

RESUMEN

El presente proyecto tiene como objetivo el estudio y análisis estadístico Multivariado.

En este proyecto se procederá Componentes Principales que se obtengan.

ÍNDICE GENERAL

INTRODUCCION

CAPÍTULO 1

1. PROYECTO DE AUTOEVALUACIÓN ESPOL 2005

1.1 INTRODUCCIÓN

1.2 RAZONES QUE LLEVARON A TOMAR LA DECISIÓN DE REALIZAR ESTE PROYECTO

1.3 ESTADISTICA DESCRIPTIVA

CAPÍTULO 2

2. ANALISIS ESTADISTICO MULTIVARIADO

2.1 Introducción

2.2 Marco Teórico

2.2.1 Matriz de datos

2.3 Matriz de varianzas y Covarianzas

2.4 ANÁLISIS DE COMPONENTES PRINCIPALES

2.4.1 Marco Teórico

2.4.2 Desarrollo

2.5 ANÁLISIS DE FACTORES

2.5.1 Marco Teórico

2.5.2 Desarrollo

2.6 ANÁLISIS DE CORRELACIÓN CANÓNICA

2.6.1 Marco Teórico

2.6.2 Desarrollo

2.7 ANÁLISIS DE CLUSTER

2.7.1 Marco Teórico

2.7.2 Desarrollo

2.8 ANÁLISIS DISCRIMINANTE

2.8.1 Marco Teórico

2.8.2. Desarrollo

2.9. ANÁLISIS DE ESCALAMIENTO DIMENSIONAL

2.9.1 Marco Teórico

2.9.2. Desarrollo

2.10 ANÁLISIS REGRESION LOGISTICA

μυπηφ

CONCLUSIONES

BIBLIOGRAFÍA

INTRODUCCIÓN

Es indudable que “La juventud ecuatoriana dedica cada vez menor tiempo a leer” es una frase que es común escuchar y más aún se la toma como cierta, pero el demostrarlo de manera científica utilizando técnicas estadísticas para descubrir las causas que nos impiden practicar más la lectura es uno de los objetivos de esta investigación para esto se han planteado algunas hipótesis, proposiciones que fueron tomadas en consideración para este proyecto.

Todos los jóvenes se ven acosados por distintos flujos de información que merman la capacidad y el interés en leer un buen libro, claro está que no todos opinan de esa manera, también es razonable pensar que hay personas que no les gusta leer, pero más que eso este proyecto está diseñado para evaluar esas causas .

Nuestra investigación está dirigida a los estudiantes de la Facultad de Ciencias Humanísticas y Administrativas “ICHE” en la cual pondremos en prácticas los conocimientos obtenidos en la clase de Muestreo y demás cursos de estadística.

Nuestro espíritu altruista al dejar constancia de nuestro conocimiento siempre en el bien común nos lleva a realizar esta investigación.

CAPITULO

1

1.2 RAZONES QUE LLEVARON A TOMAR LA DECISIÓN DE REALIZAR ESTE PROYECTO

El proyecto nace como requisito para la materia de ANÁLISIS MULTIVARIADO DE DATOS de la carrera de Ingeniería Estadística Informática.

La idea fue planteada por nuestro profesor Máster Gaudencio Zurita el cual propuso este proyecto para desarrollar nuestras destrezas y poner en práctica nuestros conocimientos acerca del Análisis Multivariado.

Para este proyecto se utilizó un estudio previo llamado “PROYECTO DE AUTOEVALUACION ESPOL 2005” los cuales fueron proporcionados por el Mgcs. Gaudencio Zurita el cual es director del **CENTRO DE ESTUDIOS E INVESTIGACIONES ESTADISTICAS ICM-ESPOL** con el fin de poner en práctica los conocimientos adquiridos en el Curso de Análisis Multivariado de Datos. utilizando las herramientas estadísticas que hemos adquirido en nuestro paso por las diferentes materias de estadística y Muestreo.

