CAPÍTULO IV

**4. SIMULACIÓN BAJO DISTINTAS CONDICIONES UNIVARIADAS Y MULTIVARIADAS**

* 1. **Introducción**

En el presente capítulo se presentan y analizan 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. El análisis se lo realiza para variables aleatorias conjuntas dependientes e independientes. Para la generación de las variables aleatorias se utiliza el programa Matlab 6.5 el cual provee de los comandos adecuados para la realización de esta tarea.

Se escogieron tamaños de muestra de 30, 50 y 100, puesto que en primera instancia se realizó simulaciones con tamaños de muestra *n*=10, de los cuales no se pudo obtener resultados dignos de comentario.

En la sección 4.2 se presentan simulaciones para distribuciones normal, poisson y exponencial, idénticamente distribuìdas e independientes, mientras que en la sección 4.3 se presentan distribuciones con variables aleatorias dependientes. Para la utilización del Método de Imputación por Regresión se desarrolló un algoritmo en Matlab 6.5 (Ver Anexo 2).

**4.2 Matrices de Datos con variables aleatorias independientes**

**4.2.1 Distribución Normal: *Tres datos faltantes* en una sola variable (2% de la matriz), tamaño de muestra n=30**

Se tiene una matriz de datos cuyas columnas son muestras tomadas de cinco poblaciones todas ellas Normal, independientes e idénticamente distribuidas, con parámetros =5 y =1, , *i= 1,2,....30*  y *j= 1,2,3*,*4,5* y se supone que tiene el 2% de datos faltantes, es decir tres datos, los que recayeron en la variable *X*1 y son: el *X*10,1=4.168, *X*14,1=6.624 y el *X*25,1=6.290. Nótese que el 2% de datos faltantes en la matriz, constituye 10% de datos faltantes en la columna que corresponde a *X*1 (Ver Tabla 4.1).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.1**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias independientes con distribución Normal (5, 1)**  Tamaño de muestra n=30 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 4.813 | 3.396 | 5.569 | 3.812 | 5.806 |
| 5.726 | 5.257 | 4.744 | 2.798 | 5.232 |
| 4.412 | 3.944 | 4.623 | 5.986 | 4.010 |
| 7.183 | 6.415 | 4.704 | 4.481 | 6.340 |
| 4.864 | 4.195 | 3.525 | 5.327 | 5.290 |
| 5.114 | 5.529 | 4.766 | 5.234 | 6.479 |
| 6.067 | 5.219 | 5.118 | 5.022 | 6.138 |
| 5.059 | 4.078 | 5.315 | 3.996 | 4.316 |
| 4.904 | 2.829 | 6.444 | 4.053 | 3.708 |
| **4.168** | 4.941 | 4.649 | 4.626 | 4.927 |
| 5.294 | 3.989 | 5.623 | 3.814 | 4.669 |
| 3.664 | 5.615 | 5.799 | 3.944 | 4.156 |
| 5.714 | 5.508 | 5.941 | 6.473 | 5.498 |
| **6.624** | 6.692 | 4.008 | 5.056 | 6.489 |
| 4.308 | 5.591 | 5.212 | 3.783 | 4.454 |
| 5.858 | 4.356 | 5.238 | 4.959 | 4.153 |
| 6.254 | 5.380 | 3.992 | 3.872 | 4.754 |
| 3.406 | 3.991 | 4.258 | 3.651 | 5.663 |
| 3.559 | 4.981 | 6.082 | 4.739 | 4.146 |
| 5.571 | 4.952 | 4.869 | 5.954 | 3.799 |
| 4.600 | 5.000 | 5.390 | 5.129 | 4.880 |
| 5.690 | 4.682 | 5.088 | 5.657 | 4.935 |
| 5.816 | 6.095 | 4.365 | 3.832 | 5.485 |
| 5.712 | 3.126 | 4.440 | 4.539 | 4.405 |
| **6.290** | 5.428 | 5.444 | 4.738 | 4.850 |
| 5.669 | 5.896 | 4.050 | 3.787 | 4.565 |
| 6.191 | 5.731 | 5.781 | 3.681 | 4.921 |
| 3.798 | 5.578 | 5.569 | 5.931 | 6.535 |
| 4.980 | 5.040 | 4.178 | 5.011 | 4.394 |
| 4.843 | 5.677 | 4.734 | 4.355 | 3.653 |

**Elaborado por**: G. Cuenca

El vector de medias de los datos originales es:



**Método de Eliminación por Filas**

Como detallamos en el Capítulo 1, el Método de Eliminación por Filas no toma en cuenta las filas donde se encuentren datos faltantes y como los datos faltantes recayeron en la variable *X*1 y son: el *X*10,1=4.168, *X*14,1=6.624 y el *X*25,1=6.290, se procede a prescindir de las filas diez, catorce y veinte y cinco. La matriz resultante con filas eliminadas se muestra en la Tabla 4.2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.2**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 4.813 | 3.396 | 5.569 | 3.812 | 5.806 |
| 5.726 | 5.257 | 4.744 | 2.798 | 5.232 |
| 4.412 | 3.944 | 4.623 | 5.986 | 4.010 |
| 7.183 | 6.415 | 4.704 | 4.481 | 6.340 |
| 4.864 | 4.195 | 3.525 | 5.327 | 5.290 |
| 5.114 | 5.529 | 4.766 | 5.234 | 6.479 |
| 6.067 | 5.219 | 5.118 | 5.022 | 6.138 |
| 5.059 | 4.078 | 5.315 | 3.996 | 4.316 |
| 4.904 | 2.829 | 6.444 | 4.053 | 3.708 |
| 5.294 | 3.989 | 5.623 | 3.814 | 4.669 |
| 3.664 | 5.615 | 5.799 | 3.944 | 4.156 |
| 5.714 | 5.508 | 5.941 | 6.473 | 5.498 |
| 4.308 | 5.591 | 5.212 | 3.783 | 4.454 |
| 5.858 | 4.356 | 5.238 | 4.959 | 4.153 |
| 6.254 | 5.380 | 3.992 | 3.872 | 4.754 |
| 3.406 | 3.991 | 4.258 | 3.651 | 5.663 |
| 3.559 | 4.981 | 6.082 | 4.739 | 4.146 |
| 5.571 | 4.952 | 4.869 | 5.954 | 3.799 |
| 4.600 | 5.000 | 5.390 | 5.129 | 4.880 |
| 5.690 | 4.682 | 5.088 | 5.657 | 4.935 |
| 5.816 | 6.095 | 4.365 | 3.832 | 5.485 |
| 5.712 | 3.126 | 4.440 | 4.539 | 4.405 |
| 5.669 | 5.896 | 4.050 | 3.787 | 4.565 |
| 6.191 | 5.731 | 5.781 | 3.681 | 4.921 |
| 3.798 | 5.578 | 5.569 | 5.931 | 6.535 |
| 4.980 | 5.040 | 4.178 | 5.011 | 4.394 |
| 4.843 | 5.677 | 4.734 | 4.355 | 3.653 |

**Elaborado por:** G. Cuenca

Nótese que la eliminación por filas, equivale a prescindir de todos los datos de los informantes porque no respondieron, por ejemplo, una pregunta.

El vector de medias para las veintisiete filas restantes es:



Como era de esperarse el vector de medias de los datos originales y de los datos con filas eliminadas no coinciden.

Ahora analicemos en el Cuadro 4.1 el efecto que causa en la *matriz de varianzas y covarianzas*, y *matriz de correlaciones*, la eliminación de tres filas, es decir la diez, la catorce y la veinticinco, con un tamaño de muestra *n*=30.

Se puede notar que la mayoría de las covarianzas entre las variables tanto en la matriz de datos originales como en la matriz con tres filas eliminadas son cercanas a cero, lo cual era de esperarse dado que las columnas son muestras tomadas de poblaciones independientes.

|  |  |
| --- | --- |
| **CUADRO 4.1**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.891 |  |  |  |  | | *X*2 | 0.299 | 0.891 |  |  |  | | *X*3 | -0.152 | -0.138 | 0.502 |  |  | | *X*4 | -0.010 | 0.034 | 0.014 | 0.756 |  | | *X*5 | 0.197 | 0.315 | -0.123 | 0.090 | 0.740 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.335 | 1.000 |  |  |  | | *X*3 | -0.227 | -0.206 | 1.000 |  |  | | *X*4 | -0.012 | 0.042 | 0.023 | 1.000 |  | | *X*5 | 0.242 | **0.388** | -0.202 | 0.120 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **(3 Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.827 |  |  |  |  | | *X*2 | 0.214 | 0.866 |  |  |  | | *X*3 | -0.147 | -0.095 | 0.510 |  |  | | *X*4 | -0.041 | 0.004 | 0.031 | 0.835 |  | | *X*5 | 0.136 | **0.247** | -0.077 | 0.073 | 0.732 | | **Matriz de Correlaciones**  **(3 Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.253 | 1.000 |  |  |  | | *X*3 | -0.226 | -0.143 | 1.000 |  |  | | *X*4 | -0.050 | 0.005 | 0.048 | 1.000 |  | | *X*5 | 0.175 | **0.311** | -0.125 | 0.094 | 1.000 | |

**Elaborado por:** G. Cuenca

La mayor covarianza en la matriz de datos originales se da entre las variables *X2*  y *X5* y es 0.315; mientras que en la matriz con tres filas eliminadas este valor disminuye a 0.247.

En la matriz de correlaciones de datos originales, la mayor correlación se da entre las variables *X2* y *X5*, y es 0.388, la que disminuye a 0.311 en la matriz de correlaciones con tres filas eliminadas.

En el Cuadro 4.2, podemos apreciar que con el 10% de datos eliminadas en la primera columna (Variable *X****1*)**, el valor de la varianza disminuyó de 0.891 a 0.827.

|  |  |
| --- | --- |
| **CUADRO 4.2**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****1*”** | |
| **Estimadores**   |  |  |  |  | | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Con el 10% de datos eliminadas en X1** | | n | | 30 | 27 | | Media | | 5,205 | 5,151 | | Mediana | | 5,205 | 5,114 | | Moda | | 3,410 | 3,410 | | Varianza | | 0,891 | 0,827 | | Desviación Estándar | | 0,944 | 0,909 | | Error Estándar | | 0,172 | 0,175 | | **Coeficiente de Asimetría** | | -0,161 | -0,176 | | Curtosis | | -0,460 | -0,105 | | Rango | | 3,780 | 3,780 | | Mínimo | | 3,410 | 3,410 | | Máximo | | 7,180 | 7,180 | | Percentiles | 25 | 4,553 | 4,600 | | 50 | 5,204 | 5,114 | | 75 | 5,827 | 5,726 | | **Diagrama de Cajas** |

**Elaborado por:** G. Cuenca

**Método de Imputación por la Media y Regresión**

A continuación se aplica el método de imputación por media y regresión a la misma matriz de datos originales, con los mismos datos faltantes, utilizada en el método de eliminación por filas.

Por medio del Método de *Imputación por la Media*, se procede a calcular la media aritmética de la variable *X1* con los tres datos faltantes, cuyo valor es 5.151, entonces reemplazamos en *X*10,1, *X*14,1 y en *X*25,1. La matriz de datos resultante con tres valores completados por imputación por la media en la variable *X1* se muestra en la Tabla 4.3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.3**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 4.813 | 3.396 | 5.569 | 3.812 | 5.806 |
| 5.726 | 5.257 | 4.744 | 2.798 | 5.232 |
| 4.412 | 3.944 | 4.623 | 5.986 | 4.010 |
| 7.183 | 6.415 | 4.704 | 4.481 | 6.340 |
| 4.864 | 4.195 | 3.525 | 5.327 | 5.290 |
| 5.114 | 5.529 | 4.766 | 5.234 | 6.479 |
| 6.067 | 5.219 | 5.118 | 5.022 | 6.138 |
| 5.059 | 4.078 | 5.315 | 3.996 | 4.316 |
| 4.904 | 2.829 | 6.444 | 4.053 | 3.708 |
| **5.151** | 4.941 | 4.649 | 4.626 | 4.927 |
| 5.294 | 3.989 | 5.623 | 3.814 | 4.669 |
| 3.664 | 5.615 | 5.799 | 3.944 | 4.156 |
| 5.714 | 5.508 | 5.941 | 6.473 | 5.498 |
| **5.151** | 6.692 | 4.008 | 5.056 | 6.489 |
| 4.308 | 5.591 | 5.212 | 3.783 | 4.454 |
| 5.858 | 4.356 | 5.238 | 4.959 | 4.153 |
| 6.254 | 5.380 | 3.992 | 3.872 | 4.754 |
| 3.406 | 3.991 | 4.258 | 3.651 | 5.663 |
| 3.559 | 4.981 | 6.082 | 4.739 | 4.146 |
| 5.571 | 4.952 | 4.869 | 5.954 | 3.799 |
| 4.600 | 5.000 | 5.390 | 5.129 | 4.880 |
| 5.690 | 4.682 | 5.088 | 5.657 | 4.935 |
| 5.816 | 6.095 | 4.365 | 3.832 | 5.485 |
| 5.712 | 3.126 | 4.440 | 4.539 | 4.405 |
| **5.151** | 5.428 | 5.444 | 4.738 | 4.850 |
| 5.669 | 5.896 | 4.050 | 3.787 | 4.565 |
| 6.191 | 5.731 | 5.781 | 3.681 | 4.921 |
| 3.798 | 5.578 | 5.569 | 5.931 | 6.535 |
| 4.980 | 5.040 | 4.178 | 5.011 | 4.394 |
| 4.843 | 5.677 | 4.734 | 4.355 | 3.653 |

**Elaborado por:** G. Cuenca

Por medio del Método de *Imputación por Regresión*, el cálculo de los valores faltantes se realiza por medio de la ecuación de predicción  y el cálculo de los coeficientes de la misma es de la forma . La matriz de datos resultante, con tres valores completados por imputación utilizando regresión en la variable *X1* se puede ver en la Tabla 4.4.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.4**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 4.813 | 3.396 | 5.569 | 3.812 | 5.806 |
| 5.726 | 5.257 | 4.744 | 2.798 | 5.232 |
| 4.412 | 3.944 | 4.623 | 5.986 | 4.010 |
| 7.183 | 6.415 | 4.704 | 4.481 | 6.340 |
| 4.864 | 4.195 | 3.525 | 5.327 | 5.290 |
| 5.114 | 5.529 | 4.766 | 5.234 | 6.479 |
| 6.067 | 5.219 | 5.118 | 5.022 | 6.138 |
| 5.059 | 4.078 | 5.315 | 3.996 | 4.316 |
| 4.904 | 2.829 | 6.444 | 4.053 | 3.708 |
| **5.294** | 4.941 | 4.649 | 4.626 | 4.927 |
| 5.294 | 3.989 | 5.623 | 3.814 | 4.669 |
| 3.664 | 5.615 | 5.799 | 3.944 | 4.156 |
| 5.714 | 5.508 | 5.941 | 6.473 | 5.498 |
| **5.714** | 6.692 | 4.008 | 5.056 | 6.489 |
| 4.308 | 5.591 | 5.212 | 3.783 | 4.454 |
| 5.858 | 4.356 | 5.238 | 4.959 | 4.153 |
| 6.254 | 5.380 | 3.992 | 3.872 | 4.754 |
| 3.406 | 3.991 | 4.258 | 3.651 | 5.663 |
| 3.559 | 4.981 | 6.082 | 4.739 | 4.146 |
| 5.571 | 4.952 | 4.869 | 5.954 | 3.799 |
| 4.600 | 5.000 | 5.390 | 5.129 | 4.880 |
| 5.690 | 4.682 | 5.088 | 5.657 | 4.935 |
| 5.816 | 6.095 | 4.365 | 3.832 | 5.485 |
| 5.712 | 3.126 | 4.440 | 4.539 | 4.405 |
| **5.726** | 5.428 | 5.444 | 4.738 | 4.850 |
| 5.669 | 5.896 | 4.050 | 3.787 | 4.565 |
| 6.191 | 5.731 | 5.781 | 3.681 | 4.921 |
| 3.798 | 5.578 | 5.569 | 5.931 | 6.535 |
| 4.980 | 5.040 | 4.178 | 5.011 | 4.394 |
| 4.843 | 5.677 | 4.734 | 4.355 | 3.653 |

**Elaborado por:** G. Cuenca

En la Tabla 4.5 se realiza una comparación entre el valor real y el valor con imputación por la media y regresión.

|  |
| --- |
| **Tabla 4.5**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 |
| ***10% de datos completados en*** *X1* ***por la Media***   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Imputación por**  **la Media** | **Error**  **| Dato Observado –**  **Dato con Imputación |** | | 4.168 | 5.151 | 0.983 | | 6.624 | 5.151 | 1.473 | | 6.290 | 5.151 | 1.139 | |
| ***10% de datos completados en*** *X1* ***por Regresión***   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Imputación por**  **Regresión** | **Error**  **| Dato Observado –**  **Dato con Imputación |** | | 4.168 | 5.245 | 1.077 | | 6.624 | 5.871 | 0.753 | | 6.290 | 5.726 | 0.564 | |

**Elaborado por:** G. Cuenca

La diferencia en valor absoluto entre el dato observado de cada variable es menor en el “*Método de Imputación por Regresión*”, con excepción del primer valor donde error por medio del Método de Imputación por Media es menor (0.983).

En los Cuadros 4.3, 4.4 y 4.5, podemos apreciar el número de imputaciones sucesivas por medio del *Método de Regresión* que se realiza a los tres datos faltantes en la variable *X1.*

|  |  |
| --- | --- |
| **CUADRO 4.3**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 para** *X*10,1=4.168   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 5.352 | 1.184 | | 2 | 5.310 | 1.142 | | 3 | 5.286 | 1.118 | | 4 | 5.263 | 1.095 | | 5 | 5.251 | 1.083 | | 6 | 5.245 | 1.077 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 6 | | Media | 5.285 | | Error Estándar | 0.017 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 6 | | Media | 1.117 | | Error Estándar | 0.017 | |

**Elaborado por:** G. Cuenca

En el Cuadro 4.3, se puede ver que el primer resultado de predicción es 5.352 ± 0.017, y el último es 5.245 ± 0.017, donde la media de los resultados de predicción es 5.285 ± 0.017.

|  |  |
| --- | --- |
| **CUADRO 4.4**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 para** *X*14,1=6.629   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 5.322 | 1.307 | | 2 | 5.464 | 1.165 | | 3 | 5.583 | 1.046 | | 4 | 5.674 | 0.955 | | 5 | 5.725 | 0.904 | | 6 | 5.871 | 0.758 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 6 | | Media | 5.607 | | Error Estándar | 0.080 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 6 | | Media | 1.023 | | Error Estándar | 0.080 | |

**Elaborado por:** G. Cuenca

En el Cuadro 4.4, se puede ver que el primer resultado de predicción es 5.322 ± 0.080, y el último es 5.871 ± 0.080, donde la media de los resultados de predicción es 5.607 ± 0.080. Mientras que la media del error de predicción es 1.023 ± 0.080.

|  |  |
| --- | --- |
| **CUADRO 4.5**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 para** *X*25,1=6.290   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 5.273 | 1.017 | | 2 | 5.321 | 0.969 | | 3 | 5.492 | 0.798 | | 4 | 5.545 | 0.745 | | 5 | 5.673 | 0.617 | | 6 | 5.726 | 0.564 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 6 | | Media | 5.505 | | Error Estándar | 0.075 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 6 | | Media | 0.785 | | Error Estándar | 0.075 | |

**Elaborado por:** G. Cuenca

En el Cuadro 4.5, se puede observar que el resultado de predicción tiene una media de 5.505 ± 0.075. Se nota también que, en general las imputaciones sucesivas a los tres datos faltantes no tienden al valor observado.

**Elaborado por:** G. Cuenca

|  |  |
| --- | --- |
| **CUADRO 4.6**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****1*”** | |
| **Estimadores**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 30 | 27 | 30 | 30 | | Media | | 5,205 | 5,151 | 5,151 | 5,193 | | Mediana | | 5,205 | 5,114 | 5,151 | 5,294 | | Moda | | 3,406 | 3,410 | 5,151 | 5,294 | | Varianza | | 0,891 | 0,827 | 0,741 | 0,763 | | Desviación Estándar | | 0,944 | 0,909 | 0,861 | 0,873 | | Error Estándar | | 0,172 | 0,175 | 0,157 | 0,159 | | **Coeficiente de Asimetría** | | -0,161 | -0,176 | -0,184 | -0,314 | | Curtosis | | -0,460 | -0,105 | 0,229 | 0,102 | | Rango | | 3,777 | 3,780 | 3,777 | 3,777 | | Mínimo | | 3,406 | 3,410 | 3,406 | 3,406 | | Máximo | | 7,183 | 7,180 | 7,183 | 7,183 | | Percentiles | 25 | 4,553 | 4,600 | 4,759 | 4,759 | | 50 | 5,204 | 5,114 | 5,151 | 5,294 | | 75 | 5,827 | 5,726 | 5,717 | 5,726 | | **Diagrama de Cajas** |

Al realizar la imputación por los métodos de “media” y “regresión” se obtuvieron los siguientes resultados (Ver Cuadro 4.6)

El valor de la media de los “datos completados” por *la media* disminuye. comparándolo con los “datos originales” y “datos completados” por *regresión.*

El valor de la varianza de los datos completados por la media disminuye de 0.891 a 0.741, mientras que en los datos completados por regresión este valor se incrementa a 0.763, comparándolo con el valor anterior.

El vector de medias con tres datos completados por la media en *X1* es:



Mientras que el vector de medias con tres datos completados por la regresión en *X1* es:



El efecto que causa en la *matriz de varianzas y covarianzas* y *matriz de correlaciones*, el completar 2% de datos faltantes en una matriz de tamaño 30, por medio de la *imputación por media y regresión*, se presenta en el Cuadro 4.7.

|  |  |
| --- | --- |
| **CUADRO 4.7**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.891 |  |  |  |  | | *X*2 | 0.299 | 0.891 |  |  |  | | *X*3 | -0.152 | -0.138 | 0.502 |  |  | | *X*4 | -0.010 | 0.034 | 0.014 | 0.756 |  | | *X*5 | 0.197 | 0.315 | -0.123 | 0.090 | 0.740 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.335 | 1.000 |  |  |  | | *X*3 | -0.227 | -0.206 | 1.000 |  |  | | *X*4 | -0.012 | 0.042 | 0.023 | 1.000 |  | | *X*5 | 0.242 | **0.388** | -0.202 | 0.120 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **10% Datos Completados por Media en “*Variable*** *X****1*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | **0.741** |  |  |  |  | | *X*2 | **0.192** | 0.891 |  |  |  | | *X*3 | **-0.132** | -0.138 | 0.502 |  |  | | *X*4 | **-0.037** | 0.034 | 0.014 | 0.756 |  | | *X*5 | **0.122** | 0.315 | -0.123 | 0.090 | 0.740 | | **Matriz de Correlaciones**  **10% Datos Completados por Media en “*Variable*** *X****1*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | **0.236** | 1.000 |  |  |  | | *X*3 | **-0.215** | -0.206 | 1.000 |  |  | | *X*4 | **-0.049** | 0.042 | 0.023 | 1.000 |  | | *X*5 | **0.164** | 0.388 | -0.202 | 0.120 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **10% Datos Completados por Regresión en “*Variable*** *X****1*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | **0.763** |  |  |  |  | | *X*2 | **0.235** | 0.891 |  |  |  | | *X*3 | **-0.143** | -0.138 | 0.502 |  |  | | *X*4 | **-0.026** | 0.034 | 0.014 | 0.756 |  | | *X*5 | **0.149** | 0.315 | -0.123 | 0.090 | 0.740 | | **Matriz de Correlaciones**  **10% Datos Completados por Regresiòn en “*Variable*** *X****1*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | **0.285** | 1.000 |  |  |  | | *X*3 | **-0.231** | -0.206 | 1.000 |  |  | | *X*4 | **-0.034** | 0.042 | 0.023 | 1.000 |  | | *X*5 | **0.199** | 0.388 | -0.202 | 0.120 | 1.000 | |

**Elaborado por:** G. Cuenca

Se puede apreciar que los únicos valores que cambian son las covarianzas de la variable *X1* con las demás variables, donde la covarianza entre *X1*  y *X2* disminuye de 0.299 a 0.192 en la matriz con 10% de datos completados por la media en la variable *X1*, mientras que la covarianza entre*X1* y *X4*  se incrementa en valor absoluto de 0.010 a 0.037.

En la matriz de varianzas y covarianzas de los datos completados por regresión, el valor de las covarianzas de variable *X1* con las demás variables se incrementa, comparándolo con la matriz de varianzas y covarianzas de los datos completados por la media.

Por otro lado, analizando el efecto que causa en la matriz de correlaciones, se nota que la mayor correlación se da entre las variables *X2*  y *X5* , es decir 0.388, seguida por 0.335 entre las variables *X1*  y *X2*. En la matriz de correlaciones con 10% de datos completados por la media, la correlación entre *X1*  y *X2* disminuye a 0.236, mientras que en la matriz de datos completados por regresión ésta tiene un ligero incremento a 0.285. Se puede apreciar también que en general las variables no están fuertemente correlacionadas entre sí.

**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**

Continuando con la matriz de datos anterior, pero ahora se tienen dos datos faltantes en la variable *X*1 y *uno* en la variable *X*4 , datos cuyas columnas son muestras tomadas de cinco poblaciones todas ellas Normal, independientes e idénticamente distribuidas, con parámetros =5 y =1. Los datos faltantes son los siguientes: *X*10,1=4.168. *X*18,4=3.651 y el *X*24,1=5.712. Nótese que el 2% de datos faltantes en la matriz, constituye 7% de datos faltantes en la columna correspondiente a *X*1 y 3% de datos faltantes en la columna *X*4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.6**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias independientes con distribución Normal (5, 1)**  Tamaño de muestra n=30 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 4.813 | 3.396 | 5.569 | 3.812 | 5.806 |
| 5.726 | 5.257 | 4.744 | 2.798 | 5.232 |
| 4.412 | 3.944 | 4.623 | 5.986 | 4.010 |
| 7.183 | 6.415 | 4.704 | 4.481 | 6.340 |
| 4.864 | 4.195 | 3.525 | 5.327 | 5.290 |
| 5.114 | 5.529 | 4.766 | 5.234 | 6.479 |
| 6.067 | 5.219 | **5.118** | 5.022 | 6.138 |
| 5.059 | 4.078 | 5.315 | 3.996 | 4.316 |
| 4.904 | 2.829 | 6.444 | 4.053 | 3.708 |
| **4.168** | 4.941 | 4.649 | 4.626 | 4.927 |
| 5.294 | 3.989 | 5.623 | 3.814 | 4.669 |
| 3.664 | 5.615 | 5.799 | 3.944 | 4.156 |
| 5.714 | 5.508 | 5.941 | 6.473 | 5.498 |
| 6.624 | 6.692 | 4.008 | 5.056 | 6.489 |
| 4.308 | 5.591 | 5.212 | 3.783 | 4.454 |
| 5.858 | 4.356 | 5.238 | 4.959 | 4.153 |
| 6.254 | 5.380 | 3.992 | 3.872 | 4.754 |
| 3.406 | 3.991 | 4.258 | **3.651** | 5.663 |
| 3.559 | 4.981 | 6.082 | 4.739 | 4.146 |
| 5.571 | 4.952 | 4.869 | 5.954 | 3.799 |
| 4.600 | 5.000 | 5.390 | 5.129 | 4.880 |
| 5.690 | 4.682 | 5.088 | 5.657 | 4.935 |
| 5.816 | 6.095 | 4.365 | 3.832 | 5.485 |
| **5.712** | 3.126 | 4.440 | 4.539 | 4.405 |
| 6.290 | 5.428 | 5.444 | 4.738 | 4.850 |
| 5.669 | 5.896 | 4.050 | 3.787 | 4.565 |
| 6.191 | 5.731 | 5.781 | 3.681 | 4.921 |
| 3.798 | 5.578 | 5.569 | 5.931 | 6.535 |
| 4.980 | 5.040 | 4.178 | 5.011 | 4.394 |
| 4.843 | 5.677 | 4.734 | 4.355 | 3.653 |

**Elaborado por:** G. Cuenca

El vector de medias de los datos originales es:



**Método de Eliminación por Filas**

Como los datos faltantes recayeron en las variables *X*1 y *X*4, se procede a prescindir de las filas diez, dieciocho y veinticuatro. La matriz de datos resultante con filas eliminadas se muestra en la Tabla 4.7.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.7**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 4.813 | 3.396 | 5.569 | 3.812 | 5.806 |
| 5.726 | 5.257 | 4.744 | 2.798 | 5.232 |
| 4.412 | 3.944 | 4.623 | 5.986 | 4.010 |
| 7.183 | 6.415 | 4.704 | 4.481 | 6.340 |
| 4.864 | 4.195 | 3.525 | 5.327 | 5.290 |
| 5.114 | 5.529 | 4.766 | 5.234 | 6.479 |
| 6.067 | 5.219 | **5.118** | 5.022 | 6.138 |
| 5.059 | 4.078 | 5.315 | 3.996 | 4.316 |
| 4.904 | 2.829 | 6.444 | 4.053 | 3.708 |
| 5.294 | 3.989 | 5.623 | 3.814 | 4.669 |
| 3.664 | 5.615 | 5.799 | 3.944 | 4.156 |
| 5.714 | 5.508 | 5.941 | 6.473 | 5.498 |
| 6.624 | 6.692 | 4.008 | 5.056 | 6.489 |
| 4.308 | 5.591 | 5.212 | 3.783 | 4.454 |
| 5.858 | 4.356 | 5.238 | 4.959 | 4.153 |
| 6.254 | 5.380 | 3.992 | 3.872 | 4.754 |
| 3.559 | 4.981 | 6.082 | 4.739 | 4.146 |
| 5.571 | 4.952 | 4.869 | 5.954 | 3.799 |
| 4.600 | 5.000 | 5.390 | 5.129 | 4.880 |
| 5.690 | 4.682 | 5.088 | 5.657 | 4.935 |
| 5.816 | 6.095 | 4.365 | 3.832 | 5.485 |
| 6.290 | 5.428 | 5.444 | 4.738 | 4.850 |
| 5.669 | 5.896 | 4.050 | 3.787 | 4.565 |
| 6.191 | 5.731 | 5.781 | 3.681 | 4.921 |
| 3.798 | 5.578 | 5.569 | 5.931 | 6.535 |
| 4.980 | 5.040 | 4.178 | 5.011 | 4.394 |
| 4.843 | 5.677 | 4.734 | 4.355 | 3.653 |

**Elaborado por:** G. Cuenca

El vector de medias para las veintisiete filas restantes es:



|  |  |
| --- | --- |
| **CUADRO 4.8**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.891 |  |  |  |  | | *X*2 | 0.299 | 0.891 |  |  |  | | *X*3 | -0.152 | -0.138 | 0.502 |  |  | | *X*4 | -0.010 | 0.034 | 0.014 | 0.756 |  | | *X*5 | 0.197 | 0.315 | -0.123 | 0.090 | 0.740 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.335 | 1.000 |  |  |  | | *X*3 | -0.227 | -0.206 | 1.000 |  |  | | *X*4 | -0.012 | 0.042 | 0.023 | 1.000 |  | | *X*5 | 0.242 | **0.388** | -0.202 | 0.120 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **(3 Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.811 |  |  |  |  | | *X*2 | 0.291 | 0.815 |  |  |  | | *X*3 | -0.227 | -0.226 | 0.521 |  |  | | *X*4 | -0.078 | -0.007 | -0.014 | 0.807 |  | | *X*5 | 0.278 | **0.339** | -0.129 | 0.125 | 0.795 | | **Matriz de Correlaciones**  **(3 Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.358 | 1.000 |  |  |  | | *X*3 | -0.350 | -0.348 | 1.000 |  |  | | *X*4 | -0.097 | -0.009 | -0.022 | 1.000 |  | | *X*5 | 0.347 | 0.422 | -0.201 | 0.156 | 1.000 | |

