

APLICACIÓN DE LA TEORÍA GEOESTADÍSTICA EN LA
ESTIMACIÓN DE LA TEMPERATURA MEDIA ANUAL EN LA
REGIÓN SURESTE DEL ESTADO DE COAHUILA, MEXICO

FRANCISCO JAVIER SÁNCHEZ ASPEYTIA

T E S I S

Presentada como requisito parcial para
obtener el grado de:

MAESTRO EN CIENCIAS
EN ESTADÍSTICA EXPERIMENTAL



Universidad Autónoma Agraria

“Antonio Narro”

Buenvista, Saltillo, Coahuila, México.

Marzo de 2011

UNIVERSIDAD AUTONOMA AGRARIA

“ANTONIO NARRO “

SUBDIRECCION DE POSTGRADO

APLICACIÓN DE LA TEORÍA GEOESTADISTICA EN LA ESTIMACIÓN DE LA
TEMPERATURA MEDIA ANUAL EN LA REGIÓN SURESTE DEL ESTADO DE
COAHUILA, MÉXICO

TESIS

Por:

FRANCISCO JAVIER SÁNCHEZ ASPEYTIA

Elaborada bajo la supervisión del comité particular de asesoría y aprobada como requisito parcial
para optar al grado de:

MAESTRO EN CIENCIAS EN ESTADÍSTICA EXPERIMENTAL

COMITÉ PARTICULAR

Asesor principal:




Dr. Javier de Jesús Cortés Bracho

Asesor:




Dr. Fernando Esquivel Bocanegra

Asesor:



M.C. Félix de Jesús Sánchez Pérez



Dr. Fernando Ruiz Zarate
Subdirector de Postgrado

Buenvista, Saltillo, Coahuila, Marzo de 2011

DEDICATORIA

A Dios

Por estar conmigo, guiarme y protegerme.

A mi Padre

Francisco Sánchez Pérez (†), quien me enseñó que con dedicación, trabajo y metas fijas se llega muy lejos en la vida.

A mi Madre

Victoria Aspeytia Herrera, que con su cariño, me ha sabido guiar en los momentos difíciles y que con sus innumerables sacrificios y esfuerzos, ha sabido hacer de mí un hombre de bien.

Por su confianza y amor les doy las gracias por todo el amor y apoyo que siempre me han brindado.

A mis Hermanos

David, Alma Delia, Leticia y Juan Pablo, por darme todo su amor y cariño que me han brindado hasta el día de hoy, además por ser parte importante en mi vida y ser mi motivo de superación.

A mis Abuelos

Agustín, Natividad, Macario y María del Carmen (†), por todo su amor, cariño y comprensión que me brindaron durante toda mi vida y que de una manera u otra contribuyen en mi formación.

A mis Tíos

Por todos los consejos y motivación que me han dado durante mi vida.

En especial a mis tíos Félix y Mary quienes me brindaron su apoyo incondicional y me dieron la oportunidad de formarme como profesionalista

A mis Primos

Por todas las experiencias alegres que en su momento llegamos a compartir, por todos los consejos y motivación que me dieron para lograr mis metas propuestas.

A mis Sobrinos

Por el amor, cariño y respeto que les profeso.

A todas aquellas personas que me han dado su cariño, amistad y apoyo

AGRADECIMIENTOS

Primeramente a Dios por darme la vida y terminar mis estudios, gracias por estar conmigo en todos los momentos de mi vida, además por darme la alegría de compartir a mi familia y amigos.

A mi “Alma Terra Mater” que siempre llevare su nombre en alto donde sea que me encuentre, por haberme brindado la oportunidad de formarme como profesionista al cumplir satisfactoriamente mis estudios en esta universidad.

A mis profesores de la Maestría en estadística experimental, por transmitirme sus conocimientos y experiencia.

A los integrantes de mi comité asesor:

Dr. Javier de Jesús Cortés Bracho, por su excelente dirección para la realización del presente trabajo.

Dr. Fernando Esquivel Bocanegra y M.C Félix de Jesús Sánchez Pérez, por sus valiosas observaciones, sugerencias y por todo el apoyo que me brindaron para concluir de manera satisfactoria este trabajo.

Al personal del Departamento de Estadística y Calculo de la U.A.A.A.N., por su extraordinaria amistad, aprecio y apoyo que me brindaron durante esta etapa de mi vida.

A mis amigos, compañeros y a todas las personas que de una u otra forma siempre me han brindado su apoyo incondicional.

ÍNDICE DE CONTENIDO

	Página
ÍNDICE DE CUADROS	ix
ÍNDICE DE FIGURAS.....	x
RESUMEN	xi
ABSTRACT.....	xiv
1.- INTRODUCCIÓN.....	1
Objetivos.....	5
2.- REVISIÓN DE LITERATURA.....	6
Reseña Histórica de la Geoestadística.....	6
Teoría de las Variables Regionalizadas.....	8
Conceptos de la Variable Aleatoria Regionalizada.....	9
Hipótesis de la Geoestadística.....	10
Estacionaridad Estricta.....	11
Estacionaridad de segundo Orden.....	11
Hipótesis Intrínseca.....	12
Etapas del Estudio Geoestadístico.....	14
Análisis Exploratorio de los datos.....	14
Conceptos Necesarios de Estadística Básica.....	15
Análisis Estructural.....	19
Variograma y Semivariograma.....	19
Covariograma y Correlograma.....	20
Estimación del Semivariograma.....	21

Modelado de Semivariogramas.....	23
Parámetros del Semivariograma.....	23
Modelos Teóricos de Semivariogramas.....	25
Estimación Kriging.....	29
Kriging Simple.....	31
Kriging Ordinario.....	34
Kriging Universal.....	40
Validación Cruzada.....	44
3.-MATERIALES Y METODOS.....	47
Localización.....	47
Clima.....	48
Recopilación de Información.....	49
Análisis de la Información.....	50
Análisis Estadístico.....	50
Análisis de Normalidad.....	50
Contrastes de Hipótesis.....	51
Prueba de Bondad de Ajuste de Kolmogorov-Smirnov.....	52
Métodos Gráficos.....	54
Datos Atípicos (Outliers).....	55
Tipos de Outliers.....	55
Identificación de Outliers.....	56
Análisis Geoestadístico.....	57
Estimación de la función de correlación espacial.....	57
Consideraciones para el computo del Variograma.....	58

Modelación de la función de Correlación espacial.....	58
Estimación espacial de la propiedad Estudiada.....	59
Validación del modelo de Variograma.....	60
Inferencias estadísticas del error.....	60
4.-RESULTADOS Y DISCUSION.....	62
Análisis Estadístico.....	62
Parámetros Estadísticos.....	62
Prueba de Normalidad.....	62
Identificación de Outliers.....	64
Prueba de Grubbs.....	65
Análisis Geoestadístico.....	66
Estimación del Variograma.....	66
Estimación espacial Kriging.....	71
Análisis de Varianza para la estimación.....	72
Validación del modelo del Variograma.....	74
Inferencias estadísticas del error.....	76
5.- CONCLUSIONES.....	80
6.- LITERATURA CITADA.....	82
7.-APÉNDICE.....	85

ÍNDICE DE CUADROS

	Página
Cuadro 2.1. Tipos de Métodos Kriging	31
Cuadro 3.1. Distribución Espacial de las estaciones Climatológicas de la región sureste del estado de Coahuila	49
Cuadro 3.2. Modelos y valores de α	53
Cuadro 3.3. Tabla de Ecuaciones de K(n)	53
Cuadro 4.1. Parámetros Estadísticos de la Temperatura Media anual de la región sureste del estado de Coahuila	62
Cuadro 4.2. Prueba de Grubbs	66
Cuadro 4.3. Parámetros Geoestadísticos de los diferentes modelos teóricos de semivariogramas ajustados.	69
Cuadro 4.4. Estimaciones Kriging	71
Cuadro 4.5. Estadísticas de la Regresión entre las Temperaturas observadas y Estimadas	72
Cuadro 4.6. Análisis de Varianza de la estimación	72
Cuadro 4.7. Errores al aplicar la metodología Kriging para las observaciones de la temperatura media anual de la región sureste de Coahuila	77
Cuadro 4.8. Cuadro de comparación de las temperaturas medias anuales observadas y estimadas de la región sureste del estado de Coahuila.	79

ÍNDICE DE FIGURAS

	Página
Figura 2.1. Parámetros del Semivariograma.....	24
Figura 2.2. Modelo de Efecto de Pepita.....	26
Figura 2.3. Modelo con función potencia.....	26
Figura 2.4. Modelo Lineal.....	27
Figura 2.5. Modelo Esférico.....	27
Figura 2.6. Modelo Exponencial.....	28
Figura 2.7. Modelo Gaussiano.....	29
Figura 2.8. ¿Cómo estimar el valor en el sitio “?”.....	30
Figura 3.1. Región Sureste del estado de Coahuila.....	48
Figura 4.1. Histograma con Distribución Normal de la TMA.....	63
Figura 4.2. Diagrama P-P de la Temperatura Media Anual.....	64
Figura 4.3. Diagrama Q-Q de la Temperatura Media Anual.....	64
Figura 4.4. Diagrama de Caja de la Temperatura Media Anual.....	64
Figura 4.5. Variograma tipo Esférico.....	68
Figura 4.6. Variograma tipo Exponencial.....	68
Figura 4.7. Variograma tipo Lineal.....	68
Figura 4.8. Variograma tipo Lineal con Umbral.....	69
Figura 4.9. Variograma tipo Gaussiano.....	69
Figura 4.10. Diagrama de Regresion de las Temperaturas Observadas y Estimadas.....	72

RESUMEN

APLICACIÓN DE LA TEORIA GEOESTADISTICA EN LA ESTIMACIÓN DE LA TEMPERATURA MEDIA ANUAL EN LA REGIÓN SURESTE DEL ESTADO DE COAHUILA, MÉXICO.

POR

FRANCISCO JAVIER SÁNCHEZ ASPEYTIA

MAESTRO EN CIENCIAS EN ESTADISTICA EXPERIMENTAL

UNIVERSIDAD AUTONOMA AGRARIA ANTONIO NARRO
BUENAVISTA, SALTILLO, COAHUILA, MEXICO. MARZO DE 2011.

- ASESOR PRINCIPAL-

DR. JAVIER DE JESUS CORTÉS BRACHO

Palabras Clave: Geoestadística, Correlación Espacial, Semivariograma, Estimación Kriging, Validación.

El objetivo de este Trabajo es mostrar como la geoestadística es una herramienta imprescindible cuando se desea analizar los patrones de distribución espacial de las variables medioambientales a partir de muestreos realizados en el área de interés. Asimismo, los procedimientos de estimación geoestadísticos, conocidos como kriging, permiten la realización de las "mejores interpolaciones" en aquellos lugares donde no se conoce la magnitud del atributo investigado., para conseguir unos buenos resultados, es preciso conocer sus fundamentos y las asunciones sobre las que se basa.

El kriging (en inglés Kriging) es una familia de métodos lineales para la estimación de cantidades con dependencia espacial y su varianza que son óptimos en el sentido de los mínimos

cuadrados. El krigado se ha desarrollado fundamentalmente en las Ciencias de la tierra, medicina, geología, biología, etc. para aquellos en el caso de que las observaciones presenten ruido, estén irregularmente distribuidas y su número no sea excesivamente grande. La ventaja fundamental del krigado respecto a otras técnicas es que utiliza las correlaciones espaciales de la cantidad medida para reducir el ruido en las estimaciones. Además, el krigado permite la determinación de la incertidumbre asociada a la medición.

Como ejemplo de aplicación de la geoestadística se estudia la estimación de la temperatura media anual en la región sureste del estado de Coahuila a través de 21 estaciones climatológicas dependientes del Servicio Meteorológico Nacional.

Para que el análisis geoestadístico sea válido, puesto que el mismo está basado en la estimación y modelación de una función que refleja la correlación espacial de la propiedad que se estudia: el semivariograma. Para poder estimarla es necesario que la muestra cumpla una serie de requisitos, como son:

- 1).-Que su distribución de probabilidad sea normal.
- 2).-Que no exista tendencia, es decir que sea estacionaria al menos la media.
- 3).-Que no se vea afectada por valores atípicos (outliers) tanto distribucionales como espaciales.
- 4).-Que tenga una distribución espacial homogénea

Se Utilizo el Programa GS+ (Geostatistics for the Environmental Sciences) Versión 5.1 para Windows para calcular los variogramas y posteriormente la interpolación Kriging.

Se genero una Base de datos que contiene las coordenadas georreferenciadas en el sistema UTM (Universal Transversal de Mercator) de cada estación y su respectiva medición de la variable.

El GS+ genero Varios tipos de modelos de Variogramas y se determino que le modelo Gaussiano

fue el de mejor ajuste con un coeficiente de Determinación R^2 de 1.000.

Con este modelo se procedió a la interpolación Kriging el cual se debe tener en cuenta el tamaño de la vecindad y el número de puntos vecinos que intervienen en la estimación. Normalmente el tamaño de la vecindad se elige cercano al alcance o radio de correlación y se considera 16 puntos como una cantidad razonable para una buena estimación.

Se considero un radio de 150 Km. y vecinos = 16 para la estimación así mismo también se considero a los puntos irregularmente espaciados tomando en cuenta las coordenadas Cartesianas.

Posterior a la interpolación se realizaron una serie de Pruebas para los valores observados y estimados tales como un análisis de regresión, su análisis de Varianza y la prueba “t” de Student para comprobar si existe Correlacion en los datos

Para Validar el modelo de estimación se calcularon medidas de precisión tales como:

- 1.- El promedio del error (E.M.) que mide el seso del estimador.
- 2.-La Varianza de los Errores o Cuadrado Medio del Error (C.M.E.) que mide la precisión del estimador.
- 3.-Como medida de Efectividad, el estimado de Efectividad de Predicción (E).

Se plantearon pruebas de hipótesis para demostrar que los errores en la interpolación kriging se distribuyen normal con media cero y varianza igual a uno.

ABSTRACT

**APPLICATION OF GEOSTATISTICAL THEORY IN THE ESTIMATION OF THE
AVERAGE ANNUAL TEMPERATURE IN THE REGION SOUTHEAST OF THE
STATE OF COAHUILA, MEXICO.**

FOR

FRANCISCO JAVIER SANCHEZ ASPEYTIA

MASTER OF SCIENCE IN EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO
BUENAVISTA, SALTILLO, COAHUILA, MEXICO. MARCH 2011.

- ADVISER-

DR. JAVIER DE JESUS CORTÉS BRACHO

Key words: Geostatistics, Spatial Correlation, Semivariogram, Kriging Estimation,
Validation.

The objective of this work is to show how geostatistics is an indispensable tool when you want to analyze the spatial distribution patterns of environmental variables from surveys conducted in the area of interest. Furthermore, geostatistical estimation procedures, known as kriging, allow the realization of "best tweens" in places where no one knows the magnitude of the attribute under investigation. In order to achieve good results, it is necessary to understand their rationale and assumptions about which is based

The kriging is a family of linear methods for estimating quantities with spatial dependence and variance that are optimal in the sense of least squares. Kriging has been developed mainly in the

Earth sciences, medicine, geology, biology, etc.. for those in the case of noise present observations, are unevenly distributed and their number is too large. The main advantage of kriging over other techniques is that it uses the spatial correlations of the measured quantity to reduce noise in the estimates. In addition, kriging allows the determination of the uncertainty associated with the measurement.

As an example of application of geostatistics studies the estimate of the average annual temperature in the state of Coahuila through 21 weather stations dependent on the National Weather Service.

For the geostatistical analysis is valid, since it is based on the estimation and modeling of a function that reflects the spatial correlation of the property being studied: the semivariogram. To estimate it is necessary that the sample meets certain requirements, such as:

- 1). - That its probability distribution is normal.
- 2). - There is no trend, ie it is stationary at least average.
- 3). - Not affected by outliers as well as distributional space.
- 4). - Have a homogeneous spatial distribution

Program was used GS + (Geostatistics for the Environmental Sciences) version 5.1 for Windows to calculate the variograms and kriging interpolation later.

It generated a database containing geo-referenced coordinates in the UTM (Universal Transverse Mercator) of each station and their respective temperature measurement.

The GS + generates various types of variogram models and Gaussian model determined that it was the best fit with a coefficient of determination r^2 of 1.000

With this model we proceeded to the Kriging interpolation which must take into account the size of the neighborhood and the number of neighboring points involved in the estimate. Normally the

size of the neighborhood is chosen close to the range or correlation radius and 16 points is considered a reasonable amount for a good estimate.

It was considered a radius of 150 km and neighbors = 16 for the estimate and it is also considered to irregularly spaced points taking into account the coordinates Cartesian.

After the interpolation is carried out a series of tests for observed and estimated values such as regression analysis, analysis of variance and the "t" of Student to check whether there is correlation in the data.

To validate the estimation model were calculated precision measurements such as:

1.-The mean error (ME) which measures the bias of the estimator.
2.-The Variance of Errors and Mean Square Error (MSE) which measures the accuracy of the estimator

3.-As a measure of effectiveness, the estimated prediction effect (E).

Tests of hypotheses were raised to show that the kriging interpolation errors are normally distributed with zero mean and variance equal to one

1.-INTRODUCCION

El estudio de fenómenos con correlación espacial, por medio de métodos geoestadísticos, surgió a partir de los años sesenta, especialmente con el propósito de predecir valores de las variables en sitios no muestreados. Como antecedentes suelen citarse trabajos de Sichel (1947; 1949) y Krige (1951). El primero observó la naturaleza asimétrica de la distribución del contenido de oro en las minas sudafricanas, la equiparó una distribución de probabilidad lognormal y desarrolló las fórmulas básicas para esta distribución. Ello permitió una primera estimación de las reservas, pero bajo el supuesto de que las mediciones eran independientes, en clara contradicción con la experiencia de que existen “zonas” más ricas que otras. Una primera aproximación a la solución de este problema fue dada por geólogo G. Krige que propuso una variante del método de medias móviles, el cual puede considerarse como el equivalente al krigeado simple que, como se verá más adelante, es uno de los métodos de estimación lineal en el espacio con mayores cualidades teóricas. La formulación rigurosa y la solución al problema de predicción (estimación en muchos textos geoestadísticos) inició con Matheron (1962) en la escuela de minas de París. En los años sucesivos la teoría se fue depurando, ampliando su campo de validez y reduciendo las hipótesis necesarias (Samper y Carrera, 1990). De la minería las técnicas geoestadísticas, se han "exportado" a muchos otros campos como hidrología, física del suelo, ciencias de la tierra y más recientemente al monitoreo ambiental y al procesamiento de imágenes de satélite.

Aunque la aplicación de la herramienta geoestadística es bastante reciente, son innumerables los ejemplos en los que se ha utilizado esta técnica en estudios ambientales con el ánimo de predecir fenómenos espaciales (Robertson, 1987; Cressie y Majure, 1995; Diggle *et al.*, 1995). La columna vertebral del análisis geoestadístico es la determinación de la estructura de autocorrelación entre los datos y su uso en la predicción a través de las técnicas conocidas como kriging y cokriging. Otros temas importantes dentro del estudio de información georreferenciada son el diseño de redes de muestreo (McBratney *et al.*, 1981), la geoestadística multivariada (Wackernagel, 1995) y la simulación (Deutsch y Journel, 1992).

La geoestadística es solo una de las áreas del análisis de datos espaciales. Es importante reconocer cuando la información georreferenciada es susceptible de ser analizada por medio de dicha metodología. En el estudio de información georreferenciada, de forma análoga a como se procede en la aplicación de muchos procedimientos estadísticos, la primera etapa que se debe cumplir es la del análisis exploratorio de datos (AED). Esta busca identificar localización, variabilidad, forma y observaciones extremas.

Se describen los procedimientos empleados para identificar de manera experimental (con base en datos muestrales) la estructura de autocorrelación espacial, para algunas distancias dadas, de un conjunto de datos de una variable. Se muestra también como generalizar dicha estructura para cualquier distancia entre los sitios de observación. Una vez detectada la autocorrelación espacial, el siguiente paso es la predicción en sitios de la región de estudio donde no se ha hecho medición de la variable de interés.

Algunos métodos de interpolación se emplean generalmente para obtener estos valores en los nodos de la red. Sin embargo, con frecuencia se descuida lo más importante: la

aplicación de un método óptimo para determinar los valores en aquellos lugares donde no se ha realizado una medida directa de la variable o atributo. Aunque muchos usuarios son capaces de manejar un software muy sofisticado, un gran porcentaje de los mismos no llegan a entender el porqué de los resultados conseguidos. Es notorio que, debido a este desconocimiento, se suele seleccionar el método de interpolación que los programas proveen por defecto, lo cual puede ser adecuado en algunos casos pero tremendamente erróneo en otros

Un primer escollo se encuentra en que los datos disponibles son, con cierta generalidad, escasos y heterogéneos (por ejemplo, datos puntuales relativos a la concentración de un soluto, posición del nivel freático, cantidad de lluvia recogida, etc.), debido, en algunos casos a la dificultad de acceso a los lugares donde se desea muestrear o, en otros, al costo que supone la obtención de datos de ciertas variables.

También, la resolución de un problema, como la distribución de la lluvia o la medición del nivel freático de un pozo de una región determinada, exige una cierta comprensión de los procesos que influyen en dicha distribución y una estimación estadística para inferir el valor de la variable estudiada en aquellos lugares donde no se haya muestreado.