1.3 ESTADISTICA DESCRIPTIVA

<i>Nº</i>	<i>Proposición</i>	<i>variable</i>
1.-	<i>Tengo confianza en aplicar mis conocimientos de álgebra para resolver problemas de Planteamiento de Ecuaciones.</i>	<i>X₁</i>
2.-	<i>Tengo confianza en aplicar mis conocimientos de trigonometría para resolver problemas sobre triángulos rectángulos.</i>	<i>X₂</i>
3.-	<i>Tengo confianza en aplicar mis conocimientos de trigonometría para resolver problemas sobre triángulos oblicuángulos.</i>	<i>X₃</i>
4.-	<i>He utilizado frecuentemente una calculadora básica.</i>	<i>X₄</i>
5.-	<i>He utilizado frecuentemente una calculadora científica.</i>	<i>X₅</i>
6.-	<i>He utilizado frecuentemente una calculadora gráfica.</i>	<i>X₆</i>
7.-	<i>He utilizado frecuentemente una calculadora programable.</i>	<i>X₇</i>
8.-	<i>He utilizado una calculadora para determinar el promedio de un conjunto de datos.</i>	<i>X₈</i>
9.-	<i>He utilizado una calculadora para determinar la desviación estándar de un conjunto de datos.</i>	<i>X₉</i>
10.-	<i>Tengo confianza en mis habilidades para graficar e interpretar un histograma de frecuencias.</i>	<i>X₁₀</i>

Tabla 1 Estadísticas Descriptivas y Grafico 1 Histograma de la variable
“Tengo confianza en aplicar mis conocimientos de álgebra para resolver problemas de Planteamiento de Ecuaciones”.

