

# Minimización de la dimensión y modelos lineales

## 3.1. Introducción

En este capítulo se hace un análisis teórico de las técnicas de minimización de la dimensión que se aplicaron para predecir  $X_a$ , así como, los modelos de regresión lineal para estimar  $Y_T$ .

En  $X_a$ , se utilizan los componentes principales, el análisis factorial y el escalamiento multidimensional. Se conjetura que  $X_a$  depende de once indicadores, por tanto, se construye un índice que predice su comportamiento, se estima la dinámica  $Y_T$  mediante la aplicación de modelos lineales, posteriormente, se predice los efectos de  $X_a$  en el tiempo sobre  $Y_T$ .

La utilización de estas técnicas matemáticas y estadísticas dan elementos necesarios para tener una fotografía de  $X_a$  y su efecto en  $Y_T$ .

## 3.2. Minimización de la dimensión

Hablar de la minimización de la dimensión, significa el identificar las técnicas y el modelo estadístico multivariable que permiten reducir el espacio conformado por un conjunto variables aleatorias, teniendo como objetivo, conservar la mayor variabilidad de los datos en un espacio de menor magnitud.

Estas herramientas permiten analizar y estudiar la relación lineal entre variables, así como, reducir el costo que implica el construir algoritmos que predicen el comportamiento de un fenómeno complejo. Entre ese modelo y técnica se encuentran las siguientes:

Técnicas	Modelo
<ul style="list-style-type: none"><li>• Componentes principales (CP).</li><li>• Escalamiento multidimensional (ED)</li></ul>	<ul style="list-style-type: none"><li>• Análisis factorial (AF)</li></ul>

**Cuadro 16.** Modelo y técnicas de minimización de la dimensión

**Fuente.** Elaboración personal

Para la construcción de  $X_a$  se utiliza la técnica que presente mejor ajuste, a partir de variables cuantitativas que inciden en  $e$  la dinámica de dicho fenómeno.

### 3.2.1. Técnica de Componentes principales (TCP)

En 1901 K. Person introdujo TCP, a partir del ajuste ortogonal del método de mínimos cuadrados, esto se trace en (Peña, 2002):

- La reducción de un espacio complejo a un espacio mínimo con  $p$  – dimensión.
- Conversión de variables aleatorias originales que tienen correlaciones altas entre ellas, a nuevas variables con cero correlacion.

El objetivo de TCP es minimizar un espacio complejo que se forma “N” variables aleatorias (correlacionadas, carecen de simetría y con alto grado de varianza), a un espacio pequeño de “W” variables aleatorias (no correlacionadas, con simetría y varianza controlada), tal que  $N \geq W$ . Es decir (Peña, 2002):

$$(X_1, X_2, X_3, \dots, X_p) \rightarrow (W_1, W_2, \dots, W_p)$$

Donde:

- $X_i$  = variables originales
- $W_i$  = componentes principales

TCP son la mezcla de combinaciones lineales de las variables aleatorias originales (Peña, 2002):

$$W_{ij} = \sum_{ij=1}^p (B_{ij})(X_{ij}) \quad (17)$$

En forma matricial (Peña, 2002):

$$\begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \vdots & \ddots & \vdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} \begin{pmatrix} B_{11} \\ \vdots \\ B_{1p} \end{pmatrix} = \begin{pmatrix} W_1 \\ \vdots \\ W_p \end{pmatrix} \quad (18)$$

Si se hace una combinación lineal entre las variables originales, se obtiene la varianza original, esto da origen al  $W_1$  (Peña, 2002):

$$W_1 = XB_1 ; \text{ tal que } X_{ij} \text{ y } B_{ij} \sim N(0, 1) \quad (19)$$

Tal que (Peña, 2002):

$$\text{Var}(W_1) = B_1' S B_1 = S B_1^2 = 1; S = \text{matriz de varianza y covarianza} \quad (20)$$

Aplicando el método de Lagrange (Peña, 2002):

$$M = SB_1^2 - \lambda(B_1^2 - 1) = 0 \quad (21)$$

Derivando,

$$\frac{\partial M}{\partial B_1} = 2SB_1 - 2\lambda B_1 = 0 \quad (22)$$

Despejando,

$$SB_1 = \lambda B_1 \rightarrow B_1^{-1} \lambda B_1 = \lambda \quad (23)$$

Donde  $\lambda$  es mayor al valor propio de S y es la varianza de  $W_1$ . Si se tiene una función objetivo a través del método de Langrange (Peña, 2002):

$$\theta = B_1' SB_1 + B_2' SB_2 - \lambda_1(B_1' B_1 - 1) - \lambda_2(B_2' B_2 - 1) \quad (24)$$

Por tanto,  $W_1 = XB_1$  y  $W_2 = XB_2$  son las máximas varianzas, donde  $B_1$  y  $B_2$  son vectores que describen el plano que proyecta a las variables X's. "Derivando e igualando a cero" (Peña, 2002):

$$\frac{\partial \theta}{\partial B_1} = 2SB_1 - 2\lambda_1 B_1 = 0 \rightarrow SB_1 = \lambda_1 B_1 \quad (25)$$

$$\frac{\partial \theta}{\partial B_2} = 2SB_2 - 2\lambda_2 B_2 = 0 \rightarrow SB_2 = \lambda_2 B_2 \quad (26)$$

"Si  $B_1$  y  $B_2$  son vectores propios de S, y sustituyendo en la expresión algebraica (24),  $\theta$  sería la función objetivo de la suma de  $\lambda_1$  y  $\lambda_2$ " (Peña, 2002). Estos componentes principales deben cumplir con las siguientes propiedades (Díaz, 2007):

