

Taller 3 – Selección de variables

Cartografía Geotécnica

Asignatura

Dubán Uribe Góez

Autor

Edier Vicente Aristizábal Giraldo

Docente



Universidad Nacional de Colombia – Sede Medellín

Facultad de Minas

2025-1

Semestre

Contents

Contents.....	2
Selección de las variables para la cuenca Río Grande	3
Análisis univariado.....	6
Análisis multivariado	8
Correlación con la variable categórica.....	10
Modelo heurístico.....	11
Clasificación de variables de pesos.....	11
Pendiente.....	11
Elevación	11
Aspecto	12
Flujo acumulado	12

Selección de las variables para la cuenca Río Grande

Para la modelación, se consideraron inicialmente diversas variables a las que se les aplicó un análisis exploratorio, con el fin de identificar su comportamiento general. Posteriormente, se llevaron a cabo análisis univariados y multivariados para profundizar en la comprensión de cada variable y sus relaciones. Las variables preliminares incluidas en el estudio fueron: pendiente, aspectos, curvatura, flujo acumulado, elevación y geología.

Se observó que la mayoría de los movimientos en masa ocurrieron dentro de una misma unidad geológica, como se evidencia en la Figure 1 Mapa geología y MenM. Por esta razón, la variable geología fue descartada, ya que no permitía una discriminación adecuada entre las celdas con y sin movimientos en masa. De igual manera, se excluyó la variable geomorfología, al considerarse redundante frente a otras variables morfométricas ya incluidas en el análisis, como pendiente, aspectos, elevación y curvatura

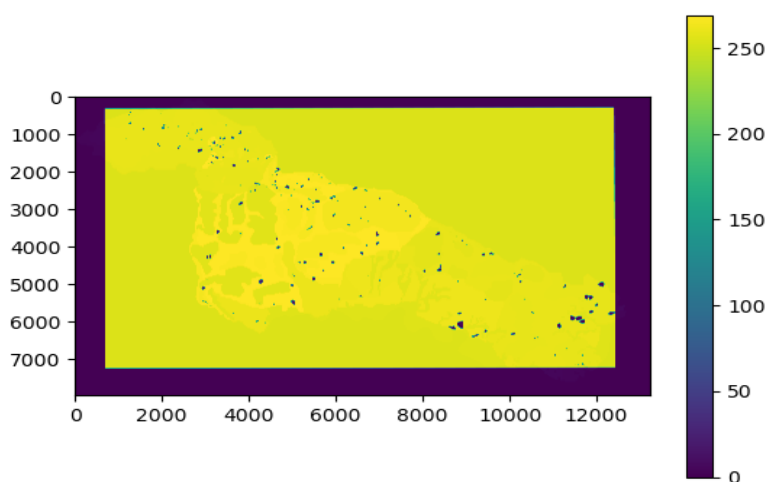


Figure 1 Mapa geología y MenM

Por lo tanto, las variables seleccionadas para los análisis y la posterior identificación de las más relevantes y representativas fueron: pendiente, aspectos, curvatura, flujo acumulado y elevación.

	count	mean	std	min	25%
inventario	155114.0	0.000000e+00	0.000000e+00	0.0	0.000000
pendiente	155114.0	1.959671e+01	1.137327e+01	0.0	9.402564
aspectos	155114.0	1.604277e+02	8.477580e+01	0.0	94.049171
curvatura	155114.0	1.270371e+02	1.274271e+00	100.0	127.000000
Flujo Acumulado	155114.0	1.248390e+06	8.787535e+07	0.0	7382.208984
elevacion	155114.0	1.021182e+02	1.126113e+02	1.0	4.000000

	50%	75%	max
inventario	0.000000	0.000000	0.000000e+00
pendiente	19.826857	28.267174	8.991143e+01
aspectos	159.363068	216.801186	3.599996e+02
curvatura	127.000000	127.000000	2.550000e+02
Flujo Acumulado	20856.406250	48347.200195	1.590894e+10
elevacion	13.000000	233.000000	2.550000e+02

