Introducción a la Metodología de la Investigación

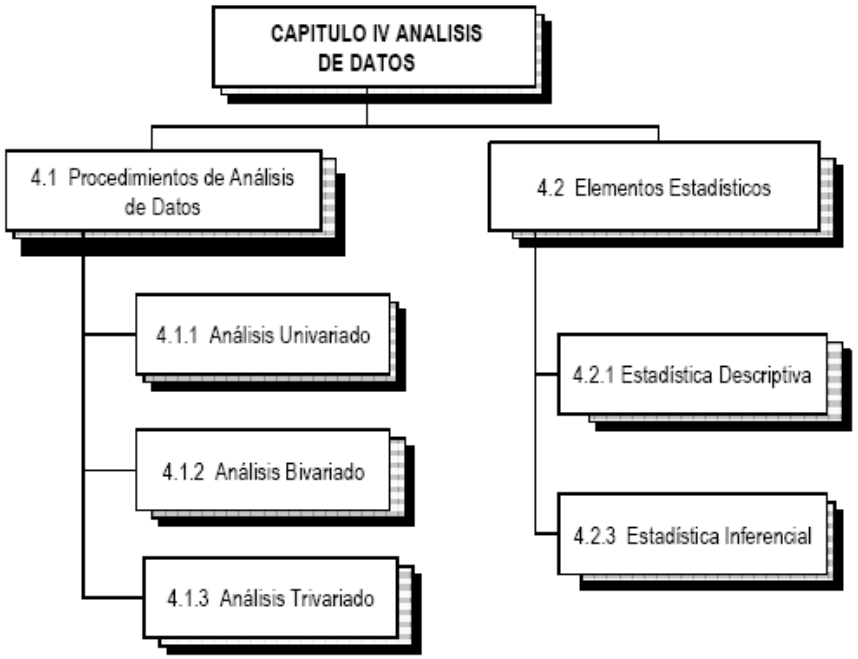
CAPÍTULO IV ANÁLISIS DE DATOS

Preferencia

	UACH	ITCC	Institución Particular	Total
Recien egresados	20	24	4	48
Próximos a egresar	16	18	8	38
Total	36	42	12	90

PLAN DEL CAPÍTULO IV

Objetivo: En el presente capítulo se expone un panorama conceptual sobre el análisis de datos. Se describen de manera no exhaustiva algunos elementos estadísticos útiles tanto para la organización y presentación de los datos como para el análisis de los resultados de investigación.



4.1 Procedimientos de Análisis de Datos

Una vez concluidas las etapas de colección y procesamiento de datos se inicia con una de las más importantes fases de una investigación: el análisis de datos. En esta etapa se determina como analizar los datos y que herramientas de análisis estadístico son adecuadas para éste propósito. El tipo de análisis de los datos depende al menos de los siguientes factores.

a) El nivel de medición de las variables (los niveles de medición fueron explicados en la sección 2.4 del capítulo II).

b) El tipo de hipótesis formulada (ver sección 2.2, capítulo II).

c) El diseño de investigación utilizado indica el tipo de análisis requerido para la comprobación de hipótesis.

El análisis de datos es el precedente para la actividad de interpretación. La interpretación se realiza en términos de los resultados de la investigación. Esta actividad consiste en establecer inferencias sobre las relaciones entre las variables estudiadas para extraer conclusiones y recomendaciones (Kerlinger, 1982). La interpretación se realiza en dos etapas:

a) Interpretación de las relaciones entre las variables y los datos que las sustentan con fundamento en algún nivel de significancia estadística.

b) Establecer un significado más amplio de la investigación, es decir, determinar el grado de generalización de los resultados de la investigación.

Las dos anteriores etapas se sustentan en el grado de validez y confiabilidad de la investigación. Ello implica la capacidad de generalización de los resultados obtenidos.

"Analizar significa establecer categorías, ordenar, manipular y resumir los datos," (Kerlinger, 1982, p. 96). En esta etapa del proceso de investigación se procede a racionalizar los datos colectados a fin de explicar e interpretar las posibles relaciones que expresan las variables estudiadas.

El diseño de tablas estadísticas permite aplicar técnicas de análisis complejas facilitando este proceso. El análisis debe expresarse de manera clara y simple utilizando lógica tanto inductiva como deductiva.

Los resultados de una investigación basados en datos muestrales requieren de una aproximación al verdadero valor de la población (Zorrilla, 1994). Para lograr lo anterior se requiere de una serie de técnicas estadísticas. Estas técnicas se derivan tanto de la estadística paramétrica como de la estadística no paramétrica. La primera tiene como supuestos que la población estudiada posee una distribución normal y que los datos obtenidos se midieron en una escala de intervalo y de razón. La segunda no establece supuestos acerca de la distribución de la población sin embargo requiere que las variables estudiadas se midan a nivel nominal u ordinal (ver Weiers, 1993).

Las tablas diseñadas para el análisis de datos se incluyen en el reporte final y pueden ser útiles para analizar una o más variables. En virtud de éste último criterio el análisis de datos puede ser univariado, bivariado o trivariado dependiendo de la cantidad de variables que se analizan.

4.1.1 Análisis Univariado.

Consiste en el análisis de cada una de las variables estudiadas por separado, es decir, el análisis esta basado en una sola variable. Las técnicas más frecuentes de análisis univariado son la distribución de frecuencias para una tabla univariada y el análisis de las medidas de tendencia central de la variable. Se utiliza únicamente en aquellas variables que se midieron a nivel de intervalo o de razón (ver Therese L. Baker, 1997). La distribución de frecuencias de la variable requiere de ver como están distribuidas las categorías de la variable, pudiendo presentarse en función del número de casos o en términos porcentuales.

4.1.2 Análisis Bivariado.

El análisis bivariado diseña tablas con tabulaciones cruzadas, es decir, las categorías de una variable se cruzan con las categorías de una segunda variable. Se les conoce como tablas de contingencia. Los requisitos que debe cubrir son:

1 El título debe reflejar la información que contiene la tabla.

2 Incluir un subtítulo para cada columna y subcolumna que se integre a la tabla.

3 Indicar el 100 % cuando la tabla se exprese en términos porcentuales.

4 Indicar al final de cada columna el número total de casos o categorías que comprende.

4.1.3 Análisis Trivariado

El análisis trivariado incluye una tercer variable que se utiliza como variable control. Esto permite analizar la asociación entre las dos variables, controlando el efecto de una tercer variable mediante la observación de las dos primeras sobre cada condición que presenta la tercera.

Por ejemplo si se analiza el ingreso económico de los ejecutivos de la micro, pequeña y mediana empresa regional con estudios de licenciatura y los ingresos de aquellos ejecutivos con estudios de posgrado (maestría), es posible incluir en el análisis la variable dicotómica sexo.

4.2 Elementos Estadísticos

El análisis e interpretación de datos requiere de un profundo conocimiento de la estadística, es decir, para que una investigación pueda arrojar luz sobre el PON, el investigador tendrá que someter los datos a la prueba estadística y para ello necesita tener conocimiento de los supuestos que involucra la metodología estadística que habrá de utilizar.

