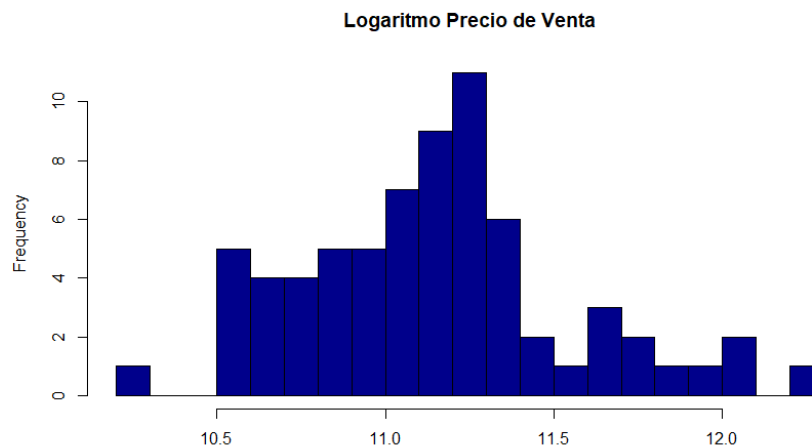


Tarea Final Estadística Aplicada

Pregunta 1

Realice un análisis exploratorio para el logaritmo del precio de venta (*logSellingPr*) sin considerar la información geográfica. Comente.

RESPUESTA



Al observar la distribución del Logaritmo del Precio de Venta (lo llamaré Precio de acá en adelante, pero siempre refiriéndome al Logaritmo del Precio de Venta) podemos notar que sí tiene una tendencia central alrededor de la cuál se esparcen las observaciones. Con un promedio de 11.15, Mediana de 11.16 y Moda de 10.54 nos podemos hacer la idea de que estamos tratando con una distribución simétrica con una diferencia en el nivel general de frecuencia del lado derecho contra el lado izquierdo: la distribución del lado izquierdo tiene frecuencias promedio de 4, mientras que el derecho de 1.

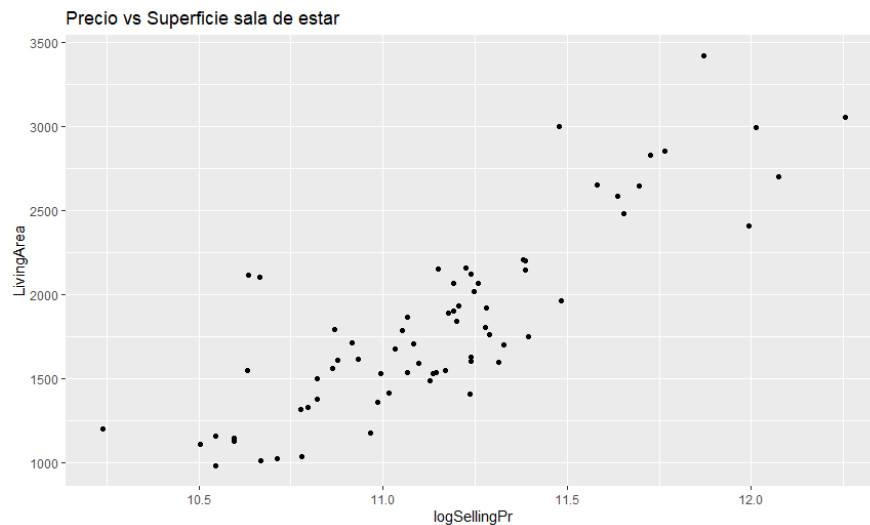
También podemos notar que, si tomamos exclusivamente los valores de 11 en adelante, existe una pequeña concentración en 12-12.5

Pregunta 2:

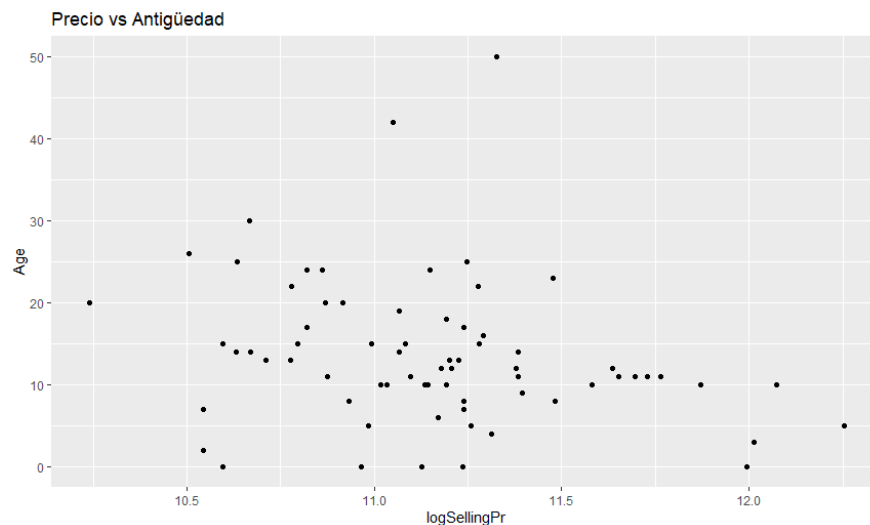
Repita lo anterior, pero ahora utilizando el resto de las variables disponibles. Comente.

