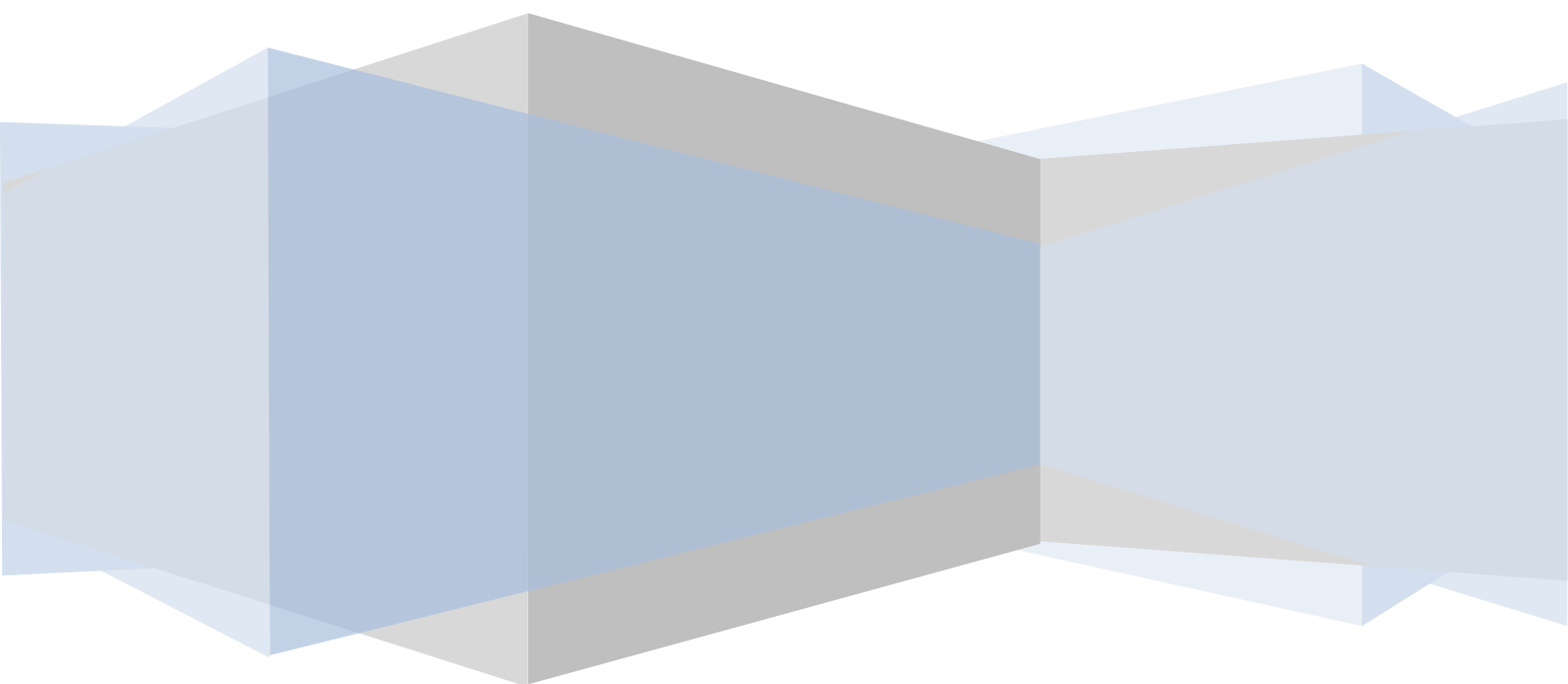


Estadística Espacial

Análisis estructural

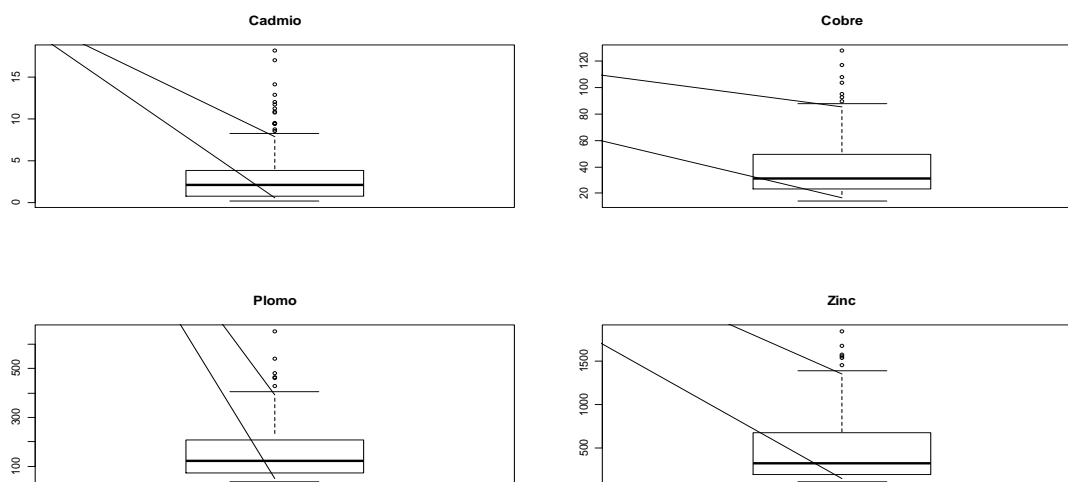
Fichero: meuse Variable: zinc

María Leyenda Rodríguez



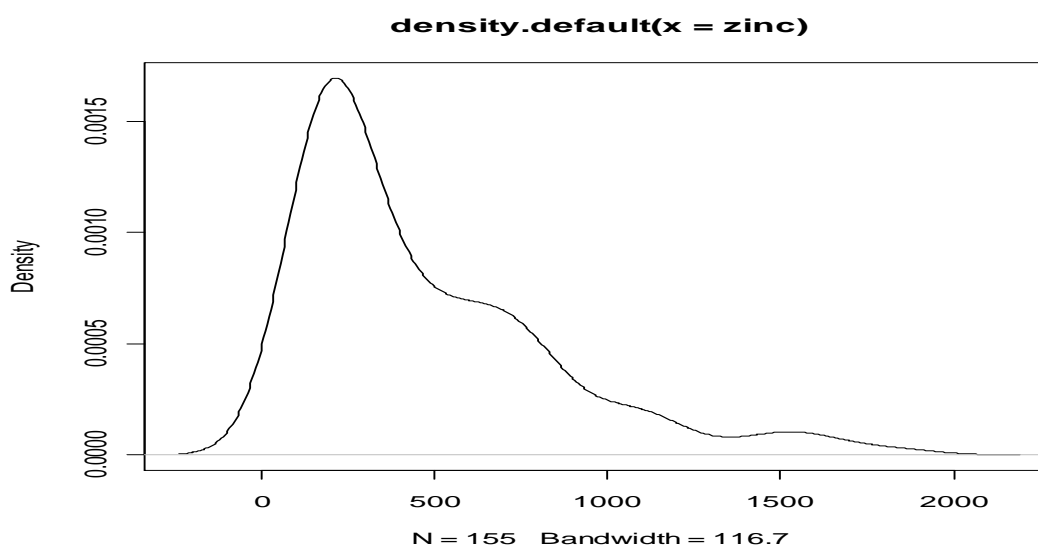
1.- Análisis de la variable zinc del fichero meuse.

En primer lugar, se aplica un diagrama de cajas a cada una de las variables del paquete meuse: cadmio, cobre, plomo y zinc. Con el fin de observar el rango de valores en el que toma valores cada una de ellas.

