CAPÍTULO 2

2. MARCO TEÓRICO

En este capítulo se introducirá el marco teórico que se utilizó en la investigación de la incidencia que tiene la Prueba de Aptitud Académica (PAA) en el rendimiento académico.

Es decir, esto corresponde a la descripción de la población estudiada, las variables con sus respectivas codificaciones y la teoría de las técnicas y métodos utilizados en este estudio.

2.1. Descripción de los datos

Se ha considerado para el estudio de la incidencia de las pruebas de aptitud académica sobre el rendimiento académico, a la población de estudiantes que ingresaron a ingeniería de la ESPOL en 1999, mediante aprobación del curso prepolitécnico y/o examen de ingreso.

Para la realización del estudio se requirió la información de las instituciones o departamentos de la ESPOL, encargadas de los registros de académicos de ingreso: Oficina de Ingresos y de los registros de las materias básicas: CRECE y CESERCOM.

Las variables tomadas en consideración están únicamente relacionadas con la identificación demográfica del estudiante y su rendimiento académico, el rendimiento académico esta basado en las calificaciones obtenidas en las pruebas de ingreso y en materias básicas dentro de la ESPOL. Estas variables serán distribuidas de la siguiente manera:

- Variables Demográficas: Son variables que identificarán demográficamente al estudiante: sexo, tipo de colegio de culminación de los estudios secundarios, especialidad, y ubicación geográfica del colegio.

- Variables consideradas en el proceso de admisión a la ESPOL: Son variables que medirán el rendimiento del estudiante en el proceso de admisión a la ESPOL: pruebas de ingreso (Física, Matemática y Química) aprobadas por el estudiante, en cualquier período, antes de su ingreso en el año 1999; y las pruebas de aptitud (Verbal y Matemática), en cualquiera de los períodos, del año 1999.

- Variables consideradas dentro de la ESPOL, Materias Básicas: Son variables que medirán el rendimiento del estudiante en su transcurso académico por la Ingeniería Básica: Cálculo (I y II), Física (I y II), Álgebra Lineal y Química I.

Cada materia básica se analizará desdoblándola en cuatro variables, cada una representará la calificación que se obtuvo en la materia tomando en cuenta las repeticiones de la materia.

TABLA I

DESCRIPCIÓN DE LAS VARIABLES DE ESTUDIO

Tipo	Nombre de la Variable	Codificación/ Valor Numérico	
DEMOGRÁFICAS	**Sexo**	Mujer	0
		Hombre	1
	Tipo de Colegio	Fiscal	1
		Particular (laico)	2
		Particular (religioso)	3
	Especialidad	Físico matemático	1
		Químico biológico	2
		Informática	3
		Eléctrica	4
		Electrónica	5
		Mecánica	6
		Otros	7
	Provincia	Provincia de donde obtuvo su título, en su respectiva ciudad	
PRUEBAS DE INGRESO	**Conocimiento**	Pruebas que han sido evaluadas de 0 (la mínima calificación) a 100 (la máxima calificación) puntos. Siendo números reales con dos decimales.	
	Matemáticas		
	Física		
	Química		
	Aptitud	Pruebas que han sido evaluadas de 0 (la mínima calificación) a 100 (la máxima calificación) puntos. Siendo números reales con dos decimales.	
	Matemática		
	Verbal		
MATERIAS BÁSICAS	**Cálculo I** **(PCI: 1,2,3,4)** **Cálculo II** **(PCII: 1,2,3,4)** **Física I** **(PFI: 1,2,3,4)** **Física II** **(PFII: 1,2,3,4)** **Álgebra** **(PA: 1,2,3,4)** **Química I** **(PQ I: 1,2,3,4)**	Pruebas que han sido evaluadas de 0 (la mínima calificación) a 100 (la máxima calificación) puntos. Siendo números reales con dos decimales.	

Antes de realizar el análisis estadístico, se debe tomar en cuenta algunas consideraciones con respecto al tamaño de la población y a algunas de las variables de nuestro estudio.

En la siguiente tabla se presenta la cantidad de alumnos que ingresaron a las Ingenierías de la ESPOL en el año 1999, tomando en cuenta los períodos en los que ingresaron: invierno y verano.