**Elaborado por:** G. Cuenca

|  |  |
| --- | --- |
| **CUADRO 4.9**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****1*” y “*Variable*** *X****4*”** | |
| **Estimadores “*Variable*** *X****1*”**   |  |  |  |  | | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Con el 7% de datos eliminadas en X1** | | n | | 30 | 27 | | Media | | 5,205 | 5,291 | | Mediana | | 5,205 | 5,294 | | Moda | | 3,410 | 3,559 | | Varianza | | 0,891 | 0,811 | | Desviación Estándar | | 0,944 | 0,900 | | Error Estándar | | 0,172 | 0,173 | | **Coeficiente de Asimetría** | | -0,161 | -0,133 | | Curtosis | | -0,460 | -0,279 | | Rango | | 3,780 | 3,624 | | Mínimo | | 3,410 | 3,559 | | Máximo | | 7,180 | 7,183 | | Percentiles | 25 | 4,553 | 4,813 | | 50 | 5,204 | 5,294 | | 75 | 5,827 | 5,858 | | **Diagrama de Cajas “*Variable*** *X****1*”** |
| **Estimadores “*Variable*** *X****4*”**   |  |  |  |  | | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Con el 3% de datos eliminadas en X4** | | n | | 30 | 27 | | Media | | 4,608 | 4,645 | | Mediana | | 4,583 | 4,738 | | Moda | | 2,800 | 2,800 | | Varianza | | 0,756 | 0,807 | | Desviación Estándar | | 0,870 | 0,898 | | Error Estándar | | 0,159 | 0,173 | | **Coeficiente de Asimetría** | | 0,266 | 0,187 | | Curtosis | | -0,447 | -0,589 | | Rango | | 3,680 | 3,680 | | Mínimo | | 2,800 | 2,800 | | Máximo | | 6,470 | 6,470 | | Percentiles | 25 | 3,828 | 3,832 | | 50 | 4,583 | 4,738 | | 75 | 5,155 | 5,234 | | **Diagrama de Cajas “*Variable*** *X****4*”** |

**Elaborado por:** G. Cuenca

**Método de Imputación por la Media y Regresión**

La matriz de datos resultante con dos valores completados por imputación por la media en la variable *X1* y uno en *X4,* se muestra en la Tabla 4.8.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.8**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 4.813 | 3.396 | 5.569 | 3.812 | 5.806 |
| 5.726 | 5.257 | 4.744 | 2.798 | 5.232 |
| 4.412 | 3.944 | 4.623 | 5.986 | 4.010 |
| 7.183 | 6.415 | 4.704 | 4.481 | 6.340 |
| 4.864 | 4.195 | 3.525 | 5.327 | 5.290 |
| 5.114 | 5.529 | 4.766 | 5.234 | 6.479 |
| 6.067 | 5.219 | **5.118** | 5.022 | 6.138 |
| 5.059 | 4.078 | 5.315 | 3.996 | 4.316 |
| 4.904 | 2.829 | 6.444 | 4.053 | 3.708 |
| **5.196** | 4.941 | 4.649 | 4.626 | 4.927 |
| 5.294 | 3.989 | 5.623 | 3.814 | 4.669 |
| 3.664 | 5.615 | 5.799 | 3.944 | 4.156 |
| 5.714 | 5.508 | 5.941 | 6.473 | 5.498 |
| 6.624 | 6.692 | 4.008 | 5.056 | 6.489 |
| 4.308 | 5.591 | 5.212 | 3.783 | 4.454 |
| 5.858 | 4.356 | 5.238 | 4.959 | 4.153 |
| 6.254 | 5.380 | 3.992 | 3.872 | 4.754 |
| 3.406 | 3.991 | 4.258 | **4.641** | 5.663 |
| 3.559 | 4.981 | 6.082 | 4.739 | 4.146 |
| 5.571 | 4.952 | 4.869 | 5.954 | 3.799 |
| 4.600 | 5.000 | 5.390 | 5.129 | 4.880 |
| 5.690 | 4.682 | 5.088 | 5.657 | 4.935 |
| 5.816 | 6.095 | 4.365 | 3.832 | 5.485 |
| **5.196** | 3.126 | 4.440 | 4.539 | 4.405 |
| 6.290 | 5.428 | 5.444 | 4.738 | 4.850 |
| 5.669 | 5.896 | 4.050 | 3.787 | 4.565 |
| 6.191 | 5.731 | 5.781 | 3.681 | 4.921 |
| 3.798 | 5.578 | 5.569 | 5.931 | 6.535 |
| 4.980 | 5.040 | 4.178 | 5.011 | 4.394 |
| 4.843 | 5.677 | 4.734 | 4.355 | 3.653 |

**Elaborado por:** G. Cuenca

Por otro lado, matriz de datos resultante con dos valores completados por imputación por regresión en la variable *X1* y uno en *X4* , se puede ver en la Tabla 4.9.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.9**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 4.813 | 3.396 | 5.569 | 3.812 | 5.806 |
| 5.726 | 5.257 | 4.744 | 2.798 | 5.232 |
| 4.412 | 3.944 | 4.623 | 5.986 | 4.010 |
| 7.183 | 6.415 | 4.704 | 4.481 | 6.340 |
| 4.864 | 4.195 | 3.525 | 5.327 | 5.290 |
| 5.114 | 5.529 | 4.766 | 5.234 | 6.479 |
| 6.067 | 5.219 | **5.118** | 5.022 | 6.138 |
| 5.059 | 4.078 | 5.315 | 3.996 | 4.316 |
| 4.904 | 2.829 | 6.444 | 4.053 | 3.708 |
| **3.543** | 4.941 | 4.649 | 4.626 | 4.927 |
| 5.294 | 3.989 | 5.623 | 3.814 | 4.669 |
| 3.664 | 5.615 | 5.799 | 3.944 | 4.156 |
| 5.714 | 5.508 | 5.941 | 6.473 | 5.498 |
| 6.624 | 6.692 | 4.008 | 5.056 | 6.489 |
| 4.308 | 5.591 | 5.212 | 3.783 | 4.454 |
| 5.858 | 4.356 | 5.238 | 4.959 | 4.153 |
| 6.254 | 5.380 | 3.992 | 3.872 | 4.754 |
| 3.406 | 3.991 | 4.258 | **3.872** | 5.663 |
| 3.559 | 4.981 | 6.082 | 4.739 | 4.146 |
| 5.571 | 4.952 | 4.869 | 5.954 | 3.799 |
| 4.600 | 5.000 | 5.390 | 5.129 | 4.880 |
| 5.690 | 4.682 | 5.088 | 5.657 | 4.935 |
| 5.816 | 6.095 | 4.365 | 3.832 | 5.485 |
| **5.238** | 3.126 | 4.440 | 4.539 | 4.405 |
| 6.290 | 5.428 | 5.444 | 4.738 | 4.850 |
| 5.669 | 5.896 | 4.050 | 3.787 | 4.565 |
| 6.191 | 5.731 | 5.781 | 3.681 | 4.921 |
| 3.798 | 5.578 | 5.569 | 5.931 | 6.535 |
| 4.980 | 5.040 | 4.178 | 5.011 | 4.394 |
| 4.843 | 5.677 | 4.734 | 4.355 | 3.653 |

**Elaborado por:** G. Cuenca

En la Tabla 4.10 se realiza una comparación entre el valor real y el valor con imputación por la media y regresión. La diferencia en valor absoluto entre el dato observado de cada variable es menor en el Método de Imputación por Regresión, es decir los datos estimados por medio de la imputación por regresión, están más cercanos a los verdaderos valores, que los de la imputación por la media.

|  |
| --- |
| **Tabla 4.10**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 |
| **Imputación por Media y Regresión en dos valores de la variable** *X****1***   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Dato Observado** | **Resultado de Imputación por**  **la Media** | **Error**  **|Dato Observado –**  **Resultado de Imputación por Media|** | **Resultado de Predicción** | **Error**  **| Dato Observado –**  **Resultado de Predicción|** | | 4.168 | 5.196 | 1.028 | 3.543 | 0.625 | | 5.712 | 5.196 | 0.516 | 5.238 | 0.474 | |
| **Imputación por Media y Regresión en un valor de la variable** *X****4***   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Dato Observado** | **Resultado de Imputación por**  **la Media** | **Error**  **|Dato Observado –**  **Resultado de Imputación por Media|** | **Resultado de Predicción** | **Error**  **| Dato Observado –**  **Resultado de Predicción|** | | 3.651 | 4.641 | 0.990 | 3.872 | 0.221 | |

**Elaborado por:** G. Cuenca

En el Cuadro 4.10, se pueden observar los resultados de realizar la imputación por medio de la media y regresión en la variable *X*1:

|  |  |
| --- | --- |
| **CUADRO 4.10**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****1*” y “*Variable*** *X****4*”** | |
| **Estimadores “*Variable*** *X****1*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 30 | 28 | 30 | 30 | | Media | | 5,205 | 5,224 | 5,224 | 5,168 | | Mediana | | 5,205 | 5,204 | 5,196 | 5,176 | | Moda | | 3,410 | 3,410 | 5,200 | 3,410 | | Varianza | | 0,891 | 0,908 | 0,845 | 0,939 | | Desviación Estándar | | 0,944 | 0,953 | 0,919 | 0,969 | | Error Estándar | | 0,172 | 0,180 | 0,168 | 0,177 | | **Coeficiente de Asimetría** | | -0,161 | -0,188 | -0,188 | -0,174 | | Curtosis | | -0,460 | -0,392 | -0,196 | -0,479 | | Rango | | 3,780 | 3,780 | 3,780 | 3,780 | | Mínimo | | 3,410 | 3,410 | 3,410 | 3,410 | | Máximo | | 7,180 | 7,180 | 7,180 | 7,180 | | Percentiles | 25 | 4,553 | 4,653 | 4,760 | 4,553 | | 50 | 5,204 | 5,204 | 5,196 | 5,176 | | 75 | 5,827 | 5,848 | 5,827 | 5,827 | | **Diagrama de Cajas “*Variable*** *X****1*”** |
| **Estimadores “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 30 | 29 | 30 | 30 | | Media | | 4,608 | 4,641 | 4,641 | 4,615 | | Mediana | | 4,583 | 4,626 | 4,634 | 4,583 | | Moda | | 2,800 | 2,800 | 2,800 | 3,870 | | Varianza | | 0,756 | 0,750 | 0,724 | 0,743 | | Desviación Estándar | | 0,870 | 0,866 | 0,851 | 0,862 | | Error Estándar | | 0,159 | 0,161 | 0,155 | 0,157 | | **Coeficiente de Asimetría** | | 0,266 | 0,209 | 0,213 | 0,274 | | Curtosis | | -0,447 | -0,393 | -0,297 | -0,398 | | Rango | | 3,680 | 3,680 | 3,680 | 3,680 | | Mínimo | | 2,800 | 2,800 | 2,800 | 2,800 | | Máximo | | 6,470 | 6,470 | 6,470 | 6,470 | | Percentiles | 25 | 3,828 | 3,852 | 3,862 | 3,862 | | 50 | 4,583 | 4,626 | 4,634 | 4,583 | | 75 | 5,155 | 5,182 | 5,155 | 5,155 | | **Diagrama de Cajas “*Variable*** *X****4*”** |

**Elaborado por**: G. Cuenca

El vector de medias con dos datos completados por la media en *X1* y uno en *X4* es:



El vector de medias con dos datos completados por regresión en *X1* y uno en *X4* es:



El efecto que causa en la *matriz de varianzas y covarianzas* y *matriz de correlaciones*, el completar 2% de datos faltantes en una matriz de tamaño 30, por medio de la imputación por media y regresión, se presenta en el Cuadro 4.11.

|  |  |
| --- | --- |
| **CUADRO 4.11**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.891 |  |  |  |  | | *X*2 | 0.299 | 0.891 |  |  |  | | *X*3 | -0.152 | -0.138 | 0.502 |  |  | | *X*4 | -0.010 | 0.034 | 0.014 | 0.756 |  | | *X*5 | 0.197 | 0.315 | -0.123 | 0.090 | 0.740 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.335 | 1.000 |  |  |  | | *X*3 | -0.227 | -0.206 | 1.000 |  |  | | *X*4 | -0.012 | 0.042 | 0.023 | 1.000 |  | | *X*5 | 0.242 | **0.388** | -0.202 | 0.120 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **10% Datos Completados por Media en “*Variable*** *X****1*” y “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | **0.845** |  |  |  |  | | *X*2 | **0.330** | 0.891 |  |  |  | | *X*3 | **-0.154** | -0.138 | 0.502 |  |  | | *X*4 | **-0.070** | **0.001** | **-0.011** | **0.724** |  | | *X*5 | **0.205** | 0.315 | -0.123 | **0.114** | 0.740 | | **Matriz de Correlaciones**  **10% Datos Completados por Media en “*Variable*** *X****1*” y “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.381 | 1.000 |  |  |  | | *X*3 | -0.236 | -0.206 | 1.000 |  |  | | *X*4 | -0.089 | 0.001 | -0.018 | 1.000 |  | | *X*5 | 0.260 | 0.388 | -0.202 | 0.156 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **10% Datos Completados por Regresión en “*Variable*** *X****1*” y “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | **0.939** |  |  |  |  | | *X*2 | **0.329** | 0.891 |  |  |  | | *X*3 | **-0.136** | -0.138 | 0.502 |  |  | | *X*4 | **-0.022** | **0.027** | **0.009** | **0.743** |  | | *X*5 | **0.206** | 0.315 | -0.123 | **0.095** | 0.740 | | **Matriz de Correlaciones**  **10% Datos Completados por Regresión en “*Variable*** *X****1*” y “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.360 | 1.000 |  |  |  | | *X*3 | -0.197 | -0.206 | 1.000 |  |  | | *X*4 | -0.027 | 0.033 | 0.014 | 1.000 |  | | *X*5 | 0.247 | 0.388 | -0.202 | 0.128 | 1.000 | |

**Elaborado por**: G. Cuenca

**4.2.3 Distribución Poisson: *Ocho datos faltantes* en una sola variable (5% de la matriz), tamaño de muestra n=30**

Se tiene una matriz de datos cuyas columnas son muestras tomadas de cinco poblaciones todas ellas Poisson, independientes e idénticamente distribuìdas, con parámetro ,, *i= 1,2,....30*  y *j= 1,2,3*,*4,5* y se supone que tiene el 5% de datos faltantes, es decir ocho datos, los que recayeron en la variable *X*5 y son: el *X*3,5=6, *X*7,5=3, *X*10,5=3, *X*14,5=4, *X*18,5=5, *X*21,5=5, *X*25,5=9 y el *X*28,5=7.

Nótese que el 5% de datos faltantes en la matriz, constituye 27% de datos faltantes en la columna que corresponde a *X*5. (Ver Tabla 4.11)

Los resultados obtenidos para este caso se presentan desde la Tabla 4.11 hasta el Cuadro 4.16

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.11**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias independientes con distribución Poisson**  Tamaño de muestra n=30 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 3 | 10 | 8 | 4 | 2 |
| 3 | 6 | 7 | 8 | 5 |
| 6 | 8 | 3 | 10 | **6** |
| 6 | 4 | 7 | 10 | 8 |
| 11 | 5 | 7 | 2 | 4 |
| 4 | 6 | 9 | 5 | 3 |
| 9 | 5 | 7 | 6 | **3** |
| 3 | 8 | 9 | 6 | 5 |
| 5 | 2 | 10 | 6 | 8 |
| 9 | 7 | 4 | 7 | **3** |
| 8 | 4 | 7 | 10 | 4 |
| 3 | 9 | 2 | 8 | 2 |
| 6 | 9 | 6 | 4 | 4 |
| 5 | 10 | 6 | 3 | **4** |
| 7 | 5 | 6 | 11 | 7 |
| 5 | 8 | 3 | 5 | 3 |
| 8 | 11 | 6 | 7 | 8 |
| 9 | 12 | 7 | 2 | **5** |
| 6 | 4 | 8 | 6 | 12 |
| 5 | 12 | 7 | 9 | 8 |
| 3 | 2 | 8 | 9 | **5** |
| 8 | 9 | 4 | 3 | 10 |
| 8 | 10 | 4 | 6 | 7 |
| 4 | 4 | 6 | 7 | 8 |
| 3 | 8 | 5 | 0 | **9** |
| 5 | 9 | 4 | 7 | 11 |
| 5 | 7 | 8 | 5 | 4 |
| 4 | 4 | 5 | 2 | **7** |
| 8 | 0 | 7 | 6 | 5 |
| 5 | 8 | 7 | 4 | 9 |

**Elaborado por**: G. Cuenca

El vector de medias de los datos originales es:



**Método de Eliminación por Filas**

Puesto que los datos faltantes recayeron en la variable *X*5 y son: el *X*3,5=6, *X*7,5=3, *X*10,5=3, *X*14,5=4, *X*18,5=5, *X*25,5=9 y el *X*28,5=7, se procede a prescindir de las filas que tienen estos valores “faltantes”, donde la matriz de datos resultante con filas eliminadas se muestra en la Tabla 4.12.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.12**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 3 | 10 | 8 | 4 | 2 |
| 3 | 6 | 7 | 8 | 5 |
| 6 | 4 | 7 | 10 | 8 |
| 11 | 5 | 7 | 2 | 4 |
| 4 | 6 | 9 | 5 | 3 |
| 3 | 8 | 9 | 6 | 5 |
| 5 | 2 | 10 | 6 | 8 |
| 8 | 4 | 7 | 10 | 4 |
| 3 | 9 | 2 | 8 | 2 |
| 6 | 9 | 6 | 4 | 4 |
| 7 | 5 | 6 | 11 | 7 |
| 5 | 8 | 3 | 5 | 3 |
| 8 | 11 | 6 | 7 | 8 |
| 6 | 4 | 8 | 6 | 12 |
| 5 | 12 | 7 | 9 | 8 |
| 8 | 9 | 4 | 3 | 10 |
| 8 | 10 | 4 | 6 | 7 |
| 4 | 4 | 6 | 7 | 8 |
| 5 | 9 | 4 | 7 | 11 |
| 5 | 7 | 8 | 5 | 4 |
| 8 | 0 | 7 | 6 | 5 |
| 5 | 8 | 7 | 4 | 9 |

**Elaborado por**: G. Cuenca

El vector de medias para las veinte y dos filas restantes es:



Como era de esperarse el vector de medias de los datos originales y de los datos con filas eliminadas no coinciden.

Ahora analicemos el efecto que causa en la *matriz de varianzas y covarianzas*, y *matriz de correlaciones*, la eliminación de ocho filas, con un tamaño de muestra *n*=30.(Ver Cuadro 4.12)

|  |  |
| --- | --- |
| **CUADRO 4.12**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.993 |  |  |  |  | | *X*2 | -0.062 | 9.361 |  |  |  | | *X*3 | -0.469 | -2.140 | 3.771 |  |  | | *X*4 | -0.221 | -2.009 | -0.156 | 7.582 |  | | *X*5 | -0.110 | -0.246 | -0.061 | 0.067 | 7.275 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.009 | 1.000 |  |  |  | | *X*3 | -0.108 | -0.360 | 1.000 |  |  | | *X*4 | -0.036 | -0.238 | -0.029 | 1.000 |  | | *X*5 | -0.018 | -0.030 | -0.012 | 0.009 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **(8 Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.494 |  |  |  |  | | *X*2 | -1.242 | 9.394 |  |  |  | | *X*3 | -0.537 | -2.532 | 4.069 |  |  | | *X*4 | -0.719 | -1.082 | -0.294 | 5.465 |  | | *X*5 | 1.208 | -0.290 | -0.013 | 0.877 | 8.374 | | **Matriz de Correlaciones**  **(8 Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.191 | 1.000 |  |  |  | | *X*3 | -0.126 | -0.410 | 1.000 |  |  | | *X*4 | -0.145 | -0.151 | -0.062 | 1.000 |  | | *X*5 | 0.197 | -0.033 | -0.002 | 0.130 | 1.000 | |

**Elaborado por**: G. Cuenca

**Elaborado por**: G. Cuenca

|  |  |
| --- | --- |
| **CUADRO 4.13**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****5*”** | |
| **Estimadores**   |  |  |  |  | | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Con el 27% de datos eliminadas en X5** | | n | | 30 | 22 | | Media | | 5,967 | 6,227 | | Mediana | | 5,000 | 6,000 | | Moda | | 4,000 | 8,000 | | Varianza | | 7,275 | 8,374 | | Desviación Estándar | | 2,697 | 2,894 | | Error Estándar | | 0,492 | 0,617 | | **Coeficiente de Asimetría** | | 0,456 | 0,296 | | Curtosis | | -0,633 | -0,858 | | Rango | | 10,000 | 10,000 | | Mínimo | | 2,000 | 2,000 | | Máximo | | 12,000 | 12,000 | | Percentiles | 25 | 4,000 | 4,000 | | 50 | 5,000 | 6,000 | | 75 | 8,000 | 8,000 | | **Diagrama de Cajas** |

**Método de Imputación por la Media y Regresión**

A continuación se aplica el método de imputación por media y regresión a la misma matriz de datos utilizada en el método de eliminación por filas, es decir se completan datos en la variable *X*5 que presenta ocho valores faltantes que son: *X*3,5=6, *X*7,5=3, *X*10,5=3, *X*14,5=4, *X*18,5=5, *X*25,5=9 y el *X*28,5=7. La matriz de datos resultante con ocho valores completados por imputación por la media en la variable *X5* se muestra en la Tabla 4.13.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.13**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 3 | 10 | 8 | 4 | 2 |
| 3 | 6 | 7 | 8 | 5 |
| 6 | 8 | 3 | 10 | **6.227** |
| 6 | 4 | 7 | 10 | 8 |
| 11 | 5 | 7 | 2 | 4 |
| 4 | 6 | 9 | 5 | 3 |
| 9 | 5 | 7 | 6 | **6.227** |
| 3 | 8 | 9 | 6 | 5 |
| 5 | 2 | 10 | 6 | 8 |
| 9 | 7 | 4 | 7 | **6.227** |
| 8 | 4 | 7 | 10 | 4 |
| 3 | 9 | 2 | 8 | 2 |
| 6 | 9 | 6 | 4 | 4 |
| 5 | 10 | 6 | 3 | **6.227** |
| 7 | 5 | 6 | 11 | 7 |
| 5 | 8 | 3 | 5 | 3 |
| 8 | 11 | 6 | 7 | 8 |
| 9 | 12 | 7 | 2 | **6.227** |
| 6 | 4 | 8 | 6 | 12 |
| 5 | 12 | 7 | 9 | 8 |
| 3 | 2 | 8 | 9 | **6.227** |
| 8 | 9 | 4 | 3 | 10 |
| 8 | 10 | 4 | 6 | 7 |
| 4 | 4 | 6 | 7 | 8 |
| 3 | 8 | 5 | 0 | **6.227** |
| 5 | 9 | 4 | 7 | 11 |
| 5 | 7 | 8 | 5 | 4 |
| 4 | 4 | 5 | 2 | **6.227** |
| 8 | 0 | 7 | 6 | 5 |
| 5 | 8 | 7 | 4 | 9 |

**Elaborado por**: G. Cuenca

Mientras que la matriz de datos resultante, con ocho valores completados por imputación utilizando regresión en la variable *X5* se puede ver en la Tabla 4.14.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.14**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias independientes que provienen de una distribución Poisson**  **Método de Imputación por Regresión**  Tamaño de muestra n=30 y 5% de datos faltantes en la matriz | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 3 | 10 | 8 | 4 | 2 |
| 3 | 6 | 7 | 8 | 5 |
| 6 | 8 | 3 | 10 | **6.287** |
| 6 | 4 | 7 | 10 | 8 |
| 11 | 5 | 7 | 2 | 4 |
| 4 | 6 | 9 | 5 | 3 |
| 9 | 5 | 7 | 6 | **5.110** |
| 3 | 8 | 9 | 6 | 5 |
| 5 | 2 | 10 | 6 | 8 |
| 9 | 7 | 4 | 7 | **3.420** |
| 8 | 4 | 7 | 10 | 4 |
| 3 | 9 | 2 | 8 | 2 |
| 6 | 9 | 6 | 4 | 4 |
| 5 | 10 | 6 | 3 | **4.310** |
| 7 | 5 | 6 | 11 | 7 |
| 5 | 8 | 3 | 5 | 3 |
| 8 | 11 | 6 | 7 | 8 |
| 9 | 12 | 7 | 2 | **6.005** |
| 6 | 4 | 8 | 6 | 12 |
| 5 | 12 | 7 | 9 | 8 |
| 3 | 2 | 8 | 9 | **5.106** |
| 8 | 9 | 4 | 3 | 10 |
| 8 | 10 | 4 | 6 | 7 |
| 4 | 4 | 6 | 7 | 8 |
| 3 | 8 | 5 | 0 | **8.873** |
| 5 | 9 | 4 | 7 | 11 |
| 5 | 7 | 8 | 5 | 4 |
| 4 | 4 | 5 | 2 | **5.517** |
| 8 | 0 | 7 | 6 | 5 |
| 5 | 8 | 7 | 4 | 9 |

**Elaborado por**: G. Cuenca

En la Tabla 4.15 se realiza una comparación entre el dato observado y el dato con imputación por la media y regresión.

|  |
| --- |
| **Tabla 4.15**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 |
| **27% de datos completados en** *X1* **por la Media**   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Imputación por Media** | **Error**  **|Dato Observado –**  **Resultado de Imputación por Media|** | | 6 | 6.227 | 0.227 | | 3 | 6.227 | 3.227 | | 3 | 6.227 | 3.227 | | 4 | 6.227 | 2.227 | | 5 | 6.227 | 1.227 | | 5 | 6.227 | 1.227 | | 9 | 6.227 | 2.773 | | 7 | 6.227 | 0.773 | |
| **27% de datos completados en** *X1* **por Regresión**   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Predicción** | **Error**  **| Dato Observado –**  **Resultado de Predicción** | | 6 | 6.287 | 0.287 | | 3 | 5.110 | 2.110 | | 3 | 3.420 | 0.420 | | 4 | 4.310 | 0.310 | | 5 | 6.005 | 1.005 | | 5 | 5.106 | 0.106 | | 9 | 8.873 | 0.127 | | 7 | 5.517 | 1.483 | |

**Elaborado por**: G. Cuenca

Por medio del Cuadro 4.14, podemos apreciar el número de imputaciones sucesivas por medio del Método de Regresión que se realiza a los ocho datos faltantes en la variable *X5.*

|  |  |
| --- | --- |
| **CUADRO 4.14**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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 para** *X*3,5=6   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 6.878 | 0.878 | | 2 | 6.754 | 0.754 | | 3 | 6.632 | 0.632 | | 4 | 6.576 | 0.576 | | 5 | 6.471 | 0.471 | | 6 | 6.323 | 0.323 | | 7 | 6.287 | 0.287 | | **Imputaciones sucesivas para** *X*7,5=3   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 6.119 | 3.119 | | 2 | 6.032 | 3.032 | | 3 | 5.971 | 2.971 | | 4 | 5.862 | 2.862 | | 5 | 5.531 | 2.531 | | 6 | 5.204 | 2.204 | | 7 | 5.110 | 2.110 | |
| **Imputaciones sucesivas para** *X*10,5=3   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 5.429 | 2.429 | | 2 | 5.210 | 2.210 | | 3 | 4.973 | 1.973 | | 4 | 4.415 | 1.415 | | 5 | 4.206 | 1.206 | | 6 | 3.843 | 0.843 | | 7 | 3.420 | 0.420 | | **Imputaciones sucesivas para** *X*14,5=4   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 5.184 | 1.184 | | 2 | 5.003 | 1.003 | | 3 | 4.852 | 0.852 | | 4 | 4.725 | 0.725 | | 5 | 4.612 | 0.612 | | 6 | 4.561 | 0.561 | | 7 | 4.310 | 0.310 | |

**Elaborado por**: G. Cuenca

**Continúa…**

**Viene…**

|  |  |
| --- | --- |
| *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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 para** *X*18,5=5   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 6.751 | 1.751 | | 2 | 6.623 | 1.623 | | 3 | 6.541 | 1.541 | | 4 | 6.432 | 1.432 | | 5 | 6.317 | 1.317 | | 6 | 6.210 | 1.210 | | 7 | 6.005 | 1.005 | | **Imputaciones sucesivas para** *X*21,5=5   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 5.749 | 0.749 | | 2 | 5.663 | 0.663 | | 3 | 5.549 | 0.549 | | 4 | 5.432 | 0.432 | | 5 | 5.316 | 0.316 | | 6 | 5.257 | 0.257 | | 7 | 5.106 | 0.106 | |
| **Imputaciones sucesivas para** *X*25,5=9   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 8.215 | 0.785 | | 2 | 8.351 | 0.649 | | 3 | 8.532 | 0.468 | | 4 | 8.673 | 0.327 | | 5 | 8.725 | 0.275 | | 6 | 8.801 | 0.199 | | 7 | 8.873 | 0.127 | | **Imputaciones sucesivas para** *X*28,5=7   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 4.364 | 2.636 | | 2 | 4.713 | 2.287 | | 3 | 4.846 | 2.154 | | 4 | 5.112 | 1.888 | | 5 | 5.235 | 1.765 | | 6 | 5.418 | 1.582 | | 7 | 5.517 | 1.483 | |

**Elaborado por**: G. Cuenca

**Elaborado por:** G. Cuenca

|  |  |
| --- | --- |
| **CUADRO 4.15**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****5*”** | |
| **Estimadores**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 30 | 22 | 30 | 30 | | Media | | 5,966 | 6,227 | 6,227 | 6,054 | | Mediana | | 5,000 | 6,000 | 6,227 | 5,314 | | Moda | | 4,000 | 8,000 | 6,230 | 8,000 | | Varianza | | 7,275 | 8,375 | 6,064 | 6,779 | | Desviación Estándar | | 2,697 | 2,894 | 2,462 | 2,604 | | Error Estándar | | 0,492 | 0,617 | 0,450 | 0,475 | | **Coeficiente de Asimetría** | | 0,456 | 0,296 | 0,339 | 0,471 | | Curtosis | | -0,633 | -0,858 | 0,010 | -0,429 | | Rango | | 10,000 | 10,000 | 10,000 | 10,000 | | Mínimo | | 2,000 | 2,000 | 2,000 | 2,000 | | Máximo | | 12,000 | 12,000 | 12,000 | 12,000 | | Percentiles | 25 | 4,000 | 4,000 | 4,000 | 4,000 | | 50 | 5,000 | 6,000 | 6,227 | 5,314 | | 75 | 8,000 | 8,000 | 8,000 | 8,000 | | **Diagrama de Cajas** |

El vector de medias con ocho “datos completados” por la media en *X5* es:



Mientras que el vector de medias con ocho “datos completados” utilizando regresión en *X5* es:



|  |  |
| --- | --- |
| **CUADRO 4.16**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.993 |  |  |  |  | | *X*2 | -0.062 | 9.361 |  |  |  | | *X*3 | -0.469 | -2.140 | 3.771 |  |  | | *X*4 | -0.221 | -2.009 | -0.156 | 7.582 |  | | *X*5 | -0.110 | -0.246 | -0.061 | 0.067 | 7.275 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.009 | 1.000 |  |  |  | | *X*3 | -0.108 | -0.360 | 1.000 |  |  | | *X*4 | -0.036 | -0.238 | -0.029 | 1.000 |  | | *X*5 | -0.018 | -0.030 | -0.012 | 0.009 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **27% Datos Completados por Media en “*Variable*** *X****5*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.993 |  |  |  |  | | *X*2 | -0.062 | 9.361 |  |  |  | | *X*3 | -0.469 | -2.140 | 3.771 |  |  | | *X*4 | -0.221 | -2.009 | -0.156 | 7.582 |  | | *X*5 | 0.875 | -0.210 | -0.009 | 0.635 | 6.064 | | **Matriz de Correlaciones**  **27% Datos Completados por Media en “*Variable*** *X****5*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.009 | 1.000 |  |  |  | | *X*3 | -0.108 | -0.360 | 1.000 |  |  | | *X*4 | -0.036 | -0.238 | -0.029 | 1.000 |  | | *X*5 | 0.159 | -0.028 | -0.002 | 0.094 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **27% Datos Completados por Regresión en “*Variable*** *X****5*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.993 |  |  |  |  | | *X*2 | -0.062 | 9.361 |  |  |  | | *X*3 | -0.469 | -2.140 | 3.771 |  |  | | *X*4 | -0.221 | -2.009 | -0.156 | 7.582 |  | | *X*5 | 0.367 | -0.033 | 0.030 | 0.198 | 6.779 | | **Matriz de Correlaciones**  **27% Datos Completados por Regresión en “*Variable*** *X****5*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.009 | 1.000 |  |  |  | | *X*3 | -0.108 | -0.360 | 1.000 |  |  | | *X*4 | -0.036 | -0.238 | -0.029 | 1.000 |  | | *X*5 | 0.063 | -0.004 | 0.006 | 0.028 | 1.000 | |

**Elaborado por**: G. Cuenca

**4.2.4 Distribución Exponencial: *Trece datos faltantes* en una sola variable (5% de la matriz), tamaño de muestra n=50**

Se tiene una matriz de datos cuyas columnas son muestras tomadas de cinco poblaciones todas ellas Exponencial, independientes e idénticamente distribuìdas, con parámetro ,, *i= 1,2,....50*  y *j= 1,2,3*,*4,5* y se supone que tiene el 5% de datos faltantes, es decir trece datos, los que recayeron en la variable *X*2 y son: el *X*3,2=0.335, *X*6,2=2.326, *X*10,2=0.158, *X*13,2=2.019, *X*18,2=1.525, *X*25,2=0.169, *X*28,2=0.606, *X*31,2=4.334, *X*33,2=0.950, *X*33,2=0.950, *X*37,2=4.403, *X*41,2=0.775, *X*46,2=0.337 y *X*49,2=2.209.