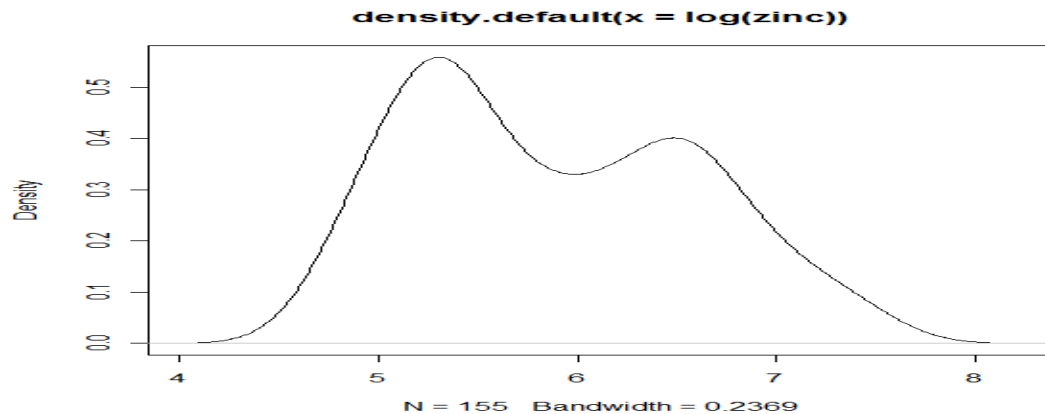


Además también aplicamos el test de Shapiro-Wilks para contrastar la hipótesis de normalidad a cada una de las variables; de dónde obtenemos que ninguna de las variables sigue distribución normal. Este informe, se va a centrar en la variable zinc. Esta variable tiene media 469.7161 y desviación típica 367.0738. Pues esta variable toma valores altos, entre 113 y 1839, y con gran dispersión.