- $E(W_i) = 0$ ; tal que  $X_{ij} \sim N(0, 1)$
- $\text{Var}(W_1) > \text{Var}(W_2) > \dots > \text{Var}(W_p) = 0$
- $\text{Var}(W_i) = \lambda_i$ ;  $i$  -ésimo valor propio
- $\text{Var}(W_i) = \sum_{i=1}^p \lambda_i = \text{Traza}$
- $\text{Cov}(W_i, W_j) = 0$ , tal que  $i \neq j$
- $\prod_{i=1}^p \text{var}(W_i) = \det(W_j)$

Existen otros elementos a considerar para TCP, tales como (Díaz, 2007):

- La proporción de variabilidad de cada componente principal es:  $\text{Var}(W_i) = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i}$ .
- Las covarianzas entre cada componente principal y las variables originales es:  
 $\text{Cov}(W_i: X_1, X_2, \dots, X_p) = \lambda_i B_i$ .

- Las correlaciones entre cada componente principal y una variable  $X_{ij}$  es:  

$$\text{Cor} (W_i; X_{ij}) = B_i \frac{\sqrt{\lambda_i}}{S_j}$$
 Tal que:  $\sqrt{\lambda_i}$  es la desviación estándar del componente y  $S_j$  es la desviación estándar de la variable original.

La TCP se aplica a través de dos vías (Díaz, 2007):

- Los datos son simétricos y/o homogéneos y no correlacionados, por tanto, su punto de partida es la matriz de covarianzas.
- Los datos no son simétricos ni homogéneos y están correlacionados, por tanto, su punto de partida es la matriz correlación.

“La selección de componentes principales se hace a través de tres criterios” (Díaz, 2007):

- A partir de que  $\lambda_i \geq 1$ , tal que  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ .
- Cuando la variabilidad total es superior al 80%.
- Cuando el gráfico *screen plot* describe un codo.

“Los componentes principales se analizan como índices o indicadores que describen el comportamiento de un fenómeno. La interpretación de esto es la siguiente” (Díaz, 2007):

- $CP_1$  (primer componente): tiene la máxima varianza y sus coeficientes son positivos.
- $CP_2$  (segundo componente): tiene coeficientes positivos y negativos, y su varianza es la segunda máxima.

“Las etapas que debe llevar un modelo de componentes principales es la siguiente” (Díaz, 2007):

- **Etapas 1:** se evalúa la mejor opción de modelaje, a partir de la matriz correlación y la matriz de varianza – covarianza.
- **Etapas 2:** con el polinomio característico (valores propios), se estiman los componentes principales.
- **Etapas 3:** mediante la comprobación de sus propiedades, se procede a la validación del modelo.
- **Etapas 4:** se hace la lectura de los componentes principales, y se hacen predicciones.

### 3.2.2. Modelo de Análisis factorial (MF)

El MF permite minimizar el espacio de un conjunto de variables que tienen alta correlación entre ellas, es decir, tiene por objetivo reducir las múltiples relaciones que hay entre “N”

variables originales a “F” factores y/o dimensiones comunes (Peña, 2012).

La reducción de información o de datos está dada mediante la matriz correlación de las variables originales, con la finalidad de identificar si existe algún patrón de relaciones, para llevarlas a nuevas variables denominadas factores (F), tal que (Peña, 2012):

$$(X_1, X_2, \dots, X_p) = (F_1, F_2, \dots, F_k); \text{ tal que } P \geq k$$

La diferencia que existe entre componentes principales y análisis factorial son las siguientes (Cuadras, 2007):

- Los componentes principales están basados en aspectos matemáticos (álgebra lineal y método de la Langrange), cuya estimación da como resultado nuevas variables que resumen relaciones lineales.
- El MF es un proceso matemático y estadístico que busca nuevas variables ficticias que requieren de la inferencia estadística para su validación y aplicación.

La estructura del modelo factorial es (Anderson, 2003):

$$X = BF + E \quad (27)$$

Donde:

- X = variables originales.
- B = coeficientes o pesos de los factores.
- F = factores comunes.
- E = factores específicos.

Tal que (Anderson, 2003):

$$\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} B_{11} & \dots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{p1} & \dots & B_{pk} \end{pmatrix} \begin{pmatrix} F_1 \\ \vdots \\ F_k \end{pmatrix} + \begin{pmatrix} E_1 \\ \vdots \\ E_k \end{pmatrix}; \text{ tal que } k < p \quad (28)$$

Dónde:  $\text{Var}(X_{ij}) = 1$

$$\text{Var}(X_{ij}) = \begin{pmatrix} B_{11} & \dots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{p1} & \dots & B_{pk} \end{pmatrix} \begin{pmatrix} B_{11} & \dots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{p1} & \dots & B_{pk} \end{pmatrix} + \begin{pmatrix} E_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & E_p^2 \end{pmatrix} \quad (29)$$

Tal que (Anderson, 2003):

$$\text{Var}(X_1) = B_{11}^2 + B_{12}^2 + \dots + B_{1k}^2 + \text{var}(E_1^2) = 1 \quad (30)$$

Donde:

$B_{11}^2 + B_{12}^2 + \dots + B_{1k}^2$	<b>Comunalidad</b>	Explica a los factores comunes y representa una parte de la varianza de Xi.
$\text{Var}(E_1^2)$	<b>Unicidad</b>	Explica a los factores específicos y representa una parte de la varianza de Xi.

