

EL MÉTODO DE BAYES APLICADO EN LA IDENTIFICACIÓN DE LAS LOCALIZACIONES MÁS PROPENSAS A SUFRIR ACCIDENTES DE TRÁNSITO

Rafael F. Feria Torres^a y Jorge A. Timaná Rojas^b

RESUMEN.

La frecuencia y la razón de accidentes en una localización particular, son variables aleatorias cuyo verdadero valor no se puede predecir con absoluta certeza, esto origina que el proceso de identificación de las localizaciones más propensas a sufrir accidentes de tránsito (APL; por sus siglas en inglés) esté sujeto a cierta incertidumbre.

Normalmente la identificación de los APL se realiza con técnicas estadísticas simples, tales como el método del intervalo de confianza y control de calidad. Sin embargo, debido a la limitación que estos métodos presentan ante fenómenos estadísticos propios de variables aleatorias, es conveniente la utilización de metodologías más elaboradas como el análisis Bayesiano que si considera los fenómenos estadísticos de regresión a la media y los efectos no relacionados.

Aun cuando el análisis Bayesiano es un método estadísticamente superior respecto de otros métodos, su uso en el análisis de accidentes de tránsito no ha sido tan difundido como en la medicina. La razón principal radica no sólo en la propia complejidad del método, sino en la dificultad de organismos locales de contar con un soporte técnico adecuado.

El presente trabajo tiene como objetivo presentar un ejemplo práctico de aplicación del método Bayesiano en la identificación los APL en la ciudad de Piura.

ABSTRACT

The frequency of and the reason for accidents in a particular location are random variables, variables whose true value cannot be predicted with absolute certainty. This means that the process of identification of the locations most prone to suffer traffic accidents (APL - accident-prone locations) is subject to uncertainty.

Normally the identification of the APL is made with simple statistical techniques, such as the method of confidence intervals and quality control. Nevertheless, due to the limitations of these methods, the use of more elaborated methodologies, like Bayesian analysis, is advised. This type of analysis takes into account the statistical phenomena of regression to the mean and the unrelated effects.

Even though Bayesian analysis is a statistically superior method, its use in the field of engineering, such as analysis of traffic accidents, is not as widespread as the use of this methodology in the field of medicine. The main reason for this is not only the complexity of the method, but also that it requires often unavailable technical support.

The objective of this paper is to present a practical example of the application of the Bayesian method in the identification of the accident-prone locations (APL) in the city of Piura

1 Antecedentes.

Los métodos estadísticos del intervalo de confianza y el control de calidad, tradicionalmente usados para la identificación de localizaciones riesgosas o propensas a sufrir accidentes de tránsito (APL; Accident Prone Location), presentan limitaciones frente a fenómenos estadísticos propios de variables aleatorias como la regresión a la media.

Un tipo de análisis que si toma en cuenta fenómenos estadísticos propios de variables aleatorias es el método de Bayes, el cual considera al número de accidentes esperado en una localización como una variable aleatoria por si misma, la cual fluctúa sobre algún valor medio no conocido.

Aunque el método bayesiano es considerado como un método estadísticamente superior a los métodos del intervalo de confianza y de control de calidad, aun no es tan ampliamente usado por las diferentes agencias que realizan análisis de accidentes, debido a su complejidad. Además, de la imposibilidad de contar con el soporte técnico adecuado.

El presente trabajo tiene como objetivo presentar la ventajas del método estadístico de Bayes frente a los métodos estadísticos tradicionales. Además de mostrar un ejemplo práctico de aplicación del método Bayesiano en la identificación de las APL en la ciudad de Piura.

Los métodos estadísticos tradicionales.

Tradicionalmente las agencias especializadas en la identificación de los APL han utilizados dos métodos o técnicas estadísticas conocidas como:

Intervalo de Confianza.

La técnica del Intervalo de Confianza, considera que el número o frecuencia accidentes observada (AF) [1], así como la razón o tasa de accidentes observada (AR) [1] tienen una distribución normal.