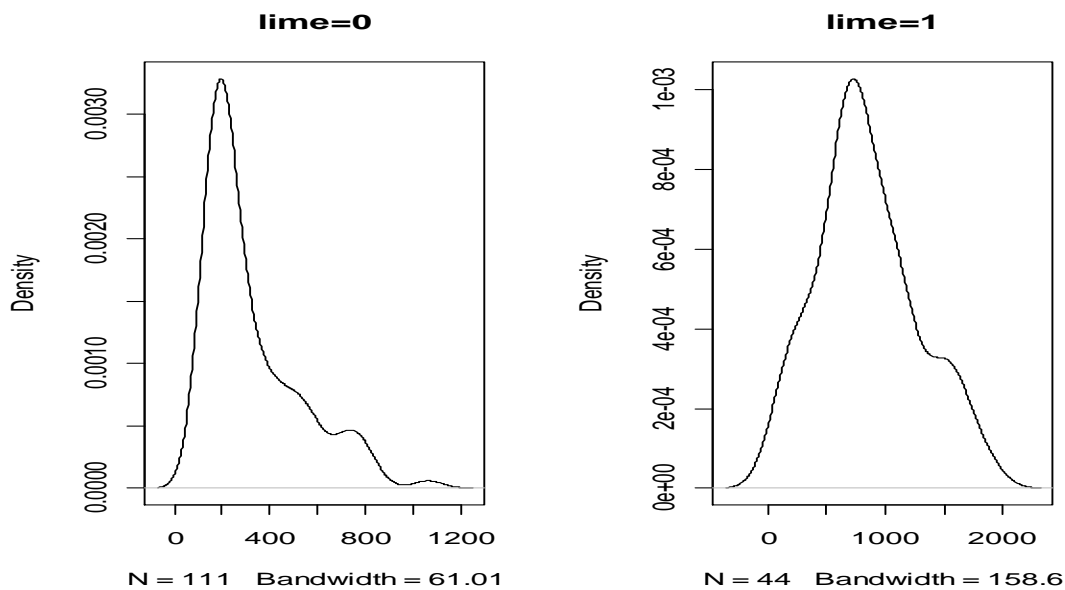
Como hemos analizado anteriormente la variable zinc no es normal, de hecho si representamos su función de densidad no es simétrica.



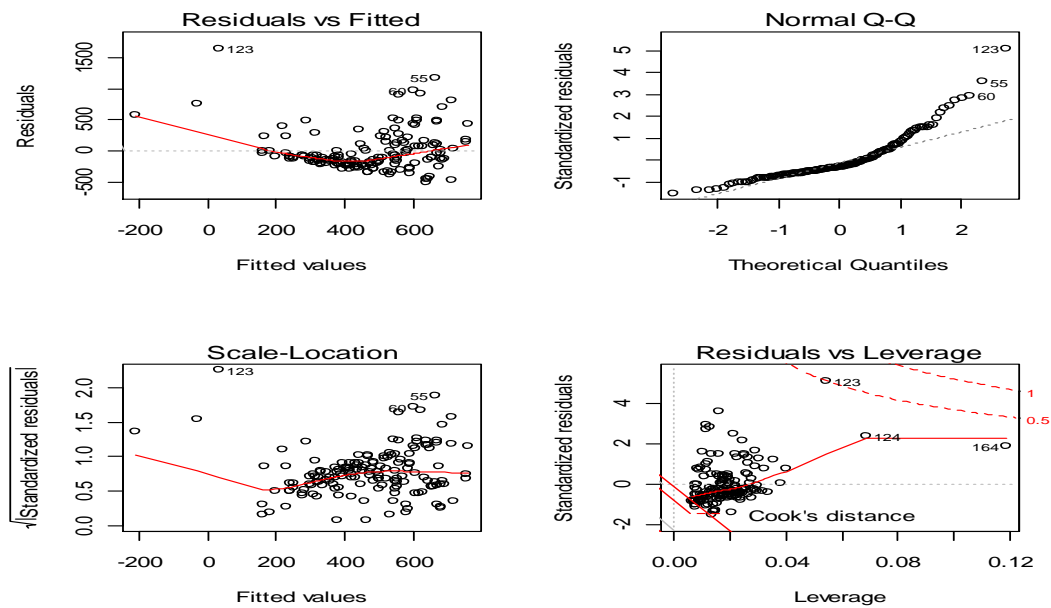
Intentamos corregir la no normalidad aplicándole la transformación logaritmo a nuestra variable; zinc. Si representamos la función de densidad de la variable transformada observamos que el test de Shapiro-Wilks también rechaza la hipótesis de normalidad. La variable transformada a pesar de tener dos modas presenta un comportamiento más “cercano” a la distribución normal que la variable sin transformar.