Estos datos no muestrales se estiman a partir de los valores conocidos en los puntos muestrales, donde se mide la variable. Debido a ello, es necesario seleccionar el método de interpolación que nos genere unos resultados que sean lo más próximo posible a los reales.

La Geoestadística proporciona un conjunto de de herramientas estadísticas para el análisis de la variabilidad espacial e interpolación espacial. Los métodos geoestadísticos

se puede utilizar para determinar los valores de los puntos donde las mediciones no se hacen o no es factible de medir debido a la consideración económica. Los métodos de estimación geoestadística son conocidos como krigeado o krigeaje (kriging en la literatura inglesa, en honor de Daniel Krige quien formulo por primera vez esta metodología en 1951). Kriging proporciona la mejor estimación objetiva lineal para la interpolación espacial.

Hasta el día de hoy, todos los interpoladores geoestadísticos están en el grupo de los Kriging (con sus variantes), los cuales ofrecen no solo predicciones y superficies de respuesta requeridas, sino también mapas de probabilidades y cuantiles (Johnston *et al.* 2001).

En el presente trabajo se mostrará como el uso adecuado de los métodos de estimación geoestadísticos es la mejor herramienta para conseguir una estimación real del atributo objeto de investigación.

Objetivos.

Por lo anteriormente mencionado, los objetivos planteados en este trabajo son:

General

.- Desarrollar una metodología que permita la distribución espacial de la variable temperatura media anual de la región sureste del estado de Coahuila.

Específicos

a).-Aplicar técnicas geoestadísticas que conduzcan a evaluar la distribución espacial de la variable temperatura media anual.

b).-Analizar la distribución espacial de la variable temperatura media anual.

c).-Validar el modelo de acuerdo a datos conocidos de la variable temperatura media anual con la aplicación del método Kriging.

2.-REVISION DE LITERATURA

Reseña Histórica de la Geoestadística

El uso de la Geoestadística surge a partir de la década de los años 50 con los estudios realizados por D.G. Krige y por otros Ingenieros mineros en las minas de oro de Rand, África del Sur, desarrollaron un procedimiento ponderado de estimación, con determinados coeficientes correctores empíricos.

El desarrollo posterior es atribuido, en paralelo, a Matèrn y Matheron, quien estuvo ligado directamente a la Escuela de Minas de París, desarrollo la teoría de las variables regionalizadas, que provee una buena base para los recursos naturales, esta teoría incluye una técnica de estimación óptima llamada Kriging, esta técnica se hizo conocida debido a que Matheron hizo las publicaciones en textos en Inglés y campos como la Hidrogeología empezaron a implementar la Geoestadística.

En la ciencia del suelo, el reconocimiento de la variación espacial de propiedades del suelo, es probablemente tan viejo como su clasificación. Este enfoque de caracterizar la variabilidad de los suelos es principalmente cualitativo

En los años 70, los investigadores empezaron los estudios de las propiedades de suelo de una manera más cuantitativa (Rogowski, 1972; Nielsen et al, 1973; Biggar y Nielsen, 1976), la mayoría de ellos confió en métodos estadísticos clásicos (p. ej. La función de la distribución de frecuencias con momentos asociados y coeficientes de variación; análisis de varianza para medir cambios sobre la distancia, etc.).

Una suposición implícita en estos análisis es; que las observaciones de una variable del suelo son independientes con respecto a su ubicación en el campo. Sin embargo, solamente el conocimiento de la distribución de frecuencias de las observaciones provee información sobre la variabilidad de las observaciones con respecto a las coordenadas del área donde se ha hecho un muestreo

La primera publicación que aplicó Geoestadística apareció en 1978 (Campbell, 1978). En 1980, Webster y su colegas iniciaron una serie de publicaciones llamadas “Interpolación Óptima y mapeo isarithmico de propiedades del suelo”, en la que Kriging era por primera vez usado en la ciencia del suelo (Burguess & Webster, 1980, Burguess et al, 1981; McBratney & Webster, 1983, webster & Oliver, 1989). Samra en 1989 realizó estimación de alturas en plantaciones de Melia azedarach, (Mandallaz, 1993), que comparó varios métodos de kriging en la estimación de densidad y área basal. (Höck et al, 1994), combinó el uso de Sistemas de Información Geográficos con técnicas geoestadísticas en la estimación de índices de sitio de Pinus radiata en Nueva Zelanda.

La palabra Geoestadística, denota el estudio estadístico de fenómenos naturales, tiene varias interpretaciones como la de Matheron, que la define como: “la aplicación del formalismo de las funciones aleatorias al reconocimiento y estimación de fenómenos naturales” (Journel & Huijbregts, 1978). El término Geoestadística nos indica, como algunos interpretan incorrectamente, “estadísticas en la geología”.

Una más reciente (y orientado al SIG) interpretación de Geoestadística, podría ser: “las estadísticas involucradas con datos geográficos”, también llamada “estadísticas espaciales”.

Chilès (1999, p. 2) la define como “la aplicación de los métodos probabilísticas a las variables regionalizadas”. Olea (1999 p. 1) la considera como “un conjunto de técnicas

que se auxilian de la caracterización espacial de los atributos, empleando modelos primarios aleatorios, de manera similar a las series temporales”. Quizás, la definición más acertada es “la ciencia que estudia las variables regionalizadas”, dada por Le Loc’h (2005). Una caracterización bastante aceptable es presentada por Dubois (1998); él destaca el papel del análisis espacial o temporal del fenómeno estudiado.

De acuerdo con Stoyan *et al.* (2000), la geoestadística proporciona un medio para definir la autocorrelación y utilizar el conocimiento de la fuerza y la escala para interpolar el valor de las variables en localizaciones no muestreadas utilizando el método de *kriging*.

Teoría de las Variables Regionalizadas

Una variable medida en el espacio de forma que presente una estructura de correlación, se dice que es una variable regionalizada. De manera más formal se puede definir como un proceso estocástico con dominio contenido en un espacio euclidiano d -dimensional \mathbb{R}^d $\{Z(x): x \in D \subset \mathbb{R}^d\}$. Si $d = 2$, $Z(x)$ puede asociarse a una variable medida en un punto x del plano (Díaz-Francés, 1993). En términos prácticos $Z(x)$ puede verse como una medición de una variable aleatoria (p.ej. medición de la temperatura) en un punto x de una región de estudio.

Recuérdese que un proceso estocástico es una colección de variables aleatorias indexadas; esto es, para cada x en el conjunto de índices D , $Z(x)$ es una variable aleatoria.

En el caso de que las mediciones sean hechas en una superficie, entonces $Z(x)$ puede interpretarse como la variable aleatoria asociada a ese punto del plano (x representa las coordenadas, planas o geográficas, y Z la variable en cada una de ellas).

Estas variables aleatorias pueden representar la magnitud de una variable ambiental medida en un conjunto de coordenadas de la región de estudio.

Conceptos de Variable Aleatoria Regionalizada

Según (Alfonso, 1989b) es importante aclarar varios conceptos en el estudio de la variables aleatorias regionalizadas, conceptos que se presentan en [Journel, 1978], [David, 1977] y que son utilizados en la mayoría de los artículos consultados donde se aplican los métodos Geoestadísticos como herramienta fundamental de trabajo.

Estos conceptos son:

a).-Región: se refiere al espacio en el cual existe y se estudia el fenómeno natural.

b).-Localización: Es el punto de una región en la cual se define una variable aleatoria regionalizada.

c.-Soporte Geométrico: Está determinado por el elemento físico sobre el cual se realiza la determinación de la variable aleatoria regionalizada, Esto no es más que la muestra unitaria, sobre la cual estudiaremos el atributo de interés.

d).-Momentos de primer orden: Si la función de distribución de $Z(x_i)$ tiene una media definida, será una función de la localización x_i :

$$m(x_i) = E \{Z(x_i)\} \quad (1)$$

e).-Momentos de segundo orden: Si la varianza (Var) de $Z(x_i)$ existe, entonces se define como el momento de segundo orden y será también una función de la localización x_i :

$$\text{Var} \{Z(x_i)\} = E \{[Z(x_i) - m(x_i)]^2\} = E [Z(x_i)^2] - m(x_i)^2 \quad (2)$$

*Si la varianza de las variables $Z(x_i)$ y $Z(x_j)$ existe entonces la covarianza (Cov) de estas también existe y es función de las localizaciones x_i y x_j :

$$\text{Cov} [Z (x_i), Z (x_j)] = E \{[Z (x_i) - m (x_i)][Z (x_j) - m (x_j)]\} \quad (3)$$

$$\text{Si } x_i = x_j; \text{ Cov} [Z (x_i), Z (x_j)] = \text{Var} \{Z (x_i)\} \quad (4)$$

f).-La función variograma o función estructural se define como la varianza de la diferencia $Z (x_i) - Z (x_j)$.

$$\text{Var} \{Z (x_i) - Z (x_j)\} = 2\gamma (x_i, x_j) \quad (5)$$

La magnitud $\gamma (x_i, x_j) = \frac{1}{2} \text{Var} \{Z (x_i) - Z (x_j)\}$ se denomina semivariograma.

También se puede definir el correlograma estandarizando la covarianza para los valores $x_i - x_j = h = 0$ como:

$$\rho (h) = C (h) / C(0) \quad -1 \leq \rho \leq 1 \quad (6)$$

donde: $C (h)$ es la covarianza a la distancia h ,

$C (0)$ es la covarianza en el origen.

Existen relaciones entre estas medidas de correlación:

$$\gamma (h) = C(0) - C(h) \quad \text{con } \gamma(0) = 0 \quad (7)$$

$$\rho (h) = 1 - \gamma(h) / C(0) \quad (8)$$

Hipótesis de la Geoestadística

Como la forma en que se presenta la información es muy diversa (Journel y Huijbregts, 1978), la geoestadística se construye asumiendo condiciones de estacionaridad. Por lo que es necesario aceptar el cumplimiento de ciertas hipótesis sobre el carácter de la función aleatoria o procesos estocásticos estudiados, llamadas Hipótesis de la Geoestadística.

Estas son según Journel y Huijbregts (1978) y David (1977): La Estacionariedad Estricta, La Estacionariedad de Segundo Orden, La Hipótesis Intrínseca.

Estacionariedad Estricta. Se dice que $Z(x)$ es estrictamente estacionaria si la función de distribución de probabilidades de las variables aleatorias regionalizadas $Z(x_i)$ son iguales entre sí, independiente de la localización x_i , lo que requiere que los momentos de distinto orden para cada variable aleatoria regionalizada sean completamente independientes de la localización x_i . Esto se expresa por medio de la probabilidad de:

$$\Pr\{Z(x_i) < z_i, \dots, Z(x_k) < z_k\} = \Pr\{Z(x_i + h) < z_i, \dots, Z(x_k + h) < z_k\} \quad (9)$$

Tal función es aleatoria. Físicamente esto significa que el fenómeno es homogéneo en el espacio y, por decirlo así, se repite así mismo en todo el espacio. En términos espaciales esto implica isotropía.

Esta condición como su nombre lo indica es demasiado restrictiva al estudiar la mayoría de los fenómenos encontrados en la práctica.

Estacionariedad de Segundo Orden. Esta condición es más frecuente en la práctica, la misma requiere que:

$$1) E\{Z(x_i)\} = m, \text{ existe y no depende de la localización } x_i \quad (10)$$

2) La función de covarianza: $\text{Cov}\{Z(x_i) - Z(x_j)\}$ exista y solo dependa de la longitud del vector $h = x_i - x_j$. o sea.

$$C(h) = \text{Cov}\{Z(x_i), Z(x_j)\} = E\{Z(x_i), Z(x_i+h)\} - m^2 \quad (11)$$

Esta hipótesis requiere la estacionariedad solo para la media y para la función de covarianza de la variable aleatoria regionalizada.

La segunda condición implica, estacionaridad de la varianza y del variograma.

$$1.- \text{Var}[Z(x_i)] = E\{[Z(x_i) - m]^2\} = C(0) \quad \forall x \quad (12)$$

$$2.- \gamma(h) = E \{ [Z(x_i)]^2 \} - E \{ Z(x_i), Z(x_i+h) \} \quad \forall x \quad (13)$$

$$\text{Como: } E [Z(x_i), Z(x_i+h)] = C(h) + m^2$$

$$\text{y } E[Z^2(x_i)] = C(0) + m^2$$

$$\gamma(h) = C(0) + m^2 - (C(h) + m^2)$$

$$\gamma(h) = C(0) - C(h). \quad (14)$$

Como se observa en la última expresión $\gamma(h)$ y $C(h)$, son dos herramientas que permiten expresar la correlación entre la variable aleatoria regionalizada $Z(x_i)$ y $Z(x_i+h)$, separadas por el vector h .

Hipótesis Intrínseca. Una función aleatoria $Z(x)$ se dice intrínseca cuando:

a) Su esperanza matemática existe y no depende de la localización X_i .

$$E \{ Z(x) \} = m \quad \forall x \quad (15)$$

b) Para todo vector h el incremento $[Z(x+h) - Z(x)]$ tiene varianza finita y no depende de la localización x_i :

$$\text{Var} \{ Z(x+h) - Z(x) \} = E \{ [Z(x+h) - Z(x)]^2 \} = 2\gamma(h) \quad \forall x \quad (16)$$

Cuando se cumple esta condición se dice que la función aleatoria $Z(x)$ es homogénea. Esta condición se encuentra con bastante frecuencia en la naturaleza, pues existen muchos procesos que no tienen varianza finita y sin embargo, poseen una función variograma finita.

La estacionaridad de segundo orden, siempre implica la condición intrínseca (homogeneidad), sin embargo la relación inversa no siempre se cumple.

Es claro que si una variable regionalizada es estacionaria fuerte entonces también será estacionaria débil.

El concepto de estacionariedad es muy útil en la modelación de series temporales (Box & Jenkins, 1976). En este contexto es fácil la identificación, puesto que sólo hay una dirección de variación (el tiempo). En el campo espacial existen múltiples direcciones y por lo tanto se debe asumir que en todos los fenómenos es estacionario. Cuando la esperanza de la variable no es la misma en todas las direcciones o cuando la covarianza o correlación dependan del sentido en que se determinan, no habrá estacionariedad.

Si la correlación entre los datos no depende de la dirección en la que esta se calcule se dice que el fenómeno es isotrópico, en caso contrario se hablará de anisotropía. En Isaaks y Srivastava (1989) se definen los posibles tipos de anisotropía y se proponen algunas soluciones. Cressie (1993) discute cual debe ser el tratamiento en caso de que la media no sea constante.

En casos prácticos resulta compleja la identificación de la estacionariedad. Suelen emplearse gráficos de dispersión de la variable respecto a las coordenadas, de medias móviles y de valores clasificados según puntos de referencia, con el propósito de identificar posibles tendencias de la variable en la región de estudio. La isotropía es estudiada a través del cálculo de funciones de autocovarianza o de semivarianza muestrales en varias direcciones. Si estas tienen formas considerablemente distintas puede no ser válido el supuesto de isotropía. Finalmente una variable regionalizada será no estacionaria si su esperanza matemática no es constante, esto es si $E [Z(x)] = m(x)$.

En la práctica según Armstrong y Carignan (1997) y Chica (1987) son dos las hipótesis que más se presentan: La Estacionariedad de Segundo Orden y la Hipótesis Intrínseca.

Estas condiciones de estacionariedad se asumen en el desarrollo teórico, en la práctica deben ser verificadas en los datos antes de comenzar un estudio geoestadístico, para lo

que se puede realizar un análisis estadístico de la información, de modo que se refleje de así el grado de confiabilidad en la aplicación de estos métodos.

Etapas del Estudio Geoestadístico.

Antes de comenzar un estudio geoestadístico se deben discutir todos los elementos que aporten conocimientos del problema a resolver o el fenómeno en estudio, organización y verificación de la información disponible y finalmente realizar el análisis exploratorio de los datos.

Una vez obtenido los datos, es necesario que se controlen integralmente a fin de verificar de una parte su exactitud y de otra su representatividad. Es importante que se esté familiarizado con los datos, discutir todos los elementos necesarios a fin de conocer el problema a resolver (Armstrong y Carignan, 1997). En la minería los resultados son muy sensibles al nivel de información usado (Carrasco-Castelli y Jara-Salame, 1998; Lantuéjoul, 1994), cualquier modificación involuntaria en la etapa inicial se refleja sistemáticamente durante todo el estudio (Armstrong y Roth, 1997; Armstrong y Carignan, 1997).

Todo Trabajo Geoestadístico tiene que llevarse a cabo en tres etapas:

Análisis Exploratorio de los Datos

Se Estudian los datos muestrales sin tener en cuenta su distribución geográfica. Sería una etapa de la aplicación de la estadística básica. Se comprueba la consistencia de los datos eliminándose los erróneos, e identificándose las distribuciones de las que provienen

Stoyan *et al.* (2000) mencionan que la estadística clásica paramétrica no puede usarse para evaluar datos autocorrelacionados sin violar el supuesto central de independencia de muestras.

Conceptos necesarios de estadística básica

Con el objetivo de conocer la información disponible se puede hacer un análisis de la estadística descriptiva (Krajewski y Gibbs, 1993; Journal y Huijbregts, 1978; David, 1977). A continuación se presenta un resumen de los conceptos necesarios de estadística básica.

A: Cálculos estadísticos o estadística descriptiva. Permiten determinar si la distribución de los datos es normal, lognormal, o si no se ajustan a una distribución estadística, lo cual implica tener conocimiento de:

1.- Numero de casos: Es el número de valores muestreados del fenómeno en estudio, representados por n y los datos por $x_i, i = 1, \dots, n$, que llamamos distribución.

2.- Rango de la distribución: Es la diferencia entre el valor máximo y el mínimo.

3.- Media: Es la media aritmética de la distribución, dado por la fórmula:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (17)$$

4.- Moda: Es el valor más frecuente de la distribución.

5.- Mediana: Es el valor para el cual la mitad de los datos son menores y la otra mitad están por encima de este valor.

Si ordenamos los datos en orden ascendente podemos calcular la posición de la mediana como.

$$M = \begin{cases} X_{(n+1)/2} & \text{si } n \text{ es impar.} \\ (X_{n/2} + X_{n/2+1})/2 & \text{si } n \text{ es par.} \end{cases} \quad (18)$$

La mediana es también llamada percentil 50, además los datos no solo se dividen en dos grupos, sino que se pueden dividir en cuatro partes, cuartiles, donde $Q_1 =$ percentil 25,

$Q_2 =$ Mediana y $Q_3 =$ percentil 75, si los datos se dividen en 10, tenemos los deciles. De forma general estas medidas se pueden calcular por: $[p(n+1)/100]$ i-ésima observación de los datos ordenados ascendentemente, donde p es el percentil que se desea calcular.

6.- Varianza: Describe la variabilidad de la distribución. Es la medida de la desviación o dispersión de la distribución y se calcula por:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (19)$$

La razón principal por la que se aboga por la división entre n-1 en la estimación de la varianza, es porque proporciona un mejor estimado; si dividimos por n-1 nos referimos a la varianza muestral S^2 como un estimador insesgado de la varianza poblacional σ^2 . Esto significa que si un experimento fuera repetido muchas veces se podría esperar que el promedio de los valores así obtenidos para S^2 igualaría a σ^2 . Por otra parte si dividimos entre n los valores obtenidos para S^2 serían como promedio demasiado pequeño.

7.- Desviación estándar: Describe la tendencia o dispersión de la distribución. Es la medida de desviación alrededor de la media. Se calcula por:

$$s = \sqrt{s^2} \quad (20)$$

8.- Coeficiente de asimetría: Describe la simetría de la distribución relativa a la distribución normal. Se calcula por:

$$\alpha_3 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3 / s^3 \quad (21)$$

En la distribución normal la asimetría tiene valor cero, un valor negativo indica una cola a la izquierda y un valor positivo indica una cola a la derecha.

9.- Curtosis: Describe el grado de esbeltez de la distribución, tomado por lo general en relación a una distribución normal, y se puede calcular por:

$$\alpha_4 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4 / S^4 \quad (22)$$

La distribución normal tiene curtosis igual a 3 y es llamada mesocúrtica. A las distribuciones más agudas, con colas relativamente anchas, se les llama leptocúrticas, tienen valores de curtosis mayores que 3, y las distribuciones más bien achatadas en el centro se llaman platicúrticas, tienen valores menores que tres, en ocasiones se acostumbra a definir la curtosis como $\alpha_4 - 3$.

10.- Error estándar de la media: Describe el grado de conocimiento de los datos y se puede calcular por:

$$\varepsilon = \sqrt{\frac{S^2}{n}} \quad (23)$$

La distribución normal tiene un valor de error estándar menor que 1.25 y la distribución lognormal o una distribución con tendencia positiva, tiene valores de error estándar mayores que 1.25.

11.- Coeficiente de variación: Es una medida de la variación relativa de los datos y puede ser calculado por:

$$c.v. = \frac{s}{\bar{X}} \quad (24)$$

Proporciona una comparación entre la variación de grandes valores y la variación de pequeños valores. Las técnicas de Geoestadística Lineal que predomina en el campo de las geociencias producen los mejores resultados cuando el coeficiente de variación es menor que uno, $CV < 1$. Para $CV > 1$ se recomiendan técnicas de Geoestadística no Lineal.

