

35. Selten R. (1994). Descriptive approaches to cooperation. In S. Hart and A. Mas-Colell (eds.), *Cooperation: game-theoretic approaches*. NATO ASI series. Springer-Verlag.
36. Selten R (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* 4 pp. 25-55.
37. Shapley L.S. (1953). A Value for N-person Games. In HW Karlin and AW Tucker (eds), *Contributions to the Theory of Games II*, Princeton University Press, pp. 307-317.
38. Tucker A.W. (1955). *Game theory and Programming*. Department of Mathematics, The Oklahoma Agricultural and Mechanical College, Stillwater (mimeo).
39. van Damme E. (1991). *Stability and perfection of Nash equilibria*, second, revised and enlarged edition. Springer-Verlag.
40. von Neumann J.(1928). Zur Theorie der Gesellschaftsspiele. *Math. Annal.* 100, 295-320.
41. von Neumann J.(1937). Über ein ökonomisches gleichungssystem und eine verallgemeinerung des Brouwerschen fixpunktsatzes. *Ergebnisse eines Mathematik Kolloquiums* 8 pp. 73-83.
42. von Neumann J. and Morgenstern O. (1944). *Theory of games and economic behavior*. Princeton University Press.
43. Zermelo E. (1913). Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. *Proceedings 5th International Congress of Mathematicians* 2 pp. 501-504.

### 3. ARTÍCULOS DE APLICACIÓN

#### LA DETERMINACIÓN DEL «LUGAR DE LA MANCHA» COMO PROBLEMA ESTADÍSTICO

F. Javier Girón y M. Jesús Ríos

Departamento de Estadística, Universidad de Málaga y Universidad Complutense de Madrid

“En un lugar de la Mancha de cuyo nombre no quiero acordarme ...”

Desde la primera frase, Cervantes establece el principio que regula toda la redacción de ‘El Quijote’: la incertidumbre. Incertidumbre autoral, incertidumbre nominativa, incertidumbre genérica y, en consecuencia, el ojo crítico de la imaginación, dándole “una segunda realidad” a lo que pasa por ser “realidad”.

Carlos Fuentes (2005)

#### 1 Introducción

Muchos —curiosos y estudiosos— han sido los que desde la publicación de la primera parte del *Quijote* por Miguel de Cervantes en

1605, han intentado situar *el lugar de la Mancha* dentro del territorio geográfico de esa zona. Algunos teóricos de la Literatura han opinado que ese lugar era puramente imaginario, por lo que no tendría sentido buscarlo. No obstante, está claro que Cervantes —desde la primera frase «*En un lugar de la Mancha de cuyo nombre no quiero acordarme [...]*», hasta uno de los últimos párrafos del último capítulo de la segunda parte de la novela «*Este fin tuvo el Ingenioso Hidalgo de la Mancha, cuyo lugar no quiso poner Cide Hamete puntualmente, por dejar que todas las villas y lugares de la Mancha contendiesen entre sí por ahijársele y tenérsele por suyo, como contendieron las siete ciudades de Grecia por Homero*»—, va dejando caer pistas que, aunque no exentas de inconsistencias en algunos casos, llevan al convencimiento de que hay realismo

geográfico en la novela.

Según Parra et al. (2005) de una lectura cuidadosa del *Quijote* se pueden obtener los siguientes datos:

- a) El *lugar de la Mancha* está dentro del Campo de Montiel (prólogo; Cap. 1; Cap. VII; Cap. LII, Primera parte y Cap. VII, Segunda Parte).
- b) El *lugar de la Mancha* está a dos días de Sierra Morena (SM), (Cap. XXIX, Primera parte).
- c) El *lugar de la Mancha* está a una noche y dos días de El Toboso (ET), (Cap. XXXVII, Primera parte).
- d) El *lugar de la Mancha* está a dos días y algunas horas de Puerto Lápice (PL), (Caps. XII a XXIV, Primera parte).
- e) El *lugar de la Mancha* está entre un día y medio y dos días del *Punto Tarfe* (Munera) (Cap. LXXII, Segunda parte).
- f) La velocidad de Rocinante/Rucio está comprendida entre los 30 y 35 kms por jornada (Caps. XI, XXVII, Primera parte).

