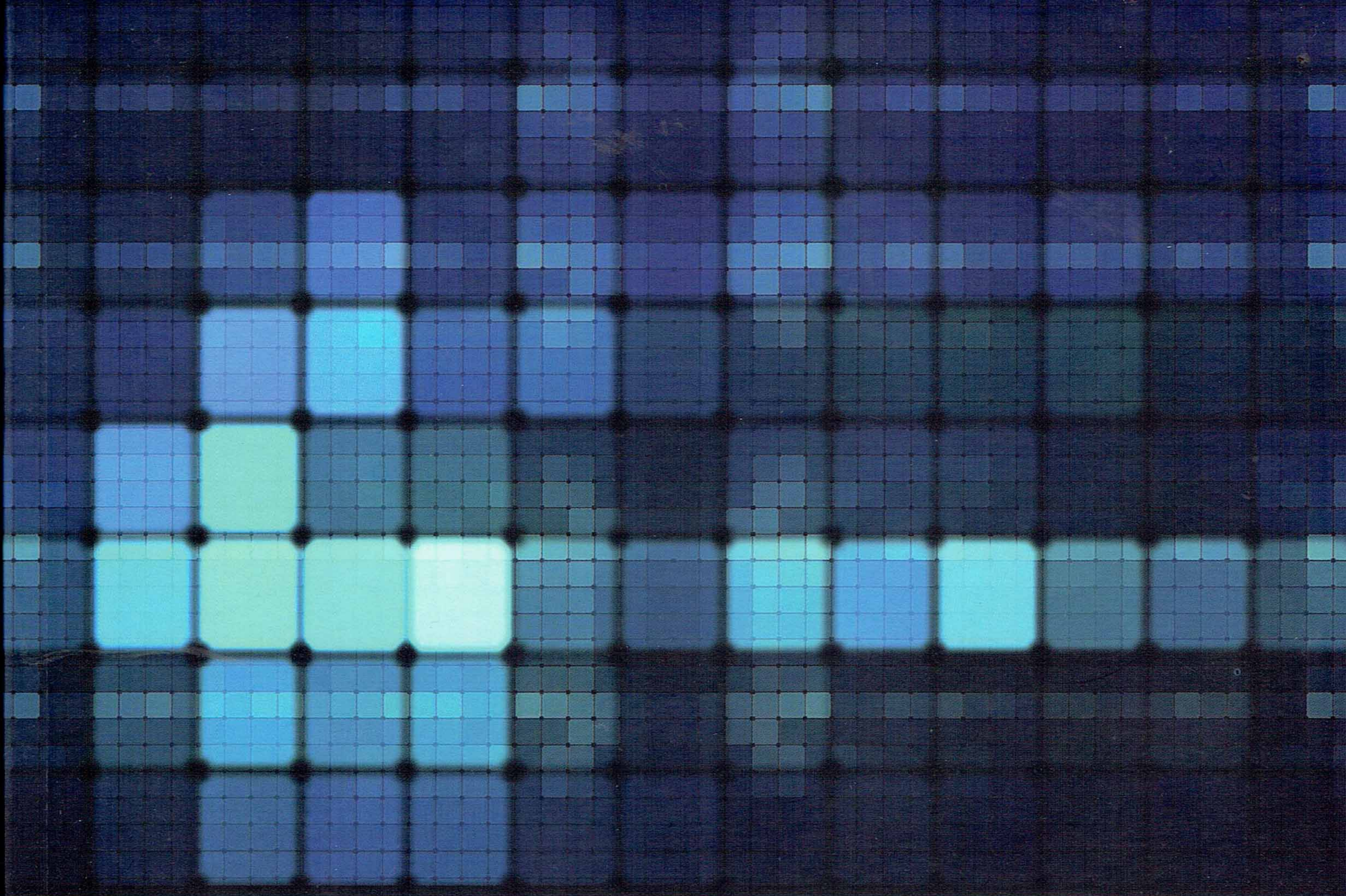


Estadística

Con aplicaciones a las Ciencias Políticas

Francisco Sánchez Espinoza

Estadística

Con aplicaciones a las Ciencias Políticas

Francisco Sánchez Espinoza

PISO 15 Editores

Primera edición, agosto 2015

14 Oriente 2827
Puebla, Puebla.

ISBN: 978-607-96963-3-7

Corrección: El Errante editor

Impreso y hecho en México / *Printed and made in Mexico.*

Estadística

Con aplicaciones a las Ciencias Políticas

Francisco Sánchez Espinoza

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA · 1578 ·

BUAP

Benemérita Universidad Autónoma de Puebla

J. Alfonso Esparza Ortiz
Rector

René Valdiviezo Sandoval
Secretario General

Ygnacio Martínez Laguna
Vicerrector de Investigación y Estudios de Posgrado

Facultad de Derecho y Ciencias Sociales, BUAP

Carlos Antonio Moreno Sánchez
Director de la Facultad de Derecho y Ciencias Sociales

Georgina Tenorio Martínez
Secretaria Académica

Omar Eduardo Mayorga Gallardo
Coordinador de Publicaciones y Comunicación Institucional

Primera edición, julio 2015

ISBN:

Corrección: El Errante editor

Impreso y hecho en México / *Printed and made in Mexico.*

ÍNDICE

1. ESTADÍSTICA DESCRIPTIVA

1.1 Frecuencias y gráficos

Para abordar los distintos temas de estadística descriptiva es necesario establecer las partes que contiene una tabla de frecuencias:

Cuadro 1

Tabla de frecuencias

Intervalo de clase	frecuencia fi	frecuencias acumuladas
18-25	20	20
26-40	29	49
41-60	35	84
61-…	16	100
SUMA	100	

FUENTE: Elaboración del autor.

Límite inferior verdadero de la marca de clase. Es 0.5 unidades atrás del límite inferior de la marca de clase.

Límite superior verdadero de la marca de clase. Es 0.5 unidades adelante del límite superior de la marca de clase.

Amplitud de la marca de clase. Es la sustracción del límite inferior de clase a partir del límite superior de clase.

Amplitud verdadera de la marca de clase. Es la sustracción del límite inferior verdadero de clase a partir del límite superior verdadero de clase.

Las frecuencias acumuladas se obtienen de ir adicionando cada una de las frecuencias de la marca de clase.

1.2 Medidas de tendencia central

Una medida de tendencia central es, generalmente, un único número que indica el centro de una serie de números a partir de los cuales se calcula.

1.2.1 Media en datos desagregados

En el cuadro 2 los datos son las edades de ciudadanos que eran potenciales votantes en una elección.

A partir de ellos calcule la media o promedio.

Cuadro 2

Edades de ciudadanos encuestados, por número de folio del cuestionario

Número de folio	Edad del votante (años)	Número de folio	Edad del votante (años)	Número de folio	Edad del votante (años)
1	18	36	25	71	60
2	28	37	24	72	45
3	46	38	57	73	49
4	39	39	24	74	30
5	33	40	59	75	50
6	34	41	22	76	57
7	25	42	50	77	79
8	55	43	39	78	52
9	27	44	40	79	82
10	29	45	41	80	29
11	57	46	20	81	64
12	26	47	19	82	63
13	23	48	43	83	53
14	30	49	44	84	69
15	19	50	47	85	37
16	19	51	63	86	74
17	52	52	47	87	71
18	43	53	58	88	55
19	28	54	18	89	70
20	41	55	28	90	52
21	42	56	56	91	66
22	67	57	27	92	64
23	26	58	42	93	19
24	22	59	26	94	23
25	21	60	33	95	28
26	46	61	24	96	62
27	44	62	45	97	30
28	27	63	35	98	43
29	21	64	32	99	29
30	41	65	44	100	19
31	66	66	72		
32	35	67	49		
33	24	68	70		
34	34	69	30		
35	60	70	29		

FUENTE: Elaboración del autor.

La fórmula para el cálculo es la siguiente:

$$Media = \frac{\sum_{i=1}^{n} x_i}{n}$$

Como se observa, la media de las edades de los votantes es de 41.54 años, es decir, un poco más de 41 años con 6 meses.

La desventaja de la media como medida de tendencia central es que puede ser influida fuertemente por un solo valor extremo.

1.2.2 Media en datos agrupados

Cuando los datos se encuentran agrupados, la media se calcula de manera diferente.[1]

Cuando se agrupa un conjunto de valores en intervalos de clase, las observaciones individuales pierden su identidad, por lo que, al calcular la media, se supone que los valores que se incluyen dentro de un intervalo de clase determinado son iguales al punto medio de ese intervalo. Puede obtenerse el punto medio de un intervalo calculando la media de los límites de clase.

Para obtener la media se suman los productos que se obtienen multiplicando la frecuencia de cada intervalo por su punto medio, y dividiendo luego este total por la suma de las frecuencias.

Si m_i es el punto medio del intervalo de clase i, entonces la fórmula de cálculo de la media es la siguiente:

$$Media = \frac{\sum_{i=1}^{k} m_i f_i}{n}$$

Donde "k" es el número de intervalos de clase y "n" es el total de las frecuencias.

En el cuadro 3 se observan los datos considerados.

Cuadro 3

Edades de ciudadanos encuestados

Intervalo de clase	frecuencia fi	punto medio mi	mifi
18-25	20	21.5	430
26-40	29	33	957
41-60	35	50.5	1767.5
61-82	16	71.5	1144
	100		4298.5

FUENTE: Elaboración del autor.

[1] Puede consultarse Wayne, Daniel, *Estadística con aplicaciones a las ciencias sociales y a la educación*. McGraw-Hill, México, 1988.

Obtenga la media:

$$Media = \frac{\sum_{i=1}^{k} m_i f_i}{n} = \frac{4298.5}{100} = 42.98$$

Entonces, la media es de 42.98 años, es decir, casi 43 años.

1.2.3. Media ponderada

La media aritmética supone que cada observación tiene la misma importancia, pero existen otras formas para calcular la media.

En un programa de estudios las materias tienen diferente número de créditos (peso específico), el cálculo del promedio aritmético supone a cada materia con el mismo peso específico, es decir, se suman todas las calificaciones y se dividen entre el total de materias; en cambio, el promedio ponderado requiere un cálculo más laborioso, siendo su resultado un indicador que toma en cuenta el peso específico de cada materia.

Podemos obtener el promedio ponderado de la estudiante Rosaura Aguirre, considerando los siguientes datos:

Cuadro 4

Calificaciones por materia de una estudiante

Nombre de la Materia	(P) Créditos	(X) Calificación	X* P
Matemáticas financieras	6	10	60
Contabilidad I	10	6	60
Contabilidad II	10	10	100
Contabilidad III	7	6	42
Contabilidad II	15	9	135
Estadística	6	10	60
Microeconomía	6	10	60
Derecho mercantil II	4	7	28
Computación	5	7	35
Laboratorio I	3	9	27
	72	84	607

FUENTE: Elaboración del autor.

Promedio aritmético:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{84}{10} = 8.4$$

Promedio ponderado:

$$\overline{X}_p = \frac{\sum_{i=1}^{n} x_i P_i}{\sum_{i=1}^{n} P}$$

Donde, P son las ponderaciones.

Los cálculos respectivos se presentan en el cuadro 5.

Cuadro 5

Calificaciones por materia (Cálculos para media ponderada)

Nombre de la Materia	(P) Créditos	(X) Calificación	X* P
Matemáticas financieras	6	10	60
Contabilidad I	10	6	60
Contabilidad II	10	10	100
Contabilidad III	7	6	42
Contabilidad II	15	9	135
Estadística	6	10	60
Microeconomía	6	10	60
Derecho mercantil II	4	7	28
Computación	5	7	35
Laboratorio I	3	9	27
	72	84	607

FUENTE: Elaboración del autor.

El cálculo con la fórmula es el siguiente:

$$\overline{Xp} = \frac{\sum_{i=1}^{n} x_i P_i}{\sum_{i=1}^{n} P} = \frac{607}{72} = 8.43$$

1.2.4 Mediana en datos desagregados

La mediana es aquel valor que se encuentra en la mitad de una serie de datos, cuando están ordenados por su magnitud.

Si el número de valores es impar, la mediana es igual al valor de la mitad; si el número de valores es par, la mediana es igual a la media de los valores que quedan en la mitad.

Cuadro 6

Edades de ciudadanos encuestados, ordenados de menor a mayor

Número de folio	Edad del votante (años)	Número de folio (modif.)	Número de folio	Edad del votante (años)	Número de folio (modif.)	Número de folio	Edad del votante (años)	Número de folio (modif.)
1	18	1	69	30	36	17	52	71
54	18	2	74	30	37	78	52	72
15	19	3	97	30	38	90	52	73
16	19	4	64	32	39	83	53	74
47	19	5	5	33	40	8	55	75
93	19	6	60	33	41	88	55	76
100	19	7	6	34	42	56	56	77
46	20	8	34	34	43	11	57	78
25	21	9	32	35	44	38	57	79
29	21	10	63	35	45	76	57	80
24	22	11	85	37	46	53	58	81
41	22	12	4	39	47	40	59	82
13	23	13	43	39	48	35	60	83
94	23	14	44	40	49	71	60	84
33	24	15	20	41	50	96	62	85
37	24	16	30	41	51	51	63	86
39	24	17	45	41	52	82	63	87
61	24	18	21	42	53	81	64	88
7	25	19	58	42	54	92	64	89
36	25	20	18	43	55	31	66	90
12	26	21	48	43	56	91	66	91
23	26	22	98	43	57	22	67	92
59	26	23	27	44	58	84	69	93
9	27	24	49	44	59	68	70	94
28	27	25	65	44	60	89	70	95
57	27	26	62	45	61	87	71	96
2	28	27	72	45	62	66	72	97
19	28	28	3	46	63	86	74	98
55	28	29	26	46	64	77	79	99
95	28	30	50	47	65	79	82	100
10	29	31	52	47	66			
70	29	32	67	49	67			
80	29	33	73	49	68			
99	29	34	42	50	69			
14	30	35	75	50	70			

FUENTE: Elaboración del autor.

Como se observa, el lugar número 50 lo ocupa una persona con el folio 20 y una edad de 41 años; el folio número 51 lo ocupa una persona con el folio 30 y una edad de 41 años. Como el número de datos es par es necesario obtener la media de lo datos 50 y 51.

$$Media = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{82}{2} = 41$$

La mediana no se ve tan afectada por los valores extremos de una serie de datos.

1.2.5 Mediana en datos agrupados

Así como la mediana para datos desagregados es el valor que se encuentra en la mitad de una serie ordenada de valores, o bien, es el punto del eje horizontal del histograma correspondiente en el cual, si se traza una línea vertical, el área comprendida bajo el histograma queda dividida en dos partes iguales, también puede obtenerse una medida para datos agregados.

De acuerdo con las anteriores definiciones, pero adaptándolas a datos agrupados, se diría que la mediana es el valor asociado al n/2-avo punto, cuando los valores han sido organizados en orden de magnitud.

Se encuentra ese valor empleando la distribución de frecuencia acumulada y suponiendo que los valores que están incluidos en cada una de las marcas de clase se distribuyen uniformemente en ellas.

La fórmula de cálculo es la siguiente:

$$Mediana = L + \frac{j}{f} w$$

Donde:

L es el límite inferior verdadero de la marca de clase que contiene a la mediana.

j es el número de observaciones que se necesitan para llegar hasta la mediana, después de que se ha llegado a la marca de clase que contiene a la mediana.

f número de observaciones que están incluidas en la marca de clase que contiene a la mediana.

w amplitud verdadera de la marca de clase que contiene a la mediana.

Cuadro 7

Edades de encuestados por marcas de clase

Intervalo de clase	frecuencia fi	frecuencias acumuladas
18-25	20	20
26-40	29	49
41-60	35	84
61-82	16	100
SUMA	100	

FUENTE: Elaboración del autor.

El valor asociado al n/2-avo es el siguiente:

$$Mediana = n/2 - avo = 100/2 = 50$$

Como puede verse, hasta la segunda marca de clase se encuentran 49 observaciones incluidas, por lo que se necesita una observación más, de las incluidas en la tercera marca de clase para llegar al n/2-avo punto. La mediana entonces está en la tercera marca de clase. Suponiendo que las 35 observaciones de la tercera marca de clase se encuentran uniformemente distribuidas en ella, se tiene que avanzar 1/35 del trayecto de la marca de clase para llegar hasta la observación 50. Como el límite inferior verdadero de clase de la marca de clase que contiene a la mediana es de 40.5:

$$Mediana = L + \frac{j}{f} w = 40.5 + \frac{1}{35}(20) = 41.07$$

1.2.6 Moda en datos desagregados

La moda es el valor que aparece con mayor frecuencia en un grupo de datos.
Pero también puede haber más de un valor modal en una serie de datos.

Cuadro 8

Edades de encuestados y número de repeticiones

Núm. de folio	Edad del votante (años)	Núm. de folio (modif.)	Núm. de Repeticiones del dato	Núm. de folio	Edad del votante (años)	Núm. de folio (modif.)	Número de Repeticiones del dato	Núm. de folio	Edad del votante (años)	Núm. de folio (modif.)	Número de Repeticiones del dato
1	18	1		69	30	36		17	52	71	
54	18	2	2	74	30	37		78	52	72	
15	19	3		97	30	38	4	90	52	73	3
16	19	4		64	32	39	1	83	53	74	1
47	19	5		5	33	40		8	55	75	
93	19	6		60	33	41	2	88	55	76	2
100	19	7	5	6	34	42		56	56	77	1
46	20	8	1	34	34	43	2	11	57	78	
25	21	9		32	35	44		38	57	79	
29	21	10	2	63	35	45	2	76	57	80	3
24	22	11		85	37	46	1	53	58	81	1
41	22	12	2	4	39	47		40	59	82	1
13	23	13		43	39	48	2	35	60	83	
94	23	14	2	44	40	49	1	71	60	84	2
33	24	15		20	41	50		96	62	85	1
37	24	16		30	41	51		51	63	86	
39	24	17		45	41	52	3	82	63	87	2
61	24	18	4	21	42	53		81	64	88	
7	25	19		58	42	54	2	92	64	89	2
36	25	20	2	18	43	55		31	66	90	
12	26	21		48	43	56		91	66	91	2
23	26	22		98	43	57	3	22	67	92	1
59	26	23	3	27	44	58		84	69	93	1
9	27	24		49	44	59		68	70	94	
28	27	25		65	44	60	3	89	70	95	2
57	27	26	3	62	45	61		87	71	96	1
2	28	27		72	45	62	2	66	72	97	1
19	28	28		3	46	63		86	74	98	1
55	28	29		26	46	64	2	77	79	99	1
95	28	30	4	50	47	65		79	82	100	1
10	29	31		52	47	66	2				
70	29	32		67	49	67					
80	29	33		73	49	68	2				
99	29	34	4	42	50	69					
14	30	35		75	50	70	2				

FUENTE: Elaboración del autor.

El valor modal, aparece 5 veces, es la edad de 19 años.

1.2.7 Moda en datos agrupados

Cuando se necesita una medida para datos agrupados análoga a la moda de los datos no agrupados, basta hablar de la “clase modal”, que es el intervalo de clase que contiene el mayor número de observaciones.

Puede ser que se tenga más de una clase modal.

Cuadro 9

Edades de encuestados por marcas de clase

Intervalo de clase	frecuencia fi
18-25	20
26-40	29
41-60	35
61-82	16
SUMA	100

FUENTE: Elaboración del autor.

Como puede observarse la clase modal es 41-60.
Resumiendo, se diría que a partir de estos datos:

LA MEDIA	Datos desagregados	41.54 años
LA MEDIA	Datos agrupados	42.98 años
LA MEDIANA	Datos desagregados	41 años
LA MEDIANA	Datos agrupados	41.07 años
LA MODA	Datos desagregados	19 años
LA MODA	Datos agrupados	Clase modal 41-60.

1.3 Medidas de variabilidad

Una medida de dispersión o variabilidad es una medida de la manera en la que los valores individuales se desvían del promedio.

1.3.1 Máximo, mínimo y rango en datos desagregados

Los siguientes datos son las edades de ciudadanos que eran potenciales votantes en una elección.

A partir de ellos obtenga el máximo, el mínimo y el rango.

El máximo es el valor más grande de todos, por lo que es necesario ordenar los datos de menor a mayor, para un más fácil manejo (véase el cuadro 6).

Como se observa, el máximo valor que adquiere la variable edad, es de 82 años; el mínimo valor que adquiere la variable es de 18 años; y, dado que el rango es la sustracción del mínimo respecto del máximo, entonces el rango es de 64. Resumiendo:

Máximo = 82
Mínimo = 18
Rango = Máximo – Mínimo = 82 – 18 = 64 años.

1.3.2 Máximo, mínimo y rango en datos agrupados

Si se parte de la tabla de frecuencias (cuadro 9):

Entonces el máximo es de 82 años

El mínimo es de 18 años

El rango es de 64 años.

Por otro lado:
El máximo inclusive o verdadero es de 82.5 años
El mínimo inclusive o verdadero es de 17.5 años
El rango inclusive o verdadero es de 65 años.

El rango tiene un valor limitado como medida de variabilidad.

En primer lugar, toma en cuenta solamente a los valores extremos de un conjunto de datos y no da ningún indicio sobre la forma como varían los valores en el interior del intervalo.

En segundo lugar, el rango de una muestra depende de su tamaño. Los valores extremos de una población, por ser menos numerosos, no son tan propensos a aparecer en las muestras pequeñas y sí en las grandes y, en consecuencia, las muestras pequeñas tienden a tener rangos pequeños y las muestras grandes rangos grandes.

1.3.3 Varianza en datos desagregados

Las desventajas del rango tratan de solucionarse con otra medida de variabilidad: la varianza.

La varianza de un conjunto de datos se obtiene restando a cada uno de los valores el valor de la media de todos los valores, elevando al cuadrado cada una de las diferencias resultantes, sumando las diferencias al cuadrado y dividiendo el total entre el número de observaciones menos 1.

En otras palabras, la varianza es la suma de las diferencias de las observaciones respecto a la media elevadas al cuadrado, dividida entre el total de observaciones, menos uno.

La fórmula de cálculo es la siguiente:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Considere los datos que se han venido trabajando, correspondientes a ciudadanos encuestados y sus edades, y calcule la *varianza*.

Cuadro 10

Edades de encuestados y operaciones para obtener la Varianza

Núm. de folio	Edad del votante (años) (X)	MEDIA DE X	X-X MEDIA	(X-X MEDIA)2
1	18	41.54	-23.54	554.13
2	28	41.54	-13.54	183.33
3	46	41.54	4.46	19.89
4	39	41.54	-2.54	6.45
5	33	41.54	-8.54	72.93
6	34	41.54	-7.54	56.85
7	25	41.54	-16.54	273.57
8	55	41.54	13.46	181.17
9	27	41.54	-14.54	211.41
10	29	41.54	-12.54	157.25
11	57	41.54	15.46	239.01
12	26	41.54	-15.54	241.49
13	23	41.54	-18.54	343.73
14	30	41.54	-11.54	133.17
15	19	41.54	-22.54	508.05
16	19	41.54	-22.54	508.05
17	52	41.54	10.46	109.41
18	43	41.54	1.46	2.13
19	28	41.54	-13.54	183.33
20	41	41.54	-0.54	0.29
21	42	41.54	0.46	0.21
22	67	41.54	25.46	648.21
23	26	41.54	-15.54	241.49
24	22	41.54	-19.54	381.81
25	21	41.54	-20.54	421.89
26	46	41.54	4.46	19.89
27	44	41.54	2.46	6.05
28	27	41.54	-14.54	211.41
29	21	41.54	-20.54	421.89
30	41	41.54	-0.54	0.29
31	66	41.54	24.46	598.29
32	35	41.54	-6.54	42.77
33	24	41.54	-17.54	307.65
34	34	41.54	-7.54	56.85
35	60	41.54	18.46	340.77
36	25	41.54	-16.54	273.57
37	24	41.54	-17.54	307.65
38	57	41.54	15.46	239.01
39	24	41.54	-17.54	307.65
40	59	41.54	17.46	304.85
41	22	41.54	-19.54	381.81
42	50	41.54	8.46	71.57
43	39	41.54	-2.54	6.45
44	40	41.54	-1.54	2.37
45	41	41.54	-0.54	0.29
46	20	41.54	-21.54	463.97
47	19	41.54	-22.54	508.05
48	43	41.54	1.46	2.13
49	44	41.54	2.46	6.05
50	47	41.54	5.46	29.81
51	63	41.54	21.46	460.53
52	47	41.54	5.46	29.81
53	58	41.54	16.46	270.93
54	18	41.54	-23.54	554.13
55	28	41.54	-13.54	183.33

Núm. de folio	Edad del votante (años) (X)	MEDIA DE X	X-X MEDIA	(X-X MEDIA)2
56	56	41.54	14.46	209.09
57	27	41.54	-14.54	211.41
58	42	41.54	0.46	0.21
59	26	41.54	-15.54	241.49
60	33	41.54	-8.54	72.93
61	24	41.54	-17.54	307.65
62	45	41.54	3.46	11.97
63	35	41.54	-6.54	42.77
64	32	41.54	-9.54	91.01
65	44	41.54	2.46	6.05
66	72	41.54	30.46	927.81
67	49	41.54	7.46	55.65
68	70	41.54	28.46	809.97
69	30	41.54	-11.54	133.17
70	29	41.54	-12.54	157.25
71	60	41.54	18.46	340.77
72	45	41.54	3.46	11.97
73	49	41.54	7.46	55.65
74	30	41.54	-11.54	133.17
75	50	41.54	8.46	71.57
76	57	41.54	15.46	239.01
77	79	41.54	37.46	1403.25
78	52	41.54	10.46	109.41
79	82	41.54	40.46	1637.01
80	29	41.54	-12.54	157.25
81	64	41.54	22.46	504.45
82	63	41.54	21.46	460.53
83	53	41.54	11.46	131.33
84	69	41.54	27.46	754.05
85	37	41.54	-4.54	20.61
86	74	41.54	32.46	1053.65
87	71	41.54	29.46	867.89
88	55	41.54	13.46	181.17
89	70	41.54	28.46	809.97
90	52	41.54	10.46	109.41
91	66	41.54	24.46	598.29
92	64	41.54	22.46	504.45
93	19	41.54	-22.54	508.05
94	23	41.54	-18.54	343.73
95	28	41.54	-13.54	183.33
96	62	41.54	20.46	418.61
97	30	41.54	-11.54	133.17
98	43	41.54	1.46	2.13
99	29	41.54	-12.54	157.25
100	19	41.54	-22.54	508.05
	4,154	**SUMAS**	5.68434E-14	27776.84
	41.54			

FUENTE: Elaboración del autor.

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{27776.84}{99} = 280.57$$

1.3.4 Varianza en datos agrupados

Con el fin de calcular la varianza a partir de datos agrupados se debe suponer, como en el caso del cálculo de la media, que las observaciones de cada marca de clase se localizan en el punto medio de la misma.

La fórmula de cálculo es la siguiente:

$$S^2 = \frac{n\sum_{i=1}^{k} m_i^2 f_i - (\sum_{i=1}^{k} m_i f_i)^2}{n(n-1)}$$

Cuadro 11

Edades de encuestados por marca de clase y operaciones para obtener la Varianza

Intervalo de clase	frecuencia fi	mi	mifi	mi2fi
18-25	20	21.5	430	9,245.0
26-40	29	33	957	31,581.0
41-60	35	50.5	1,767.5	89,258.8
61-82	16	71.5	1144	81,796.0
SUMA	100		4,298.5	211,880.75
			18,477,102.3	

FUENTE: Elaboración del autor.

$$S^2 = \frac{n\sum_{i=1}^{k} m_i^2 f_i - (\sum_{i=1}^{k} m_i f_i)^2}{n(n-1)} = \frac{100 * 211880.75 - 18477102.3}{100(99)} = 273.83$$

1.3.5 Desviación estándar en datos desagregados

La fórmula de cálculo es la siguiente:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{280.57} = 16.75$$

La diferencia entre la varianza y la desviación estándar es que la primera se da en unidades elevadas al cuadrado, en este caso años al cuadrado, en tanto que la segunda está dada en unidades reales, en este caso años.

1.3.6. Desviación estándar en datos agrupados

Con el fin de calcular la desviación estándar a partir de datos agrupados se debe suponer, como en el caso del cálculo de la media, que las observaciones de cada marca de clase se localizan en el punto medio de la misma.

Adicionalmente, se sabe que la diferencia entre la varianza y la desviación estándar es que la primera se da en unidades elevadas al cuadrado, en este caso años al cuadrado, en tanto que la segunda está dada en unidades reales, en este caso años.

La fórmula de cálculo es la siguiente:

$$S = \sqrt{s^2}$$

Entonces:

$$S = \sqrt{s^2} = \sqrt{273.83} = 16.55$$

Resumiendo, se diría que a partir de estos datos:

El *máximo*, el *mínimo* y el *rango*

Para datos desagregados:

Son	Máximo	=	82
	Mínimo	=	18
	Rango	=	64 años

Para datos agrupados:

Máximo	=	82
Mínimo	=	18
Rango	=	64 años

La *varianza*

Para datos desagregados: 280.57

Para datos agrupados: 273.83

La desviación *estándar*

Para datos desagregados: 16.75

Para datos agrupados: 16.55

1.4 Otras medidas descriptivas

1.4.1 Asimetría y curtosis

El rasgo más notorio e identificable de una distribución, a partir de un histograma o de un polígono de frecuencias, es la forma que adquiere la distribución.[2]

La primera parte de ese rasgo es el grado de apuntalamiento, es decir, qué tan alto o bajo es el pico de la distribución.

Una segunda parte de la forma de una distribución es la simetría o asimetría. La idea general de simetría es bastante sencilla. Sabemos que la mediana divide el histograma en dos áreas de la misma superficie. Pues bien, se dice que la distribución de frecuencias es simétrica cuando una de las áreas es imagen de la otra.

Cuando la curva es simétrica, la mediana coincide con la media. Si, además, la distribución de frecuencias es unimodal, la moda coincide igualmente con la media y la mediana.

Se dice que la simetría es positiva si la mayoría de las puntuaciones son bajas y la minoría altas, mientras que la simetría es negativa si sucede lo contrario.

Si la asimetría es negativa, el orden es de izquierda a derecha; es decir, primero está la media, después la mediana y, por último, la moda. Si la asimetría es positiva, el orden es el contrario, esto es, moda, mediana y media.

La desviación de las puntuaciones en relación con la media de una distribución se suele expresar así:

$$x = (x_i - \overline{x})$$

El momento de primer orden con respecto a la media aritmética es, simplemente, el promedio de la primera potencia de las desviaciones con respecto a la media; esto es:

$$m_1 = \frac{\sum x}{N}$$

Dado que la suma de las desviaciones con respecto a la media es siempre 0, el momento de primer orden es también cero, lo que representa una característica definidora de la media. Si se utilizan potencias más elevadas, se obtienen nuevas medidas que ofrecen mayor información estadística. Así, el momento de segundo orden es la varianza:

$$m_2 = \frac{\sum x^2}{N}$$

Otros dos momentos de interés estadístico son los de tercer y cuarto orden, que se definen como los promedios de las potencias de tercer y cuarto orden de las desviaciones con respecto a la media, respectivamente:

$$m_3 = \frac{\sum x^3}{N}$$

$$m_4 = \frac{\sum x^4}{N}$$

---

[2] Puede consultarse García Ferrando, Manuel. *Socioestadística. Introducción a la estadística en sociología*. Alianza Universidad Textos, Madrid, 1997.

Se definen del siguiente modo:

$$B_1 = \frac{m_3}{\sqrt{m_2^3}}$$ Asimetría

$$B_2 = \frac{m_4}{m_2^2}$$ Curtosis

Si la curva está sesgada a la derecha, B_1 tendrá un valor positivo, mientras que si el sesgo es negativo, B_1 ofrecerá un valor negativo.

Los valores pequeños de B_2 representan una curva platicúrtica (más baja que la curva normal), mientras que valores altos de B_2 indican una distribución leptocúrtica o apuntada. La curva normal, es mesocúrtica, tiene un valor de B_2 igual a 3.

Ejemplo 1

Considérense las siguientes cifras, las cuales corresponden al caso de un medicamento para personas que sufran un paro cardiaco, el cual se encuentra en etapa de experimentación. Las observaciones "x" indican la cantidad de una sustancia que se ha inyectado a distintos pacientes y las observaciones "y" indican el tiempo, en segundos, que tarda el paciente en reaccionar.

Cuadro 12
Milímetros inyectados al paciente y
tiempo de reacción medida en segundos
en pacientes que sufren un paro cardiaco

	Xi	Yi
1	0.5	12
2	0.5	22
3	0.5	30
4	1	18
5	1	32
6	1	36
7	1.5	30
8	1.5	34
9	1.5	46
10	2	40
11	2	44
12	2	50
13	2.5	44
14	2.5	60
15	2.5	64
16	3	64
17	3	68
18	3	76
	31.5	770
	1.75	42.78

FUENTE: Wayme, *Estadística*, 317. Si bien el autor lo aplica a otra herramienta estadística.

Las fórmulas para el cálculo de los momentos son las siguientes:

$$m_1 = \frac{\sum x}{N} \qquad m_2 = \frac{\sum x^2}{N}$$

$$m_3 = \frac{\sum x^3}{N} \qquad m_4 = \frac{\sum x^4}{N}$$

Para el cálculo de los momentos, tomando el tiempo de reacción, se presenta el cuadro 13.

Cuadro 13

Milímetros inyectados al paciente y tiempo de reacción medida en segundos (Cálculos para asimetría y curtosis)

	x_i	* x	** x^2	x^3	x^4
1	12	-30.78	947.27	- 29,154.91	897,323.49
2	22	-20.78	431.72	- 8,970.10	186,378.75
3	30	-12.78	163.27	- 2,086.25	26,657.62
4	18	-24.78	613.94	- 15,212.03	376,920.20
5	32	-10.78	116.16	- 1,251.95	13,493.26
6	36	-6.78	45.94	- 311.36	2,110.32
7	30	-12.78	163.27	- 2,086.25	26,657.62
8	34	-8.78	77.05	- 676.32	5,936.61
9	46	3.22	10.38	33.46	107.80
10	40	-2.78	7.72	- 21.43	59.54
11	44	1.22	1.49	1.83	2.23
12	50	7.22	52.16	376.71	2,720.72
13	44	1.22	1.49	1.83	2.23
14	60	17.22	296.60	5,108.20	87,974.49
15	64	21.22	450.38	9,558.12	202,844.59
16	64	21.22	450.38	9,558.12	202,844.59
17	68	25.22	636.16	16,045.38	404,700.17
18	76	33.22	1,103.72	36,667.90	1,218,189.12
Suma	770	0.00	5,569.11	17,580.94	3,654,923.35
Media	42.78				

* Se toma a las Yi del ejercicio anterior como Xi.

** Ahora $x = (x_i - \overline{x})$

FUENTE: Elaboración del autor con base en Wayme, *Estadística*, 317.

Los resultados de los momentos son los siguientes:

$$m_1 = \frac{\sum x}{N} = \frac{0}{18} = 0$$

$$m_2 = \frac{\sum x^2}{N} = \frac{5569.11}{18} = 309.39$$

$$m_3 = \frac{\sum x^3}{N} = \frac{17580.94}{18} = 976.72$$

$$m_4 = \frac{\sum x^4}{N} = \frac{3654923.35}{18} = 203051.29$$

A continuación se presentan las fórmulas, cálculos y resultados de asimetría y curtosis de los datos:

$$B_1 = \frac{m_3}{\sqrt{m_2{}^3}}$$ Asimetría

$$B_2 = \frac{m_4}{m_2{}^2}$$ Curtosis

Asimetría

$$B_1 = \frac{m_3}{\sqrt{m_2{}^3}} = \frac{976.72}{\sqrt{309.4^3}}$$

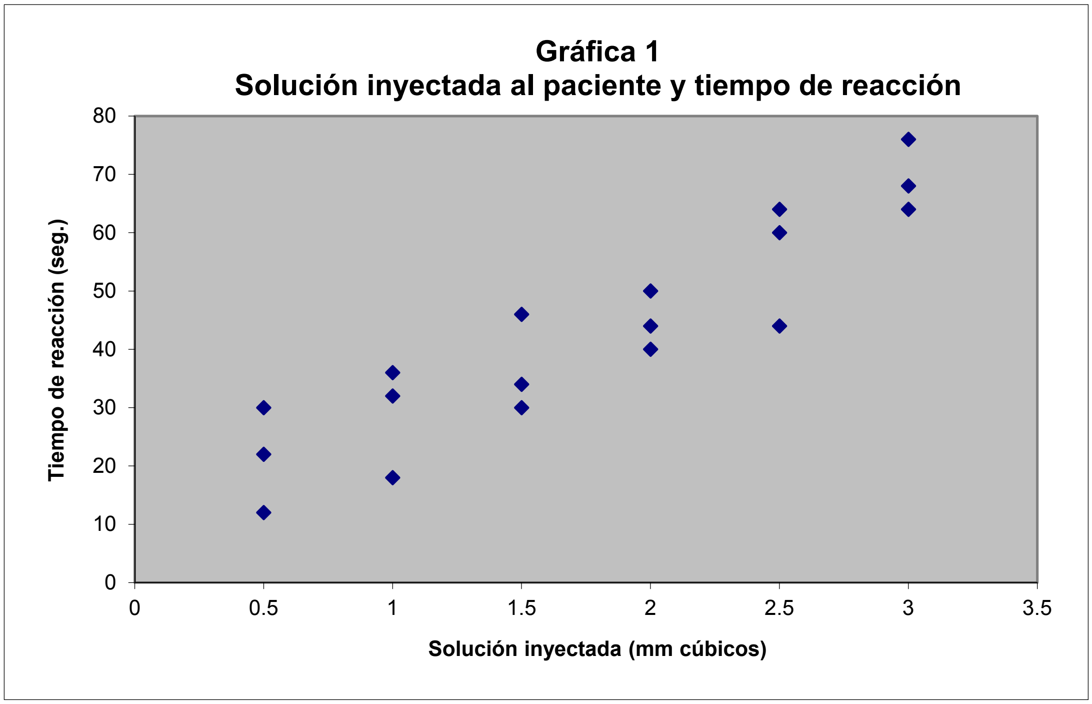
$$B_1 = \frac{m_3}{\sqrt{m_2{}^3}} = .1795$$

Curtosis

$$B_2 = \frac{m_4}{m_2{}^2} = \frac{203051.3}{309.4^2}$$

$$B_2 = \frac{m_4}{m_2{}^2} = 2.12$$

Observe la gráfica 1 y considere que deben realizarse las operaciones, ya que a simple vista resulta difícil dar cuenta de la tendencia.



FUENTE: Elaboración del autor.

Ejemplo 2

Para este ejemplo se deben considerar las cifras del cuadro 14.

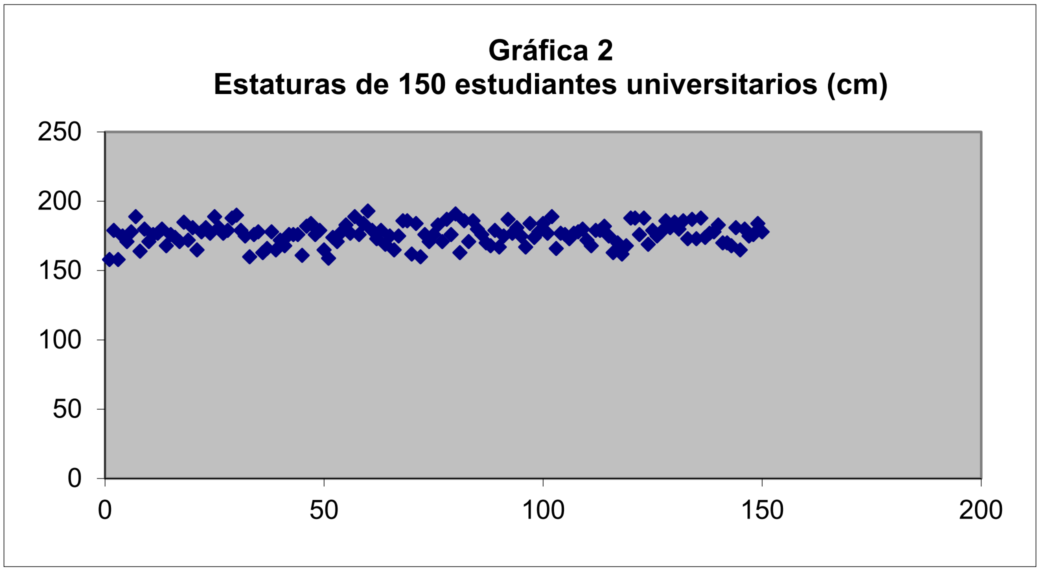
Cuadro 14

Folio y estatura en 150 estudiantes universitarios (centímetros lineales)

Folio	Estatura en cm	Folio	Estatura en cm	Folio	Estatura en cm
1	158	51	159	101	177
2	179	52	174	102	189
3	158	53	171	103	166
4	175	54	178	104	177
5	171	55	183	105	176
6	178	56	177	106	173
7	189	57	189	107	177
8	164	58	176	108	178
9	180	59	184	109	180
10	171	60	193	110	172
11	176	61	179	111	168
12	177	62	173	112	179
13	180	63	179	113	179
14	168	64	169	114	182
15	176	65	175	115	175
16	174	66	165	116	163
17	171	67	175	117	170
18	185	68	186	118	162
19	172	69	186	119	168
20	181	70	162	120	188
21	165	71	184	121	188
22	178	72	160	122	176
23	181	73	176	123	188
24	177	74	171	124	169
25	189	75	176	125	179
26	181	76	183	126	175
27	177	77	171	127	178
28	179	78	187	128	186
29	188	79	176	129	181
30	190	80	191	130	185
31	179	81	163	131	180
32	175	82	186	132	186
33	160	83	171	133	173
34	176	84	186	134	187
35	178	85	180	135	173
36	163	86	176	136	188
37	166	87	170	137	174
38	178	88	168	138	177
39	165	89	179	139	178
40	172	90	167	140	183
41	168	91	175	141	170
42	176	92	187	142	170
43	176	93	177	143	168
44	176	94	181	144	181
45	161	95	175	145	165
46	182	96	167	146	180
47	184	97	184	147	175
48	176	98	174	148	176
49	179	99	178	149	184
50	165	100	184	150	178

FUENTE: Elaboración del autor con base en Wayme, *Estadística,* 13.
El autor ha utilizado los datos para otros fines estadísticos.

Los datos se presentan en una gráfica, observe que en la nube de puntos no es posible decir qué tipo de asimetría se tiene.



FUENTE: Elaboración del autor.

Las fórmulas para el cálculo de los momentos son las siguientes:

$$m_1 = \frac{\sum x}{N} \qquad m_2 = \frac{\sum x^2}{N}$$

$$m_3 = \frac{\sum x^3}{N} \qquad m_4 = \frac{\sum x^4}{N}$$

Para el cálculo de los momentos se presenta el cuadro 15.