Figure 2 Análisis exploratorio de las variables

El siguiente paso consiste en generar un *DataFrame*, el cual permite organizar la información de manera estructurada y facilita su manipulación. En la imagen a continuación se muestra el *DataFrame* resultante. Cabe destacar que las variables utilizadas en este análisis son de tipo continuo.

	inventario	pendiente	aspectos	curvatura	Flujo Acumulado	elevacion
0	0.0	0.0	0.0	127.0	0.000000	239.0
1	0.0	0.0	0.0	128.0	0.000000	239.0
2	0.0	0.0	0.0	127.0	0.000000	239.0
3	0.0	0.0	0.0	127.0	4602.217773	239.0
4	0.0	0.0	0.0	127.0	6903.275391	239.0

Figure 3 DataFrame

Se comienza comparando las variables continuas, que en este caso son todas las elegidas. Para esto se realiza una matriz de comparación, la cual en su diagonal contiene un histograma de densidad de Kernel, que nos muestra la distribución de los datos. Se desea observar si existe alguna correlación entre ellas, a partir de tendencias. En este caso, las celdas de la matriz diferentes a la diagonal, no muestra tendencias que den indicios de correlación entre variables

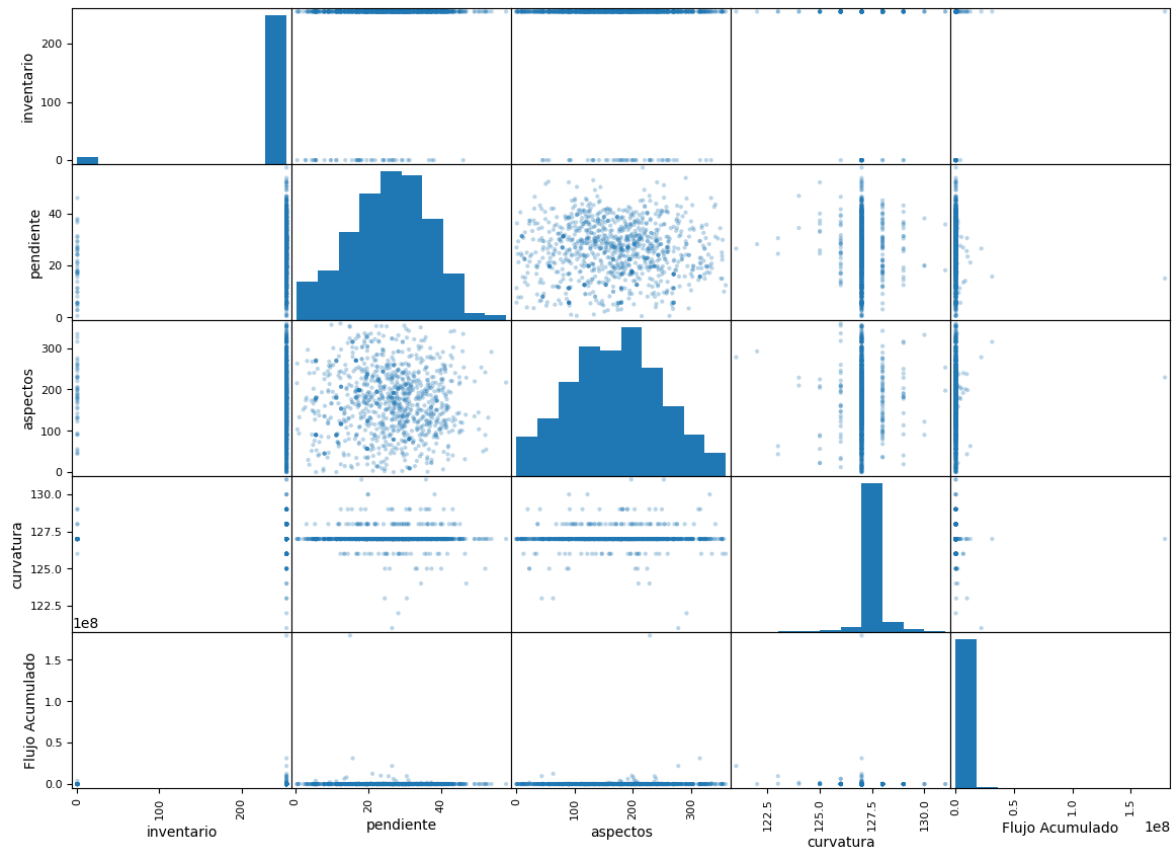


Figure 4 Matriz de correlación 1

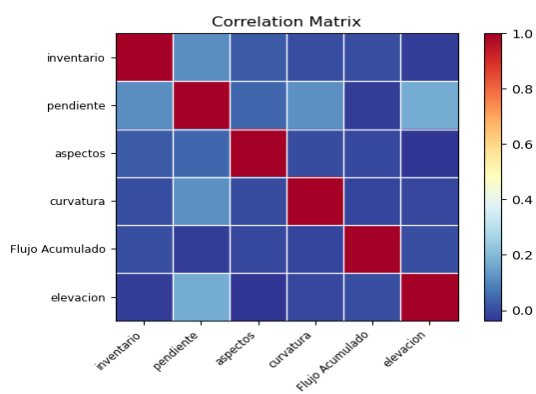


Figure 5 Matriz de correlación 2

Las matrices muestran correlaciones bajas entre la mayoría de las variables y no se observan correlaciones fuertes ni multicolinealidades significativas entre predictores, lo cual es una señal buena para ciertos modelos.

Análisis univariado

Para este caso, se lleva a cabo un análisis para cada una de las variables elegidas, para así lograr conocer si estas pueden o no mostrar la ocurrencia de los movimientos en masa.

Pendiente

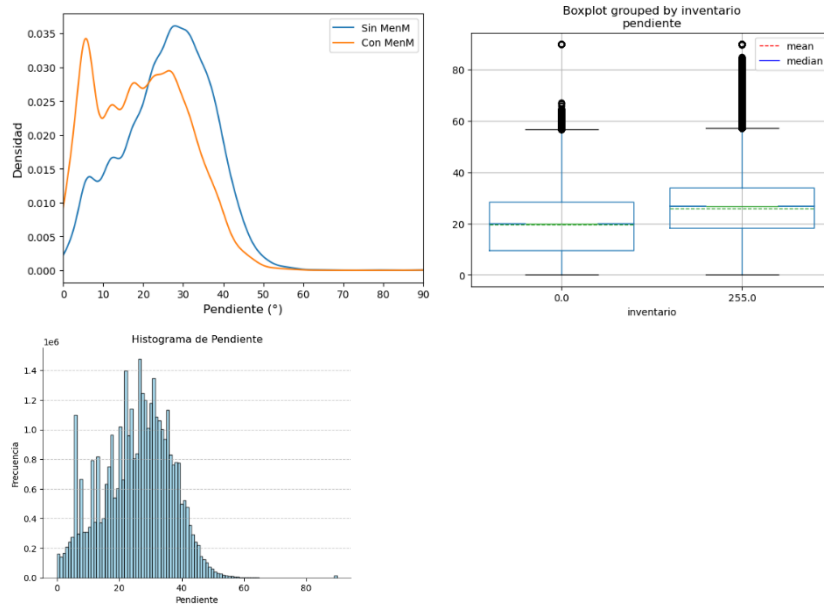


Figure 6 Análisis univariado – Pendiente

La relación entre pendiente y MenM no es estrictamente lineal: los deslizamientos no ocurren solamente en zonas muy inclinadas.

