

- Fecha esperada de terminación:

1. Considere el modelo de regresión lineal simple

$$y = \beta_0 + \beta_1 x + \epsilon$$

donde  $\epsilon \sim N(0, \sigma^2)$ .

- a) Muestre que los estimadores de mínimos cuadrados (EMC) están dados por

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}; \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

donde  $S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$  y  $S_{xx} = \sum (x_i - \bar{x})^2$ .

*Sugerencia:* Resuelva las ecuaciones normales y use el hecho que  $\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum y_i (x_i - \bar{x})$ .

- b) Muestre que los EMC  $\hat{\beta}_0$  y  $\hat{\beta}_1$  son insesgados.

- c) Muestre que

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \left( \frac{1}{S_{xx}} \right); \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

- d) Muestre que la covarianza entre los estimadores  $\hat{\beta}_0$  y  $\hat{\beta}_1$  es  $\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \frac{\bar{x}}{S_{xx}}$ .

- e) Muestre que la *respuesta media ajustada* al nivel  $x$ ,  $\hat{y}(x)$

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x \sim N \left( \beta_0 + \beta_1 x, \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \right)$$

y que para una *nueva observación* al nivel  $x$ ,  $\dot{y}(x)$ :

$$\dot{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \epsilon \sim N \left( \beta_0 + \beta_1 x, \sigma^2 \left( 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}} \right) \right)$$

2. Una empresa de televisión por cable encargó a un bufete un estudio de mercado para conocer el perfil de los clientes potenciales de una zona residencial formada por dos colonias. Las colonias constan de 12 y 25 manzanas con un total de 236 y 605 hogares, respectivamente. Mediante muestreo probabilístico (no discutido aquí) se seleccionó una muestra de ocho manzanas y cinco hogares por manzana. En cada hogar seleccionado se recabaron varias respuestas de las que presentamos solamente algunas de éstas.

	Variable	Descripción
1	Colonia	Colonia a la que pertenece el hogar de la zona residencial
2	Manzana	Número de manzana a la que pertenece el hogar
3	Adultos	Número de adultos por hogar
4	Niños	Número de niños menores de 12 años por hogar
5	Teles	Número de televisores por hogar
6	Tipo	Tipo de televisor que posee: blanco y negro (B), color (C), ambos (A)
7	TVtot	Suma del número de horas frente al televisor en la semana de todos los miembros de la familia
8	Renta	Cantidad máxima de renta que el jefe del hogar estaría dispuesto a pagar al mes por servicio de TV por cable (múltiplos de \$5)
9	Valor	Valor catastral del hogar (m\$). La respuesta se usa para dar idea aproximada del ingreso familiar

La información la encuentra en el archivo de datos `cableTV.dat` en *Comunidad. z*

- a) Ajuste por mínimos cuadrados un modelo de regresión lineal simple para la respuesta *renta*, con el *valor catastral* (en miles de pesos) como variable independiente. Calcule  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ , y  $\hat{\sigma}^2$ , y grafique los residuales  $\hat{r}_i = y_i - \hat{y}_i$ , contra el regresor  $x_i$  (gráfica de dispersión  $\{(x_i, \hat{r}_i)\}$ ).
- b) ¿Cuál es la significancia de la regresión? (Valor- $p$  del estadístico  $F$ .)
- c) Repita los incisos anteriores pero sin considerar los 2 casos donde  $y = 0$ . ¿Consideraría los nuevos coeficientes estadísticamente iguales a los anteriores? Comente.
- d) Compare los coeficientes de determinación  $R^2$  en ambos casos. Comente.

Tabla 1: Resultados parciales de la encuesta de venta de televisión por cable (Aguirre et al. 2006).

obs.	colonia	manzana	adultos	niños	teles	renta	tvttot	tipo	valor
1	2	20	3	2	2	50	68	B	79928
2	2	25	3	3	1	65	82	B	94415
3	2	20	1	2	1	45	40	A	120896
4	2	8	2	2	2	35	56	A	132867
5	2	25	1	2	0	0	0	N	141901
6	2	14	1	2	0	0	0	N	147997
7	2	22	2	1	1	65	30	A	156410
8	2	20	3	1	3	45	62	C	156841
9	2	25	3	3	2	70	82	A	157041
10	2	20	2	2	3	45	60	C	161222