CAPITULO 4

**4.- ANÁLISIS Estadístico MULTIVARIADO**

**4.1 Análisis de Correlación**

Esta técnica estadística está basada en el coeficiente de correlación , definido en términos del cociente entre la covarianza  y el producto de varianzas  y . El coeficiente de correlación proporciona una medida de la asociación lineal entre dos variables. El coeficiente de correlación entre Xj y Xk se denota por y se define por:

****

En general, si dos variables se distribuyen de manera independiente, entonces  y .

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

** =** 



=

Se puede demostrar que  Si  significa que no existe relación lineal entre las variables Xj y Xk y si  entonces existe una perfecta relación lineal.

El total de correlaciones obtenidas fueron 595, correspondientes a las combinaciones de 35 en 2, a partir de las 35 variables seleccionadas para este análisis. El 49.08% de los pares de variables utilizadas para este análisis poseen un coeficiente de correlación negativo y el 50.92% de los pares de variables poseen un coeficiente de correlación positivo.

**4.2 Análisis De Correspondencia**

Uno de los fines del análisis de correspondencias es describir las relaciones existentes entre dos variables nominales, recogidas en una tabla de correspondencias, sobre un espacio de pocas dimensiones, mientras que al mismo tiempo se describen las relaciones entre las categorías de cada variable. Para cada variable, las distancias sobre un gráfico entre los puntos de categorías reflejan las relaciones entre las categorías, con las categorías similares representadas próximas unas a otras. La proyección de los puntos de una variable sobre el vector desde el origen hasta un punto de categoría de la otra variable describe la relación entre ambas variables.

El análisis de las tablas de contingencia a menudo incluye examinar los perfiles de fila y de columna, así como contrastar la independencia a través del estadístico de chi-cuadrado. Sin embargo, el número de perfiles puede ser bastante grande y la prueba de chi-cuadrado no revelará la estructura de la dependencia. El procedimiento Tablas de contingencia ofrece varias medidas y pruebas de asociación pero no puede representar gráficamente ninguna relación entre las variables.

El análisis de correspondencias se puede utilizar para analizar cualquier tabla de medidas de correspondencia que sean positivas.

El análisis de correspondencia calculará, medidas de correspondencia, perfiles de fila y de columna, valores propios, puntuaciones de fila y de columna, inercia, masa, estadísticos de confianza para las puntuaciones de fila y de columna, estadísticos de confianza para los valores propios, gráficos de transformación, gráficos de los puntos de fila, gráficos de los puntos de columna y diagramas de dispersión biespaciales, de las diversas filas y columnas de la tabla y además nos permitirá analizar la posible relación entre las variables, las diversas categorías de las variables estarán representadas en el gráfico más próximas o alejadas en las diversas dimensiones en función de su grado de similitud o diferencia

Procedimientos relacionados. Si hay implicadas más de dos variables, utilice el análisis de homogeneidad. Si se deben escalar las variables de forma ordinal, utilice el análisis de componentes principales mediante escalamiento óptimo.

Para la medida de distancia entre las filas y columnas se puede seleccionar el estadístico Chi-Cuadrado. Utiliza una distancia ponderada entre los perfiles, donde la ponderación es la masa de las filas o de las columnas. Esta distancia es necesaria para el análisis de correspondencias típico.

# 4.2.1 Método de Normalización Simétrico

Para cada dimensión, las puntuaciones de fila son la media ponderada de las puntuaciones de columna divididas por el valor propio coincidente y las puntuaciones de columna son la media ponderada de las puntuaciones de fila divididas por el valor propio coincidente. Utilice este método si desea examinar las diferencias o similaridades entre las categorías de las dos variables.

**4.3 Prueba Chi- Cuadrado**

Chi-cuadrado (Tablas de contingencia)

Estadístico utilizado para contrastar la hipótesis de que las variables de fila y columna son independientes. No debe utilizarse si cualquiera de las casillas tiene un valor esperado menor que 1, o si más de un 20% de las casillas tienen valores esperados menores que 5. A efectos prácticos, el valor de significación es más importante que el valor real del estadístico.

Se presentará a continuación no la tabla de contingencia sino la tabla que hace aprobar o rechazar la hipótesis de dependencia, consideramos mostrar parte del análisis en tablas independientes porque luego resumiremos la dependencia en una sola tabla conjunta.

Tabla 4.1

Prueba Chi-Cuadrada para Capacidad Efectiva y Energía Entregada a EMELESA

|  |  |  |  |
| --- | --- | --- | --- |
| Pruebas de chi-cuadrado | | | |
| Capacidad Efec. Ener. Entreg a EMELESA | *Valor* | *gl* | *Sig. asintótica (bilateral)* |
| *Chi-cuadrado de Pearson* | 41,46 | 8,00 | 0,00 |
| *Razón de verosimilitud* | 49,39 | 8,00 | 0,00 |
| *Asociación lineal por lineal* | 24,07 | 1,00 | 0,00 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

H0: La Capacidad Efectiva y la Energía Entregada a EMELESA.

Vs.

