

# **Regresión Avanzada**

## Calidad del Aire en el Valle de México

Carlos Pérez & Juan Zinser

ITAM

5 de Diciembre, 2016

# Índice

- ① Introducción
- ② Descripción de los datos
- ③ Modelado
- ④ Interpretación
- ⑤ Referencias

# Introducción

- El nivel de bienestar de la población depende de la calidad de vida, en particular las condiciones ambientales.
- La calidad del aire es un tema primordial para la salud de los habitantes. La mayoría de los contaminantes atmosféricos son producto de actividades humanas.
- Las más contaminadas: Monterrey, Toluca, León, Ciudad de México. En segundo plano Guadalajara, Silao, Irapuato, Puebla y Salamanca
- PGR: 12,000 muertes por hechos criminales en 2015.
- OMS: 14.000 muertes causadas por cáncer de pulmón, enfermedades cardiopulmonares e infecciones respiratorias asociadas a contaminantes en el aire en 2015. <sup>1</sup>

---

<sup>1</sup>Fuente: Excélsior (Abril 2016)

# Datos - Variable Dependiente

## DISEASES DUE TO:

● OZONE (O<sub>3</sub>)

● PM2.5 AIR POLLUTION



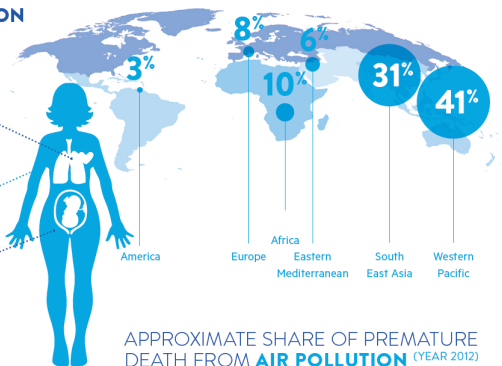
- Chronic obstructive pulmonary disease (COPD)
- Childhood pneumonia
- Ischaemic heart disease
- Stroke



- Asthma
- Breathing problems
- airway inflammation
- Chronic respiratory illness
- Reduced lung function



- Low birth weight




# Introducción

- En la Ciudad de México cada vez más frecuente se activa la alarma de Contingencia Ambiental (Mayo, Junio, Julio, Agosto).
- Aproximadamente el 50% de las partículas contaminantes son producto del tránsito vehicular. Se dividen en PM10 y PM2.5 por su masa en micrómetros.
- El objetivo de este trabajo es cuantificar el efecto del padrón vehicular sobre la calidad del aire.
- Se analizará la relación entre las partículas PM2.5 y el conteo de vehículos que transitan en la Ciudad de México.

## Physiology of PM<sub>2.5</sub> Inhalation

**How Particulate Matter Enters Our Body**



- 1** Particulate matter enters our respiratory (lung) system through the nose and throat.
- 2/3** The larger particulate matter (PM<sub>10</sub>) is eliminated through coughing, sneezing and swallowing.
- 4** PM<sub>2.5</sub> can penetrate deep into the lungs. It can travel all the way to the alveoli, causing lung and heart problems, and delivering harmful chemicals to the blood system.

<http://www.bcairquality.ca/health/air-quality-and-health.html>

# Datos - Variables Independientes

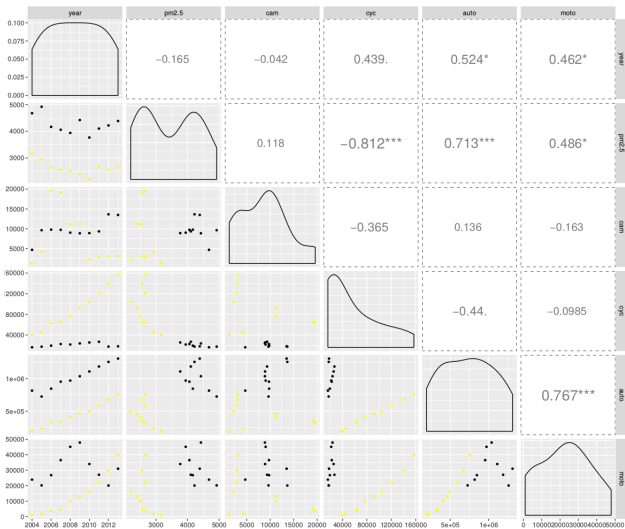
- **Variables independientes:**
- Datos de vehículos de motor registrados en circulación. A nivel municipal y por año entre 2004 y 2013  
<http://www.inegi.org.mx/est/contenidos/proyectos/registros/economicas/vehiculos/default.aspx>
- Variables:  $x_{auto}$ ,  $x_{cam}$ ,  $x_{cyc}$  y  $x_{moto}$ , las cuales representan los conteos de autos, camiones de pasajeros, camiones de carga y motocicletas. Incluyen categorías particular, oficial y público.