Cuadro 15

Folio y estatura en 150 estudiantes universitarios (centímetros lineales)

No.	x_i	* x	x^2	x^3	x^4	x_i
1	158	- 18.19	331.00	- 6,021.95	109,559.26	158
2	179	2.81	7.88	22.11	62.05	158
3	158	- 18.19	331.00	- 6,021.95	109,559.26	159
4	175	- 1.19	1.42	- 1.70	2.03	160
5	171	- 5.19	26.97	- 140.07	727.42	160
6	178	1.81	3.26	5.90	10.65	161
7	189	12.81	164.01	2,100.43	26,899.51	162
8	164	- 12.19	148.68	- 1,812.87	22,104.96	162
9	180	3.81	14.49	55.16	209.98	163
10	171	- 5.19	26.97	- 140.07	727.42	163
11	176	- 0.19	0.04	- 0.01	0.00	163
12	177	0.81	0.65	0.52	0.42	164
13	180	3.81	14.49	55.16	209.98	165
14	168	- 8.19	67.13	- 550.02	4,506.53	165
15	176	- 0.19	0.04	- 0.01	0.00	165
16	174	- 2.19	4.81	- 10.55	23.14	165
17	171	- 5.19	26.97	- 140.07	727.42	165
18	185	8.81	77.56	683.02	6,015.15	166
19	172	- 4.19	17.58	- 73.74	309.20	166
20	181	4.81	23.10	111.05	533.80	167
21	165	- 11.19	125.29	- 1,402.42	15,697.76	167
22	178	1.81	3.26	5.90	10.65	168
23	181	4.81	23.10	111.05	533.80	168
24	177	0.81	0.65	0.52	0.42	168
25	189	12.81	164.01	2,100.43	26,899.51	168
26	181	4.81	23.10	111.05	533.80	168
27	177	0.81	0.65	0.52	0.42	168
28	179	2.81	7.88	22.11	62.05	169
29	188	11.81	139.40	1,645.82	19,431.63	169
30	190	13.81	190.62	2,631.88	36,337.53	170
31	179	2.81	7.88	22.11	62.05	170
32	175	- 1.19	1.42	- 1.70	2.03	170
33	160	- 16.19	262.22	- 4,246.28	68,761.45	170
34	176	- 0.19	0.04	- 0.01	0.00	171
35	178	1.81	3.26	5.90	10.65	171
36	163	- 13.19	174.06	- 2,296.48	30,298.29	171
37	166	- 10.19	103.90	- 1,059.13	10,796.05	171
38	178	1.81	3.26	5.90	10.65	171
39	165	- 11.19	125.29	- 1,402.42	15,697.76	171
40	172	- 4.19	17.58	- 73.74	309.20	171
41	168	- 8.19	67.13	- 550.02	4,506.53	172
42	176	- 0.19	0.04	- 0.01	0.00	172
43	176	- 0.19	0.04	- 0.01	0.00	172
44	176	- 0.19	0.04	- 0.01	0.00	173
45	161	- 15.19	230.84	- 3,507.19	53,285.89	173
46	182	5.81	33.72	195.79	1,136.86	173
47	184	7.81	60.94	475.77	3,714.18	173
48	176	- 0.19	0.04	- 0.01	0.00	174
49	179	2.81	7.88	22.11	62.05	174
50	165	- 11.19	125.29	- 1,402.42	15,697.76	174
51	159	- 17.19	295.61	- 5,082.53	87,385.69	174
52	174	- 2.19	4.81	- 10.55	23.14	175
53	171	- 5.19	26.97	- 140.07	727.42	175
54	178	1.81	3.26	5.90	10.65	175
55	183	6.81	46.33	315.36	2,146.53	175
56	177	0.81	0.65	0.52	0.42	175
57	189	12.81	164.01	2,100.43	26,899.51	175
58	176	- 0.19	0.04	- 0.01	0.00	175

59	184	7.81	60.94	475.77	3,714.18	175
60	193	16.81	282.46	4,747.28	79,785.94	175
61	179	2.81	7.88	22.11	62.05	176
62	173	- 3.19	10.20	- 32.56	103.99	176
63	179	2.81	7.88	22.11	62.05	176
64	169	- 7.19	51.74	- 372.21	2,677.45	176
65	175	- 1.19	1.42	- 1.70	2.03	176
66	165	- 11.19	125.29	- 1,402.42	15,697.76	176
67	175	- 1.19	1.42	- 1.70	2.03	176
68	186	9.81	96.17	943.11	9,248.81	176
69	186	9.81	96.17	943.11	9,248.81	176
70	162	- 14.19	201.45	- 2,859.26	40,582.39	176
71	184	7.81	60.94	475.77	3,714.18	176
72	160	- 16.19	262.22	- 4,246.28	68,761.45	176
73	176	- 0.19	0.04	- 0.01	0.00	176
74	171	- 5.19	26.97	- 140.07	727.42	176
75	176	- 0.19	0.04	- 0.01	0.00	176

(Continúa)

76	183	6.81	46.33	315.36	2146.53	177
77	171	- 5.19	26.97	-140.07	727.42	177
78	187	10.81	116.78	1262.05	13638.51	177
79	176	- 0.19	0.04	-0.01	0.00	177
80	191	14.81	219.24	3246.17	48065.03	177
81	163	- 13.19	174.06	-2296.48	30298.29	177
82	186	9.81	96.17	943.11	9248.81	177
83	171	- 5.19	26.97	-140.07	727.42	177
84	186	9.81	96.17	943.11	9248.81	177
85	180	3.81	14.49	55.16	209.98	178
86	176	- 0.19	0.04	-0.01	0.00	178
87	170	- 6.19	38.36	-237.56	1471.29	178
88	168	- 8.19	67.13	-550.02	4506.53	178
89	179	2.81	7.88	22.11	62.05	178
90	167	- 9.19	84.52	-777.00	7143.19	178
91	175	- 1.19	1.42	-1.70	2.03	178
92	187	10.81	116.78	1262.05	13638.51	178
93	177	0.81	0.65	0.52	0.42	178
94	181	4.81	23.10	111.05	533.80	178
95	175	- 1.19	1.42	-1.70	2.03	179
96	167	- 9.19	84.52	-777.00	7143.19	179
97	184	7.81	60.94	475.77	3714.18	179
98	174	- 2.19	4.81	-10.55	23.14	179
99	178	1.81	3.26	5.90	10.65	179
100	184	7.81	60.94	475.77	3714.18	179
101	177	0.81	0.65	0.52	0.42	179
102	189	12.81	164.01	2100.43	26899.51	179
103	166	- 10.19	103.90	-1059.13	10796.05	179
104	177	0.81	0.65	0.52	0.42	179
105	176	- 0.19	0.04	-0.01	0.00	180
106	173	- 3.19	10.20	-32.56	103.99	180
107	177	0.81	0.65	0.52	0.42	180
108	178	1.81	3.26	5.90	10.65	180
109	180	3.81	14.49	55.16	209.98	180
110	172	- 4.19	17.58	-73.74	309.20	180
111	168	- 8.19	67.13	-550.02	4506.53	181
112	179	2.81	7.88	22.11	62.05	181
113	179	2.81	7.88	22.11	62.05	181
114	182	5.81	33.72	195.79	1136.86	181
115	175	- 1.19	1.42	-1.70	2.03	181
116	163	- 13.19	174.06	-2296.48	30298.29	181
117	170	- 6.19	38.36	-237.56	1471.29	182
118	162	- 14.19	201.45	-2859.26	40582.39	182
119	168	- 8.19	67.13	-550.02	4506.53	183
120	188	11.81	139.40	1645.82	19431.63	183
121	188	11.81	139.40	1645.82	19431.63	183

122	176	- 0.19	0.04	-0.01	0.00	184
123	188	11.81	139.40	1645.82	19431.63	184
124	169	- 7.19	51.74	-372.21	2677.45	184
125	179	2.81	7.88	22.11	62.05	184
126	175	- 1.19	1.42	-1.70	2.03	184
127	178	1.81	3.26	5.90	10.65	184
128	186	9.81	96.17	943.11	9248.81	185
129	181	4.81	23.10	111.05	533.80	185
130	185	8.81	77.56	683.02	6015.15	186
131	180	3.81	14.49	55.16	209.98	186
132	186	9.81	96.17	943.11	9248.81	186
133	173	- 3.19	10.20	-32.56	103.99	186
134	187	10.81	116.78	1262.05	13638.51	186
135	173	- 3.19	10.20	-32.56	103.99	186
136	188	11.81	139.40	1645.82	19431.63	187
137	174	- 2.19	4.81	-10.55	23.14	187
138	177	0.81	0.65	0.52	0.42	187
139	178	1.81	3.26	5.90	10.65	188
140	183	6.81	46.33	315.36	2146.53	188
141	170	- 6.19	38.36	-237.56	1471.29	188
142	170	- 6.19	38.36	-237.56	1471.29	188
143	168	- 8.19	67.13	-550.02	4506.53	188
144	181	4.81	23.10	111.05	533.80	189
145	165	- 11.19	125.29	-1402.42	15697.76	189
146	180	3.81	14.49	55.16	209.98	189
147	175	- 1.19	1.42	-1.70	2.03	189
148	176	- 0.19	0.04	-0.01	0.00	190
149	184	7.81	60.94	475.77	3714.18	191
150	178	1.81	3.26	5.90	10.65	193
SUMA	26,429	0	8,807	- 17,074	1364245.77	26429
MEDIA	176.19					

FUENTE: Elaboración del autor con base en Wayme, Estadística, 13.

Los resultados de los momentos son los siguientes:

$$m_1 = \frac{\sum x}{N} = \frac{0}{150} = 0$$

$$m_2 = \frac{\sum x^2}{N} = \frac{8807}{150} = 58.72$$

$$m_3 = \frac{\sum x^3}{N} = \frac{17074}{150} = -113.83$$

$$m_4 = \frac{\sum x^4}{N} = \frac{1364245.77}{150} = 9094.97$$

A continuación se presentan las fórmulas, cálculos y resultados de asimetría y curtosis de los datos:

$$B_1 = \frac{m_3}{\sqrt{m_2^{\ 3}}}$$ Asimetría

$$B_2 = \frac{m_4}{m_2^{\ 2}}$$ Curtosis

Asimetría

$$B_1 = \frac{m_3}{\sqrt{m_2^{\ 3}}} = \frac{-113.83}{\sqrt{58.72^3}}$$

$$B_1 = \frac{m_3}{\sqrt{m_2^{\ 3}}} = -.2529$$

Curtosis

$$B_2 = \frac{m_4}{m_2^{\ 2}} = \frac{9094.97}{58.72^2}$$

$$B_2 = \frac{m_4}{m_2^{\ 2}} = 2.64$$

Ejemplo 3

Considérense las cifras del cuadro 16.

Cuadro 16

Número de horas mensuales que universitarios ven televisión

Número	x_i	Número	x_i
1	40	54	36
2	49	55	40
3	65	56	39
4	30	57	44
5	32	58	60
6	36	59	50
7	31	60	65
8	54	61	93
9	31	62	38
10	37	63	41
11	41	64	31
12	35	65	52
13	97	66	46
14	32	67	33
15	75	68	40
16	40	69	56
17	36	70	50
18	67	71	32
19	34	72	38
20	30	73	39
21	39	74	35
22	84	75	36
23	88	76	44
24	48	77	38
25	68	78	44
26	34	79	45
27	81	80	38
28	44	81	71
29	76	82	42
30	54	83	58
31	61	84	36
32	51	85	38
33	39	86	45
34	38	87	61
35	37	88	47
36	42	89	48
37	51	90	43
38	42	91	62
39	52	92	48
40	30	93	53
41	47	94	41
42	52	95	51
43	35	96	36
44	63	97	55
45	50	98	45
46	53	99	45
47	66	100	55
48	50	SUMA	4812
49	41	MEDIA	48.12
50	50		
51	43		
52	55		
53	73		

FUENTE: Elaboración del autor.

Las fórmulas para el cálculo de los momentos son las siguientes:

$$m_1 = \frac{\sum x}{N} \qquad m_2 = \frac{\sum x^2}{N}$$

$$m_3 = \frac{\sum x^3}{N} \qquad m_4 = \frac{\sum x^4}{N}$$

Para el cálculo de los momentos se presenta el cuadro 17.

Cuadro 17

Número de horas mensuales que universitarios ven televisión (Calculos para Asimetría y Curtosis)

			*	**	
	x_i	x	x^2	x^3	x^4
1	40	-8.12	65.93	-535.39	4347.35
2	49	0.88	0.77	0.68	0.60
3	65	16.88	284.93	4809.69	81187.61
4	30	-18.12	328.33	-5949.42	107803.48
5	32	-16.12	259.85	-4188.85	67524.31
6	36	-12.12	146.89	-1780.36	21577.96
7	31	-17.12	293.09	-5017.78	85904.33
8	54	5.88	34.57	203.30	1195.39
9	31	-17.12	293.09	-5017.78	85904.33
10	37	-11.12	123.65	-1375.04	15290.41
11	41	-7.12	50.69	-360.94	2569.92
12	35	-13.12	172.13	-2258.40	29630.25
13	97	48.88	2389.25	116786.76	5708536.59
14	32	-16.12	259.85	-4188.85	67524.31
15	75	26.88	722.53	19421.72	522055.96
16	40	-8.12	65.93	-535.39	4347.35
17	36	-12.12	146.89	-1780.36	21577.96
18	67	18.88	356.45	6729.86	127059.74
19	34	-14.12	199.37	-2815.17	39750.15
20	30	-18.12	328.33	-5949.42	107803.48
21	39	-9.12	83.17	-758.55	6917.98
22	84	35.88	1287.37	46190.99	1657332.85
23	88	39.88	1590.41	63425.73	2529417.96
24	48	-0.12	0.01	0.00	0.00
25	68	19.88	395.21	7856.86	156194.42
26	34	-14.12	199.37	-2815.17	39750.15
27	81	32.88	1081.09	35546.38	1168765.10
28	44	-4.12	16.97	-69.93	288.13
29	76	27.88	777.29	21670.97	604186.58
30	54	5.88	34.57	203.30	1195.39
31	61	12.88	165.89	2136.72	27520.95
32	51	2.88	8.29	23.89	68.80
33	39	-9.12	83.17	-758.55	6917.98
34	38	-10.12	102.41	-1036.43	10488.71
35	37	-11.12	123.65	-1375.04	15290.41
36	42	-6.12	37.45	-229.22	1402.83
37	51	2.88	8.29	23.89	68.80
38	42	-6.12	37.45	-229.22	1402.83
39	52	3.88	15.05	58.41	226.63
40	30	-18.12	328.33	-5949.42	107803.48
41	47	-1.12	1.25	-1.40	1.57
42	52	3.88	15.05	58.41	226.63
43	35	-13.12	172.13	-2258.40	29630.25
44	63	14.88	221.41	3294.65	49024.34
45	50	1.88	3.53	6.64	12.49
46	53	4.88	23.81	116.21	567.13
47	66	17.88	319.69	5716.14	102204.51
48	50	1.88	3.53	6.64	12.49
49	41	-7.12	50.69	-360.94	2569.92
50	50	1.88	3.53	6.64	12.49

(Continúa)

(Continuación)

51	43	-5.12	26.2144	-134.217728	687.1947674
52	55	6.88	47.33	325.66	2240.55
53	73	24.88	619.01	15401.08	383178.83
54	36	-12.12	146.89	-1780.36	21577.96
55	40	-8.12	65.93	-535.39	4347.35
56	39	-9.12	83.17	-758.55	6917.98
57	44	-4.12	16.97	-69.93	288.13
58	60	11.88	141.13	1676.68	19918.92
59	50	1.88	3.53	6.64	12.49
60	65	16.88	284.93	4809.69	81187.61
61	93	44.88	2014.21	90397.94	4057059.65
62	38	-10.12	102.41	-1036.43	10488.71
63	41	-7.12	50.69	-360.94	2569.92
64	31	-17.12	293.09	-5017.78	85904.33
65	52	3.88	15.05	58.41	226.63
66	46	-2.12	4.49	-9.53	20.20
67	33	-15.12	228.61	-3456.65	52264.54
68	40	-8.12	65.93	-535.39	4347.35
69	56	7.88	62.09	489.30	3855.71
70	50	1.88	3.53	6.64	12.49
71	32	-16.12	259.85	-4188.85	67524.31
72	38	-10.12	102.41	-1036.43	10488.71
73	39	-9.12	83.17	-758.55	6917.98
74	35	-13.12	172.13	-2258.40	29630.25
75	36	-12.12	146.89	-1780.36	21577.96
76	44	-4.12	16.97	-69.93	288.13
77	38	-10.12	102.41	-1036.43	10488.71
78	44	-4.12	16.97	-69.93	288.13
79	45	-3.12	9.73	-30.37	94.76
80	38	-10.12	102.41	-1036.43	10488.71
81	71	22.88	523.49	11977.55	274046.39
82	42	-6.12	37.45	-229.22	1402.83
83	58	9.88	97.61	964.43	9528.57
84	36	-12.12	146.89	-1780.36	21577.96
85	38	-10.12	102.41	-1036.43	10488.71
86	45	-3.12	9.73	-30.37	94.76
87	61	12.88	165.89	2136.72	27520.95
88	47	-1.12	1.25	-1.40	1.57
89	48	-0.12	0.01	0.00	0.00
90	43	-5.12	26.21	-134.22	687.19
91	62	13.88	192.65	2674.04	37115.72
92	48	-0.12	0.01	0.00	0.00
93	53	4.88	23.81	116.21	567.13
94	41	-7.12	50.69	-360.94	2569.92
95	51	2.88	8.29	23.89	68.80
96	36	-12.12	146.89	-1780.36	21577.96
97	55	6.88	47.33	325.66	2240.55
98	45	-3.12	9.73	-30.37	94.76
99	45	-3.12	9.73	-30.37	94.76
100	55	6.88	47.33	325.66	2240.55
SUMA	4812	1.27898E-13	20648.56	377040.35	18927906.58
MEDIA	48.12				

FUENTE: Elaboración del autor.

Los resultados de los momentos son los siguientes:

$$m_1 = \frac{\sum x}{N} = \frac{0}{100} = 0$$

$$m_2 = \frac{\sum x^2}{N} = \frac{20648.56}{100} = 206.49$$

$$m_3 = \frac{\sum x^3}{N} = \frac{377040.35}{100} = 3770.49$$

$$m_4 = \frac{\sum x^4}{N} = \frac{18927906.58}{100} = 189279.07$$

A continuación se presentan las fórmulas, cálculos y resultados de asimetría y curtosis de los datos:

$$B_1 = \frac{m_3}{\sqrt{{m_2}^3}}$$ Asimetría

$$B_2 = \frac{m_4}{{m_2}^2}$$ Curtosis

Asimetría

$$B_1 = \frac{m_3}{\sqrt{{m_2}^3}} = \frac{3770.40}{\sqrt{206.49^3}}$$

$$B_1 = \frac{m_3}{\sqrt{{m_2}^3}} = 1.27$$

Curtosis

$$B_2 = \frac{m_4}{{m_2}^2} = \frac{189279.07}{206.49^2}$$

$$B_2 = \frac{m_4}{{m_2}^2} = 4.44$$

Ejemplo 4

Considérense las cifras del cuadro 18.

Cuadro 18

Votación porcentual por el PRI en las elecciones municipales de 1994 en Tlaxcala

No.	Municipio	Votación (%)
1	Amaxac	52.05
2	A.Carvajal	44.62
3	Apizaco	58.89
4	Atlangatepec	100.00
5	Altzayanca	80.38
6	Calpulalpan	33.45
7	El Carmen	80.91
8	Cuapiaxtla	59.13
9	Cuaxomulco	36.68
10	Chiautempan	40.67
11	D. Arenas	67.74
12	Españita	77.52
13	Huamantla	79.54
14	Heuyotlipan	45.97
15	Ixtacuixtla	45.13
16	Ixtenco	46.42
17	Morelos	39.49
18	J. Cuamatzi	48.76
19	Lardizabal	49.75
20	L. Cárdenas	51.54
21	M. Arista	62.29
22	Hidalgo	67.96
23	Nativitas	56.36
24	Panotla	68.86
25	S. Pablo	75.48
26	S. Cruz T.	63.18
27	Tenancingo	52.40
28	Teolocholco	58.24
29	Tepeyanco	57.16
30	Terrenate	67.31
31	Tetla	71.32
32	Tetlatlahuca	71.34
33	Tlaxcala	64.17
34	Tlaxco	76.77
35	Tocatlán	73.95
36	Totolac	66.98
37	Trinidad de SS.	77.54
38	Tzonpantepec	72.47
39	Xaloztoc	44.80
40	Xaltican	58.75
41	Xicohtencatl	50.18
42	Xicotzingo	49.89
43	Yauhquemecan	70.54
44	Zacatelco	53.24
		2,669.82
		60.68

FUENTE: Elaboración del autor con datos del IET.

Las fórmulas para el cálculo de los momentos son las siguientes:

$$m_1 = \frac{\sum x}{N} \qquad\qquad m_2 = \frac{\sum x^2}{N}$$

$$m_3 = \frac{\sum x^3}{N} \qquad m_4 = \frac{\sum x^4}{N}$$

Para el cálculo de los momentos se presenta el cuadro 19.

Cuadro 19

Votación porcentual por el PRI en las elecciones municipales de 1994 en Tlaxcala
(Cálculos para la Asimetría y Curtosis)

No.	Municipio	Votación (%)	* x	**		
1	Amaxac	52.05	-8.63	74.48	-642.74	5546.81
2	A.Carvajal	44.62	-16.06	257.92	-4142.25	66524.58
3	Apizaco	58.89	-1.79	3.20	-5.74	10.27
4	Atlangatepec	100.00	39.32	1546.06	60791.17	2390308.94
5	Altzayanca	80.38	19.70	388.09	7645.37	150613.85
6	Calpulalpan	33.45	-27.23	741.47	-20190.31	549782.06
7	El Carmen	80.91	20.23	409.25	8279.19	167487.94
8	Cuapiaxtla	59.13	-1.55	2.40	-3.72	5.77
9	Cuaxomulco	36.68	-24.00	576.00	-13824.00	331776.00
10	Chiautempan	40.67	-20.01	400.40	-8012.01	160320.24
11	D. Arenas	67.74	7.06	49.84	351.90	2484.38
12	Españita	77.52	16.84	283.59	4775.58	80420.79
13	Huamantla	79.54	18.86	355.70	6708.49	126522.21
14	Heuyotlipan	45.97	-14.71	216.38	-3183.01	46822.08
15	Ixtacuixtla	45.13	-15.55	241.80	-3760.03	58468.45
16	Ixtenco	46.42	-14.26	203.35	-2899.74	41350.25
17	Morelos	39.49	-21.19	449.02	-9514.65	201615.46
18	J. Cuamatzi	48.76	-11.92	142.09	-1693.67	20188.55
19	Lardizabal	49.75	-10.93	119.46	-1305.75	14271.86
20	L. Cárdenas	51.54	-9.14	83.54	-763.55	6978.86
21	M. Arista	62.29	1.61	2.59	4.17	6.72
22	Hidalgo	67.96	7.28	53.00	385.83	2808.83
23	Nativitas	56.36	-4.32	18.66	-80.62	348.29
24	Panotla	68.86	8.18	66.91	547.34	4477.27
25	S. Pablo	75.48	14.80	219.04	3241.79	47978.52
26	S. Cruz T.	63.18	2.50	6.25	15.63	39.06
27	Tenancingo	52.40	-8.28	68.56	-567.66	4700.25
28	Teolocholco	58.24	-2.44	5.95	-14.53	35.45
29	Tepeyanco	57.16	-3.52	12.39	-43.61	153.52
30	Terrenate	67.31	6.63	43.96	291.43	1932.21
31	Tetla	71.32	10.64	113.21	1204.55	12816.41
32	Tetlatlahuca	71.34	10.66	113.64	1211.36	12913.05
33	Tlaxcala	64.17	3.49	12.18	42.51	148.35
34	Tlaxco	76.77	16.09	258.89	4165.51	67023.05
35	Tocatlán	73.95	13.27	176.09	2336.75	31008.71
36	Totolac	66.98	6.30	39.69	250.05	1575.30
37	Trinidad de SS.	77.54	16.86	284.26	4792.62	80803.52
38	Tzonpantepec	72.47	11.79	139.00	1638.86	19322.14
39	Xaloztoc	44.80	-15.88	252.17	-4004.53	63591.93
40	Xaltican	58.75	-1.93	3.72	-7.19	13.87
41	Xicohtencatl	50.18	-10.50	110.25	-1157.63	12155.06
42	Xicotzingo	49.89	-10.79	116.42	-1256.22	13554.57
43	Yauhquemecan	70.54	9.86	97.22	958.59	9451.65
44	Zacatelco	53.24	-7.44	55.35	-411.83	3064.02
		2,669.82	- 0.10	8,813.48	32,153.71	4,811,421.11
		60.68				

FUENTE: Elaboración del autor.

Los resultados de los momentos son los siguientes:

$$m_1 = \frac{\sum x}{N} = \frac{0}{100} = 0$$

$$m_2 = \frac{\sum x^2}{N} = \frac{8813.48}{100} = 200.31$$

$$m_3 = \frac{\sum x^3}{N} = \frac{32153.71}{100} = 730.77$$

$$m_4 = \frac{\sum x^4}{N} = \frac{4811421.11}{100} = 109350.48$$

A continuación se presentan las fórmulas, cálculos y resultados de asimetría y curtosis de los datos:

$$B_1 = \frac{m_3}{\sqrt{m_2{}^3}}$$ Asimetría

$$B_2 = \frac{m_4}{m_2{}^2}$$ Curtosis

Asimetría.

$$B_1 = \frac{m_3}{\sqrt{m_2{}^3}} = \frac{730.77}{\sqrt{200.31^3}}$$

$$B_1 = \frac{m_3}{\sqrt{m_2{}^3}} = .257$$

Curtosis.

$$B_2 = \frac{m_4}{m_2{}^2} = \frac{109350.48}{200.31^2}$$

$$B_2 = \frac{m_4}{m_2{}^2} = 2.72$$

Ejemplo 5

Del ejemplo 2 se ha construido el cuadro 20 de distribución de frecuencias:

Cuadro 20
Estatura en estudiantes universitarios
por marcas de clase (Cm.)

I de C	Frecuencia
155-159	3
160-164	9
165-169	17
170-174	22
175-179	53
180-184	23
185-189	20
190-194	3
	150

FUENTE: Elaboración del autor.

Es necesario establecer el supuesto de que todas las frecuencias contenidas en cada marca de clase se encuentran en la parte media de esa marca de clase, entonces es útil la obtención de los puntos medios de las marcas de clase:

Cuadro 21
Estatura en estudiantes universitarios
por marcas de clase y puntos medios (Cm.)

I de C	Frecuencia	mi
155-159	3	157
160-164	9	162
165-169	17	167
170-174	22	172
175-179	53	177
180-184	23	182
185-189	20	187
190-194	3	192
	150	

FUENTE: Elaboración del autor.

Entonces, se puede suponer que hay tres valores 157; hay 9 observaciones de 162; 17 observaciones de 167; etcétera, de manera que pueden obtenerse, mediante el supuesto establecido, los valores desagregados, para así proceder a obtener la asimetría y curtosis:

Cuadro 21-a

Estatura en estudiantes universitarios por marcas de clase y puntos medios (cm)

		*	**		
	x_i	x	x^2	x^3	x^4
1	157	- 19.23	369.79	- 7,111.12	136,746.79
2	157	- 19.23	369.79	- 7,111.12	136,746.79
3	157	- 19.23	369.79	- 7,111.12	136,746.79
4	162	- 14.23	202.49	- 2,881.47	41,003.37
5	162	- 14.23	202.49	- 2,881.47	41,003.37
6	162	- 14.23	202.49	- 2,881.47	41,003.37
7	162	- 14.23	202.49	- 2,881.47	41,003.37
8	162	- 14.23	202.49	- 2,881.47	41,003.37
9	162	- 14.23	202.49	- 2,881.47	41,003.37
10	162	- 14.23	202.49	- 2,881.47	41,003.37
11	162	- 14.23	202.49	- 2,881.47	41,003.37
12	162	- 14.23	202.49	- 2,881.47	41,003.37
13	167	- 9.23	85.19	- 786.33	7,257.83
14	167	- 9.23	85.19	- 786.33	7,257.83
15	167	- 9.23	85.19	- 786.33	7,257.83
16	167	- 9.23	85.19	- 786.33	7,257.83
17	167	- 9.23	85.19	- 786.33	7,257.83
18	167	- 9.23	85.19	- 786.33	7,257.83
19	167	- 9.23	85.19	- 786.33	7,257.83
20	167	- 9.23	85.19	- 786.33	7,257.83
21	167	- 9.23	85.19	- 786.33	7,257.83
22	167	- 9.23	85.19	- 786.33	7,257.83
23	167	- 9.23	85.19	- 786.33	7,257.83
24	167	- 9.23	85.19	- 786.33	7,257.83
25	167	- 9.23	85.19	- 786.33	7,257.83
26	167	- 9.23	85.19	- 786.33	7,257.83
27	167	- 9.23	85.19	- 786.33	7,257.83
28	167	- 9.23	85.19	- 786.33	7,257.83
29	167	- 9.23	85.19	- 786.33	7,257.83
30	172	- 4.23	17.89	- 75.69	320.16
31	172	- 4.23	17.89	- 75.69	320.16
32	172	- 4.23	17.89	- 75.69	320.16
33	172	- 4.23	17.89	- 75.69	320.16
34	172	- 4.23	17.89	- 75.69	320.16
35	172	- 4.23	17.89	- 75.69	320.16
36	172	- 4.23	17.89	- 75.69	320.16
37	172	- 4.23	17.89	- 75.69	320.16
38	172	- 4.23	17.89	- 75.69	320.16
39	172	- 4.23	17.89	- 75.69	320.16
40	172	- 4.23	17.89	- 75.69	320.16
41	172	- 4.23	17.89	- 75.69	320.16
42	172	- 4.23	17.89	- 75.69	320.16
43	172	- 4.23	17.89	- 75.69	320.16
44	172	- 4.23	17.89	- 75.69	320.16
45	172	- 4.23	17.89	- 75.69	320.16
46	172	- 4.23	17.89	- 75.69	320.16
47	172	- 4.23	17.89	- 75.69	320.16
48	172	- 4.23	17.89	- 75.69	320.16
49	172	- 4.23	17.89	- 75.69	320.16
50	172	- 4.23	17.89	- 75.69	320.16
51	172	- 4.23	17.89	- 75.69	320.16
52	177	0.77	0.59	0.46	0.35
53	177	0.77	0.59	0.46	0.35
54	177	0.77	0.59	0.46	0.35
55	177	0.77	0.59	0.46	0.35
56	177	0.77	0.59	0.46	0.35
57	177	0.77	0.59	0.46	0.35
58	177	0.77	0.59	0.46	0.35
59	177	0.77	0.59	0.46	0.35
60	177	0.77	0.59	0.46	0.35
61	177	0.77	0.59	0.46	0.35

62	177	0.77	0.59	0.46	0.35
63	177	0.77	0.59	0.46	0.35
64	177	0.77	0.59	0.46	0.35
65	177	0.77	0.59	0.46	0.35
66	177	0.77	0.59	0.46	0.35
67	177	0.77	0.59	0.46	0.35
68	177	0.77	0.59	0.46	0.35
69	177	0.77	0.59	0.46	0.35
70	177	0.77	0.59	0.46	0.35
71	177	0.77	0.59	0.46	0.35
72	177	0.77	0.59	0.46	0.35
73	177	0.77	0.59	0.46	0.35
74	177	0.77	0.59	0.46	0.35
75	177	0.77	0.59	0.46	0.35
76	177	0.77	0.59	0.46	0.35
77	177	0.77	0.59	0.46	0.35
78	177	0.77	0.59	0.46	0.35
79	177	0.77	0.59	0.46	0.35
80	177	0.77	0.59	0.46	0.35

81	177	0.77	0.59	0.46	0.35
82	177	0.77	0.59	0.46	0.35
83	177	0.77	0.59	0.46	0.35
84	177	0.77	0.59	0.46	0.35
85	177	0.77	0.59	0.46	0.35
86	177	0.77	0.59	0.46	0.35
87	177	0.77	0.59	0.46	0.35
88	177	0.77	0.59	0.46	0.35
89	177	0.77	0.59	0.46	0.35
90	177	0.77	0.59	0.46	0.35
91	177	0.77	0.59	0.46	0.35
92	177	0.77	0.59	0.46	0.35
93	177	0.77	0.59	0.46	0.35
94	177	0.77	0.59	0.46	0.35
95	177	0.77	0.59	0.46	0.35
96	177	0.77	0.59	0.46	0.35
97	177	0.77	0.59	0.46	0.35
98	177	0.77	0.59	0.46	0.35
99	177	0.77	0.59	0.46	0.35
100	177	0.77	0.59	0.46	0.35
101	177	0.77	0.59	0.46	0.35
102	177	0.77	0.59	0.46	0.35
103	177	0.77	0.59	0.46	0.35
104	177	0.77	0.59	0.46	0.35
105	182	5.77	33.29	192.10	1,108.42
106	182	5.77	33.29	192.10	1,108.42
107	182	5.77	33.29	192.10	1,108.42
108	182	5.77	33.29	192.10	1,108.42
109	182	5.77	33.29	192.10	1,108.42
110	182	5.77	33.29	192.10	1,108.42
111	182	5.77	33.29	192.10	1,108.42
112	182	5.77	33.29	192.10	1,108.42
113	182	5.77	33.29	192.10	1,108.42
114	182	5.77	33.29	192.10	1,108.42
115	182	5.77	33.29	192.10	1,108.42
116	182	5.77	33.29	192.10	1,108.42
117	182	5.77	33.29	192.10	1,108.42
118	182	5.77	33.29	192.10	1,108.42
119	182	5.77	33.29	192.10	1,108.42
120	182	5.77	33.29	192.10	1,108.42
121	182	5.77	33.29	192.10	1,108.42
122	182	5.77	33.29	192.10	1,108.42
123	182	5.77	33.29	192.10	1,108.42
124	182	5.77	33.29	192.10	1,108.42
125	182	5.77	33.29	192.10	1,108.42
126	182	5.77	33.29	192.10	1,108.42

127	182	5.77	33.29	192.10	1,108.42
128	187	10.77	115.99	1,249.24	13,454.35
129	187	10.77	115.99	1,249.24	13,454.35
130	187	10.77	115.99	1,249.24	13,454.35
131	187	10.77	115.99	1,249.24	13,454.35
132	187	10.77	115.99	1,249.24	13,454.35
133	187	10.77	115.99	1,249.24	13,454.35
134	187	10.77	115.99	1,249.24	13,454.35
135	187	10.77	115.99	1,249.24	13,454.35
136	187	10.77	115.99	1,249.24	13,454.35
137	187	10.77	115.99	1,249.24	13,454.35
138	187	10.77	115.99	1,249.24	13,454.35
139	187	10.77	115.99	1,249.24	13,454.35
140	187	10.77	115.99	1,249.24	13,454.35
141	187	10.77	115.99	1,249.24	13,454.35
142	187	10.77	115.99	1,249.24	13,454.35
143	187	10.77	115.99	1,249.24	13,454.35
144	187	10.77	115.99	1,249.24	13,454.35
145	187	10.77	115.99	1,249.24	13,454.35
146	187	10.77	115.99	1,249.24	13,454.35
147	187	10.77	115.99	1,249.24	13,454.35
148	192	15.77	248.69	3,921.89	61,848.16
149	192	15.77	248.69	3,921.89	61,848.16
150	192	15.77	248.69	3,921.89	61,848.16
SUMA	26,435.00	0.50	8,636.83	- 21,106.32	1,389,841.04
MEDIA	176.23				

FUENTE: Elaboración del autor.

Las fórmulas para el cálculo de los momentos son las siguientes:

$$m_1 = \frac{\sum x}{N} \qquad m_2 = \frac{\sum x^2}{N}$$

$$m_3 = \frac{\sum x^3}{N} \qquad m_4 = \frac{\sum x^4}{N}$$

Los resultados de los momentos son los siguientes:

$$m_1 = \frac{\sum x}{N} = \frac{0}{150} = 0$$

$$m_2 = \frac{\sum x^2}{N} = \frac{8636.83}{150} = 57.58$$

$$m_3 = \frac{\sum x^3}{N} = \frac{-21106.32}{150} = -140.71$$

$$m_4 = \frac{\sum x^4}{N} = \frac{1389841.04}{150} = 9265.61$$

A continuación se presentan las fórmulas, cálculos y resultados de asimetría y curtosis de los datos:

$$B_1 = \frac{m_3}{\sqrt{m_2^{\ 3}}} \qquad \text{Asimetría}$$

$$B_2 = \frac{m_4}{m_2^{\ 2}} \qquad \text{Curtosis}$$

Asimetría.

$$B_1 = \frac{m_3}{\sqrt{m_2^{\ 3}}} = \frac{-140.71}{\sqrt{57.58^3}}$$

$$B_1 = \frac{m_3}{\sqrt{m_2^{\ 3}}} = -.3220$$

Curtosis.

$$B_2 = \frac{m_4}{m_2^{\ 2}} = \frac{9265.61}{57.58^2}$$

$$B_2 = \frac{m_4}{m_2^{\ 2}} = 2.79$$

Como puede verse, la asimetría con los datos desagregados fue de -.2529, en tanto que, con datos agrupados es de -.3220.

La curtosis con datos desagregados fue de 2.64 y, con datos agrupados, de 2.79.

Por lo tanto, puede verse que hay algunas diferencias, no obstante, son pequeñas y se deben a que al agregar datos se están perdiendo detalles.

1.4.2 Puntuaciones Z

Un elemento descriptivo son las puntuaciones Z, ellas son transformaciones que se pueden hacer a los valores o puntuaciones obtenidas, con el propósito de analizar su distancia respecto a la media, en unidades de desviación estándar.[3]

Una puntuación Z indica la dirección y grado en que un valor individual obtenido se aleja de la media, en una escala de unidades de desviación estándar. La medición ayuda a estandarizar la escala de una variable medida en un nivel por intervalos.

La fórmula para su cálculo es la siguiente:

$$Z = \frac{X - \overline{X}}{DS}$$

[3] Puede consultarse Hernández Sampieri, Roberto E. A. Metodología de la investigación. McGraw-Hill, México, 2000.

Donde X es la puntuación o valor a transformar, $\overline{X}$ es la media de la distribución; DS es la desviación estándar de la distribución. El resultado Z es la puntuación transformada en unidades de desviación estándar.

Supóngase que en una distribución de frecuencias se obtuvo una media de 30 y una desviación estándar de 5, y se desea comparar a una puntuación de 25 con el resto de la distribución.

Se transforma la puntuación o valor en una puntuación Z:

X = 25
$\overline{X}$ = 30
DS = 5

La puntuación Z correspondiente a un valor de 25 es:

$$Z = \frac{X - \overline{X}}{DS} = \frac{25 - 30}{5} = -1.00$$

Puede decirse que el valor 25 está localizado a una desviación estándar por debajo de la media de la distribución.

El valor 15 está situado a 3 desviaciones estándar por debajo de la media:

$$Z = \frac{X - \overline{X}}{DS} = \frac{15 - 30}{5} = -3.00$$

Estadarizar los valores permite comparar puntuaciones de dos distribuciones diferentes (la forma de medición es la misma, pero se trata de distribuciones distintas).

1.4.3 Las razones o proporciones

Razón es una comparación, la comparación es proporción.

Una razón es la relación entre dos categorías. Por ejemplo:

Categorías	Frecuencias
Hombres	230
Mujeres	115

La razón de hombres a mujeres es de:

$$H/M = \frac{No.\,de\,H}{No.\,de\,M}$$

Sustituyendo:

$$H/M = \frac{No.\,de\,H}{No.\,de\,M} = \frac{230}{115} = 2$$

Es decir, por cada dos hombres hay una mujer.

1.4.4 Las tasas

Las tasas, proviene del verbo tasar, que indica medir.

Una tasa es la relación entre el indicador de una variable, perteneciente a una categoría y el número total de observaciones, multiplicada generalmente por 100 o 1000.

La fórmula para su cálculo es la siguiente:

$$TASA = \frac{No.\ del\ indicador\ en\ un\ período}{No.\ total\ de\ eventos\ posibles} * 100$$

Ejemplo:

Obtenga la tasa del número de nacidos vivos en la ciudad, en el número de habitantes en la ciudad.

$$TASA = \frac{No.\ de\ eventos\ de\ un\ período}{No.\ total\ de\ eventos\ posibles} * 100 = \frac{No.\ de\ nacidos\ vivos\ en\ la\ ciudad}{No.\ de\ habitantes\ en\ la\ ciudad} * 100$$

$$TASA = \frac{No.\ de\ nacidos\ vivos\ en\ la\ ciudad}{No.\ de\ habitantes\ en\ la\ ciudad} * 100 = \frac{9874}{301000} * 100 = 3.28$$

Es decir, hay 3.28 nacidos vivos por cada 100 habitantes en la ciudad.

1.4.5 Índice de Gini

La fórmula para su cálculo es la siguiente:[4]

$$IG = \frac{\sum x_i (Y_{i+1}) - \sum (x_{i+1}) y_i}{10000}$$

Las literales “x” e “y” representan los porcentajes acumulados de población y de ingreso.

x_i significa el valor x en el tiempo i
x_{i+1} significa el valor de x en el tiempo i más un periodo
y_i significa el valor y en el tiempo i
y_{i+1} significa el valor de y en el tiempo i más un periodo.

Los valores posibles del índice fluctúan entre 0 y 1, donde valores cercanos a 0 indican igualdad y, valores cercanos a 1 indican desigualdad.

[4] Puede consultarse: Holguín Quiñones, Fernando. *Estadística descriptiva aplicada a las ciencias sociales*. UNAM, México, 1981.

Ejemplo 1

Los siguientes datos hipotéticos corresponden a salarios percibidos por 862 000 familias, en cuatro marcas de clase.

Cuadro 22

Ingreso por familias (salarios mínimos)

	No. de familias	% de Ingreso
PEA < 1 sm	378,000	0.17
1 sm < PEA < 2 sm	262,000	0.38
2 sm =< PEA < 5 sm	178,000	0.27
PEA => 5 sm	44,000	0.18
	862,000	1.00

FUENTE: Elaboración del autor, datos hipotéticos.

Elaborando cálculos:

Cuadro 23

Ingreso por familias (salarios mínimos)

	No. de familias	% de Ingreso	% de familias	% de Ingreso	% acum. de fam.	% acum. de ingreso			MULT. 7*8			MULT. 10*11
PEA < 1 sm	378,000	0.17	43.85	17.00	43.85	17.00		17.00		43.85		
1 sm < PEA < 2 sm	262,000	0.38	30.39	38.30	74.25	55.30	43.85	55.30	2,424.90	74.25	17.00	1,262.18
2 sm =< PEA < 5 sm	178,000	0.27	20.65	26.55	94.90	81.84	74.25	81.84	6,076.51	94.90	55.30	5,247.52
PEA => 5 sm	44,000	0.18	5.10	18.03	100.00	99.88	94.90	99.88	9,477.93	100.00	81.84	8,184.30
	862,000	1.00	100	99.88			100.00		17,979.33		99.88	14,694.00

FUENTE: Elaboración del autor.

I.G.	0.33

El indicador es 0.33

Ejemplo 2

Obtenga las curvas de Lorenz y los índices de Gini para los dos cuadros siguientes (24 y 25):

Cuadro 24
Escuela Nacional de
Enfermería, UNAM *

% acumulados	
Familias	Ingresos
26.27	9.82
48.08	23.41
81.91	57.13
91.20	71.02
97.03	84.08
98.39	88.50
99.01	91.20
99.50	94.29
99.63	95.37
100.00	100.00

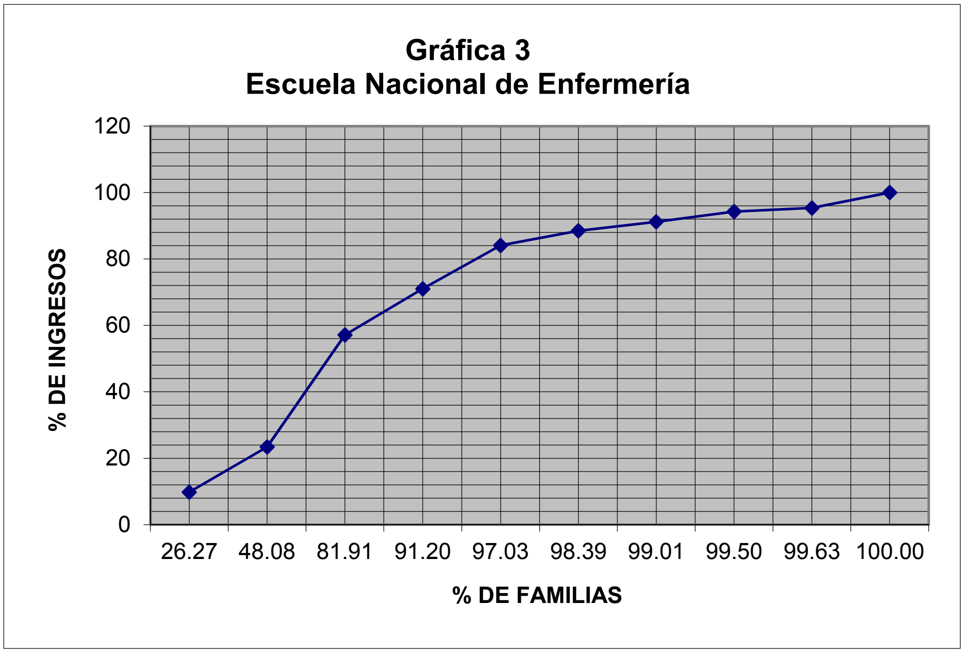
* Datos para estudiantes
activos en 1974
FUENTE: Elaboración del autor
a partir de Holguín, Estadística, 272.