Al cumplirse este año el cuarto centenario de la publicación del *Quijote*, el trabajo arriba citado llevado a cabo por un equipo multidisciplinar de la UCM (Parra et al., 2005) resuelve, desde distintos enfoques (literarios, sociológicos, geométricos, ...), el problema de situar geográficamente *el lugar de la Mancha*. En un segundo trabajo presentado en las XII JAEM (Ríos, Montero y Parra, 2005) se aborda directamente el problema desde el punto de vista de la *Teoría de la Decisión Multicriterio* asociando a cada uno de los pueblos candidatos a ser el lugar de la Mancha un vector de discrepancias, construyendo distintas funciones de valor basándonos en las distancias o métricas usuales, lo que conduce a varias posibles soluciones que proporcionan distintos órdenes de preferencia de los pueblos candidatos, teniendo además en cuenta los diversos valores de la velocidad. En las conclusiones se pone de manifiesto que, dada la incertidumbre inherente a los datos, se hace necesario emplear un modelo estadístico para valorar dicha incertidumbre que, además, permita calcular las probabilidades a

posteriori de cada uno de los pueblos candidatos a ser el lugar de la Mancha. La cita introductoria del gran escritor mexicano Carlos Fuentes sobre el papel de la incertidumbre en el *Quijote* se nos antoja como una premonición para enfocar el problema de determinar *el lugar de la Mancha* desde una disciplina como es la Estadística, que tiene como objetivo fundamental el estudio de aquellos problemas en los que hay presente incertidumbre en cualquiera de sus formas o manifestaciones.

En la sección 2 se describen los datos, haciendo la salvedad de que los correspondientes a las jornadas recorridas son solamente aproximaciones y que además estos datos contienen incertidumbre, a diferencia de lo que ocurre con las distancias entre los pueblos de origen y los cuatro destinos, que son precisas. En esta nota, por razones de espacio, no se considera el problema de incluir incertidumbre en el modelo cuando hay datos inciertos. En la sección 3 se proponen tres posibles modelos estadísticos para determinar el lugar de la Mancha. Usando técnicas de selección de modelos se elige el mejor de ellos, se estiman sus parámetros y, finalmente, se calculan las probabilidades a posteriori de los pueblos candidatos. Por último, en la sección 4, se plantean brevemente las posibles extensiones del modelo básico.

## 2 Descripción de los datos

Tal como se ha explicado en la introducción, se tiene una idea aproximada de las jornadas que necesitaron las cabalgaduras en las tres salidas a Puerto Lápice, Sierra Morena y El Toboso, más la correspondiente al denominado *punto Tarfe* que Parra et al. (2005) sitúan finalmente en Munera. Estas jornadas, medidas en días, las representamos por el vector  $\mathbf{d} = (d_1, d_2, d_3, d_4)$ , donde  $d_1=2.25$ ,  $d_2=2.00$ ,  $d_3=2.50$  y  $d_4=1.75$ .

Por otra parte, no se conoce con precisión las distancias reales a los cuatro destinos desde los veinticuatro pueblos candidatos a ser *el lugar de la Mancha* al no saber con seguridad los caminos y veredas de la época que conducían a estos destinos. Por ello, en esta primera aproximación al problema, hemos tomado como datos siguiendo a Parra et al.

las distancias en línea recta desde el pueblo de origen al destino, ya que son fáciles de medir y constituyen una cota inferior de las

distancias reales recorridas. Los datos, en kilómetros, se han obtenido de la guía Michelin y aparecen en la Tabla 1.