Nótese que el 5% de datos faltantes en la matriz, constituye 26% de datos faltantes en la columna que corresponde a *X*2.(Ver Tabla 4.16).

Los resultados correspondientes a este caso se presentan desde la Tabla 4.16 hasta el Cuadro 4.19.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.16**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias independientes con distribución Exponencial**  Tamaño de muestra n=50 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 0.308 | 5.836 | 3.978 | 0.967 | 3.134 |
| 0.399 | 4.329 | 0.284 | 1.314 | 0.790 |
| 2.807 | **0.335** | 2.222 | 2.019 | 2.838 |
| 0.216 | 3.516 | 1.435 | 0.514 | 0.656 |
| 1.008 | 6.595 | 1.681 | 0.377 | 1.833 |
| 3.936 | **2.326** | 2.690 | 2.289 | 1.863 |
| 3.649 | 0.404 | 0.034 | 1.035 | 1.776 |
| 0.249 | 2.899 | 0.070 | 0.492 | 3.798 |
| 0.043 | 1.064 | 0.106 | 6.787 | 0.260 |
| 1.017 | **0.158** | 1.589 | 0.309 | 7.656 |
| 2.033 | 2.217 | 2.207 | 0.764 | 0.158 |
| 2.927 | 3.164 | 9.718 | 0.442 | 0.916 |
| 1.598 | **2.019** | 3.594 | 0.172 | 5.409 |
| 0.883 | 0.049 | 0.377 | 1.893 | 0.916 |
| 2.811 | 0.666 | 0.882 | 1.502 | 0.565 |
| 1.519 | 0.882 | 2.265 | 4.860 | 0.524 |
| 1.397 | 0.490 | 7.271 | 0.156 | 0.426 |
| 5.182 | **1.525** | 2.069 | 0.776 | 1.220 |
| 0.532 | 2.920 | 0.592 | 0.889 | 1.447 |
| 6.186 | 2.292 | 2.417 | 0.008 | 1.636 |
| 3.602 | 0.475 | 3.416 | 1.388 | 0.935 |
| 0.504 | 0.910 | 1.758 | 1.688 | 5.108 |
| 3.137 | 0.832 | 0.285 | 0.375 | 1.217 |
| 2.758 | 1.241 | 0.940 | 4.296 | 3.130 |
| 1.850 | **0.169** | 5.794 | 1.026 | 2.619 |
| 1.465 | 0.496 | 1.812 | 3.017 | 2.464 |
| 1.350 | 2.237 | 2.395 | 1.839 | 0.392 |
| 0.941 | **0.606** | 3.764 | 2.400 | 1.776 |
| 0.938 | 1.652 | 2.348 | 0.913 | 1.281 |
| 0.018 | 1.049 | 11.453 | 0.113 | 0.166 |
| 2.048 | **4.334** | 1.654 | 10.276 | 1.940 |
| 2.575 | 2.284 | 0.782 | 0.405 | 0.896 |
| 0.777 | **0.950** | 0.390 | 0.740 | 3.500 |
| 1.352 | 0.223 | 0.560 | 0.038 | 0.482 |
| 2.569 | 0.074 | 3.632 | 5.326 | 2.012 |
| 1.094 | 7.818 | 1.188 | 6.204 | 0.505 |
| 0.390 | **4.403** | 1.288 | 0.602 | 2.145 |
| 0.121 | 1.818 | 0.168 | 0.399 | 0.512 |
| 1.622 | 4.662 | 1.633 | 2.688 | 3.823 |
| 1.720 | 2.455 | 6.211 | 0.702 | 3.818 |
| 0.008 | **0.775** | 0.072 | 2.432 | 0.896 |
| 0.975 | 0.041 | 8.616 | 4.995 | 4.742 |
| 0.115 | 1.835 | 1.188 | 0.266 | 0.148 |
| 0.184 | 0.395 | 0.136 | 5.116 | 1.447 |
| 0.409 | 4.056 | 0.214 | 0.600 | 3.625 |
| 0.743 | **0.337** | 0.963 | 4.158 | 2.572 |
| 2.351 | 2.916 | 0.714 | 1.625 | 4.066 |
| 1.166 | 5.402 | 0.126 | 1.047 | 7.526 |
| 0.903 | **2.209** | 1.588 | 0.904 | 2.928 |
| 0.525 | 1.106 | 3.467 | 1.260 | 0.336 |

El vector de medias de los datos originales es:



**Método de Eliminación por Filas**

Puesto que los datos faltantes recayeron en la variable *X*2 y son: el *X*3,2=0.335, *X*6,2=2.326, *X*10,2=0.158, *X*13,2=2.019, *X*18,2=1.525, *X*25,2=0.169, *X*28,2=0.606, *X*31,2=4.334, *X*33,2=0.950, *X*33,2=0.950, *X*37,2=4.403, *X*41,2=0.775, *X*46,2=0.337 y *X*49,2=2.209, se procede a prescindir de las filas que tienen estos valores “faltantes”, donde la matriz de datos resultante con filas eliminadas se muestra en la Tabla 4.17.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.17**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 0.308 | 5.836 | 3.978 | 0.967 | 3.134 |
| 0.399 | 4.329 | 0.284 | 1.314 | 0.790 |
| 0.216 | 3.516 | 1.435 | 0.514 | 0.656 |
| 1.008 | 6.595 | 1.681 | 0.377 | 1.833 |
| 3.649 | 0.404 | 0.034 | 1.035 | 1.776 |
| 0.249 | 2.899 | 0.070 | 0.492 | 3.798 |
| 0.043 | 1.064 | 0.106 | 6.787 | 0.260 |
| 2.033 | 2.217 | 2.207 | 0.764 | 0.158 |
| 2.927 | 3.164 | 9.718 | 0.442 | 0.916 |
| 0.883 | 0.049 | 0.377 | 1.893 | 0.916 |
| 2.811 | 0.666 | 0.882 | 1.502 | 0.565 |
| 1.519 | 0.882 | 2.265 | 4.860 | 0.524 |
| 1.397 | 0.490 | 7.271 | 0.156 | 0.426 |
| 0.532 | 2.920 | 0.592 | 0.889 | 1.447 |
| 6.186 | 2.292 | 2.417 | 0.008 | 1.636 |
| 3.602 | 0.475 | 3.416 | 1.388 | 0.935 |
| 0.504 | 0.910 | 1.758 | 1.688 | 5.108 |
| 3.137 | 0.832 | 0.285 | 0.375 | 1.217 |
| 2.758 | 1.241 | 0.940 | 4.296 | 3.130 |
| 1.465 | 0.496 | 1.812 | 3.017 | 2.464 |
| 1.350 | 2.237 | 2.395 | 1.839 | 0.392 |
| 0.938 | 1.652 | 2.348 | 0.913 | 1.281 |
| 0.018 | 1.049 | 11.453 | 0.113 | 0.166 |
| 2.575 | 2.284 | 0.782 | 0.405 | 0.896 |
| 1.352 | 0.223 | 0.560 | 0.038 | 0.482 |
| 2.569 | 0.074 | 3.632 | 5.326 | 2.012 |
| 1.094 | 7.818 | 1.188 | 6.204 | 0.505 |
| 0.121 | 1.818 | 0.168 | 0.399 | 0.512 |
| 1.622 | 4.662 | 1.633 | 2.688 | 3.823 |
| 1.720 | 2.455 | 6.211 | 0.702 | 3.818 |
| 0.975 | 0.041 | 8.616 | 4.995 | 4.742 |
| 0.115 | 1.835 | 1.188 | 0.266 | 0.148 |
| 0.184 | 0.395 | 0.136 | 5.116 | 1.447 |
| 0.409 | 4.056 | 0.214 | 0.600 | 3.625 |
| 2.351 | 2.916 | 0.714 | 1.625 | 4.066 |
| 1.166 | 5.402 | 0.126 | 1.047 | 7.526 |
| 0.525 | 1.106 | 3.467 | 1.260 | 0.336 |

**Elaborado por**: G. Cuenca

El vector de medias para las treinta y siete filas restantes es:



El efecto que causa en la *matriz de varianzas y covarianzas*, y *matriz de correlaciones*, la eliminación de trece filas, con un tamaño de muestra *n*=50, se puede ver en el Cuadro 4.17.

|  |  |
| --- | --- |
| **CUADRO 4.17**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.852 |  |  |  |  | | *X*2 | -0.355 | 3.437 |  |  |  | | *X*3 | 0.249 | -0.651 | 6.516 |  |  | | *X*4 | -0.164 | 0.072 | -0.517 | 4.472 |  | | *X*5 | -0.124 | 0.309 | -0.189 | -0.225 | 3.241 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.141 | 1.000 |  |  |  | | *X*3 | 0.072 | -0.138 | 1.000 |  |  | | *X*4 | -0.057 | 0.018 | -0.096 | 1.000 |  | | *X*5 | -0.050 | 0.092 | -0.041 | -0.059 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **(13 Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.769 |  |  |  |  | | *X*2 | -0.484 | 3.850 |  |  |  | | *X*3 | 0.123 | -0.786 | 8.060 |  |  | | *X*4 | -0.279 | -0.361 | -0.466 | 3.644 |  | | *X*5 | -0.034 | 0.821 | -0.264 | 0.152 | 3.002 | | **Matriz de Correlaciones**  **(13 Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.186 | 1.000 |  |  |  | | *X*3 | 0.033 | -0.141 | 1.000 |  |  | | *X*4 | -0.110 | -0.096 | -0.086 | 1.000 |  | | *X*5 | -0.015 | 0.242 | -0.054 | 0.046 | 1.000 | |

**Elaborado por**: G. Cuenca

**Método de Imputación por la Media y Regresión**

La matriz de datos resultante con trece valores completados por imputación por la media y regresión en la variable *X2* se muestra en la Tabla 4.18 y 4.19 respectivamente.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.18**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 0,308 | 5,836 | 3,978 | 0,967 | 3,134 |
| 0,399 | 4,329 | 0,284 | 1,314 | 0,790 |
| 2,807 | **2.197** | 2,222 | 2,019 | 2,838 |
| 0,216 | 3,516 | 1,435 | 0,514 | 0,656 |
| 1,008 | 6,595 | 1,681 | 0,377 | 1,833 |
| 3,936 | **2.197** | 2,690 | 2,289 | 1,863 |
| 3,649 | 0,404 | 0,034 | 1,035 | 1,776 |
| 0,249 | 2,899 | 0,070 | 0,492 | 3,798 |
| 0,043 | 1,064 | 0,106 | 6,787 | 0,260 |
| 1,017 | **2.197** | 1,589 | 0,309 | 7,656 |
| 2,033 | 2,217 | 2,207 | 0,764 | 0,158 |
| 2,927 | 3,164 | 9,718 | 0,442 | 0,916 |
| 1,598 | **2.197** | 3,594 | 0,172 | 5,409 |
| 0,883 | 0,049 | 0,377 | 1,893 | 0,916 |
| 2,811 | 0,666 | 0,882 | 1,502 | 0,565 |
| 1,519 | 0,882 | 2,265 | 4,860 | 0,524 |
| 1,397 | 0,490 | 7,271 | 0,156 | 0,426 |
| 5,182 | **2.197** | 2,069 | 0,776 | 1,220 |
| 0,532 | 2,920 | 0,592 | 0,889 | 1,447 |
| 6,186 | 2,292 | 2,417 | 0,008 | 1,636 |
| 3,602 | 0,475 | 3,416 | 1,388 | 0,935 |
| 0,504 | 0,910 | 1,758 | 1,688 | 5,108 |
| 3,137 | 0,832 | 0,285 | 0,375 | 1,217 |
| 2,758 | 1,241 | 0,940 | 4,296 | 3,130 |
| 1,850 | **2.197** | 5,794 | 1,026 | 2,619 |
| 1,465 | 0,496 | 1,812 | 3,017 | 2,464 |
| 1,350 | 2,237 | 2,395 | 1,839 | 0,392 |
| 0,941 | **2.197** | 3,764 | 2,400 | 1,776 |
| 0,938 | 1,652 | 2,348 | 0,913 | 1,281 |
| 0,018 | 1,049 | 11,453 | 0,113 | 0,166 |
| 2,048 | 2.197 | 1,654 | 10,276 | 1,940 |
| 2,575 | 2,284 | 0,782 | 0,405 | 0,896 |
| 0,777 | **2.197** | 0,390 | 0,740 | 3,500 |
| 1,352 | 0,223 | 0,560 | 0,038 | 0,482 |
| 2,569 | 0,074 | 3,632 | 5,326 | 2,012 |
| 1,094 | 7,818 | 1,188 | 6,204 | 0,505 |
| 0,390 | **2.197** | 1,288 | 0,602 | 2,145 |
| 0,121 | 1,818 | 0,168 | 0,399 | 0,512 |
| 1,622 | 4,662 | 1,633 | 2,688 | 3,823 |
| 1,720 | 2,455 | 6,211 | 0,702 | 3,818 |
| 0,008 | **2.197** | 0,072 | 2,432 | 0,896 |
| 0,975 | 0,041 | 8,616 | 4,995 | 4,742 |
| 0,115 | 1,835 | 1,188 | 0,266 | 0,148 |
| 0,184 | 0,395 | 0,136 | 5,116 | 1,447 |
| 0,409 | 4,056 | 0,214 | 0,600 | 3,625 |
| 0,743 | **2.197** | 0,963 | 4,158 | 2,572 |
| 2,351 | 2,916 | 0,714 | 1,625 | 4,066 |
| 1,166 | 5,402 | 0,126 | 1,047 | 7,526 |
| 0,903 | **2,197** | 1,588 | 0,904 | 2,928 |
| 0,525 | 1,106 | 3,467 | 1,260 | 0,336 |
| **Tabla 4.19**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias independientes con distribución Exponencial**  **Método de Imputación por Regresión**  Tamaño de muestra n=50 y 5% de datos faltantes en la matriz | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 0.308 | 5.836 | 3.978 | 0.967 | 3.134 |
| 0.399 | 4.329 | 0.284 | 1.314 | 0.790 |
| 2.807 | **2.070** | 2.222 | 2.019 | 2.838 |
| 0.216 | 3.516 | 1.435 | 0.514 | 0.656 |
| 1.008 | 6.595 | 1.681 | 0.377 | 1.833 |
| 3.936 | **1.403** | 2.690 | 2.289 | 1.863 |
| 3.649 | 0.404 | 0.034 | 1.035 | 1.776 |
| 0.249 | 2.899 | 0.070 | 0.492 | 3.798 |
| 0.043 | 1.064 | 0.106 | 6.787 | 0.260 |
| 1.017 | **3.682** | 1.589 | 0.309 | 7.656 |
| 2.033 | 2.217 | 2.207 | 0.764 | 0.158 |
| 2.927 | 3.164 | 9.718 | 0.442 | 0.916 |
| 1.598 | **3.246** | 3.594 | 0.172 | 5.409 |
| 0.883 | 0.049 | 0.377 | 1.893 | 0.916 |
| 2.811 | 0.666 | 0.882 | 1.502 | 0.565 |
| 1.519 | 0.882 | 2.265 | 4.860 | 0.524 |
| 1.397 | 0.490 | 7.271 | 0.156 | 0.426 |
| 5.182 | **1.151** | 2.069 | 0.776 | 1.220 |
| 0.532 | 2.920 | 0.592 | 0.889 | 1.447 |
| 6.186 | 2.292 | 2.417 | 0.008 | 1.636 |
| 3.602 | 0.475 | 3.416 | 1.388 | 0.935 |
| 0.504 | 0.910 | 1.758 | 1.688 | 5.108 |
| 3.137 | 0.832 | 0.285 | 0.375 | 1.217 |
| 2.758 | 1.241 | 0.940 | 4.296 | 3.130 |
| 1.850 | **1.978** | 5.794 | 1.026 | 2.619 |
| 1.465 | 0.496 | 1.812 | 3.017 | 2.464 |
| 1.350 | 2.237 | 2.395 | 1.839 | 0.392 |
| 0.941 | **2.117** | 3.764 | 2.400 | 1.776 |
| 0.938 | 1.652 | 2.348 | 0.913 | 1.281 |
| 0.018 | 1.049 | 11.453 | 0.113 | 0.166 |
| 2.048 | **0.907** | 1.654 | 10.276 | 1.940 |
| 2.575 | 2.284 | 0.782 | 0.405 | 0.896 |
| 0.777 | **3.181** | 0.390 | 0.740 | 3.500 |
| 1.352 | 0.223 | 0.560 | 0.038 | 0.482 |
| 2.569 | 0.074 | 3.632 | 5.326 | 2.012 |
| 1.094 | 7.818 | 1.188 | 6.204 | 0.505 |
| 0.390 | **3.011** | 1.288 | 0.602 | 2.145 |
| 0.121 | 1.818 | 0.168 | 0.399 | 0.512 |
| 1.622 | 4.662 | 1.633 | 2.688 | 3.823 |
| 1.720 | 2.455 | 6.211 | 0.702 | 3.818 |
| 0.008 | **2.484** | 0.072 | 2.432 | 0.896 |
| 0.975 | 0.041 | 8.616 | 4.995 | 4.742 |
| 0.115 | 1.835 | 1.188 | 0.266 | 0.148 |
| 0.184 | 0.395 | 0.136 | 5.116 | 1.447 |
| 0.409 | 4.056 | 0.214 | 0.600 | 3.625 |
| 0.743 | **2.395** | 0.963 | 4.158 | 2.572 |
| 2.351 | 2.916 | 0.714 | 1.625 | 4.066 |
| 1.166 | 5.402 | 0.126 | 1.047 | 7.526 |
| 0.903 | **2.891** | 1.588 | 0.904 | 2.928 |
| 0.525 | 1.106 | 3.467 | 1.260 | 0.336 |

En la Tabla 4.20 se realiza una comparación entre el dato observado y el dato con imputación por la media y regresión.

|  |
| --- |
| **Tabla 4.20**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 |
| ***26% de datos completados en*** *X2* ***por la Media***   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Imputación por Media** | **Error**  **|Dato Observado –**  **Resultado de Imputación por Media|** | | 0.335 | 2.197 | 1.862 | | 2.326 | 2.197 | 0.129 | | 0.158 | 2.197 | 2.039 | | 2.019 | 2.197 | 0.178 | | 1.525 | 2.197 | 0.672 | | 0.169 | 2.197 | 2.028 | | 0.606 | 2.197 | 1.591 | | 4.334 | 2.197 | 2.137 | | 0.950 | 2.197 | 1.247 | | 4.403 | 2.197 | 2.206 | | 0.775 | 2.197 | 1.422 | | 0.337 | 2.197 | 1.860 | | 2.090 | 2.197 | 0.107 | |
| ***26% de datos completados en*** *X2* ***por Regresión***   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Predicción** | **Error**  **| Dato Observado –**  **Resultado de Predicción** | | 0,335 | 2,070 | 1,735 | | 2,326 | 1,403 | 0,923 | | 0,158 | 3,682 | 3,524 | | 2,019 | 3,246 | 1,227 | | 1,525 | 1,151 | 0,374 | | 0,169 | 1,978 | 1,809 | | 0,606 | 2,117 | 1,511 | | 4,334 | 0,907 | 3,427 | | 0,950 | 3,181 | 2,231 | | 4,403 | 3,011 | 1,392 | | 0,775 | 2,484 | 1,709 | | 0,337 | 2,395 | 2,058 | | 2,090 | 2,891 | 0,801 | |

**Elaborado por**: G. Cuenca

**Elaborado por**: G. Cuenca

|  |  |
| --- | --- |
| **CUADRO 4.18**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****2*”** | |
| **Estimadores**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 50 | 37 | 50 | 50 | | Media | | 2,029 | 2,197 | 2,197 | 2,236 | | Mediana | | 1,589 | 1,818 | 2,197 | 2,094 | | Moda | | 0,040 | 0,040 | 2,200 | 0,040 | | Varianza | | 3,437 | 3,850 | 2,828 | 3,011 | | Desviación Estándar | | 1,854 | 1,962 | 1,682 | 1,735 | | Error Estándar | | 0,262 | 0,323 | 0,238 | 0,245 | | **Coeficiente de Asimetría** | | 1,222 | 1,164 | 1,338 | 1,158 | | Curtosis | | 1,115 | 0,879 | 2,207 | 1,487 | | Rango | | 7,780 | 7,780 | 7,780 | 7,780 | | Mínimo | | 0,040 | 0,040 | 0,040 | 0,040 | | Máximo | | 7,820 | 7,820 | 7,820 | 7,820 | | Percentiles | 25 | 0,495 | 0,581 | 0,903 | 0,901 | | 50 | 1,589 | 1,818 | 2,197 | 2,094 | | 75 | 2,917 | 3,042 | 2,566 | 3,049 | | **Diagrama de Cajas** |

El vector de medias con trece datos completados por la media en *X2* es:



Mientras que el vector de medias con trece datos completados por la regresión en *X2* es:



El efecto que causa en la *matriz de varianzas* y *covarianzas* y *matriz de correlaciones*, el completar 5% de datos faltantes en una matriz de tamaño 50, por medio de la imputación por media y regresión, se presenta en el Cuadro 4.19.

|  |  |
| --- | --- |
| **CUADRO 4.19**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.852 |  |  |  |  | | *X*2 | -0.355 | 3.437 |  |  |  | | *X*3 | 0.249 | -0.651 | 6.516 |  |  | | *X*4 | -0.164 | 0.072 | -0.517 | 4.472 |  | | *X*5 | -0.124 | 0.309 | -0.189 | -0.225 | 3.241 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.141 | 1.000 |  |  |  | | *X*3 | 0.072 | -0.138 | 1.000 |  |  | | *X*4 | -0.057 | 0.018 | -0.096 | 1.000 |  | | *X*5 | -0.050 | 0.092 | -0.041 | -0.059 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **26% Datos Completados por Media en “*Variable*** *X****2*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.852 |  |  |  |  | | *X*2 | -0.356 | 2.828 |  |  |  | | *X*3 | 0.249 | -0.578 | 6.516 |  |  | | *X*4 | -0.164 | -0.265 | -0.517 | 4.472 |  | | *X*5 | -0.124 | 0.603 | -0.189 | -0.225 | 3.241 | | **Matriz de Correlaciones**  **26% Datos Completados por Media en “*Variable*** *X****2*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.155 | 1.000 |  |  |  | | *X*3 | 0.072 | -0.135 | 1.000 |  |  | | *X*4 | -0.057 | -0.075 | -0.096 | 1.000 |  | | *X*5 | -0.050 | 0.199 | -0.041 | -0.059 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **26% Datos Completados por Regresión en “*Variable*** *X****2*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.852 |  |  |  |  | | *X*2 | -0.560 | 3.011 |  |  |  | | *X*3 | 0.249 | -0.657 | 6.516 |  |  | | *X*4 | -0.164 | -0.597 | -0.517 | 4.472 |  | | *X*5 | -0.124 | 0.901 | -0.189 | -0.225 | 3.241 | | **Matriz de Correlaciones**  **26% Datos Completados por Regresión en “*Variable*** *X****2*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | -0.237 | 1.000 |  |  |  | | *X*3 | 0.072 | -0.148 | 1.000 |  |  | | *X*4 | -0.057 | -0.163 | -0.096 | 1.000 |  | | *X*5 | -0.050 | 0.289 | -0.041 | -0.059 | 1.000 | |

**Elaborado por**: G. Cuenca

**4.3 Matrices de Datos con variables aleatorias dependientes**

En esta sección se realiza la comparación de los Métodos de Imputación, utilizando matrices de datos con variables aleatorias dependientes, con las distribuciones Normal, Poisson y Exponencial.

**4.3.1 Distribución Normal: *Trece datos faltantes* en una sola variable (5% de la matriz), tamaño de muestra n=50**

Se tiene una matriz de datos cuyas columnas son muestras tomadas de cinco poblaciones todas ellas Normal, dependientes e idénticamente distribuidas, con parámetros =10 y =1, , *i= 1,2,....50*  y *j= 1,2,3*,*4,5* y se supone que tiene el 5% de datos faltantes, es decir trece datos, los que recayeron en la variable *X*3 y son: el *X*2,3=9.010, *X*5,3=11.221, *X*6,3=10.102, *X*9,3=9.927, *X*11,3=10.718, *X*17,3=11.504, *X*21,3=12.263, *X*23,3=10.329, *X*29,3=10.655, *X*32,3=9.547, *X*37,3=9.509, *X*41,3=9.189 y el *X*46,3=9.549 . Nótese que el 5% de datos faltantes en la matriz, constituye 26% de datos faltantes en la columna que corresponde a *X*3.(Ver Tabla 4.21)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.21**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias dependientes con distribución Normal (10, 1)**  Tamaño de muestra n=50 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 10.795 | 10.399 | 10.777 | 10.610 | 11.217 |
| 9.866 | 9.975 | **9.010** | 9.863 | 10.929 |
| 7.841 | 7.267 | 8.513 | 8.214 | 8.712 |
| 11.869 | 10.340 | 11.380 | 10.312 | 11.007 |
| 10.350 | 12.547 | **11.221** | 10.324 | 10.532 |
| 9.299 | 10.392 | **10.102** | 10.320 | 9.449 |
| 10.534 | 9.264 | 10.164 | 9.067 | 9.447 |
| 10.325 | 11.979 | 11.486 | 10.526 | 11.554 |
| 10.288 | 10.920 | **9.927** | 9.554 | 11.840 |
| 9.232 | 9.984 | 10.538 | 9.633 | 9.045 |
| 11.463 | 10.285 | **10.718** | 9.156 | 9.243 |
| 9.427 | 10.861 | 9.573 | 9.717 | 8.939 |
| 10.678 | 9.843 | 10.905 | 10.302 | 9.628 |
| 9.580 | 9.948 | 9.478 | 10.324 | 9.885 |
| 9.714 | 9.214 | 9.334 | 10.042 | 9.996 |
| 8.282 | 8.433 | 9.356 | 9.677 | 8.955 |
| 11.562 | 10.166 | **11.504** | 10.953 | 10.491 |
| 9.588 | 10.713 | 10.476 | 11.278 | 11.123 |
| 9.649 | 10.292 | 9.565 | 10.365 | 9.811 |
| 10.100 | 10.191 | 9.732 | 10.977 | 9.444 |
| 12.278 | 11.190 | **12.263** | 10.723 | 11.435 |
| 9.723 | 11.318 | 11.123 | 11.680 | 10.760 |
| 10.240 | 9.289 | **10.329** | 9.904 | 9.946 |
| 9.526 | 9.516 | 11.707 | 10.888 | 10.849 |
| 9.059 | 9.980 | 8.240 | 10.071 | 10.326 |
| 8.777 | 9.674 | 9.730 | 10.410 | 9.548 |
| 10.328 | 10.406 | 10.584 | 10.678 | 10.698 |
| 10.047 | 9.038 | 9.562 | 9.427 | 9.446 |
| 10.290 | 9.460 | **10.655** | 9.544 | 9.785 |
| 9.312 | 10.242 | 9.415 | 10.194 | 9.982 |
| 9.330 | 8.964 | 9.607 | 9.561 | 9.740 |
| 9.819 | 9.472 | **9.547** | 9.324 | 9.188 |
| 9.774 | 9.301 | 10.327 | 10.016 | 9.132 |
| 9.706 | 9.902 | 10.165 | 10.196 | 10.329 |
| 9.645 | 9.857 | 10.916 | 10.587 | 9.147 |
| 11.296 | 11.196 | 10.420 | 10.252 | 10.928 |
| 9.854 | 9.483 | **9.509** | 9.731 | 10.447 |
| 9.163 | 9.153 | 11.430 | 10.506 | 10.708 |
| 9.435 | 9.901 | 9.737 | 10.184 | 10.011 |
| 10.232 | 9.714 | 9.208 | 9.834 | 9.961 |
| 9.658 | 8.187 | **9.189** | 8.847 | 9.840 |
| 9.695 | 9.276 | 10.903 | 10.868 | 10.161 |
| 11.174 | 12.345 | 11.321 | 11.366 | 11.804 |
| 9.630 | 11.485 | 11.574 | 12.158 | 11.666 |
| 9.131 | 10.067 | 9.754 | 9.340 | 9.765 |
| 10.164 | 9.141 | **9.549** | 9.524 | 10.820 |
| 9.455 | 10.444 | 9.792 | 10.016 | 10.999 |
| 10.790 | 9.637 | 9.035 | 9.795 | 9.584 |
| 11.428 | 10.079 | 11.551 | 10.164 | 10.742 |
| 10.463 | 9.852 | 9.813 | 9.842 | 10.429 |

El vector de medias de los datos originales es:



**Método de Eliminación por Filas**

Debido a que los datos faltantes recayeron en la variable *X*3 y son: el *X*2,3=9.010, *X*5,3=11.221, *X*6,3=10.102, *X*9,3=9.927, *X*11,3=10.718, *X*17,3=11.504, *X*21,3=12.263, *X*23,3=10.329, *X*29,3=10.655, *X*32,3=9.547, *X*37,3=9.509, *X*41,3=9.189 y el *X*46,3=9.549, se procede a prescindir de las filas que tienen estos valores “faltantes”, donde la matriz de datos resultante con filas eliminadas se muestra en la Tabla 4.22.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.22**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 10.795 | 10.399 | 10.777 | 10.610 | 11.217 |
| 7.841 | 7.267 | 8.513 | 8.214 | 8.712 |
| 11.869 | 10.340 | 11.380 | 10.312 | 11.007 |
| 10.534 | 9.264 | 10.164 | 9.067 | 9.447 |
| 10.325 | 11.979 | 11.486 | 10.526 | 11.554 |
| 9.232 | 9.984 | 10.538 | 9.633 | 9.045 |
| 9.427 | 10.861 | 9.573 | 9.717 | 8.939 |
| 10.678 | 9.843 | 10.905 | 10.302 | 9.628 |
| 9.580 | 9.948 | 9.478 | 10.324 | 9.885 |
| 9.714 | 9.214 | 9.334 | 10.042 | 9.996 |
| 8.282 | 8.433 | 9.356 | 9.677 | 8.955 |
| 9.588 | 10.713 | 10.476 | 11.278 | 11.123 |
| 9.649 | 10.292 | 9.565 | 10.365 | 9.811 |
| 10.100 | 10.191 | 9.732 | 10.977 | 9.444 |
| 9.723 | 11.318 | 11.123 | 11.680 | 10.760 |
| 9.526 | 9.516 | 11.707 | 10.888 | 10.849 |
| 9.059 | 9.980 | 8.240 | 10.071 | 10.326 |
| 8.777 | 9.674 | 9.730 | 10.410 | 9.548 |
| 10.328 | 10.406 | 10.584 | 10.678 | 10.698 |
| 10.047 | 9.038 | 9.562 | 9.427 | 9.446 |
| 9.312 | 10.242 | 9.415 | 10.194 | 9.982 |
| 9.330 | 8.964 | 9.607 | 9.561 | 9.740 |
| 9.774 | 9.301 | 10.327 | 10.016 | 9.132 |
| 9.706 | 9.902 | 10.165 | 10.196 | 10.329 |
| 9.645 | 9.857 | 10.916 | 10.587 | 9.147 |
| 11.296 | 11.196 | 10.420 | 10.252 | 10.928 |
| 9.163 | 9.153 | 11.430 | 10.506 | 10.708 |
| 9.435 | 9.901 | 9.737 | 10.184 | 10.011 |
| 10.232 | 9.714 | 9.208 | 9.834 | 9.961 |
| 9.695 | 9.276 | 10.903 | 10.868 | 10.161 |
| 11.174 | 12.345 | 11.321 | 11.366 | 11.804 |
| 9.630 | 11.485 | 11.574 | 12.158 | 11.666 |
| 9.131 | 10.067 | 9.754 | 9.340 | 9.765 |
| 9.455 | 10.444 | 9.792 | 10.016 | 10.999 |
| 10.790 | 9.637 | 9.035 | 9.795 | 9.584 |
| 11.428 | 10.079 | 11.551 | 10.164 | 10.742 |
| 10.463 | 9.852 | 9.813 | 9.842 | 10.429 |

**Elaborado por**: G. Cuenca

El vector de medias para las treinta y siete filas restantes es:



El vector de medias de los datos originales y de los datos con filas eliminadas no coincide.

