



UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO

ESCUELA NACIONAL DE ESTUDIOS  
PROFESIONALES  
ACATLAN

APLICACIONES DE ESTADISTICA BASICA  
EN MEXICO

T E S I S

que para obtener el título de

ACTUARIO

P R E S E N T A

LOURDES DE LA LUZ GODOY DOMINGUEZ

ACATLAN, EDO. DE MEXICO 1993



TESIS CON  
FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INTRODUCCION .....	1
CAPITULO 1. APLICACIONES EN ECONOMIA	
1.1. ACTIVIDAD PRODUCTIVA .....	6
1.1.1. Empresas manufactureras, comerciales y de servicios (I)	
1.1.2. Empresas manufactureras, comerciales y de servicios (II)	
1.1.3. Productividad de la mano de obra	
1.1.4. Recursos naturales y producción	
1.2. INFLACION Y DESEMPLEO .....	32
1.2.1. Inflación y zonas salariales	
1.2.2. Niveles de desocupación en dos ciudades	
1.3. FINANZAS PUBLICAS .....	41
1.3.1. Gasto del Gobierno Federal	
1.4. MICROECONOMIA .....	47
1.4.1. Recursos financieros en la la instalación de un sistema automatizado de información	
CAPITULO 2. APLICACIONES EN DEMOGRAFIA	
2.1. ESTRUCTURAS POR EDAD Y SEXO .....	51
2.1.1. Cambios en la estructura por edad	
2.2. FECUNDIDAD .....	57
2.2.1. Uso de métodos anticonceptivos	
2.3. MORTALIDAD .....	64
2.3.1. Causas principales de mortalidad	
2.3.2. Mortalidad infantil	
2.4. DISTRIBUCION TERRITORIAL DE LA POBLACION .....	70
2.4.1. Crecimiento de la población urbana	
CAPITULO 3. APLICACIONES EN CIENCIAS SOCIALES	
3.1. SALUD .....	66
3.1.1. Unidades medicas del sector salud	
3.2. EDUCACION .....	93
3.2.1. Elección del área profesional y el sexo del estudiante	
3.2.2. Numero de alumnos por maestro	

3.3. CULTURA .....	103
3.3.1. Uso de bibliotecas	
3.3.2. Espectáculos públicos (I)	
3.3.3. Espectáculos públicos (II)	
3.3.4. Museos y zonas arqueológicas	
3.4. SEGURIDAD Y ORDEN PUBLICO .....	135
3.4.1. Accidentes de tránsito terrestre	
CAPITULO 4. APLICACIONES EN AREAS DIVERSAS	
4.1. CONTAMINACION .....	141
4.1.1. Monóxido de Carbono y el programa "Un día sin auto"	
4.2. ECOLOGIA .....	150
4.2.1. El comportamiento reproductivo de una palma tropical	
4.2.2. Dos especies maderables en el desierto	
4.3. PSICOLOGIA .....	162
4.3.1. Conocimientos y actitudes sobre el sexo	
4.4. TRANSPORTE .....	178
4.4.1. EL METRO	
CONCLUSIONES .....	182
BIBLIOGRAFIA .....	185
APENDICE A. PRUEBAS DE HIPOTESIS .....	189
APENDICE B. PRUEBAS DE BONDAD DE AJUSTE .....	192
APENDICE C. PRUEBAS DE INDEPENDENCIA .....	193
APENDICE D. ANALISIS DE PERSESION .....	195
APENDICE E. ESTADISTICA NO PARAMETRICA .....	199

## INTRODUCCION

Resulta muy halagador que el lector se interese en leer estas líneas introductorias, pero sería aún más gratificante que este sencillo trabajo de tesis le fuera de alguna utilidad. No hay seguridad de que así sea y, aún más, ni siquiera se conoce la probabilidad de que dado que apenas ha leído poco más de cinco líneas de esta hoja, decida seguir adelante. Además, resultaría muy interesante poder estimar el número total de páginas o, tal vez, de renglones que serán leídos; e investigar si ello depende de las aspiraciones y formación profesionales del lector, de su sexo o simplemente de su estado de ánimo. Es así, en algo tan sencillo como lo que se desprende del interés porque este trabajo al menos sea hojeado, donde se puede aplicar la Estadística.

El presente trabajo está dirigido principalmente a quienes, por lo menos, poseen una base de conocimientos sobre materias de Estadística. Si esta es la situación del lector, seguramente ya se habrá percatado de que, en muy pocos casos, la bibliografía disponible relacionada con temas estadísticos, incluye aplicaciones basadas en datos reales que, además, son producto de fenómenos y situaciones ocurridas actual o recientemente en México.

Los anteriores, son elementos que se pretende estén contenidos en una serie de aplicaciones prácticas, breves y sencillas que, con el objeto de ilustrar el uso de algunas técnicas de estadística básica, constituyen este trabajo de tesis.

Se incluyen pruebas de hipótesis sobre media y varianza, bajo el supuesto de distribución Normal, considerando dos o más poblaciones. También se consideran pruebas sobre el parámetro  $P$ , suponiendo que se tiene una muestra procedente de una población con distribución Bernoulli.

Por otra parte, se considera la aplicación de pruebas de bondad de ajuste, específicamente la de Kolmogorov-Smirnov para una población y la de Lilliefors, así como la prueba gráfica de

#### Papel Normal.

En otros casos se plantean pruebas de independencia, a través de tablas de contingencia bi y tri-dimensionales.

Como una introducción a la aplicación de técnicas más avanzadas se presentan ejercicios de Análisis de Regresión Simple y Múltiple.

Adicionalmente, mostrando como alternativa el uso de técnicas de estadística no paramétrica, para problemas en los cuales los supuestos necesarios para aplicar estadística paramétrica parecen no satisfacerse, se ilustran pruebas como la de Wilcoxon, los análisis de varianza de Friedman y de Kruskal-Wallis, y medidas de asociación como el coeficiente de correlación  $r$  de Kendall y el coeficiente de Contingencia.

En general, el esquema de cada una de las aplicaciones incluye los siguientes elementos:

- a) El planteamiento del problema y de la hipótesis estadística a probar, con base en el contexto particular del mismo.
- b) La elección de la técnica estadística adecuada para probar la hipótesis planteada, considerando la naturaleza del cuestionamiento y el tipo de información de que se dispone. En algunos casos se hace necesario el uso de una prueba de bondad de ajuste.
- c) Descripción de la información a emplear.
- d) Cálculo de los estadísticos de prueba pertinentes y determinación de su probabilidad de ocurrencia, conforme a  $H_0$ .
- e) Formulación de conclusiones.
- f) Presentación de cuadros y gráficas.

Cabe mencionar que los cálculos numéricos fueron realizados con la ayuda de programas de cómputo, principalmente: SYSTAT, SPSS, LOTUS 123 y DBASE III Plus.

El nivel de significancia de las pruebas, en general, corresponde a los valores usuales de  $\alpha$ , es decir 0.1 y 0.05.<sup>1</sup>

Toda la información estadística empleada es real, corresponde a SIEGEL, Sidney: *Diseño experimental no paramétrico aplicado a las ciencias de la conducta*, trad. por: Javier Aguilar Villalobos, 1a. ed., México, Trillas, 1970, p. 27.

a situaciones y fenómenos ocurridos en México, y hasta donde fue posible, es reciente.

El presente trabajo se encuentra dividido en 4 capítulos de acuerdo a la temática abordada. Así, en el Capítulo 1 se maneja información de índole económica relacionada con temas como productividad, inflación y finanzas públicas. El Capítulo 2 incluye el planteamiento de hipótesis basadas en datos demográficos que expresan características de la población como estructura por edad y sexo, mortalidad y fecundidad.

En el Capítulo 3 se muestra una gama de aspectos sociales a través de planteamientos muy específicos en torno a temas como salud, educación, cultura y seguridad y orden públicos.

Con el objetivo de ilustrar la gran diversidad de cuestionamientos en los que es posible aplicar la Estadística, en el Capítulo 4 se maneja información ligada a temas tales como contaminación ambiental, Ecología, Psicología y transporte público.

Finalmente, después de las conclusiones y la bibliografía, se presentan los apéndices A al E. Cada uno se divide a su vez en secciones y en algunos casos en subsecciones, en las cuales se exponen las hipótesis nula y alternativa, así como la fórmula matemática para calcular el estadístico de prueba correspondiente.

## CAPITULO 1

## CAPITULO 1 APLICACIONES EN ECONOMIA

### 1.1. ACTIVIDAD PRODUCTIVA

#### 1.1.1. Empresas manufactureras, comerciales y de servicios (I)

Para llevar a cabo la actividad productiva a la cual se dedica es necesario que toda empresa realice periódicamente ciertos gastos. Si se considera la relación numérica que existe entre los gastos y los ingresos, expresada por el cociente que resulta dividir estas dos cifras, ambas correspondientes al mismo periodo, podrá tenerse algún parametro para medir si el funcionamiento de la empresa es adecuado.

La magnitud de una empresa está estrechamente relacionada con su nivel de ocupación: se conoce como empresa menor a aquella con hasta cinco empleados y como empresa mayor a la que cuenta con mas de cinco.<sup>1</sup>

A continuación se intentará dar respuesta a lo siguiente: ¿La relación gastos-ingresos<sup>2</sup> se diferencia según la magnitud de las empresas?. Para ello se consideran datos agregados por entidad federativa de los dos tipos de establecimientos<sup>3</sup> lo cual será de utilidad para verificar estadísticamente la siguiente hipótesis:

H<sub>0</sub>: A nivel estatal, la relación gastos-ingresos de las empresas dedicadas a la actividad "y" es similar en las empresas menores y en las mayores.

vs.

H<sub>1</sub>: Dicha relación es diferente.

Donde:

y = manufacturera, comercial, de servicios.

<sup>1</sup> INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *Resultados oportunos. Censos Económicos 1985*. México, INEGI, 1987, p. 41.

<sup>2</sup> *Op. cit.*, p. 18-19.

<sup>3</sup> El cociente entre gastos e ingresos se encuentra multiplicado por cien.

Analizando las estadísticas generales del Cuadro 1.1 sobresale que la actividad en la cual el cociente de gastos respecto a ingresos es más pequeño en los dos tipos de empresas (48% y 10% respectivamente) es la prestación de servicios. También se observa que a excepción de esta misma actividad, el comportamiento de las cifras es más variable en el caso de los establecimientos menores.

Para determinar la técnica estadística adecuada para probar la hipótesis planteada anteriormente, primero se investigará si existe evidencia estadística de que la distribución de los datos es Normal, para lo cual se plantea la siguiente prueba de hipótesis:

H<sub>0</sub>: La relación gastos-ingresos por cien del tipo de empresas "x" dedicadas a la actividad "y" es Normal.

vs.

H<sub>1</sub>: La distribución de dicha relación no es Normal.

Donde:

x = menores, mayores;

y = manufacturera, comercial, de servicios.

Ya que la hipótesis anterior considera la distribución Normal, la prueba de Lilliefors<sup>4</sup> parece ser adecuada. Con n=32, dicha prueba arrojó los siguientes resultados:

ACTIVIDAD	TIPO DE ESTABLECIMIENTO	d <sub>n</sub>	d(0.9) <sup>*</sup>
MANUFACTURA	MAYORES	0.083	0.142
	MENORES	0.262	
COMERCIO	MAYORES	0.144	
	MENORES	0.205	
SERVICIOS	MAYORES	0.148	
	MENORES	0.131	

<sup>\*</sup>Ver tabla 15 en CONOVER W. J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & Sons, 1971, p. 398.

Ver apéndice B, sección II.

Al parecer, solo en el caso de los establecimientos mayores manufactureros y menores de servicios se puede considerar que la distribución de los porcentajes es Normal.

Dado que dicho supuesto distribucional no se verificó en todos los casos, una técnica alternativa para probar la primera hipótesis planteada es la prueba de los Rangos de Wilcoxon<sup>5</sup>, aplicada a cada par de variables, es decir a los cocientes de gastos sobre ingresos (por cien) por entidad de las empresas mayores y menores, para cada tipo de actividad. Se recomienda esta prueba pues se dispone de muestras apareadas cuyos datos se encuentran en una escala más que ordinal.

El Cuadro 1.2 indica que en 26 estados los establecimientos manufactureros mayores realizan gastos más cuantiosos que los menores respecto a los ingresos de cada uno, obteniéndose que las sumas de los rangos asociados a las diferencias en favor de los establecimientos mayores y en favor de los menores es, respectivamente:

$$R^+ = 117 \text{ y } R^- = 81.$$

En esta ocasión  $R=R^-$ , pues si las diferencias en favor de los establecimientos pequeños tienen un valor poco significativo, se concluirá entonces que los dos tipos de empresas no presentan un nivel similar de gastos relativo a los ingresos. Dado que el tamaño de muestra es  $n=32$ , es conveniente emplear la aproximación Normal<sup>6</sup>, para lo cual resulta que  $z$  toma el valor de:

$$z = -3.42,$$

con una probabilidad muy pequeña, menor que  $\alpha=0.1$ , de cometer el error de rechazar  $H_0$  (La relación gastos-ingresos es similar en empresas menores), dado que en realidad es verdadera. Por lo tanto, con un nivel de significancia  $\alpha=0.1$ , se rechaza  $H_0$ .

Un resultado similar al anterior se obtiene en relación a la actividad comercial, al analizar el Cuadro 1.3. Se dan, 27 diferencias en favor de los establecimientos mayores, además se observa que:

---

<sup>5</sup> Ver apéndice E, sección I.

<sup>6</sup> Ver apéndice E, fórmula E1.

$$R^+ = 443 \text{ y } R^- = 85 = R.$$

El criterio para elegir R es el mismo. Si se usa la aproximación Normal se obtiene que  $z = -3.35$ , con una probabilidad de ocurrencia menor que  $\alpha$ , bajo el supuesto de que existe similitud entre las relaciones gastos-beneficios de ambos tipos de empresas. Por lo tanto también se concluye que existe evidencia para rechazar  $H_0$ .

En el sector servicios puede apreciarse un comportamiento más equilibrado, puesto que de acuerdo al Cuadro 1.4, en 19 entidades los establecimientos mayores realizan gastos superiores, mientras que en los 13 restantes, son los menores los que gastan más en relación con sus ingresos, de tal forma que:

$$R^+ = 367 \text{ y } R^- = 161 = R,$$

estas cifras revelan que existe evidencia para rechazar  $H_0$  con un nivel de confianza del 90%, dado que la probabilidad de ocurrencia conforme a  $H_0$  de  $z = -1.93$  es de 0.0268, menor que  $\alpha$ .

En conclusión, se ha encontrado que en tres de las actividades económicas más importantes para México, como son la industria manufacturera, el comercio y la prestación de servicios, la relación gastos-ingresos es diferente según el tamaño de la empresa. Tomando en cuenta la información analizada se encontraron indicios de que los gastos relativos a los ingresos son más bajos en el caso de las empresas menores, para las tres actividades.

Conviene aclarar que la información empleada para efectuar esta prueba es ciertamente limitada si se considera que la clasificación en establecimientos menores y mayores divide a estos en dos grupos bastante amplios. También se recuerda que las cifras correspondientes a gastos no incluyen remuneraciones al personal, y que en general la información corresponde a las zonas urbanas.

Finalmente, queda abierta la interrogante respecto a las demás actividades económicas bajo un análisis similar.

Cuadro 1.1. ESTADÍSTICAS GENERALES DE LOS PORCENTAJES DE GASTOS RESPECTO A INGRESOS CORRESPONDIENTES A LOS ESTABLECIMIENTOS DEDICADOS A LAS ACTIVIDADES MANUFACTURERAS, COMERCIALES Y DE SERVICIOS, EN LAS ENTIDADES FEDERATIVAS DE LA REPÚBLICA MEXICANA.

ACTIVIDAD TIPO DE ESTABLEC.	MANUFACTURA		COMERCIO		SERVICIOS	
	MAYORES	MENORES	MAYORES	MENORES	MAYORES	MENORES
NO. DE OBS.	32	32	32	32	32	32
V. MIN.	54.057	53.688	60.943	67.650	48.237	19.125
V. MAX.	96.156	112.326	106.826	121.950	94.983	75.504
MEDIA	75.786	67.898	86.011	79.174	62.961	57.266
DESV. EST.	8.717	9.800	8.806	9.448	11.053	9.200

Fuente: Resultados oportunos, Censos Económicos 1986, INEGI, México, 1987, p. 40-41, 45-46, 50-51.

Cuadro 1.2. DIFERENCIAS ENTRE LAS RELACIONES GASTOS-INGRESOS DE LOS ESTABLECIMIENTOS MANUFACTUREROS MAYORES Y MENORES

ESTADO	di	di	ra
Aguascalientes	3.936	3.936	5
Baja California	10.535	10.535	17
Baja California Sur	14.621	14.621	23
Campeche	4.957	4.957	7
Coahuila	12.239	12.239	20
Colima	5.686	5.686	9
Chiapas	3.598	3.598	2
Chihuahua	10.107	10.107	15
Distrito Federal	3.797	3.797	4
Durango	21.628	21.628	28
Guanajuato	10.189	10.189	16
Guerrero	3.692	3.692	3
Hidalgo	16.178	16.178	25
Jalisco	9.699	9.699	12
Edo. de Méx.	9.695	9.695	13
Michoacán	-25.047	25.047	(-) 31
Morales	-2.453	2.453	(-) 1
Nayarit	-7.787	7.787	(-) 10
Nvo. León	9.852	9.852	6
Oaxaca	-7.960	7.960	(-) 11
Puebla	22.346	22.346	29
Queretaro	13.014	13.014	21
Quintana Roo	-3.953	3.953	(-) 5
San Luis Potosí	10.715	10.715	18
Sinaloa	16.020	16.020	24
Sonora	9.325	9.325	14
Tabasco	25.954	25.954	30
Tamaulipas	21.237	21.237	27
Tlaxcala	-13.389	13.389	(-) 22
Veracruz	11.165	11.165	19
Yucatán	19.094	19.094	26
Zacatecas	27.247	27.247	32

Cuadro 1.3. DIFERENCIAS ENTRE LAS RELACIONES GASTOS-INGRESOS DE LOS ESTABLECIMIENTOS COMERCIALES MAYORES Y MENORES.

ESTADO	di	di	ri
Aguascalientes	14.474	14.474	26
Baja California	-43.251	43.251	(-) 32
Baja California Sur	9.527	9.527	12
Campeche	12.995	12.995	23
Coahuila	6.135	6.135	4
Colima	11.107	11.107	19
Chiapas	-7.228	7.228	(-) 6
Chihuahua	10.333	10.333	14
Distrito Federal	5.150	5.150	3
Durango	-11.010	11.010	(-) 18
Guanajuato	12.584	12.584	22
Guerrero	11.754	11.754	21
Hidalgo	8.855	8.855	9
Jalisco	8.241	8.241	8
Edo. de Mèx.	9.654	9.654	13
Michoacan	10.791	10.791	16
Morelos	8.952	8.952	10
Nayarit	0.313	0.313	1
Nvo. Leon	10.389	10.389	15
Oaxaca	14.111	14.111	25
Puebla	16.878	16.878	29
Querétaro	16.153	16.153	28
Quintana Roo	22.480	22.480	31
San Luis Potosi	9.190	9.190	11
Sinaloa	-14.534	14.534	(-) 27
Sonora	7.928	7.928	7
Tabasco	6.239	6.239	5
Tamaulipas	11.660	11.660	20
Tlaxcala	10.920	10.920	17
Veracruz	-2.359	2.359	(-) 2
Yucatán	17.170	17.170	30
Zacatecas	13.212	13.212	24

Cuadro 1.4. DIFERENCIAS ENTRE LAS RELACIONES GASTOS-INGRESOS DE LOS ESTABLECIMIENTOS DE SERVICIOS MAYORES Y MENORES.

ESTADO	di	di	ri
Aguascalientes	11.411	11.411	22
Baja California	-17.573	17.573	(-) 27
Baja California Sur	15.815	15.815	25
Campeche	5.775	5.775	14
Coahuila	-6.757	6.757	(-) 16
Colima	-0.010	0.010	(-) 1
Chiapas	5.677	5.677	13
Chihuahua	-6.222	6.222	(-) 15
Distrito Federal	-4.840	4.840	(-) 9
Durango	5.529	5.529	12
Guanaajuato	9.387	9.387	20
Guerrero	1.621	1.621	4
Hidalgo	8.270	8.270	18
Jalisco	24.788	24.788	30
Edo. de Méx.	10.674	10.674	21
Michoacan	4.399	4.399	8
Morelos	-4.277	4.277	(-) 7
Nayarit	44.681	44.681	32
Nvo. León	-0.111	0.111	(-) 2
Oaxaca	5.189	5.189	11
Puebla	-5.172	5.172	(-) 10
Queretaro	-0.462	0.462	(-) 3
Quintana Roo	-11.655	11.655	(-) 23
San Luis Potosi	7.809	7.809	17
Sinaloa	2.949	2.949	6
Sonora	-2.227	2.227	(-) 5
Tabasco	23.041	23.041	29
Tamaulipas	-14.275	14.275	(-) 24
Tlaxcala	21.716	21.716	28
Veracruz	39.067	39.067	31
Yucatán	-9.346	9.346	(-) 19
Zacatecas	17.395	17.395	26

### 1.1.2. Empresas manufactureras, comerciales y de servicios (II)

Al retomar la tematica de la sección 1.1.1, ahora surge el siguiente cuestionamiento: ¿La relación gastos-ingresos de las empresas es semejante en las actividades de manufactura, comercio y servicios?.

Para dar respuesta a lo anterior se realizará un análisis por separado para las empresas mayores y otro para las menores, lo cual permitirá comparar los resultados de un tipo de empresas con el otro.

En la sección 1.1.1 se comprobó que la información de que se dispone debe ser tratada mediante una prueba no paramétrica. Dado que se consideran tres tipos de actividad en cada entidad, las cifras correspondientes se encuentran relacionadas y por ello se incurre en el manejo de muestras relacionadas, para lo cual una de las técnicas adecuadas es el análisis de la varianza de dos clasificaciones por rangos de Friedman<sup>4</sup>. La hipótesis a probar es:

H<sub>0</sub>: A nivel estatal, la relación gastos-ingresos de las empresas  
(mayores  
menores) manufactureras, comerciales y de servicios es similar.

vs.

H<sub>1</sub>: dicha relación es diferente para cuando menos una actividad.

Atendiendo primero a la información correspondiente a los establecimientos con más de cinco empleados, según el Cuadro 1.5, los rangos por renglón indican que en 24 de los 32 estados prevalece el mismo orden: en el sector servicios se eroga el menor porcentaje de los ingresos, en la actividad comercial el mayor y en la industria manufacturera el intermedio; no obstante, se presentan otros comportamientos como en el estado de Durango donde la actividad comercial efectúa los mayores gastos y en Nayarit, cuyo ramo de servicios presenta la mayor inversión en gastos de operación.

Resultados similares, aunque más extremos, se observan al considerar los establecimientos con personal menos numeroso. En 27  
Ver apéndice E, sección II-1.

estados se observa el mismo patrón descrito anteriormente.

Así, la suma de rangos por columna ( $R_j$ ) para cada actividad correspondiente a los establecimientos mayores y menores es respectivamente:

ACTIVIDAD	$R_j$ ESTABLECIMIENTOS	
	MAYORES	MEJORES
MANUFACTURA	86	83
COMERCIO	89	94
SERVICIOS	37	35

Si  $H_0$  fuera verdadera, se esperarían sumas de rangos ( $R_j$ ) similares. La prueba de Friedman determina la diferencia significativa entre estos valores a través de la estadística  $\chi_r^2$ , que en este caso asume, para cada tamaño de establecimiento, el valor:

$$\chi_r^2 = 42.41 \text{ y } \chi_r^2 = 54.44,$$

respectivamente.

En ambos casos la probabilidad de ocurrencia de  $\chi_r^2$ , conforme a  $H_0$ , localizada en tablas de la distribución Ji-Cuadrada con  $k-1=3-1=2$  grados de libertad, es menor que cualquier nivel de significancia usual, por lo cual, con una probabilidad de error muy pequeña se rechaza  $H_0$  para las empresas mayores y para las menores.

En conclusión puede decirse que la relación gastos-ingresos es diferente para cada actividad. Ocurre lo que de antemano se sabe: para llevar a cabo sus actividades las empresas comerciales son las que gastan más en relación a sus ingresos, las de servicios las que gastan menos y las manufactureras se ubican en el grado intermedio. Adicionalmente se encontró que el patrón descrito anteriormente se presenta tanto en empresas menores como en mayores. Cabe señalar que estas condiciones podrían verse alteradas al introducir las remuneraciones al personal dentro del concepto de gastos manejado.

Cuadro 1.5. RANGOS DE LA RELACION GASTOS-INSPECCIONES PARA LAS EMPRESAS DE MANUFACTURA, COMERCIO Y SERVICIOS

Entidad Federativa	ACTIVIDAD					
	ESTAB. MAY.			ESTAB. MEN.		
	M	C	S	M	C	S
Aguascalientes	2	3	1	3	2	1 *
Baja California	3	2	1 *	2	3	1
Baja California Sur	2	3	1	2	3	1
Campeche	2	3	1	2	3	1
Coahuila	2	3	1	2	3	1
Colima	2	3	1	2	3	1
Chiapas	2	3	1	2	3	1
Chihuahua	2	3	1	2	3	1
Distrito Federal	2	3	1	2	3	1
Durango	3	1	2 *	2	3	1
Guanajuato	2	3	1	2	3	1
Guerrero	2	3	1	2	3	1
Hidalgo	2	3	1	2	3	1
Jalisco	2	3	1	2	3	1
Edo. de Mex.	2	3	1	2	3	1
Michoacan	2	3	1	3	2	1 *
Morelos	2	3	1	2	3	1
Nayarit	1	2	3 *	2	3	1
Nvo. León	2	3	1	2	3	1
Oaxaca	1	3	2 *	2	3	1
Puebla	2	3	1	1	3	2 *
Queretaro	2	3	1	2	3	1
Quintana Roo	2	3	1	2	3	1
San Luis Potosí	2	3	1	2	3	1
Sinaloa	3	2	1 *	2	3	1
Sonora	2	3	1	2	3	1
Tabasco	2	3	1	2	3	1
Tamaulipas	2	3	1	1	3	2 *
Tlaxcala	1	3	2 *	2	3	1
Veracruz	3	2	1 *	2	3	1
Yucatán	2	3	1	1	3	2 *
Zacatecas	3	2	1 *	2	3	1

\* Casos en donde no se presenta la tendencia general.

### 1.1.3. Productividad de la mano de obra

El desarrollo de una economía se da gracias a la interrelación de los agentes que la integran.

Por ejemplo, en una economía como la de México, las empresas ofrecen a los consumidores los bienes que satisfacen sus necesidades de alimentación, salud, vestido y esparcimiento. Y a su vez los hogares o consumidores ponen a disposición de las empresas la mano de obra, indispensable en el proceso productivo.

Se sabe que la productividad de la mano de obra, definida como el cociente entre el Producto Interno Bruto (PIB)<sup>2</sup> y la población que fue empleada para obtenerlo, es decir, indica el valor monetario de lo que produjo cada empleado, no es igual para los tres grandes sectores de la producción, los cuales se mencionan a continuación:

- Primario. (Agricultura, Ganadería, Silvicultura, Pesca, etc.)
- Secundario. (Industria)
- Terciario. (Servicios)

De hecho, se sabe que en las actividades correspondientes al sector primario se presenta la productividad más baja, pues el número de trabajadores que se requiere es muy grande comparado con el valor monetario de lo que estos producen. Por otro lado, la mano de obra del sector servicios es la más productiva y la correspondiente al sector industrial se encuentra en medio de los otros dos sectores.

Con el fin de ejemplificar lo anterior, se realizará una prueba de hipótesis, basada en los datos de las tres primeras columnas del Cuadro 1.6, para lo cual se tiene lo siguiente:

---

<sup>2</sup> Suma de los valores monetarios de los bienes y servicios producidos por un país en un año. INSTITUTO NACIONAL DE ESTADÍSTICA GEOGRAFÍA E INFORMATICA: *El ABC de las Cuentas Nacionales*, México, INEGI, 1995, p. 15.

H<sub>0</sub>: La productividad es similar en los tres sectores productivos de la economía mexicana.

vs.

H<sub>1</sub>: Por lo menos uno de los tres sectores presenta diferente nivel de productividad, respecto a los dos restantes.

Para conocer el tipo de prueba que debe emplearse se aplicara una prueba de bondad de ajuste con el fin de investigar si puede suponerse que los datos siguen la distribución Normal.

Con base en las Gráficas 1.1 a 1.3, de acuerdo a la prueba grafica de Papel Normal<sup>3</sup>, puede decirse que existe evidencia para suponer que la productividad de cada uno de los sectores, considerando observaciones de 1951 a 1984, no sigue en ningún caso la distribución Normal. Ya que para validar este supuesto se requiere que al trazar una línea recta sobre los puntos de la gráfica de Papel Normal, la mayoría de estos se encuentren sobre dicha línea recta.

De esta forma, una alternativa es recurrir al análisis de varianza de Friedman<sup>4</sup>, prueba no paramétrica, puesto que al parecer la información a analizar consta de tres muestras relacionadas.

El resultado es contundente, por lo cual no se incluye el cuadro correspondiente. Considerando los rangos que anualmente alcanza la productividad de cada sector con respecto a los otros dos, se observa que en todos los años el sector cuya mano de obra es más productiva es el de servicios, el que lo sigue es aquel que conjunta a las actividades industriales, y la mano de obra ocupada en el sector primario que contempla a la agricultura, pesca, silvicultura, etc., es la menos productiva de los tres sectores.

Como era de esperarse la estadística  $\chi_r^2$ , toma el valor de:

$$\chi_r^2 \cong 70.$$

dado que la suma de los rangos es de 35, 70, y 105 respectivamente.

El valor de  $\chi_r^2$  es elevado pues existen diferencias importantes entre las sumas de rangos de los tres sectores.

<sup>3</sup>Ver SEBER, G.A.S.: *Linear Regression Analysis*, E.U.A., John Wiley & sons, 1977, p. 399.

<sup>4</sup>Ver apéndice E, sección II-1.

Dado que n es "grande" (n=34) se considera valido suponer que  $\chi_r^2$  sigue una distribución Ji-Cuadrada con  $k-1=3-1=2$  grados de libertad, por lo cual es posible saber que su probabilidad de ocurrencia es muy pequeña y menor que cualquier  $\alpha$  usual. Esto significa que con una probabilidad muy pequeña de error puede rechazarse  $H_0$  y admitirse que existen diferencias en la productividad de los tres sectores de la economía.

Este resultado resalta la presencia de altos niveles de productividad en los servicios, comparados con los que prevalecen en las actividades industriales y en las actividades primarias, siendo estas ultimas las menos productivas.

Uno de los factores que posiblemente influye en estas diferencias en cuanto a productividad, es que, talvez, la incorporación de elementos tecnológicos y financieros al desarrollo de las diferentes actividades económicas, con la finalidad de elevar la productividad, no se ha dado de manera similar en los tres sectores.

Para continuar con el presente análisis, ahora pudiera ser de interés investigar si de acuerdo al tipo de actividad en la forma en que, al paso del tiempo, se ha ido incrementando la productividad. Es decir, interesa probar la hipótesis:

$H_0$ : En México, la productividad de la mano de obra crece de manera similar en los tres sectores.

vs.

$H_1$ : Por lo menos un sector muestra un crecimiento distinto en la productividad de su mano de obra, con respecto a los dos restantes.

Los datos a emplear están contenidos en las tres ultimas columnas del Cuadro 1.6. Cabe mencionar que el incremento marcado de las cifras de 1979 a 1980, se debe a que a partir de 1980 se manejan precios de 1970, siendo que anteriormente el año base es 1960.

Enseguida se aplicará la prueba de Lilliefors<sup>9</sup> para probar si se satisface el supuesto necesario para poder emplear una prueba paramétrica. Para lo cual se obtienen los siguientes cálculos:

---

Ver apéndice B, sección II.

SECTOR	n	d <sub>n</sub>	dC. 90*
PRIMARIO	34	0.344	0.138
INDUSTRIAL	34	0.345	0.138
SERVICIOS	34	0.414	0.139

\*Ver tabla 15 en CONOVER, W.J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & sons, 1971, p. 398.

Puede advertirse que en ningún caso se satisface el supuesto de distribución Normal, ya que la diferencia entre la distribución empírica y la teórica (d<sub>n</sub>) supera en todos los casos al valor crítico (dC. 90).

Como prueba alternativa, se aplicará también el análisis de varianza de Friedman, cuya elección presenta la misma justificación que en el caso anterior.

Los rangos del Cuadro 1.7 señalan que la posición que ocupan los tres sectores en cuanto al incremento de su productividad no sigue ningún patrón en particular, es decir, a lo largo del tiempo un sector ocupa diferentes lugares en cuanto al incremento de su productividad respecto a los otros dos sectores, situación que da indicios de una posible similitud en los incrementos. Lo anterior se corrobora al conocer la suma de los rangos por sector, que es de 72 para el primario y para el secundario y de 60 para el terciario, que en términos generales son similares, de tal forma que el valor

$$\chi_r^2 \cong 2.92.$$

indica que no existen grandes diferencias entre los tres sectores.

Considerando una distribución Ji-Cuadrada con dos grados de libertad, el valor de  $\chi_r^2$  tiene una probabilidad de ocurrencia, conforme a H<sub>0</sub>, de 0.244, superior a cualquier nivel de significancia usual, lo cual da alguna evidencia de que el incremento en la productividad es similar en los tres sectores.

Como ya se indicó con anterioridad, este último resultado no significa que la productividad sea igual en todos los sectores. lo que se concluye es que el incremento o decremento porcentual de la productividad en los tres sectores de la economía parece ser semejante.

Cuadro 1.6. PRODUCTIVIDAD DE LA MANO DE OEEA POR SECTORES ECONOMICOS Y SU INCREMENTO PORCENTUAL ANUAL, 1950-1984.

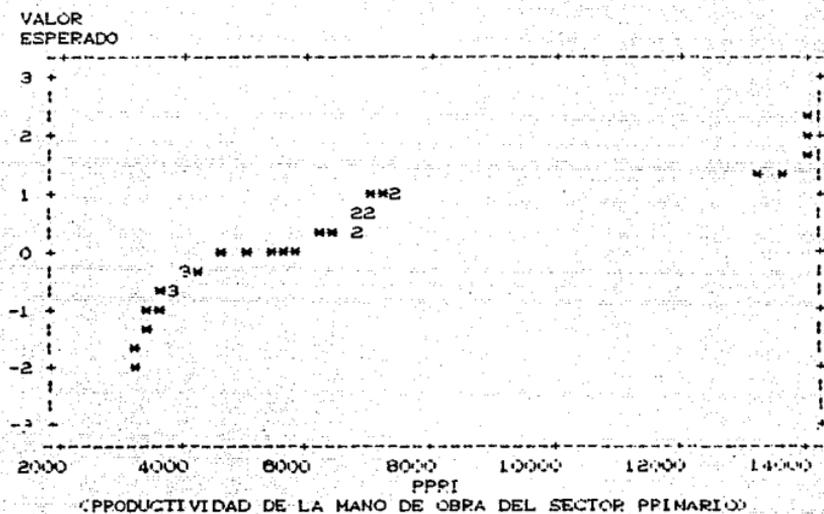
AÑO	PRODUCTIVIDAD			INC. EN LA PRODUCTIVIDAD (%)		
	SECTOR			SECTOR		
	PRIMARIO	INDUSTRIAL	SERVICIOS	PRIMARIO	INDUSTRIAL	SERVICIOS
1950	3310	16753	25501			
1951	3403	17428	26132	2.91	4.03	2.48
1952	3228	17838	26120	-5.15	1.21	-0.05
1953	3146	16544	25140	-2.55	-6.20	-3.75
1954	3593	17036	25983	14.21	2.97	3.27
1955	3828	17891	26503	6.55	5.02	2.08
1956	3652	18853	27219	-4.61	5.38	2.70
1957	3954	19307	27791	5.53	2.40	2.04
1958	4020	19199	27794	4.31	-0.56	0.05
1959	3801	19778	27027	-5.45	3.02	-2.75
1960	3901	20463	27918	2.65	3.46	3.29
1961	4048	20849	28529	3.77	1.89	2.19
1962	4280	21284	28830	5.73	2.08	1.05
1963	4588	22639	30051	7.18	6.37	4.24
1964	5025	25339	32162	9.54	11.93	7.03
1965	5397	26319	33064	7.39	3.97	2.90
1966	5592	27924	34142	3.62	6.10	3.25
1967	5853	29393	34898	4.67	5.26	2.22
1968	6146	31288	36536	5.01	6.44	4.69
1969	6329	32775	37578	2.97	4.75	2.85
1970	6766	34349	38719	6.90	4.90	3.04
1971	6827	33820	39457	0.90	-1.54	1.90
1972	6785	35501	41388	-0.62	4.97	4.89
1973	6957	37242	43507	1.05	4.90	5.12
1974	6968	38335	44898	1.62	2.93	3.19
1975	6954	38402	45924	-0.20	0.17	2.07
1976	6957	38295	45274	0.04	-0.28	-1.20
1977	7236	38522	45074	3.99	0.89	-0.44
1978	7481	40759	46565	3.40	5.81	3.31
1979	7349	43186	49055	-1.77	5.95	5.35
1980	13281	66318	89972	80.73	53.56	83.41
1981	13934	69137	94793	4.91	4.25	5.35
1982	13698	65296	92725	-1.63	-5.55	-2.18
1983	13939	57642	88545	1.76	-11.72	-4.51
1984	13960	57735	89249	0.15	0.16	0.80

Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *Estadísticas Históricas de México*, México, INEGI, 1985, t. I, p. 341.

