ESCUELA SUPERIOR **POLITECNICA** DEL **LITORAL**

Instituto de Ciencias Matemáticas

"MEDIDAS ALTERNATIVAS EN EL ANALISIS DE DATOS APLICADA AL RENDIMIENTO ESTUDIANTIL, DESARROLLO MODELO DE MARKOV"

TESIS DE GRADO

Previa a la obtencion del Titulo de:

INGENIERO EN ESTADISTICA IMFORMATICA

Presentada por:

JORGE FABRICIO CUEVARA VIEJO

GUAYAQWIL - ECUADOR

AÑO

2000

AGRADECIMIENTO

Agradezco a Dios por ser mi compañero incondicional en todo momento, a todos aquellos que de una u otra manera colaboraron en la culminación de mi carrera universitaria y en especial al Mat. Fernando Sandoya por sus sabios consejos en el desarrollo de la presente Tesis de Grado.

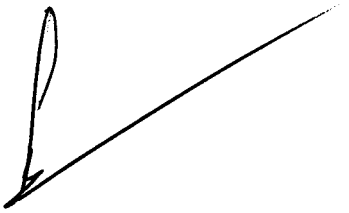
DEDICATORIA

Dedico la presente Tesis de Grado, Aquel que me regalo la vida y el mundo, Dios.

A mis padres, Jorge y Zoila, mis hermanos, Katherine y Alvaro como una muestra del caritio y respeto que siento por ellos.

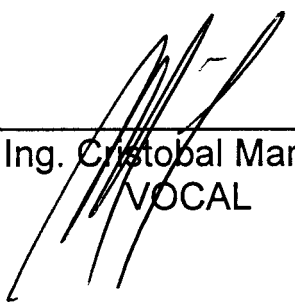
Y a todos aquellos que algun dia confiaron en mi.

TRIBUNAL DE GRADUACIÓN

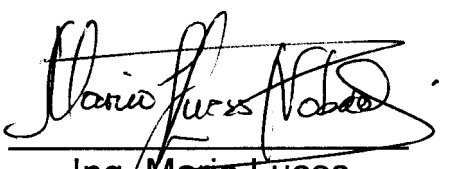


Ing. **Felix** Ramirez
DIRECTOR DEL ICM

Mat. Fernando Sandoya
DIRECTOR DE TESIS



Ing. Cristobal Mariscal
VOCAL



Ing. Mario Luces
VOCAL

DECLARACIÓN EXPRESA

"La responsabilidad del contenido de esta Tesis de Grado, me corresponden exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL"

(Reglamento de Graduación de la ESPOL).

Jorge F. Guevara V.

RESUMEN

En el presente estudio se realiza un análisis del rendimiento estudiantil en las carreras de Economia y Gestidn Empresarial e Ingeniería Estadistica Informatica mediante la aplicación del Análisis de Factores junto con nuevos criterios en la retención de factores. Luego con el conocimiento obtenido en el análisis anterior se procede al desarrollo un modelo de Markov que pronostica el número de estudiantes en cada nivel y por consiguiente los que concluyen el programa de estudios.

INDICE GENERAL

INDICE DE FIGURAS

INDICE DE TABLAS

INTRODUCCION

El presente trabajo trata sobre el uso de medidas alternativas en el análisis de datos y el desarrollo de un Modelo de Markov, con una aplicacion en el Rendimiento estudiantil, en las Carreras de Economia e Ingenieria Estadistica Informática, que son Carreras de reciente creación en la **ESPOL.**

El estudio se encuentra dividido tacitamente en dos partes. La primera parte, Capitulo I y II, en la que se expone toda la teoria referente al Analisis de Factores y las Cadenas de Markov. En la segunda parte, Capitulo III y IV, se realiza un estudio de la estructura de las carreras objeto de estudio, la aplicacion del Analisis de Factores y la presentación del numero equivalente como una medida alternativa para la selección del numero de factores a retener.

En el Capitulo IV, se desarrolla un Modelo Matematico de Markov que pronostica el numero de graduados, y a su vez el numero de estudiantes en cada nivel en cada Carrera.

Capitulo 1

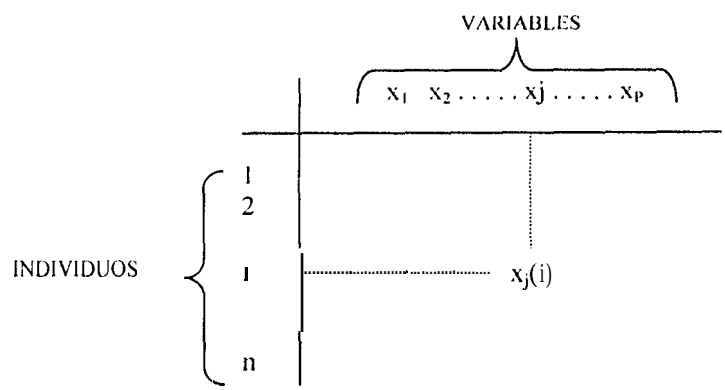
I - METODOS FACTORIALES Y DE CLASIFICACION PARA EL TRATAMIENTO DE VARIABLES CUALITATIVAS Y CUANTITATIVAS EN EL ANALISIS DE DATOS

1.1 Tipos de Tablas de Datos Multidimensionales

Llamamos datos multidimensionales al conjunto de valores de un cierto numero de variables estadisticas sobre individuos de una poblacion. Se puede considerar como la realización de un vector aleatorio definido sobre la poblacion en la que sus valores estan en un espacio preciso. Consideremos por ejemplo el sexo, la edad, el rendimiento estudiantil (promedio), nivel socioeconomico (factor p) de una persona cualquiera. Sea un individuo de sexo masculino, edad 20 años, promedio 7.26, factor socioeconomico 8. El dato multidimensionai (masc., 20, 7.26, 8) es la realizacion

del vector aleatorio (sexo, edad, promedio, factor p) definido sobre el grupo de estudio.

Una tabla de datos multidimensionales es pues asi, una muestra de un vector aleatorio: las mismas variables son medidas sobre un cierto numero de individuos y estas se presentan en la mayoría de los casos de la siguiente forma:



El termino de la i-esima línea y la j-esima columna es el valor para la variable x_j sobre el individuo i. En la practica pueden existir diferentes tipos de tablas de datos, tales como:

Tablas de Individuos por Caracteres Cuantitativos

Tablas de Contingencia

Tablas Lógicas

Tablas Disyuntivas Completas

Tablas de datos ordinales

Tablas de distancias, o de proximidades

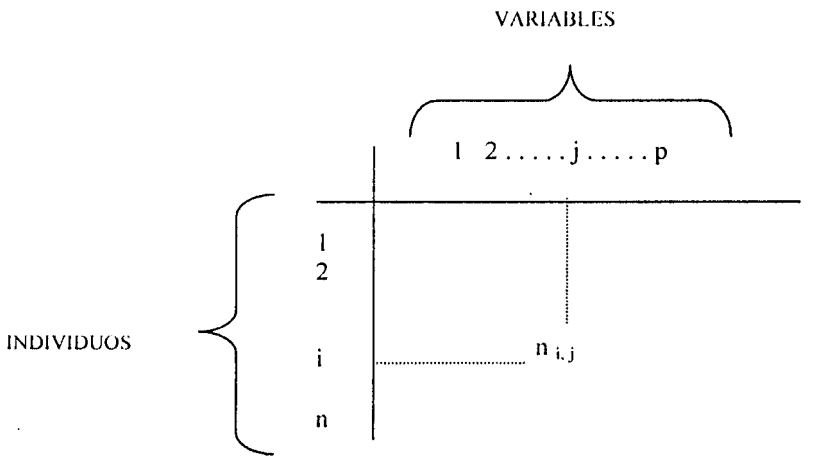
Tablas Mixtas

1.1.1 Tabla de Individuos por Caracteres Cuantitativos

Este tipo de tablas es una de las **mas** simples; las componentes de un vector aleatorio $(x_1, ... x_p)$ son variables cuantitativas o valores reales, no necesariamente continuos. El termino $x_j(i)$ es un numero real, representando la medida de la variable x_j sobre el individuo i.

1.1.2 Tablas de Contingencia

Una tabla de contingencia es la reparticion de una poblacion estadistica formada de caracteres cualitativos asignando cada una en modalidades exhaustivas y exclusivas la una de las otras. La tabla se presenta de la siguiente forma:

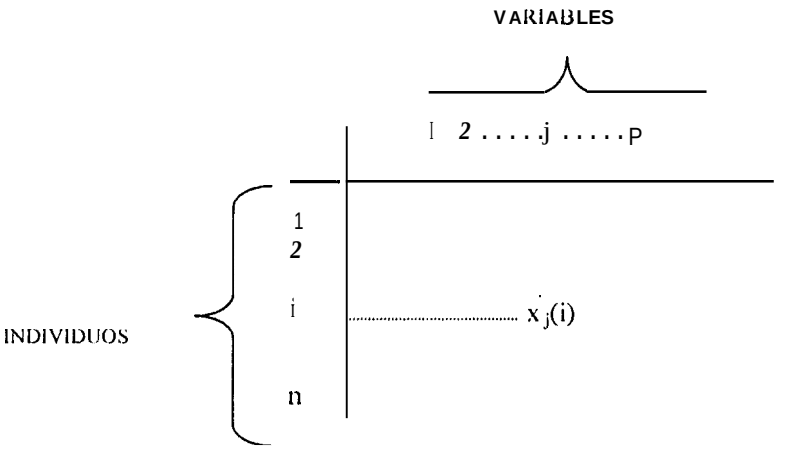


El termino $n_{i,j}$ de la i-ésima línea y la j-esima columna representa el numero de individuos que poseen la modalidad i de la caracteristica 1 y la modalidad j de la caracteristica 2. Los datos socioeconomicos se presentan frecuentemente en la forma de tablas de contingencia, pues frecuentemente las variables estudiadas son cualitativas.

1.3 Tablas Lógicas, tablas disyuntivas completas.

Las tablas Iogicas indican, para cada individuo de una poblacion estadistica, la pertenencia a un grupo particular o si este es equivalente, o la modalidad del dato cualitativo que esta posee. El codigo usado es el codigo lógico: la

pertenencia es representada por 1, la no pertenencia por 0. Cada individuo pertenece a un solo grupo, los datos se presentan de la siguiente forma:



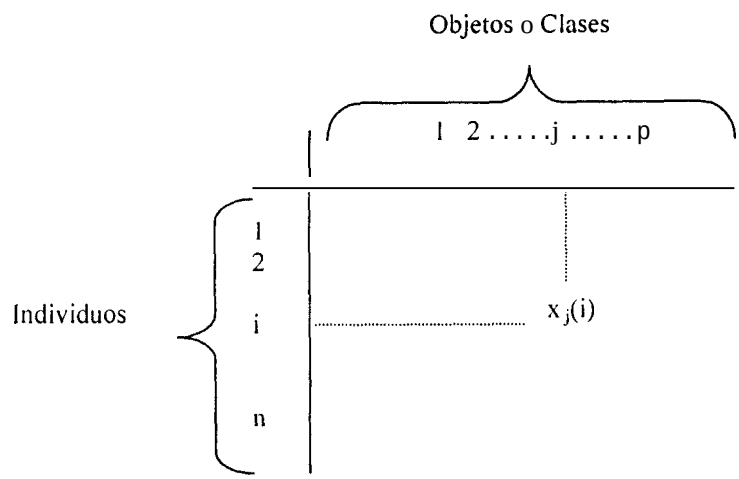
El termino $X_j(i)$ es igual a 1 o 0 segun si el individuo pertenece al grupo j (o si pertenece a la modalidad j o caracter cualitativo) o no, en cada línea un solo termino es igual a 1.

Una tabla disyuntiva completa esta formada por varias tablas lógicas. Donde cada tabla lógica corresponde a una particion del conjunto de individuos.

1.1.4 Tablas de Datos Ordinates

Las tablas de datos ordinales son frecuentemente utilizadas en técnicas de comercializacion. En efecto, los especialistas de estas tecnicas consideran frecuentemente que las respuestas de esta forma son un aporte de información mas coherente que una respuesta de datos por ejemplo sobre la forma de notas; estas son frecuentes en una encuesta.

Los datos se presentan de la siguiente forma



El termino $x_j(i)$ es la clasificacion del individuo i a la clase j .

Se tiene que:

$$\forall j \in J \qquad \forall j' \in J \qquad \forall i \in I \qquad j \neq j' \Rightarrow x_j(i) \neq x_{j'}(i)$$

$$1 \leq x_j(i) \leq p$$

La suma de todas las líneas es constante e igual a la suma de los p primeros numeros enteros (p(p+1)/2)

1.1.5 Tablas Mixtas

Se denomina tabla mixta de datos multidimensionales a las tablas en las que las variables son de diferente naturaleza, las cuales por lo general son una mezcla de variables cualitativas y cuantitativas. La tabla se obtiene por yuxtaposicion de una tabla de individuos por caracteres cuantitativos y de una tabla disyuntiva completa.

1.2 Vector Aleatorio

Un vector aleatorio se define como un arreglo de numeros reales $x_1, x_2, \ldots, x_n$, donde $x_1, x_2, \ldots\ldots, x_n$ son variables aleatorias, y este se escribe asi:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

1.2.1 Valor Esperado de un Vector Aleatorio

Sea $X=[x_1,x_2,\ldots,x_p]$ es un vector aleatorio. Entonces cada elemento de **X** es una variable aleatoria con su respectiva distribución de probabilidad. La media $\mu_i=E(X_i)$ y $\sigma_i^2=E(X_i-\mu_i)^2$ para i=1,2, ……, respectivamente.

1.3 Analisis de Factores

1.3.1 Introducción

El Analisis de Factores nacio a principios del siglo **XX** cuando investigadores como Karl Pearson, Charles Spearman entre otros buscaban un método para definir una medida de la inteligencia humana. Es por este interes en encontrar una medida de la Inteligencia que los primeros cientificos interesados en Analisis de Factores fueron Psicologos. En un principio el Analisis de Factores no podia ser utilizado debido a la gran cantidad de calculos que este requeria, pero con el desarrollo de las computadoras, el interes en el Analisis de Factores ha renacido especialmente en investigadores que buscan estudiar los aspectos teoricos y computacionales del modelo.

El proposito esencial del Analisis de Factores es describir si es posible la relación de covarianza entre muchas variables en terminos de unas pocas, pero no observables llamados factores. Es así que el modelo de Factores es motivado por el siguiente argumento:

"Las variables pueden ser agrupadas por sus correlaciones"

Es decir las variables pueden ser altamente correlacionadas en un grupo, pero estas tienen correlaciones pequeñas con las variables de otro grupo.

El Analisis de Factores puede ser visto como una extension del Analisis de Componentes Principales, pues ambos son vistos como un intento por aproximar la matriz de covarianzas $\sum$ aunque la aproximacion basada en Analisis de Factores es mas elaborada.

1.3.2 El Modelo de Factores Ortogonales

El vector aleatorio X, con p componentes, tiene media μ y matriz de covarianza Σ. El modelo de factores indica que X es linealmente dependiente de unas pocas variables

aleatorias no observables $F_1, F_2, \ldots\ldots, F_m$ llamados Factores Comunes, y p fuentes de variación $\varepsilon_1, \varepsilon_2, \ldots . \varepsilon_p$ llamadas error o algunas veces factores especificos.

El modelo de análisis de Factores se presenta asi:

$$x_1 - u_1 = \ell_{11}F_1 + \ell_{12}F_2 + \ldots\ldots\ldots + \ell_{1m}F_m + \varepsilon_1$$
$$x_2 - u_2 = \ell_{21}F_1 + \ell_{22}F_2 + \ldots\ldots\ldots + \ell_{2m}F_m + \varepsilon_2$$

$$x_p - u_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \ldots\ldots\ldots + \ell_{pm}F_m + \varepsilon_p$$

o matricialmente:

$$X - U = L \cdot F + \varepsilon$$

Donde

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \quad U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{bmatrix} \quad L = \begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \ell_{p1} & \ell_{p2} & \cdots & \ell_{pm} \end{bmatrix} \quad F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

ℓ_{ij} es el peso de la i-esima variable en el j-ésimo factor y L se denomina matriz de carga. Notese que el i-ésimo factor

especifico ε_i es asociado unicamente con la i-ésima respuesta x_i, las **p** desviaciones $X_1 - \mu_1$, $X_2 - \mu_2$, ..., $X_p - \mu_p$ son expresadas en terminos de p+m variables aleatorias F_1, F_2,....., F_m, ε_1, ε_2, ..., ε_p las cuales son no observables.

Además se tiene que F y ε son independientes, llegando a lo siguiente.

$E[F] = 0$ $\quad$ $Cov[F] = I = E[FF'] = I$

$E[\varepsilon] = 0$ $\quad$ $Cov[\varepsilon] = \Psi$ donde Ψ es matriz diagonal

$$\Psi = \begin{vmatrix} \psi_1 & 0 & . & . & 0 \\ 0 & \psi_2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \psi_p \end{vmatrix}$$

Si F y ε son independientes se tiene que

$Cov[\varepsilon,F] = E[\varepsilon,F'] = 0$

$$Cov[X] = E(X-u)(X-\mu)' = LL' + \Psi$$

$$Cov[X,F] = E(X-\mu)F' = L$$

Cabe anotar que el modelo $X-\mu = LF + \varepsilon$ es lineal en los factores comunes. Si las p respuestas X estan relacionadas por factores fundamentales, pero la relación no es lineal entonces las estructura de covarianza $LL' + \Psi$ puede no ser adecuada, pues uno de los mas importantes supuestos es el inherente a la linealidad.

El porcentaje de varianza con el que la i-esima variable contribuye a la varianza de los m factores comunes es llamado la i-esima comunalidad, mientras que la porción de $Var(X_i) = \sigma_{ii}$ debido al factor especifico se denomina varianza especifica.

$$\text{Var}(X_i) = \underbrace{\text{comunalidad}}_{} + \underbrace{\text{varianza especifica}}_{}$$

$$\sigma_{ii} = \ell^2_{i1} + \ell^2_{i2} + \ldots + \ell^2_{im} + \Psi_i$$

$$h_i^2 = \ell^2_{i1} + \ell^2_{i2} + \ldots\ldots + \ell^2_{im}$$

$$\sigma_{ii} = h_i' + \Psi_i \quad i = 1,2,\ldots\ldots p$$

El modelo de factores supone que las p(p+1)/2 varianzas y covarianzas para X pueden ser reproducidas a partir de los pm factores de carga ℓ_{ij} y las p varianzas especificas ψ_i. Cuando m=p, algunas matrices de covarianza Σ pueden ser reproducidas exactamente por L'L y la matriz Ψ puede ser la matriz 0.

De cualquier modo el Análisis de Factores es mas usado cuando el numero de factores (m) es relativamente pequeño con respecto al numero de variables (p), en este caso, el modelo de factores provee una explicación simple de la covarianza en X con pocos parametros con respecto a los p(p+1)/2 parametros en Σ.

Desafortunadamente muchas de las matrices de covarianza no pueden ser factoradas por $LL' + \Psi$, donde el numero de factores m es mucho menor que p. Estos problemas se presentan generalmente en el no cumplimiento de las condiciones de los parametros e_{ij} y ψ_i.

Cuando m>1, existe siempre alguna ambiguedad asociada con el modelo de factores. Para ver esto, sea T una matriz ortogonal $m_x m$, se tiene que $TT' = T'T = I$, entonces

$$X - \mu = LF + \varepsilon = LTT'F + \varepsilon = L^*F^* + \varepsilon$$

donde

$$L^* = LT \text{ y } F^* = TF$$

Y

$$E[F^*] = T'E[F] = 0$$

$$E[FF'] = Cov[F^*] = T'Cov[F]T = T'T = I$$

Es imposible, basados en las observaciones en X, distinguir los pesos L de los pesos L*. Esto es, los factores F y F*=T´F tienen las mismas propiedades estadisticas, y aunque los pesos L* son, por lo general, diferentes de los pesos L, ambos generan la misma matriz de covarianza $\sum$. Esto es

$$\sum = LL' + \Psi = LTT'L' + \Psi = (L^*)(L^*)' + \Psi$$

Esta ambiguedad nos da la pauta para la rotación de factores donde la matriz ortogonal T corresponde a las rotaciones del sistema coordenado X, que analizaremos mas adelante con mayor detalle.

I.3.3 Metodos de Estimación

Presentamos a continuación 2 de los mas populares metodos de estimacion de parametros considerados en el Análisis de Factores.

- El Metodo de Componentes Principales
- El Método de Maxima Verosimilitud

La estimacion por medio de estos metodos requiere calculos iterativos que gracias al desarrollo de las computadoras ha logrado que estos puedan ser realizados facilmente.

1.3.3.1 El Metodo de Componentes Principales

Se tiene que la matriz Σ tiene un par (λ_i, e_i) (valor propio, vector propio) donde $\lambda_1 >= \lambda_2 >= \ldots\ldots >= \lambda_p >= 0$.

