

**UNIVERSIDAD NACIONAL DEL ALTIPLANO-PUNO**  
**FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA**  
**ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA**



**TESIS**

**MODELO DE REGRESIÓN PARA LA PRECIPITACIÓN PLUVIAL  
EN LA CIUDAD DE PUNO PERIODO 1979 - 2012**

**PRESENTADA POR:**

**Bach. MAGDALENA JALLO APAZA**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

**INGENIERO ESTADÍSTICO E INFORMÁTICO**

**PUNO – PERU**

**2014**

**1**



**UNIVERSIDAD NACIONAL DEL ALTIPLANO - PUNO**  
**FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA**  
**ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA**



## TESIS

**MODELO DE REGRESIÓN PARA LA PRECIPITACIÓN PLUVIAL EN LA CIUDAD DE PUNO PERIODO 1979 - 2012**

**PRESENTADA POR:**

**Bach. MAGDALENA JALLO APAZA**

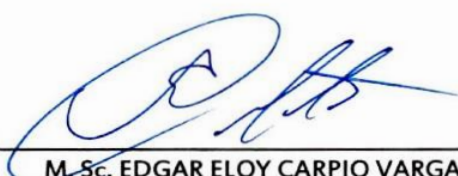


A la Coordinación de Investigación de la Facultad de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano - Puno, para optar el Título Profesional de:

**INGENIERO ESTADÍSTICO E INFORMÁTICO**

**APROBADA POR:**


**PRESIDENTE DEL JURADO**

:   
 M. Sc. EDGAR ELOY CARPIO VARGAS


**PRIMER MIEMBRO**

:   
 M. Sc. ALEJANDRO APAZA TARQUI


**SEGUNDO MIEMBRO**

:   
 Ing. ALCIDES RAMOS CALCINA

**DIRECTOR DE LA TESIS**

:   
 M. Sc. OCTAVIO MORILLOS VALDERRAMA

**ASESOR DE LA TESIS**

:   
 Dr. JUAN REYNALDO PAREDES QUISPE

Área : Estadística  
 Tema : Series de tiempo  
 Fecha de Sustentación : 23/05/2014

## DEDICATORIAS

*Con todo lo que soy y he logrado ser, a Dios por haberme permitido llegar hasta este punto, por no faltarme nunca, por estar presente en todo lugar, en todo momento, circunstancia y hasta donde permita que pueda estar yo.*

*A mis Padres Víctor y Benita a ellos que con sus obras, consejos y su cariño incentivaron mucho más cada día el verme convertida en profesional, por su incondicional apoyo, por el esfuerzo abnegado y constante, ejemplos de perseverancia y constancia, por inculcarme el amor al estudio y por demostrarme que con el esfuerzo, la perseverancia y disciplina podemos realizar todos nuestros sueños.*

*Con todo lo que he podido ser hasta hoy, a mis queridos hermanos: Cesar, Norma y Katty D. por todo su apoyo, ejemplo, cariño y complicidad. Y a mis pequeños sobrinos Jefferson A. y Juem A. por darle alegría a mis días con sus ocurrencias.*

*Por sus consejos, por la motivación constante, por estar siempre ahí cuando las necesite y por todos los momentos inolvidables que vivimos a mis amigas Evelin, Lizet, Deysi y Carla, por tener siempre una sonrisa en la cara frente a cualquier adversidad y ser la muestra de que la verdadera y sincera amistad existe.*

*Magdalena J. A.*

## AGRADECIMIENTOS

1. Mis más sinceras muestras de agradecimiento:

A los Docentes de la Facultad de Ingeniería Estadística e Informática, por compartir sus conocimientos y experiencias profesionales con mí persona.

2. Al Director del Servicio Nacional de Meteorología e Hidrología Puno, Ing. Sixto Sancho Flores, por su apoyo incondicional.

3. A mis padres y a mis hermanos, por el apoyo constante. Por que apostaron por mí y eso me ha dado fuerzas para dar siempre lo mejor de mí. Gracias a ellos pude concluir mi carrera.

4. Por último, quiero agradecer a mis amigos(as) y a todos aquellos que durante todo este tiempo compartieron muchos sueños junto conmigo, los cuales conforme fue pasando el tiempo se hicieron realidad, por su compañía y los gratos momentos. Este presente trabajo es un sueño hecho realidad.

*Magdalena J. A.*

## INDICE

RESUMEN .....	vii
ABSTRAC .....	viii
INTRODUCCIÓN .....	ix
<b>CAPÍTULO I PLAN DE INVESTIGACIÓN .....</b>	<b>1</b>
1.1. EL PROBLEMA .....	1
1.1.1. DEFINICIÓN DEL PROBLEMA .....	1
1.1.2. FORMULACIÓN DEL PROBLEMA .....	3
1.2. JUSTIFICACIÓN DE LA INVESTIGACIÓN .....	3
1.3. OBJETIVOS DE LA INVESTIGACIÓN .....	4
1.3.1. OBJETIVO GENERAL .....	4
1.3.2. OBJETIVOS ESPECÍFICOS .....	5
1.4. HIPÓTESIS .....	5
1.4.1. HIPÓTESIS GENERAL .....	5
1.5. LIMITACIONES DEL PROBLEMA .....	5
<b>CAPÍTULO II MARCO TEÓRICO .....</b>	<b>6</b>
2.1. ANTECEDENTES DE LA INVESTIGACIÓN .....	6
2.2. BASE TEÓRICA .....	8
2.3. DEFINICIÓN DE TÉRMINOS BÁSICOS .....	41
2.4. OPERACIONALIZACIÓN DE VARIABLES .....	50

<b>CAPÍTULO III MATERIALES Y MÉTODOS .....</b>	<b>51</b>
3.1. POBLACIÓN .....	51
3.2. MUESTRA.....	51
3.3. MÉTODOS DE RECOPIACIÓN DE DATOS .....	51
3.4. MÉTODOS DE TRATAMIENTOS DE DATOS .....	52
3.5. METODOLOGÍA DE LA INVESTIGACIÓN .....	52
<b>CAPITULO IV RESULTADOS Y DISCUSION.....</b>	<b>54</b>
4.1. RESULTADOS DE RECOPIACIÓN DE INFORMACIÓN.....	54
4.2. COMPORTAMIENTO DE LAS VARIABLES EN ESTUDIO .....	54
4.3. SELECCIÓN DE VARIABLES INDEPENDIENTES .....	56
4.4. VALIDACION DEL MODELO .....	58
<b>CONCLUSIONES .....</b>	<b>66</b>
<b>RECOMENDACIONES Y SUGERENCIAS .....</b>	<b>67</b>
<b>BIBLIOGRAFÍA .....</b>	<b>68</b>
<b>ANEXOS .....</b>	<b>71</b>

## ÍNDICE DE TABLAS

<b>Tabla N° 1</b>	Análisis de la varianza .....	25
<b>Tabla N° 2</b>	Promedio, desviación estándar, valores mínimos, máximos covarianza, para la precipitación pluvial en la ciudad de Puno periodo 1979 - 2012 .....	55
<b>Tabla N° 3</b>	Análisis de varianza para el modelo final para la precipitación pluvial en el distrito de Puno periodo 1979 - 2012.....	56
<b>Tabla N° 4</b>	Modelo final para la precipitación pluvial en el distrito de Puno periodo 1979-2012 .....	57
<b>Tabla N° 5</b>	Análisis de varianza mostrando la descomposición para el modelo de la precipitación pluvial en el distrito de Puno periodo 1979-2012.....	57
<b>Tabla N° 6</b>	Especificación del modelo a través del análisis de la prueba de Reset de Ramsey. ....	60
<b>Tabla N° 7</b>	El contraste de white.....	64
<b>Tabla N° 8</b>	Análisis de la prueba de autocorrelación .....	64

## INDICE DE FIGURAS

<b>FIGURA Nº 1</b> Análisis de residuos en función a la temperatura superficial del mar en la región niño 3 .....	59
<b>FIGURA Nº 2</b> Análisis de residuos en función a los vientos en la superficie del océano .....	60
<b>FIGURA Nº 3</b> Análisis de la normalidad de los errores .....	61
<b>FIGURA Nº 4</b> Análisis de los residuos cuadráticos en relación a los valores estimados de la precipitación pluvial de la ciudad de Puno. ....	63



## RESUMEN

El pronóstico de la precipitación pluvial es muy complejo y difícilmente se puede atribuir a una sola variable los resultados en otra puesto que sus orígenes en la interacción océano – atmósfera. En este sentido la presente tesis se ha desarrollado con el objetivo de “determinar las variables principales que influyen en la precipitación pluvial en la ciudad de Puno”.

Teniendo como base los datos de precipitación anual y las variables climáticas Anomalía de la Temperatura Superficial del Mar, Radiación de Onda Larga, Vientos en la Superficie del Océano e Índice de Oscilación del Sur durante el periodo 1979 - 2012. El modelo estimado se determinó, siguiendo el procedimiento de la regresión lineal múltiple, utilizando el método de mínimos cuadrados para obtener los estimadores del modelo y selección de variables a través del método de Stepwise.

Finalmente se concluyó que, de las 7 variables independientes consideradas inicialmente, solo 2 variables fueron seleccionadas, llegándose así a estimar el modelo:  $Precipitacion = 1374.1367 - 22.8173 * Niño3 - 6.7482 * Vientos$

Las variables climáticas temperatura superficial del mar de la región Niño 3 y vientos en la superficie del océano influyeron significativamente en la precipitación pluvial mostrando un valor  $P = 0.0064$ .

**PALABRAS CLAVES:** Modelo; Puno; Precipitación; Variables climáticas  
Vientos; Niño 3.

## ABSTRACT

The forecast rainfall is very complex and can hardly be attributed to a single variable results in another it has its origins in the interaction ocean - atmosphere. In this sense this thesis has been developed in order to "identify key variables that influence rainfall in the city of Puno."

On the basis of data of annual precipitation and climatic variables Anomaly of Sea Surface Temperature, Long Wave Radiation, Wind on the sea surface and the Southern Oscillation Index for the period 1979 -. 2012 The estimated model was determined following the method of multiple linear regression, using the method of least squares estimators for model and variable selection via Stepwise method.

Finally it was concluded that, of the 7 independent variables initially considered, only two variables I were selected, thus reaching estimate the model:

$$\textit{Precipitacion} = 1374.1367 - 22.8173 * \textit{Niño}3 - 6.7482 * \textit{Vientos}$$

Weather variables sea surface temperature of child 3 region and winds in the ocean surface significantly influenced the rainfall showing a value  $P = 0.0064$ .

KEYWORDS: Model; Puno; precipitation; Winds variable weather; Child 3.

## INTRODUCCIÓN

**Lavado W., Silvestre E. & Obando O. (2009)** “*Impacto de ENOS en la Hidrología del Perú (1968-2006)*”. *Revista Peruana GEO-Atmósfera RPGA*, núm. 49, pp. 48-57, en los últimos años se ha observado una variación muy marcada en el comportamiento de algunos fenómenos meteorológicos, traducidos especialmente en la frecuencia e intensidad de las lluvias, sequías, tormentas severas, entre otros.

**(SENAMHI, 1999)**, el recurso agua en forma natural no siempre se encuentra disponible en el lugar, tiempo y cantidad que se requiere, generando frecuentemente situaciones de déficit o superávits, las cuales se tornan extremas cuando ocurren fenómenos climáticos como el fenómeno el Niño, que tiene origen en la interacción océano – atmósfera. El caso de Perú esta se asocia con torrenciales lluvias e inundaciones en la costa norte y deficiencias de precipitaciones en la sierra Sur y el Altiplano es afectado con sequias en diferentes regiones.

En consecuencia, se han puesto en marcha una serie de proyectos con miras a profundizar el conocimiento de sus mecanismos genéticos y dinámica espacio-temporal, involucrando no sólo el campo climático estricto, sino caracterizando los variables que lo relacionan con el campo oceanográfico, como la temperatura superficial del Mar, Radiación de onda larga, vientos en la superficie del Océano e índice de oscilación del sur.

Las modificaciones que mostró el régimen pluviométrico de la provincia de Puno durante los últimos años, provocaron fuertes impactos ambientales con particularidades novedosas, alejadas del comportamiento habitual.

**(SENAMHI, 2009)**, en la provincia de Puno, la ganadería y agricultura juegan un papel importante no solo como elemento de su seguridad alimentaria sino también como una base esencial de su economía aprovechando sus potencialidades biogeofísicas. Un conocimiento adecuado de las condiciones climáticas es una herramienta fundamental para lograr una organización y un uso sostenible y eficiente de los recursos suelo, agua y para las actividades productivas.

Sequías prolongadas afectan la producción agrícola llevando a la malnutrición e inanición tanto para la población humana como para el ganado. Las inundaciones pueden arruinar las reservas de comida, remover la fértil capa superficial del suelo y dañar por varias estaciones posteriores la producción de comida con similares impactos, contribuyendo significativamente a la severidad y extinción de la reducción de la producción de comida llevando con esto a una severa escasez. Al respecto en el presente análisis se busca determinar qué variables principales influyen significativamente en la precipitación pluvial en la ciudad de Puno.

La estructura de la presente investigación es la siguiente:

En el Capítulo I se realizó la identificación del problema, planteamiento de objetivos e hipótesis de investigación.

En el Capítulo II se desarrollo la búsqueda de la información disponible que nos ayudó en la resolución del problema, la teoría disponible de los métodos estadísticos que se utilizaron para resolver el problema.

En el Capítulo III se observo la metodología utilizada para la resolución del problema: recolección de datos, procesamiento de datos, obtención de resultados según los objetivos planteados.

En el Capítulo IV se realizo el análisis y discusión de los datos obtenidos según los objetivos e hipótesis planteados.

Finalmente se explican las principales conclusiones de la investigación y se dan las recomendaciones y sugerencias para próximas investigaciones.

Por último se presenta la Bibliografía y los Anexos de la investigación.

## CAPÍTULO I

### PLAN DE INVESTIGACIÓN

#### 1.1. EL PROBLEMA

##### 1.1.1. DEFINICIÓN DEL PROBLEMA

Existen instituciones que dedican buena parte de sus actividades al estudio de variables climáticas observando su evolución y desarrollo, tal es el caso del Servicio Nacional de Meteorología e Hidrología Puno - SENAMHI. Pues la ciencia tiene una exigencia, responder a los requerimientos de la sociedad, de forma dinámica y permanente. El entendimiento de los procesos físicos, comportamientos naturales e impactos, es de mucha importancia pues el conocimiento de lo existente permite plantear soluciones a los impactos expresadas en decisiones y acciones.

Se está viendo que el futuro de la humanidad está en riesgo. El cambio climático es un peligro para la sociedad en general.

