

Estadística bayesiana y su naturaleza continúa en la estimación de niveles contaminantes por material particulado pm_{10} en control y monitoreo de calidad del aire en Trujillo (La Libertad).

Bayesian statistics and their continuous nature in the estimation of levels of particulate matter pm_{10} pollution control and air quality monitoring in Trujillo (La Libertad).

Elmis Jonatan García Zare¹

RESUMEN

El material particulado (PM_{10}) es uno de los principales contaminantes que deteriora la calidad de aire, convirtiéndose en factor negativo para la salud, motivando el monitoreo continuo y diversas investigaciones. En Trujillo (Perú) contamos con el Programa Nacional de Vigilancia Sanitaria de Calidad del Aire, dirigido por la Dirección General de Salud Ambiental, en la cual se siguen estándares anuales internacionales permisibles ($50\mu g/m^3$). El enfoque bayesiano determinó que los niveles promedios son variables respecto a las estimaciones promedio clásica. En los diversos puntos de monitoreo se determinó que los niveles promedio pueden ser mayores al hacer los cálculos mediante el enfoque bayesiano.

Palabras clave: estadística bayesiana, distribución normal, material particulado.

ABSTRACT

Particulate matter (PM_{10}) is one of the main pollutants that deteriorates air quality, becoming a negative factor for health, encouraging continuous monitoring, and research. In Trujillo (Peru) we have the National Health Surveillance of Air Quality, led by the Environmental Health Directorate, which keeps track of the permissible annual international standards ($50\mu g/m^3$). The Bayesian approach found that average levels are variable to the classical average estimate. In the various monitoring points has be found that average levels may be higher making the calculations using Bayesian statistics.

Keywords: Bayesian statistics, normal distribution, particulate matter.

INTRODUCCIÓN

El aire es un recurso indispensable para la vida sobre el planeta. El hombre utiliza la atmósfera como un receptáculo de muchos de los desechos que emite. Mezclas de gases y partículas son devueltas a la atmósfera como subproductos de los procesos de combustión y otras actividades propias de la transferencia de energía (Préndez, 2011). En las ciudades con industrias, miles de vehículos motorizados, acumulación de basuras y deficiente circulación de aire, se acumula en el aire una capa de gases y polvos muy concentrada denominada *smog*, siendo dañino para la salud y el medio ambiente.

Los contaminantes más relevantes en la actualidad son las partículas en suspensión en sus diferentes tamaños ($PM_{2.5}$ y PM_{10}), dióxido de nitrógeno (NO_2), monóxido de carbono (CO), dióxido de azufre (SO_2), entre otros.

¹ Universidad Nacional de Trujillo; jonatan_gz18@hotmail.com

Recibido el 14 de julio del 2012

Aceptado el 5 de setiembre del 2012

El Material particulado PM_{10} (E.I.R.L.) se define como la acumulación de gotitas de un sólido o líquido en la atmósfera ambiental generada a partir de alguna actividad antropogénica o natural (Thomas, 1996). En general, el particulado capaz de penetrar las vías respiratorias de los humanos, se divide en 2 rangos de tamaños: ($2.5 \mu\text{g}/\text{m}^3$ y $10 \mu\text{g}/\text{m}^3$). El $PM_{2.5}$ es responsable de causar los mayores daños a la salud de las personas, siendo hoy objeto de gran atención en los Estados Unidos. El material particulado PM_{10} , causa también daños a la salud, así cuando estas partículas son inhaladas, se depositan en la parte superior del sistema respiratorio y son eventualmente expulsadas a través de la garganta (Celis, 2004). Existe evidencia de que el material particulado aumenta el riesgo por muertes y produce el desarrollo de enfermedades respiratorias y cardiovasculares en las personas (Harrison, 2000). Estudios al respecto se reportan en Santiago de Chile, presentándose una metodología para estimar los impactos en salud que se pueden asociar a un programa que reduce las concentraciones de PM_{10} en esta ciudad (Sánchez et al. 1998). De igual modo el objetivo de modelar las relaciones entre la contaminación del aire por material particulado y ozono, y la mortalidad diaria en la ciudad de Rancagua y Codegua, Chile (Sanhueza, 2006). Estudios más específicos sobre enfermedades relacionadas se reportaron también en artículos como: Efecto del tiempo de exposición a PM_{10} en las urgencias por bronquitis aguda (Muñoz, 2009). Muchas otras investigaciones hacen suponer de la importancia de una buena medición de los niveles contaminantes de aire. Estas mediciones se obtienen mediante diversas técnicas e instrumentos especializados, siguiendo especificaciones, requisitos, normas y estándares de calidad que permitan la precisión de los monitoreos (DIGESA, 2011).