Entonces

$$\Sigma = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \ldots\ldots + \lambda_p e_p e_p$$

$$=[\lambda_1^{1/2}e_1 \quad \lambda_2^{1/2}e_2 \quad \ldots\ldots \lambda_p^{1/2}e_p] \begin{bmatrix} \lambda_1^{1/2}e'_1 \\ \lambda_2^{1/2}e'_2 \\ \vdots \\ \lambda_p^{1/2}e'_p \end{bmatrix}$$

Esto describe la estructura de covarianza para un modelo de factores que tiene tantos factores como variables y las varianzas especificas $\psi_i=0 \ \forall i$, entonces las matriz de carga para la j-esima columna es obtenida por:

$$\sum = LL' + 0 = LL'$$

Aunque obviamente este modelo no es frecuentemente usado pues lo que se desea es explicar la estructura de covarianza en terminos de unos pocos factores. Es decir debido a que los p-m valores propios restantes son pequeños se pueden obtener la aproximacion siguiente:

$$\sum = \lambda_1 e_1 e_1 + \lambda_2 e_2 e_2 + \ldots\ldots + \lambda_m e_m e_m$$

$$=[\lambda_1^{1/2}e_1 \quad \lambda_2^{1/2}e_2 \quad \ldots\ldots \lambda_m^{1/2}e_m]\begin{bmatrix}\lambda_1^{1/2}e'_1\\ \lambda_2^{1/2}e'_2\\ \cdot\\ \lambda_m^{1/2}e'_m\end{bmatrix}=LL'$$

En este tipo de aproximacion asumimos que los factores especificos ε son de menor importancia y pueden ser ignorados en la factorización de Σ. **Si** incluimos los factores especificos en el modelo, sus varianzas se pueden obtener de los elementos de la diagonal de (Σ - LL').

Se tendría asi que la aproximacion seria la siguiente:

$$\Sigma=\begin{bmatrix}\lambda_1^{1/2}e_1 & \lambda_2^{1/2}e_2 & \ldots\ldots & \lambda_m^{1/2}e_m\end{bmatrix}\begin{bmatrix}\lambda_1^{1/2}e'_1\\ \lambda_2^{1/2}e'_2\\ \cdot\\ \lambda_m^{1/2}e'_m\end{bmatrix}+\begin{bmatrix}\psi_1 & 0 & . & 0\\ 0 & \psi_2 & & 0\\ . & . & . & .\\ 0 & . & . & \psi_p\end{bmatrix}=LL'+\Psi$$

Donde

$$\psi_i=\sigma_{ii}-\sum_{j=1}^{m}\ell_{ij}^2 \qquad i=1,2,\ldots,p$$

Cuando F_j y ε_j estan distribuidos normalmente , las observaciones $Xj - \mu = LF_j + \varepsilon_j$ tienen tambien una distribución normal.

$$L(\mu,\Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right) tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)'\right)\right]}$$

El estimador de maxima verosimilitud de $\hat{L}$ y $\hat{\Psi}$ se obtiene por la maximización numérica de la funcion de verosimilitud, es aqui donde los programas computacionales hacen que esta estimación sea considerablemente fácil.

Sea $X_1, X_2, \ldots\ldots, X_n$ una muestra aleatoria de $N_p(\mu,\Sigma)$ donde $\Sigma = LL' + \Psi$ es la matriz de covarianza para los m factores comunes del modelo.

Los estimadores de maxima verosimilitud $\hat{L}$, $\hat{\Psi}$ y $u = x$ maximizan la funcion de verosimilitud sujetos a que $L'Y^{-1}L$ sea diagonal.

Los estimadores de maxima verosimilitud de las comunalidades son:

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \ldots\ldots + \hat{\ell}_{im}^2 \qquad i = 1,2,\ldots\ldots p$$

Donde la proporción de varianza muestral total explicada por el j-esimo factor sera igual a:

$$\frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \ldots\ldots + \hat{\ell}_{pj}^2}{s_{11} + s_{22} + \ldots\ldots + s_{pp}}$$

Si las variables son estandarizadas esto es $Z=V^{-1/2}(X-\mu)$, entonces la matriz p de covarianza tiene la representación siguiente:

$$\rho = V^{-1/2} \sum V^{-1/2} = (V^{-1/2} L)(V^{-1/2} L)' + V^{-1/2} \Psi V^{-1/2}$$

Asi tenemos que ρ es una factorización analoga a la encontrada al principio donde $L_z=V^{-1/2}L$ y la matriz de varianza específica $Y_z= V^{-1/2}Y V^{-1/2}$.

Por la propiedad de invariancia de los estimadores de maxima verosimilitud, el estimador de maxima verosimilitud de ρ es:

$$\hat{\rho} = (\hat{Y}^{1/2}\hat{L})(\hat{V}^{1/2}\hat{L})' + \hat{V}^{1/2}\hat{\Psi}\hat{V}^{1/2} = \hat{L}_z\hat{L}'_z + \hat{\Psi}_z$$

Donde $V^{-1/2}$ (estimador) y L (estimador)

$$\hat{V}^{-\frac{1}{2}} \qquad y \qquad L$$

son los estimadores de maxima verosimilitud de $V^{-1/2}$ y L, respectivamente.

Como consecuencia de esta factorización, cuando el análisis de maxima verosimilitud pertenece a la matriz de correlación, se tiene que:

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \ldots\ldots + \hat{\ell}_{im}^2 \qquad i = 1,2,\ldots\ldots,p$$

Son los estimadores de maxima verosimilitud de las comunalidades, y podemos evaluar la importancia de los factores en base a la proporción de la varianza muestral total explicada por el j-ésimo factor que viene dada por:

$$\frac{\hat{\ell}_{1j} + \hat{\ell}_{2j} + \ldots + \hat{\ell}_{pj}}{p}$$

Es muy comun que las observaciones sean estandarizadas y la matriz de correlación sea analizada por los factores.

1.4 Selección dei Numero de Factores

En el estudio del Análisis de Factores siempre existe la pregunta de cuantos factores seran necesarios retener. Esta no es una pregunta con una contestación exactamente definida, pues existen muchos aspectos que hay que considerar tales como el monto de varianza explicada, los tamaños relativos de los valores propios (que es la varianza de los componentes), y la interpretación de los componentes.

1.4.1 Criterios para decidir cuantos componentes retener

Entre los metodos que son usados para determinar el numero apropiado de factores, tenemos los siguientes:

Criterio de Kaiser

Criterio del Codo

Criterio del Total de Varianza Explicada

Criterio del Numero Equivalente

1.4.1.1 Criterio de Kaiser (1966)

Uno de los mayormente usados. Consiste en la retención de los componentes cuyos valores propios son mayores que 1. Con esta regla generalmente se retienen los factores mas importantes, pero su uso puede retener factores que no tienen significancia practica. Estudios por Cattell y Jaspers (1967), Browne (1968) y Lim (1968) han evaluado la exactitud del criterio de Kaiser. En los 3 estudios los autores determinaron que en su mayoría este criterio puede identificar el numero correcto de factores de matrices con un conocido numero de factores.

El numero de variables en los estudios fue de 10 a 40. Generalmente el criterio fue de exacto a debilmente exacto.

El criterio resultó **mas** exacto cuando el numero de variables fue pequeño (de 10 a 15) o moderado (20 a 30) y las comunalidades eran altas (mayor a 0.70).

1.4.1.2 Criterio del Codo

Propuesto por Cattell (1966). En este método la magnitud de los valores propios (eje vertical) son graficados contra los numeros ordinales. Generalmente sucede que la magnitud de los sucesivos valores propios caen bruscamente y luego tienden a disminuir lentamente. La recomendacion es retener todos los valores propios que se encuentren antes del brusco declive.

Se han realizado varios estudios referentes a la precision del Criterio del Codo, entre ellos Tucker, Koopman y Linn (1969) encontrando que este encuentra el numero correcto de

factores en 12 de 18 casos. Linn (1968) encontro que esta regla proporcionaba el correcto numero de factores en 7 de 10 casos, mientras que Cattell y Jaspers (1967) encontraron que este proporcionaba el correcto numero de factores en 6 de 8 casos.

Un estudio mas reciente del problema del numero de factores (Hakstian, Rogers & Cattell, 1982) nos da algo de información adicional. Ellos dicen que para n>250 y para un promedio de comunalidad mayor a 0.60, ambos criterios, Kaiser y el Criterio del Codo, dan una estimacion precisa para el numero de factores.

Ellos agregan que una estimacion puede ser mucho mas creible si la razon Q/P es < 0.30 (P es el numero de variables y Q es el numero de factores)

Cuando la comunalidad media es 0.30 o Q/P>0.30, la regla de Kaiser es menos precisa siendo el Criterio del Codo mucho menos preciso.

1.4.1.3 Criterio del Monto de Varianza Explicada

Este metodo consiste en retener tantos factores como sean necesarios para explicar un monto de varianza total superior a nuestras expectativas. Este metodo puede llevar a la retención de factores los cuales estan conformados por la variable especifica.

1.4.1.4 Criterio del Número Equivalente (NEQ)

Entre estos criterios aparece una medida que quizás para el lector que no tiene mucha experiencia en el Análisis Factorial le signifique de gran ayuda, siendo este definido como una medida de la información no redundante aportada por un conjunto de variables.

Se tiene pues que el numero equivalente asociado a la matriz X, respecto a la metrica M, es $Neq(X,M)=\frac{(traza VM)^2}{traza(VM)^2}$, donde V es la matriz de covarianza de las p variables x^l.

1.5 Prueba de Hipotesis Referente al Numero de Factores

El supuesto de una poblacion normal nos permite realizar una prueba de lo adecuado del modelo. Supongamos que tenemos un modelo con m factores comunes. En este caso $\Sigma=LL'+\Psi$, entonces probar lo adecuado del modelo es equivalente a probar lo siguiente:

Ho: $\Sigma = LL' + \Psi$

vs.

Ha : Σ es cualquier otra matriz definida positiva.

En este caso Σ esta restringido por $\Sigma=LL'+\Psi$. Siendo la funcion de maxima verosimilitud proporcional a

$$\left|\hat{\Sigma}\right|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}tr\left[\hat{\Sigma}^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right)\right]\right)=$$

$$=\left|\hat{L}\hat{L}'+\hat{\Psi}\right|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}ntr\left[\left(\hat{L}\hat{L}+\Psi\right)^{-1}Sn\right]\right)$$

Se tiene que el radio estadistico de la prueba de razón de verosimilitud para probar Ho es

$$-2\ln\Lambda = -2\ln\left[\frac{maxim\ verosimilud\ bajo\ Ho}{maxim\ verosimilud}\right] = -2\ln\left(\frac{\left|\hat{\Sigma}\right|}{|Sn|}\right)^{-\frac{n}{2}} + n\left[tr\left(\hat{\Sigma}^{-1}Sn\right) - p\right]$$

Donde los grados de libertad vienen dados por

$$v - v_0 = \frac{1}{2}[(p-m)^2 - p - m$$

Se tiene que si

$$tr\left(\hat{\Sigma}^{-1}Sn\right) - p = 0 \qquad entonces \qquad \hat{\Sigma} = \hat{L}\hat{L}' + \hat{\Psi}$$

Es el estimador de maxima verosimilitud de $\hat{\Sigma} = \hat{L}\hat{L}' + \hat{\Psi}$. Se tiene pues que

$$-2\ln\Lambda = n\ln\left(\frac{\left|\hat{\Sigma}\right|}{|Sn|}\right)$$

Estudios de Bartlett indican que la aproximacion de χ^2 de la distribución muestral de $-2\ln\Lambda$ puede ser mejorada remplazando **n** con el factor multiplicativo **(n-1-(2p+4m+5)/6)**.

Usando el corrector de Bartlett se rechaza Ho con un nivel de significancia a si

$$(n-1-(2p+4m+5)/6)\ln\frac{\left|\hat{L}\hat{L}'+\hat{\Psi}\right|}{|Sn|} > \chi^2_{[(p-m)^2-p-m]/2}(\alpha)$$

Al usar el test de Bartlett, se esta probando lo adecuado del modelo de m factores comunes al comparar las varianzas generalizadas $\left|\hat{L}\,L'+\Psi\right|$ y $|Sn|$

Si **n** es grande y m es relativamente pequeño con respecto a **p**, la hipotesis **Ho** generalmente es rechazada, sugiriendose la retención de mas factores comunes.

1.6 Rotacion de Factores

1.6.1 Rotacion Ortogonal

Del algebra matricial, se sabe que una transformacion ortogonal corresponde a una rotacion rigida (o reflexion) de los ejes coordenados. Por esta razon, una transformacion ortogonal de las cargas de factores es llamada rotacion de factores.

Si $\hat{L}$ es una matriz de $p_x m$ de la estimaciones de las cargas de factores obtenidas por algun método entonces

$$\hat{L}^* = \hat{L}T \qquad donde \qquad TT' = T'T = I$$

es una matriz $p_x m$ de las cargas rotadas. Y, la matriz de covarianza estimada (o de correlación) permanece sin cambios, donde

$$\hat{L}\hat{L}' + \hat{\Psi} = \hat{L}TT'\hat{L} + \hat{\Psi} = \hat{L}^*\hat{L}^{*\prime} + \hat{\Psi}$$

Esto no indica que la matriz de residuos

$$S_n - \hat{L}\hat{L}' - \hat{\Psi} = S_n - L^* L^{*\prime} - \Psi$$

Permanece sin cambios. Por otro lado las varianzas especificas y las comunalidades también permanecen sin cambios. Es decir que desde el punto de vista matematico es irrelevante si se obtiene L o L^*.

Puesto que las cargas originales pueden no ser facilmente interpretables, se practica rotarlos hasta que se logre una estructura mas simple.

Idealmente, se pretende ver un patron de cargas tales que cada variable sea altamente cargada en un solo factor y tenga de pequeñas a moderadas cargas en los restantes factores.

Esto no iridica que la matriz de residuos

$$S_n - L\,L' - \Psi = S_n - L^*\,L^{*\prime} - \Psi$$

Permanece sin cambios. Por otro lado las varianzas especificas y las comunalidades también permanecen sin cambios. Es decir que desde el punto de vista matematico es irrelevante si se obtiene L o L^*.

Puesto que las cargas originales pueden no ser facilmente interpretables, se practica rotarlos hasta que se logre una estructura mas simple.

Idealmente, se pretende ver un patron de cargas tales que cada variable sea altamente cargada en un solo factor y tenga de pequetias a moderadas cargas en los restantes factores.

Nos concentraremos en los metodos graficos y analíticos para determinar una rotación ortogonal simple para una estructura. Cuando m=2 la transformación hacia una estructura simple puede ser frecuentemente determinada graficamente. Los factores comunes son considerados como vectores unitarios a lo largo del eje coordenado.

Los ejes coordenados pueden ser rotados visualmente en un angulo θ y las nuevas cargas rotadas ℓ_{ij}*. Son determinadas de las relaciones donde

$$T = \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix} \quad con \quad sentido \quad horario$$

$$T = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} \quad con \quad sentido \quad antihorario$$

Esta relación es rara vez usada en un análisis grafico en 2 dimensiones. En este caso, los grupos de variables son identificados por simple inspección visual.

Kaiser sugiere una medida analitica de una estructura simple conocida como criterio varimax.

Definimos

$$\tilde{e}_{ij}^{*} = e^{2}_{ij}{}^{*} / \tilde{h}_{i}$$

Como los coeficientes rotados sobre la raiz cuadrada de las comunalidades. Entonces el procedimiento Varimax selección la transformación T hace que

$$V = \frac{1}{p} \sum_{j=1}^{m} \left[\frac{\sum_{j=1}^{p} \tilde{e}^{*4}_{ij} - \left(\sum_{i=1}^{p} \tilde{e}^{*2}_{ij} \right)^{2}}{p} \right]$$

Sea lo mas grande posible.

Este procedimiento tiene el efecto de darnos variables con pequetias comunalidades relativamente mas pesadas en la determinación de la estructura simple. Despues que la

transformación T es determinada, las comunalidades son preservadas al multiplicar

$$(\tilde{e}^{*}_{ij})x(\hat{h}_i)$$

V tiene una simple interpretación que no es otra cosa que:

$$Va\sum_{j=1}^{m}(\text{varianza del cuadrado de los pesos para el } j-esimo\ factor)$$

Efectivamente, maximizar V corresponde a separar el cuadrado de los pesos en cada factor lo mayormente posible. Por lo tanto, se encontraran grupos de coeficientes altos y otros grupos de coeficientes insignificantes en algunas columnas de la matriz de carga rotada L^{*}

La rotación de las cargas de factores es recomendada particularmente cuando usamos el método de maxima verosimilitud, donde los valores iniciales son restringidos para satisfacer la condición de que la matriz

$$\hat{L}'\hat{\Psi}^{-1}\hat{L}$$

Esta condición es conveniente para propositos computacionales, pero esto hace que los factores no puedan ser interpretables facilmente.

6.2 Rotacion Oblicua

La Rotacion ortogonal es apropiada para un modelo de factores en el cual los factores comunes se asumen que son independientes.

Una rotacion oblicua para llegar a una estructura simple corresponde a una rotacion no rigida del sistema coordenado tal que los ejes rotados pasen a traves de los grupos. Una rotacion oblicua busca expresar cada variable en terminos de un mínimo numero de factores, preferiblemente un solo factor.

Capitulo 2

2. MODELOS DE MARKOV

2.1 Procesos Estocasticos

Un proceso estocastico es un sistema que se desarrolla en el tiempo mientras que se ve afectado por fluctuaciones al azar. Dicho proceso estocastico se puede describir definiendo una familia de variables aleatorias, $\{£_t\ t\}$, donde $£_t$ mide en el instante t, el aspecto del sistema bajo consideración.

Así pues los valores que puede tomar $£_t$ son llamados sus estados y los cambios en el valor de $£_t$ reciben el nombre de transiciones entre sus estados. Los modelos estocasticos son aplicables a cualquier sistema que comprenda variabilidad al azar con el transcurso del tiempo. En geofisica se han usado para la predicción y localización de los terremotos; en el area de la educación en la presente tesis se esta desarrollando un

modelo de predicción del numero de graduados en las carreras de Economia en Gestion Empresarial e Ingenieria Estadistica Informática.

Definición

Se define a los procesos estocasticos como una familia de variables aleatorias $\{£_t, \in T\}$.

Considere los puntos discretos en el tiempo $\{t_k\}$, para $k=1,2,3,$ ....., y sea $£_{tk}$ la variable aleatoria que caracteriza el estado del sistema en el instante t_k. La familia de variables aleatorias $\{f_{tk}\}$ forma un proceso estocastico. Los estados en el tiempo t_k representan realmente las situaciones (exhaustivas y mutuamente excluyentes) del sistema en el tiempo especifico. Se dice entonces que el numero de estados puede ser finito o infinito. Por ejemplo en el juego de lanzar la moneda, sean k lanzamientos, cada lanzamiento se puede interpretar como un punto en el tiempo. La secuencia resultante de lanzamientos constituye un proceso estocastico. El estado del sistema en cualquier lanzamiento es cara o sello.

En la carrera universitaria de un estudiante , para que este se gradue o egrese tiene que cumplir con un programa de estudio, es decir tiene que pasar k niveles. Donde la permanencia en un nivel puede interpretarse como un punto en el tiempo. La secuencia resultante de inscribirse en los diversos niveles constituye un proceso estocastico; donde el estado del sistema en cualquier curso es aprobado o no aprobado.

2.2 Procesos de Markov

Un proceso de Markov de primer orden es un proceso estocastico en el que la ocurrencia de un estado futuro depende del estado inmediatamente precedente y solo de **el.**

Definición

Sea $t_o<t_1<.....<t_n$ (n=0,1,2,3 .....) puntos en el tiempo, la familia de variables aleatorias $\{£_{tk}\}$ es un proceso de Markov, si esta posee la siguiente Propiedad Markoviana:

$$P\{£_{tn}=x_n/£_{tn-1}= x_{n-1},....., £_{t0}=x_0\}= P\{£_{tn}=x_n/£_{tn-1}= x_{n-1}\}$$

Para todos los valores posibles de $£_{t0},£_{t1},....,£_{tn}$

Un proceso de Markov queda representado por una funcion de probabilidad de transicion que se indica como:

$$P£_{n-1,}\ £_n = P\{£_{tn}=x_n/£_{tn-1}= x_{n-1}\}$$

Que representa la probabilidad condicional de que el sistema este en X_n en t_n, dado que estaba en X_{n-1} en t_{n-1}. Esta probabilidad tambien se denomina probabilidad de transicion de un paso, ya que describe al sistema entre t_{n-1} y t_n. Podemos entonces definir la probabilidad de transicion de m pasos como:

$$P£_{n,}\ £_{n+m}= P\{£_{tn+m}=x_{n+m}/£_{tn}= x_n\}$$

2.2.1 Cadenas de Markov

Definición

Consideremos una poblacion que es observada en un conjunto discreto de instantes. Si representamos las observaciones sucesivas mediante $£_o$, $£_1$, ….$£_n$,…… Se

supone que $£_n$ es una variable aleatoria, donde su valor representa el estado del sistema fisico en el instante n. Llamamos una cadena a la sucesion {fn} si existe un numero finito o numericamente infinito de estados en los que puede estar el sistema. La sucesion {fn} es una Cadena de Markov; si cada variable aleatoria f n es discreta y satisface que: para un entero cualquiera m>2 y un conjunto de puntos cualquiera $n_1<n_2<....<n_m$ la distribución condicional de f_{n_m} para valores dados de $f_{n_1},.....,£_{n_{m-1}}$ depende solo de $£_{n_{m-1}}$ que es el valor conocido mas recientemente, tenemos entonces que en general se cumple:

$$P[£_m = x_m / £_0 = x_0,.....,£_{m-1} = x_{m-1}] = P[£_m = x_m / £_{m-1} = x_{m-1}]$$

2.2.1.1 Probabilidades de Transición y la Ecuacion de Chapman Kolmogorov

Un proceso de Markov cuyo espacio de estados es discreto se denomina una Cadena de Markov, con frecuencia se usa el conjunto de los enteros positivos como el espacio de estados de una Cadena de Markov.

