



**UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO**

**ESCUELA NACIONAL DE ESTUDIOS
PROFESIONALES
ACATLAN**



**TENDENCIA DE LOS NIVELES DE PLOMO EN SANGRE
EN MUJERES EMBARAZADAS DE LA CIUDAD DE
MEXICO RELACIONADO CON LOS NIVELES
DE PLOMO EN AIRE.**

T E S I S P R O F E S I O N A L
QUE PARA OBTENER EL TITULO DE:
L I C E N C I A D A E N M A T E M A T I C A S
A P L I C A D A S Y C O M P U T A C I O N
P R E S E N T A :
M O N T S E R R A T L I N O G O N Z A L E Z

ASESORA: DRA. SILVIA RUIZ VELASCO ACOSTA

NAUCALPAN, ESTADO DE MEXICO.

**TESIS CON
FALLA DE ORIGEN**





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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 esta 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.



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES "ACATLAN"
DIVISION DE MATEMATICAS E INGENIERIA
PROGRAMA DE ACTUARIA Y M.A.C.

SRITA. MONTSERRAT LINO GONZALEZ

Alumna de la carrera de Matemáticas Aplicadas y Computación.
Presente.

De acuerdo a su solicitud presentada con fecha 24 de abril de 1996, me complace notificarle que esta Jefatura tuvo a bien asignarle el siguiente tema de Tesis: "TENDENCIA DE LOS NIVELES DE PLOMO EN SANGRE EN MUJERES EMBARAZADAS DE LA CIUDAD DE MEXICO RELACIONADAS CON LOS NIVELES DE PB EN AIRE ", el cual se desarrollará como sigue:

INTRODUCCION.

CAP. I Marco Teórico.

CAP. II Análisis exploratorio de datos .

CAP. III Correlación .

CAP. IV Regresión lineal.

CAP. V Analisis estadísticos .

CAP. VI Resultados.

CONCLUSIONES .

BIBLIOGRAFIA.

Asimismo, fué designado como Asesor de Tesis: la DRA. SILVIA RUIZ VELASCO.

Ruego a usted tomar nota que en cumplimiento de lo especificado en la Ley de Profesiones, deberá presentar servicio social durante un tiempo mínimo de seis meses como requisito básico para sustentar examen profesional, así como de la disposición de la Coordinación de la Administración Escolar en el sentido de que se imprima en lugar visible de los ejemplares de la Tesis el título del trabajo realizado. Esta comunicación deberá imprimirse en el interior de la misma.

ATENTAMENTE
"POR MI RAZA HABLARA EL ESPIRITU"
Acatlán, Edo. Méx. Enero 27 de 1997.

ACT. LAURA MARCELA BECERRA
Jefe del Programa de Actuaría y M.A.C.

cg'

ESLEP. ACATLAN



JEFATURA DEL PROGRAMA DE
ACTUARIA Y MATEMATICAS
APLICADAS Y COMPUTACION

AGRADEZCO:

A la Dra. Silvia Ruiz Velasco Acosta por su dedicación e interés durante la dirección y asesoría de éste trabajo.

Al Dr. Mauricio Hernández Avila, quien también asesoró ésta tesis y puso ante mí la oportunidad de desarrollar éste trabajo. Gracias por el apoyo, las enseñanzas, los valiosos comentarios, aportaciones y el tiempo extra dedicado durante la realización de éste trabajo.

Al Dr. Fernando Meneses González, quien ha sido un gran maestro y ejemplo en mi desarrollo profesional; así como por su invaluable apoyo.

A la Dra. Isabelle Romieu porque gracias a sus enseñanzas del curso de Epidemiología Ambiental pude comprender muchos aspectos necesarios para esta tesis.

A la Dra. Luz Helena Sanin, a la Ing. Marlene Cortez Lugo y a la Q.F.B. Ilda Muñoz Quiles, quienes me facilitaron material bibliográfico y sus valiosas experiencias.

A las Q.F.B. Araceli Rivero Avilés y Q.F.B. María Teresa Amador Cabrera, quienes me explicaron el proceso de la determinación de plomo en sangre.

Al Instituto Nacional de Salud Pública en la Subdirección del Centro de Investigación en Salud Poblacional, a la Red Manual de Monitoreo Atmosférico, al Instituto Mexicano del Petróleo y a las mujeres embarazadas que accedieron participar en el estudio, ya que sin ellos no hubiera podido realizarse éste trabajo.

A la Universidad Nacional Autónoma de México, a la Escuela Nacional de Estudios Profesionales de Acatlán y especialmente a su profesorado por su colaboración en mi formación profesional y su esforzada e invaluable labor en beneficio de la Sociedad.

DEDICO ESTE TRABAJO

A mis padres, el señor Lic. Ramón Lino Jurado y la señora Rosa María González Almeida, por su apoyo incondicional, comprensión y cariño; así como su ejemplar manera de vivir la vida y saber guiar mis pasos a través de ella.

A mis hermanos Ramón, Ana Luisa y Rodrigo Carlos, por su ejemplo, apoyo y cariño.

A mis abuelos, tíos y primos y muy especialmente al Ing. Juan José González Almeida y a la Dra. María del Carmen González Almeida por su apoyo.

A todas aquellas personas a las que de alguna manera la realización de éste trabajo les pueda aportar algo positivo.

INDICE

INTRODUCCION.....	1
OBJETIVO GENERAL.....	3
OBJETIVOS ESPECIFICOS.....	3
I. MARCO TEORICO.....	4
A. La salud ambiental.	
B. La contaminación del aire.	
C. El plomo	
1. La contaminación por plomo y principales fuentes emisoras	
a. Aire.	
b. Suelo	
c. Agua.	
d. Alimentos.	
2. Intoxicación por plomo y efectos en la salud.	
II. ANALISIS EXPLORATORIO DE DATOS.....	12
A. Análisis exploratorio en los datos de la Red Manual de Monitoreo Atmosférico.	
B. Análisis exploratorio de las variables proporcionadas por el Instituto Mexicano del Petróleo 1995.	
C. Análisis exploratorio de la información proporcionada por INSP-CISP.	
III. CORRELACION.....	55
A. Características del coeficiente de correlación	
B. Cálculo del coeficiente de correlación de Pearson.	
C. Interpretación del coeficiente de correlación.	
1. Coeficiente de determinación.	
2. Prueba t para correlación	
3. Transformación z de Fisher	
4. Intervalos de confianza para la correlación.	
5. Coeficiente de correlación del Rango de Spearman.	

IV. REGRESION LINEAL..... 66

- A. Definición
- B. Supuestos estadísticos de la Regresión Lineal
 - 1. Existencia
 - 2. Independencia.
 - 3. Linealidad
 - a. Esquema supuestos de regresión lineal
 - 4. Homocedasticidad
 - 5. Normalidad
 - a. Gráficos residuales versus predichos.
 - 6. Transformaciones
- C. Método de mínimos cuadrados.
- D. Cálculo de la ecuación de regresión
- E. Predicción con la ecuación de regresión.
- F. Análisis de residuos y puntos influyentes.
- G. Regresión Múltiple.

V. ANALISIS ESTADISTICO..... 85

- A. Materiales y métodos
 - 1. Sistemas de evaluación de la calidad del aire en la Ciudad de México
 - a. Antecedentes
 - b. La Red Automática de Monitoreo Atmosférico
 - c. La Red Manual de Monitoreo Atmosférico
 - d. Parámetros en cada estación donde se monitorea Pb
 - e. Ubicación de estaciones donde se mide plomo
 - 2. Determinación de plomo en sangre
 - 3. Criterios de asignación de plomo en aire a plomo en sangre
 - a. Número de muestreos de Pb en aire por estación
- B. Análisis de variación de los datos de plomo en aire calculados por el programa versus los datos del IMP.
- C. Correlación
- D. Regresión Lineal

VI RESULTADOS.....	105
GLOSARIO.....	115
APENDICE A: Programa.....	118
BIBLIOGRAFIA.....	130

INTRODUCCION

La contaminación es uno de los grandes problemas que aqueja a la humanidad en general y por ende a la población que habita en el área del Valle de México.

El plomo es un elemento extraño en la biología celular de cualquier ser vivo, pues no se conocen funciones o reacciones bioquímicas en las cuales desempeñe alguna función. Es por ello que la determinación de su presencia en cualquier ser vivo, provoca enfermedades que como se comprenderá, las más estudiadas son las del ser humano.

La contaminación por plomo (Pb) que es un metal pesado, provoca efectos dañinos en la salud, principalmente en los niños, causando desde somnolencia hasta un retardo mental profundo e inclusive la muerte, si la exposición a dicho metal es crónica y los niveles de plomo en sangre son elevados. Las mujeres embarazadas con éstas características, le transmiten al feto parte de ese nivel de plomo por medio del cordón umbilical, debido en parte a los grandes cambios que presenta la madre en la etapa de gestación.

Esta tesis, parte de la hipótesis de que el plomo en sangre está relacionado con los niveles de plomo en aire de meses antes de la toma de muestra sanguínea y el objetivo principal es encontrar la relación entre los niveles de plomo en sangre con los niveles de plomo en aire, a través de la información estadística recopilada de 1992 a 1995 de plomo en sangre y de 1986 a 1995 del plomo en aire. Asimismo, siendo la gasolina la principal fuente de plomo en aire, trataremos de establecer la relación existente entre los diferentes tipos de gasolinas, la contaminación del aire por plomo y sus efectos medidos a través de plomo en sangre en el transcurso de varios años. Hay que reconocer que también los niveles de plomo en sangre, son reflejo de ciertos hábitos en la alimentación de las madres, como el uso de recipientes de barro vidriado, ya sea para cocinar o bien guardar los alimentos que ingiere. Este aspecto de la alimentación también es manejado dentro de ésta tesis.

Se utilizó correlación, debido a que se encontró una fuerte asociación entre las concentraciones de plomo en aire y consumo de gasolinas con plomo; así como una fuerte asociación en el decremento de la mediana de los niveles de plomo en sangre de las madres y la reducción en el consumo de gasolinas con plomo, disminuyendo de 10.5µg/dl en 1992 a 7.5µg/dl en 1995.

Se utilizó la regresión lineal múltiple para explicar mejor esa tendencia, añadiendo las covariables con las que se puede entender mejor la relación entre el plomo en sangre y el plomo en aire, pues reduce la variabilidad no explicada del plomo en sangre y su inclusión es determinante de la validez de los resultados.

Las variables que se utilizan son las siguientes:

- 1) Humanos, con el binomio madre-hijo. Para tal efecto se toman muestras de sangre de la madre y del cordón umbilical, en el momento del parto. Dichos datos fueron proporcionados por el Instituto Nacional de Salud Pública (INSP) con información de 1992 a 1995, así como una encuesta efectuada cuando las madres eran gestantes.
- 2) Información de los niveles de plomo en aire, proporcionada por la Red Manual de Monitoreo Atmosférico de 1987 a 1995.
- 3) Información sobre el consumo mensual en millones de litros de gasolinas Nova, Extra y Magna-Sin dentro de la Zona Metropolitana de la Ciudad de México (ZMCM), contenido de plomo y niveles atmosféricos, proporcionada por el Instituto Mexicano del Petróleo (IMP) de 1986 a 1995.

Los programas utilizados para el proceso y análisis estadístico de la información son:

- 1) El programa Dbase IV para el procesamiento de la información y la creación de un programa que asigna los niveles de plomo en aire en la fecha en que nació el producto, así como el cálculo de promedios de uno, dos y tres meses antes de la fecha del parto. (Apéndice A)
- 2) Uso del programa STATA® para realizar el análisis estadístico de la información y para las representaciones gráficas.

OBJETIVO GENERAL.

Describir la tendencia de los niveles y concentraciones de plomo en sangre de las mujeres que están embarazadas de la Zona Metropolitana de la Ciudad de México, relacionandolos con los niveles de plomo en aire.

OBJETIVOS ESPECIFICOS.

- 1) Describir como ha cambiado el contenido del plomo en el aire y si éste cambio se refleja en el contenido de plomo en la sangre.
- 2) Evaluar el cambio que los niveles de plomo en sangre pueden tener al relacionarlos con otros factores como son: el uso de barro vidriado, la edad de las mujeres embarazadas y su grado de escolaridad.
- 3) Evaluar la asociación de los niveles de plomo en aire con el consumo mensual en millones de litros de gasolinas Nova, Extra y Magna-sin.
- 4) Creación de un programa automatizado que calcule los niveles de plomo en aire para relacionarlos con los niveles de plomo en sangre, dependiendo de la fecha en que se tomo la muestra de sangre y la Delegación o Municipio donde vivía la madre, así como el cálculos de los niveles promedio de plomo en aire de uno, dos y tres meses antes de la fecha de la toma de muestra. Se eligen éstos periodos de tiempo, ya que en estudios previos se ha demostrado que la exposición ambiental que mejor refleja la exposición al plomo esta dada por las medias mensuales y trimestrales. Dicho programa se podrá actualizar conforme se tengan más mediciones de plomo en aire.
- 5) Localización de cuales son las zonas con mayor riesgo para tener niveles de plomo en sangre elevados.

CAPITULO I

MARCO TEORICO

A. La Salud Ambiental.

La salud ambiental incorpora todos los aspectos del ambiente natural y de las transformaciones sobre él efectuadas por el hombre, que entrañan algún peligro o que tienen algún tipo de efecto indeseable en la salud o en la seguridad humanas y que pueden afectar su bienestar. El grado de contaminación ambiental tiene efectos tóxicos sobre la salud humana, pero no se distribuye uniformemente en la población expuesta. Cada individuo de la comunidad sometida a exposición ambiental, tiene un riesgo diferente de experimentar efectos dañinos sobre su salud, dado por el grado de exposición al contaminante, por la medida en que sus propias defensas contrarresten el efecto tóxico y por la aminoración del efecto mediante mecanismos homeostáticos(*) y reacciones fisiológicas de adaptación. Se pueden llevar a cabo medidas correctivas de la contaminación ambiental en los sistemas ambientales, lo que daría por consecuencia una reducción de los efectos tóxicos del ambiente sobre las personas.(15)

Las formas básicas por las que los contaminantes entran al medio ambiente y se ponen en contacto con el hombre:

- 1. Exposición directa a la fuente.** Se da con mayor frecuencia en trabajadores que por su labor se encuentran expuestos a elevados niveles de sustancias tóxicas. También puede ser en la población general por la utilización del DDT como insecticida doméstico o de pinturas para interiores con alto contenido en plomo.
- 2. Descarga directa al aire o al agua.** Este tipo de descarga fué durante muchos años la manera habitual de eliminar los productos residuales, sin reparar en el efecto que ello podría tener. El humo de las chimeneas caseras e industriales arrojado al aire, las emisiones sin control de vehículos automotores y el vaciado de aguas negras sin tratamiento a los ríos son ejemplo de ello.
- 3. Los depósitos o rellenos inapropiados para desechos.** Fueron creados con el fin de no contaminar el medio ambiente; sin embargo, la lluvia arrastra sustancias peligrosas que se vierten en los ríos y lagos, los que a su vez suministran agua para el consumo humano. También por la lluvia se pasan a los mantos acuíferos.(15)

B. La Contaminación del Aire.

La contaminación del aire se define como: "...la presencia en la atmósfera exterior de uno o más contaminantes o sus combinaciones que por su cantidad y duración puedan afectar la vida humana, de animales, la propiedad o el ejercicio de las actividades..." (14)

La contaminación del aire forma parte de nuestra vida, mas no es un fenómeno reciente. En 1272 el Rey Eduardo I de Inglaterra prohibió el uso del carbón marítimo, llamado así porque se transportaba por vía marítima, para despejar los cielos llenos de humo sobre Londres. La contaminación y degradación del ambiente está asociada con los beneficios de la sociedad tecnológica y también por procesos naturales; por ejemplo incendios forestales y erupciones volcánicas. Los contaminantes producidos por el hombre, se encuentran en regiones geográficas de poca extensión y por lo tanto, la mayor parte de la contaminación la provoca él mismo.(14)

La masa de un contaminante se expresa en microgramos de contaminante por metro cúbico de aire ($\mu\text{g}/\text{m}^3$) y su clasificación general es la siguiente:

1. Materia particulada o partículas.
2. Compuestos que contienen azufre.
3. Compuestos orgánicos.
4. Compuestos que contienen nitrógeno.
5. Monóxido de carbono
6. Compuestos halogenados.
7. Compuestos radiactivos.

C. EL PLOMO (Pb)

El plomo es un metal pesado cuyos efectos sobre el ambiente y la salud humana han sido estudiados en forma muy amplia, debido a la antigüedad de su utilización, a la variedad de sus aplicaciones y a su toxicidad.

Se utiliza en procesos industriales muy diversos como en la fabricación de acumuladores eléctricos, baterías, la producción de aditivos para combustibles con alquilatos de plomo como la gasolina, la elaboración de pigmentos de pinturas y barnices, etc.(15). También se aplica en productos metálicos como soldadura y tuberías.

1. La Contaminación por plomo y principales fuentes emisoras.

En muchas industrias se manejan sustancias potencialmente tóxicas y el grado de riesgo del individuo depende de la magnitud y el tiempo de la exposición. Dentro de estos agentes químicos se encuentra el plomo, el cual se presenta como polvo o como vapores en el aire en el lugar de trabajo o en las minas. Es más común en fundidoras de plomo, en la producción o reparación de acumuladores, en la industria automotriz, en fábricas de cerámica, de vidrio, de pinturas y en talleres de impresión. También en gasolineras por ser un componente de la gasolina aunque esto ha disminuido en los últimos años.(15)

De la misma manera se utiliza en la fabricación de pinturas y pigmentos para dar la coloración amarilla a algunas pinturas. Los óxidos de plomo dan una coloración roja a las pinturas a las que se les adiciona. El sulfuro de plomo es un componente de los pigmentos azules. El plomo, en forma de carbonato, da durabilidad a la pintura y se le adiciona a ésta para acelerar el secado. Las pinturas a base de plomo, además de recubrimientos de paredes, pueden utilizarse en juguetes y artículos escolares como lápices.(2)

a. Aire.

El plomo se encuentra contenido en las partículas. Partícula es el término empleado para describir las materias sólidas, líquidas, dispersas y arrastradas por el aire. Miden de 0.0002 μ g hasta 500 μ g (1 μ g= 1 micrón = 10⁻⁴ cms). Las partículas tienen una vida media de suspensión en el aire de unos cuantos segundos hasta varios meses.

Efectos de las partículas en el aire sobre la salud humana.

Los contaminantes entran principalmente al cuerpo humano por las vías respiratorias. Los daños a los órganos respiratorios pueden presentarse directamente, ya que se ha estimado que más del 50% de las partículas entre 0.01 y 0.1 μg que penetran a los pulmones se depositarán allí (13). La absorción de plomo por vía pulmonar es más peligrosa que la oral, debido a que el tamaño de la partícula es menor a un micrón, siendo fácilmente incorporado a través de los alvéolos al torrente sanguíneo.

Las partículas sólidas que existen en la atmósfera son generadas por la combustión de carbón, combustóleo, gasolina, aceite, diesel y combustible para motores de turbina en fuentes móviles.

Según la Norma Federal(*) para la calidad del Aire Ambiente el plomo debe tener un promedio inferior a 1.5 $\mu\text{g}/\text{m}^3$ por espacio de 3 meses y el método adecuado para medirlo es el quimoluminiscente.(*).

El uso de gasolina con plomo en vehículos de motor, cuya densidad en las áreas urbanas alcanza números elevados, son los responsables de los contaminantes, tales como el monóxido de carbono, óxidos de nitrógeno, **plomo** y partículas suspendidas, siendo también la fuente predominante de contaminación del aire. Las concentraciones de dichas sustancias puede variar de acuerdo con el uso del vehículo. La liberación de éste metal también es consecuencia de la actividad industrial y el crecimiento de las ciudades.

Otras fuentes emisoras de plomo a la atmósfera son las industrias que fabrican baterías, las fábricas de pinturas, soldadura, cables, municiones, industria del hierro y el acero, la de petróleo y petroquímica, la textil, la de celulosa y papel, y la metalúrgica (2). Por consecuencia las poblaciones urbanas que habitan en las zonas que rodean a éstas industrias presentan concentraciones elevadas de *plomo en sangre* (15), debido a la inhalación de humos y a la ingestión de partículas del metal que contaminan el agua y los alimentos.

b. Suelo.

El plomo en el aire se sedimenta y forma entonces parte de las partículas y del polvo que se encuentra en las viviendas, o sedimentado en el suelo. El plomo no se biodegrada cuando se deposita en el polvo o en el suelo, por lo que puede ser una fuente de exposición a largo plazo.(2)

c. Agua.

Las aguas residuales pueden contener muchos contaminantes y entre éstos, los contaminantes químicos, específicamente los metales pesados que son electronegativos y que poseen una densidad superior a 5g/cm^3 . Entre ellos está el plomo, que al igual que los plaguicidas no son eliminados, lo que hace que a medida que evoluciona la cadena trófica(*), aumente su concentración en el cuerpo de los animales, a consecuencia del fenómeno de bioacumulación(*). (15)

El agua potable puede ser contaminada por plomo al ser transportada en tuberías hechas de dicho metal, o hechas de cobre con soldadura de plomo.

d. Alimentos.

El movimiento de los contaminantes a través del agua, del aire, del suelo y de los seres vivos, se da cuando un contaminante es incorporado al ambiente. Se dispersa en el medio que lo recibe e interactúa con los elementos que integran a ese ambiente y su concentración se modifica a su paso por los diferentes niveles en la cadena de alimentos, dando lugar a la acumulación de un contaminante en los tejidos de un organismo animal o vegetal. Los contaminantes vertidos en ríos, lagunas y mares, o depositados desde la atmósfera, pueden entrar en la cadena de alimentos por conducto de los peces o de las plantas con los que éstos se alimentan. Los moluscos, como los ostiones y las almejas, filtran grandes volúmenes de agua para obtener su alimento, por lo que concentran sustancias tóxicas o radiactivas que se encuentran en el agua en pequeñas cantidades. En México en las lagunas litorales de Tabasco, se han reportado en algunas especies, una concentración importante de hidrocarburos y metales pesados, debido a la explotación del petróleo. Tal es el caso del plomo que se encuentra en concentraciones elevadas en ostiones. (15)

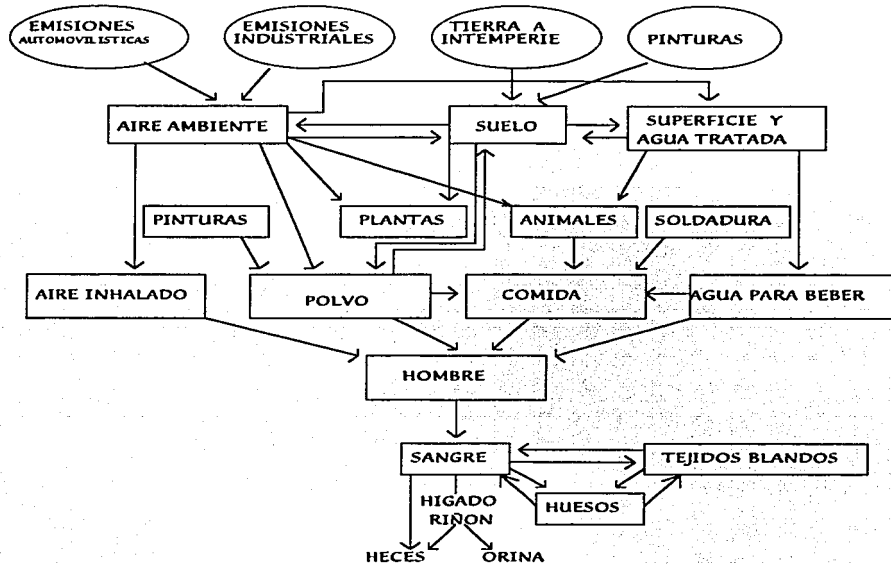
La contaminación de los alimentos también se puede dar en su manufactura. Tal es el caso de los alimentos enlatados cuando las latas se encuentran soldadas con plomo, el metal se desprende y produce contaminación. Los niveles de plomo pueden ser de cinco a veinte veces superiores a las cantidades que se encuentran en los alimentos frescos similares; sin embargo, es pertinente mencionar que en México la fabricación de latas con soldadura de plomo está prohibida. La cerámica vidriada y horneada a baja temperatura (menos de 990°C) es una fuente importante de exposición a plomo. El plomo contenido en el barniz del vidriado puede ser desprendido por acción de los alimentos, principalmente ácidos, y así contaminar los alimentos. (2)

El desprendimiento de plomo presente en el barniz se da por:

1. Tipo de cerámica
2. Tiempo de uso del utensilio.
3. Grado de acidez de los alimentos.
4. Tiempo de contacto del alimento con el vidriado. (2)

También se puede encontrar plomo en la leche materna. Una de las mayores fuentes de contaminación en frutas y legumbres es por la deposición del plomo de los automotores (por el tetraetilo de Pb usado como antidetonante) en sembradíos cercanos a carreteras o zonas industrializadas. (15)

ESQUEMA DEL CAMINO QUE SIGUE EL PLOMO DESDE EL MEDIO AMBIENTE AL INTERIOR DEL HOMBRE.



FUENTE. EPA, 1986 (17)

2. Intoxicación por plomo y efectos en la salud.

Al plomo se ha considerado como una posible explicación para la caída del Imperio Grecoromano ya que una práctica aceptada en esa época era la de usar vasijas y recipientes de plomo para almacenar alimentos y bebidas (especialmente el vino), ocasionando una *intoxicación crónica*. En la época de Hipócrates y Plinio ya eran conocidos algunos de los síntomas que ocasiona el saturnismo(*).(5)

El organismo humano absorbe plomo a través de las vías digestivas y respiratorias principalmente; en menor medida por la piel. El metal se deposita en los huesos, la médula ósea, el cerebro y los nervios periféricos a través de la sangre. La intoxicación es crónica, ya que las manifestaciones clínicas del envenenamiento se presentan después de semanas o meses de exposición. En ocasiones transcurren periodos de años para que las alteraciones sean evidentes.

La intoxicación por plomo produce los siguientes efectos en la salud:

- Disminución del desarrollo intelectual
- Problemas con eritrocitos
- Deficiencia de las delta amino levulínico dehidrasa en la síntesis del grupo hemo(*)
- Anemia
- Dolor abdominal intenso
- Encefalopatía (daño al sistema nervioso central), aguda o crónica.
- Neuropatía periférica(*)

Una exposición continua y prolongada causa lesiones en el riñón y con el tiempo puede producir insuficiencia renal. Los síntomas de la encefalopatía son vómitos esporádicos, periodos de confusión e irritabilidad; si continúa la exposición se presentan alucinaciones, excitabilidad, pérdida de la memoria y dificultades en los movimientos. Las alteraciones del sistema nervioso central son generalmente irreversibles y dejan secuelas permanentes, causando convulsiones epilépticas, pérdida de equilibrio y los nervios oculares son afectados con un decremento en la visión (15).

Los efectos del plomo son más graves en los niños. La gravedad de las alteraciones que provoca el plomo en la infancia, se debe a su capacidad para provocar lesiones permanentes en el sistema nervioso central, que abarca desde las alteraciones del comportamiento y diversos grados de disminución de las capacidades intelectuales, hasta el retardo mental profundo.

El plomo puede provocar abortos, disminuir la fertilidad femenina, producir partos prematuros y muertes fetales. Investigaciones con animales, han demostrado los efectos del plomo sobre los órganos reproductores y sus funciones. Puede causar anomalías en el espermatozoides y disminución de la fertilidad masculina en trabajadores expuestos.(15)

Actualmente se reconoce que no existe un nivel seguro de exposición, por debajo del cual se puede considerar que no existe daño. Es por esto que existe un esfuerzo mundial para disminuir y si es posible abatir la emisión de plomo al medio ambiente.

CAPITULO II

ANALISIS EXPLORATORIO DE DATOS

El análisis exploratorio de datos se utiliza en las primeras fases del análisis estadístico, éste se realiza mediante la aplicación de técnicas estadísticas y gráficas, que nos permiten tener una idea global de la información, distribución y dispersión de los datos (8). Las técnicas de análisis exploratorio sirven para:

a) Evaluar la calidad y consistencia de la información

Paso que debe ser realizado antes de iniciar el análisis estadístico. En el campo de la epidemiología ambiental se recolecta información sobre un gran número de variables, ya sea mediante cuestionarios o mediante instrumentos de medición ambiental. Si se utilizan datos de fuentes secundarias que no están sujetos a controles de calidad estrictos es necesario realizar evaluaciones completas, con el propósito de encontrar y estudiar los valores no plausibles.

b) Investigar la distribución de las variables de interés

Se pueden detectar asimetrías, rangos observados y valores máximos y mínimos.

c) Investigar adherencia a las suposiciones estadísticas necesarias, en etapas posteriores del análisis

Esta información es importante, ya que muchas de las técnicas estadísticas asumen una serie de suposiciones sobre el comportamiento y distribución en las variables en estudio.

d) Resumir información mediante diferentes estadísticas y gráficas

Con ello se tiene un panorama general de la información y se puede ver rápidamente cuales son las características y el comportamiento.

e) Evaluar la necesidad de realizar transformación de las variables de interés

Si las variables de interés no cumplen con los supuestos estadísticos que requieran posteriores pruebas estadísticas. Por ejemplo, la normalidad para la correlación y para la regresión lineal.

f) Detectar valores discrepantes (outliers) valores aberrantes o errores, no plausibles

Mediante técnicas estadísticas, se pueden detectar valores discrepantes, que son observaciones que se diferencian en forma importante de la nube de puntos. Es importante poder decidir si se trata de una observación con plausibilidad biológica o de una observación que queda fuera del rango biológico de observación.

g) Explorar formas de categorizar variables (puntos de corte)

Mediante puntos de corte para los percentiles más utilizados

A. Análisis exploratorio en los datos de la Red Manual de Monitoreo Atmosférico

El primer análisis exploratorio fué en la base proporcionada por la Red Manual de Monitoreo Atmosférico (RMMA) ya que era preciso que estos fueran consistentes y confiables de utilizar.