TABLA II

ESTUDIANTES QUE INGRESARON A LA ESPOL EN 1999

	Invierno		Verano		Total	
No están en la ESPOL.	35	***10.09%***	7	***2.02%***	42	***12.10%***
No están en ciclo básico de ingeniería ESPOL o no han sido de mayor aportación al estudio.	37	***10.66%***	54	***15.56%***	91	***26.22%***
Muestra del estudio.	175	***50.43%***	39	***11.24%***	214	***61.67%***
Total de alumnos que ingresaron 1999 en la ESPOL.	247	***71.18%***	100	***28.82%***	347	***100.00%***

Es de mencionar que el estudio se baso en el seguimiento del *61,67 %* del total de estudiantes que ingresaron en el año 1999, aceptándose como una muestra representativa para obtener una buena información estadística; así mismo es de mencionar que aunque el *12,10%* no se encuentran registrados en la ESPOL, esto puede deberse a que aún no terminan el bachillerato o simplemente tomaron otra opción de estudio universitario.

Por lo general, con respecto a las materias básicas de Ingeniería, no todos los estudiantes aprueban las materias a la primera vez que la cursan, o en la segunda o tercera; por ello se tomó en consideración las notas que se obtuvo en cada vez que fue tomada la materia, es decir, aparecerán cuatro variables por cada materia, ya que en los reglamentos de la ESPOL, determina que una materia puede ser cursada hasta por una cuarta ocasión.

En la construcción de estas nuevas variables hay que evitar que se originen vacíos en los registros, por ej.: si un estudiante aprobó en la primera vez, las otras casillas (2da, 3era y 4ta vez) estarían vacías, si un estudiante tomo el curso pero no lo aprobó y no lo tomo de nuevo, también provocaría que haya datos faltantes.

Por está razón, para evitar datos faltantes se realizó una imputación de datos, la cual consistió en completar los datos con la nota máxima obtenida por el estudiante en la materia, ya sea que la haya aprobado o no.

2.2. Análisis Descriptivo

2.2.1. Conceptos Básicos

Para entender los términos estadísticos utilizados en análisis posteriores, se debe conocer algunos conceptos básicos citados a continuación.

El proceso que se realiza para obtener una observación o medición cualquiera es lo que se denomina **Experimento**. Un experimento esta asociado a varios eventos elementales o también denominados resultados posibles.

El conjunto de todos los resultados posibles de un experimento se llama **espacio muestral,** denotado por Ω. A todo subconjunto $A \subset \Omega$ es llamado evento; Ω es un evento cierto, $\varnothing$ un evento imposible. Si $\omega \in \Omega$ entonces ω es llamado un evento simple. Una clase A de subconjunto Ω es llamada álgebra de subconjuntos de Ω, si satisface las siguientes propiedades:

1. $\Omega \in$ A
2. Si $A \in$ A entonces $A^c \in$ A
3. Si $A \in$ A y $B \in$ A entonces $A \cup B \in$ A

Además si cumple que:

4. Si $A_n \in$ A para n=1,2,3,..., entonces $\bigcup_{n=1}^{\infty} A_n \in$ A

Es llamada una σ- álgebra de subconjuntos de Ω.

Una función *P* definida en una σ- álgebra $\mathcal{A}$ es llamada medida de probabilidad en $\mathcal{A}$ o simplemente probabilidad en $\mathcal{A}$ si cumple que:

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. Si $A_1,.., A_n$ son disjuntos (2 a 2) entonces

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$$

(dos eventos son disjuntos 2 a 2, si son mutuamente excluyentes. i.e. $A_i \cap A_j = \emptyset$, $i \neq j$.

Una **variable aleatoria** *X* en un espacio de probabilidad $(\Omega, \mathcal{S}, P)$ es una función definida en Ω tal que $[X \leq x]=\{\omega \in \Omega \,/\, X(\omega)<x\}$ es un evento aleatorio para todo $x \in \mathbb{R}$; i.e., $X: \Omega \rightarrow \mathbb{R}$, es una variable aleatoria si $[X \leq x] \in \mathcal{S} \quad \forall x \in \mathbb{R}$.