Sean $E_1, E_2, \ldots\ldots, Ej$ (j=0,1,2,3,....) los estados exhaustivos y mutuamente excluyentes de un sistema en un tiempo cualquiera. Inicialmente, en el tiempo t_o, el sistema puede estar en cualquiera de esos estados. Sean $a_j^{(0)}$ (j=0,1,2,3,....) las probabilidades absolutas de que el sistema se encuentre en el estado Ej en t_o. Suponga además que el sistema es Markoviano, definimos

$$P_{ij} = P\{£_{tn}=j/£_{tn-1}= i\}$$

Como la probabilidad de transicion de un paso, de pasar del estado i en t_{n-1} al estado j en t_n y suponemos que esas probabilidades de transicion del estado E_i al estado Ej se pueden arreglar mas comunmente en forma de matriz como sigue:

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \ldots\ldots \\ p_{10} & p_{11} & p_{12} & \ldots\ldots \\ p_{20} & p_{21} & p_{22} & \ldots\ldots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{pmatrix}$$

La matriz P se denomina matriz estocastica o matriz de transicion homogenea porque todas las probabilidades de transicion, son fijas e independientes del tiempo.

Las probabilidades P_{ij} deben satisfacer lo siguiente:

$$\sum_{j} P_{i,j} \qquad \forall i$$
$$P_{i,j} \geq 0 \qquad \forall i, j$$

Tenemos pues que una matriz P de transicion junto con las probabilidades iniciales $\{a_j^{(0)}\}$, asociados con los estados Ej definen completamente una cadena de Markov.

En una cadena de Markov por lo general el comportamiento de transicion es un intervalo de tiempo igualmente espaciado. Pero existen casos donde los espaciamientos temporales dependen de las caracteristicas del sistema y por ello, pueden no ser iguales, a este tipo de cadenas se les llama *Cadenas de Markov incrustadas.*

2.2.1.2 Probabilidades Absolutas y de Transición

Dadas $\{a_j^{(0)}\}$ y P una cadena de Markov, las probabilidades absolutas del sistema despues de un numero especifico de transacciones se determina como sigue.. Sean $\{a_j^{(n)}\}$ las probabilidades absolutas del sistema despues de n transacciones, o sea en t_n. La expresion general para $\{aj^{(n)}\}$ en terminos de $\{a_j^{(0)}\}$ y P, se puede encontrar como sigue:

$$a_j^{(1)} = a_1^{(0)} p_{1,j} + a_2^{(0)} p_{2,j} + \ldots\ldots = \sum_i a_i^{(0)} p_{i,j}$$

también

$$a_j^{(2)} = \sum_i a_i^1 p_{i,j} = \sum_i \left(\sum_k a_k^{(0)} P_{k,i} \right) P_{i,j} = \sum_k a_k^{(0)} P_{k,j}^{(2)}$$

donde

$$P_{k,j}^{(2)} = \sum_i P_{k,i} P_{i,j}$$

Es la probabilidad de transición de 2do. Orden o de 2 pasos, o sea, es la probabilidad de pasar del estado k al j en exactarnente 2 pasos.

Se puede llegar por inducción a la relacion:

$$a_j^{(n)} = \sum_i a_i^{(0)} \left(\sum_k P_{i,k}^{(n-1)} P_{k,j} \right) = \sum_i a_i^{(0)} P_{i,j}^{(n)}$$

Donde $P_{i,j}^{(n)}$ es la probabilidad de transición de n pasos dada por la siguiente formula recursiva:

$$P_{i,j}^{(n)} = \sum_k P_{i,k}^{(n-1)} P_{k,j} \qquad \forall i,j$$
$$P_{i,j}^{(n)} = \sum_k P_{i,k}^{(n-m)} P_{k,j}^{(m)} \qquad 0 < m < n$$

Que son las llamadas ecuaciones de **Chapman-Kolmogorov**

Los elementos de una matriz de transición de orden superior se pueden obtener en forma directa por multiplicación matricial. Asi tenemos que:

$$P_{ij}^{(2)} = P_{ij} \quad P_{ij} = P^2$$

$$P_{ij}^{(3)} = P_{ij}^2 \quad P_{ij} = P^3$$

.

.

$$P_{ij}^{n} = P^{n-1} P = P^n$$

Por lo tanto, si definimos las probabilidades absolutas en forma vectorial

$$a^{(n)} = \{a_1^{(n)}, a_2^{(n)}, a_3^{(n)}, ...\}$$

Tendremos que:

$$a^{(n)} = a^{(0)} P^n$$

Por consiguiente, la ley de probabilidad de una cadena de Markov se determina completamente una vez que se conoce la matriz de probabilidad de transicion de un paso $P=\{P_{ij}\}$ y el vector de probabilidad incondicional $a^{(0)}$.

Así, se dice que una cadena de Markov $\{\pounds_t\}$ es una cadena de Markov finita con k estados si el numero de valores posibles de las variables aleatorias $\{\pounds_t\}$ es finita e igual a **k.** Entonces las probabilidades de transicion $P_{i,j}$ son distintas de 0 solo para un numero finito de valores de **i** y j lo que hace que la matriz de probabilidades de transicion sea una matriz de $k_x k$.

Se puede probar aqui que las probabilidades absolutas a largo plazo son independientes de $a^{(0)}$, lo que indica que las

probabilidades resultantes se denominan probabilidades de estado estable.

2.2.1.3 Descomposicion de las Cadenas de Markov en Clases Comunicantes.

A continuación estudiamos la evolución en el tiempo de una Cadena de Markov con parametro discreto $\{\pounds_{tk}\}$. Empezamos clasificando los estados de esta segun sea posible ir de un estado dado a otro.

Se dice que un estado j es accesible desde un estado I si para un entero $n \geq 1$, $P_{i,j}^{(n)} > 0$.

Existe asi comunicacion entre dos estados i y j, si i es accesible desde j y j es accesible desde i.

Así si j es accesible desde i, se escribe i→j y si i y j se comunican, escribiremos $i \leftrightarrow j$

Teorema

Si $i \to j$ y $j \to k$ entonces $i \to k$

Teorema

La comunicacion es simétrica y transitiva en el sentido de que para estados cualquiera i j y k.

$$j \leftrightarrow k \text{ implica } k \leftrightarrow j$$

$$i \leftrightarrow j \text{ y } j \leftrightarrow k \text{ implica } i \leftrightarrow k$$

Clases Comunicantes

Dado un estado i de una cadena de Markov, se define su clase comunicante c(i) como el resultado de todos los estados k en la cadena que comunican con i, es decir:

$$k \in C(j) \text{ si y solo si } k \leftrightarrow j$$

Una clase C de estados no vacia en una Cadena de Markov es una clase comunicante si, para cierto estado i, C es igual a C(i).

Clase Comunicante Cerrada

Se dice que un conjunto de estados C no vacio es cerrado si desde un estado cualquiera interior al conjunto no es accesible a ningun estado exterior al conjunto. Es decir

C es cerrado si Y solo si

$\forall j \ en \ C \ y \ \forall k \notin C, \ P_{j,k}(n) = 0 \ \forall n = 1,2,.....$

Hay que tener en cuenta que una vez que una Cadena de Markov entra en una clase cerrada permanece en ella.

2.2.1.4 Clasificacion de los Estados en las Cadenas de Markov

Al realizar el análisis de Cadenas de Markov es interesante estudiar el comportamiento del sistema en un periodo corto, que es lo que se ha hecho hasta ahora, sin embargo es necesario conocer el comportamiento del sistema a largo plazo; es decir, cuando el numero de transiciones tienda a infinito. Presentamos a continuación algunas definiciones de la clasificacion de los estados en las cadenas de Markov que son utiles en el estudio del comportamiento de los sistemas a largo plazo.

Estado sin Retorno

Un estado se denomina sin retorno si C(i) esta vacia, es decir i no se comunica con ningun otro estado, ni siquiera consigo mismo. caso contrario dicho estado se denomina con retorno.

Cadena de Markov irreducible

Se dice que una cadena de Markov es irreducible si cada estado E_j se puede alcanzar desde cualquier otro estado E_i despues de un numero finito de transiciones.

$$P_{ij}^{(n)} > 0, \quad para \quad 1 \leq n < \infty$$

es decir, todos los estados de la cadena se comunican.

Estados Absorbentes v Estados de Conjunto Cerrado

Dada una cadena de Markov, un conjunto C de estados se denomina cerrado si el sistema, una vez en uno de los estados de C, permanece en C indefinidamente. Un ejemplo especial de un conjunto cerrado es un estado particular E, que tenga una probabilidad de transición $p_{ij}=1$, cuyo estado E_j se denomina estado absorbente. Todos los estados de una

cadena irreducible forman un conjunto cerrado y ningun otro subconjunto puede ser cerrado. El conjunto cerrado C tambien debe satisfacer todas las condiciones de una cadena de Markov y por ello, puede estudiarse en forma independiente.

Tiempos de primer retorno

Dado que el sistema esta inicialmente en el estado E_j, puede retornar a Ejpor primera vez en el paso n-esimo, con n>0. El numero de pasos antes de que el sistema retorne a Ej se llama tiempo de primer retorno.

Sea $f_{ij}^{(n)}$ la probabilidad de que el primer retorno a Ejocurra en el paso n-esimo. Entonces, dada la matriz de transición

$$P = P_{ij}$$

podemos determinar la expresion para $f_{ij}^{(n)}$ de la siguiente manera

$$P_{jj} = f_{jj}^{(1)}$$

$$P_{jj}^{(n)} = f_{jj}^{(n)} + \sum_{m=1}^{n-1} f_{jj}^{(m)} P_{jj}^{(n-m)}$$

donde

$$f_{jj}^{(n)} = P_{jj}^{(n)} - \sum_{m=1}^{n-1} f_{jj}^{(m)} P_{jj}^{(n-m)}$$

La probabilidad de por lo menos un retorno al estado Ej esta dada por

$$f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)}$$

Entonces, es seguro que el sistema retorna a j si f_{jj}=1. En este caso, si u_{jj} define el tiempo medio de retorno (recurrencia),

$$u_{jj} = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$$

Si f_{jj}<1, no es seguro que el sistema retornara a Ej y, en consecuencia, $u_{jj} = \infty$

Es así que podemos clasificar los estados de una Cadena de Markov en base a los tiempos de primer retorno como sigue:

Un estado es transitorio si $f_{jj}<1$, o sea, $u_{jj}=\infty$.

Un estado es recurrente (persistente) si $f_{jj}=1$

Un estado recurrente es nulo si $u_{jj}=\infty$ y no nulo si $u_{jj}<\infty$ (finito).

Un estado es periódico con periodo t si es posible un retorno solo en los pasos t, 2t, 3t, ... Esto significa que $P_{ij}^{(n)}=0$ siempre que *n* no sea divisible entre t.

Un estado recurrente es ergodico si es no nulo y aperiodico.

2.2.1.5 Cadenas Ergódicas de Markov

Una cadena de Markov irreducible es ergodica si todos sus estados son ergodicos. En este caso la distribución de probabilidad absoluta

$$a^{(n)}=a^{(0)}P^{n}$$

siempre converge univocamente a una distribucion limite cuando $n \to \infty$, donde la distribución limite es independiente de las probabilidades iniciales $a^{(0)}$.

Teorema

Todos los estados en una cadena de Markov infinita irreducible pueden pertenecer a una, y solo una, de las siguientes tres clases: estados transitorios, estados recurrentes nulos o estados recurrentes no nulos. En cada caso, todos los estados se comunican y tienen el mismo periodo. En el caso especial cuando la cadena tenga un numero finito de estados, la cadena no puede constar solo de estados transitorios, ni tampoco puede contener algun estado nulo.

Capitulo 3

3. ANALISIS DE FACTORES APLICADO AL RENDIMIENTO ESTUDIANTIL

En el presente capitulo analizaremos el comportamiento del rendimiento estudiantil, para tal motivo se han seleccionado estudiantes de las carreras de Ingenieria Estadistica Informatica y de la Carrera de Economia en los años de 1994 hasta el Segundo Termino de 1998.

3.1 Variables de Estudio

Las variables seleccionadas son las siguientes:

- Año de Ingreso
- Edad
- Sexo
- Estado Civil
- Lugar de Origen

- Carrera
- Factor Socioeconomico
- Materias Aprobadas
- Materias Reprobadas
- Rendimiento General
- Materia
  - Año Término
  - Promedio
  - Veztomada

3.2 Definición de Variables de Estudio

Año de Inareso

Expresa el año en el que ingreso el estudiante a cursar sus estudios de pregrado en la carrera de economia ESPOL, esta variable nos permite describir la cantidad de novatos que ingresan cada año, y queda definida así:

$$1994 \leq AÑOINGRESO \leq 1999; \qquad AÑOINGRESO \in N$$

Edad

La edad es la variable que mide el tiempo transcurrido (en años completos) desde el instante que el individuo nace hasta la presente fecha, la misma que esta definida asi:

$$0 \leq EDAD \leq 99; \quad EDAD \in N$$

Sexo

Esta variable nos indica a que sexo pertenece el individuo, esta variable es de tipo cualitativo y en el presente estudio se denotara con M al sexo masculino y F para el sexo femenino.

Estado Civil

Esta variable nos indica el estado civil del individuo, es de tipo cualitativo y consta de las siguientes categorias:

S = Soltero

C = Casado

V = Viudo

U = Union libre

D = Divorciado

Ciudad de Oriuen

Esta variable nos indica el lugar de donde proviene el estudiante, esta variable es de tipo cualitativo y consta de las siguientes categorias:

Guayaquil

Ciudades Satelites de Guayaquil

Los Rios

Manabi

El Oro

Esmeraldas

Region Sierra

Oriente

Extranjero

Carrera

Esta variable nos indica la carrera a la que pertenece el estudiante, es una variable de tipo cualitativo y consta de las siguientes categorias:

Economia y Gestion Empresarial

Ing. Estadistica Informática

Factor Socioeconómico[1]

Esta variable indica el nivel socio economico del estudiante, la que viene definida de la siguiente manera

[1] Ver Apendice A

$3 \leq FACTORP \leq 40 \qquad FACTORP \in N$

Esta variable puede agruparse en las tres categorias siguientes:

[3,12]; Nivel Socioeconomico Bajo

[13,26]; Nivel Socioeconomico Medio

[26,40]; Nivel Socioeconomico Alto

Materias Aprobadas

Esta variable nos indica el numero de materias aprobadas por el estudiante hasta el momento del estudio, dicha variable es de tipo cuantitativo y se define asi:

$$MAT\_APRO \geq 0, \qquad MAT\_APRO \in N$$

Materias Reprobadas

Esta variable nos indica el numero de materias reprobadas por el estudiante hasta el momento del presente estudio, dicha variable es de tipo cuantitativo y se define asi:

$$MAT\_REPRO \geq 0, \qquad MAT\_REPRO \in N$$

Rendimiento General

Esta variable mide el rendimiento en base al promedio de todas las materias tomadas por el estudiante, esta variable es de tipo cuantitativo, y se define asi:

$$0 \leq RENDIMIENTO \leq 10 \quad RENDIMIENTO \in \Re$$

Promedio Materia X

Promedio Materia **X,** es un conjunto de variables que miden el rendimiento estudiantil de un individuo en base al promedio obtenido en una materia dada, es de tipo cuantitativo y se define asi:

$$0 \leq PMATERIAX \leq 10 \quad PMATERIAX \in \Re$$

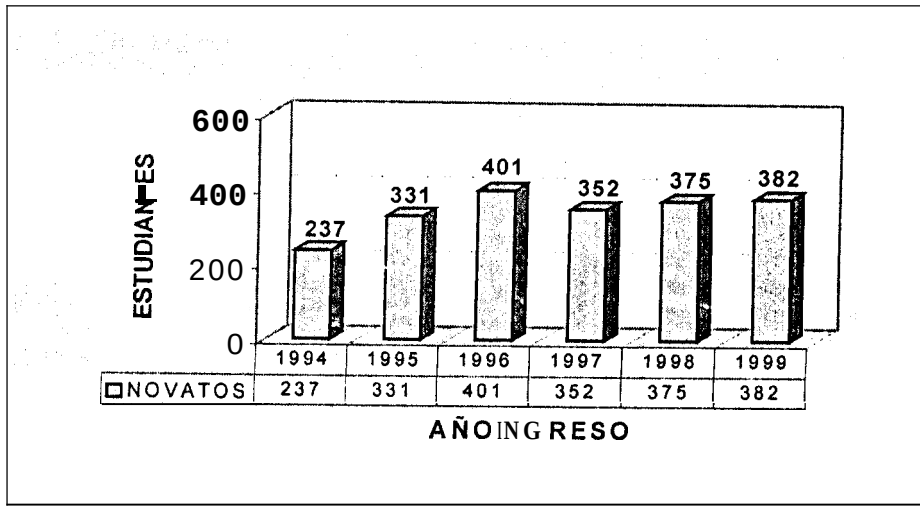
3.3 ANALISIS UNIVARIADO DE LAS VARIABLES OBJETOS DE ESTUDIO

En el analisis univariado desarrollaremos toda la estadistica descriptiva a cada una de las variables cuantitativas, analisis grafico de las variables cualitativas, y el posible ajuste de las variables a alguna distribución estadistica conocida.

3.3.1 Año de lngreso

FIGURA 3.1

NOVATOS ECONOMIA y GESTIÓN EMPRESARIAL

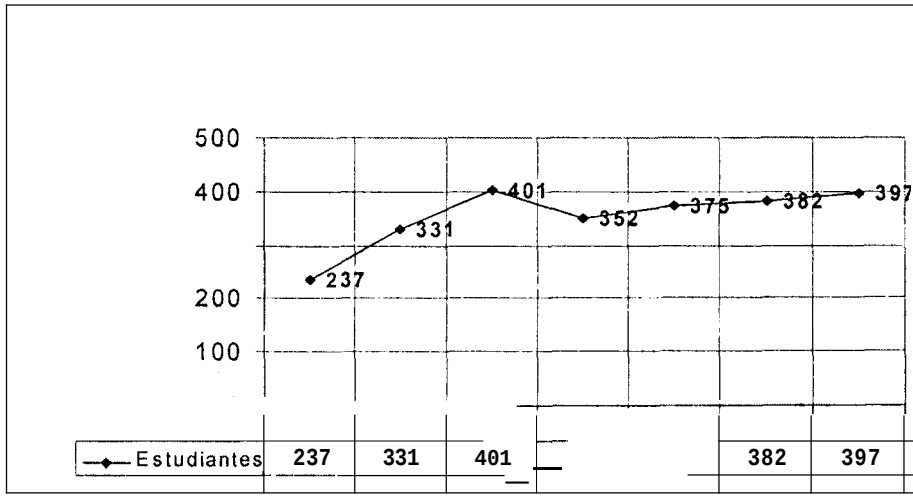


UNIDAD	ESTUDIANTES
NO. DE CASOS	6
MINIMO	237
MAXIMO	401
RANGO	164
SUMA	2078
MEDIA	363
MEDIANA	346
DESVIACION ESTANDAR	58,827
VARIANZA	3460,667
COEFICIENTE DE VARIACION	0,17
SESGOS	0,845
KURTOSIS	1,741

El promedio de nuevos ingresos a la carrera de Economia es de **346** estudiantes, con una tasa de crecimiento promedio del **11,36%** anual. La varianza es una medida que nos indica la dispersion de la variable, se tiene una desviacion estandar de 58,8. El coeficiente de variación es la desviacion estandar dividida para su muestra. El sesgo que mide el grado de simetria de la distribucion alrededor de la media nos dio 0,845, lo que nos indica que la distribucion es sesgada positivamente. La Kurtosis que mide el grado de picudez de la distribucion de la variables dio un valor de **1,741.**

FIGURA 3.2

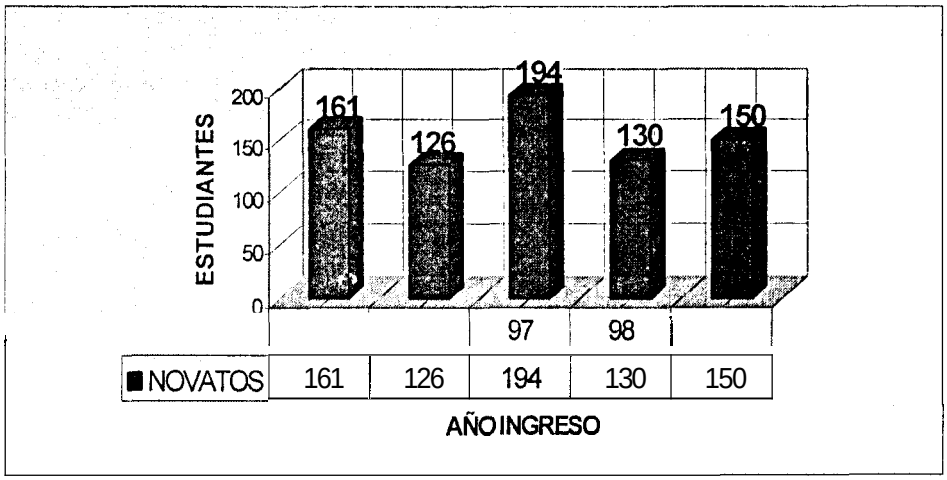
PREDICCION CRECIMIENTO NOVATOS ECONOMIA



Como vimos anteriormente la variable año de ingreso nos permite obtener el numero de novatos registrados en un termino dado; en este grafico se muestra el resultado de una regresion lineal simple efectuada para predecir el número de novatos en la carrera de economia para el año 2000, obteniendo un ajuste del 96,9 de confianza.