La herramienta utilizada para el análisis de datos es la estadística. Esta disciplina proporciona innumerables beneficios a la investigación científica y tecnológica. La estadística descriptiva se entiende como el conjunto de métodos para procesar

información en términos cuantitativos de tal forma que se les de un significado. La estadística inferencial estudia la confiabilidad de las inferencias de que los fenómenos observados en la muestra son extensivos a la población de donde se obtuvo la muestra, es decir, facilita el establecimiento de inferencias de la muestra analizada hacia la población de origen.

4.2.1 Elementos de Estadística Descriptiva

Como ya fue explicado la estadística descriptiva permite organizar y presentar un conjunto de datos de manera que describan en forma precisa las variables analizadas haciendo rápida su lectura e interpretación.

Entre los sistemas para ordenar los datos se encuentran principalmente dos: a) la distribución de frecuencias y b) la representación gráfica. Estos sistemas de organización y descripción de los datos permiten realizar un análisis de datos univariado, bivariado o trivariado, dependiendo de los objetivos y de la naturaleza de la investigación que se realiza.

Distribución de Frecuencias. Comunmente llamada tabla de frecuencias, se utiliza para hacer la presentación de datos provenientes de las observaciones realizadas en el estudio, estableciendo un orden mediante la división en clases y registro de la cantidad de observaciones correspondientes a cada clase. Lo anterior facilita la realización de un mejor análisis e interpretación de las características que describen y que no son evidentes en el conjunto de datos brutos o sin procesar. Una distribución de frecuencias constituye una tabla en el ámbito de investigación.

La distribución de frecuencias puede ser simple o agrupada. La distribución de frecuencias simple es una tabla que se construye con base en los siguientes datos: clase o variable (valores numéricos) en orden descendente o ascendente, tabulaciones o marcas de recuento y frecuencia. Por ejemplo, si se construye una distribución de frecuencias sobre los resultados finales que arrojó la evaluación de un curso de planeación estratégica para estudiantes de administración correspondientes al semestre

agosto-diciembre de 1998, se tienen los siguientes datos brutos: 86, 80, 84, 84, 74, 88, 87, 84, 74, 77, 77, 82, 68, 78, 67, 74, 66, 86, 65, 88,69 se procede a organizarlos en forma ascendente o descendente y se tiene en orden descendente: 88, 88, 87, 86, 86, 84, 84, 84, 82, 80, 78, 77, 77, 74, 74, 74, 69, 698, 67, 66, 65 posteriormente se registran en una tabla de distribución de frecuencias simple (ver Tabla 4.1). Cuando se pretende "... determinar el número de observaciones que son mayores o menores que determinada cantidad," (Webster, 1998, p. 27) se utiliza la distribución de frecuencias agrupadas también conocida como distribución de frecuencias acumuladas. La distribución de frecuencias agrupadas es una tabla que contiene las columnas siguientes: intervalo de clase, puntos medios, tabulación frecuencias y frecuencias agrupadas. Los pasos para diseñarla son:

Tabla 4.1 Distribución de Frecuencias de los Resultados Finales obtenidos de la Evaluación de Planeación Estratégica correspondientes al semestre agosto-diciembre de 1998.

Calificaciones	Tabulación	Frecuencia
88	//	2
87	/	1
86	//	2
85		0
84	///	3
83		0
82	/	1
81		0
80	/	1
79		0
78	/	1
77	//	2
76		0
75		0
74	///	3
73		0
72		0
71		0
70		0
69	/	1
68	/	1
67	/	1
66	/	1
65	/	1
Total		21

1 Se localizan el computo mas alto y el mas bajo de la serie de datos.

2 Se encuentra la diferencia entre esos dos cómputos.

3 La diferencia obtenida se divide entre números nones tratando de encontrar un cociente cercano a 15 pero no mayor. Lo anterior indica cuantas clases va a tener la distribución de frecuencias agrupadas y cuál va a ser la magnitud del intervalo de clase.

4 Se determina el primer intervalo de clase y posteriormente se van disminuyendo los límites del intervalo de clase de acuerdo al valor de la magnitud establecida previamente.

El ejemplo planteado en la distribución de frecuencias simples se utilizará tanto para efectos de ejemplificación de la distribución de frecuencias agrupadas como para el diseño de gráficas tipo polígono de frecuencias, histograma y ojiva. En la Figura 4.2 se presenta un ejemplo de una distribución de frecuencias agrupada.

Tabla 4.2 Distribución de Frecuencias Acumuladas de los Resultados Finales obtenidos de la Evaluación de Planeación Estratégica correspondientes al semestre agosto-diciembre de 1998.

Intervalo de Clase	Punto Medio	Tabulación	Frecuencias	Frecuencias Agrupadas
86-88	87	////	5	5
83-85	84	///	3	8
80-82	81	//	2	10
77-79	78	///	3	13
74-76	75	///	3	16
71-73	72		0	16
68-70	69	//	2	18
65-67	66	///	3	21
Total				21

Los computos mayor y menor son las puntuaciones 88 y 65, la diferencia es 88-65=23 y el número de intervalos de clase es 23/3= 7.68.

b) Representación Gráfica. A partir de la distribución de frecuencias se procede a presentar los datos por medio de gráficas. La información puede describirse por medio de gráficos a fin de facilitar la lectura e interpretación de las variables medidas. Los actuales sistemas computacionales como Excel, Lotus Smart Suite, Minitab, SAS-PC, Stath Graph, entre otros permiten obtener representaciones gráficas de diversos conjuntos de datos. Las gráficas pueden ser tipo histograma, polígono de frecuencias, gráfica de series de tiempo, etc,

b1) El Histograma. El histograma "... es una gráfica de barras que permite describir el comportamiento de un conjunto de datos en cuanto a su tendencia central, forma y dispersión," (Gutiérrez, 1998, p.79). De acuerdo con Glass y Stanley (1994) un histograma no debe ser demasiado plano o esculpado. El ancho es de dos tercios de su altura. Los pasos para elaborar un histograma son (ver Figura 4.1):

1 Se trazan los ejes horizontal y vertical.

2 Se registran marcas equidistantes sobre ambos ejes.

3 Se marcan los puntos medios de cada intervalo de clase sobre el eje horizontal.

b2) El Polígono de Frecuencias. Un método ampliamente utilizado para mostrar información numérica de forma gráfica es el polígono de frecuencia o gráfica de línea. La construcción es similar a la del histograma pero la diferencia radica en que para indicar la frecuencia solo se utiliza un punto sobre el punto medio de cada intervalo. Los pasos para construirlo son (ver Figura 4.2):

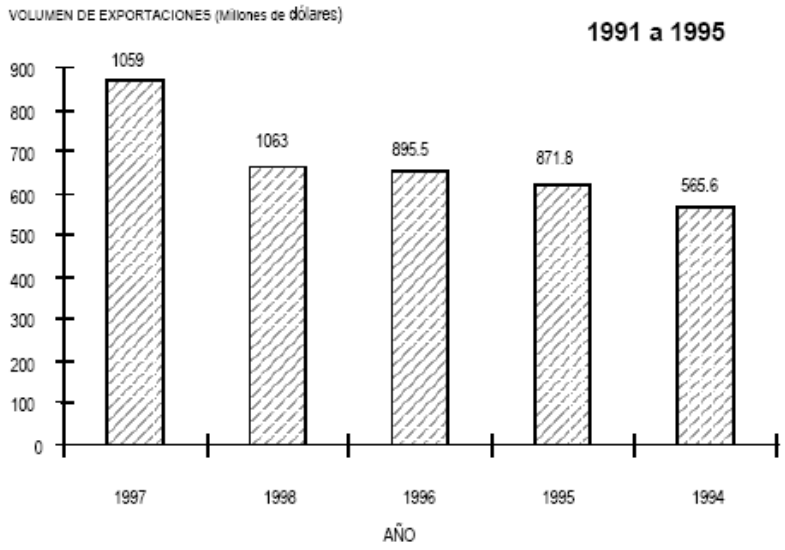


Figura 4.1 Número de Empresas de la Industria Gráfica de Estados Unidos por Segmento. (EPA, 1997).