En cambio si estudiamos la variable en cada uno de los grupos dónde está definida; obtenemos que en uno de ellos es normal (lime=1) y en otro no (lime=0).

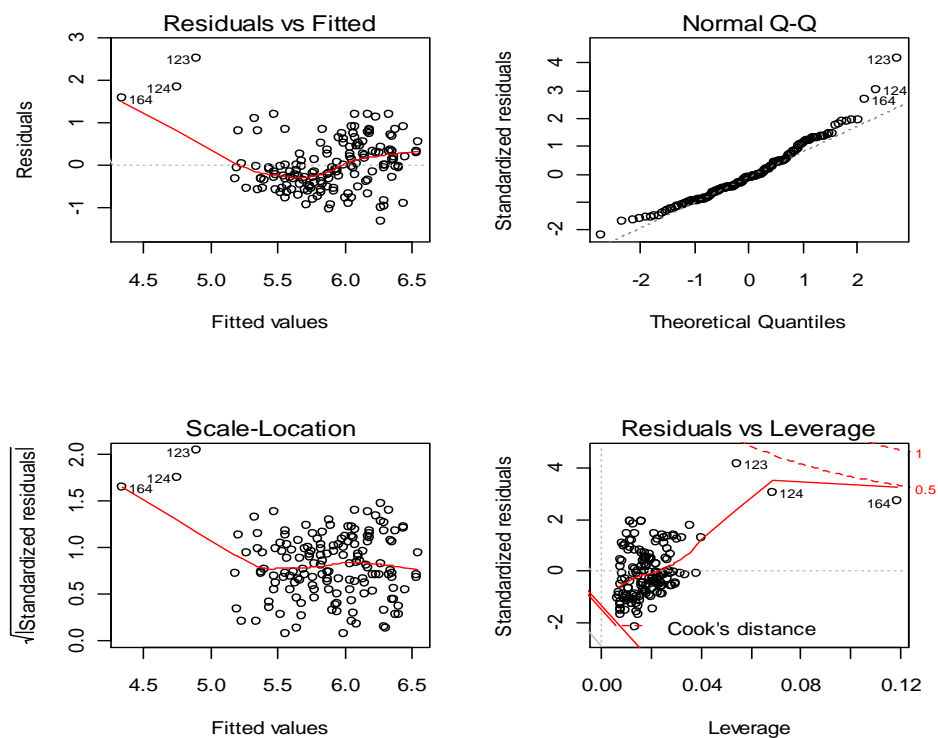


Comenzamos estudiando el modelo de regresión que tiene a la variable zinc como variable respuesta y como explicativas las coordenadas (x,y) con el fin de comprobar si influyen las localizaciones en la variable. Para ello comprobamos que los coeficientes correspondientes a las localizaciones sean significativos.



Si analizamos los residuos de este modelo de regresión observamos que estos no son normales y están muy concentrados. Además, tenemos observaciones muy influyentes en el modelo (50,65, 123).

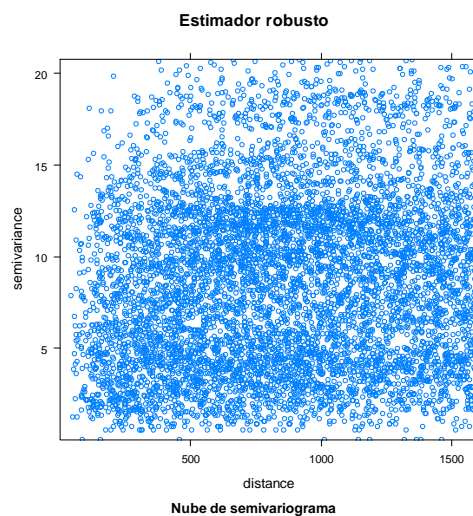
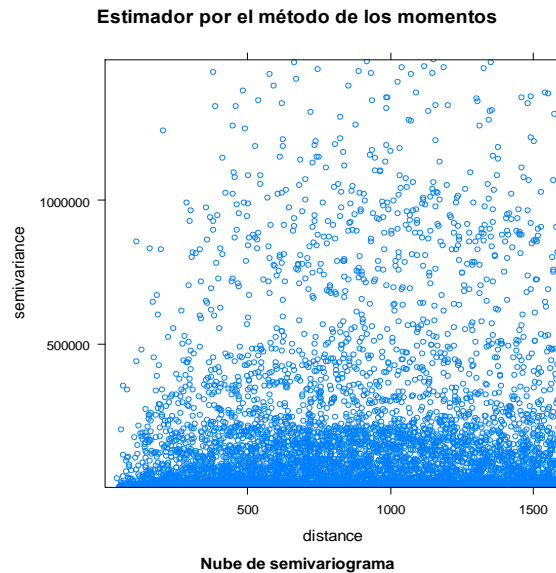
Si estudiamos el modelo de regresión que tiene a la variable logaritmo de zinc como variable respuesta y como explicativas las coordenadas (x,y). De dónde obtenemos que las localizaciones influyen en la variable.



