

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**

Instituto de Ciencias Matemáticas

**Ingeniería en Estadística Informática**

**“Efectos de la Imputación en el Análisis de Datos Multivariados”**

**TESIS DE GRADO**

**Previa la obtención del título de:**

 **INGENIERA EN ESTADÍSTICA INFORMÁTICA**

**Presentada por:**

**Marcia Gabriela Cuenca Calle**

**GUAYAQUIL – ECUADOR**

**AÑO**

**2006**

**AGRADECIMIENTO**

A Dios y la Virgen Santísima, por haberme permitido llegar hasta aquí. A mis padres y hermana, por la ayuda incondicional. A todos mis amigos: Eduardo, Emma, Patricio, Freddy, Juan, David, Mónica, Jorge, Fico, Fátima, Evelyn, etc. que con su ayuda y apoyo constante, han estado siempre presente.

A todos ellos,
Muchas Gracias

**DEDICATORIA**

A los seres que me enseñaron que la responsabilidad, el esfuerzo y la perseverancia son los únicos medios para alcanzar el éxito.

A mis padres y hermana, *ERNESTINA, MANRIQUE y PAOLA*.

**TRIBUNAL DE GRADUACIÓN**

 **Ing. Washington Armas M. Sc. Gaudencio Zurita**

 **DIRECTOR DEL ICM DIRECTOR DE TESIS**

 **Ing. Erwin Delgado Mat. Johni Bustamante**

 **VOCAL VOCAL**

DECLARACIÓN EXPRESA

“La responsabilidad del contenido de esta Tesis de Grado, me corresponde exclusivamente, y el patrimonio intelectual de la misma a la Escuela Superior Politécnica del Litoral”

(*Reglamento de Graduación de la ESPOL*)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Marcia Gabriela Cuenca Calle

**RESUMEN**

El presente trabajo consiste en un estudio estadístico acerca de los Efectos de la Imputación en el Análisis de Datos Multivariados, basados en muestras con variables aleatorias dependientes e independientes de diferentes tamaños y distribuciones, así como también el análisis de un caso real.

La tesis está conformada por cuatro capítulos más las conclusiones y recomendaciones. El primer capítulo describe los principios estadísticos relacionados con los Métodos de Imputación que son parte de esta investigación.

En el capítulo dos aborda las técnicas y principios científicos que permiten la generación de números aleatorios. El tercer capítulo ilustra las técnicas de imputación para el manejo de datos incompletos en una matriz de datos. En el siguiente capítulo se comparan los métodos de imputación por medio de simulaciones. Finalmente se muestran las conclusiones y recomendaciones basadas en los resultados obtenidos en este trabajo.

**INDICE GENERAL**

|  |  |
| --- | --- |
| RESUMEN………………………………………………………………….……. | I |
| INDICE GENERAL.…………………………………………………….……….. | II |
| SIMBOLOGÍA……………………………………………………………………. | III |
| ÍNDICE DE TABLAS……………………………………………………………. | IV |
| ÍNDICE DE GRÁFICOS………………………………………………………… | V |
| ÍNDICE DE CUADROS…………………………………………………………. | VI |
| INTRODUCCIÓN………………………………………………...……………… | VII |

|  |  |
| --- | --- |
| 1. LA PÈRDIDA DE DATOS EN UNA INVESTIGACIÒN |  |
| 1.1. Introducción……………………………………………….…………….. | 1 |
| 1.2. Matriz de Datos Multivariados………………………………………… | 2 |
| 1.3. Variables Aleatorias Univariadas y Bivariadas……………………… | 2 |
| 1.4. La Pérdida de Datos en una Investigación…………………………... | 17 |
|  1.5 Métodos que emplean toda a información disponible………………. | 18 |
|  1.5.1. Método de Eliminación por Lista………………………………. | 18 |
|  1.5.2. Método de Eliminación por Pares……………………………… | 21 |

|  |  |
| --- | --- |
| 2. MODELOS ESTOCÁSTICOS A UTILIZARSE PARA IMPUTACIÓN DE DATOS |  |
| 2.1. Introducción……………………………………………………………… | 25 |
| 2.2. Distribución Uniforme…………………………………………………... | 26 |
|  2.3. Prueba de Bondad de Ajuste ……………………………………... | 32 |
| 2.4. Prueba de Kolmogorov-Smirnov……………………………………… | 34 |
| 2.5. Generación de Números Pseudo Aleatorios U(0,1)……………….. | 37 |
|  2.5.1 Generadores Congruenciales Lineales | 38 |
|  2.6. Métodos de Generación de Variables Aleatorias No Uniformes…... | 47 |
|  2.6.1 Método de la Transformada Inversa…………………………… | 48 |