Nombre de la base: AIRE.DTA

.describe

Contains data from C:\AIRE.DTA

Obs: 3951 (max= 39857)

Vars: 13 (max= 99)

1. edate	float	↖dn/d/Y	Fecha medición Plomo aire
2. mun	float	↖9.0g	MES medición Plomo aire
3. dia	float	↖9.0g	DIA medición Plomo aire
4. aa	float	↖9.0g	AÑO medición Plomo aire
5. estacion	str4	%s	ESTACION DE MONITOREO
6. filtro	float	↖7.0f	NUMERO de filtro
7. plomo	float	↖6.3f	Nivel de PLOMO en aire µg/m'
* 8. nova	float	↖5.1f	GASOLINA NOVA CONSUMO MILL/LTS
* 9. extra	float	↖5.2f	GASOLINA EXTRA CONSUMO MILL/LTS
*10. magna	float	↖9.0g	GASOLINA MAGNANIN CONSUMO MILL/LTS
*11. total	float	↖9.0g	CONSUMO TOTAL GASOLINAS MILL/LTS
*12. ton_pb	float	↖9.0g	TONELADAS PLOMO EMITIDAS ATMOSFERA
*13. pb_atm	float	↖9.0g	µ MENSUAL DE PLOMO EMITIDO A ATM.

La instrucción **summarize** proporciona un panorama general de la información mediante estadísticas como la media, desviación estándar y los valores máximo y mínimo. Además proporciona el número de observaciones con que fueron calculadas dichas estadísticas que resumen la información.

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
edate	3951	11549.98	813.7756	9507	12868
mm	3951	5.913693	3.574366	1	12
dia	3951	15.36522	8.735291	1	31
aa	3951	1991.174	2.207937	1986	1995
filtro	744	6026577	525	6025553	6027497
plomo	3951	0.802	0.814	0.009	8.870
nova	3945	415.0	53.5	295.5	513.6
extra	1398	12.88	12.13	1.38	38.94
magna	2467	129.9199	74.99649	13.31	244.5
total	3945	500.8121	47.23267	388	586.2
ton_pb	3855	81.12053	50.28434	18.952	319.02
pb_atm	3510	.8736578	.5883062	.222	3.624

A simple vista parece no haber problemas graves con la calidad de los datos; por ejemplo, no hay mes 13, ni 32 días por mes.

La variable principal a ser estudiada con detenimiento es **Plomo** ($\mu\text{g}/\text{m}^3$), que se refiere a el nivel de plomo en aire que había en la estación de monitoreo y es reportada por la Red Manual de Monitoreo Atmosférico.

. sum plomo

Variable	Obs	Mean	Std. Dev.	Min	Max
plomo	3951	0.802	0.814	0.009	8.870

De el año de 1986 a 1995 hubo 3951 mediciones de plomo en aire en total con una media de **0.802 $\mu\text{g}/\text{m}^3$** y desviación estándar de **0.814 $\mu\text{g}/\text{m}^3$** . El valor mínimo reportado fue de **0.009 $\mu\text{g}/\text{m}^3$** y el más alto reportado fue de **8.870 $\mu\text{g}/\text{m}^3$** .

Con más detalle para poder imaginar su distribución:

. sum plomo,detail

		plomo			

	Percentiles	Smallest			
1%	0.046	0.009			
5%	0.109	0.012			
10%	0.150	0.013	Obs		3951
25%	0.263	0.014	Sum of Wgt.		3951
50%	0.600		Mean		0.802
		Largest	Std. Dev.		0.814
75%	1.042	7.600	Variance		0.663
90%	1.631	8.389	Skewness		3.198
95%	2.260	8.784	Kurtosis		20.297
99%	4.050	8.870			

Con esta instrucción tenemos información de la media (0.802) y la mediana (0.600), de la desviación estandar (0.814) y varianza (0.663), valores máximos (8.870) y mínimos (0.009); así como los puntos de corte para los percentiles más utilizados. Se enlistan los cuatro valores más pequeños (smallest) y los cuatro valores más grandes (largest). Además de ello, tenemos información sobre la simetría de la distribución (*skewness*) y el grado de curtosis (*kurtosis*).

Se observa que hay un sesgo positivo ya que la **mediana < media**. Hay asimetría a la derecha o sesgo positivo, por lo que la distribución no es normal. Para que existiera una distribución normal, la media y la mediana serían iguales y el *sesgo* tendría valor 0 (cero) y la *curtosis* un valor de 3.

El siguiente gráfico de letras (lv) evalúa el impacto de los extremos de la distribución, "de las colas" de los datos, asumiendo diferentes puntos de corte. El nombre del diagrama de letras se origina en el hecho de que a cada punto de corte se le ha asignado una letra:

M MEDIANA	0.5	1/2	JUSTO MITAD OBSERVACIONES		
F CUARTOS	0.25	1/4	PERCENTILES	25	Y 75
E OCTAVOS	0.125	1/8	PERCENTILES	12.5	Y 87.5
D DIECISEISAVOS	0.0625	1/16	PERCENTILES	6.25	Y 93.75
C TREINTAIDOSAVOS	0.03125	1/32	PERCENTILES	3.125	Y 96.87
B SESENTAICUATROAVOS	0.01562	1/64	PERCENTILES	1.56	Y 98.44
A CIENTOVEINTIOCHOAVOS	0.00781	1/128	PERCENTILES	0.78	Y 99.22
Z DOSCIENTOSCINCUENTAISEISAVOS	0.00390	1/256	PERCENTILES	0.39	Y 99.61
Y QUINIENTOSDOCEAVOS	0.00195	1/512	PERCENTILES	0.20	Y 99.80
X MILVENTICUATROAVOS	0.00097	1/1024	PERCENTILES	0.10	Y 99.90

El diagrama de letras para plomo en aire:

. lv plomo

#	3951	plomo			spread	pseudosigma
M	1976		0.600			
F	988.5	0.263	0.652	1.041	0.778	0.577
E	494.5	0.170	0.815	1.461	1.291	0.561
D	247.5	0.122	1.073	2.023	1.901	0.620
C	124	0.084	1.405	2.727	2.643	0.710
B	62.5	0.059	1.778	3.497	3.439	0.799
A	31.5	0.042	2.141	4.240	4.198	0.869
Z	16	0.030	2.760	5.490	5.460	1.028
Y	8.5	0.017	3.368	6.719	6.702	1.168
X	4.5	0.015	3.775	7.536	7.521	1.223
	1	0.009	4.439	8.870	8.861	1.240
inner fence		-0.904		2.209	# below	# above
outer fence		-2.072		3.376	0	210
					0	67

lv muestra que hay un total de 3951 observaciones con valor de plomo diferente de cero, M contiene el valor de la mediana 0.600 y son 1976 las observaciones que se encuentran debajo de la mediana. La línea de la F muestra las estadísticas asociadas a los cuartiles, el 0.263 y 1.041 marcan los valores límites para el cuartil inferior (percentil 25) y el superior (percentil 75). La cifra de 988.5 indica que por debajo y por arriba de estos puntos de corte quedan 1977 observaciones (988.5 en cada extremo). El valor de 0.652 indica el punto medio de las observaciones que quedan entre el punto de corte inferior y el punto de corte superior. Si la distribución fuese perfectamente simétrica, es decir normal, se esperaría que los puntos medios de los diferentes puntos de corte fueran iguales a la mediana.

La estadística que indica una medida de dispersión (**spread**), describe la separación entre los datos numéricos para cada punto de corte y se obtiene de calcular la diferencia entre el límite inferior del superior.

La estadística que estima la desviación estándar (**pseudosigma**) es la medida de la dispersión de los datos alrededor del punto medio para cada punto de corte. La pseudosigma se obtiene de dividir la dispersión (spread) entre la desviación derivada de una distribución normal. Si la variable tiene una distribución normal, los valores para los diferentes puntos de corte deben ser similares; en este caso la pseudosigma es creciente, lo que indica *mayor dispersión* que la distribución normal y por tanto una asimetría en la distribución.

La parte inferior proporciona información sobre los valores que se encuentran separados de la nube de puntos, es decir, separados de tal manera que se consideran errores potenciales. Será también importante detectar si las 210 observaciones por arriba de 2.209 son errores o son valores plausibles, ya que pueden tener un impacto importante en los resultados y conclusiones. Investigadores de renombre internacional (8) sugieren hacer el análisis estadístico con éstos datos y sin ellos, para evaluar la diferencia en los resultados antes de decidir borrarlos anticipadamente, pues podrían ser las observaciones más interesantes de toda la base de datos.

Lv maneja dos puntos de corte: **inner fence** (límite interno) y **outer fence** (límite externo), basados en el rango intercuartil, dado que es una medida robusta(*) que no se afecta por la presencia de valores extremos. El límite interno identifica a los puntos que podrían ser considerados como valores aberrantes y el límite externo identifica los valores que tienen una alta probabilidad de ser valores aberrantes.

Los límites internos y externos se definen (8):

Diferencia intercuartil	$DI = C75 - C25$
Límite interior inferior	$LI_i = C25 - 1.5 * DI$
Límite interior superior	$LI_s = C75 + 1.5 * DI$
Límite externo inferior	$LE_i = C25 - 3.0 * DI$
Límite externo superior	$LE_s = C75 + 3.0 * DI$

En el diagrama de letras de plomo en aire hay 210 valores aberrantes y 67 observaciones con una alta probabilidad de ser valores aberrantes. Para identificar las 210 observaciones que podrían ser aberrantes y donde estarían incluidas las 67 observaciones, le pedimos a STATA® que cuente las mediciones de plomo en aire por arriba de $2.209 \mu\text{g}/\text{m}^3$:

ESTACION DE MONITOREO

TOTAL DE OBSERVACIONES>2.209

CES	6
CFE	7
FAN	11
LPR	1
MCM	11
MER	12
PED	5
TEC	22
TLA	10
XAL	125

Se observó que el 59.5% de las mediciones fueron realizadas en la estación de monitoreo de XALOSTOC (XAL), el 10.5% con las mediciones del Cerro del Tepeyac (TEC). Ambas estaciones de monitoreo se ubican en la zona Noreste. La estación de Xalostoc está en una zona industrial, por lo que siempre se ha caracterizado por ser la que tiene niveles de plomo en aire más elevados en comparación con todas las demás estaciones de monitoreo y que constantemente rebasaba las normas permitidas ($1.5\mu\text{g}/\text{m}^3$). Por éstas razones, se consideraron valores plausibles de haber ocurrido y se dejaron para los cálculos del programa.

Frecuencias de valores de plomo mayores a $2.209\mu\text{g}/\text{m}^3$ por año en dichas estaciones de monitoreo.

. tab (aa) if plomo>2.209, summ (plomo)

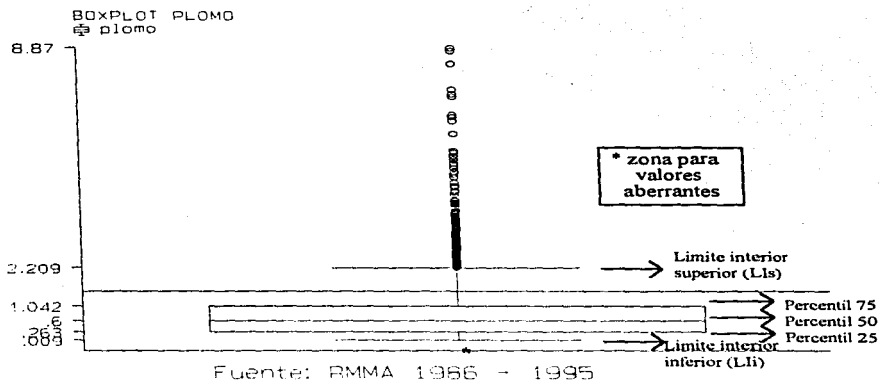
aa	Summary of plomo		Freq.
	Mean	Std. Dev.	
1986	2.247	0.000	1
1987	2.686	0.361	9
1988	3.469	1.433	77
1989	3.352	1.342	48
1990	3.239	1.062	36
1991	3.435	1.311	26
1992	3.141	0.659	8
1993	2.364	0.071	2
1994	2.720	0.623	3
Total	3.325	1.272	210

La mayoría de éstas observaciones se concentraron en los años de 1988 a 1991, años en que ese tipo de valores eran posibles y frecuentes por lo que se consideran observaciones dentro del rango esperado.

Gráficas de caja (boxplots):

Este tipo de gráfica es una representación simple de la información y visualmente permite:

- * Localizar el centro de los datos
- * Ver la dispersión de los datos
- * La simetría
- * La extensión de los extremos (colas de la distribución)
- * La existencia de valores aberrantes (outliers)



Revisando las estadísticas anteriores de **sum plomo**, **detail** o **lv plomo**, se puede observar que los datos más representativos de la variable plomo, los representa visualmente el comando **graph plomo,box**.

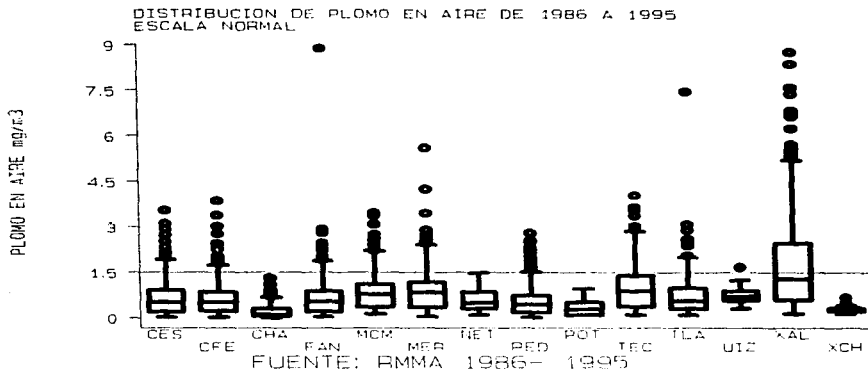
donde:

Lli = Percentil 25 - 1.5 * (diferencia intercuartil)

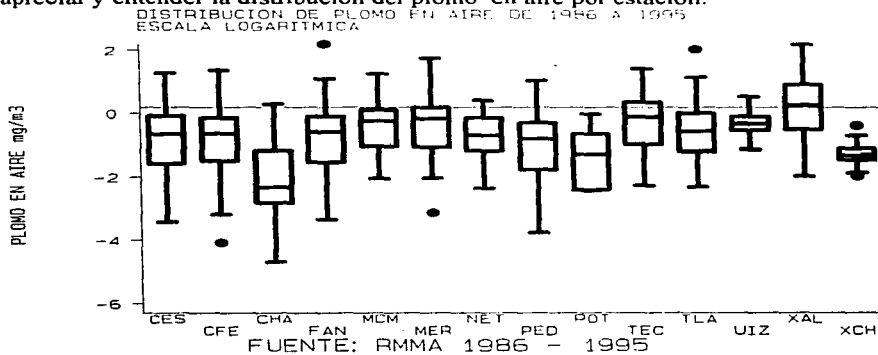
Lls = Percentil 75 + 1.5 * (diferencia intercuartil)

La ventaja del diagrama de caja basado en los rangos intercuartiles, es que es resistente al impacto de valores extremos.

Esta gráfica es una herramienta para hacer comparaciones entre diferentes categorías. La siguiente gráfica de caja permite comparar los niveles de plomo en aire ($\mu\text{g}/\text{m}^3$) en cada una de las estaciones de monitoreo de la RMMA.

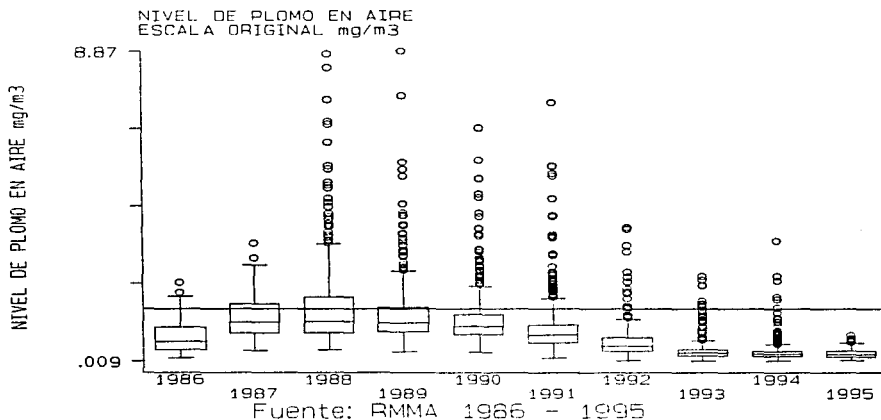


Se aprecia que en la estación de monitoreo de Xalostoc la mitad de las mediciones están por arriba de la norma, la línea que hace intersección en $1.5\mu\text{g}/\text{m}^3$ en el eje de las Y's, por ello se transformó la variable plomo a escala logarítmica para re-exresar los valores observados y lograr una dispersión similar que nos permita apreciar y entender la distribución del plomo en aire por estación.



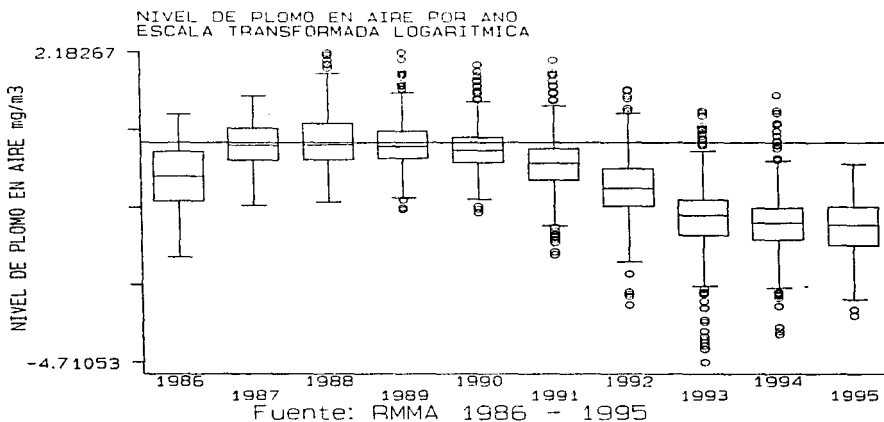
Este gráfico de caja en escala logarítmica, permite comparar de manera más adecuada los valores de plomo en aire entre las diferentes estaciones de monitoreo del aire de la Ciudad de México. La línea que intersecta el eje de las Y's es $\log(1.5)=0.18$

También basado en la exploración de los datos, se puede decir de antemano si algunas de nuestras hipótesis serán verdaderas. Ejemplo de ello es que el nivel de plomo en aire ha disminuido al pasar de los años, pues hay autoridades que están haciendo que se cumplan y respeten las normas de calidad del aire, así como una "mejoría" de la gasolina con la venta de la Magna Sin (sin plomo).



El hecho de que en 1986 y 1987, parezca que no hayan sido elevados los niveles de plomo en aire, se puede explicar por el número de mediciones que se hicieron en esos años con 65 y 95 muestreos respectivamente (figura 1, capítulo 5). Se observa también que en 1990 más del 75% de los niveles de plomo en aire se mantuvieron por debajo de la norma de $1.5\mu\text{g}/\text{m}^3$, y a partir de entonces, las concentraciones de plomo en aire han disminuido. La explicación de ello puede ser que en Octubre de 1990 se reemplazó la Gasolina EXTRA por la MAGNA SIN.

Se puede observar que existe una relación entre la dispersión y el tiempo. En años recientes se observa una dispersión menor del plomo en el aire. Este patrón podría sugerir la necesidad de una transformación, es decir, de re-expresar los valores observados para lograr una dispersión similar, una igualdad de varianzas dentro de los diferentes grupos de comparación, consiguiendo una mejor representación gráfica y datos más apropiados para los análisis estadísticos tradicionales. Por ello se transforma la variable plomo a escala logarítmica $\log(\text{plomo})$:

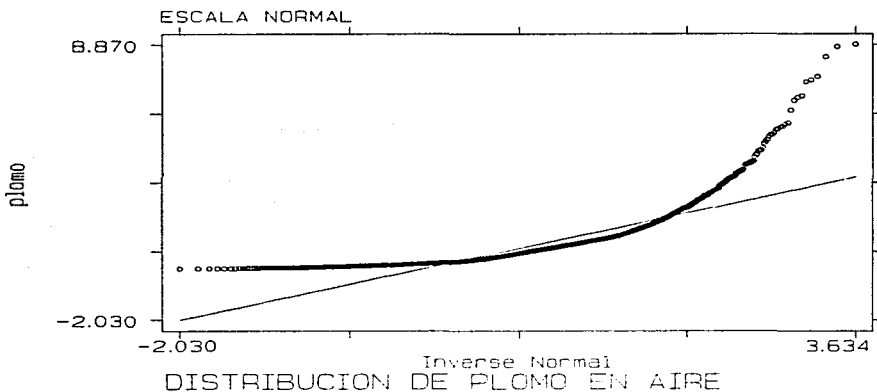


Con la escala transformada se aprecia normalidad en cada año así como la disminución de plomo en aire y se ve también que no es sino hasta 1991, cuando más del 75% de las observaciones están por debajo de la norma, lo que es más creíble que la gráfica en escala original, pues si bien fue en Octubre de 1990 que salió a la venta la gasolina Magna, el consumo de millones de litros de ésta gasolina era sólo del 4.30%, mientras que la gasolina Nova tenía el 95.70% de consumo. Gradualmente se ha ido incrementando el consumo de gasolina Magna y así, en Septiembre de 1995, los millones de litros de gasolina Magna consumidos representaban el 44.44% del consumo.

La elección de la transformación de la variable plomo como $\log(\text{plomo})$ a una variable logarítmica se explica a continuación:

Como se dijo anteriormente, con los estadísticos **summarize** y **lv**, se encontró que la variable plomo no tenía una distribución normal, pues tenía un sesgo positivo. Para el análisis estadístico posterior, como la correlación y la regresión lineal, es necesario que la variable plomo tenga una distribución normal, pero existen otros métodos para evaluar la normalidad. Probablemente el diagnóstico gráfico de normalidad más utilizado es el gráfico de la variable original, en relación a su transformación como una variable normalizada. Este gráfico se puede construir graficando la variable original versus la variable transformada ($F[(x_i - \mu_i)/\sigma]$). (8)

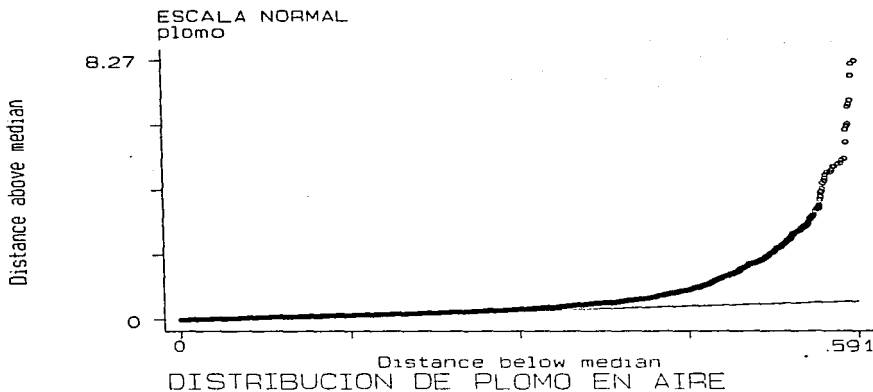
. qnorm plomo, t1(ESCALA NORMAL) title(DISTRIBUCION DE PLOMO EN AIRE)



Este gráfico de normalidad muestra una fuerte asimetría dado principalmente en las colas de la distribución. Si la distribución fuera normal, los datos de plomo (círculos de color gris) quedarían exactamente sobre la línea inclinada.

El siguiente gráfico también nos da una idea de la normalidad de la variable plomo:

. symplot plomo, t1(ESCALA NORMAL) title(DISTRIBUCION DE PLOMO EN AIRE)



Esta evaluación visual muestra a la variable plomo que no tiene una distribución normal y si lo fuera los datos de plomo (círculos de color gris) quedarían por debajo de la línea inclinada.

La estadística para realizar la prueba de normalidad es `sktest`, esta prueba se basa en la curvatura (curtosis) y la simetría (skewness). Para que la variable plomo fuera simétrica o normal la ji-cuadrada tendría un valor pequeño diferente de cero y la probabilidad de la ji-cuadrada debe ser no significativa; es decir $p > 0.05$

. `sktest plomo`

```

Skewness/Kurtosis tests for Normality
-----
Variable | Pr(Skewness) | Pr(Kurtosis) | adj chi-sq(2) | Pr(chi-sq)
-----|-----|-----|-----|-----
plomo | 0.000 | 0.000 | . | .

```

Así que le pedimos a STATA® proporcione sugerencias de que transformación haría que la variable plomo se normalice.

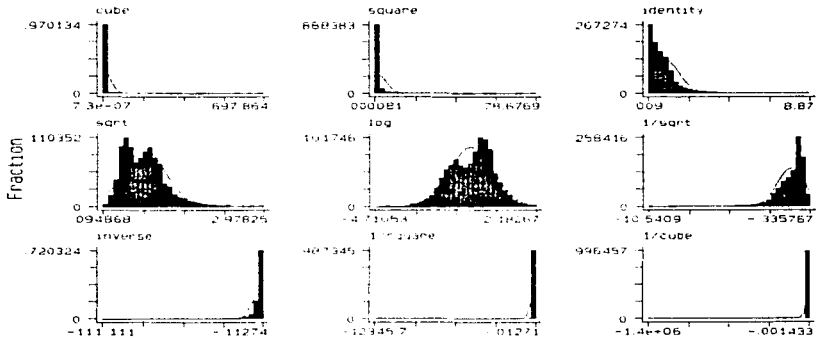
El comando **ladder** nos proporciona diferentes transformaciones que pueden re-expresar a la variable y evaluar cuál es la que logrará que tenga simetría o mejor aún lograr una distribución normal. (Ver capítulo IV 6. Transformaciones)

. ladder plomo

Transformation	formula	Chi-sq(2)	P (Chi-sq)
cube	plomo^3	.	.
square	plomo^2	.	.
raw	plomo	.	.
square-root	sqrt(plomo)	.	0.000
log	log(plomo)	60.90	0.000
reciprocal root	1/sqrt(plomo)	.	.
reciprocal	1/plomo	.	.
reciprocal square	1/(plomo^2)	.	.
reciprocal cube	1/(plomo^3)	.	.

Al parecer no existe ninguna transformación que normalice a esta variable; la transformación que lograría una distribución simétrica o normal sería aquella que en la ji-cuadrada sea menor, pero diferente de cero y su probabilidad asociada sea no significativa. Sin embargo, es muy probable que la transformación logarítmica sea la que haga que la variable plomo sea casi normal. Se pueden ver gráficamente las transformaciones antes dadas:

. gladder plomo



HISTOGRAMAS POR TRANSFORMACION

Es por ello que se seleccionó la transformación logarítmica para la variable plomo, puesto que es la que mejor ajusta los datos a una transformación normal:

($\ln\text{plomo} = \log(\text{plomo})_{\text{generated}}$)

Analizando la nueva variable generada **lnplomo** se puede ver si es simétrica y si se acerca a tener una distribución normal.

. summ lnplomo,detail

		log (plomo)			
Percentiles		Smallest			
1%	-3.079114	-4.710531		Obs	3951
5%	-2.216408	-4.422849		Sum of Wgt.	3951
10%	-1.89712	-4.342806			
25%	-1.335601	-4.268698			
50%	-.5108256			Mean	-.6376642
		Largest		Std. Dev.	.9605443
75%	.041142	2.028148			
90%	.4891934	2.126921		Variance	.9226454
95%	.8153648	2.172932		Skewness	-.3322254
99%	1.398717	2.182675		Kurtosis	3.105756

El summarize en la nueva variable generada, se puede observar que la distribución de **lnplomo** no es normal pero se acerca mucho a ser normal. La media y la mediana informan que hay un ligero sesgo negativo.