Además, obtenemos que los residuos no están tan concentrados, hay más variabilidad pero siguen sin ser normales aunque como tienen un comportamiento más similar al de la distribución normal que los del modelo anterior. En este caso, también tenemos valores muy influyentes (123, 124, 164).

Finalmente, como hemos comprobado que la variable transformada no nos soluciona completamente todos nuestros problemas, realizaremos el análisis con la variable sin transformar.

La representación gráfica de la nube de semivariograma nos da una idea de la variabilidad de los datos. Representamos la nube de semivariograma para el estimador por el método de los momentos y la nube de semivariograma para el estimador robusto.

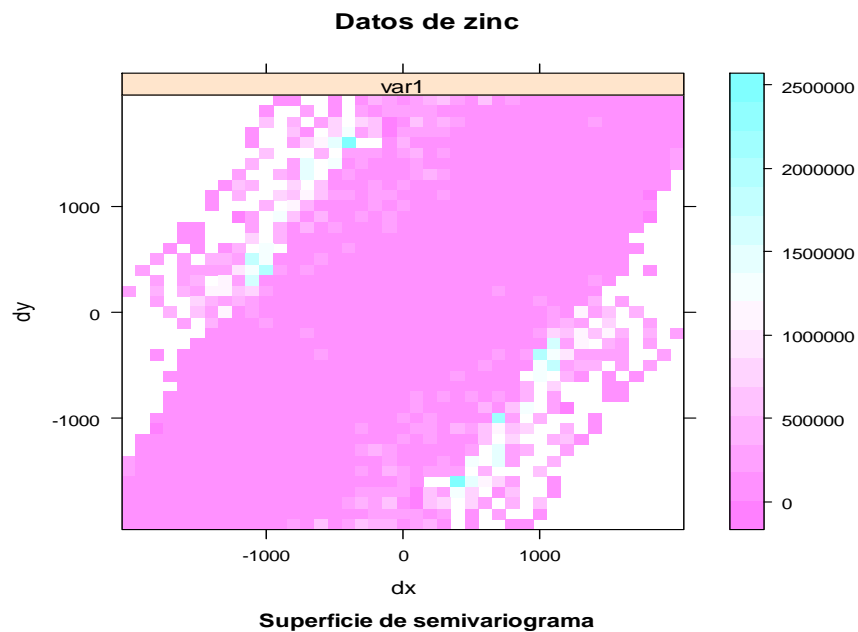
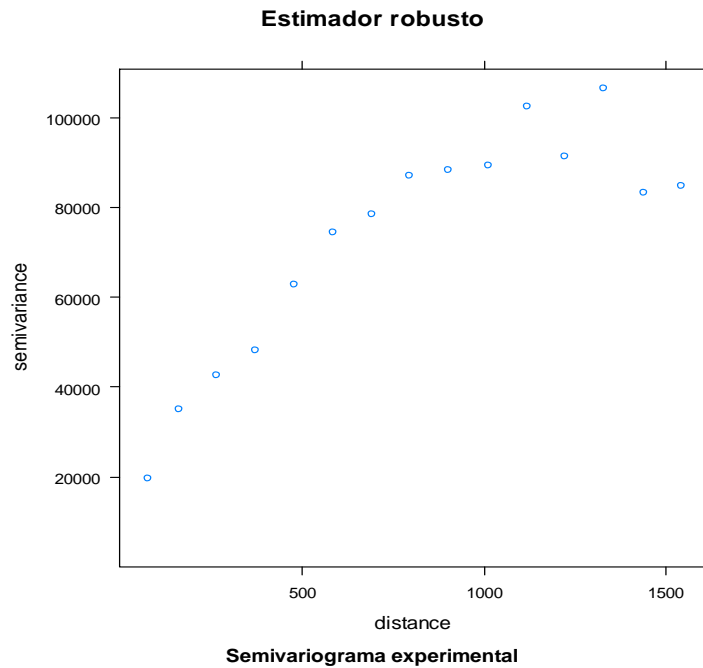


Se observa que con este estimador tenemos una menor variabilidad del semivariograma puesto que solo toma valores menores que 30.

En estas dos nubes de semivariograma se observa una tendencia creciente.

2.- Cálculo de semivariogramas experimentales para determinar el tipo de dependencia espacial.

