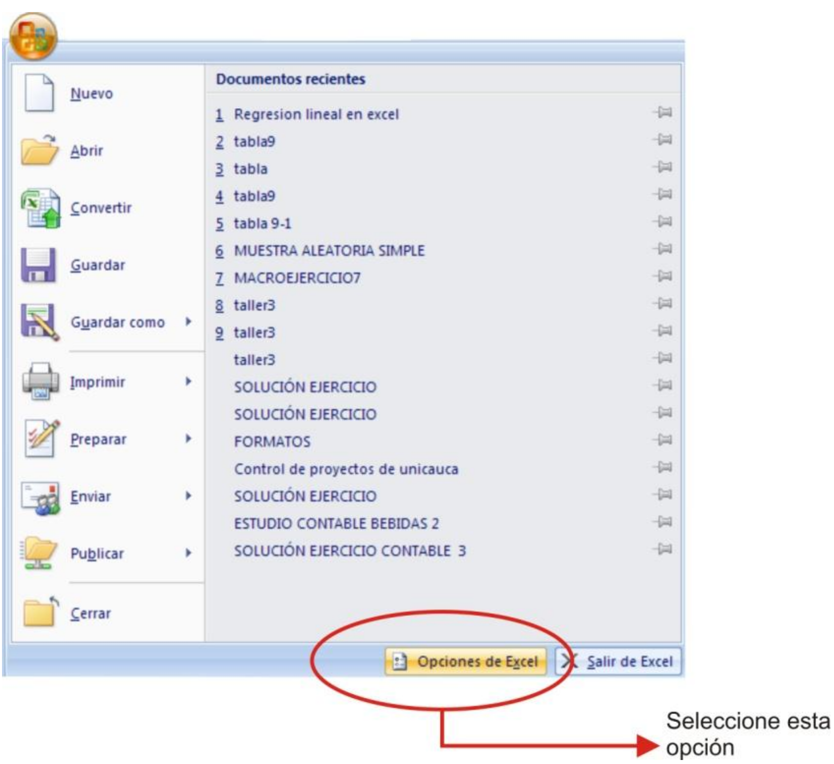
- Hacer uso del botón de office (sólo para versión de office 2007), como se observa en la grafica 9.12.

Figura 9.12. Botón de Office.



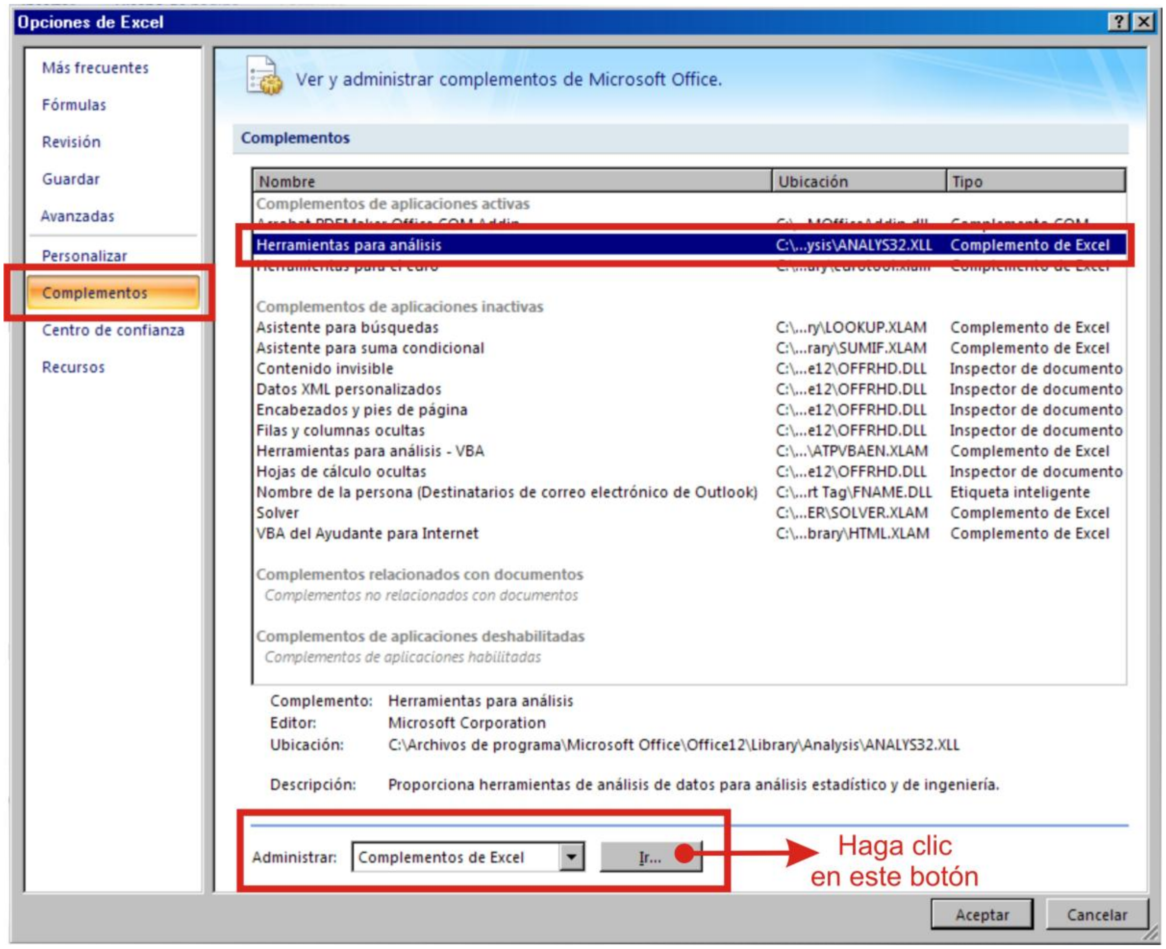
• Una vez se haga clic en el botón de Office, aparece el cuadro de diálogo (Figura 9.13).