Pueblos	PL	SM	ET	MU	Pueblos	PL	SM	ET	MU
Albaladejo	96.8	63.2	99.6	54.4	Ossa	74.4	88.0	64.0	24.0
Alcubillas	70.0	48.4	84.4	64.4	Puebla P.	96.0	51.6	104.0	64.8
Alhambra	59.6	65.2	67.6	51.2	Ruidera	64.8	80.8	59.6	35.6
Almedina	89.2	52.0	97.6	61.2	S. C. Caña.	92.4	59.6	96.8	55.2
Argamasilla	40.0	84.8	43.2	53.2	Solana	47.2	61.6	66.0	65.6
Cañamares	89.6	73.6	87.2	40.0	Terrinches	95.6	60.4	100.0	56.8
Carrizosa	67.6	63.6	73.6	49.2	T. J. Abad	89.2	45.2	102.0	70.8
Castellar	88.4	22.8	109.6	88.0	Torrenueva	76.0	26.8	100.8	87.6
Cozar	81.2	44.8	94.0	66.0	Villahermosa	82.0	64.8	84.4	46.4
Fuencollana	77.6	59.6	83.6	51.6	Villamanrique	95.2	45.2	106.4	70.4
Membrilla	40.8	62.0	66.8	74.8	Villa Fuente	96.4	76.0	93.2	41.6
Montiel	86.8	62.8	90.0	49.6	Villa Infantes	76.4	54.4	85.6	56.8

**Tabla 1.** Distancias en kms, medidas en línea recta, de los veinticuatro pueblos candidatos a los cuatro destinos. Fuente: Mapa de carreteras Michelin 576 Regional España.

Todos los pueblos que aparecen en la Tabla 1, exceptuando Argamasilla de Alba, estaban situados en la época de Cervantes en el Campo de Montiel. La razón de que hayamos incluido Argamasilla de Alba es sobre todo histórica y tiene como fin confirmar, mediante el análisis estadístico que se describe en la sección 2, la imposibilidad práctica de que Argamasilla fuera el lugar de la Mancha. No olvidemos que Rubén Darío y Azorín, con motivo del tercer centenario de la publicación de la primera parte del Quijote, *determinaron* que la cuna de don Alonso Quijano era Argamasilla de Alba.

Acerca de la velocidad de las cabalgaduras, aunque existen datos dentro del propio texto del Quijote que dan una idea de cual podría ser ésta, tal como se señala en el apartado f) de la introducción, (véase en Parra *et al.* 2005), no se han utilizado en este trabajo por varios motivos.

En primer lugar, la velocidad depende no solamente de los destinos sino también, al menos en principio, de los pueblos de origen y puede ser bastante variable. En segundo lugar, y este es el motivo principal para no incluirla como *parámetro conocido* es que en el modelo estadístico que proponemos en la sección siguiente para determinar el lugar de la Mancha —que incluye como parámetros

subsidiarios las diferentes velocidades— no es necesario conocer éstas.

No obstante, como subproducto del modelo estadístico considerado, cuya finalidad primordial es *estimar probabilísticamente* el lugar de la Mancha, también se puede dar una estimación de la cota inferior de las velocidades de las cabalgaduras.

## 2 Selección y análisis de los modelos estadísticos propuestos

El enfoque que proponemos es totalmente general en lo que se refiere al número de pueblos y de destinos.

Supongamos, pues, que hay  $m$  pueblos de origen a ser candidatos al lugar de la Mancha y  $k$  destinos. Los datos que se conocen son:

- 1.-  $\mathbf{d} = (d_1, \dots, d_j, \dots, d_k)$ , donde  $d_j$  representa la jornada empleada por las cabalgaduras desde el posible lugar de la Mancha hasta el destino  $j$ -ésimo.
- 2.-  $x_{ij}$  representa la distancia euclídea entre el  $i$ -ésimo pueblo de origen y el destino  $j$ -ésimo.