**Cuadro 17.** Comunalidad y Unicidad

**Fuente.** Elaboración personal

El ajuste del MF es a través de la comunalidad ( $C^2$ ), tal que  $0 \leq C^2 \leq 1$ :

$$C^2 = 1 - \lim_{E_1 \rightarrow 0} [\text{var}(E_1^2)] \quad (31)$$

Ese grado de ajuste es:

<b>Excelente</b>	$0.76 \leq C^2 \leq 1.00$	<b>Deficiente</b>	$0.25 \leq C^2 \leq 0.49$
<b>Bueno</b>	$0.50 \leq C^2 \leq 0.7$	<b>Malo</b>	$0.00 \leq C^2 \leq 0.24$

**Cuadro 18.** Grado de ajuste MF

**Fuente.** Elaboración personal

Las propiedades de los factores son (Cuadras, 2007):

$E[FF^{-1}] = I$	La matriz identidad es la matriz de varianza y covarianza de los factores comunes.
$E[F] = 0$	La esperanza matemática (valor esperado) de los factores comunes se aproxima a cero
$E[EE^{-1}] = \emptyset$	La correlación entre las factores comunes se aproxima a cero
$E[FE^{-1}] = 0$	Los factores (únicos y comunes) tiene covarianza que se aproxima a cero.

**Cuadro 19.** Propiedades de MF

**Fuente.** Elaboración personal

Un elemento de los factores es la rotación, pues es fundamental para la interpretación de estos. La rotación se da cuando hay un cambio del eje original, alcanzando otra posición, mediante la distribución óptima de la varianza, con la finalidad de tener un patrón eficiente y significativo del fenómeno en estudio. (Cuadras, 2007). Los tipos de rotación en MAF son (Cuadras, 2007):

<b>Varimax</b>	Reduce a su mínima expresión los vectores columna de la matriz de factores, hasta existir unos y ceros en las columnas.
<b>Quartimax</b>	Crea pocos factores a partir de la existencia de altas correlaciones entre las variables originales, mediante la minimización de la varianza de los factores.

<b>Equamax</b>	Reduce la cantidad de variables que tienen alta incidencia en un factor, así como, la cantidad de factores que se requieren para el análisis de una variable.
----------------	---

**Cuadro 20.** Tipos de rotación de MF

**Fuente.** Elaboración personal

Los supuestos que debe cumplir un MF son (Anderson, 2003):

- El total de factores “n” necesarios para predecir el modelo.
- Los scores (factores comunes) se acercan a un normal:  $F_i = N(\mu, \sigma^2)$ .
- La esperanza matemática (valor esperado) de cada factor debe aproximarse a cero:  $E(F_i) = 0$ .
- Debe aproximarse a la unidad la varianza de cada factor:  $Var(F_i) = 1$ .
- Correlación nula entre factores:  $Cor(F_i, F_j) = 0$ , tal que  $i \neq j$ .
- Debe tender a cero la covarianza entre factores:  $Cov(F_i, F_j) = 0$ , tal que  $i \neq j$ .

La construcción de un modelo factorial debe seguir las siguientes etapas (Anderson, 2003):

- **Etap 1:** formulación del problema. En esta etapa se plantean los objetivos, las preguntas e hipótesis del fenómeno en estudio, y con ello, se debe desarrollar un marco teórico.
- **Etap 2:** desarrollo del modelo. En esta etapa se plantean tres fases:
  - 1° **fase:** se construye y se describe el efecto lineal que hay entre las variables originales, a partir de la matriz correlación.
  - 2° **fase:** se hace la estimación de los parámetros de cada uno de los factores.
  - 3° **fase:** se selección de los factores y se determina la cantidad de estos para predecir el comportamiento del fenómeno en estudio.
- **Etap 3:** validación del modelo. Consiste en el cumplimiento de los supuestos, así como, la elección de las variables más latentes.
- **Etap 4:** predicción del modelo. Se analiza si el modelo para ver si requiere de la rotación de los factores, con el objetivo de mejorar la explicación del fenómeno en estudio.

### 3.2.3. Técnica de Escalamiento multidimensional (TD)

La TD es un complemento de la TCP, que representa a una matriz de distancias  $D$  de  $n \times n$  (cuadrada) de un conjunto de variables ortogonales  $(X_1, X_2, \dots, X_p)$ , con la finalidad de obtener una matriz  $X$  de  $n \times p$ , la cual, se entiende como la matriz de  $p$  variables en los  $n$  objetos,

donde la distancia entre los objetos de estudio da origen a una distancia directa (Peña, 2012).

La TD y TCP comparten el mismo objetivo, el analizar, describir e interpretar un conjunto de datos. Si existen muchos objetos de estudio, la matriz de similitudes será grande y su personificación será con pocas variables.

Como se puede ver, TD representa un complemento de TCP, ya que, TCP considera la matriz  $p \times p$  de correlación o de varianzas – covarianzas para minimizar el espacio de “N” variables originales, a “CP” de nuevas variables ficticias, en tanto, TD considera la matriz de distancias  $n \times n$  de correlación o de varianzas – covarianzas entre objetos, para minimizar el espacio “N” variables originales, a “D” (dimensiones) de nuevas variables ficticias (Peña, 2012).

La elaboración de un modelo a través de TD debe considerar a la matriz de proximidades (Guerrero & Ramírez, 2012):

$$\Delta \in B_{n \times n} \text{ tal que } \Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{nn} \end{bmatrix} \quad (32)$$

Donde:

- $\delta_{ij}$  = cada objeto en estudio.
- $\Delta$  = Proximidad entre el estímulo  $i$  y  $j$ , tal que  $i \neq j$ .

Tal que,  $X_{ij}$  es la coordenada de la dimensión “j” en el estímulo “i”. Si existe una matriz X, la distancia entre “j” e “i” se expresa como (Guerrero & Ramírez, 2012):

$$d_{ij} = \left[ \sum_{t=1}^m (x_{it} - x_{jt})^p \right]^{1/p}; \quad 1 \leq p \leq \alpha \quad (33)$$

A partir de las distancias:

$$D \in B_{n \times n}; \text{ tal que } D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \quad (33)$$

Por tanto, la respuesta que da TD es la óptima que hay en  $\Delta$  y la matriz D.