Cuadro 25
Facultad de Sociología, UNAM *

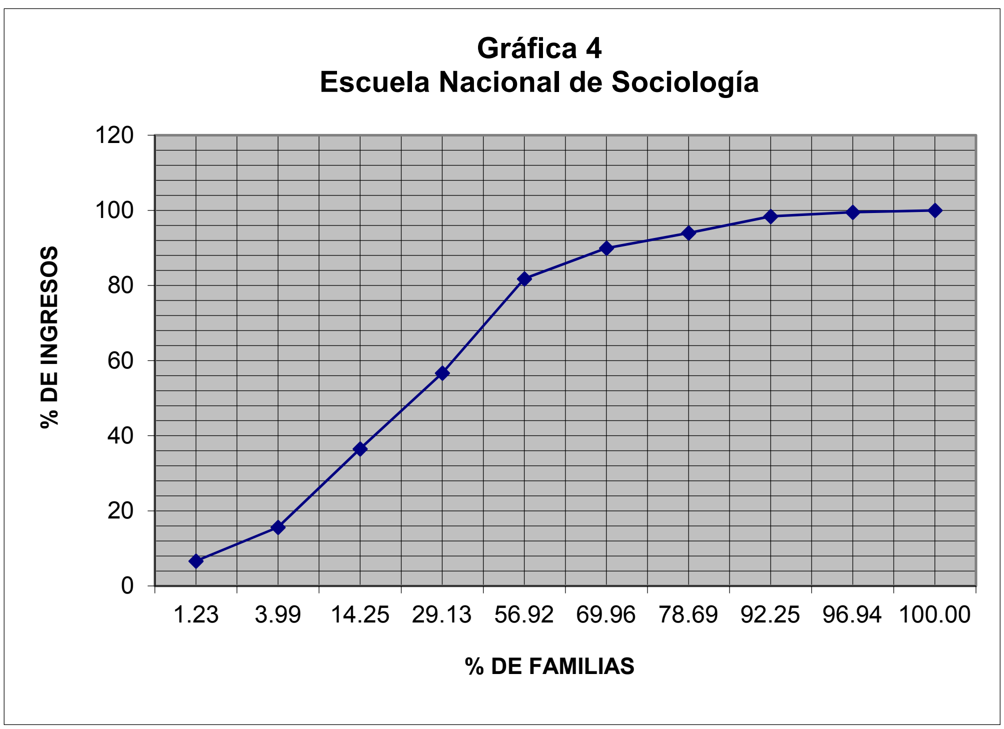
% acumulados	
Familias	Ingresos
6.66	1.23
15.63	3.99
36.50	14.25
56.67	29.13
81.79	56.92
89.94	69.96
94.00	78.69
98.41	92.25
99.50	96.94
100.00	100.00

* Datos para estudiantes
activos en 1974
FUENTE: Elaboración del autor
a partir de Holguín, Estadística, 272.

Las gráficas son las siguientes:



FUENTE: Elaboración del autor.



FUENTE: Elaboración del autor.

Tablas de resultados:

CUADRO 26

ESCUELA NACIONAL DE ENFERMERÍA

Punto medio (de las clases de ing	Frecuencia	Ingreso del grupo (miles $)	Porcentajes Familias	Porcentajes Ingresos	% acumulados Familias	% acumulados Ingresos	(Xi)	(yi+1)		(xi+1)	(yi)	
750	212	159.00	26.27	9.82	26.27	9.82		9.82		26.27		
1,250	176	220.00	21.81	13.59	48.08	23.41	26.27	23.41	614.90	48.08	9.82	472.12
2,000	273	546.00	33.83	33.72	81.91	57.13	48.08	57.13	2746.63	81.91	23.41	1917.20
3,000	75	225.00	9.29	13.90	91.20	71.02	81.91	71.02	5817.35	91.20	57.13	5210.09
4,500	47	211.50	5.82	13.06	97.03	84.08	91.20	84.08	7668.69	97.03	71.02	6891.05
6,500	11	71.50	1.36	4.42	98.39	88.50	97.03	88.50	8586.85	98.39	84.08	8273.02
8,750	5	43.70	0.62	2.70	99.01	91.20	98.39	91.20	8973.02	99.01	88.50	8762.32
12,500	4	50.00	0.50	3.09	99.50	94.29	99.01	94.29	9335.26	99.50	91.20	9074.73
17,500	1	17.50	0.12	1.08	99.63	95.37	99.50	95.37	9489.54	99.63	94.29	9393.68
25,000	3	75.00	0.37	4.63	100.00	100.00	99.63	100.00	9962.83	100.00	95.37	9536.81
	807	1619.20	100.00	100.00			100.00				100.00	
									63195.06			**59531.02**
								CG	0.37			

ESCUELA NACIONAL DE SOCIOLOGÍA

Punto medio (de las clases de ing	Frecuencia	Ingreso del grupo (miles $)	Porcentajes Familias	Porcentajes Ingresos	% acumulados Familias	% acumulados Ingresos	(Xi)	(yi+1)		(xi+1)	(yi)	
750	361	271	6.66	1.23	6.66	1.23		1.23		6.66		
1,250	486	608	8.97	2.76	15.63	3.99	6.66	3.99	26.57	15.63	1.23	19.22
2,000	1,131	2,262	20.87	10.26	36.50	14.25	15.63	14.25	222.78	36.50	3.99	145.59
3,000	1,093	3,279	20.17	14.88	56.67	29.13	36.50	29.13	1,063.38	56.67	14.25	807.75
4,500	1,361	6,124	25.12	27.79	81.79	56.92	56.67	56.92	3,225.85	81.79	29.13	2,382.67
6,500	442	2,873	8.16	13.04	89.94	69.96	81.79	69.96	5,721.74	89.94	56.92	5,119.76
8,750	220	1,925	4.06	8.74	94.00	78.69	89.94	78.69	7,078.04	94.00	69.96	6,576.38
12,500	239	2,988	4.41	13.56	98.41	92.25	94.00	92.25	8,672.11	98.41	78.69	7,744.60
17,500	59	1,032	1.09	4.68	99.50	96.94	98.41	96.94	9,539.86	99.50	92.25	9,179.43
25,000	27	675	0.50	3.06	100.00	100.00	99.50	100.00	9,950.18	100.00	96.94	9,693.70
	5,419	**22,037**	**100.00**	**100.00**			100.00				100.00	
									45,500.51			**41,669.11**
								CG	**0.38**			

FUENTE: Elaboración del autor.

Ejemplo 3

Considere los datos del cuadro 27.

Cuadro 27

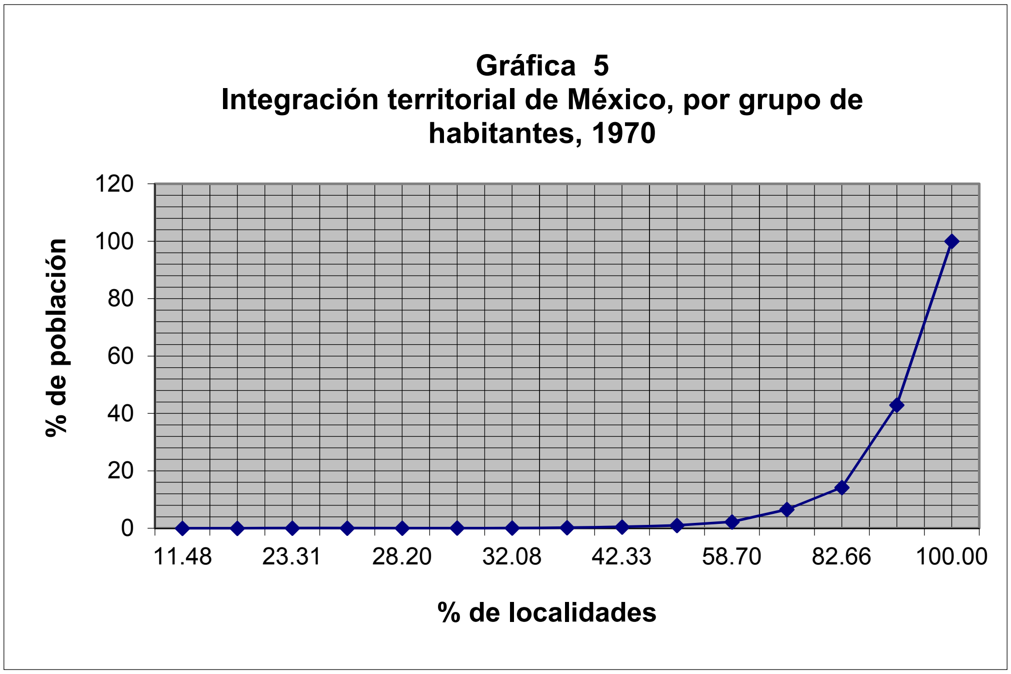
Integración territorial de México por gpo. de HBTS. 1970 *

	Localidades	Población (miles)
500000 y más	4	5,535
250000-499999	6	1,972
100000-249999	24	3,735
75000-99999	13	1,114
50000-74999	21	1,242
40000-49999	19	858
30000-39999	30	1,016
20000-29999	65	1,532
10000-19999	248	3,410
5000-9999	539	3,764
2500-4999	1,201	4,130
1000-2499	4,232	6,366
500-999	7,473	5,190
100-499	28,055	6,889
1 a 99	55,650	1,471
	97,580	48,224

* Muestra a partir de la ENYGH

FUENTE: Elaboración con datos de la ENYGH 1970.

La gráfica correspondiente es:



FUENTE: Elaboración con datos de la ENYGH 1970.

La obtención del indicador:

Cuadro 28

Integración territorial de México por grupo de habitantes. 1970*

	1	2	3	4	5	6	7	8	9	12	13	14
	X	Y	X	Y	X	Y	X	Y	(Xi)*(Yi+	X	Y	(Xi+1)*(Yi)
	Locali-dades	Población (miles)	% de loc.	% de Pob.	% Acum. Loc.	% Acum. Pob.	(Xi)	(yi+1)	MULT	(xi+1)	(yi)	MULT
500000 y más	4	5,535	0.00	11.48	0.0041	11.48		11.48		0.00		
250000-499999	6	1,972	0.01	4.09	0.01	15.57	0.00	15.57	0.06	0.01	11.48	0.12
100000-249999	24	3,735	0.02	7.75	0.03	23.31	0.01	23.31	0.24	0.03	15.57	0.54
75000-99999	13	1,114	0.01	2.31	0.05	25.62	0.03	25.62	0.89	0.05	23.31	1.12
50000-74999	21	1,242	0.02	2.58	0.07	28.20	0.05	28.20	1.36	0.07	25.62	1.79
40000-49999	19	858	0.02	1.78	0.09	29.98	0.07	29.98	2.09	0.09	28.20	2.51
30000-39999	30	1,016	0.03	2.11	0.12	32.08	0.09	32.08	2.86	0.12	29.98	3.59
20000-29999	65	1,532	0.07	3.18	0.19	35.26	0.12	35.26	4.23	0.19	32.08	5.98
10000-19999	248	3,410	0.25	7.07	0.44	42.33	0.19	42.33	7.90	0.44	35.26	15.54
5000-9999	539	3,764	0.55	7.81	0.99	50.14	0.44	50.14	22.09	0.99	42.33	42.04
2500-4999	1,201	4,130	1.23	8.56	2.22	58.70	0.99	58.70	58.29	2.22	50.14	111.50
1000-2499	4,232	6,366	4.34	13.20	6.56	71.90	2.22	71.90	159.90	6.56	58.70	385.12
500-999	7,473	5,190	7.66	10.76	14.22	82.66	6.56	82.66	542.34	14.22	71.90	1022.38
100-499	28,055	6,889	28.75	14.29	42.97	96.95	14.22	96.95	1378.54	42.97	82.66	3552.07
1 a 99	55,650	1,471	57.03	3.05	100.00	100.00	42.97	100.00	4296.99	100.00	96.95	9694.97
	97,580	48,224	100.0000	100.00			100.00				100.00	
									6477.77			14839.27
								CG	-0.8361			

* Muestra a partir de la ENYGH

FUENTE: ENYGH 1970

Ejemplo 4

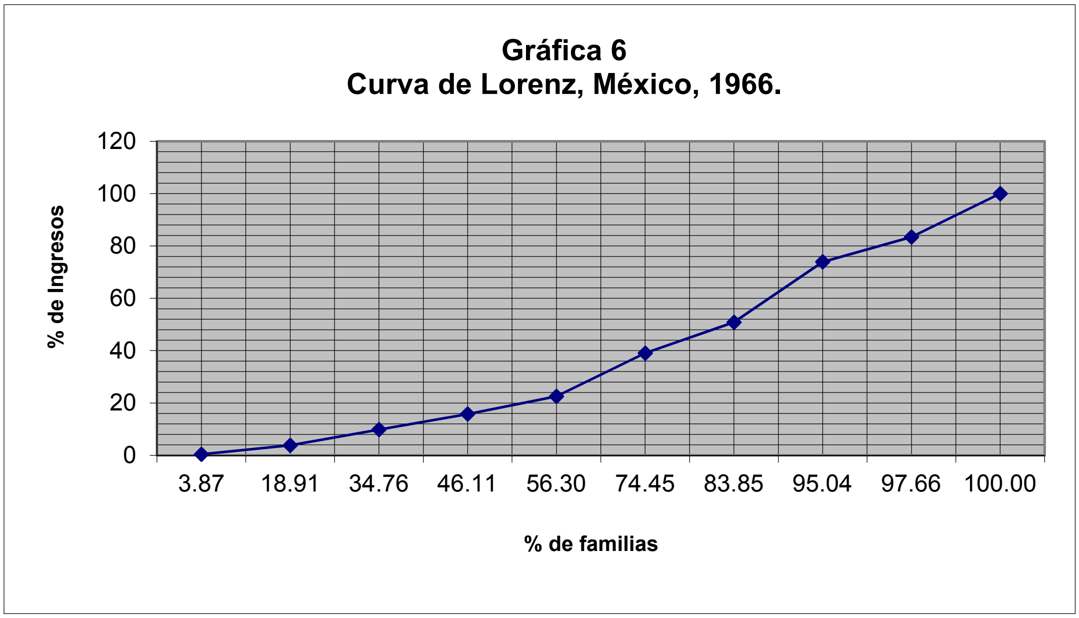
Observe los datos del cuadro 29.

Cuadro 29
Distribución del ingreso mensual familiar en México,
encuesta por muestreo (1966)

1	2	3
X	Y	X
Ingreso	Número de familias	Ingreso medio *
Menos de 100	223,411	72.20
101-200	869,602	157.79
201-300	916,060	263.35
301-400	655,904	361.78
401-500	588,552	459.47
501-750	1,049,112	629.47
751-1000	543,131	871.16
1 001-2 000	646,968	1,426.80
2 001-3 000	151,688	2,512.49
3 001-y más	134,998	4,918.28
suma	5,779,426	

* El ingreso medio se computó como un promedio de las unidades
de la muestra correspondiente a cada clase
FUENTE: INEGI.

La gráfica correspondiente es la 6:



FUENTE: INEGI.

Con los datos anteriores, el cálculo del coeficiente de Gini se muestra en el cuadro 30.

Cuadro 30

Distribución del ingreso mensual familiar en México, encuesta por muestreo (1966)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	Y	X	2 * 3 = 4	4 / 1000	Y	X	Y	X	X	Y	(Xi)*(Yi+1)	X	Y	(Xi+1)*(Yi)
Ingreso	Número de familias	Ingreso medio *	Ingreso del grupo	Ingreso del grupo (miles)	% de familias	% de Ingresos	% acum. de fam.	% acum. de Ingresos			MULT.			MULT.
Menos de 100	223,411	72.20	16,130,274.20	16,130.27	3.87	0.40	3.87	0.40		0.40		3.87		
101-200	869,602	157.79	137,214,499.58	137,214.50	15.05	3.43	18.91	3.83	3.87	3.83	14.80	18.91	0.40	7.62
201-300	916,060	263.35	241,244,401.00	241,244.40	15.85	6.03	34.76	9.85	18.91	9.85	186.38	34.76	3.83	133.13
301-400	655,904	361.78	237,292,949.12	237,292.95	11.35	5.93	46.11	15.78	34.76	15.78	548.60	46.11	9.85	454.42
401-500	588,552	459.47	270,421,987.44	270,421.99	10.18	6.75	56.30	22.54	46.11	22.54	1,039.12	56.30	15.78	888.40
501-750	1,049,112	629.47	660,384,530.64	660,384.53	18.15	16.49	74.45	39.03	56.30	39.03	2,197.09	74.45	22.54	1,677.68
751-1000	543,131	871.16	473,154,001.96	473,154.00	9.40	11.82	83.85	50.85	74.45	50.85	3,785.29	83.85	39.03	3,272.32
1 001-2 000	646,968	1,426.80	923,093,942.40	923,093.94	11.19	23.05	95.04	73.90	83.85	73.90	6,196.10	95.04	50.85	4,832.29
2 001-3 000	151,688	2,512.49	381,114,583.12	381,114.58	2.62	9.52	97.66	83.42	95.04	83.42	7,927.97	97.66	73.90	7,217.32
3 001-y más	134,998	4,918.28	663,957,963.44	663,957.96	2.34	16.58	100.00	100.00	97.66	100.00	9,766.41	100.00	83.42	8,341.76
suma	5,779,426		4,004,009,132.90	4,004,009.13	100.00	100.00			100.00				100.00	
											31661.77			26824.95

* El ingreso medio se computó como un promedio de las unidades de la muestra correspondiente a cada clase

FUENTE:INEGI

COEF. DE GINI	0.484

2. Estadística inferencial

2.1 Introducción

Falsificar hipótesis, en la idea de Popper, es el equivalente a efectuar la prueba de hipótesis. En ella se plantean hipótesis, las cuales son de muchos tipos y deben revisarse junto con la intencionalidad de ellas en distintos autores; es necesario tomar como referente un nivel de significancia que, en una tabla de distribución de puntuaciones Z, toma valores específicos; asimismo, en el capítulo se expone la idea de distribución muestral. Con el nivel de significancia y la distribución muestral se puede comprender lo que posteriormente se expone: la técnica de probar hipótesis, planteadas a partir de estadígrafos y llevándola a la generalización de los parámetros.[5]

En su última obra, Popper habla de que la teoría tradicional de la probabilidad es capaz de estudiar y dar explicación probabilística a experimentos como la moneda al aire o el dado tirado a la mesa, pero cuando el dado está cargado se muestra la impotencia de la probabilidad tradicional. Algunos, apunta Popper, dicen que es ahí donde termina la certidumbre y empieza lo incierto, donde termina lo cuantitativo e inicia lo cualitativo, ahí es el terreno de la libertad y el libre albedrío, pero esa es una suerte de “interpretación subjetiva de la probabilidad”, él en cambio propone una “interpretación objetiva de la probabilidad”.[6] Si se hace un experimento en repetidas ocasiones, puede ser con el dado cargado, entonces se puede tener una distribución de frecuencias de tal manera que se puede establecer que hay “algo” que está provocando que exista una propensión a que una cara del dado esté cayendo en una proporción mayor a lo que las condiciones del azar esperarían. Por supuesto, Popper nunca dice que ese “algo” sea inherente al objeto de estudio, en este caso al dado, las condiciones del ambiente del sistema estudiado pueden también condicionar las propensiones; por ejemplo, una superficie de mármol es distinta en sus efectos sobre la distribución de frecuencias que una superficie acolchonada y hasta corrugada. Entonces se hace necesario distinguir entre el cálculo absoluto y el condicional o relativo de la probabilidad:

> Podemos establecer un enunciado del cálculo absoluto del siguiente modo:
>
> $$p\ (a) = r$$
>
> a saber: la probabilidad del evento a es igual a r... a diferencia del enunciado de probabilidad relativa o condicional
>
> $$p\ (a, b) = r$$

[5] Una versión anterior de este capítulo aparece publicado en Sánchez Espinoza, Francisco. “La prueba de hipótesis”, en *Diké Revista del Centro de Investigaciones Jurídico Políticas.* BUAP, Puebla, marzo de 2013, pp. 145-166.

[6] Popper, Karl. *Un mundo de propensiones*. Tecnos, Madrid, España, 1992.

> a saber: la probabilidad del evento a en la situación b (o dadas las condiciones b) es igual a r.[7]

Ahora bien, la cuestión es cómo diseñar investigaciones de tal manera que la enseñanza de Popper sea aprovechada efectivamente. En muchos libros de Estadística aplicada en ciencias sociales incluso la prueba de hipótesis es tomada como “estadística inferencial”, se trata de la técnica que ayuda a considerar la generalización del estadígrafo al parámetro de la muestra a la población.

2.2. Las hipótesis

En toda investigación uno de los pasos que se sigue es el planteamiento de hipótesis, a veces se plantea incluso un conjunto de hipótesis secundarias, es necesario entonces citar algunas concepciones del término:

> Hipótesis es una proposición enunciada para responder tentativamente a un problema. Proposición es un conjunto de palabras que expresan un sujeto y sus atributos gramaticales, relacionados entre sí por un verbo. El adverbio tentativamente dice que proponemos la respuesta sin saber aún si las observaciones, hechos o datos la comprobarán o disprobarán.[8]

Es respuesta a la pregunta que constituye el problema de investigación. Si bien es una respuesta es tentativa. Además:

> Dentro de la investigación científica, las hipótesis son proposiciones tentativas acerca de las relaciones entre dos o más variables y se apoyan en conocimientos organizados y sistematizados.[9]

Es decir, se obtienen de la literatura correspondiente al tema. Además, las hipótesis deben cumplir con algunas características, distintos autores han señalado, en común, las siguientes:

1. Se refieren a situación o situaciones reales.
2. Los términos serán comprensibles y concretos.
3. La relación entre variables debe ser lógica.
4. Los términos y las relaciones deben ser observables y medibles.
5. Las relaciones entre términos deben ser susceptibles de ser tratadas con técnicas y procedimientos al alcance del investigador.[10]

Puede haber, desde luego, hipótesis pertenecientes a trabajos que no busquen falsificaciones, pero si lo son, entonces se refieren a la realidad.

La mayoría de los estudios buscan volver operacionales los conceptos, entonces volver explícitas las variables que al medirse forman indicadores que pueden ser descritos o analizados mediante herramientas diversas, por ello lo que se da en llamar “términos” en las hipótesis deben ser comprensibles y concretos.

En este caso se entiende que las variables deben estar relacionadas en un cuerpo teórico.

---

[7] *Ibidem*, p. 36.
[8] Pardinas, *Metodología*, p. 151.
[9] Hernández, *Metodología*, p. 230.
[10] Pueden tomarse a manera de ejemplo los textos, antes señalados: de Pardinas, en su capítulo 6, y; de Hernández, *Metodología*, capítulo 5.

En investigaciones teóricas puede darse cuenta de otro tipo de términos y relaciones, pero en hipótesis relacionadas a la falsificación deben ser términos susceptibles de operacionalizarse y mediante variables volverse sujetos de cuantificación mediante indicadores.

El último punto pone en relieve que sea posible para el investigador obtener los datos, procesarlos y analizarlos con lo que tiene al alcance en sus conocimientos y recursos en general.

Por otra parte, pueden hallarse distintos tipos de hipótesis, que atienden a distintas preocupaciones de los redactores de las mismas, una primera tipología es la siguiente:

- *Descriptivas*. Se trata de afirmaciones univariables, aunque desde luego siempre está referida a una constante. Es decir solamente una variable es la que cambia en sus indicadores, pero está referida a alguna población, grupo, periodo, etc., que no cambia, por lo tanto solamente se busca describir el comportamiento de lo que sí varía y no relacionar variables que están cambiando.
- *Correlacionales*. Especifican la relación y el tipo de ella entre dos o más variables. Puede haber una o más variables explicadas y lo mismo en las explicatorias, cuando son dos variables las que se relacionan se trata de una correlación bivariada y, cuando son más de dos es una correlación multivariada.
- *Diferencia de grupos*. Establece comparaciones entre grupos. Dicha comparación se hace respecto a una o más variables. Es decir, si la variable se mueve en su indicador, entonces se esperan reacciones en los grupos de manera diferenciada.
- *Causales*. Establece relaciones entre variables en términos de causa-efecto. No solamente establece relaciones sino que también el sentido de la relación en términos de dirección. En esta hipótesis hay correlación, la causa antecede al efecto y los cambios en la causa deben provocar cambios en el efecto.[11]

Otra tipología de hipótesis es la siguiente:

- Reversible. Si X entonces Y, si Y entonces X. Es decir, una variable causa a la otra y viceversa, hay bidireccionalidad causal.
- Irreversible. Si X entonces Y, si Y no se puede afirmar nada sobre X. Se reconoce solamente la unidireccionalidad, sin desconocer el otro sentido.
- Determinista. Si X siempre Y. Es contundente, si una variable aparece la otra invariablemente lo hará, hay un solo nivel de medición presencia-ausencia y es 0 %, no estar, o 100 %, aparecer.
- Estocástica. Si X probablemente Y. De X se sigue Y, pero estableciendo una probabilidad, por lo tanto, se tiene la ventaja de establecer también los casos de error.
- Suficiente. Si X entonces Y. Ya que basta con que suceda X para que Y aparezca.
- Contingente. Si X entonces Y, pero solamente que aparezca Z. Es decir, se requiere la condición de que suceda X, pero eso no basta, ya que se requiere que suceda Z.
- Necesaria. Si X, y sólo si X, entonces Y. Es la única forma en la que se establece contundentemente la relación y la clara dependencia, la no necesidad de otras variables, como Z, sin embargo, el reto es demostrarla en la investigación.
- De variable sustituible. Si X entonces Y, también si Z entonces Y. X detona a Y, pero Y también puede ser detonada por Z.
- Interdependiente. Si X cambia de X_1 a X_2 y X_2 es igual a X_1 más un incremento de X, entonces Y cambiará de Y_1 A Y_2 y Y_2 será igual a Y_1 más un incremento de Y;

[11] Véase capítulo 5 de Hernández, *Metodología*.

pero si Y cambia de Y_1 a Y_2 y Y_2 es igual a Y_1 más un incremento de Y, entonces X cambia de X_1 a X_2 y X_2 será igual a X_1 más un incremento de X.[12]

Estas dos series de tipos de hipótesis son representativas y suficientes para notar que son intenciones diferentes en ambos casos. El primero, el de las hipótesis descriptivas, correlacionales, diferencia de grupos y causales, viene de un texto amplio, que expone ideas acerca de la metodología y técnica de la investigación en general, dedica un amplísimo capítulo al "Análisis de los datos" en el cual detallan las técnicas estadísticas que pueden utilizarse en la investigación, por ello es marcada la intención de exponer los tipos de hipótesis en función de esas técnicas que veremos en capítulos posteriores. En el caso de las hipótesis reversible e irreversible; determinista y estocástica; suficiente, contingente y, necesaria; y, de variable sustituible e interdependiente, tiene la intención de ir en términos "crecientes" hacia el establecimiento de la causación. El primer par muestra la diferencia entre la reversivilidad o no; el segundo par va del determinismo hacia la probabilidad; el trío de hipótesis que sigue muestra la suficiencia, contingencia o necesidad de las variables; el último par va en busca del establecimiento de interdependencia mediante proporciones o cantidades que implican una relación de carácter causal que, incluso podría establecerse mediante una formulación que contenga una relación precisa de variación entre las variables, unívoca e irreversible.

Todas ellas son hipótesis que pueden plantearse en una investigación, pero son necesarias otro tipo de hipótesis en la investigación. Las hipótesis nulas están referidas siempre respecto de las hipótesis de investigación, y son la negación de las hipótesis; en tanto que las hipótesis alternativas constituyen otras, respecto de las principales, posibles explicaciones a las preguntas de investigación. A manera de ejemplo, puede decirse como hipótesis "la puerta es café"; como hipótesis nula "la puerta no es café"; y, como hipótesis alternativa "la puerta es azul".

2.3 Los conceptos

Muchas investigaciones deben hacerse sin considerar a la población o universo de estudio, es decir, no se obtienen datos para cada una de las entidades de estudio, solamente se considera una muestra de la población, y es que hay poblaciones de estudio tan grandes que cualquier investigador e incluso grupo de investigadores se verían impedidos en abordarlas.

El dato que proviene de una población se denomina *parámetro* y el que se deriva a partir de una muestra se llama *estadígrafo*. Por ejemplo, la media de las edades de los ciudadanos de un país es un parámetro; la media de las edades de una muestra de ciudadanos de un país es un estadígrafo.

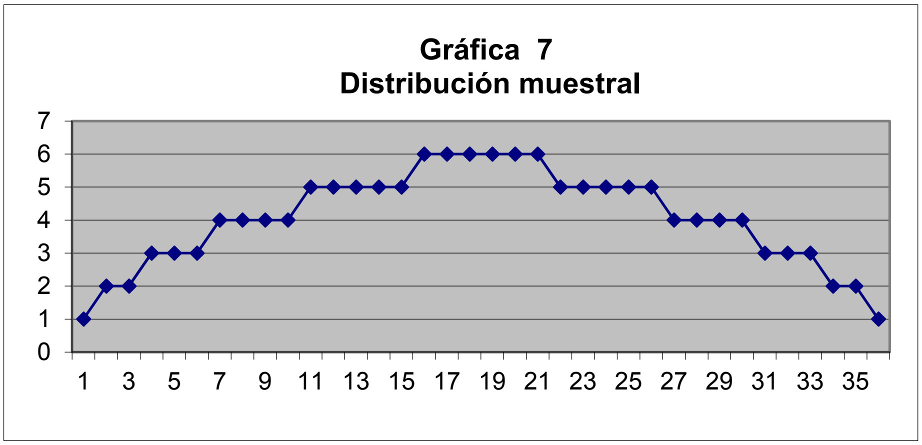
La inferencia estadística es un procedimiento mediante el cual se busca generalizar los datos y, por lo tanto, los análisis de ellos, de una muestra al total de la población, en otras palabras, equiparar los estadígrafos a los parámetros.

La hipótesis, de cualquier tipo que sea, debe plantearse considerando a la población en conjunto, y entonces la prueba de hipótesis de la estadística inferencial consiste en verificar si la hipótesis es congruente con los datos de la muestra, en cuyo caso se acepta la generalización del estadígrafo a la población, en caso contrario se rechaza la hipótesis, pero no quiere esto decir que se desechen los datos muestrales.

Una distribución de los datos de una muestra contiene los datos obtenidos mediante una muestra, por ejemplo las edades de 1600 ciudadanos con credencial para votar, la media de los datos podría ser de 51 años; si se tomaran todas las

[12] Véase capítulo 6 de Pardinas, *Metodología*.

muestras posibles de 1600 entidades de estudio, ciudadanos, de entre la población de 76 millones, entonces se podría obtener una serie de medias, una para cada muestra posible, si esas medias se disponen en un eje bidimensional, entonces el resultado es una distribución de medias o distribución muestral, desde luego, unas medias serán mayores a otras:



FUENTE: Elaboración del autor.

Si se calculara la media de todas las medias de las muestras, se obtendría el valor de la media poblacional, sin embargo, esa medición es más bien de carácter teórico, ya que lo que en realidad se llega a calcular es la media de una de las muestras posibles y la cuestión es si el estadígrafo calculado es cercano, o no, al parámetro.

Al respecto puede citarse el teorema central del límite:

> El teorema central del límite dice que si tenemos un grupo numeroso de variables independientes y todas ellas siguen el mismo modelo de distribución (cualesquiera que éste sea), la suma de ellas se distribuye según una distribución normal.[13]

Por ejemplo, si se tira una moneda al aire cien veces, y a la cara se le da valor de 1 y al escudo valor de 0, entonces el conjunto de los eventos se distribuirá según una distribución normal.

Asimismo, puede citarse:

> Si una población, no necesariamente normal, tiene una media m y una desviación estándar de s, la distribución de las medias en el muestreo aleatorio realizado en esta población tiende, al aumentar n, a una distribución normal de media m y de desviación estándar $s/\sqrt{n}$, donde n es el tamaño de la muestra.[14]

Es decir, cuando crece el tamaño de la muestra, n, entonces crece la probabilidad de que la media como estadígrafo se acerque a la media como parámetro.

La estructura de la probabilidad puede ser tomada como el área bajo una curva y, como la probabilidad de ocurrencia de un evento va de 0 a 1, donde 0 es que no ocurra y 1 que ocurra, entonces el área entre dos puntos es la probabilidad de la

[13] http://www.aulafacil.com/CursoEstadistica/Lecc-38-est.htm (Consulta: 12 de septiembre de 2012).
[14] Hernández, *Metodología*, p. 368.

distribución, y el investigador buscará evaluar si es alta o baja la probabilidad de que el estadígrafo esté cerca del parámetro. En el caso de que sea alta, se asume como correcto el generalizar; si es baja se asume dudoso que sea correcto generalizar de los estadígrafos a los parámetros.

Entonces, se obtienen datos a partir de una muestra y se trata de establecer con un porcentaje de confianza, y un porcentaje de error, la probabilidad de que sea correcto generalizar los resultados y el análisis de los estadígrafos a los parámetros. Aun cuando se haya realizado el muestreo bien, y aunque el tratamiento estadístico de muestreo trata de asegurar cercanía entre estadígrafos y parámetros, la cercanía puede no ser real, no realizarse, y esto puede deberse a errores en el proceso de obtención de estadígrafos, es decir, puede haber errores en alguna de las fases de recolección de datos.

Generalmente se utilizan dos niveles de significancia en la prueba de hipótesis: .05, es decir 95 % de confianza para generalizar, con 5 % de probabilidades de equivocarse al generalizar, y; .01, es decir 99 % de probabilidades de generalizar sin cometer error, con 1 % de probabilidades de errar al generalizar.

Tanto el nivel de significancia como la distribución muestral son áreas bajo una curva. El nivel de significancia se toma como un área bajo la distribución muestral, representando áreas de riesgo y confianza al generalizar.

2.4 El procedimiento para prueba de hipótesis

Deben seguirse los siguientes siete pasos:

1. A partir del marco teórico o estado del arte se establece una hipótesis que se constituye en el parámetro poblacional.

Por ejemplo, “Los potenciales votantes en el estado de Puebla ven, en promedio, 2.8 horas de televisión al día.”

2. Se escoge un nivel de significancia.
En este caso escogeremos $\alpha = .05$.

3. Obtención de los estadígrafos.
Por ejemplo, a partir de una encuesta, cuyo cuestionario se aplicó a 1600 personas con credencial para votar, con un 2.5 como porcentaje aceptado de error y, un 95.5 % como índice de confianza, en condiciones de máxima varianza, se obtuvo que en promedio las personas encuestadas ven 2.7 horas diarias de televisión, con una desviación estándar de 0.9 horas.

4. Con la siguiente fórmula se calcula la desviación estándar de la distribución muestral de la media:

$$S\bar{x} = \frac{s}{\sqrt{n}}$$

Con los datos del ejemplo:

$$S\bar{x} = \frac{s}{\sqrt{n}} = \frac{0.9}{\sqrt{1600}} = .0225$$

5. Se transforma la media de la muestra en una puntuación Z, es decir a unidades de desviación estándar, pero tomando en cuenta que se trata de una distribución muestral. La fórmula para la obtención de puntuaciones Z es la siguiente:

$$Z = \frac{X - \overline{X}}{DS}$$

Una observación particular menos la media de los datos, dividido el resultado y, por lo tanto, puesta en términos de desviación estándar.

Como en este caso no se trata de una observación, sino de la media de la muestra, entonces la fórmula se adecúa de la siguiente manera:

$$Z = \frac{X - \overline{X}}{S\overline{X}}$$

Es decir, la media de la muestra menos la media que se plantea como hipótesis o parámetro, dividida esa resta entre la desviación estándar de la distribución muestral de medias.

Entonces, para nuestro ejemplo:

$$Z = \frac{X - \overline{X}}{S\overline{X}} = \frac{2.7 - 2.8}{.0225} = -4.44$$

6. Se escoge un nivel de significancia de 95 %. Entonces es necesario recurrir a una tabla de áreas bajo la curva (ver tabla al final del capítulo). Debe buscarse el valor 0.025, ya que en la cuarta columna se muestra el "Área de la parte menor", es decir la cola de la derecha del área bajo la curva y, dado el nivel de significancia escogido, 2.5 queda por la derecha y 2.5 por la izquierda. En la primera columna se muestra la "Puntuación Z" correspondiente, en este caso a 1.96. La segunda columna muestra la "Distancia de Z a la media" y, la tercera, el área bajo la curva desde el inicio de la distribución, es decir el "Área de la parte mayor".

7. Se compara la media de la muestra transformada a puntuación Z con el valor en tabla, se aplica el criterio: si Z es menor a tablas se acepta H; si Z es mayor a Tablas se rechaza H.

En el ejemplo, Z = 4.44 y; Tabla = 1.96; entonces se rechaza H.

La decisión de rechazar se efectúa teniendo un 95 % a favor y, 5 % de riesgo de cometer un error.

Para manejar otro ejemplo, considérese que la media de la muestra hubiera sido de 2.76 horas, entonces los nuevos cálculos serían los siguientes:

$$Z = \frac{X - \overline{X}}{S\overline{X}} = \frac{2.76 - 2.8}{.0225} = -1.77$$

La media de la muestra estaría más cercana a la media hipotética y la hipótesis se aceptaría con una probabilidad de acierto de 95 %.

También es posible construir un intervalo en el que se localice el parámetro, dado un nivel de confianza. En otras palabras, se calcula con una probabilidad determinada el que un parámetro se localice en un determinado intervalo.

La fórmula es la siguiente:

Intervalo de confianza = Estadígrafo o ± (Puntuación que expresa el nivel de confianza elegido) (Desviación estándar de la distribución muestral correspondiente)

O bien:

I de C = Estadígrafo o ± (puntuación Z) (DS)

El estadígrafo, desde luego, proviene de la muestra, la puntuación Z proviene de las tablas (para nivel de 95, Z es de 1.96; para 99, Z es 2.58), el error estándar depende del estadígrafo.

Con el ejemplo que se manejó desde el principio:

Media = 2.7 horas
S = 0.9 horas
$S\bar{x}$ = 0.0225 (Es la ds de la distrib. muestral de la media)
Nivel de confianza = .95 (Z=1.96)
Intervalo de confianza = 2.7 +- (1.96) (0.0225)
= 2.7 +- 0.0441
= 2.7441 y 2.6559

La media poblacional se encuentra entre 2.74 y 2.65 horas, con 95 % de probabilidades de no cometer error.

Vamos a manejar un ejemplo adicional.

1. A partir de la consideración de una encuesta realizada en noviembre de 2008 en el estado de Tlaxcala, en la cual se hizo la pregunta:

"En política la gente habla a menudo de la izquierda y la derecha. En una escala de 0 a 10, en la que cero significa izquierda y 10 derecha, ¿Dónde se ubicaría usted?"

Las respuestas fueron recibidas oralmente del entrevistado y anotadas por el aplicador de la encuesta en 11 números con un círculo pequeño al lado del mismo, el cual debía tacharse.

La fórmula que se utilizó para el muestreo fue la correspondiente a poblaciones infinitas, considerando un coeficiente de confianza de 95.5 % y un error de muestreo de 5 %:

$$n = \frac{4\,pq}{E^2}$$

$$n = \frac{4 * 50 * 50}{5^2}$$

$$n = \frac{10{,}000}{25}$$

$$n = 400$$

Por lo tanto, se trataron de levantar 400 cuestionarios, más 40 adicionales para reemplazos, sin embargo, los cuestionarios válidos al final del procedimiento de vaciado de datos fueron 410.

El cálculo de la media con los datos desagregados fue el siguiente:

$$Media = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$Media = \frac{1,796}{410}$$

$$Media = 4.38$$

Presentados en una tabla de frecuencias los datos son los siguientes:

Cuadro 31

Tabla de frecuencias

Intervalo de Clase	Frecuencia Fi
0	85
1	23
2	16
3	35
4	3
5	97
6	32
7	56
8	27
9	21
10	15
SUMA	410

FUENTE: Elaboración del autor.

Si a partir de esa tabla (cuadro 31) se calcula la media, el procedimiento se sigue así:

Cuadro 32

Cálculo de la media

Intervalo de Clase	Frecuencia Fi	mi	mifi
0	85	0	-
1	23	1	23
2	16	2	32
3	35	3	105
4	3	4	12
5	97	5	485
6	32	6	192
7	56	7	392
8	27	8	216
9	21	9	189
10	15	10	150
SUMA	410		1,796
			3,225,616

FUENTE: Elaboración del autor.

$$Media = \frac{\sum_{i=1}^{k} m_i f_i}{n}$$

$$Media = \frac{1,796}{410}$$

$$Media = 4.38$$

Entonces, la media para datos agregados o desagregados es, en ambos casos, de 4.38.

El cálculo de la varianza para datos desagregados es la siguiente:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$S^2 = \frac{3,832.6439}{410-1}$$

$$S^2 = \frac{3,832.6439}{409}$$

$$S^2 = 9.3707$$

Por lo tanto, la desviación estándar es de:

$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

$$S = \sqrt{\frac{3832.6439}{410-1}}$$

$$S = \sqrt{\frac{3832.6439}{409}}$$

$$S = \sqrt{9.3707}$$

$$S = 3.06$$

Ahora el cálculo de la varianza para datos agregados es, según el cuadro 33:

Cuadro 33

Cálculo de la varianza

Intervalo de Clase	Frecuencia Fi	mi	mifi	mi2fi
0	85	0	-	-
1	23	1	23	23
2	16	2	32	64
3	35	3	105	315
4	3	4	12	48
5	97	5	485	2,425
6	32	6	192	1,152
7	56	7	392	2,744
8	27	8	216	1,728
9	21	9	189	1,701
10	15	10	150	1,500
SUMA	410		1,796	11,700
			3,225,616	

FUENTE: Elaboración del autor.

La fórmula y despejes:

$$S^2 = \frac{n\sum_{i=1}^{k} m_i^2 f_i - (\sum_{i=1}^{k} m_i f_i)^2}{n(n-1)}$$

$$S^2 = \frac{(410 * 11,700) - (1,796)^2}{410\,(409)}$$

$$S^2 = \frac{4,797,000 - 3,225,616}{167,690}$$

$$S^2 = \frac{1,571,384}{167,690}$$

$$S^2 = 9.370$$

La desviación estándar de esos mismos datos desagregados:

$$S = \sqrt{9.370}$$

$$S = 3.06$$

Entonces, tanto para datos desagregados como para agregados la desviación estándar es de 3.06.

Recapitulando lo expuesto, a la vez que se constituye la hipótesis de trabajo:

Con base en esa encuesta, se establece que los mayores de dieciocho años en el estado de Tlaxcala se autodefinen, en la escala propuesta de izquierda y derecha política, en promedio como de 4.38.

2. Se escoge un nivel de significancia.

En este caso escogeremos $\alpha = .05$

3. Obtención de los estadígrafos.
Se realizó una réplica de la encuesta anteriormente enunciada pero un año después, es decir en noviembre de 2009. En este caso se aplicó el mismo muestreo para poblaciones infinitas, mayores a 100 000 entidades en estudio, siendo el resultado que con un error de muestreo aceptado de 6 % y un nivel de confianza de 95.5 %:

$$n = \frac{4\, pq}{E^{2}}$$

$$n = \frac{4 * 50 * 50}{6^{2}}$$

$$n = \frac{10,000}{36}$$

$$n = 278$$

Y esos 278 fueron los cuestionarios validados después de vaciar los datos.

El cálculo de la media con los datos desagregados fue el siguiente:

$$Media = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$Media = \frac{1,139}{278}$$

$$Media = 4.097$$

Los datos se presentan en una tabla de frecuencias (cuadro 34).

Cuadro 34

Tabla de frecuencias

Intervalo de Clase	Frecuencia Fi
0	65
1	14
2	19
3	20
4	17
5	43
6	26
7	28
8	27
9	14
10	5
SUMA	278

FUENTE: Elaboración del autor.

Si a partir de esa tabla se calcula la media, el procedimiento se sigue como se muestra en el cuadro 35.

Cuadro 35

Cálculo de la media

Intervalo de Clase	Frecuencia Fi	mi	mifi
0	65	0	0
1	14	1	14
2	19	2	38
3	20	3	60
4	17	4	68
5	43	5	215
6	26	6	156
7	28	7	196
8	27	8	216
9	14	9	126
10	5	10	50
	278		1,139
			1,297,321

FUENTE: Elaboración del autor.

$$Media = \frac{\sum_{i=1}^{k} m_i f_i}{n}$$

$$Media = \frac{1,139}{278}$$

$$Media = 4.09$$

Entonces, la media para datos agregados o desagregados es, en ambos casos, de 4.09.