Lugar	*Cantidad*	*Porcentaje*
México	*133,584*	*100.00 %*
Estado de Chihuahua	3,728	2.79 %
Municipio de Cuauhtémoc	48	0.04 %

Figura 4.2 Resultados de la Aplicación de una Prueba Matemáticas con 100 ítems al Grupo de 2º. de Ingeniería en Sistemas.

1 Se trazan los ejes horizontal y vertical.

2 Se registran marcas equidistantes sobre el eje horizontal y se anotan debajo de cada una de ellas los puntos medios de los intervalos de clase en un orden de menor a mayor.

3 Se registran marcas equidistantes sobre el eje vertical y se anotan a la izquierda de cada una de ellas las frecuencias en orden ascendentes. A partir de ellas se diseña la cuadrícula del espacio enmarcado, trazando las abscisas y ordenadas.

4 Se representa con puntos las frecuencias de cada intervalo de clase. Se toma en cuenta el punto medio de cada intervalo de clase como base y las frecuencias como altura.

5 Se unen con línea gruesa los puntos así determinados. 6 Se registra el título expresando en resumen el asunto o cuestión sobre la que informa la gráfica.

b3) Gráfica de Series de Tiempo. Es una gráfica de línea en la que la línea horizontal representa el tiempo. Es utilizada para representar tendencias como puede ser el tipo de cambio peso-dólar, el índice de precios al consumidor, etc. (ver Figura 4.3).

los anteriores elementos de estadística descriptiva son utilizados en investigación para diseñar tablas y figuras que presenten de manera resumida y organizada n conjunto de datos obtenidos mediante la observación y medición de las variables estudiadas.

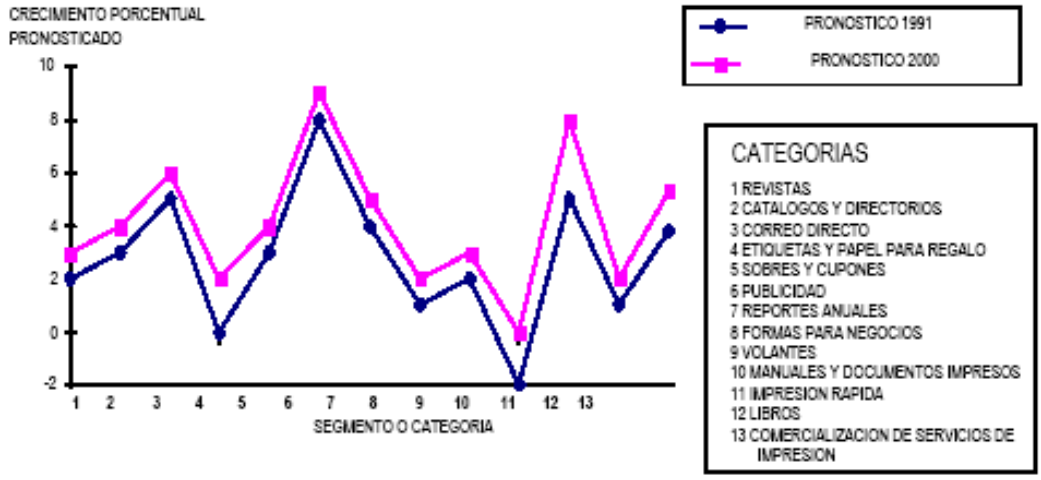


Figura 4.3 Tendencias de Crecimiento de la Industria Gráfica de Estados Unidos para el periodo 1990- 2000, (EPA, 1997).

c) Medidas de Tendencia Central. Las medidas de tendencia central son útiles para encontrar indicadores representativos de un colectivo de datos. Los tres métodos que permiten obtener el punto medio de una serie de datos son la media, la mediana y la moda.

c1) Media Aritmética. Medida de tendencia central que se define como el promedio o media de un conjunto de observaciones o puntuaciones. En aquellas situaciones en que la población de estudio es pequeña suele utilizarse la media poblacional mediante la expresión:

$$\mu = \frac{\sum_{i=1}^{N} Xi}{N}$$

(Ec. 4.1)

donde:

μ = media poblacional

ΣXi = Sumatoria de las puntuaciones

N = Número de casos

En cambio si la población de estudio es muy numerosa se procede a obtener la media muestral definida matemáticamente por la expresión:

$$X = \frac{\sum_{i=1}^{N} \overline{Xi}}{N}$$

(Ec. 4.2)

donde:

X = media muestral

ΣXi = Sumatoria de las puntuaciones

N = Número de casos

Al obtener la media alcanzada por la compañía XYZ que comercializa computadoras personales. Las ventas diarias realizadas por la compañía durante una semana indican las siguientes cantidades: 4, 12, 7, 9, 11, 7, 8, el cálculo de la media es:

$$\overline{x} = \frac{58}{7} = 8.29$$

el anterior resultado sugiere que el promedio semanal de ventas de la compañía XYZ es de 8.29 computadoras personales.

c2) La Moda. En una serie de puntuaciones se denomina moda a la observación que se presenta con mayor frecuencia. Así en el ejemplo anterior de la

compañía XYZ la moda es la puntuación 7. Para obtener la moda a partir de una distribución de frecuencias agrupadas se utiliza la expresión:

$$Mo = Lmo + \left[\frac{Da}{Db + Da}\right] i$$

(Ec. 4.3)

donde:

Mo = Moda

Lmo = Límite inferior del intervalo de clase modal

Da = Diferencia entre la frecuencia de la clase modal y la de la clase que la precede.

Db = Diferencia entre la frecuencia de la clase modal y la de la clase que l la sigue.

i = Intervalo de clase.

La moda para una distribución de frecuencias agrupadas se obtiene a partir de los datos de la Tabla 4.2:

$$Mo = 86 + \left[\frac{3}{(87 - 0) + (87 - 84)}\right]3 = 86.10$$

la moda tiene un valor de 86.10.

c3) La Mediana. También conocida como media posicional en virtud de que se localiza en el centro de un conjunto de observaciones presentadas en una serie ordenada de datos. Lo anterior sugiere que el 50 % de los casos se encuentra por encima de la mediana y el resto por debajo de ella. La posición central de la mediana se obtiene mediante la expresión matemática.

$$PMd = \frac{N + 1}{2}$$

(Ec. 4.4)

donde:

PMd = Posición de la Mediana

N = Número de casos.

el procedimiento para obtener la mediana a partir de una distribución de frecuencias simple o agrupada requiere de aplicar la expresión:

$$Md = \left[\frac{N/2 - FA}{\text{FS}}\right] \quad i$$

(Ec. 4.5)

donde:

Md = Mediana

N = Número de casos.