En Trujillo, (y demás departamentos del Perú), se lleva a cabo el Programa Nacional de Vigilancia Sanitaria de Calidad del Aire. Sus reportes anuales se pueden seguir en la web oficial de la Dirección General de Salud Ambiental – DIGESA y ejecutado por la Dirección Ejecutiva de Salud Ambiental – DESA, en la cual observamos que consideran cuatro puntos de monitoreo, las cuales solo tres se encuentran operativos (DESA, Hospital Belén de Trujillo, Red Trujillo) durante los años 2006 al 2011. Estos reportes siguen lineamientos internacionales y niveles estándares de calidad ambiental (ECA) por material particulado, aprobado por decreto supremo (MINAM, 2001). En estos reglamentos se especifica que el nivel aceptable para PM_{10} es $50 \mu\text{g}/\text{m}^3$ (anual).

Los niveles de PM_{10} reportados desde el año 2006 al 2011, presentan variaciones casi “exageradas” e inexplicables, esto puede deberse a fallas en la calibración o mantenimiento de los aparatos, inclusive por los reducidos puntos de monitoreo, haciendo que las estimaciones puedan no ser tan representativas. Además, niveles por debajo de los estándares permitidos no precisamente indican seguridad en cuanto a efectos sobre la salud (Aldunate, 2005). Por ello es necesario plantear otras opciones para inferir parámetros de medición en función de los datos ya disponibles.

La estadística bayesiana, basada como su nombre indica en el teorema de Bayes, se diferencia de la estadística frecuentista básicamente en la incorporación de información externa al estudio que se esté realizando, de manera que, según la formulación de éste teorema, si conocemos la probabilidad de que ocurra un suceso, su valor será modificado cuando dispongamos de esa información. Así pues, las fuentes de información “a priori” se ven transformadas en probabilidad “a posteriori” y se utilizan a continuación para realizar la inferencia (Bolstad, 2007). Consecuentemente, los reportes de años pasados, pueden servir para actualizar los niveles promedios de material particulado del año que se desee hacer su estimación. De esta manera las estimaciones anuales, podrían variar según comportamientos pasados.

El propósito de este trabajo es determinar mejores estimaciones respecto al promedio anual de los niveles de material particulado, utilizando provechosamente información de años pasados, basándonos en la inferencia bayesiana, mencionado líneas anteriores. Así también, generar un antecedente de la utilidad de la inferencia bayesiana para el estudio y monitoreo llevado a cabo por el Programa Nacional de Vigilancia Sanitaria de Calidad del Aire.

MATERIAL Y MÉTODOS

En el trabajo, se realizó la evaluación de los datos reportados por la Dirección General de Salud Ambiental – DIGESA, cuyos puntos de monitoreo fueron establecidos por el Programa Nacional de Vigilancia Sanitaria de Calidad del Aire. Los datos obtenidos fueron en periodo mensual, anual y en $\mu\text{g}/\text{m}^3$ como unidad de medida desde el año 2006 al 2011.

Para el estudio se tomaron en cuenta 3 puntos de monitoreo, de los 4 que inicialmente se tuvieron en cuenta: Dirección Ejecutiva de Salud Ambiental – DESA, Hospital Belén de Trujillo y Red Trujillo.

Los reportes del año 2006 fueron considerados como año base de información a priori para las estimaciones consecutivas de los parámetros a posteriori de los años siguientes, tal que, al estimar los parámetros posteriori del año 2007, éste se tomara como información a priori para las estimaciones posteriori del siguiente año, y así sucesivamente, asumiendo de esta manera la naturaleza continua del enfoque bayesiano en las estimaciones de niveles contaminantes de material particulado PM_{10} .

El análisis estadístico se realizó empleando el software R para poder estimar los parámetros mediante el proceso de inferencia bayesiana.