## Datos - Variable Dependiente

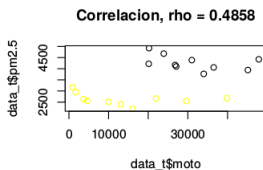
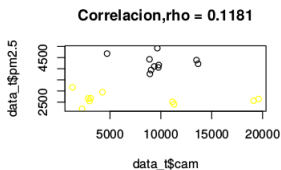
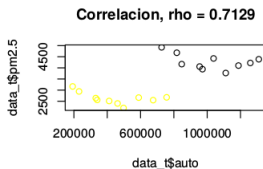
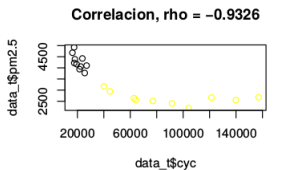
- **Variable Dependiente:**
- Se considerará como *pm2.5* a la concentración de partículas suspendidas de menos de 2.5 micrómetros.
- La información proviene de la Red Manual de Monitoreo Atmosférico (RAMA). Representa el promedio diario del contaminante en el aire desde 2004 hasta 2013 a nivel estación de monitoreo (lat,lon) y fue georreferenciada a nivel municipio y agregada a nivel anual.
- Las partículas son medidas a través de dos métodos en cada estación gravimetría y radiación beta. Asimismo puede ser descargada del sitio <http://www.aire.cdmx.gob.mx/>



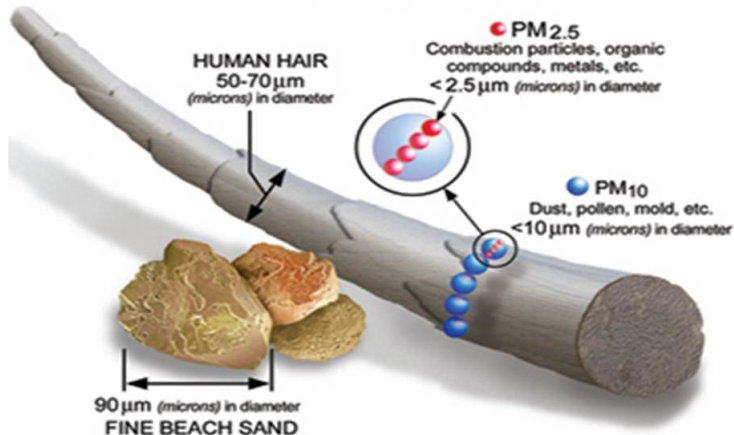
# Datos - Exploración



# Datos - Exploración



## Relative Size of Particulate Matter



# Modelado

Se desea medir el efecto que tiene el total de vehículos motorizados de distintas categoría sobre el contaminante PM2.5. En otras palabras, cuantificar el efecto del padrón vehicular sobre la calidad del aire en la Ciudad de México.

$$pm2.5_t = y_t = f(\mathbf{x}_t|\theta) = f(auto_t, cyc_t, cam_t, moto_t|\theta_t)$$

En la cual  $\theta_t$  está compuesto por

- $\beta_{t,auto}, \beta_{t,cyc}, \beta_{t,cam}, \beta_{t,moto},$
- $\beta_{t,auto}^{cdmx}, \beta_{t,cyc}^{cdmx}, \beta_{t,cam}^{cdmx}, \beta_{t,moto}^{cdmx},$
- $\alpha^{cdmx}, \alpha$

## Distribución de Observable

$$\log(y_t) \sim \Gamma(\tau * \mu_i, \tau)$$

$$\mu_i = \exp(\eta_i)$$

$$\begin{aligned} \eta_i = \alpha + & \sum_{i \in \{auto, cyc, cam, moto\}} \beta_{it} \log(x_{it}) + \alpha^{cdmx} x_{cdmx} \\ & + \sum_{i \in \{auto, cyc, cam, moto\}} \beta_{it}^{cdmx} \log(x_{it}) * x_{cdmx} \end{aligned}$$

# Modelado - Dinámica

## Distribuciones de Estado

$$\beta_{i,t}^{cdmx} \sim N(\beta_{i,t-1}^{cdmx}, \tau_i^{cdmx}); \quad i \in \{auto, cyc, cam, moto\}; \quad 2 \leq t \leq 10$$

$$\beta_{i,t} \sim N(\beta_{i,t-1}, \tau_i); \quad i \in \{auto, cyc, cam, moto\}; \quad 2 \leq t \leq 10$$