Diagrama de letras para log(plomo)

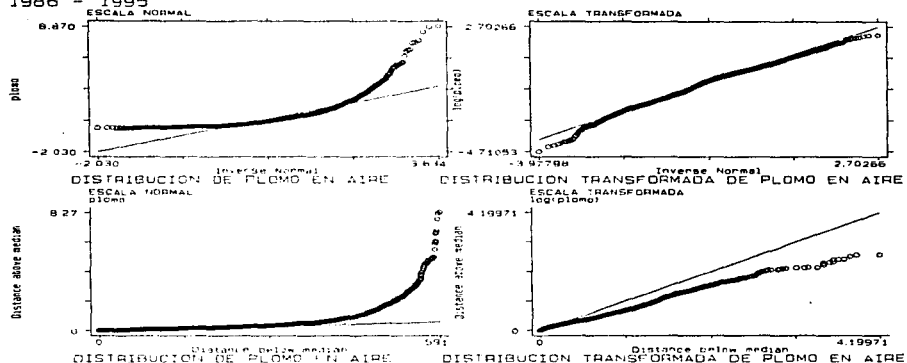
. lv lnplomo

#	3951	log (plomo)			
M	1976		-.5108256		spread
F	988.5	-1.335031	-.6472087	.0406139	1.375645
E	494.5	-1.771957	-.6964179	-.3791211	2.151078
D	247.5	-2.103325	-.6992479	-.7048286	2.808153
C	124	-2.476938	-.7368682	1.003202	3.480141
B	62.5	-2.836185	-.7920901	1.252005	4.088191
A	31.5	-3.166527	-.8609832	1.44456	4.611087
Z	16	-3.506558	-.9018149	1.702928	5.209486
Y	8.5	-4.054865	-1.074982	1.904902	5.959767
X	4.5	-4.234201	-1.107306	2.019588	6.25379
	1	-4.710531	-1.263928	2.182675	6.893206
					pseudosigma
					1.020169
					.935217
					.915488
					.9344877
					.9502439
					.9550427
					.9810547
					1.039091
					1.017062
					.9643958
					# below
inner fence	-3.398499		2.104081		18
outer fence	-5.461966		4.167549		0
					# above
					3
					0

Con la transformación se reducen en mucho los valores aberrantes. Según el límite inferior hay 18 valores que están por debajo de -3.398499 y 3 observaciones por arriba de 2.104081 .

Diagnósticos gráficos de normalidad de Plomo en aire en el período 1986 - 1995

EVALUACION DE LA DISTRIBUCION DE NORMALIDAD DEL PLOMO EN AIRE 1986 - 1995



Estas gráficas permiten comparar la distribución de la variable plomo antes, gráficos de la izquierda, qnorm plomo (arriba) y symplot plomo (abajo), y después de la transformación gráficos de la derecha, qnorm lnplomo (arriba) y symplot lnplomo (abajo).

Se toma el período de 1986 a 1995 completo pues será el lapso de tiempo con que se harán los análisis estadísticos de correlación, regresión lineal y múltiple posteriores que requieren que la distribución de plomo en aire sea normal.

B. Análisis exploratorio de las variables proporcionadas por el Instituto Mexicano del Petróleo 1995.

Empezando por los millones de litros de gasolina NOVA consumidos desde febrero de 1986 hasta Septiembre de 1995:

. summ nova,detail

		nova			
Percentiles		Smallest			
1%	294.20	278.10			
5%	295.70	294.20			
10%	331.80	295.20		Obs	116
25%	374.55	295.20		Sum of Wgt.	116
50%	428.95			Mean	412.18
		Largest		Std. Dev.	54.09
75%	447.70	498.80			
90%	477.40	501.30		Variance	2925.59
95%	483.20	511.60		Skewness	-0.58
99%	511.60	513.60		Kurtosis	2.60

Se tienen 116 observaciones, donde cada observación es la cantidad en millones de litros de gasolina *nova* consumidos en cada mes. La media de consumo fue de 412.18 millones de litros, la mediana o percentil 50 de 428.95. Hay sesgo negativo. La desviación de cada dato con respecto a la media fue de 54.09, el valor mínimo es de 278.10 y el máximo de 513.60.

Diagrama de letras para la gasolina nova:

. lv nova

#	116	nova				

M	58.5		428.95		spread	pseudosigma
F	29.5	374.55	411.12	447.70	73.15	54.41
E	15	335.50	401.65	467.80	132.30	57.77
D	8	317.50	399.55	481.60	164.10	54.45
C	4.5	295.35	394.12	492.90	197.55	54.83
B	2.5	294.70	400.58	506.45	211.75	50.83
A	1.5	286.15	399.38	512.60	226.45	48.69
	1	278.10	395.85	513.60	235.50	46.84
inner fence		264.83		557.42	# below	# above
outer fence		155.10		667.15	0	0

La distribución no es perfectamente normal. Si lo fuera, los puntos medios de cada punto de corte serían iguales a la mediana. El **pseudosigma**, estimación de la desviación estándar, confirma esto y al tener valores decrecientes se puede concluir que tiene menor dispersión que la distribución normal. Este comportamiento indica asimetrías en la distribución de la variable Nova.

Análisis exploratorio del consumo por millones de litros de gasolina EXTRA de febrero de 1987 a Septiembre de 1990:

. summ extra,detail

		extra			
Percentiles		Smallest			
1*	1.38		1.38		
5*	1.38		1.38		
10*	1.43		1.38	Obs	44
25*	1.74		1.40	Sum of Wgt.	44
50*	3.15			Mean	9.55
			Largest	Std. Dev.	11.74
75*	10.50		34.34		
90*	33.52		34.82	Variance	137.92
95*	34.82		35.99	Skewness	1.47
99*	38.94		38.94	Kurtosis	3.65

El número de observaciones de gasolina EXTRA son 44 con una media de 9.55 consumo en millones de litros y desviación estándar de 11.74; la mediana es de 3.15, por lo que hay un sesgo positivo. El valor mínimo es de 1.38 y el máximo de 38.94. La simetría y la curtosis muestran que la distribución no es normal.

Diagrama de letras para el consumo de gasolina extra:

. lv extra

#	44	extra			spread	pseudosigma
M	22.5		3.15			
F	11.5	1.74	6.12	10.50	8.76	6.55
E	6	1.48	17.50	33.52	32.04	14.09
D	3.5	1.39	17.98	34.58	33.19	11.33
C	2	1.38	18.69	35.99	34.61	9.73
B	1.5	1.38	19.42	37.47	36.09	8.31
	1	1.38	20.16	38.94	37.56	8.72
inner fence		-11.39		23.63	# below	# above
outer fence		-24.53		36.76	0	8
					0	1

El diagrama de letras igualmente muestra que no tiene una distribución normal, hay tendencia hacia los valores altos. Los posibles valores aberrantes son superiores a los 23.63 millones de litros. Hay un valor superior a los 36.76 millones de litros.

Análisis exploratorio del consumo por millones de litros de gasolina MAGNA SIN (Sin plomo) de octubre de 1990 a Septiembre de 1995:

. summ magna, detail

----- magna -----					
	Percentiles	Smallest			
1%	12.88	12.88			
5%	16.30	13.31			
10%	24.69	14.79	Obs		60
25%	62.60	17.81	Sum of Wgt.		60
50%	145.46		Mean		137.04
75%	211.30	Largest	Std. Dev.		78.44
90%	236.80	239.20	Variance		6153.13
95%	239.35	239.50	Skewness		-0.21
99%	244.50	241.30	Kurtosis		1.60
		244.50			

Hay 60 observaciones del consumo en millones de litros de gasolina Magna, con una media de 137.04 millones de litros consumidos y desviación estándar de 78.44; la mediana es de 145.46 lo que revela un sesgo negativo. La simetría y la curtosis muestran que la distribución no es normal.

Diagrama de letras para el consumo de gasolina magna:

. lv magna

#	60	----- magna -----				spread	pseudosigma
M	30.5		145.46				
F	15.5	62.60	136.95	211.30	148.70	110.94	
E	8	31.05	131.62	232.20	201.15	88.20	
D	4.5	19.16	128.97	238.78	219.62	74.05	
C	2.5	14.05	127.23	240.40	226.35	62.87	
B	1.5	13.10	128.00	242.90	229.81	55.57	
	1	12.88	128.69	244.50	231.62	50.95	
inner fence		-160.44		434.34	# below	# above	
outer fence		-383.48		657.39	0	0	
					0	0	

El lv nos reporta que hay 60 observaciones de la variable Magna. Con mediana de 145.46 millones de litros consumidos y debajo de la mediana hay 30.5 observaciones. La distribución no es perfectamente normal. Si lo fuera, los puntos medios de cada punto de corte serían iguales a la mediana. El pseudosigma, estimación de la desviación estándar, confirma esto. Al tener valores decrecientes se puede concluir que tiene menor dispersión que la distribución normal y, este comportamiento indica asimetrías en la distribución de la variable Magna.

Análisis exploratorio del TOTAL de millones de litros de gasolina consumidos de febrero de 1986 a Septiembre de 1995:

. summ total,detail

		total			
Percentiles		Smallest			
1%	391.38	388.00			
5%	399.80	391.38			
10%	418.76	393.00		Obs	116
25%	440.68	399.28		Sum of Wgt.	116
50%	490.36			Mean	486.68
		Largest		Std. Dev.	52.02
75%	530.93	569.70			
90%	555.82	577.80		Variance	2705.86
95%	564.20	581.25		Skewness	-0.03
99%	581.25	586.20		Kurtosis	1.86

Hay 116 observaciones del consumo total en millones de litros, con una media de 486.68 millones de litros de gasolina consumidos y desviación estándar de 52.02. La mediana es de 490.36, lo que revela un sesgo negativo. El sesgo y la curtosis indican que no se distribuye normalmente.

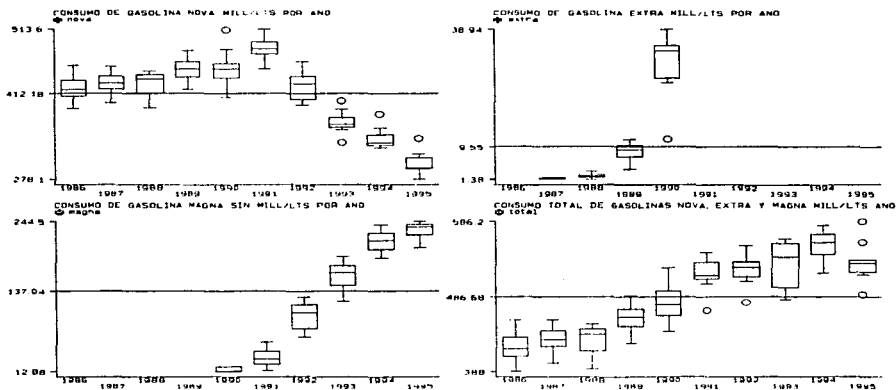
Diagrama de letras del total de consumo de gasolinas.

. lv total

#	116	total				
M	58.5	490.36	485.80	530.93	spread	pseudosigma
F	29.5	440.68	488.67	552.22	90.25	67.13
E	15	425.12	484.39	569.48	127.10	55.50
D	8	408.00	485.86	579.52	155.20	51.49
C	4.5	399.29	486.71	583.73	170.19	47.24
B	2.5	392.19	487.10	586.20	187.33	44.97
A	1.5	389.69			194.04	41.72
	1	388.00			198.20	39.42
inner fence		305.30		666.30	# below	# above
outer fence		169.93		801.68	0	0
					0	0

El lv nos reporta que hay 116 observaciones del consumo total de gasolinas. Con mediana de 490.36 millones de litros consumidos. La distribución no es perfectamente normal, ya que si lo fuera, los puntos medios de cada punto de corte serían iguales a la mediana. El pseudosigma, estimación de la desviación estándar, confirma esto al tener valores decrecientes. Se puede concluir que tiene menor dispersión que la distribución normal.

GRAFICAS DE CAJA DEL CONSUMO DE GASOLINAS NOVA, EXTRA, MAGNA Y EL TOTAL DE MILLONES DE LITROS CONSUMIDOS.



Las gráficas de caja permiten comparar de manera adecuada la variación en el consumo en millones de litros de los tres tipos de gasolinas que se venden en México. Se puede apreciar que el consumo de la gasolina Nova fue disminuyendo a partir de 1992; aunque todavía la media de consumo estaba por arriba de los 412.18 millones de litros y que la gasolina Magna para éste año tuvo un importante aumento en millones de litros consumidos. A partir de 1993 desciende significativamente el consumo de gasolina Nova, mientras que el consumo de gasolina Magna asciende.

La gasolina Extra salió a la venta en Enero de 1987 con el 0.38% del total de millones de litros consumidos, mientras que la gasolina Nova tenía el 99.62% restante. No fué sino hasta febrero de 1990 que se aumentó el consumo de la gasolina Extra con el 5.48% y la gasolina Nova quedó con el 94.52% restante. Es ésta la razón por la que algunas observaciones se sitúan por arriba de los 9.55 millones de litros, cambio que la gráfica de caja muestra en la marcada diferencia del consumo de gasolina Extra entre los años 1989 y 1990.

La gráfica de caja del consumo total de las tres gasolinas que se venden en México muestra un incremento en el consumo total de gasolinas en millones de litros. La razón quizá se deba al aumento de vehículos automotores que se dió en el sexenio pasado, producto del "hoy no circula".

En 1986 el consumo total de gasolinas era de 420.18 millones de litros en promedio y para el mes de septiembre de 1995 aumentó a 532.84 mill/lts.

Análisis exploratorio de las Toneladas de Plomo emitidas a la atmósfera de febrero de 1986 a Enero de 1995:

. summ ton_pb, detail

		ton_pb			
Percentiles		Smallest			
1-	18.99	18.95			
5-	19.19	18.99			
10-	19.96	19.12		Obs	108
25-	41.59	19.14		Sum of Wgt.	108
50%	116.90			Mean	97.65
		Largest		Std. Dev.	62.74
75%	125.15	292.43			
90%	134.71	300.83		Variance	3936.53
95%	143.81	302.93		Skewness	1.11
99%	302.93	319.02		Kurtosis	5.68

Hay 108 observaciones de las toneladas de plomo emitidas a la atmósfera, con una media de 97.65 y desviación estándar de 62.74; la mediana es de 116.90, por lo que hay un sesgo positivo. El sesgo y la curtosis indican las toneladas de plomo emitidas a la atmósfera no tienen una distribución normal.

. lv ton_pb

#	108	ton_pb					

M	54.5		116.90			spread	pseudosigma
F	27.5	41.59	83.37	125.15	83.55		62.16
E	14	20.46	75.71	130.96	110.50		48.26
D	7.5	19.34	80.55	141.76	122.42		40.67
C	4	19.14	155.78	292.43	273.29		74.79
B	2.5	19.06	160.47	301.88	282.82		68.86
A	1.5	18.97	164.97	310.97	292.00		63.52
	1	18.95	168.98	319.02	300.07		60.29
						# below	# above
inner fence		-83.73		250.47		0	5
outer fence		-209.06		375.80		0	0

El lv nos reporta que hay 108 observaciones de las Toneladas de plomo emitidas a la atmósfera. Con mediana de 116.90. La distribución no es normal. Si lo fuera, los puntos medios de cada punto de corte serían iguales a la mediana. El limite interior reporta que hay 5 observaciones por arriba de 250.47 toneladas de plomo emitidas a la atmósfera y son:

. list mm aa nova extra magna ton_pb if ton_pb>250.47

	mm	aa	nova	extra	magna	ton_pb
1.	2	1986	430.0	.	.	300.83
2.	3	1986	393.0	.	.	274.94
3.	4	1986	418.0	.	.	292.43
4.	5	1986	433.0	.	.	302.93
5.	6	1986	456.0	.	.	319.02

Estas 5 mediciones discrepantes tienen **plausibilidad biológica**, pues son de Febrero, Abril, Mayo y Junio de 1986; cuando la única gasolina que existía era la NOVA.

Análisis exploratorio del promedio de Plomo emitido a la atmósfera de Marzo de 1988 a Noviembre de 1994.

Estadísticas de resumen:

. summ pb_atm,detail

		pb_atm			

	Percentiles	Smallest			
1%	0.22	0.22			
5%	0.25	0.23			
10%	0.27	0.23	Obs		81
25%	0.37	0.24	Sum of Wgt.		81
50%	0.84		Mean		0.91
		Largest	Std. Dev.		0.61
75%	1.20	1.75			
90%	1.48	2.23	Variance		0.37
95%	1.70	2.98	Skewness		1.58
99%	3.62	3.62	Kurtosis		7.50

Hay 81 observaciones del promedio de plomo emitido a la atmósfera, con una media de $0.91\mu\text{g}/\text{m}^3$ y desviación estándar de $0.61\mu\text{g}/\text{m}^3$. La mediana es de 0.84 por lo que hay un sesgo positivo. El sesgo y la curtosis indican que ésta variable no se distribuye normalmente.

Diagrama de letras para el Plomo emitido a la atmósfera en unidades de $\mu\text{g}/\text{m}^3$:

. lv pb_atm

#	81	pb_atm			spread	pseudosigma
M	41		0.84			
F	21	0.37	0.79	1.20	0.83	0.63
E	11	0.28	0.84	1.40	1.12	0.50
D	6	0.26	0.98	1.69	1.43	0.48
C	3.5	0.23	1.11	1.99	1.76	0.50
B	2	0.23	1.61	2.98	2.75	0.67
A	1.5	0.23	1.76	3.30	3.07	0.70
	1	0.22	1.92	3.62	3.40	0.71
inner fence		-0.88		2.45	# below	# above
outer fence		-2.12		3.69	0	2
					0	0

El lv nos reporta que hay 81 observaciones de plomo emitidas a la atmósfera. Con mediana de 0.84 y por debajo de la mediana hay 41 observaciones. La distribución no es normal. Si lo fuera, los puntos medios de cada punto de corte serían iguales a la mediana.

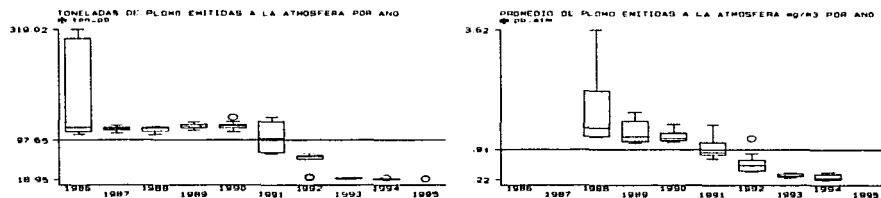
El límite interior reporta que hay 2 observaciones por arriba de $2.4\mu\text{g}/\text{m}^3$ del promedio de plomo emitido a la atmósfera y son:

. list mm aa nova extra magna pb_atm if pb_atm>2.4

	mm	aa	nova	extra	magna	pb_atm
1.	3	1988	390.0	1.38	.	3.624
2.	12	1988	447.4	3.34	.	2.977

Estos datos son plausibles ya que ocurrieron en los meses de Marzo y Diciembre de 1988 que coinciden con la temporada de secas y pudiera probarse en otra investigación que es en ésta época del año que los niveles de plomo en la atmósfera se incrementan.

GRAFICAS DE CAJA DE TONELADAS DE PLOMO EMITIDAS A LA ATMOSFERA Y EL PROMEDIO DE PLOMO EMITIDO ($\mu\text{g}/\text{m}^3$)







En el primer semestre de 1986, el promedio era de 298.03 toneladas de plomo emitidas a la atmósfera, del segundo semestre de 1986 hasta el primer semestre de 1991 se mantiene "constante" la emisión de éste contaminante con una emisión media de 123.63 toneladas de plomo durante éste período. Una razón de ello se puede deber al gradual incremento en el consumo de la gasolina Magna a partir del segundo semestre de 1991 con un promedio de 6.45% de consumo.

Para el segundo semestre de 1992 en adelante, la gasolina Magna se vende mensualmente por arriba de los 100 millones de litros, lo que puede explicar la importante reducción en las toneladas de plomo emitidas a la atmósfera.

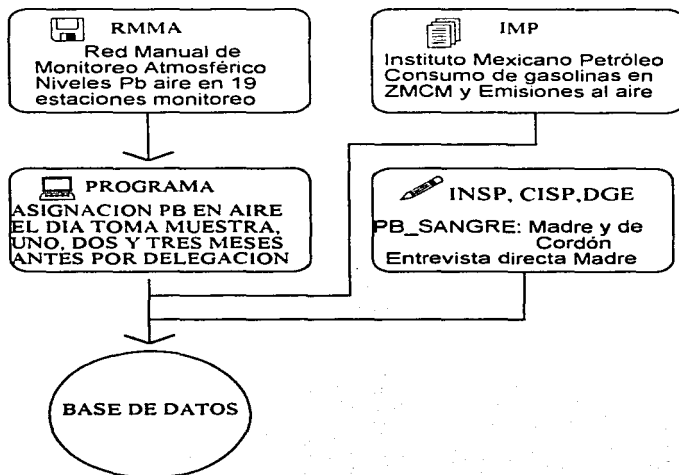
La gráfica de caja refleja un cambio a la baja en el promedio de plomo emitido a la atmósfera a través del tiempo.


C. Análisis exploratorio de la información proporcionada por el INSP-CISP.

El Instituto Nacional de Salud Pública (INSP) en la división del Centro de Investigación de Salud Poblacional (CISP) proporcionó información de mujeres embarazadas de la Ciudad de México como son: el hospital donde nació el hijo, la delegación donde vive la madre, su grado de escolaridad, hábitos de la madre de guardar y/o preparar alimentos en recipientes de barro vidriado. También proporcionó el nivel de concentración de plomo en el torrente sanguíneo de la madre y del cordón umbilical. muestras que se usaron como biomarcadores (*) en el estudio. (Ver Capítulo V, 2. Determinación de plomo en sangre).

El siguiente esquema muestra con  los datos antes mencionados. Con el símbolo  la información que proporcionó la RMMA, con el símbolo  los datos que fueron calculados por el programa de asignación de pb en aire por Delegación donde vive la madre y por el hospital donde nació el hijo y la fecha de nacimiento. Con  la información que proporcionó el Instituto Mexicano del Petróleo (IMP).

ESQUEMA DE LAS FUENTES DE INFORMACION



Las variables a las que se les hará un análisis exploratorio serán las que están en negritas y las que resultaron del programa simbolo . Las restantes se podrán apreciar en los resultados de ésta investigación.

Variable	Obs	Mean	Std. Dev.	Min	Max
folio	3833	43490	10557	10003	52438
procede	0				
fecha_na	0				
mescont1	3835	71	13	53	95
pbaire_d	445	0.314575	0.185016	0.042300	0.969000
pbdoas_d1	452	0.332155	0.167728	0.043000	0.969000
asigna_d	2118	0.293633	0.202707	0.016700	1.059500
asidos_d	1865	0.337527	0.188767	0.043000	1.059500
estima_d	2563	0.297265	0.199870	0.016700	1.059500
estdos_d	2317	0.336473	0.184928	0.043000	1.059500
pbiant_d	3346	0.312072	0.171612	0.043000	0.875200
pbldos_d	3393	0.328579	0.153282	0.043000	0.875200
pbzart_d	3599	0.315610	0.167381	0.063700	0.853000
pbzdos_d	3647	0.333357	0.147511	0.151232	0.853000
pbzart_d	3753	0.317110	0.164566	0.063700	0.892300
pbzdos_d	3801	0.336587	0.142402	0.154400	0.892300
hozp	3835	4	3	1	10
delegaci	0				
delega	3835	7	5	1	33
asign_es	3835	6	4	1	14
edad	3507	25	6	11	46
esco	3518	9	3	0	18
newesco	3518	3	1	0	5
loza	3838	2	0	1	2
mpb2	3663	10.22	6.32	0.50	66.20
cpb2	3450	9.35	6.22	0.40	50.60
date_na	3835	12253.25	399.5797	11705	12964
diad1	432	-1.36207	0.5762475	-3.162968	-2.2841327
diad1G	445	-1.327413	0.602046	-3.162968	-0.0314907
pbdiabos	452	-1.211528	0.461867	-3.146555	-0.0314907
estimado	2563	-1.415864	0.6353772	-4.082347	0.0577971
lante2d	3598	-1.30028	0.5555573	-2.753571	-1.1589857
lpb3an_d	3753	-1.279446	0.5655043	-2.753571	-1.1335229
lmpE2	3663	2.153922	0.9746	-0.6931472	4.19268
tesad	3507	2.0039114	0.021173	0.147442	3.015113
lcpb2	3450	2.036815	0.6450647	-0.9162907	3.923982
nova	3835	376.587	49.624	278.100	462.460
magna	3835	156.236	54.131	66.540	244.500
total	3835	532.703	24.507	478.960	586.200
ton_pb	3247	42.246	21.633	18.952	70.062
FB_atm	3246	0.411	0.123	0.222	1.149
anc	3835	1993.003	1.157404	1992	1995
mm	3835	7.061538	2.585294	1	12
dia	3835	15.76193	8.71013	1	31
medianam	3835	3.188344	2.571523	3.9	16
trimcont	3835	6.263963	4.370906	1	15
trimpb2	3835	9.186889	1.966017	5.8	13
ltrimpb2	3835	2.168575	1.072705	1.757858	2.564949
estacion	3835	2.646154	0.298324	1	4
mescont	3835	14.51056	6.189745	1	31
zona	3835	2.326299	1.240185	1	5
nivel	3835	2.19302	0.867436	1	3

De la variable **folio** se formó la variable **procede** que indica que Institución proporcionó la información; *Tenden* fué información proporcionada por la Dirección General de Epidemiología (DGE) de un proyecto de vigilancia epidemiológica - 2489 observaciones-, en conjunción con el INSP; e *Intervención* son datos del INSP con 1394 observaciones.

Con la variable **fecha_na**, fecha en que nació el hijo y se tomaron muestras de sangre de la madre y del cordón umbilical para determinar sus niveles de plomo en sangre, se obtuvieron las variables **date_na** - con formato de fecha - y por separado las variables **año**, **mm** y **dia**. Esto con la finalidad de poder construir la variable **mescont** que mide el tiempo mensual progresivo y la variable estación para saber a que estación del año corresponde a cada observación.

La información que se tiene fué recopilada de 1992 a 1995 y se distribuyó de la siguiente manera:

AÑO QUE NACIO EL HIJO	Freq.	Percent	Cum.
1992	1973	51.47	51.47
1993	466	12.16	63.63
1994	805	21.00	84.63
1995	589	15.37	100.00
Total	3833	100.00	

Las variables **hosp** y **delegaci** permitieron crear las variables **delega** y **asign_es** que proporcionan la Delegación donde vive la madre y la Delegación donde nació su hijo, respectivamente. Esta información resulta de importancia para ver si el plomo en sangre de la madre tiene que ver con el día que nació su bebé o está asociado a exposición a plomo meses antes, dados por el grado de contaminación ambiental de estas Delegaciones.

La distribución de las observaciones por la Delegación del hospital donde nació el hijo:

. tab hosp

No. HOSPITAL DONDE NACIO HIJO	Freq.	Percent	Cum.
1	236	6.16	6.16
2	449	11.71	17.87
3	1687	44.01	61.88
4	470	12.26	74.15
5	12	0.31	74.46
6	309	8.06	82.52
7	121	3.16	85.68
8	97	2.53	88.21
9	71	1.85	90.06
10	381	9.94	100.00
Total	3833	100.00	

Descripción del número del hospital donde nació el hijo:

No. asignado	HOSPITAL	DELEGACION
1	GINECO OBSTETRICIA 3A	GUSTAVO A. MADERO
2	GINECO OBSTETRICIA 27	CUAUHTEMOC
3	GINECO OBSTETRICIA 4	ALVARO OBREGON
4	GINECO OBSTETRICIA 2A	IZTACALCO
5	PERINATOLOGIA	MIGUEL HIDALGO
6	GEA GONZALEZ	TLALPAN
7	GRAL. DE MEXICO	CUAUHTEMOC
8	DE LA MUJER	MIGUEL HIDALGO
9	REGIONAL 20 DE NOVIEMBRE	BENITO JUAREZ
10	IGNACIO ZARAGOZA	IZTAPALAPA

La distribución de las Delegaciones donde vivían las madres:

No. asignado delegacion MAMA	Freq.	Percent	Cum.	DELEGACION
1	902	23.53	23.53	ALVARO OBREGON
2	3	0.08	23.61	AZCAPOTZALCO
3	156	4.07	27.68	BENITO JUAREZ
4	212	5.53	33.21	COYOACAN
5	66	1.72	34.93	CUAJIMALPA
6	479	12.50	47.43	CUAUHTEMOC
7	239	6.24	53.67	GUSTAVO A. MADERO
8	478	12.47	66.14	IZTACALCO
9	522	13.62	79.75	IZTAPALAPA
10	146	3.81	83.56	MAGDALENA CONTRERAS
11	45	1.17	84.74	MIGUEL HIDALGO
12	2	0.05	84.79	MILPA ALTA
13	6	0.16	84.95	TLAHUAC
14	506	13.20	98.15	TLALPAN
15	8	0.21	98.36	VENUSTIANO CARRANZA
16	40	1.04	99.40	XOCHIMILCO
25	1	0.03	99.43	CHIMALHUACAN
31	8	0.21	99.63	HUIXQUILUCAN
32	4	0.10	99.74	NAUCALPAN
33	10	0.26	100.00	NEZAHUALCOYOTL
Total	3833	100.00		

La variable esco se refiere al grado de escolaridad de la madre.

. summ esco, detail

ESCOLARIDAD DE LA MADRE					
Percentiles		Smallest			
1%	1	0			
5%	4	0			
10%	6	0	Obs		3516
25%	6	0	Sum of Wgt.		3516
50%	9		Mean		9
		Largest	Std. Dev.		3
75%	11	18			
90%	12	18	Variance		10
95%	15	18	Skewness		0
99%	17	18	Kurtosis		4

La mayoría de madres tienen una formación secundaria con el 36.35%; con preparación primaria el 18.80% y con preparación preparatoria el 13.71%.