Un punto es considerado como un APL sí:

$$N_i > X + K * s \quad (\text{ec. 1})$$

Donde "N_i" es la frecuencia o razón de accidentes

^a Programa Académico de Ingeniera Civil de la Universidad de Piura.

^b Programa Académico de Ingeniera Civil de la Universidad de Piura.

observada en una localización dada, "X" es la frecuencia o razón promedio medida de un determinado conjunto de datos, "s" es la desviación estándar y "K" es obtenido de una función de distribución normal..

La credibilidad de esta técnica estadística ha sido cuestionada por ser muy sensible a la variación de la media y desviación estándar de un conjunto de datos de frecuencia o razón, además de no considerar la especial naturaleza aleatoria de la ocurrencia de accidentes.

Control de Calidad.

El Método del control de calidad considera que el frecuencia y la razón de accidentes que ocurren en una determinada localización durante un determinado tiempo puede ser aproximados por una distribución de Poisson. Este fundamento es ampliamente aceptado entre investigadores de protección de tráfico y ha sido investigado muchas veces, consiguiendo ser sustentado con una basta cantidad de evidencia empírica. (Oppe 1982 [2]; 1992 [3]).

Nordon (1956) [4] obtuvo satisfactoriamente valores aproximados del número y razón de accidentes críticos usando las siguientes fórmulas:

$$AFc = a + k\sqrt{a} + 0.5 \quad (\text{ec. 2})$$

$$ARc = \lambda + k + \sqrt{\lambda/m} + 0.5 * m \quad (\text{ec. 3})$$

Donde k es una constante relacionada con la probabilidad P:

Las localizaciones son consideradas como APL si los valores de frecuencia o razón de accidentes observados son mayores que los AFc ó ARc (Valores críticos) calculados con las ecuaciones 8 y 9, siempre y cuando la desviación de su media esperada no puede ser razonablemente atribuida a la fluctuación aleatoria de la ocurrencia de accidentes.

1.1. Método de Bayes.

La técnica del Control de Calidad de la Razón presenta una importante limitación en el análisis variables aleatorias como los accidentes de tránsito Nordon (1956) [4], al no tomar en cuenta el efecto de fenómenos estadísticos como la Regresión a la Media (RTM.).

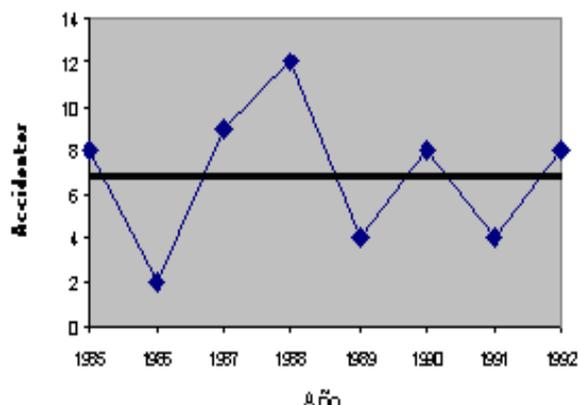


Figura N° 01. Efecto de regresión a la Media.

El fenómeno de regresión a la media se refiere a la tendencia en la cual eventos extremadamente grandes son seguidos por valores mucho menores o viceversa.

La figura N° 01 ilustra el fenómeno de regresión a la media, al considerar una localización hipotética donde los promedios de accidentes registrados en los años 1986 y 1988 varían considerablemente respecto de una frecuencia anual de accidentes promedio de 7 accidentes por año, buscando siempre regresar al promedio en el año siguiente.

El tipo de análisis que si toma en cuenta el fenómeno de regresión a la media es el Análisis de Bayesiano.