La función de distribución para una variable aleatoria X se define por:

$$F_X(x) = P(X \leq x)$$

La variable aleatoria X puede ser ***discreta*** si y sólo si le otorgan valores finito o infinito contable $\{x_1, x_2, \ldots\} \subset \mathrm{R}$ tal que $X(\omega) \in \{x_1, x_2, \ldots\}$ $\forall \omega \in \Omega$. La función de probabilidad de una variable aleatoria discreta es:

$$F_X(x) = P(X \leq x) = \sum_{t \leq x} p(t)$$

Una variable aleatoria es (absolutamente) continúa si existe una función $f(x) \geq 0$ tal que:

$$F_X(x) = \int_{-\infty}^{x} f(t)\, \partial t \ , \forall x \in \mathrm{R}$$

En este caso decimos que f es una función de probabilidad de X o simplemente la densidad de X.

Parte de esta investigación consistirá en el estudio de las observaciones y sus características que estas poseen, así pues tenemos que el conjunto total de observaciones correspondientes a una característica de interés se llama **Población**, a la cual se le extrae un subconjunto de la población que contiene las observaciones obtenidas mediante una selección a la cual se le llama **Muestra** de la población.

Una forma estadística de describir las características de la población es por medio de ciertos valores llamados **Parámetros,** en general estos parámetros no son fáciles de encontrar, por ello se utilizan ciertas variables aleatorias para estimarlos, las cuales son denominadas **Estimadores,** en general los estimadores se basan en los datos de una muestra.

Si *X* es una variable aleatoria discreta, *f(x)* es el valor de la función de probabilidad en *x* y *g(x)* es una función de *x*, se define el ***valor esperado*** de *g(x)* como una transformación de *E* sobre *g(x)*, tal que:

$$E[g(x)] = \sum_{x} f(x).g(x)dx$$

Si *X* una variable aleatoria continua con densidad de probabilidad *f(x)*, y *g(x)* una función de x, el valor esperado de *g(x)* es la transformación de *E* sobre *g(x)*, tal que:

$$E[g(x)] = \int_{-\infty}^{\infty} f(x).g(x)dx$$

El **r-ésimo momento con respecto al origen** de la variable aleatoria *X* es el valor esperado de x^r, representado por m'_r, así pues se tiene para *X* discreta:

$$\mu'_r = E(x^r) = \sum_{x} x^r . f(x) \quad , r = 0,1,2,...$$

para X continua

$$\mu'_r = E(x^r) = \int_{-\infty}^{\infty} x^r . f(x)dx$$

El **r-ésimo momento con respecto a la media** de la variable aleatoria X es el valor esperado de $(x\text{-}m)^r$, representado por m_r, se tiene para X discreta:

$$\mu_r = E\left[(x-\mu)^r\right] = \sum_x (x-\mu)^r . f(x) \quad , r = 0,1,2,...$$

para X continua:

$$\mu_r = E\left[(x-\mu)^r\right] = \int_{-\infty}^{\infty} (x-\mu)^r . f(x)dx$$

Entre los principales parámetros poblacionales tenemos a la **Media poblacional** que se define como:

$$\mu = \sum_{-\infty}^{\infty} x.P(X = x) \quad , X \text{ una variable aleatoria discreta}$$

$$\mu = \int_{-\infty}^{\infty} x.f(x)dx \quad , X \text{ es una variable aleatoria continua}$$

El estimador más usado de la media poblacional es **la media aritmética** que es el promedio de $X_1, X_2, ..., X_n$, n observaciones de una muestra de la población:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Otro parámetro poblacional es la **Mediana** que se define de la siguiente manera: Cuando *X* es una variable aleatoria continua es el valor central $X_{1/2}$ de una distribución:

$$\int_{-\infty}^{X_{1/2}} f(x)dx = 1/2$$

donde *f(x)* es la función de densidad de la variable aleatoria *X*.

Donde el estimador de la mediana es el valor que posee $X_{(n/2)}$, donde $X_{(1)}$, $X_{(2)}$,..., $X_{(n/2)}$,..., $X_{(n)}$, *n* observaciones ordenadas de una muestra, si n es impar se toma el promedio de las dos observaciones de la mitad.