Distribución de el grado de escolaridad de las madres de éste estudio:

ESCOLARIDAD	Freq.	Percent	Cum.
DE LA MADRE			
0	32	0.91	0.91
1	18	0.51	1.42
2	36	1.02	2.45
3	69	1.96	4.41
4	45	1.28	5.69
5	65	1.85	7.54
6	661	18.80	26.34
7	101	2.87	29.21
8	181	5.15	34.36
9	1278	36.35	70.71
10	133	3.78	74.49
11	158	4.49	78.98
12	482	13.71	92.69
13	35	1.00	93.69
14	22	0.63	94.31
15	51	1.45	95.76
16	78	2.22	97.98
17	40	1.14	99.12
18	31	0.88	100.00
Total	3516	100.00	

La variable **esco** se categorizó en la variable *newesco* de la siguiente manera:

- 0= 0 años (analfabeta)
- 1= 1 - 6 años (primaria)
- 2= 7 - 8 años (secundaria incompleta)
- 3= 9 años (secundaria)
- 4= 10 - 12 años (preparatoria)
- 5= 13 y + años (preparación universitaria)

. summ newesco

Variable	Obs	Mean	Std. Dev.	Min	Max
newesco	3516	3	1	0	5

En este estudio las mujeres tienen en su mayoría el grado escolar de secundaria completa.

Distribución de la escolaridad categorizada:

. tab newesco

CATEG. ESCOLARIDAD DE LA MADRE	Freq.	Percent	Cum.
0	32	0.91	0.91
1	894	25.43	26.34
2	282	8.02	34.36
3	1278	36.35	70.71
4	773	21.99	92.69
5	257	7.31	100.00
Total	3516	100.00	

La edad de la madre en años cumplidos tiene una mediana de 25 años. Hay dos mujeres con valores muy extremos: 1 mujer de 11 años de edad y otra de 46 años. Esto se cotejó con las fuentes primarias y son datos correctos por lo que se dejaron en el análisis.

. suma edad,detail

		EDAD DE LA MADRE			

	Percentiles	Smallest			
1%	15	11			
5%	17	14			
10%	18	14		Obs	3507
25%	21	14		Sum of Wgt.	3507
50%	24			Mean	25
75%	28	Largest		Std. Dev.	6
90%	33	43		Variance	31
95%	35	44		Skewness	1
99%	39	46		Kurtosis	3

Diagrama de letras de la edad de las madres:

. lv edad		EDAD DE LA MADRE			
#	3507	-----			
M	1754	21	24	spread	pseudosigma
F	877.5	21	24	7	5
E	439	19	26	13	6
D	220	18	26	17	6
C	110.5	17	27	20	5
B	55.5	16	28	23	5
A	28	15	28	25	5
Z	14.5	15	28	26	5
Y	7.5	14	28	28	5
X	4	14	28	29	5
	1	11	28	35	5
inner fence		10	38	# below	# above
outer fence		0	49	0	57
				0	0

En el límite interno superior (inner fence) aparecen 57 mujeres mayores a los 38 años de edad, que son las que hacen que la distribución tenga un sesgo a la derecha.

Distribución de la edad en las madres participantes del estudio:

```

.tab edad
  EDAD DE LA MADRE |
-----+-----
    11 |          1      0.03      0.03
    14 |          9      0.26      0.29
    15 |         26      0.74      1.03
    16 |         55      1.57      2.59
    17 |        104      2.97      5.56
    18 |        157      4.48     10.04
    19 |        248      7.07     17.11
    20 |        242      6.90     24.01
    21 |        249      7.10     31.11
    22 |        261      7.44     38.55
    23 |        237      6.76     45.31
    24 |        241      6.87     52.18
    25 |        222      6.33     58.51
    26 |        210      5.99     64.50
    27 |        195      5.56     70.06
    28 |        187      5.33     75.39
    29 |        155      4.42     79.81
    30 |        138      3.93     83.75
    31 |         99      2.82     86.57
    32 |        112      3.19     89.76
    33 |         79      2.25     92.02
    34 |         56      1.60     93.61
    35 |         61      1.74     95.35
    36 |         47      1.34     96.69
    37 |         30      0.86     97.55
    38 |         29      0.83     98.37
    39 |         22      0.63     99.00
    40 |         12      0.34     99.34
    41 |         12      0.34     99.69
    42 |          4      0.11     99.80
    43 |          4      0.11     99.91
    44 |          2      0.06     99.97
    46 |          1      0.03    100.00
-----+-----
  Total |       3507    100.00
  
```

Mediante estadísticas de normalidad como sktest y ladder se determinó que la mejor transformación para lograr una distribución normal en ésta variable, es el valor recíproco de la raíz cuadrada de la variable edad:

```

.sktest edad

Skewness/Kurtosis tests for Normality
-----+-----
Variable | Pr(Skewness)  Pr(Kurtosis)  adj chi-sq(2)  joint Pr(chi-sq)
-----+-----
edad |          0.000          0.673          0.0000
  
```

. ladder edad

Transformation	formula	Chi-sq(2)	P(Chi-sq)
cube	edad^3	.	0.000
square	edad^2	.	0.000
raw	edad	.	0.000
square-root	sqrt(edad)	.	0.000
log	log(edad)	62.34	0.000
reciprocal root	1/sqrt(edad)	48.35	0.000
reciprocal	1/edad	71.28	0.000
reciprocal square	1/(edad^2)	.	0.000
reciprocal cube	1/(edad^3)	.	0.000

. ladder edad, g(tedad)

(tedad = 1/sqrt(edad) generated)

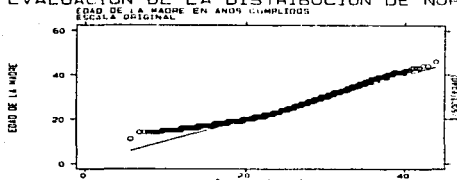
Esta transformación la selecciona STATA por ser la de menor valor en su ji-cuadrada. La transformación ideal sería aquella en que la probabilidad del estadístico ji-cuadrada es no significativa.

. summ tedad,detail

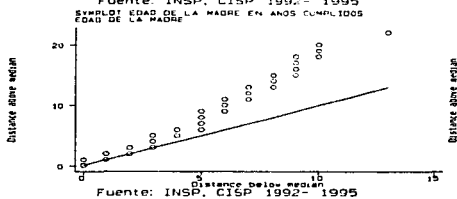
1/sqrt(edad)

Percentiles		Smallest		
1%	.1601281	.147442		
5%	.1690308	.1507557		
10%	.1740777	.1507557	Obs	3507
25%	.1889822	.1524986	Sum of Wgt.	3507
50%	.2041242		Mean	.2039114
		Largest	Std. Dev.	.0221173
75%	.2182179	.2672612		
90%	.2357023	.2672612	Variance	.0004892
95%	.2425356	.2672612	Skewness	.1309829
99%	.2581989	.3015113	Kurtosis	2.563293

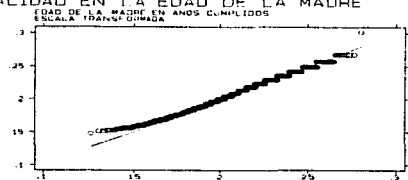
EVALUACION DE LA DISTRIBUCION DE NORMALIDAD EN LA EDAD DE LA MADRE



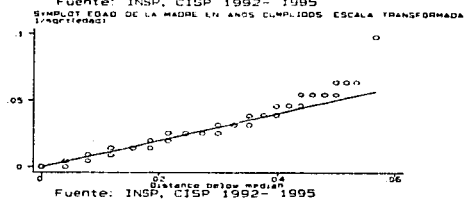
Fuente: INSP, CISP 1992-1995



Fuente: INSP, CISP 1992-1995



Fuente: INSP, CISP 1992-1995



Fuente: INSP, CISP 1992-1995

Diagrama de letras para la transformación de la variable edad:

#	lv	tedad	1/sqrt(edad)			spread	pseudosigma
M	1754		.2041242				
F	877.5	.1889822	.2036001	.2182179	.0292357	.021682	
E	439	.1767767	.2030962	.2294157	.052639	.0228865	
D	220	.1690308	.2023666	.2357023	.0666714	.0217446	
C	110.5	.164399	.2034673	.2425356	.0781366	.0209993	
B	55.5	.1601281	.2050641	.25	.0898719	.0208885	
A	28	.1581139	.2081564	.2581989	.100085	.0207297	
Z	14.5	.1561738	.2071863	.2581989	.1020251	.0192589	
Y	7.5	.153401	.2103311	.2672612	.1138603	.0198267	
X	4	.1524986	.2098799	.2672612	.1147627	.0186484	
	1	.147442	.2244767	.3015113	.1540694	.0217457	
inner fence		.1451287		.2620714	# below	# above	
outer fence		.1012753		.3059249	0	10	
					0	0	

Tanto las gráficas como el diagrama de letras muestran que con la transformación del recíproco de la raíz cuadrada de la variable edad, la distribución se asemeja a la distribución normal.

La variable **loza** reporta los hábitos de cada madre en guardar y/o preparar los alimentos en recipientes de barro vidriado.

. tab loza

USO DE BARRO VIDRIADO	Freq.	Percent	Cum.
1	1363	38.52	38.52
2	2175	61.48	100.00
Total	3538	100.00	

Las mujeres que sí utilizan recipientes de barro vidriado para guardar y/o preparar alimentos, son el 38.52% y el 61.48% no los utilizan.

Las variables **mpb2** y **cpb2** miden el nivel de plomo en la sangre de la madre y en el cordón umbilical respectivamente.

Nivel de plomo en el torrente sanguíneo de la madre (mpb2):

. summ mpb2,detail

NIVEL PLOMO SANGRE DE LA MADRE				
	Percentiles	Smallest		
1%	1.70	0.50		
5%	3.20	0.50		
10%	4.20	1.00	Obs	3663
25%	6.00	1.00	Sum of Wgt.	3663
50%	8.70		Mean	10.22
		Largest	Std. Dev.	6.32
75%	12.60	50.00		
90%	18.00	54.40	Variance	39.98
95%	22.20	58.30	Skewness	1.95
99%	33.60	66.20	Kurtosis	9.62

En la evaluación estadística se puede apreciar que la concentración de plomo en el torrente sanguíneo de la madre no se distribuye normalmente y existe un sesgo positivo. La media en el nivel de plomo en sangre es de 10.22 μ g/dl y la mediana es de 8.7 μ g/dl, valores que son muy altos considerando que la norma actual es de 10 μ g/dl.

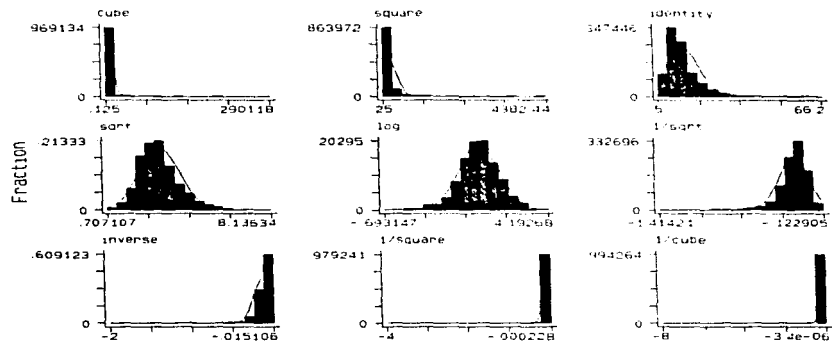
Diagrama de letras del nivel de plomo en sangre de la madre

. lv mpb2		NIVEL PLOMO SANGRE DE LA MADRE					
#	3663						
M	1832		8.70			spread	pseudosigma
F	916.5	6.00	9.30	12.60		6.60	4.89
E	458.5	4.60	10.80	17.00		12.40	5.39
D	229.5	3.50	12.22	20.95		17.45	5.69
C	115	2.70	14.15	25.60		22.90	6.15
B	58	2.00	16.25	30.50		28.50	6.63
A	29.5	1.50	18.00	34.50		33.00	6.84
Z	15	1.10	19.05	37.00		35.90	6.77
Y	8	1.00	20.75	40.50		39.50	6.90
X	4.5	1.00	24.65	48.30		47.30	7.75
	1	0.50	33.35	66.20		65.70	9.24
inner fence		-3.90		22.50		# below	# above
outer fence		-13.80		32.40		0	178
						0	44

La mediana es de 8.7 $\mu\text{g/dl}$. Los puntos medios para cada punto de corte son mayores a la mediana y se incrementan conforme la amplitud para cada punto de corte aumenta, lo que indica una tendencia hacia los niveles de plomo en sangre elevados. El límite interno superior marca 178 mujeres con elevados niveles de plomo en sangre mayores a los 22.5 $\mu\text{g/dl}$ y de éstas hay 44 mujeres (límite externo superior) que tienen niveles mayores a los 32.4 $\mu\text{g/dl}$ de plomo en sangre.

El comportamiento descrito anteriormente se puede apreciar en el gladder de plomo en sangre de las madres en la gráfica sin transformar la variable mpb2 (indentify) con un marcado sesgo hacia la derecha.

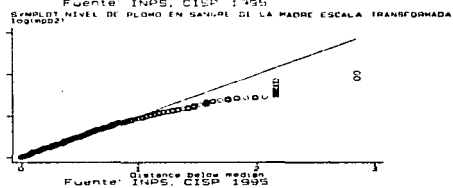
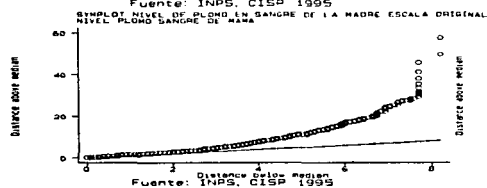
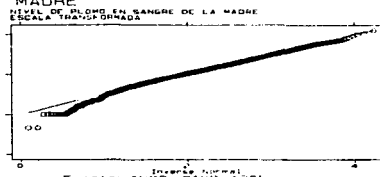
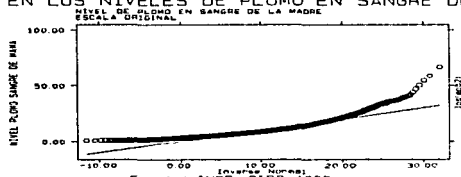
Evaluación visual para elegir la mejor transformación
 . gladder mpb2



NIVEL PLOMO SANGRE DE MAMA
 HISTOGRAMA POR TRANSFORMACION

La transformación logarítmica de la variable **mpb2**, es la que más la normaliza; así que se hizo la transformación pues ésta variables debe cumplir con los supuestos de los análisis estadísticos que se emplearán en capítulos posteriores. (Ver Capítulo IV 6, Transformaciones)

EVALUACION DE LA DISTRIBUCION DE NORMALIDAD
EN LOS NIVELES DE PLOMO EN SANGRE DE LA MADRE



. summ lmpb2, detail

lmpb2

Percentiles		Smallest		
1%	.5306283	-.6931472		
5%	1.163151	-.6931472		
10%	1.435084	0	Obs	3663
25%	1.791759	0	Sum of Wgt.	3663
50%	2.163323		Mean	2.153922
		Largest	Std. Dev.	.59746
75%	2.533697	3.912023		
90%	2.890372	3.996364	Variance	.3569585
95%	3.100092	4.065602	Skewness	-.3438459
99%	3.514526	4.19268	Kurtosis	3.892053

Como lo demuestran las gráficas y la estadística anterior, la transformación logarítmica logra casi una perfecta distribución normal.

Análisis de la distribución de los niveles de plomo en sangre del cordón umbilical:

. sum cpb2,detail

NIVEL PLOMO SANGRE DE CORDON					

	Percentiles	Smallest			
1%	1.30	0.40			
5%	2.60	1.00			
10%	3.40	1.00	Obs		3450
25%	5.00	1.00	Sum of Wgt.		3450
50%	7.90		Mean		9.35
		Largest	Std. Dev.		6.22
75%	12.00	41.00			
90%	17.00	43.60	Variance		38.71
95%	21.00	49.00	Skewness		1.82
99%	33.00	50.60	Kurtosis		7.83

La distribución de los niveles de plomo en el cordón umbilical indican que hay un sesgo a la derecha con una media de 9.35µg/dl, desviación estándar de 6.22µg/dl. No se distribuye normalmente.

. lv cpb2

NIVEL PLOMO SANGRE DE CORDON						

#	3450					
M	1725.5		7.90		spread	pseudosigma
F	863	5.00	8.50	12.00	7.00	5.19
E	432	3.70	9.60	15.50	11.80	5.13
D	216.5	2.90	11.45	20.00	17.10	5.58
C	108.5	2.10	13.35	24.60	22.50	6.04
B	54.5	1.65	15.83	30.00	28.35	6.59
A	27.5	1.20	18.00	34.80	33.60	6.96
Z	14	1.00	19.40	37.80	36.80	6.93
Y	7.5	1.00	20.75	40.50	39.50	6.89
X	4	1.00	21.00	41.00	40.00	6.51
	1	0.40	25.50	50.60	50.20	7.09
					# below	# above
inner fence		-5.50		22.50	0	148
outer fence		-16.00		33.00	0	36

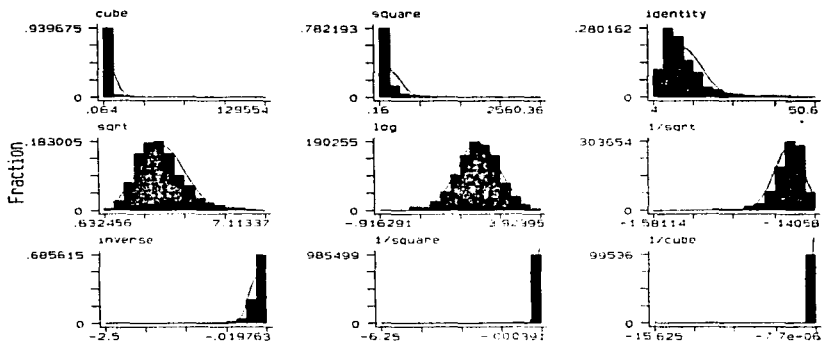
En el diagrama de letras se observa que la distribución no es normal, pues la dispersión en cada punto de corte es creciente, lo que indica que tiene mayor dispersión que la distribución normal. Además el límite interno indica que hay 148 observaciones por arriba de los 22.50µg/dl, pero debido a que fueron cotejados éstos resultados de laboratorio de éstas observaciones con sus respectivos resultados de niveles de plomo en sangre de la madre y dado que no corresponden a errores o valores aberrantes se incluyeron en el análisis.

Posibles transformaciones para normalizar la distribución en el nivel de plomo en sangre del cordón umbilical:

. ladder cpb2

Transformation	formula	Chi-sq(2)	P(Chi-sq)
cube	cpb2^3	.	.
square	cpb2^2	.	.
raw	cpb2	.	0.000
square-root	sqrt(cpb2)	.	0.000
log	log(cpb2)	39.64	0.000
reciprocal root	1/sqrt(cpb2)	.	0.000
reciprocal	1/cpb2	.	.
reciprocal square	1/(cpb2^2)	.	.
reciprocal cube	1/(cpb2^3)	.	.

. gladder cpb2



NIVEL PLOMO SANGRE DE CORDON
HISTOGRAMAS POR TRANSFORMACION

Tanto el ladder como el gladder muestran que la transformación logarítmica de la variable **cpb2** es la que puede hacer que se ajuste más a la distribución normal.

. ladder cpb2.g(1cpb2)
(1cpb2 = log(cpb2) generated)

La transformación logarítmica de ésta variable se llama **1cpb2 = log(cpb2)**

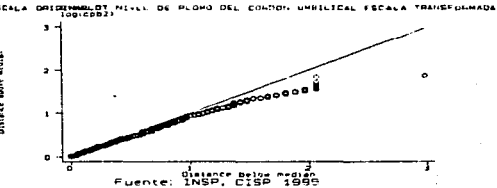
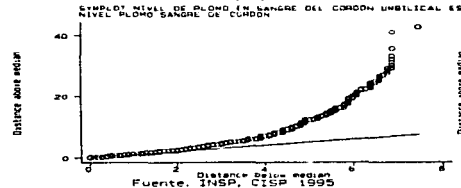
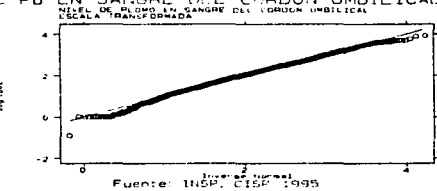
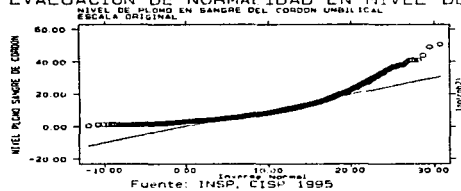
Análisis exploratorio para la transformación de plomo en cordón umbilical:

. summa lcpb2,detail

log (cpb2)

Percentiles		Smallest		
1%	.2623642	-.9162907	Obs	3450
5%	.9555114	0	Sum of Wgt.	3450
10%	1.223776	0		
25%	1.609438	0		
50%	2.066863		Mean	2.036815
		Largest	Std. Dev.	.6450647
75%	2.484907	3.713572		
90%	2.833213	3.775057	Variance	.4161085
95%	3.044523	3.89182	Skewness	-.2509805
99%	3.496508	3.923952	Kurtosis	3.295431

EVALUACION DE NORMALIDAD EN NIVEL DE Pb EN SANGRE DEL CORDON UMBILICAL



Análisis exploratorio de las variables generadas por el programa que calcula el nivel de plomo en aire por Delegación donde vive la madre y por la Delegación del hospital donde nació su hijo (Ver Capítulo V, 3. Criterios de asignación...).

NIVEL PB AIRE	DELEGACION VIVIA MAMA		DELEGACION HOSPITAL	
	DELEGA		ASIGN_ES	
Día de fecha_na	pbaire_d	$\mu=0.314$ s.d.=0.185	pbdos_di	$\mu=0.332$ s.d.=0.167
Asignado por las fechas mas cercanas a la fecha_na	asigna_d	$\mu=0.293$ s.d.=0.202	asidos_d	$\mu=0.337$ s.d.=0.188
Conjunción día de fecha_na U valor asignado cuando no hubo medición.	estima_d	$\mu=0.297$ s.d.=0.199	estdos_d	$\mu=0.336$ s.d.=0.185
1 mes antes de la fecha_na	pb1ant_d	$\mu=0.312$ s.d.=0.172	pb1dos_d	$\mu=0.328$ s.d.=0.154
2 meses antes de la fecha_na	pb2ant_d	$\mu=0.316$ s.d.=0.167	pb2dos_d	$\mu=0.333$ s.d.=0.147
3 meses antes de la fecha_na	pb3ant_d	$\mu=0.317$ s.d.=0.164 M=0.297 skewness=0.54* kurtosis=2.38*	pb3dos_d	$\mu=0.337$ s.d.=0.142 M=0.291 skewness=0.905 kurtosis=2.864

* Conforme se incrementa el lag de tiempo en meses antes de la toma de muestra se hace más fuerte la correlación y la variable adquiere por su propia naturaleza una distribución muy simétrica, casi normal.

Después de hacer el análisis exploratorio de se encontró que las variables más representativas son las correspondientes a la Delegación de la madre y no a la Delegación del hospital donde nació su hijo. Por esta razón fueron las únicas variables que se transformaron, aunque no es tan necesaria la transformación pues serán variables independientes en el modelo de regresión, en el caso que se decidiera buscar una relación del consumo de gasolinas y toneladas emitidas de Pb a la atmósfera versus las variables calculadas por el programa.

CAPITULO III

CORRELACION.

El **coeficiente de correlación** se define como la cuantificación del grado de asociación en el que dos variables continuas se relacionan, simbolizadas por **X** y **Y**, suponiendo que la relación es lineal.

A. Características del coeficiente de correlación

Las características del coeficiente de correlación nos ayudan a interpretar su valor numérico y a describir una relación. Son las siguientes:

- * El coeficiente de correlación oscila desde -1 a $+1$, en el que -1 describe la relación lineal negativa perfecta y $+1$ describe la relación lineal positiva perfecta.
- * Una relación de 0 significa que no hay relación lineal entre las dos variables.
- * Cuando la correlación se aproxima a 0 , la forma del patrón de observaciones es más o menos circular (fig.3-1A). Conforme el valor de las observaciones se aproxima a $+1$ o -1 , la forma se vuelve más elíptica u oval, pero con una pequeña correlación (fig. 3-1B, 3-1C). Cuando el valor de la correlación se aproxima a $+1$ o -1 se tiene una forma estrecha y alargada (fig. 3-1D), hasta que en $+1$ o -1 , las observaciones quedan directamente sobre la línea recta.(fig. 3-1E, 3-1F).

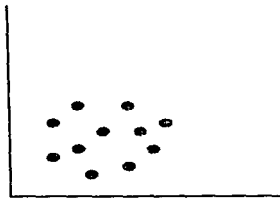


figura 3-1A $r \rightarrow 0$

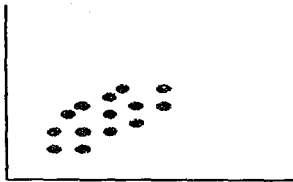


figura 3-1B $r=0.7$



figura 3-1C $r=-0.4$

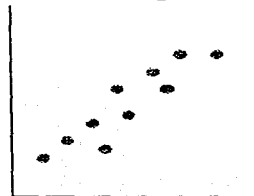


fig. 3-1D $r=+0.9$

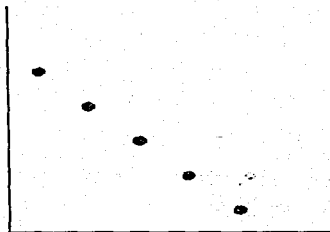


fig. 3-1E $r=-1$
Es una correlación lineal negativa perfecta

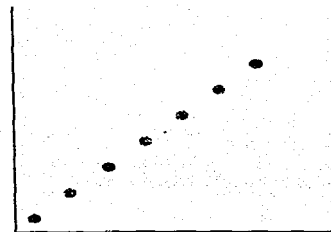


fig. 3-1 F $r=+1$
Es una correlación lineal positiva perfecta

La razón para graficar la dispersión de los datos, es para identificar *relaciones no lineales* cuando éstas existen. Si se calcula el coeficiente de correlación sin analizar los datos, es posible ignorar una *relación importante, pero no lineal*.

* La correlación se eleva al cuadrado (r^2) para formar un valor estadístico importante, llamado **coeficiente de determinación**. En el contexto de regresión lineal el coeficiente de determinación nos indica que se puede explicar o pronosticar (r^2) % de la variabilidad de la variable dependiente o respuesta (Y) por cada unidad de cambio en la variable independiente o explicatoria (X). (1)

* El valor del coeficiente de correlación es **independiente** de cualquier unidad usada para medir las variables, porque las unidades de X y Y aparecen ambas en el numerador y denominador y así se cancelan aritméticamente. Por ello se usa la correlación para ver si hay asociación entre el nivel de Pb en aire, con unidades de medición de $\mu\text{g}/\text{m}^3$, y el nivel de Pb en sangre con unidades de $\mu\text{g}/\text{dl}$. Esto se debe a que el denominador en la fórmula para el coeficiente de correlación se ajusta para la escala de unidades.

* El valor del coeficiente de correlación se altera de forma importante por la presencia de un valor alejado o distante, *ya que si se elimina una observación con estas características, puede hacer que se reduzca el coeficiente de correlación*. Por ello la correlación no proporciona una descripción adecuada de la relación entre dos variables cuando la distribución de una u otra variable incluye valores distantes o está sesgada. Cuando esto ocurra se debe hacer una transformación de los datos. Por ejemplo, *logaritmos o categórica, para cambiar la escala de medición y moderar el efecto de los puntos distantes o alejados antes de calcular la correlación*.

* El coeficiente de correlación tiende a ser menos confiable cuando se estima para muestras pequeñas. Es importante que el tamaño de la muestra sea lo suficientemente grande para que las correlaciones sean confiables.

Según Comrey: " un tamaño de muestra de 50 a 100 observaciones *es pequeño*; de 200 observaciones *es regular*; 300 *es buena*; de 500 *es muy buena* y de 1000 observaciones *es excelente*".

Otros sugieren que un tamaño de muestra de 100 a 200 es suficiente. (20)

B. Cálculo del coeficiente de correlación de PEARSON

En la población bajo estudio de la cual tenemos una muestra seleccionada de puntos (x, y) , la correlación entre X y Y se denota por ρ (rho) que es el símbolo del coeficiente de correlación en la población. La correlación cuantifica la fuerza de la relación lineal entre los resultados x y y . (19) Esto puede ser pensado como el promedio del producto de las desviaciones estándar de X y Y :

$$\rho = \text{promedio} \left[\frac{(X - \mu_x)}{\sigma_x} \cdot \frac{(Y - \mu_y)}{\sigma_y} \right]$$

El estimador de la correlación poblacional es conocido como el coeficiente de correlación de Pearson, denotado por r ; se calcula usando la fórmula:

$$\begin{aligned} r &= \frac{1}{(n-1)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) \\ &= \frac{1}{(n-1)} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \end{aligned}$$

Recordando que s_x y s_y son las desviaciones estándar de la muestra de los valores de x y y respectivamente.

$$s_x s_y = \left(\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \right) \left(\sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \right)$$

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

Sustituyendo en la fórmula anterior se obtiene:

$$= \frac{1}{n-1} \left[\frac{\sum_{i=1}^n (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}} \right]$$

La fórmula del coeficiente de correlación, simbolizado por r

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

C. Interpretación del coeficiente de correlación.

Colton da una regla para interpretar el tamaño de las correlaciones:

" correlaciones de 0 a 0.25 (o -0.25) indican correlación escasa o sin correlación; de 0.25 a 0.5 (-0.25 a -0.5), cierto grado de correlación; de 0.5 a 0.75 (-0.5 a -0.75) la relación es moderada a buena y mayores a 0.75 (-0.75) es muy buena a excelente ".(4)

Sin embargo, esta interpretación depende del tipo de estudio en cuestión. En estudios experimentales donde se controlan una buena cantidad de fuentes de error es posible aplicar esta recomendación. En estudios observacionales en los que no se controlan adecuadamente las fuentes de error correlaciones entre 0.15 a 0.25 (-0.15 a -0.25) pueden indicar asociaciones importantes.