Se consideró trabajar bajo la comprobación de que los datos siguen una distribución normal, dada la naturaleza continua de la variable respuesta ($\mu\text{g}/\text{m}^3$). Esta distribución fue considerada tanto para los datos a priori, verosimilitud y consecuentemente las estimaciones a posteriori. La comprobación de la distribución se hizo mediante el test de Shapiro-Wilks, y no se presentaron problemas de normalidad.

Función de distribución de normalidad.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Proceso de inferencia bayesiana:

$$g(\mu / y_1, \dots, y_n) = \frac{g(\mu) \times f(y_1, \dots, y_n / \mu)}{\int g(\mu) \times f(y_1, \dots, y_n / \mu) d\mu}$$

$$g(\mu) \times f(y / \mu) \propto e^{-\frac{1}{2\sigma^2 s^2 / (\sigma^2 + s^2)} \left[\mu - \frac{(\sigma^2 m + s^2 y)}{\sigma^2 + s^2} \right]^2}$$

Donde:

$g(\mu / y)$: Distribución posteriori

$g(\mu)$: Distribución a priori

$f(y / \mu)$: Verosimilitud

Parámetros a posteriori:

Media a posteriori:

Desviación estándar a posteriori:

$$m' = \frac{(1/s^2)}{(n/\sigma^2 + 1/s^2)} m + \frac{(n/\sigma^2)}{(n/\sigma^2 + 1/s^2)} \cdot y$$

$$\frac{1}{(s')^2} = \frac{1}{s^2} + \frac{n}{\sigma^2}$$

Intervalos de credibilidad:

$$m' \pm z_{\alpha/2} \times s'$$

Los parámetros a posteriori se calculan con la función normnp() que se encuentra dentro de la librería: library (Bolstad), contenedor de diversas funciones de cálculo estadístico bayesiano.

DETERMINACION DEL AJUSTE DE DISTRIBUCION: TEST SHAPIRO-WILKS.

Sea la hipótesis:

Ho: Los datos se ajustan una Normal (μ, σ^2).

Ha: Los datos NO se ajustan a una Normal (μ, σ^2).

Considérese el estadístico:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$x_{(i)}$: número que ocupa la i-esima posición de la muestra
 \bar{x} : es la media muestral
 a_i : constantes

donde: (a_1, \dots, a_n) se calcula

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

Donde:

$$m: (m_1, \dots, m_n)^T$$

siendo (m_1, \dots, m_n) son los valores medios del estadístico ordenado, de variables aleatorias independientes e idénticamente distribuidas, muestreadas de distribuciones normales. V es la matriz de covarianzas de ese estadístico de orden.

La hipótesis nula se rechaza con p-valor < 0.05 .

RESULTADOS

Tabla 1. Niveles promedio de PM₁₀ de los años 2006 al 2011 mediante estimaciones clásicas y bayesianas.

Año	Estimaciones clásicas				Estimaciones bayesianas			
	Media	Desviación estándar	Límite inferior	Límite superior	Media	Desviación estándar	Límite inferior	Límite superior
2006	64,035	15,055	55,516	72,553	-	-	-	-
2007	57,313	7,714	52,948	61,677	57,469	2,201	53,156	61,782
2008	47,521	9,010	42,423	52,619	53,319	1,680	50,026	56,612
2009	46,639	11,470	40,149	53,129	51,951	1,498	49,015	54,888
2010	52,472	16,372	43,209	61,735	51,999	1,428	49,200	54,798
2011	66,310	9,767	60,783	71,836	53,862	1,332	51,251	56,472

Fuente: Dirección Ejecutiva de Salud Ambiental y del autor.

Tabla 2. Niveles promedio de PM₁₀ de los años 2006 al 2011 por zonas de monitoreo mediante estimaciones clásicas y bayesianas.