FIGURA 3.3

NOVATOS INGENIERÍA ESTADISTICA INFORMATICA



UNIDAD	ESTUDIANTES
NO. DE CASOS	5
MINIMO	126
MAXIMO	194
RANGO	68
SUMA	76
MEDIA	152,2
MEDIANA	150
DESVIACION ESTANDAR	27,42
VARIANZA	752,2

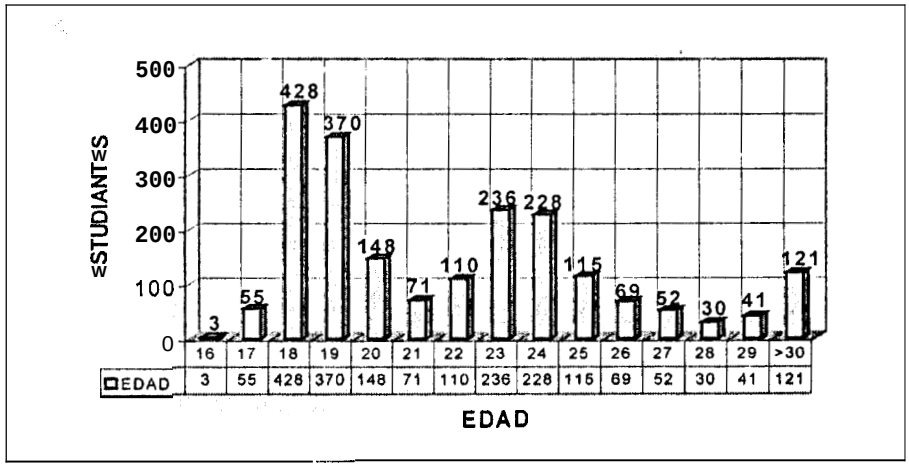
SESGOS	0,904
KURTOSIS	0,335

El promedio de ingreso de nuevos estudiantes a la carrera de Ingenieria Estadistica Informatica es de 152 estudiantes. La varianza es una medida que nos indica la dispersion de la variable, se tiene una desviacion estandar de 27,42. El coeficiente de variación es la desviacion estandar dividida para su muestra. El sesgo que mide el grado de simetria de la distribucion alrededor de la media nos dio 0,904, lo que nos indica que la distribucion es sesgada positivamente. La Kurtosis que mide el grado de picudez de la distribucion de la variables dio un valor de 0,335.

3.3.2 Edad

FIGURA 3.4

EDAD ESTUDIANTES ECONOMIA



UNIDAD	AÑOS
NO. DE CASOS	756
MINIMO	16
MAXIMO	52
RANGO	36
SUMA	16 676
MEDIA	22,058
MEDIANA	21
MODA	18
DESVIACION ESTANDAR	4,595
VARIANZA	21,117
COEFICIENTE DE VARIACION	0,208
SESGOS	0,089
KURTOSIS	0, 178

La edad promedio del estudiante de Economia es de 22 años con una desviacion estandar de 4,59, habiendo estudiantes desde los 16 años hasta los 52. La edad de moda en los

estudiantes es de 18 años. Existe un sesgo positivo y un grado de picudez de 0,178.

Efectuando la prueba de Kolmogorov-Smirnov, la cual nos permite definir hipotesis referente a la distribución estadistica de una variable estadistica continua, probamos la siguiente hipotesis referente a la distribución de la variable edad.

Ho: Edad ~ N(22,058;4,06)

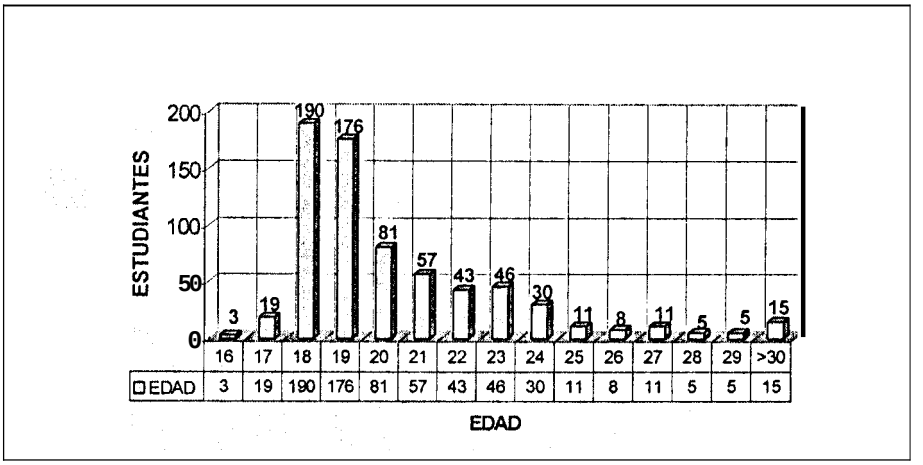
Ha: Edad ¬~N(22,058;4,06)

Rechazando la hipotesis nula con un valor p de 0,0001

FIGURA 3.5

EDAD ESTUDIANTES INGENIERÍA ESTADISTICA

INFORMATICA



UNIDAD	ANOS
NO. DE CASOS	258
MINIMO	16
MAXIMO	48
RANGO	32
SUMA	5 258
MEDIA	20,37
MEDIANA	19
DESVIACION ESTANDAR	3,39
VARIANZA	11,55
COEFICIENTE DE VARIACION	0,166
SESGOS	0,151
KURTOSIS	0,302

La edad promedio del estudiante de Ingeniería Estadistica Informática es de 20,3 años con una desviacion estandar de 3,39, habiendo estudiantes desde los 16 años hasta los 48.

La distribucion de la variable edad esta sesgada positivamente y tiene un grado de picudez de 0,302.

Efectuando la prueba de Kolmogorov-Smirnov, la cual nos permite definir hipotesis referente a la distribucion estadistica de una variable estadistica continua, probamos la siguiente hipotesis referente a la distribucion de la variable edad.

Ho: Edad ~ N(20,3;3,39)

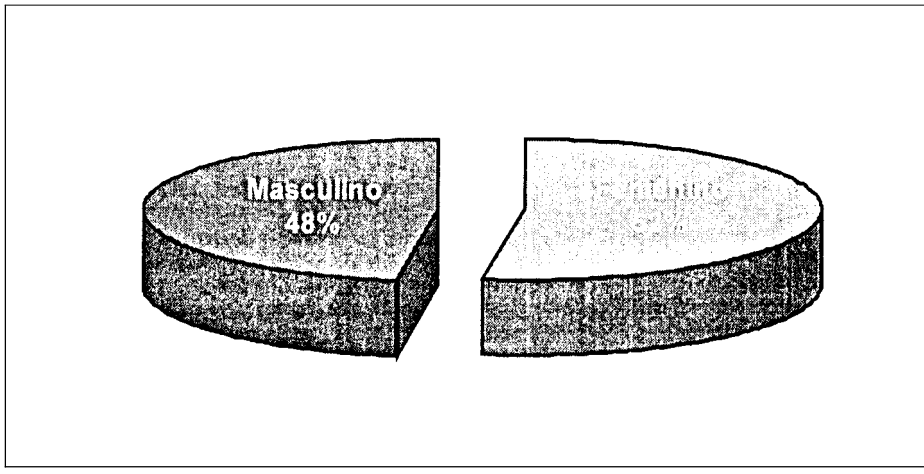
Ha: Edad ¬~N(20,3;3,39)

Rechazando la hipotesis nula con un valor p de 0,0001

3.3.3 Sexo

FIGURA 3.6

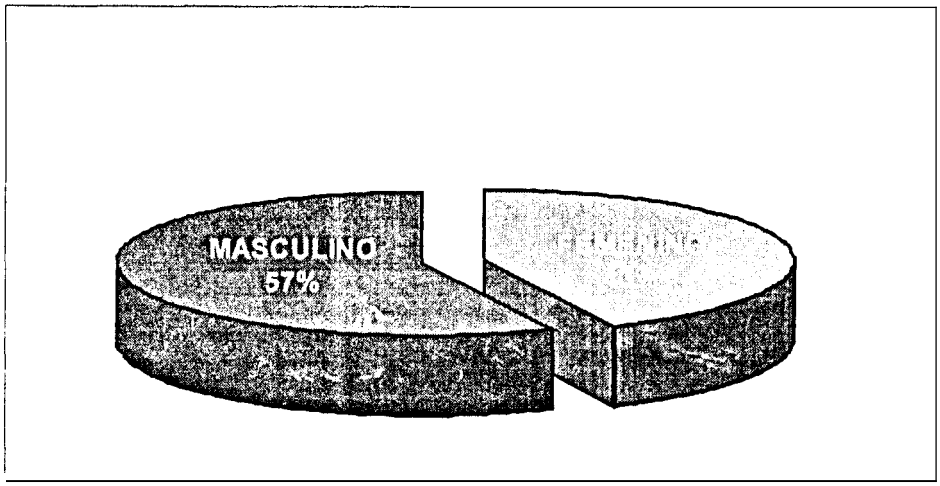
SEXO ESTUDIANTES ECONOMIA



El 52% de los estudiantes de economia pertenecen al sexo femenino, siendo una de las pocas carreras en donde el sexo femenino es mayoria.

FIGURA 3.7

SEXO ESTUDIANTES INGENIERÍA ESTADÍSTICA INFORMATICA

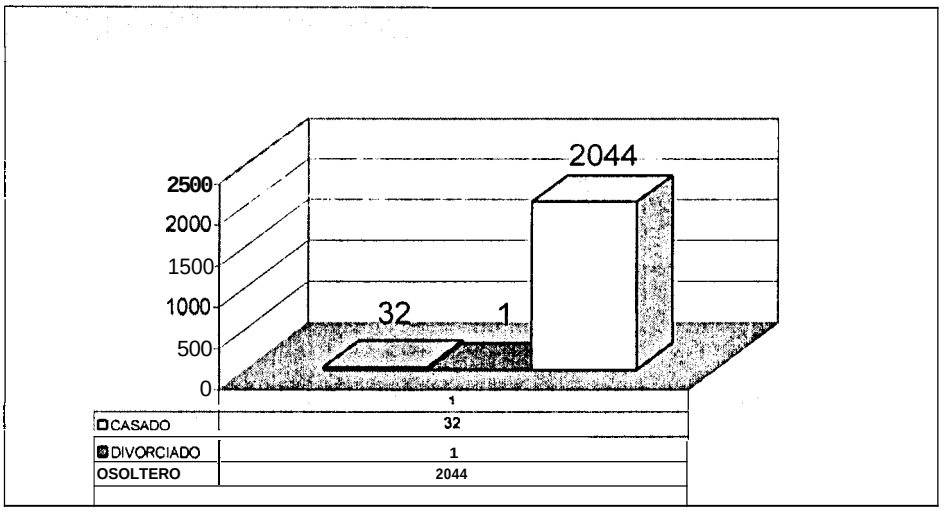


El 57% de los estudiantes de Ing. Estadistica Informática pertenecen al sexo masculino. Siguiendo la tendencia de las carreras tradicionales en la que el sexo masculino conforman la mayoria.

3.3.4 Estado Civil

FIGURA 3.8

ESTADO CIVIL ESTUDIANTES ECONOMIA



En lo referente al estado civil de los estudiantes de economia se tiene que el 1,.53% son de estado civil casado, un 0,05 son de estado civil divorciado y el 98,4% de los estudiantes son de estado civil soltero.

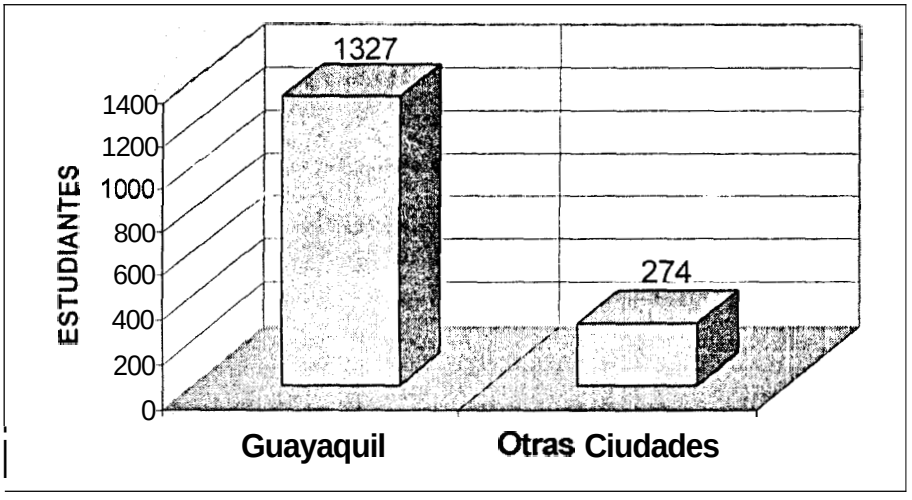
Con respecto a la carrera de Ingenieria en Estadistica e Informática el 99,8% de los estudiantes son de estado civil

soltero, mientras que el restante 0,02% son de estado civil casado (2 estudiantes).

3.3.5 Lugar de Origen

FIGURA 3.9

ORIGEN ESTUDIANTES ECONOMIA



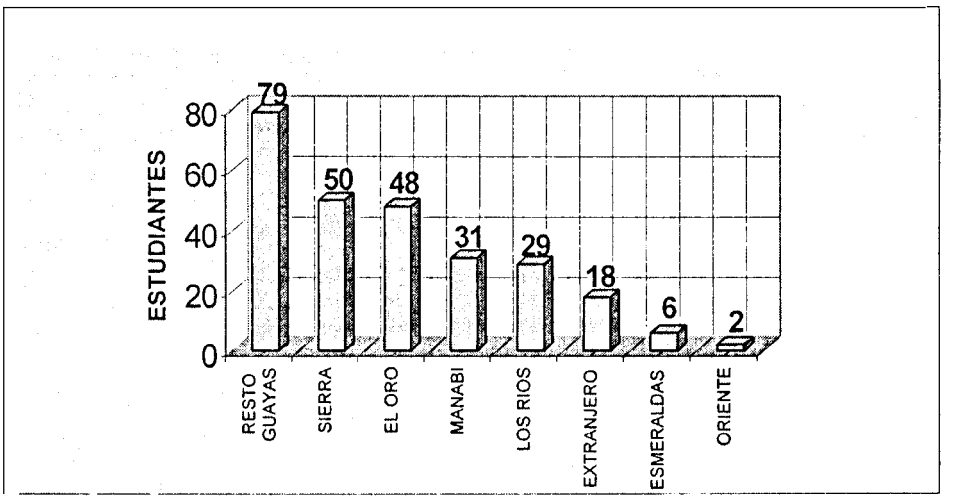
El 83% de los estudiantes de la Carrera de Economia provienen de la ciudad de Guayaquil, mientras que un 17% provienen de otras ciudades.

Para una mejor vision de la distribución del origen de los estudiantes se ha realizado la siguiente division:

- Los Rios
- El Oro
- Resto Guayas
- Manabi
- Esmeraldas
- Sierra
- Oriente
- Extranjero

FIGURA 3.9 (a)

ORIGEN ESTUDIANTES ECONOMIA

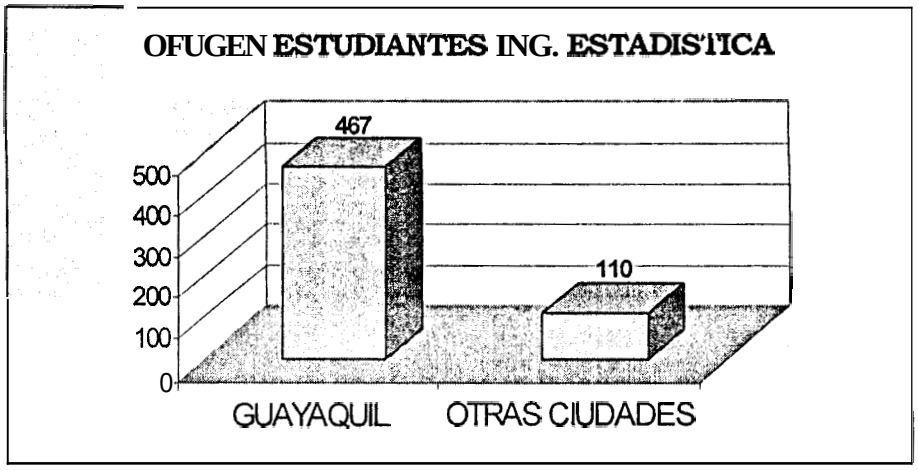


El 29% de los estudiantes cuyo origen no es Guayaquil, pertenecen al resto de ciudades de la Provincia del Guayas.

Seguidos por el 19% de estudiantes cuyo origen es la Region sierra.

FIGURA 3.10

ORIGEN ESTUDIANTES INGENIERÍA ESTADISTICA INFORMATICA

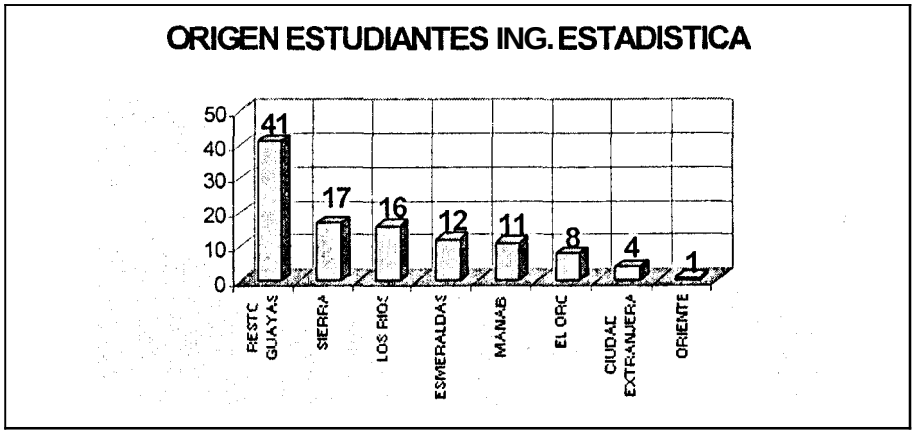


El 81% de los estudiantes de Ingeniería Estadistica Informática provienen de la ciudad de Guayaquil, mientras que un 19% provienen de otras ciudades.

FIGURA 3.10 (a)

ORIGEN ESTUDIANTES INGENIERÍA ESTADÍSTICA

INFORMATICA



El 37% de los estudiantes de Ingeniería Estadistica cuyo origen no es Guayaquil, pertenecen al resto de ciudades de la Provincia del Guayas. Seguidos por el 15% de estudiantes cuyo origen es la Región sierra.

3.3.6 Carrera

La carrera de Economia, consta de un cronograma de estudios en el cual los estudiantes hasta el nivel 200, avanzan juntos en sus estudios, tomando luego materias de

especializacion, habiendo existiendo una división natural de la carrera en 4 grupos que son:

Economia Nivel Basico

Especializacion Finanzas

Especializacion Marketing

Especializacion Sector Publico

Ingenieria en Estadistica Informatica se encuentra regido bajo un pénsum de estudio unico, no existiendo la división de la carrera en especialidades.

