**Completado y extendido de series - Imputación**

El procedimiento de completado se realiza a partir de la generación de datos sintéticos utilizando diferentes métodos estadísticos y su propósito general, es la conformación de series homogéneas y continuas para las diferentes variables en estudio.

Para el desarrollo de esta actividad, utilizaremos los métodos mostrados en pantalla, correspondientes a llenado con el valor de la media, mediana, última observación anterior, primera observación siguiente, interpolación lineal, media móvil, vecino natural, y multivariante con ecuación de encadenamiento. En la tabla, podrá observar la descripción general de cada método y su alcance.

*Tenga en cuenta que, en este video se realiza una explicación general del procedimiento a realizar, y que el procedimiento detallado para completar las actividades requeridas, se encuentra en la guía de clase disponible en el enlace de la descripción.*

**Objetivos**

* Obtener estadísticas generales de cada estación para cada parámetro hidroclimatológico.
* Identificar valores atípicos a través de diferentes métodos estadísticos.
* Representar gráficamente las series de cada parámetro visualizando también los datos atípicos identificados.
* Excluir valores atípicos de la matriz de estaciones para cada parámetro.
* Reemplazar valores atípicos por valores límite definidos a partir de un rango de confianza.
* Imputar valores atípicos con el valor de la media de cada estación.

**Procedimiento general**

El diagrama mostrado en pantalla, contiene el procedimiento general para la exclusión y aplicación de cada método de completado, luego de la imputación, obtendrá diferentes tablas de resultados y gráficas descriptivas de análisis.

Para iniciar, descargue el script Impute punto py, disponible en la carpeta punto .src de este repositorio, y guárdelo en la carpeta local punto src de su equipo.

**Funcionalidades del script**

* Completado y extendido de series por 8 métodos estadísticos.
* Permite definir la tabla dinámica del parámetro hidro-climatológico a evaluar.
* El usuario puede excluir estaciones del análisis o definir las estaciones a utilizar.
* Análisis masivo de estaciones por parámetro hidro climatológico con estadísticos, parámetros de evaluación y gráficas generales y detalladas por estación, con representación de datos completados y/o extendidos.
* Generación de reportes detallados Mark down y reporte complementario gráfico por estación, para cada parámetro hidro climatológico evaluado.
* Para cada método y cada parámetro hidroclimatológico analizado, crea tablas en formato de texto separado por comas punto c s v.

En pantalla, se muestran las instrucciones iniciales del script, en la guía de clase encontrará su contenido completo.

En la carpeta local punto datasets de su repositorio local, cree una carpeta con el nombre IDEAM Impute, y verifique que en la carpeta IDEAM \_ EDA o IDEAM \_ Outlier, del mismo directorio datasets, se encuentren las tablas dinámicas de cada parámetro hidro climatológico, las cuales fueron obtenidas en actividades anteriores. Dependiendo del tipo de parámetro y de los análisis previamente realizados en actividades anteriores, para el desarrollo de esta actividad de completado de datos, podrá utilizar las series de datos iniciales, o la series con identificación y ajuste de valores atípicos.

Desde el editor de texto, abra el script Impute punto py, y defina las variables, pivot table name, que corresponde a la tabla dinámica a procesar, min\_value, correspondiente al valor mínimo a imputar en el método 8 multivariante, y n neighbors, para definir el número de vecinos naturales a utilizar en el Método 7 de vecino natural, y en el método 8 multivariante. Tenga en cuenta que los resultados del Método 8, pueden variar de ejecución a ejecución del script, debido al tipo de imputador utilizado.

Desde la consola de comandos de Windows, ingrese a la carpeta IDEAM \_ Impute. Luego, ejecute la instrucción mostrada en pantalla, que realizará todo el proceso de análisis.

Durante la ejecución, podrá observar que en la consola se presenta el detalle de los procesos ejecutados y la previsualización de las diferentes tablas y gráficos.

Luego de la ejecución, podrá observar que en la carpeta local IDEAM \_ Impute, se han generado diferentes archivos de resultados por cada método, para la tabla de datos de precipitación mensual total.

De igual manera, en la carpeta local Graph, se ha generado para cada estación, 8 gráficas de comparación de las series iniciales y las imputadas. Para las 130 estaciones de precipitación del ejemplo, se han generado 1040 gráficas.

Durante el proceso de ejecución del script, se genera automáticamente un reporte científico integrado de resultados en formato Mark down, que contiene los resultados mostrados en pantalla, y un reporte complementario detallado, que incluye las gráficas generadas de valores completados para cada estación.

Al inicio del reporte Mark down, encontrará una descripción general del análisis y las librerías utilizadas, también información relacionada con los parámetros definidos.

A continuación, se presenta un resumen de la cabecera y cola del data frame.

Luego, los tipos de datos, el conteo de valores nulos y una gráfica mostrando las todas las series de las estaciones analizadas.