Figura 9.13. Cuadro de dialogo del Botón de Office.



- En el cuadro de dialogo del Botón de Office seleccione **Opciones de Excel.**
- Excel le presentará un cuadro de dialogo igual al de la Figura 9.14.

Figura 9.14. Cuadro de dialogo de Opciones de Excel.



- En este cuadro de dialogo seleccione la opción de COMPLEMENTOS.
- Luego seleccione HERRAMIENTAS DE ANÁLISIS.
- Luego ubíquese en la parte inferior en la opción de Administrar – Complementos de Excel y haga clic en Ir...

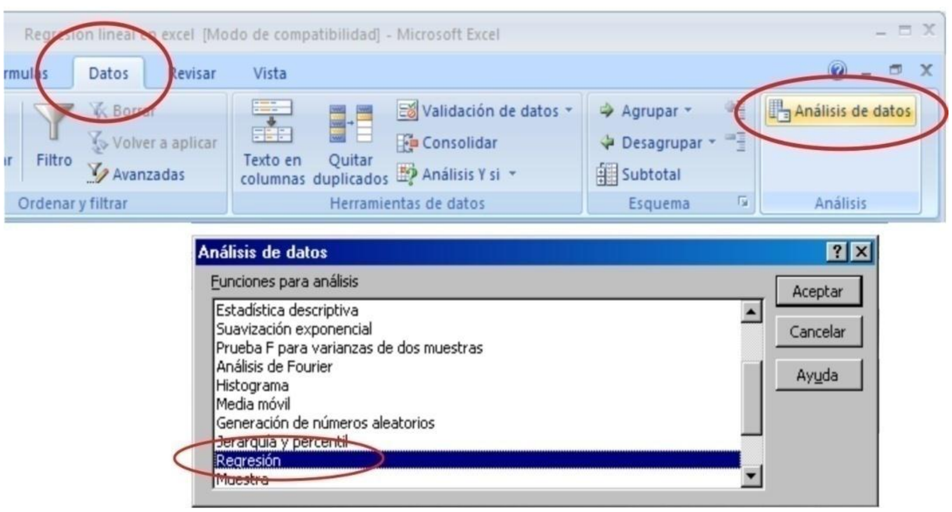
Excel le suministrara el siguiente el cuadro de dialogo de la figura 9.15.

Figura 9.15. Ventana de herramientas para análisis

Fuente: Figura 9.14.

- Haga clic en aceptar.
- Excel adiciona esta herramienta al menú del programa.
- Luego en el menú de la hoja de cálculo seleccione la opción datos, como se muestra en la figura 9.16.

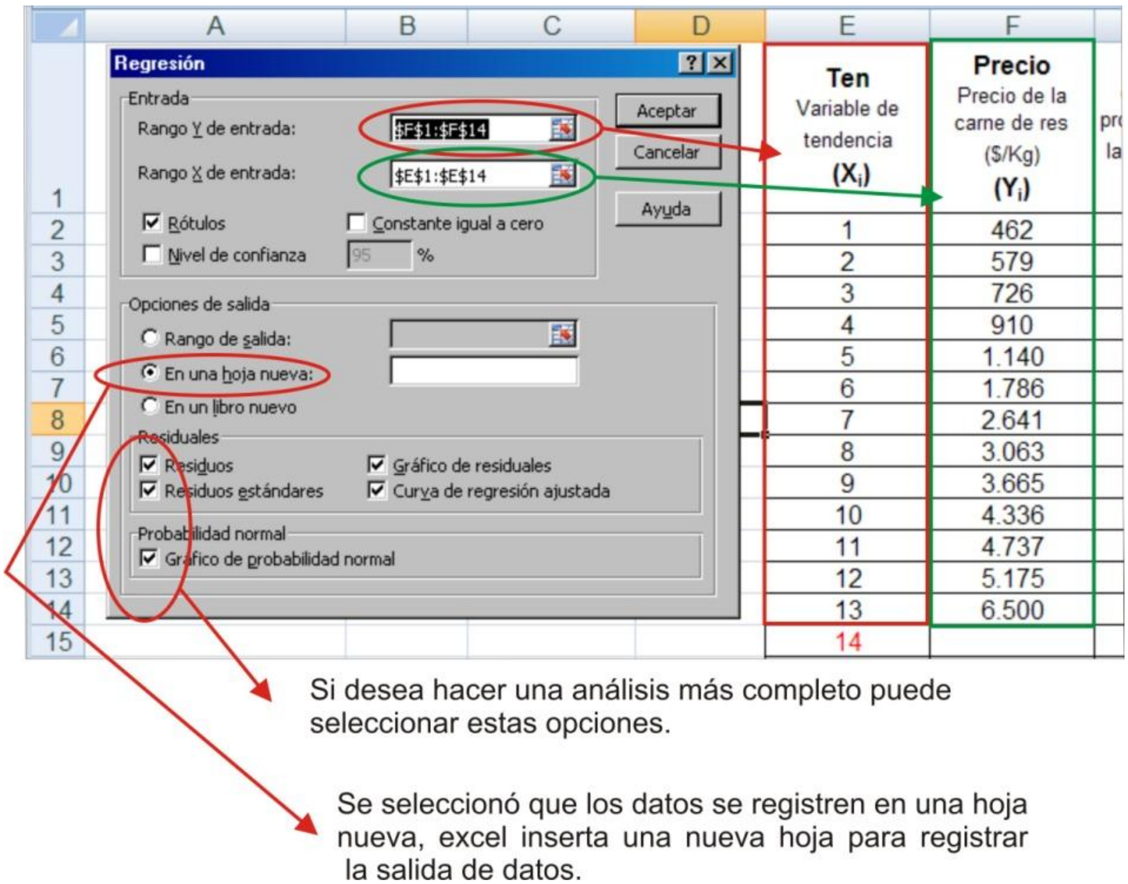
Figura 9.16. Opciones del menú DATOS.