El cálculo de la varianza para datos desagregados es la siguiente:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

$$S^2 = \frac{2,620.3777}{278-1}$$

$$S^2 = \frac{2,620.3777}{277}$$

$$S^2 = 9.4598$$

Por lo tanto, la desviación estándar es de:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$S = \sqrt{\frac{2,620.3777}{278-1}}$$

$$S = \sqrt{\frac{2,620.3777}{277}}$$

$$S = \sqrt{9.4598}$$

$$S = 3.0756$$

Ahora el cálculo de la varianza para datos agregados es, según el cuadro 36:

Cuadro 36

Cálculo de la media

Intervalo de Clase	Frecuencia Fi	mi	mifi	mi2fi
0	65	0	0	0
1	14	1	14	14
2	19	2	38	76
3	20	3	60	180
4	17	4	68	272
5	43	5	215	1,075
6	26	6	156	936
7	28	7	196	1,372
8	27	8	216	1,728
9	14	9	126	1,134
10	5	10	50	500
	278		1,139	7,287
			1,297,321	

FUENTE: Elaboración del autor.

La fórmula y despejes:

$$S^2 = \frac{n\sum_{i=1}^{k} m_i^{\,2} f_i - (\sum_{i=1}^{k} m_i f_i)^2}{n(n-1)}$$

$$S^2 = \frac{(278 * 7,287) - (1,139)^2}{278\,(277)}$$

$$S^2 = \frac{2,025,786 - 1,297,321}{77,006}$$

$$S^2 = \frac{728,465}{77,006}$$

$$S^2 = 9.4598$$

La desviación estándar de esos mismos datos desagregados:

$$S = \sqrt{9.4598}$$

$$S = 3.0756$$

Entonces, tanto para datos desagregados como para agregados la desviación estándar es de 3.0756.

Es decir, en esta nueva encuesta, la media es de 4.09 y la desviación estándar de 3.0756.

Los datos para ambas encuestas son:

Cuadro 37

Datos para ambas encuestas

	Primera	Segunda
n	410	278
Media	4.38	4.09
D.S.	3.06	3.07
I. de Confianza	95.5	95.5
Error de Muestreo	5	6

FUENTE: Elaboración del autor.

4. Con la siguiente fórmula se calcula la desviación estándar de la distribución muestral de la media:

$$S\bar{x} = \frac{s}{\sqrt{n}}$$

Con los datos del ejemplo:

$$S\bar{x} = \frac{3.07}{\sqrt{278}}$$

$$S\bar{x} = \frac{3.07}{16.6733}$$

$$S\bar{x} = .184$$

5. Se transforma la media de la muestra en una puntuación Z, es decir a unidades de desviación estándar, pero tomando en cuenta que se trata de una distribución muestral.

Para nuestro ejemplo:

$$Z = \frac{X - \overline{X}}{S\overline{X}}$$

$$Z = \frac{4.09 - 4.38}{.184}$$

$$Z = \frac{-0.29}{.184}$$

$$Z = -1.576$$

6. Se escoge un nivel de significancia de 95 %. Entonces el valor en la tabla es de 1.96.

7. Se compara la media de la muestra transformada a puntuación Z con el valor en Tabla, se aplica el criterio: si Z es menor a Tablas se acepta H; si Z es mayor a Tablas se rechaza H.
Se acepta la hipótesis.

La decisión de aceptar se efectúa teniendo un 95 % a favor y 5 % de riesgo de cometer un error.

Para manejar otro ejemplo, considérese que la media de la muestra hubiera sido de 5.0, entonces los nuevos cálculos serían los siguientes:

$$Z = \frac{X - \overline{X}}{S\overline{X}}$$

$$Z = \frac{5 - 4.38}{.184}$$

$$Z = \frac{0.62}{.184}$$

$$Z = 3.37$$

El valor Z es mayor al de la tabla y entonces la hipótesis se rechaza.

Los resultados posibles, al plantear hipótesis y llevar a cabo el procedimiento de probarlas, son:

1. Aceptar una hipótesis verdadera (decisión correcta).
2. Rechazar una hipótesis falsa (decisión correcta).
3. Aceptar una hipótesis falsa (error tipo II o Beta).
4. Rechazar una hipótesis verdadera (error tipo I o Alfa).

Dispuesto lo anterior en un cuadro, éste queda de la siguiente manera:

Cuadro 38
Decisiones correctas y tipos
error en las hipótesis

	Hipótesis Verdadera	Hipótesis Falsa
Aceptar	Decisión correcta	Error Tipo II o Beta
Rechazar	Error Tipo I o Alfa	Decisión correcta

FUENTE: Elaboración del autor.

Los errores pueden hacerse menos probables si se pone atención en algunos aspectos, que anteceden a la prueba de hipótesis y, por lo tanto, impactan su resultado:

- Conocer a la población en estudio. Entonces puede preverse su crecimiento, cambios de comportamiento, las relaciones con variables internas y externas a su sistema, etc.
- Debe buscarse que el muestreo sea el apropiado, prefiriendo el método que permita la generalización de datos y no los métodos que no sean estadísticos, los cuestionarios deben ser probados antes de aplicarse, y se debe cuidar la asignación de cuotas haciéndolo con el método apropiado.
- Hay una correspondencia entre los problemas de investigación, los objetivos, hipótesis, los datos, y a ellos deben atender las pruebas estadísticas que se utilicen.

Anexo A del capítulo 2

ÁREAS BAJO LA CURVA NORMAL			
COLUMNAS			
(I) Puntuación "Z"	(II) Distancia de "Z" a la media	(III) Área de la parte mayor	(IV) Área de la parte menor
0.00	0.0000	0.5000	0.5000
0.01	0.0040	0.5040	0.4960
0.02	0.0080	0.508	0.4920
0.03	0.0120	0.5120	0.4880
0.04	0.0160	0.5160	0.4840
0.05	0.0199	0.5199	0.4801
0.06	0.0239	0.5239	0.4761
0.07	0.0279	0.5279	0.4721
0.08	0.0319	0.5319	0.4681
0.09	0.0359	0.5359	0.4641
0.10	0.0398	0.5398	0.4602
0.11	0.0438	0.5438	0.4562
0.12	0.0478	0.5478	0.4522
0.13	0.0517	0.5517	0.4483
0.14	0.0557	0.5557	0.4443
0.15	0.0596	0.5596	0.4404
0.16	0.0636	0.5636	0.4364
0.17	0.0675	0.5675	0.4325
0.18	0.0714	0.5714	0.4286
0.19	0.0753	0.5753	0.4247
0.20	0.0793	0.5793	0.4207
0.21	0.0832	0.5832	0.4168
0.22	0.0871	0.5871	0.4129
0.23	0.0910	0.5910	0.4090
0.24	0.0948	0.5948	0.4052

0.25	0.0987	0.5987	0.4013
0.26	0.1026	0.6026	0.3974
0.27	0.1064	0.6064	0.3936
0.28	0.1103	0.6103	0.3897
0.29	0.1141	0.6141	0.3859
0.30	0.1179	0.6179	0.3821
0.31	0.1217	0.6217	0.3783
0.32	0.1255	0.6255	0.3745
0.33	0.1293	0.6293	0.3707
0.34	0.1331	0.6331	0.3669
0.35	0.1368	0.6368	0.3632
0.36	0.1406	0.6406	0.3594
0.37	0.1443	0.6443	0.3557
0.38	0.1480	0.6480	0.3520
0.39	0.1517	0.6517	0.3483
0.40	0.1554	0.6554	0.3446
0.41	0.1591	0.6591	0.3409
0.42	0.1628	0.6628	0.3372
0.43	0.1664	0.6664	0.3336
0.44	0.1700	0.6700	0.3300
0.45	0.1736	0.6736	0.3264
0.46	0.1772	0.6772	0.3228
0.47	0.1808	0.6808	0.3192
0.48	0.1844	0.6844	0.3156
0.49	0.1879	0.6879	0.3121
0.50	0.1915	0.6915	0.3085
0.51	0.1950	0.6950	0.3050
0.52	0.1985	0.6985	0.3015
0.53	0.2019	0.7019	0.2981
0.54	0.2054	0.7	0.2946
0.55	0.2088	0.7088	0.2912
0.56	0.2123	0.7123	0.2877
0.57	0.2157	0.7153	0.2843
0.58	0.2190	0.7190	0.2810
0.59	0.2224	0.7224	0.2776

0.60	0.2257	0.7257	0.2743
0.61	0.2291	0.7291	0.2709
0.62	0.2324	0.7324	0.2676
0.63	0.2357	0.7357	0.2643
0.64	0.2389	0.7389	0.2611
0.65	0.2422	0.7422	0.2578
0.66	0.2424	0.7454	0.2546
0.67	0.2486	0.7486	0.2514
0.68	0.2517	0.7517	0.2483
0.69	0.2549	0.7549	0.2451
0.70	0.2580	0.7580	0.2420
0.71	0.2611	0.7611	0.2389
0.72	0.2642	0.7642	0.2358
0.73	0.2673	0.7673	0.2327
0.74	0.2704	0.7704	0.2296
0.75	0.2734	0.7734	0.2266
0.76	0.2764	0.7706	0.2236
0.77	0.2794	0.7794	0.2206
0.78	0.2823	0.7823	0.2177
0.79	0.2852	0.7852	0.2148
0.80	0.2881	0.7881	0.2119
0.81	0.2910	0.7910	0.2090
0.82	0.2939	0.7939	0.2061
0.83	0.2967	0.7967	0.2033
0.84	0.2995	0.7995	0.2005
0.85	0.3023	0.8023	0.1977
0.86	0.3051	0.8051	0.1949
0.87	0.3078	0.8078	0.1922
0.88	0.3106	0.8106	0.1894
0.89	0.3133	0.8133	0.1867
0.90	0.3159	0.8159	0.1841
0.91	0.3186	0.8186	0.1814
0.92	0.3212	0.8212	0.1788
0.93	0.3228	0.8238	0.1762
0.94	0.3264	0.8264	0.1736

0.95	0.3289	0.8289	0.1711
0.96	0.3315	0.8315	0.1685
0.97	0.3340	0.8340	0.1660
0.98	0.3365	0.8365	0.1635
0.99	0.3389	0.8389	0.1611
			.
1.00	0.3413	0.8413	0.1587
1.01	0.3438	0.8438	0.1562
1.02	0.3461	0.8461	0.1539
1.03	0.3485	0.8485	0.1515
1.04	0.3508	0.8508	0.1492
1.05	0.3531	0.8531	0.1469
1.06	0.3554	0.8554	0.1446
1.07	0.3577	0.8577	0.1423
1.08	0.3599	0.8599	0.1401
1.09	0.3621	0.8621	0.1379
1.10	0.3643	0.8643	0.1357
1.11	0.3665	0.8665	0.1335
1.12	0.3686	0.8686	0.1314
1.13	0.3708	0.8708	0.1292
1.14	0.3729	0.8729	0.1271
1.15	0.3749	0.8749	0.1251
1.16	0.3770	0.8770	0.1230
1.17	0.3790	0.8790	0.1210
1.18	0.3810	0.8810	0.1190
1.19	0.3830	0.8830	0.1170
1.20	0.3849	0.8849	0.1151
1.21	0.3869	0.8869	0.1131
1.22	0.3888	0.8888	0.1112
1.23	0.3907	0.8907	0.1093
1.24	0.3925	0.8925	0.1075
1.25	0.3944	0.8944	0.1056
1.26	0.3962	0.8962	0.1038
1.27	0.3980	0.8980	0.1020
1.28	0.3997	0.8997	0.1003
1.29	0.4015	0.9015	0.0985

1.30	0.4032	0.9032	0.0968
1.31	0.4049	0.9049	0.0951
1.32	0.4066	0.9066	0.0934
1.33	0.4082	0.9082	0.0918
1.34	0.4099	0.9099	0.0901
1.35	0.4115	0.9115	0.0885
1.36	0.4131	0.9131	0.0869
1.37	0.4147	0.9147	0.0853
1.38	0.4162	0.9162	0.0838
1.39	0.4177	0.9177	0.0823
1.40	0.4192	0.9192	0.0808
1.41	0.4207	0.9207	0.0793
1.42	0.4222	0.9222	0.0778
1.43	0.4236	0.9236	0.0764
1.44	0.4251	0.9251	0.0749
1.45	0.4265	0.9265	0.0735
1.46	0.4279	0.9279	0.0721
1.47	0.4242	0.9292	0.0708
1.48	0.4306	0.9306	0.0694
1.49	0.4319	0.9319	0.0681
1.50	0.4332	0.9332	0.0668
1.51	0.4345	0.9345	0.0655
1.52	0.4357	0.9357	0.0643
1.53	0.4370	0.9370	0.0630
1.54	0.4382	0.9382	0.0618
1.55	0.4394	0.9394	0.0606
1.56	0.4406	0.9406	0.0594
1.57	0.4418	0.9418	0.0582
1.58	0.4429	0.9429	0.0571
1.59	0.4441	0.9441	0.0559
1.60	0.4452	0.9452	0.0548
1.61	0.4463	0.9463	0.0537
1.62	0.4474	0.9474	0.0526
1.63	0.4484	0.9484	0.0516
1.64	0.4495	0.9495	0.0505

1.65	0.4505	0.9505	0.0495
1.66	0.4515	0.9515	0.0485
1.67	0.4525	0.9525	0.0475
1.68	0.4535	0.9535	0.0465
1.69	0.4545	0.9545	0.0455
1.70	0.4554	0.9554	0.0446
1.71	0.4564	0.9564	0.0436
1.72	0.4573	0.9573	0.0427
1.73	0.4582	0.9582	0.0418
1.74	0.4591	0.9591	0.0409
1.75	0.4599	0.9599	0.0401
1.76	0.4608	0.9608	0.0392
1.77	0.4616	0.9616	0.0384
1.78	0.4625	0.9625	0.0375
1.79	0.4633	0.9633	0.0367
		.	
1.80	0.4641	0.9641	0.0359
1.81	0.4649	0.9649	0.0351
1.82	0.4656	0.9656	0.0344
1.83	0.4664	0.9664	0.0336
1.84	0.4671	0.9671	0.0329
1.85	0.4648	0.9678	0.0322
1.86	0.4686	0.9686	0.0314
1.87	0.4693	0.9693	0.0307
1.88	0.4699	0.9699	0.0301
1.89	0.4706	0.9706	0.0294
1.90	0.4713	0.9713	0.0287
1.91	0.4719	0.9719	0.0281
1.92	0.4716	0.9726	0.0274
1.93	0.4732	0.9732	0.0268
1.94	0.4738	0.9738	0.0262
1.95	0.4744	0.9744	0.0256
1.96	0.475	0.975	0.025
1.97	0.4756	0.9756	0.0244
1.98	0.4761	0.9761	0.0239
1.99	0.4767	0.9767	0.0233

2.00	0.4772	0.9772	0.0228
2.01	0.4778	0.9778	0.0222
2.02	0.4783	0.9783	0.0217
2.03	0.4788	0.9788	0.0212
2.04	0.4793	0.9793	0.0207
2.05	0.4798	0.9798	0.0202
2.06	0.4803	0.9803	0.0197
2.07	0.4808	0.9808	0.0192
2.08	0.4812	0.9812	0.0188
2.09	0.4817	0.9817	0.0183
2.10	0.4821	0.9821	0.0179
2.11	0.4826	0.9826	0.0174
2.12	0.4830	0.9830	0.0170
2.13	0.4834	0.9834	0.0166
2.14	0.4838	0.9838	0.0162
2.15	0.4842	0.9842	0.0158
2.16	0.4846	0.9846	0.0154
2.17	0.4850	0.9850	0.0150
2.18	0.4854	0.9854	0.0146
2.19	0.4857	0.9857	0.0143
2.20	0.4861	0.9861	0.0139
2.21	0.4864	0.9864	0.0136
2.22	0.4868	0.9868	0.0132
2.23	0.4871	0.9871	0.0129
2.24	0.4875	0.9875	0.0125
2.25	0.4878	0.9878	0.0122
2.26	0.4881	0.9881	0.0119
2.27	0.4884	0.9884	0.0116
2.28	0.4887	0.9887	0.0113
2.29	0.4890	0.9890	0.0110
2.30	0.4893	0.9893	0.0107
2.31	0.4896	0.9896	0.0104
2.32	0.4898	0.9898	0.0102
2.33	0.4901	0.9901	0.0099
2.34	0.4904	0.9904	0.0096

2.35	0.4906	0.9906	0.0094
2.36	0.4909	0.9909	0.0091
2.37	0.4911	0.9911	0.0089
2.38	0.4913	0.9913	0.0087
2.39	0.4916	0.9916	0.0084
2.40	0.4918	0.9918	0.0082
2.41	0.4920	0.9920	0.0080
2.42	0.4922	0.9922	0.0078
2.43	0.4925	0.9925	0.0075
2.44	0.4957	0.9927	0.0073
2.45	0.4929	0.9929	0.0071
2.46	0.4931	0.9931	0.0069
2.47	0.4932	0.9932	0.0068
2.48	0.4934	0.9934	0.0066
2.49	0.4936	0.9936	0.0064
2.50	0.4938	0.9938	0.0062
2.51	0.4940	0.9940	0.0060
2.52	0.4941	0.9941	0.0059
2.53	0.4943	0.9943	0.0057
2.54	0.4945	0.9945	0.0055
2.55	0.4946	0.9946	0.0054
2.56	0.4948	0.9948	0.0052
2.57	0.4949	0.9949	0.0051
2.58	0.4951	0.9951	0.0049
2.59	0.4952	0.9952	0.0048
2.60	0.4953	0.9953	0.0047
2.61	0.4955	0.9955	0.0045
2.62	0.4956	0.9956	0.0044
2.63	0.4957	0.9957	0.0043
2.64	0.4959	0.9959	0.0041
2.65	0.4960	0.9960	0.0040
2.66	0.4961	0.9961	0.0039
2.67	0.4962	0.9962	0.0038
2.68	0.4963	0.9963	0.0037
2.69	0.4964	0.9964	0.0036

2.70	0.4965	0.9965	0.0035
2.71	0.4966	0.9966	0.0034
2.72	0.4967	0.9967	0.0033
2.73	0.4968	0.9968	0.0032
2.74	0.4969	0.9969	0.0031
2.75	0.4970	0.9970	0.0030
2.76	0.4971	0.9971	0.0029
2.77	0.4972	0.9972	0.0028
2.78	0.4973	0.9973	0.0027
2.79	0.4974	0.9974	0.0026
2.80	0.4974	0.9974	0.0026
2.81	0.4975	0.9975	0.0025
2.82	0.4976	0.9976	0.0024
2.83	0.4977	0.9977	0.0023
2.84	0.4977	0.9977	0.0023
2.85	0.4978	0.9978	0.0022
2.86	0.4979	0.9979	0.0021
2.87	0.4979	0.9979	0.0021
2.88	0.4980	0.9980	0.0020
2.89	0.4981	0.9981	0.0019
2.90	0.4981	0.9981	0.0019
2.91	0.4982	0.9982	0.0018
2.92	0.4982	0.9982	0.0018
2.93	0.4983	0.9983	0.0017
2.94	0.4984	0.9984	0.0016
2.95	0.4984	0.9984	0.0016
2.96	0.4985	0.9985	0.0015
2.97	0.4985	0.9985	0.0015
2.98	0.4986	0.9986	0.0014
2.99	0.4986	0.9986	0.0014
3.00	0.4987	0.9987	0.0013
3.01	0.4987	0.9987	0.0013
3.02	0.4987	0.9987	0.0013
3.03	0.4988	0.9988	0.0012
3.04	0.4988	0.9988	0.0012

3.05	0.4989	0.9989	0.0011
3.06	0.4989	0.9989	0.0011
3.07	0.4989	0.9989	0.0011
3.08	0.4990	0.9990	0.0010
3.09	0.4990	0.9990	0.0010
3.10	0.4990	0.9990	0.0010
3.11	0.4991	0.9991	0.0009
3.12	0.4991	0.9991	0.0009
3.13	0.4991	0.9991	0.0009
3.14	0.4992	0.9992	0.0008
3.15	0.4992	0.9992	0.0008
3.16	0.4992	0.9992	0.0008
3.17	0.4992	0.9992	0.0008
3.18	0.4993	0.9993	0.0007
3.19	0.4993	0.9993	0.0007
3.20	0.4993	0.9993	0.0007
3.21	0.4993	0.9993	0.0007
3.22	0.4994	0.9994	0.0006
3.23	0.4994	0.9994	0.0006
3.24	0.4994	0.9994	0.0006
0.3	0.4995	0.9995	0.0005
3.4	0.4997	0.9997	0.0003
3.5	0.4998	0.9998	0.0002
3.6	0.4998	0.9998	0.0002
3.7	0.4999	0.9999	0.0001

FUENTE: N.M. Downie y R. W. Hearh (1973). Metodos Estadísticos Aplicados. México, Harla, pp. 320-327.
Fuente original: A.L. Edwars. Statistical Methods for the Behavioral Sciencies. Nueva York: Holt, Rinehart and Winston, 1954.

3. ESTADÍSTICA PARAMÉTRICA

3.1 Coeficiente de correlación de Pearson

El coeficiente de correlación de Pearson mide el grado de asociación entre dos series de datos.

Los valores fluctúan entre -1 y 1.

-1 Significa que tienen una relación perfectamente inversa
1 Significa que tienen una relación perfectamente directa
0 O valores cercanos a cero indican que no hay relación entre los indicadores de esas variables.

La fórmula mediante la cual se calcula el coeficiente de correlación (r) es la siguiente:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$n =$

$\sum xy =$

$\sum x =$

$\sum y =$

$\sum x^2 =$

$\sum y^2 =$

Los resultados intermedios, ente -1 y 1, se pueden calificar así:

-0.90	Correlación negativa muy fuerte
-0.75	Correlación negativa considerable
-0.50	Correlación negativa media
-0.10	Correlación negativa débil
0.00	Ausencia de correlación
0.10	Correlación positiva débil
0.50	Correlación positiva media
0.75	Correlación positiva considerable
0.90	Correlación positiva muy fuerte.

Cuando el coeficiente de correlación se eleva al cuadrado, se obtiene la R^2, la cual se denomina varianza de factores comunes, y es el indicador del porcentaje de la variación de una variable debido a la variación de la otra variable y viceversa.

Ejemplo 1. Los resultados, como puntaje de ingreso, de universitarios se presentan en el cuadro 39, también las calificaciones de la materia A. Se espera que el puntaje de ingreso mida lo que el perfil de ingreso de la licenciatura requiere, es decir, las capacidades para desempeñarse en las materias correspondientes al plan de estudios. Por lo anterior, se espera una alta correlación.

Verifique si eso es así.

Cuadro 39

	X Puntaje de ingreso	Y Calificación en materia A	XY	X*X	Y*Y
A	100.00	8.92	892.00	10,000.00	79.57
B	95.65	10.00	956.52	9,149.34	100.00
C	95.65	9.54	912.52	9,149.34	91.01
D	91.30	8.21	749.61	8,336.48	67.40
E	82.61	9.64	796.35	6,824.20	92.93
F	95.65	5.35	511.74	9,149.34	28.62
G	78.26	8.57	670.70	6,124.76	73.44
H	82.61	8.21	678.22	6,824.20	67.40
I	82.61	6.42	530.35	6,824.20	41.22
J	100.00	9.28	928.00	10,000.00	86.12
K	100.00	8.92	892.00	10,000.00	79.57
L	86.96	8.21	713.91	7,561.44	67.40
M	91.30	8.21	749.61	8,336.48	67.40
N	86.96	8.92	775.65	7,561.44	79.57
Ñ	82.61	6.78	560.09	6,824.20	45.97
O	100.00	7.85	785.00	10,000.00	61.62
P	100.00	8.57	857.00	10,000.00	73.44
Q	91.30	6.07	554.22	8,336.48	36.84
R	100.00	9.64	964.00	10,000.00	92.93
S	100.00	8.92	892.00	10,000.00	79.57
T	91.30	8.57	782.48	8,336.48	73.44
U	100.00	7.50	750.00	10,000.00	56.25
V	95.65	6.42	614.09	9,149.34	41.22
W	82.61	4.64	383.30	6,824.20	21.53
X	82.61	5.00	413.04	6,824.20	25.00
Y	86.96	8.92	775.65	7,561.44	79.57
Z	95.65	9.64	922.09	9,149.34	92.93
AA	78.26	3.57	279.39	6,124.76	12.74
BB	82.61	-	-	6,824.20	-
CC	73.91	1.78	131.57	5,463.14	3.17
DD	82.61	10.00	826.09	6,824.20	100.00
	2,795.65	232.27	21,247.17	254,083.18	1,917.89

COEF. DE CORREL.
0.5088

FUENTE: Elaboración del autor con datos de Carter, Undergraduate.

Se trata de una correlación media.

Ejemplo 2. Los siguientes datos son de los años de 1970 a 2004 para la inflación entre los Estados Unidos y México. La estructura económica de México indica que para exportar primero debe importar, lo cual se debe a que es productor de bienes de consumo y no se especializa en bienes de capital. Entonces, puede suponerse que la inflación en los Estados Unidos se vea reflejada en México, por lo que las inflaciones entre los dos países deben tener una correlación positiva.

Verifique esto a través del tiempo.

Cuadro 40

Inflación México-USA 1970-2005

Inflación %		
PRES/AÑO	MÉXICO	U.S.A.
GDO 1970	4.69	5.72
LEA 1971	4.96	4.38
LEA 1972	5.56	3.21
LEA 1973	21.37	6.22
LEA 1974	20.60	11.04
LEA 1975	11.30	9.13
LEA 1976	27.20	5.76
JLP 1977	20.66	6.50
JLP 1978	16.17	7.59
JLP 1979	20.02	11.35
JLP 1980	29.84	13.50
JLP 1981	28.69	10.32
JLP 1982	98.85	6.16
MMH 1983	80.78	3.21
MMH 1984	59.16	4.32
MMH 1985	63.75	3.56
MMH 1986	105.75	1.86
MMH 1987	159.17	3.65
MMH 1988	51.66	4.14
CSG 1989	19.69	4.82
CSG 1990	29.93	5.40
CSG 1991	18.79	4.21
CSG 1992	11.94	3.01
CSG 1993	8.01	2.99
CSG 1994	7.05	2.56
EZP 1995	51.97	2.83
EZP 1996	27.70	2.95
EZP 1997	15.72	2.29
EZP 1998	18.61	1.56
EZP 1999	12.32	3.25
EZP 2000	8.96	3.39
VFQ 2001	4.40	1.55
VFQ 2002	5.70	2.38
VFQ 2003	3.90	1.89
VFQ 2004	4.61	1.13
VFQ 2005	4.72	1.03

FUENTE: Elaboración del autor con datos de INEGI.

COEF. DE CORRELACIÓN	r	R2
ENTRE INF MEX E INF EU	-0.02	0.00
1 año atrás	0.06	0.00
2 años atrás	0.27	0.07
3 años atrás	0.35	0.13
4 años atrás	0.34	0.11
5 años atrás	0.51	0.26
6 años atrás	0.72	0.51

Puede observarse que si la correlación se elabora con datos para México en t y para Estados Unidos t+1, se eleva el coeficiente, lo cual se sostiene a través del tiempo a medida que se desplaza el tiempo para la correlación.

Ejemplo 3. Los siguientes datos son para el estado de Tlaxcala en 1994. La información está desagregada por municipio, una serie de datos se refiere a la votación porcentual que obtuvo el PRI en las elecciones a presidencias municipales, la otra serie de datos se refiere al nivel educativo de los municipios y se ha operacionalizado por medio de la variable que indica el porcentaje de alfabetas (saben leer y escribir) dentro del total de la población mayor a 12 años.

Obtenga la correlación entre las series de datos.

Cuadro 41

Votación porcentual por el PRI en las elecciones municipales de 1994 y nivel educativo en Tlaxcala

		Xi 1994 Educ.	Yi 1994 Votación	XiYi	Xi2	Yi2
1	Amaxac	58.39	52.05	3039.20	3409.39	2709.20
2	A.Carvajal	60.14	44.62	2683.45	3616.82	1990.94
3	Apizaco	61.33	58.89	3611.72	3761.37	3468.03
4	Atlangatepec	48.25	100.00	4825.00	2328.06	10000.00
5	Altzayanca	49.15	80.38	3950.68	2415.72	6460.94
6	Calpulalpan	55.98	33.45	1872.53	3133.76	1118.90
7	El Carmen	44.50	80.91	3600.50	1980.25	6546.43
8	Cuapiaxtla	48.41	59.13	2862.48	2343.53	3496.36
9	Cuaxomulco	54.94	36.68	2015.20	3018.40	1345.42
10	Chiautempan	56.96	40.67	2316.56	3244.44	1654.05
11	D. Arenas	53.64	67.74	3633.57	2877.25	4588.71
12	Españita	52.68	77.52	4083.75	2775.18	6009.35
13	Huamantla	51.52	79.54	4097.90	2654.31	6326.61
14	Heuyotlipan	52.54	45.97	2415.26	2760.45	2113.24
15	Ixtacuixtla	56.70	45.13	2558.87	3214.89	2036.72
16	Ixtenco	60.94	46.42	2828.83	3713.68	2154.82
17	Morelos	51.48	39.49	2032.95	2650.19	1559.46
18	J. Cuamatzi	53.97	48.76	2631.58	2912.76	2377.54
19	Lardizabal	57.98	49.75	2884.51	3361.68	2475.06
20	L. Cárdenas	52.40	51.54	2700.70	2745.76	2656.37
21	M. Arista	53.32	62.29	3321.30	2843.02	3880.04
22	Hidalgo	55.09	67.96	3743.92	3034.91	4618.56
23	Nativitas	57.45	56.36	3237.88	3300.50	3176.45
24	Panotla	59.64	68.86	4106.81	3556.93	4741.70
25	S. Pablo	48.47	75.48	3658.52	2349.34	5697.23
26	S. Cruz T.	56.89	63.18	3594.31	3236.47	3991.71
27	Tenancingo	55.02	52.40	2883.05	3027.20	2745.76
28	Teolocholco	50.74	58.24	2955.10	2574.55	3391.90
29	Tepeyanco	58.49	57.16	3343.29	3421.08	3267.27
30	Terrenate	41.48	67.31	2792.02	1720.59	4530.64
31	Tetla	53.18	71.32	3792.80	2828.11	5086.54
32	Tetlatlahuca	62.64	71.34	4468.74	3923.77	5089.40
33	Tlaxcala	61.95	64.17	3975.33	3837.80	4117.79
34	Tlaxco	49.72	76.77	3817.00	2472.08	5893.63
35	Tocatlán	58.53	73.95	4328.29	3425.76	5468.60
36	Totolac	61.90	66.98	4146.06	3831.61	4486.32
37	Trinidad de SS.	43.60	77.54	3380.74	1900.96	6012.45
38	Tzonpantepec	54.28	72.47	3933.67	2946.32	5251.90
39	Xaloztoc	52.34	44.80	2344.83	2739.48	2007.04
40	Xaltican	54.21	58.75	3184.84	2938.72	3451.56
41	Xicohtencatl	58.36	50.18	2928.50	3405.89	2518.03
42	Xicotzingo	62.55	49.89	3120.62	3912.50	2489.01
43	Yauhquemecan	58.11	70.54	4099.08	3376.77	4975.89
44	Zacatelco	58.35	53.24	3106.55	3404.72	2834.50
		2408.21	2669.82	144908.50	132927.00	170812.09

COEF CORR.
-0.39

FUENTE: Elaboración del autor con datos del IET y del INEGI.

Ejemplo 4. Los siguientes datos son para el estado de Tlaxcala en 1994. La información está desagregada por municipio, una serie de datos se refiere a la votación porcentual que obtuvo el PRI en las elecciones a presidencias municipales; otra serie de datos se refiere al nivel educativo de los municipios y se ha operacionalizado por medio de la variable que indica el porcentaje de alfabetas (saben leer y escribir) dentro del total de la población mayor a 12 años; una serie adicional de datos se refiere al nivel de urbanización que tiene cada municipio, lo que se operacionalizó como el porcentaje de la población en el municipio que vivía en localidades mayores a 2500 habitantes; otra serie se refiere a la riqueza que existe en el municipio, operacionalizada como el número de salarios mínimos per cápita; la última serie de datos es la industrialización, y se refiere al porcentaje de la *población ocupada* que labora en el sector secundario.

Obtenga la correlación entre las series de datos y analice los resultados. (Para la solución a este problema véase el Anexo B del capítulo 3).

Cuadro 42

Votación porcentual por el PRI en las elecciones municipales de 1994
y datos Socioeconómicos en Tlaxcala

		Xi1 1994 Educación	Yi 1994 Votación	Xi2 1994 Urbanización	Xi3 1994 Riqueza	Xi 4 1994 Industria
1	Amaxac	58.39	52.05	22.28	1.31	29.28
2	A.Carvajal	60.14	44.62	14.35	1.48	45.13
3	Apizaco	61.33	58.89	34.23	3.95	33.54
4	Atlangatepec	48.25	100	0	0.25	38.86
5	Altzayanca	49.15	80.38	50.02	7.61	15.14
6	Calpulalpan	55.98	33.45	0.83	2.58	42.24
7	El Carmen	44.5	80.91	29.4	2.12	49.6
8	Cuapiaxtla	48.41	59.13	3.67	1.61	17.92
9	Cuaxomulco	54.94	36.68	0	0.99	33.56
10	Chiautempan	56.96	40.67	85.79	2.95	47.26
11	D. Arenas	53.64	67.74	0	12.06	28.91
12	Españita	52.68	77.52	0	2.47	15.32
13	Huamantla	51.52	79.54	5.95	2.43	35.11
14	Heuyotlipan	52.54	45.97	8.67	2.36	25.82
15	Ixtacuixtla	56.7	45.13	8.24	1.64	40.99
16	Ixtenco	60.94	46.42	100	7.36	12.35
17	Morelos	51.48	39.49	100	3.19	63.46
18	J. Cuamatzi	53.97	48.76	13.43	0.51	73.47
19	Lardizabal	57.98	49.75	31.27	0.82	49.34
20	L. Cárdenas	52.4	51.54	9.83	2.88	29.35
21	M. Arista	53.32	62.29	2.27	1.07	39.91
22	Hidalgo	55.09	67.96	10.5	23.44	54.72
23	Nativitas	57.45	56.36	34.82	2.16	23.86
24	Panotla	59.64	68.86	29.32	6.62	27.2
25	S. Pablo	48.47	75.48	74.8	1.07	84.6
26	S. Cruz T.	56.89	63.18	8.78	1.18	40.28
27	Tenancingo	55.02	52.4	89.24	3.3	74.17
28	Teolocholco	50.74	58.24	25	0.48	68.23
29	Tepeyanco	58.49	57.16	40.74	1.61	54.22
30	Terrenate	41.48	67.31	32.4	1.13	26.78
31	Tetla	53.18	71.32	9.84	2.06	48.52
32	Tetlatlahuca	62.64	71.34	62.54	3.93	22.72
33	Tlaxcala	61.95	64.17	65.98	2.88	31.93
34	Tlaxco	49.72	76.77	1.82	1.61	33.41
35	Tocatlán	58.53	73.95	50	1.76	53.21
36	Totolac	61.9	66.98	100	2.39	33.92
37	Trinidad de SS.	43.6	77.54	17.4	0.7	10.23
38	Tzonpantepec	54.28	72.47	31.27	24.13	36.33
39	Xaloztoc	52.34	44.8	77.59	2.77	45.28
40	Xaltican	54.21	58.75	100	1.48	30.95
41	Xicohtencatl	58.36	50.18	87.06	3.1	57.11
42	Xicotzingo	62.55	49.89	100	1.42	67.19
43	Yauhquemecan	58.11	70.54	100	5.15	38.18
44	Zacatelco	58.35	53.24	37.89	2.82	55.72
		2408.21	2669.82	1707.22	158.83	1785.32

Correlación	
	Votación
Educación	-0.39
Urbanización	-0.19
Riqueza	0.16
Industrialización	-0.21

FUENTE: Elaboración del autor con datos del IET y del INEGI.

Puede verse que la única variable significativa es la de *Educación*.

Anexo A del capítulo 3

Correlación a través de Excel

1. Requiere de dos series de datos.
2. Elija en el menú principal “insertar” y de esa opción elija “función”, con lo que aparecerá un cuadro de diálogo.
3. Dicho cuadro le da la opción “O seleccionar una categoría”, presione “click” y elija “estadísticas”, con lo que aparecerá más abajo la opción “COEF.DE.CORREL” elija esa opción y verá la leyenda “Devuelve el coeficiente de correlación de dos conjuntos de datos”, elija “aceptar”.
4. Aparecerá un nuevo cuadro de diálogo, que le pedirá dos matrices, diciendo: “…es un rango de celdas de valores. Los valores deben ser números, nombres, matrices o referencias que contengan números.” Inserte entonces los rangos arrastrando el “mouse” a lo largo de las series de observaciones.
5. Presione aceptar y obtendrá el valor en la celda de la hoja de cálculo, el cual ya es visible en el cuadro de diálogo.
6. Para poder plantear correlaciones con rezagos habrá que escoger pertinentemente las celdas de las observaciones.

3.2 Coeficiente de correlación ordinal de Spearman

El coeficiente de correlación de Spearman mide el grado de asociación entre dos series de datos que tengan una forma de medición ordinal.

Los valores fluctúan entre -1 y 1:

-1 significa que tienen una relación perfectamente inversa

1 significa que tienen una relación perfectamente directa

0 o valores cercanos a cero indican que no hay relación entre los indicadores de esas variables.

Los valores intermedios son:

-0.90	Correlación negativa muy fuerte
-0.75	Correlación negativa considerable
-0.50	Correlación negativa media
-0.10	Correlación negativa débil
0.00	Ausencia de correlación
0.10	Correlación positiva débil
0.50	Correlación positiva media
0.75	Correlación positiva considerable
0.90	Correlación positiva muy fuerte

Cuando el coeficiente de correlación se eleva al cuadrado, se obtiene R_s^2, el cual se denomina varianza de factores comunes, y es el indicador del porcentaje de la variación de una variable debida a la variación de la otra variable y viceversa.

La fórmula mediante la cual se calcula el coeficiente de correlación de Spearman (r_s) es la siguiente:

$$r_s = 1 - \frac{6\sum d^2}{n^3 - n}$$

Donde:

$n =$ Es el número de datos

$\sum d^2 =$ Es la suma de las diferencias, elevadas al cuadrado,

entre cada par de datos de las dos series, una vez que se han ordenado, es decir, se han dispuesto ordinalmente (se pueden ordenar de menor a mayor o de mayor a menor, pero utilizando el mismo criterio para ordenar a ambas series de datos).

Ejemplo:

Los siguientes datos fueron recopilados mediante una muestra de los trabajadores de una secretaría, perteneciente a un gobierno estatal de México. Se recopilaron mediante un cuestionario anónimo. La situación fue al final del sexenio y cuestionaba acerca de la preferencia entre otorgar un bono sexenal a cada trabajador o acordar la venta de los autos. Los datos obtenidos son las edades, el grado de acuerdo con el bono sexenal y, el referente a la venta de autos.

Obtenga y analice la relación entre las edades y los dos grados de acuerdo con las respectivas propuestas.

Cuadro 43

		Edad	Grado de acuerdo A	Grado de acuerdo B	Promedio del grado de acuerdo
1	PRM	23	4	3	3.5
2	RHS	29	5	3	4
3	ANG	45	5	4	4.5
4	LHC	43	4	2	3
5	JAPO	27	3	1	2
6	MRB	29	2	1	1.5
7	FOO	33	2	4	3
8	GJ	34	5	4	4.5
9	NMA	34	5	5	5
10	FLH	56	5	4	4.5
		353	40	31	
		35.3			

A	Acuerdo con bono sexenal
B	Acuerdo con venta de autos

FUENTE: Elaboración del autor.

Procesando ordinales:

Cuadro 44

Ordinales para grado de acuerdo con B y por edad

	Ordinales por grado de acuerdo con B			Ordinales por edad	
	Valor Con B	Número ordinal		Edad	Número ordinal
JAPO	1	1	PRM	23	1
MRB	1	2	JAPO	27	2
LHC	2	3	RHS	29	3
PRM	3	4	MRB	29	4
RHS	3	5	FOO	33	5
ANG	4	6	GJ	34	6
FOO	4	7	NMA	34	7
GJ	4	8	LHC	43	8
FLH	4	9	ANG	45	9
NMA	5	10	FLH	56	10

Elaboración del autor.

Finalizando el cálculo:

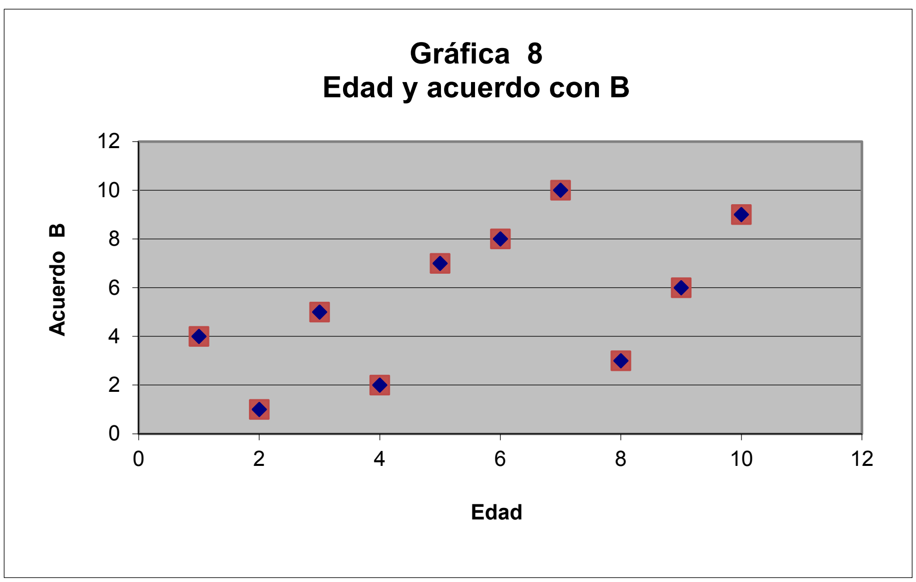
Cuadro 45

Obtención del indicador de correlación de Spearman

	Ordinal edad	Ordinal Acuerdo B	Edad-A	(Edad-A)2		
PRM	1	4	-3	9		
RHS	3	5	-2	4		
ANG	9	6	3	9		
LHC	8	3	5	25		
JAPO	2	1	1	1		
MRB	4	2	2	4		
FOO	5	7	-2	4		
GJ	6	8	-2	4		
NMA	7	10	-3	9		
FLH	10	9	1	1		
				70	10	n
				420	1000	n3
				0.4242	990	n3-n
			CS	0.5758		
			R2	0.3315		

FUENTE: Elaboración del autor.

La gráfica que enlaza ambas variables es la siguiente:



FUENTE: Elaboración del autor.

Dado que ya se tienen los ordinales por edad, se requieren los ordinales por acuerdo con A:

Cuadro 46

Ordinales para grado de acuerdo con A

	Ordinales por grado de acuerdo con A	
	Valor con A	Número ordinal
FOO	2	1
MRB	2	2
JAPO	3	3
LHC	4	4
PRM	4	5
ANG	5	6
FLH	5	7
GJ	5	8
NMA	5	9
RHS	5	10

FUENTE: Elaboración del autor.