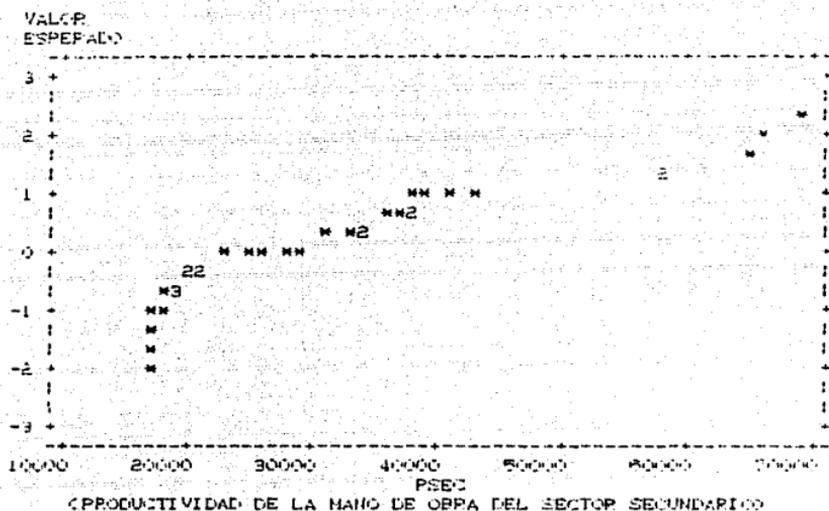
Cuadro 1.7. RANGOS ANUALES DEL INCREMENTO EN LA PRODUCTIVIDAD POR SECTOR ECONOMICO, 1951-1984.

Año	RANGOS DEL INC. EN LA PRODUCTIVIDAD		
	SECTOR		
	PRIMARIO	INDUSTRIAL	SERVICIOS
1951	2	3	1
1952	1	3	2
1953	3	1	2
1954	3	1	2
1955	3	2	1
1956	1	3	2
1957	3	2	1
1958	3	1	2
1959	1	3	2
1960	1	3	2
1961	3	1	2
1962	3	2	1
1963	3	2	1
1964	2	3	1
1965	3	2	1
1966	2	3	1
1967	2	3	1
1968	2	3	1
1969	2	3	1
1970	3	2	1
1971	2	1	3
1972	1	3	2
1973	1	2	3
1974	1	2	3
1975	1	2	3
1976	3	2	1
1977	3	2	1
1978	2	3	1
1979	1	3	2
1980	2	1	3
1981	2	1	3
1982	3	1	2
1983	3	1	2
1984	1	2	3

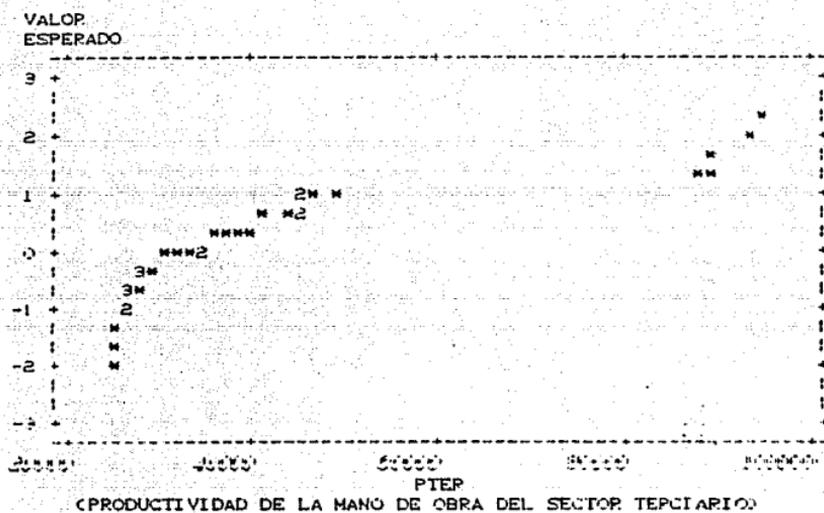
Gráfica 1.1. GRAFICA DE PAPEL NORMAL DE LA PRODUCTIVIDAD DE LA MANO DE OERA DEL SECTOR PRIMARIO (1951-1984).



GRÁFICA 1.2. GRÁFICA DE PAPEL NORMAL DE LA PRODUCTIVIDAD DE LA MANO DE OBRA DEL SECTOR SECUNDARIO (1951-1964)



Grafica 1.3. GRAFICA DE PAPEL NORMAL DE LA PRODUCTIVIDAD DE LA MANO DE OBRA DEL SECTOR TERCIARIO (1981-1984).



#### 1.1.4. Recursos naturales, industria y comercio.

Nuestro país cuenta con una enorme cantidad de recursos naturales que son extraídos, en ocasiones de forma immoderada, con el fin de sustentar la actividad económica. La variedad de esta riqueza se manifiesta con la presencia de numerosas zonas propicias para la extracción de elementos tales como recursos mineros, agrícolas, y marinos, repartidas a lo largo y ancho del territorio mexicano.

Cada estado de la República presenta características particulares en lo que respecta a sus riquezas naturales, cualitativa y cuantitativamente. Una situación similar ocurre con la actividad industrial y comercial pues cada entidad presenta su propio nivel de desarrollo, y en este sentido existen claras diferencias: Basta mencionar que sólo en el Distrito Federal, Estado de México, Jalisco, Veracruz, Puebla, Guanajuato, Michoacán y Nuevo León se asienta más del 50% de la actividad industrial y comercial nacional, según información procedente de los Censos Económicos de 1986.

Considerando la información anterior sería interesante investigar si existe relación entre la riqueza natural y el desarrollo industrial y comercial de las entidades federativas del país.

Para los fines propuestos, se tomará como fuente de información los resultados oportunos de los Censos Económicos de 1986.

Como medida de los recursos naturales con que cuenta cada entidad federativas se considerará la suma del número de

establecimientos pesqueros<sup>d</sup> y el numero de unidades mineras<sup>e</sup>, ya que la fuente consultada no presenta información sobre otros recursos como los agrícolas.

Por otro lado, para medir la actividad industrial y comercial se considera al numero de establecimientos industriales (manufactureros)<sup>f</sup> sumado al de establecimientos comerciales<sup>g</sup>.

Cabe hacer notar que la información a emplear no es la mas adecuada pues, en primer termino, la medida propuesta para la riqueza natural no es muy exacta, ya que no se incluye el sector agropecuario y ademas aun existen fuentes de recursos no explotados.

Por otro lado, no se toma en cuenta la capacidad de los centros de explotación de recursos ni de las industrias y comercios. En este sentido, podria ser mas adecuado emplear, por ejemplo, información respecto al valor monetario de los recursos extraídos, de la producción industrial y de las ventas de los comercios.

Sin embargo, al analizar la información descrita se puede obtener una aproximación de lo que se desea saber.

Dado que el objetivo de este breve analisis es el de investigar si existe relación entre la riqueza natural y el

---

<sup>d</sup> Definido como: "Unidad económica que en una sola ubicación física, asentada en un lugar permanente realiza principalmente actividades de captura y/o extracción y/o acuicultura de especies acuáticas". MEXICO, SECRETARIA DE PROGRAMACION Y PRESUPUESTO: Censos Economicos 1986 Resultados Oportunos Nacionales, México, SPP, 1987, p. 10.

<sup>e</sup> Una unidad minera es: "La unidad económica que se dedica, bajo una sola entidad propietaria, a realizar algun tipo de actividad minera pudiendo ser de explotación y/o beneficio de productos metalicos o no metalicos y cuyos centros de extracción tienen una ubicación contigua en una área geografica determinada (un cerro, un yacimiento, una cantera, etc.)". Ob. cit., p. 11.

<sup>f</sup> Es decir, "la unidad económica que en una sola ubicación física combina actividades y recursos para efectos de realizar principalmente labores de maquila, ensamble, procesamiento o transformación total o parcial de uno o varios productos". Ob. cit., p. 9.

<sup>g</sup> Esto es, "la unidad económica que en una sola ubicación física combina actividades y recursos para realizar principalmente operaciones de compra-venta de mercancías, sin transformarias, al por mayor o al por menor, ya sea por cuenta propia o ajena". Ob. cit., p. 9.

desarrollo industrial y comercial de las entidades federativas del país, parece adecuado aplicar una prueba de independencia<sup>10</sup>, donde se pretende probar la siguiente hipótesis:

H0: La actividad industrial y comercial de los estados de la República Mexicana no depende de los recursos naturales con que cuentan.

vs.

H1: Dentro de la República Mexicana, la actividad industrial y comercial y los recursos naturales se encuentran relacionados.

Para efectos de la prueba en cuestión, se propone categorizar los estados de la República según tres niveles para cada una de las características a considerar, a saber, actividad baja, media y alta, tanto para la industria y comercio como para la extracción minera y pesquera, de acuerdo al número de establecimientos de cada tipo correspondiente a cada estado. De tal forma que se obtiene la siguiente clasificación:

NO. DE ESTABLEC.	MINERIA Y PESCA INDUST. Y COMERCIO	(0, 20]	(20, 40]	MÁS DE 40	TOTAL INDUSTRIA Y COMERCIO
(0, 10]	8 (0.72)	1 (0.01)	2 (0.18)	11	
(10, 30]	3 (0.23)	6 (0.48)	1 (0.08)	10	
MÁS DE 30	2 (0.25)	4 (0.50)	2 (0.25)	8	
TOTAL MINERIA Y PESCA	13 (0.41)	11 (0.34)	5 (0.25)	29	

Fuente: Censos Económicos 1986 Resultados Oportunos Nacionales, SPP, México, 1987, p. 28-34.

<sup>10</sup> Número de establecimientos en miles.

Donde, naturalmente, las celdas de la tabla anterior contienen el número de estados que pertenece al mismo tiempo a las dos categorías señaladas, correspondientes a ambas características.

<sup>10</sup> Ver apéndice C, sección I.

Para cada característica, el criterio de categorización se determinó después de ordenar las entidades federativas según el número de establecimientos, con lo cual se formaron tres grupos, hasta donde fue posible homogéneos, procurando que en cada uno de ellos no se incluyera algún elemento que fuera demasiado diferente a los demás, en cuanto a la característica considerada.

La tabla de los valores esperados conforme a  $H_0$  es como sigue<sup>11</sup>:

$E_i \setminus j$	(0, 20)	(20, 40)	MAS DE 40	$\Sigma$
(0, 10)	4.47 (0.41)	3.78 (0.34)	2.75 (0.25)	11
(10, 30)	5.28 (0.41)	4.47 (0.34)	3.25 (0.25)	13
MAS DE 30	3.25 (0.41)	2.75 (0.34)	2.00 (0.25)	8
$\Sigma$	13 (0.41)	11 (0.34)	9 (0.25)	$\Sigma \Sigma = 32$

De los cuadros anteriores se observa lo siguiente:

Predominan los estados con actividad industrial y comercial considerada en un nivel medio, mientras que respecto a los recursos explotados sobresalen los que se ubican en el nivel más bajo, según las categorías estipuladas. También se aprecia que una cuarta parte de los estados se encuentra en la categoría más baja en los dos aspectos, con lo cual la mayoría de los estados con débil actividad industrial y comercial también cuentan con pocos recursos. Algo similar ocurre con el nivel medio en ambas categorías, en donde se encuentran 8 de las 32 entidades.

Comparando ambas tablas, se nota que las menores discrepancias corresponden a los estados con altos niveles de extracción minera y pesquera.

Elaborando los cálculos pertinentes, el estadístico  $\chi^2$ <sup>12</sup> resulta tomar el valor de:

<sup>11</sup> Ver apéndice C, sección 1.

<sup>12</sup> Ver apéndice C, fórmula C1.

$$\chi^2 = 9.039761 \cong 9.04.$$

que comparado con  $\chi^2_{(4)}(1-\alpha)$  se encuentra, para un nivel de significancia  $\alpha=0.10$  que  $\chi^2_{(4)}(0.9) = 7.78$  y se rechaza  $H_0$ , pero si  $\alpha$  disminuye, digamos  $\alpha=0.05$ , entonces  $\chi^2_{(4)}(0.95) = 9.49$  y por lo tanto, con una probabilidad de error muy pequeña, podría decirse que, aparentemente, no existe relación entre la explotación de recursos mineros y pesqueros y la actividad comercial e industrial de las entidades federativas del país.

En esta sección se ilustra que en algunas ocasiones, al intentar responder a cierto cuestionamiento, a través de la aplicación de una prueba estadística, el resultado que se obtiene no es concluyente. En este caso, al tomar  $\alpha=0.1$ , resulta que existe evidencia para considerar que el nivel de extracción minera y pesquera y la actividad industrial y comercial de las entidades federativas, se encuentran relacionados de alguna forma. Mientras que si se considera una  $\alpha=0.05$ , se concluye que, con un margen de error muy pequeño, no existe relación entre ambos aspectos.

En este caso la elección del nivel de significancia  $\alpha$  dependerá de los propósitos del análisis estadístico.

## 1.2. INFLACION Y DESEMPLEO

### 1.2.1. Inflación y zonas salariales

Uno de los principales problemas que atañe a la economía mexicana, es el correspondiente al aumento generalizado en los precios de bienes y servicios que se adquieren en el país, fenómeno conocido comúnmente como inflación, y es el Índice Nacional de Precios al Consumidor (INPC), publicado por el Banco de México, la versión oficial sobre su evolución.

Es importante mencionar que para calcular el INPC, "se recopilan cada mes 140 mil cotizaciones que abarcan alrededor de 1 200 productos y servicios, los cuales se integran en 302 conceptos genéricos que conforman la canasta del índice"<sup>1</sup>.

La cobertura del indicador comprende 35 ciudades, que están clasificadas según la zona salarial a la que pertenecen, como lo muestra el Cuadro 1.8. La zona I presenta el nivel de salario más alto y la III el más bajo.

Para establecer las categorías salariales se consideran las características económicas de cada lugar, dadas por su ubicación geográfica, recursos naturales y las principales actividades que ahí se desempeñan.

Sin lugar a dudas, la inflación es un problema que aqueja a todo el país, no obstante, es posible que se presente con diferente gravedad de acuerdo a la zona salarial. Para obtener conclusiones al respecto se efectuará una comparación entre las tasas inflacionarias de las tres zonas, tomando información anual de las ciudades de interés durante el periodo 1982-1989.

En términos generales, las cifras indican que la inflación alcanzó sus niveles más extremos en la zona III: el mínimo con casi un 16% (durante 1989 en la ciudad de Monclova, Coah.) y el máximo con aproximadamente el 144% en (Puebla, Pue., un año antes). En promedio, la inflación fue mayor en la zona I y también

<sup>1</sup> "Índices de Precios. ¿qué nos dicen?, ¿para qué nos sirven?". en: Tendencias Económicas y Financieras, Grupo Expansión, México, Ho. 10, marzo 1989, p. 11.

en ese grupo se registró la variabilidad más alta. Véase el Cuadro 1.9.

Puesto que se dispone de una cantidad considerable de datos, sería conveniente emplear una prueba paramétrica para averiguar si los niveles de inflación son similares en las tres zonas. En primera instancia, parece correcto considerar la comparación de medias para tres poblaciones, en base a la distribución Normal.

Para verificar el supuesto distribucional correspondiente se hace uso de la prueba de Lilliefors<sup>2</sup>, con la hipótesis:

H<sub>0</sub>: La tasa de inflación de la zona salarial "I" presenta una distribución Normal.

vs.

H<sub>1</sub>: La distribución de la tasa de inflación de la zona "I" no es Normal.

$$i = \overline{I, III}.$$

Efectuando las operaciones correspondientes, con un nivel de confianza del 90%, se obtienen los siguientes resultados:

ZONA	n	d <sub>n</sub>	$\alpha(0.9)^*$
I	48	0.120	0.118
II	40	0.097	0.127
III	192	0.095	0.058

\*Ver tabla 15 en CONOVER, W. J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & sons, 1971, p. 398.

De acuerdo con la forma de la región de rechazo, se determina que solo en el caso de la zona salarial II, se acepta H<sub>0</sub>, es decir que hay evidencia para suponer que la distribución de la inflación es Normal. En el caso de la zona I, el valor de d<sub>n</sub> se acerca mucho al de  $\alpha(0.9)$  y si se considerara un nivel de significancia menor, digamos del 5%, podría aceptarse H<sub>0</sub>. Estos resultados establecen la presencia de niveles inflacionarios extremos en la zona III, lo cual indica que el comportamiento de las ciudades agrupadas en

Ver apéndice B, sección II.

cada zona y a través del tiempo es heterogeneo en un indicador tan importante como lo es la inflación.

En este caso la variabilidad de los datos tiene dos fuentes: en primer término, el grado de heterogeneidad de las características económicas de las ciudades agrupadas en una misma zona, tal es el caso de la III, y en segundo término el tiempo.

Lo anterior indica que deberá recurrirse a una prueba no paramétrica y considerando que se cuenta con tres muestras independientes, cuyos datos presentan al menos una medida ordinal de la inflación, resulta conveniente emplear la prueba de Kruskal-Wallis<sup>3</sup>, para lo cual la hipótesis de interés es:

H<sub>0</sub>: Las tres zonas salariales presentan a lo largo del tiempo un nivel de inflación similar.

vs.

H<sub>1</sub>: La inflación se desarrolla de diferente manera en al menos una zona salarial.

Para calcular la estadística H<sup>4</sup> de Kruskal Wallis se considera que k=3 y N=280, y se obtiene lo siguiente:

ZONA (j)	n <sub>j</sub>	R <sub>j</sub>
I	48	7311.0
II	40	5510.0
III	192	28519.0

Donde R<sub>j</sub> es la suma de los rangos correspondiente de la zona j. Así, el valor de H es:

$$H \cong 1.23,$$

y asumiendo que H tiene una distribución  $\chi^2$  con k-1=2 grados de libertad, su probabilidad asociada es 0.540, mayor que cualquier nivel de significancia usual. Por lo tanto, con un margen de error pequeño, podría decirse que la inflación es similar en las ciudades de la República Mexicana sin importar la zona salarial a la que pertenezcan, lo cual no significa que las repercusiones de este fenómeno sean comunes a todos los lugares.

<sup>3</sup> Ver apéndice E, sección II-2.

<sup>4</sup> Ver apéndice E, fórmula E3.

Cuadro 1.9. CLASIFICACION POR ZONA SALARIAL DE LAS CIUDADES DE LA REPUBLICA MEXICANA QUE CONSTITUYEN LA COBERTURA DEL INPC.

CIUDAD	ZONA SALARIAL	CIUDAD	ZONA SALARIAL
Area Metropolitana de la Cd. de México	I	San Luis Potosí	III
Mexicali		Tapachula	
Cd. Juárez		Toluca	
Tijuana		Torreón	
Matamoros		Veracruz	
La Paz		Villahermosa	
Guadalajara		Chihuahua	
Monterrey		Monclova	
Tampico		Córdoba	
Hermosillo		Aguascalientes	
Navojca	II	Colima	
Merida		Chetumal	
Morelia		Zamora	
Acapulco		Fresnillo	
Culiacán		Iguala	
Leon		Tulancingo	
Puebla		Cortázar	
	III	Parral	

Fuente: MEXICO, COMISION NACIONAL DE LOS SALARIOS MINIMOS: Salarios mínimos vigentes a partir del primero de enero de 1969, México, 1989.

Cuadro 1.9. ESTADISTICAS GENERALES DE LAS TASAS DE INFLACION POR ZONA SALARIAL, EN EL PERIODO 1982-1989.

	ZONA I	ZONA II	ZONA III
NO. DE AÑOS	8	8	8
NO. DE CIUDADES	6	5	24
NO. DE OBS.	48	40	192
V. MIN.	17.750	17.020	15.990
V. MAX.	138.620	136.040	143.930
MEDIA	84.310	79.858	80.333
DESV. EST.	35.658	34.063	34.757

Fuente: MEXICO. BANCO DE MEXICO: *Indicadores económicos*, México, 1990.

### 1.2.2. Niveles de desocupación en dos ciudades

Al revisar los índices de desocupación<sup>5</sup> registrados durante 1997 en algunas de las ciudades más importantes del país, se encontró que las tasas más elevadas correspondían a las ciudades de Monterrey y Tampico.

También se encontró que ambas ciudades pertenecen a la zona salarial II<sup>6</sup>, de lo cual podría suponerse que existen similitudes entre ellas en el aspecto económico. De esta forma, cabría cuestionarse si presentan tasas de desocupación parecidas. Para averiguarlo se analizará la información que presenta el Cuadro 1.10.

Llevar a cabo una comparación de medias, usando la distribución Normal, no sería lo más adecuado, puesto que se cuenta con pocas observaciones en cada caso; además, como se tienen datos apareados que están representados en una escala más que ordinal, una de las pruebas no paramétricas adecuada es la de los rangos de Wilcoxon<sup>7</sup>, que se podría utilizar para verificar la hipótesis:

H<sub>0</sub>: Las ciudades de Monterrey y Tampico presentan niveles de desocupación similares.

vs.

H<sub>1</sub>: Existen diferencias entre los niveles de desocupación de las dos ciudades.

---

<sup>5</sup> Porcentaje de personas de 12 años o más que en la semana anterior a la semana en que se realizó la entrevista no tenía empleo y realizó durante los dos meses anteriores alguna actividad para encontrarlo. ZAVALA Ivan: *El año pasado 1997*, México, Plaza y Valdés, 1998, p. 61.

<sup>6</sup> Para mayor información sobre zonas salariales, ver sección 1.2.1..

<sup>7</sup> Ver apéndice E, sección I.

Para ello se obtienen lo siguientes resultados:

Mes	$d_i$	$ d_i $	$r_i$
Ene	-0.2	0.2	(-) 2
Feb	-1.6	1.6	(-) 9
Mar	1.0	1.0	(+) 8
Abr	-0.3	0.3	(-) 3
May	-0.7	0.7	(-) 4
Jun	1.5	1.5	(+) 8
Jul	-0.8	0.8	(-) 5
Ago	1.1	1.1	(+) 7
Sep	1.8	1.8	(+) 10
Oct	-0.1	0.1	(-) 1

Se observa que existen cuatro diferencias en favor de Monterrey y seis en favor de Tampico, pero los rangos indican que tan importantes son estas diferencias. Aunque existe un mayor número de ellas en favor de Tampico, las que están a favor de Monterrey son más significativas lo que arroja estadísticas  $R^+$  y  $R^-$  semejantes:

$$R^+ = 31 \text{ y } R^- = 24.$$

Si las tasas fueran similares, se esperaría encontrar algunas diferencias favoreciendo a Monterrey y algunas a Tampico, más o menos con el mismo peso; al contrario, para rechazar  $H_0$  se requeriría que  $R^+$  o  $R^-$  fuera muy pequeña. Por esto se elige:

$$R = \min(R^+, R^-).$$

y se considera que la región crítica tiene la forma:

$$R \leq r_0,$$

de tal manera que eligiendo  $R$  y buscando  $r_0$  en tablas<sup>6</sup>, para un nivel de significancia  $\alpha=0.05$ , considerando que la prueba es de dos colas, se encuentra que:

$$R = 24 \text{ y } r_0(0.05) = 9,$$

por lo tanto no existe evidencia para rechazar  $H_0$  y con una

Tabla 6 en: SIEGEL, Sidney: Diseño experimental no paramétrico aplicado a las ciencias de la conducta, trad. por Javier Aguilar Villalobos, 1a. ed., México, Editorial Trillas, 1970, p. 288.

probabilidad de error pequeña, se podría decir que los niveles de desocupación, en Monterrey y Tampico, son similares.

El resultado de la prueba muestra que existe una similitud, en términos del desenvolvimiento económico, entre las dos ciudades consideradas: ambas son incapaces de absorber la totalidad de la fuerza de trabajo con que cuentan de manera similar, pese a la importancia y variedad de las actividades industriales y de extracción que en ellas se realizan.

Cuadro 1.10. TASAS MENSUALES DE DESOCUPACION EN LAS CIUDADES DE MONTERREY Y TAMPICO, 1987 (ENERO-OCTUBRE).

MES	TASAS DE DESOCUPACION (%)	
	MONTERREY	TAMPICO
Ene	5.2	5.4
Feb	6.4	8.0
Mar	6.4	5.4
Abr	5.0	5.3
May	5.8	6.3
Jun	6.4	4.9
Jul	5.1	5.9
Ago	5.7	4.8
Sep	6.6	4.8
Oct	4.4	4.5

Fuente: ZAVALA Ivan: El año pasado 1987, México, Plaza y Valdés, 1988, p. 63.

### 1.3. FINANZAS PUBLICAS

#### 1.3.1. Gasto del Gobierno Federal

Un instrumento básico de la política de desarrollo es el gasto público, pues para lograr que se incremente la producción y el empleo se requiere del abastecimiento de insumos básicos para el crecimiento industrial, de regular el mercado nacional y de generar las obras de infraestructura y de servicio social que necesita el país<sup>1</sup>.

Los principales capítulos de gastos son los siguientes:

1000. Servicios personales	5000. Bienes Muebles e Inmuebles
2000. Materiales y suministros	6000. Obras Públicas
3000. Servicios generales	7000. Inversiones Financieras
4000. Transferencias	

El gobierno efectúa dichos gastos a través de sus dependencias, una buena parte de las cuales se encuentra relacionada ya sea con el sector primario, con los servicios sociales o bien con el sector comercio e industria como se observa en el siguiente cuadro:

<b>SECTOR PRIMARIO</b>	<b>AGRICULTURA Y RECURSOS HIDRAULICOS REFORMA AGRARIA PESCA</b>
<b>SERVICIOS SOCIALES</b>	<b>EDUCACION PUBLICA SALUBRIDAD Y ASISTENCIA TRABAJO Y PREVISION SOCIAL APORTACIONES A SEGURIDAD SOCIAL DESARROLLO URBANO Y ECOLOGIA</b>
<b>SECTOR COMERCIO E INDUSTRIA</b>	<b>COMERCIO Y FOMENTO INDUSTRIAL ENERGIA, MINAS E INDUSTRIA PARAESTATAL</b>

Dados los diferentes requerimientos de cada sector podría pensarse que el monto de lo que el Gobierno Federal gasta depende del sector al cual se destina dicha erogación. De esta manera, considerando las cifras correspondientes a la Clasificación Administrativa del Gasto Programable del Gobierno Federal para LOPEZ ROSADO, Diego G.: *Problemas económicos de México*, 6a. ed., México, UNAM, 1984, p. 241.

cada uno de los Capítulos mencionados durante 1985, se intentará probar lo siguiente:

H<sub>0</sub>: El monto de lo que el Gobierno Federal gasta es independiente del sector al cual se destina dicha erogación.

vs.

H<sub>1</sub>: El monto de lo que el Gobierno Federal gasta depende del sector al cual se destina dicha erogación.

En el Cuadro 1.11 se presentan, ya agregadas por sector o tipo de dependencia gubernamental, las cifras del gasto.

Se observa que en total, durante 1985, el Gasto Programable del Gobierno Federal ascendió a 85 688 millones de pesos, de los cuales más de la mitad correspondieron al capítulo de Transferencias y más de una tercera parte se planeó fuera destinada a servicios sociales.

También se aprecia que en los tres tipos de organismos gubernamentales, principalmente en el comercial e industrial, una importante parte de los gastos se destinaria al capítulo de Transferencias. En conclusión, podría decirse que aparentemente existe cierto grado de dependencia entre el monto del gasto y el sector.

Según el Cuadro 1.13 la estadística de prueba<sup>2</sup> toma el valor de:

$$\chi^2 = 33628.598,$$

lo cual aparentemente indicaría que H<sub>0</sub> debe ser rechazada, considerando cualquier nivel de significancia usual.

Sin embargo, uno de los factores que influye determinantemente en que el valor de  $\chi^2$  sea tan grande, es que únicamente en el sector comercial e industrial se realiza inversión financiera (capítulo 7000), siendo esta nula en los otros dos sectores. Pero debe considerarse que dicha inversión representa un porcentaje muy pequeño respecto al total.

Dado que se tiene un tamaño de muestra muy grande (85688), vale la pena calcular el coeficiente de contingencia C<sup>3</sup> que mide el grado de asociación entre las características observadas, en

2

<sup>2</sup>Ver apéndice C, fórmula C1.

<sup>3</sup>Ver apéndice E, sección II-3, fórmula E8.

este caso monto del gasto y sector al cual se destina, resultando que:

$$C = 0.2818.$$

valor que, suponiendo una distribución  $\chi^2_{(12)}$ , tiene una probabilidad de ocurrencia, conforme a  $H_0$ , superior a cualquier nivel de significancia usual, lo cual quiere decir que al parecer no existe relación entre el monto del gasto del Gobierno Federal y el sector al que este se destina.

Cuadro 1.11. GASTO PROGRAMABLE DEL GOBIERNO FEDERAL, POR OBJETO DEL GASTO Y TIPO DE DEPENDENCIA GUBERNAMENTAL, 1985.

SECTOR CAPITULO	PRIMARIO	SERVICIOS SOCIALES	COMERCIAL E INDUSTRIAL	TOTAL POR CAPITULO
C1000	3682 (0.16)	19429 (0.82)	540 (0.02)	23651
C2000	695 (0.49)	694 (0.49)	23 (0.02)	1412
C3000	883 (0.53)	676 (0.40)	111 (0.07)	1670
C4000	13785 (0.24)	13519 (0.24)	28501 (0.51)	55805
C5000	311 (0.59)	205 (0.39)	11 (0.02)	527
C6000	2132 (0.87)	215 (0.09)	107 (0.04)	2454
C7000	0 (0.00)	0 (0.00)	169 (1.00)	169
<b>TOTAL POR SECTOR</b>	<b>21488 (0.25)</b>	<b>34738 (0.41)</b>	<b>29462 (0.34)</b>	<b>85688</b>

Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *Estadísticas de Finanzas Públicas 1970-1985*, México, INEGI, 1988, p. 28-42.

Cuadro 1.12. VALORES ESPERADOS CONFORME A Ho.

$E_{ij}$	PRIMARIO	SERVICIOS SOCIALES	COMERCIAL E INDUSTRIAL	$\Sigma_j$
C1000	5930.987 (0.25)	9589.139 (0.41)	8131.894 (0.34)	23651
C2000	354.088 (0.25)	572.426 (0.41)	493.486 (0.34)	1412
C3000	418.786 (0.25)	677.020 (0.41)	574.194 (0.34)	1670
C4000	13994.233 (0.25)	22623.402 (0.41)	19187.365 (0.34)	55805
C5000	132.156 (0.25)	213.646 (0.41)	181.198 (0.34)	527
C6000	615.390 (0.25)	994.854 (0.41)	843.756 (0.34)	2454
C7000	42.380 (0.25)	68.513 (0.41)	58.107 (0.34)	169
$\Sigma_i$	21489.000 (0.25)	34738.000 (0.41)	29482.000 (0.34)	85688

Cuadro 1.13. DIFERENCIA AL CUADRADO ENTRE LO OBSERVADO Y LO ESPERADO CONFORME A Ho, ESTANDARIZADA POR LO ESPERADO.

$\frac{(E_{ij} - O_{ij})^2}{E_{ij}}$	PRIMARIO	SERVICIOS SOCIALES	COMERCIAL E INDUSTRIAL	$\Sigma$ j
C1000	852.787	10100.245	7087.753	18040.785
C2000	328.228	25.820	440.576	794.624
C3000	514.589	0.002	373.652	888.222
C4000	3.128	3663.911	4520.881	8187.921
C5000	242.028	0.350	159.888	402.242
C6000	3737.838	611.318	643.325	4992.281
C7000	42.380	68.513	211.631	322.524
$\Sigma$ i	5720.756	14470.159	13437.684	$\chi^2 = 33628.598$

## 1.4. MICROECONOMÍA

### 1.4.1. Recursos financieros en la instalación de un sistema automatizado de información

Hacia finales de 1986, las secretarías de Hacienda y Crédito Público y de Programación y Presupuesto, la Contraloría General de la Federación y el Banco de México, acordaron la creación de un sistema integral de información de los ingresos y gastos públicos, operado y controlado por un órgano auxiliar nombrado Comité Técnico de Información (CTI), cuyo objetivo fundamental sería el de: "automatizar la operación del sistema integral de información creando una red telemática que vincule a las dependencias y entidades informantes con las secretarías globalizadoras y el Banco de México"<sup>1</sup>.

Para cumplir con el objetivo anterior se requiere que tanto receptores como informantes cuenten con un hardware específico y por ello es necesario conocer las disponibilidades en este renglón con el fin de incluir dentro del presupuesto del CTI lo correspondiente al equipo faltante.

Dos de los elementos primordiales que debe poseer un organismo informante para integrarse al sistema son la microcomputadora (recepción, procesamiento y transmisión de la información) y el *modem* (necesario para llevar a cabo la teletransmisión).

El CTI estima que aproximadamente el 80% de los informantes cuenta con al menos una microcomputadora y el 30% con *modem*, lo cual significa que deberá asignarse una partida presupuestal para cubrir los requerimientos del 20% y 70% restantes, respectivamente. Para determinar si estos porcentajes son verdaderos se encuestó a 173 informantes, encontrando que 41 no poseen microcomputadora y 113 no cuentan con *modem*.

Empleando los datos obtenidos se realizarán las siguientes pruebas de Hipótesis:

---

COMITE TECNICO DE INFORMACION: Proyecto de automatización del sistema integral de información de los ingresos y gastos públicos. México, 1990, p. 5.

$$a) \quad H_0: P = 0.2 \quad \text{vs.} \quad H_1: P > 0.2$$

C) ¿Más del 20% de los usuarios no cuenta con una microcomputadora?

$$b) \quad H_0: P = 0.7 \quad \text{vs.} \quad H_1: P > 0.7$$

C) ¿Más del 70% de los usuarios no dispone de un *modem*?

En este caso se eligen pruebas de una sola cola pues es necesario evaluar si es factible que el equipo próximo a adquirir resulte insuficiente.

La elección de este tipo de hipótesis alternativa conlleva al uso de la prueba uniformemente más potente para el caso de la distribución Bernoulli<sup>2</sup> con parámetros  $P=0.2$ , en el primer caso, y  $P=0.7$ , en el segundo.

Para a), considerando que  $n=173$  y tomando  $\alpha=0.1$ , se obtiene que:

$$A' \cong \sqrt{173(0.2)(0.8)} (1.29) + 173(0.2) = 41.387,$$

y al comparar con  $\sum_{i=1}^{173} X_i=41$ , resulta que existe evidencia para considerar que el porcentaje de usuarios que carece de una microcomputadora no es mayor al 20%.

Para la segunda hipótesis, con el mismo nivel de significancia,  $A'$  toma el valor:

$$A' \cong \sqrt{173(0.7)(0.3)} (1.29) + 173(0.7) = 128.875,$$

con lo cual, recordando que  $\sum_{j=1}^{173} Y_j=113$ , también se concluye que existe evidencia para considerar que el porcentaje de usuarios que carece de un *modem* no es mayor al 70%.