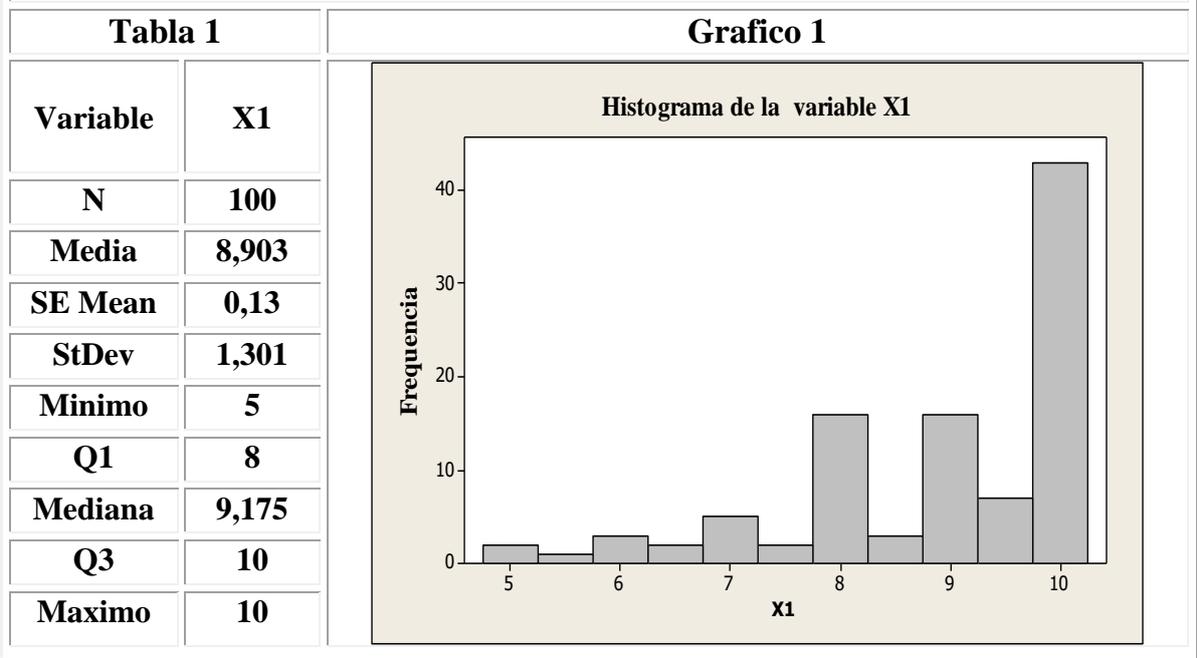
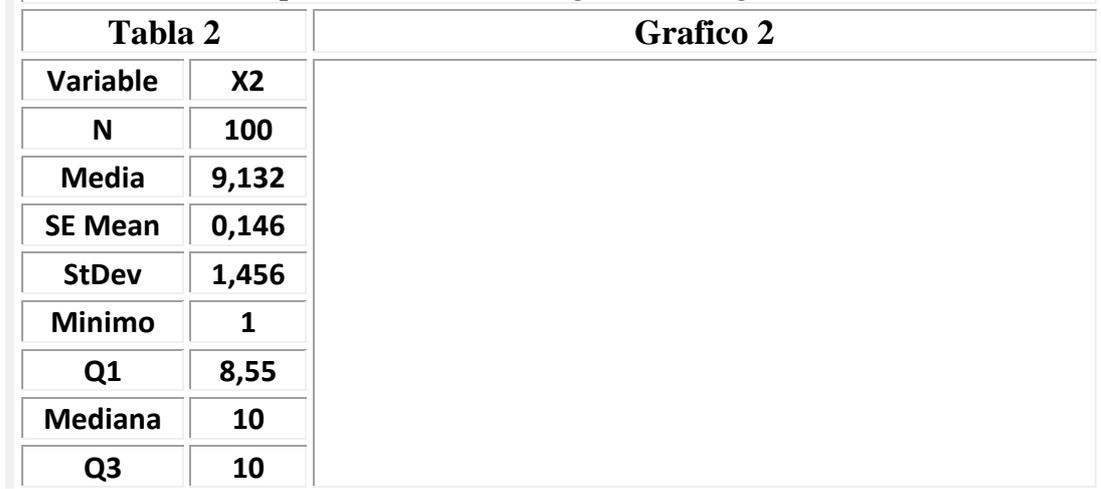
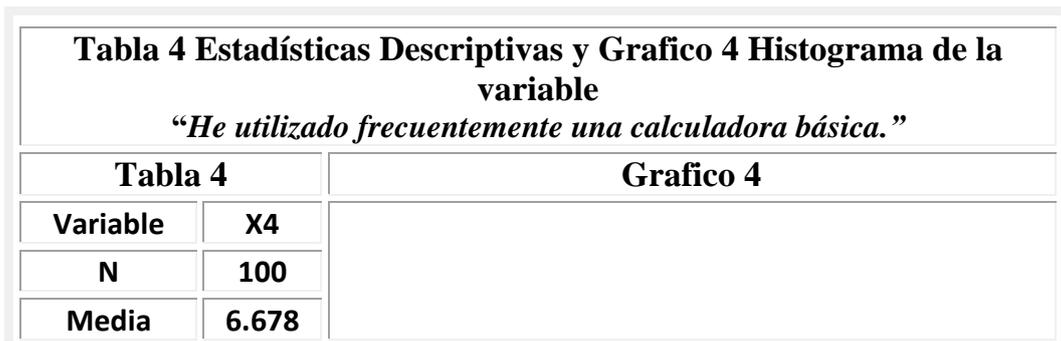
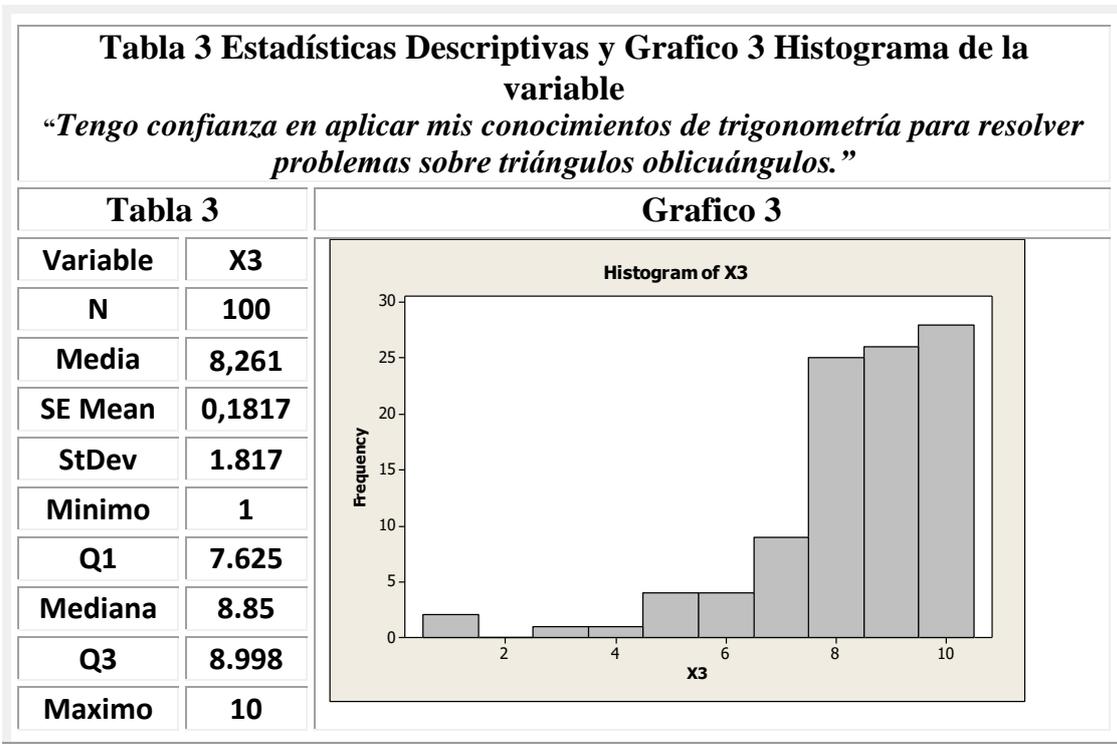
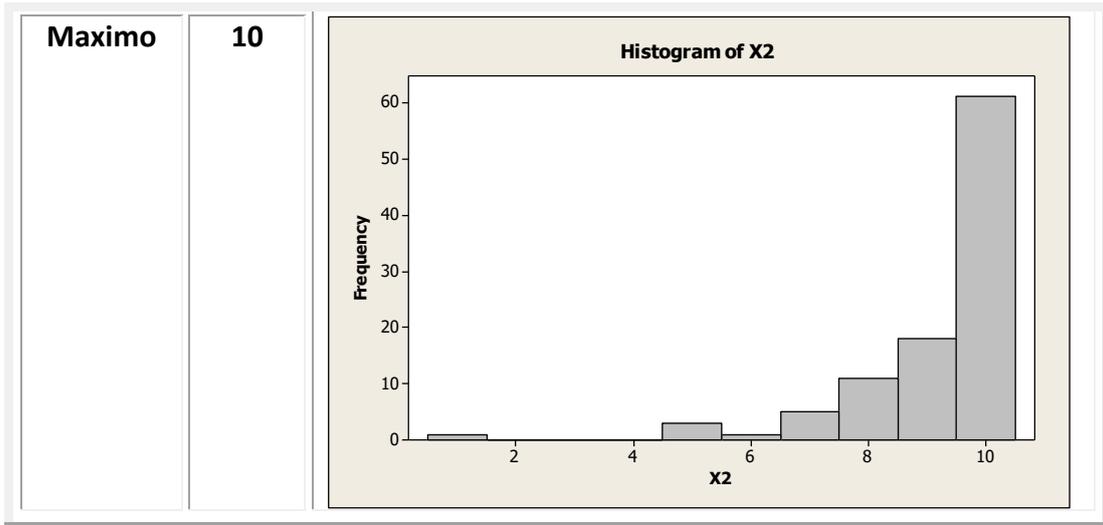