El hecho es que el estado de desarrollo de muchas de las ciencias del comportamiento es tal, que no mucha varianza de la variable dependiente se puede predecir. Una correlación pequeña de $r=.15$ implica que el coeficiente de determinación $r^2=.0225$, muestra una pequeña relación entre los pares de observaciones de X y Y. Sin embargo, muestra una relación que podía no haber sido perceptible en una observación casual. Por ello el *coeficiente de correlación depende de la naturaleza del problema que se esté estudiando.*(9)

1. Coeficiente de determinación.

El coeficiente de correlación se eleva al cuadrado para formar la medida estadística llamada coeficiente de determinación, que indica r^2 % de la variabilidad de los valores de la variable respuesta o dependiente (Y) se pueden explicar si los valores de la variable predictorica o independiente (X) se conocen (o viceversa).

El coeficiente de determinación se interpreta como la cantidad de variación en una variable que puede definirse por el conocimiento de una segunda variable; sin embargo también es conveniente presentar límites de confianza o los resultados de una prueba estadística para la significación del coeficiente de correlación.

2. Prueba t para correlación.

El símbolo para el coeficiente de correlación en la población es ρ . En una muestra aleatoria, ρ se estima según sea r . Si de una población dada se seleccionan varias muestras aleatorias del mismo tamaño y se calcula el coeficiente de correlación de cada una, entonces éstas correlaciones varían de una a otra, pero siguiendo cierto patrón de distribución alrededor del valor para la población ρ . La distribución muestral de la correlación, no se comporta de manera tan uniforme como la distribución muestral de la media.

Esto se debe principalmente al efecto de techo que causa el valor del parámetro de la población, pues cuando éste se aproxima a -1 o $+1$, entonces los valores de la muestra pueden variar desde el valor del parámetro de la correlación hasta $+1$ o hasta -1 haciendo que se produzca una distribución sesgada y ello es por que el parámetro de población es diferente de 0. De otra manera, si $\rho=0$, la distribución de la muestra se ubicaría aproximadamente de acuerdo a la distribución t . (4)

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

Por ello, la distribución t puede usarse para probar la hipótesis de que *el valor verdadero del parámetro de la población ρ es igual a cero*. La siguiente expresión matemática que emplea el coeficiente de correlación, tiene una distribución t con $n-2$ grados de libertad.

Se usa el estadístico t para probar si el valor encontrado del coeficiente de correlación (r) con n observaciones, es evidencia suficiente para concluir que el valor de la población verdadero de la correlación ρ es diferente de cero.

1° $H_0 : \rho=0$ (la correlación verdadera es cero. Ello implicaría que no hay asociación entre las dos variables estudiadas).

$H_1 : \rho \neq 0$ (la correlación verdadera no es cero).

Estas hipótesis nos conducen a una prueba de dos colas o bidireccional. Una prueba de dos colas es apropiada cuando no se espera algo a priori respecto al valor a observar en la prueba. Lo único que se quiere, es saber si la muestra es diferente de la correlación de la población en cualquier dirección.

- 2° La prueba estadística apropiada para probar las hipótesis, es la distribución t , pues puede usarse cuando se cumplen las premisas de correlación.
- 3° Seleccionar el nivel de significancia para la prueba estadística. Se designa el valor alfa α que es la probabilidad de rechazar la hipótesis nula, cuando realmente es verdadera, por lo que dicha probabilidad deberá ser pequeña. Los valores tradicionales para α son: 0.05, 0.01.
- 4° Determinar el valor que la prueba estadística debe alcanzar para declararse significativa, es decir determinar el valor crítico de aceptación y de rechazo que será determinado por el valor escogido para α . Entonces el valor de la distribución t con $n-2$ grados de libertad que divide el área en $100(1 - \alpha)\%$ central y el $(\alpha(100)\%)/2$ superior e inferior, será el que nos señale la tabla de distribución de t en el nivel de significancia α y con $n-2$ grados de libertad.

Se rechazará la hipótesis nula de no correlación, si el valor absoluto de t es mayor que el que encontramos en la tabla de distribución t .

- 5° Calcular la distribución t :

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

- 6° Obtener las conclusiones.

Si el valor que resulte de t es mayor que el obtenido en el paso 4, se rechaza la hipótesis nula de correlación igual a cero y se concluye que la asociación que se está estudiando es significativamente diferente de cero.

3. Transformación z de Fisher.

Esta transformación es utilizada para probar la correlación cuando el parámetro de la población es diferente de cero y luego se usa la distribución normal (z) estándar. La transformación es necesaria, pues cuando $\rho \neq 0$ la distribución de los valores muestrales es sesgada.

Esta prueba permite probar cualquier hipótesis nula como ρ es igual a otro valor, es mayor que el valor encontrado o bien si $\rho=0$, como en la prueba t, y también formar intervalos de confianza. (4)

La transformación z es la siguiente:

$$z(r) = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$$

La tabla de distribución normal da los valores de z para diferentes valores de r. Con tamaños de muestra moderados, esta transformación sigue una distribución normal y puede usarse la siguiente expresión para la prueba z:

$$z = \frac{z(r) - z(\rho)}{\sqrt{1/(n-3)}}$$

Ejemplo:

Si se quiere saber la relación que hay entre los niveles de plomo en sangre y los niveles de plomo en aire, es mayor de 0.80:

1° $H_0 : \rho \leq 0.80$ (La correlación verdadera es menor o igual a 0.80)

$H_1 : \rho > 0.80$ (La correlación verdadera es mayor de 0.80).

2° Con $\alpha=0.01$

3° La prueba alternativa especifica una prueba de una cola, donde el valor de la distribución z divide el área en un 99% inferior y el 1% restante superior es 2.326 (tabla distribución normal).

Se rechaza la hipótesis nula de que la correlación es 0.80 o menos, si el valor de la prueba estadística de z, es mayor que 2.326.

4° Encontrar los valores transformados para $r=0.85$ $\rho=0.80$
Estos valores son 1.256 y 1.099 respectivamente.

Calculando la prueba z:

$$z = \frac{1.256 - 1.099}{\sqrt{1/(4300 - 3)}}$$

$$z = 10.29164$$

Debido a que en éste ejemplo $z > 2.326$, se rechaza la hipótesis nula de que la correlación es de 0.80 o menos y se concluye que la correlación entre el nivel de plomo en sangre y el nivel de plomo en aire excede de 0.80, con $p \leq 0.01$.

4. Intervalos de confianza para la correlación.

Del ejemplo anterior, se puede calcular el intervalo de confianza al 99% , usando la distribución z de la tabla de distribución normal para encontrar el valor crítico para el coeficiente de correlación $r=0.85$.

El intervalo de confianza es:

$$\text{correlación transformada} \pm 2.575 * \text{error estándar}$$

$$\begin{aligned} &\text{correlación transformada} \pm 2.575 * (1/n-3)^{1/2} \\ &= 1.256 \pm (2.575) * (0.0152551) \\ &= 1.256 \pm 0.039282 \\ &= 1.2167 \text{ a } 1.2953 \end{aligned}$$

Transformando los valores 1.2167 y 1.2953 de nuevo a correlaciones por medio de la tabla de valores críticos para la distribución F: $r=0.84$ y $r=0.87$. El intervalo con 99% de confianza para la correlación observada de 0.85 es de 0.84 a 0.87. Esto quiere decir que se tiene un 99% de confianza de que el valor verdadero de la correlación en la población, se encuentra dentro de éste intervalo.

5. Coeficiente de correlación de rango de Spearman.

El coeficiente de correlación de Pearson es muy sensitivo a valores extremos o errores que hacen que la distribución se sesgue y la medida que da el coeficiente de correlación no sea la real. Lo que se quiere es calcular una medida de asociación que sea más robusta y eso se logra con el coeficiente de correlación del **rango** de Spearman, denotado por r_s , que es el coeficiente de correlación de Pearson calculado para los valores de rango de x y y .

$$r_s = \frac{\sum_{i=1}^n (x_{ri} - \bar{x}_r)(y_{ri} - \bar{y}_r)}{\sqrt{\sum_{i=1}^n (x_{ri} - \bar{x}_r)^2 \sum_{i=1}^n (y_{ri} - \bar{y}_r)^2}}$$

donde x_{ri} , y_{ri} son los rangos asociados con el i -ésimo sujeto de las observaciones. Una fórmula equivalente es:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

donde n es el número de datos que contiene la muestra, y d_i es la diferencia entre los rangos de x_i y el rango de y_i .

También se utiliza la prueba t para determinar si existe una correlación entre X y Y , probando la hipótesis nula de que no hay semejante correlación en la población bajo estudio.

El método de correlación de rangos de Spearman se usa cuando se desea medir la relación entre dos variables **ordinales**(*). También puede usarse con variables numéricas si una o las dos *no tienen una distribución normal*. De manera alterna, una media numérica puede transformarse de modo que manifieste distribución normal y luego aplicar la correlación de Pearson.

CAPITULO IV

REGRESION LINEAL

A. Definición.

Al término regresión lineal, - también se le llama regresión lineal simple o regresión de mínimos cuadrados-, se refiere al hecho de que mide sólo una relación en línea recta o lineal, entre dos variables continuas. Cuando el término simple se usa en la regresión se refiere a la situación donde sólo se usa una variable explicatoria (independiente) para predecir la otra. En la regresión múltiple, se incluye más de una variable explicatoria en la ecuación de predicción.

B. Supuestos estadísticos de la Regresión Lineal.

Si se desea usar una ecuación de regresión, los datos deben tener las siguientes propiedades:

1. Existencia.

Dado un determinado valor de la variable X , Y es una variable aleatoria con una cierta distribución probabilística con *media y varianzas finitas*.

2. Independencia.

Se supone que los valores de Y son independientes uno de otro. Esto no se cumple cuando se hacen mediciones repetidas en los mismos sujetos.

3. Linealidad.

Requiere que los valores medios de Y que corresponden a varios valores de X , caigan en una línea recta. La linealidad entre (X, Y) se estima mediante la inspección de gráficos de distribución. (fig 4-1)

4. Homocedasticidad. (Homogeneidad)

Varianza igual de las Y a través del rango entero de las X ; es decir, para cualquier valor de X , $\sigma_{Y|X}$ - la desviación estándar de los resultados de Y - no cambia. (19)

5. Normalidad.

Para cada valor de la variable X , se supone que la variable Y tiene una distribución normal y también se supone que la media de la distribución es el valor predicho o presunción, Y^* . (fig. 4-2)

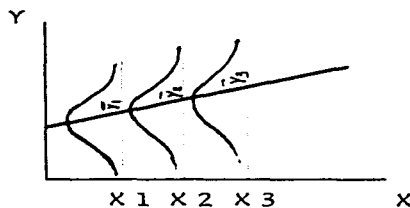


fig. 4-1. Esquema de los supuestos de Regresión Lineal.

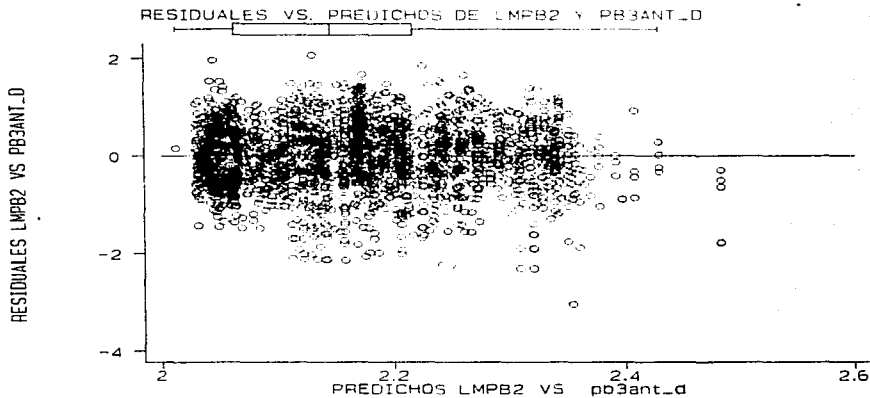


fig. 4-2. Gráfico de Residuales versus Predichos

6. Transformaciones.

Una transformación expresa los valores de observaciones en otra escala de medición. Se utiliza para estabilizar la varianza (o la desviación estándar) de modo que sea independiente del valor de la media.

Hay dos tipos de transformaciones: las lineales y las no lineales. Las lineales se usan por conveniencia, como la *transformación z*, usada para expresar la media igual a cero y la desviación estándar igual a 1.

Las transformaciones *no lineales* se utilizan cuando se necesita que las observaciones sigan una distribución normal y son las que se presentan a continuación:

* *Transformación logarítmica*. Extiende las observaciones con valores menores y comprime los valores mayores.

$$X' = \log(X)$$

* *Transformación extrayendo raíz cuadrada*. Tiene el mismo efecto que la transformación logística. (También se utiliza cuando las observaciones tienen una distribución Poisson)

$$X' = (X)^{1/2}$$

* *Transformación arcoseno*. Cuando las observaciones siguen una distribución binomial, en donde el tamaño de la varianza se relaciona con el tamaño de la media.

$$X' = \arcsen(X/n)^{1/2}$$

Los datos transformados resultantes se distribuirán en forma aproximadamente normal.

* *Transformación logística (logit)*. Se utiliza para analizar relaciones de dosis-respuesta o datos de supervivencia cuando la respuesta es una proporción.

$$\ln(p/(1-p))$$

STATA® ofrece gran variedad de transformaciones con el comando *ladder*.

* Cúbica	variable^3
* Cuadrada	variable^2
* "Cruda"	variable
* Raíz cuadrada	$\text{sqrt}(\text{variable})$
* Logarítmica	$\text{log}(\text{variable})$
*Recíproco de raíz cuadrada	$1/\text{sqrt}(\text{variable})$
*Recíproco de variable	$1/\text{variable}$
*Recíproco cuadrado de la variable	$1/\text{variable}^2$
*Recíproca cúbica de la variable	$1/\text{variable}^3$

La regresión lineal puede usarse inclusive en numerosas situaciones donde las premisas no se cumplen, siempre que las mediciones sean bastante confiables y se use el modelo correcto de análisis de regresión.

C. Método de mínimos cuadrados.

El método de mínimos cuadrados es una forma para determinar la ecuación de la línea que se ajusta con precisión a los puntos.

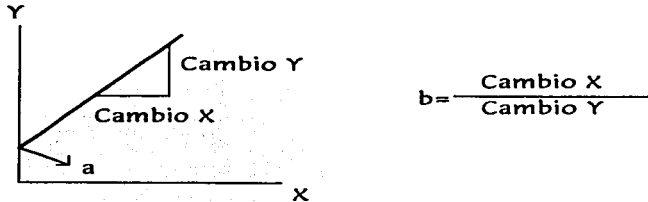


figura 4-4. Interpretación geométrica de la línea de regresión.

a es el punto donde la línea interseca el eje de las Y y **b** es la pendiente de la línea. La ecuación es:

$$Y = a + bX$$

La pendiente de la línea (**b**) mide el cambio en Y para cada unidad de cambio en X. En un modelo de regresión, la pendiente en la población se simboliza con β_1 , designado coeficiente de regresión. β_0 denota la intersección de la línea de regresión. β_0 y β_1 son los parámetros de población en la regresión.

Los valores de Y obtenidos con la ecuación de regresión, son predicciones o *valores predichos*, más que valores reales de Y. Las predicciones de Y se denotarán por Y' . Entonces la ecuación de regresión para la población es:

$$Y' = \beta_0 + \beta_1 X + e$$

El modelo de regresión incluye un término de error aleatorio **e**, se asume que éste error tiene media cero y varianza desconocida. Se supone que los errores no están correlacionados; es decir, son independientes.

Si la desviación del error, e , es positiva, entonces el punto (X, Y) se sitúa arriba de la línea; si e es negativo entonces (X, Y) se sitúa debajo de ésta. (19)

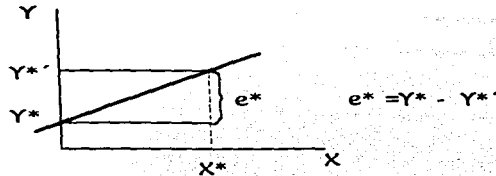


figura 4-5. Línea de regresión de cuadrados mínimos.

La diferencia entre el valor real para Y^* y la predicción, $e^* = Y^* - Y^{*'}$, proporciona un criterio para juzgar cuanto se ajusta la línea a los puntos de los datos.

El método de los mínimos cuadrados determina la línea que reduce a un mínimo la suma de las diferencias, e , al cuadrado entre los valores reales y predichos de la variable Y ; β_0 y β_1 se determinan de modo que:

$$\sum (Y - Y')^2$$

sea mínima; es decir que β_0 y β_1 tengan una desviación mínima

$$R = \sum (Y - Y')^2$$

$$= \sum (Y - \beta_0 - \beta_1 X)^2$$

(R es el cuadrado total de las desviaciones estándar)

Donde \sum denota la suma de $i=1, \dots, n$

Debido a que β_0 y β_1 son parámetros arbitrarios, se determinan en forma que minimicen a R .

El mínimo de R se obtiene si las derivadas parciales con respecto a β_0 y β_1 se anulan. (18)

$$\begin{aligned}
\frac{\delta R}{\delta \beta_0} &= (\sum Y - \beta_0 - \beta_1 X)^2 \\
&= 2 \sum \frac{\delta R}{\delta \beta_0} (Y - \beta_0 - \beta_1 X) \\
&= 2 \sum (Y - \beta_0 - \beta_1 X) \\
&= -2 \sum (Y - \beta_0 - \beta_1 X) = 0
\end{aligned}$$

$$\begin{aligned}
\frac{\delta R}{\delta \beta_1} (\sum (Y - \beta_0 - \beta_1 X)^2) &= 2 \sum \frac{\delta R}{\delta \beta_1} (Y - \beta_0 - \beta_1 X) \\
&= 2 \sum (Y - \beta_0 - (\beta_1 X + X(1))) \\
&= 2 \sum (Y - \beta_0 - \beta_1 X - X) \\
&= -2 \sum X (Y - \beta_0 - \beta_1 X) = 0
\end{aligned}$$

Después se dividen $\frac{\delta R}{\delta \beta_0}$ y $\frac{\delta R}{\delta \beta_1}$ entre -2. Se puede escribir como:

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

Donde $A_{1,1} = n$, $A_{1,2} = A_{2,1} = \sum X$; $A_{2,2} = \sum (X)^2$; $Z_1 = \sum Y$; $Z_2 = \sum X Y$
 La solución del sistema de ecuaciones se escribe como:

$$\beta_0 = \frac{A_{2,2} Z_1 - A_{1,2} Z_2}{A_{1,1} A_{2,2} - A_{1,2} A_{2,1}} \quad \beta_1 = \frac{A_{1,1} Z_2 - A_{2,1} Z_1}{A_{1,1} A_{2,2} - A_{1,2} A_{2,1}}$$

O bien se resuelve el sistema de ecuaciones en forma simultánea y se obtiene en términos de los estimados a y b de la muestra:

$$b = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} \quad ; \quad a = \bar{Y} - b \bar{X}$$

D. Cálculo de la ecuación de regresión.

Para calcular la ecuación de regresión, el método de los mínimos cuadrados usa los datos originales y las fórmulas para **a** y **b**. Otro método se apoya en la relación entre **r** y **b**, donde **r** es el coeficiente de correlación, y **b** es la pendiente de la línea de regresión. También se le llama coeficiente de regresión.

Cuando: **r** es positivo → **b** tiene pendiente positiva
 r es negativo → **b** tiene pendiente decreciente
 r es cero → **b** no tiene pendiente = 0.

Si **r** se conoce, puede multiplicarse por la razón de la desviación estándar de **Y** a la desviación estándar de **X**, s_y / s_x , para obtener **b**:

$$b = r \frac{s_y}{s_x}$$

De igual modo, si el coeficiente de regresión se conoce, **r** puede encontrarse por:

$$r = b \frac{s_x}{s_y}$$

ej. Se supone que el nivel de plomo en sangre del cordón umbilical (CPB2) depende del nivel de plomo en el torrente sanguíneo de la madre (MPB2).

El CPB2 se ha regresado sobre el nivel de plomo en sangre de la madre y la ecuación de regresión es:

$$CPB2' = 1.54 + 0.756 * MPB2$$

con $p < 0.0000$, IC 95% [0.73, 0.77] y $r^2 = 0.5928$

Interpretación: La ecuación de regresión tiene una intersección positiva de 1.54, de modo que se predice que una mamá con nivel de plomo en sangre=0 tendrá un nivel de plomo en la sangre del cordón umbilical de 1.54 (aún cuando es muy poco probable un nivel de plomo en sangre de la madre=0). La pendiente de 0.756, indica que por cada 1µg/dl de aumento en el nivel de plomo en sangre de la madre, el nivel de plomo en cordón umbilical se incrementa 0.75µg/dl. Además el 59% de la variabilidad del nivel de plomo en cordón es explicada por el nivel de plomo en sangre de la madre.

Se deben hacer pruebas estadísticas para saber si la relación es significativa o no, además de determinar que tan probable es cualquier relación entre las variables X y Y. Hay dos formas de llevar esto a cabo: 1° Mediante pruebas de hipótesis, 2° Formando intervalos de confianza.

(En *STATA*® con teclear : `regress cph2 mph2`, nos proporciona además de los datos de la ecuación de regresión, r , r^2 , error estándar, resultado del estadístico de prueba "t", así como su probabilidad, intervalos de confianza de 95% -por default- los datos de los residuales y el número de observaciones y grados de libertad con que realizó los cálculos).

Error estándar de la estimación.

Las líneas de regresión pueden variar de una muestra a otra de una misma población. La ecuación de regresión muestral es sólo una estimación de la ecuación de regresión de la población a la que pertenece la muestra. Si se calcula la ecuación de regresión para varias muestras de una misma población, las ecuaciones variarán de una muestra a otra respecto a sus pendientes y a sus intersecciones. La estimación de esa variación se le llama **error estándar de la estimación** y se simboliza por $S_{y,x}$.

$$S_{y,x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n-2}}$$

Debido a que tanto *la pendiente* como *la intersección* pueden variar, tiene valor efectuar una prueba estadística para cada una. En los dos casos puede usarse la prueba "t" mediante comprobación de hipótesis o por cálculo de intervalos de confianza.

Inferencias acerca de la intersección.

Se emplea la comprobación de hipótesis para probar si la intersección se separa en forma significativa de cero, lo que implica que al calcular varias ecuaciones de regresión de diferentes muestras en una misma población, sus intersecciones no sean iguales.

- 1° $H_0: \beta_0 = 0$ (Las intersecciones son iguales)
 $H_1: \beta_0 \neq 0$ (Las intersecciones no son iguales)

2° la prueba estadística es la razón t:

$$t = \frac{a - \beta_0}{\sqrt{S^2_{y,x} [1/(n+1) + \frac{\bar{x}^2}{\sum (x - \bar{x})^2}]}}$$

donde el denominador es el error estándar de la intersección (a)

3° Escoger el nivel de significancia α .

4° Buscar el valor de la distribución t con (n-2) grados de libertad que divide al área en $100(1 - \alpha) \%$ central y el $100(\alpha)\%$ del combinado superior e inferior en las tablas.

5° Se rechazará la H_0 de una intersección en cero, si el valor absoluto de la prueba estadística t es mayor que el observado en tablas.

6° Si $t < \text{Valor observado en tablas}$, entonces H_0 no puede rechazarse. Se concluye que no hay evidencia suficiente para demostrar que la intersección es diferente de 0 para la regresión de "Y" sobre "X": es decir, el valor observado de a (intersección) podría presentarse por casualidad cuando la intersección es *cero*.

Si $t > \text{Valor observado en tablas}$, entonces se acepta H_1 . Se concluye que la intersección se separa en forma significativa de cero. Si se quiere inferir mediante pruebas de hipótesis acerca de la pendiente de la población β_1 :

$H_0: b = \beta_1$ ($\beta_1 = 0$). Implica que no hay relación lineal entre X y Y).

$$t = \frac{(b - \beta_1) / (S^2_{y,x} / \sum (x - \bar{x})^2)^{1/2}}$$

donde el denominador es el error estándar de la pendiente, b. (19)

Inferencias acerca del coeficiente de regresión, de la población β_1 .

El intervalo de confianza conserva los efectos de variación entre muestras en la estimación estadística.

* Forma general de los intervalos de confianza:

Cada intervalo de confianza se obtiene por sustraer y añadir a la estadística estimada (o a una transformación) un múltiplo de su error estándar.

La distribución t para regresión, o la distribución normal para correlación. El múltiplo se toma como el valor que corresponde a la inclusión de $100(1-\alpha)\%$ central de la distribución.

Esto se puede hacer mediante intervalos de confianza.

$$b \pm t_{(n-2)} \sqrt{S^2_{y,x} \left[\frac{1}{\sum (X - \bar{X})^2} \right]}$$

donde:

b = Coeficiente de regresión de la muestra.

$t_{(n-2)}$ = Valor absoluto del valor observado en tablas. De la distribución t con $(n-2)$ grados de libertad.

$S^2_{y,x}$ = El cuadrado del error estándar de la estimación.

$\sum (X - \bar{X})^2$ = La suma de los cuadrados de cada dato de la variable independiente menos su media.

* Si el intervalo excluye el 0, puede tenerse $100(1-\alpha)\%$ de confianza de que el coeficiente de regresión **no** es 0; sino que está entre:

$$b - t_{(n-2)} \sqrt{S^2_{y,x} \left[\frac{1}{\sum (X - \bar{X})^2} \right]} \quad \text{y} \quad b + t_{(n-2)} \sqrt{S^2_{y,x} \left[\frac{1}{\sum (X - \bar{X})^2} \right]}$$

También se pueden calcular los intervalos de confianza para el intercepto de la población β_0 . Sin embargo, si los datos observados tienden a estar alejados de la intersección resultará muy poco práctico hacer inferencias acerca de la misma. De hecho es peligroso extrapolar la línea ajustada *más alla* del rango entero de las X ; la relación entre X y Y podría ser bastante diferente fuera de este rango. (19)

E. Predicción con la ecuación de regresión.

Se obtiene una ecuación de regresión para predecir valores futuros en grupos de sujetos o para personas independientes. En cualquiera de los dos casos, la variabilidad relacionada con la línea de regresión debe estar reflejada en la predicción. El intervalo de 95% de confianza para una media probable (predicción) Y' en un grupo de sujetos es:

$$Y' \pm t_{(n-2)} \sqrt{S^2_{y,x} \left[\frac{1}{n} + \frac{(X - \bar{X})^2}{\sum(X - \bar{X})^2} \right]}$$

El intervalo de confianza de 95% para predecir un valor individual es:

$$Y' \pm t_{(n-2)} \sqrt{S^2_{y,x} \left[1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum(X - \bar{X})^2} \right]}$$

Se le agrega la unidad al término de error estándar pues habrá mayor variación en la predicción para un solo valor, que en las de grupos de observaciones.

* El tamaño del error estándar depende de que tan próximo esté el valor X de \bar{X} y mientras más pequeña sea la distancia, *mayor* será la exactitud de la predicción de Y .

* Los errores estándar y los intervalos de confianza, toman sus valores más pequeños cuando $X = \bar{X}$. Conforme X se separa en cualquier dirección de su media, \bar{X} , los errores estándar y los intervalos de confianza se vuelven cada vez mayores, reflejando la diferencia al cuadrado entre $X - \bar{X}$. Si los intervalos de confianza se grafican como **bandas de confianza** alrededor de la línea de regresión, están más próximas a la línea en la media de X , y se curvan alejándose de ésta en ambas direcciones a cada lado de \bar{X} .

* Cuando la media de X se usa en la ecuación de regresión, la predicción Y' es la media de Y . La línea de regresión atraviesa el punto (\bar{X}, \bar{Y}) .

F. Análisis de residuos y puntos influyentes.

El análisis de residuos sirve para valorar el ajuste de la ecuación de regresión. Un residual es la diferencia entre el valor real de Y y el valor de predicción Y' o $Y - Y'$. Es la parte de Y que no se predice a partir de X (la parte sobrante o residual). Los valores residuales, $Y - Y'$, en el eje Y se grafican *contra* los valores en el eje X . La media de los residuales es 0 y debido a que la pendiente se ha sustraído en el proceso de calcular los residuales, la correlación entre éstas y los valores de X deberá de ser cero.(4)

Si el modelo de regresión proporciona un ajuste adecuado para los datos, los valores de los residuales *no* se relacionan con los X (fig. 4-6). Entonces la gráfica de residuales y valores de X debería parecerse a puntos dispersos sin patrón (fig. 4-7), donde no hay correlación entre los residuales y los valores de X .