Fuente: Figura 9.15.

- Al final del submenú de datos seleccione ANÁLISIS DE DATOS.
- En el cuadro de dialogo, busque la opción de REGRESIÓN. Haga clic en ACEPTAR.
- Marque las siguientes opciones en el cuadro de dialogo **Regresión** como aparece en la Figura 9.17.

Figura 9.17. Cuadro de dialogo de la opción REGRESIÓN

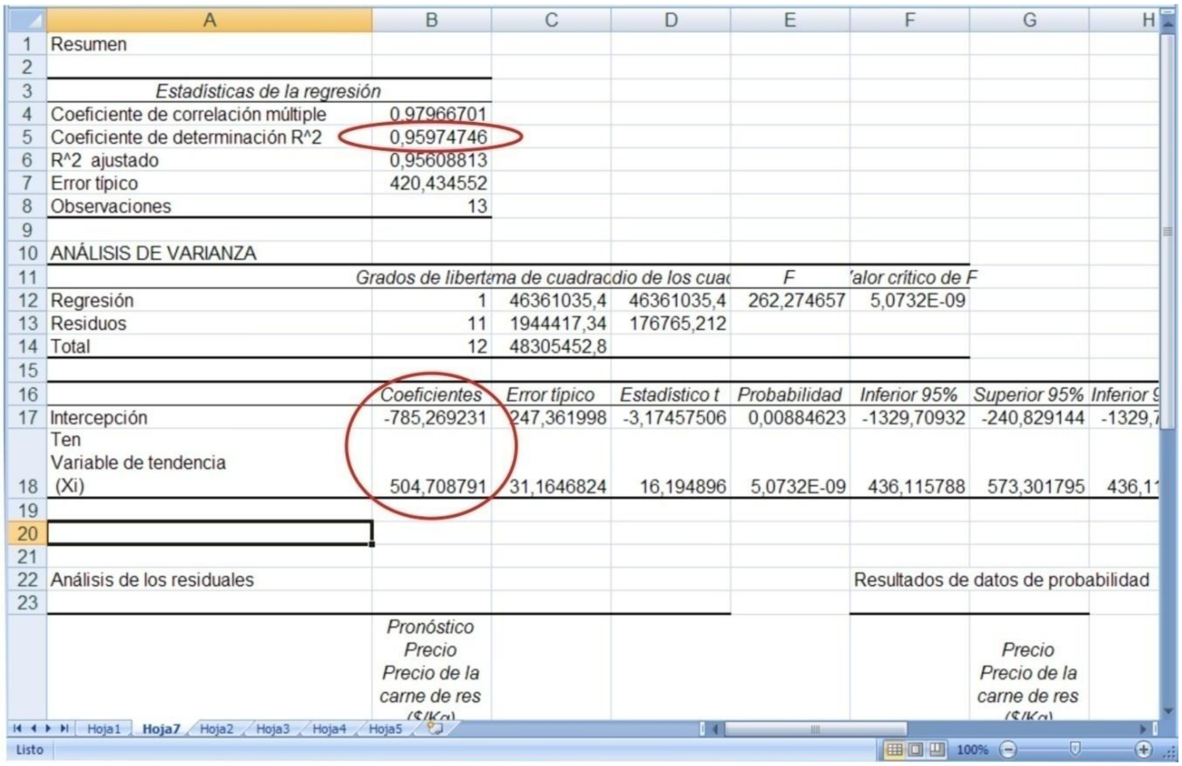


Fuente: Figura 9.16.

- Seleccionar los rangos de Y y X, tal como se muestra en la Figura 9.17.
- En la sesión de salidas, seleccione **EN UNA HOJA NUEVA**.
- En la sesión de residuales puede seleccionar todas las opciones para realizar un análisis más completo.
- En la sesión de Probabilidad normal seleccione gráfico de probabilidad normal, este le ayudará a verificar la normalidad de los residuos del modelo.

Una vez esté conforme con la información que requiere, haga clic en ACEPTAR. Los resultados se presentan en una nueva hoja como aparece en la grafica 9.18.

Figura 9.18. Salida de resultados de análisis de regresión.