El teorema de Bayes puede ser matemáticamente definido como:

Si un cierto parámetro ϕ tiene una distribución previa $P(\phi)$ y para un valor específico ϕ , la probabilidad de generar x observaciones esta dada por $P(x/\phi)$. La posterior distribución $P(\phi/x)$ de ϕ representa la determinación de la anterior dado que la observación es dada por:

$$P(\phi/x) = \frac{P(x/\phi) * P(\phi)}{\sum P(x/\phi) * P(\phi)} \quad (\text{ec. 4})$$

Típicamente, la distribución observada podría ser representada por una distribución Poisson o Binomial y una distribución a priori gamma o beta (Calvín, 1990 [6]).

Aunque el método bayesiano es considera estadísticamente como un método superior respecto al método de control de calidad, este no es usado tan ampliamente debido a su complejidad.

A continuación se desarrollara en forma detallada la metodología de Bayes y su aplicación en los análisis para la identificación de localizaciones riesgosas o APL.

1.2. Desarrollo de la Metodología de Bayes.

EL análisis Bayesiano difiere significativamente del análisis estadístico clásico de accidentes. El motivo del uso del análisis Bayesiano es el deseo de tratar la razón actual de accidentes de una localización, como una variable aleatoria y el uso de una combinación de las características regionales de accidentes y la historia de accidentes de una localización para luego determinar la probabilidad de que la localización sea un APL. De esta manera se mejora la forma de utilización de la información aprovechable en la identificación de las intersecciones más propensas a sufrir accidentes de tránsito.

El método de Bayes de puede dividir en dos pasos. En el primer paso, podemos tomar en cuenta la historia de accidentes a través de un conjunto de intersecciones (todas las localizaciones con características similares definidas dentro una región.) El resultado de este paso es una gruesa estimación de la distribución de probabilidad de la razón de accidentes a través de una región. Luego se puede usar esta distribución regional y la historia de accidentes de una determinada localización para obtener una estimación refinada de la distribución de probabilidad que es asociada con la razón de accidentes en una localización particular. Naturalmente esta estimación refinada se obtiene para todas las localizaciones

definidas dentro de la región, por lo tanto dos localizaciones con una idéntica historia de accidentes tendrán una misma estimación refinada. Con esta estimación refinada uno puede evaluar la probabilidad de que alguna localización pueda ser riesgosa.

Para formalmente definir el proceso de identificación del método Bayesiano, se requiere la siguiente notación:

- λ_{0i} = Razón de accidentes en una localización i (notar que λ_{0i} es considerada como una variable aleatoria)
- N_i = Número de accidentes registrados en la localización i durante el tiempo en cuestión.
- V_i = Número de vehículos que pasan a través de la localización i durante el tiempo en cuestión
- $f_i(\lambda | N_i, V_i)$ = Función densidad de probabilidad asociada con la razón accidentes en una localización i , dada por las observaciones N_i y V_i
- $f_R(\lambda)$ = Función densidad de probabilidad asociada con la razón de accidentes a través de la región.

De esta manera, $f_R(\lambda)$ representa una gruesa estimación de la distribución de probabilidad de la razón de accidentes a través de la región, y $f_i(\lambda | N_i, V_i)$ representa la refinada estimación de la distribución de probabilidad en la localización i como previamente se discutió. Sin embargo la función de distribución acumulada asociada con la razón de accidentes, λ_i , esta dada por la ecuación:

$$P\{\lambda_0 \leq \lambda\} = \int_0^{\lambda} f_i(\lambda | N_i, V_i) d\lambda \quad (\text{ec. 5})$$

El desarrollo de este análisis esta basado en las siguientes suposiciones, las cuales son similares a las dadas por Morín [7], Nordenet, además de Hauer, Persaud, y Glauz

A1. En una localización dada, cuando la razón de accidentes es conocida (si $\lambda_{0i} \leq \lambda$) el actual número de accidentes sigue una distribución de Poisson con un valor esperado λV_i , que es:

$$P\{N_i = n | \lambda_i = \lambda, V_i\} = \frac{(\lambda V_i)^n}{n!} * e^{-\lambda V_i} \quad (\text{ec. 6})$$

A2. La distribución de probabilidad de la razón regional de accidentes $f_R(\lambda)$ es una distribución Gamma.