Elevación

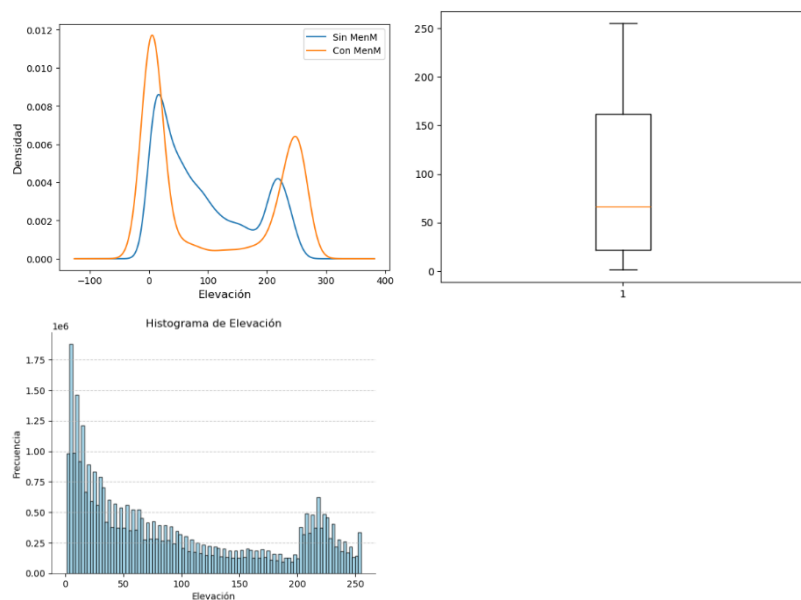


Figure 7 Análisis univariado – Elevación

La elevación muestra un patrón bimodal para los movimientos en masa, lo que sugiere que es altamente informativa para predecir este fenómeno. No es simplemente “a más altura, más riesgo”, sino que hay rangos concretos (bajo y alto) donde los eventos son más frecuentes.

Aspecto

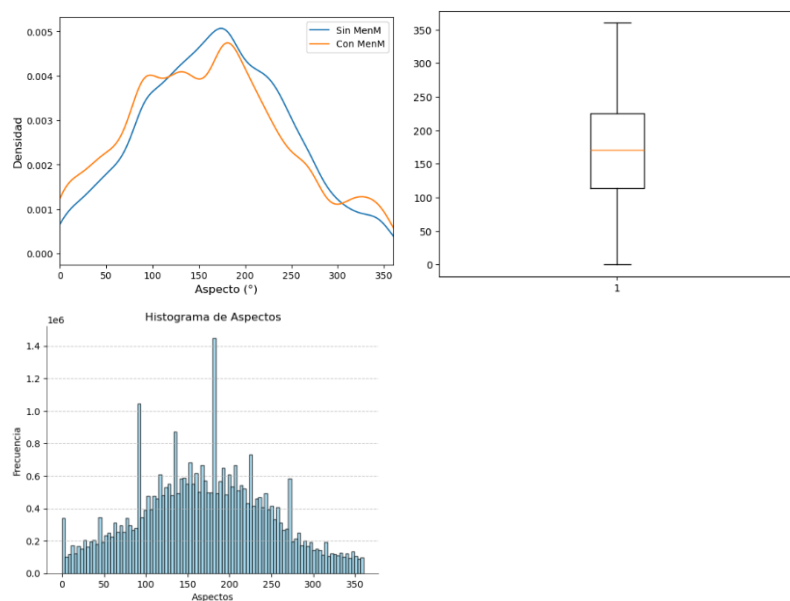


Figure 8 Análisis univariado – Aspecto

En la gráfica se observan diferencias sutiles pero consistentes entre las curvas de densidad de las celdas con MenM (movimientos en masa) y sin MenM:

Aspecto podría ser una variable relevante, especialmente si el patrón de orientación incide en la exposición solar, el viento, o la acumulación de humedad (factores que influyen en la estabilidad del terreno).