H1: No es verdad H0

Valor del estadístico de prueba *χ*2 = 41.46

Grados de libertad 8

Valor p 0.00

El valor p de la prueba es 0.00 por lo tanto se rechaza la hipótesis nula el cual nos indica que la Capacidad Efectiva Y la energía entregada a EMELESA no son independientes a todo nivel de significancia estadística.

Tabla 4.2

Prueba Chi-Cuadrada para Potencia Máxima y Generación Bruta

|  |  |  |  |
| --- | --- | --- | --- |
| Pruebas de chi-cuadrado | | | |
| Potencia Max y Gen Bruta | *Valor* | *gl* | *Sig. asintótica (bilateral)* |
| *Chi-cuadrado de Pearson* | 40,03 | 8,00 | 0,00 |
| *Razón de verosimilitud* | 38,47 | 8,00 | 0,00 |
| *Asociación lineal por lineal* | 29,03 | 1,00 | 0,00 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

H0: Potencia Máxima y Generación Bruta.

Vs.

H1: No es verdad H0

Valor del estadístico de prueba *χ*2 = 40.03

Grados de libertad 8

Valor p 0.00

El valor p de la prueba es 0.00 por lo tanto se rechaza la hipótesis nula el cual nos indica que la Potencia Máxima y Generación Bruta no son independientes a todo nivel de significancia estadística.

Tabla 4.3

Prueba Chi-Cuadrada para Potencia Mínima y Horas Periodo

|  |  |  |  |
| --- | --- | --- | --- |
| Pruebas de chi-cuadrado | | | |
| Pot. Mín. Horas Periodo | *Valor* | *gl* | *Sig. asintótica (bilateral)* |
| *Chi-cuadrado de Pearson* | 53,11 | 12,00 | 0,00 |
| *Razón de verosimilitud* | 20,36 | 12,00 | 0,06 |
| *Asociación lineal por lineal* | 10,53 | 1,00 | 0,00 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

H0: Potencia Mínima y Horas Periodo.

Vs.

H1: No es verdad H0

Valor del estadístico de prueba *χ*2 = 53.11

Grados de libertad 12

Valor p 0.00

El valor p de la prueba es 0.00 por lo tanto se rechaza la hipótesis nula el cual nos indica que la Potencia Mínima y Horas Periodo no son independientes a todo nivel de significancia estadística.

Tabla 4.4

Prueba Chi-Cuadrada para Consumo Interno UES y Consumo Interno UFS

|  |  |  |  |
| --- | --- | --- | --- |
| Pruebas de chi-cuadrado | | | |
| Consum UES y Cons UFS | *Valor* | *gl* | *Sig. asintótica (bilateral)* |
| *Chi-cuadrado de Pearson* | 91,21 | 16,00 | 0,00 |
| *Razón de verosimilitud* | 79,15 | 16,00 | 0,00 |
| *Asociación lineal por lineal* | 34,69 | 1,00 | 0,00 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

H0: Consumo Interno UES y Consumo Interno UFS

Vs.

H1: No es verdad H0

Valor del estadístico de prueba *χ*2 = 91.21

Grados de libertad 16

Valor p 0.00

El valor p de la prueba es 0.00 por lo tanto se rechaza la hipótesis nula el cual nos indica que el Consumo Interno UES y Consumo Interno UFS no son independientes a todo nivel de significancia estadística.

Tabla 4.5

Prueba Chi-Cuadrada para Generación Neta y Horas Operación

|  |  |  |  |
| --- | --- | --- | --- |
| Pruebas de chi-cuadrado | | | |
| Gen Neta y Horas Operación | *Valor* | *gl* | *Sig. asintótica (bilateral)* |
| *Chi-cuadrado de Pearson* | 151,06 | 16,00 | 0,00 |
| *Razón de verosimilitud* | 121,63 | 16,00 | 0,00 |
| *Asociación lineal por lineal* | 57,40 | 1,00 | 0,00 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

H0: Generación Neta y Horas Operación

Vs.

H1: No es verdad H0

Valor del estadístico de prueba *χ*2 = 151.06

Grados de libertad 16

Valor p 0.00

El valor p de la prueba es 0.00 por lo tanto se rechaza la hipótesis nula el cual nos indica que la Generación Neta y Horas Operación no son independientes a todo nivel de significancia estadística.

Tabla 4.6

Prueba Chi-Cuadrada para Horas Disponibles y Energía no Generada por Mantenimiento Programado

|  |  |  |  |
| --- | --- | --- | --- |
| Pruebas de chi-cuadrado | | | |
| Horas Disp. Energía no gen por Mto. Program. | *Valor* | *gl* | *Sig. asintótica (bilateral)* |
| *Chi-cuadrado de Pearson* | 170,99 | 16,00 | 0,00 |
| *Razón de verosimilitud* | 110,24 | 16,00 | 0,00 |
| *Asociación lineal por lineal* | 65,91 | 1,00 | 0,00 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

H0: Horas Disponibles y Energía no Generada por Mantenimiento Programado.

Vs.

H1: No es verdad H0

Valor del estadístico de prueba *χ*2 = 170,99

Grados de libertad 16

Valor p 0.00

El valor p de la prueba es 0.00 por lo tanto se rechaza la hipótesis nula el cual nos indica que las Horas Disponibles y Energía no Generada por Mantenimiento Programado no son independientes a todo nivel de significancia estadística.