La **varianza poblacional** (s^2) Medida de variabilidad que se define como la media del cuadrado de las diferencias de las observaciones con respecto a la media poblacional;

$$s^2 = E\left[(X - m)^2\right]$$

Una medida de dispersión de las observaciones alrededor de la media poblacional es la **Desviación Estándar**, definiéndose como la raíz cuadrada positiva de la varianza poblacional:

$$\sigma = \sqrt{\sigma^2}$$

Se define la ***covarianza*** entre *X* y *Y* como:

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

Una medida de la relación lineal entre dos variables aleatorias *X* y *Y* es el **Coeficiente de correlación,** definido por :

$$\rho = \frac{Cov[X,Y]}{\sigma_X . \sigma_Y}$$

Otro parámetro poblacional es el **Sesgo** que mide la simetría de la distribución de los datos de una población alrededor de la media.

El sesgo se calcula como: $\alpha_3 = \frac{\mu_3}{\sigma^3}$

Si el **Sesgo es positivo** quiere decir que los datos se sesgan hacia la derecha, si el **Sesgo es negativo** los datos se sesgan hacia la izquierda y si el **Sesgo es cero:** Se dice que los datos tienen distribución simétrica, entonces la media, la mediana y la moda son iguales.

La **Kurtosis** mide la picudez de la distribución de los datos de una población.

La Kurtosis se calcula como: $\alpha_4 = \frac{\mu_4}{\sigma^4}$

Si la distribución de los datos posee la forma de una distribución normal se denomina **Mesocúrtica**, si la distribución es achatada con respecto a una normal se denomina **Platicúrtica**, y **Leptocúrtica** si la distribución más puntiaguda que una normal.

Prueba de hipótesis: Se utiliza para comprobar si las conjeturas con respecto a la distribución o los parámetros de una población son ciertas o no. La conjetura que se busca aceptar o rechazar se denomina ***hipótesis nula*** **(H_0),** y la conjetura que se constratará se denomina ***hipótesis alterna (H_1)***.

La región o área que se utiliza para realizar esta inferencia se llama **región crítica de la prueba,** que representa un subconjunto R^n, tal que:

$$C = \{(x_1, x_2, ..., x_n) \in R^n \text{ / } rechaza\ H_0\}$$

Para reconocer si una prueba es rechazada o no se utiliza el **Valor p** que es el mínimo nivel de significancia de la prueba.

Una prueba estadística de hipótesis que se aplica cuando se quiere verificar si la distribución de una población es la **Prueba de Kolmogorov-Smirnov,** cuyo estadístico de prueba es:

$$D = max\left|\hat{F}(X_i) - F_0(X_i)\right|$$

donde $F_0(X_i)$ es la distribución acumulada propuesta en H_0 y $\hat{F}(X_i)$ es la distribución empírica que se define de la siguiente manera:

$$\hat{F}(X) = \begin{cases} 0 & X < X_{(1)} \\ k/n & X_{(k)} \leq X < X_{(k+1)} \\ 1 & X_{(n)} \leq X \end{cases}$$

donde $X_{(1)}$, $X_{(2)}$,.....,$X_{(n)}$ es la notación de la serie de datos ordenados.

2.3. Análisis de Varianza

El método de Análisis de Varianza trata de analizar la variación de una respuesta y de asignar porciones de esta variación a cada de las variables de un conjunto de variables independientes.

El modelo a aplicar es Análisis de Varianza de un Factor y se denomina así porque solo se considera el efecto de un factor sobre variable a ser explicada.

El procedimiento de este análisis consiste en la extracción de una muestra aleatoria *n*, a la cual esta siendo afectada por *k* niveles o tratamientos del factor, este procedimiento cumple el siguiente modelo:

Sea y_{ij} variables aleatorias independientes (muestra aleatoria) proveniente de una población normal con media m_i y varianza s^2, siendo $i=1,2, \ldots ,k$ (k número de niveles o tratamientos del factor); $j=1,2, \ldots ,n$ (n es el número de la muestra), entonces tenemos que:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

donde y_{ij} es la (ij)-ésima observación, μ es un parámetro común a todos los tratamientos denominado media global, ti es el efecto del nivel o tratamiento i-ésimo y e_{ij} es la componente aleatoria del error, bajo el supuesto de que a media de e_{ij} es igual 0 y la varianza igual a s^2.