Flujo acumulado

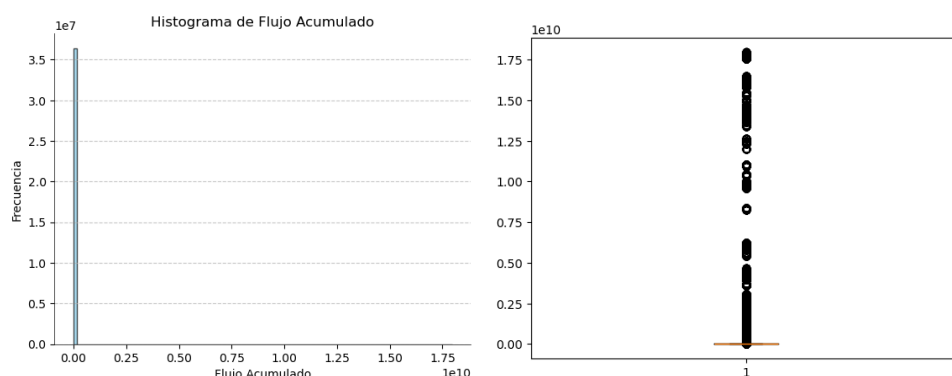


Figure 9 *Análisis univariado - Flujo acumulado*

Curvatura

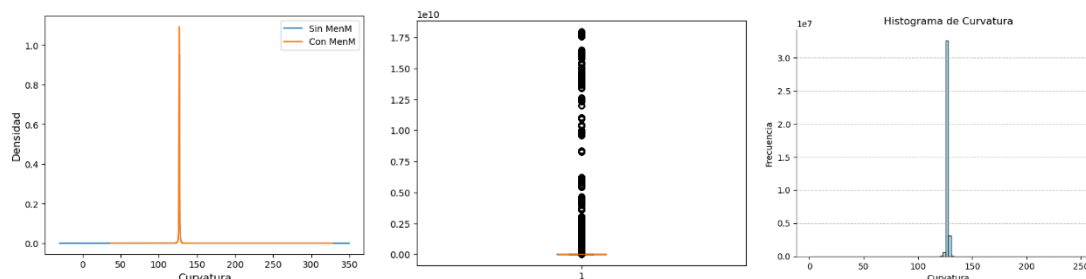


Figure 10 *Análisis univariado – Curvatura*

La distribución de curvatura es prácticamente idéntica tanto para las celdas con movimientos en masa (Con MenM) como para las que no los presentan (Sin MenM).

Curvatura no discrimina entre zonas con y sin movimientos en masa. Es decir, no aporta poder predictivo al modelo tal como está.

Análisis multivariado

En este análisis se comparan unas variables con otras con el fin de determinar si existen correlaciones entre estas variables a partir de tendencias.