En el anexo 1 se presenta una tabla que contiene todas las variables dependientes, con sus respectivos valores del estadístico de prueba, nivel de significancia y grados de libertad.

Marcado se encontrará en la tabla mencionada casillas marcadas y en negrillas para el índice Factor de Plata y las variables de las cuales él depende, estas son: Potencia Máxima, Generación Bruta, Consumo Interno UES y UFS, Generación Neta, Horas Operación, Horas Disponibles, Horas Equivalentes en Reserva, Energía no Generada, Horas fuera de servicio por mantenimiento Programado, Reserva al sistema, Energía Programada por el CENACE, Energía no generada por mantenimiento Programado, Consumo de Fuel y además con los otros Índices como: Disponibilidad, Indisponibilidad y Factor de Reserva al Sistema. Se dará un nuevo análisis para tratar de reducir estas dependencias a las más importantes. En total el índice Factor de Planta depende de 18 variables, lo que no ocurre con la Capacidad Efectiva que solo depende de 2, estas son: Decremento de Capacidad Original y Energía Entregada a EMELESA. Se puede observar que Potencia máxima, Generación Bruta y Generación Neta son las variables que mayor dependencias tienen con otras variables.

Para los Índices se realizará un Análisis de Correspondencia, teniendo presente los valores de la matriz del anexo 1.

**4.4 ANÁLISIS DE CORRESPONDENCIA SIMPLE**

**4.4.1 Análisis de Correspondencia para variables Factor de Planta e Indisponibilidad.**

Tabla 4.7

Análisis de Correspondencia para Índices Factor de Planta e Indisponibilidad

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Examen de los puntos de fila | | | | | | | | | |
| FAC. PLAN | Masa | Puntuación en  la dimensión | | Inercia | Contribución | | | | |
| De los puntos  a la inercia de la dimensión | | De la dimensión  a la inercia del punto | |  |
|  |  | 1 | 2 | 1 | 2 | Total |
| 0-0,329 | 0,26 | -1,31 | -0,46 | 0,34 | 0,64 | 0,10 | 0,91 | 0,09 | 1,00 |
| 0,33-0,659 | 0,22 | -0,08 | 1,44 | 0,26 | 0,00 | 0,78 | 0,00 | 1,00 | 1,00 |
| 0,66-0,999 | 0,52 | 0,69 | -0,37 | 0,21 | 0,36 | 0,12 | 0,81 | 0,19 | 1,00 |
| Total activo | 1,00 |  |  | 0,82 | 1,00 | 1,00 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| Examen de los puntos columna | | | | | | | | | |
| INDISP. | Masa | Puntuación en  la dimensión | | Inercia | Contribución | | | | |
| De los puntos  a la inercia de la dimensión | | De la dimensión  a la inercia del punto | |  |
|  |  | 1 | 2 | 1 | 2 | Total |
| 0-0,249 | 0,71 | 0,48 | -0,21 | 0,13 | 0,23 | 0,06 | 0,86 | 0,14 | 1,00 |
| 0,25-0,49 | 0,10 | -0,37 | 2,03 | 0,25 | 0,02 | 0,73 | 0,04 | 0,96 | 1,00 |
| 0,50-0,749 | 0,06 | -1,00 | 0,85 | 0,06 | 0,08 | 0,07 | 0,63 | 0,37 | 1,00 |
| 0,75-1,01 | 0,13 | -1,88 | -0,80 | 0,37 | 0,66 | 0,15 | 0,87 | 0,13 | 1,00 |
| Total activo | 1,00 |  |  | 0,82 | 1,00 | 1,00 |  |  |  |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

El Factor de Planta presenta una mejor explicación para su varianza en la dimensión 1 con un valor de 0,64 que corresponde a los valores entre 0 y 0,249; de 0,36 para valores entre 0,66 y 0,99. En la dimensión 2 se da entre los valores 0.33 y 0,659; para el índice de Indisponibilidad se obtiene una buena representatividad de los valores en el primer y cuarto intervalo con 0,86 y 0,87 respectivamente. Las sumas ponderadas de las distancias al centroide dividida entre la suma de las celdas en la tabla es mayor en el primer intervalo para el Factor de Planta con un valor de 0,34; para la Indisponibilidad es 0,37 en el último intervalo.

### 4.4.2 Análisis de Correspondencia Simple para las Variables Factor de Reserva al Sistema y Factor de Decremento

La siguiente tabla presenta dos variables no correlacionadas, observe los niveles bajos de inercia y el nivel de significancia alto.