Año	Media	Estimaciones clásicas			Estimaciones bayesianas			
		Desviación estándar	Límite inferior	Límite superior	Media	Desviación estándar	Límite inferior	Límite superior
Zona 1: Red Trujillo								
2006	76,000	21,477	63,848	88,152	-	-	-	-
2007	74,417	20,183	62,997	85,836	74,525	5,623	63,504	85,546
2008	69,100	12,413	62,077	76,123	70,667	3,022	64,744	76,590
2009	41,273	18,479	30,817	51,728	63,526	2,629	58,372	68,679
2010	76,200	12,059	69,377	83,023	68,130	2,098	64,017	72,242
2011	82,571	38,557	60,756	104,387	68,423	2,077	64,353	72,493
Zona 2: DESA								
2006	58,083	18,889	47,396	68,771	-	-	-	-
2007	46,083	10,858	39,940	52,227	46,405	3,092	40,344	52,466
2008	26,700	10,028	21,026	32,374	35,903	2,113	31,761	40,045
2009	38,364	8,626	33,483	43,244	36,933	1,611	33,775	40,091
2010	44,000	9,391	38,687	49,313	38,778	1,385	36,063	41,493
2011	57,333	26,964	42,077	72,590	38,930	0,996	36,978	40,882
Zona 3: Hospital Belén								
2006	58,000	29,883	41,092	74,908	-	-	-	-
2007	48,000	13,611	40,299	55,701	48,170	3,896	40,534	55,806
2008	46,600	14,316	38,500	54,700	47,431	2,835	41,875	52,987
2009	59,636	20,142	48,240	71,033	49,775	2,548	44,781	54,769
2010	54,700	21,123	42,748	66,652	50,507	2,351	45,899	55,115
2011	59,000	16,199	49,835	68,165	51,460	2,215	47,118	55,802

Fuente: Dirección Ejecutiva de Salud Ambiental y del autor.

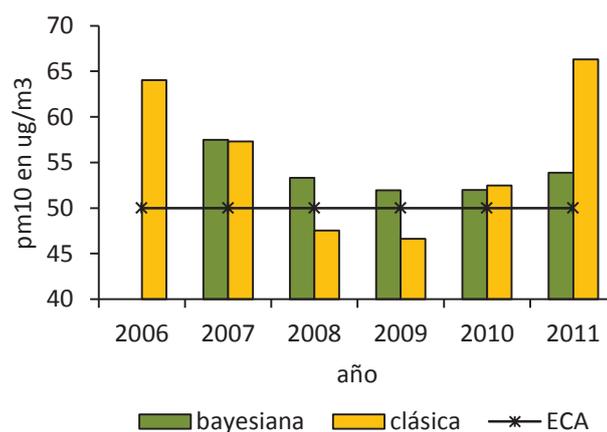


Figura 1. Comparación puntual de las estimaciones de PM₁₀ siguiendo el enfoque estadístico clásico y bayesiano.

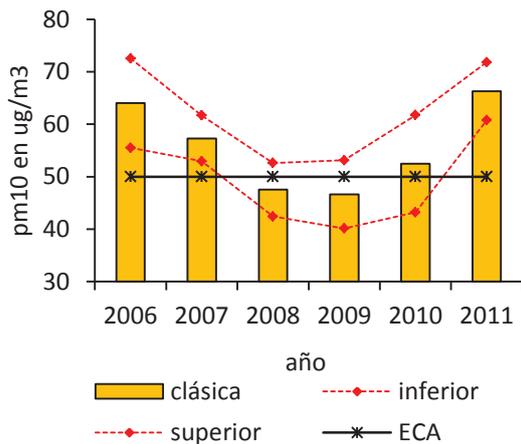


Figura 2: Estimación puntual e intervalos de confianza en los niveles promedio de PM₁₀ bajo el enfoque clásico.

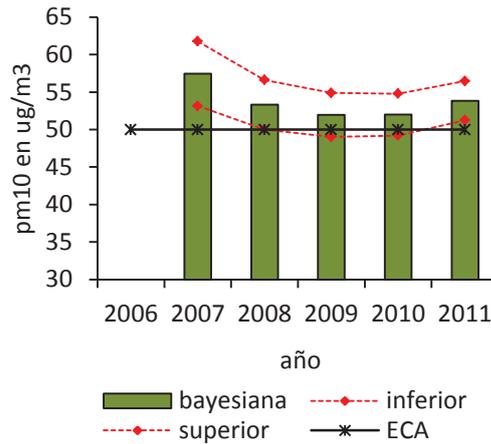


Figura 3: Estimación puntual e intervalos de confianza en los niveles promedio de PM₁₀ bajo el enfoque bayesiano.