FA = Frecuencia agrupada.

FS = Frecuencia del intervalo adyacente superior.

Al aplicar la ecuación 4.5 a los datos de la Tabla 4.2 se obtiene un valor de 83 para la mediana:

$$Md = 82.5 + \left[\frac{10.5 - 10}{3}\right]3 = 83$$

De las tres medidas de tendencia central la media es mas exacta que la mediana por ser una estadística obtenida a través de una medición ordinal o de razón mientras que la mediana se obtiene a un nivel de medición nominal.

La principal característica de la media consiste en tomar en cuenta al 100 % de las puntuaciones de una distribución de frecuencias. No obstante cuando se analizan medidas extremas esta medida pudiera ser afectada por desviaciones que se posicionan por debajo o por arriba de ella. Ni la mediana ni la moda tienen este problema (Webster, 1998; Hopkins, Hopkins y Glass 1997; Kazmier, 1998).

a) Medidas de Dispersión.

Las medidas de dispersión son índices que se utilizan para describir una distribución de frecuencias a partir de la variación de los valores obtenidos. Los índices más utilizados son el rango, la varianza y la desviación estándar.

d1) El Rango. Indice conocido como recorrido. Se le define como la diferencia existente entre la puntuación mayor y la menor en una serie de datos. Tiene como desventaja que solo toma en cuenta para su cálculo las puntuaciones extremas, es decir la mayor y la menor omitiendo el resto de los datos u observaciones. Debido a lo anterior no es una medida confiable dado que se obtiene prácticamente por inspección.

d2) La Varianza. La varianza es una medida de variabilidad que toma en cuenta el 100 % de las puntuaciones de manera individual. Webster (1998) la define como “la media

aritmética de las desviaciones respecto a la media aritmética elevada al cuadrado," (p. 83). La definición matemática de la varianza se expresa por medio de la ecuación 4.6:

$$\sigma^2 = \frac{\Sigma X^2}{N}$$

(Ec. 4.6)

donde:

σ^2 = Varianza.

Σ = Suma de

X^2= Desviación de las puntuaciones de la media (X-X)

N = Número de casos.

d3) La Desviación Estándar. Dada la dificultad inherente de interpretar el significado de una varianza en virtud de que expresa valores elevados al cuadrado, para efectos de investigación es más adecuado utilizar la desviación estándar o desviación típica, definida como la raíz cuadrada de la varianza. La desviación estándar se expresa mediante la ecuación 4.7:

$$\sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}} = \sqrt{\sigma^2}$$

(Ec. 4.7)

donde:

Σ X^2= Suma de los cuadrados de cada puntuación

$(\Sigma X)^2$ = Suma de las puntuaciones elevadas al cuadrado

N = Número de casos.

σ = Desviación Estándar

La desviación estándar es una medida obtenida mediante una escala de intervalo o de razón basada en la magnitud de las puntuaciones individuales de la distribución (D"Ary, Jacobs y Razavieh, 1982). Es de mucha utilidad en "... en conjunción con la...distribución normal," (Kazmier, 1998).

e) Correlación.

La correlación es un método estadístico que permite determinar la presencia o ausencia de asociación entre dos variables sometidas a investigación. Por ejemplo se puede medir el grado de asociación entre el rendimiento académico y el nivel socioeconómico de una muestra de unidades de observación. La correlación se describe por medio de índices estadísticos denominados coeficientes de correlación que pueden sugerir si el cambio de una variable se asocia con el cambio de la otra variable.

Los índices mas utilizados para medir la asociación entre dos variables es el coeficiente de correlación producto-momento que se aplica a escalas de medición de intervalo o de razón y el coeficiente de correlación de rangos que se utiliza en escalas de medición ordinal.

Al analizar la correlación de una serie de datos el resultado que arroja un coeficiente de correlación fluctúa entre – 1.00 y + 1.00. Una puntuación de – 1.00 sugiere una correlación negativa perfecta. Una puntuación de 0.00 sugiere ausencia de asociación entre las variables y una puntuación de + 1.00 sugiere una correlación positiva perfecta. Una correlación positiva perfecta indica que si una variable aumenta la otra también aumenta, por ejemplo cabe esperar que si el tipo de cambio peso-dólar aumenta el volumen de exportaciones del sector manufacturero del país también aumenta.

En el caso de una correlación negativa perfecta ocurre el aumento de una variable y el decremento o disminución de la otra variable. Por ejemplo ante el aumento del tipo de

cambio peso-dólar cabe esperar una disminución o decremento en el volumen de importaciones del país. Una adecuada técnica para leer e interpretar los valores de correlación son las gráficas de dispersión. La Tabla 4.3 muestra algunos valores de coeficientes de correlación con su respectiva descripción y gráfica de dispersión.

Determinar la existencia de asociación entre las variables no indica existencia de causalidad. Esto es, un coeficiente de correlación únicamente sugiere el grado de relación entre las variables y no una situación causal.

e1) Correlación Producto-Momento. La correlación producto-momento es conocida como r de Pearson en virtud de que el estadístico Karl Pearson desarrollo este procedimiento. Se define como la media de los productos

$$r_{xy} = \frac{\sum ZXZY}{N}$$

de las puntuaciones Z y se expresa matemáticamente mediante la ecuación:

donde:

rxy = coeficiente de correlación producto-momento.

ΣZxZy = Sumatoria de los productos de puntuación Z.

n = Número de casos o puntuaciones pareadas.

en situaciones en las que el conjunto de observaciones es muy numeroso se omite la aplicación de la ecuación 4.8 y es sustituida por la expresión:

$$r_{xy} = \frac{\sum XiYi - \left(\sum Xi\right)\left(\sum Yi\right)/n}{\sqrt{\left[\sum X_i^2 - \left(\sum Xi\right)^2/n\right]\left[\sum Y_i^2 - \left(\sum Yi\right)^2/n\right]}}$$

(Ec. 4.9)

donde:

rxy = coeficiente de correlación producto-momento.

n = Número de casos. ΣXi = Sumatoria de las puntuaciones de la variable X.

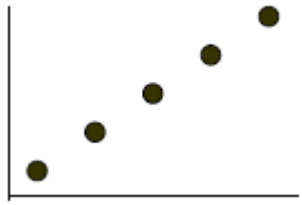
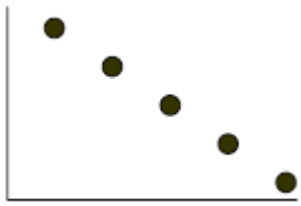
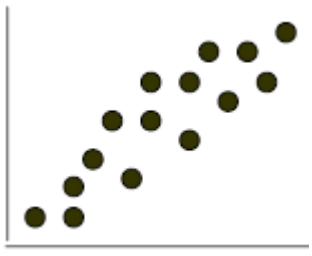
ΣYi = Sumatoria de las puntuaciones de la variable Y.