Graficos del análisis de residuales

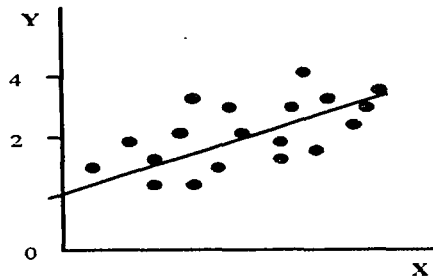


figura 4-6. Relación lineal entre X y Y

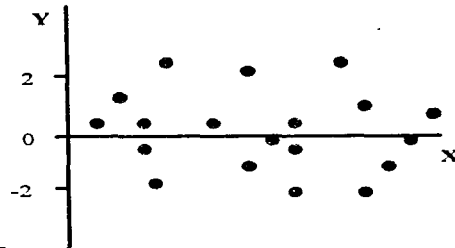


figura 4-7. Residuales contra valores de X para la relación en fig 4-6

Evaluación del modelo.

Mediante el coeficiente de determinación R^2 , que es el coeficiente de correlación de Pearson al cuadrado r^2 . El coeficiente de determinación puede ser interpretado como la proporción de la variabilidad entre los valores observados de Y que es explicada por la regresión lineal de Y en X .

Mediante el gráfico de residuales.

Otra estrategia para evaluar que tan bien se ajusta la línea de mínimos cuadrados a los datos observados es graficar los *residuales* contra los valores ajustados o *predichos* de la variable respuesta. (fig. 4-2)

El gráfico de residuales tiene tres propósitos:

1°. Puede ayudar a detectar puntos aberrantes en la muestra, o errores en la información. El método de mínimos cuadrados es muy sensible a dichas observaciones distantes en los datos. Cuando se cree que un valor distante es resultado de un error en la medición o captura de una particular observación, se borra dicha observación con lo que se perfecciona el ajuste de la línea de regresión. Sin embargo, se debe de tener cuidado de no borrar datos que están inusualmente alejados y que son verdaderos (no son errores); pues podrían ser las observaciones mas interesantes de toda la base de datos.

2° Puede sugerir error en la asunción de la homocedasticidad. Si el rango en magnitud de cada residual se incrementa o decrementa como Y' se hace más grande - produciendo una distribución que toma la forma de un abanico tal como se muestra en la (fig. 4-8) - entonces la $\sigma_y | x$ no es constante a través del rango entero de las X .

En esta situación, la regresión lineal simple no es la técnica apropiada para modelar la relación entre X y Y .

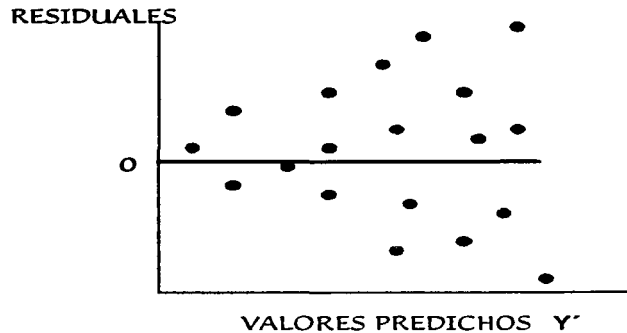


Fig. 4-8 Violación de homocedasticidad

3° Si los residuales no exhiben una distribución aleatoria, pero en su lugar siguen una tendencia distinta, esto sugiere que la verdadera relación entre X y Y podría no ser lineal. Bajo estas circunstancias, una transformación para cada X o Y o ambas debe ser calculada, y después usar la regresión lineal simple que mejor se ajuste a los datos transformados. (19)

Si hay problemas graves con la falta de linealidad de los datos, pueden seguirse varios procedimientos:

1°. *La transformación de los datos* puede volver lineal la relación y luego pueden usarse los métodos regulares de regresión en los datos transformados.

2°. *Si los datos forman una curva*, lo que se puede hacer es una regresión lineal por fragmentos.

3°. *Una regresión polinomial* si los datos forman curvas.

Principales semejanzas y diferencias entre el coeficiente de regresión y el coeficiente de correlación.

☑ El coeficiente de regresión, b , puede situarse en un rango de $-\infty \leq b \leq \infty$, y esto expresa la magnitud de un cambio en Y asociado a una unidad de cambio en X . El coeficiente de correlación va de $-1 \leq r \leq 1$.

☑ El coeficiente de correlación no es una medida de la cantidad de cambio de una variable con respecto a otra, pero sí es una medida de asociación entre las dos variables. (9)

☑ Si se conocen las medias y desviaciones estándar de las dos variables, es posible calcular el coeficiente de correlación a partir de la ecuación de regresión y viceversa. Además, si el coeficiente de correlación tiene significado estadístico, la ecuación de regresión también será estadísticamente significativa. (4)

☑ La correlación es independiente de la escala de medición. La ecuación de regresión depende de la escala que se esté utilizando.

☑ La pendiente de la línea de regresión tiene el mismo signo (+ o -) que el coeficiente de correlación.

☑ El coeficiente de regresión permanece constante bajo cambios en la variabilidad de X , mientras que el coeficiente de correlación no. (3)

G. Regresión Múltiple

El análisis de regresión múltiple es una extensión natural de la regresión lineal simple y se usa para investigar una más complicada relación entre más de dos variables explicatorias (independientes) y una variable respuesta (dependiente).

El modelo de regresión múltiple es:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_q x_q + \varepsilon$$

donde x_1, x_2, \dots, x_q son las q distintas variables explicatorias. Como en la regresión simple ε es el error aleatorio asociado con y . El intercepto de la población α es el valor medio de la variable respuesta cuando todas las variables explicatorias toman un valor de cero, la pendiente de la población β_i es el cambio estimado en y que corresponde a una unidad de incremento en x_i dado que todas las otras variables explicatorias permanecen constantes.

Supuestos que se deben cumplir antes de hacer un modelo de regresión múltiple:

1. Para valores específicos de x_1, x_2, \dots, x_q los resultados de y son normalmente distribuidos con $\mu_{y|x_1, x_2, \dots, x_q}$ y desviación estándar $\sigma_{y|x_1, x_2, \dots, x_q}$.

2. La relación entre $\mu_{y|x_1, x_2, \dots, x_q}$ y x_1, x_2, \dots, x_q se representa por la ecuación:

$$\mu_{y|x_1, x_2, \dots, x_q} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_q x_q$$

3. Para cualquier conjunto específico de valores de y y x_1, x_2, \dots, x_q , $\sigma_{y|x_1, x_2, \dots, x_q}$ es constante.

4. Los resultados de y son independientes.

Ecuación de regresión de mínimos cuadrados:

Para estimar la ecuación de regresión de la población

$$\mu y | x_1, x_2, \dots, x_n = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

se usa el método de los mínimos cuadrados para ajustar el modelo

$$y' = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Esta técnica requiere que se minimice la suma del cuadrado de los residuales, es decir:

$$\sum_{i=1}^n (y_i - y'_i)^2 = \sum_{i=1}^n (y_i - a - b_1 x_{1i} - b_2 x_{2i} - \dots - b_n x_{ni})^2$$

Cuando se tiene una sola variable explicatoria se tiene la ecuación de la recta. Con dos variables explicatorias el modelo representa un plano en el espacio tridimensional, con tres o más variables un hiperplano en un mayor espacio dimensional.

Evaluación del modelo.

El modelo se puede evaluar mediante el coeficiente de determinación y el gráfico de residuales. La inclusión de una variable adicional en un modelo nos puede causar que el coeficiente de determinación se decremente, por ello se utiliza el coeficiente de determinación ajustado que compensa la complejidad de un modelo. A diferencia de la R^2 , la R^2 ajustada se incrementa cuando la inclusión de una variable mejora la habilidad para predecir la variable respuesta y decrece cuando no la mejora. Consecuentemente la R^2 ajustada nos permite hacer una más válida comparación entre modelos que contienen diferentes variables explicatorias.

Variables indicadoras.

En el análisis de regresión además de usar variables explicatorias continuas, se puede generalizar a variables explicatorias discretas o nominales. Las variables explicatorias nominales son las que indican la presencia o ausencia de alguna característica; ejemplo: usa loza vidriada=1, no usa loza vidriada=0, este tipo de variables reciben el nombre de variables indicadoras o dummies.

Selección del modelo.

Si de una regresión particular el coeficiente β , no es significativamente diferente de cero, entonces podemos simplificar el modelo bajo estudio borrando la variable explicatoria x .

Inicialmente se debe tener un conocimiento de cuales variables podrían ser relevantes en un modelo, ejemplo seguir un modelo biológico que explica alguna enfermedad; sin embargo, para estudiar el efecto completo de cada una de las variables explicatorias, será necesario realizar un análisis de regresión para todas las posibles combinaciones de las variables. Los modelos resultantes podrían ser evaluados de acuerdo a algún criterio estadístico. Este es el método más completo, pero también el que consume más tiempo y resulta poco práctico cuando se tienen muchas variables explicatorias. Como resultado de esto se cuenta con el recurso stepwise (pasos acertados) que nos sirven para elegir el modelo de regresión. Son dos los procedimientos de stepwise: el primero es conocido como selección hacia adelante y el segundo como eliminación hacia atrás. (Ver Capítulo VI , Análisis multivariado).

Selección hacia adelante: El método consiste en ir incrementando al modelo las variables que explican mejor la proporción de la variabilidad en y . El procedimiento continúa hasta que ninguna de las variables restantes explican una cantidad adicional de la variabilidad de y .

Eliminación hacia atrás: Se incluyen todas las variables explicatorias en el modelo. Se borran las variables que explican solo una pequeña proporción de la variabilidad en y , una cada vez que se calcula la regresión. Este procedimiento se continúa hasta que no sea significativa la diferencia de R^2 ajustada resultante en dejar o eliminar una variable explicatoria del modelo. (Ver Capítulo VI, Análisis multivariado de Pb en aire 3 meses antes).

CAPITULO V

ANALISIS ESTADISTICO

A. MATERIALES Y METODOS.

Los niveles de plomo en aire, del D.F. y del Estado de México, fueron proporcionados por la **Red Manual de Monitoreo Atmosférico (RMMA)**. Corresponden a los datos de 19 estaciones de monitoreo atmosférico que miden el PLOMO en el aire en $\mu\text{g}/\text{m}^3$. Estas mediciones fueron efectuadas de Enero de 1986 a Marzo de 1995.

1. SISTEMAS DE EVALUACION DE LA CALIDAD DEL AIRE EN LA CIUDAD DE MEXICO.

a. Antecedentes.

A fines de la década de los años 50's a través de la SSA, el Ejecutivo Federal inició estudios para evaluar las condiciones del aire en la zona metropolitana de la Ciudad de México, pero fué hasta 1966 cuando se instaló la primera red de monitoreo atmosférico que contaba con 4 estaciones localizadas en el Municipio de Tlalnepantla, en el Aeropuerto Internacional, en el Centro de la Ciudad y en Tacuba.

Al establecerse la Subsecretaría de Mejoramiento del Ambiente y dentro del programa de Naciones Unidas para el Desarrollo, se adquirieron 48 nuevas estaciones de monitoreo, de las cuales 22 se instalaron en el área Metropolitana de la Ciudad de México. Dichas estaciones contaban con muestreadores de alto volumen para partículas suspendidas totales* y de burbujeadores para gases, destinados a la determinación de SO_2 , NO_x , O_3 , etc.

La Subsecretaria de Mejoramiento del Ambiente adquirió la primera Red Automática de Monitoreo Atmosférico (RAMA), de fabricación holandesa, con 20 estaciones fijas y dos móviles. Las primeras 5 contaban con monitores de partículas, SO_2 , CO^{**} , NO_x y O_3 , así como con equipo meteorológico para la

* SO_2 : Bixido de azufre

NO_x : Oxidos de nitrógeno

O_3 : Ozono

** CO : Monoxido de carbono

determinación de la dirección y velocidad del viento, la humedad relativa y la temperatura. Las estaciones móviles estaban equipadas con instrumental meteorológico y monitores de polvo, SO₂ y CO.

Toda la información de esta red, que era obtenida en forma continua, se procesaba en un minicomputador que generaba los reportes horarios y diarios. Los métodos de determinación utilizados en esta red, eran de tipo colorimétrico medidos a través de un espectrofotómetro (*) para el SO₂ y el CO, quimiluminiscente para el O₃ y de atenuación beta para las partículas menores de 5 micrómetros.

b. LA RED AUTOMÁTICA DE MONITOREO ATMOSFÉRICO.

La Red Automática de Monitoreo Atmosférico (RAMA), se diseñó en base a la experiencia sobre monitoreo atmosférico de la Subsecretaría de Mejoramiento del Ambiente y los criterios de diseño de redes de monitoreo atmosférico publicados tanto por organismos internacionales (OMS-OMM-PNUMA), como por organismos ambientales de otros países (USEPA).

La red fue diseñada con un total de 70 monitores, de los cuales 15 estaban destinados para medir partículas totales en suspensión, de donde se pueden conocer los niveles de plomo en el aire.

CONTAMINANTE	NUMERO	CRITERIO APLICADO
<i>Partículas totales en suspensión</i>	15	<i>PNUMA-OMS-OMN</i>
Bióxido de azufre	15	PNUMA-OMS-OMN/EPA
Oxidos de nitrógeno	15	Promedio ONU/EPA
Ozono	10	Promedio ONU/EPA
Monóxido de carbono	15	EPA

Diversos estudios indicaron la necesidad de contar con 10 estaciones para la medición de dirección y velocidad del viento, humedad y temperatura, así como dos radares acústicos a fin de mantener una vigilancia constante sobre el gradiente vertical de temperatura, factor dominante de la meteorología de la zona.

Con el fin de reducir los problemas de operación y mantenimiento de la red, éstos se reagruparon para ser estaciones de monitoreo combinadas en donde su ubicación fuera representativa de diversas características de las zonas metropolitanas (áreas residenciales, industriales y comerciales).

c. RED MANUAL DE MONITOREO ATMOSFERICO.

La red manual de monitoreo atmosférico fue diseñada con base en los criterios de la OMS-OMM-PNUMA, tanto para el número como para la ubicación de las estaciones.

Las estaciones con equipo de muestreo de partículas totales en suspensión y de gases fueron ubicadas en las zonas representativas de áreas residenciales, comerciales e industriales

La presencia de partículas suspendidas totales en la atmósfera es producto de la dispersión de sólidos provenientes de áreas de cultivo descubiertas de vegetación y emisiones de humo que se generan a partir de las actividades industriales.

Las partículas suspendidas menores a 10 micrómetros (PM10), son las partículas de fracción respirable que permanecen suspendidas en el aire por períodos prolongados de tiempo, con lo cual provocan disminución en la visibilidad atmosférica y reacciones químicas con otros contaminantes. Tienen efectos en la salud, pues al ser inhaladas llegan a depositarse sin dificultad en los pulmones.

Las partículas de plomo son emitidas al aire a partir de diversas fuentes, entre las que destacan la gasolina y sus productos de combustión, algunas pinturas, esmaltes y las fundidoras. Las muestras de partículas obtenidas se someten a análisis por *espectrofotometría de absorción atómica*(*), a fin de determinar las concentraciones de diversos metales pesados, entre ellos el plomo.

En las partículas suspendidas menores a 20 micrómetros se utiliza el método que consiste en la determinación de la dispersión lumínica de las partículas suspendidas en un determinado volumen de muestra.

**d. PARAMETROS REGISTRADOS EN CADA ESTACION DE
MONITOREO DONDE SE MIDE PLOMO
RED MANUAL**

ZONA	ESTACION	CARACT. CLAVE		PARAMETRO	
NOROESTE	SECRETARIA DE HACIENDA	M	SIIA	PST	PB
	TLALNEPANTLA	M	TLA	PST	PM10 PB
NORESTE	LA PRESA	I	LPR	PST	PB
	NEZAHUALCOYOTL SUR	R	NTS	PST	PB
	CERRO DEL TEPEYAC	R	TEC	PST	PB
	CHAPINGO	M	CHA	PST	PB
	XALOSTOC	I	XAL	PST	PM10 PB
CENTRO	MERCED	C	MER	PST	PM10 PB
	HANGARES	M	HAN	PST	PB
	PORTALES	C	POT	PST	PB
	MUSEO DE LA CD DE MEXICO	C	MCM	PST	PB
SUROESTE	MUSEO TEC DE LA C.F.E.	R	CFE	PST	PB
	FELIPE ANGELES	R	FAN	PST	PB
	LOMAS	R	LOM	PST	PB
	PEDREGAL	R	PED	PST	PM10 PB
SURESTE	CERRO DE LA ESTRELLA	R	CES	PST	PM10 PB
	XOCHIMILCO	M	XCH	PST	PB
	TAXQUEÑA	M	TAX	PST	PB
	U.A.M. IZTAPALAPA		UIZ	PST	PB

SIMBOLOGIA

PST PARTICULAS SUSPENDIDAS TOTALES $\mu\text{g}/\text{m}^3$
PM10 PARTICULAS SUSPENDIDAS FRACCION RESPIRABLE $\mu\text{g}/\text{m}^3$
PB PLOMO $\mu\text{g}/\text{m}^3$

CLAVE DE CARACTERIZACION:

R Habitacional.
C Comercial
I Industrial
M Mixta

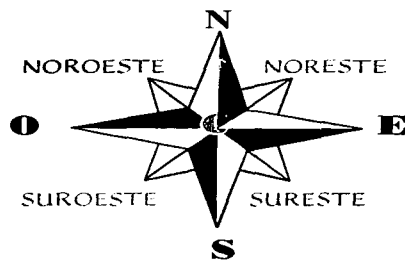
UNIDADES DE REPORTE

**e. UBICACION DE LAS ESTACIONES DE LA RED MANUAL
DE MONITOREO ATMOSFERICO DONDE SE MIDE PLOMO**

ESTACION	CLAVE	UBICACION
SECRETARIA DE HACIENDA	SHA	Laboratorio central de SHA Legaria 600 Col. Irrigación Del. Miguel Hidalgo
TLALNEPANTLA	TLA	Glorieta Atlacomulco, Avenida Toluca s/n Municipio de Tlalnepantla Edo. de México
LA PRESA	LPR	Centro de salud Av. la Presa y Excursionistas Municipio de Tlalnepantla Edo. de México
NETZAHUALCOYOTL SUR	NET	Centro de salud Bordo Nochiaca y Avenida Netzahualcoyotl Cd. Netzahualcoyotl Edo. de México
CERRO DEL TEPEYAC	TEC	Centro de salud 5 Febrero y Progreso Col. Gustavo A. Madero
XALOSTOC	XAL	Distribuidora Santa Clara, Km 13.5 Calle del Hierro, Xalostoc Ecatepec Morelos
CHAPINGO	CHA	Colegio de posgraduados Mex-Texcoco
HANGARES	HAN	Aeropuerto Internacional CIAAC Hangares No. 255 Del. Venustiano Carranza
MERCED	MER	Centro de Salud Ave. Congreso de la Union No. 148 Del. Venustiano Carranza
MUSEO DE LA CD. MEXICO	MCM	Museo de la Cd. de México entre Pino Suarez y Republica del Salvador. Del. Cuauhtemoc
PORTALES	POT	Escuela primaria "Republica de Chipre"
MUSEO TECNOLOGICO C.F.E.	CFE	Delegación Miguel Hidalgo
FELIPE ANGELES	FAN	Escuela Primaria Felipe Angeles, Del Magdalena Contreras
LOMAS	LOM	Edificio Industrial "Resistol" Cuercos No. 222 Del. Miguel Hidalgo
PEDREGAL	PED	Escuela Primaria Cañada esq. Crater
XOCHIMILCO	XCH	Calle Pino Gpe. R. y Nuevo Leon. Col. San Bernardino. Del. Xochimilco
CERRO DE LA ESTRELLA	CES	Planta de tratamiento de aguas, Cerro de la Estrella. Del. Iztapalapa
TAXQUEÑA	TAX	Escuela C. Jaramillo, Taxqueña 1811. Del. Coyoacan
UAM- IZTAPALAPA	UIZ	Escuela "Ciudad Vicentina" Sur 21 y Sur 10 Del. Iztapalapa

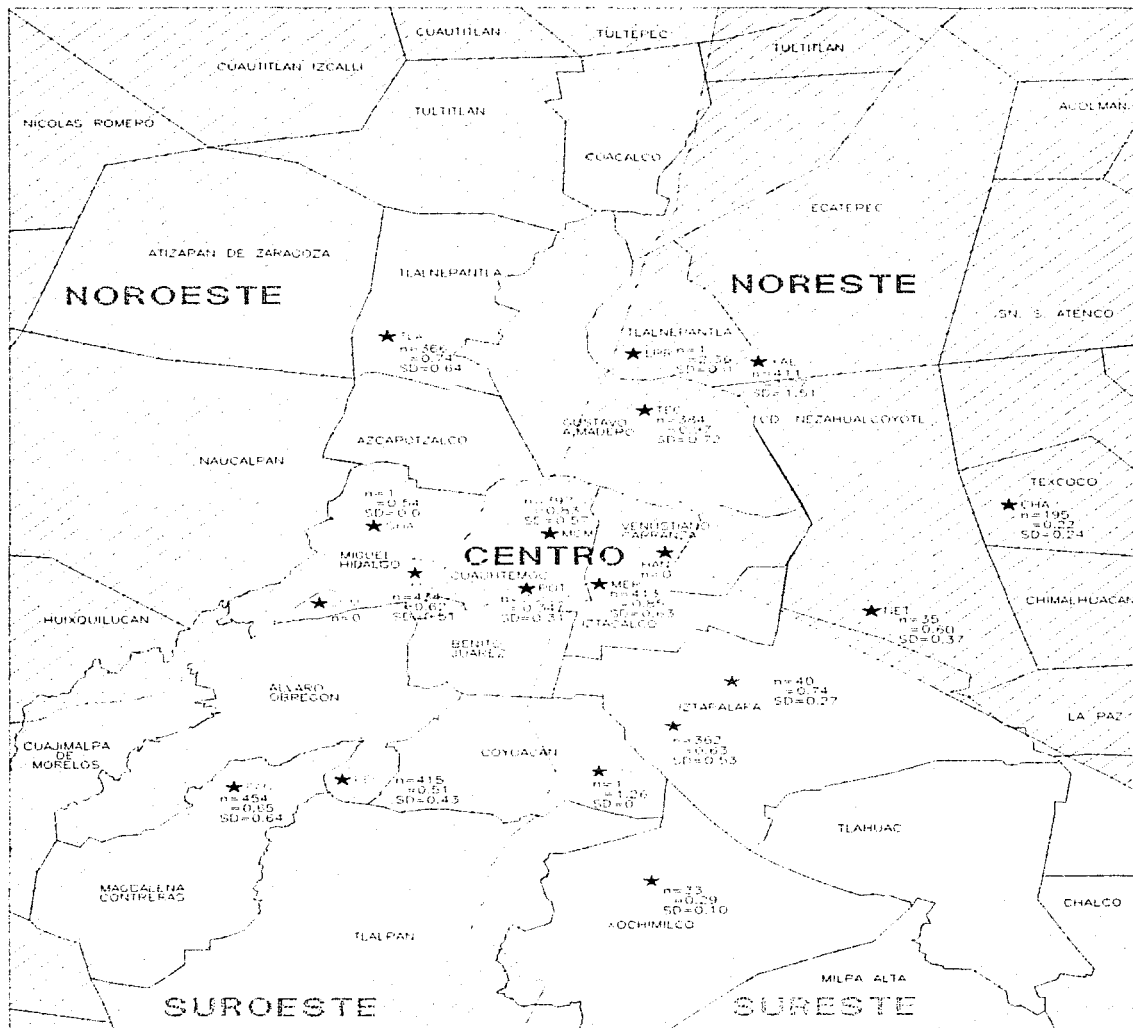
Los colores representan las zonas en donde se localizan las estaciones de monitoreo, para su más fácil identificación en el mapa, de la misma manera como lo representa la rosa de los cuatro vientos.

ZONA



El siguiente mapa muestra la localización de las estaciones de monitoreo de la Red Manual, que son las que miden plomo, así como las Delegaciones y Municipios donde vivían las mamás.

MAPA DE LOCALIZACION DE LAS ESTACIONES DE MONITOREO Pb



La RMMA tiene en promedio de 4 a 5 muestreos de plomo en aire por mes en cada estación de monitoreo, a excepción de las estaciones LPR, SHA y TAX que solo tienen un muestreo y NET, POT, UIZ y XCH que cuentan con muy pocas mediciones. (Ver figura 1 y mapa).

Se hizo el análisis exploratorio de la base de plomo en aire (RMMA) y se eliminaron tres datos de la estación Xalostoc, pues aunque es ésta estación la que siempre presenta niveles de plomo en aire muy elevados incluso rebasa la norma de $1.5\mu\text{g}/\text{m}^3$ promedio trimestral en un sólo mes, los resultados eran extremos comparados con el conjunto de resultados obtenidos en fechas cercanas.

La información que proporcionó el INSP (Instituto Nacional de Salud Pública), en la división del CISP (Centro de Investigación en Salud Poblacional) también refiere el resultado de la determinación de plomo en las muestras de sangre de la madre y del cordón umbilical, al momento del parto. Dicha muestra fué analizada en el *laboratorio de metales traza*, del Hospital ABC y el método de determinación se expresa a continuación:

2. Determinación de plomo en sangre.

1. Principios de la determinación

La determinación de la concentración de plomo se efectúa por **espectrofotometría de absorción atómica**, utilizando el método descrito por Miller y col.(16). La cuantificación se basa en la medición de la luz absorbida a 283.3 nm por átomos en estado elemental, generada por una lámpara de cátodo hueco, del metal en estudio.

En el tubo de grafito, la determinación del elemento se realiza incrementando la temperatura con una secuencia programada del potencial eléctrico, llevándose a cabo en tres etapas principales:

Etapas de secado: El principal propósito de ésta etapa es evaporar los líquidos presentes en la muestra.

Etapas de Calcinación o Mineralización: En esta etapa se quema la materia orgánica de la muestra.

Etapas de Atomización: En ésta etapa se obtienen los átomos del elemento en estado basal.

La contaminación por plomo debe ser cuidadosamente eliminada en todos los pasos de la determinación. Todo el material es monitoreado y el trabajo es realizado bajo condiciones asépticas.

Características del método para la determinación de plomo en sangre por ESPECTROFOTOMETRIA DE ABSORCIÓN ATÓMICA CON HORNO DE GRAFITO, usando modificador de matriz y plataforma de l'vov.

a) Rango de trabajo

El rango lineal de la calibración es de 5.0 - 80.0 μg de plomo/dl de sangre.

b) Sensibilidad del método

Es altamente sensible y sirve para evaluar concentraciones bajas.

c) Límite de detección

El límite de detección estimado es de 1.4 μg de plomo a sangre.

d) Precisión y exactitud del método.

La precisión y exactitud fué estimada $\pm 2-5\%$ en los niveles de concentración de 5.0 - 50 μg de plomo / dl de sangre.

e) Coeficiente de variación.

El coeficiente de variación aceptado es menor o igual al 5%.

f) Instrumentación.

Espectrofotómetro de absorción atómica Perkin-Elmer Mod. 2100 con horno de grafito HGA-700, lámpara de cátodo hueco de plomo, muestreador automático AS-70 e impresora.

2. Control de calidad

El laboratorio de *metales traza* pertenece al programa de certificación del control de calidad de la determinación de plomo de Center for Disease Control (CDC) de Atlanta, E.U.

El CDC les envió mensualmente tres muestras de sangre de bovino con plomo, para analizarlas en dicho laboratorio, y los resultados que obtuvieron no variaron más allá de $\pm 4\mu\text{g}$ Pb/dl.

En el análisis de rutina que aquí llevaron a cabo, junto con cada lote de muestras procesadas, se incuyen dos de los estándares de Atlanta, de valor conocido.

Las muestras son resultado de un proyecto de investigación en mujeres embarazadas de la Ciudad de México, por parte del INSP desde 1992 a 1995. Se tomaron muestras de sangre de 3663 mujeres embarazadas que otorgaron su consentimiento con conocimiento de causa. Se les hizo una breve encuesta para saber su edad, escolaridad, domicilio, si acostumbra guardar y/o preparar sus alimentos en ollas de barro vidriado o no. Se registró el hospital en que nació su hijo y la fecha de nacimiento. Tomaron también 3450 muestras de sangre del cordón umbilical al momento del parto.

Los criterios que se tomaron en cuenta para seleccionar a estas mujeres es que estuvieran embarazadas, fueran mujeres sanas, tanto física como mentalmente, que vivieran en la Ciudad de México o en el Estado de México.

De 3833 registros que conforman la base de datos que se analizó, 3663 (95.51%) registros tienen resultados de plomo en sangre de la madre. Las muestras faltantes se debieron a que el parto fué en la madrugada, hubo algún cambio de turno de las enfermeras o bien la mamá, después de haber aceptado decidió no dar la autorización para la toma de muestras.

La muestra se conformó bajo la condición de tener la fecha de nacimiento del hijo pues con esta fecha y la Delegación donde vivía la mamá, se pudo asignar por medio del programa *-ver apéndice A-* el nivel de plomo en aire correspondiente a su Delegación.



La fecha de nacimiento del hijo fué necesaria, ya que fué la fecha de toma de muestras de sangre de la madre y del cordón umbilical.





3. Criterios de asignación de plomo en aire a plomo en sangre.

Los niveles de plomo en aire se asignaron en forma individual para cada mamá del estudio y los condicionantes estuvieron dados por la Delegación donde vivía la mamá y la fecha en que nació su hijo.