**(SENAMHI, 2009)**, la observación meteorológica desempeña un papel de gran importancia al proporcionar información necesaria para planificar, a su vez permite proteger a la población de daños causados por fenómenos naturales catastróficos, como la precipitación (sequias

inundaciones extremas), los cuales generan cuantiosos daños y pérdidas de vidas e interrumpe la producción de bienes y servicios.

Estos fenómenos tienen sus orígenes en la interacción océano-atmósfera, que manifiesta sus efectos en el continente, alterando el régimen de precipitaciones. Según el SENAMHI, en el caso de Perú esta se asocia con torrenciales lluvias e inundaciones en la costa norte y deficiencias de precipitaciones en la sierra sur y el Altiplano es afectado con sequias en diferentes regiones.

Puno es una ciudad Altiplánica que está a 3824m.s.n.m., considerada una de las zonas más sensibles y perturbadas por la variabilidad climática, según estudios del SENAMHI, en la que actualmente se observan anomalías en el comportamiento de la precipitación, debido a que ya no se presentan en el periodo esperado, es por ello que existe el interés de conocer qué variables climáticas; (anomalía de la temperatura superficial del mar, radiación de onda larga, índice de oscilación del sur, vientos en la superficie del océano) están influyendo principalmente en dicho comportamiento.

La participación de estas variables climáticas en la presente investigación es debido a la interacción que existe entre la atmosfera y el océano, y por la teoría del efecto mariposa.

El efecto mariposa hace referencia a que una simple ala de mariposa hoy produce un diminuto cambio en el estado de la atmósfera. Después de un cierto tiempo, el comportamiento de la atmósfera cambia del que debería de haber tenido, por ejemplo muchas veces se dice que cuando

nieva en las alturas, aquí en la ciudad de Puno se siente un viento suave pero muy frío, lo mismo ocurre cuando nubes oscuras cubren el cielo y se espera una lluvia fuerte, pero basta que venga un viento fuerte para que las nubes se desplacen a otro lugar y llueva en un lugar en el que no se esperaban precipitaciones y en el que se esperaba no se presente, lo que ocurre en el otro lado del mundo puede cambiar mucho o influir en el comportamiento de una variable climática en este caso la precipitación, es por ello el interés en investigar cuál de estas variables influyen principalmente en el comportamiento de la precipitación en la ciudad de Puno.

El pronóstico de la precipitación pluvial es muy complejo y difícilmente se puede atribuir a una sola variable los resultados en otra, la realidad obliga a reconocer que para predecir con mayor precisión la precipitación pluvial, es necesario observar e integrar en el análisis otras variables que también pueden estar relacionadas.

Por lo tanto surge el interés de conocer:

### **1.1.2. FORMULACIÓN DEL PROBLEMA**

¿Cuáles son las variables principales que influyen significativamente en la precipitación pluvial en la ciudad de Puno, periodo 1979-2012?

## **1.2. JUSTIFICACIÓN DE LA INVESTIGACIÓN**

Siendo política del SENAMHI Puno, generar información que demandan los diversos sectores para la realización de sus estudios y proyectos de inversión, prestar servicios especializados, para contribuir al desarrollo sostenible, la seguridad y el bienestar nacional, mediante estudios e



investigaciones científicas. La precipitación es una parte importante del ciclo hidrológico, responsable del depósito de agua dulce en el planeta y, por ende, de la vida en nuestro planeta, tanto de animales como de vegetales que requieren del agua para vivir.

La presente investigación, se realizó para determinar las variables principales que influyen en la precipitación pluvial, para que una vez identificado se pueda prestar mayor atención en la observación de esta o estas variables de tal manera que esto ayude a mejorar los pronósticos y que estos sean más acertados, sobre la presencia o ausencia de las precipitaciones pluviales en la ciudad de Puno.

La precipitación pluvial es un evento climático muy difícil de predecir por qué no siempre se presenta en el lugar, tiempo y cantidad que se requiere, generando frecuentemente situaciones de desconcierto, debido a que existe la necesidad de planificar las actividades humanas en función del factor clima.

Si se logra identificar una relación causa-efecto entre ellos, la auscultación constante de la variable causa (predictor) permitiría pronosticar la variable efecto.

### **1.3. OBJETIVOS DE LA INVESTIGACIÓN**

#### **1.3.1. OBJETIVO GENERAL**

Determinar las variables principales que influyen en la precipitación pluvial en la ciudad de Puno, periodo 1979-2012.

### 1.3.2. OBJETIVOS ESPECÍFICOS

- Describir las variables climáticas de la precipitación pluvial en la ciudad de Puno, período 1979-2012.
- Determinar un modelo con las variables climáticas principales que influyen significativamente en la precipitación pluvial en la ciudad de Puno, periodo 1979–2012.

## 1.4. HIPÓTESIS

### 1.4.1. HIPÓTESIS GENERAL

La Anomalía de la temperatura superficial del mar de la región niño 3 y los vientos en la superficie del océano influyen significativamente en la precipitación pluvial en la ciudad de Puno, periodo 1979-2012.

## 1.5. LIMITACIONES DEL PROBLEMA

El presente trabajo de investigación tuvo como limitación el acceso a la información de las variables climáticas que influyen en la precipitación en vista de que el departamento de Puno y el Perú, no registra dicha información pues no cuenta con la tecnología ni los instrumentos necesarios para la observación y registro de los mismos.

Asimismo otra limitación fue el poco conocimiento de cada una de las variables que intervienen en la precipitación en la presente investigación, como son anomalía de la temperatura superficial del mar, radiación de onda larga, índice de oscilación del sur y vientos en la superficie del océano.

## CAPÍTULO II

### MARCO TEÓRICO

#### 2.1. ANTECEDENTES DE LA INVESTIGACIÓN

Los antecedentes al proyecto son los siguientes:

- **Mamani Sacachipana, Cristobal.** Tesis “Estimación de Intensidad de Lluvia en Zonas sin Información Pluviográfica”. Su objetivo fue Estimar las intensidades de la precipitación a partir de la precipitación máxima en 24 horas y la precipitación total máxima mensual en zonas sin información Pluviográfica. Concluye que la serie de precipitaciones máximas en 24 horas y las precipitaciones totales máximas mensuales de las estaciones Azángaro, Muñani y Putina de 37 años de información, es suficiente, homogénea y consistente para el análisis de frecuencia de eventos extremos.
- **Paredes Quispe, Juan R.** Tesis “Modelamiento Estocástico de la Precipitación Pluvial en el Distrito de Puno, periodo 1966-1990”. Su objetivo fue Ajustar los Valores de Precipitación Pluvial Promedio Anual del Distrito de Puno, de los periodos 1966 a 1990 a un Modelo Estocástico. Concluye que según el correlograma general para modelos de Markov I, II Y III y los límites de confianza calculados, se elige el modelo estocástico de Markov III como bueno para la búsqueda de mejor aproximación a los datos seriales de precipitación.

Cuyos coeficientes de autoregresión calculados fueron: 0.271935, -0.137857 y 0.213713 respectivamente los que determinan para llegar a obtener la serie residual o variable estocástica independiente del modelo de Markov III. Con una probabilidad del 95%. Cuyo modelo estocástico de tercer orden es:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \alpha_3 x_{t-3} + \epsilon_t$$

Donde:

$\alpha_1, \alpha_2$  y  $\alpha_3$ : Primer, segundo y tercer coeficiente de autorregresión y la variable estocástica dependiente:

$$\epsilon_t = \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \alpha_3 \epsilon_{t-3} + [1 - (\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + 2\alpha_1\alpha_2r_1 + 2\alpha_1\alpha_3r_2 + 2\alpha_2\alpha_3r_3)]^{1/2} \epsilon_t$$

- **Perez Quispe, Samuel D.** Tesis “Selección de la mejor ecuación de regresión para pronósticos de Temperatura extrema Mínima en función de las observaciones de Termómetro Seco y Termómetro Húmedo en el Distrito de Puno; periodo 1980-1990”. Su objetivo fue determinar la mejor ecuación de Regresión, para pronósticos de Temperatura Extremas Mínimas en función de los Termómetros seco y Termómetros húmedo. Para los datos observados en el periodo 1980-1990 de la ciudad de Puno. Concluye que las variables que dan un aporte significativo al pronóstico de las temperaturas mínimas, son las temperaturas registradas con: Termómetro Seco de las 7 a.m. ( $x_1$ ) y Termómetro Húmedo de las 7am. ( $x_4$ ) 19 p.m. ( $x_6$ ). Siendo la ecuación de pronóstico para las temperaturas extremas mínimas de:

$$\hat{y} = -0.4355 - 0.1285x_2 + 0.9329x_4 + 0.2401x_6$$

## 2.2. BASE TEÓRICA

### 2.2.1. MODELO

Es la representación o la imagen esquemática, simplificada de la realidad o expresión matemática de una determinada teoría. Constituyen instrumentos de descripción que permiten sacar conclusiones por analogía y una explicación o teoría de su funcionamiento que permite a la vez suponer que podemos predecir en algún grado su comportamiento quizá controlarlo en cierta forma.

Un modelo es una representación simplificada de un sistema complejo, que expresa relación entre variables y parámetros, pudiendo ser físicos, análogos y matemáticos.

**Kendall Buckland (1989)**, También define al modelo como una expresión formalizada de una teoría o una situación causal, que se considera haya sido la generadora de los datos observados.

### 2.2.2. MODELAMIENTO ESTADÍSTICO

Es la aproximación de ajuste de datos a una ecuación de regresión.

La regresión estudia el grado de dependencia entre dos variables o más variables, una llamada independiente o explicativa que es representada por "X" y las otras "Y" llamadas dependientes o explicada.

Usualmente se supone que la variable de regresión  $X$  es continua, controlable y fija. Fija significa que  $X$  tiene valores que son fijados es decir escogidos o predeterminados por el investigador.

### 2.2.3. MODELOS ESTADÍSTICOS DE REGRESIÓN

El análisis de regresión es una **técnica estadística** para investigar y **modelar la relación entre variables**.

Para llegar a determinar una ecuación que relacione las variables, un aspecto esencial del análisis de regresión es la recolección, recopilación o adquisición de datos que muestren los correspondientes valores de las variables consideradas en el análisis. Todo análisis de regresión es tan bueno como lo son los datos sobre los que se basa.

El paso siguiente es representar los puntos  $(X,Y)$  en el sistema de coordenadas, cuyo sistema de puntos resultantes se llama a veces diagrama de dispersión. Con el diagrama de dispersión es posible frecuentemente representar una curva que se aproxime a los datos, tal curva se llama curva de aproximación.

El problema general de encontrar ecuaciones de curvas de aproximación que se ajusten al conjunto de datos es el buscar la curva de ajuste.

**2.2.4. MODELOS LINEALES.**- Son ecuaciones de regresión de primer grado:

#### 2.2.4.1. MODELO DE REGRESIÓN LINEAL SIMPLE

Estudia el grado de dependencia lineal entre dos variables, una llamada independiente o “explicativa” que es representada por  $X$ , y la otra  $Y$  llamada variable dependiente o “explicativa”, cuyo modelo estadístico es:

$$Y = a + bX + e$$

Dónde:

$Y$ : Variable de respuesta o variable dependiente.

$X$ : Variable predictor o variable independiente.

$a$ : Es el valor de la variable dependiente cuando la variable independiente es cero.

$b$ : Es el cálculo de la inclinación de la línea, tasa de cambio de la variable dependiente cuando cambia en una unidad la variable independiente, y se denomina coeficiente de regresión.

$e$ : Error aleatorio.

Si  $y_i$  es cualquier observación de la variable dependiente con  $x_i$  observaciones independientes, entonces:

Si consideramos:

$E(y_i) = a + bx_i$ , entonces el modelo es:

$$y_i = a + bx_i + e_i$$

El modelo cumple las siguientes condiciones:

- $E(e_i) = 0$ .
- $V(e_i) = \sigma^2, \forall_i$
- La covarianza:

$$\text{Cov}(e_i, e_j) = 0 \text{ para } i \neq j$$

$$E(y_i) = a + bx_i$$

$$V(y_i) = V(e_i) = \sigma^2$$

**Calzada Benza (1998:120)**, “si una característica  $Y$  depende cuando menos en parte de otra característica, decimos que son características concaminantes. El grado de dependencia se mide con la **REGRESION**. La regresión es lineal, cuando al variar los valores de la característica independiente, los valores correspondientes de la característica dependiente tienden a variar con alguna proporcionalidad, en este caso la línea recta puede representar bastante bien la posición de todos los puntos fijados, uno con cada par de valores de ambas características.

Se presentan problemas de regresión cuando un investigador se pregunta por ejemplo, si existe influencia entre el uso del tabaco y las afecciones cardiacas, la regresión permite estudiar la dependencia de una característica respecto de la otra, para establecer como varia el promedio de la primera característica.

Si  $Y_i$  representa las medidas de la característica dependiente y  $X_i$  las de la característica independiente; pueden presentarse los casos siguientes:

- Que las medidas o niveles de  $X$  sean seleccionados o escogidos por el investigador; tal por ejemplo: si se estudia la longevidad de conejos sometidos a niveles escogidos de radiación, si se estudian los pesos de los habitantes de un país correspondientes a las alturas escogidas de los mismos,



etc. Si los niveles de  $X$  son escogidos, se puede estudiar la regresión de  $Y$  sobre  $X$ , mas no la de  $X$  sobre  $Y$ , por que los niveles de  $X$  no representan a la distribución de  $X$  de la población bivariada normal.

- Que las medidas o niveles de  $X$  sean tomados al azar; tal por ejemplo si se toma al azar habitantes de un país y de cada uno se toma la altura y el peso; si se estudia la velocidad de los autos y el numero de metros que requieren para frenar, etc. Si las medidas de  $X$  no son escogidas, se puede estudiar la regresión de  $Y$  sobre  $X$  y la de  $X$  sobre  $Y$ , porque las medidas de las dos características representan a las distribuciones de  $X$  e  $Y$  de la población bivariada normal muestreada.
- Distribución bivariada normal es aquella que tiene dos distribuciones. La distribución de los valores de  $Y$  es normal para cualquier de  $X$ , y la distribución de los valores de  $X$  es normal para cualquier valor de  $Y$ . Además, ambas distribuciones tienen varianzas homogéneas a lo largo de sus respectivas líneas de regresión.

La literatura ofrece varios métodos para estimar el grado de asociación entre dos variables, nosotros trataremos el Método de los mínimos cuadrados.

### 2.2.4.2. ESTIMACIÓN

Para la estimación de los parámetros utilizaremos el método de los mínimos cuadrados.