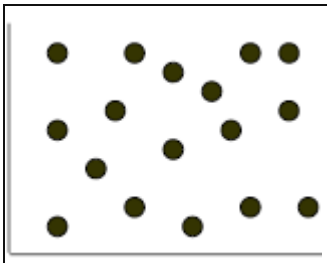
ΣXY = Sumatoria de los productos de las puntuaciones apareadas XiYi.

ΣXi^2 = Sumatoria de los cuadrados de las puntuaciones de la variable X.

ΣYi^2 = Sumatoria de los cuadrados de las puntuaciones de la variable Y.

Tabla 4.3 Ejemplo de Gráficas de Dispersión.

Gráfica de Dispersión	Coeficiente de correlación	Interpretación
	+ 1.00	Correlación positiva perfecta.
	- 1.00	Correlación negativa perfecta.
	+ 0.60	Correlación positiva moderada

	0.00	Ausencia de correlación

Para ejemplificar el coeficiente de correlación producto-momento se desarrollará el análisis de correlación al volumen de exportaciones del Sector de Servicios de Impresión (SSI) de México en el periodo comprendido entre 1991 y 1995 en relación con el tipo de cambio peso-dólar. El volumen de exportaciones se expresa en millones de dólares. La Tabla 4.4 muestra los datos del ejemplo. Al aplicar la ecuación 4.9 se obtiene:

$$r_{xy} = \left[39108323-(3372.9)/5(6123.84)/5\right]$$

$$\sqrt{\left[232922069-(3372.9)^2/5\right]\left[1867511092-(6123.8^{4} = -0.28\right.}$$

Tabla 4.4 Volumen de Exportaciones en relación con el Tipo de Cambio peso- dólar del Sector de Servicios de Impresión Mexicano.

AÑO	EXPORTACIONES	TIPO DE CAMBIO (Y)	X^2	Y^2	XY
1991	621.8	3016.69	386386.56	9100418.56	1875174.50
1992	654.8	3094.29	428763.04	9574630.6	2026141.09
1993	662.3	3.1091	438641.29	9.67	2059.16
1994	561.6	3.3751	315394.56	11.39	1895.46
1995	871.8	6.38	760035.24	40.70	5562.08
Total	3372.1	6123.84	2329220.69	18675110.92	3910832.30

Fuente: Avila, H. L. (1999). Determinación de la Productividad Total del Sector de Servicios de Impresión de Cd. Cuauhtémoc, Chih. Tesis para obtener el grado de M. C. en Comercio Exterior, Instituto Tecnológico de Cd. Juárez, Juárez, Chih.

al calcular el coeficiente de determinación (ver sección 5.3) se obtiene un valor de:

$$r_{xy}^2 = -0.28^2 = 0.06$$

el análisis de correlación arrojó un coeficiente de correlación de – 0.28 para la asociación del valor total de exportaciones con el tipo de cambio peso-dólar, esto indica una débil correlación inversa entre ambas variables, con un coeficiente de determinación de 0.06. Lo anterior sugiere la conclusión lógica de que mientras el volumen de exportaciones se incrementa, el tipo de cambio peso-dólar decrece, sin que lo anterior indique una relación causística, dado que para tal efecto sería necesario un análisis marginal con soporte en algún modelo económico. El análisis de correlación simple es susceptible de someterse a prueba de hipótesis estadística mediante la distribución t con gl = n – 2 (gl = grados de libertad). Para lo anterior se procede a:

a) Establecer la hipótesis nula expresada en términos estadísticos (ver sección 2.2 del capítulo II). La hipótesis es:

$$Ha = r_{xy} = 0$$
$$H_1 = r_{xy} \neq 0$$

b) Determinar el nivel de significancia estadística al que se someterá a contrastación la hipótesis nula y que pudiera ser en nivel de:

$$\alpha = 0.05$$
$$\alpha = 0.01$$
$$\alpha = 0.10$$

c) Calcular la prueba de signifitica mediante el

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

(Ec. 4.10)

d) estadístico t expresado en la ecuación:

donde:

t = prueba t para prueba de hipótesis de correlación simple

r = coeficiente de correlación

r^2 = coeficiente de determinación

n = número de casos

al aplicar la ecuación 4.10 al análisis de correlación anterior se obtiene un valor t calculado de:

$$t = \frac{-0.28}{\sqrt{\frac{1-(-0.28)^2}{5-2}}} = -0.90$$

el valor t calculado de – 0.90 se compara con el valor t crítico a una significación de α = 0.05 consultado en el apéndice A, procediendo previamente a obtener los grados de libertad para la distribución t con la ecuación:

gl = n - 2

(Ec. 4.11)

en el problema son cinco casos por lo que gl = 5 – 2 = 3. Con tres grados de libertad el valor t crítico es de 3.182. La regla de decisión es que si el valor t calculado es mayor que el valor t crítico entonces se rechaza la hipótesis de nulidad. En este caso se acepta la hipótesis de nulidad en virtud de que el valor to = - 0.90 < tc = 3.182 y se concluye que si existe asociación entre las variables volumen de exportaciones y tipo de cambio peso-dólar.

e2) Coeficiente de Correlación por Rangos. El coeficiente de correlación por rangos conocido como coeficiente de Spearman (rho) se obtiene por medio de la expresión:

$$\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

(Ec. 4.10)

donde:

ρ = Coeficiente de correlación por rangos.

ΣD^2= Sumatoria de los cuadrados de las diferencias entre los rangos.

N = Número de casos.

e3) Coeficiente de Determinación. El coeficiente de determinación (rXY2) se define como el cuadrado del coeficiente de correlación y se utiliza para medir la variación de la variable dependiente (Y) explicada por la variación de la variable independiente (X). Es más adecuado aplicarlo en modelos de regresión lineal para medir el poder explicativo de modelos de regresión lineal.

5.2.2 Elementos de Estadística Inferencial

a) Análisis de Varianza. El análisis de varianza (ANOVA) es una técnica estadística diseñada para comparar la varianza de dos poblaciones a partir del análisis de las varianzas de las muestras respectivas. Webster (1998) aplica el concepto de ANOVA al contexto de un experimento y la define como "... el procedimiento [que] se puede aplicar a la determinación de si un tratamiento en particular aplicado a una población tendrá efecto significativo sobre su media," (p. 595). Aplicar el ANOVA requiere cumplir con dos criterios específicos:

a1) Las poblaciones de estudio deber ser normales y tener varianzas iguales.

a2) Seleccionar las muestras independientemente.

La varianza total de todos los tratamientos (observaciones) se puede dividir en dos fuentes:

a) Variación Intermuestral. Factor que representa la variación entre los diversos tratamientos administrados durante el desarrollo de un experimento.

b) Variación Intramuestral o debida al Error. Factor que representa la variación dentro de un mismo tratamiento administrado durante la realización de un experimento.