La primera suposición indica el motivo por el cual la actual razón de accidentes es explícitamente tratada como una variable aleatoria, la distribución condicional del número de accidentes (razón de accidentes dada) es una distribución de Poisson. La segunda suposición implica que:

$$F_R\{\lambda\} = \frac{\beta^\alpha * \lambda^{\alpha-1} * e^{-\beta\lambda}}{\Gamma(\alpha)} \quad (\text{ec. 7})$$

Para unos α y β , de esta manera, el primer paso se asocia con el análisis bayesiano, debido a que la determinación de $f_R(\lambda)$, es necesaria para determinar los valores de α y β . Teniendo en cuenta que existen diferentes métodos para obtener los valores α y β . El más comúnmente usado para estimar los valores α y β es el método de los **Momentos Estimados (MME)** [8], donde α y β son elegidos para que la media y la varianza

asociadas con la distribución gamma sean iguales a la media y la varianza del ejemplo. Esto es, permitir que “ X ” es la media del ejemplo de las razones observadas, “ s^2 ” la varianza del ejemplo de las razones observadas, y “ m ” el número de localizaciones que existen en la región. Luego.

$$\lambda_p = \frac{1}{m} * \sum_{i=1}^m \frac{N_i}{V_i} \quad (\text{ec. 8})$$

$$s^2 = \frac{1}{m-1} * \sum_{i=1}^m \left(\frac{N_i}{V_i} - \lambda_p \right)^2 \quad (\text{ec. 9})$$

Usando el método de los Momentos Estimados, uno selecciona un α y β tal que $\lambda_p = \alpha/\beta$ y $s^2 = \alpha/\beta^2$, o lo equivalente, $\beta = u/s^2$ y $\alpha = \beta u$

Una vez que los valores α y β han sido determinados, el primer paso ha sido completado. En el segundo paso, la razón de accidentes observada en cada localización es usada en combinación con la gruesa estimación de la distribución de probabilidad regional para obtener una función densidad de probabilidad específica en cada localización, $f_i(\lambda | N_i, V_i)$. Esta función densidad es obtenida usando el teorema Bayes. Esto es:

$$f_i(\lambda | N_i, V_i) = \alpha f_i(N_i | \lambda, V_i) f_R(\lambda) \quad (\text{ec. 10})$$

Dentro del mecanismo de trabajo del Análisis Bayesiano, es bien conocido que bajo las suposiciones A1 y A2, el resultado es una distribución $f_i(\lambda | N_i, V_i)$. La cual es una distribución gamma. Sin embargo los parámetros asociados con esta distribución, α_i y β_i , son fácilmente obtenidos de la original elección de α y β y los datos observados, N_i y V_i de la siguiente manera:

$$\alpha_i = \alpha + N_i \quad (\text{ec. 11})$$

$$\beta_i = \beta + V_i \quad (\text{ec. 12})$$

De esta manera la función probabilidad asociada con la razón de accidentes en una localización i (λ_i). Esta dada por:

$$f_i(\lambda | N_i, V_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} * \lambda^{\alpha_i-1} * e^{-\beta_i * \lambda} \quad (\text{ec. 13})$$

Notar que N_i y V_i incrementan, los parámetros de lugares específicos (α_i y β_i) los cuales. podrían ser fácilmente determinados mediante los datos observados (N_i y V_i), y el uso de una computadora.

Con el uso de las funciones de probabilidad la identificación de los APL es más fácil, esto quiere decir que si “ λ ” es un valor límite en una “aceptable” razón de accidentes, luego se quiere identificar una localización como APL si la probabilidad es tal que “ λ_i ” excede “ λ ”. Esto es, si

$$P(\lambda_i > \lambda | N_i, V_i) > \delta \quad (\text{ec. 14})$$

Donde δ es un nivel de tolerancia, este valor de tolerancia debe ser definido en función del grado del grado de precisión que el investigador requiera.