12.- Prueba Chi-Cuadrado: Permite determinar si la distribución es normal, lognormal o alguna otra distribución probabilística, en su lugar puede ser usada la prueba “Kolmogorov Smirnov” como se refleja por muchos autores es más robusta.

B: Construcción de gráficos estadísticos: Estos gráficos permiten ilustrar y entender las distribuciones de los datos, identificar datos errados, valores extremos, los mismos incluyen:

1.- Mapa base, sección cruzada y vista en perspectiva: Son usados para visualizar la relación espacial en 2 y 3 dimensiones, permiten encontrar errores en la información.

2.- Histogramas: Son usados para ver las características descriptivas de la distribución. Es un gráfico de barras donde en las abscisas aparecen los límites de las clases y en las ordenadas las frecuencias correspondientes a cada clase.

3.-Frecuencia acumulativa: Usado para identificar el tipo de distribución muestral y ayuda a determinar si están presentes poblaciones mixtas. Es un gráfico de límite de clase contra frecuencia acumulada.

En el caso de gráficos estadísticos es útil usar los gráficos de frecuencia absoluta, relativa, acumulativa y el diagrama de dispersión, como se presenta en muchos sistemas. Todos estos elementos permiten decidir sobre las condiciones de estacionaridad vistas anteriormente. Muchos autores sólo toman como elementos fundamentales de estadística básica que: la media y la mediana tome valores próximos; el coeficiente de variación sea inferior a 1; la distribución de los datos esté próxima a la curva normal y no existan valores extremos que afecten el desarrollo del análisis estructural.

Análisis Estructural.

El análisis estructural es una de las etapas más importantes de la geoestadística puesto que se encarga de la caracterización de la estructura espacial de una propiedad o fenómeno regionalizado. Es el proceso en el marco del cual se obtiene un modelo geoestadístico para la función aleatoria que se estudia.

El análisis estructural consiste en estimar y modelar una función que refleje la correlación espacial de la variable regionalizada a partir de la adopción razonada de la hipótesis más adecuada acerca de su variabilidad.

En dependencia de las características de estacionaridad del fenómeno se modelará la función de covarianzas o la de semivarianzas.

Por su importancia y generalidad estudiaremos el proceso de estimación y modelación de la función de semivarianzas o semivariograma.

Para llevarla a cabo, con base en la información muestral, se usan tres funciones: El semivariograma, el covariograma y el correlograma.

Variograma y semivariograma

Cuando se definió la Hipótesis intrínseca se mencionó que se asumía que la varianza de los incrementos de la variable regionalizada era finita. A esta función denotada por $2\gamma(h)$ se le denomina variograma. Utilizando la definición teórica de la varianza en términos del valor esperado de una variable aleatoria, tenemos:

$$2\gamma(h) = \text{var.} [Z(x+h) - Z(x)] = E [(Z(x+h) - Z(x))^2] - [E (Z(x+h) - Z(x))]^2$$
$$2\gamma(h) = E [(Z(x+h) - Z(x))^2] \quad (25)$$

La mitad del variograma $\gamma(h)$, se conoce como la función de semivarianza y caracteriza las propiedades de dependencia espacial del proceso. Dada una realización del

fenómeno, la función de semivarianza es estimada, por el método de momentos, a través del semivariograma experimental, que se calcula mediante:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (26)$$

donde $Z(x_i)$ es el valor de la variable en un sitio x , $Z(x_i+h)$ es otro valor muestral separado del anterior por una distancia h y $N(h)$ es el número de parejas que se encuentran separadas por dicha distancia. La función de semivarianza se calcula para varias distancias h . En la práctica, debido a irregularidad en el muestreo y por ende en las distancias entre los sitios, se toman intervalos de distancia $\{[0, h], (h, 2h], (2h, 3h], \dots\}$ y el semivariograma experimental corresponde a una distancia promedio entre parejas de sitios dentro de cada intervalo y no a una distancia h específica. Obviamente el número de parejas de puntos $N(h)$ dentro de los intervalos no es constante.

Para interpretar el semivariograma experimental se parte del criterio de que a menor distancia entre los sitios mayor similitud o correlación espacial entre las observaciones. Por ello en presencia de autocorrelación se espera que para valores de h pequeños el semivariograma experimental tenga magnitudes menores a las que este toma cuando las distancias h se incrementan.

Covariograma y Correlograma.

La función de covarianza muestral entre parejas de observaciones que se encuentran a una distancia h se calcula, empleando la fórmula clásica de la covarianza muestral, por:

$$C(h) = Cov[Z(x_i + h), Z(x_i)] = \frac{\sum_{i=1}^n [Z(x_i + h) - m][Z(x_i) - m]}{n}$$

$$= \frac{\sum_{i=1}^n [Z(x_i + h)][Z(x_i)]}{n} - m^2 = C(h) \quad (27)$$

donde m representa el valor promedio en todo punto de la región de estudio y n es el número de parejas de puntos que se encuentran a una distancia h .

Asumiendo que el fenómeno es estacionario y estimando la varianza de la variable regionalizada a través de la varianza muestral, se tiene que el correlograma muestral está dado por:

$$r(h) = \frac{Cov[Z(x_i + h), Z(x_i)]^2}{(S_{x_i+h})(S_{x_i})} = \frac{C(h)}{S^2_x} = \frac{C(h)}{C(0)} \quad (28)$$

Bajo el supuesto de estacionariedad cualquiera de las tres funciones de dependencia espacial mencionadas, es decir semivariograma, covariograma o correlograma, puede ser usada en la determinación de la relación espacial entre los datos. Sin embargo como se puede observar en las fórmulas, la única que no requiere hacer estimación de parámetros es la función de semivarianza. Por esta razón, fundamentalmente, en la práctica se emplea el semivariograma y no las otras dos funciones.

Estimación del Semivariograma.

Debido a que el estimador es esencialmente una media muestral, tiene todas las desventajas comúnmente asociadas a este tipo de estimadores como es la no robustez.

A menudo el empleo de este estimador produce variogramas experimentales erráticos, lo cuál se debe a desviaciones del caso ideal para la aplicación del mismo

Estas desviaciones pueden ser enumeradas por su importancia en el orden siguiente:

- a).-Fuerte asimetría en la distribución de probabilidad.
- b).-Varianza no constante. La varianza varía con la posición.

c).-Muestreo sesgado. Los puntos de medición se encuentran distribuidos de manera no homogénea.

d).-Presencia de valores atípicos (outliers) distribucionales y espaciales

Muchos libros de Geoestadística dan la impresión descabellada de que es fácil calcular el semivariograma experimental, y ajustar este a un modelo teórico, ajuste que veremos posteriormente, pero quién ha tratado de hacerlo sabe que no es así. [Armstrong, 1984a] El primer semivariograma siempre es errático y se necesita mucho esfuerzo para lograr un semivariograma que represente realmente la variabilidad del fenómeno es estudio, se presentan en [Armstrong, 1984a] tres de los problemas que más se presentan en este proceso, que son:

a).-Pobre elección de la distancias.

b).-Población mixta presente en los datos.

c).-Valores fuera de límites o distribuciones asimétricas.

En [Krajewsky, 1993] se presentan otras razones por los que los semivariogramas son erráticos:

a).-No hay suficientes muestras.

b).-Las muestras no son representativas del fenómeno.

c).-Las clasificaciones de las muestras no son válidas.

d).-El área estudiada es no homogénea.

e).-Pequeños o largos conjuntos son necesarios.

f).-Pequeñas o largas distancia deben ser calculadas.

g).-Mas o menos distancias deben ser calculadas.

h).-Pequeñas tolerancias es necesarias.

i).-Las muestras pueden tener localizaciones incorrectas.

j).-Las valores muestreados pueden ser malos.

De acuerdo a la experiencia, el problema fundamental en la obtención de un semivariograma correcto es, la elección adecuada de los intervalos de distancias en los cuales será calculado el semivariograma, de modo que en estos la cantidad de pares encontrados sea suficiente desde el punto de vista estadístico [Armstrong,1984a], [Krajewsky,1993].

Modelado de Semivariogramas

El modelado de semivariogramas incluye dos etapas fundamentales (Xie y Myers, 1995a), una vez construido el semivariograma experimental o empírico es necesario ajustar a este un modelo teórico, con el objetivo de determinar los parámetros descriptivos del semivariograma que posteriormente serán usados en la estimación (ASCE Task, 1990; Journel y Huijbregts, 1978; David, 1977; Lamorey y Jacobsom, 1995; Pannatier, 1993; Arik, 1990; Dubrule, 1994).

Parámetros del Semivariograma

Una gráfica de $\gamma(h)$ contra h se conoce como semivariograma experimental y se usa para obtener el valor de los parámetros que lo definen y darle estructura.

Los parámetros del semivariograma caracterizan tres elementos importantes en la variabilidad de un atributo que son: la discontinuidad en el origen (existencia de efecto de pepita), el valor máximo de variabilidad (meseta), y el área de influencia de la correlación (Rango), (figura 2.1). como se presentan en Krajewski y Gibbs (1993), Journel y Huijbregts (1978), David (1977), Echaabi (1995), Lamorey y Jacobsom (1995), Wallace y Hawkims (1994), Pannatier (1993), Arik (1990), Pitard (1994), y se describen a continuación.

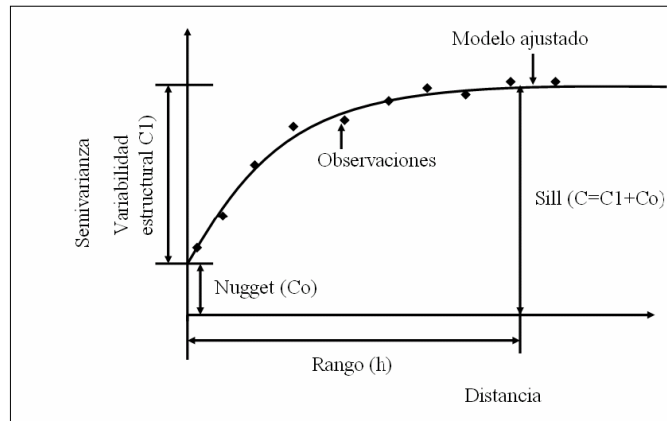


Figura 2.1.- Parámetros del Semivariograma

El Efecto Pepita (Nugget): El semivariograma por definición es nulo en el origen, pero en la práctica las funciones obtenidas pueden presentar discontinuidad en el origen, a esta discontinuidad se le llama **efecto de pepita**, en inglés (Nugget effect). Puede ser obtenido trazando una línea recta entre los primeros puntos del semivariograma empírico y extender ésta hasta que se intercepte con el eje Y. Si esta intersección ocurre por debajo de cero, el valor asumido por este efecto es cero, pues valores negativos de $\gamma(0)$ no tienen significado y no es común. El efecto pepita se representa como C_0 .

La Meseta o Umbral (Sill): Es el valor de $\gamma(h)$ para el cual con el aumento de h su valor permanece constante, se representa como $(C = C_1 + C_0)$ y se denomina meseta. Puede obtenerse trazando una línea paralela a la abscisa y que se ajuste a los puntos de mayor valor del semivariograma y su valor se lee en la intersección de esta línea con la ordenada.

El Alcance (Rango): La distancia h para la cual las variables $Z(x)$ y $Z(x+h)$ son independientes, se denomina alcance y se representa por (h) , es decir, las distancias para la cual los valores de la variable dejan de estar correlacionados, o lo que es lo mismo, la distancia para la cual el semivariograma alcanza su meseta.

El alcance siempre tiene valor positivo y puede ser obtenido a partir de la intersección de las líneas descritas en los puntos anteriores, ese punto leído en la abscisa es una fracción del propio alcance, fracción que se detallara posteriormente en la explicación de los modelos teóricos

Modelos Teóricos de Semivariogramas

Los modelos teóricos de semivariogramas admisible o autorizados más utilizados en la práctica se presentan en Journel y Huijbregts (1978) en los que coinciden Krajewski y Gibbs (1993), Deutsch y Journel (1998), Bacchi y Kottegoda (1995), Wackernagel (1995), Armstrong y Carignan (1997), Myers (1991c), Kiyono y Suzuki (1996). Atendiendo a las dos características más importantes en el modelado de semivariogramas que son según Journel y Huijbregts (1978):

- a).-Su comportamiento en el origen, el cual puede ser lineal, parabólico y con Efecto de Pepita
- b).-La presencia o ausencia de meseta.

Estos modelos son:

1.-Modelo con Efecto de Pepita: Corresponde a un fenómeno puramente aleatorio (ruido blanco), sin correlación entre las muestras, cualquiera sea la distancia que las separe, (figura 2), donde C representa el valor de la meseta.

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_1 + C_0 & h > 0 \end{cases} \quad (29)$$

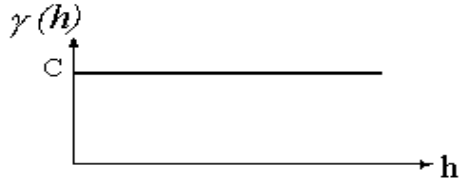


Figura 2.2.- Modelo de Efecto de Pepita

2.-Modelo Con Función Potencia: Este es un modelo sin meseta, su forma se representa en la figura 3, para valores de α correspondientes a 0.5, 1.0 y 1.5. Obviamente cuando el parámetro α es igual a 1 el modelo es lineal y k representa la pendiente de la ecuación de regresión con intercepto cero. Gráficamente se pueden representar así:

$$\gamma(h) = Kh^\alpha \quad 0 < \alpha < 2 \quad (30)$$

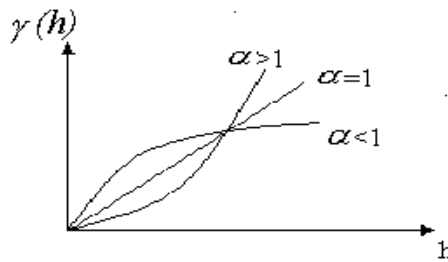


Figura 2.3. Modelo con Función Potencia

3.- Modelo Lineal: Para el valor de $\alpha=1$ en el modelo anterior se obtiene el modelo Lineal, al cual no tiene ni meseta ni Rango. Ahora por efectos prácticos, sin embargo, muchos programas informáticos denotan la pendiente del modelo lineal con la relación C/a (figura 2.4).

$$\gamma(h) = \frac{C}{a} |h| \quad (31)$$

continua aumentando mientras la distancia (h) aumenta, es utilizado para modelar fenómenos que presentan capacidad infinita de dispersión

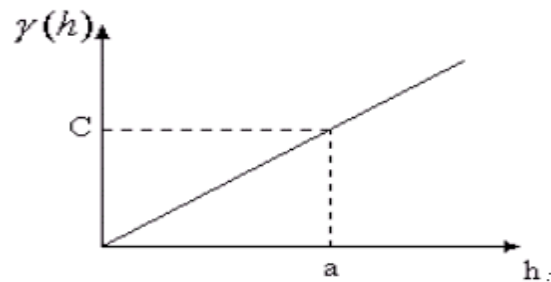


Figura 2.4. Modelo Lineal

4.-Modelo Esférico: Este modelo es probablemente el más utilizado, es una expresión polinomial simple, en su forma representada en la figura 2.5, se puede observar un crecimiento casi lineal y después a cierta distancia finita del origen se alcanza una estabilización, la meseta. La tangente en el origen encuentra a la meseta en el punto de abscisa $(2/3)a$, donde a representa el valor del Rango Su expresión es la siguiente:

$$\gamma(h) = \begin{cases} 0 & |h| = 0 \\ C_0 + C_1 \left[\frac{3}{2} \left(\frac{|h|}{a} \right) - \frac{1}{2} \left(\frac{|h|}{a} \right)^3 \right] & 0 < h < a \\ C_0 + C_1 & |h| > a \end{cases} \quad (32)$$

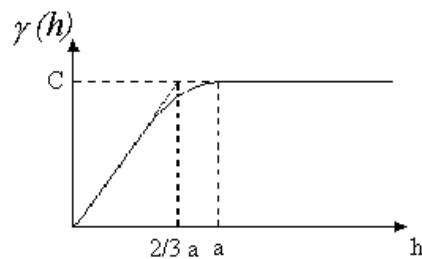


Figura 2.5. Modelo Esférico

5.- Modelo Exponencial: Este modelo a diferencia del esférico crece inicialmente más rápido y después se estabiliza de forma asintótica (figura 2.5).

Como la meseta no se alcanza a una distancia finita, se usa con fines prácticos el “alcance efectivo” o “alcance práctico” a' , valor que se obtiene en el punto de abscisa para el cual el modelo obtiene el 95% de la meseta, con un valor $a'=3a$, donde a es el parámetro de escala. La tangente en el origen encuentra a la meseta en el punto $a=(1/3)a'$. Su ecuación es la siguiente:

$$\gamma(h) = \begin{cases} 0 & |h|=0 \\ C0 + C1 \left[1 - \exp\left(-\frac{|h|}{a}\right) \right] & h \neq 0 \end{cases} \quad (33)$$

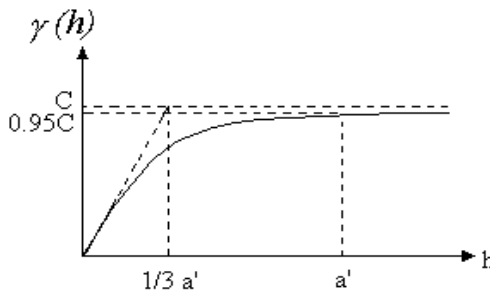


Figura 2.6. Modelo Exponencial

6.-Modelo Gaussiano: Este es un modelo extremadamente continuo (figura 7), inicialmente presenta un comportamiento parabólico en el origen, después al igual que en el modelo Exponencial se alcanza la meseta de forma asintótica. El alcance práctico tiene un valor de $a'=1.73a$, que es el valor de la abscisa donde se alcanza el 95% de la meseta. su expresión matemática es la siguiente:

$$\gamma(h) = \begin{cases} 0 & |h|=0 \\ C0 + C1 \left[1 - \exp\left(-\frac{|h|}{a}\right)^2 \right] & h \neq 0 \end{cases} \quad (34)$$

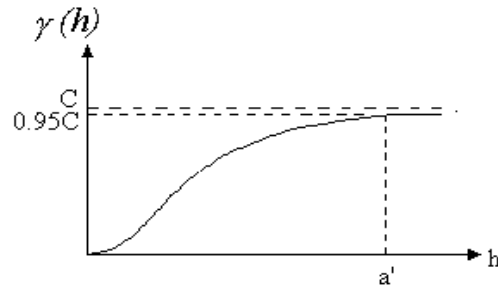


Figura 2.7 Modelo Gaussiano

La estimación de los parámetros de los modelos teóricos descritos puede ser llevada a cabo, entre otros métodos, por máxima verosimilitud (Cressie, 1993) o regresión no lineal (Gotway, 1991). Algunos paquetes de computo geoestadísticos como GS+ (Gamma Design Software, 1999) traen incorporados procedimientos iterativos como el de Gauss-Newton para llevar a cabo la estimación. Otros como GeoEAS (Englund y Sparks, 1988) sólo permiten el ajuste a sentimiento por el método de ensayo y error.

Estimación Kriging.

Inicialmente, Matheron denominó a esta técnica **Krigeage** (en francés) que en ingles se convierte en **Kriging** y en español se escribe **Krigeaje**. Este término que tiene su origen en el apellido de D.G. Krige, reconociendo de esta forma su aporte. El krigeaje es una técnica de estimación que proporciona el Mejor Estimador Lineal Insesgado (BLUE, en ingles, Best Linear Unbiased Estimator):

*Lineal: porque es una combinación lineal ponderada de los datos.

*Insesgado: porque el error de estimación tendrá una media igual a cero.

*mejor: en el sentido del error de varianza minima para un modelo dado de covarianza/variograma.(Schaug et al.,1993; Christensen et al.,1993; Abasov et al., 1990), y que además proporciona un error de estimación conocido como **varianza de krigeaje** que depende del modelo de variograma obtenido y de las localizaciones de los

datos originales (Armstrong y Carignan, 1997; Journel y Huijbregts, 1978; David, 1977; Abasov et al., 1990). Esto brinda la posibilidad de hacer análisis sobre la calidad de las estimaciones (Weerts y Bierkens, 1993; Haas, 1992).

La estimación local busca Estimar o Predecir el valor de la variable regionalizada en un sector del campo estudiado, a partir de los datos disponibles (problema de interpolación espacial). El problema se formula de la siguiente manera: estimar el valor $Z(X_0)$ (desconocido) de la variable regionalizada estudiada en un sitio X_0 a partir de los valores conocidos en sitios con datos. Para ello, se buscará asignar un ponderador a cada dato.

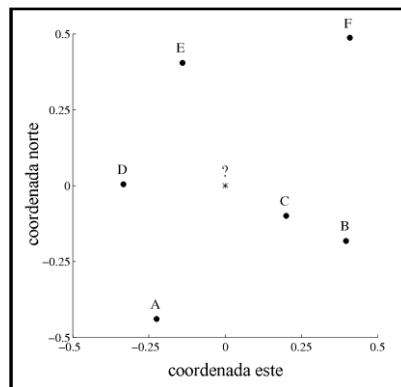


Figura 2.8. ¿Cómo estimar el valor en el sitio “?”

Razones económicas y operativas hacen imposible medir una variable regionalizada en mas que un conjunto limitado de puntos a partir de los cuales se desea calcular el valor de la variable en cualquier otro punto. Esta operación puede llamarse Estimación, predicción o interpolación según el contexto en que se realiza.