RESPUESTA

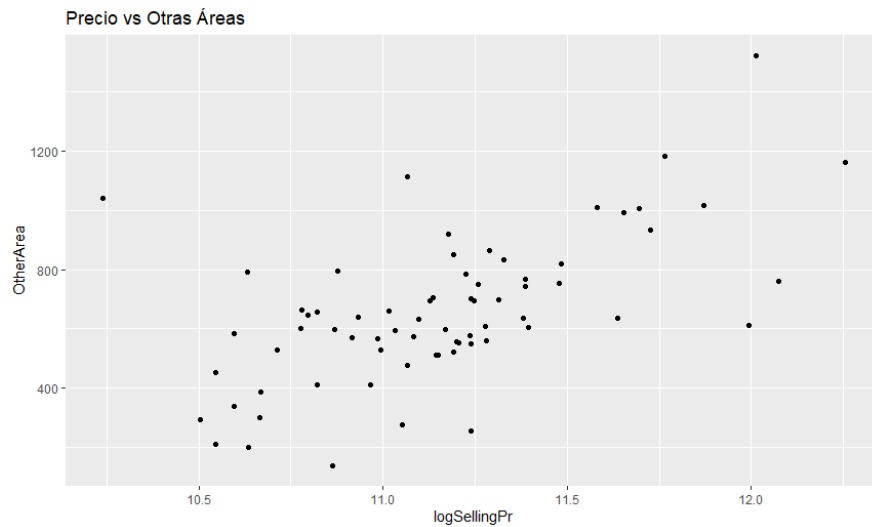
Precio contra LivingArea: Al cruzar la variable de la superficial de la sala de estar contra el precio de la vivienda, podemos ver una clara relación lineal positiva entre las dos variables. Podemos decir entonces que, en general, las casas con una sala de estar más grande tienden a tener un mayor precio.



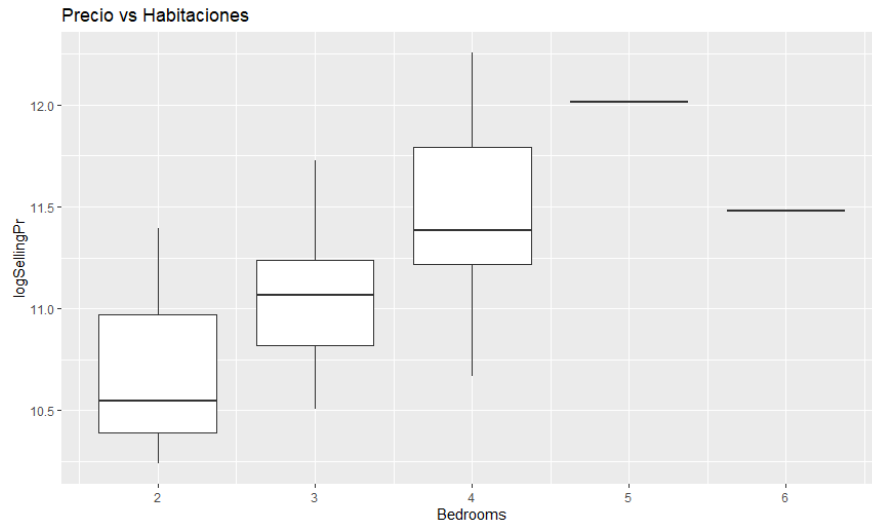
Precio contra Antigüedad de la casa: No se encuentra una relación clara entre la Antigüedad de la casa con el Precio que tiene; sí parece ser que las casas más caras tienen en promedio 10 años. Puede haber una variable intermedia que nos ayude a entender mejor la relación entre estas dos variables, como la zona de la casa.