Ahora analicemos en el Cuadro 4.20, el efecto que causa en la *matriz de varianzas y covarianzas*, y *matriz de correlaciones*, la eliminación de trece filas, con un tamaño de muestra *n*=50.

|  |  |
| --- | --- |
| **CUADRO 4.20**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.758 |  |  |  |  | | *X*2 | 0.387 | 0.953 |  |  |  | | *X*3 | 0.439 | 0.465 | 0.828 |  |  | | *X*4 | 0.135 | 0.439 | 0.396 | 0.517 |  | | *X*5 | 0.317 | 0.483 | 0.363 | 0.327 | 0.668 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.455 | 1.000 |  |  |  | | *X*3 | 0.554 | 0.524 | 1.000 |  |  | | *X*4 | 0.215 | 0.625 | 0.606 | 1.000 |  | | *X*5 | 0.445 | 0.606 | 0.488 | 0.556 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **(Trece Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.711 |  |  |  |  | | *X*2 | 0.399 | 0.898 |  |  |  | | *X*3 | 0.357 | 0.414 | 0.812 |  |  | | *X*4 | 0.163 | 0.470 | 0.411 | 0.533 |  | | *X*5 | 0.338 | 0.540 | 0.445 | 0.401 | 0.678 | | **Matriz de Correlaciones**  **(Trece Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.499 | 1.000 |  |  |  | | *X*3 | 0.470 | 0.484 | 1.000 |  |  | | *X*4 | 0.266 | 0.680 | 0.625 | 1.000 |  | | *X*5 | 0.487 | 0.693 | 0.600 | 0.667 | 1.000 | |

**Elaborado por**: G. Cuenca

Se puede apreciar que la mayor covarianza en la matriz de datos originales se da entre las variables *X2*  y *X5* es decir 0.483; mientras que en la matriz con tres filas eliminadas este valor aumenta a 0.540.

En la matriz de correlaciones de datos originales, la mayor correlación se da entre las variables *X2* y *X4*, es decir 0.625, cuyo valor se incrementa a 0.680 en la matriz de correlaciones con trece filas eliminadas. Se puede apreciar también que la mayor correlación en la matriz de datos con trece filas eliminadas se da entre las variables *X2* y*X5*, es decir 0.693. En general, se puede decir que las variables tienen una correlación fuerte.

También se realiza el análisis de la variable que presenta datos faltantes, en este caso la variable *X*3. (Ver Cuadro 4.21)

|  |  |
| --- | --- |
| **CUADRO 4.21**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****3*”** | |
| **Estimadores**   |  |  |  |  | | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Con el 26% de datos eliminadas en X3** | | n | | 50 | 37 | | Media | | 10,214 | 10,194 | | Mediana | | 10,133 | 10,164 | | Moda | | 8,240 | 8,240 | | Varianza | | 0,828 | 0,812 | | Desviación Estándar | | 0,910 | 0,901 | | Error Estándar | | 0,129 | 0,148 | | **Coeficiente de Asimetría** | | 0,173 | -0,012 | | Curtosis | | -0,639 | -0,759 | | Rango | | 4,020 | 3,470 | | Mínimo | | 8,240 | 8,240 | | Máximo | | 12,260 | 11,710 | | Percentiles | 25 | 9,549 | 9,564 | | 50 | 10,133 | 10,164 | | 75 | 10,907 | 10,911 | | **Diagrama de Cajas** |

**Elaborado por**: G. Cuenca

En el Cuadro 4.21, podemos apreciar que con el 26% de datos eliminados en la tercera columna de la matriz de datos (Variable *X****3*)**, el valor de la media aumentó de 10.214 a 10.194 , la varianza disminuyó de 0.828 a 0.812.

**Método de Imputación por la Media y Regresión**

A continuación se aplica el método de imputación por media y regresión a la misma matriz de datos utilizada en el método de eliminación por filas, es decir se completan datos en la variable *X*3 que presenta trece valores faltantes que son: el *X*2,3=9.010, *X*5,3=11.221, *X*6,3=10.102, *X*9,3=9.927, *X*11,3=10.718, *X*17,3=11.504, *X*21,3=12.263, *X*23,3=10.329, *X*29,3=10.655, *X*32,3=9.547, *X*37,3=9.509, *X*41,3=9.189 y el *X*46,3=9.549.

Por medio del *Método de Imputación por Media*, se procede a calcular la media aritmética de la variable *X3* con los trece datos faltantes, cuyo valor es 10.194, entonces reemplazamos en *X*2,3, *X*5,3, *X*6,3, *X*9,3, *X*11,3, *X*17,3, *X*21,3, *X*23,3, *X*29,3, *X*32,3, *X*37,3, *X*41,3 y en *X*46,3.

La matriz de datos resultante con trece valores completados por *imputación por la media* y *regresión* en la variable *X3* se muestra en la Tabla 4.23 y 4.24 respectivamente.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.23**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 10.795 | 10.399 | 10.777 | 10.610 | 11.217 |
| 9.866 | 9.975 | **10.194** | 9.863 | 10.929 |
| 7.841 | 7.267 | 8.513 | 8.214 | 8.712 |
| 11.869 | 10.340 | 11.380 | 10.312 | 11.007 |
| 10.350 | 12.547 | **10.194** | 10.324 | 10.532 |
| 9.299 | 10.392 | **10.194** | 10.320 | 9.449 |
| 10.534 | 9.264 | 10.164 | 9.067 | 9.447 |
| 10.325 | 11.979 | 11.486 | 10.526 | 11.554 |
| 10.288 | 10.920 | **10.194** | 9.554 | 11.840 |
| 9.232 | 9.984 | 10.538 | 9.633 | 9.045 |
| 11.463 | 10.285 | **10.194** | 9.156 | 9.243 |
| 9.427 | 10.861 | 9.573 | 9.717 | 8.939 |
| 10.678 | 9.843 | 10.905 | 10.302 | 9.628 |
| 9.580 | 9.948 | 9.478 | 10.324 | 9.885 |
| 9.714 | 9.214 | 9.334 | 10.042 | 9.996 |
| 8.282 | 8.433 | 9.356 | 9.677 | 8.955 |
| 11.562 | 10.166 | **10.194** | 10.953 | 10.491 |
| 9.588 | 10.713 | 10.476 | 11.278 | 11.123 |
| 9.649 | 10.292 | 9.565 | 10.365 | 9.811 |
| 10.100 | 10.191 | 9.732 | 10.977 | 9.444 |
| 12.278 | 11.190 | **10.194** | 10.723 | 11.435 |
| 9.723 | 11.318 | 11.123 | 11.680 | 10.760 |
| 10.240 | 9.289 | **10.194** | 9.904 | 9.946 |
| 9.526 | 9.516 | 11.707 | 10.888 | 10.849 |
| 9.059 | 9.980 | 8.240 | 10.071 | 10.326 |
| 8.777 | 9.674 | 9.730 | 10.410 | 9.548 |
| 10.328 | 10.406 | 10.584 | 10.678 | 10.698 |
| 10.047 | 9.038 | 9.562 | 9.427 | 9.446 |
| 10.290 | 9.460 | **10.194** | 9.544 | 9.785 |
| 9.312 | 10.242 | 9.415 | 10.194 | 9.982 |
| 9.330 | 8.964 | 9.607 | 9.561 | 9.740 |
| 9.819 | 9.472 | **10.194** | 9.324 | 9.188 |
| 9.774 | 9.301 | 10.327 | 10.016 | 9.132 |
| 9.706 | 9.902 | 10.165 | 10.196 | 10.329 |
| 9.645 | 9.857 | 10.916 | 10.587 | 9.147 |
| 11.296 | 11.196 | 10.420 | 10.252 | 10.928 |
| 9.854 | 9.483 | **10.194** | 9.731 | 10.447 |
| 9.163 | 9.153 | 11.430 | 10.506 | 10.708 |
| 9.435 | 9.901 | 9.737 | 10.184 | 10.011 |
| 10.232 | 9.714 | 9.208 | 9.834 | 9.961 |
| 9.658 | 8.187 | **10.194** | 8.847 | 9.840 |
| 9.695 | 9.276 | 10.903 | 10.868 | 10.161 |
| 11.174 | 12.345 | 11.321 | 11.366 | 11.804 |
| 9.630 | 11.485 | 11.574 | 12.158 | 11.666 |
| 9.131 | 10.067 | 9.754 | 9.340 | 9.765 |
| 10.164 | 9.141 | **10.194** | 9.524 | 10.820 |
| 9.455 | 10.444 | 9.792 | 10.016 | 10.999 |
| 10.790 | 9.637 | 9.035 | 9.795 | 9.584 |
| 11.428 | 10.079 | 11.551 | 10.164 | 10.742 |
| 10.463 | 9.852 | 9.813 | 9.842 | 10.429 |
| **Tabla 4.24**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de 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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 10.795 | 10.399 | 10.777 | 10.610 | 11.217 |
| 9.866 | 9.975 | **9.110** | 9.863 | 10.929 |
| 7.841 | 7.267 | 8.513 | 8.214 | 8.712 |
| 11.869 | 10.340 | 11.380 | 10.312 | 11.007 |
| 10.350 | 12.547 | **11.215** | 10.324 | 10.532 |
| 9.299 | 10.392 | **10.112** | 10.320 | 9.449 |
| 10.534 | 9.264 | 10.164 | 9.067 | 9.447 |
| 10.325 | 11.979 | 11.486 | 10.526 | 11.554 |
| 10.288 | 10.920 | **9.953** | 9.554 | 11.840 |
| 9.232 | 9.984 | 10.538 | 9.633 | 9.045 |
| 11.463 | 10.285 | **10.709** | 9.156 | 9.243 |
| 9.427 | 10.861 | 9.573 | 9.717 | 8.939 |
| 10.678 | 9.843 | 10.905 | 10.302 | 9.628 |
| 9.580 | 9.948 | 9.478 | 10.324 | 9.885 |
| 9.714 | 9.214 | 9.334 | 10.042 | 9.996 |
| 8.282 | 8.433 | 9.356 | 9.677 | 8.955 |
| 11.562 | 10.166 | **11.510** | 10.953 | 10.491 |
| 9.588 | 10.713 | 10.476 | 11.278 | 11.123 |
| 9.649 | 10.292 | 9.565 | 10.365 | 9.811 |
| 10.100 | 10.191 | 9.732 | 10.977 | 9.444 |
| 12.278 | 11.190 | **12.253** | 10.723 | 11.435 |
| 9.723 | 11.318 | 11.123 | 11.680 | 10.760 |
| 10.240 | 9.289 | **10.333** | 9.904 | 9.946 |
| 9.526 | 9.516 | 11.707 | 10.888 | 10.849 |
| 9.059 | 9.980 | 8.240 | 10.071 | 10.326 |
| 8.777 | 9.674 | 9.730 | 10.410 | 9.548 |
| 10.328 | 10.406 | 10.584 | 10.678 | 10.698 |
| 10.047 | 9.038 | 9.562 | 9.427 | 9.446 |
| 10.290 | 9.460 | **10.652** | 9.544 | 9.785 |
| 9.312 | 10.242 | 9.415 | 10.194 | 9.982 |
| 9.330 | 8.964 | 9.607 | 9.561 | 9.740 |
| 9.819 | 9.472 | **9.545** | 9.324 | 9.188 |
| 9.774 | 9.301 | 10.327 | 10.016 | 9.132 |
| 9.706 | 9.902 | 10.165 | 10.196 | 10.329 |
| 9.645 | 9.857 | 10.916 | 10.587 | 9.147 |
| 11.296 | 11.196 | 10.420 | 10.252 | 10.928 |
| 9.854 | 9.483 | **9.507** | 9.731 | 10.447 |
| 9.163 | 9.153 | 11.430 | 10.506 | 10.708 |
| 9.435 | 9.901 | 9.737 | 10.184 | 10.011 |
| 10.232 | 9.714 | 9.208 | 9.834 | 9.961 |
| 9.658 | 8.187 | **9.181** | 8.847 | 9.840 |
| 9.695 | 9.276 | 10.903 | 10.868 | 10.161 |
| 11.174 | 12.345 | 11.321 | 11.366 | 11.804 |
| 9.630 | 11.485 | 11.574 | 12.158 | 11.666 |
| 9.131 | 10.067 | 9.754 | 9.340 | 9.765 |
| 10.164 | 9.141 | **9.539** | 9.524 | 10.820 |
| 9.455 | 10.444 | 9.792 | 10.016 | 10.999 |
| 10.790 | 9.637 | 9.035 | 9.795 | 9.584 |
| 11.428 | 10.079 | 11.551 | 10.164 | 10.742 |
| 10.463 | 9.852 | 9.813 | 9.842 | 10.429 |

|  |
| --- |
| **Tabla 4.25**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 |
| ***26% de datos completados en*** *X3* ***por la Media***   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Imputación por Media** | **Error**  **|Dato Observado –**  **Resultado de Imputación por Media|** | | 9.010 | 10.194 | 1,184 | | 11.221 | 10.194 | 1,027 | | 10.102 | 10.194 | 0,092 | | 9.927 | 10.194 | 0,267 | | 10.718 | 10.194 | 0,524 | | 11.504 | 10.194 | 1,310 | | 12.263 | 10.194 | 2,069 | | 10.329 | 10.194 | 0,135 | | 10.655 | 10.194 | 0,461 | | 9.547 | 10.194 | 0,647 | | 9.509 | 10.194 | 0,685 | | 9.189 | 10.194 | 1,005 | | 9.549 | 10.194 | 0,645 | |
| ***26% de datos completados en*** *X3* ***por Regresión***   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Predicción** | **Error**  **| Dato Observado –**  **Resultado de Predicción |** | | 9.010 | 9.110 | 0,100 | | 11.221 | 11.215 | 0,006 | | 10.102 | 10.112 | 0,010 | | 9.927 | 9.931 | 0,004 | | 10.718 | 10.709 | 0,009 | | 11.504 | 11.510 | 0,006 | | 12.263 | 12.253 | 0,010 | | 10.329 | 10.333 | 0,004 | | 10.655 | 10.652 | 0,003 | | 9.547 | 9.545 | 0,002 | | 9.509 | 9.507 | 0,002 | | 9.189 | 9.181 | 0,008 | | 9.549 | 9.539 | 0,010 | |

**Elaborado por**: G. Cuenca

Se puede notar, por medio de la Tabla 4.25 que la diferencia en valor absoluto entre el dato observado de cada variable y el resultado de predicción, es menor en el *Método de Imputación por Regresión.*

|  |  |
| --- | --- |
| **CUADRO 4.22**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*2,3=9.010   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 9.980 | 0,970 | | 2 | 9.772 | 0,762 | | 3 | 9.564 | 0,554 | | 4 | 9.321 | 0,311 | | 5 | 9.110 | 0,100 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 9.549 | | Error Estándar | 0.155 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.539 | | Error Estándar | 0.155 | |

**Elaborado por**: G. Cuenca

En el Cuadro 4.22, se puede ver que el primer resultado de predicción es 9.980 ± 0.155, y el último es 9.110 ± 0.155, donde la media de los resultados de predicción es 9.549 ± 0.155.

|  |  |
| --- | --- |
| **CUADRO 4.23**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*5,3=11.221   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 12.550 | 1,329 | | 2 | 12.115 | 0,894 | | 3 | 11.731 | 0,510 | | 4 | 11.416 | 0,195 | | 5 | 11.215 | 0,006 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 11.805 | | Error Estándar | 0.240 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.587 | | Error Estándar | 0.239 | |

**Elaborado por**: G. Cuenca

En el Cuadro 4.23, se puede ver que el primer dato resultado de predicción es 12.550 ± 0.240, y el último es 11.215 ± 0.240, donde la media de los resultados de predicción es 11.805 ± 0.240 y la media del error de predicción es 0.587 ± 0.240.

|  |  |
| --- | --- |
| **CUADRO 4.24**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*6,3=10.102   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 10.390 | 0,288 | | 2 | 10.305 | 0,203 | | 3 | 10.252 | 0,150 | | 4 | 10.221 | 0,119 | | 5 | 10.112 | 0,010 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 10.256 | | Error Estándar | 0.046 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.154 | | Error Estándar | 0.046 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.24, nos muestra que el primer resultado de predicción es 10.390 ± 0.046, y el último es 10.112 ± 0.046, donde la media de los resultados de predicción es 10.256 ± 0.046 y la media del error de predicción es 0.154 ± 0.046.

|  |  |
| --- | --- |
| **CUADRO 4.25**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*9,3=9.927   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 10.541 | 0,614 | | 2 | 10.356 | 0,429 | | 3 | 10.112 | 0,185 | | 4 | 10.005 | 0,078 | | 5 | 9.931 | 0,004 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 10.189 | | Error Estándar | 0.114 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.262 | | Error Estándar | 0.114 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.25, nos muestra que el primer resultado de predicción es 10.541 ± 0.114, y el último es 9.931 ± 0.114, donde la media de los resultados de predicción es 10.189 ± 0.114 y la media del error de predicción es 0.262 ± 0.114.

|  |  |
| --- | --- |
| **CUADRO 4.26**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*11,3=10.718   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 10.290 | 0,428 | | 2 | 10.441 | 0,277 | | 3 | 10.563 | 0,155 | | 4 | 10.685 | 0,033 | | 5 | 10.709 | 0,009 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 10.538 | | Error Estándar | 0.078 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.184 | | Error Estándar | 0.078 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.26, nos muestra que el primer resultado de predicción es 10.290 ± 0.078, y el último es 10.709 ± 0.078, donde la media de los resultados de predicción es 10.538 ± 0.078 y la media del error de predicción es 0.184 ± 0.078.

|  |  |
| --- | --- |
| **CUADRO 4.27**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*17,3=11.504   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 11.835 | 0,331 | | 2 | 11.752 | 0,248 | | 3 | 11.631 | 0,127 | | 4 | 11.563 | 0,059 | | 5 | 11.510 | 0,006 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 11.658 | | Error Estándar | 0.060 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.154 | | Error Estándar | 0.060 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.27, nos muestra que el primer resultado de predicción es 11.835 ± 0.060, y el último es 11.510 ± 0.060, donde la media de los resultados de predicción es 11.658 ± 0.060 y la media del error de predicción es 0.154 ± 0.060.

|  |  |
| --- | --- |
| **CUADRO 4.28**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*21,3=12.263   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 11.630 | 0,633 | | 2 | 11.780 | 0,483 | | 3 | 11.852 | 0,411 | | 4 | 11.991 | 0,272 | | 5 | 12.253 | 0,010 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 11.901 | | Error Estándar | 0.105 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.362 | | Error Estándar | 0.105 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.28, nos muestra que el primer resultado de predicción es 11.630 ± 0.150, y el último es 12.253 ± 0.150, donde la media de los resultados de predicción es 11.901 ± 0.150 y la media del error de predicción es 0.362 ± 0.0105.

|  |  |
| --- | --- |
| **CUADRO 4.29**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*23,3=10.329   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 10.532 | 0,203 | | 2 | 10.451 | 0,122 | | 3 | 10.410 | 0,081 | | 4 | 10.365 | 0,036 | | 5 | 10.333 | 0,004 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 10.418 | | Error Estándar | 0.035 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.089 | | Error Estándar | 0.035 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.29, nos muestra que el primer resultado de predicción es 10.532 ± 0.035, y el último es 10.333 ± 0.035, donde la media de los resultados de predicción es 10.418 ± 0.035 y la media del error de predicción es 0.089 ± 0.035.

|  |  |
| --- | --- |
| **CUADRO 4.30**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*29,3=10.655   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 10.525 | 0,130 | | 2 | 10.573 | 0,082 | | 3 | 10.610 | 0,045 | | 4 | 10.630 | 0,025 | | 5 | 10.652 | 0,003 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 10.598 | | Error Estándar | 0.022 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.026 | | Error Estándar | 0.022 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.30, nos muestra que el primer resultado de predicción es 10.525 ± 0.022, y el último es 10.652 ± 0.022, donde la media de los resultados de predicción es 10.598 ± 0.022 y la media del error de predicción es 0.026 ± 0.022.

|  |  |
| --- | --- |
| **CUADRO 4.31**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*32,3=9.547   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 9.470 | 0,077 | | 2 | 9.493 | 0,054 | | 3 | 9.515 | 0,032 | | 4 | 9.532 | 0,015 | | 5 | 9.545 | 0,002 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 9.511 | | Error Estándar | 0.013 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.036 | | Error Estándar | 0.013 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.31, nos muestra que el primer resultado de predicción es 9.470 ± 0.013, y el último es 9.545 ± 0.013, donde la media de los resultados de predicción es 9.511 ± 0.013 y la media del error de predicción es 0.036 ± 0.013.

|  |  |
| --- | --- |
| **CUADRO 4.32**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*37,3=9.509   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 9.480 | 0,029 | | 2 | 9.489 | 0,020 | | 3 | 9.495 | 0,014 | | 4 | 9.501 | 0,008 | | 5 | 9.507 | 0,002 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 9.494 | | Error Estándar | 0.005 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.015 | | Error Estándar | 0.005 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.32, nos muestra que el primer resultado de predicción es 9.480 ± 0.005, y el último es 9.507 ± 0.005, donde la media de los resultados de predicción es 9.494 ± 0.005 y la media del error de predicción es 0.015±0.005.

|  |  |
| --- | --- |
| **CUADRO 4.33**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*41,3=9.189   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 8.790 | 0,399 | | 2 | 8.832 | 0,357 | | 3 | 8.951 | 0,238 | | 4 | 9.011 | 0,178 | | 5 | 9.181 | 0,008 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 8.953 | | Error Estándar | 0.069 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.236 | | Error Estándar | 0.069 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.33, nos muestra que el primer resultado de predicción es 8.790 ± 0.399, y el último es 9.181 ± 0.399, donde la media de los resultados de predicción es 8.953 ± 0.069 y la media del error de predicción es 0.253 ± 0.069.

|  |  |
| --- | --- |
| **CUADRO 4.34**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** *X*46,3=9.549   |  |  |  | | --- | --- | --- | | **Iteración** | **Resultado de**  **Predicción** | **Error**  **| Dato Observado – Resultado de Predicción |** | | 1 | 9.140 | 0,409 | | 2 | 9.326 | 0,223 | | 3 | 9.473 | 0,076 | | 4 | 9.516 | 0,033 | | 5 | 9.539 | 0,010 | | |
| **Distribución del Resultado de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Resultado de Predicción** | | Número de Iteración | 5 | | Media | 9.399 | | Error Estándar | 0.075 | | **Distribución del Error de Predicción**     |  |  | | --- | --- | | **Estimadores** | **Error de Predicción** | | Número de Iteración | 5 | | Media | 0.150 | | Error Estándar | 0.075 | |

**Elaborado por**: G. Cuenca

El Cuadro 4.34, nos muestra que el primer resultado de predicción es 9.140 ± 0.075, y el último es 9.539 ± 0.075, donde la media de los datos de predicción es 9.399 ± 0.075 y la media del error de predicción es 0.150 ± 0.075. Todos los resultados de predicción de los cuadros anteriores, tienden al dato observado.

|  |  |
| --- | --- |
| **CUADRO 4.35**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****3*”** | |
| **Estimadores**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 50 | 37 | 50 | 50 | | Media | | 10,214 | 10,194 | 10,194 | 10,216 | | Mediana | | 10,133 | 10,164 | 10,194 | 10,138 | | Moda | | 8,240 | 8,240 | 10,190 | 8,240 | | Varianza | | 0,828 | 0,812 | 0,597 | 0,823 | | Desviación Estándar | | 0,910 | 0,901 | 0,773 | 0,907 | | Error Estándar | | 0,129 | 0,148 | 0,109 | 0,128 | | **Coeficiente de Asimetría** | | 0,173 | -0,012 | -0,013 | 0,175 | | Curtosis | | -0,639 | -0,759 | 0,074 | -0,638 | | Rango | | 4,020 | 3,470 | 3,470 | 4,010 | | Mínimo | | 8,240 | 8,240 | 8,240 | 8,240 | | Máximo | | 12,260 | 11,710 | 11,710 | 12,250 | | Percentiles | 25 | 9,549 | 9,564 | 9,699 | 9,544 | | 50 | 10,133 | 10,164 | 10,194 | 10,138 | | 75 | 10,907 | 10,915 | 10,6322 | 10,908 | | **Diagrama de Cajas** |

**Elaborado por**: G. Cuenca

Al realizar la imputación por la media y regresión se obtuvieron los siguientes resultados (Ver Cuadro 4.35):

El valor de la media de los “datos completados” por *la media* disminuye. comparándolo con los “datos originales” y completados por *regresión.*

El valor de la varianza de los “datos completados” por la *media* disminuye de 0.828 a 0.597, mientras que en los “datos completados” por regresión este valor se incrementa a 0.823, comparándolo con el valor anterior y es muy cercano al valor de la varianza de los datos originales.

El vector de medias con trece datos completados por la media en *X3* es:



Mientras que el vector de medias con trece datos completados por la regresión en *X3* es:



El efecto que causa en la *matriz de varianzas* *y covarianzas* y *matriz de correlaciones*, el completar 5% de datos faltantes en una matriz de tamaño 50, por medio de la *imputación por media y regresión*, se presenta en el Cuadro 4.36.

|  |  |
| --- | --- |
| **CUADRO 4.36**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.758 |  |  |  |  | | *X*2 | 0.387 | 0.953 |  |  |  | | *X*3 | 0.439 | 0.465 | 0.828 |  |  | | *X*4 | 0.135 | 0.439 | 0.396 | 0.517 |  | | *X*5 | 0.317 | 0.483 | 0.363 | 0.327 | 0.668 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.455 | 1.000 |  |  |  | | *X*3 | 0.554 | 0.524 | 1.000 |  |  | | *X*4 | 0.215 | 0.625 | 0.606 | 1.000 |  | | *X*5 | 0.445 | 0.606 | 0.488 | 0.556 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **26% Datos Completados por Media en “*Variable*** *X****3*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.758 |  |  |  |  | | *X*2 | 0.387 | 0.953 |  |  |  | | *X*3 | 0.262 | 0.304 | 0.597 |  |  | | *X*4 | 0.135 | 0.439 | 0.302 | 0.517 |  | | *X*5 | 0.317 | 0.483 | 0.327 | 0.327 | 0.668 | | **Matriz de Correlaciones**  **26% Datos Completados por Media en “*Variable*** *X****3*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.455 | 1.000 |  |  |  | | *X*3 | 0.390 | 0.403 | 1.000 |  |  | | *X*4 | 0.215 | 0.625 | 0.544 | 1.000 |  | | *X*5 | 0.445 | 0.606 | 0.518 | 0.556 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **26% Datos Completados por Regresión en “*Variable*** *X****3*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 0.758 |  |  |  |  | | *X*2 | 0.387 | 0.953 |  |  |  | | *X*3 | 0.438 | 0.466 | 0.823 |  |  | | *X*4 | 0.135 | 0.439 | 0.396 | 0.517 |  | | *X*5 | 0.317 | 0.483 | 0.365 | 0.327 | 0.668 | | **Matriz de Correlaciones**  **26% Datos Completados por Regresiòn en “*Variable*** *X****3*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.455 | 1.000 |  |  |  | | *X*3 | 0.554 | 0.526 | 1.000 |  |  | | *X*4 | 0.215 | 0.625 | 0.607 | 1.000 |  | | *X*5 | 0.445 | 0.606 | 0.493 | 0.556 | 1.000 | |

**Elaborado por**: G. Cuenca

Analizando el Cuadro 4.36 se puede notar que la covarianza entre *X2*  y *X3* disminuye de 0.465 a 0.304 en la matriz con 26% de “datos completados” por la media en la variable *X3*, así como también la covarianza entre*X3* y *X4*  disminuye 0.396 a 0.302.