A diferencia de otros métodos en los que se intenta describir a la variable mediante una ecuación matemática prefijada, cuyos parámetros se encuentran por algún criterio de minimización, aquí la información es completamente estadística y se obtiene ya sea a través del variograma o de la covarianza.

El objetivo es encontrar un operador lineal que vincule el o los puntos en los que se desea conocer el valor de la variable con los valores medidos.

Se conocen tres tipos de métodos cuyas características se resumen en el siguiente cuadro:

Cuadro 2.1. Tipos de Métodos Kriging

Tipo	Media	Requisito Mínimo	Modelo Estadístico
Kriging Simple	Constante y conocida	Covarianza	Estacionario
Kriging Ordinario	Constante y desconocida	Variograma	Intrínseco
Kriging Universal	Variable y desconocida	Variograma	Kriging Universal

Kriging Simple.

Suponga que hay una variable regionalizada estacionaria con media (m) y covarianza conocidas. De manera análoga a como se define en modelos lineales (por ejemplo en diseño de experimentos) el modelo establecido en este caso es igual a la media más un error aleatorio con media cero. La diferencia es que en este caso los errores no son independientes.

Sea $Z(x)$ la variable de interés medida en el sitio x .

$$\begin{aligned}
 E[Z(x)] &= m \\
 Z(x) &= m + \varepsilon(x), \quad \text{con} \quad E[\varepsilon(x)] = 0
 \end{aligned}
 \tag{35}$$

El Estimador de la variable de interés en un sitio x_0 donde no se tiene información se define como:

$$\hat{Z}(x_0) = m + \varepsilon^*(x_0)
 \tag{36}$$

con $\varepsilon^*(x_0)$ que corresponde a la estimación del error aleatorio en el sitio x_0 . Despejando de la ecuación anterior

$$\varepsilon^*(x_0) = \hat{Z}(x_0) - m \quad (37)$$

El Estimador del error aleatorio se define por:

$$\varepsilon^*(x_0) = \sum_{i=1}^n \lambda_i \varepsilon(x_i) = \sum_{i=1}^n \lambda_i [Z(x_i) - m] \quad (38)$$

de donde el Estimador de la variable de estudio es:

$$\hat{Z}(x_0) = m + \left[\sum_{i=1}^n \lambda_i [Z(x_i) - m] \right] = m + \sum_{i=1}^n \lambda_i \varepsilon(x_i) \quad (39)$$

El Estimador es insesgado si:

$$E \left[\hat{Z}(x_0) \right] = E [Z(x_0)] = m \quad (40)$$

Luego el estimador será insesgado cuando $E[\varepsilon^*(x_0)] = 0$

$$E[\varepsilon^*(x_0)] = \sum_{i=1}^n \lambda_i \varepsilon(x_i) = \sum_{i=1}^n \lambda_i (0) = 0 \quad (41)$$

Por consiguiente, a diferencia del kriging ordinario, en este caso no existen restricciones para las ponderaciones tendientes al cumplimiento de la condición de insesgamiento.

La estimación de los pesos del método kriging ordinario se obtiene de tal forma que se minimice $V[\varepsilon^*(x_0) - \varepsilon(x_0)]$

$$\begin{aligned} V[\varepsilon^*(x_0) - \varepsilon(x_0)] &= E[\varepsilon^*(x_0) - \varepsilon(x_0)]^2 = E \left[\left(\sum_{i=1}^n \lambda_i \varepsilon(x_i) \right) - \varepsilon(x_0) \right]^2 \\ &= E \left[\left(\sum_{i=1}^n \lambda_i \varepsilon(x_i) \right) \left(\sum_{j=1}^n \lambda_j \varepsilon(x_j) \right) - 2E \left[\left(\sum_{i=1}^n \lambda_i \varepsilon(x_i) \right) (\varepsilon(x_0)) \right] + E(\varepsilon(x_0))^2 \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E[\varepsilon(x_i) \varepsilon(x_j)] - 2 \sum_{i=1}^n \lambda_i E[\varepsilon(x_i) \varepsilon(x_0)] + E[\varepsilon(x_0)]^2 \quad (42) \end{aligned}$$

Usando :

$$\text{i.- } E[\varepsilon(x_0)] = 0$$

$$\text{ii.- } E[\varepsilon(x_i)\varepsilon(x_j)] = \text{Cov}[\varepsilon(x_i), \varepsilon(x_j)] = C_{ij}$$

$$\text{iii.- } E[\varepsilon(x_0)]^2 = \sigma^2$$

$$V[\varepsilon^*(x_0) - \varepsilon(x_0)] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 \quad (43)$$

Derivando respecto a λ_1 se tiene:

$$\begin{aligned} \frac{\partial V[\varepsilon^*(x_0) - \varepsilon(x_0)]}{\partial \lambda_1} &= \frac{\partial}{\partial \lambda_1} \left[\lambda_1^2 C_{11} + 2\lambda_1 \sum_{j=2}^n \lambda_j C_{1j} + \sum_{i=2}^n \sum_{j=2}^n \lambda_i \lambda_j C_{ij} - 2\lambda_1 C_{10} - 2 \sum_{i=2}^n \lambda_i C_{i0} + \sigma^2 \right] \\ &= 2\lambda_1 C_{11} + 2 \sum_{j=2}^n \lambda_j C_{1j} - 2C_{10} = 2 \sum_{i=1}^n \lambda_i C_{1i} - 2C_{10} \end{aligned}$$

Igualando a cero, se tiene:

$$\sum_{i=1}^n \lambda_i C_{1i} = C_{10}$$

En general, para cualquier i , $i=1,2,3,\dots,n$, se obtiene :

$$\frac{\partial}{\partial \lambda_i} = \sum_{j=1}^n \lambda_j C_{ij} = C_{i0} \quad (44)$$

Con las n ecuaciones resultantes se construye el siguiente sistema de ecuaciones:

$$\begin{bmatrix} C_{11} & C_{12} & \cdot & \cdot & C_{1n} \\ C_{21} & C_{22} & \cdot & \cdot & C_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ C_{n1} & C_{n2} & \cdot & \cdot & C_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \lambda_{nn} \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ \cdot \\ \cdot \\ C_{n0} \end{bmatrix}$$

Varianza de Estimación Kriging Simple

Se tiene de (43) que:

$$V[\varepsilon^*(x_0) - \varepsilon(x_0)] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2$$

$$\sigma_{KS}^2 = \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2$$

Reemplazando (44) en (43) se tiene:

$$\sigma_{KS}^2 = \sum_{i=1}^n \lambda_i C_{i0} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2$$

$$\sigma_{KS}^2 = \sigma^2 - \sum_{i=1}^n \lambda_i C_{i0} \quad (45)$$

Kriging Ordinario.

Suponga que se hacen mediciones de la variable de interés Z en los puntos x_i , $i=1,2,\dots,n$, de la región de estudio, es decir se tienen realizaciones de las variables $Z(x_1), Z(x_2), \dots, Z(x_n)$, y se desea estimar $Z(x_0)$, en el punto x_0 donde no hubo medición. En esta circunstancia, el método kriging ordinario propone que el valor de la variable puede estimarse como una combinación lineal de las n variables aleatorias así:

$$\hat{Z}(x_0) = \lambda_1 Z(x_1) + \lambda_2 Z(x_2) + \lambda_3 Z(x_3) + \lambda_4 Z(x_4) + \lambda_5 Z(x_5) + \dots + \lambda_n Z(x_n)$$

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (46)$$

en donde los λ_i representan los pesos o ponderaciones de los valores originales. Dichos pesos se calculan en función de la distancia entre los puntos muestreados y el punto donde se va a hacer la correspondiente estimación. La suma de los pesos debe ser igual a uno para que la esperanza del Estimador sea igual a la esperanza de la variable. Esto

último se conoce como el requisito de insesgamiento. Estadísticamente la propiedad de insesgamiento se expresa a través de:

$$E\left[\hat{Z}(x_0)\right] = E[Z(x_0)]$$

Asumiendo que el proceso es estacionario de media m (desconocida) y utilizando las propiedades del valor esperado, se demuestra que la suma de las ponderaciones debe ser igual a uno:

$$\begin{aligned} E\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] &= m \\ \sum_{i=1}^n \lambda_i E[Z(x_i)] &= m \\ \sum_{i=1}^n \lambda_i m &= m \\ m \sum_{i=1}^n \lambda_i &= m \quad \Rightarrow \sum_{i=1}^n \lambda_i = 1 \end{aligned}$$

Se dice que $\hat{Z}(x_0)$ es el mejor estimador, lineal en este caso, porque los pesos se obtienen de tal manera que minimicen la varianza del error de estimación, es decir que minimicen la expresión: $V\left[\hat{Z}(x_0) - Z(x_0)\right]$

Esta última es la característica distintiva de los métodos kriging, ya que existen otros métodos de interpolación como el de distancias inversas o el poligonal, que no garantizan varianza mínima de estimación (Samper y Carrera, 1990). La estimación de

los pesos se obtiene minimizando $V\left[\hat{Z}(x_0) - Z(x_0)\right]$ sujeto a $\sum_{i=1}^n \lambda_i = 1$.

Se tiene que: $V\left[\hat{Z}(x_0) - Z(x_0)\right] = V\left[\hat{Z}(x_0)\right] - 2Cov\left[\hat{Z}(x_0), Z(x_0)\right] + V\left[Z(x_0)\right]$

Separando los componentes de la ecuación anterior se obtiene lo siguiente:

$$V\left[\hat{Z}(x_0)\right] = V\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}\left[Z(x_i), Z(x_j)\right]$$

en adelante se usara la siguiente notación:

$$\text{Cov}\left[Z(x_i), Z(x_j)\right] = C_{ij}$$

$$V\left[Z(x_0)\right] = \sigma^2$$

$$\text{de lo anterior: } \text{Cov}\left[\hat{Z}(x_0), Z(x_0)\right] = \text{Cov}\left[\sum_{i=1}^n \lambda_i Z(x_i), Z(x_0)\right]$$

$$\text{Cov}\left[\hat{Z}(x_0), Z(x_0)\right] = \sum_{i=1}^n \lambda_i \text{Cov}\left[Z(x_i), Z(x_0)\right] = \sum_{i=1}^n \lambda_i C_{i0}$$

entonces, reemplazando se tiene que:

$$V\left[\hat{Z}(x_0) - Z(x_0)\right] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 \quad (47)$$

Luego se debe minimizar la función anterior sujeta a la restricción $\sum_{i=1}^n \lambda_i = 1$

Este problema de minimización con restricciones se resuelve mediante el método de multiplicadores de Lagrange:

$$\sigma_{KO}^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 + 2\mu \left[\sum_{i=1}^n \lambda_i - 1 \right]$$

donde:

$2\mu = \text{Multiplicador de Lagrange}$

$$\sum_{i=1}^n \lambda_i - 1 = 0$$

Siguiendo el procedimiento acostumbrado para obtener valores extremos de una función,

se deriva e iguala a cero, en este caso con respecto a λ_i y μ :

$$\frac{\partial(\sigma_{KO}^2)}{\partial \lambda_1} = \frac{\partial \left[(\lambda_1^2 C_{11} + 2\lambda_1 \sum_{j=2}^n \lambda_j C_{1j} + \sum_{i=2}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij}) - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 + 2\mu (\sum_{i=1}^n \lambda_i - 1) \right]}{\partial \lambda_1}$$

$$\begin{aligned}
&= \left[2\lambda_1 C_{11} + 2 \sum_{j=2}^n \lambda_j C_{1j} \right] - 2C_{10} + 2\mu \\
&\quad \downarrow \\
&= 2 \sum_{j=1}^n \lambda_j C_{1j} - 2C_{10} + 2\mu = 0 \\
&\Rightarrow \sum_{j=1}^n \lambda_j C_{1j} + \mu = C_{10}
\end{aligned} \tag{48}$$

De manera análoga, se determinan las derivadas con respecto a $\lambda_2, \dots, \lambda_n$

$$\frac{\partial(\sigma_{KO}^2)}{\partial \lambda_2} = 2 \sum_{j=1}^n \lambda_j C_{2j} - 2C_{20} + 2\mu = 0 \quad \Rightarrow \sum_{j=1}^n \lambda_j C_{2j} + \mu = C_{20} \tag{49}$$

$$\frac{\partial(\sigma_{KO}^2)}{\partial \lambda_n} = 2 \sum_{j=1}^n \lambda_j C_{nj} - 2C_{n0} + 2\mu = 0 \quad \Rightarrow \sum_{j=1}^n \lambda_j C_{nj} + \mu = C_{n0} \tag{50}$$

Por ultimo derivamos con respecto a μ :

$$\frac{\partial(\sigma_{KO}^2)}{\partial \mu} = 2 \sum_{i=1}^n \lambda_i - 2 = 0 \quad \Rightarrow \sum_{i=1}^n \lambda_i = 1 \tag{51}$$

de (48), (49), (50), (51), resulta un sistema de (n+1) ecuaciones con (n+1) incógnitas que matricialmente puede ser escrito como :

$$\begin{bmatrix} C_{11} & \cdot & \cdot & C_{1n} & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ C_{n1} & \cdot & \cdot & C_{nn} & 1 \\ 1 & \cdot & \cdot & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ \cdot \\ \cdot \\ C_{n0} \\ 1 \end{bmatrix}$$

$$C_{ij} \quad \cdot \quad \lambda = C_{i0}$$

por lo cual los pesos que minimizan el error de Estimación se determinan mediante la función del covariograma a través de:

$$\lambda = C_{ij}^{-1} \cdot C_{i0}$$

Encontrando los pesos se calcula la estimación en el punto x_0

De forma análoga se procede para cada punto donde se quiera hacer estimación

Varianza de Estimación del Kriging Ordinario.

multiplicando (48),(49),(50) por λ_i se obtiene:

$$\lambda_i \left[\sum_{j=1}^n \lambda_j C_{ij} + \mu \right] = \lambda_i C_{i0} \quad \forall i = 1, 2, \dots, n$$

sumando las n ecuaciones:

$$\sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j C_{ij} + \sum_{i=1}^n \lambda_i \mu = \sum_{i=1}^n \lambda_i C_{i0}$$

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} = \sum_{i=1}^n \lambda_i C_{i0} - \sum_{i=1}^n \lambda_i \mu$$

sustituyendo la expresión anterior en (47):

$$\sigma_{KO}^2 = \sigma^2 + \sum_{i=1}^n \lambda_i C_{i0} - \sum_{i=1}^n \lambda_i \mu - 2 \sum_{i=1}^n \lambda_i C_{i0}$$

$$\sigma_{KO}^2 = \sigma^2 - \sum_{i=1}^n \lambda_i C_{i0} - \mu \tag{52}$$

Estimación de Ponderaciones por medio de la Función de Semivarianza

Los pesos λ pueden ser estimados a través de la función de semivarianza, para lo cual se requiere conocer la relación entre las funciones de covariograma y de semivarianza.

Antes de esto es conveniente tener en cuenta la siguiente notación:

$$\sigma^2 = V[Z(x)]$$

$$\gamma_{ij} = \gamma(h)$$

$$C_{ij} = C(h)$$

$h = \text{distancia entre los puntos } i \text{ y } j$

.La relación entre las 2 funciones en cuestión es la siguiente:

$$\begin{aligned}
\gamma_{ij} &= \frac{1}{2} E[(Z(x_j) - Z(x_i))^2] \\
\gamma_{ij} &= \frac{1}{2} E[(Z(x_j))^2 - 2(Z(x_j)Z(x_i)) + (Z(x_i))^2] \\
\gamma_{ij} &= \frac{1}{2} E[(Z(x_j))^2] - E[Z(x_j)Z(x_i)] + \frac{1}{2} E[(Z(x_i))^2] \\
\gamma_{ij} &= \frac{1}{2} [E(Z(x_j))^2 - k^2] + \frac{1}{2} [E(Z(x_i))^2 - k^2] - [E(Z(x_j)Z(x_i)) - k^2] \\
\gamma_{ij} &= \frac{1}{2} [V(Z(x)) + \frac{1}{2} [V(Z(x))] - Cov[Z(x_j)Z(x_i)] \\
\gamma_{ij} &= V[Z(x)] - Cov[Z(x_j)Z(x_i)] \\
\gamma_{ij} &= \sigma^2 - C_{ij} \\
C_{ij} &= \sigma^2 - \gamma_{ij} \tag{53}
\end{aligned}$$

Reemplazando (53) en (48), (49) y (50) se determinan los pesos óptimos λ en términos de la función de semivarianza:

$$\begin{aligned}
\frac{\partial(\sigma_{KO}^2)}{\partial\lambda_1} &= \sum_{j=1}^n \lambda_j C_{1j} + \mu - C_{10} = \sum_{j=1}^n \lambda_j (\sigma^2 - \gamma_{1j}) + \mu - (\sigma^2 - \gamma_{10}) \\
\frac{\partial(\sigma_{KO}^2)}{\partial\lambda_1} &= \sigma^2 \sum_{j=1}^1 \lambda_j - \sum_{j=1}^n \lambda_j \gamma_{1j} + \mu - \sigma^2 + \gamma_{10} \\
\frac{\partial(\sigma_{KO}^2)}{\partial\lambda_1} &= \sigma^2 - \sum_{j=1}^n \lambda_j \gamma_{1j} + \mu - \sigma^2 + \gamma_{10} \\
\Rightarrow \sum_{j=1}^n \lambda_j \gamma_{1j} - \mu &= \gamma_{10}
\end{aligned}$$

Similarmente :

$$\begin{aligned}
\frac{\partial(\sigma_{KO}^2)}{\partial\lambda_2} &= \sum_{j=1}^n \lambda_j \gamma_{2j} - \mu = \gamma_{20} \\
. \\
\frac{\partial(\sigma_{KO}^2)}{\partial\lambda_n} &= \sum_{j=1}^n \lambda_j \gamma_{nj} - \mu = \gamma_{n0}
\end{aligned}$$

El sistema de ecuaciones se completa con (51). De acuerdo con lo anterior, los pesos se obtienen en términos del semivariograma a través del sistema de ecuaciones:

$$\begin{bmatrix} \gamma_{11} & \cdot & \cdot & \gamma_{1n} & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma_{n1} & \cdot & \cdot & \gamma_{nn} & 1 \\ 1 & \cdot & \cdot & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \lambda_n \\ -\mu \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \cdot \\ \cdot \\ \gamma_{n0} \\ 1 \end{bmatrix}$$

Para establecer la expresión de la correspondiente varianza del error de estimación en términos de la función de semivarianza, se reemplaza (53) en (52) y se obtiene:

$$\begin{aligned} \sigma_{KO}^2 &= \sigma^2 - \left[\sum_{i=1}^n \lambda_i (\sigma^2 - \gamma_{ij}) \right] + \mu \\ \sigma_{KO}^2 &= \sigma^2 - \sigma^2 \sum_{i=1}^n \lambda_i + \sum_{i=1}^n \lambda_i \gamma_{ij} + \mu \\ \sigma_{KO}^2 &= \sum_{i=1}^n \lambda_i \gamma_{i0} + \mu \end{aligned} \quad (54)$$

Los pesos de kriging ordinario también pueden ser estimados mediante el uso del correlograma aplicando la siguiente relación: $\rho_{ij} = \frac{C_{ij}}{\sigma^2}$, caso en el que la correspondiente varianza de predicción estaría dada por (Isaaks y Srivastava, 1989):

$$\sigma_{KO}^2 = \sigma^2 \left[1 - \sum_{i=1}^n \lambda_i \gamma_{i0} + \mu \right] \quad (55)$$

Kriging Universal.

