

Probabilidad de resultar herido leve o morir en un accidente de tránsito en distintos distritos de Costa Rica, según el tipo de accidente, clasificación de la carretera y el clima, en el periodo de 2015 a 2019

Valeria Alvarado Madrigal¹, Emerson Arce Rodríguez¹, Nicole Rivera Morales¹

valeria.alvaradomadrigal@ucr.ac.cr, emerson.arce@ucr.ac.cr,

nicole.riveramorales@ucr.ac.cr

RESUMEN

Con el crecimiento demográfico de la población, crece también la necesidad de movilización en todos los medios de transporte y, por consiguiente, es posible preguntarse si este fenómeno ha implicado entonces un incremento en los percances que se dan en las vías terrestres de transporte. En este estudio, se exploran los factores tipo de accidente en los niveles atropello a persona, colisión de motocicleta y colisión de automóviles, clasificación de la carretera donde sus niveles son rural cantonal, rural nacional, urbana cantonal y urbana nacional, y el año de ocurrencia del accidente, otro factor contemplado es el estado del clima durante el accidente, sus niveles son Buen tiempo, Lluvia, Neblina y Lluvia con neblina, con el objetivo de analizar la propensión de muertes con respecto a los heridos leves por accidentes de tránsito en diferentes distritos del país del 2015 al 2019. Mediante un modelo logístico longitudinal se encuentra que el aumento de la probabilidad de morir con respecto a la probabilidad de que sean solo heridos leves, es la misma para cada distrito. Luego, a pesar de existir un efecto de la clasificación de la carretera, no se detectaron diferencias en las propensiones para cada uno de los tipos de carreteras según año. El tipo de accidente sí tiene efecto en la probabilidad de resultar herido leve o muerto en un accidente de tránsito. La probabilidad de morir es mayor en el tipo de accidente atropello (por cualquier tipo de vehículo), que en un accidente de moto o carro. Finalmente, la probabilidad de morir en una colisión de moto es mayor que en una colisión de carro.

PALABRAS CLAVES: modelo logístico, longitudinal, propensión, potencia de la prueba.

INTRODUCCIÓN

Con el crecimiento demográfico de la población, crece también la necesidad de movilización en todos los medios de transporte y, por consiguiente, es posible preguntarse si este fenómeno ha implicado entonces un incremento en los percances que se dan en las vías terrestres de transporte. Según Ramírez (2013), los accidentes de tránsito son ahora considerados un problema de salud pública, y sobre el cual interesa proponer estrategias de prevención que involucren, por ejemplo, educar a la sociedad en torno a las posibles causas relacionadas con este tipo de accidentes. Ramírez (2013) hace mención del uso del celular al conducir, la ingesta de alcohol y la falta del cinturón de seguridad, a manera de ejemplificar algunos factores coadyuvantes en los accidentes de tránsito, que corresponden a características directamente relacionadas con las víctimas. Así como también, existen factores o variables que

¹ Estudiantes de bachillerato en Estadística de la Universidad de Costa Rica

no se relacionan directamente con las personas, sino que corresponden a factores externos o ambientales, también involucrados en estos sucesos, como el tipo de ruta sobre la cual se transita, el tipo de clima al momento del accidente, e incluso si este se da en una zona urbana o rural.

El análisis de estos factores externos implica la construcción de un panorama amplio en cuanto a las condiciones físicas de las calles y espacios públicos para transitar. Tanto la infraestructura de las carreteras como la regulación del tránsito en Costa Rica se han ido deteriorando al pasar de los años y no se ha visto las mejoras correspondientes. Según los datos del INEC (2019), para el 2018 en el país circulaban 959 342 automóviles y 312 948 motos. La gran cantidad de vehículos y motos en circulación por las carreteras de Costa Rica puede implicar un aumento en el número de accidentes. En el periodo del 2015 al 2019, la cantidad de accidentes de tránsito se presenta de manera fluctuante, rondando entre 75 051 y 84 278 (COSEVI, 2021).