	A	B	C	D	E	F	G	H
1	Resumen							
2								
3	*Estadísticas de la regresión*							
4	Coeficiente de correlación múltiple	0,97966701						
5	Coeficiente de determinación R^2	0,95974746						
6	R^2 ajustado	0,95608813						
7	Error típico	420,434552						
8	Observaciones	13						
9								
10	ANÁLISIS DE VARIANZA							
11		*Grados de libert*	*ma de cuadrac*	*dio de los cuac*	*F*	*'alor crítico de F*		
12	Regresión	1	46361035,4	46361035,4	262,274657	5,0732E-09		
13	Residuos	11	1944417,34	176765,212				
14	Total	12	48305452,8					
15								
16		*Coeficientes*	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 9*
17	Intercepción	-785,269231	247,361998	-3,17457506	0,00884623	-1329,70932	-240,829144	-1329,7
18	Ten Variable de tendencia (Xi)	504,708791	31,1646824	16,194896	5,0732E-09	436,115788	573,301795	436,1
19								
20								
21								
22	Análisis de los residuales					Resultados de datos de probabilidad		
23								
		Pronóstico Precio Precio de la carne de res ($/Kg)					*Precio Precio de la carne de res ($/Kg)*	

Fuente: Figura 9.17.

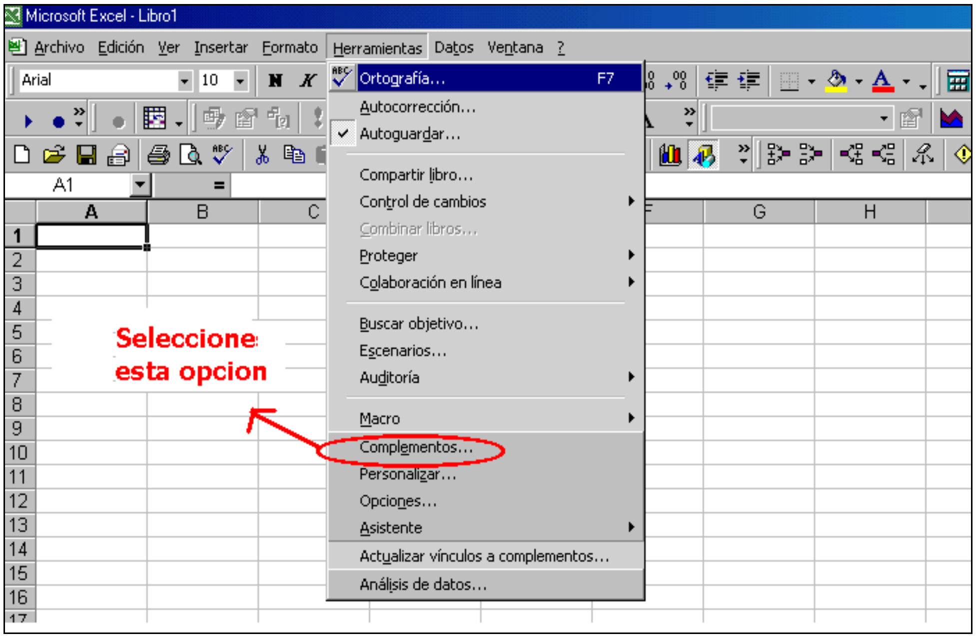
Compare estos resultados con los obtenidos con las ecuaciones 9.13, 9.14, 9.16 y 9.17.

9.6. HERRAMIENTA DE ANÁLISIS DE DATOS CON EXCEL 2003

 Herramientas	Para explicar este punto un ejemplo sencillo de 5 datos de las variables X y Y. Para trabajar en la versión de OFFICE 2003, haga los siguientes pasos:

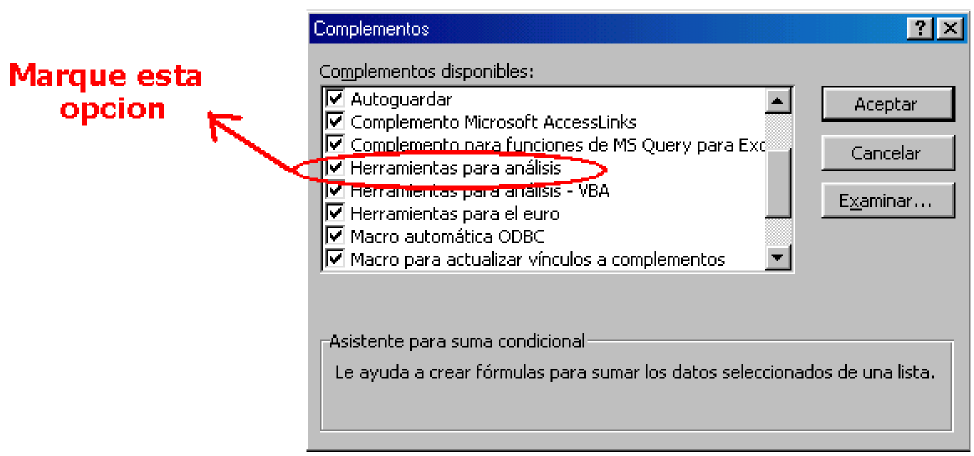
- Seleccione del menú principal la opción HERRAMIENTAS.
- En el submenú de herramientas seleccione la opción COMPLEMENTOS, como se ve en la Figura 9.19.

Figura 9.19. Opción de complementos en Excel 2003.



• En el cuadro de diálogo COMPLEMENTOS ubique HERRAMIENTA DE ANÁLISIS y active la pestaña.

Figura 9.20. Submenú de complementos.