En la matriz de varianzas y covarianzas de los datos completados por regresión, el valor de las covarianzas de variable *X3* con las demás variables se incrementa, comparándolo con la matriz de varianzas y covarianzas de los datos completados por la media.

Por otro lado, analizando el efecto que causa en la matriz de correlaciones, podemos apreciar en el Cuadro 4.36 que la mayor correlación se da entre las variables *X2*  y *X4*, es decir 0.625, seguida por 0.606 entre las variables *X2*  y *X5*. En la matriz de correlaciones con 26% de datos completados por la media, la correlación entre *X1*  y *X3* disminuye de 0.554 a 0.390, mientras que en la matriz de datos completados por regresión, este valor es igual al de la matriz de datos originales es decir 0.554.

**4.3.2 Distribución Poisson: *Cincuenta datos faltantes* en una sola variable (10% de la matriz), tamaño de muestra n=100**

Se tiene una matriz de datos cuyas columnas son muestras tomadas de cinco poblaciones todas ellas Poisson, dependientes e idénticamente distribuidas, con parámetro , , *i= 1,2,....100*  y *j= 1,2,3*,*4,5* y se supone que tiene el 10% de datos faltantes, es decir cincuenta datos, los que recayeron en la variable *X*4 y son: el *X*1,1=11, *X*2,1=15, *X*4,1=15, *X*5,1=9, *X*8,1=8, *X*9,1=13, *X*10,1=8, *X*12,1=11, *X*15,1=13, *X*16,1=10, *X*18,1=9, *X*22,1=10, *X*23,1=12, *X*24,1=12, *X*25,1=10, *X*26,1=10, *X*27,1=19, *X*28,1=9, *X*30,1=8, *X*33,1=11, *X*34,1=10, *X*36,1=10, *X*39,1=9, *X*41,1=8, *X*44,1=9, *X*45,1=8, *X*47,1=11, *X*49,1=10, *X*51,1=9, *X*54,1=6, *X*55,1=12, *X*58,1=8, *X*60,1=8, *X*62,1=10, *X*64,1=12, *X*67,1=9, *X*69,1=9, *X*70,1=12, *X*72,1=10, *X*75,1=8, *X*79,1=4, *X*82,1=12, *X*85,1=14, *X*88,1=15, *X*90,1=9, *X*93,1=13, *X*95,1=11, *X*97,1=11, *X*99,1=11 y *X*100,1=8. Nótese que el 10% de datos faltantes en la matriz, constituye 50% de datos faltantes en la columna que corresponde a *X*4.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.26**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias dependientes con distribución Poisson**  Tamaño de muestra n=100 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 11 | 10 | 9 | **11** | 9 |
| 15 | 16 | 14 | **15** | 14 |
| 9 | 7 | 8 | 6 | 6 |
| 11 | 12 | 13 | **15** | 13 |
| 9 | 9 | 8 | **9** | 9 |
| 10 | 11 | 11 | 12 | 10 |
| 10 | 12 | 11 | 12 | 11 |
| 9 | 9 | 9 | **8** | 9 |
| 13 | 11 | 12 | **13** | 12 |
| 9 | 8 | 9 | **8** | 8 |
| 11 | 13 | 13 | 12 | 11 |
| 11 | 11 | 12 | **11** | 10 |
| 9 | 8 | 10 | 9 | 10 |
| 10 | 12 | 11 | 11 | 9 |
| 13 | 13 | 14 | **13** | 12 |
| 10 | 9 | 11 | **10** | 12 |
| 8 | 8 | 7 | 7 | 8 |
| 8 | 7 | 7 | **9** | 9 |
| 11 | 13 | 12 | 11 | 12 |
| 14 | 12 | 13 | 14 | 11 |
| 8 | 9 | 10 | 10 | 8 |
| 12 | 11 | 11 | **10** | 12 |
| 11 | 10 | 13 | **12** | 11 |
| 13 | 11 | 11 | **12** | 13 |
| 9 | 9 | 11 | **10** | 11 |
| 9 | 10 | 11 | **10** | 11 |
| 8 | 8 | 8 | **9** | 10 |
| 10 | 11 | 12 | **9** | 8 |
| 11 | 13 | 11 | 12 | 13 |
| 7 | 9 | 9 | **8** | 7 |
| 9 | 10 | 11 | 10 | 11 |
| 10 | 9 | 8 | 9 | 8 |
| 10 | 11 | 9 | **11** | 9 |
| 11 | 10 | 9 | **10** | 11 |

**Elaborado por**: G. Cuenca

**Continúa…**

**Viene…**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias dependientes con distribución Poisson**  Tamaño de muestra n=100 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 10 | 13 | 13 | 11 | 13 |
| 11 | 9 | 8 | **10** | 8 |
| 12 | 10 | 9 | 10 | 8 |
| 10 | 10 | 12 | 13 | 12 |
| 10 | 12 | 12 | **9** | 11 |
| 10 | 14 | 11 | 12 | 14 |
| 9 | 8 | 10 | **8** | 11 |
| 10 | 12 | 11 | 11 | 9 |
| 9 | 7 | 8 | 8 | 9 |
| 11 | 8 | 11 | **9** | 8 |
| 8 | 11 | 10 | **8** | 11 |
| 8 | 9 | 8 | 10 | 10 |
| 11 | 12 | 10 | **11** | 11 |
| 12 | 11 | 10 | 13 | 12 |
| 12 | 12 | 13 | **10** | 10 |
| 6 | 8 | 7 | 7 | 8 |
| 9 | 10 | 10 | **9** | 9 |
| 10 | 12 | 9 | 10 | 9 |
| 7 | 8 | 6 | 7 | 6 |
| 7 | 6 | 9 | **6** | 8 |
| 11 | 10 | 12 | **12** | 14 |
| 10 | 12 | 11 | 10 | 10 |
| 9 | 8 | 9 | 9 | 7 |
| 10 | 9 | 10 | **8** | 10 |
| 10 | 14 | 10 | 14 | 14 |
| 8 | 11 | 9 | **8** | 10 |
| 5 | 5 | 8 | 7 | 8 |
| 10 | 11 | 12 | **10** | 11 |
| 8 | 8 | 9 | 10 | 9 |
| 18 | 10 | 11 | **12** | 10 |
| 10 | 9 | 12 | 13 | 9 |
| 12 | 13 | 12 | 11 | 11 |
| 9 | 12 | 11 | **9** | 8 |
| 14 | 8 | 14 | 10 | 9 |
| 8 | 11 | 11 | **9** | 12 |
| 11 | 10 | 8 | **12** | 11 |
| 11 | 9 | 8 | 11 | 11 |
| 11 | 8 | 9 | **10** | 9 |
| 10 | 12 | 13 | 11 | 10 |
| 11 | 9 | 10 | 11 | 12 |
| 11 | 8 | 11 | **8** | 8 |
| 10 | 11 | 12 | 10 | 11 |
| 11 | 12 | 13 | 10 | 10 |
| 9 | 9 | 10 | 9 | 8 |
| 4 | 5 | 5 | **4** | 4 |
| 8 | 10 | 5 | 8 | 11 |
| 9 | 11 | 12 | 9 | 8 |
| 9 | 13 | 11 | **12** | 10 |
| 12 | 12 | 9 | 11 | 10 |

**Elaborado por**: G. Cuenca

**Viene…**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias dependientes con distribución Poisson**  Tamaño de muestra n=100 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 9 | 6 | 7 | 8 | 6 |
| 10 | 11 | 14 | **14** | 12 |
| 13 | 11 | 13 | 14 | 12 |
| 8 | 9 | 7 | 7 | 8 |
| 13 | 15 | 13 | **15** | 14 |
| 15 | 14 | 11 | 12 | 11 |
| 9 | 8 | 12 | **9** | 10 |
| 8 | 9 | 9 | 10 | 11 |
| 10 | 11 | 10 | 13 | 12 |
| 9 | 8 | 12 | **13** | 10 |
| 12 | 10 | 11 | 9 | 11 |
| 11 | 12 | 10 | **11** | 11 |
| 13 | 11 | 9 | 10 | 9 |
| 13 | 11 | 13 | **11** | 13 |
| 11 | 10 | 13 | 14 | 13 |
| 10 | 10 | 11 | **11** | 12 |
| 8 | 10 | 9 | **8** | 9 |

**Elaborado por**: G. Cuenca

El vector de medias de los datos originales es:



**Método de Eliminación por Filas**

Debido a que los datos faltantes recayeron en la variable *X*4, se procede a prescindir de las filas que tienen estos valores “faltantes”, donde la matriz de datos resultante con filas eliminadas se muestra en la Tabla 4.27.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.27**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 9 | 7 | 8 | 6 | 6 |
| 10 | 11 | 11 | 12 | 10 |
| 10 | 12 | 11 | 12 | 11 |
| 11 | 13 | 13 | 12 | 11 |
| 9 | 8 | 10 | 9 | 10 |
| 10 | 12 | 11 | 11 | 9 |
| 8 | 8 | 7 | 7 | 8 |
| 11 | 13 | 12 | 11 | 12 |
| 14 | 12 | 13 | 14 | 11 |
| 8 | 9 | 10 | 10 | 8 |
| 11 | 13 | 11 | 12 | 13 |
| 9 | 10 | 11 | 10 | 11 |
| 10 | 9 | 8 | 9 | 8 |
| 10 | 13 | 13 | 11 | 13 |
| 12 | 10 | 9 | 10 | 8 |
| 10 | 10 | 12 | 13 | 12 |
| 10 | 14 | 11 | 12 | 14 |
| 10 | 12 | 11 | 11 | 9 |
| 9 | 7 | 8 | 8 | 9 |
| 8 | 9 | 8 | 10 | 10 |
| 12 | 11 | 10 | 13 | 12 |
| 6 | 8 | 7 | 7 | 8 |
| 10 | 12 | 9 | 10 | 9 |
| 7 | 8 | 6 | 7 | 6 |
| 10 | 12 | 11 | 10 | 10 |
| 9 | 8 | 9 | 9 | 7 |
| 10 | 14 | 10 | 14 | 14 |
| 5 | 5 | 8 | 7 | 8 |
| 8 | 8 | 9 | 10 | 9 |
| 10 | 9 | 12 | 13 | 9 |
| 12 | 13 | 12 | 11 | 11 |
| 14 | 8 | 14 | 10 | 9 |
| 11 | 9 | 8 | 11 | 11 |
| 10 | 12 | 13 | 11 | 10 |
| 11 | 9 | 10 | 11 | 12 |
| 10 | 11 | 12 | 10 | 11 |
| 11 | 12 | 13 | 10 | 10 |
| 9 | 9 | 10 | 9 | 8 |
| 8 | 10 | 5 | 8 | 11 |
| 9 | 11 | 12 | 9 | 8 |
| 12 | 12 | 9 | 11 | 10 |
| 9 | 6 | 7 | 8 | 6 |
| 13 | 11 | 13 | 14 | 12 |
| 8 | 9 | 7 | 7 | 8 |
| 15 | 14 | 11 | 12 | 11 |
| 8 | 9 | 9 | 10 | 11 |
| 10 | 11 | 10 | 13 | 12 |
| 12 | 10 | 11 | 9 | 11 |
| 13 | 11 | 9 | 10 | 9 |
| 11 | 10 | 13 | 14 | 13 |

El vector de medias para las cincuenta filas restantes es:



Se analiza el efecto que causa en la *matriz de varianzas y covarianzas*, y *matriz de correlaciones*, la eliminación de cincuenta filas, con un tamaño de muestra *n*=100.(Ver Cuadro 4.37)

|  |  |
| --- | --- |
| **CUADRO 4.37**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.349 |  |  |  |  | | *X*2 | 2.400 | 4.364 |  |  |  | | *X*3 | 2.421 | 2.493 | 4.091 |  |  | | *X*4 | 2.927 | 2.986 | 2.851 | 4.563 |  | | *X*5 | 2.023 | 2.679 | 2.343 | **3.113** | 3.920 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.551 | 1.000 |  |  |  | | *X*3 | 0.574 | 0.590 | 1.000 |  |  | | *X*4 | 0.657 | 0.669 | 0.660 | 1.000 |  | | *X*5 | 0.490 | 0.648 | 0.585 | 0.736 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **(Cincuenta Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 3.835 |  |  |  |  | | *X*2 | 2.356 | 4.655 |  |  |  | | *X*3 | 2.443 | 2.572 | 4.490 |  |  | | *X*4 | 2.455 | 2.897 | 2.928 | 4.194 |  | | *X*5 | 1.613 | 2.863 | 2.146 | 3.069 | 3.979 | | **Matriz de Correlaciones**  **(Cincuenta Filas Eliminadas)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.558 | 1.000 |  |  |  | | *X*3 | 0.589 | 0.563 | 1.000 |  |  | | *X*4 | 0.612 | 0.656 | 0.675 | 1.000 |  | | *X*5 | 0.413 | 0.665 | 0.508 | 0.751 | 1.000 | |

**Elaborado por**: G. Cuenca

Se puede apreciar que la mayor covarianza en la matriz de datos originales se da entre las variables *X4*  y *X5* es decir 3.113; mientras que en la matriz con cincuenta filas eliminadas este valor aumenta a 3.069.

En la matriz de correlaciones de datos originales, la mayor correlación se da entre las variables *X4* y *X5*, es decir 0.736, cuyo valor se incrementa a 0.751 en la matriz de correlaciones con cincuenta filas eliminadas y por lo tanto se convierte en la mayor correlación.

|  |  |
| --- | --- |
| **CUADRO 4.38**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****4*”** | |
| **Estimadores**   |  |  |  |  | | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Con el 50% de datos eliminadas en X4** | | n | | 100 | 50 | | Media | | 10,270 | 10,360 | | Mediana | | 10,000 | 10,000 | | Moda | | 10,000 | 10,000 | | Varianza | | 4,563 | 4,194 | | Desviación Estándar | | 2,136 | 2,048 | | Error Estándar | | 0,214 | 0,290 | | **Coeficiente de Asimetría** | | 0,039 | -0,080 | | Curtosis | | 0,058 | -0,500 | | Rango | | 11,000 | 8,000 | | Mínimo | | 4,000 | 6,000 | | Máximo | | 15,000 | 14,000 | | Percentiles | 25 | 9,000 | 9,000 | | 50 | 10,000 | 10,000 | | 75 | 12,000 | 12,000 | | **Diagrama de Cajas** |

**Elaborado por**: G. Cuenca

En el Cuadro 4.38, podemos apreciar que con el 50% de datos eliminados en la cuarta columna de la matriz de datos (Variable *X****4*)**, el valor de la media aumentó de 10.270 a 10.360. La varianza de la variable *X****4***, con 50% de datos eliminadosdisminuyó de 4.536 a 4.194.

**Método de Imputación por la Media y Regresión**

A continuación se aplica el *método de imputación por media y regresión* a la misma matriz de datos utilizada en el método de eliminación por filas, es decir se completan datos en la variable *X*4 que presenta cincuenta valores faltantes que son: el *X*1,1=11, *X*2,1=15, *X*4,1=15, *X*5,1=9, *X*8,1=8, *X*9,1=13, *X*10,1=8, *X*12,1=11, *X*15,1=13, *X*16,1=10, *X*18,1=9, *X*22,1=10, *X*23,1=12, *X*24,1=12, *X*25,1=10, *X*26,1=10, *X*27,1=19, *X*28,1=9, *X*30,1=8, *X*33,1=11, *X*34,1=10, *X*36,1=10, *X*39,1=9, *X*41,1=8, *X*44,1=9, *X*45,1=8, *X*47,1=11, *X*49,1=10, *X*51,1=9, *X*54,1=6, *X*55,1=12, *X*58,1=8, *X*60,1=8, *X*62,1=10, *X*64,1=12, *X*67,1=9, *X*69,1=9, *X*70,1=12, *X*72,1=10, *X*75,1=8, *X*79,1=4, *X*82,1=12, *X*85,1=14, *X*88,1=15, *X*90,1=9, *X*93,1=13, *X*95,1=11, *X*97,1=11, *X*99,1=11 y *X*100,1=8.

Por medio del *Método de Imputación por Media*, se procede a calcular la media aritmética de la variable *X4* con los cincuenta datos faltantes, cuyo valor es 10.360 y se reemplaza en los datos faltantes descritos anteriormente. La matriz de datos resultante con cincuenta valores completados por *imputación por la media* y *regresión* en la variable *X4* se muestra en la Tabla 4.28 y 4.29 respectivamente.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tabla 4.28**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 11 | 10 | 9 | **10.360** | 9 |
| 15 | 16 | 14 | **10.360** | 14 |
| 9 | 7 | 8 | 6 | 6 |
| 11 | 12 | 13 | **10.360** | 13 |
| 9 | 9 | 8 | **10.360** | 9 |
| 10 | 11 | 11 | 12 | 10 |
| 10 | 12 | 11 | 12 | 11 |
| 9 | 9 | 9 | **10.360** | 9 |
| 13 | 11 | 12 | **10.360** | 12 |
| 9 | 8 | 9 | **10.360** | 8 |
| 11 | 13 | 13 | 12 | 11 |
| 11 | 11 | 12 | **10.360** | 10 |
| 9 | 8 | 10 | 9 | 10 |
| 10 | 12 | 11 | 11 | 9 |
| 13 | 13 | 14 | **10.360** | 12 |
| 10 | 9 | 11 | **10.360** | 12 |
| 8 | 8 | 7 | 7 | 8 |
| 8 | 7 | 7 | **10.360** | 9 |
| 11 | 13 | 12 | 11 | 12 |
| 14 | 12 | 13 | 14 | 11 |
| 8 | 9 | 10 | 10 | 8 |
| 12 | 11 | 11 | **10.360** | 12 |
| 11 | 10 | 13 | **10.360** | 11 |
| 13 | 11 | 11 | **10.360** | 13 |
| 9 | 9 | 11 | **10.360** | 11 |
| 9 | 10 | 11 | **10.360** | 11 |
| 8 | 8 | 8 | **10.360** | 10 |
| 10 | 11 | 12 | **10.360** | 8 |
| 11 | 13 | 11 | 12 | 13 |
| 7 | 9 | 9 | **10.360** | 7 |
| 9 | 10 | 11 | 10 | 11 |
| 10 | 9 | 8 | 9 | 8 |
| 10 | 11 | 9 | **10.360** | 9 |
| 11 | 10 | 9 | **10.360** | 11 |
| 10 | 13 | 13 | 11 | 13 |
| 11 | 9 | 8 | **10.360** | 8 |
| 12 | 10 | 9 | 10 | 8 |
| 10 | 10 | 12 | 13 | 12 |
| 10 | 12 | 12 | **10.360** | 11 |
| 10 | 14 | 11 | 12 | 14 |
| 9 | 8 | 10 | **10.360** | 11 |
| 10 | 12 | 11 | 11 | 9 |
| 9 | 7 | 8 | 8 | 9 |
| 11 | 8 | 11 | **10.360** | 8 |
| 8 | 11 | 10 | **10.360** | 11 |
| 8 | 9 | 8 | 10 | 10 |
| 11 | 12 | 10 | **10.360** | 11 |
| 12 | 11 | 10 | 13 | 12 |
| 12 | 12 | 13 | **10.360** | 10 |

**Continúa…**

**Viene…**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 6 | 8 | 7 | 7 | 8 |
| 9 | 10 | 10 | **10.360** | 9 |
| 10 | 12 | 9 | 10 | 9 |
| 7 | 8 | 6 | 7 | 6 |
| 7 | 6 | 9 | **10.360** | 8 |
| 11 | 10 | 12 | **10.360** | 14 |
| 10 | 12 | 11 | 10 | 10 |
| 9 | 8 | 9 | 9 | 7 |
| 10 | 9 | 10 | **10.360** | 10 |
| 10 | 14 | 10 | 14 | 14 |
| 8 | 11 | 9 | **10.360** | 10 |
| 5 | 5 | 8 | 7 | 8 |
| 10 | 11 | 12 | **10.360** | 11 |
| 8 | 8 | 9 | 10 | 9 |
| 18 | 10 | 11 | **10.360** | 10 |
| 10 | 9 | 12 | 13 | 9 |
| 12 | 13 | 12 | 11 | 11 |
| 9 | 12 | 11 | **10.360** | 8 |
| 14 | 8 | 14 | 10 | 9 |
| 8 | 11 | 11 | **10.360** | 12 |
| 11 | 10 | 8 | **10.360** | 11 |
| 11 | 9 | 8 | 11 | 11 |
| 11 | 8 | 9 | **10.360** | 9 |
| 10 | 12 | 13 | 11 | 10 |
| 11 | 9 | 10 | 11 | 12 |
| 11 | 8 | 11 | **10.360** | 8 |
| 10 | 11 | 12 | 10 | 11 |
| 11 | 12 | 13 | 10 | 10 |
| 9 | 9 | 10 | 9 | 8 |
| 4 | 5 | 5 | **10.360** | 4 |
| 8 | 10 | 5 | 8 | 11 |
| 9 | 11 | 12 | 9 | 8 |
| 9 | 13 | 11 | **10.360** | 10 |
| 12 | 12 | 9 | 11 | 10 |
| 9 | 6 | 7 | 8 | 6 |
| 10 | 11 | 14 | **10.360** | 12 |
| 13 | 11 | 13 | 14 | 12 |
| 8 | 9 | 7 | 7 | 8 |
| 13 | 15 | 13 | **10.360** | 14 |
| 15 | 14 | 11 | 12 | 11 |
| 9 | 8 | 12 | **10.360** | 10 |
| 8 | 9 | 9 | 10 | 11 |
| 10 | 11 | 10 | 13 | 12 |
| 9 | 8 | 12 | **10.360** | 10 |
| 12 | 10 | 11 | 9 | 11 |
| 11 | 12 | 10 | **10.360** | 11 |
| 13 | 11 | 9 | 10 | 9 |
| 13 | 11 | 13 | **10.360** | 13 |
| 11 | 10 | 13 | 14 | 13 |
| 10 | 10 | 11 | **10.360** | 12 |
| 8 | 10 | 9 | **10.360** | 9 |
| **Tabla 4.29**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 11 | 10 | 9 | **10,979** | 9 |
| 15 | 16 | 14 | **15,064** | 14 |
| 9 | 7 | 8 | 6 | 6 |
| 11 | 12 | 13 | **14,987** | 13 |
| 9 | 9 | 8 | **9,057** | 9 |
| 10 | 11 | 11 | 12 | 10 |
| 10 | 12 | 11 | 12 | 11 |
| 9 | 9 | 9 | **8,514** | 9 |
| 13 | 11 | 12 | **12,995** | 12 |
| 9 | 8 | 9 | **8,091** | 8 |
| 11 | 13 | 13 | 12 | 11 |
| 11 | 11 | 12 | **11,015** | 10 |
| 9 | 8 | 10 | 9 | 10 |
| 10 | 12 | 11 | 11 | 9 |
| 13 | 13 | 14 | **13,048** | 12 |
| 10 | 9 | 11 | **10,031** | 12 |
| 8 | 8 | 7 | 7 | 8 |
| 8 | 7 | 7 | **8,982** | 9 |
| 11 | 13 | 12 | 11 | 12 |
| 14 | 12 | 13 | 14 | 11 |
| 8 | 9 | 10 | 10 | 8 |
| 12 | 11 | 11 | **10,018** | 12 |
| 11 | 10 | 13 | **11,924** | 11 |
| 13 | 11 | 11 | **12,081** | 13 |
| 9 | 9 | 11 | **10,005** | 11 |
| 9 | 10 | 11 | **10,012** | 11 |
| 8 | 8 | 8 | **9,071** | 10 |
| 10 | 11 | 12 | **9,100** | 8 |
| 11 | 13 | 11 | 12 | 13 |
| 7 | 9 | 9 | **8,005** | 7 |
| 9 | 10 | 11 | 10 | 11 |
| 10 | 9 | 8 | 9 | 8 |
| 10 | 11 | 9 | **10,985** | 9 |
| 11 | 10 | 9 | **10,972** | 11 |
| 10 | 13 | 13 | 11 | 13 |
| 11 | 9 | 8 | **9,901** | 8 |
| 12 | 10 | 9 | 10 | 8 |
| 10 | 10 | 12 | 13 | 12 |
| 10 | 12 | 12 | **9,172** | 11 |
| 10 | 14 | 11 | 12 | 14 |
| 9 | 8 | 10 | **8,051** | 11 |
| 10 | 12 | 11 | 11 | 9 |
| 9 | 7 | 8 | 8 | 9 |
| 11 | 8 | 11 | **9,053** | 8 |
| 8 | 11 | 10 | **8,003** | 11 |
| 8 | 9 | 8 | 10 | 10 |
| 11 | 12 | 10 | **11,072** | 11 |
| 12 | 11 | 10 | 13 | 12 |

**Continúa…**

**Sigue…**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* |
| 12 | 12 | 13 | **10,030** | 10 |
| 6 | 8 | 7 | 7 | 8 |
| 9 | 10 | 10 | **9,022** | 9 |
| 10 | 12 | 9 | 10 | 9 |
| 7 | 8 | 6 | 7 | 6 |
| 7 | 6 | 9 | **5,987** | 8 |
| 11 | 10 | 12 | **12,101** | 14 |
| 10 | 12 | 11 | 10 | 10 |
| 9 | 8 | 9 | 9 | 7 |
| 10 | 9 | 10 | **7,983** | 10 |
| 10 | 14 | 10 | 14 | 14 |
| 8 | 11 | 9 | **8,003** | 10 |
| 5 | 5 | 8 | 7 | 8 |
| 10 | 11 | 12 | **10,002** | 11 |
| 8 | 8 | 9 | 10 | 9 |
| 18 | 10 | 11 | **11,978** | 10 |
| 10 | 9 | 12 | 13 | 9 |
| 12 | 13 | 12 | 11 | 11 |
| 9 | 12 | 11 | **9,062** | 8 |
| 14 | 8 | 14 | 10 | 9 |
| 8 | 11 | 11 | **9,051** | 12 |
| 11 | 10 | 8 | **11,971** | 11 |
| 11 | 9 | 8 | 11 | 11 |
| 11 | 8 | 9 | **10,101** | 9 |
| 10 | 12 | 13 | 11 | 10 |
| 11 | 9 | 10 | 11 | 12 |
| 11 | 8 | 11 | **8,106** | 8 |
| 10 | 11 | 12 | 10 | 11 |
| 11 | 12 | 13 | 10 | 10 |
| 9 | 9 | 10 | 9 | 8 |
| 4 | 5 | 5 | **4,031** | 4 |
| 8 | 10 | 5 | 8 | 11 |
| 9 | 11 | 12 | 9 | 8 |
| 9 | 13 | 11 | **11,931** | 10 |
| 12 | 12 | 9 | 11 | 10 |
| 9 | 6 | 7 | 8 | 6 |
| 10 | 11 | 14 | **13,920** | 12 |
| 13 | 11 | 13 | 14 | 12 |
| 8 | 9 | 7 | 7 | 8 |
| 13 | 15 | 13 | **14,933** | 14 |
| 15 | 14 | 11 | 12 | 11 |
| 9 | 8 | 12 | **9,010** | 10 |
| 8 | 9 | 9 | 10 | 11 |
| 10 | 11 | 10 | 13 | 12 |
| 9 | 8 | 12 | **12,915** | 10 |
| 12 | 10 | 11 | 9 | 11 |
| 11 | 12 | 10 | **10,993** | 11 |
| 13 | 11 | 9 | 10 | 9 |
| 13 | 11 | 13 | **11,061** | 13 |
| 11 | 10 | 13 | 14 | 13 |
| 10 | 10 | 11 | **11,076** | 12 |
| 8 | 10 | 9 | **8,003** | 9 |
| **Tabla 4.30**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | | |
| ***50% de datos completados en X3 por la Media***   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Imputación por Media** | **Error**  **|Dato Observado –**  **Resultado de Imputación por Media|** | | 11 | 10.360 | 0,64 | | 15 | 10.360 | 4,64 | | 15 | 10.360 | 4,64 | | 9 | 10.360 | 1,36 | | 8 | 10.360 | 2,36 | | 13 | 10.360 | 2,64 | | 8 | 10.360 | 2,36 | | 11 | 10.360 | 0,64 | | 13 | 10.360 | 2,64 | | 10 | 10.360 | 0,36 | | 9 | 10.360 | 1,36 | | 10 | 10.360 | 0,36 | | 12 | 10.360 | 1,64 | | 12 | 10.360 | 1,64 | | 10 | 10.360 | 0,36 | | 10 | 10.360 | 0,36 | | 9 | 10.360 | 1,36 | | 9 | 10.360 | 1,36 | | 8 | 10.360 | 2,36 | | 11 | 10.360 | 0,64 | | 10 | 10.360 | 0,36 | | 10 | 10.360 | 0,36 | | 9 | 10.360 | 1,36 | | 8 | 10.360 | 2,36 | | 9 | 10.360 | 1,36 | | 8 | 10.360 | 2,36 | | 11 | 10.360 | 0,64 | | 10 | 10.360 | 0,36 | | 9 | 10.360 | 1,36 | | 6 | 10.360 | 4,36 | | 12 | 10.360 | 1,64 | | 8 | 10.360 | 2,36 | | 8 | 10.360 | 2,36 | | 10 | 10.360 | 0,36 | | 12 | 10.360 | 1,64 | | 9 | 10.360 | 1,36 | | 9 | 10.360 | 1,36 | | 12 | 10.360 | 1,64 | | 10 | 10.360 | 0,36 | | 8 | 10.360 | 2,36 | | 4 | 10.360 | 6,36 | | 12 | 10.360 | 1,64 | | 14 | 10.360 | 3,64 | | 15 | 10.360 | 4,64 | | 9 | 10.360 | 1,36 | | 13 | 10.360 | 2,64 | | 11 | 10.360 | 0,64 | | 11 | 10.360 | 0,64 | | 11 | 10.360 | 0,64 | | 8 | 10.360 | 2,36 | | | | | | |

**Viene…**

|  |
| --- |
| **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 |
| *50% de datos completados en X3 por Regresión*   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Predicción** | **Error**  **| Dato Observado –**  **Resultado de Predicción |** | | 11 | 10.979 | 0,021 | | 15 | 15.064 | 0,064 | | 15 | 14.987 | 0,013 | | 9 | 9.057 | 0,057 | | 8 | 8.514 | 0,514 | | 13 | 12.995 | 0,005 | | 8 | 8.091 | 0,091 | | 11 | 11.015 | 0,015 | | 13 | 13.048 | 0,048 | | 10 | 10.031 | 0,031 | | 9 | 8.982 | 0,018 | | 10 | 10.018 | 0,018 | | 12 | 11.924 | 0,076 | | 12 | 12.081 | 0,081 | | 10 | 10.005 | 0,005 | | 10 | 10.012 | 0,012 | | 9 | 9.071 | 0,071 | | 9 | 9.100 | 0,100 | | 8 | 8.005 | 0,005 | | 11 | 10.985 | 0,015 | | 10 | 10.972 | 0,972 | | 10 | 8.901 | 1,099 | | 9 | 9.172 | 0,172 | | 8 | 8.051 | 0,051 | | 9 | 9.053 | 0,053 | | 8 | 8.003 | 0,003 | | 11 | 11.072 | 0,072 | | 10 | 10.030 | 0,030 | | 9 | 9.022 | 0,022 | | 6 | 5.987 | 0,013 | | 12 | 12.101 | 0,101 | | 8 | 7.983 | 0,017 | | 8 | 8.003 | 0,003 | | 10 | 10.002 | 0,002 | | 12 | 11.978 | 0,022 | | 9 | 9.062 | 0,062 | | 9 | 9.051 | 0,051 | | 12 | 11.971 | 0,029 | | 10 | 10.101 | 0,101 | | 8 | 8.106 | 0,106 | | 4 | 4.031 | 0,031 | | 12 | 11.931 | 0,069 | | 14 | 13.920 | 0,080 | | 15 | 14.933 | 0,067 | | 9 | 9.010 | 0,010 | | 13 | 12.915 | 0,085 | | 11 | 10.993 | 0,007 | | 11 | 11.061 | 0,061 | | 11 | 11.076 | 0,076 | | 8 | 8.003 | 0,003 | |

Se puede notar, por medio de la Tabla 4.30 que la diferencia en valor absoluto entre el valor observado de cada variable, es menor en el *Método de Imputación por Regresión.*

|  |  |
| --- | --- |
| **CUADRO 4.39**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****4*”** | |
| **Estimadores**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 100 | 50 | 100 | 100 | | Media | | 10,270 | 10,360 | 10,360 | 10,295 | | Mediana | | 10,000 | 10,000 | 10,360 | 10,004 | | Moda | | 10,000 | 10,000 | 10,360 | 10,000 | | Varianza | | 4,563 | 4,194 | 2,076 | 4,510 | | Desviación Estándar | | 2,136 | 2,048 | 1,441 | 2,124 | | Error Estándar | | 0,214 | 0,290 | 0,144 | 0,212 | | **Coeficiente de Asimetría** | | 0,039 | -0,080 | -0,111 | 0,016 | | Curtosis | | 0,058 | -0,500 | 2,022 | 0,087 | | Rango | | 11,000 | 8,000 | 8,000 | 11,030 | | Mínimo | | 4,000 | 6,000 | 6,000 | 4,030 | | Máximo | | 15,00 | 14,000 | 14,000 | 15,060 | | Percentiles | 25 | 9,000 | 9,000 | 10,000 | 9,000 | | 50 | 10,000 | 10,000 | 10,360 | 10,004 | | 75 | 12,000 | 12,000 | 10,360 | 11,961 | | **Diagrama de Cajas** |

**Elaborado por**: G. Cuenca

Al realizar la imputación por la media y regresión se obtuvieron los siguientes resultados (Ver Cuadro 4.39):

El valor de la media de los “datos completados” por *la media* aumenta, comparándolo con los “datos originales” y completados por *regresión.*

El valor de la varianza de los “datos completados” por la *media* disminuye de 4.563 a 2.076, mientras que en los datos completados por regresión este valor se incrementa a 4.510, comparándolo con el valor anterior y es muy cercano al valor de la varianza de los “datos originales”.