La evaluación de variables externas o ambientales como posibles factores importantes en la propensión de accidentes a través del tiempo, permite abrir una discusión sobre la importancia que tiene transitar en un ambiente seguro y sustenta el propósito de este trabajo, que se propone como objetivo general: analizar la propensión de muertes con respecto a los heridos leves por accidentes de tránsito en diferentes distritos del país del 2015 al 2019. Y como objetivos específicos, en primer lugar, identificar si la clasificación de la carretera donde ocurrió el accidente tiene efecto sobre la gravedad de la víctima en el accidente. En segundo lugar, se busca determinar cuál tipo de accidente (heridos leves o muertos) tiene mayor propensión de morir del año 2015 al 2019.

Este estudio contempla como hipótesis que la propensión de accidentes con personas muertas es mayor en la clasificación urbana-nacional y urbana-cantonal que en las carreteras que se encuentran en la clasificación rural-nacional y rural-cantonal, considerando que la circulación vial en las zonas urbanas es mayor. Como segunda hipótesis se plantea que la propensión de que las personas mueran en un accidente de tipo atropello es mayor que en un accidente de tipo colisión de carro o moto.

METODOLOGÍA

Costa Rica cuenta con el Consejo de Seguridad Vial (COSEVI), la cual es una institución encargada de administrar el dinero para invertir en la seguridad vial de las carreteras del país, esta institución cuenta con datos abiertos para información al público. El presente trabajo se basó en los datos del *Consolidado de accidentes con víctimas* del COSEVI, que según la opinión del experto Deiby Solano Cambroner, estadístico del COSEVI, contiene datos importantes para el estudio. Este archivo ofrece información referente a los accidentes registrados en el periodo del 2015 al 2019 en Costa Rica, y se extrae una muestra de 2513 accidentes.

Las variables que se contemplaron en este estudio son el *resultado del accidente* como variable respuesta, la cual hace referencia a si en el accidente hubo sólo heridos leves o muertos. Asimismo, se tomó como factor de diseño el *tipo de accidente*, la cual es una variable que categoriza los accidentes en *atropello a persona*, *colisión de motocicleta* y *colisión de automóviles*. También, como un segundo factor de diseño se contó con la *clasificación* de la carretera donde sus niveles son *rural cantonal*, *rural nacional*, *urbana cantonal* y *urbana nacional*, cabe destacar que se extraen 5 distritos por cada uno de estos niveles, como elemento aleatorio. Además, se tomó en cuenta otro factor importante el cual es el *estado del clima* durante el accidente, sus niveles son *Buen tiempo*, *Lluvia*, *Neblina* y el tipo de accidente *Lluvia*

con neblina. Y como el interés es ver una tendencia en el tiempo, el año se trabaja como una variable continua.

El diseño de esta investigación es longitudinal, pues los datos se han ido recolectando a través del tiempo y además la variable respuesta consta de dos niveles, de forma que sigue una distribución Bernoulli, con parámetro $\pi_{A,jkli}$. Dicho esto, se construye un modelo lineal generalizado saturado inicial de tipo logístico longitudinal, que contempla todos los factores y la interacción entre ellos:

$$\log\left(\frac{\pi_{A,jkli}}{1 - \pi_{A,jkli}}\right) = \beta_0 + \beta_1 A + \alpha_j + \tau_k + \gamma_l + \beta_j^1 A + \beta_k^1 A + \beta_l^1 A + (\alpha\tau)_{jk} + (\alpha\gamma)_{jl} + (\tau\gamma)_{kl} + (\alpha\tau\gamma)_{jkl} + b_{0,i} + b_{1,i}A$$

Donde:

β_1 = coeficiente asociado al año

α_j = j-ésimo efecto del tipo de accidente, donde $\alpha_1 = 0$

τ_k = k-ésimo efecto del clima, donde $\tau_1 = 0$

γ_l = l-ésimo efecto de la clasificación, donde $\gamma_1 = 0$

β_j^1 = interacción entre el año y el accidente

β_k^1 = interacción entre el año y el clima

β_l^1 = interacción entre el año y la clasificación

$(\alpha\tau)_{jk}$ = interacción entre el tipo de accidente y el clima

$(\alpha\gamma)_{jl}$ = interacción entre el tipo de accidente y la clasificación

$(\tau\gamma)_{kl}$ = interacción entre el clima y la clasificación

$(\alpha\tau\gamma)_{jkl}$ = interacción triple entre el tipo de accidente, clima y la clasificación