Tabla 4.8

Análisis de Correspondencia para Índices Factor de Reserva al Sistema y Factor de Decremento

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Resumen | | | | | Proporción de inercia | | Confianza para el Valor propio |  |
|  | Valor propio | Inercia | Chi-cuadrado | Sig. | Explicada | Acumulada | Desviación típica | Correlación |
| Dimensión |  |  | 7,73 | 0,56 |  |  |  | 2 |
| 1,00 | 0,31 | 0,10 | 0,86 | 0,86 | 0,12 | -0,17 |
| 2,00 | 0,13 | 0,02 | 0,14 | 1,00 | 0,04 |  |
| Total |  | 0,11 | 1,00 | 1,00 |  |
| 9 grados de libertad | | |  | |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

Tabla 4.9

Análisis de Correspondencia para Índices Factor de Reserva al Sistema y Factor de Decremento

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Examen de los puntos de fila | | | | | | | | | |
| FAC. RESR. | Masa | Puntuación  en la dimensión | | Inercia | Contribución | | | | |
| De los puntos  a la inercia de la dimensión | | De la dimensión  a la inercia del punto | |  |
|  |  | 1 | 2 | 1 | 2 | Total |
| 0-0,249 | 0,72 | -0,24 | -0,15 | 0,02 | 0,14 | 0,14 | 0,86 | 0,14 | 1,00 |
| 0,25-0,49 | 0,17 | 1,21 | 0,04 | 0,08 | 0,82 | 0,00 | 1,00 | 0,00 | 1,00 |
| 0,50-0,749 | 0,01 | -0,34 | 1,03 | 0,00 | 0,01 | 0,12 | 0,21 | 0,79 | 1,00 |
| 0,75-1,01 | 0,09 | -0,34 | 1,03 | 0,01 | 0,03 | 0,74 | 0,21 | 0,79 | 1,00 |
| Total activo | 1,00 |  |  | 0,11 | 1,00 | 1,00 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| Examen de los puntos columna | | | | | | | | | |
| FAC. DECR. | Masa | Puntuación  en la dimensión | | Inercia | Contribución | | | | |
| De los puntos  a la inercia de la dimensión | | De la dimensión  a la inercia del punto | |  |
|  |  | 1 | 2 | 1 | 2 | Total |
| 0-0,0219 | 0,86 | -0,11 | 0,13 | 0,00 | 0,03 | 0,11 | 0,62 | 0,38 | 1,00 |
| 0,022-0,0439 | 0,09 | 0,78 | -0,71 | 0,02 | 0,17 | 0,35 | 0,75 | 0,25 | 1,00 |
| 0,044-0,0659 | 0,04 | -0,78 | -1,23 | 0,02 | 0,09 | 0,52 | 0,50 | 0,50 | 1,00 |
| 0,066-0,088 | 0,01 | 3,91 | 0,32 | 0,07 | 0,71 | 0,01 | 1,00 | 0,00 | 1,00 |
| Total activo | 1,00 |  |  | 0,11 | 1,00 | 1,00 |  |  |  |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

Podemos observar que la mejor explicación para la variabilidad de los datos es 1 en el intervalo de 0,066 y 0,08, pero las inercias son bajas, se debe tomar en cuenta la gráfica individual de cada variable ya que no están correlacionadas fuertemente.

**4.4.3 Análisis de Correspondencia Simple para las Variables Factor de Falla e Factor de Decremento.**

Tabla 4.10

Análisis de Correspondencia para Índices Factor de Falla y Factor de Decremento

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Examen de los puntos de fila | | | | | | | | | |
| FAC. FALLA | Masa | Puntuación en  la dimensión | | Inercia | Contribución | | | | |
| De los puntos a  la inercia de la dimensión | | De la dimensión  a la inercia del punto | |  |
|  |  | 1 | 2 | 1 | 2 | Total |
| 0-0,04 | 0,97 | -0,07 | 0,02 | 0,00 | 0,01 | 0,02 | 0,99 | 0,01 | 1,00 |
| 0,04-0,08 | 0,01 | 5,16 | -0,12 | 0,15 | 0,98 | 0,01 | 1,00 | 0,00 | 1,00 |
| 0,08-0,12 | 0,00 | , | , | , | , | , | , | , | , |
| 0,16-0,20 | 0,01 | -0,49 | -1,49 | 0,00 | 0,01 | 0,98 | 0,57 | 0,43 | 1,00 |
| Total activo | 1,00 |  |  | 0,16 | 1,00 | 1,00 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| Examen de los puntos columna | | | | | | | | | |
| FAC. DEC. | Masa | Puntuación en  la dimensión | | Inercia | Contribución | | | | |
| De los puntos a la  inercia de la dimensión | | De la dimensión a  la inercia del punto | |  |
|  |  | 1 | 2 | 1 | 2 | Total |
| 0-0,0219 | 0,86 | -0,19 | -0,05 | 0,01 | 0,08 | 0,06 | 0,99 | 0,01 | 1,00 |
| 0,022-0,0439 | 0,09 | 2,03 | 0,00 | 0,14 | 0,91 | 0,00 | 1,00 | 0,00 | 1,00 |
| 0,044-0,0659 | 0,04 | -0,18 | 0,73 | 0,00 | 0,00 | 0,70 | 0,41 | 0,59 | 1,00 |
| 0,066-0,088 | 0,01 | -0,18 | 0,73 | 0,00 | 0,00 | 0,23 | 0,41 | 0,59 | 1,00 |
| Total activo | 1,00 |  |  | 0,16 | 1,00 | 1,00 |  |  |  |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

Resultan independientes este par de variables con un nivel de significancia 0,29 y es estadístico Chi-Cuadrado de 10,97; pues bien, la variabilidad con respecto a Factor de Falla para la dimensión 1 es 0,98 en que pertenece al intervalo 0,04 y 0,08 este último valor sin incluir; para la dimensión 2 es 0,98 en el último intervalo que es 0,16 a 0,20. En términos practico esto resulta impositivo; porque si recordamos la Meta de este índice es 0, y si los datos se dan en estos intervalos con tal frecuencia, no se están cumpliendo las metas, Para la segunda parte de esta tabla tenemos el Factor de Decremento que se explica muy bien en el segundo datos es 1 en el intervalo.