Tabla 2 Estadísticas Descriptivas y Grafico 2 Histograma de la variable
“Tengo confianza en aplicar mis conocimientos de trigonometría para resolver problemas sobre triángulos rectángulos.”





SE Mean	0.365
StDev	3.645
Minimo	1
Q1	7.625
Mediana	8.85
Q3	8.998
Maximo	10
IQR	7

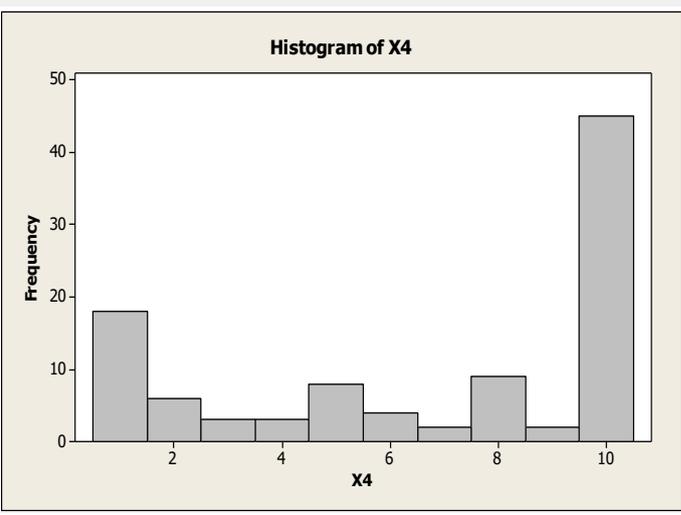


Tabla 5 Estadísticas Descriptivas y Grafico 5 Histograma de la variable
“He utilizado frecuentemente una calculadora científica.”