Para todas las estaciones, se presenta una gráfica con los valores faltantes en la serie, y las posiciones donde se encuentra el menor y mayor número de datos para cada instante de tiempo.

A continuación, podrá encontrar los estadísticos característicos de los datos de entrada de cada estación.

Para cada método, podrá encontrar una descripción general y una gráfica donde se muestra la serie original, la serie imputada y el total de valores faltantes e imputados. Para el método 1, correspondiente a imputación a partir de la media de cada estación, se completan todos los valores faltantes.

Luego de realizada la imputación por el método 1, se obtienen los estadísticos característicos de las series de cada estación. Estos valores podrán ser comparados con los estadísticos de las series originales, y con los estadísticos de los demás métodos.

En el método 2, correspondiente a imputación utilizando el valor de la mediana, también se completan todos los datos faltantes.

En el método 3, correspondiente al llenado a partir de la última observación anterior, se han imputado 8947 datos, de los 10157 faltantes.

En el método 4, correspondiente al llenado a partir de la primera observación siguiente, se han imputado 3143 datos, de los 10157 faltantes.

En el método 5, correspondiente al llenado a partir de la interpolación lineal a partir de las dos observaciones más próximas, se han imputado 8947 datos, de los 10157 faltantes.

En el método 6, correspondiente al llenado a partir de la media móvil, se han imputado 8947 datos, de los 10157 faltantes.

En el método 7, correspondiente al llenado por vecino natural, se han imputado todos los datos faltantes.

Al igual que en el método 7, en el método 8, correspondiente al llenado multivariante con ecuación de encadenamiento, se han imputado todos los datos faltantes.

Para conocer los estadísticos característicos detallados obtenidos en cada método y para cada estación, consulte la guía de clase y el reporte detallado.

El reporte complementario de resultados, presenta para cada estación, 8 gráficas de análisis donde se pueden visualizar los valores existentes en la serie, y los valores completados. Las dos primeras gráficas contienen los valores imputados por media y mediana.

Las dos siguientes gráficas, las imputaciones con última y siguiente observación encontrada.

Luego, dos gráficas con las imputaciones por interpolación lineal y media móvil.

Y por último, las gráficas de imputación por vecinos naturales y multivariante.

En la guía de clase, se presenta una tabla resumen con el conteo de las imputaciones realizadas, y para las variables evaluadas. Las columnas M1 a M8, contienen el número de datos imputados por cada método utilizado.

Al revisar los estadísticos característicos, por ejemplo, de la estación 15 0 1 50 20, correspondiente a datos de precipitación total mensual en milímetros, podrá observar los siguientes valores de media y desviación estándar presentados en pantalla.

Al revisar los estadísticos característicos, por ejemplo, de la estación 15 0 6 70 80, correspondiente a datos de caudal medio mensual en metros cúbicos por segundo, podrá observar los siguientes valores de media y desviación presentados en pantalla.

Al revisar los estadísticos característicos, por ejemplo, de la estación 15 0 1 50 20, correspondiente a datos de temperatura mínima diaria en grados centígrados, podrá observar los siguientes valores de media y desviación presentados en pantalla.

Al revisar los estadísticos característicos, por ejemplo, de la estación 15 0 1 50 20, correspondiente a datos de temperatura máxima diaria en grados centígrados, podrá observar los siguientes valores de media y desviación presentados en pantalla.

La conclusión general es que existen diferentes metodologías estadísticas para el completado y extendido de series, su aplicación en hidrología depende del tipo de parámetro hidro-climatológico a estudiar, del número de datos faltantes, del número de estaciones simultáneas evaluadas, y de la ventana de tiempo definida para los análisis. Si bien existen metodologías geoestadísticas, en las que se evalúan las relaciones espaciales, (basadas en distancia y o proximidad, bandas de elevación, y correlación con otros parámetros entre las estaciones utilizadas, con los métodos estadísticos vecino natural, y multivariante, se pueden obtener datos sintéticos, que mantienen la tendencia general de la zona estudiada a partir de las estaciones definidas.

Para el desarrollo de las actividades posteriores de este curso, usaremos las series de datos completadas y extendidas por el Método 8 Multivariante, con ecuación de encadenamiento, debido a que permite mantener la tendencia general de los datos zonales, y valores similares de media y desviación estándar.

**Actividades complementarias**

*En la guía de clase, se encuentran listadas las actividades adicionales que los estudiantes deben desarrollar y documentar para complementar los conocimientos y alcances definidos en este curso. También encontrarás información y referencias adicionales, te recomiendo realizar estas lecturas, que te ayudarán a fortalecer y complementar tus conocimientos.*

*Para terminar el completado y extendido de series - Imputación, consulta la guía de clase detallada de esta actividad. Si necesitas ayuda, da clic en el enlace Ayuda o Colabora, que se encuentra en el enlace adjunto de la descripción.*