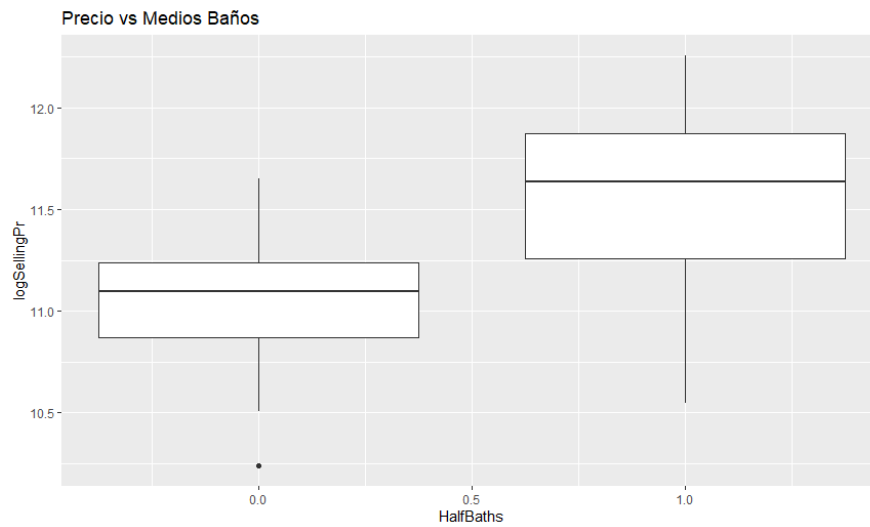
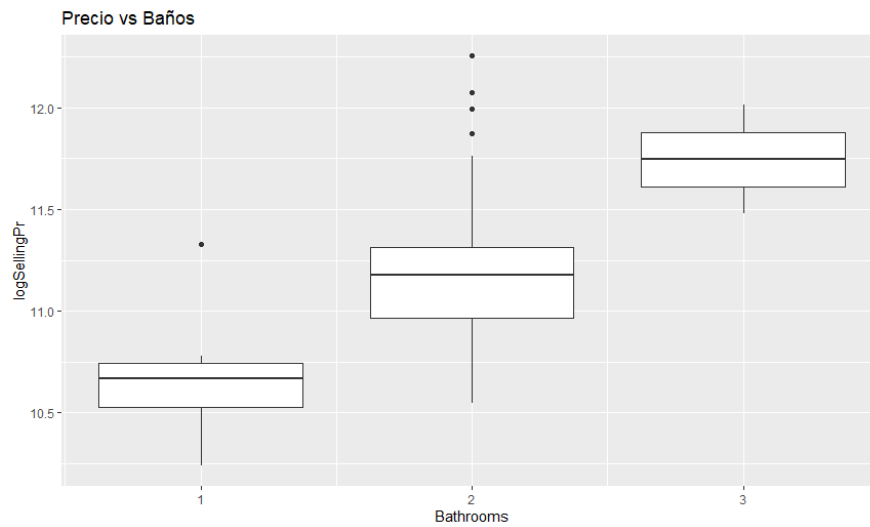
Precio contra Otras Áreas: podemos ver una clara relación lineal positiva entre las dos variables. Podemos decir entonces que, en general, las casas con mayor área tienden a tener un mayor precio.



Precio contra Número de habitaciones: Como el número de habitaciones es una variable que puede considerar como categóricas para fines de visualización, podemos concluir que, en general, las casas con más Habitaciones tienden a ser más caras.



Precio contra Baños y Medios Baños: Para la variable de Baños y Medios Baños, también podemos decir que hay una relación positiva contra el precio, es decir, a mayor número de baños o ante la presencia de medios baños, el precio de la casa sube.



Pregunta 3

Defina las coordenadas mediante la siguiente sentencia: `coordinates(data) <- ~Latitude+Longitude`

RESPUESTA

Ver código de R.

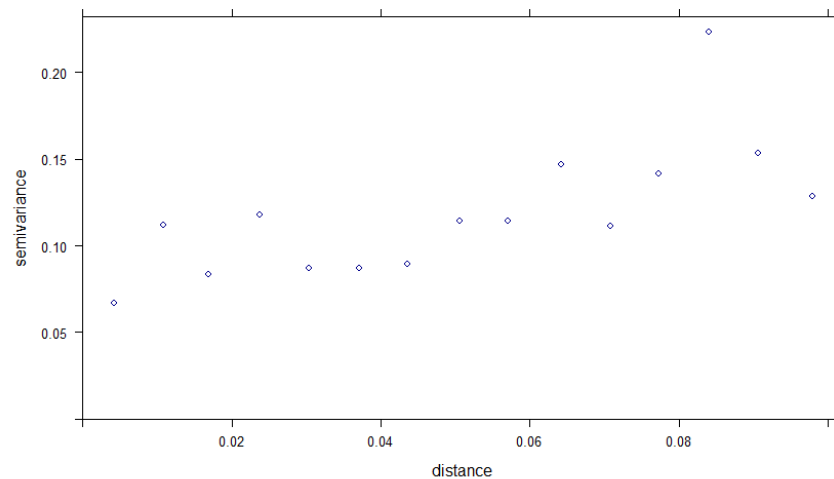
Pregunta 4

Utilizando la función *variogram()* de R, obtenga una representación del variograma empírico o experimental para el logaritmo del precio de venta. Comente.