Los parámetros que interviene en el modelo

son:

- 3.-  $\theta=(\theta_1,\dots,\theta_j, \dots,\theta_k)$  punto del espacio paramétrico  $\Omega = \mathbb{R}^{+k}$ , donde  $\theta_j$  representa la distancia euclídea desde un punto genérico —no necesariamente un pueblo— cualquiera del mapa del Campo de Montiel y sus entornos al destino  $j$ -ésimo.
- 4.-  $v_{\theta_j}$  representa la *velocidad* de las cabalgaduras en kms/jornada, desde un pueblo genérico arbitrario  $\theta$  al destino  $j$ -ésimo.
- 5.- representa el *factor de inflación de la distancia* desde un pueblo genérico arbitrario  $\theta$  al destino  $j$ -ésimo. Representa el factor (mayor que 1) que hay que aplicar a la distancia euclídea  $\theta_j$  para obtener la distancia real entre el pueblo  $\theta$  y el destino  $j$ -ésimo.
- 6.-  $\sigma^2$  es un factor de variabilidad.

La ecuación básica que han de cumplir los datos y los parámetros es la siguiente, que nos dice que la distancia real recorrida entre el pueblo  $i$ -ésimo y el destino  $j$ -ésimo es aproximadamente igual al producto de la velocidad por la jornada correspondiente, es decir:

$$\gamma_{\theta_j} x_{ij} \approx v_{\theta_j} d_j \text{ y } \gamma_{\theta_j} \theta_j \approx v_{\theta_j} d_j$$

para todo  $i=1,\dots,m; j=1,\dots,k$  y  $\theta \in \Theta$ . (3.1)

Tenemos, por consiguiente, una única observación  $d$  que depende de los parámetros  $\theta$ ,  $v_{\theta_j}$ ,  $\gamma_{\theta_j}$  y  $\sigma^2$ , que mediría el grado de aproximación de la fórmula (3.1). El modelo estadístico que proponemos, basado en esta fórmula, es el siguiente: Para  $j =1,\dots,k$  se tiene

$$d_j | \theta_j, v_{\theta_j}, \gamma_{\theta_j}, \sigma^2 \sim N(d_j | \theta_j / v_{\theta_j} \theta_j, \sigma^2 \theta_j^2). \quad (3.2)$$

La justificación del modelo se basa en que la hipótesis sobre la esperanza de  $d_j$  se corresponde con la ecuación (3.1), mientras que la varianza de la observación  $d_j$  depende del factor de varianza  $\sigma^2$  y del parámetro  $\theta_j$ , de

tal forma que el coeficiente de variación no depende de éste último. Es decir, estamos suponiendo que a las distancias más largas corresponde mayor variabilidad que a las más cortas.

La última hipótesis sobre el modelo, que representa que las salidas son independientes, es que para  $j =1,\dots,k$  las  $d_j$  son condicionalmente independientes, con lo cual el modelo (3.2) se puede escribir como una normal  $k$  variante

$$\mathbf{d} | \theta, \mathbf{M}_\theta, \sigma^2 \sim N_k(\mathbf{d} | \mathbf{M}_\theta \theta, \sigma^2 \mathbf{D}(\theta)), \quad (3.3)$$

donde  $\mathbf{M}_\theta$  es una matriz diagonal de dimensiones  $k \times k$ , en la que cada elemento de la diagonal principal es  $\gamma_{\theta_j} / v_{\theta_j}$  y  $\mathbf{D}(\theta)$  es una matriz diagonal de dimensiones  $k \times k$ , cuyos elementos son  $\theta_j^2$ .