Son Independientes también el Consumo Específico y el Decremento de la Capacidad Original, presenta un nivel de significancia de 0,999; los valores de inercia en las tablas de correspondencias son 0. Hay independencia entre el Consumo Específico y la Disponibilidad.

## 4.5 Análisis de Homogeneidad

El análisis de homogeneidad cuantifica los datos (categóricos) nominales mediante la asignación de valores numéricos a los casos (los objetos) y a las categorías. El análisis de homogeneidad se conoce también por el acrónimo HOMALS, del inglés Homogeneity Analysis by Means of Alternating Least Squares (Análisis de Homogeneidad Mediante Mínimos Cuadrados Alternantes).

El objetivo de HOMALS es describir las relaciones entre dos o más variables nominales en un espacio de pocas dimensiones que contiene las categorías de las variables así como los objetos pertenecientes a dichas categorías. Los objetos pertenecientes a la misma categoría se representan cerca los unos de los otros, mientras que los objetos de diferentes categorías se representan alejados los unos de los otros. Cada objeto se encuentra lo más cerca posible de los puntos de categoría para las categorías a las que pertenece dicho objeto.

El análisis de homogeneidad es similar al análisis de correspondencias, pero no está limitado a dos variables. Es por ello que el análisis de homogeneidad se conoce también como el análisis de correspondencias múltiple. También se puede ver el análisis de homogeneidad como un análisis de componentes principales para datos nominales.

El análisis de homogeneidad es más adecuado que el análisis de componentes principales típico cuando puede que no se conserven las relaciones lineales entre las variables, o cuando las variables se miden a nivel nominal. Además, la interpretación del resultado es mucho más sencilla en HOMALS que en otras técnicas categóricas, como pueden ser las tablas de contingencia y los modelos loglineales. Debido a que las categorías de las variables son cuantificadas, se pueden aplicar sobre las cuantificaciones técnicas que requieren datos numéricos, en análisis subsiguientes.

El análisis de homogeneidad calculara, frecuencias, autovalores, historial de iteraciones, puntuaciones de objeto, cuantificaciones de categoría, medidas de discriminación, gráficos de las puntuaciones de objeto, gráficos de las cuantificaciones de categoría, gráficos de las medidas de discriminación.

Todas las variables del análisis tienen cuantificaciones de categoría que pueden diferir para cada dimensión (nominal múltiple). En el análisis, sólo se utiliza un conjunto de variables. El número máximo de dimensiones utilizado en el procedimiento es el más pequeño entre el número total de categorías menos el número de variables sin datos perdidos y el número de casos menos 1. Por ejemplo, si una variable dispone de cinco categorías y la otra de cuatro (sin datos perdidos), el número máximo de dimensiones es siete ((5+4) - 2). Si especifica un número superior al máximo, se utilizará el valor máximo.

Procedimientos relacionados. Para dos variables, el Análisis de homogeneidad es análogo al Análisis de correspondencias. Si piensa que las variables poseen propiedades ordinales o numéricas, se deben utilizar Componentes principales mediante escalamiento óptimo. Si hay conjuntos de variables que son de interés, se debe utilizar el Análisis de correlación canónica no lineal.´

* + 1. **Análisis de Homogeneidad para los Índices y Consumo de Fuel Oil**

Autovalores

|  |  |  |
| --- | --- | --- |
| Dimensión | Autovalores |  |
| 1 | ,388 |  |
| 2 | ,316 |  |

El porcentaje de varianza es explicado por la dimensión uno con 0,388 y 0,316 para la dimensión 2.

Tabla 4.11

Medidas de discriminación para Índices y Fuel

|  |  |  |  |
| --- | --- | --- | --- |
|  | Dimensión |  |  |
|  | 1 | 2 |  |
| Confiabilidad | ,236 | ,002 |  |
| Factor de Planta | ,873 | ,647 |  |
| Disponibilidad | ,614 | ,864 |  |
| Indisponibilidad | ,614 | ,864 |  |
| Factor Resv. | ,250 | ,146 |  |
| Factor Decremento | ,120 | ,038 |  |
| Factor Falla | ,022 | ,156 |  |
| Decremento Cap. Org | ,076 | ,032 |  |
| Consumo Especìfico | ,276 | ,017 |  |
| Consumo Fuel | ,742 | ,314 |  |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

La tabla 4.11 presenta los valores discriminantes; las variables que discriminan mejor en la dimensión 1, son el Factor de Planta, Consumo de Fuel a Disponibilidad e Indisponibilidad, en la dimensión 2 Factor de Planta, Disponibilidad e Indisponibilidad.

Gráfico 4.1



Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

El gráfico muestra mejor lo descrito en el párrafo anterior, las variables que se encuentran encerradas en el círculo son los índices que se dan en los primeros intervalos en su mayoría, más alejados se encuentra el consumo de fuel, Factor de Planta que son valores más altos, es decir se requiere de altas cantidades de combustible para mantener un Factor de planta alto y relativamente la confiabilidad de la misma forma. Son variables alejadas de la tendencia.