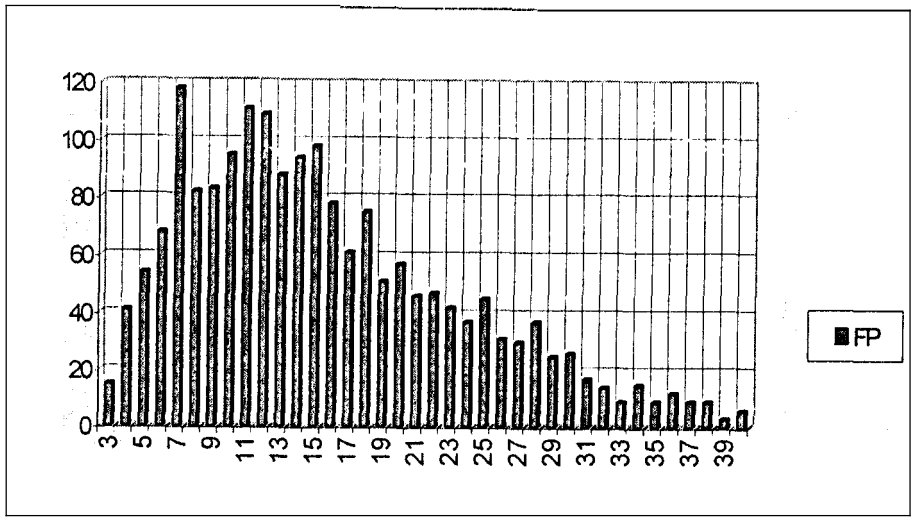
3.3.7 Factor Socioeconomico

Economia en Gestion Empresarial

UNIDAD	FACTOR SOCIOECONOMIC0
NO. DE CASOS	1 847
MINIMO	3
MAXIMO	40
RANGO	37
SUMA	2,89660E+04
MEDIA	15,682729
MEDIANA	14
MODA	12
DESVIACION ESTANDAR	8,0617
VARIANZA	64,992
COEFICIENTE DE VARIACION	0,514
SESGOS	0,056
KURTOSIS	0,113

FIGURA 3.11

FACTOR SOCIOECONOMIC0 ECONOMIA



La media del Factor Socioeconomico en los estudiantes de la carrera de Economía es de 15,68, con una desviacion estandar de 8,061. Siendo el sesgo 0,746 lo que nos indica un sesgo positivo de la distribucion del factor socioeconomico, y el grado de picudez de la misma fue de 0,113.

Lanzando la siguiente hipotesis referente a la distribucion de los datos

Ho: RENDIMIENTO GENERAL ~ N(15,68 ; 8,061)

vs.

Ha: ¬ Ho

Rechazando la hipotesis nula con un valor p de 0,00001.

Para un mejor de interpretación del Factor P se ha efectuado la siguiente agrupacion:

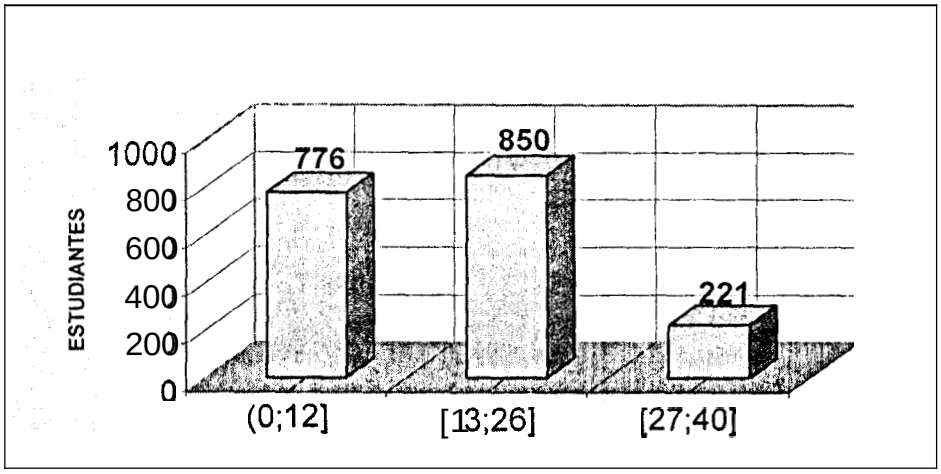
[3,12] Clase Baja

[13,26] Clase Media

[27, 40] Clase Alta

FIGURA 3.11 (a)

FACTOR SOCIOECONÓMICO ECONOMIA



La Carrera de Economia representa el 20,4% de la poblacion estudiantil en la ESPOL de los cuales el 11,9% de los estudiantes pertenecen a la clase economicamente alta, un 46,02% pertenece a la clase media y 42,01% pertenece a la clase baja. Del total de clase alta en la ESPOL, el 50% pertenecen a Economia.

Ingenieria Estadistica Informatica

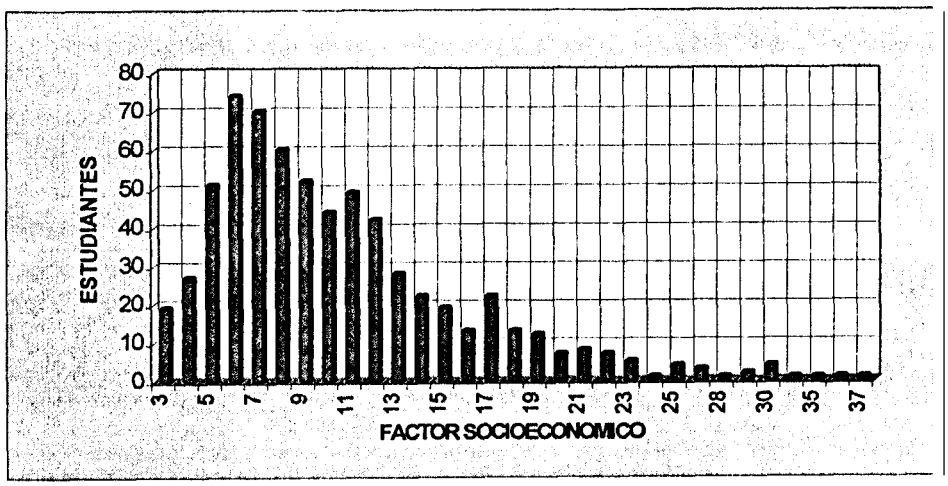
UNIDAD	FACTOR SOCIOE
NO. DE CASOS	663
MINIMO	3
MAXIMO	37
RANGO	34
SUMA	6 913
MEDIA	10,426
MEDIANA	9
MODA	6
DESVIACION ESTÁNDAR	5,557
VARIANZA	30,885
COEFICIENTE DE VARIACION	0,532
SESGOS	0,094
KURTOSIS	0,189

La media del Factor Socioeconomico en los estudiantes de la carrera de Ing. Estadistica Informatica es de 10,426, con una desviacion estandar de 5,557. Siendo el sesgo 0,094 lo que nos indica un sesgo positivo de la distribución de la variable correspondiente al factor socioeconomico, y el grado de picudez de la misma fue de 0,189.

FIGURA 3.12

FACTOR SOCIOECONÓMICO

INGENIERÍA ESTADISTICA INFORMATICA



Lanzando la siguiente hipotesis referente a la distribución de los datos

Ho: RENDIMIENTO GENERAL ~ N(10.42; 5.55)

vs.

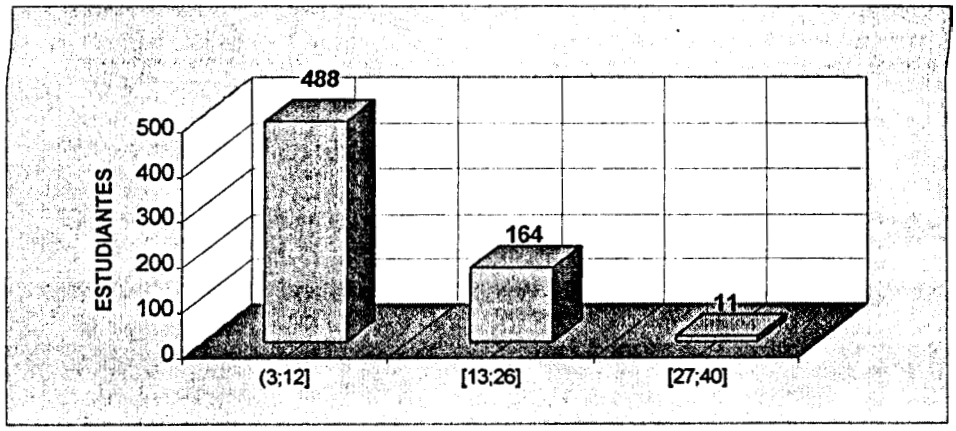
Ha: ⌉ Ho

Rechazando la hipotesis nula con valor p de 0.00001.

FIGURA 3.12 (a)

FACTOR SOCIOECONÓMICO

INGENIERÍA ESTADÍSTICA INFORMATICA



La Carrera de Ingeniería en Estadistica Informatica representa el 7,77 % de la población estudiantil en la ESPOL. Socioeconómicamente, la poblacion estudiantil se encuentra en un 73,6% en la clase economicamente baja, el 24,7% en la clase economicamente media, y un 1,6% pertenece a la clase económica alta.

3.3.8 Materias Aprobadas

Carrera de Economia en Gestion Empresarial

UNIDAD	MATERIAS APROBADAS
NO. DE CASOS	2 078
MINIMO	0
MAXIMO	55
RANGO	55

SUMA	4,177670E+04
MEDIA	20,10
MEDIANA	16
DESVIACION ESTÁNDAR	17,29
VARIANZA	299,09
COEFICIENTE DE VARIACION	0,860
SESGOS	0,053
KURTOSIS	0,107

El pensum de estudio de la Carrera de Economia consta de 52 materias, siendo este el requisito mínimo para poder egresar, con respecto a esta variable el promedio de materias aprobadas en la carrera de economia ***es*** de 20,10 con una desviación estandar de 17,29. Para una mejor descripción de esta variable, haremos un detalle del comportamiento de esta variable tomando en cuenta el año de ingreso a la ESPOL.

Así, tenemos que para los estudiantes cuyo ingreso a la ESPOL se dio en el año 1994; la media de materias aprobadas es de 34,64 con una desviacion estandar de 18,57, con un maximo de 55 materias aprobadas por estudiante **y** un mínimo de 0 materias aprobadas.

Para los estudiantes que ingresaron en 1995 la media de materias aprobadas es de 6,12 con una desviacibn estándar de 15,46, con un máximo de 55 materias aprobadas, y un minimo de 0 materias aprobadas.

Los estudiantes ingresados en 1996 la media de materias aprobadas es de 26,85 con un desviacion estándar de 11,57, con un máximo de 47 materias aprobadas, y un minimo de 0 materias aprobadas.

Estudiantes que ingresaron en 1997, media de materias aprobadas 18,07 y una desviacion estandar de 7,61, con un maximo de 35 materias aprobadas y un minimo de 0 materias aprobadas.

En el grupo de estudiantes que ingresaron a la ESPOL en 1998 la media de materias aprobadas fue de 9,69 con una desviacion estandar de 4,36, un maximo de 33 materias aprobadas y un minimo de 0 materias aprobadas.

Ingenieria Estadistica Informática

UNIDAD	MATERIAS APROBADAS
NO. DE CASOS	761
MINIMO	0
MAXIMO	48
RANGO	48
SUMA	9 377
MEDIA	12,32
MEDIANA	8
DESVIACION ESTÁNDAR	13,32
VARIANZA	177,52
COEFICIENTE DE VARIACION	1,081
SESGOS	0,088
KURTOSIS	0,177

Para los estudiantes que ingresaron en 1995 la media de materias aprobadas es de 27,11 con una desviacion estandar de 15,54, con un maximo de 48 materias aprobadas, y un minimo de 0 materias aprobadas.

Los estudiantes ingresados en 1996 la media de materias aprobadas es de 19,27 con un desviacion estandar de 11,16, con un maximo de 39 materias aprobadas, y un mínimo de 0 materias aprobadas.

Estudiantes que ingresaron en 1997, media de materias aprobadas 10,06 y una desviacion estandar de 5,804, con un

maximo de 25 materias aprobadas y un mínimo de 0 materias aprobadas.

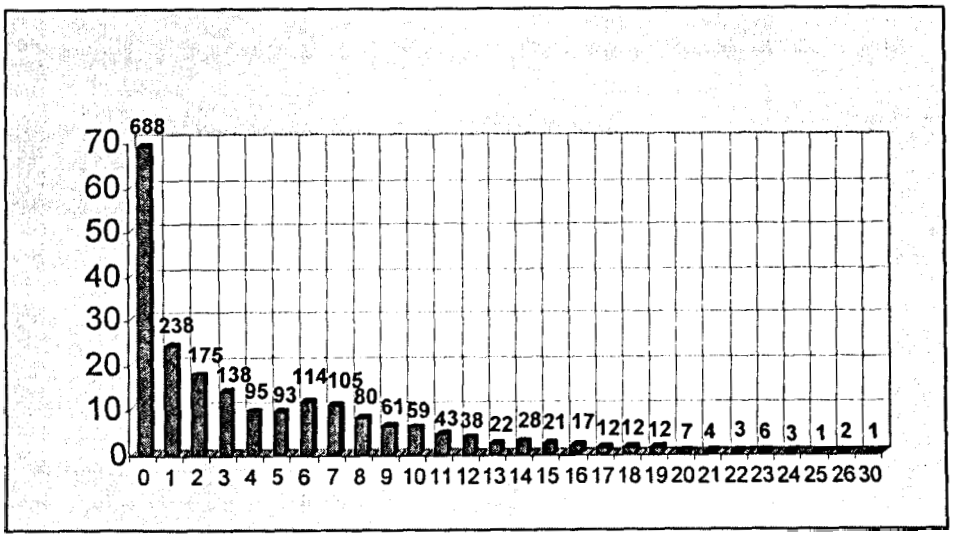
En el grupo de estudiantes que ingresaron a la ESPOL en 1998 la media de materias aprobadas fue de 4,82 con una desviacibn estándar de 3,74, un maximo de 16 materias aprobadas y un minimo de 0 materias aprobadas.

3.3.9 Materias Reprobadas

Carrera de Economia y Gestibn Empresarial

FIGURA 3.13

MATERIAS REPROBADAS ECONOMIA



UNIDAD	MATERIAS REPROBADAS

NO. DE CASOS	0
MINIMO	30
MAXIMO	30
RANGO	8 494
SUMA	4,08
MEDIA	2
MEDIANA	4,95
DESVIACION ESTÁNDAR	24,53
VARIANZA	1,211
COEFICIENTE DE VARIACION	0,053
SESGOS	0,107
KURTOSIS	

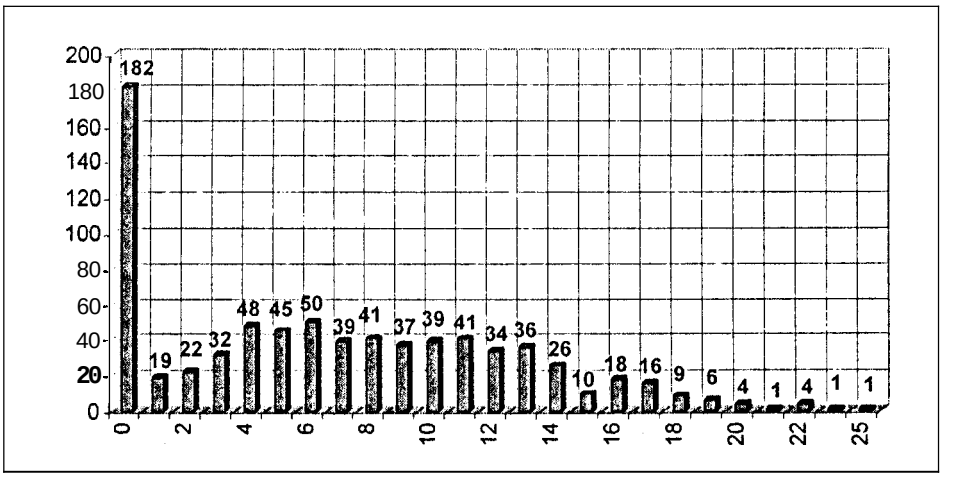
La media de materias reprobadas en los estudiantes de Economia es de 4,08 con una desviación estandar de 4,95, siendo el sesgo de la distribución de la variable 0,053 y la Kurtosis 0,107. Se tiene que el 66,89% de los estudiantes han reprobado por lo menos una materia a lo largo de su carrera estudiantil.

Ingeniería Estadistica Informática

FIGURA 3.14

MATERIAS REPROBADAS

INGENIERÍA ESTADISTICA INFORMATICA



UNIDAD	MATERIAS REPROBADAS
NO. DE CASOS	761
MINIMO	0
MAXIMO	25
RANGO	25
SUMA	5 114
MEDIA	6,72
MEDIANA	6
DESVIACION ESTANDAR	5,60
VARIANZA	31,428
COEFICIENTE DE VARIACION	0,834
SESGOS	0,088
KURTOSIS	0,177

La media de materias reprobadas en los estudiantes de Ingenieria Estadistica Informática es de 6,72 con una desviacion estandar de 5,60, siendo el sesgo de la distribución de la variable 0,088 y la Kurtosis 0,177. Se tiene

que el 76% de los estudiantes han reprobado por lo menos una materia a lo largo de su carrera estudiantil.

3.3.10 Rendimiento General

UNIDAD	PROMEDIO
NO. DE CASOS	1 466
MINIMO	600
MAXIMO	973
RANGO	373
SUMA	1,11368E+06
MEDIA	759,673
MEDIANA	752
DESVIACION ESTANDAR	61,463
VARIANZA	3 777,719
COEFICIENTE DE VARIACION	0,080
SESGOS	0,063
KURTOSIS	0,1277

El promedio de la carrera de Economia es de 7,596 con una desviacion estandar de 0,0614. El sesgo que mide el grado de simetria de la distribucion alrededor de la media es de 0,063, lo que significa que la distribucion esta ligeramente sesgada en sentido positivo, y la Kurtosis que mide el grado de picudez de la distribucion es de 0,1277.

Efectuando una prueba de Kolmogorov para probar la hipotesis referente a lo siguiente:

Ho: RENDIMIENTO GENERAL ~ N(7,596 ; 0,6056)

vs.

Ha: ¬ Ho

Se rechaza Ho con un valor p de 0,0001

UNIDAD	PROMEDIO
NO. DE CASOS	544
MINIMO	600
MAXIMO	940
RANGO	340
SUMA	3,82058E+05
MEDIA	702,31
MEDIANA	696
DESVIACION ESTANDAR	46,053
VARIANZA	2 120,95
COEFICIENTE DE VARIACION	0,065
SESGOS	0,104
KURTOSIS	0,2090

El promedio de Ingenieria en Estadistica Informática es de 7,0231 con una desviacion estandar de 0,046053. El sesgo que mide el grado de simetria de la distribucion alrededor de la media es de 0,104, lo que significa que la distribucion esta

ligeramente sesgada en sentido positivo, y la Kurtosis que mide el grado de picudez de la distribución es de 0,2090.

Efectuando una prueba de Kolmogorov para probar la hipotesis referente a lo siguiente:

Ho: RENDIMIENTO GENERAL ~ N(7,0231 ; 0,4888)

vs.

Ha: ¬ Ho

Se rechaza Ho con un valor **p** 0,025.

3.4 Análisis Bivariado

Para la realización del análisis multivariado hemos empleado el paquete estadistico **SPSS** 8.0.

3.4.1 Prueba de Medias

El estadistico a usar para probar nuestras hipotesis es el siguiente: $z = \frac{(x_1 - x_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$. El cual es un valor de una variable aleatoria que tiene la distribución normal estandar

Donde el criterio a usar para rechazar o aceptar la hipotesis nula se presenta segun el tipo de hipotesis. Tenemos pues para la prueba de medias las siguientes hipotesis alternas:

$\mu_1 - \mu_2 < \delta$
$\mu_1 - \mu_2 > \delta$
$\mu_1 - \mu_2 \neq \delta$

Donde los criterios para rechazar la hipotesis nula son respectivamente los siguientes:

$z < -z_\alpha$
$z > z_\alpha$
$z < -z_{\alpha/2}$ o $z > z_{\alpha/2}$

donde a es el nivel de significancia.

3.4.1.1 Aprovechamiento segun sexo

La hipotesis a plantear consiste en probar si la diferencia entre las medias del aprovechamiento del sexo femenino y masculino es significativa, probando la hipotesis de que el rendimiento en el sexo femenino es superior que el sexo masculino en la carrera de Ingenieria en Estadistica e Informática.

Por lo que nuestras hipotesis a probar seran

Hipotesis nula: $u_1 - u_2 = 0$

Hipotesis alterna: $u_1 - u_2 \neq 0$

Probaremos nuestra hipotesis con un nivel de significancia de 0.01. teniendo que rechazar nuestra hipotesis nula con un nivel de significancia de 0.01 y un valor $p<0.005$.

Realizando la prueba de medias en la Carrera de Economia se rechazo la hipotesis nula con un nivel de significancia de

0.01 y un valor $p<0.005$. Observandose que la presencia de altos valores en el sentido positivo en los estudiantes de sexo masculino. Es decir los alumnos que ocupan los primeros puestos en su mayoría pertenecen al sexo masculino.

3.4.1.2 Aprovechamiento segun nivel Socioeconomico

A continuación probaremos si existe diferencia estadisticamente significativa entre las medias del aprovechamiento del nivel socioeconomico bajo, con los niveles socioeconomicos medio y alto.

Para esto usaremos el Análisis de Varianza

Por lo que nuestras hipotesis a probar seran

Hipotesis nula: $u_1 = u_2 = u_3$

Hipotesis alterna: $u_1 \neq u_2 \neq u_3$

Donde u_1 es la media del rendimiento estudiantil de los estudiantes en Ingenieria Estadistica Informática cuyo nivel socioeconomico es bajo y u_2 es la media del rendimiento

estudiantil de los estudiantes en Ingeniería Estadistica Informática cuyo nivel socioeconomico es medio, de igual forma u_3 corresponde a la media del rendimiento estudiantil en los estudiantes cuyo nivel socioeconomico es alto.