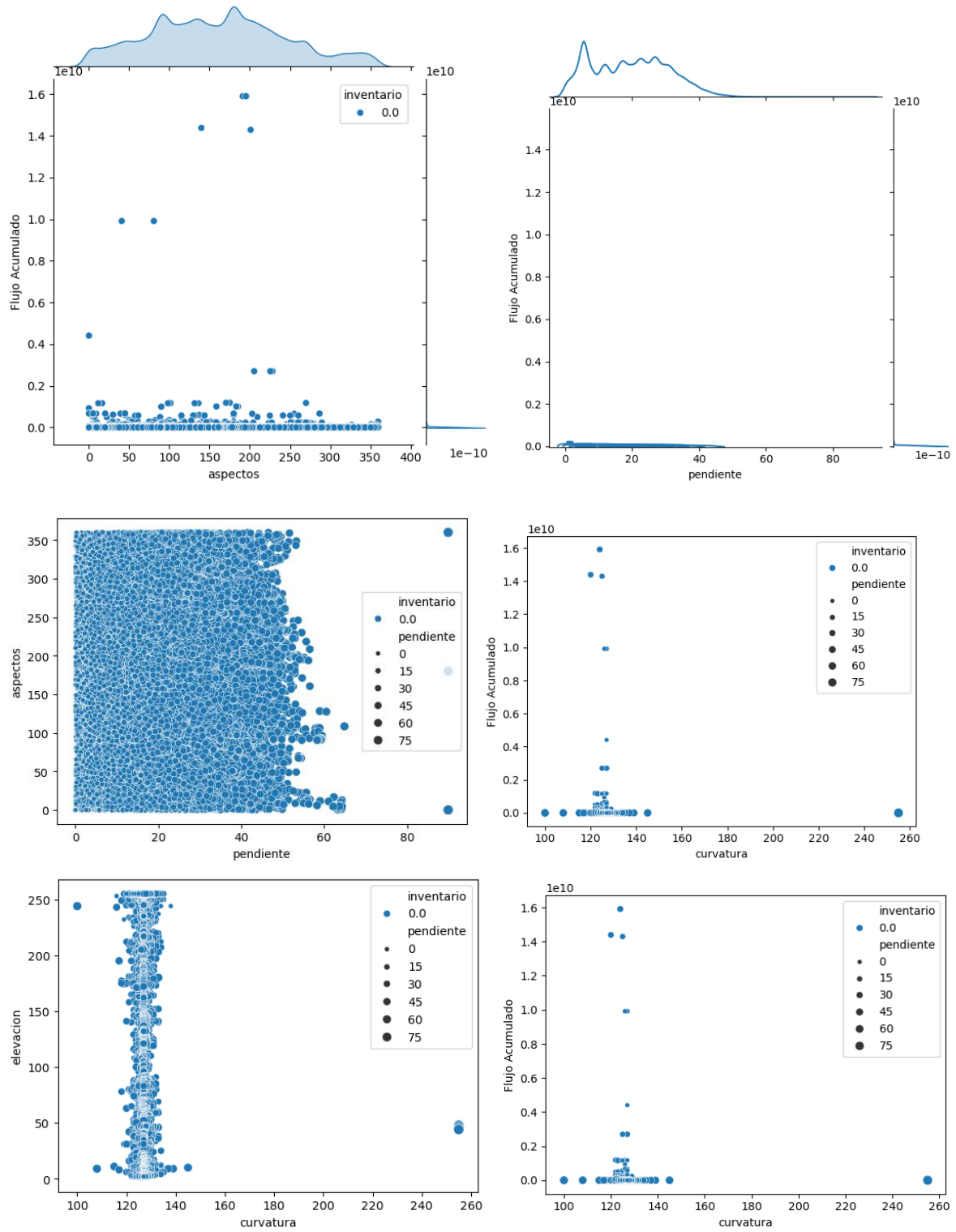


Figure 10 Análisis univariado – Curvatura

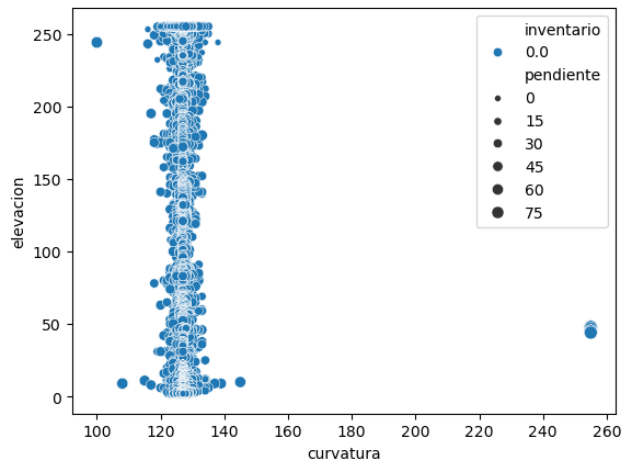


Figure 11 análisis multivariado

Correlación con la variable categórica

En este caso se realiza una correlación de las variables independientes elegidas, teniendo en cuenta el inventario de movimientos en masa.

	pendiente	aspectos	curvatura	Flujo Acumulado	elevacion
inventario					
0.0	19.624375	160.269557	127.035604	1.005468e+06	102.319911
255.0	25.886898	170.236485	127.094739	3.584399e+06	92.427134

Tabla 1 Medidas agrupadas por la variable MenM

La tabla anterior nos muestra lo siguiente

Pendiente: Mayor en zonas con inventario = 255 (25.89 vs 19.62)

Aspectos: Diferencia moderada (170° vs 160°)

Curvatura: Muy similar (127.09 vs 127.04). No parece discriminante

Flujo Acumulado: Mucho mayor en zonas con inventario (3.58M vs 1.00M)

Elevación: Menor en zonas con inventario (92.43 vs 102.32)

Después de hacer el análisis de las estadísticas, se determinó que las variables que mejor me representan la susceptibilidad en el área de estudio son La pendiente, la elevación, el flujo acumulado, y el aspecto.