DISCUSIÓN

Estimaciones anuales (2006 – 2011).

Las estimaciones anuales de los niveles de contaminación de aire por material particulado PM₁₀ fueron calculadas de la manera clásica y bajo el enfoque bayesiano, con sus respectivos intervalos al 95% de confianza (Tabla 1). Con el método clásico, en los años 2008 y 2009, los niveles promedio están por debajo del nivel anual permisible (50ug/m³), reportan 47,521ug/m³ y 46.639ug/m³ respectivamente. En el resto de años, los niveles son superiores al 50ug/m³. Otra característica es que los intervalos de confianza son amplios, “poco prácticos” para las interpretaciones, esto debido a las desviaciones estándar de cada año, por ejemplo, en el 2008 puede interpretarse que el nivel promedio de PM₁₀ está estimado entre 42.423ug/m³ y 52.619ug/m³ con un 95% de confianza, lo cual nos hace pensar que nos encontramos, bajo una probabilidad de que estemos en un nivel aceptable pero a la vez excesivamente inaceptable de contaminación. Según las estimaciones bayesianas, estas tienden a “estabilizarse” por efecto de la continuidad en el proceso de estimación. El año 2006 se consideró como año base para información a priori y actualizar (a posteriori) la información del año 2007, consecuentemente los parámetros bayesianos del 2007 pasaron a ser a priori para las estimaciones a posteriori del 2008, y así sucesivamente, adoptando de esta manera la naturaleza continua del enfoque bayesiano. En tal sentido, la estimación bayesiana del año 2008 se actualizó mostrando 53.319ug/m³ de PM₁₀ y con intervalos desde 50.026ug/m³ hasta 56.612ug/m³. Las estimaciones anuales restantes también presentaron variaciones con tendencias idénticas en todos los casos. Además la desviación estándar de cada año se redujo y por ende los límites de confianza (que en el caso bayesiano se denominan *limites de credibilidad*) redujeron su amplitud, haciendo mas factible su interpretación. Cabe resaltar que por procedimiento, el año 2006 no tiene estimación bayesiana. Se puede apreciar en la Figura 1, como es que ambos enfoques difieren en las estimaciones, aun así la tendencia de disminución y aumento en el tiempo se mantiene. La diferencia más importante está en que el enfoque clásico sostiene que en el 2008 y 2009 los niveles promedio de contaminación se encuentran “bajo control”; mientras que el enfoque bayesiano, estos niveles promedio se encuentran por encima de 50ug/m³. Si observamos los intervalos de confianza marcados en la

Figura 2, hallaríamos cierta ambigüedad en interpretarlos, puesto que estos intervalos son amplios y poco prácticos en un proceso de estimación, mientras que bajo el enfoque bayesiano la variabilidad en las estimaciones es más estable y estrecha.

Estimaciones por puntos de monitoreo (Zonas).