TABLA I

PROMEDIO RENDIMIENTO ESTUDIANTIL SEGUN NIVEL SOCIOECONOMIC0

INGENIERÍA EN ESTADÍSTICA INFORMATICA

	MEDIA	N	Desviacion Estandar	Media Std. Error
Nivel Socioeconomico Bajo	7,0406	373	0,459	2,377E-0.2
Nivel Socioeconomico Medio	7,0461	148	0,4810	3,954E-0.2
Nivel Socioeconornico Altn	7,4167	9	0,5942	0,1981

Existiendo incertidumbre en la decision, al obtener un valor p=0,059. Es asi que en base a lo observado pasamos e lanzar la siguiente hipotesis.

Hipotesis nula: $u_1 - u_3 = 0$

Hipotesis alterna: $u_1 - u_3 \neq 0$

Rechazando la hipotesis nula con un valor p=0,0301. Es decir existe diferencia estadisticamente significativa entre el rendimiento estudiantil del nivel socioeconomico bajo con el nivel socioeconomico alto.

TABLA II

PROMEDIO RENDIMIENTO ESTUDIANTIL SEGUN NIVEL SOCIOECONOMICO

ECONOMIA Y GESTION EMPRESARIAL

	MEDIA	N	Desviacion Estandar	Media Std. Error
ivel Socioeconomico ajo	7'5351	568	0,6140	2,5766E-0.2
ivel Socioeconomico ledio	7'6352	738	0,5724	2,1072E-0.2
ivel Socioeconomico lto	7'6724	204	0,5705	3,9944E-0.2

Probaremos la siguiente hipotesis

Hipotesis nula: $u_1 = u_2 = u_3$

Hipotesis alterna: $u_1 \neq u_2 \neq u_3$

Donde u_1 es la media del rendimiento estudiantil de los estudiantes de Economia cuyo nivel socioeconomico es bajo

y u_2 es la media del rendimiento estudiantil de los estudiantes de Economia cuyo nivel socioeconomico es medio, de igual forma u_3 corresponde a la media del rendimiento estudiantil en los estudiantes cuyo nivel socioeconomico es alto.

Rechazando la hipotesis nula con un valor p=0,001, es decir existe diferencia estadisticamente significativa entre el promedio del rendimiento estudiantil entre los distintos niveles socioeconomicos.

3.5 Tablas de Contingencias

Una tabla de contingencia es el caso en que se extrae de una poblacion; y se clasifica cada unidad de acuerdo con dos categorias. La tabla de contingencia nos permitira probar la hipotesis nula de que las variables aleatorias representadas por dos clasificaciones son independientes. La hipotesis alternativa establece que los dos criterios de clasificación no son independientes. Donde rechazamos la hipotesis nula si el valor del estadistico excede $\chi^2_{\alpha,(r-1)(c-1)}$ donde a es el nivel

de significancia y (r-1)(c-1) son los grados de libertad (r = filas, c = columnas).

El estadistico viene dado por la siguiente formula $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij} - e_{ij})^2}{e_{ij}}$, donde o_{ij} es la frecuencia observada en la celda, y e_{ij} son las frecuencias esperadas en la celda.

TABLA III

ANÁLISIS CONTINGENCIA

RENDIMIENTO ESTUDIANTIL Y NIVEL SOCIOECONOMIC0

ECONOMIA

	CLASE ECONOMICA BAJA	CLASE ECONOMICA MEDIA	CLASE ECONOMICA ALTA
ACEPTABLE	11	7	1
SATISFACTORIO	86	55	18
BUENO	360	510	131
MUY BUENO	100	149	53
EXCELENTE	13	21	3

Confrontando estas dos variables lo que deseamos es probar la siguiente hipotesis:

Hipotesis nula: el rendimiento estudiantil en economia y el nivel socioeconomico son independientes.
Hipotesis Alterna: el rendimiento estudiantil en economia y el nivel socioeconomico no son independientes.

Probaremos nuestras hipotesis con un nivel de significancia α=0.01

El criterio a usar es el siguiente rechazamos la hipotesis nula si χ^2>20.090, efectuando los calculo obtuvimos que el valor del estadistico fue de 30.620, por lo que se rechaza la hipotesis nula; concluyendo que existe dependencia entre el rendimiento estudiantil y el nivel socioeconomico con un valor p<0.005.

Efectuando el mismo análisis para los estudiantes de la Carrera de Ingenieria en Estadistica Informatica se rechazo la hipotesis nula; es decir concluimos que existe dependencia entre el rendimiento estudiantil y el nivel socioeconomico con un valor $p<0.005$.

TABLA IV

ANÁLISIS CONTINGENCIA

RENDIMIENTO ESTUDIANTIL Y SEXO

ECONOMIA

	MASCULINO	FEMENINO
ACEPTABLE	9	9
SATISFACTORIO	87	48
BUENO	340	390
MUY BUENO	90	145
EXCELENTE	13	17

Confrontando estas dos variables lo que deseamos es probar la siguiente hipotesis

Hipotesis nula: el rendimiento estudiantil en economia y el sexo son independientes.

Hipotesis Alterna: el rendimiento estudiantil en economia y el sexo no son independientes.

El criterio a usar es el siguiente: rechazamos la hipotesis nula si $\chi^2>13.277$, efectuando los calculos obtuvimos que el valor del estadistico fue de **23.91,** por lo que se rechaza la hipotesis nula; concluyendo que existe dependencia entre el rendimiento estudiantil y el sexo con un valor $p<0.005$.

Efectuando el mismo análisis para los estudiantes de la Carrera de Ingenieria en Estadistica Informática se rechazo la hipotesis nula; es decir concluimos que existe dependencia entre el rendimiento estudiantil y sexo con un valor $p<0.005$.

TABLA V

ANÁLISIS CONTINGENCIA

RENDIMIENTO ESTUDIANTIL Y LUGAR DE ORIGEN

ECONOMIA

	GUAYAQUIL	FUERA DE GUAYAQUIL
ACEPTABLE	22	4
SATISFACTORIO	130	27
BUENO	645	118
MUYBUENO	206	40
EXCELENTE	28	3

Confrontando estas dos variables lo que deseamos es probar la siguiente hipotesis

Hipotesis nula: el rendimiento estudiantil en economia y la ciudad de origen son independientes.

Hipotesis Alterna: el rendimiento estudiantil en economia y la ciudad de origen no son independientes.

El criterio a usar es el siguiente: rechazamos la hipotesis nula si χ^2>13.277, efectuando los calculos obtuvimos que el valor del estadistico fue de 1.207, por lo que se acepta la hipotesis nula; concluyendo que el rendimiento estudiantil y el lugar de

origen son independientes con un nivel de significancia de 0.01 y un valor p>0.1.

Efectuando el mismo análisis para los estudiantes de la Carrera de Ingenieria en Estadistica Informatica no se rechazo la hipotesis nula; es decir concluimos que existe independencia entre el rendimiento estudiantil y la ciudad de origen con un valor p>0.1.

TABLA VI

ANÁLISIS CONTINGENCIA

NIVEL SOCIOECONOMIC0 Y LUGAR DE ORIGEN

ECONOMIA

	GUAYAQUIL	FUERA DE GUAYAQUIL
BAJO	478	175
MEDIO	567	63
ALTO	145	10

Confrontando estas dos variables lo que deseamos es probar la siguiente hipotesis

Hipotesis nula: el nivel socioeconomico en economia y la ciudad de origen son independientes.

Hipotesis Alterna: el nivel socioeconomico en economia y la ciudad de origen no son independientes.

Probaremos nuestras hipotesis con un nivel de significancia $\alpha=0.01$

El criterio a usar es el siguiente: rechazamos la hipotesis nula si $\chi^2>9.210$, efectuando los calculos obtuvimos que el valor del estadistico fue de 77.59, por lo que se rechaza la hipotesis nula; concluyendo que el nivel socioeconomico y el lugar de origen son dependientes con un valor $p<0.005$.

Efectuando el mismo análisis para los estudiantes de la Carrera de Ingenieria en Estadistica Informática se rechazo la hipotesis nula; es decir concluimos que existe dependencia entre el nivel socioeconomico y la ciudad de origen con un valor $p<0.005$.

3.6 Clasificacion Materias segun dificultad

Generalmente cuando se va a tomar un curso, el estudiante se hace la pregunta que tan difícil es el curso. Como un análisis muy útil previo al desarrollo del Modelo de Markov se han analizado las probabilidades de aprobacion en cada materia, en la Carrera de Economia en Gestion Empresarial y en Ingenieria Estadistica e Informática.

Así pues se ha dividido las materias según su dificultad en tres grupos que son:

Alta Dificultad

Mediana Dificultad

Baja Dificultad

La clasificacion se realizo de acuerdo a la probabilidad de aprobar el curso (Pa), donde Pa es igual a:

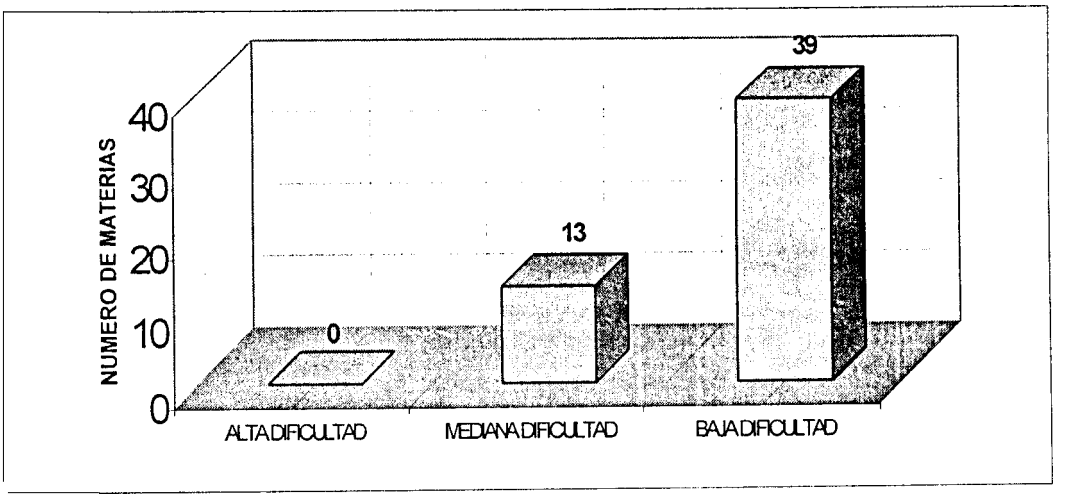
$$Pa = \frac{No.\ estudiantes\ que\ aprueba\ el\ curso}{Total\ estudiantes\ que\ toman\ el\ curso}$$

En la Carrera de Economia la dificultad de las materias se presenta así:

FIGURA 3.15

DIFICULTAD MATERIAS

ECONOMIA



Alta dificultad	$0 < P_a < 0.50$
Mediana Dificultad	$0.50 < P_a < 0.80$
Baja Dificultad	$0.80 < P_a < 1$

A continuación mostramos un cuadro en el que se indica la materia con su respectiva probabilidad de aprobacion.

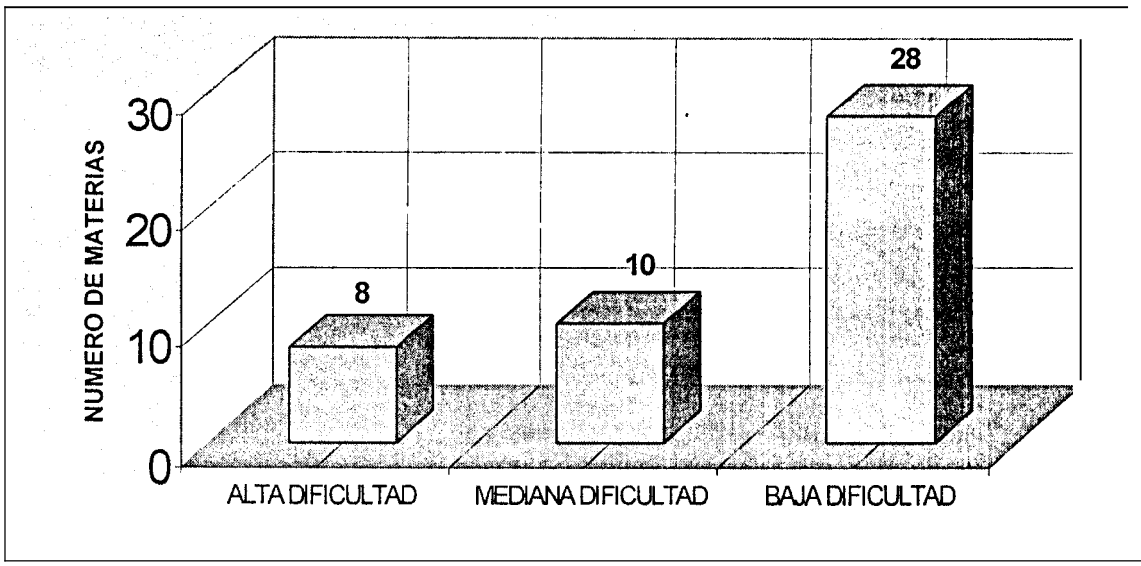
	MATERIA	PROBABILIDAD
MEDIANA DIFICULTAD	ALGEBRA LINEAL (1)	0.59
	MATEMATICAS FINANCIERAS (1)	0.59
	ESTADISTICA INFERENCIAL (1)	0.65
	ESTADISTICA DESCRIPTIVA (1)	0.65
	ECONOMETRIA II (1)	0.71
	ECONOMETRIA I (1)	0.73
	MICROECONOMIA I (1)	0.75
	MATEMATICAS I (1)	0.75
	MATEMATICAS II (1)	0.75
	MATEMATICAS III (1)	0.77
	CUENTAS NACIONALES (1)	0.77
	INGENIERIA ECONOMICA (1)	0.79
	ECONOMIA PUBLICA (1)	0.79
BAJA DIFICULTAD	MACROECONOMIA II (1)	0.83
	CONTABILIDAD FINANCIERA (1)	0.84
	ECON. MONETAR. ABIERTA (1)	0.84
	MICROECONOMIA II (1)	0.84
	ADMINISTRACION DE OPERACIONES (1)	0.85
	ECOLOGIA Y EDUC. AMBIENTAL (1)	0.85
	INTRODUCCION A LA ECONOMIA (1)	0.86
	SIMULACION Y MUESTREO (1)	0.87
	MACROECONOMIA I (1)	0.87
	FINANZAS I (1)	0.87
	ECONOMIA INTERNAC. (1)	0.88
	PRESUPUESTO (1)	0.89
	FINANZAS II (1)	0.90
	ETICA COMP. ORGANIZ. (1)	0.90
	FORM. EVAL. PROYECTO I (1)	0.90
	DESARROLLO ECON. SOSTEN. (1)	0.91
	MACROECON. DINAMICA (1)	0.91
	ECON. MERCADOS Y REGULAC. (1)	0.91
	ECONOMIA MATEMATICA (1)	0.92
	CONTABILIDAD DE COSTOS (1)	0.92
	HISTORIA Y FILOSOFIA (1)	0.93
	MARCO LEGAL EMPRESA (1)	0.93
	HISTORIA DEL PENSAM. ECONOMIC@(1)	0.93
	DERECHO I (1)	0.93
	FINANZAS WBUCAS (1)	0.94
	HISTORIA CIENCIAS POLITICAS (1)	0.94
	ADMINISTRACION (1)	0.94
	ADM. RECURSOS HUMANOS (1)	0.94
	DERECHO II (1)	0.95
	FORMUL. EVAL. PROYECTOS II (1)	0.95
	FUNDAMENTOS DE MERCADEO (1)	0.95
	POUT. ECON. TEOR. BINIEST. (1)	0.96
	FINANZAS INTERNACIONALES (1)	0.96
	TEORIA DEL ARTE (1)	0.96
	COMERC. INTERN. DER. ECONOM. (1)	0.96
	TEC. EXP. ORAL. (1)	0.96
	INVESTIGAC. MERCADO (1)	0.98
	ADMINISTRAC. PUBLICA (1)	1.00
	SOCIOECON. ANLS. ECONOMICO (1)	1.00

En la carrera de Ingenieria Estadistica Informatica la dificultad de las materias se presenta asi:

FIGURA 3.16

DIFICULTAD MATERIAS

ING. ESTADÍSTICA INFORMATICA



Alta dificultad	0 <P_a < 0.50
Mediana Dificultad	0.50 <Pa < 0.80
Baja Dificultad	0.80 <Pa < 1

A continuación mostramos un cuadro en el que se indica la materia con su respectiva probabilidad de aprobacion.

	MATERIA	PROBABILIDAD
ALTA DIFICULTAD	ESTRUCT. ALGEBRAICAS I	0.26
	CALCULO I	0.27
	ESTRUCT.ALGEBRAICAS It	0.29
	ESTADIST.MATEMAT.II	0.32
	ESTADIST.MATEMAT I	0.37
	FUNDAMENTOS COMPUTACION	0.43
	ANALIS.VARIAB REAL	0.48
	INVESTIG.OPERAC. II	0.49
MEDIANA DIFICULTAD	CALCULO II	0.51
	TRATAM.ESTAD.DATOS	0.51
	CALCULO III	0.57
	MATEMATICAS SUPERIORES	0.59
	INGENIW A DE LA CALIDAD	0.60
	MATEMATICAS ACTUARIALES	0.62
	ESTADISTICA COMPUTAC.	0.63
	ANALIS.ALGORIT.ESTR.DATOS	0.63
	METOD.NUMERICOS	0.64
	MUESTREO	0.65
BAJA DIFICULTAD	INVESTIGAC.OPERAC.I	0.69
	CONTABILIDAD GRAL.	0.70
	ANLS.MULTIVARIADO Y DISENO EXP	0.73
	ELEMENTOS FINITOS	0.76
	SIMULACION MATEMATICA	0.79
	ADM.EMPRESAS	0.79
	SISTEMAS EXPERTOS	0.85
	FORM.EVAL. PROYECTOS	0.86
	MATEMATICAS FINANCIERAS	0.86
	MACROECONOMIA	0.86
	ECOLOG.EDUCAC.AMB.	0.87
	ANLS.SERIES TIEMPO	0.88
	POLITICA EMPRESARIAL	0.89
	UTILIT.INFORMAT.II	0.90
	ARCH.Y BASES DATOS	0.90
	ADMINIST.REDES	0.92
	CONTAB.COSTOS	0.93
	ANALISIS FINANCIERO	0.94
	MICROECONOMIA	0.95
	TEC.EXP.ORAL Y ESCR.	0.96
	DESARROLLO APLICAC.COMPUTACION	0.97
	UTILITARIOS INFORMAT.I	0.98
	INGENIERIA DE SOFTWARE	1.00
	INV. DE MERCADO	1.00
	MARCO LEGAL EMPRESA	1.00
	MARKETING	1.00
	MONEDA Y BANCA	1.00
	ORGANIZACION Y METODOS	1.00

3.7 ANALISIS MULTIVARIADO

3.7.1 Análisis de Factores para las variables referentes al entorno personal y académico de los estudiantes de las Carreras de Ingeniería en Estadistica e Informatica y Economia en Gestion Empresarial

El objetivo de aplicar el análisis de factores a las variables de tipo personal y de rendimiento estudiantil es para encontrar dos o tres factores que expliquen la información contenida. En la tabla 3.10 presentamos la estadistica descriptiva de las variables de tipo personal y academicas. El método a utilizar para la extraccion es el de Componentes Principales

TABLA VII

ESTADÍSTICA DESCRIPTIVA VARIABLES PERSONALES Y ACADEMICAS

ESTUDIANTES INGENIERÍA ESTADISTICA INFORMATICA

	Media	Desviación Estandar	Datos Analizados	Datos Perdidos
Edad	20.27	3.34	243	272
Mat. Reprobadas	5.73	5.28	515	0
Mat. Aprobadas	10.19	11.23	515	0
F. Socioeconomico	9.52	6.02	514	1
Rend. Estudiantil	7.0531	0.4888	366	149

La tabla 3.10 muestra el porcentaje de explicación de cada variable. Mostrando en la primera columnas los valores caracteristicos por medio del cual obtenemos el porcentaje de explicación.

TABLA VIII

TOTAL DE VARIANZA EXPLICADA

ANÁLISIS DE FACTORES INGENIERÍA ESTADISTICA INFORMATICA

	Total	% de Varianza	% Acumulado
1	2.950	36.871	36.871
2	1.312	16.397	53.268
3	1.056	13.204	66.472
4	0.956	11.951	78.423
5	0.804	10.053	88.476
6	0.501	6.264	94.740
7	0.357	4.463	99.203
8	6.379E-02	0.797	100

El numero de factores a retener depende segun el criterio de retencion a usar, como dijimos en el capitulo 1 existen varios métodos para la retencion de factores. Si se escoge el criterio de los valores propios mayores que. 1 el numero de factores a retener sera de 3. Aunque podría extenderse bajo nuestro criterio a **4** factores pues el cuarto valor propio esta muy cerca de 1.