El primer paso para obtener las probabilidades a posteriori de los  $m$  pueblos candidatos a ser el lugar de la Mancha es *estimar* los parámetros  $v_{\theta_j}$ ,  $\gamma_{\theta_j}$  y  $\sigma^2$  del modelo con los datos  $d_j$  y  $x_{ij}$ . La hipótesis de independencia, para todo  $i = 1,\dots,m$  y  $j = 1,\dots,k$ , es del todo plausible en nuestro caso, de modo que tenemos el modelo

$$d_j | x_{ij}, v_{ij}, \gamma_{ij}, \sigma^2 \sim N(d_j | \mu_{ij} x_{ij}, \sigma^2 x_{ij}^2), \quad (3.4)$$

donde  $\mu_{ij} = \gamma_{ij} / v_{ij}$ . Definiendo los nuevos datos  $y_{ij} = d_j / x_{ij}$ , el modelo anterior lo podemos escribir como

$$y_{ij} | \mu_{ij}, \sigma^2 \sim N(y_{ij} | \mu_{ij}, \sigma^2). \quad (3.5)$$

La primera conclusión que se extrae del modelo (3.5) es que los parámetros  $v_{ij}$  y  $\gamma_{ij}$  *no son identificables* -solamente lo son los  $\mu_{ij}$ -, de modo que no se pueden estimar por separado, a no ser que se tenga información a priori sobre ellos.

Como el número de parámetros de (3.5) es muy elevado,  $m \times k$ , establecemos las siguientes hipótesis simplificadoras sobre los  $\mu_{ij}$  que dan lugar a tres modelos posibles.

- i)  $\mu_{ij} = \mu$ , es decir, el cociente entre el factor de inflación y la velocidad es constante y no depende ni del origen ni del destino.
- ii)  $\mu_{ij} = \mu_j$ , es decir, el cociente entre el factor

de inflación y la velocidad solamente depende del destino  $j$ -ésimo.

iii)  $\mu_{ij} = \alpha_i + \beta_j$ , es decir, la contribución del origen  $i$ -ésimo y del destino  $j$ -ésimo al cociente entre el factor de inflación y la velocidad se supone separable y aditiva.

Los tres modelos para los datos  $y_{ij}$ , de menos a más complejo, se corresponden con un modelo normal ordinario, un ADEVA de una vía y un ADEVA de dos vías, respectivamente.

Los datos se analizaron para los tres modelos y tras eliminar las siete observaciones anómalas por pueblos, básicamente coincidentes para los tres modelos, y comprobar la homogeneidad del factor de varianza  $\sigma^2$  para los destinos, se aplicó la técnica bayesiana descrita en Girón, Moreno y Martínez (2005) para seleccionar el mejor de los tres modelos, que resultó ser el segundo, es decir

$$y_{ij} | \mu_j, \sigma^2 \sim N(y_{ij} | \mu_j, \sigma^2)$$

para  $i=1, \dots, m'$ ;  $j=1, \dots, k$ , (3.6)

donde  $m'$  es el número de pueblos una vez eliminados los anómalos que resultaron ser, por orden lexicográfico: Argamasilla de Alba, Castellar, Membrilla, Ossa, Ruidera, Solana y Torrenueva.

La selección del segundo modelo como el que mejor describe los datos es muy ilustrativa ya que básicamente implica que *la velocidad de las cabalgaduras y/o el factor de inflación solamente dependen del destino y no del pueblo de origen*; además, al rechazarse también el primer modelo, *la velocidad no se puede suponer constante para todos los destinos*, como reflejan las estimaciones.

El análisis bayesiano del problema arroja los siguientes resultados sobre las cotas inferiores de las velocidades a los cuatro destinos, representada por  $\underline{v}_j = 1/\mu_j$ , para  $j=1,2,3,4$ , que se obtienen haciendo  $\gamma_j=1$  para todo  $j$ .