|  |  |
| --- | --- |
| 3. TÈCNICAS DE IMPUTACIÒN APLICABLES |  |
| 3.1. Introducción……………………………………………………………… | 53 |
| 3.2. Imputación de Datos……………………………………………………. | 54 |
| 3.3. Métodos de Imputación………………………………………………… | 54 |
| 3.3.1. Imputación por la Media Muestral……………………..………. | 54 |
| 3.3.2. Modelo de Regresión Lineal Múltiple…………….……………. | 69 |
| 3.3.3. Imputación por Regresión……………………………………….. | 74 |

|  |  |
| --- | --- |
| 4. SIMULACIÒN BAJO DISTINTAS CONDICIONES UNIVARIADAS Y MULTIVARIADAS |  |
| 4.1. Introducción……………………………………………………………… | 101 |
| 4.2. Matrices de datos con variables aleatorias independientes……...… | 102 |
|  4.2.1. Distribución Normal: *Tres datos faltantes* en una sola variable (2% de la matriz), tamaño de muestra n=30…………………… | 102 |
|  4.2.2. Distribución Normal*: Tres datos faltantes*, *dos* en la variable *X*1 y *uno* en la variable *X*4 (2% de la matriz), tamaño de muestra n=30 | 117 |
|  4.2.3 Distribución Poisson: *Ocho datos faltantes* en una sola variable (5% de la matriz), tamaño de muestra n=30………….............. | 127 |
|  4.2.4 Distribución Exponencial: *Trece datos faltantes* en una sola variable (5% de la matriz), tamaño de muestra n=50…………………… | 140 |
| 4.3. Matrices de Datos con variables aleatorias dependientes…………. | 150 |
|  4.3.1 Distribución Normal: *Trece datos faltantes* en una sola variable (5% de la matriz), tamaño de muestra n=50…………………… | 150 |
|  4.3.2 Distribución Poisson: *Cincuenta datos faltantes* en una sola variable (10% de la matriz), tamaño de muestra n=100………………… | 176 |
|  4.3.3 Distribución Exponencial: *Cincuenta datos faltantes*: Veinticinco en *X*3  y veinticinco en *X*8 (10% de la matriz), tamaño de muestra n=100 ………………………………………………………………. | 193 |

|  |  |
| --- | --- |
| Conclusiones y Recomendaciones |  |
| Conclusiones…………………………………………………………………. | 216 |
| Recomendaciones……………………………………………………………. | 220 |

|  |  |
| --- | --- |
|  ANEXOS |  |
|  BIBLIOGRAFÍA |  |

**SIMBOLOGÍA**

 Matriz de datos multivariada.

P Población

 Conjunto de todos lo resultados posibles del experimento

**** Es el **-** álgebra de subconjuntos de 

 Conjunto de los Números reales.

*X* Variable Aleatoria

 Media Poblacional

 Varianza Poblacional

*MX(t)* Función Generadora de Momentos

 Número de subconjuntos, de tamaño n, entre N objetos disponibles.

 Media Muestral.

 Estimador insesgado de la media poblacional 

*s2* Varianza Muestral.

 Nivel de confianza al 100 %.

n Tamaño de muestra.

N Tamaño de la población.

 Coeficiente de correlación lineal entre las variables *i* y *k.*

 Desviación Estándar de la población.

 Vector Aleatorio *p-*variado.

 Matriz de varianzas y covarianzas.

*sij* Matriz muestral de varianzas y covarianzas.

 Covarianza entre las variables *i* y *j.*

***D****s* Matriz Diagonal

 Función de densidad.

 Distribución Uniforme con parámetros y .

H0  Hipótesis Nula

H1  Hipótesis Alternativa.

 Estadístico Ji Cuadrado.

 Estadístico K-S tabulado.

 Número Pseudos aleatorio.

 Sucesor de un número aleatorio.

*F* Función Acumulada.

 Valor que se coloca, “o imputa”, en la variable con datos faltantes.

 Media para datos incompletos.

**ÍNDICE DE TABLAS**

|  |  |  |
| --- | --- | --- |
| **Capítulo I** |  |  |
| Tabla 1.1 | Matriz de datos de variables aleatorias independientes con distribución Poisson , tamaño de muestra n=5………. | 19 |
| Tabla 1.2 | Matriz de datos de variables aleatorias independientes con distribución Poisson , Método de Eliminación por Filas, tamaño de muestra n=5, 13% de datos faltantes en la matriz…………………………………………………………. | 20 |
| Tabla 1.3 | Matriz de datos de variables aleatorias independientes con distribución Poisson , Método de Eliminación por Pares, tamaño de muestra n=5, 13% de datos faltantes en la matriz…………………………………………………………. | 22 |
| Tabla 1.4 | Variables aleatorias independientes con distribución Poisson , Método de eliminación por Pares, tamaño de muestra n=5, 13% de datos faltantes en la matriz……… | 24 |