Hay dos herramientas para abordar la TD (Guerrero & Ramírez, 2012); primero, la Técnica de Escalamiento Métrico (TDM); y en segundo, la Técnica de Escalamiento no métrico (TDN).

En la TDM, se parte del hecho de  $X_i \in R$ , y las distancias de las  $X_i$  dependen de las proximidades [ $d_{ij} = f(\delta_{ij})$ ] y tienen relaciones lineales [ $d_{ij} = a + b\delta_{ij}$ ], por tanto, su proceso es el siguiente (Guerrero & Ramírez, 2012):



- De  $\Delta \in M_{n \times n}$  se obtiene  $B \in M_{n \times n}$  que es el producto escalar de vectores.
- Transformación de  $\Delta \in M_{n \times n}$  a  $B \in M_{n \times n}$ , bajo las siguientes propiedades de la distancia (Guerrero & Ramírez, 2012):

$$\begin{aligned} d_{ij} &\geq 0 = d_{ii} && \text{No negatividad} \\ d_{ij} &= d_{ji} && \text{Simetría} \\ d_{ij} &\leq d_{ik} + d_{ik} && \text{Triangular desigual} \end{aligned}$$

Con base en  $d_{ij}$ , predice la constante:

$$c_{\min} = \max_{(i,j,k)} \{ \delta_{ij} - \delta_{ik} - \delta_{kj} \} \quad (34)$$

tal que, las distancias es el resultado de la suma  $\delta_{ij}$  y  $c$ :

$$d_{ij} = \delta_{ij} + c \quad (35)$$

Si  $D \in M_{n \times n}$  y se convierte en  $B \in M_{n \times n}$  (Guerrero & Ramírez, 2012):

$$b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2) \quad (36)$$

Donde:

$$d_{i.}^2 = \frac{1}{n} \sum_{j=1}^n d_{ij}^2 \rightarrow \text{Distancia cuadrática media por fila} \quad (37)$$

$$d_{.j}^2 = \frac{1}{n} \sum_{i=1}^n d_{ij}^2 \rightarrow \text{Distancia cuadrática media por columna} \quad (38)$$

$$d_{..}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2 \rightarrow \text{Distancia cuadrática media de la matriz} \quad (39)$$

Si  $B \in M_{n \times n}$  se transforma en  $X \in M_{n \times m}$ , entonces:

- $B = X * X'$
- $X$  = coordenadas de cada objeto en los “n” estímulos de las “m” dimensiones.

El TD, las proximidades y las distancias no tienen relaciones lineales, en ellas, existe relación monótona, por tanto, el proceso de un modelo TD es el siguiente (Guerrero & Ramírez, 2012):

- Transformación de la matriz de proximidades en una matriz de rangos: 1 a  $\frac{n(n-1)}{2}$ .
- Obtención de la matriz  $X \in M_{n \times m}$  de coordenadas aleatorias, que da las distancias entre los estímulos.
- Comparación de las proximidades con las distancias, dando como resultado las disparidades  $d'_{ij}$ .

- Definición y minimización del Stress.

La bondad de ajuste de un modelo a través de TD se hace mediante el indicador de Stress, es decir (Guerrero & Ramírez, 2012):

- $d_{ij} = f(\delta_{ij})$ .
- A las transformaciones de las proximidades son distorsiones.

Por tanto, el error cuadrático del modelo es (Guerrero & Ramírez, 2012):

$$e_{ij}^2 = [f(\delta_{ij}) - d_{ij}]^2 \quad (40)$$

Partiendo de  $e^2$ , el Stress se define como (Guerrero & Ramírez, 2012):

$$\text{Stress} = \sqrt{\frac{\sum_{ij} [f(\delta_{ij}) - d_{ij}]^2}{\sum_{ij} d_{ij}^2}} \quad (41)$$

Tal que:

“A mayor diferencia entre las distancias y disparidades, mayor Stress”, por tanto, menor eficiencia del modelo, es decir (Guerrero & Ramírez, 2012):

0.20 → Pobre ajuste	0.025 → Muy buen ajuste
0.10 → Aceptable ajuste	0.00 → Excelente ajuste
0.05 → Buen ajuste	

Donde, su coeficiente de correlación al cuadrado (RSQ) es el grado de varianza que conserva el modelo TD sobre los datos:

$$\text{RSQ} = \frac{\left[ \sum_i \sum_j (d_{ij} - d_{..})(f(d_{ij}) - f(d_{..})) \right]^2}{\left[ \sum_i \sum_j (d_{ij} - d_{..})^2 \right] \left[ \sum_i \sum_j (f(d_{ij}) - f(d_{..}))^2 \right]} \quad (42)$$

Tal que:

- Si  $\text{RSQ} = 1$  indica que el modelo es bueno.
- Si  $\text{RSQ} = 0$  indica que el modelo es malo.

Las etapas que debe cumplir la elaboración de un modelo de TD son las siguientes (Repiso, 2013):

- **Etapa 1:** planteamiento del problema. Consiste en la identificación de las variables que intervienen en el comportamiento del fenómeno en estudio.