Estas estimaciones, basadas en la moda  $\tilde{v}_j$ , y en la media armónica  $\hat{v}_j$  de sus correspondientes densidades a posteriori, expresadas en kilómetros por jornada, son, respectivamente:

$$\tilde{v}_1=36.83, \tilde{v}_2=28.33, \tilde{v}_3=35.83, \tilde{v}_4=31.01.$$

$$\hat{v}_1=36.97, \hat{v}_2=28.40, \hat{v}_3=35.96, \hat{v}_4=31.09.$$

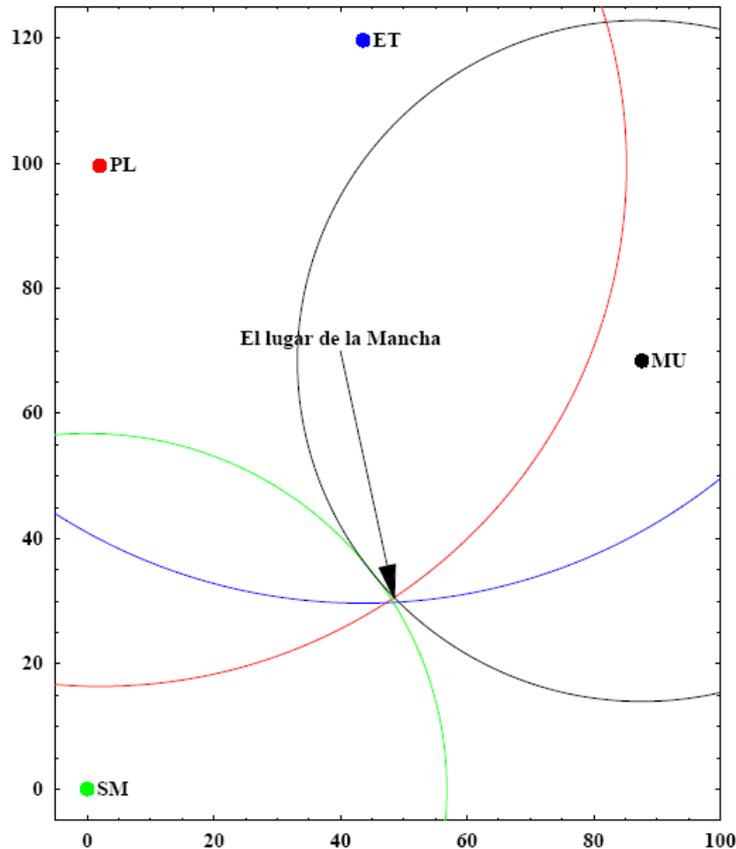
Aunque no debe sorprendernos, llama particularmente la atención el hecho de que la cota inferior de la velocidad a Sierra Morena sea menor que al resto de los destinos, probablemente debido a la mayor dificultad del camino. Además, estas estimaciones son consistentes con la velocidad que Cervantes sugiere implícitamente en la primera parte del Quijote.

La información sobre el parámetro  $\theta$  vendría dada por su distribución a posteriori que, por cierto, no es una distribución estándar a la que hemos denominado  $t$ -multivariante invertida. La moda  $\tilde{\theta}$  y la media armónica  $\hat{\theta}_h$  de esta distribución son, respectivamente

$$\tilde{\theta} = (78.18, 54.68, 84.75, 52.01),$$

$$\hat{\theta}_h = (83.19, 56.80, 89.90, 54.40).$$

Estas estimaciones, que representarían la distancia desde el *punto o lugar ideal* a los cuatro destinos, sirven para localizar geoméricamente el lugar de la Mancha, tal como se ilustra en la Figura 1.



**Figura 1.** Representación geométrica de la estimación del *lugar de la Mancha* usando la media armónica como estimador de los parámetros  $\theta_j$ , que son los radios de sus respectivos círculos. Las coordenadas del lugar, de acuerdo con el convenio de tomar como origen de coordenadas *Sierra Morena* y las distancias expresadas en kms, son aproximadamente (48.6,30.4).