$b_{0,i} \sim N(0, \sigma_0^2)$ = interceptos aleatorios

$b_{1,i} \sim N(0, \sigma_1^2)$ = pendientes aleatorias

Se evalúan los efectos aleatorios, interacciones y efectos marginales sobre la variable respuesta mediante la prueba de razón de verosimilitudes (LRT, sigas en inglés) con la distribución chi cuadrado, se obtiene el siguiente modelo teórico final que contempla los factores de diseño, el factor clima, la interacción entre año y clasificación y los interceptos aleatorios.

$$\log\left(\frac{\pi_{A,jki}}{1 - \pi_{A,jkli}}\right) = \beta_0 + \beta_1 A + \alpha_j + \tau_k + \gamma_l + \beta_l^1 A + b_{0,i}$$

Por último, se realiza una simulación en R para obtener la probabilidad de detectar diferencias entre las razones de propensión entre cada tratamiento del año y la clasificación, si realmente las hay. Así como también, con esta simulación se busca obtener un tamaño de muestra que proporcione una potencia aceptable. Para esto, se plantea una diferencia relevante de $\log(2)$, es decir 0.7, entre los tratamientos de año y clasificación y se ejecuta iteradamente 1000 veces.

Para obtener los resultados se utiliza el software estadístico R (R core team, 2021), en la versión 4.1.1. Las librerías que se utilizan en este estudio son: *lattice* (Sarkar, D., 2008), *car* (Fox, J. y Weisberg, S., 2019), *lme4* (Bates et al., 2015)

RESULTADOS

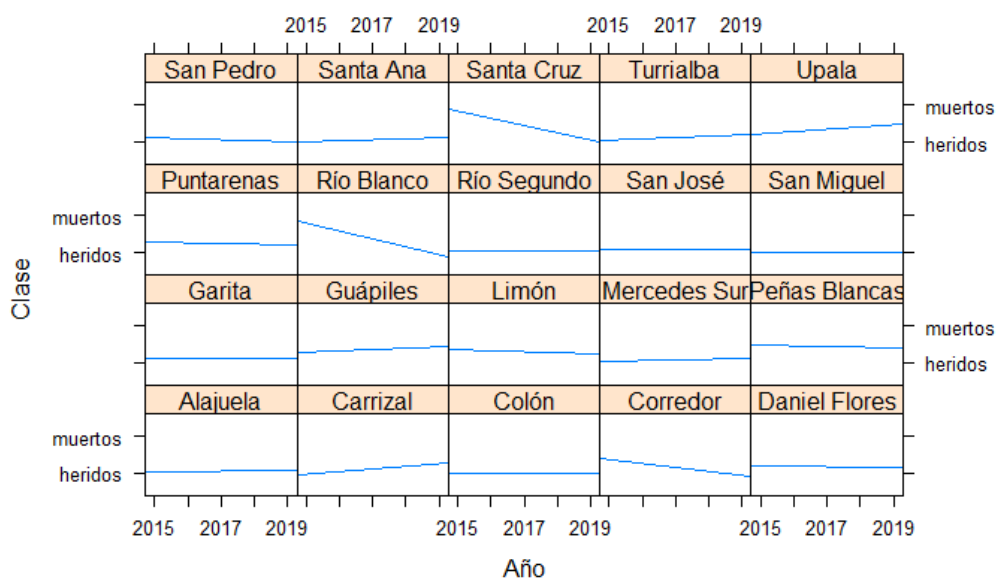
Con el propósito de conocer el comportamiento inicial de los datos, es importante contextualizarse con algunas características y tendencias de los factores que contempla este trabajo, previo a la construcción del modelo. Para esto, se construye una tabla resumen (ver tabla 1 en Anexos) con la distribución porcentual de los heridos leves y las muertes, según clasificación de la carretera, tipo de accidente y año de ocurrencia. Se obtiene que los niveles de los factores y el año con mayor porcentaje de heridos leves son la clasificación de carretera urbana-cantonal (59.0%), colisión con moto (52.5%) y el año 2016 (21.5%). Mientras que los niveles de los factores y el año con mayor porcentaje de muertes son la clasificación de carretera urbano-nacional (37.8%), la colisión de moto (52.1%) y el año 2018 (21.6%).