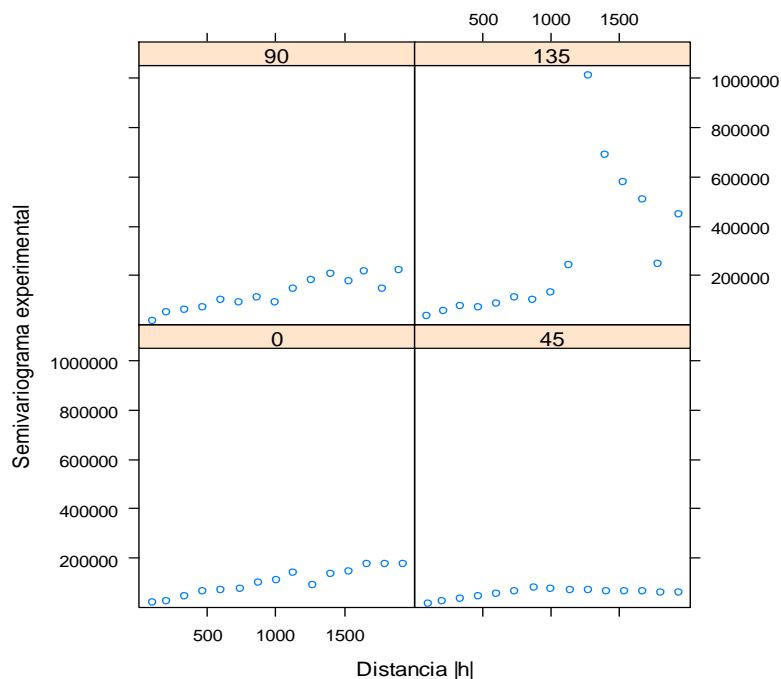
En este caso, como la variable zinc no presenta un comportamiento parecido a la distribución normal, es más aconsejable la utilización del estimador robusto de la función de semivariograma.



Esta representación de la superficie del variograma muestra una dependencia espacial que depende de $|h|$ y de su dirección. Además, en este gráfico puede apreciarse fácilmente La dependencia de mayor alcance está en la dirección 45 grados y la de menor alcance en 135

3.- Comportamiento anisotrópico.

A continuación, se representan los semivariogramas experimentales en las direcciones de 0° , 45° , 90° , 135° .



Obtenemos los cuantiles del semivariograma experimental en estas direcciones. Con el fin de que nos sirvan de referencia a la hora de fijar las partes del semivariograma:

0%	25%	50%	75%	100%
----	-----	-----	-----	------

14780.69	62235.56	87696.14	152553.79	1009409.81
----------	----------	----------	-----------	------------

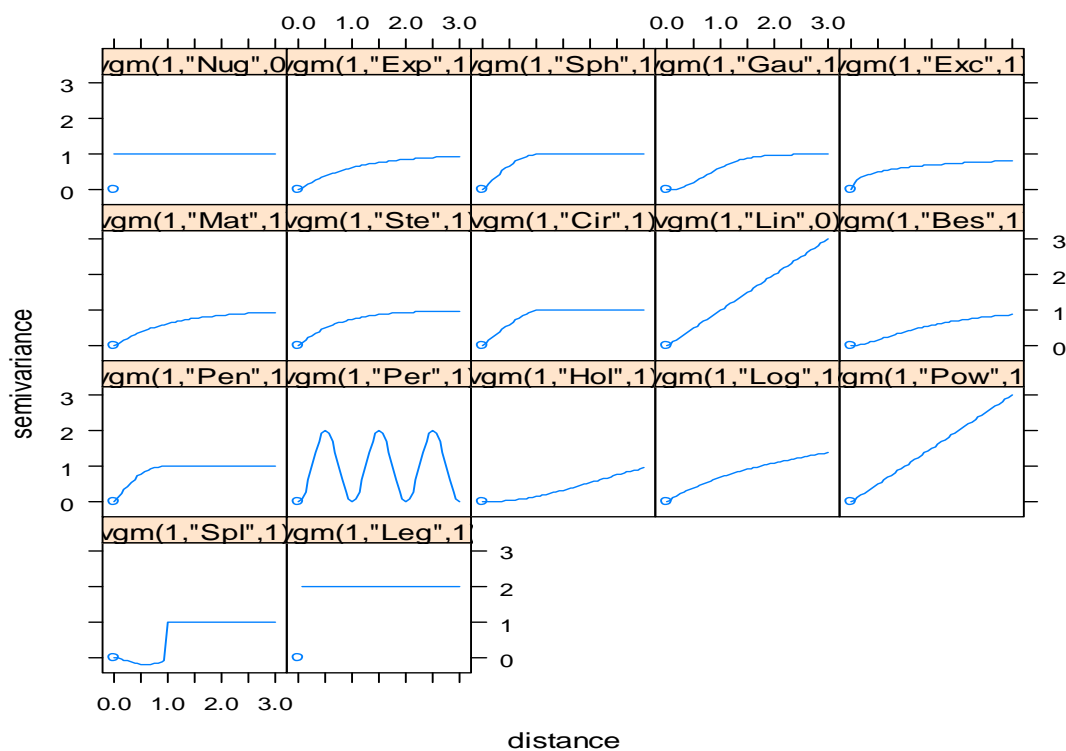
Por tanto, las partes del semivariograma son:

- Pepita: 14000
 - El valor de semivariograma en el cero
- Meseta: 90000
 - La cota superior del semivariograma.