Finalmente, a partir de esta distribución también se pueden calcular las probabilidades a posteriori de cada uno de los  $m$  pueblos candidatos a ser el lugar de la Mancha, que se presentan en la Tabla 2, en la que observamos que Villanueva de los Infantes es el pueblo más probable seguido, muy de cerca, por Fuenllana.

#### 4 Conclusiones y posibles extensiones del modelo estadístico

Aunque el análisis bayesiano de datos positivos con modelos normales, como hemos hecho con los datos del Quijote, tiene la ventaja de permitir obtener resultados analíticos, un análisis más ajustado a la realidad emplearía modelos estadísticos para los datos con soporte positivo como, p. ej., los modelos gamma o gamma-invertido. Hemos realizado un análisis análogo al aquí presentado con estos modelos, que no

incluimos por razones de espacio y por su mayor complejidad —ya que este análisis requiere técnicas bayesianas basadas en métodos MCMC—, que conduce a resultados similares a los obtenidos con los modelos normales.

Otro aspecto de gran interés que ofrecen estos datos es el elevado número de observaciones anómalas —siete pueblos de un total de veinticuatro que están concentrados en una área geográfica relativamente pequeña— que se debe fundamentalmente a que *los datos* no se han obtenido a partir de un diseño estadístico previamente planificado. Obviamente, como revela la Tabla 2, estos datos se identifican con los pueblos más improbables de ser el lugar de la Mancha.

Pueblos	Probabilidades	Pueblos	Probabilidades
Villa Infantes	0.15992	T. J. Abad	0.00747
Fuencollana	0.15433	Villamanrique	0.00530
S. C. Caña.	0.12063	Villa Fuente	0.00344
Montiel	0.11323	Alhambra	0.00325
Terrinches	0.08799	Cañamares	0.00313
Almedina	0.08632	Solana	0.00001
Albaladejo	0.07385	Ruidera	0.00000
Villahermosa	0.06554	Membrilla	0.00000
Puebla P.	0.03758	Torrenueva	0.00000
Alcubillas	0.03304	Argamasilla	0.00000
Carrizosa	0.02982	Ossa	0.00000
Cozar	0.01515	Castellar	0.00000

**Tabla 2.** Probabilidades a posteriori de cada uno de los 24 pueblos candidatos a ser el lugar de la Mancha, ordenadas en orden decreciente. Las probabilidades menores que 0.00001 se han redondeado a 0.

Por último, la inclusión de incertidumbre en la duración de las jornadas  $d$  y en las distancias reales entre los pueblos de origen y los destinos (mayores, por definición, que las  $x_{ij}$ ), así como la incorporación de información a priori sobre los factores de inflación y las velocidades se deja, como en el Quijote, para una segunda parte..., aunque esperamos no tardar diez años en escribirla, como le sucedió a Cervantes.

## 5 Referencias

Girón, F. J., Moreno, E. and Martínez, M. L. (2005). *An objective Bayesian procedure for*

*variable selection in regression*. In *Advances on distribution theory, order statistics and inference*. Eds. N. Balakrishnan et al., Birkhauser Boston.

Parra, F. et al. (2005). *El lugar de la Mancha es...: El Quijote como un sistema de distancias/tiempos*. Editorial Complutense Madrid.

Ríos, M. J., Montero, F. J. y Parra, F. (2005). *Encontrando el «lugar de la Mancha» con las Matemáticas*. XII, JAEM: Albacete.

## LA ESTADÍSTICA, UN INSTRUMENTO PARA EL PROGRESO DE LA SOCIEDAD

Carles M. Cuadras  
Departamento de Estadística  
Universidad de Barcelona

### 1. INTRODUCCIÓN

La Estadística tiene cada vez más influencia en la sociedad. En los periódicos aparecen diariamente resultados estadísticos sobre

economía, salud, opinión, política. ¿Qué hay de cierto en una estadística? Podemos desvirtuar la verdad con estadísticas, pero podemos especular más sin estadísticas. La