La estimación por el método de los mínimos cuadrados minimiza la suma de cuadrados de desviaciones de los valores observados de  $y_i$ , esto es: (Minimiza el error).

$$e'e = \sum_{i=1}^N e_i^2$$

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_N \end{bmatrix}$$

El error aleatorio denotado por  $e$  cuya distribución de probabilidad es normal con media cero y varianza  $\sigma^2$ ,  $e \sim N(0, \sigma^2)$ .

**Ecuaciones Normales:**

$$N\hat{a} + \hat{b} \sum x_i = \sum y_i$$

$$\hat{a} \sum x_i + \hat{b} \sum x_i^2 = \sum x_i y_i$$

Para estimar  $\hat{a}$  y  $\hat{b}$ :

$$\hat{a} = \bar{Y} - \hat{b}\bar{X}$$

$$\hat{b} = \frac{\sum x_i y_i - N\bar{Y}\bar{X}}{\sum x_i^2 - N\bar{X}^2}$$

Donde:

$$\bar{X} \cdot \frac{\sum x_i}{N} \bar{Y} = \frac{\sum y_i}{N} .”$$

### 2.2.4.3. REGRESIÓN LINEAL MÚLTIPLE

**Anderson, Hair, Tatham, Black (1999. P. 11)**, “la regresión múltiple es el método de análisis apropiado cuando el problema del investigador incluye una única variable métrica dependiente que se supone está relacionada con una o más variables métricas independientes. El objetivo del análisis de regresión múltiple es predecir los cambios en la variable dependiente en respuesta a cambios en varias de las variables independientes. Este objetivo se consigue muy a menudo a través de la regla estadística de los mínimos cuadrados.

La regresión múltiple es útil siempre que el investigador este interesado en predecir la cantidad o la magnitud de la variable dependiente. Por ejemplo, se puede hacer la predicción de los gastos mensuales de cenar fuera de casa (variable dependiente) con información referente a la renta familiar, su tamaño y la edad del cabeza de familia (variables independientes). De la misma forma, el investigador puede intentar predecir las ventas de una campaña a partir de información sobre sus gastos en publicidad, el número de vendedores y el número de tiendas que distribuyen sus productos.

## ¿Qué es el análisis de regresión múltiple?

**Anderson, Hair, Tatham, Black (1999: 144–145)**, “el análisis de regresión múltiple es una técnica que puede utilizarse para analizar la relación entre una única **variable criterio (criterio)** y varias **variables independientes (predictores)**. El objetivo del análisis de regresión múltiple es usar las variables independientes cuyos valores son conocidos para predecir la única variable criterio seleccionada por el investigador. Cada variable predictor es ponderada, de forma que las ponderaciones indican su contribución relativa a la predicción conjunta. Al calcular las ponderaciones, el procedimiento del análisis de regresión asegura la máxima predicción a partir del conjunto de variables. Estas ponderaciones facilitan también la interpretación de la influencia de cada variable en la realización de la predicción, aunque la correlación entre las variables independientes complica el proceso de interpretación. El conjunto de variables independientes ponderadas es conocido también como valor teórico **de la regresión**, una combinación lineal de las variables independientes que predice mejor la variable criterio la ecuación de regresión, también denominado como el valor teórico de la regresión, es el ejemplo de valor teórico más ampliamente reconocido entre todas las técnicas multivariantes.

Como ya se menciona anteriormente el análisis de regresión múltiple es una técnica de dependencia. Por tanto, al utilizarla, deberíamos de ser capaces de dividir las variables entre

independientes y dependientes. El análisis de regresión es también una herramienta estadística que debería utilizarse solo cuando tanto las variables dependientes como las independientes son métricas.

Sin embargo, bajo ciertas circunstancias, es posible incluir datos no métricos para las variables independientes (transformando los datos ordinales o los nominales en variables ficticias) o la variable criterio (mediante el uso de una medida binaria en la técnica especial de la regresión logística).

En resumen, al aplicar el análisis de regresión múltiple, los datos deben ser métricos o apropiadamente transformados.

El objetivo del análisis de regresión es predecir una única variable criterio a partir del conocimiento de una o más variables independientes. Cuando el problema implica una única variable independiente, la técnica estadística se denomina regresión simple. Cuando el problema implica dos o más variables independientes, se denomina regresión múltiple

El análisis de regresión múltiple es una técnica de dependencia”.

**Modelos de regresión múltiple (cf. 2013: 39)**, La regresión lineal múltiple es una técnica que intenta modelar probabilísticamente el valor esperado de una variable Y, a partir de los valores de dos o más predictores.

Determinar la posibilidad de predecir a través de una expresión muy simple el valor de la respuesta de interés, a partir de los

valores observados de una serie de variables. ¿Son los predictores propuestos adecuados para modelar en forma lineal la respuesta de interés?

Determinar la importancia relativa de la asociación lineal entre la respuesta y un predictor respecto a la asociación entre ella y otro predictor. ¿Cuáles de los predictores propuestos son los eficaces para modelar la respuesta en forma sencilla?

Estimar la relación lineal entre los predictores y la variable respuesta a partir de nuestros datos: ¿Cuál sería el modelo lineal que recomendaríamos más adecuado, sencillo, pero relativamente preciso?

La regresión lineal múltiple es matemáticamente similar a la regresión lineal simple.

#### 2.2.4.4. MODELO DE REGRESION LINEAL MULTIPLE

**Montgomery C., Peck Elizabeth. & Vining Geoffrey (2005:78–79)**, explica que “un modelo de regresión múltiple es formulado:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} \\ + e_i,$$

$$i = 1, 2, \dots, n \quad (1)$$

Donde:

$$x_{0i} = 1$$

- $i$ : Subíndice para la  $i$ -ésima observación ( $i=1,2,\dots,n$ )
- $Y$ : Precipitación (mm). (Variable respuesta o variable dependiente).
- $X_{1i}$ : Es la variable anomalía de la temperatura superficial del mar región Niño 1+2 ( $^{\circ}\text{C}$ ) (variable predictor o independiente), para la  $i$ -ésima observación.
- $X_{2i}$ : Es la variable anomalía de la temperatura superficial del mar región Niño 3 ( $^{\circ}\text{C}$ ) (variable predictor o independiente), para la  $i$ -ésima observación.
- $X_{3i}$ : Es la variable anomalía de la temperatura superficial del mar región Niño 3.4 ( $^{\circ}\text{C}$ ) (variable predictor o independiente), para la  $i$ -ésima observación.
- $X_{4i}$ : Es la variable anomalía de la temperatura superficial del mar región Niño 4 ( $^{\circ}\text{C}$ ) (variable predictor o independiente), para la  $i$ -ésima observación.
- $X_{5i}$ : Es la variable radiación de onda larga ( $\text{W}/\text{m}^2$ ) (variable predictor o independiente), para la  $i$ -ésima observación.
- $X_{6i}$ : Es la variable vientos en la superficie del Océano ( $\text{m}/\text{s}$ ) (variable predictor o independiente), para la  $i$ -ésima observación.

$X_{7i}$ : Es la variable índice de oscilación del Sur(hPa) (variable predictor o independiente), para la  $i$ -ésima observación.

$Y_i$ :  $i$ -ésima observación para la Precipitación.

$n$ : Tamaño de la muestra.

$\beta_0, \beta_1, \dots, \beta_k$ : Parámetros Lineales de regresión desconocidos.

Los parámetros  $\beta_j$ ,  $j = 0, 1, \dots, 7$ , se llaman coeficientes de regresión. El parámetro  $\beta_j$  representa el cambio esperado en la respuesta  $Y$  por cambio unitario en  $X_j$  cuando todas las demás variables regresoras  $X_i$  ( $i \neq j$ ) se mantienen constantes. Por esta razón a los parámetros  $\beta_j$  se les llama con frecuencia coeficientes de regresión parcial.

$e_i$ :  $i$ -ésimo Error aleatorio no observable asociado con  $Y_i$ ,  $\text{lose}_i \sim DNI(0, \sigma^2)$ .

Ahora definimos las siguientes matrices:

$$x = \begin{bmatrix} 1 & x_{11} & \dots & x_{17} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{n7} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_7 \end{bmatrix}$$

La Ecuación (1) puede escribirse en forma matricial:

$$Y = X\beta + e$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{17} \\ 1 & x_{21} & \dots & x_{27} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{n7} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_7 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$



Para estimar  $\beta$  utilizaremos el método de los mínimos cuadrados.

La ecuación de regresión estimada usando el método de mínimos cuadrados es:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_K X_{ki} \quad i = 1, 2, \dots, n$$

Ecuaciones normales:

$$X'Y = (X'X)\hat{\beta}$$

De aquí:  $\hat{\beta} = (X'X)^{-1}X'Y$

Matricialmente el modelo general es escrito como:

Donde:

$$X'Y = \begin{bmatrix} \sum X_{i0} Y_i \\ \sum X_{i1} Y_i \\ \sum X_{i2} Y_i \\ \vdots \\ \sum X_{ik} Y_i \end{bmatrix}, X'X = \begin{bmatrix} \sum X_{i0}^2 & \sum X_{i0} X_{i1} & \sum X_{i0} X_{i2} & \sum X_{i0} X_{i3} \dots \sum X_{i0} X_{ik} \\ \sum X_{i0} X_{i1} & \sum X_{i1}^2 & \sum X_{i2} X_{i1} & \sum X_{i3} X_{i1} \dots \sum X_{ik} X_{i1} \\ \sum X_{i0} X_{i2} & \sum X_{i1} X_{i2} & \sum X_{i2}^2 & \sum X_{i3} X_{i2} \dots \sum X_{ik} X_{i2} \\ \dots & \dots & \dots & \dots \\ \sum X_{i0} X_{ik} & \sum X_{i1} X_{ik} & \sum X_{i2} X_{ik} & \sum X_{i3} X_{ik} \dots \sum X_{ik}^2 \end{bmatrix}$$

Siendo:

$$X_{i0} = 1, \forall i=1, \dots, n$$

$$\sum X_{i0}^2 = n$$

$$\sum X_{i0} X_{i1} = \sum X_{i1} = X_{.1}$$

$$\sum X_{i0} Y_i = \sum Y_i = Y_{.}$$

$$X'Y = \begin{bmatrix} \sum Y_i \\ \sum X_{i1}Y_i \\ \sum X_{i2}Y_i \\ \sum X_{i3}Y_i \\ \vdots \\ \sum X_{ik}Y_i \end{bmatrix}, X'X = \begin{bmatrix} N & \sum X_{i1} & \sum X_{i2} & \sum X_{i3} & \dots & \sum X_{ik} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i2}X_{i1} & \sum X_{i3}X_{i1} & \dots & \sum X_{ik}X_{i1} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 & \sum X_{i3}X_{i2} & \dots & \sum X_{ik}X_{i2} \\ \sum X_{i3} & \sum X_{i1}X_{i3} & \sum X_{i2}X_{i3} & \sum X_{i3}^2 & \dots & \sum X_{ik}X_{i3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \sum X_{ik} & \sum X_{i1}X_{ik} & \sum X_{i2}X_{ik} & \sum X_{i3}X_{ik} & \dots & \sum X_{ik}^2 \end{bmatrix}$$

Las matrices de varianzas y covarianzas estimadas para los parámetros estimados se distribuyen; sabiendo que:

$$\hat{\beta}_i \sim N[\beta_i, a_{ii}\sigma^2],$$

Donde:

$a_{ii}$ : Es el elemento  $i$  – ésimo ( $i=0, \dots, k$ ) de la diagonal de  $(X'X)^{-1}$ , de aquí

$$var[\hat{\beta}_i] = a_{ii}\sigma^2$$

Siendo su estimador:

$$\widehat{Var}(\hat{\beta}_i) = a_{ii}\hat{\sigma}^2$$

**2.2.4.5. PRUEBA DE HIPOTESIS EN REGRESION LINEAL MULTIPLE:**

Hay varios procedimientos de prueba de hipótesis que demuestran su utilidad para contestar estas preguntas. Las pruebas formales requieren que los errores aleatorios sean independientes y tengan una distribución normal con promedio  $E(\varepsilon_i) = 0$  y una  $Var(\varepsilon_i) = \sigma^2$ .

#### 2.2.4.6. PRUEBA DE LA SIGNIFICANCIA DE LA REGRESION

La prueba de la **significancia de la regresión**, es para determinar si hay una **relación lineal** entre la respuesta  $Y$  y cualquiera de las variables regresoras  $X_1, X_2, \dots, X_k$ . Este procedimiento suele considerarse como una prueba general o global de la adecuación del modelo. Las hipótesis pertinentes son:

1. Prueba de significancia global del modelo de regresión múltiple estimado, es decir, encontrar si todos los coeficientes parciales son iguales a cero simultáneamente.

$$H_0: \beta_0 = \beta_1 = \beta_2, \dots, = \beta_k = 0$$

Esta hipótesis nula es conjunta de que  $\beta_1, \beta_2, \dots, \beta_k$  son iguales a cero en forma conjunta o simultánea. Una prueba de tal hipótesis se denomina prueba de **significancia global** de la recta de regresión observada o estimada, es decir si  $Y$  está relacionada o no linealmente con  $X_1, X_2, \dots, X_K$  a la vez.

$$H_1: \beta_k \neq 0 \text{ al menos para un } k.$$

El rechazo de la hipótesis nula implica que al menos uno de los regresores  $X_1, X_2, \dots, X_K$  contribuye al modelo en forma significativa. El procedimiento de prueba se realiza mediante el análisis de varianza, donde la suma total de cuadrados  $SS_T$  se divide en una suma de cuadrados debido a la regresión,  $SS_R$  y una suma de cuadrados de residuales  $SS_{Res}$  así:

$$SS_T = SS_R + SS_{Res}$$

Se demuestra que si  $H_0$  es cierta, entonces  $SS_{R/\sigma^2}$  tiene una distribución  $\chi_k^2$ , con la misma cantidad de grados de libertad que la cantidad de variables regresoras en el modelo. También se demuestra que  $SS_{Res/\sigma^2}$  tiene una distribución  $\chi_{n-k-1}^2$  y que  $SS_{Res}$  y  $SS_R$  son independientes.

El estadístico de prueba:

$$F_0 = \frac{SS_R/k}{SS_{Res}/(n-k-1)} = \frac{MS_R}{MS_{Res}}$$

Tiene la distribución  $F_{k,n-k-1}$ .