La delegación fue clave para poder determinar que estaciones de monitoreo de plomo en aire de la RMMA eran las más cercanas y/o representativas de la calidad del aire en dicha Delegación y/o Municipio donde vivía la mamá:

DELEGA #	DELEGACIÓN O MUNICIPIO	ESTACIONES DE RMMA MAS CERCANAS			
1	ALVARO OBREGON	LOM	FAN	PED	CFE
2	AZCAPOTZALCO	SHA	TLA	MCM	
3	BENITO JUAREZ	POT	CFE	MER	
4	COYOACAN	TAX	PED	CES	
5	CUAJIMALPA	FAN			
6	CUAUHTEMOC	POT	MCM	MER	CFE
7	GUSTAVO A. MADERO	LPR	TEC	TLA	
8	IZTACALCO	HAN	UIZ	MER	NET
9	IZTAPALAPA	CES	UIZ		

10	MAGDALENA CONTRERAS	FAN			
11	MIGUEL HIDALGO			CFE	
12	MILPA ALTA	XCH			
13	TLAHUAC	XCH			
14	TLALPAN	FAN		PED	XCH
15	VENUSTIANO CARRANZA	MCM		HAN	MER
16	XOCHIMILCO	XCH			
21	ATIZAPAN DE ZARAGOZA	TLA			
23	CHALCO				
24	CHICOLOAPAN				
25	CHIMALHUACAN	CHA		NET	
26	COACALCO	TLA		LPR	XAL
28	CUAUTITLAN IZCALLI				
30	ECATEPEC				

- 31 HUIXQUILUCAN  CFE
- 32 NAUCALPAN TLA   CFE
- 33 NEZAHUALCOYOTL XAL CHA NET
- 34 TEXCOCO
- 35 TLALNEPANTLA  TLA XAL
- 36 TOLUCA

NOTA: LOS SIMBOLOS  INDICAN QUE FUERON REEMPLAZADOS POR NO TENER MUESTREOS SUFICIENTES POR LAS ESTACIONES CON LETRAS EN NEGRILLAS (Ver mapa).

a. NUMERO DE MUESTRAS DE PLOMO EN AIRE
AÑO DE MONITOREO

ESTACION DE MONITOREO	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	TOTAL
CFE	7	18	54	49	51	71	52	59	60	15	436
CES	0	0	8	53	50	70	48	58	60	15	362
CHA	0	0	0	0	0	21	44	58	58	14	195
FAN	15	19	56	55	54	69	52	59	60	15	454
LPR	0	0	1	0	0	0	0	0	0	0	1
MCM	16	18	56	48	53	65	49	18	59	15	397
MER	0	0	53	55	48	69	53	61	60	14	413
NET	13	20	2	0	0	0	0	0	0	0	35
PED	0	0	55	52	53	71	51	59	59	15	415
POT	0	0	0	0	0	0	0	7	0	0	7
SHA	1	0	0	0	0	0	0	0	0	0	1
TAX	0	0	1	0	0	0	0	0	0	0	1
TEC	13	20	6	53	53	71	52	41	60	15	384
TLA	0	0	8	52	52	70	50	59	60	15	366
UIZ	0	0	40	0	0	0	0	0	0	0	40
XAL	0	0	48	53	53	71	52	59	60	15	411
XCH	0	0	0	0	0	28	5	0	0	0	33
TOTAL	65	95	388	470	467	676	508	538	596	148	3951

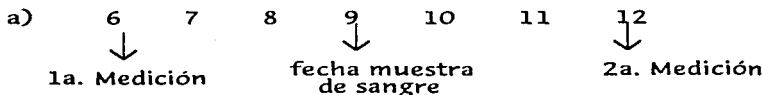
FIGURA 1. FUENTE: RMMA

Se presentaron los siguientes casos al momento de hacer las asignaciones de plomo en aire a plomo en sangre, ya que no todos los días se monitorea el aire por ser muy caro, además de que no variaría demasiado debido a las características del particulado de plomo $1.5\mu\text{g}/\text{m}^3$ de suspensión en el aire por tres meses.

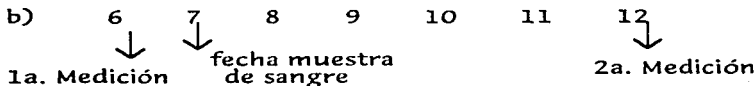
1° Coincide la fecha de toma de muestras sanguíneas con la fecha en que se monitoreó el plomo en aire en las estaciones que corresponden a la Delegación dada. Se asignó el promedio de plomo en aire de las estaciones que representan a dicha Delegación.

2° No hubo medición de plomo en aire en la fecha de toma de muestras de sangre en la Delegación dada. Entonces se aplicaron los siguientes *criterios de asignación de pb en aire en donde no se tenía medición.*

Debido a que se hacen 4 o 5 muestreos por mes, el programa encontraba la medición en aire inmediata anterior a la fecha de toma de muestra de sangre, y también la medición posterior a dicha fecha y se promediaron esos valores para asignarlos a la variable *asigna_d* en la base, dependiendo de los siguientes casos:

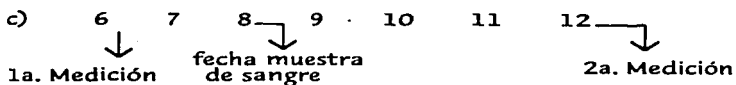


Criterio: se le asignó el promedio de las 2 mediciones
(1a. Medición + 2a. Medición)/2



Criterio: se le asignó el promedio de la 1a medición

Nota. Si la fecha de la muestra de sangre estaba en la posición 11, se le asignó la 2a. medición



Criterio: se le asignó el promedio de las 2 mediciones debido a la cercanía que tiene con la posición "9". Lo mismo se hizo si la fecha de la muestra de sangre estaba en la posición "10".

El programa también calcula el nivel de plomo en aire 1, 2 y 3 meses antes de la toma de muestra de plomo en sangre. Se promedió el nivel diario por delegación y con este se obtuvo después el promedio uno, dos y tres meses antes de la toma de muestra de sangre.

También se corrió el programa en función de la Delegación del hospital donde nació el hijo, por el supuesto de que asisten al Centro de Salud que le corresponde a su zona, y de alguna manera con ello se pudo constatar que el nivel de plomo en sangre, si tiene que ver con lo que se respiró meses atrás, y no con una concentración de plomo en ese preciso día que se tomó la muestra de sangre, lo que vuelve a hacer patente las características de las partículas de plomo en aire: $1.5\mu\text{g}/\text{m}^3$ de suspensión en aire promedio trimestral.

Se agregaron los datos de consumo de gasolinas en la Zona Metropolitana de la Ciudad de México (ZMCM) Contenido de plomo y niveles atmosféricos de Febrero de 1986 a Septiembre de 1995. Información proporcionada por el Instituto Mexicano del Petróleo, referente al Consumo en millones de litros de gasolina "NOVA", "EXTRA" y "MAGNA SIN", toneladas de plomo emitido a la atmósfera y el promedio de plomo emitido.

El software utilizado para realizar el programa de asignación de los niveles de plomo en aire a plomo en sangre por Delegación, se hizo en el manejador de archivos DBASE IV; y STATA (Duxbury Press, Belmont, California 1993) para el análisis estadístico.

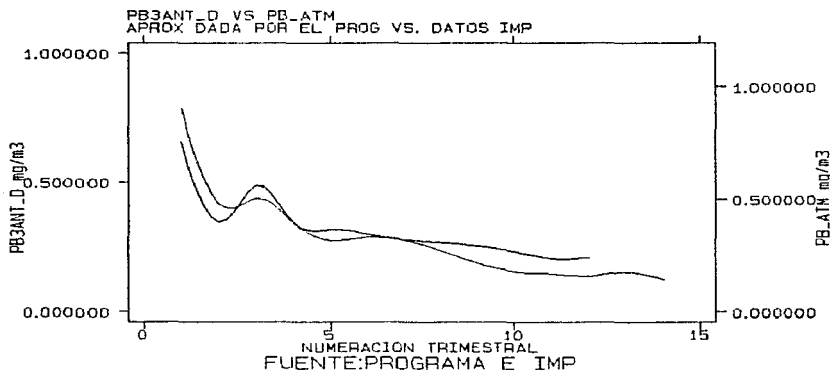
B. Análisis de variación de los datos de plomo en aire calculados por el programa versus los del IMP.

Fueron 444 observaciones las que coincidieron con la fecha en que se monitoreo plomo en aire en la Delegación donde vivia la madre y 2117 observaciones las que se estimaron mediante el programa en los días que no se monitoreó

La correlación estimada por el programa pb3ant_d versus la información que proporcionó el IMP de pb_atm fue de

Correlación	Significancia	n	β	IC 95%
0.7728	0.0000	3205	0.9966	0.9682,1.0249

El grado de correlación es muy alto: 77% por lo que se puede decir que las aproximaciones son muy buenas.



- PB3ANT_D. VARIABLE CALCULADA POR EL PROGRAMA
- PB-ATM. INFORMACION PROPORCIONADA POR IMP 1985

C. CORRELACION

1. Correlaciones entre el logaritmo del nivel de plomo en sangre de la madre (LMPB2) versus las variables que pueden explicar su variación. (Ver capítulo II)

variable	n	correlación	significancia	β
PBAIRE_D	425	0.0017*	0.9726	0.0057106
ASIGNA_D	2043	0.1150	0.0000	0.3133438
ESTIMA_D	2468	0.0929	0.0000	0.2702318
PBIANT_D	3220	0.1399	0.0000	0.4960124
PB2ANT_D	3429	0.1563	0.0000	0.5652065
PB3ANT_D	3581	0.1559	0.0000	0.568669
EDAD	3344	0.0652	0.0002	0.0069675
ESCO	3352	-0.0010*	0.9516	-0.0001992
SI LOZA	3372	0.0984	0.0000	0.1330185
NOVA	3661	0.1753	0.0000	0.0021159
MAGNA	3661	-1.1704	0.0000	-0.0018777
TOTAL	3661	-0.0231*	0.1627	-0.0005603
TON_PB	3088	0.1450	0.0000	0.0041187
PB_ATM	3087	0.1655	0.0000	0.8188848
AÑO	3661	-0.2077	0.0000	-0.1070074
TRIMCONT	3661	-0.1655	0.0000	-0.0225676

* No significativo

Cabe hacer mención que las variables en la delegación del hospital donde nació el hijo, que fue el día de la toma de muestra resultaron ser menos significativas que las de la Delegación en donde vive la madre; además, la variable se hace más robusta con el lag de tiempo en 3 meses antes de la toma de muestra PB3ANT_D.

2. Correlaciones entre el logaritmo del nivel de plomo en aire de tres meses antes por la Delegación donde vive la madre (LPB3ANT_D) versus las variables que pueden explicar su variación.

variable	n	correlación	significancia	β
NOVA	3751	+0.7912	0.0000	+0.0089431
MAGNA	3751	-0.8793	0.0000	-0.0091276
TOTAL	3751	-0.3325	0.0000	-0.0075189
TON_PB	3206	+0.7022	0.0000	+0.0163671
AÑO	3751	-0.8611	0.0000	-0.4184534

El logaritmo en el nivel de plomo en aire en tres meses antes está explicado por el consumo de las gasolinas NOVA Y MAGNA, es importante notar que el consumo de la gasolina Nova genera un incremento en los niveles de plomo en aire de un 79%; mientras que el consumo en millones de litros de la gasolina Magna Sin lo reduce en un 87% , es protector. La variable AÑO dice que el nivel de plomo en aire se ha ido reduciendo al pasar de los años, consecuencia del aumento del consumo de gasolina Magna y reducción de la Nova.

D. REGRESION LINEAL

El análisis bivariado se hizo en base a las correlaciones que fueron significativas.

Regresión del logaritmo en el nivel de plomo en sangre de la madre versus el nivel de plomo en aire 3 meses antes en la delegación de la madre:

. regress lmpb2 pb3ant_d

Source	SS	df	MS		Number of obs =	3581
Model	31.3154848	1	31.3154848		F(1, 3579) =	89.10
Residual	1267.95495	3579	.351482243		Prob > F =	0.0000
					R-squared =	0.0243
					Adj R-squared =	0.0240
Total	1299.27043	3580	.360131406		Root MSE =	.59286

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lmpb2						
pb3ant_d	.568669	.0602465	9.439	0.000	.4505481	.6867898
_cons	1.976837	.02139	92.419	0.000	1.934899	2.018775

La ecuación predictiva de éste modelo es: $\text{Log}(\text{mpb2}) = 1.97 + 0.568 * \text{pb3ant_d}$, es decir, por cada $1\mu\text{g}/\text{m}^3$ de incremento en los niveles de plomo en aire de tres meses antes, aumenta $0.56\mu\text{g}/\text{dl}$ el logaritmo del nivel de plomo en sangre de la madre. ($p < 0.0000$, IC 95% [0.45, 0.68]) (Véase sección de resultados)

Respecto a los hábitos de la madre de guardar y/o preparar alimentos en trastos de barro vidriado, se regreso el logaritmo del nivel de plomo en la sangre de la madre versus el uso de loza vidriada:

. regress lmpb2 bv1

Source	SS	df	MS			
Model	11.2553617	1	11.2553617	Number of obs	=	3372
Residual	1151.3624490	3370	.341651928	F(1, 3370)	=	32.94
Total	1162.62230	3371	.344889457	Prob > F	=	0.0000
				R-squared	=	0.0097
				Adj R-squared	=	0.0094
				Root MSE	=	.58451

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lmpb2					
bv1	.1187702	.0026928	5.740	0.000	.0781984 .1593410
_cons	2.121103	.0018288	165.342	0.000	2.09595 .1146.56

La ecuación predictiva de este modelo: $\text{Log}(\text{mpb2}) = 2.12 + 0.1187 * \text{bv1}$, lo que quiere decir que si la madre si utiliza trastos de barro vidriado se incrementará $0.1187 \mu\text{g}/\text{dl}$ el logaritmo del nivel de plomo en sangre de ésta; es decir, se incrementa $1.12 \mu\text{g}/\text{dl}$ su nivel de plomo en sangre comparado con aquellas mamás que no usan loza vidriada.

Se regresó el logaritmo del nivel de plomo en cordón umbilical versus el logaritmo del nivel de plomo en sangre de la madre:

. regress lcpb2 lmpb2

Source	SS	df	MS			
Model	787.788524	1	787.788524	Number of obs	=	3360
Residual	609.257273	3357	.180950185	F(1, 3367)	=	4353.62
Total	1397.04578	3358	.414800415	Prob > F	=	0.0000
				R-squared	=	0.5639
				Adj R-squared	=	0.5638
				Root MSE	=	.42538

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lcpb2					
lmpb2	.8036350	.1211705	65.982	0.000	.779750 .8275154
_cons	1.0790793	.1177062	10.971	0.000	.2456278 .1355324

Se tiene que el 56% de la variación en los niveles de plomo en sangre del cordón umbilical estan explicados por la variación de los niveles de plomo en sangre de la madre. La ecuación predictor es: $\text{log}(\text{CPB2}) = 0.299 + 0.8036 * \text{log}(\text{MPB2})$, lo que quiere decir que por cada $\mu\text{g}/\text{dl}$ de aumento en el nivel de plomo en sangre de la madre aumenta $0.80 \mu\text{g}/\text{dl}$ el logaritmo del nivel de plomo en sangre del cordón umbilical. ($p < 0.000$, IC 95% [0.779, 0.827])

CAPITULO VI

RESULTADOS

La mediana de los niveles de plomo en sangre de la madre bajaron de 10.5µg/dl en 1992 a 7.45 µg/dl en 1995 (Tabla 1). Se encontró una fuerte asociación de los niveles de plomo en sangre con los niveles de plomo en aire de tres meses antes de la toma de muestra de sangre ($r=0.1559$, $p<0.0000$) Tabla 2.

Tabla 1. Mediana de los niveles de plomo en sangre de las mujeres embarazadas de la Ciudad de México por año .

AÑO	# REGISTROS	EDAD MEDIA	MEDIANA MPB2
1992	1839	25	10.5
1993	458	25	7.5
1994	790	24	7.8
1995	574	25	7.45

Tabla 2. Grado de correlación, nivel de significancia, β e intervalos de confianza del análisis bivariado de la transformación logarítmica de los niveles de plomo en sangre de la madre versus parámetros que influyen en el nivel de éste.

PARAMETRO	CORRELACION	p	β	IC 95%
PB3ANT_D	0.1559	0.0000	0.568669	.4505481, .6867898
ESCO	-0.0010	0.9516*	-0.001992	-.0066331, .0062348
USA LOZA	0.0984	0.0000	0.1187702	.0781984, .1593419
NOVA	0.1753	0.0000	0.0021159	.0017307, .0025011
MAGNA	-0.1704	0.0000	-0.0018777	-.0022296,-.0015259
TON_PB	0.1450	0.0000	0.0041187	.0031266, .0051107
PB_ATM	0.1655	0.0000	0.8188848	.6465876, .991182
AÑO	-0.2077	0.0000	-0.1070074	-.1233459,-.0906689

* No Significativo

Se encontró que sin importar la categoría de edad a que pertenece la madre la mediana de su nivel de plomo en sangre disminuyó $3\mu\text{g/dl}$ de 1992 a 1995 (figura 1, Tabla 3) y a partir de 1993 se encuentra por debajo del nivel que recomienda la norma ($10\mu\text{g/dl}$).

Figura 1.

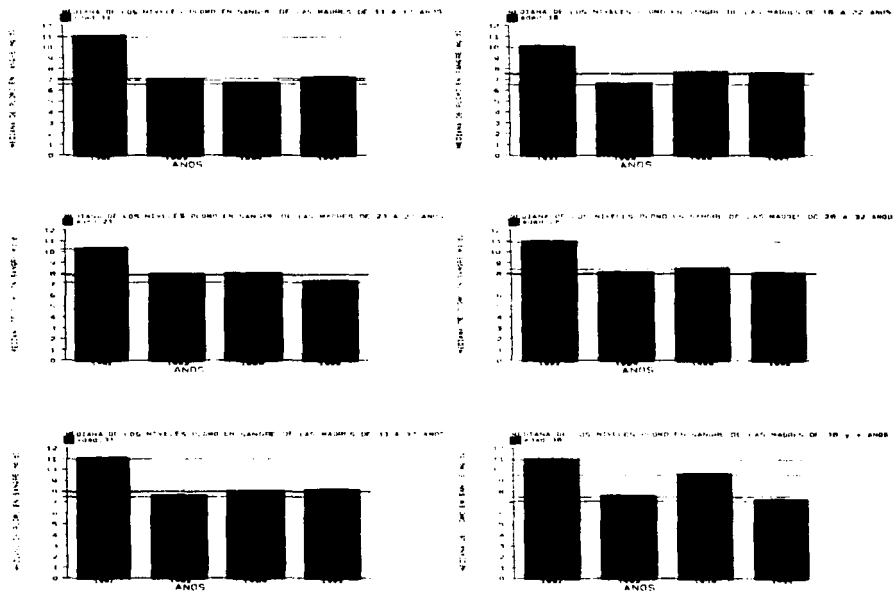
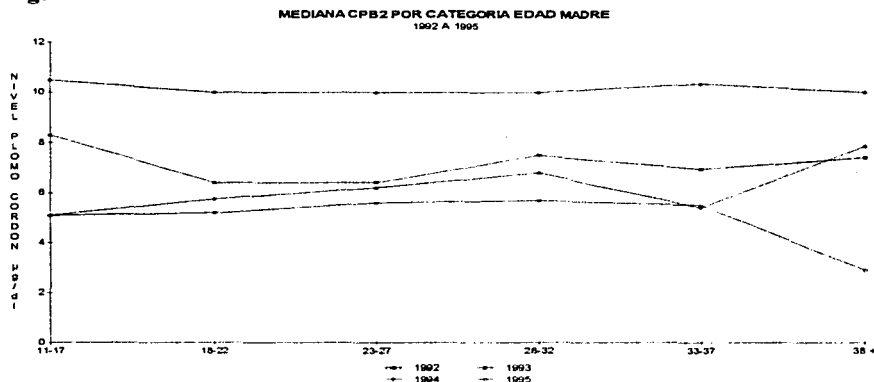


Tabla 3. Mediana de los niveles de plomo en sangre madre por grupos de edad

AÑO	GRUPOS DE EDAD, µg/dl, (n)					
	11-17	18-22	23-27	28-32	33-37	38 Y +
1992	10.95(108)	10.0 (547)	10.2 (511)	10.9 (381)	11.0 (159)	10.9 (133)
1993	6.95 (8)	6.55 (78)	7.85 (64)	8.0 (49)	7.5 (14)	7.5 (245)
1994	6.60 (51)	7.6 (292)	7.9 (263)	8.35 (124)	7.9 (48)	9.5 (12)
1995	7.10 (19)	7.5 (191)	7.2 (217)	7.95 (104)	8.0 (35)	7.10 (8)

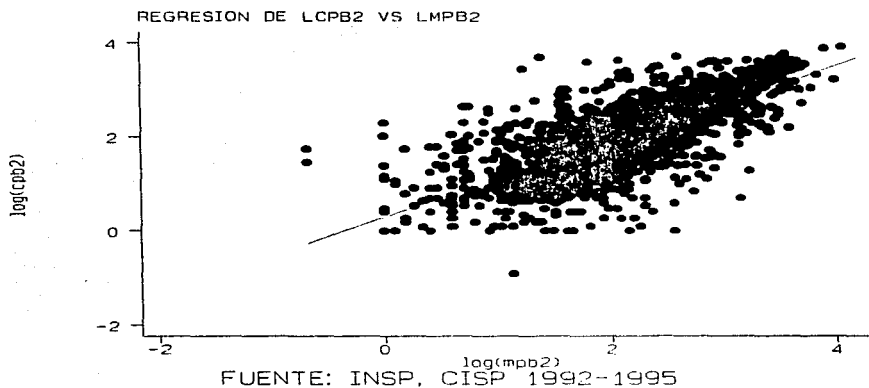
En el nivel de plomo en cordón umbilical se observó la misma tendencia (figura 2) y en promedio el nivel de plomo en sangre del cordón bajo de 10.0µg/dl en 1992 a 5.1µg/dl en 1995; con excepción del grupo de edad de las madres de 38 y más años cuyo mínimo en 1995 fue de 2.9µg/dl.

Figura 2.



El grado de correlación entre el nivel de plomo en sangre del cordón umbilical y el nivel de plomo en sangre de la madre es muy alto (figura 3), hecho que tiene una explicación biológica que es: la sangre de la madre por medio de cordón umbilical nutre y forma a su hijo. (correlación=0.7509, $p < 0.0000$, $\beta = 0.8036352$, IC 95% [0.77975, 0.8275154]). En la regresión bivariada $LCPB2 = 0.2990759 + 0.8036352 * LMPB2$; es decir, por cada 1.0 µg/dl de aumento en el logaritmo del nivel de plomo en la sangre de la madre aumenta 0.8µg/dl el logaritmo del nivel de plomo en sangre del cordón umbilical, dicho resultado es muy alto y es el nivel de plomo que va a pasar a la sangre del producto y por ende le causará daños a su salud.

Figura 3.



Se encontró que el nivel de plomo en sangre de la madre es explicado, en parte, por el nivel de plomo en aire de tres meses antes de la toma de muestra sanguínea por la delegación donde vive y no por la concentración de plomo que había el día que se tomó dicha muestra.(figura 4).

$$\text{LMPB2} = 1.976837 + 0.568669 * \text{PB3ANT_D}$$

La ecuación tiene una intersección positiva de 1.98. La pendiente de 0.56 indica que por cada $1\mu\text{g}/\text{m}^3$ de aumento en el nivel de plomo en aire, el logaritmo del nivel de plomo en sangre se aumenta $0.56\mu\text{g}/\text{dl}$ (figura 4).

figura 4. Regresión lineal del nivel de plomo en sangre de la madre versus el nivel de plomo en aire 3 meses antes de la toma de muestra con bandas de confianza al 95% para los valores individuales de las predicciones.

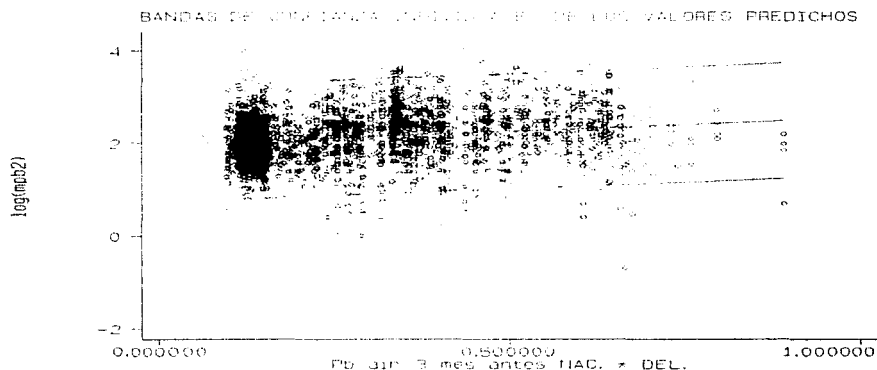
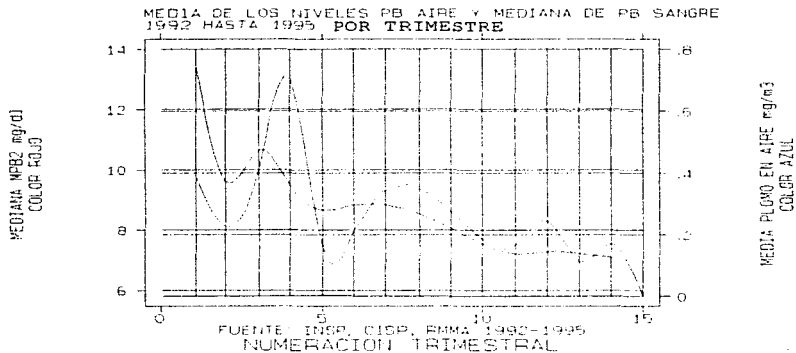


Figura 5. Media de los niveles de plomo en aire y mediana de los niveles de plomo en sangre por trimestre, 1992 a 1995.



La figura 5 permite fácilmente responder a la primera hipótesis de la presente tesis, en efecto los niveles de plomo en aire han disminuido al paso de los años, debido en parte al aumento en el consumo de millones de litros de gasolina Magna sin (Tabla 4), y este decremento en el nivel de plomo en aire se ve reflejado en el nivel de plomo en sangre.

TABLA 4. Media de los niveles de plomo en aire, las toneladas de Pb emitidas a la atmósfera y el consumo en millones de litros de las gasolinas Nova y Magna sin.

AÑO	PB EN AIRE	NOVA	MAGNA	TON_PB
1992	0.42924	417.689	109.909	56.545
1993	0.28444	374.074	149.537	21.331
1994	0.14811	336.070	220.680	19.392
1995	0.13018	296.282	229.074	19.770

Para encontrar las zonas de mayor riesgo para tener elevados niveles de plomo en sangre se obtuvo la siguiente tabla comparativa. (Tabla 5).

TABLA 5. Comparación de niveles de plomo en sangre de la madre (MPB2) y plomo en sangre de cordón umbilical (CPB2) de acuerdo a la zona donde viven las madres.

ZONAS	CONCENTRACIONES DE MPB2 ($\mu\text{g}/\text{dl}$; sd,n)	DE PLOMO EN SANGRE CPB2($\mu\text{g}/\text{dl}$;sd,n)
SE	9.00 \pm 6.93 (1159)	8.40 \pm 6.59 (1131)
SO	8.60 \pm 6.12 (1781)	7.50 \pm 6.17 (1626)
C	8.45 \pm 5.97 (472)	8.25 \pm 6.01 (454)
NO	6.80 \pm 5.18 (15)	5.20 \pm 2.10 (11)
NE	8.90 \pm 5.26 (234)	8.30 \pm 4.99 (226)

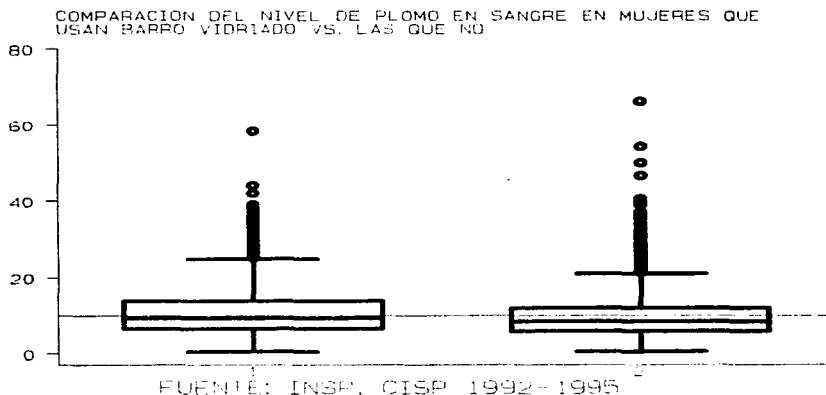
La tabla 5 muestra las medianas de las concentraciones de plomo en sangre de las madres de este estudio así como en el cordón umbilical y se aprecia que para las diferentes zonas de la Ciudad de México las concentraciones son muy similares, a excepción de la zona Noroeste en donde se reclutaron sólo a 15 mujeres y es muy probable que ésta sea la razón por la que difiere de las otras zonas.

TABLA 6. Comparaciones de hábitos de la madre en uso de loza vidriada y concentraciones en niveles de plomo en sangre.