Tabla 5	
Variable	X5
N	100
Media	8.566
SE Mean	0.236
StDev	2.36
Minimo	1
Q1	8
Mediana	10
Q3	10
Maximo	10
IQR	2

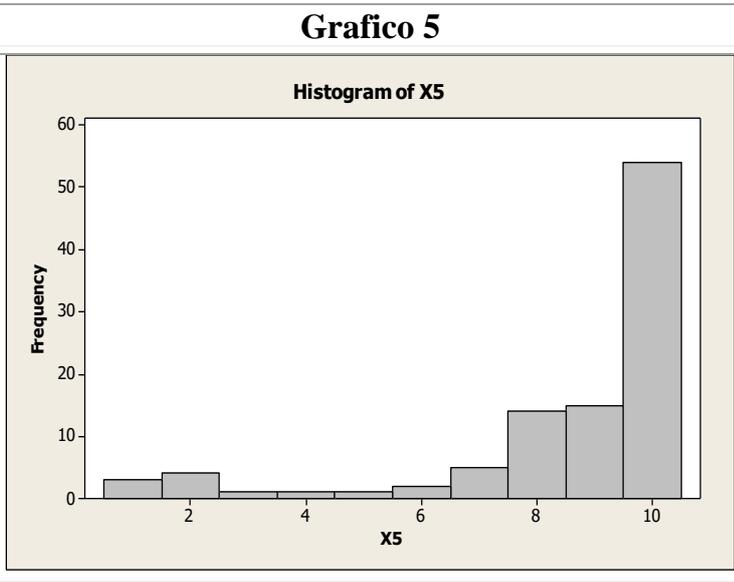


Tabla 6 Estadísticas Descriptivas y Grafico 6 Histograma de la variable

“He utilizado frecuentemente una calculadora gráfica.”

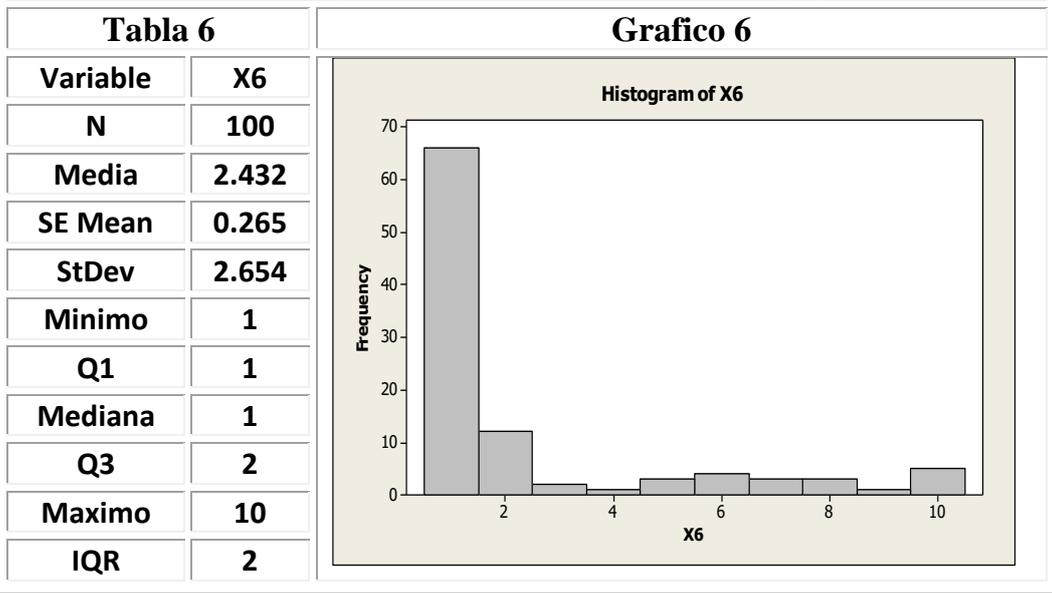


Tabla 7 Estadísticas Descriptivas y Grafico 7 Histograma de la variable

“He utilizado frecuentemente una calculadora programable.”

