

TALLER 1: REPASO DE RLM CON VARIABLES INDICADORAS

Problema: En cada una de tres zonas residenciales de cierta ciudad fueron seleccionadas al azar cinco casas recientemente vendidas, con el fin es estudiar la relación entre el precio de venta de la propiedad (Y) y el valor catastral de la propiedad (X). Los datos se muestran en la siguiente tabla (en miles de dólares):

Propiedad	Valor catastral	Precio de venta	Zona
1	33.10	42.50	A
2	42.00	36.80	A
3	47.80	42.60	A
4	53.40	41.20	A
5	59.60	48.60	A
6	63.90	75.20	B
7	68.40	83.40	B
8	72.30	83.30	B
9	77.80	116.80	B
10	80.80	114.30	B
11	96.50	122.80	C
12	101.80	125.60	C
13	106.20	132.50	C
14	112.60	127.40	C
15	120.50	147.80	C

1. Analice gráficamente la relación entre el precio de venta y el valor catastral teniendo en cuenta la información de la zona ¿Qué se puede concluir acerca de la relación entre el precio de venta y el valor catastral? ¿difiere según la zona? ¿tendrá sentido realizar una regresión lineal con las 15 observaciones ignorando la información de la zona?
2. Si se considera que la relación lineal entre el precio de venta vs. el valor catastral puede diferir según la zona residencial, postule el modelo de regresión indicado con variables indicadoras para modelar la situación planteada, tome como zona de referencia
 - a. La zona A.
 - b. La zona B
 - c. La zona C

Escriba en cada caso tanto la ecuación general como la ecuación de las rectas de precio de venta vs. el valor catastral en cada una de las tres zonas, recuerde incluir los supuestos del modelo. Use en la notación de las indicadora a las variables I_1 para la zona A, I_2 para la zona B e I_3 para la zona C. Defina cada una de estas variables.

3. Tome los resultados del modelo ajustado en 2-a. Escriba la ecuación general ajustada así como las ecuaciones ajustadas para cada una de las tres zonas.
4. De nuevo con base en los resultados del modelo ajustado en 2-a, ¿Existe alguna razón para creer que las pendientes en las zonas A y B son diferentes? Realice el test de hipótesis para concluir.
5. Ajuste el modelo donde se considera que el efecto del valor de catastro sobre el precio de venta no difiere según la zona aunque el precio medio de venta es diferente en cada zona. Analice los residuales de este modelo.

PROGRAMA R

```
datos=data.frame(scan(what=list(Valor.Catastro=0,Precio.Venta=0,Zona=""))
33.1 42.5 A
42.0 36.8 A
47.8 42.6 A
53.4 41.2 A
59.6 48.6 A
63.9 75.2 B
68.4 83.4 B
72.3 83.3 B
77.8 116.8 B
80.8 114.3 B
96.5 122.8 C
101.8 125.6 C
106.2 132.5 C
112.6 127.4 C
120.5 147.8 C

datos

attach(datos)

plot(Valor.Catastro,Precio.Venta,pch=as.numeric(Zona),col=as.numeric(Zona),xlab="Valor catastral propiedad (X)",
      ylab="Precio de venta propiedad (Y)",cex=2.0)
legend("topleft",legend=c("Zona A","Zona B","Zona C"),pch=1:3,col=1:3,cex=2)

#USANDO POR DEFECTO COMO ZONA DE REFERENCIA LA A
###MODELO GENERAL: RECTAS DIFERENTES TANTO EN PENDIENTE COMO EN INTERCEPTO###
modelo1=lm(Precio.Venta~Valor.Catastro*Zona)
summary(modelo1)
```

TALLER 1: REPASO DE RLM CON VARIABLES INDICADORAS

```
confint(modelo1)
```

```
#USANDO COMO ZONA DE REFERENCIA LA B
```

```
###MODELO GENERAL: RECTAS DIFERENTES TANTO EN PENDIENTE COMO EN INTERCEPTO###
```

```
Zona2=relevel(Zona,ref="B")
```

```
modelo2=lm(Precio.Venta~Valor.Catastro*Zona2)
```

```
summary(modelo2)
```

```
confint(modelo2)
```

```
#USANDO COMO ZONA DE REFERENCIA LA C
```

```
###MODELO GENERAL: RECTAS DIFERENTES TANTO EN PENDIENTE COMO EN INTERCEPTO###
```

```
Zona3=relevel(Zona,ref="C")
```

```
modelo3=lm(Precio.Venta~Valor.Catastro*Zona3)
```

```
summary(modelo3)
```

```
confint(modelo3)
```

```
###MODELO CON RECTAS DIFERENTES SÓLO EN EL INTERCEPTO; NIVEL DE REFERENCIA ZONA A###
```

```
modelo1b=lm(Precio.Venta~Valor.Catastro+Zona)
```

```
summary(modelo1b)
```

```
confint(modelo1b)
```

```
#ANÁLISIS RESIDUALES EN modelo1b (RESIDUALES COMUNES)
```

```
plot(Valor.Catastro,residuals(modelo1b),pch=as.numeric(Zona),cex=2,col=as.numeric(Zona),  
      ylim=c(-20,20),cex.lab=2)
```

```
abline(h=c(-2*summary(modelo1b)$sigma,0,2*summary(modelo1b)$sigma),lty=2)
```

```
legend("topright",legend=c("A","B","C"),pch=c(1:3),col=c(1:3),cex=2)
```

```
plot(as.numeric(Zona), residuals(modelo1b),pch=as.numeric(Zona),cex=2,col=as.numeric(Zona),ylim=c(-20,20),  
      xlab="Zona",xaxt="n",cex.lab=2)
```

```
abline(h=c(-2*summary(modelo1b)$sigma,0,2*summary(modelo1b)$sigma),lty=2)
```

```
axis(1,at=c(1,2,3),c("A","B","C"))
```

```
legend("topright",legend=c("A","B","C"),pch=c(1:3),col=c(1:3),cex=2)
```

```
plot(fitted(modelo1b), residuals(modelo1b),pch=as.numeric(Zona),cex=2,col=as.numeric(Zona),  
      ylim=c(-20,20),cex.lab=2)
```

```
abline(h=c(-2*summary(modelo1b)$sigma,0,2*summary(modelo1b)$sigma),lty=2)
```

```
legend("topright",legend=c("A","B","C"),pch=c(1:3),col=c(1:3),cex=2)
```

```
test=shapiro.test(residuals(modelo1b))
```

```
test
```

```
qqnorm(residuals(modelo1b),pch=as.numeric(Zona),cex=2,col=as.numeric(Zona))
```

```
qqline(residuals(modelo1b))
```

```
legend("topleft",legend=rbind(names(test),test),cex=0.8)
```

```
legend("bottomright",legend=c("A","B","C"),pch=c(1:3),col=c(1:3),cex=2)
```

```
detach(datos)
```