- **Etapa 2:** obtención de datos. Una vez identificadas el fenómeno en estudio y sus variables de incidencia, en esta fase, se procede a la busque de los datos a partir de distintas fuentes de información.
- **Etapa 3:** elección del procedimiento del modelo de escalamiento multidimensional. Con base a la estructura de los datos, se hace la selección del tipo escalamiento (métrico o no métrico).
- **Etapa 4:** decisión sobre el número de dimensiones. Mediante la matriz de distancias y de similitudes, se predicen sus parámetros, con la finalidad de determinar el tamaño de dimensiones.
- **Etapa 5:** validación. Mediante los indicadores de Stress y RSQ se procede la estimación del grado de ajuste el modelo.
- **Etapa 6:** interpretación de resultados. Una vez identificado el tamaño dimensional y la estimación del ajuste del modelo, se procede al estudio y explicación de los resultado obtenidos.

### 3.3. Modelos lineales

Si YT es una variable aleatoria continua (v.a.c.), entonces existe una  $f(YT)$  de nominada función de densidad de probabilidad que cumple con lo siguiente (Wackely, Mendenhall y Scheaffer, 2010):

$$f(YT) \geq 0 \quad (43)$$

$$\int_{-\infty}^{\infty} f(YT) dYT = 1 \quad (44)$$

$$P(a \leq YT \leq b) = \int_a^b f(YT) dYT \quad (45)$$

Donde su esperanza matemática y su varianza son las siguientes (Wackely, Mendenhall y Scheaffer, 2010):

**Esperanza matemática**  $E(YT) = \int_{-\infty}^{\infty} Tpf(YT)dYT \quad (46)$

**Varianza**  $Var(YT) = \int_{-\infty}^{\infty} (YT - \overline{YT})^2 f(YT)DYT \quad (47)$

En forma de regresión lineal (Rencher y Bruce, 2008):

$$YT = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + u_i \quad (48)$$

Tal que:

- $YT$  = Temperatura Media del Planeta.
- $X_j$  = variables que inciden en el comportamiento de  $X_a$ .
- $B_j$  = parámetros a estimar (a través de máxima verosimilitud).
- $u_i$  = margen de error.

Si la  $YT$  es una v.a.c., sus opciones de ser modelada son las regresiones Gauss y Gamma. Como funciones de distribución, sus características son las siguientes (Canavos, 1988):

	<b>Gauss</b>	<b>Gamma</b>
<b>Función de distribución</b>	$f(YT) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(YT-\mu)^2}{2\sigma^2}}; -\infty \leq YT \leq \infty$	$f(Tp) = \frac{\lambda(\lambda YT)^{\alpha-1} e^{-\lambda YT}}{\Gamma(\alpha)}; YT \in (0, \infty)$
<b>Función de densidad</b>	$F(YT) = \int_{-\infty}^{YT} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(YT-\mu)^2}{2\sigma^2}}$	$f(YT) = \int_0^{Tp} \frac{\lambda(\lambda YT)^{\alpha-1} e^{-\lambda YT}}{\Gamma(\alpha)}$
<b>Esperanza matemática</b>	$E(YT) = \mu_{YT}$	$E(YT) = \frac{\alpha}{\lambda}$
<b>Varianza</b>	$Var(YT) = \sigma_{YT}^2$	$Var(YT) = \frac{\alpha}{\lambda^2}$

**Cuadro 21.** Estimadores de las distribuciones Gauss y Gamma

**Fuente.** Probabilidad y Estadística. Aplicaciones y Métodos (Canavos, 1988)

Partiendo del cuadro 21, en las siguientes secciones se plantea los pasos, la estructura, las etapas, los supuestos y los procesos matemáticos, estadísticos y de probabilidad que engloban en cada sistema de modelamiento.

### 3.3.1. Modelo de Regresión Lineal Múltiple (RLM)

El modelo de Gauss hace referencia a la regresión lineal, la cual tendría la siguiente estructura, a partir de  $YT$  (Rencher y Bruce, 2008):

<b>Regresión lineal simple</b>	$YT = B_0 + B_1X + u_i$
<b>Regresión lineal múltiple</b>	$YT = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + u_i$

**Cuadro 22.** Regresión lineal simple y múltiple

**Fuente.** Probabilidad y Estadística. Aplicaciones y Métodos (Rencher y Bruce, 2008).

Desde el contexto de YT, el modelo sería una regresión lineal múltiple, donde la predicción de sus estimadores es mediante el método de mínimos cuadrados.

$$YT = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_k X_k + u_i; \quad 0 \leq R^2 \leq 1 \quad (49)$$

Tal que  $R^2$  se conoce como coeficiente de determinación (cuanto conserva de varianza de los datos).

- Si  $R^2 = 1$ , excelente ajuste
- Si  $R^2 = 0.5$ , moderado ajuste
- Si  $R^2 = 0$ , bajo ajuste

La aplicación de este modelo de regresión múltiple parte del cumplimiento de una serie puestos de inferencia, tales como ((Rencher y Bruce, 2008):

**Supuesto 1:** relación lineal entre las variables dependientes y las variables independientes, mediante la matriz correlación.

$$-1 \leq \text{Cor}(YT, X_i) \leq 1 \quad (50)$$

Tal que:

- Si la  $\text{Cor}(YT, X_i) = 1$ , el efecto es positivo: “un incremento de  $X_i$ , conlleva a un incremento de YT”.
- Si la  $\text{Cor}(YT, X_i) = -1$ , el efecto es negativo: “un incremento de  $X_i$ , conlleva a un decremento de YT”.
- Si la  $\text{Cor}(YT, X_i) = 0$ , no existe mucha evidencia de la relación lineal.

**Supuesto 2:** Las variables independientes deben ser significativas para predecir la variable dependiente. Prueba de hipótesis:

$$H_0: X_i \text{ es significativa en YT} \quad \text{vs.} \quad H_a: X_i \text{ no es significativa en YT}$$

Si el P – valor es menor 0.05, se va a aceptar  $H_0$ , por lo tanto,  $X_i$  es significativa para predecir YT.