El vector de medias con cincuenta datos completados por la media en *X4* es:



Mientras que el vector de medias con cincuenta datos completados por la regresión en *X4* es:



El efecto que causa en la *matriz de varianzas* y *covarianzas* y *matriz de correlaciones*, el completar 10% de datos faltantes en una matriz de tamaño 100, por medio de la imputación por media y regresión, se presenta en el Cuadro 4.40.

|  |  |
| --- | --- |
| **CUADRO 4.40**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.349 |  |  |  |  | | *X*2 | 2.400 | 4.364 |  |  |  | | *X*3 | 2.421 | 2.493 | 4.091 |  |  | | *X*4 | 2.927 | 2.986 | 2.851 | 4.563 |  | | *X*5 | 2.023 | 2.679 | 2.343 | **3.113** | 3.920 | | **Matriz de Correlaciones**  **(Datos Originales)**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.551 | 1.000 |  |  |  | | *X*3 | 0.574 | 0.590 | 1.000 |  |  | | *X*4 | 0.657 | 0.669 | 0.660 | 1.000 |  | | *X*5 | 0.490 | 0.648 | 0.585 | 0.736 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **50% Datos Completados por Media en “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.349 |  |  |  |  | | *X*2 | 2.400 | 4.364 |  |  |  | | *X*3 | 2.421 | 2.493 | 4.091 |  |  | | *X*4 | 1.215 | 1.434 | 1.449 | 2.076 |  | | *X*5 | 2.023 | 2.679 | 2.343 | 1.519 | 3.920 | | **Matriz de Correlaciones**  **50% Datos Completados por Media en “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.551 | 1.000 |  |  |  | | *X*3 | 0.574 | 0.590 | 1.000 |  |  | | *X*4 | 0.404 | 0.476 | 0.497 | 1.000 |  | | *X*5 | 0.490 | 0.648 | 0.585 | 0.532 | 1.000 | |
| **Matriz de Varianzas y Covarianzas**  **50% Datos Completados por Regresión en “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 4.349 |  |  |  |  | | *X*2 | 2.400 | 4.364 |  |  |  | | *X*3 | 2.421 | 2.493 | 4.091 |  |  | | *X*4 | 2.931 | 2.976 | 2.834 | 4.510 |  | | *X*5 | 2.023 | 2.679 | 2.343 | 3.118 | 3.920 | | **Matriz de Correlaciones**  **50% Datos Completados por Regresiòn en “*Variable*** *X****4*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | | *X*1 | 1.000 |  |  |  |  | | *X*2 | 0.551 | 1.000 |  |  |  | | *X*3 | 0.574 | 0.590 | 1.000 |  |  | | *X*4 | 0.662 | 0.671 | 0.660 | 1.000 |  | | *X*5 | 0.490 | 0.648 | 0.585 | 0.742 | 1.000 | |

**Elaborado por**: G. Cuenca

La covarianza entre *X4*  y *X5* disminuye de 3.113 a 1.159 en la matriz con 50% de “datos completados” por la media en la variable *X4*, así como también disminuye la covarianza entre*X4*  con las otras variables.

En la matriz de varianzas y covarianzas de los datos completados por regresión, el valor de las covarianzas de variable *X4* con las demás variables se incrementa, comparándolo con la matriz de varianzas y covarianzas de los “datos completados” por *la media.*

Por otro lado, analizando el efecto que causa en la matriz de correlaciones, podemos apreciar en le Cuadro 4.40 que también los únicos valores que cambian son los de la correlación de *X4* con las demás variables, puesto que a esta variable se le completó datos por medio de los métodos de imputación; donde la mayor correlación se da entre las variables *X4*  y *X5* , es decir 0.736, seguida por 0.669 entre las variables *X2*  y *X4*. En la matriz de correlaciones con 50% de datos completados por la media, la correlación entre *X4*  y *X5* disminuye de 0.736 a 0.532, mientras que en la matriz de datos completados por regresión, este valor es 0.742.

**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**

Se tiene una matriz de datos cuyas columnas son muestras tomadas de diez poblaciones todas ellas Exponencial, dependientes e idénticamente distribuidas, con parámetro , , *i= 1,2,....100* y *j= 1,2,3*,*…,10* y se supone que tiene el 5% de datos faltantes, es decir cincuenta datos, los que recayeron en las variables *X*3 y *X*8 y son: el *X*3,3=2.851, *X*9,3=1.414, *X*15,3=1.069, *X*18,3=6.462, *X*21,3=3.914, *X*24,3=1.131, *X*31,3=6.562, *X*33,3=2.254, *X*39,3=1.689, *X*42,3=1.432, *X*43,3=3.693, *X*47,3=3.960, *X*48,3=3.420, *X*52,3=2.683, *X*55,3=6.730, *X*58,3=0.860, *X*59,3=6.406, *X*67,3=3.578, *X*69,3=5.157, *X*71,3=4.083, *X*74,3=2.061, *X*79,3=1.148, *X*81,3=3.359, *X*84,3=1.913, *X*86,3=1.351, *X*6,8=2.390, *X*12,8=1.060, *X*17,8=1.383, *X*23,8=1.219, *X*30,8=2.582, *X*34,8=5.997, *X*37,8=3.952, *X*41,8=19.664, *X*46,8=5.859, *X*50,8=5.255, *X*53,8=9.518, *X*60,8=2.947, *X*61,8=2.566, *X*62,8=0.929, *X*63,8=4.580, *X*75,8=2.080, *X*77,8=3.767, *X*87,8=4.930, *X*88,8=6.314, *X*92,8=0.704, *X*93,8=5.413, *X*97,8=3.183, *X*98,8=4.859, *X*99,8=4.800 y *X*100,8=5.525.

Nótese que el 5% de datos faltantes en la matriz, constituye 25% de datos faltantes en la columna que corresponde a *X*3 y 25% de datos faltantes en la columna *X*8 (Ver Tabla 4.31)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tabla 4.31**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias dependientes con distribución Exponencial**  Tamaño de muestra n=100 | | | | | | | | | |
| *X1* | *X2* | ***X3*** | *X4* | *X5* | *X6* | *X7* | ***X8*** | *X9* | *X10* |
| 6.726 | 6.168 | 3.447 | 4.124 | 4.017 | 4.550 | 5.149 | 4.957 | 6.743 | 3.346 |
| 1.168 | 1.763 | 0.622 | 2.786 | 4.782 | 3.397 | 3.994 | 1.921 | 1.714 | 2.373 |
| 3.238 | 4.557 | 2.851 | 3.335 | 0.641 | 10.599 | 10.406 | 11.662 | 0.222 | 10.237 |
| 0.283 | 0.163 | 0.814 | 2.302 | 1.101 | 2.715 | 0.470 | 0.462 | 0.814 | 2.980 |
| 3.054 | 1.277 | 3.099 | 1.934 | 0.206 | 1.929 | 0.575 | 1.089 | 0.289 | 1.435 |
| 3.483 | 3.547 | 3.129 | 5.710 | 3.334 | 3.645 | 5.478 | 2.390 | 4.686 | 3.469 |
| 0.668 | 1.180 | 3.188 | 2.429 | 4.009 | 3.122 | 2.252 | 1.105 | 4.255 | 2.085 |
| 2.576 | 2.268 | 3.545 | 1.127 | 2.069 | 3.408 | 3.349 | 3.863 | 2.491 | 3.414 |
| 4.385 | 1.285 | 1.414 | 1.937 | 1.812 | 2.162 | 2.081 | 4.421 | 4.249 | 4.599 |
| 1.589 | 1.276 | 2.751 | 0.819 | 2.093 | 2.700 | 2.421 | 2.740 | 2.224 | 2.820 |
| 0.706 | 1.523 | 4.851 | 1.602 | 4.022 | 1.399 | 1.671 | 2.287 | 4.115 | 1.108 |
| 1.721 | 3.194 | 1.051 | 3.420 | 1.406 | 3.575 | 1.586 | 1.060 | 1.712 | 3.696 |
| 1.535 | 1.701 | 1.466 | 1.192 | 2.600 | 3.875 | 2.265 | 1.995 | 1.767 | 3.724 |
| 3.876 | 1.856 | 1.723 | 1.872 | 2.278 | 1.143 | 1.079 | 2.902 | 1.891 | 2.860 |
| 0.737 | 2.047 | 1.069 | 2.488 | 1.351 | 1.041 | 2.934 | 2.882 | 1.617 | 1.052 |
| 2.750 | 5.298 | 2.372 | 5.287 | 5.913 | 4.634 | 4.520 | 3.012 | 4.673 | 3.123 |
| 1.373 | 1.996 | 3.664 | 1.678 | 3.197 | 1.797 | 2.731 | 1.383 | 2.728 | 1.343 |
| 3.386 | 1.849 | 6.462 | 5.218 | 6.036 | 2.054 | 6.604 | 2.182 | 1.310 | 2.984 |
| 4.755 | 3.972 | 1.879 | 3.576 | 2.127 | 2.750 | 1.792 | 1.623 | 2.187 | 3.749 |
| 2.650 | 2.213 | 1.241 | 2.986 | 2.135 | 1.215 | 1.608 | 1.562 | 1.126 | 1.524 |
| 5.571 | 3.181 | 3.914 | 5.382 | 3.060 | 3.755 | 1.035 | 4.237 | 5.737 | 5.339 |
| 1.530 | 2.504 | 2.470 | 2.068 | 1.122 | 0.344 | 3.872 | 1.045 | 3.311 | 1.349 |
| 4.779 | 4.420 | 3.471 | 4.447 | 0.445 | 4.719 | 3.270 | 1.219 | 4.179 | 3.091 |
| 2.452 | 4.650 | 1.131 | 2.951 | 4.005 | 0.832 | 2.911 | 2.574 | 2.371 | 1.803 |
| 2.565 | 2.414 | 0.923 | 2.062 | 5.526 | 2.385 | 1.990 | 2.036 | 2.973 | 2.421 |
| 1.439 | 3.829 | 1.334 | 1.294 | 1.279 | 2.422 | 2.949 | 2.741 | 1.932 | 2.659 |
| 3.888 | 1.524 | 3.675 | 4.748 | 7.131 | 7.411 | 7.808 | 1.854 | 5.252 | 5.882 |
| 1.603 | 1.507 | 4.001 | 2.180 | 1.244 | 1.084 | 2.942 | 1.930 | 2.045 | 1.612 |
| 2.633 | 1.371 | 1.907 | 2.073 | 1.416 | 1.304 | 2.665 | 3.206 | 1.354 | 1.596 |
| 2.086 | 1.962 | 1.252 | 1.197 | 1.661 | 1.713 | 2.182 | 2.582 | 2.399 | 2.791 |
| 2.800 | 1.987 | 6.562 | 1.832 | 6.257 | 1.129 | 6.075 | 7.053 | 1.242 | 6.120 |
| 7.423 | 6.601 | 6.400 | 3.976 | 3.149 | 1.643 | 7.398 | 7.141 | 4.436 | 6.879 |
| 3.786 | 6.453 | 2.254 | 6.418 | 6.050 | 5.496 | 3.591 | 6.079 | 1.401 | 3.806 |
| 1.755 | 6.641 | 1.837 | 5.535 | 3.645 | 5.206 | 3.588 | 5.997 | 3.233 | 1.775 |
| 0.804 | 2.132 | 5.803 | 3.424 | 2.305 | 3.475 | 7.773 | 7.824 | 2.168 | 4.732 |
| 1.661 | 1.418 | 2.400 | 3.917 | 4.567 | 1.186 | 1.240 | 3.133 | 1.511 | 1.656 |
| 4.292 | 4.003 | 3.284 | 4.179 | 3.924 | 4.342 | 4.589 | 3.952 | 1.153 | 4.109 |
| 4.955 | 2.839 | 4.372 | 3.730 | 3.567 | 3.045 | 3.825 | 5.077 | 3.874 | 2.255 |
| 4.301 | 1.327 | 1.689 | 2.704 | 3.954 | 2.647 | 4.671 | 2.970 | 1.283 | 2.873 |
| 2.509 | 1.469 | 3.747 | 3.180 | 7.432 | 4.313 | 7.123 | 4.382 | 7.261 | 4.588 |
| 1.275 | 9.904 | 1.865 | 1.178 | 6.441 | 3.053 | 1.436 | 19.664 | 0.179 | 1.579 |
| 4.694 | 3.156 | 1.432 | 7.665 | 6.024 | 4.361 | 4.524 | 2.119 | 6.514 | 2.655 |
| 0.705 | 3.267 | 3.693 | 0.557 | 2.272 | 2.904 | 1.237 | 2.449 | 1.013 | 2.028 |
| 2.262 | 4.162 | 3.531 | 1.048 | 1.417 | 1.594 | 3.558 | 1.702 | 1.956 | 1.286 |
| 3.973 | 3.493 | 1.691 | 3.246 | 2.600 | 4.683 | 3.667 | 4.641 | 3.274 | 4.739 |
| 1.411 | 1.568 | 0.709 | 1.908 | 2.580 | 1.461 | 2.729 | 5.859 | 2.888 | 0.146 |
| 2.416 | 1.431 | 3.960 | 1.198 | 1.046 | 2.869 | 6.104 | 3.508 | 4.971 | 6.288 |
| 3.240 | 1.273 | 3.420 | 1.785 | 3.923 | 4.030 | 2.579 | 4.832 | 3.118 | 4.303 |
| 1.458 | 2.949 | 2.079 | 3.588 | 1.777 | 3.941 | 1.778 | 1.587 | 1.203 | 1.796 |
| 4.904 | 5.356 | 5.279 | 5.169 | 10.262 | 5.529 | 10.492 | 5.255 | 5.913 | 10.542 |

**Elaborado por**: G. Cuenca

**Continúa…**

**Sigue…**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias dependientes con distribución Exponencial**  Tamaño de muestra n=100 | | | | | | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* | *X6* | *X7* | *X8* | *X9* | *X10* |
| 0.919 | 4.446 | 1.333 | 4.688 | 2.057 | 4.830 | 0.712 | 4.278 | 0.169 | 4.367 |
| 0.177 | 3.187 | 2.683 | 2.848 | 0.209 | 2.757 | 0.875 | 2.298 | 2.544 | 2.163 |
| 6.596 | 7.337 | 7.012 | 6.574 | 7.968 | 6.449 | 9.439 | 9.518 | 11.604 | 11.447 |
| 0.511 | 0.453 | 2.859 | 3.076 | 0.471 | 1.660 | 3.090 | 3.111 | 2.044 | 3.462 |
| 7.019 | 5.192 | 6.730 | 1.971 | 7.147 | 1.580 | 7.974 | 5.108 | 3.734 | 3.566 |
| 5.082 | 5.470 | 6.423 | 10.571 | 10.914 | 6.174 | 5.790 | 4.342 | 6.142 | 7.809 |
| 5.216 | 0.577 | 0.070 | 4.630 | 5.805 | 6.604 | 6.580 | 2.890 | 6.806 | 0.555 |
| 5.508 | 5.838 | 0.860 | 1.988 | 0.277 | 8.785 | 3.236 | 0.196 | 8.269 | 9.167 |
| 7.531 | 7.868 | 6.406 | 6.971 | 10.050 | 5.283 | 10.384 | 5.845 | 6.612 | 5.274 |
| 1.219 | 2.384 | 2.895 | 1.906 | 5.324 | 2.125 | 4.701 | 2.947 | 2.949 | 2.660 |
| 7.965 | 5.846 | 8.665 | 5.610 | 5.002 | 4.962 | 4.533 | 2.566 | 6.117 | 4.267 |
| 0.773 | 2.720 | 1.633 | 2.129 | 1.205 | 0.556 | 0.720 | 0.929 | 0.521 | 0.184 |
| 0.449 | 2.003 | 4.027 | 2.725 | 2.785 | 2.466 | 4.397 | 4.580 | 4.170 | 2.684 |
| 6.455 | 7.185 | 7.863 | 3.065 | 4.945 | 2.619 | 1.508 | 1.379 | 1.302 | 1.192 |
| 4.833 | 1.780 | 2.271 | 2.454 | 1.586 | 2.595 | 2.939 | 1.324 | 1.128 | 4.257 |
| 1.653 | 2.624 | 0.779 | 0.238 | 0.172 | 1.338 | 2.313 | 1.290 | 1.440 | 2.493 |
| 5.029 | 3.679 | 3.578 | 4.295 | 3.063 | 5.534 | 4.939 | 4.058 | 5.257 | 4.231 |
| 3.027 | 6.997 | 3.002 | 3.647 | 1.625 | 2.274 | 1.651 | 3.216 | 4.641 | 1.289 |
| 4.947 | 4.069 | 5.157 | 4.715 | 5.132 | 4.946 | 4.934 | 0.827 | 4.110 | 4.323 |
| 1.047 | 1.023 | 4.330 | 3.551 | 4.398 | 2.603 | 1.513 | 1.317 | 4.113 | 1.171 |
| 2.160 | 3.286 | 4.083 | 5.008 | 5.835 | 4.443 | 5.692 | 6.458 | 6.420 | 6.410 |
| 3.437 | 4.315 | 2.402 | 3.724 | 4.977 | 2.237 | 3.348 | 3.577 | 4.924 | 3.505 |
| 2.650 | 4.631 | 4.361 | 2.749 | 4.810 | 4.374 | 2.653 | 2.303 | 2.003 | 4.456 |
| 4.387 | 4.031 | 2.061 | 1.303 | 2.059 | 3.308 | 2.004 | 4.271 | 4.820 | 3.195 |
| 3.349 | 2.733 | 2.041 | 4.734 | 3.214 | 3.010 | 2.136 | 2.080 | 1.895 | 2.561 |
| 5.950 | 5.100 | 5.241 | 8.751 | 8.797 | 5.607 | 6.784 | 5.941 | 8.083 | 5.750 |
| 2.180 | 4.490 | 1.422 | 3.254 | 2.905 | 3.984 | 4.586 | 3.767 | 4.684 | 5.501 |
| 1.096 | 3.067 | 1.154 | 3.048 | 2.318 | 2.521 | 2.126 | 1.073 | 4.016 | 4.150 |
| 2.541 | 3.118 | 1.148 | 1.888 | 3.642 | 1.282 | 3.155 | 0.424 | 3.997 | 1.188 |
| 1.782 | 2.558 | 1.205 | 1.638 | 2.784 | 3.678 | 3.476 | 1.468 | 1.700 | 1.718 |
| 5.260 | 2.272 | 3.359 | 1.292 | 4.339 | 4.104 | 2.877 | 3.287 | 3.006 | 2.248 |
| 3.872 | 3.320 | 1.821 | 3.069 | 1.131 | 3.017 | 1.615 | 1.421 | 3.691 | 2.732 |
| 0.977 | 5.323 | 3.878 | 5.360 | 1.664 | 1.563 | 3.183 | 1.979 | 1.301 | 3.020 |
| 1.538 | 0.544 | 1.913 | 1.379 | 4.166 | 1.871 | 2.308 | 4.817 | 3.755 | 1.849 |
| 3.097 | 3.744 | 2.224 | 2.974 | 2.029 | 3.689 | 3.154 | 0.622 | 2.684 | 3.376 |
| 2.264 | 1.749 | 1.351 | 1.056 | 2.011 | 1.089 | 1.400 | 1.754 | 2.505 | 2.449 |
| 1.922 | 1.135 | 2.030 | 2.992 | 1.665 | 1.782 | 3.061 | 4.930 | 3.322 | 3.144 |
| 5.309 | 1.632 | 5.489 | 0.409 | 6.785 | 5.881 | 5.931 | 6.314 | 7.342 | 5.589 |
| 2.024 | 2.555 | 3.541 | 3.185 | 1.807 | 1.535 | 2.964 | 3.691 | 1.676 | 1.626 |
| 1.962 | 1.450 | 2.667 | 3.870 | 4.081 | 1.627 | 3.066 | 4.395 | 4.515 | 3.001 |
| 3.514 | 6.951 | 1.244 | 2.751 | 2.468 | 2.018 | 2.323 | 1.230 | 4.707 | 1.959 |
| 0.750 | 0.800 | 0.449 | 1.177 | 1.890 | 1.178 | 2.311 | 0.704 | 0.035 | 1.687 |
| 0.537 | 1.374 | 1.158 | 5.727 | 1.508 | 5.355 | 1.709 | 5.413 | 1.359 | 1.518 |
| 10.796 | 10.465 | 10.521 | 8.610 | 8.558 | 8.599 | 8.032 | 10.596 | 8.551 | 8.123 |
| 1.056 | 1.232 | 2.367 | 1.325 | 0.526 | 1.676 | 2.971 | 2.542 | 2.939 | 2.814 |
| 8.118 | 8.500 | 9.924 | 9.254 | 10.405 | 9.636 | 10.348 | 9.218 | 9.805 | 10.046 |
| 4.040 | 4.244 | 3.613 | 3.099 | 4.680 | 6.852 | 2.452 | 3.183 | 4.986 | 5.924 |
| 4.949 | 7.182 | 4.366 | 3.236 | 4.999 | 3.716 | 3.930 | 4.859 | 7.191 | 4.460 |
| 3.296 | 7.442 | 7.542 | 4.733 | 4.720 | 3.237 | 3.182 | 4.800 | 5.351 | 3.424 |
| 5.786 | 6.620 | 6.717 | 5.252 | 5.305 | 5.491 | 6.526 | 5.525 | 6.722 | 4.637 |

**Elaborado por**: G. Cuenca

El vector de medias de los datos originales es:



**Método de Eliminación por Filas**

Debido a que los datos faltantes recayeron en las variables *X*3 y *X*8 es decir en: *X*3,3=2.851, *X*9,3=1.414, *X*15,3=1.069, *X*18,3=6.462, *X*21,3=3.914, *X*24,3=1.131, *X*31,3=6.562, *X*33,3=2.254, *X*39,3=1.689, *X*42,3=1.432, *X*43,3=3.693, *X*47,3=3.960, *X*48,3=3.420, *X*52,3=2.683, *X*55,3=6.730, *X*58,3=0.860, *X*59,3=6.406, *X*67,3=3.578, *X*69,3=5.157, *X*71,3=4.083, *X*74,3=2.061, *X*79,3=1.148, *X*81,3=3.359, *X*84,3=1.913, *X*86,3=1.351, *X*6,8=2.390, *X*12,8=1.060, *X*17,8=1.383, *X*23,8=1.219, *X*30,8=2.582, *X*34,8=5.997, *X*37,8=3.952, *X*41,8=19.664, *X*46,8=5.859, *X*50,8=5.255, *X*53,8=9.518, *X*60,8=2.947, *X*61,8=2.566, *X*62,8=0.929, *X*63,8=4.580, *X*75,8=2.080, *X*77,8=3.767, *X*87,8=4.930, *X*88,8=6.314, *X*92,8=0.704, *X*93,8=5.413, *X*97,8=3.183, *X*98,8=4.859, *X*99,8=4.800 y *X*100,8=5.525, se procede a prescindir de las filas que tienen estos valores “faltantes”(Ver Tabla 4.32).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tabla 4.32**  *Efectos de la Imputación en el análisis de datos multivariados*  **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** | | | | | | | | | |
| *X1* | *X2* | ***X3*** | *X4* | *X5* | *X6* | *X7* | ***X8*** | *X9* | *X10* |
| 6.726 | 6.168 | 3.447 | 4.124 | 4.017 | 4.550 | 5.149 | 4.957 | 6.743 | 3.346 |
| 1.168 | 1.763 | 0.622 | 2.786 | 4.782 | 3.397 | 3.994 | 1.921 | 1.714 | 2.373 |
| 0.283 | 0.163 | 0.814 | 2.302 | 1.101 | 2.715 | 0.470 | 0.462 | 0.814 | 2.980 |
| 3.054 | 1.277 | 3.099 | 1.934 | 0.206 | 1.929 | 0.575 | 1.089 | 0.289 | 1.435 |
| 0.668 | 1.180 | 3.188 | 2.429 | 4.009 | 3.122 | 2.252 | 1.105 | 4.255 | 2.085 |
| 2.576 | 2.268 | 3.545 | 1.127 | 2.069 | 3.408 | 3.349 | 3.863 | 2.491 | 3.414 |
| 1.589 | 1.276 | 2.751 | 0.819 | 2.093 | 2.700 | 2.421 | 2.740 | 2.224 | 2.820 |
| 0.706 | 1.523 | 4.851 | 1.602 | 4.022 | 1.399 | 1.671 | 2.287 | 4.115 | 1.108 |
| 1.535 | 1.701 | 1.466 | 1.192 | 2.600 | 3.875 | 2.265 | 1.995 | 1.767 | 3.724 |
| 3.876 | 1.856 | 1.723 | 1.872 | 2.278 | 1.143 | 1.079 | 2.902 | 1.891 | 2.860 |
| 4.755 | 3.972 | 1.879 | 3.576 | 2.127 | 2.750 | 1.792 | 1.623 | 2.187 | 3.749 |
| 2.650 | 2.213 | 1.241 | 2.986 | 2.135 | 1.215 | 1.608 | 1.562 | 1.126 | 1.524 |
| 1.530 | 2.504 | 2.470 | 2.068 | 1.122 | 0.344 | 3.872 | 1.045 | 3.311 | 1.349 |
| 2.565 | 2.414 | 0.923 | 2.062 | 5.526 | 2.385 | 1.990 | 2.036 | 2.973 | 2.421 |
| 1.439 | 3.829 | 1.334 | 1.294 | 1.279 | 2.422 | 2.949 | 2.741 | 1.932 | 2.659 |
| 3.888 | 1.524 | 3.675 | 4.748 | 7.131 | 7.411 | 7.808 | 1.854 | 5.252 | 5.882 |
| 1.603 | 1.507 | 4.001 | 2.180 | 1.244 | 1.084 | 2.942 | 1.930 | 2.045 | 1.612 |
| 2.633 | 1.371 | 1.907 | 2.073 | 1.416 | 1.304 | 2.665 | 3.206 | 1.354 | 1.596 |
| 7.423 | 6.601 | 6.400 | 3.976 | 3.149 | 1.643 | 7.398 | 7.141 | 4.436 | 6.879 |
| 0.804 | 2.132 | 5.803 | 3.424 | 2.305 | 3.475 | 7.773 | 7.824 | 2.168 | 4.732 |
| 1.661 | 1.418 | 2.400 | 3.917 | 4.567 | 1.186 | 1.240 | 3.133 | 1.511 | 1.656 |
| 4.955 | 2.839 | 4.372 | 3.730 | 3.567 | 3.045 | 3.825 | 5.077 | 3.874 | 2.255 |
| 2.509 | 1.469 | 3.747 | 3.180 | 7.432 | 4.313 | 7.123 | 4.382 | 7.261 | 4.588 |
| 2.262 | 4.162 | 3.531 | 1.048 | 1.417 | 1.594 | 3.558 | 1.702 | 1.956 | 1.286 |
| 3.973 | 3.493 | 1.691 | 3.246 | 2.600 | 4.683 | 3.667 | 4.641 | 3.274 | 4.739 |
| 1.458 | 2.949 | 2.079 | 3.588 | 1.777 | 3.941 | 1.778 | 1.587 | 1.203 | 1.796 |
| 0.919 | 4.446 | 1.333 | 4.688 | 2.057 | 4.830 | 0.712 | 4.278 | 0.169 | 4.367 |
| 0.511 | 0.453 | 2.859 | 3.076 | 0.471 | 1.660 | 3.090 | 3.111 | 2.044 | 3.462 |
| 5.082 | 5.470 | 6.423 | 10.571 | 10.914 | 6.174 | 5.790 | 4.342 | 6.142 | 7.809 |
| 5.216 | 0.577 | 0.070 | 4.630 | 5.805 | 6.604 | 6.580 | 2.890 | 6.806 | 0.555 |
| 6.455 | 7.185 | 7.863 | 3.065 | 4.945 | 2.619 | 1.508 | 1.379 | 1.302 | 1.192 |
| 4.833 | 1.780 | 2.271 | 2.454 | 1.586 | 2.595 | 2.939 | 1.324 | 1.128 | 4.257 |
| 1.653 | 2.624 | 0.779 | 0.238 | 0.172 | 1.338 | 2.313 | 1.290 | 1.440 | 2.493 |
| 3.027 | 6.997 | 3.002 | 3.647 | 1.625 | 2.274 | 1.651 | 3.216 | 4.641 | 1.289 |
| 1.047 | 1.023 | 4.330 | 3.551 | 4.398 | 2.603 | 1.513 | 1.317 | 4.113 | 1.171 |
| 3.437 | 4.315 | 2.402 | 3.724 | 4.977 | 2.237 | 3.348 | 3.577 | 4.924 | 3.505 |
| 2.650 | 4.631 | 4.361 | 2.749 | 4.810 | 4.374 | 2.653 | 2.303 | 2.003 | 4.456 |
| 5.950 | 5.100 | 5.241 | 8.751 | 8.797 | 5.607 | 6.784 | 5.941 | 8.083 | 5.750 |
| 1.096 | 3.067 | 1.154 | 3.048 | 2.318 | 2.521 | 2.126 | 1.073 | 4.016 | 4.150 |
| 1.782 | 2.558 | 1.205 | 1.638 | 2.784 | 3.678 | 3.476 | 1.468 | 1.700 | 1.718 |
| 3.872 | 3.320 | 1.821 | 3.069 | 1.131 | 3.017 | 1.615 | 1.421 | 3.691 | 2.732 |
| 0.977 | 5.323 | 3.878 | 5.360 | 1.664 | 1.563 | 3.183 | 1.979 | 1.301 | 3.020 |
| 3.097 | 3.744 | 2.224 | 2.974 | 2.029 | 3.689 | 3.154 | 0.622 | 2.684 | 3.376 |
| 2.024 | 2.555 | 3.541 | 3.185 | 1.807 | 1.535 | 2.964 | 3.691 | 1.676 | 1.626 |
| 1.962 | 1.450 | 2.667 | 3.870 | 4.081 | 1.627 | 3.066 | 4.395 | 4.515 | 3.001 |
| 3.514 | 6.951 | 1.244 | 2.751 | 2.468 | 2.018 | 2.323 | 1.230 | 4.707 | 1.959 |
| 10.796 | 10.465 | 10.521 | 8.610 | 8.558 | 8.599 | 8.032 | 10.596 | 8.551 | 8.123 |
| 1.056 | 1.232 | 2.367 | 1.325 | 0.526 | 1.676 | 2.971 | 2.542 | 2.939 | 2.814 |
| 8.118 | 8.500 | 9.924 | 9.254 | 10.405 | 9.636 | 10.348 | 9.218 | 9.805 | 10.046 |
| 6.726 | 6.168 | 3.447 | 4.124 | 4.017 | 4.550 | 5.149 | 4.957 | 6.743 | 3.346 |

**Elaborado por**: G. Cuenca

El vector de medias para las cincuenta filas restantes es:



El vector de medias de los datos originales y de los datos con filas eliminadas no coincide.