Como en el caso anual, las estimaciones por cada zona de monitoreo también presenta variaciones en las estimaciones puntuales e intervalos en ambos enfoques (Tabla 2). Para la Zona 1 (Red Trujillo) en el año 2009 el nivel promedio de PM_{10} reporta $41.273\mu\text{g}/\text{m}^3$, pero al observar su intervalo de confianza, resulta contradictorio pensar que esa estimación oscila (con un 95% de confianza) entre $30.817\mu\text{g}/\text{m}^3$ y $51.728\mu\text{g}/\text{m}^3$. Estos intervalos, en su mayoría son muy amplios. En la zona 2 (DESA), en el año 2011 el nivel promedio de PM_{10} es $57.333\mu\text{g}/\text{m}^3$, y su intervalo de confianza, está entre $42.077\mu\text{g}/\text{m}^3$ y $72.590\mu\text{g}/\text{m}^3$. En la zona 3 (Hospital Belén de Trujillo) en el año 2011 el nivel promedio de PM_{10} reporta $59\mu\text{g}/\text{m}^3$ con un intervalo de confianza desde $49.835\mu\text{g}/\text{m}^3$ a $68.165\mu\text{g}/\text{m}^3$. Para calcular los estimadores del año 2006 bajo en enfoque bayesiano es necesario contar con información a priori a esta. Al emplear el enfoque bayesiano, las estimaciones puntuales e intervalos se estabilizan por efecto de la continuidad en el proceso de estimación bayesiano, y los intervalos de credibilidad son más prácticos. Por ejemplo en el año 2009 en la zona 1: Red Trujillo, la estimación puntual bayesiana indica que el nivel promedio de PM_{10} es $63.526\mu\text{g}/\text{m}^3$ y su intervalo de credibilidad está entre $58.372\mu\text{g}/\text{m}^3$ y $68.679\mu\text{g}/\text{m}^3$, lo cual es más factible que la estimación clásica de $41.273\mu\text{g}/\text{m}^3$ con intervalo de confianza de $30.817\mu\text{g}/\text{m}^3$ y $51.728\mu\text{g}/\text{m}^3$. Además el enfoque bayesiano, advierte que el nivel promedio esta por encima del estándar permisible, lo cual sucede en muchas de las otras estimaciones de las 3 zonas de estudio en cada año. En la zona 2: DESA, en el año 2011 el estimador bayesiano, indica que el nivel promedio es de $38.930\mu\text{g}/\text{m}^3$ y su intervalo de credibilidad muestra oscilar entre $36.978\mu\text{g}/\text{m}^3$ y $40.882\mu\text{g}/\text{m}^3$.

CONCLUSIONES

La estadística bayesiana permite estimar con similitud a la estadística clásica, los niveles de material particulado y los intervalos de credibilidad hacen más factible estas estimaciones. Las variaciones que permite observar el enfoque bayesiano da pie a reconsiderar sobre los reales niveles de contaminación por material particulado, siendo esto muy importante referente a la influencia negativa que tiene en la salud pública.

REFERENCIAS

- Aldunate, P. (2005). *Los efectos de la contaminación atmosférica por PM_{10} sobre la salud en la ciudad de la Paz, Bolivia*. Disponible en http://www.swisscontact.org.pe/PRAL/15_pablo_aldunate.pdf
- Bolstad, W. (2007). *Introduction to Bayesian Statistics* (2ª Ed.). Canadá: Editorial Wiley.
- Celis, H. (2004). *Aspectos Generales de la contaminación atmosférica por material particulado*. Disponible en <http://www.ciencia-ahora.cl/Revista14/PM10.pdf>
- DIGESA, (2011). Dirección General de Salud Ambiental, *Protocolo de Monitoreo de la calidad del aire y gestión de los datos*. Lima, Perú.
- E.I.R.L. *Análisis Efectos en Salud por Material Particulado Respirable (PM_{10}) y Ozono (O_3) en la VI Región*.

Disponible en

<http://sinca.conama.cl/uploads/documentos/eeee9695e3c6b8ccdb57fcb7eb94c794.pdf>

Harrison, R. y Yin, J. (2000). Particulate matter in the atmosphere: Which particle properties are important for its effects on health? *Sci. Total Environ.*

MINAM. (2001). Decreto Supremo N° 074-2001-PCM (22/Jun/01). *Reglamento de Estándares Nacionales de Calidad Ambiental del Aire*. Disponible en http://www.minam.gob.pe/dmdocuments/d.s_074-2001-pcm_eca_para_aire.pdf

Muñoz, F. (2009). Efecto del tiempo de exposición a PM10 en las urgencias por bronquitis aguda. *The Scientific Electronic Library Online*. Disponible en <http://www.scielo.br/pdf/csp/v25n3/08.pdf>

Préndez, M. (2011). *La contaminación atmosférica de Santiago*. Disponible en <http://entornoperu.tripod.com/atmosferica.htm>

Sánchez, M., Sebastian, V. y Ostro, B. (1998). Los Efectos en salud de la contaminación atmosférica por PM10 en Santiago. *Centro de Estudios Públicos de Chile*. Disponible en www.cepchile.cl/dms/archivo_1632_364/rev69_sanchez.pdf

Sanhueza, P. (2006). Asesorías en Ingeniería Ambiental. Santiago de Chile, Chile.

Thomas, G. y William, M. (1996). *Química Medioambiental* (2ª Ed.). Madrid, España: Pearson Prentice Hall.