## Distribuciones Iniciales

$$\tau \sim \Gamma(0.001, 0.001); \quad i \in \{auto, cyc, cam, moto\}$$

$$\beta_{i,1} \sim N(0, 0.001); \quad i \in \{auto, cyc, cam, moto\}$$

$$\beta_{i,1}^{cdmx} \sim N(0, 0.001); \quad i \in \{auto, cyc, cam, moto\}$$

$$\alpha \sim N(0, \tau_a)$$

$$\alpha^{cdmx} \sim N(0, \tau_{cdmx})$$

## Parámetros de Suavizamiento

$$\lambda = 15$$

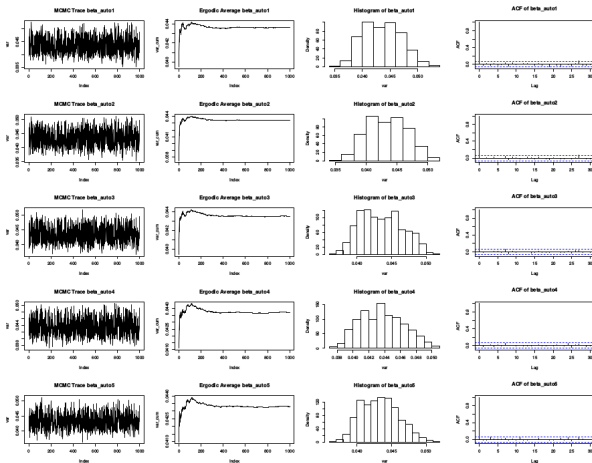
$$\tau_i = \lambda * \tau; \quad i \in \{auto, cyc, cam, moto\}$$

$$\tau_a = \lambda * \tau$$

$$\tau_{cdmx} = \lambda * \tau$$

$$\tau_i^{cdmx} = \lambda * \tau$$

# Resultados - Convergencia



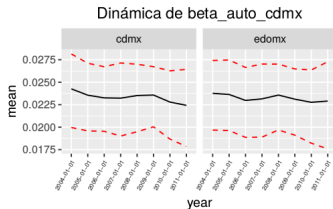
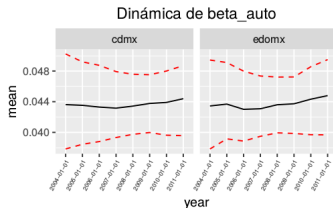
**Figure:** beta auto

Convergencia MCMC mediante análisis gráfico



# Resultados - Dinámica de betas

## Efectos positivos y decrecientes



# Resultados - Coeficientes

## Efectos CDMX y Efectos EDOMX

$$\frac{\beta_{auto,t} + \beta_{auto,t}^{cdmx}}{x_{auto,t}} = 0.0048387$$

$$\frac{\beta_{cyc,t} + \beta_{cyc,t}^{cdmx}}{x_{cyc,t}} = 0.0174772$$

$$\frac{\beta_{moto,t} + \beta_{moto,t}^{cdmx}}{x_{moto,t}} = -0.0091395$$

$$\frac{\beta_{cam,t} + \beta_{cam,t}^{cdmx}}{x_{cam,t}} = 0.0010684$$

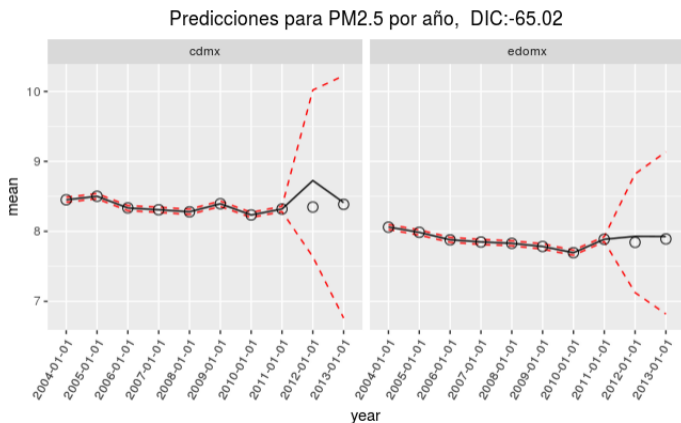
$$\frac{\beta_{auto,t}}{x_{auto,t}} = 0.0033559$$

$$\frac{\beta_{cyc,t}}{x_{cyc,t}} = 0.0185979$$

$$\frac{\beta_{moto,t}}{x_{moto,t}} = -0.009259$$

$$\frac{\beta_{cam,t}}{x_{cam,t}} = -0.004819$$

# Resultado - Predicción



**Figure:** Predicción

# Conclusión

- Modelos Bayesianos tienen mucho poder para no linealidades y escasez de datos, aunque es tardado el mejoramiento de los mismos.
- Modelos complejos permiten probar hipótesis que revelan información sobre los procesos.
- El rubro de autos y camiones de carga son los dos sectores más importantes que deben ser analizados para la actuar contra la contaminación atmosférica.
- Es necesario realizar evaluaciones con más y mejores datos para comprobar la validación de resultados.

# Referencia

- <http://www.aire.df.gob.mx>
- <http://www.inegi.org.mx/est/contenidos/proyectos/registros>
- Nieto-Barajas, L. E. de Alba, E. (2014). Bayesian regression models. En Predictive Modeling Applications in Actuarial Science. E.W. Frees, R.A. Derrig G. Meyers (eds.) Cambridge University Press, pp 334-366.
- <https://www.sharelatex.com>