Seguidamente se realiza la prueba respectiva para descartar la posibilidad de correlación entre las pendientes e interceptos de cada distrito, y se obtiene como resultado una probabilidad de 0.5 que no constituye evidencia estadística suficiente para rechazar la hipótesis nula de no correlación, por lo que se asume que los interceptos y las pendientes por cada distrito son independientes entre sí. Esto equivale a decir, que el ritmo al que incrementa la probabilidad de resultar herido leve o muerto no depende de la probabilidad en el primer año del periodo estudiado.

Luego se evalúa si las pendientes entre los distritos son iguales, para esto, se grafican las pendientes (gráfico 1), en el cual se observa que la mayoría de las pendientes de los distritos son, por lo general, similares. Y a manera de complemento, se ejecuta la prueba de razón de verosimilitud, obteniendo probabilidad asociada de aproximadamente 1, por lo que se podría asumir que las pendientes de los distritos son iguales, es decir, el aumento de la probabilidad de morir con respecto a la probabilidad de que sean solo heridos, es la misma para cada distrito.

Gráfico 1.

Tendencia de la probabilidad de muertes o heridos, por distrito en Costa Rica, del 2015 al 2019

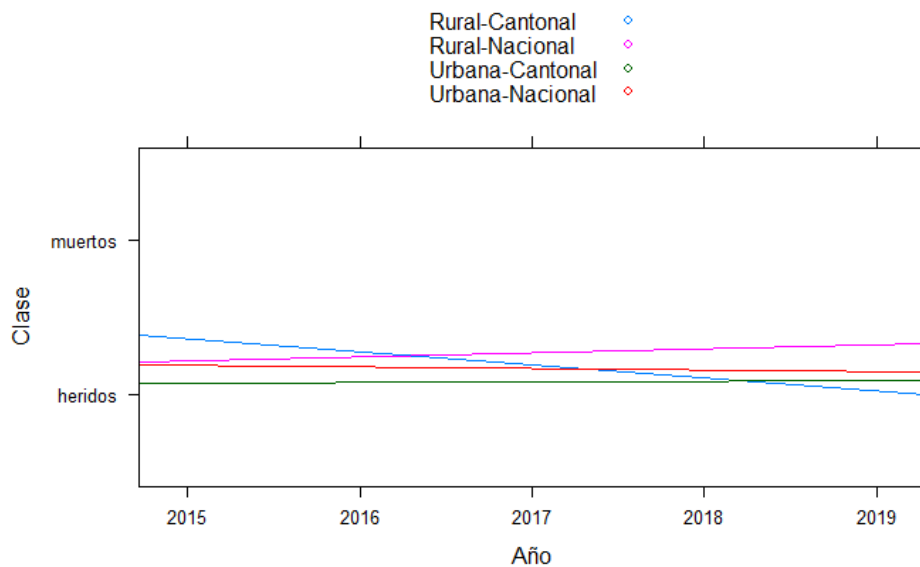


Por otro lado, como no se encontraron interacciones del año, clasificación y el tipo de accidente con el clima, se analiza el efecto marginal de este factor y se obtiene una probabilidad asociada de 0.9, lo que indica que el clima no tiene efecto sobre la variable respuesta. Seguidamente, al evaluar la existencia de interacción entre las variables independientes,

realizando la prueba correspondiente, se obtiene que la única interacción existente se da entre el año y la clasificación, con una probabilidad asociada de 0.04, por lo que se asume que la probabilidad de resultar muerto o herido leve en un accidente varía conforme avanza el tiempo, para al menos una de las clasificaciones de la carretera. Ante esto, se visualiza y verifica esta interacción también por medio del gráfico 2, el cual revela que la clasificación rural-nacional va aumentando conforme avanzan los años, pero contrariamente, la rural-cantonal va disminuyendo.

Gráfico 2.

Interacción entre el año y la clasificación



Posteriormente, como se prueba la existencia de interacción entre la clasificación de la carretera y el año, se realizan las comparaciones entre los niveles de la clasificación de la carretera para los años 2015, 2017 y 2019. Sin embargo, no se detectan diferencias significativas para ninguna de las comparaciones, lo cual implica que no resulta útil cuantificar las diferencias entre las propensiones.