Una fórmula de cálculo para  $SS_R$  se deduce partiendo de:

$$SS_{Res} = y'y - \hat{\beta}'x'y$$

Y ya que

$$SS_T = \sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} = y'y - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

Se puede escribir la ecuación anterior en la forma

$$SS_{Res} = y'y - \frac{(\sum_{i=1}^n y_i)^2}{n} = \left[ \hat{\beta}'xy - \frac{(\sum_{i=1}^n y_i)^2}{n} \right]$$

O bien

$$SS_{Res} = SS_T - SS_R$$

Por consiguiente, la suma de cuadrados de la regresión es:

$$SS_R = \hat{\beta}' X' y - \frac{\left( \sum_{i=1}^n y_i \right)^2}{n}$$

La suma de cuadrados de residuales es

$$SS_{Res} = y' y - \hat{\beta}' X' y$$

Y la suma total de cuadrados es

$$SS_T = y' y - \frac{\left( \sum_{i=1}^n y_i \right)^2}{n}$$

El procedimiento de prueba se resume normalmente en una tabla de análisis de varianza”.

**El método del análisis de varianza en las pruebas de significancia global de una regresión múltiple observada: la prueba F**

**Searle, S.R., (1971: 95)** explica “esta hipótesis conjunta puede ser probada por la técnica de análisis de varianza (ANOVA)”.

Tabla Nº 1 Análisis de la varianza

FUENTE DE VARIACIÓN	GRADOS DE LIBERTAD	SUMA DE CUADRADOS	CUADRADO S MEDIOS	ESTADÍSTICO
Regresión	$k$	$SS_R$	$MS_R$	$\frac{MS_R}{MS_{Res}}$
Residuales	$n - k - 1$	$SS_{Res}$	$MS_{Res}$	
Total	$n - 1$	$SS_T$		

Fuente: SEARLE, S.R., Linear Models, Wiley, Nueva York, 1971.

Donde:

k: Numero de variables independientes en el modelo.

#### 2.2.4.6.1. PRUEBA SOBRE COEFICIENTES INDIVIDUALES DE REGRESIÓN

**Gujarati N. (2004:209)**, sostiene “una vez determinado que al menos uno de los regresores es importante, la pregunta lógica es ¿cuál(es) sirve(n) de ellos? Si se agrega una variable a un modelo de regresión, la suma de cuadrados de la regresión aumenta, y la suma de cuadrados residuales disminuye.

Se debe decidir si el aumento de la suma de cuadrados de la regresión es suficiente para garantizar el uso del regresor adicional en el modelo. La adición de un regresor también aumenta la varianza del valor ajustado  $\hat{y}$ , por lo que se debe tener cuidado de incluir sólo regresores que tengan valor para

explicar la respuesta. Además, si se agrega un regresor no importante se puede aumentar el cuadrado medio de residuales, y con eso se disminuye la utilidad del modelo.

La hipótesis para probar la significancia de cualquier coeficiente individual de regresión, como por ejemplo  $\beta_k$  son:

$$H_0: \beta_k = 0$$

$$H_1: \beta_k \neq 0$$

Si no se rechaza  $H_0: \beta_k = 0$ , quiere decir que se puede eliminar el regresor  $x_k$  del modelo. El estadístico de prueba para esta hipótesis es:

$$t_0 = \frac{\hat{\beta}_k}{\sqrt{\hat{\sigma}^2 c_{kk}}} = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)}$$

Donde  $c_{kk}$  es el elemento diagonal de  $(X'X)^{-1}$  que corresponde a  $\hat{\beta}_k$ . Se rechaza la hipótesis nula  $H_0: \beta_k = 0$  si

$$|t_0| > t_{\alpha/2, n-k-1}.$$

#### 2.2.4.7. MEDIDAS DE LA BONDAD DE AJUSTE

Una vez que se ha realizado el ajuste por mínimos cuadrados, conviene disponer de algún indicador que permita medir el grado de ajuste entre el modelo y los datos. En el caso de que se hayan estimado varios modelos alternativos podría utilizarse medidas de este tipo, a las que se denomina medidas de la bondad de ajuste, para seleccionar el modelo más adecuado”.

#### 2.2.4.8. COEFICIENTE DE DETERMINACIÓN

**Montgomery C., Peck Elizabeth y Vining Geoffrey (2005: 37-38)**, sostiene que “el coeficiente de determinación se basa en la descomposición de la varianza de la variable endógena, a la que denominaremos varianza total. Cuando se aplica el método de mínimos cuadrados y existe un término independiente en el modelo, se puede establecer que:

Varianza total = varianza explicada + varianza residual

Una propiedad importante de  $R^2$  es que es una función no decreciente del número de variables explicativas o de regresoras presentes en el modelo; a medida que aumenta el número de regresoras,  $R^2$  aumenta casi invariablemente y nunca disminuye. Planteado de otra forma, una variable adicional  $X$  no reducirá  $R^2$ .

Analíticamente, y prescindiendo de los términos del denominador, la igualdad anterior se puede expresar de esta forma:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$

Se llama coeficiente de determinación. Como  $SS_T$  es una medida de la variabilidad de  $Y$  sin considerar el efecto de la variable regresora  $X$  y  $SS_{Res}$  es una medida de la variabilidad de  $Y$  que queda después de haber tenido en consideración a  $X$ ,  $R^2$  se llama, con frecuencia, la proporción de la variación explicada por el regresor  $X$ . Ya que  $0 \leq SS_{Res} \leq SS_T$ , entonces  $0 \leq R^2 \leq 1$ . Los



valores de  $R^2$  cercanos a 1 implican que la mayor parte de la variabilidad de  $Y$  esta explicada por el modelo de regresión.

Hay otras ideas erróneas acerca de  $R^2$ . En general  $R^2$  no mide la magnitud de la pendiente de la línea de regresión. Un valor grande de  $R^2$  no implica que la pendiente sea grande, además,  $R^2$  no mide la adecuación del modelo lineal, porque con frecuencia  $R^2$  es grande aunque  $X$  y  $Y$  no tengan relación lineal”.

#### 2.2.4.8.1. COEFICIENTE DE DETERMINACIÓN CORREGIDO

Montgomery C., Peck Elizabeth y Vining Geoffrey (2005:82), “el coeficiente de determinación corregido permite comparar modelos con distinto número de regresores. Analíticamente viene dado por:

$$R_{Adj}^2 = 1 - \frac{SS_{Res}/(n-r)}{SS_T/(N-1)}$$

#### ***R<sup>2</sup> y R<sup>2</sup> Ajustada***

Gujarati Damodar (2004:210), “en vista de que  $SS_{Res}/(n-r)$  es el cuadrado medio de residuales y  $SS_T/(n-1)$  es constante, independientemente de cuantas variables hay en el modelo,  $R_{Adj}^2$  solo aumentara al agregar una variable al modelo si esa adición reduce el cuadrado medio residual. La ***R<sup>2</sup> Ajustada*** penaliza la adición de términos que no son útiles,

además que es ventajoso para evaluar y comparar los modelos posibles de regresión.

En vista de esto, al comparar dos modelos de regresión con la misma variable dependiente pero un número diferente de variables  $X$ , se debe de tener mucho cuidado al escoger el modelo con la más alta  $R^2$ .

Para comparar dos términos  $R^2$ , se debe tener en cuenta el número de variables  $X$  presentes en el modelo. Esto puede hacerse fácilmente si se considera un coeficiente alternativo, que es el  $R_{Adj}^2$

El  $R^2$  así definido se conoce como  $R^2$  ajustado, denotado por  $R_{Adj}^2$  el término ajustado significa ajustado por los grados de libertad asociados con las sumas de los cuadrados”.

#### 2.2.4.8.2. EL JUEGO DE MAXIMIZACION DE $R_{Adj}^2$

**Gujarati Damodar (2004:208)**, “para concluir esta sección debe hacerse una advertencia: algunas veces los investigadores juegan maximizar el  $R_{Adj}^2$ , es decir, escogen el modelo que da el  $R_{Adj}^2$  más elevado. Pero esto puede ser peligroso, ya que en el análisis de regresión el objetivo no es obtener un  $R_{Adj}^2$  más elevado sino más bien obtener estimados de los verdaderos coeficientes de regresión poblacional de los cuales se pueda depender y sea posible realizar inferencia estadística sobre ellos. En el análisis empírico es inusual

obtener un  $R_{Adj}^2$  muy elevado, sino encontrar que algunos de los coeficientes de regresión no son estadísticamente significativos o muestran signos contrarios a los esperados a priori. Por consiguiente el investigador debe preocuparse más por la relevancia lógica o teórica que tienen las variables explicativas para la variable dependiente y por su significancia estadística.

Si en este proceso se obtiene un  $R_{Adj}^2$  elevado, muy bien; por otra parte, si  $R_{Adj}^2$  es bajo, esto no significa que el modelo sea necesariamente malo.

De hecho, Goldberger es muy crítico sobre el papel de  $R^2$  ha dicho: desde nuestra perspectiva, el  $R^2$  tiene un papel muy modesto en el análisis de regresión, y es una medida de la bondad del ajuste de una regresión lineal MC (mínimos cuadrados) de una muestra en un cuerpo de datos.

Nada en el modelo exige que  $R^2$  sea elevado. Por tanto, un  $R^2$  elevado es evidencia en favor del modelo y un  $R^2$  bajo no es evidencia en su contra.

De hecho lo más importante sobre el  $R^2$  es que este no es importante en el modelo. El modelo tiene que ver con parámetros en una población, no con la bondad de ajuste en la muestra. Si se insiste en una medida de predecir el éxito (o más bien el fracaso), entonces  $\sigma^2$  es el error de predicción esperado al cuadrado que resultaría si la población fuera

utilizada como predictor, en forma alterna, el error estándar de predicción elevado al cuadrado para valores relevantes de  $x$  (regresores) puede ser informativo”.

#### 2.2.4.9. REGRESION STEPWISE

**Netter J, & Wasserman W. (1990: 234)**, “la regresión Stepwise, forma una secuencia de regresiones agregando o eliminando variables independientes en cada etapa, el criterio para agregar o eliminar una variable puede ser en base estadística  $F^*$  o en base a coeficientes de correlaciones parciales. El procedimiento se puede resumir en lo siguiente:

- a. Se calculan todas las regresiones simples para cada una de las

$k - 1$  Variables independientes potenciales.

Para cada una de las ecuaciones de regresión simple se obtiene la estadística:

$$F^* = \frac{MSR_{(x_k)}}{MSE_{(x_k)}}$$

La variable independiente con el mayor valor  $F^*$  es considerada, si este valor  $F^*$  excede a un nivel predeterminado, la variable independiente debe ser considerada en la regresión, de otra, manera el programa termina, considerándose que ninguna variable

independiente brinda la suficiente ayuda como para ser considerada en el modelo de regresión.

- b. Supongamos que la variable  $x_7$  fue la variable independiente considerando para entrar en la primera etapa, la rutina de regresión Stepwise ahora calcula todas las regresiones con dos variables independientes donde  $x_7$  es una de ellas. Para cada una de tales regresiones se calcula la estadística:

$$F^* = \frac{MSR_{(x_k/x_7)}}{MSE_{(x_k, x_7)}}$$

Esta es la estadística que se utiliza para probar la hipótesis

$H_0: \beta_k = 0$  Cuando  $x_7$  y  $x_k$  son las variables independientes del modelo, las variables independientes con el mayor valor  $F^*$  es considerada en la segunda etapa. Si este valor  $F^*$  excede a un nivel predeterminado, se tiene la segunda variable a ser introducida en el modelo, de otra manera el programa termina.

- c. Supongamos que  $x_7$  fue la variable que es agregada en el segundo paso, ahora la rutina Stepwise examina si cualquiera de las otras variables independientes que ya están en el modelo deben ser eliminadas. En nuestra ilustración hay hasta en esta etapa solamente una variable independientes en el modelo  $x_7$ , de manera que solamente la estadística:

$$F^* = \frac{MSR_{(x_7/x_3)}}{MSE_{(x_7, x_3)}}$$

Debe ser estimada.

- d. Supongamos  $x_7$  fue retenida, de manera que ambas  $x_3$  y  $x_7$  están consideradas en el modelo, ahora la rutina Stepwise examina que variable independiente es la próxima candidata a ser agregada, luego examina cuál de las variables independientes que ya están en el modelo deben ser eliminadas y así sucesivamente hasta que no exista variable independiente que pueda ser introducida o eliminada, en este punto la búsqueda termina”.

#### **2.2.4.10. PRUEBA DE SUPUESTOS ESTADÍSTICOS PARA LA REGRESIÓN**

Gran parte de los procesos estadísticos para su buena aplicación, hacen uso de algunas suposiciones estadísticas, estas suposiciones son de mucha importancia porque determinará el éxito o el fracaso del procedimiento estadístico aplicado, para la validez en la aplicación y elaboración de pronósticos con la ecuación de regresión lineal múltiple.

##### **A. PRUEBA PARA LA ESPECIFICACIÓN DEL MODELO**

##### **PRUEBA RESET DE RAMSEY:**

**Damodar Gujarati (1994:455)**, resalta “Ramsey ha propuesto una prueba general de errores de especificación conocida como

RESET de Ramsey (Prueba del error de especificación en regresión). Aquí se ilustrara solamente la versión más simple de la prueba. Suponiendo una función lineal de la siguiente forma:

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i}$$

Donde:

$Y$ : Es la variable dependiente.

$X$ : Es la variable independiente., si se grafican los residuales  $\hat{u}_i$  obtenidos de esta regresión frente a  $\hat{Y}_i$ , el estimado de  $Y_i$  de este modelo. Y si el incremento en el  $R^2$  es estadísticamente significativo, esto sugerirá que la función lineal estaba mal especificada. Esta es esencialmente la idea que soporta la prueba de RESET. Los pasos considerados en RESET son los siguientes:

1. A partir del modelo seleccionado, obténgase el  $Y_i$  estimado, es decir  $\hat{Y}_i$ .
2. Nuevamente, efectúese la regresión introduciendo  $\hat{Y}_i$ , en alguna forma, como uno o varios regresores adicionales.
3. Sea el  $R^2$  obtenido " $R_{nuevo}^2$ ", y aquel obtenido con el modelo sea " $R_{viejo}^2$ ". Entonces, se puede utilizar la prueba  $F$  introducida inicialmente:

$$F = \frac{(R_{nuevo}^2 - R_{viejo}^2) / \text{número de regresores nuevos}}{(1 - R_{nuevo}^2) / (n - \text{número de parámetros en el modelo nuevo})}$$

Para averiguar si el incremento en el  $R^2$  obtenido, es estadísticamente significativo.

4. Si el valor  $F$  calculado es significativo, por ejemplo, al nivel del 5%, se puede aceptar la hipótesis de que el modelo está mal especificado”.