Nota: Se sugiere utilizar la función *variogram()* de la librería *gstat*, sin embargo, pueden utilizar otra que les sea más conveniente.

RESPUESTA

A continuación, se presenta el variograma:



Con el variograma Podemos observar que existe una relación positiva entre la distancia y la semivarianza. En términos prácticos, esto quiere decir que la autocorrelación espacial entre las observaciones que están cerca es mayor a la autocorrelación espacial de las observaciones que están más lejos, lo que puede implicar que la zona en la que se encuentran las casas influye en el precio de estas.

Además, si descartamos el punto con semivarianza mayor a 20, podemos inferir que un modelo esférico es un buen candidato para modelar este variograma pues presenta una disminución progresiva de la autocorrelación espacial.

Pregunta 5

Utilizando la función `fit.variogram()`, ajuste el variograma teórico considerando los siguientes modelos matemáticos y sus respectivas especificaciones:

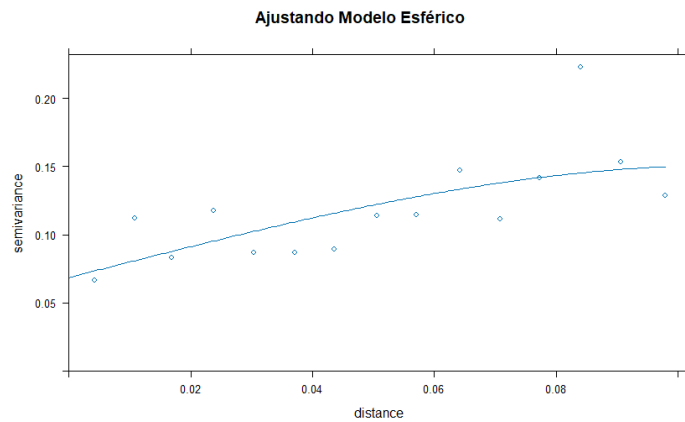
- Esférico con un $\text{psill} = 0.09$, $\text{range} = 0.15$ y $\text{nugget} = 0.03$.
- Exponencial con un $\text{psill} = 0.08$ y $\text{nugget} = 0.06$.

Nota: Dentro de la función `fit.variogram()`, se debe especificar el modelo teórico a través de la función `vgm()`, la cual recibe como argumentos los parámetros antes descritos en el siguiente orden `vgm(psill, model, range, nugget)`. Adicionalmente si no se especifican algunos de esos parámetros, se utiliza, por ejemplo, la sentencia `fit.range=F`.

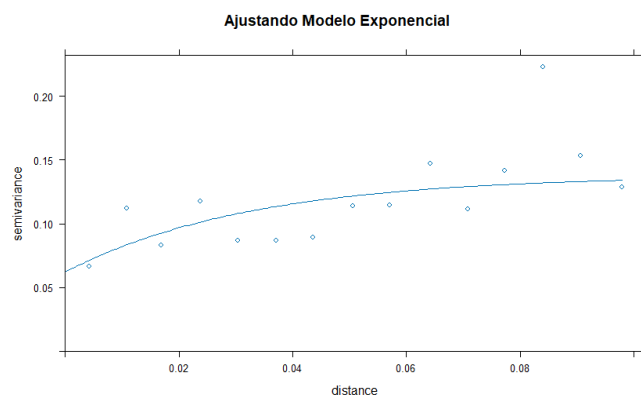
RESPUESTA

A continuación, se presentan los variogramas teóricos

MODELO ESFÉRICO



MODELO EXPONENCIAL



Pregunta 6

A través del método kriging ordinario, realice una predicción del logaritmo del precio de venta, según los modelos teóricos utilizados en el apartado anterior (esférico y exponencial).

Nota: Se recomienda que utilice la función *kriging()* de la librería *kriging*, la cual recibe como argumentos las coordenadas (*Latitude* y *Longitude*), la variable de interés (*logSellingPr*) y el modelo teórico usado (*spherical* o *exponential*).

RESPUESTA

Ver Código R.

Pregunta 7

El modelo exponencial da un precio mayor a comparación del modelo esférico:

Predicción para Modelo Esférico: 11.07415

Predicción para Modelo Exponencial: 11.09257

Sin embargo, el Modelo Exponencial nos da una varianza menor.

PREGUNTA 8

A continuación, se presenta un gráfico que representa la densidad de los puntos en la muestra ponderados por el precio. Con este gráfico podemos ver que hay vecindarios más homogéneos que otros. Esto se puede deber al tamaño de los puntos cercanos. Este es otro insumo para comprobar el que precio de las casas están correlacionados con el área dónde están y el precio de las muestras cercanas.