Al parecer, el GTI no estaba equivocado al suponer, en la formulación de su presupuesto, que el 20% y el 70% de los futuros usuarios del sistema automatizado de información que implantará, carecen de microcomputadora y *modem*, respectivamente.

El interés por realizar la prueba anterior se debe a la importancia de los beneficios que la cristalización de este proyecto brindará al generar la simplificación y agilización en el envío de información, dando a los usuarios facilidad en su

---

<sup>2</sup> Ver apéndice A, sección II.

procesamiento y análisis, para así lograr la optimización de los recursos actualmente dedicados al control y captura, haciendo posible una vigilancia más estricta del cumplimiento y calidad de la información.

## CAPITULO 2

## CAPITULO 2. APLICACIONES EN DEMOGRAFIA

### 2.1. ESTRUCTURAS POR EDAD Y SEXO.

#### 2.1.1. Cambios en la estructura por edad.

La estructura por edad y sexo de la población así como su dinámica futura se encuentran determinadas por el comportamiento pasado de la mortalidad, de la fecundidad y de la migración. Su estudio es de gran importancia ya que una estructura dada actúa como determinante de otros procesos sociales como los relacionados con la atención escolar, la situación del empleo y la vivienda, entre otros<sup>1</sup>.

Los grupos en los cuales se divide la estructura por edad de la población se aproximan a categorías específicas de población. Por ejemplo: "Los menores de 5 años equivaldrían a la población en edades preescolares; el grupo 5-14 a la población en edad escolar; el grupo 15-44 o el 15-49, amujeres en edades fértiles, al potencial reproductivo en un momento dado; el grupo 15-64 a la población en edades de trabajo; el grupo 65 o más a la población en retiro de la actividad económica"<sup>2</sup>.

En México, hacia 1940, la población inició una etapa de rejuvenecimiento, el cual alcanzó su máximo en la década de los 70, para posteriormente dar lugar a un proceso de incipiente envejecimiento. En consecuencia, se estima que entre 1985 y el año 2010 ocurrirán una serie de cambios en la estructura por edad de la población.<sup>3</sup> Tal estimación puede ser apreciada en el Cuadro 2.1, donde además es posible establecer comparaciones con la estructura dada en 1985.

Sería interesante investigar si el tiempo tendrá el mismo efecto sobre la estructura por edad de las poblaciones de ambos

<sup>1</sup>ALBA, Francisco: *La población de México, evolución y dilemas*, 3a. ed., México, El Colegio de México, 1989, p. 59.

<sup>2</sup>Ob. cit., p. 59.

<sup>3</sup>MEXICO, CONSEJO NACIONAL DE POBLACION: *Población y desarrollo en México y el mundo*, México, CONAPO, 1988, v. I. p. 46.

sexos.

El Cuadro 2.2 muestra el cambio estimado, en porcentajes, que sufrirá la estructura por edad de la población de cada sexo de 1985 al año 2010. Se consideran cifras relativas, pues las absolutas podrían no ser muy claras para expresar los cambios en la estructura por edad, ya que son elevadas.

Para ambos sexos se advierte que, respecto al total, la población entre los 0 y los 24 años disminuirá, registrándose el máximo decremento en el grupo de 10 a 14 años. A partir del grupo 25-29 los cambios serán positivos, aunque para este grupo y para los que se abarca a partir de los 65 años se estima que dicho cambio sea menor al 1%, mostrándose un comportamiento similar en las poblaciones masculina y femenina. Esta similitud podría ser comprobada estadísticamente considerando que:

Ho: Los cambios porcentuales que de 1985 a 2010 se estima ocurrirán en la estructura por edad de las poblaciones masculina y femenina son semejantes.

vs.

Hi: Estos cambios son diferentes.

Una comparación de medias en base a la distribución Normal parece la prueba adecuada, pero primero habrá de verificarse el supuesto:

Ho: La distribución del cambio porcentual estimado, de 1985 a 2010, en la estructura por edad de la población masculina (femenina) es Normal.

vs.

Hi: La distribución no es Normal.

La prueba de bondad de ajuste que se usará en este caso es la de Kolmogorov-Smirnov<sup>4</sup>, la cual se basa en el cálculo de la máxima diferencia entre la distribución empírica y la teórica. Los resultados obtenidos son los siguientes:

---

<sup>4</sup> Ver apéndice B, sección I.

CAMBIO EN LA ESTRUCTURA POR EDAD		
	POBLACION	
	MASCULINA	FEMENINA
n	18	18
dn	0.228	0.277
dn(0.9) <sup>a</sup>	0.278	0.278

<sup>a</sup>Ver valor en tabla de: SIEGEL, Sidney: *Diseño Experimental no Paramétrico aplicado a las ciencias de la conducta*, trad. por: Javier Aguilar Villalobos, 1a. ed., México, Trillas, 1970, p. 285.

Es posible concluir que en ambos casos la diferencia entre la distribución empírica y la teórica no es lo suficientemente grande como para rechazar el supuesto de distribución Normal con una confianza del 90%, aunque en el caso de la población femenina si se fijara un nivel  $\alpha$  mayor, digamos  $\alpha=0.15$ , entonces  $dn(0.85)=0.259$  y  $H_0$  sería rechazada; lo anterior se validaría si se deseara que la probabilidad de rechazar el supuesto de distribución Normal dado que en realidad está presente fuera mayor de lo que se considera usualmente ( $\alpha=0.10$ ).

Validado el supuesto de distribución Normal, ahora puede ser aplicada la comparación de medias<sup>5</sup> con base en esta distribución. La información que se requiere para construir la estadística de prueba  $t^6$  es la siguiente:

CAMBIO EN LA ESTRUCTURA POR EDAD		
	POBLACION	
	MASCULINA	FEMENINA
m, n	18	18
$\bar{X}_i$	0.001	0.002
$\sum_j (X_j - \bar{X}_i)^2$	140.477	142.436

<sup>5</sup>Ver apéndice A sección I-1..

<sup>6</sup>Ver apéndice A, fórmula A1.

Por lo tanto:

$$t = -0.0010,$$

y a continuación se localiza el valor de  $t_{34}(0.95)$  en tablas de la distribución  $t$  de Student y conociendo la forma de  $\mathcal{L}$ , se concluye que, al parecer, no habrá grandes diferencias en los efectos que tendrá el proceso de envejecimiento de la población, en la estructura por edad de las poblaciones masculina y femenina, hacia el año 2010.

Cuadro 2.1. ESTRUCTURA POR EDAD Y SEXO DE LA POBLACION DE MEXICO, 1985 Y 2010. (PORCENTAJES)

Grupo de edad	Total nacional			
	1985		2010	
	Hombres	Mujeres	Hombres	Mujeres
0-4	13.45	13.21	8.79	8.39
5-9	13.45	13.17	8.63	8.29
10-14	13.83	13.54	8.40	8.09
15-19	12.08	11.84	8.17	7.93
20-24	9.87	9.78	8.47	8.30
25-29	8.02	8.02	8.60	8.51
30-34	6.36	6.38	8.67	8.56
35-39	5.28	5.27	8.89	8.80
40-44	4.24	4.27	7.88	7.85
45-49	3.42	3.49	6.25	6.30
50-54	2.82	2.91	5.00	5.12
55-59	2.27	2.40	3.83	3.99
60-64	1.75	1.92	3.00	3.20
65-69	1.23	1.39	2.20	2.45
70-74	0.85	0.99	1.53	1.80
75-79	0.57	0.71	0.99	1.27
80-84	0.32	0.42	0.56	0.80
85 y más	0.18	0.25	0.32	0.53
Total	100.00	100.00	100.00	100.00

Fuente: MEXICO, CONSEJO NACIONAL DE POBLACION: *Población y desarrollo de México y el mundo*, CONAPO, 1988, v. IV, p. 1007.

Cuadro 2.2. CAMBIO ESTIMADO EN LA ESTRUCTURA POR EDAD DE LA POBLACION MASCULINA Y FEMENINA ENTRE 1985 Y 2010. (PORCENTAJES).

Grupo de edad	Diferencia entre las estructuras por edad de 1985 y 2010	
	Hombres	Mujeres
0-4	-4.66	-4.82
5-9	-4.82	-4.88
10-14	-5.43	-5.45
15-19	-3.89	-3.91
20-24	-1.40	-1.48
25-29	0.58	0.49
30-34	2.31	2.18
35-39	3.61	3.53
40-44	3.44	3.39
45-49	2.83	2.81
50-54	2.18	2.21
55-59	1.56	1.59
60-64	1.25	1.28
65-69	0.97	1.06
70-74	0.69	0.81
75-79	0.42	0.56
80-84	0.24	0.38
85 y más	0.14	0.28

## 2.2. FECUNDIDAD

### 2.2.1. Uso de métodos anticonceptivos.

Uno de los factores que ha influido en gran medida sobre el descenso experimentado por la fecundidad en México desde hace algunos años es la aplicación de los programas de planificación familiar organizados por el Estado desde 1974.<sup>1</sup>

Se sabe que en México, los métodos anticonceptivos más usados son:<sup>2</sup>

Quirúrgico. Tiene por objeto impedir la fecundación del óvulo; si la operación se realiza en la mujer, la obstrucción se efectúa en las trompas de Falopio.

DIU (Dispositivo Intrauterino). Su uso no evita que el óvulo sea fecundado, pero imposibilita su implantación en la matriz y, por tanto, el desarrollo del embarazo.

Hormonal. Impide la fecundación al inhibir la ovulación de la mujer.

También se sabe que dichos métodos son de uso más generalizado en las áreas urbanas que en las rurales y que además existen grandes diferencias en cuanto a preferencias por los diversos métodos. De acuerdo con el Instituto Mexicano del Seguro Social (IMSS), las mujeres de las ciudades prefieren el método quirúrgico y el DIU, el que menos utilizan es el hormonal; las mujeres del campo utilizan este último método y recurren muy poco a la cirugía o al uso del dispositivo para regular su fecundidad.<sup>3</sup>

Para complementar la información anterior se plantea la siguiente prueba de hipótesis:

---

MEXICO. CONSEJO NACIONAL DE POBLACION: *Población y desarrollo en México y el mundo*, México, CONAPO, 1988, v. I, p. 162, 163.

<sup>2</sup>Op. cit., p. 156.

<sup>3</sup>Op. cit., p. 169.

Ho: El grado de heterogeneidad en el uso de un método anticonceptivo determinado entre las áreas rurales del país, es similar al grado de heterogeneidad que se presenta entre las áreas urbanas.

v/s.

Hi: El grado de heterogeneidad es mayor en las áreas rurales.

Se tomara en consideración la información del Cuadro 2.3 del porcentaje registrado por el IMSS en 1985 de las usuarias de cada uno de los tres métodos anticonceptivos más comunes en México, correspondientes a las áreas rural y urbana de cada entidad federativa. (Los porcentajes están calculados respecto al total de usuarias del área correspondiente)

Las estadísticas generales (Cuadro 2.4) revelan que, en promedio, el método más usado en las áreas rurales es el hormonal (casi 59 de cada 100 usuarias), mientras que en las urbanas predomina el método quirúrgico (50.71%) y el hormonal sólo es usado por 11 de cada 100 usuarias. Sobresale una mayor variabilidad dentro del país en las áreas rurales.

Es importante resaltar que en el área rural del estado de Baja California se presentan los casos más extremos: casi la totalidad de las usuarias emplea el método hormonal (93.31%), mientras que ninguna usa el quirúrgico. Por otro lado, en Campeche, también en las comunidades rurales, ocurre lo mismo respecto al método quirúrgico, pero casi hay el mismo número de usuarias en los dos métodos restantes.

Las cifras también revelan que el uso del método hormonal presenta una variabilidad menor que los otros dos métodos en el caso de las ciudades (2.75), y mayor en las áreas rurales (15.58).

El análisis consistirá en comparar la variabilidad de los porcentajes del área urbana con respecto a la rural, para cada tipo de método. En consecuencia es necesario, como primer paso, aplicar una prueba de bondad de ajuste en relación a la distribución Normal para determinar el tipo de prueba que es posible emplear.

En esta ocasión, es conveniente utilizar la prueba de

Lilliefors<sup>4</sup>. El Cuadro 2.5 muestra la máxima diferencia entre la función de distribución empírica y la teórica para cada caso, las cuales servirán para probar la hipótesis:

H<sub>0</sub>: En la República Mexicana, la distribución del porcentaje, por entidad federativa, de usuarias del método 

Quirúrgico
DIU
Hormonal

 del área 

rural
urbana

 es Normal.

vs.

H<sub>1</sub>: La distribución de este porcentaje no es Normal.

Fijando  $\alpha=0.05$ , resulta que se rechaza H<sub>0</sub> para el porcentaje de usuarias urbanas del método hormonal, al que corresponde la mayor diferencia ( $d_n = 0.184$ ).

Si se desea que la probabilidad de rechazar la hipótesis de que la distribución de los datos es Normal, dado que en verdad lo es, sea muy pequeña, es decir si se considera que  $\alpha=0.01$ , entonces, con una probabilidad de error pequeña, puede aceptarse el supuesto de distribución Normal en todos los casos, sin olvidar que la probabilidad de error Tipo II ( $\beta$ ) se ha elevado.

En este caso es deseable que  $\alpha$  sea muy pequeña, pues es importante hallar cualquier evidencia de que la distribución de los datos es Normal, con lo cual sería válido aplicar una prueba paramétrica.

De esta forma se determina que puede emplearse una prueba paramétrica para verificar la hipótesis planteada inicialmente. Si se desea comparar la variabilidad entre los datos de dos poblaciones, parece conveniente efectuar una comparación sobre las varianzas<sup>5</sup> bajo el supuesto de distribución Normal, para lo cual se consideran muestras aleatorias de:

X<sub>10</sub>: Porcentaje de usuarias rurales del método quirúrgico.

X<sub>11</sub>: " " " " " " " " DIU.

X<sub>12</sub>: " " " " " " " " Hormonal.

X<sub>20</sub>: " " " urbanas " " " " quirúrgico.

X<sub>21</sub>: " " " " " " " " DIU.

X<sub>22</sub>: " " " " " " " " Hormonal.

<sup>4</sup> Ver apéndice B, sección II..

<sup>5</sup> Ver apéndice A, sección I-3.

Para construir la estadística  $f^6$  de cada una de las tres pruebas basta con saber que:

VARIABLE <sup>6</sup>	$\sum_{i=1}^n (X_i - \bar{X})^2$
XRD	3212.088
XRD	3017.989
XRH	7035.001
XUD	2421.066
XUD	2355.294
XUH	218.471

<sup>6</sup>El tamaño de muestra en todos los casos es de 30.

Por lo tanto:

METODO	f
Quirúrgico	1.3267
DIU	1.2799
Hormonal	32.2015

Y al comparar cada uno de estos valores con:<sup>7</sup>

$$F_{(m-1, n-1)}(1-\alpha) = F_{(29, 29)}(0.9) \cong 1.63.$$

resulta ser que, de acuerdo con la forma de la región crítica para esta prueba, con una probabilidad de error pequeña, puede aceptarse  $H_0$  en el caso de los métodos quirúrgico y DIU, mientras que existe evidencia para rechazarla en el caso del método hormonal.

La estadística  $f$  se encarga de medir la relación que guardan las dispersiones existentes entre los datos de dos muestras. En esta ocasión resultó que esa relación no era lo suficientemente lejana al valor de 1, en el caso de los métodos quirúrgico y DIU,

<sup>6</sup>Ver apéndice A, fórmula A5.

<sup>7</sup>Se interpoló linealmente para hallar el valor de  $F_{(29, 29)}(0.9)$ .

como para suponer que las varianzas poblacionales de los porcentajes de usuarias son diferentes en las zonas rurales y en las urbanas. No obstante, este no es el caso del método hormonal, para el cual el elevado valor de  $f$  revela evidencias para considerar que las varianzas no son iguales, considerando cualquier nivel de significancia usual.

El resultado obtenido indica que el uso de los métodos anticonceptivos de tipo quirúrgico y el DIU presentan niveles de heterogeneidad similares en las zonas urbanas y en las rurales, de una entidad federativa a otra. Sin embargo el método hormonal, que es predominante en el campo, se usa de forma más heterogénea en este tipo de comunidades en el país.

Cuadro 2.3. PORCENTAJES DE USUARIAS DE METODOS ANTICONCEPTIVOS SEGUN METODO UTILIZADO, AREA URBANA Y AREA RURAL, POR ENTIDAD FEDERATIVA, INSTITUTO MEXICANO DEL SEGURO SOCIAL, MEXICO, 1985.

Entidad Federativa*	Area Urbana			Area Rural		
	Quirurg.	DIU	Hormon.	Quirurg.	DIU	Hormon.
Aguascalientes	39.42	45.01	15.57	23.32	29.63	47.01
Baja California	50.41	36.09	13.50	0.00	6.69	93.31
Baja California Sur	50.38	35.42	14.18	0.00	45.00	56.00
Campeche	62.19	24.40	13.41	18.30	6.83	74.87
Coahuila	53.62	35.13	11.25	39.31	23.13	37.56
Colima	56.15	35.91	7.93	11.11	14.02	74.94
Chiapas	69.57	21.63	8.80	22.42	10.68	66.88
Chihuahua	69.65	21.58	8.78	17.48	19.69	62.83
Durango	50.11	37.82	12.08	15.53	28.05	56.41
Guanajuato	36.02	53.93	10.04	9.28	23.52	67.19
Guerrero	47.21	43.94	8.85	10.13	54.59	35.28
Hidalgo	47.57	44.71	7.74	25.43	29.00	45.57
Jalisco	46.87	36.96	16.27	16.54	19.22	64.24
México	48.85	42.15	9.00	13.03	21.97	65.00
Michoacán	40.15	50.08	9.77	18.42	26.00	55.59
Morelos	56.09	36.14	7.77	39.19	27.75	33.06
Nayarit	47.26	37.49	15.25	32.33	30.31	37.34
Oaxaca	52.33	34.13	13.56	27.90	22.87	49.25
Puebla	50.52	40.60	9.87	18.98	19.31	63.74
Querétaro	48.50	42.30	9.21	23.84	34.66	41.20
Quintana Roo	53.65	40.65	9.75	4.69	15.15	79.97
San Luis Potosí	34.95	50.30	14.74	19.76	24.24	55.98
Sinaloa	52.37	39.29	9.34	16.95	19.76	53.21
Sonora	51.64	37.25	11.10	20.52	17.49	61.99
Tabasco	58.16	24.92	16.92	3.14	11.50	85.36
Tamaulipas	53.21	36.94	9.85	19.55	22.04	58.40
Tlaxcala	32.62	56.78	10.62	33.51	29.43	37.06
Veracruz	55.00	34.98	10.02	3.43	14.92	81.65
Yucatán	65.90	21.21	12.89	29.27	17.22	53.51
Zacatecas	41.00	45.31	13.69	14.03	28.05	57.90

Fuente: MEXICO, CONSEJO NACIONAL DE POBLACION: *Población y desarrollo de México y el mundo*, México, CONAPO, 1988, v. IV, p. 1043.

\* No se incluye al Distrito Federal y al estado de Nuevo León, por no presentar usuarias rurales dentro del programa del IMSS.

Cuadro 2.4. ESTADISTICAS GENERALES DE LOS PORCENTAJES DE USUARIAS DE METODOS ANTICONCEPTIVOS, SEGUN METODO UTILIZADO, AREA URBANA Y AREA RURAL.

Estadística	Area Urbana			Area Rural		
	Quirurgico	DIU	Hormonal	Quirugico	DIU	Hormonal
VALOR MINIMO	32.62	21.21	7.74	0.00	6.69	33.06
VALOR MAXIMO	69.65	56.78	16.92	39.31	51.59	93.31
MEDIA	50.71	38.07	11.36	18.19	23.10	58.74
DESV. EST.	9.14	9.03	2.75	10.52	10.20	15.58

Cuadro 2.5. CALCULOS CORRESPONDIENTES A LA PRUEBA DE LILLIEFORS SOBRE LOS PORCENTAJES DE USUARIAS DE METODOS ANTICONCEPTIVOS, SEGUN METODO UTILIZADO, AREA URBANA Y AREA RURAL.

METODO	dn		d(1- $\alpha$ ) <sup>*</sup>	
	Area Rural	Area Urbana	$\alpha = 0.5$	$\alpha = 0.01$
QUIRURGICO	0.078	0.109		
DIU	0.140	0.133	0.144	0.187
HORMONAL	0.094	0.184		

\* Ver tabla 15 en CONOVER, W.J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & sons, 1971, p. 398.

## 2.3. MORTALIDAD

### 2.3.1. Principales causas de mortalidad

A nivel nacional, las causas que producen el mayor número de defunciones, de acuerdo con la clasificación internacional de enfermedades, son en el siguiente orden:

- VII. Enfermedades del aparato circulatorio.
- XVII. Causas extremas de traumatismo y envenenamiento.
  - I. Enfermedades infecciosas y parasitarias.
- VIII. Enfermedades del aparato respiratorio.
  - II. Tumores.

Cada estado de la República Mexicana presenta sus propias características en cuanto a la mortalidad por causas en 1986. (Ver Cuadro 2.6)

Dadas las diferencias socioeconómicas y geográficas que prevalecen dentro del país, es factible suponer que el porcentaje de víctimas que cobran las principales causas de muerte varía para cada una de ellas de un estado a otro; pero, ¿esta variación a lo largo del país será similar para todas las causas mencionadas?

El Cuadro 2.7 indica que, durante el año de 1986, mientras que la proporción de muertes debida a enfermedades respiratorias se comportó de la forma más homogénea (con una desviación estándar de 2.44), la que se debió a enfermedades infecciosas y parasitarias presentó la máxima variabilidad (5.16), dado que por ejemplo, habiendo causado más de la cuarta parte de las defunciones totales en el estado de Chiapas, produjo 5 de cada 100 de las que sucedieron en Durango.

Así se plantea la siguiente prueba de hipótesis:

H<sub>0</sub>: Dentro de la República Mexicana, el porcentaje de defunciones producidas por las cinco principales causas presentan grados de variación similares de uno a otro estado

vs.

H<sub>1</sub>: Por lo menos una causa de muerte presenta una variabilidad diferente a las restantes.

En primer término se verificará si se cumple el supuesto de distribución Normal, es decir,

Ho: La distribución del porcentaje de defunciones por entidad federativa debidas a la causa "x" es Normal.

vs.

Hi: La distribución del porcentaje de defunciones por entidad federativa debidas a la causa "x" no es Normal.

En las gráficas de Papel Normal<sup>1</sup> sobre los porcentajes por cada causa de muerte (Gráficas 2.1 a 2.5) se observa que podría trazarse una línea recta que pasara muy cerca de la mayoría de los puntos en lo que corresponde a las enfermedades del aparato circulatorio y a los tumores.

Si se consideran las enfermedades infecciosas y parasitarias (I), las enfermedades del aparato respiratorio (VIII) y las causas extremas de traumatismo y envenenamiento (XVII), habría menos indicios de que la distribución de los datos en Normal.

Sin embargo, aunque existe la posibilidad de que no sea así, se considerará que se satisface el supuesto de distribución Normal en los cinco casos, pues dado que es más potente, es preferible aplicar una prueba paramétrica que una no paramétrica.

De esta forma se puede proceder a efectuar la comparación de varianzas<sup>2</sup>, para lo cual se calcula la estadística de prueba correspondiente<sup>3</sup>, empleando los siguientes datos:

CAUSA DE MORTALIDAD *	$\sigma_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2$
VII	25.7602
XVII	7.8613
I	13.6693
VIII	8.7636
II	20.8289

<sup>1</sup> SEBER, G.A.F.: *Linear Regression Analysis*, E.U.A., John Wiley & Sons, 1977, p. 395.

<sup>2</sup> Ver apéndice A, sección I-4.

<sup>3</sup> Ver apéndice A, fórmula A7.

\* Para todos los casos:  $n_j = 32$ .

Por lo tanto,

$$-2 \log \lambda = 10.4426,$$

que resulta ser superior a

$$\chi_4^2(0.99) = 7.78,$$

pero si se toma  $\alpha = 0.01$ , entonces se tiene que:

$$\chi_4^2(0.999) = 13.3.$$

De esta forma, antes de concluir sobre la prueba sería necesario fijar la probabilidad de error tipo I ( $\alpha$ ) que se está dispuesto a aceptar, lo cual generalmente está ligado al uso que se dará a las conclusiones que se obtengan.

Cuadro 2.6 . PORCENTAJE DE DEFUNCIONES, CAPITULOS DE CAUSAS DE MORTALIDAD, SEGUN LA CLASIFICACION INTERNACIONAL DE ENFERMEDADES, POR ENTIDAD FEDERATIVA, 1986.

Entidad Federativa	Porcentaje de defunciones por causa				
	I	II	VII	VIII	XVII
Aguascalientes	9.84	10.27	18.23	10.76	14.59
Baja California	5.85	11.40	21.76	8.19	18.55
Baja California Sur	5.51	15.22	20.35	8.78	14.38
Campeche	11.03	10.07	17.49	6.82	16.53
Coahuila	7.86	12.93	22.44	10.85	12.65
Colima	11.48	11.25	19.76	7.61	15.77
Chiapas	26.50	6.71	12.11	7.95	12.61
Chihuahua	7.00	12.22	23.16	9.56	16.66
Distrito Federal	5.46	12.83	25.05	12.17	13.91
Durango	5.03	10.56	24.62	8.86	19.29
Guanajuato	13.34	7.18	16.22	14.70	14.35
Guerrero	17.35	7.06	14.84	6.82	23.74
Hidalgo	12.09	6.43	14.91	11.61	13.31
Jalisco	6.88	11.20	22.18	10.69	14.73
México	11.14	6.51	13.66	14.70	18.97
Michoacán	9.26	9.56	19.66	9.13	19.40
Morelos	9.61	9.18	19.82	8.34	19.70
Nayarit	7.10	12.92	21.20	6.25	21.94
Nuevo León	7.33	15.61	24.31	8.92	9.53
Oaxaca	25.89	5.16	12.37	8.34	13.96
Puebla	13.46	6.12	14.19	12.54	12.58
Querétaro	14.40	6.76	15.79	12.45	15.43
Quintana Roo	10.03	8.07	13.30	6.87	29.55
San Luis Potosí	11.40	8.89	18.75	10.60	16.52
Sinaloa	6.94	12.65	20.69	8.28	27.87
Sonora	6.81	12.12	18.59	7.59	15.92
Tabasco	15.94	8.21	17.09	9.23	18.94
Tamaulipas	5.79	13.76	24.57	7.73	18.77
Tlaxcala	12.41	5.79	16.08	15.69	13.40
Veracruz	11.49	9.13	18.20	7.90	14.49
Yucatán	10.25	11.77	21.77	6.38	7.19
Zacatecas	7.81	9.80	20.46	10.69	15.17

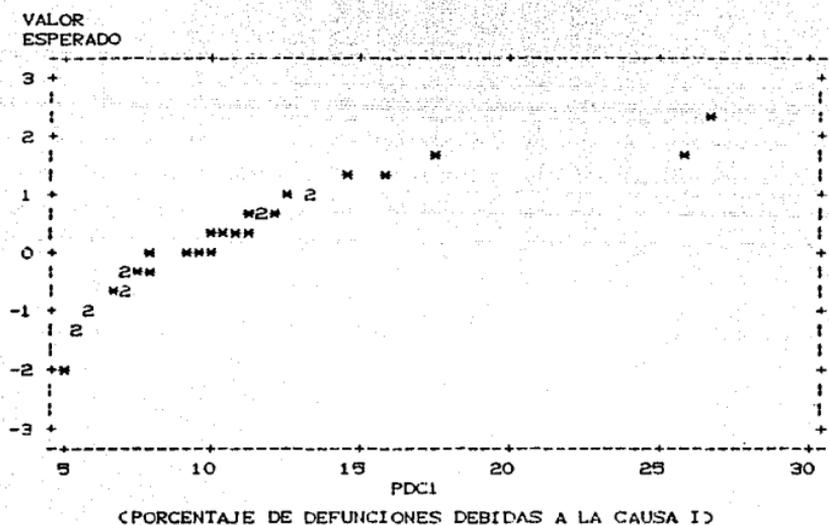
Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA; Anuario Estadístico de los Estados Unidos Mexicanos 1988-1989, México, 1990, INEGI, p. 338-357.

Cuadro 2.7 . ESTADÍSTICAS GENERALES CORRESPONDIENTES A LOS PORCENTAJES DE DEFUNCIONES POR CAUSA.

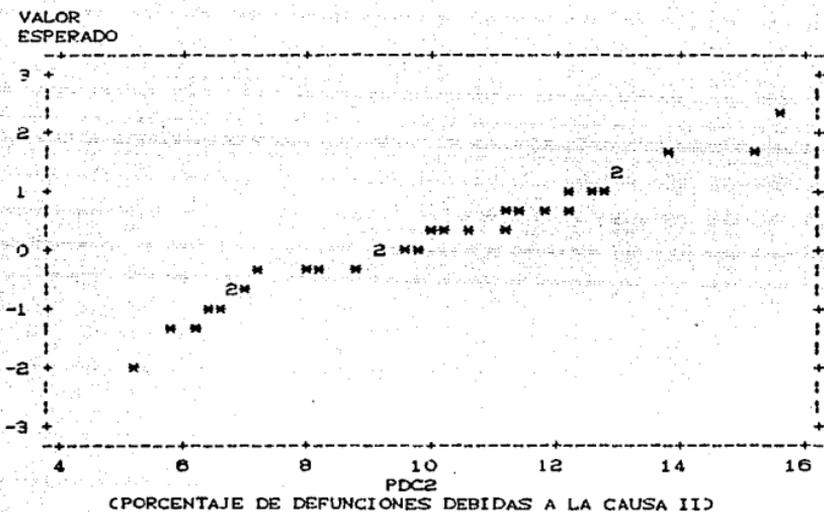
Estadísticas*	Porcentaje de defunciones por causa				
	I	II	VII	VIII	XVII
V. MIN.	5.03	5.16	12.11	6.25	7.19
V. MAX.	20.50	15.81	25.05	15.69	29.55
MEDIA	10.70	9.92	18.83	9.66	16.58
DESV. EST.	5.16	2.85	3.76	2.44	4.64

\*En todos los casos n=32.

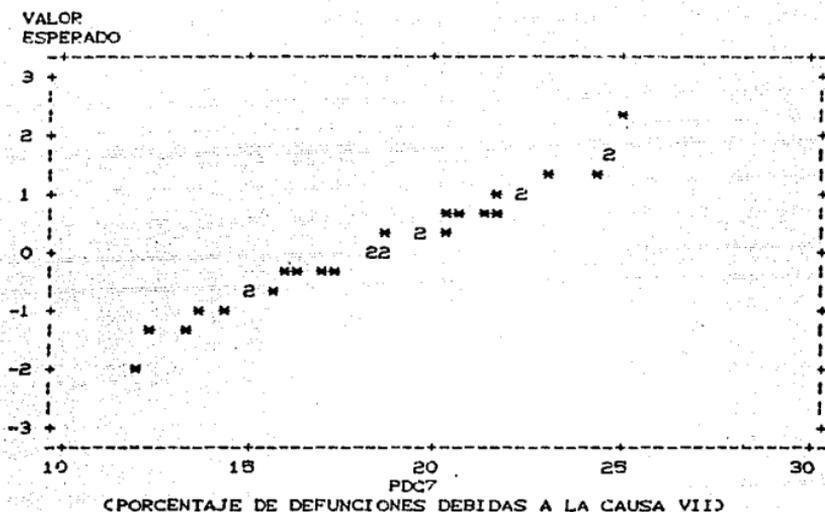
Gráfica 2.1 . GRAFICA DE PAPEL NORMAL DEL PORCENTAJE DE DEFUNCIONES POR ENTIDAD FEDERATIVA DEBIDAS A LA CAUSA DE MORTALIDAD I.



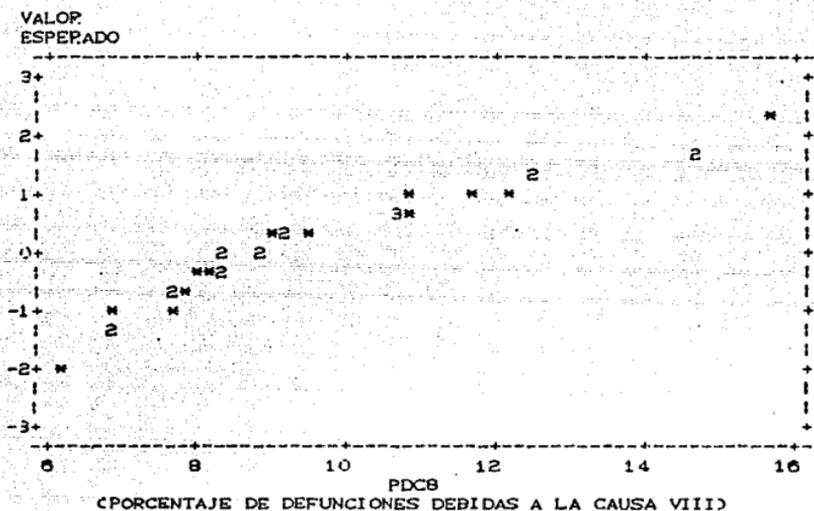
Gráfica 2.2 . GRAFICA DE FAPEL NORMAL DEL PORCENTAJE DE DEFUNCIONES POR ENTIDAD FEDERATIVA DEBIDAS A LA CAUSA DE MORTALIDAD II.



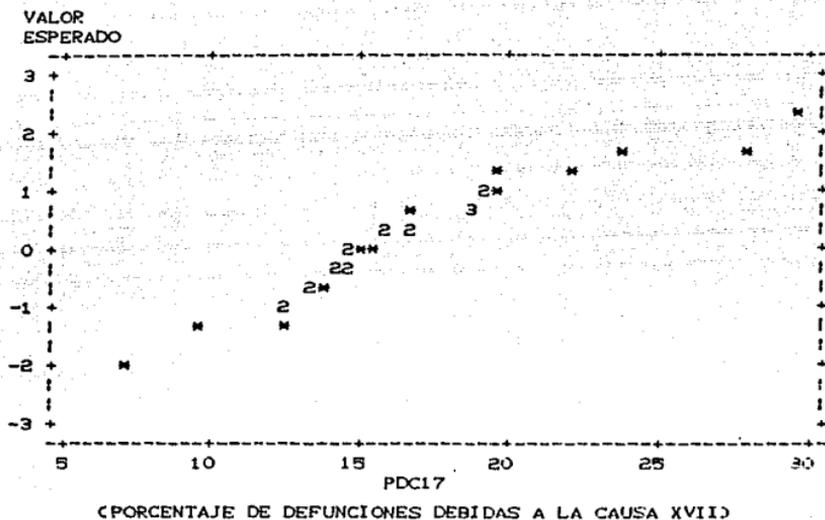
Gráfica 2.3 . GRAFICA DE PAPEL NORMAL DEL PORCENTAJE DE DEFUNCIONES POR ENTIDAD FEDERATIVA DEBIDAS A LA CAUSA DE MORTALIDAD VII.



Gráfica 2.4 . GRAFICA DE PAPEL NORMAL DEL PORCENTAJE DE DEFUNCIONES POR ENTIDAD FEDERATIVA DEBIDAS A LA CAUSA DE MORTALIDAD VIII.



Gráfica 2.5 . GRAFICA DE PAPEL NORMAL DEL PORCENTAJE DE DEFUNCIONES POR ENTIDAD FEDERATIVA DEBIDAS A LA CAUSA DE MORTALIDAD XVII.



### 2.3.2. Mortalidad infantil

Durante los últimos años la mortalidad infantil<sup>1</sup> ha descendido notablemente en comparación con la mortalidad general. Esto se debe principalmente a la implantación de programas de vacunación, rehidratación oral y salud materno-infantil.<sup>2</sup>

La población con menos de un año es la más sensible a los factores del medio externo, tales como dotación de alimentos, condiciones de higiene y disposición de servicios médicos. Dichos factores presentan diferencias de una comunidad urbana a una rural. Es por ello que resultaría interesante realizar la siguiente prueba de hipótesis:

H<sub>0</sub>: El decremento porcentual anual registrado de 1985 a 1988 en el número de defunciones menores de un año es similar en los medios rural y urbano de las entidades federativas.

vs.