## B. PRUEBA DE NORMALIDAD DEL ERROR

**Anderson, L., Hair, F., Tatham E.& Black c., (1999: 63)**, sostiene “el supuesto fundamental del análisis multivariante es la **normalidad** de los datos, en referencia al perfil de la distribución de los datos para una única variable métrica y su correspondencia con una **distribución normal**, punto de referencia de los métodos estadísticos. Si la variación respecto de la distribución normal es suficientemente amplia, todos los test estadísticos resultantes no son validos, dado que se requiere la normalidad para el uso de los estadísticos de la  $t$  y de la  $F$ . tanto los métodos estadísticos univariantes como los multivariantes analizados en este texto se basan en el supuesto de la normalidad univariante, suponiendo también los multivariantes la normalidad multivariante. La normalidad univariante para una única variable es fácil de contrastar, siendo posible varias mediadas correctoras. Dicho de forma sencilla, la normalidad multivariante (la combinación de dos o más variables) implica que las variables individuales son normales en un sentido univariante y que sus combinaciones también sean normales. Por tanto, si una variable es una normal multivariante, es también normal univariante. Sin embargo, lo



contrario no es necesariamente cierto (dos o más variables normales univariantes no son necesariamente normal multivariante). Por tanto, una situación en la que todas las variables exhiben normalidad univariante ayudara a obtener normalidad multivariante, aunque no lo garantiza. La normalidad multivariante es mucho más difícil de contrastar, aunque existen varios test para situaciones en las que la técnica multivariante se ve particularmente afectada por una violación de los supuestos”.

**Análisis grafico de la normalidad.-** la prueba de la normalidad se realizó de forma grafica, mediante probability plot, y nos apoyaremos en el teorema del límite central que garantiza la normalidad por tener datos  $n > 30$ .

### **C. PRUEBA JI-CUADRADA DE LA BONDAD DE AJUSTE**

**Bernard Ostle (1990:151),**“Una cosa que se hace frecuente, sin más justificación que el decir que parece razonable, es suponer que la variable aleatoria en discusión sigue una distribución particular. Por ejemplo frecuentemente se supone que los datos son muestras de una población normal y con justa razón objetar esa suposición. En esta ocasión se presentara un procedimiento útil para comprobar la validez de tales suposiciones.

El procedimiento consiste en hacer comparaciones entre el número real de observaciones y el número esperado de las mismas (esperados bajo la “suposición”), para varios valores de la variable aleatoria. Los números esperados usualmente se

calculan usando la distribución supuesta, haciendo los parámetros iguales a sus muestras estimadas. La estadística **ji-cuadrada** se calcula de acuerdo con la ecuación:

$$\chi^2 = \sum_{i=1}^4 \frac{(O_{ij} - E_i)^2}{E_i}$$

Y los grados de libertad serán  $k - p - 1$ , donde  $p$  representa al número de parámetros estimados por la muestra estadística. Por ejemplo, si una suposición de normalidad estuviese bajo prueba,  $\mu$  y  $\sigma^2$  se estimarían por  $\bar{X}$  y  $S^2$ , y los grados de libertad serían  $k - 3$ , donde  $k$  representa el número de intervalos de clase usados a fijar la distribución”.

#### **D. PRUEBA DE HOMOGENEIDAD DE VARIANZAS DE LOS ERRORES (HOMOCEASTICIDAD)**

**Anderson, L., Hair, F., Tatham E.&Black c., (1999: 67)**, la prueba de homocedasticidad de dos variables métricas se evalúa mejor gráficamente. La aplicación más común de esta forma de evaluación se produce en la regresión múltiple, en relación con la dispersión de la variable dependiente a lo largo de las variables independientes métricas. Dado que el eje del análisis de las regresiones es el valor teórico, el gráfico de residuos se usa para revelar la presencia de homocedasticidad (o su opuesto, heterocedasticidad, desigual dispersión de la varianza).

Para determinar si los errores tienen o no una variación homogénea de la forma gráfica.

**Análisis gráfico de la Homocedasticidad.-** consiste en preparar un diagrama de dispersión para los errores al cuadrado en el eje de las ordenadas, y los valores pronosticados ( $\hat{Y}_i$ ) en el eje de las abscisas. Si la nube de puntos presenta un patrón tendencial, se dice que existe heteroscedasticidad, en caso contrario existirá homocedasticidad”.

## E. PRUEBA DEL CONTRASTE DE WHITE

**Alfonso Novales (1993:205),**“sostiene que White (1980) ha propuesto un contraste general, que no precisa especificar la forma que puede adoptar la heteroscedasticidad:

- Estimar el modelo original por MCO, ignorando la posible heteroscedasticidad.
- Estimar una regresión del cuadrado de los residuos mínimo cuadráticos anteriores, sobre una constante, los regresores del modelo original, sus cuadrados y productos cruzados de segundo orden.
- Al aumentar el tamaño muestral, el producto  $TR^2$ , donde  $T$  es el tamaño muestral y  $R^2$  el coeficiente de determinación de la última regresión, sigue una distribución **ji-cuadrado** con  $p - 1$  grados de libertad, donde  $p$  es el número de regresores del modelo estimado en b).

Mientras que el tamaño muestral crece con el número de observaciones, el coeficiente de determinación tendera a cero

bajo la hipótesis nula de homocedasticidad. Solo cuando la varianza del termino de error depende (aunque sea de forma no lineal) de las variables explicativas del modelo, el coeficiente  $R^2$  no tendera a cero. En tal caso, el producto  $TR^2$  permanecera a un cierto nivel, lejos de cero, y es de esperar que supere al valor de tablas de la distribución ***ji-cuadrado***. Como puede verse, el contraste no solo requiere que los regresores no tengan un alto poder explicativo de  $\sigma_t^2$ , sino que el coeficiente de determinación de dicha regresión tienda hacia cero rápidamente”.

### **PRUEBA DE LA INDEPENDENCIA DEL ERROR**

Un supuesto importante del modelo lineal clásico, es que no existe autocorrelación o relación serial entre los errores. Si se viola este supuesto tenemos el problema de correlación serial o autocorrelación.

La autocorrelación se presenta por diferentes razones, como por ejemplo, la inercia de la mayoría de las series económicas de tiempo, el sesgo de especificación que resulta de excluir algunas variables relevantes del modelo o de la utilización de una forma funcional incorrecta, la exclusión de variables rezagadas y la manipulación de datos.

Aunque los estimadores mínimo cuadráticos, continúan siendo insesgados y consistentes en presencia de correlación, ellos dejan de ser eficientes. Como resultado, las pruebas de significancia usual t y F no pueden aplicarse legítimamente.

Para determinar si los errores están correlacionados, se utilizará la prueba de Durbin-Watson, para lo cual se seguirá el procedimiento siguiente:

$H_0$  : No hay correlación serial positiva

$H_1$  : Hay autocorrelación positiva.

Calculamos el estadístico Durbin-Watson “d”

$$d = \frac{\sum_{t=1}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad 0 \leq d \leq 4$$

Para un “n” dado y “k” (número de variables independientes) vamos a la tabla y elegimos un valor  $d_u$  (superior) y  $d_l$  (inferior). Tomar la siguiente decisión:

- i) si  $d < d_l$ , existe autocorrelación positiva
- ii) si  $d > d_u$ , no existe autocorrelación positiva
- iii) si  $d_l < d < d_u$ , la prueba no es concluyente.

Si los valores muestrales dedson superiores a 2, probamos la existencia de autocorrelación negativa considerando  $(4 - d)$  en lugar de “d”, análogamente al caso anterior.

## MULTICOLINEALIDAD

Se sospecha que la colinealidad está presente en situaciones en que  $R^2$  es alto (por ejemplo, entre 0.7 y 1) y cuando las

correlaciones de orden cero son altas y a la vez ninguno o pocos de los coeficientes de regresión parcial son individualmente significativos, con base en la prueba “t” convencional. Si  $\epsilon R^2$  es alto quiere decir que la prueba F del análisis de varianza, en la mayoría de los casos, rechazara la hipótesis nula de que el valor verdadero de todos los coeficientes parciales de la pendiente sea simultáneamente cero, independientemente de la prueba t.

### 2.3. DEFINICIÓN DE TÉRMINOS BÁSICOS

#### ADMINISTRACIÓN NACIONAL DE LA ATMÓSFERA Y DEL OCÉANO

**SENAMHI-Puno**, (cf. 2010:2), definen “parte del departamento de comercio (DOC) de los Estados Unidos. Ubicado en Silver Spring, Estado de Maryland. Es la oficina matriz del Servicio Nacional de Meteorología, con la responsabilidad de protección al medio ambiente con énfasis en los recursos marinos y atmosféricos”.

#### ADAPTACIÓN

**SENAMHI-Puno**, (cf. 2010:2), “ajuste de los sistemas humanos o naturales frente a entornos nuevos o cambiantes. La adaptación al cambio climático se refiere a los ajustes en sistemas humanos o naturales como respuesta a estímulos climáticos proyectados o reales, o sus efectos, que pueden moderar el daño o aprovechar sus aspectos beneficiosos. Se pueden distinguir varios tipos de adaptación, entre ellas la preventiva y la reactiva, la pública y privada, o la autónoma y la planificada”.

## ANOMALÍA

**SENAMHI-Puno, (cf. 2010:2),**“término genérico para identificar una condición diferente del valor medio climatológico. Se expresa como la diferencia (absoluta o normalizada) con respecto a la media”.

## ANOMALÍA DE LA TEMPERATURA SUPERFICIAL DEL MAR

**SENAMHI-Puno, (cf. 2010:2),**Anomalía de Temperatura de la Superficie del Mar, en las zonas más calientes, los movimientos ascendentes de aire húmedo dan nacimiento a extensas formaciones nubosas cuya condensación produce fuertes precipitaciones, mientras que en otras partes la subsistencia de aire seco y frío desde la tropósfera superior genera una especie de capa que impide la formación de nubes susceptibles de dar lluvias de consideración.

Los índices de temperaturas superficiales del mar, intentan captar las características del calentamiento oceánico en regiones que son escogidas de acuerdo a determinados criterios.

La región El Niño 1 está situada cerca de las costas de Perú y Ecuador, mientras que El Niño 2 se ubica en las cercanías de las Islas Galápagos. El efecto combinado de las anomalías de temperaturas en estas dos regiones se considera un indicador altamente representativo de los cambios inducidos por El Niño en los patrones de variabilidad de la costa del Pacífico de América del Sur. Las regiones El Niño 3 y El Niño 4 están situadas más alejadas de las costas de Suramérica pero coinciden las zonas donde se observan las máximas anomalías de temperaturas superficiales del mar. Estudios posteriores demostraron que una región

que comprendía parte de la región El Niño 3 y parte de la región El Niño 4, constituía un índice muy adecuado debido a que esa zona coincidía con la de mayor correlación entre la temperaturas superficiales del mar y el índice de Oscilación del Sur, por lo que definió una nueva subregión definida como El Niño 3.4.

## **ALTITUD**

**SENAMHI-Puno, (cf. 2010:2)**, Es la distancia vertical entre un punto situado sobre la superficie terrestre y el nivel medio del mar.

## **ANEMÓMETRO TOTALIZADOR**

**SENAMHI-Puno, (cf. 2010:2)**, Instrumento mecánico que determina el recorrido del viento y con ello la velocidad media del viento en un periodo determinado. Se expresa en km/h y m/s.

## **CALENTAMIENTO GLOBAL**

**SENAMHI-Puno, (cf. 2010:2)**, Es el aumento en la temperatura de la atmosfera terrestre y de los océanos en las últimas décadas causado por la acumulación de contaminantes en la atmosfera. Estos forman una capa gruesa que atrapa el calor del sol y causa el calentamiento del planeta. Los principales agentes contaminantes son el vapor de agua ( $H_2O$ ) y el dióxido de carbono ( $CO_2$ ), emitido por la quema de combustibles fósiles.

## **CAMBIO GLOBAL**

**SENAMHI-Puno, (cf. 2010:2)**, Conjunto de cambios ambientales afectados por la actividad humana, con especial referencia a cambios en



los procesos que determinan el funcionamiento del sistema Tierra. Se incluyen en este término aquellas actividades que, aunque ejercidas localmente, tienen efectos que trascienden el ámbito local y regional.

### **CAMBIO CLIMÁTICO**

**SENAMHI-Puno, (cf. 2010:2),** Son las variaciones medias de los valores de los elementos meteorológicos (temperatura, precipitación, humedad, etc.) de una amplia región, a lo largo de un periodo de tiempo, que provocan alteraciones en el clima de esa zona.

### **CLIMA**

**SENAMHI-Puno, (cf. 2010:2),** Se puede definir como las condiciones predominantes en el espacio atmosfera-superficie, las cuales están controladas por los variables radiactivos (radiación solar y efecto invernadero), por la interacción de los componentes del sistema Tierra (atmosfera, litosfera, hidrosfera, biosfera y antroposfera) y por características fisicogeograficas (latitud, altitud sobre el nivel del mar, orografía, etc.) del lugar o región específica.

### **ESTACIÓN METEOROLÓGICA**

**SENAMHI-Puno, (cf. 2010:2),** Una estación meteorológica, es una unidad de información básica. Es una zona destinada a explorar un sector atmosférico más o menos extenso. Por ello es fundamental seleccionar el emplazamiento y las condiciones de medida, a fin de que

las observaciones realizadas sean representativas de la climatología regional.

## ÍNDICE DE OSCILACIÓN DEL SUR

**SENAMHI-Puno, (cf. 2010:2)**, Es un valor que se obtiene de la diferencia de los valores superficiales de presión atmosférica entre la isla de Tahití y Darwin.

Es una medida comparativa de cómo cambia la presión atmosférica en dos grandes regiones, occidental y central-oriental, del Océano pacífico tropical.

## METEOROLOGÍA

**SENAMHI-Puno, (cf. 2010:2)**, Ciencia que estudia la atmósfera y los fenómenos atmosféricos. Comprende el estudio del tiempo y el clima y se ocupa del estudio físico, dinámico y químico de la atmósfera terrestre.

En tal sentido, la meteorología estudia las causas, naturaleza, evolución y relación de los fenómenos atmosféricos, así como las condiciones dominantes en un periodo de tiempo corto que es llamado tiempo atmosférico”.

## MUESTREO

**Cáceda F., Perez s. (2001: 83)**, define “es una parte representativa de la población. Para que una muestra sea representativa, y por lo tanto útil; debe de reflejar las similitudes y diferencias encontradas en la población;

es decir debe ejemplificar las características de la misma. Cuando decimos que la muestra es representativa indicamos que reúne aproximadamente las características de la población que son importantes para la investigación”.

## **OBSERVACIÓN**

**SENAMHI-Puno (cf. 2010:2)**, “en meteorología, es la evaluación de uno ó más elementos meteorológicos, como la temperatura, la presión, o el viento, que describen el estado de la atmósfera ya sea sobre o por encima de la superficie de la tierra. Un observador es quien registra las evaluaciones de los elementos meteorológicos.