En este contexto se entiende que la variación total es igual a la variación intermuestral + la variación intramuestral o debida al error. Para obtener la comprobación de una hipótesis de nulidad mediante el ANOVA se tienen que calcular los siguientes factores:

a) La suma total de cuadrados expresada por la ecuación:

$$\sum xt^2 = \frac{\sum X^2 - \left(\sum X\right)^2}{N}$$

(Ec. 4.11)

donde:

$\sum xt^2$ = Suma total de cuadrados

$\sum x^2$ =Suma de los cuadrados de las puntuaciones

$(\sum x)^2$ = Suma de las puntuaciones elevadas al cuadrado

N = Número de casos.

b) La suma de los cuadrados entre grupos (varianza intermuestral) se expresa por la ecuación:

$$\sum xi^2 = \frac{\left(\sum X_1\right)^2}{n_1} + \frac{\left(\sum X_2\right)^2}{n_2} + \ldots + \frac{\left(\sum X_K\right)^2}{n_k}$$

(Ec. 4.12)

donde:

Σxi^2 = Suma de los cuadrados entre los grupos

$(\Sigma X1)^2$ = Suma de las puntuaciones elevadas al cuadrado del tratamiento1

n = Número de casos

c) La suma de cuadrados dentro de grupos (varianza intramuestral) se expresa por la ecuación:

$$\sum Xd^2 = \sum X_1^2 - \frac{\left(\sum X_1\right)^2}{n_1} + \ldots$$

(Ec. 4.13)

donde:

ΣXd^2 = Suma de cuadrados dentro de grupos

$\Sigma X1^2$ = Suma de los cuadrados de las puntuaciones del grupo 1

$(\Sigma X1)^2$ = Suma de las puntuaciones elevadas al cuadrado del tratamiento1

Lo anterior refleja que se cuenta con tres varianzas y solo es posible realizar la comparación de la varianza intermuestral con la varianza intramuestral mediante el análisis del comportamiento de las mismas con respecta a la distribución F que supone la independencia de las varianzas. La distribución F se expresa por la ecuación:

$$F = \frac{La\ mayor\ e\ stimacion\ \ de\ la\ \ \text{var}\ ianza}{La\ menor\ e\ stimación\ \ de\ la\ \ \text{var}\ ianza}$$

(Ec. 4.14)

donde:

F = Distribución F.

d) Los grados de libertad para la varianza del error se obtienen mediante la ecuación:

gl =c-1

(Ec. 4.15)

donde:

gl = grados de libertad

c = columnas

e) Los grados de libertad para la varianza intramuestral se obtienen por medio de la ecuación:

gl =n-c

(Ec. 4.16)

donde:

gl = grados de libertad

c = columnas

n = número de casos

Para ejemplificar el ANOVA se tomarán los datos siguientes: Con el propósito de determinar que las medias de las puntuaciones obtenidas por tres grupos de menonitas provenientes de los campos menonitas del municipio de Riva Palacio, Chih., en un experimento de lectura veloz en indioma español utilizando un vocabulario técnico-científico. El rendimiento de cada uno de los grupos se muestra en la Tabla 4.5.

Tabla 4.5 Resultados de un experimento de lectura veloz con una muestra de menonitas del Municipio de Riva Palacio, Chih.

A	A^2	B	B^2	C	C^2
60	3600	90	8100	80	6400
80	6400	85	7225	70	4900
74	5476	84	7056	100	10000
90	8100	76	5776	74	5476
66	4356				
79	6241				
Total **449**	**34173**	**245**	**28157**	**324**	**26776**

$$\sum xb^2 = 89106 - \frac{(1018)^2}{14} = 15082.86$$

$$\sum xt^2 = \frac{(449)^2 + (245)^2 + (324)^2 - (1018)^2}{} = 827.50$$

$$\sum Xd^2 = 15082.86 - 827.50 = 14255.36$$

Para obtener la razón F se recomienda elaborar la siguiente tabla a fin de facilitar el análisis de resultados:

Tabla 4.6 Formato para Obtener la Razón F.

Fuente de Variación	Suma de Cuadrados	126	Cuadrado Medio	*F* Calculada	α
Intermuestral	827.50	2	413.75	0.44	0.05
Intramuestral	14255.36	15	950.36		

Valor crítico de *F* para α = 0.05. $_{0.95}F_{2,15} = 3.68$

La razón F crítica para 2 y 15 grados de libertad a un nivel α = 0.05 se obtiene consultando el apéndice B. Para el presente caso tiene un valor de F = 3.68. La interpretación requiere de aplicar la regla de decisión: Si la razón F calculada es mayor que la razón F crítica entonces se rechaza la hipótesis nula, en caso contrario se acepta.

Para el problema anterior la razón F calculada es de 0.44 valor que esta muy por debajo de la razón F crítica con valor de 3.68 para α = 0.05 por consiguiente es posible aceptar la hipótesis de nulidad concluyendo que no existe evidencia de que las tres medias de calificaciones obtenidas por los grupos en lectura veloz sean diferentes.

b) Análisis Multifactorial de Varianza. El análisis multifactorial de varianza (ANCOVA) también denominado análisis de covarianza permite la comparación de más de dos variables entre si con el propósito de comprobar tanto el efecto de las variables como el efecto de interacción entre ellas.

El ANCOVA es utilizado para analizar los resultados de investigaciones de tipo experimental que aplican un diseño factorial (ver sección 3.4 del capítulo III). En este tipo de diseños se analizan los efectos combinados de dos o más variables independientes. Para realizar un ANCOVA se necesita obtener:

b1) La suma total de cuadrados mediante la ecuación 4.11.

b2) La suma de cuadrados entre grupos mediante la ecuación 4.12.

b3) La suma de cuadrados dentro de grupos mediante la ecuación 4.13.

b4) La suma de cuadrados entre columnas que se define por la ecuación:

$$\sum X_{ec}^2 = \frac{\left(\sum X_{c1}\right)^2}{n_{c1}} + \frac{\left(\sum X_{c2}\right)^2}{n_{c2}} + \ldots - \frac{\left(\sum X\right)^2}{N}$$

(Ec. 4.17)

donde:

ΣX_{ec}^2 = suma de cuadrados entre columnas

n_{c1} = número de casos en la columna 1.

N = número total de casos

b5) La suma de cuadrados entre hileras que se expresa matemáticamente mediante la ecuación:

$$\sum X_{er}^2 = \frac{\left(\sum X_{r1}\right)^2}{n_{c1}} + \frac{\left(\sum X_{r2}\right)^2}{n_{c2}} + \ldots - \frac{\left(\sum X\right)^2}{N}$$

(Ec. 4.18)

donde:

ΣX_{ec}^2 = suma de cuadrados entre columnas

nr1 = número de casos en la hilera 1.