Tabla 4.12

Medidas de Saturación para Índices

|  |  |  |  |
| --- | --- | --- | --- |
|  | Dimensión |  |  |
|  | 1 | 2 |  |
| Confiabilidad | ,546 | ,265 |  |
| Fac. Plan. | ,902 | -,317 |  |
| Disponibil | ,875 | ,410 |  |
| Indisponibil | -,875 | -,410 |  |
| Fac. Resv | -,222 | ,892 |  |
| Fac Decre | ,282 | -,205 |  |
| Fac Falla | ,104 | -,229 |  |
| Dec Cap Org | -,175 | ,758 |  |
| Consu Espe | ,473 | -,346 |  |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

Gráfico 4.2



Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

La tabla 4.12 junto a la gráfica 4.2 muestran que las saturaciones de las componentes son menores a uno, pero cuatro de estas están sobre 0,5; posiblemente debería redefinirse la codificación o quitarlas del análisis.

* + 1. **Análisis de Homogeneidad Principales Variables**

En el gráfico 4.3 se resume las puntuaciones de las variables que intervienen el en análisis, adjuntas a él. Vemos que en el círculo derecho, se concentra la mayor cantidad de datos, los girasoles tienen

Gráfico 4.3



|  |  |
| --- | --- |
| Cap. Efectiva |  |
| Pot. Máxima |  |
| Gen Bruta |  |
| Con Interno UES |  |
| Horas Operación |  |
| Hora Falla |  |
| Con Fuel |  |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

un máximo de 14 pétalos y mínimo uno. En el círculo de la izquierda hasta 3 pétalos máximos por girasol, en el rectángulo los girasoles no tienen pétalos. La concentración de datos derecha es superior a 0 en la dimensión 1 y entre –1 y 1 en la dimensión 2.

Las saturaciones se explican en el siguiente gráfico, 4.4 y la tabla 4.13, pretenden determinar el grado de aproximación o lejanía entre las variable, la más lejana de ellas es el consumo de Fuel.

Tabla 4.13

Medidas de Saturación para variables

|  |  |  |  |
| --- | --- | --- | --- |
|  | Dimensión |  |  |
|  | 1 | 2 |  |
| Cap. Efectiva | ,479 | ,385 |  |
| Pot. Máxima | ,859 | ,019 |  |
| Gen Bruta | ,978 | -,049 |  |
| Con Interno UES | ,978 | -,014 |  |
| Horas Operación | ,976 | -,010 |  |
| Hora Falla | ,019 | ,940 |  |
| Con Fuel | ,960 | -,154 |  |

Gráfico 4.4



Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

**4.6 Análisis Factorial para las variables no Índices**

El análisis con las 26 variables se reduce a un considerado número de 6 componentes que nos darán luces para determinar claramente los grupos de variables. Si tomamos los 3 primeros explicaremos el 64,32 % de la variabilidad de los datos.

Tabla 4.14

Autovalores

|  |  |  |
| --- | --- | --- |
| Sumas de las saturaciones al  cuadrado de la extracción |  |  |
| Total | % de la varianza | % acumulado |
| 10,08 | 38,77 | 38,77 |
| 3,92 | 15,07 | 53,84 |
| 2,72 | 10,48 | 64,32 |
| 2,23 | 8,59 | 72,92 |
| 1,88 | 7,23 | 80,15 |
| 1,38 | 5,31 | 85,46 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

Gráfico 4.5

Autovalores

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

Tabla 4.15

Matriz de Componentes Rotados

|  |  |  |  |
| --- | --- | --- | --- |
| Matriz de componentes rotados | | | |
|  | Componente | | |
| 1 | 2 | 3 |
| Cap. Efec | 0,17 | 0,29 | -0,09 |
| Pomas | 0,58 | 0,41 | 0,06 |
| Pmin | 0,22 | 0,09 | -0,04 |
| Gbru | **0,90** | 0,33 | -0,02 |
| Cues | **0,91** | 0,27 | 0,05 |
| Cufs | -0,70 | -0,10 | 0,00 |
| GN | **0,90** | 0,31 | -0,01 |
| Eeem | 0,14 | 0,13 | 0,06 |
| Hper | -0,21 | -0,13 | 0,07 |
| Hopera | **0,93** | 0,24 | 0,06 |
| Herese | -0,19 | **-0,95** | -0,04 |
| Ener no por Dec | 0,14 | -0,05 | 0,70 |
| Ener no Mto. Pro | **-**0,89 | 0,42 | -0,09 |
| Ener no Mat. Core | 0,01 | 0,08 | **0,89** |
| E no generada | **-0,91** | -0,30 | 0,02 |
| Res al Sist | -0,17 | **-0,95** | -0,06 |
| H Mto Pro | **-0,89** | 0,42 | -0,09 |
| H Mto Cor | 0,01 | 0,08 | **0,89** |
| H Fall | -0,06 | 0,04 | -0,02 |
| H Equi Decre | 0,08 | -0,05 | **0,79** |
| H Dispo | 0,89 | -0,43 | 0,02 |
| H Indispo | -0,89 | 0,43 | -0,02 |
| E Disp | -0,06 | 0,10 | 0,01 |
| E Pro CENACE | 0,58 | 0,19 | 0,11 |
| E no por Falla | -0,02 | 0,04 | 0,00 |
| Consumo fuel | 0,86 | 0,35 | -0,02 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

Se realizó una rotación a los datos para alcanzar mayor precisión, los valores marcados en negrillas serán los conjuntos que sirvan para predecir el comportamiento de los índices.