Lo que se desea es probar hipótesis de que hay o no algún efecto de los tratamientos del factor sobre la variable explicada y_{ij}, es decir:

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0$$
$$vs$$
$$H_1 : \neg H_0$$

La tabla donde se muestran los resultados se denomina *Tabla ANOVA* la cual contiene las sumas cuadráticas (SC) y medias cuadráticas (MC) y grados de libertad, para los tratamientos del Factor y el Error producido por el modelo; el estadístico de la prueba, F con distribución de Fisher, y el nivel de significancia, con el 95 % de confianza. A continuación se presenta en el siguiente cuadro:

Fuente Variación	Suma Cuadrática	Grados de Libertad	Media Cuadrática	F	Valor p
Tratamientos	SCTr	*k-1*	SCT/(*k-1)*	SCTr/ SCE	Valor de significancia con el 95% de confianza
Error	SCE	*n-k*	SCE/ *n-1*		
Total	**SCT**	*n-1*			

Donde,

$$SCT = \sum_{i}^{k} \sum_{j}^{n} \left(y_{ij} - \bar{y} \right)^2$$

$$SCTr = \sum_{i}^{k} \sum_{j}^{n} \left(\bar{y}_{i.} - \bar{y} \right)^2$$

$$SCE = \sum_{i}^{k} \sum_{j}^{n} \left(y_{ij} - \bar{y}_{i.} \right)^2$$

$$\bar{y}_{i.} = \sum_{j}^{n} \left(\frac{y_{ij}}{n} \right)$$

2.4. Análisis Factorial

2.4.1. Conceptos Básicos

Vector aleatorio

Sean $X_1, X_2, \ldots, X_p$, p variables aleatorias de un estudio. Se define un vector p variado $\mathbf{X} \in R^p$, como aquel que está compuesto por p variables aleatorias:

$$\mathbf{X}^t = \begin{bmatrix} X_1 & X_2 & \cdots & X_p \end{bmatrix}$$

Matriz de datos

Sea $X_1, X_2, \ldots, X_n$ es una muestra tomada de una población de tamaño n que tiene p variables o características de interés.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & . & . & . & x_{1n} \\ x_{21} & x_{22} & . & . & . & x_{2n} \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ x_{p1} & x_{p2} & . & . & . & x_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} & \mathbf{X_2} \cdots \mathbf{X_n} \end{bmatrix}, \quad \mathbf{X_i} \in R^P$$

Vector de medias

Sea: $\boldsymbol{X}$ un vector p variado, es decir, compuesto por p variables aleatorias, se define a su vector de medias como:

$$\boldsymbol{\mu} = E[\mathbf{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_p] \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_p \end{bmatrix}$$

Matriz de varianzas y covarianzas

Sea un $\boldsymbol{X}$ vector p variado, se define para éste la matriz de varianzas y covarianzas como:

$$\mathbf{S} = E\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t\right]$$

$$\mathbf{S} = \begin{bmatrix} \operatorname{cov}(X_1, X_1) & \operatorname{cov}(X_1, X_2) & . & . & . & \operatorname{cov}(X_1, X_p) \\ \operatorname{cov}(X_2, X_1) & \operatorname{cov}(X_2, X_2) & . & . & . & \operatorname{cov}(X_2, X_p) \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \operatorname{cov}(X_p, X_1) & \operatorname{cov}(X_p, X_2) & . & . & . & \operatorname{cov}(X_p, X_p) \end{bmatrix} .$$

$$\mathbf{S} = \begin{bmatrix} s_1^2 & s_{12} & . & . & . & s_{1p} \\ s_{21} & s_2^2 & . & . & . & s_{2p} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ s_{p1} & s_{p2} & . & . & . & s_p^2 \end{bmatrix}$$

donde $\sigma_{ij}=\sigma_{ji}$, por lo tanto, S es simétrica y por tanto diagonalizable ortogonalmente definida positivamente.

Matriz de correlación

Sea S la matriz de varianzas y covarianzas de un vector aleatorio $\boldsymbol{X} \in$ R^P, defínase $\boldsymbol{V^{1/2}}$ como la matriz de desviaciones estándar de $\boldsymbol{X}$, como sigue:

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{s_{11}} & 0 & . & . & . & 0 \\ 0 & \sqrt{s_{22}} & . & . & . & 0 \\ . & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ 0 & 0 & . & . & . & \sqrt{s_{pp}} \end{bmatrix}$$

donde: $\sqrt{s_{ii}}$ es la desviación estándar de la variable aleatoria xi, entonces se puede deducir que:

$$\mathrm{r} = \begin{bmatrix} 1 & r_{12} & . & . & . & r_{1p} \\ r_{21} & 1 & . & . & . & r_{2p} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ r_{p1} & r_{p2} & . & . & . & 1 \end{bmatrix} = \left(\mathbf{V}^{1/2}\right)^{-1} \mathrm{S} \left(\mathbf{V}^{1/2}\right)^{-1}$$

donde: ρ_{ij} es el coeficiente de correlación entre la variable X_i i=1,2,...,p y X_j, j=1,2,...,p.