H<sub>1</sub>: En el medio rural el decremento porcentual es menos significativo que en el urbano.

La información que se empleará para probar la hipótesis está contenida en el Cuadro 2.8, donde los porcentajes positivos corresponden a decrementos y los negativos a incrementos.

Se observa que con excepción de la zona rural correspondiente a los estados de Campeche, Coahuila, Colima y Chiapas, en el resto de las zonas rurales del país, de 1985 a 1988 disminuyó el número de muertes con menos de un año, considerando cifras porcentuales. Como se puede notar, las cuatro únicas entidades federativas en las cuales se incrementó la mortalidad infantil, se encuentran en orden alfabético continuo. Esto podría ser coincidental, pero existe la posibilidad de que se trate de un error en los datos.

También se aprecia que en el área urbana de los cuatro estados mencionados, junto con los de Baja California Sur, Chihuahua y Michoacán, hubo incrementos, siendo el más importante superior al 40%, ocurrido en esta última entidad. En contraste, a

---

<sup>1</sup> Número de defunciones de menores de un año por mil nacimientos ocurridos durante un mismo lapso, generalmente un año.

<sup>2</sup> CONSEJO NACIONAL DE POBLACION: *Población y desarrollo en México y el mundo*, México, CONAPO, 1988, v. I, p. 78.

las ciudades de Nayarit correspondió el máximo decremento porcentual registrado (23.16%).

Por otro lado, en la parte rural del Distrito Federal el descenso fue del 100% aunque debe considerarse que según los datos originales durante 1985 el número de muertes con menos de un año fue de 3. Este caso, hace pensar que posiblemente resultaría más adecuado analizar datos absolutos y no porcentuales.

Asimismo, en las comunidades rurales de Coahuila se observó el máximo incremento (16.54%) de la República Mexicana en este medio.

El tipo de prueba que se utilizará será determinado después de verificar la hipótesis:

$H_0$ : La distribución del decremento porcentual registrado en el número de muertes con menos de un año de 1985 a 1986 en las zonas { rurales } del país es Normal.

vs.

$H_1$ : La distribución no es Normal.

Aplicando la prueba de Lilliefors<sup>8</sup> se obtiene la siguiente información:

DECREMENTO PORCENTUAL	n,m	ds <sub>2</sub>	ds <sub>2</sub> (0.95) <sup>*</sup>	ds <sub>2</sub> (0.99) <sup>*</sup>
RURAL	32	0.181	0.157	0.182
URBANO	32	0.140		

<sup>\*</sup>Ver tabla 15 en CONOVER, W.J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & sons, 1971, p.393.

A partir de la cual se concluye que, dado que es deseable hallar cualquier evidencia de que la distribución de los datos es Normal, para justificar el uso de una prueba paramétrica, considerando  $\alpha=0.01$ , puede asumirse que la distribución del decremento porcentual registrado en el número de muertes con menos de un año de 1985 a 1986 en las zonas rurales y urbanas del país es Normal.

De esta forma, se efectuará una comparación de medias con

<sup>8</sup> Ver apéndice B, sección II.

base en la distribución Normal<sup>4</sup>.

Para calcular  $t^5$ , se toma en cuenta que:

DECREMENTO PORCENTUAL	m, n	$\bar{X}$	$\sum_{i=1}^{m, n} (X_i - \bar{X})^2$
RURAL	32	12.494	11562.932
URBANO	32	4.494	4792.116

Por lo tanto:

$$t = 1.9835,$$

y de acuerdo con la forma de la región crítica para la prueba de una cola, se determina que existe evidencia para rechazar  $H_0$  con un nivel de significancia  $\alpha=0.1$ , dado que

$$t_{(62)}(0.90) \cong 1.296.$$

Sin embargo, si se fijara una  $\alpha$  más pequeña, por ejemplo de 0.01, entonces,

$$t_{(62)}(0.99) \cong 2.389$$

y podría concluirse que al parecer existe similitud entre los porcentajes de las áreas rural y urbana.

Como se ha visto, el resultado de esta prueba depende de la elección del nivel de significancia.

Esta situación se podría deber a que, como ya se mencionó al principio, probablemente existen errores en la información manejada. Lo cual conduciría a recomendar que antes de realizar cualquier análisis estadístico, debe evaluarse la calidad de la información a emplear.

<sup>4</sup> Ver apéndice A, sección I-1.

<sup>5</sup> Ver apéndice A, fórmula A1.

Cuadro 2.8 . DECREMENTO PORCENTUAL REGISTRADO DE 1985 A 1986 EN EL NUMERO DE DEFUNCIONES CON MENOS DE UN AÑO, MEDIO URBANO Y MEDIO RURAL POR ENTIDAD FEDERATIVA.

Entidad Federativa	Decremento porcentual	
	Medio Urbano	Medio Rural
Aguascalientes	0.00	20.59
Baja California	0.87	3.38
Baja California Sur	-26.27	5.88
Campeche	-2.14	-6.58
Coahuila	-7.24	-18.54
Colima	-0.71	-1.64
Chiapas	-4.32	-7.30
Chihuahua	-2.22	3.29
Distrito Federal	20.12	100.00
Durango	14.98	17.24
Guanajuato	6.57	10.76
Guerrero	2.03	3.04
Hidalgo	7.08	12.47
Jalisco	7.77	7.78
México	15.91	19.75
Michoacán	-40.57	6.93
Morelos	6.78	0.00
Nayarit	23.16	23.16
Nuevo León	4.17	26.62
Oaxaca	11.88	9.69
Puebla	12.77	11.32
Querétaro	3.37	4.27
Quintana Roo	12.14	30.77
San Luis Potosí	8.31	6.23
Sinaloa	8.13	11.36
Sonora	11.45	4.72
Tabasco	1.13	19.34
Tamaulipas	11.60	32.14
Tlaxcala	12.80	15.07
Veracruz	1.44	0.49
Yucatán	3.07	18.68
Zacatecas	18.05	8.94

Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: Anuario Estadístico de Los Estados Unidos Mexicanos 1988-1989, México, 1990, INEGI, p. 326-328.

## 2.4. DISTRIBUCIÓN TERRITORIAL DE LA POBLACION

### 2.4.1. Crecimiento de la población urbana

La alta concentración poblacional sobre algunas zonas es resultado del proceso de urbanización que ha transformado a México a través de los años.

Debido al movimiento revolucionario que sacudió al país de 1910 a 1917, sólo se registraron crecimientos positivos en la población urbana,<sup>1</sup> pues, en búsqueda de condiciones más seguras, se presentó la salida de volúmenes importantes de población rural rumbo a las ciudades, volúmenes que, en gran parte, retornaron a su lugar de origen al concluir el movimiento.<sup>2</sup>

Posteriormente, de 1930 a 1940, el impulso que dio Lázaro Cárdenas a la Reforma Agraria, ocasionó que por única vez en el siglo XX, el crecimiento poblacional en las ciudades se contrajera. Sin embargo, esta tendencia se invirtió completamente a partir de 1940, cuando el modelo de desarrollo adoptado favoreció la industrialización, el mejoramiento de las vías de comunicación y transporte, así como el acelerado crecimiento natural de la población, llegando al grado de que, entre 1980 y 1985, tres cuartas partes de los incrementos totales de población ocurrieron en el Área urbana.<sup>3</sup>

A continuación se analizará brevemente la forma en que ha evolucionado cuantitativamente la población urbana en las

---

<sup>1</sup> MEXICO, CONSEJO NACIONAL DE POBLACION: *Población y desarrollo en México y el mundo*, México, CONAPO, 1988, v. I, p.210.  
Se considera como población urbana a aquella que se ubica en localidades de 15 000 y más habitantes.

<sup>2</sup> *Op. cit.*, p. 212.

<sup>3</sup> *Op. cit.*, p. 218.

distintas regiones geoeconómicas<sup>4</sup> del país, para saber si tal evolución es similar entre ellas.

El Cuadro 2.9 presenta las tasas promedio de crecimiento de la población urbana por entidad federativa y agrupadas por región geoeconómica, por décadas de 1910 a 1985. Este indicador se refiere a la diferencia entre nacimientos y muertes sumada al saldo migratorio, relativa a la población total.

En lo que corresponde a la región Noroeste, se observa que únicamente la población urbana establecida en Sinaloa registró incrementos desde la primera década considerada, mientras que la de Baja California Sur empezó a crecer solamente a partir de 1950. En Sonora se presentan los valores extremos de toda la región: Un decremento del 0.51% en el periodo 1930-1940 y un incremento del 20.86% durante el siguiente decenio.

En ninguna entidad de la zona Norte y en ningún periodo, se mantuvo sin cambio la población citadina. En San Luis Potosí y en Zacatecas, durante los primeros años se presentan tasas negativas, siendo la correspondiente a Zacatecas (-4.86%), la menor que se observó en esta zona, para que durante la siguiente década (1920-1930) alcanzara el máximo crecimiento, también de toda la zona (7.83%).

En el Noreste se han registrado únicamente tasas positivas durante todo el periodo considerado, destacando un crecimiento anual del 13.71% durante los primeros 10 años en el estado de Tamaulipas.

La población de las ciudades del Occidente de México se ha mantenido en crecimiento, principalmente la que radica en

---

<sup>4</sup> Op. cit. V. IV, p. 1055.

Las regiones geoeconómicas son:

Noroeste: Baja California, Baja California Sur, Nayarit, Sinaloa y Sonora.

Norte: Coahuila, Chihuahua, Durango, San Luis Potosí y Zacatecas.

Noreste: Nuevo León y Tamaulipas.

Occidente: Aguascalientes, Colima, Guanajuato, Jalisco y Michoacán.

Centro: Hidalgo, Morelos, Puebla, Querétaro y Tlaxcala.

Golfo: Tabasco y Veracruz.

Metropolitana: México y Distrito Federal.

Sur: Chiapas, Guerrero y Oaxaca.

Península de Yucatán: Campeche, Quintana Roo y Yucatán.

Aguascalientes y Jalisco, pues en las restantes existe algun decremento durante los dos primeros periodos tomados en cuenta.

La zona que en promedio ha presentado el menor incremento es la del Centro: casi en todos los estados que la componen, la poblacion urbana se mantuvo constante o decreció durante los primeros periodos; incluso en el estado de Tlaxcala se registró un crecimiento positivo únicamente a partir del periodo 1960-1970. En contraste, desde 1920, en Puebla se ha dado un incremento sostenido.

Mientras que la poblacion de las ciudades de Tabasco empieza a crecer a partir del tercer decenio revisado, la que corresponde a Veracruz crece constantemente desde 1910, dentro de la región del Golfo.

No sorprende el hecho de que en la zona Metropolitana se haya registrado el máximo incremento promedio de todo el país, lo cual ocurrió en el Estado de México durante la década de los 50.

Las 3 entidades que conforman la región Sur registraron crecimientos nulos o negativos durante los decenios iniciales, siendo que el más notable (de -2.7%) lo sufrió la población urbana del estado de Oaxaca de 1910 a 1920. Por otro lado, tal vez debido al factor turístico, en Guerrero en la década de los 60 se dio el incremento más importante de toda la zona.

Quintana Roo es el estado cuya población urbana se mantuvo cuantitativamente estable por más tiempo, presentando incrementos significativos solamente a partir de 1970, posiblemente por motivos turísticos también; mientras que en las dos entidades restantes de la Península de Yucatán el incremento se ha sostenido.

La hipótesis que se pretende probar es:

H<sub>0</sub>: El crecimiento anual promedio, por entidad federativa que ha experimentado la población urbana asentada en las diferentes regiones geoeconómicas del país es similar.

vs.

H<sub>1</sub>: Por lo menos en una región geoeconómica, el crecimiento de la población urbana ha sido distinto al registrado en las demás.

El tipo de prueba adecuado podrá ser determinado después de

aplicar una de Bondad de ajuste para saber si existe evidencia para considerar que los datos siguen una distribución Normal, es decir, para  $i=1, \overline{9}$ , se tiene que:

$H_0$ : La tasa anual promedio de crecimiento, por entidad federativa que ha experimentado la población urbana que radica en la región geoeconómica "i" presenta una distribución Normal.

vs.

$H_1$ : Esta distribución no es Normal.

Donde:

- 1 = Noroeste
- 2 = Norte
- 3 = Noreste
- 4 = Occidente
- 5 = Centro
- 6 = Golfo
- 7 = Metropolitana
- 8 = Sur
- 9 = Península de Yucatán

En el Cuadro 2.10 se observa la máxima diferencia por región entre la distribución empírica de los datos originales, pero estandarizados, y la distribución teórica, de acuerdo al procedimiento indicado por la prueba de Lilliefors<sup>1</sup>. Considerando  $\alpha=0.01$ , con la intención de hallar cualquier evidencia de que la distribución es Normal, puede concluirse que existe evidencia para aceptar tal supuesto en los nueve casos, no sin antes comentar que, de acuerdo a la magnitud de la diferencia entre la distribución empírica y la teórica, en las regiones Noroeste, Metropolitana y Sur la población urbana tuvo el comportamiento más extremo.

De acuerdo a las condiciones que debe reunir un grupo de datos para ser analizados mediante una comparación de medias con base en la distribución Normal<sup>2</sup>, en este caso es factible emplearla.

Así, la estadística de prueba  $r^3$ , cuya magnitud indica si las muestras consideradas provienen de poblaciones semejantes, se construye tomando en cuenta los resultados contenidos en el Cuadro

<sup>1</sup>Ver apéndice B, sección II.

<sup>2</sup>Ver apéndice A, sección I-2.

<sup>3</sup>Ver apéndice A, fórmula A2.

2.11, donde se observa que las regiones que en promedio registraron los mayores incrementos fueron la Noreste y la Metropolitana (donde se encuentran dos de las ciudades más importantes del país: Monterrey y la Ciudad de México) con valores superiores al 5%. Por otro lado, la región del Centro presentó el menor crecimiento promedio, inferior al 3%. De esta forma,  $r$  toma el valor de:

$$r = \frac{93,673/(9-1)}{3242,278/(256-9)} = 0.892,$$

que al ser comparado con

$$F_{(p-1, 256-p)}(0.9) = F_{(8, 248)}(0.9) = 1.67,$$

resulta ser menor, por lo cual con un nivel de significancia  $\alpha=0.1$ , se concluye que existe evidencia para considerar que el crecimiento de la población urbana se ha dado de manera semejante en todas las regiones geoeconómicas de la República Mexicana.

La conclusión derivada de este breve análisis no significa que el tamaño de las poblaciones establecidas en las ciudades de todas las regiones del país sea igual, lo que se obtuvo es que su dinámica es semejante de una región a otra, considerando además que los factores que impulsan este crecimiento son particulares a cada región.

Con la conclusión anterior parecería contradecirse el conocimiento de que el crecimiento urbano ha sido sumamente concentrador. Sin embargo, no es así por dos razones: en primer término, se está comparando el crecimiento de la población urbana por región geoeconómica y no por ciudad; por otro lado, parecería ser que los bajos incrementos ocurridos hasta 1940, equilibran el crecimiento global de las nueve regiones.

Cuadro 2.9 . TASA DE CRECIMIENTO MEDIO ANUAL DE LA POBLACION URBANA, SEGUN REGIONES GEOECONOMICAS Y ENTIDADES FEDERATIVAS DE MEXICO, 1910-1985. (POR CIEND. HOJA 1)

PERIODO	REGION					
	Noroeste	Norte	Noreste	Occidente	Centro	
1910-1920	BC	Coah	HL	Ags	Hgo	0.00
1920-1930						0.00
1930-1940						0.00
1940-1950						14.07
1950-1960						9.48
1960-1970						6.22
1970-1980						2.89
1980-1985	1.61	4.15	0.96	0.65	0.50	
1910-1920	BCS	Chlh	Tamps	Col	Mor	0.00
1920-1930						0.00
1930-1940						0.00
1940-1950						0.00
1950-1960						0.00
1960-1970						6.72
1970-1980						9.03
1980-1985	6.56	5.31	13.71	1.13	0.00	
1910-1920	Nay	Lgo	Sto	Pue	Pue	0.00
1920-1930						0.00
1930-1940						1.82
1940-1950						3.29
1950-1960						7.70
1960-1970						7.21
1970-1980						6.84
1980-1985	1.99	0.17	6.55	-2.69	0.00	
1910-1920	Sin	SLP	Jal	Gro	Gro	0.00
1920-1930						2.06
1930-1940						0.17
1940-1950						-0.17
1950-1960						7.02
1960-1970						2.92
1970-1980						4.79
1980-1985	2.78	0.90	4.88	0.91	0.00	
1910-1920	Son	Zac	Mich	Hax	Hax	0.00
1920-1930						6.19
1930-1940						2.08
1940-1950						1.18
1950-1960						7.30
1960-1970						6.61
1970-1980						6.77
1980-1985	5.99	4.85	3.93	11.24	1.77	
1910-1920	Con	Zac	Hax	Hax	Hax	0.00
1920-1930						0.00
1930-1940						-0.51
1940-1950						0.66
1950-1960						8.27
1960-1970						5.67
1970-1980						3.94
1980-1985	2.77	4.72	3.28	3.45	11.70	
1910-1920	Con	Zac	Hax	Hax	Hax	0.00
1920-1930						0.00
1930-1940						0.00
1940-1950						0.00
1950-1960						6.72
1960-1970						9.03
1970-1980						6.56
1980-1985	2.77	1.35	1.54	4.08	3.76	
1910-1920	Con	Zac	Hax	Hax	Hax	0.00
1920-1930						0.00
1930-1940						0.00
1940-1950						0.00
1950-1960						6.19
1960-1970						4.40
1970-1980						2.45
1980-1985	3.05	3.05	3.05	4.16	4.59	
1910-1920	Con	Zac	Hax	Hax	Hax	0.00
1920-1930						0.00
1930-1940						0.00
1940-1950						0.00
1950-1960						6.72
1960-1970						9.03
1970-1980						6.56
1980-1985	2.77	1.35	1.54	4.08	3.76	

Cuadro 2.9 . TASA DE CRECIMIENTO MEDIO ANUAL DE LA POBLACION URBANA, SEGUN REGIONES GEOECONOMICAS Y ENTIDADES FEDERATIVAS DE MEXICO, 1910-1985. (POR CIENTO). (HOJA 2)

PERIODO	REGION					
	Golfo	Metropolitana	Sur	P. de Yucatán		
1910-1920	Tab	Mex	Chis	Cam	0.00	
1920-1930					-0.65	
1930-1940					5.11	
1940-1950					3.07	
1950-1960					4.25	
1960-1970	8.28	21.57	8.97	7.10		
1970-1980	7.42	15.02	6.72	1.89		
1980-1985	6.84	9.52	4.35	6.45		
		5.55	3.54	7.90		
1910-1920	Ver	DF	Gro	Q. Roo	0.00	
1920-1930					1.05	
1930-1940					3.54	
1940-1950					2.70	
1950-1960					3.38	
1960-1970	7.29	2.45	0.00	0.00		
1970-1980	6.57	4.41	0.00	0.00		
1980-1985	4.78	3.82	0.00	0.00		
	4.46	6.20	0.00	0.00		
		4.84	6.72	0.00		
		3.65	11.50	0.00		
		2.66	5.01	18.22		
		1.45	5.09	12.16		
1910-1920			Oax	Yuc	-2.70	
1920-1930					1.29	2.17
1930-1940					-1.29	2.10
1940-1950					4.87	0.21
1950-1960					6.72	3.88
1960-1970			8.16	4.14		
1970-1980			5.43	2.45		
1980-1985			3.34	6.79		
				6.91		

Fuente: MEXICO, CONSEJO NACIONAL DE POBLACION: *Poblacion y desarrollo en México y el mundo*, Mexico, CONAPO, 1982, v. IV, p. 1058, 1059.

Cuadro 2.10. CALCULOS CORRESPONDIENTES A LA PRUEBA DE LILLIEFORS  
 SOBRE LAS TASAS DE CRECIMIENTO DE LA POBLACION URBANA POR  
 REGION GEOECONOMICA.

REGION	n	dn	$d_nCO.95)^{**}$	$d_nCO.99)^{**}$
Nordeste		0.149 <sup>*</sup>		
Norte	40	0.066	0.140	0.163
Occidente		0.072		
Centro		0.195		
Noreste		0.143		
Golfo	16	0.074	0.213	0.250
Metropolitana		0.234 <sup>*</sup>		
Sur	24	0.178 <sup>*</sup>	0.173	0.200
P. de Yucatán		0.160		

\* Se rechaza el supuesto de distribución Normal con  $\alpha=0.05$ .

\*\* Ver tabla 19 en CONOVER, W. J.: *Practical Nonparametric Statistics*,  
 E.U.A., John Wiley & sons, 1971, p. 398.

Cuadro 2.11. CALCULOS CORRESPONDIENTES A LA COMPARACION DE  $k$  MEDIAS CON BASE EN LA DISTRIBUCION NORMAL, SOBRE LAS TASAS DE CRECIMIENTO DE LA POBLACION URBANA POR REGION GEOECONOMICA.

REGION ( $k=9$ )	$n_j$	$\bar{X}_j$	$\sum_{i=1}^{n_j} X_{ji}$	$n_j(\bar{X}_j - \bar{X})^2$	$\sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2$
Noroeste	40	4.183	167.32	6.498	789.124
Norte	40	3.433	137.32	4.816	256.139
Noreste	16	5.018	80.288	24.522	137.548
Occidente	40	3.457	138.28	4.173	337.808
Centro	40	2.987	119.48	25.154	385.656
Golfo	16	4.257	68.112	3.640	106.063
Metropolitana	16	5.015	80.24	24.404	542.295
Sur	24	3.685	88.44	0.217	304.381
P. de Yucatán	24	3.678	88.272	0.250	384.266
$\sum_j$	$n=256$		967.752	93.672	3242.278
		$\rightarrow \bar{X} = 3.780$			

## CAPITULO 3

## CAPITULO 3. APLICACIONES EN CIENCIAS SOCIALES

### 3.1. SALUD

#### 3.1.1. Unidades médicas del sector salud

Dentro del Sector Salud, las unidades médicas se encuentran clasificadas en tres grandes grupos: hospitalarias generales, hospitalarias especiales y de consulta externa.

La Secretaría de Salud (SSA) y el Instituto Mexicano del Seguro Social (IMSS) poseen la mayor parte de las unidades médicas que funcionan en el país, la SSA más de la mitad (5734) y el IMSS más de un tercio (3940). Por otro lado, de un total de 10983 unidades, 10250 (más del 90%) son de consulta externa (Ver Cuadro 3.1)

Cabe destacar que las unidades del DDF son principalmente hospitalarias especiales (20 de un total de 34). En contraste, la EEDENA sólo cuenta con una unidad dentro de ese mismo grupo, ya que la gran mayoría son de consulta externa.

Si se analiza el Cuadro 3.1, aparentemente existe cierta relación entre la institución y el tipo de unidad, lo cual puede ser verificado mediante una prueba de independencia<sup>1</sup> donde la hipótesis en cuestión es:

H<sub>0</sub>: Dentro del Sector Salud, el tipo de unidades médicas en servicio no depende de la institución que las posee

vs.

H<sub>1</sub>: De acuerdo a la institución de que se trate es el tipo de unidades médicas que posee.

---

<sup>1</sup> Ver apéndice C, sección I.

Considerando los datos contenidos en los Cuadros 3.1 a 3.3, se obtiene que:

$$\chi^2 = 1643.754,$$

que resulta ser mayor que  $\chi^2_{(4)(2)}(1-\alpha) = \chi^2_{(8)}(1-\alpha)$ , para cualquier nivel de significancia  $\alpha$  usual.

Dado que se tiene un tamaño de muestra muy grande (10983) conviene calcular el coeficiente de contingencia  $C^2$  para obtener conclusiones más certeras, el cual toma el valor de:

$$C = 0.1302,$$

donde la probabilidad de ocurrencia conforme a  $H_0$  (no hay relación entre el tipo de unidad médica y la institución) es elevada y superior que  $\alpha$ , para cualquier valor usual, lo que significa que la relación entre el tipo de unidad médica y la institución del Sector Salud que la posee no es estadísticamente significativa.

Observando el Cuadro 3.3 podrá notarse que la mayor discrepancia cuadrada entre el valor observado y el esperado conforme a  $H_0$ , estandarizada por el valor esperado, corresponde al número de unidades hospitalarias especiales del Departamento del Distrito Federal (DDDF), dado que bajo el supuesto de independencia se esperaría que fuera de cero.

Es precisamente en el caso de las unidades médicas hospitalarias especiales donde en conjunto se registra la mayor discrepancia (1325.706). Y por institución, los números que presenta el DDF, contribuyen de manera importante (con 1340.008) para que el valor de  $\chi^2$  sea tan elevado.

Si las conclusiones se hubieran basado en el valor de  $\chi^2$ , entonces se habría encontrado que el tipo de unidad médica se relaciona con la institución; situación que parece no ocurrir en las unidades de consulta externa, que representan el 93% de la población total; pero, por manejarse un tamaño de muestra grande, la  $\chi^2$  resulta ser elevada.

---

2-  
Ver apéndice E, sección III-3.

Cuadro 3.1. UNIDADES MEDICAS EN SERVICIO, POR TIPO DE UNIDAD, SEGUN INSTITUCIONES DEL SECTOR SALUD, 1987.

TIPO DE U. M. INSTITUCION	Hospitalarias		De consulta externa	TOTAL POR INSTITUCION
	Generales	Especiales		
SSA	145 (0.03)	30 (0.01)	5559 (0.97)	5734
IMSS	408 (0.10)	23 (0.01)	3512 (0.89)	3943
ISSSTE	52 (0.05)	23 (0.02)	973 (0.93)	1048
SEDENA	25 (0.11)	1 (0.00)	201 (0.89)	227
DDF	9 (0.28)	20 (0.59)	5 (0.15)	34
<b>TOTAL POR TIPO DE U. M.</b>	<b>638 (0.08)</b>	<b>97 (0.01)</b>	<b>10250 (0.99)</b>	<b>10985</b>

Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA:  
Anuario Estadístico de los Estados Unidos Mexicanos 1985-1989,  
México, INEGI, 1990, p. 374.

Cuadro 3.2. VALORES ESPERADOS, CONFORME A Ho.

E <sub>i</sub>	Hospitalarias		De consulta externa	Σ <sub>i</sub>
	Generales	Especiales		
SS	332.043 (0.06)	50.642 (0.01)	5351.315 (0.93)	5734
IMSS	228.156 (0.08)	34.797 (0.01)	3677.046 (0.93)	3940
ISSSTE	60.687 (0.06)	9.256 (0.01)	978.057 (0.93)	1048
SEDENA	13.145 (0.06)	2.005 (0.01)	211.850 (0.93)	227
DDF	1.969 (0.06)	0.300 (0.01)	31.731 (0.93)	34
Σ <sub>i</sub>	636 (0.06)	97 (0.01)	10250 (0.93)	10983

Cuadro 3.3. DIFERENCIA AL CUADRADO, ENTRE LO OBSERVADO Y LO ESPERADO CONFORME A  $H_0$ , ESTANDARIZADA POR EL VALOR ESPERADO.

$\frac{O_{ij} - E_{ij}}{E_{ij}}^2$	Hospitalarias		De consulta externa	$\Sigma$
	Generales	Especiales		
SS	105.363	8.414	8.050	121.837
IMSS	137.071	3.990	7.408	148.479
ISSSTE	1.244	20.409	0.028	21.679
SEDENA	10.891	0.504	0.558	11.953
DDF	25.109	1292.380	22.519	1340.008
$\Sigma$	279.479	1325.706	38.565	$\chi^2 = 1643.754$

### 3.2. EDUCACION

#### 3.2.1. Elección del área profesional y el sexo del estudiante

Es comun escuchar comentarios respecto a que existen carreras "propias" para hombres y otras adecuadas para mujeres, con base en el tipo de actividades desempeñadas en cada una de ellas y su marco de desarrollo.

Se intentará verificar estadísticamente si existe alguna relación entre la carrera y el sexo del estudiante, considerando para ello estadísticas por carrera correspondientes al ciclo 1988-1989 del conjunto de facultades y Escuelas Nacionales de Artes Plásticas, Enfermería y Obstetricia, Musica y Trabajo Social de la Universidad Nacional Autónoma de México (UNAM). En estas escuelas y facultades se encuentran todas las carreras impartidas por la UNAM.

Dado el objetivo del análisis, una de las pruebas que puede ser util es la de independencia<sup>1</sup>, donde se establece:

Ho: La elección del área de la carrera no depende del sexo del estudiante.

vs.

H1: La elección de Área de la carrera depende del sexo del estudiante.

Como se aprecia en el Cuadro 3.4, los estudiantes están categorizados de acuerdo al sexo y al tipo de carrera que estudian.

En las cifras totales por sexo se observa que hay un mayor numero de estudiantes varones, constituyendo casi el 58% de la población estudiantil. Tomando en cuenta los totales por tipo de carrera puede verse la marcada preferencia por el Área Social-Humanística, contrastando con la Artística; mientras que en los tipos 1 a 3 no existen diferencias muy marcadas entre si.

Es muy notorio el hecho de que casi toda la población

---

<sup>1</sup> Ver apéndice C, sección I.

femenina se reparte entre los tipos de carrera 2 a 5, tan solo 800 de 25399 mujeres estudian alguna carrera del Área Físico-Matemática.

Según la información proporcionada por los cuadros 3.4 y 3.5, la estadística  $X^2$  de Pearson resulta tomar el valor de:

$$X^2 = 11471.72.$$

Como se observa  $X^2$  es mayor, por mucho, que  $\chi^2_{(4)}(1 - \alpha)$  para cualquier  $\alpha$ . No obstante, en este caso es conveniente calcular el coeficiente de contingencia  $C^2$ , para obtener resultados más precisos, dado que se ha considerado un tamaño de muestra grande (83474) lo cual podría hacer parecer que la diferencia entre lo observado y lo esperado conforme a  $H_0$  es demasiado grande. Así se tiene que:

$$C = 0.1208,$$

valor que no es estadísticamente significativo, suponiendo una distribución  $\chi^2_{(4)}$  para cualquier valor de  $\alpha$  usual.

De esta forma, no existe evidencia de alguna relación entre la elección de una área profesional y el sexo de los estudiantes de la UNAM.

Pese al resultado anterior, comparando los cuadros, se observa una poca frecuente inclinación por parte de las estudiantes del sexo femenino hacia las carreras del área Físico-Matemática, situación contraria con lo que respecta al sexo masculino.

El resultado previo indica que entre los estudiantes del nivel superior, hombres y mujeres se incorporan por igual a todo tipo de carreras profesionales, exceptuando a las que corresponden al área Físico-Matemática, la cual tiene poca aceptación entre las estudiantes.

---

2  
Ver apéndice E, sección III-3.

**Cuadro 3.4. NUMERO DE ESTUDIANTES DE LA UNAM POR TIPO DE CARRERA Y POR SEXO**

SEXO TIPO DE CARRERA *	MASCULINO	FEMENINO	TOTAL POR CARRERA
1	15005 (0.95)	800 (0.05)	15805
2	9248 (0.52)	8508 (0.48)	17757
3	9784 (0.52)	9099 (0.48)	18883
4	12713 (0.45)	15682 (0.55)	28395
5	1345 (0.51)	1309 (0.49)	2654
<b>TOTAL POR SEXO</b>	<b>48078 (0.58)</b>	<b>35309 (0.42)</b>	<b>83474</b>

Fuente: UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO: *Agencia Estadística 1989*, Mexico, UNAM, 1989, p. 31-34.

\* Los tipos de carrera corresponden a las áreas:

1. Físico-Matemática. (Actuaría, Ingeniería, Matemáticas, etc.)
2. Químico-Biológica. (Química, Medicina, Odontología, etc.)
3. Económico-Administrativa. (Contaduría, Economía, Informática, etc.)
4. Social-Humanística. (Derecho, Filosofía, Pedagogía, etc.)
5. Artística. (Canto, Instrumentación, Piano, etc.)

Cuadro 3.5. VALORES ESPERADOS CONFORME A Ho.

Ei j	MASCULINO	FEMENINO	$\Sigma$ j
1	9102.54 (0.58)	6702.46 (0.42)	15805
2	10226.75 (0.58)	7530.25 (0.42)	17757
3	10863.73 (0.58)	7999.27 (0.42)	18863
4	16353.47 (0.58)	12041.53 (0.42)	28395
5	1528.51 (0.58)	1125.49 (0.42)	2654
$\Sigma$ i	48075 (0.58)	35399 (0.42)	83474

### 3.2.2. Número de alumnos por maestro

Una medida relevante para obtener alguna información sobre las características que presenta el Sistema Educativo Nacional podría ser el número de alumnos por maestro (NAPM). Ciertamente, la interpretación de este parámetro debe ser realizada con cautela, pues sólo es de carácter cuantitativo y de ninguna forma resalta las características particulares de cada región y menos aun de cada maestro.

La distribución por edad de la población mexicana, entre otros factores, podrían hacer pensar que el número promedio de alumnos que atiende un maestro varía según el nivel educativo. Se puede intentar verificar estadísticamente este supuesto considerando la información del Cuadro 3.6

Las estadísticas generales (Cuadro 3.7) revelan que el promedio anual del NAPM es cercano a 35 en el nivel preescolar, es aproximadamente de 45 para la primaria y de ahí en adelante tiende a disminuir conforme el nivel es más elevado hasta llegar a ser cercano a 10 para el nivel superior. Se muestra también que el NAPM alcanzó su máximo nivel en el año de 1960 en el nivel primaria y el mínimo en 1968 en la enseñanza profesional.

Se observa que en la educación preescolar el indicador no ha sufrido grandes variaciones a través del tiempo, presentando una desviación estándar de 2.537; algo similar ocurre para el nivel superior. La mayor variabilidad se manifiesta en el nivel medio superior con 3.888.

Para saber qué tipo de prueba es susceptible de ser utilizada en este caso, se verificara si los datos disponibles satisfacen el supuesto de distribución Normal.

Así:

Para cada nivel educativo,

H<sub>0</sub>: El número anual de alumnos por maestro presenta una Distribución Normal.

vs.

H<sub>1</sub>: La Distribución del número anual de alumnos por maestro no es Normal.

Según la prueba de Lilliefors<sup>3</sup> y el Cuadro 3.8, puede concluirse que en los casos de los niveles primaria y medio superior, existe evidencia para rechazar H<sub>0</sub>, aun tomando un nivel de significancia muy pequeño ( $\alpha=0.01$ ), en los demás casos el supuesto se acepta incluso con un valor de  $\alpha$  superior ( $\alpha=0.10$ ). Este resultado da una idea de la variabilidad en la evolución del indicador en los niveles primaria y Medio superior.

Dado que el supuesto de distribución Normal no se cumple en las cinco variables, se aplicará una prueba no paramétrica, en particular el análisis de varianzas de Friedman,<sup>4</sup> puesto que se dispone de cinco muestras relacionadas medidas en una escala más que ordinal, para decidir:

H<sub>0</sub>: En el Sistema Educativo Nacional, el número anual de alumnos por maestro es semejante en todos los niveles.

vs.

H<sub>1</sub>: El número anual de alumnos por maestro es diferente, de acuerdo al nivel educativo que se trate.

El rango que ocupan las observaciones anuales para cada nivel respecto al número de alumnos por maestro, se observa en el Cuadro 3.9.

A lo largo de todo el periodo examinado, los maestros de los niveles preescolar y primaria son los que comparativamente tienen a su cargo a un número mayor de alumnos; a estos, en la mayoría de los casos, les sigue el nivel secundaria, después el medio superior y finalmente son los profesores del nivel superior, quienes dentro de todo el Sistema Educativo Nacional, atienden grupos de alumnos menos numerosos. La situación de estos tres

---

<sup>3</sup> Ver apéndice B, sección II.

<sup>4</sup> Ver apéndice A, sección I-2.

últimos niveles presenta algunas variaciones a través de los años, siendo esta tendencia la más acusada. Sin embargo, resalta el comportamiento observado en 1957, cuando las posiciones entre los niveles secundaria y superior se invirtieron.

En términos generales puede decirse que el nivel educativo y el número de alumnos por maestro se encuentran relacionados inversamente.