**4.7 Regresión mediante Escalamiento Óptimo (CATREG)**

La regresión mediante escalamiento óptimo cuantifica los datos categóricos mediante la asignación de valores numéricos a las categorías, obteniéndose una ecuación de regresión lineal óptima para las variables transformadas. La regresión mediante escalamiento óptimo se conoce también por el acrónimo CATREG, del inglés categorical regression with optimal scaling (regresión categórica mediante escalamiento óptimo).

El análisis de regresión lineal ordinario implica minimizar las diferencias de la suma de los cuadrados entre una variable de respuesta (la dependiente) y una combinación ponderada de las variables predictoras (las independientes). Las variables son normalmente cuantitativas, con los datos categóricos (nominales) recodificados como variables binarias o de contraste. Como resultado, las variables categóricas sirven para separar grupos de casos y la técnica estima conjuntos separados de parámetros para cada grupo. Los coeficientes estimados reflejan cómo los cambios en los predictores afectan a la respuesta. El pronóstico de la respuesta es posible para cualquier combinación de los valores predictores.

Una aproximación alternativa incluye la regresión de la respuesta respecto a los propios valores predictores categóricos. Como consecuencia, se estima un coeficiente para cada variable. Sin embargo, para las variables categóricas, los valores categóricos son arbitrarios. La codificación de las categorías de diferentes maneras proporciona diferentes coeficientes, dificultando las comparaciones entre los análisis de las mismas variables.

CATREG amplía la aproximación típica mediante un escalamiento de las variables nominales, ordinales y numéricas simultáneamente. El procedimiento cuantifica las variables categóricas de manera que las cuantificaciones reflejen las características de las categorías originales. El procedimiento trata a las variables categóricas cuantificadas como si fueran variables numéricas. La utilización de transformaciones no lineales permite a las variables ser analizadas en varios niveles para encontrar el modelo que más se ajusta.

*Estadístico R*

Medida de asociación lineal entre dos variables. Los valores de R van de -1 (una relación negativa perfecta en la que todos los puntos caen sobre una recta de pendiente negativa) y +1 (una relación positiva perfecta en la que todos los puntos caen sobre una recta de pendiente positiva). Un valor de O indica que no existe relación lineal.

**4.7.1 Regresión para Predecir Índice Decremento Capacidad Original**

|  |  |  |
| --- | --- | --- |
| Resumen del modelo |  |  |
| R múltiple | R cuadrado | R cuadrado corregida |
| 0,69 | 0,47 | 0,42 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA | | | | | |
|  | Suma de cuadrados | gl | Media cuadrática | F | Sig. |
| Regresión | 32,63 | 6,00 | 5,44 | 9,27 | 0,00 |
| Residual | 36,37 | 62,00 | 0,59 |  |  |
| Total | 69,00 | 68,00 |  |  |  |

En las tablas anteriores se puede apreciar un valor de R múltiple el cual indica la relación de las variables con el índice, que en este caso es aceptada porque los confirma la siguiente tabla ANOVA con un nivel de significancia de 0,00 y 6 grados de libertad.

Se presentará a continuación la tabla del modelo y su tolerancia.

Tabla 4.16

Coeficientes de Modelo para Factor de Decremento Capacidad Original

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coeficientes tipificados | | F | Importancia | Tolerancia | |
| Beta | Error típ. | Después de  la transformación | Antes de la  transformación |
| Gen Bruta | -0,48 | 0,32 | 2,22 | 0,27 | 0,08 | 0,04 |
| Cues | -0,13 | 0,39 | 0,11 | 0,10 | 0,06 | 0,03 |
| H Opera | -0,82 | 0,38 | 4,64 | 0,62 | 0,06 | 0,03 |
| E no generad | -0,52 | 0,32 | 2,68 | -0,20 | 0,08 | 0,05 |
| E no Gen pro | -0,61 | 0,15 | 17,44 | 0,19 | 0,40 | 0,03 |
| H Dispo | 0,19 | 0,13 | 2,12 | 0,01 | 0,48 | 0,03 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

La tabla 4.16 presenta los coeficientes tipificados para el modelo de regresión, con las variables, Generación Bruta, Consumo Interno UES, Horas Operación, Energía no Generada, Energía no Generada por Mantenimiento Programado, Horas Disponibles.