|  |  |  |
| --- | --- | --- |
| **Capítulo II** |  |  |
| Tabla 2.1 | Prueba de Bondad de Ajuste…………………………………... | 34 |
| Tabla 2.2 | Matriz de Datos de variables aleatorias independientes con distribución Normal (0, 1), tamaño de muestra n=4 | 37 |
| Tabla 2.3 | Prueba de Kolmogorov-Smirnov………………………………. | 37 |
| Tabla 2.4 | Método Congruencial Mixto, números pseudos aleatorios del generador …………………………….. | 41 |
| Tabla 2.5 | Método Congruencial Mixto, números pseudoaleatorios del generador ………………………………. | 44 |
| Tabla 2.6 | Método Congruencial Multiplicativo, números pseudo aleatorios del generador ……………………. | 47 |

|  |  |  |
| --- | --- | --- |
| **Capítulo III** |  |  |
| Tabla 3.1 | Matriz de datos de variables aleatorias independientes con distribución Poisson, tamaño de muestra n=10, 3% de datos faltantes en la matriz……………………………………. | 59 |
| Tabla 3.2 | Matriz de datos de variables aleatorias independientes con distribución Poisson, tamaño de muestra n=10, 5% de datos faltantes en la matriz……………………………………. | 64 |
| Tabla 3.3 | Matriz de datos de variables aleatorias independientes con distribución Poisson, tamaño de muestra n=10, 13% de datos faltantes en la matriz……………………………………. | 67 |
| Tabla 3.4 | Matriz de datos de variables aleatorias independientes con distribución Poisson, Método de Imputación por la Media, tamaño de muestra n=10, 13% de datos completados en la matriz……………………………………………………………. | 68 |
| Tabla 3.5 | Matriz de datos de variables aleatorias dependientes con distribución Normal, tamaño de muestra n=10, 7% de datos faltantes en la matriz…………………………………… | 78 |
| Tabla 3.6 | Matriz de datos de variables aleatorias dependientes con distribución Normal, tamaño de muestra n=10, 7% de datos faltantes en la matriz, matriz particionada……………. | 79 |
| Tabla 3.7 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 7% de datos faltantes en la matriz, primeros valores estimados………………………….. | 82 |
| Tabla 3.8 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 7% de datos faltantes en la matriz, segundos valores estimados…………………………. | 83 |
| Tabla 3.9 | Matriz de datos de variables aleatorias dependientes con distribución Normal, tamaño de muestra n=10, 10% de datos faltantes en la matriz……………………………………. | 87 |
| Tabla 3.10 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Media, tamaño de muestra n=10, 10% de datos faltantes en la matriz……………………………………………………………. | 88 |
| Tabla 3.11 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, primeros valores estimados………………………….. | 90 |
| Tabla 3.12 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, segundos valores estimados…………………………. | 91 |
| Tabla 3.13 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, terceros valores estimados………………………….. | 92 |
| Tabla 3.14 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, cuartos valores estimados………………………….. | 93 |
| Tabla 3.15 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, quintos valores estimados………………………….. | 94 |
| Tabla 3.16 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, sextos valores estimados………………………….. | 95 |
| Tabla 3.17 | Matriz de datos de variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, séptimos valores estimados………………………….. | 96 |