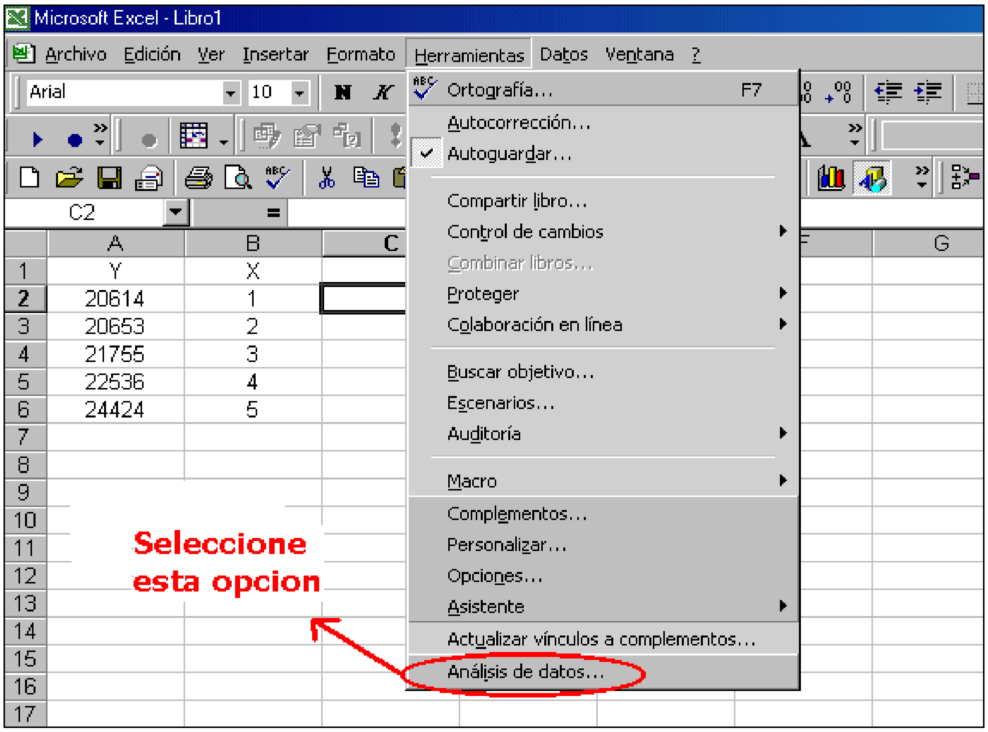


• De esta forma usted activa la opción de análisis de datos en el menú de herramientas. Como se muestra en la Figura 9.21.

• Digite los datos como aparece en la figura 9.21.

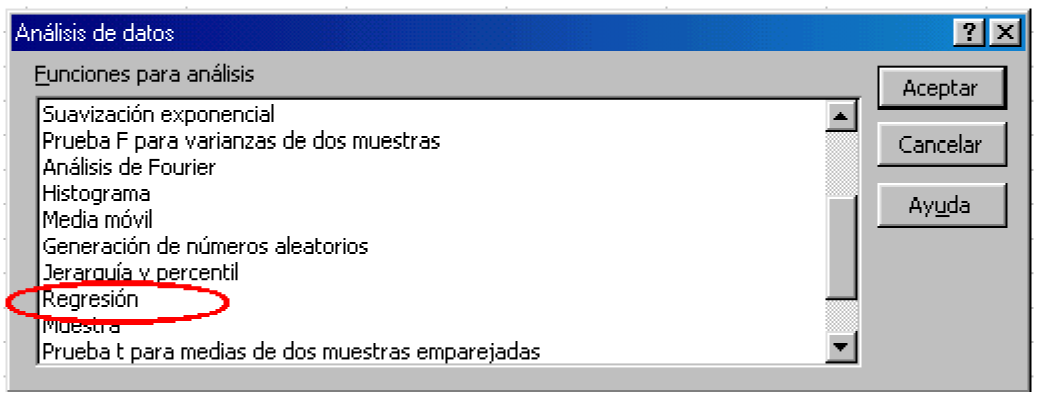
- Marque la opción ANÁLISIS DE DATOS del menú herramientas (Figura 9.21).

Figura 9.21. Opción de Análisis de Datos del menú herramientas.



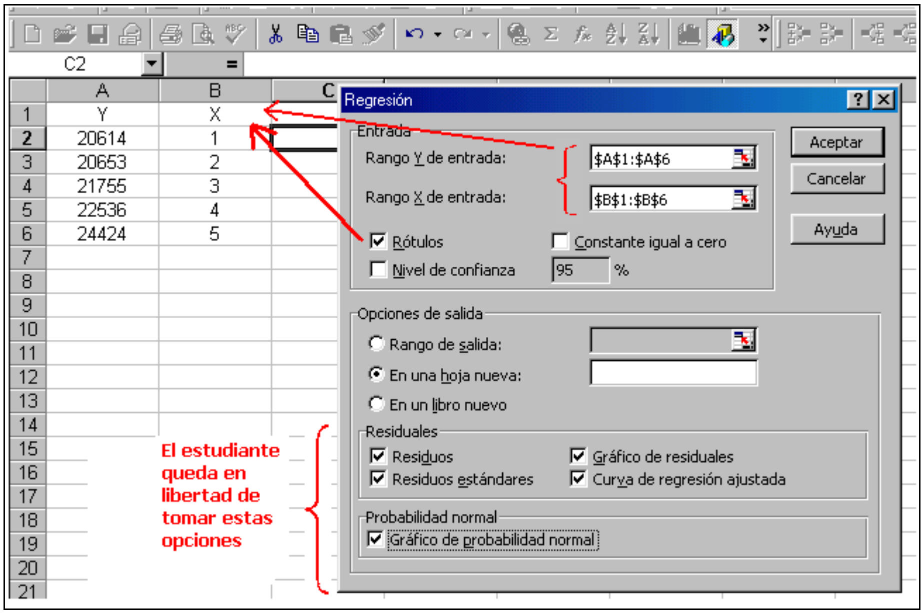
- Seleccione REGRESIÓN del submenú **Análisis de datos** y haga clic en aceptar.