**Supuesto 3:** Los residuales deben aproximarse a una distribución normal. Prueba de hipótesis:

$$H_0: u_i \sim N(\mu, \sigma^2) \quad \text{vs.} \quad H_a: u_i \neq N(\mu, \sigma^2) \quad (51)$$

Mediante las pruebas de Kolmogorov, si el P – valor es mayor a 0.05, se acepta  $H_0$ , por lo tanto,  $u_i$  se aproximan a distribución normal.

**Supuesto 4:** Los residuales deben tener varianza controla, es decir, deben ser homocedásticos. Prueba de hipótesis:

$$H_0: \text{Var}(u_i) \sim \sigma^2 \quad \text{vs.} \quad H_a: \text{Var}(u_i) \neq \sigma^2 \quad (52)$$

A través de la prueba Breusch – Pagan, si el P – valor es mayor 0.05, se acepta Ho, por lo tanto, los  $u_i$  son homocedásticos.

**Supuesto 5.** La correlación entre los residuales debe aproximarse a cero, es decir, no debe existir autocorrelación entre los residuales: Prueba de hipótesis:

$$\begin{aligned} \text{Ho: } \text{Cor}(u_i, u_j) \sim 0 \text{ vs. } \text{Ha: } \text{Cor}(u_i, u_j) \sim 1; \\ \text{tal que } i \neq j \end{aligned} \quad (53)$$

Mediante la prueba Durbin – Watson, si el P – valor es mayor a 0.05, se acepta Ho, por lo tanto, los  $u_i$  no tiene autocorrelación.

**Supuesto 6:** La correlación entre las variables independiente debe aproximarse a cero, es decir, no debe existir colinealidad entre las variables independiente. Prueba de hipótesis:

$$\begin{aligned} \text{Ho: } \text{Cor}(X_i, X_j) \sim 0 \text{ vs. } \text{Ha: } \text{Cor}(X_i, X_j) \sim 1; \\ \text{tal que } i \neq j \end{aligned} \quad (54)$$

Si el VIF (Factor de Influencia de la Varianza) es mayor a 5, se acepta Ha y se rechaza Ho, es decir, hay colinealidad entre las variables independientes.

Si el modelo cumple con estos supuestos se aplica, si ocurre el caso contrario, se tiene que transformar YT, para tanto, se debe aplicar las transformaciones Box – Cox. Si el modelo no cumple con los supuestos de normalidad y de homocedasticidad, una posible solución son las transformaciones Box – Cox, las cuales son aplicadas a la variable dependiente, en forma de potencia (Atkinson, 2020):

$$YT^\lambda = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + u_i \quad (55)$$

Tal que YT tiene que ser mayor a cero, es decir tiene que ser positiva, en caso contrario, se le tiene que sumar una constante. Sus transformaciones más habituales se describen en el cuadro 23, donde  $\lambda \in \mathbb{R}$ :

$\lambda$	$-\lambda$	$0$	$\lambda$
<b>Transformación</b>	$\frac{1}{YT^\lambda}$	$\text{Ln}(YT)$	$(YT)^\lambda$

**Cuadro 23.** Transformaciones Box - Cox

**Fuente.** Transformaciones de BOX – COX (Atinson, 2020).

Con base a las transformaciones Box – Cox, el modelo deja de ser lineal, por tanto, su forma general sería de la siguiente forma:

$$E(YT) = \sqrt{\lambda \widehat{B}_0 + \widehat{B}_1 X_1 + \widehat{B}_2 X_2 + \dots + \widehat{B}_k X_k + u_i}; \text{ tal que } u_i \sim 0 \quad (56)$$

Donde la interpretación de sus parámetros es:

Si  $X_1 = X_2 = \dots = X_k = 0$ , la esperanza matemática de la Temperatura Media del Planeta (YT) será de  $\sqrt{\lambda \widehat{B}_0}$ .

$$E(YT) = \sqrt{\lambda \widehat{B}_0 + \widehat{B}_1(0) + \widehat{B}_2(0) + \dots + \widehat{B}_k(0)} \quad (57)$$

Por tanto:

$$E(YT) = \sqrt{\lambda \widehat{B}_0} \quad (58)$$

Si  $X_i = 1$ , y las demás variables independientes permanecen constantes, el incremento que experimentaría la Temperatura Media del Planeta (YT) sería de  $\Delta E(YT)$ .

$$E(YT)_{X_i} = \sqrt{\lambda \widehat{B}_0 + \widehat{B}_i X_i} = \sqrt{\lambda \widehat{B}_0 + \widehat{B}_i(1)} \quad (59)$$

Por tanto:

$$\Delta E(YT) = E(YT)_{X_i} - E(YT) \quad (60)$$

Como se puede observar, el análisis de regresión lineal múltiple permite entender la relación entre dos o más variables. Esto implica una variable dependiente y más de dos variables independientes, que, al ser estimada la variable dependiente, se puede predecir los cambios que producen las variables independientes sobre la dependiente.

### 3.3.2. Modelo de Regresión Gamma (MRG)

Si la YT es modelada mediante la regresión de Gauss, y esta no cumple con los supuestos de inferencia (normalidad y homocedasticidad en los  $u_i$ ), a pesar de las transformaciones Box – Cox, lo recomendable es buscar otra alternativa de modelamiento.