Las sumas de rangos ( $R_j$ ) reflejan las consideraciones previas y al calcular  $\chi_r^2$  resulta que:

$$\chi_r^2 = 123.95,$$

valor que señala que hay diferencias importantes en los niveles, con una probabilidad de ocurrencia conforme a  $H_0$  de cero hasta tres cifras significativas. Comparando esta probabilidad con cualquier nivel de significancia usual resulta que se tienen bases para concluir que existe evidencia para rechazar  $H_0$  y considerar que el número de alumnos por maestro varía según el nivel educativo.

... De alguna manera, esta aplicación ilustra uno entre los innumerables problemas que acarrea una estructura por edades como la que prevalece en México: la dificultad para satisfacer las necesidades educativas de la población en edad escolar, principalmente en los grados iniciales, donde el NAPM presenta los niveles máximos.

Cuadro 3.6. NUMERO DE ALUMNOS POR MAESTRO SEGUN NIVEL EDUCATIVO, 1952-1984.

Año	NIVEL				
	PREESC.	PRIM.	SECUND.	MED. SUP.	SUP.
1952	32.309	44.542	6.732	5.751	5.632
1953	36.672	44.733	9.269	6.902	5.306
1954	38.424	45.678	9.628	7.467	5.499
1955	35.129	45.887	10.803	9.793	7.792
1956	32.719	45.859	9.529	7.339	4.836
1957	33.026	45.909	7.428	8.629	8.760
1958	33.228	46.014	10.085	8.392	7.237
1959	32.642	48.383	12.214	10.909	5.921
1960	34.481	50.009	11.759	13.188	7.312
1961	34.573	48.653	11.565	10.125	7.728
1962	34.151	47.912	12.659	9.438	7.862
1963	34.905	47.645	13.142	8.419	7.360
1964	35.181	46.955	12.194	9.644	7.609
1965	35.232	47.891	12.977	10.353	10.272
1966	35.045	47.717	13.594	9.845	10.815
1967	33.583	47.450	13.715	10.402	11.945
1968	33.091	46.718	13.962	10.260	12.821
1969	32.898	46.259	13.957	10.410	10.130
1970	38.021	47.671	16.272	14.733	10.827
1971	37.876	46.726	16.795	15.299	11.262
1972	37.129	45.842	16.584	15.841	11.445
1973	37.686	45.449	16.466	16.309	10.264
1974	37.726	45.269	17.208	17.538	11.260
1975	38.165	44.782	17.112	18.381	11.427
1976	38.693	44.060	16.795	16.798	13.204
1977	38.395	42.431	17.781	16.112	11.681
1978	37.625	42.378	18.105	17.256	11.873
1979	34.129	40.698	18.074	17.205	12.215
1980	33.107	39.087	17.996	17.144	12.709
1981	32.421	37.414	18.084	16.982	13.438
1982	30.228	36.644	18.555	16.465	12.002
1983	31.076	35.923	18.715	15.843	12.143
1984	29.692	34.787	18.874	14.607	11.273

Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *Estadísticas Históricas de México*, México, INEGI, 1990. t. I, p. 86-87.

Cuadro 3.7. ESTADISTICAS GENERALES DEL NUMERO ANUAL DE ALUMNOS POR MAESTRO, SEGUN NIVEL EDUCATIVO, 1952-1984.

Estadística*	NIVEL				
	PREESC.	PRIM.	SECUND.	MED. SUP.	SUP.
V. MIN.	29.692	34.787	6.732	5.751	4.835
V. MAX.	38.693	50.009	19.874	18.351	13.438
MEDIA	34.828	44.584	14.167	12.638	9.753
DESV. EST.	2.537	3.861	3.600	3.888	2.609

\*En todos los casos n=33.

Cuadro 3.8. CALCULOS CORRESPONDIENTES A LA PRUEBA DE LILLIEFORS PARA EL NUMERO DE ALUMNOS POR MAESTRO DE CADA NIVEL EDUCATIVO.

NIVEL	n	dn	dn(0.90)**	dn(0.99)**
PREESCOLAR		0.107		
PRIMARIA		0.193*		
SECUNDARIA	33	0.145	0.140	0.179
MEDIO SUPERIOR		0.193*		
SUPERIOR		0.143		

\*Caso en el que se rechaza el supuesto de distribución normal con  $\alpha=0.01$ .

\*\*Ver tabla 15 en CONOVER, W. J.: *Practical Nonparametric Statistics*, E. U. A., John Wiley & sons, 1971, p. 398.

Cuadro 3.9. RANGOS PARA EL ANALISIS DE VARIANZA DE DOS CLASIFICACIONES DE FRIEDMAN PARA EL NUMERO DE ALUMNOS POR MAESTRO Y POR NIVEL EDUCATIVO, 1952-1984.

Año	NIVEL				
	PREESC.	PRIM.	SECUND.	MED. SUP.	SUP.
1952	4	5	3	2	1
1953	4	5	3	2	1
1954	4	5	3	2	1
1955	4	5	3	2	1
1956	4	5	3	2	1
1957	4	5	1	2	3
1958	4	5	3	2	1
1959	4	5	3	2	1
1960	4	5	2	3	1
1961	4	5	3	2	1
1962	4	5	3	2	1
1963	4	5	3	2	1
1964	4	5	3	2	1
1965	4	5	3	2	1
1966	4	5	3	1	2
1967	4	5	3	1	2
1968	4	5	3	1	2
1969	4	5	3	2	1
1970	4	5	3	2	1
1971	4	5	3	2	1
1972	4	5	3	2	1
1973	4	5	3	2	1
1974	4	5	2	3	1
1975	4	5	2	3	1
1976	4	5	2	3	1
1977	4	5	3	2	1
1978	4	5	3	2	1
1979	4	5	3	2	1
1980	4	5	3	2	1
1981	4	5	3	2	1
1982	4	5	3	2	1
1983	4	5	3	2	1
1984	4	5	3	2	1
Pj	132	165	93	67	38

\* Casos en los que no se presenta la tendencia general.

### 3.3. CULTURA

#### 3.3.1. Uso de bibliotecas

Como fuente de conocimiento, los libros poseen gran valor, y es por ello que las bibliotecas constituyen un importante medio de acceso a la cultura.

Podría pensarse que dentro de todos los factores que determinan el número de obras consultadas en una biblioteca (OBC), durante cierto periodo, está el número de lectores registrados (LEC), durante ese mismo periodo.

Por ejemplo, considerando cifras por entidad federativa de la República Mexicana (ver Cuadro 3.10), según las estadísticas generales, durante 1988 se consultaron en promedio 3,685,463 obras disponibles en las bibliotecas por entidad, lo cual fue hecho por 2,950,738 lectores. En el D. F. se registraron los valores máximos de ambas variables y en el estado de Nayarit los mínimos, observándose una amplia variabilidad en los datos, pues mientras que en Nayarit el número de obras consultadas y el de lectores no llegó a 300,000, en el D. F. en los dos casos se superó a los 25 millones.

En la Gráfica 3.1 puede observarse que existe cierta relación lineal entre ambas variables (OBC y LEC), por lo cual es probable que la variable LEC resulte adecuada para explicar OBC a través de la siguiente ecuación de un modelo de regresión lineal simple<sup>1</sup>:

$$OBC_i = \beta_0 + \beta_1 LEC_i + \epsilon_i$$

$i = 1, 32.$

Según el Cuadro 3.11, el valor que toma la  $R^2$  indica que con el modelo planteado se explica más del 96% de la variación total corregida por la media de la variable OBC.

De acuerdo al método de estimación de mínimos cuadrados ordinarios, se obtiene que la constante ( $\beta_0$ ) es negativa y como era de esperarse el término asociado a la variable explicativa

---

<sup>1</sup> Ver apéndice D, sección I.

$(\beta_1)$  es mayor que cero, esto es, a mayor número de lectores, mayor número de obras consultadas y puede decirse que de acuerdo al valor de  $\beta_1$  cada lector consulta alrededor de un libro en promedio.

El resultado de las pruebas  $t$  sobre los parámetros, indica que con cualquier nivel de significancia usual, con una probabilidad de error muy pequeña, puede aceptarse  $H_0 (\beta_1=0)$  en el caso de  $\beta_0$ , dado que la probabilidad de dos colas (0.788) asociada con la ocurrencia de ese valor de  $t$ , conforme a  $H_0$ , es superior que  $\alpha$ . Para  $\beta_1$  se rechaza  $H_0$ .

Las pruebas  $t$  y  $F$  llevan a la misma conclusión cuando se trata de un modelo de regresión lineal simple. En este caso la prueba  $F$  también señala indicios de que el modelo propuesto parece explicar adecuadamente el comportamiento de la variable OBC, dado que la probabilidad de ocurrencia, suponiendo una distribución  $F_{1, 90}$ , es de cero hasta tres cifras decimales, por lo cual existe evidencia para rechazar  $H_0$ .

Los resultados anteriores, así como el cálculo de intervalos de confianza para los parámetros y los pronósticos carecen de validez antes de verificar los supuestos del modelo mediante el análisis de residuales, lo cual se realizará a continuación.

Como se trata de un modelo con una sola variable explicativa (LEC) el supuesto de no multicolinealidad se valida.

Uno de los métodos para detectar la presencia de heterocedasticidad es el que se vale de gráficas de los residuales ( $e_i$ ) contra los valores estimados ( $\hat{Y}_i$ ) y contra las variables explicativas que en este caso es solo una ( $X_i$ ). El patrón de comportamiento de los residuales señalado por las Gráficas 3.2 y 3.3 indica que a medida que aumenta el valor de la abscisa ( $\hat{Y}_i$  o  $X_i$ ) también crece la varianza de los residuales, por lo cual se sospecha que los residuales no tienen varianza constante.

Por otro lado, la estadística  $D$  de Durbin-Watson con valor de 1.884 superior al valor crítico  $D_{u1.57}^2$  para un tamaño de muestra de 32 y una variable explicativa en el modelo, señala que aparentemente no existe autocorrelación, cuando menos de primer

Ver Tabla D.5a en: SUJARATI Damodar, Econometría, trad. J. Manuel Mesa, México, 1986, p. 426.

orden. Esto significa que el valor de un residual no se encuentra influenciado por el residual anterior. El problema de la autocorrelación es más común cuando se utilizan series de tiempo que cuando se usa información de corte transversal<sup>3</sup> como en este caso.

Para intentar validar el supuesto de distribución Normal de los residuales, obsérvese la Gráfica 3.4. Se trata de una gráfica de papel Normal en la cual puede encontrarse evidencia de que se viola dicho supuesto, dado que no podría decirse que la ubicación de los puntos describe una línea recta, como se requeriría para aceptar el supuesto de distribución Normal, del cual se duda a partir de que se detecta la heterocedasticidad.

Una de las opciones para corregir esta falla es transformar los datos antes de estimar el modelo, en este caso se aplicará logaritmo natural pues esta transformación comprime las escalas en que están medidas las variables y así frecuentemente se reduce la heterocedasticidad<sup>4</sup>. Por lo tanto la nueva ecuación es:

$$\text{LOBC}_i = \beta_0 + \beta_1 \text{LLEC}_i + \epsilon_i$$
$$i = 1, 32,$$

considerando los demás supuestos de un modelo de regresión lineal.

El Cuadro 3.13 contiene los datos transformados y en el 3.14 se observa que  $\bar{R}^2=0.884$  ( $R^2$  ajustada por el tamaño de muestra y por el número de variables explicativas incluidas en el modelo), es inferior al coeficiente de determinación del modelo inicial (0.981) y sin embargo sigue indicando que el modelo propuesto explica de manera adecuada el comportamiento del número de obras consultadas en bibliotecas por entidad federativa.

Las gráficas 3.5 y 3.6 revelan que aparentemente se ha corregido en buena medida la heterocedasticidad, ya que en cada gráfica podría dibujarse una banda horizontal que incluyera a todos los residuales, lo cual estaría indicando que la varianza de estos es constante independientemente de los valores estimados o del número de lectores (LEC).

Hay indicios de que el supuesto de no autocorrelación no se viola, dado que  $D = 2.177$ .

<sup>3</sup>Op. cit. p. 221.

<sup>4</sup>Op. cit. p. 207.

Por otro lado, analizando la gráfica de papel Normal de los residuales (Gráfica 3.7), podría sospecharse que existe evidencia para aceptar que siguen la distribución Normal, ya que si se traza una línea recta sobre los puntos, esta abarca a la mayoría.

En resumen, al parecer se satisfacen los supuestos del modelo sobre  $\beta_0$  después de haber realizado la transformación logarítmica, lo cual valida los resultados contenidos en el Cuadro 3.14. De esta forma, según las pruebas  $t$ , el modelo puede prescindir de  $\beta_0$  ya que la probabilidad asociada a  $t=0.75$  es de 0.458, suponiendo una distribución  $t_{(31)}$ . Por lo que, con una probabilidad de error muy pequeña, puede aceptarse  $H_0$  con cualquier nivel de significancia usual.

Respecto al coeficiente  $\beta_1$ , las pruebas  $t$  y  $F$  correspondientes señalan que existe evidencia para rechazar  $H_0$ . Por lo tanto el modelo a estimar ahora es:

$$LOBC_i = \beta_1 LLEC_i + \epsilon_i \quad i = \overline{1, 32},$$

junto con los supuestos restantes.

Según el Cuadro 3.15, redondeando cifras, tanto  $R^2$  como  $\bar{R}^2$  alcanzan el valor de 1, lo cual indica que la variable LOBC es explicada casi al 100% por el modelo aplicado.

La recta ajustada, además de pasar por el origen ( $\hat{\beta}_0 = 0$ ), tiene una pendiente positiva ( $\hat{\beta}_1 = 1.014$ ), es decir los estados que registran un elevado número de lectores en sus bibliotecas, también presentan una gran cantidad de obras consultadas. Como se trata de logaritmos la interpretación de  $\hat{\beta}_1$  no es directa, lo que puede decirse es que  $\hat{\beta}_1$  indica el incremento porcentual en el número de obras consultadas (OBC) cuando el número de lectores (LEC) crece en un 1% de una entidad a otra<sup>5</sup>.

Puesto que la varianza fue estabilizada la pendiente no está sujeta a una grandes variaciones. El intervalo de confianza del 95% tiene la siguiente forma<sup>6</sup>:

$$\beta_1 \in [1.014 - 0.004(2.042) , 1.014 + 0.004(2.042)] \\ \Leftrightarrow \beta_1 \in (1.006 , 1.022).$$

El Cuadro 3.16 contiene los residuales así como los valores estimados producto de este último modelo. Analizando las gráficas

<sup>5</sup>Op. cit. p. 207.

<sup>6</sup>Ver apéndice D, sección I-c.

3.8 a 3.10, se observa que al parecer se satisfacen los supuestos.

Para complementar las conclusiones anteriores, valdria la pena realizar un analisis similar, excluyendo al Distrito Federal, para saber si el dato correspondiente, ademas de ser discrepante, tambien es influyente. Y si este es el caso, conocer de que manera cambian los resultados obtenidos, con respecto a los del analisis anterior

Definitivamente se ha planteado un modelo muy sencillo, en el cual podrian introducirse otras variables que proporcionaran informacion mas completa, o bien, formular otros modelos que captaran las diferencias culturales de un estado a otro introduciendo variables que detallaran el tipo de obras consultadas, si es que existe informacion al respecto.

Cuadro 3.10. OBRAS CONSULTADAS Y LECTORES DE LAS BIBLIOTECAS DE LA REPUBLICA MEXICANA, 1988.

Entidad Federativa	Obras consultadas	Lectores
Aguascalientes	1680921	1692986
Baja California	1945527	1682569
Baja California Sur	1360823	571342
Campeche	487408	501841
Coahuila	3802932	3016218
Colima	825282	481270
Chiapas	2812819	2572397
Chihuahua	1011227	1632373
Distrito Federal	33911579	25073499
Durango	2263863	1748585
Guanajuato	2984186	2016935
Guerrero	2435090	1983665
Hidalgo	1849491	1855175
Jalisco	5372821	4727491
México	5934983	5712753
Michoacán	3359518	2950553
Morelos	3351049	1556267
Nayarit	224278	291200
Nuevo León	5254385	6181377
Oaxaca	1611010	1503498
Puebla	5360450	3819761
Querétaro	7101392	2059219
Quintana Roo	978448	735374
San Luis Potosí	2137157	1666557
Sinaloa	2241073	1691569
Sonora	5173799	5690781
Tabasco	3421757	2631380
Tamaulipas	1436527	1474998
Tlaxcala	1458932	1022470
Veracruz	3179974	2530045
Yucatán	1288650	1199411
Zacatecas	1567516	1272130

Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA; Anuario Estadístico de los Estados Unidos Mexicanos, 1988-1989, México. INEGI, 1990, p. 485.

Cuadro 3.11. ESTIMACIONES CORRESPONDIENTES AL MODELO A:

$$OBC_i = \beta_0 + \beta_1 LEC_i + \varepsilon_i$$

VAR. DEP.: OBC		n=32	R = 0.981	R <sup>2</sup> = 0.962
R <sup>2</sup> = 0.961, GL = 30				
VARIABLE	COEFICIENTE $\hat{\beta}_i$	ERROR ESTANDAR e. e. ( $\hat{\beta}_i$ )	T	PC DOS COLAS
CONSTANTE	-65866.156	242691.567	-0.27	0.798
LEC	1.265	0.046	27.67	0.000

Cuadro 3.12. VALORES ESTIMADOS Y RESIDUALES, CORRESPONDIENTES AL MODELO A.

Entidad Federativa	Estimados	Residuales
Aguascalientes	2074059.051	-394038.051
Baja California	2061811.616	-496284.616
Baja California Sur	656819.325	724203.675
Campeche	568732.449	-81328.449
Coahuila	3748263.154	-145331.154
Colima	517428.749	307853.251
Chiapas	3313489.874	-500870.874
Chihuahua	1998336.708	-987109.708
Distrito Federal	32905100	1006432.217
Durango	2142724.548	121139.482
Guanajuato	2484631.189	499554.911
Guerrero	2417256.436	17833.564
Hidalgo	2027170.775	-177679.775
Jalisco	5912240.846	-539419.846
México	7158145.208	-1223192.20
Michoacán	3677974.999	-318256.999
Morelos	1914742.949	1436306.051
Nayarit	302368.233	-78092.233
Nuevo León	7750739.549	-2496354.54
Oaxaca	1833365.171	-224355.171
Puebla	4764378.889	596071.111
Queretaro	2539101.048	4563290.952
Quintana Roo	864044.542	114403.458
San Luis Potosí	2041563.783	96593.217
Sinaloa	2073192.487	167880.513
Sonora	7130360.709	-1956561.70
Tabasco	3261622.189	160134.911
Tamaulipas	1799329.541	-262802.541
Tlaxcala	1227089.254	231742.746
Veracruz	3133479.909	46494.091
Yucatán	1450838.433	-162188.433
Zacatecas	1542794.603	24721.397

Cuadro 3.13. LOGARITMO NATURAL DEL NO. DE OBRAS CONSULTADAS Y  
Y DEL NO. DE LECTORES DE LAS BIBLIOTECAS DE LA REPUBLICA  
MEXICANA, 1988.

Entidad Federativa	LOBC	LEEC
Aguascalientes	14.33	14.34
Baja California	14.28	14.44
Baja California Sur	14.14	13.26
Campeche	13.10	13.12
Coahuila	15.10	14.92
Colima	13.62	13.04
Chiapas	14.85	14.80
Chihuahua	13.83	14.31
Distrito Federal	17.34	17.09
Durango	14.63	14.37
Guanajuato	14.91	14.52
Guerrero	14.71	14.49
Hidalgo	14.43	14.32
Jalisco	15.50	15.37
México	15.80	15.58
Michoacán	15.03	14.90
Morelos	15.02	14.28
Nayarit	12.32	12.58
Nuevo León	15.47	15.64
Oaxaca	14.29	14.22
Puebla	15.49	15.16
Querétaro	15.78	14.54
Quintana Roo	13.79	13.51
San Luis Potosí	14.57	14.33
Sinaloa	14.62	14.34
Sonora	15.48	15.55
Tabasco	15.05	14.78
Tamaulipas	14.18	14.20
Tlaxcala	14.19	13.94
Veracruz	14.97	14.74
Yucatán	14.07	14.00
Zacatecas	14.27	14.06

Cuadro 3.14. ESTIMACIONES Y TABLA ANDEVA CORRESPONDIENTES AL MODELO B:  $LOBC_i = \beta_0 + \beta_1 LLEC_i + \epsilon_i$ .

VAR. DEP.: LOBC		n=32	R = 0.932	R <sup>2</sup> = 0.869	
R <sup>2</sup> = 0.864, GL = 30					
VARIABLE	COEFICIENTE $\hat{\beta}_i$	ERROR ESTANDAR $e.e.(\hat{\beta}_i)$	T	PUNOS COLAS	
CONSTANTE	0.745	0.990	0.75	0.459	
LLEC	0.962	0.088	14.07	0.000	
ANALISIS DE VARIANZA					
FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	CUADRADO MEDIO	F	P
EXPLICADA POR REGRESION	21.651	1	21.651	198.002	0.000
NO EXPLICADA POR REGRESION	3.280	30	0.109		

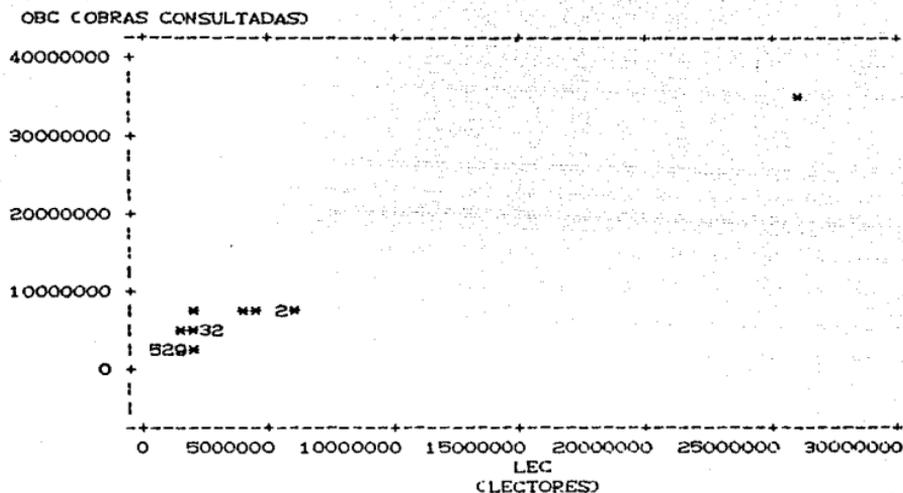
Cuadro 3.15. ESTIMACIONES Y TABLA ANDEVA CORRESPONDIENTES AL MODELO C:  $LOBC_i = \beta_1 LLEC_i + \epsilon_i$ .

VAR. DEP.: LOBC		n=32	R = 1.000	P <sup>2</sup> = 1.000	
R <sup>2</sup> = 1.000, GL = 31					
VARIABLE	COEFICIENTE $\hat{\beta}_i$	ERROR ESTANDAR $e.e.(\hat{\beta}_i)$	T	PUNOS COLAS	
LLEC	1.014	0.004	252.85	0.000	
ANALISIS DE VARIANZA					
FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	CUADRADO MEDIO	F	P
EXPLICADA POR REGRESION	6893.612	1	6893.612	83937.290	0.000
NO EXPLICADA POR REGRESION	3.342	31	0.108		

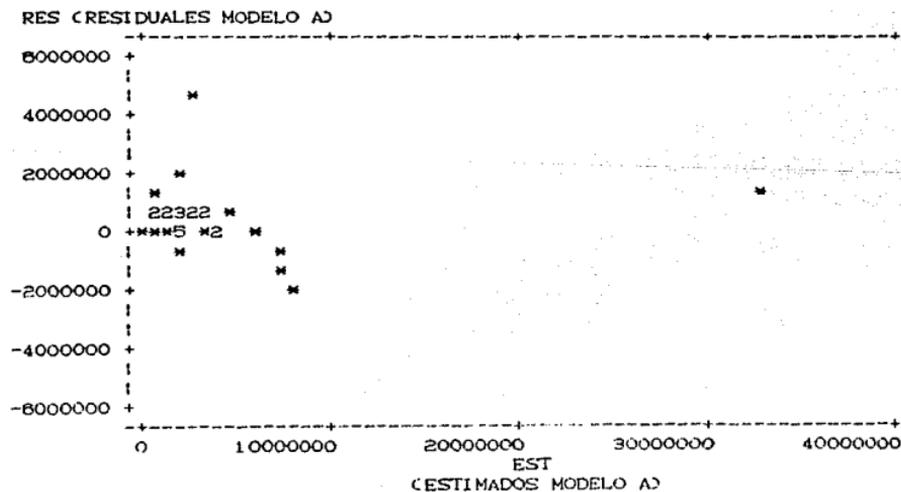
Cuadro 3.16. RESIDUALES Y VALORES ESTIMADOS CORRESPONDIENTES A LOS MODELOS B Y C.

Entidad Federativa	MODELO B		MODELO C	
	EST	PES	EST	PES
Aguascalientes	14.546	-0.216	14.537	-0.207
Baja California	14.546	-0.266	14.537	-0.257
Baja California Sur	13.506	0.634	13.442	0.699
Campeche	13.381	-0.291	13.31	-0.21
Coahuila	15.104	-0.004	15.125	-0.025
Colima	13.294	0.325	13.219	0.401
Chiapas	14.988	-0.139	15.003	-0.153
Chihuahua	14.517	-0.687	14.507	-0.677
Distrito Federal	17.182	0.158	17.215	0.025
Durango	14.574	0.066	14.569	0.062
Guanajuato	14.719	0.191	14.72	0.19
Guerrero	14.69	0.02	14.689	0.021
Hidalgo	14.525	-0.095	14.517	-0.087
Jalisco	15.537	-0.037	15.581	-0.091
México	15.72	-0.12	15.774	-0.174
Michoacán	15.084	-0.054	15.105	-0.075
Morelos	14.469	0.551	14.456	0.564
Nayarit	12.852	-0.532	12.753	-0.433
Nuevo León	15.797	-0.327	15.855	-0.385
Oaxaca	14.43	-0.14	14.415	-0.125
Puebla	15.335	0.155	15.369	0.122
Querétaro	14.738	1.042	14.74	1.04
Quintana Roo	13.747	0.043	13.695	0.094
San Luis Potosí	14.535	0.034	14.527	0.043
Sinaloa	14.546	0.074	14.537	0.083
Sonora	15.71	-0.25	15.764	-0.304
Tabasco	14.969	0.081	14.983	0.067
Tamaulipas	14.411	-0.231	14.395	-0.215
Tlaxcala	14.064	0.126	14.03	0.15
Veracruz	14.93	0.04	14.943	0.027
Yucatán	14.218	-0.148	14.192	-0.122
Zacatecas	14.275	-0.005	14.253	0.017

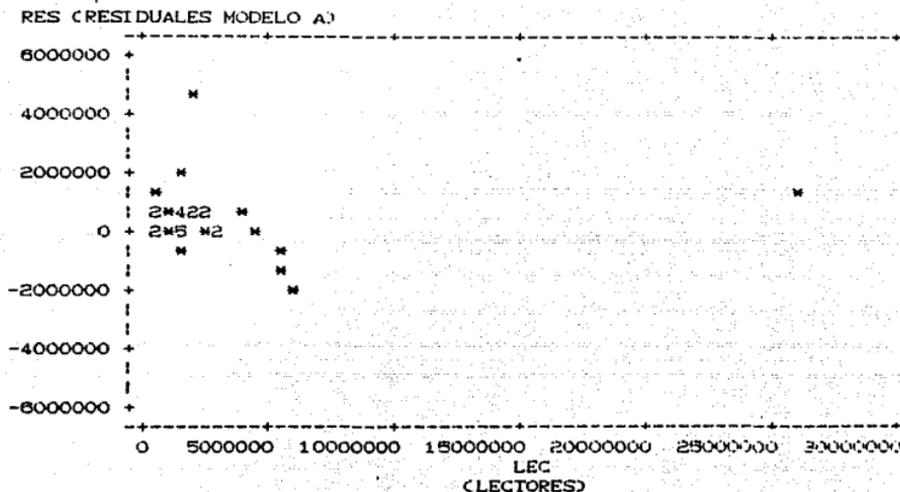
Gráfica 3.1 . OBRAS CONSULTADAS VS. LECTORES DE LAS BIBLIOTECAS,  
POR ENTIDAD FEDERATIVA, 1988.



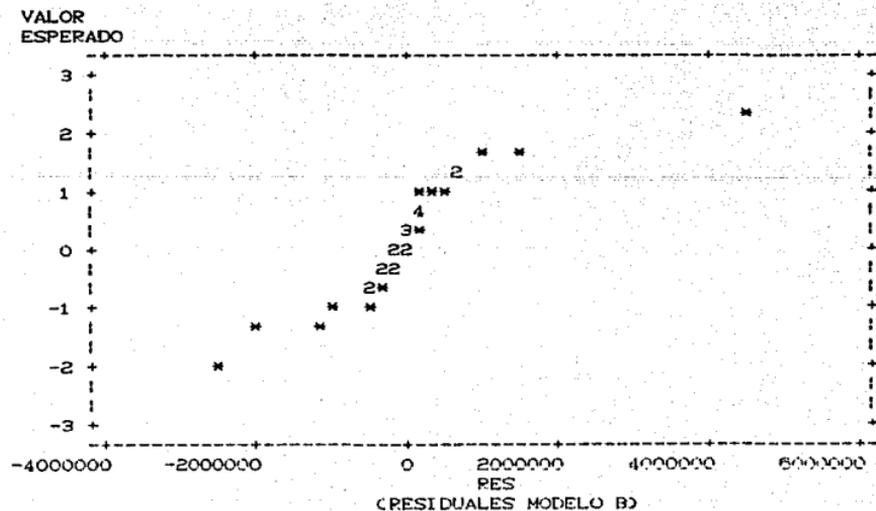
Gráfica 3.2. RESIDUALES VS. VALORES ESTIMADOS DEL MODELO A.



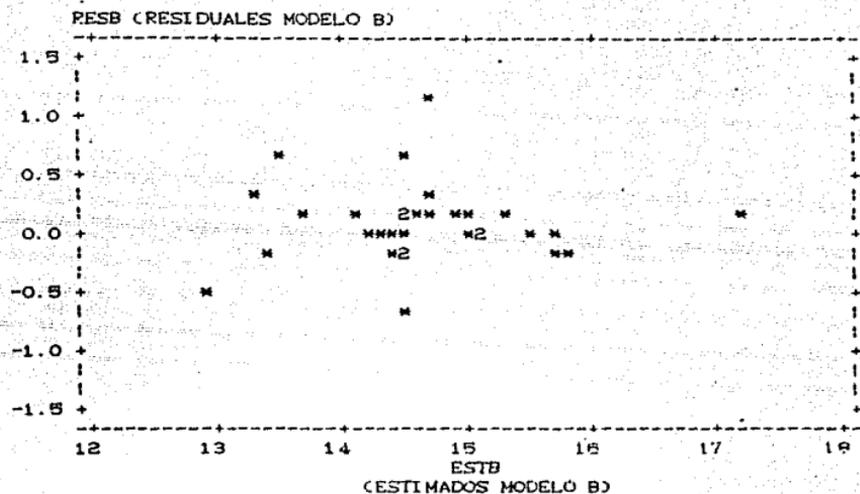
Gráfica 3.3. RESIDUALES DEL MODELO A VS. NO. DE LECTORES.



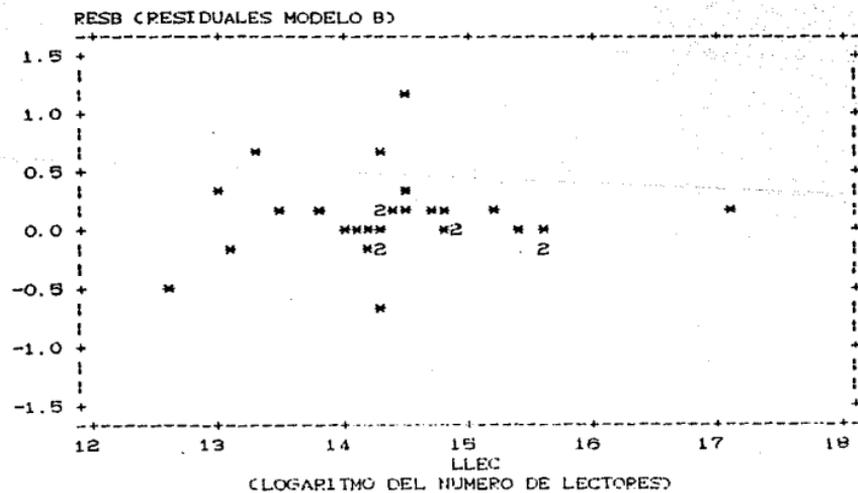
Gráfica 3.4. GRAFICA DE PAPEL NORMAL DE LOS RESIDUALES DEL MODELO A



Gráfica 3.5. RESIDUALES VS. VALORES ESTIMADOS DEL MODELO B.

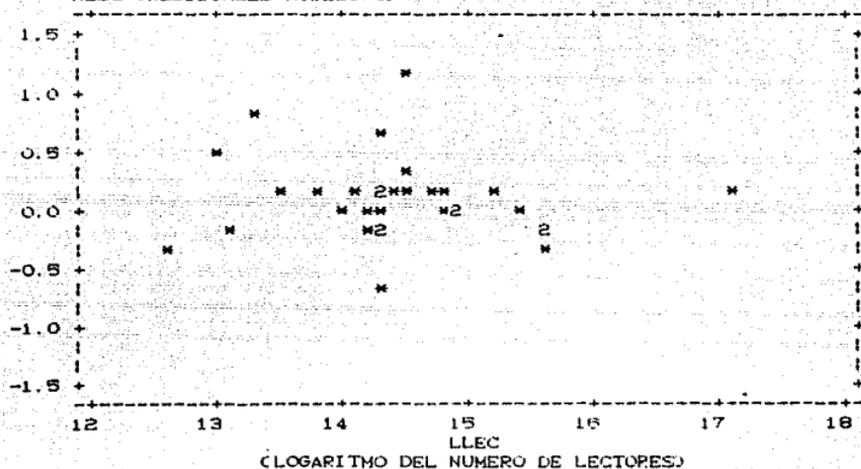


Gráfica 3.6. RESIDUALES DEL MODELO B VS. LOG. DEL NO. DE LECTORES.

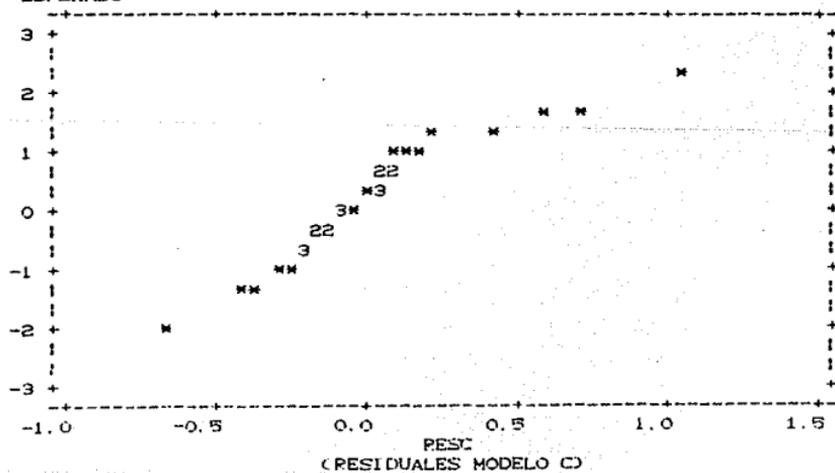




Gráfica 3.9. RESIDUALES DEL MODELO C VS. LOG. DEL NÚ. DE LECTOPES.  
 RESC (RESIDUALES MODELO C)



Gráfica 3.10. GRAFICA DE PAPEL NORMAL DE LOS RESIDUALES DEL MODELO C.  
 VALOR ESPERADO



### 3.3.2. Espectáculos públicos I

En México, dos de los medios de esparcimiento más populares son los espectáculos cinematográficos y los deportivos, ya que en comparación con otras formas de entretenimiento, por ejemplo el teatro comercial, en muchas ocasiones son económicamente más accesibles.