Figura 9.22. Selección de la opción Regresión del submenú de análisis de datos.



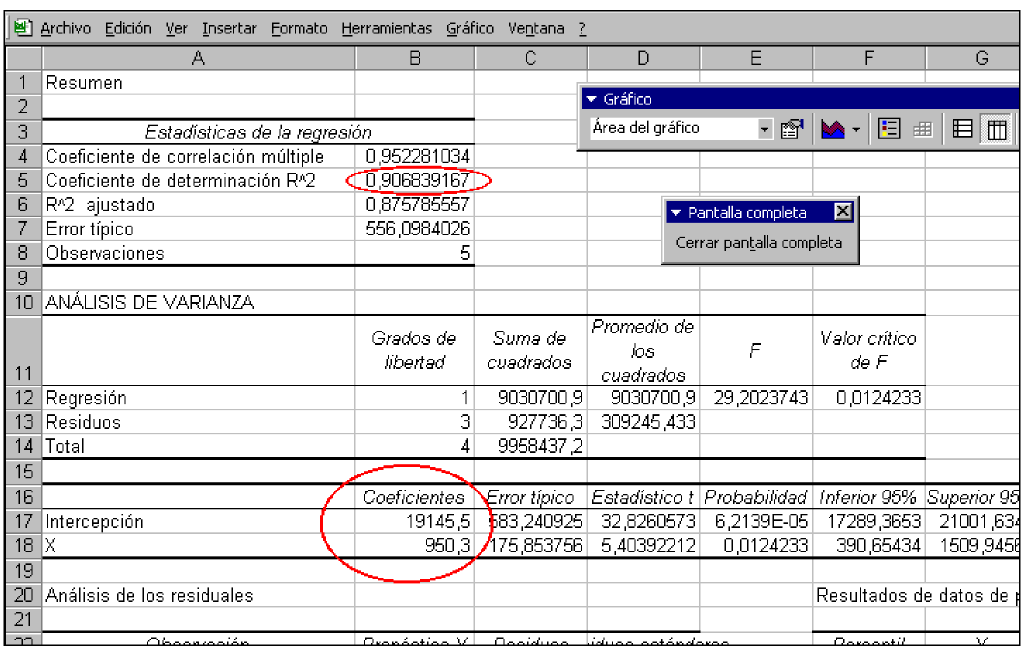
- Complete el cuadro de diálogo como se muestra en la Figura 9.23.
- Haga clic en ACEPTAR.

Figura 9.23. Cuadro de diálogo de Regresión.



• Excel suministra la información que se muestra en las Figuras 9.24 y 9.25.

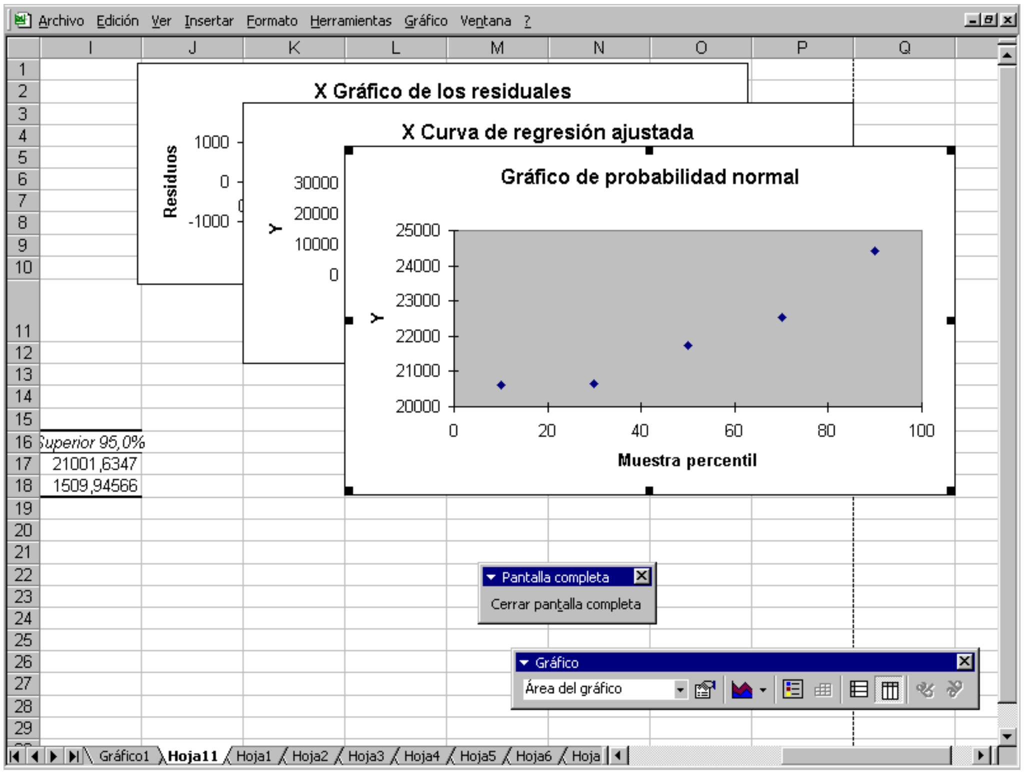
Figura 9.24. Salida de resultados de regresión.