- Rango:1200
 - Es el rango en la dirección de dependencia de mayor alcance.
 - El rango corresponde a la distancia a partir de la cual dos observaciones son independientes. El rango se interpreta como zona de influencia. Por ello cuanto más pequeño sea el rango más cerca se está del modelo de independencia espacial.
- Ratio de anisotropía: 0.4

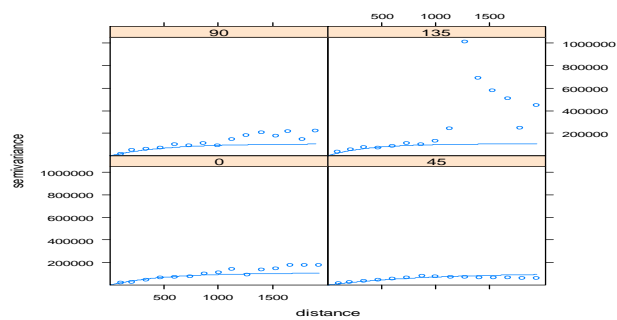
4.-Selección de modelos de semivariogramas teóricos

A continuación, están representados varios modelos de semivariogramas teóricos.

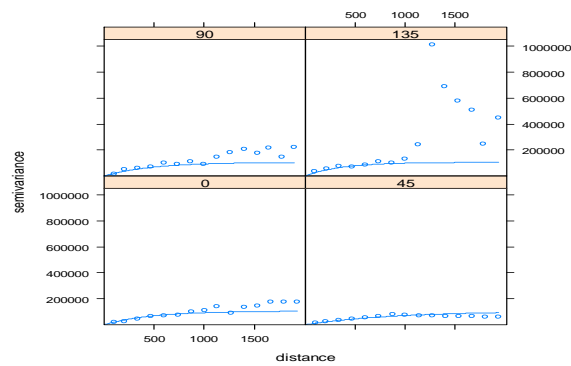


A la vista de los semivariogramas experimentales, parece que los modelos que más se ajustan son:

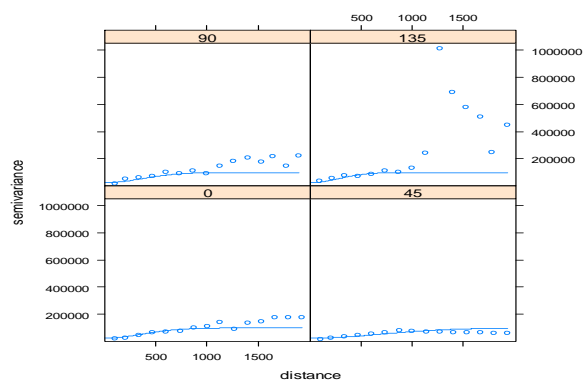
- Sph: Modelo esférico.



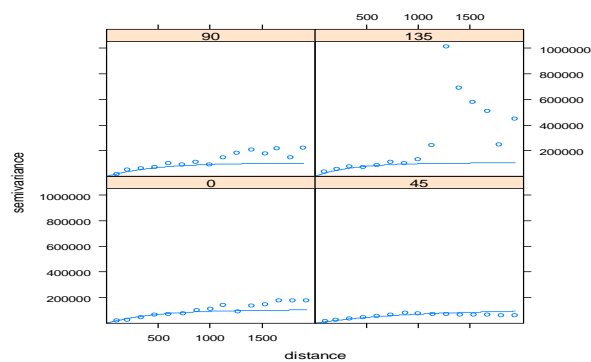
- Exp: Modelo exponencial.



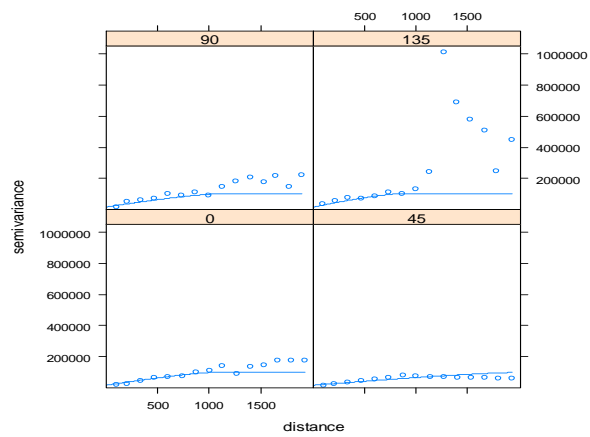
- Gau: Modelo gaussiano.



