

Técnicas de Análisis Cuantitativas y Cualitativas

Resolución del ejercicio de evaluación 2

Marcos Rial Docampo

4 de noviembre de 2015

En este ejercicio de evaluación se nos presentan datos de superficie agrícola abandonada, densidad de población y altitud media de una serie de 50 observaciones tomadas en otros tantos municipios gallegos. Se destaca la importancia que tiene la densidad de población o la elevación sobre los cambios de uso de suelo que afectan a superficie agrícola.

En los gráficos de la figura 1 se presenta la relación existente entre el abandono de la superficie agrícola y las otras dos variables: elevación y densidad de población. A primera vista podemos observar como aparentemente la relación entre el abandono y la densidad de población no ofrece indicios de correlación entre ambas. Lo que no ocurre con la segunda relación entre el abandono y la elevación, donde sí puede haber correlación. Pero esta es una valoración a priori solo observando el aspecto de las relaciones en una gráfica. Podría ser que alguna de las variables necesitara ser transformada para que en la gráfica se ofreciera una visión más fidedigna, pero no se ve necesario.

Comprobamos la existencia de correlación entre las variables abandono y elevación y densidad de población. Partimos de la hipótesis nula (H_0) de que no existe correlación entre las dos variables estudiadas en cada caso. Para ello empleamos el comando de R *cor.test()* como se muestra en la figura 2. Vemos en la figura 3 que los resultados nos arrojan una alta correlación entre la variable abandono y la de elevación (coeficiente de Pearson de 0,8718). Mientras que para la otra relación, entre el abandono y la densidad de población, la correlación no existe (coeficiente de Pearson de -0,5379). Se confirma lo expuesto en el párrafo anterior.

Podemos obtener un modelo de regresión lineal múltiple entre variables obteniendo de este el plano de regresión:

$$y = -0,1375 + 0,0002x_1 + 0,0007x_2 \quad (1)$$

siendo y la variable abandono y x_1 y x_2 las variables densidad de población y elevación.

De dicho modelo de regresión múltiple extraemos la información que nos permitirá saber que variable es significativa y cual no, en el caso de haberlas, respecto de la variable abandono. Partiendo con la H_0 de que no hay relación entre las variables, observamos los coeficientes de regresión ofrecidos de el cuadro 1 donde vemos que existe una fuerte relación del abandono con la variable elevación, mientras que con la variable densidad