## **PLUVIÓMETRO**

**SENAMHI-Puno, (cf. 2010:2)**, Instrumento que mide la cantidad de lluvia que ha caído. La unidad de medida es en milímetros.

## **PRECIPITACIÓN**

**SENAMHI-Puno, (cf. 2010:2)**, Es un conjunto de partículas acuosas, líquidas o sólidas, cristalizadas o amorfas, que caen de una nube o de un conjunto de nubes y que alcanzan el suelo. Esto incluye la lluvia, llovizna, granizo, la nieve, el rocío, la escarcha y bolillas de nieve.

## **POBLACIÓN**

**Cáceda F., Perez s. (2001: 99)**, Es la totalidad del universo que interesa considerar, y que es necesario que este bien definido, para que se sepa en todo momento que elementos lo componen”.

## **PRONÓSTICO**

**SENAMHI-Puno, (cf. 2010:2)**, Pronunciamento sobre sucesos futuros. El pronóstico del tiempo incluye el uso de modelos objetivos basados en algunos parámetros atmosféricos, unidos a la habilidad y experiencia del meteorólogo. También se conoce como predicción.

## **RADIACIÓN DE ONDA LARGA**

**SENAMHI-Puno, (cf. 2010:2)**, La atmósfera está compuesta por gases que presentan una banda de fuerte absorción en la región de onda larga (módulo de Teoría de la Radiación); éstos son, principalmente, dióxido de carbono, vapor de agua y ozono. Esto supone que, en condiciones de cielo despejado, estos gases emiten radiación de onda larga en las longitudes de onda correspondientes a esas bandas de absorción. La radiación es emitida en todas direcciones.

La cantidad de radiación de onda larga recibida en la superficie terrestre depende de la distribución vertical de temperatura, y de las concentraciones de dióxido de carbono, vapor de agua y ozono. Parece que la mayor parte de la radiación de onda larga del "cielo despejado" recibida por la superficie, se origina en los primeros cientos de metros de atmósfera. Más aún, las nubes emiten radiación como cuerpos negros en la región de onda larga.

## **RESIDUAL**

**Cáceda F., Perez s. (2001: 98)**, Es la diferencia entre el valor real observado y su valor de pronóstico.

## **SEQUÍA**

**SENAMHI-Puno, (cf. 2010:2)**, Condición climática anormalmente seca en un área específica que se prolonga debido a la falta de agua y causa un serio desbalance hidrológico.

## **TEMPERATURA**

**SENAMHI-Puno, (cf. 2010:2)**, Medida del movimiento molecular o el grado de calor de una sustancia. Se mide usando una escala arbitraria a partir del cero absoluto, donde las moléculas teóricamente dejan de moverse. Es también el grado de calor y de frío. En observaciones de la superficie, se refiere principalmente al aire libre o temperatura ambiental cerca a la superficie de la tierra.

## **TEMPERATURA SUPERFICIAL DEL MAR**

**SENAMHI-Puno, (cf. 2010:2)**, Desde fines del siglo XIX las observaciones de temperatura superficial del mar son rutinarias en los barcos. En la actualidad, gran parte de este tipo de información se obtiene mediante observaciones remotas desde satélites. Desde hace unos 15 años funciona a lo largo del Pacífico ecuatorial una red de boyas (unas 60 en total) ancladas al fondo marino, que entre otras variables atmosféricas y oceánicas permiten mantener un seguimiento continuo (diario) de las condiciones térmicas superficiales y sub-superficiales en esa región (ref. programa TAO/NOAA).

## TERMÓMETRO

**SENAMHI-Puno, (cf. 2010:2)**, Instrumento que sirve para medir la temperatura. Las diferentes escalas usadas en meteorología son: Celsius, Fahrenheit y Kelvin o Absoluta.

## VARIABLE

**Cáceda F., Perez s. (2001: 98)**, “es una función que asocia a cada elemento de la población; la medición particularmente de una característica que se puede observar”.

## VIENTOS DE LA SUPERFICIE DEL OCÉANO

**SENAMHI-Puno, (cf. 2010:2)**, Cuando el viento sopla sobre el océano, mueve el agua oceánica debido a la fricción con la superficie. Si la Tierra no girara, el agua se movería en la misma dirección del viento. Sin embargo, la Tierra gira, lo que complica un poco las cosas. La rotación de la Tierra hace que el agua de la superficie se mueva a la derecha de la dirección del viento en el hemisferio norte y hacia la izquierda de la dirección del viento en el hemisferio sur.

## VELOCIDAD DEL VIENTO

**SENAMHI-Puno, (cf. 2010:2)**, Es el promedio del movimiento del aire durante un periodo de tiempo preestablecido.

## VIENTO

**SENAMHI-Puno, (cf. 2010:2)**, Movimiento del aire en relación a la superficie terrestre, generalmente de manera horizontal. Hay cuatro aspectos del viento que se miden: dirección, velocidad, tipo (ráfagas y rachas) y cambios. Los cambios superficiales se miden con veletas y anemómetros mientras que los de gran altitud se detectan con globos piloto, radio vientos o reportes de la aeronáutica civil”.

**2.4. OPERACIONALIZACIÓN DE VARIABLES**

VARIABLES INDEPENDIENTES	INDICADOR	ÍNDICE O CARACTERÍSTICA
<ul style="list-style-type: none"> <li>Anomalía de la temperatura superficial del mar</li> </ul>	Niño 1+2	°C
	Niño 3	°C
	Niño 4	°C
	Niño 3.4	°C
<ul style="list-style-type: none"> <li>Radiación de onda larga</li> </ul>	Radiación de onda larga	$W/m^2$
<ul style="list-style-type: none"> <li>Vientos en la superficie del océano</li> </ul>	Vientos en la superficie del océano	m/s
<ul style="list-style-type: none"> <li>Índice de oscilación del sur</li> </ul>	Índice de oscilación del sur	hPa
VARIABLE DEPENDIENTE Precipitación pluvial	Precipitación	mm.

## CAPÍTULO III

### MATERIALES Y MÉTODOS

#### 3.1. POBLACIÓN

Estuvo constituida por la información de precipitación registrada por el Servicio Nacional de Meteorología e Hidrología SENAMHI Puno desde el año 1964 al 2012 y los eventos climáticos (anomalía temperatura del mar, radiación de onda larga, vientos en la superficie del océano e índice de oscilación), obtenidos de la base de datos de la Administración Nacional de la Atmósfera y Océanos (NOAA), registrado por dicha institución a través del tiempo.

#### 3.2. MUESTRA

La muestra no probabilística, estuvo comprendida por la precipitación registrada por el Servicio Nacional de Meteorología e Hidrología SENAMHI Puno correspondiente al periodo 1979-2012.

#### 3.3. MÉTODOS DE RECOPIACIÓN DE DATOS

Los datos de precipitación para la presente investigación fueron obtenidos de la Oficina General de Estadística e Informática del Servicio Nacional de Meteorología e Hidrología SENAMHI PUNO y la



Administración Nacional de la Atmósfera y Océanos (NOAA), los cuales corresponden al periodo 1979-2012.

La información fue obtenida en archivos, los cuales, se trabajó en base a promedios por año, para el periodo en estudio, para las variables precipitación, anomalía temperatura del mar, radiación de onda larga, vientos en la superficie del océano e índice de oscilación.

### **3.4. MÉTODOS DE TRATAMIENTOS DE DATOS**

Toda la información obtenida se registró en una base de datos, el tratamiento de los datos se realizó en base a promedios anuales para las variables predictoras (anomalía temperatura del mar, radiación de onda larga, vientos en la superficie del océano e índice de oscilación) y para la precipitación.

### **3.5. METODOLOGÍA DE LA INVESTIGACIÓN**

La metodología sigue el siguiente procedimiento:

- Recolección de información.
- Analizar el comportamiento de las variables en estudio.
- Determinación del Modelo: Esto de manera gráfica, para ver si los datos se ajustan a un modelo lineal, o no lineal.
- Ecuación de regresión Adecuada para pronósticos, siguiendo el procedimiento de la regresión Lineal Múltiple y la utilización del método de mínimos cuadrados para obtener los estimadores del modelo.

- Selección de variables independientes que están influyendo significativamente en la precipitación pluvial, para ello se usó el procedimiento de selección de variables STEPWISE.
- Verificación de supuestos estadísticos para la regresión lineal múltiple (Normalidad, Homogeneidad de las varianzas, Independencias, y Multicolinealidad); se utilizaran las pruebas de supuestos estadísticos para la regresión múltiple.

## CAPITULO IV

### RESULTADOS Y DISCUSION

#### 4.1. RESULTADOS DE RECOPIACIÓN DE INFORMACIÓN

Para el análisis de datos, se ha utilizado información de la precipitación pluvial, anomalías de la temperatura superficial del Mar en las regiones niño (1+2, 3,3.4 y 4) en la ciudad de Puno obtenidas de la Oficina General de Estadística e Informática del Servicio Nacional de Meteorología e Hidrología SENAMHI PUNO y la Administración Nacional de la Atmósfera y Océanos (NOAA) los cuales corresponden al periodo 1979-2012 Ver (Anexo N° 01).

#### 4.2. COMPORTAMIENTO DE LAS VARIABLES EN ESTUDIO

Una descripción de las variables en estudio se muestra en la tabla N°1, en los cuales se presenta el promedio, la desviación estándar, los valores mínimo y máximo de la precipitación pluvial.

**Tabla N° 2 Promedio, desviación estándar, valores mínimos, máximos covarianza, para la precipitación pluvial en la ciudad de Puno periodo 1979 - 2012**

LA VARIABLE	$\bar{X}$	$\sigma$	MIN	MAX	C.V.
TSM NIÑO 1+2	0.2376	11.21	-15.46	30.26	47.18
TSM NIÑO 3	1.2974	8.44	-14.01	20.2	6.51
TSM NIÑO 3.4	1.4594	7.62	-13.02	17.11	5.22
TSM NIÑO 4	2.2344	6.12	-11.14	11.73	2.74
ROL	495.61	119.08	240.05	706.7	0.24
VIENTOS	98.0853	20.55	49.3	130.3	0.21
IOS	-5.1765	12.82	-26.2	21.4	-2.47
PRECIPITACION	682.635	250.967	48.68	1051.9	0.37

Fuente: Preparado en base a los datos obtenidos del SENAMHI Puno, periodo 1979 – 2012.

**INTERPRETACIÓN:** En la tabla N° 2 se puede observar que la precipitación pluvial media en el periodo de estudio fue de 682.635 mm, con una desviación estándar de 250.967, oscilando la precipitación entre 1051.9 mm. y 48.68 mm.

Asimismo se aprecia que el promedio de la Temperatura Superficial del Mar de la Región Niño 1+2 en el periodo de estudio fue de 0.24°C, con una Temperatura Superficial del Mar de la Región Niño 1+2 máxima de 30.26 °C y una mínima de -15.46°C.

Por otro lado se puede apreciar que el promedio de la Temperatura Superficial del Mar de la Región Niño 3 en el periodo de estudio fue de 1.2974 °C. Asimismo se observa que la Temperatura Superficial del Mar de la Región Niño 3 varía entre 20.2 °C y -14.01°C.

También se observa que el promedio de la velocidad de los Vientos en la superficie del Océano en el periodo de estudio fue de 98.0853 m/s. con

una desviación estándar de 20.55 y una variación entre 130.3 m/s y 49.3 m/s.

#### 4.3. SELECCIÓN DE VARIABLES INDEPENDIENTES

**INTERPRETACIÓN:** En el modelo inicial fueron incluidas 7 variables independientes. Al aplicar el método de Stepwise y utilizando el Software Statgraphics Versión 6.1, este programa al igual que el SPSS versión 20, selecciono 2 de estas variables independientes las cuales fueron:

- Anomalías de la Temperatura Superficial del Mar Región Niño 3.
- Vientos en la superficie del Océano.

**Tabla Nº 3 Análisis de varianza para el modelo final para la precipitación pluvial en el distrito de Puno periodo 1979 - 2012**

FUENTE DE VARIACION	SUMA DE CUADRADOS	G.L.	CUADRADOS MEDIOS	F	P
MODELO	577355	2	288677	5.96154	0.0064
ERROR	1501110	31	48422.9		
TOTAL	2078465	33			

R-SQ. (ADJ.)= 0.231185      SE= 220.052      DurbWat = 1.77268

Los coeficientes de Regresión estimados y los elementos necesarios que justificaron su presencia en el modelo a través del nivel de significancia se muestran en los siguientes cuadros.

**Tabla N° 4 Modelo final para la precipitación pluvial en el distrito de Puno periodo 1979-2012**

VARIABLE INDEPENDIENTE	COEFICIENTES	ERROR ESTÁNDAR	t	P
CONSTANTE	1374.136713	275.209881	4.9931	0.0000
NIÑO 3	-22.817261	6.608317	-3.4528	0.0016
VIENTOS	-6.748205	2.715124	-2.4854	0.0185

R-SQ. (ADJ.)= 0.2312      SE= 220.052    DurbWat = 1.773

**INTERPRETACIÓN:** Como se observa en la Tabla N° 4, el modelo final fue estimado con un coeficiente de determinación  $R^2 = 23.12\%$ , valor supuestamente pequeño para un modelo. Sin embargo el análisis de varianza mostrado en la Tabla N° 3 proporcionó un valor  $p=0.0064$ , lo que indicó de que al menos una de estas variables explican significativamente la precipitación.

**Tabla N° 5 Análisis de varianza mostrando la descomposición para el modelo de la precipitación pluvial en el distrito de Puno periodo 1979-2012**

FUENTE DE VARIACIÓN	SUMA DE CUADRADOS	G.L.	CUADRADOS MEDIOS	F	P
MODELO	577355	2	288677	5.96154	
NIÑO 3	278234.215	1	278234.215	5.75	0.0064
VIENTOS	299120.738	1	299120.738	6.18	0.0227
ERROR	1501110	31	48422.9		0.0185
TOTAL	2078465	33			

**INTERPRETACIÓN:** La Tabla N° 5, nos muestra la descomposición de la Fuente de Variación correspondiente a la regresión, el cual muestra un aporte adicional significativo de cada variable cuando las restantes ya están incluidas en el modelo.

### **EFFECTO DE LA TEMPERATURA SUPERFICIAL DEL MAR DE LA REGION NIÑO 3:**

El coeficiente de Regresión asociada a esta variable se estimo - 22.817261, lo cual indicó que por cada grado celsius de Temperatura adicional la precipitación disminuye en 22.817261 mm.

### **EFFECTO DE LOS VIENTOS EN LA SUPERFICIE DEL OCEANO:**

Del mismo modo el coeficiente de Regresión, correspondiente a los vientos se estimo en -6.748205, lo cual indica que por cada incremento en m/s de la velocidad de los vientos en la superficie del Océano la precipitación disminuye en 6.748205 mm.