1.2.1. Aplicación de Método de Bayes.

En la práctica existen dos comunes criterios para la identificación de localizaciones riesgosas Los cuales se definen de la siguiente manera:

Criterio 1 (C1).

Siendo λo_i la razón de accidentes observada de una localización i , y λp la media de las razones de accidentes observadas en toda la región, la localización es considerada una APL si la probabilidad de que λo_i exceda a λp , es mayor que δ .

Es necesario tener en cuenta que el Método de Bayes trata a la razón de accidentes en una determinada región como una variable aleatoria y obtiene una refinada estimación de esta distribución de probabilidad, tal que, si

$$P(\lambda o_i > \lambda p / N_i, V_i) > \delta \quad (\text{ec. 15}).$$

Se puede decir que una localización es un APL. En consecuencia, la identificación de localizaciones riesgosas usando el criterio C1 envuelve la computación de:

$$P(\lambda_i > \lambda / N_i, V_i) = 1 - P(\lambda o_i \leq \lambda p) \quad (\text{ec. 16}).$$

$$P(\lambda_i > \lambda / N_i, V_i) = 1 - \int_0^{\lambda p} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} * \lambda^{\alpha_i-1} * e^{-\beta_i * \lambda} \partial \lambda \quad (\text{ec. 17}).$$

Si el valor computado es mayor que δ , la localización es definida como un APL.

Otra forma de expresar este criterio 1 es:

$$P = 1 - \int_0^{\lambda p} \frac{\beta_i^{\alpha_i + \lambda_{cr} * V_i}}{\Gamma(\alpha_i)} * \lambda^{(\alpha_i + \lambda_{cr} * V_i) - 1} * e^{(-\beta_i * \lambda)} \partial \lambda \quad (\text{ec. 18}).$$

Donde:

λ_{cr} = tasa crítica de accidentes en cada intersección

De esta forma se puede calcular la razón crítica de cada localización (λ_{cr}), el cual es el valor de razón, para el cual la probabilidad de que la razón observada λo_i sea mayor a λ es igual a δ (0.90, 0.95 o 0.99). Este método permite hacer una tabla de valores λo_i y λ_{cr} con la cual se puede elaborar un grafico como comparación.

Criterio 2 (C2)

Siendo " λo_i " la razón de accidentes observada en una localización " i ", y " λr " la razón de accidentes regional observada en toda la región, la localización " i " es considerada un APL si la probabilidad de que " λo_i " exceda " λr " es mayor a δ .

Donde la razón de accidentes regional observada es definida como:

$$\lambda_r = \sum \frac{N_i}{V_i} \quad (\text{ec. 19}).$$

Lo que significa que, bajo este criterio una localización es considerada un APL cuando:

$$P(\lambda o_i > \lambda r / N_i, V_i) > \delta \quad (\text{ec. 20}).$$

Igual a:

$$P = 1 - \int_0^{\lambda r} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} * \lambda^{(\alpha_i-1)} * e^{(-\beta_i * \lambda)} \partial \lambda$$

Si el valor computado es mayor que δ , la localización es considerada como riesgosa o APL.

1.2.2. Ranking para priorizar el estudio de las localizaciones riesgosas o APL .

Una vez ubicadas las localizaciones más propensas a sufrir accidentes de tránsito, es necesario definir un ranking de peligrosidad. Esto permitirá establecer un orden de prioridad. El cual será necesario para definir que localizaciones deben ser estudiadas y analizadas en primera instancia, con el fin de establecer sus deficiencias y las medidas correctivas que son necesarias para atenuar la ocurrencia de accidentes en estos puntos.

Existen dos criterios para definir un ranking de riesgo de ocurrencia de accidentes, los cuales se pueden aplicar tanto al Método de Control de Calidad como al Método de Bayes, estos se definen de la siguiente manera:

- **Criterio 1 (AR/ARc>1).**

Donde:

AR = Es la razón de accidentes observada en cada localización, esta notación es valida tanto para el ranking por método de Control de Calidad como para el Método de Bayes.