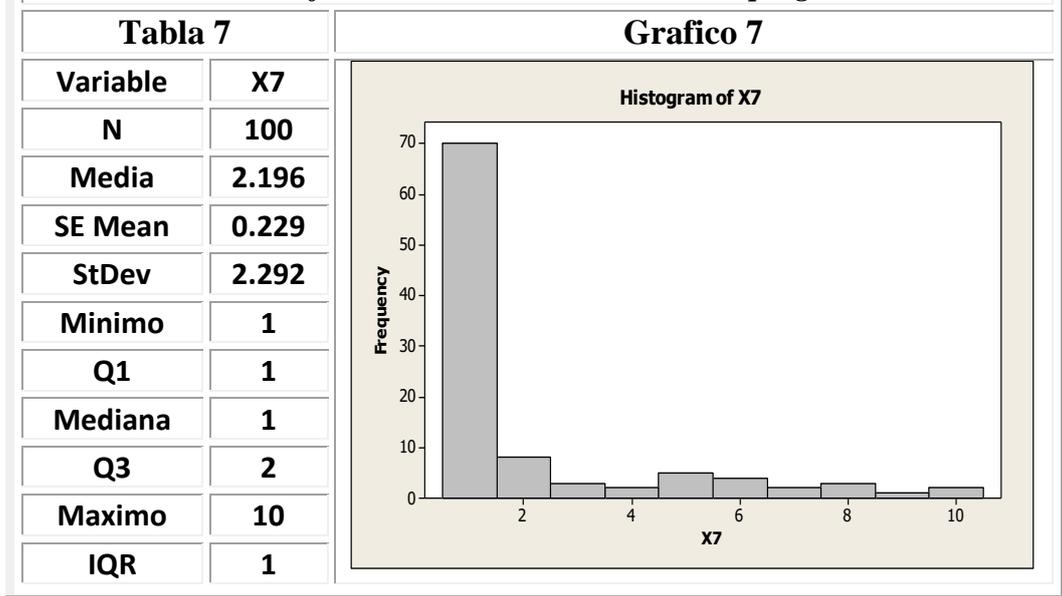


Tabla 8 Estadísticas Descriptivas y Grafico 8 Histograma de la variable

“He utilizado una calculadora para determinar el promedio de un conjunto de datos.”

Tabla 8		Grafico 8
Variable	X8	
N	100	
Media	5.74	
SE Mean	0.347	
StDev	3.467	
Minimo	1	
Q1	1	
Mediana	7	
Q3	9	
Maximo	10	
IQR	8	

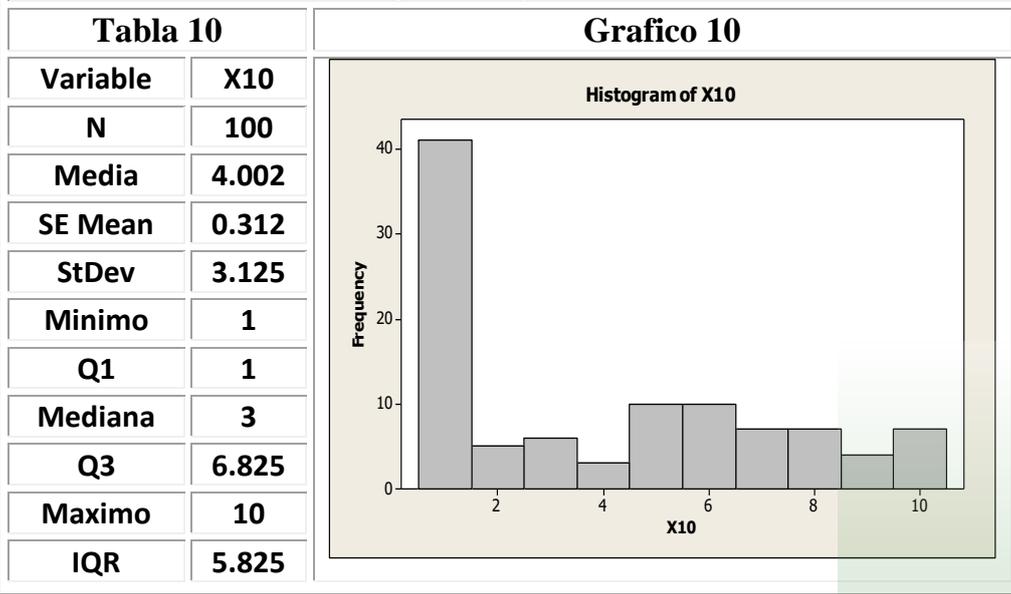
Tabla 9 Estadísticas Descriptivas y Grafico 9 Histograma de la variable

“He utilizado una calculadora para determinar el promedio de un conjunto de datos.”

Tabla 9		Grafico 9
Variable	X9	
N	100	
Media	2.752	
SE Mean	0.289	
StDev	2.89	
Minimo	1	
Q1	1	
Mediana	1	
Q3	4	
Maximo	10	
IQR	3	

Tabla 10 Estadísticas Descriptivas y Grafico 10 Histograma de la variable

“Tengo confianza en mis habilidades para graficar e interpretar un histograma de frecuencia.”



CAPITULO

2

2. ANALISIS ESTADISTICO MULTIVARIADO

Introducción.