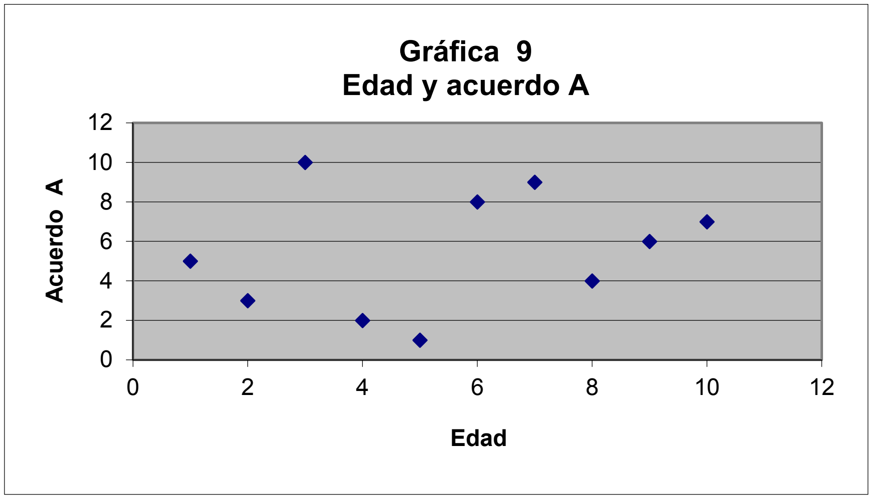
Finalizando el cálculo del indicador:

Cuadro 47

Obtención del indicador de correlación de Spearman

	Ordinal edad	Ordinal acuerdo A	Edad-A	(Edad-A)2		
PRM	1	5	-4	16		
RHS	3	10	-7	49		
ANG	9	6	3	9		
LHC	8	4	4	16		
JAPO	2	3	-1	1		
MRB	4	2	2	4		
FOO	5	1	4	16		
GJ	6	8	-2	4		
NMA	7	9	-2	4		
FLH	10	7	3	9		
				128	10	n
				768	1000	n3
				0.7758	990	n3-n
			CS	0.2242		
			R2	0.0503		

FUENTE: Elaboración del autor.



FUENTE: Elaboración del autor.

La correlación entre edad y el bono sexenal es de .22; edad con venta de autos es de 0.58.

3.3 Regresión

La regresión lineal establece la relación entre series de datos, estableciendo la relación de cambio y, adicionalmente, un punto de intercepto.

Los posibles valores fluctúan desde –n hasta +n.

El formato para obtener los valores es:

$$y_i = b_0 + b_1 \, x$$

La pendiente de la línea de regresión se obtiene con la siguiente fórmula:

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

El intercepto con Y de la línea de regresión se obtiene con la siguiente fórmula:

$$b_0 = \bar{y} - b_1 \, \bar{x}$$

Donde la barra superior en y y x indica que se trata de los promedios. [15]

[15] Wayne, *Estadística*.

Entonces, los elementos necesarios son los siguientes:

$\sum xy =$

$\sum x =$

$\sum y =$

$\sum x^2 =$

$n =$

$\bar{y} =$

$\bar{x} =$

También pueden usarse las siguientes fórmulas:

$$\beta_1 = \frac{\sum (x_t - \bar{x})\, y_t}{\sum (x_t - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_2 \bar{x}$$ [16]

Ejemplo 1

Los siguientes datos son de las familias de alumnos de una universidad de Estados Unidos, y corresponden al ingreso semanal y los gastos en alimentos, fueron tomados en 1998.

Obtenga la función de regresión muestral.

[16] R. Carter E. A., *Undergraduate econometrics*. John Willey and Sons Inc., United States of America, 2000.

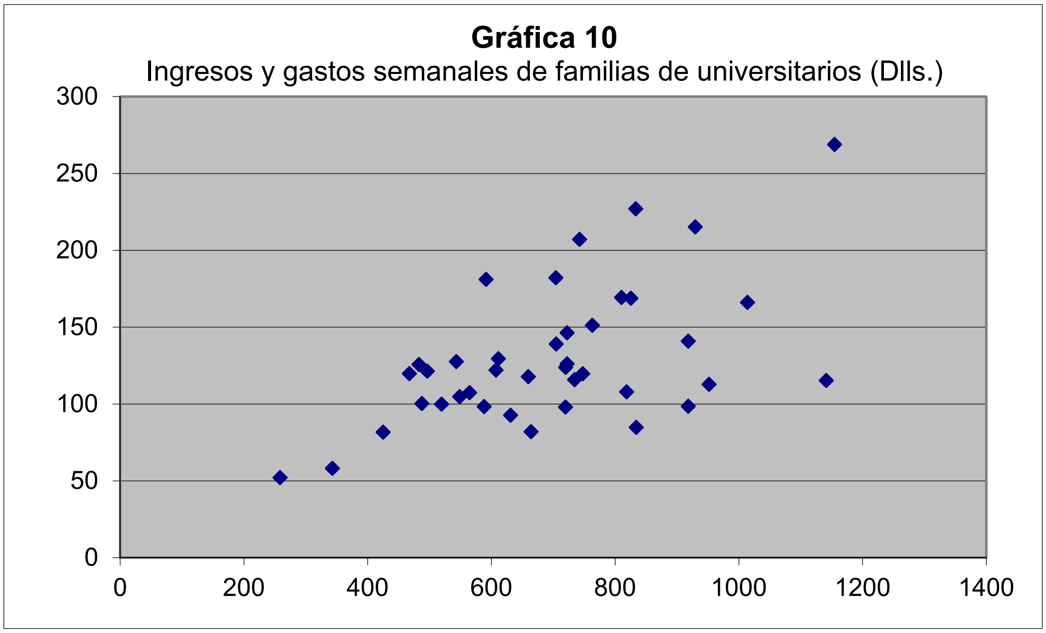
Cuadro 48

Ingresos y gastos de familias de universitarios semanales (Dlls.)

Observación	Gasto en alimentos	Ingreso
1	52.25	258.30
2	58.32	343.10
3	81.79	425.00
4	119.90	467.50
5	125.80	482.90
6	100.46	487.70
7	121.51	496.50
8	100.08	519.40
9	127.75	543.30
10	104.94	548.70
11	107.48	564.60
12	98.48	588.30
13	181.21	591.30
14	122.23	607.30
15	129.57	611.20
16	92.84	631.00
17	117.92	659.60
18	82.13	664.00
19	182.28	704.20
20	139.13	704.80
21	98.14	719.80
22	123.94	720.00
23	126.31	722.30
24	146.47	722.30
25	115.98	734.40
26	207.23	742.50
27	119.80	747.70
28	151.33	763.30
29	169.51	810.20
30	108.03	818.50
31	168.90	825.60
32	227.11	833.30
33	84.94	834.00
34	98.70	918.10
35	141.06	918.10
36	215.40	929.60
37	112.89	951.70
38	166.25	1,014.00
39	115.43	1,141.30
40	269.03	1,154.60
	5,212.52	27,920.00
	130.31	698.00

FUENTE: Elaboración del autor con datos de Carter, Undergraduate.

La gráfica es la siguiente:



FUENTE: Elaboración del autor.

Para el cálculo de las Betas:

Cuadro 49

Ingresos y gastos de familias de universitarios semanales (Dlls.)

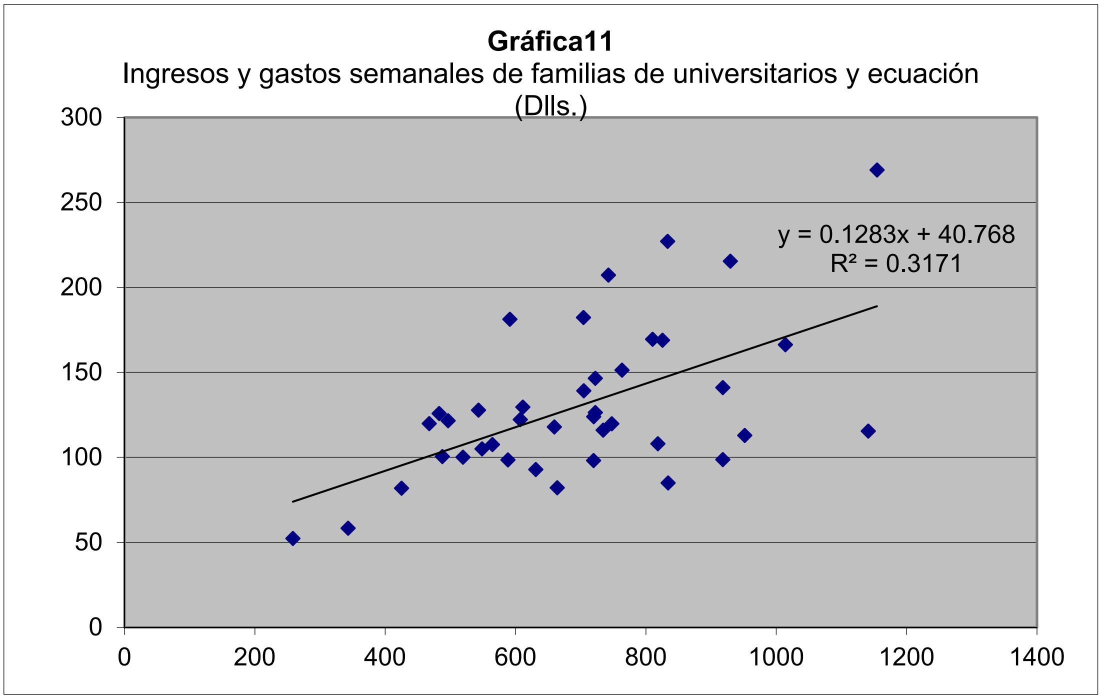
Cálculo de la línea de regresión

Obs	Food Expenditure Y	Income X	NUMERADOR	DENOMINADOR		
1	52.25	258.30	- 22,974.33	193,336.09		
2	58.32	343.10	- 20,697.77	125,954.01		
3	81.79	425.00	- 22,328.67	74,529.00		
4	119.90	467.50	- 27,636.95	53,130.25		
5	125.80	482.90	- 27,059.58	46,268.01		
6	100.46	487.70	- 21,126.74	44,226.09		
7	121.51	496.50	- 24,484.27	40,602.25		
8	100.08	519.40	- 17,874.29	31,897.96		
9	127.75	543.30	- 19,762.93	23,932.09		
10	104.94	548.70	- 15,667.54	22,290.49		
11	107.48	564.60	- 14,337.83	17,795.56		
12	98.48	588.30	- 10,803.26	12,034.09		
13	181.21	591.30	- 19,335.11	11,384.89		
14	122.23	607.30	- 11,086.26	8,226.49		
15	129.57	611.20	- 11,246.68	7,534.24		
16	92.84	631.00	- 6,220.28	4,489.00		
17	117.92	659.60	- 4,528.13	1,474.56		
18	82.13	664.00	- 2,792.42	1,156.00		
19	182.28	704.20	1,130.14	38.44		
20	139.13	704.80	946.08	46.24		
21	98.14	719.80	2,139.45	475.24		
22	123.94	720.00	2,726.68	484.00		
23	126.31	722.30	3,069.33	590.49		
24	146.47	722.30	3,559.22	590.49		
25	115.98	734.40	4,221.67	1,324.96		
26	207.23	742.50	9,221.74	1,980.25		
27	119.80	747.70	5,954.06	2,470.09		
28	151.33	763.30	9,881.85	4,264.09		
29	169.51	810.20	19,019.02	12,588.84		
30	108.03	818.50	13,017.62	14,520.25		
31	168.90	825.60	21,551.64	16,281.76		
32	227.11	833.30	30,727.98	18,306.09		
33	84.94	834.00	11,551.84	18,496.00		
34	98.70	918.10	21,723.87	48,444.01		
35	141.06	918.10	31,047.31	48,444.01		
36	215.40	929.60	49,886.64	53,638.56		
37	112.89	951.70	28,640.19	64,363.69		
38	166.25	1,014.00	52,535.00	99,856.00		
39	115.43	1,141.30	51,170.12	196,514.89		
40	269.03	1,154.60	122,839.10	208,483.56	b2	b1
	5,212.52	27,920.00	196,597.54	1,532,463.02	0.13	40.77
	130.31	698.00				

corr	r*r
0.563	0.317

FUENTE: Elaboración del autor con datos de Carter, Undergraduate.

El resultado puede verificarse mediante la gráfica 11 (para su obtención véase el anexo B del capítulo 3.)



FUENTE: Elaboración del autor.

Ejemplo 2

Los siguientes datos corresponden a pacientes que sufren un infarto y a los que se les inyecta una sustancia en distintas dosis, además se da el dato de los segundos que tardan en reaccionar.

Obtenga los valores de intercepto y la razón de cambio.

Obtenga el error estándar.

Obtenga valores interpolados y extrapolados diversos.

Cuadro 50

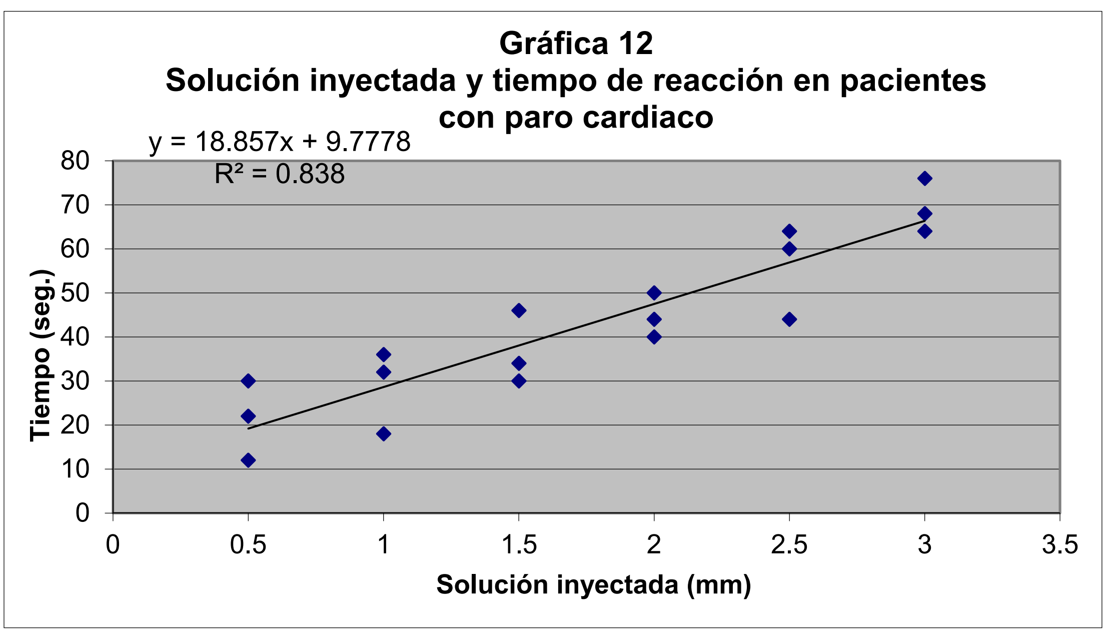
Milímetros inyectados al paciente y tiempo de reacción medido en segundos en pacientes que sufren un paro cardíaco

	Xi	Yi	Xi2	Yi2	XiYi
1	0.5	12	0.25	144	6
2	0.5	22	0.25	484	11
3	0.5	30	0.25	900	15
4	1	18	1	324	18
5	1	32	1	1024	32
6	1	36	1	1296	36
7	1.5	30	2.25	900	45
8	1.5	34	2.25	1156	51
9	1.5	46	2.25	2116	69
10	2	40	4	1600	80
11	2	44	4	1936	88
12	2	50	4	2500	100
13	2.5	44	6.25	1936	110
14	2.5	60	6.25	3600	150
15	2.5	64	6.25	4096	160
16	3	64	9	4096	192
17	3	68	9	4624	204
18	3	76	9	5776	228
	31.5	770	68.25	38508	1595
	1.75	42.7777778			

Correlación	0.92
R Cuadrada	0.84

FUENTE: Wayme, Estadística, 317. Si bien el autor lo aplica a otra herramienta estadística.

La gráfica correspondiente es la 12:



Gráfica 12
Solución inyectada y tiempo de reacción en pacientes con paro cardiaco

FUENTE: Elaboración del autor.

Cálculo del error estándar:

Cuadro 51

Milímetros inyectados al paciente y tiempo de reacción medido en segundos en pacientes que sufren un paro cardiaco

	Xi	Yi	Y	Y-Yi	(Y-Yi)*(Y-Yi)
1	0.5	12	19.206	7.206	51.931
2	0.5	22	19.206	-2.794	7.805
3	0.5	30	19.206	-10.794	116.504
4	1	18	28.635	10.635	113.099
5	1	32	28.635	-3.365	11.325
6	1	36	28.635	-7.365	54.246
7	1.5	30	38.063	8.063	65.017
8	1.5	34	38.063	4.063	16.510
9	1.5	46	38.063	-7.937	62.991
10	2	40	47.492	7.492	56.127
11	2	44	47.492	3.492	12.193
12	2	50	47.492	-2.508	6.291
13	2.5	44	56.920	12.920	166.934
14	2.5	60	56.920	-3.080	9.485
15	2.5	64	56.920	-7.080	50.122
16	3	64	66.349	2.349	5.517
17	3	68	66.349	-1.651	2.726
18	3	76	66.349	-9.651	93.146
	31.5	770			901.968
					50.109
				E.S.	7.079

FUENTE: Wayme, Estadística, 317. Si bien el autor lo aplica a otra herramienta estadística.

Cuadro 52

Milímetros inyectados al paciente y tiempo de reacción medido en segundos en pacientes que sufren sufren un paro cardiaco. Interpolación y extrapolación

	Xi	Yi	Y
1	0.5	12	19.206
2	0.5	22	19.206
3	0.5	30	19.206
4	1	18	28.635
5	1	32	28.635
6	1	36	28.635
7	1.5	30	38.063
8	1.5	34	38.063
9	1.5	46	38.063
10	2	40	47.492
11	2	44	47.492
12	2	50	47.492
13	2.5	44	56.920
14	2.5	60	56.920
15	2.5	64	56.920
16	3	64	66.349
17	3	68	66.349
18	3	76	66.349
	2.3		53.1489
	3.2		70.1202
	0.03		10.344

FUENTE: Elaboración del autor.

Es decir que dados los valores de las betas, la interpolación correspondiente a 2.3 mm cúbicos es de 53 segundos; la correspondiente a 0.03 mm cúbicos es de 10 segundos, y; para 3.2 mm cúbicos es de 70.12 segundos.

Anexo B del capítulo 3

Manual de interpolación y extrapolación con Excell

I. Datos

1. Meter los datos a Excell.

II. Gráfico de líneas

2. Obtener un gráfico de "líneas".
3. "Clic", imponer orden de pestaña de "selección de serie", obteniendo así el "asistente para gráficos".
4. Tipo Estándar, de líneas.
5. Finalizar.

III. Sobre el gráfico

1. Hacer "clic derecho" sobre la línea, cuidando de no hacerlo sobre la línea de fila, hacer "clic" en línea de tendencia.
2. Aparecen varios tipos de gráfico: lineal, logarítmica, polinomial, potencial, exponencial y media móvil.
3. Escoger pestaña "opciones".
4. Extrapolar, darle periodos atrás y adelante.
5. En "señalar intersección" se refiere al punto de intercepto, y se refiere a forzar el valor a en $y = b_0 + b_1x$, esto podría variar la línea de tendencia porque ella buscará pasar por entre los datos, ya que es la línea de mínimos cuadrados ordinarios.
6. Dar "palomita" en "presentar ecuación en el gráfico", lo mismo en "presentar el valor R cuadrado en el gráfico". Dar "aceptar".
7. Dar "clic derecho" en el área del gráfico.
8. Escoger "ubicación", y escoger entre "en hoja nueva" o "como objeto" y dar "clic" en la escogida.
9. Se obtiene $y = b_1x + b_0$ y no $y = b_0 + b_1x$. La fórmula y el R cuadrado pueden agrandarse o arrastrarse seleccionándolos.

IV. Para mejorar la regresión

1. "Clic derecho" en la curva de M. C. O. y seleccionar "formato de línea de tendencia", aparece un cuadro anterior, escoger en él la opción "tipo" y se escoge "logarítmica". Los valores de M. C. O. y de R cuadrado variarán.
2. De la misma forma que en 1, se siguen tomando las otras opciones de "formato de línea de tendencia", que se escoge con "clic derecho" en la línea de tendencia y se escoge "tipo" para escoger el gráfico.
3. Con 2 se está buscando el R cuadrado más elevado. Recuérdese que el R cuadrado indica qué tanto se ajusta la tendencia a los datos reales, es decir,

qué tanto la línea de datos calculados se ajusta a la línea de tendencia de los datos reales, ambos aparecen sobre el gráfico.

4. Considérense los siguientes casos: 1) cuando se utiliza la media móvil aparece "periodo", el cual contiene un máximo y un mínimo, es necesario darle orden. No presenta R cuadrado ni ecuación, se presenta según los periodos que se hayan escogido. Todos los demás tipos sí presentan fórmula y R cuadrado, pero si se ha utilizado anteriormente la media móvil se borra la orden de presentarlos, por lo que es necesario volver a dar la orden de darlos; 2) En la polinomial al aumentar los "orden" aumentan las b y se eleva el R cuadrado, el mínimo de "orden" que acepta es 2.

Nota 1

A pesar de encontrar R cuadrado = 1 o muy cercano a 1 hay que evaluar lógicamente.

Nota 2

Si no tuviera PC haría como si fuera lineal y calcularía a y b mediante fórmulas, considerando los valores de x. Si no tengo datos para x y para y, tomo los dados por default.

V. Interpolación y extrapolación

1. Ya obtenida la línea de tendencia, seleccionando un R cuadrado y una ecuación, escogiendo o tratando de escoger una ecuación con R cuadrado alto, con las salvedades mencionadas en las notas anteriores. Con la ecuación se pueden obtener los datos intermedios o exteriores al periodo de observaciones, con una calculadora de bolsillo.
2. Pero si se desea hacerlo con Excell: a) Seleccionar la ecuación y pegarla en la hoja de cálculo donde están los datos originales, se exponen los valores para x (se deben hacer explícitos por *default*) y se elabora la y estimada adecuando la función y rellenando hacia abajo.
3. Se puede obtener el error: e = y – y estimada. Hay que evaluar por qué en algunos puntos el error es alto, puede deberse a que haya puntos de inflexión en los datos observados.

Según se desee interpolar o extrapolar deben darse los espacios necesarios y, los datos para x (considerando los dados por *default*), y se pegará la función rellenando hacia abajo.

4. MATEMÁTICAS FINANCIERAS

Interés. Es la cantidad que se cobra o se paga por el uso del dinero.

Ahorradores ⟶ Intermediarios financieros ⟶ Prestatarios

En el interés simple no hay recapitalización de los intereses.

En el interés compuesto hay recapitalización de los intereses, es decir, los intereses se aúnan al monto.

Interés compuesto. Es aquél en el que al final de cada periodo se agregan los intereses al capital, es decir, se reinvierte (se capitaliza).

4.1 Interés simple

Variables respecto del interés simple:

Capital (C). Representa el capital inicial, llamado también principal, es la cantidad de dinero que se invierte o presta al inicio de una transacción (es el valor presente).

Monto (S). Representa el capital final, llamado también monto o dinero incrementado (es el valor futuro de C).

Interés (I). Representa el interés, y es la cantidad de dinero que se paga por el uso de C, es decir, es la diferencia entre S y C.

Tasa de interés (i). Representa el tanto por ciento que se paga por cada unidad monetaria y siempre se da anualmente, salvo que se especifique de manera diferente.

Tiempo (t). Representa el tiempo que dura una transacción.

Las fórmulas de interés simple:

1. $S = C + I$
2. $I = Cit$
3. $S = C(1 + it)$
4. $C = \dfrac{S}{1 + it}$

5. $i = \dfrac{\frac{S}{C} - 1}{t}$

6. $i = \dfrac{I}{Ct}$

7. $t = \dfrac{\frac{S}{C} - 1}{i}$

8. $t = \dfrac{I}{Ci}$

Ejercicios

1. Roberto otorga un préstamo a Alberto por $ 3,400.00 por un tiempo de 4 años y con una tasa de interés simple de 7 % ¿Cuánto recibe Roberto al final del lapso? ¿Cuánto ganó de intereses?

C = 3,400.00
i = 7 % (las tasas de interés siempre se dividen entre 100 para operaciones)
t = 4 años

Entonces:
S = X
I = X

$$S = C(1 + it)$$

S = 3, 400 (1+.07*4)
= 4, 352.00

Asimismo, se puede usar lo siguiente:

$$I = Cit$$

I = 3,400 *.07 * 4
= 952.00

Y como:

$$S = C + I$$

Entonces:
S = 3, 400 + 952
= 4, 352.00

2. Calcule la tasa de interés que se cobró, considere que se prestaron $ 87,000.00 y, después de 5 años, le devuelven $ 150,000.00

C = 87,000.00
T = 5 años
S = 150,000.00
i = X

$$i = \frac{\frac{S}{C} - 1}{t}$$

$$i = \frac{\frac{150000}{87000} - 1}{5}$$

= 0.1448 * 100 = 14.48 %
(Las tasas siempre se multiplicarán por 100 para dar resultados)

Otra forma de solución es:

$$i = \frac{I}{Ct}$$

$$i = \frac{I}{Ct} = \frac{S - C}{Ct}$$

$$i = \frac{I}{Ct} = \frac{S - C}{Ct} = \frac{150000 - 87000}{87000 * 5} = .1448 * 100 = 14.48$$

3. Francisco debe $ 450, 000.00, esa cantidad la deberá pagar al cabo de 3 años y medio. Si él decidiera pagar hoy la deuda, considerando que pactó una tasa de interés del 5.6 % ¿Cuánto pagaría?

S = 450, 000
t = 3.5
i = 5.6
C = X

$$C = \frac{S}{1 + it}$$

$$C = \frac{S}{1 + it} = \frac{450000}{1 + .056 * 3.5} = 376254.18$$

El resultado anterior significa que $ 376,254.18 es equivalente a tener $ 450,000.00 dentro de 3.5 años con una tasa de interés del 5.6 %

4. ¿En cuánto tiempo un capital se cuadruplica a una tasa de interés de 5.5 %?

i = 5.5
C = 1
S = 4
t = X

$$t = \frac{\frac{S}{C} - 1}{i}$$

$$t = \frac{\frac{S}{C} - 1}{i} = \frac{\frac{4}{1} - 1}{0.055} = 54.54$$

Son 54 años y medio los que se necesitan para cuadruplicar el capital inicial.

Una forma alternativa de solución es la siguiente:

$$t = \frac{I}{Ci}$$

$$t = \frac{I}{Ci} = \frac{S - C}{Ci} = \frac{4 - 1}{1 * 0.055} = 54.54$$

5. Gabriela ha prestado a Eleazar $ 10,000.00 a 8 meses, con una tasa de interés simple del 8 % ¿Cuánto le cobrará de intereses?

t = 8/12
C = 10, 000
i = 8 %

$$I = Cit$$

$$I = Cit = 10000 * .08\left(\frac{8}{12}\right) = 533.33$$

6. El promedio de gastos de un estudiante de posgrado en México asciende a $9,084.00, es decir, $ 109,008.00 anuales.[17]

Si un padre desea que su hijo curse estudios de maestría sin depender de la beca, entonces depositará en el banco una cantidad que sea suficiente para que los intereses paguen sus estudios de maestría.

[17] http://www.qro.cinvestav.mx (Consulta: 28 de diciembre de 2014). Es la cantidad que el CINVESTAV otorga mediante becas Conacyt a sus estudiantes de maestría.

Al hijo le faltan 8 años para terminar la licenciatura y suponiendo una tasa de interés simple del 27 % ¿Qué cantidad necesita depositar el día de hoy para poder cumplir su promesa?

Este problema se resuelve en dos partes:

Primero se necesita saber qué cantidad se debe invertir para que cada año produzca $ 109, 008.00 de interés.

t = 1
i = .27
I = 109, 008

C = X

I = Cit

Cit = I

C = I/it

C = 109, 008 / .27 * 1

= 403, 733.33

Es decir, si se depositan $ 403,733.33 esta cantidad generará $ 109,008.00 de intereses anualmente.

Para la segunda parte de solución del problema tiene que pensarse en buscar la cantidad de dinero que si se invierte hoy se convertirá en $ 403,733.33. Debe considerarse que el hijo requerirá la primera cantidad dentro de 8 años, por lo que ella tiene 7 años para formarse. Esto es que ahora S = $ 403,733.33:

$$C = \frac{S}{1 + it}$$

$$C = 403,\ 733.33 / 1 + .27 * 7$$

$$C = 139,\ 700.11$$

Es decir, el día de hoy el padre deposita $ 139,700.11 dentro de 7 años se convertirán en $ 403,733.33, cantidad que generará $ 109,008.00 de interés anual.

4.2 Interés compuesto

Es aquél en el que al final de cada periodo el interés se agrega al capital, es decir, se reinvierte o capitaliza.

Hay periodos de capitalización anual y fraccionaria. La primera es cuando hay una sola capitalización por año; la segunda sucede cuando la reinversión de interés se efectúa varias veces al año, al suceder lo anterior las fórmulas del interés compuesto quedan de la siguiente manera:

1. $S = C(1+i)^n$ Capitalización anual

2. $S = C(1+\frac{i}{r})^{nr}$ Capitalización fraccionaria

3. $C = \frac{S}{(1+i)^n}$ Capitalización anual

4. $C = \frac{S}{(1+\frac{i}{r})^{nr}}$ Capitalización fraccionaria

5. $i = \sqrt[n]{\frac{S}{C}} - 1$ Capitalización anual

6. $i = \left[\sqrt[nr]{\frac{S}{C}} - 1\right] r$ Capitalización fraccionaria

7. $n = \frac{Log(\frac{S}{C})}{Log(1+i)}$ Capitalización anual

8. $n = \frac{\frac{Log(\frac{S}{C})}{Log(1+\frac{i}{r})}}{r}$ Capitalización fraccionaria

En donde r representa las veces que se capitaliza el capital. Por ejemplo:

Si la capitalización es mensual,	r = 12
Si la capitalización es trimestral,	r = 4
Si la capitalización es semestral,	r = 2
Si la capitalización es bimestral,	r = 6
Si la capitalización es cuatrimestral,	r = 3
Si la capitalización es anual,	r = 1

Asimismo, debe considerarse que hay un cambio de notación, respecto a las fórmulas de interés simple: n es ahora el tiempo, anteriormente representado por t.

7. Silverio deposita en el banco $ 200,000.00, dicha institución paga una tasa de interés del 5.17 % ¿Cuánto recibirá después de 2 años, si la capitalización es:

a) Anual
b) Semestral

c) Mensual

f) Analizar los resultados obtenidos.

Datos generales:

C = 200, 000
i = 5.17
n = 2 años
S = X

a) Capitalización anual

n = 2

$$S = C(1+i)^n$$

$$S = C(1+i)^n = 200000\,(1+.0517)^2 = 221214.58$$

I = S – C

I = 221, 214.58 – 200, 000.00

I = 21, 214.58

El interés ganado es de $ 21, 214.58

b) Capitalización semestral

r = 2

$$S = C(1+\frac{i}{r})^{nr}$$

$$S = C\,(1+\frac{i}{r})^{nr} = 200000(1+\frac{.0517}{2})^{2*2} = 221495.78$$

I = S – C

I = 221, 495.78 – 200, 000.00

I = 21, 495.78

c) Capitalización mensual

r = 12

$$S = C(1+\frac{i}{r})^{nr}$$

$$S = C(1+\frac{i}{r})^{nr} = 200\,000\,(1+\frac{.0517}{12})^{\,2*12} = 221\,737.72$$

I = S - C

I = 221, 737.72 – 200, 000.00

I = 21, 737.72

f) Análisis

a. El monto mayor se obtiene cuando los intereses se capitalizan un mayor número de veces.

b. El factor de capitalización es mayor cuando r es mayor (número de veces que se capitalizan los intereses por año).

8. Guadalupe tiene $ 405,000.00, que llevará a invertir durante dos años a un banco que ofrece diversos planes de inversión:

a) Si la capitalización es anual, la tasa de interés es del 115 %
b) Si la capitalización es semestral, la tasa de interés es del 105 %
c) Si la capitalización es mensual, la tasa de interés es del 79 %

¿Cuál es el mejor plan?

C = 405, 000
i = varía 115, 105 y, 79
n = 2
r = varía 1, 2 y, 12
S = X

a) $S = C(1+i)^n$

$$S = C\,(1+i)^{\,n} = 405,\,000\,(1+1.15)^{\,2} = 1,\,872,\,112.5$$

b) $S = C(1+\frac{i}{r})^{nr}$

$$S = C(1+\frac{i}{r})^{nr} = 405,\,000\,(1+\frac{1.05}{2})^{\,2*2} = 2,\,190,\,455.31$$

c) $S = C(1+\frac{i}{r})^{nr}$

$$S = C(1+\frac{i}{r})^{nr} = 405,000\,(1+\frac{.79}{12})^{\,2*12} = 1,870,674.32$$

Por lo tanto, el plan que resulta más atractivo es el b), con una tasa de interés del 105 % y la capitalización semestral, pues el monto que se recibe es de $ 2,190,455.31 y es mayor que el de los otros planes.

9. Paco ha pedido un préstamo a Joshua, por ello deberá pagar $ 4,000,000.00 dentro de 4 años. Si él quisiera liquidar la obligación hoy y suponiendo que la tasa de interés es del 9 % y la capitalización es trimestral ¿Qué cantidad pagaría?

Y si la capitalización fuera anual ¿Qué cantidad pagaría?

S = 4,000,000.00
n = 4
i = .09
r = 4
C = X

$$C = \frac{S}{(1+\frac{i}{r})^{nr}}$$

$$C = \frac{S}{(1+\frac{i}{r})^{nr}} = \frac{4{,}000{,}000}{(1+\frac{.09}{4})^{4*4}} = 2,801,863.18$$

Es decir que Paco pagaría $ 2,801,863.18 el día de hoy, cantidad que es equivalente a pagar $ 4,000,000.00 dentro de 4 años, pues si esta cantidad se invirtiera bajo las mismas condiciones se tendrían $ 4,000,000.00

$$S = C(1+\frac{i}{r})^{nr}$$

$$S = C(1+\frac{i}{r})^{nr} = 2{,}801{,}863.18\,(1+\frac{.09}{4})^{\,4*4} = 4{,}000{,}000.00$$

Ahora bien, si la capitalización fuera anual:

$$C = \frac{S}{(1+i)^{n}}$$

$$C = \frac{S}{(1+i)^{n}} = \frac{4{,}000{,}000}{(1+.09)^{4}} = 2{,}833{,}700.84$$

Es decir, el pago sería de $ 2,833,700.84

10. Gabriela quiere invertir $ 85,000.00 y que en 3 años este capital se convierta en $ 120,000.00, la cuestión es a qué tasa de interés esto sucedería si la capitalización se hace:

a) Anual

b) Semestral

C = 85, 000
S = 120, 000
n = 3
r = 1 ó 2
i = X

a) Anual

$$i = \sqrt[n]{\frac{S}{C}} - 1$$

$$i = \sqrt[n]{\frac{S}{C}} - 1 = \sqrt[3]{\frac{120,000}{85,000}} - 1 = \sqrt[3]{1.411765} - 1 = 1.1218 - 1 = .1218$$

Si se multiplica el resultado por 100, el resultado final es una tasa de 12.18 %.

b) Semestral

$$i = \left[\sqrt[nr]{\frac{S}{C}} - 1\right]r$$

$$i = \left[\sqrt[nr]{\frac{S}{C}} - 1\right]r = \left[\sqrt[3*2]{\frac{120,000}{85,000}} - 1\right]*2 = .1183$$

Si se multiplica el resultado por 100, el resultado final es una tasa de 11.83 %.

11. Rocío se pregunta cuánto tiempo debe pasar para que $ 5000.00 se conviertan en $ 5,800.00, si se sabe que la tasa de interés es del 4.3 % y la capitalización es:

a) Anual,
b) Semestral.

C = 5, 000
S = 5, 800
i = .043
r = 1 ó 2
n = X

a) Anual

$$n = \frac{Log(\frac{S}{C})}{Log(1+i)}$$

$$n = \frac{Log(\frac{S}{C})}{Log(1+i)} = \frac{Log\frac{5,800}{5,000}}{Log(1+.043)} = \frac{Log(1.16)}{Log(1.043)} = \frac{.064457}{.018284} = 3.52$$

Son 3.52 años.

b) Semestral

$$n = \frac{\frac{Log(\frac{S}{C})}{Log(1+\frac{i}{r})}}{r}$$

$$n = \frac{\frac{Log(\frac{S}{C})}{Log(1+\frac{i}{r})}}{r} = \frac{\frac{Log\frac{5,800}{5,000}}{Log(1+\frac{.043}{2})}}{2} = (\frac{.064458}{.00923837})/2 = 3.49$$

Son 3.49 años.

Ejercicios:[18]

1. El señor David Juárez firmó un pagaré por $ 5,000,000.00 a 2 años, con una tasa de interés en la transacción del 75 % ¿Cuánto pagó el señor Juárez al finalizar el plazo?

t = 2 años
i = 75 %
C = 5,000,000

$$S = C(1+it)$$

$$S = C(1+it) = 5,000,000 * (1 + (.75 * 2)) = 12,500,000$$

S = $ 12,500,000.00

[18] Los cinco ejercicios siguientes se han obtenido de Hernández, Abraham. *Matemáticas financieras. Teoría y práctica*. ECAFSA Thomson Learning, México, 2002, pp. 102-103. Los ejercicios en ese texto se plantean y se da el resultado final, aquí se presenta el procedimiento para solución de los ejercicios.

2. Con respecto al problema anterior, ¿cuánto pagó el señor Juárez de intereses?

$$S = C + I$$

$$I = S - C$$

$$I = S - C = 12{,}500{,}000 - 5{,}000{,}000 = 7{,}500{,}000$$

$I =$ \$ 7,500,000.00

3. El señor Ruiz quiere saber cuál es el valor presente de un pagaré 9 meses antes de su vencimiento, si fue firmado a 3 años, con una tasa de interés del 92 % sobre los \$ 950,000.00 que le prestaron.

t = 3 años t = 9 meses
C = 950, 000
i = 92 %

$$S = C(1 + it)$$

$$S = C(1 + it) = 950,\,000\,(1 + (.92 * 3)) = 3,\,572,\,000$$

$$C = \frac{S}{1 + it}$$

$$C = \frac{S}{1 + it} = 3,\,572,\,000 \,/\, (1 + (.92 * 9/12) = 2,\,113,\,609.47$$

C = \$ 2,113,609.47

4. El señor Hernández firmó un pagaré por \$ 12,000 000.00 a 1.5 años con una tasa de interés del 97 % ¿cuánto pagó de intereses el señor Hernández?

C = \$12,000,000.00
t = 1.5
I = 97 %

$$I = Cit$$

$$I = Cit = 12{,}000{,}000 * .97 * 1.5 = 17{,}460{,}000$$

También:

$$S = C(1 + it)$$

$$S = C(1+it) = 12{,}000{,}000\ (1 + (.97 * 1.5)) = 29{,}460{,}000$$

I = S – C = 29,460,000 – 12,000,000 = 17,460,000

I = $ 17,460,000.00

5. ¿En qué precio podría el señor Fernández vender un pagaré con un valor de vencimiento de $ 15,000,000.00, 8 meses antes de su vencimiento de tal forma que no pierda ni gane, si la tasa de interés es del 80 %?

S = 15,000,000.00
t = 8 meses
i = 80 %

$$C = \frac{S}{1+it}$$

$$C = \frac{S}{1+it} = \frac{15,\ 000,\ 000}{1+(.8*\frac{8}{12})} = 9,\ 782,\ 608.7$$

C = $ 9, 782, 608.70

6. Calcular el valor presente de un pagaré por $ 700,000.00 firmado a 3 años, con una tasa de interés del 75 %, siete meses antes de su vencimiento.

C = 700,000
t = 3 t = 7 meses
i = 75 %

$$S = C(1+it)$$

$$S = C(1+it) = 700{,}000\ (1 + (.75 * 3)) = 2{,}275{,}000$$

$$C = \frac{S}{1+it}$$

$$C = \frac{S}{1+it} = 2{,}275{,}000 / (1 + (.75 * 7/12) = 1{,}582{,}608.70$$

C = $ 1,582,608.70

7. Una persona tiene $ 48,000.00 y desea depositarlos en el banco para obtener algunos rendimientos. Solamente puede depositarlos por un período de un año. Los planes que ofrecen los bancos son los siguientes:

a) Banco Olmeca
De $ 5,000.00… i = 5.5 % trimestralmente

b) Bansur
De $ 25,000.00-50,000.00 i = 3.02 % mensualmente

c) BAVB
De $ 100,000.00-250,000.00 i = 1.65 % trimestralmente

d) EXI
De $ 10,000.00-50,000.00 i = 2.15 % trimestralmente

a) Banco Olmeca

De $ 5,000.00… i = 5.5 % trimestralmente

$$S = C(1+\frac{i}{r})^{nr}$$

$$S = C(1+\frac{i}{r})^{nr} = 48,000\,(1+\frac{.055}{4})^{1*4} = 50,694.95$$

b) Bansur

De $ 25,000.00-50, 000.00 i = 3.02 % mensualmente

$$S = C(1+\frac{i}{r})^{nr}$$

$$S = C(1+\frac{i}{r})^{nr} = 48,000\,(1+\frac{.0302}{12})^{1*12} = 49,469.83$$

c) BAVB

De $ 100, 000.00-250, 000.00 i = 1.65 % trimestralmente

$$S = C(1+\frac{i}{r})^{nr}$$

$$S = C(1+\frac{i}{r})^{nr} = 48,000\,(1+\frac{.0165}{4})^{1*4} = 48,796.91$$

d) EXI

De $ 10,000.00-50,000.00 i = 2.15 % trimestralmente

$$S = C(1+\frac{i}{r})^{nr}$$

$$S = C(1+\frac{i}{r})^{nr} = 48{,}000\,(1+\frac{.0215}{4})^{1*4} = 49{,}040.35$$

Cuantitativamente conviene invertir en Banco Olmeca, ya que da mayor S. Aunque deberá valorarse el plazo y el riesgo.

8. En la publicidad de una cadena de tiendas se anuncia que está a la venta una computadora portátil. Se puede comprar en riguroso contado a \$ 11,995.00, pero también se puede adquirir en 24 pagos semanales de \$ 776.00, el primer pago debe efectuarse de inmediato.

Considere que la tasa de interés promedio es de 5.5 % trimestral.

Compare las dos formas de pagar.

i	= 5.5 %
C	= \$ 11, 995.00
t	= 24 semanas

$$C = \frac{S}{(1+\frac{i}{r})^{nr}}$$

$$C = \frac{S}{(1+\frac{i}{r})^{nr}} = \frac{776}{(1+\frac{.055}{48})^{((23/48)*48)}} = 755.83$$

Con ello se ha trasladado el último pago a valor presente. Se requiere, además, establecer los valores de cada uno de los pagos en valor presente y luego sumarlos:

Cuadro 53

Cálculo del valor presente de los 24 pagos

No. de semana de pago		C
23	1	755.83
22	2	756.69
21	3	757.56
20	4	758.43
19	5	759.30
18	6	760.17
17	7	761.04
16	8	761.91
15	9	762.78
14	10	763.66
13	11	764.53
12	12	765.41
11	13	766.29
10	14	767.16
9	15	768.04
8	16	768.92
7	17	769.80
6	18	770.69
5	19	771.57
4	20	772.45
3	21	773.34
2	22	774.22
1	23	775.11
0	24	776.00
SUMA		18,380.92

FUENTE: Elaboración del autor.

Puesto que el pago de contado es de $ 11,995.00 y la suma de los valores presentes, a una tasa de interés de 5.5 %, es de 18, 380. 92, entonces conviene más pagar al contado, aunque debe considerarse la tasa de inflación.

La diferencia entre las dos formas de pagar es: 18,380.92 – 11,995.00 = 6,385.92 pesos aunque, desde luego estaría posponiendo el consumo del bien y debe suponerse que la tasa de inflación sería de 0, o que la capacidad de compra se mantenga.

9. Una persona desea adquirir un departamento que tiene un valor de $ 250,000.00 Para ello desea pedir un financiamiento a 15 años, por lo que acude a solicitar un crédito hipotecario. Una institución financiera le ofrece una tasa anual de 14.5 %, con una comisión por apertura de 2.5 % y el porcentaje del monto a financiar es del 80 %.

Elabore el plan, abordando:

a) Cuánto necesita al principio
b) Cuánto necesita en cada uno de los años.

a) El 2.5 % de 250 000 es: $ 6, 250.00, cantidad que deberá pagar por la apertura.