### **MODELO FINAL PARA LA PRECIPITACION**

El modelo de Regresión fue seleccionado de entre varios modelos, siendo la función de la Regresión estimada:

$$Precipitacion = 1374.1367 - 22.8173 * Niño3 - 6.7482 * Vientos$$

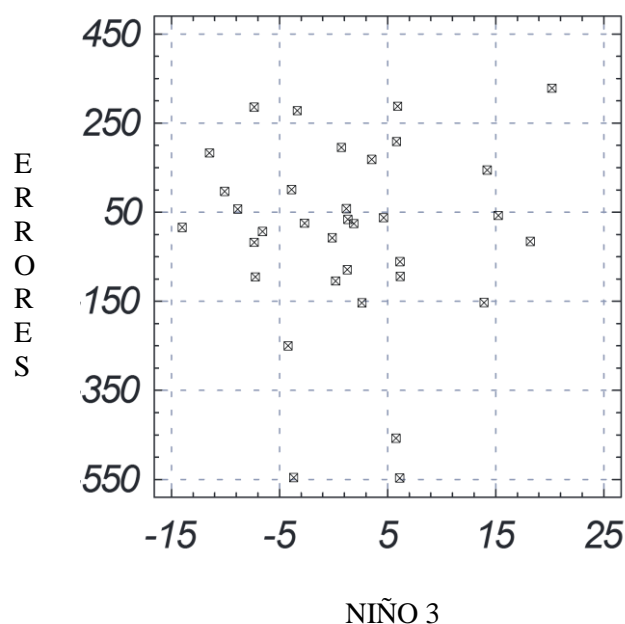
$$R^2_{Ajust} = 0.2312$$

#### **4.4. VALIDACION DEL MODELO**

Se procedió a la validación de las suposiciones correspondientes al modelo de regresión estimado, los resultados fueron:

## ESPECIFICACIÓN DEL MODELO

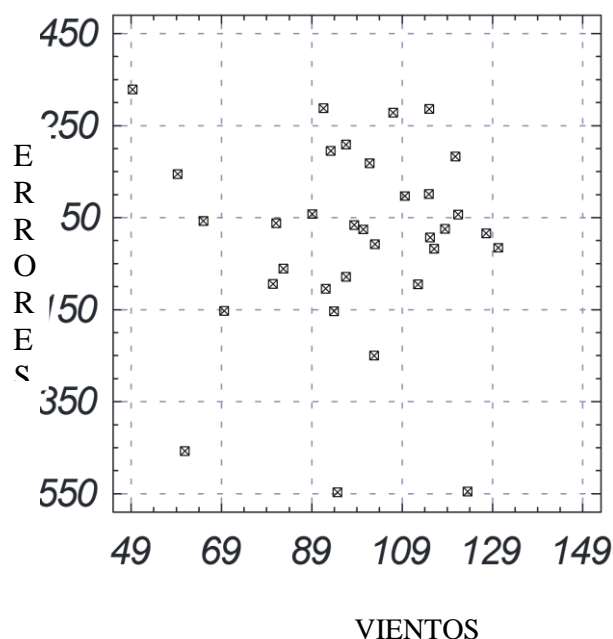
**FIGURA N° 1** Análisis de residuos en función a la temperatura superficial del mar en la región niño 3



**INTERPRETACIÓN:** Se sabe que al establecer una relación funcional indebida, se puede cometer el error de especificación del modelo. Para analizar este problema se procedió a graficar los residuales con cada una de las variables explicativas seleccionadas en el modelo, tal como se observa en las figuras N° 1 y N° 2 y deducimos la no existencia de patrones característicos, por lo que concluimos que el modelo está bien especificado.



**FIGURA N° 2 Análisis de residuos en función a los vientos en la superficie del océano**



**Tabla N° 6 Especificación del modelo a través del análisis de la prueba de Reset de Ramsey.**

VARIABLE INDEPENDIENTE	COEFICIENTE	STD. ERROR	T-VALOR	NIVEL SIG.
CONSTANTE	1491.512612	659.889157	2.2637	0.8318
LLUVIA.X2	-25.69146	16.831468	-1.6826	0.1195
LLUVIA.X6	-7.418454	4.374142	-1.6968	0.1002
VARIABLES.FITTED*VARIABLES.FITTED	-1.394E-7	7.0629E-7	-81974	0.8448

R-SQ. (ADJ.)= 0.2578 SE= 223.5443 DurbWat = 1.881

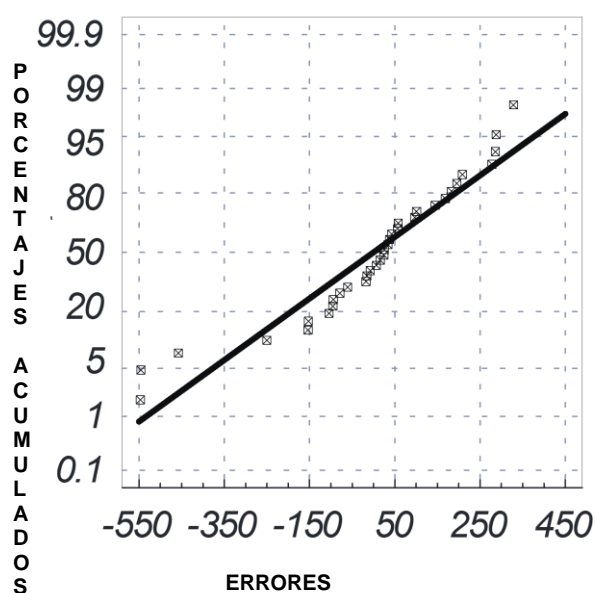
**INTERPRETACIÓN:** En la tabla N° 6 se observa el análisis numérico realizado mediante la prueba de RESET de Ramsey que estima la regresión original introduciendo además como regresor adicional a la estimación cuadrática, en este caso precipitación anual, esto debido a una relación cuadrática entre  $Y_i$  y los  $\hat{Y}_i$  esta prueba arrojó un coeficiente de determinación  $R^2 = 0.2578$ .

La estadística de prueba proporciona un valor  $F = 1.075$  el cual fue comparado con  $F_{Tabular (1,30,0.05)=4.17}$ , tal como se muestra en el cuadro

N° 07, que determinó decidir la no significancia en el aumento de  $R^2$  al utilizar la nueva regresión y aceptar la hipótesis de que el modelo está bien especificado.

## PRUEBA DENORMALIDAD

**FIGURA N° 3 Análisis de la normalidad de los errores**



**INTERPRETACIÓN:** El Gráfico de la prueba de la normalidad a través de errores presentado en la Figura N° 3, muestra algunas desviaciones de normalidad en las colas. Sin embargo, el problema de la no normalidad no se considero serio, ya que el tamaño de muestra que se utilizo nos garantiza con el apoyo del teorema del límite central este supuesto y así validar las pruebas “t” y “F”.

### **JI-CUADRADAPRUEBA DE LA BONDAD DE AJUSTE:**

#### **Hipótesis:**

$H_0$ : En la población de la que se extrajo la muestra, los errores siguen una distribución normal.

$H_a$ : En la población de la que se extrajo la muestra, los errores no siguen una distribución normal.

**Nivel de significancia:** 5%

**Estadística de prueba:**

$$\chi_c^2 = \sum_{i=1 \dots k}^4 \frac{(O_{ij} - E_i)^2}{E_i}$$

**Distribución de la estadística de prueba:**

Cuando  $H_0$  es verdadero  $\chi_c^2$  sigue aproximadamente una  $\chi_{k-p-1}^2$ , grados de libertad donde  $k$  representa al número de intervalos de clases usados a fijar la distribución,  $p$  representa el número de parámetros estimados por la muestra estadística.

**Decisión:**

**Sí:**

$$\chi_c^2 < \chi_T^2 \text{ aceptar } H_0$$

$$\chi_c^2 \geq \chi_T^2 \text{ rechazar } H_0$$

**INTERPRETACIÓN:** Se realizó la prueba de la bondad de ajuste utilizando el paquete STATGRAPHICS, el cual arrojó un valor  $\chi_c^2 = 4.46325$ , con 2 grados de libertad, 5 intervalos, la tabla arrojó un valor  $p = 0.10735$ , lo cual condujo a aceptar la  $H_0$ . Por tanto los errores siguen una distribución normal.

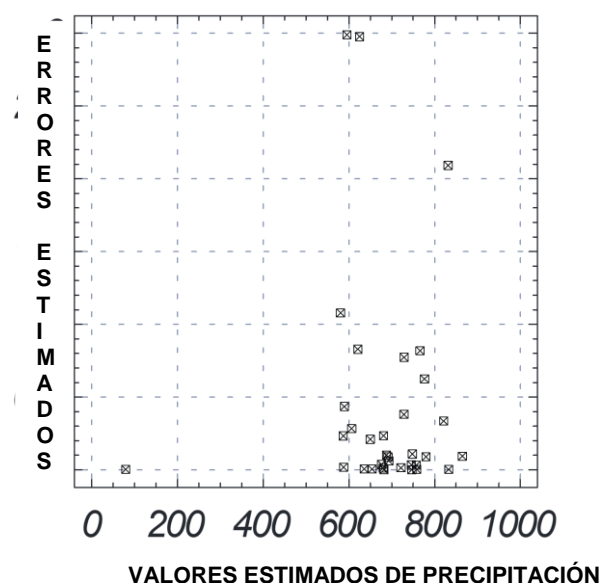
## MULTICOLINEALIDAD

**INTERPRETACIÓN:** Individualmente la correlación de las variables independientes con la precipitación pluvial fue de -0.3659 para la variable Niño 3 y de 0.0055 para la variable vientos.

Los factores VIF mayores a un valor 10 implican multivolinealidad. Mostrados en el Anexo N° 2, evidentemente la variable que presentó una mayor correlación fue la Temperatura Superficial del Mar en la Región Niño 3 a excepción de la variable vientos, el cual se correlaciona significativamente con la precipitación con un valor  $p= 0.333$ . Además para fundamentar lo dicho anteriormente el coeficiente de determinación  $R^2 = 23.12\%$  no indica una multicolinealidad que nos deba preocupar.

## HETEROCEDASTICIDAD

**FIGURA N° 4** Análisis de los residuos cuadráticos en relación a los valores estimados de la precipitación pluvial de la ciudad de puno.



**INTERPRETACIÓN:** El análisis gráfico de los residuos cuadráticos frente a los respectivos valores estimados por la precipitación mostrado

en la figura N° 4, permitió deducir la no existencia de un patrón sistemático entre las 2 variables, lo cual sugirió la inexistencia de la Heteroscedasticidad.

**Tabla N° 7 El contraste de white**

VARIABLE INDEPENDIENTE	COEFICIENTE	STD. ERROR	T-VALOR	NIVEL SIG.
CONSTANTE	9.6499	5.4278	1.7779	0.0863
LLUVIA.X2	11639.6551	1.5763	0.7384	0.4664
LLUVIA.X6	-2.0919	1.0701	-1.9548	0.0687
LLUVIA.X2*X6	-69.7654	125.2644	-0.5569	0.5820

R-SQ. (ADJ.)= 0.2578 SE= 223.5443 DurbWat = 1.881

**INTERPRETACIÓN:** Después del análisis gráfico, el análisis numérico realizado a través de la prueba de White proporcionó un coeficiente de determinación  $R^2 = 0.2578$ , produciendo de esta manera un valor  $TR^2 = 34(0.2578) = 8.7652$ , que comparado con una  $\chi^2_{Tabular}$  con 5 grados de libertad y con un valor de condianza del 97.5% determina un valor de 12.8, de esta manera se determinó que hay ausencia de heteroscedasticidad.

**AUTOCORRELACION**

**Tabla N° 8 Análisis de la prueba de autocorrelación**

**RESUMEN DEL MODELO<sup>c</sup>**

MODELO	R	R CUADRADO	R CUADRADO CORREGIDA	ERROR TÍP. DE LA ESTIMACIÓN	DURBIN-WATSON
1	,366 <sup>a</sup>	,134	,107	237,18601	
2	,527 <sup>b</sup>	,278	,231	220,05204	1,773

- a. Variables predictoras: (Constante), NIÑO3
- b. Variables predictoras: (Constante), NIÑO3, VIENTOS
- c. Variable dependiente: PRECIPITACION

En el cuadro N° 01 se muestra la hipótesis y el estadístico de Durbin Watson, para analizar la autocorrelación de los errores en el modelo estimado.

**CUADRO N° 01: PRUEBA “d” DURBIN WATSON**

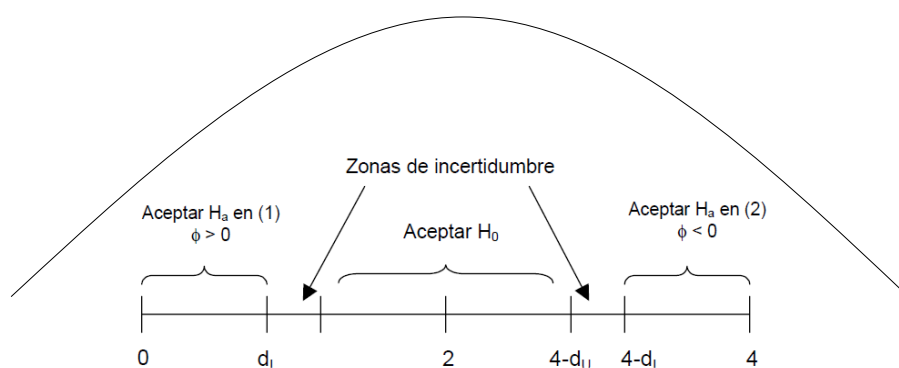
$H_0 = \text{No Autocorrelación}$

$H_1 = \text{Autocorrelación}$

$n = 34$  ,  $d = 1.773$

$d_L = 1.58$  ,  $d_U = 1.33$

*Como  $d_U < d < 4 - d_U$ , aceptamos  $H_0$*



**INTERPRETACIÓN:** De la tabla de Durbin Watson se tiene que para 34 observaciones y 2 variables explicativas excluyendo el término de intersección  $d_L$  y  $d_U = a$  un nivel de significancia del 5%. Puesto que el valor estimado fue de 1.773, esta por encima de 1.33, se acepta la hipótesis de que no existe autocorrelación positiva en los residuos, significando esta ausencia que no existirá correlación entre la precipitación de periodo a periodo.