Por otro lado, realizando la prueba respectiva se encontró que sí hay un efecto por parte del tipo de accidente sobre la probabilidad de morir, dado que se obtuvo una probabilidad asociada de casi 0. Lo anterior implica que sí hay diferencias en las propensiones por cada tipo de accidente, por lo que se procede a realizar contrastes para verificar cuáles diferencias son significativas y seguidamente cuantificarlas.

En primera instancia, todas las comparaciones, observables en la tabla 1, resultaron significativas ya que las probabilidades asociadas obtenidas son muy bajas ($p < 0.01$). Por lo tanto, con un 95% confianza global se encuentra que la propensión de morir en un atropello es 85.9% mayor que la propensión de morir en una colisión de carro y un 8.5% mayor que la propensión de morir en una colisión de moto, mientras que, la propensión de morir en una colisión de moto es 29.7% mayor que la propensión de morir en una colisión de carro.

Tabla 1.

Comparaciones de propensiones por tipo de accidente

Comparaciones	Probabilidad	OR	Cota inferior
Atropello-Moto	< 0.01*	1.5	1.1
Atropello-Carro	< 0.01*	2.8	1.9
Moto-Carro	< 0.01*	1.8	1.3

Nota: *Comparaciones por corrección de 0.05/3

Por último, por medio de la simulación realizada, es posible observar que existe una potencia suficiente para detectar diferencias entre la razón de propensiones entre cada tratamiento del año y la clasificación. Es decir, si realmente existen diferencias de al menos $\log(2)$ en la propensión de morir con respecto a resultar herido leve, para cada uno de los tipos de carretera, conforme avanza el tiempo y se repite el experimento muchas veces, en el 99% de los casos se concluiría que sí hay diferencias. Cabe destacar que, aunque haya suficiente potencia para detectar diferencias entre los niveles de las clasificaciones de la carretera por año, estas no se encuentran porque probablemente no existen.

CONCLUSIONES

A manera de síntesis, es posible concluir al respecto del factor clima, que resultó mínimamente relevante a lo largo de este estudio, pues no se encontró que tuviese un efecto significativo, pero se conservó por formar parte del diseño inicial. Ahora bien, se concluye con respecto al primer objetivo que sí se encontró que la clasificación de la carretera tuviese efecto sobre la probabilidad de resultar herido leve o muerto, pero resulta muy interesante que no fue posible detectar diferencias estadísticamente significativas entre las diferentes clasificaciones de la carretera, de acuerdo con diferentes años. Es valioso, sin embargo, que, a pesar de no detectar dichas diferencias, la potencia obtenida para este caso es bastante alta y refleja la alta posibilidad de detectar dichas diferencias siempre y cuando estas sean reales. En suma, al comparar los resultados con la primera hipótesis, no es posible determinar en cuál de las clasificaciones es mayor la propensión de muerte.

Seguidamente, en respuesta al segundo objetivo planteado, se logró detectar que hay un efecto por parte del tipo de accidente sobre la probabilidad de morir o de ser herido leve. Y de la mano con este hallazgo, se logró también encontrar diferencias entre todas las comparaciones, y se logra probar la segunda hipótesis, ya que la propensión de morir en un atropello es mayor que en los otros dos niveles del tipo de accidente.

Una de las limitaciones presentadas, consiste en que el marco de datos utilizado no contempla variables de características individuales de las personas afectadas en los accidentes, como, por ejemplo, su estado de ebriedad o la cantidad de personas involucradas por cada accidente.

Se destaca también que, como una posible recomendación a futuro de este trabajo, se podría realizar el experimento variando los distritos o aumentando la cantidad de estos. Inclusive, se podría ampliar el periodo de estudio y de esta forma verificar si bajo este escenario se logran comprobar o no, los resultados obtenidos en esta investigación. En la misma línea de las recomendaciones, se insta a profundizar el estudio también en la metodología implementada en el cálculo de la potencia de la prueba. Pues al tratarse de la simulación sujeta a las condiciones

que establezca el programador, se abre la posibilidad de plantear más interrogantes de relevancia para estudios sobre la seguridad en carreteras, como ¿qué cambiaría si se simula el remuestreo entre los mismos distritos, o, si se remuestrea más distritos, con características diferentes entre sí?