**4.7.2 Regresión para Predecir Índice Factor de Reserva al Sistema.**

Es casi cercano a uno el valor de R, por consiguiente podemos decir que casi están perfectamente alineados los valores de las variables en la tabla 4.17 con el índice en cuestión, también se puede verificar esa afirmación por los valores del nivel de significancia en la tabla ANOVA.

|  |  |  |
| --- | --- | --- |
| Resumen del modelo |  |  |
| R múltiple | R cuadrado | R cuadrado corregida |
| 0,98 | 0,96 | 0,96 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA | | | | | |
|  | Suma de cuadrados | Gl | Media cuadrática | F | Sig. |
| Regresión | 66,33 | 6,00 | 11,06 | 257,05 | 0,00 |
| Residual | 2,67 | 62,00 | 0,04 |  |  |
| Total | 69,00 | 68,00 |  |  |  |

Tabla 4.17

Coeficientes de Modelo para Factor de Reserva al Sistema

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coeficientes tipificados | | F | Importancia | Tolerancia | |
| Beta | Error típ. | Después de  la transformación | Antes de  a transformación |
| Gen Bruta | -0,70 | 0,35 | 4,09 | 0,40 | 0,01 | 0,04 |
| Cues | 0,34 | 0,36 | 0,93 | -0,19 | 0,00 | 0,03 |
| H Opera | -0,82 | 0,12 | 49,24 | 0,44 | 0,05 | 0,03 |
| E no generad | 0,26 | 0,06 | 17,10 | 0,12 | 0,16 | 0,05 |
| E no Gen pro | 0,19 | 0,06 | 9,48 | -0,04 | 0,16 | 0,03 |
| H Dispo | 1,37 | 0,07 | 420,82 | 0,27 | 0,14 | 0,03 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

**4.7.3 Regresión para Predecir Índice Disponibilidad**

|  |  |  |
| --- | --- | --- |
| Resumen del modelo |  |  |
| R múltiple | R cuadrado | R cuadrado corregida |
| 1,00 | 1,00 | 1,00 |

Es perfecta la alineación del índice Disponibilidad con los datos variables, es valor R es 1, la ANOVA presenta el estadístico de prueba alto y el nivel de significancia óptimo, para aprobar la dependencia de las variables. El cuadro 4.18 se muestran los coeficientes beta del modelo con mínimas desviaciones.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA | | | | | |
|  | Suma de cuadrados | gl | Media cuadrática | F | Sig. |
| Regresión | 68,98 | 6,00 | 11,50 | 34706,76 | 0,00 |
| Residual | 0,02 | 62,00 | 0,00 |  |  |
| Total | 69,00 | 68,00 |  |  |  |

Tabla 4.18

Coeficientes de Modelo para Disponibilidad

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coeficientes tipificados | | F | Importancia | Tolerancia | |
| Beta | Error típ. | Después de  la transformación | Antes de  la transformación |
| Gbru | 0,05 | 0,08 | 0,42 | 0,03 | 0,00 | 0,04 |
| Cues | 0,04 | 0,07 | 0,34 | 0,02 | 0,00 | 0,03 |
| Hopera | -0,10 | 0,03 | 10,48 | -0,06 | 0,00 | 0,03 |
| Enogenera | -0,01 | 0,01 | 4,78 | 0,01 | 0,19 | 0,05 |
| Enogenmtopro | -0,10 | 0,01 | 81,85 | 0,09 | 0,04 | 0,03 |
| Hdispo | 0,91 | 0,01 | 7667,43 | 0,91 | 0,04 | 0,03 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

**4.7.4 Regresión para Predecir Índice Factor de Planta.**

Es un modelo casi perfecto por lo que muestra el valor de R, la tabla ANOVA, con sus parámetros y la tabla 4.19, con los niveles de tolerancia y desviaciones típicas baja. Presentamos aquí los coeficientes del modelo, que para futuro podremos simular y determinar el comportamiento relativo de las variables.

|  |  |  |
| --- | --- | --- |
| Resumen del modelo |  |  |
| R múltiple | R cuadrado | R cuadrado corregida |
| 0.99 | 0.99 | 0.99 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA | | | | | |
|  | Suma de cuadrados | gl | Media cuadrática | F | Sig. |
| Regresión | 68,91 | 6,00 | 11,48 | 7622,70 | 0,00 |
| Residual | 0,09 | 62,00 | 0,00 |  |  |
| Total | 69,00 | 68,00 |  |  |  |

Tabla 4.19

Coeficientes de Modelo para Disponibilidad

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coeficientes tipificados | | F | Importancia | Tolerancia | |
| Beta | Error típ. | Después de  la transformación | Antes de  la transformación |
| Gbru | 0,61 | 0,12 | 24,60 | 0,59 | 0,00 | 0,04 |
| cues | -0,50 | 0,12 | 17,87 | -0,48 | 0,00 | 0,03 |
| Hopera | 0,79 | 0,03 | 834,84 | 0,79 | 0,03 | 0,03 |
| enogenera | -0,11 | 0,02 | 47,02 | 0,09 | 0,09 | 0,05 |
| enogenmtopro | -0,21 | 0,03 | 41,89 | 0,14 | 0,02 | 0,03 |
| hdispo | -0,20 | 0,03 | 39,76 | -0,13 | 0,02 | 0,03 |

Fuente: Informe Ejecutivo de Proceso de Producción

Elaboración: Katty Helen Delgado

Para los demás índices no presentan dependencia con las variables que hemos tomado para la regresión, es claro que repitiendo el proceso para varias combinaciones se puede encontrar el modelo adecuado.