En los supuestos hechos hasta ahora respecto a los métodos kriging se ha asumido que la variable regionalizada es estacionaria (al menos se cumple con la hipótesis intrínseca). En muchos casos, la variable no satisface estas condiciones y se caracteriza por exhibir una tendencia. Por ejemplo en hidrología los niveles piezométricos de una acuífero pueden mostrar una pendiente global en la dirección del flujo (Samper y Carrera, 1990). Para tratar este tipo de variables es frecuente

descomponer la variable $Z(x)$ como la suma de la tendencia, tratada como una función determinística, más una componente estocástica estacionaria de media cero. Asuma que:

$$Z(x) = m(x) + \varepsilon(x)$$

con:

$$E[\varepsilon(x)] = 0$$

$$V[\varepsilon(x)] = \sigma^2$$

$$E[Z(x)] = m(x)$$

La tendencia puede expresarse mediante:

$$m(x) = \sum_{l=1}^p a_l f_l(x)$$

donde las funciones $f_l(x)$ son conocidas y p es el número de términos empleados para ajustar $m(x)$

El estimador Kriging Universal se define como:

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \tag{56}$$

este será insesgado si:

$$E\left[\hat{Z}(x_0)\right] = m(x_0)$$

$$E\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] = m(x_0)$$

$$\left[\sum_{i=1}^n \lambda_i m(x_i)\right] = m(x_0)$$

$$\sum_{i=1}^n \lambda_i \left[\sum_{l=1}^p a_l f_l(x_i)\right] = \sum_{l=1}^p a_l f_l(x_0)$$

$$\sum_{l=1}^p a_l \left[\sum_{i=1}^n \lambda_i f_l(x_i)\right] = \sum_{l=1}^p a_l f_l(x_0)$$

$$\Rightarrow \sum_{i=1}^n \lambda_i f_l(x_i) = \sum_{l=1}^p f_l(x_0)$$

La obtención de los pesos en el Kriging Universal, análogo a los otros métodos Kriging se hace de tal forma que la varianza del error de estimación sea

$$\begin{aligned}
 V\left[\hat{Z}(x_0) - Z(x_0)\right] &= E\left[\left(\hat{Z}(x_0) - Z(x_0)\right)^2\right] \\
 V\left[\hat{Z}(x_0) - Z(x_0)\right] &= E\left[\left(\sum_{i=1}^n \lambda_i (m(x_i) - \varepsilon(x_i))\right) - (m(x_0) - \varepsilon(x_0))\right]^2 \\
 V\left[\hat{Z}(x_0) - Z(x_0)\right] &= E\left[\left(\sum_{i=1}^n \lambda_i m(x_i) - m(x_0)\right) + \left(\sum_{i=1}^n \lambda_i \varepsilon(x_i) - \varepsilon(x_0)\right)\right]^2 \\
 V\left[\hat{Z}(x_0) - Z(x_0)\right] &= E\left[\left(\sum_{i=1}^n \lambda_i \varepsilon(x_i) - \varepsilon(x_0)\right)^2\right] \\
 V\left[\hat{Z}(x_0) - Z(x_0)\right] &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E\left[\varepsilon(x_i) \varepsilon(x_j)\right] - 2 \sum_{i=1}^n \lambda_i E\left[\varepsilon(x_i) \varepsilon(x_0)\right] + E\left(\varepsilon(x_0)\right)^2
 \end{aligned}$$

usando:

$$C_{ij} = Cov\left[\varepsilon(x_i), \varepsilon(x_j)\right]$$

$$\sigma^2 = E\left(\varepsilon(x_0)\right)^2$$

se tiene:

minima.
$$V\left[\hat{Z}(x_0) - Z(x_0)\right] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2$$

Luego incluyendo la restricción dada por la condición de insesgamiento se debe minimizar

$$\sigma_{KU}^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 + \sum_{l=1}^n \mu_l \left[\sum_{i=1}^n \lambda_i f_l(x_i) - f_l(x_0) \right]$$

o en términos de la función de semivarianza:

$$\sigma_{KU}^2 = - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma_{ij} + 2 \sum_{i=1}^n \lambda_i \gamma_{i0} + \sum_{l=1}^n \mu_l \left[\sum_{i=1}^n \lambda_i f_l(x_i) - f_l(x_0) \right]$$

derivando la expresión anterior respecto a $\lambda_1, \lambda_2, \dots, \lambda_n$; $\mu_1, \mu_2, \dots, \mu_p$ e igualando a cero,

las correspondientes derivadas se obtienen las siguientes ecuaciones:

$$\sum_{j=1}^n \lambda_j \gamma_{1j} + \sum_{l=1}^p \mu_l f_l(x_i) = \gamma_{i0} \quad i = 1, 2, \dots, n$$

$$\sum_{j=1}^n \lambda_j f_l(x_j) = f_l(x_0) \quad j = 1, 2, \dots, p$$

En términos matriciales:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdot & \cdot & \gamma_{1n} & f_{11} & \cdot & \cdot & f_{p1} \\ \gamma_{21} & \gamma_{22} & \cdot & \cdot & \gamma_{2n} & f_{12} & \cdot & \cdot & f_{p2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma_{n1} & \gamma_{n2} & \cdot & \cdot & \gamma_{nn} & f_{1n} & \cdot & \cdot & f_{pn} \\ f_{11} & f_{12} & \cdot & \cdot & f_{1n} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ f_{p1} & f_{p2} & \cdot & \cdot & f_{pn} & 0 & \cdot & \cdot & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \lambda_n \\ \mu_1 \\ \cdot \\ \cdot \\ \mu_n \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \cdot \\ \cdot \\ \gamma_{n0} \\ f_{10} \\ \cdot \\ \cdot \\ f_{p0} \end{bmatrix}$$

donde $f_{ij} = f_l(x_j)$ es la l -ésima función en el punto j -ésimo

La varianza de estimación del Kriging Universal esta dado por (Samper y Carrera, 1990):

$$\sigma_{KU}^2 = \sum_{i=1}^n \lambda_i \gamma_{i0} + \sum_{l=1}^p \mu_l f_l(x_0) \quad (57).$$

Nótese que si $p=1$ y $f_l(x)=1$, el sistema de ecuaciones del Kriging Universal y la varianza de estimación coinciden con el de Kriging Ordinario

En este orden de ideas puede decirse que el Kriging Ordinario es un caso particular del Kriging Universal.

Validación Cruzada.

Sea $Z(x)$ una función aleatoria estacionaria con semivariograma $\gamma(h)$, su función de covarianza $C(h)$ viene dada por $C(h) = \sigma^2 - \gamma(h)$ donde σ^2 es la varianza de $Z(x)$.

Sea $Z_{x_1}, Z_{x_2}, \dots, Z_{x_n}$ los valores de $Z(x)$ en n puntos medidos. La validación cruzada consiste en suprimir el i -ésimo valor medido $Z(x_i)$ y estimarlo a partir del resto de los datos. El valor estimado $\hat{Z}(x_0)$ se calcula por kriging, procedimiento explicado anteriormente. Si se repite este proceso para los n puntos, se pueden calcular n errores de validación:

$$e(x_i) = \hat{Z}(x_i) - Z(x_i) \quad i = 1, 2, \dots, n$$

Así se van probando diferentes valores de los parámetros del semivariograma hasta que los errores de validación cumplen los siguientes criterios estadísticos: (Journel y Huijbregts, 1978; David, 1977; Armstrong y Carignan, 1997).

Se puede procurar satisfacer lo mejor posible los siguientes criterios estadísticos;

1).- La media de los errores mide el Sesgo del estimador, y debe ser cercana a cero.

De hecho, este criterio resulta secundario con respecto al modelamiento del variograma y la elección de la vecindad de kriging pues los valores obtenidos por kriging son insesgados por construcción.

Así, en la práctica, la media de los errores y la media de los errores estandarizados tenderán hacia cero.

$$E.M. = \frac{1}{n} \sum_{i=1}^n \left[\hat{Z}(x_i) - Z(x_i) \right] \cong 0 \quad (58)$$

$$E.M.E. = \frac{1}{n} \sum_{i=1}^n \left[\frac{\hat{Z}(x_i) - Z(x_i)}{\sigma(x_i)} \right] \cong 0 \quad (59)$$

2).- La varianza de los errores, que mide la Precisión del estimador, debe ser mínima.

$$C.M.E. = \frac{1}{n} \sum_{i=1}^n \left[\hat{Z}(x_i) - Z(x_i) \right]^2 \cong \text{minima} \quad (60)$$

3).- El error estandar de estimación debe ser cercano a cero

$$E.E. = \frac{\sqrt{C.M.E.}}{n} \cong 0 \quad (61)$$

4).-El Porcentaje del error debe ser menor o igual al 10%

$$P.E. = \frac{\sqrt{\frac{\sum_{i=1}^n \left(\hat{Z}(x_i) - Z(x_i) \right)^2}{n-1}}}{\frac{\sum_{i=1}^n Z(x_i)}{n}} \leq 10\% \quad (62)$$

5).-El coeficiente de correlación entre los valores estimados y los valores de los datos debe ser lo más cercano posible a 1.

$$\rho = \left[\hat{Z}(x_i), Z(x_i) \right] \cong 1 \quad (63)$$

6).-El coeficiente de efectividad de predicción (E) el cual es un estimado de que tan efectiva fue la predicción en comparación con el uso solamente del promedio general de los datos (Schloeder et al 2001, Gotmay et al 1996). Este parámetro se calcula con la formula siguiente:

$$E = \left[1 - \frac{\left[\frac{\sum_{i=1}^n \left[\hat{Z}(x_i) - Z(x_i) \right]^2}{n} \right]}{\left[\frac{\sum_{i=1}^n \left[Z(x_i) - \bar{Z} \right]^2}{n} \right]} \right] \times 100 \quad (64)$$

donde \bar{Z} es el promedio de las estimaciones. Cuando E es igual al 100% indica una predicción perfecta y cuando el valor es negativo indica que hubiese sido mejor usar la media general que la predicción. De acuerdo con Kravchenko (2003), si la estructura espacial es pobre y no es viable hacer un muestreo intensivo para crear un buen mapa, es mejor trabajar con la media general de los datos.

Una técnica similar a la validación cruzada es el **Jack-knife**, el cual no considera una reposición de los datos: se divide los datos en dos grupos y se estima los datos de un grupo a partir de los datos del otro grupo.

3.-MATERIALES Y METODOS

Ejemplo de Aplicación.

Se aplicara la metodología Kriging para realizar estimaciones del valor de la variable temperatura media anual de la región sureste del estado de Coahuila en estaciones climatológicas irregularmente espaciadas.

Localización.

El estudio se desarrollo en la región sureste del estado de Coahuila la cual esta conformada por 5 municipios los cuales son: Saltillo, Ramos Arizpe, Arteaga, General Cepeda y Parras de la Fuente. Saltillo, capital del estado de Coahuila, limita al norte con los municipios de Ramos Arizpe, Arteaga y General Cepeda; al este con el municipio de Arteaga y el estado de Nuevo León; al sur con los estados de Nuevo León y Zacatecas, al oeste con el estado de Zacatecas y los municipios de Parras de la Fuente y General Cepeda. Cuenta con una superficie de 6,837 kilómetros cuadrados, que representan el 4.51% del total de la superficie del estado. Al oeste se localiza la sierra Playa Madero, que abarca también la parte del sureste de Parras de la Fuente. En el suroeste se localiza la sierra El Laurel, que forma parte también del citado municipio. La sierra de Zapalinamé se levanta al este, y la sierra Hermosa está localizada en el suroeste. Al sur se encuentran la presa de San Pedro y la de los Muchachos.

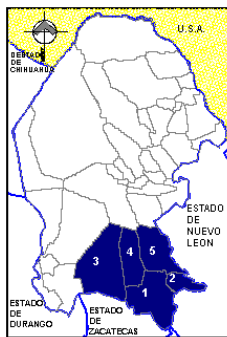


Figura 3.1.-Región Sureste del estado de Coahuila

Clima.

En la Región Sureste el clima es caluroso en primavera y verano, la estación lluviosa es en julio y agosto, en invierno el tiempo es frío y brumoso. El estado de Coahuila está situado, en su mayor parte, en el oriente de una gran área climática denominada como Desierto de Chihuahua, o Desierto del Norte de México. Se caracteriza por poseer climas continentales, secos y muy secos, que van desde los semicálidos, predominantes en los bolsones coahuilenses, hasta los templados de las partes más altas y las más septentrionales. Tres son las áreas en las que puede dividirse al estado por sus climas, de un modo general: El occidente muy seco; el centro y sur, en los que se asocian climas desde los muy secos y secos semicálidos de sus bolsones y valles hasta los semisecos templados y los templados subhúmedos de las cumbres serranas, con predominancia de climas secos y por último el noreste semiseco y seco con influencia marítima más notoria.

En el estado de Coahuila, la mitad de su territorio (49%) presenta clima seco y semi-seco, el 46% tiene clima muy seco y el 5% restante registra clima templado Subhúmedo, localizado en las partes altas de las sierras del sur: San Antonio y Tampiquillo.

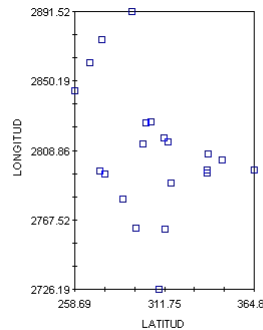
La temperatura media anual oscila entre los 18⁰ a 22⁰ C. La temperatura mas alta, mayor de 30⁰ C., se presenta en los meses de Mayo a Agosto y la mas baja en Enero que es

alrededor de 4⁰ C. Las lluvias son muy escasas, se presentan durante el verano, la precipitación total anual es alrededor de 400 mm.

Recopilación de información.

Para este trabajo se seleccionaron los datos disponibles de temperatura media anual del período 1971-2000 de las estaciones climatológicas de la región sureste del estado de Coahuila del Servicio Meteorológico Nacional-Comisión Nacional del Agua (Normales Climatológicas).

Cuadro 3.1.- Distribución espacial de las Estaciones Climatológicas



ESTACION	MUNICIPIO	LATITUD	LONGITUD	TMA
Alto de Norias	Ramos Arizpe	25 58 37	101 14 56	18.7
Arteaga	Arteaga	25 26 0	100 51 0	17.4
Purísima	Arteaga	25 20 17	100 31 49	15.2
La Sauceda	Ramos Arizpe	25 51 0	101 19 0	19.5
Carneros	Saltillo	25 7 9	101 6 40	15.5
Derramadero	Saltillo	25 16 0	101 15 0	18.9
Ejido Reata	Ramos Arizpe	26 7 41	101 4 35	20.5
El Tunal	Arteaga	25 27 4	100 52 38	13.9
Gómez Farias	Saltillo	24 57 50	101 1 49	17.5
Hipólito	Ramos Arizpe	25 41 54	101 24 17	20
Huachichil	Arteaga	25 12 29	100 49 47	14.6
Jame	Arteaga	25 22 0	100 37 0	12.4
La Ventura	Saltillo	24 38 17	100 53 23	18.5
Las Hormigas	Saltillo	24 57 37	100 51 38	16
Potrero	Arteaga	25 17 4	100 20 34	16.9
R. Arizpe(DGE)	Ramos Arizpe	25 32 21	100 57 3	18.1
R. Arizpe(SMN)	Ramos Arizpe	25 32 0	100 59 0	18.2
Alazanas(DGE)	Arteaga	25 16 0	100 37 0	14
Alazanas(SMN)	Arteaga	25 17 0	100 37 0	16.9
Saltillo(DGE)	Saltillo	25 25 0	101 0 0	18.2
Vaquería	Saltillo	25 15 17	101 13 12	17

ANALISIS DE LA INFORMACION.

Análisis Estadístico.

Aquí se trata de caracterizar a la muestra, de forma tal que se obtenga la mayor información posible a partir de los datos que se disponen.

Esta etapa es fundamental para que el análisis geoestadístico sea válido, puesto que el mismo está basado en la estimación y modelación de una función que refleja la correlación espacial de la propiedad que se estudia: el semivariograma. Para poder estimarla es necesario que la muestra cumpla una serie de requisitos, como son:

- a.-Que su distribución de probabilidad sea normal,
- b.-Que no exista tendencia, es decir, que sea estacionaria al menos la media,
- c.-Que no se vea afectada por valores atípicos (outliers) tanto distribucionales como espaciales,
- d.-Que tenga una distribución espacial homogénea.

Tomando en cuenta lo anteriormente planteado, de lo que se trata es de explorar las características de la muestra con el fin de modificar en la medida de lo posible aquellas que no satisfagan los requisitos exigidos. Además nos permite decidir cual procedimiento sería más adecuado aplicar en la estimación espacial.

Mediante el análisis univariado se obtendrán los estadísticos básicos: media, varianza, desviación Estándar, coeficiente de variación, kurtosis, asimetría, valores máximos y mínimos, cuartiles (0.25, 0.75), rango, tamaño, moda, mediana, Error Estándar

Análisis de Normalidad

Muchos métodos estadísticos se basan en la hipótesis de normalidad de la variable objeto de estudio. De hecho, si la falta de normalidad de la variable es

suficientemente fuerte, muchos de los contrastes utilizados en los análisis estadístico-inferenciales no son válidos. Incluso aunque las muestras grandes tiendan a disminuir los efectos perniciosos de la no normalidad, el investigador debería evaluar la normalidad de todas las variables incluidas en el análisis.

Existen varios métodos para evaluar la normalidad de un conjunto de datos que pueden dividirse en dos grupos: los Contrastes de Hipótesis y los Métodos Gráficos.

Contrastes de Hipótesis

La primera de las formas para comprobar la normalidad de una distribución se efectúa a través de un *contraste de hipótesis*. No existe un contraste óptimo para probar la hipótesis de normalidad. La razón es que la potencia relativa depende del tamaño muestral y de la verdadera distribución que genera los datos. Desde un punto de vista poco riguroso, el contraste de Shapiro y Wilks es, en términos generales, el más conveniente en muestras pequeñas ($n < 30$), mientras que el contraste de Kolmogorov-Smirnov, en la versión modificada de Lilliefors es adecuado para muestras grandes.

En el *test de Kolmogorov-Smirnov* la hipótesis nula que se pone a prueba es que los datos proceden de una población con distribución normal frente a una alternativa de que no es así. Este contraste calcula la distancia máxima entre la función de distribución empírica de la muestra y la teórica. Si la distancia calculada es mayor que la encontrada en las tablas, fijado un nivel de significación, se rechaza el modelo normal.

Otros contrastes muy utilizados son los **tests de asimetría y curtosis** cuyos estadísticos muestrales vienen dados por:

$$Z_{asimetria} = \frac{Asimetria}{\sqrt{\frac{6}{N}}} \quad Z_{Curtosis} = \frac{Curtosis}{\sqrt{\frac{24}{N}}} \quad (65)$$

donde N es el tamaño muestral. Si es cierta la hipótesis de normalidad ambos se distribuyen asintóticamente según una $N(0,1)$.

Prueba de Bondad de Ajuste de Kolmogorov-Smirnov (KS).

Hipótesis a contrastar:

H_0 : Los datos analizados siguen una distribución Normal.

H_a : Los datos analizados no siguen una distribución Normal.

Estadístico de contraste:

$$D = \sup_{1 \leq i \leq n} \left| \hat{F}_n(x_i) - F_0(x_i) \right| \quad (66)$$

donde:

x_i es el i -ésimo valor observado en la muestra (cuyos valores se han ordenado previamente de menor a mayor).

$\hat{F}_n(x_i)$ es un estimador de la probabilidad de observar valores menores o iguales que x_i
 $F_0(x_i)$ es la probabilidad de observar valores menores o iguales que x_i cuando H_0 es cierta. Así pues, D es la mayor diferencia absoluta observada entre la frecuencia acumulada observada $\hat{F}_n(x_i)$ y la frecuencia acumulada teórica $F_0(x_i)$ obtenida a partir de la distribución de probabilidad que se especifica como hipótesis nula.

Si los valores observados $\hat{F}_n(x_i)$ son similares a los esperados $F_0(x_i)$, el valor de D será pequeño. Cuanto mayor sea la discrepancia entre la distribución empírica $\hat{F}_n(x_i)$ y la distribución teórica, mayor será el valor de D .

Por tanto, el criterio para la toma de la decisión entre las dos hipótesis será de la forma:

$$\text{Si } D \leq D_\alpha \Rightarrow \text{Aceptar } H_0 \qquad \text{Si } D > D_\alpha \Rightarrow \text{Rechazar } H_0$$

donde el valor D_α se elige de tal manera que:

$P(\text{Rechazar } H_0 / H_0 \text{ es cierta}) = P(D > D_\alpha / \text{Los datos siguen una distribución Normal}) = \alpha$

Siendo α el nivel de significancia del contraste.

Para el cálculo práctico del estadístico D deben obtenerse:

$$D^+ = \max_{1 \leq i < n} \left\{ \frac{i}{n} - F_0(x_i) \right\} \quad D^- = \max_{1 \leq i < n} \left\{ F_0(x_i) - \frac{i-1}{n} \right\} \quad (67)$$

y a partir de estos valores .

$$D = \max \{ D^+, D^- \} \quad (68)$$

A su vez, el valor de D_α depende del tipo de distribución a probar y se encuentra tabulado. En general es de la forma:

$$D_\alpha = \frac{C_\alpha}{k(n)} \quad (69)$$

Donde C_α y $k(n)$ se encuentran en las tablas siguientes:

Cuadro 3.2.- Modelo y Valores de α

Cα	A		
	0.1	0.05	0.01
General	1.224	1.358	1.628
Normal	0.819	0.895	1.035
Exponencial	0.99	1.094	1.308
Weibull n=10	0.76	0.819	0.944
Weibull n=20	0.779	0.843	0.973
Weibull n=50	0.79	0.856	0.988
Weibull n= ∞	0.803	0.874	1.007

Cuadro 3.3.- Ecuaciones de $K(n)$

DISTRIBUCION QUE SE CONTRASTA	K(n)
General. Parámetros Conocidos	$k(n) = \sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}$
Normal	$k(n) = \sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}}$
Exponencial	$k(n) = \sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}$
Weibull	$k(n) = \sqrt{n}$

Modo alternativo de realizar la prueba de Kolmogorov Smirnov.

La toma de la decisión en el contraste anterior puede llevarse a cabo también mediante el empleo del p-valor asociado al estadístico D observado. El p-valor se define como:

$$\text{P-valor} = P(D > D_{\text{abs}} / H_0 \text{ es cierta})$$

Si el p-valor es grande significa que, siendo cierta la hipótesis nula, el valor observado del estadístico D era esperable. Por tanto no hay razón para rechazar dicha hipótesis. Asimismo, si el p-valor fuera pequeño, ello indicaría que, siendo cierta la hipótesis nula, era muy difícil que se produjera el valor de D que efectivamente se ha observado. Ello obliga a poner muy en duda, y por tanto a rechazar, la hipótesis nula. De esta forma, para un nivel de significación α , la regla de decisión para este contraste es:

$$\text{Si p-valor} \geq \alpha \Rightarrow \text{Aceptar } H_0$$

$$\text{Si p-valor} < \alpha \Rightarrow \text{Rechazar } H_0$$

Obviamente, la obtención del p-valor requiere conocer la distribución de D bajo la hipótesis nula y hacer el cálculo correspondiente.