Ahora analicemos el efecto que causa en la *matriz de varianzas y covarianzas*, y *matriz de correlaciones*, la eliminación de cincuenta filas, con un tamaño de muestra *n*=100.

|  |
| --- |
| **CUADRO 4.41**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 4.386 |  |  |  |  |  |  |  |  |  | | *X*2 | 2.700 | 4.854 |  |  |  |  |  |  |  |  | | *X*3 | 2.701 | 2.336 | 4.528 |  |  |  |  |  |  |  | | *X*4 | 2.165 | 2.247 | 2.041 | 3.978 |  |  |  |  |  |  | | *X*5 | 2.780 | 2.346 | 3.072 | 2.968 | 6.029 |  |  |  |  |  | | *X6* | 2.252 | 1.997 | 1.489 | 2.269 | 2.240 | 4.084 |  |  |  |  | | *X7* | 2.706 | 1.857 | 2.925 | 2.329 | 3.696 | 2.695 | 5.563 |  |  |  | | *X8* | 1.637 | 2.954 | 2.226 | 1.497 | 2.897 | 2.173 | 3.059 | 7.626 |  |  | | *X9* | 3.019 | 2.133 | 2.366 | 2.365 | 3.272 | 2.508 | 3.039 | 1.543 | 5.322 |  | | *X10* | 2.552 | 2.045 | 2.371 | 2.044 | 2.521 | 3.192 | 3.685 | 2.716 | 2.939 | 5.072 | |
| **Matriz de Varianzas y Covarianzas**  **(Cincuenta Filas Eliminadas)**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 5.136 |  |  |  |  |  |  |  |  |  | | *X*2 | 3.729 | 5.163 |  |  |  |  |  |  |  |  | | *X*3 | 3.001 | 3.035 | 4.813 |  |  |  |  |  |  |  | | *X*4 | 2.889 | 2.730 | 2.849 | 4.413 |  |  |  |  |  |  | | *X*5 | 3.226 | 2.320 | 3.247 | 4.180 | 6.452 |  |  |  |  |  | | *X6* | 2.606 | 1.939 | 2.009 | 2.879 | 3.690 | 3.844 |  |  |  |  | | *X7* | 2.925 | 1.975 | 2.881 | 2.848 | 3.717 | 3.004 | 4.956 |  |  |  | | *X8* | 3.041 | 2.650 | 3.279 | 2.847 | 2.932 | 2.358 | 3.560 | 4.705 |  |  | | *X9* | 3.303 | 2.498 | 2.551 | 3.175 | 4.367 | 2.924 | 3.744 | 3.001 | 5.132 |  | | *X10* | 2.468 | 2.255 | 2.569 | 2.878 | 3.074 | 2.689 | 3.055 | 2.958 | 2.433 | 3.854 | |

**Elaborado por**: G. Cuenca

Analizando el Cuadro 4.41, se puede apreciar que la mayor covarianza en la matriz de datos originales se da entre las variables *X5*  y *X9* es decir 3.272; mientras que en la matriz con cincuenta filas eliminadas este valor aumenta a 4.367.

|  |
| --- |
| **CUADRO 4.42**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 1.000 |  |  |  |  |  |  |  |  |  | | *X*2 | 0.585 | 1.000 |  |  |  |  |  |  |  |  | | *X*3 | 0.606 | 0.498 | 1.000 |  |  |  |  |  |  |  | | *X*4 | 0.518 | 0.511 | 0.481 | 1.000 |  |  |  |  |  |  | | *X*5 | 0.541 | 0.434 | 0.588 | 0.606 | 1.000 |  |  |  |  |  | | *X6* | 0.532 | 0.448 | 0.346 | 0.563 | 0.451 | 1.000 |  |  |  |  | | *X7* | 0.548 | 0.357 | 0.583 | 0.495 | 0.638 | 0.565 | 1.000 |  |  |  | | *X8* | 0.283 | 0.486 | 0.379 | 0.272 | 0.427 | 0.389 | 0.470 | 1.000 |  |  | | *X9* | 0.625 | 0.420 | 0.482 | 0.514 | 0.578 | 0.538 | 0.559 | 0.242 | 1.000 |  | | *X10* | 0.541 | 0.412 | 0.495 | 0.455 | 0.456 | 0.701 | 0.694 | 0.437 | 0.566 | 1.000 | |
| **Matriz de Correlaciones**  **(Cincuenta Filas Eliminadas)**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 1.000 |  |  |  |  |  |  |  |  |  | | *X*2 | 0.724 | 1.000 |  |  |  |  |  |  |  |  | | *X*3 | 0.604 | 0.609 | 1.000 |  |  |  |  |  |  |  | | *X*4 | 0.607 | 0.572 | 0.618 | 1.000 |  |  |  |  |  |  | | *X*5 | 0.560 | 0.402 | 0.583 | 0.783 | 1.000 |  |  |  |  |  | | *X6* | 0.586 | 0.435 | 0.467 | 0.699 | 0.741 | 1.000 |  |  |  |  | | *X7* | 0.580 | 0.390 | 0.590 | 0.609 | 0.657 | 0.688 | 1.000 |  |  |  | | *X8* | 0.619 | 0.538 | 0.689 | 0.625 | 0.532 | 0.554 | 0.737 | 1.000 |  |  | | *X9* | 0.643 | 0.485 | 0.513 | 0.667 | 0.759 | 0.658 | 0.742 | 0.611 | 1.000 |  | | *X10* | 0.555 | 0.505 | 0.596 | 0.698 | 0.616 | 0.699 | 0.699 | 0.695 | 0.547 | 1.000 | |

**Elaborado por**: G. Cuenca

En la matriz de correlaciones de datos originales, la mayor correlación se da entre las variables *X7* y *X10*, es decir 0.701, cuyo valor se disminuye a 0.699 en la matriz de correlaciones con cincuenta filas eliminadas. La mayor correlación en la matriz con cincuenta filas eliminadas es entre las variables *X4* y *X5,* es decir 0.783. En general, se puede decir que la correlación entre las variables, se incrementó en la matriz con 50 filas eliminadas.

**Elaborado por**: G. Cuenca

|  |  |
| --- | --- |
| **CUADRO 4.43**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****3*” y “*Variable*** *X****8*”** | |
| **Estimadores “*Variable*** *X****3*”**   |  |  |  |  | | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Con el 25% de datos eliminadas en X3** | | n | | 100 | 50 | | Media | | 3,206 | 3,158 | | Mediana | | 2,801 | 2,709 | | Moda | | 0,070 | 3,450 | | Varianza | | 4,528 | 4,813 | | Desviación Estándar | | 2,128 | 2,194 | | Error Estándar | | 0,213 | 0,310 | | **Coeficiente de Asimetría** | | 1,194 | 1,559 | | Curtosis | | 1,351 | 2,943 | | Rango | | 10,450 | 10,450 | | Mínimo | | 0,070 | 0,070 | | Máximo | | 10,520 | 10,520 | | Percentiles | 25 | 1,508 | 1,635 | | 50 | 2,801 | 2,709 | | 75 | 4,020 | 3,909 | | **Diagrama de Cajas “*Variable*** *X****3*”** |
| **Estimadores “*Variable*** *X****8*”**   |  |  |  |  | | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Con el 25% de datos eliminadas en X8** | | n | | 100 | 50 | | Media | | 3,588 | 3,059 | | Mediana | | 2,959 | 2,423 | | Moda | | 0,200 | 4,960 | | Varianza | | 7,626 | 4,705 | | Desviación Estándar | | 2,762 | 2,169 | | Error Estándar | | 0,276 | 0,307 | | **Coeficiente de Asimetría** | | 2,576 | 1,619 | | Curtosis | | 11,269 | 2,870 | | Rango | | 19,470 | 10,130 | | Mínimo | | 0,200 | 0,460 | | Máximo | | 19,660 | 10,600 | | Percentiles | 25 | 1,715 | 1,456 | | 50 | 2,959 | 2,423 | | 75 | 4,813 | 4,294 | | **Diagrama de Cajas “*Variable*** *X****8*”** |

En el Cuadro 4.43, podemos apreciar que con el 25% de datos eliminados en la tercera columna de la matriz de datos (Variable *X****3*)**, el valor de la media y la mediana disminuyó de 3.206 a 3.158 y de 2.801 a 2.709, respectivamente. La varianza de la variable *X****3***, con 25% de datos eliminadosaumentó de 4.528 a 4.813. En la variable *X****8***, el valor de la media y la mediana disminuyeron su valor, así como también el valor de la varianza.

**Método de Imputación por la Media y Regresión**

Estos métodos se aplican a la misma matriz de datos utilizada en el método de eliminación por filas, es decir se completan datos en la variable *X*3 y *X*8, que presentan veinte y cinco valores faltantes cada una. A través del Método de Imputación por Media, se procede a calcular la media aritmética de la variable *X3* con los veinticinco datos faltantes, cuyo valor es 3.219, entonces reemplazamos en *X*3,3, *X*9,3, *X*15,3, *X*18,3, *X*21,3, *X*24,3, *X*31,3, *X*33,3, *X*39,3, *X*42,3, *X*43,3, *X*47,3, *X*48,3, *X*52,3, *X*55,3, *X*58,3, *X*59,3, *X*67,3, *X*69,3, *X*71,3, *X*74,3, *X*79,3, *X*81,3, *X*84,3, *X*86,3, también se calcula el valor de la media de la variable *X8* , 3.298, mismo que se remplaza en *X*6,8, *X*12,8, *X*17,8, *X*23,8, *X*30,8, *X*34,8, *X*37,8, *X*41,8, *X*46,8, *X*50,8, *X*53,8, *X*60,8, *X*61,8, *X*62,8, *X*63,8, *X*75,8, *X*77,8, *X*87,8, *X*88,8, *X*92,8, *X*93,8, *X*97,8, *X*98,8, *X*99,8 y en *X*100,8. La matriz de datos resultante con cincuenta valores completados por *imputación por la media* y *regresión*, se muestra en la Tabla 4.33 y 4.34 respectivamente.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tabla 4.33**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | | | | | | |
| *X1* | *X2* | ***X3*** | *X4* | *X5* | *X6* | *X7* | ***X8*** | *X9* | *X10* |
| 6.726 | 6.168 | 3,447 | 4.124 | 4.017 | 4.550 | 5.149 | 4.957 | 6.743 | 3.346 |
| 1.168 | 1.763 | 0,622 | 2.786 | 4.782 | 3.397 | 3.994 | 1.921 | 1.714 | 2.373 |
| 3.238 | 4.557 | **3,219** | 3.335 | 0.641 | 10.599 | 10.406 | 11.662 | 0.222 | 10.237 |
| 0.283 | 0.163 | 0,814 | 2.302 | 1.101 | 2.715 | 0.470 | 0.462 | 0.814 | 2.980 |
| 3.054 | 1.277 | 3,099 | 1.934 | 0.206 | 1.929 | 0.575 | 1.089 | 0.289 | 1.435 |
| 3.483 | 3.547 | 3,129 | 5.710 | 3.334 | 3.645 | 5.478 | **3.298** | 4.686 | 3.469 |
| 0.668 | 1.180 | 3,188 | 2.429 | 4.009 | 3.122 | 2.252 | 1.105 | 4.255 | 2.085 |
| 2.576 | 2.268 | 3,545 | 1.127 | 2.069 | 3.408 | 3.349 | 3.863 | 2.491 | 3.414 |
| 4.385 | 1.285 | **3,219** | 1.937 | 1.812 | 2.162 | 2.081 | 4.421 | 4.249 | 4.599 |
| 1.589 | 1.276 | 2,751 | 0.819 | 2.093 | 2.700 | 2.421 | 2.740 | 2.224 | 2.820 |
| 0.706 | 1.523 | 4,851 | 1.602 | 4.022 | 1.399 | 1.671 | 2.287 | 4.115 | 1.108 |
| 1.721 | 3.194 | 1,051 | 3.420 | 1.406 | 3.575 | 1.586 | **3.298** | 1.712 | 3.696 |
| 1.535 | 1.701 | 1,466 | 1.192 | 2.600 | 3.875 | 2.265 | 1.995 | 1.767 | 3.724 |
| 3.876 | 1.856 | 1,723 | 1.872 | 2.278 | 1.143 | 1.079 | 2.902 | 1.891 | 2.860 |
| 0.737 | 2.047 | **3,219** | 2.488 | 1.351 | 1.041 | 2.934 | 2.882 | 1.617 | 1.052 |
| 2.750 | 5.298 | 2,372 | 5.287 | 5.913 | 4.634 | 4.520 | 3.012 | 4.673 | 3.123 |
| 1.373 | 1.996 | 3,664 | 1.678 | 3.197 | 1.797 | 2.731 | **3.298** | 2.728 | 1.343 |
| 3.386 | 1.849 | **3,219** | 5.218 | 6.036 | 2.054 | 6.604 | 2.182 | 1.310 | 2.984 |
| 4.755 | 3.972 | 1,879 | 3.576 | 2.127 | 2.750 | 1.792 | 1.623 | 2.187 | 3.749 |
| 2.650 | 2.213 | 1,241 | 2.986 | 2.135 | 1.215 | 1.608 | 1.562 | 1.126 | 1.524 |
| 5.571 | 3.181 | **3,219** | 5.382 | 3.060 | 3.755 | 1.035 | 4.237 | 5.737 | 5.339 |
| 1.530 | 2.504 | 2,470 | 2.068 | 1.122 | 0.344 | 3.872 | 1.045 | 3.311 | 1.349 |
| 4.779 | 4.420 | 3,471 | 4.447 | 0.445 | 4.719 | 3.270 | **3.298** | 4.179 | 3.091 |
| 2.452 | 4.650 | **3,219** | 2.951 | 4.005 | 0.832 | 2.911 | 2.574 | 2.371 | 1.803 |
| 2.565 | 2.414 | 0,923 | 2.062 | 5.526 | 2.385 | 1.990 | 2.036 | 2.973 | 2.421 |
| 1.439 | 3.829 | 1,334 | 1.294 | 1.279 | 2.422 | 2.949 | 2.741 | 1.932 | 2.659 |
| 3.888 | 1.524 | 3,675 | 4.748 | 7.131 | 7.411 | 7.808 | 1.854 | 5.252 | 5.882 |
| 1.603 | 1.507 | 4,001 | 2.180 | 1.244 | 1.084 | 2.942 | 1.930 | 2.045 | 1.612 |
| 2.633 | 1.371 | 1,907 | 2.073 | 1.416 | 1.304 | 2.665 | 3.206 | 1.354 | 1.596 |
| 2.086 | 1.962 | 1,252 | 1.197 | 1.661 | 1.713 | 2.182 | **3.298** | 2.399 | 2.791 |
| 2.800 | 1.987 | **3,219** | 1.832 | 6.257 | 1.129 | 6.075 | 7.053 | 1.242 | 6.120 |
| 7.423 | 6.601 | 6,400 | 3.976 | 3.149 | 1.643 | 7.398 | 7.141 | 4.436 | 6.879 |
| 3.786 | 6.453 | **3,219** | 6.418 | 6.050 | 5.496 | 3.591 | 6.079 | 1.401 | 3.806 |
| 1.755 | 6.641 | 1,837 | 5.535 | 3.645 | 5.206 | 3.588 | **3.298** | 3.233 | 1.775 |
| 0.804 | 2.132 | 5,803 | 3.424 | 2.305 | 3.475 | 7.773 | 7.824 | 2.168 | 4.732 |
| 1.661 | 1.418 | 2,400 | 3.917 | 4.567 | 1.186 | 1.240 | 3.133 | 1.511 | 1.656 |
| 4.292 | 4.003 | 3,284 | 4.179 | 3.924 | 4.342 | 4.589 | **3.298** | 1.153 | 4.109 |
| 4.955 | 2.839 | 4,372 | 3.730 | 3.567 | 3.045 | 3.825 | 5.077 | 3.874 | 2.255 |
| 4.301 | 1.327 | **3,219** | 2.704 | 3.954 | 2.647 | 4.671 | 2.970 | 1.283 | 2.873 |
| 2.509 | 1.469 | 3,747 | 3.180 | 7.432 | 4.313 | 7.123 | 4.382 | 7.261 | 4.588 |
| 1.275 | 9.904 | 1,865 | 1.178 | 6.441 | 3.053 | 1.436 | **3.298** | 0.179 | 1.579 |
| 4.694 | 3.156 | **3,219** | 7.665 | 6.024 | 4.361 | 4.524 | 2.119 | 6.514 | 2.655 |
| 0.705 | 3.267 | **3,219** | 0.557 | 2.272 | 2.904 | 1.237 | 2.449 | 1.013 | 2.028 |
| 2.262 | 4.162 | 3,531 | 1.048 | 1.417 | 1.594 | 3.558 | 1.702 | 1.956 | 1.286 |
| 3.973 | 3.493 | 1,691 | 3.246 | 2.600 | 4.683 | 3.667 | 4.641 | 3.274 | 4.739 |
| 1.411 | 1.568 | 0,709 | 1.908 | 2.580 | 1.461 | 2.729 | **3.298** | 2.888 | 0.146 |
| 2.416 | 1.431 | **3,219** | 1.198 | 1.046 | 2.869 | 6.104 | 3.508 | 4.971 | 6.288 |
| 3.240 | 1.273 | **3,219** | 1.785 | 3.923 | 4.030 | 2.579 | 4.832 | 3.118 | 4.303 |
| 1.458 | 2.949 | 2,079 | 3.588 | 1.777 | 3.941 | 1.778 | 1.587 | 1.203 | 1.796 |
| 4.904 | 5.356 | 5,279 | 5.169 | 10.262 | 5.529 | 10.492 | **3.298** | 5.913 | 10.542 |

**Elaborado por**: G. Cuenca

**Viene…**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Efectos de la Imputación en el análisis de datos multivariados*  **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 | | | | | | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* | *X6* | *X7* | *X8* | *X9* | *X10* |
| 0.919 | 4.446 | 1,333 | 4.688 | 2.057 | 4.830 | 0.712 | 4.278 | 0.169 | 4.367 |
| 0.177 | 3.187 | **3,219** | 2.848 | 0.209 | 2.757 | 0.875 | 2.298 | 2.544 | 2.163 |
| 6.596 | 7.337 | 7,012 | 6.574 | 7.968 | 6.449 | 9.439 | **3.298** | 11.604 | 11.447 |
| 0.511 | 0.453 | 2,859 | 3.076 | 0.471 | 1.660 | 3.090 | 3.111 | 2.044 | 3.462 |
| 7.019 | 5.192 | **3,219** | 1.971 | 7.147 | 1.580 | 7.974 | 5.108 | 3.734 | 3.566 |
| 5.082 | 5.470 | 6,423 | 10.571 | 10.914 | 6.174 | 5.790 | 4.342 | 6.142 | 7.809 |
| 5.216 | 0.577 | 0,070 | 4.630 | 5.805 | 6.604 | 6.580 | 2.890 | 6.806 | 0.555 |
| 5.508 | 5.838 | **3,219** | 1.988 | 0.277 | 8.785 | 3.236 | 0.196 | 8.269 | 9.167 |
| 7.531 | 7.868 | **3,219** | 6.971 | 10.050 | 5.283 | 10.384 | 5.845 | 6.612 | 5.274 |
| 1.219 | 2.384 | 2,895 | 1.906 | 5.324 | 2.125 | 4.701 | **3.298** | 2.949 | 2.660 |
| 7.965 | 5.846 | 8,665 | 5.610 | 5.002 | 4.962 | 4.533 | **3.298** | 6.117 | 4.267 |
| 0.773 | 2.720 | 1,633 | 2.129 | 1.205 | 0.556 | 0.720 | **3.298** | 0.521 | 0.184 |
| 0.449 | 2.003 | 4,027 | 2.725 | 2.785 | 2.466 | 4.397 | **3.298** | 4.170 | 2.684 |
| 6.455 | 7.185 | 7,863 | 3.065 | 4.945 | 2.619 | 1.508 | 1.379 | 1.302 | 1.192 |
| 4.833 | 1.780 | 2,271 | 2.454 | 1.586 | 2.595 | 2.939 | 1.324 | 1.128 | 4.257 |
| 1.653 | 2.624 | 0,779 | 0.238 | 0.172 | 1.338 | 2.313 | 1.290 | 1.440 | 2.493 |
| 5.029 | 3.679 | **3,219** | 4.295 | 3.063 | 5.534 | 4.939 | 4.058 | 5.257 | 4.231 |
| 3.027 | 6.997 | 3,002 | 3.647 | 1.625 | 2.274 | 1.651 | 3.216 | 4.641 | 1.289 |
| 4.947 | 4.069 | **3,219** | 4.715 | 5.132 | 4.946 | 4.934 | 0.827 | 4.110 | 4.323 |
| 1.047 | 1.023 | 4,330 | 3.551 | 4.398 | 2.603 | 1.513 | 1.317 | 4.113 | 1.171 |
| 2.160 | 3.286 | **3,219** | 5.008 | 5.835 | 4.443 | 5.692 | 6.458 | 6.420 | 6.410 |
| 3.437 | 4.315 | 2,402 | 3.724 | 4.977 | 2.237 | 3.348 | 3.577 | 4.924 | 3.505 |
| 2.650 | 4.631 | 4,361 | 2.749 | 4.810 | 4.374 | 2.653 | 2.303 | 2.003 | 4.456 |
| 4.387 | 4.031 | **3,219** | 1.303 | 2.059 | 3.308 | 2.004 | 4.271 | 4.820 | 3.195 |
| 3.349 | 2.733 | 2,041 | 4.734 | 3.214 | 3.010 | 2.136 | **3.298** | 1.895 | 2.561 |
| 5.950 | 5.100 | 5,241 | 8.751 | 8.797 | 5.607 | 6.784 | 5.941 | 8.083 | 5.750 |
| 2.180 | 4.490 | 1,422 | 3.254 | 2.905 | 3.984 | 4.586 | **3.298** | 4.684 | 5.501 |
| 1.096 | 3.067 | 1,154 | 3.048 | 2.318 | 2.521 | 2.126 | 1.073 | 4.016 | 4.150 |
| 2.541 | 3.118 | **3,219** | 1.888 | 3.642 | 1.282 | 3.155 | 0.424 | 3.997 | 1.188 |
| 1.782 | 2.558 | 1,205 | 1.638 | 2.784 | 3.678 | 3.476 | 1.468 | 1.700 | 1.718 |
| 5.260 | 2.272 | **3,219** | 1.292 | 4.339 | 4.104 | 2.877 | 3.287 | 3.006 | 2.248 |
| 3.872 | 3.320 | 1,821 | 3.069 | 1.131 | 3.017 | 1.615 | 1.421 | 3.691 | 2.732 |
| 0.977 | 5.323 | 3,878 | 5.360 | 1.664 | 1.563 | 3.183 | 1.979 | 1.301 | 3.020 |
| 1.538 | 0.544 | **3,219** | 1.379 | 4.166 | 1.871 | 2.308 | 4.817 | 3.755 | 1.849 |
| 3.097 | 3.744 | 2,224 | 2.974 | 2.029 | 3.689 | 3.154 | 0.622 | 2.684 | 3.376 |
| 2.264 | 1.749 | **3,219** | 1.056 | 2.011 | 1.089 | 1.400 | 1.754 | 2.505 | 2.449 |
| 1.922 | 1.135 | 2,030 | 2.992 | 1.665 | 1.782 | 3.061 | **3.298** | 3.322 | 3.144 |
| 5.309 | 1.632 | 5,489 | 0.409 | 6.785 | 5.881 | 5.931 | **3.298** | 7.342 | 5.589 |
| 2.024 | 2.555 | 3,541 | 3.185 | 1.807 | 1.535 | 2.964 | 3.691 | 1.676 | 1.626 |
| 1.962 | 1.450 | 2,667 | 3.870 | 4.081 | 1.627 | 3.066 | 4.395 | 4.515 | 3.001 |
| 3.514 | 6.951 | 1,244 | 2.751 | 2.468 | 2.018 | 2.323 | 1.230 | 4.707 | 1.959 |
| 0.750 | 0.800 | 0,449 | 1.177 | 1.890 | 1.178 | 2.311 | **3.298** | 0.035 | 1.687 |
| 0.537 | 1.374 | 1,158 | 5.727 | 1.508 | 5.355 | 1.709 | **3.298** | 1.359 | 1.518 |
| 10.796 | 10.465 | 10,521 | 8.610 | 8.558 | 8.599 | 8.032 | 10.596 | 8.551 | 8.123 |
| 1.056 | 1.232 | 2,367 | 1.325 | 0.526 | 1.676 | 2.971 | 2.542 | 2.939 | 2.814 |
| 8.118 | 8.500 | 9,924 | 9.254 | 10.405 | 9.636 | 10.348 | 9.218 | 9.805 | 10.046 |
| 4.040 | 4.244 | 3,613 | 3.099 | 4.680 | 6.852 | 2.452 | **3.298** | 4.986 | 5.924 |
| 4.949 | 7.182 | 4,366 | 3.236 | 4.999 | 3.716 | 3.930 | **3.298** | 7.191 | 4.460 |
| 3.296 | 7.442 | 7,542 | 4.733 | 4.720 | 3.237 | 3.182 | **3.298** | 5.351 | 3.424 |
| 5.786 | 6.620 | 6,717 | 5.252 | 5.305 | 5.491 | 6.526 | **3.298** | 6.722 | 4.637 |
| **Tabla 4.34**  *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias dependientes con distribución Exponencial**  **Método de Imputación por Regresiòn**  Tamaño de muestra n=100 y 5% de datos faltantes en la matriz | | | | | | | | | |
| *X1* | *X2* | ***X3*** | *X4* | *X5* | *X6* | *X7* | ***X8*** | *X9* | *X10* |
| 6.726 | 6.168 | 3,447 | 4.124 | 4.017 | 4.550 | 5.149 | 4,957 | 6.743 | 3.346 |
| 1.168 | 1.763 | 0,622 | 2.786 | 4.782 | 3.397 | 3.994 | 1,921 | 1.714 | 2.373 |
| 3.238 | 4.557 | **2,849** | 3.335 | 0.641 | 10.599 | 10.406 | 11,662 | 0.222 | 10.237 |
| 0.283 | 0.163 | 0,814 | 2.302 | 1.101 | 2.715 | 0.470 | 0,462 | 0.814 | 2.980 |
| 3.054 | 1.277 | 3,099 | 1.934 | 0.206 | 1.929 | 0.575 | 1,089 | 0.289 | 1.435 |
| 3.483 | 3.547 | 3,129 | 5.710 | 3.334 | 3.645 | 5.478 | **2,386** | 4.686 | 3.469 |
| 0.668 | 1.180 | 3,188 | 2.429 | 4.009 | 3.122 | 2.252 | 1,105 | 4.255 | 2.085 |
| 2.576 | 2.268 | 3,545 | 1.127 | 2.069 | 3.408 | 3.349 | 3,863 | 2.491 | 3.414 |
| 4.385 | 1.285 | **1,403** | 1.937 | 1.812 | 2.162 | 2.081 | 4,421 | 4.249 | 4.599 |
| 1.589 | 1.276 | 2,751 | 0.819 | 2.093 | 2.700 | 2.421 | 2,740 | 2.224 | 2.820 |
| 0.706 | 1.523 | 4,851 | 1.602 | 4.022 | 1.399 | 1.671 | 2,287 | 4.115 | 1.108 |
| 1.721 | 3.194 | 1,051 | 3.420 | 1.406 | 3.575 | 1.586 | **1,102** | 1.712 | 3.696 |
| 1.535 | 1.701 | 1,466 | 1.192 | 2.600 | 3.875 | 2.265 | 1,995 | 1.767 | 3.724 |
| 3.876 | 1.856 | 1,723 | 1.872 | 2.278 | 1.143 | 1.079 | 2,902 | 1.891 | 2.860 |
| 0.737 | 2.047 | **1,057** | 2.488 | 1.351 | 1.041 | 2.934 | 2,882 | 1.617 | 1.052 |
| 2.750 | 5.298 | 2,372 | 5.287 | 5.913 | 4.634 | 4.520 | 3,012 | 4.673 | 3.123 |
| 1.373 | 1.996 | 3,664 | 1.678 | 3.197 | 1.797 | 2.731 | **1,374** | 2.728 | 1.343 |
| 3.386 | 1.849 | **6,399** | 5.218 | 6.036 | 2.054 | 6.604 | 2,182 | 1.310 | 2.984 |
| 4.755 | 3.972 | 1,879 | 3.576 | 2.127 | 2.750 | 1.792 | 1,623 | 2.187 | 3.749 |
| 2.650 | 2.213 | 1,241 | 2.986 | 2.135 | 1.215 | 1.608 | 1,562 | 1.126 | 1.524 |
| 5.571 | 3.181 | **3,909** | 5.382 | 3.060 | 3.755 | 1.035 | 4,237 | 5.737 | 5.339 |
| 1.530 | 2.504 | 2,470 | 2.068 | 1.122 | 0.344 | 3.872 | 1,045 | 3.311 | 1.349 |
| 4.779 | 4.420 | 3,471 | 4.447 | 0.445 | 4.719 | 3.270 | **1,207** | 4.179 | 3.091 |
| 2.452 | 4.650 | **1,098** | 2.951 | 4.005 | 0.832 | 2.911 | 2,574 | 2.371 | 1.803 |
| 2.565 | 2.414 | 0,923 | 2.062 | 5.526 | 2.385 | 1.990 | 2,036 | 2.973 | 2.421 |
| 1.439 | 3.829 | 1,334 | 1.294 | 1.279 | 2.422 | 2.949 | 2,741 | 1.932 | 2.659 |
| 3.888 | 1.524 | 3,675 | 4.748 | 7.131 | 7.411 | 7.808 | 1,854 | 5.252 | 5.882 |
| 1.603 | 1.507 | 4,001 | 2.180 | 1.244 | 1.084 | 2.942 | 1,930 | 2.045 | 1.612 |
| 2.633 | 1.371 | 1,907 | 2.073 | 1.416 | 1.304 | 2.665 | 3,206 | 1.354 | 1.596 |
| 2.086 | 1.962 | 1,252 | 1.197 | 1.661 | 1.713 | 2.182 | **2,601** | 2.399 | 2.791 |
| 2.800 | 1.987 | **6,554** | 1.832 | 6.257 | 1.129 | 6.075 | 7,053 | 1.242 | 6.120 |
| 7.423 | 6.601 | 6,400 | 3.976 | 3.149 | 1.643 | 7.398 | 7,141 | 4.436 | 6.879 |
| 3.786 | 6.453 | **2,226** | 6.418 | 6.050 | 5.496 | 3.591 | 6,079 | 1.401 | 3.806 |
| 1.755 | 6.641 | 1,837 | 5.535 | 3.645 | 5.206 | 3.588 | **6,003** | 3.233 | 1.775 |
| 0.804 | 2.132 | 5,803 | 3.424 | 2.305 | 3.475 | 7.773 | 7,824 | 2.168 | 4.732 |
| 1.661 | 1.418 | 2,400 | 3.917 | 4.567 | 1.186 | 1.240 | 3,133 | 1.511 | 1.656 |
| 4.292 | 4.003 | 3,284 | 4.179 | 3.924 | 4.342 | 4.589 | **4,007** | 1.153 | 4.109 |
| 4.955 | 2.839 | 4,372 | 3.730 | 3.567 | 3.045 | 3.825 | 5,077 | 3.874 | 2.255 |
| 4.301 | 1.327 | **1,673** | 2.704 | 3.954 | 2.647 | 4.671 | 2,970 | 1.283 | 2.873 |
| 2.509 | 1.469 | 3,747 | 3.180 | 7.432 | 4.313 | 7.123 | 4,382 | 7.261 | 4.588 |
| 1.275 | 9.904 | 1,865 | 1.178 | 6.441 | 3.053 | 1.436 | **19,618** | 0.179 | 1.579 |
| 4.694 | 3.156 | **1,429** | 7.665 | 6.024 | 4.361 | 4.524 | 2,119 | 6.514 | 2.655 |
| 0.705 | 3.267 | **3,688** | 0.557 | 2.272 | 2.904 | 1.237 | 2,449 | 1.013 | 2.028 |
| 2.262 | 4.162 | 3,531 | 1.048 | 1.417 | 1.594 | 3.558 | 1,702 | 1.956 | 1.286 |
| 3.973 | 3.493 | 1,691 | 3.246 | 2.600 | 4.683 | 3.667 | 4,641 | 3.274 | 4.739 |
| 1.411 | 1.568 | 0,709 | 1.908 | 2.580 | 1.461 | 2.729 | **5,832** | 2.888 | 0.146 |
| 2.416 | 1.431 | **3,952** | 1.198 | 1.046 | 2.869 | 6.104 | 3,508 | 4.971 | 6.288 |
| 3.240 | 1.273 | **3,411** | 1.785 | 3.923 | 4.030 | 2.579 | 4,832 | 3.118 | 4.303 |
| 1.458 | 2.949 | 2,079 | 3.588 | 1.777 | 3.941 | 1.778 | 1,587 | 1.203 | 1.796 |
| 4.904 | 5.356 | 5,279 | 5.169 | 10.262 | 5.529 | 10.492 | **5,243** | 5.913 | 10.542 |