En el Cuadro 3.17 se observa el número promedio de localidades que vendió cada centro de espectáculos públicos del tipo cinematográfico y del tipo deportivo, por entidad federativa, durante 1988. Estos datos de alguna forma reflejan el grado de aceptación que tiene cada actividad en los diferentes puntos del país.

Según las estadísticas generales (Cuadro 3.18.), en promedio por estado se vendieron 117 localidades por cine y 90 por centro de espectáculo deportivo.

Zacatecas es el estado en el que se vendió el menor número de boletos por establecimiento para ambos tipos de espectáculo de toda la república, tan sólo 8 y 11, respectivamente.

Con 272 localidades vendidas por cine, a Baja California le corresponde el primer lugar entre todos los estados. La misma posición le corresponde al Distrito Federal, pero en espectáculos deportivos, con 324 localidades.

Al parecer existe mayor variabilidad en las cifras correspondientes a espectáculos deportivos que en las relacionadas con cinematográficos, a juzgar por los valores de la desviación estándar: 80.755 y 50.621, respectivamente.

Existen varios factores, tales como los climatológicos y geográficos, que podrían originar cierta variabilidad, de una región a otra, en la preferencia y en la práctica de las actividades deportivas. En cambio, parecería que la preferencia por espectáculos cinematográficos es más homogénea, dadas las características propias de este medio de esparcimiento.

Una forma para intentar probar la suposición anterior consistió en realizar una comparación de la variabilidad registrada de una entidad a otra entre el número promedio de localidades

ventas por uno y otro tipo de centro de espectáculo público. Previo a dicho análisis es necesario probar:

H<sub>0</sub>: La distribución del número promedio de localidades vendidas por cada centro de espectáculos deportivos (cinematográficos) por entidad federativa es Normal.

vs.

H<sub>1</sub>: La distribución no es Normal.

A continuación se muestra la máxima diferencia hallada entre la distribución empírica y la distribución Normal para cada muestra además del valor crítico de acuerdo con la prueba de Bondad de Ajuste de Lilliefors:<sup>7</sup>

TIPO DE ESTABLECIMIENTO	n	dn	dn(0.99) <sup>*</sup>
CINEMATOGRAFICO	27	0.176	0.195
DEPORTIVO	32	0.113	0.192

<sup>\*</sup>Ver tabla 15 en CONOVER, W.J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & sons, 1971, p. 398.

Se concluye que, con una probabilidad de error muy pequeña, puede suponerse que la distribución de la información que se posee, en ambos casos, es Normal, lo cual satisface el requisito para poder aplicar una prueba paramétrica, en este caso una comparación entre las varianzas de dos poblaciones, considerando que:

H<sub>0</sub>: Dentro de la República Mexicana, la variación existente de un estado a otro en el número promedio de localidades vendidas por establecimiento de espectáculos deportivos es similar a la que se presenta en el número promedio de localidades vendidas por cine.

vs.

H<sub>1</sub>: La variación es mayor en el caso de los espectáculos deportivos.

<sup>7</sup> Ver apéndice B, sección II.

Para construir la estadística de prueba  $f^6$  se requiere de la siguiente información:

TIPO DE ESTABLECIMIENTO	n, m	$\sum (X_i - \bar{X})^2$
CINEMATOGRAFICO	27	169559.953
DEPORTIVO	32	79436.469

De esta forma:

$$f = \frac{(27-1) 169559.953}{(32-1) 79436.469} = 1.7903,$$

y al ser comparada con

$$F_{(26,31)}(0.9) \cong 1.61,$$

resulta mayor, por lo que según la forma de la región de rechazo para la prueba de una cola, se concluye que existe evidencia para rechazar  $H_0$ , con un nivel de significancia  $\alpha=0.1$ .

Como se observó, la diferencia entre la estadística de prueba y el valor crítico, es muy pequeña. Si se tomara una  $\alpha$  más pequeña, digamos  $\alpha=0.05$ , se aceptaría la hipótesis de que las varianzas son semejantes, pero de esa forma aumentaría la probabilidad de error tipo II ( $\beta$ ), es decir de aceptar  $H_0$ , dado que es falsa, y lo que interesa es encontrar cualquier evidencia de que dadas las características particulares de cada tipo de espectáculo, la variabilidad de una entidad a otra en cuanto a la preferencia se acentúa más en los espectáculos deportivos.

Finalmente cabe comentar que la variabilidad en los datos de ambas muestras, en gran parte, se encuentra determinada por los tamaños de población y por el número de cines, estadios, etc. que operan en cada estado, presentando algún espectáculo de los considerados en esta sección.

<sup>6</sup> Ver apéndice A, fórmula A6.

Cuadro 3.17. NUMERO PROMEDIO DE LOCALIDADES VENDIDAS POR CENTRO DE ESPECTACULOS PUBLICOS CINEMATOGRAFICOS Y DEPORTIVOS, POR ENTIDAD FEDERATIVA, 1988.

Entidad Federativa	Localidades vendidas por centro de espectáculos	
	Cinematográficos	Deportivos
Aguascalientes	128	72
Baja California	272	252
Baja California Sur	108	18
Campeche	118	27
Coahuila	118	105
Colima	101	38
Chiapas	118	-
Chihuahua	148	128
Distrito Federal	172	324
Durango	79	-
Guanajuato	181	92
Guerrero	72	58
Hidalgo	94	132
Jalisco	127	159
México	184	73
Michoacán	56	12
Morolos	82	43
Nayarit	50	29
Nuevo León	135	222
Oaxaca	76	-
Puebla	187	123
Querétaro	128	83
Quitana Roo	112	-
San Luis Potosi	114	43
Sinaloa	60	86
Sonora	184	20
Tabasco	103	45
Tamaulipas	132	30
Tlaxcala	75	-
Veracruz	182	29
Yucatán	97	202
Zacatecas	8	11

Fuente: INSTITUIO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: Anuario Estadístico de los Estados Unidos Mexicanos, 1986-1989, México, INEGI, 1990, p. 466-467.

**Cuadro 3.18. ESTADISTICAS GENERALES DEL NUMERO PROMEDIO DE LOCALIDADES VENDIDAS POR CENTRO DE ESPECTACULOS PUBLICOS CINEMATOGRAFICOS Y DEPORTIVOS, POR ENTIDAD FEDERATIVA, 1968.**

Estadística	LOCALIDADES VENDIDAS POR CENTRO DE ESPECTACULOS	
	CINEMATOO.	DEPORTIVOS
NO. DE CASOS	32	27
VALOR MINIMO	8	11
VALOR MAXIMO	272	324
MEDIA	117.281	90.000
DESV. EST.	50.621	80.758

### 3.3.3. Espectáculos públicos II

Siendo los espectáculos teatrales y deportivos, medios de esparcimiento tan diferentes entre sí, principalmente desde el punto de vista cultural, sería factible pensar que, considerando cifras estatales, en el país no existe relación alguna en la participación de la población en uno y otro tipo de espectáculos. De alguna manera, la siguiente prueba estadística dará información al respecto, tomando en cuenta que:

Ho: En la República Mexicana, el número promedio de localidades vendidas por centro de espectáculos deportivos (NPLVD) no se encuentra correlacionado con el número promedio de localidades vendidas por centro de espectáculos teatrales (NPLVT).

vs.

H1: Ambas variables se encuentran correlacionadas positivamente. Es decir, las entidades donde se registra un NPLVD elevado también presentan un NPLVT alto, mientras que donde el NPLVD es bajo, el correspondiente NPLVT también lo es.

El Cuadro 3.19 muestra los datos que serán empleados. Puede observarse que no se cuenta con información de todas las entidades, pero sí de la mayoría de éstas (del 70% aproximadamente).

Según las estadísticas generales (Cuadro 3.20), el número promedio, a nivel nacional, de localidades vendidas es mayor en los espectáculos deportivos (88.273) y aunque el promedio en los teatros es menor (35.636) la variabilidad de un estado a otro no difiere mucho entre ambos medios de esparcimiento.

En dos de las entidades más importantes de la República Mexicana se registran los valores máximos de ambas variables: en Nuevo León para el teatro y en el Distrito Federal para los deportes. En Campeche se vendió el menor número de boletos por establecimiento para el teatro de todo el país, con únicamente 2 localidades; en cuanto a los deportes este número fue de 12 y se registró en el estado de Michoacán.

Para saber qué tipo de coeficiente de correlación es adecuado

aplicar, se debe probar primero:

H<sub>0</sub>: La variable NPLVD (NPLVT) por entidad federativa presenta la distribución Normal.

vs.

H<sub>1</sub>: La distribución del NPLVD (NPLVT) no es Normal.

Según la prueba de Lilliefors<sup>9</sup> y el siguiente cuadro:

TIPO DE ESTABLECIMIENTO	n	d <sub>n</sub>	d <sub>n</sub> (0.99) <sup>*</sup>
DEPORTIVO	22	0.167	0.219
TEATRAL	22	0.262	

<sup>\*</sup> Ver tabla 15 en CONOVER, W.J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & sons, 1971, p. 398.

se concluye que existe evidencia para rechazar el supuesto de distribución Normal en el caso de los centros teatrales (LVT) ya que la diferencia entre la distribución empírica y la teórica (d<sub>n</sub>) es lo suficientemente grande para rechazar H<sub>0</sub>, con un nivel de significancia  $\alpha=0.01$ . Se usa un  $\alpha$  tan pequeña pues se deseaba hallar cualquier evidencia para aceptar H<sub>0</sub>.

Por lo tanto, debe ser empleada una prueba no paramétrica.

Ya que se tienen observaciones apareadas correspondientes a dos variables medidas en una escala más que ordinal y se desea probar si existe un grado de asociación significativo entre ellas, es conveniente calcular el coeficiente de correlación  $r$  de Kendall,<sup>10</sup> para lo cual se requiere de la información contenida en el Cuadro 3.21.

Como se sabe, para calcular  $r$  se asignan rangos por separado a las observaciones de cada variable; después, ambos conjuntos de rangos se ordenan tomando en cuenta el orden natural de una de las variables, en este caso NPLVD. A continuación se asigna el valor de 1 a cada par de observaciones del NPLVT que se encuentren en el orden natural y -1 a cada par que no se encuentre en ese orden.

El valor de S se calcula sumando el conjunto de valores

<sup>9</sup> Ver apéndice B, sección II.

<sup>10</sup> Ver apéndice E, sección III-2.

asignados a cada par de observaciones de LVT que se obtuvo y de esta forma:

$$S = 98,$$

siendo positiva ya que predominaron las parejas de observaciones del NPLVT que se encontraban en el orden natural.

De esta forma se obtiene que:

$$r = 0.419,$$

valor obtenido mediante la fórmula corregida, por presentarse observaciones ligadas.

Para determinar la significancia de  $r$ , considerando que  $N=22$ , se calcula  $z$ , que resulta tomar el valor de:

$$z = 2.73,$$

que tiene una probabilidad de ocurrencia conforme a  $H_0$ , de 0.0032, según tablas de la distribución Normal, por lo cual siendo esta probabilidad menor que cualquier valor usual de  $\alpha$ , se concluye que existe evidencia para rechazar  $H_0$ , hasta con un nivel de significancia de 0.0032, para la prueba de una cola.

En otras palabras, existe evidencia de que, contrario a lo que se supuso inicialmente, en las entidades donde existe una importante asistencia del público a espectáculos deportivos también se vende un buen número de boletos para el teatro y por otro lado, donde uno de los dos medios de esparcimiento citados no es muy popular, tampoco lo es el otro.

Definitivamente existen otras variables que al ser analizadas brindaran conclusiones más certeras en relación con el cuestionamiento planteado, por ejemplo el total de localidades vendidas en cierto periodo referidas al tamaño de población de la entidad.

Cuadro 3.19. *NUMERO PROMEDIO DE LOCALIDADES VENDIDAS POR CENTRO DE ESPECTACULOS PUBLICOS DEPORTIVOS Y TEATRALES, POR ENTIDAD FEDERATIVA, 1988.*

Entidad Federativa	Localidades vendidas por centro de espectáculo	
	Deportivos	Teatrales
Aguascalientes	72	36
Campeche	27	2
Coahuila	105	12
Colima	36	8
Chihuahua	126	16
Distrito Federal	324	72
Guanajuato	92	14
Guerrero	58	17
Jalisco	159	38
México	73	18
Michoacán	12	19
Nayarit	29	12
Nuevo León	222	299
Puebla	123	50
Querétaro	63	8
San Luis Potosí	43	9
Sinaloa	86	60
Sonora	20	13
Tamaulipas	30	3
Veracruz	29	15
Yucatán	202	45
Zacatecas	11	20

Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *Anuario Estadístico de los Estados Unidos Mexicanos, 1988-1989*, México, INEGI, 1990, p. 466-467.

Cuadro 3.20. ESTADISTICAS GENERALES DEL NUMERO PROMEDIO DE LOCALIDADES VENDIDAS POR CENTRO DE ESPECTACULOS PUBLICOS DEPORTIVOS Y TEATRALES, POR ENTIDAD FEDERATIVA, 1988.

Estadística	LOCALIDADES VENDIDAS POR CENTRO DE ESPECTACULOS	
	DEPORTIVOS	TEATRALES
NO. DE CASOS	22	22
V. MIN.	12	2
V. MAX.	324	299
MEDIA	88.273	35.638
DESV. EST.	79.215	61.757

Cuadro 3.21. RANGOS DE LAS VARIABLES NPLVD Y NPLVT, ORDENADOS DE ACUERDO AL ORDEN NATURAL DE NPLVD Y VALOR DE LA ASIGNACION DADA AL ORDEN DE LAS PAREJAS DE OBSERVACIONES DE NPLVT.

Entidad Federativa	Rango		Asignación	
	LVD	LVT	(+)	(-)
Zacatecas	1	15	7	14
Michoacan	2	14	7	13
Sonora	3	9	12	7
Campeche	4	1	18	
Nayarit	5.5	6.5	12	4
Veracruz	5.5	10	10	6
Tamaulipas	7	2	15	
Colima	8	3.5	13	
San Luis Potosi	9	5	12	1
Guerrero	10	12	8	4
Queretaro	11	3.5	11	
Aguascalientes	12	16.5	5	4
Mexico	13	13	6	3
Sinaloa	14	20	2	6
Guanajuato	15	9	6	1
Coahuila	16	6.5	6	
Puebla	17	19	2	3
Chihuahua	18	11	4	
Jalisco	19	16.5	3	
Yucatan	20	18	2	
Nuevo Leon	21	22	1	
Distrito Federal	22	21		
			$\Sigma$	66

### 3.3.4. Museos y zonas arqueológicas

México es un país que cuenta con muy diversos atractivos turísticos y culturales como son los museos y zonas arqueológicas localizados en los distintos estados, donde cada año se da una importante afluencia de visitantes nacionales y extranjeros.

El Cuadro 3.22 contiene el porcentaje de visitantes de origen extranjero de zonas arqueológicas y museos, por entidad federativa, registrado durante 1988. Cabe aclarar que se carece de la información correspondiente a varios estados.

Podría resultar interesante investigar si estos dos tipos de atractivos turístico-culturales presentan un porcentaje de visitantes extranjeros similares. Es decir, se intentará probar estadísticamente lo siguiente:

H<sub>0</sub>: Los porcentajes, por entidad federativa, de visitantes extranjeros de zonas arqueológicas y de museos son similares.

vs.

H<sub>1</sub>: Los porcentajes son diferentes.

Analizando las estadísticas generales (Cuadro 3.23), se puede ver que los porcentajes promedio de visitantes extranjeros de zonas arqueológicas y de museos son muy similares, 22.5% y 22.4% respectivamente. Tampoco existe mucha diferencia entre las desviaciones estándar.

Dos estados de la península de Yucatán presentan el mayor porcentaje de visitantes extranjeros: Yucatán con 50 de cada 100 visitantes en sus zonas arqueológicas y Quintana Roo con 83 de cada cien en sus museos. En los estados de Guerrero y Tamaulipas, ninguno de los visitantes de zonas arqueológicas y museos son de origen extranjero.

Primeramente, con el fin de validar el supuesto de distribución Normal para efectuar la prueba de hipótesis correspondiente, se probará lo siguiente:

H<sub>0</sub>: Dentro de la República Mexicana, el porcentaje, por entidad federativa, de visitantes de nacionalidad extranjera de zonas arqueológicas (museos) presenta la distribución Normal.

vs.

H<sub>1</sub>: La distribución de este porcentaje no es Normal.

Observando el siguiente cuadro:

N DE VISITANTES EXTRANJEROS	n	dn	dn(0.99) <sup>*</sup>
ZONAS ARG.	17	0.198	0.245
MUSEOS	26	0.172	0.197

\*Ver tabla 15 en CONOVER, W. J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & sons, 1971, p. 398.

y de acuerdo a la prueba de Lilliefors<sup>11</sup> se indica que, con una probabilidad de error muy pequeña, puede considerarse que la distribución de los datos es Normal en ambos casos; con un nivel de significancia pequeño, es decir  $\alpha=0.01$ , pues con el fin de aplicar una prueba paramétrica es importante hallar cualquier evidencia de que la distribución de los datos es Normal.

De esta forma, se realizará una comparación de medias. Para lo cual, al construir la estadística de prueba<sup>12</sup> se considera que:

N DE VISITANTES EXTRANJEROS	m, n	$\bar{X}$	$\sum_{i=1}^{m,n} (X_i - \bar{X})^2$
ZONAS ARG.	17	22.554	4922.559
MUSEOS	26	22.386	9722.421

por lo cual:

$$t = 0.0295,$$

que, por mucho, resulta ser menor comparada con el valor crítico

<sup>11</sup> Ver apéndice B, sección II.

<sup>12</sup> Ver apéndice A, fórmula A1.

correspondiente:

$$t_{(17+26-2)}(0.95) = t_{(41)}(0.95) \cong 1.684.$$

Por lo tanto, con un nivel de confianza del 90%, se concluye que existe evidencia para suponer que, durante 1988, el porcentaje de visitantes extranjeros que registraron las zonas arqueológicas es similar al registrado por los museos de la República Mexicana.

Cuadro 3.22. PORCENTAJE DE VISITANTES EXTRANJEROS DE ZONAS ARQUEOLÓGICAS Y MUSEOS POR ENTIDAD FEDERATIVA, 1988.

Entidad Federativa	Porcentaje de visitantes extranjeros	
	Zonas Arqueológicas	Museos
Aguascalientes		16.50
Baja California Sur		53.02
Campeche	47.92	28.85
Coahuila		3.37
Colima		13.19
Chiapas	30.33	18.47
Chihuahua	9.03	19.31
Distrito Federal	42.02	22.28
Guanajuato		8.08
Guerrero	0.16	26.05
Hidalgo	17.08	1.25
Jalisco		39.47
México	9.87	28.78
Michoacán	3.99	22.24
Morelia	17.88	17.38
Nayarit	28.10	10.35
Nuevo León		28.85
Oaxaca	40.59	37.45
Puebla	10.24	10.41
Queretaro		3.78
Quintana Roo	49.41	82.98
San Luis Potosí		21.44
Tabasco	3.44	22.78
Tamaulipas		0.13
Tlaxcala	12.30	
Veracruz	11.09	4.24
Yucatán	50.21	49.98

Fuente: INSTITUTO NACIONAL DE ESTADÍSTICA GEOGRAFÍA E INFORMÁTICA: Anuario Estadístico de los Estados Unidos Mexicanos, 1988-1989. México, INEGI, 1990, p. 477-481.

**Cuadro 3.23. ESTADÍSTICAS GENERALES DEL PORCENTAJE DE VISITANTES  
EXTRANJEROS DE ZONAS ARQUEOLÓGICAS Y MUSEOS POR ENTIDAD  
FEDERATIVA, 1988.**

Estadística	% DE VISITANTES EXTRANJEROS	
	ZONAS ARG.	MUSEOS
NO. DE CASOS	17	26
V. MIN.	0.180	0.130
V. MAX.	50.210	82.980
MEDIA	22.554	22.388
DESV. EST.	17.540	18.879

### 3.4. SEGURIDAD Y ORDEN PUBLICO

#### 3.4.1. Accidentes de tránsito terrestre

Las consecuencias de un accidente de tránsito están determinadas por una gran cantidad de factores, entre ellos, la forma en que este sucedió. Aunque podría considerarse que las características del lugar donde ocurrió dicho accidente también son importantes.

Por ejemplo, un determinado choque ocurrido en la carretera pudo haber causado una o más muertes, mientras que otro sucedido en cierta calle de alguna ciudad, pudo haber causado solo daños materiales. O bien, una persona que ha sido atropellada y gravemente herida, puede salvar la vida, si se le traslada rápidamente a un hospital.

En este país, el tipo o el estado de carreteras, caminos, o calles, así como la disponibilidad de servicios médicos y auxiliares de urgencia, en el caso de algún accidente de tránsito terrestre, son características que posiblemente originen alguna relación entre las consecuencias que presenta un accidente de tránsito terrestre y el tipo de zona, ya sea rural o urbana, en que este sucedió.

La suposición anterior puede ser probada estadísticamente mediante la siguiente prueba de independencia<sup>1</sup>:

H<sub>0</sub>: Las consecuencias de un accidente de tránsito terrestre no dependen de la zona, urbana o rural, en que este se sucedió.

vs.

H<sub>1</sub>: Las consecuencias y la zona sí dependen.

Según el Cuadro 3.24, durante 1988, en total se registraron más de 200,000 accidentes terrestres en el país, de los cuales, como es natural dada la densidad del flujo vehicular, cerca de dos terceras partes (151,852) ocurrieron en la zona urbana y el tercio restante en la rural (52,004).

Por otra parte, la gran mayoría de los accidentes (150,637)

---

<sup>1</sup> Ver apéndice C, sección I.

ocasionaron solo daños materiales, menos de una cuarta parte (46972) se consideran como no fatales, es decir que causaron heridas graves o leves a una o más personas sin ocurrencia de muerte, y 6047 como fatales, ocasionando la muerte de por lo menos una persona.

También se observa que el número de accidentes fatales es muy similar en ambas zonas, aunque las cifras relativas son muy diferentes; en el caso de los no fatales se guarda una proporción similar a la observada en los totales entre las cifras de las dos áreas, lo cual también se presenta en los datos correspondientes a accidentes fatales.

Según los valores esperados conforme a  $H_0$  (Cuadro 3.25) el número de accidentes de consecuencias fatales debió haber sido menor en la zona rural y mayor en la urbana. Las diferencias cuadradas entre los valores observados y los esperados conforme a  $H_0$ , estandarizadas por estos últimos (ver Cuadro 3.26), indican que en los casos donde al parecer existen evidencias de que las consecuencias dependen de la zona, es en los accidentes que ocasionaron por lo menos una muerte.

La estadística de prueba<sup>2</sup> toma el valor de:

$$\chi^2 = 2139.047,$$

valor que, por mucho, es superior a  $\chi^2_{(1)(2)(72)}(1-\alpha)$ , para cualquier nivel de significancia usual.

Dado que se tiene un tamaño de muestra muy grande (203,658) se corre el riesgo de que el valor tan elevado que se obtuvo de  $\chi^2$ , no se deba a que no existe independencia, sino al tamaño de muestra. Para comprobarlo se calcula el coeficiente de contingencia  $C^3$  y resulta que:

$$C = 0.0104,$$

lo cual indica que al parecer el tipo de accidente de tránsito no depende de la zona en la cual ocurrió.

Sin embargo, cabe comentar que en la zona rural existe una alta incidencia de accidentes fatales, en comparación con la ciudad, lo cual, como ya se comentó, puede deberse a las diferencias entre ambas zonas en cuanto a características tales

<sup>2</sup>Ver apéndice C, fórmula C1.

<sup>3</sup>Ver apéndice E, sección, III-3.

como el tipo o el estado de carreteras, caminos, o calles y la disponibilidad de servicios medicos y auxiliares de urgencia, en el caso de algun accidente de tránsito terrestre.

Cuadro 3.24. ACCIDENTES DE TRANSITO TERRESTRE, SEGUN ZONA Y TIPO DE ACCIDENTE, 1988.

TIPO ZONA	FATAL	NO FATAL	SOLO DAÑOS	TOTAL POR ZONA
URBANA	3012 (0.02)	36311 (0.24)	112329 (0.74)	151652
RURAL	3035 (0.08)	10861 (0.21)	38308 (0.74)	52004
TOTAL POR TIPO	6047 (0.03)	46972 (0.23)	150637 (0.74)	203656

Fuente: INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: Anuario Estadístico de los Estados Unidos Mexicanos, 1988-1989, México, INEGI, 1990, p. 540.

Cuadro 3.25. VALORES ESPERADOS CONFORME A Ho.

$E_{ij}$	FATAL	NO FATAL	SOLO DAÑOS	$\Sigma_j$
URBANO	4502.865 (0.03)	34977.598 (0.23)	112171.516 (0.74)	151652
RURAL	1544.115 (0.03)	11994.402 (0.23)	38465.484 (0.74)	52004
$\Sigma_i$	6047 (0.03)	46972 (0.23)	150637 (0.74)	203656

Cuadro 3.26. DIFERENCIAS CUADRADAS ENTRE LOS VALORES OBSERVADOS Y LOS ESPERADOS CONFORME A  $H_0$ , ESTANDARIZADAS POR ESTOS ULTIMOS.

$\frac{CE_{ij} - O_{ij}}{E_{ij}}^2$	FATAL	NO FATAL	SOLO DAÑOS	$\Sigma_j$
URBANA	493.626	50.831	0.221	544.678
RURAL	1439.491	148.233	0.645	1588.369
$\Sigma_i$	1933.117	199.064	0.866	2133.047

## CAPITULO 4

## CAPITULO 4. APLICACIONES EN AREAS DIVERSAS

### 4.1. CONTAMINACION

#### 4.1.1. Monóxido de Carbono y el programa "Un día sin auto"<sup>1</sup>

Uno de los problemas más serios que aquejan a la Ciudad de México es, sin lugar a dudas, la contaminación ambiental, en particular la del aire. Dicho fenómeno puede entenderse como "la adición de cualquier sustancia que altere las propiedades físicas y químicas del aire".

Ciertamente existen fuentes naturales de contaminación tales como las erupciones volcánicas, las tolvaneras y los incendios forestales, entre otras; pero el hombre, a través de algunas de las actividades que realiza, origina las fuentes antropogénicas de contaminación.

Estas últimas se clasifican en fuentes fijas o de actividad industrial y fuentes móviles (vehículos automotores). Específicamente, procesos industriales como la fundición, la producción de cemento, asbesto y sustancias químicas y la refinación petrolera son capaces de provocar el deterioro del aire. Sin embargo la combustión, empleada para generar calor, energía eléctrica o movimiento, es el proceso más significativo en cuanto a emisión de contaminantes.

Los principales contaminantes atmosféricos son el monóxido de carbono (CO), los óxidos de nitrógeno (NO<sub>x</sub>), los hidrocarburos (HC), los óxidos de azufre (SO<sub>x</sub>), los oxidantes fotoquímicos y partículas que pueden tener muy diversos componentes como silicatos, sulfatos o metales pesados.

Anualmente las emisiones de contaminantes en el país son superiores a los 16 millones de toneladas, de los cuales 35% proviene de fuentes industriales y 65% son de origen vehicular.

<sup>1</sup>MEXICO, SECRETARIA DE DESARROLLO URBANO Y ECOLOGIA: *Informe sobre contaminación*, Mexico, SEDUE, 1990, p. 33-37.

En la Ciudad de México se genera el 23.6% de dichas emisiones, en Guadalajara el 3.5% y en Monterrey el 3%; los otros centros industriales del país generan el 70% restante.

El gobierno ha tomado varias medidas con el fin de abatir los niveles de contaminación principalmente en el área metropolitana de la Ciudad de México (AMCM). Entre estas medidas destaca la implementación del programa "Un día sin auto" a partir del 20 de noviembre de 1989 en el AMCM. Dicho programa consiste en dejar fuera de circulación por un día a la semana a cada uno de los vehículos automotores privados que transiten el área, con vigencia del día lunes al viernes. El día que se debe evitar la circulación depende de la terminación de las placas del vehículo.

Como ya se mencionó, el monóxido de carbono (CO) es uno de los principales contaminantes del aire. Se trata de un gas incoloro e inodoro originado a partir de la combustión incompleta de combustibles y otras sustancias que contienen carbono, así como de incendios. Sus efectos principales en la salud del hombre y de los animales consisten en que al combinarse con la hemoglobina de la sangre se forma una sustancia denominada carboxihemoglobina, la cual afecta al sistema nervioso central y provoca cambios funcionales cardíacos y pulmonares, dolor de cabeza, fatiga, somnolencia, afecciones respiratorias y, en casos extremos, la muerte.

En la presente sección se realizará un breve análisis en torno a este último contaminante. Se intentará conocer que efecto ha tenido el programa de "Un día sin auto" sobre el nivel de CO en el aire.

La hipótesis que interesa probar es:

H<sub>0</sub>: El nivel de CO en el aire no se vio alterado por la presencia del programa "Un día sin auto".

vs.

H<sub>1</sub>: A partir de la implementación de dicho programa, el nivel de CO en el aire disminuyó.

Se cuenta con observaciones que cubren el siguiente periodo: del 10. de enero de 1988 al 28 de febrero de 1990, excluyendo fines de semana y días oficialmente feriados en los que

generalmente no opera el programa.

En las gráficas 4.1 y 4.2 se muestra el comportamiento del nivel de CO en el aire durante el periodo especificado.

El Cuadro 4.1 contiene las estadísticas correspondientes a las observaciones registradas antes de que entrara en vigor el programa y las que corresponden al periodo posterior a su implementación. Se observa que durante el primer periodo existe una mayor variabilidad (3.18) que en el segundo (2.79), es notable que antes de "Un día sin auto" se registró el índice más alto de CO (34.00) y el más bajo (1.80). Por otro lado, cabe destacar la similitud entre los valores promedio.

La primera prueba que al parecer es susceptible de ser utilizada es una comparación de medias en base a la distribución Normal, dado que se cuenta con un número considerable de observaciones. Para validar el uso de esta prueba es necesario verificar si, para cada periodo de interés, se cumple el supuesto distribucional correspondiente:

$H_0$ : El nivel de CO en el aire presenta una distribución Normal.

vs.

$H_1$ : La distribución del nivel de CO en el aire no es Normal.

Las gráficas de papel Normal<sup>2</sup> correspondientes (gráficas 4.3 y 4.4) señalan que existe alguna evidencia para suponer que cada variable presenta una distribución Normal cuyos parámetros desconocidos. Esto se concluye después de observar que, en ambos casos, puede trazarse una línea recta que abarque a la mayoría de los puntos.

De esta forma se procederá a construir la estadística de prueba<sup>3</sup> considerando la siguiente información:

---

<sup>2</sup> Ver SEBER, G.A.F.: *Linear Regression Analysis*, E.U.A., John Wiley & sons, 1977, p.395.

<sup>3</sup> Ver apéndice A, fórmula A1.

	"Un día sin auto"	
	Antes	Después
m, n	483	69
$\bar{X}_t$	8.974	9.313
$\sum_j (X_j - \bar{X}_t)^2$	12 808.969	527.918

De tal manera que  $t$  toma el valor de:

$$t = -0.5390,$$

y de acuerdo a la región de rechazo, en este caso:

$$\bar{X} = ( \bar{X} \mid t > t_{(500)}(0.90) ),$$

donde  $t = -s$ ,  $s = 0.5390$ , se obtiene que:

$$\bar{X} = ( \bar{X} \mid s < -t_{(500)}(0.90) )$$

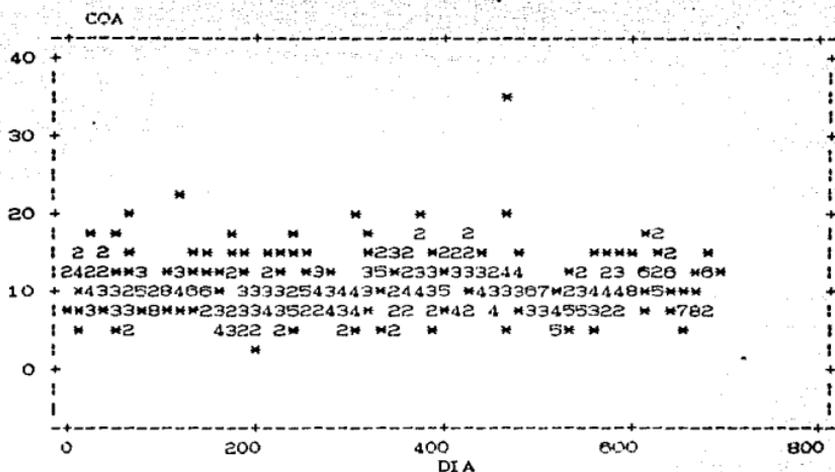
y como  $t_{(500)}(0.90) = 1.282$ , se concluye que existe evidencia para considerar, con una probabilidad de error muy pequeña, que los niveles de CO en el aire son similares en ambos periodos.

Aparentemente, el nivel de monóxido de carbono se ha mantenido sin alteración aun con la presencia del programa de "Un día sin auto". Sin embargo, con este resultado, de ninguna forma se sugiere que el programa no ha cumplido con su objetivo de abatir los niveles de contaminación, pues para emitir conclusiones de mayor alcance se requeriría efectuar un análisis bastante más complejo en el que intervinieran otras variables más, de las cuales se contara con información actualizada y haciendo uso de técnicas estadísticas más avanzadas como es el caso de las Series de Tiempo.

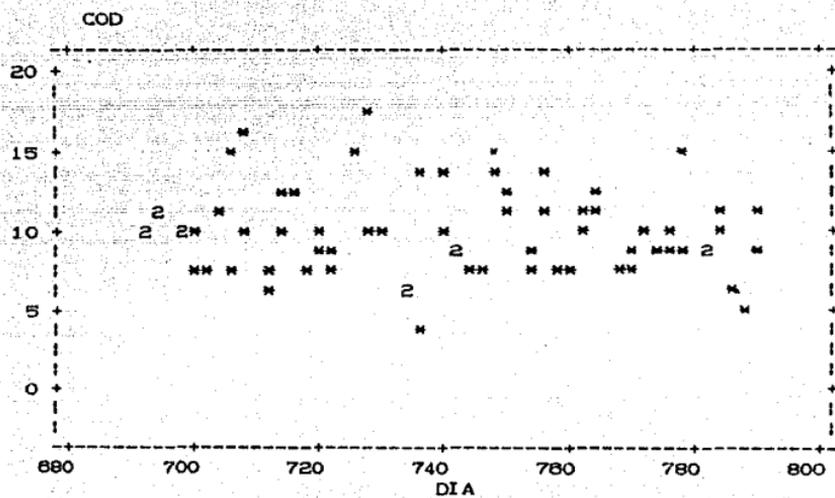
Cuadro 4.1. ESTADISTICAS GENERALES DE LOS NIVELES DE CO EN EL AIRE ANTES Y DESPUES DE HABER SIDO IMPLANTADO EL PROGRAMA "Un dia sin auto". (10. de enero de 1988 - 28 de febrero de 1990)

	"Un dia sin auto"	
	Antes	Despues
n	483	69
V. MIN	1.80	3.50
V. MAX	34.00	17.20
MEDIA	8.97	9.31
DESV. EST.	3.18	2.79

Gráfica 4.1. NIVEL DIARIO DE CO EN EL AIRE DEL 1-ENERO-1988 AL 19-NOV-1989.

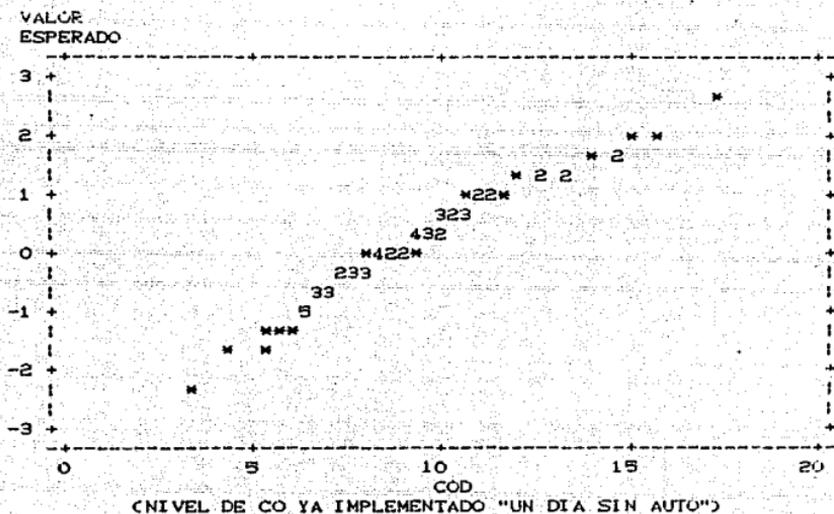


Gráfica 4.2. NIVEL DIARIO DE CO EN EL AIRE DEL 20-NOV-1988 AL 28-FEB-1990.