N = número total de casos

b6) La suma de la interacción de los cuadrados "... es la parte de la desviación entre las medias de los grupos y la media total que no se debe ni a las diferencias de las hileras ni a las diferencias de las columnas," (D"Ary, Jacobs y Razavieh, 1982, p. 164). Se define matemáticamente mediante la expresión:

$$\sum X_{\text{int}}^2 = \sum X_i^2 - \left(\sum X_{ec}^2 + \sum X_{er}^2\right)$$

(Ec. 4.19)

donde:

ΣX_{int}^2 = suma de cuadrados entre hileras

ΣX_i^2 = suma de cuadrados entre grupos

ΣX_{ec}^2 = suma de cuadrados entre columnas

ΣX_{er}^2 = suma de cuadrados entre hileras

b7) Determinar el número de grados de libertad asociados a cada puntuación de variación:

1 Suma de cuadrados entre columnas utilizando la ecuación 4.15

2 Suma de cuadrados entre hileras mediante la ecuación:

gl = r-1

(Ec. 4.20)

donde:

gl = grados de libertad

r = hileras

3 Suma de la interacción de los cuadrados por medio de la expresión:

gl =(c-1)(r-1)

(Ec. 4.21)

donde:

gl = grados de libertad

c = columnas

r = hileras

4 Suma de cuadrados entre grupos mediante la ecuación:

gl =G-1

(Ec. 4.22)

donde:

gl = grados de libertad

G = grupos

5 Suma total de cuadrados definido por la expresión:

gl =N-1

(Ec. 4.23)

donde:

gl = grados de libertad

N = columnas

b8) Obtención de la razón F mediante la ecuación 4.14.

Para ejemplificar el ANCOVA se aplicará el procedimiento al siguiente caso: Se desea investigar como influye un programa de incentivación económica en la productividad de la mano de obra de una compañía de servicios de impresión, formando cuatro grupos de trabajadores aleatoriamente. Los integrantes de dos de los grupos son menores de 24 años y los integrantes del resto del grupo mayores de 24 años. Los datos obtenidos se muestran en la Tabla 4.7.

Tabla 4.7 Valores de la Productividad de la Mano de Obra de la Empresa de Servicios de Impresión.

	Incentivación: Alta		Incentivación: Baja		
Edad: Menos de 24	GRUPO 1 22 22 21 21 19 $\Sigma X = 105$	$\Sigma X^2 = 2211$	GRUPO 2 25 23 22 21 20 $\Sigma X = 111$	$\Sigma X^2 = 2479$	$\Sigma X_{r1} = 216$
Edad: Mas de 24	GRUPO 3 23 22 21 20 19 $\Sigma X = 105$	$\Sigma X^2 = 2215$	GRUPO 4 19 17 16 15 13 $\Sigma X = 80$	$\Sigma X^2 = 1300$	$\Sigma X_{r2} = 185$
	$\Sigma X_{c1} = 210$		$\Sigma X_{c2} = 191$		

los cálculos son:

$$\sum Xt^2 = 8205 - \frac{(401)^2}{20} = 164.95$$

$$\sum X_i^2 = \frac{(105)^2}{5} + \frac{(111)^2}{5} + \frac{(105)^2}{5} + \frac{(80)^2}{5} - \frac{(401)^2}{20} = 114$$

para obtener la razón F se sugiere diseñar la siguiente tabla con el propósito de facilitar

$$\sum Xd^2 = 164.95 - 114 = 50.95$$

$$\sum X_{ec}^2 = \frac{(210)^2}{10} + \frac{(191)^2}{10} - \frac{(401)^2}{20} = 18.10$$

$$\sum X_{er}^2 = \frac{(216)^2}{10} + \frac{(185)^2}{10} - \frac{(401)^2}{20} = 48.10$$

$$\sum X_{\text{int}}^2 = 114 - (18.10 + 48.10) = 47.80$$

el análisis:

Tabla 4.8 Formato para Obtener la Razón F.

Fuente de Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	*F* Calculada	α
Entre columnas	18.10	1	18.10	5.69	0.05
Entre hileras	48.10	1	48.10	15.13	
Interacción	47.80	1	47.80	15.03	
Intermuestral	827.50	3	38.00		
Intramuestral	14255.36	16	3.18		

Valor crítico de *F* para $\alpha = 0.05$. $_{0.95}F_{1,16} = 4.49$

Para la varianza entre columnas la razón Fo = 5.69 > Fc = 4.49 por consiguiente no se acepta la hipótesis de nulidad. La razón F calculada (Fo) es significativa a nivel α = 0.05. Para la varianza entre hileras la razón Fo = 15.13 > Fc = 4.49 por lo que no se acepta la hipótesis nula. La razón F es altamente significativa a nivel α = 0.05. Para la varianza de la interacción la razón Fo = 15.03 > Fc = 4.49 por lo que no se acepta la hipótesis de nulidad. La razón F es altamente significativa a nivel α = 0.05. Los anteriores resultados permiten concluir que existe evidencia estadística para establecer como conclusión que la incentivación económica tiene influencia significativa en el aumento de la productividad de los empleados de la compañía de servicios de impresión. Este efecto se presenta tanto en trabajadores menores de 24 años como en los mayores de 24 años.

c) La Distribución Xi2. La Xi2 (chi cuadrada) es una prueba de estadística no paramétrica que se utiliza para la contrastación de hipótesis. De acuerdo con Webster (1998) "las pruebas no paramétricas son procedimientos estadísticos que se pueden utilizar para contrastar hipótesis cuando no es posible fijar ningún supuesto sobre parámetros o distribuciones poblacionales," (p. 836). Las aplicaciones de la prueba Xi2 son dos: c1) las pruebas de bondad del ajuste y c2) las pruebas de independencia.

c1) Xi2 de bondad del ajuste. Esta prueba se utiliza para apreciar si las distribuciones observadas se ajustan a las esperadas. La prueba es adecuada para realizar pruebas de

variancia sin que interese el tipo de distribución que tiene (Glass y Stanley, 1994; Kazmier, 1998).

Lo anterior significa que esta prueba permite determinar si los datos empíricos de alguna distribución específica corresponde a una distribución teórica como la binomial, la poisson o la normal. Se emplea en el muestreo con el propósito de precisar si los valores obtenidos de una muestra corresponden a las frecuencias poblacionales (ver Hopkins, Hopkins y Glass, 1997; Kazmier, 1998).

Para Webster (1998) presenta una definición muy completa de las pruebas de bondad del ajuste. "... estas pruebas miden el grado en que los datos muestrales observados cumplen una distribución hipotética determinada. Si el grado de cumplimiento es razonable, se puede deducir que la distribución hipotética existe," (p. 838).

La hipótesis de nulidad en la prueba de bondad del ajuste se expresa:

Ho: fo = fe. (No hay diferencia entre las frecuencias observadas y las esperadas)

H1: fo ǂ fe. (Existe diferencia entre las frecuencias observadas y las esperadas.