En el caso particular de la prueba de Kolmogorov Smirnov, la mayoría de los paquetes de software estadístico realizan este cálculo y proporcionan el p-valor directamente.

Métodos gráficos

El método gráfico univariante más simple para diagnosticar la normalidad es una comprobación visual del *histograma* que compare los valores de los datos observados con una distribución normal. Aunque atractivo por su simplicidad, este método es problemático para muestras pequeñas, donde la construcción del histograma puede distorsionar la representación visual de tal forma que el análisis sea poco fiable. Otras

posibilidades, también basadas en información gráfica, consisten en realizar *diagramas de cuantiles* (Q-Q plots).

Los diagramas de cuantiles comparan en un sistema de coordenadas cartesianas, los cuantiles muestrales (eje X) con los cuantiles esperados bajo la hipótesis normalidad. Si la distribución de partida es normal dichos diagramas tenderán a ser rectas que pasan por el origen. Cuanto más se desvíen de una recta, menos normales serán los datos.

Datos Atípicos (Outliers)

Los casos atípicos son observaciones con características diferentes de las demás. Este tipo de casos no pueden ser caracterizados categóricamente como benéficos o problemáticos sino que deben ser contemplados en el contexto del análisis y debe evaluarse el tipo de información que pueden proporcionar. Su principal problema radica en que son elementos que pueden no ser representativos de la población pudiendo distorsionar seriamente el comportamiento de los contrastes estadísticos.

Por otra parte, aunque diferentes a la mayor parte de la muestra, pueden ser indicativos de las características de un segmento válido de la población y, por consiguiente, una señal de la falta de representatividad de la muestra.

Tipos de Outliers

Los casos atípicos pueden clasificarse en 4 categorías:

La primera categoría contiene aquellos casos atípicos que surgen de un error de procedimiento, tales como la entrada de datos o un error de codificación. Estos casos atípicos deberían subsanarse en el filtrado de los datos, y si no se puede, deberían eliminarse del análisis o recodificarse como datos ausentes.

La segunda clase es la observación que ocurre como consecuencia de un acontecimiento extraordinario. En este caso, el outlier no representa ningún segmento válido de la población y puede ser eliminado del análisis.

La tercera clase contiene las observaciones cuyos valores caen dentro del rango de las variables observadas pero que son únicas en la combinación de los valores de dichas variables. Estas observaciones deberían ser retenidas en el análisis pero estudiando qué influencia ejercen en los procesos de estimación de los modelos considerados.

La cuarta y última clase comprende las observaciones extraordinarias para las que el investigador no tiene explicación. En estos casos lo mejor que se puede hacer es replicar el análisis con y sin dichas observaciones con el fin de analizar su influencia sobre los resultados. Si dichas observaciones son influyentes el analista debería reportarlo en sus conclusiones y debería averiguar el por qué de dichas observaciones.

Identificación de Outliers

Los casos atípicos pueden identificarse desde una perspectiva univariante o multivariante.

La perspectiva univariante examina la distribución de observaciones para cada variable, seleccionando como casos atípicos aquellos casos cuyos valores caigan fuera de los rangos de la distribución. La cuestión principal consiste en el establecimiento de un umbral para la designación de caso atípico. Esto se puede hacer gráficamente mediante histogramas o diagramas de caja o bien numéricamente, mediante el cálculo de puntuaciones tipificadas. Para muestras pequeñas (de 80 o incluso menos observaciones), las pautas sugeridas identifican como atípicos aquellos casos con valores estándar de 2.5 o superiores. Cuando los tamaños muestrales son mayores, las pautas sugieren que el valor umbral sea 3.

Los paquetes estadísticos son conjuntos de programas que implementan diversas técnicas estadísticas en un entorno común. Algunos de los más utilizados son SAS, BMDP, SPSS, SYSTAT, STATISTICA, STATA y últimamente MINITAB, S-PLUS, EViews, STATGRAPHICS y MATLAB.

Para los análisis Estadísticos se utilizara el Programa Statistica Versión 9 para Windows.

Análisis Geoestadístico.

Para llevar a cabo los análisis geoestadísticos se dispuso del programa GS+ versión 5.1 (Geostatistics for the Environmental Sciences). La base de datos disponible consta de las coordenadas UTM (Latitud., Longitud) de las 21 estaciones Climatológicas de la región sureste del Estado de Coahuila. y de la Variable Temperatura media Anual (Z).

Estimación de la función de correlación espacial

En el sentido más amplio se debe estimar una función que describa el grado de correlación espacial de la propiedad que se estudia. En dependencia de las características de la muestra, ésta puede ser la función de covarianzas, si la propiedad es estacionaria de segundo orden (sus estadígrafos de 1er y 2do orden no dependen de la posición), o la función de semivarianzas, cuando es intrínseca (las diferencias son estacionarias de segundo orden).

Usualmente se trata de estimar el semivariograma, conocido simplemente como variograma, puesto que supone que se cumpla la hipótesis más débil o dicho de otro modo exige menos requisitos a la propiedad bajo estudio.

Consideraciones para el cómputo del variograma

- Los pares de las observaciones se agrupan según la distancia dentro de un intervalo con una tolerancia y dentro de una dirección con una tolerancia angular.
- Se estima para valores menores que la mitad de la distancia máxima.
- Se considera que un número máximo de 25 intervalos es suficiente para cualquier propósito, y un mínimo de 10 debe ser usado para determinar con precisión el rango y la meseta del variograma.
- Se considera que debe haber entre 30 y 50 pares de puntos como mínimo por intervalo.
- Los valores estimados para cada intervalo se deben graficar contra la distancia promedio de todos los pares que se encuentran dentro de dicho intervalo.

Modelación de la función de correlación espacial

La modelación del variograma consiste en buscar una función analítica que represente adecuadamente los valores estimados del variograma. Esta función no puede ser cualquiera y de hecho existen un número reducido de modelos teóricos que pueden ser usados. Entre los modelos más conocidos se encuentran: el esférico, el gaussiano, el exponencial, el lineal con meseta y sin meseta. Estos modelos se caracterizan fundamentalmente por tres parámetros: el alcance o radio de correlación espacial, la meseta (sill) y la varianza pepita (nugget). Entonces el proceso de modelación se reduce a determinar cuál modelo y con que parámetros se ajusta mejor a los valores estimados del variograma.

Estimación espacial de la propiedad estudiada

Kriging Ordinario

Si no existe tendencia significativa de manera que se puede considerar que la variable es estacionaria de segundo orden o al menos que sea intrínseca, es decir que las diferencias sean estacionarias de segundo orden, entonces se puede estimar la propiedad mediante Kriging Ordinario. Es el método más apropiado, para situaciones medioambientales, este método asume que las medias locales, no están necesariamente relacionadas lo más cercanamente a la media poblacional, por lo cual solo usa las muestras en la vecindad local para la estimación.

Kriging en presencia de tendencia

En el caso de existencia de tendencia, esta se puede tener en cuenta al estimar la propiedad si se aplica el método de Kriging Universal. Pero este método tiene el inconveniente que requiere del conocimiento de la función de semivarianzas y de la forma de la tendencia lo cual no es posible conocer *a priori* y en general es un método de prueba y error algorítmicamente complicado. Es preferible en tal caso usar el método conocido como Kriging Residual, el cual consiste en determinar de manera iterativa la tendencia como una función polinomial y luego se aplica el método de Kriging ordinario a los residuos obtenidos que deben ser estacionarios. De manera que el valor estimado en un punto sería la suma de la tendencia más la estimación por Kriging Ordinario en dicho punto.

En ambos casos se debe tener en cuenta el tamaño de las vecindades y el número de puntos vecinos que intervienen en la estimación. Normalmente el tamaño de la vecindad se elige cercano al alcance o radio de correlación y se considera 16 puntos vecinos como una cantidad razonable para una buena estimación.

Validación del modelo de variograma

Se realiza un análisis integral de los estadígrafos de los errores y se aplican de manera combinada los siguientes criterios:

- El valor medio de los errores (Error Medio =E.M.) y de los errores estandarizados (Error medio Estandarizado=E.M.E.) deben ser cercanos a cero.
- La varianza de los errores (Cuadrado Medio del Error = C.M.E.) debe ser la mas baja posible.
- El CME se usara para calcular el error estándar o de estimación (EE) el cual deberá ser cercano a cero
- El Porcentaje de Error (P.E.) deber ser menor o igual al 10%.
- La correlación muestral entre el valor real y el estimado debe ser próxima a la unidad,
- El Coeficiente de efectividad de Predicción (E) debe ser cercano al 100%

Inferencias Estadísticas del Error.

Bajo la asunción de distribución normal de los errores $e(x_i)$, una prueba “t” de Student puede ser realizada sobre el error medio, m , para probar la significancia de la asunción $m=0$. Para la varianza de los errores es esperado que sea igual a 1, y una prueba de χ^2 puede ser realizada. Se plantean las siguientes hipótesis:

$$H_0: m=0$$

$$H_a: m \neq 0$$

Donde m corresponde a la media de los errores, de tal forma que:

$$t_c = \frac{\sqrt{n}(\bar{X} - m)}{\sigma} \tag{70}$$

Siendo n el número total de errores, \bar{X} la media y σ la desviación estándar de la serie de los errores. Bajo la teoría de normalidad, t_c se distribuye como una t-student con $n-1$ grados de libertad (g.l.) si y solo si $H_0: m=0$ es cierta. La regla de decisión es: Rechazar $H_0: m=0$ si $|t_c| \geq t_{\alpha/2, (n-1)g.l.}$ (Ostle, 1983). α será el nivel de significancia.

Un intervalo de confianza para m , basado sobre la distribución t-student con una confianza del 95% (Snedecor y Cochran, 1967) es:

$$\bar{X} - t_{0.05, (n-1)g.l.} \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + t_{0.05, (n-1)g.l.} \frac{\sigma}{\sqrt{n}} \quad (71)$$

De la misma forma, bajo asunción de normalidad, se plantean las hipótesis:

$$H_0: \sigma^2 = 1$$

$$H_a: \sigma^2 \neq 1$$

Donde σ^2 corresponde a la varianza de los errores que bajo la teoría de normalidad:

$$\chi_c^2 = \frac{(n-1)\sigma}{\sigma^2} \quad (72)$$

De tal manera que χ_c^2 se distribuirá como una χ_c^2 con $(n-1)g.l.$ si y solo si $H_0: \sigma^2=1$ es cierta, la regla de decisión es: se rechaza $H_0: \sigma^2 = 1$ si $\chi_c^2 \geq \chi_{\alpha/2, (n-1)g.l.}^2$ o si

$$\chi_c^2 \leq \chi_{1-\alpha/2, (n-1)g.l.}^2 \quad (\text{Ostle, 1983}).$$

Bajo la teoría de normalidad, un intervalo de confianza para σ^2 con un 95% de confiabilidad (Snedecor y Cochran, 1967) es:

$$\frac{\sum_{i=1}^n [e(x_i)]^2}{\chi_{(0.025), (n-1)g.l.}^2} < \sigma^2 < \frac{\sum_{i=1}^n [e(x_i)]^2}{\chi_{(0.975), (n-1)g.l.}^2} \quad (73)$$

4.-RESULTADOS Y DISCUSION

Análisis Estadístico.

Parámetros Estadísticos.

Mediante el análisis univariado se obtuvieron los estadísticos básicos de la muestra de 21 estaciones Climatologicas de la región sureste del Estado de Coahuila y la variable a evaluar es la Temperatura Media Anual. Estos estadísticos se obtuvieron mediante el programa STATISTICA versión 9 para Windows.

Cuadro 4.1 Parámetros Estadísticos de la Temperatura Media Anual
De la región sureste de Coahuila

Tamaño Muestral	21
Media	17.042
Mediana	17.4
Moda	Bimodal(16.9)
Mínimo	12.4
Máximo	20.5
Cuartil Inferior	15.5
Cuartil Superior	18.5
Rango	8.1
Varianza	4.637
Desviación Estándar	2.153
Coficiente de Variación	12.635
Error Estándar	0.469
Asimetría	-0.467
Curtosis	-0.410

Se observa que los datos tienen Asimetría Negativa (<0) es decir, la mayor concentración de valores se encuentran a la izquierda de la media

En lo referente a la Curtosis (<3) la distribución es Platicurtica.

Prueba de Normalidad.

En el test de Kolmogorov-Smirnov, en la versión modificada de Lilliefors es adecuado para muestras grandes. La hipótesis nula que se pone a prueba es que los datos proceden de una población con distribución normal frente a una alternativa de que no es así.

Hipótesis a contrastar:

H₀: Los datos analizados siguen una distribución Normal.

H₁: Los datos analizados no siguen una distribución Normal.

$$D_{\alpha} = \frac{C_{\alpha}}{k(n)}$$

C_α = 0.895 (Tabla 3.2)

$$k(n) = \sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}} \text{ (Tabla 3.3.)}$$

Sustituyendo:

$$D_{\alpha} = \frac{0.895}{\sqrt{21} - 0.01 + \frac{0.85}{\sqrt{21}}} = \frac{0.895}{4.758} = 0.188$$

El Valor Calculado de D por el programa Statistica proporciona un Valor de

D= 0.14022

Regla de Decisión:

Por lo tanto, el criterio para la toma de la decisión entre las dos hipótesis será de la forma:

Si $D \leq D_{\alpha} \Rightarrow$ Aceptar H₀

Si $D > D_{\alpha} \Rightarrow$ Rechazar H₀

Como el valor $D = 0.14022 < 0.188$, se Acepta H₀ y se Concluye que los datos se distribuyen Normalmente. Gráficamente:

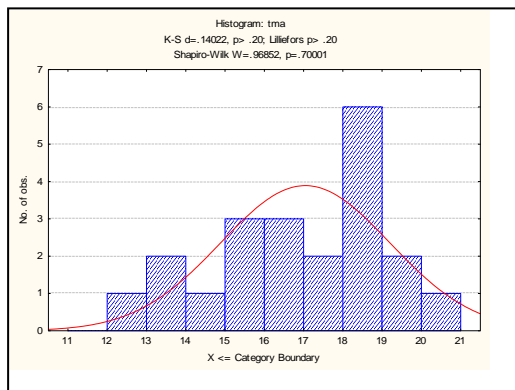


Figura 4.1. Histograma con distribución Normal de la Temperatura Media Anual Con prueba de Kolmogorov-Smirnov

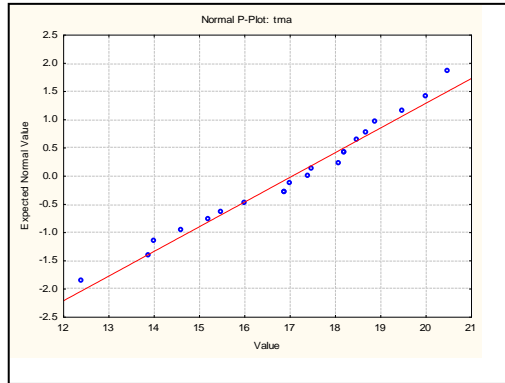


Figura 4.2 Diagrama P-P de la Temperatura Media Anual

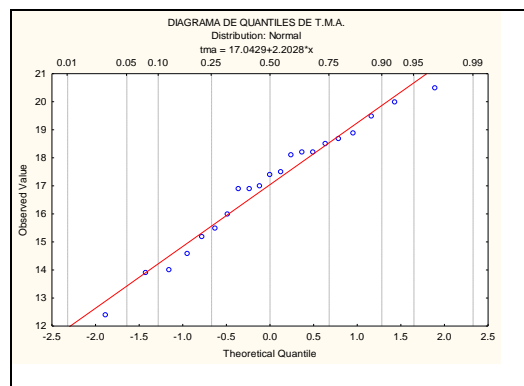


Figura 4.3 Diagrama Q-Q de la Temperatura Media Anual

En la figura 4.2 y 4.3 se puede notar la proximidad de los valores observados a la recta lo cual indica que los datos se pueden considerar con distribución normal.

Identificación de outliers.

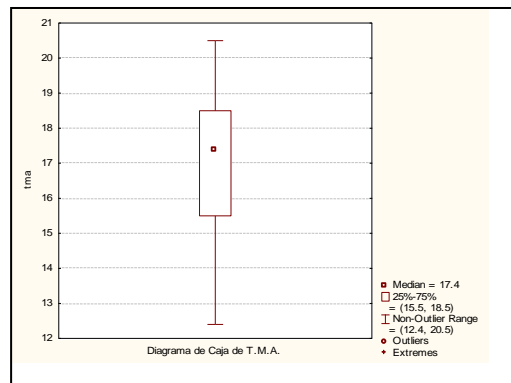


Figura 4.4. Diagrama de Caja (Box-plot) de la Temperatura Media Anual

Prueba de Grubbs.

Esta prueba detecta los valores extremos de las distribuciones normales. Los datos para la prueba son los valores mínimo y máximo. El resultado es una probabilidad que indica que los datos pertenecen al núcleo de población.

La prueba se basa en la diferencia de la media de la muestra y el dato más extremo considerando a la desviación estándar (Grubbs, 1950, 1969).

Hipótesis a Contrastar:

H_0 : No hay valores atípicos en el conjunto de datos

H_a : Hay por lo menos un valor atípico en el conjunto de datos

El Estadístico de prueba de Grubbs se define como:

$$G = \frac{\max_{i=1, \dots, N} |Y_i - \bar{Y}|}{\sigma} \quad (74)$$

Con \bar{Y} y σ denotan la media muestral y la desviación estándar, respectivamente.

El estadístico de prueba de Grubbs es la mayor desviación absoluta de la media de la muestra en unidades de la desviación estándar muestral

Para probar si el valor mínimo es un caso aparte, el estadístico de prueba es

$$G = \frac{\bar{Y} - Y_{\min}}{\sigma} \quad (75)$$

Con Y_{\min} que indica el valor mínimo.

Para probar si el valor máximo es un caso aparte, el estadístico de prueba es

$$G = \frac{Y_{\max} - \bar{Y}}{\sigma} \quad (76)$$

Con Y_{\max} que indica el valor máximo.

Para la prueba de dos colas, la hipótesis de no outliers se rechaza al nivel de significancia α si :

$$G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t_{\alpha/(2N), N-2}^2}{N-2 + t_{\alpha/(2N), N-2}^2}} \quad (77)$$

con $t_{\alpha/(2N), N-2}$ que denota el alto valor crítico de la distribución t de Student con $N-2$ grados de libertad y un nivel de significancia de $\alpha/(2N)$. Para la pruebas de una cola, sustituir $\alpha/(2N)$ con α/N .

El Resultado de la Prueba es la siguiente

Cuadro 4.2. Prueba de Grubbs

	N	Media	Grubbs Test	p-value	Minino	Máximo	Std.Dev.
T.M.A.	21	17.04286	2.155957	0.479669	12.4	20.5	2.153502

Decisión: Como $G=2.155 < 2.58$ (tabla A-4), Se Acepta la Hipótesis Nula, No hay Valores Atípicos en el conjunto de datos.

Análisis Geoestadístico.

Estimación del Variograma.

El **variograma** o **semivariograma** es una herramienta que permite analizar el comportamiento espacial de una variable sobre un área definida, obteniendo como resultado la influencia de los datos a diferentes distancias. A partir de los datos proporcionados por el variograma teórico se realizará la estimación por krigeaje.

El procedimiento correcto para su uso es:

1. Eliminación de cualquier comportamiento tendencial.

2. Selección de la h (*lag* en inglés) y del ángulo en el caso de que se trabaje con variogramas direccionales.
3. Creación del variograma experimental basado en la muestra seleccionada.
4. Selección de un variograma teórico que se adecúe al experimental, pues el variograma experimental no es una función donde se puedan realizar interpolaciones.

Para la creación del variograma experimental, se generó una base de datos con el programa GS+ (Geostatistics for the environmental sciences) versión 5.1, la cual incluye:

- .- La coordenada X (Longitud) en Unidades UTM del área de estudio.
- .-La coordenada Y (Latitud) en unidades UTM del área de estudio.
- .- La Variable Z (Temperatura Media Anual) del área de estudio.

Tomando en cuenta las consideraciones para su cómputo:

- .-considerar un número máximo de 25 intervalos y un mínimo de 10 para determinar con precisión el rango y la meseta del variograma.
- .-Se considera que debe haber entre 30 y 50 pares de puntos como mínimo por intervalo.

Este Variograma experimental se ajustará a un modelo teórico de variograma. Entre los modelos más conocidos se encuentran: el esférico, el gaussiano, el exponencial, el lineal con umbral y sin umbral. Estos modelos se caracterizan fundamentalmente por tres parámetros: el alcance o radio de correlación espacial, el umbral (sill) y la varianza pepita (nugget). Entonces el proceso de modelación se reduce a determinar cuál modelo y con qué parámetros se ajusta mejor a los valores estimados del variograma.

