METODO MULTIVARIADO

Jose Manuel Sepulveda Rueda

Taller 6

Edier Aristizábal

Universidad Nacional de Colombia

Facultad de Minas

Cartografía geotécnica

Noviembre 2023

Taller 6

Descripción de la cuenca

La cuenca de la Quebrada Quebradona en el sector occidental del municipio de Ituango, ubicado en el departamento de Antioquia, Colombia, desempeña un papel esencial dentro de la extensa red fluvial del Río Ituango al fungir como uno de sus afluentes.

En su cuenca alta y media, se caracteriza por extensas áreas de bosque que contribuyen significativamente a su biodiversidad. Destacan los siguientes afluentes, como la Quebrada Las Mellizas, la Quebrada Santa Lucía, y la Quebrada Quindío, junto con otros afluentes de menor tamaño, que contribuyen al caudal y la singularidad de esta región.

Además de su relevancia ecológica, la cuenca de la Quebrada Quebradona desempeña un papel crucial en la sustentabilidad de las comunidades locales, que dependen de sus recursos hídricos y disfrutan de los beneficios que brinda a la agricultura y la vida silvestre. Por consiguiente, es imperativo asegurar su conservación y protección, no solo como un valioso ecosistema, sino también como un activo fundamental para el bienestar de aquellos que residen en su entorno.

CARACTERISTICA	DETALLE
ÁREA	76.152497 km^2
PERÍMETRO	46.431991 km
ALTITUD MÁXIMA	3117 msnm
ALTITUD MÍNIMA	1269 msnm
ALTURA PROMEDIO	2252 msnm
LONG AXIAL LARGO	11.8 km
LONG AXIAL ANCHO	11.64 km
PENDIENTE PROMEDIO	30°
LONGITUD DEL CAUCE PRINCIPAL	8.66 km

Tabla 1. Características generales de la cuenca

Métodos Multivariados

Los métodos estadísticos multivariados analizan la relación conjunta entre la variable dependiente (ocurrencia de movimientos en masa) y todas las variables independientes (predictoras) simultáneamente. Uno de estos métodos es la regresión logística, ampliamente utilizado a nivel mundial para evaluar la susceptibilidad a movimientos en masa.

Regresión logística

La regresión logística es un método estadístico multivariado utilizado para predecir la probabilidad de que ocurra un evento binario, es decir, un evento que puede tener dos resultados posibles, como sí (Ocurrencia de MenM) /no (No ocurrencia de MenM), etc. Es especialmente útil cuando la variable que queremos predecir es categórica.

Funciona de la siguiente manera: la regresión logística utiliza una función logística para modelar la relación entre una o más variables independientes (predictoras) y la probabilidad de que ocurra el evento de MenM. Esta función logística transforma la suma ponderada de las variables independientes utilizando una curva en forma de "S", que va de 0 a 1. La salida de esta función representa la probabilidad estimada del evento.

En términos más simples, la regresión logística busca encontrar la mejor combinación de coeficientes para las variables independientes de manera que la curva logística se ajuste de la mejor manera posible a los datos observados. Una vez que se ha ajustado el modelo, podemos usarlo para predecir la probabilidad de que ocurran MenM para nuevos conjuntos de datos.

Python

Nota: para visualizar el código de Python de cada una de las figuras, dirigirse al documento anexo a este en la carpeta.

Se crea un Dataframe con todas las variables y el inventario de MenM

	inventario	pendiente	Curvatura	flujo_acum	aspecto	geologia
0	0.0	14.068556	1.28	0.0	118.610458	510.0
1	0.0	18.160751	1.28	0.0	127.568596	510.0
2	0.0	17.216877	-0.00	0.0	121.809387	510.0
3	0.0	19.485619	0.64	1.0	137.290604	510.0
4	0.0	20.243055	0.64	1.0	139.398712	510.0

Figura 1. Creación del Dataframe

Se procede a normalizar los valores

	inventario	pendiente	aspecto	geologia
0	0.0	-1.582421	-0.051931	14.0
1	0.0	-1.460616	-0.062760	14.0
2	0.0	-1.442916	-0.110142	14.0
3	0.0	-1.383993	-0.021063	14.0
4	0.0	-1.302091	0.062773	14.0

Figura 2. Normalización de los valores

Se usa la librería statsmodels para mirar un resumen de los resultados y métricas del modelo

```
                           Logit Regression Results
==============================================================================
Dep. Variable:             inventario   No. Observations:               484040
Model:                          Logit   Df Residuals:                   484031
Method:                           MLE   Df Model:                            8
Date:                Sun, 19 Nov 2023   Pseudo R-squ.:                 0.02178
Time:                        18:18:31   Log-Likelihood:                -73572.
converged:                      False   LL-Null:                       -75210.
Covariance Type:            nonrobust   LLR p-value:                     0.000
=========================================================================================
                            coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept               -15.4983    114.670     -0.135      0.892    -240.248     209.252
C(geologia)[T.1.0]       11.1913    114.670      0.098      0.922    -213.559     235.941
C(geologia)[T.510.0]     12.2125    114.670      0.107      0.915    -212.537     236.962
C(geologia)[T.765.0]     11.5078    114.670      0.100      0.920    -213.242     236.258
C(geologia)[T.1020.0]    -7.7059   1462.234     -0.005      0.996   -2873.631    2858.220
pendiente                 0.2337      0.008     30.221      0.000       0.219       0.249
Curvatura                -0.0016      0.004     -0.454      0.650      -0.009       0.005
flujo_acum                0.0080      0.001      6.773      0.000       0.006       0.010
aspecto                  -0.2161      0.008    -26.561      0.000      -0.232      -0.200
=========================================================================================
```