Para someter a prueba estas hipótesis se utiliza la expresión matemática:

$$Xi^2 = \frac{\sum_{i=1}^{k} (fo - fe)^2}{fe}$$

(Ec. 4.24)

donde:

Xi^2= prueba chi cuadrada

k = número de categorías o clases

fo = frecuencias observadas

fe = frecuencias esperadas

Para ejemplificar la Xi2 de bondad del ajuste se utilizarán los siguientes datos: El Sr. David Neufeld es gerente de ventas de la fábrica de queso menonita tipo chester ubicada en la Colonia Manitoba en la región noroeste del Estado de Chihuahua. En particular el Sr. Neufeld tiene que desplazar la producción de queso en el mercado nacional. Recientemente se da cuenta de la existencia de una fuerte competencia de otras marcas de queso provenientes de otras entidades del país y del extranjero. Le resulta cada vez más difícil comercializar la producción de queso y decide someter a comprobación la hipótesis de nulidad a un nivel $\alpha = 0.05$:

Ho: fo = fe. La demanda real es uniforme a la esperada

H1: fo ǂ fe. La demanda real no es uniforme a la esperada.

el Sr. Neufeld toma como muestra el volumen de ventas mensual en toneladas de queso correspondientes a un periodo de 12 meses. Las frecuencias son:

Tabla 4.9 Frecuencias Esperadas y Observadas de las Ventas de Queso

Mes	Frecuencias	
	Esperadas	Observadas
Enero	*60*	*43*
Febrero	60	41
Marzo	60	75
Abril	60	71
Mayo	60	59
Junio	60	69
Julio	60	45
Agosto	60	51
Septiembre	60	61
Octubre	60	65
Noviembre	60	50
Diciembre	60	90
Total	720	720

El valor de Xi^2 es:

$$Xi^2 = \sum_{i=1}^{1} \left[\begin{array}{l} \frac{(43-60)^2}{60} + \frac{(41-60)^2}{60} + \frac{(75-60)^2}{60} + \frac{(71-60)^2}{60} \\ + \frac{(71-59)^2}{60} + \frac{(69-60)^2}{60} + \frac{(45-60)^2}{60} + \frac{(51-60)^2}{60} \\ + \frac{(61-60)^2}{60} + \frac{(65-60)^2}{60} + \frac{(50-60)^2}{60} + \frac{(90-60)^2}{60} \end{array} \right] = 42.57$$

Con el propósito de analizar e interpretar el valor encontrado de Xi^2 se necesita obtener los grados de libertad. Para ello se utiliza la expresión matemática:

gl = n - 1

donde:

gl = grados de libertad

n = número de casos

para el anterior problema los gl = 11. Posteriormente se obtiene el valor de Xi2 crítica(consultar apéndice C) aplicando la regla de decisión: Se rechaza la hipótesis nula sí Xi2 calculada > Xi2 crítica, no se rechace en caso contrario. Para el caso anterior se puede concluir que Xi2 calculada = 42.57 > Xi2 crítica =19.675 a nivel α = 0.05 por consiguiente no se acepta la hipótesis de nulidad que expresa que la demanda de queso menonita tipo chester en el mercado nacional es uniforme. Las diferencias entre la demanda observada y la esperada son significativas por lo que es posible refutar la hipótesis de nulidad.

c2) Xi^2 de Independencia. Es una excelente herramienta estadística para comprobar la independencia de variables categóricas. Analiza dos factores con el propósito de

determinar la existencia o no de relación entre ellos. Para lo anterior utiliza tablas de tabulaciones cruzadas o de contingencia (ver sección 4.3 del capítulo IV).

Así por ejemplo si se analizará el rendimiento de alumnos con resultados por arriba o por debajo del promedio en la prueba coeficiente intelectual se estarían comparando dos factores: rendimiento y coeficiente intelectual. La Xi2 de independencia aplica la ecuación 4.23 para analizar la diferencia entre las frecuencias observadas y las esperadas.

Para ilustrar esta prueba se utilizará un ejemplo según el cual a una muestra aleatoria de 90 estudiantes recién egresados y próximos a egresar de educación media superior se les pregunta si prefieren estudiar una carrera profesional en la Universidad Autónoma de Chihuahua (UACH), en el Instituto Tecnológico de Cd. Cuauhtémoc (ITCC) o si tienen preferencia por alguna Institución de Educación Superior Particular. Los resultados se muestran en la Tabla 4.9.

Tabla 4.10 Frecuencias Observadas para la Preferencia por Educación Superior.

como puede observarse la Tabla 4.9 contiene 6 casillas integradas por tres columnas y dos hileras. Para realizar la comparación se plantea la hipótesis:

Ho: fo = fe. No existe preferencia por alguna institución específica

H1: fo ǂ fe. Existe preferencia por alguna institución específica.

para someter a comprobación la hipótesis de nulidad se elige un nivel de $\alpha = 0.05$ procediéndose a obtener las frecuencias esperadas en función de las frecuencias observadas por medio de operaciones aritméticas. Se multiplica el valor de cada casilla por el total de la columna dividido por el total de casos.

Así por ejemplo la frecuencia esperada de los estudiantes que prefieren la UACH es 20 (36/90) = 8. Se realiza la misma operación para obtener el resto de frecuencias esperadas. La Tabla 4.10 presenta las frecuencias esperadas.

$$Xi^2 = \sum_{i=1}^{2}\left[\frac{(20-8)^2}{8} + \frac{(24-11.20)^2}{11.20} + \frac{(4-0.53)^2}{0.53} + \frac{(16-6.4)^2}{6.4} + \frac{(18-8.46)^2}{8.46} + \frac{(8-1.04)^2}{1.04}\right] = 87.20$$

Tabla 4.11 Frecuencias Esperadas para la Preferencia por Educación Superior.

	Preferencia			
	UACH	ITCC	Institución Particular	Total
Recien egresados	8	11.20	0.53	19.73
Próximos a egresar	6.4	8.46	1.04	15.90
Total	14.4	19.66	1.57	35.63

el valor de Xi^2 es:

Finalmente se compara el valor de Xi^2 observada con el valor de Xi^2 crítica. Para lo cual se requiere obtener los grados de libertad mediante la expresión:

gl = (r-1) (c-1)

(Ec. 4.25)

donde:

gl = grados de libertad

r = número de renglones en la tabla

c = número de columnas en la tabla

por consiguiente los grados de libertad son gl = (2 – 1) (3 – 1) = 2 que a un nivel α = 0.05 el valor de Xi^2 crítica es de 5.99. Como Xi^2=87.20 > Xi^2 crítica = 5.99 es posible afirmar con cierto grado de confianza que existen diferencias significativas acerca de la relación de la variable estudiantes recién egresados y próximos a egresar de educación media superior

y la variable preferencia por alguna institución de educación superior. La hipótesis de nulidad no se acepta a nivel $\alpha = 0.05$.

RESUMEN DEL CAPITULO

La etapa de análisis de datos es una de las más importantes en el proceso de investigación en virtud de que se procede a racionalizar los datos colectados con el propósito de explicar las posibles relaciones que expresan las variables estudiadas. El análisis puede ser univariado, bivariado o trivariado. El análisis e interpretación requiere del conocimiento de la estadística.

La estadística proporciona innumerables beneficios a la investigación científica y tecnológica. Esta disciplina aporta elementos estadísticos descriptivos e inferenciales. Los primeros representan un conjunto de procedimientos que permiten procesar y presentar la información de manera organizada y resumida. Los segundos facilitan el establecimiento de inferencias de la muestra estudiada hacia la población de origen a través de una serie de pruebas de hipótesis aplicando estadística paramétrica y no paramétrica.

Fuente: Introducción a la Metodología de la Investigación. Autor: Héctor Luis Ávila Baray URL: http://www.cyta.com.ar