BIBLIOGRAFÍA

- Bates, D., Maechler, M., Ben Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01
- COSEVI. (2021). *Cantidad de accidentes de tránsito según categoría*. <https://datosabiertos.csv.go.cr/dataviews/234875/ACCID-DEL-ANO-2015/>
- INEC. (2019). *Vehículos automotores en circulación, según estilo 2017-2018*. <https://www.inec.cr/documento/social-22-costa-rica-vehiculos-automotores-en-circulacion-segun-estilo-2017-2018>
- John Fox and Sanford Weisberg. (2019). *An {R} Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Ramírez Muñoz, J. (2013). Accidentes de tránsito terrestre. *Revista Medicina Legal de Costa Rica - Edición Virtual*. 30(2). <art09v30n2.pdf> (scielo.sa.cr)
- Sarkar, Deepayan. (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5

ANEXOS

Tabla 1.

Distribución porcentual de heridos y muertos según factores de interés

Factores	Estado	
	Heridos	Muertos
Total	100	100
Clasificación de la carretera		
Rural-Cantonal	1.8	2.4
Rural-Nacional	11.1	26.8
Urbana-Cantonal	59.0	32.9
Urbana-Nacional	28.1	37.8
Tipo de accidente		
Atropello	16.3	22.6
Colisión de carro	31.2	25.3
Colisión de moto	52.5	52.1
Año		
2015	19.8	20.1
2016	21.5	21.3
2017	20.1	19.5
2018	19.1	21.6
2019	19.5	17.4

Anexo 2.

Simulación para el cálculo de la potencia

Código	Explicación
<pre>library(lme4) mod = glmer(Clase~Clima+Accidente+Año1*Clasificación+(1 Distrito),family = "binomial", accidentes) summary(mod)</pre>	<p>Librería lme4, se usa para realizar modelos utilizando factores con efectos aleatorios.</p> <p>mod es el modelo final, se utiliza para obtener la desviación estándar de los distritos, la cual se obtiene de la función summary.</p>
<pre>n=table(accidentes\$Distrito) b0=rnorm(length(n),0,0.7484) b0i=NULL for (i in 1:20) { b0i=c(b0i, rep(b0[i], n[i])) }</pre>	<p>El n calcula la cantidad de observaciones por cada distrito.</p> <p>El b0 genera 20 interceptos aleatorios con una distribución normal, uno por cada distrito.</p> <p>El b0i es un vector vacío que va a almacenar los interceptos para cada distrito.</p> <p>En el ciclo for almacena en el vector b0i todos los interceptos por cada observación de cada distrito.</p>

<pre> k = 1000 p = NULL for(i in 1:k){ X=model.matrix(mod8) b=numeric(13) b[11]=log(2) pl=X%*%b+b0i pi=exp(pl)/(1+exp(pl)) y = rbinom(n=nrow(accidentes), size=1, prob = pi) moda = glmer(y~Clima+Accidente+Año1*Clasificación+(1 Distrito), family="binomial", accidentes) p[i]=drop1(moda, test = "Chisq")[4,4] } mean(p<0.05) </pre>	<p>k es una variable que contiene la cantidad de iteraciones por realizar en la simulación.</p> <p>p es el vector vacío donde se van a almacenar las probabilidades.</p> <p>X es la matriz estructura del modelo final.</p> <p>b son los coeficientes del modelo, para este caso se pone 0 en todos los coeficientes excepto en el de la interacción a comparar.</p> <p>En la posición (11,12 o 13) de b, se pone log(2) (delta) para obtener la potencia de las comparaciones por cada una de las interacciones con respecto al tratamiento de referencia.</p> <p>pl representa la parte lineal del modelo, se utiliza para obtener la probabilidad pi para generar las nuevas variables respuesta mediante la función rbinom.</p> <p>moda consiste en el modelo donde se van a obtener las probabilidades, con la variable respuesta generada aleatoriamente.</p> <p>drop1 es la función donde se extrae la probabilidad de la interacción. Para luego ser almacenada en el vector p.</p> <p>Por último, se promedian todas las probabilidades, almacenadas en el vector p, menores a 0.05 y se obtiene la potencia</p>
---	---