En el presente capítulo se realizará el análisis estadístico multivariado a las características de interés de los entrevistados.

El estudio será para 100 entes de observación a los cuales se les realizo mediciones de 10 características de interés. El análisis estará descrito por las técnicas estadísticas, como son el análisis de correlación, componentes principales, Análisis de Factores y

Correlación Canónica, Análisis discriminante, Análisis de Cluster, Análisis de Escalamiento Dimensional.

En la *Sección 2.1* serán definidas las técnicas a utilizar. En la *Sección 2.3* se realizará el análisis de correlación entre las variables de estudio. La *Sección 2.4* se realizará el análisis de Componentes Principales La *Sección 2.5*. Se realizará el Análisis de Factores En la *Sección 2.6* se realizará el análisis de Correlación Canónica. En la *Sección 2.7* se realizará el Análisis discriminante, en la *Sección 2.8* se realizará el Análisis de Cluster Finalmente, en la *Sección 2.9* se presenta el Análisis de Escalamiento Dimensional

Es importante mencionar que para este análisis, se utilizó el software estadístico SPSS versión 11, MINITAB versión 14, SYSTAT versión 11 y MATLAB versión 6.1

2.2 Marco Teórico

2.2.1 Matriz de Datos

Los datos deben ser representados en forma de un arreglo rectangular de n filas y p columnas denotado por \mathbf{X} , a esta matriz se la denomina matriz de datos; en donde cada elemento de X_{ij} de la matriz representa la i -ésima observación de la j -ésima variable, así:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1p} \\ X_{21} & X_{22} & X_{23} & \cdots & X_{2p} \\ X_{31} & X_{32} & X_{33} & & X_{3p} \\ \vdots & \vdots & & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & \cdots & X_{np} \end{bmatrix} \in M_{n \times p} \quad \begin{array}{l} \text{para } i = 1, 2, \dots, n \\ j = 1, 2, \dots, p \end{array}$$

La matriz \mathbf{X} contiene todas las observaciones de todas las variables.

En nuestro estudio nuestra matriz de datos contiene 100 filas (antes de investigación) y diez columnas (características de interés a investigar) es la que se muestra en la Tabla 11- Anexo 1

Vector de Medias

Sea $\mathbf{X}^T = (X_1, X_2, \dots, X_p)$ un vector aleatorio, se define el **vector de medias** $\boldsymbol{\mu}$, correspondiente al vector \mathbf{X} , donde:

$$\boldsymbol{\mu} = E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_k) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \\ \vdots \\ \mu_p \end{bmatrix} \in \mathfrak{R}^p$$

Donde el valor μ_k representa el valor esperado de la variable aleatoria X_k ,

es decir $\mu_k = E(X_k)$ para $k = 1, 2, 3, \dots, p$.

Para nuestro estudio el vector de medias es el siguiente:

$$\boldsymbol{\mu}_{10} = E(\mathbf{X}_{10}) = \begin{pmatrix} 8,903 \\ 9,132 \\ 8,261 \\ 6,678 \\ 8,566 \\ 2,432 \\ 2,196 \\ 5,74 \\ 2,752 \\ 4,022 \end{pmatrix}$$

Expresión 1.1

Después de haber definido el vector de medias $\boldsymbol{\mu}$, se explica el cálculo de la matriz de varianzas y covarianzas $\boldsymbol{\Sigma}$.

2.3 Matriz de Varianzas y Covarianzas

Siendo $\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$, para $i, j = 1, 2, 3, \dots, p$, donde σ_{ij} representa la covarianza entre X_i y X_j , por lo tanto $\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$, el mismo que se expresa como:

$$\Sigma = E \left\{ \begin{array}{l} \left[\begin{array}{c} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{array} \right] \left[\begin{array}{cccc} X_1 - \mu_1 & X_2 - \mu_2 & \dots & X_p - \mu_p \end{array} \right] \end{array} \right\} \in M_{p \times p}$$

$$\Sigma = \begin{bmatrix} E[(X_1 - \mu_1)^2] & E[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & E[(X_1 - \mu_1)(X_p - \mu_p)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & E[(X_2 - \mu_2)^2] & \cdots & E[(X_2 - \mu_2)(X_p - \mu_p)] \\ \vdots & \vdots & \ddots & \vdots \\ E[(X_p - \mu_p)(X_1 - \mu_1)] & E[(X_p - \mu_p)(X_2 - \mu_2)] & \cdots & E[(X_p - \mu_p)^2] \end{bmatrix} \in M_{p \times p}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1j} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2j} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \sigma_{i1} & \sigma_{i2} & \cdots & \sigma_{ij} & \cdots & \sigma_{ip} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pj} & \cdots & \sigma_{pp} \end{bmatrix} \in M_{p \times p}$$