Grafica 4.4. GRAFICA DE PAPEL NOPHAL DEL NIVEL DIARIO DE CO EN EL AIRE DEL 20-NOV-1988 AL 28-FEB-1990.



## 4.2. ECOLOGIA

### 4.2.1. El comportamiento reproductivo de una palma tropical<sup>1</sup>

La variabilidad en el crecimiento y los patrones reproductivos de las especies tropicales es tema de gran interés para los estudiosos de la Ecología.

Un estudio reciente llevado a cabo en la selva de Los Tuxtlas, en el estado de Veracruz, sobre la variabilidad espacial, temporal e individual en la supervivencia, tasas de crecimiento y comportamiento reproductivo de una especie de palma denominada científicamente *Chamaedorea tepejilote*, arrojo valiosa información. En esta sección solo se abordará el aspecto reproductivo.

*C. tepejilote* es una palma que cuenta con un tronco único cuya altura generalmente rebasa los 5 m., florece entre octubre y diciembre y se ha sugerido que la polinización se realiza gracias al viento. En julio de 1981 todos los individuos de dicha especie localizados en el lugar de estudio fueron marcados, se delimitaron tres zonas, donde cada una de estas presenta ciertas condiciones naturales específicas, quedando 324 individuos dentro de la zona 1, 341 en la 2 y tan solo 145 en la 3. La observación sobre los aspectos de interés para la investigación se extendió de julio de 1981 hasta julio de 1985, sin el registro correspondiente a 1983.

Por medio de una prueba de independencia los investigadores obtuvieron que el número de veces que la palma atravesó por el proceso reproductivo, durante los 4 años observados no depende de su sexo. Sin embargo, existen 16 posibles secuencias reproductivas diferentes; por ejemplo, se encontró que 21 plantas masculinas y 22 femeninas se reprodujeron en 2 de los 4 años considerados, pero algunas lo hicieron en 2 años consecutivos, al principio, en medio o al final del periodo y otras dejaron de reproducirse por uno o 2

---

<sup>1</sup> OYAMA, K.: "Variation in growth and reproduction in the Neotropical dioecious palm *Chamaedorea tepejilote*", en: *The Journal of Ecology*, British Ecological Society, Londres, sept., 1990, vol. 78, no. 3, p. 648-663.

años seguidos.

De esta manera, resultaría interesante investigar si para cada uno de los tres grupos factibles, esto es las plantas que se reprodujeron 1, 2 o 3 veces respectivamente, la secuencia reproductiva depende del sexo, en otras palabras se intentara probar:

H<sub>0</sub>: Dentro del grupo de las plantas de *C. tepalote* que se reprodujeron en  $k$  de los 4 años observados, la secuencia reproductiva no depende del sexo.

vs.

H<sub>1</sub>: Dichas características se encuentran relacionadas.

$$k = \overline{1,3}.$$

Según el Cuadro 4.5, de las 40 plantas que se reprodujeron sólo una vez durante todo el periodo examinado, predominaron las del sexo femenino. Por otra parte, el primer año de observación (1981) fue en el cual se registró un mayor número de reproducciones.

Las cifras contenidas en los cuadros 4.6 y 4.7, revelan que no existe gran diferencia entre lo observado y lo esperado conforme a H<sub>0</sub> y por tanto<sup>2</sup>:

$$X^2 = 4.5848,$$

resultando menor que  $\chi^2_{9\%}(1-\alpha) = \chi^2_9(1-\alpha)$ , para cualquier nivel de significancia usual, lo cual indica que existe evidencia para pensar que no hay relación entre la secuencia reproductiva y el sexo.

Entre las plantas que se reprodujeron en dos ocasiones (un total de 43) casi la mitad son masculinas y, considerando ambos sexos, la gran mayoría de ellas (23) lo hizo en los primeros dos periodos. En este caso, las diferencias entre lo que se observó y lo que se esperaba de acuerdo al supuesto de independencia, son menores aun, dado que:

$$X^2 = 2.1166,$$

que al ser comparada con  $\chi^2_{4\%}(1-\alpha) = \chi^2_4(1-\alpha)$  es más pequeña, por lo que la conclusión es la misma que en el caso anterior.

Para el grupo de los 38 individuos que dejaron de

---

Ver apéndice C, fórmula C1.

reproducirse solo durante un periodo, se observa que casi dos terceras partes corresponden al sexo masculino. Tambien se encontro que  $\chi^2$  no es significativa al obtenerse que su valor es de

$$\chi^2 = 1.4175,$$

por lo tanto, con una probabilidad de error muy pequena, se acepta  $H_0$ , como en los casos anteriores.

En suma, se concluye que ademas de ser el numero de reproducciones durante el periodo de estudio independiente del sexo de la planta, no hay una secuencia reproductiva particular que se manifieste predominantemente en plantas de algun sexo. Es decir, la secuencia reproductiva que se presenta en la palma *C. Tepejilote* no depende estadisticamente de su sexo en los tres casos considerados. Se hace hincapié en que se trata de independencia estadistica, ya que faltaria considerar la opinion de un experto en la materia.

Tambien se observó que mientras mayor es el numero de reproducciones la diferencia entre lo observado y lo esperado conforme a  $H_0$  es más pequeña.

Cuadro 4.5 . FRECUENCIA DE QUINCE<sup>a</sup> POSIBLES SECUENCIAS REPRODUCTIVAS PARA PLANTAS MASCULINAS Y FEMENINAS DE *Chamaedorea Tepezilote* EN LOS TUXTLAS DUFANTE CUATRO AÑOS (1981-1985, SIN REGISTRO PARA 1983): 0 = NO REPRODUCCION; 1 = REPRODUCCION; - = SIN INFORMACION (PARA 1983). DATOS CORRESPONDIENTES A LAS PALMAS SOBREVIVIENTES EN LAS ZONAS 1 Y 2

SECUENCIA DE RE- PRODUCCION ANUAL	FRECUENCIA		TOTAL
	SEXO MASCULINO	SEXO FEMENINO	
Reproductiva en 1 de 4 años			
10_00	3	11	14
01_00	6	4	10
00_10	7	6	13
00_01	1	2	3
TOTAL	17	23	40
Reproductiva en 2 de 4 años			
11-00	11	12	23
10-10	2	1	3
10-01	2	5	7
01-01	2	1	3
00-11	4	3	7
TOTAL	21	22	43
Reproductiva en 3 de 4 años			
11-10	4	2	6
11-01	12	7	19
10-11	3	1	4
01-11	3	4	7
TOTAL	22	14	36

Fuente: *Journal of Ecology*, British Ecological Society, Londres, Vol. 78, No. 3, 1990, p. 658.

<sup>a</sup> Son un total de 16 posibles secuencias reproductivas, pero el caso de reproducción en 4 de 4 años es único por lo que no es aplicable una prueba de independencia.

Cuadro 4.8 . VALORES ESPERADOS CONFORHE A Ho.

E <sub>ij</sub>	MASCULINO	FEMENINO	
			$\Sigma$ j
10-00	5.9500	8.0500	14
01-00	4.2500	5.7500	10
00-10	5.9250	7.4750	13
00-01	1.2750	1.7250	3
$\Sigma$ i	17	23	40
			$\Sigma$ j
11-00	11.2328	11.7674	23
10-10	1.4651	1.5349	3
10-01	3.4188	3.5814	7
01-01	1.4651	1.5349	3
00-11	3.4188	3.5814	7
$\Sigma$ i	21	22	43
			$\Sigma$ j
11-10	3.6687	2.3333	6
11-01	11.6111	7.3889	19
10-11	2.4444	1.5556	4
01-11	4.2778	2.7222	7
$\Sigma$ i	22	14	36

Cuadro 4.7 . DIFERENCIA AL CUADRADO ENTRE LO OBSERVADO Y LO ESPERADO CONFORME A Ho. ESTANDARIZADA POR LO ESPERADO.

$\frac{CO_{ij} - E_{ij}}{E_{ij}}^2$	MASCULINO	FEMENINO	
$E_{ij}$			$\Sigma_j$
10_00	1.4626	1.0811	2.5437
01_00	0.7206	0.5326	1.2532
00_10	0.3938	0.2911	0.6849
00_01	0.0593	0.0438	0.1032
$\Sigma_i$	2.6363	1.9486	$\chi^2_1 = 4.5848$
			$\Sigma_j$
11_00	0.0048	0.0048	0.0094
10_10	0.1953	0.1864	0.3817
10_01	0.5887	0.5619	1.1506
01_01	0.1953	0.1864	0.3817
00_11	0.0989	0.0944	0.1933
$\Sigma_i$	1.0829	1.0337	$\chi^2_2 = 2.1168$
			$\Sigma_j$
11_10	0.0303	0.0476	0.0779
11_01	0.0130	0.0205	0.0335
10_11	0.1263	0.1984	0.3247
01_11	0.3817	0.5998	0.9814
$\Sigma_i$	0.5513	0.8663	$\chi^2_3 = 1.4175$

#### 4.2.2. Dos especies maderables en el desierto<sup>4</sup>

En algunas áreas desérticas del mundo se ha observado la presencia de arcos de vegetación alternando con zonas áridas. Se cree que estos patrones se deben a una distribución diferenciada del agua proveniente de las lluvias, donde las zonas de escasa o nula vegetación actúan como almacenadoras del agua que será evaporada por los arcos.

A las fronteras entre arcos de vegetación y zonas áridas se las denomina *ecotonos* (cada arco posee un *ecotono* pendiente arriba y otro pendiente abajo). Se ha inferido la migración de las especies vegetales pendiente arriba de los arcos, sin embargo, no existe información sobre la dinámica de las especies maderables.

Un estudio realizado en la región del Bolson de Mapimí, dentro del desierto del estado de Chihuahua, donde el patrón de arcos de vegetación cubre aproximadamente el 32% del territorio, tuvo la finalidad de verificar la hipótesis de migración de arco, investigando la distribución de plantas jóvenes y plantas adultas de dos especies maderables y su cobertura dentro de los arcos.

Las dos especies maderables consideradas fueron las más comunes dentro del Bolson de Mapimí: *Prosopis glandulosa* y *Flourensia cernua*.

El Área de estudio fue una banda de 600mx2m, perpendicular a los arcos, donde se realizó un gran número de mediciones. Una de ellas fue el número de individuos adultos<sup>5</sup> de ambas especies a lo largo del trayecto clasificados por su ubicación (arco de vegetación o zona árida) y por su condición (vivo o muerto), obteniéndose la información del Cuadro 4.8.

*P. glandulosa* fue más abundante que *F. cernua*, particularmente en la zona árida, donde, de esta última especie, sólo se encontraron 8 plantas y además muertas.

MONTANA C., LOPEZ-PORTILLO J. y MAUCHAMP A.: "The response of two woody species to the conditions created by a shifting ecotone in an arid ecosystem", en: *The Journal of Ecology*, British Ecological Society, Londres, sept., 1990, vol. 78, no. 3, p. 648-663.

<sup>5</sup> Se considera como adulto al individuo cuya altura es de por lo menos 50cm en el caso de *F. cernua* y de 1m para *F. glandulosa*.

Se observa una clara relación entre la condición y la ubicación para ambas especies, ya que los arcos de vegetación se caracterizan por la presencia de individuos con vida, mientras que en la zona seca la mayoría de las plantas que se encontraron, estaban muertas, la totalidad de ellas, en el caso de *F. cernua*.

Estos datos fueron analizados y tuvieron su importancia dentro del estudio. En la presente sección se efectuará una prueba diferente que puede resultar interesante. Se intentará probar:

H<sub>0</sub>: Dentro de una región cuyo patrón de vegetación es de arcos, considerando las dos especies maderables más abundantes en ella, la condición y la ubicación no dependen de la especie.

vs.

H<sub>1</sub>: Bajo el patrón de arcos de vegetación, la condición y la ubicación de las plantas dependen de la especie.

Por lo tanto, se incurre en el uso de una tabla de contingencia de tres dimensiones<sup>6</sup>, donde la característica A es la especie, B es la condición y C la ubicación, cada una con dos clasificaciones.

Con esta misma información podrían realizarse otras pruebas, por ejemplo, que las tres características fueran independientes entre sí. Sin embargo, se decidió probar si el comportamiento de las plantas en cuanto a sobrevivencia y ubicación se relaciona con la especie, es decir, si B y C son independientes de A.

Llevando a cabo los cálculos correspondientes se obtiene que:

$$\chi^2 = 11.7666,$$

por lo cual se concluye que existe evidencia para rechazar H<sub>0</sub>, de acuerdo a la forma de la región crítica para esta prueba pues  $\chi^2$  cuadrada resulta ser superior a

$$\chi^2_{((2 \times 2) - 1) \times (2 - 1)} (1 - \alpha) = \chi^2_{(1)} (1 - \alpha),$$

tomando un nivel de significancia de hasta 0.01.

Al considerar los cuadros 4.9 y 4.10, es notable la sobrevivencia en el medio árido de *P. glandulosa*; mientras que en *F. cernua* es un fenómeno que no se presenta y precisamente en estos cruces de la tabla es donde se obtienen las mayores discrepancias.

---

Ver apéndice C, sección II.

Aunque las dos especies estudiadas se consideran como maderables, cada una observa un desarrollo particular bajo un patrón de vegetación consistente en arcos, donde la condición y ubicación de las plantas depende de la especie.

Cuadro 4.8 . PLANTAS ADULTAS DE LAS ESPECIES *Prosopis glandulosa* y *Flourensia cernua* HALLADAS A LO LARGO DE UN TERRENO DE 600m x 2m, PERPENDICULAR A LOS ARCOS DE VEGETACION. LOS RENGLONES INDICAN LA CONDICION DE LAS PLANTAS Y LAS COLUMNAS LA UBICACION.

*Prosopis glandulosa*

UBICACION CONDICION	ARCO	ZONA ARIDA	TOTAL
VIVA	36	9	45
MUERTA	1	19	20
TOTAL	37	28	65

*Flourensia cernua*

UBICACION CONDICION	ARCO	ZONA ARIDA	TOTAL POR UBICACION
VIVA	36	0	36
MUERTA	1	6	7
TOTAL POR CONDICION	37	6	43

Fuente: *Journal of Ecology*, British Ecological Society, Londres, Vol. 78, No. 3, 1990, p. 794.

Cuadro 4.9 . VALORES ESPEFADOS CONFORME A Ho.

*Prosopis glandulosa*

$E_{ijk}$	ARCO	ZONA ARIDA	$\Sigma_k$
VIVA	43.3333	5.4167	48
MUERTA	1.2037	15.0463	20
$\Sigma_j$	37	28	64

*Fluorensia cernua*

$E_{zjk}$	ARCO	ZONA ARIDA	$\Sigma_k$
VIVA	28.6667	3.5833	38
MUERTA	0.7963	9.9537	7
$\Sigma_j$	37	6	43

Cuadro 4.10. DIFERENCIA AL CUADREDO ENTRE LO OBSERVADO Y LO ESPERADO CONFORME A  $H_0$ , ESTANDARIZADA POR LO ESPERADO.

*Prosopis glandulosa*

$\frac{(O_{ijk} - E_{ijk})^2}{E_{ijk}}$	ARCO	ZONA ARIDA	$\Sigma$ v
VIVA	1.2410	2.3705	3.6115
MUERTA	0.0345	1.0389	1.0734
$\Sigma$ j	1.2755	3.4094	4.6849

*Flourensia cernua*

$\frac{(O_{2jk} - E_{2jk})^2}{E_{2jk}}$	ARCO	ZONA ARIDA	$\Sigma$ k
VIVA	1.8759	3.5833	5.4592
MUERTA	0.0521	1.5704	1.6225
$\Sigma$ j	1.9280	5.1537	7.0817

$$\chi^2 = \sum_i \sum_j \sum_k = 11.7668$$

### 4.3. PSICOLOGIA

#### 4.3.1. Conocimientos y actitudes sobre la sexualidad humana

Existen múltiples investigaciones en torno al tema de la sexualidad humana, sin embargo, buena parte de ellas atienden únicamente aspectos biológicos, dejando de lado los factores psicológico y social.<sup>1</sup>

Tomando en consideración precisamente estos dos aspectos, fue realizada la tesis intitulada "Elaboración y aplicación de un cuestionario de conocimientos y actitudes hacia la sexualidad humana en el maestro". Cuyo objetivo general consiste en conocer si existe algún grado de asociación entre las actitudes que manifiestan los maestros de educación primaria hacia el sexo y los conocimientos que estos mismos poseen sobre el tema en cuestión.

Como actitud se define: el grado de aceptación o el grado de rechazo hacia la sexualidad humana. Y conocimiento se refiere al nivel de respuestas correctas obtenidas a través del cuestionario correspondiente.

Es decir, se pretendía saber si aquellos profesores con un alto nivel de conocimientos, de acuerdo a la calificación obtenida al evaluar el cuestionario respectivo, a su vez manifestaban un alto grado de aceptación hacia la sexualidad, así como quienes obtenían una calificación muy baja en los conocimientos, también mostraban niveles bajos de aceptación en relación al sexo.

Las variables de interés (CON=Conocimientos, ACT=Actitudes) son representadas por la calificación total que cada maestro obtuvo al responder a cada uno de los dos cuestionarios aplicados.<sup>2</sup>

---

LOZOYA TELLEZ, Rosario: *Elaboración y aplicación de un cuestionario de conocimientos y actitudes acerca de la sexualidad humana en el maestro*. México, s/ed., 1991, p. I.

<sup>2</sup>Para mayor información sobre los cuestionarios de conocimientos y actitudes favor de referirse al documento ya citado en esta sección.

El estudio se basa en una muestra de 81 profesores, pertenecientes a 3 escuelas primarias ubicadas en una zona del D.F., correspondiente al estrato socioeconómico bajo. El criterio de selección, tanto de escuelas como de profesores, no es probabilístico, por lo cual los resultados de esta investigación son limitados, pues con ellos se logra caracterizar, a lo más, a la planta docente de las escuelas seleccionadas.

El análisis estadístico aplicado a la información obtenida de la encuesta incluyó, entre otras pruebas, el cálculo del coeficiente de correlación de Pearson sobre las variables ACT y CON, el cual no resultó ser estadísticamente significativo con una confianza del 90%. En otras palabras, se encontró evidencia de que los conocimientos y las actitudes de los profesores hacia el sexo no se encuentran relacionados significativamente.

Posteriormente y al margen de la investigación mencionada, se aplicaron los mismos cuestionarios a 23 profesores del Colegio de Ciencias y Humanidades (CCH), plantel Vallejo. De nueva cuenta, el proceso de selección de las unidades muestrales no fue aleatorizado. El Cuadro 4.11 contiene la información que se obtuvo.

A continuación se presenta un breve análisis de estos últimos datos.

La gráfica 4.5 señala que sería válido sugerir una relación lineal entre las variables ACT y CON, por lo tanto se estimará el siguiente modelo de regresión lineal simple:<sup>3</sup>

$$\text{ACT}_i = \beta_0 + \beta_1 \text{CON}_i + \epsilon_i$$
$$i = \overline{1, 23}$$

considerando también los supuestos del modelo sobre  $\epsilon_i$ .

Aplicando el método MCO se obtienen los resultados que muestra el Cuadro 4.12.

El valor de  $R^2$  indica que con el modelo se logra explicar poco menos de la mitad de la variación corregida por la media de la variable ACT. Por otro lado, según  $\hat{\beta}_0$ , en general, a mayor nivel de conocimientos se observa una actitud más positiva hacia el sexo.

---

Ver apéndice D, sección I.

Si los supuestos sobre  $\epsilon_i$  se cumplieran, las pruebas T y F serian validas, al indicar que el coeficiente asociado con la variable CON es estadisticamente significativo.

De cualquier manera, el coeficiente de determinación es relativamente bajo, lo cual sugiere que no se trata de un muy buen ajuste.

Al parecer, analizando la información correspondiente a profesores del CCH se obtienen conclusiones similares a las de la investigación sobre profesores de primaria.

Sin embargo, como lo menciona la tesis citada, la sexualidad humana se encuentra estrechamente ligada con infinidad de factores de indole sociocultural. Estas consideraciones harian pensar que el modelo planteado podria mejorar si, además de los conocimientos, se introdujeran variables relacionadas con características socioculturales de los profesores, por ejemplo: edad, sexo y escolaridad.

Observando la Gráfica 4.6 no parece existir alguna relación funcional en particular entre las actitudes y la edad. Se aprecia que tanto jóvenes como viejos pueden llegar a manifestar actitudes muy negativas hacia el sexo. Sin embargo, todos los profesores que presentan las calificaciones mas altas en actitudes tienen menos de 45 años.

Las variables SEXO y escolaridad (ESC) se construyen de la siguiente manera:

$$SEXO_i = \begin{cases} 1 & \text{si es masculino} \\ 0 & \text{si es femenino} \end{cases}$$

$$ESC_i = \begin{cases} 1 & \text{con Licenciatura} \\ 0 & \text{con estudios superiores (Maestria y/o Doctorado)} \end{cases}$$

$i = \overline{1, 23}$ .

De acuerdo con la Gráfica 4.7, el sexo del profesor parece no tener influencia en las actitudes, aunque existe una pequeña diferencia en favor de los profesores del sexo masculino.

Por otro lado, si se revisa la Gráfica 4.8, resulta evidente que un nivel de escolaridad superior se relaciona con actitudes muy positivas hacia la sexualidad.

Enseguida se incluirán las variables EDAD, SEXO y ESC en el

modelo anterior. Es modelo resultante es:

$$ACT_i = \beta_1 + \beta_2 CON_i + \beta_3 EDAD_i + \beta_4 SEXO_i + \beta_5 ESC_i + \epsilon_i$$
$$i = 1, 23, \dots$$

tomando también en consideración los supuestos sobre  $\epsilon_i$ .

Revisando el Cuadro 4.13, destaca un incremento considerable en el valor de  $R^2$ ,<sup>5</sup> de 0.483 para el modelo inicial a 0.789 para el último, lo cual significa que las nuevas variables incluidas contribuyen a dar un mejor ajuste.

Al parecer, confirmando lo observado en las gráficas 4.9 a 4.11, considerando un nivel de significancia  $\alpha=0.1$ , los coeficientes asociados a las variables EDAD y SEXO no son estadísticamente significativos, según su respectiva prueba  $T^6$ . Aunque en el caso del SEXO, si se fijara un  $\alpha=0.18$  podría rechazarse  $H_0: \beta_4=0$ , lo cual indicaría que el sexo es una variable de utilidad para explicar el comportamiento de las actitudes hacia la sexualidad de los maestros.

La prueba  $F^7$  indica que existen evidencias de que el modelo proporciona un buen ajuste, ya que con un nivel de significancia de hasta 0.001 se puede rechazar  $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ , y aceptar la hipótesis alternativa de que por lo menos uno de los cuatro coeficientes es estadísticamente significativo.

Para validar los resultados anteriores es necesario llevar a cabo el análisis de residuales para verificar si se satisfacen los supuestos del modelo.

Un indicio de que existe el problema de la multicolinealidad se da al obtenerse un coeficiente de determinación muy alto y al mismo tiempo las pruebas  $T$  indican que ninguno o pocos coeficientes son significativos. Bajo este criterio al parecer no se presenta dicho problema.

Existe evidencia gráfica de que el supuesto de homocedasticidad no parece ser violado. En las gráficas 4.9 a 4.11 se aprecia que no hay un patrón bien definido que pudiera asociar la varianza de los residuales con alguna de las variables

<sup>4</sup>Ver apéndice D, sección II.

<sup>5</sup>Ver apéndice D, sección II-b.

<sup>6</sup>Ver apéndice D, sección II-d.

<sup>7</sup>Ver apéndice D, sección II-e.

explicativas o con los valores estimados. En los tres casos se podría trazar una franja horizontal que abarcara todos los puntos.

En cuanto al supuesto de no autocorrelación, parece satisfacerse dado que el valor de la estadística de Durbin-Watson<sup>9</sup> es  $D=1.932$ , siendo  $D_U=1.79^0$ , por lo cual se cumple la desigualdad:

$$D_U < D < 4-D_U.$$

de donde existe evidencia para aceptar la hipótesis nula de que no existe autocorrelación positiva o negativa con una confianza del 95%.

Finalmente, la prueba de Lilliefors<sup>10</sup> sobre los residuales, señala que puede aceptarse el supuesto de distribución Normal con una confianza del 90%, puesto que  $d_{23}=0.143$  y  $d_{23}(0.1)=0.164^{11}$ .

Una vez que se ha realizado el análisis de residuales y aparentemente ninguno de los supuestos del modelo de regresión lineal múltiple se viola, se puede retomar el Cuadro 4.13.

La prueba T correspondiente al coeficiente asociado con la variable EDAD, señala que dicha variable no debe ser incluida en el modelo ya que su  $\beta$  asociada no es estadísticamente significativa. Esta conclusión estadística debe ser tomada con cautela, pues sería importante saber si la variable no significativa es muy importante conceptualmente.

En el caso de la variable SEXO, si se asume un riesgo mayor de rechazar  $H_0$  dado que en realidad es cierta, considerando por ejemplo  $\alpha=0.18$ , entonces se podría concluir que existe evidencia para rechazar  $H_0$  y aceptar  $H_1: \beta_4 \neq 0$ . Se ha tomado esta decisión por considerarse que omitir una variable que en realidad es importante en el modelo sería más grave que incluirla aunque en verdad no fuera tan necesaria.

Tomando en cuenta lo anterior ahora el nuevo modelo a estimar es:

$$ACT_i = \beta_1 + \beta_2 CON_i + \beta_3 SEXO_i + \beta_4 ESC_i + \varepsilon_i.$$

$$i = 1, 23,$$

<sup>9</sup> Ver apéndice D, sección II-f.

<sup>10</sup> Ver Tabla D.5 en GUJARATI, Damodar: *Econometría*, trad. Juan Manuel Mesa, México, McGraw-Hill, 1981, p. 436-437.

<sup>11</sup> Ver apéndice B, sección II.

<sup>12</sup> Ver Tabla 15 de CONOVER, W.J.: *Practical Nonparametric Statistics*, E. U. A., John Wiley & Sons, 1971, p. 398.

considerando también los supuestos sobre  $\epsilon_1$ .

Según el Cuadro 4.15, la  $F^2$  no registró variación respecto a los resultados anteriores, es decir que con el modelo propuesto se sigue explicando cerca del 80% de la variación total, corregida por la media, de las actitudes hacia el sexo.

Sin embargo, como el número de variables explicativas disminuyó, entonces el coeficiente de determinación ajustado aumenta de 0.741 a 0.754. Por lo tanto, el último modelo propuesto da un buen ajuste.

Dado que  $\beta_2 > 0$ , puede decirse que cuando se tiene un conocimiento más amplio sobre el sexo, en general, se presenta un mayor nivel de aceptación hacia él.

Nuevamente, la prueba T sobre el parámetro asociado al SEXO, hace pensar que tal vez esta variable no se encuentra muy relacionada con ACT. En el caso de considerar  $\alpha = 0.132$ , se aceptaría Ho lo cual significa que no existe diferencia en el nivel de aceptación de la sexualidad entre hombres y mujeres.

La escolaridad (ESC) es la variable que cuantitativamente tiene la mayor influencia sobre la actitud hacia la sexualidad, ya que  $\hat{\beta}_4 = -24.718$ . Dado que ha resultado estadísticamente significativa y que además es negativa, puede concluirse que los profesores con estudios superiores a la Licenciatura, manifiestan un mayor nivel de aceptación hacia la sexualidad que aquellos cuyo grado máximo es la Licenciatura.

La variabilidad de los estimadores no es grande, por lo cual los intervalos de confianza del 95% tampoco son amplios y tienen la siguiente forma<sup>12</sup>:

$$\beta_1 \in (54.279, 124.745)$$

$$\beta_2 \in (2.168, 4.370)$$

$$\beta_3 \in (-0.380, 17.120)$$

$$\beta_4 \in (-35.350, -14.077).$$

Según la prueba F, con un nivel de significancia de hasta 0.001, el modelo proporciona un buen ajuste.

Como conclusión general puede decirse que el modelo propuesto para explicar las actitudes de los maestros hacia el sexo es

<sup>12</sup>

Ver apéndice D, sección, II-c.

adecuado. Si bien el sexo del profesor podría no ser incluido, se considera que es preferible tomarlo en cuenta, que correr el riesgo de descartarlo dado que en realidad tiene un efecto significativo sobre las actitudes.

A través de este breve análisis, de alguna manera se apoya la hipótesis de que los factores socioculturales influyen, en gran medida, en el nivel de aceptación de los profesores hacia la sexualidad humana. Por esta razón, podría resultar de gran interés incluir otras variables de este tipo tales como religión y nivel de ingreso, entre otras.

Cuadro 4.11. EVALUACION DE DOS CUESTIONARIOS DE CONOCIMIENTOS Y ACTITUDES HACIA LA SEXUALIDAD HUMANA, EDAD, SEXO Y ESCOLARIDAD DE LOS PROFESORES ENCUESTADOS.

Folio	ACT	CON	EDAD	SEXO	ESC
1	171	36	31	1	1
2	173	30	48	1	1
3	131	24	48	0	1
4	169	30	30	0	1
5	184	32	33	1	1
6	148	27	40	1	1
7	205	32	40	1	0
8	171	27	48	1	1
9	170	31	30	1	1
10	159	31	34	0	1
11	199	28	27	1	0
12	176	30	32	1	1
13	184	33	27	1	1
14	187	33	42	1	1
15	187	31	33	0	1
16	198	34	40	1	1
17	176	34	27	0	1
18	159	27	29	0	1
19	152	24	35	1	1
20	202	36	39	1	1
21	140	21	30	1	1
22	194	30	38	0	0
23	184	32	30	1	0

Cuadro 4.12. ESTIMACIONES Y TABLA ANDEVA CORRESPONDIENTES AL  
 MODELO  $ACT_t = \beta_0 + \beta_1 CON + \epsilon_t$ .

VAR. DEP.: ACT      n=23      R = 0.695      p <sup>2</sup> = 0.483				
R <sup>2</sup> = 0.458, GL = 21				
VARIABLE	COEFICIENTE $\hat{\beta}_i$	ERROR ESTANDAR $e.e.(\hat{\beta}_i)$	T	PCDOS COLAS)
CONSTANT	67.193	23.991	2.80	0.011
CON	3.528	0.797	4.43	0.000

ANALISIS DE VARIANZA					
FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	CUADRADO MEDIO	F	P
EXPLICADA POR REGRESION	3840.651	1	3840.651	19.581	0.000
NO EXPLICADA POR REGRESION	4119.001	21	196.143		

Cuadro 4.13. ESTIMACIONES Y TABLA ANDEVA CORRESPONDIENTES AL

$$\text{MODELO ACT}_i = \beta_1 + \beta_2 \text{CON}_i + \beta_3 \text{EDAD}_i + \beta_4 \text{SEXO}_i + \beta_5 \text{ESC}_i + \varepsilon_i$$

VAR. DEP.: ACT                      n=23                      P = 0.888                       $R^2 = 0.788$   
 $\bar{R}^2 = 0.741$ , GL = 18

VARIABLE	COEFICIENTE $\hat{\beta}_i$	ERROR ESTANDAR $e.e.(\hat{\beta}_i)$	T	P(LOS COLAS)
CONSTANT	87.095	23.211	3.75	0.001
CON	3.307	0.600	5.51	0.000
EDAD	0.041	0.258	0.16	0.874
SEXO	8.218	4.502	1.82	0.085
ESC	-24.831	5.395	-4.60	0.000

## ANALISIS DE VARIANZA

FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	CUADRADO MEDIO	F	P
EXPLICADA POR REGRESION	6272.410	4	1568.103	18.720	0.000
NO EXPLICADA POR REGRESION	1687.242	18	93.736		

Cuadro 4.14. VALORES ESTIMADOS Y RESIDUALES CORRESPONDIENTES AL

$$\text{MODELO } ACT_i = \beta_1 + \beta_2 \text{CON}_i + \beta_3 \text{EDAD}_i + \beta_4 \text{SEXO}_i + \beta_5 \text{ES}_i + \epsilon_i$$

Folio	Estimados (Y <sub>i</sub> )	Residuales (e <sub>i</sub> )
1	190.90	-19.80
2	171.58	1.43
3	143.60	-12.60
4	162.70	6.30
5	177.65	6.35
6	161.41	-13.41
7	202.77	2.23
8	161.66	9.35
9	174.22	-4.22
10	166.17	-7.17
11	169.01	9.99
12	171.00	9.00
13	160.71	3.29
14	161.33	6.67
15	166.13	0.87
16	164.71	3.29
17	175.80	0.20
18	152.74	6.26
19	152.52	-0.52
20	161.13	10.87
21	141.16	-1.16
22	187.96	6.14
23	202.36	-18.36

Cuadro 4.15. ESTIMACIONES Y TABLA ANDEVA CORRESPONDIENTES AL

$$\text{MODELO ACT}_i = \beta_1 + \beta_2 \text{CON}_i + \beta_3 \text{SEXO}_i + \beta_4 \text{ESC}_i + \varepsilon_i$$

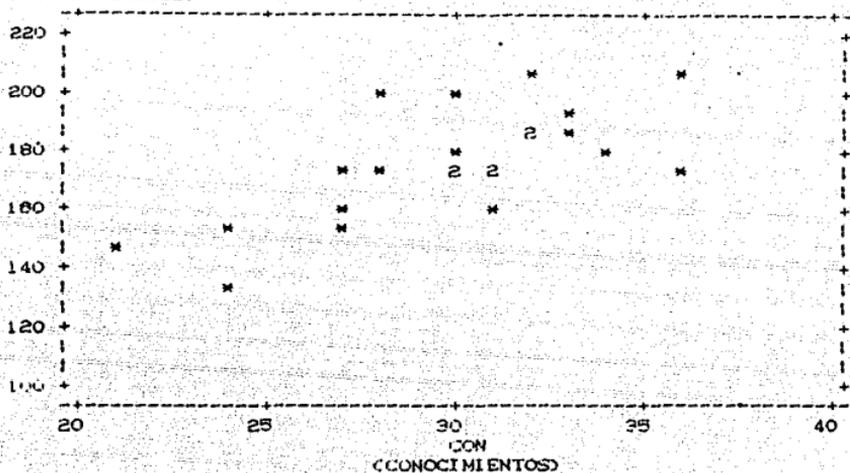
VARIABLE	COEFICIENTE $\hat{\beta}_i$	ERROR ESTANDAR $e.e.(\hat{\beta}_i)$	T	PCDOS COLAS
CONSTANT	89.512	17.254	5.19	0.000
CON	3.269	0.539	6.06	0.000
SEXO	8.370	4.285	1.95	0.068
ESC	-24.718	5.211	-4.74	0.000

## ANALISIS DE VARIANZA

FUENTE DE VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	CUADRADO MEDIO	F	P
EXPLICADA POR REGRESION	6289.975	3	2099.992	23.501	0.000
NO EXPLICADA POR REGRESION	1889.877	19	88.930		

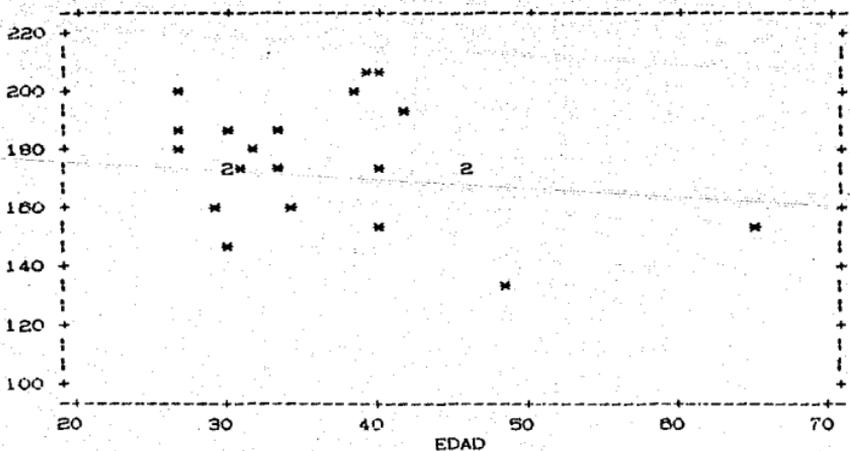
Grafica 4.5 . ACTITUDES VS. CONOCIMIENTOS.