Si la YT es una v.a.c., tal que se aproxima a distribución gamma, es decir:  $YT \sim \Gamma(\alpha, \lambda)$ , su forma de modelamiento se basaría en la regresión lineal gamma, la cual, forma parte de los GLM (Modelos Lineales Generalizados). Los GLM son una continuación de los modelos lineales y se aplican en datos donde no hay normalidad y homocedasticidad. Los GLM tiene tres componentes (aleatorio, sistemático y la función link) (Dobson y Barnett, 2008):

**Componente aleatorio:** si YT es fexp, su función de densidad es:

$$f_x(X, \Theta) = h(X) * \exp [\eta(\Theta) T(X) - A(\Theta)] \quad (61)$$

Talque T(X), h(X), h(θ) y A(θ) son funciones conocidas (logístico, probit, binomial negativa, poisson, gamma, y exponencial, entre otros).

**Componente sistemático:** si YT es fexp, su forma de regresión lineal es:

$$E(YT) = g^{-1}(XB) \quad (62)$$

Dónde:

- E (YT) es la esperanza matemática (valor esperado) de YT.
- $XB = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k$  (componente lineal).
- $g^{-1}$  es la función link.

**Función link:** Permite modelar relaciones no lineales, sus funciones más comunes son las siguientes:

Función link	Identidad	Logaritmo natural	Logit	Reciproca	Raiz cuadrada	Exp.
Fórmula	$\mu_{Tp}$	$\ln(\mu_{Tp})$	$\ln\left(\frac{\mu_{Tp}}{n - \mu_{Tp}}\right)$	$\frac{1}{\mu_{Tp}}$	$\sqrt{\mu_{Tp}}$	$\mu_{Tp}^2$

**Cuadro 24.** Funciones link

**Fuente.** An Introduction to Generalized Linear Models (Dobson y Barnett, 2008).

Los GLM deben cumplir con los siguientes supuestos (McCullagh y Nelder, 1983):

**Supuesto 1:** selección del modelo de mayor ajuste, a través del AIC (Criterio de Información Akaike).

$$AIC = 2(P - LM) \quad (63)$$

Dónde:

- P = número de parámetros estimados.
- LM = verosimilitud del modelo actual.

El AIC compara modelos, “a menor AIC, mayor será su ajuste”.

**Supuesto 2:** los  $u_i$  (residuales) de la devianza se acercan a sus grados de libertad. Prueba de hipótesis:

$$H_0: u_i \sim gl \text{ vs. } H_a: u_i \neq gl$$



Si el P – valor es superior a 0.05 se acepta Ho, por tanto, los ui se aproximan a sus grados de libertad.

**Supuesto 3:** no de existir colinealidad entre las variables independientes (solo se aplica en los casos donde las variables independientes son paramétricas). Prueba de hipótesis:

$$H_0: \text{Cor}(X_i, X_j) \sim 0 \text{ vs. } H_a: \text{Cor}(X_i, X_j) \neq 0$$

Mediante el criterio del VIF (Factor de Inflación de Devianza), si el VIF es mayor a 5, existe colinealidad entre las variables independientes.

Con base en los supuestos de inferencia, el grado de ajuste de los modelos GLM es mediante la devianza (McCullagh y Nelder, 1983):

$$D^2 = \frac{D_N - D_R}{D_N}; \quad 0 \leq D^2 \leq 1 \quad (64)$$

Dónde:

- $D_N$  = devianza nula.
- $D_R$  = devianza de los residuales.

Si  $YT \sim \Gamma(\infty, \lambda)$  tal que  $0 \leq f(YT) \leq 1$ , su forma de regresión lineal es (McCullagh y Nelder, 1983):

$$YT = g^{-1} \left[ B_0 + \left( \sum_{i=1}^k B_i X_i \right) + ui \right]; \quad i = 1, 2, 3, \dots, k \quad (65)$$

Donde:

- $g^{-1}$  Es la función link.
- $B_0 + \left( \sum_{i=1}^k B_i X_i \right)$  Son los parámetros a estimar.

Mediante máxima verosimilitud, se tiene que (McCullagh y Nelder, 1983):

$$g^{-1}(YT) = \left[ B_0 + \left( \sum_{i=1}^k B_i X_i \right) + ui \right]; \quad i = 1, 2, 3, \dots, k \quad (66)$$

Tal que (McCullagh y Nelder, 1983):

<b>Función link</b>	<b>Ecuación</b>	
<b>Logaritmo natural</b>	$\ln(YT) = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + u_i$	(67)
<b>Raíz</b>	$\sqrt{YT} = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + u_i$	(68)
<b>Inversa</b>	$\frac{1}{YT} = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + u_i$	(69)
<b>Identidad</b>	$YT = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k + u_i$	(70)

**Cuadro 25.** Funciones link de la regresión gamma.

**Fuente.** Generalized Liner Models (McCullagh y Nelder, 1983).

Con el análisis de regresión gamma, se puede predecir los efectos que producen los factores de  $X_a$  sobre  $YT$ . Esto implica que se tenga que hacer una selección adecuada del modelo de mejor ajuste, a partir de las distintas funciones link que se puede utilizar, y del cumplimiento del supuesto de inferencia que engloban a los Modelos Lineales Generalizados (GLM).

### 3.4. Regresión Lineal Aditiva

La regresión lineal aditiva es un complemento de los Modelos Lineales (LM) y de los Modelos Lineales Generalizados (GLM), también conocidos como Modelos Aditivos Generalizados (GAM), cuya técnica estadística tiene su origen en 1990 por Robert Tibshirani y Trevor Hastie, donde el objetivo central es predecir la relación no lineal entre la variable dependiente ( $Y$ ) y las variables independientes ( $X$ 's) (Mercedes, 2009).

Si  $Y$  es una variable aleatoria (discreta o continua) que está en función de  $X$  o  $X$ 's variables independiente, donde la relación no es lineal y existe una función desconocida entre ellas, es recomendable la utilización de funciones suavizadas, es decir, funciones específicas de predicción (García, 2022).