|  |  |  |
| --- | --- | --- |
| **Capítulo IV** |  |  |
| Tabla 4.1 | Matriz de Datos de variables aleatorias independientes con distribución Normal (5, 1), tamaño de muestra n=30…………………………………………………………… | 105 |
| Tabla 4.2 | Matriz de Datos de variables aleatorias independientes con distribución Normal (5,1), tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Matriz de datos con tres filas eliminadas…………………………………………. | 106 |
| Tabla 4.3 | Matriz de Datos de variables aleatorias independientes con distribución Normal (5, 1), Método de Imputación por la Media, tamaño de muestra n=30 y 2% de datos faltantes en la matriz………………………………………… | 110 |
| Tabla 4.4 | Matriz de Datos de variables aleatorias independientes con distribución Normal (5, 1), Método de Imputación por Regresión, tamaño de muestra n=30 y 2% de datos faltantes en la matriz………………………………………… | 111 |
| Tabla 4.5 | Variables aleatorias independientes con distribución Normal (5,1), Comparación de los Métodos de Imputación, tamaño de muestra n=30 y 2% de datos faltantes en la matriz……………………………………..….. | 112 |
| Tabla 4.6 | Matriz de Datos de variables aleatorias independientes con distribución Normal (5, 1), tamaño de muestra n=30……………………………………………………………. | 120 |
| Tabla 4.7 | Matriz de Datos de variables aleatorias independientes con distribución Normal (5,1), tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Matriz de datos con tres filas eliminadas………………………………………….. | 121 |
| Tabla 4.8 | Matriz de Datos de variables aleatorias independientes con distribución Normal (5, 1), Método de Imputación por la Media, tamaño de muestra n=30 y 2% de datos faltantes en la matriz…………………………………………. | 124 |
| Tabla 4.9 | Matriz de Datos de variables aleatorias independientes con distribución Normal (5, 1), Método de Imputación por la Regresión, tamaño de muestra n=30 y 2% de datos faltantes en la matriz…………………………………………. | 125 |
| Tabla 4.10 | Variables aleatorias independientes con distribución Normal (5,1), Comparación de los Métodos de Imputación, tamaño de muestra n=30 y 2% de datos faltantes en la matriz…………………………………………. | 126 |
| Tabla 4.11 | Matriz de Datos de variables aleatorias independientes con distribución Poisson , tamaño de muestra n=30... | 130 |
| Tabla 4.12 | Matriz de Datos de variables aleatorias independientes con distribución Poisson , tamaño de muestra n=30 y 5% de datos faltantes en la matriz, Matriz de datos con ocho filas eliminadas………………………………………… | 132 |
| Tabla 4.13 | Matriz de Datos de variables aleatorias independientes con distribución Poisson , Método de Imputación por la Media, tamaño de muestra n=30 y 5% de datos faltantes en la matriz………………………………………… | 135 |
| Tabla 4.14 | Matriz de Datos de variables aleatorias independientes con distribución Poisson , Método de Imputación por la Regresión, tamaño de muestra n=30 y 5% de datos faltantes en la matriz…………………………………………. | 136 |
| Tabla 4.15 | Variables aleatorias independientes con distribución Poisson , Comparación de los Métodos de Imputación , tamaño de muestra n=30 y 5% de datos faltantes en la matriz………………………………………………………….. | 137 |
| Tabla 4.16 | Matriz de Datos de variables aleatorias independientes con distribución Exponencial , tamaño de muestra n=50……………………………………………………………. | 143 |
| Tabla 4.17 | Matriz de Datos de variables aleatorias independientes con distribución Exponencial , tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Matriz de datos con trece filas eliminadas……………………………. | 145 |
| Tabla 4.18 | Matriz de Datos de variables aleatorias independientes con distribución Exponencial , Método de Imputación por la Media, tamaño de muestra n=50 y 5% de datos faltantes en la matriz………………………………………… | 147 |
| Tabla 4.19 | Matriz de Datos de variables aleatorias independientes con distribución Exponencial , Método de Imputación por la Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz………………………………… | 148 |
| Tabla 4.20 | Variables aleatorias independientes con distribución Exponencial , Comparación de los Métodos de Imputación, tamaño de muestra n=50 y 5% de datos faltantes en la matriz………………………………………… | 149 |
| Tabla 4.21 | Matriz de Datos de variables aleatorias dependientes con distribución Normal (10, 1), Tamaño de muestra n=50……. | 153 |
| Tabla 4.22 | Matriz de Datos de variables aleatorias dependientes con distribución Normal (10, 1), tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Matriz de datos con trece filas eliminadas………………………………………… | 155 |
| Tabla 4.23 | Matriz de Datos de variables aleatorias dependientes con distribución Normal (10, 1), Método de Imputación por la Media, tamaño de muestra n=50 y 5% de datos faltantes en la matriz……………………………………………………. | 159 |
| Tabla 4.24 | Matriz de Datos de variables aleatorias dependientes con distribución Normal (10, 1), Método de Imputación por la Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz………………………………………… | 160 |
| Tabla 4.25 | Variables aleatorias dependientes con distribución Normal (10,1), Comparación de los Métodos de Imputación, tamaño de muestra n=50 y 5% de datos faltantes en la matriz………………………………………… | 161 |
| Tabla 4.26 | Matriz de Datos de variables aleatorias dependientes con distribución Poisson , tamaño de muestra n=100….. | 179 |
| Tabla 4.27 | Matriz de Datos de variables aleatorias dependientes con distribución Poisson , tamaño de muestra n=100 y 10% de datos faltantes en la matriz, Matriz de datos con cincuenta filas eliminadas…………………………………… | 182 |
| Tabla 4.28 | Matriz de Datos de variables aleatorias dependientes con distribución Poisson , Método de Imputación por la Media, tamaño de muestra n=100 y 10% de datos faltantes en la matriz………………………………………… | 186 |
| Tabla 4.29 | Matriz de Datos de variables aleatorias dependientes con distribución Poisson , Método de Imputación por Regresión, tamaño de muestra n=100 y 10% de datos faltantes en la matriz………………………………………… | 188 |
| Tabla 4.30 | Variables aleatorias dependientes con distribución Poisson , Comparación de los Métodos de Imputación, tamaño de muestra n=100 y 10% de datos faltantes en la matriz………………………………………… | 190 |
| Tabla 4.31 | Matriz de Datos de variables aleatorias dependientes con distribución Exponencial , Tamaño de muestra n=100 | 197 |
| Tabla 4.32 | Matriz de Datos de variables aleatorias dependientes con distribución Exponencial , tamaño de muestra n=100 y 5% de datos faltantes en la matriz, Matriz de datos con cincuenta filas eliminadas…………………………………… | 200 |
| Tabla 4.33 | Matriz de Datos de variables aleatorias dependientes con distribución Exponencial , Método de Imputación por Media, tamaño de muestra n=100 y 5% de datos faltantes en la matriz………………………………………… | 206 |
| Tabla 4.34 | Matriz de Datos de variables aleatorias dependientes con distribución Exponencial , Método de Imputación por Regresión, tamaño de muestra n=100 y 5% de datos faltantes en la matriz………………………………………… | 208 |
| Tabla 4.35 | Variables aleatorias dependientes con distribución Exponencial , Comparación de los Métodos de Imputación, tamaño de muestra n=100 y 5% de datos faltantes en la matriz………………………………………… | 210 |