ACT (ACTITUDES)

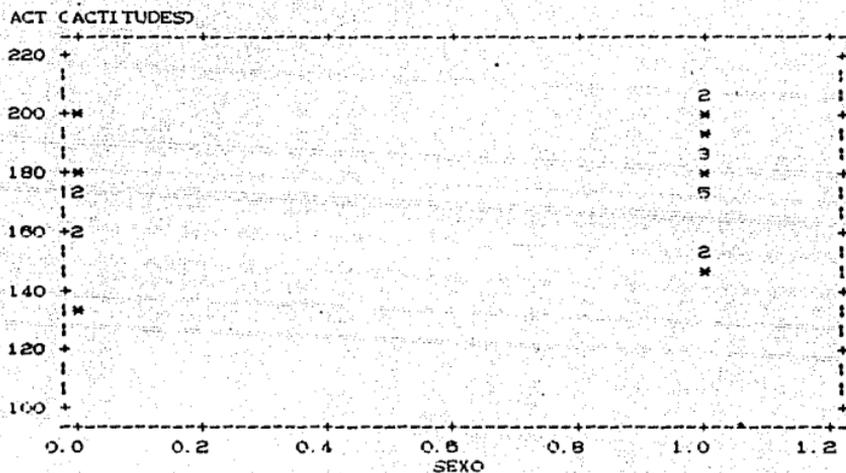


Grafica 4.6 . ACTITUDES VS. EDAD.

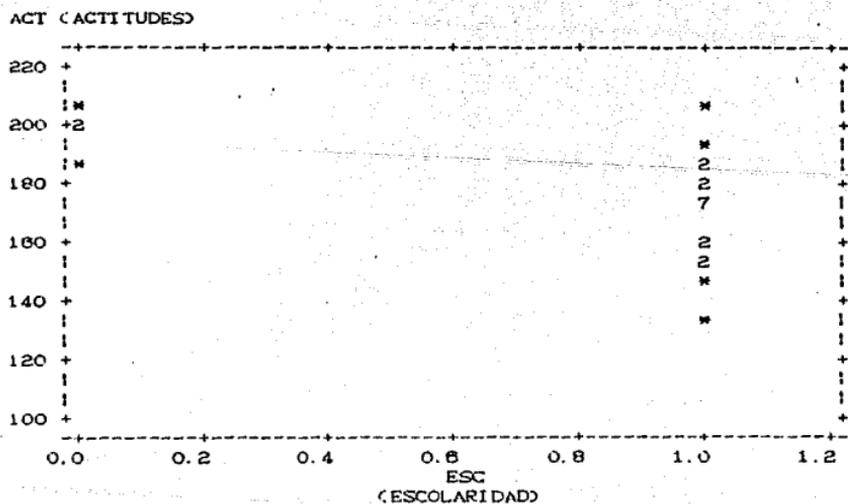
ACT (ACTITUDES)



Gráfica 4.7. ACTITUDES VS. SEXO.

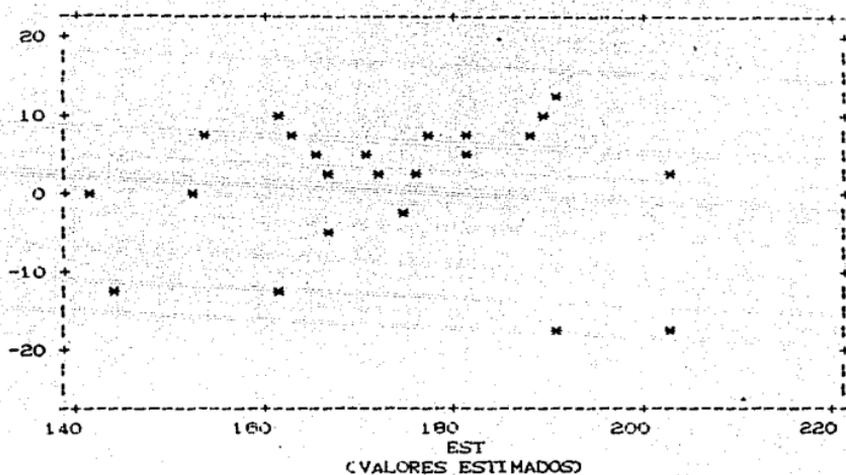


Gráfica 4.8. ACTITUDES VS. ESCOLARIDAD.



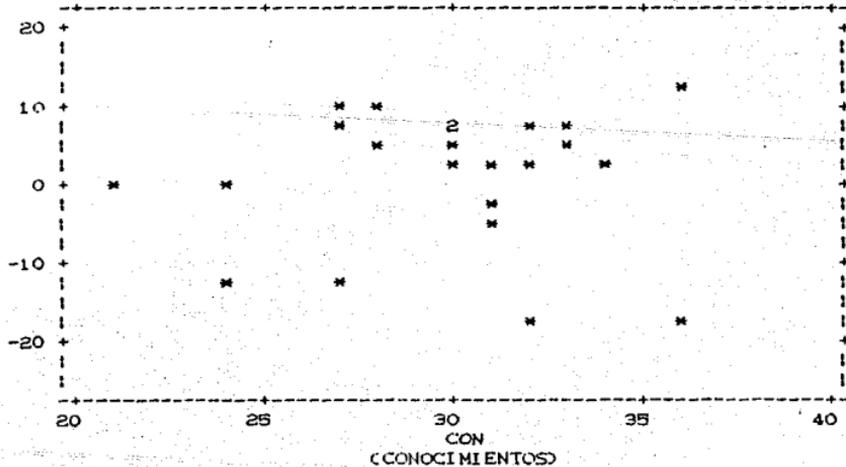
Gráfica 4.9. RESIDUALES VS. VALORES ESTIMADOS.

RES (RESIDUALES)



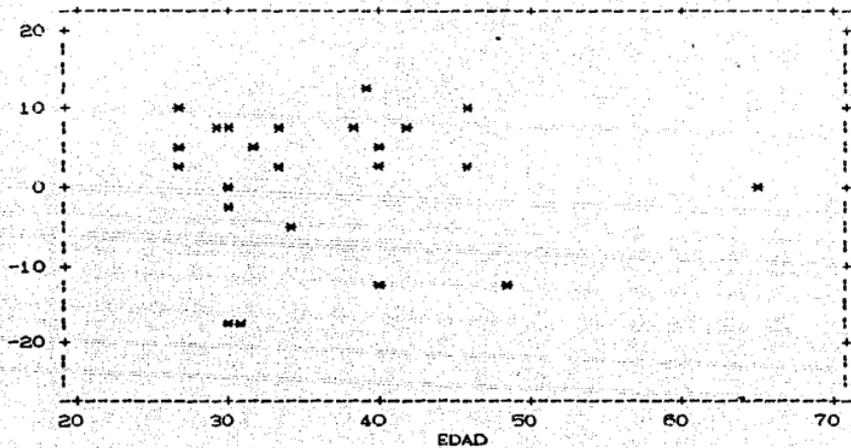
Gráfica 4.10. RESIDUALES VS. CONOCIMIENTOS.

RES (RESIDUALES)



Gráfica 4.11. RESIDUALES VS. EDAD.

RES (RESIDUALES)



#### 4.4. TRANSPORTE

##### 4.4.1. El METRO

Entre todos los medios de transporte público y privado que operan en el Área Metropolitana de la Ciudad de México, el Sistema de Transporte Colectivo (METRO) destaca por su gran capacidad, tanto en recorridos como en pasajeros y porque, debido a las características de su funcionamiento, no contribuye a incrementar los niveles de contaminación, además de constituir un forma de desplazamiento económica, en comparación con otros medios.

La afluencia de pasajeros del METRO no se mantiene constante a lo largo del día, ni conforme va transcurriendo la semana. Y esto es producto de las mismas actividades desarrolladas en la ciudad. Así, podrán encontrarse marcadas diferencias de un día entre semana a un sábado o domingo.

Dejando al margen los días oficialmente feriados y los fines de semana, y considerando la gran diversidad de factores que determinan el que una persona haga o no uso del METRO un día específico, cabe cuestionarse si existen diferencias en el número de usuarios entre los diferentes días de la semana. Es decir, interesa comprobar si:

H<sub>0</sub>: El número de usuarios del METRO es similar de lunes a viernes.  
vs.

H<sub>1</sub>: Por lo menos un día, de lunes a viernes, el número de usuarios del METRO es distinto al de los días restantes.

La información a manejar consiste de 552 observaciones diarias que abarcan el periodo comprendido entre el 1-enero-1988 y el 28-feb-1990.

Según el cuadro de estadísticas generales (Cuadro 4.16) el lunes es el día en el cual se transporta un mayor número de pasajeros, en promedio 4 887 571, y el jueves el menor: 4 437 910. De miércoles a viernes se registran las mayores fluctuaciones, particularmente el jueves (468 753).

Para saber que tipo de prueba es conveniente emplear se

aplicará primero la de Lilliefors<sup>1</sup> para probar la siguiente hipótesis:

H<sub>0</sub>: El número de usuarios del METRO, el día "x" presenta una distribución Normal.

vs.

H<sub>1</sub>: La distribución del número de usuarios del METRO el día "x" no es Normal.

Para concluir sobre la hipótesis anterior se tienen los siguientes datos:

DIA	n	d <sub>n</sub>	d(0.99)*
LUNES	107	0.1755	0.0997
MARTES	112	0.1579	0.0974
MIÉRCOLES	112	0.2522	0.0974
JUEVES	111	0.3122	0.0979
VIERNES	110	0.2615	0.0983

\*Ver tabla 15 en CONOVER, W. J.: *Practical Nonparametric Statistics*, E.U.A., John Wiley & sons, 1971, p. 398.

Como es claro, en ninguno de los 5 casos existe evidencia para aceptar H<sub>0</sub>, ya que d<sub>n</sub> es mayor que d(0.99), considerando un nivel de significancia  $\alpha = 0.01$ . Por lo tanto habrá que hacer uso de una prueba no paramétrica.

Puesto que se tienen muestras que al parecer son independientes, cuyos datos se encuentran en una escala más que ordinal podría ser conveniente aplicar la prueba de Kruskal-Wallis<sup>2</sup> que consiste en un análisis de varianza no paramétrico. A continuación se presentan los resultados obtenidos:

<sup>1</sup>Ver apéndice B, sección II.

<sup>2</sup>Ver apéndice E, sección II-2.

DIA	$R_j/n_j$
LUNES	332.00
MARTES	275.85
MIÉRCOLES	263.08
JUEVES	206.45
VIERNES	307.43

De esta forma  $H$ , que mide la desigualdad entre las sumas de rangos de cada muestra, toma el valor de<sup>9</sup>:

$$H = 39.2998,$$

la cual se supone tiene una distribución  $\chi^2_{(4)}$  con una probabilidad de ocurrencia, conforme a  $H_0$ , muy cercana a cero. Esto significa que, con un nivel de significancia  $\alpha = 0.1$ , puede concluirse que existe evidencia para rechazar  $H_0$ .

En resumen, podría decirse que aun entre los días laborables de la semana existen diferencias con relación a la afluencia de pasajeros del METRO. Aparentemente, conforme la semana avanza disminuye el número de usuarios, para volver a incrementarse a partir del miércoles.

---

<sup>9</sup> Ver apéndice D, fórmula E3.

Cuadro 4.4. ESTADISTICAS GENERALES DEL NUMERO DE USUARIOS DEL METRO POR DIA. (1o. de enero de 1988 - 28 de febrero de 1990.)  
 Excluye sábados, domingos y días oficialmente feriados.

DIA	n	MEDIA	DESV. EST.
LUNES	107	4 687 570.5	244 448.3
MARTES	112	4 622 275.6	254 507.3
MIERCOLES	112	4 560 985.1	423 900.3
JUEVES	111	4 437 910.4	573 022.3
VIERNES	110	4 620 266.8	468 753.3

## CONCLUSIONES

No hay duda en que uno de los comentarios finales más importantes que puede hacerse respecto al presente trabajo, es que en ningún momento su intención fue la de obtener "conclusiones brillantes" en materia demográfica o económica, por ejemplo, sino únicamente la de resaltar que las diferentes técnicas estadísticas pueden dar apoyo en la resolución de cuestionamientos de muy diversa índole.

No está por demás hacer notar que resulta muy conveniente no abusar del uso de la Estadística cuando se llega a la interpretación de resultados, en ninguno de los dos sentidos siguientes: Al dejar expresada una conclusión únicamente en términos numéricos sin retomar la problemática planteada inicialmente y vincularla con dichos números, o bien, generar conclusiones demasiado ambiciosas, cuando la herramienta estadística o la naturaleza misma del problema sugieren una interpretación más conservadora, en términos de la medida en que responden a lo que se ha cuestionado en un inicio.

Ligado al párrafo anterior, seguramente se habrá notado que en muchos casos la técnica aplicada no era la opción idónea sino que existen otras, incluso más avanzadas, a través de las cuales se hubieran obtenido resultados más completos y confiables. Si se reparó en ello, quiere decir que se cumplió uno de los objetivos que persigue este trabajo: El de presentar las técnicas de Estadística básica como una introducción al estudio de técnicas más avanzadas.

Uno de los aspectos que no se abordó, pero que es esencial en la vida profesional, es el de evaluar la calidad de la información estadística a analizar. Ya que el valorarse de datos incompletos o erróneos puede conducir a obtener conclusiones falsas.

El hecho de haber empleado datos verídicos y relacionados con este país, fue con la intención de acercar un poco más a la Estadística con la realidad del lector y que de esta manera sea sencillo imaginar otras aplicaciones.

Para finalizar, vale la pena comentar que las aplicaciones

realizadas fueron concebidas para cumplir un fin didáctico (lo cual espero se haya logrado en alguna medida) por lo que la solución a los cuestionamientos pudo parecer directa y sencilla. Sin embargo, muchos de los problemas de la vida real y en particular de la vida profesional, presentan un alto grado de complejidad y más aun cuando sucede, como en muchos casos, que no se dispone de la información estadística necesaria, o bien esta es limitada o poco confiable.

## BIBLIOGRAFIA

ALBA Francisco: *La población de México. Evolución y dilemas*, 3a. ed., México, El Colegio de México, 1989, 189 p.

COMITE TECNICO DE INFORMACION: *Proyecto de automatización del sistema integral de información de los ingresos y gastos públicos*, México, 1990, 15 p.

CONOVER W. J.: *Practical Nonparametric Statistics*, U.S.A., John Wiley & sons, 1971, 462p.

CONSEJO NACIONAL DE POBLACION: *Población y desarrollo en México y el Mundo*, México, CONAPO, 1988, I al IV, 1291 p.

DRAPER N. R., SMITH H.: *Applied Regression Analysis*, 2a. ed., E.U.A., John Wiley & sons, 1981, 707 p.

GUJARATI Damodar: *Econometria*, trad. Juan Manuel Mesa, 1a. ed., México, Mc Graw-Hill, 1986, 463 p.

"Índices de Precios, ¿qué nos dicen?, ¿para qué nos sirven?", en: *Tendencias Económicas y Financieras*, Grupo Expansión, México, No. 10, marzo 1989, p. 11.

INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *Anuario Estadístico de los Estados Unidos Mexicanos 1988-1989*, México, INEGI, 1990.

INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *El ABC de las Cuentas Nacionales*, México, INEGI, 1985, 35 p.

INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *Estadísticas Históricas de México*, México, INEGI, 1983, t. I.

INSTITUTO NACIONAL DE ESTADISTICA GEOGRAFIA E INFORMATICA: *Resultados oportunos, Censos Económicos 1986*, México, INEGI, 1987, 58 p.

LOPEZ ROSADO, Diego G.: *Problemas económicos de México*, 5a. ed., México, UNAM, 1984.

LOZOYA TELLEZ, Rosario: *Elaboración y aplicación de un cuestionario de conocimientos y actitudes acerca de la sexualidad humana en el maestro*, México, 1991, 105 p.

MEXICO, BANCO DE MEXICO: *Indicadores económicos*, México, 1990.

MEXICO, COMISION NACIONAL DE LOS SALARIOS MINIMOS: *Salarios Minimos vigentes a partir del 10. de enero de 1989*, México, 1989.

MEXICO, SECRETARIA DE PROGRAHACION Y PRESUPUESTO: *Estadísticas de Finanzas Públicas 1970-1985*, México, SPP, 1988, 112 p.

MONTAÑA C., LOPES-PORTILLO J. y MAUGHAM A.: "The response of two woody species to the conditions created by a shifting ecotone in an arid ecosystem", en: *The Journal of Ecology*, British Ecological Society, Londres, sept., 1990, vol. 78, no. 3, p. 648-663.

MOOD Alexander M., GRAYBILL Franklin A. y BOES Juan C.: *Introduction to the theory of statistics*, 3a. ed., Singapur, Mc Graw-Hill, 1968, 564 p.

OYAMA K.: "Variation in growth and reproduction in the Neotropical dioecious palm *Chamaedorea tepehlate*", en: *The Journal of Ecology*, British Ecological Society, Londres, sept., 1990, vol. 78, no. 3., p. 649-663.

SEBEK G. A. F.: *Linear Regression Analysis*, Nueva York, John Wiley & sons, 1977.

SIEGEL Sidney: *Diseño experimental no paramétrico aplicado a las ciencias de la conducta*, trad. por: Javier Aguilar Villalobos, 1a. ed., México, Trillas, 1970, 348 p.

TABORGA Huascar: *Como hacer una tesis*, 11a. ed., México, Grijalbo, 1982, 220 p.

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO: *Agenda Estadística 1989*, México, UNAM, 1989, 99 p.

ZAVALA Iván: *El año pasado 1987*, 1a. ed., México, Plaza y Janés, 1988, 450 p.

## APPENDICES

## APENDICE A. PRUEBAS DE HIPOTESIS

### I. DISTRIBUCION NORMAL.

#### 1. Sobre $\mu$ y con $k=2$ poblaciones.

a)  $H_0: \mu_1 = \mu_2$  vs.  $H_1: \mu_1 \neq \mu_2$

$$\mathcal{C} = \{ (X_1, X_2) \mid t > t_{(m+n-2)}^{(1-\alpha/2)}, t < -t_{(m+n-2)}^{(1-\alpha/2)} \}$$

b)  $H_0: \mu_1 = \mu_2$  vs.  $H_1: \mu_1 < \mu_2$

$$\mathcal{C} = \{ (X_1, X_2) \mid t < -t_{(m+n-2)}^{(1-\alpha)} \}$$

c)  $H_0: \mu_1 = \mu_2$  vs.  $H_1: \mu_1 > \mu_2$

$$\mathcal{C} = \{ (X_1, X_2) \mid t > t_{(m+n-2)}^{(1-\alpha)} \}$$

$$A1. t = \frac{\sqrt{\frac{mn(m+n-2)}{m+n}} (\bar{X}_1 - \bar{X}_2)}{\sqrt{\sum_{i=1}^m (X_{1i} - \bar{X}_1)^2 + \sum_{j=1}^n (X_{2j} - \bar{X}_2)^2}}$$

#### 2. Sobre $\mu$ y con $k$ poblaciones.

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$  vs.  $H_1$ : por lo menos  $\mu_k$  distinta p.a.  $k$

$$\mathcal{C} = \{ (X_1, X_2, \dots, X_k) \mid r \leq F_{(k-1, n-k)}(\alpha), r \geq F_{(k-1, n-k)}^{(1-\alpha)} \}$$

$$A2. r = \frac{\sum_{j=1}^k n_j (\bar{X}_j - \bar{X})^2 / (k-1)}{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2 / (n-k)}$$

Donde:

$$A3. \bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji}, \quad j = \overline{1, k}.$$

$$A4. \bar{X} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} X_{ji}$$

$$A5. n = \sum_{j=1}^k n_j$$

3. Sobre  $\sigma^2$  y con  $k=2$  poblaciones.

a) Ho:  $\sigma_1^2 = \sigma_2^2$  vs.  $H_1: \sigma_1^2 \neq \sigma_2^2$

$$\mathcal{C} = \langle \langle \bar{X}_1, \bar{X}_2 \rangle \mid f \rangle \in F_{(m-1, n-1)}^{(1-\alpha/2)}, f \in F_{(m-1, n-1)}^{(\alpha/2)}$$

b) Ho:  $\sigma_1^2 = \sigma_2^2$  vs.  $H_1: \sigma_1^2 < \sigma_2^2$

$$\mathcal{C} = \langle \langle \bar{X}_1, \bar{X}_2 \rangle \mid f \rangle \in F_{(m-1, n-1)}^{(\alpha)}$$

c) Ho:  $\sigma_1^2 = \sigma_2^2$  vs.  $H_1: \sigma_1^2 > \sigma_2^2$

$$\mathcal{C} = \langle \langle \bar{X}_1, \bar{X}_2 \rangle \mid f \rangle \in F_{(m-1, n-1)}^{(1-\alpha)}$$

$$A6. f = \frac{(n-1) \sum_{i=1}^m (X_{1i} - \bar{X}_1)^2}{(m-1) \sum_{j=1}^n (X_{2j} - \bar{X}_2)^2}$$

4. Sobre  $\sigma^2$  y con  $k$  poblaciones.

Ho:  $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$  vs.  $H_1$ : por lo menos  $\sigma_k^2$  distinta p.a.  $k$

$$\mathcal{C} = \langle \langle \bar{X}_1, \bar{X}_2, \dots, \bar{X}_k \rangle \mid -2 \log \lambda \rangle, \chi^2_{(k-1)}^{(1-\alpha)}, -2 \log \lambda > \chi^2_{(k-1)}^{(\alpha)}$$

$$A7. \lambda = \frac{\prod_{j=1}^k \hat{\sigma}_j^2 n_j / 2}{\left( \sum_{j=1}^k n_j \hat{\sigma}_j^2 / \sum_{j=1}^k n_j \right) \sum n_j / 2}$$

Donde:

$$A8. \hat{\sigma}_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2$$

$$A9. \bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji}$$

## II. DISTRIBUCIÓN BERNOLLI.

Sobre  $p$  y con una población.

$$H_0: p = p_0 \quad \text{vs.} \quad H_1: p > p_0$$

$$\mathcal{C} = \langle CX \mid \sum_{i=1}^n X_i \geq A' \rangle$$

$$A10. \quad A' = \sqrt{np_0(1-p_0)} \quad z(1-\alpha) + np_0$$

## APENDICE B. PRUEBAS DE BONDAD DE AJUSTE

### I. KOLMOGOROV-SMIRNOV.

$H_0: X \sim F_X$  vs.  $H_1: X \sim F_X$  ( $F_X$  continua)

$\mathcal{C} = \{ (X) \mid d_n \geq d(1-\alpha) \}$

$$B1. d_n = \sup_x |F_n(x) - F_X(x)|$$

donde  $F_n(x)$  es la función de distribución empírica.

### II. LILLIEFORS.

$H_0: X \sim N$  vs.  $H_1: X \sim N$ ,  
con media y varianza desconocida.

$\mathcal{C} = \{ (X) \mid d_n \geq d(1-\alpha) \}$

$$B2. d_n = \sup_z |F_n(z) - F_X(z)|$$

donde  $d_n$  es calculada usando los datos estandarizados:

$$Z_i = \frac{X_i - \bar{X}}{s} \quad i = 1, \dots, n.$$

## APENDICE C. PRUEBAS DE INDEPENDENCIA

### I. TABLAS DE CONTINGENCIA EN DOS DIMENSIONES

Sean  $X_1, \dots, X_n$  individuos u objetos con 2 características A y B. Los cuales pueden ser clasificados en  $p$  categorías, según el criterio A y en  $q$  categorías de acuerdo al B.

$$H_0: p_{ij} = p_i \cdot p_j \quad \forall (i, j) \quad \text{vs.} \quad H_1: p_{ij} \neq p_i \cdot p_j \quad \wedge (i, j)$$

$$i=1, \dots, p \quad \& \quad j=1, \dots, q$$

$$\chi^2 = \left( \chi^2_{(p-1)(q-1)} \mid 1-\alpha \right)$$

$$C1. \chi^2 = \sum_{i=1}^p \sum_{j=1}^q \frac{(E_{ij} - O_{ij})^2}{E_{ij}}$$

$$\text{donde: } E_{ij} = \frac{n_i \cdot n_j}{n}$$

$$n_i = \sum_{j=1}^q n_{ij}$$

$$n_j = \sum_{i=1}^p n_{ij}$$

$$n = \sum_{i=1}^p \sum_{j=1}^q n_{ij}$$

### II. TABLAS DE CONTINGENCIA DE TRES DIMENSIONES

Sean  $X_1, \dots, X_n$  individuos u objetos con 3 características A, B y C. Los cuales pueden ser clasificados en  $p$  categorías, según el criterio A, en  $q$  categorías de acuerdo al B y en  $r$  según el C.

$$H_0: p_{ijk} = p_i \cdot p_j \cdot p_k \quad \forall (i, j, k)$$

$$\text{vs.}$$

$$H_1: p_{ijk} \neq p_i \cdot p_j \cdot p_k \quad \wedge (i, j, k)$$

$$i=1, \dots, p; \quad j=1, \dots, q \quad \text{y} \quad k=1, \dots, r.$$

(i.e.: la característica A es independiente de las características B y C)

$$\chi^2 = \left( \chi^2_{(p-1)(q-1)(r-1)} \mid 1-\alpha \right)$$

$$C2. \chi^2 = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r \frac{(E_{ijk} - O_{ijk})^2}{E_{ijk}}$$

donde 
$$E_{ij} = \frac{n_{.jk} n_{i.}}{n}$$

$$n_{.jk} = \sum_{i=1}^p n_{ijk}$$

$$n_{i.} = \sum_{j=1}^q \sum_{k=1}^r n_{ijk}$$

$$n = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r n_{ijk}$$

## APENDICE D. ANALISIS DE REGRESION

### I. MODELO DE REGRESION LINEAL SIMPLE

#### a) Modelo

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$$E(\varepsilon_i) = 0$$

$$\text{var}(\varepsilon_i) = \sigma^2$$

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$i = \overline{1, n.}$$

#### b) Estimadores y coeficiente de determinación

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{SCXY}{SCX}$$

$$\hat{\sigma}^2 = \frac{SCR}{n-2}$$

$$\text{var}(\hat{\beta}_0) = \hat{\sigma}^2 \frac{\sum_{i=1}^n X_i^2}{n SCX}$$

$$\text{var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{SCX}$$

$$R^2 = \frac{SCY - SCR}{SCY}$$

$$\bar{R}^2 = 1 - \frac{(n-1)(1-R^2)}{n-2}$$

Donde:

$$SCXY = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

$$SCX = \sum_{i=1}^n (X_i - \bar{X})^2; \quad SCY = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$SCR = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

c) Intervalos de confianza

$$\beta_1 \in \left[ \hat{\beta}_1 - \sqrt{\widehat{\text{var}}(\hat{\beta}_1)} t_{(n-2), (1-\alpha/2)}, \hat{\beta}_1 + \sqrt{\widehat{\text{var}}(\hat{\beta}_1)} t_{(n-2), (1-\alpha/2)} \right]$$

d) Prueba T

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

$$\hat{\sigma} = ( \chi^2 \mid t > t_{(n-2), (1-\alpha/2)} - t < t_{(n-2), (1-\alpha/2)} )$$

$$t = \frac{\hat{\beta}_1}{\sqrt{\widehat{\text{var}}(\hat{\beta}_1)}}$$

e) Prueba F

$$H_0: \beta_1 = 0 \text{ vs. } H_1: \beta_1 \neq 0$$

$$\hat{\sigma} = ( \chi^2 \mid f > F_{(1, n-2), (1-\alpha/2)} )$$

$$f = \frac{\text{CMReg}}{\text{CMR}}$$

Donde:

$$\text{CMReg} = \frac{\text{SCReg}}{1}$$

$$\text{CMR} = \frac{\text{SCR}}{n-2}$$

$$\text{SCReg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

$$\text{SCR} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

f) Prueba de Durbin-Watson<sup>1</sup>

H<sub>0</sub>: No hay correlación serial

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

<sup>1</sup> GUJARATI, Damodar: *Econometría*, trad. Juan Manuel Mesa, México, Mc Graw-Hill, 1986, p. 231-235.

## 11. MODELO DE REGRESION LINEAL MULTIPLE

### a) Modelo

En notación matricial:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad Y_{n \times 1} = X_{n \times k} \beta_{k \times 1} + \varepsilon_{n \times 1}$$

$$E(\varepsilon_i) = 0 \quad E(\varepsilon) = 0$$

$$\text{var}(\varepsilon_i) = \sigma^2 \quad \text{var-cov}(\varepsilon) = \sigma^2 I_n$$

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$r(X) = k < n$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

$$i = \overline{1, n}$$

### b) Ecuaciones Normales, estimadores y coeficiente de determinación

$$X'X\hat{\beta} = X'Y$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{\sigma}^2 = \frac{e'e}{n-k}$$

$$\text{var-cov}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}$$

$$R^2 = \frac{SCY - SCR}{SCY}$$

$$\bar{R}^2 = 1 - \frac{(n-1)(1-R^2)}{n-2}$$

Donde:

$$SCY = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

$$SCR = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

### c) Intervalos de confianza

$$\beta_i \in \left( \hat{\beta}_i - \sqrt{\text{var}(\hat{\beta}_i)} t_{(n-k)}(1-\alpha/2), \hat{\beta}_i + \sqrt{\text{var}(\hat{\beta}_i)} t_{(n-k)}(1-\alpha/2) \right)$$

$$i = \overline{1, k}$$

d) Prueba T

Ho:  $\beta_1 = 0$  vs.  $H_1: \beta_1 \neq 0$

$$\hat{\beta}_1 \in (X' | t \cdot t_{(n-k)}^{(1-\alpha/2)} - t < t_{(n-k)}^{(1-\alpha/2)})$$

$$t = \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}}, \quad 1 = 1, k.$$

e) Prueba F

Ho:  $\beta_2 = \beta_3 = \dots = \beta_k = 0$  vs.  $H_1: \beta_i \neq 0$  p.a. 1

$$\hat{\beta} \in (X' | F_{(k-1, n-k)}^{(1-\alpha/2)})$$

$$f = \frac{\text{CMReg}}{\text{CMR}}$$

Donde:

$$\text{CMReg} = \frac{\text{SCReg}}{1}$$

$$\text{CMR} = \frac{\text{SCP}}{n-2}$$

$$\text{SCReg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

$$\text{SCP} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

f) Prueba de Durbin-Watson<sup>3</sup>

Ho: No hay correlación serial

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

<sup>3</sup> Para mayor información acerca de la prueba de Durbin-Watson ver: GUJARATI, Damodar: *Econometría*, trad. Juan Manuel Mesa, México, Mc Graw-Hill, 1986, p. 231-235.

## APENDICE E. ESTADISTICA NO PARAMETRICA

### I. COMPARACION DE DOS POBLACIONES

Prueba de los rangos de Wilcoxon

Se tienen dos muestras relacionadas.

$H_0$ : Las variables X y Y son similares

$D_i = X_i - Y_i$

$R_i$  = Rango de  $D_i$

La estadística de prueba R se elige de acuerdo a la forma de  $H_1$ , entre:

$$R^+ = \sum D_i \quad \forall D_i > 0$$

$$R^- = \sum D_i \quad \forall D_i < 0$$

Para muestras pequeñas ( $n \leq 25$ ):

$$z = ((X, Y) \mid R \leq r_0)$$

Para muestras grandes ( $n > 25$ ) se calcula:

$$E1. z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N_{(0,1)}$$

Si la probabilidad asociada a un valor tan extremo como la z observada, usando tablas de la distribución Normal, es menor o igual que  $\alpha$  se rechaza  $H_0$ .

### II. COMPARACION ENTRE MAS DE DOS POBLACIONES.

1. Análisis de la Varianza de dos clasificaciones por rangos de Friedman.

Se tienen k muestras relacionadas correspondientes a las variables  $X_1, \dots, X_k$ .

$H_0$ : Las variables  $X_1, \dots, X_k$  son similares entre si.

vs.

$H_1$ : Por lo menos una X es distinta a las demás.

La estadística de prueba es como sigue:

$$E2. \chi_r^2 = \frac{12}{n(k)(k+1)} \sum_{j=1}^k (R_j)^2 - 3n(k+1) \sim \chi_{(k-1)}^2$$

$R_j$  = Suma de los rangos de la columna  $j$

Para muestras grandes:

Si la probabilidad asociada a un valor tan extremo como la  $\chi_r^2$  observada, usando tablas de la distribución Ji-Cuadrada, es menor o igual que  $\alpha$  se rechaza  $H_0$ .

2. Análisis de varianza de una clasificación por rangos de Kruskal-Wallis.

Se tienen  $k$  muestras independientes correspondientes a las variables  $X_1, \dots, X_k$ .

$H_0$ : Las variables  $X_1, \dots, X_k$  son similares entre sí.

vs.

$H_1$ : Por lo menos una  $X$  es distinta a las demás.

La estadística de prueba es como sigue:

$$E3. H = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(n+1) \sim \chi_{(k-1)}^2$$

Donde:

$k$  = No. de muestras

$n_j$  = No. de casos de la  $j$ -ésima muestra;  $\sum_{j=1}^k n_j = n$

$R_j$  = Suma de rangos de la  $j$ -ésima muestra

$$j = \overline{1, k.}$$

Si la probabilidad asociada a un valor tan extremo como la  $\chi_r^2$  observada, usando tablas de la distribución Ji-Cuadrada, es menor o igual que  $\alpha$  se rechaza  $H_0$ .

### III. MEDIDAS DE ASOCIACION

#### 1. Coeficiente de correlacion de rango de Spearman.

Ho: Las variables X y Y no se encuentran relacionadas en la poblacion

vs  
 H1: Las variables X y Y si se encuentran relacionadas en la poblacion.

$$E4. r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n^3 - n}$$

$$D_i = X_i - Y_i$$

Si la probabilidad asociada con la ocurrencia de un valor tan extremo como la  $r_s$  observada es igual o menor que  $\alpha$ , puede rechazarse Ho.

#### 2. Coeficiente de correlacion de rango de Kendall.

Ho: Las variables X y Y no se encuentran relacionadas en la poblacion

vs  
 H1: Las variables X y Y si se encuentran relacionadas en la poblacion.

$$E5. r = \frac{S}{\frac{1}{2} n(n-1)}$$

Ajustando por observaciones ligadas:

$$E6. r = \frac{S}{\sqrt{\frac{1}{2} n(n-1) - T_x} \sqrt{\frac{1}{2} n(n-1) - T_y}}$$

S = Diferencia entre el numero de pares de rangos que se encuentran en el orden natural y el numero de pares que no se encuentra en este orden, considerando la variable cuyos rangos fueron ordenados de acuerdo al orden natural de la otra variable.

$T_x = \frac{1}{2} \sum t(t-1)$ ; t=No. de obs. ligadas en cada grupo de ligas de X

$T_y = \frac{1}{2} \sum t(t-1)$ ; t=No. de obs. ligadas en cada grupo de ligas de Y

Para determinar la significancia de  $r$  si no se

$$E7. z = \frac{r}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \sim N(0,0)$$

Si la probabilidad asociada de un valor tan grande como la  $z$  observada es igual o menor que  $\alpha$ , usando tablas de la distribución Normal, puede rechazarse  $H_0$ .

### 3. Coeficiente de Contingencia.

Usando la información de una tabla de contingencia de  $p \times q$  (Apendice C):

$H_0$ : Las características A y B son independientes

vs.

$H_1$ : Las características A y B son dependientes

$$\chi^2 = (C) \chi^2_{(p-1)(q-1)} (1-\alpha)$$

$$E8. C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$