Usando el criterio del numero equivalente el numero de factores a retener es de **4.**

Con lo que se obtendria el 78.42% de explicación con el modelo de 4 factores.

TABLA IX

MATRIZ DE COMPONENTES

INGENIERÍA ESTADISTICA INFORMATICA

	COMPONENTES			
	1	2	3	4
SEXO	3.603E-02	-0.290	0.734	0.959
CIUDAD	-0.105	-0.368	-.569	4.305E-02
EDAD	0.736	-0.108	-8.026E-02	0.142
MAT. APROBADA	0.805	0.341	-8.832E-02	-4.072E-02
INGRESO	-0.950	-5.155E-02	9.293E-02	5.934E-02
REND. ESTUDIAN	-9.196E-02	0.864	-0.112	-7.940E-02
MAT. REPROBAD	0.802	-0.383	-6.298E-02	-6.519E-02
F. SOCIOECONOMIC0	0.439	0.260	0.393	0.277

En la tabla 3.11 consta la matriz de componentes en donde cada columna contiene los vectores propios correspondientes a cada componente, el metodo utilizado para la extracción de los factores es el de Componentes Principales. Donde se puede observar el primer componente el peso significativo de variables como son edad, numero de materias aprobadas y año de ingreso, en las componentes restantes la explicación de los componentes esta un poco dificil por lo que se procede a la rotacion, para poder darle una mejor interpretación a las componentes. Para lo que hemos optado por usar el metodo de rotacion Varimax.

TABLA X

MATRIZ DE COMPONENTES ROTADA

METODO ROTACION VARIMAX

INGENIERÍA ESTADISTICA INFORMATICA

	COMPONENTES			
	1	2	3	4
SEXO	-1.359E-02	-8.020E-02	7.568E-04	0.959
CIUDAD	2.972E-02	1.516E-02	0.940	4.305E-02
EDAD	0.757	-7.263E-02	0.145	0.142
MAT. APROBADA	0.837	0.263	-0.120	-4.072E-02
INGRESO	-0.945	7.366E-02	0.115	5.934E-02
REND. ESTUDIAN	1.871E-02	0.934	-4.191E-02	-7.940E-02
MAT. REPROBAD	0.744	-0.531	-6.803E-02	-6.519E-02
F. SOCIOECONÓMICO	0.398	0.161	-0.393	0.277

El primer factor puede ser llamado como “Experiencia Estudiantil”. Se llega a esta conclusion al observar la presencia de altos pesos en las variables edad, materias aprobadas, materias reprobadas e Año de Ingreso, aunque esta ultima tiene un peso negativo el cual es explicable pues mientras mas reciente es el ingreso menor serán, las tres variables restantes.

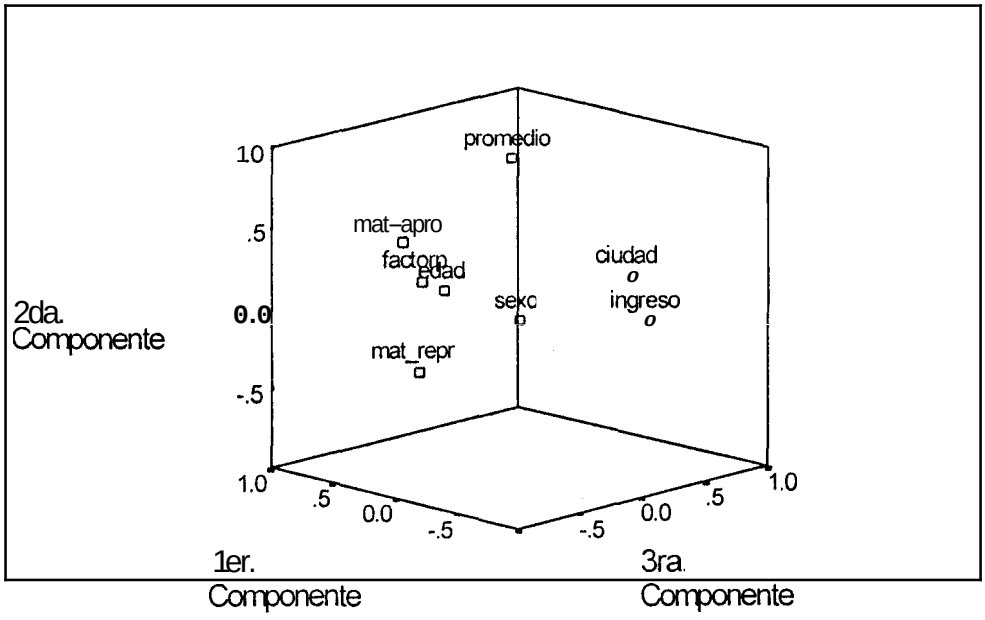
El segundo factor lo llamaremos “El rendimiento Estudiantil”, llegandose a esta conclusion al observar un peso alto en la variable rendimiento estudiantil. Recordando que la variable rendimiento estudiantil es el resultado del promedio de todas las materias cursadas por el estudiante hasta el momento de estudio.

El Tercer Factor se lo puede calificar como “Lugar de Origen” y el cuarto factor que es el sexo.’

FIGURA 3.17

COMPONENTES EN EL ESPACIO ROTADO

INGENIERÍA ESTADISTICA INFORMATICA



En el grafico se observa como las variables presentan su agrupacion y conforrnan las distintas zonas en el espacio, haciendo mas obvia la agrupacion de las variables en los factores.

TABLA XI

ESTADÍSTICA DESCRIPTIVA VARIABLES PERSONALES Y ACADEMICAS ESTUDIANTES ECONOMIA en GESTION EMPRESARIAL

	Media	Desviacion Estandar	Datos Analizados	Datos Perdidos
Edad	21.20	4.06	640	827
Mat. Reprobadas	3.56	4.80	1467	0
Mat. Aprobadas	16.63	15.53	1467	0
F. Socioeconómico	14.83	8.09	1466	1
Rend. Estudiantil	7.5966	.6056	1108	359

TABLA XII

TOTAL DE VARIANZA EXPLICADA

ANÁLISIS DE FACTORES ECONOMIA en GESTION EMPRESARIAL

	Valores Propios		
	Total	**% de Varianza**	**% Acumulado**
1	2.821	35.262	35.262
2	1,470	18.379	53.641
3	1.140	14.249	67.890
4	.967	12.085	79.975
5	,786	9.821	89.796
6	.452	5.648	. 95.445
7	.300	3.744	99.189
8	6.487E-02	.811	100.000

En la tabla 3.15 se presentan los valores propios, y la varianza explicada por cada uno de los factores.

Segundo Factor: Esta conformado por las variables, materias reprobadas y materias aprobadas, al que llamaremos "Avance", pues observamos en el la relación inversa existente entre el numero de materias aprobadas y el numero de materias reprobadas, concluyendose que a medida que aumentan las materias reprobadas, el numero de materias aprobadas disminuirá.

Tercer Factor: Lo llamaremos rendimiento estudiantil, y esta conformado por la variable Rendimiento Estudiantil.

Cuarto Factor: que es el factor socioeconomico.

Aplicando el metodo de retencion de Kaiser, sugiere la retencion de tres factores, con un 67.890% de explicación. El metodo del numero equivalente sugiere la retencion de cuatro factores, con un 79.975% de explicación.

TABLA XIII

MATRIZ DE COMPONENTES

ECONOMIA en GESTION EMPRESARIAL

	COMPONENTES			
	1	2	3	4
SEXO	.155	-.321	.639	.467
CIUDAD	.802	-5.500E-04	-9.302E-02	8.602E-02
EDAD	-.937	-.127	,129	3.035E-02
MAT. APROBADA	.683	-.576	-.114	-9.506E-02
INGRESO	.832	.416	-9.869E-02	1.745E-02
REND. ESTUDIAN	.307	.195	.667	.145
MAT. REPROBAD	-9.516E-02	.898	5.417E-02	.106
F. SOCIOECONOMIC0	-.116	-3.761E-02	-.486	.836

En la matriz de componentes podemos apreciar las cargas aportadas por cada una de las variables a los factores, de donde se puede observar las variables que tienen un impacto significativo en el componente. Sugiriendose que para una mejor apreciacion de la incidencia de las cargas en el factor, aplicar uno de los metodos de rotación, tales como: Varimax, Quartimax, u otros.

En la tabla 3.16 se observa la presencia de las componentes luego de haber aplicado el método de rotación Varimax.

TABLA XIV

MATRIZ DE COMPONENTES ROTADA

METODO ROTACION VARIMAX

ECONOMIA en GESTION EMPRESARIAL

	COMPONENTES			
	1	2	3	4
SEXO	-4.190E-02	-.223	.830	.121
CIUDAD	.784	-.174	.113	4.662E-02
EDAD	-.947	9.804E-02	-4.017E-02	6.034E-02
MAT. APROBADA	.537	-.726	4.907E-02	-4.749E-02
INGRESO	,909	.212	3.795E-02	-4.529E-02
REND. ESTUDIAN	.213	.207	.677	-.228
MAT. REPROBAD	.110	.904	-1.717E-02	7.253E-03
F. SOCIOECONÓMICO	-1.309E-02	3.668E-02	-4.296E-02	.973

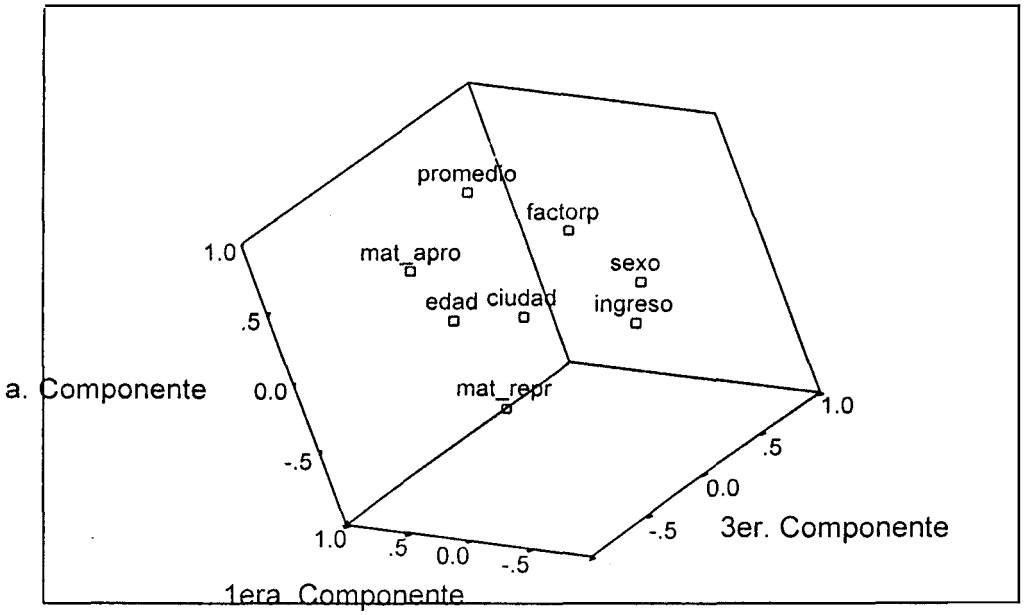
En esta matriz de componentes rotadas, se ha mejorado la visualización de la incidencia de las variables en los distintos factores, pudiendo notar los siguientes factores.

Primer Factor: Las variables, ciudad, materias aprobadas, y año de ingreso, conforman este factor, llamandole a este el factor de "Experiencia Estudiantil", pues este involucra el tiempo que lleva el estudiante en la carrera y las materias que lleva aprobadas.

FIGURA 3.18

COMPONENTES EN EL ESPACIO ROTADO

ECONOMIA EN GESTIÓN EMPRESARIAL



En el grafico se observa como las variables presentan su agrupacion y conforman las distintas zonas en el espacio, haciendo mas obvia la agrupacion de las variables en los factores.

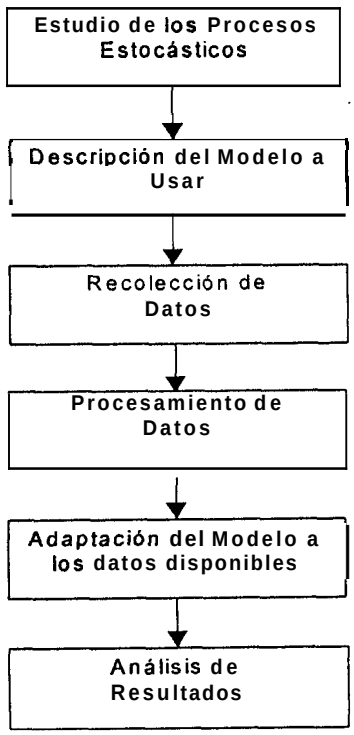
CAPÍTULO 4

4. ELABORACIÓN DE UN MODELO MATEMÁTICO CON CADENAS DE MARKOV QUE PRONOSTIQUE EL NUMERO DE GRADUADOS

4.1 Introducción

El objetivo del presente capitulo es el desarrollo de una Cadena de Markov como medio para predecir el numero de estudiantes graduados, para las Carreras de Economia en Gestion Empresarial e Ingenieria en Estadistica Informática. La teoria en que se basa el modelo de Markov fue ya expuesta en el Capitulo II.

Los pasos a seguir para el desarrollo del modelo son:



4.2 Descripcion del Modelo

El objetivo es crear un modelo de Markov para el sistema educacional de las carreras de Ingenieria Estadistica Informática y Economia que prediga el numero de graduados, haciendo enfasis que en este modelo lo basico es *el progreso de los estudiantes a* lo largo *de* sus años de

estudio, pudiendo utilizar las cadenas de Markov para modelar esto.

Consideraremos estudiantes graduados aquellos estudiantes que aprueben todas las materias correspondientes al pensum de estudio, es decir en realidad lo que el modelo predice son los estudiantes egresados caaa año.

Tenemos entonces que:

M sera el numero de niveles que el estudiante necesita aprobar para graduarse. Asumiendo que no hay deserciones.

P_{ij} (j = i, i+1, ..., m) sera la probabilidad que un individuo que estando en un estado i en un período sea encontrado en un estado j en el siguiente periodo. Por ejernplo P_{12} es la probabilidad de que un individuo que esta en el nivel 100 pase al nivel 200 en el año siguiente.

$C_i(n)$ sera el numero de estudiantes en el estado i despues de n periodos y $E[C_i(n)] = \gamma_i(n)$. Por ejemplo $\gamma_i(4)$ seria el valor

esperado del numero de estudiantes que se encuentran en el estado i luego de 4 años.

N_n sera el numero de nuevos participantes al sistema durante el n-esimo periodo, con $N_n\rho_i$ de estos, entrados al n-esimo estado siguiente, que cumplen las relaciones periodicas:

$$\gamma_i(n) = \sum_{r=1}^{m} \Pr i\gamma_r(n-1) + Nn\rho_i \qquad (1)$$

Q sera la matriz traspuesta de la matriz de probabilidades de transición (que es subestocastica) tal que:

$$Q = \begin{vmatrix} P_{11} & 0 & 0 & \ldots\ldots & 0 \\ P_{12} & P_{22} & 0 & \ldots\ldots & 0 \\ 4\ 3 & \blacksquare & P_{3m} & \ldots\ldots & 0 \\ \ldots\ldots & \ldots\ldots & \ldots\ldots\ldots & & \ldots\ldots \\ P_{1m} & P_{2m} & P_{3m} & & P_{mm} \end{vmatrix}$$

Si se escribe

$$\gamma_n = \begin{vmatrix} \gamma_1(n) \\ \gamma_2(n) \\ \gamma_3(n) \\ \vdots \\ \gamma_m(n) \end{vmatrix} \qquad P = \begin{vmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \\ \rho_m \end{vmatrix}$$

donde se puede escribir así:

$$\gamma_n = Q\gamma_{(n-1)} + N_n\rho$$

de donde:

$$\gamma_n = Q(Q\gamma_{(n-2)} + N_{(n-1)}\rho) + N_n\rho$$

y procediendo en forma recurrente se tiene:

$$\gamma_n = Q_{k+1}\gamma_{(n-k-1)} + \sum_{j=0}^{k} N_{(n-j)}Q^j\rho$$

y cuando k → ∝, Q → 0 debido a su naturaleza subestocastica. Otra forma de escribir lo mismo es:

$$\gamma_n = \sum_{j=0}^{\infty} N_{(n-j)} Q^j \rho \qquad (2)$$

Conociendo Q, ρ y N_r (r=0,1,2,3.....), la ecuación (2) puede ser usada para predecir el numero esperado de individuos en todos los estados despues de n periodos.

Para el sistema educacional en la ESPOL y en la mayoría de las organizaciones es razonable asumir que Pij=0 para i > j. Es decir el avance del estudiante en la carrera es en un solo sentido o caso contrario repetir el curso. Luego, la matriz Q seria:

$$Q = \begin{pmatrix} P_{11} & 0 & 0 & & 0 \\ P_{12} & P_{22} & 0 & ..... & 0 \\ 0 & P_{23} & P_{3m} & ..... & 0 \\ ..... & & & ..... & ..... \\ 0 & 0 & 0 & P_{(m-1)m} & P_{mm} \end{pmatrix}$$

4.3 Recoleccion y Procesamiento de Datos

Se recolectaron datos de la historia académica de los estudiantes que cursan la carrera desde sus inicios. Siendo esta la parte mas dificultosa del proceso, debido a el gran manejo de datos. En esta etapa se desarrollo una base de datos, la misma que esta conformada por toda la información correspondiente a lo académico y datos personales.

4.4 Adaptación del Modelo a los datos disponibles

En el caso de la ESPOL, existen ciertas situaciones que se deben tomar en cuenta para un mejor ajuste del modelo a la realidad.

Las distintas carreras de la ESPOL, tienen una division natural que es denominada niveles. Estos niveles optimos deberian ser aprobados en 1 año (2 semestres) tanto para la Carrera de Economia en Gestion Empresarial e Ingeniería en Estadistica Informática.

Carrera de Ingeniería en Estadistica Informática

- Nivel 100 de 0 a 9 materias aprobadas
- Nivel 200 de 10 a 21 materias aprobadas
- Nivel 300 de 22 a 34 materias aprobadas
- Nivel 400 de 35 a 45 materias aprobadas
- Nivel Salida mas de 45 materias aprobadas

Carrera de Economia y Gestión Empresarial

- Nivel 100 de 0 a 9 materias aprobadas
- Nivel 200 de 10 a 22 materias aprobadas
- Nivel 300 de 23 a 36 materias aprobadas
- Nivel 400 de 37 a 49 materias aprobadas
- Nivel Salida mas de 50 materias aprobadas

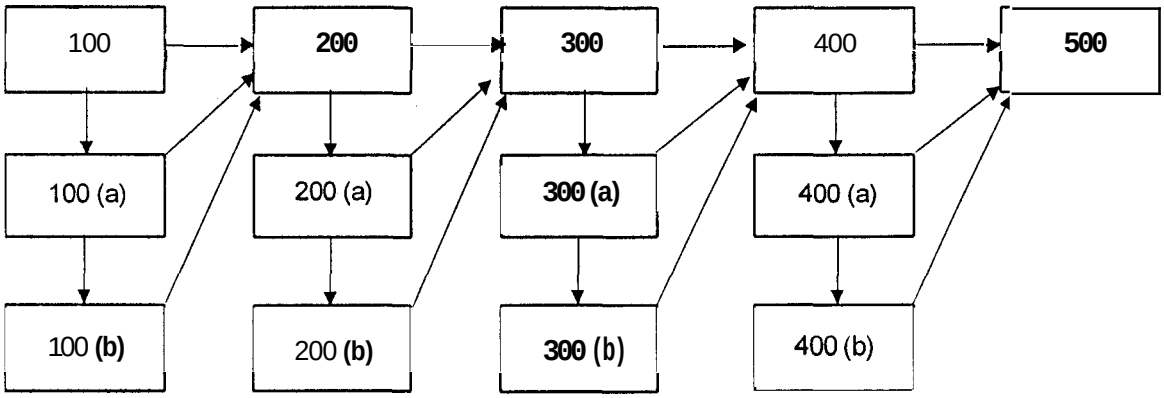
La formula con la que realizaremos la predicción es la determinada por la ecuacion (2), bajo las observaciones realizadas en esta parte. Destacando que en esta ecuacion se requiere de una matriz de probabilidad de transición,

disponiendo de la información de los cursos dictados desde la creación de las carreras.

4.5 Descripción Grafica del Modelo

FIGURA 4.1

DESCRIPCIÓN GRAFICA DEL MODELO



Se observa aqui los distintos estados (niveles), asi un estudiante al ingresar a la Carrera se ubica en el nivel 100 y este tiene dos opciones pasar al nivel 200 o quedarse otro año en el nivel 100, lo que en el modelo es considerado como pasar al estado 100 (a), si un estudiante pasa al nivel 100 (a) el estudiante tiene nuevamente 2 opciones pasar al nivel 200 o quedarse otro atio en el nivel 100 que en el

modelo es considerado 100 **(b).** Igual sucede con los otros niveles.

El nivel (c) que representaria el quedarse tres años en el mismo nivel, es considerado estadisticamente despreciable, por lo que en el presente modelo no es considerado.