2.4.2. Métodos Factoriales

Es una técnica estadística que tiene como por objetivo general reducir la información obtenida por un número de variables originales en un pequeño grupo de nuevas variables construidas con una pérdida mínima de información.

El Modelo Ortogonal de Factores

Sea un vector aleatorio $\boldsymbol{X}$ de p variables con un vector de medias m y matriz de varianzas y covarianzas S.

El modelo de factores postula que X es linealmente dependiente sobre un grupo de variables no observables $\boldsymbol{F_1, F_2, \ldots, F_m}$, llamados factores y p fuentes de variación, $e_1, e_2, \ldots, e_p$, llamados errores o factores específicos

El modelo de análisis de factores es:

$$
\begin{gathered}
X_1 - m_1 = l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + e_1 \\
X_2 - m_1 = l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + e_1 \\
\vdots \\
X_p - m_p = l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + e_1
\end{gathered}
$$

En notación matricial,

$$
\underset{(p\times 1)}{X - m} = \underset{(p\times m)}{L}\;\underset{(m\times 1)}{F} + \underset{(p\times 1)}{e}
$$

El coeficiente l_{ij} es llamada la carga de la *i-ésima* variable en el *j-ésimo* factor, donde la matriz L es la matriz de cargas.

Donde F y e son independientes y:

E(F) = 0, *Cov (F)* = I, I es la matriz de identidad.

E(e) = 0, *Cov* (e) = Y, donde Y es una matriz diagonal.

Estructura de la Matriz de Varianzas y Covarianza para el modelo de factor ortogonal.

1. *Cov* (**X**)= S = $\mathbf{LL}'$ + Y, ó

$$Var(X_i) = \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i$$
$$Cov(X_i, X_j) = \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km}$$

2. *Cov(**X, F**) =* ***L,*** ó

$$Cov(X_i, F_j) = \ell_{ij}$$

Método de Componentes Principales

Es un método multivariado que tiene por objetivo la reducción de datos. En este estudio se lo utilizará como método de estimación de los parámetros ***L*** (matriz de cargas) y *Y* (varianza especifica), de la ecuación que representa la estructura de la matriz de varianzas y covarianzas del modelo de factor ortogonal.

Sea matriz Σ de varianzas y covarianzas o ***r*** la matriz de correlación, y sea *l* y ***e*** su valor y vector propio respectivamente con $l_{\mathbf{1}}{}^3\, l_{\mathbf{2}}{}^3$ ¼ $^3\, l_{\mathbf{p}}{}^3$ ***0,*** donde m<p es el número de factores. Entonces la matriz cargas de Los factores esta dad por:

$$\tilde{L} = \left[\sqrt{\tilde{\lambda}_1}e_1, \sqrt{\tilde{\lambda}_2}e_2, \ldots, \sqrt{\tilde{\lambda}_m}e_m\right]$$

La matriz de varianzas específicas está provista por Los elementos diagonales de la matriz **R-LL,** *R* es la matriz estimada de correlación.

$$\tilde{\Psi} = \begin{bmatrix} \tilde{\psi}_1 & 0 & \dots & 0 \\ 0 & \tilde{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \tilde{\psi}_m \end{bmatrix} \quad con \quad \tilde{\psi}_i = r_{ii} - \sum_{j=1}^{m} \tilde{l}\, ij^{\,2}$$

Rotación de Factores por Varimax

La rotación de factores tiene por objetivo redistribuir la varianza en factores, para un mejor entendimiento de estos. Teóricamente la matriz de carga multiplicada por la una matriz ortogonal representa una matriz rotada.

Se utilizó para este estudio el método de rotación Varimax que maximiza la suma de varianzas de las cargas requeridas en la matriz de factores, por medio de la rotación de los factores teniendo como finalidad en acercamiento de entre altas cargas (+1 o -1) y de algunas cargas cercanas a cero.