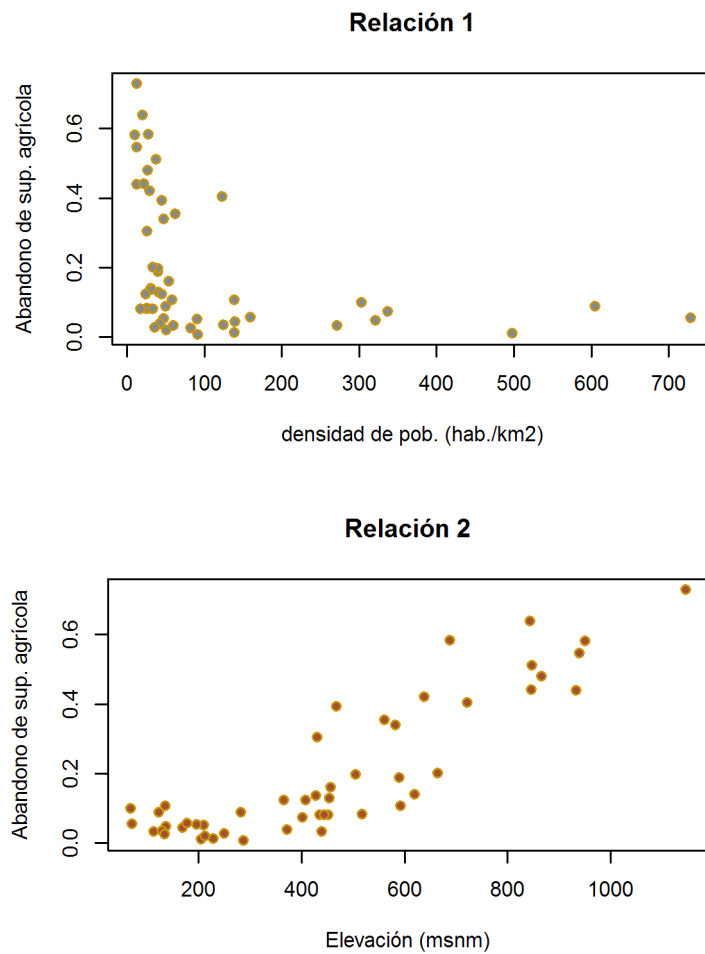


Figura 1: Relación entre el abandono de superficie agrícola y las variables densidad de población y elevación.

```

1 # Estudio de correlacion entre variables
2 # H0 = las variables no estan correlacionadas
3 # Abandono vs densidad de poblacion
4 cor.test(datos$abandon.uaa, datos$pop.dens, alternative = "greater",
5         method = "pearson", conf.level = 0.95)
6 # Abandono vs elevacion
7 cor.test(datos$abandon.uaa, datos$elevation, alternative = "greater",
8         method = "pearson", conf.level = 0.95)

```

Figura 2: Empleo del comando `cor.test` con el método de Pearson y nivel de confianza al 95 %.

<pre> Pearson's product-moment correlation data: datos\$abandon.uaa and datos\$pop.dens t = -2.6908, df = 48, p-value = 0.9951 alternative hypothesis: true correlation is greater than 0 95 percent confidence interval: -0.5505354 1.0000000 sample estimates: cor -0.3620322 </pre>
<pre> Pearson's product-moment correlation data: datos\$abandon.uaa and datos\$elevation t = 12.328, df = 48, p-value < 2.2e-16 alternative hypothesis: true correlation is greater than 0 95 percent confidence interval: 0.8006674 1.0000000 sample estimates: cor 0.8717672 </pre>

Figura 3: Resultados del test de correlación.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.375e-01	3.732e-02	-3.684	0.000592
pop.dens	1.967e-04	1.071e-04	1.837	0.072527
elevation	6.987e-04	6.005e-05	11.635	1.94e-15

Cuadro 1: Resultados del análisis de significación de la regresión múltiple.

de población es menor. Por lo tanto, con un p-valor de 1.94e-15 (muy cercano a cero) rechazamos la hipótesis nula y aceptamos que la variable elevación es significativa en relación con la variable abandono.

Analizamos los supuestos de partida (figura).

Ahora, una vez aplicado un modelo de regresión múltiple y extraída la información obtendremos la recta de regresión entre la variable abandono y elevación (puesto que es la única relación que presenta correlación) obtenemos la gráfica de la figura 4 y observamos gracias a invocar el valor “modelo2”, creado para aplicar la función de regresión lineal $lm()$ como en el caso anterior, que devuelve los coeficientes de significación que nos permiten obtener la recta de regresión:

$$y = -0,0896 + 0,0006x \quad (2)$$

siendo y la variable abandono y x la variable elevación.

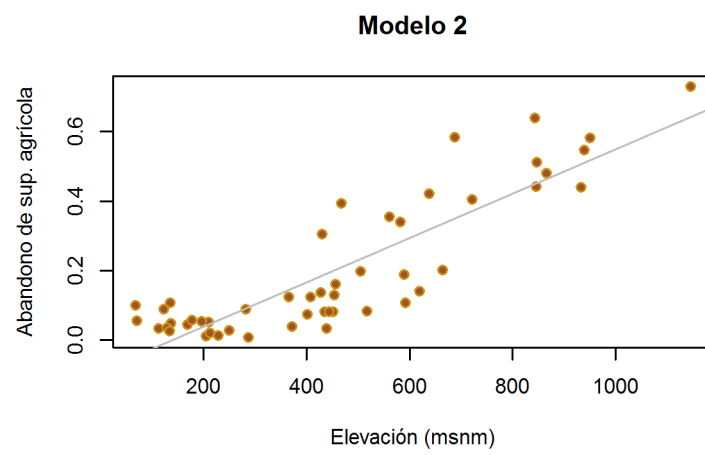


Figura 4: Modelo de regresión lineal para la relación entre abandono y elevación.