Como solamente le financiarán 80 % del costo total, 20 % de 250,000 es: $ 50,000.00

Entonces, necesita inicialmente desembolsar $ 56,250.00

b) El monto del financiamiento es, 80 %, $ 200,000.00

$$S = C(1+it)$$

$$S = C(1+it) = 200,000*(1+.145*15) = 635,000$$

En 15 años deberá $ 635,000.00. Si se supone que el primer pago debe hacerse de inmediato:

Cuadro 54

Cálculo de los pagos anuales para cubrir el total del préstamo

1	21,009.10	14	63,657.57
2	21,009.10	13	60,611.25
3	21,009.10	12	57,564.93
4	21,009.10	11	54,518.61
5	21,009.10	10	51,472.30
6	21,009.10	9	48,425.98
7	21,009.10	8	45,379.66
8	21,009.10	7	42,333.34
9	21,009.10	6	39,287.02
10	21,009.10	5	36,240.70
11	21,009.10	4	33,194.38
12	21,009.10	3	30,148.06
13	21,009.10	2	27,101.74
14	21,009.10	1	24,055.42
15	21,009.10	0	21,009.10
		SUMA	635,000.05

FUENTE: Elaboración del autor.

Debe pagar $ 21,009.10 por año.

5. MUESTREO

Este capítulo es una recopilación y revisión de distintas formas de obtener una muestra a partir de una población determinada. Como se sabe, si fuera necesario obtener información directamente de las entidades de estudio, en muchas ocasiones no sería posible, lo cual sucede más cuando la población o universo de estudio es grande. Para poder realizar este tipo de investigaciones se utiliza el muestreo, método con el cual es posible acortar el tiempo y los recursos financieros que la investigación requiere. Se exponen seis textos, cada uno con una formulación distinta, en los cuales se encuentra esta necesidad de tomar muestras para aplicar cuestionarios y obtener información como fuente primaria de investigación.[19]

Se inicia con el formato de muestreo utilizado por Francisco Javier Barranco Sáiz en su libro *Técnicas de Marketing Político*, el cual dedica el capítulo 2, denominado "Investigación del mercado político", a la obtención de información por fuentes primarias. Presenta cuatro ejemplos, uno para cada una de las fórmulas que expone.

El segundo texto es el de Roberto Hernández Sampieri, su título es *Metodología de la investigación*, en el capítulo 8, "¿Cómo seleccionar una muestra?", expone un par de ejemplos.

Otro formato de muestreo lo expone María Isabel Pérez Enríquez en su libro *El impacto de las migraciones y expulsiones indígenas de Chiapas*, en el anexo es donde aborda el muestreo, el cual es aplicado en el estudio como forma de obtener información.

Felipe Párdinas, en su libro *Metodología y técnicas de investigación en ciencias sociales*, analiza, en el capítulo 7, otro formato de muestreo, ofreciendo dos ejemplos de aplicación.

En un quinto texto revisado, David Aaker y Georges Day en su libro *Investigación de mercados* dedican el capítulo 11, "Tamaño de la muestra y teoría estadística", al muestreo, exponiendo el tipo probabilístico y dando una fórmula para obtener el tamaño de la muestra, además dedican un ejemplo que va variando para exponer distintos niveles de error, confianza, tamaño poblacional y desviación estándar.

Por último, Iván Ximitl, en su tesis de grado *"Organización campesina y comercialización de café: estudio en dos municipios de la sierra norte de Puebla"*, expone una fórmula distinta para tomar el tamaño de la muestra que utiliza en su investigación.

En la séptima parte, después de explicar las formulaciones, se confrontan con fines de evaluación, para lo cual se destacan sus características y se establece la preferencia de algunos formatos sobre otros, mostrándose la primacía de uno de ellos.

---

[19] Una versión anterior de este capítulo aparece publicado en Sánchez Espinoza, Francisco. "La técnica utilizada en los estudios de cultura política. Una revisión de distintas formulaciones de la técnica de muestreo", en Fabiola Coutiño (coordinadora), *Perspectivas teóricas y metodológicas de la cultura política en México*. BUAP, México, 2011, pp. 103-130.

5.1 Barranco Sáiz

Francisco Barranco[20] expone una forma de muestreo basada en una serie de elementos y en la aplicación de cuatro fórmulas.

Los elementos necesarios para el diseño de una muestra son los siguientes:

1. Tipos de población
   a) Infinita, es mayor a 100 000 entidades
   b) Finita, es menor o igual a 100 000 entidades

2. Coeficiente de fiabilidad
   1) 95.5 %
   2) 99.7 %

3. Error de muestreo (E)

Puede escogerse dentro de un margen de 0 a 100 %, por supuesto, se prefieren errores pequeños. El 0 % de error implicaría encuestar al 100 % de la población.

4. Porcentajes

Los porcentajes o proporcionalidades son dos valores, p y q que sumados dan como resultado 100 %. Se pueden obtener a partir de una encuesta piloto que puede ser tomada en 100 a 200 entidades. Cuando no es posible, sobre todo por cuestiones financieras, hacer esa encuesta piloto, entonces se asigna un valor de 50 a cada una de las proporciones.[21]

Con los dos primeros elementos, tipos de población y coeficiente de fiabilidad, se escoge una de cuatro fórmulas; los otros dos elementos se encuentran en la constitución de la fórmula. Entonces, las posibles fórmulas son las siguientes:

1) Población infinita con coeficiente de fiabilidad de 99.7 %

$$n = \frac{9pq}{E^2}$$

2) Población infinita con coeficiente de fiabilidad de 95.5 %

$$n = \frac{4pq}{E^2}$$

3) Población finita con coeficiente de fiabilidad de 99.7 %

$$n = \frac{9pqN}{E^2(N-1)+9pq}$$

N, es el tamaño de la población o universo.

[20] Barranco, Francisco J., *Técnicas de marketing político*, Red Editorial Iberoamericana, México, 1994.

[21] Cuando ya se ha encuestado con anterioridad a la misma población entonces es posible tomar los valores de los porcentajes de esa encuesta.

4) Población finita con coeficiente de fiabilidad de 95.5 %

$$n = \frac{4pqN}{E^2(N-1)+4pq}$$

A continuación se presenta un ejemplo de aplicación de cada una de las fórmulas.[22]

Para el caso de una población infinita con un coeficiente de fiabilidad del 99.7 %, considérese que un partido político desea hacer un sondeo a nivel nacional,[23] con el objeto de conocer la imagen que tiene entre los futuros electores. Para ello encuesta, inicialmente, a 150 votantes potenciales, obteniendo que 60 tienen intención de votarle. ¿A cuántos tendrá que encuestar, si fija un coeficiente de fiabilidad del 99.7 % y un error de muestreo de +- 1.5 %?

Los valores de p y q se obtendrán del pretest realizado:

$$p = \frac{60}{150} * 100 = 40\,\%$$

$$q = 100 - 40 = 60\,\%$$

Entonces los elementos son:

Población infinita		
Coeficiente de fiabilidad	99.7	%
Error	1.5	%
p	40	%
q	60	%

Sustituyendo en la expresión se tiene lo siguiente:

$$n = \frac{9pq}{E^2}$$

$$n = \frac{9*40*60}{1.5^2} = \frac{21,600}{2.25} = 9,600$$

Se deberá entrevistar a 9,600 individuos de la población en estudio.

Para ejemplificar el caso de una población infinita con coeficiente de fiabilidad del 95.5 % se considera que se desea realizar una encuesta en una región,[24] con objeto de analizar las aspiraciones del electorado, de cara a una futura confrontación electoral. ¿A cuántos posibles votantes se deberá entrevistar, si el coeficiente de fiabilidad a utilizar es del 95.5 %, el error de muestreo del +- 4 %, y se considera que las condiciones de muestreo son desfavorables?

Entonces los elementos son:

---

[22] Estos cuatro ejemplos se han adecuado para su exposición a partir de Barranco, *Técnicas…*, pp. 58-64.

[23] Se supone que al tratarse de un sondeo nacional la población a encuestar excede a las 100, 000 entidades de estudio.

[24] Como en el caso anterior, se supone que las entidades en estudio son más de 100,000.

Población infinita
Coeficiente de fiabilidad 95.5 %
Error 4 %
p 50 %
q 50 %

Sustituyendo en la expresión se tiene lo siguiente:

$$n = \frac{4pq}{E^2}$$

$$n = \frac{4*50*50}{4^2} = \frac{10,000}{16} = 625$$

Ahora se ejemplifica el caso de una población finita con un coeficiente de fiabilidad del 99.7 %, se trata de que un partido político desea hacer una encuesta de opinión entre los integrantes de un colegio profesional, formado por 87,520 colegiados, ¿a cuántos deberá entrevistar si desea trabajar con un coeficiente de fiabilidad del 99.7 %, un error de muestreo del +- 4.5 % y condiciones desfavorables?

Entonces los elementos son:
Población finita
Coeficiente de fiabilidad 99.7 %
Error 4.5 %
p 50 %
q 50 %
N 87,520

Sustituyendo en la fórmula:

$$n = \frac{9pqN}{E^2(N-1)+9pq}$$

$$n = \frac{9*50*50*87520}{4.5^2*(87520-1)+9*50*50} = \frac{1969200000}{1794759.75} = 1097.19$$

Se entrevistará a 1,907 integrantes del colegio.[25]

Para el caso de la población finita con un coeficiente de fiabilidad del 95.5 %, considérense los datos del ejemplo anterior, pero sabiendo que se ha llevado a cabo un pretest a 180 colegiados, de los cuales 105 denotan afinidad ideológica hacia nuestro partido. El coeficiente de fiabilidad a aplicar será de 95.5 % y el error de muestreo del +- 4 %.

Primero se calculan las proporciones:

$$p = \frac{105}{180}*100 = 58.33\,\%$$

[25] El resultado correcto es 1,097, en el citado texto, Barranco, *Técnicas*…, se da como resultado 1,799.

$$q = 100 - 58.33 = 41.67\ \%$$

Por lo tanto los elementos son los siguientes:

Población finita

Coeficiente de fiabilidad	95.5	%
Error	4	%
p	58.33	%
q	41.67	%
N	87,520	

Aplicando la fórmula se tiene lo siguiente:

$$n = \frac{4pqN}{E^2(N-1)+4pq}$$

$$n = \frac{4*58.33*41.67*87520}{4^2*(87520-1)+4*58.33*41.67} = \frac{850908333.9}{1410026.444} = 603.47$$

Tablas de Harvard

Este método, constituido por cuatro fórmulas predeterminadas para obtener el tamaño de n, tiene gran sencillez. Solamente debe tenerse cuidado en el caso de que el valor escogido para el coeficiente de error sea de 0 y la población sea infinita, ya que en ese caso, cualquiera que sea el índice de confianza, el resultado de la fórmula se indefine; no sucede lo anterior cuando la población es finita, ya que en ese caso n es igual a N, resultado que es correcto, ya que si se desea un error de 0 se tendría que encuestar a toda la población. El caso de indefinición en la fórmula no se puede resolver, sin embargo, se sabe que un error de 0 requiere que se iguale n a N, es decir, el tamaño de la muestra es el total de la población, por lo que quedaría salvado el problema, no estadísticamente pero sí en términos prácticos.

La sencillez de las fórmulas permite la elaboración de las llamadas “tablas de Harvard”. Se trata de cuatro tablas, una para cada caso de fórmula: población infinita con coeficiente de confianza de 95.5 % y otra para 99.7 %; una más para población finita con coeficiente de confianza de 95.5 % y otra para 99.7 %.[26]

[26] Las tablas pueden ser consultadas en Barranco, *op. cit.*, pp. 62-65.

Tabla 1. Población infinita con coeficiente de confianza de 95.5 %

Límites de error en porcentajes	Valores posibles de p y q ($p + q = 100$)													
	1/99	2/98	3/97	4/96	5/95	10/90	15/85	20/80	25/75	30/70	35/65	40/60	45/55	50/50
0,1	39.600	78.400	116.400	153.600	190.000	360.000	510.000	640.000	750.000	840.000	910.000	960.000	990.000	1.000.000
0,2	9.900	19.600	29.100	38.400	47.500	90.000	127.500	160.000	187.500	210.000	227.500	240.000	247.500	250.000
0,3	4.400	8.711	12.933	17.067	21.111	40.000	56.667	71.111	83.333	93.333	101.111	106.667	110.000	111.111
0,4	2.475	4.900	7.275	9.600	11.875	22.500	31.875	40.000	46.875	52.500	56.875	60.000	61.875	62.500
0,5	1.584	3.136	4.656	6.144	6.600	13.400	20.400	25.600	30.000	33.600	36.400	38.400	39.600	40.000
0,6	1.100	2.178	3.233	4.267	5.278	10.000	14.167	17.778	20.833	23.333	25.278	26.667	27.500	27.778
0,7	808	1.600	2.376	3.135	3.878	7.347	10.408	13.061	15.306	17.143	28.577	19.592	20.204	20.408
0,8	619	1.225	1.819	2.400	2.969	5.625	7.969	10.000	11.719	13.125	14.219	15.000	15.469	15.625
0,9	489	968	1.437	1.896	2.346	4.444	6.296	7.901	9.259	10.370	11.235	11.852	12.222	12.346
1,0	396	784	1.164	1.536	1.900	3.600	5.100	6.400	7.500	8.400	9.100	8.600	9.900	10.000
1,5	176	348	517	683	844	1.600	2.267	2.844	3.333	3.733	4.044	4.267	4.400	4.444
2,0	99	196	291	384	475	900	1.275	1.600	1.875	2.100	2.275	2.400	2.475	2.500
2,5	63	125	186	246	304	576	816	1.024	1.200	1.344	1.456	1.536	1.584	1.600
3,0	44	87	129	171	211	400	517	711	833	933	1.011	1.067	1.100	1.111
3,5	32	64	95	125	155	294	416	522	612	686	743	784	808	816
4,0	25	49	73	96	119	225	310	400	469	525	569	600	619	625
4,5	20	39	57	76	94	178	252	316	370	415	449	474	489	494
5,0	16	31	47	61	76	144	204	256	300	336	364	384	396	400
6,0	11	22	32	43	53	100	142	178	208	233	253	267	275	278
7,0	8	16	24	31	39	73	104	131	153	171	186	196	202	204
8,0	6	12	18	24	30	56	80	100	117	131	142	150	155	156
9,0	5	10	14	19	23	44	63	79	93	104	112	119	122	123
10,0	4	8	12	15	19	36	51	64	75	83	91	96	99	100
15,0	2	3	5	7	8	16	23	28	33	37	40	43	44	45
20,0	1	2	3	4	5	9	13	16	19	21	23	24	25	25
25,0	0,6	1	2	2	3	6	8	12	12	13	15	15	16	16

El manejo de la tabla incluye en las columnas distintos valores posibles para los porcentajes p y q. Por otro lado, incluye en las filas distintos valores para el límite de error. El tamaño de la muestra está dado en el cruce de la fila con la columna.

Tabla 2. Población infinita con coeficiente de confianza de 99.7 %

Límites de error en porcentaje	Valores posibles de p y q ($p + q = 100$)													
	1/99	2/98	3/97	4/96	5/95	10/90	15/85	20/80	25/75	30/70	35/65	40/60	45/55	50/50
0,1	89.100	176.400	261.900	345.600	427.850	810.000	1.147.500	1.440.000	1.687.500	1.890.000	2.047.500	2.160.000	2.227.500	2.250.000
0,2	22.275	44.100	65.475	86.400	106.875	202.500	286.875	360.000	421.875	472.500	511.875	540.000	556.875	562.500
0,3	9.900	19.600	29.100	38.400	47.500	90.000	127.400	160.000	187.500	210.000	227.500	240.500	247.500	250.000
0,4	5.569	11.025	16.369	21.600	26.719	50.625	71.719	90.000	105.469	118.125	127.969	135.000	139.219	140.625
0,5	3.564	7.056	10.476	13.824	17.100	32.400	45.000	57.600	67.500	75.600	81.900	86.400	89.100	90.000
0,6	2.475	4.900	7.275	9.600	11.875	22.500	31.875	40.000	46.875	52.500	56.875	60.000	61.875	62.500
0,7	1.818	3.600	5.345	7.053	8.724	16.531	23.418	29.388	34.439	38.571	41.786	44.082	45.459	45.918
0,8	1.392	2.756	4.092	5.400	6.680	12.656	17.930	22.500	27.367	29.531	31.992	33.750	34.805	35.156
0,9	1.100	2.178	3.233	4.267	5.278	10.000	14.167	17.778	20.833	25.278	26.667	26.667	27.500	27.778
1,0	891	1.764	2.619	3.456	4.275	8.100	11.475	14.400	16.875	18.900	20.475	21.600	22.275	22.500
1,5	396	784	1.164	1.356	1.900	3.600	5.100	6.400	7.500	8.400	9.100	9.600	9.900	10.000
2,0	223	441	655	864	1.069	2.025	2.869	3.600	4.219	4.725	5.119	5.400	5.569	5.627
2,5	143	282	419	553	684	1.296	1.836	2.304	2.700	3.024	3.276	3.456	3.564	3.600
3,0	99	196	291	384	475	900	1.275	1.600	1.875	2.100	2.275	2.400	2.475	2.500
3,5	73	144	241	282	349	661	937	1.176	1.378	1.543	1.671	1.763	1.818	1.837
4,0	56	110	164	216	267	506	717	900	1.055	1.181	1.280	1.350	1.392	1.406
4,5	44	87	129	171	211	400	567	711	833	933	1.011	1.067	1.100	1.111
5,0	36	71	105	138	171	324	459	576	675	756	819	864	891	900
6,0	25	49	73	96	119	225	319	400	469	525	569	600	619	625
7,0	18	36	53	71	87	165	234	294	344	386	418	441	455	459
8,0	14	28	41	54	67	127	179	225	364	295	320	338	348	352
9,0	11	22	32	43	53	100	142	178	208	233	263	267	275	278
10,0	9	18	26	35	43	81	115	144	169	189	205	216	223	225
15,0	4	8	12	15	19	36	51	64	75	84	91	96	99	100
20,0	2	4	7	9	11	20	29	36	42	47	51	54	56	56
25,0	1	3	4	6	7	3	18	23	27	30	33	35	36	36
30,0	1	2	3	4	5	9	13	16	19	21	23	24	25	25
35,0	0,7	1	2	3	3	7	9	12	14	15	17	18	18	18
40,0	0,6	1	2	2	3	5	7	9	11	12	13	14	14	14

El manejo de la tabla incluye en las columnas distintos valores posibles para los porcentajes p y q. Por otro lado, incluye en las filas distintos valores para el límite de error.

Tabla 3. Población finita con coeficiente de confianza de 95.5 %

Amplitud del universo	$p = q = 50$					
	± 1%	± 2%	± 3%	± 4%	± 5%	± 10%
—	—	—	—	—	222	83
1.000	—	—	—	385	286	91
1.500	—	—	638	441	316	94
2.000	—	—	714	476	333	95
2.500	—	1.250	769	500	345	96
3.000	—	1.364	811	517	353	97
3.500	—	1.458	843	530	359	97
4.000	—	1.538	870	541	364	98
4.500	—	1.607	891	549	367	98
5.000	—	1.667	909	556	370	98
6.000	—	1.765	938	566	375	98
7.000	—	1.842	949	574	378	99
8.000	—	1.905	976	580	381	99
9.000	—	1.957	989	584	383	99
10.000	5.000	2.000	1.000	588	383	99
15.000	6.000	2.143	1.034	600	390	99
20.000	6.667	2.222	1.053	606	392	100
25.000	7.143	2.273	1.064	610	394	100
50.000	8.333	2.381	1.087	617	397	100
100.000	9.091	2.439	1.099	621	398	100
∞	10.000	2.500	1.111	625	400	100

El manejo de la tabla incluye en las columnas distintos valores posibles para el coeficiente de error. Por otro lado, incluye en las filas distintos valores para la amplitud del universo. Los valores de p y q en todos los casos son de 50. En esta tabla y en la siguiente cuando no aparece dato en el cruce entre fila y columna es porque la amplitud de la muestra es superior a la mitad de la población y no parece viable el muestreo.

Tabla 4. Población finita con coeficiente de confianza de 99.7 %

Amplitud del universo	p = q = 50				
	± 1%	± 2%	± 3%	± 4%	± 5%
500	—	—	—	—	—
1.000	—	—	—	—	474
1.500	—	—	—	726	563
2.000	—	—	—	826	621
2.500	—	—	—	900	662
3.000	—	—	1.364	958	692
3.500	—	—	1.458	1.003	716
4.000	—	—	1.539	1.041	735
4.500	—	—	1.607	1.071	750
5.000	—	—	1.667	1.098	763
6.000	—	2.903	1.765	1.139	783
7.000	—	3.119	1.842	1.171	798
8.000	—	3.303	1.905	1.196	809
9.000	—	3.462	1.957	1.216	818
10.000	—	3.600	2.000	1.233	826
15.000	—	4.091	2.143	1.286	849
20.000	—	4.390	2.222	1.314	861
25.000	11.842	4.592	2.273	1.331	869
50.000	15.517	5.056	2.381	1.368	884
100.000	18.367	5.325	2.439	1.387	892
∞	22.500	5.625	2.500	1.406	900

El manejo de la tabla incluye en las columnas distintos valores posibles para el coeficiente de error. El valor considerado para p y q es de 50. Por otro lado, incluye en las filas distintos valores para la amplitud del universo.

5.2 Hernández Sampieri

Otra forma de tomar el tamaño de una muestra la expone Roberto Hernández.[27] Dada una población de tamaño N, debe preguntarse por el mínimo número de unidades muestrales que deberán ser tomadas en cuenta y que serán n, para que se tenga un error estándar menor a .01.

Al tomar el tamaño de una muestra se desea que el valor de n en una variable, representado por $\bar{y}$, se acerque al valor de la misma variable en N, representado por $\bar{Y}$

El tamaño de la muestra se obtiene en dos pasos. El primero es aplicar la siguiente fórmula:

$$n' = \frac{S^2}{V^2}$$

O lo que es lo mismo:

$$Tamaño\ provisional\ de\ n = \frac{Varianza\ de\ la\ muestra}{Varianza\ de\ la\ población}$$

[27] Hernández, *Metodología*.

Dado que se trata de una medida provisional, debe ajustarse, si se conoce el tamaño de la población, lo cual se hace con la fórmula:

$$n = \frac{n'}{1 + \frac{n'}{N}}$$

A continuación se presentan algunos ejemplos para ilustrar la aplicación de las fórmulas.

Se parte de considerar a todos los directores generales de empresas industriales y comerciales que en 1983 tenían un capital social superior a 30 millones de pesos, además de que sus ventas hayan sido superiores a 100 millones de pesos y/o con más de 300 personas empleadas. Eso dio como resultado a 1, 176 directores generales, es decir, en el mismo número de empresas se cumplían las características enunciadas. La pregunta es a cuántos directores se deberá entrevistar para tener un error estándar menor a 1.5 %.[28]

Es decir:

N Población 1,176

Se Error estándar .015

V^2 Varianza de la población. Su definición (Se) cuadrado del error estándar.

S^2 Varianza de la muestra. Expresada como probabilidad de ocurrencia, es decir, $S^2 = p\ (1\text{-}p)$

Sustituyendo en la fórmula:

$$n' = \frac{S^2}{V^2}$$

Y como,

$$S^2 = p\,(1 - p) = .9\,(1 - .9) = .09$$

$$V^2 = (.015)^2 = .000225$$

Entonces:

$$n' = \frac{S^2}{V^2} = \frac{.09}{.000225} = 400$$

Si ahora se sustituye en la fórmula de ajuste:

[28] El ejemplo se ha adecuado a partir de Hernández, *Metodología*, p. 211. Adopto señalar el error estándar como 1.5 % y no como se hace en el texto citado, 0.015 %. También debe considerarse que en el texto se señala "V" cuando se refiere al error estándar elevado al cuadrado, por lo que debiera señalarse como "V^2".

$$n = \frac{n^{'}}{1 + \frac{n^{'}}{N}}$$

$$n = \frac{n^{'}}{1 + \frac{n^{'}}{N}} = \frac{400}{1 + \frac{400}{1176}} = 298$$

Entonces se deberá entrevistar a 298 directores generales.

Para un ejemplo adicional[29] se plantea que una radiodifusora local desea hacer un estudio para saber cómo usan la radio los adultos de la ciudad: a qué hora la escuchan, qué prefieren, qué opinan del noticiario, etc. La ciudad tiene 2,500,000 adultos, la población a considerar son todos los sujetos, hombres o mujeres, que tengan más de 21 años, y que vivan en una casa o departamento propio o rentado de la ciudad. Se desconoce el número de sujetos con dichas características, por lo que se toma en cuenta un mapa de la ciudad, denotando que tiene 5, 000 cuadras. La pregunta es en cuántas cuadras deberá levantarse la encuesta si se desea un error estándar no mayor a 1.5 % y con una probabilidad de ocurrencia del 50 %.

Entonces se toma en cuenta la fórmula:

$$n^{'} = \frac{S^2}{V^2} = \frac{p\,(1-p)}{.015} = \frac{.5\,(1-.5)}{.000225} = \frac{.25}{.000225} = 1111.11$$

Ajustando la muestra:

$$n = \frac{n^{'}}{1 + \frac{n^{'}}{N}}$$

$$n = \frac{n^{'}}{1 + \frac{n^{'}}{N}} = \frac{1111.11}{1 + \frac{1111.11}{5000}} = 909.0902$$

5.3 Pérez Enríquez

María Isabel Pérez Enríquez en su libro *El impacto de las migraciones y expulsiones indígenas de Chiapas*,[30] elabora una encuesta que abarca, en la aplicación de cuestionarios, los municipios de San Pedro Chenalhó y San Andrés Larráinzar.

Para el “cálculo del muestreo” se sigue la siguiente fórmula:

---

[29] El ejemplo se ha adecuado a partir de Hernández, *op. cit.*, pp. 214-215. Debe entenderse en el texto referenciado que se trabajará con un error estándar no mayor a 0.015 y no, como dice, a 0.15, ya que esa es la notación que utiliza en el otro ejemplo que presenta. Asimismo, al efectuar operaciones con la primera fórmula se obtiene un error estándar de 0.000225 y no, como dice el texto, 0.00025.

[30] Pérez Enríquez, María Isabel, *El impacto de las migraciones y expulsiones indígenas de Chiapas.* UNACH, Tuxtla Gutiérrez, 1998.

$$n/cn = \frac{N\,t_2\,s_2}{N\,d_2 + t_2\,s_2}$$

Para un ejemplo, considérese el siguiente dato, con el fin de obtener el número de cuestionarios que deben aplicarse: en el municipio de San Andrés había 11,050 habitantes.

Además, debe tenerse en cuenta que la autora considera una probabilidad del 95 % del nivel de confianza en relación a la "t" Student, considerando a t = 1.96 y el nivel de precisión de un 20 % del promedio premuestral para la población, donde S_2 es la varianza de la variable. Con la fórmula y sustituyendo valores en ella:

$$n/cn = \frac{N\,t_2\,s_2}{N\,d_2 + t_2\,s_2}$$

$$n/cn = \frac{11,050 * 1.96 * (50 * 50)}{11,050 * 20 + 1.96 * (50 * 50)}$$

$$n/cn = \frac{54,145,000}{225,900} = 239.68 \approx 240$$

Es decir, deben aplicarse 240 cuestionarios en San Andrés.

El texto indica que las variables a considerar para el muestreo son demográficas, económicas, de salud, políticas e ideológicas (educación familiar, escolaridad, religión). Asimismo, considera homogéneas a las comunidades indígenas al interior de los parajes, y dice que se aplicó el muestreo simple-aleatorio cualitativo y cuantitativo, de acuerdo a la asesoría proporcionada por el maestro en Estadística, Guillermo Peláez Gramajo.[31]

En cuanto a la variable cuantitativa considerada apunta:

> Fue necesario ver una variable cuantitativa, que de acuerdo con la relevancia del análisis, estuvo constituida por la tenencia de la tierra, y en particular el número de hectáreas por familia, en contraste con variables cualitativas de orden político, organizativo y religioso.[32]

Por lo tanto, la fórmula es:

$$n/cn = \frac{N\,t_2\,s_2}{N\,d_2 + t_2\,s_2}$$

En donde:

n/cn	Muestra cuantitativa
N	Tamaño de la población
S_2	Considerando la varianza máxima (p = q)
t_2	Distribución "t" de Student (1.96)
d_2	Nivel de precisión o error aceptado.

Por último, debe considerarse que la autora señala que bajo la forma de muestreo que presenta se utiliza la varianza máxima (p = q) y el nivel de confianza de 95 %.

---

[31] Pérez Enríquez, María Isabel, *op. cit.*

[32] *Ibidem*, p. 183.

5.4 Felipe Pardinas

Para estimar el tamaño de la muestra Párdinas[33] utiliza la siguiente fórmula:

$$n = \frac{N\, z_1^2 - \frac{\alpha}{2}\, \sigma^2}{z_1 - \frac{\alpha}{2}\, \sigma^2 \;\; + (N-1)\, e^2}$$

N = tamaño de la población

$z_1 - \frac{\alpha}{2}$ = nivel de confianza de la prueba que diseñamos.

σ^2 = variancia de la población, conocida o estimada por otros estadísticos.

e = imprecisión que podemos tolerar.

En la fórmula Pardinas incluye el término $z^2{}_1 - \frac{\alpha}{2}$, en tanto que después habla del nivel de confianza como $z_1 - \frac{\alpha}{2}$, en realidad se trata del mismo elemento.

Para ejemplificar con la fórmula tómense los datos siguientes:

N = 125
z de .95 = 1.96
e = 1.7
σ^2 = 4

$$n = \frac{N\, z_1^2 - \frac{\alpha}{2}\, \sigma^2}{z_1^2 - \frac{\alpha}{2}\, \sigma^2 \;\; + (N-1)\, e^2} = \frac{125 * 1.96^2 * 4^2}{1.96^2 * 4^2 + 124 * 1.7^2} = \frac{7,683.2}{419.83} = 18.3 \approx 18$$

La fracción $\frac{n}{N}$, aquí $\frac{18}{125}(100) = 14.4\,\%$ es el porcentaje de la población requerido por la muestra.

[33] Pardinas, Felipe, Metodología y técnicas de investigación en ciencias sociales. Siglo XXI editores, México, 1985, p. 178.

Supóngase que se eleva N a 1, 250 (125 * (10))

$$n = \frac{N\ z_1^2 - \frac{\alpha}{2}\ \sigma^2}{z_1^2 - \frac{\alpha}{2}\ \sigma^2\ \ + (N-1)\ e^2} = \frac{1,250 * 1.96^2 * 4^2}{1.96^2 * 4^2 + 1249 * 1.7^2} = \frac{76,832}{3,671.07} = 21$$

$$\frac{n}{N}(100) = \frac{21}{1,250}(100) = 1.68$$

Con esto, Pardinas advierte atinadamente que no debe optarse por establecer un porcentaje de población como tamaño de muestra.

5.5 David Aaker

Para tomar el tamaño de la muestra es necesario conocer las características que tiene una población: la media, la varianza y la desviación estándar.[34]

La población de votantes, de ciudadanos, de habitantes, etc., tiene una media en sus preferencias electorales, sus actitudes políticas y sus valoraciones en general acerca de lo político. Se trata del promedio simple, entendido como la suma de las observaciones dividida entre el número de entidades tomadas en cuenta.

El mismo tipo de población de entidades tiene diferencias en sus opiniones y preferencias, esas diferencias pueden cuantificarse con la varianza, la cual es la sumatoria de las diferencias de cada observación respecto a la media o promedio elevadas al cuadrado, y dividida, la sumatoria, entre el número de entidades menos la unidad. La desviación estándar es la raíz cuadrada de la varianza. Las unidades en las que se mide la varianza están elevadas al cuadrado, dicha elevación responde a la necesidad de evitar que la sumatoria de las diferencias de las observaciones respecto a la media sea cero, por lo que son unidades de medición que no se pueden usar en el análisis, las unidades de desviación estándar evitan ese problema retornando las unidades de expresión a las unidades de medición originales. Si, por ejemplo, la unidad de medida original de las observaciones fueran kilogramos, entonces las unidades de varianza estarán en kilogramos al cuadrado, que es una medida abstracta; para retornar las medidas a su forma original, en kilogramos, es necesario obtener las unidades de desviación estándar.

El problema es cuando la media de la población no es conocida y debe ser estimada a partir de una muestra, por lo que la media, la varianza y la desviación estándar muestrales tienen alguna distancia respecto de esas mismas medidas en la población, y ese es un problema que no puede evitarse cuando por razones de tiempo y recursos económicos debe sacrificarse tener las medidas poblacionales sustituyéndolas por las muestrales.

Desde luego, no todas las muestras posibles de una población tienen las mismas magnitudes en su media, varianza y desviación estándar. Intuitivamente, es razonable pensar que la variación en la media será más grande a medida que la varianza de la población también lo sea. En extremo, si no hay variación en la

[34] Ver Aaker, David y Day, George S., *Investigación de mercados.* Mc Graw-Hill, segunda edición en español, México, 1989. El texto dedica el capítulo 11 especialmente al muestreo.

población, tampoco habrá variación en la media. También es razonable pensar que, a medida que aumenta el tamaño de la muestra, la variación en la media disminuirá. Cuando la muestra es pequeña, se necesitan sólo uno o dos valores extremos para afectar sustancialmente a la media muestral, generando de este modo una media relativamente grande o pequeña. A medida que aumenta el tamaño de la muestra, estos valores extremos tendrán un menor impacto cuando aparezcan, porque serán promediados con más valores. La variación en la media es medida por su error estándar, el cual es:

$$ES_X = \frac{\overline{ES} * X}{\sqrt{n}}$$

ES, es el error estándar.

Entonces, el error estándar de la media muestral depende de n, el tamaño de la muestra, por lo que si n se altera el error estándar cambiará de acuerdo con ello, aun cuando el error estándar de x permanezca inalterado.

La variable X tiene una distribución de probabilidad, la media de la muestra también la tiene. Se acostumbra suponer que la variación de la media muestral de muestra a muestra seguirá la distribución normal, también llamada curva de campana o campana de Gauss. Tal supuesto no es tan extremoso como puede parecerlo. La distribución de las medias muestrales será normal si la distribución de la población es normal, o si el tamaño de la muestra aumenta. La última condición constituye el teorema del límite central.

La media de la muestra es usada para estimar la media de la población desconocida. Debido a que la media de la muestra varía de muestra a muestra, no es igual a la media de la población. Existe un error de la muestra. Es útil proporcionar una estimación de intervalo en torno de la media de la muestra, el cual refleje el juicio acerca del alcance de este error muestral:

Media muestral $\pm$ error de la muestra = estimación de intervalo de la media poblacional

Así:

$$\overline{X} \pm Error \quad muestral \quad , \quad ó \quad \overline{X} = \pm \frac{ZDSX}{\sqrt{n}}$$

Donde:

Z = 2 para un nivel de confianza del 95 %
Z = 5/3 para un nivel de confianza del 90 %
DSX = Desviación estándar de la población (será usada la muestral si la poblacional es desconocida)
n = Es el tamaño de la muestra.

De este modo, el tamaño de la estimación del intervalo dependerá de tres factores. El primero es el nivel de confianza. Si estamos dispuestos a tener menos confianza de que la estimación del intervalo incluya a la media de la población verdadera y desconocida, entonces el intervalo será más pequeño. El segundo factor es la desviación estándar de la población. Si hay poca variación de la población entonces la estimación del intervalo de la media poblacional será más pequeña. El tercero es el tamaño de la muestra. Conforme el tamaño de la muestra aumenta, el error de la muestra se ve reducido y el intervalo se volverá más pequeño.

Mediante esos conceptos es que se puede finalmente especificar cómo se determina el tamaño de una muestra. Para proceder al análisis deben especificarse dos cosas: el tamaño del error de la muestra que se desea y, el nivel de confianza.

Esta especificación dependerá de la intercompensación entre el valor de la información más exacta y el costo por un incremento en el tamaño de la muestra. Para un nivel de confianza dado, un error muestral más pequeño tendrá un costo, en términos de un tamaño de la muestra más grande. Similarmente, para un error de la muestra dado, un error muestral más pequeño tendrá un costo, en términos de un tamaño de la muestra más grande.

Usando la fórmula general para la estimación del intervalo:

$$\overline{X} \pm Error \quad muestral \quad , ó \ \overline{X} \pm \frac{ZDS}{\sqrt{n}}$$

Además, se sabe que:

$$Error \quad de \ la \ muestra \ = \frac{ZDS}{\sqrt{n}}$$

Dividiendo toda la expresión entre el error de la muestra y multiplicando por la raíz cuadrada de n:

$$\sqrt{n} = \frac{ZDS}{Error \quad Muestral}$$

Y elevando al cuadrado ambos lados de la ecuación, obtenemos una expresión para el tamaño de la muestra:

$$n = \frac{Z^2 DS^2}{(Error \quad de \ la \ Muestra)^2}$$

Por lo tanto, si se ha decidido con qué nivel de confianza trabajar, y por consiguiente Z, y si también se ha decidido el error de la muestra que se permitirá, entonces el tamaño de la muestra necesario queda especificado por la fórmula anterior.

El procedimiento que se acaba de mostrar supone que la desviación estándar de la población o de la muestra es conocida. En la mayor parte de las situaciones prácticas, no es conocida y debe ser estimada usando uno de varios enfoques disponibles. Un método es usar una desviación estándar de la muestra obtenida de una encuesta anterior comparable o de una encuesta piloto. Otro enfoque es tomar una situación del “peor de los casos”. Por ejemplo, considerar que la mitad de la población tiene la opinión exactamente opuesta a la que tendría la otra mitad de la población, obviamente, la muestra se torna más grande.

A manera de ejemplo supóngase que se quiere estar 95 % seguros de que el error muestral al estimar la media de la población no exceda de 0.3, además se conoce que la desviación estándar de la población es de 1.49, entonces:

$$n = \frac{Z^2 DS^2}{(Error \quad de \ la \ Muestra)^2}$$

$$n = \frac{Z^2 DS^2}{(Error \ \ de \ la \ Muestra)^2} = \frac{2^2 * 1.49^2}{0.3^2} = \frac{8.8804}{0.09} = 98.67 \approx 99$$

Si se varía el nivel de confianza a 90 % entonces bajará el tamaño de n:[35]

$$n = \frac{Z^2 DS^2}{(Error \ \ de \ la \ Muestra)^2}$$

$$n = \frac{Z^2 DS^2}{(Error \ \ de \ la \ Muestra)^2} = \frac{5/3^2 * 1.49^2}{0.3^2} \frac{6.16694}{0.09} = 68.52 \approx 69$$

Si ahora se varía el error permitido a .5 entonces n bajará, aunque se sostenga el nivel de confianza de 95 %:

$$n = \frac{Z^2 DS^2}{(Error \ \ de \ la \ Muestra)^2}$$

$$n = \frac{Z^2 DS^2}{(Error \ \ de \ la \ Muestra)^2} = \frac{2^2 * 1.49^2}{0.5^2} \frac{8.8804}{0.25} = 35.52 \approx 36$$

En cuanto al tamaño de la población Aaker apunta que el cálculo del tamaño de la muestra es determinado de la misma manera, con la misma fórmula, independientemente de si el tamaño de la población es de 1, 000 o de 100, 000.

5.6 Gómez Aguilar y Ximitl Islas

La fórmula que presenta Iván Ximitl[36] en su trabajo, la cual es tomada de Roberto Gómez,[37] es la siguiente:

$$n = \frac{N \sum_{i=1}^{k} N_i S_i^2}{N^2 V + \sum_{i=1}^{k} N_i S_i^2}$$

Donde:

N Tamaño de la población

S_i^2 Varianza del distrito i

Para otras literales debe considerarse la fórmula:

[35] El resultado que el texto, incorrectamente, señala es 65.5.

[36] Ximitl Islas, Iván, *Organización campesina y comercialización de café: estudio en dos municipios de la sierra norte de Puebla*. Tesis de maestría en Estrategias para el Desarrollo Agrícola Regional, Colegio de Postgraduados, Puebla, 2006.

[37] Gómez Aguilar, Roberto, *Introducción al muestreo*. Tesis de maestría en Ciencias en Estadística, Colegio de Postgraduados, Chapingo, 1979.

$$V = \frac{d^2}{Z^2}$$

La asignación de tamaño de muestra a cada uno de los estratos se hace con la siguiente fórmula:

$$n_i = \frac{N_i}{N} n$$

Como ejemplo podemos considerar que en la referenciada tesis de maestría en Estrategias para el Desarrollo Regional Agrícola del Colegio de Postgraduados, cuyo título es *Organización campesina y comercialización de café: estudio en dos municipios de la sierra norte de Puebla*, Iván Ximitl toma en cuenta, entre sus fuentes primarias, la aplicación de cuestionarios. En Puebla localiza 54,725 predios dedicados al cultivo de café. Hay 5 Distritos de Desarrollo Rural: Huauchinango, Libres, Tehuacan, Teziutlán y Zacatlán. De ese territorio se consideran los municipios en que predominan grupos étnicos: nahuas en Cuetzalan y totonacos en Huehuetla.

La variable que se considera para el muestreo es la superficie por predio en términos de la varianza, tal como aparece en el cuadro siguiente[38]:

Cuadro 55

Medidas de la superficie por predio

Municipio	Promedio	N	Desviación Estándar	Suma	Mínimo	Máximo	Varianza
Cuetzalan del Progreso	0.898	5490	1.232	4931.3	0.2	4.0	1.518
Huehuetla	0.742	1754	0.584	1302.2	0.3	6.0	0.341

Fuente: ver nota 38.

Se considera una precisión de 15 % y una confiabilidad de 90 %, los datos completos se presentan a continuación:

Población	Región	N	=	493
Población estrato 1	Cuetzalan	N_1	=	305
Población estrato 2	Huehuetla	N_2	=	188
Varianza estrato 1	Cuetzalan	$(S^2)_1$	=	1.518
Varianza estrato 2	Huehuetla	$(S^2)_2$	=	0.341
Precisión		d	=	0.15
Confiabilidad		Z	=	1.64

Sustituyendo los valores en la fórmula:

38 El cuadro en la tesis da la siguiente referencia: Ramírez Valverde, Benito. Proyecto: Estudio sobre estrategias para el desarrollo sustentable de la sierra Nor-oriente de Puebla con la participación de productores, organizaciones e instituciones como respuesta a las condiciones de pobreza y marginación. Primer informe semestral. Fondo mixto de investigación científica y tecnológica Conacyt-Gobierno del estado de Puebla. Puebla, México, 2003.

$$n = \frac{N \sum_{i=1}^{k} N_i S_i^2}{N^2 V + \sum_{í=1}^{k} N_i S_i^2} = \frac{493 * [(305 * 1.518) + (188 * 0.341)]}{493^2 * \frac{0.15^2}{1.64^2} + [(305 * 1.518) + (188 * 0.341)]} = 101$$

La asignación de tamaño de muestra a cada uno de los estratos:

$$n_1 = \frac{N_1}{N} n = \frac{305}{493} * 101 = 62$$

$$n_2 = \frac{N_2}{N} n = \frac{188}{493} * 101 = 39$$

Entonces:

62 + 39 = 101

Análisis comparativo

La técnica de Hernández Sampieri contiene dos fórmulas, primero debe obtenerse el tamaño de muestra y luego la muestra ajustada por el tamaño de la población.

En cuanto a los conceptos se utilizan la “varianza de la muestra” y “varianza de la población”. Como se ha dicho, S^2 o sea la “varianza de la muestra”, es tomada en términos de probabilidad de ocurrencia, es decir, la probabilidad de que un evento ocurra es p, la probabilidad de que no ocurra es 1 - p. Esta cuestión es equivalente a lo que Barranco maneja como los “porcentajes” p y q.

Asimismo, V^2 es la “varianza de la población”, pero es tomada para todo efecto práctico, como el cuadrado del error estándar, el cual define el investigador. Se trata del máximo error que se está dispuesto a aceptar, Barranco lo maneja como “error de muestreo” el cual define como el error que se comete por el hecho de extraer un grupo pequeño de individuos de un grupo mayor.

Pérez Enríquez indica que considera solamente la “varianza” máxima, S_2, es decir, cuando la probabilidad de ocurrencia de un fenómeno es de 50 % y la probabilidad de que no ocurra es de 50 %. Asimismo, utiliza sólo una confianza de 95 %. Los términos correspondientes en Barranco son los “porcentajes”, para la “varianza” de Pérez; y el “coeficiente de fiabilidad” para la “confianza”.

Por otro lado, Pérez habla del “nivel de precisión”, d, que es el “error” aceptado en Barranco.

En ambos autores N, simboliza el tamaño de la población, si bien en Pérez se utiliza en su única fórmula, en tanto que en Barranco se utiliza en dos de sus cuatro fórmulas.