ARc = Es la razón crítica de accidentes en cada intersección en el caso del Método de Control de Calidad y como " λ_{cr} " para el Método de Bayes

- **Criterio 2 ((AR-ARc)*m)**

Donde:

AR = Es la razón de accidentes observada en cada intersección, esta notación es valida tanto para el ranking por método de Control de Calidad como para el Método de Bayes.

ARc = Es la razón crítica de accidentes en cada intersección en el caso del Método de Control de Calidad y como " λ_{cr} " para el Método de Bayes.

$$m = \frac{TMDA * t * 365}{1'000,000} \quad (\text{ec. 21}).$$

donde:

TMDA = Promedio de volumen de tráfico medio diario anual.

t = tiempo de observación.

2 Aplicación del Método de Bayes en la ciudad de Piura.

En el año 2000 se elaboró el trabajo de tesis [9], donde se explicaban las diferentes técnicas utilizadas en el análisis de accidentes de tránsito, especialmente el método de Bayes. La tesis buscaba identificar las intersecciones más riesgosas o propensas a sufrir accidentes (APL) de Tránsito del distrito de Piura.

Los datos utilizados para el análisis correspondieron a 1,152 accidentes registrados en 279 intersecciones del distrito de Piura durante los años 1998 y 1999.

El análisis de accidentes encontró que en la ciudad de Piura se producen entre 2 y 3 accidentes por día de los cuales 1 produce daños personales. El 59% de los accidentes de tránsito registrados en intersecciones no produjeron ningún daño personal, el 40% produjeron al menos un herido y el 1% dejaron muertos. También se registró que 612 personas sufrieron lesiones, de las cuales 35% eran conductores, 31% pasajeros y 34% peatones. Así mismo, de las 14 personas muertas, el 43% eran conductores, el 21% pasajeros y el 36% eran peatones.

El estudio permitió observar entre otras cosas que el control y registro de los accidentes de tránsito eran deficiente ya que la policía enfoca el problema, únicamente desde el punto de vista policial y judicial, lo que genera que mucha información necesaria para un buen análisis de accidentes de tránsito no sea recopilada de la manera adecuada.

Aplicando el método del control de calidad, junto al criterio de la razón de accidentes y con una confiabilidad de 95% se identificaron con riesgosas o APL 25 intersecciones. Mientras que aplicando el método de Bayes se identificaron como riesgosas o APL 20 intersecciones.

bien ambas curvas tienen una tendencia logarítmica, podemos observar que la curva del método de control de calidad permite que más valores de razón observada (25), la superen; si lo comparamos con la curva obtenida con el método de Bayes (20). La consideración antes mencionada tiene una relación directa con el método de regresión a la media, el cual contribuye para que la curva del método de Bayes sea más conservadora que el método del control de calidad.

La aplicación de los dos criterios para el ranqueo de intersecciones riesgosas definieron a la intersección Av. Sánchez Cerro – Gullman, como la intersección más propensa a sufrir accidentes de tránsito en la ciudad de Piura.

Del diagnóstico de los resultados de la aplicación del Método de Bayes, considerado por algunos especialistas como el método más adecuado para este tipo de análisis, se verificó que 13 de las 20 intersecciones semaforizadas, es decir el 65%, eran intersecciones de alto riesgo de accidentes de tránsito. (Ver Figura N° 03)

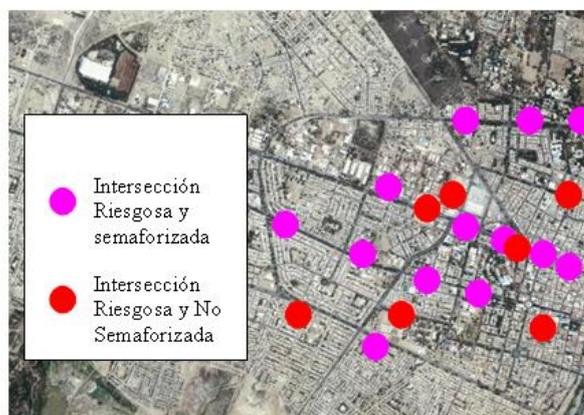


Figura N° 03. Intersecciones Riesgosas y Semaforizadas.