Modelo heurístico

La forma general del modelo es la siguiente:

$$f(x) = b * x_1 + c * x_2 + d * x_3 + e * x_4$$

Donde x_1 = Clase de pendiente

Donde x_2 = Clase de elevación

Donde x_3 = Clase de aspecto

Donde x_4 = Clase de flujo acumulado

Los coeficientes b, c, d, e representan la importancia de cada variable, que se ajusta para que la función final esté normalizada entre 0 y 1

Clasificación de variables de pesos

Pendiente

Rango (°)	Clase	Peso x_1
0–5	baja	0.1
5–15	media	0.4
15–30	alta	0.6
>30	muy alta	0.1

Elevación

Rango (m)	Clase	Peso x_2
0–10	muy baja	0.1
10–90	ideal	0.6
90–200	moderada	0.2
200–300	secundaria	0.4
>300	alta	0.1

Aspecto

Rango (°)	Clase	Peso x_3
60–230	favorable	0.6
0–60	intermedio	0.3
230–360	menos relevante	0.1

Flujo acumulado

Rango (valor)	Clase	Peso x_4
< 1e5	bajo	0.1
1e5 – 1e6	medio	0.3
1e6 – 3e6	alto	0.5
> 3e6	muy alto	0.6

Función normalizada

Se asigna coeficientes de peso (importância global de cada Variable)

b=0.35 para pendiente

c=0.25 para elevación

d=0.2 para aspecto

e=0.2 para flujo acumulado

Asegurando que $b+c+d+e=1$

Así pues la función heurística de susceptibilidad es:

$$f(x) = 0.35 * x_1 + 0.25 * x_2 + 0.2 * x_3 + 0.2 * x_4$$

Se calcula el radio de consistencia usando la matriz de comparación

	pendiente	elevación	aspecto	flujo
pendiente	1.000	1.400	1.750	1.750
elevación	0.714	1.000	1.250	1.250
aspecto	0.571	0.800	1.000	1.000
flujo	0.571	0.800	1.000	1.000

Vector de pesos normalizado

Variable	Peso normalizado
Pendiente	0.35
Elevación	0.25
Aspecto	0.20
Flujo	0.20

$\lambda_{\max}=4$

$n= 0$

$CI= 0$

La matriz es consistente ya que el CI es menor a 0,1

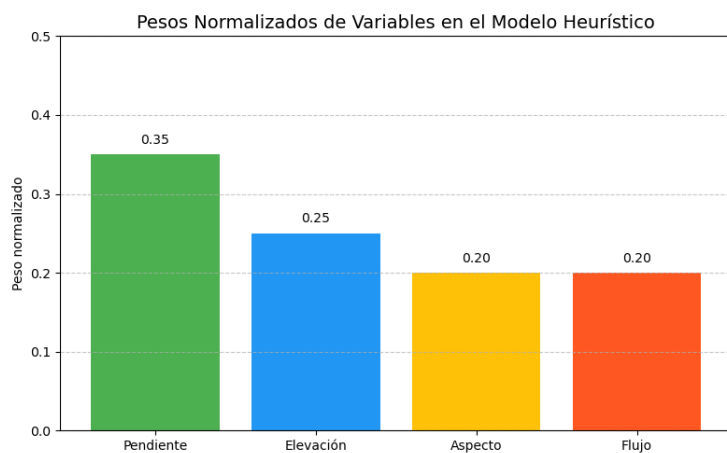


Figure 11 pesos normalizados de variables modelo heurístico