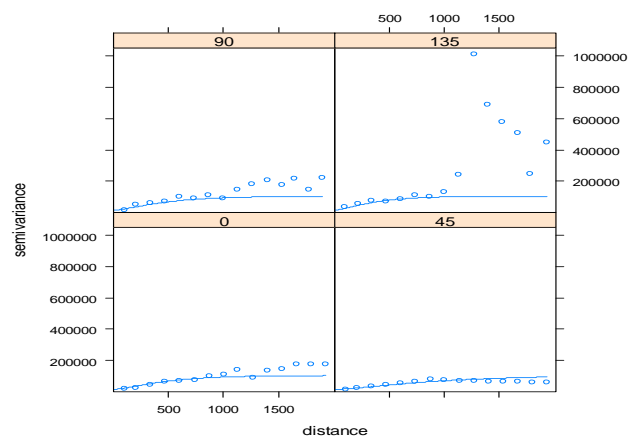
- Ste: Modelo matern, con la parametrización de M. Stei



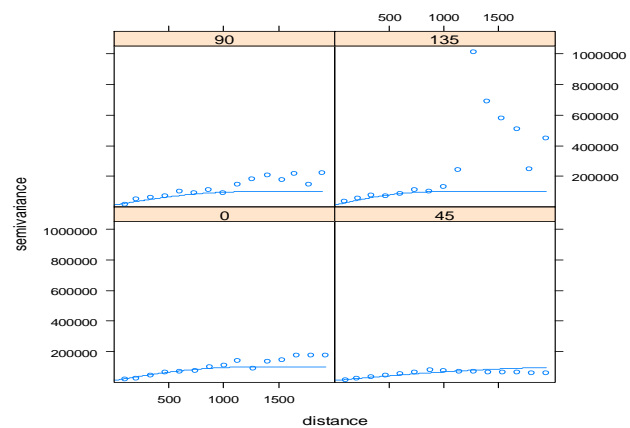
- Cir: Modelo circular.



- Bes: Modelo de Bessel.



- Pen : Modelo pentaesférico.



5.-Ajuste de las funciones mediante diversos métodos y comparación de resultados.

Modelo	Meseta parcial	Rango	Ratio de anisotropia	Pepita	SSE	Fit	Rango efectivo
Esférico	89015.45	2325.835	0.4	13261.30	23631465	7	89015.45
	86289.41	2303.101	0.4	14046.48	3908.996	2	86289.41
Exponencial	107944.773	1188.750	0.4	6542.709	23747599	7	35981.59
	105331.5	902.5492	0.4	0	3614.004	2	35110.5
Gausiano	72672.48	23998588	0.4	21826.86	23998588	7	41957.48
	73597.26	986.508	0.4	23152.21	4087.759	2	42491.4
M.Stein	107955.466	1681.728	0.4	6547.939	23747599	7	
	104770.8	1273.054	0.4	0	3530.271	2	
Circular	14156.71	2047.288	0.4	14156.71	23673028	7	
	84307.82	2098.952	0.4	16184.04	3897.179	2	
Bessel	90086.26	674.609	0.4	16687.42	23683441	7	
	88052.06	617.3699	0.4	15013.53	4587.72	2	
Pentaesféric	90128.49	2805.309	0.4	12375.86	23619585	7	
	88375.20	2697.474	0.4	11894.07	3894.925	2	

Notemos que el rango efectivo es la distancia para la cual el semivariograma alcanza el 95% de la meseta. Este es útil cuando no existe una distancia finita para la cual dos observaciones sean independientes. En el caso del modelo esférico, el rango coincide con el rango efectivo. En cambio, en el modelo exponencial el rango efectivo es rango/3 y en el modelo gausiano es rango/ $\sqrt{3}$.

Las ponderaciones para el ajuste de semivariograma experimental son

- Ajustada por gstat con peso $\frac{N_j}{\{\gamma(h_j)\}^2}$ {fit=2}
- Ajustada por gstat con peso $\frac{N_j}{h_j^2}$ (fit=7)

La suma de los errores cuadráticos del ajuste del modelo (SSE) es menor usando las ponderaciones del ajuste 2 ya que los valores del semivariograma son mayores que las distancias empleadas.

El modelo matern, con la parametrización de M. Stei con ponderación $\frac{N_j}{\{\gamma(h_j)\}^2}$, y el modelo exponencial, con la ponderación $\frac{N_j}{\{\gamma(h_j)\}^2}$, son los que tienen menor SSE; por tanto, serían los semivariogramas teóricos escogidos.