3 Conclusiones.

- Los métodos estadísticos tradicionales no toman en cuenta fenómenos estadísticos como la regresión a la media y los efectos no relacionados.
- El método de Bayes es el más adecuado para la identificación de localizaciones con alto riesgo de ocurrencia de accidentes de tránsito. Es incluso más conservador que el Método de Control de Calidad.
- A pesar que el método de Bayes no es muy utilizado por las agencias de análisis de accidentes debido a su complejidad, Si es posible su aplicación con el adecuado soporte técnico, por lo tanto son las Universidades y los profesionales especializados los llamados a difundir su uso y aplicación.
- La instalación de semáforo nos garantiza la reducción de accidentes.

Localización de Intersecciones Riesgosas

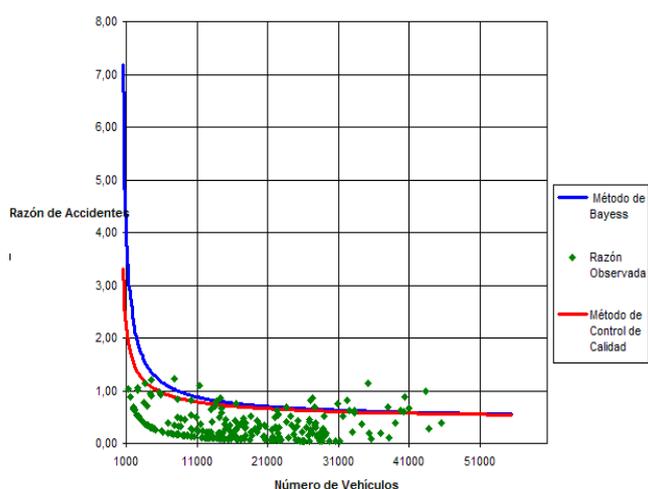


Figura N° 02: Ciudad de Piura, Unidad Metropolitana.

La figura N° 02 nos muestra la diferencia existente entre los métodos de control de calidad y método de Bayes. Si

REFERENCIAS BIBLIOGRÁFICAS

1. Zeeger, C.V, and Deen, R.C (1977). "Identification of hazardous locations on city streets", traffic Quarterly 31, pp 549-570.
2. Oppe, S., (1982). "Detection and analysis of black spots with even small accident figures". R82-15, SWOV, Leidschendam, The Netherlands.
3. Oppe, S., (1992). "A comparison of some statistical techniques for road accident analysis". Accident Analysis and Prevention, 24(4), pp. 397-423.
4. Nordon, M., Orlandsky, J., and Jacobs, H., (1956) "Applications of statistical quality control techniques to analysis of highway accident data", Bulletin 117, Highway Research Board, National Research Council, Washington, D.C., pp. 17-31.
5. Canavos, J (1988), "Probabilidad y Estadística", Capítulo 8.5, pp. 285-287.
6. Calvin, T.M., (1990). "Bayesian analysis", in Handbook of Statistical Methods for Engineers and Scientists, Wadsworth H, Editor, Mcgraw-Hill
7. Morin, D.A.(1967) "Application of statistical concepts to accident data". Highway Research Record 187, HRB, National Research Council, Washington, D.C., pp 72-79.
8. Hagle, J.L., Witkowsky, J.M. ."Bayesian Identification of Hazardous Location". Transportation Research Record 11185, Transportation Research Board, National Research Council, Washington, D.C., pp 24-36.
9. Ing. Rafael F. Feria Torres UDEP (2001). Tesis para obtener el Título de Ingeniero Civil en la Universidad de Piura, "Técnicas de análisis de accidentes de tránsito con aplicación a la ciudad de Piura".

E-mail: rferia@udep.edu.pe y jtimana@udep.edu.pe