El Resultado es el siguiente:

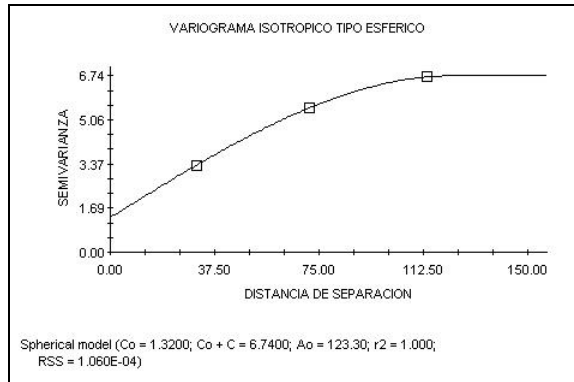


Figura 4.5. Variograma tipo Esférico.

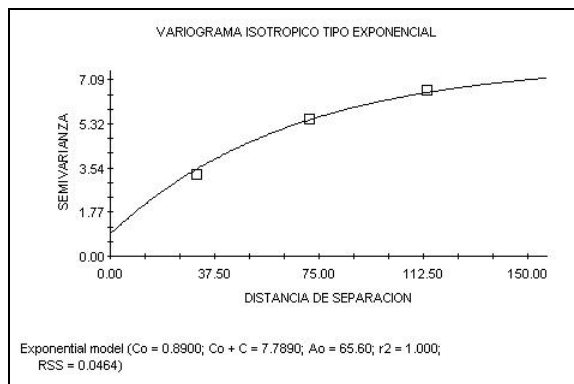


Figura 4.6. Variograma tipo Exponencial

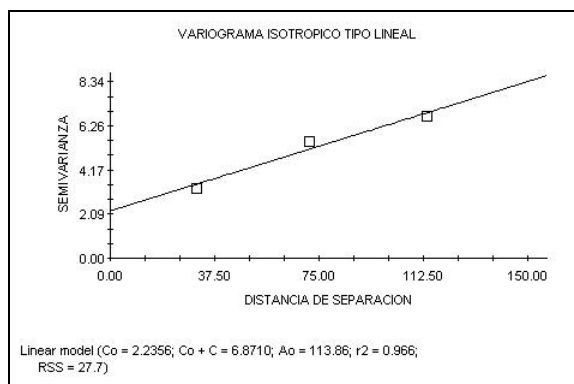


Figura 47. Variograma tipo Lineal

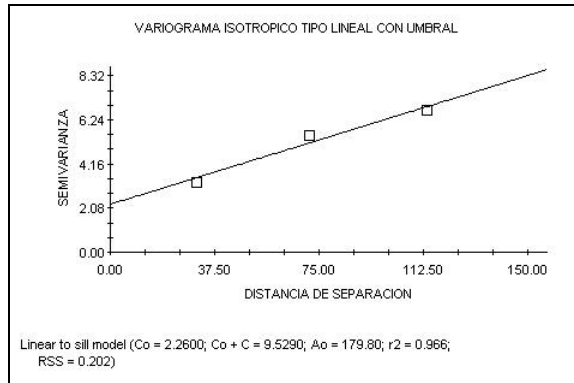


Figura 4.8. Variograma tipo Lineal con Umbral

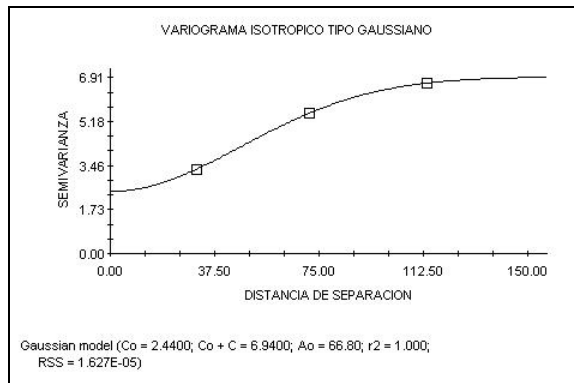


Figura 4.9. Variograma tipo Gaussiano.

Cuadro 4.3. Parámetros geoestadísticos de los diferentes modelos teóricos de semivariograma ajustados

Model	Nugget Co	Sill Co + C	Range Parameter Ao	Effective Range	Proportion C/(Co+C)	r2	RSS
<input type="radio"/> Spherical	1.32000	6.74000	123.3000	123.3000	0.804	1.000	1.060E-04
<input type="radio"/> Exponential	0.89000	7.78900	65.6000	196.8000	0.886	1.000	0.0464
<input type="radio"/> Linear	2.23555	6.87100	113.8626	113.8626	0.675	0.966	27.7
<input type="radio"/> Linear to sill	2.26000	9.52900	179.8000	179.8000	0.763	0.966	0.202
<input checked="" type="radio"/> Gaussian	2.44000	6.94000	66.8000	115.7010	0.648	1.000	1.627E-05

Analizando los Modelos Teóricos ajustados, se observa que el modelo Gaussiano tiene mayor valor del coeficiente de determinación, $r^2=1.000$, que implicara un valor alto del coeficiente de correlación, $r=1.000$, el cual es una estimación del coeficiente de correlación poblacional ρ y que describe la correlación entre h y $\hat{\gamma}(h)$.

Interpretación:

.- Efecto Nugget (C_0)= 2.44Km: Generalmente el variograma no tiende a cero como lo hace la distancia h , el Efecto Nugget indica la discontinuidad del semivariograma para distancias que sean menores, que la menor distancia que se da entre los puntos muestrales. Parte de esta discontinuidad puede también ser dividida en errores de medición o una pobre precisión analítica.

.- Umbral (Sill) = $(C_1 + C_0)$ =6.94Km: Es el valor alcanzado por el variograma en el Rango. El Sill de un variograma, puede ser denotado como la varianza de la muestra, $\text{Var } Z(x_i)$. El cambiar el valor de Sill, no cambia los valores de las estimaciones, por lo que los mapas de estimaciones seguirán siendo los mismo, afecta a la variación de las estimaciones, un sill más alto indica, mayor variación en las estimaciones.

.-Rango (h)=66.8 Km.: es la distancia dentro de la cual, las muestras están correlacionadas espacialmente, es decir es la distancia que determina la dependencia espacial entre las observaciones de la temperatura media anual. Un valor grande para el rango (h), significa un comportamiento más continuo.

.-Rango Efectivo =115.70 Km. es la distancia para la cual el semivariograma alcanza el 95% de la meseta (sill).

.- Proporción de la estructura espacial o $C / (C_0 + C)$ =0.648 esta estadística proporciona una medida de la proporción de la varianza de la muestra ($C_0 + C$) que se explica por la variación espacial estructurada C .

Por lo tanto, se seleccionara para la interpolación Kriging al modelo Gaussiano cuya grafica se presenta en la figura 4.9.

Estimación Espacial Kriging

En la interpolación Kriging se debe tener en cuenta el tamaño de la vecindad y el número de puntos vecinos que intervienen en la estimación. Normalmente el tamaño de la vecindad se elige cercano al alcance o radio de correlación y se considera 16 puntos como una cantidad razonable para una buena estimación.

Se considero un Lag de 150 Km. y vecinos = 16 para la estimación así mismo también se considero a los puntos irregularmente espaciados tomando en cuenta las coordenadas Cartesianas, los resultados de las estimaciones fueron:

Cuadro 4.4 Estimaciones Kriging

GS+ Output: Puntual Kriging Interpolation File v3.0				
Set: Estimación de T.M.A. de la región sureste de Coahuila				
Dimensions: 2; Interval source: File				
File: Tesis				
rango: 15.197 - 19.489				
Z-sd: T.M.,A. (SD)				
rango: 0.540 - 0.912				
Mean Z-estimate (sd): 17.034 (2.030)				
Valid N: 21; Missing N: 0; Missing Value Indicator: -99				
Coordenada X	Coordenada Y	Z(T.M.A.) Estimada	EstStdDev	N(Vecinos)
274.86	2875.07	19.461	0.725	16
313.96	2814.22	16.194	0.597	16
345.99	2803.26	15.265	0.54	16
267.82	2861.12	19.489	0.722	16
287.14	2779.81	16.897	0.687	16
273.41	2796.38	17.81	0.712	16
292.4	2891.52	19.324	0.894	16
311.25	2816.23	16.391	0.579	16
295.04	2762.48	16.78	0.729	16
258.69	2844.47	19.382	0.813	16
315.65	2789.24	15.633	0.755	16
337.34	2806.53	15.3	0.571	16
308.73	2726.19	17.776	0.912	16
312.17	2761.84	16.326	0.78	16
364.81	2797.12	15.766	0.733	16
303.99	2826.09	17.311	0.688	16
300.71	2825.49	17.492	0.692	16
337.2	2795.46	15.197	0.608	16
337.22	2797.3	15.204	0.597	16
298.84	2812.59	17.06	0.657	16
276.41	2795	17.649	0.693	16

Así mismo se realizó una regresión Lineal Simple de valores observados y valores estimados, el resultado es el siguiente:

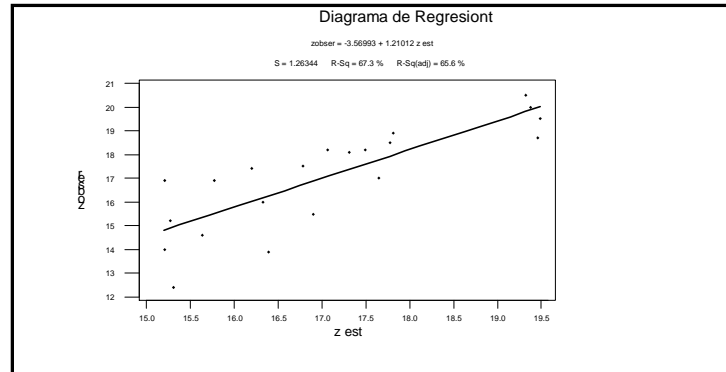


Figura 4.10. Diagrama de Regresión entre las temperaturas observadas y estimadas

Cuadro 4.5. Estadísticas de la Regresión entre las Temperaturas observadas y estimadas

<i>Estadísticas de la regresión</i>				
Coeficiente de correlación múltiple		0.82036693		
Coeficiente de determinación R ²		0.6730019		
R ² ajustado		0.65579148		
Error típico		1.26344435		
Observaciones		21		
	<i>Coeficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>
Intercepción	3.56993213	3.30779268	-1.07924905	0.2939875
Variable X 1	1.2101205	0.19351574	6.2533442	5.2656E-06

Análisis de Varianza para la Estimación.

Hipótesis a contrastar:

H₀ = No Existe Regresión lineal entre Z actual y Z estimada.

H₁ = Si Existe Regresión lineal entre Z actual y Z estimada.

Cuadro 4.6.- Análisis de Varianza

	<i>G.L.</i>	<i>S.C.</i>	<i>C.M.</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	1	62.4218879	62.4218879	39.1043137	5.26565E-06
Residuos	19	30.3295407	1.59629161		
Total	20	92.7514286			

Decisión:

Como el valor $F=39.104 > 5.265E-06$ se rechaza la Hipótesis nula y se concluye que el valor actual de Z (T.M.A.) está influenciado, en forma lineal, por la Z estimada de la T.M.A. con un nivel de confianza del 95 %.

Además, la presencia lineal de la Temperatura media anual Estimada en el modelo de regresión lineal explica el 67 % de la variación total en los valores observados de la temperatura media anual. (Coeficiente de Determinación r^2). Se tiene también una asociación lineal entre la Temperatura media anual estimada y la observada del 82 % (coeficiente de correlación r). Lo que indica que existe una relación lineal entre las variables pues este valor a pesar de no estar muy lejos de cero, nos dice que existen otros factores que también pudieran estar condicionando la temperatura media anual observada.

Para ello es necesario realizar una inferencia acerca del coeficiente de correlación a través de una prueba de hipótesis utilizando el estadígrafo t student con un 95 % de confianza a $\alpha = 0.05$

Hipótesis a contrastar:

$H_0: \rho = 0$ (No existe Correlación Lineal)

$H_a: \rho \neq 0$ (Existe Correlación Lineal)

El estadístico de prueba es:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (78)$$

Se rechaza la H_0 si $t \geq t_{(1-\frac{\alpha}{2})(n-2)}$ en caso contrario se acepta H_a

Cálculos:

$$t = \frac{0.82\sqrt{19}}{\sqrt{1-0.82^2}} = \frac{3.574}{0.572} = 6.248 \quad t_{\frac{0.05}{2}, (21-2)} = 2.0930 \text{ (Valor de tabla A-5)}$$

Decisión:

Como el valor de $t = 6.248 > 2.0930$ excede el valor crítico obtenido de la tabla se deduce que hay evidencia suficiente para señalar que existe correlación lineal entre la temperatura media anual observada y la temperatura media anual estimada.

Validación del modelo del Variograma.

Se puede procurar satisfacer lo mejor posible los siguientes criterios estadísticos;

1).- La media de los errores mide el Sesgo del estimador, y deben ser cercanas a cero.

De hecho, este criterio resulta secundario con respecto al modelamiento del variograma y la elección de la vecindad de kriging pues los valores obtenidos por kriging son insesgados por construcción. Así, en la práctica, la media de los errores (Error Medio=E.M.) y la media de los errores estandarizados (Error medio Estandarizado=E.M.E.) tenderán hacia cero.

$$E.M. = \frac{1}{n} \sum_{i=1}^n \left[\hat{Z}(x_i) - Z(x_i) \right] \cong 0$$

$$E.M. = \frac{-0.2}{21} = -0.009$$

$$E.M.E. = \frac{1}{n} \sum_{i=1}^n \left[\frac{\hat{Z}(x_i) - Z(x_i)}{\sigma(x_i)} \right] \cong 0$$

$$E.M.E. = \frac{1.56}{21} = 0.074$$

2).- La varianza de los errores (Cuadrado Medio del Error=C.M.E.), que mide la Precisión del estimador, debe ser mínima.

$$C.M.E. = \frac{1}{n} \sum_{i=1}^n \left[\hat{Z}(x_i) - Z(x_i) \right]^2 \cong \text{mínima}$$

$$C.M.E. = \frac{32.21}{21} = 1.534$$

3).- El error estándar de estimación (E.E) debe ser cercano a cero

$$E.E. = \frac{\sqrt{C.M.E.}}{n} \cong 0$$

$$E.E. = \frac{\sqrt{1.534}}{21} = 0.059$$

4).- El porcentaje del error debe ser menor o igual al 10%

$$P.E. = \frac{\sqrt{\frac{\sum_{i=1}^n \left(\hat{Z}(x_i) - Z(x_i) \right)^2}{n-1}}}{\frac{\sum_{i=1}^n Z(x_i)}{n}} \leq 10\%$$

$$P.E. = \frac{\sqrt{\frac{32.21}{20}}}{\frac{357.9}{21}} = \frac{1.269}{17.043} = 0.074 \times 100 = 7.4\%$$

En términos generales se considera una buena estimación cuando el porcentaje del error es menor o igual al 10%.

5).-El coeficiente de correlación entre los valores estimados y los valores observados debe ser lo más cercano posible a 1.

$$\rho = \left[\hat{Z}(x_i), Z(x_i) \right] \cong 1$$

$$\rho = \left[\hat{Z}(x_i), Z(x_i) \right] = 0.82$$

6).- El Coeficiente de efectividad de Predicción (E) debe ser cercano al 100%

$$E = \left[1 - \frac{\sum_{i=1}^n \left[\hat{Z}(x_i) - Z(x_i) \right]^2}{\sum_{i=1}^n \left[Z(x_i) - \bar{Z} \right]^2} \right] \times 100$$

$$E = \left[1 - \left[\frac{32.21}{92.75} \right] \right] = 0.65 \times 100 = 65\%$$

Idealmente todos los estadígrafos anteriores deben satisfacerse simultáneamente, pero en la práctica una mejoría en uno de ellos puede degradar a otro. Por lo que es recomendable hacer un análisis integral de los estadígrafos de las diferencias.

Estos criterios permiten comprobar el desempeño del kriging y comparar la calidad de Diferentes ajustes posibles para el variograma.

Inferencias estadísticas del error

Los resultados correspondientes a los errores para las observaciones de la temperatura media anual de la región sureste de Coahuila utilizando el método kriging son:

Cuadro 4.7. Errores al aplicar la metodología Kriging para las observaciones de la temperatura media anual de la región sureste de Coahuila

Estación	$Z(x_i)$	$\hat{Z}(x_i)$	$\sigma(x_i)$	$e(x_i)$	$e(x_i)^2$
1	18.7	19.461	0.725	0.761	0.57912
2	17.4	16.194	0.597	-1.206	1.45444
3	15.2	15.265	0.54	0.065	0.00423
4	19.5	19.489	0.722	-0.011	0.00012
5	15.5	16.897	0.687	1.397	1.95161
6	18.9	17.81	0.712	-1.090	1.1881
7	20.5	19.324	0.894	-1.176	1.38298
8	13.9	16.391	0.579	2.491	6.20508
9	17.5	16.78	0.729	-0.720	0.5184
10	20	19.382	0.813	-0.618	0.38192
11	14.6	15.633	0.755	1.033	1.06709
12	12.4	15.3	0.571	2.900	8.41
13	18.5	17.776	0.912	-0.724	0.52418
14	16	16.326	0.78	0.326	0.10628
15	16.9	15.766	0.733	-1.134	1.28596
16	18.1	17.311	0.688	-0.789	0.62252
17	18.2	17.492	0.692	-0.708	0.50126
18	14	15.197	0.608	1.197	1.43281
19	16.9	15.204	0.597	-1.696	2.87642
20	18.2	17.06	0.657	-1.140	1.2996
21	17	17.649	0.693	0.649	0.4212
SUMAS	357.9	357.7	14.68	-0.2	32.21

\bar{X} =media de la serie de los $e(x_i)$ =-0.009

σ =desviación estándar de la serie de los $e(x_i)$ =1.2691

n=numero total de la serie de los $e(x_i)$ =21

De la tabla 4.7, asumiendo distribución normal de los errores, se plantean las hipótesis

de:

$$H_0: m=0$$

$$H_a: m \neq 0$$

Siendo m la media de los errores, de tal forma que, aplicando la ecuación (70) se tiene:

$$t_c = \frac{\sqrt{21}(-0.009 - 0)}{1.2691} = -0.032$$

La regla de decisión es:

Rechazar $H_0: m=0$ si $|t_c| \geq t_{\alpha/2, (n-1)g.l.}$ si el nivel de significancia $\alpha=0.05$, $t_{\alpha/2, (n-1)g.l.}$ será igual a $t_{0.025, (20)g.l.} = 2.423$. Por lo tanto no se rechaza $H_0: m=0$ bajo un nivel de significancia α de 0.05.

Un intervalo de confianza basado sobre la distribución t-student, con una confiabilidad del 95%, es, de acuerdo a la ecuación (71):

$$(-0.009) - (2.086)\left(\frac{1.2691}{\sqrt{21}}\right) \leq m \leq (-0.009) + (2.086)\left(\frac{1.2691}{\sqrt{21}}\right)$$

$$-0.5867 \leq m \leq 0.5687$$

También asumiendo normalidad de los errores se plantean las hipótesis de:

$$H_0: \sigma^2 = 1$$

$$H_a: \sigma^2 \neq 1$$

Donde σ^2 corresponde a la varianza de los errores. Así, aplicando la ecuación (72):

$$\chi_c^2 = \frac{(20)(1.2691)^2}{1} = 25.382$$

La regla de decisión es: se rechaza $\sigma^2 = 1$ si: $\chi_c^2 \geq \chi_{\alpha/2, (n-1)g.l.}^2$ o si $\chi_c^2 \leq \chi_{1-\alpha/2, (n-1)g.l.}^2$.

Si el nivel de significancia $\alpha=0.05$

$$\chi_{\alpha/2, (n-1)g.l.}^2 = \chi_{0.025, 20g.l.}^2 = 34.17$$

$$\chi_{1-\alpha/2, (n-1)g.l.}^2 = \chi_{0.975, 20g.l.}^2 = 9.59$$

Por lo que no se rechaza $H_0: \sigma^2 = 1$ bajo un nivel de significancia α de 0.05

Un intervalo de confianza para σ^2 con un 95% de confiabilidad de acuerdo a la ecuación (73) es:

$$\frac{32.21}{34.17} < \sigma^2 < \frac{32.21}{9.59}$$

$$0.9426 < \sigma^2 < 3.3587$$

Cuadro 4.8.- Cuadro de comparación de las temperaturas medias anuales observadas y estimadas de la región sureste del estado de Coahuila

No	Estación	Municipio	T.m.a.- obs	T.m.a.-est
1	Alto de Norias	Ramos Arizpe	18.7	19.461
2	Arteaga	Arteaga	17.4	16.194
3	Purísima	Arteaga	15.2	15.265
4	La Saucedá	Ramos Arizpe	19.5	19.489
5	Carneros	Saltillo	15.5	16.897
6	Derramadero	Saltillo	18.9	17.81
7	Ejido Reata	Ramos Arizpe	20.5	19.324
8	El Tunal	Arteaga	13.9	16.391
9	Gomez Farias	Saltillo	17.5	16.78
10	Hipólito	Ramos Arizpe	20	19.382
11	Huachichil	Arteaga	14.6	15.633
12	Jame	Arteaga	12.4	15.3
13	La Ventura	Saltillo	18.5	17.776
14	Las Hormigas	Saltillo	16	16.326
15	Potrero	Arteaga	16.9	15.766
16	R. Arizpe(DGE)	Ramos Arizpe	18.1	17.311
17	R. Arizpe(SMN)	Ramos Arizpe	18.2	17.492
18	Alazanas(DGE)	Arteaga	14	15.197
19	Alazanas(SMN)	Arteaga	16.9	15.204
20	Saltillo(DGE)	Saltillo	18.2	17.06
21	Vaquería	Saltillo	17	17.649

Promedio de la temperatura media anual observada = 17.042^0 C.