**ÍNDICE DE GRÁFICOS**

|  |  |  |
| --- | --- | --- |
| **Capítulo II** |  |  |
| Gráfico 2.1 | Densidad de la Distribución Uniforme……………………… | 27 |
| Gráfico 2.2 | Media de la distribución uniforme………………………….. | 28 |
| Gráfico 2.3 | Números en el intervalo *X Є …………………………..* | 30 |

**ÍNDICE DE CUADROS**

|  |  |  |
| --- | --- | --- |
| **Capítulo I** |  |  |
| Cuadro 1.1 | Matriz de Datos Multivariados……………………………… | 2 |
| Cuadro 1.2 | Variables aleatorias independientes con distribución Poisson , Método de eliminación por Filas, tamaño de muestra n=5, 13% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas………………………… | 20 |
| Cuadro 1.3 | Variables aleatorias independientes con distribución Poisson , Método de eliminación por Pares, tamaño de muestra n=5, 13% de datos faltantes en la matriz, pares de observaciones disponibles para s12 …………….. | 23 |
| Cuadro 1.4 | Variables aleatorias independientes con distribución Poisson , Método de eliminación por Pares, tamaño de muestra n=5, 13% de datos faltantes en la matriz, pares de observaciones disponibles para s13 …………….. | 23 |
| Cuadro 1.5 | Variables aleatorias independientes con distribución Poisson , Método de eliminación por Pares, tamaño de muestra n=5, 13% de datos faltantes en la matriz, pares de observaciones disponibles para s23 …………….. | 24 |

|  |  |  |
| --- | --- | --- |
| **Capítulo II** |  |  |
| Cuadro 2.1 | Contraste de Hipótesis de la Prueba de Bondad de Ajuste………………………………………………………….. | 33 |
| Cuadro 2.2 | Prueba de Bondad de Ajuste……………………………….. | 34 |
| Cuadro 2.3 | Contraste de Hipótesis de la Prueba de Kolmogorov-Smirnov……………………………………………………….. | 36 |
| Cuadro 2.4 | Prueba de Kolmogorov-Smirnov…………………………… | 37 |