Sea  $Y$  un vector aleatorio  $n$  – dimensional de distribución de probabilidad desconocida y  $X$  un vector  $n$  – dimensional de constantes conocidas, tal que  $f : R \rightarrow R$ , que se llamaría función suavizada desconocida (García, 2022).

Por definición, si  $C_1, C_2, \dots, C_k$  son una base de funciones de modo que toda  $f$  se puede representar como una combinación lineal de los elementos de la base (García,2022):

$$f(X) = \sum_{j=1}^k C_j(X)B_j \quad (71)$$

Tal que  $B_j$  son  $k$  parámetros a estimar.

Si  $P_n[X]$  es el espacio de polinomios de orden  $n$ , con dimensión  $n + 1$  y base canónica  $X^n, X^{n+1}, \dots, X$ , por tanto (García, 2022):

$$f(X) = B_0 + B_1X + B_2X^2 + \dots + B_nX^{n+1} + u_i \quad (72)$$

Cuya base polinómica es dar ajustes puntuales, no del comportamiento general de un conjunto amplio de datos. Si se extiende el fenómeno a más variables, se asume la interacción entre variables, teniendo como efecto, superar el tamaño de la muestra.

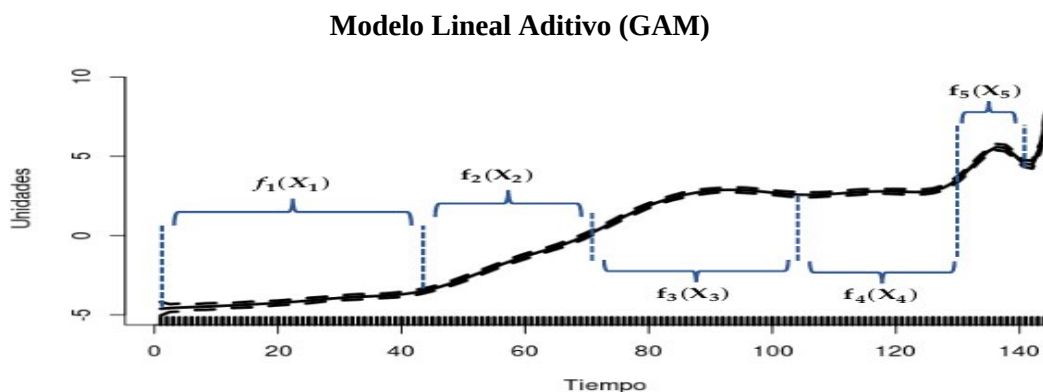
Si el fenómeno en estudio tiene una suma de efectos con relaciones no lineales entre algunos elementos, dicha relación deben ser una suma para predecir el comportamiento de dicho fenómeno, es decir, los efectos pueden ser modelados por una suma de funciones de cada elemento que conforman al objeto de estudio (Molner, 2021):

$$E(Y) = B_0 + \sum_{i=1}^p f_i(X_i) \quad (73)$$

Donde:

- $E(Y)$  es el valor esperado de  $Y$ .
- $B_0$  es la ordenada al origen, o bien, el valor esperado de  $Y$  cuando las demás funciones permanecen constantes.
- $f_i(X_i)$  es el conjunto de funciones que predicen al fenómeno en estudio.

Gráficamente esto es (Molner, 2021):



**Imagen 21.** Gráfico de GAM

**Fuente.** Interpretable Machine Learning 2021

Cuya estructura es (Molner, 2021):

$$E(Y) = B_0 + f_1(X_1) + f_2(X_1) + f_3(X_1) + \dots + f_p(X_1) \quad \text{Corte longitudinal}$$

$$E(Y) = B_0 + f_1(X_1) + f_2(X_2) + f_2(X_1) + \dots + f_p(X_p) \quad \text{Corte transversal}$$

Donde el ajuste del modelo es de la siguiente forma (Molner, 2021):

$$D^2 = \frac{D_N - D_R}{D_N}; \quad 0 \leq D^2 \leq 1 \quad (74)$$

Dónde:

- $D_N$  = devianza nula.
- $D_R$  = devianza de los residuales.

Partiendo de la estructura del modelo, debe cumplir con los siguientes supuestos (Molner, 2021):

Corte longitudinal	Corte transversal
<ul style="list-style-type: none"> <li>• Los <math>u_i</math> deben aproximarse a una distribución normal.</li> <li>• Los <math>u_i</math> deben ser homocedasticos</li> </ul>	<ul style="list-style-type: none"> <li>• Los <math>u_i</math> deben aproximarse a una distribución normal.</li> <li>• Los <math>u_i</math> deben ser homocedasticos.</li> <li>• Las variables independientes (X's) no deben tener colinealidad.</li> </ul>

Esta técnica (GAM) dará los elementos necesarios para hacer estimaciones en tiempo del Índice de Actividad Humana (IAH), y con ello, pronosticar sus efectos en la Temperatura Media del Planeta (YT).

### 3.5. Conclusiones

En este capítulo se presentaron todas las técnicas estadísticas y matemáticas teóricas que se utilizan para la construcción de dos modelos:

- En el primer modelo, se utilizan las técnicas y el modelo de la minimización de la dimensión, con ellas, se elige el de mejor ajuste, con la finalidad de obtener un indicador de  $X_a$ .
- En el segundo modelo, se predice los efectos que producen los factores de  $X_a$  sobre YT, por tanto, se hacen dos regresiones, de las cuales se selecciona el de mejor ajuste, a partir del cumplimiento de los supuestos de inferencia.
- En el tercer modelo, se estima el grado del IAH en el tiempo y sus efectos en YT.