

**Primer Examen Parcial**

1. (8 pts.) Los siguientes datos corresponden al número de kioscos existentes en diversos sectores de Caracas:

33	47	11	51	429
473	23	225	29	47
11	80	14	3	43

- a) (3 pts.) Construya un diagrama de caja para los datos anteriores
- b) (2 pts.) ¿Puede afirmarse que estos datos provienen de una distribución aproximadamente normal? Justifique su respuesta.
- c) (3 pts.) Construya un nuevo diagrama de caja para el logaritmo de los datos. ¿Qué cambios observa ahora?
2. (12 pts.) El dueño de una agencia de servicios de correos contrata personal adicional cada vez que hay una demanda pico que excede la carga de trabajo de sus tres empleados. Para probar si esto es necesario, él cuenta el número de paquetes (Y) enviados en días *pico* y en días normales.

No. Paquetes (Y)	50	110	90	150	140	180	190	310	330	340	360	380	360
No. Empleados (X)	1*	2*	2*	3	3	3	4	6	6	7	8	10	10

(\*) Empleados enfermos o de vacaciones

- a) (4 pts.) Estime los parámetros del modelo de regresión lineal simple  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
- b) (3 pts.) Haga los cálculos que considere convenientes para probar la hipótesis:

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0$$

- c) (2 pts.) Calcule los valores de  $R^2$  y  $R^2$ -ajustado para el modelo anterior y concluya sobre los resultados obtenidos.
- d) (3 pts.) Al ajustar un modelo de la forma  $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$  se obtiene una suma de cuadrados del error  $SSE = 3502$ . Haga las pruebas que considere convenientes para determinar cuál es el mejor modelo para los datos.

Los siguientes cálculos pueden ser de utilidad:  $\sum x_i = 65$ ;  $\sum y_i = 2990$ ;  $\sum x_i^2 = 437$ ;  $\sum y_i^2 = 857500$ ;  $\sum x_i y_i = 19120$ ;

$$(X^t X)^{-1} = \begin{pmatrix} 0,300137362 & -0,044643 \\ -0,044643 & 0,00892857 \end{pmatrix}$$

$$(X^t Y) = \begin{pmatrix} 2990 \\ 19120 \end{pmatrix}$$

## Primer Examen Parcial CO2124

3. (10 pts.) En tres urbanizaciones de Caracas se cuenta el número de accidentes viales que ocurren desde las 6:00 pm del viernes hasta las 6:00 pm del domingo por cuatro semanas consecutivas.

Urbanización	Semanas				Total
	1	2	3	4	
A	13	22	18	39	92
B	16	24	17	44	101
C	5	4	1	22	32
Total	34	50	36	105	225

- a) (5 pts.) Asumiendo que los datos siguen un modelo de la forma  $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$  con  $\sum_i \alpha_i = 0$  y  $\varepsilon_i \sim N(0, \sigma^2)$ . construya la Tabla ANOVA para determinar si existen diferencias entre el número de accidentes promedio para las distintas urbanizaciones. Los siguientes cálculos pueden ser de utilidad:  $\sum y_{ij} = 225$ ;  $\sum y_{ij}^2 = 6081$ ;  $\bar{y}_{..} = 18,75$ .
- b) (3 pts.) Estime los parámetros de este modelo.
- c) (2 pts.) Se sospecha que existe un efecto de las semanas en el análisis y que estas deben ser consideradas como *bloques*. Plantee un modelo alternativo para los datos que incluya esta consideración y enumere cuáles son las suposiciones de este modelo.