|  |  |  |
| --- | --- | --- |
| **Capítulo III** |  |  |
| Cuadro 3.1 | Variables aleatorias independientes con distribución Poisson*,* Método de Imputación por la Media, tamaño de muestra n=10 y 3% de datos faltantes en la matriz*,* Tabla y Diagrama de la “*Variable X4*”……………………………… | 60 |
| Cuadro 3.2 | Variables aleatorias independientes con distribución Poisson*,* Método de Imputación por Media*,* tamaño de muestra n=10 y 3% de datos faltantes en la matriz, matriz de varianzas y covarianzas………………………………… | 62 |
| Cuadro 3.3 | Variables aleatorias independientes con distribución Poisson*,* Método de Imputación por la Media*,* tamaño de muestra n=10 y 5% de datos faltantes en la matriz*,* Tabla y Diagrama de la “*Variable X1*”……………………………… | 65 |
| Cuadro 3.4 | Variables aleatorias independientes con distribución Poisson*,* Método de Imputación por Media*,* tamaño de muestra n=10 y 5% de datos faltantes en la matriz, matriz de varianzas y covarianzas………………………………… | 66 |
| Cuadro 3.5 | Variables aleatorias independientes con distribución Poisson, Método de Imputación por la Media, tamaño de muestra n=10 y 13% de datos faltantes en la matriz, Tablas y Diagramas de las “*Variables X1* y *X3* ”…………… | 69 |
| Cuadro 3.6 | Variables aleatorias independientes con distribución Poisson , Método de Imputación por la Media, tamaño de muestra n=10 y 13% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas………………………… | 71 |
| Cuadro 3.7 | Variables aleatorias dependientes con distribución Normal*,* Método de Imputación por Regresión, tamaño de muestra n=10 y 7% de datos faltantes en la matriz…………………......................................................... | 78 |
| Cuadro 3.8 | Variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión(Variable Dependiente *X*2)……………………………………………… | 80 |
| Cuadro 3.9 | Variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 7% de datos faltantes en la matriz, Imputaciones sucesivas……………………………………… | 84 |
| Cuadro 3.10 | Variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 7% de datos faltantes en la matriz, Matriz de varianzas y covarianzas………………………………….. | 85 |
| Cuadro 3.11 | Variables aleatorias dependientes con distribución Normal*,* Método de Imputación por Regresión (Variable dependiente *X1*)………………………………………………. | 88 |
| Cuadro 3.12 | Variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión (Variable dependiente *X*2)……………………………………………… | 89 |
| Cuadro 3.13 | Variables aleatorias dependientes con distribución Normal*,* Método de Imputación por Regresión (Variable dependiente *X3*)……………………………………………… | 90 |
| Cuadro 3.14 | Variables aleatorias dependientes con distribución Normal , Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, Imputaciones Sucesivas *X*21………………………………… | 97 |
| Cuadro 3.15 | Variables aleatorias dependientes con distribución Normal , Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, Imputaciones Sucesivas *X*22………………………………… | 98 |
| Cuadro 3.16 | Variables aleatorias dependientes con distribución Normal , Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, Imputaciones Sucesivas *X*23………………………………… | 99 |
| Cuadro 3.17 | Variables aleatorias dependientes con distribución Normal, Método de Imputación por Regresión, tamaño de muestra n=10, 10% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas………………………… | 101 |

