

Estadística (Q) - Clase 20 (17/06/2014)

Mínimos cuadrados pesados

1. Varias inmobiliarias han recolectado datos viendo el valor real (Y en miles de pesos) de las ventas de departamentos que han realizado en el último año en el barrio de Palermo y los metros cuadrados de los departamentos (X). Desean plantear un modelo lineal que relacione dichos valores, con el objeto de estimar mediante mínimos cuadrados la media del valor del metro cuadrado en dicho barrio. A continuación un gráfico

Si planteamos el modelo

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

2. ¿hay alguno de los supuestos del modelo lineal del que se pueda sospechar con el gráfico de XY observado?
3. Se sospecha que $s.d.(\varepsilon_i)$ depende linealmente de X_i . ¿Cómo se podría intentar estimar dicha dependencia?
4. Supongamos que se sabe que $S.D.(\varepsilon_i) = 3X_i$. Transforme los datos para armar un modelo lineal que cumpla las hipótesis y estime los parámetros del nuevo modelo y úselos para construir la recta de mínimos cuadrados del viejo modelo. ¿Cuál es finalmente la recta de mínimos cuadrados que elegiría para estimar?

```
w<-1/(3*x)^2 #es el vector de pesos
ajuste1<-lm(y~x)
ajuste2<-lm(y~x,weights=w)
ajuste3<-lm(y2~x2-1,weights=w)
```

```
> shapiro.test(rstandard(ajuste1))
```

Shapiro-Wilk normality test

```
data:  rstandard(ajuste1)
W = 0.956, p-value = 0.01414
```

```
> shapiro.test(rstandard(ajuste2))
```

Shapiro-Wilk normality test

```
data:  rstandard(ajuste2)
W = 0.9892, p-value = 0.8065
```

```
> shapiro.test(rstandard(ajuste3))
```

Shapiro-Wilk normality test

```
data:  rstandard(ajuste3)
```

W = 0.9892, p-value = 0.8052

```
> summary(ajuste2)
```

Call:

```
lm(formula = y ~ x, weights = w)
```

Weighted Residuals:

Min	1Q	Median	3Q	Max
-2.32830	-0.55453	0.02148	0.69910	2.32115

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	78.086	56.599	1.380	0.172
x	8.858	1.039	8.522	2.22e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9955 on 69 degrees of freedom

Multiple R-squared: 0.5128, Adjusted R-squared: 0.5058

F-statistic: 72.63 on 1 and 69 DF, p-value: 2.216e-12

Call:

```
lm(formula = y ~ x - 1, weights = w)
```

Weighted Residuals:

Min	1Q	Median	3Q	Max
-2.27248	-0.69065	0.00596	0.62764	2.15472

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
x	10.2063	0.3567	28.61	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.002 on 70 degrees of freedom

Multiple R-squared: 0.9212, Adjusted R-squared: 0.9201

F-statistic: 818.6 on 1 and 70 DF, p-value: < 2.2e-16