## CONCLUSIONES

- En el modelo fueron incluidos 7 variables independientes los cuales pasaron por un proceso de selección. Se utilizo el método de Stepwise usando los programas del Statgraphics y SPSS, los cuales seleccionaron 2 variables, Temperatura Superficial del Mar de la Región Niño 3 y Vientos en la Superficie del Océano.
- El modelo Uniecuacional estimado para la Precipitación pluvial del Distrito de Puno es:

$$\hat{Y} = 1374.1367 - 22.8173 * Niño3 - 6.7482 * Vientos$$

$$t \quad (4.9931)(-3.4528) \quad (-2.4854)$$

$$p \quad (0.0000)(0.0016) \quad (0.0185)$$

- El análisis de varianza proporciono un valor  $P = 0.0064$  lo cual indicó de que al menos una de estas variables explicó significativamente, asimismo el coeficiente de determinación explicó el 23.12 % de la variabilidad de la precipitación pluvial del distrito de Puno.
- No existieron evidencias de evasión de los supuestos inherentes al modelo estimado

## RECOMENDACIONES Y SUGERENCIAS

Después de haber cumplido con los objetivos propuestos y con el ánimo de obtener conocimientos más amplios acerca de Los variables que influyen significativamente en la precipitación pluvial del Distrito de Puno, se hacen las siguientes sugerencias:

- Buscar un modelo alternativo con el propósito de que aumente la explicación de los variables que influyen significativamente en la precipitación pluvial, ya sea efectuando un estudio considerando zonas determinadas de las regiones del Perú para efectos de muestreo en el presente estudio o ingresando variables adicionales.
- Realizar un nuevo estudio posterior referente a la precipitación pluvial en el distrito de Puno a fin de observar si posteriormente el modelo presentado en este estudio persiste o en caso contrario hacer un análisis para determinar los variables que influyeron en el cambio.
- En la presente investigación se utilizó Regresión Múltiple, apoyándonos en el método de mínimos cuadrados ordinarios, sin embargo existen muchos otros métodos aplicables a series de tiempo, por lo que se recomienda realizar investigación con distintos instrumentos tales como Procesos estocásticos.



**BIBLIOGRAFÍA**

- ANDERSON, R.L., HAIR, J.F., TATHAM R.E.& BLACK W. C. (1999).***Análisis Multivariante*. Quinta Edición, Prentice Hall Iberia, Madrid, España. 832p.
- ANGELES CABALLERO, CESAR A. (2002).***La Monografía Investigación y elementos*.8va edición, Lima, Perú, San Marcos. 186p.
- CÁCEDA, D. F., PEREZ Q. S.(2001).***Procedimientos Metodológicos y Analíticos para desarrollar Investigación Científica Introducción a la metodología de la investigación*.UNA, Puno,215p.
- CARRASCAL, A. U., GONZALES, G. Y. y RODRIGUEZ P. B. (2001).***Análisis Econométrico con Eviews*. España, RA-MA,337p.
- FERNANDEZ GARCÍA, FELIPE. (1996).***Manual de Climatología Aplicada, Clima Medio Ambiente y Planificación*.Primera reimpresión, España, SINTESIS,285p.
- GUJARATI DAMODAR,N.(2004).***Econometría*.Cuarta edición, México, Editorial McGraw-Hill,.972p.
- HANKE, J. E., REITSCH A. G. (1996).***Pronósticos en los Negocios*. Quinta edición, México, Editorial PRENTICE HALL HISPANOAMERICANA,.605p.
- MONTGOMERY Douglas C., PECK Elizabeth A. & VINING G. Geoffrey, (2005).***Introducción al Análisis de Regresión Lineal*. Primera Edición en Español, México, Compañía Editorial Continental,.567p.

**NETTER, J. & WASSERMAN, W.** *Applied Linear Statistical Models*. Boston.  
Editorial Richard D. Irwin. Inc. Third Edition.

**SEARLE, S.R., (1971).** *Linear Models*, Wiley, Nueva York.

**SERVICIO NACIONAL DE METEOROLOGIA E HIDROLOGIA (SENAMHI-PUNO), Dirección Regional de Puno.(2010).** “Folleto de Divulgación N° 001, Observación de Variables Meteorológicas. 27p.”

**URIEL, J. E. & ALDAS, M. J. (2005).** *Análisis Multivariante Aplicado*, España, PARANINFO, 531p.

**VALDERRAMA, MENDOZA, SANTIAGO. (2006).** *Pasos para Elaborar Proyectos y Tesis de Investigación Científica*. Segunda reimpresión, Lima, Perú, San Marcos, 310p.

## TESIS

**BENAVIDES CAMACHO, VLADIMIR. (2000)** *Modelos de Predicción aplicables a los cambios climáticos del Distrito de Puno; 1964-1998. Tesis (para optar el título profesional de: Ingeniero Estadístico)*. Puno, Perú. Universidad Nacional del Altiplano, Facultad de Ingeniería Estadística e Informática, 78 p.

**PAREDES QUISPE, JUAN R. (1992)** *Modelamiento Estocástico de la Precipitación Pluvial en el Distrito de Puno periodos 1966-1990. Tesis (para optar el título profesional de: Ingeniero Estadístico)*. Puno, Perú. Universidad nacional del Altiplano, Facultad Ingeniería Estadística e Informática, Escuela Profesional Ingeniería Estadística e Informática, 72 p.

**PEREZ QUISPE, SAMUEL D. (1999)** *Selección de la mejor ecuación de regresión para pronostico de Temperatura Extrema Mínima en función de las observaciones de Termómetro Seco y Termómetro Húmedo en el Distrito de Puno; periodo 1980-1990. Tesis (para optar el título profesional de: Ingeniero Estadístico).* Puno, Perú. Universidad Nacional del Altiplano, Facultad Ingeniería Estadística e Informática, Escuela Profesional Ingeniería Estadística e Informática, 76 p.

### FUENTES INFORMÁTICAS

- Modelos de Regresión Múltiple [en línea]: Perú. 2013. [01 de Enero]. Disponible en <http://www.bioestadistica.uma.es/baron/apuntes/ficheros/cap06.pdf>
- Modelos de Regresión Múltiple [en línea]: Perú. 2013. [02 de Enero]. Disponible en <http://www.cienciaytrabajo.cl/pdfs/22/pagina%20185.pdf>
- Modelos de Regresión Múltiple [en línea]: Perú. 2013. [02 de Enero]. Disponible en <http://www.cienciaytrabajo.cl/pdfs/23/pagina%2039.pdf>
- Modelos de Regresión Múltiple [en línea]: Perú. 2013. [03 de Febrero]. Disponible en [http://www.uam.es/personal\\_pdi/ciencias/abaillo/AmbEst/Tema4.pdf](http://www.uam.es/personal_pdi/ciencias/abaillo/AmbEst/Tema4.pdf)
- Regresión múltiple [en línea]: Perú. 2013. [05 de Marzo]. Disponible en [http://mit.ocw.universia.net/17.87/s02/lecture-notes/pdf/multiple\\_regression.pdf](http://mit.ocw.universia.net/17.87/s02/lecture-notes/pdf/multiple_regression.pdf)
- Términos Meteorológicos [en línea]: Perú. 2013. [07 de Marzo]. Disponible en <http://www.snet.gob.sv/Educacion/glosariomet.pdf>

# ANEXOS

**ANEXO N° 1**

**DATOS DE PRECIPITACIÓN REGISTRADA POR EL SENAMHI PUNO Y LAS  
VARIABLES CLIMÁTICAS REGISTRADA POR LA NOAA A TRAVÉS DEL  
TIEMPO**

AÑO	A. Niño 1+2	A. Niño 3	A. Niño 3.4	A. Niño 4	ROL	VIENTOS	IOS	PRECIPITACION
1979	2,87	2,62	2,77	2,5	487	93,9	-2	527,3
1980	-0,21	1,27	2,75	3,97	480,9	96,6	-7,4	614,4
1981	-4,79	-3,89	-2,95	-1,37	518,3	114,9	0,1	788,6
1982	8,39	14,21	13,43	7,96	389,5	59,3	-25,6	794,5
1983	29,48	13,92	7,33	1,45	468,5	69,6	-22,1	434,1
1984	-5,55	-7,35	-6,91	-5,77	638,2	115	-2,4	1051,9
1985	-12,99	-10,08	-7,26	-3,73	580	109,6	0,9	961,5
1986	-2,51	0,73	2,57	4,23	473,6	93,2	-8	923,9
1987	14,06	15,23	17,11	11,73	240,5	65	-26,2	630,7
1988	-11,66	-14,01	-12,05	-6,06	628,8	127,6	13,8	848,6
1989	-4,68	-7,23	-8,57	-7,06	629,9	112,5	11	684,7
1990	-4,68	0,19	3	6,92	414,8	92,1	-7,8	643,8
1991	1,73	6,16	8,87	8,06	352	80,4	-18,3	596,8
1992	6,78	5,77	8,95	7,77	302,5	60,9	-24,1	374,1
1993	5,12	4,62	5,64	7,77	310	81,1	-19,9	759,2
1994	-3,5	1,19	6,05	3,3	419,7	89,1	-24,2	803,6
1995	-4,96	-4,21	-1,25	-1,59	560,3	102,8	-5,7	526,5
1996	-13,07	-6,58	-4,58	9,14	659,2	115,2	10	753,5
1997	30,26	20,2	16,14	-2,7	316,8	49,3	-21,3	908,9
1998	21,41	6,14	-0,59	-11,14	594,8	82,7	-6,2	615,1
1999	-10,35	-11,48	-13,02	-7,99	706,7	120,8	14,6	1003,9
2000	-6,58	-7,35	-9,49	1,28	691,7	116,1	14,1	740,6
2001	-6,78	-3,36	-1,83	11,13	523,3	107	0,1	1006,8
2002	0,55	5,92	9,99	8,52	349,7	91,6	-12,3	908,8
2003	-6,88	1,32	4,8	9,25	481,1	98,4	-7,5	714,1
2004	-5,83	1,87	5,77	6,45	450,6	100,4	-10,8	678,4
2005	-6,9	-0,13	2,6	5,05	485,5	103	-9,5	674,5
2006	3,54	3,53	3,21	0,27	508,7	101,8	-4,6	775,0
2007	-15,46	-8,87	-5,13	-8,53	571,8	121,4	0,8	814,2
2008	3,21	-2,69	-7,66	5,53	685	118,5	21,4	661,3
2009	4,33	5,82	5,65	0,46	522,9	96,6	-1,4	798,1
2010	-5,88	6,12	4,73	6,54	533,1	94,7	-4,4	48,68
2011	-6,75	18,2	2,8	2,18	475,2	130,3	-3,1	64,17
2012	15,46	-3,69	-3,25	0,45	500	123,5	12	79,34

**ANEXO N° 2**

**Regresión**

**VARIABLES INTRODUCIDAS/ELIMINADAS**

Modelo	VARIABLES INTRODUCIDAS	VARIABLES ELIMINADAS	Método
1	NIÑO3		Por pasos (criterio: Prob. de F para entrar <= ,050, Prob. de F para salir >= ,100).
2	VIENTOS		Por pasos (criterio: Prob. de F para entrar <= ,050, Prob. de F para salir >= ,100).

a. Variable dependiente: PRECIPITACION

**Resumen del modelo**

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,366 <sup>a</sup>	,134	,107	237,18601
2	,527 <sup>b</sup>	,278	,231	220,05204

a. Variables predictoras: (Constante), NIÑO3

b. Variables predictoras: (Constante), NIÑO3, VIENTOS

**ANOVA<sup>a</sup>**

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	278234,215	1	278234,215	4,946	,033 <sup>b</sup>
	Residual	1800230,582	32	56257,206		
	Total	2078464,797	33			
2	Regresión	577354,953	2	288677,477	5,962	,006 <sup>c</sup>
	Residual	1501109,844	31	48422,898		
	Total	2078464,797	33			

**Coefficientes<sup>a</sup>**

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	696,747	41,169		16,924	,000
	NIÑO3	-10,878	4,891	-,366	-2,224	,033
2	(Constante)	1374,137	275,210		4,993	,000
	NIÑO3	-22,817	6,608	-,767	-3,453	,002
	VIENTOS	-6,748	2,715	-,552	-2,485	,019

a. Variable dependiente: PRECIPITACION

**Variables excluidas<sup>a</sup>**

Modelo		Beta dentro	t	Sig.	Correlación parcial	Estadísticos de colinealidad
						Tolerancia
1	NIÑO1.2	,039 <sup>b</sup>	,168	,868	,030	,516
	NIÑO3.4	,564 <sup>b</sup>	1,672	,105	,288	,225
	NIÑO4	-,041 <sup>b</sup>	-,225	,823	-,040	,846
	ROL	-,212 <sup>b</sup>	-,883	,384	-,157	,473
	VIENTOS	-,552 <sup>b</sup>	-2,485	,019	-,408	,472
	IOS	-,451 <sup>b</sup>	-1,962	,059	-,332	,469
2	NIÑO1.2	-,195 <sup>c</sup>	-,841	,407	-,152	,438
	NIÑO3.4	,096 <sup>c</sup>	,227	,822	,041	,135
	NIÑO4	-,092 <sup>c</sup>	-,542	,592	-,099	,833
	ROL	,084 <sup>c</sup>	,323	,749	,059	,355
	IOS	-,130 <sup>c</sup>	-,415	,681	-,076	,244

a. Variable dependiente: PRECIPITACION

b. Variables predictoras en el modelo: (Constante), NIÑO3

c. Variables predictoras en el modelo: (Constante), NIÑO3, VIENTOS

ANEXO N° 3

ESTADÍSTICO d DE DURBIN-WATSON

DURBIN-WATSON: Los puntos de significancia  $d_L$  y  $d_U$  al nivel de significancia 5%

n	k*=1		k=2		k=3		k=4		k=5		k=6		k=7		k=8		k=9		k=10	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
6	0.610	1.400	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
7	0.700	1.356	0.467	1.896	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
8	0.763	1.332	0.559	1.777	0.367	2.287	----	----	----	----	----	----	----	----	----	----	----	----	----	----
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	----	----	----	----	----	----	----	----	----	----	----	----
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	----	----	----	----	----	----	----	----	----	----
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	----	----	----	----	----	----	----	----
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	----	----	----	----	----	----
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	----	----	----	----
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	----	----
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.258	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.681	2.396
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.796	2.281
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.079	1.891	1.015	1.978	0.950	2.069	0.885	2.162	0.821	2.257
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.876	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216



37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.197
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098	0.912	2.180
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.952	2.149
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986	1.110	2.044
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.838	1.369	1.874	1.337	1.910	1.305	1.948
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898
150	1.720	1.747	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.846	1.608	1.862	1.593	1.877
200	1.758	1.779	1.748	1.789	1.738	1.799	1.728	1.809	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874

K\* : Representa el número de regresoras que excluyen al intercepto.