|  |  |  |
| --- | --- | --- |
| **Capítulo IV** |  |  |
| Cuadro 4.1 | Variables aleatorias independientes con distribución Normal (5,1), Método de Eliminación por Filas, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y Correlaciones……... | 108 |
| Cuadro 4.2 | Variables aleatorias independientes con distribución Normal (5,1), Método de Eliminación por Filas, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X1*”……………………………… | 109 |
| Cuadro 4.3 | Variables aleatorias independientes con distribución Normal (5,1), Método de Imputación por Regresión, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Imputaciones sucesivas *X10,1 ……………………….* | 113 |
| Cuadro 4.4 | Variables aleatorias independientes con distribución Normal (5,1), Método de Imputación por Regresión, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Imputaciones sucesivas *X14,1 ……………………….* | 114 |
| Cuadro 4.5 | Variables aleatorias independientes con distribución Normal (5,1), Método de Imputación por Regresión, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Imputaciones sucesivas *X25,1 ……………………….* | 115 |
| Cuadro 4.6 | Variables aleatorias independientes con distribución Normal (5,1), Método de Imputación por la Media y Regresión, Tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X1*”………………………………………………………………. | 116 |
| Cuadro 4.7 | Variables aleatorias independientes con distribución Normal (5,1), Método de Imputación por la Media y Regresión, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y Correlaciones……………………………………………….. | 118 |
| Cuadro 4.8 | Variables aleatorias independientes con distribución Normal (5,1), Método de Eliminación por Filas, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y Correlaciones……... | 122 |
| Cuadro 4.9 | Variables aleatorias independientes con distribución Normal (5,1), Método de Eliminación por Filas, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X1*” y “*Variable X4*”………………………………………………………………. | 123 |
| Cuadro 4.10 | Variables aleatorias independientes con distribución Normal (5,1), Método de Imputación por la Media y Regresión, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X1*” y “*Variable X4*”…………………………………………….. | 127 |
| Cuadro 4.11 | Variables aleatorias independientes con distribución Normal (5,1), Método de Imputación por la Media y Regresión, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y Correlaciones……………………………………………..… | 129 |
| Cuadro 4.12 | Variables aleatorias independientes con distribución Poisson , Método de Eliminación por Filas, tamaño de muestra n=30 y 2% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y Correlaciones…………………………………………………. | 133 |
| Cuadro 4.13 | Variables aleatorias independientes con distribución Poisson , Método de Eliminación por Filas, tamaño de muestra n=30 y 5% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X5*”……………………………………………………………… | 134 |
| Cuadro 4.14 | Variables aleatorias independientes con distribución Poisson , Método de Imputación por Regresión, tamaño de muestra n=30 y 5% de datos faltantes en la matriz, Imputaciones sucesivas…………………………….. | 138 |
| Cuadro 4.15 | Variables aleatorias independientes con distribución Poisson , Método de Imputación por la Media y Regresión, tamaño de muestra n=30 y 5% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X5*”……………………………………………………………… | 140 |
| Cuadro 4.16 | Variables aleatorias independientes con distribución Poisson , Método de Imputación por la Media y Regresión, tamaño de muestra n=30 y 5% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y Correlaciones……………………………………………….. | 141 |
| Cuadro 4.17 | Variables aleatorias independientes con distribución Exponencial , Método de Eliminación por Filas, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y de Correlaciones…………………………………………………. | 146 |
| Cuadro 4.18 | Variables aleatorias independientes con distribución Exponencial , Método de Imputación por la Media y Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X2*”……………………………………………………………… | 150 |
| Cuadro 4.19 | Variables aleatorias independientes con distribución Exponencial , Método de Imputación por la Media y Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y de Correlaciones……………………………………………. | 151 |
| Cuadro 4.20 | Variables aleatorias dependientes con distribución Normal (10, 1), Método de Eliminación por Filas, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y de Correlaciones…………………………………………………. | 156 |
| Cuadro 4.21 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Eliminación por Filas, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X3*”………………………. | 157 |
| Cuadro 4.22 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X2,3* ………………….. | 162 |
| Cuadro 4.23 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X5,3* ………………….. | 163 |
| Cuadro 4.24 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X6,3* ………………….. | 164 |
| Cuadro 4.25 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X9,3* ………………….. | 165 |
| Cuadro 4.26 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X11,3* ………………… | 166 |
| Cuadro 4.27 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X17,3* ………………… | 167 |
| Cuadro 4.28 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X21,3* ………………… | 168 |
| Cuadro 4.29 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X23,3* ………………… | 169 |
| Cuadro 4.30 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X29,3* ………………… | 170 |
| Cuadro 4.31 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X32,3* ………………… | 171 |
| Cuadro 4.32 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X37,3* ………………… | 172 |
| Cuadro 4.33 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X41,3* ………………… | 173 |
| Cuadro 4.34 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Imputaciones sucesivas para *X46,3* ………………… | 174 |
| Cuadro 4.35 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por la Media y Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz , Tabla y Diagrama de la “*Variable X3*”………………………………………………………………. | 175 |
| Cuadro 4.36 | Variables aleatorias dependientes con distribución Normal (10,1), Método de Imputación por la Media y Regresión, tamaño de muestra n=50 y 5% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y de Correlaciones…………………………………………… | 177 |
| Cuadro 4.37 | Variables aleatorias dependientes con distribución Poisson , Método de Eliminación por Filas, tamaño de muestra n=100 y 10% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y de Correlaciones….. | 183 |
| Cuadro 4.38 | Variables aleatorias dependientes con distribución Poisson , Método de Eliminación por Filas, tamaño de muestra n=100 y 10% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X4*”………………………. | 184 |
| Cuadro 4.39 | Variables aleatorias dependientes con distribución Poisson , Método de Imputación por la Media y Regresión, tamaño de muestra n=100 y 10% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X4*”……………………………………………………………… | 192 |
| Cuadro 4.40 | Variables aleatorias dependientes con distribución Poisson , Método de Imputación por la Media y Regresión, tamaño de muestra n=100 y 10% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas y de Correlaciones…………………………………………… | 194 |
| Cuadro 4.41 | Variables aleatorias dependientes con distribución Exponencial , Método de Eliminación por Filas, tamaño de muestra n=100 y 5% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas………………... | 202 |
| Cuadro 4.42 | Variables aleatorias dependientes con distribución Exponencial , Método de Eliminación por Filas, tamaño de muestra n=100 y 5% de datos faltantes en la matriz, Matriz de Correlaciones…………...………………... | 203 |
| Cuadro 4.43 | Variables aleatorias dependientes con distribución Exponencial , Método de Eliminación por Filas, tamaño de muestra n=100 y 5% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X3*” y “*Variable X8*”………………………………………………………………. | 204 |
| Cuadro 4.44 | Variables aleatorias dependientes con distribución Exponencial , Método de Imputación por la Media y Regresión, tamaño de muestra n=100 y 5% de datos faltantes en la matriz, Tabla y Diagrama de la “*Variable X3*” y “*Variable X8*”…………………………………………….. | 212 |
| Cuadro 4.45 | Variables aleatorias dependientes con distribución Exponencial , Método de Imputación por Media y Regresión, tamaño de muestra n=100 y 5% de datos faltantes en la matriz, Matriz de Varianzas y Covarianzas…………………………………………………… | 215 |
| Cuadro 4.46 | Variables aleatorias dependientes con distribución Exponencial , Método de Imputación por Media y Regresión, tamaño de muestra n=100 y 5% de datos faltantes en la matriz, Matriz de Correlaciones………….... | 216 |