	A	B	C	D	E	F	G
1	Resumen						
2							
3	*Estadísticas de la regresión*						
4	Coeficiente de correlación múltiple	0,952281034					
5	Coeficiente de determinación R^2	0,906839167					
6	R^2 ajustado	0,875785557					
7	Error típico	556,0984026					
8	Observaciones	5					
9							
10	ANÁLISIS DE VARIANZA						
11		*Grados de libertad*	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*	
12	Regresión	1	9030700,9	9030700,9	29,2023743	0,0124233	
13	Residuos	3	927736,3	309245,433			
14	Total	4	9958437,2				
15							
16		*Coeficientes*	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95*
17	Intercepción	19145,5	583,240925	32,8260573	6,2139E-05	17289,3653	21001,634
18	X	950,3	175,853756	5,40392212	0,0124233	390,65434	1509,9456
19							
20	Análisis de los residuales					Resultados de datos de p	
21							

.

Figura 9.25. Salidas de gráficas de análisis de regresión.



Excel también ofrece las funciones estadísticas avanzadas de TENDENCIAS y PRONÓSTICO, que operan sobre matrices de números. A continuación se explica estas funciones con el ejemplo de consumo de carne de res.

9.7. FUNCIONES TENDENCIA Y PRONÓSTICO.

9.6.1. Función tendencia. Devuelve valores que resultan de una tendencia lineal. Ajusta una recta (calculada con el método de mínimos cuadrados) a los valores de las matrices definidas por los argumentos conocido_ y y conocido_ x.

Para hacer uso de esta función haga los siguientes pasos:

Digite la información, es decir, los valores conocidos de X_i y Y_i, como se aprecia en la Figura 9.26.

Figura 9.26. Funciones tendencia y pronóstico.

	E	F	G	H	I
1	**Ten** Variable de tendencia **(X_i)**	**Precio** Precio de la carne de res ($/Kg) **($Y_i$)**	Valores estimados y proyectados con la función lineal **Y_{est}**	Valores estimados y proyectados con la función lineal **Utilizando la función tendencia** **Y_{est}**	Valores estimados y proyectados con la función lineal **Utilizando la función pronostico** **Y_{est}**
2	1	462	=$C$17+($C$16*E2)	=TENDENCIA(F2:F14;E2:E14)	
3	2	579	=$C$17+($C$16*E3)	=TENDENCIA(F2:F14;E2:E14)	
4	3	726	=$C$17+($C$16*E4)	=TENDENCIA(F2:F14;E2:E14)	
5	4	910	=$C$17+($C$16*E5)	=TENDENCIA(F2:F14;E2:E14)	
6	5	1140	=$C$17+($C$16*E6)	=TENDENCIA(F2:F14;E2:E14)	
7	6	1786	=$C$17+($C$16*E7)	=TENDENCIA(F2:F14;E2:E14)	
8	7	2641	=$C$17+($C$16*E8)	=TENDENCIA(F2:F14;E2:E14)	
9	8	3063	=$C$17+($C$16*E9)	=TENDENCIA(F2:F14;E2:E14)	
10	9	3665	=$C$17+($C$16*E10)	=TENDENCIA(F2:F14;E2:E14)	
11	10	4336	=$C$17+($C$16*E11)	=TENDENCIA(F2:F14;E2:E14)	
12	11	4737	=$C$17+($C$16*E12)	=TENDENCIA(F2:F14;E2:E14)	
13	12	5175	=$C$17+($C$16*E13)	=TENDENCIA(F2:F14;E2:E14)	
14	13	6500	=$C$17+($C$16*E14)	=TENDENCIA(F2:F14;E2:E14)	
15	14		=$C$17+($C$16*E15)	=TENDENCIA(F2:F14;E2:E14;E15:E19)	=PRONOSTICO(E15:E19;F2:F14;E2:E14)
16	15		=$C$17+($C$16*E16)	=TENDENCIA(F2:F14;E2:E14;E15:E19)	=PRONOSTICO(E15:E19;F2:F14;E2:E14)
17	16		=$C$17+($C$16*E17)	=TENDENCIA(F2:F14;E2:E14;E15:E19)	=PRONOSTICO(E15:E19;F2:F14;E2:E14)
18	17		=$C$17+($C$16*E18)	=TENDENCIA(F2:F14;E2:E14;E15:E19)	=PRONOSTICO(E15:E19;F2:F14;E2:E14)
19	18		=$C$17+($C$16*E19)	=TENDENCIA(F2:F14;E2:E14;E15:E19)	=PRONOSTICO(E15:E19;F2:F14;E2:E14)
20					

- Ubíquese en la celda H2, seleccione el rango de H2 hasta H14.

- Escriba la formula = TENDENCIA(F2:F14;E2:E14).
- Presione conjuntamente CTRL+MAY+ENTER, dado que el formato es matricial.
- Ubíquese en la celda H15, seleccione el rango H15 hasta H19.
- Escriba la formula = TENDENCIA(F2:F14;E2:E14;E15:E19).
- Presione conjuntamente CTRL+MAY+ENTER.