**Continúa…**

**Viene…**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Efectos de la Imputación en el análisis de datos multivariados*  **Matriz de Datos de variables aleatorias dependientes con distribución Exponencial**  **Método de Imputación por Regresiòn**  Tamaño de muestra n=100 y 5% de datos faltantes en la matriz | | | | | | | | | |
| *X1* | *X2* | *X3* | *X4* | *X5* | *X6* | *X7* | *X8* | *X9* | *X10* |
| 0.919 | 4.446 | 1,333 | 4.688 | 2.057 | 4.830 | 0.712 | 4,278 | 0.169 | 4.367 |
| 0.177 | 3.187 | **2,689** | 2.848 | 0.209 | 2.757 | 0.875 | 2,298 | 2.544 | 2.163 |
| 6.596 | 7.337 | 7,012 | 6.574 | 7.968 | 6.449 | 9.439 | **9,492** | 11.604 | 11.447 |
| 0.511 | 0.453 | 2,859 | 3.076 | 0.471 | 1.660 | 3.090 | 3,111 | 2.044 | 3.462 |
| 7.019 | 5.192 | **6,713** | 1.971 | 7.147 | 1.580 | 7.974 | 5,108 | 3.734 | 3.566 |
| 5.082 | 5.470 | 6,423 | 10.571 | 10.914 | 6.174 | 5.790 | 4,342 | 6.142 | 7.809 |
| 5.216 | 0.577 | 0,070 | 4.630 | 5.805 | 6.604 | 6.580 | 2,890 | 6.806 | 0.555 |
| 5.508 | 5.838 | **0,853** | 1.988 | 0.277 | 8.785 | 3.236 | 0,196 | 8.269 | 9.167 |
| 7.531 | 7.868 | **6,397** | 6.971 | 10.050 | 5.283 | 10.384 | 5,845 | 6.612 | 5.274 |
| 1.219 | 2.384 | 2,895 | 1.906 | 5.324 | 2.125 | 4.701 | **3,003** | 2.949 | 2.660 |
| 7.965 | 5.846 | 8,665 | 5.610 | 5.002 | 4.962 | 4.533 | **2,572** | 6.117 | 4.267 |
| 0.773 | 2.720 | 1,633 | 2.129 | 1.205 | 0.556 | 0.720 | **0,919** | 0.521 | 0.184 |
| 0.449 | 2.003 | 4,027 | 2.725 | 2.785 | 2.466 | 4.397 | **4,489** | 4.170 | 2.684 |
| 6.455 | 7.185 | 7,863 | 3.065 | 4.945 | 2.619 | 1.508 | 1,379 | 1.302 | 1.192 |
| 4.833 | 1.780 | 2,271 | 2.454 | 1.586 | 2.595 | 2.939 | 1,324 | 1.128 | 4.257 |
| 1.653 | 2.624 | 0,779 | 0.238 | 0.172 | 1.338 | 2.313 | 1,290 | 1.440 | 2.493 |
| 5.029 | 3.679 | **3,562** | 4.295 | 3.063 | 5.534 | 4.939 | 4,058 | 5.257 | 4.231 |
| 3.027 | 6.997 | 3,002 | 3.647 | 1.625 | 2.274 | 1.651 | 3,216 | 4.641 | 1.289 |
| 4.947 | 4.069 | **4,993** | 4.715 | 5.132 | 4.946 | 4.934 | 0,827 | 4.110 | 4.323 |
| 1.047 | 1.023 | 4,330 | 3.551 | 4.398 | 2.603 | 1.513 | 1,317 | 4.113 | 1.171 |
| 2.160 | 3.286 | **4,052** | 5.008 | 5.835 | 4.443 | 5.692 | 6,458 | 6.420 | 6.410 |
| 3.437 | 4.315 | 2,402 | 3.724 | 4.977 | 2.237 | 3.348 | 3,577 | 4.924 | 3.505 |
| 2.650 | 4.631 | 4,361 | 2.749 | 4.810 | 4.374 | 2.653 | 2,303 | 2.003 | 4.456 |
| 4.387 | 4.031 | **2,075** | 1.303 | 2.059 | 3.308 | 2.004 | 4,271 | 4.820 | 3.195 |
| 3.349 | 2.733 | 2,041 | 4.734 | 3.214 | 3.010 | 2.136 | **2,078** | 1.895 | 2.561 |
| 5.950 | 5.100 | 5,241 | 8.751 | 8.797 | 5.607 | 6.784 | 5,941 | 8.083 | 5.750 |
| 2.180 | 4.490 | 1,422 | 3.254 | 2.905 | 3.984 | 4.586 | **3,642** | 4.684 | 5.501 |
| 1.096 | 3.067 | 1,154 | 3.048 | 2.318 | 2.521 | 2.126 | 1,073 | 4.016 | 4.150 |
| 2.541 | 3.118 | **1,129** | 1.888 | 3.642 | 1.282 | 3.155 | 0,424 | 3.997 | 1.188 |
| 1.782 | 2.558 | 1,205 | 1.638 | 2.784 | 3.678 | 3.476 | 1,468 | 1.700 | 1.718 |
| 5.260 | 2.272 | **3,347** | 1.292 | 4.339 | 4.104 | 2.877 | 3,287 | 3.006 | 2.248 |
| 3.872 | 3.320 | 1,821 | 3.069 | 1.131 | 3.017 | 1.615 | 1,421 | 3.691 | 2.732 |
| 0.977 | 5.323 | 3,878 | 5.360 | 1.664 | 1.563 | 3.183 | 1,979 | 1.301 | 3.020 |
| 1.538 | 0.544 | **1,922** | 1.379 | 4.166 | 1.871 | 2.308 | 4,817 | 3.755 | 1.849 |
| 3.097 | 3.744 | 2,224 | 2.974 | 2.029 | 3.689 | 3.154 | 0,622 | 2.684 | 3.376 |
| 2.264 | 1.749 | **1,348** | 1.056 | 2.011 | 1.089 | 1.400 | 1,754 | 2.505 | 2.449 |
| 1.922 | 1.135 | 2,030 | 2.992 | 1.665 | 1.782 | 3.061 | **4,910** | 3.322 | 3.144 |
| 5.309 | 1.632 | 5,489 | 0.409 | 6.785 | 5.881 | 5.931 | **6,289** | 7.342 | 5.589 |
| 2.024 | 2.555 | 3,541 | 3.185 | 1.807 | 1.535 | 2.964 | 3,691 | 1.676 | 1.626 |
| 1.962 | 1.450 | 2,667 | 3.870 | 4.081 | 1.627 | 3.066 | 4,395 | 4.515 | 3.001 |
| 3.514 | 6.951 | 1,244 | 2.751 | 2.468 | 2.018 | 2.323 | 1,230 | 4.707 | 1.959 |
| 0.750 | 0.800 | 0,449 | 1.177 | 1.890 | 1.178 | 2.311 | **0,697** | 0.035 | 1.687 |
| 0.537 | 1.374 | 1,158 | 5.727 | 1.508 | 5.355 | 1.709 | **5,407** | 1.359 | 1.518 |
| 10.796 | 10.465 | 10,521 | 8.610 | 8.558 | 8.599 | 8.032 | 10,596 | 8.551 | 8.123 |
| 1.056 | 1.232 | 2,367 | 1.325 | 0.526 | 1.676 | 2.971 | 2,542 | 2.939 | 2.814 |
| 8.118 | 8.500 | 9,924 | 9.254 | 10.405 | 9.636 | 10.348 | 9,218 | 9.805 | 10.046 |
| 4.040 | 4.244 | 3,613 | 3.099 | 4.680 | 6.852 | 2.452 | **3,192** | 4.986 | 5.924 |
| 4.949 | 7.182 | 4,366 | 3.236 | 4.999 | 3.716 | 3.930 | **4,846** | 7.191 | 4.460 |
| 3.296 | 7.442 | 7,542 | 4.733 | 4.720 | 3.237 | 3.182 | **4,782** | 5.351 | 3.424 |
| 5.786 | 6.620 | 6,717 | 5.252 | 5.305 | 5.491 | 6.526 | **5,493** | 6.722 | 4.637 |

En la Tabla 4.35 se realiza una comparación entre el valor real y el valor con imputación por la media y regresión.

|  |  |
| --- | --- |
| **Tabla 4.35**  *Efectos de la Imputación en el análisis de datos multivariados*  **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 | |
| *25% de datos completados en X3 por la Media*   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Imputación por Media** | **Error**  **|Dato Observado –**  **Resultado de Imputación por Media|** | | 2.851 | 3.219 | 0,368 | | 1.414 | 3.219 | 1,805 | | 1.069 | 3.219 | 2,150 | | 6.462 | 3.219 | 3,243 | | 3.914 | 3.219 | 0,695 | | 1.131 | 3.219 | 2,088 | | 6.562 | 3.219 | 3,343 | | 2.254 | 3.219 | 0,965 | | 1.689 | 3.219 | 1,530 | | 1.432 | 3.219 | 1,787 | | 3.693 | 3.219 | 0,474 | | 3.960 | 3.219 | 0,741 | | 3.420 | 3.219 | 0,201 | | 2.683 | 3.219 | 0,536 | | 6.730 | 3.219 | 3,511 | | 0.860 | 3.219 | 2,359 | | 6.406 | 3.219 | 3,187 | | 3.578 | 3.219 | 0,359 | | 5.157 | 3.219 | 1,938 | | 4.083 | 3.219 | 0,864 | | 2.061 | 3.219 | 1,158 | | 1.148 | 3.219 | 2,071 | | 3.359 | 3.219 | 0,140 | | 1.913 | 3.219 | 1,306 | | 1.351 | 3.219 | 1,868 | | *25% de datos completados en X3 por Regresión*   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Predicción** | **Error**  **| Dato Observado –**  **Resultado de Predicción |** | | 2.851 | 2.849 | 0,002 | | 1.414 | 1.403 | 0,011 | | 1.069 | 1.057 | 0,012 | | 6.462 | 6.399 | 0,063 | | 3.914 | 3.909 | 0,005 | | 1.131 | 1.098 | 0,033 | | 6.562 | 6.554 | 0,008 | | 2.254 | 2.226 | 0,028 | | 1.689 | 1.673 | 0,016 | | 1.432 | 1.429 | 0,003 | | 3.693 | 3.688 | 0,005 | | 3.960 | 3.952 | 0,008 | | 3.420 | 3.411 | 0,009 | | 2.683 | 2.689 | 0,006 | | 6.730 | 6.713 | 0,017 | | 0.860 | 0.853 | 0,007 | | 6.406 | 6.397 | 0,009 | | 3.578 | 3.562 | 0,016 | | 5.157 | 4.993 | 0,164 | | 4.083 | 4.052 | 0,031 | | 2.061 | 2.075 | 0,014 | | 1.148 | 1.129 | 0,019 | | 3.359 | 3.347 | 0,012 | | 1.913 | 1.922 | 0,009 | | 1.351 | 1.348 | 0,003 | |

**Elaborado por**: G. Cuenca

**Continúa…**

**Viene…**

|  |  |
| --- | --- |
| *Efectos de la Imputación en el análisis de datos multivariados*  **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 | |
| *25% de datos completados en X8 por la Media*   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Imputación por Media** | **Error**  **|Dato Observado –**  **Resultado de Imputación por Media|** | | 2.390 | 3.298 | 0,908 | | 1.060 | 3.298 | 2,238 | | 1.383 | 3.298 | 1,915 | | 1.219 | 3.298 | 2,079 | | 2.582 | 3.298 | 0,716 | | 5.997 | 3.298 | 2,699 | | 3.952 | 3.298 | 0,654 | | 19.664 | 3.298 | 16,366 | | 5.859 | 3.298 | 2,561 | | 5.255 | 3.298 | 1,957 | | 9.518 | 3.298 | 6,22 | | 2.947 | 3.298 | 0,351 | | 2.566 | 3.298 | 0,732 | | 0.929 | 3.298 | 2,369 | | 4.580 | 3.298 | 1,282 | | 2.080 | 3.298 | 1,218 | | 3.767 | 3.298 | 0,469 | | 4.930 | 3.298 | 1,632 | | 6.314 | 3.298 | 3,016 | | 0.704 | 3.298 | 2,594 | | 5.413 | 3.298 | 2,115 | | 3.183 | 3.298 | 0,115 | | 4.859 | 3.298 | 1,561 | | 4.800 | 3.298 | 1,502 | | 5.525 | 3.298 | 2,227 | | *25% de datos completados en X8 por Regresión*   |  |  |  | | --- | --- | --- | | **Dato Observado** | **Resultado de Predicción** | **Error**  **| Dato Observado –**  **Resultado de Predicción |** | | 2.390 | 2.386 | 0,004 | | 1.060 | 1.102 | 0,042 | | 1.383 | 1.374 | 0,009 | | 1.219 | 1.207 | 0,012 | | 2.582 | 2.601 | 0,019 | | 5.997 | 6.003 | 0,006 | | 3.952 | 4.007 | 0,055 | | 19.664 | 19.618 | 0,046 | | 5.859 | 5.832 | 0,027 | | 5.255 | 5.243 | 0,012 | | 9.518 | 9.492 | 0,026 | | 2.947 | 3.003 | 0,056 | | 2.566 | 2.572 | 0,006 | | 0.929 | 0.919 | 0,010 | | 4.580 | 4.489 | 0,091 | | 2.080 | 2.078 | 0,002 | | 3.767 | 3.642 | 0,125 | | 4.930 | 4.910 | 0,020 | | 6.314 | 6.289 | 0,025 | | 0.704 | 0.697 | 0,007 | | 5.413 | 5.407 | 0,006 | | 3.183 | 3.192 | 0,009 | | 4.859 | 4.846 | 0,013 | | 4.800 | 4.782 | 0,018 | | 5.525 | 5.493 | 0,032 | |

**Elaborado por**: G. Cuenca

Se puede notar, por medio de la Tabla 4.35 que la diferencia en valor absoluto entre el dato observado y el estimado de cada variable es menor en el Método de Imputación por Regresión.

|  |  |
| --- | --- |
| **CUADRO 4.44**  *Efectos de la Imputación en el Análisis de Datos Multivariados*  **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*** *X****3*” y “*Variable*** *X****8*”** | |
| **Estimadores “*Variable*** *X****3*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 100 | 75 | 100 | 100 | | Media | | 3,206 | 3,219 | 3,219 | 3,201 | | Mediana | | 2,801 | 2,751 | 3,219 | 2,800 | | Moda | | 0,070 | 0,070 | 3,220 | 0,070 | | Varianza | | 4,528 | 4,901 | 3,663 | 4,518 | | Desviación Estándar | | 2,128 | 2,214 | 1,914 | 2,126 | | Error Estándar | | 0,213 | 0,256 | 0,191 | 0,213 | | **Coeficiente de Asimetría** | | 1,194 | 1,294 | 1,486 | 1,198 | | Curtosis | | 1,351 | 1,630 | 3,139 | 1,372 | | Rango | | 10,450 | 10,450 | 10,450 | 10,450 | | Mínimo | | 0,070 | 0,070 | 0,070 | 0,070 | | Máximo | | 10,520 | 10,520 | 10,520 | 10,520 | | Percentiles | 25 | 1,508 | 1,633 | 1,886 | 1,508 | | 50 | 2,801 | 2,751 | 3,219 | 2,800 | | 75 | 4,021 | 4,027 | 3,596 | 4,021 | | **Diagrama de Cajas “*Variable*** *X****3*”** |
| **Estimadores “*Variable*** *X****8*”**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Estimadores** | | **Datos Originales** | **Datos Incompletos** | **Datos Completados por la Media** | **Datos Completados por Regresión** | | n | | 100 | 75 | 100 | 100 | | Media | | 3,588 | 3,298 | 3,298 | 3,585 | | Mediana | | 2,959 | 2,882 | 3,298 | 2,987 | | Moda | | 0,200 | 0,200 | 3,300 | 0,200 | | Varianza | | 7,626 | 5,164 | 3,860 | 7,599 | | Desviación Estándar | | 2,762 | 2,273 | 1,965 | 2,757 | | Error Estándar | | 0,276 | 0,262 | 0,197 | 0,276 | | **Coeficiente de Asimetría** | | 2,576 | 1,484 | 1,704 | 2,573 | | Curtosis | | 11,269 | 2,742 | 4,597 | 11,234 | | Rango | | 19,470 | 11,470 | 11,470 | 19,420 | | Mínimo | | 0,200 | 0,200 | 0,200 | 0,200 | | Máximo | | 19,660 | 11,660 | 11,660 | 19,620 | | Percentiles | 25 | 1,715 | 1,623 | 1,983 | 1,715 | | 50 | 2,959 | 2,882 | 3,298 | 2,987 | | 75 | 4,813 | 4,382 | 3,820 | 4,808 | | **Diagrama de Cajas “*Variable*** *X****8*”** |

**Elaborado por**: G. Cuenca

Al realizar la imputación por la media y regresión se obtuvieron los siguientes resultados en la variable *X****3*** (Ver Cuadro 4.44):

El valor de la media de los “datos completados” por *la media* aumenta, comparándolo con los “datos originales” y completados por *regresión.*

El valor de la varianza de los “datos completados” por la *media* disminuye de 4.528 a 3.663, mientras que en los datos completados por regresión este valor se incrementa a 4.518, comparándolo con el valor anterior y es muy cercano al valor de la varianza de los datos originales.

Mientras que en la variable *X****8,*** el valor de la media de los “datos completados” por *la media* aumenta, comparándolo con los “datos originales” y completados por *regresión.*

El valor de la varianza de los “datos completados” por la *media* disminuye de 7.626 a 3.860. Esta variable presenta valores atípicos.

El vector de medias con veinticinco datos completados por la media en *X3* y veinticinco en *X8* es:



Mientras que el vector de medias con veinticinco datos completados por la regresión en *X3* y veinticinco en *X8* es:



El efecto que causa en la *matriz de varianzas y covarianzas* y *matriz de correlaciones*, el completar 10% de datos faltantes en una matriz de tamaño 100, por medio de la imputación por media y regresión, se presenta en el Cuadro 4.45.

|  |
| --- |
| **CUADRO 4.45**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 4.386 |  |  |  |  |  |  |  |  |  | | *X*2 | 2.700 | 4.854 |  |  |  |  |  |  |  |  | | *X*3 | 2.701 | 2.336 | 4.528 |  |  |  |  |  |  |  | | *X*4 | 2.165 | 2.247 | 2.041 | 3.978 |  |  |  |  |  |  | | *X*5 | 2.780 | 2.346 | 3.072 | 2.968 | 6.029 |  |  |  |  |  | | *X6* | 2.252 | 1.997 | 1.489 | 2.269 | 2.240 | 4.084 |  |  |  |  | | *X7* | 2.706 | 1.857 | 2.925 | 2.329 | 3.696 | 2.695 | 5.563 |  |  |  | | *X8* | 1.637 | 2.954 | 2.226 | 1.497 | 2.897 | 2.173 | 3.059 | 7.626 |  |  | | *X9* | 3.019 | 2.133 | 2.366 | 2.365 | 3.272 | 2.508 | 3.039 | 1.543 | 5.322 |  | | *X10* | 2.552 | 2.045 | 2.371 | 2.044 | 2.521 | 3.192 | 3.685 | 2.716 | 2.939 | 5.072 | |
| **Matriz de Varianzas y Covarianzas**  **25% Datos Completados por Media en “*Variable*** *X****3*” y 25% en “*Variable*** *X****8*”**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 4.386 |  |  |  |  |  |  |  |  |  | | *X*2 | 2.700 | 4.854 |  |  |  |  |  |  |  |  | | *X*3 | 2.402 | 2.182 | 3.663 |  |  |  |  |  |  |  | | *X*4 | 2.165 | 2.247 | 1.826 | 3.978 |  |  |  |  |  |  | | *X*5 | 2.780 | 2.346 | 2.401 | 2.968 | 6.029 |  |  |  |  |  | | *X6* | 2.252 | 1.997 | 1.509 | 2.269 | 2.240 | 4.084 |  |  |  |  | | *X7* | 2.706 | 1.857 | 2.224 | 2.329 | 3.696 | 2.695 | 5.563 |  |  |  | | *X8* | 1.611 | 1.419 | 1.629 | 1.522 | 1.735 | 1.728 | 2.697 | 3.860 |  |  | | *X9* | 3.019 | 2.133 | 2.404 | 2.365 | 3.272 | 2.508 | 3.039 | 1.139 | 5.322 |  | | *X10* | 2.552 | 2.045 | 2.138 | 2.044 | 2.521 | 3.192 | **3.685** | 2.296 | 2.939 | 5.072 | |
| **Matriz de Varianzas y Covarianzas**  **25% Datos Completados por Regresión en “*Variable*** *X****3*” y 25% en “*Variable*** *X****8*”**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 4.386 |  |  |  |  |  |  |  |  |  | | *X*2 | 2.700 | 4.854 |  |  |  |  |  |  |  |  | | *X*3 | 2.697 | 2.335 | 4.518 |  |  |  |  |  |  |  | | *X*4 | 2.165 | 2.247 | 2.037 | 3.978 |  |  |  |  |  |  | | *X*5 | 2.780 | 2.346 | 3.065 | 2.968 | 6.029 |  |  |  |  |  | | *X6* | 2.252 | 1.997 | 1.487 | 2.269 | 2.240 | 4.084 |  |  |  |  | | *X7* | 2.706 | 1.857 | 2.919 | 2.329 | 3.696 | 2.695 | 5.563 |  |  |  | | *X8* | 1.639 | 2.950 | 2.228 | 1.498 | 2.895 | 2.172 | 3.056 | 7.599 |  |  | | *X9* | 3.019 | 2.133 | 2.367 | 2.365 | 3.272 | 2.508 | 3.039 | 1.536 | 5.322 |  | | *X10* | 2.552 | 2.045 | 2.369 | 2.044 | 2.521 | 3.192 | 3.685 | 2.713 | 2.939 | 5.072 | |

**Elaborado por**: G. Cuenca

|  |
| --- |
| **CUADRO 4.46**  *Efectos de la Imputación en el análisis de datos multivariados*  **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**  **(Datos Originales)**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 1.000 |  |  |  |  |  |  |  |  |  | | *X*2 | 0.585 | 1.000 |  |  |  |  |  |  |  |  | | *X*3 | 0.606 | 0.498 | 1.000 |  |  |  |  |  |  |  | | *X*4 | 0.518 | 0.511 | 0.481 | 1.000 |  |  |  |  |  |  | | *X*5 | 0.541 | 0.434 | 0.588 | 0.606 | 1.000 |  |  |  |  |  | | *X6* | 0.532 | 0.448 | 0.346 | 0.563 | 0.451 | 1.000 |  |  |  |  | | *X7* | 0.548 | 0.357 | 0.583 | 0.495 | 0.638 | 0.565 | 1.000 |  |  |  | | *X8* | 0.283 | 0.486 | 0.379 | 0.272 | 0.427 | 0.389 | 0.470 | 1.000 |  |  | | *X9* | 0.625 | 0.420 | 0.482 | 0.514 | 0.578 | 0.538 | 0.559 | 0.242 | 1.000 |  | | *X10* | 0.541 | 0.412 | 0.495 | 0.455 | 0.456 | 0.701 | 0.694 | 0.437 | 0.566 | 1.000 | |
| **Matriz de Correlaciones**  **25% Datos Completados por Media en “*Variable*** *X****3*” y 25% en “*Variable*** *X****8*”**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 1.000 |  |  |  |  |  |  |  |  |  | | *X*2 | 0.585 | 1.000 |  |  |  |  |  |  |  |  | | *X*3 | 0.599 | 0.517 | 1.000 |  |  |  |  |  |  |  | | *X*4 | 0.518 | 0.511 | 0.478 | 1.000 |  |  |  |  |  |  | | *X*5 | 0.541 | 0.434 | 0.511 | 0.606 | 1.000 |  |  |  |  |  | | *X6* | 0.532 | 0.448 | 0.390 | 0.563 | 0.451 | 1.000 |  |  |  |  | | *X7* | 0.548 | 0.357 | 0.493 | 0.495 | 0.638 | 0.565 | 1.000 |  |  |  | | *X8* | 0.392 | 0.328 | 0.433 | 0.388 | 0.360 | 0.435 | 0.582 | 1.000 |  |  | | *X9* | 0.625 | 0.420 | 0.544 | 0.514 | 0.578 | 0.538 | 0.559 | 0.251 | 1.000 |  | | *X10* | 0.541 | 0.412 | 0.496 | 0.455 | 0.456 | 0.701 | 0.694 | 0.519 | 0.566 | 1.000 | |
| **Matriz de Correlaciones**  **25% Datos Completados por Regresión en “*Variable*** *X****3*” y 25% en “*Variable*** *X****8*”**   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | *X*1 | *X*2 | *X*3 | *X*4 | *X*5 | *X6* | *X7* | ***X8*** | *X9* | *X10* | | *X*1 | 1.000 |  |  |  |  |  |  |  |  |  | | *X*2 | 0.585 | 1.000 |  |  |  |  |  |  |  |  | | *X*3 | 0.606 | 0.499 | 1.000 |  |  |  |  |  |  |  | | *X*4 | 0.518 | 0.511 | 0.480 | 1.000 |  |  |  |  |  |  | | *X*5 | 0.541 | 0.434 | 0.587 | 0.606 | 1.000 |  |  |  |  |  | | *X6* | 0.532 | 0.448 | 0.346 | 0.563 | 0.451 | 1.000 |  |  |  |  | | *X7* | 0.548 | 0.357 | 0.582 | 0.495 | 0.638 | 0.565 | 1.000 |  |  |  | | *X8* | 0.284 | 0.486 | 0.380 | 0.272 | 0.428 | 0.390 | 0.470 | 1.000 |  |  | | *X9* | 0.625 | 0.420 | 0.483 | 0.514 | 0.578 | 0.538 | 0.559 | 0.242 | 1.000 |  | | *X10* | 0.541 | 0.412 | 0.495 | 0.455 | 0.456 | 0.701 | 0.694 | 0.437 | 0.566 | 1.000 | |

**Elaborado por**: G. Cuenca

Se puede apreciar en el Cuadro 4.45, que los únicos valores que cambian son las covarianzas de la variable *X3* y *X8* con las demás variables, donde la covarianza entre *X3*  y *X5* , disminuye de 3.072 a 2.401.

En la matriz de varianzas y covarianzas de los datos completados por regresión, el valor de las covarianzas de variable *X3* y *X8* con las demás variables se incrementa, comparándolo con la matriz de varianzas y covarianzas de los “datos completados” por *la media.*

Por otro lado, analizando el efecto que causa en la matriz de correlaciones, podemos apreciar en le Cuadro 4.46 que también los únicos valores que cambian son los de la correlación de *X3* y *X8* con las demás variables, puesto que a estas variables se les completó datos por medio de los métodos de imputación; donde la mayor correlación se da entre las variables *X6*  y *X10*, es decir 0.701, seguida por 0.694 entre las variables *X7*  y *X10*. En la matriz de correlaciones con 25% de datos completados por la media en *X3* y 25% en *X8*, la correlación entre *X3*  y *X5* disminuye de 0.588 a 0.511, mientras que en la matriz de datos completados por regresión, este valor es 0.587, es decir tiende al valor observado.