Promedio de la temperatura media anual estimada = 17.034^0 C.

Por lo tanto, la temperatura media anual de la región sureste del estado de Coahuila es del orden de los 17^0 C.

5.- CONCLUSIONES

Al realizar el análisis Estadístico univariado para la variable temperatura media anual de la región sureste del estado de Coahuila se observó que existe un mejor ajuste a una normal en coordenadas naturales, también se apreció que no hay valores atípicos (outliers) detectables en la información analizada.

De la base de datos generada para la variable mencionada, el análisis geoestadístico realizado mediante el programa GS+ (Geostatistics for Environmental Sciences) se crearon diferentes semivariogramas experimentales los cuales fueron ajustados a modelos teóricos, el cual el mejor ajuste resultó de tipo Gaussiano con un coeficiente de Determinación $r^2=1.000$ demostrando con ello que los datos analizados de temperatura media anual de la región sureste del estado de Coahuila se comporta como una variable regionalizada distribuida espacialmente. Tanto del análisis estadístico y geoestadístico de la variable regionalizada analizada se concluye que su media es del orden de 17^0 C. Con un análisis de regresión de valores observados y estimados de dicha variable presentó un coeficiente de determinación de $r^2=0.673$ y un coeficiente de correlación de $r=0.82$, lo cual indica una asociación lineal entre la temperatura media anual observada y estimada.

Con el modelo de semivariograma seleccionado se procedió a la interpolación Kriging, la cual requirió un lag de 150 km y 16 vecinos involucrados.

Para la validación del modelo del semivariograma, se procuro satisfacer los diferentes criterios estadísticos debido a los errores en la interpolación Kriging los cuales fueron: la media de los errores, promedio de los errores estandarizados y el error

estándar de estimación deben ser cercanos a cero; el cuadrado medio del error deberá ser mínimo; el porcentaje del error debe ser menor o igual al 10% y el coeficiente de efectividad de estimación debe ser cercano al 100%. Se evaluó el método kriging mediante los errores y aplicando pruebas de hipótesis se demostró que éstos se distribuyen normal con media cero y varianza uno.

Con el ejemplo se ilustra como la geoestadística constituye una herramienta muy útil en el estudio de las distribuciones espaciales de las variables naturales, lo cual es esencial y recomendable cuando se deben tomar decisiones que afectan al medio ambiente.

Las ventajas del krigeado con respecto a otros métodos de interpolación más sencillos es que, mediante los procedimientos geoestadísticos, se conoce la bondad de la estimación en toda el área de estudio. Además el kriging proporciona un análisis más elaborado y con un fundamento estadístico por lo que este método puede ser el preferido cuando la intensidad de muestreo es mayor.

Con los resultados obtenidos se permiten detectar varios temas directamente relacionados con el enfoque propuesto que serían muy interesante analizar mediante los procedimientos geoestadísticos, entre ellos los siguientes: Generación de curvas de nivel, simulación geoestadística, muestreo espacial y métodos geoestadísticos multivariantes

6.-LITERATURA CITADA

- Análisis Geoestadístico. Documento en línea: <http://ocw.upm.es/proyectos-de-ingenieria/sistemas-de-informacion-geografica-tecnicas-cuantitativas>.
- Camarero, J.J. Rozas, V. 2006. Técnicas de Análisis Espacial de patrones de superficies y Detección de Fronteras Aplicadas en Ecología Forestal. Investigaciones Agrarias. Sistemas de Recursos Forestales.: 15(1), pp.: 66-87.
- Castillo Daniel I., Corral José R., Eguiarte Diego G., Garnica José F., Padilla Gabriel D. 2009. Distribución Espacial del PH de los Suelos Agrícolas de Zapopan Jalisco, México. Agricultura Técnica en México: 35(3) pp.:267-276.
- Cortez Adriana, Ovallez Francisco A., Rodríguez Maria F., Rey Juan C., Núñez Maria C. 2005. Análisis Geoestadístico de la Variabilidad Temporal a una Escala anual y mensual de parámetros Climáticos en un campo Experimental, Estado de Aragua, Venezuela. Agronomía Tropical: 55(3).
- Cuador Gil, J.Q. 1997. Elementos de Geoestadística. Universidad de Pinar del Río, Cuba. Documento en línea <http://www.monografias.com>.
- Díaz Viera, M.A. 2002. Geoestadística Aplicada. Instituto de Geofísica, UNAM. Instituto de Geofísica y Astronomía, CITMA, Cuba.
- Díaz Viera, M.A. 2002. Metodología para el Análisis Geoestadístico del Comportamiento de los Niveles piezométricos en un Acuífero. Instituto de Geofísica, UNAM. Instituto de Geofísica y Astronomía, CITMA, Cuba.
- Estado de Coahuila. Monografía. Mayo de 2009. <http://www.aregional.com>

- Figueras, S.; Gargallo, P. 2003. Análisis Exploratorio de Datos (en línea) Disponible:
<http://www.5campus.com/leccion/aed>.
- Funes, Nicolás, H. 2004. Métodos Estadísticos en Problemas Espaciales. Tesis. Magíster en Estadística Aplicada. Instituto de Investigaciones Estadísticas. Facultad de Ciencias Económicas. Universidad Nacional de Tucumán, San Miguel de Tucumán
- Giraldo, H.R. 2009. Estadística Espacial. Notas de Clase. Departamento Estadística. Universidad Nacional de Colombia, Sede Bogota.
- Giraldo, H.R. 2009. Introducción a la Geoestadística: Teoría y Aplicación. Departamento de Estadística. Universidad Nacional de Colombia. Sede Bogota.
- Matheron, G. 1969. Curso de Geoestadística. Traducido al Español por Marco Alfaro 2005. Cuadernos del Centro de Morfología Matemática de Fontainebleau. Fascículo 2 Centro de Geoestadística de la Escuela de Minas de Paris, Francia.
- Moral García, F.J. 2004. Aplicación de la Geoestadística en las Ciencias Ambientales. Revista Ecosistemas: 13(1), pp. 78-86.
- Ostle, B. 1965. Estadística Aplicada 1ª Ed. LIMUSA, México. 629 p.
- Snedecor, G.W. and W.G. Cochran, 1967. Statistical methods 65h ed. Iowa State University, press, Ames. 593 p.
- Sommer C. Irene, Fernández L. Pilar, Rivas S. Hilda, Gutiérrez R. Margarita .2000. La Geoestadística como herramienta en estudios de contaminación de suelos, Análisis de Caso: Afectación por Arsénico, Plomo y Cadmio contenidos en jales Mineros. Revista Internacional de Contaminación Ambiental: 16(4) pp. : 205-214.

- Umbría N. Igle, Mora M. Luís, Jegat Herve, Trezza P. Ricardo, Mejias E. Jesús, Araujo F. Luís, Rodríguez B. Nelson. 2002. Análisis Espacial de la Salinidad del Acuífero en la Planicie Aluvial del Río Motatan-Trujillo. *Geoenseñanza*: 7(1-2) pp.:74-87.
- Valdez Cepeda, R.D.1997. Análisis Estocástico Espacial de Procesos Edáficos y de Plantas en *Zea mays L* Tesis Doctoral en ciencias Agrícolas con Especialidad en Agua-Suelo .Universidad Autónoma de Nuevo León, Facultad de Agronomía, Marín N.L., México.
- Vanderlinden, Karl. 2001. Análisis de Procesos Hidrológicos a diferentes escalas Espacio-Temporales. Tesis Doctoral. Universidad de Córdoba. Escuela Técnica Superior de Ingenieros Agrónomos y Montes, Departamento de Agronomía, Córdoba.
- Vázquez Polo J. R. 2009. Caracterización de la Variabilidad Espacial de las Propiedades Físicas y Químicas de los Suelos de la granja Experimental de la Universidad de Magdalena. Tesis de Maestría en Ciencias Agrarias con Énfasis en Suelos.Santa Marta, Colombia
- Villatoro Mario, Henríquez Carlos, Sancho Freddy. 2008. Comparación de los Interpoladores IDW y Kriging en la Variación Espacial de PH., Ca, CICE, y P del Suelo. *Agronomía Costarricense*: 32(1) pp.: 99-105. Universidad de Costa Rica, San José de Costa Rica.

APENDICE

A.1 Hoja de cálculo para los diferentes parámetros de la Validación Kriging

Estación	$Z(x_i)$	$\hat{Z}(x_i)$	$\sigma(x_i)$	$e(x_i)$	$e(x_i)^2$	$\frac{e(x_i)}{\sigma(x_i)}$	$Z(x_i) - \bar{Z}_{2\sigma}$	$(Z(x_i) - \bar{Z}_{2\sigma})^2$
1	18.7	19.461	0.725	-0.761	0.57912	-1.0497	1.666	2.77556
2	17.4	16.194	0.597	1.206	1.45444	2.0201	0.366	0.13396
3	15.2	15.265	0.54	-0.065	0.00423	-0.1204	-1.834	3.36356
4	19.5	19.489	0.722	0.011	0.00012	0.01524	2.466	6.08116
5	15.5	16.897	0.687	-1.397	1.95161	-2.0335	-1.534	2.35316
6	18.9	17.81	0.712	1.090	1.1881	1.5309	1.866	3.48196
7	20.5	19.324	0.894	1.176	1.38298	1.31544	3.466	12.0132
8	13.9	16.391	0.579	-2.491	6.20508	-4.3022	-3.134	9.82196
9	17.5	16.78	0.729	0.720	0.5184	0.98765	0.466	0.21716
10	20	19.382	0.813	0.618	0.38192	0.76015	2.966	8.79716
11	14.6	15.633	0.755	-1.033	1.06709	-1.3682	-2.434	5.92436
12	12.4	15.3	0.571	-2.900	8.41	-5.0788	-4.634	21.474
13	18.5	17.776	0.912	0.724	0.52418	0.79386	1.466	2.14916
14	16	16.326	0.78	-0.326	0.10628	-0.4179	-1.034	1.06916
15	16.9	15.766	0.733	1.134	1.28596	1.54707	-0.134	0.01796
16	18.1	17.311	0.688	0.789	0.62252	1.1468	1.066	1.13636
17	18.2	17.492	0.692	0.708	0.50126	1.02312	1.166	1.35956
18	14	15.197	0.608	-1.197	1.43281	-1.9688	-3.034	9.20516
19	16.9	15.204	0.597	1.696	2.87642	2.84087	-0.134	0.01796
20	18.2	17.06	0.657	1.140	1.2996	1.73516	1.166	1.35956
21	17	17.649	0.693	-0.649	0.4212	-0.9365	-0.034	0.00116
SUMAS	357.9	357.7	14.68	0.2	32.21	-1.56	0.186	92.75
media de los errores =0.009 desviación estándar de los errores =1.2691								

A.2 Tabla de Valores Críticos de D de la Prueba de Kolmogorov-Smirnov

TABLA 8. *Contraste de Kolmogorov-Smirnov*

Valores críticos de $D = |F_n(x) - F(x)|$ donde $F_n(x)$ es la distribución muestral de tamaño n y $F(x)$ la distribución teórica.

Tamaño muestral n	Nivel de significación				
	0,20	0,15	0,10	0,05	0,01
1	0,900	0,925	0,950	0,975	0,995
2	0,684	0,726	0,776	0,842	0,929
3	0,565	0,597	0,642	0,706	0,828
4	0,494	0,525	0,564	0,624	0,733
5	0,446	0,474	0,510	0,565	0,669
6	0,410	0,436	0,470	0,521	0,618
7	0,381	0,405	0,438	0,486	0,577
8	0,358	0,381	0,411	0,457	0,543
9	0,339	0,360	0,388	0,432	0,514
10	0,322	0,342	0,368	0,410	0,490
11	0,307	0,326	0,352	0,391	0,468
12	0,295	0,313	0,338	0,375	0,450
13	0,284	0,302	0,325	0,361	0,433
14	0,274	0,292	0,314	0,349	0,418
15	0,266	0,283	0,304	0,338	0,404
16	0,258	0,274	0,295	0,328	0,392
17	0,250	0,266	0,286	0,318	0,381
18	0,244	0,259	0,278	0,309	0,371
19	0,237	0,252	0,272	0,301	0,363
20	0,231	0,246	0,264	0,294	0,356
25	0,21	0,22	0,24	0,27	0,32
30	0,19	0,20	0,22	0,24	0,29
35	0,18	0,19	0,21	0,23	0,27
>35	$\frac{1,07}{\sqrt{n}}$	$\frac{1,14}{\sqrt{n}}$	$\frac{1,22}{\sqrt{n}}$	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$

n es el tamaño de la muestra.

A.3.- Tabla de Valores Críticos de la Prueba de Kolmogorov-Smirnov (Lilliefors)

TABLA 9. *Contraste Kolmogorov-Smirnov (Lilliefors)*

Tablas de $D_n = |F_n(x) - F(x)|$ para contrastar la hipótesis de normalidad cuando la media y la varianza poblacionales son estimadas por sus valores muestrales.

Tamaño muestral n	Nivel de significación				
	0,20	0,15	0,10	0,05	0,01
4	0,300	0,319	0,352	0,381	0,417
5	0,285	0,299	0,315	0,337	0,405
6	0,265	0,277	0,294	0,319	0,364
7	0,247	0,258	0,276	0,300	0,348
8	0,233	0,244	0,261	0,285	0,331
9	0,223	0,233	0,249	0,271	0,311
10	0,215	0,224	0,239	0,258	0,294
11	0,206	0,217	0,230	0,249	0,284
12	0,199	0,212	0,223	0,242	0,275
13	0,190	0,202	0,214	0,234	0,268
14	0,183	0,194	0,207	0,227	0,261
15	0,177	0,187	0,201	0,220	0,257
16	0,173	0,182	0,195	0,213	0,250
17	0,169	0,177	0,189	0,206	0,245
18	0,166	0,173	0,184	0,200	0,239
19	0,163	0,169	0,179	0,195	0,235
20	0,160	0,166	0,174	0,190	0,231
25	0,149	0,153	0,165	0,180	0,203
30	0,131	0,136	0,144	0,161	0,187
>30	$\frac{0,736}{\sqrt{n}}$	$\frac{0,768}{\sqrt{n}}$	$\frac{0,805}{\sqrt{n}}$	$\frac{0,886}{\sqrt{n}}$	$\frac{1,031}{\sqrt{n}}$

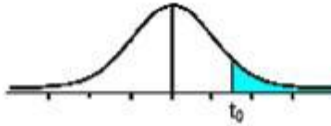
A.4.- Tabla de Valores Críticos para la Prueba de Grubbs.

Grubbs' critical value table:

N	0.1	0.075	0.05	0.025	0.01	N	0.1	0.075	0.05	0.025	0.01
3	1.15	1.15	1.15	1.15	1.15	53	0	0	2.961	3.151	999
4	1.42	1.44	1.46	1.48	1.49	54	0	0	2.968	3.158	999
5	1.6	1.64	1.67	1.71	1.75	55	0	0	2.995	3.165	999
6	1.73	1.77	1.82	1.89	1.94	56	0	0	3.002	3.172	999
7	1.83	1.88	1.94	2.02	2.1	57	0	0	3.009	3.179	999
8	1.91	1.96	2.03	2.13	2.23	58	0	0	3.016	3.186	999
9	1.96	2.04	2.11	2.21	2.32	59	0	0	3.023	3.193	999
10	2.03	2.1	2.18	2.29	2.41	60	0	0	3.03	3.2	999
11	2.09	2.14	2.23	2.36	2.48	61	0	0	3.036	3.206	999
12	2.13	2.2	2.29	2.41	2.55	62	0	0	3.042	3.212	999
13	2.17	2.24	2.33	2.46	2.61	63	0	0	3.048	3.218	999
14	2.21	2.28	2.37	2.51	2.66	64	0	0	3.054	3.224	999
15	2.25	2.32	2.41	2.55	2.71	65	0	0	3.06	3.23	999
16	2.28	2.35	2.44	2.59	2.75	66	0	0	3.066	3.236	999
17	2.31	2.38	2.47	2.62	2.79	67	0	0	3.072	3.242	999
18	2.34	2.41	2.5	2.65	2.82	68	0	0	3.078	3.248	999
19	2.36	2.44	2.53	2.68	2.85	69	0	0	3.084	3.254	999
20	2.38	2.46	2.55	2.71	2.88	70	0	0	3.09	3.26	999
21	0	0	2.58	2.73	2.91	71	0	0	3.095	3.265	999
22	0	0	2.6	2.76	2.94	72	0	0	3.1	3.27	999
23	0	0	2.62	2.78	2.96	73	0	0	3.105	3.275	999
24	0	0	2.64	2.8	2.99	74	0	0	3.11	3.28	999
25	0	0	2.66	2.82	3.01	75	0	0	3.115	3.285	999
26	0	0	2.68	2.84	999	76	0	0	3.12	3.29	999
27	0	0	2.7	2.86	999	77	0	0	3.125	3.295	999
28	0	0	2.72	2.88	999	78	0	0	3.13	3.3	999
29	0	0	2.73	2.9	999	79	0	0	3.135	3.305	999
30	0	0	2.75	2.91	999	80	0	0	3.14	3.31	999
31	0	0	2.76	2.93	999	81	0	0	3.144	3.314	999
32	0	0	2.78	2.95	999	82	0	0	3.148	3.318	999
33	0	0	2.79	2.96	999	83	0	0	3.152	3.322	999
34	0	0	2.81	2.97	999	84	0	0	3.156	3.326	999
35	0	0	2.82	2.98	999	85	0	0	3.16	3.33	999
36	0	0	2.83	2.992	999	86	0	0	3.164	3.334	999
37	0	0	2.84	3.004	999	87	0	0	3.168	3.338	999
38	0	0	2.85	3.016	999	88	0	0	3.172	3.342	999
39	0	0	2.86	3.028	999	89	0	0	3.176	3.346	999
40	0	0	2.87	3.04	999	90	0	0	3.18	3.35	999
41	0	0	2.88	3.05	999	91	0	0	3.183	3.353	999
42	0	0	2.89	3.06	999	92	0	0	3.186	3.356	999
43	0	0	2.9	3.07	999	93	0	0	3.189	3.359	999
44	0	0	2.91	3.08	999	94	0	0	3.192	3.362	999
45	0	0	2.92	3.09	999	95	0	0	3.195	3.365	999
46	0	0	2.928	3.096	999	96	0	0	3.198	3.368	999
47	0	0	2.936	3.106	999	97	0	0	3.201	3.371	999
48	0	0	2.944	3.114	999	98	0	0	3.204	3.374	999
49	0	0	2.952	3.122	999	99	0	0	3.207	3.377	999
51	0	0	2.967	3.137	999	100	0	0	3.21	3.38	999
52	0	0	2.974	3.144	999						

A.5.- Tabla de Valores Críticos para la Prueba “t” de Student

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7849	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8985	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8186
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3007	1.6794	2.0141	2.4121	2.6896
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800

Continúa A.5.

50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778
51	0.6793	1.2984	1.6753	2.0078	2.4017	2.6757
52	0.6792	1.2980	1.6747	2.0066	2.4002	2.6737
53	0.6791	1.2977	1.6741	2.0057	2.3988	2.6718
54	0.6791	1.2974	1.6736	2.0049	2.3974	2.6700
55	0.6790	1.2971	1.6730	2.0040	2.3961	2.6682
56	0.6789	1.2969	1.6725	2.0032	2.3948	2.6665
57	0.6788	1.2966	1.6720	2.0025	2.3936	2.6649
58	0.6787	1.2963	1.6716	2.0017	2.3924	2.6633
59	0.6787	1.2961	1.6711	2.0010	2.3912	2.6618
60	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603
61	0.6785	1.2956	1.6702	1.9996	2.3890	2.6589
62	0.6785	1.2954	1.6698	1.9990	2.3880	2.6575
63	0.6784	1.2951	1.6694	1.9983	2.3870	2.6561
64	0.6783	1.2949	1.6690	1.9977	2.3860	2.6549
65	0.6783	1.2947	1.6686	1.9971	2.3851	2.6536
66	0.6782	1.2945	1.6683	1.9966	2.3842	2.6524
67	0.6782	1.2943	1.6679	1.9960	2.3833	2.6512
68	0.6781	1.2941	1.6676	1.9955	2.3824	2.6501
69	0.6781	1.2939	1.6672	1.9949	2.3816	2.6490
70	0.6780	1.2938	1.6669	1.9944	2.3808	2.6479
71	0.6780	1.2936	1.6666	1.9939	2.3800	2.6469
72	0.6779	1.2934	1.6663	1.9935	2.3793	2.6458
73	0.6779	1.2933	1.6660	1.9930	2.3785	2.6449
74	0.6778	1.2931	1.6657	1.9925	2.3778	2.6439
75	0.6778	1.2929	1.6654	1.9921	2.3771	2.6430
76	0.6777	1.2928	1.6652	1.9917	2.3764	2.6421
77	0.6777	1.2926	1.6649	1.9913	2.3758	2.6412
78	0.6776	1.2925	1.6646	1.9908	2.3751	2.6403
79	0.6776	1.2924	1.6644	1.9905	2.3745	2.6395
80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9852	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2903	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