Una diferencia fundamental en ambos autores es que Pérez utiliza la distribución t para el nivel de confianza, en tanto que Barranco sólo tiene dos opciones para el nivel de confianza, 99.5 y 99.7 %.

Para Felipe Pardinas “e” es la “imprecisión” que se puede tolerar, en tanto que en Barranco E es el “error” de muestreo. Asimismo, N es la “población” en ambos autores.

También Pardinas utiliza el nivel de confianza $z_1^2 - \frac{\alpha}{2}$ y la varianza σ^2

Un asunto notorio en Pardinas es que cuando presenta el ejemplo que se ha incluido, en el que después establece variaciones en N y se vuelve a calcular el tamaño de n, en el primer momento n es el 14 % de N y, en el segundo momento, es el 1.68 %, él dice que: "Advertimos que dar a n prejuiciadamente un porcentaje fijo de la población no es un buen criterio para calcular n."[39]

En tanto que Pérez dice:

> Aun cuando se muestra una ligera diferencia en cuanto a la población total en ambos municipios, tratamos de encuestar al 10 % del total de las familias, en las comunidades seleccionadas para el estudio. Realizamos 721 encuestas para tratar de precisar las condiciones socioeconómicas de los municipios de origen en donde se dieron las expulsiones...[40]

Es decir, no tiene sentido tomar el tamaño de muestra si se ha establecido prejuiciadamente una cifra.

Ahora bien, Pardinas hace una reflexión muy atinada en torno a la existencia de varias fórmulas de muestreo:

> Existen varias fórmulas para calcular el tamaño de una muestra porque aunque todas ellas demandan como factores un grado de precisión elegido por el investigador –diferencia entre parámetro y estadístico encontrado-, simbolizado por e; el número de veces en que no deseamos equivocarnos dada la distribución de la estimada que buscamos –nivel de confianza-; el error estándar de esa estimada y la prueba que deseamos de uno o dos extremos; las fórmulas del tamaño de la muestra varían, porque varían los parámetros que buscamos.[41]

En Aaker se utiliza el "nivel de confianza", señala dos niveles de desviación estándar, 95 y 90 % con Z igual a 2 y 5/3 respectivamente, en Barranco también se usa el concepto, pero las fórmulas son para 99.7 y 95.5 %. Usa el "error de la muestra" y en Barranco haya su correspondiente en el mismo sentido. Asimismo, usa la "desviación estándar de la población", pero aclara que si no se conoce la poblacional deberá usarse el de la muestra.

El cuestionamiento de la desviación estándar o de la varianza en cuanto a qué referente debe elaborarse la medición, es abordado de la siguiente manera por Aaker:

> Generalmente, un instrumento de encuesta o un experimento no se basará sólo en una pregunta. Algunas veces, cientos de ellas pueden estar involucradas. Normalmente no valdrá la pena pasar por todo este proceso en todas las preguntas. Un enfoque razonable sería tomar unas cuantas preguntas representativas y determinar el tamaño de la muestra a partir de ellas. Deberían incluirse las más cruciales con la varianza esperada más alta.[42]

Es decir, se hace respecto de alguna variable, la cual escoge el investigador.

---

[39] Pardinas, *op. cit.*, p. 180.
[40] Pérez, *op. cit.*, p. 13.
[41] Pardinas, *op. cit.*, p. 178.
[42] Aaker, *op. cit.*, p. 334.

Por otro lado, Aaker corrobora lo dicho por otros autores cuando critican que no debe manejarse un porcentaje de la población como muestra:

> Debería notarse que el cálculo del tamaño de la muestra es completamente independiente del tamaño de la población. Una mala concepción común es que una "buena" muestra debería tener un porcentaje relativamente alto con respecto a la estructura muestral incluida. En realidad, el tamaño de la muestra será determinado de la misma manera, independientemente de si el tamaño de la población es de 1000 o de 100 000.[43]

Entonces, la lógica de cálculo es la misma cuando se trata de una población pequeña o de una grande.

Ximitl utiliza la "población" en el mismo sentido de Barranco. Además utiliza la "varianza" y el tamaño de cada estrato. Asimismo, usa la "precisión" y "confiabilidad"; la primera refiere al nivel de "error" que se permite y, la segunda al "nivel de confianza" en Barranco.

Esta fórmula considera la existencia de estratos, es decir, la población está dividida. Por ello escribe Ximitl:

> Al analizar la información se determinó que el esquema de muestreo adecuado a estos datos era un muestreo estratificado aleatorio, y debido a las diferencias en los tamaños de la población de cada estrato se decidió por la asignación proporcional a cada uno de los municipios.[44]

Entonces, desde el tamaño general de la muestra se toma en cuenta el tamaño de cada uno de los estratos.

Este autor muestra en su trabajo que es posible obtener la varianza a partir de cifras oficiales.

Por su parte Barranco propone una serie de cuatro fórmulas, son sencillas en su manejo y distinguen entre tipos de población, sin tener que ajustar el resultado. Ellas no exigen el conocimiento del tamaño de los estratos, ni siquiera estratificar a la población. Además se han creado tablas que facilitan aún más el manejo de la herramienta estadística de muestreo, sin siquiera usar una fórmula. En los casos de población infinita no se requiere conocer el tamaño de N.

Si bien el formato sólo permite el manejo de dos niveles de confianza, el error aceptado puede variarse en la fórmula.

[43] *Ibidem*, p. 332.
[44] Ximitl, *Organización*, p. 64.

6. ESTADÍSTICA NO PARAMÉTRICA

En ocasiones no se dispone de largas listas con los valores de las variables para cada una de las múltiples entidades en estudio, por lo tanto, no es posible analizar los datos mediante la utilización de herramientas de la estadística paramétrica, como la correlación y la regresión.

Asimismo, muchos estudios, la mayoría, tienen como fuentes de información a las de tipo secundario y, con salvedades disponibles en la red Internet en páginas web gubernamentales, de algunas organizaciones no gubernamentales y, de investigadores universitarios, las cuales ponen a disposición del público listados de datos presentados en forma desagregada, la mayoría de las fuentes secundarias son revistas y libros, que solamente incluyen cuadros con muy pocas filas y columnas. Se trata de información agregada que tiene diferentes formas para presentarse, pero que en el fondo son distribuciones de frecuencias.

Cuando se presenta ese caso, con cuadros que agregan datos, pueden utilizarse herramientas de la estadística no paramétrica, las cuales tienen las siguientes características:

> 1. La mayoría de estos análisis no requieren de presupuestos acerca de la forma de la distribución poblacional. Aceptan distribuciones no normales.
>
> 2. Las variables no necesariamente deben estar medidas en un nivel por intervalos o de razón, pueden analizar datos nominales u ordinales. De hecho, si se quieren aplicar análisis no paramétricos a datos por intervalos o razón, éstos deben ser resumidos a categorías discretas (a unas cuantas). Las variables deben ser categóricas.[45]

Al puntualizar el contenido de un cuadro se busca establecer la relación que tiene el comportamiento de una variable y sus categorías con la otra variable y sus categorías, generalmente esa relación tiene su anclaje en alguna teoría. La relación entre variables puede tener visa en la estadística descriptiva, simplemente se da cuenta de qué valores son mayores y menores, pretendiendo observar relaciones entre ellos. Pero la mejor forma de mostrar relaciones entre las variables es establecer lo significativa o no que sea esa relación, lo que puede hacerse mediante las herramientas que se exponen en este capítulo.

En primer lugar se expone el coeficiente Q de Yule, el cual puede entenderse como una medida de asociación entre variables, análoga a la correlación. Su recorrido va desde -1 a +1. Se dan ejemplos de aplicación a partir de publicaciones consultadas y otros que se han construido para este propósito.

Posteriormente se expone el coeficiente Fi, que es muy parecido al descrito en el párrafo anterior. Tiene el mismo recorrido y finalidad. Se aprovecha esta condición para reutilizar los ejemplos del Q de Yule.

Los coeficientes anteriores son útiles para tablas de 2 por 2, es decir dos filas y dos columnas, dichas tablas tienen dos variables y cada una de ellas se divide en dos categorías. Para cuando se requiere un análisis de significación de la relación entre

[45] Hernández, *op. cit.*, p. 400.

variables que tengan más de dos categorías, entonces se requiere de un análisis más sofisticado, el que se incluye en este capítulo es la X cuadrada.

Se ofrecen al final conclusiones del capítulo y fuentes para el mismo.

6.1 Coeficiente Q de Yule

El coeficiente Q de Yule, cuyo símbolo es una Q, mide el grado de asociación en tablas de dos por dos y el determinante varía entre -1 y +1. La fórmula para calcularlo es la siguiente:

$$Q = \frac{\Delta}{AD + BC}$$

O lo que es lo mismo:

$$Q = \frac{AD - BC}{AD + BC}$$

Análogamente con el coeficiente de correlación si las variables, por ejemplo X y Y, son estadísticamente independientes, entonces el valor del numerador es de cero, $\Delta = 0$, y por tanto el valor del coeficiente también es cero, $Q = 0$.

Por otro lado, el coeficiente alcanza su máximo valor cuando al menos uno de los valores es de cero.

Dicho lo anterior, se entiende que hay algunas distribuciones de frecuencias que dan como resultado un indicador del Q de 1, son los siguientes:

Cuadro 56

Distribución de frecuencias "A" en la que se maximiza el valor de Q de yule

Y/X	X	X´	
Y	A	0	A + B
Y´	C	D	C + D
	A + C	B + D	n

FUENTE: Elaboración del autor.

Cuadro 57

Distribución de frecuencias "B" en la que se maximiza el valor de Q de yule

Y/X	X	X´	
Y	A	B	A + B
Y´	0	D	C + D
	A + C	B + D	n

FUENTE: Elaboración del autor.

En el cuadro 56 se cumplen las siguientes condiciones:

$$A = (A + B)$$

$$B = 0$$

$$D = (B + D)$$

El resultado, al reemplazar esas condiciones en la fórmula de cálculo del coeficiente, es el siguiente:

$$Q = \frac{AD - CB}{AD + CB}$$

$$Q = \frac{(A + B) * (B + D) - C * 0}{(A + B) * (B + D) + C * 0}$$

$$Q = \frac{(A + B) * (B + D) - 0}{(A + B) * (B + D) + 0}$$

$$Q = \frac{(A + B) * (B + D)}{(A + B) * (B + D)}$$

$$Q = \frac{AD}{AD}$$

$$Q = 1$$

Asimismo, en el cuadro 57 se cumplen las siguientes condiciones:

$$A = (A + C)$$

$$C = 0$$

$$D = (C + D)$$

Sustituyendo esas igualdades en la fórmula de cálculo el resultado es el siguiente:

$$Q = \frac{AD - CB}{AD + CB}$$

$$Q = \frac{(A + C) * (C + D) - 0 * B}{(A + C) * (C + D) + 0 * B}$$

$$Q = \frac{(A+C)*(C+D)-0}{(A+C)*(C+D)+0}$$

$$Q = \frac{(A+C)*(C+D)}{(A+C)*(C+D)}$$

$$Q = \frac{AD}{AD}$$

$$Q = 1$$

Entonces, este índice de asociación alcanza su valor máximo si:

$$B = 0$$

O cuando:

$$C = 0$$

También será igual a 1, valor máximo, cuando se cumpla que:

$$B = C = 0$$

Ya que si se parte de la fórmula:

$$Q = \frac{AD - CB}{AD + CB}$$

Y si se sustituyen en esa fórmula los valores de B y C, que es cero, entonces se tiene que:

$$Q = \frac{AD - 0*0}{AD + 0*0}$$

$$Q = \frac{AD}{AD}$$

$$Q = 1$$

El coeficiente adquiere el valor mínimo, es decir -1, cuando una frecuencia de la diagonal principal tiene valor 0.

Cuadro 58

Distribución de frecuencias "C" en la que se minimiza el valor de Q de yule

Y/X	X	X´	
Y	0	B	A + B
Y´	C	D	C + D
	A + C	B + D	n

FUENTE: Elaboración del autor.

Cuadro 59

Distribución de frecuencias "D" en la que se minimiza el valor de Q de yule

Y/X	X	X´	
Y	A	B	A + B
Y´	C	0	C + D
	A + C	B + D	n

FUENTE: Elaboración del autor.

En el cuadro 58 se cumple que:

$$B = A + B$$

$$A = 0$$

$$C = A + C$$

Al sustituir estas igualdades en la fórmula de cálculo se tiene que:

$$Q = \frac{AD - CB}{AD + CB}$$

$$Q = \frac{0 \; * D - (A + C) * (A + B)}{0 \; * D + (A + C) * (A + B)}$$

$$Q = \frac{0 - (A + C) * (A + B)}{0 + (A + C) * (A + B)}$$

$$Q = \frac{0 - (0 + C) * (0 + B)}{0 + (0 + C) * (0 + B)}$$

$$Q = \frac{0 - CB}{0 + CB}$$

$$Q = \frac{-CB}{CB}$$

$$Q = -1$$

En el cuadro 59 se cumple que:

$$C = C + D$$

$$D = 0$$

$$B = B + D$$

Al sustituir dichas igualdades en la fórmula de cálculo se tiene que:

$$Q = \frac{AD - CB}{AD + CB}$$

$$Q = \frac{A \ * 0 - (C + D) * (B + D)}{A \ * 0 + (C + D) * (B + D)}$$

$$Q = \frac{0 - (C + 0) * (B + 0)}{0 + (C + 0) * (B + 0)}$$

$$Q = \frac{-C \ B}{C \ B}$$

$$Q = -1$$

Entonces, este índice de asociación alcanza su valor mínimo si:

$$A = 0$$

O cuando:

$$D = 0$$

También será igual a 1, valor máximo, cuando se cumpla que:

$$A = D = 0$$

Ya que si se parte de la fórmula:

$$Q = \frac{AD - CB}{AD + CB}$$

Y si se sustituyen en esa fórmula los valores de A y D, que es cero, entonces se tiene que:

$$Q = \frac{0*0 - CB}{0*0 + CB}$$

$$Q = \frac{-CB}{CB}$$

$$Q = -1$$

Si ambas frecuencias son nulas entonces también Q= -1.

Los valores potenciales de la medida de asociación del coeficiente Q fluctúan entre -1 y +1:

$$-1 \leq Q \leq 1$$

El valor máximo indica una relación perfectamente directa entre los valores de las variables. El valor mínimo indica máxima relación inversa. El valor 0 indica que hay independencia estadística.

Q es una forma de cerrar el determinante y mide la asociación como lejanía a la independencia.

Las distribuciones de frecuencias de los cuadros 56 y 57 permiten sostener que el coeficiente Q de Yule mide, en una escala limitada por 0 y 1, el grado de aproximación entre la distribución de los datos y las proposiciones empíricas siguientes:

Proposición P_1:

$$X \longrightarrow (Y \text{ o } Y')$$

$$X' \longrightarrow Y'$$

Proposición P_2:

$$X' \longrightarrow Y'$$

$$X' \longrightarrow (Y \text{ o } Y')$$

Cuando Q tiene un valor de 0 o cercano a él pero positivo, entonces hay cercanía a la independencia y se aleja de ambas proposiciones.

Si el coeficiente adquiere un valor grande, cercano a 1, entonces habrá cercanía entre la distribución de frecuencias y la proposición P_1 ó P_2.

El coeficiente Q de Yule no permite al analista poder distinguir su correspondencia entre una proposición y otra, entonces debe analizarse cuidadosamente la distribución de frecuencias para llegar a una conclusión.

Ejemplo

Con el fin de ilustrar lo que se ha dicho en los párrafos inmediatos anteriores, tómese en cuenta la siguiente distribución de observaciones y calcule el Coeficiente Q de Yule.

Cuadro 60
Distribución de frecuencias
bidimensional "A"

Y/X	X	X	
Y	65	6.5	71.5
Y´	26	13	39
	91	19.5	110.5

FUENTE: Elaboración del autor.

$$Q = \frac{AD - CB}{AD + CB}$$

$$Q = \frac{65 * 13 \ - \ 26 * 6.5}{65 * 13 \ + \ 26 * 6.5}$$

$$Q = \frac{845 - 69}{845 + 69}$$

$$Q = \frac{676}{1,014}$$

$$Q = 0.666$$

Ahora se realizará el mismo cálculo para la siguiente tabla de distribución de frecuencias:

Cuadro 61
Distribución de frecuencias
bidimensional "B"

Y/X	X	X´	
Y	65	26	91
Y´	6.5	13	19.5
	71.5	39	110.5

FUENTE: Elaboración del autor.

$$Q = \frac{AD - CB}{AD + CB}$$

$$Q = \frac{65 * 13 \ - \ 6.5 * 26}{65 * 13 \ + \ 6.5 * 26}$$

$$Q = \frac{845 - 69}{845 + 69}$$

$$Q = \frac{676}{1,014}$$

$$Q = 0.666$$

Los resultados son idénticos, sin embargo, la configuración del cuadro 60 tiende a mostrar un mayor grado de concordancia con P_1 que con P_2.

En el caso de que las variables X y Y estén positivamente asociadas tenemos dos proposiciones empíricas cuyo grado de acuerdo con una distribución de frecuencias no se puede determinar sólo a partir del valor de Q por cuanto es incapaz de distinguir entre P_1 y P_2. Por tanto, para juzgar la bondad de ajuste entre una proposición empírica y una distribución de frecuencias específica hay que considerar la configuración de los casos.

Se ejemplificará con el texto de Ana Jaramillo[46] "Movimiento obrero y acumulación de capital: el caso argentino",[47]

Se trata de una investigación que analiza el carácter de la lucha obrera en función del tipo de desarrollo económico. Se plantea que los sectores estratégicos o fundamentales para los diferentes modelos de acumulación son los que adquieren mayor peso político en tanto organización sindical. Generalmente son vanguardia de las luchas sindicales y, por otra parte, tienen mayor poder de negociación con respecto al aparato de dominación estatal.

Se analizará ahora esta relación en el modelo de acumulación agro-exportador implementado desde 1915 hasta 1930 para corroborar la hipótesis.

Todo el sistema de transporte terrestre se construía en función del proyecto exportador. El ferrocarril, que era el medio de transporte más eficaz del momento, se desarrolló construyendo todas las redes hacia las ciudades puerto.

Se intentará demostrar cómo el sector de transportes que constituía el sector económico estratégico del modelo de acumulación del proceso económico, es el que se convierte en vanguardia, en términos políticos. O por lo menos es el que posee mayor poder político y de movilización de sus agremiados.

Vamos a corroborar técnicamente que este sector fue el más combativo; por otra parte, que sus luchas no fueron específicamente salariales, sino que en su mayoría se entablaron por motivos de organización y de mejoras en las condiciones de trabajo.

El artículo dice que a partir de 1930 se produce una reorientación del modelo de acumulación y que hay una clara relación entre el proceso de industrialización y el

[46] En *Revista Mexicana de Ciencias Políticas y Sociales*, no. 89, julio-septiembre, 1977.

[47] El ejemplo ha sido desarrollado en Cortés, Fernando y Rubalcava, Rosa María. *Métodos estadísticos aplicados a la investigación en ciencias sociales. Análisis de asociación*. El Colegio de México, México, 1987.

detrimento de la ganancia en el sector transportes. Otro será el sector económicamente estratégico en el nuevo modelo de acumulación, y el capital británico tratará de vender el sector que hasta ahora había sido la columna vertebral de su dominación.

Este cambio que se produce en la organización y estrategia del capital tiene repercusiones en la organización y estrategia del trabajo. El sector de vanguardia dejará de ser el de transportes, dejará de ser el más conflictivo y la industrialización tomará su lugar, trayendo como consecuencias fundamentales: incorporación masiva de fuerza de trabajo a la industria y, por otra parte, el proceso de urbanización.

Los sectores de sustitución de importaciones en los cuales se basaba la industrialización, pasan a ser los más conflictivos. Analizando entonces el periodo 1915-1930 y el posterior 1930-l945, se explicará el papel de los sectores estratégicos de la economía y su relación con el proceso de acumulación y el proceso político sindical. Se analizará a partir de su forma específica de lucha: el conflicto, la huelga.

La autora sostiene que debe haber una relación entre "hegemonía económica" y "proceso político sindical", el análisis histórico del tipo de desarrollo que ha seguido Argentina le permite concluir que hay un punto de inflexión al inicio de la cuarta década del siglo XX, cuando se pasó de un modelo orientado hacia afuera a otro volcado hacia el mercado interno; por tanto, para que la distribución de las observaciones concuerde con la proposición debe haber un cambio en los sectores específicos de donde provienen los obreros que conforman la vanguardia.

Se define como variable explicativa (X) a las ramas de la actividad productiva y supone que ella sólo puede asumir dos valores: ferrocarriles y vestido. El primero es el sector hegemónico indiscutido en el modelo agroexportador y en el segundo descansa el ideario del modelo industrializador.

Como en el texto se sostiene que el proceso político sindical se objetiva a través de la lucha: el conflicto, la huelga, la variable explicada o dependiente (Y), será el tipo de lucha, categorizada según su naturaleza en salarial y organizacional.

Tenemos entonces que la variable independiente es "ramas de la actividad productiva", con ferrocarriles y vestido, y que la variable explicada es "conflictos laborales", dicotomizada en salariales y organizacionales.

La proposición teórica dice que los conflictos desarrollados por las organizaciones obreras de las actividades hegemónicas obedecen fundamentalmente a motivos de organización.

En función de la hipótesis manejada se define el modelo:

$$X \quad \begin{cases} B & \text{Ferrocarriles} \\ B' & \text{Vestido} \end{cases}$$

$$Y \quad \begin{cases} A & \text{Salario} \\ A' & \text{Organización} \end{cases}$$

Entonces la proposición empírica válida para el periodo 1915-30 será:

$$B \longrightarrow A'$$

$$B' \longrightarrow A \text{ ó } A'$$

Con el coeficiente Q de Yule se realizará el análisis. En el estudio de referencia se encuentra la distribución del número de huelgas, según sus causas (tipo de lucha) por ramas de la actividad productiva (hegemonía sectorial) en Argentina, 1919:

Cuadro 62

		Rama de actividad productiva		
	Y/X	Ferrocarriles	Vestido	Total
Causas	Salario	2	53	55
	Organización	28	18	46
	Total	30	71	101

FUENTE: Construido a partir de: Cortés, Fernando y Rubalcava, Rosa María. *Métodos estadísticos aplicados a la investigación en ciencias sociales. Análisis de asociación*. El Colegio de México, México, 1987.

Si se observa cuidadosamente el cuadro puede notarse que la distribución de conflictos por rama es mucho mayor en el sector del vestido, que es el no hegemónico.

Por otro lado, la distribución por causas es equilibrado, solamente un poco mayor en la reivindicación salarial.

La proposición empírica dice que la asociación entre X y Y debe medirse a través de Q y su valor debe ser negativo:

$$Q = \frac{AD - CB}{AD + CB}$$

$$Q = \frac{2 * 18 \; - \; 53 * 28}{2 * 18 \; + \; 53 * 28}$$

$$Q = \frac{36 - 1484}{36 + 1484}$$

$$Q = \frac{1, \quad 520}{1,448}$$

$$Q = -\,0.9526$$

El resultado es cercano al -1. Se trata de una relación indirecta. Hay una muy grande correspondencia entre la propuesta empírica y la relación intrínseca a los datos: los conflictos en los ferrocarriles son básicamente de índole organizacional.

Un complemento a este análisis, que daría un carácter más concluyente, sería que, cuando se hubiera producido el traslado al modelo de acumulación basado en el desarrollo del mercado interno, se estableciera el valor de Q de Yule, y la relación se esperaría que fuera directa y fuerte.

6.2 Coeficiente fi (ϕ)

Se trata de un coeficiente cuyos valores oscilan entre -1 y +1, lo que logra normalizando la relación con la raíz cuadrada del producto de las frecuencias marginales:

$$\phi = \frac{\Delta}{\sqrt{(A+B)\,(C+D)\,(A+C)\,(B+D)}}$$

Lo que es:

$$\phi = \frac{A\,D - B\,C}{\sqrt{(A+B)\,(C+D)\,(A+C)\,(B+D)}}$$

Si las variables X y Y son estadísticamente independientes, entonces el numerador será de 0 y el coeficiente Fi también, $\Delta = 0$ y $\phi = 0$.

Si la distribución de frecuencias es tal que todas las observaciones se ubican en la diagonal principal, entonces se tiene lo siguiente:

Cuadro 63

Distribución de frecuencias en la que se

Maximiza el valor de Fi

Y/X	X	X´	
Y	A	0	A + B
Y´	0	D	C + D
	A + C	B + D	n

FUENTE: Elaboración del autor.

Y se cumplen las siguientes propuestas:

$$A = (A+B) = (A+C)$$

$$D = (C+D) = (B+D)$$

$$B = C = 0$$

Efectuando el reemplazo de esas igualdades en la ecuación del coeficiente Fi:

$$\phi = \frac{A\,D - B\,C}{\sqrt{(A+B)\,(C+D)\,(A+C)\,(B+D)}}$$

$$\phi = \frac{A\,D - 0 * 0}{\sqrt{A\,D\,A\,D}}$$

$$\phi = \frac{A\,D}{A\,D}$$

$$\phi = 1$$

Es tanto como si se tuviera X en el numerador y en el denominador $\sqrt{X^2}$, de lo que queda $\frac{X}{X} = 1$

Entonces puede concluirse que cuando las frecuencias de la diagonal secundaria son nulas, el coeficiente ϕ alcanza su valor máximo.

Supóngase ahora que las frecuencias de la diagonal principal son nulas:

Cuadro 64

Distribución de frecuencias en la que se

Minimiza el valor de Fi

Y/X	X	X´	
Y	0	B	A + B
Y´	C	0	C + D
	A + C	B + D	n

FUENTE: Elaboración del autor.

Las frecuencias de esta tabla satisfacen las siguientes relaciones:

$$C = (A + C) = (C + D)$$

$$B = (A + B) = (B + D)$$

$$A = D = 0$$

Efectuando el reemplazo de esas igualdades en la ecuación del coeficiente Fi:

$$\phi = \frac{A\,D - B\,C}{\sqrt{(A + B)\,(C + D)\,(A + C)\,(B + D)}}$$

$$\phi = \frac{0 * 0 - B\,C}{\sqrt{B\,C\,B\,C}}$$

$$\phi = \frac{-B\,C}{B\,C}$$

$$\phi = -1$$

Entonces, el coeficiente ϕ asume su valor más pequeño cuando los casos se ubican sobre la diagonal secundaria y en la principal sólo hay valores nulos.

Se tiene entonces que:

$$-1 \leq \phi \leq 1$$

Cuando el valor es de 1 hay una máxima relación directa o positiva.

Cuando el valor es de -1 la relación es inversa o negativa.

Valores cercanos a 0 indican ausencia de asociación.

Cuando hay máxima relación directa deducimos que se trata del valor más alejado de la independencia estadística, pero no se sabe muy bien a qué se aproxima por cuanto no se ha definido conceptualmente el sentido de los límites del coeficiente.

Con base en la distribución de frecuencias del cuadro 63, el coeficiente ϕ mide, en una escala que va de 0 a 1, el grado de aproximación entre la proposición empírica:

$$P_3: \quad \begin{array}{ccc} X & \longrightarrow & Y \\ X' & \longrightarrow & Y' \end{array}$$

En la distribución efectiva de los datos.

Valores de ϕ positivos y cercanos a cero indican no sólo independencia estadística sino también lejanía respecto a P_3.

Los valores cercanos a 1 indican alejamiento de la independencia, pero también un alto grado de adecuación entre P_3 y la distribución bidimensional de los datos.

Un examen similar, pero esta vez aplicado al cuadro 20 permite afirmar que los valores de ϕ entre 0 y -1 indican el grado de adecuación entre P_3 y la distribución bidimensional de frecuencias (ϕ=-1).

Las relaciones son las siguientes:

$$\phi = \begin{cases} 1 & \text{Si las observaciones se distribuyen según:} \\ & P_3: \quad \begin{array}{ccc} X & \longrightarrow & Y \\ X' & \longrightarrow & Y' \end{array} \\ 0 & \text{Si X y Y son estadísticamente independientes} \\ -1 & \text{Si los casos se distribuyen de acuerdo con:} \\ & P_4: \quad \begin{array}{ccc} X & \longrightarrow & Y' \\ X' & \longrightarrow & Y \end{array} \end{cases}$$

El uso de este coeficiente puede ejemplificarse considerando el siguiente texto: Boudon, Raymond. *Education, opportunity and social inequality: changing prospects in western society*. John Wiley and Sons, Nueva York, 1974.[48]

[48] El ejemplo ha sido desarrollado en Cortés, *Métodos*.

Raymond Boudon, dice:

> Consideremos el siguiente problema. En la perspectiva generalmente aceptada de que las sociedades industriales son ampliamente meritocráticas, el nivel educativo alcanzado es uno de los principales determinantes del status. A partir de esto, es grande la tentación de concluir que el logro educativo debería ser también el principal factor de movilidad social. En otras palabras, si el logro educativo es realmente un poderoso determinante del status, la probabilidad de que un individuo tenga mayor status que su padre debería ser mayor, en la medida que su nivel educativo es mayor. Inversamente, la probabilidad de que un individuo caiga a un status más bajo que el de su padre debería ser mayor, en tanto menor sea su nivel educativo.[49]

Se mide la variable logro educativo en dos dimensiones: el nivel educativo del hijo respecto del padre, cuando es mayor y, cuando es igual o menor con relación al del padre, lo que se llamará variable X. Por otra parte, el status socioeconómico tiene las siguientes dos dimensiones: el status del hijo respecto del padre, cuando es mayor y, cuando es igual o menor, que se llamará variable Y. Las variables se dicotomizan de la siguiente manera:

$$X = \begin{cases} B & \text{Mayor nivel educacional del hijo en relación con el del padre} \\ B' & \text{Igual o menor nivel educacional del hijo en relación con el del padre} \end{cases}$$

$$Y = \begin{cases} A & \text{Mayor status socioeconómico del hijo con relación al del padre} \\ A' & \text{Igual o menor status socioeconómico del hijo con relación al del padre.} \end{cases}$$

La propuesta de Boudon que se deriva de su planteamiento es el siguiente:

$$B \longrightarrow A$$

$$B' \longrightarrow A'$$

Es decir, en las sociedades industriales modernas si los hijos alcanzan un nivel educativo mayor que el del padre, entonces su status socioeconómico también será mayor. Si, por el contrario, logran un grado de estudio, menor o igual, dicho status debería ser igual o menor que el del padre.

La información que se utilizará para verificar la concordancia entre la distribución de los datos y la proposición empírica fue construida a partir del siguiente texto: Centers, Robert. "Education and Occupational Mobility", en American Sociological Review, No. 14, 1949, Pp. 143-144, y muestra el status socioeconómico del hijo con respecto al del padre como función de su nivel educativo en relación con el del padre.

[49] Boudon, Raymond. Education, opportunity and social inequality: changing prospects in western society. John Wiley and Sons, Nueva York, 1974, p. 3.

Cuadro 65

Nivel educacional y status del hijo con relación al padre

	Nivel educacional del hijo en relación con el del padre			
	Y/X	Mayor	Igual o Menor	Total
Status socioeconómico del hijo en relación con el del padre	Mayor	134	30	164
	Igual o Menor	157	95	252
		291	125	416

FUENTE: Reelaboración a partir de: Centers, Robert. "EDUCATION AND OCCUPATIONAL MOBILITY", en *American Sociological Review,* No. 14, 1949

Se iniciará el análisis con el examen de las distribuciones marginales para luego establecer el grado de relación entre X y Y. La variable X muestra el resultado de la expansión del sistema educativo en la sociedad industrial del siglo XX: la escolaridad ha aumentado continuamente, lo que conduce a que una alta proporción (70 %) de los hijos tenga un nivel de educación mayor que el de sus padres.

Un fenómeno totalmente opuesto ocurre con el status socioeconómico; pareciera que el desarrollo de los Estados Unidos proporciona o pone a disposición de la sociedad una cantidad decreciente de ubicaciones sociales mejores para los hijos que para los padres.

El proceso educativo sigue facultando para obtener posiciones sociales mejores, pero la sociedad económica no ofrece mayores espacios. Hay 291 hijos que superan el nivel educativo de los padres, pero de ellos solamente 164 se ubican en un status mejor que el de sus padres.

El análisis de marginales alerta en el sentido de que no puede haber una correspondencia perfecta entre la proposición empírica y la distribución de los datos al no cumplirse el requisito aritmético de que:

N(A) = N(B)   Y   N(A´) = N(B´)

El proceso de cálculo del coeficiente ϕ es el siguiente:

$$\phi = \frac{A\,D - B\,C}{\sqrt{(A+B)\,(C+D)\,(A+C)\,(B+D)}}$$

$$\phi = \frac{134 * 95 - 30 * 157}{\sqrt{164 * 252 * 291 * 125}}$$

$$\phi = \frac{12,730 - 4,710}{38,772.49}$$

$$\phi = \frac{8,020}{38,772.49}$$

$$\phi = 0.21$$

El valor del coeficiente es bajo, entonces parece que el nivel educativo no es importante para explicar la movilidad social.

No obstante, la meritocracia no es invalidada por el resultado anterior: la educación puede ser a nivel individual la principal forma en la que se trata de forjar una mejora en el status, pero a nivel de la sociedad económica se disponen de menos espacios de los que se requieren para dar cabida a la población que ha mejorado el nivel educativo respecto del de sus padres, por lo que deben conformarse con posiciones sociales menores o iguales que las de sus padres.

6.3 X cuadrada (X^2)

Ésta es una prueba de índole no paramétrica y, por lo tanto, evalúa la relación entre dos variables categóricas. El símbolo de la prueba es X^2. El sentido es correccional y bivariable.

La materia prima para esta prueba es una tabla de contingencia o de tabulación cruzada. Se trata de una tabla bidimensional, en cada una de las dimensiones, filas y columnas, se registran las observaciones cruzadas, es decir, que inciden a la vez en ambas dimensiones. A su vez, cada variable se subdivide en dos o más categorías.

Un ejemplo de tabla de contingencia o tabulación cruzada es la siguiente:

Cuadro 66

Orientación del voto por sexo

		Orientación del voto	
		Candidato A	Candidato B
Sexo	Masculino	37	54
	Femenino	48	32

FUENTE: Elaboración del autor. Son datos ficticios solamente para ejemplificar.

Las variables, *sexo* y *orientación* del voto se señalan en la parte izquierda y en la parte alta de la tabla, de cada una de ellas se desprenden sus respectivas categorías, para sexo son masculino y femenino y, para la orientación del voto, son el voto por el candidato A y voto por el candidato B.

En este caso se trata de una tabla de dimensión 2 * 2 (fila * columna), donde cada dígito significa una variable y el valor de éste indica el número de categorías de la variable:

2 * 2

↓ ↓

Variable X, con dos categorías

Variable Y, con dos categorías

En la tabla de contingencia, materia prima obtenida a partir de la realidad concreta, se anotan las frecuencias observadas. La siguiente tabla tiene dimensiones de 3 * 3, tómese a manera de ejemplo:

Cuadro 67

Orientación del voto según zona de distrito

		Zona a la que pertenece el distrito electoral			
		Norte	Centro	Sur	Total
Orientación del voto	Partido de Derecha	90	50	72	212
	Partido de Centro	95	140	118	353
	Partido de Izquierda	82	60	71	213
	Total	267	250	261	778

FUENTE: Elaboración del autor. Son datos ficticios solamente para ejemplificar.

El procedimiento de cálculo de la X^2 implica una comparación entre la tabla de contingencia real y la tabla de frecuencias esperadas, la cual está constituida por los valores que se tendrían si las variables fueran estadísticamente independientes, es decir, no estuvieran relacionadas.

Esta prueba presupone que las variables no están relacionadas y entonces trata de evaluarse si es así o no, esto se logra analizando si las frecuencias observadas, las de la Tabla de contingencia real, son diferentes de lo que pudiera esperarse en caso de ausencia de correlación.

La reflexión es la siguiente:

> Si no hay relación entre las variables, debe tenerse una tabla como la de frecuencias esperadas. Si hay relación el resultado de la investigación debe ser muy diferente a la tabla de frecuencias esperadas.[50]

La fórmula con la que se calculan las frecuencias esperadas de cada frecuencia observada es la siguiente:

$$fe = \frac{Total\ de\ la\ \ fila * Total\ de\ la\ columna}{N}$$

Donde N es el número total de frecuencias observadas.

Ya que se han obtenido todas las frecuencias esperadas, constituyentes de la Tabla de frecuencias esperadas, se procede a aplicar la fórmula para obtener el valor de X^2:

$$X^2 = \sum \frac{(O-E)^2}{E}$$

Donde O, es la frecuencia observada en cada celda; E, es la frecuencia esperada en cada celda.

Entonces, de cada frecuencia observada se resta la frecuencia esperada, ese cálculo se lleva a cabo para todas las frecuencias; luego se elevan al cuadrado los resultados de esas restas; posteriormente se divide cada una de las restas entre el valor esperado; por último, se suman los resultados. El número resultante es el valor de X^2.

La X^2 procede de una distribución muestral, denominada distribución X^2, por otro lado, los resultados obtenidos con el proceso expuesto están identificados por los

[50] Hernández, *Metodología*, pp. 402-403.

grados de libertad; entonces, para saber si un valor de X^2 es o no significativo, se deben calcular los grados de libertad. Estos se obtienen mediante la siguiente fórmula:

$$Gl = (f-1)*(c-1)$$

Donde f es el número de filas de la tabla de contingencia y, c es el número de columnas.

Con los grados de libertad correspondientes se acude a la Tabla de distribución de X^2 que se encuentra al final del texto.

Se elige un nivel de confianza, .05 o .01. Entonces, si el valor calculado de X^2 es igual o superior al de la tabla, las variables están relacionadas y X^2 es significativa.

Puntualizando el procedimiento:

1. Se parte de una tabla de frecuencias observadas
2. Se obtiene la tabla de frecuencias esperadas
3. Se obtiene el valor de X^2
4. Se obtienen los grados de libertad
5. En la tabla de distribución de X^2 se elige un nivel de confianza y se obtiene el "valor de la tabla" con los grados de libertad
6. Se compara el Valor de X^2 con el "valor de la tabla", entonces:

Cuadro 68

Criterio de evaluación de X cuadrada

Valor de X cuadrada	Valor en tablas	Relación Estadística	Tesis
Menor	Mayor	No significativa	Variables no relacionadas
Mayor	Menor	Significativa	Variables relacionadas

FUENTE: Elaboración del autor.

A manera de ejemplo de aplicación se toman los siguientes datos. Provienen de una contienda electoral. Se presentan categorizados. Por un lado, la variable "Zona del distrito electoral" se ha categorizado en Norte y Sur; por otro lado, la variable "Identificación política", se ha categorizado en Partido derechista, Partido del centro y Partido izquierdista.[51]

Se desea saber si las variables están relacionadas, como para poder explicar a una con la otra o, por el contrario, si hay ausencia de relación.

Evalúe, entonces, la relación entre las variables, mediante la X^2.

La tabla de Contingencia:

Cuadro 69

Tabla de contingencia de frecuencias observadas

	Zonal del distrito electoral		
	Norte	Sur	Total
Partido Derechista	180	100	280
Partido del Centro	190	280	470
Partido Izquierdista	170	120	290
Total	540	500	1040

[51] El ejemplo se desarrolla en: Hernández, *op. cit.*, pp. 401-406.

FUENTE: Elaboración del autor.

Con la aplicación de la fórmula para obtención de las frecuencias esperadas se obtiene la siguiente:

Cuadro 70

Tabla de contingencia de frecuencias esperadas

	Zonal del distrito electoral		
	Norte	Sur	Total
Partido Derechista	145.4	134.6	280
Partido del Centro	244.0	226.0	470
Partido Izquierdista	150.6	139.4	290
Total	540	500	1040

FUENTE: Elaboración del autor.

Con la aplicación de la fórmula para la obtención del valor de X^2:

Cuadro 71

CELDA	O	E	O-E	(O-E)2	(O-E)2/E
ZN/PD	180	145.4	34.6	1197.2	8.23
ZN/PC	190	244.4	-54.4	2959.4	12.11
ZN/PI	170	150.6	19.4	376.4	2.50
ZS/PD	100	134.6	-34.6	1197.2	8.89
ZS/PC	280	226.0	54.0	2916.0	12.90
ZS/PI	120	139.4	-19.4	376.4	2.70
				X2	47.34

FUENTE: Elaboración del autor.

Ahora se calculan los grados de libertad:

$$Gl = (f-1)*(c-1)$$

$$Gl = (3-1)*(2-1)$$

$$Gl = 2*1$$

$$Gl = 2$$

El valor en la tabla correspondiente es:

Cuadro 72

Valor en tablas de ejemplo 1		
Nivel de confianza		
Grados libertad	0.05	0.01
2	5.991	9.21

FUENTE: Elaboración del autor.

En ambos casos, dos niveles de confianza, X^2 es mayor al valor en Tablas. La relación es significativa, las variables están relacionadas, por lo tanto, puede considerarse que hay una potencial causalidad entre las variables y otras pruebas estadísticas que pudieran utilizarse serían más concluyentes.

Un segundo ejemplo con datos no hipotéticos, que provienen de una encuesta aplicada en el Estado de Puebla tres meses antes de las elecciones federales de 2000 para la presidencia de la república. Se presentan categorizados y en términos porcentuales. Por un lado, la variable "Partidos políticos" se ha categorizado en PAN, PRI y PRD; por otro lado, la variable "Edad", se ha categorizado en, primero, electores que tengan entre 18 y 30 años, segundo, electores que tengan entre 31 y 42 años, los que tengan entre 43 y 52 años y, por último, electores de 53 o más años.[52]

Se desea saber si las variables están relacionadas como para poder explicar a una con la otra.

Evalúe, entonces, la relación entre las variables, mediante la X^2.

La tabla de inicio se presenta en el cuadro 73.

Cuadro 73

Tabla de contingencia de frecuencias observadas

	Partidos políticos			
Edad	PAN	PRI	PRD	Total
18-30	43.45	32.52	8.35	84.32
31-42	43.85	31.57	5.26	80.68
43-52	33.33	59.25	3.79	96.37
53-…	49.01	40	5.45	94.46
TOTAL	169.64	163.34	22.85	355.83

FUENTE: Elaboración del autor.

La tabla de frecuencias esperadas:

Cuadro 74

Tabla de contingencia de frecuencias esperadas

	Partidos políticos			
Edad	PAN	PRI	PRD	Total
18-30	40.20	38.71	5.41	84.32
31-42	38.46	37.04	5.18	80.68
43-52	45.94	44.24	6.19	96.37

[52] La encuesta fue llevada a cabo en las cabeceras distritales locales del estado de Puebla, con un nivel de confianza de 95.5 %, un error de muestreo de 2.5 %, y asignación de cuotas por el método proporcional. Bajo la dirección del autor.

53-...	45.03	43.36	6.07	94.46
Total	169.64	163.34	22.85	355.83

FUENTE: Elaboración del autor.

El cálculo del valor de X^2:

Cuadro 75

CELDA	O	E	O-E	(O-E)2	(O-E)2/E
PAN1	43.45	40.20	3.25	10.57	0.26
PAN2	43.85	38.46	5.39	29.01	0.75
PAN3	33.33	45.94	-12.61	159.11	3.46
PAN4	49.01	45.03	3.98	15.81	0.35
PRI1	32.52	38.71	-6.19	38.27	0.99
PRI2	31.57	37.04	-5.47	29.87	0.81
PRI3	59.25	44.24	15.01	225.37	5.09
PRI4	40.00	43.36	-3.36	11.30	0.26
PRD1	8.35	5.41	2.94	8.62	1.59
PRD2	5.26	5.18	0.08	0.01	0.00
PRD3	3.79	6.19	-2.40	5.75	0.93
PRD4	5.45	6.07	-0.62	0.38	0.06
				SUMA	14.57

FUENTE: Elaboración del autor.

El cálculo de los grados de libertad:

$$Gl = (f-1)*(c-1)$$

$$Gl = (4-1)*(3-1)$$

$$Gl = 3*2$$

$$Gl = 6$$

El valor en la tabla correspondiente es:

Cuadro 76

Valor en tablas de ejemplo 2		
Nivel de confianza		
Grados de libertad	0.05	0.01
6	12.592	16.812

FUENTE: Elaboración del autor.