TABLA XV

MATRIZ DE TRANSICION

ECONOMIA Y GESTION EMPRESARIAL

	1	1(a)	1(b)	2	2(a)	2(b)	3	3(a)	3(b)	4	4(a)	4(b)	5
1	0	0	0	0	0	0	0	0	0	0	0	0	0
1(a)	0.3182	0	0	0	0	0	0	0	0	0	0	0	0
1(b)	0	0.4834	0.8449	0	0	0	0	0	0	0	0	0	0
2	0.6818	0.5166	0.1551	0	0	0	0	0	0	0	0	0	0
2(a)	0	0	0	0.3544	0	0	0	0	0	0	0	0	0
2(b)	0	0	0	0	0.4083	0.7401	0	0	0	0	0	0	0
3	0	0	0	0.6456	0.5917	0.2599	0	0	0	0	0	0	0
3(a)	0	0	0	0	0	0	0.4317	0	0	0	0	0	0
3(b)	0	0	0	0	0	0	0	0.2588	0.6809	0	0	0	0
4	0	0	0	0	0	0	0.5683	0.7412	0.3191	0	0	0	0
4(a)	0	0	0	0	0	0	0	0	0	0	0	0	0
4(b)	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	1	0	0	1

Se observa aqui la matriz de transición transpuesta correspondiente a la Carrera de Economia en Gestión Empresarial; los rotulos de cada columna corresponden a los distintos estados.

TABLA XVI

MATRIZ DE TRANSICION

INGENIERÍA EN ESTADÍSTICA INFORMATICA

	1	1(a)	1(b)	2	2(a)	2(b)	3	3(a)	3(b)	4	4(a)	4(b)	5
1	0	0	0	0	0	0	0	0	0	0	0	0	0
1(a)	0.8207	0	0	0	0	0	0	0	0	0	0	0	0
1(b)	0	0.5548	0.8317	0	0	0	0	0	0	0	0	0	0
2	0.1793	0.4452	0.1683	0	0	0	0	0	0	0	0	0	0
2(a)	0	0	0	0.4417	0	0	0	0	0	0	0	0	0
2(b)	0	0	0	0	0.532	0.6935	0	0	0	0	0	0	0
3	0	0	0	0.5583	0.468	0.3065	0	0	0	0	0	0	0
3(a)	0	0	0	0	0	0	0.1982	0	0	0	0	0	0
3(b)	0	0	0	0	0	0	0	0.4231	0.77	0	0	0	0
4	0	0	0	0	0	0	0.8018	0.5769	0.23	0	0	0	1
4(a)	0	0	0	0	0	0	0	0	0	0	0	0	0
4(b)	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	1	1	1	0

De igual modo presentamos la matriz de transición transpuesta correspondiente a la Carrera de Ingeniería Estadistica Informática, en donde se observan las probabilidades de pasar de un estado i a un estado j.

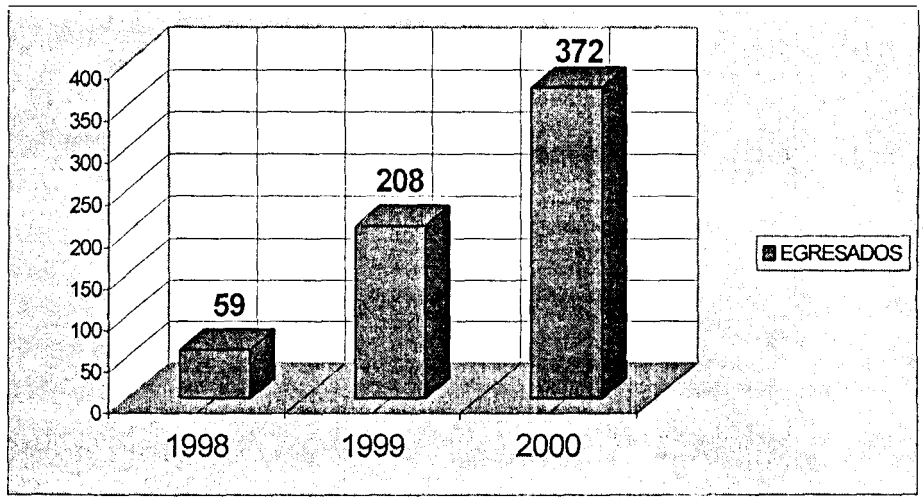
4.6 Resultados

Como resultado de la aplicacion de la formula (2), y usando las matrices de transición de las carreras de Economia y Ingeniería Estadistica Informática, así como el numero de novatos para los distintos años, se obtuvo los resultados para las carreras objeto de estudio.

4.6.1 Resultados Economia en Gestion Empresarial

FIGURA 4.2

PREDICCION EGRESADOS ESTUDIANTES ECONOMÍA Y GESTIÓN EMPRESARIAL



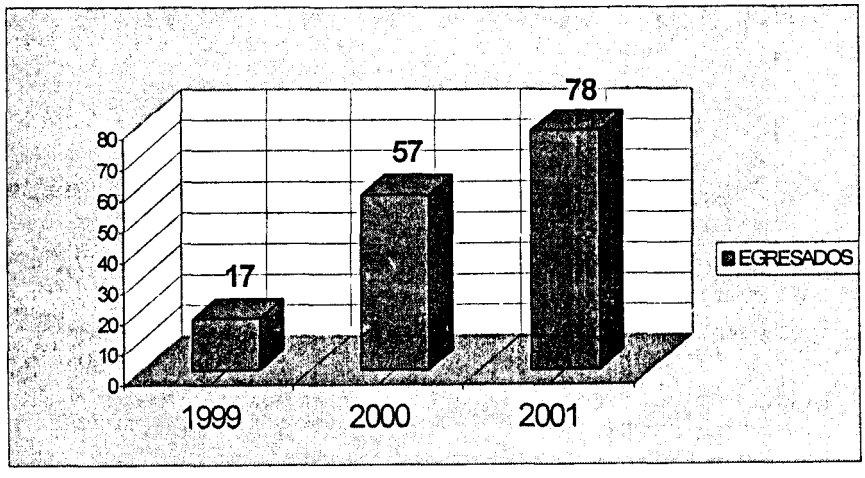
Como resultado de la aplicacion del Modelo de Markov planteado en los parrafos anteriores, se obtuvo la prediccion correspondiente al numero de Estudiantes graduados en la Carrera de Economia y Gestion Empresarial, que esta dado en el presente grafico. Resultando un valor de 59 estudiantes que terminaron su pensum de estudios para el año 1998, ya para el año 1999 se obtuvo una prediccion de 208 estudiantes. Cabe anotar que el numero de estudiantes egresados en las promociones 1999 y 2000 tiene un gran aporte por parte de los estudiantes que se han ido quedando rezagados, siendo esa una de las razon para que los egresados en estos años se incremente, considerandose que este crecimiento llegara a estabilizarse.

En la 1era. Promoción de Estudiantes de la Carrera de Economia en Gestion Empresarial el valor real de estudiantes que salieron del sistema fue de 64 estudiantes mientras que la prediccion del modelo fue de 59 estudiantes.

4.6.2 Resultados Ingenieria en Estadística Informatica

FIGURA 4.3

PREDICCION EGRESADOS ESTUDIANTES INGENIERÍA ESTADISTICA INFORMATICA



Como resultado de la Aplicacion del modelo se obtuvo una predicción de 17 alumnos egresados en el año 1999, 57 egresados para el año 2000 y 78 egresados para el año 2001. Igual que en la Carrera de Economia se observa que a partir de la segunda promoción el numero de estudiantes graduados, se ve afectado por estudiantes rezagados de las otras promociones.

Cabe mencionar que en la primera promoción de Ingenieria en Estadistica Informatica el numero de estudiantes que se graduaron fue de 29 estudiantes, mientras que para la 2da. Promoción el numero de estudiantes que se encuentran en proceso de graduación es de 56 estudiantes.

TABLA XVII

NUMERO DE ESTUDIANTES EN CADA ESTADO

ECONOMÍA EN GESTIÓN EMPRESARIAL

		AÑO						
		1994	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
	100	237	329	392	352	359	339	0
	100 (a)	0	75	105	125	112	114	108
N	100 (b)	0	0	36	81	129	163	193
	200	0	162	263	327	317	323	315
I	200 (a)	0	0	57	93	116	112	114
	200 (b)	0	0	0	23	55	88	111
V	300	0	0	104	204	272	288	298
	300 (a)	0	0	0	45	88	118	124
E	300 (b)	0	0	0	0	12	31	51
	400	0	0	0	59	149	224	260
L	400 (a)	0	0	0	0	0	0	0
	400 (b)	0	0	0	0	0	0	0
	500	0	0	0	0	59	209	373

En esta tabla se muestra el resultado del modelo en los distintos años, así tenemos que en el año 1994 existían 237 estudiantes en el nivel 100, mientras que ya en el año 1995 habia un total de 404 estudiantes en el nivel 100 de los cuales el 18.5% eran estudiantes rezagados que en la tabla se los puede identificar por hallarse en la celda correspondiente al nivel 100 (a), es decir estos estudiantes permanecieron otro año en el mismo nivel, en el mismo atio en el nivel 200 habian 162 estudiantes. De estos 162 estudiantes en el año 1996, 104 pasaron al nivel 300 y al final de 1997, 59 de estos se encontraban en nivel 400.

TABLA XVIII

NUMERO DE ESTUDIANTES EN CADA ESTADO

INGENIERÍA ESTADISTICA INFORMATICA

		AÑO						
		1995	1996	1997	1998	1999	2000	2001
	100	213	163	202	127	150	0	0
	100 (a)	0	175	134	166	104	123	0
N	100 (b)	0	0	97	155	221	241	269
	200	0	38	107	112	123	110	95
I	200 (a)	0	0	17	47	50	54	49
	200 (b)	0	0	0	9	31	48	62
V	300	0	0	21	68	57	101	102
	300 (a)	0	0	0	4	13	17	20
E	300 (b)	0	0	0	0	2	7	13
	400	0	0	0	17	57	95	149
L	400 (a)	0	0	0	0	0	0	0
	400 (b)	0	0	0	0	0	0	0
	500	0	0	0	0	17	57	78

En la presente tabla mostramos el resultado del modelo en los distintos años, tenemos pues que en el año 1995 en el nivel 100 se encontraban 213 estudiantes, el siguiente año, en 1996, en nivel 100 habian 338 de los cuales el 51.7% de estos fueron estudiantes rezagados.

En 1997 en nivel 100 habia 433 estudiantes en nivel 100 de los cuales el 53.3% de estos correspondia a estudiantes rezagados de las promociones anteriores.

Conclusiones y Recomendaciones

De los resultados obtenidos cabe resaltar lo siguiente:

La Carrera de Economia en Gestión Empresarial presenta una población en su mayoria joven y perteneciente al sexo femenino con una tendencia positiva de crecimiento, mientras que en la Carrera de Ingenieria Estadistica Informatica con una población en su mayoria joven y perteneciente al sexo masculino presenta una tendencia de crecimientos y decrecimientos alternados.

La Carrera de Economia tiene una estructura socioeconomica distinta al resto de carreras, pues su poblacion en su mayoria pertenece a la clase media y alta, mientras que en Ingenieria Estadistica Informatica su estructura socioeconomica se asemeja al de las carreras tradicionales, es decir con una poblacion en su mayoria perteneciente a la clase media baja y baja.

Mayoritariamente los estudiantes cuyo lugar de origen no es Guayaquil pertenecen a la clase socioeconomica baja.

Se observa un mejor rendimiento en el sexo femenino en ambas carreras, además de observarse una dependencia (no lineal) entre el rendimiento y el Factor Socioeconomico.

Con respecto a los criterios en la selección del numero de factores a retener el numero equivalente se presenta como un buen criterio, presentando la ventaja de no necesitar demasiada experiencia del investigador en el momento de su aplicacion.

Se observa una mayor dificultad en la aprobacion de materias correspondientes a los primeros niveles de estudio, tanto en Economia como en Ingenieria Estadistica Informatica, siendo el grado de dificultad mas acentuado en la carrera de Ingenieria Estadistica. Asi en Economia la probabilidad de aprobar el primer nivel es de 0.68 disminuyendo ligeramente en los niveles siguientes, en Estadistica Informatica el comportamiento es distinto pues la probabilidad de aprobar el primer nivel es de 0.18 aumentando en los niveles siguientes.

Por medio de la información recopilada, el adecuado procesamiento de los datos, y los resultados presentados y analizados en los capitulos anteriores, es procedente concluir que el modelo matematico propuesto constituye un acercamiento a la realidad del proceso estudiantil en la, ESPOL, y que la proyección del numero de estudiantes graduados es fundamental para la planificacion de la acogida del mercado para estos futuros profesionales.

Es importante diferenciar los escenarios establecidos para la prediccion de graduados en la ESPOL, ya que estos sufren modificaciones y distintos comportamientos.

El modelo sugerido constituye una herramienta valedera para posteriores estudios sobre la prediccion de estudiantes graduados en otras carreras por lo siguiente:

- Los modelos clasicos de prediccion e inferencia estadistica consideran la independencia de las variables, hipotesis que muchas veces no se cumple, mientras que el modelaje por medio de cadenas de Markov permite incorporar la dependencia entre las

variables, lo que resulta una ventaja en el instante de modelar el sistema educativo.

- Los resultados expresados tienen similitud con la realidad ya establecida, en lo referente al numero de graduados en las anteriores promociones.

- Es una excelente herramienta en la planificacion academica pues además de la predicción de egresados por año, provee el numero de estudiantes en cada nivel.

Basandose en los resultados obtenidos y las dificultades presentadas en el estudio se considera pertinente recomendar lo siguiente:

Establecer un sistema global que recolecte toda la información academica de la ESPOL, la misma que permita el desarrollo de estudios de tipo educativo que permitan una planificacion eficiente y eficaz.

El Modelo Matematico de Markov aqui presentado constituye un buen predictor del numero de graduados, y del numero de estudiantes por nivel y puede ser complementado con un estudio de mercado en el que se

establezca la capacidad del mercado laboral para receptar estos futuros profesionales.

El desarrollo de un estudio que investigue las causas que inciden en el bajo rendimiento estudiantil observado en los estudiantes novatos.

Con respecto al modelo se debe realizar un ajuste semestral en lo referente a las probabilidades de transición, con el objeto de actualizar el modelo a los cambios que pueden surgir en el programa de estudios.

BIBLIOGRAFIA

1. FREUND JOHN E.:WALPOLE RONALD E.. Estadistica Matematica con Aplicaciones, Prentice-Hall Hispanoamericana, Cuarta Edición. 1990.
2. HAMDY A. Taha., lnvestigaciones de Operaciones. Editorial ALFAOMEGA, 1992.
3. JAMES STEVEN., Applied Multivariate Statistics For the Social Sciences, LAWRENCE ERLBAUM ASSOCIATES, 1996.
4. JOHNSON RICHARD A.: WICHERN DEAN W., Applied Multivariate Statistical Analysis. Editorial Prentice-Hall, Año 1992.
5. MAKRIDAKIS-WHEELWRIGTH, Metodos de Pronosticos. Editorial LIMUSA S.A., 1998.
6. McCLAVE SHEAFFER, Probabilidad y Estadistica para Ingenieria, Grupo Editorial Iberoamericana, 1993.
7. ROSS SHELDON M. Probability random variables and stochastics processes, Editorial McGraw-Hill, 1965.
8. SAKRISON, DAVID J., Procesos Estocasticos Aplicaciones. 1970.

APENDICE A

DETERMINACIÓN DEL FACTOR SOCIOECONOMIC0

El valor del Factor Socioeconomico es la suma de cinco puntajes provenientes de:

a) **Colegio de procedencia,** que se clasifica en función de la pension oficial del colegio (FACTOR COL). El puntaje por colegio (COL) tiene un rango de **1** hasta **12** puntos SEGÚN la TABLA de COLEGIOS. A los estudiantes de colegios fiscales se les asignara COL igual a uno. A los estudiantes de colegios particulares, que hubiesen tenido beca completa en el colegio, se les asignara un FACTOR COL igual a la mitad del que le corresponda al colegio. Si el estudiante hubiera tenido solo media beca, se le asignara puntaje del 75% del que le corresponda al colegio.

b) **La ubicacion de la vivienda del Grupo Familiar del estudiante.** El grupo familiar del estudiante lo constituyen las personas que residen en la vivienda donde habita la persona de quien dependa economicamente el estudiante, excluyendo el personal de servicio. El estudiante podría ser considerado independientemente si habita fuera del nucleo familiar y certifica ingresos propios superiores a 6 SMV. Se establece un puntaje

(UBI) de acuerdo al sector residencial de la ciudad donde habita el grupo familiar y tendra un valor entre 0 y 14 puntos. Para Guayaquil y cantones aledaños se establecen seis categorias: Residencial Exclusivo (RE), Residencial Alto (RA), Residencial Medio Alto (RD), Residencial Medio (RM), Residencial de lnteres Social (RS), Sector Popular (SP). Para Quito se estableceran tres categorias: Residencial Alto (RA), Residencial Medio Alto (RM), Sector Popular (SP). Para Galapagos solo habra una categoria, Residencial de lnteres Social (RS). Para las capitales de otras provincias, cabeceras cantonales y sectores rurales habra dos categorias: Residencial Medio (RM) y Sector Popular (SP).

c) **La densidad habitacional de la vivienda del Grupo Familiar del estudiante.** Este puntaje (DEN) se establecera en base al factor que resulta de dividir los metros cuadrados construidos de la vivienda para el numero de miembros del grupo Familiar (FACTOR DEN), y tomara valores entre 1 y **4** puntos.

d) El lngreso per capita del Grupo Familiar del estudiante. Para su calculo se sumaran todos los ingresos mensuales de los miembros del grupo familiar, de los que se restara los egresos por arriendo o hipoteca de la vivienda y este resultado se dividira para el numero de personas que conforman el Grupo Familiar del estudiante (FACTOR IPC). El puntaje

(IPC) estara dado segun el FACTOR IPC y tomara valores entre 0 y 4 puntos.

e) **El consumo per capita de energia electrica de la vivienda del Grupo Familiar del estudiante.** Este puntaje (CEE) se calculara en función del valor per capita del consumo mensual de energia eléctrica tomando un promedio de los tres ultimos meses. El puntaje CEE tomara valores en 1 y 6 puntos. El puntaje Total P socioeconómico, asignado a cada estudiante, puede tomar valores entre 3 y 40 puntos y viene dado de la suma:

$$P = COL + UBI + DEN + IPC + CEE$$

Este puntaje podra ser revisado durante el primer año de estudios del alumno. Para solicitar revision el estudiante abonara una tasa de 50.000 sucres por la verificación.

A los estudiantes extranjeros no residentes se les asignara un factor P=40 puntos.

A los estudiantes extranjeros residentes que no sean graduados a nivel superior se les asignara un P igual al de un nacional.

A los estudiantes que ya tienen titulo profesional, sean estos nacionales o extranjeros residentes, se les asignara un factor P=25 puntos.

A los estudiantes que, por convenios o por decision institucional, estan sujetos al PAGO MÍNIMO en sus valores de registro, se les asignara un P= 7 puntos.

TABLA PARA CLASIFICACIÓN DE LOS COLEGIOS

RANGO DE PENSIONES MENSUALES	COL
Colegios fiscales y particulares hasta 100000 sucres	1
Mas de 60000 sucres hasta 200000 sucres	2
Mas de 200000 sucres hasta 300000 sucres	3
Mas de 300000 sucres hasta 400000 sucres	4
Mas de 400000 sucres hasta 500000 sucres	5
Mas de 500000 sucres hasta 600000 sucres	6
Mas de 600000 sucres hasta 700000 sucres	7
Mas de 700000 sucres hasta 800000 sucres	8
Mas de 800000 sucres hasta 900000 sucres	9
Mas de 900000 sucres hasta 1200000 sucres	10
Mas de 1200000 sucres hasta 1500000 sucres	11
Mas de 1500000 sucres	12

TABLA DE CLASIFICACIÓN DE LA UBICACIÓN

SECTOR	UBI
Sector Popular (SP)	0
Residencial de Interes Social (RS)	2
Residencial Medio (RM)	5
Residencial Medio Alto (RD)	8
Residencial Alto (RA)	11
Residencial Exclusivo (RE)	14

TABLA DE DENSIDAD HABITACIONAL

FACTOR DEN	DEN
Hasta 18 m^2	1
Mas de 18 m^2 hasta 36 m^2	2
Mas de 36 m^2 hasta 54 m^2	3
Mas de 54 m^2	4

TABLA DE INGRESO PER CAPITA

FACTOR IPC	IPC
Mas de 100000 sucres hasta 400000 sucres	0

Mas de 400000 sucres hasta 1000000 sucres	2
Mas de 1000000 sucres hasta 2000000 sucres	3
Mas de 2000000 sucres	4

TABLA DE CONSUMO DE ENERGÍA ELECTRICA

VALOR MENSUAL PER CAPITA	CEE
Hasta 100000 sucres	1
Más de 100000 sucres hasta 200000 sucres	2
Más de 200000 sucres hasta 300000 sucres	3
Mas de 300000 sucres hasta 400000 sucres	4
Mas de 400000 sucres hasta 500000 sucres	5
Mas de 500000 sucres	6