Aquí es importante detenernos para ampliar el concepto de formulas matriciales que maneja EXCEL.

Una fórmula matricial opera sobre dos o más conjuntos de valores, denominados argumentos matriciales, para devolver un resultado simple o un resultado múltiple.

Reglas de las fórmulas matriciales.

- Para introducir fórmulas matriciales, en primer lugar seleccionamos la celda o rango de que va a contener el resultado. Si la fórmula devuelve resultados múltiples, tenemos que seleccionar un rango con el mismo tamaño y forma, que el rango o rangos que emplearemos para realizar los cálculos.

• Par fijar una fórmula matricial, pulsamos conjuntamente Control+Mayús+Enter. Excel coloca un par de llaves alrededor de la fórmula, para indicar que es una fórmula matricial. No debemos introducir las llaves; si se hace, Excel interpreta que se ha introducido un rótulo.

• Para editar, borrar o desplazar celdas individuales en un rango matricial, se debe tratar las celdas de una matriz como una única entidad y editarlas todas a la vez. En un rango matricial, no se puede editar, borrar o desplazar celdas individuales, ni tampoco insertar o eliminar celdas.

• Para editar o borrar una matriz, seleccionamos la matriz completa y activamos la barra de fórmulas (las llaves que engloban a una fórmula desaparecen). Editamos o borramos la fórmula y luego pulsamos Control+Mayús+Enter.

9.6.2. Función pronóstico. Calcula o pronostica un valor futuro a través de los valores existentes. La predicción del valor puede ser una matriz de valores o un valor **Y** teniendo en cuenta un valor **X**. Los valores conocidos son valores **X** y valores **Y** existentes, y el nuevo o nuevos valores se pronostica utilizando regresión lineal. Esta función se puede utilizar para realizar previsiones de ventas, establecer requisitos de inventario o tendencias de los consumidores.

Para hacer uso de esta función haga los siguientes pasos:

• Ubíquese en la celda I14.
• Seleccione las celdas I14 hasta I19.
• Escriba la siguiente formula =PRONOSTICO(E15:E19;F2:F14;E2:E14).
• Presione conjuntamente CTRL+MAY+ENTER.

También se puede utilizar ésta función como una fórmula para calcular un pronóstico de la variable Y, dado un valor de X.

Por ejemplo. Se quiere saber cuál será el precio de la carne de res para el año 2000, es decir, para el valor de X = 15. Entonces se debe digitar en la celda que desee que aparezca el cálculo la siguiente fórmula = PRONOSTICO(15;F2:F14;E2:E14), El valor devuelto es $6.785.

Para verificar los valores y comparar los datos de la columna H con los datos de la columna I, de la Figura 9.27.

Figura 9.27. Resultados de las funciones Tendencia y pronóstico.

	E	F	G	H	I
1	**Ten** Variable de tendencia **(X_i)**	**Precio** Precio de la carne de res ($/Kg) **($Y_i$)**	Valores estimados y proyectados con la función lineal **Y_{est}**	Valores estimados y proyectados con la función lineal **Utilizando la función tendencia** Y_{est}	Valores estimados y proyectados con la función lineal **Utilizando la función pronostico** **Y_{est}**
2	1	462	-281	-281	
3	2	579	224	224	
4	3	726	729	729	
5	4	910	1.234	1.234	
6	5	1.140	1.738	1.738	
7	6	1.786	2.243	2.243	
8	7	2.641	2.748	2.748	
9	8	3.063	3.252	3.252	
10	9	3.665	3.757	3.757	
11	10	4.336	4.262	4.262	
12	11	4.737	4.767	4.767	
13	12	5.175	5.271	5.271	
14	13	6.500	5.776	5.776	
15	14		6.281	6.281	6.281
16	15		6.785	6.785	6.785
17	16		7.290	7.290	7.290
18	17		7.795	7.795	7.795
19	18		8.299	8.299	8.299
20					

PREGUNTAS DE ANÁLISIS

1. Defina:

- Pronóstico.
- Función de regresión lineal.

2. ¿Cuáles son las limitaciones de información que se presentan en la investigación?
3. ¿En qué consiste la predicción en base a la relación entre variables?
4. ¿En qué consiste la predicción en base a la información subjetiva?
5. ¿En qué consiste la predicción con base a series de tiempo?
6. ¿Cómo se define la función de regresión lineal?
7. ¿Explique brevemente el método de mínimos cuadrados ordinarios (MCO)?
8. ¿Qué explica el coeficiente de determinación r^2?

9. Con los datos de la tabla 9.1 proyecte mediante el método de mínimos cuadrados los precios al detal de la carne de aves para los años 1999 a 2003. Utilice la siguiente función:

$$Ave = \beta_1 + \beta_2 T_i + u_i$$

Es decir, que:

Variable dependiente = Ave (precio de la carne de aves)
Variable independiente =T_i (variable de tendencia)

a) Interprete los parámetros β_2 y r^2.

10. Con los datos de la tabla 9.1, corra el modelo de demanda especificado a continuación:

$$\mathrm{Re}\, s_i = \beta_1 + \beta_2 \,\mathrm{Pr}\, ecio_i + u_i$$

Es decir que:

Variable dependiente = Res (Consumo per-cápita de carne de res (Tn/año))
Variable independiente = Precio (precio al detal de la carne de res ($/kg)

a) ¿Qué signo se espera del parámetro β_2 ? β_2
b) Interprete los parámetros β_2 y r^2.