Las variables están relacionadas a un nivel de confianza de .05, y no están relacionadas al nivel de confianza de .01, lo que indica que no está garantizado que su relación sea causal, pero tampoco pueden desecharse.

Un ejemplo adicional es construido a partir de: Bizberg Guter, Ilán. ”La acción obrera en Las Truchas.” Tesis de maestría en ciencias políticas, CEI, El Colegio de México, Enero 1981.

Se trata de una encuesta aplicada a los trabajadores del sindicato de la siderúrgica “Las Truchas”, algunos de los cuales provenían de trabajos que desempeñaban en el sector agrícola, otros provenían de lugares urbanos y no tenían experiencia en el trabajo agrícola. Por otro lado, se incluye la variable calificación en el trabajo, es decir, la calificación que tienen en su trabajo de “Las truchas”.

Las variables son la “Experiencia en el trabajo agrícola”, variable categorizada en Sí tienen experiencia y, No tienen experiencia; la otra variable es la “Calificación en el trabajo anterior”, categorizada en Calificado, Semicalificado y No calificado.

Se desea saber si las variables están estadísticamente relacionadas como para poder explicar a una con la otra.

Evalúe, entonces, la relación entre las variables, mediante la X^2.

La tabla de contingencia de frecuencias observadas de la cual parte el análisis se representa en el cuadro 77.

Cuadro 77

Tabla de contingencia de frecuencias observadas

	Experiencia en el trabajo agrícola		
CEET ACTUAL	SI	NO	Total
CALIFICADO	14	35	49
SEMICALIF.	25	40	65
NO CALIF.	61	25	86
TOTAL	100	100	200

FUENTE: Elaboración del autor.

CEET. Calificación en el trabajo.
SEMICALIF. Semicalificado.
NO CALIF. No calificado.

La tabla resultante de la aplicación de la fórmula de cálculo de frecuencias esperadas es:

Cuadro 78

Tabla de contingencia de frecuencias esperadas

	Experiencia en el trabajo agrícola		
CEET	SI	NO	Total
CALIFICADO	24.5	24.5	49
SEMICALIF.	32.5	32.5	65
NO CALIF.	43.0	43.0	86
TOTAL	100	100	200

FUENTE: Elaboración del autor.

Calculando el valor de la X^2:

Cuadro 79

CELDA	O	E	O-E	(O-E)2	(O-E)2/E
C-S	14	24.50	-10.50	110.25	4.50
C-N	35	24.50	10.50	110.25	4.50
S-S	25	32.50	-7.50	56.25	1.73
S-N	40	32.50	7.50	56.25	1.73
N-S	61	43.00	18.00	324.00	7.53
N-N	25	43.00	-18.00	324.00	7.53

27.53

FUENTE: Elaboración del autor.

El cálculo de los grados de libertad:

$$Gl = (f-1)*(c-1)$$

$$Gl = (3-1)*(2-1)$$

$$Gl = 2*1$$

$$Gl = 2$$

El valor en la tabla correspondiente es:

Cuadro 80

Valor en tablas de ejemplo3		
Nivel de confianza		
Grados de libertad	0.05	0.01
2	5.991	9.21

FUENTE: Elaboración del autor.

Las variables están relacionadas por lo que su relación es potencialmente causal. Las variables podrían explicarse la una con la otra.

Ahora considérese el caso construido a partir de: Madrazo, Julio y Owen, Diana. "1994 ¿Quién votó por cuál partido y por qué?", en *Nexos*, núm. 202, octubre de 1994, pp. 20-23.

Se trata de los resultados de una encuesta "exit poll" realizada por la Cámara Nacional de la Industria de la Radio y la Televisión y por Warren Mitofsky Internacional. Desafortunadamente, no se dan más datos de la técnica de encuesta, sin embargo, se supone que se trata de una muestra nacional en las casillas instaladas en el país.

Las variables son "Preferencias electorales", categorizada en PAN, PRI y PRD; la otra variable es el "Nivel de escolaridad", categorizada en Ninguna, Primaria, Secundaria, Preparatoria y, Universidad o más.

Se desea saber si las variables están estadísticamente relacionadas como para poder explicar a una con la otra.

Evalúe, entonces, la relación entre las variables, mediante la X^2.

La tabla de inicio es la siguiente:

Cuadro 81

Tabla de contingencia de frecuencias observadas

		Preferencias electorales			
		PAN	PRI	PRD	
Nivel de escolaridad	Ninguno	7	14	9	30
	Primaria	31	42	35	108
	Secundaria	20	18	18	56
	Preparatoria	21	13	19	53
	Universidad o +	21	13	20	54

100	100	101	301

FUENTE: Elaboración del autor.

Las frecuencias esperadas:

Cuadro 82

Tabla de contingencia de frecuencias esperadas

		Preferencias electorales			
		PAN	PRI	PRD	
Nivel de escolaridad	Ninguno	9.97	9.97	10.07	30
	Primaria	35.88	35.88	36.24	108
	Secundaria	18.60	18.60	18.79	56
	Preparatoria	17.61	17.61	17.78	53
	Universidad o +	17.94	17.94	18.12	54
		100	100	101	301

FUENTE: Elaboración del autor.

El cálculo del valor de X^2:

Cuadro 83

CELDA	O	E	O-E	(O-E)2	(O-E)2/E
N-PAN	7	9.97	-2.97	8.80	0.88
P-PAN	31	35.88	-4.88	23.82	0.66
S-PAN	20	18.60	1.40	1.95	0.10
P-PAN	21	17.61	3.39	11.51	0.65
U-PAN	21	17.94	3.06	9.36	0.52
N-PRI	14	9.97	4.03	16.27	1.63
P-PRI	42	35.88	6.12	37.45	1.04
S-PRI	18	18.60	-0.60	0.37	0.02
P-PRI	13	17.61	-4.61	21.23	1.21
U-PRI	13	17.94	-4.94	24.41	1.36
N-PRD	9	10.07	-1.07	1.14	0.11
P-PRD	35	36.24	-1.24	1.54	0.04
S-PRD	18	18.79	-0.79	0.63	0.03
P-PRD	19	17.78	1.22	1.48	0.08
U-PRD	20	18.12	1.88	3.54	0.20
					8.56

FUENTE: Elaboración del autor.

El cálculo de los grados de libertad:

$$Gl = (f-1)*(c-1)$$

$$Gl = (5-1)*(3-1)$$

$$Gl = 4 * 2$$

$$Gl = 8$$

El valor en la tabla correspondiente es:

Cuadro 84

Valor en tablas ejemplo 4		
	Nivel de confianza	
Grados de libertad	0.05	0.01
8	15.507	20.09

FUENTE: Elaboración del autor.

Las variables no están relacionadas por lo que su relación no es potencialmente causal. Las variables no podrían explicarse la una con la otra.

Considérense, para otro ejemplo, los datos que provienen de la obra: Buendía Laredo, Jorge y Zuckerman Behar, Leo. "El blues de los abstencionistas", en *Nexos*, núm. 200, agosto de 1994, pp. 74-78.

Para los resultados oficiales no se especifica la fuente. Para los datos electorales la fuente es una encuesta realizada por Opinión profesional, S.A. de C.V., aplicada del 29 de agosto al 4 de septiembre de 1991. El diseño fue probabilístico, con 5 000 entrevistas en hogares y con representatividad nacional, el responsable es Ricardo Vernon, director general de la firma.

Las variables son "Decisión electoral", categorizada en abstencionistas y votantes; la otra variable es "Edad", categorizada en 18-25, 26-40, 41-60 y, 61 y más.

Se desea saber si las variables están estadísticamente relacionadas como para poder explicar a una con la otra.

Evalúe, entonces, la relación entre las variables, mediante la X^2.

La tabla de partida:

Cuadro 85

Tabla de contingencia de frecuencias observadas

		Decisión electoral		
		Abstencionista	Votante	
Edad	18-25	34	66	100
	26-40	28.6	71.4	100
	41-60	26.2	73.8	100
	61-...	31.6	68.4	100
		120.4	279.6	400

FUENTE: Elaboración del autor.

La tabla de frecuencias esperadas:

Cuadro 86

Tabla de contingencia de frecuencias esperadas

		Decisión electoral		
		Abstencionista	Votante	
Edad	18-25	30	70	100
	26-40	30	70	100
	41-60	30	70	100

	61-...	30	70	100
		120	280	400

FUENTE: Elaboración del autor.

El cálculo de X^2:

Cuadro 87

CELDA	O	E	O-E	(O-E)2	(O-E)2/E
18-25/A	34	30.1	3.90	15.21	0.51
26-40/A	28.6	30.1	-1.50	2.25	0.07
41-60/A	26.2	30.1	-3.90	15.21	0.51
61-.../A	31.6	30.1	1.50	2.25	0.07
18-25/V	66	69.9	-3.90	15.21	0.22
26-40/V	71.4	70	1.50	2.25	0.03
41-60/V	73.8	70	3.90	15.21	0.22
61-.../V	68.4	70	-1.50	2.25	0.03
					1.66

FUENTE: Elaboración del autor.

El cálculo de los grados de libertad:

$$Gl = (f-1)*(c-1)$$

$$Gl = (4-1)*(2-1)$$

$$Gl = 3*1$$

$$Gl = 3$$

El valor en la tabla correspondiente es:

Cuadro 88

Valor en tablas ejemplo 5		
Nivel de confianza		
Grados de libertad	0.05	0.01
3	7.815	11.345

FUENTE: Elaboración del autor.

Las variables no están relacionadas, la relación no es significativa, por lo que la relación no es potencialmente causal. Las variables no podrían explicarse la una con la otra.

De la obra del ejemplo anterior: Buendía Laredo, Jorge y Zuckerman Behar, Leo, *op. cit.*, pp. 74-78, es posible extraer otro ejemplo.

Las variables son “Decisión electoral”, categorizada en Abstencionistas y Votantes; la otra variable es “Educación”, categorizada en Ninguna, Primaria, Secundaria o preparatoria y, Universidad o más.

Se desea saber si las variables están estadísticamente relacionadas como para poder explicar a una con la otra.

Evalúe, entonces, la relación entre las variables, mediante la X^2.

La tabla inicial es:

Cuadro 89

Tabla de contingencia de frecuencias observadas

		Decisión electoral		
		Abstencionista	Votante	
Educación	Ninguna	35.4	64.6	100
	Primaria	32.6	67.4	100
	Sec-Prepa	27.5	72.5	100
	Universidad o +	27.1	72.9	100
		122.6	277.4	400

FUENTE: Elaboración del autor.

La tabla de frecuencias esperadas:

Cuadro 90

Tabla de contingencia de frecuencias esperadas

		Decisión electoral		
		Abstencionista	Votante	
Educación	Ninguna	30.65	69.35	100
	Primaria	30.65	69.35	100
	Sec-Prepa	30.65	69.35	100
	Universidad o +	30.65	69.35	100
		122.6	277.4	400

FUENTE: Elaboración del autor.

El cálculo del valor de X^2:

Cuadro 91

CELDA	O	E	O-E	(O-E)2	(O-E)2/E
N/A	35.4	30.65	4.75	22.56	0.74
P/A	32.6	30.65	1.95	3.80	0.12
SP/A	27.5	30.65	-3.15	9.92	0.32
U-.../A	27.1	30.65	-3.55	12.60	0.41
N/V	64.6	69.35	-4.75	22.56	0.33
P/V	67.4	69.35	-1.95	3.80	0.05
SP/V	72.5	69.35	3.15	9.92	0.14
U-.../V	72.9	69.35	3.55	12.60	0.18
					2.30

FUENTE: Elaboración del autor.

El cálculo de los grados de libertad:

$$Gl = (f-1)*(c-1)$$

$$Gl = (4-1)*(2-1)$$

$$Gl = 3*1$$

$$Gl = 3$$

El valor en la tabla correspondiente es:

Cuadro 92

Valor en tablas ejemplo 6		
Nivel de confianza		
Grados de libertad	0.05	0.01
3	7.815	11.345

FUENTE: Elaboración del autor.

Las variables no están relacionadas, su relación es no significativa, por lo que la relación no es potencialmente causal. Las variables no podrían explicarse la una con la otra.

Un séptimo ejemplo se extrae de la misma fuente: Buendía Laredo, Jorge y Zuckerman Behar, Leo, *op. cit.*, pp. 74-78.

Las variables son “Decisión electoral”, categorizada en abstencionistas y votantes; la otra variable es “Ingreso”, categorizada en 0-1, 1-3, 3-7 y, 7-… salarios mínimos.

Se desea saber si las variables están estadísticamente relacionadas como para poder explicar a una con la otra.

Evalúe, entonces, la relación entre las variables, mediante la X^2.

La primera tabla es:

Cuadro 93

Tabla de contingencia de frecuencias observadas

		Decisión electoral		
		Abstencionista	Votante	
Ingreso	0-1 s. m.	34.2	65.8	100
	1-3 s. m.	29.6	70.4	100
	3-7 s. m.	25.1	74.9	100
	7-… s. m.	29.6	70.4	100
		118.5	281.5	400

FUENTE: Elaboración del autor.

La de frecuencias esperadas:

Cuadro 94

Tabla de contingencia de frecuencias esperadas

		Decisión electoral		
		Abstencionista	Votante	
Ingreso	0-1 s. m.	29.625	70.375	100
	1-3 s. m.	29.625	70.375	100
	3-7 s. m.	29.625	70.375	100
	7-… s. m.	29.625	70.375	100
		118.5	281.5	400

FUENTE: Elaboración del autor.

El cálculo del valor de X^2:

Cuadro 95

CELDA	O	E	O-E	(O-E)2	(O-E)2/E
0-1/A	34.2	29.625	4.58	20.93	0.7065
1-3/A	29.6	29.625	-0.02	0.00062	0.0000211
3-7/A	25.1	29.625	-4.53	20.48	0.6912
7-…/A	29.6	29.625	-0.02	0.00	0.0000211
0-1/V	65.8	70.375	-4.58	20.93	0.2974
1-3/V	70.4	70.375	0.03	0.00	0.000009
3-7/V	74.9	70.375	4.53	20.48	0.2910
7-…/V	70.4	70.375	0.03	0.00	0.000009
					1.99

FUENTE: Elaboración del autor.

El cálculo de los grados de libertad:

$$Gl = (f-1)*(c-1)$$

$$Gl = (4-1)*(2-1)$$

$$Gl = 3*1$$

$$Gl = 3$$

El valor en la tabla correspondiente es:

Cuadro 96

Valor en tablas ejemplo 7		
Nivel de confianza		
Grados de libertad	0.05	0.01
3	7.815	11.345

FUENTE: Elaboración del autor.

Las variables no están relacionadas, su relación es no significativa, por lo que la relación no es potencialmente causal. Las variables no podrían explicarse la una con la otra.

De los últimos siete ejemplos, presentados para X cuadrada, las significancias son las siguientes:

Cuadro 97

Significancia de los ejemplos presentados		
Número del ejemplo	Significativo	No significativo
1	S	
2	S	NS
3	S	
4	NS	

5	NS	
6	NS	
7	NS	

FUENTE: Elaboración del autor.

Como puede recordarse, el ejemplo dos es significativo o no dependiendo del nivel de confianza escogido.

Queda así manifiesta la utilidad de las herramientas de la estadística no paramétrica. En muchas ocasiones se arman cuadros con información cruzada, tal como se han presentado en los ejemplos, posteriormente se describen suponiendo que las variables implicadas están relacionadas. Sin embargo, y tal como se ha mostrado, puede haber un alto porcentaje de casos de relaciones no significativas.

Se han abordado tres herramientas estadísticas no paramétricas, las dos primeras, coeficiente Q de Yule y coeficiente Fi, son útiles solamente para cuadros de 2 * 2; para ese mismo caso y para cuadros más amplios se debe utilizar la X cuadrada.

Entonces, no utilizar herramientas de asociación en casos considerados propicios para la estadística no paramétrica, puede llevar a incurrir en errores que se cometen mucho más por desconocimiento de esas herramientas que por hacerlo deliberadamente. Es necesario que decrezcan, aunque sea paulatinamente, esos casos.

La expuesta en este capítulo es estadística que mide asociación entre variables, pero en la mayoría de casos será, según el caso de los coeficientes Q de Yule y Fi, menor a 1, lo que indica que no es una asociación perfecta y, por lo tanto, también constituye una medida del error que se acepta al efectuar el análisis; en el caso de la X cuadrada la significación, como se ha expuesto, se elabora a los niveles de 0.05 y 0.01, por lo que la relación no se asegura, solamente se analiza aceptando un grado de significación, que no es del 100 por ciento, entonces también constituye la aceptación implícita de un error en el análisis. Desde luego es mucho mejor efectuar análisis calculando el error aceptado que suponer certeza procediendo a describir datos de variables que posiblemente no estén siquiera relacionadas estadísticamente.

Anexo A del capítulo 6

DISTRIBUCIÓN DE X CUADRADA

GRADOS DE LIBERTAD (gl)	0.05	0.01
1	3.841	6.635
2	5.991	9.210
3	7815	11345
4	9.488	13.277
5	11.07	15.086
6	12.592	16.812
7	14.067	18.475
8	15.507	20.09
9	16.919	21466
10	18.307	23.209
11	19.675	24.725
12	21.026	26.217
13	22.362	27.688
14	23.685	29.141
15	24.996	30.578
16	26.296	32.000
17	27.587	33.409
18	28.869	34.805
19	30.144	36.191
20	31.410	37.566

21	32.671	38.932
22	33.924	40.289
23	35.170	41.638
24	36,415	42.980
25	37.652	44.314
26	38.885	45.642
27	40.113	46.963
28	41.337	48.278
29	42.557	49.588
30	43.773	50.892
35	49.802	57.342
40	55.758	63.691
45	61.656	69.957
50	67.505	76.154
60	79.082	88.379
70	90.531	100.425
80	101.879	112.329
90	113.145	124.116
100	124342	135.807

Fuente: Wayne W. Daniel (1977). Estadística con Aplicaciones a las Ciencias Sociales y a la Educación, México: McGraw-Hill.
Fuente original: "A table of Perecentage Points of the x2 Distribution". Skandinavisk Aktuarietidskrift, 33 (1950), 168-175.

7. ELEMENTOS BÁSICOS DE LA TEORÍA DE DECISIONES

Iniciaremos por exponer y analizar tres casos: el dilema del prisionero, mafias y, el caso de individuos solidarios.

1. Dilema del prisionero

Por qué no hay cooperación:

Porque los intereses de los agentes son incompatibles
La solución solamente puede ser un óptimo de Pareto.

El dilema del prisionero:

En el conocido dilema del prisionero dos criminales son enjuiciados, cada uno sabe que, si ninguno confiesa entonces les darán 30 días de prisión a ambos. Si ambos confiesan les darán un año a cada uno. Entonces hay incentivos para no confesar, pero el juez ha prometido, separadamente y sin que el otro lo sepa, que si el otro confiesa saldrá libre y, el que no confiese tendrá cinco años de prisión.

Decisión individual:

Ninguno confiesa → 1 mes a cada uno
Ambos confiesan → 1 año a cada uno.

La estrategia dominante es No confesar, es decir cooperar.

Decisión colectiva:

IA. Si B confiesa sale libre; A No confiesa → 5 años a A
IB. Si A confiesa sale libre; B No confiesa 5 años a B.

Entonces, la estrategia no parece clara, hay un dilema, lo que subyace a él es la confianza o desconfianza.

Cuadro 98

		Dilema de prisionero	
		Individuo A	
		Cooperar	No cooperar
Cooperar	Individuo	(3,3)	(1,4)

No cooperar	B	(4,1)	(2,2)

Orden de preferencias: 4 > 3 > 2 > 1

FUENTE: Elaboración del autor.

A:	B Cooperar	nc	>	c	4>3
	B No cooperar	nc	>	c	2>1

(1, 4)
(3, 3)
(2, 2)
(4, 1)

B:	A Cooperar	nc	>	c	4>3
	A No cooperar	nc	>	c	2>1

(4, 1)
(3, 3)
(2, 2)
(1, 4)

El caso de las mafias

Las limitaciones del modelo de elección racional son la anomia y la ética.

En el modelo se crea un tipo ideal, en el sentido de que se imagina un mundo donde todos son egoístas y, además, actuarán racionalmente, maximizando utilidades o minimizando las pérdidas.

Además, se pasan por alto las desigualdades en el poder de los actores.

Otra dificultad es llevar a una generalización social el comportamiento egoísta individual.

Si, en el dilema del prisionero, se supone que ambos pertenecen a una mafia entonces no confesarán, pero ello no indica que sean irracionales. Lo que sucede es que la estructura de pérdidas y ganancias es diferente: el que confiese está asegurando su muerte al salir de la cárcel.

Cuadro 99

Mafias

		Individuo A	
		Cooperar	No cooperar
Cooperar	Individuo	RR (4,4)	ST (3,2)
No cooperar	B	TS (2,3)	PP (1,1)

FUENTE: Elaboración del autor.

A.	B	Coopera	Coopera	(4 > 2)
	B	No coopera	Coopera	(3 > 1)
B	A	Coopera	Coopera	(4 > 2)
	A	No coopera	Coopera	(3 > 1)

Es útil considerar en la reflexión de los modelos anteriores, y el siguiente, los conceptos:

Anomia, es la ruptura entre el sujeto y la sociedad global. Involucra un factor de pertenencia (o integración) y no pertenencia. Tiende a aprehender una noción de desintegración a un marco social de referencia, lo cual no necesariamente implica la aceptación de otras situaciones.

La anomia puede implicar una no-aceptación de cualquier clase de situaciones, debe distinguirse de la participación simbólica, la cual, en cambio, es una medida de alienación que implica el rechazo de la situación inmediata pero con otras alternativas. Esto es, hay la percepción de que, dentro de otras condiciones, se podría llegar a una integración más completa en un marco social de referencia determinado. Más aún, se sugiere que la alienación no es una característica de personalidad sino situacional.

Esos conceptos parecen estar relacionados con un individuo que coopera o no con sus pares, con sus iguales.

Individuos éticos o solidarios

Un individuo ético también puede alterar el orden de preferencias. Se supone un individuo altruista preferirá sacrificarse para que otro individuo salga mejor librado.

En el caso en el que decide cooperar con los iguales, el orden de preferencias es como sigue:

Cuadro 100

		Individuos éticos	
		Individuo A	
		Cooperar	No cooperar
Cooperar	Individuo B	(4,3)	(3,4)
No cooperar		(2,1)	(1,2)

FUENTE: Elaboración del autor.

A	B Coopera	No coopera	(4 > 3)
	B No coopera	No coopera	(2 > 1)
B	A Coopera	Coopera	(4 > 2)
	A No coopera	Coopera	(3 > 1)

Aparentemente no hay estrategia dominante en el juego, porque la estrategia de A es no cooperar, en tanto que el individuo altruista B decide cooperar. No obstante, las ganancias de las posibilidades no suman igual número, en dos de ellas la ganancia es de 1, pero cuando A coopera B tiene ganancias extraordinarias, ya que espera fervientemente confiar y cooperar con A. Además, es inflexible por definición, es decir, cooperará. Por lo tanto, la estrategia del juego es cooperar.

Racionalidad de corto plazo

Un ejemplo de reflexión en México abarca el estudio de las prácticas políticas en las elecciones.[53]

El electorado tiene cálculos racionales utilitaristas, si bien estos son de corto plazo. Wayne Cornelius documenta, para las elecciones de 2000 a la presidencia de la república, la compra y coacción del voto.

Los datos provienen de un panel del periódico Reforma con auspicio de la Fundación Nacional para la Ciencia de Estados Unidos (n = 1, 260). También de una Encuesta de Mitofsky a la salida de la casilla (n = 6, 145). Asimismo, contribuye una encuesta realizada por Investigaciones Sociales Aplicadas S. C. para la FLACSO.

Se presenta un análisis de correlación logística planteando como variables explicadas las distintas parejas formadas por los tres candidatos con más alta votación, y propuestas las variables explicatorias en: sexo, edad y, educación, como variables de control; ruralidad y, nivel socioeconómico, como variables económicas; si fue visitado en su domicilio por algún representante del PAN, PRI o PRD, si recibió algún regalo por parte de alguno de esos partidos y, si son beneficiarios de Progresa, Procampo o Liconsa, como variables políticas.

El análisis de los resultados indica que Labastida, candidato del PRI, es quien tiene la menor eficacia, solamente son significativas en su favor, las variables de los beneficiarios de Progresa y Liconsa.

Para Cárdenas son significativos el voto femenino, las visitas por representantes del PRD, y sobre todas las variables, el recibir un regalo por parte del partido.

Por el contrario, a Fox le son favorables y significativas las variables de ruralidad, ser visitado por un representante del partido pero, sobre todo, la geografía del caso indica que el voto favorable y coaccionado o comprado del PAN se distribuye de manera uniforme en el país, en tanto que el PRI y PRD se hacen notorios solamente en donde controlan el gobierno estatal, lo que parece estar relacionado con el hecho de que la estructura de "amigos de Fox" tenía independencia del partido postulante.

Óptimo de Pareto

Ejemplo adicional, también en México, pero que cito para mostrar el óptimo de Pareto. Es el de Molinar, Weldon y Sánchez. Trata el caso de la elección para diputados federales en 1994.[54]

En la arena de la legitimidad de la elección se considera que:

Sólo PRI puede o no cometer fraude.

PRD y el PAN podrían denunciar o no el fraude, independientemente de que existiera.

Posibilidades:

A. PRI hace fraude y el PRD lo acusa.
B. PRI hace fraude y el PRD no lo denuncia.
C. PRI no comete fraude, pero el PRD de cualquier manera denuncia la elección.
D. PRI no comete fraude y el PRD no denuncia la elección.

El esquema del juego es el siguiente:

[53] Wayne Cornelius. "La eficacia en la compra y coacción del voto en las elecciones mexicanas de 2000", en Perfiles Latinoamericanos, No. 20, junio de 2002. Pp. 11-31.

[54] Molinar, Juan, Sánchez, Arturo y Weldon, Jeffrey. "El año que votamos en más peligro", en Nexos, No. 200, agosto de 1994. Pp. 65-73.

Cuadro 101

Tabla para PRI y PRD

		PRD	
		Denunciar	No denunciar
PRI	Fraude	A	B
	No Fraude	C	D

FUENTE: Elaboración del autor.

Si se reemplazan las letras A, B, C y D por números, y donde 4 > 3 > 2 > 1; y donde el primer número de cada par corresponde a las filas, PRI, y el segundo corresponde a las columnas, PRD, entonces las preferencias se esquematizan así:

Cuadro 102

Tabla para PRI y PRD

		PRD	
		Denunciar	No denunciar
PRI	Fraude	(2, 2)	(4, 1)
	No Fraude	(1, 4)	(3, 3)

FUENTE: Elaboración del autor.

Si el PRD denuncia, el PRI obtendría 2 siendo fraudulento y sólo 1 si no lo hace.

Si el PRD no denuncia, el PRI obtendría 4 haciendo fraude, mientras que solamente obtendría 3 no haciéndolo.

La estrategia dominante del PRI es hacer fraude.

Si el PRI hace fraude, el PRD obtendría 2 denunciando y sólo 1 no denunciando.

Si el PRI no hace fraude, el PRD obtendría 4 denunciando y sólo 3 no denunciando.

La estrategia dominante del PRD es denunciar.

Se trata de un juego del dilema del prisionero. Dadas las estrategias dominantes en su conjunción se obtendría (2, 2), ambos saben que ganarían más con (3, 3), pero se tienen desconfianza.

Los autores suponen que el juego entre PAN y PRI es diferente:

Cuadro 103

Tabla para PRI y PAN

		PAN	
		Denunciar	No denunciar
PRI	Fraude	(2, 2)	(4, 1)
	No Fraude	(1, 3)	(3, 4)

FUENTE: Elaboración del autor.

Para el PRI todo es igual al caso anterior, su estrategia dominante es cometer fraude.

Si el PRI comete fraude, el PAN obtiene 2 denunciando y sólo 1 no haciéndolo. Denunciará.

Si el PRI no comete fraude, el PAN obtiene 3 denunciando y 4 no denunciando. No denunciará.

El PAN no tiene estrategia dominante, ya que si hay fraude denuncia, si no lo hay no denuncia y legitima la elección.

No obstante, el equilibrio del juego sigue siendo (2, 2) ya que el PRI sí tiene estrategia dominante: hacer fraude.

En este caso la única forma de que se produzca un desenlace sin fraude y legitimada, al menos en parte ya que el PRD denunciará, es que el PRI perciba que le conviene una elección legitimada y sin fraude, entonces las estrategias serían las que siguen:

Cuadro 104

Tabla para PRI y PAN

		PAN	
		Denunciar	No denunciar
PRI	Fraude	(2, 2)	(3, 1)
	No Fraude	(1, 3)	(4, 4)

FUENTE: Elaboración del autor.

En este caso hay dos equilibrios: (2, 2) que es ineficiente en el sentido de Pareto, ya que en él se requiere que la utilidad total sea mayor o que todos los jugadores obtengan más de lo que tenían en la situación anterior; y, (4, 4) que es eficiente en el sentido de Pareto.

Desde luego, la forma de que el PRI decidiera que era conveniente la eficiencia Paretiana está condicionada porque tuviera grandes probabilidades de ganar.

Costos externos

El caso expuesto por Molinar para México supone que la cuestión de la legitimidad funciona como la adopción de decisiones por la regla de la mayoría, es decir, supone que basta con que acepten la elección dos de los tres agentes.

En estos casos cabe la posibilidad de que algunos individuos sean perjudicados en sus preferencias, éstos serían los agentes que conforman la minoría.

El uso de una regla menos exigente que la de la unanimidad impone un coste a aquellas personas que resultan perjudicadas por la adopción de la decisión. Ese coste podría ser evitado dedicando el tiempo y el esfuerzo adicional que se necesite para poder redefinir el objeto de discusión de modo que su aprobación sea benéfico para todos. Ese coste es la diferencia entre la utilidad que se ha conseguido y la que se obtendría mediante la toma de decisiones por la regla de la unanimidad.

Tullock y Buchanan[55] llaman costes externos a los que resultan de la deliberación necesaria para poder llegar a dialogar, a establecer los criterios de decisión y tomar acuerdos respecto al objeto de discusión, esos costes pueden ser altos en función de cuestiones tales como:

El número de agentes decisores
La polarización de las opiniones.

Respecto a los costes externos la regla óptima sería la de unanimidad puesto que los minimiza. Sin embargo, el tiempo, que no forma parte de los costes externos, tiende al infinito en sus costes al aumentar el número de agentes del comité decisor. La cuestión entonces es determinar, cuántos agentes se necesitan para que una decisión se tome. Para ello, sugiere Mueller,[56] deben compararse los costes externos de aprobar una decisión en contra de algunos individuos y los costes correspondientes al tiempo empleado en la reelaboración de la decisión.

La intensidad de las preferencias

Del tiempo necesario para redefinir el objeto surge un problema adicional, el de la intensidad en las preferencias. Cada uno siente sobre el mismo objeto de decisión con diferente intensidad.

---

[55] Buchanan, James y Tullock, Gordon, *El cálculo del consenso*. Planeta-Agostini, Barcelona, España, 1993.

[56] Mueller, Dennis C., *Elección pública.* Alianza Editorial, edición en español, primera edición en inglés 1979, Madrid, España, 1984.

Tener en cuenta las intensidades puede ayudar a explicar por qué la regla de las mayorías ya no es completamente aceptada: hay grupos minoritarios que tienen una pequeña intensidad y ceden ante la regla de la mayoría, pero una minoría no lo hace.

La cuestión es menos grave en elecciones en las que se elige quiénes tomarán decisiones, el proceso continúa y no es un juego de suma cero. Pero cuando se trata de elementos de democracia directa como el referéndum, el voto no decide quién decidirá, sino elige el hecho de política, por lo que es un juego de suma cero.

La consecuencia de no considerar la intensidad de las preferencias puede ser recomendar una indiscriminada mayor decisión directa del demos y, por consiguiente, democracias refrendarias, no considerando que se debilita peligrosamente la probabilidad de aceptación de la regla de la mayoría.

Institucionalismo

John Orbell[57] explora el caso del dilema del prisionero y muestra que ante situaciones de este tipo el institucionalismo puede ser una alternativa para lograr algún tipo de solución a tal dilema. Se considera que hay dos conformaciones institucionales: primero, el individualismo sin coordinación y el dictador egoísta; segundo, democracia mayoritaria.

El individualismo sin coordinación conduce de manera inequívoca al dilema del prisionero, es cierto que pueden existir casos desviantes, pero la mayoría de los actores se comportan racionalmente, de forma egoísta y desconfiada, lo que conduce inevitablemente a resultados subóptimos.

El dictador egoísta implica algo diferente a la idea de un déspota benévolo, éste implica que el déspota descubre el desarrollo que produce el mejor resultado posible para la sociedad y entonces, usando su capacidad, la implementa. El dictador egoísta, en cambio, descubre el desarrollo que le producirá el máximo beneficio particular y utiliza la capacidad que tiene para implementarlo, lo cual lleva a soluciones subóptimas para la sociedad.

Si se reflexiona cuidadosamente en los dos casos anteriores, parece, según lo ha planteado la filosofía política, que debe buscarse la manera de atar las manos al dictador para que no pueda actuar solamente en su favor. Eso solamente puede hacerse mediante el diseño institucional, por lo que el autor analiza el caso de la democracia mayoritaria.

En el mundo real existen empresarios que tienen la capacidad de obligar o influir en otros para obligarles a cooperar en su favor, para sus preferencias, por lo que esa no puede ser considerada una sociedad libre. La diferencia entre este tipo de sociedad y la del individualismo egoísta parece ser solamente formal. Una de las implicaciones sería que en este tipo de gobierno sería deseable que fuera incapaz.

Democracia mayoritaria implicaría una mejora al considerar que se trata de un grupo de decisores y ya no de uno solo, lo que vuelve más competitivo al sistema. Otra mejora es hacer que los costos para cooperar sean muy bajos, lo que hará que más personas participen y se vuelva más competitivo el sistema. Aumentar la competitividad es lo mismo que elevar la representación de la democracia. El último y lógico paso para lograr hacer más competitivo al sistema es una mayor igualdad en el reparto del producto, por lo que la cuestión económica, si bien no forma parte de la democracia, sí debe ser considerada como algún tipo de causa de ella.

[57] Orbell, John M. and Wilson, L.A. II. "Institutional solutions to the n-prisoners dilemma", en *American Political Science Review*, Vol. 72, No. 2, june 1978, pp. 411-421.

Bibliografía

Aaker, David y Day, George S. *Investigación de mercados*. Mc Graw-Hill, segunda edición en español, México, 1989.

Aldrich, John. "Turnout and rational choice." Duke University Program in Political Economy, Working paper, No. 100, North Carolina, 1990.

Arya, Jagdish y Lardner, Robin. *Matemáticas aplicadas*. Prentice Hall, México, 1992.

Asher, Herbert B. *Causal modeling*. Sage publications, United States of America, 1988.

Barranco, Francisco J. *Técnicas de marketing político*. Red Editorial Iberoamericana, México, 1994.

Blalock, Hubert M. Causal models in the social sciences. University of Washington, United States of America, 1971.

Boudon, Raymond y Lazarsfeld, Paul. Metodología de las ciencias sociales. Editorial Laia, Barcelona, España, 1985.

Boudon, Raymond. Education, opportunity and social inequality: changing prospects in western society. John Wiley and Sons, Nueva York, 1974.

Brams, Steven J. Game theory and politics. Free Press, New York, 1975.

Buchanan, James y Tullock, Gordon. El cálculo del consenso. Planeta-Agostini, Barcelona, España, 1993.

Buendía Laredo, Jorge y Zuckerman Behar, Leo. "El blues de los abstencionistas", en *Nexos*, núm. 200, agosto de 1994, pp. 74-78.

Bunge, Mario. *La causalidad. El principio de causalidad en la ciencia moderna*. Editorial Sudamericana, primera edición en inglés E. U. 1959, Buenos Aires, Argentina, 1997.

Campbell, Donald y Stanley, Julian. *Diseños experimentales y cuasiexperimentales en la investigación social*. Amorrortu Editores, Buenos Aires, Argentina, 1995.

R. Carter E. A. *Undergraduate econometrics*. John Willey and Sons Inc. United States of America, 2000.

Coleman, James. *Foundations of social theory*. Harvard University Press, USA, 1990.

Cortés, Fernando y Rubalcava, Rosa María. *Métodos estadísticos aplicados a la investigación en ciencias sociales*. El Colegio de México, México, 1987.

Coutiño, Fabiola (coordinadora). *Perspectivas teóricas y metodológicas de la cultura política en México*. BUAP, México, 2011.

Chao, Lincoln. *Introducción a la estadística*. Compañía Editorial Continental S. A., México, 1985.

Díez Medrano, Juan. *Métodos de análisis causal*. Centro de Investigaciones Sociológicas, Madrid, España, 1992.

Fiorina, Morris. *Retrospective voting in american nation elections*. Yale University Press, New Haven, 1981.

García Ferrando, Manuel. Socioestadística. Introducción a la estadística en sociología. Alianza Universidad Textos, Madrid, 1997.

Gates, Scott. *Games, information and politics: applying game theoretic models to political science*. University of Michigan, Michigan, 1997.

Gómez Aguilar, Roberto. *Introducción al muestreo*. Tesis de maestría en Ciencias en Estadística, Colegio de Postgraduados, Chapingo, 1979.

Hernández Sampieri, Roberto E. A. *Metodología de la investigación*. McGraw-Hill, México, 2000.

Holguín Quiñones, Fernando. *Estadística descriptiva aplicada a las ciencias sociales*. UNAM, México, 1981.

http://www.aulafacil.com/CursoEstadistica/Lecc-38-est.htm (Consulta: 12 de septiembre de 2012).

http://www.qro.cinvestav.mx (Consulta: 28 de diciembre de 2013).

Jaramillo, Ana. "Movimiento obrero y acumulación de capital: el caso argentino", en *Revista Mexicana de Ciencias Políticas y Sociales*, núm. 89, julio-septiembre, 1977.

Krassner, Stephen. *Las teorías sobre las instituciones y la política internacional. Soberanía hipocresía organizada*. Paidós, España, 2001.

Lipset, Seymour Martin. *El hombre político. Las bases sociales de la política*. Red Editorial Iberoamericana, México, (1959) 1997.

Lipset, Seymour Martin. "Some social requisites of democracy: economic development and political legitimacy", en *American Political Science Review*, Vol. LIII, No. 1, march 1959, pp. 69-105.

Owen, Diana y Madrazo Julio. "1994 ¿Quién votó por cuál partido y por qué?", en *Nexos*, núm. 202, octubre de 1994, pp. 20-23.

Martínez, Sergio F. *De los efectos a las causas. Sobre la historia de los patrones de explicación científica*. Paidós, UNAM, México, 1997.

Mendenhall, William y Reinmuth, James E. *Estadística para administración y economía.* Grupo Editorial Iberoamérica, México, 1990.

Molinar Horcasitas, Juan. "El año que votamos en peligro", en *Nexos*, núm. 127, junio de 1988, pp. 49-55.

Molinar, Juan, Sánchez, Arturo y Weldon, Jeffrey. "El año que votamos en más peligro", en *Nexos*, núm. 200, agosto de 1994, pp. 65-73.

Molinar Horcasitas, Juan. *El tiempo de la legitimidad. Elecciones, autoritarismo y democracia en México*. Cal y Arena, segunda edición, primera edición 1991, México, 1993.

Molinar, Juan y Weldon, Jeffrey. "Elecciones de 1988 en México: crisis del autoritarismo", en *Revista Mexicana de Sociología,* UNAM, año LII, núm. 4, octubre-diciembre de 1990, pp. 229-262.

Mueller, Dennis C. *Elección pública*. Alianza Editorial, Madrid, España, (primera edición en inglés 1979) 1984.

Orbell, John M. and Wilson, L.A. II. "Institutional solutions to the n-prisoners dilemma", en *American Political Science Review*, Vol. 72, No. 2, june 1978, pp. 411-421.

Ordeshook, Peter C. *Game theory and political theory: an introduction*. Cambridge University Press, New York, 1986.

Ordeshook, Peter y Riker, William. "A theory of calculus of voting", en *American Political Science Review*, No. 62, marzo de 1968, pp. 25-42.

Pardinas, Felipe. *Metodología y técnicas de investigación en ciencias sociales*. Siglo XXI editores, México, 1985.

Pérez Enríquez, María Isabel. *El impacto de las migraciones y expulsiones indígenas de Chiapas*. UNACH, Tuxtla Gutiérrez, 1998.

Pindick, Robert y Rubinfeld, Daniel. *Econometría. Modelos y pronósticos*. Mc Graw-Hill, cuarta edición, México, 2000.

Popper, Karl. *La lógica de la investigación científica*. Tecnos, Madrid, 1973.

Popper, Karl. Un mundo de propensiones. Tecnos, Madrid, España, 1992.

Ramírez Valverde, Benito. "Proyecto: Estudio sobre estrategias para el desarrollo sustentable de la sierra Nor-oriente de Puebla con la participación de productores, organizaciones e instituciones como respuesta a las condiciones de pobreza y marginación." Primer informe semestral. Fondo mixto de investigación científica y tecnológica Conacyt-Gobierno del estado de Puebla. Puebla, México, 2003.

Reyna, José Luis. "Anomia y participación simbólica en un área rural", en *Revista Mexicana de Ciencias Políticas y Sociales,* UNAM, año XV, núm. 58, octubre-diciembre de 1969, pp. 499-518.

Sánchez Espinoza, Francisco. "Un ejercicio de correlación Pearson a partir de "La democracia en México" de González Casanova", en *Diké,* Revista del Centro de Investigaciones Jurídico Políticas de la BUAP, febrero 2007.

Sánchez Espinoza, Francisco. "La prueba de hipótesis", en *Diké,* Revista del Centro de Investigaciones Jurídico Políticas de la BUAP, Puebla, marzo de 2013, pp. 145-166.

Wayne Cornelius. "La eficacia en la compra y coacción del voto en las elecciones mexicanas de 2000", en *Perfiles Latinoamericanos*, núm. 20, junio de 2002, pp. 11-31.

Wayne, Cornelius. "El mexicano feo: México y Estados Unidos en la década de los ochentas", en *Nexos*, núm. 89, mayo de 1985, pp. 20-26.

Wayne, Daniel. *Estadística con aplicaciones a las ciencias sociales y a la educación*. Mc Graw-Hill, México, 1988.

Weber, Max (editor). *The methodology of the social sciences*. Shils and Fluch Free Press, New York, 1949.

Ximitl Islas, Iván. "Organización campesina y comercialización de café: estudio en dos municipios de la sierra norte de Puebla." Tesis de maestría en Estrategias para el Desarrollo Agrícola Regional, Colegio de Postgraduados, Puebla, 2006.

Estadística. Con aplicaciones a las Ciencias Políticas
de Francisco Sánchez Espinoza
se terminó de imprimir en julio de 2015
en los talleres de El Errante editor S.A. de C.V.,
ubicado en Privada Emiliano Zapata 5947
Col. San Baltazar, Puebla, Pue.

El tiro consta de 1000 ejemplares.

Estadística. Con aplicaciones a las Ciencias Políticas
de Francisco Sánchez Espinoza
se terminó de imprimir en agosto de 2015
en los talleres de El Errante editor S.A. de C.V.,
ubicado en Privada Emiliano Zapata 5947
Col. San Baltazar, Puebla, Pue.

El tiro consta de 1000 ejemplares.

ISBN: 978-607-96963-3-7

9 786079 696337

Francisco Sánchez Espinoza. Licenciado en Economía, maestro en Ciencias Políticas y doctor en Sociología. Profesor e investigador de la BUAP. Ha impartido cursos en la Licenciatura en Ciencias Políticas, Maestría en Ciencias Políticas (PNPC), Maestría en Sociología (PNPC) y, en el Doctorado en Economía Política del Desarrollo (PNPC). Miembro del Sistema Nacional de investigadores Nivel I e investigador visitante de El Colegio de México. Ha dictado conferencias, de carácter nacional e internacional, en México, Chile, El Salvador y, España. Ha publicado diversos artículos en revistas especializadas y en obras colectivas de Ciencia Política y Sociología. Libros individuales: *La medición del fenómeno político*, publicado por fomento editorial UAP, 2010; *Manual de SPSS con aplicaciones a las Ciencias Sociales*, Plaza y Valdés 2012; *El error en la ciencia*, Gernika, 2013.

PISO 15 Editores