**INTRODUCCIÓN**

La presente tesis tiene como objetivo efectuar en un estudio estadístico acerca de los Efectos de la Imputación en el Análisis de Datos Multivariados, el mismo que se basa en la generación de muestras con variables aleatorias dependientes e independientes de diferentes tamaños y distribuciones, así como también el análisis de un caso real.

El primer capítulo describe los principios estadísticos relacionados con los Métodos de Imputación que son parte de esta investigación, para esto presenta los conceptos relacionados con matrices de datos multivariados, y la “Pérdida de Datos” en una Investigación.

El capítulo dos aborda el tema de las técnicas y principios científicos que permiten la generación de números aleatorios, los mismos que son necesarios para la simulación de sistemas que se explican estocásticamente.

En el capítulo tres se ilustran las técnicas de imputación para el manejo de datos incompletos en una matriz de datos, para lo cual se define “Imputación de Datos” y los “Métodos de Imputación”.

Por otro lado el capítulo cuatro, presenta y analiza los resultados obtenidos al comparar los métodos de imputación utilizando diferentes tamaños de muestras: 30, 50 y 100 así como distintas distribuciones continuas y discretas tales como: normal, poisson y exponencial.

En el último capítulo se muestran las conclusiones y las recomendaciones obtenidas del análisis de los resultados en este estudio.

Referencias Bibliográficas

**[1] Azarang**, **M.** **& García, E** (1996) “***Simulación y Análisis de Modelos Estocásticos”,*** Editorial McGraw-Hill Interamericana Editores, México-México.

**[2] Coss, R.** (1991) ***“Simulación”*,** Un enfoque práctico, Editorial Limusa, México-México.

**[3] Freund, J., Miller, I., Miller, M.** (2000) ***“Estadística Matemática con Aplicaciones”***, Editorial Pearson Educación, México D.F., México.

**[4]** **Martínez, W.; Martínez, A.** (2002) “***Computational Statistics Handbook with Matlab***”, Chapman & Hall/CRC, Boca Raton, United Sates of America.

**[5]** **Mendenhall, W., Wackerly, D., & L-Scheaffer, R.** (2002) ***“Estadística Matemática con aplicaciones”*,** Thomson, Sexta Edición, México-México.

**[6] Rencher**, **A** (1998) ***Multivariate Statistical Inference and Aplications,*** WileySeries in Probability and statistics, New York- United States of America.

**[7]** **Rial, A., Varela, J., & Rojas, A.** (2001) ***“Depuración y Análisis Preliminares de Datos en SPSS”*,** Sistemas Informatizados para la investigación del comportamiento, Edición RA-MA, Madrid-España.

**[8] Pérez, C.** (2000) ***Técnicas de Muestreo Estadístico****,* Teoría y Práctica y Aplicaciones Informáticas, Editorial Alfaomega, Madrid- España.

**[9] Gómez, J. & Palarea, J** (2003) “***Inferencia basada en imputación múltiple en problemas con información incompleta”,***http://www.udc.es/dep/mate/biometria2003/Archivos/ot83.pdf, Fecha de Última Visita: febrero de 2006, Guayaquil-Ecuador.

**[10] Herrero, F. & Cuesta, M** (2004) “***Introducción al Álgebra Matricial***”,http://www.psico.uniovi.es/Dpto\_Psicologia/metodos/tutor.3/vector.html, Fecha de Última Visita: marzo de 2006, Guayaquil-Ecuador.

**[11] Kennedy, W & Gentle, J. “*Generación de números aleatorios*”** http://math.uprm.edu/~edgar/LEC9COMP.PDF, Fecha de Última Visita: abril de 2006, Guayaquil-Ecuador.

**[12] López, V.** (2005) **“*Comparación de los métodos de imputación con respecto al poder de separación del modelo de regresión logística*”**, http://grad.uprm.edu/tesis/lopezvazquez.pdf,Fecha de Última Visita: febrero de 2006, Guayaquil-Ecuador.

**[13] Tarifa, E.** (2002) ***“Teoría de Modelos y Simulación”*** http://www.modeladoeningenieria.edu.ar/unj/tms/apuntes/cp3.pdf, Fecha de Última Visita: marzo de 2006, Guayaquil-Ecuador.