USO BARRO VIDRIADO	NIVEL BAJO <math><10\mu\text{g}/\text{dl}</math>	NIVEL ALTO >=10μg/dl
SI	6.60 (706)	16.42 (590)
NO	6.31 (1268)	15.35 (808)

La tabla 6 muestra que tomando como punto de corte la norma actual para los niveles de plomo en sangre, 10μg/dl, las mujeres que utilizan barro vidriado para guardar y/o preparar alimentos tienen concentraciones de plomo en sangre mayores a los que tienen las mujeres que no los utilizan. Estas diferencias se dieron por abajo y por arriba del punto de corte.

En la figura 6 se observan visualmente estas diferencias. En el eje X el número 1 denota a las mujeres que usan barro vidriado y 2 las que no. El eje Y representa las concentraciones de plomo en sangre de las mujeres en μg/dl.



Análisis multivariado

Análisis multivariado de el nivel de plomo en sangre de la madre:

. stepwise lmpb2 bv1 esta421 esta422 esta423 edad magna pb3ant_d zona1
zona2 zona4 zona5, forward stepwise fstay(3.92) fenter(3.92)

```
Adding: esta423 F= 143
Adding: magna F= 111.7
Adding: bv1 F= 49.86
Adding: esta421 F= 47.89
Adding: zona2 F= 18.02
Adding: pb3ant_d F= 8.069
Adding: esta422 F= 8.495
Adding: edad F= 5.496
Adding: zona1 F= 4.724
```

Source	SS	df	MS
Model	127.96521	9	14.1739122
Residual	997.977941	3229	309067154
Total	1125.94315	3238	347604401

(stepwise)	
Number of obs	3239
F(9, 3229)	45.836
Prob > F	0.0000
R-squared	0.1131
Adj R-squared	0.1109
Root MSE	555.94

	Coeff.	Std. Err.	t	Prob> t	[95% Conf. Interval]
lmpb2					
bv1	1468284	2020155	7.263	0.000	1071918 1864649
esta421	1089066	10427337	2.527	0.012	1324186 1917948
esta422	2371516	10434375	5.460	0.000	1519838 3223104
esta423	4454774	1041666	9.973	0.000	3579005 5330543
edad	1004986	10017737	2.302	0.021	1006973 1907898
magna	1001287	1000537	-3.487	0.000	-1001635 -1000494
pb3ant_d	13794477	1235931	3.070	0.002	1130128 1617766
zona1	1627764	1028837	2.173	0.030	1001441 1154087
zona2	1659178	1030537	5.365	0.000	1167457 212655
zona4	1740348	1127361	15.443	0.000	1561306 1901389

$$\log(\text{lmpb2}) = 1.74 + 0.146 * \text{bv1} + 0.108 * \text{esta421} + 0.237 * \text{esta422} + 0.445 * \text{esta423} \\ + 0.004 * \text{edad} - 0.0011 * \text{magna} + 0.379 * \text{pb3ant_d} + 0.0627 * \text{zona1} + 0.165 * \text{zona2}$$

El logaritmo del nivel de plomo en sangre de la madre está explicado el 11% de su variación por el uso de barro vidriado para guardar y/o preparar alimentos y por las covariables (*) y variables confusoras (*) como son: las estaciones de primavera, verano y otoño, la edad de la madre, el consumo en millones de litros de gasolina magna sin y las zonas SE y SO.

Análisis multivariado de el nivel de plomo en sangre del cordón umbilical:

```
. stepwise lcpb2 mpb2 pb3ant_d nova bv1 zona2 zona4 zona5 esta421 esta422
> esta423,forward stepwise Zstat(3.92) fenter(3.92)
```

```
Adding: mpb2 F= 3327
Adding: nova F= 213
Adding: pb3ant_d F= 17.06
Adding: esta423 F= 12.97
Adding: zona2 F= 15.49
Adding: esta422 F= 12.9
Adding: bv1 F= 7.348
Adding: esta421 F= 4.655
```

Source	SS	df	MS	Number of obs =	(stepwise)
Model	696.216548	8	87.0270685	FI B, 3007) =	490.69
Residual	533.308077	3007	.177355529	Prob > F =	0.0000
Total	1229.52462	3015	.407802529	R-squared =	0.5662
				Adj R-squared =	0.5651
				Root MSE =	.42114

lcpb2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mpb2	.066504	.0018893	51.620	0.000	.0639779 .0690301
pb3ant_d	.4978459	.0786748	6.328	0.000	.3435841 .6521077
nova	.0011036	.0002453	4.499	0.000	.0004227 .0015845
bv1	.0438415	.0160318	2.735	0.006	.0124071 .0752759
zona2	.0756303	.0192356	3.932	0.000	.037914 .1133496
esta421	.0738076	.0342081	2.158	0.031	.0067339 .1408813
esta422	.1344001	.0368055	3.754	0.000	.0641943 .2046059
esta423	.1803366	.0369488	4.897	0.000	.1081911 .2533861
_cons	.6043187	.0752997	8.026	0.000	.4566746 .7519628

$$\log(\text{CPB2}) = 0.604 + 0.066 * \text{MPB2} + 0.497 * \text{PB3ANT_D} + 0.0011 * \text{NOVA} + 0.0438 * \text{BV1} + 0.0756 * \text{ZONA2} + 0.0738 * \text{ESTA421} + 0.1344 * \text{ESTA422} + 0.1809 * \text{ESTA423}$$

La variación en el logaritmo del nivel de plomo en sangre del cordón umbilical está explicada en un 56% por la concentración de plomo en el torrente sanguíneo de la madre y por las covariables que pueden ser modificadoras de efecto como son: el nivel de plomo en aire tres meses antes del nacimiento, por el consumo de gasolina nova, la Zona SO y las estaciones de primavera, verano y otoño.

Análisis multivariado de el nivel de plomo en aire de 3 meses antes de la toma de muestra:

. stepwise pb3ant_d magna esta421 esta422 esta423 zonal zona2 zona4 zona5,backward
stepwise fstay(3.92) fentex(3.92)

Source	SS	df	MS	Number of obs	(stepwise)
Model	78.3776697	8	9.79720871	F(8, 3742)	= 1579.19
Residual	23.2181141	3742	.006203932	Prob > F	= 0.0000
				R-squared	= 0.7515
				Adj R-squared	= 0.7419
Total	101.592784	3750	.027091409	Root MSE	= .07877

pb3ant_d	Coef.	Std. Err.	t	P> t	195% Conf. Interval	Lower	Upper
magna	-.0000392	.0000263	-17.627	0.000	-.0000908	-.0001977	.0001183
esta421	.0101041	.0047318	2.134	0.033	.0008111	-.0193871	.0398953
esta422	.0469453	.0047865	9.787	0.000	.0374608	.0564298	.0564298
esta423	.0261955	.0056045	4.678	0.000	.0163245	.0360665	.0360665
zonal	-.0991141	.0055604	-17.825	0.000	-.1100167	-.0882125	-.0882125
zona2	-.1974967	.0055266	-35.740	0.000	-.2083308	-.1866625	-.1866625
zona4	-.1206238	.0117858	-10.237	0.000	-.1633571	-.0778905	-.0778905
zona5	-.1100893	.0061711	-17.840	0.000	-.1221881	-.0979904	-.0979904
_cons	.7486477	.0072204	103.688	0.000	.7344915	.7628039	.7628039

pb3ant_d = 0.748-0.002*magna+0.010*esta421+0.0468*esta422+0.026*esta423 - 0.099*zonal - 0.197*zona2 -0.1206*zona4 - 0.11*zona5

La variación en el nivel de plomo en aire de tres meses antes de la toma de muestra está explicada por el consumo de gasolina magna sin, las estaciones del año: primavera, verano, otoño y las zonas: SE (Zona 1), SO(Zona 2), NO(Zona 4) y C (Zona 5) en un 77%.

GLOSARIO.

Bandas de confianza.

Líneas punteadas a cada lado de la línea o curva de regresión con una probabilidad dada de incluir la línea o curva de la población.

Bioacumulación.

Partes orgánicas o partes vivas que se están acumulando en un área específica dentro de los organismos vivos y pueden intoxicarlo o retardar su desarrollo.

Biomarcadores.

Muestras biológicas que miden la exposición de cada individuo a un contaminante y sirven para evitar una mala clasificación.

Cadena trófica.

Cadena completa.

Colorimétrico.

Técnica que mediante aplicación de reactivos se obtiene un cambio de coloración que es detectado en un espectrofotómetro obteniendo la concentración del contaminante en estudio.

Confusora.

Información que se debe incluir en el modelo para tener una estimación correcta -no sesgada- de la relación entre la variable dependiente y la independiente. Su inclusión es determinante de la validez de los resultados.

Covariable.

Información que se incluye en el modelo por que es importante para comprender mejor la relación entre la variable independiente y la dependiente. Reduce la variabilidad no explicada de la variable dependiente. Mejora la precisión de la medición de la variable independiente.

Deficiencia en las delta amino levulinico dehidrasa en la síntesis del grupo hemo.

Cuando hay deficiencia de esta enzima, no hay suficiente retención de oxígeno en los glóbulos rojos.

Espectrofotómetro.

Aparato con el que se determinan cuantitativamente las intensidades relativas de las luces coloreadas que forman parte de un espectro.

Falacia ecológica.

Una asociación que se encuentra a nivel poblacional puede no existir cuando los individuos son estudiados. Las asociaciones que se encuentran a nivel de población pueden ser directas, mientras que aquellas que se establecen con base a las características de los individuos pueden ser indirectas.

Homeostático.

Cuando todo está balanceado y regulado.

Línea de plomo.

Es de color gris o negro azulado de pimentación por sulfuro, aparece en la encía, pero es algo más difusa que la del Bismuto. También se encuentra a veces en otras áreas de la cavidad bucal. También son molestias comunes la salivación excesiva y el sabor metálico, así como la hinchazón de las glándulas salivales.

Atshuller y colaboradores informaron que el plomo se deposita en los dientes deciduos de los niños que sufrieron de envenenamiento por plomo, y que éstos dientes pueden servir como un índice de la cantidad de plomo que se encuentra en el cuerpo.

Neuropatía periférica.

Una alteración o enfermedad en los nervios periféricos.

Norma Federal Mexicana que regulan el uso de plomo y que son publicadas en el diario Oficial de la Federación. 18 de Enero de 1993.

Aire.

NOM-0026-SSA1-1993. Criterios para evaluar la calidad del aire, plomo(Pb). El valor permisible para la concentración de plomo en el aire ambiente, como medida de protección a la salud de la población.

Se estableció el valor de 1.5 $\mu\text{g}/\text{m}^3$ en un período de tres meses.

Envases metálicos.

NOM-002-SSA1-1993. Salud ambiental-Bienes y servicios. Envases metálicos para alimentos y bebidas. Especificaciones de la costura. Requisitos sanitarios.

Prohíbe el uso de soldadura que contiene plomo para envases de alimentos y bebidas, los fabricantes deben utilizar soldadura eléctrica, costura con pegamento o cementada.

Prohíbe la importación y comercialización de alimentos y bebidas cerrados con soldadura estaño-plomo.

Alfarería y cerámica.

NOM-011-SSA1-1993. Salud ambiental. Límites de plomo y cadmio solubles en artículos de alfarería vidriados.

Los métodos tradicionales de vidriado a base de plomo, usados en la fabricación de loza de barro, serán sustituidos gradualmente por procesos que no requieren plomo o que evite su desprendimiento.

Evite cocinar o almacenar alimentos y bebidas ácidas en trastes o cazuelas de barro vidriado.

Puede utilizar estos artículos en ornato o decoración.

Los artículos de barro vidriado contendrán la leyenda:
"No usar para alimentos o bebidas"

Quimoluminescencia.

Propiedad de despedir luz sin elevación de temperatura y visible sólo casi en la oscuridad. "Cuando se expone un trozo de fósforo a la oxidación por el aire libre".

Residual.

Diferencia entre el valor probable (predicción) y el valor real de la variable dependiente (resultado o respuesta) en regresión.

Robusto.

Término usado para describir en método estadístico a cuyo valor no le afectan en forma importante, la violación de las premisas de la prueba.

Saturnismo.

Intoxicación provocada por sales de plomo.

Sistema nervioso central.

Lo constituye el cerebro y la médula ósea. Todos los nervios que derivan sustituye el sistema nervioso periférico.

Apéndice A: Programa

Programa de asignación de plomo en aire a plomo en sangre

TESIS.PRG

El módulo llamado TESIS.PRG es el programa principal cuya función es determinar para cada registro si tiene una delegación o municipio, que está representado por un número ya antes mencionado en el capítulo 5 de la presente tesis, una vez determinada la delegación busca si tiene fecha de toma de muestra sanguínea. Si el programa no detecta la Delegación y/o la fecha de muestra se pasa al siguiente registro.

En base a la fecha de toma de muestra el programa procede a determinar las fechas correspondientes a uno, dos y tres meses antes de dicha muestra, mismas que son almacenadas temporalmente en variables de memoria.

Por último el programa principal da el mando a el subprograma que corresponde a la delegación que está procesando y que hará los cálculos correspondientes para las fechas obtenidas y a las estaciones de monitoreo que sean las más cercanas y/o representativas a dicha delegación.

```

*****
*                               T E S I S . P R G                               *
* PROGRAMA DE ASIGNACION DE LOS NIVELES DE PLOMO EN AIRE A                       *
* SANGRE DEPENDIENDO DE LA FECHA DE LA TOMA DE MUESTRA.                       *
*****

```

```

SET TALK OFF
SELECT 1
USE PRACT.DBF
SELECT 2
USE AIRE.DBF

```

```

SELECT 1

```

```

GO TOP

```

```

DO WHILE .NOT. EOF()

```

```

DO WHILE ASIGN_ESTA=0

```

```

  SKIP
  IF ASIGN_ESTA<>0
    EXIT
  ELSE
    ENDIF

```

```

ENDDO

```

```

  STORE ASIGN_ESTA TO DELE
  STORE FECHAHUESO TO HUESO
  STORE FECHAHUESO TO DATE1
  STORE FECHAHUESO TO DATE2

```

```

IF FECHAHUESO={00/00/00}

```

```

  SKIP
  ELSE

```

```

DO CASE

```

```

  CASE MONTH(FECHAHUESO)=1.OR.MONTH(FECHAHUESO)=3.OR.
    MONTH(FECHAHUESO)=5.OR.MONTH(FECHAHUESO)=7.OR.
    MONTH(FECHAHUESO)=8.OR.MONTH(FECHAHUESO)=10.OR.
    MONTH(FECHAHUESO)=12
    STORE (FECHAHUESO-31) TO HUESO1
  CASE MONTH(FECHAHUESO)=2
    STORE YEAR(FECHAHUESO) TO ANO

```

```

    STORE INT(ANO/4) TO COC1
    STORE MOD(ANO,4) TO RES1
  IF RES1=0
    STORE (FECHAHUESO-29) TO HUESO1
  ELSE
    STORE (FECHAHUESO-28) TO HUESO1
  ENDIF
  CASE MONTH(FECHAHUESO)=4.OR.MONTH(FECHAHUESO)=6.OR.
    MONTH(FECHAHUESO)=9.OR.MONTH(FECHAHUESO)=11
    STORE (FECHAHUESO-30) TO HUESO1
  ENDCASE
  DO CASE
    CASE MONTH(HUESO1)=1.OR.MONTH(HUESO1)=3.OR.
      MONTH(HUESO1)=5.OR.MONTH(HUESO1)=7.OR.
      MONTH(HUESO1)=8.OR.MONTH(HUESO1)=10.OR.
      MONTH(HUESO1)=12
      STORE (HUESO1-31) TO HUESO2
    CASE MONTH(HUESO1)=2
      STORE YEAR(HUESO1) TO ANO2
      STORE INT(ANO2/4) TO COC2
      STORE MOD(ANO2,4) TO RES2
      IF RES2=0
        STORE (HUESO1-29) TO HUESO2
      ELSE
        STORE (HUESO1-28) TO HUESO2
      ENDIF
    CASE MONTH(HUESO1)=4.OR.MONTH(HUESO1)=6.OR.
      MONTH(HUESO1)=9.OR.MONTH(HUESO1)=11
      STORE (HUESO1-30) TO HUESO2
    ENDCASE
  DO CASE
    CASE MONTH(HUESO2)=1.OR.MONTH(HUESO2)=3.OR.
      MONTH(HUESO2)=5.OR.MONTH(HUESO2)=7.OR.
      MONTH(HUESO2)=8.OR.MONTH(HUESO2)=10.OR.
      MONTH(HUESO2)=12
      STORE (HUESO2-31) TO HUESO3
    CASE MONTH(HUESO2)=2
      STORE YEAR(HUESO2) TO ANO3
      STORE INT(ANO3/4) TO COC3
      STORE MOD(ANO3,4) TO RES3
      IF RES3=0

```

```
        STORE (HUESO2-29) TO HUESO3
    ELSE
        STORE (HUESO2-28) TO HUESO3
    ENDIF
    CASE MONTH(HUESO2)=4.OR.MONTH(HUESO2)=6.OR.
        MONTH(HUESO2)=9.OR.MONTH(HUESO2)=11
        STORE (HUESO2-30) TO HUESO3
    ENDCASE
```

```
STORE RECNO() TO NOREG
```

DO CASE

***** DELEGACIONES *****

```
    CASE DELE=1
        DO OBREGON
    CASE DELE=2
        DO AZCAPOT
    CASE DELE=3
        DO JUAREZ
    CASE DELE=4
        DO COYOACAN
    CASE DELE=5
        DO CUAJIMAL
    CASE DELE=6
        DO CUAUHTEM
    CASE DELE=7
        DO MADERO
    CASE DELE=8
        DO IZTACAL
    CASE DELE=9
        DO IZTAPAL
    CASE DELE=10
        DO CONTRERA
    CASE DELE=11
        DO HIDALGO
    CASE DELE=12
        DO MILPALT
    CASE DELE=13
        DO TLAHUAC
    CASE DELE=14
        DO TLALPAN
    CASE DELE=15
```



```
DO VENUS
CASE DELE=16
DO XOCHIMIL
**** MUNICIPIOS ****
CASE DELE=21
DO ATIZAPAN
CASE DELE=23
DO CHALCO
CASE DELE=24
DO CHICO
CASE DELE=25
DO CHIMAL
CASE DELE=26
DO COACALCO
CASE DELE=28
DO IZCALLI
CASE DELE=30
DO ECATEPEC
CASE DELE=31
DO HUIXQUI
CASE DELE=32
DO NAUCALPA
CASE DELE=33
DO NEZAHUA
CASE DELE=34
DO TEXCOCO
CASE DELE=35
DO TLALNE
CASE DELE=36
DO TOLUCA
CASE DELE=40
DO ROMERO
CASE DELE=41
DO TULTEPEC
OTHERWISE
SKIP
ENDCASE
SELECT 1
GO NOREG
SKIP
ENDIF
ENDDO
```

```
*****
*   ASIGNACION DE PLOMO EN AIRE PARA LA DELEGACION:   *
*   1. ALVARO OBREGON                                   *
*****
```

```
SET TALK OFF
SELECT 1
USE PLACT.DBF
SELECT 2
USE AIRE.DBF
```

```
SELECT 2
GO TOP
```

```
SUM FOR
((FECHA>HUESO)).AND.((ESTACION="CFE".OR.ESTACION="PED".OR.
ESTACION="FAN").AND.(PLOMO<0)) TO OK
SUM FOR
((FECHA>=HUESO1).AND.(FECHA<=HUESO)).AND.((ESTACION="CFE".
OR.ESTACION="PED".OR.ESTACION="FAN").AND.(PLOMO<0)) TO HAY
```

```
*** ASIGNA EL PROMEDIO DE ESE DIA EN LA DELEGACION ALVARO OBREGON ***
*** LAS ESTACIONES DE MONITOREO QUE LE CORRESPONDEN SON CFE,FAN,PED ***
```

```
AVERAGE PLOMO FOR
(FECHA=HUESO).AND.((ESTACION="CFE".OR.ESTACION="PED".OR.
ESTACION="FAN").AND.(PLOMO<0)) TO DIA
IF .NOT.(DIA<0)
```

```
SET TALK ON
```

```
*** EN CASO DE NO HABER UNA MEDICION DE PLOMO EN AIRE PARA ESE DIA ***
```

```
STORE 0 TO ANTES
STORE 0 TO DESPUES
STORE 0 TO DAY1
STORE 0 TO DAY2
STORE 0 TO JUNTOS
```

```
DO WHILE .NOT.(DAY1<0)
STORE (DATE1-1) TO DATE1
AVERAGE PLOMO FOR FECHA=DATE1).AND.
(ESTACION="CFE".OR.ESTACION="PED".OR.ESTACION="FAN")
.AND.(PLOMO<0)) TO DAY1
STORE (ANTES+1) TO ANTES
```

```

IF ANTES=7
  EXIT
ELSE
  ENDF
ENDDO
DO WHILE .NOT. (DAY2<=0)
  STORE (DATE2+1) TO DATE2
  AVERAGE PLOMO FOR (FECHA=DATE2).AND.
  ((ESTACION="CFE".OR.ESTACION="PED".OR.ESTACION="FAN")
  .AND.(PLOMO<=0)) TO DAY2
  STORE (DESPUES+1) TO DESPUES
  IF DESPUES=7
    EXIT
  ELSE
    ENDF
  ENDDO

STORE 0 TO CACHE
IF ((DAY1<=0.AND.DAY2=0).OR.(DAY2<=0.AND.DAY1=0))
  STORE 1 TO CACHE
ELSE
  ENDF

IF OK=0
  STORE DAY1 TO P
ELSE
  STORE ((DAY1+DAY2)/2) TO P
  STORE ((ANTES+DESPUES)/2) TO JUNTOS
ENDF

IF CACHE<=1
  SELECT 1
  GO NOREG

```

*** CRITERIO DE ASIGNACION DE PB EN AIRE EN DONDE NO SE TENIA MEDICION ***

DO CASE

```
CASE ANTES=DESPUES
  REPLACE ASIGNADIAH WITH P
CASE ANTES=1
  REPLACE ASIGNADIAH WITH DAY1
CASE DESPUES=1
  REPLACE ASIGNADIAH WITH DAY2
CASE ((ANTES<JUNTOS).AND.(ANTES<>1))
  IF DAY2=0
    REPLACE ASIGNADIAH WITH DAY1
  ELSE
    REPLACE ASIGNADIAH WITH P
  ENDIF
CASE ((DESPUES>JUNTOS).AND.(DESPUES<>1))
  REPLACE ASIGNADIAH WITH P
```

ENDCASE

```
ELSE
ENDIF
ELSE
ENDIF
```

SET TALK ON
SELECT 2

** PROMEDIO DE 1 MES ANTES **

STORE 0 TO CALCULO1
STORE 0 TO CUENTA1
STORE 0 TO VIGILA
STORE HUESO1 TO DATE

USE TEMPO.DBF

APPEND FROM AIRE.DBF FOR
(FECHA>=HUESO1).AND.(FECHA<HUESO)).AND.((ESTACION="CFE".OR
.ESTACION="PED".OR.ESTACION="FAN").AND.(PLOMO<0))

DO WHILE (DATE<HUESO)
STORE HUESO1+VIGILA TO DATE
IF DATE=HUESO
EXIT
ELSE
AVERAGE PLOMO FOR (FECHA=DATE) TO PROM1
IF PROM1<0
STORE (CALCULO1+PROM1) TO CALCULO1
STORE CUENTA1+1 TO CUENTA1
ELSE
ENDIF
STORE VIGILA+1 TO VIGILA
ENDIF
ENDDO
GO TOP
DO WHILE .NOT. EOF()=.T.
DELETE
SKIP
ENDDO
PACK

SELECT 2

** PROMEDIO 2 MESES ANTES **

STORE 0 TO CALCULO2
STORE 0 TO CUENTA2
STORE 0 TO VIGILA2
STORE HUESO2 TO DATE

USE TEMPO.DBF

APPEND FROM AIRE.DBF FOR
(FECHA>=HUESO2).AND.(FECHA<HUESO)).AND.((ESTACION="CFE".OR
.ESTACION="PED".OR.ESTACION="FAN").AND.(PLOMO<>0))

DO WHILE (DATE<>HUESO)
STORE HUESO2+VIGILA2 TO DATE
IF DATE=HUESO
EXIT
ELSE
AVERAGE PLOMO FOR (FECHA=DATE) TO PROM2
IF PROM2<>0
STORE (CALCULO2+PROM2) TO CALCULO2
STORE CUENTA2+1 TO CUENTA2
ELSE
ENDIF
STORE VIGILA2+1 TO VIGILA2
ENDIF
ENDDO

GO TOP
DO WHILE .NOT. EOF()=T.
DELETE
SKIP
ENDDO
PACK

SELECT 2

** PROMEDIO 3 MESES ANTES **

STORE 0 TO CALCULO3
STORE 0 TO CUENTA3
STORE 0 TO VIGILA3
STORE HUESO3 TO DATE

USE TEMPO.DBF

APPEND FROM AIRE.DBF FOR
((FECHA>=HUESO3).AND.(FECHA<HUESO)).AND.((ESTACION="CFE".OR
.ESTACION="PED".OR.ESTACION="FAN").AND.(PLOMO<0))

DO WHILE (DATE<>HUESO)
STORE HUESO3+VIGILA3 TO DATE
IF DATE=HUESO
EXIT
ELSE
AVERAGE PLOMO FOR (FECHA=DATE) TO PROM3
IF PROM3<0
STORE (CALCULO3+PROM3) TO CALCULO3
STORE CUENTA3+1 TO CUENTA3
ELSE
ENDIF
STORE VIGILA3+1 TO VIGILA3
ENDIF
ENDDO

GO TOP
DO WHILE .NOT. EOF()=.T.
DELETE
SKIP
ENDDO
PACK

****** ASIGNANDO LOS NIVELES DE PLOMO EN AIRE ******

```
SELECT 1
GO NOREG
IF DIA<0
  REPLACE PBAIR_HUES WITH DIA
ELSE
  ENDIF
```

```
IF CALCULO1<0
  SELECT 1
  REPLACE PB1ANTHUES WITH (CALCULO1/CUENTA1)
ELSE
  ENDIF
```



```
IF CALCULO2<0
  SELECT 1
  REPLACE PB2ANTHUES WITH (CALCULO2/CUENTA2)
ELSE
  ENDIF
```

```
IF CALCULO3<0
  SELECT 1
  REPLACE PB3ANTHUES WITH (CALCULO3/CUENTA3)
ELSE
  ENDIF
```

```
RETURN
```


BIBLIOGRAFIA

1. DOUGLAS, ALTMAN, et. al., "Calculating confidence intervals for regression and correlation", British Medical Journal, vol. 296, 30 April 1988, pp 1238-1242.
2. CARREON T, LOPEZ L, ROMIEU I. Manual de procedimiento de muestras biológicas y ambientales para determinar niveles de plomo. ECO/OPS. Metepec, Edo. de Mex. 1995. pp 85.
3. COHEN JACOB. Statistical power Analysis for the Behavioral Sciences. Ed. Lawrence Erlbaum Associates Publishers. Hillsdale, New Jersey 1988. pp 350.
4. DAWSON-SAUNDERS, G. TRAPP. Bioestadística médica. Ed. Manual Moderno. México 1995. pp 384.
5. VALLE VEGA P. Toxicología de alimentos. ECO/OPS. Metepec, Edo. de Mex. 2a ed. 1991. pp 145.
6. GLANTZ S.A., SLINKER B.K. Primer of applied regression and analysis of variance. Mc- Graw Hill, 1990
7. HAMILTON C. LAWRENCE. Regression with graphics.
8. HERNANDEZ-AVILA MAURICIO. Bioestadística intermedia. Primer programa anual de verano en Salud Pública y Epidemiología. Ed. INSP-CISP, Cuernavaca Morelos. pp 120.
9. H. ZAR, JERROLD. Biostatistical analysis. Prentice Hall INC. Englewood 1984.
10. Información proporcionada por el Instituto Nacional de Salud Pública (INSP), en la subdivisión del Centro de Investigaciones en Salud Poblacional (CISP).

11.  Información proporcionada por el Instituto Mexicano del Petróleo (IMP 1995)
12.  Información proporcionada por la Red Manual de Monitoreo Atmosférico (RMMA)
13. BURCHARD, et. al., "The significance of particle emissions", J. Air Pollu. Control Assoc. 24 No. 12 (1974); 114
14. KENNET WARK, CECIL F. WARNER. Contaminación del aire, origen y control. Ed. Limusa. 1990. pp 650.
15. LOPEZ ACUÑA DANIEL, GONZALEZ DE LEON DEYANIRA, MORENO SANCHEZ ANA ROSA. La Salud ambiental en México. Ed. Universo veintiuno, México 1987, pp 244.
16. MILLER, D.T., PASCHAL, D.C., GUNTER, E.W., STROUD, P.E., D'ANGELO. "Determination of blood lead with electrothermal atomic absorption using a L'vov platform and matrix modifier". ANALYST, 1987; 112:1701-4.
17. MORTON LIPPMAN. "REVIEW 1989 Alice Hamilton Lecture. Lead and Human Health: Background and Recent Findings". Environmental Research 51, 1990. 1-24.
18. NAKAMURA SHOICHIRO. Métodos Numéricos aplicados con Software. Prentice Hall, México 1992. pp 570.
19. PAGANO. Principles of biostatistics. Duxbury Press. 1990.