En nuestro estudio determinamos el estimador **S** de la matriz de Varianzas y Covarianzas dado que se tiene una muestra tamaño 100 obteniendo la siguiente matriz 10x10:

Cuadro 1										
Proyecto final de Análisis de Multivariado de Datos										
Estimador de la matriz de varianzas y Covarianzas										
S =	1,69272	1,09305	1,18874	0,08656	0,62765	0,46193	0,40751	0,34912	0,49679	-0,22967
	1,09305	2,11889	1,97102	-0,49855	0,68582	0,36306	0,49247	0,92913	0,52296	-0,09668
	1,18874	1,97102	3,30266	-0,03908	0,55791	0,93232	0,7241	0,79506	1,10382	0,52935
	0,08656	-0,49855	-0,03908	13,2886	1,42753	1,80246	2,30112	1,97099	1,11059	-0,45547
	0,62765	0,68582	0,55791	1,42753	5,57095	0,38707	0,95461	1,71538	0,27187	-0,18531
	0,46193	0,36306	0,93232	1,80246	0,38707	7,04445	2,55693	0,8009	3,43919	2,28475
	0,40751	0,49247	0,7241	2,30112	0,95461	2,55693	5,25392	0,7322	2,25839	1,64302
	0,34912	0,92913	0,79506	1,97099	1,71538	0,8009	0,7322	12,02219	2,81701	1,67094
	0,49679	0,52296	1,10382	1,11059	0,27187	3,43919	2,25839	2,81701	8,35371	4,4633
	-0,22967	-0,09668	0,52935	-0,45547	-0,18531	2,28475	1,64302	1,67094	4,4633	9,768

De la matriz obtenida S simétrica. En su diagonal principal Diag. (S) se encuentran la varianzas de cada variable. Mientras que i diferente de j se encuentran las covarianzas entre cada par de variables.

2.3.1 Matriz de Correlación

La matriz de correlación está basada en el coeficiente de correlación ρ_{ij} , definido en términos del cociente entre la covarianza ij σ y el producto de varianzas σ_{ii} y σ_{jj} .

Se tiene un arreglo de p filas y p columnas que agrupa todas las medidas de las relaciones de tipo lineal que existen entre las p variables investigadas, denominadas Matriz de Correlación (ρ).

Se define ρ (matriz de correlación) así:

$$\rho = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \dots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \dots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \dots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix} \in M_{p \times p}$$

$$= \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \rho_{22} & \dots & \rho_{2p} \\ \rho_{31} & \rho_{32} & 1 & \dots & \rho_{3p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{p1} & \rho_{p2} & \rho_{p3} & \dots & \rho_{pp} \end{bmatrix} \in M_{p \times p}$$

Siendo $\sigma_{ij} = \sigma_{ji}$, se concluye que ρ es simétrica, tal como lo es Σ .

En nuestro estudio determinamos el estimador **R** de la matriz de Correlación que se tiene una muestra tamaño 100 obteniendo la siguiente matriz 10x10:

Cuadro 2										
Proyecto final de Análisis de Multivariado de Datos										
Estimador de la matriz de Correlación										
R=	1	0,577	0,503	0,018	0,204	0,134	0,137	0,077	0,132	-0,056
	0,577	1	0,745	-0,094	0,2	0,094	0,148	0,184	0,124	-0,021
	0,503	0,745	1	-0,006	0,13	0,193	0,174	0,126	0,21	0,093
	0,018	-0,094	-0,006	1	0,166	0,186	0,275	0,156	0,105	-0,04
	0,204	0,2	0,13	0,166	1	0,062	0,176	0,21	0,04	-0,025
	0,134	0,094	0,193	0,186	0,062	1	0,42	0,087	0,448	0,275
	0,137	0,148	0,174	0,275	0,176	0,42	1	0,092	0,341	0,229
	0,077	0,184	0,126	0,156	0,21	0,087	0,092	1	0,281	0,154
	0,132	0,124	0,21	0,105	0,04	0,448	0,341	0,281	1	0,494
	-0,056	-0,021	0,093	-0,04	-0,025	0,275	0,229	0,154	0,494	1

2.4. ANÁLISIS DE COMPONENTES PRINCIPALES

2.4.1 Marco Teórico

2.4.2 Desarrollo