Figura 3. Resumen del modelo

Parametrización del modelo

	pendiente	Curvatura	flujo_acum	aspecto	geologia
0	-1.601045	1.28	0.0	-0.637368	510.0
1	-1.198484	1.28	0.0	-0.545700	510.0
2	-1.291335	-0.00	0.0	-0.604634	510.0
3	-1.068153	0.64	1.0	-0.446214	510.0
4	-0.993641	0.64	1.0	-0.424642	510.0

Figura 4. Parametrización del modelo

Transformación de las variables categóricas a variables dummy/indicadoras

	pendiente	Curvatura	flujo_acum	aspecto	geo_0.0	geo_1.0	geo_510.0	geo_765.0	geo_1020.0
0	-1.601045	1.28	0.0	-0.637368	0	0	1	0	0
1	-1.198484	1.28	0.0	-0.545700	0	0	1	0	0
2	-1.291335	-0.00	0.0	-0.604634	0	0	1	0	0
3	-1.068153	0.64	1.0	-0.446214	0	0	1	0	0
4	-0.993641	0.64	1.0	-0.424642	0	0	1	0	0

Figura 4. Trasformación de las variables categóricas a indicadoras

Inicialmente, se importa la función para RL, y se construye el modelo con los hiperparámetros

HIPERPARÁMETRO	FUNCIÓN
LASS_WEIGHT = BALANCED	Asignación más peso a las celdas minoritarias
SOLVER='NEWTON-CG'	Utiliza el tipo de algoritmo para resolver la RL

Tabla 2. Hiperparámetros

```
LogisticRegression
LogisticRegression(class_weight='balanced', solver='newton-cg')
```

Figura 5. Modelo con los hiperparámetros

Luego de tener el modelo construido, se le asignan los datos con la función fit, en este caso, primero la matriz con las variables predictoras, y luego la variable y. De esta forma el modelo se entrena con los datos y se pueden obtener los resultados de los valores de los coeficientes con la función coef_

```
[[ 2.2395691e-01 -4.0270612e-03  6.5371278e-03 -2.6309347e-01
  -1.9937090e+00  1.0510597e+00  2.0646276e+00  1.4127628e+00
  -4.3238163e+00]]
```

Figura 6. Valores de los coeficientes

Se obtienen los resultados del modelo antes de clasificarlo como (0,1). es decir, la probabilidad de cada celda de ser 0 o de ser 1

```
array([[0.5497858 , 0.4502142 ],
       [0.5333918 , 0.46660823],
       [0.53342545, 0.46657458],
       ...,
       [0.53011453, 0.46988547],
       [0.51694584, 0.48305413],
       [0.49200678, 0.5079932 ]], dtype=float32)
```

Figura 7. Resultados del modelo sin clasificarlos con MenM o sin MenM

Ya se tienen los resultados y la predicción para todas las celdas. Sin embargo, esto se tiene para un vector que al principio solo contiene las celdas dentro de la cuenca, es decir eliminó todas las celdas por fuera. Esto impide la construcción de cuenca a partir de este vector

Para este caso se hace un nuevo Dataframe, pero en este caso con el vector completo, sin eliminar las celdas por fuera de la cuenca, y las celdas por fuera de la cuenca, en lugar de tener un valor de NaN les daremos un valor de 0. Las funciones de sklearn no corren cuando encuentra dentro de los valores NaN.

Se reconstruye la cuenca con los valores de susceptibilidad. Para eso utilizaremos como máscara el mapa de pendiente.

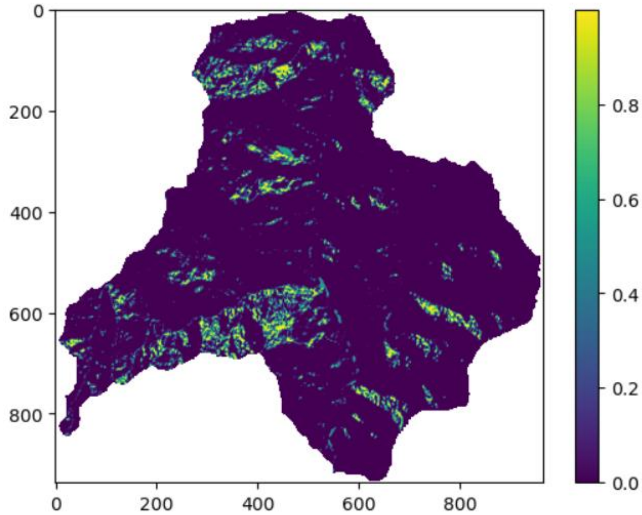

Figura 8. Cuenca con los valores de susceptibilidad